Search for the rare B0 -> K*0 tau+ tau- decay at the LHCb experiment

Jacopo Cerasoli

NNT/NL : 2021AIXM0487/063ED352

THÈSE DE DOCTORAT

Soutenue à Aix-Marseille Université
le 3 décembre 2021 par

Jacopo CERASOLI

Search for the rare $B^0 \to K^{*0}\tau^+\tau^-$ decay at the LHCb experiment

Discipline
Physique et Sciences de la Matière

Spécialité
Physique des Particules et Astroparticules

École doctorale
Physique et Sciences de la Matière (ED 352)

Centre de Physique des Particules de Marseille (CPPM) UMR 7346

Composition du jury

Cristinel DIACONU Aix Marseille Univ, CNRS/IN2P3, CPPM, Marseille, France	Président du jury
Patrick ROBBE IJCLab (IN2P3/CNRS)	Rapporteur
Marcella BONA School of Physics and Astronomy, Queen Mary University	Rapporteure
Claudia CECCHI Università di Perugia e INFN	Examinatrice
Giampiero MANCINELLI Aix Marseille Univ, CNRS/IN2P3, CPPM, Marseille, France	Directeur de thèse
Julien COGAN Aix Marseille Univ, CNRS/IN2P3, CPPM, Marseille, France	Co-directeur de thèse

I, undersigned Jacopo Cerasoli, hereby declare that the work presented in this manuscript is my own work, carried out under the scientific direction of Giampiero Mancinelli and Julien Cogan, in accordance with the principles of honesty, integrity and responsibility inherent to the research mission. The research work and the writing of this manuscript have been carried out in compliance with both the french national charter for Research Integrity and the Aix-Marseille University charter on the fight against plagiarism.

This work has not been submitted previously either in this country or in another country in the same or in a similar version to any other examination body.

Marseille
February 23, 2022

Abstract

A search for the rare $B^0 \to K^{*0}\tau^+\tau^-$ decay is performed using 9 fb^{-1} of proton-proton (pp) collision data collected at the LHCb experiment at CERN. The decay is reconstructed using two final states: the fully hadronic one, in which both τ leptons decay in the $\tau^- \to \pi^-\pi^+\pi^-(\pi^0)\nu_\tau$ channel, and the mixed hadronic-leptonic, in which one τ decays in the aforementioned channel and the other in the $\tau^- \to \mu^-\nu_\tau\overline{\nu}_\mu$ channel. The measure of the branching ratio is performed relatively to that of the normalization mode $B^0 \to D^-(\to \pi^-K^+\pi^-)D_s^+(\to K^+K^-\pi^+)$. The full event selection makes use of Boosted Decision Trees (BDT) based on kinematic and isolation variables in order to reduce the amount of background. A binned maximum likelihood fit on the output of a BDT classifier is performed to extract the number of signal events in the data. The data distribution is still blind, however an expected upper limit in the case where no signal is observed is computed to be $3.2 \cdot 10^{-4}$ at 95% CL.

Acknowledgements

As I write these words I am about to leave Marseille and the LHCb collaboration. It seems like yesterday when I arrived in Marseille for my PhD, without knowing a single word of french, scared and wet (it was raining like hell). The past three years have been intense, and gave me countless opportunities of personal growth.

I would like to thank all the people I met during this long journey, in particular my thesis supervisors Giampiero and Julien, to whom I am extremely grateful.

Un ringraziamento speciale va alla mia famiglia, che mi ha sempre regalato la possibilità di scegliere liberamente il mio cammino.

Marseille
February 23, 2022

Résumé en Français

Cadre théorique

Le modèle standard (MS) de la physique des particules représente la théorie la plus avancée pour décrire le comportement des particules élémentaires et leurs interactions. Malgré son impressionnant pouvoir prédictif, plusieurs faits expérimentaux suggèrent l'existence d'une théorie plus générale des particules élémentaires :

énergie et matière noire : environ 95 % de l'univers est composé d'énergie noire et de matière noire, qui ne sont pas décrites par le MS [1] ;

asymétrie matière-antimatière : l'univers est principalement composé de matière, alors que l'antimatière n'est observée que sporadiquement dans les rayons cosmiques ou produite artificiellement dans les accélérateurs de particules. Cette asymétrie n'est pas prise en compte par le modèle [2] ;

masse des neutrinos : des observations récentes montrent que les neutrinos ont une masse non nulle, alors que le MS les prédit sans masse [3, 4].

Les recherches d'une nouvelle physique au-delà du MS sont menées selon deux approches principales : les recherches directes et indirectes. Les premières consistent à détecter directement de nouvelles particules, tandis que les secondes consistent à mesurer des observables précisément calculées afin de mettre en évidence des divergences avec la théorie, ce qui indiquerait la présence d'une nouvelle physique. Les recherches indirectes impliquant des désintégrations de mésons B sont particulièrement intéressantes. Parmi les six quarks, le b est le quark le plus lourd capable d'hadroniser avant de se désintégrer, permettant ainsi une grande variété d'études. De plus, la durée de vie moyenne relativement longue des mésons B les rend particulièrement adaptés pour être étudiés dans les accélérateurs de particules.

Récemment, des signes de divergences par rapport aux prédictions du MS ont été observées dans plusieurs mesures expérimentales impliquant des désintégrations de hadrons B. Parmi les plus importantes, figurent les mesures suivantes :

- les mesures $R_{K^{*0}} = \frac{\bar{\mathcal{B}}(B^0 \to K^{*0}\mu^+\mu^-)}{\bar{\mathcal{B}}(B^0 \to K^{*0}e^+e^-)}$ et $R_K = \frac{\bar{\mathcal{B}}(B^+ \to K^+\mu^+\mu^-)}{\bar{\mathcal{B}}(B^+ \to K^+e^+e^-)}$ effectuées par LHCb [5, 6] dévient par environ 2.4σ et 3.1σ de la prédiction du MS ;
- les analyses angulaires de $B^+ \to K^{*+}\mu^+\mu^-$ et $B^0 \to K^{*0}\mu^+\mu^-$ [7, 8] montrent des tensions locales de l'ordre de 3σ pour les observables P_2 et P_5' ;

- la mesure combinée de $R(D)$ et $R(D^*)$ [9, 10, 11] montre un écart d'environ 3.4σ par rapport à la prédiction du MS.

Les écarts observés par rapport aux prédictions théoriques, appelés *anomalies de saveur*, suggèrent une possible violation de l'universalité de la saveur leptonique (LFU), qui stipule que les trois leptons chargés sont identiques, à l'exception des valeurs de leurs masses. Les indices d'une nouvelle physique dans les désintégrations semileptoniques impliquant des leptons τ dans l'état final suggèrent que des effets de violation de LFU pourraient apparaître dans les courants neutres $b \to s\tau^+\tau^-$. Leurs rapports d'embranchement devraient être amplifiés de plusieurs ordres de grandeur par rapport à ceux prévus par le MS pour expliquer la valeur centrale de $R(D^{(*)})$. À l'heure actuelle, les désintégrations rares de B avec des leptons τ dans l'état final sont encore peu étudiées, les résultats expérimentaux étant [12, 13, 14] :

$$\begin{aligned} \bar{\mathcal{B}}(B^0 \to \tau^+\tau^-) &< 2.1 \cdot 10^{-3} \text{ à } 95\% \text{ CL}, \\ \bar{\mathcal{B}}(B_s^0 \to \tau^+\tau^-) &< 6.8 \cdot 10^{-3} \text{ à } 95\% \text{ CL}, \\ \bar{\mathcal{B}}(B^+ \to K^+\tau^+\tau^-) &< 2.25 \cdot 10^{-3} \text{ à } 90\% \text{ CL}, \\ \bar{\mathcal{B}}(B^0 \to K^{*0}\tau^+\tau^-) &< 2.0 \cdot 10^{-3} \text{ à } 90\% \text{ CL}. \end{aligned} \tag{1}$$

Dans ce document, la première tentative de recherche de la désintégration rare $B^0 \to K^{*0}\tau^+\tau^-$ à LHCb est présentée. La désintégration a un rapport de branchement d'environ 10^{-7} [15], qui pourrait être amélioré de plusieurs ordres de grandeur grâce aux contributions des particules de nouvelle physique.

Le détecteur LHCb au LHC

Le Large Hadron Collider (LHC) est un accélérateur pp d'une circonférence de 26.7 km situé à l'Organisation Européenne pour la Recherche Nucléaire (CERN). Le tunnel est situé entre 45 et 170 m sous terre. Deux faisceaux de protons sont accélérés dans des directions opposées et entrent en collision dans quatre points d'interaction, où les principales expériences (ATLAS, CMS, ALICE et LHCb) sont installées et recueillent des données. Au cours de la période de prise de données du Run 1 (2011-2012), les protons ont été accélérés à des énergies de 3.5 TeV, tandis que pendant le Run 2 (2015-2018), ils ont été accélérés à des énergies de 6.5 TeV, correspondant respectivement à des energies dans le centre de masse des collisions pp de $\sqrt{s} = 7$ et 13 TeV. L'analyse décrite dans ce document exploite les données de collision de Run 1 et Run 2.

L'objectif principal de l'expérience LHCb est d'étudier les processus de violation de CP et d'étudier les désintégrations rares des hadrons beaux et charmés. L'appareil est constitué d'un spectromètre à un seul bras d'une couverture angulaire autour de la ligne de faisceau de 10 mrad à 300 mrad dans le plan de courbure. Une vue schématique de la disposition du détecteur est illustrée sur la Figure 1. Les principaux composants sont :

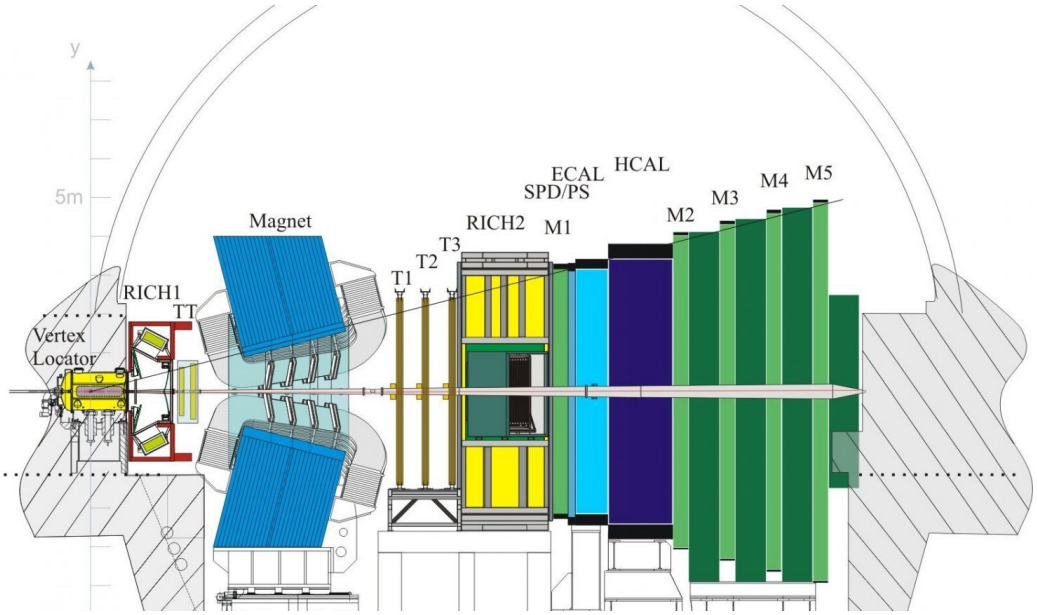


FIGURE 1 : Disposition du détecteur LHCb.

- le VErtex LOcator (VELO) utilisé pour mesurer avec précision la position des vertex ;
- un aimant dipolaire fournissant le champ de courbure pour le spectromètre ;
- les stations du système trajectographique pour effectuer des mesures de position et de quantité de mouvement des particules ;
- deux détecteurs Ring Imaging CHerenkov (RICH) effectuant l'identification des particules ;
- un système trajectographique fournissant des informations de déclenchement et d'identification pour les muons.

Analyse expérimentale

L'analyse est effectuée à l'aide de l'ensemble de données complet de LHCb enregistré à ce jour (Run 1–2). L'objectif est de mesurer le rapport d'embranchement de la désintégration rare $B^0 \to K^{*0}\tau^+\tau^-$ en utilisant la désintégration $B^0 \to D^- D_s^+$ comme canal de normalisation :

$$\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-) \propto \frac{N^{\text{obs}}_{B^0\to K^{*0}\tau^+\tau^-}}{\epsilon_{B^0\to K^{*0}\tau^+\tau^-}} \mathcal{B}_{\text{norm}} \frac{\epsilon_{\text{norm}}}{N^{\text{obs}}_{\text{norm}}}. \tag{2}$$

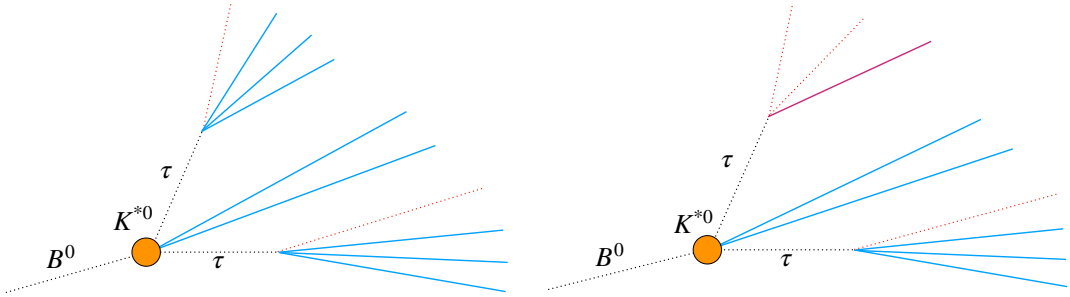
En l'absence de signal, une limite supérieure pour le rapport d'embranchement sera calculée.

Dans cette analyse, deux états finaux différents sont considérés :

- **état final $3\pi 3\pi$** : les deux leptons τ sont reconstruits en utilisant leur désintégration en trois pions chargés, qui a un rapport de branchement de $\mathcal{B}(\tau^- \to \pi^-\pi^+\pi^-\nu_\tau) = (9.31 \pm 0.05)\%$, ou dans la désintégration avec un pion neutre supplémentaire, qui a un rapport d'embranchement $\mathcal{B}(\tau^- \to \pi^-\pi^+\pi^-\pi^0\nu_\tau) = (4.62 \pm 0.05)\%$ [16].

- **état final $3\pi\mu$** : un lepton est reconstruit en utilisant sa désintégration en trois pions chargés ou celle avec le pion neutre, tandis que l'autre est reconstruit en utilisant son mode de désintégration leptonique $\tau^- \to \mu^-\nu_\mu\nu_\tau$ qui a un rapport d'embranchement $\mathcal{B}(\tau^- \to \mu^-\nu_\mu\nu_\tau) = (17.39 \pm 0.04)\%$ [16].

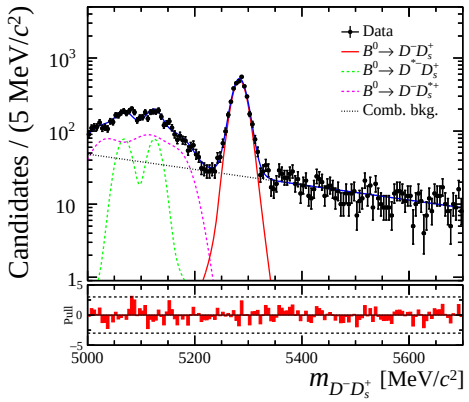
Dans les deux états finaux le K^{*0} est reconstruit via sa désintégration $K^{*0} \to K^-\pi^+$ dont le rapport d'embranchement est $\mathcal{B}(K^{*0} \to K^-\pi^+) \sim \frac{2}{3}$. Une vue schématique des deux états finaux est rapportée sur la Figure 2.



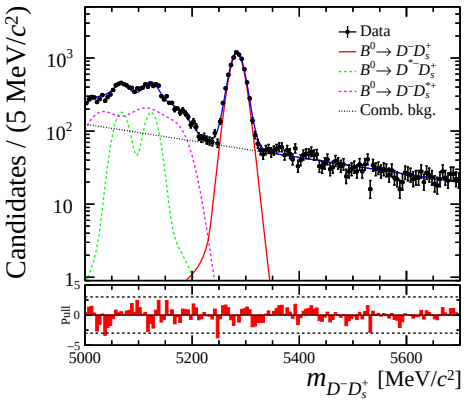
(a) État final $3\pi 3\pi$ (b) État final $3\pi\mu$

FIGURE 2 : Vue schématique des deux états finaux. Le B^0 et les particules intermédiaires sont indiqués sur la figure, tandis que les lignes bleues représentent les hadrons dans l'état final, la ligne violette indique le muon et les lignes pointillées rouges indiquent les neutrinos. Le pion neutre est omis.

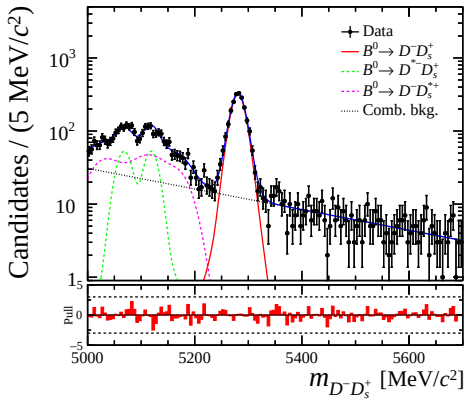
Les candidats pour le mode de normalisation sont sélectionnés à l'aide d'un ensemble de coupures sur des variables d'isolation et cinématiques, en plus des coupures sur la masse invariante des mésons D. L'ajustement sur les données du canal de normalisation de la masse invariante est effectué séparément pour chaque année de prise de données et est illustré sur la Figure 3.



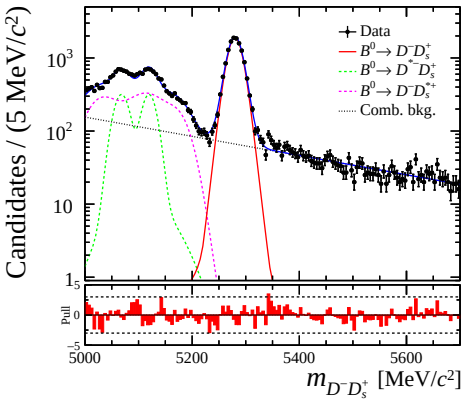
(a) 2011



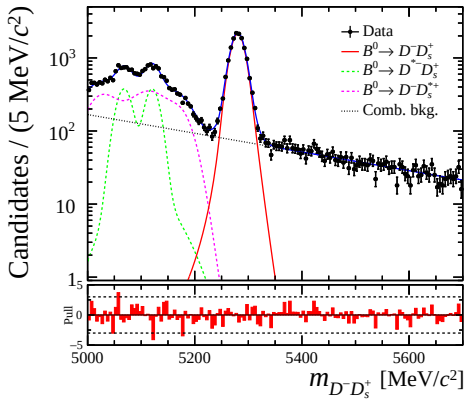
(b) 2012



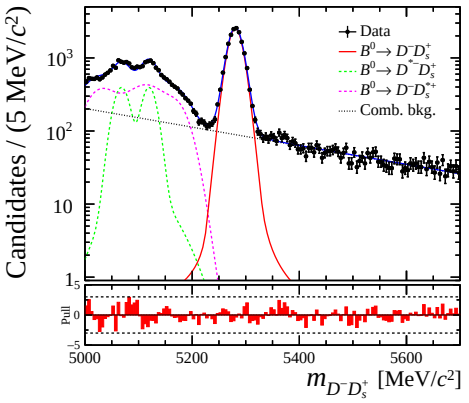
(c) 2015



(d) 2016



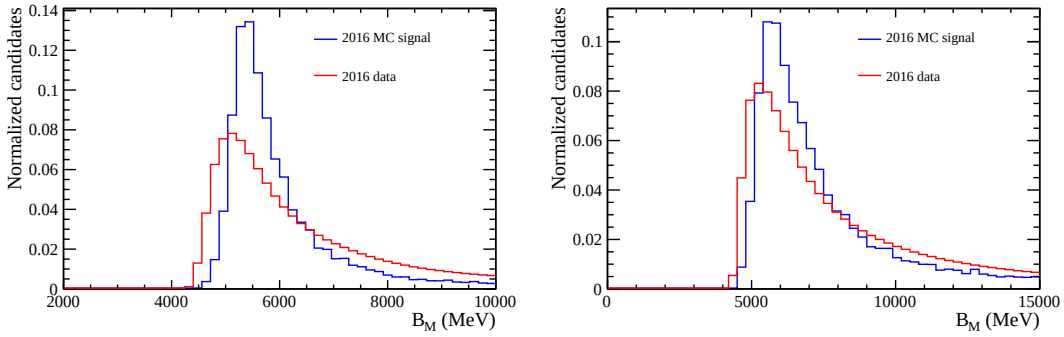
(e) 2017



(f) 2018

FIGURE 3 : Ajustement des données du canal de normalisation sur la masse invariante. Le signal est représenté en rouge, les fonds sont représentés avec les lignes pointillées et la distribution totale est représentée en bleu.

En plus de la masse invariante calculée à partir des particules chargées dans l'état final, il est possible de reconstruire analytiquement la quantité de mouvement manquante des neutrinos en appliquant des contraintes de masse sur les leptons τ. La distribution de masse reconstruite est illustrée sur la Figure 4. Elle ne montre



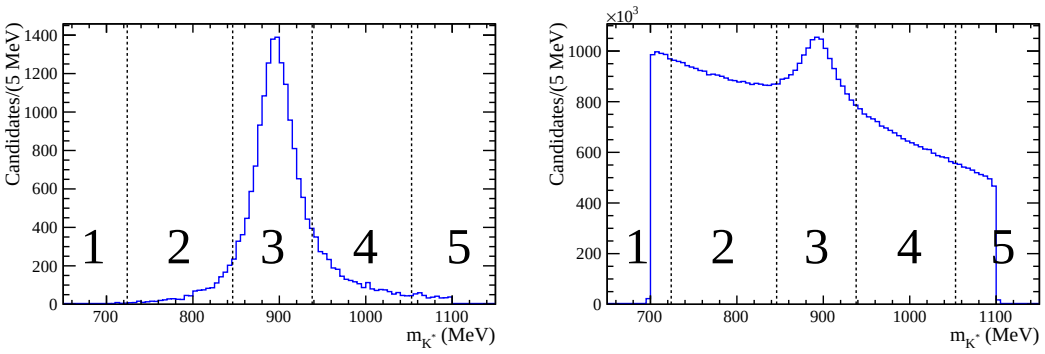
(a) État final $3\pi3\pi$ (b) État final $3\pi\mu$

FIGURE 4 : Masse reconstruite analytiquement pour des évènements de signal simulés (bleu) et données (rouge) pour les deux états finaux en 2016. Les autres années de prise de données présentent des distributions similaires.

pas un pouvoir de discrimination suffisant entre le signal et le bruit de fond. Il a donc été décidé d'effectuer l'ajustement final sur la distribution d'un classificateur d'analyse multivariée Boosted Decision Tree (BDT).

L'analyse utilise la masse invariante du K^{*0} afin de séparer l'échantillon de données en sous-régions avec des rapports signal/fond différents. Le méson K^{*0} a une masse invariante $m_{K^{*0}} = 891.66 \pm 0.26\,\text{MeV}$. La distribution de masse est illustrée sur la Figure 5 pour la simulation et les données. En utilisant quatre coupures, cinq régions différentes de la distribution de masse invariante sont definies. Elles sont étiquetées de "1" à "5" sur la Figure 5 et sont les mêmes pour les deux états finaux. L'échantillon de données peut être divisé en trois sous-échantillons mutuellement exclusifs :

- **région de signal** : la masse invariante du K^{*0} doit se situer dans la région 3. Les candidats sont ceux qui ressemblent le plus au signal et forment le sous-échantillon sur lequel l'ajustement final est effectué ;
- **région de fond** : la masse invariante du K^{*0} doit se trouver dans les régions 1 ou 5. La contamination du signal sur les données de ce sous-échantillon est négligeable et les candidats sont donc utilisés comme proxy pour le bruit de fond ;
- **région de contrôle** : la masse invariante du K^{*0} doit se trouver dans les régions 2 ou 4. Les candidats de ce sous-échantillon sont utilisés pour modéliser le bruit de fond dans l'ajustement final. Cependant, étant donné qu'une



(a) (b)

FIGURE 5 : Distribution de masse invariante du K^{*0} 2016 pour les évènements de signaux simulés (a) et données (b) pour l'état final $3\pi 3\pi$. Les distributions pour l'autre état final et les autres années de prise de données sont similaires.

quantité de signal non négligeable est présente, la contamination du signal doit être prise en compte.

Afin de séparer le signal du bruit de fond, une procédure de sélection a été mise en place. Elle est composée de deux étapes principales, la première étant une sélection par coupures et la seconde une sélection basée sur une analyse multivarié :

- sélection par coupures : des variables d'isolation et cinématiques qui présentent un pouvoir discriminant particulièrement élevé sont utilisées à ce stade. Cette première étape de sélection vise à rejeter le fond le plus trivial tout en conservant une quantité de signal élevée. De plus, elle rejette les évènements parasites qui se situent loin dans les queues des distributions ;

- sélection multivariée : elle est composée de deux BDT en séquence, où le second est entraîné après la sélection sur le premier. Ils sont entraînés à l'aide d'évènements simulés de signaux et de données de la région de fond. Les BDT sont entraînés séparément pour chaque état final à l'aide de variables cinétiques et d'isolation. Cette étape de sélection vise à supprimer la majeure partie du bruit de fond.

Les efficacités de sélection sont évaluées sur la simulation, mais comme la simulation ne reproduit pas parfaitement la distribution de certaines variables, des techniques spécifiques basées sur les données ont été utilisées pour calculer des corrections et/ou des incertitudes systématiques. De plus, le canal de normalisation a été utilisé pour calculer les corrections à appliquer aux efficacités de sélection en raison des différences entre les données et les simulations dans les variables utilisées dans la sélection hors ligne.

Après la phase de sélection, un autre BDT est entraîné en exploitant le pouvoir discriminant restant. Ce dernier BDT est utilisé pour effectuer un ajustement afin

d'extraire le nombre d'évènements de signal dans les données. Le modèle utilisé dans l'ajustement est défini comme suit :

$$\mathbf{Data}_{\boldsymbol{y}} = f_y \mu \cdot \mathbf{Sig}_{\boldsymbol{y}} + \frac{s_y}{c_y} n_y^{\mathrm{ctl}} \cdot \mathbf{Ctl}_{\boldsymbol{y}} - f_y \mu \frac{s_y}{c_y} \frac{\epsilon_y^{\mathrm{ctl}}}{\epsilon_y^{\mathrm{sig}}} \cdot \mathbf{Cont}_{\boldsymbol{y}}, \tag{3}$$

où l'indice y fait référence à une année spécifique de prise de données d'un état final donné et par souci de clarté, les PDF normalisés ont été mis en évidence en bleu tandis que les paramètres ajustés ont été mis en évidence en rouge. Les différents termes sont :

$\mathbf{Data}_{\boldsymbol{y}}$: histogramme représentant le nombre total d'évènements de signal et de bruit de fond ;

$\mathbf{Sig}_{\boldsymbol{y}}$: distribution décrivant le BDT du signal, extraite de la simulation dans la région de signal ;

$\mathbf{Ctl}_{\boldsymbol{y}}$: distribution décrivant le fond, extraite des données de la région de contrôle ;

$\mathbf{Cont}_{\boldsymbol{y}}$: distribution décrivant la contamination des évènements de signal dans la région de contrôle, extrait de la simulation ;

$\boldsymbol{\mu}$: rapport de branchement du signal, le paramètre d'intérêt partagé entre les états finaux et les années de prise de données ;

$\frac{\boldsymbol{s_y}}{\boldsymbol{c_y}}$: rapport du nombre (inconnu) d'évènements de fond dans la région de signal et de contrôle ;

$\boldsymbol{f_y}$: fraction d'évènements de signal pour une année de prise de données et un état final ;

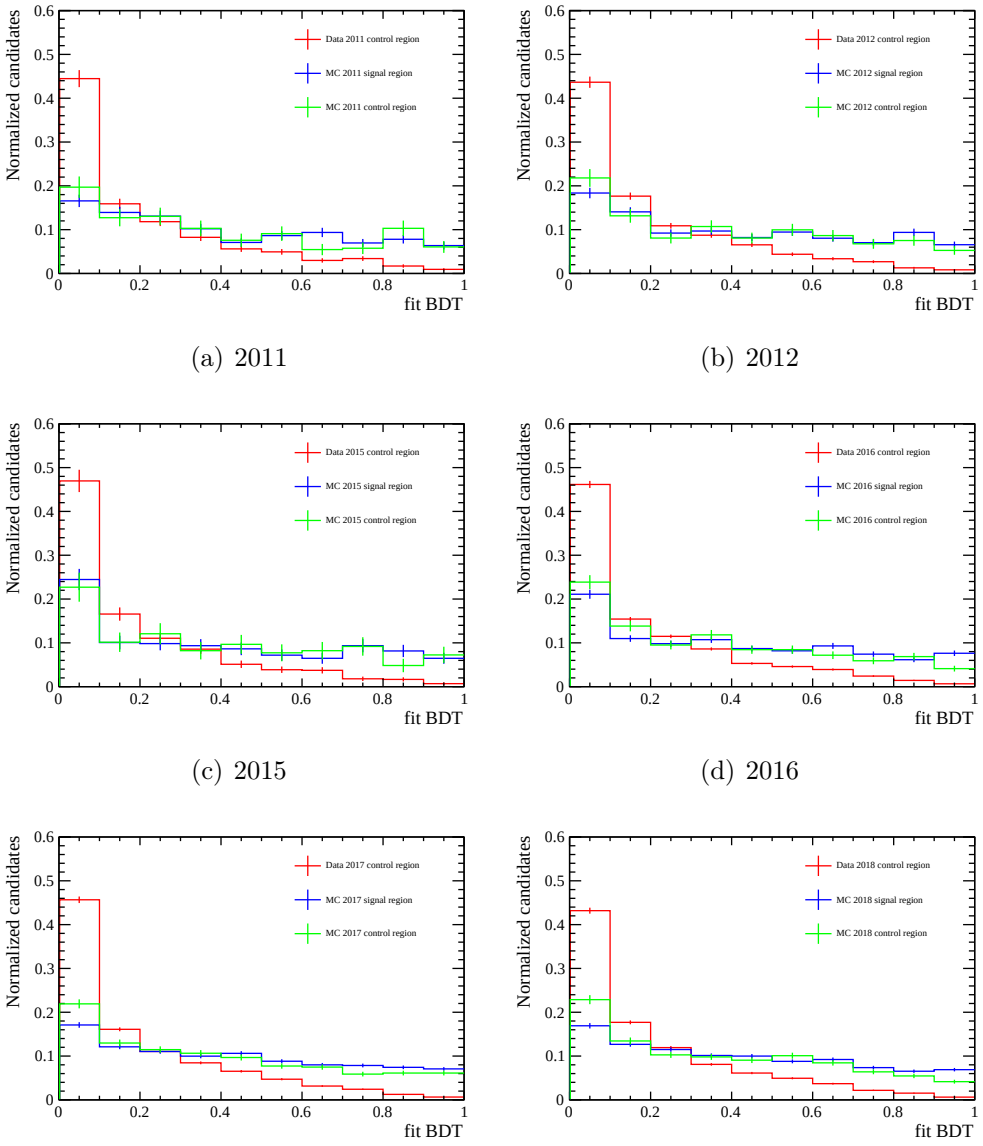
$\frac{\boldsymbol{\epsilon_y^{\mathrm{ctl}}}}{\boldsymbol{\epsilon_y^{\mathrm{sig}}}}$: rapport des efficacités dans la région de contrôle et du signal ;

$\boldsymbol{n_y^{\mathrm{ctl}}}$: nombre d'évènements dans la région de contrôle.

Les différentes distributions sont montrées sur les Figures 6 et 7.

La distribution des données dans la région de signal est encore masquée. Une limite supérieure dans le cas où aucun signal n'est observé a été calculée avec la méthode CL_s [17] comme étant :

$$\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-) < 3.2 \cdot 10^{-4} \text{ à } 95\% \text{ CL}. \tag{4}$$

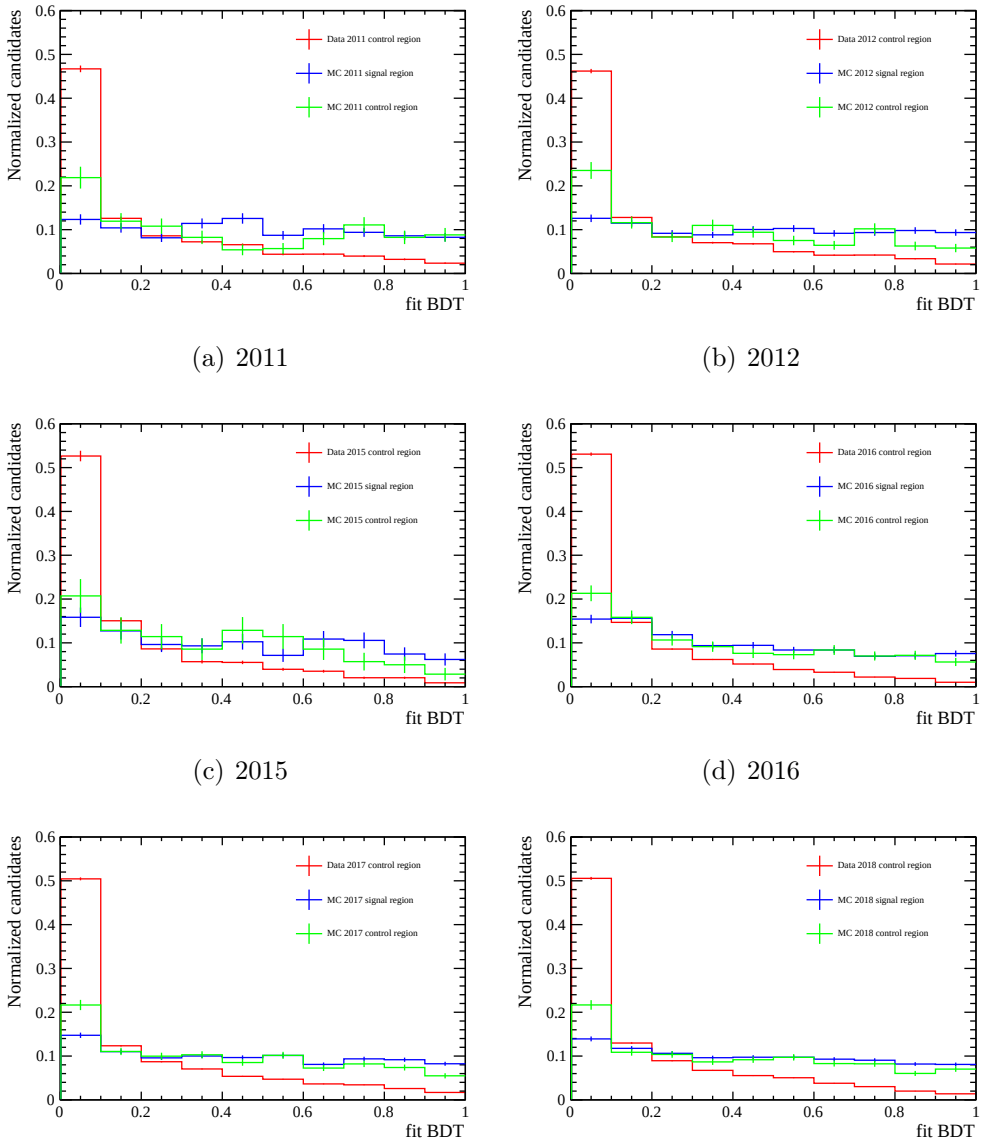


(a) 2011 (b) 2012

(c) 2015 (d) 2016

(e) 2017 (f) 2018

FIGURE 6 : Distribution du BDT dans l'état final $3\pi 3\pi$ pour les données dans la région de contrôle (rouge), simulation dans la région de signal (bleu) et simulation dans la région de contrôle (vert) pour chaque année de prise de données.



(a) 2011 (b) 2012

(c) 2015 (d) 2016

(e) 2017 (f) 2018

FIGURE 7 : Distribution du BDT dans l'état final $3\pi\mu$ pour les données dans la région de contrôle (rouge), simulation dans la région de signal (bleu) et simulation dans la région de contrôle (vert) pour chaque année de prise de données.

Contents

Introduction

The effort of the particle physics community aimed at developing a comprehensive theory of elementary particles has led to the foundation of the Standard Model (SM). Despite the astonishing predictive power of the theory, the limitations of the SM motivate the quest for a more general description of the behavior of elementary particles. Chapter 1 of this thesis presents the main features of the SM and the motivations behind the search for new physics (NP). It focuses the attention on indirect searches in the flavor sector, more specifically on the study of rare B hadron decays. Because of their low values, the branching ratios of such decays could be sensitive to small corrections due to NP contributions. In particular, rare B hadron decays involving τ leptons in the final state might play a crucial role in the search for NP, since they are still poorly investigated and their branching ratios might be enhanced by NP contributions up to several orders of magnitude [15].

Searches for NP are now pursued at the Large Hadron Collider (LHC), where four experiments (ATLAS, CMS, ALICE and LHCb) are installed and take data. The LHCb experiment is optimized for the detection of B hadrons, and is therefore well-suited to perform the aforementioned studies. Chapter 2 describes the CERN accelerator complex, with a particular focus on the LHCb experiment.

Chapters 3 to 10 describe the search for the rare $B^0 \to K^{*0}\tau^+\tau^-$ decay, performed on 9 fb^{-1} of data collected at the LHCb experiment. An event selection is established, which involves the use of multivariate classifiers in order to drastically reduce the amount of background. Successively, a binned maximum likelihood fit is used to extract the number of signal events in data. Since the reconstructed mass distribution offers poor discriminating power, the fit is performed on the output of a multivariate classifier. Despite the fact that the data distribution is still blind, an expected upper limit in the case where no signal is observed is computed.

Part I

Theoretical framework and experimental apparatus

Chapter 1

The Standard Model and beyond

This chapter focuses on the description of the SM of particle physics and the search for a beyond the SM theory. In Sec. 1.1 the main features of the SM are presented, together with a brief summary of the theoretical and experimental motivations for the search for a more fundamental theory; Sec. 1.2 shows an overview of the theory of electroweak interactions, with particular attention on the Higgs mechanism and the quark mixing; in Sec. 1.3 the modern theoretical methods to investigate the presence of new physics are introduced, together with the most up-to-date experimental results in the field of flavor physics; finally Sec. 1.4 is dedicated to a brief introduction about the phenomenology beyond the SM of rare B hadron decays with τ leptons in the final state.
For further information on the SM the reader is referred to [18].

1.1 Particles and interactions

The SM is the theory describing elementary particles and their interactions. It was built during the 60's and 70's within the Quantum Field Theory (QFT) framework and has been experimentally tested since then, showing its extremely powerful predicting nature. The theory describes phenomena from the scale of $\mathcal{O}(1\,\text{eV})$ to $\mathcal{O}(100\,\text{GeV})$, giving a consistent description of three out of the four forces known in nature: the strong, electromagnetic and weak interactions. Despite its incredible predictive power, the SM does not account for some well-established experimental facts:

dark energy and matter: around 95% of the universe is known to be composed of dark energy and dark matter, which are not described by the SM [1];

matter - anti-matter asymmetry: the universe is composed primarily of matter, while anti-matter is only observed sporadically in cosmic rays or produced artificially in particle accelerators. This asymmetry is not accounted for by the model [2];

neutrino mass: recent observations show that neutrinos have a non-zero mass, while the SM predicts them to be massless [3, 4].

Besides these experimental facts, numerous theoretical prejudices affect the theory: the Higgs boson mass is expected to be much larger than the measured value unless a powerful fine-tuning cancellation happens between the bare mass and the quantum corrections. Is there a deeper motivation for this cancellation to happen? Why are the masses of elementary particles so different (and hence why does the theory require a large number of free parameters)? Is it possible to include the gravitational force in this framework? The ensemble of these motivations inspire the search for a more comprehensive and fundamental theory of elementary particles.

The fundamental entities described by the SM are dynamic *fields*, *i.e.* mathematical functions defined at each point of the space-time $x \equiv (t, \vec{x})$, where x is a compact notation which includes both the temporal and the spatial coordinates. The field equations are derived from a Lagrangian density $\mathcal{L}$ by applying the principle of least action to the path followed by a given field $\varphi(x)$, leading to the equations

$$\frac{\partial \mathcal{L}}{\partial \varphi(x)} - \frac{\partial}{\partial x^\alpha}\left(\frac{\partial \mathcal{L}}{\frac{\partial}{\partial x^\alpha}\varphi(x)}\right) = 0, \tag{1.1}$$

where the Einstein summation convention for the α indices is used. The fields are quantized by interpreting them as *field operators* (*i.e.* operators defined at each point of the space-time) and imposing canonical relations between them and their conjugate momenta $\pi(x) \equiv \frac{\partial \mathcal{L}}{\partial \dot{\varphi}(x)}$, where $\dot{\varphi}(x)$ stands for the time derivative of the field. In this scheme, particles are interpreted as excitations of the field at a given point of the space-time. In particular, the elementary particles are divided into two classes: *bosons* and *fermions*. Bosons are particles with integer spin and their fields $\phi(x)$ are quantized imposing commutation relations:

$$[\phi(t, \vec{x}), \pi_\phi(t, \vec{x}')] = i\,\delta(\vec{x} - \vec{x}'), \tag{1.2}$$

where $[A, B] \equiv AB - BA$, with A and B two operators, π_ϕ is the conjugate momentum of the field ϕ and $\delta(\vec{x} - \vec{x}')$ indicates the 3-dimensional Dirac delta function. Fermions are particles with half-integer spin and their associated fields $\psi(x)$ are quantized imposing anti-commutation relations:

$$[\psi(t, \vec{x}), \pi_\psi(t, \vec{x}')]_+ = i\,\delta(\vec{x} - \vec{x}'), \tag{1.3}$$

where $[A, B]_+ \equiv AB + BA$ and π_ψ is the conjugate momentum of the field ψ. Fermions obey the Pauli exclusion principle which states that two identical fermions can not occupy the same quantum state simultaneously. This does not apply to bosons.
The Lagrangian of the SM is invariant under local gauge transformations of the fields with respect to three symmetry groups:

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y, \tag{1.4}$$

where the subscripts C, L and Y stand for color, weak isospin and weak hyper-charge quantum numbers respectively. These symmetries accommodate the description of the particles interactions:

Electroweak interaction: it is the unified description of the electromagnetic interaction (the *Quantum Electrodynamics*, or QED) and the weak interaction. The unification is achieved under the $SU(2)_L \otimes U(1)_Y$ symmetry group, where all the particles are assigned a value of L_3 (third component of the weak isospin) and Y. The electroweak symmetry is spontaneously broken via the Higgs mechanism: $SU(2)_L \otimes U(1)_Y \to U(1)_Q$, where Q indicates the electric charge defined as $Q = \frac{1}{2}Y + L_3$, the only conserved quantum number. More details on the electroweak interactions and the Higgs mechanism are given in Sec. 1.2.

Strong interaction: it is described by the *Quantum Chromodynamics* (QCD) which relies on the $SU(3)_C$ symmetry group. In this scheme particles are assigned a *color* charge C, which can acquire three different values, labelled green, blue and red in analogy with the primary colors. Every observable state has to be a color singlet, hence particles carrying color come in bound states in order to satisfy this constraint.

The elementary particles of the SM are represented in Fig. 1.1 and are described in the following.

Bosons

Bosons are the mediators of the fundamental interactions, and are referred to also as *gauge bosons*. The known elementary bosons are three massive spin-1 (vector) bosons, the $W^{\pm}$ and Z, mediators of the weak force, and two massless spin-1 bosons, the photon (γ), mediator of the electromagnetic force, and the gluon (g), mediator of the strong force. In addition the Higgs boson (H) is a massive spin-0 (scalar) boson associated to the Higgs field, which ensures the spontaneous symmetry breaking required to make the $W^{\pm}$ and the Z massive.
The Lagrangian of a scalar boson field and its adjoint $\phi^{\dagger}(x)$ reads as:

$$\mathcal{L} = \partial_{\mu}\phi^{\dagger}\partial^{\mu}\phi - m^2\phi^{\dagger}\phi, \tag{1.5}$$

where μ spans over the four space-time indices and m is a parameter representing the particle mass. Using Eq. 1.1 the equations of motion, known as the Klein-Gordon equations, are obtained:

$$\begin{aligned} \left(\Box + m^2\right)\phi &= 0, \\ \left(\Box + m^2\right)\phi^{\dagger} &= 0, \end{aligned} \tag{1.6}$$

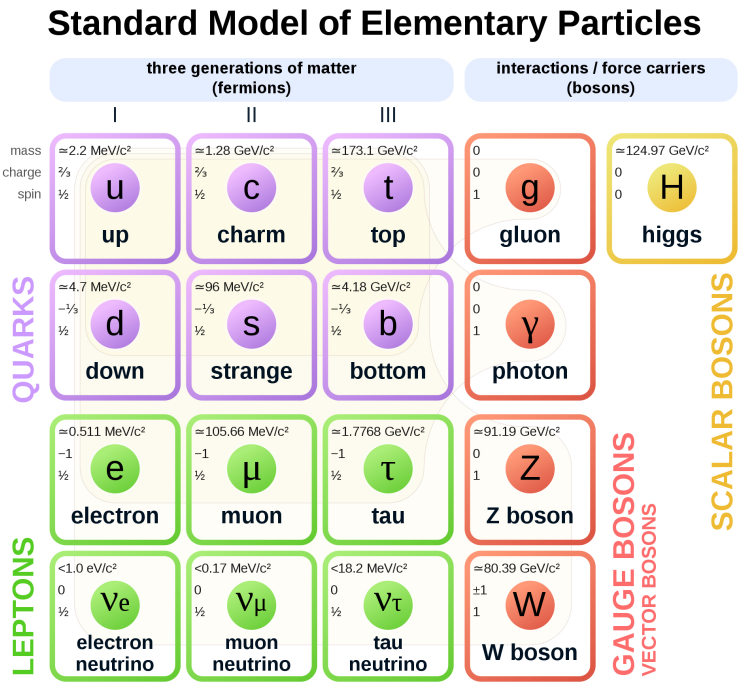


Figure 1.1: Elementary particles in the SM with some properties [19].

where $\Box \equiv \partial_\mu \partial^\mu$ is the d'Alembert operator. The solution of the Klein-Gordon equations, which describes the motion of the free scalar boson, reads as:

$$\phi(x) = \frac{1}{(2\pi)^{3/2}} \int \frac{\mathrm{d}\boldsymbol{p}}{\sqrt{2p^0}} \left(a(\boldsymbol{p}) e^{-ipx} + a^\dagger(\boldsymbol{p}) e^{ipx} \right), \tag{1.7}$$

where p^0, $\boldsymbol{p}$ and p are the temporal component, the spatial part and the total four-momentum respectively, and $a^{(\dagger)}$ is the annihilation (creation) operator. The generalization for the vector case is given by the Lagrangian:

$$\mathcal{L} = -\frac{1}{2} \left(\partial_\mu W_\nu^\dagger - \partial_\nu W_\mu^\dagger \right) \left(\partial^\mu W^\nu - \partial^\nu W^\mu \right) + m^2 W_\mu^\dagger W^\mu, \tag{1.8}$$

and the solutions of the equations of motion are:

$$\begin{aligned} W_\mu(x) &= \frac{1}{(2\pi)^{3/2}} \sum_\lambda \int \frac{\mathrm{d}\boldsymbol{p}}{\sqrt{2p^0}} \epsilon_\mu^\lambda(\boldsymbol{p}) \left(a(\boldsymbol{p}) e^{-ipx} + b^\dagger(\boldsymbol{p}) e^{ipx} \right), \\ W_\mu^\dagger(x) &= \frac{1}{(2\pi)^{3/2}} \sum_\lambda \int \frac{\mathrm{d}\boldsymbol{p}}{\sqrt{2p^0}} \epsilon_\mu^\lambda(\boldsymbol{p}) \left(b(\boldsymbol{p}) e^{-ipx} + a^\dagger(\boldsymbol{p}) e^{ipx} \right), \end{aligned} \tag{1.9}$$

where ϵ_μ^λ are the polarization vectors and the sum is taken over all the polarization states λ. The two different types of operators a and b reflect the fact that particles and anti-particles do not coincide.
The Lagrangian of a massless vector boson reduces to:

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu}, \tag{1.10}$$

with $F_{\mu\nu} = \partial_\nu A_\mu - \partial_\mu A_\nu$, and the solution of the equations of motion is:

$$\begin{aligned} A_0(x) &= 0, \\ A_i(x) &= \frac{1}{(2\pi)^{3/2}} \sum_\lambda \int \frac{\mathrm{d}\boldsymbol{p}}{\sqrt{2p^0}} \epsilon_i^\lambda(\boldsymbol{p}) \left(e^{-ipx} a_\lambda(\boldsymbol{p}) + e^{ipx} a_\lambda^\dagger(\boldsymbol{p}) \right), \end{aligned} \tag{1.11}$$

where the sum is taken over all the polarization states, which reduce to two in case of $m = 0$. The subscript i refers to the three spatial components of the field, while $A_0(x) = 0$ is the result of the particular gauge choice.

Fermions

Fermions are the building blocks constituting matter and are divided into *quarks*, which carry color charge and are therefore sensitive to the strong force, and *leptons*, which instead carry no color charge. Both leptons and quarks are divided into three families, in turn composed of two spin-1/2 particles each. A quark family is composed of a u-type (u, c or t) and a d-type (d, s or b) quark, with an electric charge of $+2/3$ and $-1/3$ respectively, while a lepton family is composed of a negative-charged lepton (e^-, μ^-, τ^-) and a neutral lepton, the neutrino (ν_e, ν_μ, ν_τ). Every fermion has a corresponding anti-particle with opposite quantum numbers.

Quarks are bound together by the strong force to form *hadrons*. In order to form color singlets, either quark - anti-quark pairs or bound states of three quarks exist, called respectively *mesons* and *baryons*. Moreover, exotic structures composed of four or five quarks [20] have been recently observed by the LHCb collaboration [21, 22].
The Dirac Lagrangian, describing the fermion field, is defined as follows:

$$\mathcal{L} = \bar{\psi}(x)(i\partial\!\!\!/ - m)\psi(x), \tag{1.12}$$

where $\partial\!\!\!/ \equiv \gamma^\mu \partial_\mu$, γ^μ are the Dirac gamma matrices, m is the mass of the fermion and $\bar{\psi}(x) \equiv \psi^\dagger \gamma^0$. By using Eq. 1.1 one obtains the equations of motion, known as the Dirac equations:

$$\begin{aligned} \left(i\partial\!\!\!/ - m\right) \psi(x) &= 0, \\ \bar{\psi}(x) \left(i\partial\!\!\!/ - m\right) &= 0. \end{aligned} \tag{1.13}$$

The solutions of the Dirac equations, which describe the free fermion, are:

$$\begin{aligned}\psi(x) &= \frac{1}{(2\pi)^{3/2}} \sum_{r=-\frac{1}{2},\frac{1}{2}} \int \mathrm{d}\boldsymbol{p} \sqrt{\frac{m}{p^0}} \left(a_r(\boldsymbol{p}) u_r(\boldsymbol{p}) e^{-ipx} + b_r^\dagger(\boldsymbol{p}) v_r(\boldsymbol{p}) e^{ipx}\right), \\ \bar{\psi}(x) &= \frac{1}{(2\pi)^{3/2}} \sum_{r=-\frac{1}{2},\frac{1}{2}} \int \mathrm{d}\boldsymbol{p} \sqrt{\frac{m}{p^0}} \left(b_r(\boldsymbol{p}) \bar{v}_r(\boldsymbol{p}) e^{-ipx} + a_r^\dagger(\boldsymbol{p}) \bar{u}_r(\boldsymbol{p}) e^{ipx}\right),\end{aligned} \tag{1.14}$$

where the sum is taken on the spin values r, $a_r^{(\dagger)}$ and $b_r^{(\dagger)}$ are the annihilation (creation) operators for fermions and anti-fermions respectively and u_r and v_r are 4-component *spinors*.
The Dirac field can be decomposed into components of left (ψ_L) and right (ψ_R) *chirality*:

$$\psi(x) = \underbrace{\frac{1-\gamma^5}{2}\psi(x)}_{\psi_L} + \underbrace{\frac{1+\gamma^5}{2}\psi(x)}_{\psi_R}, \tag{1.15}$$

where $\gamma^5 \equiv i\gamma^0\gamma^1\gamma^2\gamma^3$ and $\frac{1\pm\gamma^5}{2}$ are projection operators. In the limit $m = 0$ chirality and *helicity*, *i.e.* the projection of the spin along the momentum direction, coincide.

Particle interactions

The Lagrangians presented in Sec. 1.1 describe the behavior of the free, non-interacting fields. In the following the theory in the case of interacting particles is briefly described. Of particular interest for particle physics is the case in which particles are scattered. The initial state is described by the vector $|i\rangle$ long before the scattering happens, *i.e.* when $t_i = -\infty$, in which the particles are far apart and do not interact with each other. Quantum field theory allows to compute the probability for the state to evolve into the state $|f\rangle$ at $t_f = +\infty$, long after the particles have come close to each other, have scattered and have propagated far apart again. To calculate such a quantity one should solve the coupled non-linear equations of the fields. This is an extremely difficult problem, which can be solved only in perturbation theory, in which the Lagrangian is written as the sum of the one of the free fields plus an interaction term, which must be sufficiently weak to be treated as a perturbation.

The Hamiltonian density of a system is defined as:

$$\mathcal{H} = \pi(x)\dot{\phi}(x) - \mathcal{L}(\phi(x)), \tag{1.16}$$

where $\pi(x)$ is the conjugate momentum of the field ϕ. The energy of the system is then obtained integrating Eq. 1.16 over all space. If the Lagrangian can be decomposed into a free term and an interaction term, then the same applies for

the Hamiltonian density:

$$\mathcal{H} = \mathcal{H}_{\text{free}} + \mathcal{H}_I. \tag{1.17}$$

The amplitude of a scattering process is given by:

$$\mathcal{M} = \langle f|S|i\rangle, \tag{1.18}$$

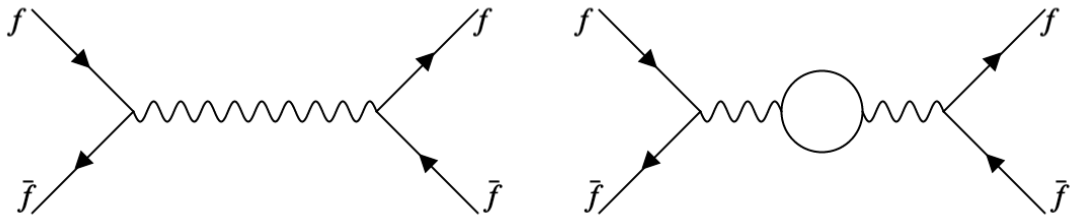
where S is called *S-Matrix* and is defined as:

$$S = I + \sum_{n=1}^{\infty} \frac{(-i)^n}{n!} \int \mathrm{d}^4x_1 \int \mathrm{d}^4x_2 ... \int \mathrm{d}^4x_n \mathrm{T}\{\mathcal{H}_I(x_1)\mathcal{H}_I(x_2)...\mathcal{H}_I(x_n)\}, \tag{1.19}$$

where I is the identity matrix, T indicates the time-ordered product of the fields and the integrals extend from $-\infty$ to $+\infty$ for the four space-time coordinates. The probability for the process to happen is then obtained as $|\mathcal{M}|^2$. Every term of the series in Eq. 1.18 can be associated to a diagram in a one-to-one correspondence by means of simple rules called *Feynman rules*, which allow to write down the expression of the scattering amplitude directly from the diagrams, avoiding tedious calculations of the S-Matrix terms. In these *Feynman diagrams* the propagating particles are represented as external lines, while energy and momentum exchanges are mediated by intermediate *virtual* particles represented as internal lines. The rules state that for each vertex in the diagram the amplitude of the process must be multiplied by the coupling constant of the interaction. Therefore the more "complex" the diagram is, the less it will contribute to the total amplitude. The simplest diagram, shown in Fig. 1.2(a), is obtained by joining two vertices with one virtual particle, and it corresponds to the first order term in perturbation theory, called *tree-level*. When computing higher order terms of the S-Matrix, *loops* are introduced in the diagrams. An example is shown in Fig. 1.2(b). Each loop is associated with a divergent integral. The process to eliminate such divergences and obtain finite results is called *renormalization*. It consists in redefining the values of some physical quantities, like masses and coupling constants, in terms of effective values measured in experiments, and transferring the divergences into the relations between the effective and the bare values of the non-interacting (and hence unobservable) particles.

1.2 The flavor sector

The theory of electroweak interactions was developed during the 60's with the aim of unifying the descriptions of the electromagnetic and weak interactions. The electromagnetic interactions are responsible for processes like the emission of light from an accelerated charge or the decay of the π^0 into two photons, while the weak interaction is responsible for processes like the β-decay ($n \to pe^-\overline{\nu}_e$) or the muon decay ($\mu^- \to e^-\nu_\mu\overline{\nu}_e$). The former affects all charged particles, while the



(a) (b)

Figure 1.2: Examples of a tree-level (a) and a higher order (b) Feynman diagrams.

latter affects all known elementary particles. The weak interaction violates some fundamental symmetries:

Parity (P): it is obtained by inverting all the spatial coordinates of a system. It is violated in weak processes [23] because the $W^{\pm}$ only couples with the left-handed component of the fermion fields;

Charge conjugation (C): it is obtained by exchanging all particles of the system in their associated anti-particles. The symmetry does not change the chirality of the particle, hence it is violated by the weak interaction. Both P and C are maximally violated in the SM;

CP symmetry: it is the product of parity and charge conjugation, and it was discovered to be violated at a very low level in some weak processes [24] (see Sec. 1.2). Because of the CPT theorem [25], that establishes the conservation of the CPT symmetry, obtained from the product of the CP symmetry and the time reversal (T), also the latter has to be violated in weak processes.

The part of the SM which describes the masses and mixing of quarks and leptons is called *flavor sector*.

The Higgs mechanism

It was observed experimentally that the weak force has a short range of interaction. This suggests that, contrary to the photon, the $W^{\pm}$ and Z bosons are massive particles. In fact, using the uncertainty principle [26] $\Delta E \Delta t \gtrsim \frac{\hbar}{2}$ with $\Delta t \sim \frac{r}{c}$ and $\Delta E = mc^2$ (using the mass of the Z), one gets $r \sim \frac{\hbar}{2mc} \sim 10^{-3}$ fm. However, adding a mass term to the Lagrangian causes an explicit breaking of the fundamental symmetries of Equation 1.4: a term such as $\frac{1}{2} M^2_{W^{\pm}} (W^{\pm}{}_{\mu})^2$, with $W^{\pm}{}_{\mu}$ the gauge field associated with the $W^{\pm}$ boson, is clearly not invariant under phase transformations. The mass terms are generated by means of the Higgs mechanism [27], which introduces a new scalar field ϕ to provide a *spontaneous symmetry breaking*, whose

Lagrangian is constructed from Eq. 1.5:

$$\mathcal{L} = \partial_\mu \phi^\dagger \partial^\mu \phi - V(\phi^\dagger \phi), \tag{1.20}$$

where

$$V(\phi^\dagger \phi) = -\mu^2(\phi^\dagger \phi) + \lambda(\phi^\dagger \phi)^2, \tag{1.21}$$

with $\mu > 0$ and $\lambda > 0$ two constants. The potential V has a local maximum for $\phi = 0$ and it is invariant under global phase transformations, being symmetric under rotation along the V axis. The minimum of the potential, which is interpreted as the ground state of the theory (the vacuum), is not unique and the field acquires a vacuum expectation value $v = \frac{\mu}{\sqrt{2\lambda}}$ as shown in Fig. 1.3. The Higgs field is a weak isospin doublet:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \begin{pmatrix} \frac{\pi_1 + i\pi_2}{\sqrt{2}} \\ \frac{\sigma + i\pi_3}{\sqrt{2}} \end{pmatrix}, \tag{1.22}$$

where the $^+$ and ^{0} stand for the electric charge and σ and π_i $(i = 1, 2, 3)$ are real fields, with

$$\sigma(x) = \tilde{\sigma}(x) + v. \tag{1.23}$$

The term $\tilde{\sigma}$ represents the fluctuations of the field in the radial direction, which

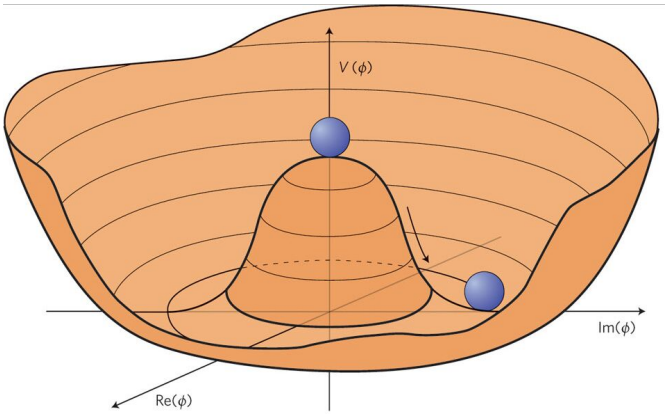


Figure 1.3: Schematic view of the Higgs potential [28].

are interpreted as the presence of a new particle, the Higgs boson. The π_i fields are instead massless spin-0 states called *Goldstone bosons*, whose presence in the theory is motivated by the spontaneous symmetry breaking, as stated by the Goldstone theorem [29].
The Lagrangian in Eq. 1.20 is invariant under global phase transformations, but to make it invariant under *local* gauge transformations it is necessary to replace the

usual derivative with the covariant derivative:

$$\partial_\mu \phi \to D_\mu \phi = (\partial_\mu - \frac{i}{2} g_w W_\mu^A \tau^A - \frac{i}{2} g_y B_\mu)\phi, \tag{1.24}$$

where τ^A (with $A = 1, 2, 3$) are the Pauli matrices, W_μ^A and B_μ are four complex vector fields and g_w and g_y are the coupling constants of the $SU(2)_L$ and $U(1)_Y$ groups respectively. Replacing Eq. 1.24 in Eq. 1.20 it is possible to extract the term

$$\mathcal{L}_{\text{mass}} = g_w^2 \frac{v^2}{8}[(W_\mu^1)^2 + (W_\mu^2)^2] + \frac{v^2}{8}(g_w W_\mu^3 - g_y B_\mu)^2. \tag{1.25}$$

These are mass terms related to the four fields, and by defining the physical fields as

$$\begin{aligned} W_\mu^\pm &\equiv \frac{W_\mu^1 \pm i W_\mu^2}{\sqrt{2}}, \\ Z_\mu &\equiv \frac{g_w W_\mu^3 - g_y B_\mu}{\sqrt{g_w^2 + g_y^2}}, \\ A_\mu &\equiv \frac{g_y W_\mu^3 + g_w B_\mu}{\sqrt{g_w^2 + g_y^2}}, \end{aligned} \tag{1.26}$$

the $W^\pm$ and the Z acquire masses equal to

$$\begin{aligned} M_{W^\pm}^2 &= g_w^2 \frac{v^2}{4}, \\ M_Z^2 &= g_w^2 \frac{v^2}{4}(1 + \frac{g_y^2}{g_w^2}) = M_{W^\pm}^2 / \cos^2 \theta_W, \end{aligned} \tag{1.27}$$

with $\cos\theta_W = \frac{g_w}{\sqrt{g_w^2 + g_y^2}}$, while the field A_μ associated with the photon remains massless. Applying some particular gauge transformations it is possible to eliminate the π_i fields of Eq. 1.22, whose degrees of freedom become the longitudinal polarizations of the massive bosons.

The Higgs mechanism provides also a way to give mass to leptons and quarks. Considering only one generation of leptons for simplicity, and since $Q = I + Y/2$, the quantum numbers are assigned in this way:

$$L = \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L^{Y=-1} \qquad R = (e)_R^{Y=-2}, \tag{1.28}$$

where L (R) is a doublet (singlet) of weak isospin and represents the left (right) component of the field. Since for the Higgs field $Y = 1$, the following term is an invariant:

$$\mathcal{L}_{\text{leptons}} = \mathcal{Y}_e(\bar{L}\phi R + \bar{R}\phi_c L), \tag{1.29}$$

where $\mathcal{Y}_e$ is the Yukawa coupling for electrons and $\phi_c = i\tau^2\phi^*$ (with τ^2 the second Pauli matrix). Considering only the terms due to the vacuum expectation value, the result is:

$$\mathcal{L}_{\text{leptons}} = \mathcal{Y}_e \frac{v}{\sqrt{2}}(\bar{e}_L e_R + \bar{e}_R e_L) = \mathcal{Y}_e \frac{v}{\sqrt{2}} \bar{e}e. \tag{1.30}$$

Analogously, for one generation of quarks the quantum numbers are:

$$Q_L = \begin{pmatrix} u \\ d \end{pmatrix}_L^{Y=1/3} \qquad (u)_R^{Y=4/3} \qquad (d)_R^{Y=-2/3}. \tag{1.31}$$

Therefore another invariant term to be added to the Lagrangian is:

$$\mathcal{L}_{\text{quarks}} = \mathcal{Y}_d(\bar{Q}_L \phi d_R + \bar{d}_R \phi_c Q_L) + \mathcal{Y}_u(\bar{Q}_L \phi_c u_R + \bar{u}_R \phi Q_L), \tag{1.32}$$

where $\mathcal{Y}_u$ and $\mathcal{Y}_d$ are the Yukawa couplings for up and down quarks respectively. Considering only the terms due to the vacuum expectation value, the result is:

$$\begin{aligned} \mathcal{L}_{\text{quarks}} &= \mathcal{Y}_d \frac{v}{\sqrt{2}}(\bar{d}_L d_R + \bar{d}_R d_L) + \mathcal{Y}_u \frac{v}{\sqrt{2}}(\bar{u}_L u_R + \bar{u}_R u_L) \\ &= \mathcal{Y}_d \frac{v}{\sqrt{2}} \bar{d}d + \mathcal{Y}_u \frac{v}{\sqrt{2}} \bar{u}u. \end{aligned} \tag{1.33}$$

Introducing the three families of quarks using the upper indices $i, j = 1, 2, 3$, the most general coupling to the Higgs field can be written as:

$$\mathcal{L}_{\text{quarks}} = \mathcal{Y}_d^{ij}(\bar{Q}_L^i \phi d_R^j + \bar{d}_R^i \phi_c Q_L^j) + \mathcal{Y}_u^{ij}(\bar{Q}_L^i \phi_c u_R^j + \bar{u}_R^i \phi Q_L^j). \tag{1.34}$$

Quark mixing and the CKM matrix

Fermions are assigned the following flavor quantum numbers:

Lepton flavor Every lepton is assigned a quantum number $L_l = 1$, where $l = e, \mu, \tau$ is the lepton family, while every anti-lepton is assigned $L_{\bar{l}} = -1$;

Quark flavor Four different quark quantum numbers are defined: *strangeness* (S), *charmness* (C), *bottomness* (B) and *topness* (T). For a given hadron they are assigned as:

- $\text{S} = -(n_s - n_{\bar{s}})$,
- $\text{C} = (n_c - n_{\bar{c}})$,
- $\text{B} = -(n_b - n_{\bar{b}})$,
- $\text{T} = (n_t - n_{\bar{t}})$,

where n_q ($n_{\bar{q}}$) is the (anti-)quark content of the hadron. Lepton flavor is conserved in the SM as a result of a so-called *accidental symmetry*. Therefore processes who violates lepton flavor conservation, like $J/\psi \to e^-\mu^+$, are forbidden and have never

been observed experimentally. Quark flavour is instead not conserved at tree-level in processes in which a $W^{\pm}$ is exchanged, like the $\Lambda \to p\pi^-$ decay.
In addition, a property of the SM called *lepton flavor universality* (LFU) states that the three families of leptons share the same weak coupling constant, while the only difference between them is their interaction with the Higgs field, which results in leptons having different masses. Therefore, theoretical predictions of physical observables do not depend on the nature of the lepton, and any eventual difference in final results is only due to the different values of their masses.

It was observed that processes involving a variation of strangeness $\Delta S = 1$, for example $K^- \to \mu^- \overline{\nu}_\mu$, are suppressed by a factor of ~ 20 with respect to processes with $\Delta S = 0$, for example $\pi^- \to \mu^- \overline{\nu}_\mu$. To preserve the universality of the weak interaction, the Cabibbo angle θ_c was introduced, allowing the mass eigenstates to be rotated with respect to the flavor eigenstates [30]. According to the theory, the three quarks known at the time (u, d, s), are organized into a doublet:

$$\begin{pmatrix} u \\ d_c \end{pmatrix} = \begin{pmatrix} u \\ d\cos\theta_c + s\sin\theta_c \end{pmatrix}. \tag{1.35}$$

In this way the $\Delta S = 1$ amplitudes are suppressed by a factor $\sin^2\theta_c$, while the $\Delta S = 0$ amplitudes are suppressed by a factor $\cos^2\theta_c$. *Flavor changing neutral currents* (FCNC), *i.e.* processes in which a quark decays into a different quark with same charge (for examples a $b \to s\ell\ell$ transition) were not observed experimentally. According to the Cabibbo theory however, these processes are allowed:

$$\begin{pmatrix} \overline{u} & \overline{d}_c \end{pmatrix} \begin{pmatrix} u \\ d_c \end{pmatrix} = \underbrace{\overline{u}u + (\overline{d}d\cos^2\theta_c + \overline{s}s\sin^2\theta_c)}_{\text{No flavor changing}} + \underbrace{(\overline{d}s + \overline{s}d)\sin\theta_c\cos\theta_c}_{\text{FCNC}}, \tag{1.36}$$

where the coupling constants and the space-time structure operator $\gamma^\mu(1-\gamma_5)$ are omitted.
A fourth quark (c) was introduced in order to cancel out the FCNC contributions [31]:

$$\begin{pmatrix} u \\ d_c \end{pmatrix} = \begin{pmatrix} u \\ d\cos\theta_c + s\sin\theta_c \end{pmatrix}, \quad \begin{pmatrix} c \\ s_c \end{pmatrix} = \begin{pmatrix} c \\ s\cos\theta_c - d\sin\theta_c \end{pmatrix}, \tag{1.37}$$

$$\begin{gathered} \begin{pmatrix} \overline{u} & \overline{d}_c \end{pmatrix} \begin{pmatrix} u \\ d_c \end{pmatrix} + \begin{pmatrix} \overline{c} & \overline{s}_c \end{pmatrix} \begin{pmatrix} c \\ s_c \end{pmatrix} = \\ \overline{u}u + \overline{c}c + (\overline{d}d + \overline{s}s)\cos^2\theta_c + (\overline{s}s + \overline{d}d)\sin^2\theta_c. \end{gathered} \tag{1.38}$$

Therefore these processes can only happen through higher order loops in perturbation theory and are hence suppressed in the SM.

The introduction of a third generation of quarks (b and t) is needed in order to

describe the observed CP violation in the weak sector. The Cabibbo mechanism is generalized using the Cabibbo-Kobayashi-Maskawa (CKM) matrix, which allows to write the expression for the quark couplings as:

$$J_W = \left(\overline{u}\ \overline{c}\ \overline{t}\right) V \begin{pmatrix} d \\ s \\ b \end{pmatrix}, \tag{1.39}$$

where V is the CKM matrix, a unitary N × N complex matrix, with N the number of quark generations, a free parameter of the SM. The V matrix can be seen also as the matrix needed to diagonalize the mass matrix in Eq. 1.34. Experimental measurements [32] have excluded the existence of a fourth-generation neutrino up to a mass of $\frac{m_Z}{2}$, where $m_Z \sim 90\,\text{GeV}$ is the mass of the Z boson, therefore constraining N to the value N = 3.
The number of free parameters of a unitary N × N complex matrix is:

$$(\mathrm{N}-1)^2 = \underbrace{\frac{1}{2}\mathrm{N}(\mathrm{N}-1)}_{\text{Mixing angles}} + \underbrace{\frac{1}{2}(\mathrm{N}-1)(\mathrm{N}-2)}_{\text{Complex phases}}. \tag{1.40}$$

According to Eq. 1.40 and using N = 3, three mixing angles and one complex phase appear. The CKM matrix can be written as [16]:

$$\begin{aligned} V = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} = \\ \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} = \\ \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix}, \end{aligned} \tag{1.41}$$

where $s_{ij} \equiv \sin\theta_{ij}$, $c_{ij} \equiv \cos\theta_{ij}$ and δ is the complex phase responsible for CP violation. Since it is known experimentally that $s_{13} \ll s_{23} \ll s_{12} \ll 1$, the matrix can be rewritten in a form that highlights this hierarchy, called Wolfenstein parametrization [33], by defining

$$\begin{aligned} \lambda = \sin(\theta_c) &= s_{12}, \\ A\lambda^2 &= s_{23}, \\ A\lambda^3(\rho - i\eta) &= s_{13}e^{i\delta} \end{aligned} \tag{1.42}$$

and

$$\begin{aligned}\bar{\rho} &= \rho(1-\lambda^2/2+\mathcal{O}(\lambda^4)),\\ \bar{\eta} &= \eta(1-\lambda^2/2+\mathcal{O}(\lambda^4)).\end{aligned} \tag{1.43}$$

The CKM matrix expressed in terms of λ, A, $\bar{\rho}$ and $\bar{\eta}$ is unitary to all orders in λ. It reads as:

$$V = \begin{pmatrix} 1-\lambda^2/2 & \lambda & A\lambda^3(\rho-i\eta) \\ -\lambda & 1-\lambda^2/2 & A\lambda^2 \\ A\lambda^3(1-\rho-i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4). \tag{1.44}$$

Since experimentally $\lambda = 0.22658 \pm 0.00044$ [16], the elements on the diagonal are approximately 1, while the couplings between quarks of different families become smaller as the difference in mass increases. From the unitarity of the matrix it follows that

$$\begin{aligned}\sum_k V_{ik}V_{jk}^* &= \delta_{ij},\\ \sum_k V_{ki}^*V_{kj} &= \delta_{ij}.\end{aligned} \tag{1.45}$$

The six vanishing combinations of Eq. 1.45 can be seen as constraints on the sides of triangles in the complex plane, called *unitary triangles*. The most widely used is described by the relation:

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0. \tag{1.46}$$

Fig. 1.4 shows the most updated values of the unitary triangle parameters. The measurements come from different sectors: the ϵ_K parameter is the CP violation in the $K^0 - \overline{K}^0$ mixing, while $\sin(2\beta)$ is related to the time-dependent CP asymmetry of neutral B decays. α is measured using $B \to \pi\pi$, $\rho\pi$ and $\rho\rho$ decay modes, while the γ parameter is measurable from tree level B decays. $|V_{ub}|$ is measured from inclusive $B \to X_u\ell\nu_\ell$, where X_u is any light hadron containing u quarks, using $B \to \tau\nu_\tau$, or from the ratio of $\Lambda_b \to p\mu^-\overline{\nu}_\mu$ and $\Lambda_b \to \Lambda_c^+\mu^-\overline{\nu}_\mu$ which allows to exctract $\left|\frac{V_{ub}}{V_{cb}}\right|$. Currently the measured values of the magnitudes of the CKM matrix elements are [16]:

$$V = \begin{pmatrix} 0.97401 \pm 0.00011 & 0.22650 \pm 0.00048 & 0.00361^{+0.00011}_{-0.00009} \\ 0.22636 \pm 0.00048 & 0.97320 \pm 0.00011 & 0.04053^{+0.00083}_{-0.00061} \\ 0.00854^{+0.00023}_{-0.00016} & 0.03978^{+0.00082}_{-0.00060} & 0.999172^{+0.000024}_{-0.000035} \end{pmatrix}, \tag{1.47}$$

while the values of the three angles and the complex phase are:

$$\begin{aligned} \sin\theta_{12} &= 0.22650 \pm 0.00048, \\ \sin\theta_{23} &= 0.04053^{+0.00083}_{-0.00061}, \\ \sin\theta_{13} &= 0.00361^{+0.00011}_{-0.00009}, \\ \delta &= 1.196^{+0.045}_{-0.043}. \end{aligned} \tag{1.48}$$

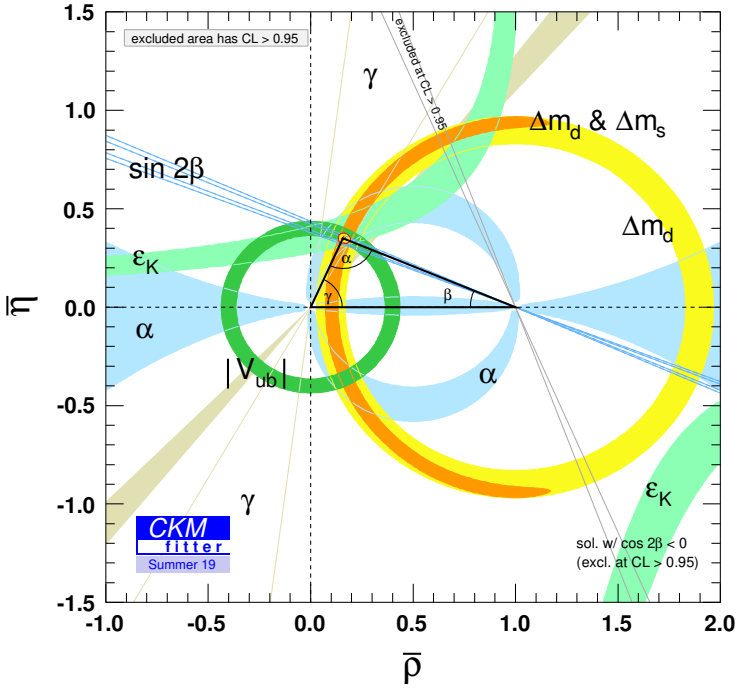


Figure 1.4: Constraints on the $(\bar{\rho}, \bar{\eta})$ plane [34].

Contrary to quarks, mixing effects for leptons are forbidden in the SM since the neutrinos are massless and only left-handed neutrino couplings are allowed. However, experimental measurements show that neutrino can oscillate between different flavor eigenstates [3, 4], which implies that the neutrino masses are different for the three families, and therefore that at least two of them must be non-zero. To account for neutrino oscillations the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) unitary matrix was introduced, which describes the mixing between mass and flavor eigenstates. One consequence is that lepton flavor violation can take place at loop level via neutrino oscillations, as shown in Fig. 1.5. However the rate of such processes is suppressed by a factor $(\frac{m_\nu}{m_W})^4 \sim \mathcal{O}(10^{-48})$, and it can be considered negligible in flavor physics studies.

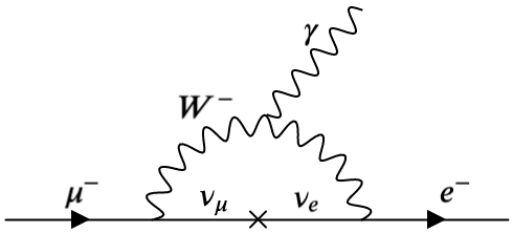


Figure 1.5: Feynman diagram for the lepton flavor violating decay $\mu^- \to e^-\gamma$. The neutrino oscillation allows the mixing between different flavors in the lepton sector.

1.3 Searches for new physics beyond the Standard Model

The search for new physics beyond the SM is motivated by the reasons exposed in Sec. 1.1 and consists of two main approaches: the *direct* and *indirect* searches. The former consists of directly detecting new particles, which could be either present in the environment (*e.g.* WIMPs and axions [35, 36]) or produced with particle accelerators (*e.g.* SUSY [37]); the latter consists instead of measuring the values of observables and comparing them to the theoretical predictions, aiming at finding discrepancies with the theory, which would indicate the presence of new physics. Of particular interest for this thesis are the indirect searches involving B meson decays. Among the six quarks, the b is the heaviest which hadronizes before decaying, allowing to study many decay channels. Moreover the relatively long mean lifetimes of B mesons make them particularly suitable to be investigated in particle accelerators.

Effective field theories

Theoretical predictions for B-hadron decays are challenging because of the multiple energy scales involved. The weak interaction responsible for the FCNC processes is governed by the electroweak scale $M_W \sim 80\,\text{GeV}$, while the strong interaction responsible for binding together the quarks in the final state hadrons is governed by the strong scale $\Lambda_{QCD} \sim 0.2\,\text{GeV}$. The energy scale of B hadron decays is approximately $m_b \sim 4\,\text{GeV}$, small compared to the electroweak scale but large compared to the strong scale. The two contributions can be factorized with the help of an effective field theory (EFT) approach. An example of EFT is the Fermi theory of β-decay: the neutron decay $n \to pe^-\overline{\nu}_e$ is mediated by the exchange of a W^-, but since the mass difference between proton and neutron $\Delta m = m_n - m_p \sim 1.3\,\text{MeV}$ is small compared to the electroweak scale, the presence of the W^- can be neglected and incorporated into an effective coupling constant, the Fermi constant G_F. The interaction is therefore described as a four-body interaction between the initial and

final state particles.

Modern EFT theories make use of the local *operator product expansion* (OPE) [38]: the FCNC processes are treated as point-like interactions and encoded in Wilson coefficients $\mathcal{C}_i(\mu)$ computed in perturbation theory, while the long-distance physics governed by QCD is described by effective operators $Q_i(\mu)$ calculated with non-perturbative methods like lattice QCD. The μ parameter indicates the energy scale, which is arbitrary and normally set at the value of the mass of the decaying particle $\mu = m_B$. Since this choice is arbitrary, a cancellation must happen in the EFT between the operators and the Wilson coefficients, so that the amplitude does not depend on μ. In this scheme, a B hadron decay amplitude reads as:

$$A(B \to f) = \langle f|\mathcal{H}_{eff}|B\rangle = \frac{G_F}{\sqrt{2}} \sum_i \lambda \mathcal{C}_i(m_B)\langle f|Q_i(m_B)|B\rangle, \tag{1.49}$$

where $\mathcal{H}_{eff}$ is the effective hamiltonian, the index i runs over all the possible pair of Wilson coefficients and effective operators entering the calculation of the amplitude and λ encodes the relevant CKM parameters.
The effective hamiltonian for rare B decays in the SM contains terms related to $b \to q\gamma$, $b \to q\ell^-\ell^+$ and $b \to \nu\overline{\nu}$ quark-level transitions, with $q = s$ or d. The Wilson coefficients of these operators are calculated from the Feynman diagrams in Fig. 1.6. The effective hamiltonian reads as:

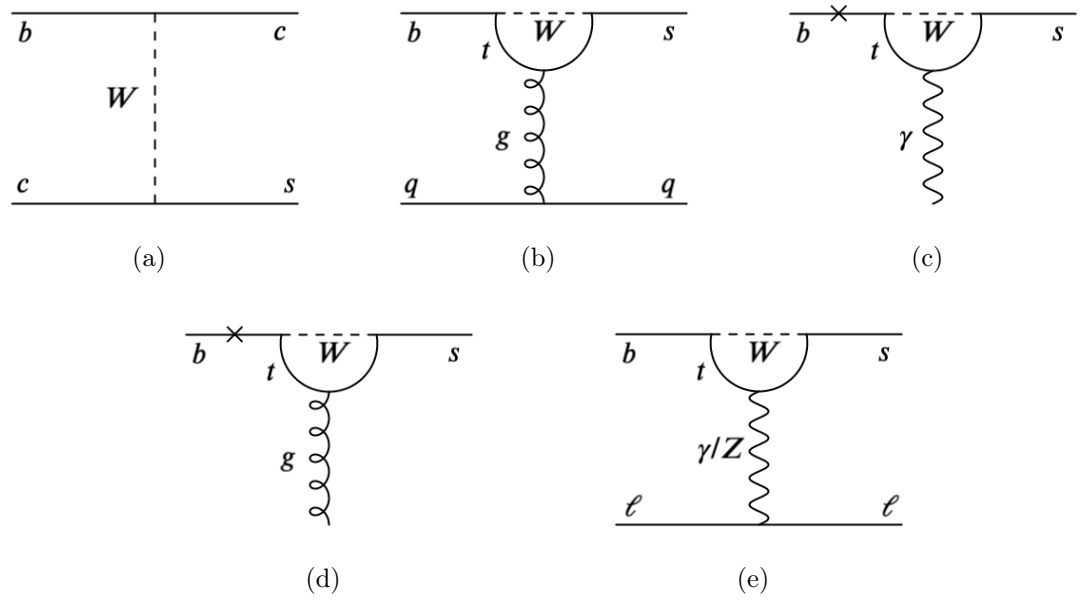


Figure 1.6: Feynman diagrams entering the calculation of the Wilson coefficients for $b \to s$ transitions in the SM. The cross indicates a chirality flip.

$$\mathcal{H}_{eff}^{b\to q} = \frac{4G_F}{\sqrt{2}}\left(\lambda_u^q \sum_{i=1}^{2}\mathcal{C}_i Q_i^u + \lambda_c^q \sum_{i=1}^{2}\mathcal{C}_i Q_i^c - \lambda_t^q \sum_{i=3}^{10}\mathcal{C}_i Q_i - \lambda_t^q \mathcal{C}_\nu Q_\nu\right), \tag{1.50}$$

where $\lambda_p^q = V_{pb}V_{pq}^*$, with p an up-type quark and a sum over the hermitian conjugate terms is implied. The operators in the sum are defined as follows:

- current-current operators (Fig. 1.6(a)):

$$Q_1^p = (\bar{q}_L\gamma_\mu T^a p_L)(\bar{p}_L\gamma^\mu T^a b_L), \tag{1.51}$$
$$Q_2^p = (\bar{q}_L\gamma_\mu p_L)(\bar{p}_L\gamma^\mu b_L); \tag{1.52}$$

- QCD penguins operators (Fig. 1.6(b)):

$$Q_3 = (\bar{q}_L\gamma_\mu b_L)\sum_p(\bar{p}\gamma^\mu p), \tag{1.53}$$
$$Q_4 = (\bar{q}_L\gamma_\mu T^a b_L)\sum_p(\bar{p}\gamma^\mu T^a p), \tag{1.54}$$
$$Q_5 = (\bar{q}_L\gamma_\mu\gamma_\nu\gamma_p b_L)\sum_p(\bar{p}\gamma^\mu\gamma^\nu\gamma^p p), \tag{1.55}$$
$$Q_6 = (\bar{q}_L\gamma_\mu\gamma_\nu\gamma_p T^a b_L)\sum_p(\bar{p}\gamma^\mu\gamma^\nu\gamma^p T^a p); \tag{1.56}$$

- electromagnetic and chromomagnetic dipole operators (Figs. 1.6(c), 1.6(d)):

$$Q_7 = \frac{e}{16\pi^2}m_b(\bar{q}_L\sigma^{\mu\nu}b_R)F_{\mu\nu}, \tag{1.57}$$
$$Q_8 = \frac{g_s}{16\pi^2}m_b(\bar{q}_L\sigma^{\mu\nu}T^a b_R)G^a_{\mu\nu}; \tag{1.58}$$

- semi-leptonic operators (Fig. 1.6(e)):

$$Q_9 = \frac{e^2}{16\pi^2}(\bar{q}_L\gamma_\mu b_L)\sum_l(\bar{\ell}\gamma^\mu\ell), \tag{1.59}$$
$$Q_{10} = \frac{e^2}{16\pi^2}(\bar{q}_L\gamma_\mu b_L)\sum_l(\bar{\ell}\gamma^\mu\gamma^5\ell), \tag{1.60}$$
$$Q_\nu = \frac{e^2}{8\pi^2}(\bar{q}_L\gamma_\mu b_L)\sum_l(\bar{\nu}_{\ell L}\gamma^\mu\nu_{\ell L}). \tag{1.61}$$

The $_L$ and $_R$ subscripts refer to left and right chirality projections, T^a are the generators of the $SU(3)$ group, e is the electric charge, g_s is the strong coupling constant, $F_{\mu\nu}$ is the electromagnetic tensor and $G^a_{\mu\nu}$ is the strong tensor.
The dominant contributions for $b \to q\ell^+\ell^-$ transitions are given by the semileptonic

operators Q_9 and Q_{10}, and the photon pole encoded in Q_7. The corresponding Wilson coefficients are computed at the energy scale $\mu = m_B$ and are usually expressed in the form of effective coefficients:

- $\mathcal{C}_7^{eff}(\mu) = \mathcal{C}_7(\mu) + \sum_{i=1}^{6} y_i \mathcal{C}_i(\mu)$,
- $\mathcal{C}_9^{eff}(\mu, q^2) = \mathcal{C}_9(\mu) + \sum_{i=1}^{6} z_i(q^2) \mathcal{C}_i(\mu)$,

where $y = [0, 0, -\frac{1}{3}, -\frac{4}{9}, -\frac{20}{3}, -\frac{80}{9}]$ and $z_i(q^2)$ are coefficients which depend on the transferred momentum of the dilepton system q^2.
The expression for the effective hamiltonian of $b \to q\ell^+\ell^-$ transitions reads as:

$$\mathcal{H}_{eff}^{b\to q\ell^+\ell^-} = \frac{G_F \alpha_e}{\sqrt{2}\pi} V_{tq}^* V_{tb} \left(\mathcal{C}_{9,q}^{eff}(q^2) Q_{9,q} + \mathcal{C}_{10,q}(\mu) Q_{10,q} - \frac{2i}{q^2} \mathcal{C}_{7,q}^{eff}(\mu) Q_{7,q} \right), \quad (1.62)$$

where α_e is the electromagnetic coupling constant and the subscript q in the Wilson coefficients and operators refers to the particular quark in the final state.
Besides $b \to q\ell^+\ell^-$ decays, theoretical predictions are computed for $B_q \to \ell^+\ell^-$ decays, which are forbidden at tree level in the SM, are CKM- and helicity-suppressed. The theoretical branching ratios of such processes read as:

$$\mathcal{B}(B_q \to \ell^+\ell^-) = \tau_{B_q} \frac{G_F^2 \alpha_e^2}{16\pi^2} f_{B_q}^2 m_l^2 m_{B_q} \sqrt{1 - \frac{4m_l^2}{m_{B_q^2}}} |V_{tb} V_{tq}^*|^2 |\mathcal{C}_{10}|^2, \quad (1.63)$$

where τ_{B_q} is the B meson lifetime, f_{B_q} is the fragmentation fraction of the B meson, and the term $\sqrt{1 - \frac{4m_l^2}{m_{B_q^2}}}$ is responsible for the helicity suppression. This class of decays is particularly clean since they only depend on Q_{10} and are free from QCD penguin diagram contributions. However, it is useful to rewrite Eq. 1.63 for the time-integrated decay [39]:

$$\bar{\mathcal{B}}(B_q \to \ell^+\ell^-) = \frac{1 + \mathcal{A} y_q}{1 - y_q^2} \mathcal{B}(B_q \to \ell^+\ell^-), \quad (1.64)$$

with

$$\begin{aligned} y_q &= \frac{\Gamma_H^q - \Gamma_L^q}{\Gamma_H^q + \Gamma_L^q}, \\ \mathcal{A} &= \frac{\Gamma_H^{q,\ell} - \Gamma_L^{q,\ell}}{\Gamma_H^{q,\ell} + \Gamma_L^{q,\ell}}, \end{aligned} \quad (1.65)$$

where $\Gamma_{L(H)}^q$ and $\Gamma_{L(H)}^{q,\ell}$ are the $B_q \to \ell^+\ell^-$ total and partial widths for the light (heavy) B mass eigenstates. In the following only time-integrated decays will be considered.

