Dissertation
submitted to the
Combined Faculty of Mathematics, Engineering and Natural Sciences
of Heidelberg University, Germany
for the degree of
Doctor of Natural Sciences

Put forward by
Giulia Frau
Born in: Cagliari, Italy
Oral Examination: 26/07/2023

Search for the lepton flavour violating decay $\tau^+ \rightarrow \mu^+\mu^-\mu^+$ with the LHCb experiment

Referees: Prof. Dr. Stephanie Hansmann-Menzemer
Prof. Dr. Klaus Reygers

Kurzfassung

Diese Dissertation präsentiert eine Suche nach dem Leptonflavour verletzenden Zerfall $\tau^+ \to \mu^+\mu^-\mu^+$ mit Hilfe einer Analyse von Daten, die vom LHCb-Experiment bei Proton-Proton-Kollisionen mit einer Schwerpunktsenergie von 13 TeV aufgezeichnet wurden. Der Datensatz entspricht einer integrierten Luminosität von 5.4 fb^{-1}. Bisher ist keine Leptonflavour-Verletzung im geladenen Sektor beobachtet worden. Bei der Berücksichtigung von Neutrinooszillationen im Standardmodell (SM) werden solche Prozesse mit nicht messbaren Zerfallsraten erwartet, weshalb die Beobachtung von $\tau^+ \to \mu^+\mu^-\mu^+$ ein klares Zeichen für neue Physik wäre. Selbst falls der Zerfall nicht beobachtet wird, würde eine Verbesserung der Obergrenze des Verzweigungsverhältnisses von $\tau^+ \to \mu^+\mu^-\mu^+$ uns erlauben, neue Theorien von Physik jenseits des SM einzuschränken, welche ein Verzweigungsverhältnis messbar mit der Sensitivität der heutigen experimentellen Einrichtungen vorhersagen.

Die primäre Herausforderung der Analyse ist die Identifikation und Unterdrückung der Untergrundquellen. Diese stammen von zufälligen Kombinationen von Spuren, die fälschlicherweise dem τ Zerfall zugeordnet wurden, und von der Fehlidentifikation von Hadronen ($\pi^\pm$, $K^\pm$), die aus D^+ oder D_s Zerfällen stammen. Die Analyse ist so gestaltet, dass zunächst die Signalregion entfernt wird, um Voreingenommenheit in der Optimierung der Signalauswahl zu vermeiden. Die Unterdrückung des Untergrundes wird durch multivariate Klassifikationsmodelle erreicht, die darauf trainiert werden, zwischen Signal- und Untergrundereignissen zu unterscheiden. Die Anzahl von τ Leptonen, die in der Akzeptanz von LHCb produziert werden, wird durch die Anzahl der $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ Zerfälle normalisiert. Eine erwartete Obergrenze von

$$\mathcal{B}(\tau^+ \to \mu^+\mu^-\mu^+) < 1.8(2.1) \times 10^{-8}$$

bei einem Konfidenzniveau von 90%(95%) wird mit Hilfe der CL_s-Methode bestimmt.

Abstract

This thesis presents a search for the lepton flavour violating decay $\tau^+ \to \mu^+\mu^-\mu^+$ using data collected by the LHCb experiment in proton-proton collisions at a centre-of-mass energy of 13 TeV, corresponding to an integrated luminosity of 5.4 fb^{-1}. As of today, no lepton flavour violation has been observed in decays of charged leptons. When accounting for neutrino oscillations in the Standard Model, such processes are expected at undetectable decay rates, thus the observation of the $\tau^+ \to \mu^+\mu^-\mu^+$ decay would be a clear sign of new physics. Even if the decay is not observed, improving the upper limit on the branching fraction of $\tau^+ \to \mu^+\mu^-\mu^+$ would allow us to constrain theories of physics beyond the Standard Model that predict a decay rate within the sensitivity of the current experimental facilities.

The main challenge of the analysis is the identification and rejection of the background sources, originating from accidental combinations of tracks wrongly associated with a τ decay and from the misidentification of hadrons ($\pi^\pm$, $K^\pm$) coming from D^+ or D_s decays. The analysis is designed by removing the signal region in data in order to avoid biases in the optimisation of the signal selection. The background suppression is achieved by means of multivariate classification models trained to distinguish between signal and background events. The number of τ leptons produced in the LHCb acceptance is normalised to the yield observed in the $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ channel. An expected upper limit of

$$\mathcal{B}(\tau^+ \to \mu^+\mu^-\mu^+) < 1.8(2.1) \times 10^{-8}$$

is found at 90%(95%) confidence level using the CL_s method.

Contents

Introduction

The Standard Model (SM) of particle physics represents our present best knowledge of the basic building blocks of the universe and of their fundamental interactions. Despite being incomplete -it does not include gravity-, the SM has successfully predicted the outcome of a plethora of experiments with remarkable accuracy. However, there is a series of phenomena that the SM leaves unexplained: the excess of matter over antimatter in the universe, the presence of dark matter and dark energy and the non-zero neutrino mass.

In the last decades, the theory community has developed alternative theories and extensions of the SM that contemplate such phenomena, and with the advent of high-energy experiments, a considerable effort has been dedicated to the search for new physics signatures. The observation of neutrino oscillations in 1998 by the Super-Kamiokande experiment [1] has shown that lepton flavour can be violated in the neutral sector, opening the possibility for new physics scenarios. Unlike neutrinos, transitions among charged leptons with different lepton family numbers are forbidden in the SM. Even in an extended version of the SM including neutrino oscillations, such processes are predicted at undetectable decay rates ($< \mathcal{O}(10^{-50})$). Theories of physics beyond the SM enhance the branching fraction of processes involving charged lepton flavour violation to the sensitivity level of the present experimental facilities. The search for such decays performed at particle colliders represents one way to test, and possibly constrain, these new physics models.

This thesis presents a search for the lepton flavour violating decay $\tau^+ \to \mu^+\mu^-\mu^+$[1], carried out analysing data collected in 2016-2018 (Run 2) at the LHCb experiment. The LHCb experiment is located at the Large Hadron Collider (CERN) and designed for the study of beauty and charm decays. The aim of the analysis is the first observation of the decay or, in case of the observation of no events, the improvement of the limit on its branching fraction. The current most stringent limit on the branching fraction of the decay, namely 2.1×10^{-8}, was set by the Belle experiment at 90% confidence level (CL) [2]. Despite the abundant production of τ leptons at LHCb, the identification of $\tau^+ \to \mu^+\mu^-\mu^+$ decays in hadronic collisions is actually more challenging compared to electron-positron experiments because of the higher background contamination.

The LHCb collaboration already provided an upper limit for the branching fraction of $\tau^+ \to \mu^+\mu^-\mu^+$ of 4.6×10^{-8} at 90% CL after analysing data collected during Run 1[2] [3]. The higher luminosity and production cross section of Run 2 leads to an expected improvement on the Run 1 limit of $\sim$40%. With the analysis presented in this thesis, an expected upper limit of 1.8×10^{-8} is determined at 90% CL. If the observed limit will correspond to the expected, this analysis would set the world's most stringent limit on the branching fraction of the $\tau^+ \to \mu^+\mu^-\mu^+$ decay.

---

[1] Charge conjugation is implied henceforth for all decay modes throughout this thesis.

[2] First LHCb data-taking period (2011-2012).

The thesis is structured as follows. In Chapter 1, an introduction to lepton flavour violation and to the $\tau^+ \to \mu^+\mu^-\mu^+$ decay is presented. In Chapter 2 the experimental setup and the data acquisition system of the LHCb experiment are described with a focus on the sub-detectors playing a crucial role in the reconstruction and selection of $\tau^+ \to \mu^+\mu^-\mu^+$ decays. The analysis strategy is defined in Chapter 3, while in Chapter 4 the recorded and simulated data sets are described. One of the novelties of this analysis is the inclusion of events where only two of the tracks in the final state are identified as muons. In Chapter 5, the preliminary cuts applied to select signal events are presented in detail. Chapter 6 is dedicated to the calibration of the simulated sample performed exploiting the $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ [1] channel. The strategy followed to discriminate between signal and background events is depicted in Chapter 7. A detailed study of the possible background sources and the estimated size of their contribution is presented in Chapter 8. The efficiencies of the applied selection are evaluated on simulated events and reported in Chapter 9. In Chapter 10, an overview of the systematic uncertainties affecting the final measurement is outlined. The expected limit on the branching fraction is finally presented in Chapter 11. The analysis and the results are summarised in Chapter 12, while supplementary material is given in the Appendixes A, B, C, D.

This thesis project has been carried out within the LHCb collaboration, an organisation composed of 1599 members from 96 institutes spread over 21 countries. The analysed data have been collected and processed by means of software architectures and analysis tools developed by current and former collaboration members. In particular, the analysis here presented relies on the fruitful collaboration of LHCb members affiliated with the University of Milano Bicocca, the University of Rome Tor Vergata and the Heidelberg University. The author is responsible for the preparation of the simulated samples presented in Chapter 4, for the application of the signal selection outlined in Chapter 5 and for the calibration of the $\tau^+ \to \mu^+\mu^-\mu^+$ mass shape illustrated in Chapter 6. The study of the sample made of events where only two of the final-state tracks are identified as muons has been entirely performed by the author, including the strategy adopted to suppress the background contamination, outlined in Chapter 7, and the estimate of its residual contribution, presented in Chapter 8. The evaluation of the normalisation factor discussed in Chapter 9, the effect of the systematic uncertainties in Chapter 10 and the final estimate of the expected upper limit presented in Chapter 11 have been carried out by the author. The optimisation of the trigger selection in Chapter 5, the calibration of the kinematic features in the simulated sample in Chapter 6 and the training of the classifiers in the sample with three tracks identified as muons discussed in Section 7.2, is the result of the work of a former PhD student of the University of Milano Bicocca. The calibration of the particle identification variables in Chapter 6 and the estimate of the residual contamination of background originating from misidentified tracks presented in Chapter 8.2 have been carried out by a post-doctoral researcher of the University of Milano Bicocca. No part of this study would have been possible without the fruitful discussion within the members of the analysis working group and the constant support of Dr. Flavio Archilli.

Throughout the doctorate program, the author has contributed to the software upgrade of the LHCb experiment for Run 3[3] by working on the development of the trigger lines dedicated to the evaluation of the track reconstruction efficiency of the

---

[3] Third LHCb data-taking period started in 2022.

detector. A detailed description of the method used to evaluate the track reconstruction efficiency at LHCb is given in Section 9.5.1. For two years the author served as liaison for "Tracking and Alignment" for the Rare Decays Working Group of the LHCb collaboration. In addition, the author has been involved in the commissioning of the Scintillating Fibre (SciFi) Tracker [4] for the LHCb upgrade and filled the role of data manager during data-taking shifts.

Chapter 1

Theoretical and experimental background

1.1 The Standard Model

The Standard Model (SM) [5–7] of particle physics is a theoretical framework based on symmetry principles representing our present best knowledge of the elementary particles and their mutual interactions.

The matter particles included in the SM, also known as *fermions* because of their spin 1/2, are classified into six leptons and six quarks depending on the force ruling their interactions. Leptons include three charged particles, namely the electron (e), muon (μ) and tau (τ), and three neutral particles labelled electron-neutrino (ν_e), muon-neutrino (ν_μ) and tau-neutrino (ν_τ). Among the quarks, three have electric charge of $+2/3$, namely the up (u), charm (c) and bottom (b) quarks, while the down (d), strange (s) and top (t) quarks have electric charge of $-1/3$.

Each particle has its corresponding anti-particle and they all carry the charge of the weak interaction, known as weak isospin, and can therefore all undergo the weak interaction. Having no electric charge, neutrinos are the only *fermions* not subject to the electromagnetic interaction. The main property of quarks instead is that, unlike leptons, they carry color charge and can therefore interact also via the strong interaction.

As shown in Table 1.1, *fermions* are grouped into three *generations*, or *families*, each containing two leptons (one charged and one neutrino) and two quarks. With the exception of neutrinos, whose masses have not been determined yet, each member of a higher *generation* have greater mass than the corresponding *fermion* of the previous generation. All leptons carry the so-called lepton charge, or lepton number, defined in the SM as $L = N_{leptons} - N_{anti-leptons}$. The lepton number is one of the two accidental symmetries of the SM and therefore expected to be conserved in

TABLE 1.1: The matter elementary particles of the SM divided into leptons and quarks. The masses are taken from the Particle Data Group (PDG) [8].

	Leptons	**Q**	**Mass [MeV/c^2]**	**Quarks**	**Q**	**Mass [GeV/c^2]**
First generation	electron (e^-)	-1	0.51 ± 0.00	down (d)	-1/3	$0.0047^{+0.5}_{-0.2}$
	neutrino (ν_e)	0	$< 10^{-6}$	up (u)	+2/3	$0.0022^{+0.5}_{-0.3}$
Second generation	muon (μ^-)	-1	105.67 ± 0.00	strange (s)	-1/3	$0.093^{+0.009}_{-0.003}$
	neutrino (ν_μ)	0	$< 10^{-6}$	charm (c)	+2/3	1.27 ± 0.02
Third generation	tau (τ^-)	-1	1776.86 ± 0.12	top (t)	-1/3	172.69 ± 0.30
	neutrino (ν_τ)	0	$< 10^{-6}$	bottom (b)	+2/3	$4.18^{+0.03}_{-0.02}$

TABLE 1.2: Lepton number and lepton family numbers for leptons and their corresponding anti-particles.

Leptons	L	L_e	L_μ	L_τ
e^- (e^+)	+1 (-1)	+1 (-1)	0	0
ν_e ($\bar{\nu}_e$)	+1 (-1)	+1 (-1)	0	0
μ^- (μ^+)	+1 (-1)	0	+1 (-1)	0
ν_μ ($\bar{\nu}_\mu$)	+1 (-1)	0	+1 (-1)	0
τ^- (τ^+)	+1 (-1)	0	0	+1 (-1)
ν_τ ($\bar{\nu}_\tau$)	+1 (-1)	0	0	+1 (-1)

interactions [9]. In addition, leptons belonging to the same *family* share the same lepton family number, also conserved in interactions, named L_e for the electron and the electron-neutrino, L_μ for the muon and the muon-neutrino, and L_τ for the tau and the tau-neutrino (see Table 1.2).

Mathematically speaking, the SM is a gauge invariant quantum field theory based on the symmetry group $SU(3) \times SU(2) \times U(1)$, with $SU(3)$ being the colour group for the strong interaction and $SU(2) \times U(1)$ for the electroweak interaction. The strong interaction between quarks is predicted by the theory of quantum chromodynamics (QCD) to be mediated via a four-vector gluon field characterised by the propagation of eight gluons, each with a different colour charge. Because of the nature of the QCD interactions, particles with colour-charge, such as quarks and gluons, can never be observed as free particles, but are always confined to color-neutral bound states called *hadrons*. Depending on their quark composition, *hadrons* are divided into *mesons*, made of one quark and one anti-quark, and *baryons*, made of three quarks. The second and last accidental symmetry of the SM is the so-called baryon number, defined as $\frac{1}{3}(N_{quarks} - N_{anti-quarks})$ and therefore assuming value $+1(-1)$ for *baryons* (anti-*baryons*) and 0 for *mesons*.

Intended as a unified description of the electromagnetic and weak forces, the electroweak interaction is mediated via the boson fields W_i $(i = 1, 2, 3)$ and B.

The spontaneous symmetry breaking of the electroweak symmetry $SU(2) \times U(1)$ to $U(1)_{em}$ via the Higgs mechanism generates the observed physical particles associated to these fields, namely the $W^\pm$ and Z^0 vector bosons, and the photon γ, mediators of the weak and electromagnetic interaction, respectively. The neutral particles Z^0 and γ are related to the boson fields B and W_3 via the rotation

$$\begin{pmatrix} \gamma \\ Z^0 \end{pmatrix} = \begin{pmatrix} cos\theta_W & sin\theta_W \\ -sin\theta_W & cos\theta_W \end{pmatrix} \begin{pmatrix} B \\ W_3 \end{pmatrix}, \tag{1.1}$$

where θ is the weak mixing angle, also known as Weinberg angle. The mass of the Z^0 particle can be written in terms of the mass of the $W^\pm$ particles as

$$m_Z = \frac{m_W}{cos\theta_W}. \tag{1.2}$$

The massive charged bosons $W^\pm$ are instead a combination of the W_1 and W_2 bosons:

$$W^\pm = \frac{1}{\sqrt{2}}(W_1 \mp iW_2). \tag{1.3}$$

The electric charge Q is defined as a linear combination of the third component of the weak isospin (I_3) and the weak hypercharge (Y_W) of the fermions by the Gell-Mann–Nishijima formula

$$Q = I_3 + \frac{Y_W}{2} \,. \tag{1.4}$$

The SM has been successfully tested by various experiments over the last decades; nevertheless, there are many important questions which remain unexplained. First of all, the SM does not include gravity which makes it an incomplete theory. The charge conjugation and CP violation observed so far and predicted by the SM is insufficient to explain the excess of matter over antimatter in our universe. None of the particles included in the SM represent a good candidate for dark matter; even dark energy does not find an explanation in the SM. In addition, the experimental observation of neutrino oscillations by SNO and other experiments [1] have shown that neutrinos have small but yet non-zero masses as otherwise stated by the SM.

The weak *eigenstates* ν_e, ν_μ and ν_τ are related to the three mass *eigenstates* ν_1, ν_2 and ν_3 by the unitary PMNS (Pontecorvo-Maki-Nakagawa-Sakata) mixing matrix.

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix} \tag{1.5}$$

Neutrino oscillations can only occur if the masses of the ν_1, ν_2 and ν_3 *eigenstates* are not the same and therefore if at least one neutrino has a non-zero mass. The observation of this phenomenon also demonstrates that lepton flavour is not strictly conserved, opening the possibility for an extended model of particle physics.

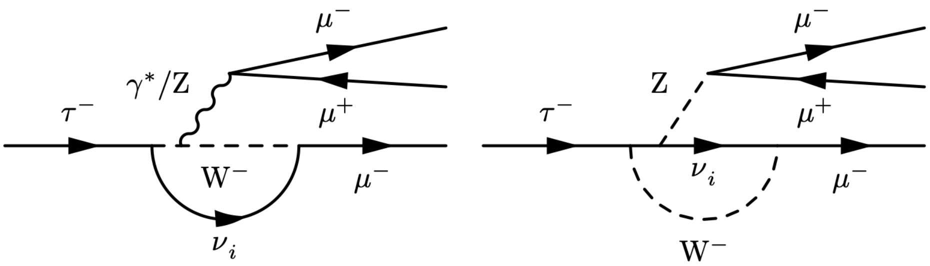
1.2 Lepton Flavour Violation in the charged sector

The mixing of neutrinos leads to lepton flavour violation in processes where no neutrino is involved in the initial and final state. This is referred to as lepton flavour violation in the charged sector (cLFV) and it occurs in processes that conserve the total lepton number L but violate the lepton number associated to the flavour of the decay particles [10]. Depending on which lepton number is violated, cLFV processes can be sorted into three groups:

- $\Delta(L_e - L_\mu) = 2$, involving transitions from the second to the first family;
- $\Delta(L_e - L_\tau) = 2$, involving transitions from the third to the first family;
- $\Delta(L_\mu - L_\tau) = 2$, involving transitions from the third to the second family.

The abundance of muons from cosmic radiation and at dedicated detectors makes some of the processes belonging to the first group the most searched ones. Among them, it is worth to mention the $\mu^+ \to e^+\gamma$, $\mu^- N \to e^- N$ and $\mu^+ \to e^+e^-e^+$ decays. The present best upper limit on the branching fraction of $\mu^+ \to e^+\gamma$ was set in 2016 by the MEG experiment (PSI) at 4.2×10^{-13} @ 90% confidence level (CL) [11]. Back in 1988, the SINDRUM experiment (PSI) published what is still today the best limit on the branching fraction of the $\mu^+ \to e^+e^-e^+$ decay, namely 1.0×10^{-12} @ 90% CL [12]. The Mu3e experiment, to be realised at the PSI and designed for the search for the $\mu^+ \to e^+e^-e^+$ decay, aims at reaching a 10^{-16} sensitivity within its three operational phases [13].

The large mass of the tau lepton allows for many LFV processes in comparison with muon decays. In this thesis I will focus on the search for the $\tau^+ \to \mu^+\mu^-\mu^+$



(a) Photonic penguin **(b) Z penguin**

FIGURE 1.1: Feynman diagrams for the $\tau^- \to \mu^-\mu^+\mu^-$ decay in the Standard Model extended with massive neutrinos allowing for flavour changing neutral current at a very low rate; ν_j indicates a neutrino mass eigenstate. Taken from Ref. [14].

decay, however, in addition to purely leptonic processes, like $\tau^+ \to l^+\gamma$ or $\tau^+ \to l^+l^-l^+$ ($l \in \{e, \mu\}$), there is a variety of decays involving hadrons that can be tested in the search for cLFV, such as $\tau^+ \to l^+\pi^0$, $\tau^+ \to l^+\pi^+\pi^-$.

Despite being theoretically more appealing, τ decays are actually more challenging from the experimental side. There is indeed no "direct" way to produce τ leptons, if not as decay products at proton or electron colliders.

1.2.1 The $\tau^+ \to \mu^+\mu^-\mu^+$ decay

In a massive neutrino scenario, cLFV can occur via one-loop diagrams with a couple of $Wl\nu_l$ ($l = e^-, \mu^-, \tau^-$) vertices with different flavour neutrinos each. Figure 1.1 shows two possible Feynman diagrams contributing to the $\tau^+ \to \mu^+\mu^-\mu^+$ decay in the SM with massive neutrinos allowing for flavour-changing neutral currents. According to what is reported in Ref. [15], due to the unitarity of the mixing matrix, the GIM mechanism [16] produces for these processes a suppression of the form

$$\sigma \propto \left| \sum_i U_{\tau i} U^*_{\mu i} \log \frac{m^2_{\nu_i}}{M^2_W} \right|^2 , \tag{1.6}$$

where m_{ν_i} (with $i = 1, 2, 3$) and M_W are the masses of the three neutrinos and of the W boson, respectively, and U_{li} are the corresponding PMNS matrix elements. This suppression leads to a branching fraction of the order of 10^{-14}. A different prediction is proposed in Ref. [17] where the suppression assumes the form of

$$\sigma \propto \left| \sum_i U_{\tau i} U^*_{\mu i} \frac{m^2_{\nu_i}}{M^2_W} \log \frac{m^2_{\nu_i}}{M^2_W} \right|^2 \tag{1.7}$$

and leads to a branching fraction of the order of 10^{-54}, well below the current and foreseen experimental sensitivity. The assumptions made in Ref. [15] and [17] are both disproved by a study presented in a recent paper [18], where a suppression of the form of

$$\sigma \propto \left| \sum_i U_{\tau i} U^*_{\mu i} \frac{m^2_{\nu_i}}{M^2_W} \right|^2 \tag{1.8}$$

TABLE 1.3: Predictions for the branching fraction of $\tau^+ \to \mu^+\mu^-\mu^+$ in different BSM models.

Model	$\mathcal{B}(\tau^+ \to \mu^+\mu^-\mu^+)$	**Ref.**
Standard Model + neutrino oscillations	10^{-55}	[18]
SM with right-handed heavy Majorana neutrino	$\sim 10^{-10}$	[19]
left-right SUSY	10^{-7}	[20]
SUSY with neutral Higgs	$\lesssim 4 \times 10^{-10}$	[21]
SUSY with Higgs triplet	10^{-7}	[22]
Non universal $Z^{'}$ (technicolor)	10^{-8}	[23]

is instead proposed. The prediction for the branching fraction of the $\tau^+ \to \mu^+\mu^-\mu^+$ decay in a massive neutrino scenario presented in Ref. [18] is 10^{-55}, even smaller than the one predicted in Ref. [17].

On the other hand, there are several theories of physics beyond the Standard Model (BSM) in which the branching fraction of lepton flavour violating decays in the charged sector is predicted to be of the order of $10^{-10} - 10^{-7}$, definitely more in line with the sensitivity of existing experimental facilities. Among these theories it is worth to mention supersymmetry (SUSY) in combination with different types of symmetry breaking, models including heavy neutrinos or models including an extra neutral gauge boson Z'. In Table 1.3, an overview of some of these BSM models is reported, together with the predicted BF for the $\tau^+ \to \mu^+\mu^-\mu^+$ decay.

Depending on whether charged-lepton flavour changing decays are forbidden or allowed at the tree level, these extensions to the SM can be separated in two main classes.

- If forbidden at the tree level, charged-lepton flavor violation can occur at the loop level where, for instance, ν_j or γ, Z (see Figure 1.1) are replaced with new particles, like right-handed heavy Dirac or Majorana neutrinos (see Table 1.3). Another example is the (minimal) supersymmetric extension of the SM including heavy right-handed neutrinos, responsible for the generation of low energy neutrino masses via the see-saw mechanism, which predicts for the $\tau^+ \to \mu^+\mu^-\mu^+$ a branching fraction of the order of 10^{-10}.

- In the second category, the charged lepton flavour violation can occur at the tree level through the exchange of new heavy particles. An example are the Higgs-mediated decays in supersymmetric seesaw models, whose representation in terms of Feynman diagrams is given in Figure 1.2.

1.3 State of the art

To date no LFV in the charged sector has been observed yet. Thus on the one hand, the observation of the $\tau^+ \to \mu^+\mu^-\mu^+$ decay would be a clear hint of new physics, on the other hand, the improvement of the existing limit on its branching fraction would allow to constrain different models of physics BSM.

The LHCb collaboration already provided a first measurement of the branching fraction of $\tau^+ \to \mu^+\mu^-\mu^+$ after analysing data collected in 2011-2012 and corresponding to an integrated luminosity of 3 fb^{-1} [3]. An upper limit of $\mathcal{B}(\tau^+ \to \mu^+\mu^-\mu^+) < 4.6(5.6) \times 10^{-8}$ at 90% (95%) confidence level (CL) was measured. The BaBar collaboration also set a limit on the $\tau^+ \to \mu^+\mu^-\mu^+$ branching fraction of

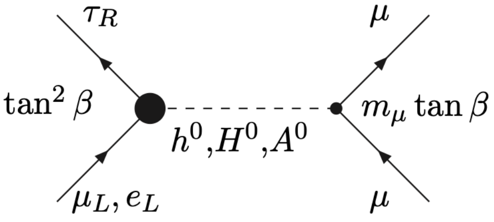


FIGURE 1.2: Dominant Higgs penguin diagrams contributing to $\tau^+ \to \mu^+\mu^-\mu^+$ decays in supersymmetric seesaw models at large $\tan\beta = v_1/v_2$, where v_1 and v_2 are the vacuum expectation values of the exchanged particle. Taken from Ref. [21].

5.3×10^{-8} at 90% CL using 376 fb^{-1} of data collected at the SLAC PEP-II *B-factory* [24].

In 2016, the ATLAS experiment published its sensitivity to the $\tau^+ \to \mu^+\mu^-\mu^+$ decay after analysing 20.3 fb^{-1} of *p-p* collision data at a centre-of-mass energy of 8 TeV. An upper limit on branching fraction was set at $3.76(3.94) \times 10^{-7}$ at 90% (95%) CL [25]. The CMS experiment also performed a search for the $\tau^+ \to \mu^+\mu^-\mu^+$ decay on 33.2 fb^{-1} of data collected at LHC in 2016, which lead to an upper limit on the branching fraction of 8.0×10^{-8} at 90% CL [26].

At present, the world best experimental upper limit on this branching fraction from a single experiment is the one provided by the Belle collaboration who measured an upper limit of 2.1×10^{-8} @ 90% CL using 782 fb^{-1} of data collected at the KEKB asymmetric-energy e^+e^- collider [2]. The best limit from the average of the BaBar, Belle and LHCb results is 1.1×10^{-8} at 90% CL [27].

1.4 Future prospects

The future of the search for the lepton flavour violation in the charged sector will be strongly determined by the High-Luminosity of LHC (HL-LHC) [28]. The peak instantaneous luminosity is expected to reach $7.5 \times 10^{34}\ \mathrm{cm}^{-2}\mathrm{s}^{-1}$ with *p-p* collisions happening at a centre-of-mass energy of $\sqrt{s}$ =14 TeV.

During the HL-LHC, the LHCb experiment will be able to bring the upper limit on the $\tau^+ \to \mu^+\mu^-\mu^+$ decay down to $\mathcal{O}(10^{-9})$ after collecting 300 fb^{-1} of data [29]. The CMS and ATLAS experiments also published their prediction on the expected exclusion limit at 90% CL. In particular, the CMS projection is at 3.7×10^{-9} and at 4.3×10^{-9} in the absence of additional muon coverage [30]. The ATLAS projections instead are 5.4×10^{-9} for τ produced in $W^+ \to \tau^+\nu_\tau$ decays and 1.0×10^{-9} for τ produced in *c*- or *b*-hadron decays, dominated by $D_s^+ \to \tau^+\nu_\tau$ [31].

The creation of the Future Circular Collider (FCC) would also play a crucial role in the search for cLFV. A first simulation study has shown that FCC-ee, the first stage of FCC, would be able to reach a sensitivity of $\mathcal{O}(10^{-10})$ on the branching fraction of the $\tau^+ \to \mu^+\mu^-\mu^+$ decay [32].

All these predictions exceed the limits set by existing BSM theories opening the possibility for further studies.

Chapter 2

The LHCb experiment

In this Chapter a description of the LHCb detector and of its data-acquisition system is presented. Specific attention is given to the sub-detectors playing a crucial role in the reconstruction and identification of the signal decay.

2.1 The Large Hadron Collider

The LHCb detector is placed at the Large Hadron Collider (LHC), the world's largest and highest-energy particle collider [33] designed for the exploration of the Standard Model at TeV energy scale and for the search for potential new physics signatures. As illustrated in Figure 2.1, LHC consists of a 27 km ring located at a depth from 50 to 175 m underground on the Franco-Swiss border near Geneva, Switzerland. The facility has been realised in the tunnel originally built for the Large Electron-Positron (LEP) collider.

Before being injected in the main ring, the beams (protons or ions) are accelerated by a four-level accelerating system in two separate beam-pipes. The first stage is a source and linear accelerator, LINAC2, followed by the Proton Synchroton Booster (PSB) where beams are injected with an energy of 50 MeV and accelerated up to an energy of 1.4 GeV. In the Proton Synchroton (PS) and in the Super Proton Synchroton (SPS) the beams reach an energy of 450 GeV; the final stage takes place in the LHC ring where the two beams are accelerated in parallel until they reach the required

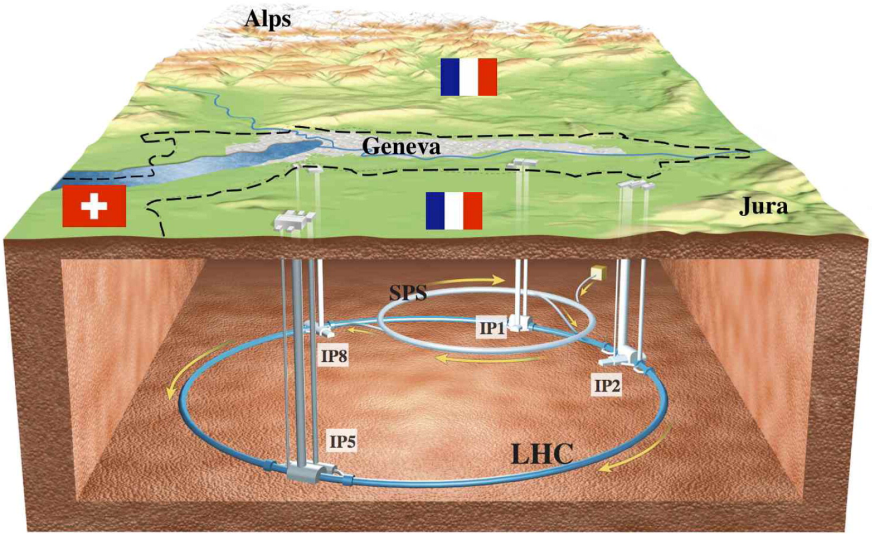


FIGURE 2.1: Illustration of the Large Hadron Collider. Taken from Ref. [33].

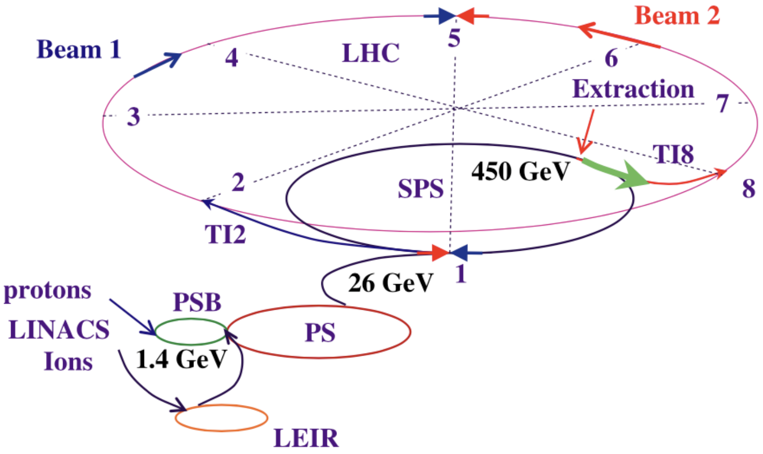


FIGURE 2.2: Schematic view of the LHC with its injector. Taken from Ref. [33].

energy. A schematic overview of the accelerating system of the LHC is displayed in Figure 2.2.

The hadron collisions take place at four experimental insertions: two general purpose detectors, ATLAS [34] and CMS [35], using proton-proton collisions ($\mathcal{L} \geq 10^{34}$ cm^{-2} s^{-1}); one forward spectrometer dedicated to beauty and charm physics, LHCb [36], exploiting proton-proton collisions at medium luminosity ($\mathcal{L} \sim 10^{32}$ cm^{-2} s^{-1}); one experiment dedicated to heavy-ion physics, ALICE [37], operating at lower luminosity ($\mathcal{L} \sim 10^{29}$ cm^{-2} s^{-1}).

In the first operational period of LHC (2010-2013), called Run 1, proton beams reached an energy collision of 8 TeV. During a two-year break, called First Long Shutdown (LS1), LHC has been adapted to enable proton collisions at 13 TeV. The second operational period (2015-2018) or Run 2, was followed by a second Long Shutdown (LS2), where both the accelerator and the experiments have been upgraded in view of the higher collision energy and luminosity of the third operational run (Run 3). Due to the pandemic, Run 3 only started in April 2022 and is expected to continue until 2026 [38].

2.2 The LHCb detector

The LHCb detector is a single arm forward spectrometer covering the pseudorapity[1] range of $1.8 < \eta < 4.9$. The experiment is dedicated to the precision measurements of CP violation and of rare decays involving mesons or baryons containing b or c quarks. As illustrated in Figure 2.3, the coordinates system of the LHCb detector consists of a right-handed Cartesian triplet where the z-axis is aligned with the beam direction, the y-axis is pointing vertically upwards and the x-axis is pointing towards the outside of the LHC ring. Cylindrical polar coordinates (r, θ, z) are also used when needed. The forward geometry of the detector is motivated by the fact that, at high energies, $b\bar{b}$ pairs are produced in the same backward or forward cone; processes of interest are indeed symmetrical with respect to the x-y plane (see Figure 2.4).

---

[1] The pseudorapidity is defined as $\eta \equiv -\log\left[\tan\left(\frac{\theta}{2}\right)\right]$, where θ is the angle between the particle momentum and the positive direction of the beam axis [39].

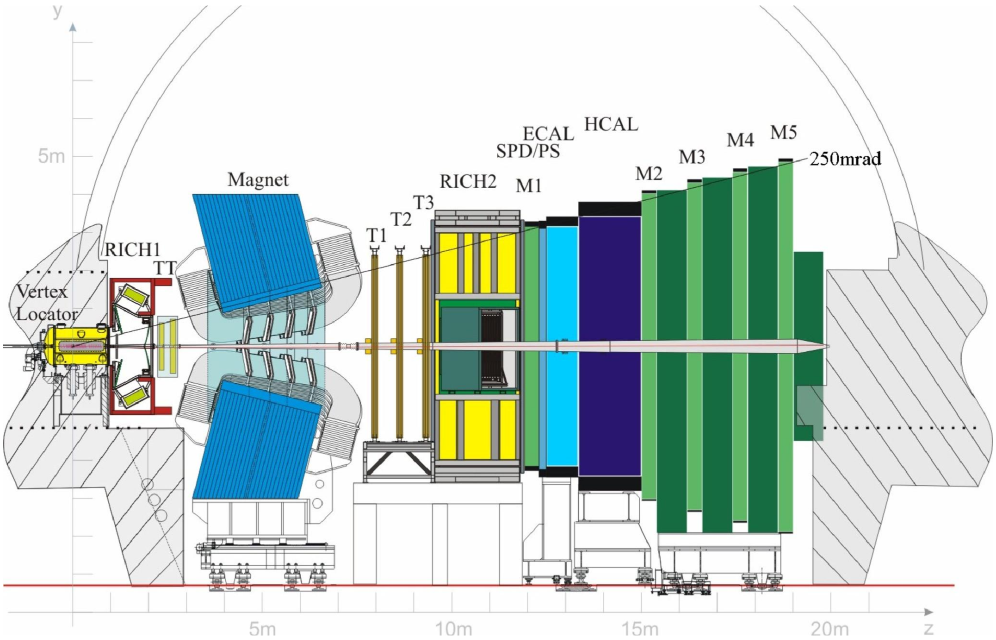


FIGURE 2.3: Illustration of the LHCb detector in the y-z plane. Taken from Ref. [36].

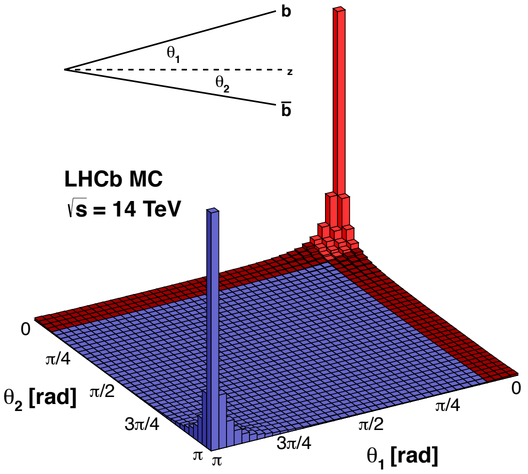


FIGURE 2.4: Production angle of the $b\bar{b}$ pair at the centre-of-mass energy of 14 TeV. 27% of all produced b or $\bar{b}$ quarks are in the LHCb acceptance (red color). In a standard general-purpose detector, covering the range $|\eta| < 2.4$, 49% of b or $\bar{b}$ quarks are produced in its acceptance. Taken from Ref. [40].

As previously mentioned, LHCb runs at a two order of magnitude lower luminosity with respect to ATLAS and CMS (see Figure 2.5). This operational choice has the advantage of having a single proton-proton interaction per bunch crossing, which is easier to analyse than events with multiple interactions, a low detector occupancy and limited radiation damage. At LHCb, the luminosity can be tuned

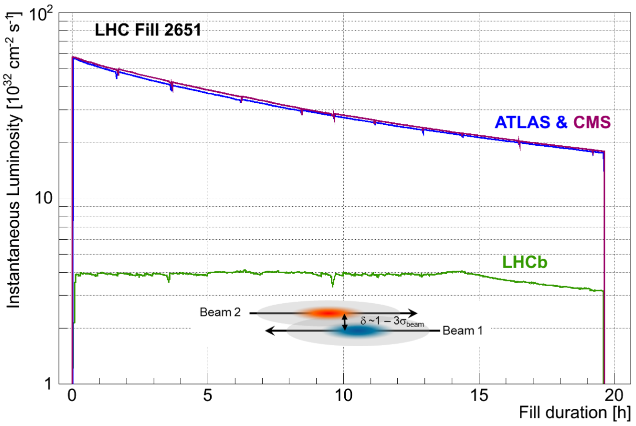


FIGURE 2.5: Development of the instantaneous luminosity for ATLAS, CMS and LHCb during LHC fill 2651. Taken from Ref. [41].

through a system of magnetic lenses that slightly changes the beam focus[2] at its interaction point independently from the other experiments [41].

The physics purpose of LHCb would not be feasible without a good tracking system and excellent particle identification performance. The different sub-detectors are described separately in the following.

2.2.1 Tracking system

An excellent tracking system is essential for providing a good decay time resolution, needed in the study of time dependent effects, such as time dependent CP violation or mixing, and a good invariant mass resolution for the suppression of the combinatorial background. The tracking system of LHCb consists of the VErtex LOcator, surrounding the interaction region, and four tracking stations, namely the Trigger Tracker (TT), placed upstream of the dipole magnet, and T1-T3, placed downstream of the magnet.

Vertex locator

The VErtex LOcator (VELO) [36, 42] is the closest sub-detector to the interaction region. Its task consists in the reconstruction of the primary vertex (PV), *i.e.* the p-p collision vertex, and of the decay vertex of b hadrons. It is a silicon strip detector composed by 42 semicircular tracking stations positioned along the beam direction. Each station is able to measure both the radial and angular coordinates of a particle traversing the sensor. Its resolution is about 10 μm and 42 μm along the directions perpendicular and parallel to the beam, respectively. In addition, the impact parameter of particles emitted with high transverse momentum is reconstructed with a resolution of 20 μm.

The VELO has been designed to minimise the amount of material traversed by a charged particle and, consequently, of the uncertainties caused by multiple scattering. In Figure 2.6, a schematic representation of the VELO and its two possible configurations are shown. The first configuration (Fully Open) is adopted in case of unstable beams in order to avoid the damage of the detectors. The second one (Fully Closed) is realised once the stability is reached and LHCb is ready for data taking.

---

[2] The beam focus at the collision point is commonly refereed to as beta function, β^*.

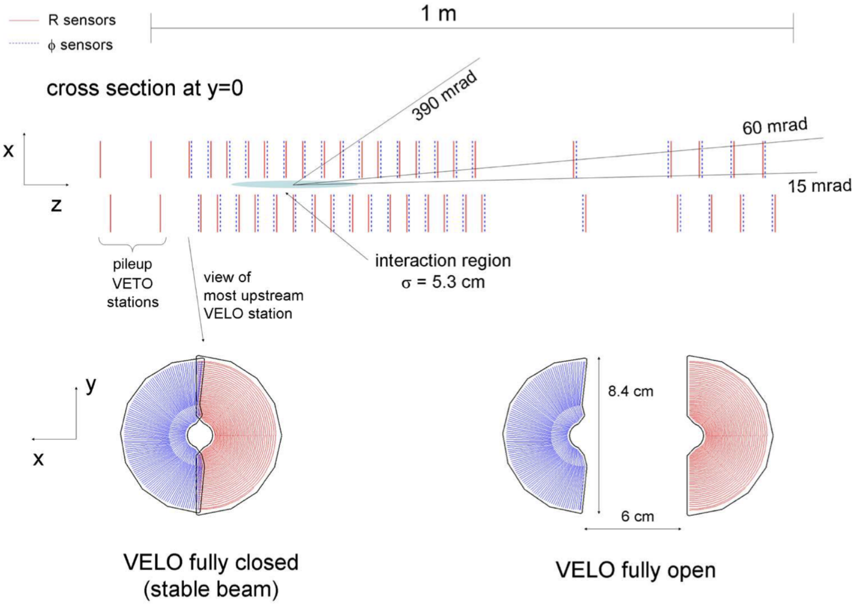


FIGURE 2.6: Cross section of the VELO in the x-z plane at y=0. The front face of the first modules is also illustrated in both the closed and open positions. Taken from Ref. [36].

Dipole magnet

A dipole magnet, placed between RICH1 and the tracking stations T1, T2 and T3, is used to measure the momentum of charged particles. It consists of two conical saddle-shaped coils placed symmetrically to each other and crossed by a nominal current of 5.85 kA. The resulting magnetic field has a magnitude of 4 Tm for tracks of 10 m length and is oriented in such a way that particles are deflected in the x-z plane. The design had to accommodate the requirement of having a less than 2 Tm field inside the RICH detectors and a field as high as possible in the region between the VELO and the tracking stations. The magnet acceptance is ±250 mrad vertically and ±300 mrad horizontally. A perspective view of the magnet is displayed in Figure 2.7, while in Figure 2.8 the map of the magnetic field along the z-axis is shown.

In order to reduce the systematic uncertainties introduced by the potential asymmetry of the detector, the direction of the magnetic field is periodically changed and the two configurations are referred to as Up and Down, to indicate the magnet polarity.

Silicon Tracker

The Trigger Tracker (TT) and the Inner Tracker (IT) constitute the Silicon Trackers (ST) as they both make use of silicon microstrip sensors with a strip pitch of $\sim 200\ \mu$m [36, 41, 43]. The TT station, located upstream of the dipole magnet, covers the whole LHCb acceptance while the IT, placed downstream, covers a cross shape region 120 cm wide and 40 cm high. Each station of the ST consists of four detection layers: in the first and in the last layer the strips are displaced vertically, while in the second and in the third layer strips are rotated by a stereo angle of -5°and +5°, respectively (see Figure 2.9).

Both the TT and IT have a single-hit resolution of $\sim$50 μm, which gets worse in proximity of the beam pipe, and a hit efficiency above 99%. The four hits provided

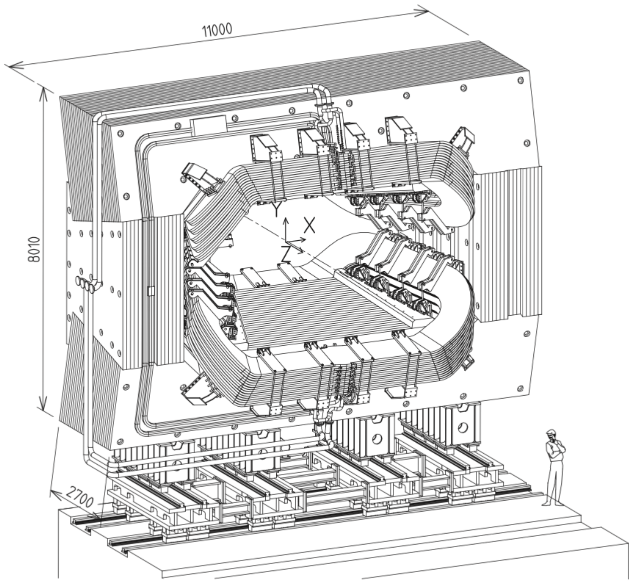


FIGURE 2.7: Perspective view of the LHCb dipole magnet. Taken from Ref. [36].

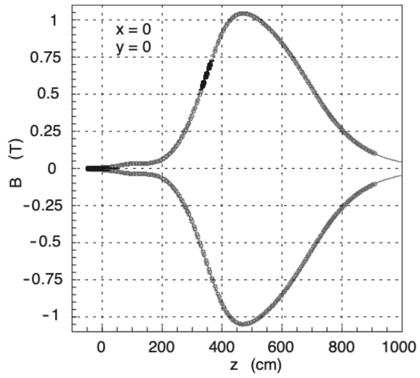
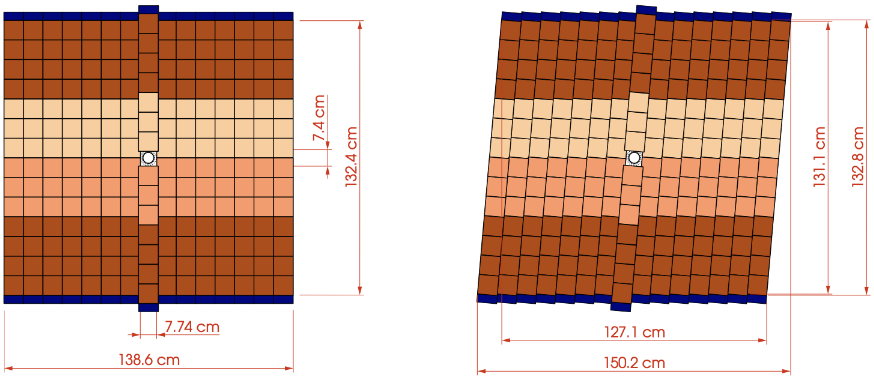


FIGURE 2.8: Magnetic field along the z-axis. Taken from Ref. [36].

by the TT are combined with those produced in the VELO in order to obtain a more accurate reconstruction of the tracks. Moreover, it is used to measure the momentum of those particles that are deflected by the magnet outside the acceptance of the detectors positioned further downstream.

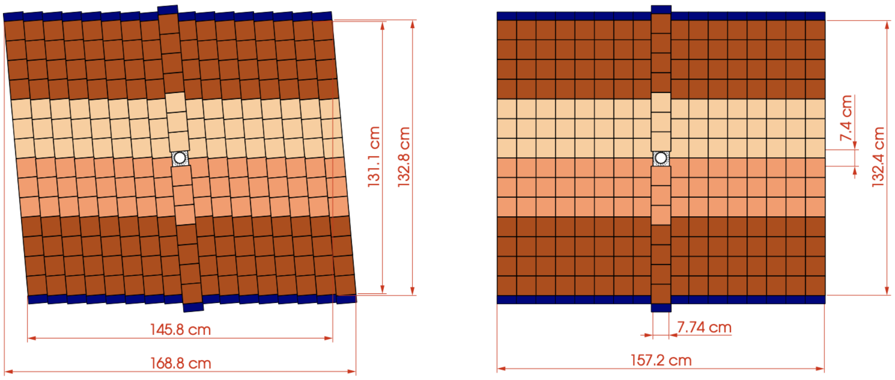
Outer Tracker

The Outer Tracker (OT) [36, 41] represents the outer parts of the tracking stations T1, T2 and T3. It is designed as an array of individual straw-tube modules each containing two staggered layers of drift-tubes with inner diameters of 4.9 mm. A mixture of Argon (70%) and CO_2 (30%) is chosen to guarantee a drift time below 50 ns and a drift-coordinate resolution of 200 μm. Each module is made of three stations, each consisting of fours layers arranged as the layers in the ST stations (see Figure 2.10). The OT acceptance goes from 300 mrad in the magnet bending plane (horizontal) to 250 mrad in the non-bending plane (vertical), and it is therefore used to reconstruct trajectories with high angular aperture.



(1^{st} detection layer)

(2^{nd} detection layer)



(3^{rd} detection layer)

(4^{th} detection layer)

FIGURE 2.9: Layout of the four layers of the TT station. Different shadings indicate different readout sectors. Taken from Ref. [43].

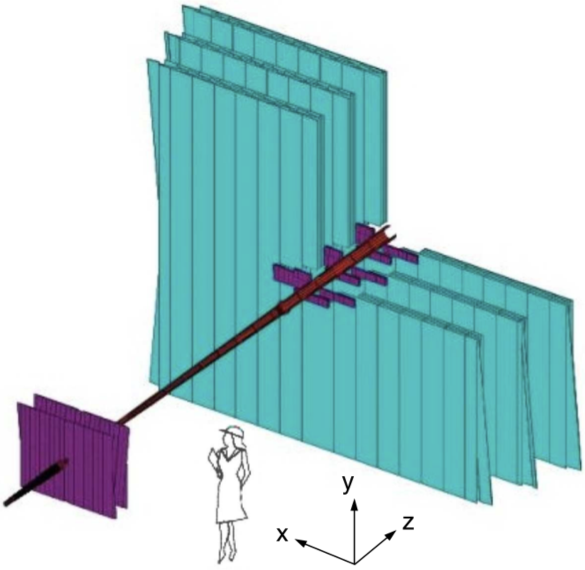


FIGURE 2.10: Arrangement of OT straw-tube modules in layers and stations. Taken from Ref. [36].

2.2.2 Particle identification system

RICH detectors

The capability of the LHCb experiment to identify effectively kaons and pions is an essential prerequisite in the reconstruction of B and D decays with hadronic final states. The identification of such particles is performed by the Ring Imaging CHerenkov Detectors RICH1 and RICH2 [36, 44] whose functioning is based on the Cherenkov effect. When a charged particle crosses a dielectric medium with a velocity v greater than the velocity of light in the medium, it emits electromagnetic radiation inside a cone whose aperture angle θ is related to the parameter $\beta = v/c$:

$$\cos\theta = \frac{1}{n\beta}, \tag{2.1}$$

where n is the refractive index of the medium. The spectrum of the emitted radiation, mainly concentrated in the UV band, has a characteristic shape depending on the medium. The two RICH detectors measure the aperture angle of the cone and so the ratio $\beta = p/E$ of the particle. Combining this result with the momentum measured by the tracking system, it is possible to estimate the mass of the particle and therefore assign a Particle IDentity (PID).

As shown in Figure 2.3, RICH1 and RICH2 are positioned in front and at the exit of the dipole magnet, respectively. The first one is able to identify particles with low momentum (up to 60 GeV/c) and has an angular acceptance between 30 and 300 mrad. The second one is dedicated to the identification of particles with high momentum (up to 100 GeV/c) and with a deflection angle between 15 and 120 mrad. Figure 2.11 shows the reconstructed Cherenkov angle as a function of the particle momentum. A distinct band is visible also for muons, suggesting that RICH detectors, despite being mainly used for hadron indentification, provide useful information also for muon identification.

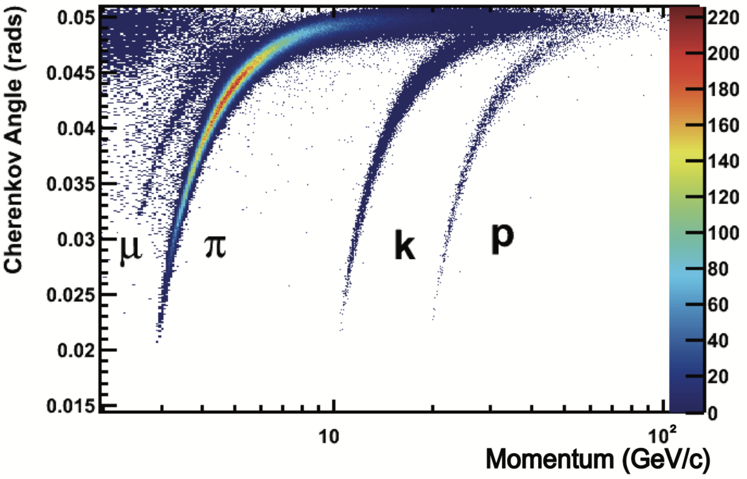


FIGURE 2.11: Reconstructed Cherenkov angle as a function of track momentum. Taken from Ref. [45].

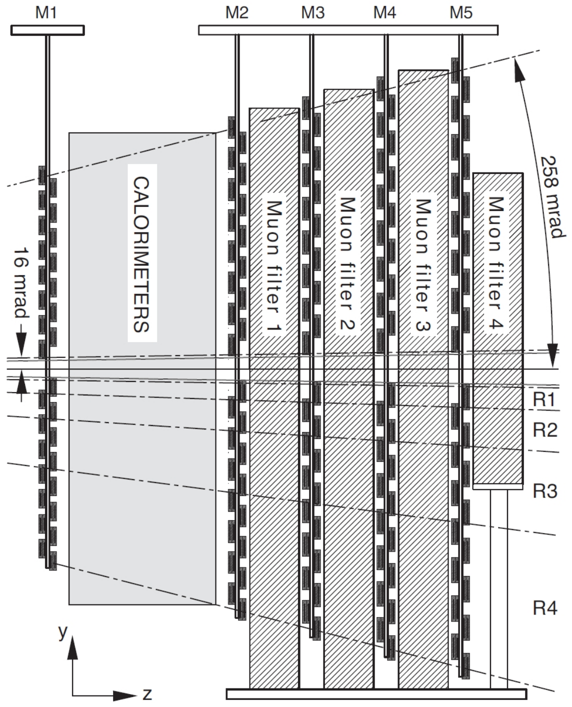


FIGURE 2.12: Side view of the muon identification system. Taken from Ref. [36].

Calorimeters

The LHCb calorimeter system [36], used to identify and measure the energy of photons, electrons and hadrons, consists of three main parts: the Pre-Shower/Scintillator Pad Detector (PS/SPD), the Electromagnetic Calorimeter (ECAL) and the Hadronic Calorimeter (HCAL). The SPD detector is positioned in front of the two calorimeters and its task is to discriminate between neutral and charged particles. The PD, located between the SPD and the ECAL, is instead able to detect photons and electromagnetic showers.

Both calorimeters are made of layers of scintillator material alternate with layers of absorbing material. When a particle passes through or stops in the absorbing material, it deposits part of its energy in the form of charged particles and photons. This energy is then converted into photons in the scintillating material. The electromagnetic radiation produced, which is proportional to the original particle's energy, is then detected by photomultipliers.

Muon stations

Muon identification is a fundamental requirement given the presence of muon in the final state of many rare decays, especially flavour-changing neutral current decays, whose observation would be a clear hint of new physics. The muon identification system [36, 46, 47] consists of five tracking stations. The first station, called M1, is placed in front of the calorimeters and provides the transverse momentum measurement of muon tracks used in the first hardware level of LHCb. The remaining four stations M2, M3, M4, M5 are placed downstream and are separated by iron filters (see Figure 2.12).

The muon stations have a projective geometry, meaning that their transverse size scales with the distance from the interaction point. The technology adopted is the MWPC (multi wire proportional chamber) which is used for all regions except for the inner region of station M1 where triple-GEM detectors are used.

2.2.3 Trigger system

Events detected by LHCb are subject to an immediate pre-selection, called *trigger* [36, 48], organised in two levels:

- Hardware Level-0 (L0) trigger: it combines the information coming from the first muon detector (M1) and from the electromagnetic and hadronic calorimeters. It reduces the beam crossing rate from 40 MHz (nominal frequency of the bunch crossing) to 1 MHz. In order to guarantee a fast selection - it operates at a speed comparable with that of the collisions - the selections have to be extremely easy and not very restrictive. It uses information such as the energy released by particles in calorimeters and the momentum of muons potentially produced.

- Software Level trigger or High Level Trigger (HLT): it can be divided into two selection stages named HLT1 and HLT2. The first one (HLT1) is responsible for the reconstruction of the primary vertex and of the trajectory of charged particles traversing the whole tracking system with a p_T larger than 500 MeV/c. Because of the clear signatures produced by muons in the detector, muon identification can be performed already in HLT1 exploiting information coming from the VELO and from the muon stations. In HLT1 the rate is reduced to 40-80 kHz. In the second trigger stage (HLT2) the full event reconstruction takes place. It consists in the track reconstruction of charged particles, in the reconstruction of neutral particles and in the particle identification. Information coming from the RICH detectors and from the calorimeters are added as input to the muon identification algorithm. With HLT2 the rate is reduced to few kHz.

In 2012, on the occasion of the increasing luminosity, another trigger system, called Deferred trigger, has been introduced. It is able to record to disk 20% of data coming from L0. These data are temporary saved and ready to be subsequently processed by HLT1. This process allows to optimise the use of computing resources and to reduce processing times. Data acquired are written to disk for further analyses. A layout of the LHCb trigger strategy for Run 2 is show in Figure 2.13.

2.2.4 Simulation and data flow

Monte Carlo (MC) simulations are fundamental to understand the experimental conditions and the detector performance. The LHCb simulation application [49–52], called Gauss, consists of two independent stages.

- The *Generator Phase*: here the generation of the primary event takes place. It is divided into two steps. The first one is the production of particles from the *p*-*p* collision, simulated using PYTHIA [53]. The second one includes the decay and time evolution of the produced particles and is performed using EvtGen [54].

- The *Simulation Phase*: here the interaction of particles with the detector (tracking) is simulated via GEANT4 [55].

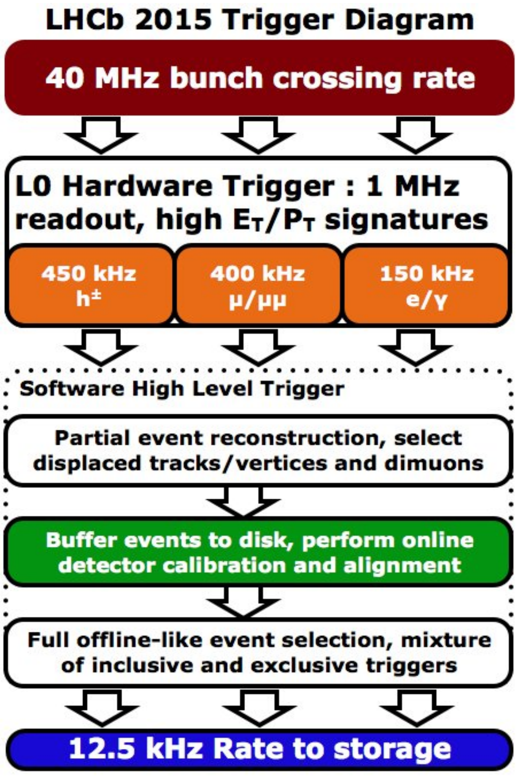


FIGURE 2.13: LHCb trigger strategy for Run 2. Taken from Ref. [48].

A second application, called Boole, is used to simulate the detector response to the passage of particles as well as the response of the hardware trigger L0. From this point on, simulation follows the same data flow as recorded data. The application responsible for the high-level-trigger (HLT) is Moore. Triggered events and raw data are reconstructed by Brunel, an object oriented application which transforms the detector hits in event objects, such as tracks and clusters, and stores them into 'DST' files. Before being accessible to users, DST files are centrally filtered through a set of selections, called *stripping*, controlled by the DaVinci application. The requirements of the *stripping* selection are specified in the so-called *stripping lines*, defined by analysts for the decay of interest. The output from DaVinci is stored into DST (or μDST) files accessible via ROOT [56] and containing the full event information. A schematic illustration of the LHCb data flow is shown in Figure 2.14.

2.2.5 Track reconstruction

Depending on the sub-detectors in which the track has measurements, a different nomenclature is adopted for the type of reconstructed tracks [57]. The different track types are illustrated in Figure 2.15 and defined as follows:

- Long tracks: traverse the full tracking set-up, from the VELO to the last tracking stations. Due to the high number of hits produced, it is possible to measure their momentum with high precision. Only tracks of this type are used in this work.
- Downstream tracks: have hits only in TT, T1, T2 and T3 tracking stations. They allow the reconstruction of decay products of long-lived particles (K_s^0, Λ) decaying out of the VELO.

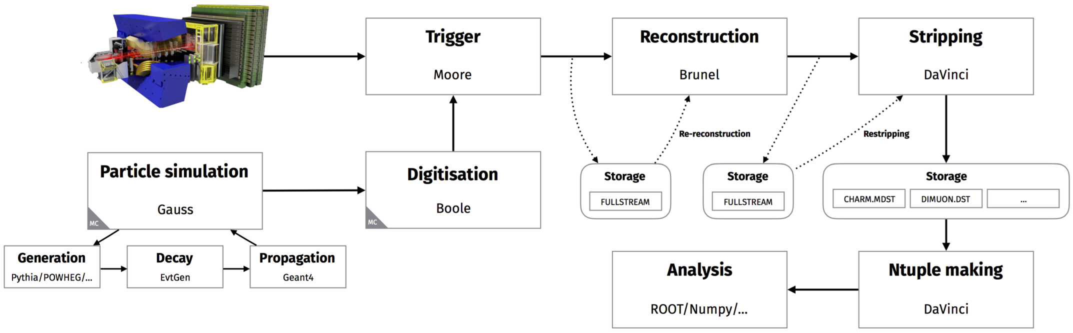


FIGURE 2.14: LHCb data processing applications and data flow. Taken from Ref. [49].

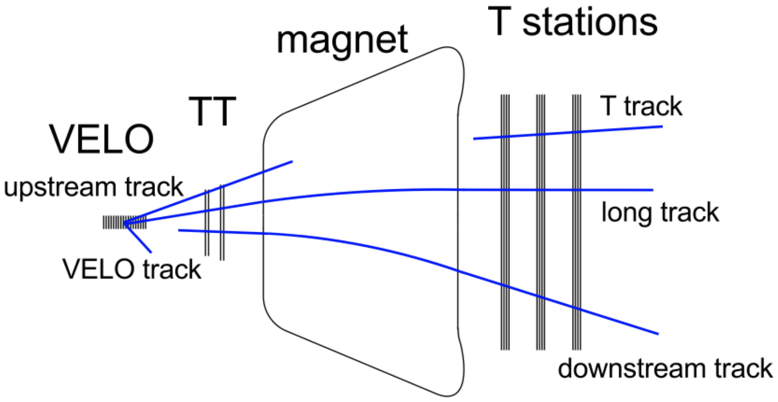


FIGURE 2.15: Schematic illustration of the LHCb track types reconstructed by the different tracking algorithms. Taken from Ref. [57].

- Upstream Tracks: have hits in the VELO and in the TT tracking station. They belong to particles with low momentum deflected by the magnet outside of the angular acceptance of the remaining detectors.
- T tracks: have hits only in the T1, T2 and T3 tracking stations. They often originate from very long-lived particles or material interactions.
- VELO tracks: have hits only in the VELO detector and are mainly used to reconstruct the primary vertex.

2.2.6 Particle identification

The search for the $\tau^+ \to \mu^+\mu^-\mu^+$ decay presented in this thesis, strongly relies on the excellent particle identification (PID) capabilities of LHCb. Misidentification of pions and kaons represents one of the major sources of background for rare decays with muons in the final state. For this reason, it is important to maintain the probability of hadron misidentification as low as possible while ensuring a high muon identification efficiency.

The information provided by the two RICH detectors, the calorimeters and the muon stations are combined into high level variables that can be exploited in the discrimination between signal and background events. For each sub-detector, the

TABLE 2.1: Overview of the required stations with hits within the FoI depending on the momentum of the track [59].

Momentum range	**Required stations**
$p < 3\,\mathrm{GeV}/c$	*always false*
$p < 6\,\mathrm{GeV}/c$	M2 & M3
$6\,\mathrm{GeV}/c < p < 10\,\mathrm{GeV}/c$	M2 & M3 & (M4 || M5)
$p > 10\,\mathrm{GeV}/c$	M2 & M3 & M4 & M5

particle identification information for a given track is stored as difference between the log-likelihood value for a given particle type hypothesis and a pion hypothesis[3] [36, 45]. The likelihood information produced by the various sub-detectors are combined in two different ways to provide two types of variables used for the identification of charged particles [58]. The first one is a linear combination of the log-likelihoods of the different sub-detectors, $\Delta \log \mathcal{L}_{comb}(X - \pi)$ (or $\mathrm{DLL}_{X\pi}$), where X can be a kaon (K), a muon (μ), an electron (e) or a proton (p). The second variable or ProbNNX, is the output of neural networks (NN) that combine particle identification information with tracking information. Apart from these, muons can be identified via another identification variable, called `isMuon`, expressed in terms of a *boolean* decision and representing the compatibility of a track with the muon hypothesis [59]. Depending on its momentum, a track receives the attribute `isMuon` if it has hits in a minimum number of muon stations within a certain field of interest (FOI) defined around the track extrapolation. An overview of the required stations depending on the momentum range can be found in Table 2.1.

Particle identification performance clearly depends on the momentum of the considered track. In general, an excellent separation is achieved in a wide momentum range. Muons are identified with an efficiency of $\epsilon(\mu \to \mu) \sim 97\%$ above 10 GeV/c and a pion misidentification rate of $\epsilon(\pi \to \mu) \sim 1 - 3\%$ [59]. The efficiency for kaon identification, averaged over the momentum range 2-100 GeV/c, is $\epsilon(K \to K) \sim 95\%$, with a corresponding pion misidentification rate of $\epsilon(\pi \to K) \sim 5\%$ [45].

---

[3] Charged pions represent the most abundant particles at LHCb.

Chapter 3

Analysis outline and tools

This thesis presents a search for the lepton flavour violating decay $\tau^+ \to \mu^+\mu^-\mu^+$ carried out on data collected at LHCb in 2016-2018 in *p*-*p* collisions at a centre-of-mass energy of 13 TeV. The aim of this Chapter is to present an overview of the analysis strategy (Section 3.1) and outline the ingredients needed in the measurement of the branching fraction of the $\tau^+ \to \mu^+\mu^-\mu^+$ decay (Section 3.2).

Rare decay searches, either resulting in a handful of signal candidates or in a limit on their branching fraction, strongly rely on the thorough implementation of a series of statistical tools. The second part of this Chapter is dedicated to the introduction of some of the statistical methods adopted in the analysis, including the TISTOS method (Section 3.3), exploited in the evaluation of the correction to the trigger selection efficiency, the multivariate classification models (Section 3.4) used to discriminate between signal and background events, and the CL_s method (Section 3.5), used to evaluate the limit on the branching fraction. Other statistical tools are discussed when needed throughout the thesis.

3.1 Analysis overview

The branching fraction of the $\tau^+ \to \mu^+\mu^-\mu^+$ decay is evaluated relative to the known branching fraction of the $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ decay, chosen as reference channel because of the similarity in topology and kinematic aspects, and because of the presence of two muons in the final state.

Two exclusive sub-samples are considered in the analysis. The main one consists of τ candidates made of three tracks identified as muons by the muon detector (`isMuon`=1 for all final-state tracks). In the second one, instead, only two of the three tracks are identified as muons (`isMuon`=0 for one track). The analysis proceeds in parallel for the two sub-samples, referred to as 3μ sample and 2μ sample, respectively. Finally, both sub-samples are used in the determination of the branching fraction.

The selection of signal candidates, discussed in Chapter 5, begins with the stripping selection, followed by some basic cuts and vetoes removing badly reconstructed events and some peaking backgrounds, and by the trigger selection. Both in the 2μ and 3μ sample, the separation between signal and background events is achieved by means of two multivariate classifiers (see Chapter 7). The first one, based on track isolation and on the 3-body kinematics of the event, is specifically trained against the combinatorial background. The second one, based on muon identification variables, is instead trained to suppress background originating from wrongly identified tracks.

The three muons invariant mass is divided into three regions.

- Signal region: $|m_{\mu\mu\mu} - m_\tau| < 20\ \text{MeV}/c^2$ where $m_\tau = 1776.86\ \text{MeV}/c^2$ is the PDG value of the τ mass [8]. This region is blinded in data, meaning that candidates with an invariant mass between $m_\tau - 20\ \text{MeV}/c^2$ and $m_\tau + 20\ \text{MeV}/c^2$ are removed from the data set until the analysis is complete. As outlined in Ref. [60], this is done in order to avoid biases in the optimisation of the analysis. In this region the signal selection efficiency and the expected number of background events are evaluated.
- Inner (or middle) sideband region: $20\,\text{MeV}/c^2 < |m_{\mu\mu\mu} - m_\tau| < 30\ \text{MeV}/c^2$. This region is used in the optimisation of the trigger selection, in the training phase of the multivariate classifiers and in the optimisation of the binning scheme.
- Outer sideband region: $|m_{\mu\mu\mu} - m_\tau| > 30\ \text{MeV}/c^2$. This region is used, together with the inner sidebands, to fit the background and obtain an expectation of the background yield in the signal region.

Selection efficiencies are evaluated on simulated $\tau^+ \to \mu^+\mu^-\mu^+$ events and corrected for differences between data and simulation using the reference channel $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ (see Chapter 9). After some loose cuts on the output of the two classifiers trained in the 3μ sample, events in the signal region are binned in three variables: the output of the two classifiers and the invariant mass of the three muon candidates.

The limit on the branching fraction is evaluated using the CL_s method in the aforementioned bins. Due to the limited statistics available in the simulated sample, not further binning is considered in the 2μ sample. Systematic uncertainties on the efficiencies and background modelling are included in the limit evaluation. In case of the observation of no events, the combination of the Run 2 limit with the one obtained in the Run 1 analysis [3] is foreseen.

3.2 Branching fraction measurement

The branching fraction of a particular decay mode $X \to Y$ of an initial state particle X, is given by the decay rate to the mode $X \to Y$ relative to the total decay rate:

$$\mathcal{B}(X \to Y) = \frac{\Gamma(X \to Y)}{\Gamma(X \to anything)}\,. \tag{3.1}$$

At a collider experiment, such as the LHC, the branching fraction of a given decay-mode $X \to Y$ can be measured by counting how many of the X particles produced at the collision point decay via the process $X \to Y$:

$$\mathcal{B}(X \to Y) = \frac{N(X \to Y)}{N(pp \to X)}\,. \tag{3.2}$$

The number of particles $N(pp \to X)$ produced in the p-p collisions is given by

$$N(pp \to X) = \sigma(pp \to X) \times \mathcal{L}_{int}\,, \tag{3.3}$$

where $\sigma(pp \to X)$ is the *production cross section* of X in the p-p collision, and $\mathcal{L}_{int}$ the *time-integrated luminosity* quantifying the amount of p-p collisions. In practice, the number of decay processes $N(X \to Y)$ that happened in the experiment has to be corrected by the detection and selection efficiency $\epsilon(X \to Y)$, evaluated on simulated data, to obtain the number of measured decay processes $N(X \to Y)^{meas}$. The branching fraction of $X \to Y$ can therefore be expressed as

$$\mathcal{B}(X \to Y) = \frac{N(X \to Y)^{meas}}{\sigma(pp \to X) \times \mathcal{L}_{int} \times \epsilon(X \to Y)} . \tag{3.4}$$

The cross section $\sigma(pp \to X)$ and integrated luminosity $\mathcal{L}_{int}$ are affected by large uncertainties that can limit the precision of the measurement. To overcome this dependency, the number of events measured in the *signal mode* $X \to Y$, is normalised to the number of events observed in another decay $X' \to Y'$, referred to as the *normalisation mode*, whose branching fraction is well known.

As previously mentioned, the decay used as a normalisation channel is the $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ decay, also referred to as the reference channel. As characterised by a similar decay topology, the $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ decay is also used in the calibration and correction of some of the features of the $\tau^+ \to \mu^+\mu^-\mu^+$ channel exploited in the discrimination between signal and background events.

According to Eq. 3.4, the branching fraction of the signal mode $\tau^+ \to \mu^+\mu^-\mu^+$ can be written as

$$\mathcal{B}(\tau^+ \to \mu^+\mu^-\mu^+) = \frac{N(\tau^+ \to \mu^+\mu^-\mu^+)^{meas}}{\sigma(pp \to \tau^+) \times \mathcal{L}_{int} \times \epsilon_\tau} , \tag{3.5}$$

where $N(\tau^+ \to \mu^+\mu^-\mu^+)^{meas}$ is the number of events observed in the $\tau^+ \to \mu^+\mu^-\mu^+$ channel, $\sigma(pp \to \tau^+)$ the production cross section of τ leptons at LHCb and ϵ_τ the detection efficiency of the $\tau^+ \to \mu^+\mu^-\mu^+$ decay evaluated on Monte Carlo (MC) events. Similarly, the branching fraction of the $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ decay can be expressed as

$$\mathcal{B}(D_s^+ \to \phi(\mu^+\mu^-)\pi^+) = \frac{N(D_s^+ \to \phi(\mu^+\mu^-)\pi^+)^{meas}}{\sigma(pp \to D_s^+) \times \mathcal{L}_{int} \times \epsilon_{D_s}} , \tag{3.6}$$

where $N(D_s^+ \to \phi(\mu^+\mu^-)\pi^+)^{meas}$ is the number of candidates measured in the reference channel. The only way a D_s meson can decay into a τ lepton is via the $D_s^+ \to \tau^+\nu_\tau$ decay, thus, $\sigma(pp \to D_s^+)$ can be written in terms of $\sigma(pp \to \tau^+)$ via the equation

$$\sigma(pp \to D_s^+) = \frac{\sigma(pp \to \tau^+)}{\mathcal{B}(D_s^+ \to \tau^+\nu_\tau)} \times f_{D_s}^{\tau} , \tag{3.7}$$

where $f_{D_s}^{\tau}$ is the fraction of τ produced via D_s decays. Exploiting Eq. 3.6 and Eq. 3.7, the branching fraction of $\tau^+ \to \mu^+\mu^-\mu^+$ is given by

$$\mathcal{B}(\tau^+ \to \mu^+\mu^-\mu^+) = \alpha \times N(\tau^+ \to \mu^+\mu^-\mu^+)^{meas} , \tag{3.8}$$

where the normalisation factor α, also know as the *single event sensitivity*, is expressed as

$$\alpha = \frac{\mathcal{B}(D_s^+ \to \phi(\mu^+\mu^-)\pi^+)}{\mathcal{B}(D_s^+ \to \tau^+\nu_\tau)} \times f_{D_s}^{\tau} \times \frac{\epsilon_{D_s}}{\epsilon_\tau} \times \frac{1}{N(D_s^+ \to \phi(\mu^+\mu^-)\pi^+)^{meas}} . \tag{3.9}$$

The ratio $\epsilon_{D_s}/\epsilon_\tau$ represents the ratio between the overall efficiency of the selection performed in the reference channel and the one performed in the signal channel, both evaluated on simulated events. The selection efficiency is factorised into several terms discussed in detail in Chapter 9. These include the detector acceptance, the reconstruction and stripping selection efficiency, the offline selection efficiency, the trigger selection efficiency and, in case of the signal channel, the efficiency of the selection applied on the output of the multivariate classifiers trained to suppress the different background sources. It should be noted that the choice of the sequence of the selection stages is motivated by the need to make the efficiency determination the most reliable and easy to determine.