New Physics in EFT

New physics contributions could modify the values of the SM Wilson coefficients or could generate new operators which are not present in the SM. It is possible to write a large number of FCNC four-quark operators, however their effect enters as a higher-order correction and new physics effects are diluted by the presence of the current-current operator Q_2 which is generated at three-level. Instead, by considering dipole operators and operators containing two quarks and two leptons, it is possible to build a large class of models sensitive to new physics contributions. In general, these operators can be different for b or c quarks. The right-handed operators (not present in the SM) are indicated with a $'$ superscript. New physics dipole operators read as:

$$\begin{aligned} Q_{7,q}^{(\prime)} &= \frac{e}{16\pi^2} m_b (\bar{q}_{L(R)} \sigma^{\mu\nu} b_{R(L)}) F_{\mu\nu}, \\ Q_{8,q}^{(\prime)} &= \frac{g_s}{16\pi^2} m_b (\bar{q}_{L(R)} \sigma^{\mu\nu} T^a b_{R(L)}) G^a_{\mu\nu}. \end{aligned} \tag{1.66}$$

In new physics models, LFU could be violated and therefore the operators could be dependent on the specific lepton ℓ considered:

$$\begin{aligned} Q_{9,q}^{(\prime)\ell} &= \frac{e^2}{16\pi^2} (\bar{q}_{L(R)} \gamma_\mu b_{L(R)}) \bar{\ell} \gamma^\mu \ell, \\ Q_{10,q}^{(\prime)\ell} &= \frac{e^2}{16\pi^2} (\bar{q}_{L(R)} \gamma_\mu b_{L(R)}) \bar{\ell} \gamma^\mu \gamma^5 \ell, \\ Q_{\nu,q}^{(\prime)\ell} &= \frac{e^2}{8\pi^2} (\bar{q}_{L(R)} \gamma_\mu b_{L(R)}) \bar{\nu}_{\ell L} \gamma^\mu \nu_{\ell L}. \end{aligned} \tag{1.67}$$

New scalar (Q_S), pseudo-scalar (Q_P) and tensor (Q_T) operators not present in the SM read as:

$$\begin{aligned} Q_{S,q}^{(\prime)\ell} &= \frac{e^2}{16\pi^2} (\bar{q}_{L(R)} b_{R(L)}) \bar{\ell} \ell, \\ Q_{P,q}^{(\prime)\ell} &= \frac{e^2}{16\pi^2} (\bar{q}_{L(R)} b_{R(L)}) \bar{\ell} \gamma^5 \ell, \\ Q_{T,q}^{(\prime)\ell} &= \frac{e^2}{16\pi^2} (\bar{q}_{L(R)} \sigma^{\mu\nu} b_{R(L)}) \bar{\ell}_{L(R)} \sigma_{\mu\nu} \ell_{R(L)}, \end{aligned} \tag{1.68}$$

while in lepton flavor violating models, operators which couple with different lepton families ℓ_i and ℓ_j can be present:

$$Q_{9,q}^{(\prime)\ell_i\ell_j} = \frac{e^2}{16\pi^2} (\bar{q}_{L(R)} \gamma_\mu b_{L(R)}) \bar{\ell}_i \gamma^\mu \ell_j, \tag{1.69}$$

and analogously for $Q_{[10,S,P,T],q}^{(\prime)\ell_i\ell_j}$.
Relations between the different Wilson coefficients can be defined in case of models

with particular symmetries:

- in *minimal flavor violating* (MFV) models [40] $C_{i,s} = C_{i,d}$ and $C'_i \approx 0$;
- in constrained MFV models [41] $C^{(\prime)}_{S,P,T} = 0$ *i.e.* only SM operators are present;
- in models involving LFU $C^{(\prime)e}_{i,q} = C^{(\prime)\mu}_{i,q} = C^{(\prime)\tau}_{i,q}$;
- in models involving lepton flavor conservation $C^{(\prime),\ell_j\ell_k}_{i,q} = 0$;
- in models with weakly coupled new physics [42] $C^{\ell}_{S,q} = -C^{\ell}_{P,q}$, $C^{\prime,\ell}_{S,q} = C^{\prime\ell}_{P,q}$, and $C^{(\prime)\ell}_{T,q} = 0$.

The state of the art of experimental searches in the flavor sector

Rare and LFV dileptonic B decays

SM predictions for ${B^0}_{(s)} \to \ell^+\ell^-$ decays are computed using Eq. 1.64 and are reported in Table 1.1 [43]:

Table 1.1: SM predictions for ${B^0}_{(s)} \to \ell^+\ell^-$ decays. The first column indicates the nature of the B meson while the second, third and fourth columns provide the values of the branching ratios depending on the final state. The branching ratios of B^0 decay modes are lower than the corresponding B^0_s modes due to CKM suppression, while the effect of helicity suppression is visible by comparing the rates of different final states for the same B meson.

B meson	e^+e^-	$\mu^+\mu^-$	$\tau^+\tau^-$
B^0	$(2.48 \pm 0.21) \cdot 10^{-15}$	$(1.06 \pm 0.09) \cdot 10^{-10}$	$(2.22 \pm 0.19) \cdot 10^{-8}$
B^0_s	$(8.54 \pm 0.55) \cdot 10^{-14}$	$(3.65 \pm 0.23) \cdot 10^{-9}$	$(7.73 \pm 0.49) \cdot 10^{-7}$

The best experimental limits on the $B^0 \to e^-e^+$ and $B^0_s \to e^-e^+$ decays were set by the LHCb collaboration [44]:

$$\bar{\mathcal{B}}(B^0 \to e^-e^+) < 3.0 \cdot 10^{-9} \text{ at } 95\% \text{ CL}, \tag{1.70}$$
$$\bar{\mathcal{B}}(B^0_s \to e^-e^+) < 11.2 \cdot 10^{-9} \text{ at } 95\% \text{ CL}. \tag{1.71}$$

The expected branching ratios of these modes are highly suppressed from the helicity factor, which puts them beyond any present and near future experimental sensitivity (in case they should be SM-like).

The CMS and LHCb collaborations have published in 2014 the result of a joint analysis [45] presenting the first measurement of the $B^0_s \to \mu^+\mu^-$ branching ratio and the first evidence of the $B^0 \to \mu^+\mu^-$ decay. The LHCb collaboration has

updated the measurements in 2021 with additional data [46], obtaining the world best results from a single experiment:

$$\bar{\mathcal{B}}(B^0 \to \mu^-\mu^+) < 2.6 \cdot 10^{-10} \text{ at } 95\% \text{ CL}, \tag{1.72}$$
$$\bar{\mathcal{B}}(B^0_s \to \mu^-\mu^+) = \left(3.09^{+0.46+0.15}_{-0.43-0.11}\right) \cdot 10^{-9}, \tag{1.73}$$

where the first uncertainty is statistical and the second systematic.

The branching ratios of the ${B^0}_{(s)} \to \tau^+\tau^-$ decays are the highest among the three lepton families mainly because of the large value of the τ mass which mitigates the helicity suppression. However the experimental searches of modes with τ leptons are highly complex due to the many final states the τ decays into, reducing the overall reconstruction efficiency. Moreover at least one neutrino is present for each τ decay, which escapes the apparatus undetected. The first limit on the $B^0 \to \tau^+\tau^-$ branching ratio was set by the BaBar collaboration [47], but the world best limit on this mode and the first limit on the $B^0_s \to \tau^+\tau^-$ branching ratio were set by the LHCb collaboration [12]:

$$\bar{\mathcal{B}}(B^0 \to \tau^+\tau^-) < 2.1 \cdot 10^{-3} \text{ at } 95\% \text{ CL}, \tag{1.74}$$
$$\bar{\mathcal{B}}(B^0_s \to \tau^+\tau^-) < 6.8 \cdot 10^{-3} \text{ at } 95\% \text{ CL}. \tag{1.75}$$

Several searches for LFV decays, highly suppressed in the SM ($\mathcal{O}(10^{-48})$), have been performed and no evidence of such decays has ever been reported. The world best limits on the ${B^0}_{(s)} \to \tau^\pm\mu^\mp$ and ${B^0}_{(s)} \to e^\pm\mu^\mp$ branching fractions have been set by the LHCb collaboration [48, 49]:

$$\bar{\mathcal{B}}(B^0 \to \tau^\pm\mu^\mp) < 1.4 \cdot 10^{-5} \text{ at } 95\% \text{ CL}, \tag{1.76}$$
$$\bar{\mathcal{B}}(B^0_s \to \tau^\pm\mu^\mp) < 4.2 \cdot 10^{-5} \text{ at } 95\% \text{ CL}, \tag{1.77}$$
$$\bar{\mathcal{B}}(B^0 \to e^\pm\mu^\mp) < 1.3 \cdot 10^{-9} \text{ at } 95\% \text{ CL}, \tag{1.78}$$
$$\bar{\mathcal{B}}(B^0_s \to e^\pm\mu^\mp) < 7.2 \cdot 10^{-9} \text{ at } 95\% \text{ CL}. \tag{1.79}$$

More results on LFV modes can be found in [50, 51, 52].

Rare semileptonic B decays

Transitions of the type $B \to H\ell^+\ell^-$, where H is a meson containing an s or a d quark, are powerful probes for new physics beyond the SM. These transitions happen as $b \to q\ell^+\ell^-$ at quark level and only at higher order in perturbation theory, as described in Sec. 1.2. They are therefore sensitive to the hypothetical presence of new particles entering the loops, which could enhance the branching ratios with respect to the SM predictions. Contrary to ${B^0}_{(s)} \to \ell^+\ell^-$ decays, they suffer from larger theoretical uncertainties due to the evaluation of hadronic form factors, however they depend on different Wilson coefficients (as shown in Eq. 1.63) and allow to probe a larger set of new physics models.

In order to cancel out theoretical and experimental uncertainties, ratios of branching fractions of the form

$$R_H = \frac{\bar{\mathcal{B}}(B \to H\ell^+\ell^-)}{\bar{\mathcal{B}}(B \to H\ell'^+\ell'^-)} \tag{1.80}$$

are defined, where ℓ and ℓ' indicates two different lepton families. Moreover, because of LFU, when $\ell^{(\prime)} = e,\ \mu$ the ratio is very close to 1. Differences arise when considering τ decays, since the τ mass is considerably higher than the ones of electrons and muons. The $R_{K^{*0}}$ and R_K ratios are defined as follows:

$$R_{K^{*0}} = \frac{\bar{\mathcal{B}}(B^0 \to K^{*0}\mu^+\mu^-)}{\bar{\mathcal{B}}(B^0 \to K^{*0}e^+e^-)}, \tag{1.81}$$

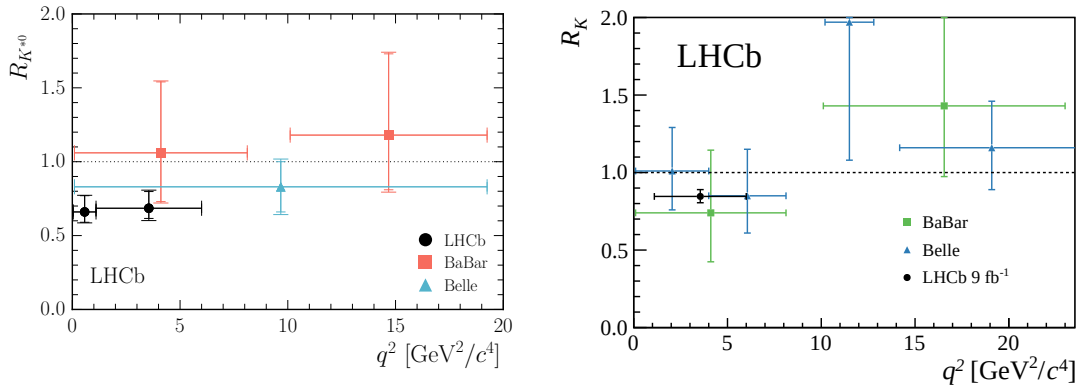
$$R_K = \frac{\bar{\mathcal{B}}(B^+ \to K^+\mu^+\mu^-)}{\bar{\mathcal{B}}(B^+ \to K^+e^+e^-)}. \tag{1.82}$$

The SM predictions for these ratios are 1 with an $\mathcal{O}(1\%)$ uncertainty [53].
The most precise measurements of such ratios have been performed by the LHCb collaboration [5, 6] and are shown in Fig. 1.7:

$$R_{K^{*0}} = \begin{cases} 0.66^{+0.11}_{-0.07} \pm 0.03 & \text{for } 0.045 < q^2 < 1.1\,\text{GeV}^2 \\ 0.69^{+0.11}_{-0.07} \pm 0.05 & \text{for } 1.1 < q^2 < 6.0\,\text{GeV}^2 \end{cases}, \tag{1.83}$$

$$R_K = 0.846^{+0.042\,+0.013}_{-0.039\,-0.012} \quad \text{for } 1.1 < q^2 < 6.0\,\text{GeV}^2. \tag{1.84}$$

The measurements show a tension with respect to the SM expectation of respectively $\sim 2.4\sigma$ and $\sim 3.1\sigma$.

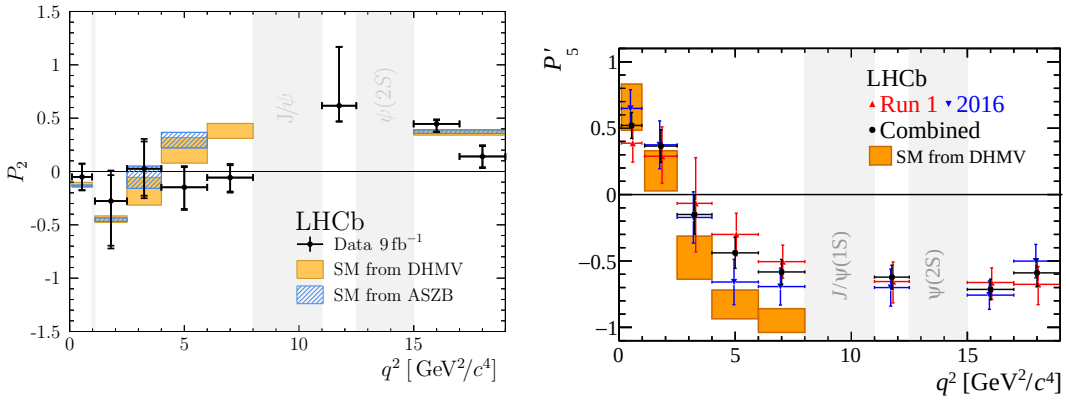


(a) (b)

Figure 1.7: Measurements of $R_{K^{*0}}$ (a) and R_K (b). The measurements by the Belle and BaBar collaborations are in agreement with the SM predictions but are less precise.

A different approach to the study of $B \to H\ell^+\ell^-$ transitions is represented by

angular analyses. Such decays are described by the di-muon invariant mass squared q^2 and a set of decay angles $\vec{\Omega}$. The angular distribution of the decay can be parametrized in terms of angular observables. The so-called *optimized observables* $P_i^{(\prime)}$ are defined as ratios of angular observables in such a way to minimize the effect of theoretical uncertainties. More details about angular observables can be found in [54, 55].
The results of angular analyses of the $B^0 \to K^{*0}\mu^+\mu^-$ and $B^+ \to K^{*+}\mu^+\mu^-$ decays [7, 8] are in agreement with the SM predictions for most of the angular observables, however local discrepancies are observed. The larger ones refer to the P_2 observable in the q^2 interval $6.0 - 8.0\,\text{GeV}^2$ measured in the $B^+ \to K^{*+}\mu^+\mu^-$ analysis and to the P_5' observable in the same q^2 interval measured in the $B^0 \to K^{*0}\mu^+\mu^-$ analysis. They are presented in Fig. 1.8 and they show a tension of respectively 3.0σ and 2.9σ with respect to the theoretical predictions computed in [56, 57].



(a) (b)

Figure 1.8: Experimental measurements of the P_2 (a) and P_5' (b) oservables for the $B^+ \to K^{*+}\mu^+\mu^-$ and $B^0 \to K^{*0}\mu^+\mu^-$ analyses respectively.

Tree-level semileptonic B decays

Analogously to Eq. 1.80, it is possible to define ratios of branching fractions for tree-level semileptonic B decays, governed by the $b \to c\ell^-\overline{\nu}_\ell$ quark-level transition. The $R(D^{(*)})$ ratios are defined as:

$$R(D) = \frac{\bar{\mathcal{B}}(B^0 \to D^-\tau^+\nu_\tau)}{\bar{\mathcal{B}}(B^0 \to D^-\mu^+\nu_\mu)}, \tag{1.85}$$

$$R(D^*) = \frac{\bar{\mathcal{B}}(B^0 \to D^{*-}\tau^+\nu_\tau)}{\bar{\mathcal{B}}(B^0 \to D^{*-}\mu^+\nu_\mu)}. \tag{1.86}$$

The SM predictions for the two ratios are [58]

$$\begin{aligned} R(D) &= 0.298 \pm 0.003, \\ R(D^*) &= 0.252 \pm 0.005. \end{aligned} \tag{1.87}$$

The reason of the difference with respect to unity is the value of the τ mass, which is substantially higher than the μ mass. A combination of the measurements of the two ratios is performed using inputs from various experiments, whose most updated values can be found in [9, 10, 11]. The result is shown in Fig. 1.9 and is equal to

$$\begin{aligned} R(D) &= 0.339 \pm 0.026 \pm 0.014, \\ R(D^*) &= 0.295 \pm 0.010 \pm 0.010. \end{aligned} \tag{1.88}$$

The result show a tension with respect to the SM prediction of about 3.4σ.

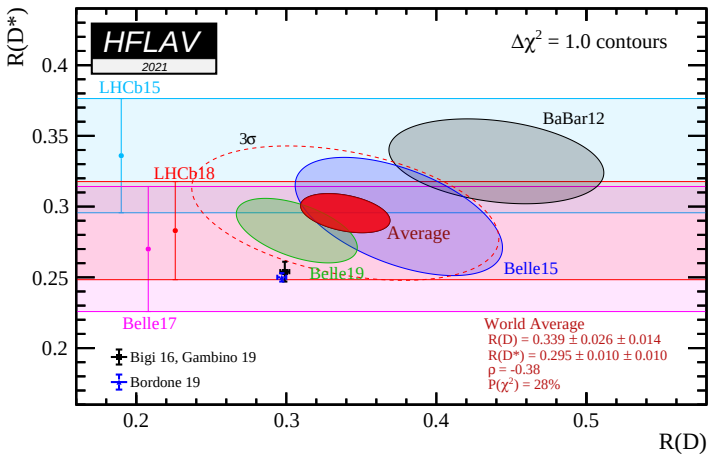


Figure 1.9: Combined measurements of the $R(D)$ and $R(D^*)$ ratios. The red oval shows the averaged measurement and is in tension with respect to the SM predictions of about 3.4σ [58].

The $R(J/\psi)$ ratio is defined as

$$R(J/\psi) = \frac{\bar{\mathcal{B}}(B_c^+ \to J/\psi \tau^+ \nu_\tau)}{\bar{\mathcal{B}}(B_c^+ \to J/\psi \mu^+ \nu_\mu)}, \tag{1.89}$$

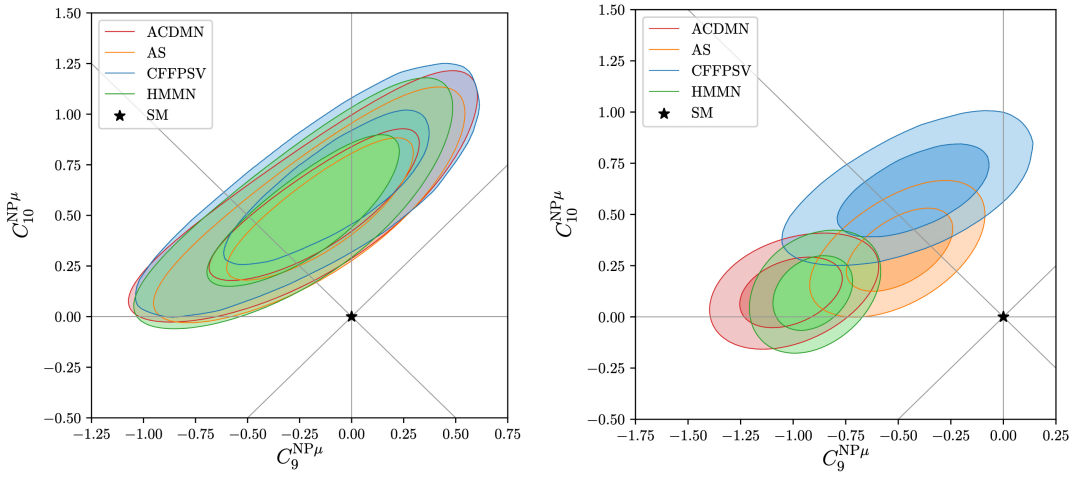
for which the predicted SM value lies in the range between 0.25 and 0.28, depending on the modeling approach for the form factors. The ratio has been measured by the LHCb collaboration [59]:

$$R(J/\psi) = 0.71 \pm 0.17 \pm 0.18, \tag{1.90}$$

where the first uncertainty is statistical and the second systematic. The result is in tension with respect to the SM predictions at the level of about 2σ.

Global fits

Experimental measurements are combined together with the use of *global fits*, which allow to take into account the effect of different measurements on the Wilson coefficients. Fig. 1.10 shows the result of the global fit in a scenario in which new physics contributions in $b \to s\ell^+\ell^-$ transitions are encoded into the $C^{\mu}_{9,s}$ and $C^{\mu}_{10,s}$ Wilson coefficients, performed by different theory groups [60, 61, 62, 63]. Fig. 1.10(a) shows the result of the global fit taking into account LFU



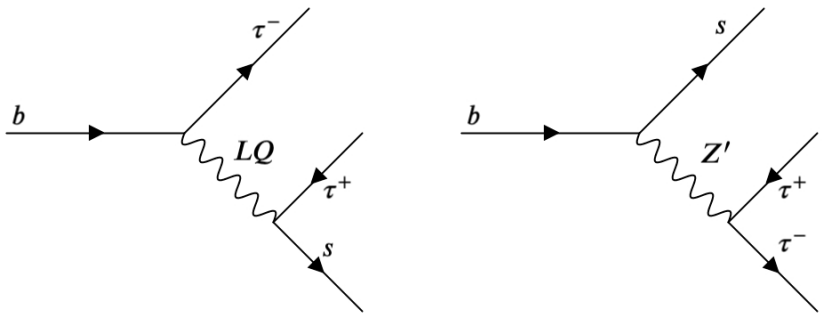
(a) (b)

Figure 1.10: Results of the global fit to LFU and $B_s^0 \to \mu^+\mu^-$ observables (a) and to a larger set of $b \to s\ell^+\ell^-$ observables including angular analyses (b). The notation $\mathcal{C}^{\mathrm{NP}\mu}_{9(10)}$ refers to the new physics contributions in the $\mathcal{C}^{\mu}_{9(10),s}$ Wilson coefficient.

observables and the measurements of the properties of $B_s^0 \to \mu^+\mu^-$. Fig. 1.10(b) shows the result of the global fit taking into account a larger set of $b \to s\ell^+\ell^-$ observables, including angular analyses. The fits performed by different groups show a remarkable agreement and a preference for new physics contributions. The differences in Fig. 1.10(b) are mainly due to different approaches used to compute non-local hadronic contributions.

Several new physics models have been developed in order to account for the observed anomalies in the flavor sector. Among the most cited ones are the *leptoquark* (LQ) and *heavy vector boson* (Z') models.
A LQ is a particle that carries baryonic and leptonic number at the same time, and therefore couples to both leptons and quarks. Moreover, they can be scalar or vector states and are specified by the set of quantum numbers $(SU(3)_C, SU(2)_L)_Y$.

A Z' boson is a neutral vector boson heavier than the Z boson which have different tree-level couplings with the three families of leptons and quarks.
Examples of $b \to s\ell^+\ell^-$ transitions involving a LQ and a Z' boson are shown in Fig. 1.11.



(a) (b)

Figure 1.11: Examples of tree-level $b \to s\ell^+\ell^-$ transitions mediated by a LQ (a) and a Z' boson (b).

1.4 Phenomenology of $b \to s\tau^+\tau^-$ decays beyond the Standard Model

The hints of new physics in semileptonic decays involving τ leptons in the final state suggest large LFU violation effects in $b \to s\tau^+\tau^-$ neutral currents, with enhancements of up to three orders of magnitude for their branching ratios to explain the central value of $R(D^{(*)})$ [15]. Currently, rare B decays with τ leptons in the final state are still poorly investigated. In fact, in addition to the limits on $\bar{\mathcal{B}}(B^0{}_{(s)} \to \tau^+\tau^-)$ in Eqs. 1.74 and 1.75, the available experimental results concern the searches for the $B^+ \to K^+\tau^+\tau^-$ and $B^0 \to K^{*0}\tau^+\tau^-$ decays by the BaBar and Belle collaborations respectively [13, 14]:

$$\bar{\mathcal{B}}(B^+ \to K^+\tau^+\tau^-) < 2.25 \cdot 10^{-3} \text{ at } 90\% \text{ CL}, \tag{1.91}$$
$$\bar{\mathcal{B}}(B^0 \to K^{*0}\tau^+\tau^-) < 2.0 \cdot 10^{-3} \text{ at } 90\% \text{ CL}. \tag{1.92}$$

This makes such class of decays particularly interesting, since not many constraints have been established and therefore large room for beyond the SM effects is still present.
The SM values for the branching fractions of such decays are computed using Eq. 1.62 considering only the contribution due to $Q^{\tau\tau}_{9,s}$ and $Q^{\tau\tau}_{10,s}$. Averaging over the

charged and neutral modes for $B \to K^{(*)}\tau^+\tau^-$ gives the results [15]:

$$\bar{\mathcal{B}}(B \to K\tau^+\tau^-)^{[15,22]}_{\rm SM} = (1.20 \pm 0.12) \cdot 10^{-7}, \tag{1.93}$$

$$\bar{\mathcal{B}}(B \to K^*\tau^+\tau^-)^{[15,19]}_{\rm SM} = (0.98 \pm 0.10) \cdot 10^{-7}, \tag{1.94}$$

$$\bar{\mathcal{B}}(B_s^0 \to \phi\tau^+\tau^-)^{[15,18.8]}_{\rm SM} = (0.86 \pm 0.06) \cdot 10^{-7}, \tag{1.95}$$

where the superscripts indicate the q^2 range for the dilepton system.
The new physics contributions are be encoded in the Wilson coefficients as

$$C^{\tau\tau}_{9(10),s} \approx C^{\tau\tau,\,\rm SM}_{9(10),s} - (+)\Delta, \tag{1.96}$$

where

$$\Delta = \frac{2\pi}{\alpha_e}\frac{V_{cb}}{V_{tb}V_{ts}^*}\left(\sqrt{\frac{R_x}{R_x^{\rm SM}}} - 1\right). \tag{1.97}$$

The expression for the Wilson coefficients depends on the ratio $\frac{R_x}{R_x^{\rm SM}}$, where R_x can be either $R(J/\psi)$ or $R(D^{(*)})$. Since a contribution to the SM coefficients $C^{\tau\tau,\,\rm SM}_{9,s}$ and $C^{\tau\tau,\,\rm SM}_{10,s}$ has as a result only a redefinition of the Fermi constant for $b\to s\ell^+\ell^-$ transitions, the $\frac{R_x}{R_x^{\rm SM}}$ ratios are independent of the particular choice made for x.
The factor multiplying the parentheses in Eq. 1.97 is $\frac{2\pi}{\alpha_e}\frac{V_{cb}}{V_{tb}V_{ts}^*} \sim 860$ and using the central value for $R(D^{(*)})$ the Δ factor turns out to be of $\mathcal{O}(100)$, which represents the dominant contribution in the Wilson coefficients, overwhelming completely the SM predictions. Neglecting the SM contributions, the branching fractions shown in Eqs. 1.93, 1.94 and 1.95 read as:

$$\bar{\mathcal{B}}(B \to K\tau^+\tau^-) = (8.8 \pm 0.8) \cdot 10^{-9}\Delta^2, \tag{1.98}$$

$$\bar{\mathcal{B}}(B \to K^*\tau^+\tau^-) = (10.1 \pm 0.8) \cdot 10^{-9}\Delta^2, \tag{1.99}$$

$$\bar{\mathcal{B}}(B_s^0 \to \phi\tau^+\tau^-) = (9.1 \pm 0.5) \cdot 10^{-9}\Delta^2. \tag{1.100}$$

Therefore such decays could be enhanced up to branching ratios of $\mathcal{O}(10^{-4})$.

Conclusions

Despite the extraordinary predictive power of the SM, some well established experimental facts and compelling theoretical arguments motivate the presence of a more fundamental theory of elementary particles. A set of experimental measurements in the flavor sector pointing towards a violation of LFU suggests that new physics particles could couple differently to the three families of leptons, and could manifest themselves by introducing differences in branching fractions of hadron decays involving different leptons in the final state. Some new physics models suggest that the third generation of leptons could be the most sensitive to

the presence of new physics. Moreover, B decays with τ leptons in the final state are still largely unexplored, which leaves lots of room for the presence of beyond the SM effects.

In the near future, many experimental results are expected to shed light on flavor anomalies, in particular the LHCb and Belle II experiments are expected to bring significant insights of the structure of the electroweak interaction.

Chapter 2

The LHCb experiment at the LHC

The search for the rare $B^0 \to K^{*0}\tau^+\tau^-$ decay is performed on data collected by the LHCb experiment [64, 65], a dedicated B and D physics experiment installed at the Large Hadron Collider (LHC) [66]. The facility is part of the European Organization for Nuclear Research (CERN), at the French-Swiss border near the city of Geneva.

A brief introduction to B physics at particle colliders is presented in Sec. 2.1, while Sec. 2.2 describes the LHC accelerator complex and Sec. 2.3 is dedicated to the description of the LHCb detector.

2.1 B physics at colliders

The main interest in studying B physics arises from the fact that the b quark is the heaviest one to hadronize before decaying, allowing for a plethora of physics studies. Moreover, B hadrons have a relatively long lifetime and form displaced decay vertices, a "signature" of their presence.

The number of b quarks produced at a particle accelerator (N_b) can be expressed as:

$$N_b = \sigma_{b\bar{b}} \int \mathcal{L}(t)\mathrm{d}t, \tag{2.1}$$

where $\sigma_{b\bar{b}}$ is the production cross-section of a $b\bar{b}$ pair and $L_{\text{int}} = \int \mathcal{L}(t)\mathrm{d}t$ is the integrated luminosity, *i.e.* the amount of data collected by the experiment in a given amount of time. The instantaneous luminosity $\mathcal{L}(t)$ refers to the amount of data per unit of time that the accelerator provides.
There exist two main classes of accelerators used to produce b quarks in large quantities: B factories and pp (or $p\bar{p}$) colliders.

B Factories

B factories collide e^+e^- pairs at the center-of-mass energies ($\sqrt{s}$) of the $\Upsilon(4S)$ and $\Upsilon(5S)$ resonances in order to maximize the number of produced B mesons. The main advantage of colliding elementary particles is a very clean environment, with a low amount of background. In such machines, the *trigger* system, *i.e.* the

system which decides whether a given event is interesting to the physics program and therefore recorded by the detector, has an efficiency of almost 100%. The main disadvantage of B factories is the relatively low $\sigma_{b\bar{b}}$, of $\sim 1\,\text{nb}$. Moreover, the $\sqrt{s}$ that can be reached at an e^+e^- circular accelerator is limited by the energy loss due to photon emission by light charged particles. In this type of accelerators, B hadrons are produced isotropically, therefore the detectors have full solid angle coverage which allows to account for missing energy due to neutral particles. To study time-dependent processes the two beams need to have different energies in such a way to boost the B hadron.
The main B factories are Belle at KEKB [67], which took data from 1999 to 2010, its successor Belle II at SuperKEKB [68], which started the data-taking in early 2019 and BaBar at PEP-II [69], which ran from 1999 to 2008.

pp ($p\bar{p}$) colliders

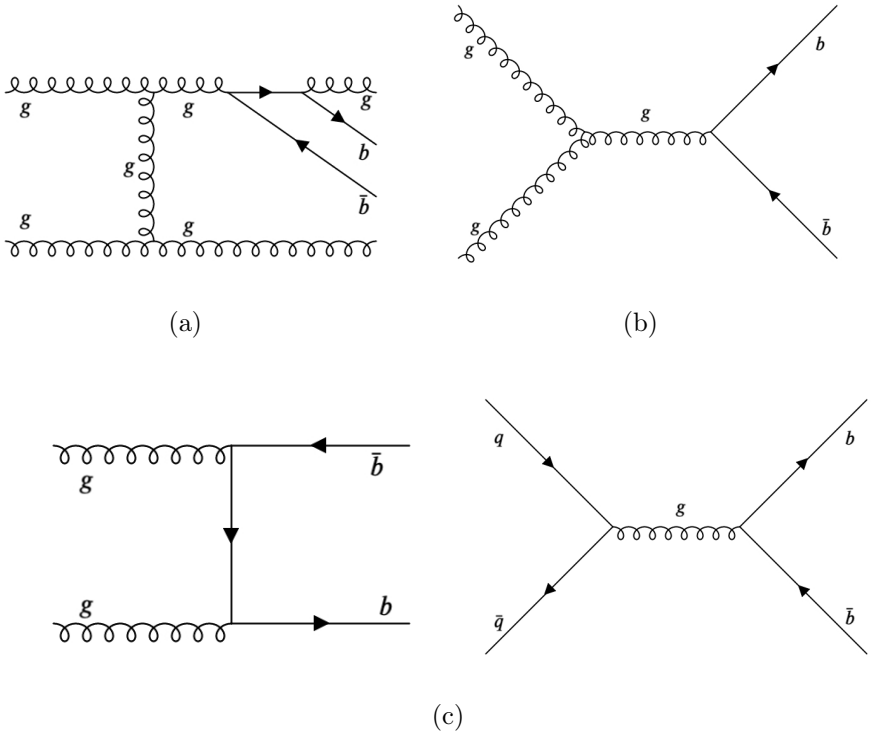
Proton-proton (and proton - anti-proton) colliders have the great advantage of running at energies well above the ones of e^+e^- accelerators, and since in such machines the $b\bar{b}$ production cross-section grows linearly with the energy, it reaches very high values of $\mathcal{O}(100\,\mu\text{b})$. In particular in the case of $\sqrt{s} = 7$ and 13 TeV, the cross-section is $(72.0 \pm 0.3 \pm 6.8)\,\mu\text{b}$ and $(144 \pm 1 \pm 21)\,\mu\text{b}$ [70]. Moreover, in $p\bar{p}$ collisions where the actual collision happens at the *parton* level, a large variety of B hadrons is produced, allowing the study of a very broad physics program. The main processes through which a $b\bar{b}$ pair is produced are shown in Fig. 2.1 and are:

- **flavor excitation**, which contributes with $\sim 54\%$ of the total production;
- **gluon splitting**, which contributes with $\sim 27\%$ of the total production;
- **pair production**, which contributes with $\sim 16\%$ of the total production.

Contrary to e^+e^- colliders, the hadron containing the $\bar{b}$ quark is not necessarily the antiparticle of the one containing the b quark.
The drawback of such an environment is the large amount of background produced by the parton interactions, which reduces the trigger performances.
The main experiments operating at hadron colliders are the D0 and CDF collaborations at Tevatron ($p\bar{p}$) [71, 72], who ran from 1985 to 2011, and LHCb, ALICE, ATLAS and CMS at the LHC (pp), which started data-taking in 2008.

2.2 The Large Hadron Collider

The LHC is a 26.7 km-long proton-proton accelerator installed in the tunnel previously built to host the Large Electron-Positron Collider (LEP), located between 45 and 170 m underground. The hadrons are accelerated in two separate series of cavities and are collided in four interaction points, where the main experiments are



(a) (b) (c)

Figure 2.1: Feynman diagrams for flavor excitation (a), gluon splitting (b) and pair production (c).

installed and take data. Since the tunnel is 3.7 m-wide, it was impossible to install two separate proton rings. Therefore the twin-bore magnet design was chosen, that allows to build a more compact structure but leads to the disadvantage that the two rings are magnetically coupled. Protons are accelerated to energies of 3.5 and 6.5 TeV, corresponding to $\sqrt{s} = 7$ and 13 TeV respectively.

The four experiments have different features and goals:

ATLAS and CMS: they are general-purpose detectors taking data at an instantaneous luminosity of $\mathcal{O}(10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1})$ [73, 74]. They have a broad scientific program and they share the merit of the discovery of the Higgs boson in 2012 [75, 76];

ALICE: it is a dedicated experiment for the study of lead-lead collisions taking data at a luminosity of $\mathcal{O}(10^{27}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1})$ [77];

LHCb: it is a dedicated b physics experiment taking data at a luminosity of $\mathcal{O}(10^{32}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1})$. It is described in detail in Sec. 2.3.

A schematic view of the CERN accelerator complex is shown in Fig. 2.2. Protons are extracted from hydrogen gas and are accelerated to an energy of 50 MeV in a

linear accelerator. They are then injected into the Proton Synchrotron Booster and accelerated to an energy of 1.4 GeV, before going into the Proton Synchrotron, where they reach an energy of 25 GeV, and into the Super Proton Synchrotron (SPS), in which are boosted to an energy of 450 GeV. Finally they are injected into the LHC where they are further accelerated and collided. Each beam is composed of a series of bunches, up to a maximum value of 2556, with about 10^{11} protons per bunch. The period of time between an injection and a beam dump is called *fill*. The nominal bunch crossing frequency is 40 MHz, which corresponds to a collision every 25 ns, but the effective collision rate is decreased due to some additional spacing between bunches needed in order to dump the beam safely. Protons are accelerated using RadioFrequency (RF) chambers, each delivering an electric field of 5 MV/m, and are bent using 1232 superconductive magnets, each delivering a magnetic field of 8.3 T.

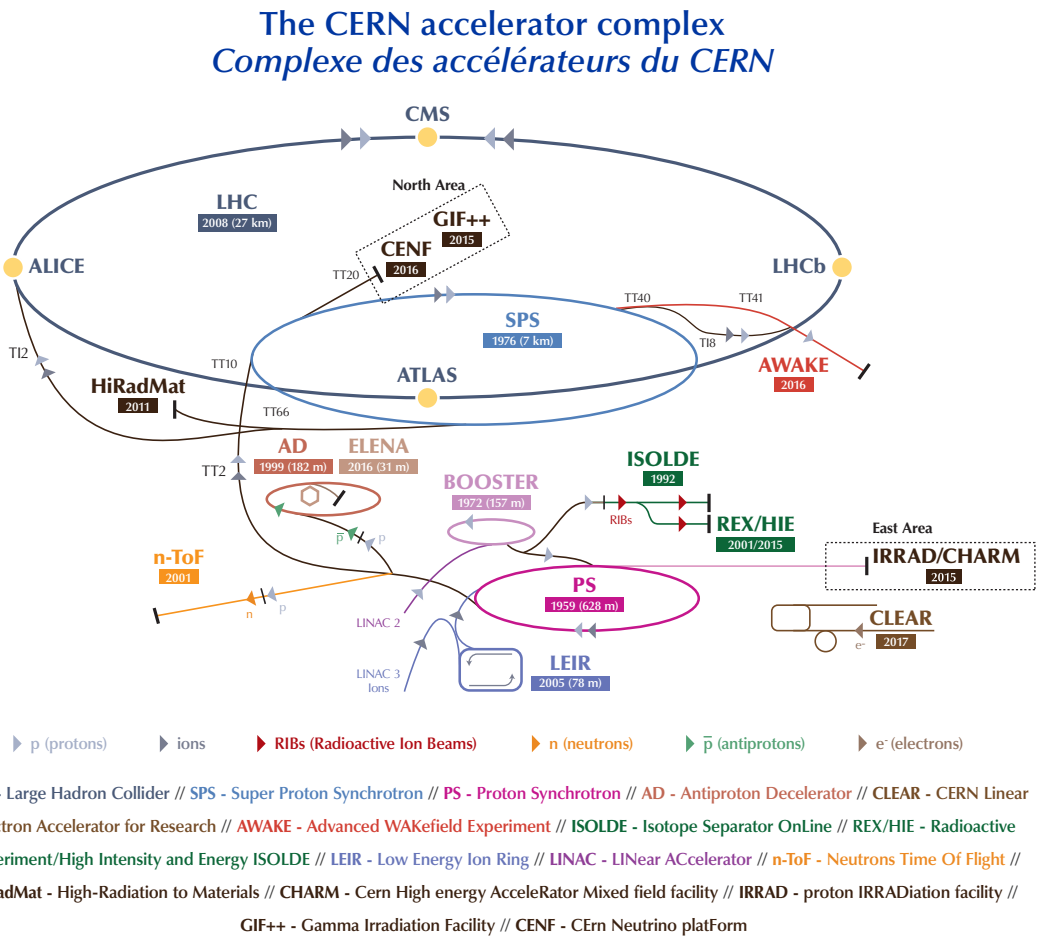


Figure 2.2: Schematic view of the CERN accelerator complex [78].

The expression for the instantaneous luminosity reads as:

$$\mathcal{L} = \frac{n^2 N_b f}{4\pi\sigma_{xy}} F, \tag{2.2}$$

where n is the number of particles in one bunch, N_b is the number of bunches in the accelerator, f is the revolution frequency, σ is the size of the bunches in the transverse plane and F is a factor taking into account the crossing angle between the two beams. The size of the bunches can be written as $\sigma_{xy} = \epsilon\beta$, where ϵ describes the geometry of the beam and β is determined by the magnet configuration. In order to maximize the luminosity, the beams must be squeezed in the transverse plane, *i.e.* β must be minimized. The minimum value for β is indicated with β^*. The LHCb experiment requires the instantaneous luminosity to be tuned in order to have about 1–2 collisions per bunch crossing, such that the detector is not overwhelmed with multiple interactions which would degrade its performance. In order to do so, a luminosity leveling mechanism controls the transverse spacing between two colliding bunches in order to keep the luminosity constant during a fill.

2.3 The LHCb experiment

The main goal of the LHCb experiment is to investigate CP-violating processes and study rare decays of beauty and charm hadrons. The apparatus is a single-arm spectrometer covering the region around the beam line from 10 to 300(250) mrad in the bending (non-bending) plane, defined as the plane in which particles are (not) deflected by the magnetic field, and corresponding to a pseudorapidity interval of $2 < \eta < 5$. This very peculiar geometry reflects the fact that most of the B hadrons are produced at small angles in the forward direction, as shown in Fig. 2.3. The integrated luminosity recorded from the detector is shown in Fig. 2.4. The period 2011–2012 is called "Run 1" while the period 2015–2018 is called "Run 2" (2015–2018). The analysis described in this thesis exploits the full LHCb dataset, from both Run 1 and 2.

The layout of the experiment is shown in Fig. 2.5. The main components are:

- the VErtex LOcator (VELO) used to precisely measure the position of displaced vertices;
- a dipole magnet providing the bending field for the spectrometer;
- the tracking stations used to perform measurements of the particles' momenta;
- two Ring Imaging CHerenkov (RICH) detectors performing particle identification;

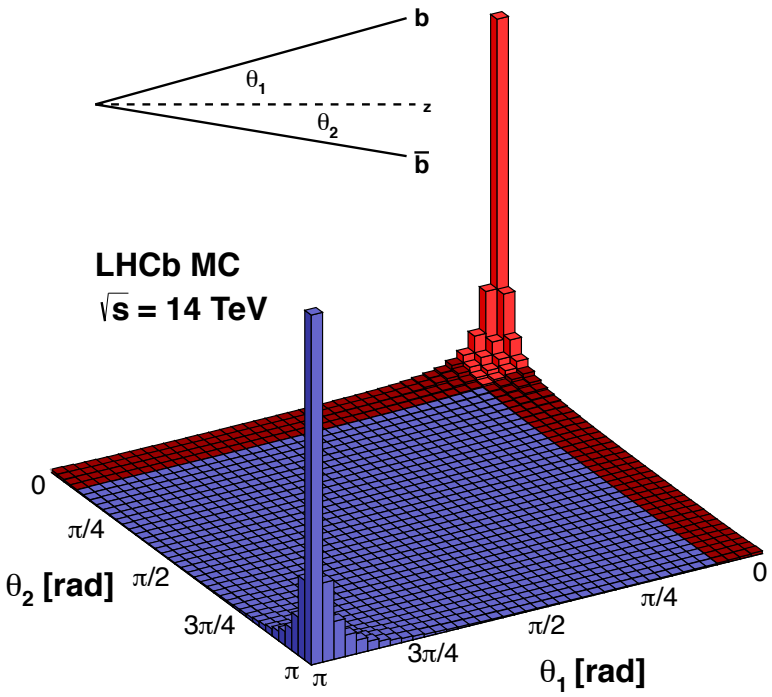


Figure 2.3: Simulated angular distribution of $b\bar{b}$ pairs produced at LHCb at $\sqrt{s} = 14\,\text{TeV}$. The red region indicates the LHCb acceptance.

- a calorimeter system used to provide trigger and particle identification information for hadrons, electrons and photons;
- a muon system providing trigger and particle identification information for muons.

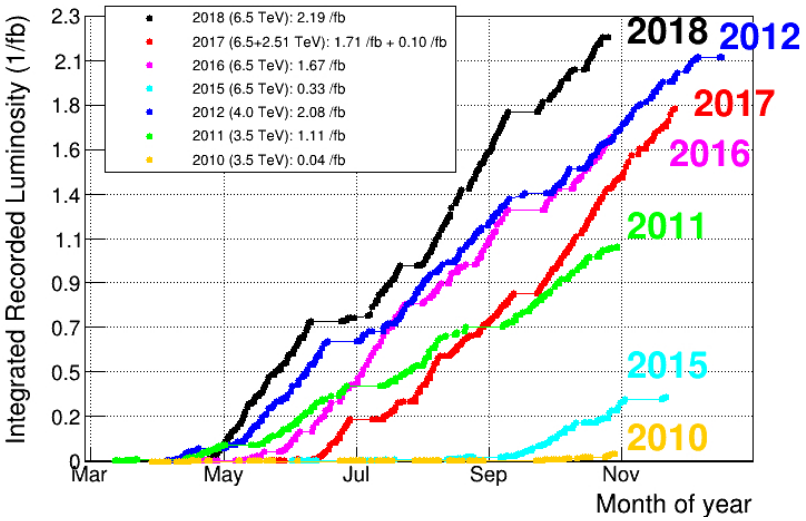


Figure 2.4: Integrated luminosity recorded by the LHCb experiment for each year of data-taking.

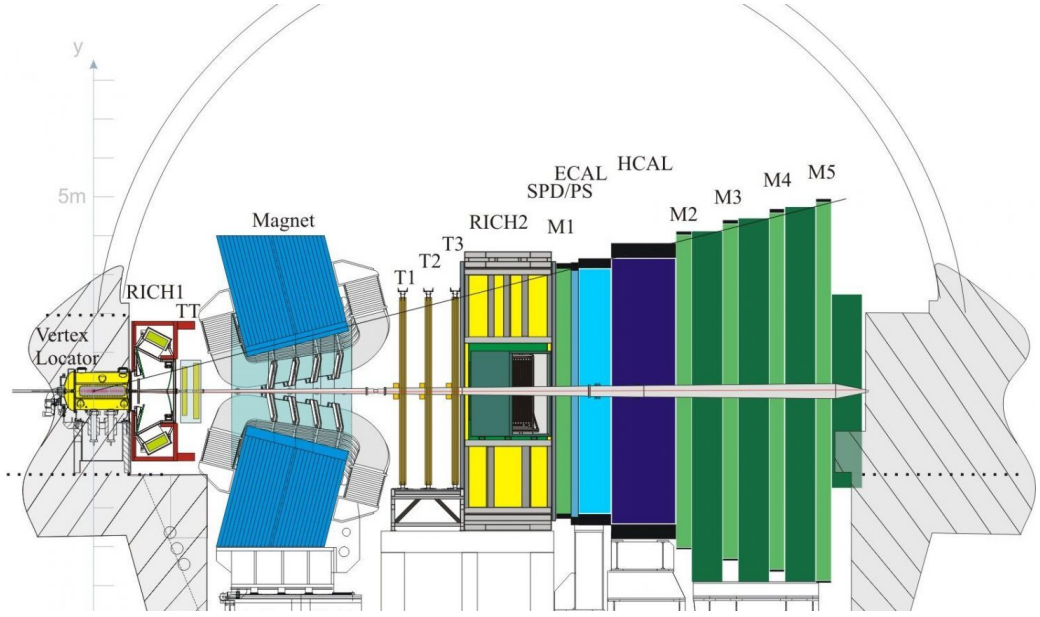


Figure 2.5: Layout of the LHCb detector.

Vertex Locator

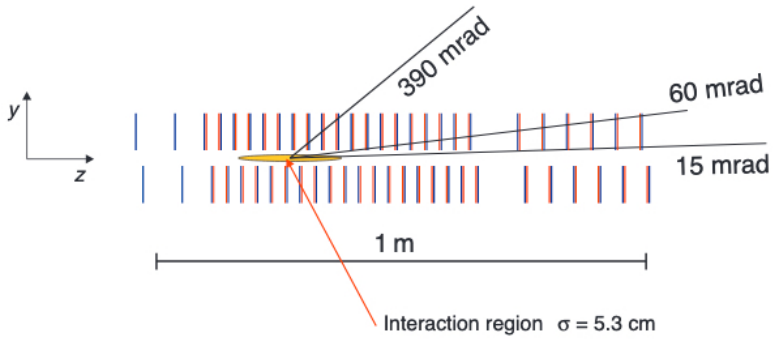
The VELO is a silicon tracker providing measurements of the position of charged particles very close to the interaction region. These measurements are of crucial importance since they allow to reconstruct the position of displaced decay vertices of B and D hadrons. The resolution on the hadron flight distance is $\sim 20\,\mu\text{m}$ for a $3\,\text{GeV}$ particle, which corresponds to an uncertainty of $\sim 50\,\text{fs}$ on the measure of its lifetime.
It is composed of 21 horizontal silicon modules covering the region $z = -18 - 80\,\text{cm}$ along the beam direction z, where $z = 0$ indicates the interaction point, and $R = 8 - 42\,\text{mm}$, where R indicates the radial distance from the beam axis. Each module is composed of two silicon layers, the R-sensor and the ϕ-sensor, which measure the radial distance from the beam and the azimuthal position around the beam respectively. The pitch of the R-sensor increases linearly with the distance from the beam axis in the range $38 - 102\,\mu\text{m}$. The ϕ-sensor is divided in an inner and an outer part. The former covers the region $8 - 17\,\text{mm}$ from the beam axis and its pitch increases linearly in the range $38 - 78\,\mu\text{m}$, while the latter covers the region $17 - 42\,\text{mm}$ with a pitch in the range $39 - 97\,\mu\text{m}$. The sensors are $300\,\mu\text{m}$-thick and they overlap, ensuring the full coverage of the azimuthal angle. A schematic view of the VELO is presented in Fig. 2.6.
The VELO is split in two halves which are brought close together during the data-taking period, in the so-called *closed* configuration. This allows the VELO to reach the innermost position close to the proton beam. To preserve the integrity of the detector and to reduce the radiation damage, the two halves are moved at a distance of $6\,\text{cm}$ with respect to each other during the beam injection and acceleration, the so-called *open* configuration.

Magnet

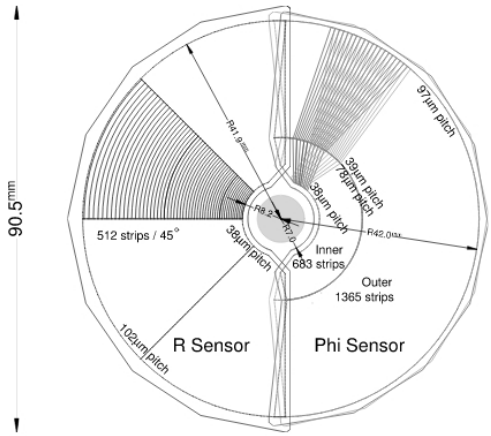
The dipole magnet is used to bend charged particles in the horizontal plane in order to measure their momenta. It is composed of two saddle-shaped coils supported by an iron yoke. Contrary to the other major LHC experiments, the LHCb magnet is a warm magnet, allowing the polarity to be inverted. This is done periodically, such that data are taken with both polarities in order to reduce systematic uncertainties that may arise because of the left-right asymmetry. The integrated magnetic field produced by the magnet is equal to $4\,\text{T m}$ for tracks of $10\,\text{m}$ length. The field mapping is performed using 60 sensors in order to achieve the required momentum resolution.

Tracking stations

Four tracking stations are used to perform precise measurements of particles' momenta: the Tracker Turicensis (TT), positioned before the magnet (upstream tracker), and the T1, T2 and T3 stations, positioned after the magnet (downstream



(a)



(b)

Figure 2.6: Schematic representation of the VELO (a) and of one of the silicon modules (b).

tracker). Each of the downstream stations is divided in two regions: the Inner Tracker (IT) and the Outer Tracker (OT). The IT and the TT are based on silicon strips and together form the Silicon Tracker (ST), while the OT uses straw tubes. Even though the IT represents only 1.3% of the total IT and OT region, it contains about 20% of the total tracks because of the much greater flux of particles in the region close to the beam pipe. Each of the ST stations is composed of four layers: the first and the fourth have vertical stripes, while the second and the third have strips rotated by $+5°$ and $-5°$ respectively, in order to measure the position on the perpendicular plane.
The measured hits are translated to actual tracks by means of a fitting algorithm based on a Kalman filter, which combines the information on the momentum and its resolution, the geometry of the detector and the magnetic field.
A measured track belongs to one of the following categories (Fig. 2.7):

long tracks: tracks which have hits in the VELO and in all the tracking stations. The momentum measurement is very precise;

VELO tracks: tracks which have hits only in the VELO. They are used to measure the position of the interaction point, the so-called *primary vertex* (PV). In general more than one PV can be present in an event, and the closest one to the track associated with the decay head is called "best" PV. Throughout this thesis the term PV refers to the best PV, unless differently specified;

upstream tracks: low-momentum tracks which have hits in the VELO and in the TT and are bent away from the magnet;

downstream tracks: tracks that have hits in all the tracking stations but not in the VELO. They are used to study long-lived particles;

T tracks: tracks that have hits only in the tracking stations after the magnet. They are due to long-lived particles or material interactions.

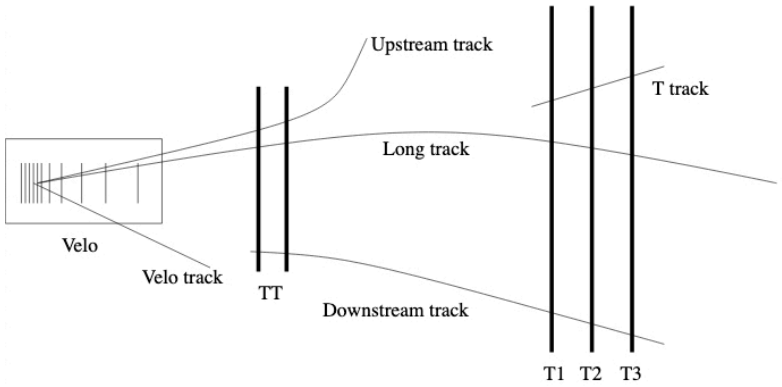


Figure 2.7: Schematic representation of the different types of tracks measured by the LHCb detector.

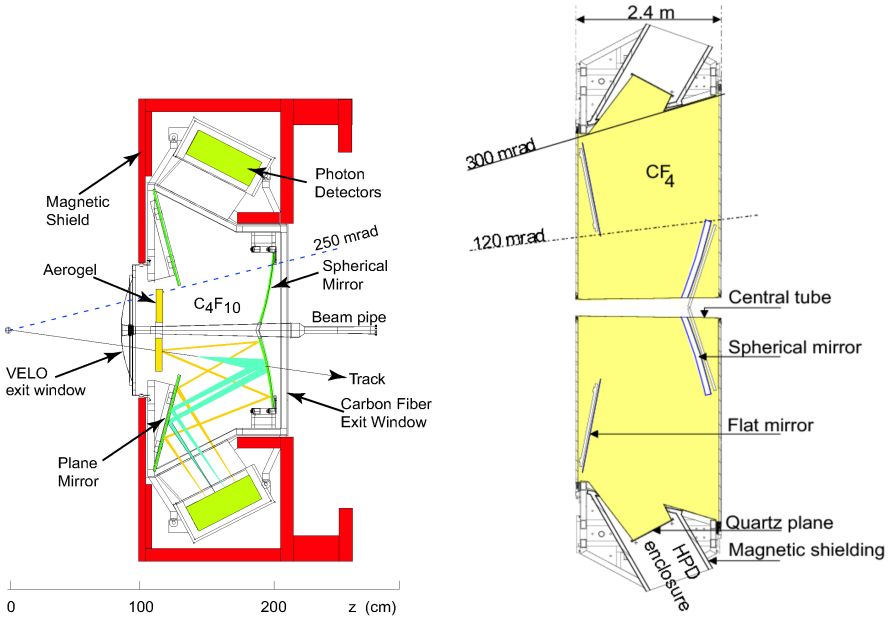
The resolution of the overall tracking system on the momentum of long tracks varies from 0.4% at 5 GeV to 1.0% at 200 GeV, while the resolution on the impact parameter is $(15 + 29\,\mathrm{GeV}/p_\mathrm{T})\,\mu\mathrm{m}$, where p_T is the transverse momentum in GeV. All the tracks considered in the analysis described in this thesis are long tracks.

RICH

The two RICH detectors provide particle identification information using Cherenkov light emitted by a particle traversing a medium with a speed greater than the speed of light in that medium. The Cherenkov light is emitted in a cone, whose aperture is related to the particle speed v via the formula:

$$\cos\theta = \frac{1}{\beta n}, \tag{2.3}$$

where $\beta = v/c$, c is the speed of light in a vacuum, n is the refraction index of the medium and θ is the Cherenkov angle, namely the angle between the particle's momentum and the plane orthogonal to the wavefront of the Cherenkov radiation. The photons emitted are reflected out of the spectrometer acceptance using a combination of spherical and flat mirrors, and are detected using Hybrid Photon Detectors (HPDs) in the wavelength range $200 - 600\,\text{nm}$. The ring resolution is proportional to $\Delta\theta/\sqrt{N}$ where $\Delta\theta$ is the uncertainty on θ and N is the number of photons in the ring. The information on the measured Cherenkov angle is used to perform a likelihood fit used to assign a mass hypothesis to the particle.
The upstream detector, RICH1, is located before the TT and covers the full LHCb solid angle and the low momentum range $\sim 1 - 60\,\text{GeV}$ using aerogel and C_4F_{10} as radiators. RICH2 is located after the tracking stations and covers the momentum range $\sim 15 - 100\,\text{GeV}$, using CF_4 as radiator. A schematic view of the two RICH detectors is presented in Fig. 2.8, while in Fig. 2.9 the Cherenkov angle as a function of the momentum is shown for different particles. Fig. 2.10 shows the performance of the RICH identification measured with background-subtracted data collected in 2012 and 2016.



(a) (b)

Figure 2.8: Representation of RICH1 (a) and RICH2 (b).

Calorimeters

The calorimeter system is composed of multiple layers as shown in Fig. 2.11. It measures the incoming particle's energy and mainly provides particle identification

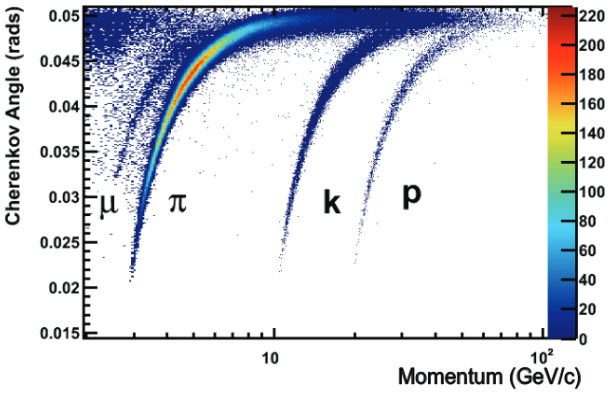
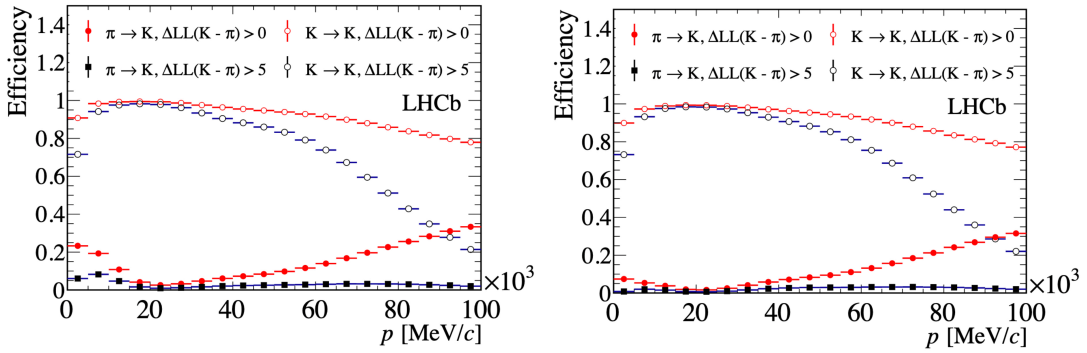


Figure 2.9: Cherenkov angle as a function of the particle's momentum in the C_4F_{10} radiator in data.



(a) (b)

Figure 2.10: Efficiency (empty dots) and fake rate (filled dots) of the RICH system as a function of the particle momentum measured in 2012 (a) and 2016 (b) for two different particle identification requirements.

and trigger information. It is composed of the Scintillator Pad Detector (SPD), the Pre-Shower Detector (PS), the Electromagnetic Calorimeter (ECAL) and the Hadronic Calorimeter (HCAL). The SPD/PS system is composed of two identical planes of rectangular scintillator pads of high granularity separated by a lead converter of 2.5 X_0, where $X_0 \approx 0.56\,\text{cm}$ is the radiation length for lead. It is 7.6 m-wide and 6.2 m-high and the scintillation light is transmitted via wavelength-shifting (WLS) fibers to multianode photomultiplier tubes, located outside the detector acceptance. Its main purpose is to distinguish between electron and neutral pions or photons.

The ECAL is composed of 66 layers of lead absorber 2 mm-thick and scintillator tiles 4 mm-thick, readout using WLS fibers. It is 42 cm-long, corresponding to 25 X_0, with a Molière radius of 3.5 cm. The outer acceptance is 300 (250) mrad in

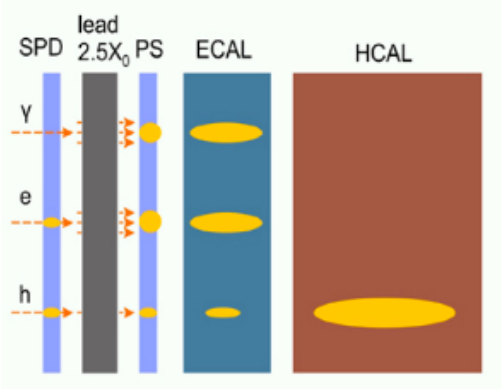


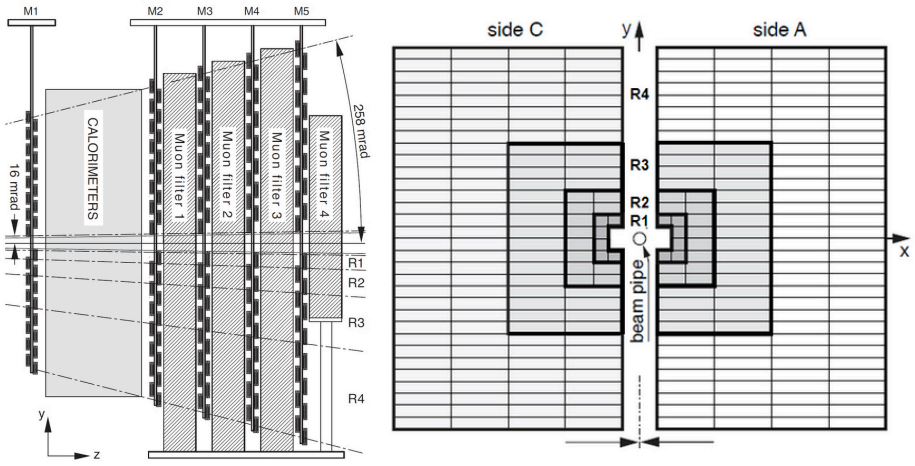
Figure 2.11: Schematic representation of the different layers of the calorimeter system, with examples of interaction by photons (γ), electrons (e) and hadrons (h).

the bending (non-bending) plane, while the inner is greater than 25 mrad, in order to reduce the radiation damage. The energy resolution is $\sigma_E/E = 10\%/\sqrt{E} \oplus 1\%$, with E in GeV.
The HCAL is composed of layers of iron and scintillating tiles, 16 mm and 4 mm-thick respectively. The overall structure measures 8.4 m in height, 6.8 m in width and 1.65 m in depth, the latter corresponding to $5.6\lambda_I$, where λ_I is the hadronic interaction length of iron. The energy resolution is $\sigma_E/E = 69\%/\sqrt{E} \oplus 9\%$, with E in GeV.

Muon stations

The muon stations provide highly efficient muon identification and trigger, a feature of the LHCb experiment which allows to perform some of its most important measurements. Given the high penetrating power of muons, the muon system is placed at the end of the experimental apparatus. It is composed of five stations (M1–5), one of them (M1) placed before the calorimeters in order to improve p_{T} measurements for the trigger and four of them (M2–5) located after the calorimeters and separated by iron blocks 80 cm-thick. The minimum momentum necessary for a muon to penetrate all the five stations is approximately 6 GeV and the track is required to have aligned hits in every station above a certain p_{T} threshold in order to be identified as a muon (at the L0 trigger level, see Sec. 2.3). The iron blocks ensure the absorption of hadrons, electrons and photons before the last muon station.
The muon system structure is shown in Fig. 2.12. Every station is composed of four regions (R1–4) of increasing cell size (such that the occupancy is about the same in the whole detector) made up of Multiwire Proportional Chambers, except for the inner region of M1 in which the particle flux is greater and where the triple-GEM technology is adopted. In order to obtain an overall hit reconstruction efficiency of $\sim$ 95%, the efficiency of most regions must exceed 99%. The hit

efficiency is measured for each region of each station using muon candidates triggered independently of the muon system during data-taking runs in 2010 and 2011–2012, and the results are shown in Fig. 2.13.



(a) (b)

Figure 2.12: Schematic side view of the muon system (a) and structure of a single station (b).

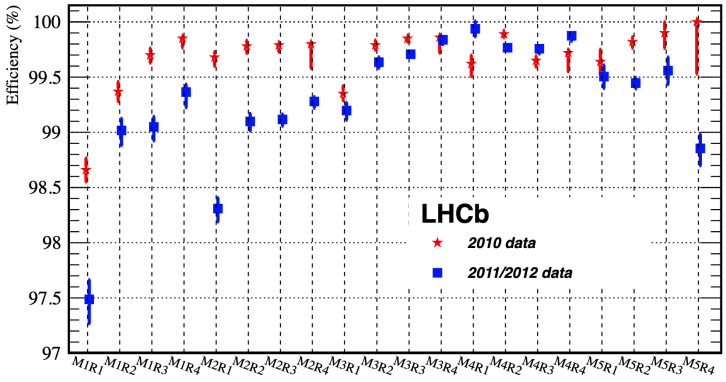


Figure 2.13: Average measured hit efficiency for each region of the muon stations [79].

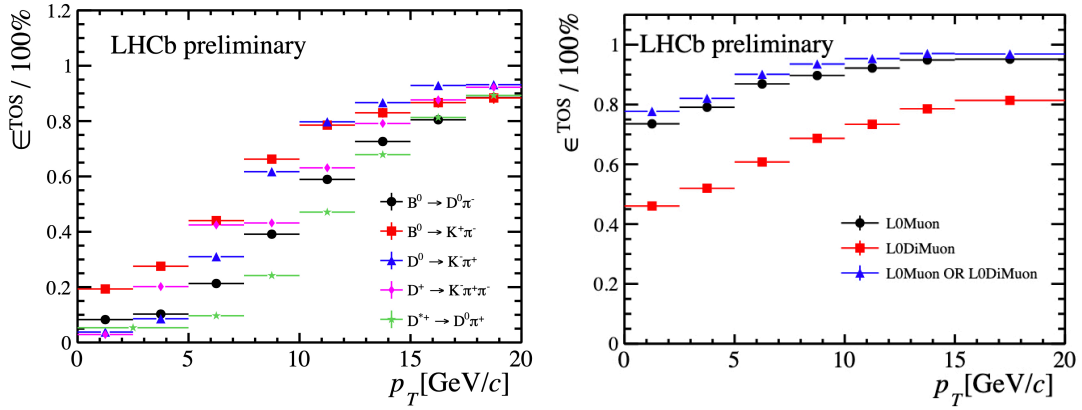
Trigger

The trigger system reduces the amount of data to be processed by the experiment by selecting physically interesting events. As already mentioned in Sec. 2.2, the nominal crossing rate at the LHC is 40 MHz, reduced to $\sim$ 30 MHz because of additional spacing required to allow dump kickers, and has to be further reduced to a few kHz, so that data can be written to storage. The trigger system is composed

of two stages: the Level-0 Trigger (L0) and the High Level Trigger (HLT), further divided into HLT1 and HLT2. It operates in real time, defined as the time between the collisions and the moment in which the data are permanently stored.
The L0 is a hardware stage that reduces the rate from $\sim 30\,\text{MHz}$ to $\sim 1\,\text{MHz}$, which is the maximum allowed rate for the detector read-out system. It selects events with:

- calorimeter clusters above a certain E_T threshold;
- tracks in the muon system above a certain p_T threshold.

Moreover, a pile-up system in the VELO estimates the number of pp collisions, while the calorimeter measures the total energy deposit and the number of charged tracks in the SPD. The information from the three detectors is collected by the L0 Decision Unit (DU) which takes the final decision. Measurements of trigger efficiency for hadrons and muons are shown in Fig. 2.14.
The HLT is a software stage executed on an Event Filter Farm. Two different

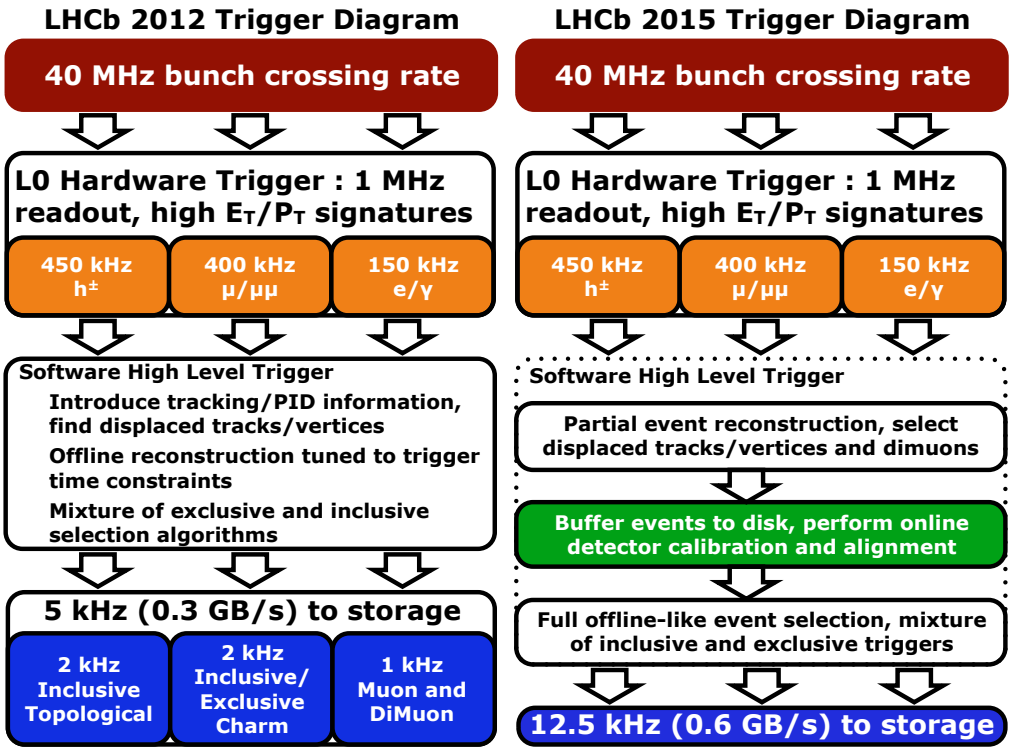


(a) (b)

Figure 2.14: Trigger efficiency measured on background-subtracted data collected during Run 1 for hadrons (a) and muons (b) [80]. The notation ϵ^{TOS} refers to the TOS (Triggered On Signal) efficiency, the probability that the signal alone (*i.e.* without the rest of the event) fires the trigger.

workflows are used in Run 1 and Run 2.
In Run 1 the HLT performed a partial reconstruction of the event, while the full reconstruction was performed offline given the large computing time required (about 2 seconds per event, while the HLT needs to operate around 50 times faster). The HLT1 algorithm reconstructed the VELO tracks and matched some of them to hits in the tracking system. In the HLT2 the time requirement was large enough to match every track with $p > 3\,\text{GeV}$ and $p_\text{T} > 300\,\text{MeV}$ to hits in the tracking system.

In Run 2, the entire HLT1 output is buffered on the Event Filter Farm, in order to perform online calibration and alignment before the data are further processed. The HLT2 performs then a full reconstruction during periods with no beams. The HLT1 is now able to match every VELO track to hits in the TT, and performs a first estimation of the charge and momentum, the latter with an uncertainty of $\sim 20\%$. It selects events on the basis of the reconstructed transverse momentum and impact parameter with respect to the primary vertex, and reduces the rate of events from $\sim 1\,\text{MHz}$ to $\sim 110(80)\,\text{kHz}$ in Run 2 (Run 1). Several changes are introduced in the HLT2 reconstruction algorithms, that allow to obtain a full reconstruction without loss of quality with respect to Run 1. It selects events with multi-tracks detached vertices compatible with coming from B or D hadrons, and reduces the rate of events from $\sim 80\,\text{kHz}$ to $\sim 12.5(5)\,\text{kHz}$ in Run 2 (Run 1).
Fig. 2.15 shows the trigger workflows for Run 1 and Run 2.



(a) (b)

Figure 2.15: Trigger workflow in Run 1 (a) and Run 2 (b).

Offline analysis

In Run 1 the raw event was stored and the reconstruction was performed offline, while in Run 2 the fully reconstructed events from HLT2 are organized into three

data streams:

Full: used for analyses that need to run dedicated event reconstructions. The raw data information is retained;

Turbo: used for analyses that do not need further offline reconstruction. Among all particles belonging to the whole event only specific candidates are saved and the raw data is discarded;

TurboCalib: used to collect calibration data in order to perform online monitoring, measurement of particle identification and tracking efficiencies. The raw data information is retained and both the online and offline reconstructions are performed.

The main advantages of implementing the online reconstruction and saving the events via the Turbo stream is the reduced computing time needed to run the reconstruction algorithms and the smaller size of the final dataset.
The offline reconstruction is performed by the BRUNEL program, which creates physics objects and performs the calculation of the associated quantities (track quality, likelihoods, ...).

The stripping process

Reconstructed events from the Turbo stream are selected and organized into *stripping lines*. These are algorithms used to build candidates and apply a first offline selection in order for the analysts to save space and processing time. The stripping process is run centrally during so-called *stripping campaigns*, and in order to perform specific selections the analysts can submit new lines or modify existing ones.

Events and candidates

The terms "event" and "candidate" refer to different concepts throughout the document. An *event* is defined as the whole set of reconstructed tracks from a pp collision. A *candidate* (or, more specifically, a *signal candidate*) is defined as the set of reconstructed tracks which are hypothesized to be part of the same underlying decay chain the analyst is interested in. Therefore, an event can contain one or more candidates.

Simulation

Monte Carlo (MC) simulations are an essential data analysis tool in order to study the properties of signal events and their interactions with the detector. To produce such simulated samples, the LHCb collaboration uses a series of dedicated tools. The PYTHIA package [81] simulates pp interactions such as the ones in

Fig. 2.1 in a specific LHCb configuration. Generated events are then forced to decay into final states of interest for the analysis. In particular, decays of hadronic particles, including meson mixing, are generated with the EvtGen tool [82], and the final state radiation is simulated with Photos [83]. Finally, the Geant4 tool [84] simulates the interaction of the particles with the detector. The result of the interactions undergoes the digitization procedure, which translates energy deposits in the detector into electronic signals. Successively, the L0 trigger decision is emulated and the HLT and reconstruction algorithms are applied.

Experimental prospects for the LHCb detector

The LHC is scheduled to deliver data for four additional runs (Run 3–6). Run 3 will begin in 2022 and will last until the end of 2024, when the accelerator will start a major upgrading phase, the Long Shutdown 3 (LS3, 2025–2027). During this period the High-Luminosity LHC (HL-LHC) will be installed, which will deliver pp collisions at a nominal luminosity ∼10 times higher than the current one, starting from Run 4.

The LHCb detector is currently in its Phase-1 upgrade, which will allow to take data at a luminosity of $\mathcal{O}(10^{33})\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. The major improvement of this phase is the dismantling of the L0 hardware trigger and the implementation of a full software trigger, increasing the efficiency selection of hadronic final states. The VELO will be replaced with a new one in order to avoid degradation of performances due to ageing of the detector, while a new downstream Scintillating Fiber (Sci-Fi) detector will replace the current downstream tracker. Moreover, most of the front-end electronics will be replaced with newer and more efficient one.
A Phase-2 upgrade, if approved, will be installed starting from 2031, which will allow to take data at a luminosity of $\mathcal{O}(10^{34})\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ during Run 5 and 6. To cope with a foreseen average number of pp interactions per bunch-crossing of ∼50, a precise track timing system will be installed in order to disentangle b-hadron decays from combinatorial background. At the end of Run 6 the LHCb detector will have accumulated $300\,\mathrm{fb}^{-1}$.

The present timeline for the LHC and LHCb upgrades is shown in Fig. 2.16.

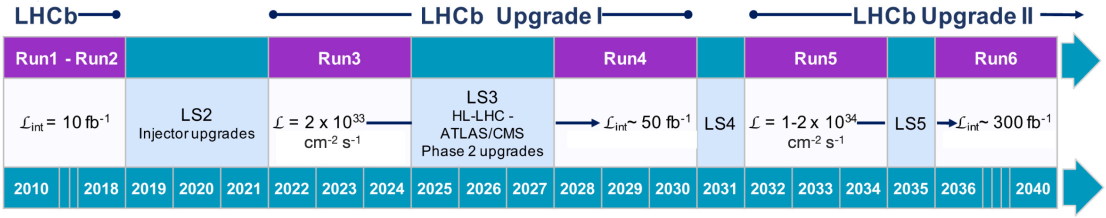


Figure 2.16: Timeline for the LHC.

Part II

Search for the rare $B^0 \to K^{*0}\tau^+\tau^-$ decay

Chapter 3

Introduction and analysis strategy

The second part of this thesis describes the search for the rare $B^0 \to K^{*0}\tau^+\tau^-$ decay performed on $9\,\text{fb}^{-1}$ of proton-proton collision data recorded with the LHCb experiment at centre-of-mass energies of $\sqrt{s} = 7\,\text{TeV}$, $\sqrt{s} = 8\,\text{TeV}$ and $\sqrt{s} = 13\,\text{TeV}$ in 2011, 2012 and the period 2015–2018 respectively, for a total of 6 data-taking years. It represents the first attempt in LHCb to search for the aforementioned decay.

In this analysis, two different final states are considered:

- **$3\pi 3\pi$ final state**: both τ leptons are reconstructed through the 3-prong decay into three charged pions $\tau^- \to \pi^-\pi^+\pi^-\nu_\tau$ or into the decay with an additional neutral pion $\tau^- \to \pi^-\pi^+\pi^-\pi^0\nu_\tau$;
- **$3\pi\mu$ final state**: one τ lepton is reconstructed through its hadronic decay into pions, while the other is reconstructed using its leptonic decay mode $\tau^- \to \mu^-\overline{\nu}_\mu\nu_\tau$.

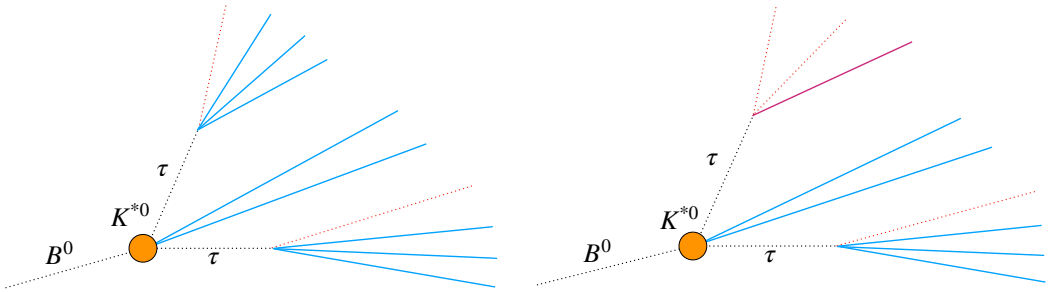
In both final states the K^{*0} is reconstructed via its decay $K^{*0} \to K^-\pi^+$. A schematic view of the two final states is reported in Fig. 3.1.

The observed number of $B^0 \to K^{*0}\tau^+\tau^-$ events $N^{\text{obs}}_{K^{*0}\tau^+\tau^-}$ is related to the branching ratio via the expression

$$\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-) = \frac{N^{\text{obs}}_{K^{*0}\tau^+\tau^-}}{\epsilon_{K^{*0}\tau^+\tau^-} N_{B^0}} \equiv \alpha N^{\text{obs}}_{K^{*0}\tau^+\tau^-}, \tag{3.1}$$

where $\epsilon_{K^{*0}\tau^+\tau^-}$ is the total signal efficiency and N_{B^0} is the total number of produced B^0 hadrons. Since the measurements of luminosity and cross-section are affected by large uncertainties at hadron colliders, the total number of produced B^0 hadrons is determined using a second decay with a relatively high branching fraction, called *normalization channel*. The chosen normalization channel for this analysis is the $B^0 \to D^-(\to \pi^-K^+\pi^-)D_s^+(\to K^+K^-\pi^+)$ decay. The normalization factor α for the $3\pi 3\pi$ ($3\pi\mu$) final state reads as:

$$\alpha = \frac{\epsilon_{D^-D_s^+} \cdot \mathcal{B}(B^0 \to D^-D_s^+) \cdot \mathcal{B}(D^+ \to \pi^+\pi^+K^-) \cdot \mathcal{B}(D_s^+ \to K^+K^-\pi^+)}{N^{\text{obs}}_{D^-D_s^+} \cdot \epsilon_{K^{*0}\tau^+\tau^-} \cdot \mathcal{B}(K^{*0} \to K^-\pi^+) \cdot \mathcal{F}_{3\pi 3\pi(3\pi\mu)}}, \tag{3.2}$$



(a) $3\pi 3\pi$ final state (b) $3\pi\mu$ final state

Figure 3.1: Schematic view of the two final states. The B^0 and the intermediate particles are indicated in the figure, while the blue lines represent the hadrons in the final state, the purple line indicates the muon and the dotted red lines indicate the neutrinos. The neutral pion is omitted.

where $\epsilon_{D^-D_s^+}$ is the normalization efficiency, $N^{\text{obs}}_{D^-D_s^+}$ is the normalization yield extracted from a fit to the data,

$$\begin{aligned} \mathcal{F}_{3\pi 3\pi} &= (\mathcal{B}(\tau^- \to \pi^-\pi^+\pi^-\nu_\tau) + \mathcal{B}(\tau^- \to \pi^-\pi^+\pi^-\pi^0\nu_\tau))^2, \\ \mathcal{F}_{3\pi\mu} &= 2 \cdot (\mathcal{B}(\tau^- \to \pi^-\pi^+\pi^-\nu_\tau) + \mathcal{B}(\tau^- \to \pi^-\pi^+\pi^-\pi^0\nu_\tau)) \cdot \\ &\quad \mathcal{B}(\tau^- \to \mu^-\overline{\nu}_\mu\nu_\tau), \end{aligned} \tag{3.3}$$

and the branching fractions are taken from the PDG [16] and reported in Table 3.1. The value of $\mathcal{B}(K^* \to (K\pi)^0)$ quoted in the PDG includes contribution from the $K^-\pi^+$ and $K^0\pi^0$ final states, and in absence of further experimental information the value $\mathcal{B}(K^{*0} \to K^-\pi^+) = \mathcal{B}(K^* \to (K\pi)^0) \cdot \frac{2}{3}$ is used, where the additional $\frac{2}{3}$ factor is the squared Clebsch-Gordan coefficient relating an isospin state $|\frac{1}{2}, \frac{1}{2}\rangle$ to the direct sum of the states $|1, 1\rangle$ and $|\frac{1}{2}, -\frac{1}{2}\rangle$. The value of the normalization

Table 3.1: Branching ratios of signal and normalization modes entering the calculation of the normalization factor α.

Channel	Branching ratio (%)
$K^* \to (K\pi)^0$	99.754 ± 0.021
$\tau^- \to \pi^-\pi^+\pi^-\nu_\tau$	9.31 ± 0.05
$\tau^- \to \pi^-\pi^+\pi^-\pi^0\nu_\tau$	4.62 ± 0.05
$\tau^- \to \mu^-\overline{\nu}_\mu\nu_\tau$	17.39 ± 0.04
$B^0 \to D^-D_s^+$	0.72 ± 0.08
$D^+ \to \pi^+\pi^+K^-$	9.38 ± 0.16
$D_s^+ \to K^+K^-\pi^+$	5.39 ± 0.15

factor α is reported in Sec. 8.2, after the evaluation of the signal and normalization

efficiencies and the extraction of the normalization yield.

3.1 Workflow

The analysis workflow is common to both final states and is divided into three main steps: the event selection, the likelihood fit, and, in case no signal is observed, the transformation of the fit result into an upper limit on the $B^0 \to K^{*0}\tau^+\tau^-$ branching ratio.

Selection

1. The first step of the selection procedure is the stripping process, which has been briefly introduced in Sec. 2.3. The stripping requirements are listed in Sec. 5.4 and are applied on events which fire *any* trigger line. Specific trigger requirements are then applied, as reported in Sec. 5.5.

2. The further step of the selection consists of a loose cut-based selection described in Sec. 5.6, which rejects the most trivial backgrounds while retaining most of the signal. The list of variables and their cut values are chosen using the $B^0 \to K^{*0}\tau^+\tau^-$ MC samples only.

3. The cut-based stage is followed by a multivariate (MVA) selection involving two *boosted decision trees* (BDT) in sequence, described in Sec 5.7. The BDTs are trained with simulated signal events and data from a background-dominated region in the K^{*0} mass sidebands.

4. The poor mass resolution from missing the neutrinos makes it very hard to differentiate signal from background in a mass fit. A final BDT, called *fitBDT*, is trained and its output is used as an input for the likelihood fit, as described in Chapter 8. The MC training samples used for the fitBDT are required to pass the same selection as for the data, including the cuts on the selection BDTs. The fitBDT is flattened on the signal sample, and its output falls in the range $[0, 1]$.

5. The BDTs combine variables chosen with a procedure aiming at maximizing a weighted area under the ROC curve. The optimization of the selection and fit procedures involves the choice of the BDT variables, the tuning of the BDT hyper-parameters and the value of the cuts on the selection BDTs, as explained in Chapter 9.

Likelihood fit

To determine the number of signal decays in the data, a binned maximum likelihood fit is performed on the output of the fitBDT. The framework for this

fit uses histogram templates to describe the signal and background components. The signal template is taken from simulation, while the background template is obtained with a data-driven method from the K^{*0} mass sidebands. The final fit is performed simultaneously on all the data-taking years of both final state, for a total of 12 samples. For the sake of completeness, also the separate fits to both final states are reported.

Branching ratio limit

The event yield obtained with the likelihood fit is translated into a branching ratio measurement through the use of the normalization mode $B^0 \to D^-(\to \pi^- K^+ \pi^-) D_s^+(\to K^+ K^- \pi^+)$, studied in Chapter 6. This mode has three plus three charged tracks in the final state, proceeds via two intermediate particles, and has a relatively high branching ratio. It has previously been studied by LHCb in [85]. The drawback consists in the absence of an equivalent for the K^{*0} decay and for the muon in the $3\pi\mu$ final state.
In the case where no excess of events is observed with respect to the expected background, a limit on the branching ratio of the signal is set with the use of the CL_s method [17].

3.2 Subsamples

The K^{*0} mass distribution is used to define sub-regions of the data with different values of signal-to-background ratio. An alternative strategy involving a "pseudo-Dalitz" plane[1], built upon the values of the invariant masses of oppositely charged pions from τ lepton decays, has been discarded because of the lower sensitivity obtained and is reported in Appendix A for the sake of completeness.

The K^{*0} meson has an invariant mass of $m_{K^{*0}} = 891.66 \pm 0.26\,\text{MeV}$. The K^{*0} mass distribution is shown in Fig. 3.2 for simulated $B^0 \to K^{*0}\tau^+\tau^-$ events and data. Using four cuts, five different ranges of the K^{*0} invariant mass distribution are defined. They are labelled from "1" to "5" in Fig. 3.2 and are the same for the two final states. The data sample is split in three mutually exclusive sub-samples:

signal region (K^{*0} mass in range 3):
the invariant mass of the K-π system is required to lie around the nominal value of the K^{*0} invariant mass. These signal candidates are the most signal-like, and form the sub-sample on which the final fit is performed. The boundaries of this region are defined in such a way to contain about 75% of the total amount of well-reconstructed signal events:

$$846 \leq m_{K^{*0}} < 938\,\text{MeV}; \tag{3.4}$$

[1] For ease of use it will be named "Dalitz" plane throughout this thesis, even though a Dalitz plot is classically defined in terms of the squared masses.

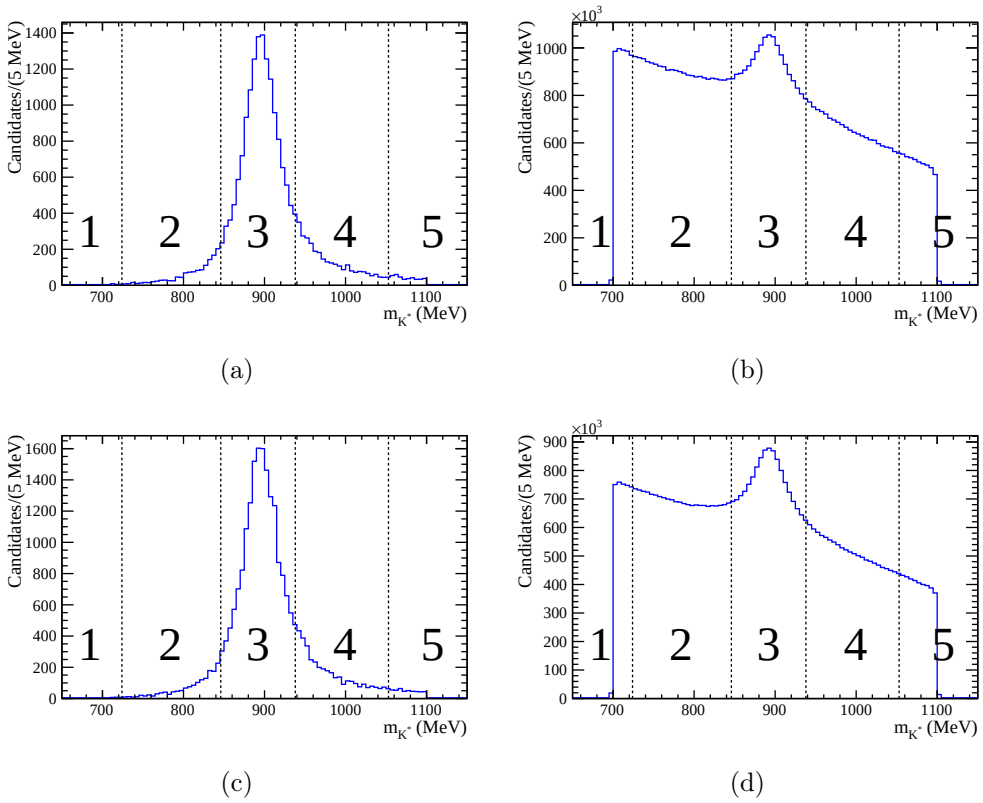


Figure 3.2: K^{*0} invariant mass distribution for 2016 simulated $B^0 \to K^{*0}\tau^+\tau^-$ $3\pi3\pi$ events (a), $3\pi3\pi$ data (b), simulated $B^0 \to K^{*0}\tau^+\tau^-$ $3\pi\mu$ events (c) and $3\pi\mu$ data (d) after the stripping selection. The MC samples used to generate the distributions do not contain the neutral pions component.

background region (K^{*0} mass in range 1 or in range 5):
the invariant mass of the K-π system is required to lie in the lowest or uppermost sidebands. The contamination from signal on data in this sub-sample is negligible and therefore the candidates are used as proxy for background. As shown in Chapter 5, these candidates are used as background samples for the training of the MVA classifiers. The boundaries of this region are chosen to contain about 15% of the total data:

$$700 \leq m_{K^{*0}} < 724\,\text{MeV} \text{ or } 1053 \leq m_{K^{*0}} < 1100\,\text{MeV}; \tag{3.5}$$

control region (K^{*0} mass in range 2 or in range 4):
the invariant mass of the K-π system is required to lie in the middle sidebands. The candidates in this sub-sample are used as proxy for the background in the final fit. However, since a non-negligible amount of signal is present, the contamination from signal is taken into account, as shown in Sec. 8. The

boundaries of this region are:

$$724 \leq m_{K^{*0}} < 846\,\text{MeV} \text{ or } 938 \leq m_{K^{*0}} < 1053\,\text{MeV}. \tag{3.6}$$

The fraction of candidates in each region are reported in Sec. 4.4.