3.3 The TISTOS method

As previously mentioned, the efficiency of the trigger selection applied both in the signal and in the reference channel represents one of the terms contributing to the overall selection efficiency needed in the evaluation of the normalisation fraction.

Ideally, the (conditional) trigger efficiency for a given decay channel would be defined as the fraction of events within the detector acceptance selected by the trigger:

$$\epsilon_{Trig|Acc} \equiv \frac{N_{Trig|Acc}}{N_{Acc}}. \tag{3.10}$$

The number of events in the detector acceptance (N_{Acc}) is however not available, as the detector only records events generating a positive trigger response. The strategy is therefore to evaluate the number of trigger-accepted events relative to the final sample of selected events:

$$\epsilon_{Trig|Sel} \equiv \frac{N_{Trig|Sel}}{N_{Sel}}. \tag{3.11}$$

This definition allows us to evaluate the trigger efficiency exploiting quantities available in the data sample by means of the so-called TISTOS method [61]. Depending on which part of the event has fired the trigger, the triggered events can be split into three categories.

1. Triggered On Signal (TOS): events for which the presence of the signal is sufficient to generate a positive trigger decision.

2. Triggered Independent of Signal (TIS): the “rest” of the event, defined through an operational procedure consisting in removing the signal and all detector hits belonging to it, is sufficient to generate a positive trigger decision.

3. Triggered On Both (TOB): events that are neither TIS nor TOS; neither the presence of the signal alone nor the rest of the event alone are sufficient to generate a positive trigger decision, but rather both are necessary.

The three categories are illustrated in the diagram shown in Figure 3.1. It should be noted that a single event can be simultaneously TIS and TOS (TISTOS) whenever both the presence of the signal alone as well as the rest of the event alone are sufficient to generate a positive trigger response. The trigger efficiency defined in Eq. 3.11, can be expressed in terms of the aforementioned categories as

$$\epsilon_{Trig} = \frac{N_{Trig}}{N_{Sel}} = \frac{N_{Trig}}{N_{\mathrm{TIS}}} \times \frac{N_{\mathrm{TIS}}}{N_{Sel}} = \frac{N_{Trig}}{N_{\mathrm{TIS}}} \times \epsilon_{\mathrm{TIS}}, \tag{3.12}$$

where the "$|Sel$" has been omitted assuming that all efficiencies are defined on a sample of selected events. In Eq. 3.12, the quantity ϵ_{TIS} cannot be directly measured

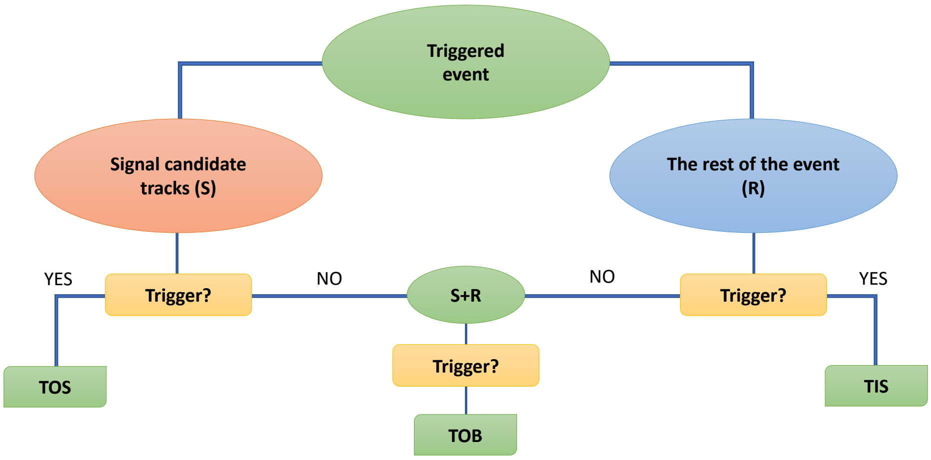


FIGURE 3.1: Diagram explaining the logic behind the categorisation of events into the three trigger categories Trigger On Signal (TOS), Trigger Independent of Signal (TIS) and Trigger On Both (TOB). Note that an event can be TIS and TOS simultaneously. Adapted from Ref. [61].

from data. Provided that the TIS trigger efficiency of any sub-sample of the triggered events is the same as that of the whole sample of selected events, ϵ_{TIS} can be measured on the TOS sub-sample:

$$\epsilon_{\text{TIS}} \equiv \epsilon_{\text{TIS}\,|\,\text{TOS}} = \frac{N_{\text{TISTOS}}}{N_{\text{TOS}}}. \tag{3.13}$$

The trigger efficiency can therefore be written in terms of quantities all measurable from data as

$$\epsilon_{Trig} = \frac{N_{Trig}}{N_{\text{TIS}}} \times \frac{N_{\text{TISTOS}}}{N_{\text{TOS}}}. \tag{3.14}$$

The condition that the TIS efficiency is independent of the chosen sub-sample represents the main assumption of the TISTOS method. The consequences of this assumption are discussed in Section 9.5.3.

Note that the TIS efficiency has to be independent of the signal sample. In case of no correlation between the signal candidate and the underlying event, also the TISTOS efficiency would be independent of the chosen signal sample and Eq. 3.3 satisfied. The trigger selection is, however, mainly based on cuts on the transverse momentum (p_T) and on the impact parameter (IP). This introduces a correlation between the signal candidate and the underlying event. As a workaround, the TISTOS efficiency can be evaluated in small volumes of the signal phase space where the properties of the signal and of the underlying event can be assumed to be largely uncorrelated. The trigger efficiency on selected events would be given by

$$\epsilon_{Trig} = \frac{N_{Trig|Sel}}{\sum\limits_i N^i_{Sel}} = \frac{N_{Trig|Sel}}{\sum\limits_i \frac{N^i_{\text{TIS}\,|\,Sel}}{\epsilon^i_{\text{TIS}}}} = \frac{N_{Trig|Sel}}{\sum\limits_i \frac{N^i_{\text{TIS}\,|\,Sel} N^i_{\text{TOS}\,|\,Sel}}{N^i_{\text{TISTOS}\,|\,Sel}}}, \tag{3.15}$$

where the index i runs over all the bins in the phase space of the signal candidate. As later discussed in Section 9.5.3, the correlation between the signal candidate and the rest of the event can be studied on $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ simulated events, where the trigger efficiency can be computed both with the definition in Eq. 3.11 and with the TISTOS method.

3.4 Boosted Decision Tree classifiers

This analysis exploits multivariate classification models to discriminate between signal decays and the various background sources. These can include background originating from tracks wrongly associated with a τ decay, also called combinatorial background, and D^+ and D_s decays where one or more tracks are misidentified as muons. The advantage of adopting methods based on multivariate analysis is that the information of several variables, or *features*, can be combined in a single discriminating variable exploited in the separation between signal and background events. Among them, this analysis makes use of *Boosted Decision Tree* (BDT) classifiers, available in the TMVA toolkit [62]. BDT classifiers belong to the family of "supervised learning" algorithms, as they are trained on events for which the outcome of the classification is known. In this analysis, $\tau^+ \to \mu^+\mu^-\mu^+$ simulated events are used as a proxy for the signal, while events from the data inner sidebands are used as a proxy for the background.

A *decision tree* consists in a binary tree-structured classifier, like the one displayed in Figure 3.2. Starting from the *root node*, the tree consists of a consecutive set of *nodes* splitting the data sample by means of rectangular cuts with binary outcome. In each node the cut is applied on the variable giving, at that node, the best separation between signal and background events. The optimal separation cut is determined by a specific metric function. One of the most used metric is the so-called *Gini-index* [63], which measures the separation as $p \cdot (1-p)$, with p being the purity of the partitions determined as the fraction of signal events in the node. In the optimisation of the *decision tree*, the maximisation of the separation power goes in parallel with the minimisation of the *loss function*, which quantifies the inaccuracy of predictions in a classification problem. Depending on the majority of training events ending up in the final node, called *leaf node*, the phase space is split into different regions eventually classified as signal or background.

Single decision trees are easy to interpret and relatively fast to train, though they are unstable with respect to statistical fluctuations in the training sample. To improve the stability and the classification power of the algorithm, the concept of a single decision tree is extended from one to several trees to form a *forest*. This process is called *boosting* of a decision tree. The trees making up the *forest* are all derived from the same training sample by re-weighting the events and, eventually, a single classifier is obtained as weighted average of the individual decision trees.

In the 2μ sample (Section 7.3), the *GradientBoost* algorithm is employed in the training of the decision trees [64]. With this technique, the minimisation of the loss function is performed by calculating the gradient of the loss function, and then growing a regression tree whose leaf values are adjusted to match the mean value of the gradient in each region defined by the tree structure. In the 3μ sample (Section 7.2), the algorithm adopted in the training of the decision tree is the so-called *eXtreme Gradient Boosting* (XGB) [65], a scalable machine learning system for tree boosting.

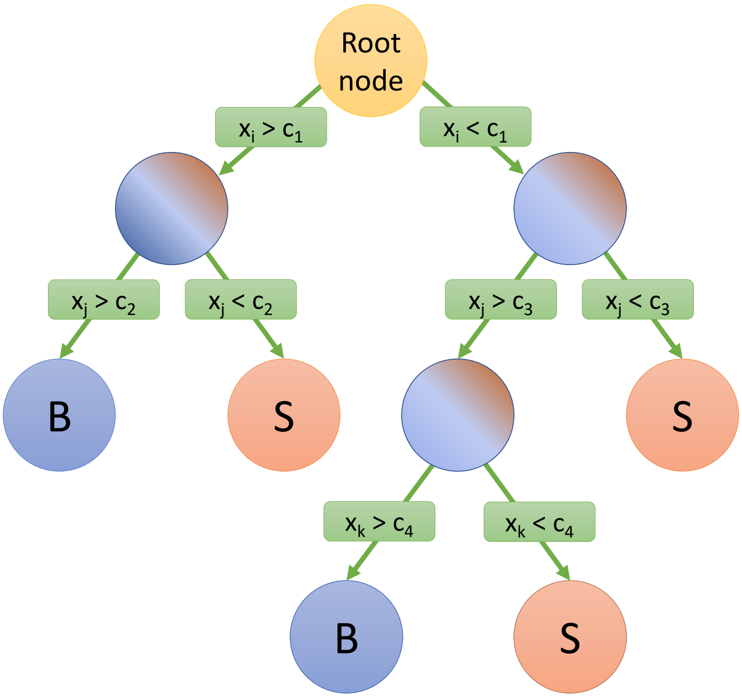


FIGURE 3.2: Schematic view of a *decision tree* adapted from Ref. [62]. Starting from the root node, a consecutive set of nodes splits the data sample by means of rectangular cuts with binary outcome. In each node the cut is applied on the discriminating variable x_i giving, at that node, the best separation between signal and background events. The same variable can be used in several nodes or not used at all. The final leaf nodes are labeled S (signal) or B (background) depending on the majority of events ending up in the respective nodes.

3.5 The CL_s method

In the search for decays strongly suppressed in the Standard Model (SM), such as the $\tau^+ \to \mu^+\mu^-\mu^+$ decay, there are two possible scenarios that analysts may encounter [66]. The first one is the "discovery" and it occurs when observation disagrees with the SM prediction, which is commonly referred to as Null-Hypothesis (H_0). In the second scenario, or "exclusion", the Alternative-Hypothesis (H_1) is excluded by the observed data. Both outcomes are to be interpreted as successful, as also the "exclusion" scenario provides important information that theorists or experimentalists can exploit for future searches. This argument applies to the search presented in this thesis; improving the upper limit on the branching fraction (BF) of the $\tau^+ \to \mu^+\mu^-\mu^+$ decay would allow to constrain, and therefore to exclude, theories of physics beyond the SM.

The evaluation of the expected upper limit is performed by means of the CL_s method, a technique commonly used in high energy physics to set upper/lower limits [14, 66–68]. In this specific analysis, H_1 would correspond to the observation of an excess in the signal region, and can therefore be referred to as the signal+background ($s + b$) hypothesis. H_0, instead, describes the observation of no signal or of a signal that is too small to be seen, and is referred to as the background-only (b) hypothesis.

The best way to test the two hypotheses is to evaluate their likelihood ratio

$$\lambda = \frac{\mathcal{L}_{s+b}(n_{obs})}{\mathcal{L}_b(n_{obs})}, \tag{3.16}$$

where $\mathcal{L}_{s+b(b)}(n_{obs})$ represents the probability to observe n_{obs} events under the hypothesis $s+b$ (b). Taking advantage of the central limit theorem[1], instead of the likelihood ratio λ, one can consider the *test statistic* Q defined as

$$Q_{obs} = -2\ln\lambda = -2\ln\frac{\mathcal{L}_{s+b}(n_{obs})}{\mathcal{L}_b(n_{obs})}. \tag{3.17}$$

Q_{obs} has the advantage of being additive, meaning that in case of experiments with multiple measurements of n_{obs}, the combined *test statistic* would be the sum of the individual *test statistics*. In a counting experiment, a natural choice for $\mathcal{L}_{s+b}$ and $\mathcal{L}_b$ is a Poisson distribution. Therefore, for an experiment with N independent measurements of a discriminating variable x, the observed *test statistic* can be written as

$$Q_{obs} = -2ln\left[\frac{\prod\limits_{i=1}^{N}\frac{e^{-(s_i+b_i)}(s_i+b_i)^{n_i}}{n_i!}}{\prod\limits_{i=1}^{N}\frac{e^{-(b_i)}(b_i)^{n_i}}{n_i!}} \times \frac{\prod\limits_{j=1}^{n_i}\frac{s_iS_i(x_{i,j})+b_iB_i(x_{i,j})}{s_i+b_i}}{\prod\limits_{j=1}^{n_i}B_i(x_{i,j})}\right], \tag{3.18}$$

and simplified to

$$Q_{obs} = -2ln\left[e^{-s_{tot}}\prod_{i=1}^{N}\prod_{j=1}^{n_i}\left(\frac{s_iS_i(x_{i,j})+b_iB_i(x_{i,j})}{b_iB_i(x_{i,j})}\right)\right] \tag{3.19}$$

$$= 2s_{tot} - 2\sum_{i=1}^{N}\sum_{j=1}^{n_i} ln\left(\frac{s_iS_i(x_{i,j})+b_iB_i(x_{i,j})}{b_iB_i(x_{i,j})}\right), \tag{3.20}$$

where s_{tot} is the total signal rate for all measurements, n_i is the number of candidates observed in each measurement i, s_i (b_i) is the number of signal (background) candidates in channel i, $x_{i,j}$ is the value of x measured for each candidate j, and $S_i(x_{i,j})$ ($B_i(x_{i,j})$) the signal (background) probability density function (pdf) of the discriminating variable in channel i.

Once the *test statistic* is defined, one should establish under which conditions observations are compatible with a discovery or an exclusion. Practically speaking, the *confidence level* (CL) for the exclusion has to be defined. The confidence level for the signal+background hypothesis, CL_{s+b}, is defined as the probability of observing a *test statistic* greater than or equal to Q_{obs},

$$CL_{s+b} = P_{s+b}(Q \geq Q_{obs}), \tag{3.21}$$

with

$$P_{s+b} = \int\limits_{Q_{obs}}^{+\infty} f(Q|s+b)dQ \tag{3.22}$$

and being $f(Q|s+b)$ the pdf of the *test statistic* for the *s+b* hypothesis as indicated

---

[1] According to the central limit theorem the distribution of a certain variable approximates to a Gaussian distribution in the high-statistic limit [69].

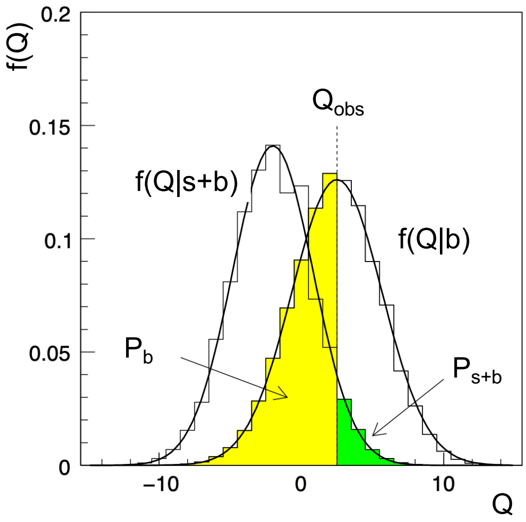


FIGURE 3.3: Distribution of the *test statistic* under the $s+b$ and b hypotheses. Adapted from Ref. [68].

in Figure 3.3. It can happen that if the signal is small and the background fluctuates downwards, the observed *test statistic* Q_{obs} is barely compatible with the signal+background hypothesis but also with the background-only hypothesis. This scenario occurs when the Q distributions for the $s+b$ and b hypotheses almost overlap with each other. In order to avoid a *false rejection* of the signal hypothesis, the confidence level for the $s+b$ hypothesis is normalised to the confidence level observed for the b hypothesis, CL_b, defined as

$$CL_b = 1 - P_b(Q \leq Q_{obs}), \tag{3.23}$$

where

$$P_b = \int_{-\infty}^{Q_{obs}} f(Q|b)dQ \tag{3.24}$$

and being $f(Q|b)$ the pdf of the *test statistic* for the b hypothesis. In this *modified frequentist* approach, the confidence (or better ratio of confidences) for the signal hypothesis would be

$$CL_s \equiv \frac{CL_{s+b}}{CL_b}. \tag{3.25}$$

The signal hypothesis is considered excluded at the confidence level CL when

$$1 - CL_s \leq CL. \tag{3.26}$$

Common values for the confidence level are 90% and 95%. As discussed in the following, the confidence level for the signal hypothesis CL_s is computed for BF values within a certain range. The BF value for which the condition in Eq. 3.26 holds will be quoted as the *confidence limit*. Note that upper limits computed with the CL_s method are always conservative, since $CL_b \leq 1$ implies that $CL_s \geq CL_{s+b}$.

3.5.1 The expected upper limit

As mentioned in Section 3.1, the analysis is designed hiding the signal region in order to avoid possible biases in the optimisation of the analysis. Thus, before the unblinding of the signal region, only the expected *confidence limit* for the exclusion of the signal hypothesis can be evaluated. The CL_s is evaluated for BF values within a range that is compatible with the *confidence limit*. The extrapolation of the expected upper limit from the Run 1 to the Run 2 data set (see Section 11.1), simply based on the larger luminosity and higher production cross section, is exploited to decide on the BF range for which the scan is performed.

For each BF value, the computation of CL_{s+b} and C_b is performed by means of pseudo-experiments ("toy Monte Carlo"). For each "toy", the *test statistic* for the $s+b$ hypothesis, Q_{s+b}, and for the b hypothesis, Q_b, are evaluated. The number of signal events in channel i (s_i) for a given BF value $\mathcal{B}(\tau^+ \to \mu^+\mu^-\mu^+)$ is extracted as $\mathcal{B}(\tau^+ \to \mu^+\mu^-\mu^+)/\alpha_i$, where α_i is the value of the normalisation factor (or *single event sensitivity*) computed in the same channel. The number of background events in the single measurement (b_i) is also unknown. An expected value for b_i is estimated by fitting the invariant mass distribution in the inner and outer sidebands and extrapolating the resulting background shape to the signal region to finally evaluate the corresponding yield (discussed in Chapter 8). The number of candidates n_i observed in the signal region for each measurement is generated as a Poissonian random number with mean $s_i + b_i$ and b_i for the computation of CL_{s+b} and CL_b, respectively.

The pdf of Q_{s+b} and Q_b are normalised to the number of toys. Finally, CL_{s+b} and CL_b are computed according to Eq. 3.22 and Eq. 3.24, respectively, taking, instead of Q_{obs}, the median of the pdf of the *test statistic* for the b hypothesis as upper limit of the integrals. Note that CL_b is by definition equal to 0.5.

The estimated number of background events b_i in the signal region, as well as the normalisation factor α_i, are of course affected by an uncertainty. A distinction is made between the statistical and systematic uncertainty as two different strategies are adopted to take them into account in the limit evaluation.

The statistical uncertainty is simply a consequence of the size of the data sample for b_i and of the simulated sample for α_i. Its effect on the expected number of background events b_i is considered by generating a random number r from a Gaussian distribution (with $\mu = 0$ and $\sigma = 1$), and fluctuating b_i by $r \times \sigma_{b_i}$ where σ_{b_i} is the error from the fit. The same strategy is used to fluctuate the normalisation factor α_i around its central value.

The systematic uncertainty is instead introduced by the choice of the pdf used to describe the background and the signal shape. As outlined in Section 3.1, the expected limit is evaluated in bins of the invariant mass distribution. The fraction of candidates falling in the single mass bin, evaluated on simulated events, clearly depends on the function used to model the signal peak. This systematic uncertainty is taken into account by generating n-times fluctuated pdfs and randomly extracting one of them. The same procedure is followed to consider the systematic uncertainty introduced by the choice of the pdf used to model the background shape. The fluctuation of the signal and background pdfs is described in detail in Section 10.2 and 10.3.

For the sake of simplicity, the uncertainty introduced by the external inputs and efficiency corrections that enter the evaluation of α_i, is combined with the statistical uncertainty and treated as a unique uncertainty in the evaluation of the expected limit. The various sources of systematic uncertainties are discussed in Chapter 10.

Finally, the distribution of the median of CL_s computed for each BF value that is being tested, is shown in a plot together with its $\pm 1\sigma$ and $\pm 2\sigma$ error bands, being σ the standard deviation. The color choice that is usually adopted for the two bands (green for $\pm 1\sigma$ and yellow for $\pm 2\sigma$) is the reason why this plot is commonly known as "Brazilian plot".

Chapter 4

Description of data sets

This Chapter is dedicated to the description of the data sets used in the analysis, including the simulated samples on which the selection efficiencies are evaluated. The preparation of the simulated samples through the application of the *truth-matching* is presented in Section 4.2.1, while Section 4.2.2 reports the evaluation of the τ and D_s production rates.

4.1 Data sample

The search for the $\tau^+ \to \mu^+\mu^-\mu^+$ decay is performed analysing data collected by the LHCb experiment during Run 2 (2016-2018) at $\sqrt{s}$ =13 TeV corresponding to an integrated luminosity of 5.4 fb^{-1}.

In case of the observation of no events, the upper limit on the branching fraction of $\tau^+ \to \mu^+\mu^-\mu^+$ is combined with the result obtained in the Run 1 analysis [3], performed on data collected in 2011-2012 at $\sqrt{s}$ =7-8 TeV. The integrated luminosity for each data-taking year of Run 1 and Run 2 is reported in Table 4.1.

TABLE 4.1: Integrated luminosity of the LHCb experiment for the different data-taking years [70].

	2011	**2012**	**2016**	**2017**	**2018**
$\mathcal{L}_{int}$ [fb^{-1}]	1	2	1.6	1.7	2.1

4.2 Simulation samples

As stated in Section 3.2, the detection and selection efficiency for a specific decay process is evaluated on simulated events. In addition to the MC sample produced for the signal and reference mode, simulated data have been generated also for the $D_s^+ \to \eta(\mu^+\mu^-\gamma)\mu^+\nu_\mu$ channel as the dominant background process with three muons in the final state (see Section 8.1.1). Tables 4.2, 4.3 and 4.4 show the number of simulated candidates that have been generated for the three decay modes, separately for each data-taking year and polarity. In case of the signal and reference channel, the simulated candidates are further split into the different modes contributing to the τ and D_s production, respectively. As later discussed in Section 4.2.2, τ can be produced via D_s-, D^+- or B-meson decays, where D_s and D^+ mesons can be either produced *prompt*, at the collision point, or *secondary*, in b-hadron decays.

It is important to mention that data - be real or simulated data - are organised "per candidate" and not "per event". This means that for a given event, namely a specific p-p interaction, it can happen that more than one candidate, *i.e.* more than

TABLE 4.2: Number of candidates simulated for each sub-channel, data-taking year and polarity for the signal mode $\tau^+ \to \mu^+\mu^-\mu^+$.

Sub-channel	**2016**		**2017**		**2018**	
	MD	**MU**	**MD**	**MU**	**MD**	**MU**
$D_s^+ \to \tau^+\nu_\tau$	129972	129942	111092	129776	114959	118229
$\bar{b} \to D_s^+ \to \tau^+\nu_\tau$	40789	41695	36367	36326	39175	41860
$D^+ \to \tau^+\nu_\tau$	15880	13526	15076	16583	12728	11448
$\bar{b} \to D^+ \to \tau^+\nu_\tau$	4013	4034	3984	3864	3924	4000
$\bar{b} \to \tau^+\nu_\tau$	51745	51192	51264	54110	60311	57946
Total	482788		458442		464580	

TABLE 4.3: Number of candidates simulated for each sub-channel, polarity and data-taking year for the $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ mode.

Sub-channel	**2016**		**2017**		**2018**	
	MD	**MU**	**MD**	**MU**	**MD**	**MU**
prompt D_s^+	72664	79078	86831	83087	69240	63887
$\bar{b} \to D_s^+$	25954	26252	27671	27329	24477	26850
Total	203948		224918		184454	

TABLE 4.4: Number of candidates simulated for each data-taking year and polarity for the $D_s^+ \to \eta(\mu^+\mu^-\gamma)\mu^+\nu_\mu$ mode.

Channel	**2016**		**2017**		**2018**	
	MD	**MU**	**MD**	**MU**	**MD**	**MU**
$D_s^+ \to \eta(\mu^+\mu^-\gamma)\mu^+\nu_\mu$	16034	15437	14665	15170	14988	15597
Total	31471		29835		30595	

one combination of particles, may be compatible with the designed selection. This gives rise to a candidate multiplicity that, after the final selection, should be as close to unity as possible, since at most one candidate per event corresponds to reality.

4.2.1 Truth-matching in the simulated samples

Simulated events are reconstructed by means of the same algorithms adopted in the reconstruction of recorded data. This implies that even in the simulation sample generated for a certain decay process, there can be cases where the signal candidate is reconstructed from tracks belonging to the underlying event wrongly associated with the given process. To make sure that the selection efficiency is evaluated on a pure sample of signal candidates, true origin, identity and momenta of particles making up the candidate can be exploited in order to remove the unwanted candidates.

Both in the signal and in the reference channel, signal candidates are selected by means of the `Background Category` tool [71, 72]. This tool studies the true information of the particles in the decay chain and returns a value (`BKGCAT`) for each candidate. In the signal and reference channel, only events with `BKGCAT` equal 50, 10 or 0 are selected.

The meaning of these background categories is explained below.

- `BKGCAT`=50: all of the final state tracks are associated to simulated particles with the correct particle ID (no misidentification), but the parent decay has not been fully reconstructed. Also, the true common parent must have a mass at most 100 MeV/c^2 above the mass of what is intended to be reconstructed.

- `BKGCAT`=10: the decay is correctly and fully reconstructed but an intermediate resonance or particle is misidentified.

- `BKGCAT`=0: none of the previous condition is satisfied (pure signal).

In the $D_s^+ \to \eta(\mu^+\mu^-\gamma)\mu^+\nu_\mu$ simulated sample, pure signal candidates are selected by means of the true identity of the final-state particles, of their *mother* and *grandmother* particles. In addition, `BKGCAT`=40 is required in order to select events where the parent decay has not been fully reconstructed and the low mass condition (`BKGCAT`=50) is not satisfied. More on the truth-matching in the $D_s^+ \to \eta(\mu^+\mu^-\gamma)\mu^+\nu_\mu$ decay can be found in Section 8.1.1.

4.2.2 Production rates of τ leptons and D_s mesons

At LHCb τ leptons are almost entirely produced via *c*- and *b*-hadron decays, *e.g.* D_s-, D^+- and *B*-meson decays. The $Z \to \tau^+\tau^-$ and $W^+ \to \tau^+\nu_\tau$ production cross section are three orders of magnitude smaller and can therefore be neglected [73, 74].

Since the MC sample is produced separately for each source with an arbitrary luminosity, when events from different sources are combined, the number of events produced for each sub-channel needs to be weighted in order to agree with the LHCb measurements of the charm and beauty cross section at 13 TeV and the known values of the involved branching fractions.

The strategy used to evaluate the weights that will be assigned to the MC sample produced for each sub-channel, referred to as *MC mixing method*, is summarised in Table 4.5 and Table 4.6 for the signal and reference channel, respectively.

TABLE 4.5: Fractional contribution of the different sub-channels to the τ production in 4π solid angle at 13 TeV.

Channel	σ^{meas} [μb]	$\sigma^{4\pi}$ [μb]	$\mathcal{B}_D$ [%]	$\mathcal{B}_\tau$ [%]	$\sigma^{4\pi} \times \mathcal{B}_{tot}$ [μb]	**Calc**$^{4\pi}$ **13 TeV [%]**
$D_s^+ \to \tau^+$	353 ± 76	1732 ± 373	-	5.32 ± 0.11	92.1 ± 19.9	68.66 ± 3.71
$\bar{b} \to D_s^- \to \tau^-$ $\bar{b} \to D_s^+ \to \tau^+$		$2 \times (495 \pm 52)$	14.7 ± 2.1 10.1 ± 3.1	5.32 ± 0.11	13.1 ± 2.4	9.73 ± 0.53
$D^- \to \tau^-$	834 ± 78	4054 ± 379	-	0.12 ± 0.027	4.9 ± 1.2	3.62 ± 0.20
$\bar{b} \to D^- \to \tau^-$ $\bar{b} \to D^+ \to \tau^+$		$2 \times (495 \pm 52)$	22.7 ± 1.6 0.5 ± 0.5	0.12 ± 0.027	0.3 ± 0.1	0.21 ± 0.01
$\bar{b} \to \tau^+$		$2 \times (495 \pm 52)$	-	2.41 ± 0.23	23.9 ± 3.4	17.78 ± 0.96
Total					134.2 ± 20.4	100

TABLE 4.6: Fractional contribution of the different sub-channels to the D_s production in 4π solid angle at 13 TeV.

Channel	σ^{meas} [μb]	$\sigma^{4\pi}$ [μb]	$\mathcal{B}_D$ [%]	$\sigma^{4\pi} \times \mathcal{B}_D$ [μb]	**Calc**$^{4\pi}$ **13 TeV [%]**
prompt D_s^+	353 ± 76	1732 ± 373	-	1732 ± 373	87.58 ± 2.70
$\bar{b} \to D_s^-$ $\bar{b} \to D_s^+$		$2 \times (495 \pm 52)$	14.7 ± 2.1 10.1 ± 3.1	246 ± 45	12.42 ± 0.38
Total				1977 ± 376	100

TABLE 4.7: Overview of the evaluation of the conversion factor $f^{4\pi}$. Efficiencies are evaluated on 100K generated events.

Decay	ϵ_{GEN} [%]	$\epsilon_{\text{GEN}\,|\,\text{CUT}}$ [%]	$f^{4\pi}$
$D_s^+ \to K^-K^+\pi^+$	18.50 ± 0.05	3.77 ± 0.02	4.91 ± 0.03
$D^+ \to K^-K^+\pi^+$	99.983 ± 0.004	20.57 ± 0.08	4.86 ± 0.02
$D_s^+ \to K^-K^+\pi^+$ (PHSP)	99.984 ± 0.004	20.32 ± 0.08	4.92 ± 0.02
$D^+ \to K^-K^+\pi^+$ (PHSP)	99.988 ± 0.003	20.46 ± 0.08	4.89 ± 0.02

The first column in Table 4.5 shows the five sub-channels that contribute to the τ production, while the second column contains the LHCb measurements of the prompt charm production cross section in the kinematic ranges $1\ \text{GeV}/c < p < 8\ \text{GeV}/c$ and $2 < y < 4.5$ [75]. The corresponding cross section in 4π ($\sigma^{4\pi}$) is obtained by multiplying the measured cross section (σ^{meas}) by a conversion factor ($f^{4\pi}$) evaluated with PYTHIA [53] as ratio between the generation-only efficiency and the efficiency of the cuts applied at the generator level of the inclusive production of prompt D_s and D^+ mesons.

More specifically, 100K simulated events have been generated for the inclusive D_s production, with $D_s^+ \to K^-K^+\pi^+$, removing the requirement on the decay products to be in the acceptance of LHCb to evaluate the generation-only efficiency ϵ_{GEN}, and keeping the requirement to get the efficiency of the generator level cuts $\epsilon_{\text{GEN}|\text{CUT}}$. The same is done for the inclusive production of D^+, generating 100K simulated events for the $D^+ \to K^+K^-\pi^+$ decay. In Table 4.7 the values of the efficiencies and of the corresponding conversion factors are shown for the D_s and D^+ production.

As a cross check, the same procedure was repeated generating 100K events for the $D_s^+ \to K^-K^+\pi^+$ and $D^+ \to K^-K^+\pi^+$ decays, but with a phase-space (PHSP) decay model. The resulting conversion factors, also reported in Table 4.7, are in agreement with the results obtained for the inclusive production. The prompt beauty production cross section in 4π is instead taken from Ref. [76].

The cross section in 4π is multiplied by the total branching fraction of the corresponding decay process to evaluate the fractional contribution of each sub-channel to the τ production. The branching fractions needed at this stage ($\mathcal{B}_D$ and $\mathcal{B}_\tau$ in Tables 4.5 and 4.6) are listed here and can be found in the PDG [8], with some exceptions:

- $\mathcal{B}(\bar{b} \to D_s^-) = (14.7 \pm 2.1)\%$
- $\mathcal{B}(\bar{b} \to D_s^+) = (10.1 \pm 3.1)\%$
- $\mathcal{B}(\bar{b} \to D^-) = (22.7 \pm 1.6)\%$
- $\mathcal{B}(\bar{b} \to D^+) = (0.5 \pm 0.5)\%$[1]
- $\mathcal{B}(D_s^+ \to \tau^+\nu_\tau) = (5.32 \pm 0.11)\%$ [2]
- $\mathcal{B}(D^+ \to \tau^+\nu_\tau) = (1.20 \pm 0.27) \times 10^{-3}$
- $\mathcal{B}(\bar{b} \to \tau^+) = (2.41 \pm 0.23)\%$

---

[1] The branching fraction for $\bar{b} \to D^+$ is unobserved, still it can be extrapolated multiplying the branching fraction for $\bar{b} \to D_s^+$ by $(|V_{cd}|^2/|V_{cs}|^2)$.

[2] The only way a $D_{(s)}$ meson can decay into a τ lepton is via the process $D_{(s)}^+ \to \tau^+\nu_\tau$.

The fractional contributions of the different production channels are normalised to unity to finally obtain the percentage contribution shown in the last column of the same Table. As can be seen, most of the τ produced at LHCb come from prompt D_s production.

The same procedure is followed in case of the $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ channel, to determine the fractional contributions of D_s production at LHCb in 4π solid angle at 13 TeV. Similarly to the signal channel, Table 4.6 shows the processes contributing to the D_s production, the cross sections and branching fractions involved, as well as the final fractional contributions. The prompt D_s production is again the dominant production channel.

The fractional contribution of each sub-channel is now translated into a weight that can be directly applied to the simulated events produced for each sub-channel[3]. More specifically, the percentage contributions evaluated so far are weighted for the efficiency of the cuts applied at the generator level for each sub-channel, which indeed depends on the kinematics of the corresponding process. This cut efficiency, or ϵ_{CUT}, is obtained as ratio between $\epsilon_{\text{GEN|CUT}}$ and ϵ_{GEN}, being the former the fraction of generated τ produced in a specific channel of interest, that pass the cuts at the generator level and decay within the LHCb acceptance, and being the latter the efficiency for a τ to be produced in a specific sub-channel. As already described, ϵ_{GEN} can be obtained by preparing a generator-level-only sample for each sub-channel with no cuts applied. All these efficiencies are listed in Table 4.8 and Table 4.9 for the signal and for the reference channel, respectively.

The weighted percentage contribution, or w_{Gauss}, is obtained for each sub-channel as

$$w_{\text{Gauss},i} = \frac{\epsilon_{\text{CUT},i} \times \text{Calc}_i^{4\pi}}{\sum_j \epsilon_{\text{CUT},j} \times \text{Calc}_j^{4\pi}}\,. \tag{4.1}$$

The final *per-sub-channel* weight depends on the number of events N_{prod} which have been generated for the corresponding sub-channel and can therefore be computed as

$$w_{\text{fraction,i}} = \frac{N^*_{\text{prod}}}{w^*_{\text{Gauss}}} \times \frac{w_{\text{Gauss},i}}{N_{\text{prod},i}}\,, \tag{4.2}$$

where N^*_{prod} is the number of events generated for the sub-channel with the minimum value of $N_{\text{prod}}/w_{\text{Gauss}}$ and w^*_{Gauss} its corresponding value of w_{Gauss}. As an example, the last column of Tables 4.8 and 4.9 shows the weights computed for the 2018 MagDown simulated sample, though the same procedure is repeated for each data-taking year and polarity.

From now on each candidate in the simulated samples is weighted with the corresponding w_{fraction} so that all generated events are considered in the analysis without losing statistics.

---

[3] Simulated events belonging to the same sub-channel share the same weight.

TABLE 4.8: *MC mixing method* for $\tau^+ \to \mu^+\mu^-\mu^+$ at 13 TeV. (*) indicates the sub-channel with the minimum value of $N_{\text{prod}}/w_{\text{Gauss}}$.

Decay Chain	**Calc$^{4\pi}$ 13 TeV [%]**	$\epsilon_{\text{GEN\|CUT}}$ **[%]**	ϵ_{GEN} **[%]**	ϵ_{CUT} **[%]**	w_{Gauss} **[%]**	N_{prod}	w_{fraction}
$D_s^+ \to \tau^+$	68.66 ± 3.71	10.1 ± 0.3	90.5 ± 0.9	11.1 ± 0.3	71.43 ± 1.17	647132^*	1.00
$\bar{b} \to D_s^+ \to \tau^+$	9.73 ± 0.53	1.04 ± 0.03	10.9 ± 0.3	9.5 ± 0.4	8.67 ± 0.14	102205	0.77
$D^+ \to \tau^+$	3.62 ± 0.20	10.2 ± 0.3	91.0 ± 0.9	11.2 ± 0.3	3.79 ± 0.06	50220	0.68
$\bar{b} \to D^+ \to \tau^+$	0.21 ± 0.01	0.91 ± 0.03	9.2 ± 0.3	9.8 ± 0.4	0.189 ± 0.003	9999	0.17
$\bar{b} \to \tau^+$	17.78 ± 0.96	3.1 ± 0.1	32.4 ± 0.8	9.6 ± 0.4	15.92 ± 0.26	170468	0.85

TABLE 4.9: *MC mixing method* for $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ at 13 TeV. (*) indicates the sub-channel with the minimum value of $N_{\text{prod}}/w_{\text{Gauss}}$.

Decay Chain	**Calc$^{4\pi}$ 13 TeV [%]**	$\epsilon_{\text{GEN\|CUT}}$ **[%]**	ϵ_{GEN} **[%]**	ϵ_{CUT} **[%]**	w_{Gauss} **[%]**	N_{prod}	w_{fraction}
prompt D_s^+	87.58 ± 2.70	10.8 ± 0.3	88.6 ± 0.9	12.2 ± 0.4	88.5 ± 0.6	1043995	0.93
$\bar{b} \to D_s^+$	12.42 ± 0.38	1.23 ± 0.03	11.1 ± 0.3	11.2 ± 0.5	11.49 ± 0.08	125433^*	1.00

Chapter 5

Selection

The data samples used in this analysis are selected online by the L0, HLT1 and HLT2 trigger system, and offline by a dedicated stripping selection and by a set of preliminary cuts tailored for this analysis. The details of the offline selection applied to the $\tau^+ \to \mu^+\mu^-\mu^+$ and $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ modes are presented in the following. In addition, events are required to satisfy a specific trigger condition as discussed in Section 5.3.

The application of the offline selection has been carried out by the author, while the search for the most efficient trigger selection has been performed by a former PhD student of the analysis group.

5.1 Stripping and offline selection

The stripping lines used to select signal events in the signal and reference channel are the `StrippingTau23MuTau23MuLine` and `StrippingTau23MuDs2PhiPiLine`, respectively, both developed within the analysis working group. The aim of the stripping selection is to select well reconstructed tracks for the relevant decay by cutting on the vertex and track quality of the *mother* and *daughter* particles. The details of the stripping selection are reported in Table 5.1 for the two channels. In addition, well identified tracks are selected by cutting on particle identification variables (see Section 2.2.6). As pointed out later, the stripping selection applied in the signal channel only requires to at least two of the three final-state tracks to be identified as muons.

After the stripping selection, additional cuts are applied to specifically select the decay of interest both in the signal and in the reference channel. These cuts are summarised in Table 5.2. In the signal channel, a veto is applied to the ϕ mass to reject events from the $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ decay, as it represents a background source for the signal mode. The opposite requirement is instead applied to the reference channel to select the ϕ resonance. Events from the $D_s^+ \to \eta(\mu^+\mu^-\gamma)\mu^+\nu_\mu$ decay are rejected by requiring the invariant mass of the opposite-sign dimuon system $m_{\mu^+\mu^-}$ to be greater than 450 MeV/c^2; more details on the $D_s^+ \to \eta(\mu^+\mu^-\gamma)\mu^+\nu_\mu$ decay can be found in Section 8.1.1. The invariant mass of the same-sign dimuon system $m_{\mu^+\mu^+}$, instead, is required to be greater than 250 MeV/c^2 to reject cloned tracks. The known values of τ, D_s and $\phi(1020)$ mass are used as central values for all mass window cuts [8]. A set of cuts is also applied to remove badly reconstructed tracks; the angle α indicates here the orientation of the momentum of the τ (D_s) candidate with respect to the line connecting the primary and its decay vertex.

TABLE 5.1: Details of the stripping selection applied to the signal and reference channel. The invariant mass $m_{\mu\mu\mu}$ is meant as $m_{\mu\mu\pi}$ in the reference channel.

Stage	Variable	$\tau^+ \to \mu^+\mu^-\mu^+$	$D_s^+ \to \phi(\mu^+\mu^-)\pi^+$
μ pairs	$m_{\mu^+\mu^-}$	> 240 MeV$/c^2$, < 2000 MeV$/c^2$	> 970 MeV$/c^2$, < 1070 MeV$/c^2$
$\mu^\pm, \pi^+$	p_T	> 300 MeV$/c$	
	Track χ^2/ndf	< 3	
	IP χ^2/ndf	>9	
	track ghost probability	<0.45	
τ^+, D_s^+	$\lvert m_{\mu\mu\mu} - m_{\text{PDG}}\rvert$	<250 MeV$/c^2$	
	Vertex χ^2	<15	
	$c\tau$	>100 μm	
	Vertex IP χ^2	<225	

TABLE 5.2: Offline cuts applied to the signal and reference channel. The invariant mass $m_{\mu\mu\mu}$ is meant as $m_{\mu\mu\pi}$ in the reference channel.

Stage	Variable	$\tau^+ \to \mu^+\mu^-\mu^+$	$D_s^+ \to \phi(\mu^+\mu^-)\pi^+$
μ pairs	$\lvert m_{\mu^+\mu^-} - m_\phi\rvert$	> 20 MeV$/c^2$	< 20 MeV$/c^2$
	$m_{\mu^+\mu^-}$	> 450 MeV$/c^2$	-
	$m_{\mu^+\mu^+}$	> 250 MeV$/c^2$	-
τ^+, D_s^+	$\lvert m_{\mu\mu\mu} - m_{\text{PDG}}\rvert$	-	< 50 MeV$/c^2$
	$cos\alpha$	> 0.99	
	decay time	>-0.01 ns, <0.025 ns	

5.2 Selection of the three and two muon samples

In the analysis performed by LHCb on Run 1 data [3], the stripping line required all muon candidates to satisfy the `isMuon` condition (see Section 2.2.6). The stripping line `StrippingTau23MuTau23MuLine` used in this analysis, instead, requires that at least two of the three tracks in the final state satisfy the `isMuon` condition, meaning that also candidates where one track does not satisfy the `isMuon` condition can pass the stripping selection. Instead of rejecting *a priori* these candidates, they are included in the data sample in order to increase the available statistics.

The signal sample is therefore split into two sub-samples depending on the number of tracks with `isMuon=true` [1]. These are in the following referred to as the 2μ and the 3μ samples. According to the simulation, 30% of the signal candidates selected by the stripping and offline selection outlined in Section 5.1 belongs to the 2μ sample.

As discussed in the following, the 2μ sample is characterised by a peculiar kinematics. Figure 5.1 shows on the left-hand side a comparison between the momentum distributions obtained for the muons satisfying and not satisfying the `isMuon` condition. The same comparison is shown on the right-hand side of the same Figure for the transverse momentum of the two tracks. As can be seen, tracks not detected by

---

[1] In the following, `isMuon=true` is replaced with the shorter `isMuon`. The opposite condition is indicated with `!isMuon`.

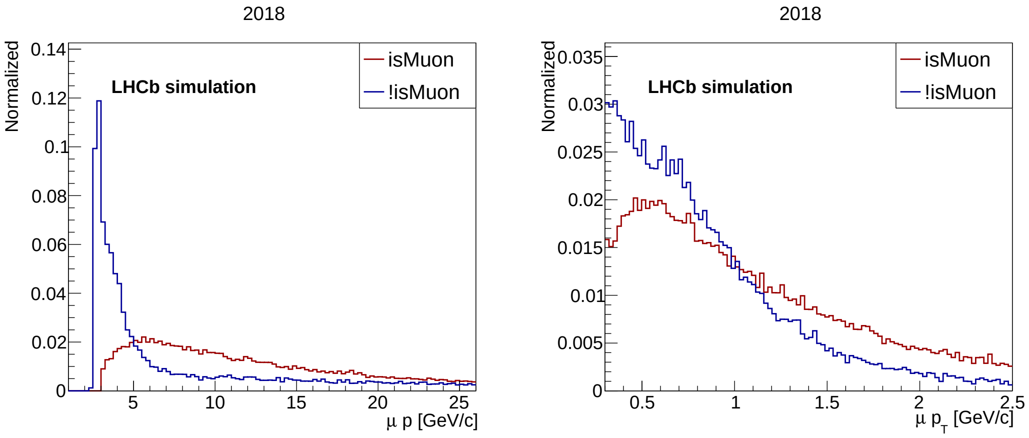


FIGURE 5.1: On the left, comparison between the momentum of the muon with `isMuon` and of the muon with `!isMuon`. On the right, same comparison for the transverse momentum. Both distributions are obtained for simulated events (2018), weighted with w_{fraction}.

the muon stations have in general very low momenta, compared to the tracks satisfying the `isMuon` requirement. A significant portion of these tracks have momentum lower than 3 GeV/c, the minimum momentum required to turn on the `isMuon` flag (see Table 2.1). By looking at the comparison in the p_T variable, it can be noticed that in the distribution obtained for the muons with `!isMuon` there are two contributions, one at very low p_T and one with p_T comparable to the p_T of muons satisfying the `isMuon` condition.

The presence of two contributions in the p_T distribution becomes more visible in Figure 5.2, where, on the right-hand side, the p_T distribution of the muons with `!isMuon` is shown as a function of the momentum p. It can happen that tracks produced at momentum higher than 3 GeV/c, and therefore eligible for the `isMuon` flag, also have a p_T such that they end up outside the muon detector acceptance before being able to trigger enough stations. On the other hand, there can be muons that, despite being produced at a sufficiently high momentum, have anyway too low p_T; these tracks are more easily affected by multiple scattering and end up in the beam pipe without generating enough hits in the muon stations. The comparison with the p_T vs. p distribution obtained for the muon satisfying the `isMuon` condition is shown on the left-hand side of Figure 5.2.

This feature of the 2μ sample will prove to be very useful in the signal over background discrimination, since, as shown in Figure 5.3, the two contributions are not visible in the p_T vs. p distribution obtained for the data outer sidebands.

Despite no information from the muon stations is available for these *bad* muons, the response of the RICH detectors can still be used to select signal events in the 2μ sample. As anticipated in Section 2.2.6, data acquired by the particle identification system are processed using a global likelihood approach. In the case of the RICH detectors, the pattern of hit pixels observed in the photodetectors is matched to that expected from the track reconstruction under a certain set of particle hypotheses. The PID information per track is then stored in the form of differences between the log-likelihood value for a given particle type hypothesis and a pion hypothesis for that track [77]. These differences are named RichDLLX, where X stands for kaon (K), muon (μ) or proton (p). From now on, the 2μ and 3μ samples will be treated separately and finally combined in the evaluation of the expected limit.

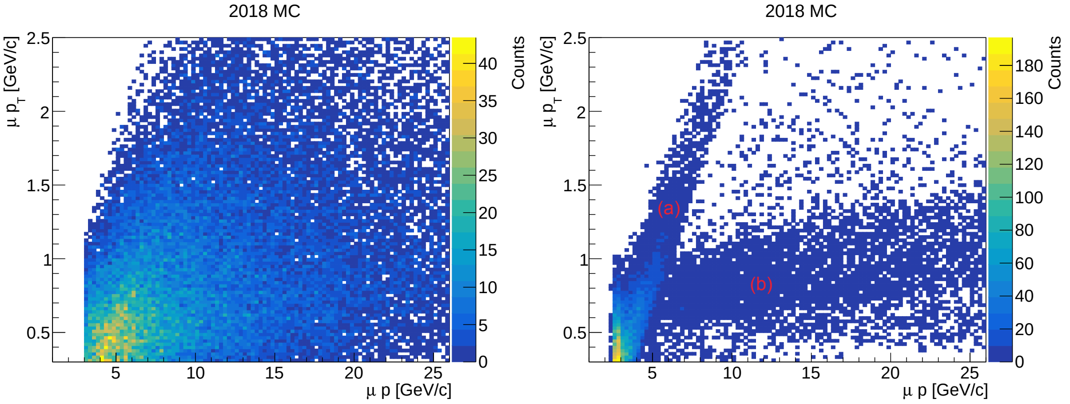


FIGURE 5.2: Transverse momentum versus momentum distribution for one of the muons with `isMuon` (left) and for the muon with `!isMuon` (right). (a) indicates the category of muons with `!isMuon` that, due to their large p_T, fall outside the muon detector acceptance. (b) indicates the category of muons with `!isMuon` that, due to their low p_T, are affected by muon scattering and end up in the beam pipe. Both distributions are obtained for 2018 simulated events, weighted with w_{fraction}.

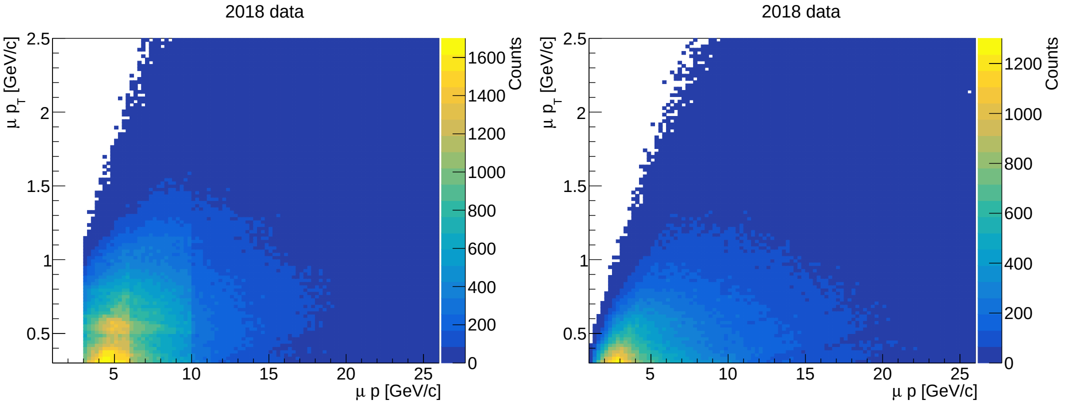


FIGURE 5.3: Transverse momentum versus momentum distribution for one of the muons with `isMuon` (left) and for the muon with `!isMuon` (right). Both distributions are obtained for events from the 2018 data outer sidebands.

5.3 Trigger selection

For both the signal and the reference channel, the choice of the trigger lines is performed on truth-matched events that passed the stripping and offline selection described in Section 5.1. The optimisation is performed separately for the 2μ and 3μ sample. The aim is to find the combination of trigger lines that maximises the efficiency of the selection.

At each stage of the trigger selection (L0, HLT1 and HLT2) the efficiency of the single line i (ϵ_i^{single}) and the efficiency of the `OR` between the line i and the line with highest efficiency (ϵ_i^{two}) are evaluated. The `OR` condition is requested between at most two lines in order to minimise the complexity of the selection. No significant improvement is observed in the total efficiency when adding a third line.

The single-line efficiency ϵ_i^{single} and the two-lines efficiency ϵ_i^{two} are computed as the ratio between the number of events that fire that specific trigger condition

TABLE 5.3: Trigger configuration for the signal and for the reference channel. For each trigger stage the specified lines are selected in `OR`. (*Not available in 2016, it is replaced with `Hlt2TopoMuMu3Body_TOS`)

Stage	**3μ sample**	**2μ sample**	**$D_s^+ \to \phi(\mu^+\mu^-)\pi^+$**
L0	`L0Muon_TOS` `L0DiMuon_TOS`		
HLT1	`Hlt1TrackMuon_TOS` `Hlt1DiMuonLowMass_TOS`		
HLT2	`Hlt2TriMuonTau23Mu_TOS`	`Hlt2TopoMuMu2Body_TOS` `Hlt2DiMuonDetached_TOS`*	

TABLE 5.4: L0 trigger requirements [78]. SPD indicates the number of hits in the scintillating-pad detectors constituting the calorimeter system.

L0 trigger	**p_T**			**SPD**
	2016	**2017**	**2018**	
`Muon`	> 1.8 GeV/c	> 1.35 GeV/c	> 1.75 GeV/c	< 450
`DiMuon`	> 2.25 (GeV/c)^2	> 1.69 (GeV/c)^2	> 3.24 (GeV/c)^2	< 900

and the number of events that passed the immediately prior selection. This means that the HLT1 lines efficiency is evaluated on events that fired the optimal L0-level configuration, just like the HLT2 lines efficiency is evaluated on events passing the most efficient HLT1 configuration. The optimal configuration is reported for each trigger level in Table 5.3.

The requirements of the chosen trigger lines are reported in Table 5.4 for L0, in Table 5.5 for HLT1 and in Table 5.6 for HLT2. The two signal sub-samples and the reference channel share the same L0 and HLT1 trigger selection.

The L0-muon trigger searches for straight-line tracks in the muon stations [78]. The track direction is used to estimate the p_T of the muon candidate. The p_T requirement is either applied to the muon candidate with the largest p_T (`L0Muon`), or to the product of the largest and second largest p_T values (`L0DiMuon`). The requirement on the SPD (scintillating-pad detectors) hits in the calorimeter system is applied in order to reduce the complexity of the event and speed up the reconstruction. In Run 2, the HLT1 muon lines are mainly organised as single muon and dimuon lines, both based on the same muon identification algorithm, *i.e.* `isMuon` [79]. Other cuts are applied on the kinematics and track quality variables.

A dedicated HLT2 trigger line was introduced in Run 2 to select 3μ candidates and can therefore be used in the corresponding sample. For the 2μ sample and for the reference channel, the `OR` between two HLT2 lines selecting two good muon candidates is instead required. The `Hlt2DiMuonDetached_TOS` line is not available for 2016 and it is therefore replaced with the `Hlt2TopoMuMu3Body_TOS` line. Note that all HLT2 trigger lines listed in Table 5.6 require the muon identification to either two or three final-state tracks. The muon identification is based on the `isMuon` requirement and on cuts on the ProbNNmu variable (see Section 2.2.6).

TABLE 5.5: HLT1 trigger requirements [80, 81].

HLT1 trigger	`DiMuonLowMass`	`TrackMuon`
$p(\mu)$	> 3 GeV/c	> 6 GeV/c
$p_T(\mu)$	> 0 GeV/c	> 1.1 GeV/c
Track χ^2/ndf	< 4	< 3
ProbNNghost	< 0.2	< 0.2
IP χ^2	> 4	> 35
`isMuon`	true	true
Other	opposite charge or $m_{\mu^+\mu^+} > 220$ MeV/c^2	VELO track compatible with hits in muon stations

TABLE 5.6: HLT2 trigger requirements [80]. Both "topological" trigger lines, `TopoMuMu2Body` and `TopoMuMu3Body`, rely on a multivariate classifier.

HLT2 trigger	**Cuts**
`TriMuonTau23Mu`	$\lvert m_{\mu\mu\mu} - m_{\tau,PDG}\rvert < 225$ MeV/c^2 $m_{\mu\mu} > 2 \times m_{\mu,PDG} + 14$ MeV/c^2 Vertex $\chi^2 < 25$ $c\tau > 45$ μm
`TopoMuMu2Body`	BDT classifier > 0.99
`DiMuonDetached`	$p_T > 600$ MeV/c p_T(at least one μ) > 300 MeV/c Vertex $\chi^2 < 9$ IP $\chi^2 > 9$ Decay length $\chi^2 > 7$
`TopoMuMu3Body`	BDT classifier > 0.99

The selected trigger lines are all TOS lines (Triggered On Signal). The inclusion of TIS (Triggered Independently of Signal) lines is found to bring no significant improvement to the trigger efficiency. More details on the definition of TOS and TIS lines can be found in Section 3.3.

Chapter 6

Calibration of the simulated sample

This analysis makes use of multivariate classification models to discriminate between signal and background events. In the training of the multivariate classifiers, simulated events are used as a proxy for the signal, while events from the data outer sidebands are used as a proxy for the background. The simulation, however, is not perfect and data and simulation often show non-negligible discrepancies. Therefore, it is important to align the simulation to the data in order to avoid any bias in the training of the classifiers.

The $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ decay is used to check for data/MC differences and derive necessary correction factors that are applied to $\tau^+ \to \mu^+\mu^-\mu^+$ simulated events in the training of the classifiers. In addition, the $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ decay is exploited in the calibration of the $\tau^+ \to \mu^+\mu^-\mu^+$ mass shape.

The evaluation of the corrections for the data/MC differences in the kinematic and track quality variables (Section 6.2) and the calibration of the particle identification variables (Section 6.3) have been performed by other members of the analysis group. The calibration of the signal mass shape (Section 6.4) has been carried out by the author.

6.1 Background subtraction

Despite being efficient in the identification of real $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ events, the selection applied in the reference channel can still accept events whose candidates are built from an accidental combination of tracks wrongly associated with the decay products of a D_s particle. This background contribution is commonly known as combinatorial background.

Since the $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ is used to evaluate corrections for possible differences between simulation and data, it is important to disentangle true $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ events from this background component. This is done by means of the *sPlot* technique [82], a statistical method that is able to unfold the distributions of several contributions coexisting in a single data sample, exploiting one or more *discriminating variables* for which the distribution of all the sources of events are known. This technique returns a set of *per-event* weights, called *sWeights*, that quantify how likely it is for a given event to belong to a certain contribution.

In this case, the coexisting contributions are the signal and the combinatorial background components. The $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ invariant mass distribution is used as a discriminating variable, where the aforementioned components can be easily identified as shown in the following.

The signal peak in the reference channel $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ is described by a Johnson's S_U[1] function [83].

$$J(x,\mu,\lambda,\gamma,\delta) = \frac{\delta}{\lambda\sqrt{2\pi}} \frac{1}{\sqrt{1+\left(\frac{x-\mu}{\lambda}\right)^2}} exp\left[-\frac{1}{2}\left(\gamma + \delta sinh^{-1}\left(\frac{x-\mu}{\lambda}\right)\right)^2\right] \tag{6.1}$$

An extended maximum likelihood fit [84] is performed to the $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ invariant mass obtained for simulated events (combining both polarities) to fix the shape parameters γ and δ. The fit is then performed on the mass distribution obtained for recorded data adding an exponential function to model the combinatorial background.

$$\text{pdf}_{tot}(x) = N_{sig} \cdot J(x,\mu,\lambda,\gamma,\delta) + N_{comb.bkg} \cdot F(x,\alpha) \tag{6.2}$$

with

$$F(x,\alpha) = exp(\alpha \cdot x)\,. \tag{6.3}$$

The number of signal and background events are extracted from the fit to the data distributions and the *sWeights* are computed separately for each data-taking year. The results of the fit to the invariant mass distribution in simulation and data are shown in Figure 6.1 for the three years.

---

[1] S_U stands for "unbounded system".

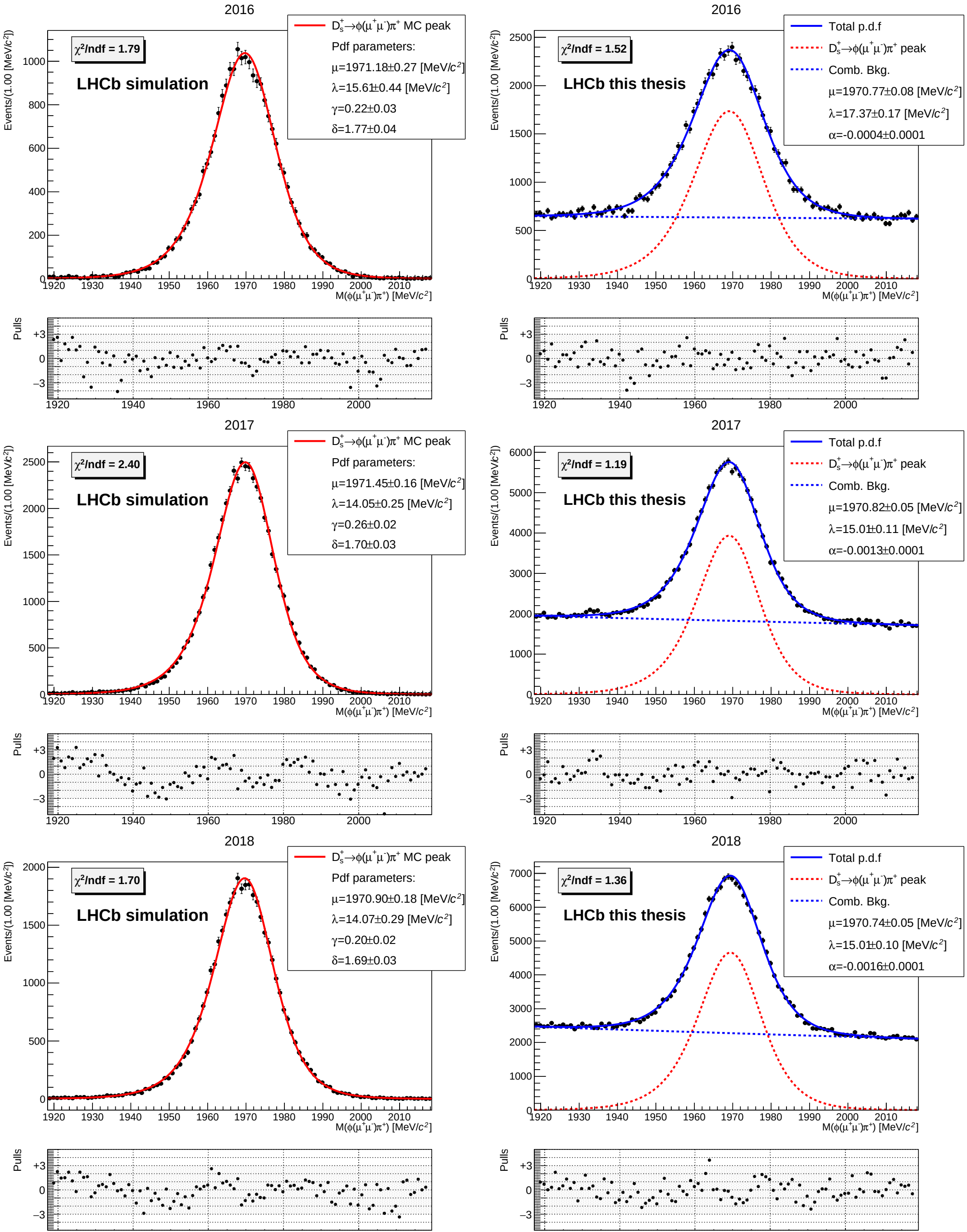


FIGURE 6.1: Fit to the $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ invariant mass distribution obtained for simulated events (left) and recorded data (right). Simulated events are weighted with w_{fraction}. The mass shape is modelled using a Johnson's SU function to describe the signal peak and an exponential distribution to described the combinatorial background in data. The fit parameters γ and δ describing the signal peak in data are fixed from the fit to simulated events. The plots are shown for the three data-taking years.

TABLE 6.1: Input variables of the GBReweighter model trained to reduce data/MC differences.

Variable	Definition
ln(D_s_P)	Logarithm of the D_s momentum
ln(D_s_PT)	Logarithm of the D_s transverse momentum
nTracks	Number of tracks in the event
D_s_ENDVERTEX_CHI2	χ^2 fit of the decay vertex
ln(D_s_IPCHI2_OWNPV)	Logarithm of the χ^2 of the impact parameter with respect to the primary vertex of the D_s track
arccos(D_s_DIRA_OWNPV)	Angle between the D_s momentum and the vector from the primary to the decaying vertex

6.2 Corrections for data/MC differences

As previously stated, some variables are not well reproduced by the simulation. The reference channel $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ can be exploited to check for differences between data and simulation and evaluate the necessary corrections. A *reweighting* technique [85] is used to generate a set of *per-event* weights that minimise the discrepancies between data and simulation. These weights are evaluated by comparing and adapting the data and MC distributions of one or more variables.

Available in the *hep_ml* package [86], a GBReweighter model based on Boosted Decision Trees (see Section 3.4) is trained on the reference channel $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$. To avoid unbiased predictions the *k-fold* technique is used [87]. Both the MC and data samples are split into *k folds* and then, iteratively, the model trained on $k-1$ *folds* is applied to the remaining one to estimate its performance. The variables used as input features to the GBReweighter are listed in Table 6.1.

Once the *k-folded* models are trained and tested on the reference channel, their average is taken and applied to simulated events in the 3μ and 2μ samples to derive the corrections. The resulting *per-events* weights, referred to as $w_{\text{Data/MC}}$, are applied to simulated events in order to minimise the discrepancies between data and simulation in the training phase of the classifiers used to discriminate between signal and background events. Figure 6.2 shows the comparison between the $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ data and MC distributions of the variables used as input features to the GBReweighter, before and after applying the $w_{\text{Data/MC}}$ corrections. To appreciate the validity of the method, the same comparison is also shown for the reduced χ^2 of the final state tracks fit (right column).

Henceforth simulated events are weighted with w_{fraction} and $w_{\text{Data/MC}}$; w_{fraction} is itself weighted with a correction, referred to as w_{fracCorr}, applied in order to minimise differences between data and simulation in the fractional contribution of the different τ and D_s production channels. The reason for this correction and its evaluation is discussed in Appendix A.

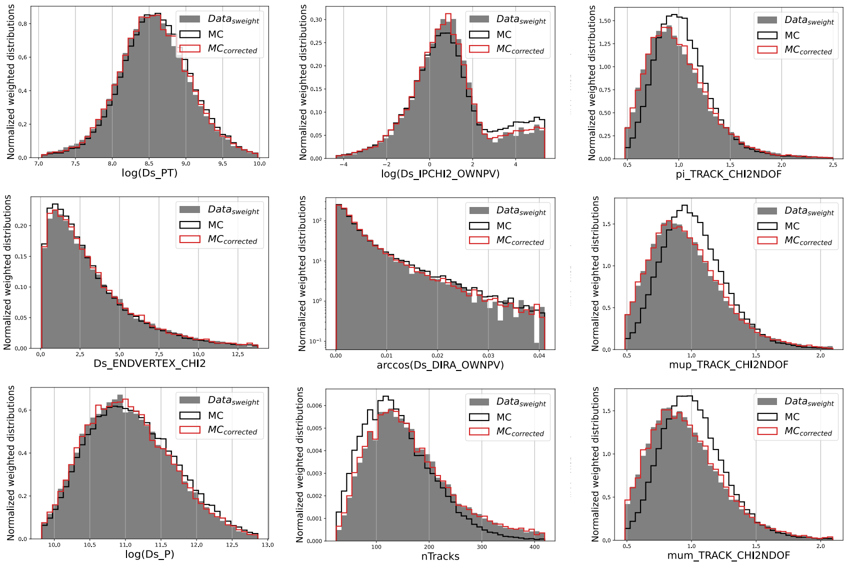


FIGURE 6.2: Distributions of the kinematic and track quality variables used as input features to the GBReweighter in data and simulation, before (black) and after (red) applying the $w_{\mathrm{Data/MC}}$ corrections. The plots are generated for 2017 data.

6.3 Calibration of particle identification variables

Both in the 2μ and 3μ sample, a multivariate classifier based on particle identification (PID) information is trained on simulated candidates to reject background arising from misidentified (mis-ID) tracks. Like kinematic and track quality variables, also PID variables are not accurate enough in simulation and need to be corrected to account for differences between data and simulation. Due to its similar decay topology, the $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ channel can be again exploited to evaluated the PID corrections applied to $\tau^+ \to \mu^+\mu^-\mu^+$ simulated events in the training phase of the PID classifier.

The calibration of the PID variables consists in creating a 4-dimensional (4D) binned map and extracting a global *per-event* weight as the ratio between the maps filled with $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ data and simulated events. The set of variables in which the maps are built differs depending on the PID variable that is being calibrated. The weights are then assigned to simulated events, both in the signal and in the reference channel. The details of the strategy adopted for the calibration of the ProbNNmu variable, in the 2μ and 3μ sample, and of the RichDLLmu variable, in the 2μ sample only, are discussed in the following.

6.3.1 Calibration of ProbNNmu

For the calibration of the ProbNNmu variable the 4D maps are built in bins of the following variables:

- min(ProbNNmu): minimum ProbNNmu value among the muons in the decay;
- p_μ: momentum of the muon with minimum ProbNNmu;
- η_μ: pseudorapidity of the muon with minimum ProbNNmu;
- *nTracks*: track multiplicity in the event.

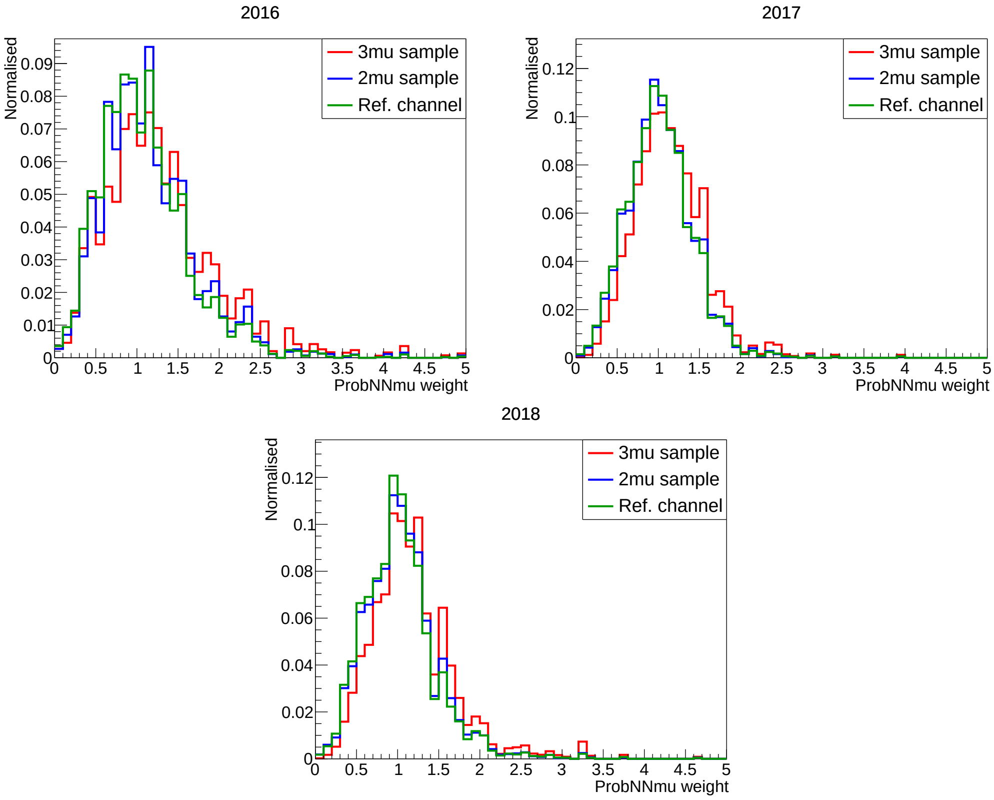


FIGURE 6.3: Distribution of the weights correcting for data/MC differences in the ProbNNmu variable in the reference channel, 3μ-signal and 2μ-signal MC samples. The weights are extracted from the data/MC ratio of the 4D maps determined on the reference channel for the three data-taking periods.

The binning scheme in the min(ProbNNmu) variable is chosen to have a similar number of events in each bin. The binning scheme in the remaining three variables differs for each min(ProbNNmu) bin in order to take into account the correlation with the min(ProbNNmu) variable, and is as well chosen such that all bins have similar statistics. The final binning scheme is made of 10 bins in min(ProbNNmu) in its whole range [0.,1.0], 10 bins in p_μ in the range [0., 500.0] MeV/c, 5 bins in η_μ in the range [1.5, 5.0], and 2 bins in *nTracks* in the range [0, 1000].

The 4D map is filled first with $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ data and then with $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ simulated events, and for each $\tau^+ \to \mu^+\mu^-\mu^+$ simulated candidate the final weight is extracted as the ratio between the two maps, using the bin value determined according to the four variables information of that candidate. In order to take into account the correlation between the D_s invariant mass and the PID variable, the *sWeights* (see Section 6.1) are again computed in each min(ProbNNmu) bin to correctly suppress the combinatorial background affecting the data sample.

The procedure followed to extract the correction to the ProbNNmu variable is the same for the 3μ and 2μ samples. The distribution of the ProbNNmu corrections are shown for the two sub-samples and for the reference channel in Figure 6.3; the plots are reported for the three data-taking years.

The choice of building the map in the ProbNNmu variable of only one of the three muons in the decay chain allows to evaluate a single *per-event* weight instead of three (two in the 2μ sample) *per-track* weights. Moreover, the reweighting obtained with this strategy does not aim at correcting directly the ProbNNmu variable, rather at adapting the efficiency of the simulated sample in bins of min(ProbNNmu) to what is observed in data. As shown in Figure 6.4, it turns out that also other PID variables, such as ProbNNk, are well described after this reweighting.

6.3.2 Calibration of RichDLLmu

In the 2μ sample, only two of the final-state tracks are detected in the muon stations, meaning that for one track no ProbNNmu variable is available. As discussed in Section 5.2, this PID information can be replaced with the one provided by the RichDLLmu variable, based on the response of the RICH detectors.