In order not to introduce an experimenter bias, a blinding procedure is applied. In particular, the most sensitive bins of the fitBDT distribution (corresponding to the BDT range $[0.7, 1]$) are masked in the signal region of the data.

Chapter 4

Event samples

In this chapter the event samples used to perform the analysis are introduced.

4.1 Data

A summary of the recorded luminosity for each data-taking year is given in Table 4.1.

Table 4.1: Recorded luminosity corresponding to each data-taking year, where a relative error of 2% is implied. The center-of-mass energy is also reported.

Year	$\sqrt{s}$(TeV)	Recorded luminosity (fb^{-1})
2011	7	1.0
2012	8	2.0
2015	13	0.3
2016	13	1.6
2017	13	1.7
2018	13	2.1

Different datasets are used in the analysis, each of them selected by a specific stripping line:

- Opposite-Sign (OS) $3\pi 3\pi$ data: used to perform the search for signal candidates of the $3\pi 3\pi$ final state;
- OS $3\pi\mu$ data: used to perform the search for signal candidates of the $3\pi\mu$ final state;
- OS normalization data: used to obtain the $B^0 \to D^- D_s^+$ normalization sample;
- Same-Sign (SS) $3\pi 3\pi$ data: SS data are selected requiring both reconstructed τ candidates to have the same charge and are used in the analysis only to perform cross-checks;
- SS $3\pi\mu$ data.

Attempts to use the SS data to model the background have been performed, however the poor statistics after the full selection makes them less suitable for this purpose. Besides this limitation, also the physics background composition differs with respect to OS data: an extra track from the rest of the event is always needed in SS data in order to build a signal candidate.

4.2 Monte Carlo simulation

MC samples are used to study the properties of the signal and normalization modes. An overview is given in Tables 4.2, 4.3 and 4.4. In order to save CPU

Table 4.2: Overview of the different MC samples used in the analysis for the $3\pi3\pi$ final state. The data-taking year, stripping mode and the number of generated events in the LHCb geometrical acceptance are shown.

Channel	Year	Mode	Events
$3\pi^{\pm}\ 3\pi^{\mp}$	2011	filtered	0.9 M
$3\pi^{\pm}\ 3\pi^{\mp}$	2012	filtered	1.8 M
$3\pi^{\pm}\ 3\pi^{\mp}$	2015	filtered	0.5 M
$3\pi^{\pm}\ 3\pi^{\mp}$	2016	filtered	1.8 M
$3\pi^{\pm}\ 3\pi^{\mp}$	2017	filtered	3.8 M
$3\pi^{\pm}\ 3\pi^{\mp}$	2018	filtered	4.8 M
$3\pi^{\pm}(\pi^0)\ 3\pi^{\mp}(\pi^0)$	2011	filtered	2.1 M
$3\pi^{\pm}(\pi^0)\ 3\pi^{\mp}(\pi^0)$	2012	filtered	4.1 M
$3\pi^{\pm}(\pi^0)\ 3\pi^{\mp}(\pi^0)$	2015	filtered	1.1 M
$3\pi^{\pm}(\pi^0)\ 3\pi^{\mp}(\pi^0)$	2016	filtered	4.1 M
$3\pi^{\pm}(\pi^0)\ 3\pi^{\mp}(\pi^0)$	2017	filtered	8.6 M
$3\pi^{\pm}(\pi^0)\ 3\pi^{\mp}(\pi^0)$	2018	filtered	11 M
$3\pi^{\pm}(\pi^0)\ 3\pi^{\mp}(\pi^0)$	2011	flagged	0.2 M
$3\pi^{\pm}(\pi^0)\ 3\pi^{\mp}(\pi^0)$	2012	flagged	0.4 M
$3\pi^{\pm}(\pi^0)\ 3\pi^{\mp}(\pi^0)$	2015	flagged	0.1 M
$3\pi^{\pm}(\pi^0)\ 3\pi^{\mp}(\pi^0)$	2016	flagged	0.4 M
$3\pi^{\pm}(\pi^0)\ 3\pi^{\mp}(\pi^0)$	2017	flagged	0.8 M
$3\pi^{\pm}(\pi^0)\ 3\pi^{\mp}(\pi^0)$	2018	flagged	1.1 M

time, all samples generated for the signal are produced with the requirement $p_{\mathrm{T}} > 250\,\mathrm{MeV}$ on the final state particles, while the samples for the normalization mode are produced with the requirement $p > 1600\,\mathrm{MeV}$.
The $\tau^- \to \pi^-\pi^+\pi^-\nu_\tau$ and $\tau^- \to \pi^-\pi^+\pi^-\pi^0\nu_\tau$ decays are modelled using the `TAUOLA` model tuned on BaBar measurements [86].
Two different configurations are used to generate events for the $3\pi3\pi$ final state: one including the 3-prong charged pions component only, and one including the

Table 4.3: Overview of the different MC samples used in the analysis for the $3\pi\mu$ final state. The data-taking year, stripping mode and the number of generated events in the LHCb geometrical acceptance are shown.

Channel	Year	Mode	Events
$\mu^{\pm}\ 3\pi^{\mp}$	2011	filtered	0.7 M
$\mu^{\pm}\ 3\pi^{\mp}$	2012	filtered	1.4 M
$\mu^{\pm}\ 3\pi^{\mp}$	2015	filtered	0.6 M
$\mu^{\pm}\ 3\pi^{\mp}$	2016	filtered	1.4 M
$\mu^{\pm}\ 3\pi^{\mp}$	2017	filtered	2.9 M
$\mu^{\pm}\ 3\pi^{\mp}$	2018	filtered	3.8 M
$\mu^{\pm}\ 3\pi^{\mp}(\pi^0)$	2011	filtered	1.2 M
$\mu^{\pm}\ 3\pi^{\mp}(\pi^0)$	2012	filtered	2.4 M
$\mu^{\pm}\ 3\pi^{\mp}(\pi^0)$	2015	filtered	0.5 M
$\mu^{\pm}\ 3\pi^{\mp}(\pi^0)$	2016	filtered	1.7 M
$\mu^{\pm}\ 3\pi^{\mp}(\pi^0)$	2017	filtered	3.6 M
$\mu^{\pm}\ 3\pi^{\mp}(\pi^0)$	2018	filtered	5.0 M
$\mu^{\pm}\ 3\pi^{\mp}(\pi^0)$	2011	flagged	0.11 M
$\mu^{\pm}\ 3\pi^{\mp}(\pi^0)$	2012	flagged	0.21 M
$\mu^{\pm}\ 3\pi^{\mp}(\pi^0)$	2015	flagged	0.08 M
$\mu^{\pm}\ 3\pi^{\mp}(\pi^0)$	2016	flagged	0.24 M
$\mu^{\pm}\ 3\pi^{\mp}(\pi^0)$	2017	flagged	0.60 M
$\mu^{\pm}\ 3\pi^{\mp}(\pi^0)$	2018	flagged	0.74 M

components $3\pi 3\pi$, $3\pi\pi^0 3\pi / 3\pi 3\pi\pi^0$ and $3\pi\pi^0 3\pi\pi^0$ in respectively $\sim 44\%$, $\sim 44\%$ and $\sim 12\%$ of the cases.
Similarly, two configurations are used to produce the MC samples for the $3\pi\mu$ final state: one including the 3-prong charged pions component only, and one including the components $3\pi\mu$ and $3\pi\pi^0\mu$ in respectively $\sim 67\%$ and $\sim 33\%$ of the cases.
In order to save storage capacity, the majority of the samples are generated in *filtered* mode, *i.e.* keeping only those candidates which pass the stripping requirements, while the information about the others is lost. However, the stripping lines include some *particle identification* (PID) requirements which are not well modelled by the simulation. In order to compute the signal efficiency for the PID cuts, a data-driven technique is used and unbiased MC samples are needed, on which the PID requirements are not applied. For this reason, smaller additional MC samples are produced in *flagged* mode, which retains the information on all the candidates and allows to apply a modified version of the stripping selection which does not contain the PID requirements. The samples for the normalization channel are instead produced entirely in flagged mode.

Table 4.4: Overview of the different MC samples used in the analysis for the study of the normalization mode. The data-taking year, stripping mode and the number of generated events in the LHCb geometrical acceptance are shown.

Channel	Year	Mode	Events
$B^0 \to D^- D_s^+$	2011	flagged	0.5 M
$B^0 \to D^- D_s^+$	2012	flagged	1.0 M
$B^0 \to D^- D_s^+$	2015	flagged	0.4 M
$B^0 \to D^- D_s^+$	2016	flagged	1.0 M
$B^0 \to D^- D_s^+$	2017	flagged	1.0 M
$B^0 \to D^- D_s^+$	2018	flagged	1.0 M
$B^0 \to D^{*-} D_s^+$	2012	flagged	0.5 M
$B_s^0 \to D_s^- D_s^{*+}$	2012	flagged	0.5 M

4.3 Truth-matching

In order to reject badly reconstructed candidates present in the MC, a so-called *truth-matching* procedure is used. A reconstructed MC track is matched to a "true" generated particle if at least 70% of the detector hits match between the true and reconstructed MC tracks, while for neutral objects the true generated particle must contribute for at least 50% of the reconstructed cluster energy [87].

For the $B^0 \to K^{*0}\tau^+\tau^-$ signal MC, different truth-matching requirements are defined:

fully-matched candidates: all the final state particles are assigned the correct mass hypothesis and come from the same ancestor whose mass hypothesis matches the one of the B^0. Moreover, the decay chain must be fully reconstructed or, in case this condition is not fulfilled, the ancestor must have a mass at most 100 MeV above the nominal mass of the B^0. This last condition assures that decays with radiative photons are included in the truth-matched sample. Such criterion is used to extract the signal samples for the BDT training;

loose-matched candidates: in addition to the candidates selected by the fully-matched criterion defined above, also candidates in which at least one track is not matched to any of the objects in the true generated MC event are included. Such tracks are most likely the result of a pion decaying in-flight;

swapped-pions candidates: all the final state particles are assigned the correct mass hypothesis and come from the same ancestor whose mass hypothesis matches the one of the B^0, but at least two pions in the final state are not associated to the correct intermediate particle (*i.e.* K^{*0}, τ^+ or τ^-). This requirement is fulfilled by signal candidates in which two or more pions are swapped between the intermediate particles.

The selection efficiencies, discussed in detail in Chapter 7, are computed on events containing at least one candidate passing the logical OR of the loose-matched and swapped-pions criteria. This requirement allows more than one candidate per event to pass the truth-matching. However, in case multiple candidates are present after the full selection, only one is kept, chosen randomly among those belonging to the same event.

Candidates for the normalization mode are selected applying an equivalent of the loose matching, retaining either fully reconstructed candidates or candidates in which at least one particle is not matched to any MC true particle. The rate of swapped-pions candidates is negligible.

4.4 Candidates in K^{*0} mass regions

The K^{*0} mass distributions for fully-matched simulated signal events and data are shown in Fig. 3.2. The fraction of candidates in each K^{*0} mass region after the stripping selection are reported in Tables 4.5 and 4.6. The MC samples used to compute these fractions include only the charged pions component (and not the one with neutral pions), and only fully-matched candidates are considered.

Table 4.5: Fraction of candidates belonging to each K^{*0} mass region for fully-matched MC events and OS data for the $3\pi3\pi$ final state after the stripping selection.

Sample	Signal region (%)	Control region (%)	Background region (%)
OS 2011	27.0331 ± 0.0098	58.041 ± 0.011	14.9251 ± 0.0077
OS 2012	27.0056 ± 0.0064	58.090 ± 0.0071	14.9041 ± 0.0051
OS 2015	27.130 ± 0.010	57.737 ± 0.012	15.1327 ± 0.0086
OS 2016	26.8033 ± 0.0038	58.0779 ± 0.0043	15.1192 ± 0.0031
OS 2017	26.8355 ± 0.0037	58.0829 ± 0.0041	15.0820 ± 0.0030
OS 2018	26.7652 ± 0.0034	58.1637 ± 0.0037	15.0710 ± 0.0027
MC 2011	74.69 ± 0.46	23.16 ± 0.44	2.17 ± 0.15
MC 2012	74.90 ± 0.31	23.01 ± 0.31	2.10 ± 0.10
MC 2015	74.34 ± 0.59	23.43 ± 0.57	2.30 ± 0.20
MC 2016	75.11 ± 0.31	22.70 ± 0.30	2.20 ± 0.10
MC 2017	74.99 ± 0.21	22.93 ± 0.21	2.080 ± 0.070
MC 2018	74.60 ± 0.20	23.28 ± 0.19	2.165 ± 0.066

Table 4.6: Fraction of candidates belonging to each K^{*0} mass region for fully-matched MC events and OS data for the $3\pi\mu$ final state after the stripping selection.

Sample	Signal region (%)	Control region (%)	Background region (%)
OS 2011	28.116 ± 0.011	57.136 ± 0.012	14.7519 ± 0.0084
OS 2012	27.9245 ± 0.0069	57.3040 ± 0.0078	14.7717 ± 0.0054
OS 2015	28.071 ± 0.012	57.102 ± 0.014	14.8263 ± 0.0098
OS 2016	27.7761 ± 0.0044	57.3696 ± 0.0049	14.8540 ± 0.0035
OS 2017	27.8354 ± 0.0045	57.3616 ± 0.0049	14.8029 ± 0.0035
OS 2018	27.7730 ± 0.0040	57.4390 ± 0.0045	14.7884 ± 0.0032
MC 2011	74.51 ± 0.35	23.38 ± 0.34	2.12 ± 0.11
MC 2012	74.64 ± 0.24	23.22 ± 0.23	2.141 ± 0.081
MC 2015	75.38 ± 0.44	22.67 ± 0.43	1.96 ± 0.14
MC 2016	74.61 ± 0.29	23.26 ± 0.28	2.135 ± 0.095
MC 2017	74.45 ± 0.21	23.50 ± 0.20	2.086 ± 0.068
MC 2018	74.77 ± 0.18	23.15 ± 0.18	2.080 ± 0.060

Chapter 5

Event selection

This chapter describes in detail the event selection used to suppress the amount of background present in the dataset.

5.1 Boosted decision trees

The analysis makes extensive use of BDTs, which are a class of multivariate techniques able to classify events in different categories, in this case signal and background. They rely on two training samples, one for the signal and one for the background, which are used to learn to discriminate between the two species, and on a list of discriminating variables. The building block of a BDT is a *decision tree* (DT). In a DT, starting from a root node which contains the training samples, the events are split in two sub-samples using the variable that maximizes the discriminating power between signal and background. The process is repeated iteratively on the sub-samples until a stopping criterion is met (for example the size of the node reaches its minimum, or the DT reaches its maximum depth). The final nodes at the bottom of the DT are labelled as "signal" or "background" nodes depending on the majority of events ending up in the node. These nodes are used to assign a numerical output to events belonging to them, *e.g.* +1 for signal and −1 for background. An illustration of a DT is shown in Fig. 5.1. Several DTs are combined together to form a BDT, where at each stage the events misclassified by the previous tree are assigned a higher weight, such that the current tree "focuses" to learn on them. The weight is defined as

$$\alpha = \frac{1 - f_{\text{err}}}{f_{\text{err}}}, \tag{5.1}$$

where f_{err} is the misclassification rate from the previous DT.
Finally, the outputs of all DTs are combined together:

$$\text{BDT}(x) = \frac{1}{N} \sum_{i}^{N} \ln \alpha_i \cdot \text{DT}_i(x), \tag{5.2}$$

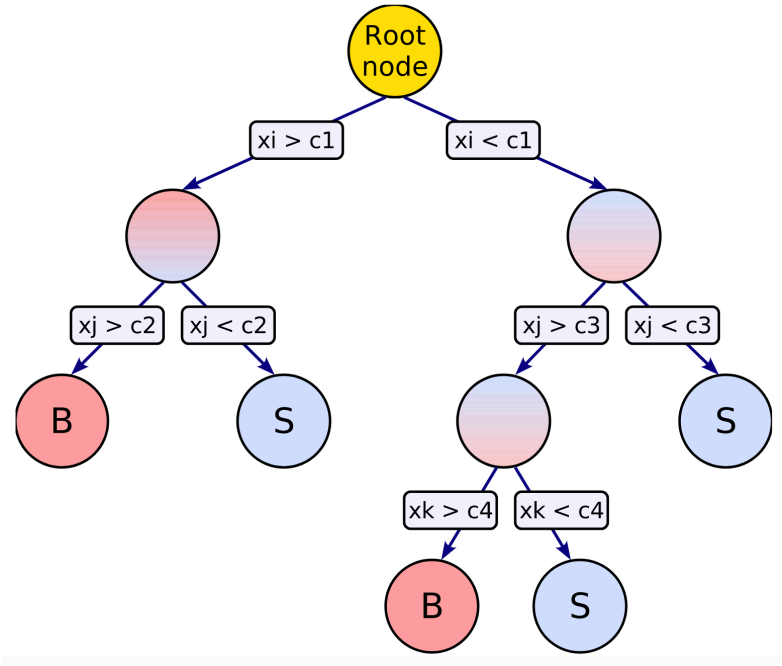


Figure 5.1: Illustration of the logic of a DT [88]

where x represents the input values, N is the total number of DTs in the algorithm and the other terms are self-explanatory. The performance of a BDT can be further increased by forcing a "slow learning" of the algorithm with the use of a *learning rate*, *i.e.* an exponent β to the weight: $\alpha \to \alpha^{\beta}$.

5.2 Discriminating variables

Besides standard kinematic and geometric variables (p, p_{T}, η, vertex χ^2, ...), several isolation and $B^0 \to K^{*0}\tau^+\tau^-$ specific variables are used in the analysis. They are defined and described in the following.

Particles naming convention

Since the analysis does not distinguish between B^0 and $\overline{B}^0$ decays, talking of positive- or negative-charged τ may lead to inconsistencies. Throughout the thesis, a naming convention is assumed to unambiguously refer to the τ lepton candidates: the τ candidate with the same charge of the K from the K^{*0} will be named "positive" τ and labelled as `tauP` or τ^+, while the one with opposite charge will be named "negative" and labelled as `tauM` or τ^-. In this way there is no ambiguity, since the charge of the K from the K^{*0} univocally determines the flavor of the B meson prior to the decay.

Moreover, a numbering convention is assumed for charged pions from τ decays:

the label "1" refers to the pion having the same charge of its ancestor and the smallest p_{T} among the two pions with the same charge originating from the same τ; the label "2" refers to the pion having opposite charge with respect to its ancestor; the label "3" refers to the pion having the same charge of its ancestor and the highest p_{T} among the two pions with the same charge originating from the same τ. The variables associated to pions from τ decays follow this convention, *e.g.* the transverse momentum of the positive-charged pion from the τ^- is indicated as $\pi_2^+ p_{\mathrm{T}}$.

Track isolation variables

The isolation variables measure the activity in terms of other particles in the proximity of the selected track or displaced vertex, *i.e.* how likely it is for an extra track in the vicinity of the reconstructed candidate to be actually part of the same underlying decay. With more activity near the signal candidate, it becomes less likely that the candidate is a true $B^0 \to K^{*0}\tau^+\tau^-$ decay.

Three types of track isolation variables are used in the selection. These variables were initially conceived for the $B_s^0 \to \mu^+\mu^-$ analysis, successively re-optimized for the $B_s^0 \to \tau^+\tau^-$ analysis [89] and adopted for the present search. They are derived on the outcome of a BDT, giving an estimation of the isolation level for each track of the signal candidate. For the training of this BDT an inclusive $b\bar{b}$ Monte Carlo has been used to select $B_s^0 \to \tau^+\tau^-$ signal candidates (defined as candidates passing a loose selection for the $B_s^0 \to \tau^+\tau^-$ search). Let us now consider a specific candidate's track, called *signal track* (ST). For each event all the other tracks, called *non-signal tracks* (NST), are divided into two categories: *non-isolating* and *isolating* tracks. The non-isolating tracks are all tracks coming from displaced B and D decay vertices that are part of the same true decay chain as the ST. The isolating tracks are all remaining tracks, coming either directly from the PV or from another B or D decay in the event, which are essentially unrelated to the ST. For each of the NST of the event, the common vertex $\mathcal{V}$ with the ST is defined as the midpoint between the two tracks along the line of closest approach[1]. The BDT combines the following seven discriminating variables:

- the minimum l/σ_l of the NST with respect to all PV's in the event, where l and σ_l are the distance and its uncertainty of the NST from a given PV;
- the transverse momentum of the NST;
- the angle between the NST and the ST;
- the parameter $f_c = \dfrac{|p^{\mathrm{ST}} + p^{\mathrm{NST}}|\alpha}{|p^{\mathrm{ST}} + p^{\mathrm{NST}}|\alpha + p_{\mathrm{T}}^{\mathrm{ST}} + p_{\mathrm{T}}^{\mathrm{NST}}}$, where p and p_{T} are the momentum and transverse momentum of the tracks, α is the angle between

[1] An actual refit of this vertex for every track in the event is too CPU intense, for likely not much gain.

the momentum of the combined NST + ST system and the vector between the PV and the vertex $\mathcal{V}$;

- the distance of closest approach of the NST and ST;
- the distance between the vertex $\mathcal{V}$ and the B decay vertex;
- the distance between the vertex $\mathcal{V}$ and the PV.

The BDT is trained using the non-isolating tracks as background target, and the isolating tracks as signal target.
The trained BDT is then used to compute the isolation level in $B^0 \to K^{*0}\tau^+\tau^-$ events[2]: for each long track in the event which is *not* part of the signal candidate, the BDT response is calculated. Therefore tracks with high BDT values (isolating tracks) are more likely to come from other B or D decays, whereas tracks with low BDT values (non-isolating tracks) are more likely to come from the same decay chain as the considered candidate track, which makes the candidate more likely to be part of the background. Let a, b and c be the number of tracks with a BDT value smaller than -0.09, -0.05 and 0, respectively. For each signal candidate track the isolation variables are defined as:

IsoBDTFirstValue: the sum $a + 100 \times b + 1000 \times c$. The values of a, b and c are recovered for the selection;

IsoBDTSecondValue: the sum $\sum_{\mathrm{BDT}(x)<-0.05} \mathrm{BDT}(x)$, of BDT values for all tracks x with a BDT output smaller than -0.05;

IsoBDTThirdValue: in Run 1, the sum of the previous variable and the minimal BDT value of all tracks in the event with a BDT output in the range $[-0.05, 0]$. In Run 2 the definition has slightly changed in order to get smoother distributions. It is defined as the sum of the previous variable and the minimal BDT value of all tracks in the event, regardless of what value it assumes.

Examples of distributions are shown in Figs. 5.10 and 5.11 in Sec. 5.6.

Vertex isolation variables

Five types of vertex isolation variables are available. For their construction, the tracks making up the τ, K^{*0} or B decay vertex $\mathcal{V}$ are combined with a single other track from the event, and fitted together into a new vertex $\mathcal{V}^*$. The new vertex isolation variables are then defined as:

VtxIsoNumVtx: the number of other tracks in the event for which the χ^2 of the new vertex fit is less than nine, $\chi^2_{\mathcal{V}^*} < 9$;

[2]The BDT has not been optimized specifically for the $B^0 \to K^{*0}\tau^+\tau^-$ search.

VtxIsoDeltaChi2OneTrack: the smallest difference in χ^2 between the $\mathcal{V}$ and $\mathcal{V}^*$ vertex fits $\Delta\chi^2 \equiv \chi^2_{\mathcal{V}} - \chi^2_{\mathcal{V}^*}$, among all the other tracks in the event;

VtxIsoDeltaChi2TwoTrack: the smallest difference in χ^2 between the $\mathcal{V}$ and $\mathcal{V}^{**}$ vertex fits, where $\mathcal{V}^{**}$ is constructed iteratively by adding one track to the $\mathcal{V}^*$ vertex that itself has the smallest $\Delta\chi^2$;

VtxIsoDeltaChi2MassOneTrack: the invariant mass of the tracks used to build the **VtxIsoDeltaChi2OneTrack** variable;

VtxIsoDeltaChi2MassTwoTrack: the invariant mass of the tracks used to build the **VtxIsoDeltaChi2TwoTrack** variable.

Isolation from neutral objects

Several isolation variables that target neutral objects are used in the analysis. These variables are only available for Run 2 data, since the stripping lines used for Run 1 do not compute these quantities. For their construction, let us define a cone of solid angle $\Omega = 0.5$ around a given candidate track. The neutral cone isolation variables are then defined as:

vPT: the vector-summed transverse momenta of the neutral objects inside the cone;

sPT: the scalar-summed transverse momenta of the neutral objects inside the cone;

$\mathbf{P_{asym}}$: the asymmetry between the momentum of the track and the momentum of the neutral objects inside the cone $\mathrm{P_{asym}} = \frac{p^{\mathrm{track}} - p^{\mathrm{neutral}}}{p^{\mathrm{track}} + p^{\mathrm{neutral}}}$;

$\mathbf{PT_{asym}}$: the asymmetry between the transverse momentum of the track and the transverse momentum of the neutral objects inside the cone $\mathrm{PT_{asym}} = \frac{p_{\mathrm{T}}^{\mathrm{track}} - p_{\mathrm{T}}^{\mathrm{neutral}}}{p_{\mathrm{T}}^{\mathrm{track}} + p_{\mathrm{T}}^{\mathrm{neutral}}}$;

IT: ratio between the transverse momentum of the track and the transverse momentum of the system track + neutral objects inside the cone $\mathrm{IT} = \frac{p_{\mathrm{T}}^{\mathrm{track}}}{\sqrt{(p_{\mathrm{x}}^{\mathrm{track}} + p_{\mathrm{x}}^{\mathrm{neutral}})^2 + (p_{\mathrm{y}}^{\mathrm{track}} + p_{\mathrm{y}}^{\mathrm{neutral}})^2}}$;

DELTAETA: the difference between the pseudorapidity η of the track and the pseudorapidity of the neutral objects inside the cone.

PID variables

Two different types of PID variables are used in LHCb: the `PID` and `ProbNN`. The `PID`, also called `DLL` from *difference log-likelihood*, use the information from the calorimeter, muon and RICH systems to form a set of likelihoods relative to the pion mass hypothesis, that are multiplied:

$$\texttt{PID}_x = \ln \mathcal{L}(x) - \ln \mathcal{L}(\pi), \tag{5.3}$$

where

$$\mathcal{L} = \mathcal{L}_{\text{RICH}} \cdot \mathcal{L}_{\text{calo}} \cdot \mathcal{L}_{\text{muon}}, \tag{5.4}$$

and x can be the electron, muon, kaon or proton mass hypothesis.
The `ProbNN` variables are built using multivariate techniques. The input variables of the classifier are PID information from each sub-system and tracking information. An additional boolean muon identification variable, called `IsMuon` is defined. A track is assigned the label `IsMuon` if a minimum number of muon stations (depending on the track momentum) have hits inside a region of interest around the extrapolated track.

Other discriminating variables

Ghost probability: probability for a given track to be a fake track built up from random hits in the detector, based on the output of a neural network algorithm [90];

Impact Parameter (IP): the minimum distance l of a track from a given vertex. Unless differently specified, the PV is used in this thesis;

IP $\boldsymbol{\chi^2}$: l/σ_l, where $l \pm \sigma_l$ is the IP of a track from a given vertex. Unless differently specified, the PV is used in this thesis;

MINimum Impact Parameter (MINIP) $\boldsymbol{\chi^2}$: minimum IP χ^2 of a given track with respect to a set of vertices. The primary vertices of the event are used in this thesis unless differently specified;

Distance Of Closest Approach (DOCA): distance between two given tracks at the point of closest approach;

DIRection Angle (DIRA): angle between a track's momentum and the vector pointing from a given vertex to the particle's decay vertex. In this thesis it is expressed in terms of its cosine;

corrected mass: $M_{\text{corr}} = \sqrt{M_B^2 + (p_B \cdot \sin(\theta))^2} + p_B \cdot \sin(\theta)$, where M_B is the invariant mass of the charged particles in the final state, and θ is the angle

between the momentum of the B and its flight direction[3];

CDFIso: CDFIso $= \dfrac{p_{\mathrm{T}}^{B}}{p_{\mathrm{T}}^{B} + \sum_i p_{\mathrm{T}}^{i}}$, where p_{T}^{B} indicates the B transverse momentum and the sum is done over the transverse momenta of all the tracks i inside a cone of solid angle $\Omega = 1.0$ around the B track [91];

rhoMass: for a given τ candidate, rhoMass is defined as the minimum absolute difference between the combined mass of oppositely charged pions building up the candidate and the central value of the ρ^0 mass (in MeV), *e.g.*: τ^+ rhoMass $\equiv \min(|m_{\pi_1^+\pi_2^-} - 775.49|, |m_{\pi_2^-\pi_3^+} - 775.49|)$;

resMass: for a given τ candidate, resMass is defined as the maximum absolute difference between the combined mass of oppositely charged pions building up the candidate and the value of the ρ^0 mass (in MeV), *e.g.*: τ^+ resMass $\equiv \max(|m_{\pi_1^+\pi_2^-} - 775.49|, |m_{\pi_2^-\pi_3^+} - 775.49|)$;

vertex 3D distance: 3-dimensional distance between two vertices V_1 and V_2.

5.3 B^0 mass reconstruction

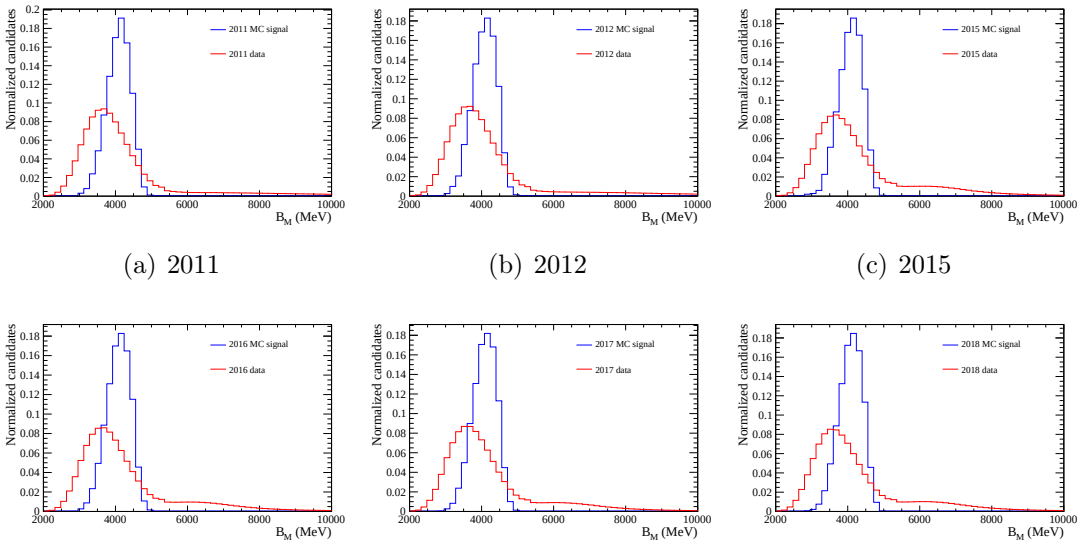
Several different ways of computing the B^0 mass are considered in the analysis and none of them shows sufficient discriminating power between signal and background to perform the final fit. However, despite the fact that the fit is performed on the output of a BDT, some of the different definitions of B^0 mass have been used as discriminating variables in the selection and are therefore reported in the following.

Visible mass The visible B^0 mass is defined as the invariant mass of the charged particles in the final state, without any corrections for the missing energy from neutrinos. The distributions are reported in Figs. 5.2 and 5.3.

Corrected mass The definition of corrected mass has been given in Sec. 5.2. The distributions are reported in Figs. 5.4 and 5.5.

Analytically reconstructed mass An analytic reconstruction of the B^0 mass is performed for both final states applying mass constraints to the system: the τ mass is fixed to its nominal value $M_\tau = 1777\,\mathrm{MeV}$ and the neutrino mass is fixed to 0.
For the $3\pi3\pi$ final state, the decay vertices of the τ lepton are reconstructed using the information on the three pion tracks while the B^0 decay vertex is reconstructed

[3]The expression $M = \sqrt{m^2 + p^2} + p$ represents the invariant mass of a particle decaying into a particle of mass m and a massless particle, with p the momentum of the decay products in the center-of-mass frame.



(a) 2011 (b) 2012 (c) 2015

(d) 2016 (e) 2017 (f) 2018

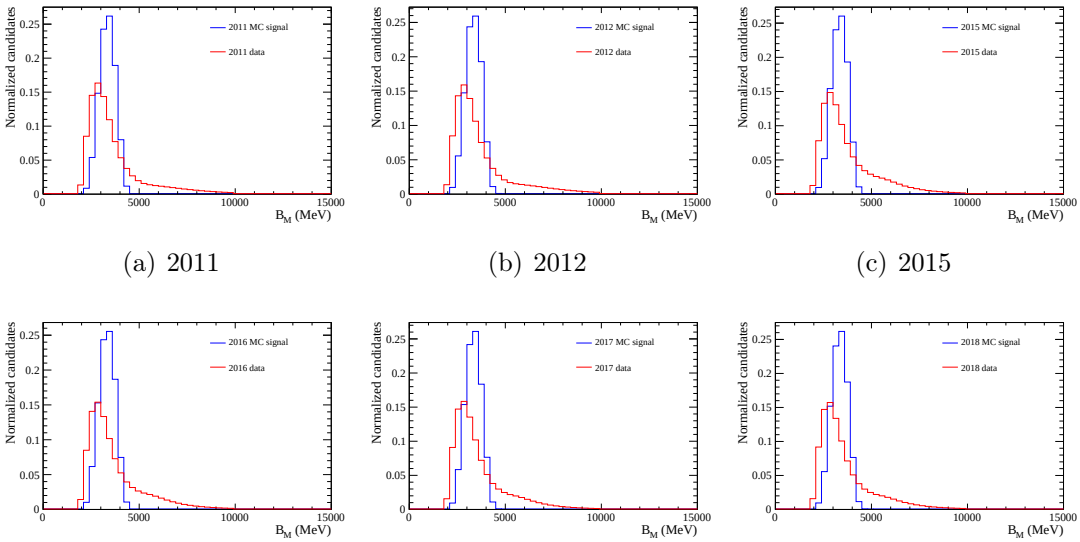
Figure 5.2: Visible B^0 mass for simulated $3\pi3\pi$ signal events (blue) and data (red) after stripping. The whole K^{*0} mass spectrum is considered.

combining the information on the K-π system and on the pions from the two τ's, which allow to derive each τ flight direction $\vec{u}_\tau$. By measuring the three-pions system four-momentum $p_{3\pi} = (E_{3\pi}, \vec{p}_{3\pi})$ the magnitude of the τ momentum is calculated as

$$|\vec{p}_\tau| = \frac{(m_\tau^2 + m_{3\pi}^2)|\vec{p}_{3\pi}|\cos\theta \pm E_{3\pi}\sqrt{(m_\tau^2 - m_{3\pi}^2)^2 - 4m_\tau^2|\vec{p}_{3\pi}|^2\sin^2\theta}}{2(E_{3\pi}^2 - |\vec{p}_{3\pi}|^2\cos^2\theta)}\,, \tag{5.5}$$

where θ is the angle between the τ and the three-pions system flight direction, $\cos\theta = \vec{u}_\tau \cdot \frac{\vec{p}_{3\pi}}{|\vec{p}_{3\pi}|}$ and $m_{3\pi}$ is the invariant mass of the three-pions system $m_{3\pi}^2 = E_{3\pi}^2 - |\vec{p}_{3\pi}|^2$. Due to the quadratic nature of the mass constraint, Eq. 5.5 gives two mathematically valid solutions. Therefore the B^0 mass is reconstructed with a four-fold ambiguity. However the ambiguity is eliminated using the information on the B^0 flight direction, defined as the unit vector pointing from the PV to the B^0 decay vertex: among the four analytic solutions for the B^0 momentum $\vec{p_B}$, the one with the minimum DIRA with respect to the PV is called "optimal" solution.
For the $3\pi\mu$ final state the vertex $\tau \to \mu\nu\nu$ is not reconstructed, however its position is recovered using topological arguments. The τ decaying hadronically is fully reconstructed, with a two-fold ambiguity on its momentum, using Eq. 5.5. Since the transverse momentum with respect to the B^0 flight direction must balance, the transverse momentum of the τ decaying leptonically $p_{\mathrm{T}}^{\tau\to\mu}$ is given by

$$p_{\mathrm{T}}^{\tau\to\mu} = -p_{\mathrm{T}}^{\tau\to3\pi} - p_{\mathrm{T}}^{K^{*0}}\,, \tag{5.6}$$



(a) 2011 (b) 2012 (c) 2015

(d) 2016 (e) 2017 (f) 2018

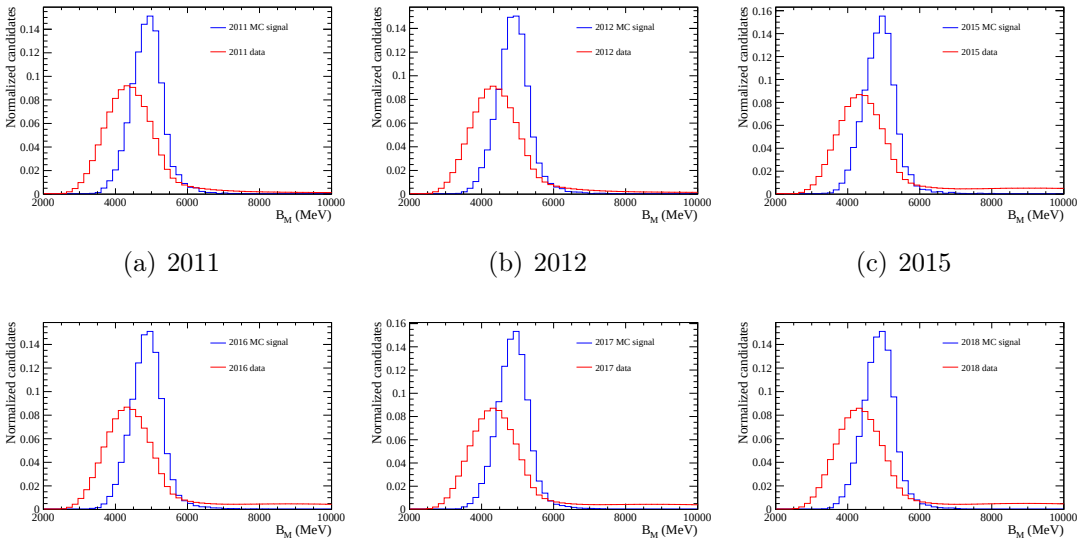
Figure 5.3: Visible B^0 mass for simulated $3\pi\mu$ signal events (blue) and data (red). The whole K^{*0} mass spectrum is considered.

where $p_{\mathrm{T}}^{\tau\to 3\pi}$ and $p_{\mathrm{T}}^{K^{*0}}$ are the transverse momenta with respect to the B^0 flight direction of the τ decaying hadronically and the K^{*0} respectively. Once $p_{\mathrm{T}}^{\tau\to\mu}$ is known, let us define the plane containing $p_{\mathrm{T}}^{\tau\to\mu}$ and the B^0 flight direction. This plane must also contain the missing τ decay vertex, which is identified as the intersection between the μ flight direction and the plane, with a two-fold ambiguity resulting from the τ mass constraint. The distributions are reported in Figs. 5.6 and 5.7. For Eq. 5.5 to be well-defined, the following condition must be satisfied:

$$\Delta = (m_\tau^2 - m_{3\pi}^2)^2 - 4m_\tau^2|\vec{p}_{3\pi}|^2\sin^2\theta \geq 0\,. \tag{5.7}$$

However, due mostly to vertex resolution, Eq. 5.7 is satisfied by only $\sim 32\%$ ($\sim 54\%$) of simulated $3\pi3\pi$ ($3\pi\mu$) signal candidates. Therefore, since variables from the analytic reconstruction are only used to provide more discrimination power and not to perform the final fit, Δ is set to 0 for candidates which do not satisfy Eq. 5.7, in order to retain 100% of signal candidates.

In this analysis, a naming convention is assumed to indicate the variables derived from the analytic reconstruction: the quantities derived from Eq. 5.5 with the positive "+" solution are labelled with the suffix `_a`, while the ones derived with the negative "-" solution are labelled with the suffix `_b` (e.g. the momentum of the positive τ with positive solution is called `tauP_P_a`). The quantities derived by combining information on the two τ's are labelled with the combined suffixes, (e.g. the analytic solution for the B^0 mass obtained considering the positive solution for the positive τ and the negative solution for the second τ is `B_M_ab`), while



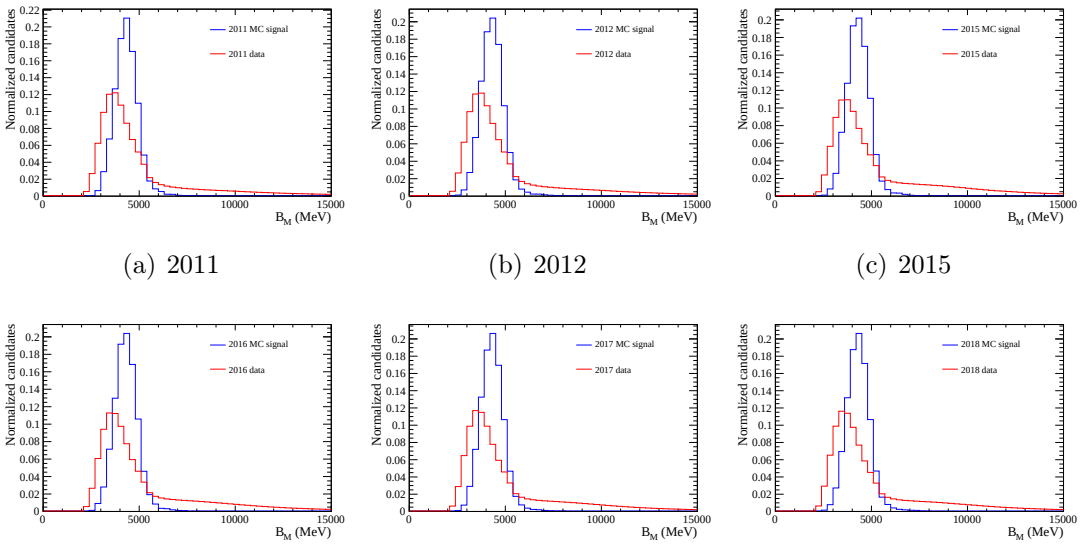
(a) 2011 (b) 2012 (c) 2015

(d) 2016 (e) 2017 (f) 2018

Figure 5.4: Corrected B^0 mass for simulated $3\pi3\pi$ signal events (blue) and data (red). The whole K^{*0} mass spectrum is considered.

the "optimal" solution in the $3\pi3\pi$ final state is labelled with the suffix `_opt`. Specifically, the variables used in the analysis are:

- B^0 analytically reconstructed mass from optimal solution `B_M_opt`;
- B^0 analytically reconstructed mass from the two τ "+" solutions `B_M_aa`;
- τ analytically reconstructed momentum from optimal solution `tauP(M)_P_opt`;
- τ analytically reconstructed momentum from "+" solution `tauP(M)_P_a`;
- τ z component of analytically reconstructed momentum from "+" solution `tauP(M)_PZ_a`;
- τ y component of analytically reconstructed momentum from "+" solution `tauP(M)_PY_a`;
- τ y component of analytically reconstructed momentum from "-" solution `tauP(M)_PY_b`;
- τ transverse component of analytically reconstructed momentum from "-" solution `tauP(M)_PT_b` (or `tauP(M)_P_tr_b`);
- τ transverse component of analytically reconstructed momentum from "+" solution `tauP(M)_PT_a` (or `tauP(M)_P_tr_a`).



(a) 2011 (b) 2012 (c) 2015

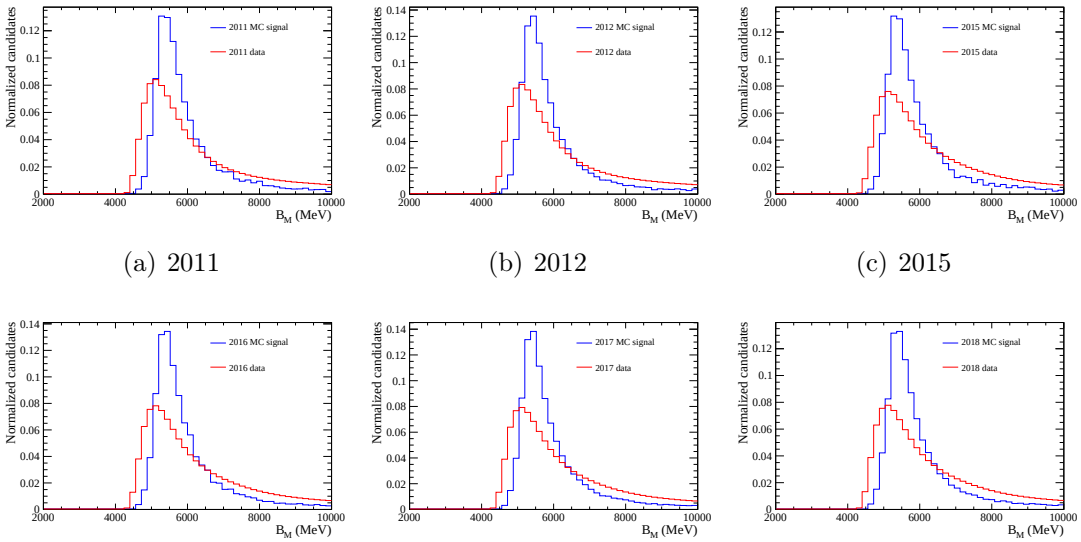
(d) 2016 (e) 2017 (f) 2018

Figure 5.5: Corrected B^0 mass for simulated $3\pi\mu$ signal events (blue) and data (red). The whole K^{*0} mass spectrum is considered.

Refitted mass The "refitted" mass is derived from a refitting of the decay chain performed using a Kalman filter [92]. The tool has been modified to include the neutrinos in the decay chain and to initialize their momenta using the analytic solution. Moreover, the fit is constrained using the values of the τ and ν_τ masses. However, due to the fit failing, some candidates show unphysical negative values for the τ mass. Therefore only events in which the refitting of the decay chain is successful and no negative masses are present are retained. This requirement has an efficiency of $\sim 98\%$ on $3\pi3\pi$ simulated signal events and is included in the cut-based preselection discussed in Sec. 5.6. On the other hand, the refitting of the decay chain is more challenging for the $3\pi\mu$ final state, given the presence of three neutrinos in the final state. The efficiency of the refitting procedure drops to $\sim 50\%$ and hence the use of refitted variables is avoided in order to retain signal efficiency. The distributions are reported in Figs. 5.8 and 5.9. The variables derived from the refitting of the decay chain are labelled with the prefix `B_OPT_`:

- τ refitted decay length `B_OPT_tauP(M)_decayLength`[4];
- τ refitted decay length error `B_OPT_tauP(M)_decayLengthErr`;
- refitting procedure χ^2 `B_OPT_chi2`;

[4]This variable is not well reproduced by the simulation in Run 2. In the final version of the analysis it is replaced by the standard decay length measured from the positions of the decay vertices.



(a) 2011 (b) 2012 (c) 2015

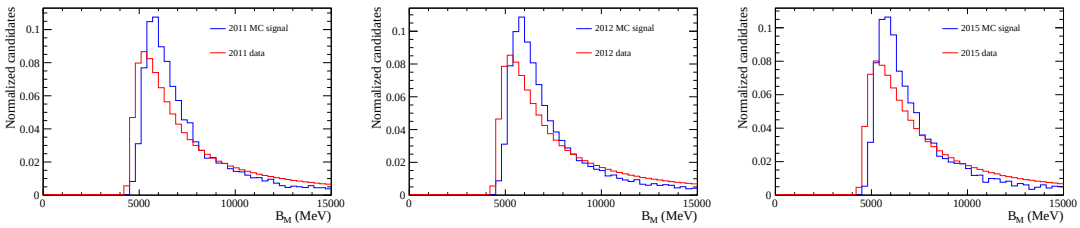
(d) 2016 (e) 2017 (f) 2018

Figure 5.6: Analytically reconstructed B^0 mass for simulated $3\pi3\pi$ signal events (blue) and data (red). The whole K^{*0} mass spectrum is considered.

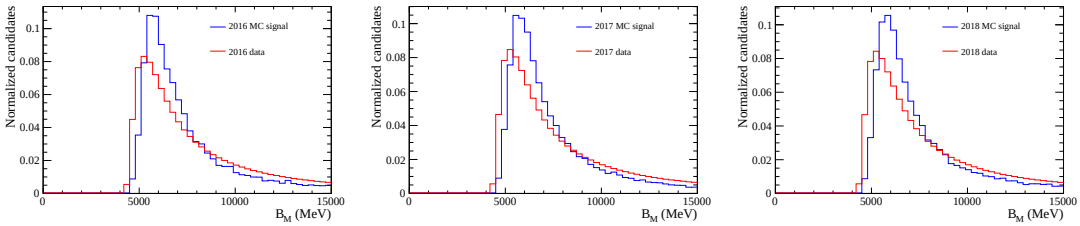
- B^0 refitted decay length `B_OPT_decayLength`;
- number of iterations of the refitting procedure `B_OPT_nIter`.

5.4 Stripping selection

An overview of the stripping selection cuts applied on $3\pi3\pi$, $3\pi\mu$ and normalization signal candidates is given in Tables 5.1, 5.2 and 5.3 respectively.
The number of events and candidates passing the stripping selection is given in Tables 5.4 and 5.5. The high multiplicity is predominantly caused by the many permutations among the pion tracks from which the candidate is reconstructed that also pass the selection requirements.

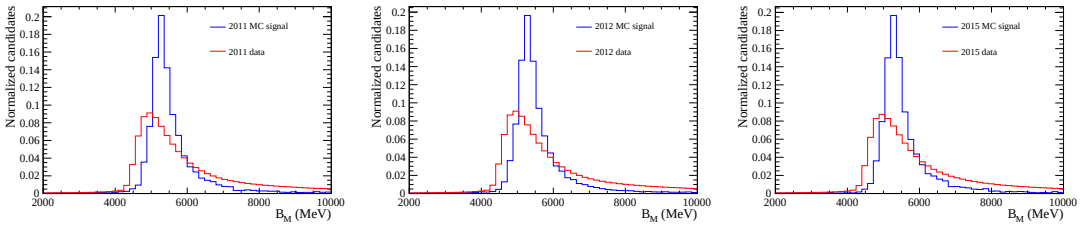


(a) 2011 (b) 2012 (c) 2015

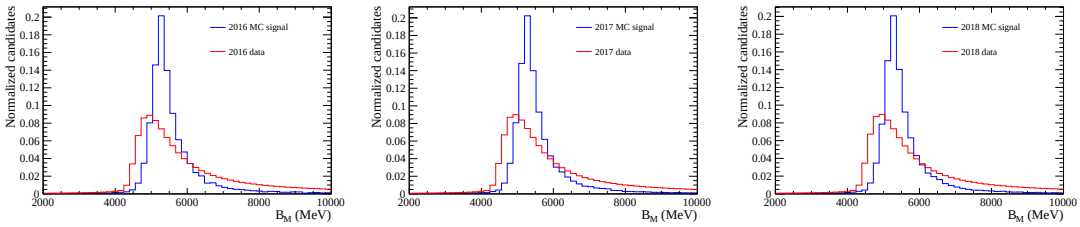


(d) 2016 (e) 2017 (f) 2018

Figure 5.7: Analytically reconstructed B^0 mass for simulated $3\pi\mu$ signal events (blue) and data (red). The whole K^{*0} mass spectrum is considered.

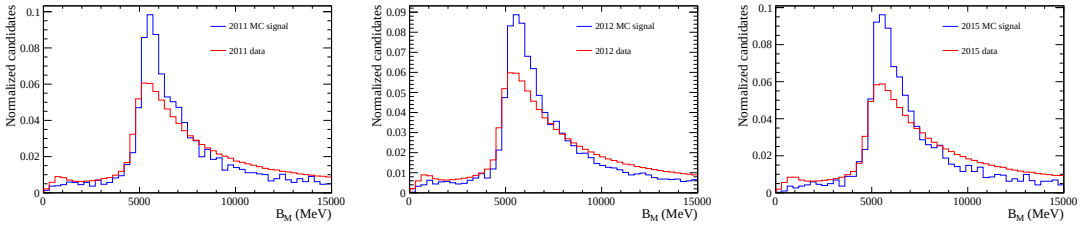


(a) 2011 (b) 2012 (c) 2015



(d) 2016 (e) 2017 (f) 2018

Figure 5.8: Refitted B^0 mass for simulated $3\pi3\pi$ signal events (blue) and data (red). The whole K^{*0} mass spectrum is considered.



(a) 2011 (b) 2012 (c) 2015

(d) 2016 (e) 2017 (f) 2018

Figure 5.9: Refitted B^0 mass for simulated $3\pi\mu$ signal events (blue) and data (red). The whole K^{*0} mass spectrum is considered.

Table 5.1: Overview of the stripping selection cuts for the $3\pi 3\pi$ final state.

Cut	on	value
Visible mass	B^0	$[2-10]\,\text{GeV}$
Invariant mass of charged π's from τ's		$< 5\,\text{GeV}$
Decay vertex χ^2		< 100
Flight distance χ^2		> 80
Flight distance		$< 40\,\text{mm}$
Charged decay products with $p_\text{T} > 0.8\,\text{GeV}$	τ	≥ 1
Decay vertex χ^2		< 16
Maximum DOCA of charged decay products		$< 0.2\,\text{mm}$
Charged decay products' invariant mass		$[500-2000]\,\text{MeV}$
Distance χ^2 between PV and decay vertex		> 16
Tr. distance between PV and decay vertex		$[1-7]\,\text{mm}$
Lon. distance between PV and decay vertex		$> 5.0\,\text{mm}$
DIRA with respect to PV		> 0.99
p_T		$> 1\,\text{GeV}$
Track χ^2/ndf	π from τ	< 4
Ghost probability		< 0.4
MINIP χ^2		> 16
p_T		$> 250\,\text{MeV}$
p		$> 2\,\text{GeV}$
ProbNNπ		> 0.55
Mass	K^{*0}	$[700-1100]\,\text{MeV}$
p_T		$> 1000\,\text{MeV}$
Distance between PV and decay vertex		$> 3\,\text{mm}$
Decay vertex χ^2		< 15
ProbNNπ	π from K^{*0}	> 0.55
ProbNNK	K from K^{*0}	> 0.2
Track χ^2/ndf	Both	< 4
p_T		$> 250\,\text{MeV}$

Table 5.2: Overview of the stripping selection cuts for the $3\pi\mu$ final state.

Cut	on	value
Visible mass	B^0	$[2-10]\,\text{GeV}$
Invariant mass of μ and charged π's from τ's		$< 5\,\text{GeV}$
Vertex χ^2		< 150
Flight distance		$[3-70]\,\text{mm}$
Charged decay products with $p_\text{T} \geq 0.8\,\text{GeV}$	τ	≥ 1
Decay vertex χ^2		< 16
Maximum DOCA of decay products		$< 0.2\,\text{mm}$
Charged decay products' invariant mass		$[500-2000]\,\text{MeV}$
Distance χ^2 between PV and decay vertex		> 16
Tr. distance between PV and decay vertex		$[1-7]\,\text{mm}$
Lon. distance between PV and decay vertex		$> 5.0\,\text{mm}$
DIRA with respect to PV		> 0.99
p_T		$> 1\,\text{GeV}$
Track χ^2/ndf	π from τ	< 4
Ghost probability		< 0.4
MINIP χ^2		> 16
p_T		$> 250\,\text{MeV}$
p		$> 2\,\text{GeV}$
ProbNNπ		> 0.55
p_T	μ	$> 1\,\text{GeV}$
Track χ^2/ndf		< 4
IsMuon		True
PIDμ		> -3
Mass	K^{*0}	$[700-1100]\,\text{MeV}$
p_T		$> 1000\,\text{MeV}$
Distance between PV and decay vertex		> 3
Decay vertex χ^2		< 15
ProbNNπ	π from K^{*0}	> 0.55
ProbNNK	K from K^{*0}	> 0.2
Track χ^2/ndf	Both	< 4
p_T		$> 250\,\text{MeV}$

Table 5.3: Overview of the stripping selection cuts for the normalization mode.

Cut	on	value
Mass	B^0	$[5-7]\,\text{GeV}$
p_T		$> 2\,\text{GeV}$
Corrected mass		$< 10\,\text{GeV}$
Decay vertex χ^2		< 90
DIRA with respect to PV		> 0.99
Flight distance χ^2		> 225
Flight distance		$< 90\,\text{mm}$
Number of D/D_s with $p_\text{T} > 4\,\text{GeV}$		≥ 1
Number of K/π with $p_\text{T} > 2\,\text{GeV}$		≥ 1
max(D MINIP χ^2, D_s MINIP χ^2)		> 150
min(D MINIP χ^2, D_s MINIP χ^2)		> 16
max{min(K/π from D MINIP χ^2), min(K/π from D_s MINIP χ^2)}		> 20
$\sum p_\text{T}$ of K/π		$> 7\,\text{GeV}$
Mass	D	$[1800-2030]\,\text{MeV}$
Decay products with $p_\text{T} \geq 0.8\,\text{GeV}$		≥ 1
Decay vertex χ^2		< 16
Maximum DOCA of decay products		$< 0.2\,\text{mm}$
Distance χ^2 between PV and decay vertex		> 16
Tr. distance between PV and decay vertex		$[0.1-7]\,\text{mm}$
Lon. distance between PV and decay vertex		$> 5.0\,\text{mm}$
DIRA with respect to PV		> 0.99
p_T		$> 1\,\text{GeV}$
ProbNNπ	π from D	> 0.55
PIDK	K from D	> -5
Track χ^2/ndf	Both	< 3
Ghost probability		< 0.3
MINIP χ^2		> 16
p_T		$> 250\,\text{MeV}$
p		$> 2\,\text{GeV}$

Table 5.4: Statistics for the data samples after the $3\pi3\pi$ stripping selection.

Year	Polarity	Candidates	Events	Multiplicity
2011	MagDown	12495299	1384143	9.027 ± 0.008
2011	MagUp	8702273	963299	9.034 ± 0.010
2012	MagDown	24483963	2808744	8.717 ± 0.005
2012	MagUp	24012984	2758122	8.706 ± 0.006
2015	MagDown	10915691	1092239	9.994 ± 0.010
2015	MagUp	6950073	709633	9.794 ± 0.012
2016	MagDown	69681180	6971258	9.995 ± 0.004
2016	MagUp	63971752	6435226	9.941 ± 0.004
2017	MagDown	73889825	7150940	10.333 ± 0.004
2017	MagUp	68663120	6695577	10.255 ± 0.004
2018	MagDown	85420403	8295917	10.297 ± 0.004
2018	MagUp	89139822	8721523	10.221 ± 0.004

Table 5.5: Statistics for the data samples after the $3\pi\mu$ stripping selection.

Year	Polarity	Candidates	Events	Multiplicity
2011	MagDown	10141490	3954523	2.565 ± 0.002
2011	MagUp	7162780	2795352	2.562 ± 0.002
2012	MagDown	21594076	8589655	2.514 ± 0.001
2012	MagUp	21411210	8522930	2.512 ± 0.001
2015	MagDown	7539632	2734522	2.757 ± 0.002
2015	MagUp	5344105	1962941	2.722 ± 0.002
2016	MagDown	51527771	18212970	2.829 ± 0.001
2016	MagUp	50301436	17784150	2.828 ± 0.001
2017	MagDown	52364489	18245983	2.870 ± 0.001
2017	MagUp	49794260	17432280	2.856 ± 0.001
2018	MagDown	59563431	20777543	2.867 ± 0.001
2018	MagUp	63274231	22164996	2.855 ± 0.001

5.5 Trigger

The successfully reconstructed $3\pi3\pi$ and $B^0 \to D^- D_s^+$ candidates are required to pass the trigger cuts in Table 5.6, while the trigger requirements for the $3\pi\mu$ candidates are shown in Table 5.7. The TOS (Triggered On Signal) conditions require that the properties of the tracks composing the candidate are enough to fire the trigger, whereas the TIS (Triggered Independent of Signal) conditions require that the rest of the event alone is enough to fire the trigger (an event could be TIS and TOS at the same time). The L0 lines accept events with large energy deposits in the calorimeter or with hits in the muon system, the HLT1 lines accept events with large p_{T} and IP χ^2, while the HLT2 lines accept events in which a 2-, 3- or 4-body decay is present.

Table 5.6: Trigger requirements applied on $3\pi3\pi$ and normalization channels.

Level	Run 1	Run 2
L0	`L0Hadron` TOS || `L0Global` TIS	`L0Hadron` TOS || `L0Global` TIS
HLT1	`TrackAllL0` TOS	`TrackMVA` TOS
HLT2	`Topo[2||3||4]BodyBBDT` TOS	`Topo[2||3||4]Body` TOS

Table 5.7: Trigger requirements applied on the $3\pi\mu$ channel. All the lines are TOS.

Level	Run 1	Run 2
L0	`L0Hadron` || `L0Muon`	`L0Hadron` || `L0Muon`
HLT1	`TrackAllL0` || `TrackMuon`	`TrackMVA` || `TrackMuon`
HLT2	`Topo[2||3||4]BodyBBDT` || `TopoMU[2||3||4]BodyBBDT`	`Topo[2||3||4]Body` || `TopoMU[2||3||4]Body`

5.6 Cut-based selection

The cut-based phase of the offline selection is aimed at rejecting the most trivial background while keeping a relatively high amount of signal and at cutting away events which lie far in the tails of variable distributions.

$3\pi3\pi$ final state

The cut-based selection for the $3\pi3\pi$ final state is composed of three main sets of cuts:

- isolation-based selection;

- refitting quality requirements;
- multiple candidates requirements.

Isolation-based selection Track isolation variables, described in Sec. 5.2, show a particularly high discriminating power, which allows them to individually reject around 40% of background events while retaining more than 98% of signal events. The cuts are applied on each track in the final state, on the τ tracks and on the K^{*0} vertex[5], and are the same for each data-taking year. They are:

- IsoBDTFirstValue % 100 ≤ 2,
- IsoBDTSecondValue ≥ -0.25,
- IsoBDTThirdValue ≥ -0.2.

The distributions are shown in Fig. 5.10 for the pion from the K^{*0} in 2016 data and MC. The distributions are similar for the other data-taking years and among all the particles (and correlated).

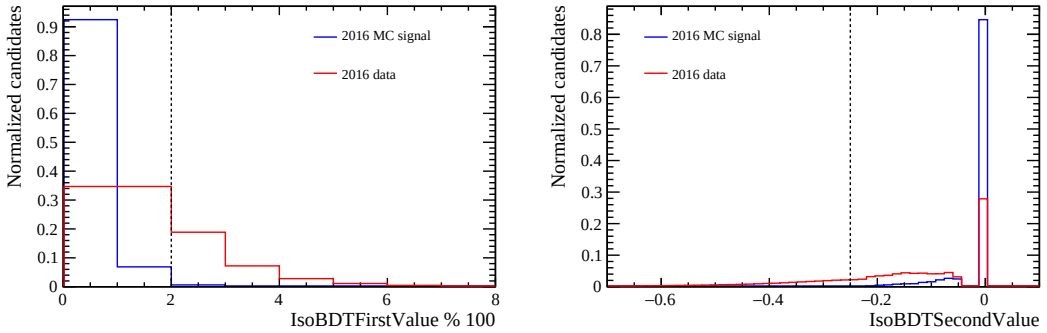
Refitting quality requirements As described in Sec. 5.3, events for which the refitting procedure fails are rejected in the cut-based selection.

Multiple candidates requirements In both MC and data the events are classified into three categories:

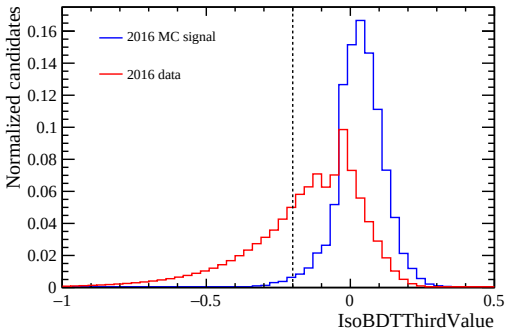
1. Events containing only one candidate, referred to as category 1;
2. Events containing multiple candidates built with the same 8 tracks in the final state (hence swapping two or more tracks between candidates), referred to as category 2;
3. Events containing multiple candidates built with more than 8 tracks in the final state, referred to as category 3.

The fraction of candidates belonging to each category is reported in Tables 5.8 and 5.9 for OS data and signal MC respectively. The amount of simulated signal candidates in category 3 is low compared to the total amount of candidates. Moreover, it has been observed that after the full selection the amount of candidates falling in this category is negligible. For this reason only the much more abundant candidates in events belonging to categories 1 and 2 are retained in order to concentrate the discriminating power of the MVA selection on them.

[5]Even though the K^{*0} candidate has not a track associated to it, it is still possible to compute the values of the input variables for the track isolation BDT.



(a) (b)



(c)

Figure 5.10: Distribution of IsoBDTFirstValue % 100 (a), IsoBDTSecondValue (b) and IsoBDTThirdValue (c) for the pion from the K^{*0} for the $3\pi3\pi$ final state in 2016 data (red) and fully-matched MC (blue).

Table 5.8: Fraction of candidates belonging to each multiplicity category in $3\pi3\pi$ OS data after stripping, trigger and previously reported cut-based requirements.

Year	Fraction cat. 1 (%)	Fraction cat. 2 (%)	Fraction cat. 3 (%)
2011	2.822 ± 0.009	90.555 ± 0.021	6.620 ± 0.017
2012	3.000 ± 0.007	90.615 ± 0.014	6.385 ± 0.014
2015	1.915 ± 0.007	85.870 ± 0.021	12.220 ± 0.021
2016	2.135 ± 0.007	83.860 ± 0.007	14.005 ± 0.007
2017	1.925 ± 0.007	84.805 ± 0.007	13.270 ± 0.007
2018	1.935 ± 0.007	84.865 ± 0.007	13.195 ± 0.007

Table 5.9: Fraction of candidates belonging to each multiplicity category in signal MC after stripping, trigger and previously reported cut-based requirements.

Year	Fraction cat. 1 (%)	Fraction cat. 2 (%)	Fraction cat. 3 (%)
2011	18.40 ± 0.84	80.22 ± 0.87	1.38 ± 0.25
2012	19.46 ± 0.62	78.95 ± 0.63	1.60 ± 0.19
2015	11.06 ± 0.75	86.22 ± 0.83	2.72 ± 0.39
2016	11.91 ± 0.40	84.96 ± 0.45	3.14 ± 0.22
2017	12.57 ± 0.28	84.41 ± 0.30	3.01 ± 0.14
2018	12.55 ± 0.27	83.75 ± 0.31	3.70 ± 0.16

$3\pi\mu$ final state

The cut-based selection for the $3\pi\mu$ final state is composed of three main sets of cuts:

- isolation-based selection;
- additional cut-based requirements;
- multiple candidates requirements.

Isolation-based selection Similarly to what done for the $3\pi3\pi$ final state, the isolation requirements are:

- IsoBDTFirstValue % 100 ≤ 2,
- IsoBDTSecondValue ≥ -0.15,
- IsoBDTThirdValue ≥ -0.2.

The distributions are shown in Fig. 5.11 for the pion from the K^{*0} in 2016 data and MC. The distributions are similar for the other data-taking years and among all the particles (and correlated).

Additional cut-based requirements Additional requirements shown in Table 5.10 are applied in order to further reduce trivial background and to reject harmful events which could degrade the successive MVA performances (*i.e.* events far in the tails of the distributions).

Multiple candidates requirements The multiple candidates requirements are analogous to those of the $3\pi3\pi$ final state. The proportions of each category are reported in Tables 5.11 and 5.12. Candidates belonging to category 3 are rejected.

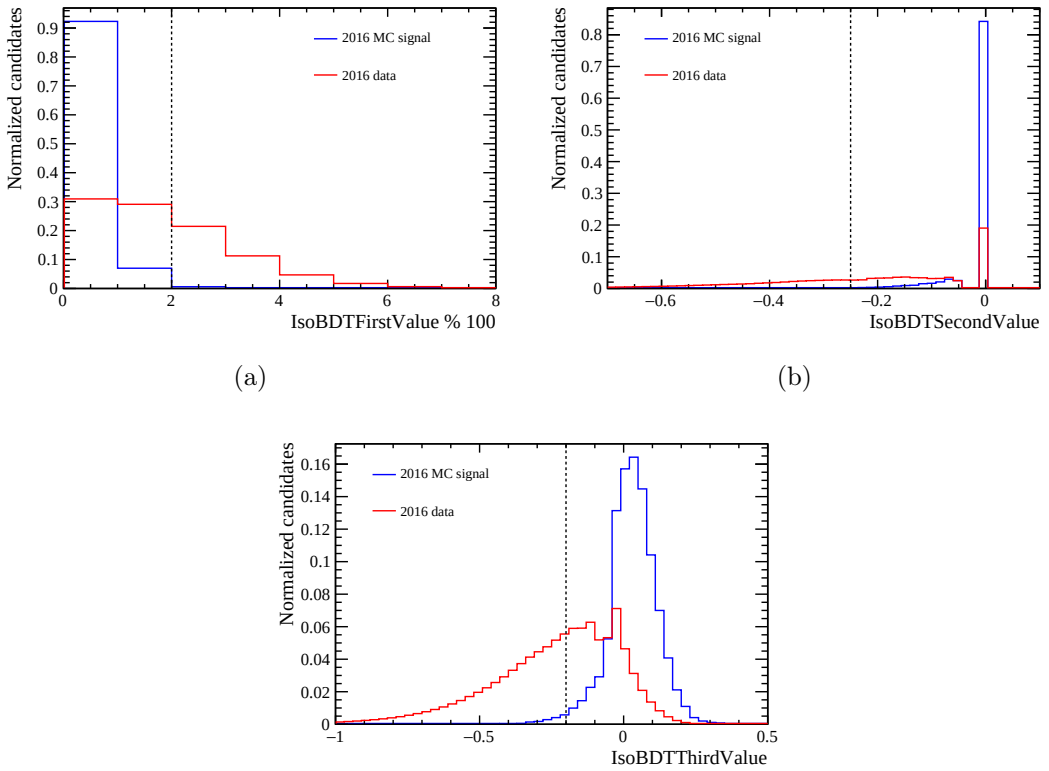


Figure 5.11: Distribution of IsoBDTFirstValue%100 (a), IsoBDTSecondValue (b) and IsoBDTThirdValue (c) for the pion from the K^{*0} for the $3\pi\mu$ final state in 2016 data (red) and fully-matched MC (blue).

Table 5.10: Cut-based requirements applied on $3\pi\mu$ final state candidates, in addition to the isolation-based preselection. The true B^0 and K^{*0} vertices coincide, therefore a non-zero K^{*0} flight distance is due to resolution effects.

Candidate	Requirements
B^0	Visible mass $\in [2500, 4300]$ MeV
τ	Visible mass $\in [900, 1550]$ MeV $M_{\pi_1^\pm\pi_2^\mp} \in [400, 1200]$ MeV $M_{\pi_3^\pm\pi_2^\mp} \in [400, 1200]$ MeV $M_{\pi_1^\pm\pi_3^\pm} < 1200$ MeV Flight distance < 100 mm IP $\chi^2 < 5000$ mm
τ and μ	Invariant mass $\in [1100, 2900]$
K^{*0}	Flight distance < 100 mm
K^{*0} and τ	DOCA < 10 mm

Table 5.11: Fraction of candidates belonging to each multiplicity category in $3\pi\mu$ OS data after stripping, trigger and previously reported cut-based requirements.

Year	Fraction cat. 1 (%)	Fraction cat. 2 (%)	Fraction cat. 3 (%)
2011	67.310 ± 0.070	31.630 ± 0.070	1.060 ± 0.014
2012	67.565 ± 0.049	31.415 ± 0.049	1.020 ± 0.007
2015	67.040 ± 0.110	31.800 ± 0.110	1.159 ± 0.024
2016	67.785 ± 0.028	30.385 ± 0.028	1.830 ± 0.007
2017	67.300 ± 0.028	30.990 ± 0.028	1.710 ± 0.007
2018	67.215 ± 0.028	31.170 ± 0.028	1.615 ± 0.007

Table 5.12: Fraction of candidates belonging to each multiplicity category in signal MC after stripping, trigger and previously reported cut-based requirements.

Year	Fraction cat. 1 (%)	Fraction cat. 2 (%)	Fraction cat. 3 (%)
2011	74.09 ± 0.62	25.61 ± 0.62	0.30 ± 0.08
2012	73.65 ± 0.44	26.11 ± 0.44	0.23 ± 0.05
2015	70.98 ± 1.00	28.87 ± 1.00	0.15 ± 0.08
2016	71.24 ± 0.53	28.22 ± 0.52	0.54 ± 0.09
2017	70.79 ± 0.36	28.72 ± 0.36	0.49 ± 0.06
2018	71.48 ± 0.34	27.99 ± 0.33	0.53 ± 0.05

5.7 Multivariate selection

The second stage of the offline selection consists of a multivariate analysis based on BDTs, aimed at rejecting most of the background. This is done using the TMVA package [88]. The selection proceeds through two BDTs, named BDT1 and BDT2, the second of which is trained after the selection on the first one.

Training samples

The training samples for each MVA stage are obtained applying the same selection sequence as the real data, *i.e.* stripping, trigger and cut-based selection for BDT1 and stripping, trigger, cut-based and BDT1 selection for BDT2 and fitBDT. The signal samples consist of the fully-matched signal MC's which *do not* include the neutral pions component, in order to train the BDT on the cleanest candidates, while the background samples consist of events from the background region of the K^{*0} mass distribution defined in Sec. 3.2.
It has to be stressed that the choice of the training samples assures that no bias due to overtraining is introduced. In fact the signal MC samples (not including the neutral pions component) used for the training are separate and independent from the MC samples (including the neutral pions component) used to assess the BDT performances. Moreover, the background region of the data used to extract the background samples for the training is independent of the control and signal regions, used in the likelihood fit.
Prior to the training, the background region is further split into three mutually exclusive subsamples of $\sim 1\%$, 47% and 52% of the total statistics, used for the training of BDT1, BDT2 and fitBDT respectively. This choice is dictated from the fact that the selection is applied to these samples prior to the training, hence the samples for BDT2 need to be sufficiently large to allow enough events to survive the cut on BDT1. Similarly, the samples for the fitBDT must be sufficiently large such that enough events survive the cuts on BDT1 and BDT2.

The training procedure is done separately for Run 1 (obtained by merging the 2011 and 2012 signal and background samples in order to increase the available statistics), 2016, 2017 and 2018. Because of the poor statistics for the 2015 MC, the BDTs trained with the 2016 samples are used also on 2015 data. This choice is also dictated from the fact that 2015 and 2016 have different data-taking conditions, therefore the variable distributions vary between the two periods and merging the two samples would degrade the performance of the BDT. The samples for the simultaneous fit are kept separate for each data-taking year. The possibility of merging some of them in the fit is investigated in Sec. 8.4, but no improvement is found.
After the training, the BDT output is flattened with respect to the signal sample, as explained in the following.

Flattening of the BDT output

The output of the BDT is a non-physical variable in the range $x \in [-1, 1]$. For conveniency the BDT output is redefined such that its distribution is flat for signal. For each event e, its BDT output x is transformed as follows:

$$x \to \frac{\int_{-1}^{x} \mathrm{BDT}(y)\mathrm{d}y}{\int_{-1}^{+1} \mathrm{BDT}(y)\mathrm{d}y}, \tag{5.8}$$

where $\mathrm{BDT}(y)$ is the BDT distribution of the training sample. As a consequence, the BDT output falls in the range $[0, 1]$ and its distribution for the fully-matched signal MC candidates used for the training is flat after this transformation.

$3\pi3\pi$ final state

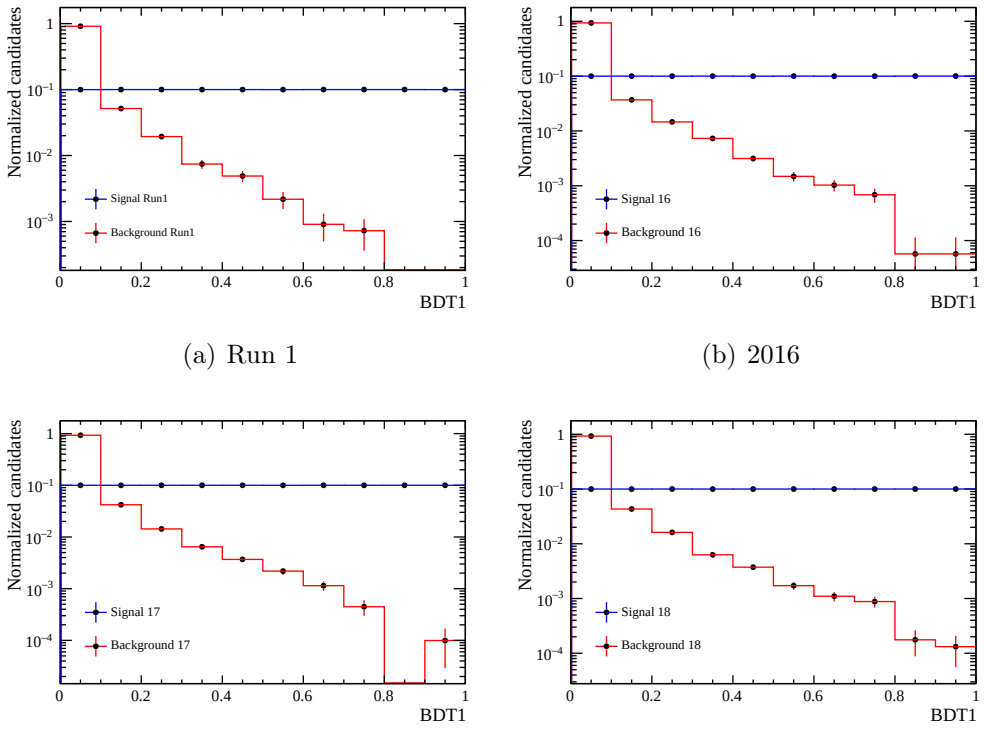
In the $3\pi3\pi$ final state BDT1 combines variables not correlated with the K^{*0} mass, in such a way to be suitable to be used as input variable for the fitBDT; while BDT2 exploits the remaining discriminating power, coming from variables correlated with the K^{*0} mass. In order not to introduce a bias in the background model, this last BDT will not be used as input variable for the fitBDT.
More details about the choice of the input variables are given in Sec. 9.2.

Stage 1 The variables used for the training of the first stage of the MVA selection are:

- max(τ^+ rhoMass, τ^- rhoMass);
- min(τ^+ refitted flight distance, τ^- refitted flight distance);
- max(τ^+ flight distance χ^2, τ^- flight distance χ^2);
- max(τ^+ visible mass, τ^- visible mass);
- B^0 visible mass;
- B^0 neutral cone isolation $\mathrm{P}_{\mathrm{asym}}$[6];
- min(τ^+ $M_{\pi_1^+\pi_2^-}$, τ^- $M_{\pi_1^-\pi_2^+}$);
- B^0 analytically reconstructed mass using optimal solution (`B_M_opt`);
- min(π^- from τ^+ IsoBDTThirdValue, π^+ from τ^- IsoBDTThirdValue);
- max(τ^+ max DOCA of decay products, τ^- max DOCA of decay products).

[6]This variable is not available in Run 1.

The specific rankings are reported in Appendix B, together with the variable distributions. The flattened BDT distributions for signal and background samples are reported in Fig. 5.12. Table 5.13 shows the size of the training samples used.



(a) Run 1 (b) 2016

(c) 2017 (d) 2018

Figure 5.12: Distributions of BDT1 for $3\pi3\pi$ signal and background training samples. The flat transformation is applied.

Table 5.13: Number of candidates used for the training of BDT1 for the $3\pi3\pi$ final state.

Year	Signal candidates	Background candidates
Run1	5385	5478
2016	5475	17400
2017	12267	19995
2018	12251	22615

The chosen working point for the first stage of the MVA selection is:

$$\text{BDT1} > 0.6. \tag{5.9}$$

The working point is chosen in order to minimize the expected upper limit in the case where no signal is observed, following the optimization procedure described in Chapter 9. Table 5.14 shows the efficiency of the BDT1 cut on the signal training sample and its rejection power on OS data. More details on the selection efficiencies are given in Chapter 7.

Table 5.14: Efficiency of BDT1 cut on $3\pi3\pi$ fully-matched MC events and rejection power on OS data.

Year	Efficiency (%)	Rejection power (%)
Run1	40.02 ± 0.66	99.6514 ± 0.0023
2016	40.01 ± 0.66	99.7960 ± 0.0010
2017	40.01 ± 0.44	99.7958 ± 0.0010
2018	40.02 ± 0.44	99.8006 ± 0.0010

Stage 2 Before proceeding to the training of the second stage of the MVA selection, the cut in Eq. 5.9 is applied on the training samples. The variables used for the training are:

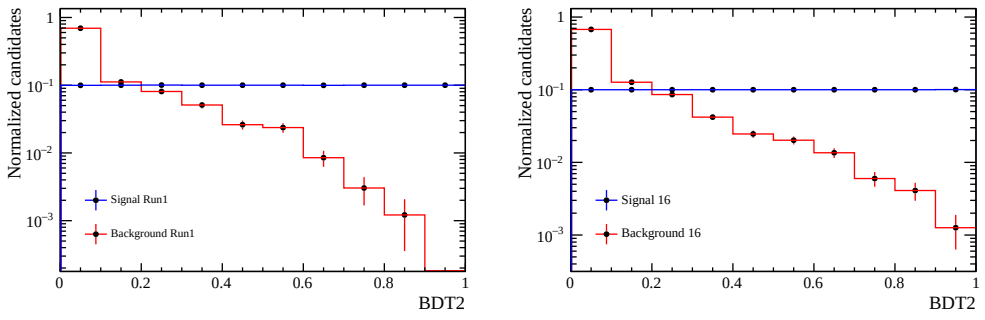
- Refitting procedure χ^2 (`B_OPT_chi2`);
- max(τ^+ VtxIsoDeltaChi2OneTrack, τ^- VtxIsoDeltaChi2OneTrack);
- K^{*0} flight distance[7];
- B^0 decay vertex χ^2;
- min(τ^+ VtxIsoDeltaChi2TwoTrack, τ^- VtxIsoDeltaChi2TwoTrack);
- K^{*0} IsoBDTSecondValue;
- π (from K^{*0}) IsoBDTSecondValue;
- K (from K^{*0}) IsoBDTSecondValue;
- min(τ^+ VtxIsoDeltaChi2MassOneTrack, τ^- VtxIsoDeltaChi2MassOneTrack);
- B^0 DIRA with respect to PV;
- max(τ^+ neutral cone isolation IT, τ^- neutral cone isolation IT)[8];
- max(τ^+ decay vertex χ^2, τ^- decay vertex χ^2);

[7]The true B^0 and K^{*0} vertices coincide, therefore a non-zero K^{*0} flight distance is due to resolution effects.

[8]This variable is not available in Run 1.

- τ^+ and τ^- decay vertices 3D distance;
- K^{*0} IP χ^2 with respect to PV;
- K^{*0} VtxIsoDeltaChi2OneTrack;
- K^{*0} DIRA with respect to PV;

The specific rankings are reported in Appendix B, together with the variable distributions. The BDT distributions for signal and background samples are reported in Fig. 5.13. Table 5.15 shows the size of the training samples used.



(a) Run1 (b) 2016

(c) 2017 (d) 2018

Figure 5.13: Distributions of BDT2 for $3\pi3\pi$ signal and background training samples. The flat transformation is applied.

The chosen working point for the second stage of the MVA selection is:

$$\text{BDT2} > 0.1. \tag{5.10}$$

The working point is chosen in order to cut away the bin with the highest background-signal ratio in Fig. 5.13. Details on the optimization procedure are given in Chapter 9. Table 5.16 shows the efficiency of the BDT2 cut on the

Table 5.15: Number of candidates used for the training of BDT2 for the $3\pi3\pi$ final state.

Year	Signal candidates	Background candidates
Run1	2087	1496
2016	2179	3069
2017	4852	3466
2018	4850	3794

signal training sample and its rejection power on OS data. More details on the selection efficiencies are given in Chapter 7.

Table 5.16: Efficiency of BDT2 cut on $3\pi3\pi$ fully-matched MC events and rejection power on OS data.

Year	Efficiency (%)	Rejection power (%)
Run1	90.01 ± 0.64	64.96 ± 0.32
2016	89.97 ± 0.64	63.80 ± 0.23
2017	90.00 ± 0.43	63.82 ± 0.22
2018	90.00 ± 0.43	63.11 ± 0.21

$3\pi\mu$ final state

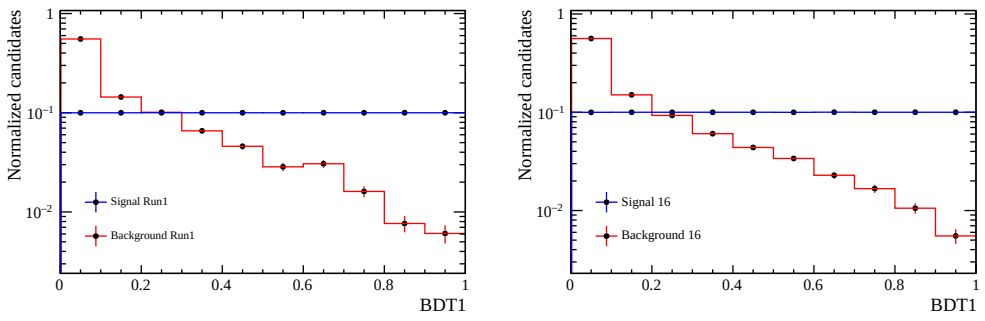
In the $3\pi\mu$ final state the selection and fit strategy is complicated by the presence of a higher degree of correlation of the discriminating variables with the K^{*0} mass. However, a similar selection strategy involving two BDTs in series is adopted, none of them is used as input variable for the fitBDT because of the aforementioned correlations. The BDT1 variables are chosen in order to maximize its discriminating power after the cut-based selection (even though the variables which are found to be the most discriminating are kept for the fitBDT), while BDT2 is used in order to recover the discriminating power of variables which are found to be correlated with the K^{*0} mass after the selection on BDT1, and therefore excluded from the fitBDT. Moreover, BDT1 is used as input variable for BDT2. More details about the choice of the variables can be found in Sec. 9.3.

Stage 1 The variables used for the training of the first stage of the MVA selection are:

- π (from K^{*0}) p_{T};
- μ p_{T};
- B^0 visible mass;

- τ^+ visible mass;
- τ^+ $m_{\pi_2^-\pi_3^+}$;
- π and K (from K^{*0}) DOCA;
- K (from K^{*0}) IsoBDTSecondValue;
- K^{*0} DIRA with respect to B^0 decay vertex[9];
- K^{*0} flight distance[9];
- τ^+ flight distance.

The specific rankings are reported in Appendix B, together with the variable distributions. The flattened BDT distributions for signal and background samples are reported in Fig. 5.14. Table 5.17 shows the size of the training samples used.



(a) Run 1 (b) 2016

(c) 2017 (d) 2018

Figure 5.14: Distributions of BDT1 for $3\pi\mu$ signal and background training samples. The flat transformation is applied.

[9] As explained previously, the B^0 decay vertex and K^{*0} origin vertex are distinct because of resolution effects.

Table 5.17: Number of candidates used for the training of BDT1 for the $3\pi\mu$ final state.

Year	Signal candidates	Background candidates
Run1	14008	3789
2016	6819	6342
2017	14472	6445
2018	16835	7378

The chosen working point for the first stage of the MVA selection is:

$$\text{BDT1} > 0.6. \tag{5.11}$$

The working point is chosen in order to minimize the expected limit in the case where no signal is observed, following the optimization procedure described in Chapter 9. Table 5.18 shows the efficiency of the BDT1 cut on the signal training sample and its rejection power on OS data. More details on the selection efficiencies are given in Chapter 7.

Table 5.18: Efficiency of BDT1 cut on $3\pi\mu$ fully-matched MC events and rejection power on OS data.

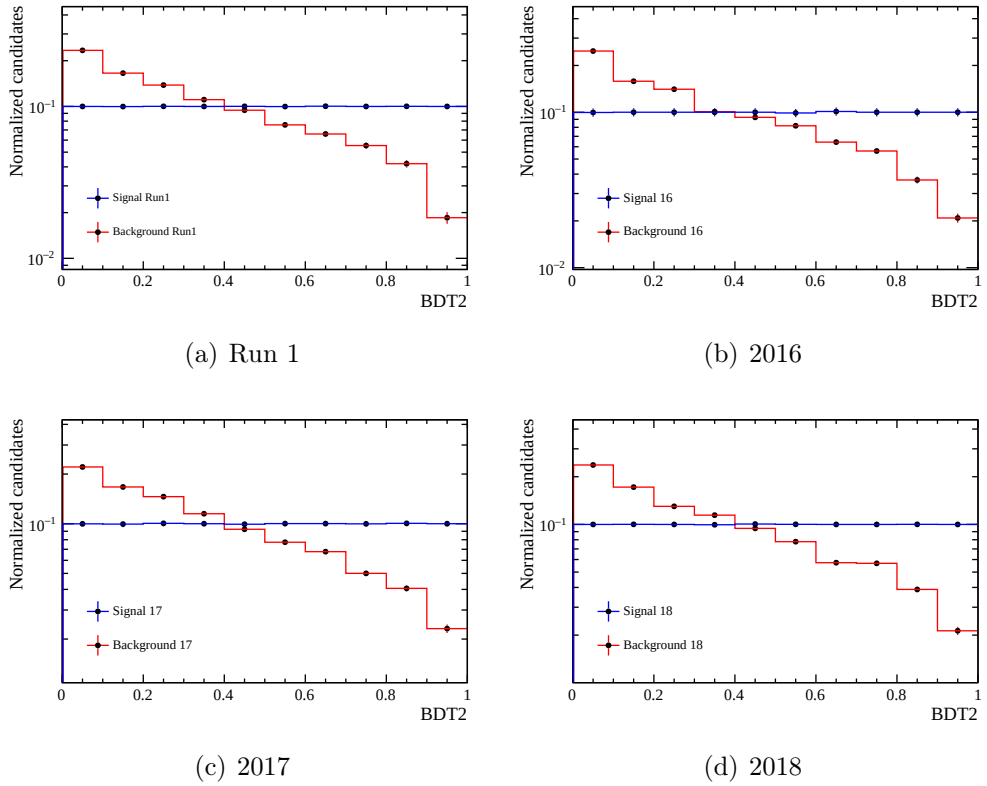
Year	Efficiency (%)	Rejection power (%)
Run1	40.01 ± 0.41	92.447 ± 0.022
2016	40.02 ± 0.59	93.339 ± 0.016
2017	40.00 ± 0.41	92.715 ± 0.017
2018	40.01 ± 0.38	93.019 ± 0.015

Stage 2 Before proceeding to the training of the second stage of the MVA selection, the cut in Eq. 5.11 is applied on the training samples. The variables used for the training are:

- BDT1;
- K^{*0} flight distance[10];
- K^{*0} DIRA with respect to B^0 decay vertex[10];
- τ^+ and μ DOCA;
- τ^+ flight distance χ^2.

[10] As explained previously, the B^0 decay vertex and K^{*0} origin vertex are distinct because of resolution effects.

The specific rankings are reported in Appendix B, together with the variable distributions. The BDT distributions for signal and background samples are reported in Fig. 5.15. Table 5.19 shows the size of the training samples used.



(a) Run 1 (b) 2016

(c) 2017 (d) 2018

Figure 5.15: Distributions of BDT2 for $3\pi\mu$ signal and background training samples. The flat transformation is applied.

Table 5.19: Number of candidates used for the training of BDT2 for the $3\pi\mu$ final state.

Year	Signal candidates	Background candidates
Run1	5604	7077
2016	2729	10582
2017	5789	12273
2018	6735	13383

The chosen working point for the second stage of the MVA selection is:

$$\text{BDT2} > 0.3. \tag{5.12}$$

The working point is chosen in order to minimize the expected limit in the case where no signal is observed, following the optimization procedure described in

Chapter 9. Table 5.20 shows the efficiency of the BDT2 cut on the signal training sample and its rejection power on OS data. More details on the selection efficiencies are given in Chapter 7.

Table 5.20: Efficiency of BDT2 cut on $3\pi\mu$ fully-matched MC events and rejection power on OS data.

Year	Efficiency (%)	Rejection power (%)
Run1	70.01 ± 0.61	51.96 ± 0.15
2016	70.01 ± 0.88	53.34 ± 0.13
2017	70.00 ± 0.60	51.72 ± 0.12
2018	70.00 ± 0.56	52.99 ± 0.12

Chapter 6

Normalization channel

In this chapter, the procedure to extract the normalization yield $N^{\text{obs}}_{D^-D_s^+}$ is presented.

6.1 Event selection

The stripping and trigger lines used to select $B^0 \to D^- D_s^+$ candidates are given in Tables 5.3 and 5.6 respectively. In order to reject other $B \to DD$ decays, additional requirements on the masses of the D mesons are applied: the D^- must have an invariant mass $m_D \in [1855, 1885]\,\text{MeV}$, while the D_s^+ must have an invariant mass $m_{D_s} \in [1955, 1985]\,\text{MeV}$. The rate of multiple candidates per event passing the selection requirements is negligible. The selection efficiencies are discussed in Chapter 7.

6.2 Invariant mass fit

The fit to the invariant mass distribution of $B^0 \to D^- D_s^+$ candidates contains the following components:

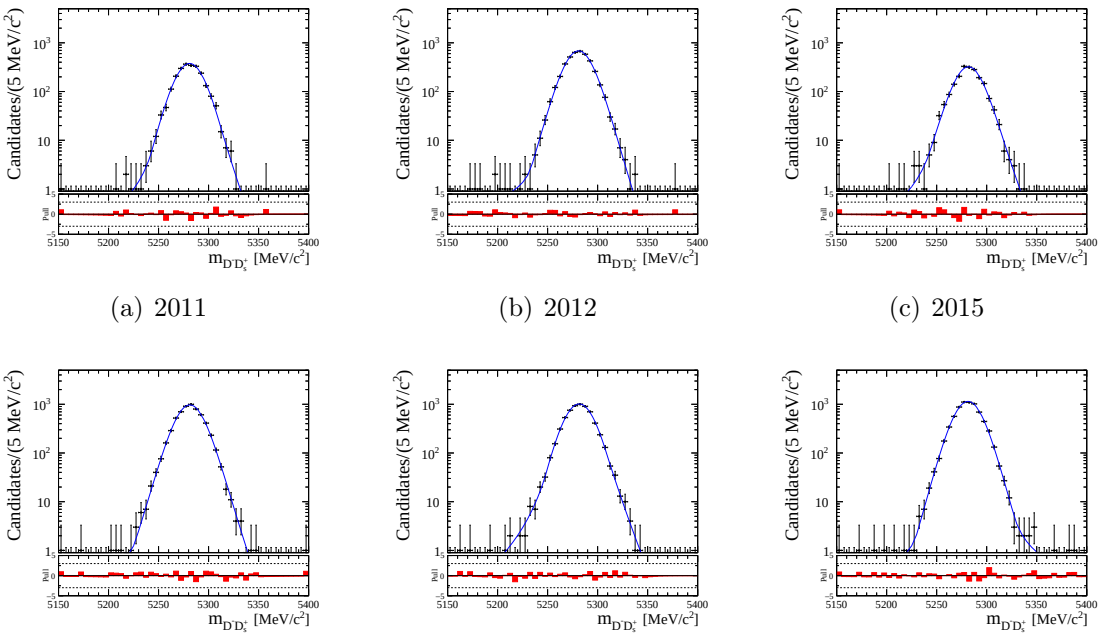
$B^0 \to D^- D_s^+$ signal: the signal distribution is modelled with a double-sided Hypatia function [93]. The fit on simulated $B^0 \to D^- D_s^+$ events is shown in Fig. 6.1 and the results are given in Table 6.1. The parameters β and ζ are fixed to 0. When fitting to data, the parameters describing the tails (a_A, a_B, n_A, n_B and λ) are fixed from simulation, while the parameters describing the width (σ) and position (μ) of the mass peak are initialized with the values obtained from simulation and left floating;

combinatorial background: the combinatorial background is described by an exponential function, whose rate parameter is fitted on data;

$B^0 \to D^{*-} D_s^+$ peaking background: the background from $B^0 \to D^{*-} D_s^+$ has contributions from the decays $D^{*-} \to D^- \pi^0$, which causes the *double-horn* structure due to angular momentum conservation [85], and $D^{*-} \to D^- \gamma$. The mass pdf is obtained from simulation, where truth-matched events are

required to pass the same selection requirements as the $B^0 \to D^- D_s^+$ signal. The shape is modelled using three Gaussian functions, two of them describing the π^0 component and the broader one describing the γ component. The fit on MC is shown in Fig. 6.2 and the parameters are given in Table 6.2. The fit is performed on 2012 MC. To improve the quality of the fit, the parameter μ_R describing the position of the right Gaussian is left floating when fitting to data, while the other parameters are fixed to the values obtained from the fit to simulation;

$B^0 \to D^- D_s^{*+}$ peaking background: the $B^0 \to D^- D_s^{*+}$ mass shape is taken from simulation using the decay $B_s^0 \to D_s^- D_s^{*+}$ and correcting for the B^0-B_s^0 mass difference. The fit is performed on 2012 MC. Truth-matched events are required to pass the same stripping and trigger lines as the $B^0 \to D^- D_s^+$ signal (the PID requirements are the same for the $D_s^{\pm(*)}$ mesons) and $D_s^{\pm(*)}$ mesons are required to decay into $K^+K^-\pi^+$ with invariant mass $m = [1955, 1985]$ MeV. The fit model is composed of three Gaussian functions, and the fit to simulated data is shown in Fig. 6.3, while the results are given in Table 6.3. When fitting to data, the mean of each gaussian is shifted by the known B^0-B_s^0 mass difference, and the mean of the left Gaussian is left floating, while the other parameters are fixed to the values obtained from the fit to simulation.



(a) 2011 (b) 2012 (c) 2015

(d) 2016 (e) 2017 (f) 2018

Figure 6.1: Fit to the invariant mass of simulated $B^0 \to D^- D_s^+$ events with a double-sided Hypatia function for each year of data-taking.

Table 6.1: Results of the fit to the invariant mass of simulated $B^0 \to D^- D_s^+$ events for each year of data-taking. The parameters β and ζ are fixed to 0, the parameters a_i and n_i (with $i = 1, 2$) describe the left and right tails, the parameter μ and σ describe the value and width of the mass peak, while the parameter λ is an additional parameter required by the distribution.

(a) 2011

Name	Fit value
a_A	2.94 ± 0.42
a_B	2.38 ± 0.55
λ	-8 ± 2
μ	5281.00 ± 0.27
n_A	2 ± 1
n_B	25 ± 14
σ	12.76 ± 0.26

(b) 2012

Name	Fit value
a_A	3.59 ± 0.22
a_B	5 ± 2
λ	-6 ± 2
μ	5280.90 ± 0.20
n_A	0.75 ± 0.23
n_B	0 ± 18
σ	13.00 ± 0.19

(c) 2015

Name	Fit value
a_A	2.90 ± 0.54
a_B	2.70 ± 0.91
λ	-5 ± 1
μ	5281.80 ± 0.30
n_A	2 ± 1
n_B	25 ± 22
σ	13.20 ± 0.33

(d) 2016

Name	Fit value
a_A	4.10 ± 0.28
a_B	4.44 ± 0.89
λ	-5 ± 1
μ	5281.30 ± 0.17
n_A	0.75 ± 0.37
n_B	1 ± 1
σ	13.21 ± 0.15

(e) 2017

Name	Fit value
a_A	2.50 ± 0.25
a_B	2.06 ± 0.18
λ	-10 ± 8
μ	5281.60 ± 0.17
n_A	4 ± 1
n_B	25 ± 20
σ	12.89 ± 0.18

(f) 2018

Name	Fit value
a_A	3.96 ± 0.24
a_B	3.55 ± 0.38
λ	-7 ± 1
μ	5281.30 ± 0.15
n_A	0.75 ± 0.24
n_B	1.23 ± 0.83
σ	12.91 ± 0.13

The nominal model used to perform the fit to data is the sum of the aforementioned components. The fit is shown in Fig. 6.4 and the results are reported in Table 6.4. Table 6.5 highlights the normalization yield $N^{\text{obs}}_{D^- D_s^+}$ for each year of data-taking.

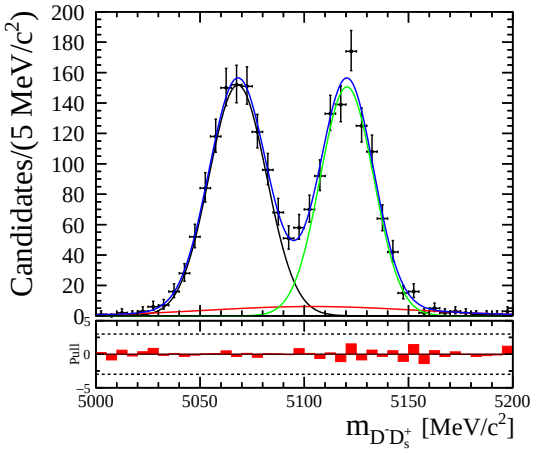


Figure 6.2: Fit to the invariant mass distribution of simulated $B^0 \to D^{*-}D_s^+$ events with 2012 data-taking conditions. The red line corresponds to the contribution from the $D^{*-} \to D^-\gamma$ component, while the black and green lines correspond to the contribution from the $D^{*-} \to D^-\pi^0$ component which causes the double-horn structure. The sum of the three components is shown in blue.

Table 6.2: Results of the fit to the invariant mass of simulated $B^0 \to D^{*-}D_s^+$ events with 2012 data-taking conditions. f_L and f_R indicate the fraction of events in the left and right gaussians respectively. The parameters μ_i and σ_i, with $i = L$, M and R standing for left, medium and right, refer to the mean and width of the i-th gaussian.

Name	Fit value
f_L	0.486 ± 0.016
f_R	0.446 ± 0.015
μ_L	5068.10 ± 0.57
μ_M	5104 ± 8
μ_R	5120.50 ± 0.56
σ_L	13.80 ± 0.53
σ_M	50 ± 9
σ_R	12.79 ± 0.49

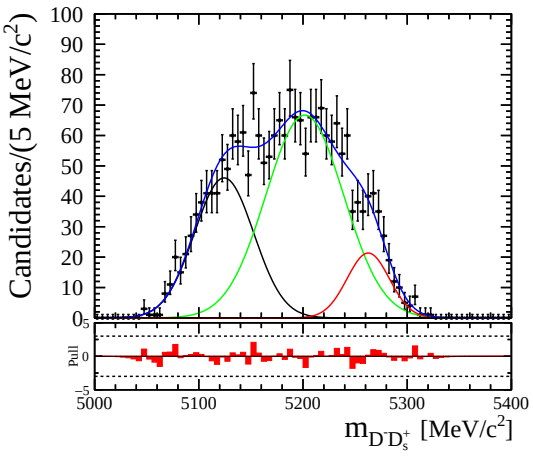
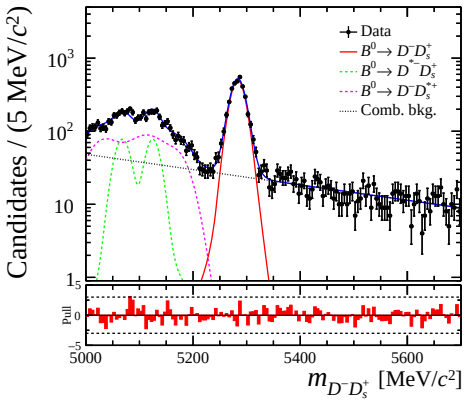


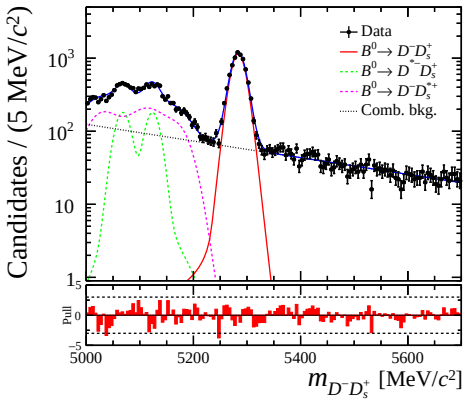
Figure 6.3: Fit to the invariant mass distribution of simulated $B_s^0 \to D_s^- D_s^{*+}$ with 2012 data-taking conditions. The black, green and red components are gaussian functions, while the sum of the three components is shown in blue.

Table 6.3: Results of the fit to the invariant mass of simulated $B_s^0 \to D_s^- D_s^{*+}$ events with 2012 data-taking conditions. f_L and f_M indicate respectively the fraction of events in the left and medium gaussians, the sum of the three being normalized to 1. The parameters μ_i and σ_i, with $i = L$, M and R standing for left, medium and right, refer to the mean and width of the i-th gaussian.

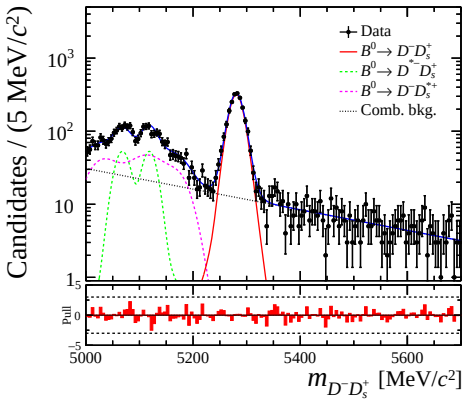
Name	Fit value
f_L	0.31 ± 0.08
f_M	0.60 ± 0.10
μ_L	5124 ± 7
μ_M	5202 ± 12
μ_R	5263 ± 12
σ_L	28 ± 3
σ_M	37 ± 5
σ_R	20 ± 8



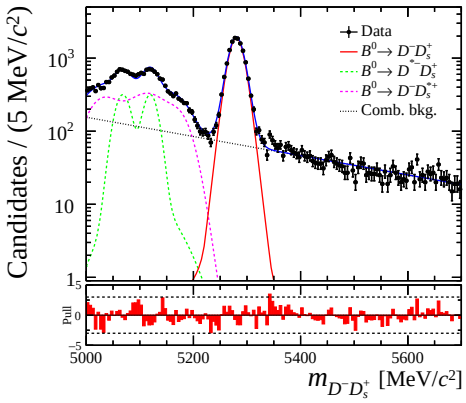
(a) 2011



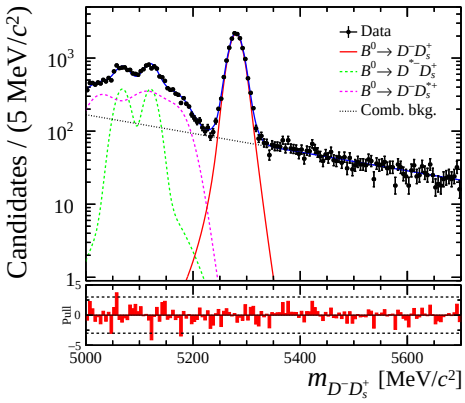
(b) 2012



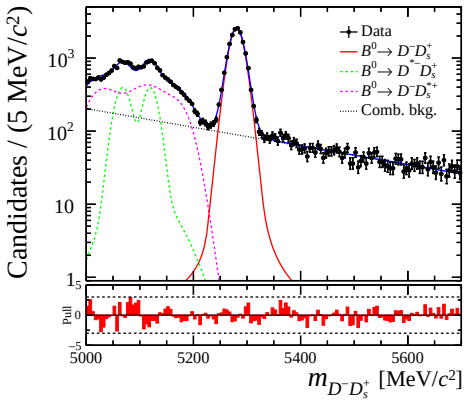
(c) 2015



(d) 2016



(e) 2017



(f) 2018

Figure 6.4: Fit to the $D^- D_s^+$ invariant mass for each year of data-taking.

Table 6.4: Results of the fit to the B^0 invariant mass. $\mu_{R(L)}$ is the mean of the right (left) gaussian of the partially reconstructed background, λ is the rate parameter of the combinatorial background and μ and σ are the position and width of the signal mass peak. The other parameters are the yields of the four components.

(a) 2011

Name	Fit value
$D^{*-}D_s^+$ μ_R	5125 ± 2
$D^-D_s^{*+}$ μ_L	5031 ± 3
λ	-0.00241 ± 0.00021
μ	5283.40 ± 0.27
σ	13.89 ± 0.26
Part. $D^{*-}D_s^+$	1074 ± 113
Part. $D^-D_s^{*+}$	2786 ± 175
Combinatorial	3268 ± 162
Signal	3308 ± 64

(b) 2012

Name	Fit value
$D^{*-}D_s^+$ μ_R	5124 ± 1
$D^-D_s^{*+}$ μ_L	5029 ± 2
λ	-0.00263 ± 0.00013
μ	5283.70 ± 0.17
σ	13.74 ± 0.17
Part. $D^{*-}D_s^+$	2500 ± 170
Part. $D^-D_s^{*+}$	6513 ± 269
Combinatorial	7949 ± 247
Signal	7445 ± 96

(c) 2015

Name	Fit value
$D^{*-}D_s^+$ μ_R	5118 ± 2
$D^-D_s^{*+}$ μ_L	5030 ± 4
λ	-0.00325 ± 0.00031
μ	5280.70 ± 0.33
σ	14.08 ± 0.33
Part. $D^{*-}D_s^+$	731 ± 81
Part. $D^-D_s^{*+}$	1481 ± 137
Combinatorial	1704 ± 133
Signal	2066 ± 50

(d) 2016

Name	Fit value
$D^{*-}D_s^+$ μ_R	5119 ± 1
$D^-D_s^{*+}$ μ_L	5030 ± 2
λ	-0.00303 ± 0.00014
μ	5280.10 ± 0.14
σ	14.32 ± 0.13
Part. $D^{*-}D_s^+$	4371 ± 207
Part. $D^-D_s^{*+}$	10363 ± 326
Combinatorial	9008 ± 304
Signal	12652 ± 123

(e) 2017

Name	Fit value
$D^{*-}D_s^+$ μ_R	5121 ± 1
$D^-D_s^{*+}$ μ_L	5026 ± 1
λ	-0.00295 ± 0.00013
μ	5280.20 ± 0.12
σ	13.40 ± 0.11
Part. $D^{*-}D_s^+$	5206 ± 204
Part. $D^-D_s^{*+}$	11092 ± 337
Combinatorial	9944 ± 306
Signal	13971 ± 128

(f) 2018

Name	Fit value
$D^{*-}D_s^+$ μ_R	5120 ± 1
$D^-D_s^{*+}$ μ_L	5029 ± 1
λ	-0.00294 ± 0.00012
μ	5280.60 ± 0.12
σ	13.69 ± 0.11
Part. $D^{*-}D_s^+$	5442 ± 230
Part. $D^-D_s^{*+}$	13486 ± 372
Combinatorial	12065 ± 342
Signal	16328 ± 139

Table 6.5: Recap of the fit results for the normalization yield for each year of data-taking. The uncertainty is statistical only.

Year	$N^{\text{obs}}_{D^- D_s^+}$
2011	3308 ± 64
2012	7445 ± 96
2015	2066 ± 50
2016	12652 ± 123
2017	13971 ± 128
2018	16328 ± 139

Chapter 7

Selection efficiencies

The selection efficiency for signal ϵ_{Sig} and normalization ϵ_{Norm} can be written as the product of the intermediate efficiencies:

$$\epsilon_{\text{Sig}} = \epsilon_{\text{Gen}} \times \epsilon_{\text{RecoStrip}} \times \epsilon_{\text{PID}} \times \epsilon_{\text{Trigger}} \times \epsilon_{\text{CutBased}} \times \epsilon_{\text{BDT}} \times \epsilon_{\text{Region}}, \tag{7.1}$$

$$\epsilon_{\text{Norm}} = \epsilon_{\text{Gen}} \times \epsilon_{\text{RecoStrip}} \times \epsilon_{\text{PID}} \times \epsilon_{\text{Trigger}} \times \epsilon_{\text{CutBased}}. \tag{7.2}$$

The intermediate efficiencies are primarily determined from simulation. However, since the simulation is not perfect, corrections and/or systematic uncertainties need to be evaluated. The various terms are:

ϵ_{Gen}: it is due to the requirements applied during the generation phase, in particular the geometrical acceptance of the LHCb detector, *i.e.* requiring that all the final state particles fly through the detector. This efficiency is determined during the generation of the simulated MC events, and no systematic uncertainties or corrections are assigned;

$\epsilon_{\text{RecoStrip}}$: it takes into account the track reconstruction efficiency and the stripping efficiency. It is obtained from simulations without further corrections;

ϵ_{PID}: it takes into account the PID efficiency. Corrections to the efficiency extracted from simulations are computed with a data-driven method in Sec. 7.2;

$\epsilon_{\text{Trigger}}$: it is the efficiency of the trigger selection. It can be further decomposed into the efficiencies of L0, HLT1 and HLT2 trigger levels. The efficiencies are obtained from simulations, however corrections to the L0 and HLT1 efficiencies are computed using a data-driven method, and are not discussed in this thesis. No corrections are assigned to the HLT2 efficiency;

$\epsilon_{\text{CutBased}}$: it is the efficiency of the cut-based stage of the offline selection. For the rare mode, it can be further decomposed into an *IsoCut* part, which includes kinematic and isolation cuts, a *DTFCut* part, which includes the requirement on the refitting procedure described in Sec. 5.3 (only for the $3\pi3\pi$ final state), and a *MultCut* part, which includes the requirement on the event multiplicity. The efficiencies are obtained from simulations without further corrections;

$\epsilon_{\mathbf{BDT}}$: it is the efficiency of the BDT selection, and it is further split into *BDT1* and *BDT2* parts. No counterpart for the normalization channel exists since no BDT is used. The efficiencies are obtained from simulations without further corrections;

$\epsilon_{\mathbf{Region}}$: after the full selection, one candidate per event is randomly chosen in the full K^{*0} mass spectrum. This term represents the efficiency for the selected candidate to fall inside the K^{*0} mass signal region. The efficiencies are obtained from simulations without further corrections. No counterpart for the normalization channel exists.

The cut-and-count efficiencies, obtained directly from simulations without applying any correction, are presented in Sec. 7.1. Since the evaluation of the corrections is still in progress, for the time being these are the efficiencies used for the final fit.

7.1 Cut-and-count efficiency

The uncertainties on the selection efficiencies are calculated in the following way:

$$\epsilon = \bar{\epsilon} \pm \sigma_\epsilon = \frac{p}{n} \pm \sqrt{\frac{p/n \cdot (1 - p/n)}{n}}, \tag{7.3}$$

where p is the number of events containing at least one matched candidate passing the selection requirement, while n is the number of events containing at least one matched candidate before the cut. The only exception is represented by the acceptance efficiency ϵ_{Gen}, which is provided by the MC generator statistics. The samples used to compute the efficiency for the rare modes include the neutral pions component. The truth-matching condition for the rare mode is the logical OR of the loose-matched and swapped-pions criteria (defined in Sec. 4.3). On the other hand, since the rate of multiple candidates is negligible for the normalization mode, the efficiency is computed directly on the truth-matched candidates.

As explained earlier, for each year of data-taking, two separate samples are used to perform the evaluation of the cut-and-count efficiencies for the rare mode: the one used up to the PID step is produced in flagged mode in order to disentangle the contribution of the PID from the stripping, while from the L0 step on, a filtered sample is used in order to increase the available statistics. For the normalization mode instead, the same flagged sample is used throughout the whole process.

$3\pi3\pi$ final state

Tables 7.1, 7.2, 7.3, 7.4, 7.5, 7.6 summarize the results of the cut-and-count estimation of the efficiency for the $3\pi3\pi$ final state.

Table 7.1: Number of accepted events on MC and efficiency for each selection step for the 2011 $3\pi3\pi$ final state. The line between PID and L0 efficiency separates the filtered from the flagged samples.

Step	Accepted events	Efficiency
Generator	223968	$0.020\,671 \pm 0.000\,034$
RecoStrip	12071	$0.053\,89 \pm 0.000\,48$
TruthMatching	8144	0.6747 ± 0.0043
PID	2591	0.3181 ± 0.0052
L0	15045	0.4663 ± 0.0028
HLT1	11534	0.7666 ± 0.0034
HLT2	5996	0.5199 ± 0.0047
IsoCut	5343	0.8910 ± 0.0040
DTFCut	5253	0.9832 ± 0.0018
MultCut	5199	0.9897 ± 0.0014
BDT1	1475	0.2837 ± 0.0063
BDT2	1207	0.820 ± 0.010
Region	831	0.688 ± 0.013
Total		$(6.16 \pm 0.24) \cdot 10^{-6}$

Table 7.2: Number of accepted events on MC and efficiency for each selection step for the 2012 $3\pi3\pi$ final state. The line between PID and L0 efficiency separates the filtered from the flagged samples.

Step	Accepted events	Efficiency
Generator	419105	$0.021\,645 \pm 0.000\,038$
RecoStrip	20941	$0.049\,96 \pm 0.000\,34$
TruthMatching	14454	0.6902 ± 0.0032
PID	4172	0.2886 ± 0.0038
L0	23048	0.4275 ± 0.0021
HLT1	18672	0.8101 ± 0.0026
HLT2	9877	0.5290 ± 0.0037
IsoCut	8528	0.8634 ± 0.0035
DTFCut	8356	0.9798 ± 0.0015
MultCut	8241	0.9862 ± 0.0013
BDT1	2330	0.2830 ± 0.0050
BDT2	1869	0.8021 ± 0.0083
Region	1278	0.684 ± 0.011
Total		$(5.11 \pm 0.16) \cdot 10^{-6}$

Table 7.3: Number of accepted events on MC and efficiency for each selection step for the 2015 $3\pi3\pi$ final state. The line between PID and L0 efficiency separates the filtered from the flagged samples.

Step	Accepted events	Efficiency
Generator	112131	$0.025\,590 \pm 0.000\,052$
RecoStrip	5361	$0.047\,81 \pm 0.000\,64$
TruthMatching	3473	0.6479 ± 0.0065
PID	1203	0.3464 ± 0.0081
L0	6772	0.4518 ± 0.0041
HLT1	5962	0.8804 ± 0.0039
HLT2	3964	0.6649 ± 0.0061
IsoCut	3648	0.9203 ± 0.0043
DTFCut	3597	0.9860 ± 0.0019
MultCut	3505	0.9744 ± 0.0026
BDT1	821	0.2342 ± 0.0072
BDT2	657	0.800 ± 0.014
Region	412	0.627 ± 0.019
Total		$(7.55 \pm 0.43) \cdot 10^{-6}$

Table 7.4: Number of accepted events on MC and efficiency for each selection step for the 2016 $3\pi3\pi$ final state. The line between PID and L0 efficiency separates the filtered from the flagged samples.

Step	Accepted events	Efficiency
Generator	414703	$0.025\,674 \pm 0.000\,053$
RecoStrip	21713	$0.052\,36 \pm 0.000\,35$
TruthMatching	14028	0.6461 ± 0.0032
PID	5235	0.3732 ± 0.0041
L0	28184	0.4837 ± 0.0021
HLT1	24514	0.8700 ± 0.0020
HLT2	17295	0.7055 ± 0.0029
IsoCut	15793	0.9132 ± 0.0021
DTFCut	15578	$0.986\,39 \pm 0.000\,92$
MultCut	15118	0.9705 ± 0.0014
BDT1	3708	0.2453 ± 0.0035
BDT2	3043	0.8207 ± 0.0063
Region	1958	0.6434 ± 0.0087
Total		$(1.089 \pm 0.029) \cdot 10^{-5}$

Table 7.5: Number of accepted events on MC and efficiency for each selection step for the 2017 $3\pi3\pi$ final state. The line between PID and L0 efficiency separates the filtered from the flagged samples.

Step	Accepted events	Efficiency
Generator	871470	$0.025\,669 \pm 0.000\,054$
RecoStrip	45729	$0.052\,47 \pm 0.000\,24$
TruthMatching	29805	0.6518 ± 0.0022
PID	10858	0.3643 ± 0.0028
L0	58319	0.5174 ± 0.0015
HLT1	50578	0.8673 ± 0.0014
HLT2	35762	0.7070 ± 0.0020
IsoCut	32724	0.9150 ± 0.0015
DTFCut	32313	$0.987\,44 \pm 0.000\,62$
MultCut	31300	$0.968\,65 \pm 0.000\,97$
BDT1	7857	0.2510 ± 0.0025
BDT2	6585	0.8381 ± 0.0042
Region	4214	0.6399 ± 0.0059
Total		$(1.196 \pm 0.022) \cdot 10^{-5}$

Table 7.6: Number of accepted events on MC and efficiency for each selection step for the 2018 $3\pi3\pi$ final state. The line between PID and L0 efficiency separates the filtered from the flagged samples.

Step	Accepted events	Efficiency
Generator	1111655	$0.025\,587 \pm 0.000\,053$
RecoStrip	58340	$0.052\,48 \pm 0.000\,21$
TruthMatching	37728	0.6470 ± 0.0020
PID	13879	0.3679 ± 0.0025
L0	61465	0.4325 ± 0.0013
HLT1	53620	0.8724 ± 0.0013
HLT2	38041	0.7090 ± 0.0020
IsoCut	34725	0.9128 ± 0.0014
DTFCut	34227	$0.985\,66 \pm 0.000\,64$
MultCut	33245	0.9713 ± 0.0009
BDT1	8325	0.2504 ± 0.0024
BDT2	6923	0.8316 ± 0.0041
Region	4468	0.6454 ± 0.0057
Total		$(1.004 \pm 0.017) \cdot 10^{-5}$

$3\pi\mu$ final state

Tables 7.7, 7.8, 7.9, 7.10, 7.11, 7.12 summarize the results of the cut-and-count estimation of the efficiency for the $3\pi\mu$ final state.

Table 7.7: Number of accepted events on MC and efficiency for each selection step for the 2011 $3\pi\mu$ final state. The line between PID and L0 efficiency separates the filtered from the flagged samples.

Step	Accepted events	Efficiency
Generator	108977	$0.040\,473 \pm 0.000\,061$
RecoStrip	13520	0.124 ± 0.001
TruthMatching	7020	0.5192 ± 0.0043
PID	1992	0.2838 ± 0.0054
L0	15375	0.7017 ± 0.0031
HLT1	13801	0.8976 ± 0.0024
HLT2	8907	0.6454 ± 0.0041
IsoCut	5715	0.6416 ± 0.0051
MultCut	5701	$0.997\,55 \pm 0.000\,65$
BDT1	1963	0.3443 ± 0.0063
BDT2	1274	0.649 ± 0.011
Region	882	0.692 ± 0.013
Total		$(2.98 \pm 0.12) \cdot 10^{-5}$

Table 7.8: Number of accepted events on MC and efficiency for each selection step for the 2012 $3\pi\mu$ final state. The line between PID and L0 efficiency separates the filtered from the flagged samples.

Step	Accepted events	Efficiency
Generator	213445	$0.042\,080 \pm 0.000\,061$
RecoStrip	26356	$0.123\,48 \pm 0.000\,71$
TruthMatching	14361	0.5449 ± 0.0031
PID	3457	0.2407 ± 0.0036
L0	26875	0.6218 ± 0.0023
HLT1	25012	0.9307 ± 0.0015
HLT2	17273	0.6906 ± 0.0029
IsoCut	10719	0.6206 ± 0.0037
MultCut	10679	$0.996\,27 \pm 0.000\,59$
BDT1	3746	0.3508 ± 0.0046
BDT2	2430	0.6487 ± 0.0078
Region	1715	0.7058 ± 0.0092
Total		$(2.704 \pm 0.079) \cdot 10^{-5}$

Table 7.9: Number of accepted events on MC and efficiency for each selection step for the 2015 $3\pi\mu$ final state. The line between PID and L0 efficiency separates the filtered from the flagged samples.

Step	Accepted events	Efficiency
Generator	74491	$0.047\,645 \pm 0.000\,092$
RecoStrip	9583	0.1286 ± 0.0012
TruthMatching	4453	0.4647 ± 0.0051
PID	1407	0.316 ± 0.007
L0	4350	0.4078 ± 0.0048
HLT1	4203	0.9662 ± 0.0027
HLT2	2932	0.6976 ± 0.0071
IsoCut	1849	0.6306 ± 0.0089
MultCut	1846	$0.998\,38 \pm 0.000\,94$
BDT1	692	0.375 ± 0.011
BDT2	472	0.682 ± 0.018
Region	320	0.678 ± 0.022
Total		$(2.70 \pm 0.16) \cdot 10^{-5}$

Table 7.10: Number of accepted events on MC and efficiency for each selection step for the 2016 $3\pi\mu$ final state. The line between PID and L0 efficiency separates the filtered from the flagged samples.

Step	Accepted events	Efficiency
Generator	227170	$0.047\,670 \pm 0.000\,092$
RecoStrip	30741	$0.135\,32 \pm 0.000\,72$
TruthMatching	14304	0.4653 ± 0.0028
PID	4987	0.349 ± 0.004
L0	22388	0.5874 ± 0.0025
HLT1	21163	0.9453 ± 0.0015
HLT2	15983	0.755 ± 0.003
IsoCut	10319	0.6456 ± 0.0038
MultCut	10270	$0.995\,25 \pm 0.000\,68$
BDT1	3580	0.3486 ± 0.0047
BDT2	2337	0.653 ± 0.008
Region	1599	0.6842 ± 0.0096
Total		$(4.39 \pm 0.12) \cdot 10^{-5}$

Table 7.11: Number of accepted events on MC and efficiency for each selection step for the 2017 $3\pi\mu$ final state. The line between PID and L0 efficiency separates the filtered from the flagged samples.

Step	Accepted events	Efficiency
Generator	601964	$0.047\,530 \pm 0.000\,092$
RecoStrip	80745	$0.134\,13 \pm 0.000\,44$
TruthMatching	37296	0.4619 ± 0.0018
PID	12739	0.3416 ± 0.0025
L0	56124	0.6766 ± 0.0016
HLT1	52886	$0.942\,31 \pm 0.000\,98$
HLT2	39205	0.7413 ± 0.0019
IsoCut	24962	0.6367 ± 0.0024
MultCut	24856	$0.995\,75 \pm 0.000\,41$
BDT1	8844	0.356 ± 0.003
BDT2	5790	0.6547 ± 0.0051
Region	3997	0.6903 ± 0.0061
Total		$(4.847 \pm 0.087) \cdot 10^{-5}$

Table 7.12: Number of accepted events on MC and efficiency for each selection step for the 2018 $3\pi\mu$ final state. The line between PID and L0 efficiency separates the filtered from the flagged samples.

Step	Accepted events	Efficiency
Generator	759371	$0.047\,622 \pm 0.000\,088$
RecoStrip	102077	$0.134\,42 \pm 0.000\,39$
TruthMatching	47293	0.4633 ± 0.0016
PID	16294	0.3445 ± 0.0022
L0	62818	0.5924 ± 0.0015
HLT1	59280	$0.943\,68 \pm 0.000\,92$
HLT2	44807	0.7559 ± 0.0018
IsoCut	28749	0.6416 ± 0.0023
MultCut	28625	$0.995\,69 \pm 0.000\,39$
BDT1	9989	0.3490 ± 0.0028
BDT2	6574	0.6581 ± 0.0047
Region	4488	0.6827 ± 0.0057
Total		$(4.324 \pm 0.072) \cdot 10^{-5}$

Normalization

Tables 7.13, 7.14, 7.15, 7.16, 7.17, 7.18 summarize the results of the cut-and-count estimation of the efficiency for the normalization mode.

Table 7.13: Number of accepted events on MC and efficiency for each selection step for the 2011 normalization channel.

Step	Accepted events	Efficiency
Generator	512704	$0.112\,88 \pm 0.000\,21$
RecoStrip	8812	$0.017\,19 \pm 0.000\,18$
TruthMatching	8165	0.9266 ± 0.0028
PID	6280	0.7693 ± 0.0047
L0	3390	0.5398 ± 0.0063
HLT1	3158	0.9322 ± 0.0043
HLT2	2719	0.8611 ± 0.0062
CutBased	2293	0.8430 ± 0.0070
Total		$(5.05 \pm 0.11) \cdot 10^{-4}$

Table 7.14: Number of accepted events on MC and efficiency for each selection step for the 2012 normalization channel.

Step	Accepted events	Efficiency
Generator	1004845	$0.115\,62 \pm 0.000\,21$
RecoStrip	15724	$0.015\,65 \pm 0.000\,12$
TruthMatching	14647	0.9320 ± 0.0020
PID	11413	0.7792 ± 0.0034
L0	5633	0.4936 ± 0.0047
HLT1	5290	0.9395 ± 0.0032
HLT2	4846	0.9163 ± 0.0038
CutBased	4165	0.8590 ± 0.0050
Total		$(4.796 \pm 0.075) \cdot 10^{-4}$

Table 7.15: Number of accepted events on MC and efficiency for each selection step for the 2015 normalization channel.

Step	Accepted events	Efficiency
Generator	378804	$0.125\,10 \pm 0.000\,30$
RecoStrip	7187	$0.018\,97 \pm 0.000\,22$
TruthMatching	6762	0.9409 ± 0.0028
PID	5225	0.7727 ± 0.0051
L0	2710	0.5187 ± 0.0069
HLT1	2688	0.9919 ± 0.0017
HLT2	2371	0.8827 ± 0.0062
CutBased	2032	0.8573 ± 0.0072
Total		$(6.71 \pm 0.15) \cdot 10^{-4}$

Table 7.16: Number of accepted events on MC and efficiency for each selection step for the 2016 normalization channel.

Step	Accepted events	Efficiency
Generator	1011373	$0.125\,44 \pm 0.000\,22$
RecoStrip	19585	$0.019\,36 \pm 0.000\,14$
TruthMatching	18158	0.9271 ± 0.0019
PID	14204	0.7823 ± 0.0031
L0	7548	0.5314 ± 0.0042
HLT1	7497	$0.993\,25 \pm 0.000\,94$
HLT2	6971	0.9301 ± 0.0029
CutBased	5908	0.8475 ± 0.0043
Total		$(7.330 \pm 0.096) \cdot 10^{-4}$

Table 7.17: Number of accepted events on MC and efficiency for each selection step for the 2017 normalization channel.

Step	Accepted events	Efficiency
Generator	1006343	$0.125\,47 \pm 0.000\,31$
RecoStrip	19396	$0.019\,27 \pm 0.000\,14$
TruthMatching	18182	0.9376 ± 0.0017
PID	14139	0.7776 ± 0.0031
L0	8184	0.5788 ± 0.0042
HLT1	8124	$0.992\,68 \pm 0.000\,94$
HLT2	7486	0.9210 ± 0.0030
CutBased	6341	0.8471 ± 0.0042
Total		$(7.90 \pm 0.10) \cdot 10^{-4}$

Table 7.18: Number of accepted events on MC and efficiency for each selection step for the 2018 normalization channel.

Step	Accepted events	Efficiency
Generator	1010676	$0.126\,42 \pm 0.000\,61$
RecoStrip	20089	$0.019\,88 \pm 0.000\,14$
TruthMatching	18828	0.9374 ± 0.0017
PID	14859	0.7890 ± 0.0030
L0	8942	0.6020 ± 0.0040
HLT1	8863	$0.991\,56 \pm 0.000\,97$
HLT2	8237	0.9294 ± 0.0027
CutBased	6985	0.8480 ± 0.0040
Total		$(8.74 \pm 0.11) \cdot 10^{-4}$

7.2 PID efficiency

The particle identification requirements applied at the stripping level are summarized in Table 7.19[1].

Table 7.19: Summary of the PID requirements used in the analysis.

Decay channel	Particle	PID requirements
$3\pi 3\pi$	π's from τ	ProbNN$\pi > 0.55$
	π from K^{*0}	ProbNN$\pi > 0.55$
	K from K^{*0}	ProbNN$K > 0.2$
$3\pi\mu$	π's from τ	ProbNN$\pi > 0.55$
	π from K^{*0}	ProbNN$\pi > 0.5$
	K from K^{*0}	ProbNN$K > 0.2$
	μ	PID$\mu > -3$ and IsMuon = True
Normalization	π	ProbNN$\pi > 0.55$
	K	PID$K > -5$

PID variables are known to be not well reproduced by the MC, hence the selection efficiencies of these cuts might differ between data and simulation. A data-driven method [94] is therefore used to determine the efficiency of the PID selection cuts and the systematic uncertainties on it.
An efficiency map is built using large *calibration samples*, consisting of $D^{*+} \to D^0(\to K^-\pi^+)\pi^+$, $D_s^+ \to \phi(\to K^+K^-)\pi^+$, $K^0_S \to \pi^+\pi^-$, $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$ and $J/\psi \to \mu^+\mu^-$ decays identified in data [95]. This map is a function of the particle's momentum (p) and its pseudorapidity (η), and is built separately for the two magnet polarities. The binning schemes vary for signal and normalization modes and depending on the mass hypothesis of the final state particles, spanning the ranges $p \in [0, 200]$ GeV and $\eta \in [1.5, 5]$. The map is used to assign an efficiency value to each track of a *reference sample* (the sample whose efficiency is being computed) according to the momentum and pseudorapidity bin the track belongs to. The PID efficiency is then computed as:

$$\epsilon_{\text{PID}} = \frac{1}{N} \sum_i \prod_j \epsilon_j^i, \tag{7.4}$$

where ϵ_j^i is the efficiency assigned to the j-th track of the i-th event and N is the total number of events in the reference sample.
The reference samples are obtained with a modified version of the stripping selection

[1] The cut ProbNN$\pi > 0.5$ applied on the π from the K^{*0} for the $3\pi\mu$ final state is tightened to ProbNN$\pi > 0.55$ in the final version of the analysis, in order to have a homogeneous set of cuts on the pions in the final state.

in which the PID requirements are removed in order not to introduce biases in the procedure. They include the neutral pion component for the rare mode, and the logical OR of loose-matched and swapped-pions requirements is applied. Since the event multiplicity is greater than unity on the rare mode, one candidate per event is randomly chosen. For the normalization mode, truth-matched candidates are used.
In order to assess a systematic uncertainty on the choice of the binning scheme, the procedure is repeated using maps in which the number of momentum and pseudorapidity bins is doubled or halved with respect to the standard configuration. The uncertainty is chosen to be half the difference between the highest and lowest efficiency values obtained with the three versions of the binning schemes. The uncertainty on the size of the calibration samples is neglected since the samples are much larger than the MC used as reference samples, while the uncertainty on the size of the reference samples is taken into account in the estimation of the stripping efficiency, and is not double-counted. The results are reported in Tables 7.20, 7.21, 7.22.

Table 7.20: PID efficiency for the $3\pi3\pi$ final state obtained from the data-driven method. The results are shown for the default ϵ^{PID}, the doubled $\epsilon^{\text{PID}}_{\text{double}}$ and the halved $\epsilon^{\text{PID}}_{\text{half}}$ binning schemes. The uncertainty includes only the binning systematics.

Year	Polarity	ϵ^{PID}	$\epsilon^{\text{PID}}_{\text{double}}$	$\epsilon^{\text{PID}}_{\text{half}}$	Uncertainty
2011	Up	0.533	0.519	0.549	0.015
2012	Up	0.5305	0.535	0.5452	0.0074
2015	Up	0.6216	0.61151	0.6281	0.0065
2016	Up	0.640	0.630	0.650	0.010
2017	Up	0.6696	0.6619	0.6786	0.0083
2018	Up	0.6569	0.6501	0.6647	0.0073
2011	Down	0.533	0.519	0.548	0.015
2012	Down	0.542	0.524	0.559	0.017
2015	Down	0.6375	0.6319	0.6434	0.0057
2016	Down	0.6440	0.6370	0.6530	0.0080
2017	Down	0.6833	0.6766	0.6919	0.0077
2018	Down	0.6574	0.6504	0.6656	0.0076

Comparing the efficiencies with the ones obtained from simulations (Sec. 7.1), large discrepancies are observed for the rare $3\pi3\pi$ and $3\pi\mu$ channels. To understand the nature of the differences, a preliminary study has been performed using the 2016 MagUp $3\pi3\pi$ MC simulation (without neutral pions component). The PID efficiencies obtained with the data-driven method and the cut-and-count procedure are compared for three different categories of candidates:

fully-matched OR swapped-pions (category A): candidates which contain only tracks matched to a true MC particle;

Table 7.21: PID efficiency for the $3\pi\mu$ final state obtained from the data-driven method. The results are shown for the default ϵ^{PID}, the doubled $\epsilon^{\text{PID}}_{\text{double}}$ and the halved $\epsilon^{\text{PID}}_{\text{half}}$ binning schemes. The uncertainty includes only the binning systematics.

Year	Polarity	ϵ^{PID}	$\epsilon^{\text{PID}}_{\text{double}}$	$\epsilon^{\text{PID}}_{\text{half}}$	Uncertainty
2011	Up	0.683	0.670	0.691	0.011
2012	Up	0.6883	0.6768	0.6924	0.0078
2015	Up	0.7368	0.7322	0.7377	0.0027
2016	Up	0.7626	0.7578	0.7648	0.0035
2017	Up	0.7845	0.7815	0.7858	0.0022
2018	Up	0.7790	0.7760	0.7800	0.0020
2011	Down	0.6883	0.6778	0.6944	0.0083
2012	Down	0.6942	0.6811	0.7010	0.0099
2015	Down	0.7459	0.7411	0.7469	0.0029
2016	Down	0.7680	0.7646	0.7713	0.0034
2017	Down	0.7920	0.7900	0.7940	0.0020
2018	Down	0.7797	0.7772	0.7809	0.0018

loose-matched OR swapped-pions (category B): candidates selected with the same truth-matching requirement as the reference sample;

"ghosts" (category C): candidates fulfilling the requirement of belonging to category `B AND NOT A`, therefore containing at least one track *not* matched to any MC true track.

The results are summarized in Table 7.23. The study shows that the two methods are in good agreement for category A, while significant differences are found for categories B and C. Since the differences arise only when tracks not matched to any MC true particle are present and appear too large to be caused by genuine mismodelling of the PID efficiency in the simulation, they are most probably due to the fact that the calibration samples are not representative of the reference sample when pion decaying in-flight are present (producing tracks not matched to any MC true particle). Therefore the results of this study suggest to compute the corrections to the PID efficiencies using only candidates matched to a true MC particle, and to apply them to the PID efficiency of the whole reference sample. This strategy is adopted in the final version of the analysis.

Table 7.22: PID efficiency for the normalization mode obtained from the data-driven method. The results are shown for the default ϵ^{PID}, the doubled $\epsilon^{\text{PID}}_{\text{double}}$ and the halved $\epsilon^{\text{PID}}_{\text{half}}$ binning schemes. The uncertainty includes only the binning systematics.

Year	Polarity	ϵ^{PID}	$\epsilon^{\text{PID}}_{\text{double}}$	$\epsilon^{\text{PID}}_{\text{half}}$	Uncertainty
2011	Up	0.7719	0.7659	0.7741	0.0041
2012	Up	0.7631	0.7572	0.7645	0.0037
2015	Up	0.7691	0.7673	0.7719	0.0023
2016	Up	0.7892	0.7861	0.7926	0.0032
2017	Up	0.8048	0.8023	0.8071	0.0024
2018	Up	0.8030	0.8009	0.8044	0.0018
2011	Down	0.7732	0.7676	0.7759	0.0042
2012	Down	0.7694	0.7632	0.7710	0.0039
2015	Down	0.7978	0.7960	0.7990	0.0015
2016	Down	0.7950	0.7930	0.7970	0.0020
2017	Down	0.8143	0.8122	0.8159	0.0019
2018	Down	0.8019	0.8009	0.8031	0.0011

Table 7.23: PID efficiency computed with the data-driven and the standard cut-and-count methods for candidates classified in category A, B and C of the 2016 MagUp $3\pi3\pi$ MC simulation.

Computation method	Category A	Category B	Category C
Cut-and-count	68.23 ± 0.82	51.68 ± 0.71	20.43 ± 0.90
Data-driven	68.55 ± 0.55	66.52 ± 0.72	62.80 ± 1.11

7.3 Systematic uncertainty from data-MC comparison

Iterative reweighting

The differences between data and MC simulation are checked using the $B^0 \to D^- D_s^+$ control channel for the $3\pi3\pi$ and $3\pi\mu$ final states. The drawback consists in the lack of an analogous for the K^{*0} decay and for the muon. The variable distributions in data are weighted using the *sPlot* technique [96] in order to obtain background-subtracted data. The comparison for the variables used in the cut-based and MVA stages of the selection is shown in Appendix C.
For a given variable, data-MC agreement is quantified using a reduced χ^2 variable:

$$\chi^2/\text{N}_{\text{dof}} = \frac{1}{n_{\text{bins}}} \sum_{i=1}^{n_{\text{bins}}} \frac{(f_i^{\text{MC}} - f_i^{\text{data}})^2}{\delta f_i^{\text{MC}^2} + \delta f_i^{\text{data}^2}}, \tag{7.5}$$

where

$$f_i^x \pm \delta f_i^x \equiv N_i^x/N^x \pm \sqrt{N_i^x}/N^x \tag{7.6}$$

is the fraction of events in bin i, x refers to either normalization MC or data, n_{bins} is the total number of bins, N_i^x is the number of events in bin i passing the selection requirements and N^x is the total number of events after selection. The fractions of events in each bin of data and MC are used to compute a set of weights:

$$w_i \pm \delta w_i \equiv \frac{f_i^{\text{data}}}{f_i^{\text{MC}}} \pm \sqrt{\left(\frac{\delta f_i^{\text{data}}}{f_i^{\text{MC}}}\right)^2 + \left(\frac{f_i^{\text{data}}}{f_i^{\text{MC}^2}}\delta f_i^{\text{MC}}\right)^2}. \tag{7.7}$$

The signal MC samples of the rare modes can then be reweighted on an event-by-event basis by assigning each event a weight depending on the bin it belongs to. A corrected selection efficiency is defined as

$$\epsilon \pm \delta\epsilon \equiv \frac{\sum_i w_i n_i}{\sum_i w_i N_i} \pm \frac{\sum_i [w_i^2(1-2\epsilon)\delta n_i^2 + w_i^2\epsilon^2\delta N_i^2 + (n_i - N_i\epsilon)^2\delta w_i^2]}{\sum_i (w_i N_i)^2}, \tag{7.8}$$

where n_i is the number of events in bin i after the full selection and the other terms have been already defined.

Due to correlations between the variables, the data-MC agreement of a sample reweighted with a given variable can impact the data-MC agreement of other variables. Therefore, an iterative procedure is used in order to compute the corrected efficiency:

1. the variable showing the worst $\chi^2/\text{N}_{\text{dof}}$ for each data-taking year is considered and weights are computed using the control channel;

2. the weights are applied to the signal sample and the corrected efficiency is computed;

3. the $\chi^2/\text{N}_{\text{dof}}$ is recomputed on the reweighted sample;

4. steps 1–3 are iterated. The weights for a given iteration are computed as the product of the weights at each step.

If multiple candidates per event are present in the dataset, Eq. 7.8 is ill-defined. In fact, if on one hand the weights to be assigned to the event after the selection are uniquely identified (because one candidate per event is randomly chosen after the full selection), the weights to be assigned to the events before the selection can not be determined, since multiple choices are possible, one for each candidate belonging to the event. To be able to apply the iterative procedure described above, the MC samples are divided in three subsamples: fully-matched candidates, swapped-pions candidates and candidates containing at least one track not matched to any MC true particle ("ghosts"). Moreover one candidate per event is randomly

chosen in each subsample *before* the selection, in order to be able to use Eq. 7.8 for the efficiency computation. The iterative procedure is applied separately to each subsample: three iterations are performed and the value of the efficiency after the third one is compared to the default value. The difference is used in the final version of the analysis to assign a systematic uncertainty on the signal efficiency for a given subsample[2].

Corrections to the $3\pi3\pi$ final state

The χ^2/N_{dof} distributions of the variables used for the $3\pi3\pi$ final state are shown in Fig. 7.1 for each data-taking year, while the variables with the highest χ^2/N_{dof} are reported in Table 7.24.

Table 7.24: $3\pi3\pi$ final state variables with the highest χ^2/N_{dof} for each data-taking years before any correction.

Year	Variable	χ^2/N_{dof}
2011	τ^-IsoBDTFirstValue%100	4.23
2012	$\tau^+\pi_1^+$ IsoBDTThirdValue	3.94
2015	B_OPT_TauP_decayLength	4.32
2016	τ^-IsoBDTFirstValue%100	3.53
2017	$\tau^-\pi_1^-$IsoBDTFirstValue%100	5.20
2018	$\tau^+\pi^-$IsoBDTFirstValue%100	4.47

The χ^2/N_{dof} distributions are shown in Figs. 7.2, 7.3 and 7.4 after the first, second and third iteration respectively, while the variables showing the worst χ^2/N_{dof} are reported in Tables 7.25, 7.26 and 7.27.

Table 7.25: $3\pi3\pi$ final state variables with the highest χ^2/N_{dof} for each data-taking years after the first iteration.

Year	Variable	χ^2/N_{dof}
2011	$\tau^-\pi^+$IsoBDTFirstValue%100	3.84
2012	τ^+IsoBDTFirstValue%100	7.75
2015	B_OPT_TauM_decayLength	4.40
2016	$\tau^+\pi^-$IsoBDTFirstValue%100	4.14
2017	B_OPT_chi2	3.69
2018	B_OPT_chi2	3.98

[2]In the latest version of the analysis the uncertainties on the subsamples are combined in a weighted average to obtain a global systematic uncertainty. In addition to assigning a systematic uncertainty on the efficiencies, the weights obtained with this procedure are used to compute a systematic uncertainty on the shape of the fitBDT.

Table 7.26: $3\pi3\pi$ final state variables with the highest χ^2/N_{dof} for each data-taking years after the second iteration.

Year	Variable	χ^2/N_{dof}
2011	τ^- IsoBDTThirdValue	2.57
2012	$\tau^-\pi_1^-$IsoBDTFirstValue%100	3.97
2015	$\tau^+\pi_2^+$IsoBDTFirstValue%100	3.20
2016	B_IP_OWNPV	3.17
2017	B_BPVVDRHO	3.09
2018	τ^+ DOCAMAX	2.66

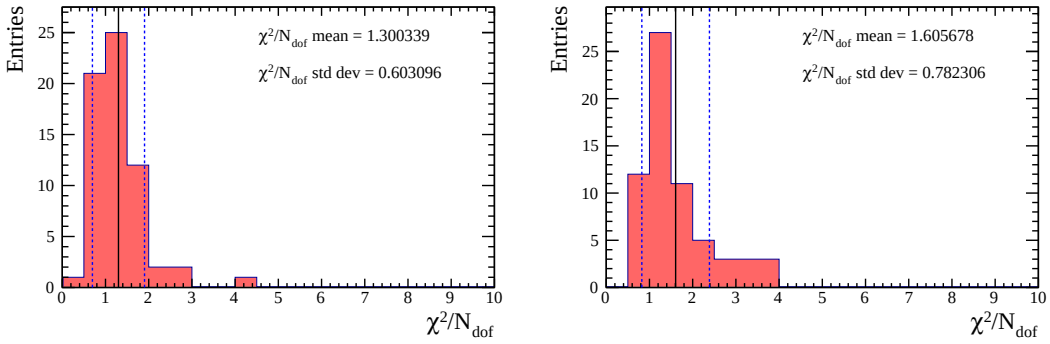
Table 7.27: $3\pi3\pi$ final state variables with the highest χ^2/N_{dof} for each data-taking years after the third iteration.

Year	Variable	χ^2/N_{dof}
2011	$\tau^+\pi_1^+$IsoBDTFirstValue%100	3.46
2012	τ^-IsoBDTFirstValue%100	4.84
2015	$\tau^-\pi^+$IsoBDTFirstValue%100	3.34
2016	$\tau^-\pi_2^-$IsoBDTFirstValue%100	3.15
2017	$\tau^+\pi_1^+$IsoBDTFirstValue%100	2.97
2018	B_BPVVDRHO	3.14

The corrected efficiencies are shown in Figs. 7.5, 7.6 and 7.7 for the three subsamples defined above. Table 7.28 shows the absolute difference between the default efficiency value and the corrected one after the third iteration for each subsample and year of data-taking.

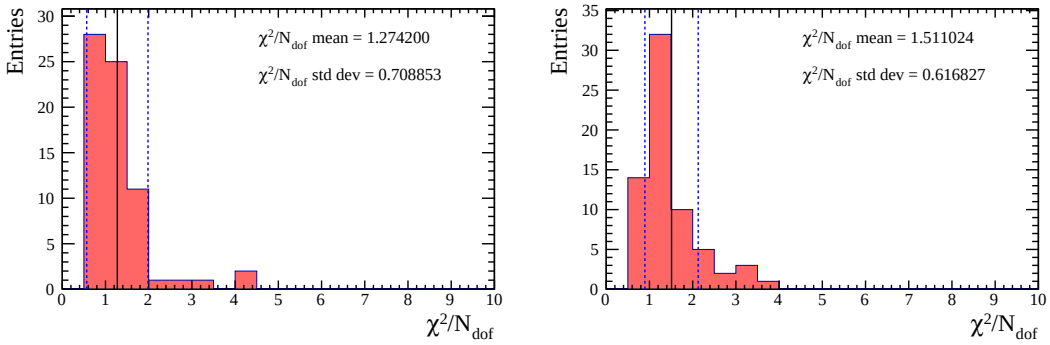
Table 7.28: Absolute difference between the default and the corrected efficiency value after the third iteration. The results are shown for each data-taking year and each subsample of the $3\pi3\pi$ final state.

Year	Fully-matched	Swapped-pions	ghosts
2011	0.0131	0.0021	0.0182
2012	0.0033	0.0013	0.0020
2015	0.0195	0.0047	0.0177
2016	0.0035	0.0007	0.0032
2017	0.0110	0.0031	0.0145
2018	0.0021	0.0	0.0063



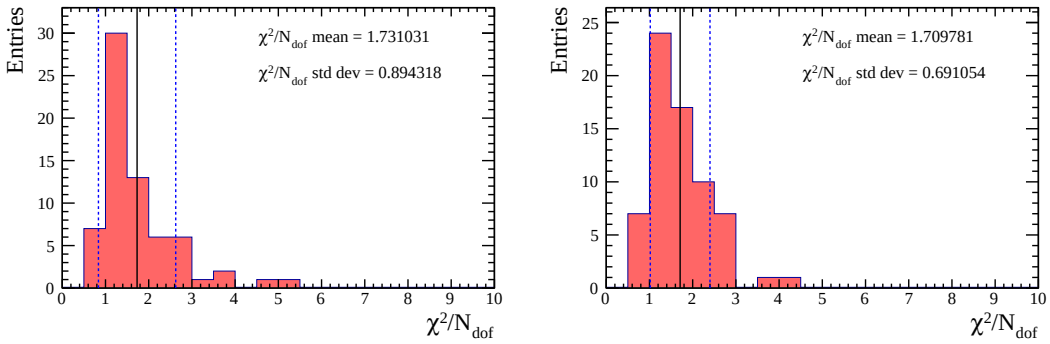
(a) 2011

(b) 2012



(c) 2015

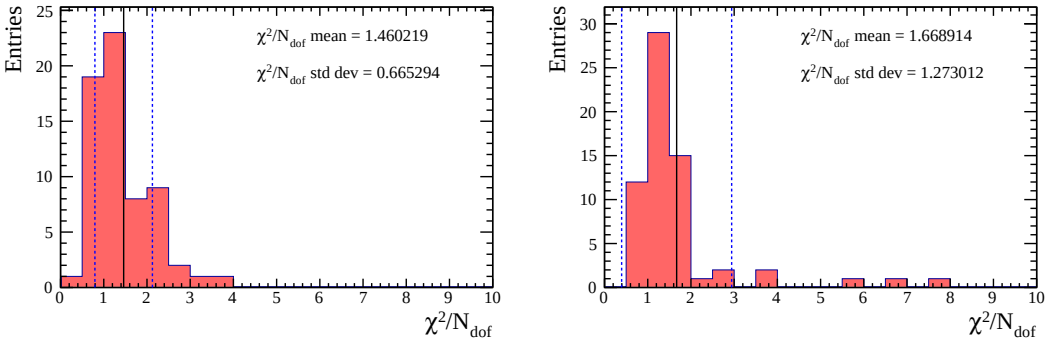
(d) 2016



(e) 2017

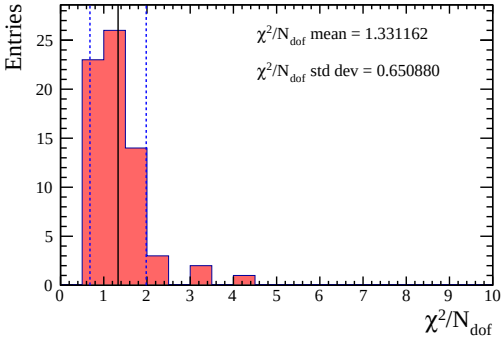
(f) 2018

Figure 7.1: χ^2/N_{dof} distributions of the variables used for the $3\pi3\pi$ final state before any correction. The mean and standard deviation of the distributions are reported and visualized with a black solid line and two blue dashed lines (referring respectively to the mean and $\pm1\sigma$ interval).

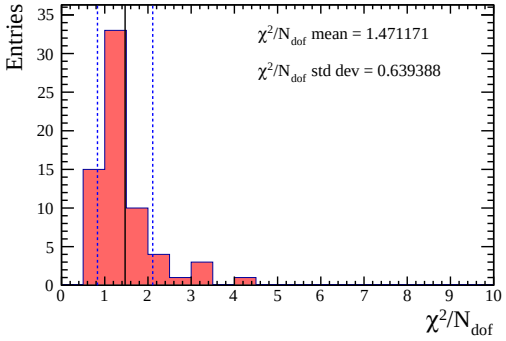


(a) 2011

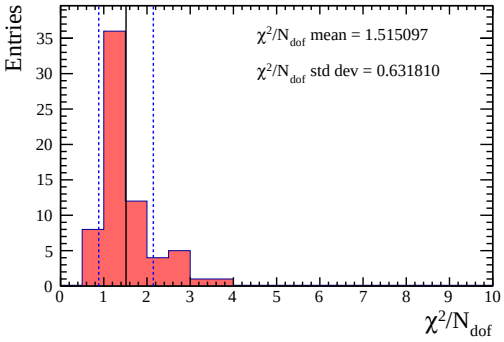
(b) 2012



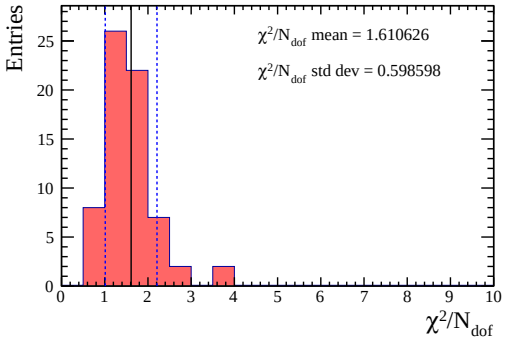
(c) 2015



(d) 2016



(e) 2017



(f) 2018

Figure 7.2: χ^2/N_{dof} distributions of the variables used for the $3\pi3\pi$ final state after the first iteration. The mean and standard deviation of the distributions are reported and visualized with a black solid line and two blue dashed lines (referring respectively to the mean and $\pm1\sigma$ interval).

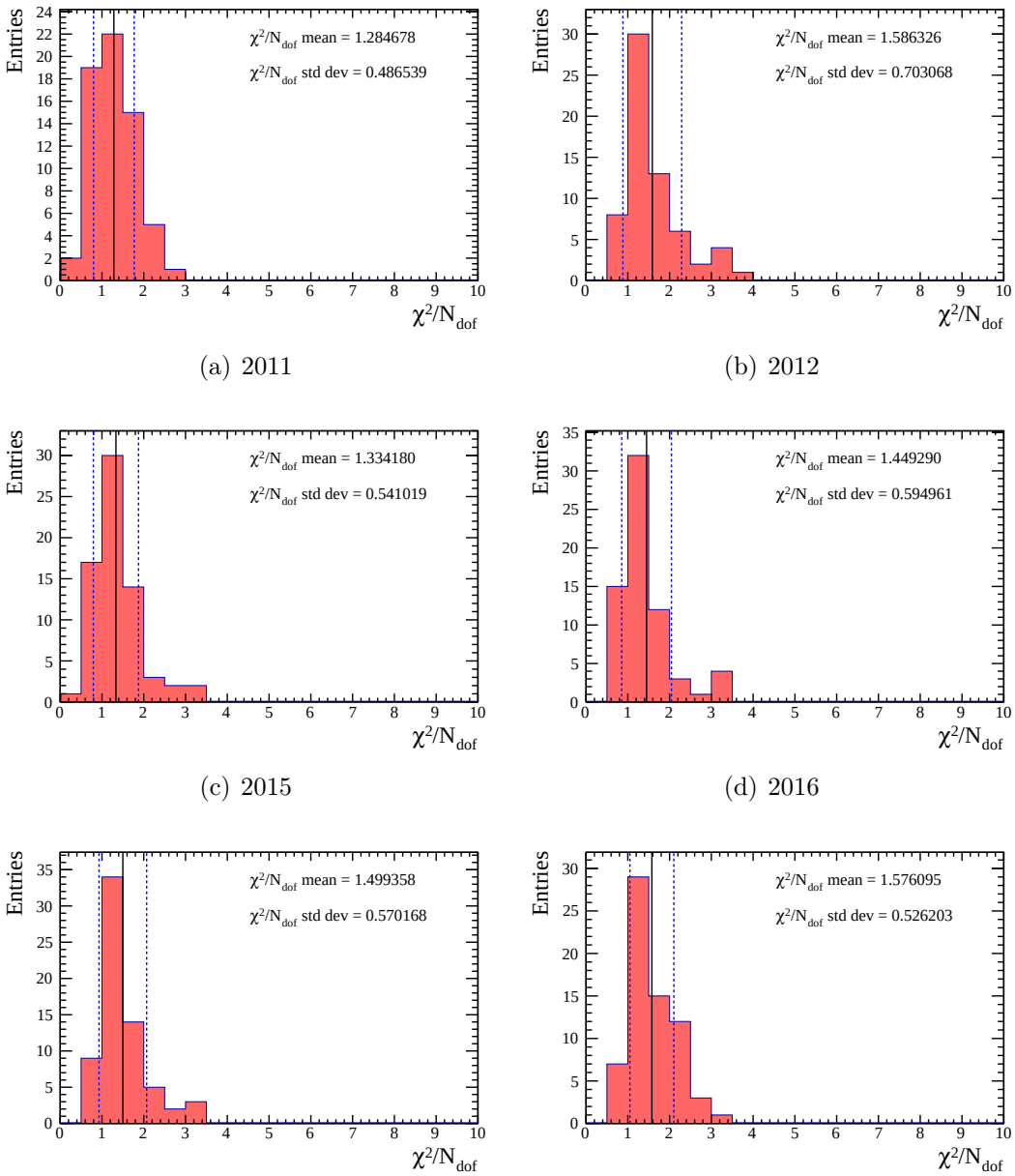


Figure 7.3: χ^2/N_{dof} distributions of the variables used for the $3\pi3\pi$ final state after the second iteration. The mean and standard deviation of the distributions are reported and visualized with a black solid line and two blue dashed lines (referring respectively to the mean and $\pm1\sigma$ interval). The rightmost variable in the 2018 dataset has been already corrected in the previous iteration and its relatively high χ^2/N_{dof} is due to differences between data and MC in low-statistics bins in the tail of the distribution that can not be further corrected. Therefore it is ignored and the variable with the second highest χ^2/N_{dof} is used to compute the weight.

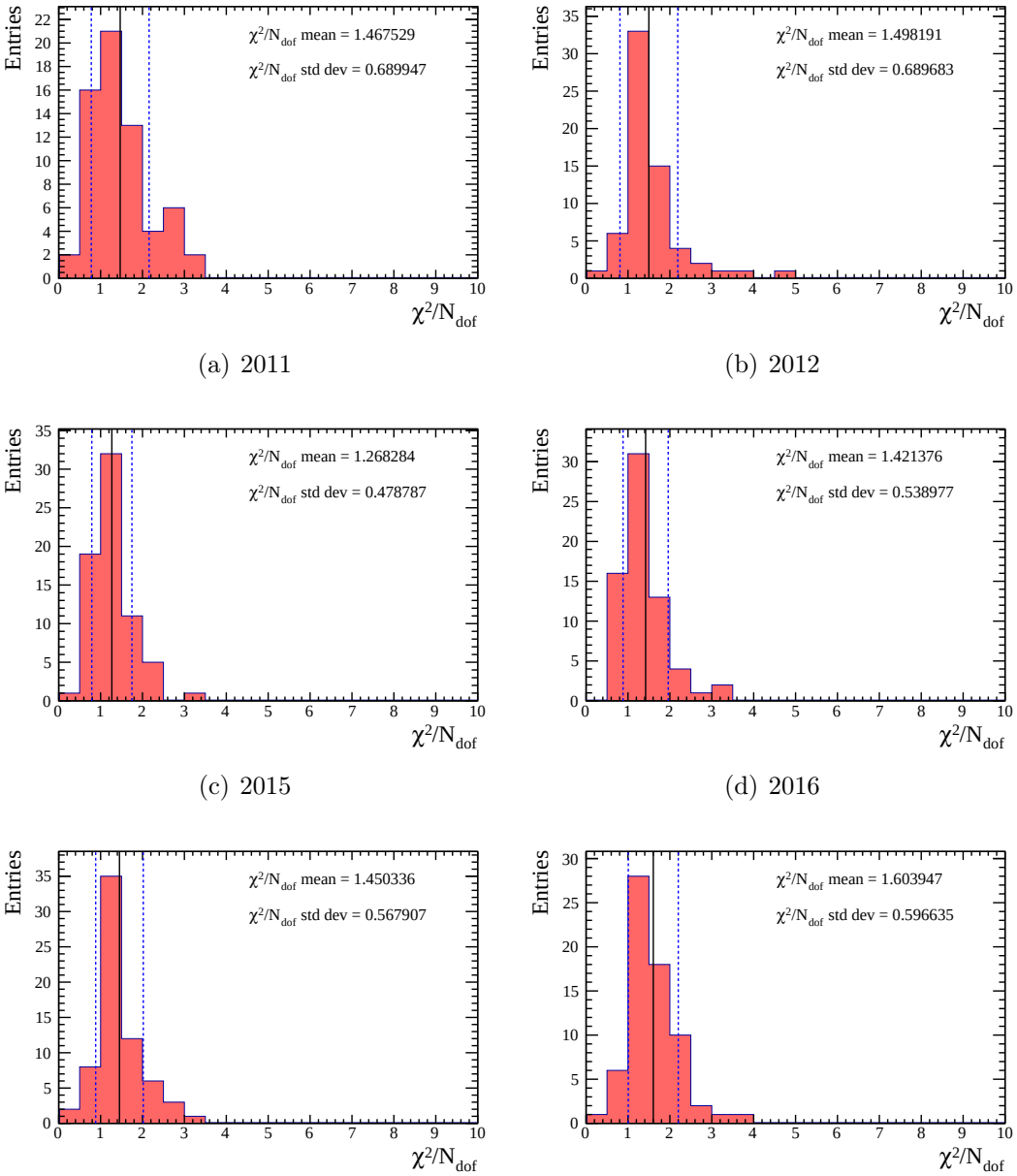
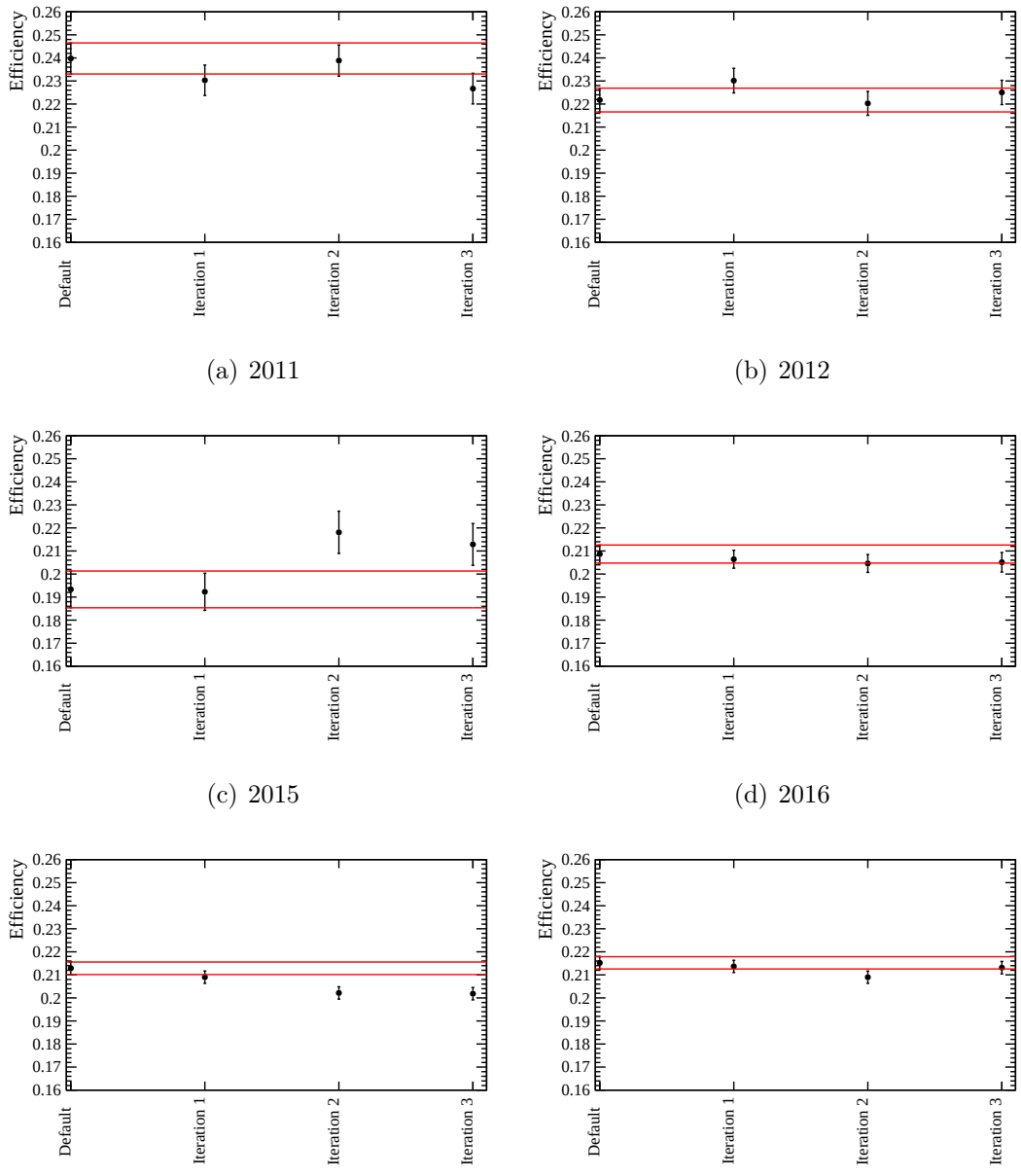


Figure 7.4: χ^2/N_{dof} distributions of the variables used for the $3\pi3\pi$ final state after the third iteration. The mean and standard deviation of the distributions are reported and visualized with a black solid line and two blue dashed lines (referring respectively to the mean and $\pm1\sigma$ interval). The rightmost variables in the 2017 and 2018 datasets have been already corrected in the first iteration and their relatively high χ^2/N_{dof} are due to differences between data and MC in low-statistics bins in the tail of the distributions that can not be further corrected. Therefore they are ignored and the variables with the second highest χ^2/N_{dof} are considered.

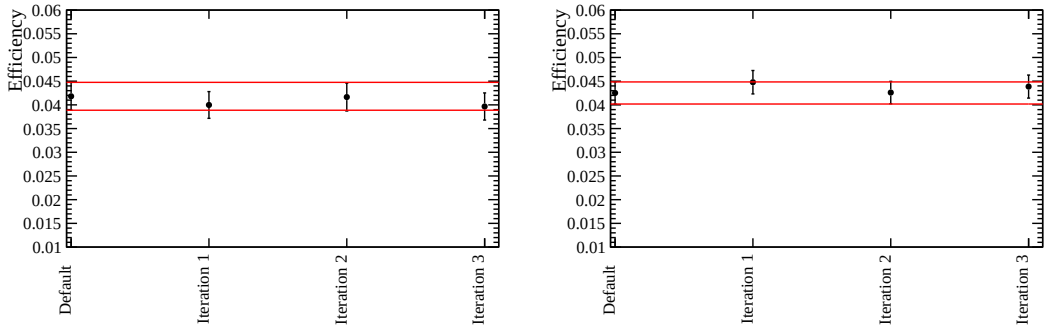


(a) 2011 (b) 2012

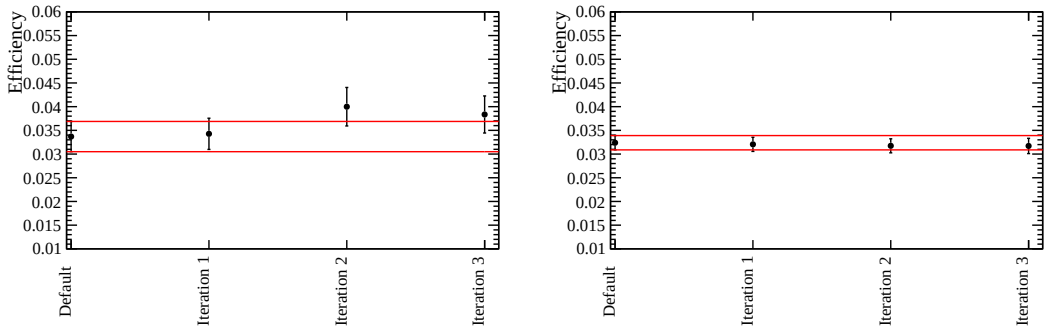
(c) 2015 (d) 2016

(e) 2017 (f) 2018

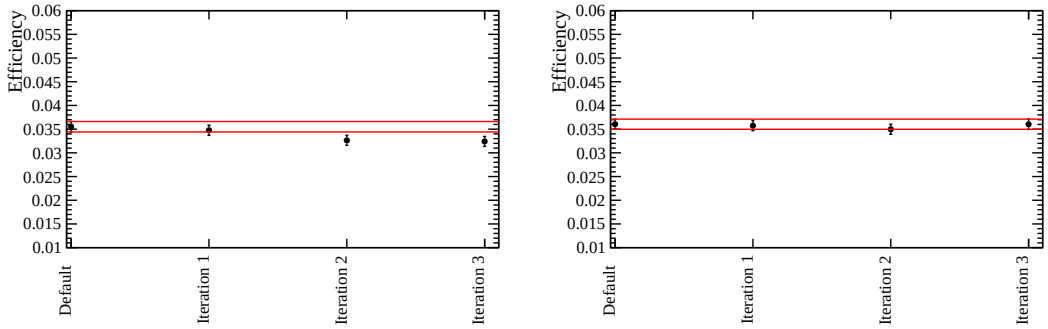
Figure 7.5: Evolution of the selection efficiency for the fully-matched candidates of the $3\pi 3\pi$ final state as a function of the correction iterations. The red lines correspond to the $\pm 1\sigma$ interval around the central value of the default efficiency.



(a) 2011 (b) 2012

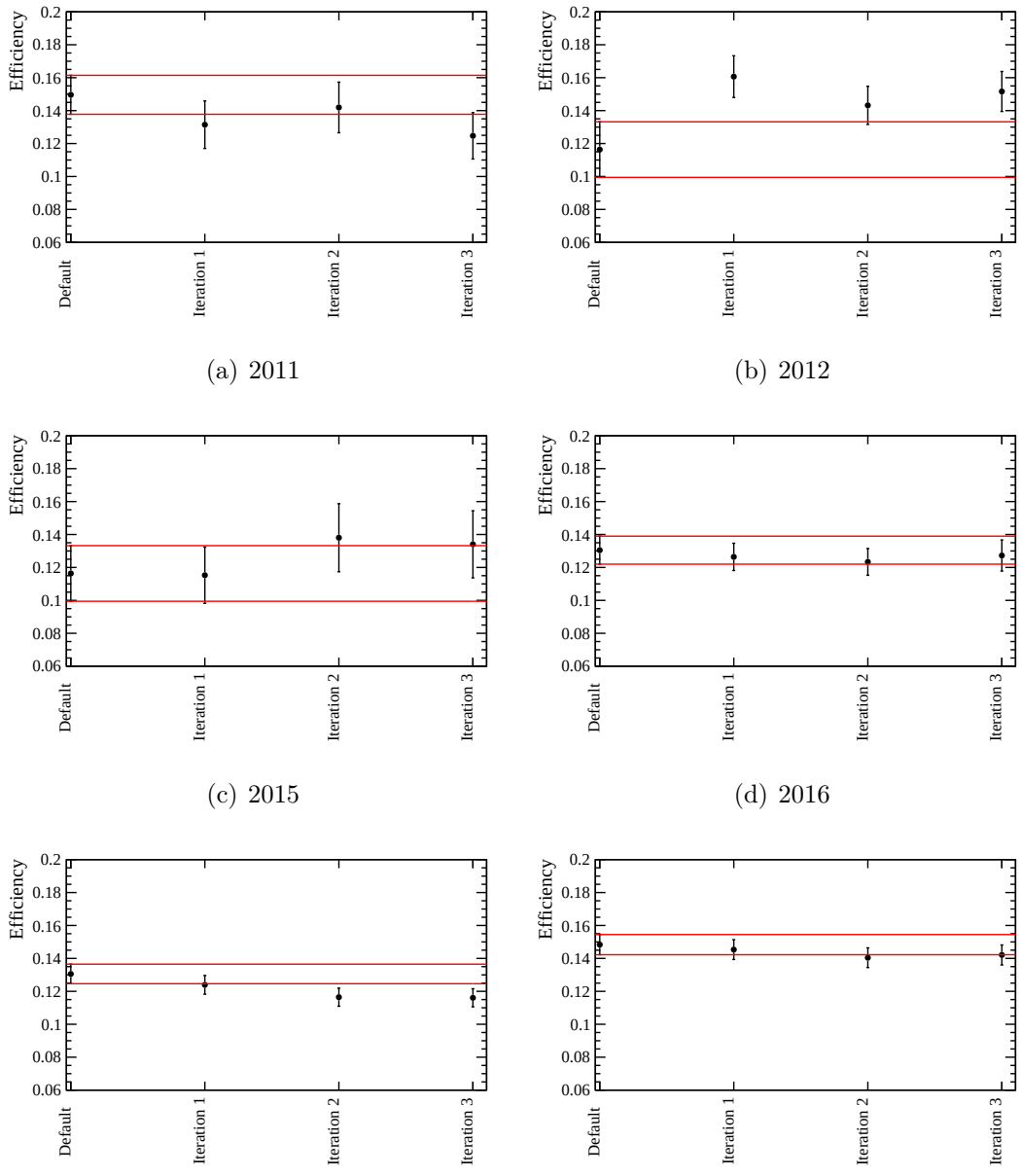


(c) 2015 (d) 2016



(e) 2017 (f) 2018

Figure 7.6: Evolution of the selection efficiency for the swapped-pions candidates of the $3\pi 3\pi$ final state as a function of the correction iterations. The red lines correspond to the $\pm 1\sigma$ interval around the central value of the default efficiency.



(a) 2011

(b) 2012

(c) 2015

(d) 2016

(e) 2017

(f) 2018

Figure 7.7: Evolution of the selection efficiency for the "ghosts" candidates of the $3\pi3\pi$ final state as a function of the correction iterations. The red lines correspond to the $\pm1\sigma$ interval around the central value of the default efficiency.

Corrections to the $3\pi\mu$ final state

The χ^2/N_{dof} distributions of the variables used for the $3\pi\mu$ final state are shown in Fig. 7.8 for each data-taking year, while the variables with the highest χ^2/N_{dof} are reported in Table 7.29.

Table 7.29: $3\pi\mu$ final state variables with the highest χ^2/N_{dof} for each data-taking years before any correction.

Year	Variable	χ^2/N_{dof}
2011	τ^+IsoBDTFirstValue%100	2.96
2012	$\tau^+\pi_1^+$ IsoBDTThirdValue	3.94
2015	$\tau^+\pi_1^+$IsoBDTFirstValue%100	2.80
2016	$\tau^+\pi^-$IsoBDTFirstValue%100	2.91
2017	$\tau^+\pi_1^+$IsoBDTFirstValue%100	3.95
2018	$\tau^+\pi^-$IsoBDTFirstValue%100	4.47

The χ^2/N_{dof} distributions are shown in Figs. 7.9, 7.10 and 7.11 after the first, second and third iteration respectively, while the variables showing the worst χ^2/N_{dof} are reported in Tables 7.30, 7.31 and 7.32.

Table 7.30: $3\pi\mu$ final state variables with the highest χ^2/N_{dof} for each data-taking years after the first iteration.

Year	Variable	χ^2/N_{dof}
2011	$\tau^+\pi_2^+$IsoBDTThirdValue	2.36
2012	τ^+IsoBDTFirstValue%100	7.75
2015	τ^+ IsoBDTSecondValue	1.85
2016	τ^+IsoBDTFirstValue%100	1.94
2017	B_ENDVERTEX_CHI2	1.89
2018	τ^+ IsoBDTSecondValue	2.16

The corrected efficiencies are shown in Figs. 7.12, 7.13 and 7.14 for each of the three subsamples defined above. Table 7.33 shows the absolute difference between the default efficiency value and the corrected one after the third iteration for each subsample and year of data-taking.

Table 7.31: $3\pi\mu$ final state variables with the highest $\chi^2/\mathrm{N_{dof}}$ for each data-taking years after the second iteration.

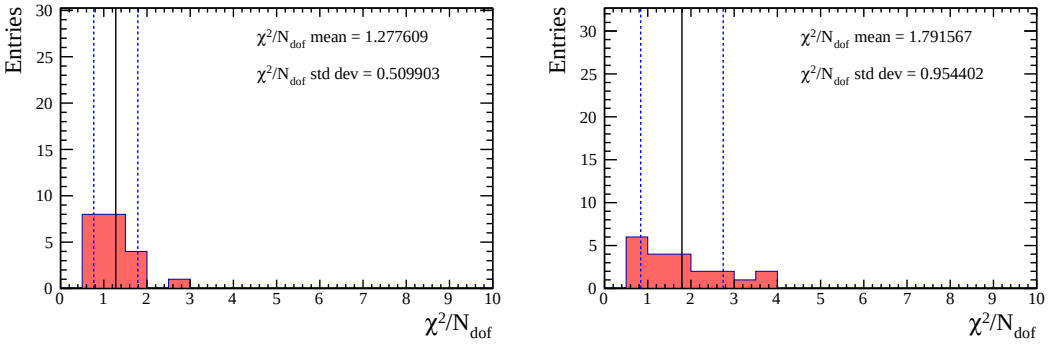
Year	Variable	$\chi^2/\mathrm{N_{dof}}$
2011	τ^+IsoBDTThirdValue	2.26
2012	τ^+IsoBDTThirdValue	3.30
2015	$\tau^+ M_{\pi_1\pi_3}$	1.48
2016	$\tau^+\pi^+{}_1$IsoBDTFirstValue%100	1.69
2017	τ^+IsoBDTThirdValue	1.86
2018	B^0 neutral isolation $\mathrm{P_{asym}}$	2.17

Table 7.32: $3\pi\mu$ final state variables with the highest $\chi^2/\mathrm{N_{dof}}$ for each data-taking years after the third iteration.

Year	Variable	$\chi^2/\mathrm{N_{dof}}$
2011	$\tau^+\pi^+{}_1$IsoBDTFirstValue%100	2.67
2012	$\tau^+\pi^-$IsoBDTFirstValue%100	4.10
2015	$\tau^+ M_{\pi_1\pi_2}$	1.44
2016	τ^+IsoBDTSecondValue	1.68
2017	$\tau^+\pi_2^+$IsoBDTSecondValue	1.96
2018	$\tau^+\pi_1^+$IsoBDTFirstValue%100	2.16

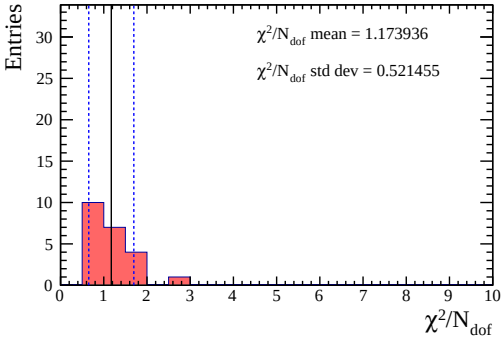
Table 7.33: Absolute difference between the default and the corrected efficiency value after the third iteration. The results are shown for each data-taking year and each subsample of the $3\pi\mu$ final state.

Year	Fully-matched	Swapped-pions	ghosts
2011	0.0061	0.0008	0.0218
2012	0.0038	0.0009	0.0113
2015	0.0168	0.0035	0.0067
2016	0.0024	0.0005	0.0056
2017	0.0074	0.0019	0.0168
2018	0.0021	0.0004	0.0107

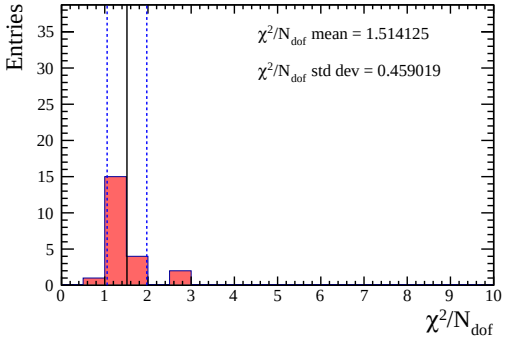


(a) 2011

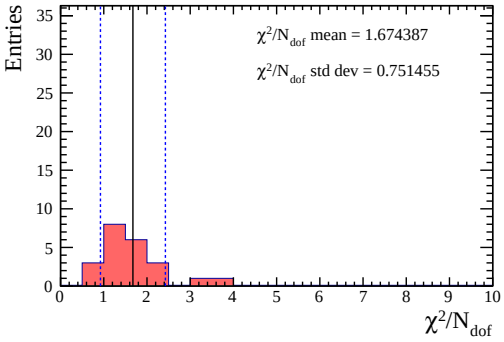
(b) 2012



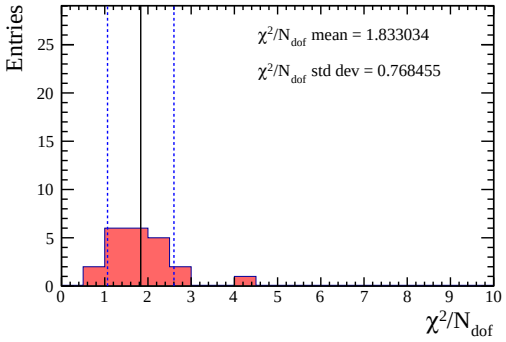
(c) 2015



(d) 2016



(e) 2017



(f) 2018

Figure 7.8: χ^2/N_{dof} distributions of the variables used for the $3\pi\mu$ final state before any correction. The mean and standard deviation of the distributions are reported and visualized with a black solid line and two blue dashed lines (referring respectively to the mean and $\pm 1\sigma$ interval).