The reference channel is used to extract the correction to the RichDLLmu variable with a procedure similar to the one used for the calibration of the ProbNNmu variable. The stripping line used to select signal events in the reference channel, however, requires both muons in the final state to activate the `isMuon` flag. The RichDLLmu variable of the muon not satisfying the `isMuon` requirement in the 2μ sample is calibrated considering for each event the RichDLLmu of the muon with the lowest value of ProbNNmu. In fact, as shown in Figure 6.5, the distribution of the RichDLLmu variable for the track with `!isMuon` is similar to the RichDLLmu distribution obtained for the track with the lowest value of ProbNNmu. The 4D maps are therefore generated in bins of:

- RichDLLmu of the muon track with minimum ProbNNmu;
- p_μ: momentum of the muon with minimum ProbNNmu;
- η_μ: pseudorapidity of the muon with minimum ProbNNmu;
- *nTracks*: track multiplicity in the event.

The binning scheme in the RichDLLmu variable consists of 10 bins with same statistics in the range [-150.0, 200.0]. In each of the 10 bins, the binning in the remaining three variables is analogous to the one chosen for the calibration of the ProbNNmu variable.

The effect of the RichDLLmu correction on simulated events is shown in Figure 6.4 for one of the final-state muons. The distributions of the RichDLLmu corrections extracted for the reference channel and for the 2μ simulated samples are shown in Figure 6.6 for the three years.

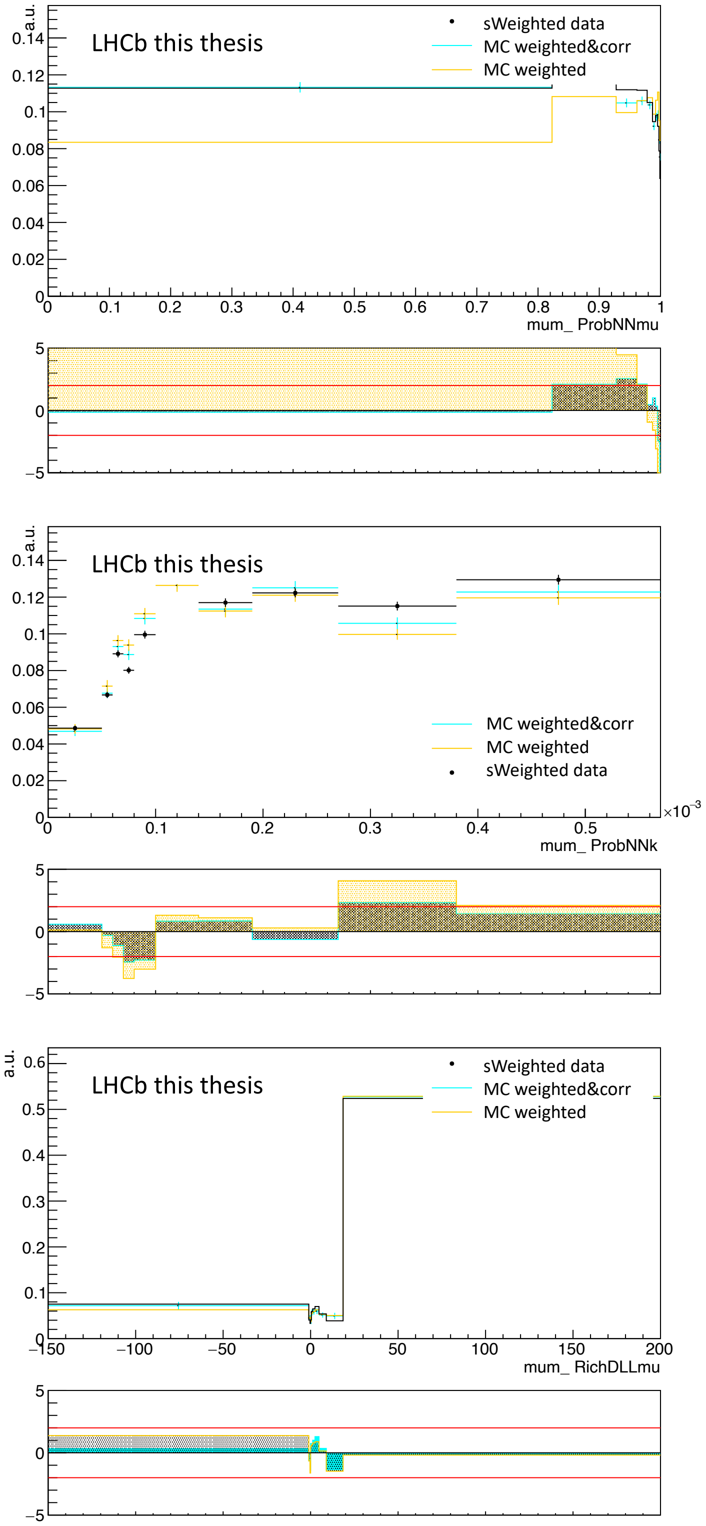


FIGURE 6.4: Comparison between the ProbNNmu (top), ProbNNk (middle) and RichDLLmu (bottom) distributions obtained for 2018 $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ *sWeighted* data and simulated events with and without applying the PID corrections.

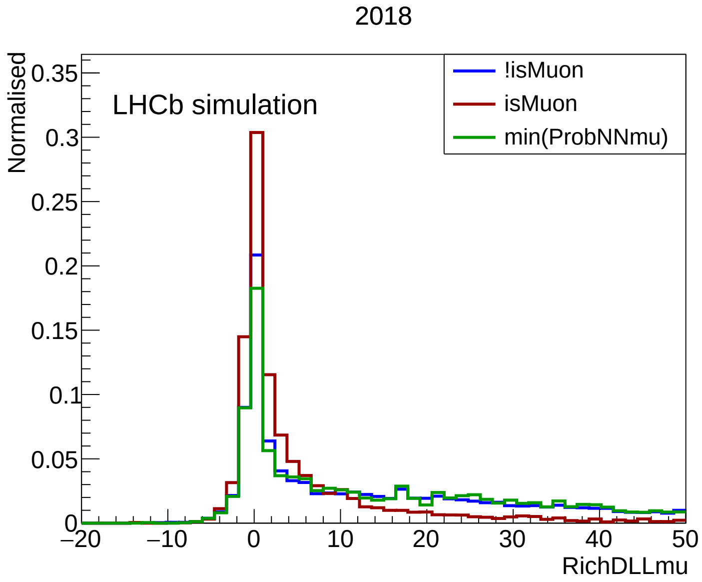


FIGURE 6.5: Comparison between the RichDLLmu distribution obtained for 2018 simulated events for the track with `!isMuon`, for the track with `isMuon` and for the track with min(ProbNNmu) in the 2μ sample.

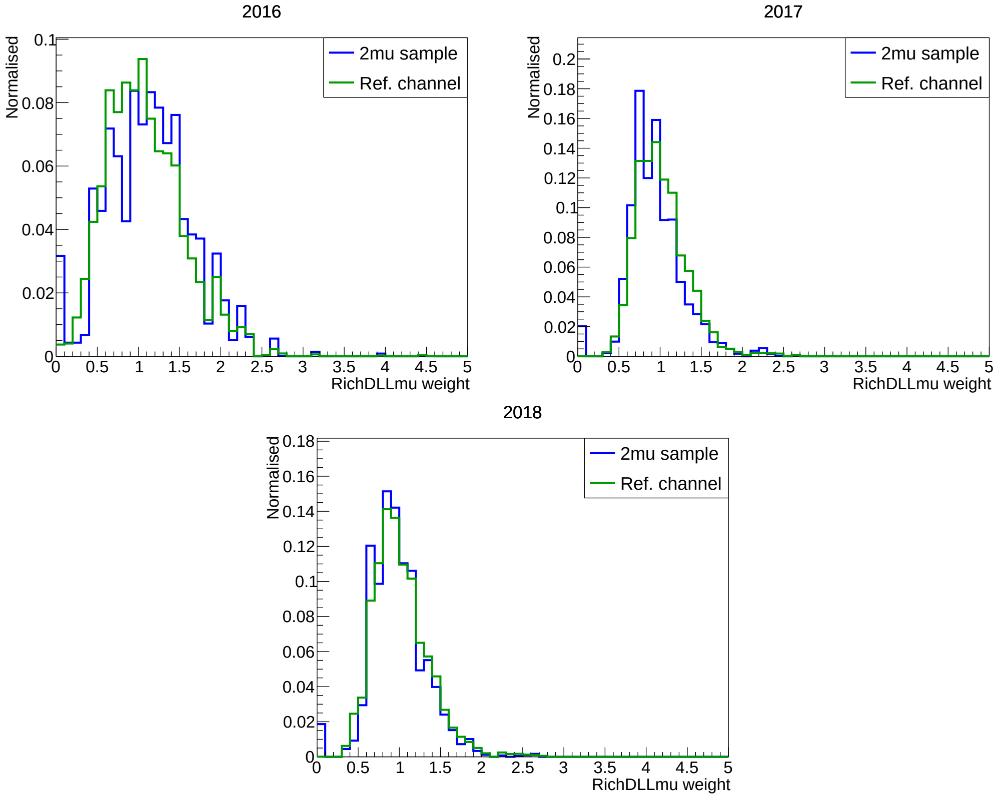


FIGURE 6.6: Distribution of the weights correcting for data/MC differences in the RichDLLmu variable in the reference and 2μ-signal MC samples. The weights are extracted from the data/MC ratio of the 4D maps determined on the reference channel for the three data-taking periods.

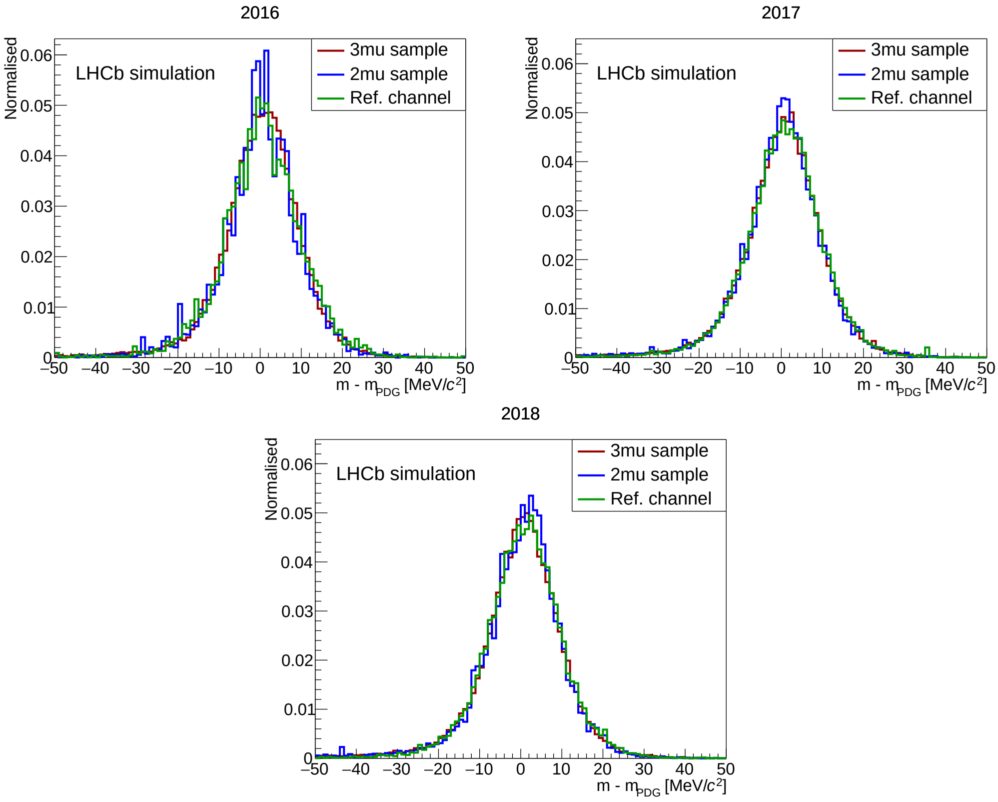


FIGURE 6.7: Comparison between the invariant mass distribution obtained for simulated candidates in the reference channel, in the 2μ and in the 3μ sample after the trigger selection. All distributions are subtracted from the τ and D_s PDG mass value in the signal and reference channel, respectively, to facilitate the comparison of the invariant mass resolution in the two channels.

6.4 Calibration of signal mass shape

The reference channel $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ is also used to extract the shape of the invariant mass distribution in the signal channel. As shown in Figure 6.7, the invariant mass distributions obtained for simulated events in the reference channel, in the 3μ and in the 2μ samples all have similar resolutions. To facilitate the comparison, the difference between the reconstructed mass of the τ (D_s) candidate and the τ (D_s) nominal mass is shown. The plots shown in Figure 6.7 are obtained for events passing the offline and trigger selection and weighted with w_{fraction}, w_{fracCorr} and $w_{\text{Data/MC}}$.

As mentioned in Section 6.1, the invariant mass spectrum of the reference channel is modelled using a Johnson's S_U function (see Eq. 6.1). A fit is performed on the invariant mass distribution of the reference channel MC sample after the trigger selection and the result is shown for each data-taking year on the left-hand side of Figure 6.8. The Johnson's S_U function is combined with an exponential distribution to describe the combinatorial background (see Eq. 6.2) and the result of the fit to the distribution in data is shown on the right-hand side of Figure 6.8.

The parameters of the fit performed on $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ data and simulated events are summarised in Table 6.2 for the three years.

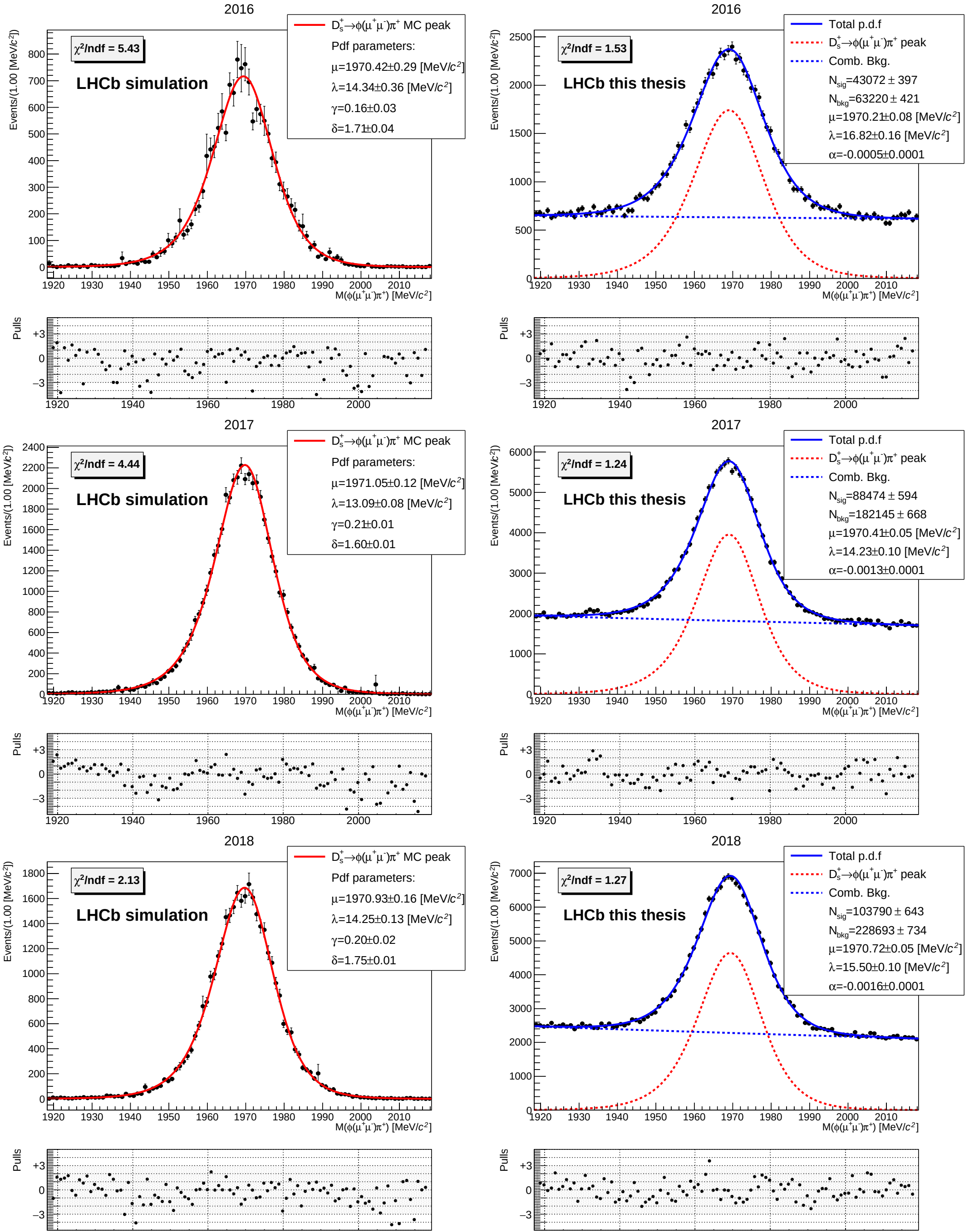


FIGURE 6.8: Fit to the $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ invariant mass distribution obtained for simulated events (left) and recorded data (right). Simulated events are weighted with w_{fraction}, w_{fracCorr} and $w_{\text{Data/MC}}$. The mass shape is modelled using a Johnson's S_U function to describe the signal peak and an exponential distribution to describe the combinatorial background in data. The fit parameters γ and δ describing the signal peak in data are fixed from the fit to simulated events. The plots are shown for the three data-taking years.

TABLE 6.2: Results of the fit to the invariant mass distribution of the reference channel in simulation and data. The parameters γ and δ are fixed from the fit to the invariant mass distribution in simulation.

	2016		**2017**		**2018**	
	MC	**Data**	**MC**	**Data**	**MC**	**Data**
μ **[MeV/c^2]**	1970.42 ± 0.29	1970.21 ± 0.08	1971.05 ± 0.12	1970.41 ± 0.05	1970.93 ± 0.16	1970.72 ± 0.05
λ **[MeV/c^2]**	14.34 ± 0.36	16.82 ± 0.16	13.09 ± 0.08	14.2 ± 0.1	14.23 ± 0.10	15.5 ± 0.1
γ	0.16 ± 0.03		0.21 ± 0.01		0.20 ± 0.02	
δ	1.71 ± 0.04		1.60 ± 0.01		1.75 ± 0.01	
α	-	-0.0005 ± 0.0001	-	-0.0013 ± 0.0001	-	-0.0016 ± 0.0001
χ^2/n_{dof}	5.43	1.53	4.44	1.24	2.13	1.27

TABLE 6.3: Correction factors of the fit parameters μ and λ of the function used to model the $\tau^+ \to \mu^+\mu^-\mu^+$ invariant mass in simulation.

	2016	**2017**	**2018**
c_μ	0.99989 ± 0.00015	0.99968 ± 0.00007	0.99989 ± 0.00008
c_λ	1.173 ± 0.032	1.087 ± 0.010	1.087 ± 0.012

From the result of the fit to the invariant mass distribution of the reference channel, it can be observed that the MC and data distributions show different resolutions. To extrapolate the signal pdf used later in the limit evaluation, the central values obtained from the fit to the MC mass shape of $\tau^+ \to \mu^+\mu^-\mu^+$ need to be corrected for the factors $c_\mu = \mu_{data}/\mu_{MC}$ and $c_\lambda = \lambda_{data}/\lambda_{MC}$ evaluated on the $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ channel. The values of the correction factors c_μ and c_λ are reported in Table 6.3 separately for the three years.

Chapter 7

Signal and background discrimination

Both in the 3μ and in the 2μ sample, the background rejection is performed by means of two multivariate classifiers. The first one, trained on kinematic variables and topological information, is used to suppress the combinatorial background. The second one is instead trained on particle identification (PID) information to suppress background originating from tracks misidentified as muons.

The multivariate classifiers employed in the 3μ sample to discriminate between signal and background events (Section 7.2) are the result of the work of a former PhD student of the analysis group. The binning in the invariant mass distribution in the 3μ sample (Section 7.2.4) and the separation between signal and background events in the 2μ sample (Section 7.3) have been carried out by the author.

7.1 Isolation variables

Among the kinematic and topological information that can be exploited in the discrimination of signal and background events, there is a category of isolation variables, originally developed for the $B^0_{(s)} \to \mu^+\mu^-$ analysis [88], that turn out to be very powerful in the suppression of the combinatorial background. The isolation of the muon tracks is indeed very relevant as the combinatorial background mainly arises from semileptonic *b*- or *c*-hadron decays, where other charged particles produced in the same event are reconstructed close to the signal muon candidate and therefore wrongly associated with the τ decay. In other words, the more isolated the three muon tracks are from the other tracks of the event, the more likely it is that these muons originated from a τ decay.

The isolation variables quantify how close other tracks are to the muon candidate (excluding the other two muon candidates). Depending on whether the muon isolation is evaluated against a long or a VELO track, the corresponding variable is called LONG or VELO isolation variable, respectively. The proximity of each muon candidate to another track, whether long or VELO track, is measured using a dedicated Boosted Decision Tree (BDT) classifier[1] with slightly different input features for the two categories. Both BDTs return an isolation score value for each track-muon pair in the event; the "closer" is the track to the muon candidate, the higher the output value of the classifiers. For each muon candidate, the maximum value of a given classifier over all the track-muon pairs in the event is taken ($I(\mu^+)$). The isolation variable is built, for both categories, as the sum of the maximum score between two of the three muon candidates ($I(\mu^+) + I(\mu^\pm)$). Given that in the signal channel there

---

[1] The BDT is developed using the *Adaboost* boosting algorithm available in the TMVA toolkit [62].

are three muons in the final state, for each of the two categories three BDT isolation variables are computed.

The isolation variables are calibrated for tracks identified as muons. Therefore, in the 2μ sample, only the LONG and VELO isolation variables evaluated for the tracks with `isMuon` can be used in the suppression of the combinatorial background [2]. For both categories (LONG and VELO), three types of isolation variables are saved.

7.2 Signal and background separation in the 3μ sample

In the 3μ sample, the discrimination of signal over background events is performed by means of two multivariate classifiers, one trained on kinematic and topological information to suppress the combinatorial background, and one trained on PID information against background originating from misidentified tracks. As mentioned in Section 3.4, both classifiers are based on a binary *eXtreme Gradient Boosting* (XGB) model [65], a scalable machine-learning technique for tree boosting.

To distinguish the two classifiers, the one based on PID information is named SuperPNN (or SuperProbNN), while the other one is simply referred to as XGBoost (or XGB). The features and performances of the two classifiers are described in the following.

7.2.1 XGBoost classifier

The XGBoost model trained against the combinatorial background uses $\tau^+ \to \mu^+\mu^-\mu^+$ simulated events as a signal proxy and data events from the middle sidebands around the blinded region as a proxy for the background. The 26 variables used as input features to the classifier are listed in Table 7.1, while in Figure 7.1, the distributions of these variables are shown for 2018 data (middle sidebands) and simulated events. The final LONG and VELO isolation variables used in the training of the classifier are computed as average among the BDT isolation variables of the three pairs of muons in the final state.

Three hyperparameters of the XGBoost model, namely the *learning_rate*, the maximum depth of a tree (*max_depth*) and the number of estimators (*n_estimators*), are tuned using a cross-validated *grid search* method[3]. In order to exploit all the events available in the data set the *k-fold* technique introduced in Section 6.2 is used. The search for the optimal set of hyperparameters is performed on a training set corresponding to 80% of the total data set. The set of hyperparameters corresponding to the configuration providing the greatest balanced accuracy (efficiency to identify each class averaged among the classes) are selected as the "best parameters". Once trained, the model is applied to the test set (20% of the total data set) to evaluate its performance.

The simulated events used in the training of the classifier are weighted with w_{fraction} (see Section 4.2.2), w_{fracCorr} (see Appendix A) and $w_{\text{Data/MC}}$ (see Section 6.2), and with an additional weight, referred to as w_{balance}. Its role is to balance the model training by correcting for the different size of the MC and data samples. Identical for all simulated events, w_{balance} is simply derived as the ratio between the number of background events and the weighted sum of simulated events.

---

[2] The BDT isolation variables are filled with 0 for the track not satisfying the `isMuon` condition.

[3] In a grid search, the search for the optimal set of hyperparameters is performed through a manually specified subset of the hyperparameter space of the targeted algorithm [89].

TABLE 7.1: Input features to the XGBoost model.

Variable	**Definition**
tau_DIRA_OWNPV	Cosine of the angle between the τ momentum and the vector from the primary to the decaying vertex
tau_ENDVERTEX_CHI2	χ^2 fit of the decay vertex
tau_TAU	Particle decay time
μ_TRACK_CHI2NDOF	χ^2 of the track fit divided by the degrees of freedom, for each of the daughter particles
IPCHI2_OWNPV	Difference in the vertex-χ^2 of the PV reconstructed with and without the particle, for τ and daughters particles
PT	Transverse momentum of τ and daughters particles.
tau_ConeMultMuon	Number of tracks within a cone built around a muon track
tau_ConePtAsymMuon	Asymmetry between the transverse momentum of a muon and the sum of the transverse momenta of all the particles within a cone built around that muon
LONGMAX1(2,3)	Average of LONG isolation variables computed for each pair of daughter particles
VELOMAX1(2,3)	Average of VELO isolation variables computed for each pair of daughter particles

TABLE 7.2: XGB classifier average performance obtained with the chosen set of hyperparameters quantified by different metrics. The last two scores are the signal efficiency obtained requiring a background rejection of 95% and the background rejection obtained requiring a signal efficiency of 85%. The results are averaged over the predictions made on the 5 *folds* (train:test ratio is 80%:20%).

Scores [%]	**2018**	**2017**	**2016**
Accuracy	91.26 ±0.14	90.90 ±0.15	91.83 ±0.20
Signal efficiency (ϵ_S)	90.69 ±0.18	90.37 ±0.29	89.76 ±0.41
Background rejection ($1-\epsilon_B$)	91.83 ±0.16	91.44 ±0.11	92.91 ±0.10
AUC	97.28 ±0.07	97.05 ±0.04	97.39 ±0.12
ϵ_S @ $1-\epsilon_B = 95\%$	95.45 ±0.12	94.95 ±0.15	95.54 ±0.17
$1-\epsilon_B$ @ $\epsilon_S = 85\%$	85.99 ±0.34	84.88 ±0.32	86.35 ±0.40

The performance of the classifier is expressed in terms of different metrics, such as the *signal efficiency* ϵ_S, the *background rejection* defined as $1-\epsilon_B$, and the AUC, namely the area under the *receiver operating characteristic* (ROC) curve. A summary of the classifier performance is reported in Table 7.2. For each year, the final performance is evaluated as the average across the 5 *folds*.

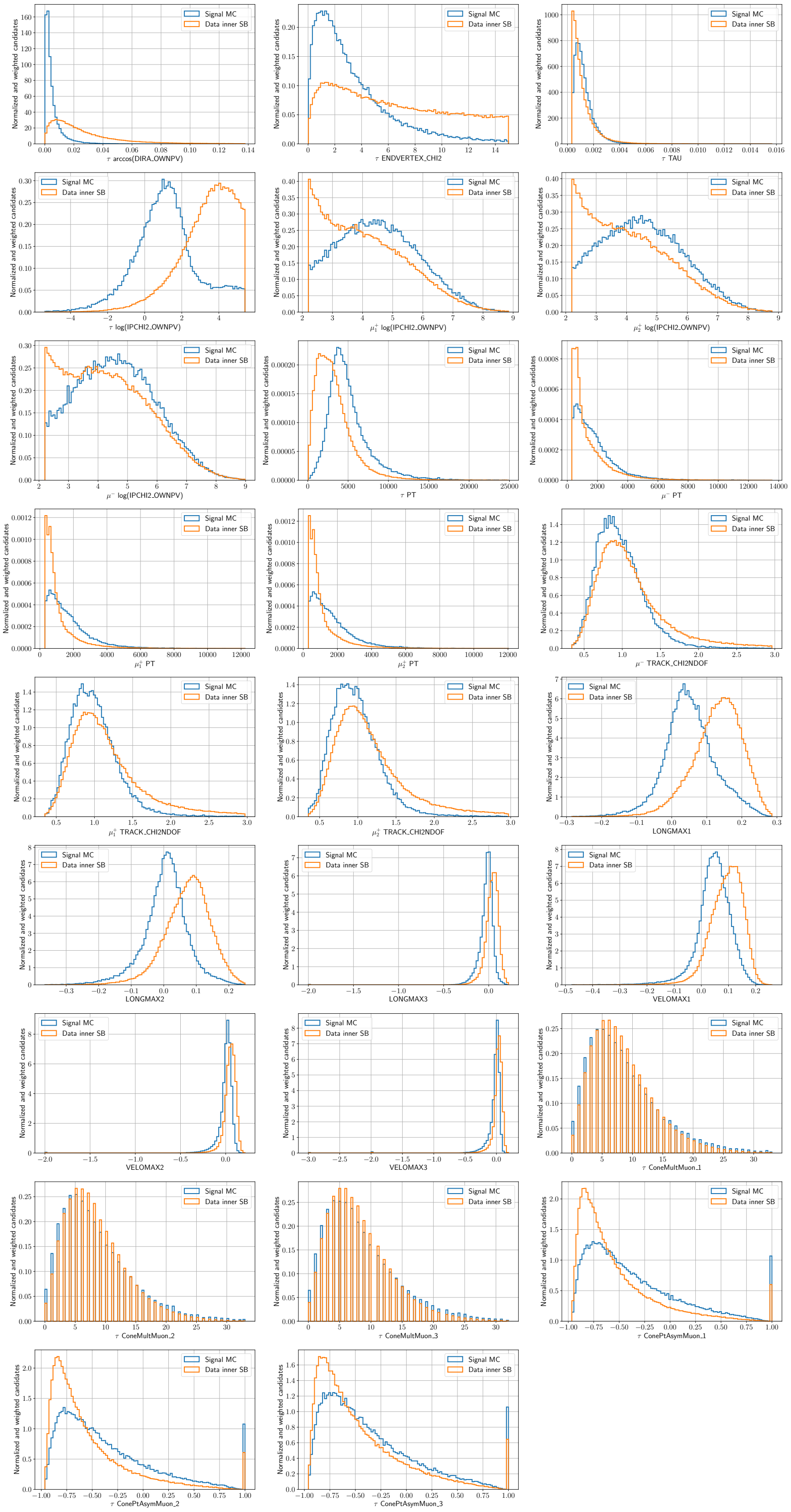


FIGURE 7.1: Distributions of the input features of the XGBoost classifier, for signal (simulation) and background (data middle sidebands). The histograms are normalised to unit area.

Calibration of the XGBoost output

Once trained and tested, the XGBoost model is again applied to signal simulated events and to the data outer sidebands. In order to check for possible differences between data and simulation, the same model is applied to the reference channel $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$. As already mentioned, the simulated events used as a signal proxy in the training of the classifier are weighted with $w_{\mathrm{Data/MC}}$ to minimise differences between data and simulation. Nevertheless, the distribution of some of the input features of the classifier produced for $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ simulated events, still show some discrepancies with respect to the distributions obtained in data (see Figure 7.2). Because of the presence of two single final-state muons, the $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ isolation variables are computed, for each category (LONG and VELO), as average between the isolation variables available for the two muon candidates.

The corrections for the residual discrepancies between data and simulation in the output of the XGBoost classifier are evaluated as the ratio between the data and MC distributions obtained for the reference channel, both normalised to unity. The ratio, shown in Figure 7.3 for all the three data-taking years, is evaluated in 30 bins whose edges are computed in such a way that all bins have the same amount of simulated events. Each signal and reference channel simulated event will be weighted with the bin content corresponding to its XGB value.

The output of the XGB classifier is shown for the signal channel in simulation and data, separately for the inner and outer sidebands, in Figure 7.4. In the same plots, the distributions obtained for $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ *sWeighted* data and simulated events, are also shown. Note that simulated events are here weighted also with the PID corrections discussed in Section 6.3.1 (w_{PID}) and with the correction for data/MC differences in the XGBoost classifier output (w_{XGBcorr}).

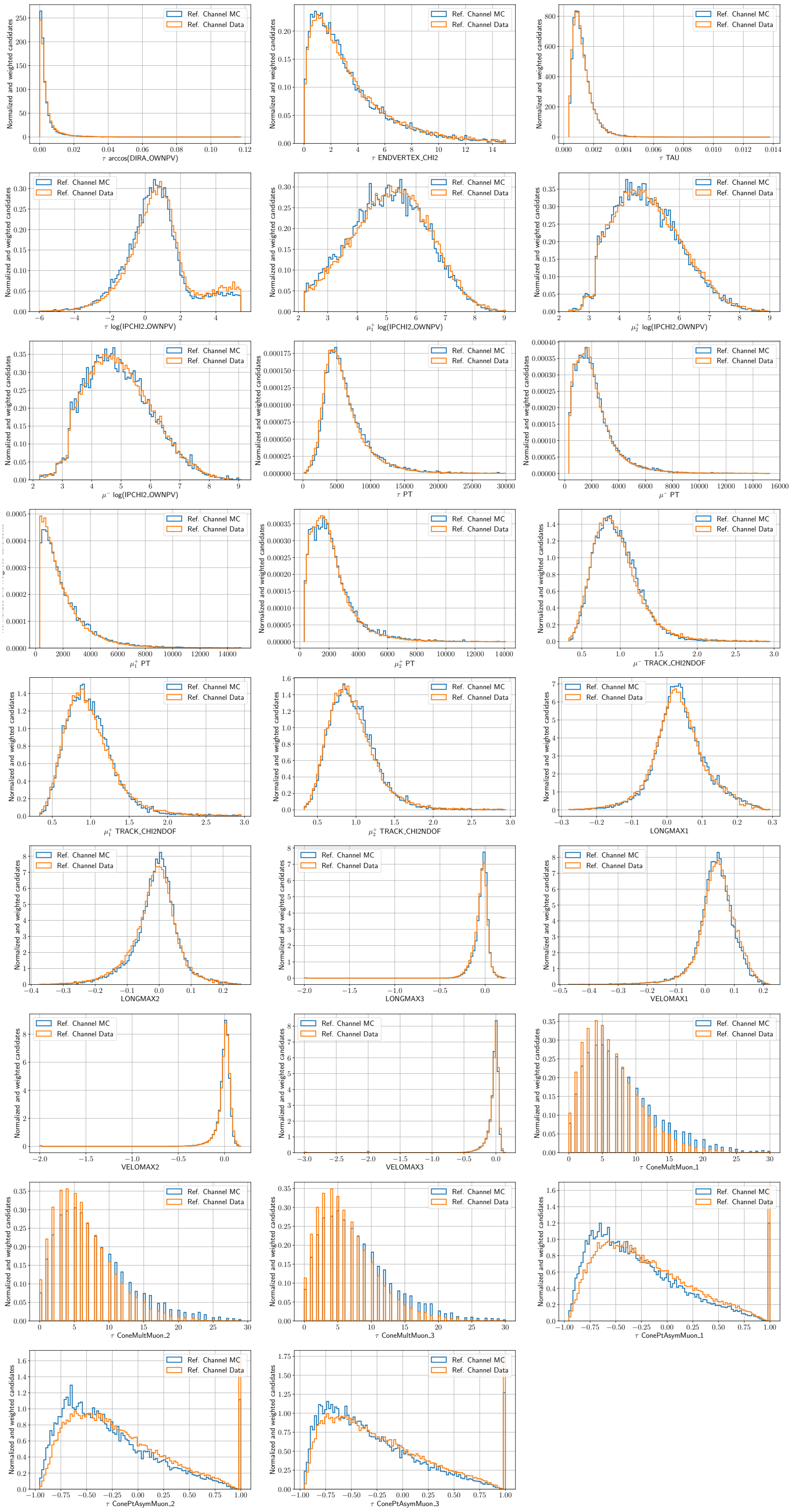


FIGURE 7.2: Distributions of the input features of the XGBoost classifier for $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ *sWeighted* data and simulated events. The features of the particles $D_s^+, \pi^+, \mu^-, \mu^+$ are matched with the ones of $\tau^+, \mu_1^+, \mu^-, \mu_2^+$. The distributions obtained for the pion are labeled as μ_1^+ in the x-axis. The histograms are normalised to unit area.

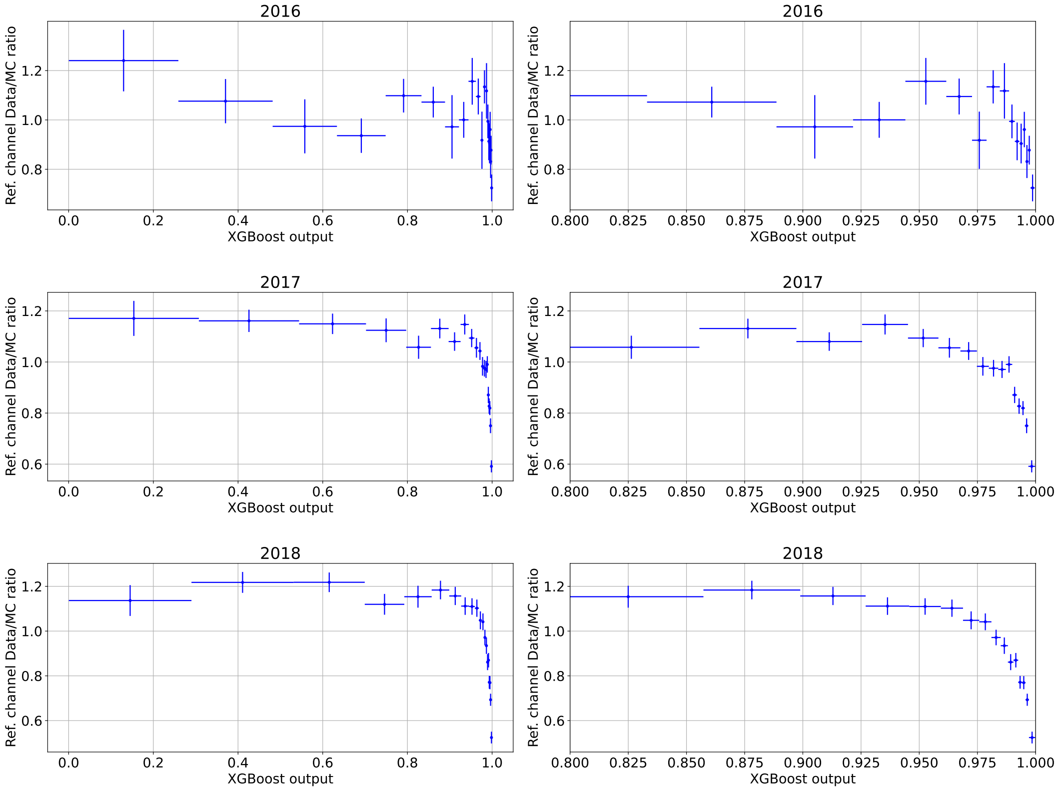


FIGURE 7.3: On the left, bin-by-bin ratio between the data and MC distribution of the classifier output used to reweight the signal and reference channel simulated events. On the right, zoom in of the region at high values of the XGBoost output. The bins edges are computed to have uniform content in the simulated sample.

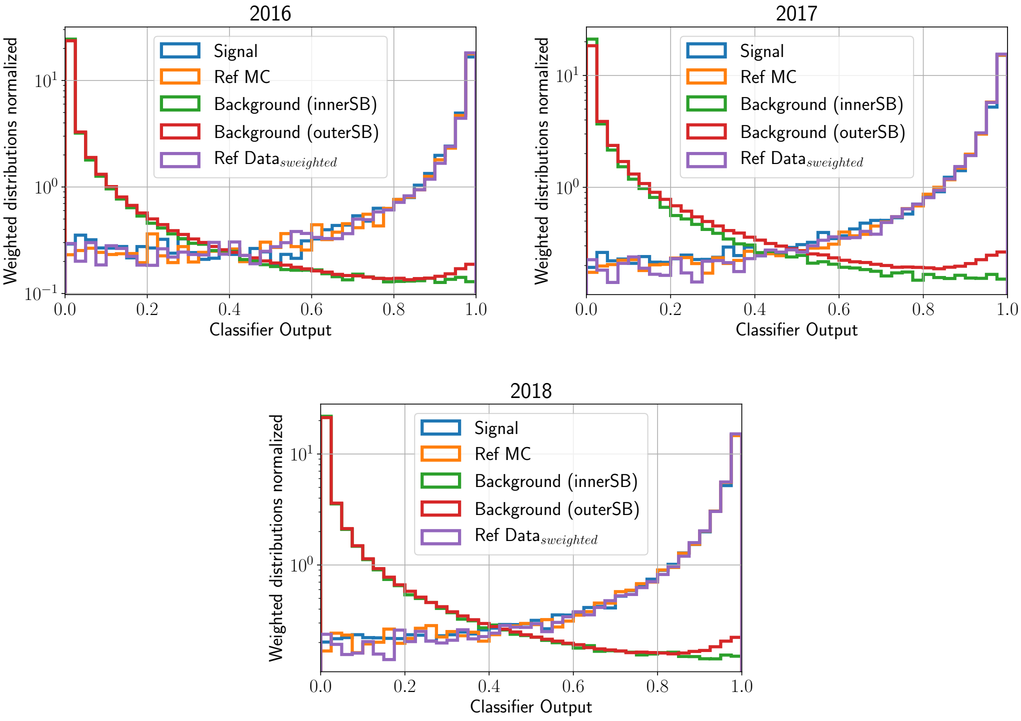


FIGURE 7.4: XGBoost output distributions obtained for the signal channel simulation and data (inner and outer sidebands), and for the reference channel simulation and *sWeighted* data. Simulated events are weighted with w_{fraction}, w_{fracCorr}, $w_{\text{Data/MC}}$, w_{XGBcorr}, and with the PID corrections introduced in Section 6.3.1. Distributions are normalised to unity.

TABLE 7.3: Input features to the SuperPNN model.

Variable	Definition
ProbNNmu	Probability of the tracks to be identified as a muon
ProbNNk	Probability of the tracks to be identified as a kaon
η	Track's pseudorapidity

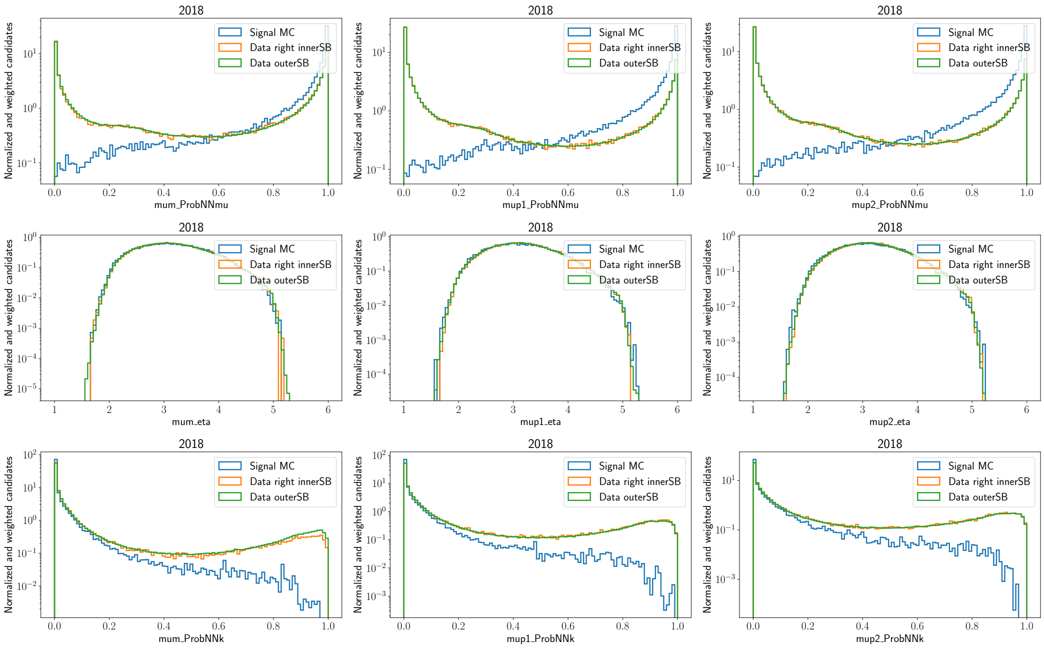


FIGURE 7.5: Distribution of the input features of the SuperPNN classifier, for the signal (simulation) and background (right inner sideband) sample. The histograms are normalised to unit area.

7.2.2 SuperPNN classifier

With a strategy similar to the one implemented for the suppression of the combinatorial background, a second XGB classifier is trained on PID information to suppress background originating from misidentified tracks. The variables given as input to the classifier are listed in Table 7.3.

The classifier, called SuperPNN (or SuperProbNN), is trained using $\tau^+ \to \mu^+\mu^-\mu^+$ simulated events as a signal proxy, and data events from the right middle sideband as a proxy for the background. The left middle sideband is excluded in order to minimise the contribution from the mis-ID background $D^+ \to K^-\pi^+\pi^+$. As later shown in Section 8.2, the $D^+ \to K^-\pi^+\pi^+$ mode is expected to peak in the left inner sideband. Simulated events are weighted with w_{fraction}, w_{fracCorr}, $w_{\text{Data/MC}}$ and w_{PID}. The distributions of the input features for simulated events and for the data sidebands are reported in Figure 7.5 for the 2018 sample.

The optimal set of hyperparameters (*learning_rate*, *max_depth*, *min_sample_leaf*, *n_estimators*) are selected as the ones providing the highest balanced accuracy. To make use of whole data sample, the training and testing of the classifier is performed by means of the *k-fold* technique. The 5 *folds* result into five independent models.

TABLE 7.4: SuperPNN classifier average performance obtained with the chosen set of hyperparameters quantified by different metrics. The last two scores are the signal efficiency obtained requiring a background rejection of 95% and the background rejection obtained requiring a signal efficiency of 85%. The results are averaged over the predictions made on the 5 *folds* (train:test ratio is 80%:20%).

Scores [%]	**2018**	**2017**	**2016**
Accuracy	88.86 ±0.38	89.17 ±0.38	86.75 ±1.25
Signal efficiency (ϵ_S)	82.53 ±0.69	84.67 ±0.62	77.65 ±2.00
Background rejection ($1-\epsilon_B$)	96.77 ±0.12	96.35 ±0.13	97.35 ±0.03
AUC	97.97 ±0.08	97.92 ±0.04	97.81 ±0.20
ϵ_S @ $1-\epsilon_B = 95\%$	89.04 ±0.39	89.31 ±0.18	88.59 ±0.57
$1-\epsilon_B$ @ $\epsilon_S = 85\%$	96.21 ±0.14	96.22 ±0.14	95.94 ±0.32

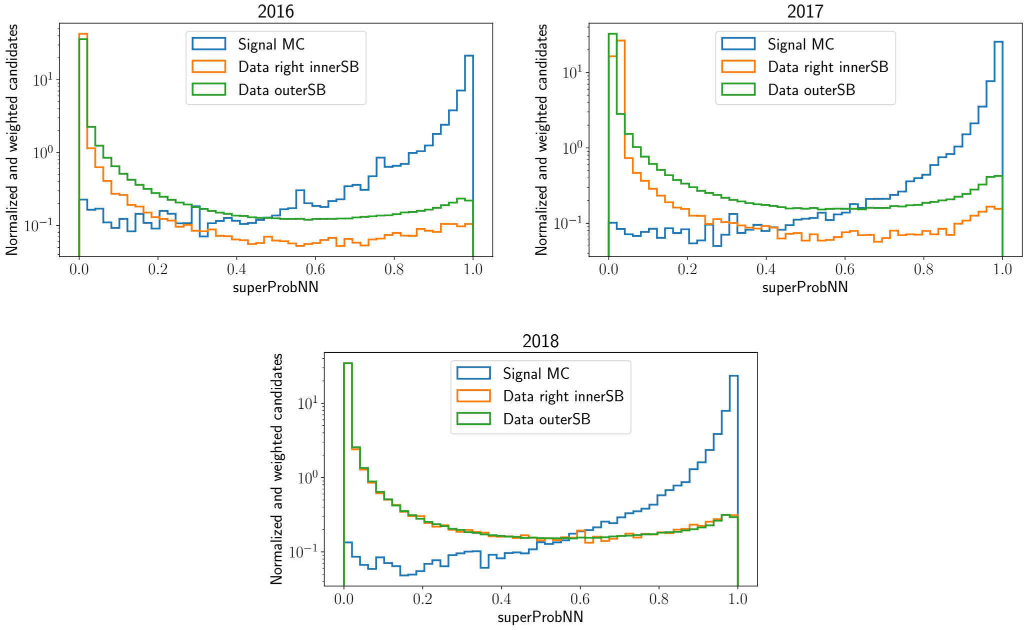


FIGURE 7.6: SuperPNN output distributions obtained for the signal channel simulation and data (separately for the right inner sideband and outer sidebands). Simulated events are weighted with w_{fraction}, w_{fracCorr}, $w_{\text{Data/MC}}$ and w_{PID}. Distributions are normalised to unity.

The performance of the SuperPNN classifier is again expressed in terms of the same metrics used for the XGB classifier, and the resulting values are reported in Table 7.4.

Data events from the outer sidebands are assigned the average of the five SuperPNN values predicted for the five different models. The output of the SuperPNN classifier is shown for signal simulated events and for the data sidebands (separately for the right inner sideband and for the outer sidebands) in Figure 7.6 for the three data-taking years.

Calibration of the SuperPNN output

Once trained and tested, the SuperPNN model is applied to the reference channel to evaluate the correction for a potential disagreement between data and simulation.

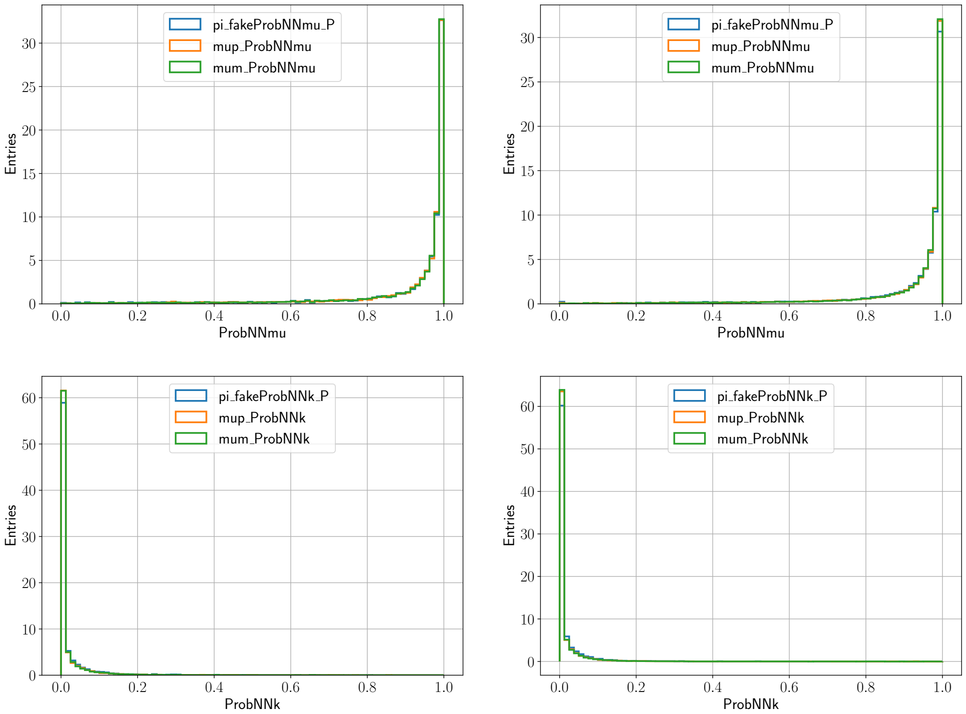


FIGURE 7.7: On the left, MC distributions of the ProbNNmu and ProbNNk variables for the two muons and the reconstructed ones for the pion. On the right, same distributions obtained for *sWeighted* data. The plots are produced for the 2018 data sample.

The SuperPNN model is trained to discriminate events with three real muon tracks from events where one or more of the final-state tracks have been wrongly identified as muons. In order to be able to apply the SuperPNN model to the reference channel, it is clear that the PID information of the pion cannot be used as it is. The followed strategy consists in replacing the ProbNNmu and ProbNNk variables of the pion with *fake* ProbNNmu and ProbNNk, made up from the PID information of the two opposite-sign muons. Separately for data and simulated events, the pion momentum distribution is divided into 10 bins with same statistics. In each of these bins, the histograms with the ProbNNmu and ProbNNk of the muon with charge opposite to the one of the mother particle are produced. The *fake* ProbNNmu and ProbNNk assigned to the pion is randomly extracted from the histogram produced in the same momentum range. The distributions of the ProbNNmu and ProbNNk variables for the muons and of the reconstructed ones for the pion are reported in Figure 7.7 for simulated and data events.

The ProbNNmu and ProbNNk variables reconstructed for the pion are then used as input features, together with the PID information of the two muons, in the application of the SuperPNN model to the reference channel. The ratio between the SuperPNN distribution obtained for *sWeighted* data and the one obtained for simulated events is then used to extract the corrections to the SuperPNN output assigned to $\tau^+ \to \mu^+\mu^-\mu^+$ and $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ simulated events to minimise data/MC discrepancies. The ratio is evaluated in 20 bins with almost uniform content in the simulated sample and its distribution is shown in Figure 7.8 for the three data-taking years.

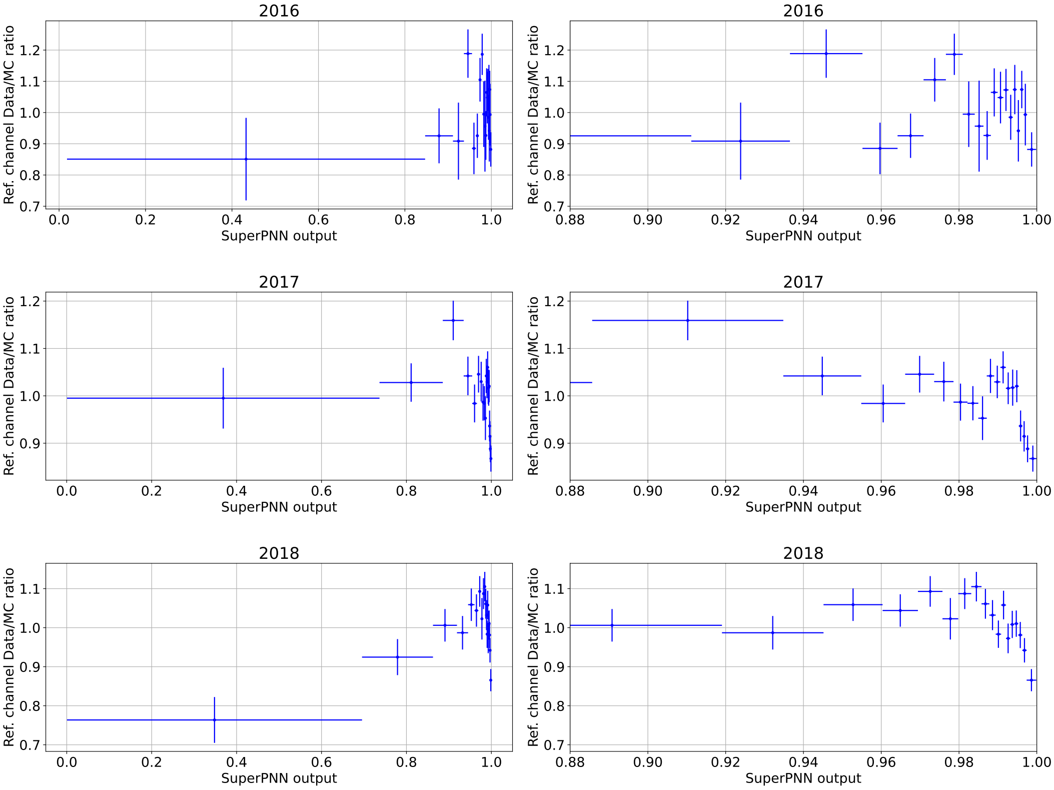


FIGURE 7.8: On the left, bin-by-bin ratio between the data and MC distribution of the classifier output used to reweight the signal and reference channel simulated events. On the right, zoom in of the region at high values of the SuperPNN output. The bins edges are computed to have uniform content in the simulated sample.

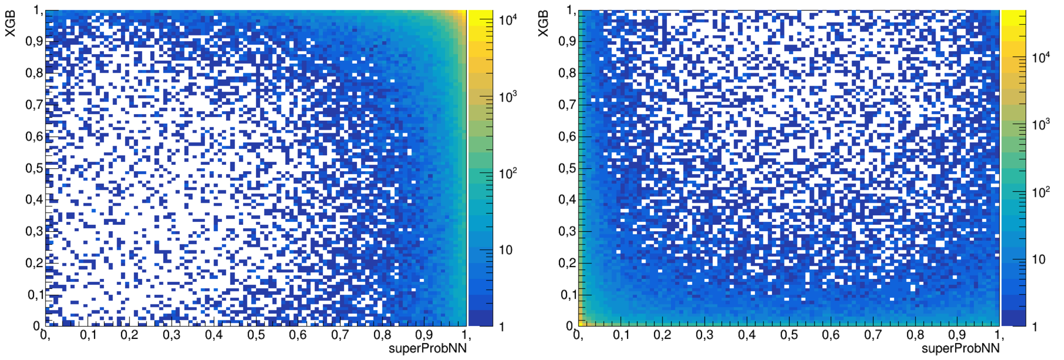


FIGURE 7.9: Correlation between the output of the XGBoost and of the SuperPNN classifiers for signal simulation (left) and for the right inner sideband in data (right). Plots are shown for the 2018 samples.

7.2.3 Binning scheme in XGBoost and SuperPNN

In Figure 7.9, the output of the XGB classifier is shown as a function of the output of the SuperPNN classifier, both for 2018 simulated events and for data events from the right inner sideband (similar distributions can be obtained for 2016 and 2017). No correlation is observed between the two classifiers, meaning that the selection in XGBoost would be independent of the selection in SuperPNN, and vice versa. Applying a tight cut on XGB and SuperPNN would certainly allow to get rid of most of the background contribution, but would have the disadvantage of cutting away potential signal candidates. In order to increase the sensitivity to the signal decay in the 3μ sample, the data set is split in bins of the XGBoost and SuperPNN classifiers output.

A technique developed for the $B_s^0 \to \mu^+\mu^-$ analysis and presented in Ref. [90], is used to find the optimal binning scheme in the two discriminating variables. The optimisation criteria is based on

$$\Delta LQ = 2\ln Q_{SB} - 2\ln Q_B, \tag{7.1}$$

where Q_{SB} and Q_B represent the *test statistics* computed for the Signal+Background (SB) and for the Background-only (B) hypothesis, respectively. The two *test statistics* are defined as

$$Q_{SB} = \prod_{i=1}^{n} \frac{\mathcal{P}(s_i + b_i, s_i + b_i)}{\mathcal{P}(s_i + b_i, b_i)}, \quad Q_B = \prod_{i=1}^{n} \frac{\mathcal{P}(b_i, s_i + b_i)}{\mathcal{P}(b_i, b_i)}, \tag{7.2}$$

with $\mathcal{P}(x, \lambda)$ being a Poissonian distribution of mean λ calculated in x, and $s_i + b_i$ (b_i) the expected number of candidates in the Signal+Background (Background) hypothesis in bin i. In Ref. [90] the authors demonstrate, through pseudo-experiments, that the separation power ΔLQ can be also expressed in terms of the median of the test statistics as

$$\Delta LQ \simeq 2\ln Q_{SB}^{med} - 2\ln Q_B^{med}. \tag{7.3}$$

A practical way to obtain this quantity without generating any pseudo-experiment is to calculate the two approximators $\widetilde{Q_{SB}^{med}}$ and $\widetilde{Q_B^{med}}$ from the definition of *test statistic*

TABLE 7.5: Binning scheme that optimises the signal event sensitivity to $\tau^+ \to \mu^+\mu^-\mu^+$ decays where all the three final-state tracks are identified as muons.

XGB	SuperPNN
$[0.800, 0.951]$	$[0.880, 0.972, 0.988, 0.994, 1.000]$
$[0.951, 0.983]$	$[0.880, 0.975, 0.989, 0.995, 1.000]$
$[0.983, 0.993]$	$[0.880, 0.979, 0.991, 0.995, 1.000]$
$[0.993, 1.000]$	$[0.880, 0.980, 0.990, 0.995, 1.000]$

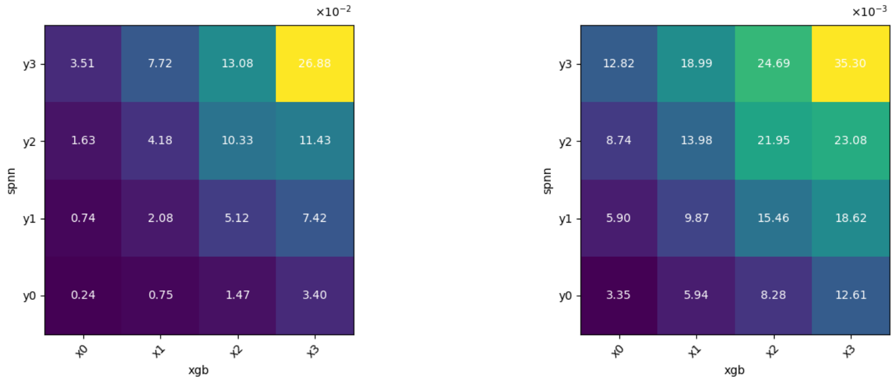


FIGURE 7.10: Individual bin score (left) and signal sensitivity (right) for the optimised binning scheme. The XGB and SuperPNN bins are displayed in the x and y axis, respectively, as defined in Table 7.5.

given in Eq. 7.2.

$$\widetilde{Q^{med}_{SB}} = \prod_{i=1}^{n} \frac{\mathcal{P}(s_i + b_i, s_i + b_i)}{\mathcal{P}(s_i + b_i, b_i)}, \quad \widetilde{Q^{med}_{B}} = \prod_{i=1}^{n} \frac{\mathcal{P}(b_i, s_i + b_i)}{\mathcal{P}(b_i, b_i)}. \tag{7.4}$$

The optimal binning is determined by maximising the ΔLQ quantity for the expected number of candidates in each bin for the two hypotheses. As no signal events are expected, the signal sample is normalised to 1 for the sake of the score calculation. The search for the optimal binning is performed following an adaptive approach, spanning from 4 (2x2) to 49 (7x7) bins with equal amount of expected signal candidates. To ensure the rejection of the most background-like candidates, the search for the optimal binning is performed on candidates with SuperPNN>0.88 and XGB>0.8. The final binning is reported in Table 7.5, while in Figure 7.10 the corresponding individual bin score and signal sensitivity are shown.

7.2.4 Binning in the invariant mass distribution

The shape of the invariant mass distribution in the signal channel is taken from a fit to the reference channel $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$, whose distribution is modelled with a Johnson's S_U function (see Eq. 6.1) as described in detail in Section 6.4. The signal pdf is extrapolated by fitting the MC invariant mass distribution and scaling the central values obtained from the fit by the correction factors c_μ and c_λ evaluated from the fit to the reference channel and reported in Table 6.3. The results of the fit to the MC invariant mass in the 3μ sample are shown on the left-hand side of Figure 7.11 for the three data-taking years; the fit parameters are reported in the legend. The fit

TABLE 7.6: Fraction of events falling in each mass bin evaluated on the extrapolated signal pdf for the three data-taking years. The fraction of events in the signal region is also reported (Total).

Mass bin	**2016**	**2017**	**2018**
1	0.0422 ± 0.0003	0.0387 ± 0.0003	0.0368 ± 0.0003
2	0.0781 ± 0.0007	0.0760 ± 0.0005	0.0721 ± 0.0006
3	0.1334 ± 0.0009	0.1380 ± 0.0007	0.1323 ± 0.0008
4	0.1903 ± 0.0006	0.2058 ± 0.0006	0.2022 ± 0.0007
5	0.2015 ± 0.0013	0.2165 ± 0.0010	0.2195 ± 0.0011
6	0.1503 ± 0.0005	0.1511 ± 0.0006	0.1567 ± 0.0007
7	0.0848 ± 0.0007	0.0774 ± 0.0006	0.0811 ± 0.0007
8	0.0411 ± 0.0004	0.0341 ± 0.0003	0.0360 ± 0.0004
Total	0.9217 ± 0.0012	0.9375 ± 0.0009	0.9366 ± 0.0010

is performed on simulated events selected by the minimum cut on XGB (>0.8) and SuperPNN (>0.88) and by the veto on ProbNNk (later discussed in Section 8.2.1). The extrapolated mass shapes are instead shown on the right-hand side of Figure 7.11 for the three years.

In addition to the sampling in the XGB and SuperPNN output, a further binning is considered in the signal region of the invariant mass distribution to increase the sensitivity to the signal decay. In this case, 8 bins with a width of 5 MeV/c^2 each are considered in the ± 20 MeV/c^2 invariant mass window around the PDG τ mass. The fraction of events falling in each mass bin is reported in Table 7.6 for the three years. Note that the fraction of events falling in each mass bin sum up to the fraction of events falling in the signal region (last row of Table 7.6).

As shown in Figures 7.12 and 7.13, no correlation is observed between the invariant mass distribution and the output of the XGB and SuperPNN classifiers. The fraction of events falling in each mass bin can therefore be evaluated before binning in the two variables.

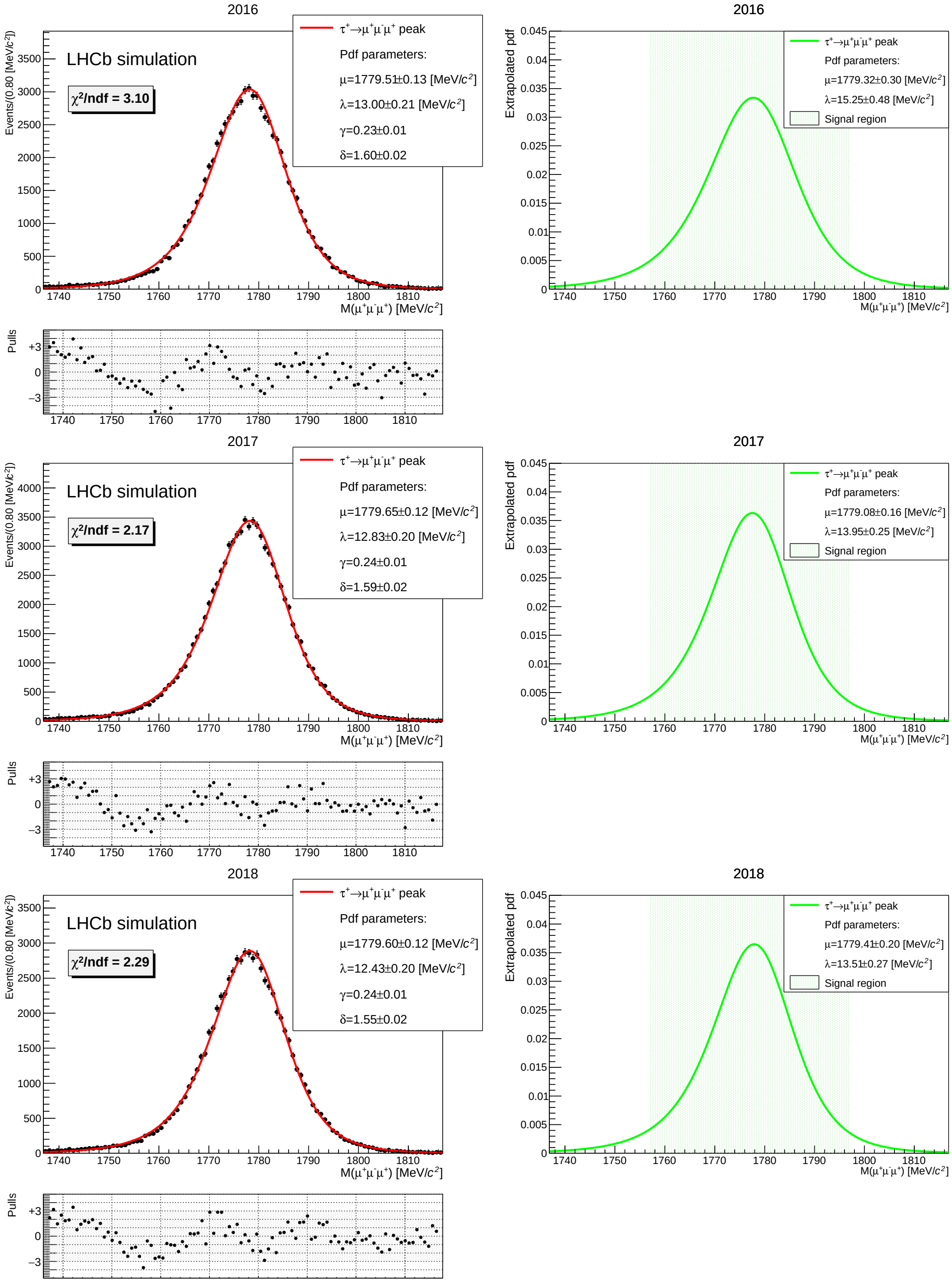


FIGURE 7.11: On the left, fit to the $\tau^+ \to \mu^+\mu^-\mu^+$ invariant mass distribution obtained for simulated events in the 3μ sample after applying the minimum cut on XGB and SuperPNN and the veto on ProbNNk. The mass shape is modelled with a Johnson's S_U function; the fit parameters are reported in the legend. On the right, extrapolated $\tau^+ \to \mu^+\mu^-\mu^+$ mass shape obtained from the fit to the MC invariant mass where the central values (μ and λ) are corrected from the fit to the reference channel. The plots are reported for the three data-taking years.

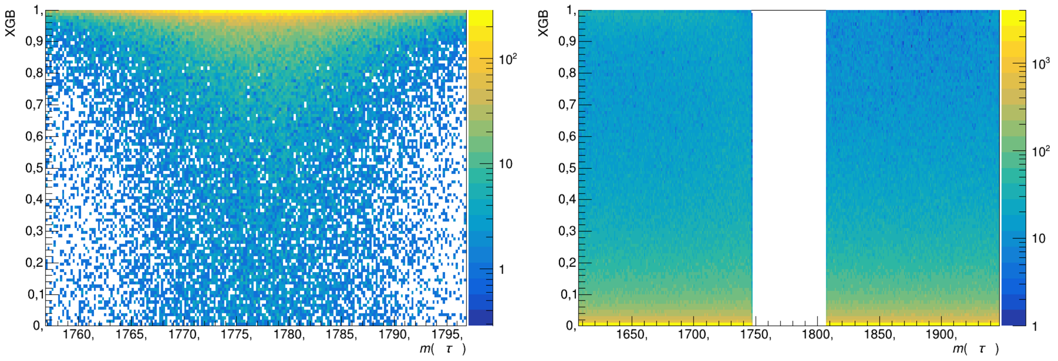


FIGURE 7.12: Correlation between the output of the XGBoost classifier and the invariant mass distribution for $\tau^+ \to \mu^+\mu^-\mu^+$ simulated events (left) and for the outer sidebands in data (right). Plots are shown for the 2018 samples.

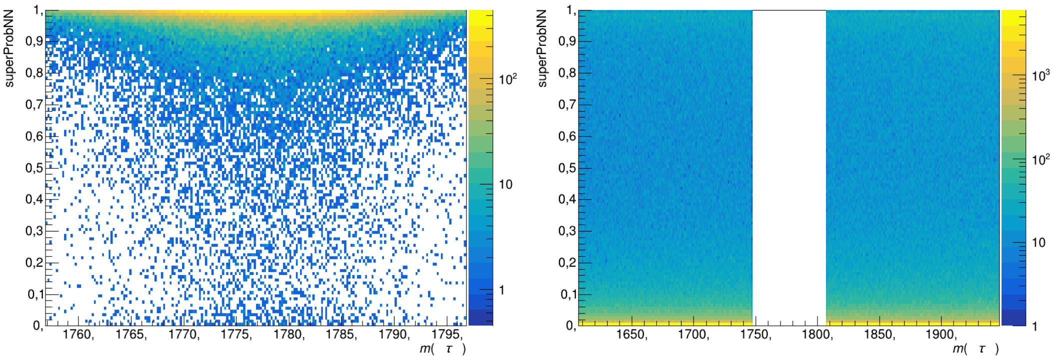


FIGURE 7.13: Correlation between the output of the SuperPNN classifier and the invariant mass distribution for $\tau^+ \to \mu^+\mu^-\mu^+$ simulated events (left) and for the outer sidebands in data (right). Plots are shown for the 2018 samples.

7.3 Signal and background separation in the 2μ sample

A strategy similar to the one used in the 3μ sample is followed in the 2μ sample to discriminate between signal and background events. A multivariate classifier based on kinematic and topological information is trained against the combinatorial background, while a second one, based on PID information, is trained to reduce the background originating from misidentified tracks. As mentioned in Section 3.4, the chosen algorithm is, for both classifiers, a *Gradient Boosted Decision Tree* (BDTG) available in the TMVA toolkit [62].

The BDTG classifier based on kinematic and topological information is refereed to as BDTGKT, while BDTGPID is the name assigned to the classifier based on PID information. The features and performances of both classifiers are discussed in the following.

7.3.1 BDTGKT classifier

The BDTG classifier trained to suppress the combinatorial background takes as input features the variables listed in Table 7.7. Regarding the variables of the daughter particles, a distinction between the tracks identified as muons and the one not satisfying the `isMuon` condition is always made. As already mentioned, in the 2μ

TABLE 7.7: Input features to the BDTGKT model. Note that when the variable is associated to the daughter particles, a distinction between tracks satisfying and not satisfying the `isMuon` condition is always made.

Variable	**Definition**
tau_IPCHI2_OWNPV	Difference in the vertex-χ^2 of the PV reconstructed with and without the particle, for τ and daughter particles
tau_PT	τ tranverse momentum
tau_TAU	τ decay time
tau_ENDVERTEX_CHI2	χ^2 fit of the decay vertex
LONGMAX1(2,3)	tau_TRKISOBDTLONG computed for the pair of muons `isMuon`
VELOMAX1(2,3)	tau_TRKISOBDTVELO computed for the pair of muons with `isMuon`
mu_PT	Transverse momentum of each daughter particle
mu_TRACK_CHI2NDOF	χ^2 of the track fit divided by the degrees of freedom for the track with `!isMuon`
mu_THETA	θ angle of the track with `!isMuon`
mu_PHI	Absolute value of the ϕ angle of each daughter particle

sample only two of three LONG and VELO isolation variables can be exploited in the training of the BDTG model. The average between these two is computed and used as input to the classifier. Simulated events are used as a proxy for the signal, while events from the inner sidebands are used as a proxy for the background. The distributions of the input features are shown for the two samples in Figure 7.14 for 2018; same distributions for 2016 and 2017 can be found in Appendix B. Simulated events are weighted with w_{fraction}, w_{fracCorr}, $w_{\text{Data/MC}}$ and w_{PID}. The latter includes both the correction of the ProbNNmu variables for the tracks with `isMuon` and of the RichDLLmmu variable for the track with `!isMuon` (see Section 6.3).

Variables related to the geometry of the decay, especially the ϕ and θ angle of the track not detected within the muon detector acceptance, turn out to be very powerful in discriminating between signal and background events. The two separate contributions visible in the MC distribution of the θ angle of the track with `!isMuon`, find their explanation in the low momentum of the muon track. As introduced in Section 5.2, muons with `!isMuon` are either produced at low p_T, and therefore end up in the beam pipe, or at higher p_T, and therefore deflected by the magnet outside the muon detector acceptance.