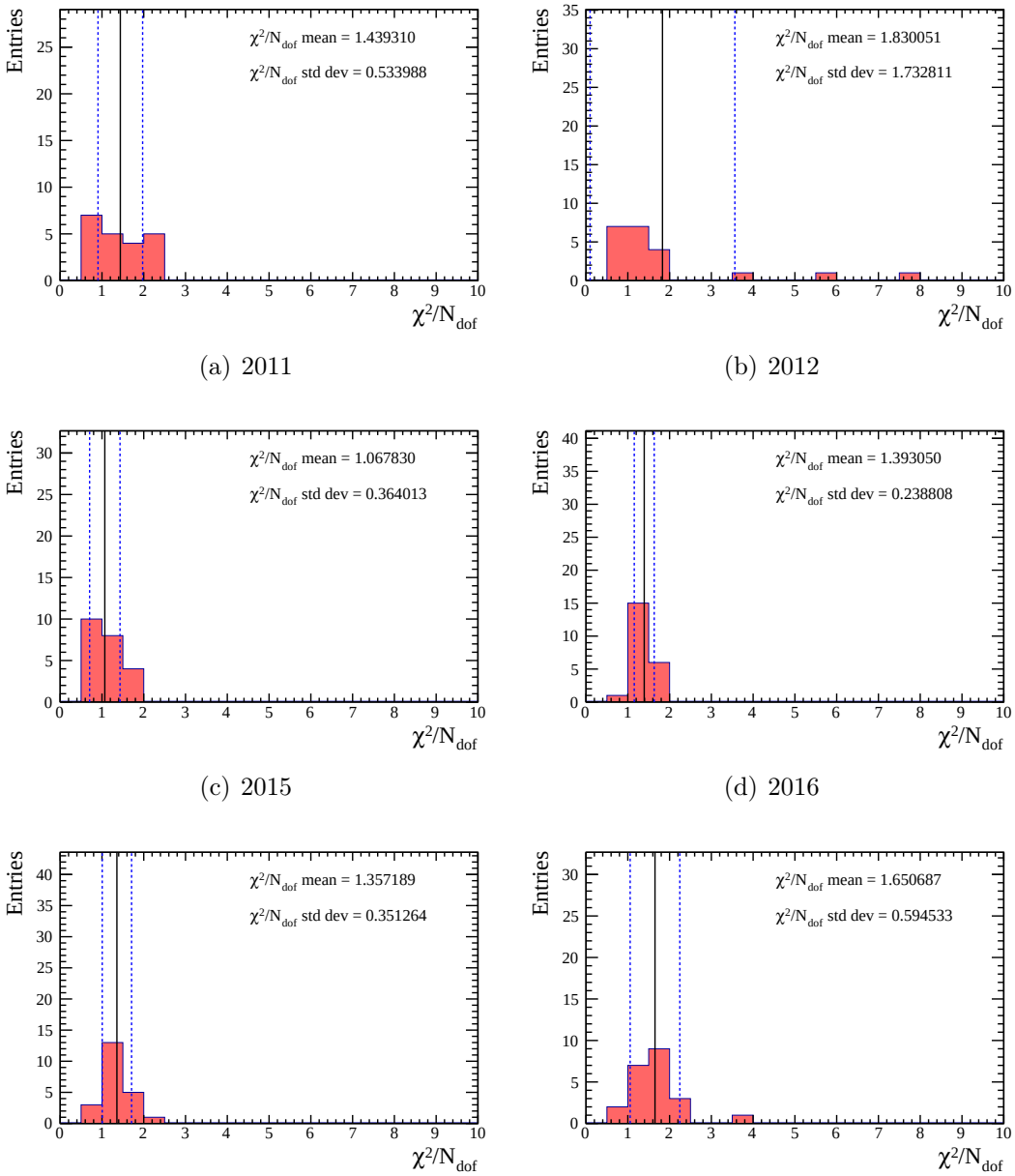


Figure 7.9: χ^2/N_{dof} distributions of the variables used for the $3\pi\mu$ final state after the first iteration. The mean and standard deviation of the distributions are reported and visualized with a black solid line and two blue dashed lines (referring respectively to the mean and $\pm 1\sigma$ interval). The rightmost variables in the 2017 and 2018 datasets have been already corrected in the first iteration and their relatively high χ^2/N_{dof} are due to differences between data and MC in low-statistics bins in the tail of the distributions that can not be further corrected. Therefore they are ignored and the variables with the second highest χ^2/N_{dof} are considered.

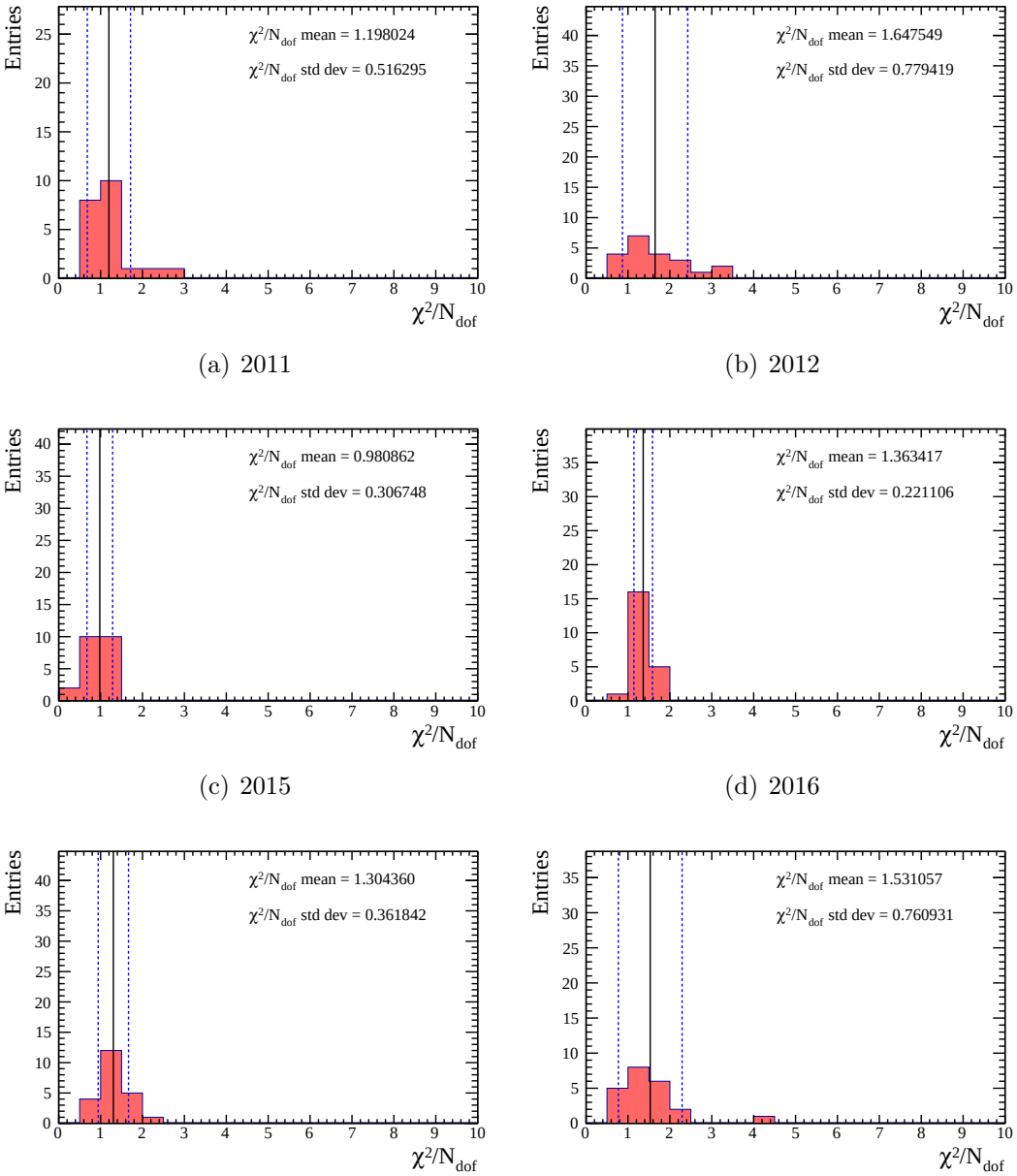


Figure 7.10: χ^2/N_{dof} distributions of the variables used for the $3\pi\mu$ final state after the second iteration. The mean and standard deviation of the distributions are reported and visualized with a black solid line and two blue dashed lines (referring respectively to the mean and $\pm 1\sigma$ interval). The rightmost variables in the 2011, 2017 and 2018 datasets have been already corrected in the first iteration and their relatively high χ^2/N_{dof} are due to differences between data and MC in low-statistic bins in the tail of the distributions that can not be further corrected. Therefore they are ignored and the variables with the second highest χ^2/N_{dof} are considered.

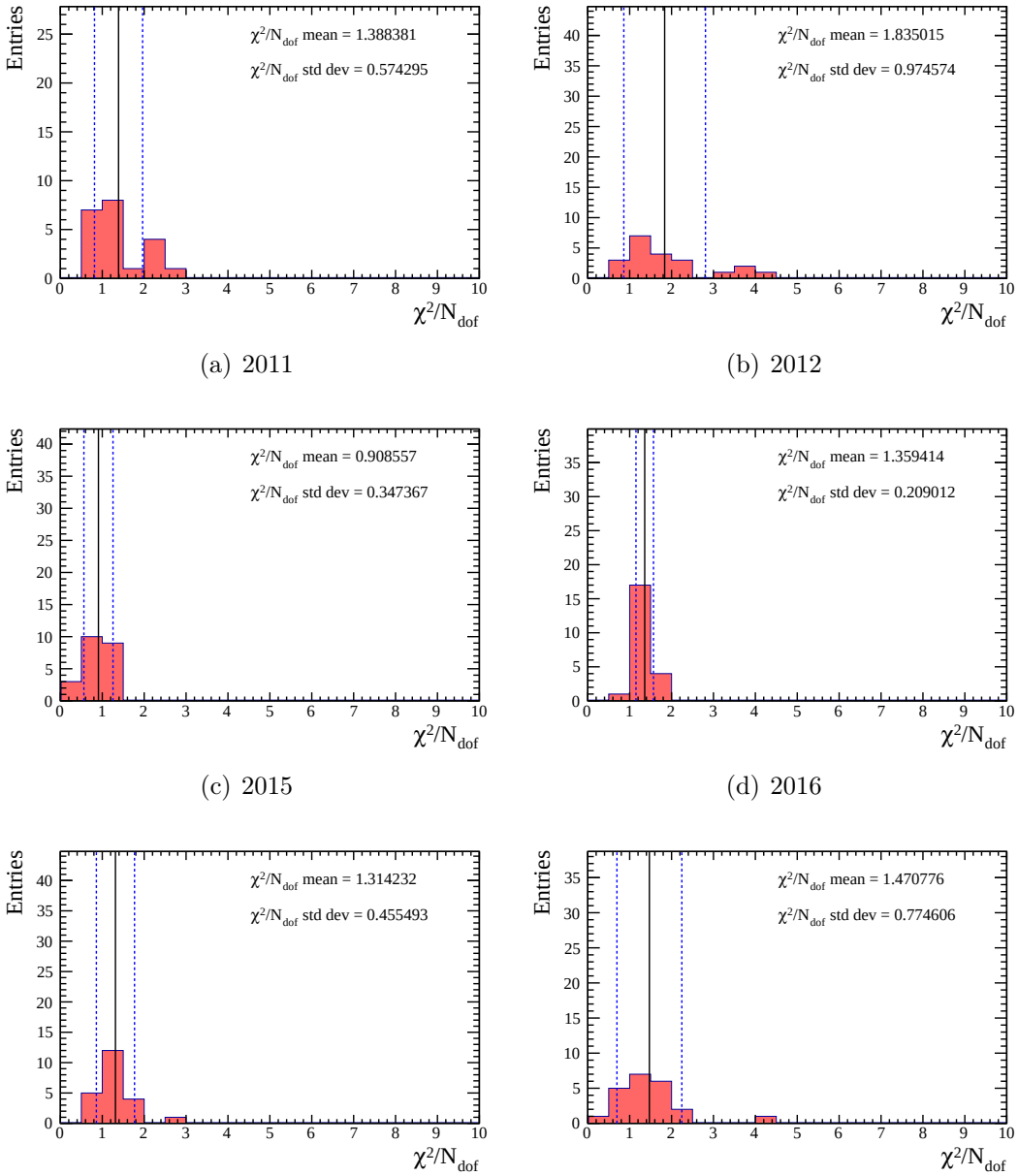
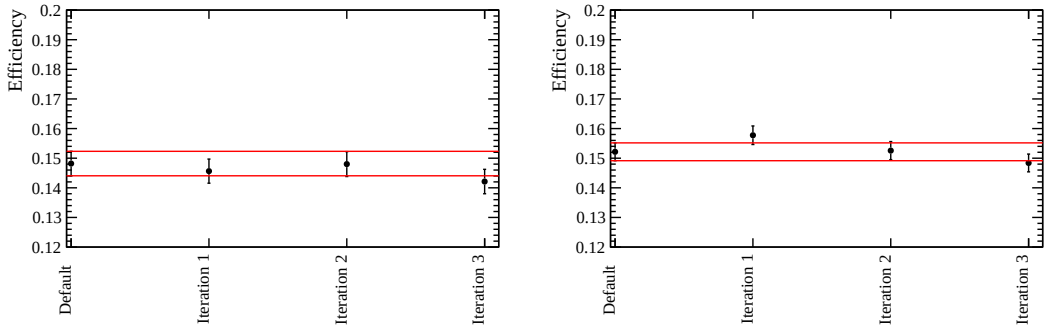
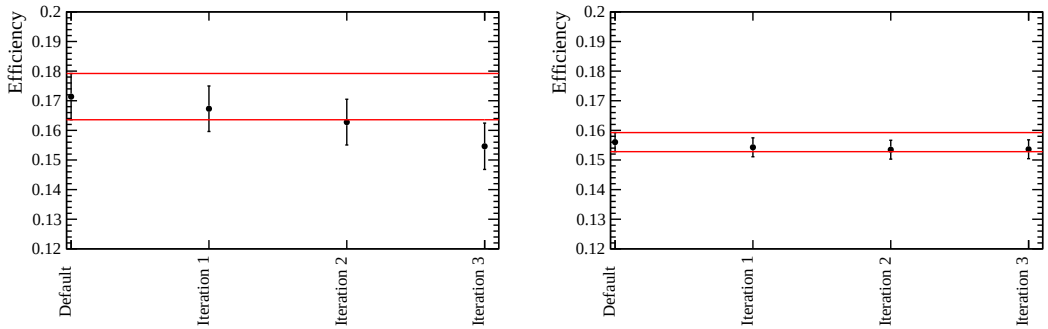


Figure 7.11: χ^2/N_{dof} distributions of the variables used for the $3\pi\mu$ final state after the third iteration. The mean and standard deviation of the distributions are reported and visualized with a black solid line and two blue dashed lines (referring respectively to the mean and $\pm 1\sigma$ interval). The rightmost variables in the 2017 and 2018 datasets have been already corrected in the first iteration and their relatively high χ^2/N_{dof} are due to differences between data and MC in low-statistics bins in the tail of the distributions that can not be further corrected. Therefore they are ignored and the variables with the second highest χ^2/N_{dof} are considered.



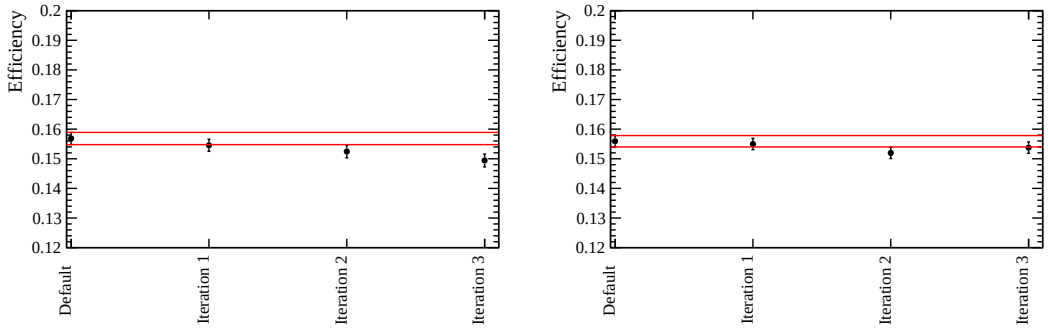
(a) 2011

(b) 2012



(c) 2015

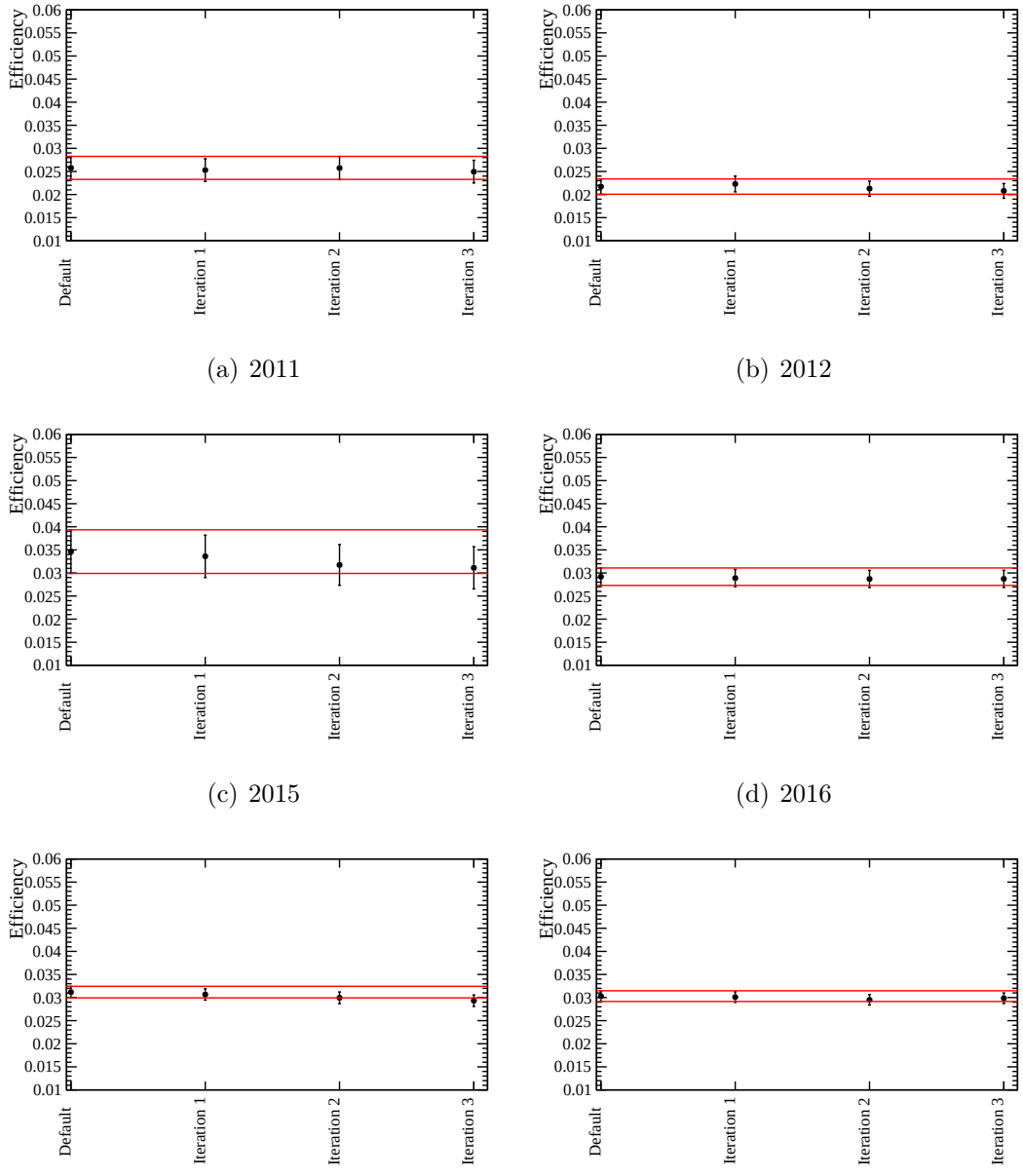
(d) 2016



(e) 2017

(f) 2018

Figure 7.12: Evolution of the selection efficiency for the fully-matched candidates of the $3\pi\mu$ final state as a function of the correction iterations. The red lines correspond to the $\pm 1\sigma$ interval around the central value of the default efficiency.

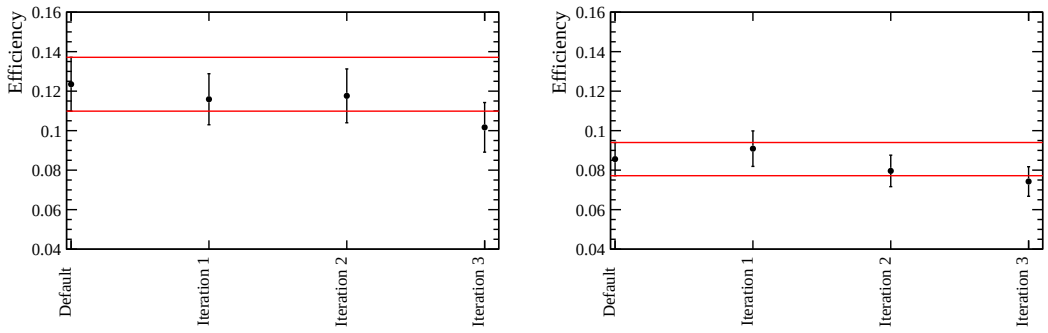


(a) 2011 (b) 2012

(c) 2015 (d) 2016

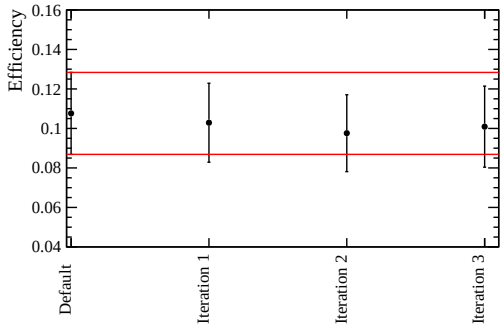
(e) 2017 (f) 2018

Figure 7.13: Evolution of the selection efficiency for the swapped-pion candidates of thes $3\pi\mu$ final state as a function of the correction iterations. The red lines correspond to the $\pm 1\sigma$ interval around the central value of the default efficiency.

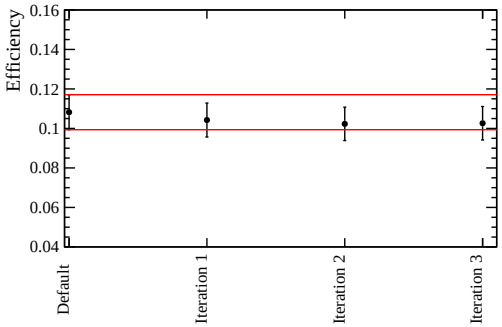


(a) 2011

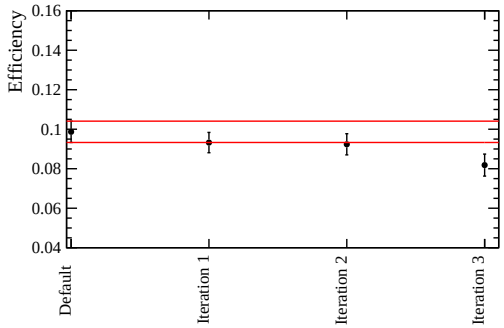
(b) 2012



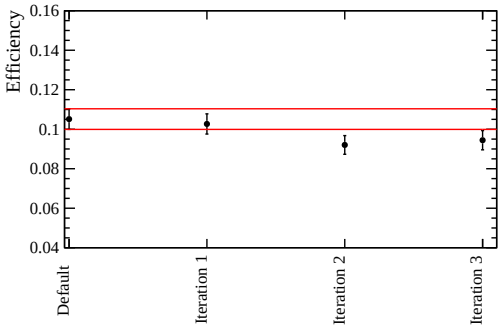
(c) 2015



(d) 2016



(e) 2017



(f) 2018

Figure 7.14: Evolution of the selection efficiency for the "ghosts" candidates of the $3\pi\mu$ final state as a function of the correction iterations. The red lines correspond to the $\pm 1\sigma$ interval around the central value of the default efficiency.

Chapter 8

Signal fit

After the full selection one candidate is randomly chosen between those belonging to the same event (therefore the terms "event" and "candidate" assume the same meaning in the context of the likelihood fit).
A binned maximum likelihood fit is then performed to the output of the fitBDT, in order to determine the number of signal decays in the data. After the training of the BDT, the signal template is extracted from MC events in the K^{*0} signal region, while the background template is extracted from data in the control region. In order for the latter to correctly describe the background distribution in the signal region, the fitBDT must be uncorrelated with the K^{*0} mass.
The HistFactory package [97] in RooFit [98], which implements various utilities to handle binned PDFs in a simple way, is used to perform the binned fit. The likelihood minimization and the limit calculation, using the CL_s method [99], is done using the RooStats package [100].

8.1 Fit BDT

The fitBDT is trained on fully-matched signal MC events (from the samples without the neutral pions component) and data from the background region, after applying the full selection on the training samples.

$3\pi3\pi$ final state

The variables used for the training of the fitBDT are:

- BDT1;
- max(longitudinal distance between τ^+ decay vertex and PV, longitudinal distance between τ^- decay vertex and PV);
- B^0 flight distance χ^2;
- B^0 flight distance projection along z;
- min(τ^+ analytically reconstructed momentum from optimal solution, τ^- analytically reconstructed momentum from optimal solution);

- min(τ^+ analytically reconstructed momentum from "+" solution, τ^- analytically reconstructed momentum from "+" solution);
- min(τ^+ visible mass, τ^- visible mass);
- min(τ^+ rhoMass, τ^- rhoMass);
- min(τ^+ resMass, τ^- resMass);
- min(τ^+ flight distance, τ^- flight distance);
- max(τ^+ flight distance, τ^- flight distance);
- B^0 IP with respect to PV;
- min(τ^+ neutral cone isolation sPT, τ^- neutral cone isolation sPT)[1];
- min(π_1^+ (from τ^+) IsoBDTThirdValue, π_1^- (from τ^-) IsoBDTThirdValue);
- max(π_1^+ (from τ^+) IsoBDTSecondValue, π_1^- (from τ^-) IsoBDTSecondValue);
- max(τ^+ IsoBDTSecondValue, τ^- IsoBDTSecondValue).

The specific rankings are reported in Appendix B, together with the variable distributions. The flattened BDT distributions for signal and background samples are reported in Fig. 8.1. Table 8.1 shows the size of the training samples used.

Table 8.1: Number of candidates used for the training of the fitBDT for the $3\pi3\pi$ final state.

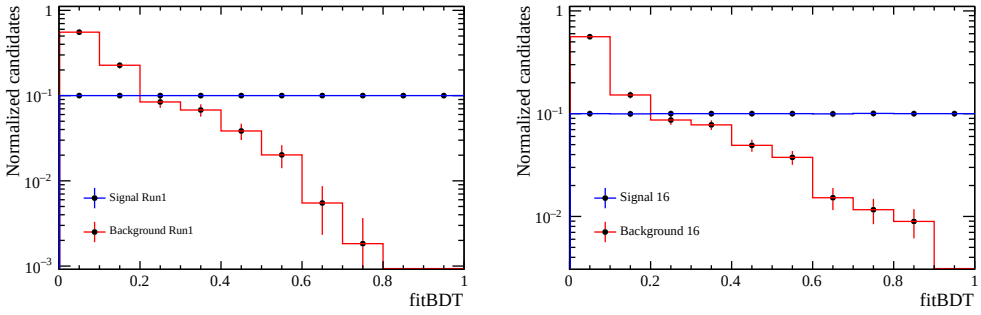
Year	Signal candidates	Background candidates
Run1	1965	529
2016	1980	1102
2017	4449	1237
2018	4449	1383

$3\pi\mu$ final state

The variables used for the training of the fit BDT are:

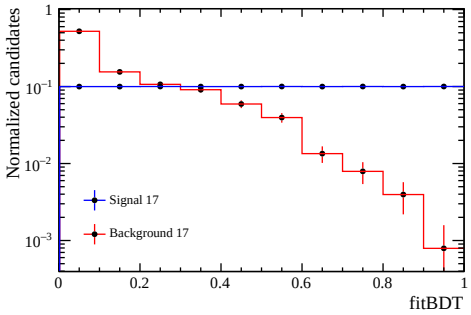
- τ^+ IsoBDTSecondValue;
- μ associated PV χ^2;
- B^0 corrected mass;

[1]This variable is not available in Run 1

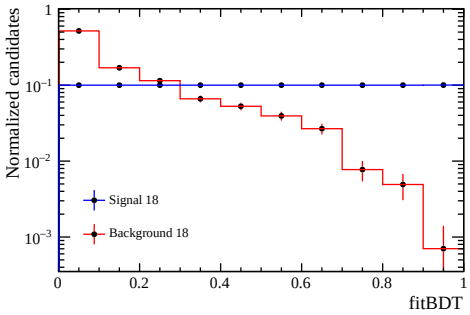


(a) Run 1

(b) 2016



(c) 2017



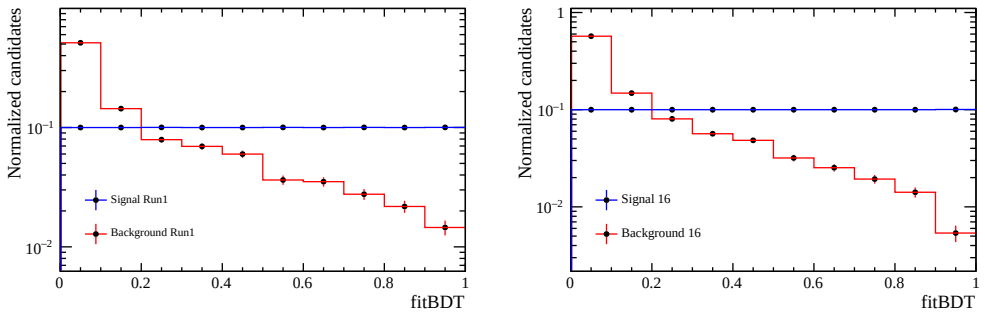
(d) 2018

Figure 8.1: Distributions of fitBDT for $3\pi 3\pi$ signal and abckground training samples. The flat transformation is applied.

- τ^+ maximum DOCA of decay products;
- τ^+ $m_{\pi_1^+\pi_2^-}$;
- B^0 decay vertex χ^2;
- B^0 neutral cone isolation $\mathrm{P}_{\mathrm{asym}}$[2];
- K^{*0} DIRA with respect to PV;
- τ^+ minimum IP χ^2 with respect to all PVs in the event;
- τ^+ and K^{*0} DOCA;
- τ^+ DIRA with respect to B^0 decay vertex.

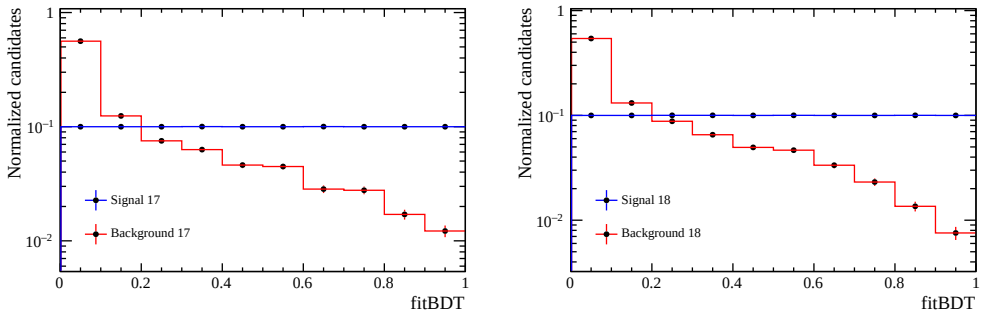
The specific rankings are reported in Appendix B, together with the variable distributions. The flattened BDT distributions for signal and background samples are reported in Fig. 8.2. Table 8.2 shows the size of the training samples used.

[2]This variable is not available in Run 1.



(a) Run 1

(b) 2016



(c) 2017

(d) 2018

Figure 8.2: Distributions of fitBDT for $3\pi\mu$ signal and abckground training samples. The flat transformation is applied.

8.2 Fit model

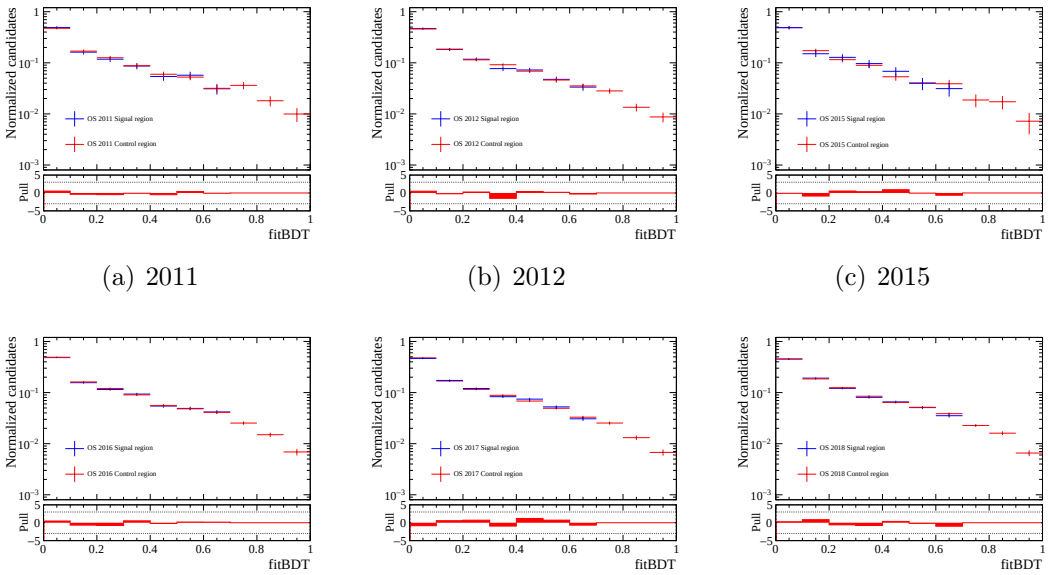
The fit is performed simultaneously over the six data-taking years and the two final states, for a total of 12 samples. The value of the signal branching ratio is shared between the samples. For the sake of completeness, also the results for the separate fits over the two final states are shown. For each sample, the template consists of a histogram of ten equal size bins in the range $[0, 1]$.
The signal template is extracted from the full MC samples including the neutral pion component. Because of the presence of the neutral pion component and of pion in-flight decays, the fitBDT distribution for the signal is not totally flat, with a higher concentration of events at low BDT values. A small contamination from badly reconstructed events is present, but its effect on the signal shape is negligible (only $\sim$1% of MC events are neither truth-matched or swapped pions after the full selection).
The background template is extracted from data in the K^{*0} mass control region. In order to check that the template extracted from the control region is representative of the background in the signal region, the fitBDT distribution is compared for events in the two regions. However, in order to blind the most sensitive bins, the

Table 8.2: Number of candidates used for the training of the fitBDT for the $3\pi\mu$ final state.

Year	Signal candidates	Background candidates
Run1	3921	3430
2016	1907	5009
2017	4051	5977
2018	4711	6476

cut fitBDT < 0.7 is applied on events in the signal region of the OS data.

The comparison of the $3\pi3\pi$ and $3\pi\mu$ fitBDT distributions in signal and control region is shown in Figs. 8.3 and 8.4 respectively. The comparison shows a good

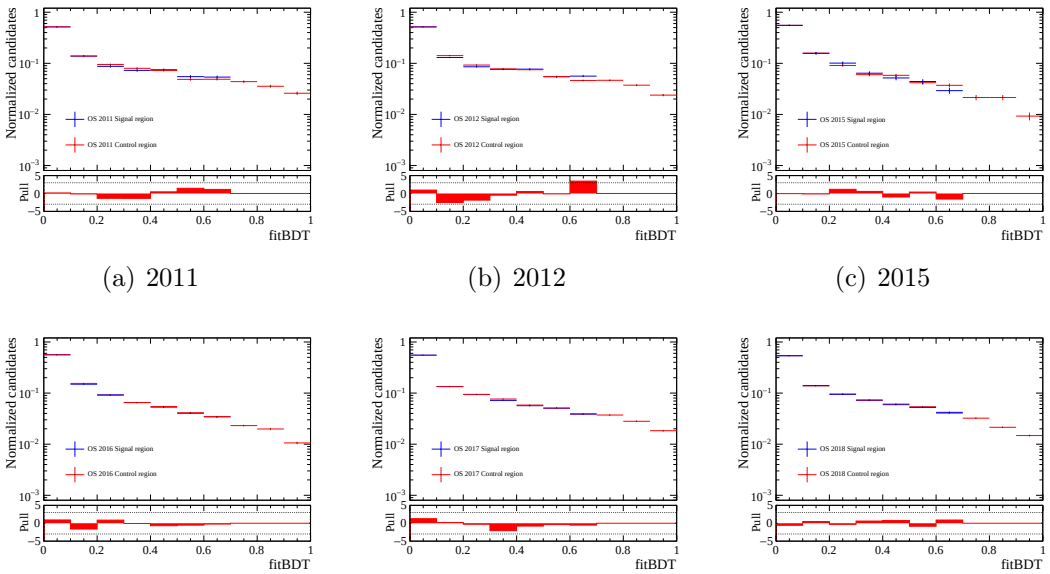


(a) 2011 (b) 2012 (c) 2015

(d) 2016 (e) 2017 (f) 2018

Figure 8.3: Comparison of the $3\pi3\pi$ fitBDT distribution in signal and control regions in logarithmic scale. The distribution in the signal region is shown for values of fitBDT < 0.7. The dashed lines in the pull correspond to the $\pm3\sigma$ difference.

agreement between the two regions, which validates the use of the control region as proxy for background distribution in the signal region.



(a) 2011 (b) 2012 (c) 2015

(d) 2016 (e) 2017 (f) 2018

Figure 8.4: Comparison of the $3\pi\mu$ fitBDT distribution in signal and control regions in logarithmic scale. The distribution in the signal region is shown for values of fitBDT < 0.7. The dashed lines in the pull correspond to the $\pm 3\sigma$ difference.

In order to derive an analytical expression for the fit model, let us consider two subsamples of a dataset, labelled 1 and 2, each containing a certain number of events $n_{1(2)} = s_{1(2)} + b_{1(2)}$, where s stands for signal and b stands for background. The distributions of events in the two regions can be written as:

$$\begin{aligned} \mathrm{H}_1 &= n_1 \cdot \mathcal{H}_1 = s_1 \cdot \mathcal{S}_1 + b_1 \cdot \mathcal{B}, \\ \mathrm{H}_2 &= n_2 \cdot \mathcal{H}_2 = s_2 \cdot \mathcal{S}_2 + b_2 \cdot \mathcal{B}, \end{aligned} \tag{8.1}$$

where $\mathcal{S}$, $\mathcal{B}$ and $\mathcal{H}$ represent respectively the signal, background and total PDFs (*i.e.* normalized to unity). It has to be noticed that *by hypothesis* the background distribution $\mathcal{B}$ is identical for the two subsamples. The latter can be written as:

$$\mathcal{B} = \frac{1}{b_2}(n_2 \cdot \mathcal{H}_2 - s_2 \cdot \mathcal{S}_2). \tag{8.2}$$

Using the relation $s_2/\epsilon_2 = s_1/\epsilon_1$, where ϵ indicates the signal efficiency in a given subsample, the distribution of events in the subsample 1 can be written as:

$$\mathrm{H}_1 = s_1 \cdot \mathcal{S}_1 + \frac{b_1}{b_2} n_2 \cdot \mathcal{H}_2 - \frac{b_1}{b_2} s_1 \frac{\epsilon_2}{\epsilon_1} \cdot \mathcal{S}_2. \tag{8.3}$$

By interpreting the subsamples 1 and 2 as the signal and control regions respectively,

the fit model assumes the form:

$$\mathbf{Data}_{\boldsymbol{y}} = f_y\mu \cdot \mathbf{Sig}_{\boldsymbol{y}} + \frac{s_y}{c_y} n_y^{\text{ctl}} \cdot \mathbf{Ctl}_{\boldsymbol{y}} - \frac{s_y}{c_y} f_y\mu \frac{\epsilon_y^{\text{ctl}}}{\epsilon_y^{\text{sig}}} \cdot \mathbf{Cont}_{\boldsymbol{y}}, \tag{8.4}$$

where the subscript y refers to one specific year of data-taking of a given final state and for the sake of clarity normalized PDFs are highlighted in blue while fitted parameters are highlighted in red. The various terms are:

$\mathbf{Data}_{\boldsymbol{y}}$: histogram representing the total number of signal and background events;

$\mathbf{Sig}_{\boldsymbol{y}}$: fitBDT PDF extracted from events in the signal region of the full MC;

$\mathbf{Ctl}_{\boldsymbol{y}}$: fitBDT PDF extracted from events in the control region of data;

$\mathbf{Cont}_{\boldsymbol{y}}$: fitBDT PDF extracted from events in the control region of the full MC to take into account the "contamination" of signal in the background template;

$\boldsymbol{\mu}$: $B^0 \to K^{*0}\tau^+\tau^-$ branching ratio, the parameter of interest of the fit which is shared between the years of data-taking and the two final states;

$\frac{s_y}{c_y}$: ratio of the (unknown) number of background events in signal and control region, a scaling factor treated as a nuisance parameter of the fit;

$\boldsymbol{f_y}$: defined as $f_y \equiv \frac{1}{\alpha_y}$, where α_y is the normalization factor introduced in Chapter 6. The uncertainty on these parameters is taken into account with a gaussian constraint. Its values are shown in Tables 8.3 and 8.4;

$\frac{\epsilon_y^{\text{ctl}}}{\epsilon_y^{\text{sig}}}$: ratio of efficiencies in control and signal region, whose uncertainty is taken into account with a gaussian constraint. Its values are shown in Tables 8.3 and 8.4;

$\boldsymbol{n_y^{\text{ctl}}}$: number of events in the data control region, a fixed parameter. Its values are shown in Tables 8.3 and 8.4.

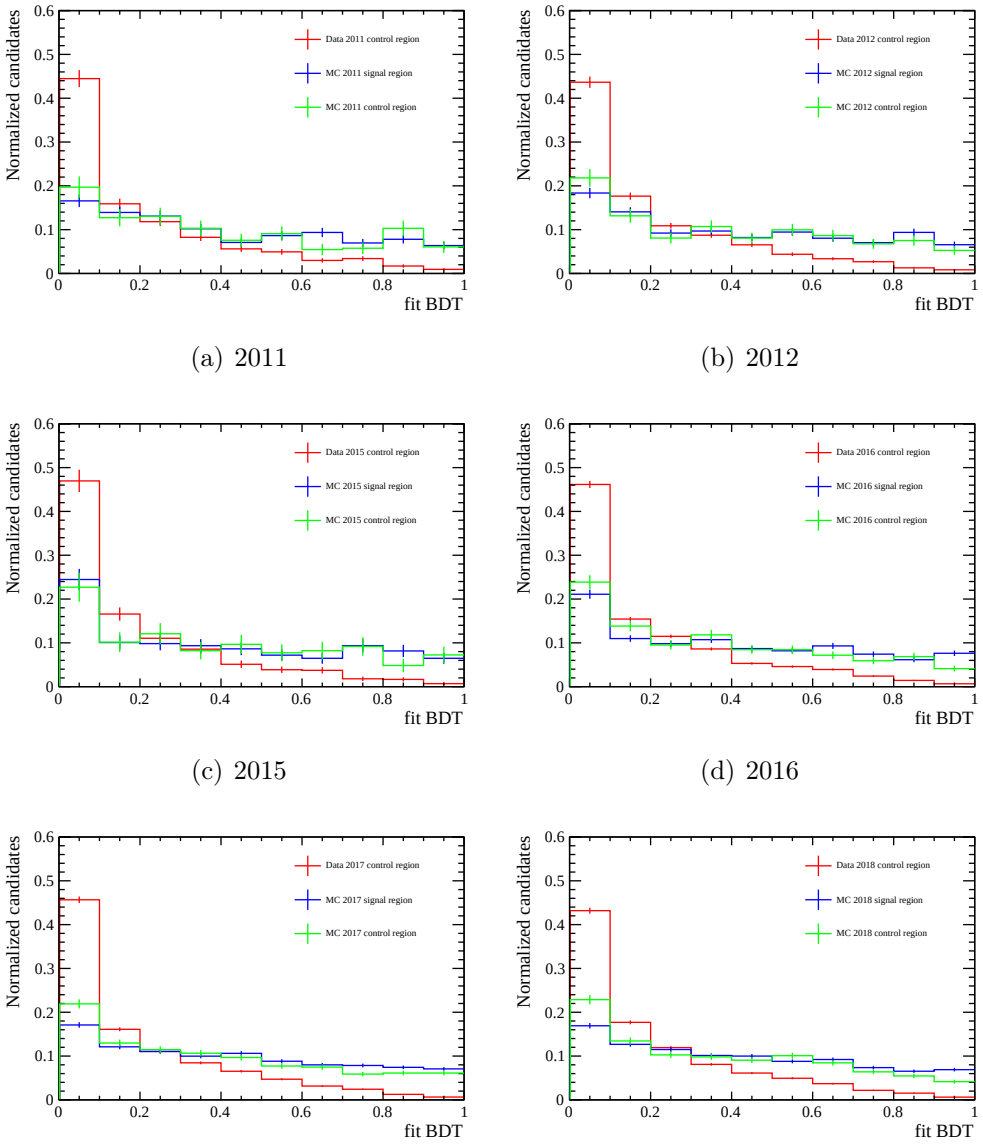
The distributions of the fitBDT for data in the control region ($\mathbf{Ctl}_{\boldsymbol{y}}$), MC in the signal region ($\mathbf{Sig}_{\boldsymbol{y}}$) and MC in the control region ($\mathbf{Cont}_{\boldsymbol{y}}$) are shown in Figs. 8.5 and 8.6 for the $3\pi3\pi$ and $3\pi\mu$ final states respectively.

Table 8.3: Values of the input parameters for the likelihood fit for the $3\pi3\pi$ final state. Shown are the normalization factor α_y, its inverse f_y, the ratio of efficiencies in control and signal region $\epsilon_y^{\text{ctl}}/\epsilon_y^{\text{sig}}$ and the number of events in data control region. Moreover, reported are also the number of events in the data signal region with fitBDT < 0.7 n_y^{blind} and the number of events in the data control region with fitBDT < 0.7 $n_y^{\text{ctl, blind}}$.

Year	α_y (10^{-5})	f_y	$\epsilon_y^{\text{ctl}}/\epsilon_y^{\text{sig}}$	n_y^{ctl}	n_y^{blind}	$n_y^{\text{ctl, blind}}$
2011	7.10 ± 0.89	14083 ± 1761	0.396 ± 0.026	1176	612	1105
2012	3.59 ± 0.44	27865 ± 3386	0.416 ± 0.021	2646	1285	2519
2015	12.4 ± 1.6	8094 ± 1060	0.496 ± 0.042	724	352	694
2016	1.48 ± 0.18	67517 ± 8094	0.481 ± 0.019	6992	3594	6677
2017	1.31 ± 0.16	76550 ± 9084	0.487 ± 0.013	8209	4043	7854
2018	1.48 ± 0.17	67796 ± 8042	0.478 ± 0.013	9246	4564	8845

Table 8.4: Values of the input parameters for the likelihood fit for the $3\pi\mu$ final state. Shown are the normalization factor α_y, its inverse f_y, the ratio of efficiencies in control and signal region $\epsilon_y^{\text{ctl}}/\epsilon_y^{\text{sig}}$ and the number of events in data control region. Moreover, reported are also the number of events in the data signal region with fitBDT < 0.7 n_y^{blind} and the number of events in the data control region with fitBDT < 0.7 $n_y^{\text{ctl, blind}}$.

Year	α_y (10^{-6})	f_y	$\epsilon_y^{\text{ctl}}/\epsilon_y^{\text{sig}}$	n_y^{ctl}	n_y^{blind}	$n_y^{\text{ctl, blind}}$
2011	5.95 ± 0.73	168160 ± 20781	0.398 ± 0.025	8043	4335	7276
2012	2.60 ± 0.31	385336 ± 46000	0.370 ± 0.017	18749	10022	16925
2015	13.6 ± 1.8	73568 ± 9706	0.435 ± 0.044	3630	1951	3450
2016	1.52 ± 0.18	658345 ± 78319	0.410 ± 0.019	38607	21008	36645
2017	1.31 ± 0.15	762637 ± 89401	0.394 ± 0.012	44797	23304	41340
2018	1.41 ± 0.16	711235 ± 83252	0.405 ± 0.011	47702	24683	44645

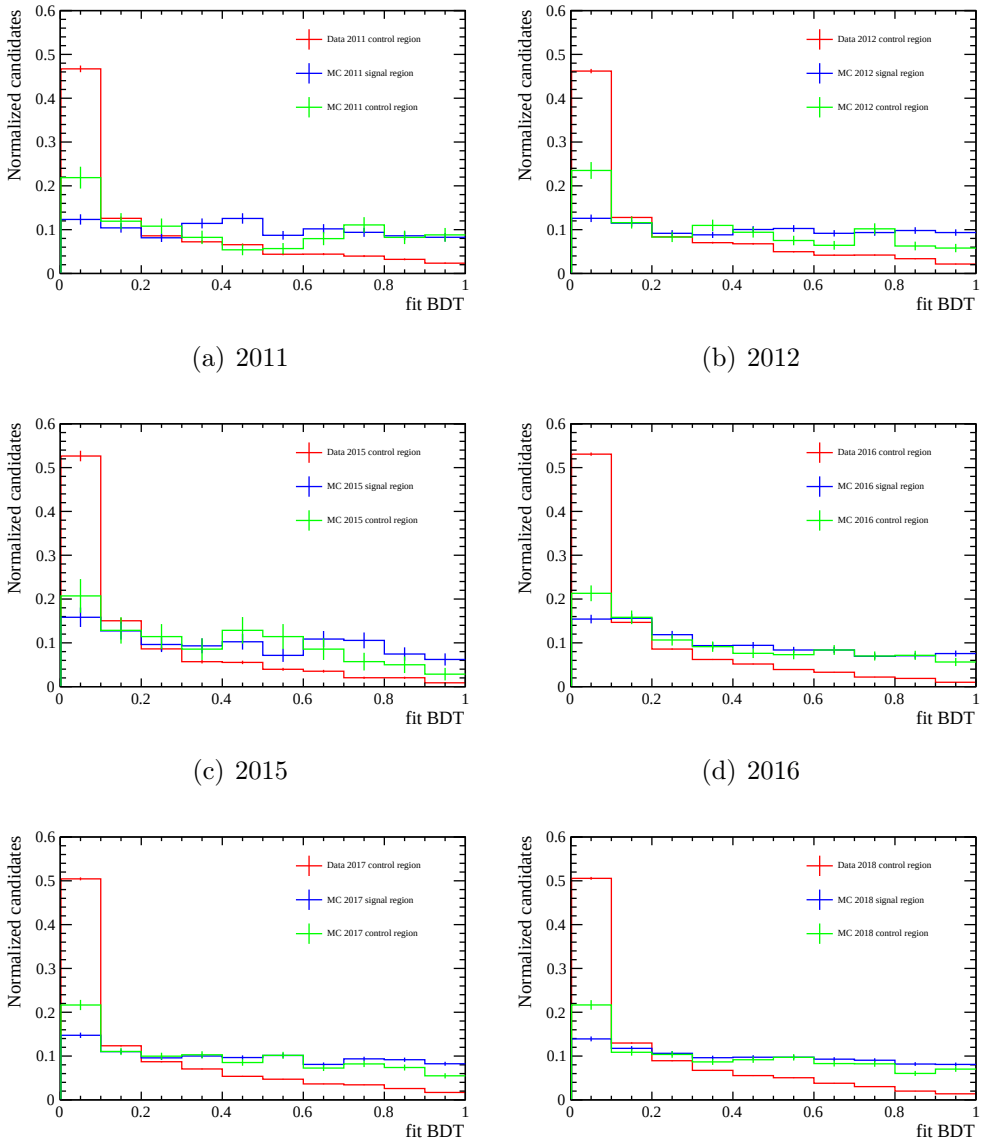


(a) 2011 (b) 2012

(c) 2015 (d) 2016

(e) 2017 (f) 2018

Figure 8.5: Distribution of the fitBDT in the $3\pi 3\pi$ channel for data in the control region (red), MC in the signal region (blue) and MC in the control region (green) for each year of data-taking.



(a) 2011 (b) 2012

(c) 2015 (d) 2016

(e) 2017 (f) 2018

Figure 8.6: Distribution of the fitBDT in the $3\pi\mu$ channel for data in the control region (red), MC in the signal region (blue) and MC in the control region (green) for each year of data-taking.

8.3 Toy studies

The fit strategy is validated using toy simulations. In order to generate pseudo-data, it is necessary to know the distribution of the fitBDT for signal events in the signal region ($\mathbf{Sig}_y$) and, ideally, the "true" (*i.e.* signal-free) distribution of background events in the signal region. Of course, the latter is unknown and an approximated distribution must be used. The best possible assumption is to use the fitBDT distribution of data in the control region ($\mathbf{Ctl}_y$), in the hypothesis that the effect of the possible presence of signal events on the distribution is negligible, since the signal-to-background ratio is much lower than in the signal region. Therefore, the model used to generate the pseudo-data for a given final state and data-taking year y is defined as:

$$\mathbf{GenData}_y = f_y\mu_0 \cdot \mathbf{Sig}_y + \left(n_y^{\mathrm{OS}} - f_y\mu_0\right) \cdot \mathbf{Ctl}_y, \tag{8.5}$$

where μ_0 is the signal branching ratio defined by the analyst and n_y^{OS} is the number of events in the signal region of the OS data. However, the value of n_y^{OS} is unknown due to the blinding procedure, and is estimated as

$$n_y^{\mathrm{OS}} \approx n_y^{\mathrm{blind}} \cdot \frac{n_y^{\mathrm{ctl}}}{n_y^{\mathrm{ctl,\ blind}}}, \tag{8.6}$$

where n_y^{blind} is the number of events in the fitBDT range $[0, 0.7]$ in the signal region of the OS data, n_y^{ctl} is the number of events in the full fitBDT range in the OS control region and $n_y^{\mathrm{ctl,\ blind}}$ is the number of events in the fitBDT range $[0, 0.7]$ in the OS control region. The values of n_y^{ctl}, n_y^{blind} and $n_y^{\mathrm{ctl,\ blind}}$ are shown in Tables 8.3 and 8.4.

The signal ($\mathbf{Sig}_y$) and background ($\mathbf{Ctl}_y$) PDFs are extended PDFs, therefore the number of generated signal and background events vary for each toy and fluctuate following a poissonian distribution around the central values $f_y\mu_0$ and $n_y^{\mathrm{OS}} - f_y\mu_0$. The central value for the number of generated background events is chosen in such a way to keep the (central) total amount of generated events equal to the number of observed events in the signal region n_y^{OS}.

In order to take into account the presence of signal events in the model used to describe the background, the toys are fitted with the model

$$\mathbf{PseudoData}_y = f_y\mu \cdot \mathbf{Sig}_y + \frac{s_y}{c_y} n_y^{\mathrm{ctl}} \cdot \mathbf{Ctl}'_y - f_y\mu \frac{s_y}{c_y} \frac{\epsilon_y^{\mathrm{ctl}}}{\epsilon_y^{\mathrm{sig}}} \cdot \mathbf{Cont}_y, \tag{8.7}$$

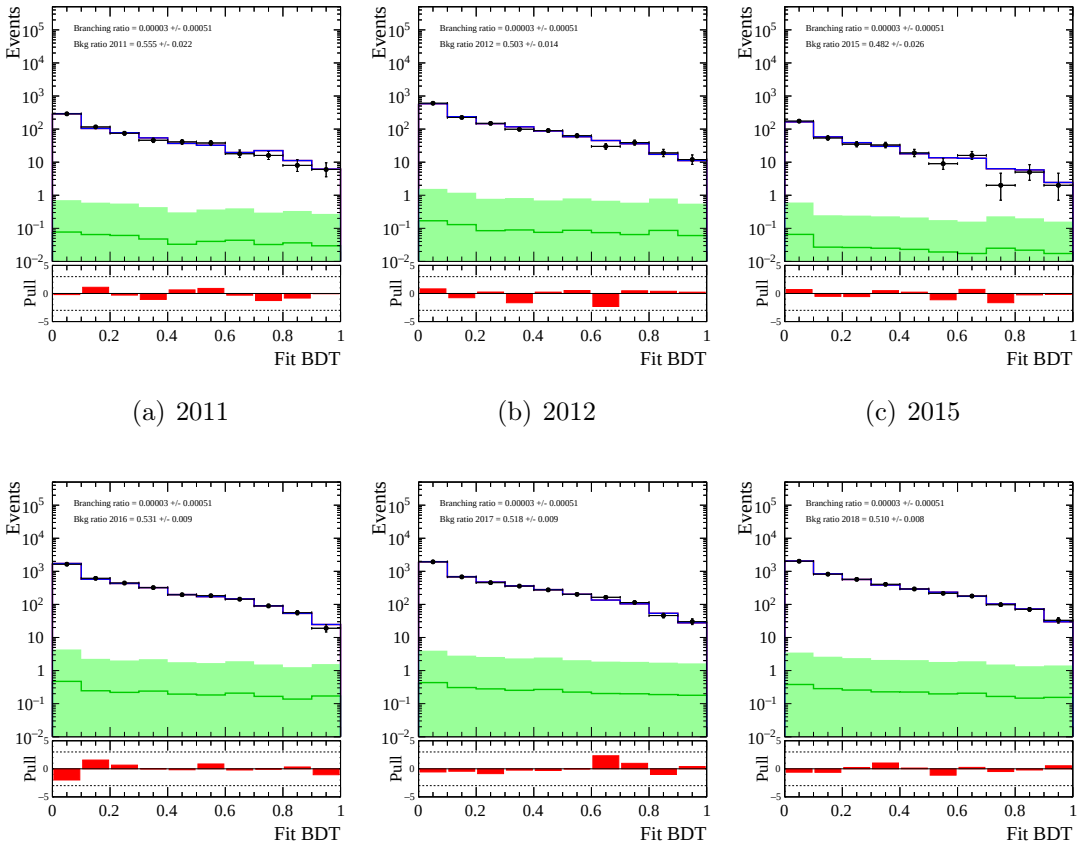
where

$$\mathbf{Ctl}'_y = \frac{1}{n_y^{\mathrm{ctl}}} \left[\left(n_y^{\mathrm{ctl}} - f_y\mu_0 \frac{\epsilon_y^{\mathrm{ctl}}}{\epsilon_y^{\mathrm{sig}}} \right) \cdot \mathbf{Ctl}_y + f_y\mu_0 \frac{\epsilon_y^{\mathrm{ctl}}}{\epsilon_y^{\mathrm{sig}}} \cdot \mathbf{Cont}_y \right]. \tag{8.8}$$

The validation of the fit model is based on 1000 toys, and only the statistical uncertainty is considered. In the following, toy studies are shown for the $3\pi3\pi$ and $3\pi\mu$ final states separately and for the simultaneous fit over the two final states.

$3\pi3\pi$ final state

As an example, the results of the likelihood fit on a single toy generated with initial branching ratio set to 0 are shown in Fig. 8.7.



(a) 2011 (b) 2012 (c) 2015

(d) 2016 (e) 2017 (f) 2018

Figure 8.7: Result of the likelihood fit on a single toy experiment generated with initial signal branching ratio set to 0 for the $3\pi3\pi$ final state. Shown are the generated data (black points), the signal (green), the background (red) and the total PDF (blue). The dashed lines in the pull correspond to the $\pm3\sigma$ difference.

The distributions of the fit results, the errors and the pulls are shown in Fig. 8.8 for the signal branching ratio, while the pull distributions are shown in Fig. 8.9 for the nuisance parameters ($\frac{s_y}{c_y}$), in the case of initial branching ratio $\mu_0 = 0$. Table

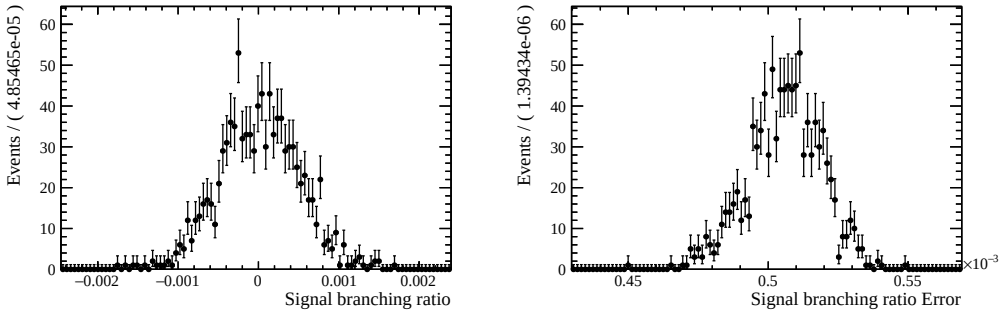


Figure 8.8: signal branching ratio fit results for toys generated with $\mu_0 = 0$ for the $3\pi 3\pi$ final state. The fit with a standard gaussian is shown in blue.

8.5 summarizes the results.

The distributions of the fit results, the errors and the pulls are shown in Fig. 8.10 for the signal branching ratio, while the pull distributions are shown in Fig. 8.11 for the nuisance parameters $(\frac{s_y}{c_y})$, in the case of initial branching ratio $\mu_0 = 0.0005$.

Table 8.6 summarizes the results.

Table 8.5: Results of the fit validation study for the $3\pi3\pi$ final state, based on 1000 toys generated with signal branching ratio $\mu_0 = 0$. Shown are the initial values, the average outcome of the fit, the pull mean and the pull variance of the signal branching ratio μ and nuisance parameters $\frac{s_y}{c_y}$ (indicated as p_y in the table).

Parameter	Initial value	Average fit value	Pull mean	Pull Std
μ	0.0	0.00001 ± 0.00051	0.013 ± 0.031	0.985 ± 0.022
$p_{11;3\pi3\pi}$	0.554	0.554 ± 0.022	-0.028 ± 0.032	1.005 ± 0.022
$p_{12;3\pi3\pi}$	0.510	0.509 ± 0.014	-0.054 ± 0.032	0.997 ± 0.022
$p_{15;3\pi3\pi}$	0.507	0.508 ± 0.027	-0.009 ± 0.032	1.012 ± 0.023
$p_{16;3\pi3\pi}$	0.5383	0.5382 ± 0.0095	-0.008 ± 0.031	0.989 ± 0.022
$p_{17;3\pi3\pi}$	0.5148	0.5147 ± 0.0087	-0.015 ± 0.032	1.017 ± 0.023
$p_{18;3\pi3\pi}$	0.5160	0.5150 ± 0.0080	-0.068 ± 0.032	1.004 ± 0.022

Table 8.6: Results of the fit validation study for the $3\pi3\pi$ final state, based on 1000 toys generated with signal branching ratio $\mu_0 = 0.0005$. Shown are the initial values, the average outcome of the fit, the pull mean and the pull variance of the signal branching ratio μ and nuisance parameters $\frac{s_y}{c_y}$ (indicated as p_y in the table).

Parameter	Initial value	Average fit value	Pull mean	Pull Std
μ	0.0005	0.00051 ± 0.00052	0.008 ± 0.031	0.976 ± 0.022
$p_{11;3\pi3\pi}$	0.549	0.549 ± 0.022	-0.008 ± 0.032	1.006 ± 0.022
$p_{12;3\pi3\pi}$	0.506	0.505 ± 0.015	-0.041 ± 0.032	1.000 ± 0.022
$p_{15;3\pi3\pi}$	0.503	0.504 ± 0.027	0.009 ± 0.032	1.000 ± 0.022
$p_{16;3\pi3\pi}$	0.5347	0.5349 ± 0.0095	0.02 ± 0.031	0.976 ± 0.022
$p_{17;3\pi3\pi}$	0.5113	0.5114 ± 0.0087	0.014 ± 0.032	1.001 ± 0.022
$p_{18;3\pi3\pi}$	0.5130	0.5130 ± 0.0080	-0.036 ± 0.032	1.001 ± 0.022

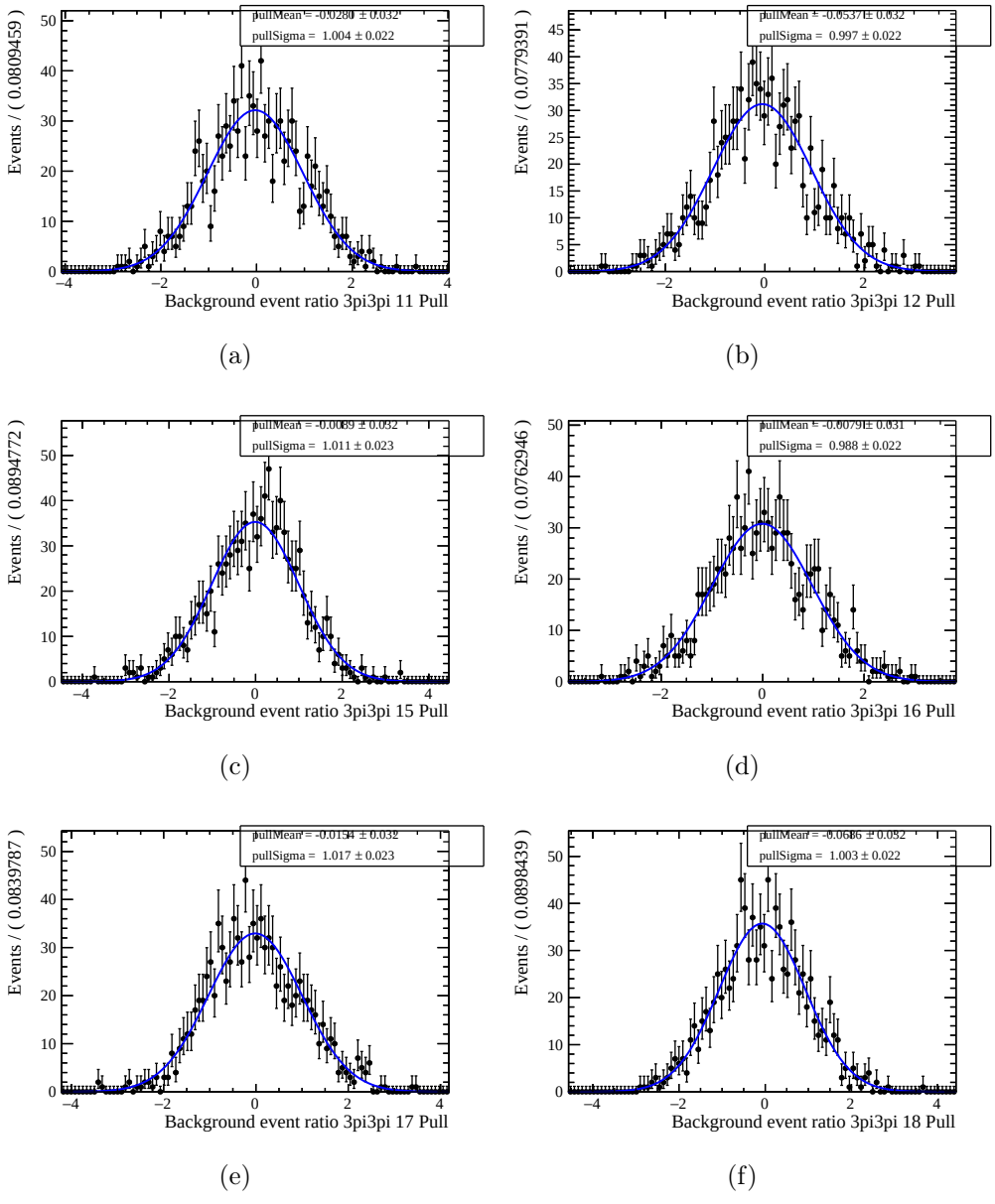


Figure 8.9: Nuisance parameters pulls for toys generated with $\mu_0 = 0$ for the $3\pi3\pi$ final state. The fit with a standard gaussian is shown in blue.

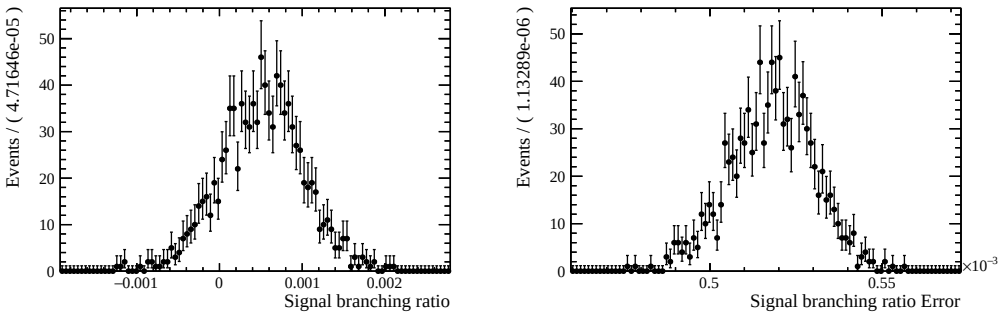


Figure 8.10: signal branching ratio fit results for toys generated with $\mu_0 = 0.0005$ for the $3\pi3\pi$ final state. The fit with a standard gaussian is shown in blue.

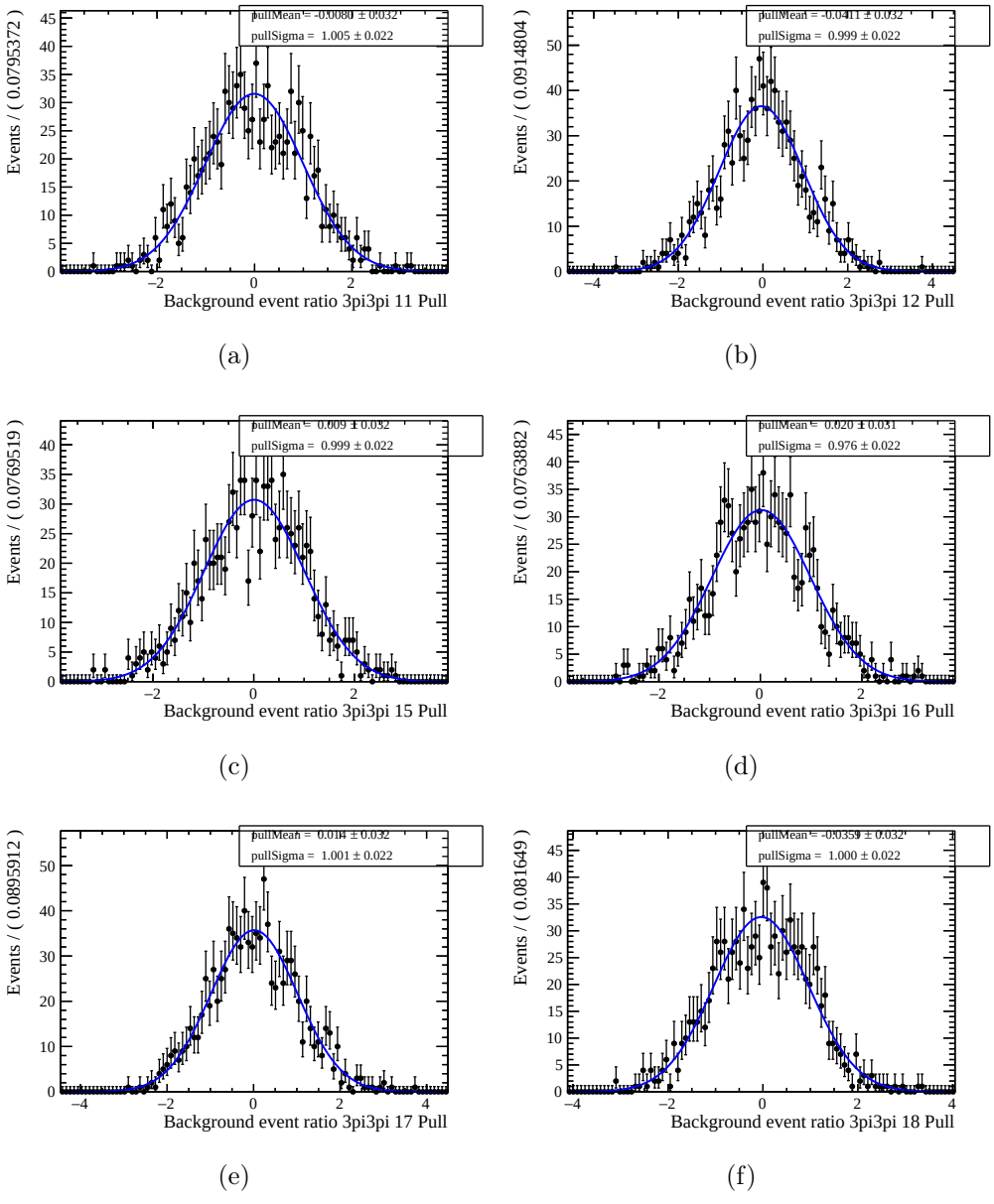
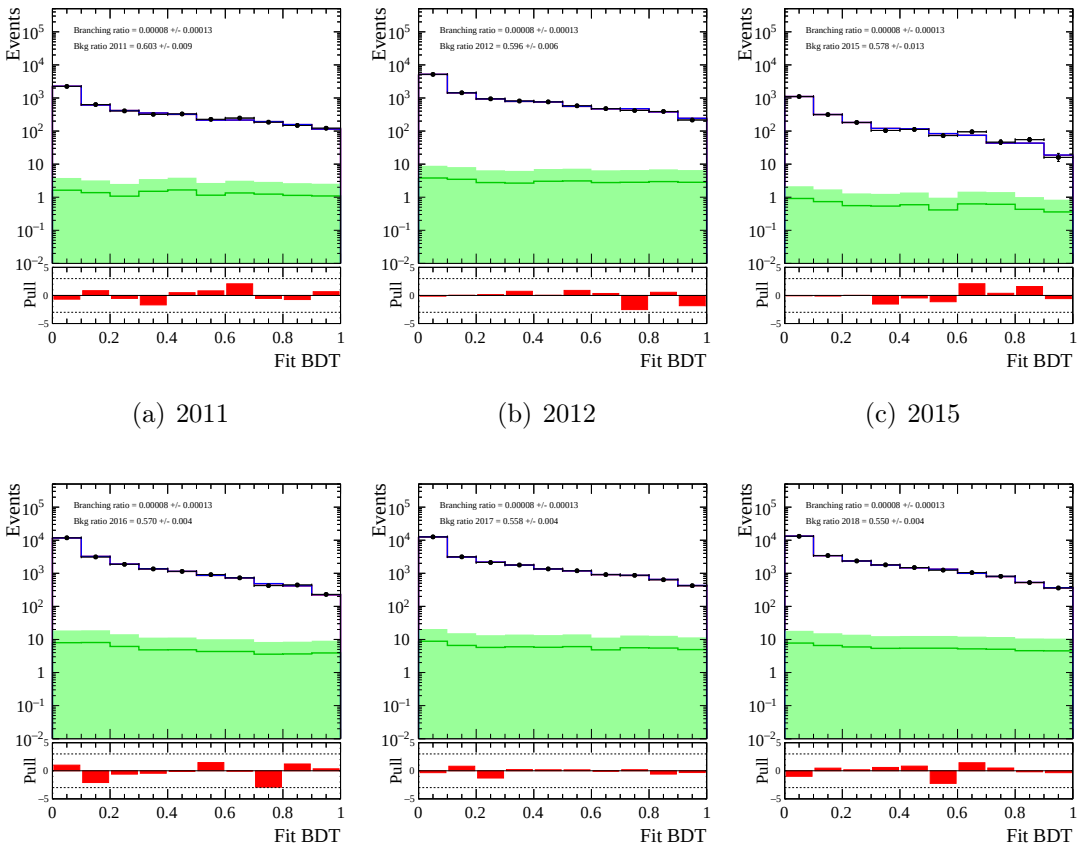


Figure 8.11: Nuisance parameters pulls for toys generated with $\mu_0 = 0.0005$ for the $3\pi3\pi$ final state. The fit with a standard gaussian is shown in blue.

$3\pi\mu$ final state

As an example, the results of the likelihood fit on a single toy generated with initial branching ratio set to 0 are shown in Fig. 8.12.



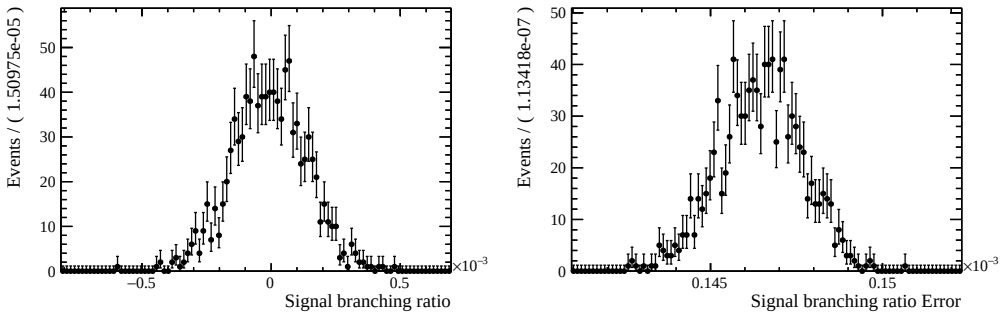
(a) 2011 (b) 2012 (c) 2015

(d) 2016 (e) 2017 (f) 2018

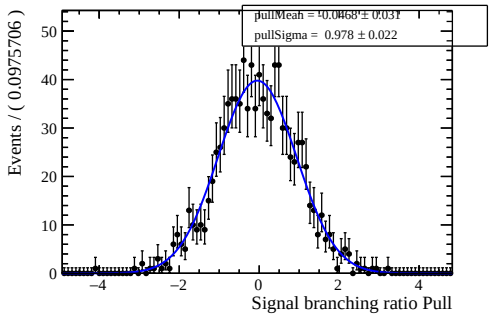
Figure 8.12: Result of the likelihood fit on a single toy experiment generated with initial signal branching ratio set to 0 for the $3\pi\mu$ final state. Shown are the generated data (black points), the signal (green), the background (red) and the total PDF (blue).

The distributions of the fit results, the errors and the pulls are shown in Fig. 8.13 for the signal branching ratio, while the pull distributions are shown in Fig. 8.14 for the nuisance parameters $(\frac{s_y}{c_y})$, in the case of initial branching ratio $\mu_0 = 0$. Table 8.7 summarizes the results.

The distributions of the fit results, the errors and the pulls are shown in Fig. 8.15 for the signal branching ratio, while the pull distributions are shown in Fig. 8.16 for the nuisance parameters $(\frac{s_y}{c_y})$, in the case of initial branching ratio $\mu_0 = 0.0005$. Table 8.8 summarizes the results.



(a) (b)



(c)

Figure 8.13: signal branching ratio fit results for toys generated with $\mu_0 = 0$ for the $3\pi\mu$ final state. The fit with a standard gaussian is shown in blue.

Table 8.7: Results of the fit validation study for the $3\pi\mu$ final state, based on 1000 toys generated with signal branching ratio $\mu_0 = 0$. Shown are the initial values, the average outcome of the fit, the pull mean and the pull variance of the signal branching ratio μ and nuisance parameters $\frac{s_y}{c_y}$ (indicated as p_y in the table).

Parameter	Initial value	Average fit value	Pull mean	Pull Std
μ	0.0	0.00000 ± 0.00013	0.016 ± 0.031	0.984 ± 0.022
$p_{11;3\pi\mu}$	0.5958	0.5955 ± 0.0089	-0.042 ± 0.032	1.010 ± 0.023
$p_{12;3\pi\mu}$	0.5920	0.5920 ± 0.0060	-0.032 ± 0.032	1.019 ± 0.023
$p_{15;3\pi\mu}$	0.566	0.566 ± 0.013	-0.014 ± 0.032	0.996 ± 0.022
$p_{16;3\pi\mu}$	0.5733	0.5733 ± 0.0042	0.003 ± 0.032	1.012 ± 0.023
$p_{17;3\pi\mu}$	0.5640	0.5640 ± 0.0040	0.006 ± 0.031	0.984 ± 0.022
$p_{18;3\pi\mu}$	0.5529	0.5530 ± 0.0037	0.028 ± 0.031	0.986 ± 0.022

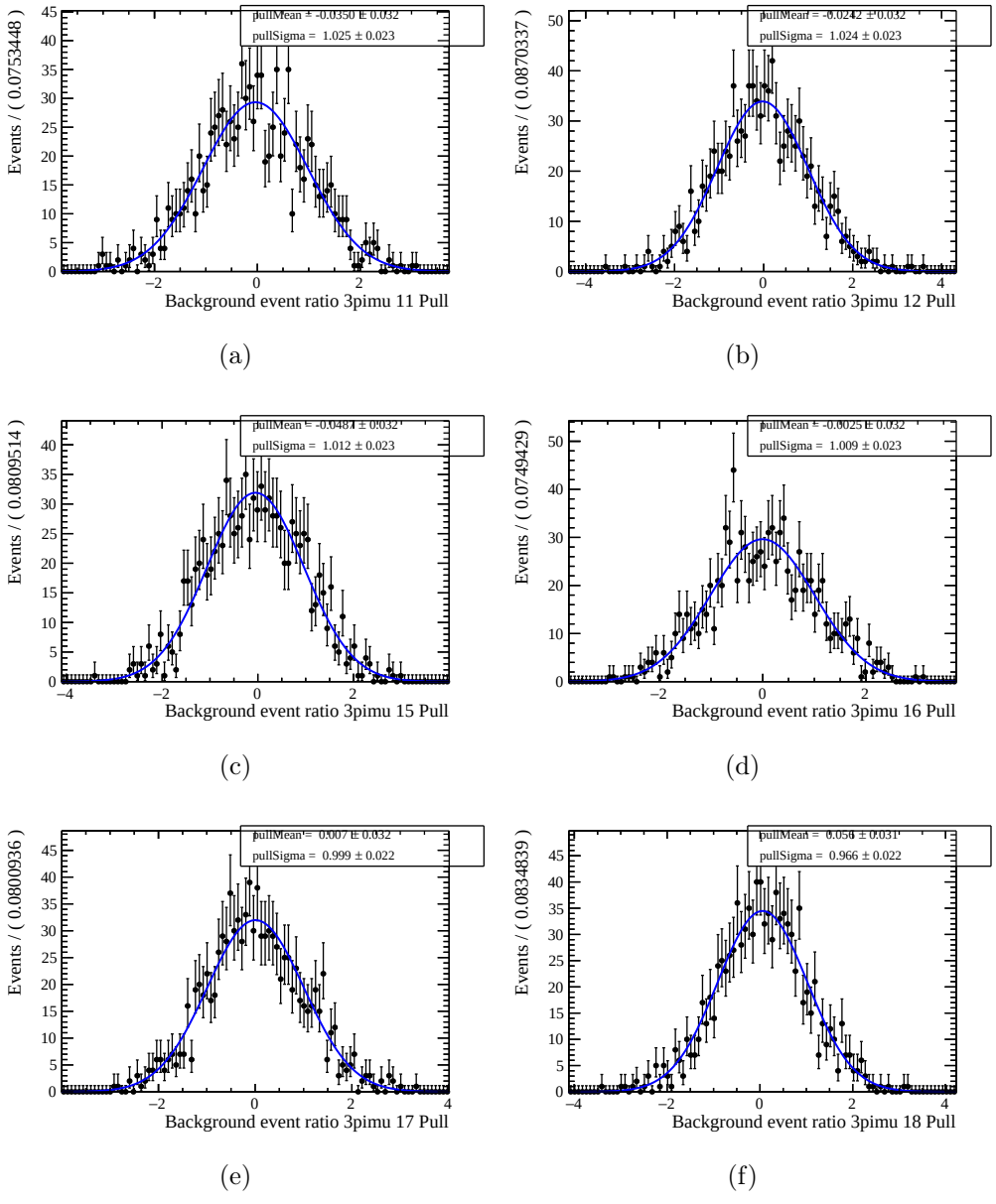


Figure 8.14: Nuisance parameters pulls for toys generated with $\mu_0 = 0$ for the $3\pi\mu$ final state. The fit with a standard gaussian is shown in blue.

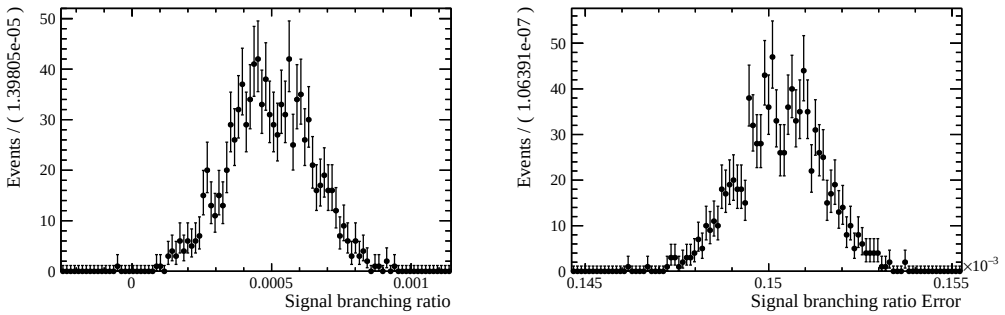


Figure 8.15: signal branching ratio fit results for toys generated with $\mu_0 = 0.0005$ for the $3\pi\mu$ final state. The fit with a standard gaussian is shown in blue.

Table 8.8: Results of the fit validation study for the $3\pi\mu$ final state, based on 1000 toys generated with signal branching ratio $\mu_0 = 0.0005$. Shown are the initial values, the average outcome of the fit, the pull mean and the pull variance of the signal branching ratio μ and nuisance parameters $\frac{s_y}{c_y}$ (indicated as p_y in the table).

Parameter	Initial value	Average fit value	Pull mean	Pull Std
μ	0.0005	0.00050 ± 0.00013	0.023 ± 0.032	0.999 ± 0.022
$p_{11;3\pi\mu}$	0.5878	0.5874 ± 0.0089	-0.051 ± 0.033	1.042 ± 0.023
$p_{12;3\pi\mu}$	0.5840	0.5840 ± 0.0060	-0.032 ± 0.032	1.012 ± 0.023
$p_{15;3\pi\mu}$	0.558	0.558 ± 0.013	-0.009 ± 0.031	0.975 ± 0.022
$p_{16;3\pi\mu}$	0.5667	0.5667 ± 0.0042	-0.002 ± 0.032	1.024 ± 0.023
$p_{17;3\pi\mu}$	0.5570	0.5570 ± 0.0040	-0.008 ± 0.033	1.028 ± 0.023
$p_{18;3\pi\mu}$	0.5471	0.5471 ± 0.0037	0.018 ± 0.031	0.975 ± 0.022

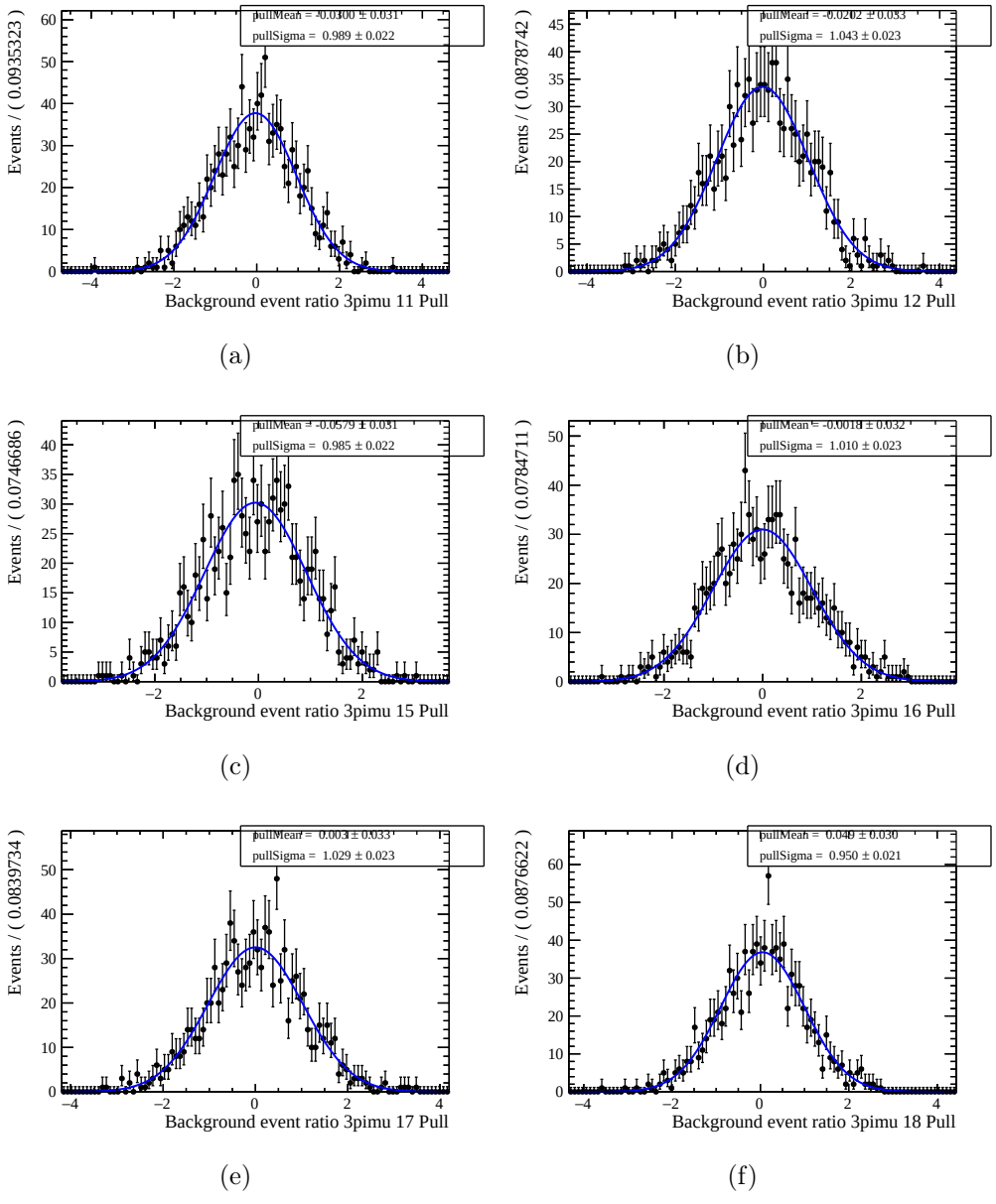


Figure 8.16: Nuisance parameters pulls for toys generated with $\mu_0 = 0.0005$ for the $3\pi\mu$ final state. The fit with a standard gaussian is shown in blue.

Combined final states

The distributions of the fit results, the errors and the pulls are shown in Fig. 8.17 for the signal branching ratio, while the pull distributions are shown in Fig. 8.18 for the nuisance parameters $(\frac{s_y}{c_y})$, in the case of initial branching ratio $\mu_0 = 0$. Table 8.9 summarizes the results.