To make use of the whole MC sample for the training and testing of the classifier, and to avoid biased results, the *k-folding* is applied. The signal and background samples are both split into 5 *folds* with same statistics and the BDTG model is iteratively trained on 4 sub-sets and tested on the remaining one to evaluate its performance. Each of the 5 *folds* leads to an independent BDTG model.

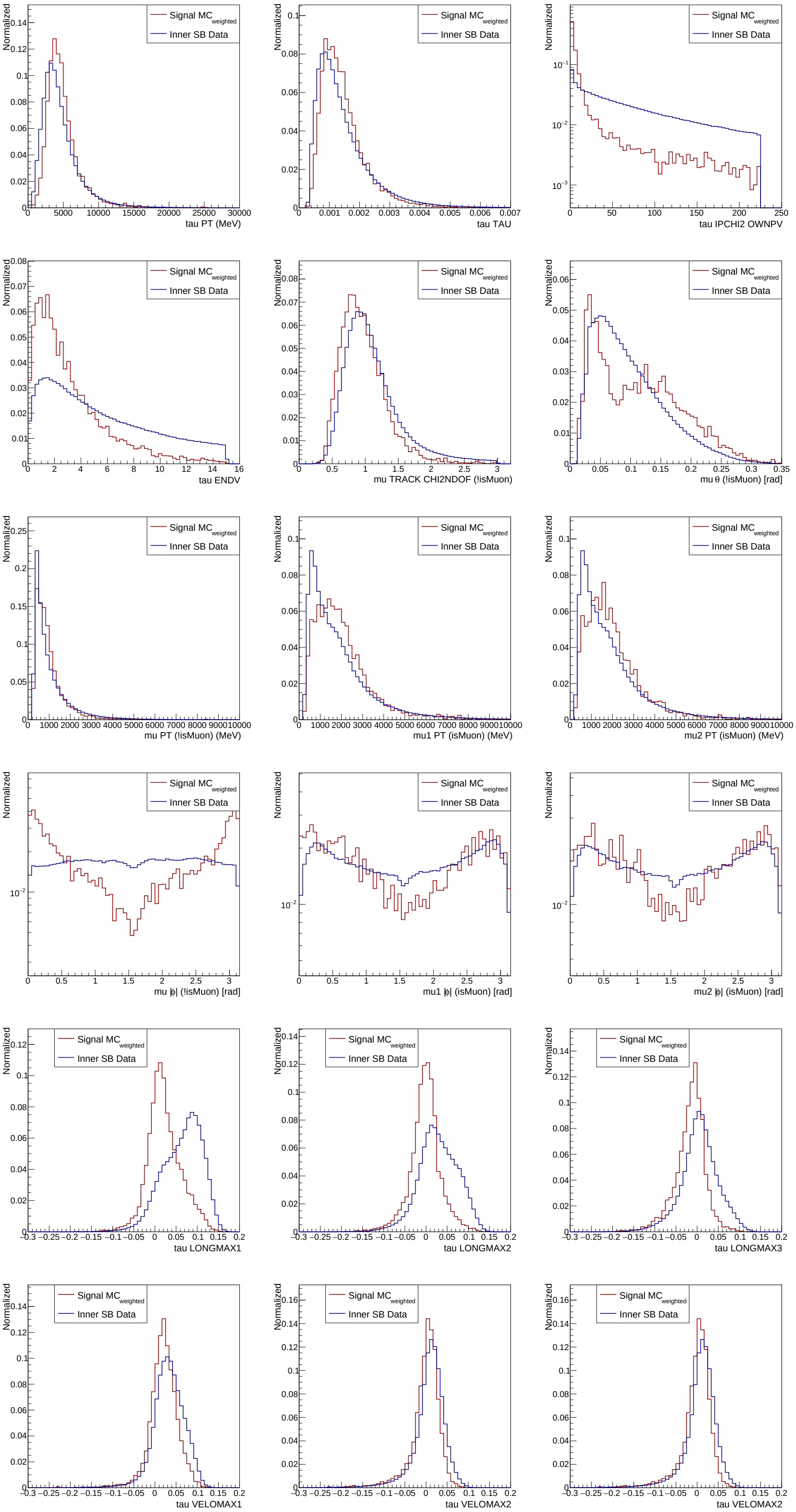


FIGURE 7.14: Distribution of the input features of the BDTGKT classifier for 2018 $\tau^+ \to \mu^+\mu^-\mu^+$ simulated events and inner sidebands data. The histograms are normalised to unit area.

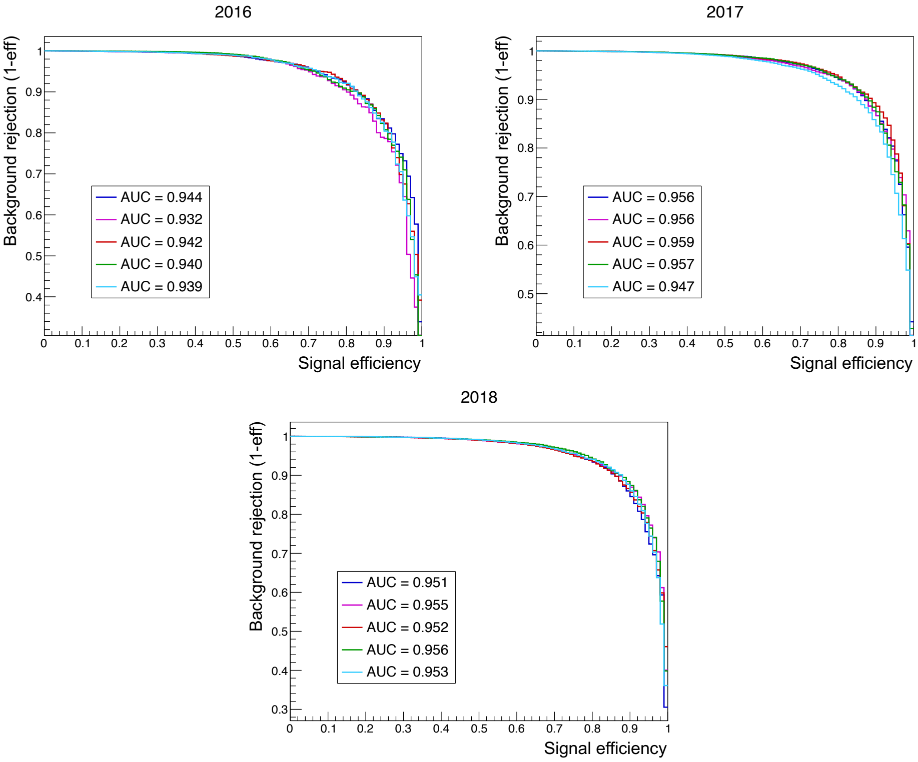


FIGURE 7.15: ROC curves of the five BDTG models trained for the three data-taking years. The corresponding AUCs are also displayed.

The search for the optimal setting of the BDTG model is performed by varying the following hyperparameters:

- *Ntrees*: number of trees in the forest;
- *MaxDepth*: maximum depth of the decision tree allowed;
- *MinNodeSize*: minimum percentage of training events required in a leaf node;
- *nCuts*: number of grid points in the variable range used in finding the optimal cut in the node splitting.

The configuration providing the highest AUC (area under the *receiver operating characteristic*, or ROC, curve) is selected as the optimal set of hyperparameters. The performances of the five BDTG models, expressed in terms of the AUC, are shown in Figure 7.15 together with the corresponding ROC curves for the three data-taking years. The BDTG models trained for 2016 show worse performances than the ones trained on the 2017 and 2018 data samples. This is most likely due to the lower statistics available in the 2016 MC sample, consequence of the different, and less efficient, trigger selection (see Section 5.3). Once trained and tested, the *folded* models are applied to the MC sample, making sure that each model is applied to the sub-set used to test its performance. Whenever the classifier is applied to events used neither for the training or for testing the classifier, one model is randomly extracted from the five trained models and applied to the single event. This is the case for events in the data outer sidebands and in the $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ data and simulation samples.

Calibration of the BDTGKT output

In order to check for residual discrepancies between data and simulation in the output of the BDTGKT classifier, the *folded* models are applied to the reference channel data and simulated events. Figure 7.16 shows how the features used as input to the BDTGKT classifier are distributed in the reference channel for 2018 data and simulation; same distributions for 2016 and 2017 can be found in Appendix B.

The distributions in Figure 7.16 are obtained by requiring the final-state pion to fall outside the muon detector acceptance. This is done in order to select pions with a kinematics similar to the one of muons with `!isMuon`. This requirement reduces the $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ data and MC sample to 5%, though the performance of the BDTGKT classifier are expected to be similar to the one obtained for the signal channel.

With a strategy similar to the one used to correct for the residual data/MC differences observed in XGB and SuperPNN, the corrections to the output of the BDTGKT classifier are extracted from the ratio between the distribution of BDTGKT in the reference channel in data and simulation. The ratio is evaluated in 30 bins whose edges are chosen such that all bins have have roughly the same amount of simulated events. The corresponding distributions are shown in Figure 7.17 for the three years, together with a zoom in the region at high BDTGKT values.

The BDTGKT output distributions for the *sWeighted* data and for the MC sample, before and after applying the BDTGKT corrections, are reported in Figure 7.18 for the three years.

The corrections for data/MC differences in the BDTGKT output are then applied to the signal MC sample. A comparison between the BDTGKT output in the signal channel before and after applying the corrections to simulated events is shown in Figure 7.19 for the three years, together with the BDTGKT output distributions obtained for the data outer sidebands.

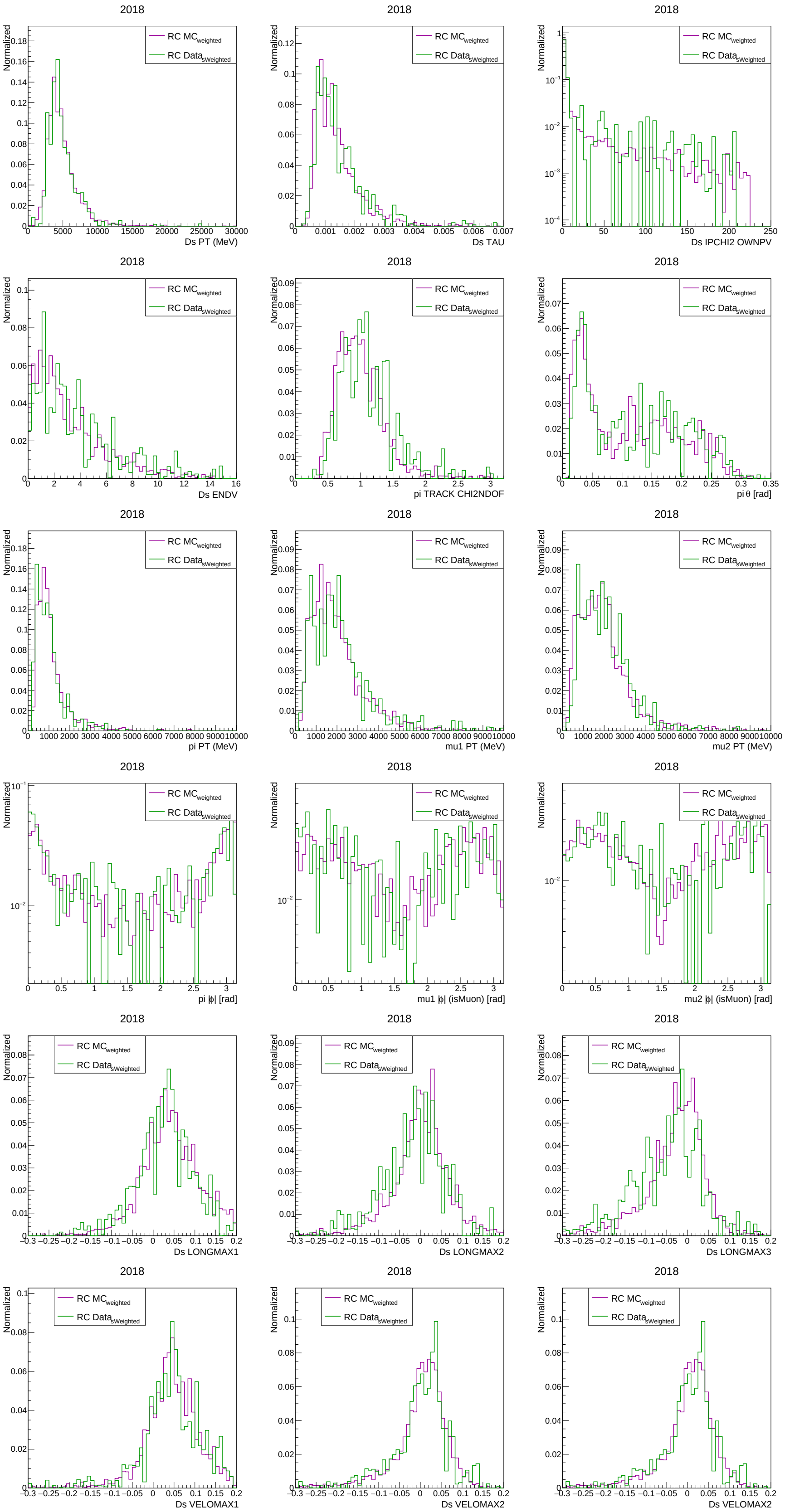


FIGURE 7.16: Distributions of the input features of the BDTGKT classifier for 2018 $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ simulated events and *sWeighted* data. The histograms are normalised to unit area.

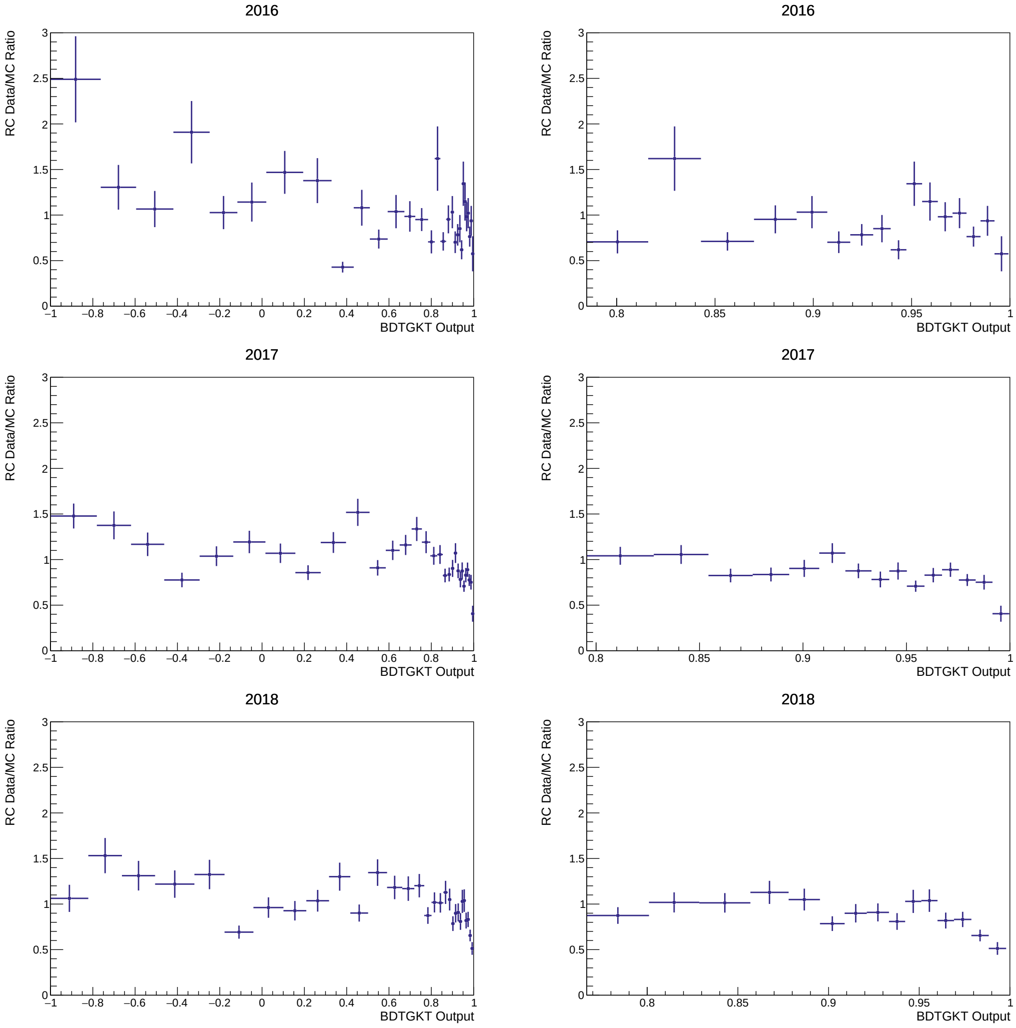


FIGURE 7.17: On the left, bin-by-bin ratio between the data and MC distributions of the classifier output used to reweight the signal and reference channel simulated events. On the right, zoom in of the region at high values of the BDTGKT output. The bins edges are computed to have uniform content in the simulated sample.

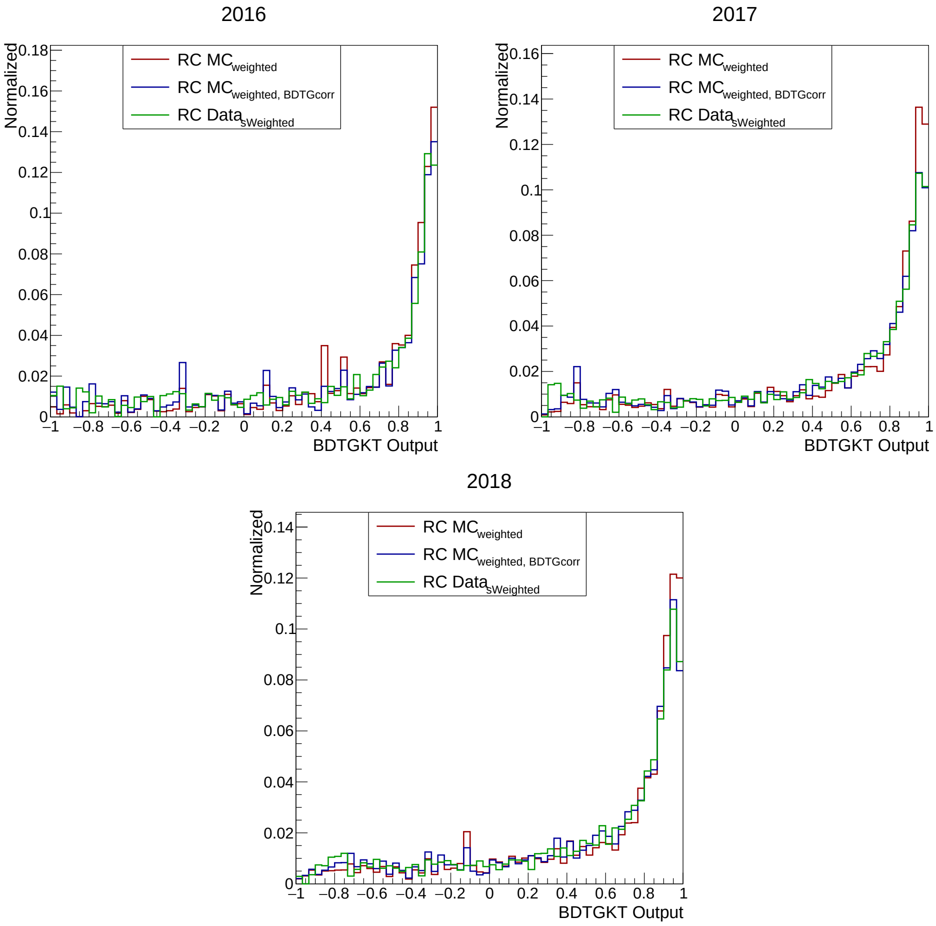


FIGURE 7.18: BDTGKT output distribution for reference channel (RC) *sWeighted* data and simulated events, before and after applying the BDTGKT corrections for the three data-taking years.

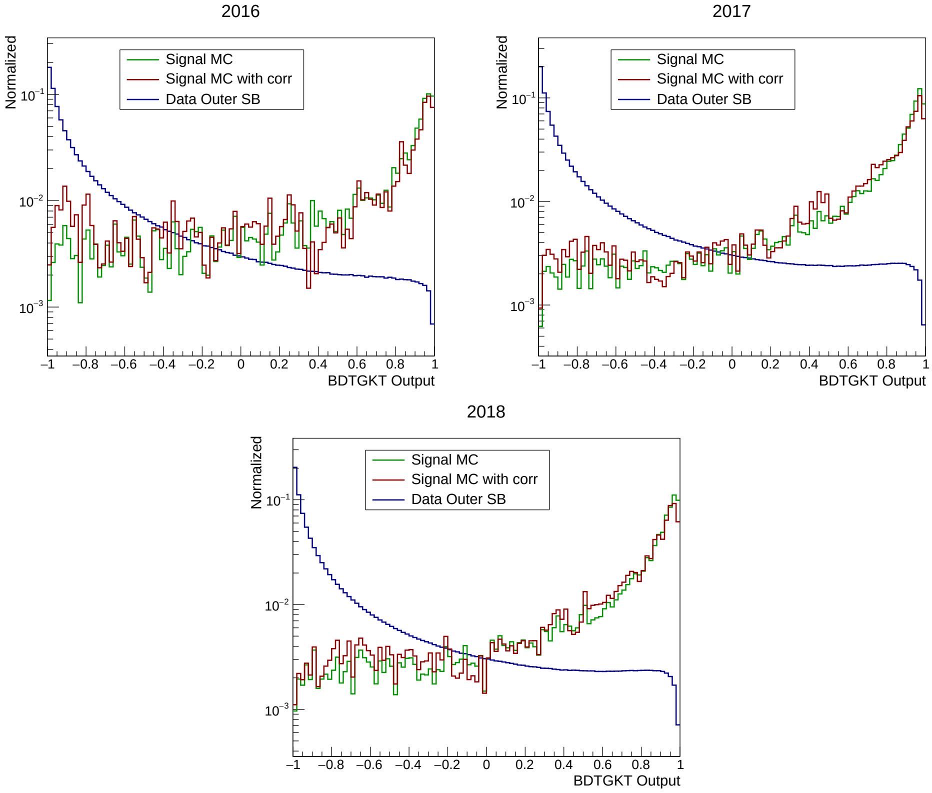


FIGURE 7.19: BDTGKT output distribution for the outer sidebands in data and for the signal channel simulated events, before and after applying the BDTGKT corrections for the three data-taking years.

7.3.2 BDTGPID classifier

The features used as input to the classifier trained against mis-ID tracks are listed in Table 7.8. Since no information from the muon detector is available for the muon not satisfying the `isMuon` condition, for this track the ProbNNmu information is replaced with the RichDLLmu variable. A distinction between tracks satisfying and not satisfying the `isMuon` condition is also made for the ProbNNk and ProbNNpi variables.

The training of the BDTG model is performed using simulated events, weighted with w_{fraction}, w_{fracCorr}, $w_{\text{Data/MC}}$ and w_{PID}, as a proxy for the signal and data events from the middle sidebands as a proxy for the background. A comparison between the signal and background distributions of the input features is shown in Figure 7.20 for 2018; same distributions for 2016 and 2017 can be found in Appendix B. All distributions are displayed in log-scale to better appreciate their discrimination power. The ProbNNk and ProbNNpi distributions obtained for the muon with `!isMuon` differ from the ones obtained for the muons with `isMuon`. Especially the ProbNNpi distribution peaking at large values suggests that, due to its peculiar kinematics, the track that is not detected by the muon detector it is more likely identified as a pion than as a muon.

Since the simulation sample used for the training and testing of the classifier coincides with the sample on which the model is applied, the usage of *k-folding* is required. The optimisation of the models is again performed on the four hyperparameters *Ntrees*, *MaxDepth*, *MinNodeSize* and *nCuts*, and the configuration maximising the AUC is chosen as the optimal one. Each *fold* leads to an independent BDTG model with its own optimal set of hyperparameters and performance. The ROC curves obtained for the five models are shown in Figure 7.21 for the three data-taking years; the corresponding AUCs are also provided.

The BDTG models trained on 2016 data turn out to be less capable in discriminating between signal and background events than the models trained on 2017 and 2018 data, which instead show similar performance. This is again a consequence of the different HLT2 trigger selection applied in 2016. As discussed in Section 5.3, the `Hlt2DiMuonDetached_TOS` trigger line that is required to be fired in the 2μ sample, is not available for 2016 and it is therefore replaced with the trigger line `Hlt2TopoMuMu3Body_TOS`. The `Hlt2TopoMuMu3Body_TOS` trigger line is, however, less

TABLE 7.8: Input features to the BDTGPID model. In case of ProbNNk and ProbNNpi a distinction between tracks satisfying and not satisfying the `isMuon` condition is always made.

Variable	Definition
ProbNNmu	Probability of the track with `isMuon` to be identified as a muon
RichDLLmu	Logarithm of the ratio between the likelihood for the muon and the likelihood of the pion hypothesis for the track with `!isMuon`
ProbNNk	Probability of the three tracks to be identified as a kaon
ProbNNpi	Probability of the three tracks to be identified as a pion

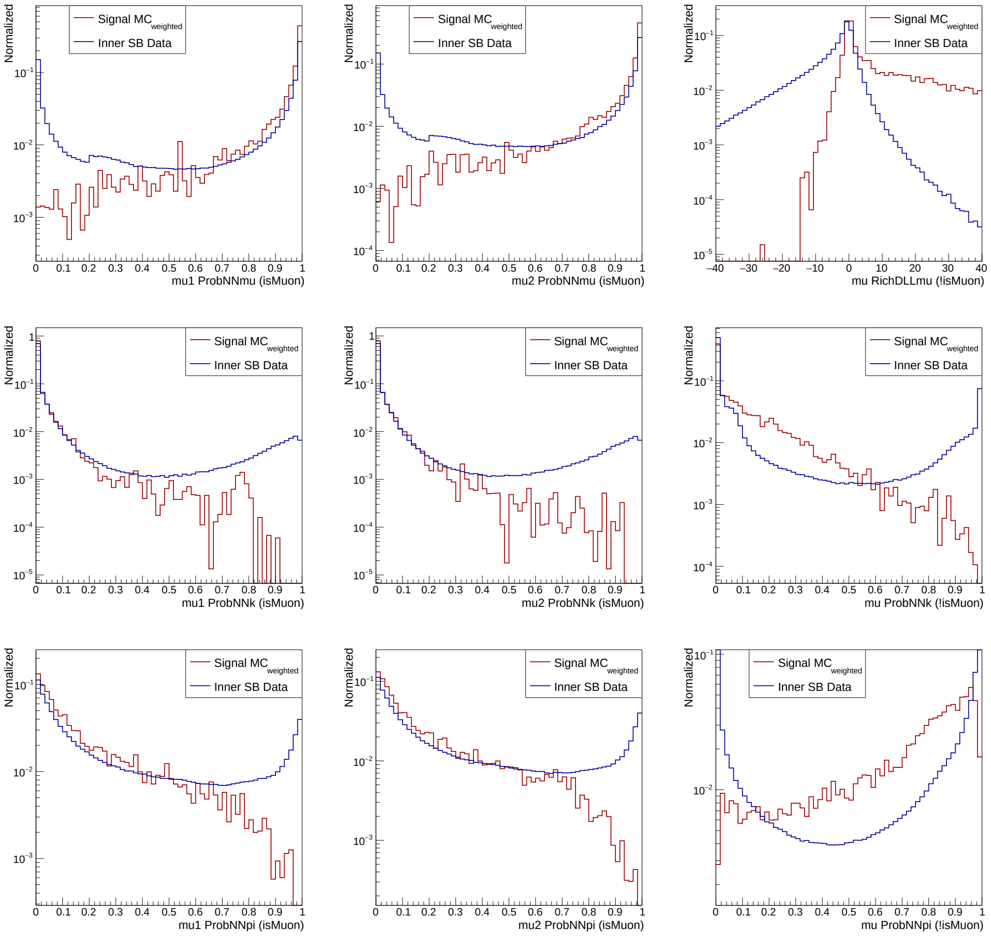


FIGURE 7.20: Distribution of the input features of the BDTGPID classifier for 2018 $\tau^+ \to \mu^+\mu^-\mu^+$ simulated events and inner sidebands data. The histograms are normalised to unit area.

efficient than the line `Hlt2DiMuonDetached_TOS` and based on a cut on ProbNNmu. In particular, it requires the ProbNNmu of the tracks in the final state to be greater than 0.2, affecting the ProbNNk and ProbNNpi variables as well. Since no ProbNNmu is available for the track with `!isMuon`, its PID information is not affected by the different trigger selection.

The ProbNNX distributions obtained for 2016 are shown in Figure 7.22 for one of two muons satisfying the `isMuon` condition. By comparing these plots with the ones shown in Figure 7.20, it can be seen how the cut at ProbNNmu>0.2 also affects the ProbNNk and ProbNNpi making their data distributions similar to the ones obtained for simulated events.

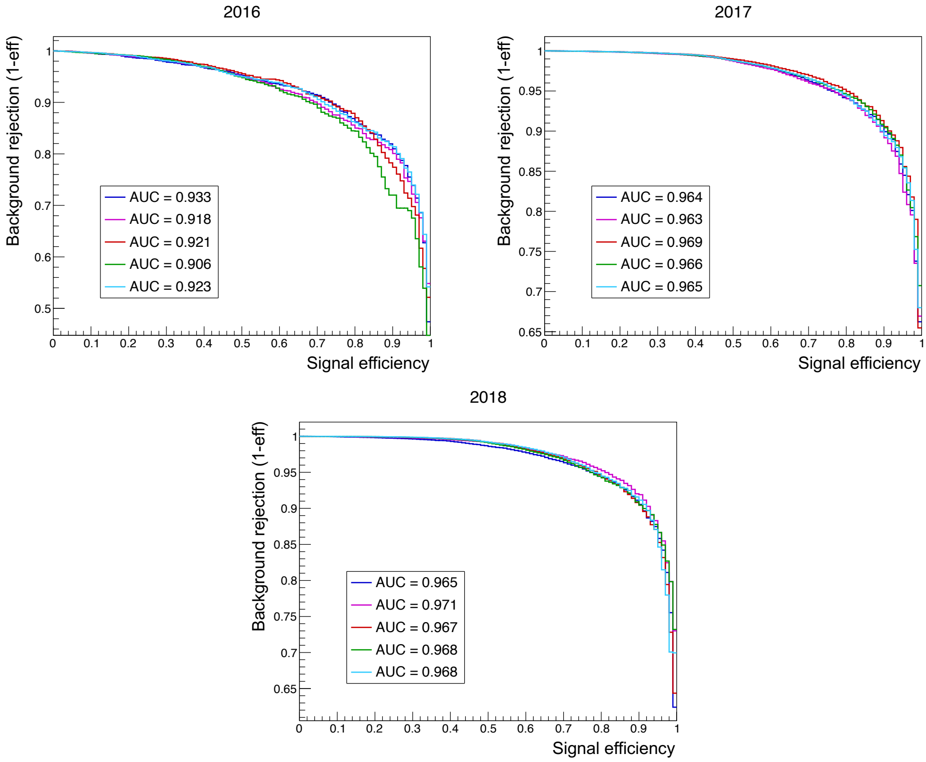


FIGURE 7.21: ROC curves of the 5 BDTG models trained for the three data-taking years. The corresponding AUCs are also displayed.

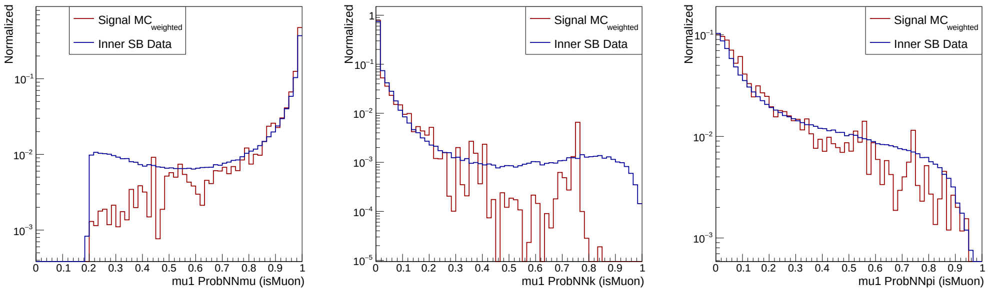


FIGURE 7.22: ProbNNX distributions for one of the two final-state muons satisfying the `isMuon` condition. Distributions are obtained for 2016 data.

Calibration of the BDTGPID output

Despite the PID corrections, some discrepancies between data and simulation could still arise in the output of the BDTGPID classifier. The reference channel is again exploited to check for this disagreement and extract the necessary corrections, with some constraints. In Figure 7.23 the distributions of the input features of the BDTG-PID classifier are shown for the reference channel in 2018 simulation and data; same distributions for 2016 and 2017 can be found in Appendix B.

The information of the track not satisfying the `isMuon` condition is here replaced with the PID features of the track identified as a pion. However, the distributions of the PID variables of this track only resemble the ones obtained for the tracks with `!isMuon` if the pion is required to fall outside the muon detector acceptance. As

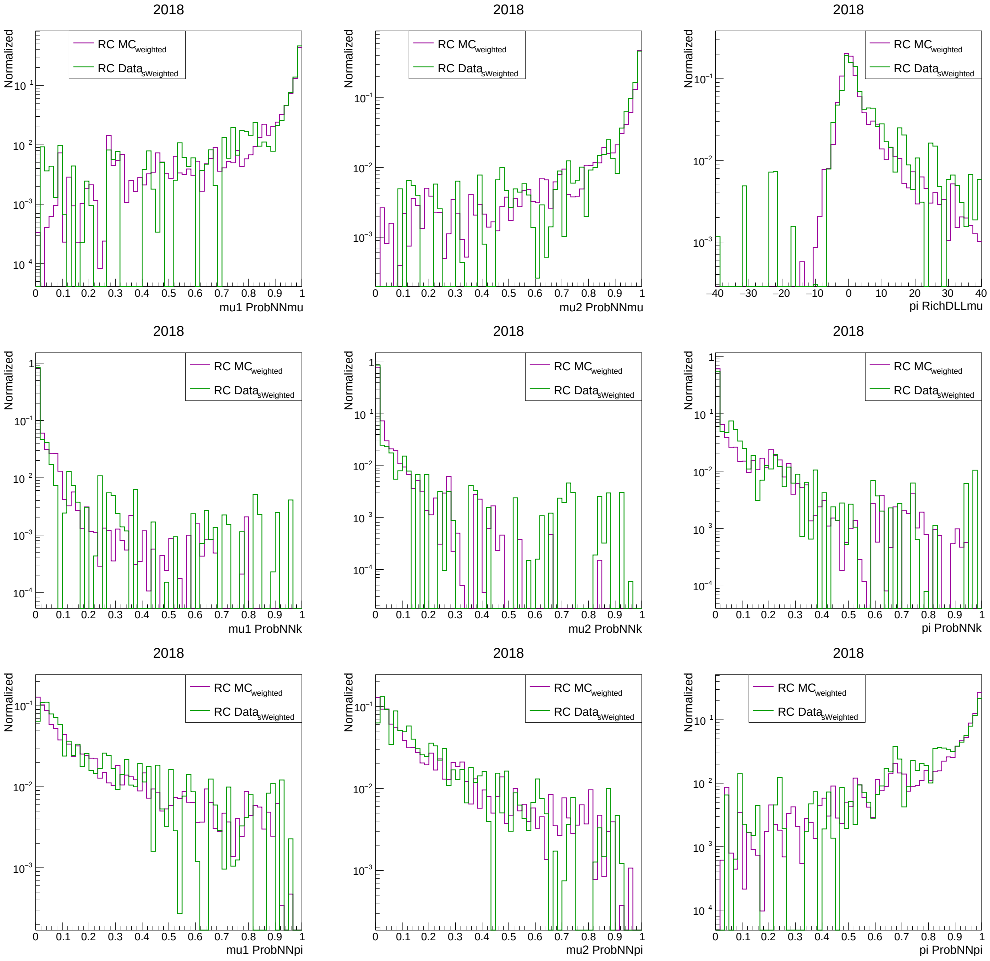


FIGURE 7.23: Distributions of the input features of the BDTGPID classifier for 2018 $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ simulated events and *sWeighted* data. The histograms are normalised to unit area.

pointed out in Section 7.3.1, the statistics available for the reference channel is considerably reduced, with the advantage for the reference channel BDTGPID output to be more similar to the one obtained for the signal channel. The BDTGPID classifier is applied to the reference channel data and simulated events by randomly selecting, event-by-event, one of the five trained models.

The calibration of the BDTPID classifier is performed by filling a histogram with the ratio between the BDTGPID output obtained for data and the one obtained for simulated events. The histogram is made up of 30 bins whose edges are such to have same amount of simulated events in each bin. The data/MC ratios are shown in Figure 7.24 for the three years, while Figure 7.25 shows the BDTGPID output distributions for the reference channel data and simulation, before and after applying the BDTGPID corrections.

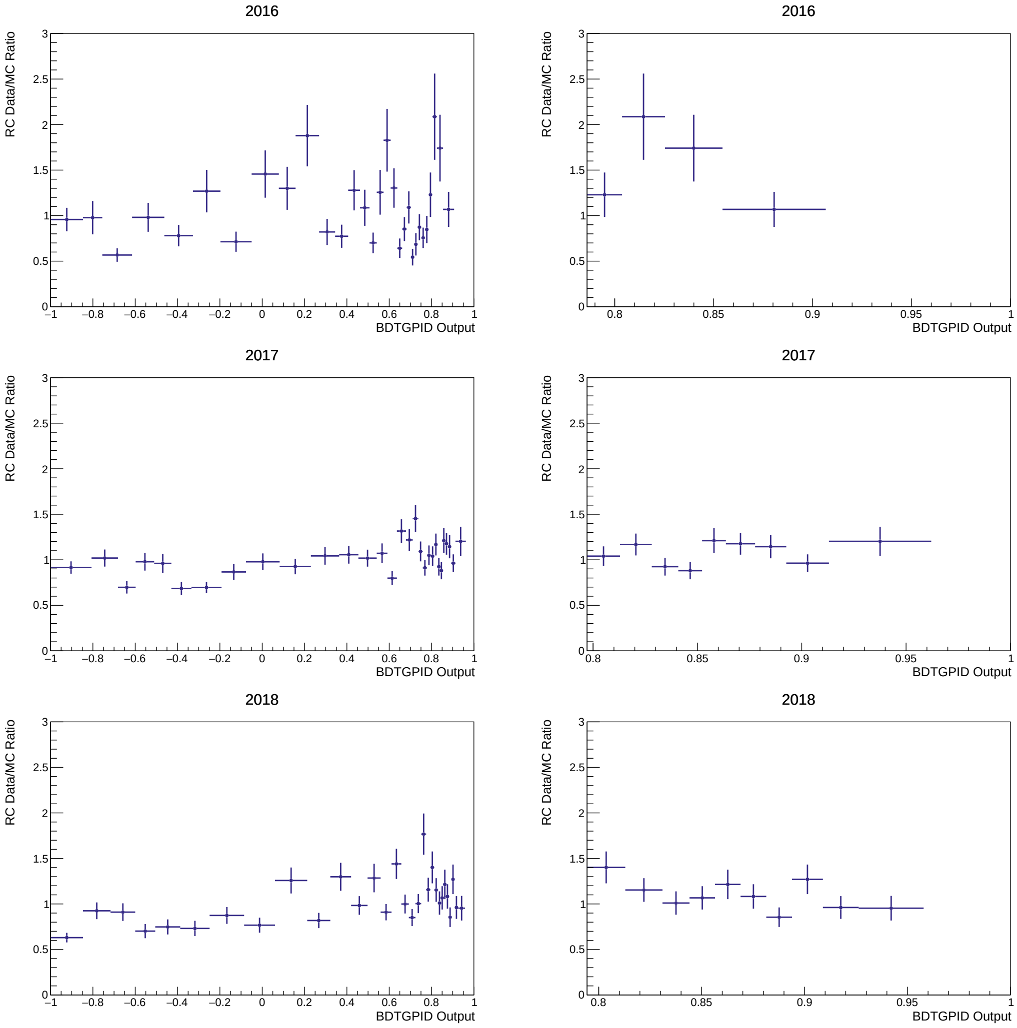


FIGURE 7.24: On the left, bin-by-bin ratio between the data and MC distributions of the classifier output in the reference channel used to reweight the signal simulated events. On the right, zoom in of the region at high values of the BDTGPID output. The bins edges are computed to have uniform content in the simulated sample.

Despite the requirement on the pion, the performance of the BDTGPID model on the reference channel are such that the classifier output does not cover the whole range $[-1, 1]$. In other words, the BDTGPID model trained on the signal channel never assigns to $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ events BDTGPID greater than a certain *threshold* of roughly 0.9. This is particularly visible from the zoom in shown on the right-hand side of Figure 7.24, where the contribution at high values of BDTGPID is missing.

The chosen strategy is to assign unit weight to signal simulated events whose BDTGPID output exceeds the given *threshold*, being the latter 0.907 for 2016, 0.962 for 2017 and 0.958 for 2018. The distributions of the BDTGPID output for the signal channel simulation are shown in Figure 7.26 for the three years before and after applying the BDTGPID corrections; the BDTGPID output distributions for the data outer sidebands are also shown.

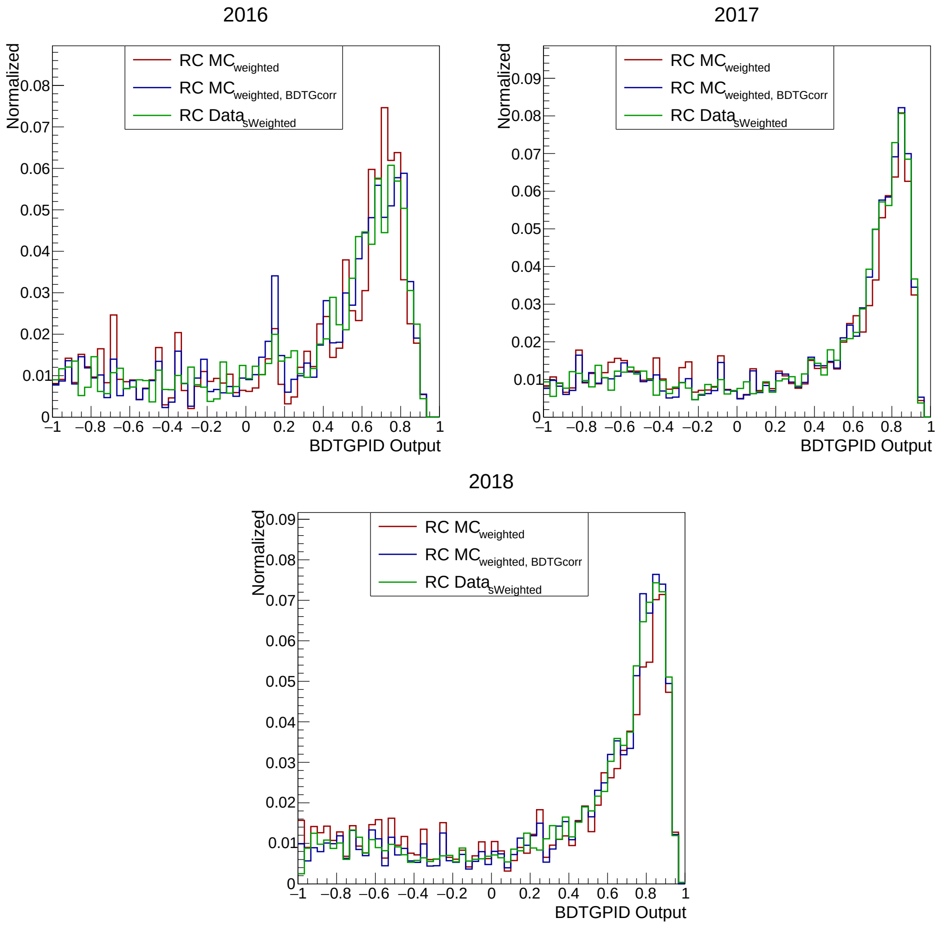


FIGURE 7.25: BDTGPID output distribution for reference channel (RC) *sWeighted* data and simulated events, before and after applying the BDTGPID corrections.

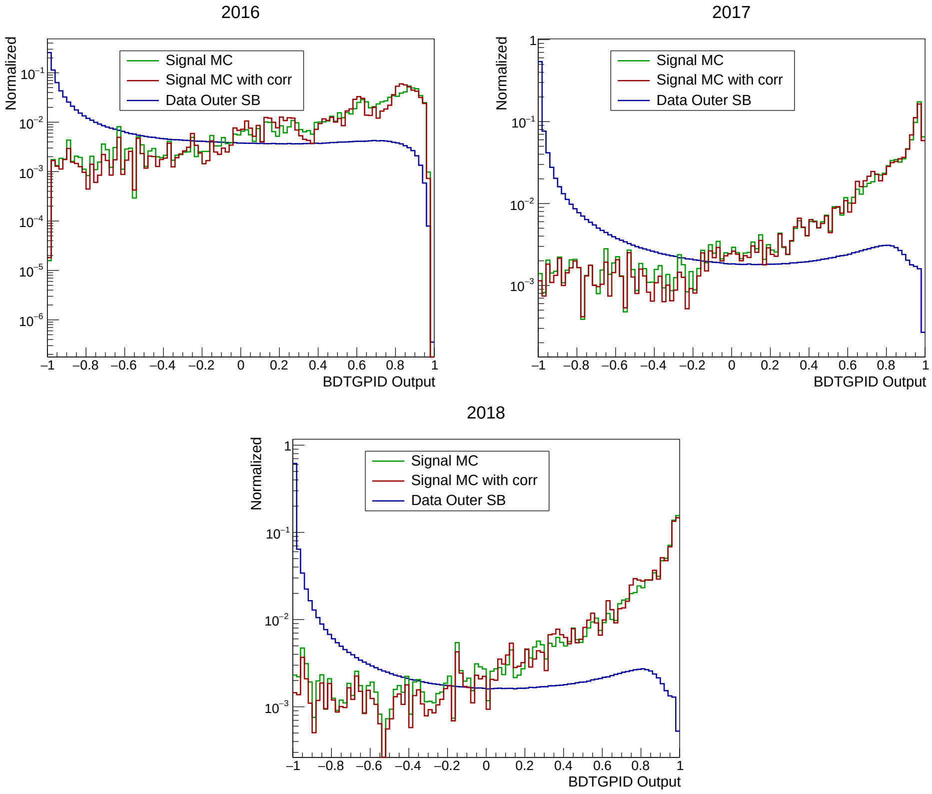


FIGURE 7.26: BDTGPID output distribution for the outer sidebands in data and for the signal channel simulated events, before and after applying the BDTGPID corrections for the three data taking years.

7.3.3 Selection in BDTGKT and BDTGPID

A first significant reduction of the background contamination in the 2μ sample is achieved by cutting on the output of the BDTGKT and BDTGPID classifiers. In Figure 7.27, the output of the BDTGPID classifier is shown as a function of the BDTGKT classifier for simulated events on the left-hand side, and for the data outer sidebands on the right-hand side. The distributions are shown in the range $[0, 1]$ as the search for the optimal cut is anyway performed at large BDTGPID and BDTGKT values.

As can be seen, no correlation is visible between the two classifiers, meaning that the selection applied on BDTGPID is independent of the selection applied on BDTGKT and vice versa. The threshold of the cut applied in both classifiers is chosen

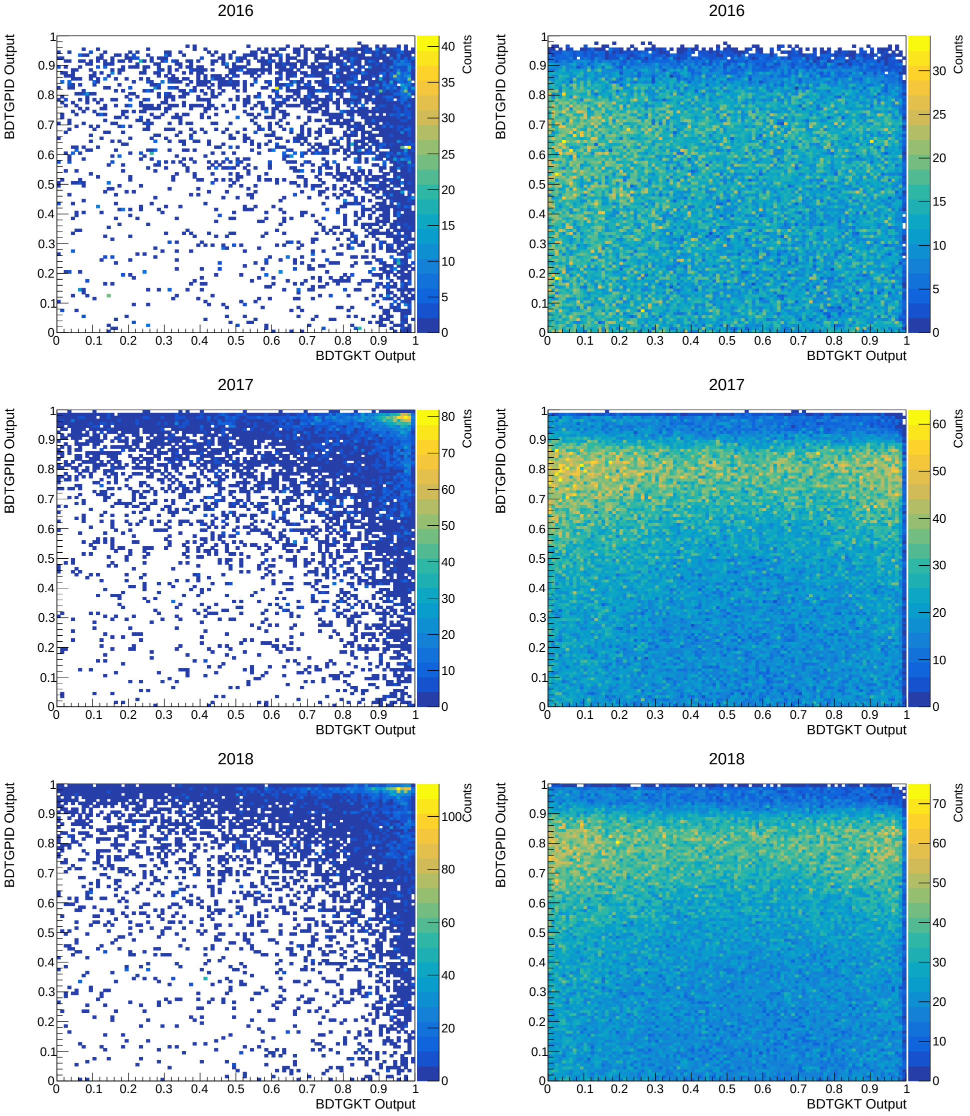


FIGURE 7.27: Correlation between the output of the BDTGKT and BDTGPID classifiers for signal simulated events (left) and for the data outer sidebands (right). The distributions are shown for the three data-taking years.

as the one that maximises the Punzi Figure of Merit (FoM) [91] defined as

$$\text{Punzi FoM} = \frac{\epsilon_S(t)}{3/2 + \sqrt{B(t)}}, \tag{7.5}$$

where $\epsilon_S(t)$ is the efficiency of the cut applied at the threshold t evaluated on simulated events and $B(t)$ the corresponding number of background events expected in the signal region. The latter is determined by fitting the data outer sidebands with an exponential distribution and extrapolating the number of events in the signal region.

For 2017 and 2018 the search for the optimal cut is performed with a step of 0.01 in the range $[0.85, 0.99]$ for the BDTGPID and with the same step in the range $[0.75, 0.95]$ for the BDTGKT. For 2016 the scanned ranges are $[0.7, 0.9]$ for BDTGPID and $[0.75, 0.95]$ for BDTGKT, both with a step of 0.01. In Figure 7.28, the FoM distributions obtained for BDTGPID (left) and BDTGKT (right) are shown for the three data-taking years. The optimal BDTGPID and BDTGKT cuts are listed in Table 7.9.

TABLE 7.9: Optimal set of cuts on the output of the BDTGPID and BDTGKT classifiers for the three data-taking years.

Classifier output	**2016**	**2017**	**2018**
BDTGPID	> 0.79	> 0.93	> 0.96
BDTGKT	> 0.93	> 0.90	> 0.81

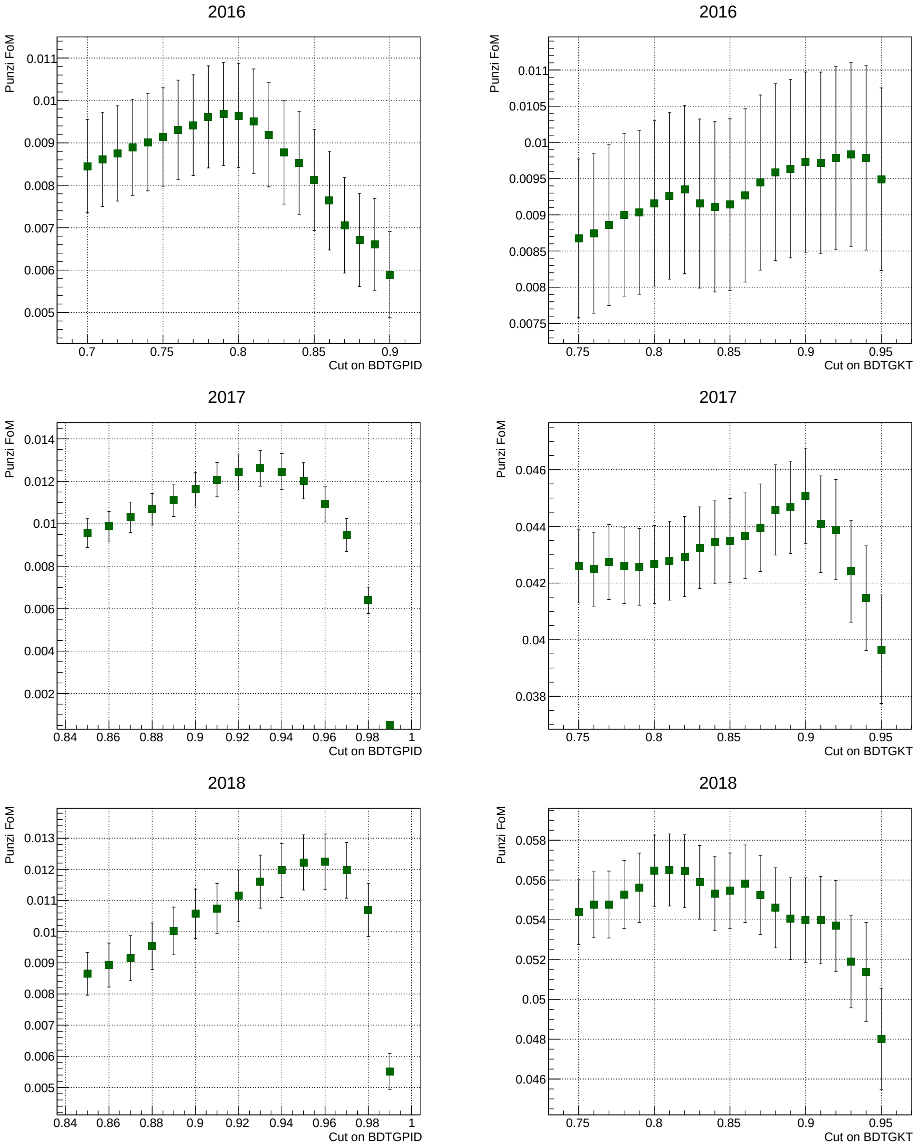


FIGURE 7.28: Distribution of the Punzi FoM evaluated at different thresholds of BDTGPID (left) and BDTGKT (right) for the three data-taking years.

Chapter 8

Background estimate

In this Chapter, the various sources of physical background are discussed, namely the ones originating from physical processes wrongly selected as $\tau^+ \to \mu^+\mu^-\mu^+$ decays. Among them, a distinction is made between decays with three muons in the final state and decays where one or more tracks are misidentified as muons. To the former category belongs the $D_s^+ \to \eta(\mu^+\mu^-\gamma)\mu^+\nu_\mu$ decay, discussed in detail in Section 8.1.1. Among the background sources originating from misidentified tracks, the $D^+ \to K^-\pi^+\pi^+$, $D^+ \to \pi^-\pi^+\pi^+$ and $D_s^+ \to \pi^-\pi^+\pi^+$ decays represent the dominant contributions. The final estimate of the contribution expected in the signal region from the various background sources is presented in the following.

The study of the background contamination from decays with three muons in the final state, especially the $D_s^+ \to \eta(\mu^+\mu^-\gamma)\mu^+\nu_\mu$ decay, has been carried out by the author. The study of background originating from misidentified tracks and the final fit to the data sidebands (see Section 8.2), is the result of the collaboration of the author with a postdoctoral researcher of the analysis group.

8.1 Background processes with three muons in the final state

The processes with three real muons of total charge ± 1 in the final state include all the semileptonic decays of D_s, D^+ and B mesons into a muon and a muon-neutrino, in combination with a light unflavoured meson decaying into $\mu^+\mu^-$ or $\mu^+\mu^-\gamma$. Because of the smaller production rate of B mesons compared to D_s and D^+, and because of the three order or magnitude smaller decay rate of B mesons than D_s and D^+ into light unflavoured mesons (e.g. $\mathcal{B}(B^+ \to \eta l^+\nu_l) = (3.9 \pm 0.5) \times 10^{-5}$ and $\mathcal{B}(D_s^+ \to \eta\mu^+\nu_\mu) = (2.4 \pm 0.5)\%$ [8]), only the D_s and D^+ decays are considered in the background evaluation.

A list of the semileptonic decays of D_s and D^+ into $3\mu + X$ was presented in Ref. [3] and it is here shown in Table 8.1. The given values for the branching fractions are taken from the PDG [8]. The expected production cross section $\sigma(3\mu + X)$ is computed by multiplying the total branching fraction $\mathcal{B}_1 \times \mathcal{B}_2$ by the prompt D_s and D^+ production cross section in 4π (see Table 4.8) and by the corresponding detector acceptance efficiency evaluated on simulated events. The values for the detector acceptance efficiency are taken from Ref. [3] as the aim of this study is mainly to identify which of the D_s or D^+ decays into three real muons could represent a background source for the signal mode. The channels with the higher expected production cross section are the $D_s^+ \to \phi(\mu^+\mu^-)\mu^+\nu_\mu$ and $D_s^+ \to \eta(\mu^+\mu^-\gamma)\mu^+\nu_\mu$; while the former decay is easily rejected by the veto applied to the ϕ mass in the offline selection (see Section 5.1), for the latter, a dedicated study is presented in the following.

TABLE 8.1: List of the branching fractions of the semileptonic decays of D^+ and D_s into $3\mu + X$ as in Ref. [3]. All branching fractions have been updated to their latest measurement according to the PDG [8]. The expected production cross section is computed by multiplying the total branching fraction by the prompt 4π production cross section of D_s (1732 μb) or D^+ (4054 μb) and by the corresponding detector acceptance efficiency taken from Ref. [3] ($\epsilon_\eta = 0.183$ for decays into η or η', $\epsilon_\phi = 0.19$ for decays with ϕ as intermediate state, $\epsilon_\omega = 0.20$ for decays into ω, $\epsilon_{\rho^0} = 0.19$ for decays into ρ^0).

$D^+_{(s)}$ **decay**	$\mathcal{B}_1$	**Secondary decay**	$\mathcal{B}_2$	$\mathcal{B}_1 \times \mathcal{B}_2$	$\sigma(3\mu + X)$ **[nb]**
$D_s^+ \to \eta\mu^+\nu_\mu$	2.4×10^{-2}	$\eta \to \mu^+\mu^-$	5.8×10^{-6}	1.4×10^{-7}	0.04
		$\eta \to \mu^+\mu^-\gamma$	3.1×10^{-4}	7.5×10^{-6}	2.36
		$\eta \to \pi^0\mu^+\mu^-\gamma$	$< 3 \times 10^{-6}$	$< 7.2 \times 10^{-8}$	0.02
$D_s^+ \to \eta'\mu^+\nu_\mu$	1.1×10^{-2}	$\eta' \to \mu^+\mu^-\gamma$	1.13×10^{-4}	1.2×10^{-6}	0.39
$D_s^+ \to \phi\mu^+\nu_\mu$	1.9×10^{-2}	$\phi \to \mu^+\mu^-$	2.85×10^{-4}	5.4×10^{-6}	1.78
		$\phi \to \mu^+\mu^-\gamma$	1.4×10^{-5}	2.7×10^{-7}	0.09
		$\phi \to \mu^+\mu^-\pi^0$	1.3×10^{-5} $(*)$	2.5×10^{-7}	0.08
$D^+ \to \eta\mu^+\nu_\mu$	1.04×10^{-3}	$\eta \to \mu^+\mu^-$	5.8×10^{-6}	6.0×10^{-9}	0.004
		$\eta \to \mu^+\mu^-\gamma$	3.1×10^{-4}	3.2×10^{-7}	0.24
		$\eta \to \pi^0\mu^+\mu^-\gamma$	$< 3 \times 10^{-6}$	$< 3.1 \times 10^{-9}$	0.002
$D^+ \to \eta'\mu^+\nu_\mu$	2.0×10^{-4} $(*)$	$\eta' \to \mu^+\mu^-\gamma$	1.13×10^{-4}	2.3×10^{-8}	0.02
$D^+ \to \omega\mu^+\nu_\mu$	1.77×10^{-3}	$\omega \to \mu^+\mu^-$	7.4×10^{-5}	1.3×10^{-7}	0.11
		$\omega \to \mu^+\mu^-\pi^0$	1.3×10^{-4}	2.4×10^{-7}	0.10
$D^+ \to \rho^0\mu^+\nu_\mu$	2.4×10^{-3}	$\rho \to \mu^+\mu^-$	4.55×10^{-5}	1.1×10^{-7}	0.08
$D^+ \to \phi\mu^+\nu_\mu$	$< 1.3 \times 10^{-5}$ $(*)$	$\phi \to \mu^+\mu^-$	2.85×10^{-4}	$< 3.7 \times 10^{-9}$	0.003

$(*)$ given branching fractions are from the electron mode.

TABLE 8.2: Branching fractions of different channels.

Channel	$\mathcal{B}$
$D_s^+ \to \eta\mu^+\nu_\mu$	$(2.4 \pm 0.5) \times 10^{-2}$
$\eta \to \mu^+\mu^-\gamma$	$(3.1 \pm 0.4) \times 10^{-4}$
$D_s^+ \to \eta(\mu^+\mu^-\gamma)\mu^+\nu_\mu$	$(7.5 \pm 1.8) \times 10^{-6}$

8.1.1 The $D_s^+ \to \eta(\mu^+\mu^-\gamma)\mu^+\nu_\mu$ channel

The branching fraction (BF) of the $D_s^+ \to \eta(\mu^+\mu^-\gamma)\mu^+\nu_\mu$ decay is obtained by multiplying the BF of the $D_s^+ \to \eta\mu^+\nu_\mu$ and of the $\eta \to \mu^+\mu^-\gamma$ decay, whose values are taken from the PDG [8] and listed again in Table 8.2 with their uncertainty. The resulting $\mathcal{B}(D_s^+ \to \eta(\mu^+\mu^-\gamma)\mu^+\nu_\mu)$ is also reported.

Because of its large BF - two orders of magnitude greater than the extrapolated limit for the BF of $\tau^+ \to \mu^+\mu^-\mu^+$ (see Section 11.1) - the $D_s^+ \to \eta(\mu^+\mu^-\gamma)\mu^+\nu_\mu$ decay can be considered as the dominant background channel with three muons in the final state.

As already mentioned in Section 5.1, a cut on the invariant mass of the opposite-sign muons is applied to reject events originating from the $D_s^+ \to \eta(\mu^+\mu^-\gamma)\mu^+\nu_\mu$ decay ($m_{\mu^+\mu^-} > 450$ MeV/c^2). Nevertheless, some $D_s^+ \to \eta(\mu^+\mu^-\gamma)\mu^+\nu_\mu$ events can survive this cut and contaminate the signal region. In the following, the strategy used to estimate the remaining number of $D_s^+ \to \eta(\mu^+\mu^-\gamma)\mu^+\nu_\mu$ events in the signal region is presented.

About 3M simulated events are generated for the three data-taking years and 9×10^4 candidates are selected by the stripping line applied to the signal channel $\tau^+ \to \mu^+\mu^-\mu^+$ (`StrippingTau23MuTau23MuLine`). A pure sample of $D_s^+ \to \eta(\mu^+\mu^-\gamma)\mu^+\nu_\mu$ events is obtained requiring the D_s background category to be equal to 40; this allows to select events where the mother particle is not fully reconstructed and whose mass value is more than 100 MeV/c^2 lower than the expected mass value. More details on the background category can be found in Section 4.2.1. In addition, a truth matching based on the TRUE parental ID of the particles in the final state is applied: in particular, when the mother ID of the opposite-sign muons matches the η ID (±221[1]), their grandmother ID is required to be equal to the D_s ID (±431[1]), while the remaining track in the final state is required to originate from a D_s. For the sake of completeness the truth matching of the D_s itself is required. The number of signal candidates selected by the truth-matching is 11029.

Figure 8.1 shows, on the left-hand side, the invariant mass distribution of one of the two possible opposite-sign muon pairs in the decay. The peak at low mass values corresponds to the η resonance, whose central value is shifted from the nominal one (547.86 ± 0.02 MeV/c^2 [8]) because of the missing γ. The structure at higher mass values is originated from the combination of one of the η's muons with the non resonant one. On the right-hand side of Figure 8.1, the corresponding Dalitz plot is displayed.

The offline, `isMuon` and trigger selections applied to the signal channel and described in Section 5 are applied to this background channel as well. The simulated sample is therefore split in its 3μ component, where all the three charged tracks in the final state are identified as muons, and the 2μ component, where only two of the three charged tracks are identified as muons. The efficiencies of each step of the selection are reported in Table 8.3 for each year.

---

[1] Monte Carlo particle numbering scheme [92].

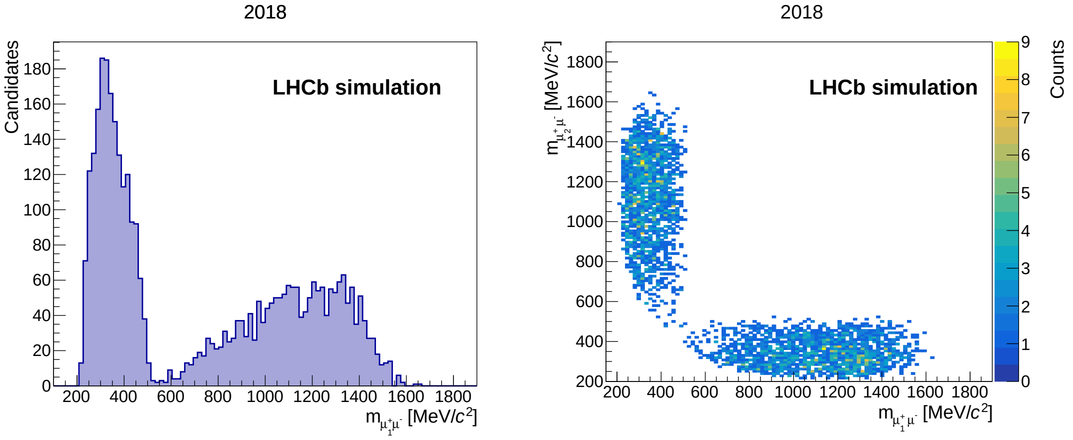


FIGURE 8.1: On the left, invariant mass distribution of one of the two possible opposite-sign muon pairs in the $D_s^+ \to \eta(\mu^+\mu^-\gamma)\mu^+\nu_\mu$ channel after the truth-matching. On the right, Dalitz plot of the two possible combination of opposite-sign muons after the truth-matching. The distributions are obtained for 2018 simulated events.

TABLE 8.3: Efficiency of the different stages of the selection applied to $D_s^+ \to \eta(\mu^+\mu^-\gamma)\mu^+\nu_\mu$ simulated events for the three data-taking years. The uncertainties are purely statistical.

Stage	**Eff. 2016 [%]**		**Eff. 2017 [%]**		**Eff. 2018 [%]**	
	3μ	**2μ**	**3μ**	**2μ**	**3μ**	**2μ**
Acceptance	12.460 ± 0.005		12.64 ± 0.04		12.55 ± 0.08	
Filtering	49.85 ± 0.05		49.86 ± 0.05		49.765 ± 0.005	
Stripping and Reco.	2.72 ± 0.02		2.74 ± 0.02		2.75 ± 0.02	
Offline Selection	6.9 ± 0.4		8.1 ± 0.5		7.6 ± 0.4	
Veto on $m_{\mu^+\mu^-}$	7.3 ± 0.4		8.5 ± 0.5		7.9 ± 0.5	
`isMuon` Selection	79 ± 3	21 ± 3	76 ± 3	24 ± 3	76 ± 3	24 ± 3
Trigger Selection	72 ± 3	9 ± 4	78 ± 3	31 ± 6	72 ± 3	25 ± 5
Classifier Selection	77 ± 3	20 ± 18	76 ± 3	9 ± 6	80 ± 3	35 ± 12

The geometrical acceptance and the efficiency of the filtering applied at the generator level, are averaged over the two polarities. The offline selection efficiency already includes the efficiency of the veto applied to the invariant mass of the two oppositely-charged muons, shown in the row right below. The effect of the veto can be visualised in Figure 8.2 where the invariant mass distribution of the three final-state muons is shown before and after applying the veto.

The $D_s^+ \to \eta(\mu^+\mu^-\gamma)\mu^+\nu_\mu$ mode is not expected to peak in proximity of the signal because of the missing γ and ν_μ, though its right tail can still contaminate the left sideband and the signal region, represented by the red area in Figure 8.2. The two multivariate classifiers trained on the 3μ sample against the combinatorial and mis-ID background are applied to the $D_s^+ \to \eta(\mu^+\mu^-\gamma)\mu^+\nu_\mu$ sub-sample with three charged tracks identified as muons and the efficiency of the minimum requirement on XGB (> 0.8) and SuperPNN (> 0.88) is evaluated separately for the three years.

The same approach is followed for events where only two tracks are identified as muons; the BDTGKT and BDTGPID classifiers trained on the 2μ signal sample are applied to the corresponding $D_s^+ \to \eta(\mu^+\mu^-\gamma)\mu^+\nu_\mu$ sub-sample and the efficiency

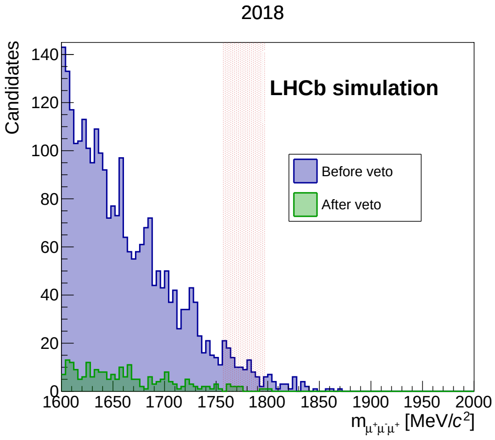


FIGURE 8.2: Invariant mass distribution of $D_s^+ \to \eta(\mu^+\mu^-\gamma)\mu^+\nu_\mu$ before and after applying the veto on the opposite-sign dimuon mass. The red area represents the signal region.

of the selection in the output of the two classifiers is evaluated. The computed efficiencies are reported in the last row of Table 8.3 separately for the three years and for the two sub-samples. The large uncertainties associated to the efficiency evaluated on the 2μ sub-sample are due to the limited statistics. For each year, less than 10 events survive the aforementioned selection in the 2μ sample.

The expected number of $D_s^+ \to \eta(\mu^+\mu^-\gamma)\mu^+\nu_\mu$ in the $\tau^+ \to \mu^+\mu^-\mu^+$ invariant mass spectrum is evaluated using the $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ as a normalisation channel via the formula

$$N_{D_s,Bkg} = N_{D_s,Ref} \times \frac{\mathcal{B}(D_s^+ \to \eta(\mu^+\mu^-\gamma)\mu^+\nu_\mu)}{\mathcal{B}(D_s^+ \to \phi(\mu^+\mu^-)\pi^+)} \times \frac{\epsilon_{D_s,Ref}}{\epsilon_{D_s,Bkg}}, \tag{8.1}$$

where $N_{D_s,Ref}$ represents the number of events observed in the reference channel $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ (see Section 9.3) and $\epsilon_{D_s,Ref}/\epsilon_{D_s,Bkg}$ the ratio between the overall efficiency evaluated in the reference channel and the one evaluated on $D_s^+ \to \eta(\mu^+\mu^-\gamma)\mu^+\nu_\mu$ simulated events. The values used in the evaluation of $N_{D_s,Bkg}$ are listed in Table 8.4.

TABLE 8.4: Evaluation of the number of $D_s^+ \to \eta(\mu^+\mu^-\gamma)\mu^+\nu_\mu$ events expected in the invariant mass spectrum of $\tau^+ \to \mu^+\mu^-\mu^+$, separately for the 3μ and 2μ samples and for the three years. For the 3μ sample, the number of events expected in the signal region (last row) is also reported.