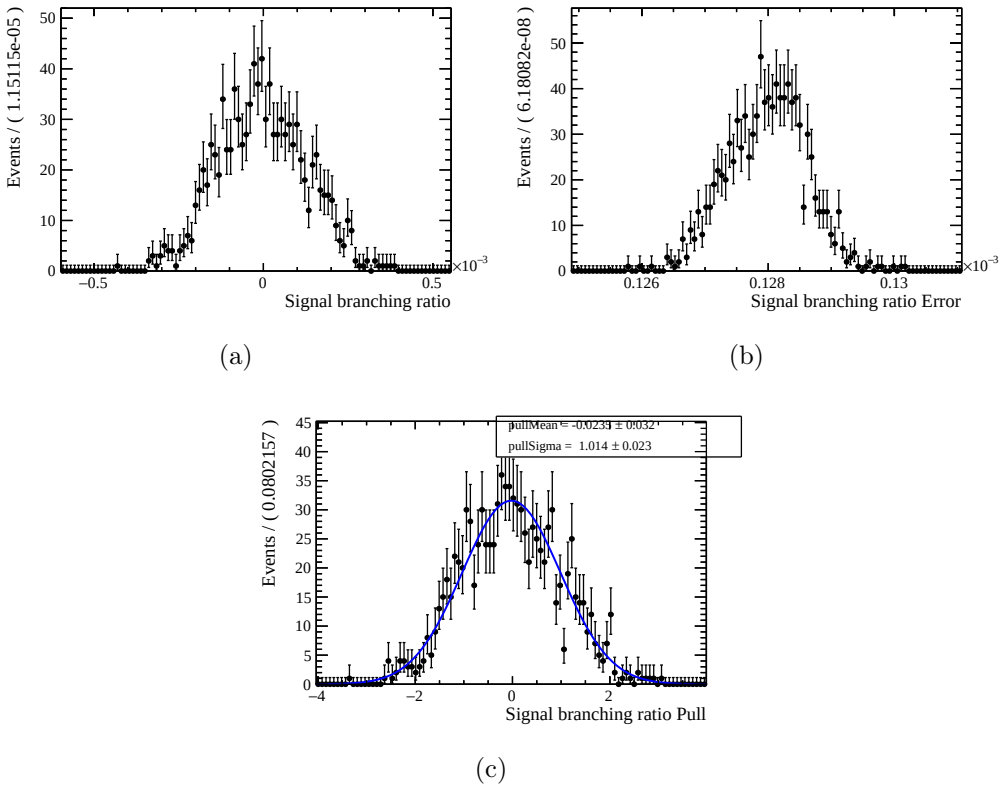


Figure 8.17: signal branching ratio fit results for toys generated with $\mu_0 = 0$ for the simultaneous fit over both modes. The fit with a standard gaussian is shown in blue.

The distributions of the fit results, the errors and the pulls are shown in Fig. 8.19 for the signal branching ratio, while the pull distributions are shown in Fig. 8.20 for the nuisance parameters $(\frac{s_y}{c_y})$, in the case of initial branching ratio $\mu_0 = 0.0005$. Table 8.10 summarizes the results.

Table 8.9: Results of the fit validation study for both final state, based on 1000 toys generated with signal branching ratio $\mu_0 = 0$. Shown are the initial values, the average outcome of the fit, the pull mean and the pull variance of the signal branching ratio μ and nuisance parameters $\frac{s_y}{c_y}$ (indicated as p_y in the table).

Parameter	Initial value	Average fit value	Pull mean	Pull Std
μ	0.0	0.00000 ± 0.00013	-0.030 ± 0.032	1.016 ± 0.023
$p_{11;3\pi3\pi}$	0.554	0.554 ± 0.022	-0.026 ± 0.033	1.038 ± 0.023
$p_{12;3\pi3\pi}$	0.510	0.510 ± 0.014	-0.01 ± 0.03	0.956 ± 0.021
$p_{15;3\pi3\pi}$	0.507	0.508 ± 0.026	0.012 ± 0.032	1.025 ± 0.023
$p_{16;3\pi3\pi}$	0.5383	0.5384 ± 0.0088	0.008 ± 0.032	1.016 ± 0.023
$p_{17;3\pi3\pi}$	0.515	0.515 ± 0.008	0.003 ± 0.031	0.994 ± 0.022
$p_{18;3\pi3\pi}$	0.516	0.5160 ± 0.0075	-0.012 ± 0.031	0.981 ± 0.022
$p_{11;3\pi\mu}$	0.5958	0.5956 ± 0.0088	-0.023 ± 0.033	1.047 ± 0.023
$p_{12;3\pi\mu}$	0.5920	0.5920 ± 0.0060	0.023 ± 0.031	0.989 ± 0.022
$p_{15;3\pi\mu}$	0.566	0.566 ± 0.013	0.039 ± 0.032	1.025 ± 0.023
$p_{16;3\pi\mu}$	0.5733	0.5735 ± 0.0042	0.041 ± 0.032	1.026 ± 0.023
$p_{17;3\pi\mu}$	0.5637	0.5638 ± 0.0039	0.015 ± 0.032	1.003 ± 0.022
$p_{18;3\pi\mu}$	0.5529	0.5531 ± 0.0037	0.052 ± 0.032	1.018 ± 0.023

Table 8.10: Results of the fit validation study for both final state, based on 1000 toys generated with signal branching ratio $\mu_0 = 0$. Shown are the initial values, the average outcome of the fit, the pull mean and the pull variance of the signal branching ratio μ and nuisance parameters $\frac{s_y}{c_y}$ (indicated as p_y in the table).

Parameter	Initial value	Average fit value	Pull mean	Pull Std
μ	0.0005	0.00049 ± 0.00013	-0.043 ± 0.032	1.018 ± 0.023
$p_{11;3\pi3\pi}$	0.549	0.549 ± 0.022	-0.039 ± 0.034	1.061 ± 0.024
$p_{12;3\pi3\pi}$	0.506	0.506 ± 0.014	-0.030 ± 0.030	0.964 ± 0.022
$p_{15;3\pi3\pi}$	0.503	0.504 ± 0.026	0.011 ± 0.033	1.039 ± 0.023
$p_{16;3\pi3\pi}$	0.5347	0.5347 ± 0.0088	-0.004 ± 0.032	1.012 ± 0.023
$p_{17;3\pi3\pi}$	0.5110	0.5110 ± 0.0080	-0.008 ± 0.031	0.973 ± 0.022
$p_{18;3\pi3\pi}$	0.5132	0.5131 ± 0.0075	-0.019 ± 0.031	0.988 ± 0.022
$p_{11;3\pi\mu}$	0.5878	0.5875 ± 0.0089	-0.035 ± 0.034	1.064 ± 0.024
$p_{12;3\pi\mu}$	0.5840	0.5840 ± 0.0060	0.004 ± 0.031	0.991 ± 0.022
$p_{15;3\pi\mu}$	0.558	0.558 ± 0.013	0.032 ± 0.032	1.003 ± 0.022
$p_{16;3\pi\mu}$	0.5667	0.5668 ± 0.0042	0.019 ± 0.032	1.00 ± 0.022
$p_{17;3\pi\mu}$	0.5571	0.5571 ± 0.0039	-0.008 ± 0.032	1.023 ± 0.023
$p_{18;3\pi\mu}$	0.5471	0.5472 ± 0.0037	0.041 ± 0.032	1.021 ± 0.023

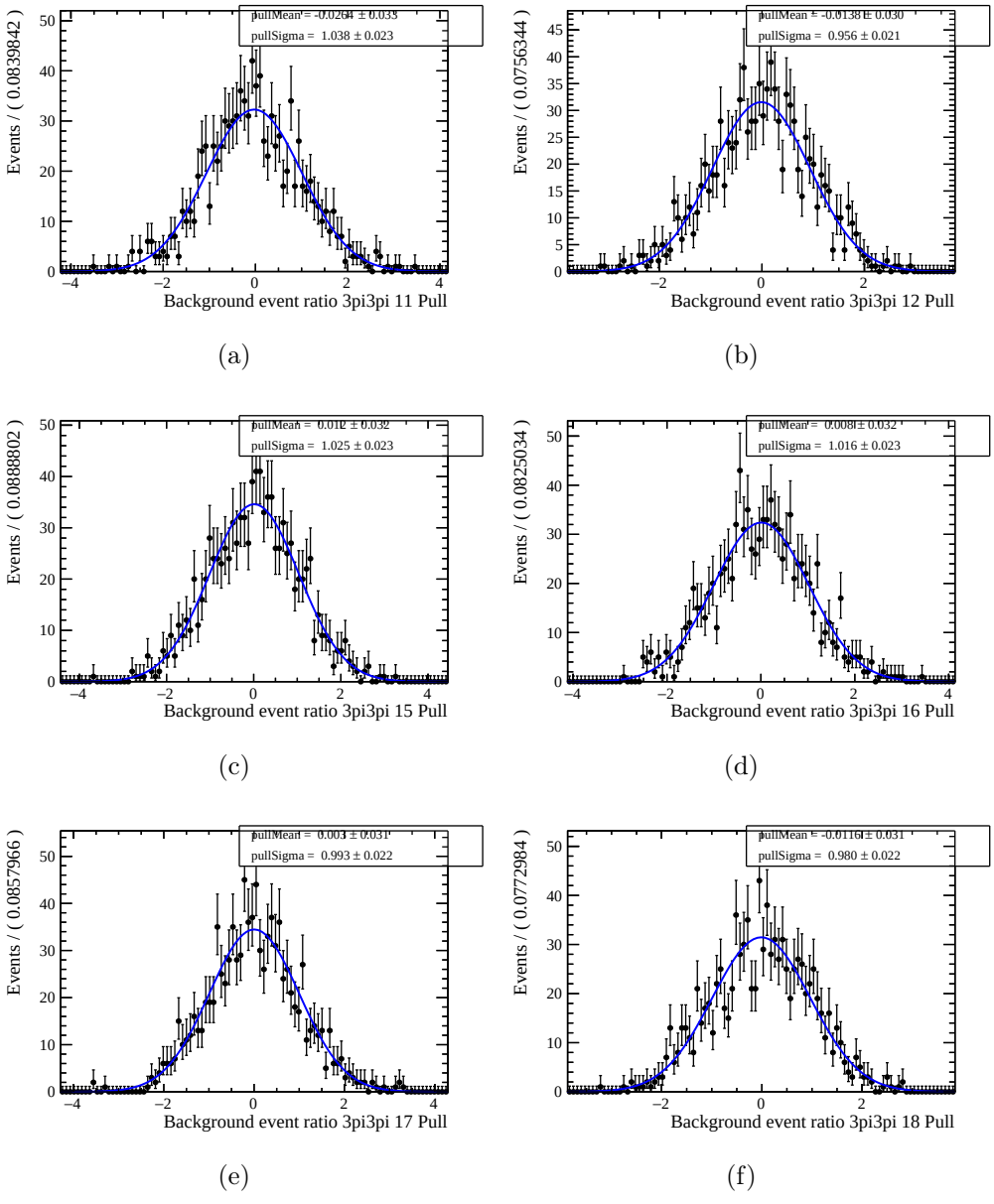


Figure 8.18: Nuisance parameters pulls for toys generated with $\mu_0 = 0$ for the simultaneous fit over both modes. The fit with a standard gaussian is shown in blue.

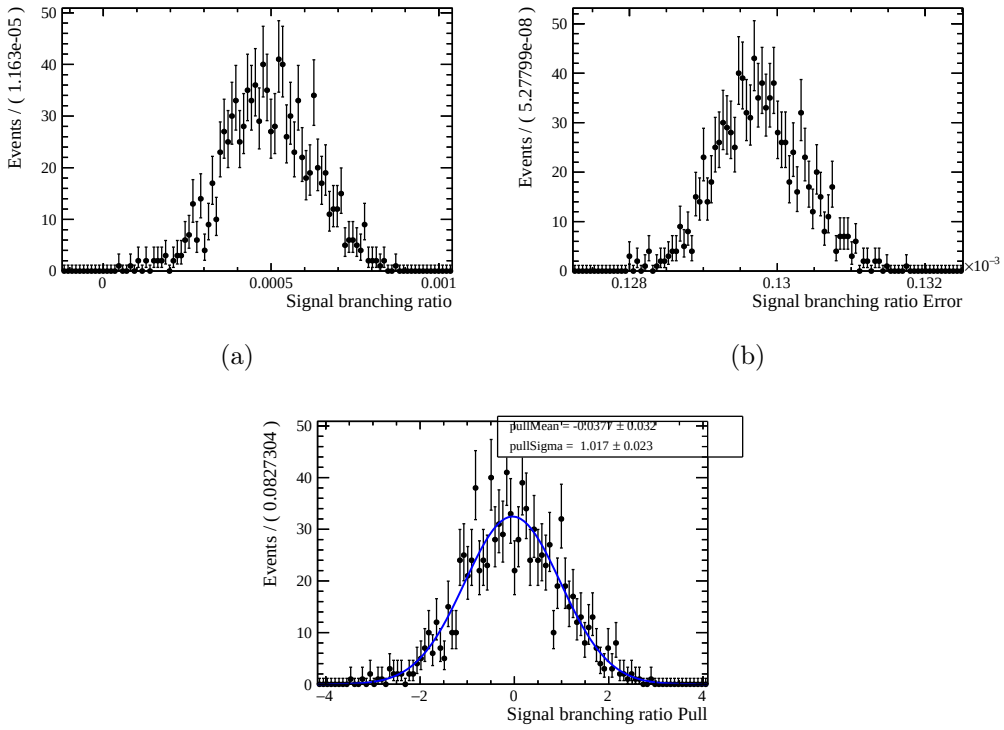
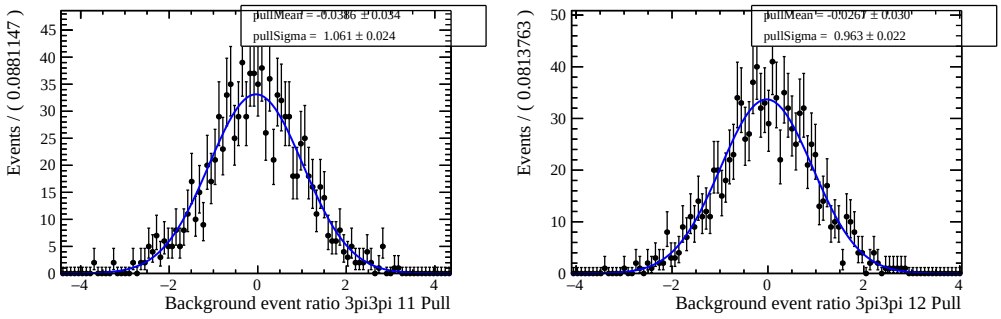
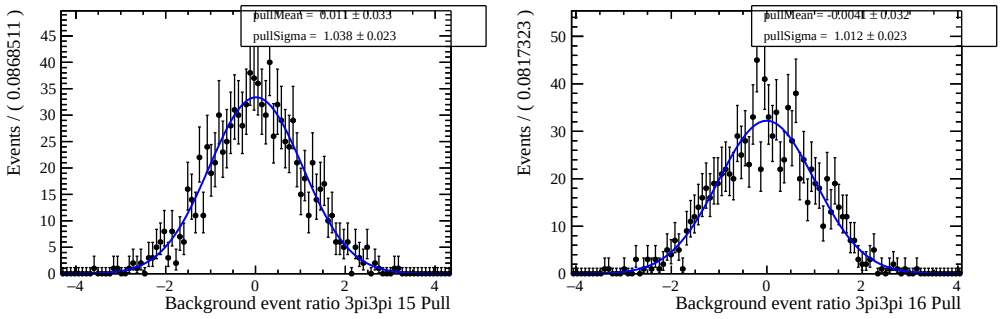


Figure 8.19: signal branching ratio fit results for toys generated with $\mu_0 = 0.0005$ for the simultaneous fit over both modes. The fit with a standard gaussian is shown in blue.



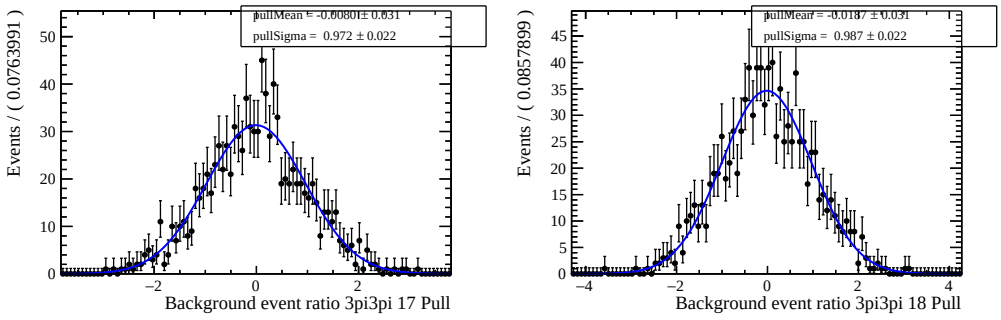
(a)

(b)



(c)

(d)



(e)

(f)

Figure 8.20: Nuisance parameters pulls for toys generated with $\mu_0 = 0.0005$ for the simultaneous fit over both modes. The fit with a standard gaussian is shown in blue.

8.4 Expected limit with the CL_s method

The $B^0 \to K^{*0}\tau^+\tau^-$ signal yield obtained from the fit is translated into a one-sided upper limit on its branching ratio, in case the result is compatible with the background hypothesis, using the CL_s method [17]. To calculate the upper limit, the distribution of the *profile likelihood ratio* test statistics is computed for the "background only" and the "signal + background" hypotheses using the asymptotic formulae [101] for different values of the signal branching ratio. The upper limit corresponds to the branching ratio which gives a p-value for the signal + background hypothesis divided by the p-value of the background only hypothesis of 5%. In the case of the expected upper limit, the median value of the distribution for the background only hypothesis is used in order to compute the p-values.

For each configuration ($3\pi3\pi$ only, $3\pi\mu$ only and combined fit), the expected upper limit at 95% CL is computed considering first the statistical uncertainty only and then combining the statistical and systematic uncertainties in the fit. The systematic uncertainties taken into account are:

fit parameters: it consists of the uncertainties on the fit parameters (f_y, $\frac{\epsilon_y^{\rm ctl}}{\epsilon_y^{\rm sig}}$) which are taken into account as gaussian constraints. The uncertainty on f_y incorporates the uncertainties on the selection efficiencies, on the normalization yield and on the input branching ratios;

signal template: it consists of the statistical uncertainties on the signal templates ($\mathbf{Sig_y}$ and $\mathbf{Cont_y}$) which are taken into account using gaussian constraints. This is done using the Beeston-Barlow light technique [102], consisting in applying a gaussian constraint on the total yield of each bin, depending on the relative uncertainty on that bin;

background template: it consists of the statistical uncertainty on the background template ($\mathbf{Ctl_y}$) which is again taken into account using the Beeston-Barlow light technique.

$3\pi3\pi$ final state

The scan of the branching ratio for the $3\pi3\pi$ final state is done for 30 points in the range $\mathcal{B} \in [0, 0.0025]$. The results are summarized in Table 8.11, while Fig. 8.21 shows the exclusion plot of the expected limit when considering all sources of uncertainties.

Table 8.11: Expected limit at 95% CL computed with the CL_s method for the $3\pi3\pi$ final state only. The central value is reported, as well as the the boundaries of the $\pm1\sigma$ confidence interval. The last column provides the relative increase of the expected limit when the corresponding systematic uncertainty is included, with respect to the limit obtained considering only the statistical uncertainty.

Uncertainty	Expected limit	$+1\sigma$	-1σ	Increase (%)
Statistical only	0.001 019	0.001 431	0.000 732	0.0
Fit parameters	0.001 028	0.001 458	0.000 736	0.9
Signal template	0.001 021	0.001 435	0.000 733	0.2
Background template	0.001 230	0.001 712	0.000 886	20.7
Sig + Bkg template	0.001 279	0.001 807	0.000 913	25.5
Total	0.001 291	0.001 844	0.000 918	26.7

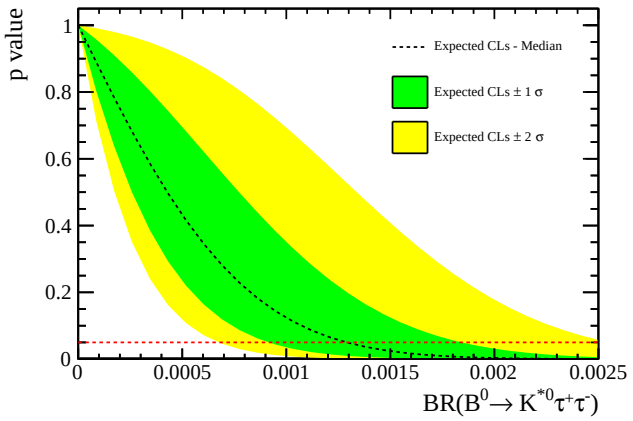


Figure 8.21: Exclusion plot of the expected limit computed for the $3\pi3\pi$ final state only. The total statistical and systematic uncertainties are included.

$3\pi\mu$ final state

The scan of the branching ratio for the $3\pi\mu$ final state is done for 30 points in the range $\mathcal{B} \in [0, 0.001]$. The results are summarized in Table 8.12, while Fig. 8.22 shows the exclusion plot of the expected limit when considering all sources of uncertainties.

Table 8.12: Expected limit at 95% CL computed with the CL_s method for the $3\pi\mu$ final state only. The central value is reported, as well as the the boundaries of the $\pm 1\sigma$ confidence interval. The last column provides the relative increase of the expected limit when the corresponding systematic uncertainty is included, with respect to the limit obtained considering only the statistical uncertainty.

Uncertainty	Expected limit	$+1\sigma$	-1σ	Increase (%)
Statistical only	0.000 262	0.000 365	0.000 190	0.0
Fit parameters	0.000 265	0.000 371	0.000 191	1.1
Signal template	0.000 263	0.000 365	0.000 190	0.4
Background template	0.000 326	0.000 451	0.000 235	24.4
Sig + Bkg template	0.000 329	0.000 457	0.000 236	25.6
Total	0.000 332	0.000 466	0.000 237	26.7

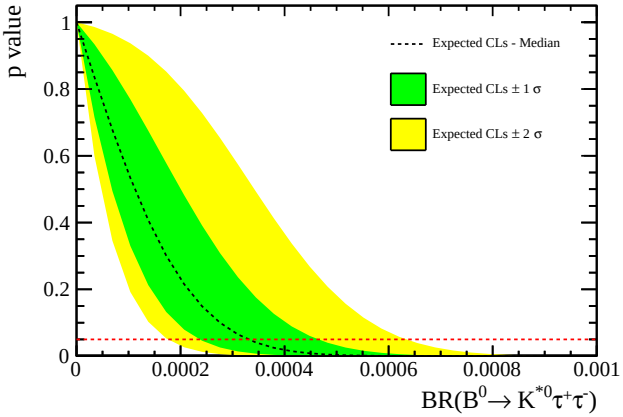


Figure 8.22: Exclusion plot of the expected limit computed for the $3\pi\mu$ final state only. The total statistical and systematic uncertainties are included.

Combined fit over the two modes

The scan of the branching ratio for the simultaneous fit over the two modes is done for 30 points in the range $\mathcal{B} \in [0, 0.001]$. The results are summarized in Table 8.13, while Fig. 8.23 shows the exclusion plot of the expected limit when considering all sources of uncertainties. The expected upper limit is computed to be $3.2(2.7) \cdot 10^{-4}$ at 95(90)% CL, which represents an improvement of one order of magnitude with respect to the current experimental limit.

Table 8.13: Expected limit at 95% CL computed with the CL_s method for the combined fit over the two final states. The central value is reported, as well as the the boundaries of the $\pm 1\sigma$ confidence interval. The last column provides the relative increase of the expected limit when the corresponding systematic uncertainty is included, with respect to the limit obtained considering only the statistical uncertainty.

Uncertainty	Expected limit	$+1\sigma$	-1σ	Increase (%)
Statistical only	0.000 254	0.000 352	0.000 183	0.0
Fit parameters	0.000 256	0.000 358	0.000 184	0.8
Signal template	0.000 254	0.000 353	0.000 184	0.0
Background template	0.000 314	0.000 437	0.000 228	23.6
Sig + Bkg template	0.000 317	0.000 442	0.000 229	24.8
Total	0.000 320	0.000 448	0.000 230	26.0

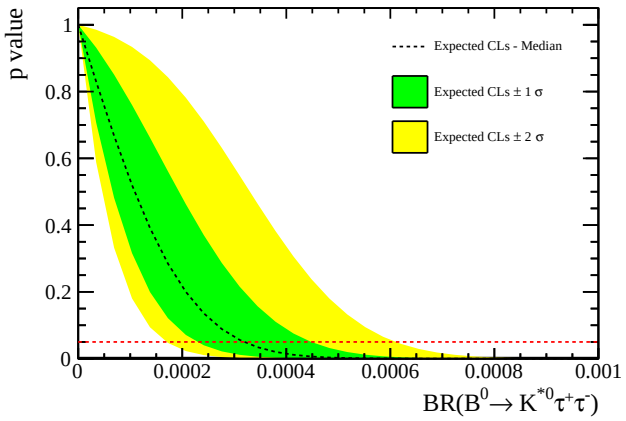


Figure 8.23: Exclusion plot of the expected limit computed for the fit on both modes. The total statistical and systematic uncertainties are included.

Fit on merged data-taking years

Tables 8.11, 8.12 and 8.13 show that the major contribution to the systematic uncertainty on the upper limit comes from the uncertainty on the background shape. In order to mitigate this uncertainty the possibility of merging data-taking years in the final fit is investigated. The expected upper limit is computed on the dataset obtained by merging years with similar data-taking conditions, corresponding to the periods 2011-2012, 2015-2016 and 2017-2018. The results are shown in Tables 8.14, 8.15, 8.16 for the $3\pi3\pi$, $3\pi\mu$ and combined final states respectively. The results show no improvements with respect to the case in which all the data-taking years are kept separate.

Table 8.14: Expected limit at 95% CL computed with the CL_s method for the $3\pi3\pi$ final state only. The data-taking years corresponding to the periods 2011-2012, 2015-2016 and 2017-2018 are merged.

Uncertainty	Expected limit	$+1\sigma$	-1σ	Increase (%)
Statistical only	0.001 046	0.001 466	0.000 751	0.0
Fit parameters	0.001 065	0.001 518	0.000 757	1.8
Signal template	0.001 047	0.001 471	0.000 751	0.1
Background template	0.001 263	0.001 760	0.000 911	20.7
Sig + Bkg template	0.001 314	0.001 856	0.000 936	25.6
Total	0.001 336	0.001 917	0.000 943	27.7

Table 8.15: Expected limit at 95% CL computed with the CL_s method for the $3\pi\mu$ final state only. The data-taking years corresponding to the periods 2011-2012, 2015-2016 and 2017-2018 are merged.

Uncertainty	Expected limit	$+1\sigma$	-1σ	Increase (%)
Statistical only	0.000 266	0.000 369	0.000 193	0.0
Fit parameters	0.000 269	0.000 379	0.000 194	1.1
Signal template	0.000 266	0.000 370	0.000 193	0.0
Background template	0.000 330	0.000 458	0.000 238	24.1
Sig + Bkg template	0.000 333	0.000 464	0.000 239	25.2
Total	0.000 338	0.000 478	0.000 241	27.1

Table 8.16: Expected limit at 95% CL computed with the CL_s method for the combined fit over the two final states. The data-taking years corresponding to the periods 2011-2012, 2015-2016 and 2017-2018 are merged.

Uncertainty	Expected limit	$+1\sigma$	-1σ	Increase (%)
Statistical only	0.000 258	0.000 358	0.000 186	0.0
Fit parameters	0.000 261	0.000 368	0.000 188	1.2
Signal template	0.000 258	0.000 358	0.000 187	0.0
Background template	0.000 319	0.000 443	0.000 231	23.6
Sig + Bkg template	0.000 322	0.000 448	0.000 232	24.8
Total	0.000 327	0.000 461	0.000 234	26.7

Chapter 9

Selection optimization

In this chapter the optimization procedure used to choose the BDT input variables, the BDT hyper-parameters and the cuts on the BDT output is presented[1]. The procedure aims at minimizing the expected limit in the case where no signal is observed. The parameters that are taken into account for the optimization of the MVA selection are:

1. the BDT1 input variables;
2. the cut on the flattened output of BDT1;
3. the BDT2 input variables;
4. the cut on the flattened output of BDT2;
5. the fitBDT input variables;
6. the hyper-parameters of the fitBDT: the number of trees in each decision tree (NTrees), the minimum size of each node (MinSize, as fraction of the total number of events), the maximum depth of each tree (MaxDepth), the learning rate (LR).

The hyper-parameters of BDT1 and BDT2 have not been optimized and their values are the TMVA default ones:

$$\begin{aligned} \text{NTrees} &= 300, \\ \text{MinSize} &= 3, \\ \text{MaxDepth} &= 2, \\ \text{LR} &= 0.5. \end{aligned} \tag{9.1}$$

The hyper-parameters are the same for the different data-taking periods.

[1] The optimization has been performed in a previous version of the analysis, using signal MC samples without the neutral pions component. Since the procedure has not been repeated on the new MC samples including the neutral pions component, some of the results look oddly better than the ones quoted in Sec. 8.4.

9.1 The variable finding procedure

The optimization makes use of an iterative procedure in order to choose the input variables for a given stage of the MVA selection [103]. The iterative procedure aims at finding the set of variables maximizing the weighted area under the receiver operating characteristic (ROC) curve:

$$\mathrm{FoM} = \frac{\sum_i w_i F_i}{\sum_i w_i A_i}, \tag{9.2}$$

where the subscript i runs over the signal efficiency bins whose boundaries are $[0, 0.4, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 1]$ (chosen in such a way not to have empty bins at low efficiency), $A_i = \epsilon_i^{\mathrm{sup}} - \epsilon_i^{\mathrm{inf}}$, with $\epsilon_i^{\mathrm{sup\ (inf)}}$ being the upper (lower) boundary of the i-th bin, F_i is the area under ROC curve in the i-th bin, and w_i is a weight defined using the Punzi figure of merit [104] in such a way to give more importance to low efficiency (hence high purity) bins:

$$w_i = \frac{A_i}{\frac{5}{2} + \sqrt{B_i}}, \tag{9.3}$$

where $B_i = B \cdot (\zeta_i^{\mathrm{sup}} - \zeta_i^{\mathrm{inf}})$, with $\zeta_i^{\mathrm{sup\ (inf)}}$ being the corresponding background rejection for efficiency $\epsilon_i^{\mathrm{inf\ (sup)}}$ and B the number of background events in the K^{*0} mass signal region after the selection (estimated by rescaling the number of events in the control region after a preliminary selection).
The variable finding procedure works as follows:

1. a list of usable variables $\mathrm{L_{usable}}$ is considered;
2. an empty list of final BDT variables $\mathrm{L_{final}}$ is defined;
3. a BDT is trained using the variables in the list $\mathrm{L_{temp}}$ built with the variables in $\mathrm{L_{final}}$ and a variable $v \in \mathrm{L_{usable}}$, for each element of $\mathrm{L_{usable}}$;
4. the variable $v \in \mathrm{L_{usable}}$ corresponding to the BDT with the maximum FoM is added to $\mathrm{L_{final}}$;
5. points 3 and 4 are repeated until the FoM reaches a plateau.

9.2 Optimization for the $3\pi3\pi$ final state

For the $3\pi3\pi$ final state, BDT1 and fitBDT use variables uncorrelated with the K^{*0} mass, in order to be able to build the background model from the output of the fitBDT on events in the K^{*0} mass sidebands. Because of that, the output of BDT1 for events that survive the selection can be used as input variable for the fitBDT. For this reason, and since the BDTs are trained in sequence, it is difficult

to define a unique strategy in order to tune in the best possible way all parameters without overcomplicating the optimization. Therefore, some simplifications are implemented in order to cope with the complexity of the procedure. In the first part of the optimization the BDT2 is removed, in such a way to obtain a simpler selection involving only BDT1 before the training of the fitBDT.

BDT1 variables The variables of BDT1, listed in Sec. 5.7, are chosen using the variable finding procedure among variables uncorrelated with the K^{*0} mass.

BDT1 cut and fitBDT variables Since the fitBDT is trained after the cut on BDT1, different cuts will lead to different sets of variables which maximize the FoM in Eq. 9.2. Therefore, five different cuts are applied to the training samples of the 2016 fitBDT and for each of them the variable finding procedure is used to extract a different set of variables, as shown in Table 9.1. For each set of fitBDT variables,

Table 9.1: Variable sets obtained using the variable finding procedure on the training samples of the fitBDT after different cuts on BDT1 for the $3\pi3\pi$ final state. The colors match the ones in Fig. 9.1.

Cut on training samples	Variable set
BDT1 > 0.1	A
BDT1 > 0.2	B
BDT1 > 0.3	C
BDT1 > 0.4	D
BDT1 > 0.5	E

the expected limit at 95% CL is computed as a function of the cut on BDT1, where only the statistical uncertainty is considered. The results are reported in Table 9.2 and Fig. 9.1. The variables in set B give the best expected limit, and are therefore chosen as input variables for the fitBDT. They are reported in Sec. 8.1. The discarded sets of variables are reported in Tables 9.3, 9.4, 9.5, 9.6. Moreover, the cut at BDT1 > 0.6 gives the best limit estimation, and is therefore chosen as BDT1 cut.

Table 9.2: Expected limit in units of 10^{-4} for the $3\pi 3\pi$ final state as a function of the cut on BDT1 for different sets of fitBDT variables. The colors match the ones in Fig. 9.1.

	Set A	Set B	Set C	Set D	E
BDT1 > 0.2	17.7	15.0	17.0	17.0	17.9
BDT1 > 0.3	16.5	14.0	16.0	16.2	16.5
BDT1 > 0.4	16.2	13.4	15.4	15.8	16.4
BDT1 > 0.5	16.2	13.4	15.4	15.6	16.2
BDT1 > 0.6	15.9	12.8	14.9	14.7	16.2
BDT1 > 0.7	15.8	13.2	15.3	15.5	16.7
BDT1 > 0.8	16.7	13.2	15.7	14.7	17.1

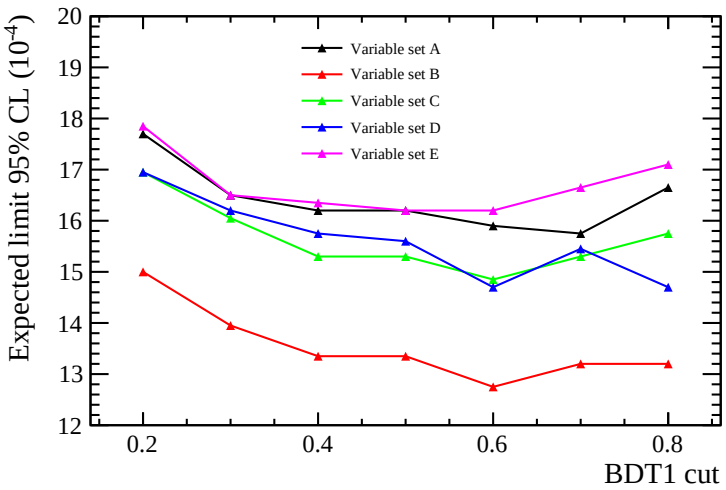


Figure 9.1: Evolution of the expected limit as a function of the cut on BDT1 for different sets of fitBDT variables.

Table 9.3: FitBDT variables: discarded set A.

min(π_3^+ IsoBDTSecondValue, π_3^- IsoBDTSecondValue)
max(τ^+ IsoBDTSecondValue, τ^- IsoBDTSecondValue)
max(τ^+ decay vertex and PV distance, τ^- decay vertex and PV distance)
max(τ^+ DIRA wrt B^0 decay vertex, τ^- DIRA wrt B^0 decay vertex)
min(tauP_P_opt, tauM_P_opt)
max(tauP_P_opt, tauM_P_opt)
min($M_{\pi_2^-\pi_3^+}$, $M_{\pi_2^+\pi_3^-}$)
max($M_{\pi_1^+\pi_2^-}$, $M_{\pi_1^-\pi_2^+}$)
max(τ^+ p_{T}, τ^- p_{T})
B^0 analytically reconstructed mass from optimal solution
B^0 corrected mass
B_DIRA_aa
B^0 IP
max(π_1^+ IP χ^2 wrt PV, π_1^- IP χ^2 wrt PV)
BDT1

Table 9.4: FitBDT variables: discarded set C.

max(τ^+ IsoBDTSecondValue, τ^- IsoBDTSecondValue)
B^0 DeltaChi2MassOneTrack
max(π_1^+ IP wrt PV, π_1^- IP wrt PV)
max(τ^+ DIRA wrt B^0 decay vertex, τ^- DIRA wrt B^0 decay vertex)
max(τ^+ decay vertex and PV distance, τ^- decay vertex and PV distance)
max(τ^+ flight distance χ^2, τ^- flight distance χ^2)
min(τ^+ max DOCA of decay products, τ^- max DOCA of decay products)
min(DOCA of π_2^- and π_3^+, DOCA of π_2^+ and π_3^-)
max(τ^+ IP, τ^- IP)
min(τ^+ resMass, τ^- resMass)
min(τ^+ visible mass, τ^- visible mass)
max($M_{\pi_2^-\pi_3^+}$, $M_{\pi_2^+\pi_3^-}$)
max(tauP_PZ_a, tauM_PZ_a)
max(τ^+ p_{T}, τ^- p_{T})
B^0 IP
B_OPT_decayLength
PV χ^2
BDT1

Table 9.5: FitBDT variables: discarded set D.

min(τ^+ IsoBDTSecondValue, τ^- IsoBDTSecondValue)
max(τ^+ IsoBDTSecondValue, τ^- IsoBDTSecondValue)
max(τ^+ IsoBDTThirdValue, τ^- IsoBDTThirdValue)
min(π_3^+ IsoBDTThirdValue, π_3^- IsoBDTThirdValue)
min(τ^+ neutral cone isolation vPT, τ^- neutral cone isolation vPT)
B^0 CDFIso
B^0 visible mass
B^0 corrected mass
$M_{\tau^+\tau^-}$
min(τ^+ resMass, τ^- resMass)
min(τ^+ visible mass, τ^- visible mass)
min(τ^+ DOCA of decay products, τ^- DOCA of decay products)
min(τ^+ flight distance χ^2, τ^- flight distance χ^2)
max(τ^+ flight distance, τ^- flight distance)
min(τ^+ IP χ^2 wrt PV, τ^- IP χ^2 wrt PV)
PV χ^2
max(tauP_P_opt, tauM_P_opt)
B^0 p_{T}
BDT1

Table 9.6: FitBDT variables: discarded set E.

min(τ^+ IsoBDTSecondValue, τ^- IsoBDTSecondValue)
B^0 neutral cone isolation DELTAETA
B^0 corrected mass
max(τ^+ flight distance χ^2, τ^- flight distance χ^2)
max(τ^+ flight distance, τ^- flight distance)
min(τ^+ DIRA wrt PV, τ^- DIRA wrt PV)
max(τ^+ DIRA wrt B^0 decay vertex, τ^- DIRA wrt B^0 decay vertex)
min(tauP_P_opt, tauM_P_opt)
max(tauP_PZ_a, tauM_PZ_a)
min(τ^+ visible mass, τ^- visible mass)
max(π_3^+ IP wrt PV, π_3^- IP wrt PV)
PV χ^2
min(π_2^- IP χ^2 wrt PV, π_2^+ IP χ^2 wrt PV)

BDT2 variables and cut After the choice of the variables for BDT1 and fitBDT, and the cut on BDT1 chosen, the BDT2 is added to the selection procedure. The variables are chosen using the iterative procedure after applying the cut BDT1 > 0.6 on the BDT2 training samples and are reported in Sec. 5.7. The cut is chosen in such a way to discard the bin with the lowest signal/background ratio in Fig. 8.1, corresponding to a cut on BDT2 > 0.1.

fitBDT hyper-parameters In order to find the best set of hyper-parameters for the fitBDT, a first rough scan of the expected limit at 95% CL is performed by varying the values of the hyper parameters. Once the starting point is defined, a more precise scan in the following subsets is performed:

- NTrees $\in [100, 150, 200, 250]$,
- MinSize (%) $\in [2, 2.5, 3, 3.5]$,
- MaxDepth $\in [2, 3, 4]$,
- LR $\in [0.12, 0.14, 0.16, 0.18, 0.2, 0.22, 0.24, 0.26, 0.28]$,

for a total of 432 iterations. Moreover, an additional scan of NTrees in the range $[220, 240, 260, 280, 300]$ is performed fixing the values of the other hyper parameters to the ones that give the best expected limit. The final set of parameters giving the best limit estimation is:

$$\begin{aligned} \text{NTrees} &= 240, \\ \text{MinSize} &= 3, \\ \text{MaxDepth} &= 3, \\ \text{LR} &= 0.2. \end{aligned} \tag{9.4}$$

9.3 Optimization for the $3\pi\mu$ final state

The optimization of the $3\pi\mu$ final state is complicated by a high degree of correlations of the BDT variables with the K^{*0} mass. Therefore, the output of BDT1 and BDT2 is not used as input variable for the fitBDT. Similarly to what done for the $3\pi 3\pi$ final state, the BDT2 is initially removed in order to simplify the procedure.

BDT1 variables The variables of BDT1, listed in Sec. 5.7, are chosen using the variable finding procedure. The three best-ranking variables are removed from BDT1 in order to save discriminating power for the fitBDT.

BDT2 and fitBDT variables The fitBDT variables are chosen using the variable finding procedure after applying on the training samples the preliminary cut BDT1 > 0.5. The correlations between the K^{*0} mass and the variables obtained from the procedure are checked. The variables which show a high degree of correlation with the K^{*0} mass are removed from the fitBDT. Moreover, the correlation between the output of the fitBDT and the K^{*0} mass is checked by removing one by one the input variables from the fitBDT. The variables which introduce correlations between the fitBDT and the K^{*0} mass are removed, in such a way to obtain the final list of fitBDT variables presented in Sec. 8.1. All variables discarded from the fitBDT are used to train BDT2. Moreover, the output of BDT1 is used as input variable for BDT2. The final list of variables for BDT2 is shown in Sec. 5.7.

BDT1 and BDT2 cuts The cut on BDT2 is chosen by making a scan of the average fit error from toys with 0 generated signal events and the expected limit at 95% CL in the case where no signal is observed as a function of the cut on BDT2. The preliminary cut BDT1 > 0.5 is applied. The results are shown in Table 9.7. The cut BDT2 > 0.3 gives the smallest average fit error and expected limit, and is

Table 9.7: Average fit error and expected limit at 95% CL with the CLs method for the $3\pi\mu$ final state as a function of the cut on BDT2. The preliminary cut BDT1 > 0.5 is applied.

BDT2 cut	Average fit error (10^{-5})	Expected limit (10^{-5})
BDT2 > 0	14.64	28.5
BDT2 > 0.1	14.63	28.5
BDT2 > 0.2	14.51	28.5
BDT2 > 0.3	14.40	28.5
BDT2 > 0.4	14.43	28.5
BDT2 > 0.5	15.26	28.5
BDT2 > 0.6	15.42	30.0
BDT2 > 0.7	17.48	33.0
BDT2 > 0.8	19.62	37.5
BDT2 > 0.9	20.25	51.0

therefore chosen as cut for BDT2.
Similarly, the cut on BDT1 is chosen with a scan of the average fit error from toys with 0 generated signal events (and the expected limit at 95% CL in the case where no signal is observed) as a function of the cut on BDT2. The cut BDT2 > 0.3 is applied. The results are shown in Table 9.8. The cut BDT1 > 0.6 gives the smallest average fit error (and expected limit), and is therefore chosen as cut for BDT1.

Table 9.8: Average fit error and expected limit at 95% CL with the CLs method for the $3\pi\mu$ final state as a function of the cut on BDT1. The cut BDT2 > 0.3 is applied.

BDT1 cut	Average fit error (10^{-5})	Expected limit (10^{-5})
BDT1 > 0.5	14.64	28.5
BDT1 > 0.6	14.40	28.5
BDT1 > 0.7	15.21	28.5

fitBDT hyper-parameters Similarly to the $3\pi3\pi$ final state, after performing a rough scan of the expected limit at 95% CL in order to find a good starting point, a more precise optimization is performed varying the values of the hyper-parameters of the fitBDT in the following subsets:

- NTrees $\in [100, 150, 200, 250]$,
- MinSize (%) $\in [0.5, 1, 1.5, 2]$,
- MaxDepth $\in [4, 5, 6]$,
- LR $\in [0.02, 0.04, 0.06, 0.08, 0.1, 0.12, 0.14, 0.16, 0.18]$,

for a total of 432 iterations. Moreover, an additional scan of NTrees in the range $[220, 240, 260, 280, 300]$ is performed fixing the values of the other hyper parameters to the ones that give the best expected limit. The final set of parameters giving the best limit estimation is:

$$\begin{aligned} \text{NTrees} &= 280, \\ \text{MinSize} &= 1, \\ \text{MaxDepth} &= 5, \\ \text{LR} &= 0.1. \end{aligned} \tag{9.5}$$

Chapter 10

Cross-checks

This chapter describes the cross-checks performed in order to validate the analysis strategy.

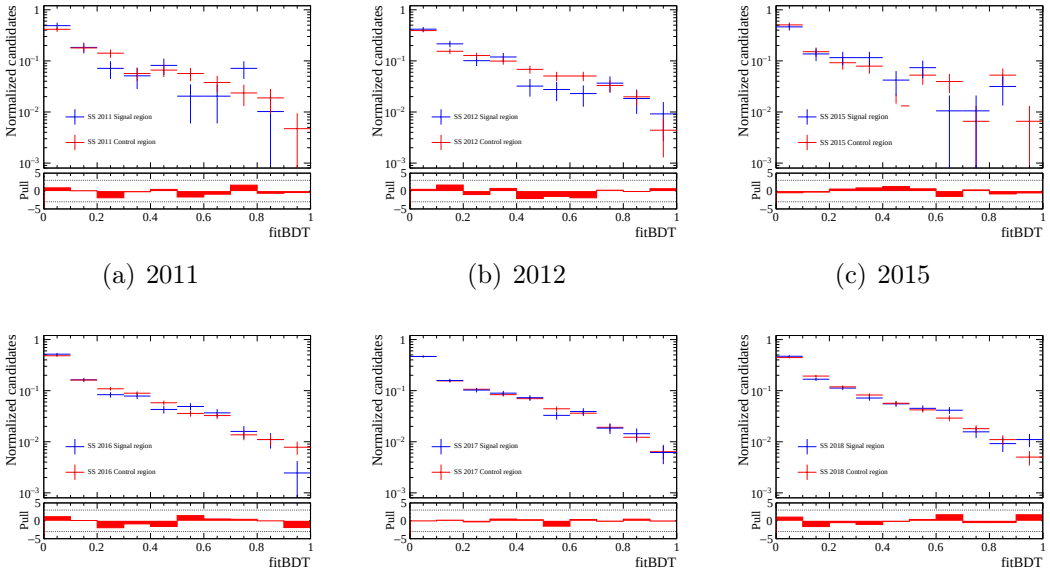
10.1 Fit to the Same-Sign data

As described in Sec. 4.1, SS data are selected requiring both τ lepton candidates to have the same charge. In order to validate the analysis strategy, the selection and fit procedures are applied to the SS data. Since no signal is present in such a dataset, there is no need to blind the most sensitive bins of the fitBDT, which allows to check the fit procedure in the full BDT spectrum. Moreover, the fitted signal branching ratio must be compatible with 0.

$3\pi3\pi$ final state

The comparison of the $3\pi3\pi$ fitBDT distribution in signal and control region for SS data after the full selection is shown in Fig. 10.1. The comparison shows good agreement between the two regions.

The result of the simultaneous fit over the data-taking years is reported in Table 10.1 and in Fig. 10.2, taking into account the statistical and systematic uncertainties.



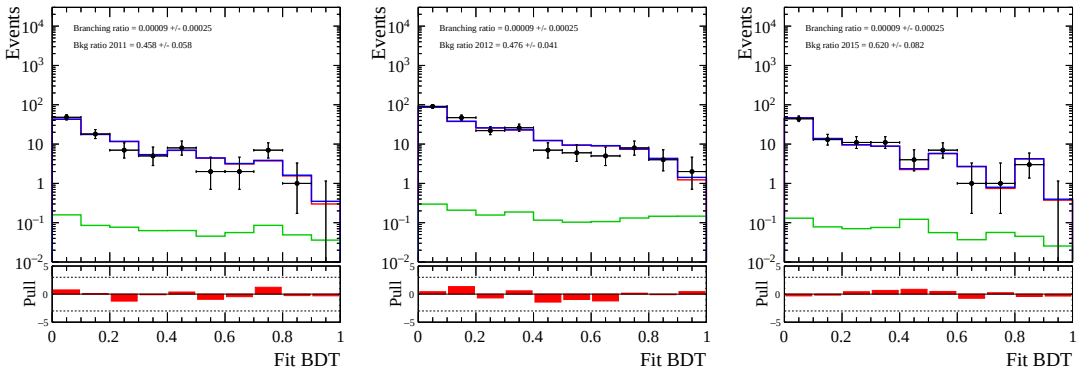
(a) 2011 (b) 2012 (c) 2015

(d) 2016 (e) 2017 (f) 2018

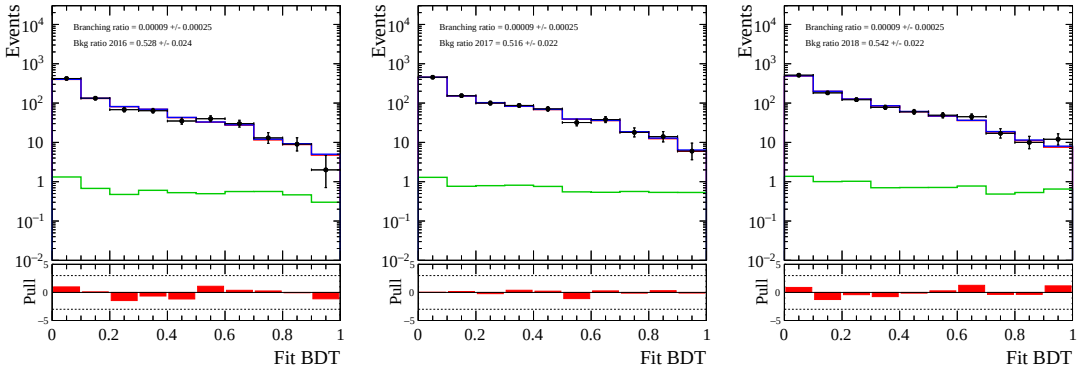
Figure 10.1: Comparison of the $3\pi3\pi$ fitBDT distribution in signal and control regions for SS data in logarithmic scale.

Table 10.1: Result of the simultaneous fit over the data-taking years of $3\pi3\pi$ SS data. Shown are the fitted values of the signal branching ratio μ and nuisance parameters $\frac{s_y}{c_y}$ (indicated as p_y in the table).

Parameter	Fit result
μ	$0.000\,09 \pm 0.000\,25$
$p_{11;3\pi\mu}$	0.458 ± 0.058
$p_{12;3\pi\mu}$	0.476 ± 0.041
$p_{15;3\pi\mu}$	0.620 ± 0.082
$p_{16;3\pi\mu}$	0.528 ± 0.024
$p_{17;3\pi\mu}$	0.516 ± 0.022
$p_{18;3\pi\mu}$	0.542 ± 0.022



(a) 2011 (b) 2012 (c) 2015

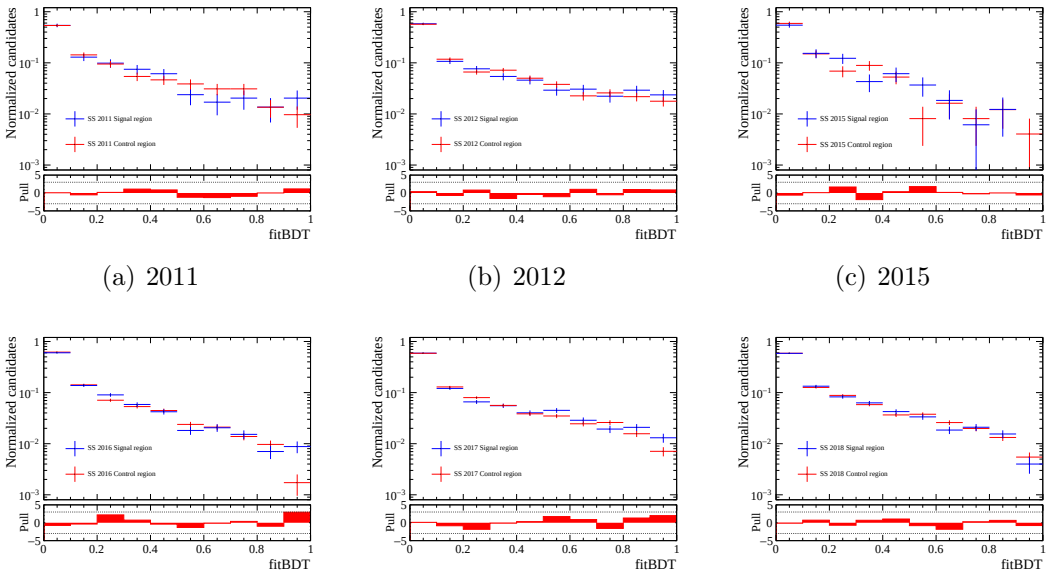


(d) 2016 (e) 2017 (f) 2018

Figure 10.2: Result of the simultaneous fit over the data-taking years of $3\pi 3\pi$ SS data. Shown are the data (black points), the signal (green), the background (red) and the total PDF (blue).

$3\pi\mu$ final state

The comparison of the $3\pi\mu$ fitBDT distribution in signal and control region for SS data after the full selection is shown in Fig. 10.3. The comparison shows good agreement between the two regions.



(a) 2011 (b) 2012 (c) 2015

(d) 2016 (e) 2017 (f) 2018

Figure 10.3: Comparison of the $3\pi\mu$ fitBDT distribution in signal and control regions for SS data in logarithmic scale.

The result of the simultaneous fit over the data-taking years is reported in Table 10.2 and in Fig. 10.4, taking into account the statistical and systematic uncertainties. The fit result is compatible with 0 at the level of $\sim 2.8\sigma$. To better understand the nature of the difference with respect to the background only hypothesis, the fit is performed separately for each data-taking years. The results are summarized in Table 10.3. A $\sim 2\sigma$ difference with respect to the null hypothesis is obtained for the 2016 and 2017 datasets, while for the remaining data-taking years the values of the branching ratio are compatible with 0 within one standard deviation. This suggests that the results of the simultaneous fit are most probably driven by fluctuations on the most sensitive bins of the aforementioned datasets[1].

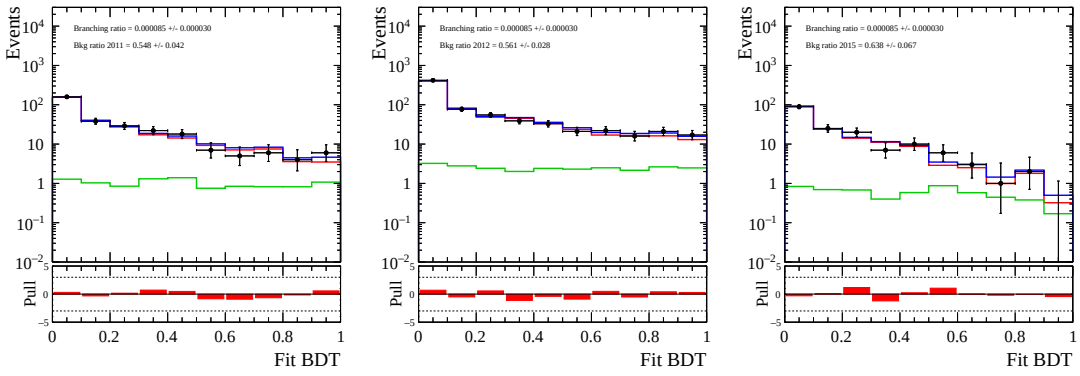
[1]In the latest version of the analysis the result is indeed compatible with 0 within one standard deviation.

Table 10.2: Result of the simultaneous fit over the data-taking years of $3\pi\mu$ SS data. Shown are the fitted values of the signal branching ratio μ and nuisance parameters $\frac{s_y}{c_y}$ (indicated as p_y in the table).

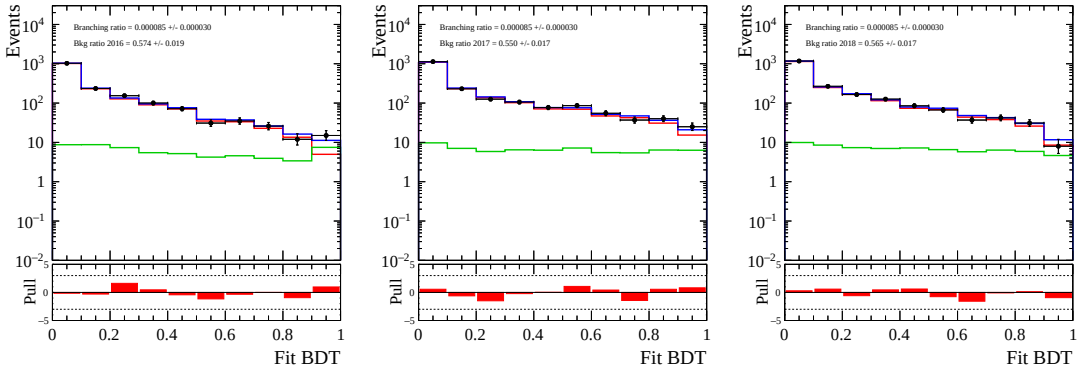
Parameter	Fit result
μ	$0.000\,085 \pm 0.000\,030$
$p_{11;3\pi\mu}$	0.548 ± 0.042
$p_{12;3\pi\mu}$	0.561 ± 0.028
$p_{15;3\pi\mu}$	0.638 ± 0.067
$p_{16;3\pi\mu}$	0.574 ± 0.019
$p_{17;3\pi\mu}$	0.550 ± 0.017
$p_{18;3\pi\mu}$	0.565 ± 0.017

Table 10.3: Results of the fit on SS data for the $3\pi\mu$ final state separately for each data-taking year.

Year	μ	$p_{y;3\pi\mu}$
2011	0.00006 ± 0.00011	0.554 ± 0.049
2012	0.000053 ± 0.000082	0.568 ± 0.033
2015	0.00018 ± 0.00025	0.617 ± 0.082
2016	0.000126 ± 0.000062	0.568 ± 0.020
2017	0.000126 ± 0.000066	0.543 ± 0.019
2018	-0.000026 ± 0.000064	0.580 ± 0.019



(a) 2011 (b) 2012 (c) 2015

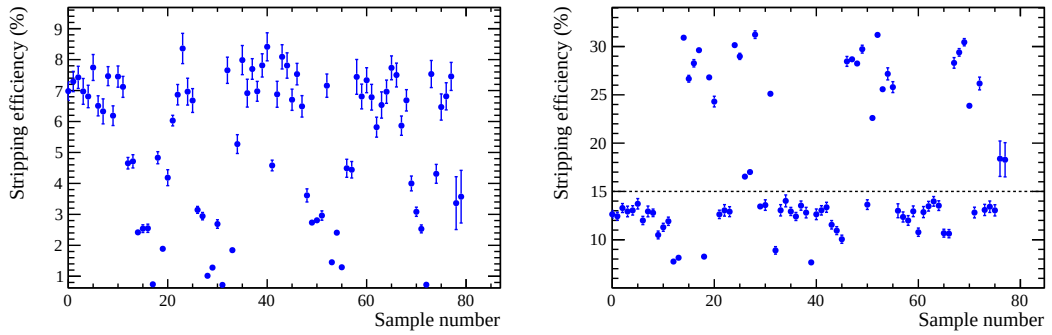


(d) 2016 (e) 2017 (f) 2018

Figure 10.4: Result of the simultaneous fit over the data-taking years of $3\pi\mu$ SS data. Shown are the data (black points), the signal (green), the background (red) and the total PDF (blue).

10.2 $b\bar{b}$ inclusive MC

In order to investigate the presence of possible sources of background, the selection procedure is applied on a $b\bar{b}$ inclusive MC. Since large inclusive MC's are not available in LHCb for the time being, an alternative strategy is used to select B decays among already available MC samples. The stripping selection is applied on a large number of MC samples of B meson decays. The samples having a stripping selection efficiency significantly higher than the average value are removed, since this is an indication of the fact that those particular decay modes are being selected by the stripping requirements. On the other hand, selection efficiencies around or below an approximately constant value are indicative of the fact that the stripping is randomly selecting the other B meson produced in the event. Only these samples are retained, and merged together in order to obtain a generic $b\bar{b}$ MC. Fig. 10.5 shows the $3\pi 3\pi$ and $3\pi\mu$ stripping selection efficiencies on the subsamples used to obtain the $b\bar{b}$ MC. For the $3\pi\mu$ stripping line, samples with stripping efficiency greater than 15% are removed. The full selection procedure



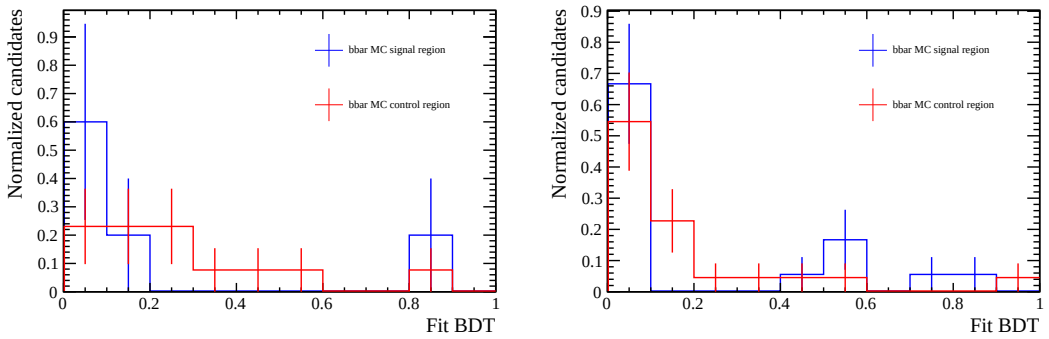
(a) (b)

Figure 10.5: $3\pi 3\pi$ (a) and $3\pi\mu$ (b) stripping selection efficiencies for the subsamples used to obtain the generic $b\bar{b}$ MC. The dashed line in the $3\pi\mu$ case indicates the value of the efficiency above which samples are discarded.

is applied on the remaining events and a summary of the selection efficiencies is reported in Table 10.4. The distribution of the fitBDT after the full selection is shown in Fig 10.6. The remaining statistics is not sufficiently large to come to a clear conclusion, however, as it is expected, the majority of the surviving events are concentrated in the first bin of the BDT spectrum. Moreover, the fitBDT distributions in the signal and control region are in good agreement.

Table 10.4: Statistics and selection efficiencies for the generic $b\bar{b}$ MC.

Generic $b\bar{b}$ MC	$3\pi3\pi$	$3\pi\mu$
Events after stripping	36577	43289
Events after selection in signal region	5	18
Selection efficiency in signal region (10^{-4})	1.37 ± 0.61	4.16 ± 0.98
Events after selection in control region	13	22
Selection efficiency in control region (10^{-4})	3.55 ± 0.99	5.08 ± 1.08



(a) (b)

Figure 10.6: FitBDT distribution for the generic $b\bar{b}$ MC after full $3\pi3\pi$ (a) and $3\pi\mu$ (b) selections.

Conclusions and prospects

The search for the rare $B^0 \to K^{*0}\tau^+\tau^-$ decay described in this thesis is performed on 9 fb^{-1} of data collected at the LHCb experiment. The decay is reconstructed either in the hadronic final state, in which the two τ leptons decay into pions, or in the mixed hadronic-leptonic final state, in which one τ decays into pions and the other into a muon. Several selection steps are performed, some of them making use of BDTs, in order to reduce the amount of background. The efficiency of the selection is evaluated using MC simulations, and systematic uncertainties and/or corrections due to effects not well reproduced by simulations are being evaluated. Since the reconstructed mass offers poor discriminating power, a binned maximum likelihood fit on the output of a BDT classifier is performed to extract the number of signal events in data. The data distribution is still blind, however an expected upper limit in the case where no signal is observed is computed to be $3.2(2.7) \cdot 10^{-4}$ at 95(90)% CL, which represents an improvement of one order of magnitude with respect to the current experimental limit.
Fig. 10.7 shows the expected evolution of the upper limit as a function of the increase in recorded luminosity at LHCb in the next years of data-taking. Two scenarios are considered: the first is a conservative one in which the upper limit obtained with the current dataset is scaled by the expected increase in luminosity, while the second is a more optimistic one in which, in addition to the increase in luminosity, a few improvements are considered. The following hypotheses are made in order to obtain the second scenario:

- from Run 3 (2022-2025) an additional 40% increase in the available statistics is assumed since the L0 trigger will be removed;
- from Run 3 the available statistics is increased of a factor 2, due to improvements in analysis techniques (*e.g.* more sophisticated MVA algorithms);
- from Run 5 (2032-2035) an additional 40% increase in the available statistics is assumed since tracking chambers inside the magnet will be installed and will contribute to the detection of slow pions.

Over the long term the upper limit is expected to be reduced by another order of magnitude, which could help discarding more new physics models predicting LFU violation. However, the sensitivity of the present machine is far too low to observe a branching ratio of the order of the SM prediction, hence more powerful particles accelerators (FCC [105], CLIC [106], ...) would be needed to reach the necessary sensitivity (in the case where no new physics contribution is present).

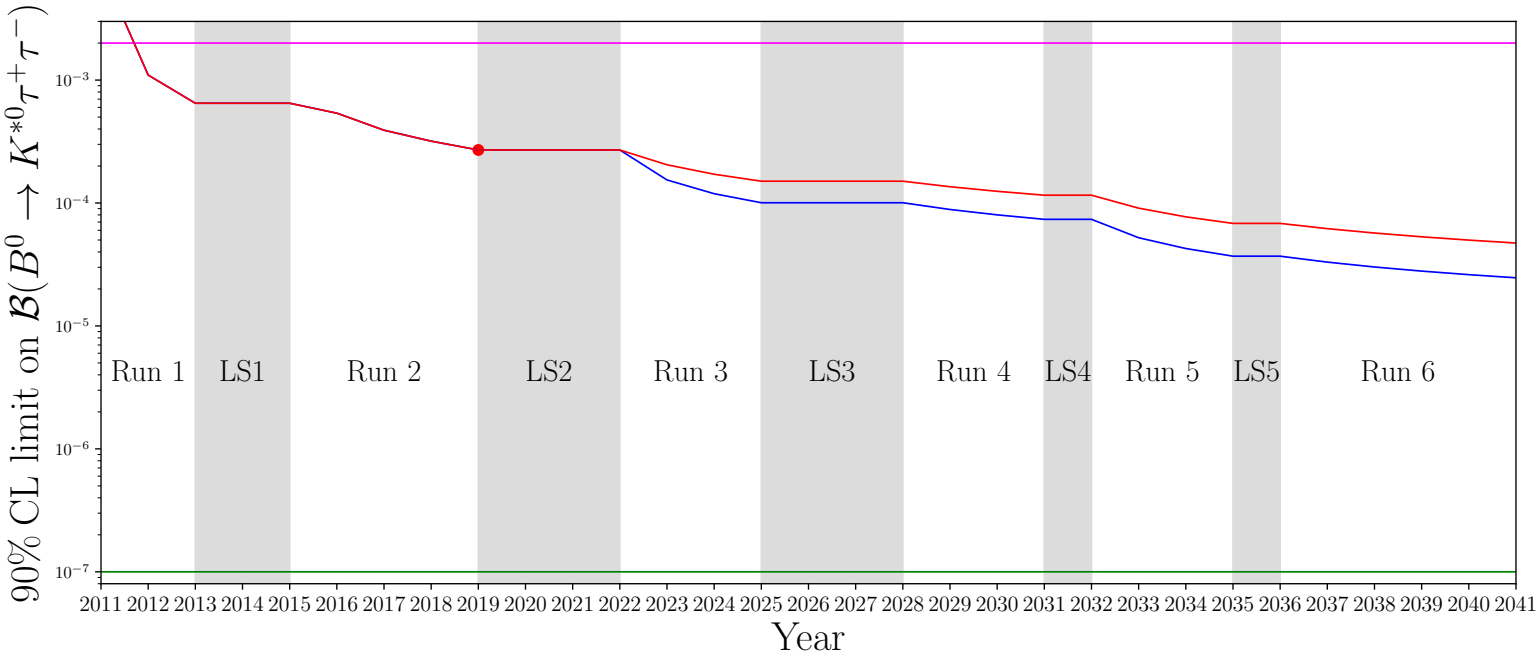


Figure 10.7: Expected evolution of the upper limit in the case where no signal is observed as a function of the data-taking years. The red solid dot indicates the expected upper limit obtained with the present dataset, while the red and blue lines represent the evolution of the limit in the conservative and optimistic scenarios, respectively. The violet line indicates the present experimental upper limit while the green line represents the SM prediction. The upper limit is computed at 90% CL in order to compare it to the present experimental result.

Appendices

Appendix A

Dalitz plane strategy

The τ 3-prong decay proceeds predominantly through the intermediate $\rho^0(770)$ resonance, therefore $\tau^{\pm} \to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu_{\tau}$ events form a "plus" shape in the Dalitz plane, as illustrated in Fig. A.1. Using two cuts on the two invariant masses, the

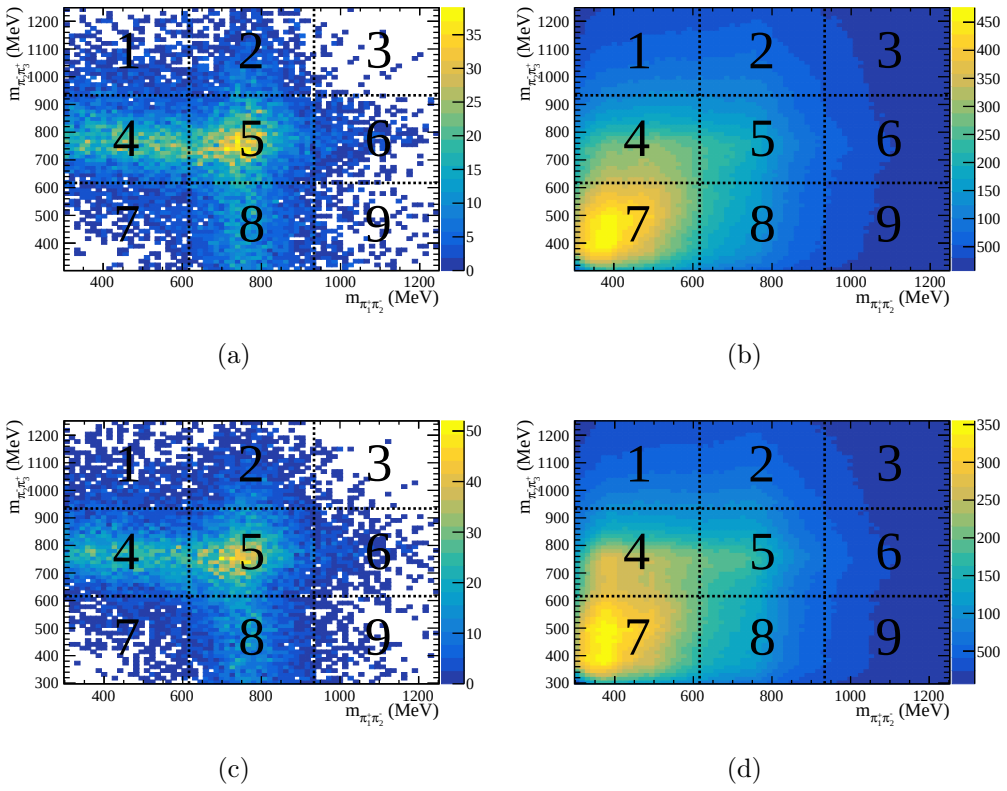


Figure A.1: Dalitz plane for simulated $B^0 \to K^{*0}\tau^+\tau^-$ $3\pi3\pi$ events (a), 2016 $3\pi3\pi$ data (b), simulated $B^0 \to K^{*0}\tau^+\tau^-$ $3\pi\mu$ events (c), 2016 $3\pi\mu$ data (d) after the stripping selection. The MC samples used to produce these plots do not include the neutral pion component.

Dalitz plot is divided into nine boxes. To simplify the notation, these boxes are referred to by their position on a classic dial pad. As an example, box "5" is defined

as

$$\text{box 5} = \left(m_{\text{low}} \leq m_{\pi_1^+\pi_2^-} \leq m_{\text{high}}\right) \text{ and } \left(m_{\text{low}} \leq m_{\pi_2^-\pi_3^+} \leq m_{\text{high}}\right), \tag{A.1}$$

for suitably defined values of m_{low} and m_{high}. The data is then split into four mutually exclusive subsamples[1]:

signal region ($\tau^{\pm}$ in box 5) [$\tau^{\mp}$ in box 5]:
the τ candidates are required to lie inside the middle box, denoted as "5" in the figure. These events are the most signal-like;

background region ($\tau^{\pm}$ in boxes 1, 3, 7 or 9):
at least one τ candidate must end up in one of the four corners of the Dalitz plot. These events are the least signal-like, and hence can be used as a proxy for the background;

control region ($\tau^{\pm}$ in boxes 2 or 6) [$\tau^{\mp}$ in boxes 2, 4, 5, 6 or 8]:
the events in this region are also used to represent the background inside the signal region;

leftover region ($\tau^{\pm}$ in boxes 4 or 8) [$\tau^{\mp}$ in boxes 4, 5 or 8]:
these events make up the remainder of the full data sample. They are considered as an alternative to the control region.

The trigger, stripping and cut-based steps of the selection are the same as those of the default analysis strategy involving the K^{*0} mass. Moreover, a BDT-based selection is implemented in order to suppress most of the background, in this case involving only one BDT (the possibility of adding a second BDT is not investigated). The BDT is trained on simulated MC signal events (without neutral pion component) and on data from the background region of the Dalitz plane. After the full selection, the fitBDT is trained exploiting the remaining discriminating power. The leftover region is used to extract the background model for the fit, since it is the most similar to the signal region, and the sensitivity is estimated using the CL_s method.

A.1 $3\pi3\pi$ final state

The variables used to train the selection BDT are chosen using the iterative procedure presented in Sec. 9.1. Moreover the variables related to τ candidates are symmetrized using the minimum and maximum functions. They are:

- K^{*0} mass;

[1]The conditions in square brackets apply for the $3\pi3\pi$ final state only.

- min(τ^+ flight distance from refitting procedure, τ^- flight distance from refitting procedure);
- max(τ^+ flight distance from refitting procedure, τ^- flight distance from refitting procedure;
- min(τ^+ error on flight distance from refitting procedure, τ^- error on flight distance from refitting procedure);
- max(τ^+ error on flight distance from refitting procedure, τ^- error on flight distance from refitting procedure);
- B^0 visible mass;
- B^0 analytically reconstructed mass using τ "+" solutions (`B_M_aa`);
- invariant mass of π's from τ;
- B^0 neutral cone isolation vPT;
- min(τ^+ DeltaChi2MassOneTrack, τ^- DeltaChi2MassOneTrack);
- max(τ^+ DeltaChi2MassOneTrack, τ^- DeltaChi2MassOneTrack);
- B^0 maximum DOCA of decay products;
- min(τ^+ y component of analytically reconstructed momentum from "+" solution, τ^- y component of analytically reconstructed momentum from "+" solution);
- max(τ^+ y component of analytically reconstructed momentum from "+" solution, τ^- y component of analytically reconstructed momentum from "+" solution).

The variable rankings are similar among the data-taking years, therefore only the one for 2016 MC and background data is shown in Table A.1. Figure A.2 shows the 2016 variable distributions, while the flattened BDT distributions for signal and background samples are reported in Fig. A.3.

The chosen working point for the MVA selection is:

$$\text{selection BDT} > 0.4. \tag{A.2}$$

The working point is not optimized.

The total selection efficiencies are reported in Table A.2. Table A.3 gathers the input parameters for the fit and the statistics after the full selection.

Table A.1: Selection BDT variables ranking for the 2016 $3\pi3\pi$ final state, using the Dalitz plane strategy.

Rank	Name	Separation
1	B_OPT_tau_decayLengthMax	0.393
2	B_OPT_tau_decayLengthMin	0.3216
3	Kst_M	0.278
4	B_M	0.2059
5	tau_DeltaChi2MassOneTrackMin	0.1823
6	B_NC_VPTTAUMin	0.1501
7	B_NC_VPTTAUMax	0.1407
8	B_Mtautau	0.1401
9	B_M_aa	0.08985
10	tau_PY_aMax	0.06313
11	tau_DeltaChi2MassOneTrackMax	0.06286
12	tau_PY_aMin	0.06239
13	B_DOCAMAX	0.02707
14	B_OPT_nIter	0.02071
15	B_OPT_tau_decayLengthErrMin	0.01868
16	B_OPT_tau_decayLengthErrMax	0.01011

Table A.2: Selection efficiency for the $3\pi3\pi$ final state, for strategy involving the Dalitz plane.

Year	Efficiency ($\times10^{-6}$)
2011	1.40 ± 0.10
2012	1.21 ± 0.10
2015	2.10 ± 0.20
2016	2.63 ± 0.12
2017	2.88 ± 0.10
2018	2.29 ± 0.10

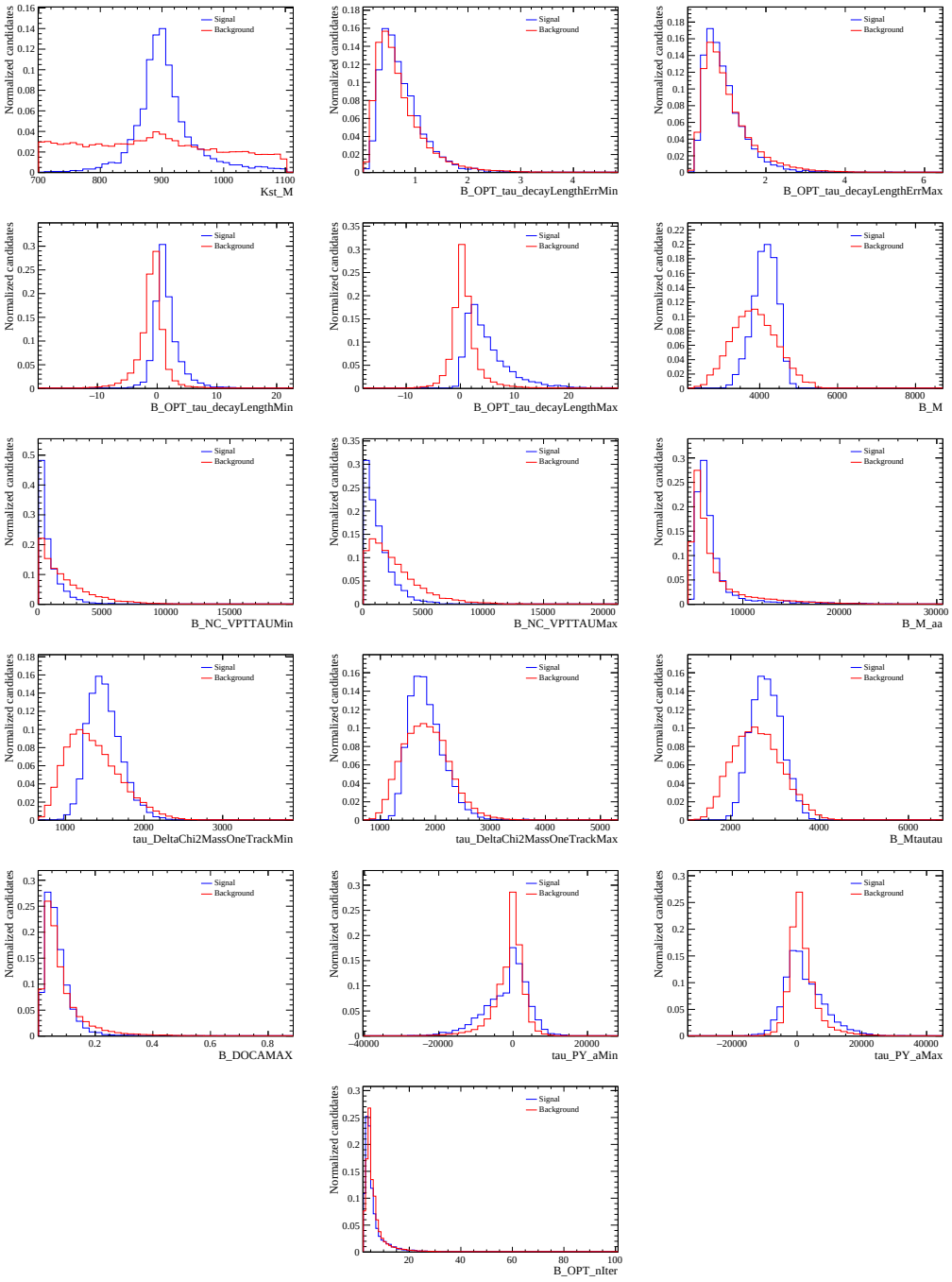
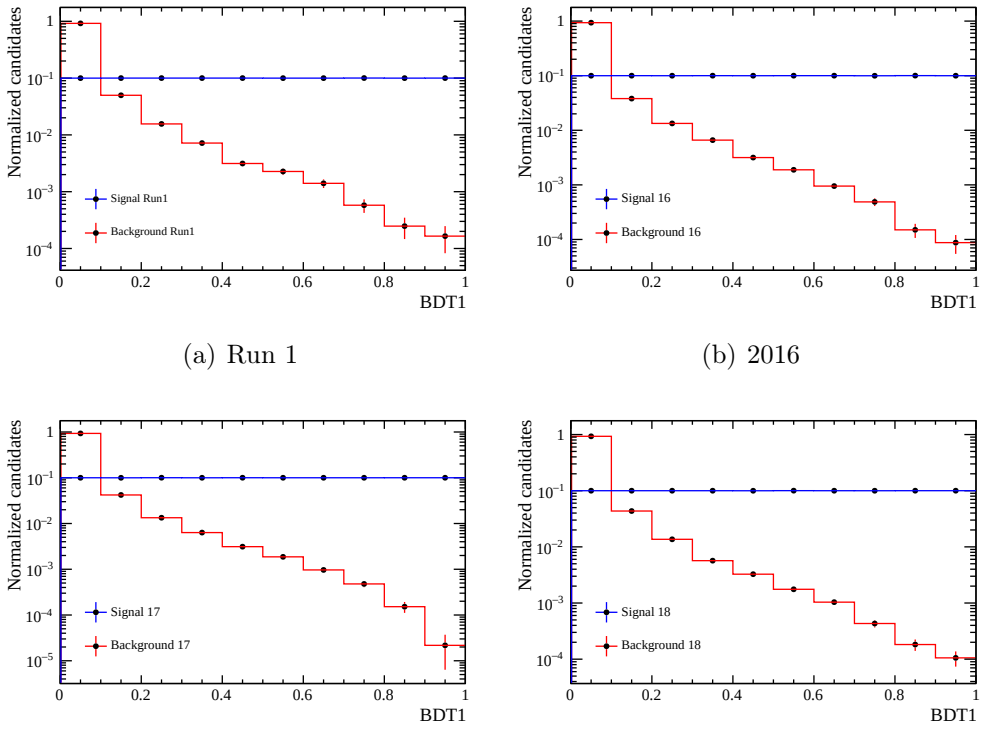


Figure A.2: Selection BDT variable distributions for the 2016 $3\pi3\pi$ final state, using the Dalitz plane strategy.



(a) Run 1 (b) 2016

(c) 2017 (d) 2018

Figure A.3: Selection BDT distribution for $3\pi3\pi$ signal and background training samples, using the Dalitz plane strategy. The flat transformation is applied.

Table A.3: Values of the input parameters for the likelihood fit for the $3\pi3\pi$ final state, using the Dalitz plane strategy. Shown are the normalization factor α_y, its inverse f_y, the ratio of efficiencies in leftover and signal region $\epsilon_y^{\text{lfv}}/\epsilon_y^{\text{sig}}$ and the number of events in data leftover region. Moreover, reported are also the number of events in the data signal region with fitBDT < 0.7 n_y^{blind} and the number of events in the data leftover region with fitBDT < 0.7 $n_y^{\text{lfv, blind}}$.

Year	α_y (10^{-5})	f_y	$\epsilon_y^{\text{lfv}}/\epsilon_y^{\text{sig}}$	n_y^{lfv}	n_y^{blind}	$n_y^{\text{lfv, blind}}$
2011	30.8 ± 4.3	3241 ± 453	6.13 ± 0.48	4632	462	4571
2012	15.0 ± 2.0	6637 ± 874	5.92 ± 0.37	9902	997	9765
2015	43.8 ± 6.7	2284 ± 352	5.89 ± 0.59	3501	379	3464
2016	6.23 ± 0.79	16040 ± 2025	6.16 ± 0.31	31609	3364	31280
2017	5.56 ± 0.68	17997 ± 2190	5.90 ± 0.20	35485	3770	35129
2018	6.60 ± 0.80	15117 ± 1836	6.32 ± 0.21	41132	4268	40711

The variables used to train the fitBDT are chosen using the iterative procedure after applying the full selection on the training samples, and are symmetrized for the two τ candidates using the minimum and maximum functions. They are:

- min(τ^+ DeltaChi2OneTrack, τ^- DeltaChi2OneTrack);
- max(τ^+ DeltaChi2OneTrack, τ^- DeltaChi2OneTrack);
- min(τ^+ DeltaChi2TwoTrack, τ^- DeltaChi2TwoTrack);
- max(τ^+ DeltaChi2TwoTrack, τ^- DeltaChi2TwoTrack);
- K^{*0} DeltaChi2MassTwoTrack;
- min(τ^+ IsoBDTSecondValue, τ^- IsoBDTSecondValue);
- max(τ^+ IsoBDTSecondValue, τ^- IsoBDTSecondValue);
- K (from K^{*0}) IsoBDTSecondValue;
- min(τ^+ VtxIsoNumVtx, τ^- VtxIsoNumVtx);
- max(τ^+ VtxIsoNumVtx, τ^- VtxIsoNumVtx);
- min(τ^+ decay vertex χ^2, τ^- decay vertex χ^2);
- max(τ^+ decay vertex χ^2, τ^- decay vertex χ^2);
- K^{*0} mass;
- B^0 analytically reconstructed mass using "+" solutions (`B_M_aa`);
- B^0 analytically reconstructed mass using "optimal" solution (`B_M_opt`);
- B^0 neutral cone isolation $\mathrm{P_{asym}}$;
- B^0 maximum DOCA of decay products;
- DOCA of π and K (from K^{*0});
- min(τ^+ y component of analytically reconstructed momentum from "-" solution (`tauP_PY_b`), τ^- y component of analytically reconstructed momentum from "-" solution (`tauM_PY_b`));
- max(τ^+ y component of analytically reconstructed momentum from "-" solution (`tauP_PY_b`), τ^- y component of analytically reconstructed momentum from "-" solution (`tauM_PY_b`));
- abs(τ^+ decay vertex z - τ^- decay vertex z);

- selection BDT.

Similarly to the selection BDT, only the variable ranking for 2016 MC and background data is shown in Table A.4. Figures A.4 and A.5 show the 2016 variable distributions, while the flattened BDT distributions for signal and background samples are reported in Fig. A.6.

Table A.4: FitBDT variables ranking for the 2016 $3\pi3\pi$ final state, using the Dalitz plane strategy.

Rank	Name	Separation
1	flat_First3piBDT	0.2004
2	tau_DeltaChi2OneTrackMax	0.1421
3	tau_DeltaChi2TwoTrackMax	0.1207
4	tau_DeltaChi2TwoTrackMin	0.09695
5	tau_DeltaChi2OneTrackMin	0.08275
6	tau_IsoBDTSecondValueMin	0.07906
7	K_IsoBDTSecondValue	0.07403
8	tau_ENDVERTEX_CHI2Max	0.07117
9	tau_IsoBDTSecondValueMax	0.0699
10	tau_ENDVERTEX_CHI2Min	0.05351
11	tau_P_bMin	0.03447
12	Kst_DeltaChi2MassTwoTrack	0.03153
13	tau_VtxIsoNumVtxMax	0.03026
14	tau_P_bMax	0.02444
15	Kst_DOCAKpi	0.02329
16	B_M_opt	0.02241
17	tau_VtxIsoNumVtxMin	0.02136
18	B_NC_PASYM	0.02118
19	tau_VertexDeltaZ	0.01538
20	B_M_aa	0.01273
21	Kst_M	0.009676
22	B_DOCAMAX	0.008236

The comparison between the fitBDT distribution in the signal and leftover regions is shown in Fig. A.7. For all but one data-taking years, a trend is present in the pull of the distributions, which suggests the presence of a correlation between the fitBDT and the invariant masses used to define the Dalitz plane.

The expected upper limit in the case where no signal is observed is computed using the CL_s method:

$$\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-) < 2.62 \cdot 10^{-3} \text{ at } 95\% \text{ CL}, \tag{A.3}$$

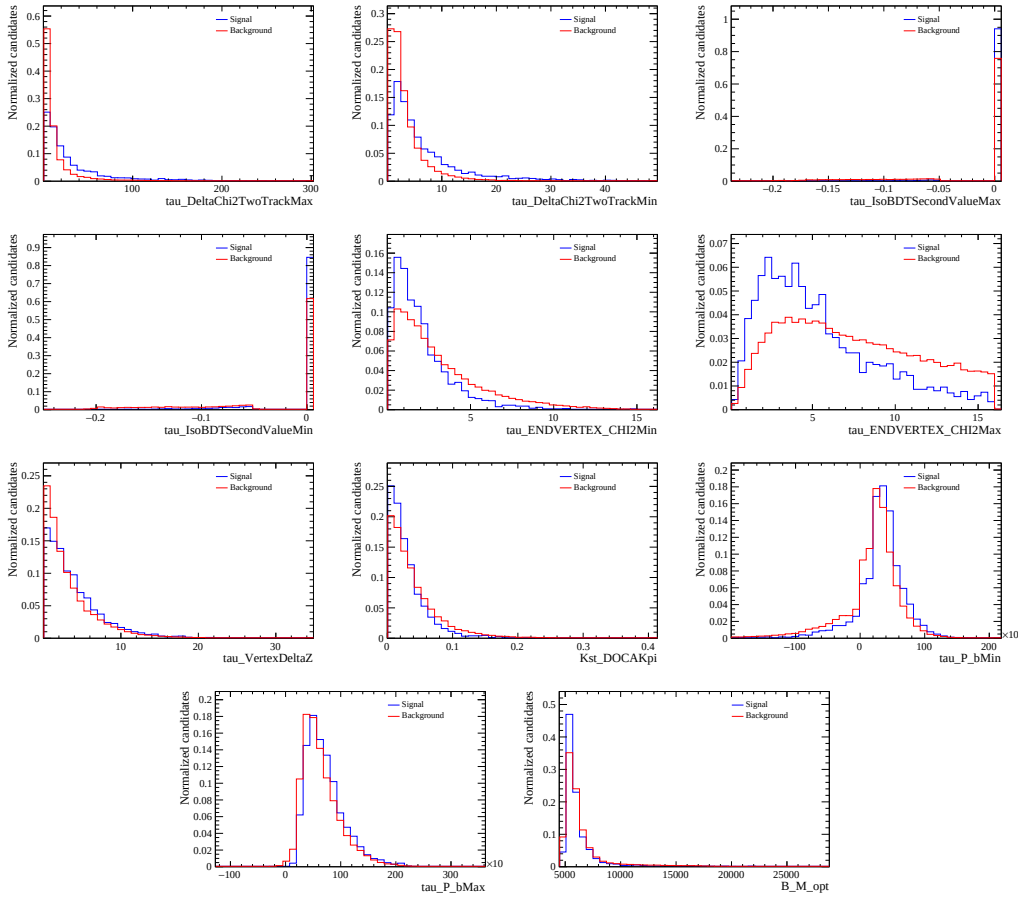


Figure A.4: FitBDT variable distributions for the 2016 $3\pi3\pi$ final state, using the Dalitz plane strategy (part 1).

where the above result takes into account only the statistical uncertainty. The upper limit is ~ 2.5 times higher than the one obtained with the default K^{*0} mass strategy shown in Table 8.11. For this reason, and because of the fact that the fitBDT background distribution is not well described by the events in the leftover region, the Dalitz plane strategy is discarded.

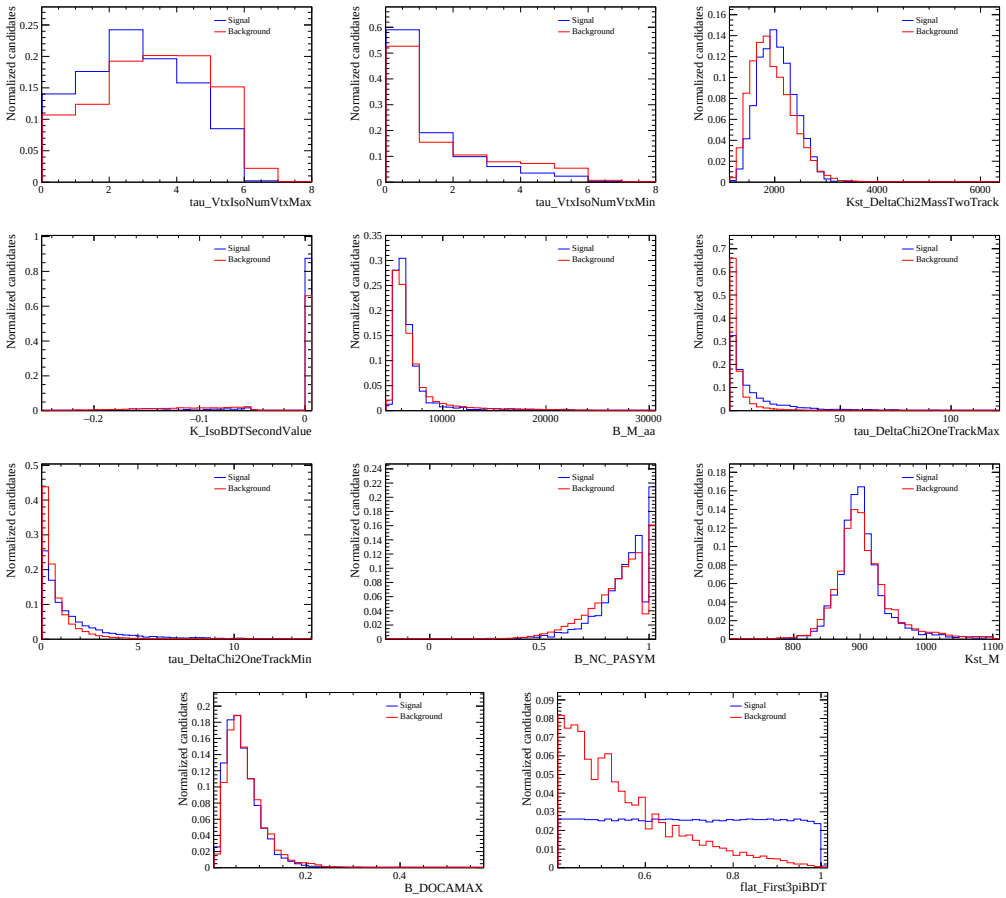
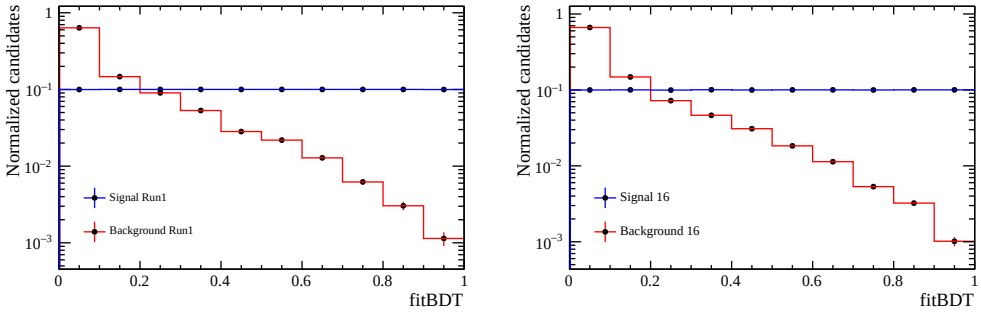
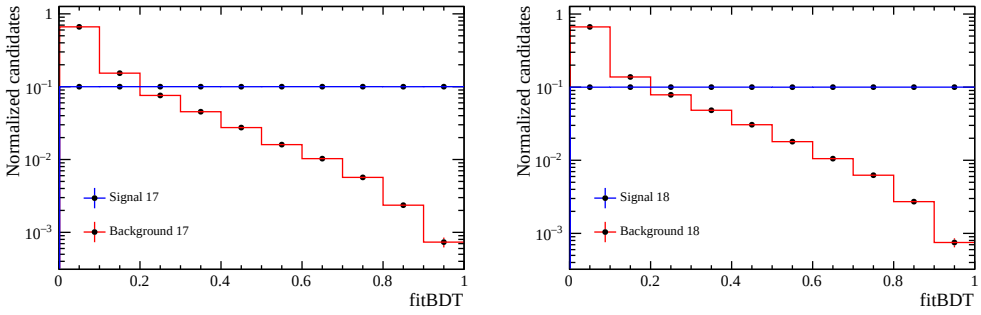


Figure A.5: FitBDT variable distributions for the 2016 $3\pi3\pi$ final state, using the Dalitz plane strategy (part 2).

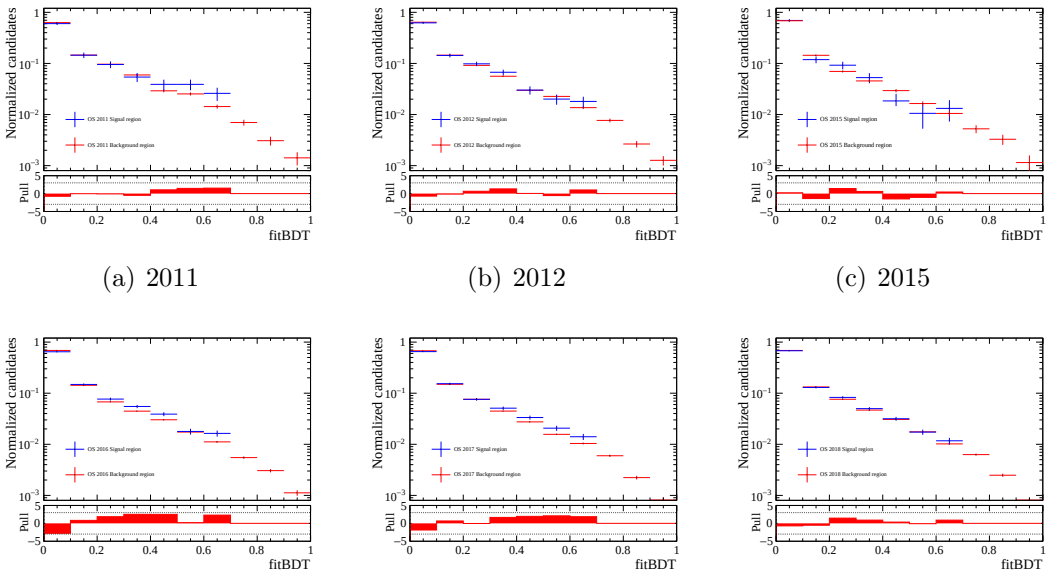


(a) Run 1 (b) 2016



(c) 2017 (d) 2018

Figure A.6: FitBDT distribution for $3\pi 3\pi$ signal and background training samples, using the Dalitz plane strategy. The flat transformation is applied.



(a) 2011 (b) 2012 (c) 2015

(d) 2016 (e) 2017 (f) 2018

Figure A.7: Comparison of the $3\pi3\pi$ fitBDT distribution in signal and leftover regions in logarithmic scale, using the Dalitz plane strategy. The distribution in the signal region is shown for values of fitBDT < 0.7. The dashed lines in the pull correspond to the $\pm3\sigma$ difference.

A.2 $3\pi\mu$ final state

The variables used to train the selection BDT are chosen using the iterative procedure presented in Sec. 9.1. They are:

- τ^+ ($\to \pi\pi\pi$) analytically reconstructed momentum using "-" solution (`tauP_P_b`);
- μ neutral cone isolation $\mathrm{PT}_{\mathrm{asym}}$;
- K^{*0} DeltaChi2MassTwoTrack;
- K^{*0} DeltaChi2MassOneTrack;
- τ^+ DeltaChi2OneTrack;
- μ p_{T};
- K^{*0} flight distance[2];
- K^{*0} IsoBDTThirdValue;
- π (from K^{*0}) IsoBDTThirdValue;
- K (from K^{*0}) IsoBDTThirdValue;
- τ^+ decay vertex χ^2;
- τ^- ($\to \mu\nu\nu$) analytically reconstructed transverse momentum using "+" solution (tauM_PT_a).

The variable rankings are similar among the data-taking years, therefore only the one for 2016 MC and background data is shown in Table A.5. Figure A.8 shows the 2016 variable distributions, while the flattened BDT distributions for signal and background samples are reported in Fig. A.9.

The chosen working point for the MVA selection is:

$$\text{selection BDT} > 0.5. \tag{A.4}$$

The working point is not optimized.

The total selection efficiencies are reported in Table A.6. Table A.7 gathers the input parameters for the fit and the statistics after the full selection.

[2]The B^0 and K^{*0} decay vertices are distinct because of resolution effects.

Table A.5: Selection BDT variables ranking for the 2016 $3\pi3\pi$ final state, using the Dalitz plane strategy.

Rank	Name	Separation
1	tauP_P_b	0.1335
2	B_NC_PTASYMMUON	0.1001
3	Kst_DeltaChi2MassTwoTrack	0.09145
4	Kst_DeltaChi2MassOneTrack	0.08535
5	tauP_DeltaChi2OneTrack	0.08444
6	mu_PT	0.05847
7	Kst_FD_ORIVX	0.05474
8	Kst_IsoBDTThirdValue	0.03228
9	pi_IsoBDTThirdValue	0.02705
10	K_IsoBDTThirdValue	0.0256
11	tauP_ENDVERTEX_CHI2	0.02425
12	tauM_P_tr_a	0.01079
13	B_BKSTTAUTAUCDFISO	0.01033

Table A.6: Selection efficiency for the $3\pi\mu$ final state, for strategy involving the Dalitz plane.

Year	Efficiency ($\times10^{-5}$)
2011	3.39 ± 0.11
2012	3.12 ± 0.10
2015	2.79 ± 0.16
2016	4.70 ± 0.11
2017	5.37 ± 0.10
2018	4.82 ± 0.10

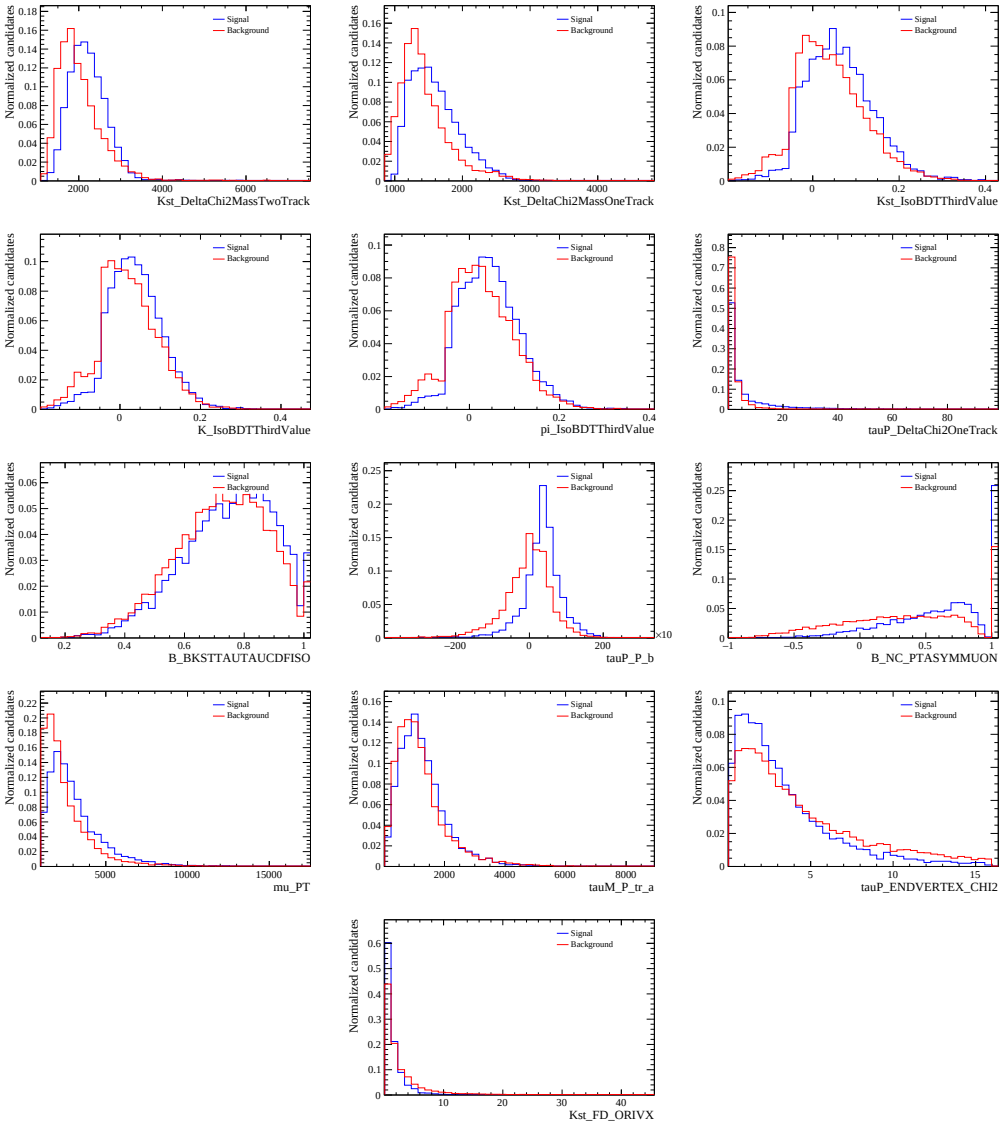
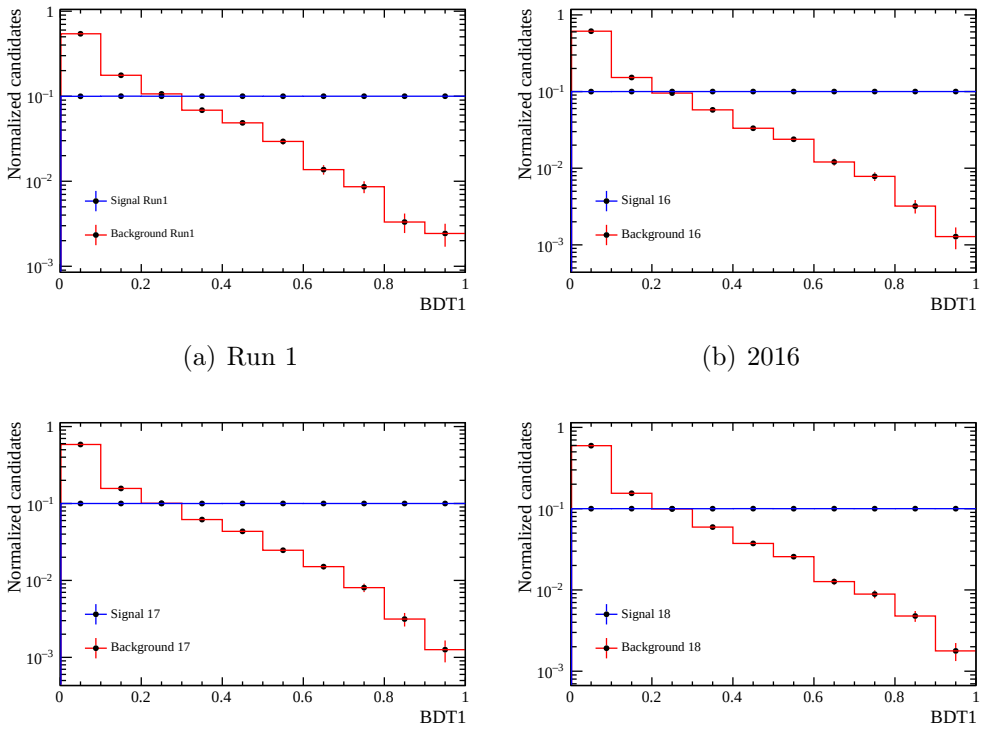


Figure A.8: 2016 $3\pi\mu$ final state BDT1 variable distributions using the Dalitz plane strategy.



(a) Run 1 (b) 2016

(c) 2017 (d) 2018

Figure A.9: Selection BDT distribution for $3\pi\mu$ signal and background training samples, using the Dalitz plane strategy. The flat transformation is applied.

Table A.7: Values of the input parameters for the likelihood fit for the $3\pi\mu$ final state, using the Dalitz plane strategy. Shown are the normalization factor α_y, its inverse f_y, the ratio of efficiencies in leftover and signal region $\epsilon_y^{\text{lfv}}/\epsilon_y^{\text{sig}}$ and the number of events in data leftover region. Moreover, reported are also the number of events in the data signal region with fitBDT < 0.7 n_y^{blind} and the number of events in the data leftover region with fitBDT < 0.7 $n_y^{\text{lfv, blind}}$.

Year	α_y (10^{-6})	f_y	$\epsilon_y^{\text{lfv}}/\epsilon_y^{\text{sig}}$	n_y^{lfv}	n_y^{blind}	$n_y^{\text{lfv, blind}}$
2011	5.09 ± 0.63	196574 ± 24184	0.953 ± 0.042	12592	8606	11720
2012	2.33 ± 0.28	429158 ± 51125	0.970 ± 0.032	28665	19794	26756
2015	13.2 ± 1.7	76020 ± 10088	0.952 ± 0.075	5526	3700	5355
2016	1.39 ± 0.17	717917 ± 85093	1.012 ± 0.034	60126	41672	57962
2017	1.19 ± 0.14	840889 ± 98488	0.962 ± 0.021	63273	44265	60427
2018	1.25 ± 0.15	796874 ± 93193	0.974 ± 0.019	71139	50383	68013

The variables used to train the fitBDT are chosen using the iterative procedure after applying the full selection on the training samples. They are:

- τ^+ DealtaChi2OneTrack;
- τ^+ decay vertex $z - B^0$ decay vertex z;
- τ^+ neutral cone isolation IT;
- τ^+ neutral cone isolation $\mathrm{P_{asym}}$;
- π and K (from K^{*0}) DOCA;
- K^{*0} flight distance[3] χ^2;
- B^0 decay vertex χ^2;
- μ IsoBDTSecondValue;
- τ^- ($\rightarrow\mu\nu\nu$) z component of analytically reconstructed momentum using "+" solution (`tauM_PZ_a`);
- selection BDT.

Similarly to the selection BDT, only the variable ranking for 2016 MC and background data is shown in Table A.8. Figure A.10 shows the 2016 variable distributions, while the flattened BDT distributions for signal and background samples are reported in Fig. A.11.

Table A.8: FitBDT variables ranking for the 2016 $3\pi\mu$ final state, using the Dalitz plane strategy.

Rank	Name	Separation
1	flat_FirstmuBDT	0.1572
2	tauP_DeltaChi2OneTrack	0.07944
3	B_tauP_DistZ	0.05873
4	B_NC_ITTAUP	0.05151
5	B_NC_PASYMTAUP	0.05139
6	Kst_DOCAKpi	0.03413
7	Kst_FDCHI2_ORIVX	0.02108
8	B_ENDVERTEX_CHI2	0.01415
9	mu_IsoBDTSecondValue	0.008815
10	tauM_PZ_a	0.002582

[3]The B^0 and K^{*0} decay vertices are distinct because of resolution effects.

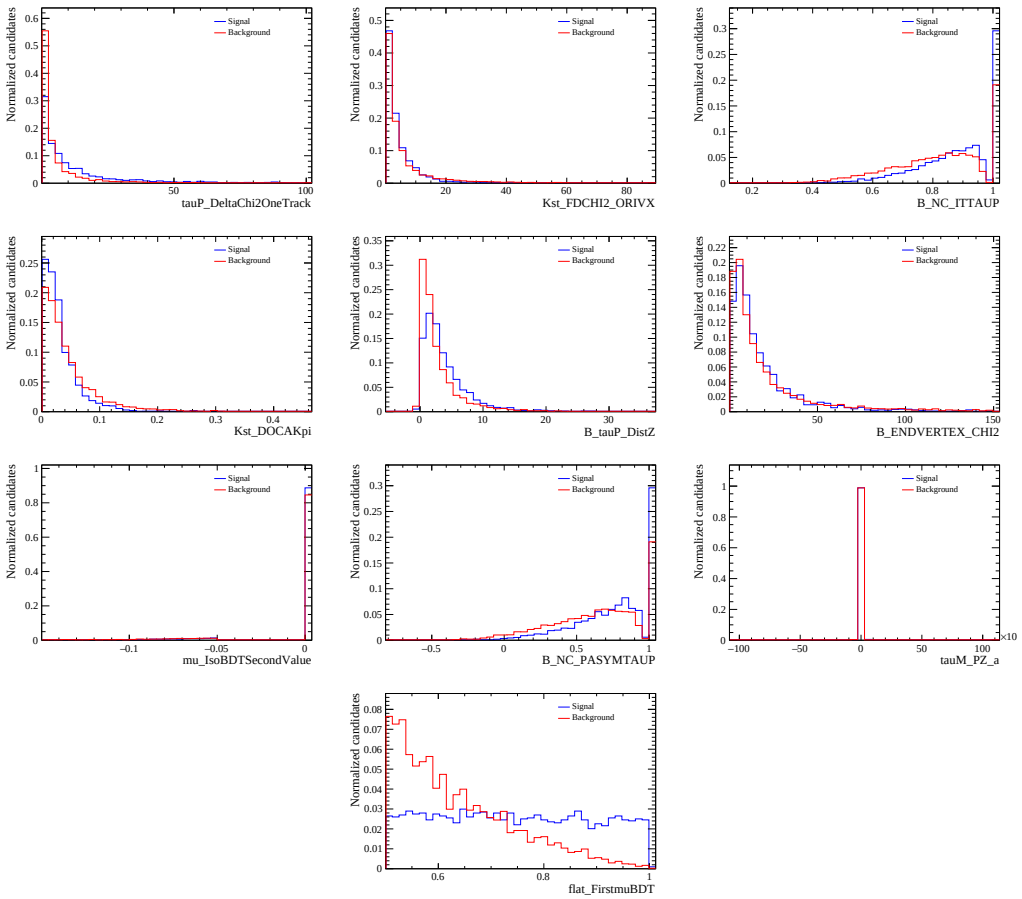


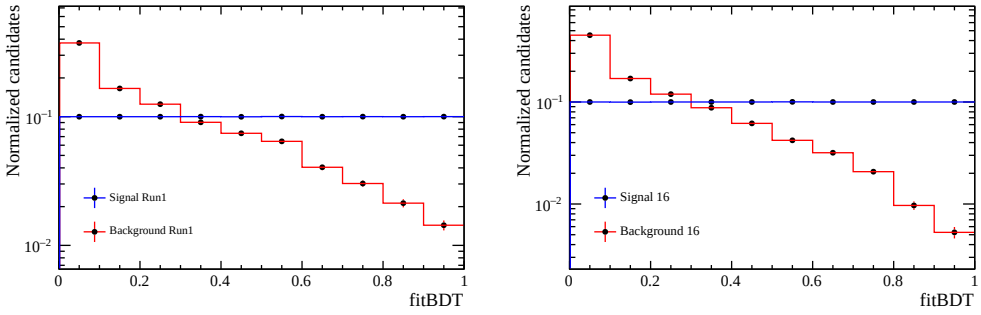
Figure A.10: FitBDT variable distributions for the 2016 $3\pi\mu$ final state, using the Dalitz plane strategy.

The comparison between the fitBDT distribution in the signal and leftover regions is shown in Fig. A.12. Also in this case a trend is present in the pull of the distributions, which suggests the presence of a correlation between the fitBDT and the invariant masses used to define the Dalitz plane.

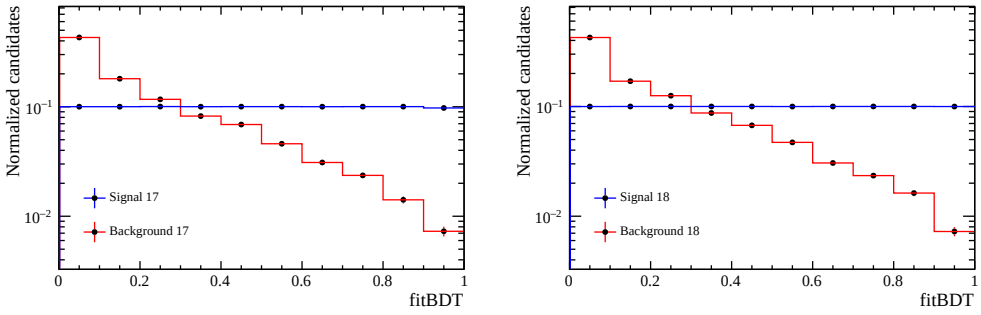
The expected upper limit in the case where no signal is observed is computed using the CL_s method:

$$\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-) < 4.77 \cdot 10^{-4} \text{ at } 95\% \text{ CL}, \tag{A.5}$$

where the above result takes into account only the statistical uncertainty. The upper limit is ~ 1.8 times higher than the one obtained with the default K^{*0} mass strategy shown in Table 8.12. For this reason, and because of the fact that the fitBDT background distribution is not well described by the events in the leftover region, the Dalitz plane strategy is discarded.

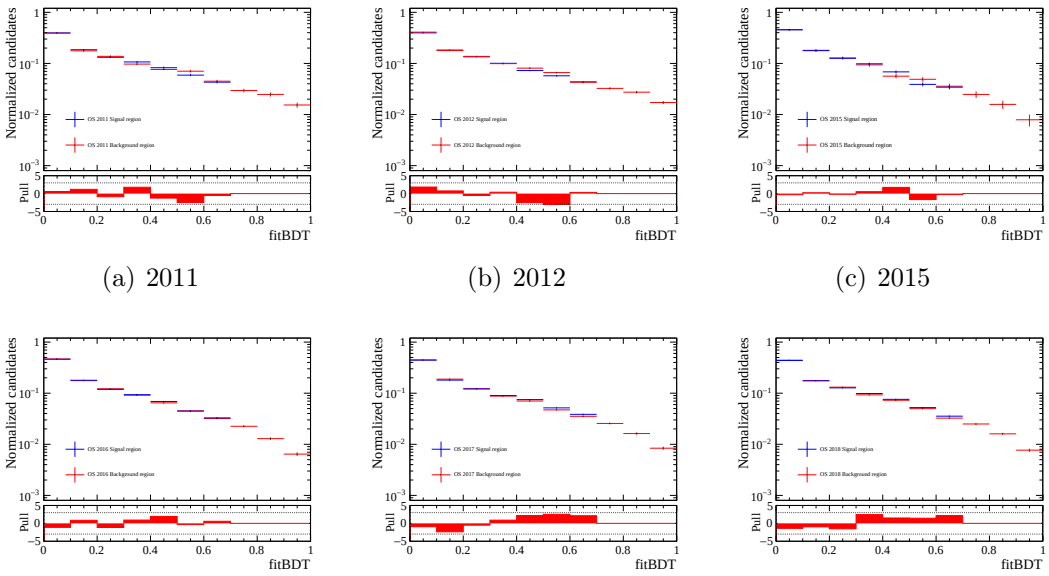


(a) Run 1 (b) 2016



(c) 2017 (d) 2018

Figure A.11: FitBDT distribution for $3\pi\mu$ signal and background training samples, using the Dalitz plane strategy. The flat transformation is applied.



(a) 2011 (b) 2012 (c) 2015

(d) 2016 (e) 2017 (f) 2018

Figure A.12: Comparison of the $3\pi\mu$ fitBDT distribution in signal and leftover regions in logarithmic scale, using the Dalitz plane strategy. The distribution in the signal region is shown for values of fitBDT < 0.7. The dashed lines in the pull correspond to the $\pm 3\sigma$ difference.

Appendix B

BDT complementary information

In this Appendix complementary information on the BDTs used in the analysis is reported.

B.1 $3\pi 3\pi$ final state

BDT1

The BDT1 variable rankings are reported in Tables B.1, B.2, B.3, B.4 for each data-taking period.
The BDT1 variable distributions are reported in Figs. B.1, B.2, B.3, B.4 for each data-taking period.

Table B.1: Run 1 BDT1 variables ranking for the $3\pi 3\pi$ final state.

Rank	Name	Separation
1	B_OPT_tau_decayLengthMin	0.3249
2	tau_FDCHI2_ORIVXMax	0.3141
3	rhoMassMax	0.3118
4	tau_MMax	0.2249
5	B_M	0.1351
6	tau_M12Min	0.06809
7	B_M_opt	0.06603
8	piM_IsoBDTThirdValueMin	0.05827
9	tau_DOCAMAXMax	0.02455

Table B.2: 2016 BDT1 variables ranking for the $3\pi3\pi$ final state.

Rank	Name	Separation
1	rhoMassMax	0.3473
2	B_OPT_tau_decayLengthMin	0.3113
3	tau_FDCHI2_ORIVXMax	0.2636
4	tau_MMax	0.224
5	B_M	0.1777
6	B_NC_PASYM	0.1718
7	tau_M12Min	0.08475
8	B_M_opt	0.06619
9	piM_IsoBDTThirdValueMin	0.06602
10	tau_DOCAMAXMax	0.03923

Table B.3: 2017 BDT1 variables ranking for the $3\pi3\pi$ final state.

Rank	Name	Separation
1	rhoMassMax	0.3455
2	B_OPT_tau_decayLengthMin	0.3163
3	tau_FDCHI2_ORIVXMax	0.2633
4	tau_MMax	0.2294
5	B_M	0.179
6	B_NC_PASYM	0.1748
7	tau_M12Min	0.08621
8	piM_IsoBDTThirdValueMin	0.06196
9	B_M_opt	0.05781
10	tau_DOCAMAXMax	0.02171

Table B.4: 2018 BDT1 variables ranking for the $3\pi3\pi$ final state.

Rank	Name	Separation
1	rhoMassMax	0.3484
2	B_OPT_tau_decayLengthMin	0.2984
3	tau_FDCHI2_ORIVXMax	0.263
4	tau_MMax	0.2178
5	B_M	0.1872
6	B_NC_PASYM	0.1839
7	tau_M12Min	0.09561
8	piM_IsoBDTThirdValueMin	0.06609
9	B_M_opt	0.05201
10	tau_DOCAMAXMax	0.0254

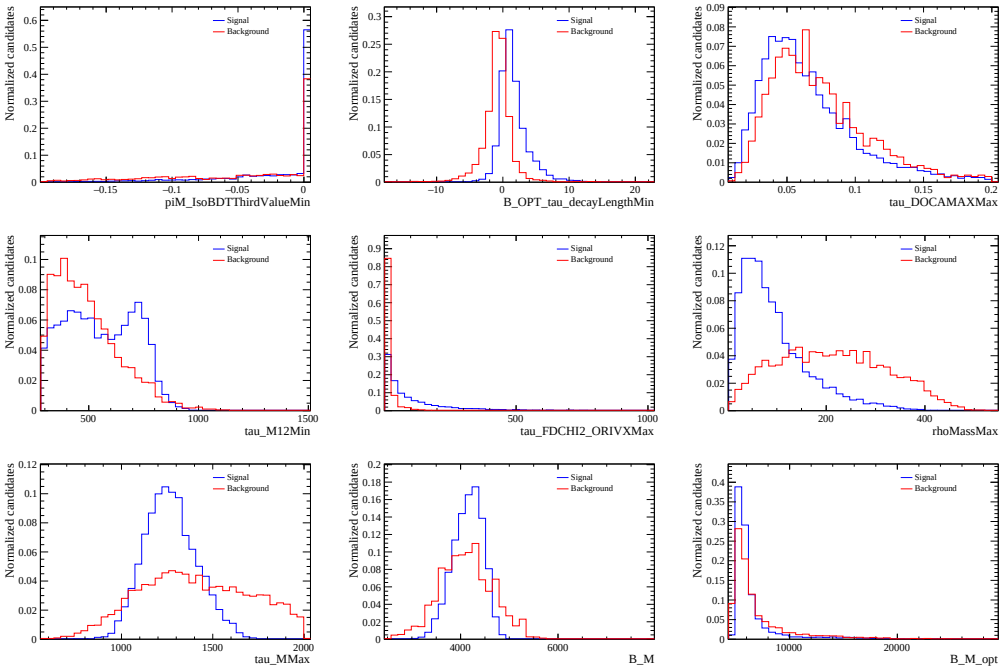


Figure B.1: Run 1 BDT1 variable distributions for the $3\pi 3\pi$ final state.

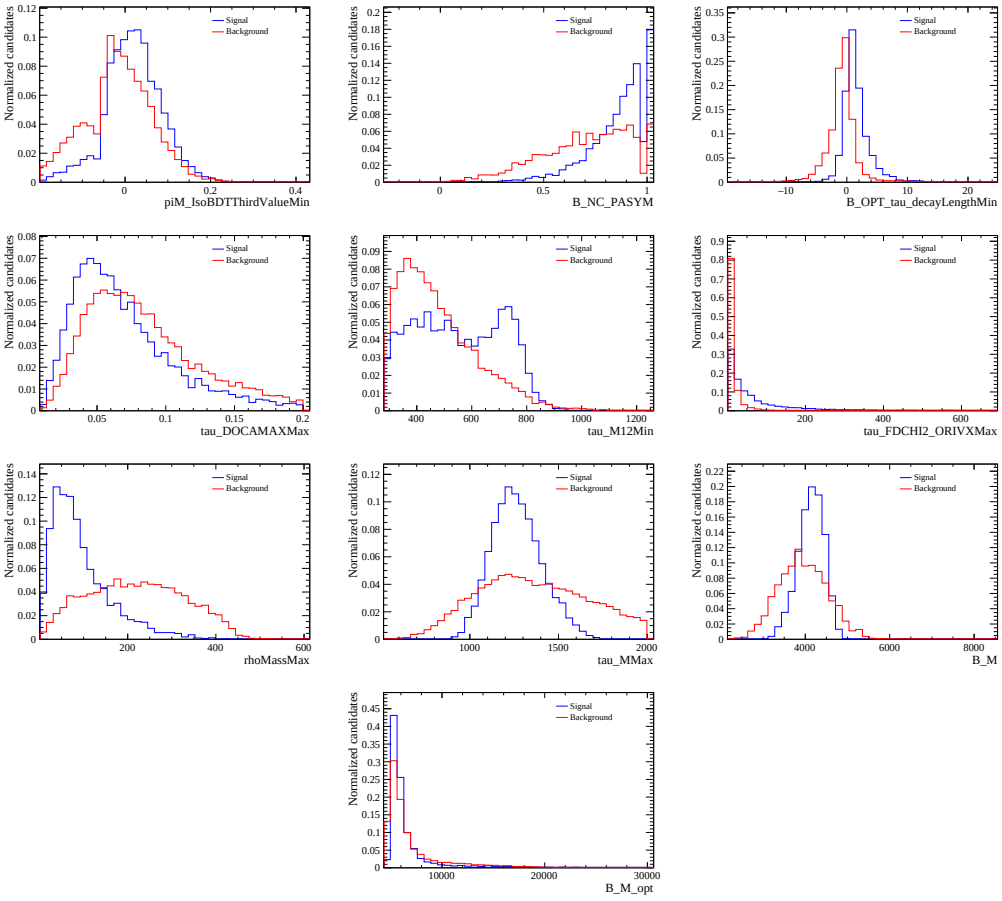


Figure B.2: 2016 BDT1 variable distributions for the $3\pi3\pi$ final state.

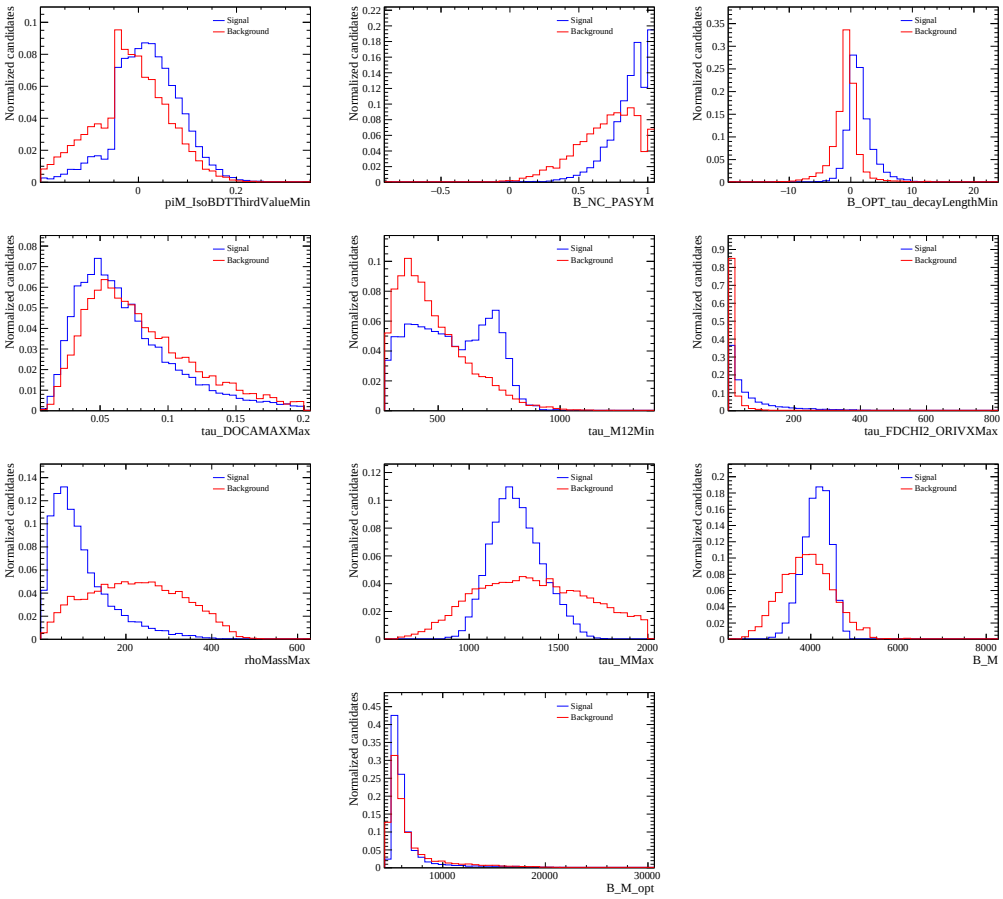


Figure B.3: 2017 BDT1 variable distributions for the $3\pi3\pi$ final state.

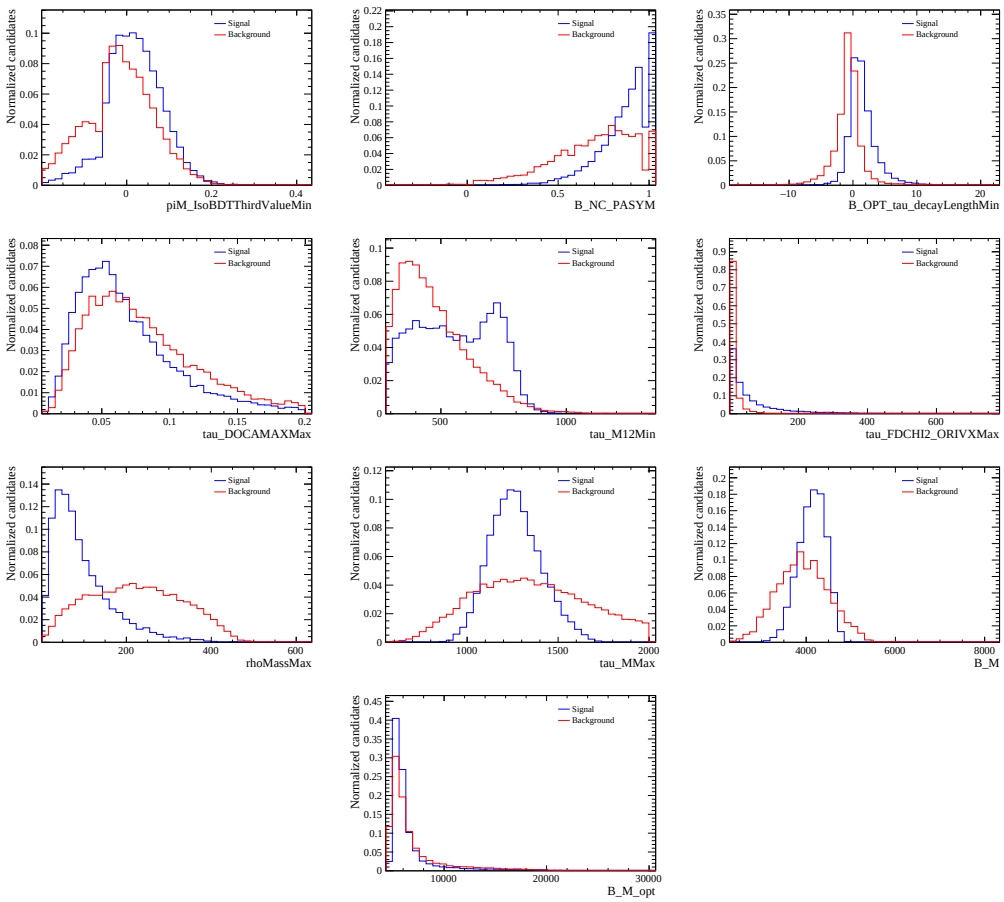


Figure B.4: 2018 BDT1 variable distributions for the $3\pi3\pi$ final state.

BDT2

The BDT2 variable rankings are reported in Tables B.5, B.6, B.7, B.8 for each data-taking period.
The BDT2 variable distributions are reported in Figs. B.5, B.6, B.7, B.8 for each data-taking period.

Table B.5: Run 1 BDT2 variables ranking for the $3\pi3\pi$ final state.

Rank	Name	Separation
1	B_OPT_chi2	0.1747
2	tau_DeltaChi2OneTrackMax	0.1623
3	tau_DeltaChi2TwoTrackMin	0.09527
4	Kst_FD_ORIVX	0.06397
5	tau_ENDVERTEX_CHI2Max	0.04972
6	B_ENDVERTEX_CHI2	0.04075
7	B_DIRA_OWNPV	0.0374
8	K_IsoBDTSecondValue	0.02786
9	pi_IsoBDTSecondValue	0.02669
10	Kst_IsoBDTSecondValue	0.02628
11	tau_DeltaChi2MassOneTrackMin	0.02593
12	Kst_MINIPCHI2	0.02158
13	B_OPT_tau_decayLengthErrMin	0.02043
14	Kst_DeltaChi2OneTrack	0.01829
15	tau_ENDVERTEX_dist3D	0.01349

Table B.6: 2016 BDT2 variables ranking for the $3\pi 3\pi$ final state.

Rank	Name	Separation
1	B_OPT_chi2	0.1668
2	tau_DeltaChi2OneTrackMax	0.122
3	tau_DeltaChi2TwoTrackMin	0.07155
4	Kst_FD_ORIVX	0.0679
5	tau_ENDVERTEX_CHI2Max	0.04859
6	B_ENDVERTEX_CHI2	0.04412
7	K_IsoBDTSecondValue	0.03804
8	Kst_IsoBDTSecondValue	0.03387
9	B_NC_ITTAUMax	0.03357
10	tau_DeltaChi2MassOneTrackMin	0.02576
11	pi_IsoBDTSecondValue	0.02425
12	B_DIRA_OWNPV	0.02277
13	B_OPT_tau_decayLengthErrMin	0.01902
14	tau_ENDVERTEX_dist3D	0.009936
15	Kst_MINIPCHI2	0.009417
16	Kst_DeltaChi2OneTrack	0.008241

Table B.7: 2017 BDT2 variables ranking for the $3\pi 3\pi$ final state.

Rank	Name	Separation
1	B_OPT_chi2	0.1599
2	tau_DeltaChi2OneTrackMax	0.1314
3	Kst_FD_ORIVX	0.08463
4	tau_DeltaChi2TwoTrackMin	0.07398
5	B_ENDVERTEX_CHI2	0.04206
6	tau_ENDVERTEX_CHI2Max	0.03822
7	pi_IsoBDTSecondValue	0.03599
8	B_DIRA_OWNPV	0.03582
9	Kst_IsoBDTSecondValue	0.03553
10	K_IsoBDTSecondValue	0.03274
11	tau_DeltaChi2MassOneTrackMin	0.0235
12	B_NC_ITTAUMax	0.01928
13	tau_ENDVERTEX_dist3D	0.012
14	Kst_DeltaChi2OneTrack	0.01186
15	Kst_MINIPCHI2	0.01166
16	B_OPT_tau_decayLengthErrMin	0.01101

Table B.8: 2018 BDT2 variables ranking for the $3\pi 3\pi$ final state.

Rank	Name	Separation
1	B_OPT_chi2	0.1841
2	tau_DeltaChi2OneTrackMax	0.1223
3	tau_DeltaChi2TwoTrackMin	0.0822
4	Kst_FD_ORIVX	0.07556
5	tau_ENDVERTEX_CHI2Max	0.05581
6	B_ENDVERTEX_CHI2	0.05237
7	B_DIRA_OWNPV	0.02918
8	Kst_IsoBDTSecondValue	0.02897
9	K_IsoBDTSecondValue	0.02798
10	pi_IsoBDTSecondValue	0.02725
11	tau_DeltaChi2MassOneTrackMin	0.02196
12	B_NC_ITTAUMax	0.02191
13	Kst_DeltaChi2OneTrack	0.01388
14	tau_ENDVERTEX_dist3D	0.01298
15	B_OPT_tau_decayLengthErrMin	0.01167
16	Kst_MINIPCHI2	0.008038

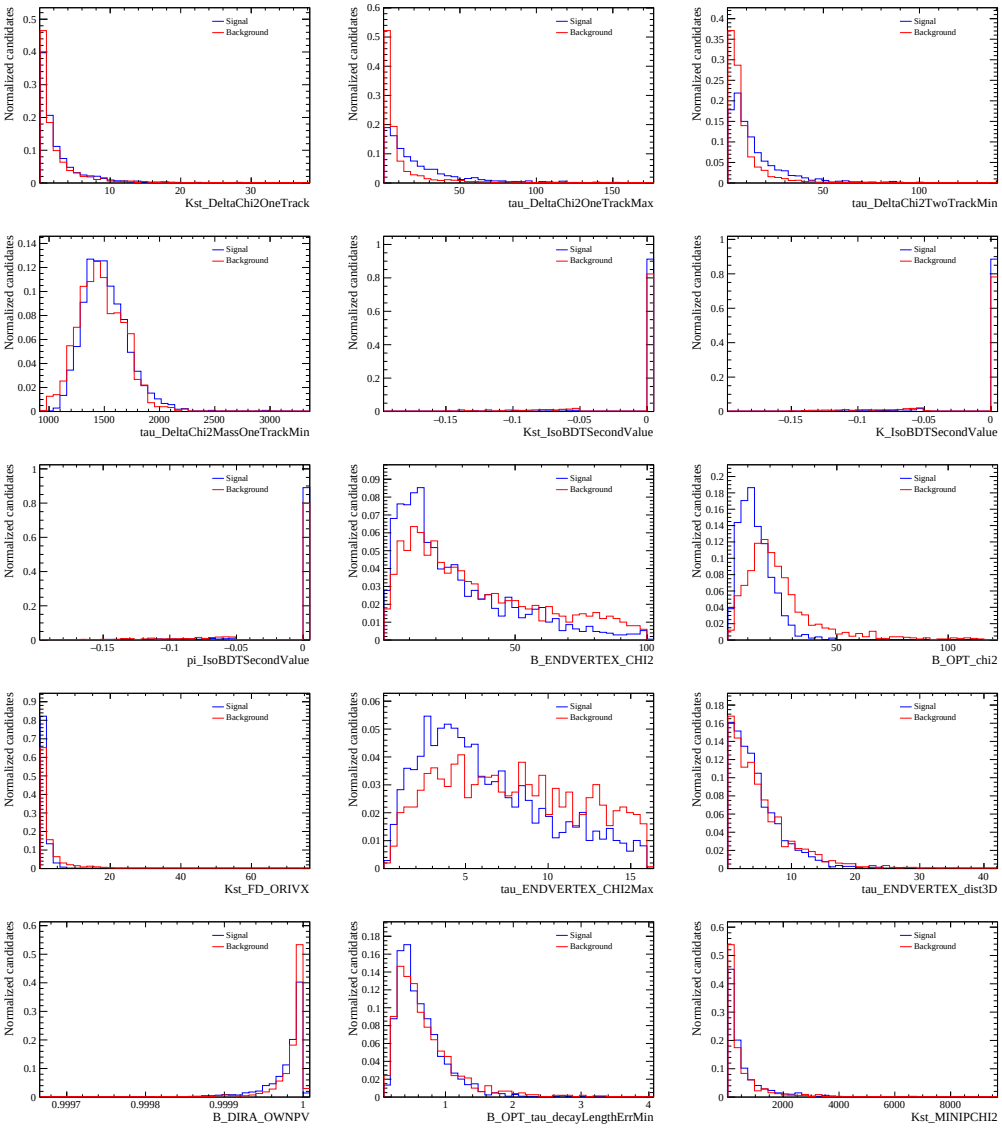


Figure B.5: Run 1 BDT2 variable distributions for the $3\pi 3\pi$ final state.

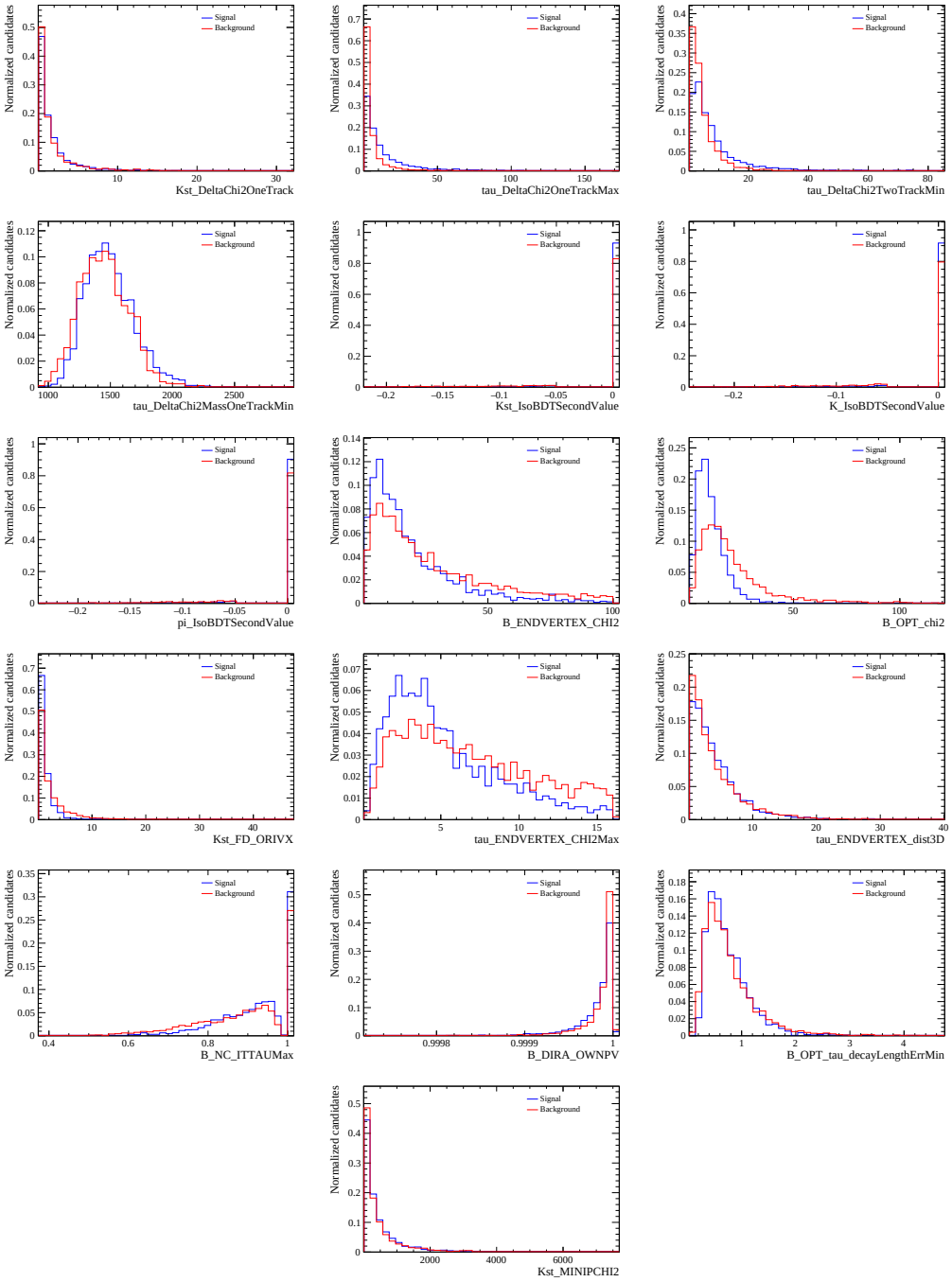


Figure B.6: 2016 BDT2 variable distributions for the $3\pi 3\pi$ final state.

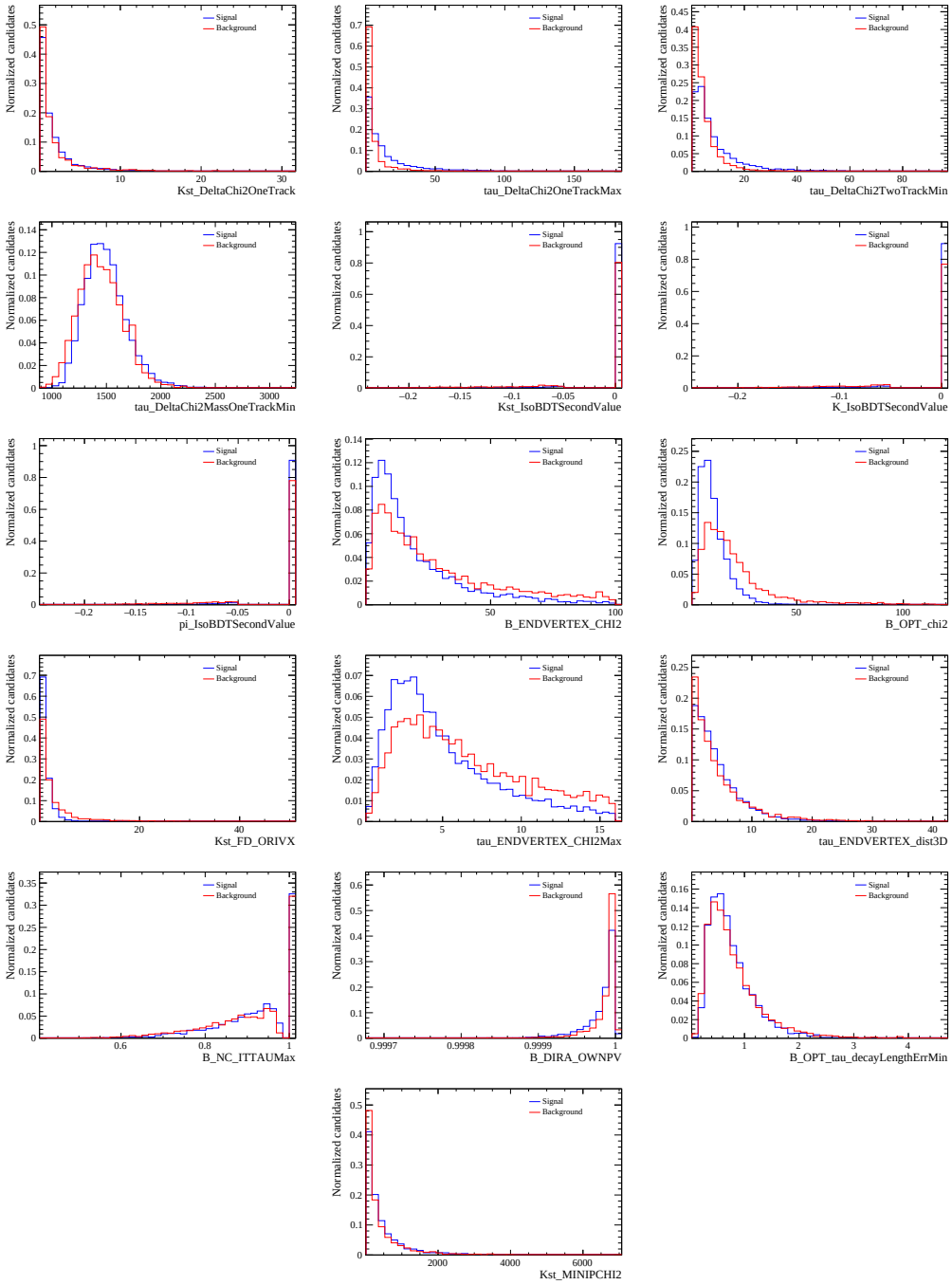


Figure B.7: 2017 BDT2 variable distributions for the $3\pi3\pi$ final state.

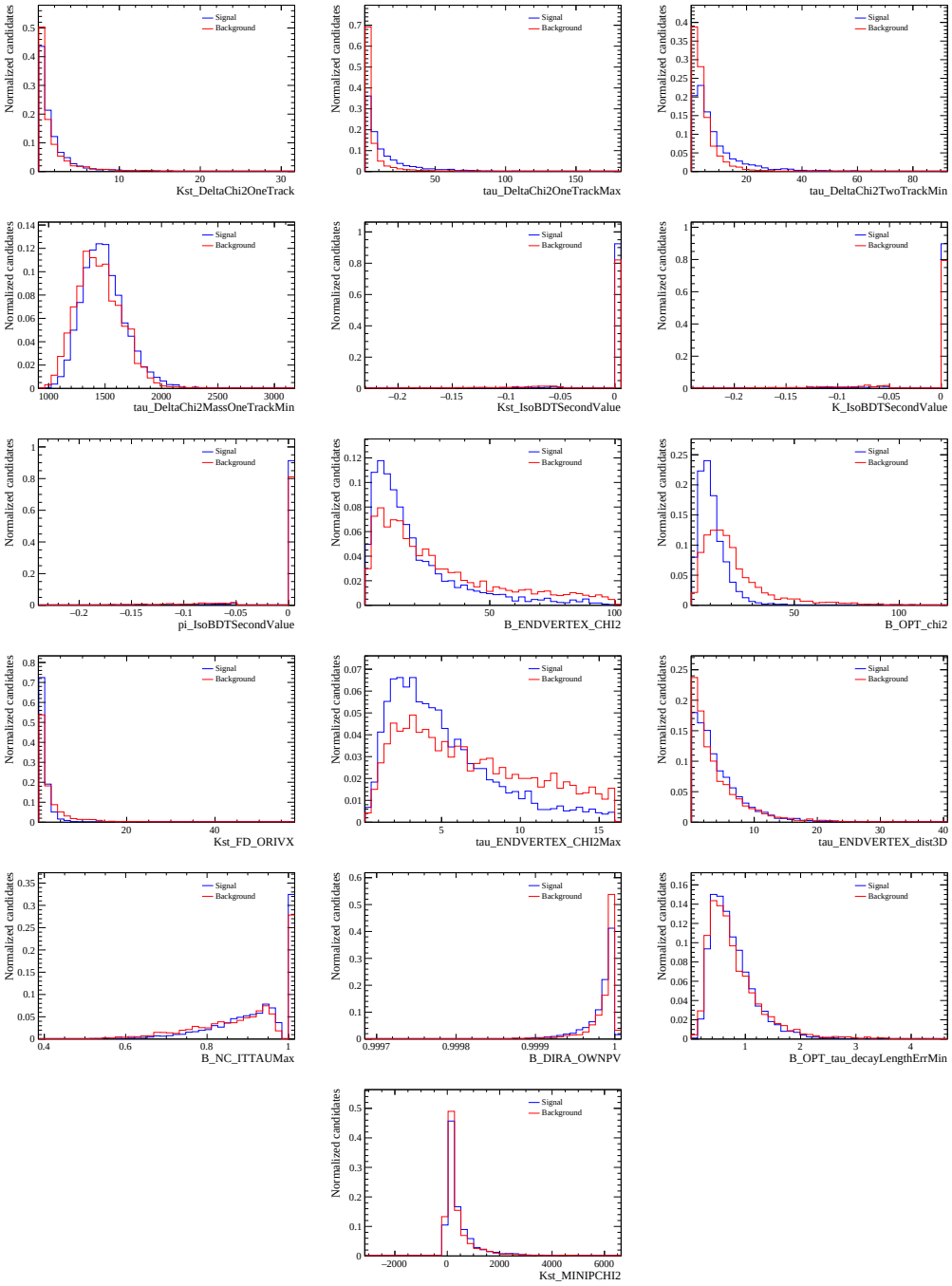


Figure B.8: 2018 BDT2 variable distributions for the $3\pi3\pi$ final state.

FitBDT

The fitBDT variable rankings are reported in Tables B.9, B.10, B.11, B.12 for each data-taking period.
The fitBDT variable distributions are reported in Figs. B.9, B.10, B.11, B.12 for each data-taking period.

Table B.9: Run 1 fitBDT variables ranking for the $3\pi 3\pi$ final state.

Rank	Name	Separation
1	flat_First3piBDT	0.128
2	B_OPT_tau_decayLengthMin	0.1004
3	resMassMin	0.05291
4	tau_P_optMin	0.05248
5	rhoMassMin	0.04714
6	tau_MMin	0.03564
7	tau_P_aMax	0.0331
8	B_OPT_tau_decayLengthMax	0.02932
9	B_IP_OWNPV	0.02786
10	piP1_IsoBDTThirdValueMax	0.02195
11	B_BPVVDRHO	0.02194
12	B_FDCHI2_OWNPV	0.02065
13	piP1_IsoBDTSecondValueMin	0.02038
14	tau_BPVVDZMax	0.01768
15	tau_IsoBDTSecondValueMax	0.01161

Table B.10: 2016 fitBDT variables ranking for the $3\pi 3\pi$ final state.

Rank	Name	Separation
1	flat_First3piBDT	0.1063
2	B_OPT_tau_decayLengthMin	0.07326
3	B_OPT_tau_decayLengthMax	0.04968
4	tau_MMin	0.04711
5	resMassMin	0.03536
6	tau_P_aMax	0.03256
7	rhoMassMin	0.03207
8	tau_P_optMin	0.02815
9	B_BPVVDRHO	0.01637
10	B_IP_OWNPV	0.01608
11	piP1_IsoBDTSecondValueMin	0.0143
12	B_NC_SPTTAUMin	0.01277
13	tau_IsoBDTSecondValueMax	0.01198
14	tau_BPVVDZMax	0.01179
15	B_FDCHI2_OWNPV	0.0106
16	piP1_IsoBDTThirdValueMax	0.01039

Table B.11: 2017 fitBDT variables ranking for the $3\pi 3\pi$ final state.

Rank	Name	Separation
1	flat_First3piBDT	0.1361
2	B_OPT_tau_decayLengthMin	0.0802
3	B_OPT_tau_decayLengthMax	0.0452
4	tau_MMin	0.03285
5	rhoMassMin	0.03056
6	tau_P_optMin	0.03043
7	tau_P_aMax	0.02588
8	resMassMin	0.02183
9	B_IP_OWNPV	0.01829
10	piP1_IsoBDTThirdValueMax	0.01727
11	B_BPVVDRHO	0.01307
12	B_FDCHI2_OWNPV	0.01126
13	tau_BPVVDZMax	0.01107
14	piP1_IsoBDTSecondValueMin	0.009764
15	B_NC_SPTTAUMin	0.007229
16	tau_IsoBDTSecondValueMax	0.007175

Table B.12: 2018 fitBDT variables ranking for the $3\pi 3\pi$ final state.

Rank	Name	Separation
1	flat_First3piBDT	0.1145
2	B_OPT_tau_decayLengthMin	0.0674
3	B_OPT_tau_decayLengthMax	0.04162
4	tau_MMin	0.03079
5	tau_P_optMin	0.02779
6	tau_P_aMax	0.02749
7	rhoMassMin	0.02086
8	resMassMin	0.01967
9	piP1_IsoBDTSecondValueMin	0.01613
10	B_IP_OWNPV	0.01441
11	piP1_IsoBDTThirdValueMax	0.01133
12	tau_BPVVDZMax	0.01078
13	B_BPVVDRHO	0.0107
14	B_FDCHI2_OWNPV	0.009125
15	B_NC_SPTTAUMin	0.009122
16	tau_IsoBDTSecondValueMax	0.007391

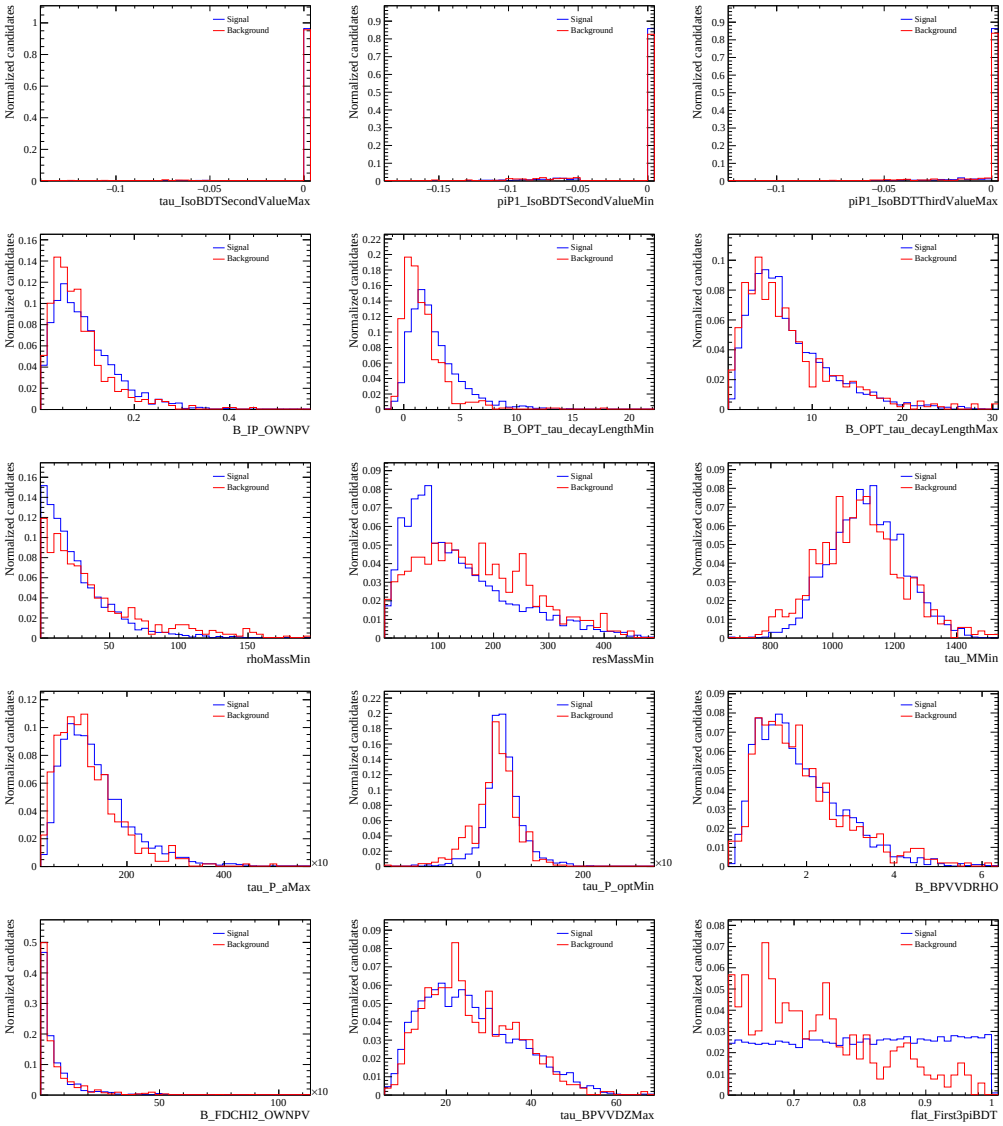


Figure B.9: Run 1 fitBDT variable distributions for the $3\pi3\pi$ final state.

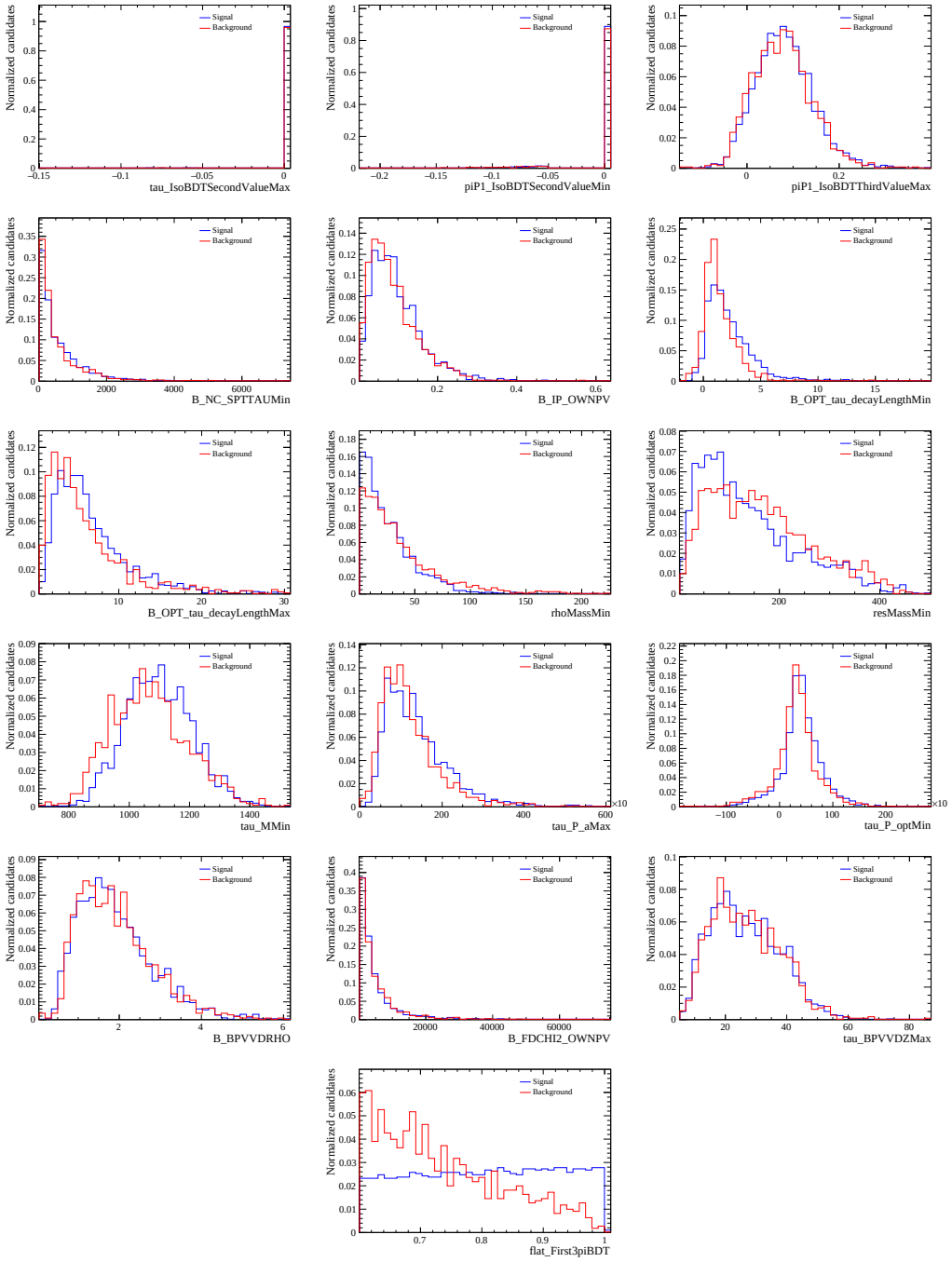


Figure B.10: 2016 fitBDT variable distributions for the $3\pi 3\pi$ final state.

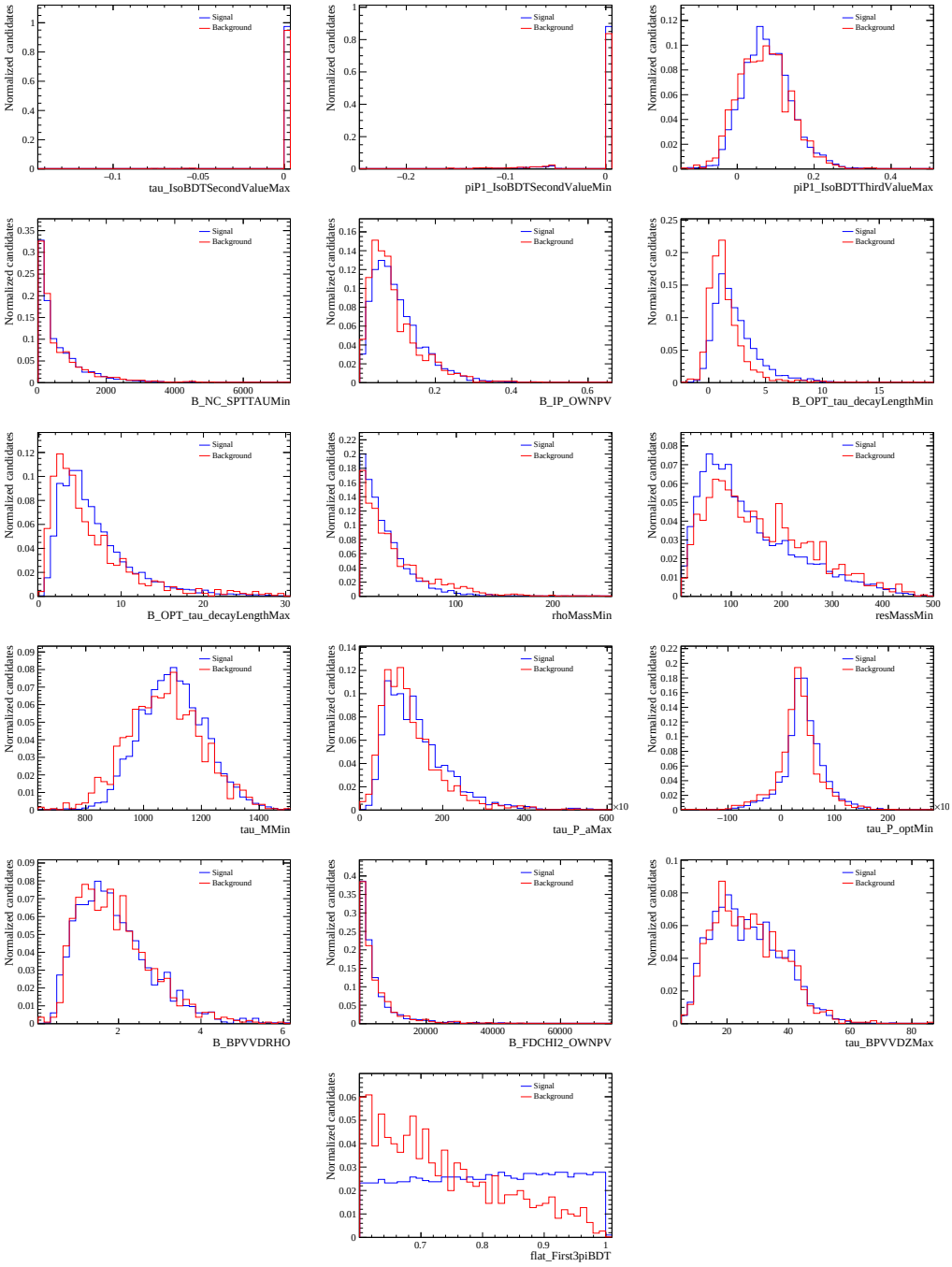


Figure B.11: 2017 fitBDT variable distributions for the $3\pi3\pi$ final state.

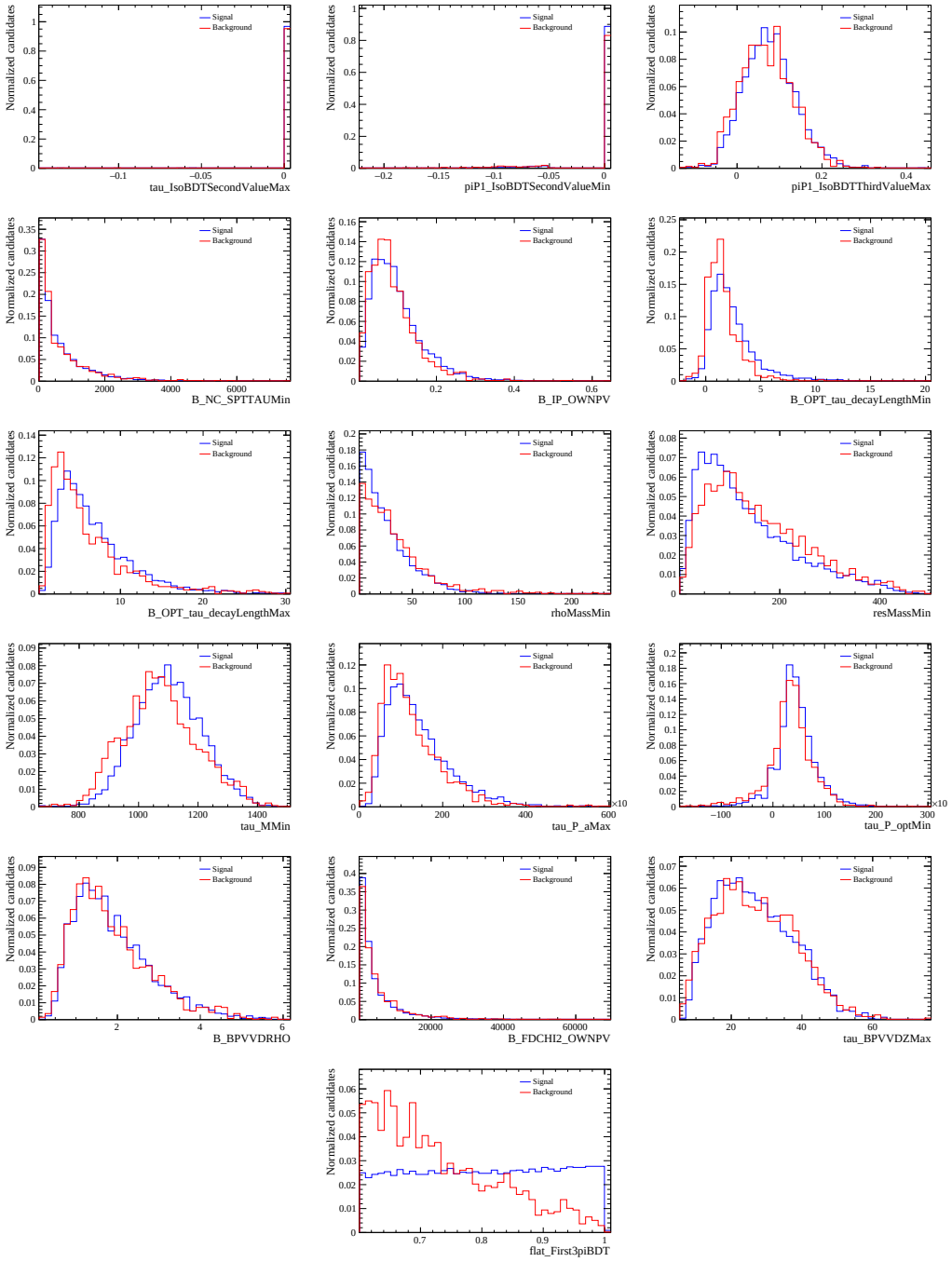


Figure B.12: 2018 fitBDT variable distributions for the $3\pi3\pi$ final state.

B.2 $3\pi\mu$ final state

BDT1

The BDT1 variable rankings are reported in Tables B.13, B.14, B.15, B.16 for each data-taking period.
The BDT1 variable distributions are reported in Figs. B.13, B.14, B.15, B.16 for each data-taking period.

Table B.13: Run 1 BDT1 variables ranking for the $3\pi\mu$ final state.

Rank	Name	Separation
1	Kst_DIRA_ORIVX	0.07935
2	Kst_FD_ORIVX	0.07318
3	tauP_M	0.05278
4	pi_PT	0.04167
5	B_M	0.03697
6	Kst_DOCAKpi	0.03308
7	tauP_M23	0.02841
8	tauP_FD_ORIVX	0.02528
9	K_IsoBDTSecondValue	0.01952
10	mu_PT	0.009725

Table B.14: 2016 BDT1 variables ranking for the $3\pi\mu$ final state.

Rank	Name	Separation
1	Kst_FD_ORIVX	0.07009
2	pi_PT	0.05982
3	Kst_DIRA_ORIVX	0.05976
4	tauP_M	0.05178
5	Kst_DOCAKpi	0.05028
6	B_M	0.04118
7	mu_PT	0.0411
8	tauP_M23	0.03284
9	K_IsoBDTSecondValue	0.02676
10	tauP_FD_ORIVX	0.02549

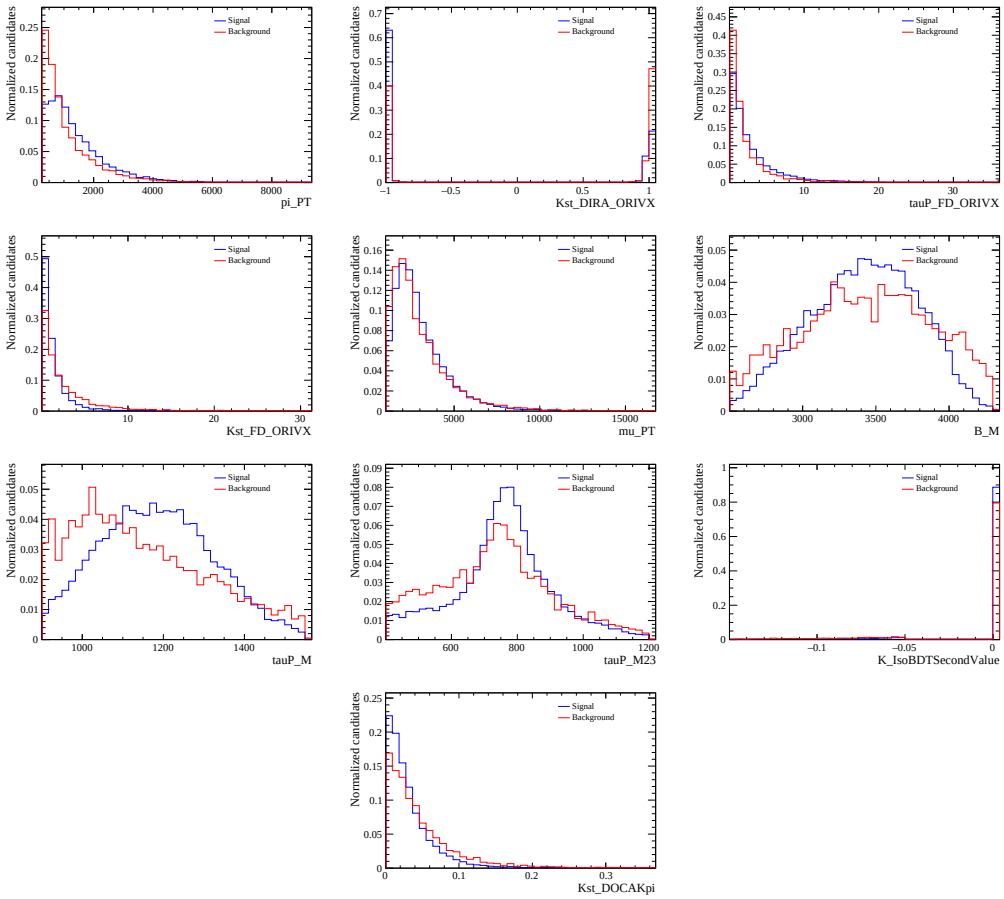


Figure B.13: Run 1 BDT1 variable distributions for the $3\pi\mu$ final state.

Table B.15: 2017 BDT1 variables ranking for the $3\pi\mu$ final state.

Rank	Name	Separation
1	Kst_DIRA_ORIVX	0.0735
2	Kst_FD_ORIVX	0.06666
3	pi_PT	0.04991
4	tauP_M	0.04634
5	Kst_DOCAKpi	0.03523
6	B_M	0.03462
7	tauP_M23	0.03236
8	tauP_FD_ORIVX	0.0242
9	K_IsoBDTSecondValue	0.01894
10	mu_PT	0.01412

Table B.16: 2018 BDT1 variables ranking for the $3\pi\mu$ final state.

Rank	Name	Separation
1	Kst_DIRA_ORIVX	0.0744
2	Kst_FD_ORIVX	0.06132
3	pi_PT	0.04816
4	tauP_M	0.0456
5	B_M	0.03871
6	Kst_DOCAKpi	0.03611
7	tauP_M23	0.03592
8	tauP_FD_ORIVX	0.02369
9	K_IsoBDTSecondValue	0.02052
10	mu_PT	0.01438

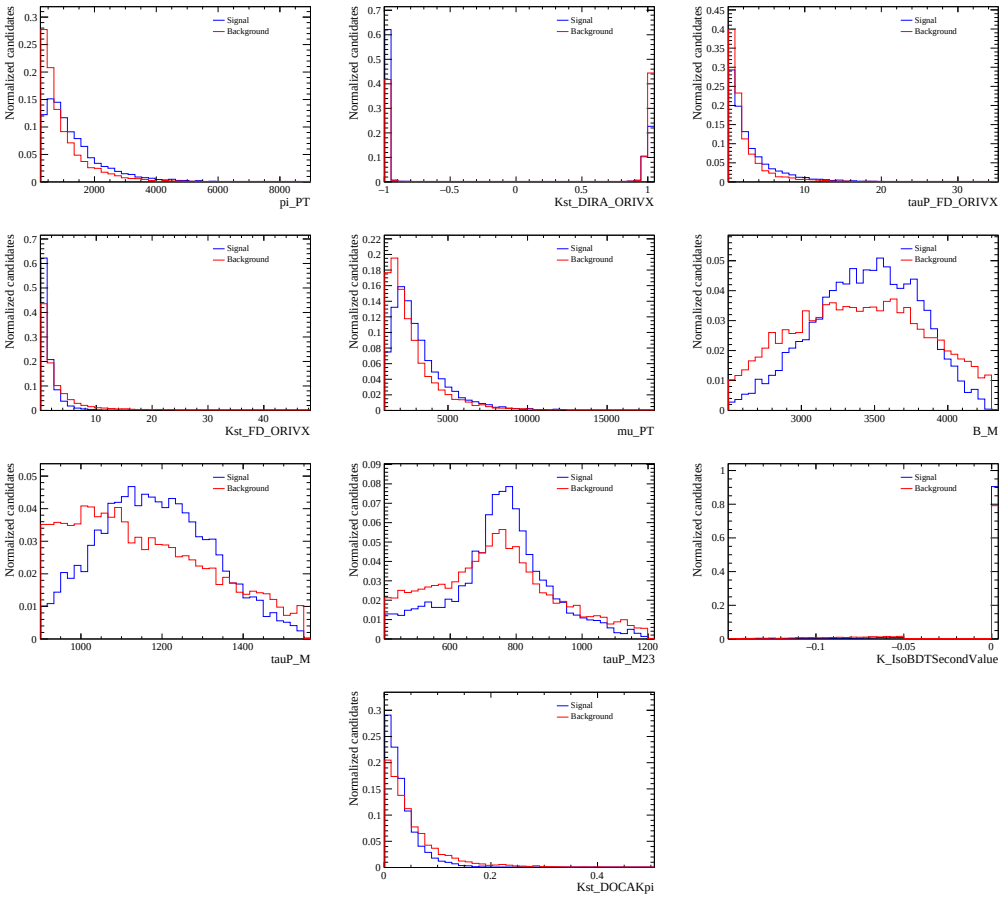


Figure B.14: 2016 BDT1 variable distributions for the $3\pi\mu$ final state.

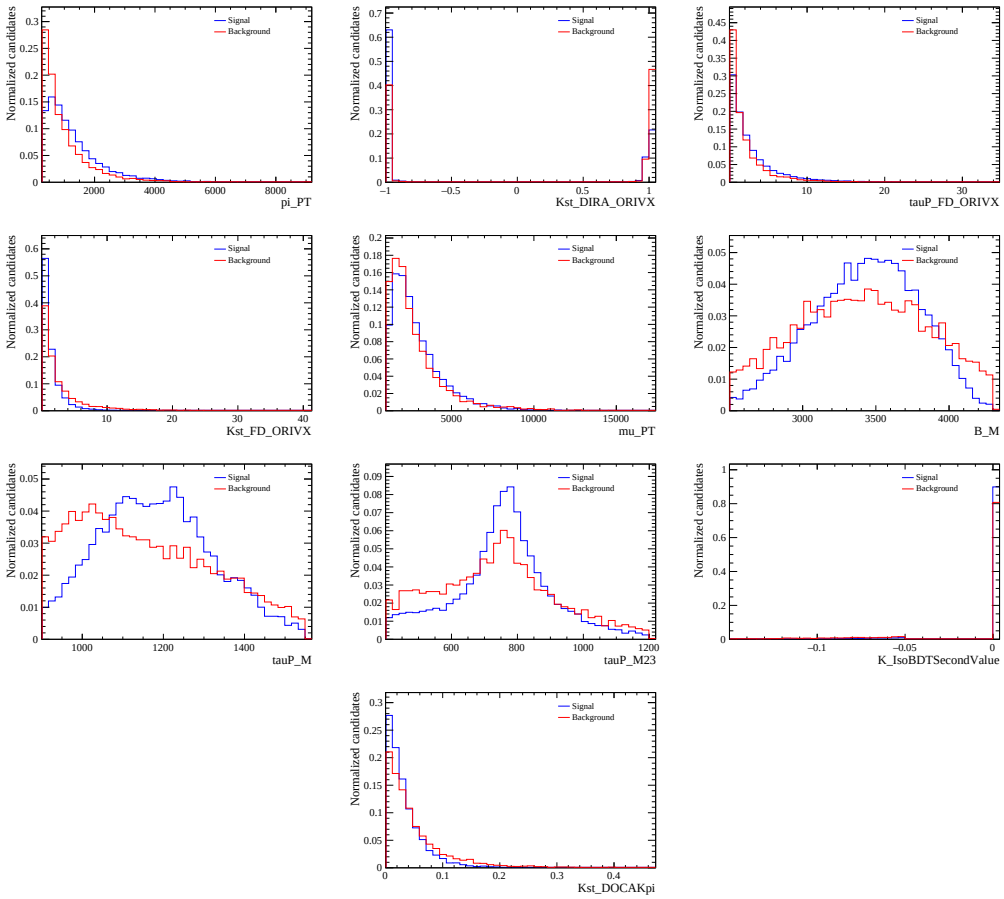


Figure B.15: 2017 BDT1 variable distributions for the $3\pi\mu$ final state.