	2016		**2017**		**2018**	
	3μ sample	**2μ sample**	**3μ sample**	**2μ sample**	**3μ sample**	**2μ sample**
$\mathcal{B}(D_s^+ \to \phi(\mu^+\mu^-)\pi^+)$	$(1.29 \pm 0.09) \times 10^{-5}$					
$\mathcal{B}(D_s^+ \to \eta(\mu^+\mu^-\gamma)\mu^+\nu_\mu)$	$(7.5 \pm 1.8) \times 10^{-6}$					
$\epsilon_{D_s,Ref}$	$(13.1 \pm 0.4) \times 10^{-4}$		$(28.3 \pm 0.8) \times 10^{-4}$		$(24.6 \pm 0.7) \times 10^{-4}$	
$N_{D_s,Ref}$	43072 ± 397		88474 ± 594		103790 ± 643	
$\epsilon_{D_s,Bkg}$	$(5.1 \pm 0.5) \times 10^{-5}$	$(4.4 \pm 4.4) \times 10^{-7}$	$(6.2 \pm 0.5) \times 10^{-5}$	$(9.6 \pm 6.8) \times 10^{-7}$	$(5.6 \pm 0.5) \times 10^{-5}$	$(2.7 \pm 1.1) \times 10^{-6}$
$N_{D_s,Bkg}$	980 ± 267	11 ± 11	1129 ± 305	22 ± 17	1384 ± 375	84 ± 41
$N_{D_s,Bkg}^{Sig.Reg.}$	43 ± 22	–	61 ± 28	–	56 ± 29	–

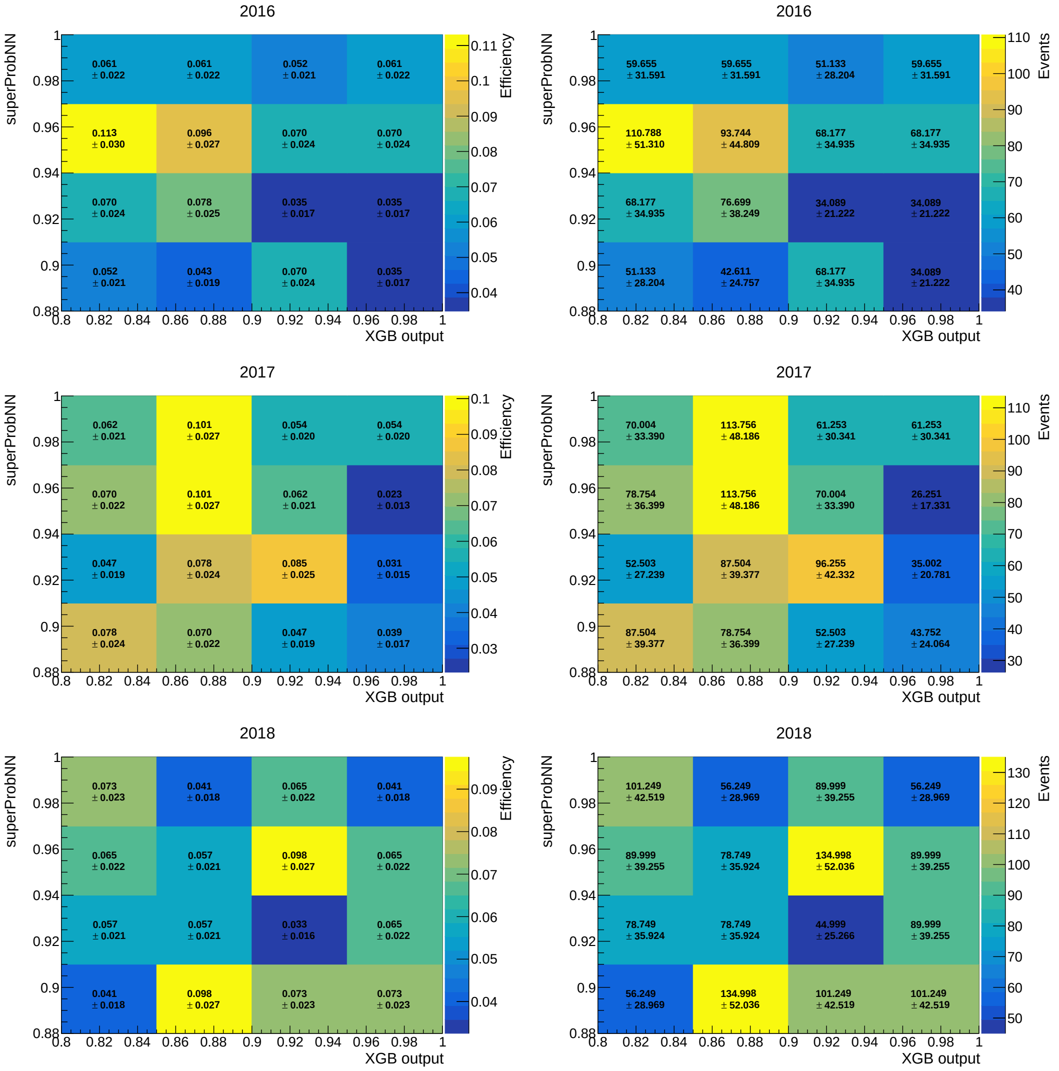


FIGURE 8.3: On the left, map containing the fraction of $D_s^+ \to \eta(\mu^+\mu^-\gamma)\mu^+\nu_\mu$ events falling in each XGB and SuperPNN bin. The fractions are evaluated on simulated events. On the right, corresponding number of $D_s^+ \to \eta(\mu^+\mu^-\gamma)\mu^+\nu_\mu$ events expected in the invariant mass spectrum of the 3μ signal sample. The maps are shown for the three data-taking years.

As described in Section 7.2.3, the limit evaluation in the 3μ sample is performed in bins of XGB and SuperPNN to increase the signal sensitivity. The fraction of $D_s^+ \to \eta(\mu^+\mu^-\gamma)\mu^+\nu_\mu$ events falling in each XGB and SuperPNN bin is evaluated on the simulated sample and reported in the maps shown on the left-hand side of Figure 8.3 for the three data-taking years. The $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ channel is once more used as a normalisation channel to evaluate the corresponding number of $D_s^+ \to \eta(\mu^+\mu^-\gamma)\mu^+\nu_\mu$ events expected in each bin, shown in the maps reported on the right-hand side of Figure 8.3. These values are compared with the total number of background events expected in the mass spectrum, to evaluate whether the contamination of the $D_s^+ \to \eta(\mu^+\mu^-\gamma)\mu^+\nu_\mu$ channel can be considered negligible or has to be modelled in the final fit.

In the last row of Table 8.4, the number of events surviving the minimum requirement on XGB and SuperPNN expected to fall in the signal region are reported for the

TABLE 8.5: List of the decays contributing to the misidentified background with their relative branching fraction. The number of events produced in 5.4 fb^{-1} in the LHCb acceptance shown in the last column is estimated using the LHCb measured inclusive 4π cross section (summed over charm and beauty contributions), the PDG branching fraction and a 20% acceptance efficiency.

Decay	$\mathcal{B}$ [10^{-2}]	$\sigma^{4\pi}(pp \to c\bar{c} + b\bar{b} \to ...)[\mu b]$	**N** [10^9]
$D^+ \to K^-\pi^+\pi^+$	9.38 ±0.16	5044 ±131	511
$D^+ \to K^-K^+\pi^+$	0.968 ±0.018	5044 ±131	53
$D^+ \to \pi^+\pi^+\pi^-$	0.327 ±0.018	5044 ±131	18
$D^+ \to K^-\pi^+\pi^+\pi^0$	6.25 ±0.18	5044 ±131	340
$D^+ \to \pi^+\pi^+\pi^-\pi^0$	1.16 ±0.08	5044 ±131	63
$D^+ \to K^-\pi^+\mu^+\nu$	3.65 ±0.34	5044 ±131	199
$D^+ \to K^-K^+\pi^+\pi^0$	0.662 ±0.032	5044 ±131	36
$D_s^+ \to K^-K^+\pi^+$	5.39 ±0.15	2722 ±387	158
$D_s^+ \to K^+\pi^+\pi^-$	0.65 ±0.04	2722 ±387	19
$D_s^+ \to \pi^-\pi^+\pi^+$	1.08 ±0.04	2722 ±387	32
$D_s^+ \to K^-K^+\pi^+\pi^0$	6.2 ±0.6	2722 ±387	183
$D^{*+} \to D^0(K^-\pi^+\pi^0)\pi^+$	9.75 ±0.35	5044 ±131	531
$\tau^+ \to \pi^+\pi^-\pi^+\nu_\tau$	9.31 ±0.05	137.4 ±21.3	14

3μ sample and for the three data-taking years. Due to the limited statistics available in the $D_s^+ \to \eta(\mu^+\mu^-\gamma)\mu^+\nu_\mu$ simulated sample with two final-state tracks identified as muons, no estimate of the number of events expected in the signal region can be provided for this sub-sample. Nevertheless, the number of $D_s^+ \to \eta(\mu^+\mu^-\gamma)\mu^+\nu_\mu$ events expected in the invariant mass spectrum of the 2μ signal sample is negligible compared to the number of events expected from the combinatorial background, as later shown in Section 8.2.5.

8.2 Background from misidentified tracks

The probability for hadrons to be misidentified as muons is not negligible and leads to a large set of physical background processes that have to be taken into account in the estimate of the background contribution.

A list of the charm decays contributing to the mis-ID background is reported in Table 8.5 with the corresponding branching fractions taken from the PDG [8]. The third column shows the charm production cross section in 4π obtained by summing the prompt and secondary contributions listed in Table 4.5. A rough estimate of the events expected from each of the listed channels can be obtained using Eq. 3.4, assuming an integrated luminosity of 5.4 fb^{-1} and an acceptance efficiency of 20%.

The dominant contributions come from the $D^+ \to K^-\pi^+\pi^+$ and $D^{*+} \to D^0(\to K^-\pi^+\pi^0)\pi^+$ decays. However, since the D^0 can only be partially reconstructed because of the missing π^0, the contribution from the $D^{*+} \to D^0(\to K^-\pi^+\pi^0)\pi^+$ decay can be excluded *a priori* as it is expected to fall far from the signal mass range. A first qualitative estimate of the presence of mis-ID background in the data sample can be obtained by plotting the invariant mass distribution of the signal candidates under different mass hypotheses, such as $K^-\pi^+\pi^+$, $K^-K^+\pi^+$ and $\pi^-\pi^+\pi^+$. The mass distributions are plotted for events surviving the preselection outlined in Chapter 5 and where all the three tracks satisfy the `isMuon` condition. In order to suppress the contribution from the combinatorial background a further requirement is applied on

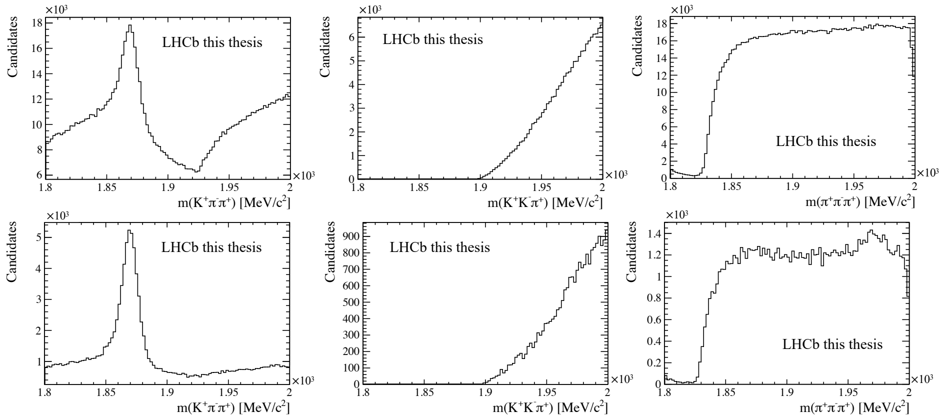


FIGURE 8.4: Distribution of the three muons invariant mass in the 2018 data sample under the $K^-\pi^+\pi^+$ (left), $K^-K^+\pi^+$ (middle) and $\pi^-\pi^+\pi^+$ (right) mass hypotheses. The distributions in the first row are obtained by applying the preselection cuts reported in Chapter 5 while in the second row an additional requirement on XGB (> 0.5) is applied in order to reduce the contamination of the combinatorial background.

the output of the XGB classifier removing all the events with XGB>0.5. The resulting invariant mass distributions for the three mass hypotheses are reported in Figure 8.4 for 2018 data with and without the requirement on XGB.

As expected from the large branching fraction of the $D^+ \to K^-\pi^+\pi^+$ channel (see Table 8.5), a clear peak is visible only in the $K^-\pi^+\pi^+$ mass hypothesis. In the $\pi^+\pi^-\pi^+$ mass hypothesis two small peaks around 1870 MeV/c^2 and 1970 MeV/c^2 become visible after the requirement on XGB.

8.2.1 Strategy for suppressing the $D^+ \to K^-\pi^+\pi^+$ contribution in the 3μ sample

In the 3μ sample, the suppression of $D^+ \to K^-\pi^+\pi^+$ events is achieved by means of a cut applied on the ProbNNk variable of the track with a charge opposite to the mother particle's, referred to as the *negative* track. The distribution of the kaon's ProbNNk for $D^+ \to K^-\pi^+\pi^+$ simulated events is shown in Figure 8.5 together with the distribution of the same variable obtained for the *negative* track in the $\tau^+ \to \mu^+\mu^-\mu^+$ simulated sample.

Requiring ProbNNk<0.05 to the *negative* track in the D^+ invariant mass range [1800, 1900] MeV/c^2 reduces the $D^+ \to K^-\pi^+\pi^+$ contamination to a negligible contribution with an efficiency greater than 99% (see Table 9.9) on the signal channel and smaller than $(0.0001 \pm 0.0007)\%$[2] on the $D^+ \to K^-\pi^+\pi^+$ mode. The performance of this requirement has been tested in combination with the selection applied on XGB and SuperPNN. The effect of this veto on ProbNNk is shown in Figure 8.6 for different invariant mass distributions. As can be seen from the plot on the top-left, no events are left in the $D^+ \to K^-\pi^+\pi^+$ simulated sample when the veto on ProbNNk is applied. On the top-right, the invariant mass distribution in data is shown, under the $K^-\pi^+\pi^+$ mass hypothesis, after each step of the selection. The requirement XGB>0.8 is able to suppress most of the combinatorial background contribution (plots are shown in log scale for a better understanding) and accentuate the $D^+ \to K^-\pi^+\pi^+$ peak. After the cut on SuperPNN (>0.88) and the veto

---

[2] Efficiency estimated as average over the three data-taking years.

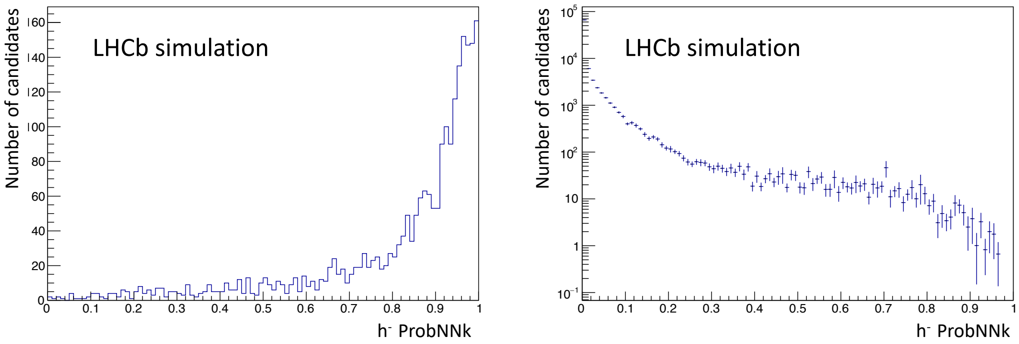


FIGURE 8.5: Distribution of the ProbNNk variable for the track with a charge opposite to the charge of the mother particle in the $D^+ \to K^-\pi^+\pi^+$ (left) and $\tau^+ \to \mu^+\mu^-\mu^+$ (right) simulated samples.

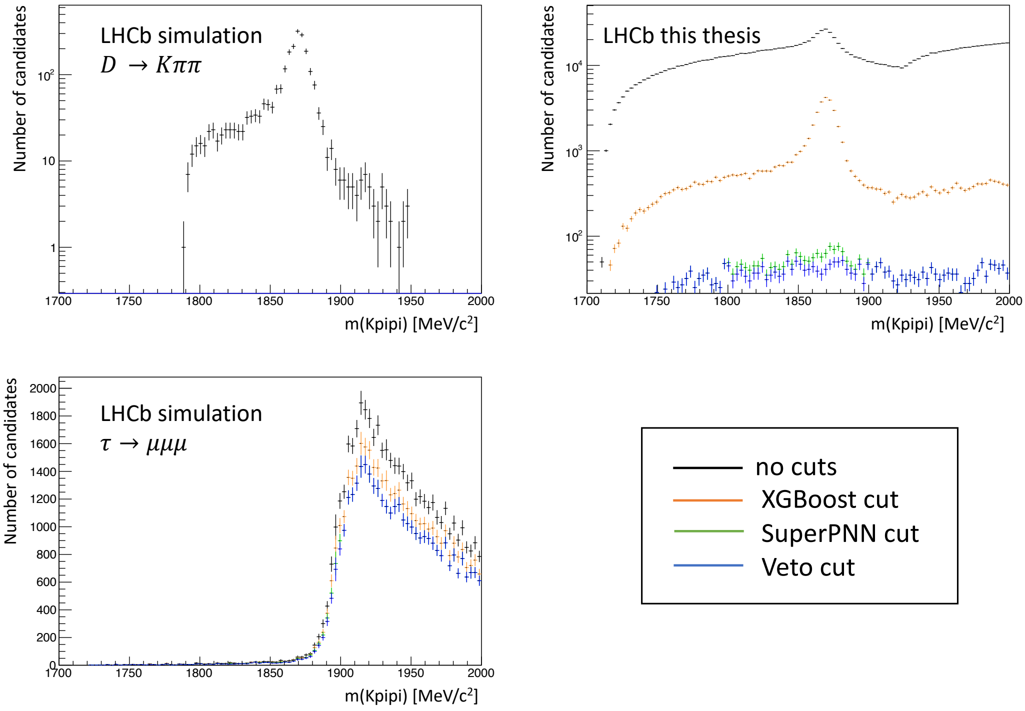


FIGURE 8.6: Invariant mass distribution under the $K^-\pi^+\pi^+$ mass hypothesis for the simulated $D^+ \to K^-\pi^+\pi^+$ (left), data (right) and simulated signal (bottom) sample. Three different requirements are applied: XGB>0.8, SuperPNN>0.88 and the veto on ProbNNk. In the $D^+ \to K^-\pi^+\pi^+$ simulated sample, no XGB requirement is applied since the requirement on SuperPNN already cut out all simulated events.

on ProbNNk, though, only a small peak of $D^+ \to K^-\pi^+\pi^+$ candidates is visible. No significant effect is observed in the $\tau^+ \to \mu^+\mu^-\mu^+$ simulated sample, as shown in the bottom plot. The strategy used to estimate the residual contribution of the $D^+ \to K^-\pi^+\pi^+$ mode in the signal region is presented in the following.

8.2.2 Estimate of mis-ID background contamination in the signal region of the 3μ sample

As previously mentioned, in addition to the selection described in Chapter 5, events in the data sample are required to satisfy the minimum requirement on XGB (>0.8) in order to reduce the contribution of the combinatorial background. When the requirement on XGB is applied, three structures become visible in the data sidebands as later shown in Figure 8.10, the $D^+ \to K^-\pi^+\pi^+$ peak in the left sideband and the $D^+ \to \pi^-\pi^+\pi^+$ and $D_s^+ \to \pi^-\pi^+\pi^+$ peaks in the right sideband. Therefore, the residual contamination of the three charmed modes in the signal region should be evaluated. This is done by fitting the 3μ invariant mass distribution in the data sidebands and extrapolating the corresponding yields to the signal region.

The minimum requirement on SuperPNN (>0.88) significantly reduces the contribution of the charmed modes. Therefore, in order to obtain a reliable estimate of their mass shape, the fit to the data sidebands is performed first without any requirement on SuperPNN and then requiring SuperPNN<0.88. The final yield in the signal region is computed, for each mis-ID background, as difference between the yields evaluated with the two fit results.

The $D^+ \to K^-\pi^+\pi^+$ mass shape is described with a double-sided Crystal Ball function defined as

$$C(m,\mu,\sigma,\alpha_{L,R},n_{L,R}) = \begin{cases} e^{-\frac{\alpha_L^2}{2}}\left(\frac{n_L}{\alpha_L}\right)^{n_L}\left(\frac{n_L}{\alpha_L}-\alpha_L-\frac{x-\mu}{\sigma}\right)^{-n_L} & \frac{x-\mu}{\sigma} < -\alpha_L \\ \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]\sigma & \frac{x-\mu}{\sigma} < \alpha_R \\ e^{-\frac{\alpha_R^2}{2}}\left(\frac{n_R}{\alpha_R}\right)^{n_R}\left(\frac{n_R}{\alpha_R}-\alpha_R+\frac{x-\mu}{\sigma}\right)^{-n_R} & \frac{x-\mu}{\sigma} > \alpha_R \end{cases} \tag{8.2}$$

where μ is the mean, σ is the resolution, and $\alpha_{L(R)}$ and $n_{L(R)}$ the left (right) tail parameters. The $D^+ \to \pi^-\pi^+\pi^+$ and $D_s^+ \to \pi^-\pi^+\pi^+$ peaks are both described with a Johnson's S_U function (see Eq. 6.1), while an exponential distribution is used to model the combinatorial background (see Eq. 6.3).

The tail parameters of the Crystal Ball and of the two Johnson's S_U functions are extrapolated from a fit to simulated events where all the three final-state tracks are required to satisfy the `isMuon` condition. No requirement on SuperPNN is applied on simulated events. Because of the limited statistics available in the $D^+ \to \pi^-\pi^+\pi^+$ simulated sample, its tail parameters are taken from the fit to the $D_s^+ \to \pi^-\pi^+\pi^+$ mode. The results of the fit to the invariant mass distribution of the $D^+ \to K^-\pi^+\pi^+$ and $D_s^+ \to \pi^-\pi^+\pi^+$ modes under the 3μ mass hypothesis are reported in Tables 8.6 and 8.7 and displayed in Figure 8.7.

It should be noted that, because of the requirement on `isMuon`, the statistics of the simulated sample available for the single data-taking year would not be sufficient to provide a reliable fit result. For this reason, the samples generated for the three years are combined and a unique fit is performed on the whole simulated sample for the two charmed modes. Since no differences are observed in the fit parameters when fitting the three years separately, no systematics is assigned for using a unique signal description.

The validation of the mass shape parameters is performed by fitting the $\tau^+ \to \mu^+\mu^-\mu^+$ data sidebands under different requirements on ProbNNk to separate the single charmed components. First, the fit is performed requiring the ProbNNk of the *negative* track to be greater than 0.05 in order to reject $D^+ \to \pi^-\pi^+\pi^+$ and $D_s^+ \to \pi^-\pi^+\pi^+$ candidates. In this way, the $D^+ \to K^-\pi^+\pi^+$ mass shape extrapolated

TABLE 8.6: List of the fit parameters used to described the $D^+ \to K^-\pi^+\pi^+$ pdf. The results of the fit performed on the full Run 2 simulation sample without any requirement on SuperPNN are reported.

Parameter	$D^+ \to K^-\pi^+\pi^+$
Yield	1401 ± 91
μ [MeV/c^2]	1741.6 ± 1.0
σ [MeV/c^2]	10.86 ± 0.94
α_L	0.15 ± 0.02
n_L	2.78 ± 0.65
α_R	1.52 ± 0.28
n_R	1.27 ± 0.71

TABLE 8.7: List of the fit parameters used to described the $D_s^+ \to \pi^-\pi^+\pi^+$ pdf. The results of the fit performed on the full Run 2 simulation sample without any requirement on SuperPNN are reported.

Parameter	$D_s^+ \to \pi^-\pi^+\pi^+$
Yield	484 ± 22
μ [MeV/c^2]	1953.92 ± 3.24
λ [MeV/c^2]	12.08 ± 1.93
γ	1.27 ± 0.30
δ	1.14 ± 0.17

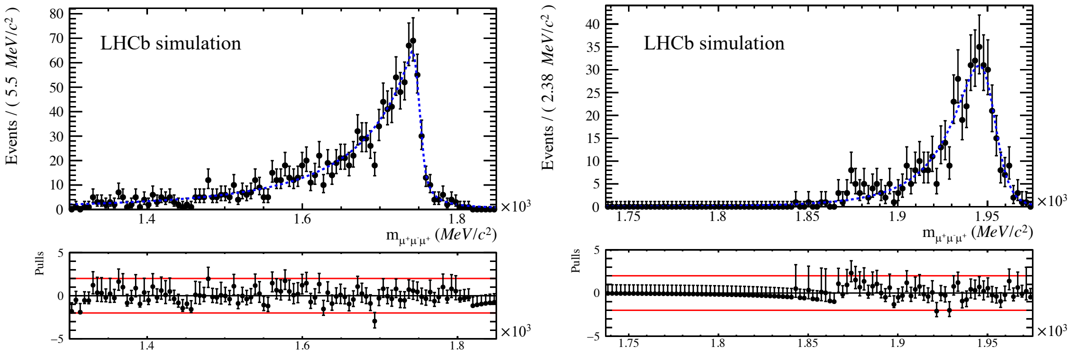


FIGURE 8.7: Invariant mass distribution of the $D^+ \to K^-\pi^+\pi^+$ (left) and $D_s^+ \to \pi^-\pi^+\pi^+$ (right) modes under the 3μ mass hypothesis for simulated events. The results of the fit are superimposed.

from the fit to the simulated sample can be compared with the one extrapolated from the fit to the data sample. When the reverse cut on ProbNNk is applied, the $D^+ \to K^-\pi^+\pi^+$ is suppressed and the $D^+ \to \pi^-\pi^+\pi^+$ and $D_s^+ \to \pi^-\pi^+\pi^+$ modes can be isolated to validate their mass shapes. The results of the fit to the data sidebands obtained requiring ProbNNk>0.05 and ProbNNk<0.05 in the invariant mass range [1800, 1900] MeV/c^2 to the *negative* track are reported in Table 8.8 and Table 8.9, respectively.

Following Eq. 3.4, the yield of the $D^+ \to \pi^-\pi^+\pi^+$ mode is fixed by scaling the yield determined for the $D_s^+ \to \pi^-\pi^+\pi^+$ mode by a factor of 0.56, computed as

$$r_{D^+} = \frac{\mathcal{B}(D^+ \to \pi^-\pi^+\pi^+) \times \sigma^{4\pi}(b \to D^+)}{\mathcal{B}(D_s^+ \to \pi^-\pi^+\pi^+) \times \sigma^{4\pi}(b \to D_s^+)}\,. \tag{8.3}$$

TABLE 8.8: List of the fit parameters used to describe the mass pdf in the data sidebands. The results of the fit performed applying the minimum requirement on XGB (>0.8), ProbNNk>0.05 to the *negative* track and without any requirement on SuperPNN, are reported for the three data-taking years.

Parameter	**2016**	**2017**	**2018**
$n_{D^+\to K^-\pi^+\pi^+}$	30485 ± 131	28343 ± 131	32694 ± 125
μ [MeV/c^2]	1732.0 ± 1.0	1730.4 ± 1.1	1733.6 ± 0.4
σ [MeV/c^2]	14.69 ± 0.97	15.92 ± 1.07	16.86 ± 0.89
α_L	0.18 ± 0.07	0.10 ± 0.03	0.14 ± 0.04
n_L	0.14 ± 0.04	0.12 ± 0.02	0.11 ± 0.03
α_R	4.58 ± 3.56	1.56 ± 0.10	1.66 ± 0.09
n_R	0.80 ± 1.53	1.02 ± 1.26	0.26 ± 1.81
$n_{comb.bkg}$	10565 ± 275	7834 ± 247	10834 ± 250
α	0.0001 ± 0.0003	0.0004 ± 0.0004	0.0003 ± 0.0003

The values of the two branching fractions are reported in Table 8.5, while the cross section for the inclusive production of D^+ and D_s in 4π are obtained by summing the prompt and secondary contributions listed in Table 4.5. The projections of the fits on the invariant mass distribution are displayed in Figure 8.8 and Figure 8.9 for the three data-taking years. In both cases, a good agreement is observed with the results obtained from the fit to the simulated sample, with the exception of n_L, whose value is fixed from the fit to the data sample. For the remaining parameters the value from the fit to the simulation is used in the final fit.

TABLE 8.9: List of the fit parameters used to describe the mass pdf in the data sidebands. The results of the fit performed applying the minimum requirement on XGB (>0.8), ProbNNk<0.05 to the *negative* track and without any requirement on SuperPNN, are reported for the three data-taking years. The parameters λ_D, γ_D and δ_D are shared between the $D^+ \to \pi^-\pi^+\pi^+$ and $D_s^+ \to \pi^-\pi^+\pi^+$ modes.

Parameter	**2016**	**2017**	**2018**
$n_{D_s^+\to\pi^-\pi^+\pi^+}$	2017 ± 15	1942 ± 12	2052 ± 16
μ_{D_s} [MeV/c^2]	1957.16 ± 1.17	1956.88 ± 1.76	1955.66 ± 1.08
λ_D [MeV/c^2]	10.68 ± 1.18	10.41 ± 1.53	9.75 ± 1.17
γ_D	0.64 ± 0.51	1.10 ± 0.87	0.95 ± 0.43
δ_D	1.42 ± 0.64	1.77 ± 0.98	1.59 ± 0.85
r_{D^+}		0.56	
μ_{D^+} [MeV/c^2]	1856.33 ± 1.54	1853.62 ± 1.99	1856.87 ± 1.54
$n_{comb.bkg}$	54322 ± 377	44822 ± 281	54285 ± 412
α	0.0003 ± 0.0001	-0.0001 ± 0.0001	0.0001 ± 0.0001

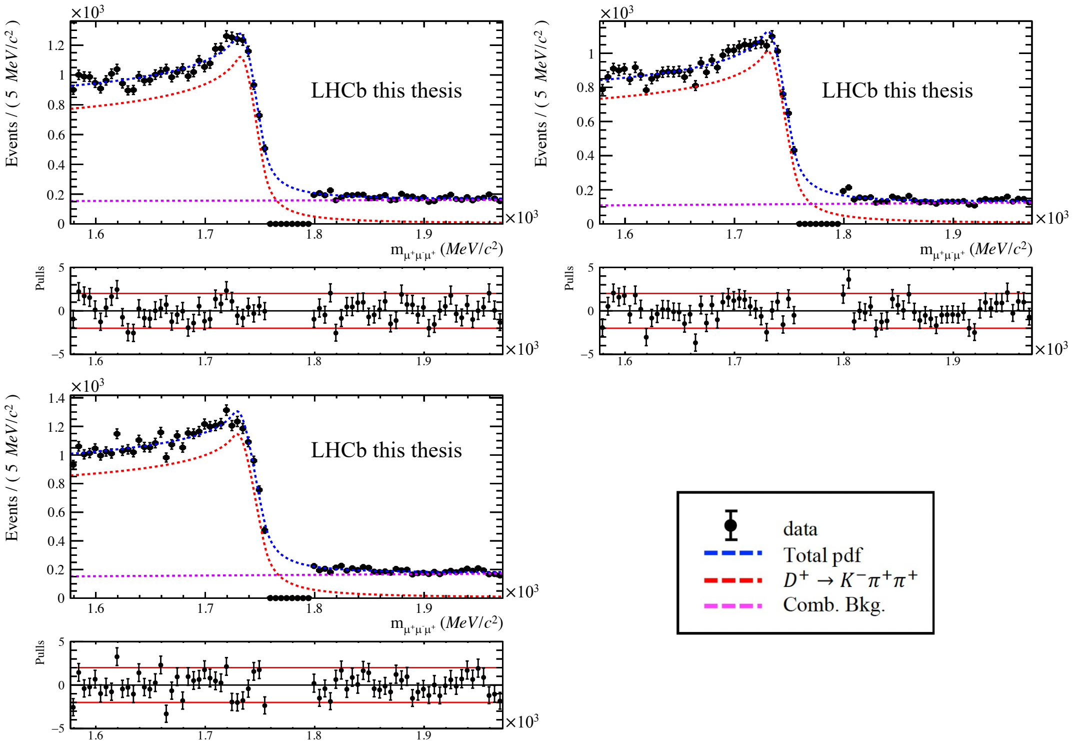


FIGURE 8.8: Data sidebands invariant mass distribution in the 3μ sample after requiring ProbNNk>0.05 to the *negative* track and without any requirement on SuperPNN. The results of the fits are superimposed and shown separately for the three data-taking periods: 2016 (top-left), 2017 (top-right), 2018 (bottom).

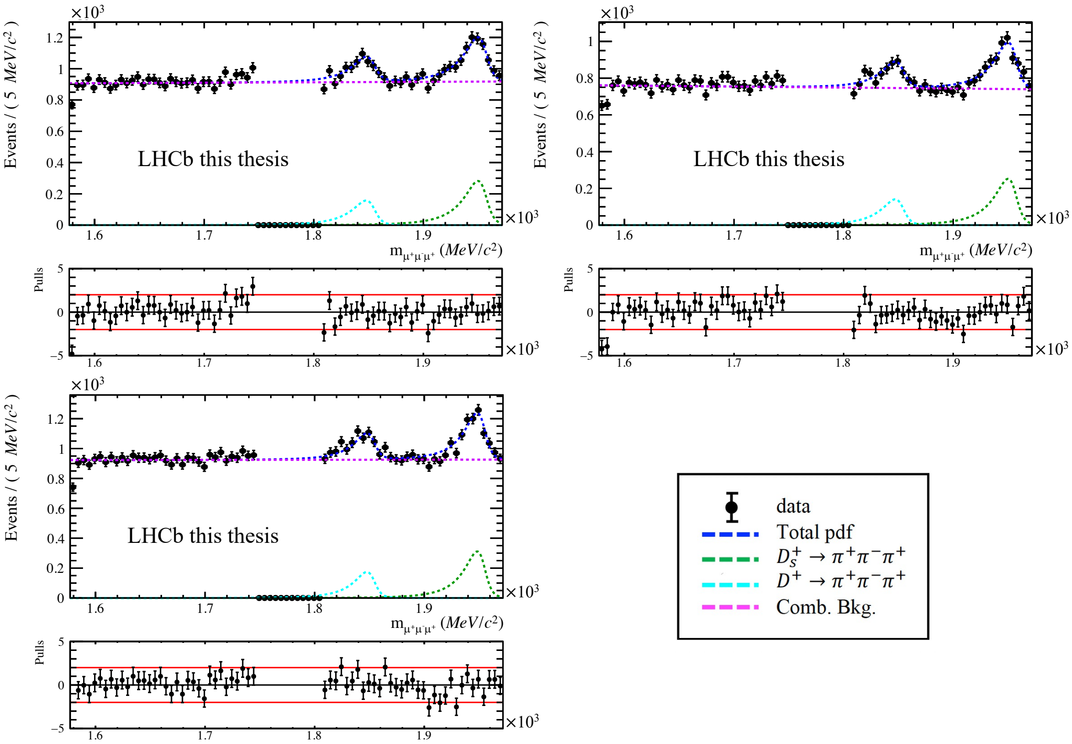


FIGURE 8.9: Data sidebands invariant mass distribution in the 3μ sample after requiring ProbNNk<0.05 to the *negative* track and without any requirement on SuperPNN. The results of the fits are superimposed and shown separately for the three data-taking periods: 2016 (top-left), 2017 (top-right), 2018 (bottom).

Once the mass shape of the charmed modes is validated, the final fit to the data sidebands is performed first, applying only the requirement on XGB, and second, adding the reverse cut on SuperPNN. The result of the two fits are displayed in Figure 8.10 for the three data-taking years. The final yield expected in the signal region, reported in Table 8.10, is evaluated as difference between the yields extrapolated from the two fits. For each background source, the uncertainty on the expected yield in the signal region is determined as:

$$\sigma^{sig} = \sqrt{(\sigma^{sig}_{XGB})^2 + (\sigma^{sig}_{XGB\&PID})^2 - 2\rho\sigma^{sig}_{XGB}\sigma^{sig}_{XGB\&PID}}\,, \tag{8.4}$$

where σ^{sig}_{XGB} is the statistical uncertainty resulting from the fit performed with the XGB requirement only, $\sigma^{sig}_{XGB\&PID}$ is the statistical uncertainty resulting from the fit performed with the additional requirement on SuperPNN and ProbNNk, and ρ is the correlation between the two fits assumed to be equal to $\rho = n^{tot}_{XGB\&PID}/n^{tot}_{XGB}$. The yields obtained for the three charmed modes are all compatible with 0 within two standard deviations, meaning that their residual contribution in the signal region is negligible compared to the contribution of the combinatorial background.

TABLE 8.10: Number of candidates for the $D^+ \to K^-\pi^+\pi^+$, $D^+ \to \pi^-\pi^+\pi^+$ and $D^+_s \to \pi^-\pi^+\pi^+$ modes and for the combinatorial background surviving the minimum requirements on XGB and SuperPNN and the veto on ProbNNk. These values are determined as differences between the yields expected in the signal region evaluated without any requirement on SuperPNN and with the reverse cut on SuperPNN. The statistical uncertainties are estimated taking into account the correlation between the two fits. Since the fit is performed without any veto requirement, the yields of the $D^+ \to K^-\pi^+\pi^+$ component have to be rescaled by the veto efficiency which is smaller than $(0.0001 \pm 0.0007)\%$ as reported in Section 8.2.1.

Parameter	**2016**	**2017**	**2018**
$n_{D^+\to K^-\pi^+\pi^+}$	100 ± 44	156 ± 51	66 ± 36
$n_{D^+\to K^-\pi^+\pi^+}$ (with veto)	$< 0.10 \pm 0.04$	$< 0.16 \pm 0.05$	$< 0.07 \pm 0.04$
$n_{D^+\to \pi^-\pi^+\pi^+}$	17 ± 14	15 ± 12	1 ± 3
$n_{D^+_s\to \pi^-\pi^+\pi^+}$	2 ± 3	2 ± 4	0 ± 3
$n^{sig}_{comb.bkg}$	608 ± 37	614 ± 33	754 ± 43

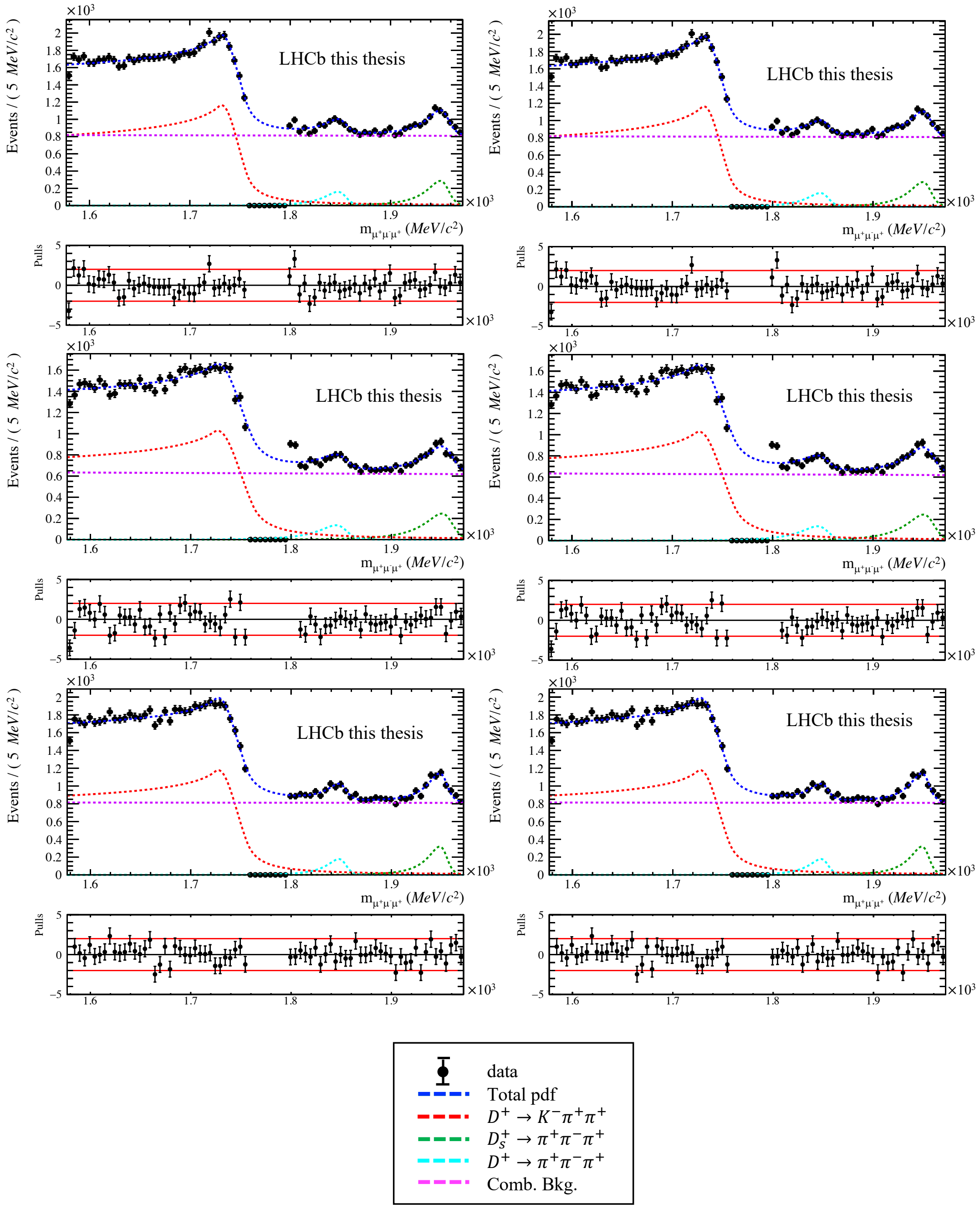


FIGURE 8.10: Data sidebands invariant mass distribution in the 3μ sample after applying the XGB requirement only (left) and adding the reverse cut on SuperPNN (right). The results of the fits are superimposed and shown separately for the three data-taking periods: 2016 (top), 2017 (middle), 2018 (bottom).

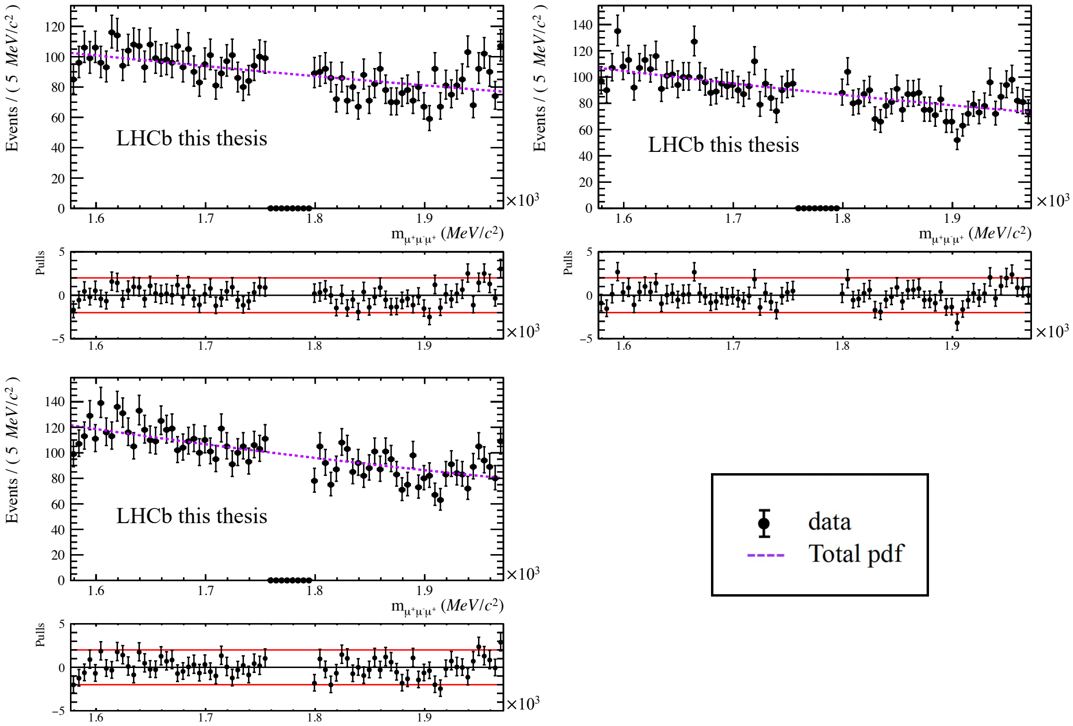


FIGURE 8.11: Fit to the data sidebands invariant mass distribution in the 3μ sample after the minimum requirements on XGB and SuperPNN and the veto on ProbNNk assuming the contribution of the combinatorial background only. Plots are shown for the three data-taking years: 2016 (top-left), 2017 (top-right), 2018 (bottom).

8.2.3 Fit to the data sidebands in the final binning scheme

The invariant mass distribution for the 3μ sample is reported in Figure 8.11 for events satisfying the minimum requirement on XGB and SuperPNN and the veto on ProbNNk. An extended maximum likelihood fit is performed to the outer sidebands using an exponential distribution (see Eq. 6.3); the result of the fits are superimposed. When looking at the pull distributions, a trend becomes visible in correspondence of the D^+ and D_s mass peaks, suggesting the presence of the $D^+ \to \pi^-\pi^+\pi^+$ and $D_s^+ \to \pi^-\pi^+\pi^+$ mis-ID backgrounds in the right sideband. Tightening the PID selection does not remove these residual mis-ID contributions without rejecting a significant fraction of signal candidates. For this reason, the final fit is performed including the two charmed modes in the total pdf in order to avoid any bias in the determination of the slope of the exponential distribution used to describe the combinatorial background. The projection of the final fit on the 3μ invariant mass is shown in Figure 8.12 for the three data-taking years; the D^+ and D_s mass peaks are described using the models illustrated in Section 8.2.2. In the final fit no contribution from the $D^+ \to K^-\pi^+\pi^+$ mode is considered since, as shown in Figure 8.12, no excess is visible in the left sideband once the veto requirement is applied.

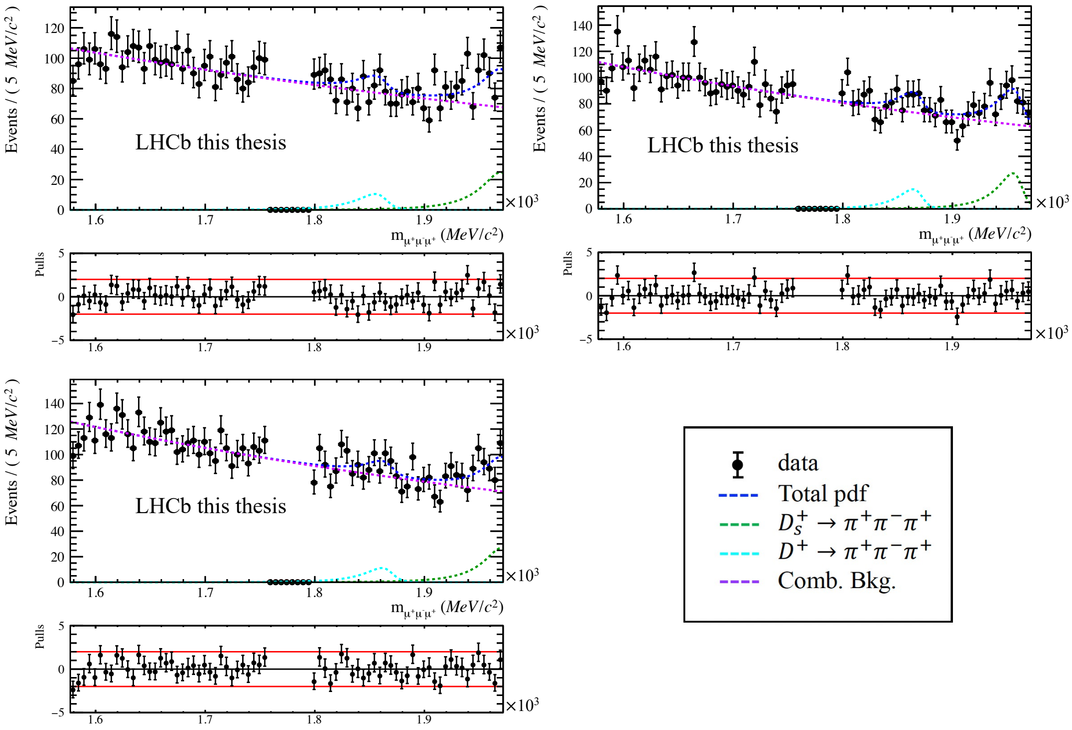


FIGURE 8.12: Fit to the data sidebands invariant mass distribution in the 3μ sample after the minimum requirements on XGB and SuperPNN and the veto on ProbNNk assuming the contribution of the combinatorial background and of the $D^+ \to \pi^-\pi^+\pi^+$ and $D_s^+ \to \pi^-\pi^+\pi^+$ modes. Plots are shown for the three data-taking years: 2016 (top-left), 2017 (top-right), 2018 (bottom).

The list of the fit parameters used to model the outer sidebands in the 3μ sample assuming the contribution of the combinatorial background only and adding the contribution of the $D^+ \to \pi^-\pi^+\pi^+$ and $D_s^+ \to \pi^-\pi^+\pi^+$ modes is reported in Table 8.11 for the three data-taking periods. The number of $D_s^+ \to \eta(\mu^+\mu^-\gamma)\mu^+\nu_\mu$ events expected to fall in the signal region (see Table 8.4) is negligible compared to the number of events expected from the combinatorial background. Therefore, no contribution from the $D_s^+ \to \eta(\mu^+\mu^-\gamma)\mu^+\nu_\mu$ mode is taken into account in the final fit.

TABLE 8.11: List of the fit parameters used to describe the mass pdf assuming the combinatorial-only and the combinatorial+mis-ID hypotheses for the three data-taking years. The fit is performed to the data sidebands in the 3μ sample after the minimum requirements on XGB and SuperPNN and the veto on ProbNNk. The parameters λ_D, γ_D and δ_D are shared between the $D^+ \to \pi^-\pi^+\pi^+$ and $D_s^+ \to \pi^-\pi^+\pi^+$ decay modes.

Parameter	**2016**	**2017**	**2018**
Combinatorial background			
$n_{comb.bkg}$	5977 ± 50	5978 ± 50	6668 ± 53
$n^{sig}_{comb.bkg}$	709 ± 9	706 ± 8	787 ± 11
α	-0.0007 ± 0.0001	-0.0010 ± 0.0001	-0.0011 ± 0.0001
Combinatorial + mis-ID background			
$n_{D_s^+ \to \pi\pi\pi}$	164 ± 6	186 ± 7	161 ± 8
$n^{sig}_{D_s^+ \to \pi\pi\pi}$	1 ± 1	1 ± 1	1 ± 1
$\mu_{D_s^+}$ [MeV/c^2]	1979.14 ± 7.09	1963.29 ± 5.17	1978.70 ± 7.82
λ_D [MeV/c^2]	14.16 ± 6.03	11.37 ± 4.59	13.13 ± 5.12
γ_D		1.27	
δ_D		1.14	
r_{D^+}		0.56	
$n^{sig}_{D^+ \to \pi^-\pi^+\pi^+}$	6 ± 4	3 ± 3	4 ± 3
μ_{D^+} [MeV/c^2]	1865.24 ± 6.08	1872.19 ± 5.17	1870.56 ± 5.43
$n_{comb.bkg}$	5742 ± 105	5710 ± 106	6437 ± 99
$n^{sig}_{comb.bkg}$	677 ± 10	668 ± 9	753 ± 12
α	-0.0011 ± 0.0002	-0.0015 ± 0.0002	-0.0015 ± 0.0002

The same fit is repeated in the 16 XGB and SuperPNN bins to finally extrapolate the number of background events expected in the signal region later used in the evaluation of the expected limit. In the various bins, the $D^+ \to \pi^-\pi^+\pi^+$ and $D_s^+ \to \pi^-\pi^+\pi^+$ modes are described with two Johnson's S_U functions fixing the shape parameters to the values reported in Table 8.11 and scaling the yields according to the single bin efficiency of the two modes.

The single bin efficiency is evaluated by fitting the 3μ invariant mass distribution in each bin and dividing the yields obtained for the $D^+ \to \pi^-\pi^+\pi^+$ and $D_s^+ \to \pi^-\pi^+\pi^+$ modes by the yields reported in Table 8.11. As shown in Table 7.5, the optimisation of the binning scheme lead to different, yet very similar, SuperPNN schemes depending on the requirement on XGB. At this step, a unique SuperPNN scheme is assumed for all XGB bins. The single bin efficiency is reported for the $D_s^+ \to \pi^-\pi^+\pi^+$ mode in Table 8.12.

As shown in Figure 8.12, the contribution of the $D^+ \to \pi^-\pi^+\pi^+$ after the minimum requirement on XGB and SuperPNN is very low. For this reason, and because of the similarity between the two modes, the $D^+ \to \pi^-\pi^+\pi^+$ mode is assumed to have the same single bin efficiencies of the $D_s^+ \to \pi^-\pi^+\pi^+$ mode. The yield of the $D^+ \to \pi^-\pi^+\pi^+$ mode in the single XGB and SuperPNN bin is obtained by scaling the yields determined for the $D_s^+ \to \pi^-\pi^+\pi^+$ mode by a factor of 0.56 (see Eq. 8.3).

TABLE 8.12: Efficiencies of the $D_s^+ \to \pi^-\pi^+\pi^+$ mode in bins of XGB and SuperPNN, defined with about the same ranges as in the default final binning scheme. The single bin efficiencies for the $D^+ \to \pi^-\pi^+\pi^+$ mode are assumed to be the same as the $D_s^+ \to \pi^-\pi^+\pi^+$ ones, given the similarity of the two decay modes.

	XGB bins			
	[0.800, 0.951]	[0.951, 0.983]	[0.983, 0.993]	[0.993, 1.000]
$D_{(s)}^+ \to \pi^-\pi^+\pi^+$	$(83 \pm 14)\%$	$(10 \pm 3)\%$	$(5 \pm 2)\%$	$(2 \pm 1)\%$
	SuperPNN bins			
	[0.880, 0.976]	[0.976, 0.990]	[0.990, 0.995]	[0.995, 1.000]
$D_{(s)}^+ \to \pi^-\pi^+\pi^+$	$(90 \pm 12)\%$	$(6 \pm 2)\%$	$(3 \pm 1)\%$	$(1 \pm 1)\%$

The maps containing the number of $D^+ \to \pi^-\pi^+\pi^+$ and $D_s^+ \to \pi^-\pi^+\pi^+$ candidates expected in the signal region in the 3μ sample are reported in Figure 8.13 for the three data-taking periods.

Once the yields of two mis-ID modes are fixed, the final fit to the outer sidebands is performed simultaneously in the 16 XGB and SuperPNN bins as defined in Table 7.5. The expected number of background events in the signal region is extrapolated from the fit and reported in Figure 8.14 for the various bins. The result of the fits in the 16 bins are reported in Figure 8.15 for 2018 and in Appendix C for 2017 and 2016.

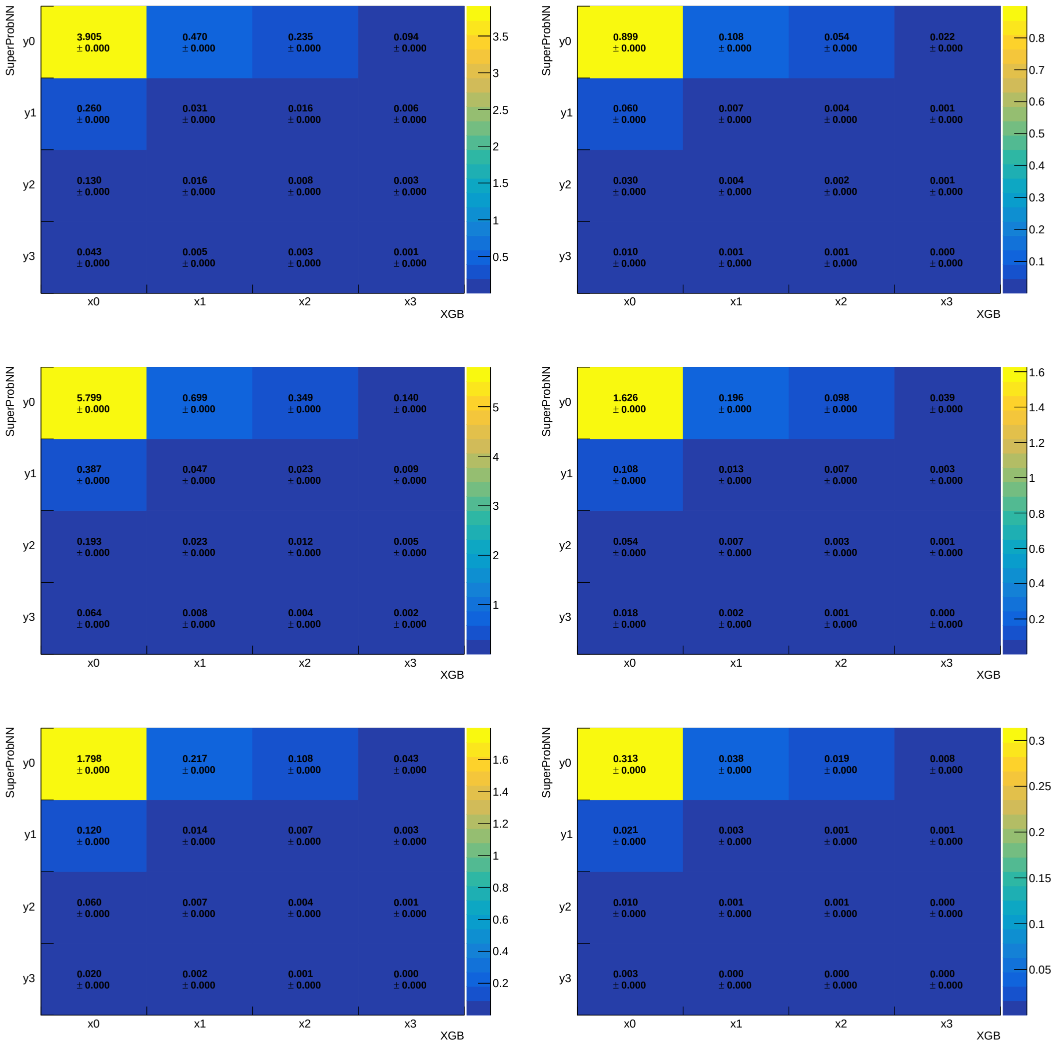
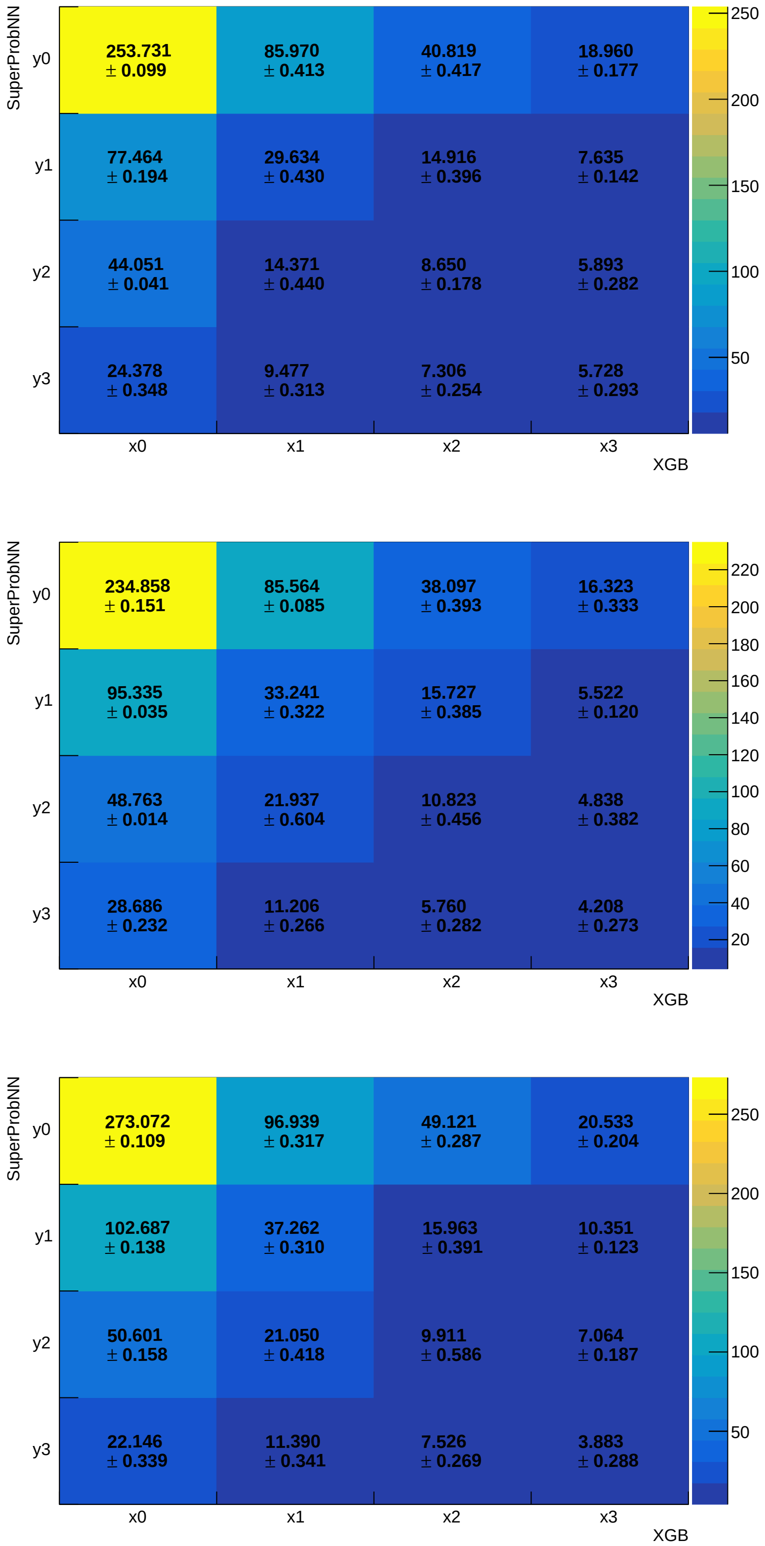


FIGURE 8.13: Number of $D^+ \to \pi^-\pi^+\pi^+$ (left) and $D_s^+ \to \pi^-\pi^+\pi^+$ (right) candidates expected in the signal region in the 3μ sample in bins of XGB and SuperPNN. The maps are reported for the three data-taking periods: 2016 (top), 2017 (middle) and 2018 (bottom).



SuperProbNN	x0	x1	x2	x3
y0	253.731 ± 0.099	85.970 ± 0.413	40.819 ± 0.417	18.960 ± 0.177
y1	77.464 ± 0.194	29.634 ± 0.430	14.916 ± 0.396	7.635 ± 0.142
y2	44.051 ± 0.041	14.371 ± 0.440	8.650 ± 0.178	5.893 ± 0.282
y3	24.378 ± 0.348	9.477 ± 0.313	7.306 ± 0.254	5.728 ± 0.293

SuperProbNN	x0	x1	x2	x3
y0	234.858 ± 0.151	85.564 ± 0.085	38.097 ± 0.393	16.323 ± 0.333
y1	95.335 ± 0.035	33.241 ± 0.322	15.727 ± 0.385	5.522 ± 0.120
y2	48.763 ± 0.014	21.937 ± 0.604	10.823 ± 0.456	4.838 ± 0.382
y3	28.686 ± 0.232	11.206 ± 0.266	5.760 ± 0.282	4.208 ± 0.273

SuperProbNN	x0	x1	x2	x3
y0	273.072 ± 0.109	96.939 ± 0.317	49.121 ± 0.287	20.533 ± 0.204
y1	102.687 ± 0.138	37.262 ± 0.310	15.963 ± 0.391	10.351 ± 0.123
y2	50.601 ± 0.158	21.050 ± 0.418	9.911 ± 0.586	7.064 ± 0.187
y3	22.146 ± 0.339	11.390 ± 0.341	7.526 ± 0.269	3.883 ± 0.288

FIGURE 8.14: Total background yield expected in the signal region in the 3μ sample in each XGB and SuperPNN bin. The maps are reported for the three data-taking periods: 2016 (top-left), 2017 (top-right) and 2018 (bottom).

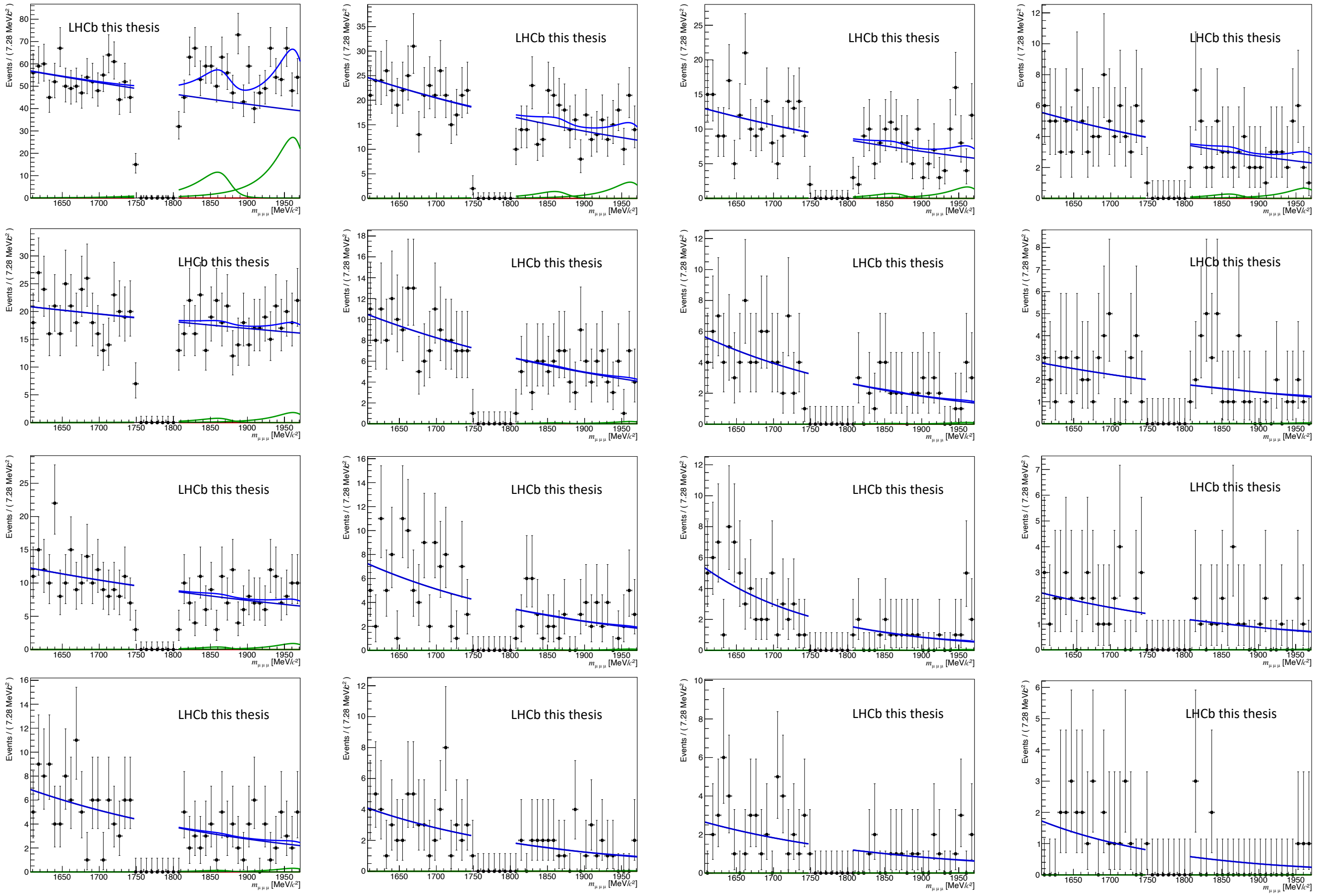


FIGURE 8.15: Results of the simultaneous fit performed in the 16 XGB and SuperPNN bins on 2018 data outer sidebands in the 3μ sample assuming the contribution of the combinatorial background (blue line) and of the $D^+ \to \pi^-\pi^+\pi^+$ and $D_s^+ \to \pi^-\pi^+\pi^+$ modes (green lines). Plots are displayed following the same layout of the maps in Figure 8.14.

The simultaneous fit is validated by means of 200 pseudo-experiments where the yields of the different background components are varied according to a Poisson distribution. For each pseudo-experiment the fit is performed using the same fitting model and the pull distributions for the background shape and for the normalisation parameters are produced in order to check for any deviation from a Gaussian shape. The pull distributions, reported in Appendix C for the three data-taking periods, do not show any significant deviation from a Gaussian distribution.

8.2.4 Background from misidentified tracks in the 2μ sample

As done in the 3μ sample, also in the 2μ sample the outer sidebands invariant mass distribution is plotted under different mass hypotheses, *i.e.* $K^-\pi^+\pi^+$, $K^-K^+\pi^+$ and $\pi^-\pi^+\pi^+$, to obtain a qualitative estimate of the possible mis-ID contributions. The mass distributions are reported in Figure 8.16 for the 2018 data sample. The plots in the first row are obtained for candidates surviving the optimal requirement on BDTGKT (see Table 7.9) whose effect is to suppress the combinatorial background.

The observations are once again in agreement with the branching fractions listed in Table 8.5: a sharp peak is visible in the distribution obtained in the $K^-\pi^+\pi^+$ mass hypothesis, where the kaon mass is assigned to the track whose charge is opposite to the one of the mother particle. No structure is visible in the $K^-K^+\pi^+$ mass hypothesis, while two excesses are noticeable in the $\pi^-\pi^+\pi^+$ mass hypothesis in correspondence of the D^+ and D_s mass.

In the bottom row, the same distributions are obtained adding the requirement on the output of the BDTGPID classifier trained to suppress backgrounds originated from misidentified tracks. The overall effect is to reduce the number of candidates surviving the BDTGKT requirement by more than 90%. Both in the $K^-\pi^+\pi^+$ and $\pi^-\pi^+\pi^+$ mass hypothesis, the remaining statistics does not allow to disentangle the presence of residual mass peaks from statistical fluctuations.

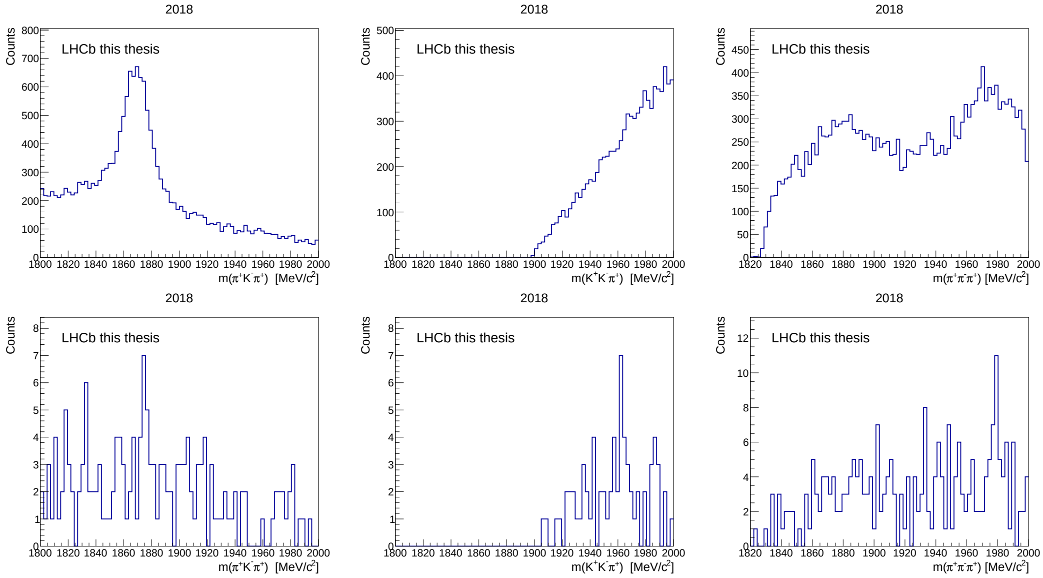


FIGURE 8.16: Outer sidebands invariant mass distribution in the 2μ sample under different mass hypotheses ($K^-\pi^+\pi^+$, $K^-K^+\pi^+$ and $\pi^-\pi^+\pi^+$) after the requirement on BDTGKT (top) and with the additional requirement on BDTGPID (bottom). The distributions are obtained for the 2018 data sample.

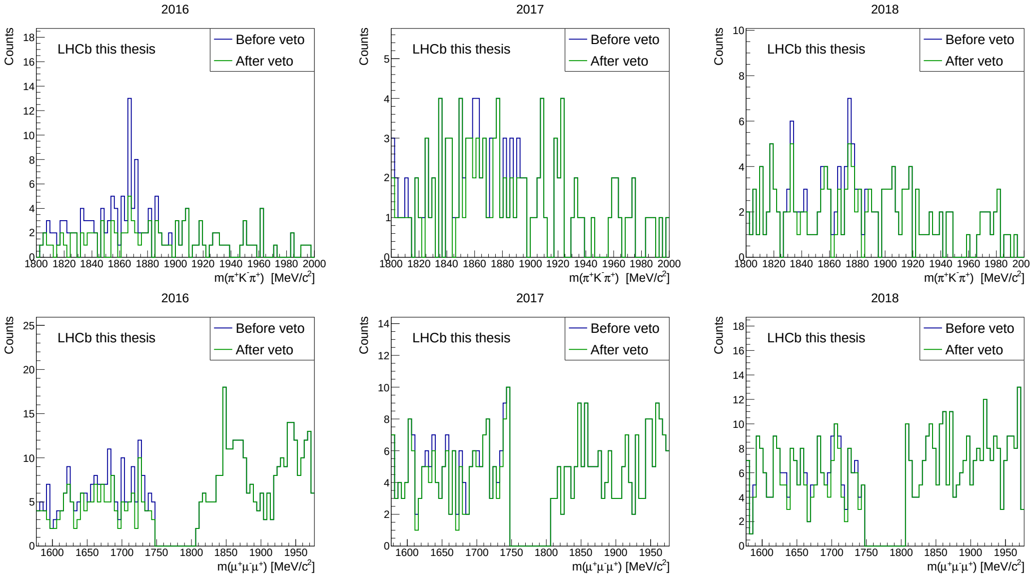


FIGURE 8.17: Outer sidebands invariant mass distribution in the 2μ sample under the $K^-\pi^+\pi^+$ (top) and 3μ (bottom) mass hypotheses before and after applying the veto on ProbNNk on the track whose charge is opposite to the one of the mother particle.

In the 3μ sample, the suppression of the $D^+ \to K^-\pi^+\pi^+$ mode is achieved by requiring ProbNNk<0.05 in the D^+ mass range [1800,1900] MeV/c^2 to the track with a charge opposite to the one of the mother particle. The effect of this cut in the 2μ sample is shown in Figure 8.17 for the three data-taking years. The outer sidebands invariant mass distribution is shown under the $K^-\pi^+\pi^+$ (top) and 3μ (bottom) mass hypotheses before and after applying the veto on ProbNNk. No significant effect is observed for 2017 and 2018 suggesting that the BDTGPID classifier is able to reduce the $D^+ \to K^-\pi^+\pi^+$ contribution to a negligible contribution. Different is the case for 2016, where the veto on ProbNNk removes a peak in correspondence of the D^+ mass in the distribution obtained under the $K^-\pi^+\pi^+$ mass hypothesis. The different behaviour among the three years is a consequence of the performance of the BDTGPID classifier which, in 2016, is worsened due to the different trigger selection as discussed in Section 7.3.2. The requirement on ProbNNk is therefore applied only on 2016 data.

8.2.5 Final fit to the data sidebands in the 2μ sample

The invariant mass distribution in the 2μ sample is reported in Figure 8.18 for events satisfying the optimal requirements on BDTGPID and BDTGKT (see Table 7.9) and, for 2016 data, the veto on ProbNNk. An extended maximum likelihood fit is performed using an exponential distribution (see Eq. 6.3) assuming the contribution of the combinatorial background only; the result of the fit is superimposed. The fit parameters are listed in Table 8.13 together with the number of background events expected in the signal region.

Especially in 2016 and 2017, a clear trend is visible in the right sideband suggesting the presence of the $D^+ \to \pi^-\pi^+\pi^+$ and $D_s^+ \to \pi^-\pi^+\pi^+$ contributions. As in the 3μ sample, the two charmed modes are included in the final fit in order to have a reliable estimate of the expected number of background events in the signal region. No significant contribution is expected in the signal region from the $D^+ \to \pi^-\pi^+\pi^+$

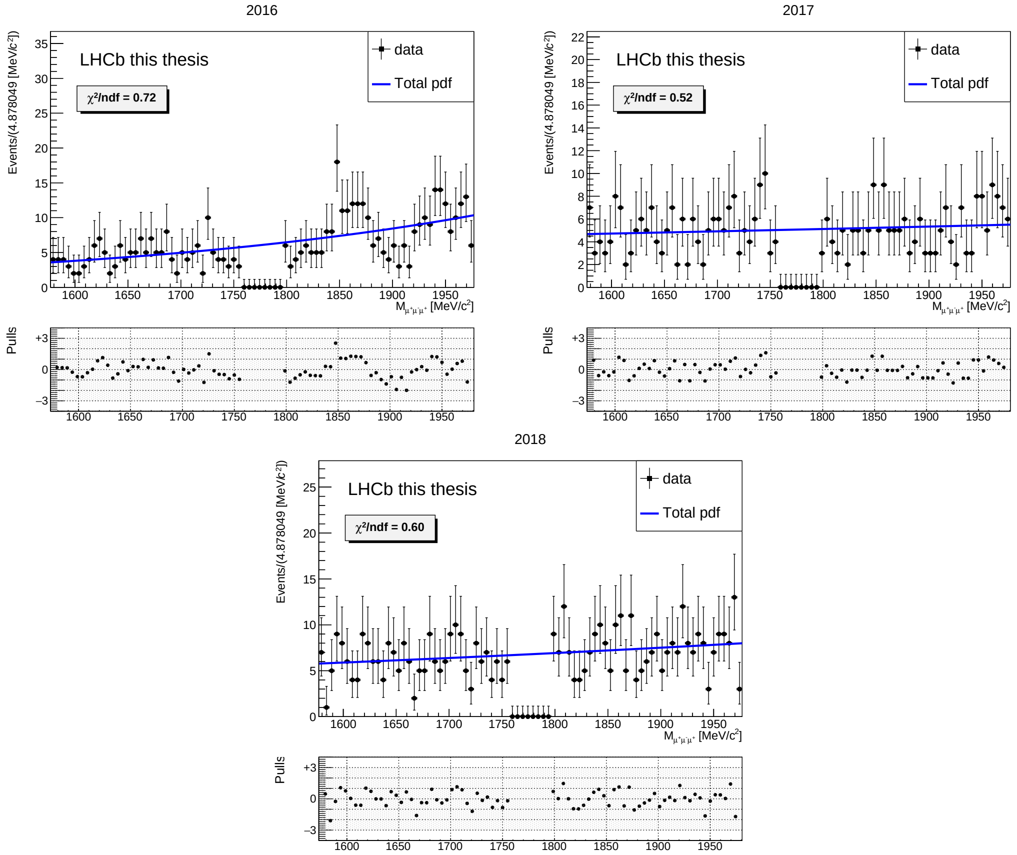


FIGURE 8.18: Data sidebands invariant mass distribution in the 2μ sample after applying the optimal requirements on BDTGKT and BDTGPID and, only for 2016, the veto on ProbNNk. The result of the fit assuming the contribution of the combinatorial background only is superimposed.

TABLE 8.13: List of the fit parameters used to describe the mass pdf in the 2μ sample assuming the combinatorial-only hypothesis for the three data-taking years. The fit is performed on the data sidebands after the optimal requirements on BDTGKT and BDTGPID. In the 2016 data sample the veto on ProbNNk to the *negative* track is also applied.

Parameter	**2016**	**2017**	**2018**
$n_{comb.bkg}$	522.72 ± 23.92	416.58 ± 21.41	559.61 ± 24.82
$n^{sig}_{comb.bkg}$	49.92 ± 2.38	41.61 ± 2.14	55.72 ± 2.48
α	0.0027 ± 0.0004	0.0004 ± 0.0004	0.0008 ± 0.0004

and $D_s^+ \to \pi^-\pi^+\pi^+$ modes, though the presence of the two peaks in the right sideband causes a bias in the determination of the slope of the exponential distribution used to describe the combinatorial background.

The shape parameters of the $D^+ \to \pi^-\pi^+\pi^+$ and $D_s^+ \to \pi^-\pi^+\pi^+$ modes are fixed from a fit to simulated events. The invariant mass distribution of the $D_s^+ \to \pi^-\pi^+\pi^+$ mode is reported in Figure 8.19 under the 3μ mass hypothesis for events surviving the selection outlined in Chapter 5; only two of the three final-state tracks are required to satisfy the `isMuon` condition. The number of simulated events surviving the optimal BDTGPID cuts reported in Table 7.9 would not be sufficient

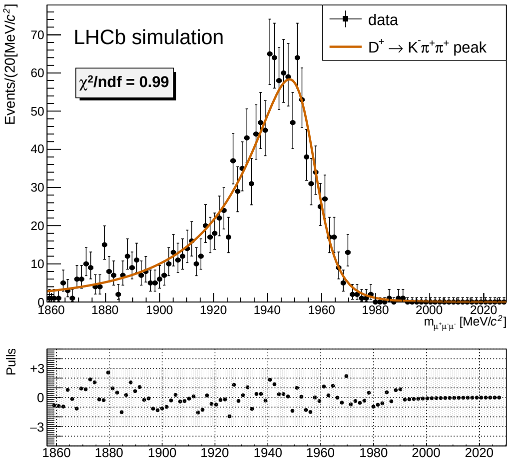


FIGURE 8.19: Invariant mass distribution of the $D_s^+ \to \pi^-\pi^+\pi^+$ mode under the 3μ mass hypothesis for simulated events where only two of the three final-state tracks are required to satisfy the `isMuon` condition. The selection outlined in Chapter 5 for the 2μ sample is applied, as well as a loose cut on BDTGPID, namely BDTGPID>0. The result of the fit is superimposed.

TABLE 8.14: List of the fit parameters used to describe the $D_s^+ \to \pi^-\pi^+\pi^+$ mass peak under the $\mu^-\mu^+\mu^+$ mass hypothesis.

Parameter	**Fit value**
μ_{D_s} [MeV/c^2]	1956.52 ± 1.54
λ_{D_s} [MeV/c^2]	11.99 ± 1.16
γ_{D_s}	1.23 ± 0.12
δ_{D_s}	1.01 ± 0.09

to provide a reliable fit result. For this reason, the requirement on BDTGPID is loosened and the fit is performed on events with BDTGPID>0. The $D_s^+ \to \pi^-\pi^+\pi^+$ mass shape is not expected to vary depending on the requirement on BDTGPID since, as shown in Figure 8.20, no correlation is observed between the mass and the output of the classifier. It should be noted that, due to the limited statistics, the simulated samples generated for the $D_s^+ \to \pi^-\pi^+\pi^+$ mode for the three data-taking years are merged and a unique fit is performed to determine the mass shape. The signal peak is described with a Johnson's S_U function (see Eq. 6.1). The result of the fit is reported in Figure 8.19, while the list of the fit parameters is given in Table 8.14.

Because of the limited statistics available in the $D^+ \to \pi^-\pi^+\pi^+$ simulated sample, its tail parameters are taken from the fit to $D_s^+ \to \pi^-\pi^+\pi^+$ simulated events. The final fit to the invariant mass distribution in the 2μ sample is shown in Figure 8.21 for the three years assuming the contribution of the combinatorial background and of the $D^+ \to \pi^-\pi^+\pi^+$ and $D_s^+ \to \pi^-\pi^+\pi^+$ modes; the result of the fit is superimposed. The yield of the $D^+ \to \pi^-\pi^+\pi^+$ mode is constrained to be a factor of 0.56 smaller than the $D_s^+ \to \pi^-\pi^+\pi^+$ yield, as discussed in Section 8.2.3.

The list of the fit parameters are reported in Table 8.15 for the three years, together with the number of events expected in the signal region from the combinatorial background (n^{sig}_{comb}), from the $D^+ \to \pi^-\pi^+\pi^+$ mode ($n^{sig}_{D^+\to\pi^-\pi^+\pi^+}$) and from the $D_s^+ \to \pi^-\pi^+\pi^+$ mode ($n^{sig}_{D_s^+\to\pi^-\pi^+\pi^+}$). For the three years, the contribution of the

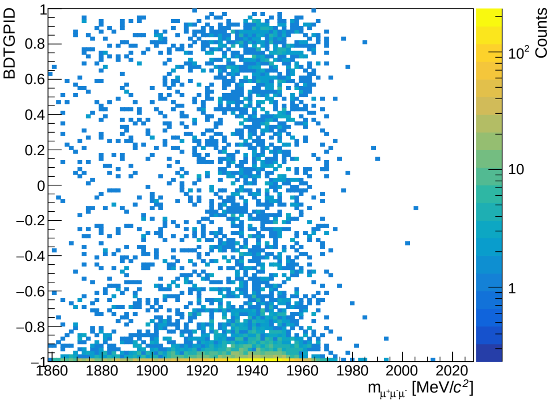


FIGURE 8.20: Invariant mass distribution of the $D_s^+ \to \pi^-\pi^+\pi^+$ modes under the 3μ mass hypothesis versus BDTGPID for simulated events.

charmed modes in the signal region is compatible with zero or with the uncertainty associated to the total number of background events in the signal region (n_{tot}^{sig}). Nevertheless, the inclusion of the $D^+ \to \pi^-\pi^+\pi^+$ and $D_s^+ \to \pi^-\pi^+\pi^+$ modes in the final fit determines a variation of the slope of the exponential distribution used to describe the combinatorial background, as well as the decrease of n_{tot}^{sig} with respect to the values reported in Table 8.13.

As already mentioned in Section 8.1.1, the number of $D_s^+ \to \eta(\mu^+\mu^-\gamma)\mu^+\nu_\mu$ events expected in the whole mass spectrum in the 2μ sample ($N_{D_s,Bkg}$ in Table 8.4) is negligible compared to the number of background events expected from the combinatorial background (n_{comb} in Table 8.15). This is true for all data-taking years.

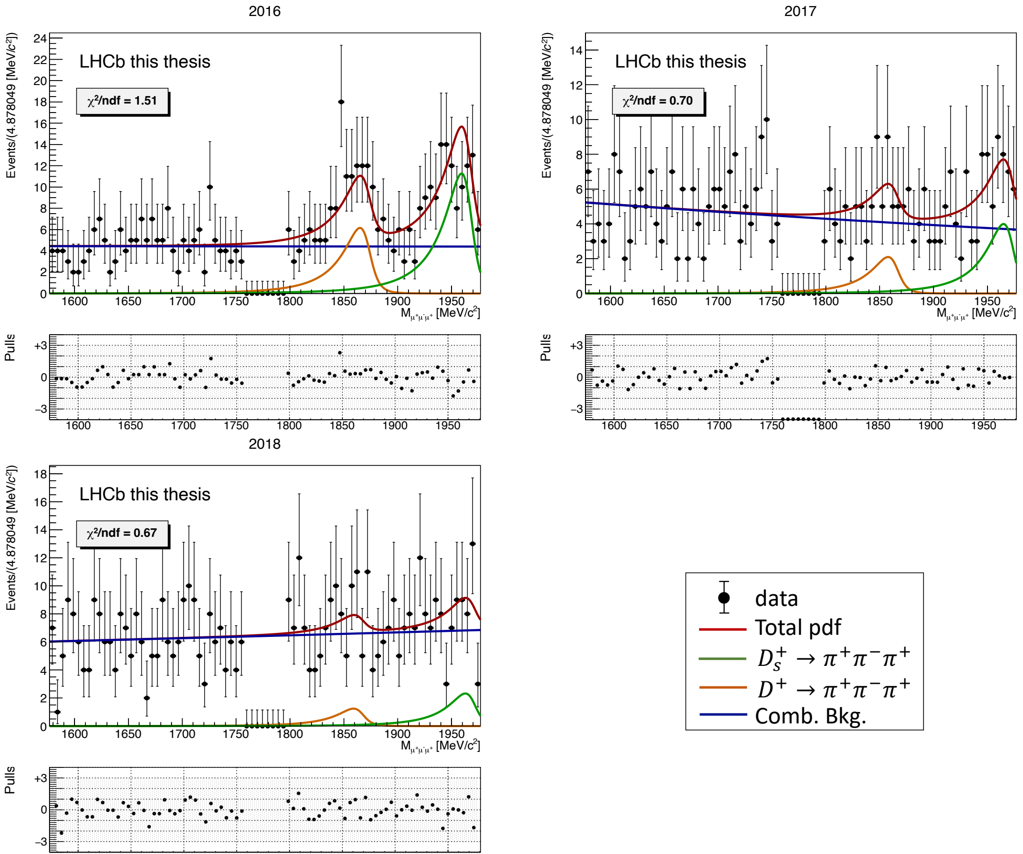


FIGURE 8.21: Final fit to the outer sidebands invariant mass distribution in the 2μ sample assuming the contribution of the combinatorial background and of the $D^+ \to \pi^-\pi^+\pi^+$ and $D_s^+ \to \pi^-\pi^+\pi^+$ modes. The mass shapes of the charmed modes are fixed from a fit to simulated events. The fit is shown for the three data-taking years.

TABLE 8.15: List of the fit parameters used to describe the mass pdf in the 2μ sample assuming the contribution of the combinatorial background and of the $D^+ \to \pi^-\pi^+\pi^+$ and $D_s^+ \to \pi^-\pi^+\pi^+$ modes for the three data-taking years. The fit is performed on events satisfying the optimal requirements on BTDGKT and BDTGPID and, in 2016, the veto on ProbNNk. The parameters λ_{D_s}, γ_{D_s} and δ_{D_s} are shared between the two charmed modes and fixed from the fit to simulated events. The yield of the $D^+ \to \pi^-\pi^+\pi^+$ is fixed at $0.56 \times n_{D_s^+ \to \pi^-\pi^+\pi^+}$, according to Eq. 8.3.

Parameter	**2016**	**2017**	**2018**
μ_{D_s} [MeV/c^2]	1967.72 ± 3.13	1973.40 ± 5.89	1971.20 ± 7.04
λ_{D_s} [MeV/c^2]		11.99	
γ_{D_s}		1.23	
δ_{D_s}		1.01	
$n_{D_s^+ \to \pi^-\pi^+\pi^+}$	96.07 ± 20.34	$32.69.33 \pm 12.96$	19.41 ± 16.25
$n^{sig}_{D_s^+ \to \pi^-\pi^+\pi^+}$	0.65 ± 0.88	0.21 ± 0.08	0.13 ± 0.11
μ_{D^+} [MeV/c^2]	1873.39 ± 3.59	1866.25 ± 9.11	1867.40 ± 11.89
$n_{D^+ \to \pi^-\pi^+\pi^+}$	53.80	18.31	10.87
$n^{sig}_{D^+ \to \pi^-\pi^+\pi^+}$	2.72 ± 0.42	1.13 ± 0.55	0.65 ± 0.61
α_{comb}	-0.0002 ± 0.0008	-0.0009 ± 0.0007	0.0003 ± 0.0006
n_{comb}	362.76 ± 35.51	361.35 ± 28.91	527.22 ± 36.46
n^{sig}_{comb}	36.28 ± 3.54	35.95 ± 2.90	52.67 ± 3.66
n^{sig}_{tot}	39.65 ± 2.61	37.29 ± 2.63	53.46 ± 3.04

Chapter 9

Normalisation

As described in Section 3.2, the branching fraction (BF) of $\tau^+ \to \mu^+\mu^-\mu^+$ can be expressed as

$$\mathcal{B}(\tau^+ \to \mu^+\mu^-\mu^+) = \alpha \times N(\tau^+ \to \mu^+\mu^-\mu^+)^{meas}, \tag{9.1}$$

where

$$\alpha = \frac{\mathcal{B}(D_s^+ \to \phi(\mu^+\mu^-)\pi^+)}{\mathcal{B}(D_s^+ \to \tau^+\nu_\tau)} \times f_{D_s}^{\tau} \times \frac{\epsilon_{D_s}}{\epsilon_\tau} \times \frac{1}{N(D_s^+ \to \phi(\mu^+\mu^-)\pi^+)^{meas}}. \tag{9.2}$$

The various terms entering the definition of the normalisation factor α, also known as the *single event sensitivity*, are discussed separately in the following.

The content of this Chapter has been carried out by the author, with the exception of the efficiency corrections discussed in Section 9.5.2 and 9.5.3 which have been evaluated by a former PhD student of the analysis group.

9.1 Fraction of τ produced via D_s decays

The fraction of τ produced via D_s decays is easily obtained as the ratio between the cross section of τ production via D_s decays and the total τ production cross section at 13 TeV.

$$f_{D_s}^{\tau} = \frac{\sigma(D_s^+ \to \tau^+) + \sigma(\bar{b} \to D_s^+ \to \tau^+)}{\sigma(pp \to \tau^+ X)} \tag{9.3}$$

The D_s meson can be either produced *prompt*, namely at the collision point, or *secondary*, via b decays. The cross section of τ production via D_s decays can therefore be expressed as the sum of the two contributions. The evaluation of the cross sections of τ production in the full phase space (4π) is reported in Section 4.2.2 (see Column 6 in Table 4.5) and the resulting value for $f_{D_s}^{\tau}$ is $(78.4 \pm 3.7)\%$.

9.2 Input branching fractions

The BF of the $D_s^+ \to \tau^+\nu_\tau$ decay, also reported in Section 4.2.2, is $(5.32 \pm 0.11)\%$. The BF of the $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ decay can be computed using the equation

$$\mathcal{B}(D_s^+ \to \phi(\mu^+\mu^-)\pi^+) = \frac{\mathcal{B}(D_s^+ \to \phi(K^+K^-)\pi^+)}{\mathcal{B}(\phi \to K^+K^-)} \mathcal{B}(\phi \to \mu^+\mu^-), \tag{9.4}$$

where $\mathcal{B}(D_s^+ \to \phi(K^+K^-)\pi^+)$ is obtained from the Dalitz-plot analysis of the $D_s^+ \to K^+K^-\pi^+$ decay [8]. The non-resonant contribution to the BF of the $D_s^+ \to \mu^+\mu^-\pi^+$ decay is assumed to be negligible given the exclusion limits on this contribution

TABLE 9.1: BFs used in the evaluation of $\mathcal{B}(D_s^+ \to \phi(\mu^+\mu^-)\pi^+)$. The values are taken from the PDG [8].

Channel	$\mathcal{B}$
$D_s^+ \to \phi(\to K^+K^-)\pi^+$	$(2.22 \pm 0.06) \times 10^{-2}$
$\phi \to K^+K^-$	$(49.1 \pm 0.5) \times 10^{-2}$
$\phi \to \mu^+\mu^-$	$(2.85 \pm 0.19) \times 10^{-4}$

reported in Ref. [93]. The BFs used in Eq. 9.4 are listed in Table 9.1 and lead to a final BF of $(1.289 \pm 0.094) \times 10^{-5}$ for the $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ decay.

9.3 Number of events observed in the $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ channel

The number of events observed in the reference channel, $N(D_s^+ \to \phi(\mu^+\mu^-)\pi^+)^{meas}$, is extracted from the fit to the invariant mass distribution in data as the integral of the pdf used to model the signal peak. The fit results are shown in Figure 6.8, while in Table 9.2 the values of $N(D_s^+ \to \phi(\mu^+\mu^-)\pi^+)^{meas}$ are reported for the three data-taking years.

TABLE 9.2: Number of events observed in the reference channel $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ for each data-taking year.

Yield	**2016**	**2017**	**2018**
$N(D_s^+ \to \phi(\mu^+\mu^-)\pi^+)^{meas}$	43072 ± 397	88474 ± 594	103790 ± 643

9.4 Efficiencies

The factor $\epsilon_{D_s}/\epsilon_\tau$ represents the ratio between the overall efficiency evaluated on the reference channel and the one evaluated on the signal channel. Both efficiencies are evaluated on simulated events and can be factorised into stages, each conditional on the immediately prior stage.

$$\frac{\epsilon_{D_s}}{\epsilon_\tau} = \frac{\epsilon_{D_s}^{ACC} \times \epsilon_{D_s}^{FILT} \times \epsilon_{D_s}^{REC} \times \epsilon_{D_s}^{SEL} \times \epsilon_{D_s}^{TRIG}}{\epsilon_\tau^{ACC} \times \epsilon_\tau^{FILT} \times \epsilon_\tau^{REC} \times \epsilon_\tau^{SEL} \times \epsilon_\tau^{TRIG} \times \epsilon_\tau^{clfs}} \tag{9.5}$$

For both channels the overall efficiency includes the detector acceptance ϵ^{ACC}, the efficiency of the cuts applied at the filtering level ϵ^{FILT}, the combined reconstruction and stripping selection efficiency ϵ^{REC}, the offline selection efficiency ϵ^{SEL} and the trigger selection efficiency ϵ^{TRIG}. For the signal channel, the `isMuon` selection efficiency is included in the offline selection efficiency. Moreover, both in the 2μ and in the 3μ sample a selection based on the output of the two multivariate classifiers (*clfs*) is also applied, therefore an additional term ϵ^{clfs} is included in the determination of the overall efficiency. In the next sections, these factors contributing to the overall efficiency are discussed separately.

9.4.1 Acceptance, reconstruction and stripping selection efficiencies

In order to save disk space and CPU time, only simulated events generating final-state tracks within the geometrical acceptance of LHCb are stored. The efficiency ϵ^{ACC} simply represents the fraction of generated events falling within the detector acceptance. For the simulated sample, it is computed as average of the efficiencies of the single sub-channel ϵ_{CUT} (see Table 4.8 and 4.9) weighted with the production fraction (w_{fraction}) of the relative sub-channel.

$$\epsilon^{ACC} = \frac{\sum_{i=1}^{N_{Ch}} w_{\text{fraction},i} \cdot \epsilon_{CUT,i}}{\sum_{i=1}^{N_{Ch}} w_{\text{fraction},i}} \tag{9.6}$$

The resulting efficiencies, reported in Table 9.3, are comparable for the signal and for the reference channel due to the similar topology.

TABLE 9.3: Geometrical acceptance for the mixed MC samples.

Channel	ϵ^{ACC} **[%]**
$\tau^+ \to \mu^+\mu^-\mu^+$	10.72 ± 0.43
$D_s^+ \to \phi(\mu^+\mu^-)\pi^+$	12.04 ± 0.33

The error on the geometrical acceptance is evaluated using *toys*. As no differences are expected between the different years, this efficiency is evaluated once for 2018 and used also for 2016 and 2017.

The generated events are filtered through some loose requirements on particle identification information and on the invariant mass range. The corresponding efficiency ϵ^{FILT} is reported in Table 9.4 for each data-taking year.

TABLE 9.4: Efficiencies of the cuts applied at the filtering level for the three data-taking years.

Channel	**2016 [%]**	**2017 [%]**	**2018 [%]**
$\tau^+ \to \mu^+\mu^-\mu^+$	57.19 ± 0.04	57.11 ± 0.04	57.05 ± 0.04
$D_s^+ \to \phi(\mu^+\mu^-)\pi^+$	47.53 ± 0.03	47.62 ± 0.03	47.50 ± 0.03

The reconstruction efficiency is evaluated as ratio between the number of reconstructed and the number of filtered events. These values are given in Table 9.5.

TABLE 9.5: Reconstruction and stripping selection efficiencies for the three data-taking years.

Channel	**2016 [%]**	**2017 [%]**	**2018 [%]**
$\tau^+ \to \mu^+\mu^-\mu^+$	39.73 ± 0.05	39.65 ± 0.05	39.77 ± 0.05
$D_s^+ \to \phi(\mu^+\mu^-)\pi^+$	16.94 ± 0.04	17.05 ± 0.03	17.09 ± 0.03

9.4.2 Offline selection efficiency

The efficiency of the offline selection, whose requirements are described in Section 5.1, is evaluated on truth-matched events and is reported in Table 9.6 for the two channels and for the three years of data-taking.

TABLE 9.6: Offline selection efficiencies for the three data-taking years.

Channel	2016 [%]	2017 [%]	2018 [%]
$\tau^+ \to \mu^+\mu^-\mu^+$	75.84 ± 0.07	75.91 ± 0.07	75.80 ± 0.07
$D_s^+ \to \phi(\mu^+\mu^-)\pi^+$	91.12 ± 0.07	91.18 ± 0.07	91.11 ± 0.07

The $\tau^+ \to \mu^+\mu^-\mu^+$ data set is split into the 2μ and the 3μ samples. The efficiencies of the cut on `isMuon` are reported in Table 9.7 for the two samples and for the three data-taking years.

TABLE 9.7: `isMuon` selection efficiencies for the three data-taking years.

Channel	2016 [%]	2017 [%]	2018 [%]
$\tau^+ \to \mu^+\mu^-\mu^+$ (2μ sample)	29.32 ± 0.09	29.31 ± 0.09	29.23 ± 0.09
$\tau^+ \to \mu^+\mu^-\mu^+$ (3μ sample)	70.68 ± 0.09	70.69 ± 0.09	70.72 ± 0.09

9.4.3 Trigger selection efficiency

The trigger selection efficiency ϵ^{TRIG} is evaluated on truth-matched events surviving the offline selection. The trigger efficiencies are reported in Table 9.8 separately for the reference channel and for the 2μ and the 3μ samples, for the three data-taking years and for the different stages of the trigger selection. Note that only the L0&HLT1&HLT2 efficiency enters the evaluation of the efficiency ratio $\epsilon_{D_s}/\epsilon_{\tau}$.

It can be observed that, for the 3μ sample, the main limitation to the trigger efficiency is due to the hardware trigger L0, as both the HLT1 and HLT2 selections are extremely efficient ($> 99\%$ relative efficiency). On the other hand, both the reference channel and the 2μ sample show lower L0 selection efficiency ($\sim 45\%$); also the HLT2 selection efficiency, $\sim 50\%$ in the 2μ sample and $\sim 65\%$ in the reference channel, is worsened in 2016 due to the replacement of the `Hlt2DiMuonDetached_TOS` line with the `Hlt2TopoMuMu3Body_TOS` line (see Table 5.3). In Run 3, the removal of the hardware trigger L0 [94] could lead to a significant improvement of the selection efficiency. Moreover, the development of new trigger lines *ad hoc* for the selection of events with only two tracks identified as muons could enhance the HLT2 selection efficiency in the 2μ sample.

The trigger selection represents the last stage of the selection applied to the reference channel $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$. All the efficiencies shown so far have been evaluated on simulated events weighted with w_{fraction}, namely for the correct τ production fraction (see Section 4.2.2).

TABLE 9.8: Trigger efficiencies calculated on simulated events surviving the truth-matching and offline selection for each year of data taking and for each trigger stage. All values are in percent.

Year	Trigger Level	3μ sample	2μ sample	$D_s^+ \to \phi(\mu^+\mu^-)\pi^+$
2016	L0	60.22 ± 0.12	43.26 ± 0.13	46.62 ± 0.13
	L0&HLT1	58.48 ± 0.12	37.78 ± 0.13	42.30 ± 0.13
	L0&HLT1&HLT2	58.38 ± 0.12	8.17 ± 0.10	14.79 ± 0.09
2017	L0	69.20 ± 0.11	51.91 ± 0.19	55.24 ± 0.12
	L0&Hlt1	66.78 ± 0.11	44.96 ± 0.18	49.58 ± 0.12
	L0&Hlt1&Hlt2	66.77 ± 0.11	22.35 ± 0.15	31.76 ± 0.11
2018	L0	59.47 ± 0.12	42.97 ± 0.18	46.65 ± 0.13
	L0&Hlt1	57.67 ± 0.12	37.61 ± 0.18	42.09 ± 0.13
	L0&Hlt1&Hlt2	57.65 ± 0.12	19.05 ± 0.15	27.59 ± 0.12

9.4.4 Efficiency of the selection on the classifiers output

The requirement on ProbNNk applied to the *negative* track in the 3μ sample to reduce the contribution of the $D^+ \to K^-\pi^+\pi^+$ channel (see Section 8.2.1) has almost no effect on signal events. Its efficiency is reported in Table 9.9 for the three years.

TABLE 9.9: Efficiency of the veto requirement applied on the ProbNNk variable of the *negative* track in the 3μ sample for the three data-taking years.

Channel	2016 [%]	2017 [%]	2018 [%]
$\tau^+ \to \mu^+\mu^-\mu^+$ (3μ sample)	99.65 ± 0.30	99.56 ± 0.28	99.59 ± 0.30

In the 2μ sample there is a small fraction of events where the track not satisfying the `isMuon` requirement falls outside the RICH detectors acceptance. For these tracks no RichDLLmu variable is available, therefore no PID information can be exploited in the suppression of the background from mis-ID tracks. These events are excluded from the analysis. The remaining fraction of events, evaluated on events fulfilling the trigger requirements, is reported in Table 9.10 for the three years.

TABLE 9.10: Efficiency of the selection of events with RichDLLmu in the 2μ sample for the three data-taking years.

Channel	2016 [%]	2017 [%]	2018 [%]
$\tau^+ \to \mu^+\mu^-\mu^+$ (2μ sample)	97.25 ± 0.21	97.99 ± 0.11	98.11 ± 0.12

In the 2μ sample a cut is applied on the output of the two classifiers trained to suppress the combinatorial background (BDTGKT) and the background from mis-ID tracks (BDTGPID) (see Section 7.3.3). The efficiency of the combined selection on the output of the two classifiers (ϵ^{clfs}) is reported in Table 9.11, together with the combined efficiency of the minimum cuts applied on the output of the classifiers trained in the 3μ sample. As discussed in Section 8.2.4, the veto on ProbNNk applied in the 3μ sample is applied also in the 2μ sample on 2016 data only. The corresponding efficiency, $(99.98 \pm 0.05)\%$, is included in ϵ^{clfs}.

TABLE 9.11: Efficiency of the selection applied on the BDTGKT and BDTGPID output in the 2μ sample and of the minimum cuts applied on the XGB and SuperPNN output in the 3μ sample for the three data-taking years. The value reported for the 2016 2μ sample also includes the efficiency of the veto on ProbNNk.

Channel	**2016 [%]**	**2017 [%]**	**2018 [%]**
$\tau^+ \to \mu^+\mu^-\mu^+$ (2μ sample)	10.50 ± 0.41	13.76 ± 0.28	15.12 ± 0.31
$\tau^+ \to \mu^+\mu^-\mu^+$ (3μ sample)	61.68 ± 0.14	62.90 ± 0.15	62.96 ± 0.16

The efficiencies of the selection applied on the output of the classifiers, as well as the efficiency of the veto requirement applied on the ProbNNk variable, are evaluated on simulated events weighted with w_{fraction}, $w_{\text{Data/MC}}$, to minimise the differences between data and simulation in the kinematic variables, w_{fracCorr}, namely the correction to w_{fraction} (see Appendix A), the correction to the PID variables and the weights evaluated to minimise the differences between data and simulation in the output of the two classifiers. Apart from w_{fraction}, which is evaluated on generated events, all the aforementioned corrections are evaluated on events that pass the trigger selection and are therefore taken into account in the evaluation of the efficiency of the selection stages that follow the trigger selection. The uncertainties on the selection efficiencies shown so far are purely statistical.

9.5 Efficiency corrections

In addition to the corrections mentioned so far, there is a set of corrections which can be directly evaluated on the efficiency ratio $\epsilon_{D_s}/\epsilon_\tau$. These include the correction to the track reconstruction efficiency, the correction to the trigger efficiency, evaluated using the so-called TISTOS method (see Section 3.3), and the correction to the efficiency of the cut applied on $m_{\mu^+\mu^-}$ in the reference channel to select the $\phi(1020)$ resonance.

9.5.1 Correction to the track reconstruction efficiency

This Section is dedicated to the evaluation of the correction to the track reconstruction efficiency of the $\tau^+ \to \mu^+\mu^-\mu^+$ and $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ channels. The strategy used to evaluate the track reconstruction efficiency at LHCb, developed within the collaboration, represented for the author the starting point for the development of the trigger lines adopted in the evaluation of the track reconstruction efficiency for Run 3. For this reason, a detailed description of the method is given in this Section.

The track reconstruction efficiency at LHCb is evaluated on the $J/\psi \to \mu^+\mu^-$ channel with a tag-and-probe approach [95]. Among the two J/ψ decay products one, the "tag", is fully reconstructed as a long track, while the other one, the "probe", is only partially reconstructed. The track reconstruction efficiency is evaluated as the fraction of probe tracks that match to a long track.

The advantage of using the $J/\psi \to \mu^+\mu^-$ channel is that, for both final-state particles, the information provided by the muon detector can be exploited in the reconstruction of the probe track. The track reconstruction efficiency is, however, evaluated using three separate methods depending on which sub-detector information is used in the reconstruction of the probe track, namely the VELO method, the T station method and the long method.

In the VELO method the probe track is reconstructed as a downstream track, namely using information from the TT and T stations (see Section 2.2.5). If there is a long track with 50% common hits in the T stations, then the probe track is defined as efficient. The T station method uses as probes tracks reconstructed combining information from the VELO and from the muon stations. The probe track is considered to be matched to a long track if the two tracks share the same VELO seed and at least two hits in the muon stations. The long method uses tracks with hits in the TT and in the muon stations as probe tracks. The reconstruction of long tracks does not require the presence of hits in the TT stations, meaning that a track can be reconstructed as a long track even if no hits are found in the TT stations. If the probe track has hits both in the TT and in the muon stations, the matching to a long track is efficient only if they share at least 60% of the hits in the TT stations and more than 70% of the hits in the muon stations. If no hits are detected in the TT stations only the second requirement is needed. The three methods are illustrated in Figure 9.1.

An unbinned extended maximum likelihood fit is performed to the invariant mass of the tag and probe tracks to evaluate the number of reconstructed J/ψ decays and the track reconstruction efficiency is computed as the fraction of reconstructed J/ψ where the probe track can be matched to a long track. The efficiency of the VELO and T station method are assumed to be uncorrelated and can therefore be multiplied to obtain the efficiency for finding long tracks. The combined method and the long method provide two independent, even if slightly different, ways to probe the long track reconstruction efficiency.

To account for data/MC differences in the track reconstruction efficiency, the ratio between the efficiency measured in data ϵ_{Data} and the one evaluate on simulated events ϵ_{MC} is computed. Since the track reconstruction efficiency clearly depends on the track kinematics, the efficiency ratio is evaluated in bins of the track momentum and pseudo-rapidity. The correction to the track reconstruction efficiency used in this analysis is the one provided by the long method. The corresponding correction maps containing the efficiency ratio $\epsilon_{\text{Data}}/\epsilon_{\text{MC}}$ produced by the tracking group with the `TrackCalib` tool [96] are displayed in Figure 9.2 for the three data-taking years.

As described in Ref. [97], the correction to the track reconstruction efficiency for a given channel is computed as

$$C^{ch} = \prod_{tr} C^{ch}_{tr} = \prod_{tr} \left\{ \sum_i (f_i^{\text{MC},tr} C_i) \beta_{RW} \beta^{tr}_{mat} \right\}, \tag{9.7}$$

where the product runs over the tracks in the final state. The factor $f_i^{\text{MC},tr}$ is the fraction of reconstructed and selected events in bin i of the distribution obtained for the track tr, defined as

$$f_i^{\text{MC}} = \frac{N_i^{Rec,Sel}}{N^{Rec,Sel}}, \tag{9.8}$$

while C_i is the content of the correction maps in bin i. The factor β_{RW} represents a *per-track* uncertainty that comes out from different data/MC reweighting techniques, and its values for the different data-taking years are listed in the following.

$$\beta_{RW} = 1.000 \pm 0.004 \quad \text{(for Run 1)} \tag{9.9a}$$

$$\beta_{RW} = 1.000 \pm 0.008 \quad \text{(for 2015/2016)} \tag{9.9b}$$

$$\beta_{RW} = 1.000 \pm 0.004 \quad \text{(for 2017/2018)} \tag{9.9c}$$

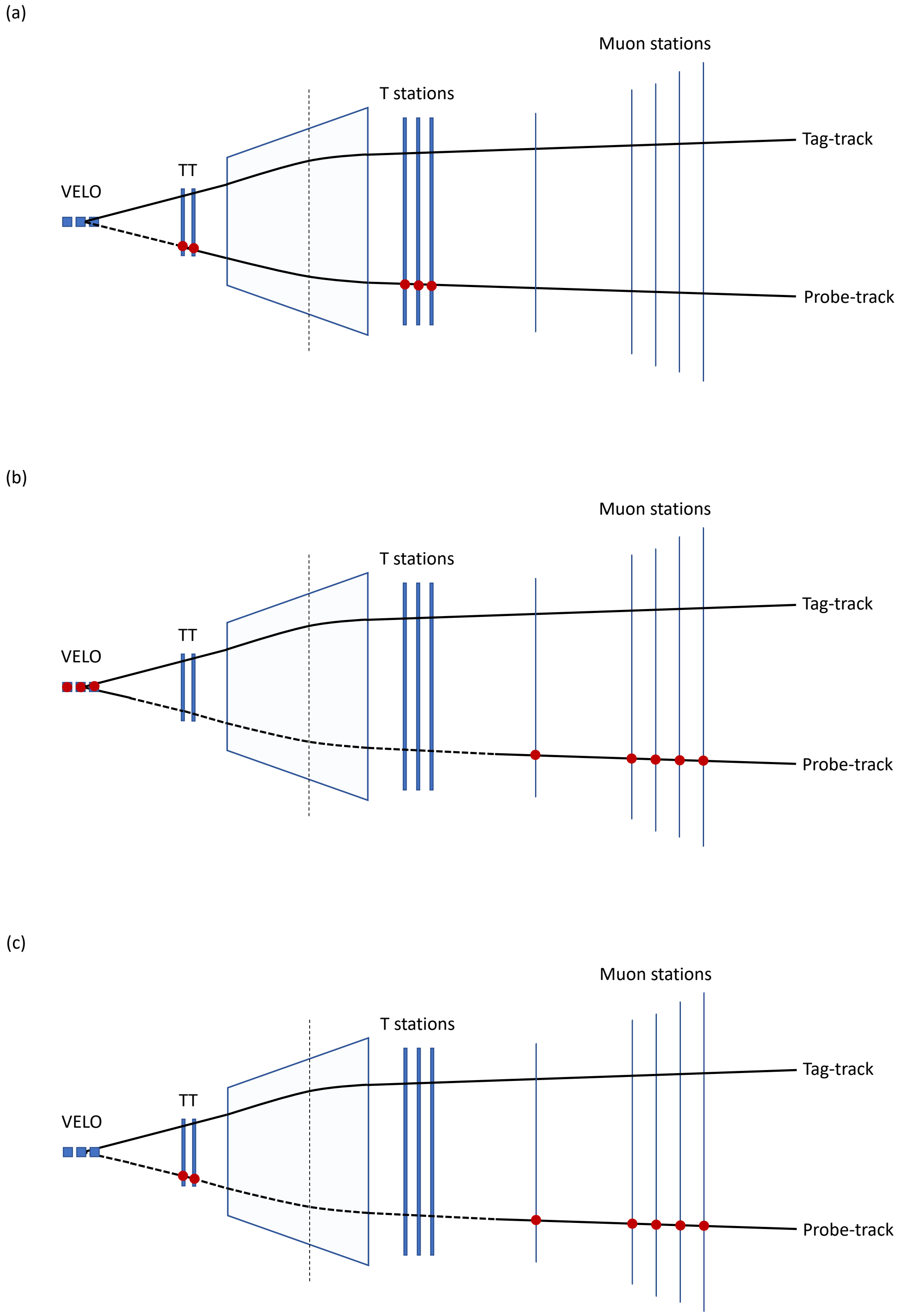


FIGURE 9.1: Illustration of the three tag-and-probe methods used to evaluate the track reconstruction efficiency at LHCb: the VELO method (a), the T station method (b) and the long method (c). The red dots are placed at the sub-detectors whose information is used in the reconstruction of the probe track.

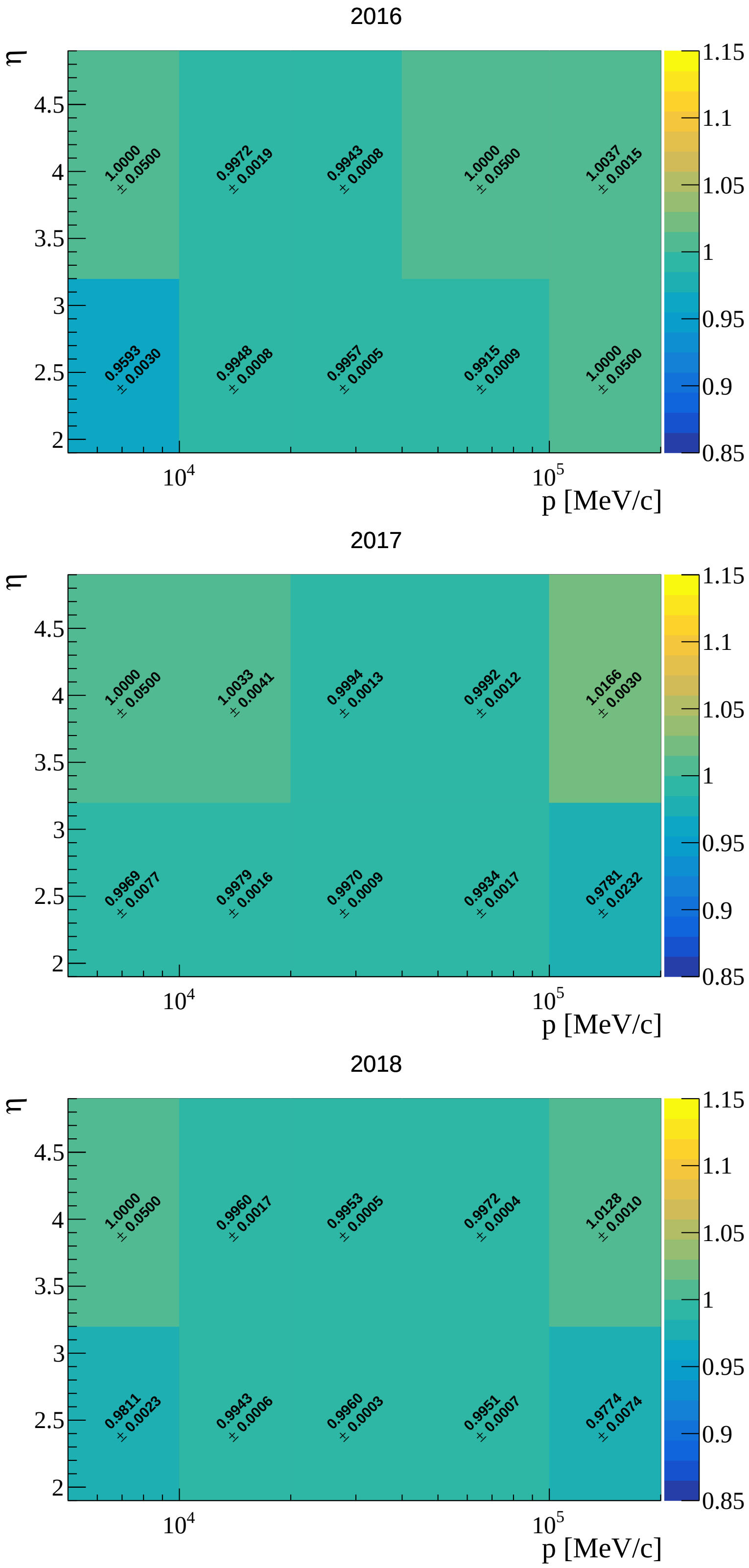


FIGURE 9.2: Maps containing the correction to the track reconstruction efficiency evaluated using the long method in bins of track momentum and pseudo-rapidity. The maps are reported for the three data-taking years.

TABLE 9.12: Correction to the track reconstruction efficiency for all tracks in the signal and in the reference channel for the different data-taking years.

	2016	**2017**	**2018**
$c_{Track,\tau}$	0.958 ± 0.017	0.997 ± 0.011	0.976 ± 0.010
c_{Track,D_s}	0.962 ± 0.022	0.997 ± 0.017	0.978 ± 0.017

TABLE 9.13: Ratio between the correction to the track reconstruction efficiency evaluated on the reference channel and the one evaluated on the signal channel for the different data-taking years. The former uncertainty is the statistical one, while the latter is the systematic uncertainty.

	2016	**2017**	**2018**
R_{Track}	$1.004 \pm 0.005 \pm 0.014$	$1.000 \pm 0.005 \pm 0.014$	$1.002 \pm 0.005 \pm 0.014$

The hadronic interaction of kaons and pions with the detector material leads to an additional uncertainty (β^{tr}_{mat}) whose values are:

$$\beta^{K}_{mat} = 1.000 \pm 0.011, \tag{9.10a}$$

$$\beta^{\pi}_{mat} = 1.000 \pm 0.014. \tag{9.10b}$$

In case of muon tracks this uncertainty is set to 1, with no further uncertainty, as no hadronic interaction has to be considered.

As shown in Figure 5.1, most of the signal events are characterised by a momentum smaller than 5 GeV/c while the efficiency maps provided by the tracking group are only given in the ranges 5 GeV/c $< p <$ 201 GeV/c and $1.9 < \eta < 4.9$. In order to reduce the percentage of outliers, the lower limit of the momentum range is extended from 5 to 3 GeV/c and events with a momentum smaller than 5 GeV/c are assigned the correction of the closest bin.

The track reconstruction efficiency correction for the signal and for the reference channel is shown in Table 9.12 for the three data-taking years. The reported uncertainties include the statistical uncertainty, the factor β_{RW} and the hadron uncertainty.

The ratio between the correction to the track reconstruction efficiency evaluated on the reference channel and the one evaluated on the signal channel is computed and shown in Table 9.13 for the three years. The efficiency ratio $\epsilon_{D_s}/\epsilon_{\tau}$ in Eq. 9.2 is then multiplied by this factor to include the correction to the track reconstruction efficiency.

9.5.2 Correction to the $\phi(1020)$ mass cut efficiency

As discussed in Section 5.1, a cut is applied on the invariant mass of the final-state muons with opposite charge to reject events from the reference channel $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$, *i.e.* $|m_{\mu^+\mu^-} - m_{\phi(1020)}| > 20$ MeV/c^2. The opposite cut is instead applied in the reference channel to select the $\phi(1020)$ resonance. The efficiency of this cut is not correctly reproduced in simulation since the $\phi \to \mu^+\mu^-$ resonance is not simulated according to the $\phi \to \mu^+\mu^-$ line shape. The $\phi(1020)$ intermediate state is indeed generated for masses within $\pm 15\Gamma$ (from 955.7 to 1083.2 MeV/c^2), being Γ the decay width of the $\phi(1020)$ resonance [8]. Therefore, a correction factor is evaluated and applied to the $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ selection efficiency.

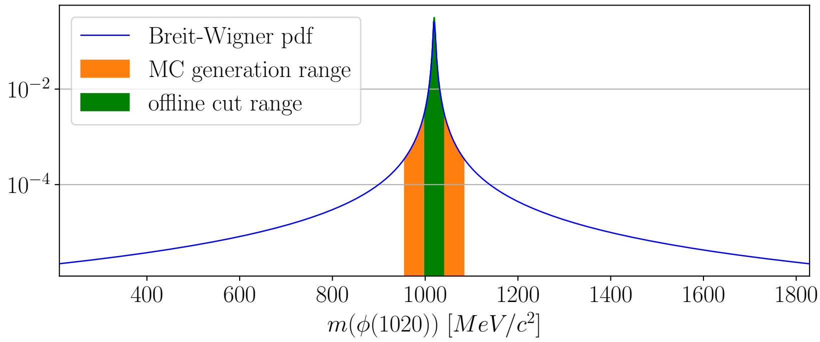


FIGURE 9.3: Non-relativistic Breit-Wigner mass distribution of the $\phi(1020)$ resonance. The green area is the portion selected by the offline cut, while the orange area shows the portion excluded by the cut and used, together with the green one, in the generation of the simulated sample. The distribution is shown in the feasible range in data, namely from $2m_\mu$ up to $m_{D_s^+} - m_\pi$.

As shown in Figure 9.3, a non-relativistic Breit-Wigner (BW) line shape [98], limited by the available phase space in the $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ decay, is used to measure the fraction of events falling outside the $\phi(1020)$ peak. In the truncated MC distribution the excluded area corresponds to the 4.7% of the generated area. For the full BW from $2m_\mu$ up to $m_{D_s} - m_\pi$, the region outside the signal peak is the 6.6% of the total area. To account for the difference in the area excluded from the fit to the invariant mass distribution between data and simulation, the $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ selection efficiency is corrected by a factor of 1.018 ± 0.009, evaluated as $1.066/1.047$. A conservative systematic uncertainty of 50% of its deviation from 1 is considered.

9.5.3 Correction to the trigger efficiency

The trigger efficiency is evaluated on simulated events and as such needs to be corrected to account for possible differences between data and simulation. Such discrepancies are only expected in L0 and HLT1 triggers, while for HLT2, a software trigger running on offline reconstructed variables, very similar performances are expected in data and simulation.

As described in Section 3.3, the TISTOS method [61] allows to evaluate the trigger efficiency by exploiting quantities available in the data sample by assuming that the TOS efficiency can be determined within the TIS sub-sample.

$$\epsilon_{\text{TOS}} = \epsilon_{\text{TOS|TIS}} = \frac{N_{\text{TISTOS}}}{N_{\text{TIS}}} \tag{9.11}$$

The TIS sub-sample is made up of events satisfying the offline selection in Section 5 and firing a combination of L0 and HLT1 TIS trigger lines, mainly based on PID requirements and on loose cuts on kinematics and track quality information [78–80].

```
(L0Hadron|L0Muon|L0Photon|L0Electron)&
    (Hlt1SingleMuonHighPT|Hlt1DiMuonLowMass|Hlt1TwoTrackMVA|
            Hlt1TrackMuon|Hlt1TrackMVA|Hlt1TrackMuonMVA)
```
(9.12)

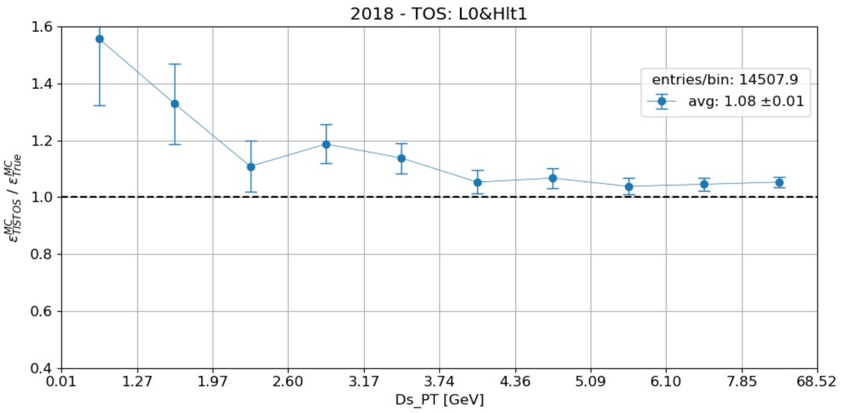


FIGURE 9.4: Ratio between the TISTOS efficiency and true trigger efficiency (N_{TOS}/N_{Sel}) computed on $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ simulated events up to L0&HLT1, in bins of D_s^+ p_T.

The correlation between the signal candidate and the underlying event can be studied on $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ simulated events, where the trigger efficiency can be computed both with the standard definition (N_{TOS}/N_{Sel}) and with the TISTOS method (Eq. 9.11). The ratio between these two efficiencies is displayed in Figure 9.4 in different bins of the D_s transverse momentum. As can be seen, the efficiency ratio shows a dependency on p_T, meaning that the signal candidate and the underlying events are not totally uncorrelated and that a bias is expected in the evaluation of the TISTOS efficiency. As shown in the following, the reference channel $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ can be exploited to evaluate a correction that accounts for this correlation.

The ratio between the trigger efficiency evaluated in the reference channel and in the signal channel on simulated events has to be multiplied by a correction factor, referred to as $R(D_s, \tau)$.

$$\frac{\epsilon^{D_s}_{\text{Data}}}{\epsilon^{\tau}_{\text{Data}}} = \frac{\epsilon^{D_s}_{\text{MC,TOS}}}{\epsilon^{\tau}_{\text{MC,TOS}}} \cdot R(D_s, \tau) \tag{9.13}$$

The double ratio is defined as

$$R(D_s, \tau) = \frac{\epsilon^{D_s}_{\text{Data,TISTOS}}}{\epsilon^{D_s}_{\text{MC,TISTOS}}} \cdot \frac{\epsilon^{\tau}_{\text{MC,TISTOS}}}{\epsilon^{\tau}_{\text{Data,TISTOS}}} = \frac{\epsilon^{D_s}_{\text{Data,TISTOS}}}{\epsilon^{D_s}_{\text{MC,TISTOS}}} \cdot \frac{\sum_i f_i^{\tau} \epsilon^{D_s,i}_{\text{MC,TISTOS}}}{\sum_i f_i^{\tau} \epsilon^{D_s,i}_{\text{Data,TISTOS}}}, \tag{9.14}$$

where $\epsilon^{D_s}_{\text{Data,TISTOS}}$ and $\epsilon^{D_s}_{\text{MC,TISTOS}}$ are the trigger efficiency of L0&HLT1 evaluated in the reference channel using the TISTOS method. As described in Section 6.1, each data event in the reference channel is weighted with the so-called *sWeight* (*sWgt*) to extract the signal-only component. The TISTOS trigger efficiency on preselected data is therefore defined as

$$\epsilon^{D_s}_{\text{Data,TISTOS}} = \frac{\sum w^{\text{TIS\&TOS}\,|\,Sel}_{sWgt}}{\sum w^{\text{TIS}\,|\,Sel}_{sWgt}}, \tag{9.15}$$

where TIS | *Sel* represents the preselected events that triggered the selection of TIS lines in Eq. 9.12, while TIS&TOS | *Sel* represents the preselected events that also triggered the TOS lines from Table 5.3. Simulated events are weighted with w_{fraction},

TABLE 9.14: Data/MC correction factor (calculated considering L0&HLT1 efficiencies) for the trigger efficiency ratio for both the 2μ and the 3μ samples and for the three years. The first uncertainty is the statistical one, while the second one is the systematic uncertainty coming from the choice of the binning scheme.

	2016	**2017**	**2018**
$R(D_s, \tau_{3\mu})$	$1.036 \pm 0.043 \pm 0.004$	$0.964 \pm 0.030 \pm 0.007$	$0.974 \pm 0.032 \pm 0.007$
$R(D_s, \tau_{2\mu})$	$0.992 \pm 0.048 \pm 0.004$	$0.961 \pm 0.033 \pm 0.011$	$0.977 \pm 0.036 \pm 0.009$

thus the MC trigger efficiencies are given by

$$\epsilon_{\mathrm{MC,TOS}} = \frac{\sum w^{\mathrm{TOS}\,|\,Sel}_{\mathrm{fraction}}}{\sum w^{Sel}_{\mathrm{fraction}}}, \quad \epsilon_{\mathrm{MC,TISTOS}} = \frac{\sum w^{\mathrm{TIS\&TOS}\,|\,Sel}_{\mathrm{fraction}}}{\sum w^{\mathrm{TIS}\,|\,Sel}_{\mathrm{fraction}}}. \tag{9.16}$$

As defined in Eq. 9.14, the TISTOS efficiencies in data and simulation are obtained from the convolution of the corresponding TISTOS efficiencies evaluated on the reference channel with the phase space distribution of the signal events. The convolution is performed in bins of the transverse momentum (p_T) and χ^2 of the impact parameter ($IP\chi^2$) of the final-state track with greatest p_T[1], and f_i^τ corresponds to the fraction of signal events falling in the single bin i. The number of bins in each dimension is such to have an efficiency ratio as close as possible to unity, while making sure to have enough statistics in each bin.

The correction factors to the trigger efficiency are reported in Table 9.14 for the 3μ and 2μ samples, and for the three data-taking years. The former uncertainty is the statistical one, while the latter is the systematic uncertainty originated from the choice of the number of p_T and $IP\chi^2$ bins.

9.5.4 Normalisation summary

The various efficiencies and correction factors discussed so far are summarised in Table 9.15 for the three data-taking years. The `isMuon` selection efficiency is included in the offline selection efficiency ϵ_τ^{SEL}. The term $\epsilon_{2\mu}^{RICH}$ is the efficiency of the selection of events where the track with `!isMuon` falls within the RICH detectors acceptance, as discussed in Section 9.4.4. The combined efficiency $\epsilon_\tau^{clfs\&PID}$ includes both the efficiency of the selection in the output of the two classifiers and the efficiency of the veto applied on ProbNNk in the 3μ sample and in the 2016 2μ sample.

The final values for the normalisation factor α are shown in Table 9.16 separately for the 2μ and for the 3μ samples. The single terms entering the definition of α are also shown. The efficiency ratio $\epsilon_{D_s}/\epsilon_\tau$ is already corrected for the correction factors reported in Table 9.15.

---

[1] In the 2μ sample the muon with higher p_T is selected among the two tracks satisfying the `isMuon` requirement.

TABLE 9.15: Summary of the efficiencies and relative corrections evaluated on the signal channel, separately for the 2μ and the 3μ sample, and on the reference channel. The first and the second uncertainties are the statistical and systematic uncertainty, respectively; when a single uncertainty is reported this is purely statistical.

	2016	**2017**	**2018**
ϵ_{τ}^{ACC} [%]	10.72 ± 0.43		
$\epsilon_{D_s}^{ACC}$ [%]	12.04 ± 0.33		
ϵ_{τ}^{FILT} [%]	57.19 ± 0.04	57.11 ± 0.04	57.05 ± 0.04
$\epsilon_{D_s}^{FILT}$ [%]	47.53 ± 0.03	47.62 ± 0.03	47.50 ± 0.03
ϵ_{τ}^{REC} [%]	39.73 ± 0.05	39.65 ± 0.05	39.77 ± 0.05
$\epsilon_{D_s}^{REC}$ [%]	16.94 ± 0.04	17.05 ± 0.03	17.09 ± 0.03
$\epsilon_{\tau,2\mu}^{SEL}$ [%]	22.24 ± 0.07	22.25 ± 0.07	22.19 ± 0.07
$\epsilon_{\tau,3\mu}^{SEL}$ [%]	53.60 ± 0.09	53.66 ± 0.09	53.61 ± 0.09
$\epsilon_{D_s}^{SEL}$ [%]	91.12 ± 0.07	91.18 ± 0.07	91.11 ± 0.07
$\epsilon_{\tau,2\mu}^{TRIG}$ [%]	8.17 ± 0.10	22.35 ± 0.15	19.05 ± 0.15
$\epsilon_{\tau,3\mu}^{TRIG}$ [%]	58.38 ± 0.12	66.77 ± 0.11	57.65 ± 0.12
$\epsilon_{D_s}^{TRIG}$ [%]	14.79 ± 0.09	31.76 ± 0.11	27.59 ± 0.12
$\epsilon_{\tau,2\mu}^{RICH}$ [%]	97.25 ± 0.21	97.99 ± 0.11	98.11 ± 0.12
$\epsilon_{\tau,2\mu}^{clfs}$ [%]	10.50 ± 0.41	13.76 ± 0.28	15.12 ± 0.31
$\epsilon_{\tau,3\mu}^{clfs\&PID}$ [%]	61.46 ± 0.14	62.62 ± 0.14	62.71 ± 0.15
R_{Track}	$1.004 \pm 0.005 \pm 0.014$	$1.000 \pm 0.005 \pm 0.014$	$1.002 \pm 0.005 \pm 0.014$
c_{ϕ}	1.018 ± 0.009		
$R(D_s, \tau_{2\mu})$	$0.992 \pm 0.048 \pm 0.004$	$0.961 \pm 0.033 \pm 0.011$	$0.977 \pm 0.036 \pm 0.009$
$R(D_s, \tau_{3\mu})$	$1.036 \pm 0.043 \pm 0.004$	$0.964 \pm 0.030 \pm 0.007$	$0.974 \pm 0.032 \pm 0.007$

TABLE 9.16: Summary of the factors entering in the normalisation factor, together with their combined statistical and systematic uncertainties. The final values for α in the 2μ and in the 3μ samples are also shown.

	2016	**2017**	**2018**
$\mathcal{B}(D_s^+ \to \phi(\mu^+\mu^-)\pi^+)$		$(1.289 \pm 0.094) \times 10^{-5}$	
$f_{D_s}^{\tau}$		$(78.4 \pm 3.7) \times 10^{-2}$	
$\mathcal{B}(D_s^+ \to \tau^+\nu_\tau)$		$(5.32 \pm 0.11) \times 10^{-2}$	
$N(D_s^+ \to \phi(\mu^+\mu^-)\pi^+)^{meas}$	43072 ± 397	88474 ± 594	103790 ± 643
$\epsilon_{D_s}/\epsilon_{\tau,2\mu}$	29.31 ± 2.40	17.03 ± 1.14	16.05 ± 1.09
$\epsilon_{D_s}/\epsilon_{\tau,3\mu}$	0.30 ± 0.02	0.51 ± 0.03	0.52 ± 0.03
$\alpha_{2\mu}$	$(12.923 \pm 1.571) \times 10^{-8}$	$(3.654 \pm 0.409) \times 10^{-8}$	$(2.937 \pm 0.330) \times 10^{-8}$
$\alpha_{3\mu}$	$(1.302 \pm 0.146) \times 10^{-9}$	$(1.095 \pm 0.119) \times 10^{-9}$	$(0.947 \pm 0.103) \times 10^{-9}$

9.6 Single bin efficiency in the 3μ sample

The upper limit on the BF of the $\tau^+ \to \mu^+\mu^-\mu^+$ decay is evaluated in bins of XGB and SuperPNN and in bins of the invariant mass distribution. Therefore, the normalisation factor α evaluated in the 3μ sample, has to be divided by the single bin efficiency that can be factorised as

$$\epsilon_{i,j} = \epsilon_{\mathrm{XGB\&PID},i} \times \epsilon_{\mathrm{mass},j} \,. \tag{9.17}$$

$\epsilon_{\mathrm{XGB\&PID},i}$ is the XGB and SuperPNN bin efficiency, while $\epsilon_{\mathrm{mass},j}$ represents the fraction of signal events in the single mass bin. The single mass bin efficiencies are reported in Table 7.6 of Section 7.2.4. The single XGB and SuperPNN bin efficiencies are evaluated as the fraction of events falling in each XGB and SuperPNN bin relative to the number of events passing the minimum cut in XGB (> 0.8) and SuperPNN (> 0.88) and the veto on ProbNNk. The maps containing the final single bin efficiency are shown in Figure 9.5 for each data-taking year.

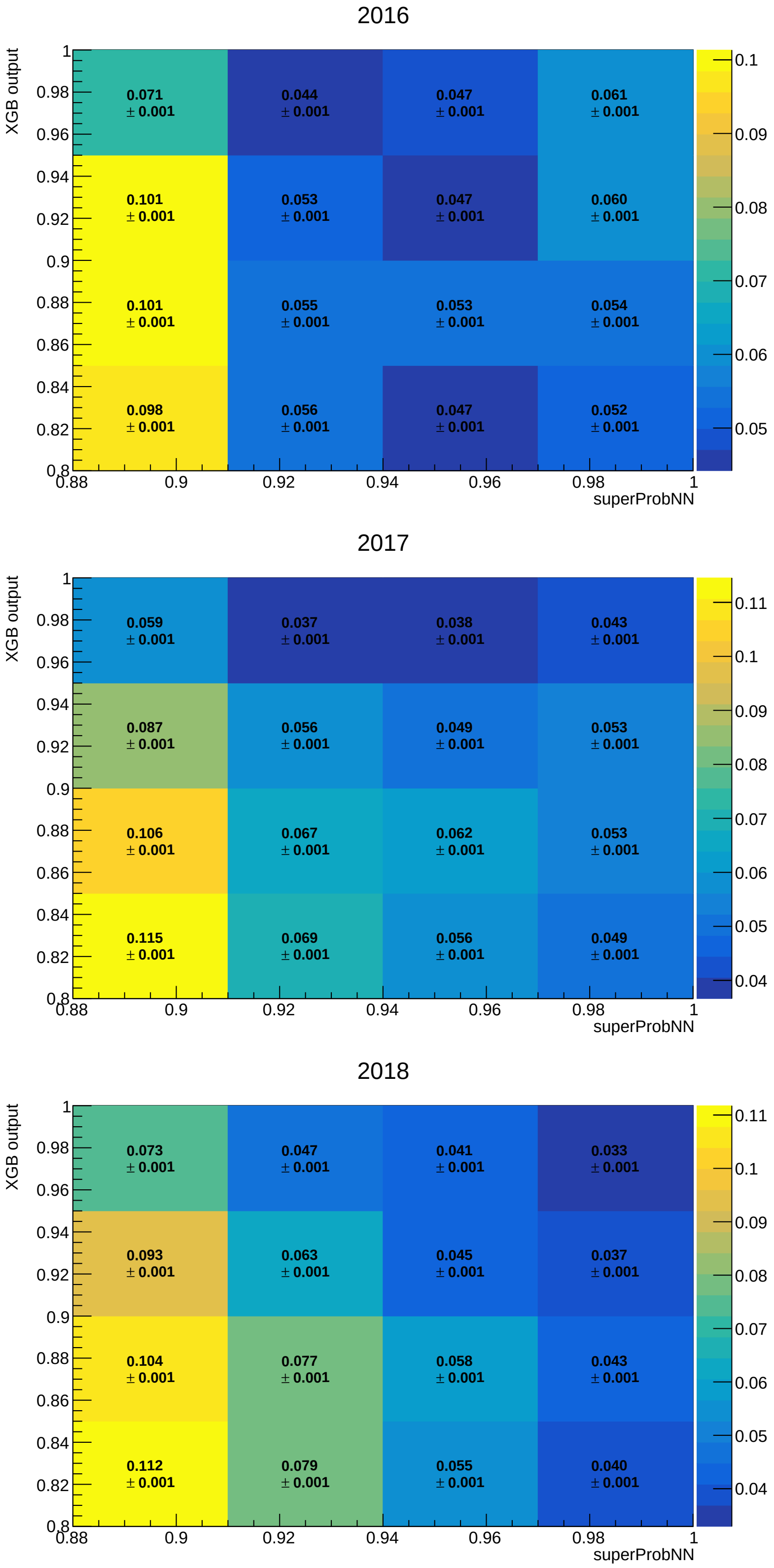


FIGURE 9.5: Maps containing the fraction of events falling in each XGB and SuperPNN bin evaluated on simulated events in the 3μ sample for the three years.

Chapter 10

Systematic uncertainties

In this Chapter, the various sources of systematic uncertainties are described in detail. These include the systematics on the normalisation factor α (Section 10.1) and the systematics introduced by the choice of the pdfs used to describe the signal peak (Section 10.2) and the background shape (Section 10.3).

The simulation used in this analysis has been generated assuming a uniform phase space distribution for the $\tau^+ \to \mu^+\mu^-\mu^+$ decay. However, the amplitude structure of the decay may depend on various beyond the Standard Model effects. The systematic uncertainty introduced by different decay models on the selection efficiency is discussed in Appendix D.

10.1 Systematics on the normalisation

All the quantities used in the evaluation of the normalisation factor α come with an uncertainty. Table 10.1 shows an overview of the relative uncertainty on α in the 3μ and 2μ samples for the three years, followed by the relative uncertainty of each term entering its evaluation.

It is clear that the major contribution to the uncertainty on α comes from external inputs, namely from the branching fraction of the $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ decay and from $f^{\tau}_{D_s}$, which is itself affected by the uncertainty on the cross sections and branching fractions used in its evaluation. The statistics of the simulated samples

TABLE 10.1: Relative uncertainties on the normalisation factor α and of all terms entering its evaluation, for the 2μ and for the 3μ samples and for the three years. The ratios of efficiencies $\epsilon_{D_s}/\epsilon_{\tau}$ do not include the correction factors defined in Section 9.5.

	Relative uncertainty [%]		
	2016	**2017**	**2018**
$\alpha_{2\mu}$	12.15	11.19	11.24
$\alpha_{3\mu}$	11.20	10.85	10.90
$f^{\tau}_{D_s}$		4.77	
$\mathcal{B}(D_s^+ \to \phi(\mu^+\mu^-)\pi^+)$		7.27	
$\mathcal{B}(D_s^+ \to \tau^+\nu_\tau)$		2.07	
$N(D_s^+ \to \phi(\mu^+\mu^-)\pi^+)^{meas}$	0.92	0.67	0.62
$\epsilon_{D_s}/\epsilon_{\tau,2\mu}$	6.39	5.36	5.37
$\epsilon_{D_s}/\epsilon_{\tau,3\mu}$	4.93	4.91	4.91
R_{Track}	0.50(stat) 1.39(syst)	0.50(stat) 1.40(syst)	0.50(stat) 1.40(syst)
c_ϕ		0.89	
$R(D_s, \tau_{2\mu})$	4.80(stat) 0.40(syst)	3.44(stat) 1.12(syst)	3.67(stat) 0.95(syst)
$R(D_s, \tau_{3\mu})$	4.16(stat) 0.37(syst)	3.11(stat) 0.76(syst)	3.29(stat) 0.72(syst)

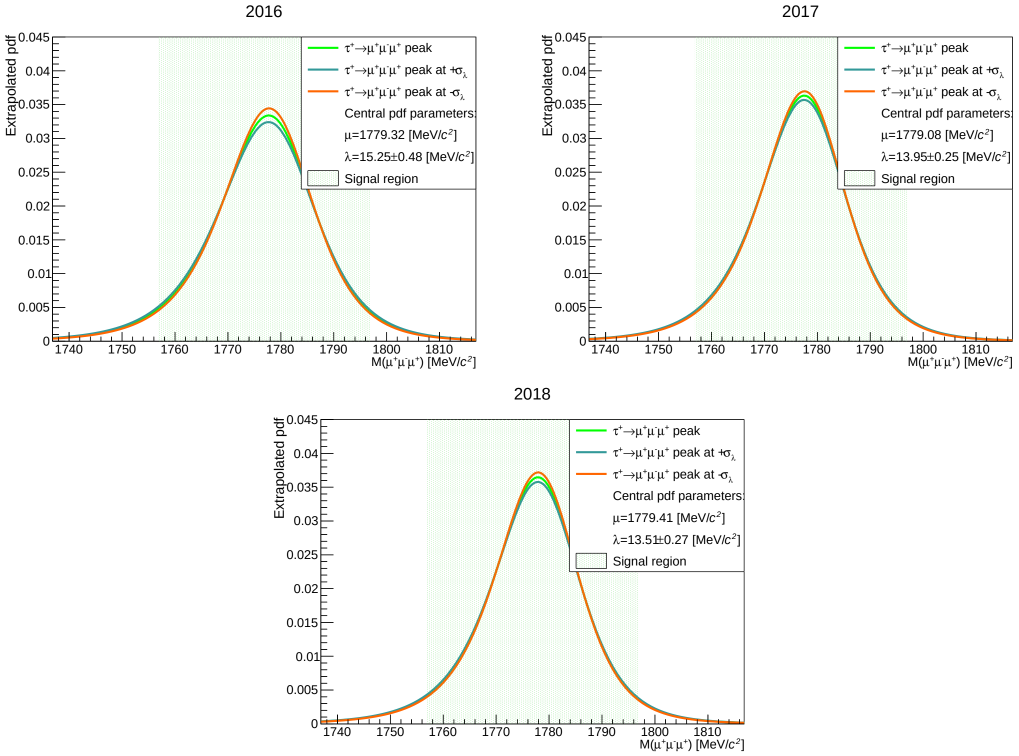


FIGURE 10.1: Fluctuation of the extrapolated pdf in the 3μ sample for the the three data-taking years. The cyan and orange curves are obtained by varying the λ parameter by $+\sigma_\lambda$ and $-\sigma_\lambda$, respectively, where σ_λ is obtained from the fit to the MC distribution and corrected with c_λ.

also introduces a large uncertainty, as it can be seen from the relative uncertainty on the efficiency ratio and on the correction to the trigger efficiency. The statistical and systematic uncertainties affecting α are treated as a unique uncertainty in the limit evaluation.

10.2 Systematics on the signal shape

The systematic uncertainty introduced by the choice of the pdf used to model the signal shape is evaluated by varying the λ parameter of the Johnson's S_U function used to describe the signal peak by $\pm\sigma_\lambda$. The uncertainty on λ is taken from the an extended maximum likelihood fit to the invariant mass distribution in simulation, then corrected for the c_λ factor evaluated from the fit to the reference channel (see Table 6.3). In the 3μ sample, the fit to the MC distribution (see Figure 7.11) is performed on events selected by the minimum requirements on XGB and SuperPNN, and the veto on ProbNNk. The fluctuated extrapolated pdfs are shown in Figure 10.1 for the three data-taking years.

In the 2μ sample, the signal pdf is extrapolated from the fit performed on simulated events after the selection in the BDTGKT and BDTGPID output, and the veto on ProbNNk in 2016 only. The result of the fit to the invariant mass distribution in the 2μ sample is shown in Figure 10.2, while the corresponding fluctuated pdfs are reported in Figure 10.3 for the three years.

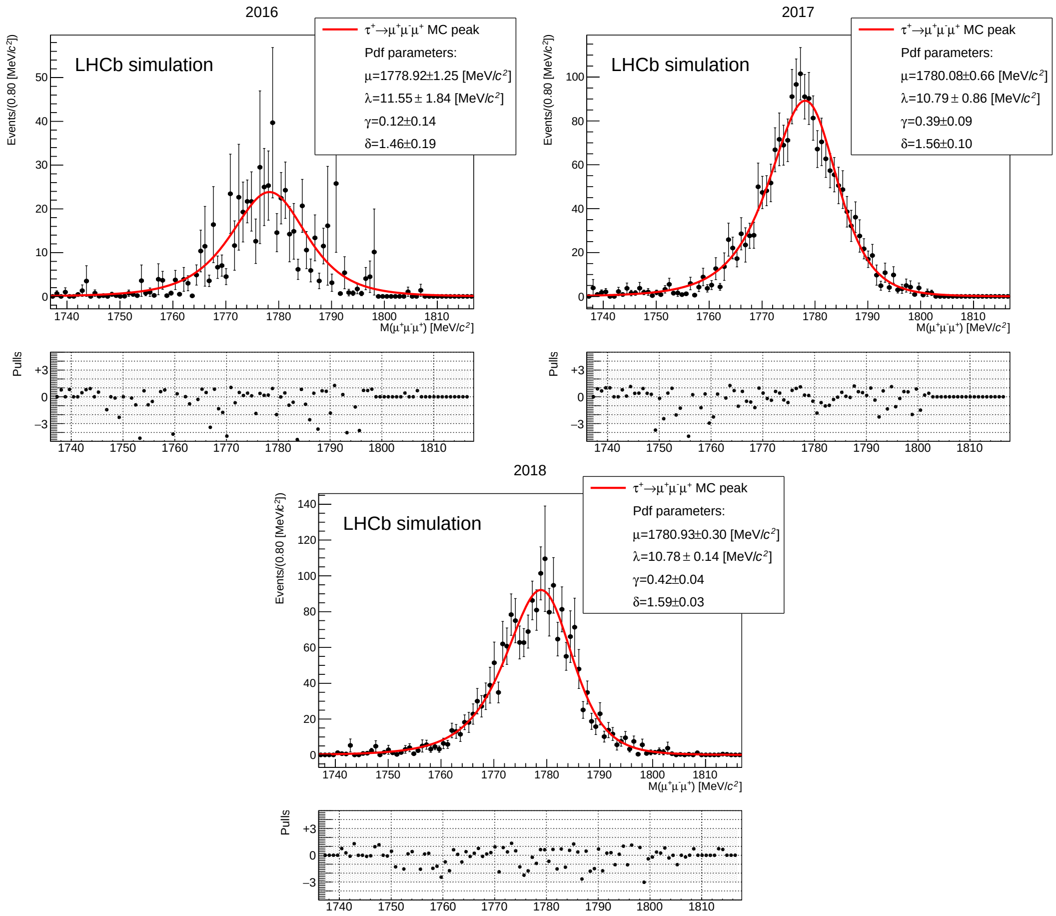


FIGURE 10.2: Fit to the invariant mass distribution obtained for simulated events in the 2μ sample after the selection in the BDTGKT and BDTGPID output and the veto on ProbNNk (2016 only) for the three data-taking years.

The fluctuation of the signal pdf reflects in the fluctuation of the fraction of events falling in the signal region and, consequently, in the fluctuation of the fraction of events falling in the single mass bin. The latter fluctuation is only relevant for the 3μ sample, as no binning is performed in the 2μ sample because of the limited statistics of the simulated sample. The fraction of events in the mass bins is reported in Table 10.2 and in Table 10.3 for the $+\sigma_\lambda$ and $-\sigma_\lambda$ fluctuation, respectively.

The fluctuation of the fraction of events in the signal region in the 2μ sample is reported in Table 10.4 together with the central value for each data-taking year.

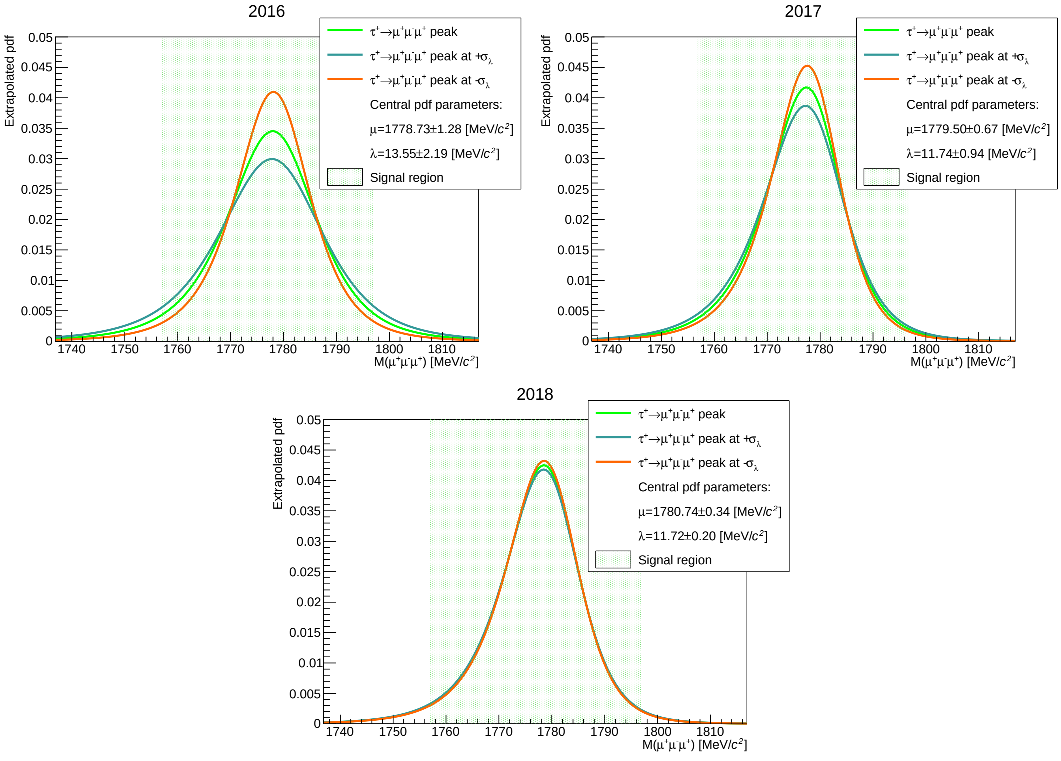


FIGURE 10.3: Fluctuation of the extrapolated pdf in the 2μ sample for the the three data-taking years. The cyan and orange curves are obtained by varying the λ fit parameter by $+\sigma_\lambda$ and $-\sigma_\lambda$, respectively, where σ_λ is obtained from the fit to the MC distribution and corrected with c_λ.

TABLE 10.2: Fraction of events falling in each mass bin evaluated on the extrapolated signal pdf fluctuated by $+\sigma_\lambda$ for the three data-taking years. The fraction of events in the signal region is also reported (Total).

Mass bin	**2016**	**2017**	**2018**
1	0.0442 ± 0.0004	0.0399 ± 0.0003	0.0381 ± 0.0003
2	0.0799 ± 0.0007	0.0772 ± 0.0005	0.0734 ± 0.0006
3	0.1331 ± 0.0010	0.1381 ± 0.0008	0.1325 ± 0.0008
4	0.1861 ± 0.0006	0.2031 ± 0.0006	0.1995 ± 0.0007
5	0.1957 ± 0.0013	0.2129 ± 0.0011	0.2154 ± 0.0012
6	0.1480 ± 0.0005	0.1500 ± 0.0006	0.1553 ± 0.0006
7	0.0857 ± 0.0007	0.0781 ± 0.0006	0.0818 ± 0.0007
8	0.0427 ± 0.0004	0.0351 ± 0.0003	0.0371 ± 0.0004
Total	0.9154 ± 0.0013	0.9344 ± 0.0010	0.9332 ± 0.0010

TABLE 10.3: Fraction of events falling in each mass bin evaluated on the extrapolated signal pdf fluctuated by $-\sigma_\lambda$ for the three data-taking years. The fraction of events in the signal region is also reported (Total).

Mass bin	**2016**	**2017**	**2018**
1	0.0402 ± 0.0003	0.0375 ± 0.0003	0.0356 ± 0.0003
2	0.0761 ± 0.0006	0.0747 ± 0.0005	0.0707 ± 0.0005
3	0.1335 ± 0.0008	0.1379 ± 0.0007	0.1320 ± 0.0008
4	0.1947 ± 0.0006	0.2084 ± 0.0007	0.2049 ± 0.0007
5	0.2076 ± 0.0012	0.2202 ± 0.0010	0.2236 ± 0.0011
6	0.1527 ± 0.0006	0.1522 ± 0.0006	0.1581 ± 0.0007
7	0.0838 ± 0.0007	0.0765 ± 0.0006	0.0802 ± 0.0006
8	0.0394 ± 0.0004	0.0332 ± 0.0003	0.0350 ± 0.0003
Total	0.9280 ± 0.0010	0.9407 ± 0.0008	0.9400 ± 0.0009

TABLE 10.4: Fraction of events falling in the signal region evaluated in the 2μ sample as the integral of the extrapolated pdf. The values obtained fluctuating the fit parameter λ by $\pm\sigma_\lambda$ are also reported.

λ fluctuation	**2016**	**2017**	**2018**
None	0.9221 ± 0.0258	0.9561 ± 0.0285	0.9611 ± 0.0591
$+\sigma_\lambda$	0.8907 ± 0.0330	0.9450 ± 0.0308	0.9590 ± 0.0605
$-\sigma_\lambda$	0.9508 ± 0.0159	0.9662 ± 0.0260	0.9632 ± 0.0577

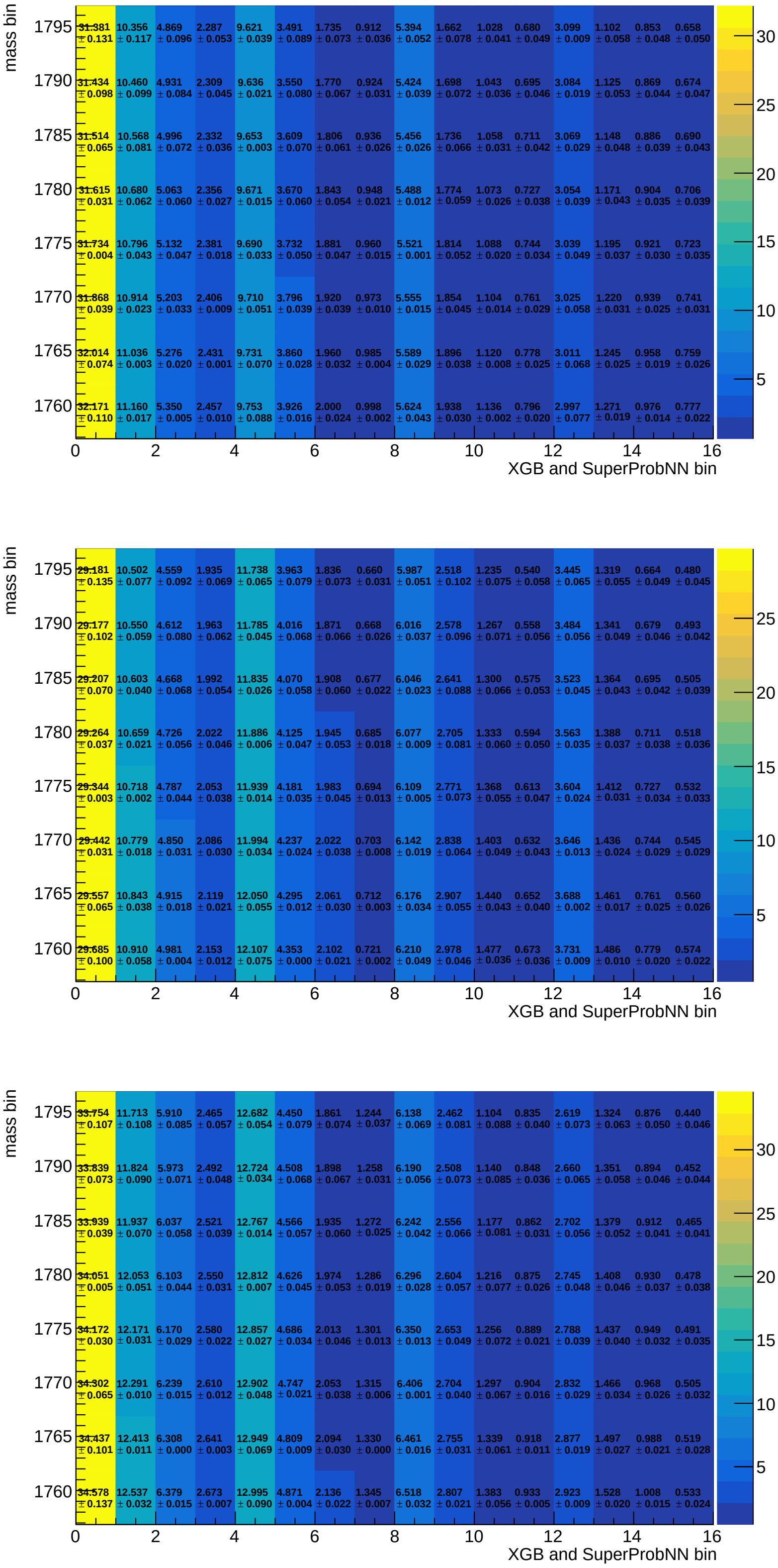
10.3 Systematics on the background shape

The pdf used to model the background shape also introduces a systematic uncertainty. As outlined in Section 8.2.3 for the 3μ sample and in Section 8.2.5 for the 2μ sample, the final fit to the data sidebands is performed by including the contribution of the combinatorial background and of the two charmed modes $D^+ \to \pi^-\pi^+\pi^+$ and $D_s^+ \to \pi^-\pi^+\pi^+$. The total number of background events expected in the signal region is extrapolated from the fit to the outer sidebands.

In the 3μ sample, a simultaneous fit is performed in the 16 XGB and SuperPNN bins and, in each of these bins, the integral of the resulting pdf in the single mass bin is evaluated. The number of background events expected in the single XGB, SuperPNN and mass bin is reported in Figure 10.4 for the three data-taking years.

In the 2μ sample, no binning is considered due to the limited statistics. The result of the fit to the data sidebands is reported in Figure 8.21, while the extrapolated number of background events expected from each background source is reported in Table 8.15.

Both in the 3μ and in the 2μ sample, the fit to the data sidebands assuming the contribution of the combinatorial background only is used to assign a systematic uncertainty to the expected number of background events in the signal region, that accounts for the modelling of the background distribution. The maps containing the expected number of background events in each XGB, SuperPNN and mass bin considered in the 3μ sample is reported in Figure 10.5 for the three years. For the 2μ sample, the same numbers are reported in Table 8.13 for the three data-taking years.

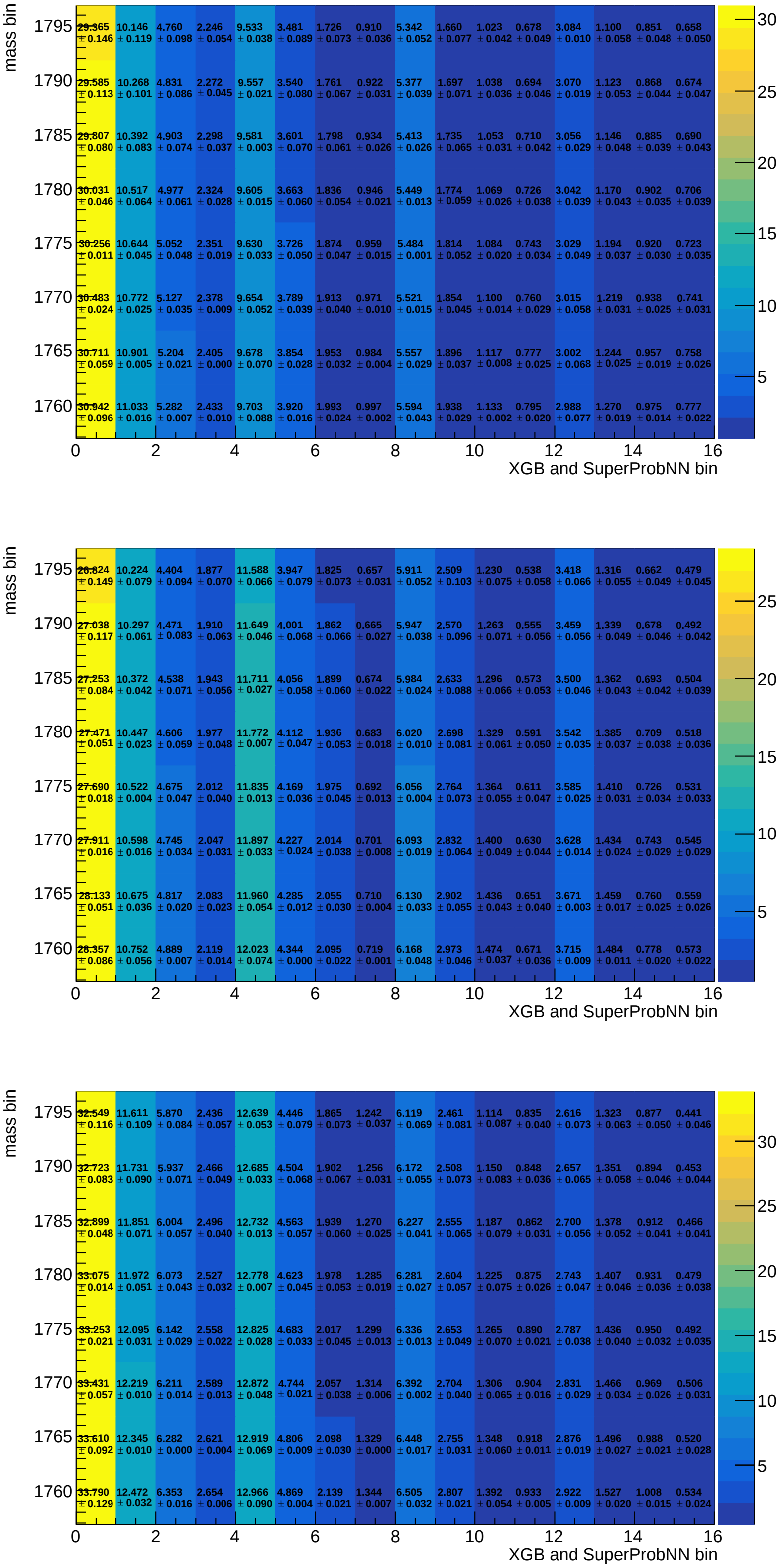


1795	31.381 ± 0.131	10.356 ± 0.117	4.869 ± 0.096	2.287 ± 0.053	9.621 ± 0.039	3.491 ± 0.089	1.735 ± 0.073	0.912 ± 0.036	5.394 ± 0.052	1.662 ± 0.078	1.028 ± 0.041	0.680 ± 0.049	3.099 ± 0.009	1.102 ± 0.058	0.853 ± 0.048	0.658 ± 0.050
1790	31.434 ± 0.098	10.460 ± 0.099	4.931 ± 0.084	2.309 ± 0.045	9.636 ± 0.021	3.550 ± 0.080	1.770 ± 0.067	0.924 ± 0.031	5.424 ± 0.039	1.698 ± 0.072	1.043 ± 0.036	0.695 ± 0.046	3.084 ± 0.019	1.125 ± 0.053	0.869 ± 0.044	0.674 ± 0.047
1785	31.514 ± 0.065	10.568 ± 0.081	4.996 ± 0.072	2.332 ± 0.036	9.653 ± 0.003	3.609 ± 0.070	1.806 ± 0.061	0.936 ± 0.026	5.456 ± 0.026	1.736 ± 0.066	1.058 ± 0.031	0.711 ± 0.042	3.069 ± 0.029	1.148 ± 0.048	0.886 ± 0.039	0.690 ± 0.043
1780	31.615 ± 0.031	10.680 ± 0.062	5.063 ± 0.060	2.356 ± 0.027	9.671 ± 0.015	3.670 ± 0.060	1.843 ± 0.054	0.948 ± 0.021	5.488 ± 0.012	1.774 ± 0.059	1.073 ± 0.026	0.727 ± 0.038	3.054 ± 0.039	1.171 ± 0.043	0.904 ± 0.035	0.706 ± 0.039
1775	31.734 ± 0.004	10.796 ± 0.043	5.132 ± 0.047	2.381 ± 0.018	9.690 ± 0.033	3.732 ± 0.050	1.881 ± 0.047	0.960 ± 0.015	5.521 ± 0.001	1.814 ± 0.052	1.088 ± 0.020	0.744 ± 0.034	3.039 ± 0.049	1.195 ± 0.037	0.921 ± 0.030	0.723 ± 0.035
1770	31.868 ± 0.039	10.914 ± 0.023	5.203 ± 0.033	2.406 ± 0.009	9.710 ± 0.051	3.796 ± 0.039	1.920 ± 0.039	0.973 ± 0.010	5.555 ± 0.015	1.854 ± 0.045	1.104 ± 0.014	0.761 ± 0.029	3.025 ± 0.058	1.220 ± 0.031	0.939 ± 0.025	0.741 ± 0.031
1765	32.014 ± 0.074	11.036 ± 0.003	5.276 ± 0.020	2.431 ± 0.001	9.731 ± 0.070	3.860 ± 0.028	1.960 ± 0.032	0.985 ± 0.004	5.589 ± 0.029	1.896 ± 0.038	1.120 ± 0.008	0.778 ± 0.025	3.011 ± 0.068	1.245 ± 0.025	0.958 ± 0.019	0.759 ± 0.026
1760	32.171 ± 0.110	11.160 ± 0.017	5.350 ± 0.005	2.457 ± 0.010	9.753 ± 0.088	3.926 ± 0.016	2.000 ± 0.024	0.998 ± 0.002	5.624 ± 0.043	1.938 ± 0.030	1.136 ± 0.002	0.796 ± 0.020	2.997 ± 0.077	1.271 ± 0.019	0.976 ± 0.014	0.777 ± 0.022

1795	29.181 ± 0.135	10.502 ± 0.077	4.559 ± 0.092	1.935 ± 0.069	11.738 ± 0.065	3.963 ± 0.079	1.836 ± 0.073	0.660 ± 0.031	5.987 ± 0.051	2.518 ± 0.102	1.235 ± 0.075	0.540 ± 0.058	3.445 ± 0.065	1.319 ± 0.055	0.664 ± 0.049	0.480 ± 0.045
1790	29.177 ± 0.102	10.550 ± 0.059	4.612 ± 0.080	1.963 ± 0.062	11.785 ± 0.045	4.016 ± 0.068	1.871 ± 0.066	0.668 ± 0.026	6.016 ± 0.037	2.578 ± 0.096	1.267 ± 0.071	0.558 ± 0.056	3.484 ± 0.056	1.341 ± 0.049	0.679 ± 0.046	0.493 ± 0.042
1785	29.207 ± 0.070	10.603 ± 0.040	4.668 ± 0.068	1.992 ± 0.054	11.835 ± 0.026	4.070 ± 0.058	1.908 ± 0.060	0.677 ± 0.022	6.046 ± 0.023	2.641 ± 0.088	1.300 ± 0.066	0.575 ± 0.053	3.523 ± 0.045	1.364 ± 0.043	0.695 ± 0.042	0.505 ± 0.039
1780	29.264 ± 0.037	10.659 ± 0.021	4.726 ± 0.056	2.022 ± 0.046	11.886 ± 0.006	4.125 ± 0.047	1.945 ± 0.053	0.685 ± 0.018	6.077 ± 0.009	2.705 ± 0.081	1.333 ± 0.060	0.594 ± 0.050	3.563 ± 0.035	1.388 ± 0.037	0.711 ± 0.038	0.518 ± 0.036
1775	29.344 ± 0.003	10.718 ± 0.002	4.787 ± 0.044	2.053 ± 0.038	11.939 ± 0.014	4.181 ± 0.035	1.983 ± 0.045	0.694 ± 0.013	6.109 ± 0.005	2.771 ± 0.073	1.368 ± 0.055	0.613 ± 0.047	3.604 ± 0.024	1.412 ± 0.031	0.727 ± 0.034	0.532 ± 0.033
1770	29.442 ± 0.031	10.779 ± 0.018	4.850 ± 0.031	2.086 ± 0.030	11.994 ± 0.034	4.237 ± 0.024	2.022 ± 0.038	0.703 ± 0.008	6.142 ± 0.019	2.838 ± 0.064	1.403 ± 0.049	0.632 ± 0.043	3.646 ± 0.013	1.436 ± 0.024	0.744 ± 0.029	0.545 ± 0.029
1765	29.557 ± 0.065	10.843 ± 0.038	4.915 ± 0.018	2.119 ± 0.021	12.050 ± 0.055	4.295 ± 0.012	2.061 ± 0.030	0.712 ± 0.003	6.176 ± 0.034	2.907 ± 0.055	1.440 ± 0.043	0.652 ± 0.040	3.688 ± 0.002	1.461 ± 0.017	0.761 ± 0.025	0.560 ± 0.026
1760	29.685 ± 0.100	10.910 ± 0.058	4.981 ± 0.004	2.153 ± 0.012	12.107 ± 0.075	4.353 ± 0.000	2.102 ± 0.021	0.721 ± 0.002	6.210 ± 0.049	2.978 ± 0.046	1.477 ± 0.036	0.673 ± 0.036	3.731 ± 0.009	1.486 ± 0.010	0.779 ± 0.020	0.574 ± 0.022

1795	33.754 ± 0.107	11.713 ± 0.108	5.910 ± 0.085	2.465 ± 0.057	12.682 ± 0.054	4.450 ± 0.079	1.861 ± 0.074	1.244 ± 0.037	6.138 ± 0.069	2.462 ± 0.081	1.104 ± 0.088	0.835 ± 0.040	2.619 ± 0.073	1.324 ± 0.063	0.876 ± 0.050	0.440 ± 0.046
1790	33.839 ± 0.073	11.824 ± 0.090	5.973 ± 0.071	2.492 ± 0.048	12.724 ± 0.034	4.508 ± 0.068	1.898 ± 0.067	1.258 ± 0.031	6.190 ± 0.056	2.508 ± 0.073	1.140 ± 0.085	0.848 ± 0.036	2.660 ± 0.065	1.351 ± 0.058	0.894 ± 0.046	0.452 ± 0.044
1785	33.939 ± 0.039	11.937 ± 0.070	6.037 ± 0.058	2.521 ± 0.039	12.767 ± 0.014	4.566 ± 0.057	1.935 ± 0.060	1.272 ± 0.025	6.242 ± 0.042	2.556 ± 0.066	1.177 ± 0.081	0.862 ± 0.031	2.702 ± 0.056	1.379 ± 0.052	0.912 ± 0.041	0.465 ± 0.041
1780	34.051 ± 0.005	12.053 ± 0.051	6.103 ± 0.044	2.550 ± 0.031	12.812 ± 0.007	4.626 ± 0.045	1.974 ± 0.053	1.286 ± 0.019	6.296 ± 0.028	2.604 ± 0.057	1.216 ± 0.077	0.875 ± 0.026	2.745 ± 0.048	1.408 ± 0.046	0.930 ± 0.037	0.478 ± 0.038
1775	34.172 ± 0.030	12.171 ± 0.031	6.170 ± 0.029	2.580 ± 0.022	12.857 ± 0.027	4.686 ± 0.034	2.013 ± 0.046	1.301 ± 0.013	6.350 ± 0.013	2.653 ± 0.049	1.256 ± 0.072	0.889 ± 0.021	2.788 ± 0.039	1.437 ± 0.040	0.949 ± 0.032	0.491 ± 0.035
1770	34.302 ± 0.065	12.291 ± 0.010	6.239 ± 0.015	2.610 ± 0.012	12.902 ± 0.048	4.747 ± 0.021	2.053 ± 0.038	1.315 ± 0.006	6.406 ± 0.001	2.704 ± 0.040	1.297 ± 0.067	0.904 ± 0.016	2.832 ± 0.029	1.466 ± 0.034	0.968 ± 0.026	0.505 ± 0.032
1765	34.437 ± 0.101	12.413 ± 0.011	6.308 ± 0.000	2.641 ± 0.003	12.949 ± 0.069	4.809 ± 0.009	2.094 ± 0.030	1.330 ± 0.000	6.461 ± 0.016	2.755 ± 0.031	1.339 ± 0.061	0.918 ± 0.011	2.877 ± 0.019	1.497 ± 0.027	0.988 ± 0.021	0.519 ± 0.028
1760	34.578 ± 0.137	12.537 ± 0.032	6.379 ± 0.015	2.673 ± 0.007	12.995 ± 0.090	4.871 ± 0.004	2.136 ± 0.022	1.345 ± 0.007	6.518 ± 0.032	2.807 ± 0.021	1.383 ± 0.056	0.933 ± 0.005	2.923 ± 0.009	1.528 ± 0.020	1.008 ± 0.015	0.533 ± 0.024

FIGURE 10.4: Maps containing the expected number of background events in the single XGB, SuperPNN and mass bin in the 3μ sample. The yields are extracted from a fit to the outer sidebands including the contribution of the combinatorial background and of the $D^+ \to \pi^-\pi^+\pi^+$ and $D_s^+ \to \pi^-\pi^+\pi^+$ modes. The maps are shown for the three data-taking years.



1795	29.365 ± 0.146	10.146 ± 0.119	4.760 ± 0.098	2.246 ± 0.054	9.533 ± 0.038	3.481 ± 0.089	1.726 ± 0.073	0.910 ± 0.036	5.342 ± 0.052	1.660 ± 0.077	1.023 ± 0.042	0.678 ± 0.049	3.084 ± 0.010	1.100 ± 0.058	0.851 ± 0.048	0.658 ± 0.050
1790	29.585 ± 0.113	10.268 ± 0.101	4.831 ± 0.086	2.272 ± 0.045	9.557 ± 0.021	3.540 ± 0.080	1.761 ± 0.067	0.922 ± 0.031	5.377 ± 0.039	1.697 ± 0.071	1.038 ± 0.036	0.694 ± 0.046	3.070 ± 0.019	1.123 ± 0.053	0.868 ± 0.044	0.674 ± 0.047
1785	29.807 ± 0.080	10.392 ± 0.083	4.903 ± 0.074	2.298 ± 0.037	9.581 ± 0.003	3.601 ± 0.070	1.798 ± 0.061	0.934 ± 0.026	5.413 ± 0.026	1.735 ± 0.065	1.053 ± 0.031	0.710 ± 0.042	3.056 ± 0.029	1.146 ± 0.048	0.885 ± 0.039	0.690 ± 0.043
1780	30.031 ± 0.046	10.517 ± 0.064	4.977 ± 0.061	2.324 ± 0.028	9.605 ± 0.015	3.663 ± 0.060	1.836 ± 0.054	0.946 ± 0.021	5.449 ± 0.013	1.774 ± 0.059	1.069 ± 0.026	0.726 ± 0.038	3.042 ± 0.039	1.170 ± 0.043	0.902 ± 0.035	0.706 ± 0.039
1775	30.256 ± 0.011	10.644 ± 0.045	5.052 ± 0.048	2.351 ± 0.019	9.630 ± 0.033	3.726 ± 0.050	1.874 ± 0.047	0.959 ± 0.015	5.484 ± 0.001	1.814 ± 0.052	1.084 ± 0.020	0.743 ± 0.034	3.029 ± 0.049	1.194 ± 0.037	0.920 ± 0.030	0.723 ± 0.035
1770	30.483 ± 0.024	10.772 ± 0.025	5.127 ± 0.035	2.378 ± 0.009	9.654 ± 0.052	3.789 ± 0.039	1.913 ± 0.040	0.971 ± 0.010	5.521 ± 0.015	1.854 ± 0.045	1.100 ± 0.014	0.760 ± 0.029	3.015 ± 0.058	1.219 ± 0.031	0.938 ± 0.025	0.741 ± 0.031
1765	30.711 ± 0.059	10.901 ± 0.005	5.204 ± 0.021	2.405 ± 0.000	9.678 ± 0.070	3.854 ± 0.028	1.953 ± 0.032	0.984 ± 0.004	5.557 ± 0.029	1.896 ± 0.037	1.117 ± 0.008	0.777 ± 0.025	3.002 ± 0.068	1.244 ± 0.025	0.957 ± 0.019	0.758 ± 0.026
1760	30.942 ± 0.096	11.033 ± 0.016	5.282 ± 0.007	2.433 ± 0.010	9.703 ± 0.088	3.920 ± 0.016	1.993 ± 0.024	0.997 ± 0.002	5.594 ± 0.043	1.938 ± 0.029	1.133 ± 0.002	0.795 ± 0.020	2.988 ± 0.077	1.270 ± 0.019	0.975 ± 0.014	0.777 ± 0.022

1795	26.824 ± 0.149	10.224 ± 0.079	4.404 ± 0.094	1.877 ± 0.070	11.588 ± 0.066	3.947 ± 0.079	1.825 ± 0.073	0.657 ± 0.031	5.911 ± 0.052	2.509 ± 0.103	1.230 ± 0.075	0.538 ± 0.058	3.418 ± 0.066	1.316 ± 0.055	0.662 ± 0.049	0.479 ± 0.045
1790	27.038 ± 0.117	10.297 ± 0.061	4.471 ± 0.083	1.910 ± 0.063	11.649 ± 0.046	4.001 ± 0.068	1.862 ± 0.066	0.665 ± 0.027	5.947 ± 0.038	2.570 ± 0.096	1.263 ± 0.071	0.555 ± 0.056	3.459 ± 0.056	1.339 ± 0.049	0.678 ± 0.046	0.492 ± 0.042
1785	27.253 ± 0.084	10.372 ± 0.042	4.538 ± 0.071	1.943 ± 0.056	11.711 ± 0.027	4.056 ± 0.058	1.899 ± 0.060	0.674 ± 0.022	5.984 ± 0.024	2.633 ± 0.088	1.296 ± 0.066	0.573 ± 0.053	3.500 ± 0.046	1.362 ± 0.043	0.693 ± 0.042	0.504 ± 0.039
1780	27.471 ± 0.051	10.447 ± 0.023	4.606 ± 0.059	1.977 ± 0.048	11.772 ± 0.007	4.112 ± 0.047	1.936 ± 0.053	0.683 ± 0.018	6.020 ± 0.010	2.698 ± 0.081	1.329 ± 0.061	0.591 ± 0.050	3.542 ± 0.035	1.385 ± 0.037	0.709 ± 0.038	0.518 ± 0.036
1775	27.690 ± 0.018	10.522 ± 0.004	4.675 ± 0.047	2.012 ± 0.040	11.835 ± 0.013	4.169 ± 0.036	1.975 ± 0.045	0.692 ± 0.013	6.056 ± 0.004	2.764 ± 0.073	1.364 ± 0.055	0.611 ± 0.047	3.585 ± 0.025	1.410 ± 0.031	0.726 ± 0.034	0.531 ± 0.033
1770	27.911 ± 0.016	10.598 ± 0.016	4.745 ± 0.034	2.047 ± 0.031	11.897 ± 0.033	4.227 ± 0.024	2.014 ± 0.038	0.701 ± 0.008	6.093 ± 0.019	2.832 ± 0.064	1.400 ± 0.049	0.630 ± 0.044	3.628 ± 0.014	1.434 ± 0.024	0.743 ± 0.029	0.545 ± 0.029
1765	28.133 ± 0.051	10.675 ± 0.036	4.817 ± 0.020	2.083 ± 0.023	11.960 ± 0.054	4.285 ± 0.012	2.055 ± 0.030	0.710 ± 0.004	6.130 ± 0.033	2.902 ± 0.055	1.436 ± 0.043	0.651 ± 0.040	3.671 ± 0.003	1.459 ± 0.017	0.760 ± 0.025	0.559 ± 0.026
1760	28.357 ± 0.086	10.752 ± 0.056	4.889 ± 0.007	2.119 ± 0.014	12.023 ± 0.074	4.344 ± 0.000	2.095 ± 0.022	0.719 ± 0.001	6.168 ± 0.048	2.973 ± 0.046	1.474 ± 0.037	0.671 ± 0.036	3.715 ± 0.009	1.484 ± 0.011	0.778 ± 0.020	0.573 ± 0.022

1795	32.549 ± 0.116	11.611 ± 0.109	5.870 ± 0.084	2.436 ± 0.057	12.639 ± 0.053	4.446 ± 0.079	1.865 ± 0.073	1.242 ± 0.037	6.119 ± 0.069	2.461 ± 0.081	1.114 ± 0.087	0.835 ± 0.040	2.616 ± 0.073	1.323 ± 0.063	0.877 ± 0.050	0.441 ± 0.046
1790	32.723 ± 0.083	11.731 ± 0.090	5.937 ± 0.071	2.466 ± 0.049	12.685 ± 0.033	4.504 ± 0.068	1.902 ± 0.067	1.256 ± 0.031	6.172 ± 0.055	2.508 ± 0.073	1.150 ± 0.083	0.848 ± 0.036	2.657 ± 0.065	1.351 ± 0.058	0.894 ± 0.046	0.453 ± 0.044
1785	32.899 ± 0.048	11.851 ± 0.071	6.004 ± 0.057	2.496 ± 0.040	12.732 ± 0.013	4.563 ± 0.057	1.939 ± 0.060	1.270 ± 0.025	6.227 ± 0.041	2.555 ± 0.065	1.187 ± 0.079	0.862 ± 0.031	2.700 ± 0.056	1.378 ± 0.052	0.912 ± 0.041	0.466 ± 0.041
1780	33.075 ± 0.014	11.972 ± 0.051	6.073 ± 0.043	2.527 ± 0.032	12.778 ± 0.007	4.623 ± 0.045	1.978 ± 0.053	1.285 ± 0.019	6.281 ± 0.027	2.604 ± 0.057	1.225 ± 0.075	0.875 ± 0.026	2.743 ± 0.047	1.407 ± 0.046	0.931 ± 0.036	0.479 ± 0.038
1775	33.253 ± 0.021	12.095 ± 0.031	6.142 ± 0.029	2.558 ± 0.022	12.825 ± 0.028	4.683 ± 0.033	2.017 ± 0.045	1.299 ± 0.013	6.336 ± 0.013	2.653 ± 0.049	1.265 ± 0.070	0.890 ± 0.021	2.787 ± 0.038	1.436 ± 0.040	0.950 ± 0.032	0.492 ± 0.035
1770	33.431 ± 0.057	12.219 ± 0.010	6.211 ± 0.014	2.589 ± 0.013	12.872 ± 0.048	4.744 ± 0.021	2.057 ± 0.038	1.314 ± 0.006	6.392 ± 0.002	2.704 ± 0.040	1.306 ± 0.065	0.904 ± 0.016	2.831 ± 0.029	1.466 ± 0.034	0.969 ± 0.026	0.506 ± 0.031
1765	33.610 ± 0.092	12.345 ± 0.010	6.282 ± 0.000	2.621 ± 0.004	12.919 ± 0.069	4.806 ± 0.009	2.098 ± 0.030	1.329 ± 0.000	6.448 ± 0.017	2.755 ± 0.031	1.348 ± 0.060	0.918 ± 0.011	2.876 ± 0.019	1.496 ± 0.027	0.988 ± 0.021	0.520 ± 0.028
1760	33.790 ± 0.129	12.472 ± 0.032	6.353 ± 0.016	2.654 ± 0.006	12.966 ± 0.090	4.869 ± 0.004	2.139 ± 0.021	1.344 ± 0.007	6.505 ± 0.032	2.807 ± 0.021	1.392 ± 0.054	0.933 ± 0.005	2.922 ± 0.009	1.527 ± 0.020	1.008 ± 0.015	0.534 ± 0.024

FIGURE 10.5: Maps containing the expected number of background events in the single XGB, SuperPNN and mass bin in the 3μ sample. The yields are extracted from a fit to the outer sidebands assuming the contribution of the combinatorial background only. The maps are shown for the three data-taking years.

Chapter 11

Limit estimate

In this Chapter, the evaluation of the expected upper limit on the branching fraction of the $\tau^+ \to \mu^+\mu^-\mu^+$ decay by means of the CL_s method is presented.

11.1 Extrapolated limit from Run 1

The observed upper limit (UL) obtained in the Run 1 analysis turned out to be slightly smaller than the expected upper limit as, after the unblinding procedure, less background than expected was measured. The expected limit on Run 1 data at $5.0(6.1) \times 10^{-8}$ at 90%(95%) confidence level (CL), can be used to extrapolate the upper limit to the current statistics.

According to Eq. 3.8, the branching fraction (BF) of the $\tau^+ \to \mu^+\mu^-\mu^+$ decay can be written as $\alpha \times N_\tau$, where N_τ is the number of events measured in the signal channel. The normalisation factor α is itself proportional to $1/N_{D_s}$ (see Eq. 3.9), being N_{D_s} the number of events observed in the reference channel $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$. Therefore, the upper limit (UL) on the BF is also proportional to the ratio $N_\tau^{\mathrm{UL}}/N_{D_s}$, where N_τ^{UL} represents the upper limit on the number of observed signal events.

In the hypothesis where no signal event is observed, N_τ^{UL} is equal to the uncertainty on the number of background events expected in the signal region $N_{exp.bkg}$, which can be naively assumed to be $\sqrt{N_{exp.bkg}}$. Both $N_{exp.bkg}$ and N_{D_s} are proportional to $\mathcal{L}_{int} \times \sigma$, being $\mathcal{L}_{int}$ the integrated luminosity and σ the production cross section of the relevant mother particle. Therefore, the expected upper limit can be expressed as

$$\mathrm{UL}_{exp} \propto \frac{\sqrt{\mathcal{L}_{int} \times \sigma_{bkg}}}{\mathcal{L}_{int} \times \sigma_{D_s}} . \tag{11.1}$$

Assuming that the cross section of the combinatorial background scales with the same factor as the cross section of the reference channel, σ_{bkg} can be replaced with σ_{D_s} and therefore have that

$$\mathrm{UL}_{exp} \propto \frac{1}{\sqrt{\mathcal{L}_{int} \times \sigma_{D_s}}} . \tag{11.2}$$

The extrapolated limit for Run 2 can be written in terms of the expected limit for Run 1 as

$$\mathrm{UL}_{extrap}^{\mathrm{Run\,2}} = \mathrm{UL}_{exp}^{\mathrm{Run\,1}} \frac{\sqrt{\mathcal{L}_{int}^{2011} \times \sigma_{D_s}^{2011} + \mathcal{L}_{int}^{2012} \times \sigma_{D_s}^{2012}}}{\sqrt{\mathcal{L}_{int}^{\mathrm{Run\,2}} \times \sigma_{D_s}^{\mathrm{Run\,2}}}} , \tag{11.3}$$

where the integrated luminosity and cross section of the different data taking periods are listed in Table 11.1. D_s mesons can be either produced prompt at the collision point or secondary via b-hadron decays. Therefore, the D_s production cross section

TABLE 11.1: Integrated luminosity and charm production cross section for each data taking period. The charm production cross section for 2012 is obtained by scaling the 2011's for the ratio between the corresponding centre-of-mass energies, *i.e.* 8/7. The values of the integrated luminosity are taken from Ref. [70].

	2011	**2012**	**Run 2**
$\mathcal{L}_{int}$ [fb^{-1}]	1	2	5.4
σ_{D_s} [μb]	1119 ± 155	1279 ± 178	1977 ± 375

is given by the sum of the two contributions.

$$\sigma_{D_s} = \sigma(pp \to D_s^+ X) + \sigma(pp \to b\bar{b}X) \times \mathcal{B}(\bar{b} \to D_s^+) \tag{11.4}$$

For Run 2, the needed quantities are listed in Table 4.5. The measurement of the prompt charm and beauty production cross sections at $\sqrt{s} = 7$ TeV are reported in Ref. [99] and Ref. [100], respectively, while $\mathcal{B}(\bar{b} \to D_s^+)$ is taken from Ref. [3]. The charm production cross section for 2012 is obtained by scaling the 2011's for the ratio between the corresponding centre-of-mass energies, *i.e.* 8/7. The extrapolated upper limit from Run 1 to the Run 2 data set would therefore be

$$\mathrm{UL}_{extrap}^{\mathrm{Run\,2}} = 2.9(3.6) \times 10^{-8} \text{ at } 90\%(95\%) \text{ CL.} \tag{11.5}$$

Analogously, one could evaluate the extrapolated limit from Run 1 to the combined Run 1+Run 2 data set by adapting Eq. 11.3 as

$$\mathrm{UL}_{extrap}^{\mathrm{Run\,1+Run\,2}} = \mathrm{UL}_{exp}^{\mathrm{Run\,1}} \frac{\sqrt{\mathcal{L}_{int}^{2011} \times \sigma_{D_s}^{2011} + \mathcal{L}_{int}^{2012} \times \sigma_{D_s}2012}}{\sqrt{\mathcal{L}_{int}^{2011} \times \sigma_{D_s}^{2011} + \mathcal{L}_{int}^{2012} \times \sigma_{D_s}^{2012} + \mathcal{L}_{int}^{\mathrm{Run\,2}} \times \sigma_{D_s}^{\mathrm{Run\,2}}}} . \tag{11.6}$$

An extrapolated limit of

$$\mathrm{UL}_{extrap}^{\mathrm{Run\,1+Run\,2}} = 2.5(3.1) \times 10^{-8} \text{ at } 90\%(95\%) \text{ CL} \tag{11.7}$$

is found.

11.2 Final fit to the invariant mass

As previously described, the limit evaluation is performed in bins of the output of the two multivariate classifiers trained in the 3μ sample, namely XGB and SuperPNN. Once the analysis is optimised and the signal region unblinded, the number of signal events is measured, in each bin, with an unbinned extended maximum likelihood fit [84] to the invariant mass spectrum. The fit is performed simultaneously in the 16 bins, independently for the three data-taking years. The total pdf in the single bin i is the sum of the pdf used to describe the signal peak and the one used to describe the background shape.

$$\mathrm{pdf}_i^{tot} = n_i^{sig} \cdot \mathrm{pdf}_i^{sig} + n_i^{bkg} \cdot \mathrm{pdf}_i^{bkg} \tag{11.8}$$

As described in Section 7.2.4, the $\tau^+ \to \mu^+\mu^-\mu^+$ signal peak is modelled with a Johnson's S_U function extrapolated from a fit to simulated events. The fit parameters λ, δ and γ of the Johnson's S_U function are in common among the bins. The modelling of the background shape is instead discussed in Section 8.2.3: an exponential distribution is used to describe the combinatorial background, while the residual contribution of the $D^+ \to \pi^-\pi^+\pi^+$ and $D_s^+ \to \pi^-\pi^+\pi^+$ modes is modelled with two Johnson's S_U functions. The correlation of the signal yields among the different bins is taken into account and the number of signal events in the single bin i is evaluated as

$$n_i^{sig} = \epsilon_i N^{sig}\,, \tag{11.9}$$

where ϵ_i represents the XGB and SuperPNN bin efficiency (see Figure 9.5) and N^{sig} the total number of events measured for the signal channel. The sum of the efficiencies ϵ_i is constrained to be equal to the total efficiency $\epsilon_{sig.reg.}$, representing the fraction of events falling in the signal region, as reported in Table 7.6.

$$\epsilon_{sig.reg.} = \sum_i \epsilon_i \tag{11.10}$$

The efficiencies and the fit parameters λ, δ and γ are allowed to fluctuate according to a gaussian distribution whose central value and resolution are obtained from the fit to simulated events. Note that the μ and λ fit parameters are corrected with c_μ and c_λ, respectively, as described in Section 6.4. The number of background events in the different bins is assumed to be uncorrelated. In the 2μ sample, no further binning is considered due to the limited statistics of the simulated sample. Therefore, the fit to the invariant mass spectrum is performed once for each data-taking year.

Before the unblinding of the signal region, only an expected upper limit on the branching fraction of the $\tau^+ \to \mu^+\mu^-\mu^+$ decay can be provided. As shown in Section 8.2.3, a simultaneous fit is performed to the data outer sidebands in the 16 XGB and SuperPNN bins. The number of background events expected in the signal region is extrapolated from the fit. As described in Section 3.5.1, the number of signal events is evaluated, for each year, as the ratio between the BF value that is being tested and the normalisation factor for that year.

11.3 Evaluation of the expected upper limit

The evaluation of the expected upper limit on the BF of $\tau^+ \to \mu^+\mu^-\mu^+$ is performed using the CL_s method, which has been introduced and discussed in Section 3.5. The method is implemented in a C++ code based on the ROOT library and developed for this specific analysis. The following quantities are given as input to the code for each data-taking year.

- Normalisation factor α together with its uncertainty. Final values for α are reported in Table 9.16 for the 3μ and 2μ samples.

- Maps containing the single XGB, SuperPNN and mass bin efficiency for the 3μ sample (see Section 9.6). Single-bin maps containing the fraction of events falling in the signal region (evaluated as integral of the extrapolated pdf) are provided for the 2μ sample (see Table 10.4).

- Maps containing the number of background events expected in each XGB, SuperPNN and mass bin for the 3μ sample (see Figure 10.4). Single-bin maps with the number of background events expected in the signal region are given as input for the 2μ sample (see Table 8.15).

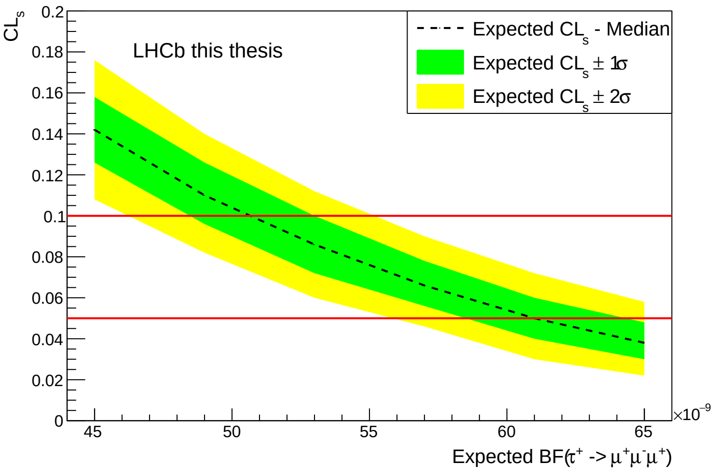


FIGURE 11.1: Expected CL_s curve obtained on Run 1 data under the hypothesis to observe background events only. The yellow and green curves represent the $\pm 1\sigma$ and $\pm 2\sigma$ error bands of the CL_s median, respectively. The upper and lower red lines indicate the 90% and 95% confidence level, respectively.

- For both samples, maps containing the fluctuated values of the single bin efficiency and of the expected number of background events in the signal region, as reported in Section 10.2 and Section 10.3, respectively.

11.3.1 Cross-check with Run 1 data

The correct implementation of the CL_s method is verified by checking if the code developed for this analysis is able to reproduce the Run 1 result for the expected upper limit, namely $5.0(6.1) \times 10^{-8}$ at 90%(95%) CL [3].

Similarly to what is done in the Run 2 analysis, the Run 1 upper limit on the BF of $\tau^+ \to \mu^+\mu^-\mu^+$ was evaluated in bins of the output of two multivariate classifiers, one trained to separate muons from other tracks, and a second one designed to identify displaced three-body tracks. The normalisation factors and the maps containing the single bin efficiency and the expected number of background events in the signal region for the two data-taking years are given as input to the limit code.

The search for the *confidence limit* is performed in the BF range $[4.5, 6.5] \times 10^{-8}$ with steps of 4×10^{-9}. The CL_s values obtained for the different BFs are reported in the plot displayed in Figure 11.1. An expected upper limit on the BF of $5.07(6.10) \times 10^{-8}$ is obtained at 90%(95%) CL. The agreement of this result with the former expected limit confirms the correct implementation of the CL_s method.

As previously mentioned, the aim is to provide a combined limit for the Run 1 and Run 2 data sets. To do that, the normalisation factor evaluated on Run 1 data needs to be adjusted in order to agree with the latest measurements of the external inputs used in its evaluation. According to Eq. 3.9, these are $\mathcal{B}(D_s^+ \to \phi(\mu^+\mu^-)\pi^+)$ and $\mathcal{B}(D_s^+ \to \tau^+\nu_\tau)$. A summary of the old and new values for the involved BFs is reported in Table 11.2. The resulting correction factor $\alpha_{corr}^{\text{Run 1}}$, computed as

$$\alpha_{corr}^{\text{Run 1}} = \frac{\mathcal{B}(D_s^+ \to \phi(\mu^+\mu^-)\pi^+)_{new}}{\mathcal{B}(D_s^+ \to \phi(\mu^+\mu^-)\pi^+)_{old}} \times \frac{\mathcal{B}(D_s^+ \to \tau^+\nu_\tau)_{old}}{\mathcal{B}(D_s^+ \to \tau^+\nu_\tau)_{new}}, \tag{11.11}$$

is 1.0329.

TABLE 11.2: Values of the branching fractions involved in the evaluation of α used in the Run 1 analysis together with their most recent measurements [8].

	Value used in Run 1	**Latest measurement**
$\mathcal{B}(D_s^+ \to \phi(\mu^+\mu^-)\pi^+)$	$(1.317 \pm 0.099) \times 10^{-5}$	$(1.290 \pm 0.094) \times 10^{-5}$
$\mathcal{B}(D_s^+ \to \tau^+\nu_\tau)$	$(5.61 \pm 0.24)\%$	$(5.32 \pm 0.11)\%$

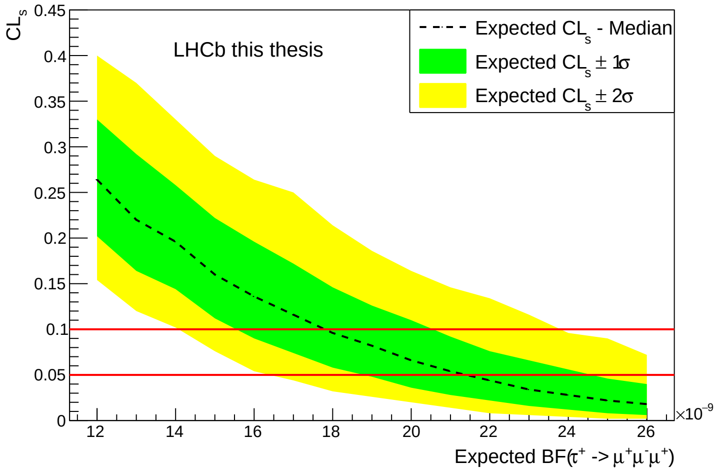


FIGURE 11.2: Expected CL_s curve obtained on Run 2 data (including the 3μ and the 2μ sample) under the hypothesis to observe background events only. The yellow and green curves represent the $\pm 1\sigma$ and $\pm 2\sigma$ error bands of the CL_s median, respectively. The upper and lower red lines indicate the 90% and 95% confidence level, respectively.

It should be noted that this correction is only applied when the Run 1 data set is combined with the Run 2 data set. The fraction of τ produced via D_s decays ($f^{\tau}_{D_s}$) also represents an external input in the evaluation of α, though its value is not affected by the variation of any BF measurement.

11.3.2 Expected upper limit with Run 2 data

The quantities listed at the beginning of Section 11.3 are given as input to the limit code in order to evaluate the expected upper limit on the BF of the $\tau^+ \to \mu^+\mu^-\mu^+$ decay on Run 2 data. The search for the *confidence limit* is performed in the BF range $[1.2, 2.6] \times 10^{-8}$ with steps of 1.0×10^{-9} and the distribution of CL_s as a function of the scanned BF values is reported in Figure 11.2.

An expected upper limit on the BF of $\tau^+ \to \mu^+\mu^-\mu^+$ of

$$\mathcal{B}(\tau^+ \to \mu^+\mu^-\mu^+)_{\text{Run 2}} < 1.8(2.1) \times 10^{-8}$$

is found at 90%(95%) CL. It turns out that the inclusion of the 2μ sample leads to a marginal improvement. As stated in Section 5.2, the fraction of simulated events where only two of the final-state tracks are identified as muons represents 30% of the total number of events surviving the stripping and offline selection. However, the efficiency of the trigger selection applied in the 2μ sample is smaller than the efficiency of the trigger selection applied in the 3μ sample by a factor of ~ 0.3 in

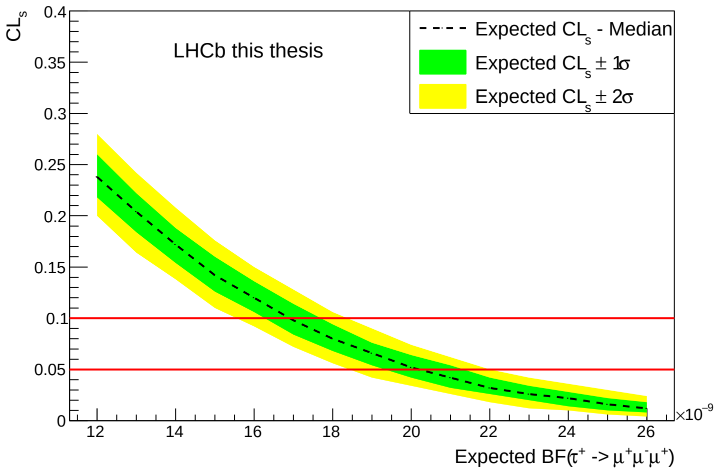


FIGURE 11.3: Expected CL_s curve obtained on Run 1+Run 2 data under the hypothesis to observe background events only. The yellow and green curves represent the $\pm 1\sigma$ and $\pm 2\sigma$ error bands of the CL_s median, respectively. The upper and lower red lines indicate the 90% and 95% confidence level, respectively. No fluctuation on the systematic uncertainties is included.

2017 and 2018, and by a factor of ~ 0.14 in 2016, as reported in Table 9.15. Moreover, the higher background contamination in the 2μ sample requires harder cuts on the discriminating variables (output of the classifiers) exploited in the suppression of the various background sources, leading to a further reduction of the total efficiency.

In Run 3, the removal of the hardware trigger (L0) in favour of a fully software trigger system [94] allows the development of new trigger lines tailored for the selection of the 2μ sample in $\tau^+ \to \mu^+\mu^-\mu^+$ decays. As previously outlined, tracks with `!isMuon` are usually produced at very low transverse momenta; relaxing the cut on the transverse momentum in the trigger selection could potentially lead to higher selection efficiency and, therefore, to a lower *single event sensitivity*.

Overall, the value obtained for the expected limit on the whole Run 2 data set shows an improvement of around 40% with respect to the value for the expected limit extrapolated from the higher luminosity and greater cross section of Run 2 (Eq. 11.5).

The expected CL_s distribution obtained for the combined Run 1+Run 2 data set is shown in Figure 11.3 without including fluctuations on the systematic uncertainties. The search for the *confidence limit* is performed in the same BF range and with the same step employed for the Run 2 limit. An expected upper limit of

$$\mathcal{B}(\tau^+ \to \mu^+\mu^-\mu^+)_{\text{Run 1+Run 2}} < 1.7(2.0) \times 10^{-8}$$

is found at 90%(95%) CL. An improvement on the expected limit of around 7% is achieved when the Run 1 data set is combined with the Run 2 data set.

According to what is presented in Section 11.1, the combination of the Run 1 and Run 2 data sets leads to an improvement on the Run 2 extrapolated limit of around 14%. The reason why adding the Run 1 contribution only brings an improvement of 7% on the expected limit indicates that the significance reached by the Run 2 analysis on the Run 2 data set is higher than the one reached by the Run 1 analysis on the Run 1 data set.

Chapter 12

Conclusions

This thesis presents a search for the lepton flavour violating decay $\tau^+ \to \mu^+\mu^-\mu^+$ on data collected at a centre-of-mass energy of 13 TeV by the LHCb experiment during Run 2, corresponding to an integrated luminosity of 5.4 fb^{-1}. The aim of the analysis is the first observation of the decay or, in the case where no events are observed, to set an upper limit on its branching fraction (BF). Improving the current best experimental limit on the BF of the $\tau^+ \to \mu^+\mu^-\mu^+$ decay (2.1×10^{-8} at 90% CL, Belle experiment [2]) would allow us to constrain theories of physics beyond the Standard Model that predict for this decay branching fractions in reach of the sensitivity of current experimental facilities.

The search for a very rare decay such as the $\tau^+ \to \mu^+\mu^-\mu^+$ decay is of course limited by the number of τ leptons produced in proton-proton collisions. In order to increase the available statistics, events where only two of the final-state tracks are identified as muons are included in the analysis. Two parallel, yet very similar, analyses are performed on this ensemble of events, referred to as the 2μ sample, and on events where all the final-state tracks are identified as muons, called the 3μ sample.

The identification and rejection of the various background sources represent the main challenge of the analysis. Multivariate classification models are trained to suppress background originating from tracks wrongly associated with a τ decay and from the misidentification of hadrons, such as pions or kaons, produced in D^+ and D_s decays. The analysis is designed while blinding the signal region in data in order to avoid biases in the optimisation of the selection. The efficiency of the applied selection is evaluated on simulated events and a fit to the data sidebands is performed to extrapolate the expected number of background events in the signal region. The CL_s method is used to evaluate the limit on the branching fraction of $\tau^+ \to \mu^+\mu^-\mu^+$ in bins of the output of the two multivariate classifiers trained in the 3μ sample and in bins of the invariant mass. The contribution of the 2μ sample for the three years is included without considering a further binning due to the limited statistics of the simulated sample.

At the moment of the writing of this thesis, the signal region is still blinded. Therefore, only an expected value for the upper limit on the branching fraction of $\tau^+ \to \mu^+\mu^-\mu^+$ on the Run 2 data set can be provided. An expected limit of

$$\mathcal{B}(\tau^+ \to \mu^+\mu^-\mu^+) < 1.8(2.1) \times 10^{-8}$$

is determined at 90%(95%) CL. The inclusion of the 2μ sample in the limit evaluation brings a marginal improvement, due to the low efficiency of the trigger selection. An overall improvement of 40% is achieved with respect to the expected limit extrapolated from the higher luminosity and greater cross section of Run 2 with respect to Run 1.

Finally, for the combined Run 1 and Run 2 data sets, corresponding to an integrated luminosity of 8.4 fb^{-1}, an expected upper limit of

$$\mathcal{B}(\tau^+ \to \mu^+\mu^-\mu^+) < 1.7(2.0) \times 10^{-8}$$

is found at 90%(95%) CL. If confirmed on data, this result would improve the current best experimental limit on the BF of the $\tau^+ \to \mu^+\mu^-\mu^+$ decay by around 20%.

Appendix A

Correction of the production fractions

As discussed in detail in Section 4.2.2, the number of simulated events generated for each of the τ and D_s production sources, need to be weighted with w_{fraction} to reproduce the correct relative contributions.

The τ production sources can be grouped into two categories. To the former category belong the so-called *prompt* decays where the τ's mother particle is produced at the collision point; these include the $D_s^+ \to \tau^+$, $D^+ \to \tau^+$ and $\bar{b} \to \tau^+$ decays. The latter category include the so-called *secondary* processes, where the τ's mother particle (D_s or D^+) is produced via b decays.

The reference channel $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ can be used to evaluate the relative fraction of prompt and secondary decays in the MC sample and compare it with the relative fraction observed in *sWeighted* data. The contribution of prompt and secondary decays can be measured by fitting the logarithm of the χ^2 of the mother particle's impact parameter with respect to the primary vertex (`log(Ds_IPCHI2_OWNPV)`), where the two contributions can be easily distinguished. The `log(Ds_IPCHI2_OWNPV)` distribution is fitted with a total pdf defined as

$$\text{pdf}_{tot} = frac_{prompt} \cdot \text{pdf}_{prompt} + (1 - frac_{prompt}) \cdot \text{pdf}_{sec.} \tag{A.1}$$

and the result is shown for simulated and data events on the left and on the right-hand side of Figure A.1, respectively, separately for the three data-taking years. The shape parameters of pdf_{prompt} and $\text{pdf}_{sec.}$, both described by a Johnson's S_U function (see Eq. 6.1), are fixed by first fitting separately the prompt and secondary contributions in simulation.

The only free parameter in pdf_{tot} is therefore $frac_{prompt}$ and its value is reported for the fit to the data and MC samples, and for the three years, in Table A.1. The last raw of Table A.1 shows, for the three years, the value of $frac_{prompt}$ obtained as

$$frac_{prompt,w_{\text{fr}}} = \frac{N_{prompt,w_{\text{fr}}}}{N_{prompt,w_{\text{fr}}} + N_{sec.,w_{\text{fr}}}}, \tag{A.2}$$

TABLE A.1: Values of $frac_{prompt}$ obtained from the fit to MC and data events and the one computed from w_{fraction}.

$frac_{prompt}$	**2016 [%]**	**2017 [%]**	**2018 [%]**
Data	72.11 ± 0.28	80.38 ± 0.17	81.71 ± 0.15
MC	66.27 ± 0.37	72.09 ± 0.24	72.42 ± 0.28
w_{fraction}	65.87 ± 0.31	71.88 ± 0.19	72.21 ± 0.22

where $N_{prompt,w_{\mathrm{fr}}}$ and $N_{sec.,w_{\mathrm{fr}}}$ are the number of prompt and secondary events weighted with w_{fraction} that passed the trigger selection. The values of $frac_{prompt}$ obtained from the fit to the MC distribution are in agreement with the ones evaluated using w_{fraction}, suggesting the validity of the *MC mixing method* (4.2.2), yet they differ from the ones obtained from the fit to the data distribution of `log(Ds_IPCHI2_OWNPV)`. For this reason, w_{fraction} is weighted with the ratio between $frac_{prompt,Data}$ and $frac_{prompt,MC}$, referred to as w_{fracCorr}, to correct for the disagreement between data and simulation in the relative contribution of prompt and secondary decays.

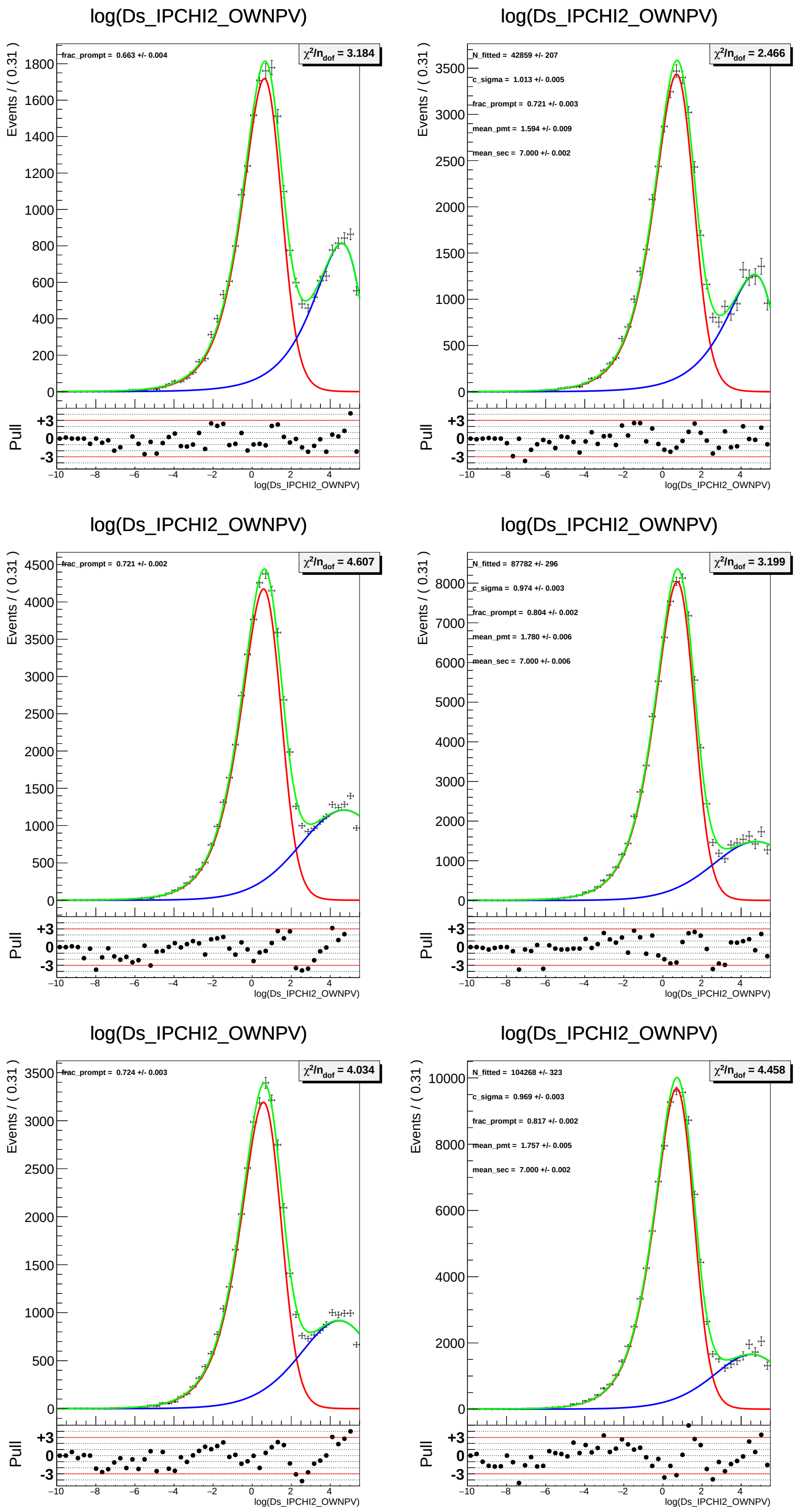


FIGURE A.1: Fit to the D_s `log(Ds_IPCHI2_OWNPV)` for simulated events (left) and *sWeighted* data (right) for 2016 (top), 2017 (middle) and 2018 (bottom). The prompt (red) and secondary (blue) contributions are both fitted with a Johnson's S_U function.

Appendix B

BDTG features

In the following, distributions of the features of the BDTGKT and BDTGPID classifiers for 2016 and 2017 data. Each plot shows the comparison between inner sideband data and simulated events in the signal channel, or between *sWeighted* data and simulated events in the reference channel.

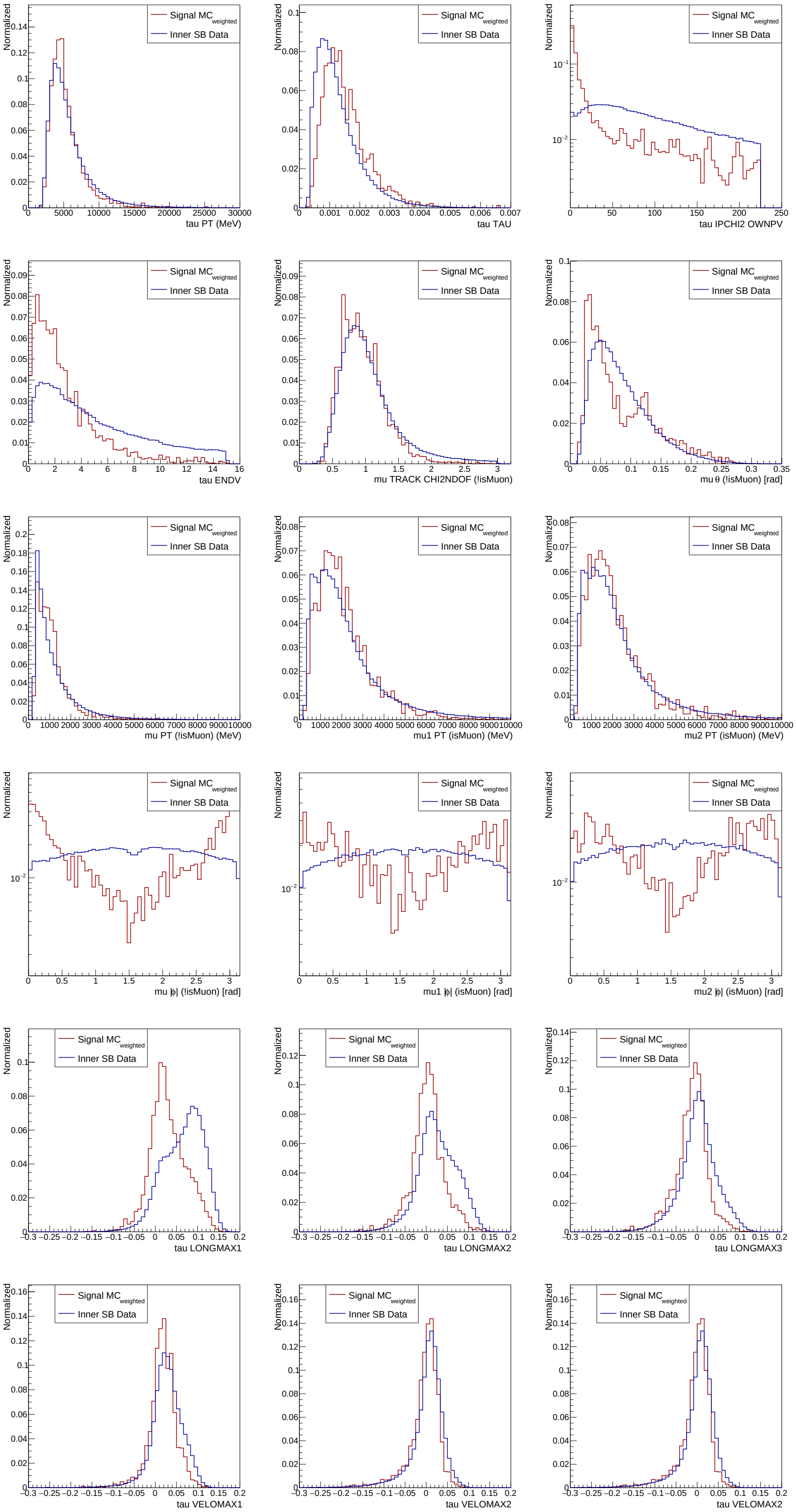


FIGURE B.1: Distributions of the input features of the BDTGKT classifier for 2016 $\tau^+ \to \mu^+\mu^-\mu^+$ simulated events and inner sidebands data. The histograms are normalised to unit area.

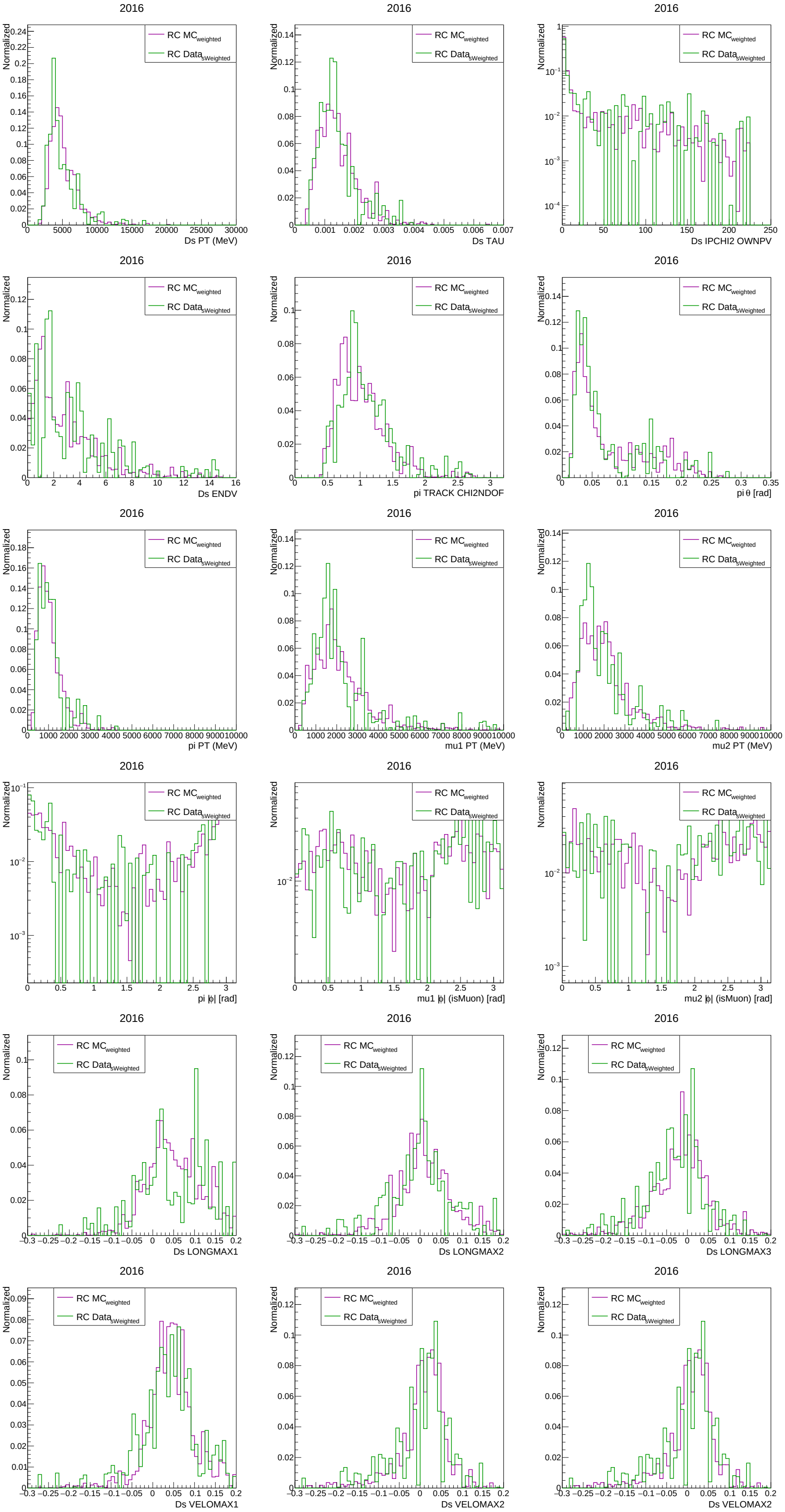


FIGURE B.2: Distributions of the input features of the BDTGKT classifier for 2016 $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ simulated events and *sWeighted* data. The histograms are normalised to unit area.

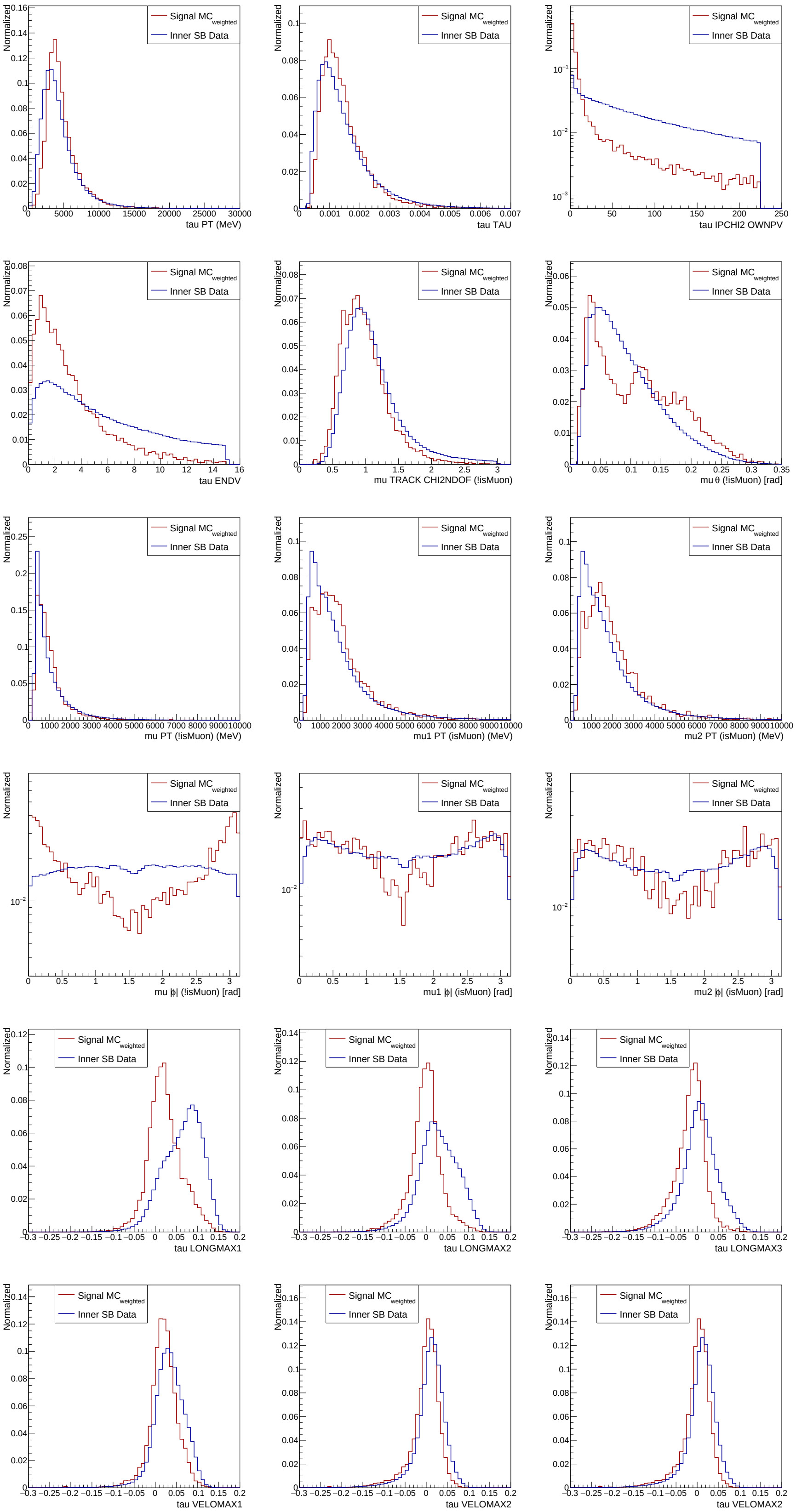


FIGURE B.3: Distributions of the input features of the BDTGKT classifier for 2017 $\tau^+ \to \mu^+\mu^-\mu^+$ simulated events and inner sidebands data. The histograms are normalised to unit area.

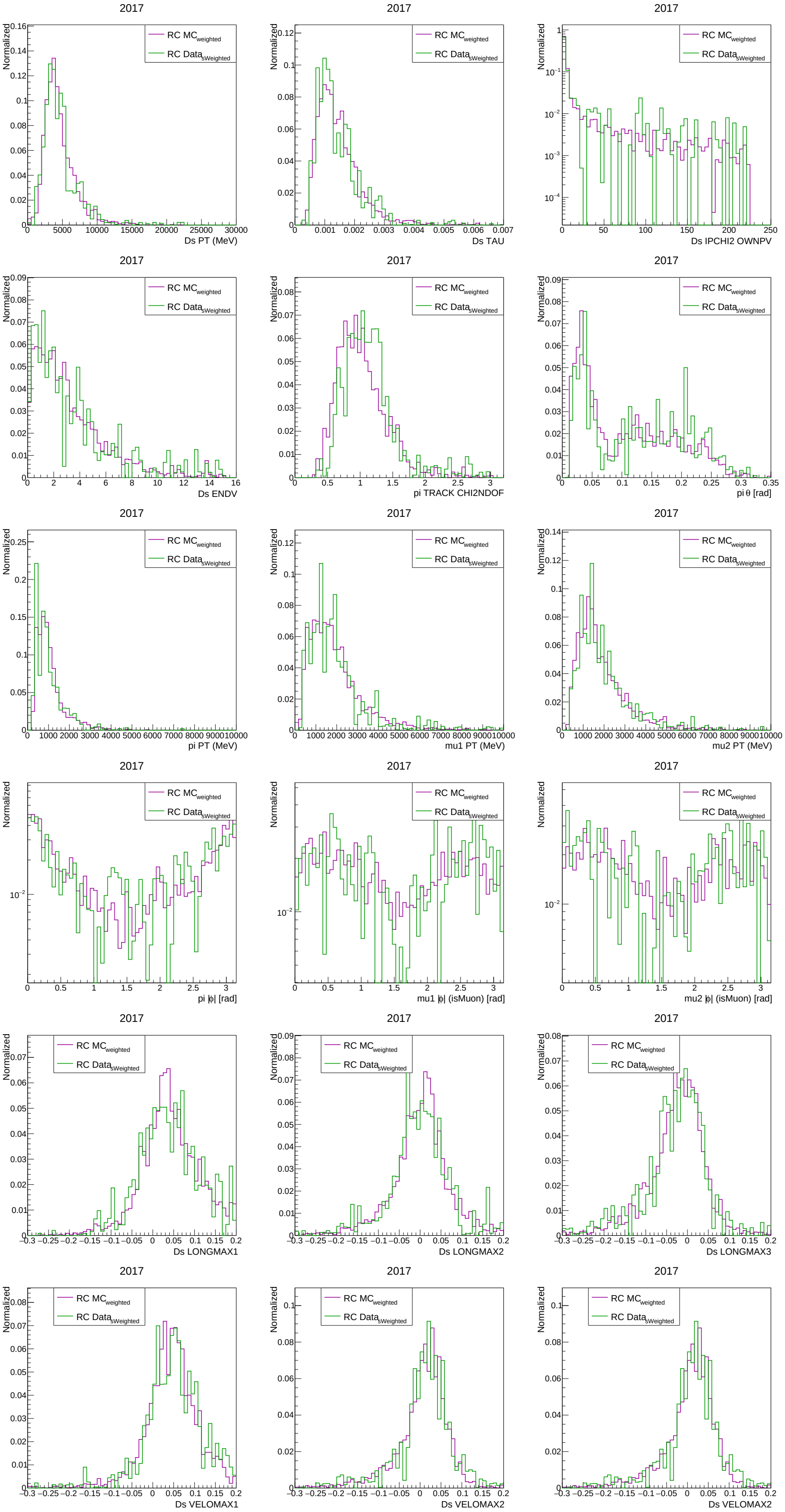


FIGURE B.4: Distributions of the input features of the BDTGKT classifier for 2017 $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ simulated events and *sWeighted* data. The histograms are normalised to unit area.

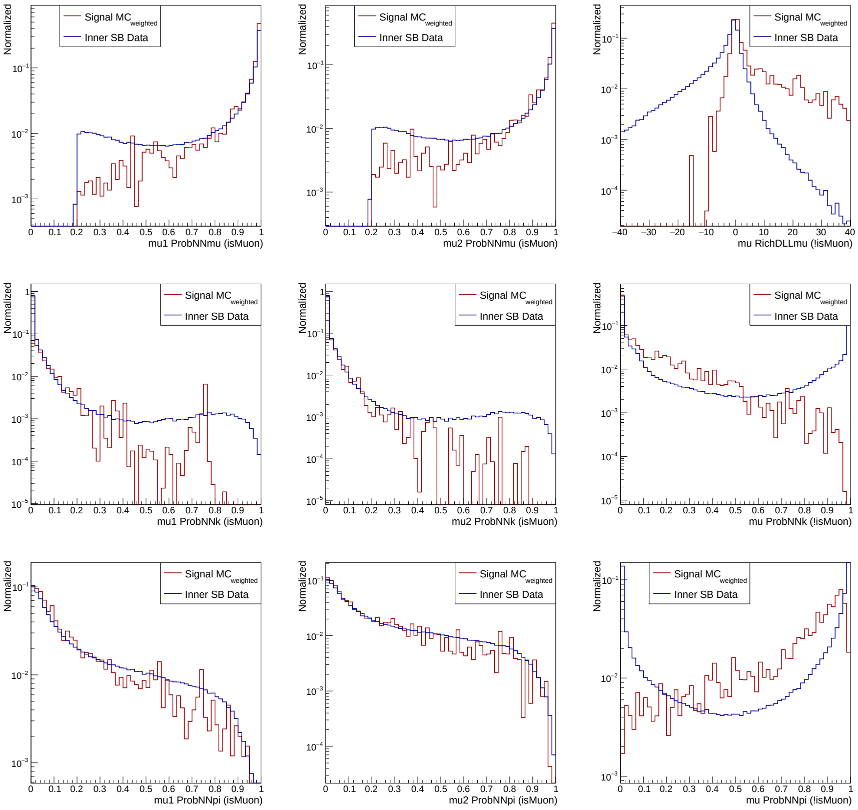


FIGURE B.5: Distributions of the input features of the BDTGPID classifier for 2016 $\tau^+ \to \mu^+\mu^-\mu^+$ simulated events and inner sidebands data. The histograms are normalised to unit area.

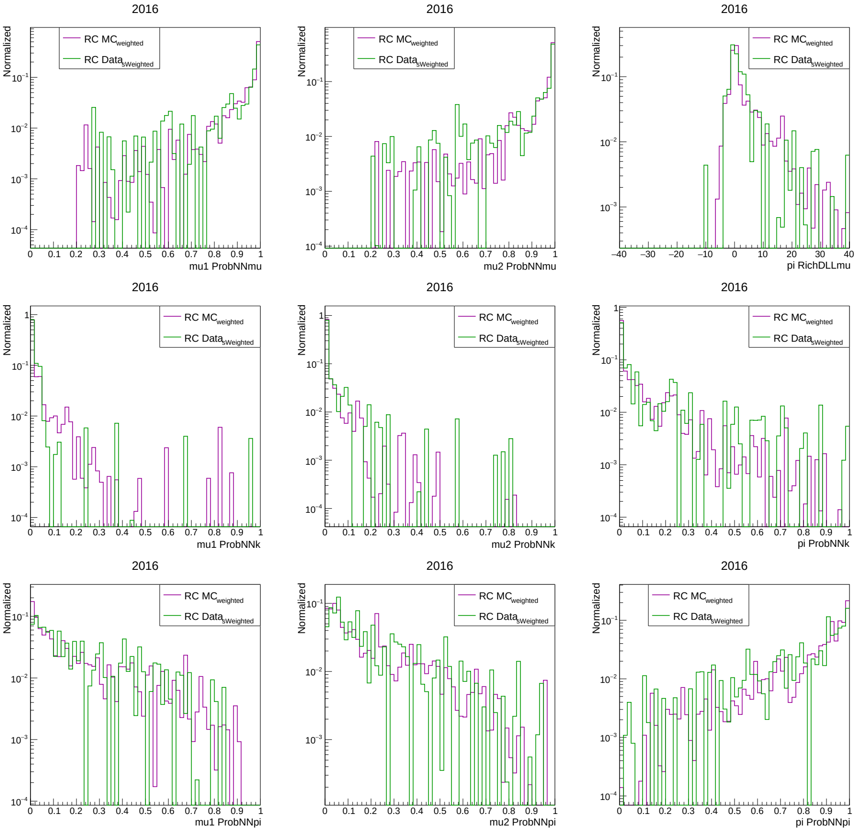


FIGURE B.6: Distributions of the input features of the BDTGPID classifier for 2016 $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ simulated events and *sWeighted* data. The histograms are normalised to unit area.

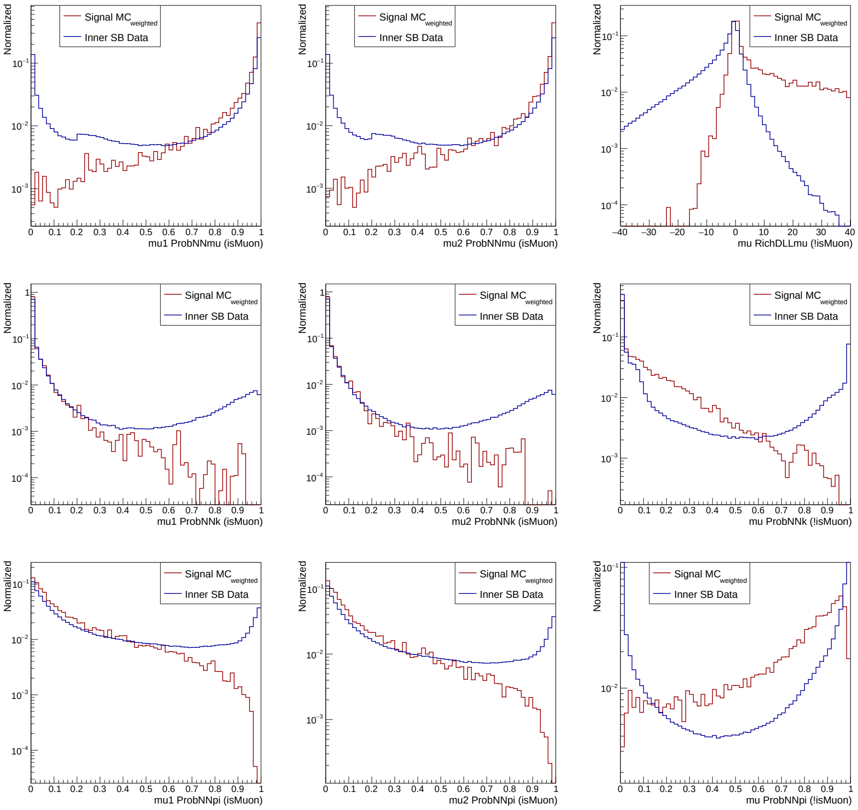


FIGURE B.7: Distributions of the input features of the BDTGPID classifier for 2017 $\tau^+ \to \mu^+\mu^-\mu^+$ simulated events and inner sidebands data. The histograms are normalised to unit area.

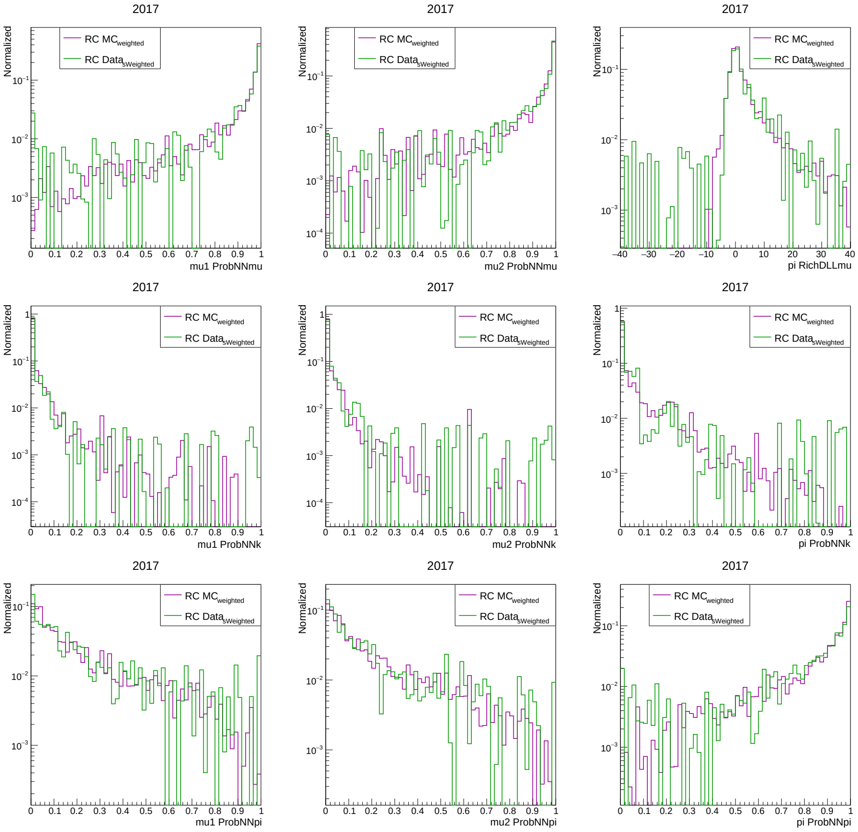


FIGURE B.8: Distributions of the input features of the BDTGPID classifier for 2017 $D_s^+ \to \phi(\mu^+\mu^-)\pi^+$ simulated events and *sWeighted* data. The histograms are normalised to unit area-

Appendix C

Fit to data sidebands

In the following, results of the simultaneous fit performed in the 16 XGB and SuperPNN bins to the data outer sidebands assuming the contribution of the combinatorial background and of the $D^+ \to \pi^-\pi^+\pi^+$ and $D_s^+ \to \pi^-\pi^+\pi^+$ modes. Plots are displayed in Figure C.1 for 2016 data and in Figure C.2 for 2017 data.

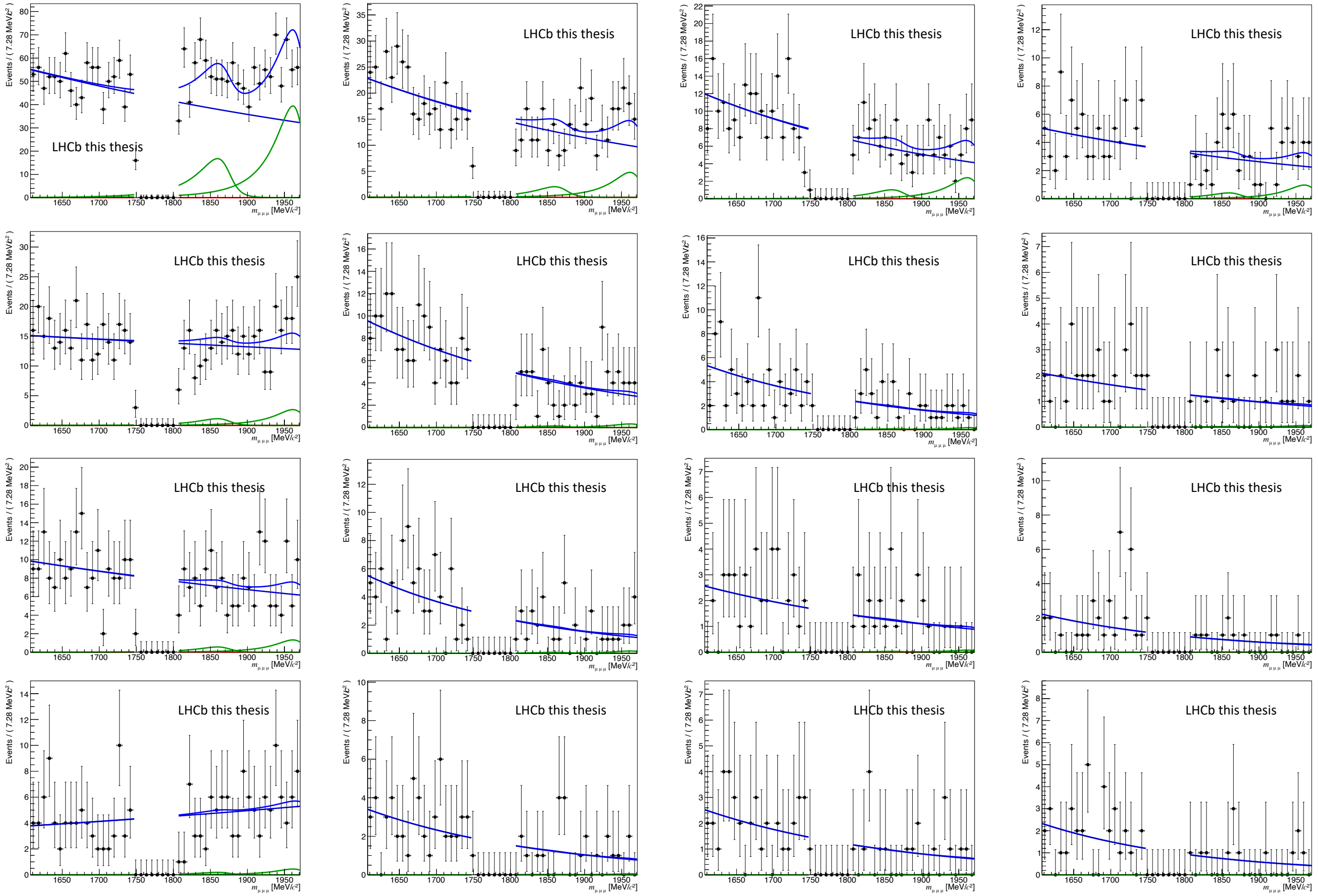


FIGURE C.1: Results of the simultaneous fit performed in the 16 XGB and SuperPNN bins on 2016 data outer sidebands in the 3μ sample assuming the contribution of the combinatorial background (blue line) and of the $D^+ \to \pi^-\pi^+\pi^+$ and $D_s^+ \to \pi^-\pi^+\pi^+$ modes (green lines). Plots are displayed following the same layout of the maps in Figure 8.14.

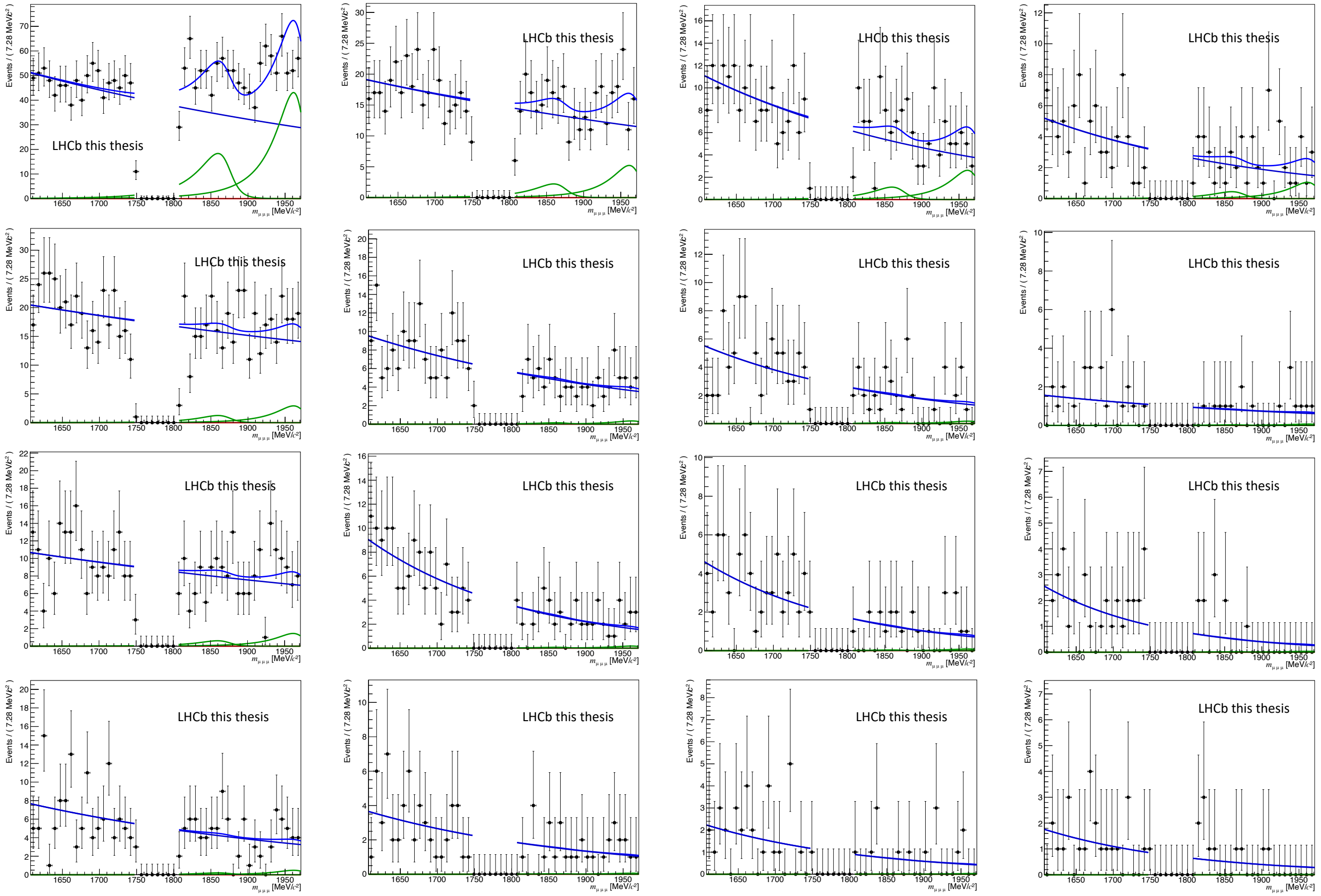


FIGURE C.2: Results of the simultaneous fit performed in the 16 XGB and SuperPNN bins on 2017 data outer sidebands in the 3μ sample assuming the contribution of the combinatorial background (blue line) and of the $D^+ \to \pi^-\pi^+\pi^+$ and $D_s^+ \to \pi^-\pi^+\pi^+$ modes (green lines). Plots are displayed following the same layout of the maps in Figure 8.14.

The simultaneous fit performed in the 16 XGB and SuperPNN bins is validated by means of a set of 200 pseudo-experiments, as described in Section 8.2.3. The pull distributions for the combinatorial background shape parameters and yields are here reported for the three data-taking years.

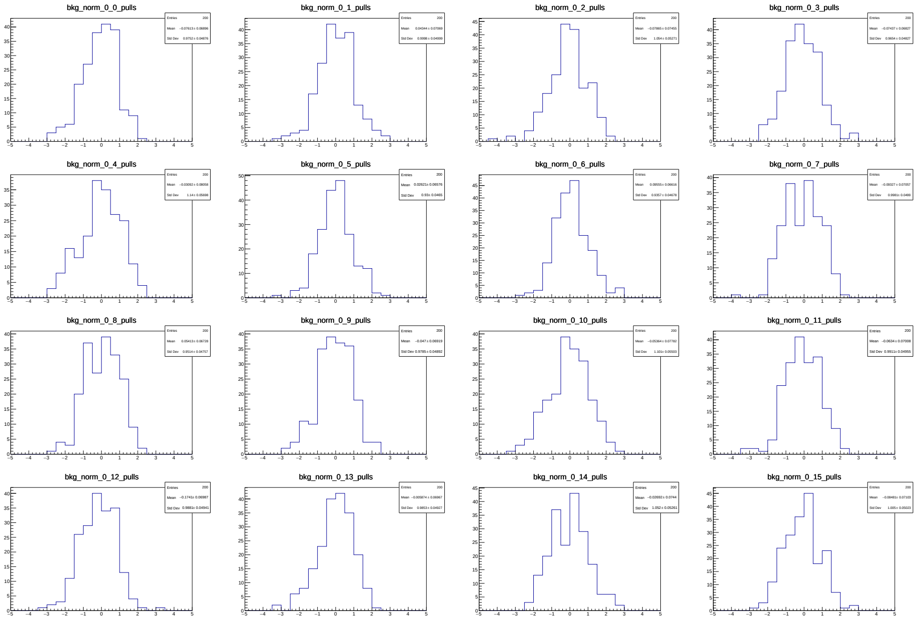


FIGURE C.3: Pull distribution for the normalisation parameters obtained in the 3μ sample in each XGB and SuperPNN bin for 2016 data.

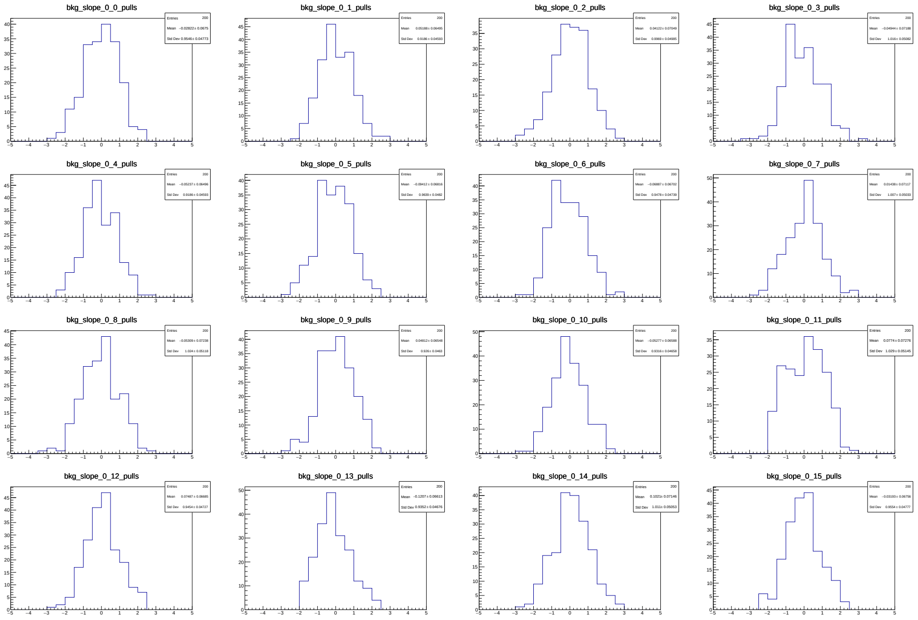


FIGURE C.4: Pull distribution for the combinatorial background shape obtained in the 3μ sample in each XGB and SuperPNN bin for 2016 data.

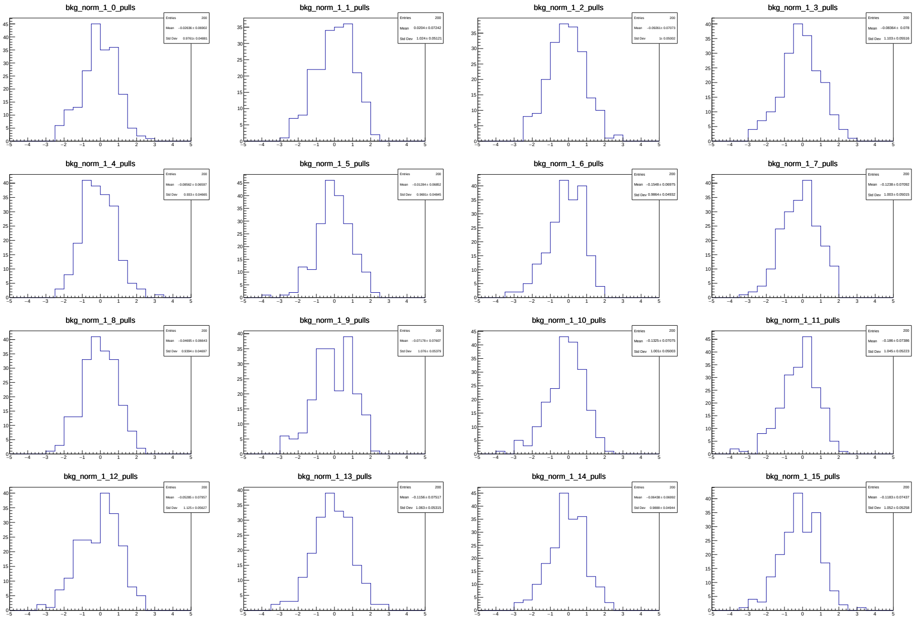


FIGURE C.5: Pull distribution for the normalisation parameters obtained in the 3μ sample in each XGB and SuperPNN bin for 2017 data.

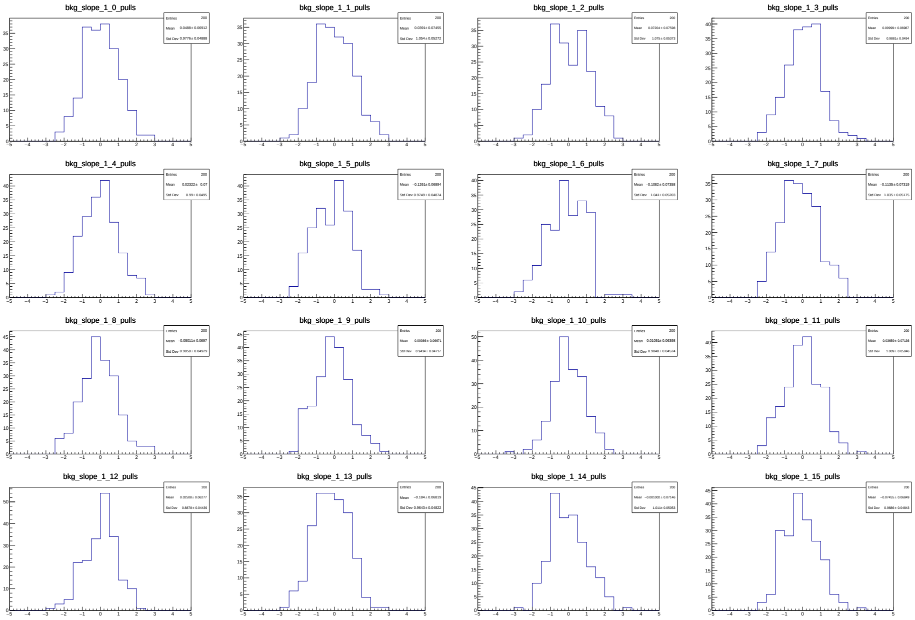


FIGURE C.6: Pull distribution for the combinatorial background shape obtained in the 3μ sample in each XGB and SuperPNN bin for 2017 data.

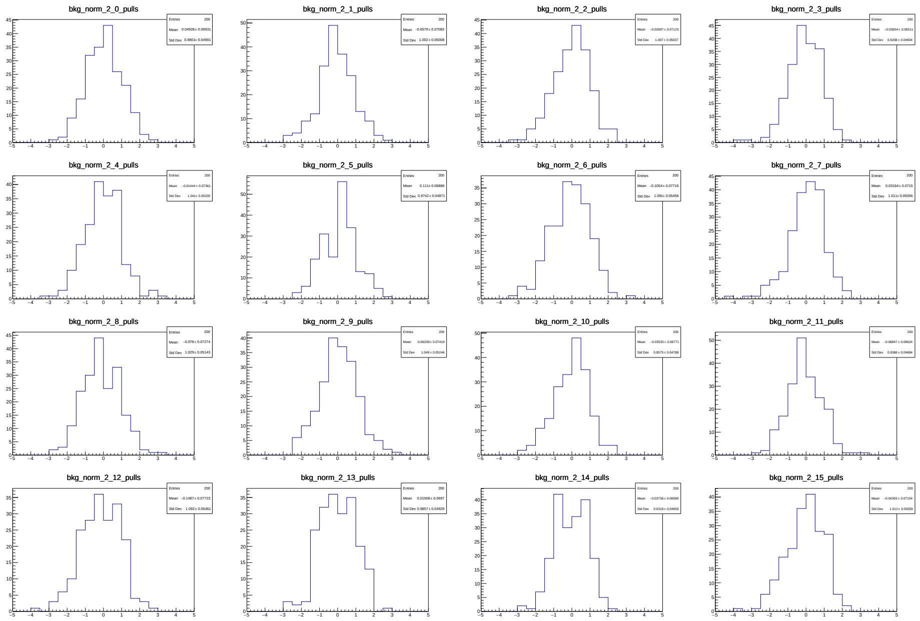


FIGURE C.7: Pull distribution for the normalisation parameters obtained in the 3μ sample in each XGB and SuperPNN bin for 2018 data.

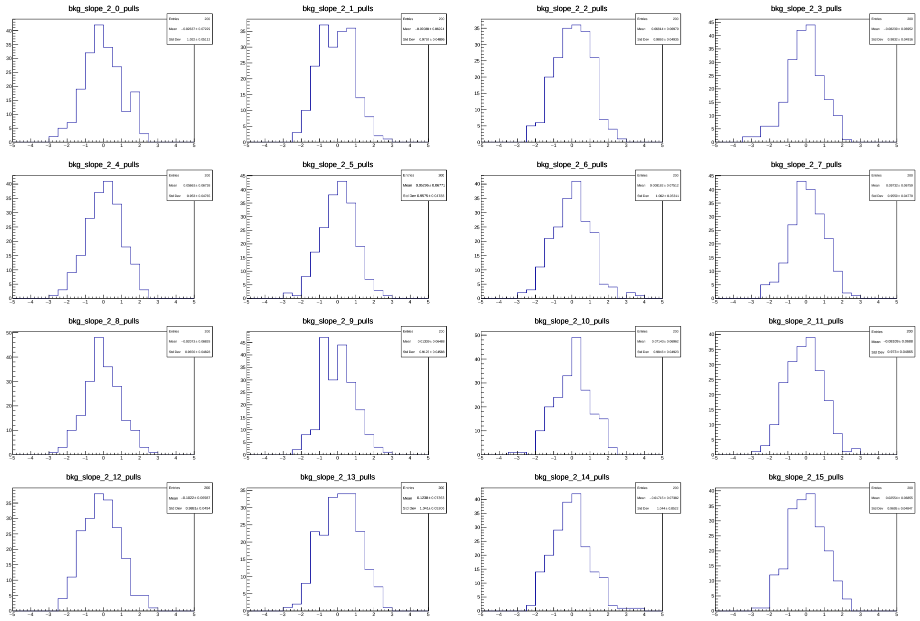


FIGURE C.8: Pull distribution for the combinatorial background shape obtained in the 3μ sample in each XGB and SuperPNN bin for 2018 data.

Appendix D

Model dependence

The simulation of $\tau^+ \to \mu^+\mu^-\mu^+$ decays used in this analysis has been generated assuming a flat phase space distribution. This is of course a simplistic assumption, as the amplitude structure of the decay may depend on various beyond the Standard Model effects. The reconstruction and selection efficiency is not uniform along the phase space, meaning that the global efficiency depends on the phase space distribution of the signal decay. This effect is taken into account by considering various amplitude models allowing for cLFV and reweighting the MC sample accordingly to study the variation of the overall efficiency.

As described in Ref. [101], the alternative models are based on effective-field-theory methods where four chirality operators of dim=6 are combined together with two radiative operators of dim=6 to produce the decay amplitudes. The chirality operators only contain helicity conserving currents and are defined as

$$
\begin{aligned}
O_1 &= \left(\overline{L}\gamma_\mu L\right)\left(\overline{L}\gamma^\mu L\right)\\
O_2 &= \left(\overline{L}\tau^\alpha\gamma_\mu L\right)\left(\overline{L}\tau^\alpha\gamma^\mu L\right)\\
O_3 &= \left(\overline{R}\gamma_\mu R\right)\left(\overline{R}\gamma^\mu R\right)\\
O_4 &= \left(\overline{R}\gamma_\mu R\right)\left(\overline{L}\gamma^\mu L\right)
\end{aligned}
\tag{D.1}
$$

with

$$
L = \begin{pmatrix} \nu_L \\ l_L \end{pmatrix} \quad R = \begin{pmatrix} 0 \\ l_R \end{pmatrix}. \tag{D.2}
$$

The radiative operators are

$$
\begin{aligned}
R_1 &= g'\left(\overline{L}H\sigma_{\mu\nu}R\right)B^{\mu\nu}\\
R_2 &= g(\overline{L}\tau^\alpha H\sigma_{\mu\nu}R)W^{\mu\nu,\alpha}.
\end{aligned}
\tag{D.3}
$$

with

$$
H = \frac{1}{\sqrt{2}}\begin{pmatrix} v + h_0 + i\chi_0 & \sqrt{2}\phi_+ \\ -\sqrt{2}\phi_- & v + h_0 - i\chi_0 \end{pmatrix}. \tag{D.4}
$$

The operators above are combined to obtain the following amplitude structures for the various processes.

- 4-fermion $LL \to LL$ (same for $RR \to RR$)

$$
\frac{\mathrm{d}\Gamma_V^{(LL)(LL)}}{\mathrm{d}m_{23}^2\mathrm{d}m_{12}^2} = \frac{|g_V^{(L_\mu L^\tau)(L_\mu L^\mu)}|^2}{\Lambda^4}\frac{(m_\tau^2 - m_\mu^2)^2 - (2m_{12}^2 - m_\tau^2 - 3m_\mu^2)^2}{256\pi^3 m_\tau^3} \tag{D.5}
$$

- 4-fermion $LL \to RR$ (same for $RR \to LL$ with $g_V^{(L_\mu L^\tau)(R_\mu R^\mu)} \to g_V^{(R_\mu R^\tau)(L_\mu L^\mu)}$)

$$\frac{\mathrm{d}\Gamma_V^{(LL)(RR)}}{\mathrm{d}m_{23}^2\mathrm{d}m_{12}^2} = \frac{|g_V^{(L_\mu L^\tau)(R_\mu R^\mu)}|^2}{\Lambda^4}\left[\frac{(m_\tau^2 - m_\mu^2)^2 - 4m_\mu^2(m_\tau^2 + m_\mu^2 - m_{12}^2)^2}{512\pi^3 m_\tau^3}\right.$$
$$\left. - \frac{(2m_{13}^2 - m_\tau^2 - 3m_\mu^2)^2 + (2m_{23}^2 - m_\tau^2 - 3m_\mu^2)^2}{1024\pi^3 m_\tau^3}\right] \tag{D.6}$$

- Radiative

$$\frac{\mathrm{d}\Gamma_{\mathrm{rad}}^{(LR)}}{\mathrm{d}m_{23}^2\mathrm{d}m_{12}^2} = \alpha_{\mathrm{em}}^2\frac{|g_{\mathrm{rad}}^{(L_\mu R^\tau)}|^2}{\Lambda^4}\left[\frac{m_\mu^2(m_\tau^2 - m_\mu^2)^2}{128\pi^3 m_\tau^3}\left(\frac{1}{m_{13}^4} + \frac{1}{m_{23}^4}\right)\right.$$
$$+ \frac{m_\mu^2(m_\tau^4 - 3m_\tau^2 m_\mu^2 + 2m_\mu^4)^2}{128\pi^3 m_\tau^3 m_{13}^2 m_{23}^2} + \frac{(2m_{12}^2 - 3m_\mu^2)^2}{128\pi^3 m_\tau^3}$$
$$\left. + \frac{(m_{13}^2 + m_{23}^2)\left(m_{12}^4 + m_{13}^4 + m_{23}^4 - 6m_\mu^2(m_\tau^2 + m_\mu^2)\right)}{256\pi^3 m_\tau^3 m_{13}^2 m_{23}^2}\right] \tag{D.7}$$

- Interference of Radiative and 4-fermion $LL \to LL$ (same for $RR \to RR$)

$$\frac{\mathrm{d}\Gamma_{\mathrm{mix}}^{(LL)(LL)}}{\mathrm{d}m_{23}^2\mathrm{d}m_{12}^2} = \alpha_{\mathrm{em}}^2\frac{2v\mathrm{Re}(g_V^{(L_\mu L^\tau)(L_\mu L^\mu)}g_{\mathrm{rad}}^{*(L_\mu R^\tau)})}{\Lambda^4}\times$$
$$\left[\frac{m_{12}^2 - 3m_\mu^2}{64\pi^3 m_\tau^2} + \frac{m_\mu^2(m_\tau^2 - m_\mu^2)(m_{13}^2 + m_{23}^2)}{128\pi^3 m_\tau^2 m_{13}^2 m_{23}^2}\right] \tag{D.8}$$

- Interference of Radiative and 4-fermion $LL \to RR$ (same for $RR \to LL$)

$$\frac{\mathrm{d}\Gamma_{\mathrm{mix}}^{(LL)(RR)}}{\mathrm{d}m_{23}^2\mathrm{d}m_{12}^2} = \alpha_{\mathrm{em}}^2\frac{2v\mathrm{Re}(g_V^{(L_\mu L^\tau)(R_\mu R^\mu)}g_{\mathrm{rad}}^{*(L_\mu R^\tau)})}{\Lambda^4}\times$$
$$\left[\frac{m_\tau^2 - m_{12}^2 - 3m_\mu^2}{256\pi^3 m_\tau^2} + \frac{m_\mu^2(m_\tau^2 - m_\mu^2)(m_{13}^2 + m_{23}^2)}{256\pi^3 m_\tau^2 m_{13}^2 m_{23}^2}\right] \tag{D.9}$$

The relevant distributions of the events in the phase space are shown in Figure D.1 for the different amplitude models and are normalised in such a way that the evaluation of the weights used to redistribute the events only depends on the phase-space.

For each amplitude model, the weights are evaluated as ratio between the Dalitz plot of the relevant model and the one of the phase-space-generated events. To avoid border effects, squared Dalitz plots are considered. The obtained weights distributions are shown in Figure D.2 for each amplitude model.

Once the weights distributions are determined, the effect of the reweighting is calculated by integrating them after the selection cuts on the output of the multivariate classifiers (XGB>0.8 and SuperPNN>0.88) and after correcting for data/MC differences and particle identification. The ratios of the integrals for each amplitude model with respect to the phase-space are shown in Table D.1, where the last column (Total) merges the results of each simulation sample after correcting for the fraction of the relevant production process (as taken from Table 4.8).

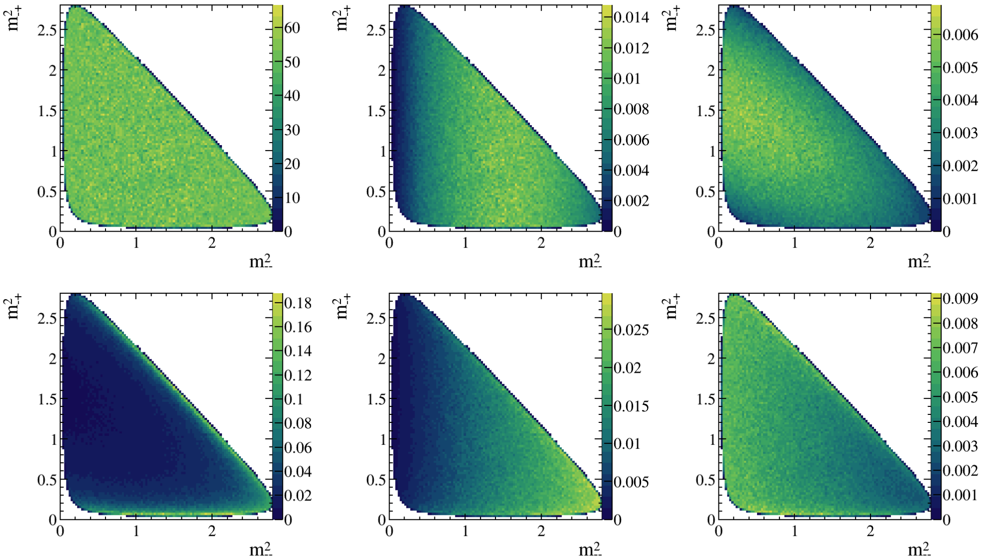


FIGURE D.1: Dalitz plots of the distribution of the events generated in the phase space (top-left) and reweighted by the amplitude of the processes calculated with effective-field theory: 4-fermion $LL \to LL$ (top-centre), 4-fermion $LL \to RR$ (top-right), radiative (bottom-left), interference of radiative and 4-fermion $LL \to LL$ (bottom-centre), interference of radiative and 4-fermion $LL \to RR$ (bottom-right). The Dalitz-plot variables are $m^2_{--} = m^2(\mu^-\mu^-)$ and $m^2_{-+} = m^2(\mu^-\mu^+)$.

TABLE D.1: Ratio of the sum of weights between the various amplitude models and the phase space in the simulated samples after applying the selection. The final column shows the combination of the results of the simulated samples after correcting for the fraction of the relevant production process.

Model	$D^+ \to \tau^+\nu_\tau$		$D_s^+ \to \tau^+\nu_\tau$		$B^+ \to \tau^+ X$	Total
	from B^+	Prompt	from B^+	Prompt		
4-fermion $LL \to LL$	1.09	1.11	1.08	1.10	1.09	1.10
4-fermion $LL \to RR$	1.05	1.02	1.04	1.04	1.04	1.04
Radiative	0.80	0.86	0.79	0.81	0.79	0.81
Rad+4f $LL \to LL$	1.02	1.06	1.01	1.04	1.02	1.04
Rad+4f $LL \to RR$	0.95	0.94	0.96	0.94	0.95	0.95

The assumption on the amplitude model introduces an efficiency variation ranging from −20% in the case of the purely radiative process to +10% for the 4-fermion $LL \to LL$. This source of systematics will be taken into account in the evaluation of the observed upper limit.

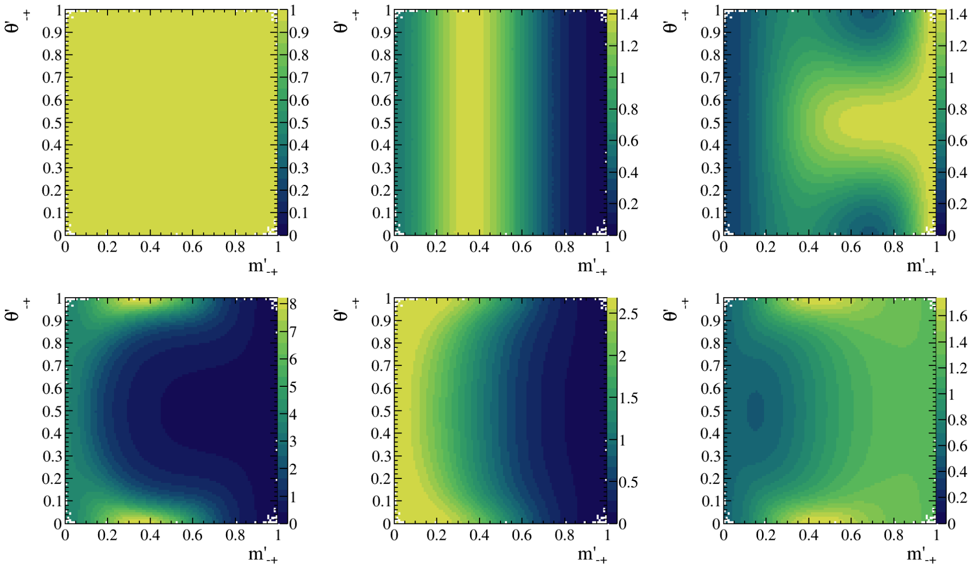


FIGURE D.2: Squared Dalitz plots showing the distribution of the weights obtained by calculating the bin-by-bin ratio of the normalised amplitudes calculated with effective-field theory and the phase space. They are shown as follows: the phase space as a cross-check (top-left), 4-fermion $LL \to LL$ (top-centre), 4-fermion $LL \to RR$ (top-right), radiative (bottom-left), interference of radiative and 4-fermion $LL \to LL$ (bottom-centre), interference of radiative and 4-fermion $LL \to RR$ (bottom-right). The squared Dalitz plot variables are $m'_{-+} = \frac{1}{\pi} \arccos\left(2\frac{m(\mu^-\mu^+) - m(\mu^-\mu^+)_{\text{min}}}{m(\mu^-\mu^+)_{\text{max}} - m(\mu^-\mu^+)_{\text{min}}} - 1\right)$ and $\theta'_{-+} = \frac{1}{\pi}\theta_{\mu^-\mu^+}$, where $\theta_{\mu^-\mu^+}$ is the helicity angle of the $\mu^-\mu^+$ system, namely the angle between the μ^+ momentum in the $\mu^-\mu^+$ rest frame and the opposite of the momentum of the $\mu^-\mu^+$ system in the τ^+ rest frame.

Bibliography

[1] Yoshiyuki Fukuda et al. "Evidence for oscillation of atmospheric neutrinos". In: *Physical review letters* (1998). URL: https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.81.1562.

[2] K Hayasaka et al. "Search for lepton-flavor-violating τ decays into three leptons with 719 million produced $\tau^+\tau^-$ pairs". In: *Physics Letters B* (2010). DOI: 10.1016/j.physletb.2010.03.037.

[3] LHCb Collaboration. "Search for the lepton flavour violating decay $\tau^- \to \mu^-\mu^+\mu^-$". In: *Journal of High Energy Physics* (2015). DOI: https://doi.org/10.1007/JHEP02(2015)121.

[4] Plamen Hopchev. "SciFi: A large Scintillating Fibre Tracker for LHCb". In: *5th Large Hadron Collider Physics Conference*. 2017. arXiv: 1710.08325 [physics.ins-det].

[5] Mark Thomson. *Modern Particle Physics*. Cambridge University Press, 2013. DOI: 10.1017/CBO9781139525367.

[6] Mary K. Gaillard, Paul D. Grannis, and Frank J. Sciulli. "The standard model of particle physics". In: *Rev. Mod. Phys.* (1999). DOI: 10.1103/RevModPhys.71.S96. URL: https://link.aps.org/doi/10.1103/RevModPhys.71.S96.

[7] Steven Weinberg. "A Model of Leptons". In: *Physical Review Letters* 19 (1967). DOI: 10.1103/PhysRevLett.19.1264. URL: https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.19.1264.

[8] R. L. Workman and Others. "Review of Particle Physics". In: *PTEP* (2022). DOI: 10.1093/ptep/ptac097.

[9] Yvette Kosmann-Schwarzbach, Bertram E. Schwarzbach, and Yvette Kosmann-Schwarzbach. *The Noether Theorems*. Springer, 2011. DOI: 10.1007/978-0-387-87868-3_3. URL: https://doi.org/10.1007/978-0-387-87868-3_3.

[10] Julian Heeck. "Interpretation of lepton flavor violation". In: *Physical Review D* (2017). DOI: 10.1103/physrevd.95.015022. URL: https://doi.org/10.1103\%2Fphysrevd.95.015022.

[11] A. M. Baldini et al. "Search for the lepton flavour violating decay $\mu^+ \to e^+\gamma$ with the full dataset of the MEG experiment". In: *Eur. Phys. J. C* (). DOI: 10.1140/epjc/s10052-016-4271-x. arXiv: 1605.05081 [hep-ex].

[12] U. Bellgardt et al. "Search for the Decay $\mu^+ \to e^+e^-e^+$". In: *Nuclear Physics B* (1988). DOI: https://doi.org/10.1016/0550-3213(88)90462-2. URL: https://www.sciencedirect.com/science/article/pii/0550321388904622.

[13] Niklaus Berger. "The Mu3e Experiment". In: *Nuclear Physics B - Proceedings Supplements* (2014). DOI: https://doi.org/10.1016/j.nuclphysbps.2014.02.007. URL: https://www.sciencedirect.com/science/article/pii/S0920563214000085.

[14] Paul Seyfert. "The search for $\tau \to \mu\mu\mu$ at LHCb". 2015. URL: https://cds.cern.ch/record/2002363.

[15] Xuan-Yem Pham. “Lepton flavor changing in neutrinoless τ decays”. In: *The European Physical Journal C* (1999). DOI: `10.1007/s100529901088`. URL: `https://doi.org/10.1007\%2Fs100529901088`.

[16] S. L. Glashow, J. Iliopoulos, and L. Maiani. “Weak Interactions with Lepton-Hadron Symmetry”. In: *Phys. Rev. D* (1970). DOI: `10.1103/PhysRevD.2.1285`. URL: `https://link.aps.org/doi/10.1103/PhysRevD.2.1285`.

[17] S. T. Petcov. “The Processes $\mu \to e + \gamma, \mu \to e + \bar{e}, \nu' \to \nu + \gamma$ in the Weinberg-Salam Model with Neutrino Mixing”. In: *Sov. J. Nucl. Phys.* (1977). [Erratum: Sov.J.Nucl.Phys. 25, 698 (1977), Erratum: Yad.Fiz. 25, 1336 (1977)].

[18] Patrick Blackstone, Matteo Fael, and Emilie Passemar. “$\tau \to \mu\mu\mu$ at a rate of one out of 10^{14} tau decays?” In: *The European Physical Journal C* (2020). DOI: `10.1140/epjc/s10052-020-8059-7`.

[19] G Cvetič et al. “Lepton flavor violation in tau decays”. In: *Physical Review D* (2002).

[20] A. G. Akeroyd, Mayumi Aoki, and Yasuhiro Okada. “Lepton Flavor Violating τ Decays in the Left-Right Symmetric Model”. In: *Physical Review D* (2007). DOI: `10.1103/physrevd.76.013004`. URL: `https://doi.org/10.1103\%2Fphysrevd.76.013004`.

[21] Athanasios Dedes, John Ellis, and Martti Raidal. “Higgs-mediated $B^0_{s,d} \to \mu\tau, e\tau$ and $\tau \to 3\mu, e\mu\mu$ decays in supersymmetric seesaw models”. In: *Physics Letters B* (2002). ISSN: 0370-2693. DOI: `https://doi.org/10.1016/S0370-2693(02)02900-3`. URL: `https://www.sciencedirect.com/science/article/pii/S0370269302029003`.

[22] K. S. Babu and Christopher Kolda. “Higgs-Mediated $\tau \to 3\mu$ in the Supersymmetric Seesaw Model”. In: *Physical Review Letters* 89.24 (2002). DOI: `10.1103/physrevlett.89.241802`. URL: `https://doi.org/10.1103\%2Fphysrevlett.89.241802`.

[23] Chong-xing Yue, Yan-ming Zhang, and Lan-jun Liu. “Nonuniversal gauge bosons Z-prime and lepton flavor violation tau decays”. In: *Phys. Lett. B* 547 (2002). DOI: `10.1016/S0370-2693(02)02781-8`. arXiv: `hep-ph/0209291`.

[24] Bernard Aubert et al. “Improved limits on the lepton-flavor violating decays $\tau^- \to \ell^-\ell^+\ell^-$”. In: *Phys. Rev. Lett.* 99 (2007). DOI: `10.1103/PhysRevLett.99.251803`. arXiv: `0708.3650 [hep-ex]`.

[25] Georges Aad et al. “Probing lepton flavour violation via neutrinoless $\tau \longrightarrow 3\mu$ decays with the ATLAS detector”. In: *Eur. Phys. J. C* 76 (2016). DOI: `10.1140/epjc/s10052-016-4041-9`. arXiv: `1601.03567 [hep-ex]`.

[26] Albert M Sirunyan et al. “Search for the lepton flavor violating decay $\tau \to 3\mu$ in proton-proton collisions at $\sqrt{s} = 13$ TeV”. In: *JHEP* 01 (2021), p. 163. DOI: `10.1007/JHEP01(2021)163`. arXiv: `2007.05658 [hep-ex]`.

[27] Swagato Banerjee et al. “Snowmass 2021 White Paper: Charged lepton flavor violation in the tau sector”. In: *arXiv:2203.14919* (2022). URL: `https://doi.org/10.48550/arXiv.2203.14919`.

[28] I. Zurbano Fernandez et al. “High-Luminosity Large Hadron Collider (HL-LHC): Technical design report”. In: 10/2020 (Dec. 2020). Ed. by I. Béjar Alonso et al. DOI: `10.23731/CYRM-2020-0010`.

[29] Roel Aaij et al. “Physics case for an LHCb Upgrade II-Opportunities in flavour physics, and beyond, in the HL-LHC era”. In: (2018). URL: `https://arxiv.org/abs/1808.08865v4`.

[30] The CMS Collaboration et al. *The Phase-2 Upgrade of the CMS Muon Detectors*. Tech. rep. Geneva: CERN, 2017. URL: `https://cds.cern.ch/record/2283189`.

[31] The ATLAS Collaboration et al. *Prospects for lepton flavour violation measurements in $\tau \to 3\mu$ decays with the ATLAS detector at the HL-LHC*. Tech. rep. Geneva: CERN, 2018. URL: `https://cds.cern.ch/record/2647956`.

[32] Mogens Dam. “Tau-lepton Physics at the FCC-ee circular e^+e^- Collider”. In: *SciPost Phys. Proc.* 1 (2019). DOI: `10.21468/SciPostPhysProc.1.041`. arXiv: `1811.09408 [hep-ex]`.

[33] O. Brüning, H. Burkhardt, and S. Myers. “The large hadron collider”. In: *Progress in Particle and Nuclear Physics* (2012). ISSN: 0146-6410. DOI: `https://doi.org/10.1016/j.ppnp.2012.03.001`. URL: `https://www.sciencedirect.com/science/article/pii/S0146641012000695`.

[34] The ATLAS Collaboration et al. “The ATLAS Experiment at the CERN Large Hadron Collider”. In: *Journal of Instrumentation* (2008). DOI: `10.1088/1748-0221/3/08/S08003`. URL: `https://dx.doi.org/10.1088/1748-0221/3/08/S08003`.

[35] The CMS Collaboration et al. “The CMS experiment at the CERN LHC”. In: *Journal of Instrumentation* (2008). DOI: `10.1088/1748-0221/3/08/S08004`. URL: `https://dx.doi.org/10.1088/1748-0221/3/08/S08004`.

[36] The LHCb Collaboration et al. “The LHCb Detector at the LHC”. In: *Journal of Instrumentation* (2008). DOI: `10.1088/1748-0221/3/08/S08005`. URL: `https://dx.doi.org/10.1088/1748-0221/3/08/S08005`.

[37] The ALICE Collaboration et al. “The ALICE experiment at the CERN LHC”. In: *Journal of Instrumentation* (2008). DOI: `10.1088/1748-0221/3/08/S08002`. URL: `https://dx.doi.org/10.1088/1748-0221/3/08/S08002`.

[38] Stephane Fartoukh et al. *LHC Configuration and Operational Scenario for Run 3*. Tech. rep. Geneva: CERN, 2021. URL: `https://cds.cern.ch/record/2790409`.

[39] Cheuk-Yin Wong. *Introduction to high-energy heavy-ion collisions*. World scientific, 1994. DOI: `https://doi.org/10.1142/1128`.

[40] Christian Elsasser. *$\bar{b}b$ production angle plots*. URL: `https://lhcb.web.cern.ch/speakersbureau/html/bb_productionangles.html`.

[41] Roel Aaij et al. “LHCb Detector Performance”. In: *Int. J. Mod. Phys. A* (2015). DOI: `10.1142/S0217751X15300227`. arXiv: `1412.6352 [hep-ex]`.

[42] LHCb Collaboration. *LHCb VELO Upgrade Technical Design Report*. Tech. rep. 2013. URL: `https://cds.cern.ch/record/1624070`.

[43] J Gassner, M Needham, and O Steinkamp. *Layout and Expected Performance of the LHCb TT Station*. Tech. rep. Geneva: CERN, 2004. URL: `https://cds.cern.ch/record/728548`.

[44] S. Amato et al. *LHCb RICH : Technical design report*. Tech. rep. 2000. URL: `https://hal.in2p3.fr/in2p3-00011869`.

[45] M. Adinolfi et al. "Performance of the LHCb RICH detector at the LHC". In: *Eur. Phys. J. C* 73 (2013). DOI: `10.1140/epjc/s10052-013-2431-9`. arXiv: `1211.6759 [physics.ins-det]`.

[46] P R et al. Barbosa-Marinho. *LHCb muon system: Technical Design Report*. Technical design report. LHCb. Geneva: CERN, 2001. URL: `http://cds.cern.ch/record/504326`.

[47] X. Cid Vidal. "Muon Identification in the LHCb experiment". In: (2010). Comments: Proceedings for the Moriond 2010 EW. arXiv: `1005.2585`. URL: `https://cds.cern.ch/record/1265848`.

[48] Giulio Dujany and Barbara Storaci. "Real-time alignment and calibration of the LHCb Detector in Run II". In: *Journal of Physics: Conference Series* 664 (2015). DOI: `10.1088/1742-6596/664/8/082010`.

[49] LHCb Starterkit. *The lhcb data flow*. 2015-2020. URL: `https://lhcb.github.io/starterkit-lessons/first-analysis-steps/dataflow.html` (visited on 04/05/2023).

[50] M. Clemencic et al. "The LHCb simulation application, Gauss: Design, evolution and experience". In: *J. Phys. Conf. Ser.* 331 (2011). Ed. by Simon C. Lin. DOI: `10.1088/1742-6596/331/3/032023`.

[51] I. Belyaev et al. "Handling of the generation of primary events in Gauss, the LHCb simulation framework". In: *2010 IEEE Nuclear Science Symposium, Medical Imaging Conference, and 17th Room Temperature Semiconductor Detectors Workshop*. 2010. DOI: `10.1109/NSSMIC.2010.5873949`.

[52] G. Corti et al. "Software for the LHCb experiment". In: *IEEE Trans. Nucl. Sci.* 53 (2006). DOI: `10.1109/TNS.2006.872627`.

[53] Torbjörn Sjöstrand. "The PYTHIA Event Generator: Past, Present and Future". In: *Comput. Phys. Commun.* 246 (2020). DOI: `10.1016/j.cpc.2019.106910`. arXiv: `1907.09874 [hep-ph]`.

[54] D. J. Lange. "The EvtGen particle decay simulation package". In: *Nucl. Instrum. Meth. A* 462 (2001). Ed. by S. Erhan, P. Schlein, and Y. Rozen. DOI: `10.1016/S0168-9002(01)00089-4`.

[55] S. Agostinelli et al. "GEANT4–a simulation toolkit". In: *Nucl. Instrum. Meth. A* 506 (2003), pp. 250–303. DOI: `10.1016/S0168-9002(03)01368-8`.

[56] I. Antcheva et al. "ROOT: A C++ framework for petabyte data storage, statistical analysis and visualization". In: *Comput. Phys. Commun.* 180 (2009), pp. 2499–2512. DOI: `10.1016/j.cpc.2009.08.005`. arXiv: `1508.07749 [physics.data-an]`.

[57] Barbara Storaci. "Optimization of the LHCb track reconstruction". In: *J. Phys. Conf. Ser.* 664.7 (2015). DOI: `10.1088/1742-6596/664/7/072047`.

[58] Carla Marin Benito. "PID strategy and performance at LHCb in Run 2". In: *PoS* ICHEP2018 (2019). DOI: `10.22323/1.340.0687`.

[59] F. Archilli et al. "Performance of the Muon Identification at LHCb". In: *JINST* 8 (2013). DOI: `10.1088/1748-0221/8/10/P10020`. arXiv: `1306.0249 [physics.ins-det]`.

[60] P F Harrison. "Blind analysis". In: *Journal of Physics G: Nuclear and Particle Physics* (2002). DOI: `10.1088/0954-3899/28/10/312`. URL: `https://dx.doi.org/10.1088/0954-3899/28/10/312`.

[61] S Tolk et al. *Data driven trigger efficiency determination at LHCb*. Tech. rep. Geneva: CERN, 2014. URL: `https://cds.cern.ch/record/1701134`.

[62] Andreas Höcker et al. "TMVA : The Toolkit for Multivariate Data Analysis eith ROOT". In: (2008). DOI: `10.5170/CERN-2008-001.184`. URL: `https://cds.cern.ch/record/1099990`.

[63] C. Gini. *Variabilità e mutabilità: contributo allo studio delle distribuzioni e delle relazioni statistiche. [Fasc. I.]* Studi economico-giuridici pubblicati per cura della facoltà di Giurisprudenza della R. Università di Cagliari. Tipogr. di P. Cuppini, 1912. URL: `https://books.google.de/books?id=fqjaBPMxB9kC`.

[64] Jerome H. Friedman. "Greedy function approximation: A gradient boosting machine." In: *The Annals of Statistics* (2001). DOI: `10.1214/aos/1013203451`. URL: `https://doi.org/10.1214/aos/1013203451`.

[65] Tianqi Chen and Carlos Guestrin. "Xgboost: A scalable tree boosting system". In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 2016. DOI: `https://doi.org/10.1145/2939672.2939785`. URL: `https://dl.acm.org/doi/abs/10.1145/2939672.2939785`.

[66] Cesare Bini. *Data analysis in Particle Physics*. 2016. URL: `https://www.roma1.infn.it/~bini/StatEPP_new.pdf`.

[67] Alexander L. Read. "Presentation of search results: The CL_s technique". In: *J. Phys. G* 28 (2002). Ed. by M. R. Whalley and L. Lyons. DOI: `10.1088/0954-3899/28/10/313`.

[68] Glen Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". In: *Eur. Phys. J. C* 71 (2011). [Erratum: Eur.Phys.J.C 73, 2501 (2013)]. DOI: `10.1140/epjc/s10052-011-1554-0`. arXiv: `1007.1727 [physics.data-an]`.

[69] Kim Jong Hae Kwak Sang Gyu. "Central limit theorem: the cornerstone of modern statistics". In: *kja* (2017). DOI: `10.4097/kjae.2017.70.2.144`. URL: `http://www.e-sciencecentral.org/articles/?scid=1156667`.

[70] LHCb Collaboration. *Standard set of performance numbers*. URL: `https://lhcb.web.cern.ch/speakersbureau/html/PerformanceNumbers.html` (visited on 04/05/2023).

[71] Vladimir V Gligorov. *Reconstruction of the Channel $B_d^0 \to D^+\pi^-$ and Background Classification at LHCb (revised)*. Tech. rep. Geneva: CERN, 2007. URL: `https://cds.cern.ch/record/1035682`.

[72] V. Gligorov. *The Background Category tool for background classification*. PPTS Meeting, CERN. 2013. URL: `https://twiki.cern.ch/twiki/pub/LHCb/Particle2MC/BkgCat_211013.key.pdf`.

[73] Georges Aad et al. "Measurement of the $W \to \tau\nu_\tau$ cross section in pp collisions at $\sqrt{s}$ =7 TeV with the ATLAS experiment". In: *Phys. Lett. B* 706 (2012). DOI: `10.1016/j.physletb.2011.11.057`. arXiv: `1108.4101 [hep-ex]`.

[74] Georges Aad et al. "Measurement of the $Z \to \tau\tau$ cross Section with the ATLAS Detector". In: *Phys. Rev. D* 84 (2011). DOI: `10.1103/PhysRevD.84.112006`. arXiv: `1108.2016 [hep-ex]`.

[75] Roel Aaij et al. "Measurements of prompt charm production cross-sections in pp collisions at $\sqrt{s} = 13$ TeV". In: *JHEP* 03 (2016). [Erratum: JHEP 09, 013 (2016), Erratum: JHEP 05, 074 (2017)]. DOI: `10.1007/JHEP03(2016)159`. arXiv: `1510.01707 [hep-ex]`.

[76] Roel Aaij et al. "Measurement of forward J/ψ production cross-sections in pp collisions at $\sqrt{s} = 13$ TeV". In: *JHEP* 10 (2015). [Erratum: JHEP 05, 063 (2017)]. DOI: `10.1007/JHEP10(2015)172`. arXiv: `1509.00771 [hep-ex]`.

[77] Artem Maevskiy et al. "Fast Data-Driven Simulation of Cherenkov Detectors Using Generative Adversarial Networks". In: *J. Phys. Conf. Ser.* 1525.1 (2020). DOI: `10.1088/1742-6596/1525/1/012097`. arXiv: `1905.11825 [physics.ins-det]`.

[78] Roel Aaij et al. "Design and performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC". In: *JINST* 14.04 (2019). DOI: `10.1088/1748-0221/14/04/P04013`. arXiv: `1812.10790 [hep-ex]`.

[79] Roel Aaij et al. *Optimization of the muon reconstruction algorithms for LHCb Run 2*. Tech. rep. Geneva: CERN, 2017. URL: `https://cds.cern.ch/record/2253050`.

[80] LHCb Collaboration. *LHCb Hlt GitLab repository*. URL: `https://gitlab.cern.ch/lhcb/Hlt/-/tree/2018-patches/` (visited on 04/25/2023).

[81] Rowina Caspary. "Search for the lepton flavour violating decay $\tau^- \to \mu^-\mu^+\mu^-$ with the LHCb experiment". MA thesis. Heidelberg University, Germany, 2020. URL: `https://www.physi.uni-heidelberg.de/Publications/masterthesis_RowinaCaspary.pdf`.

[82] Muriel Pivk and Francois R. Le Diberder. "SPlot: A Statistical tool to unfold data distributions". In: *Nucl. Instrum. Meth. A* 555 (2005). DOI: `10.1016/j.nima.2005.08.106`. arXiv: `physics/0402083`.

[83] N. L. Johnson. "Systems of Frequency Curves Generated by Methods of Translation". In: *Biometrika* (1949). ISSN: 00063444. URL: `http://www.jstor.org/stable/2332539` (visited on 04/24/2023).

[84] Roger J. Barlow. "Extended maximum likelihood". In: *Nucl. Instrum. Meth. A* 297 (1990). DOI: `10.1016/0168-9002(90)91334-8`.

[85] A. Rogozhnikov. "Reweighting with Boosted Decision Trees". In: *J. Phys. Conf. Ser.* 762.1 (2016). Ed. by Luis Salinas and Claudio Torres. DOI: `10.1088/1742-6596/762/1/012036`. arXiv: `1608.05806 [physics.data-an]`.

[86] Alex Rogozhnikov and contributors. *hep_ml documentation*. URL: `https://arogozhnikov.github.io/hep_ml/` (visited on 04/25/2023).

[87] Davide Anguita et al. "The 'K' in K-fold Cross Validation." In: *ESANN*. 2012. URL: `https://web.archive.org/web/20170829124122id_/https://www.elen.ucl.ac.be/Proceedings/esann/esannpdf/es2012-62.pdf`.

[88] "Measurement of the $B_s^0 \to \mu^+\mu^-$ decay properties and search for the $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-\gamma$ decays". In: *Phys. Rev. D* (1 2022). DOI: `10.1103/PhysRevD.105.012010`. URL: `https://link.aps.org/doi/10.1103/PhysRevD.105.012010`.

[89] Simon Chan and Philip Treleaven. "Chapter 5 - Continuous Model Selection for Large-Scale Recommender Systems". In: *Big Data Analytics*. Ed. by Venu Govindaraju, Vijay V. Raghavan, and C.R. Rao. Handbook of Statistics. Elsevier, 2015. DOI: `https://doi.org/10.1016/B978-0-444-63492-4.00005-8`. URL: `https://www.sciencedirect.com/science/article/pii/B9780444634924000058`.

[90] Mathieu Perrin-Terrin and Giampiero Mancinelli. *Optimisation of the binning of the discriminating variables used in the computation of* $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ *upper limits with the modified frequentist approach.* Tech. rep. Geneva: CERN, 2012. URL: `https://cds.cern.ch/record/1419784`.

[91] Giovanni Punzi. "Sensitivity of searches for new signals and its optimization". In: *eConf* C030908 (2003). Ed. by L. Lyons, R. P. Mount, and R. Reitmeyer, MODT002. arXiv: `physics/0308063`. URL: `https://www.slac.stanford.edu/econf/C030908/papers/MODT002.pdf`.

[92] L. Garren et al. "Monte carlo particle numbering scheme". English. In: *The European Physical Journal.C, Particles and Fields.* (Mar. 2000). URL: `https://www.proquest.com/scholarly-journals/monte-carlo-particle-numbering-scheme/docview/2293987757/se-2`.

[93] R. Aaij et al. "Search for $D^+_{(s)} \to \pi^+\mu^+\mu^-$ and $D^+_{(s)} \to \pi^-\mu^+\mu^+$ decays". In: *Physics Letters B* (2013). ISSN: 0370-2693. DOI: `https://doi.org/10.1016/j.physletb.2013.06.010`. URL: `https://www.sciencedirect.com/science/article/pii/S0370269313004747`.

[94] J Albrecht et al. "The upgrade of the LHCb trigger system". In: *Journal of Instrumentation* (2014). DOI: `10.1088/1748-0221/9/10/C10026`. URL: `https://dx.doi.org/10.1088/1748-0221/9/10/C10026`.

[95] Roel Aaij et al. "Measurement of the track reconstruction efficiency at LHCb". In: *JINST* 10.02 (2015). DOI: `10.1088/1748-0221/10/02/P02007`. arXiv: `1408.1251 [hep-ex]`.

[96] F. Archilli, G. Frau, and R. Kopecna. *TrackCalib Package Twiki.*

[97] Flavio Archilli et al. "Measurement of $B^0_{s,d} \to \mu^+\mu^-$ decays with the Run 1 and Run 2 datasets". In: (2020). URL: `https://cds.cern.ch/record/2725940`.

[98] G. Breit and E. Wigner. "Capture of Slow Neutrons". In: *Physical Review* (1936). DOI: `10.1103/PhysRev.49.519`.

[99] R. Aaij et al. "Prompt charm production in pp collisions at $\sqrt{s} = 7$ TeV". In: *Nuclear Physics B* (2013). ISSN: 0550-3213. DOI: `https://doi.org/10.1016/j.nuclphysb.2013.02.010`. URL: `https://www.sciencedirect.com/science/article/pii/S0550321313000965`.

[100] R. Aaij et al. "Measurement of J/ψ production in pp collisions at $\sqrt{s} = 7$ TeV". In: *Eur. Phys. J. C* 71 (2011), p. 1645. DOI: `10.1140/epjc/s10052-011-1645-y`. arXiv: `1103.0423 [hep-ex]`.

[101] Benjamin M. Dassinger et al. "Model-independent analysis of lepton flavour violating τ decays". In: *Journal of High Energy Physics* (2007). DOI: `10.1088/1126-6708/2007/10/039`. URL: `https://dx.doi.org/10.1088/1126-6708/2007/10/039`.