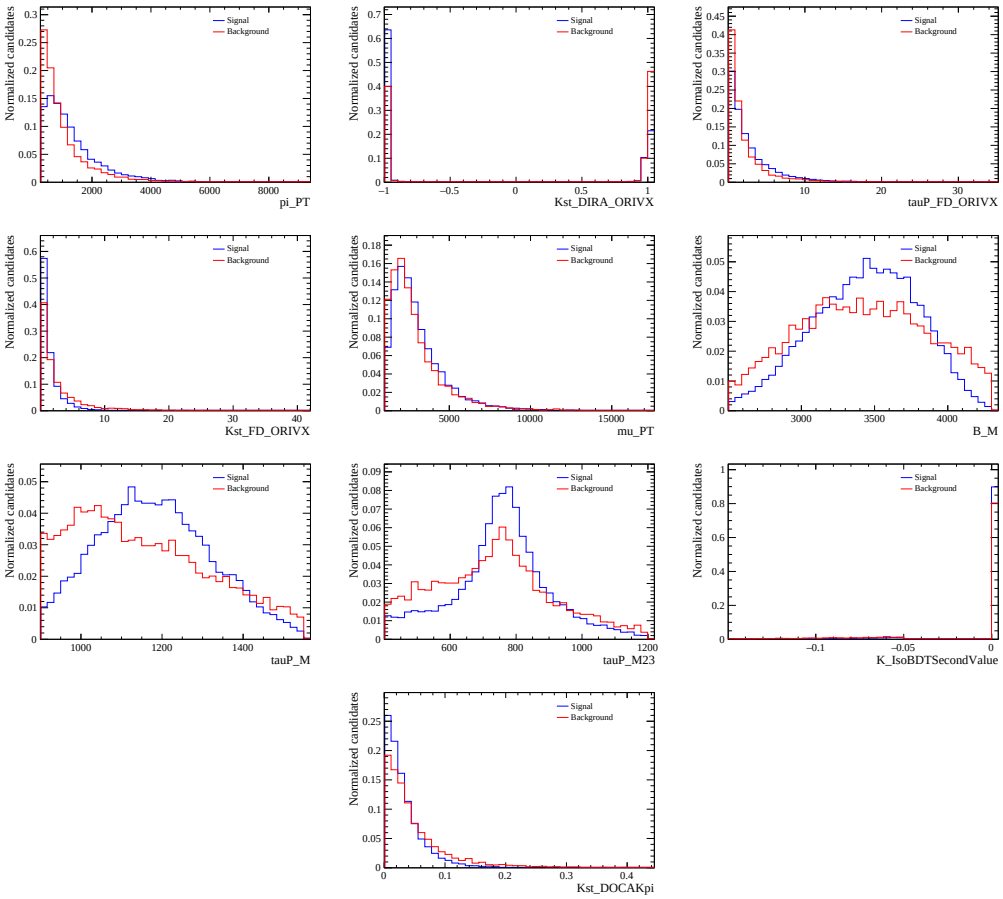


Figure B.16: 2018 BDT1 variable distributions for the $3\pi\mu$ final state.

BDT2

The BDT2 variable rankings are reported in Tables B.17, B.18, B.19, B.20 for each data-taking period.
The BDT2 variable distributions are reported in Figs. B.17, B.18, B.19, B.20 for each data-taking period.

Table B.17: Run 1 BDT2 variables ranking for the $3\pi\mu$ final state.

Rank	Name	Separation
1	flat_FirstmuBDT	0.06876
2	tauP_FDCHI2_ORIVX	0.02402
3	B_DOCAtauPtauM	0.009868
4	Kst_DIRA_ORIVX	0.003898
5	Kst_FD_ORIVX	0.003715

Table B.18: 2016 BDT2 variables ranking for the $3\pi\mu$ final state.

Rank	Name	Separation
1	flat_FirstmuBDT	0.07082
2	tauP_FDCHI2_ORIVX	0.02067
3	B_DOCAtauPtauM	0.007188
4	Kst_FD_ORIVX	0.006785
5	Kst_DIRA_ORIVX	0.004946

Table B.19: 2017 BDT2 variables ranking for the $3\pi\mu$ final state.

Rank	Name	Separation
1	flat_FirstmuBDT	0.06685
2	tauP_FDCHI2_ORIVX	0.01426
3	B_DOCAtauPtauM	0.007371
4	Kst_FD_ORIVX	0.004982
5	Kst_DIRA_ORIVX	0.002519

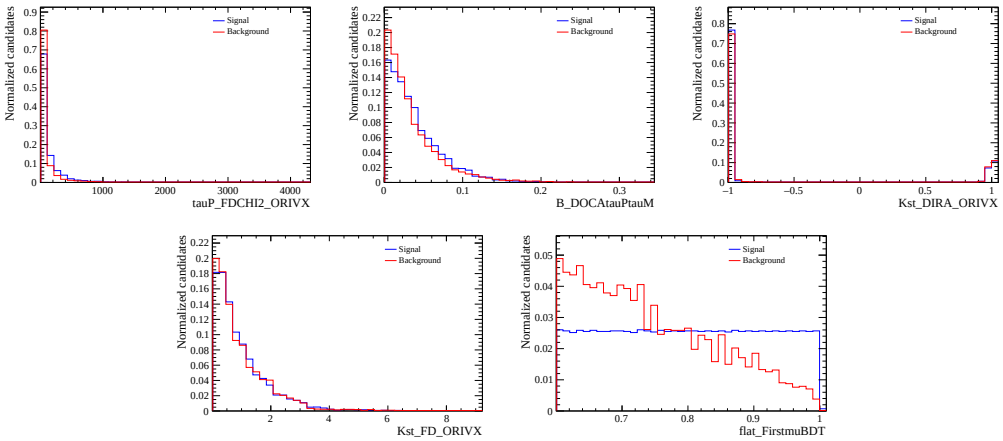


Figure B.17: Run 1 BDT2 variable distributions for the $3\pi\mu$ final state.

Table B.20: 2018 BDT2 variables ranking for the $3\pi\mu$ final state.

Rank	Name	Separation
1	flat_FirstmuBDT	0.06823
2	tauP_FDCHI2_ORIVX	0.01325
3	B_DOCAtauPtauM	0.00999
4	Kst_FD_ORIVX	0.00357
5	Kst_DIRA_ORIVX	0.00349

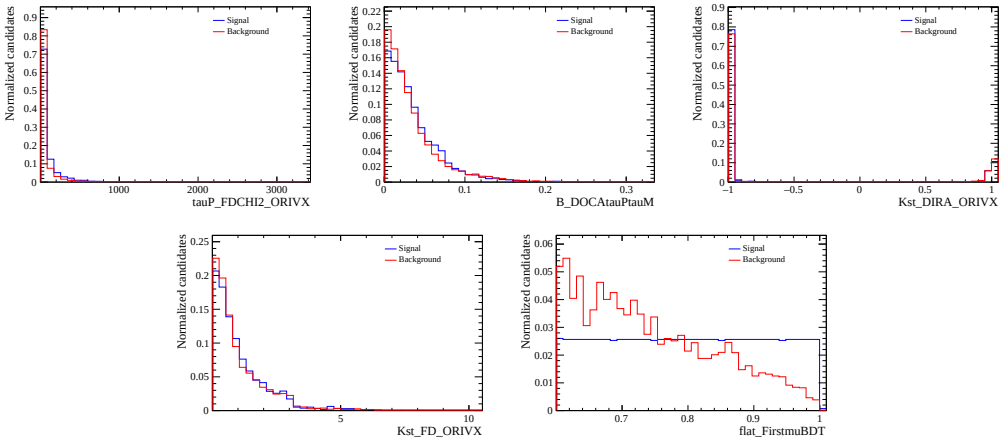


Figure B.18: 2016 BDT2 variable distributions for the $3\pi\mu$ final state.

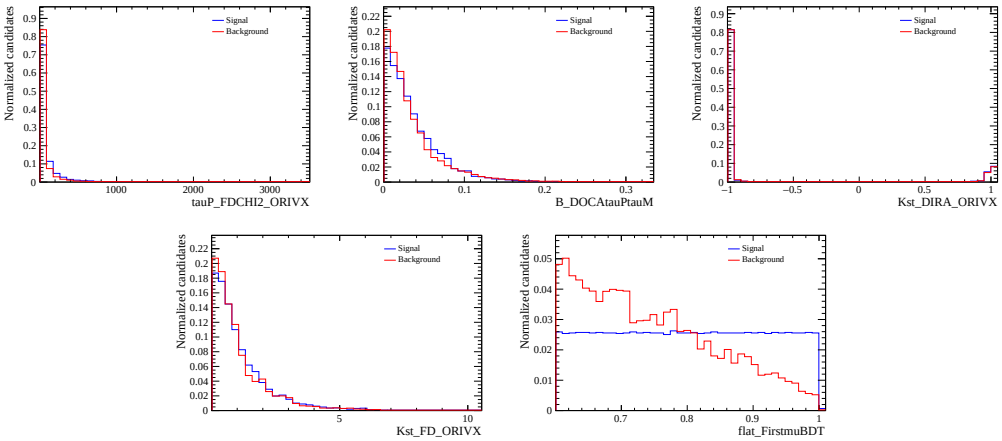


Figure B.19: 2017 BDT2 variable distributions for the $3\pi\mu$ final state.

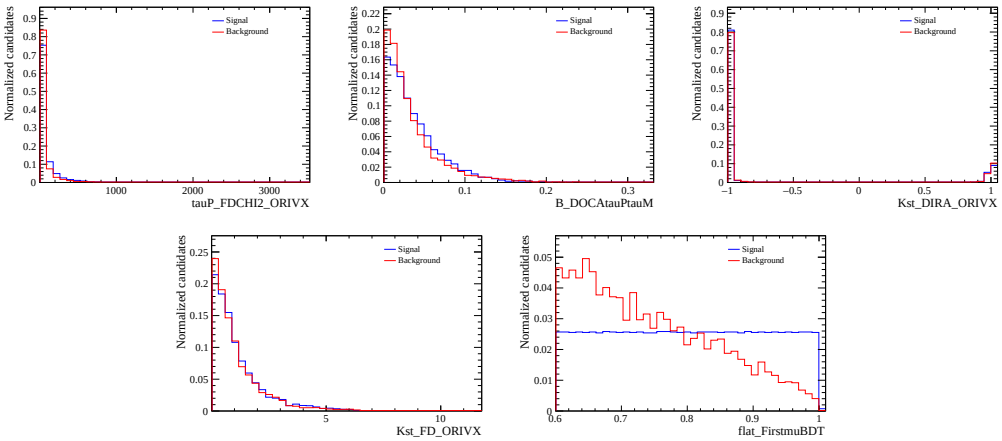


Figure B.20: 2018 BDT2 variable distributions for the $3\pi\mu$ final state.

FitBDT

The fitBDT variable rankings are reported in Tables B.21, B.22, B.23, B.24 for each data-taking period.
The fitBDT variable distributions are reported in Figs. B.21, B.22, B.23, B.24 for each data-taking period.

Table B.21: Run1 fitBDT variables ranking for the $3\pi\mu$ final state.

Rank	Name	Separation
1	tauP_DIRA_ORIVX	0.1574
2	tauP_DOCAMAX	0.02661
3	tauP_M12	0.02156
4	B_Mcor	0.02117
5	Kst_DIRA_OWNPV	0.01619
6	tauP_MINIPCHI2	0.01589
7	B_DOCAKsttauP	0.0137
8	B_ENDVERTEX_CHI2	0.01331
9	tauP_IsoBDTSecondValue	0.0123
10	mu_OWNPV_CHI2	0.008332

Table B.22: 2016 fitBDT variables ranking for the $3\pi\mu$ final state.

Rank	Name	Separation
1	tauP_DIRA_ORIVX	0.1423
2	B_NC_PASYM	0.105
3	tauP_DOCAMAX	0.03273
4	tauP_M12	0.02605
5	Kst_DIRA_OWNPV	0.02393
6	B_DOCAKsttauP	0.01835
7	B_Mcor	0.01772
8	tauP_MINIPCHI2	0.01668
9	B_ENDVERTEX_CHI2	0.01263
10	tauP_IsoBDTSecondValue	0.01227
11	mu_OWNPV_CHI2	0.01059

Table B.23: 2017 fitBDT variables ranking for the $3\pi\mu$ final state.

Rank	Name	Separation
1	tauP_DIRA_ORIVX	0.1499
2	B_NC_PASYM	0.1023
3	B_ENDVERTEX_CHI2	0.02458
4	tauP_DOCAMAX	0.02123
5	B_DOCAKsttauP	0.01804
6	B_Mcor	0.01739
7	tauP_M12	0.01658
8	Kst_DIRA_OWNPV	0.01184
9	tauP_IsoBDTSecondValue	0.01043
10	mu_OWNPV_CHI2	0.01007
11	tauP_MINIPCHI2	0.007632

Table B.24: 2018 fitBDT variables ranking for the $3\pi\mu$ final state.

Rank	Name	Separation
1	tauP_DIRA_ORIVX	0.1372
2	B_NC_PASYM	0.1
3	tauP_M12	0.02474
4	tauP_DOCAMAX	0.02206
5	B_Mcor	0.01742
6	B_ENDVERTEX_CHI2	0.01742
7	B_DOCAKsttauP	0.01673
8	tauP_IsoBDTSecondValue	0.0101
9	mu_OWNPV_CHI2	0.009681
10	Kst_DIRA_OWNPV	0.008193
11	tauP_MINIPCHI2	0.006073

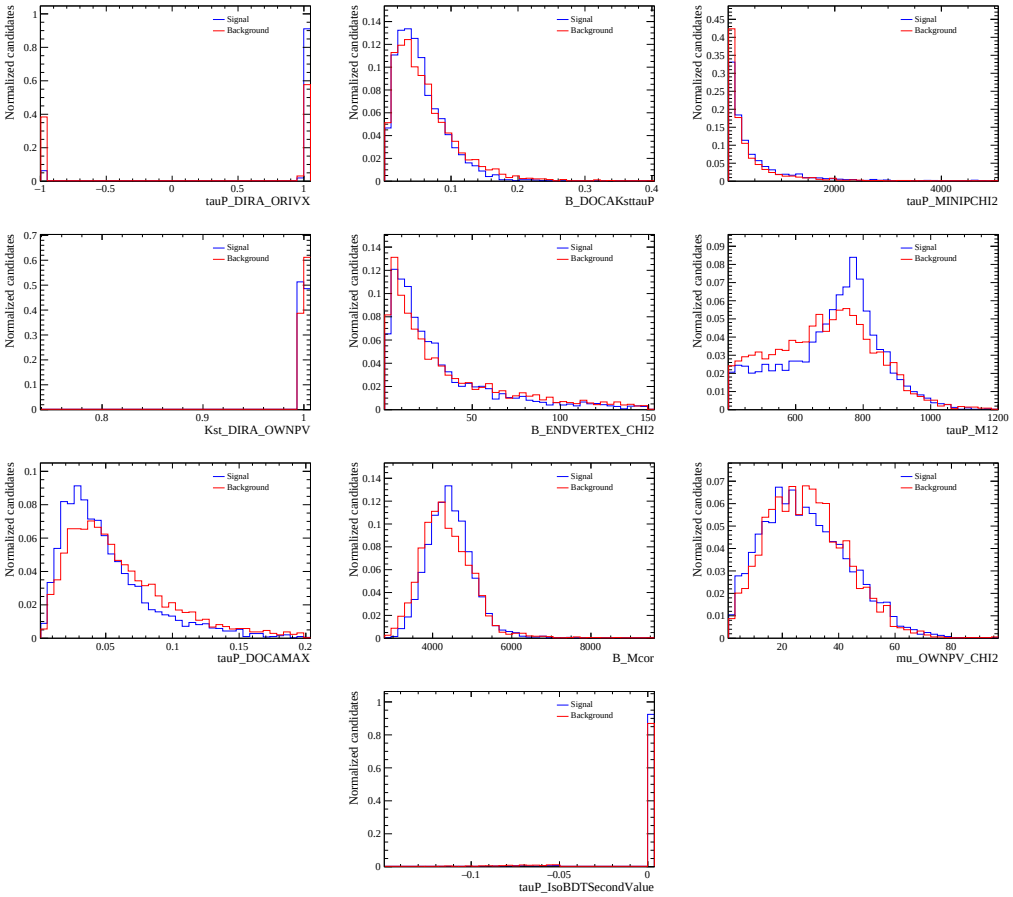


Figure B.21: Run 1 fitBDT variable distributions for the $3\pi\mu$ final state.

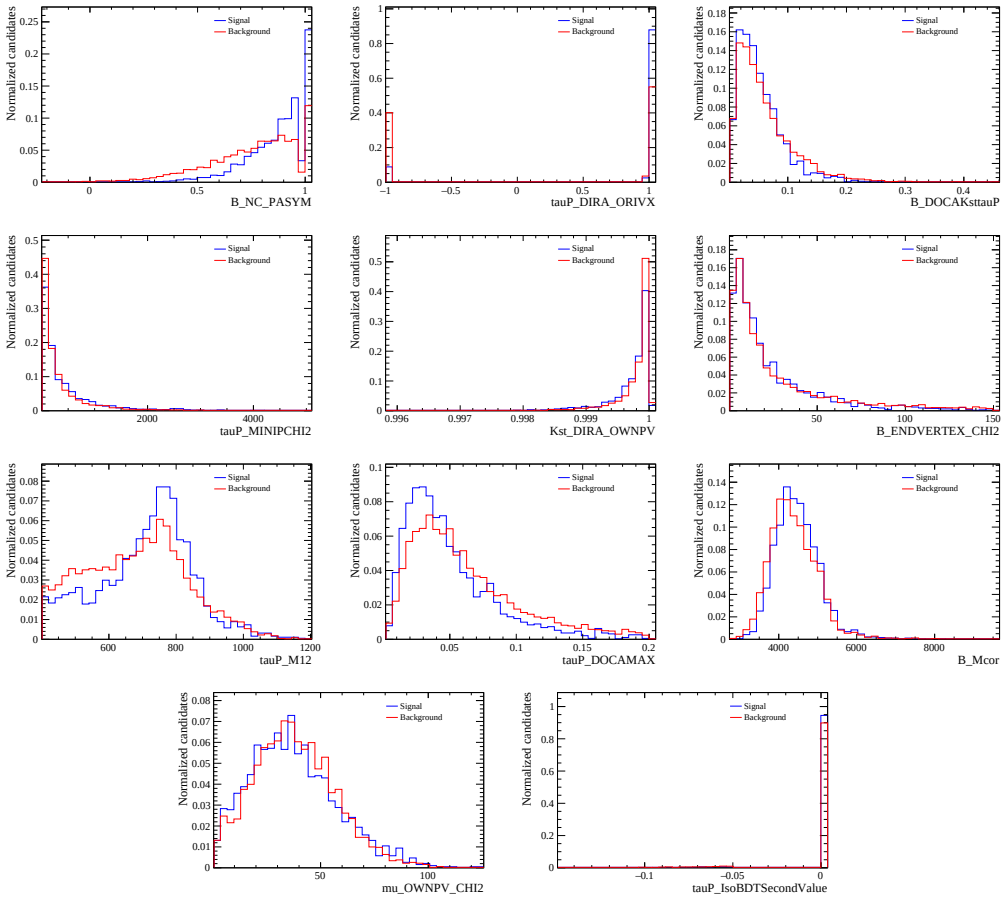


Figure B.22: 2016 fitBDT variable distributions for the $3\pi\mu$ final state.

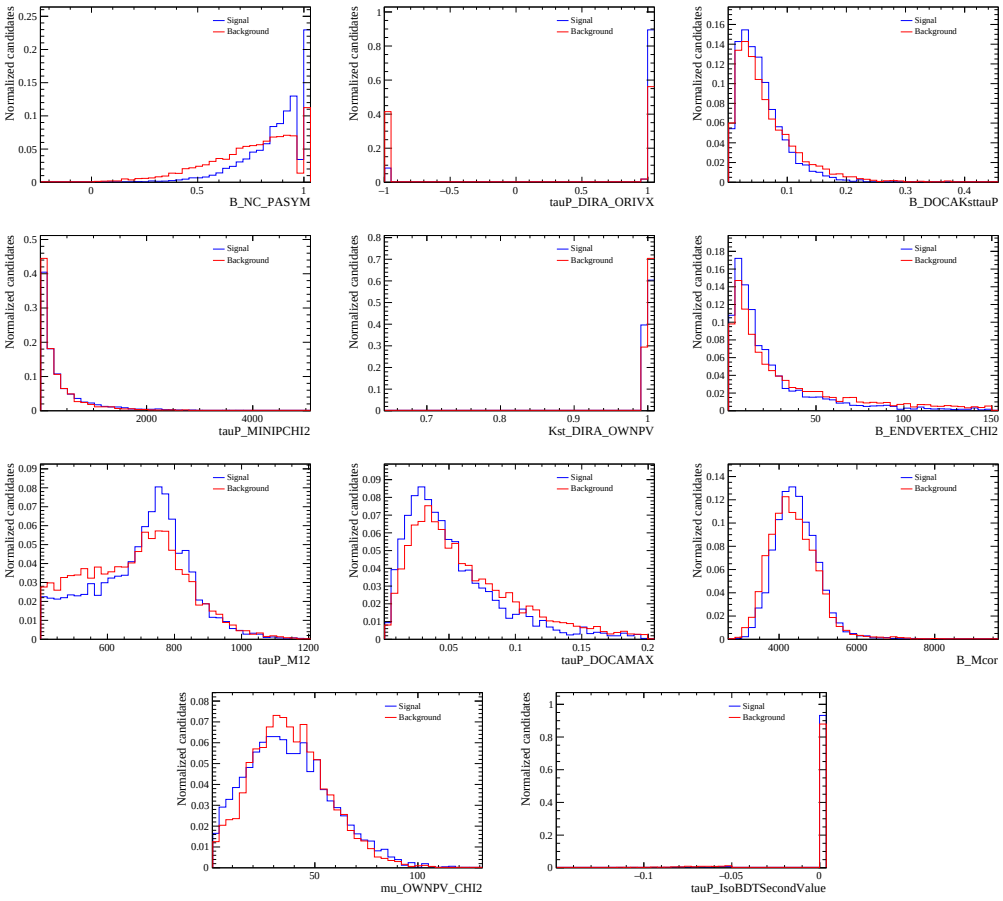


Figure B.23: 2017 fitBDT variable distributions for the $3\pi\mu$ final state.

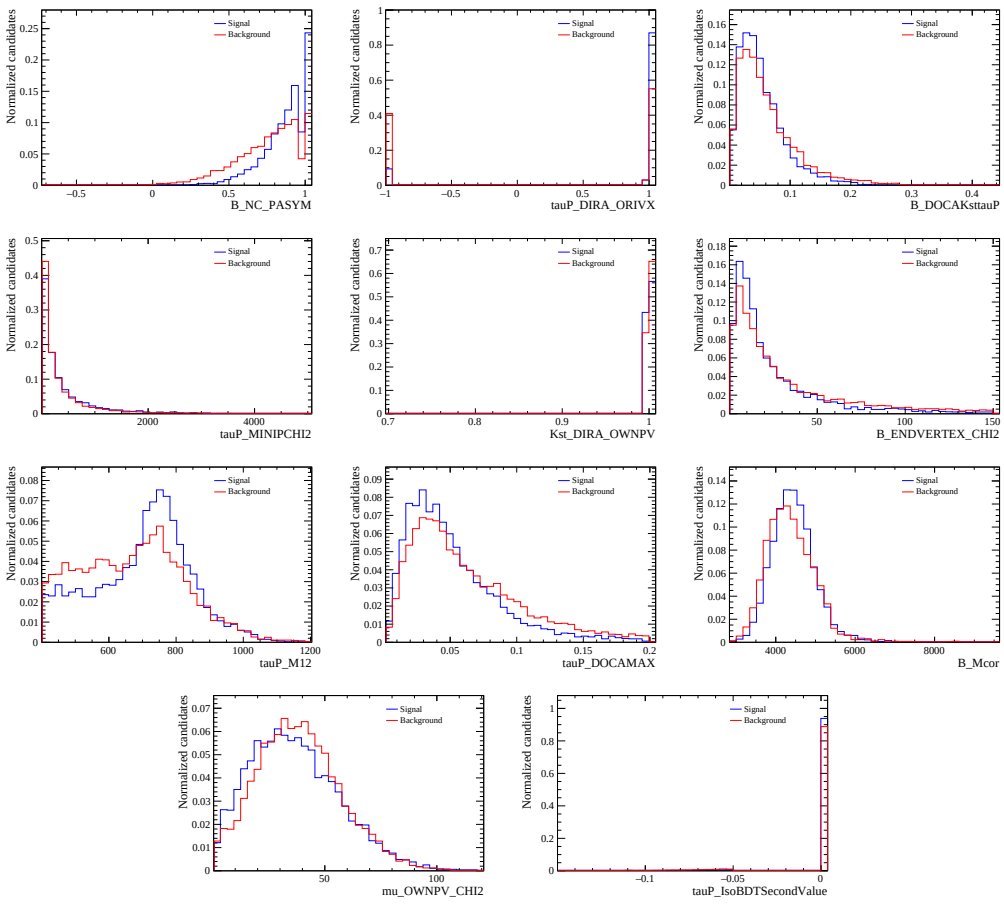


Figure B.24: 2018 fitBDT variable distributions for the $3\pi\mu$ final state.

Appendix C

Data-MC comparison

In this Appendix the comparison plots between data and MC are shown. The variables labelled as τ variables refer to the D mesons.

The data-MC comparison for the 2011 dataset is shown in Figs. C.1, C.2, C.3, C.4, C.5 and C.6.
The data-MC comparison for the 2012 dataset is shown in Figs. C.7, C.8, C.9, C.10, C.11 and C.12.
The data-MC comparison for the 2015 dataset is shown in Figs. C.13, C.14, C.15, C.16, C.17 and C.18.
The data-MC comparison for the 2016 dataset is shown in Figs. C.19, C.20, C.21, C.22, C.23 and C.24.
The data-MC comparison for the 2017 dataset is shown in Figs. C.25, C.26, C.27, C.28, C.29 and C.30.
The data-MC comparison for the 2018 dataset is shown in Figs. C.31, C.32, C.33, C.34, C.35 and C.36.

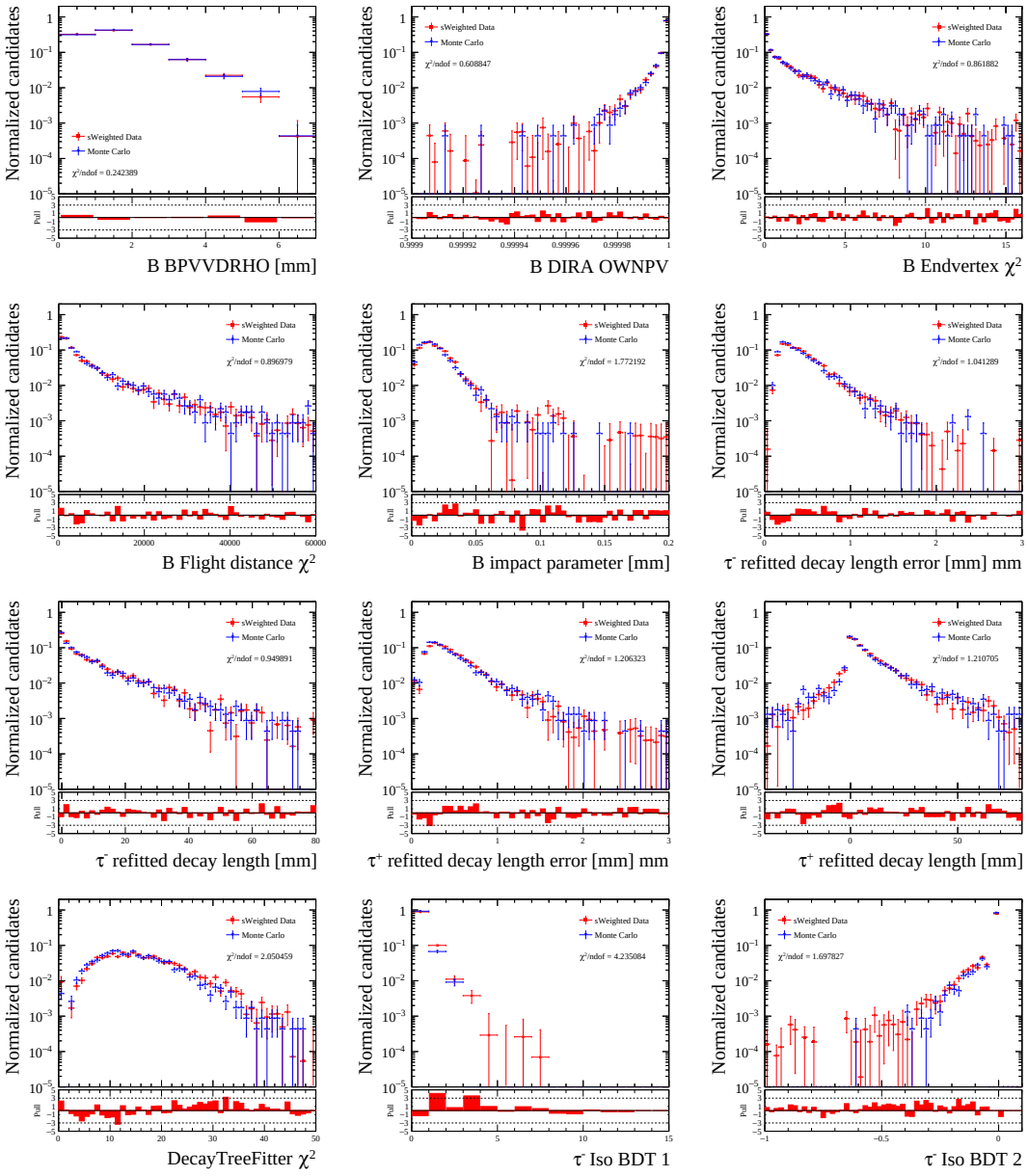


Figure C.1: Data-MC comparison for the 2011 dataset using the $B^0 \to D^- D_s^+$ channel (part 1).

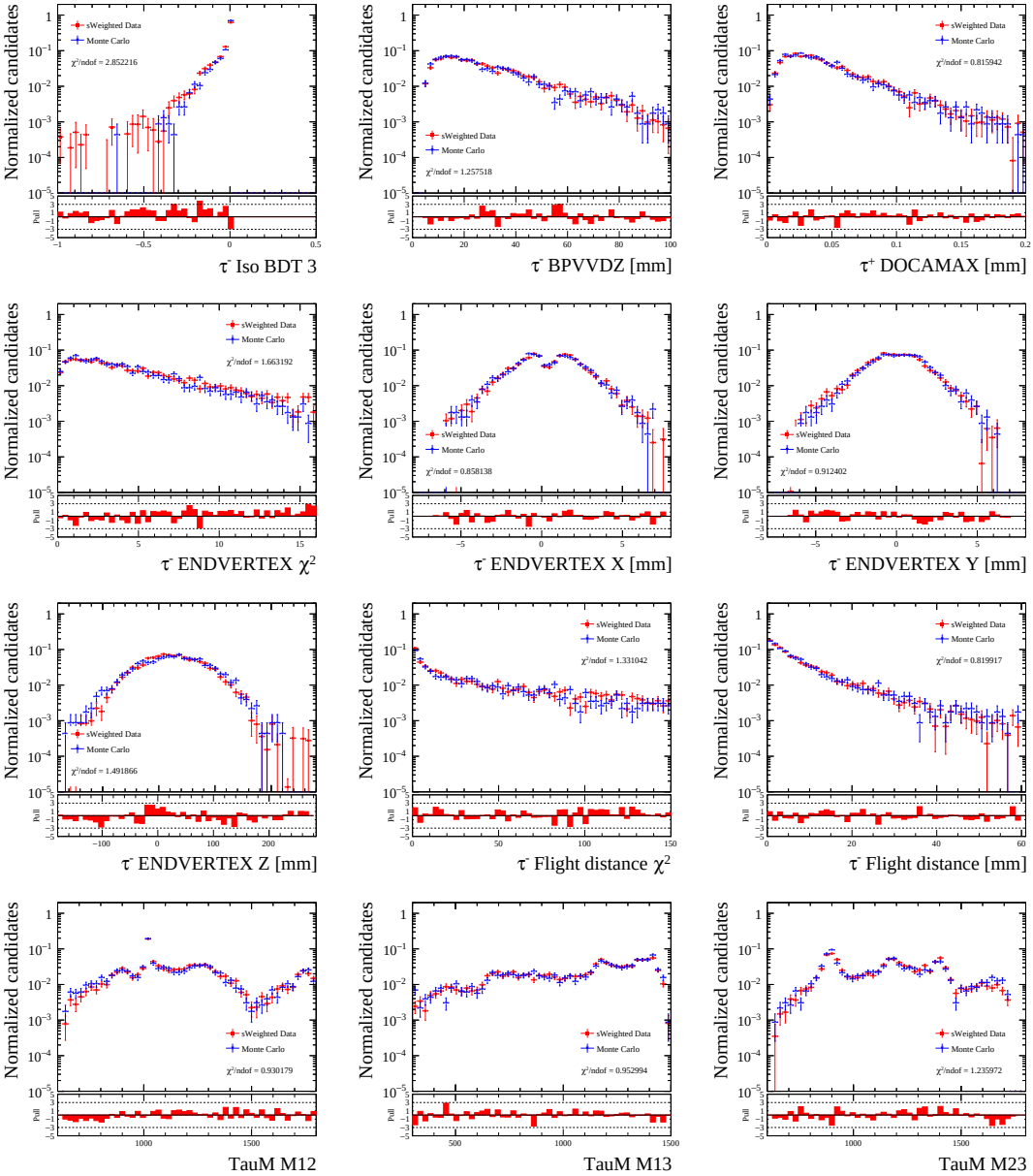


Figure C.2: Data-MC comparison for the 2011 dataset using the $B^0 \to D^- D_s^+$ channel (part 2).

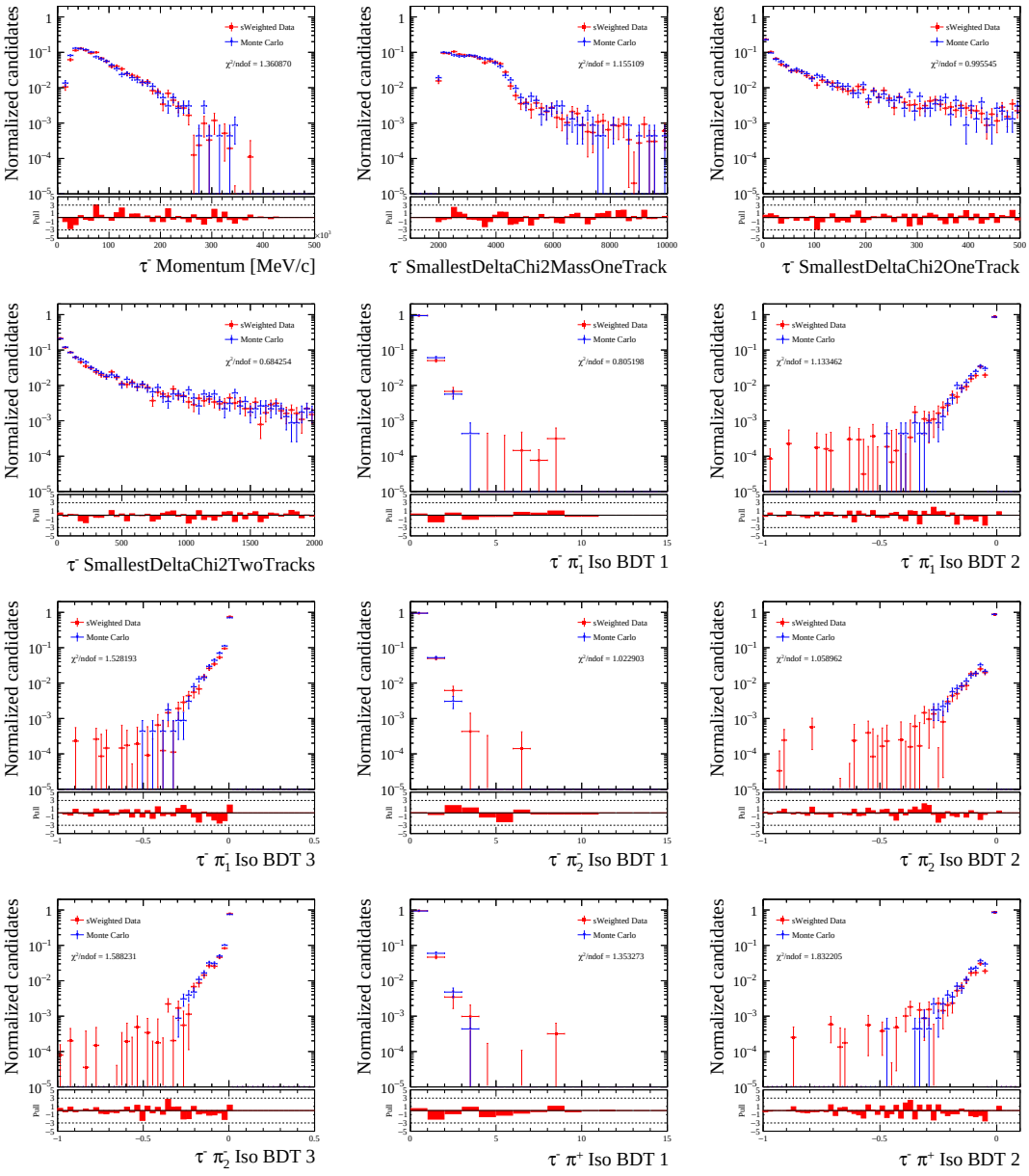


Figure C.3: Data-MC comparison for the 2011 dataset using the $B^0 \to D^- D_s^+$ channel (part 3).

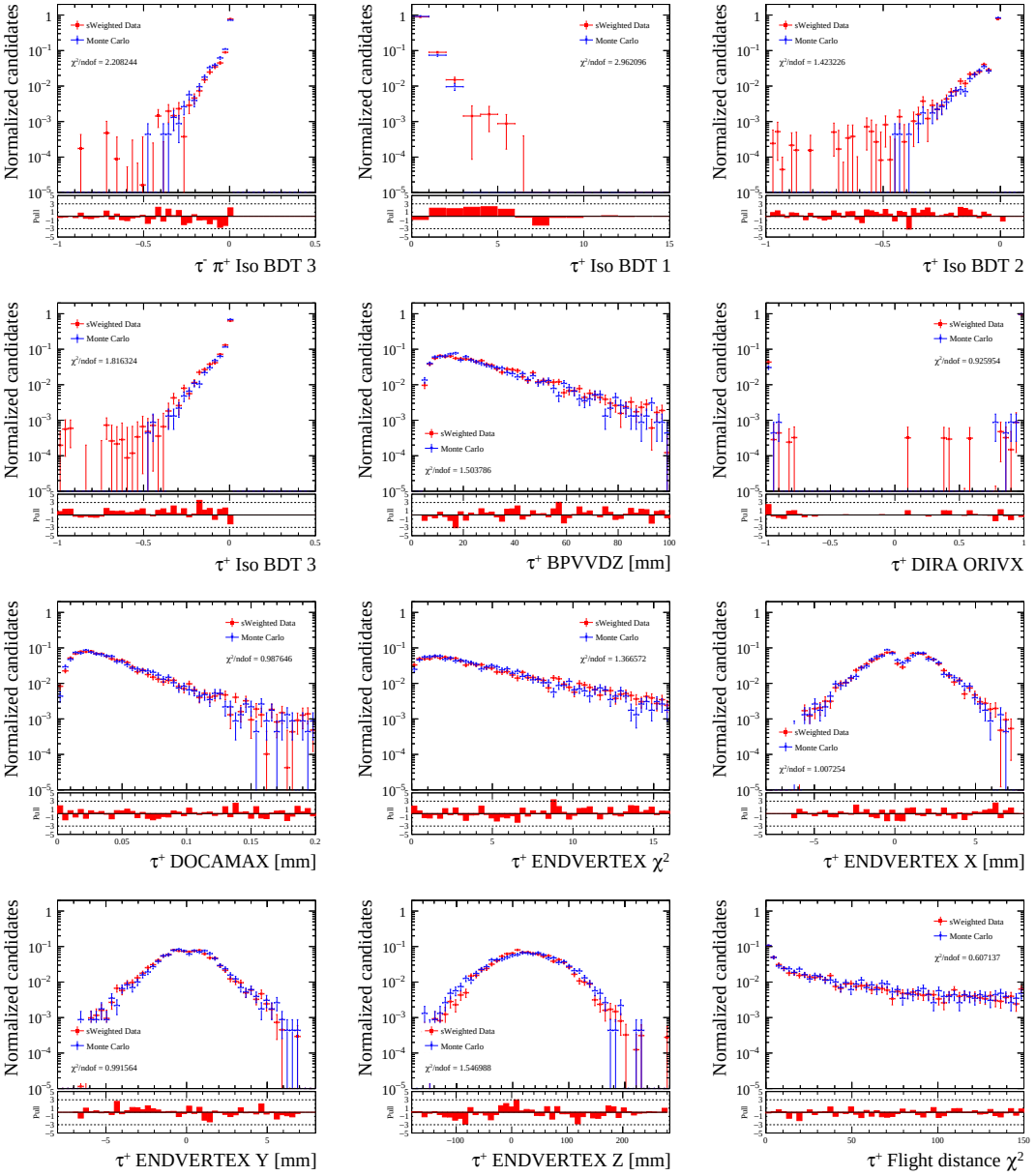


Figure C.4: Data-MC comparison for the 2011 dataset using the $B^0 \to D^- D_s^+$ channel (part 4).

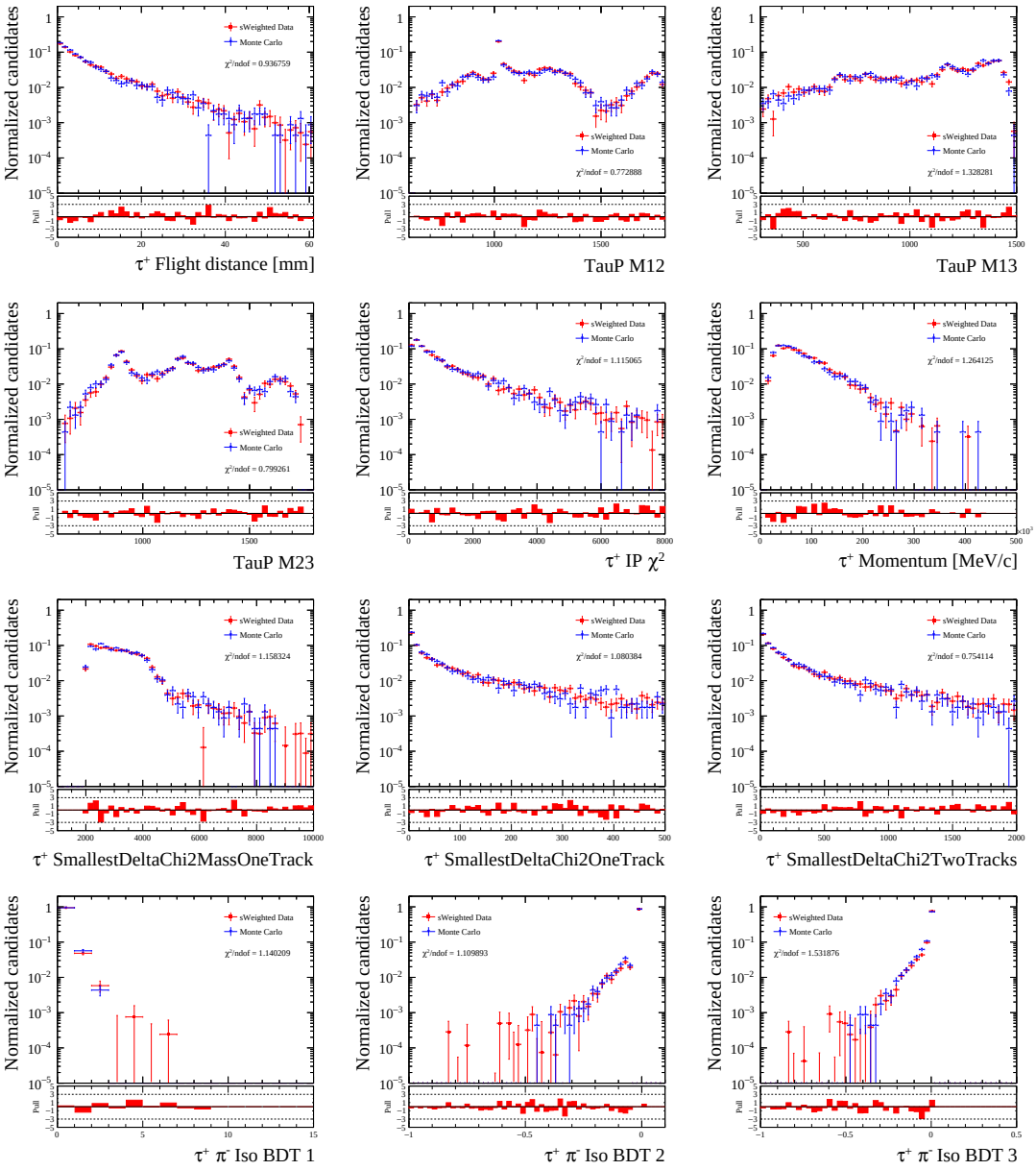


Figure C.5: Data-MC comparison for the 2011 dataset using the $B^0 \to D^- D_s^+$ channel (part 5).

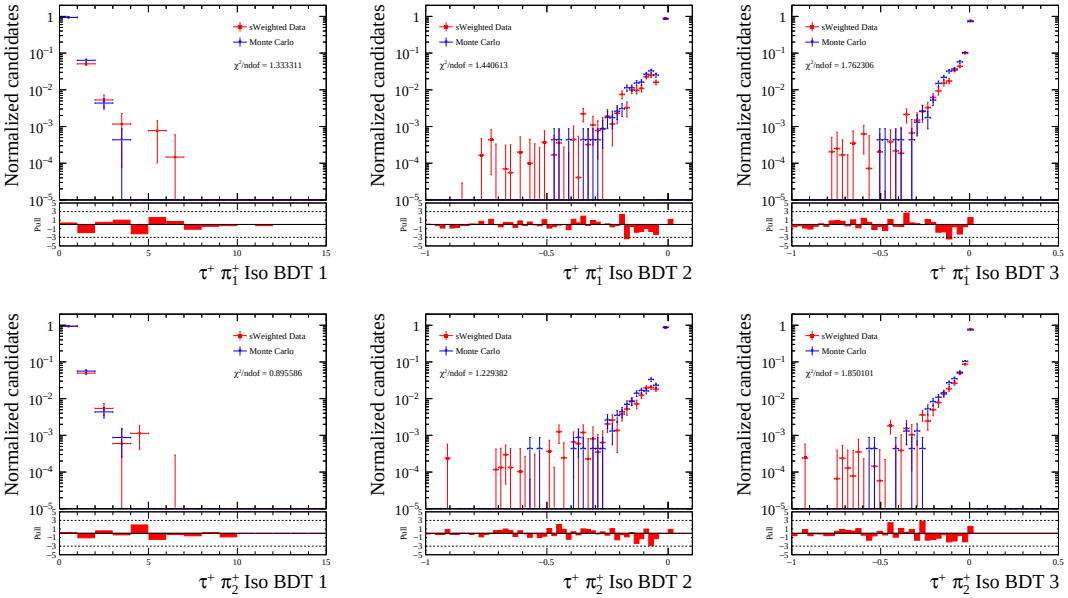


Figure C.6: Data-MC comparison for the 2011 dataset using the $B^0 \to D^- D_s^+$ channel (part 6).

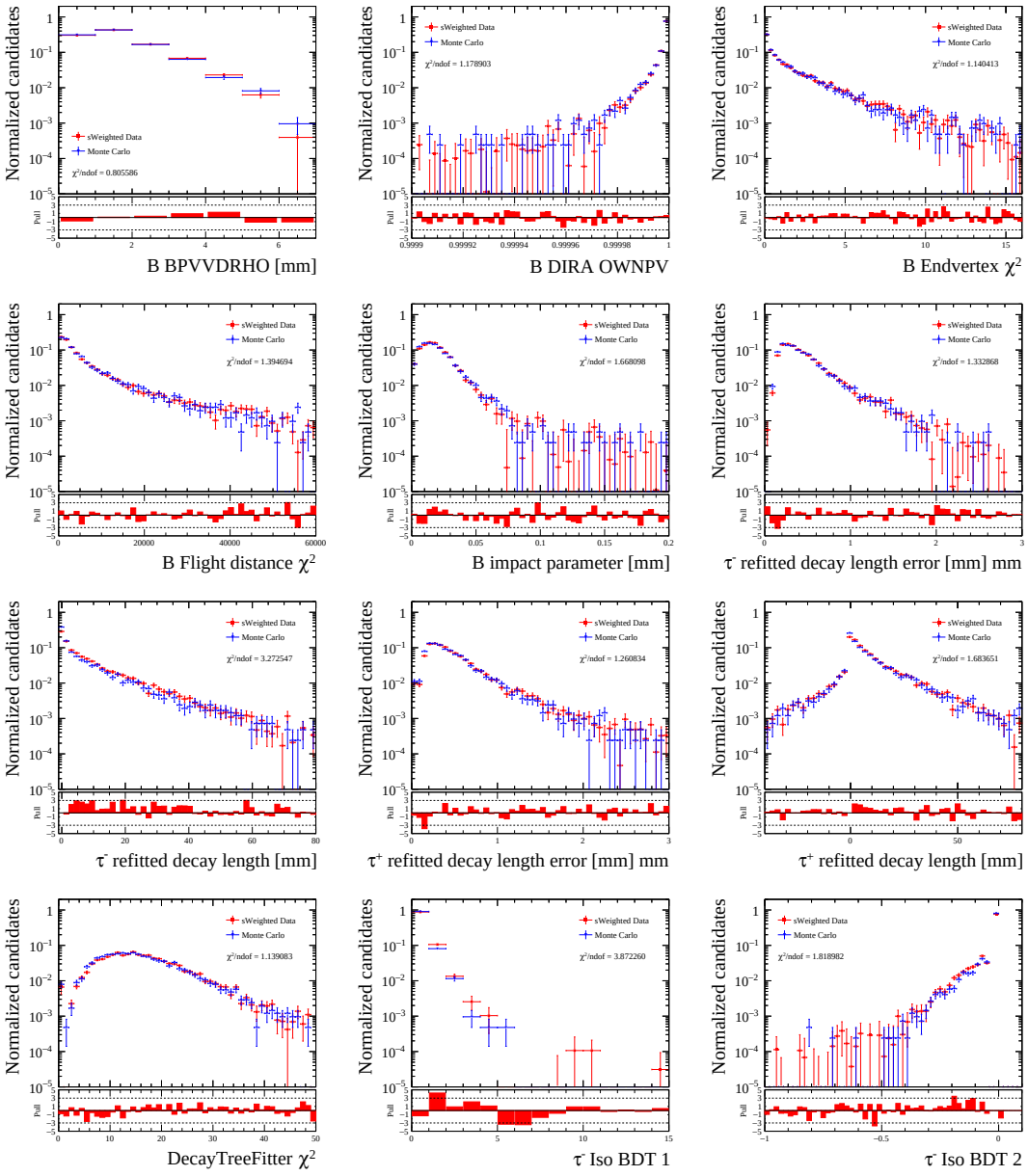


Figure C.7: Data-MC comparison for the 2012 dataset using the $B^0 \to D^- D_s^+$ channel (part 1).

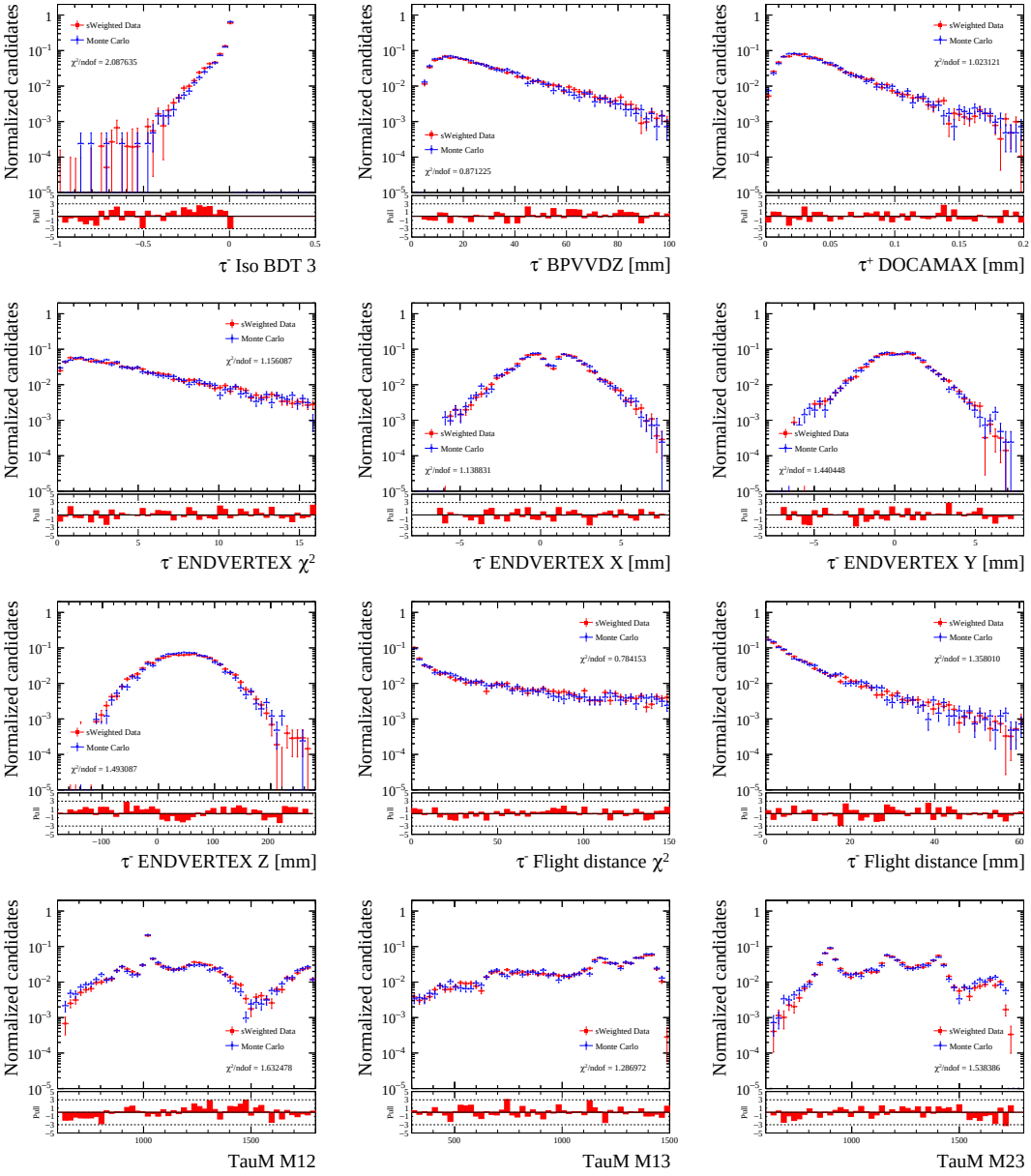


Figure C.8: Data-MC comparison for the 2012 dataset using the $B^0 \to D^- D_s^+$ channel (part 2).

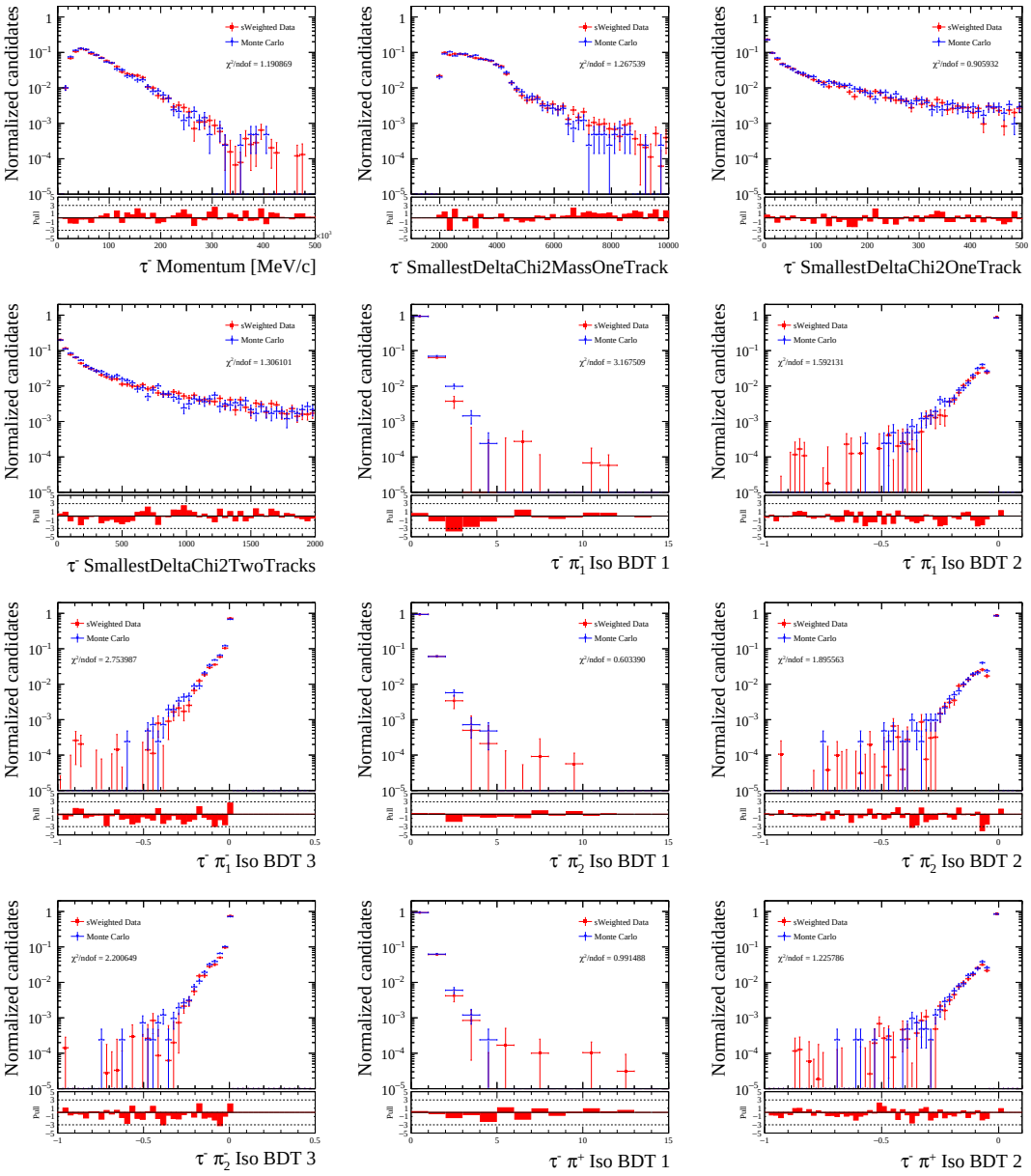


Figure C.9: Data-MC comparison for the 2012 dataset using the $B^0 \to D^- D_s^+$ channel (part 3).

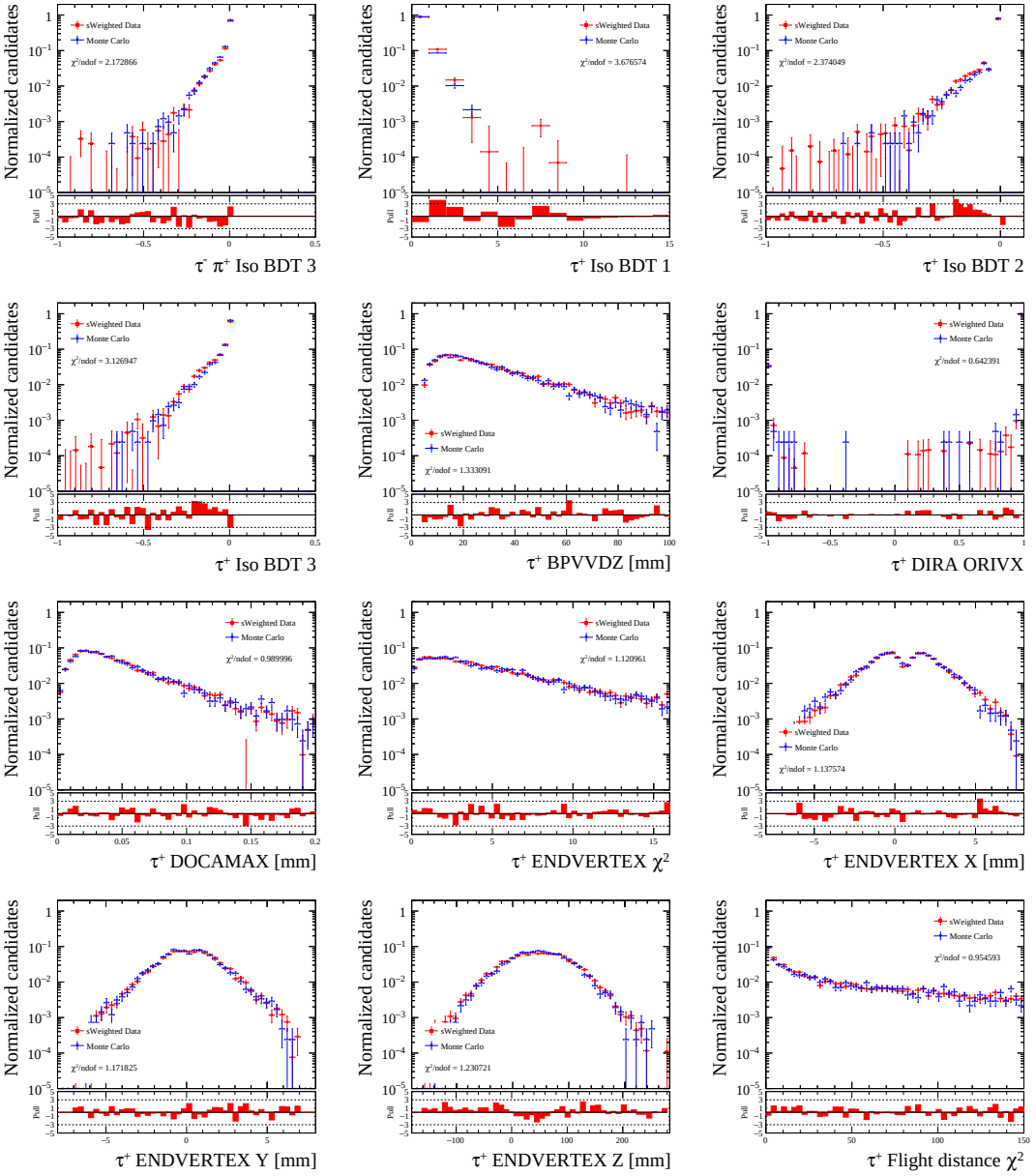


Figure C.10: Data-MC comparison for the 2012 dataset using the $B^0 \to D^- D_s^+$ channel (part 4).

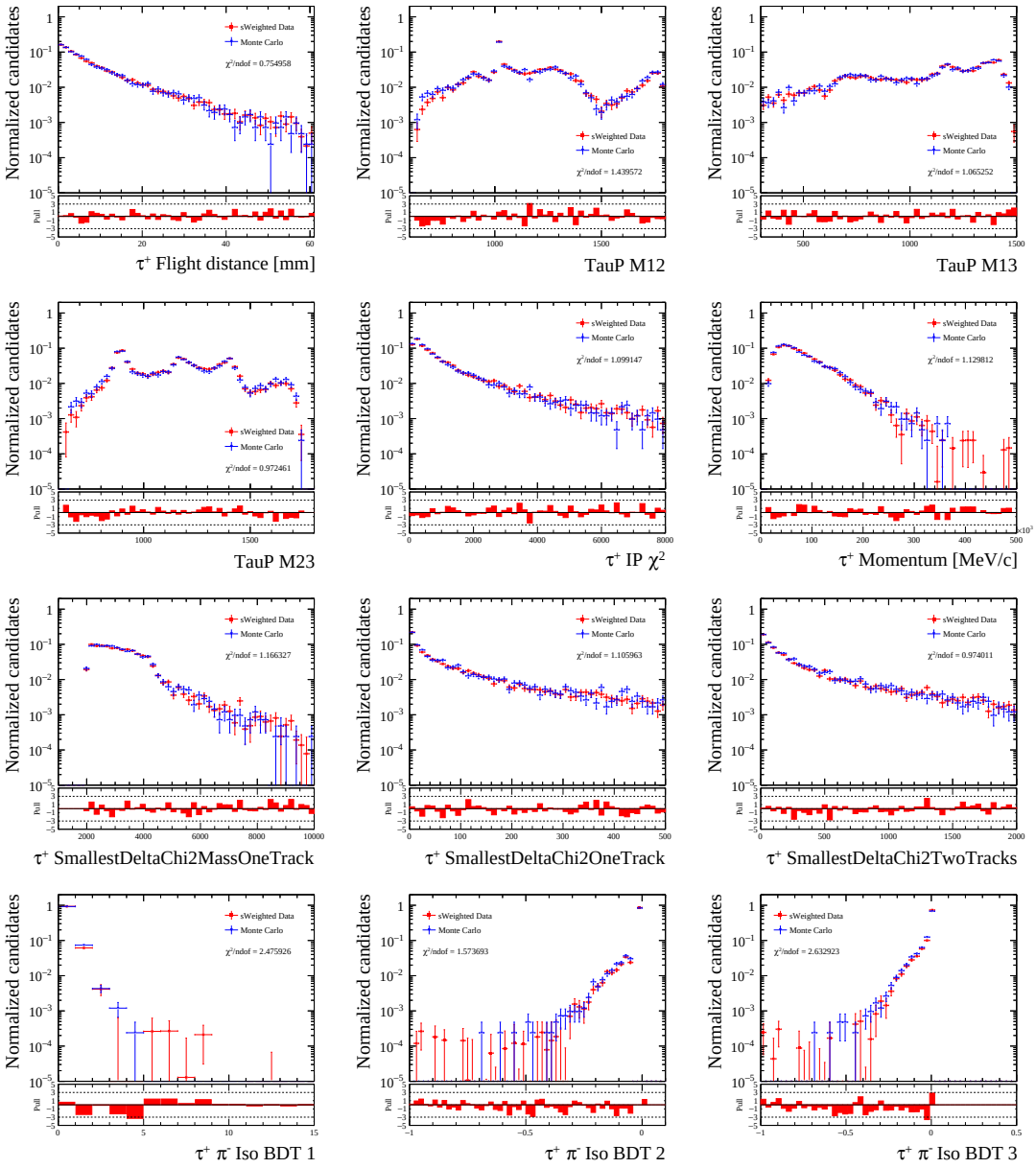


Figure C.11: Data-MC comparison for the 2012 dataset using the $B^0 \to D^- D_s^+$ channel (part 5).

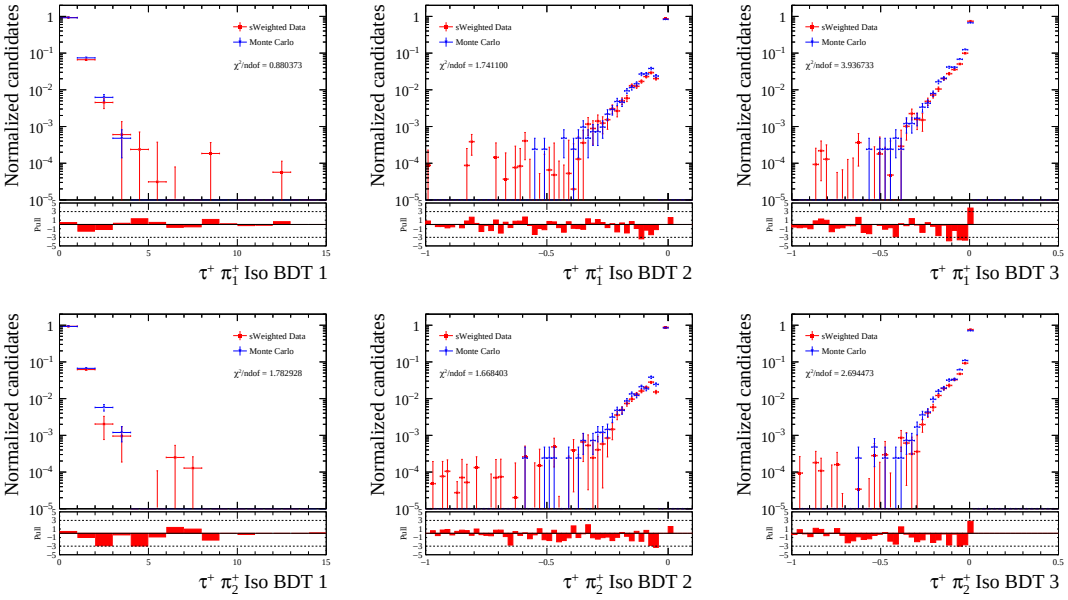


Figure C.12: Data-MC comparison for the 2012 dataset using the $B^0 \to D^- D_s^+$ channel (part 6).

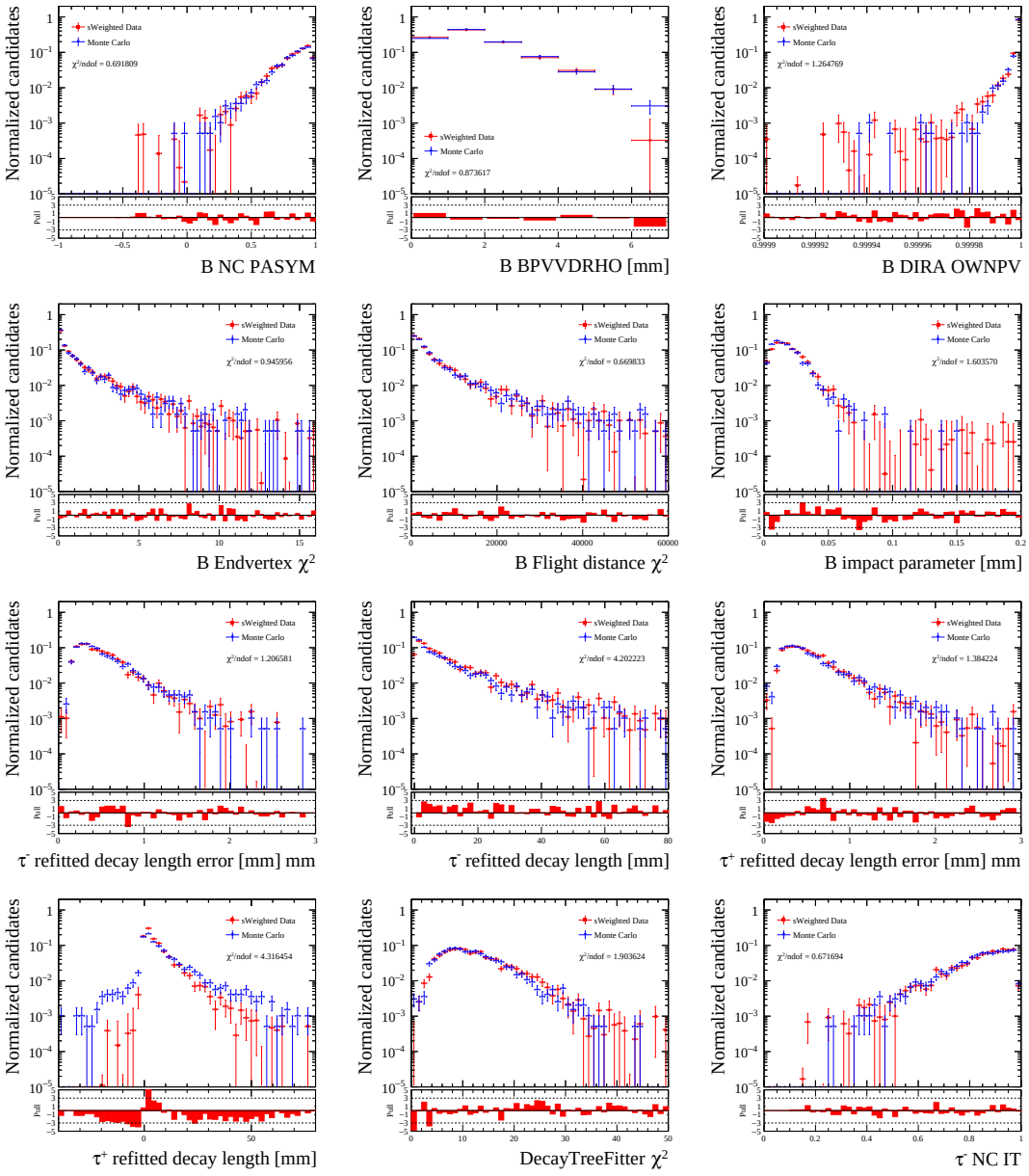


Figure C.13: Data-MC comparison for the 2015 dataset using the $B^0 \to D^- D_s^+$ channel (part 1).

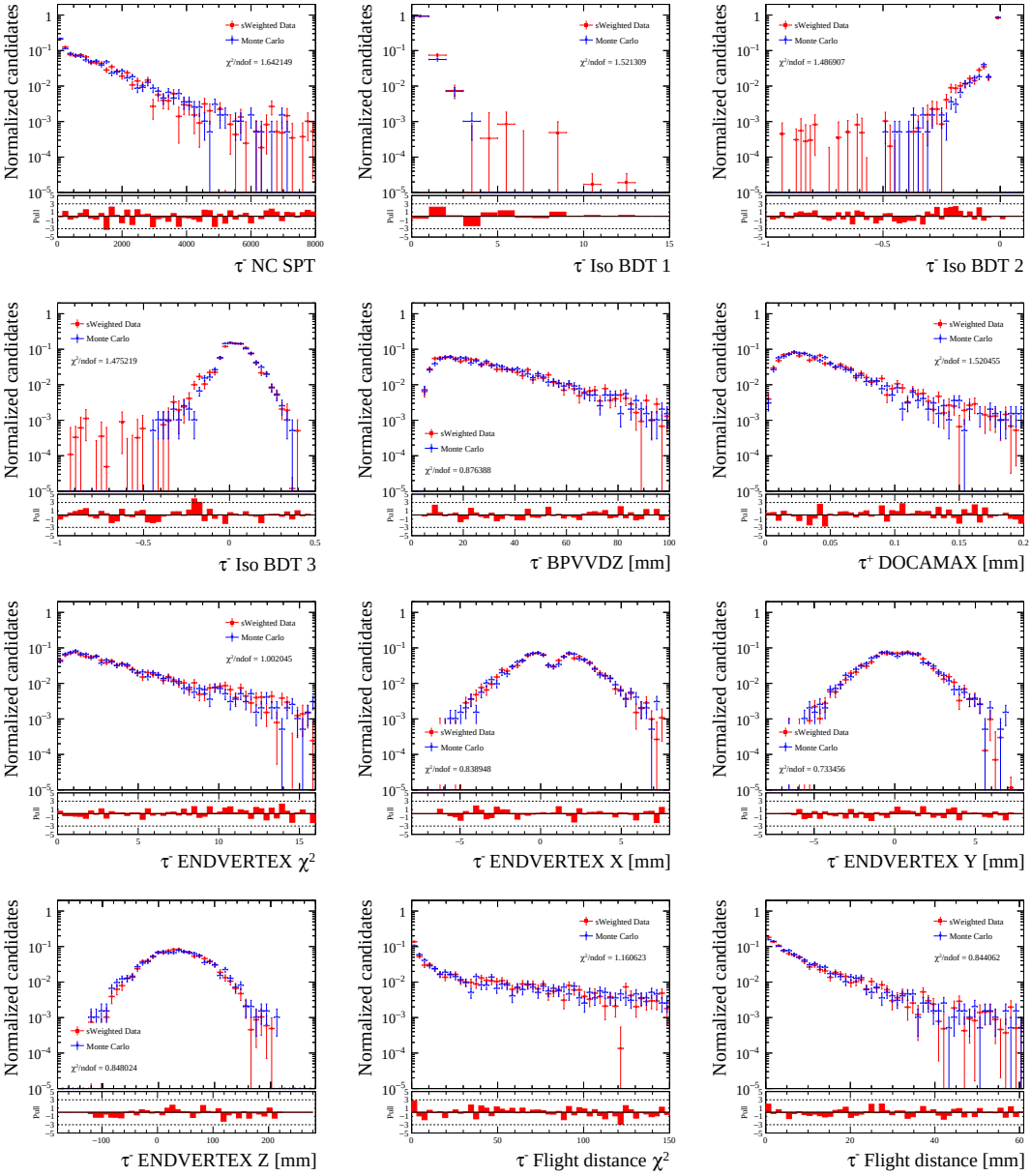


Figure C.14: Data-MC comparison for the 2015 dataset using the $B^0 \to D^- D_s^+$ channel (part 2).

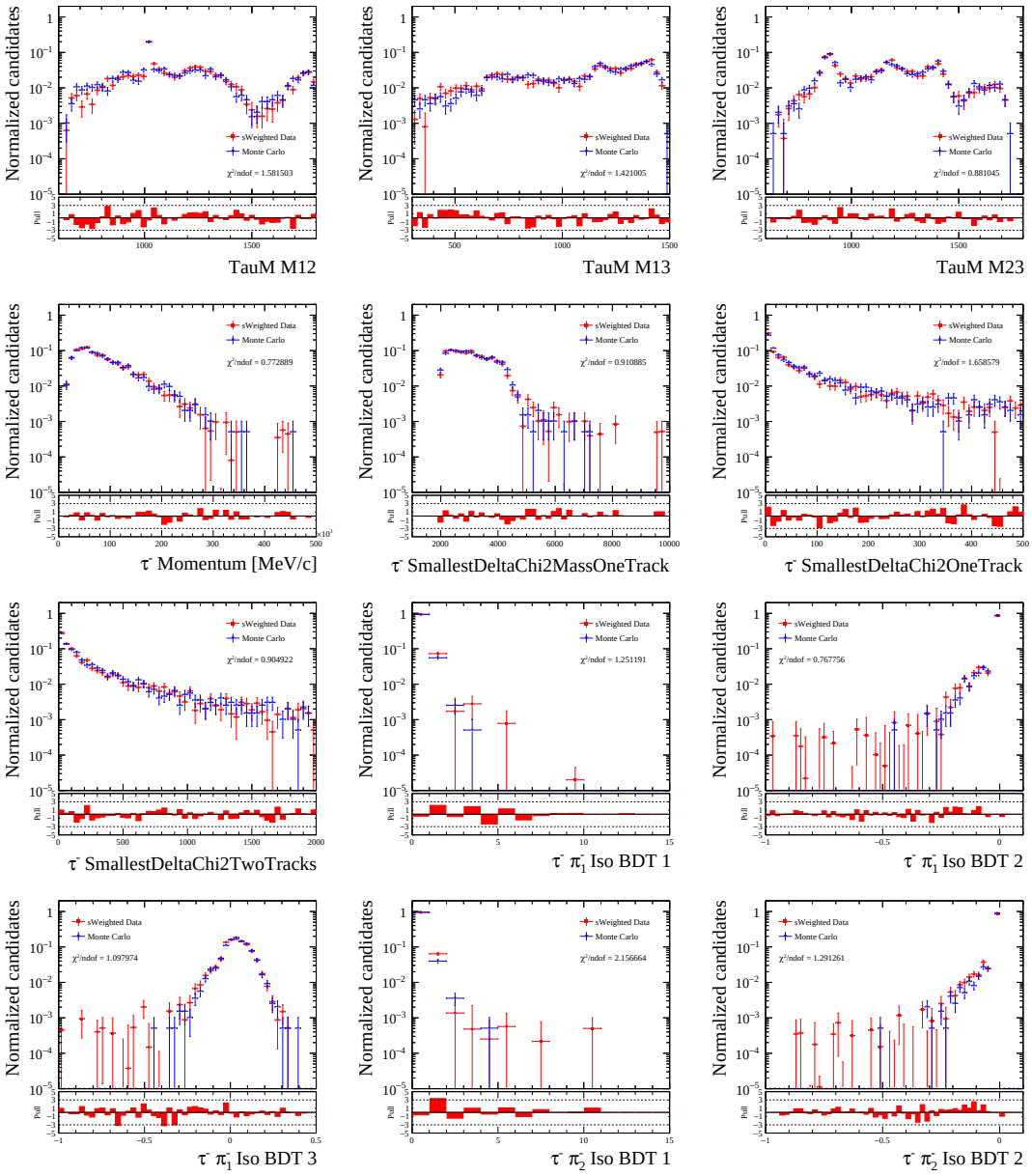


Figure C.15: Data-MC comparison for the 2015 dataset using the $B^0 \to D^- D_s^+$ channel (part 3).

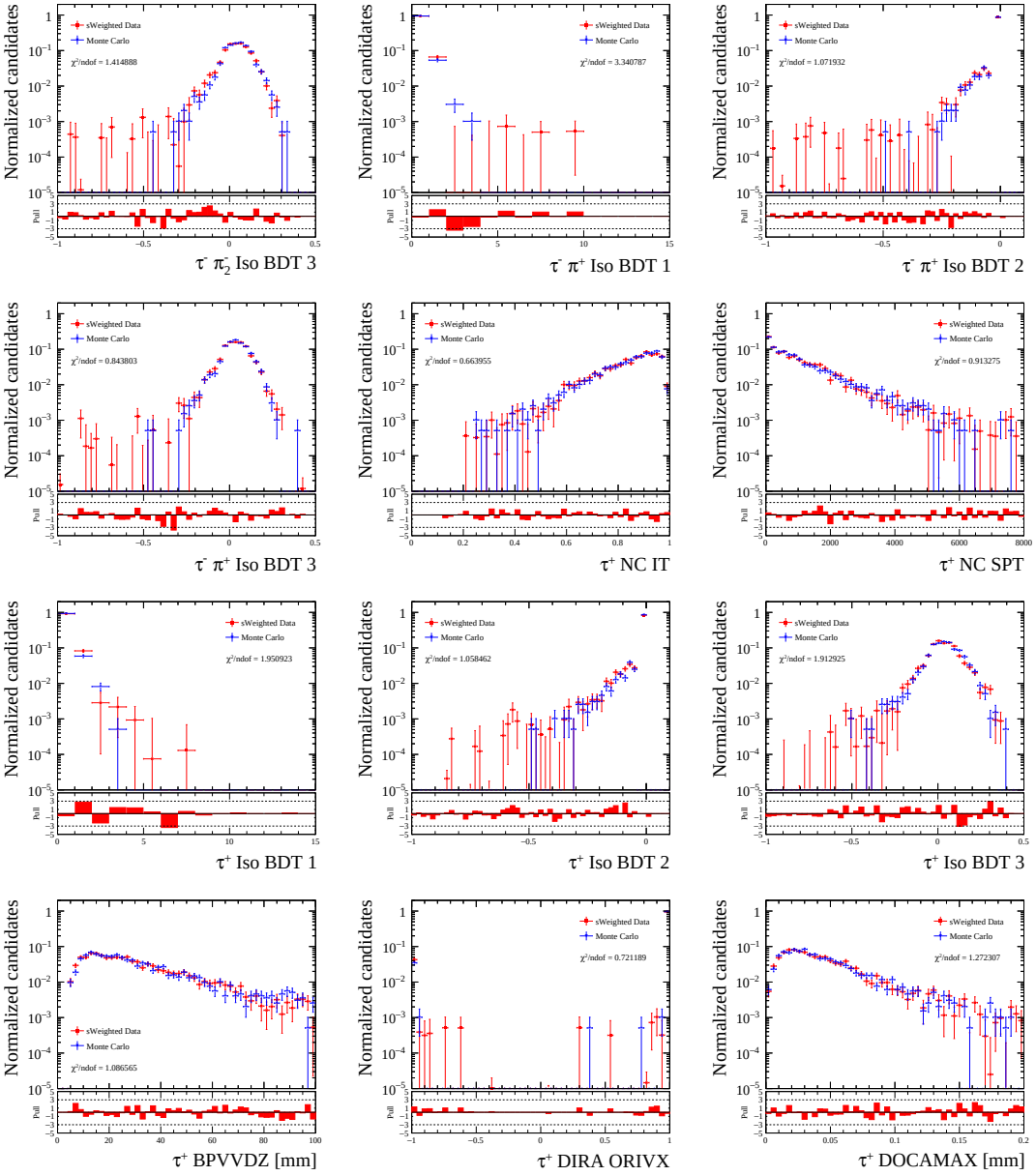


Figure C.16: Data-MC comparison for the 2015 dataset using the $B^0 \to D^- D_s^+$ channel (part 4).

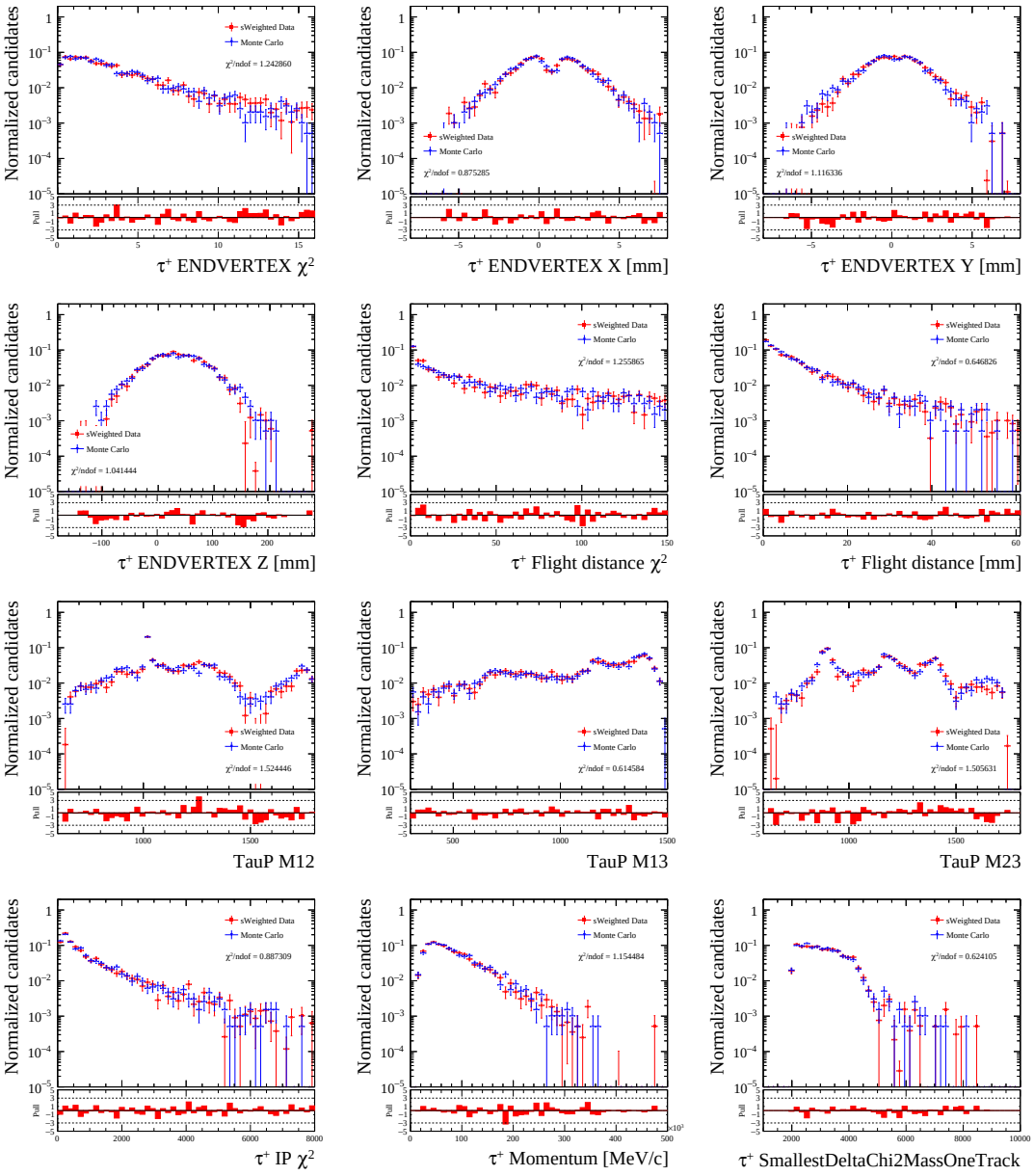


Figure C.17: Data-MC comparison for the 2015 dataset using the $B^0 \to D^- D_s^+$ channel (part 5).

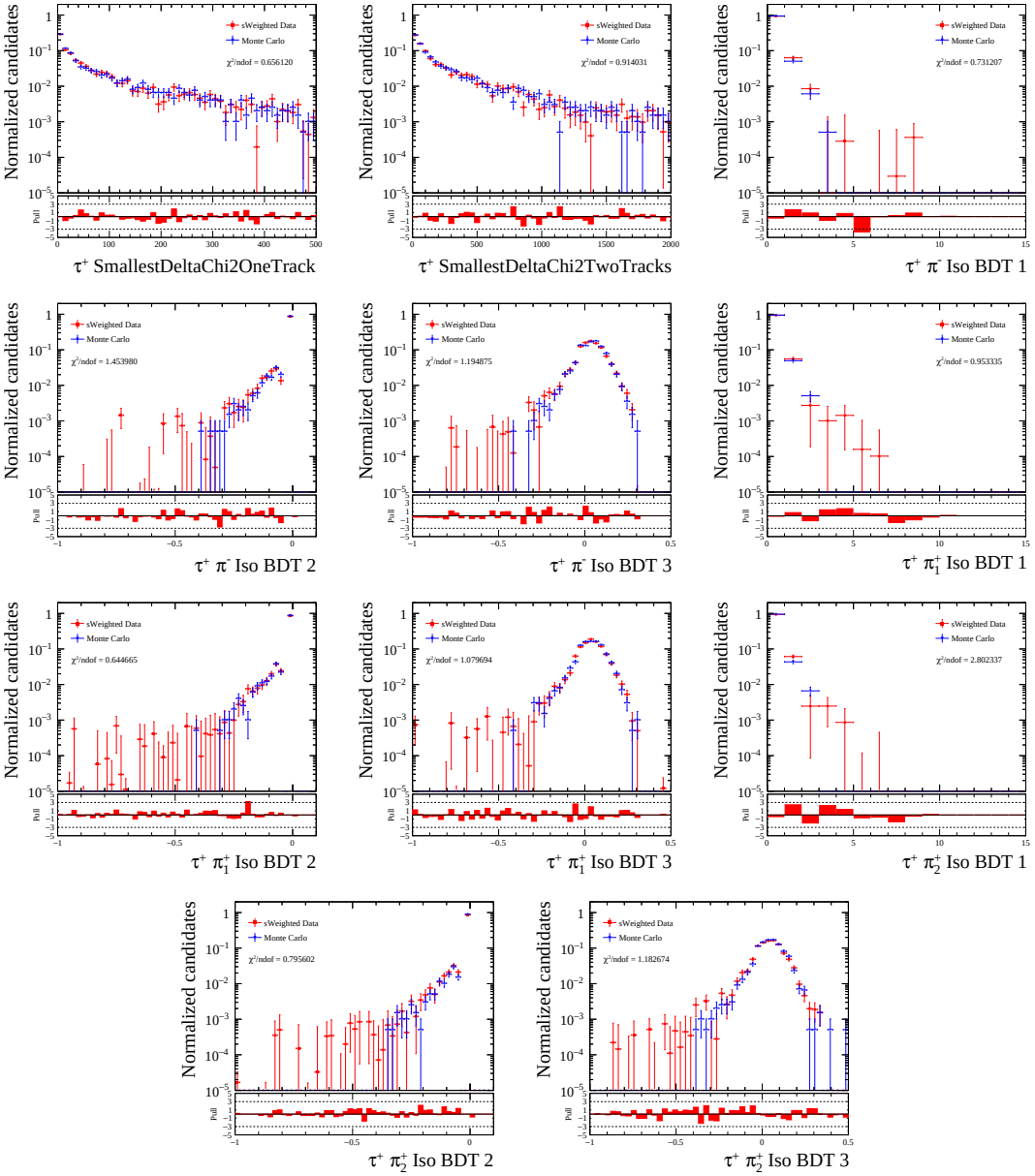


Figure C.18: Data-MC comparison for the 2015 dataset using the $B^0 \to D^- D_s^+$ channel (part 6).

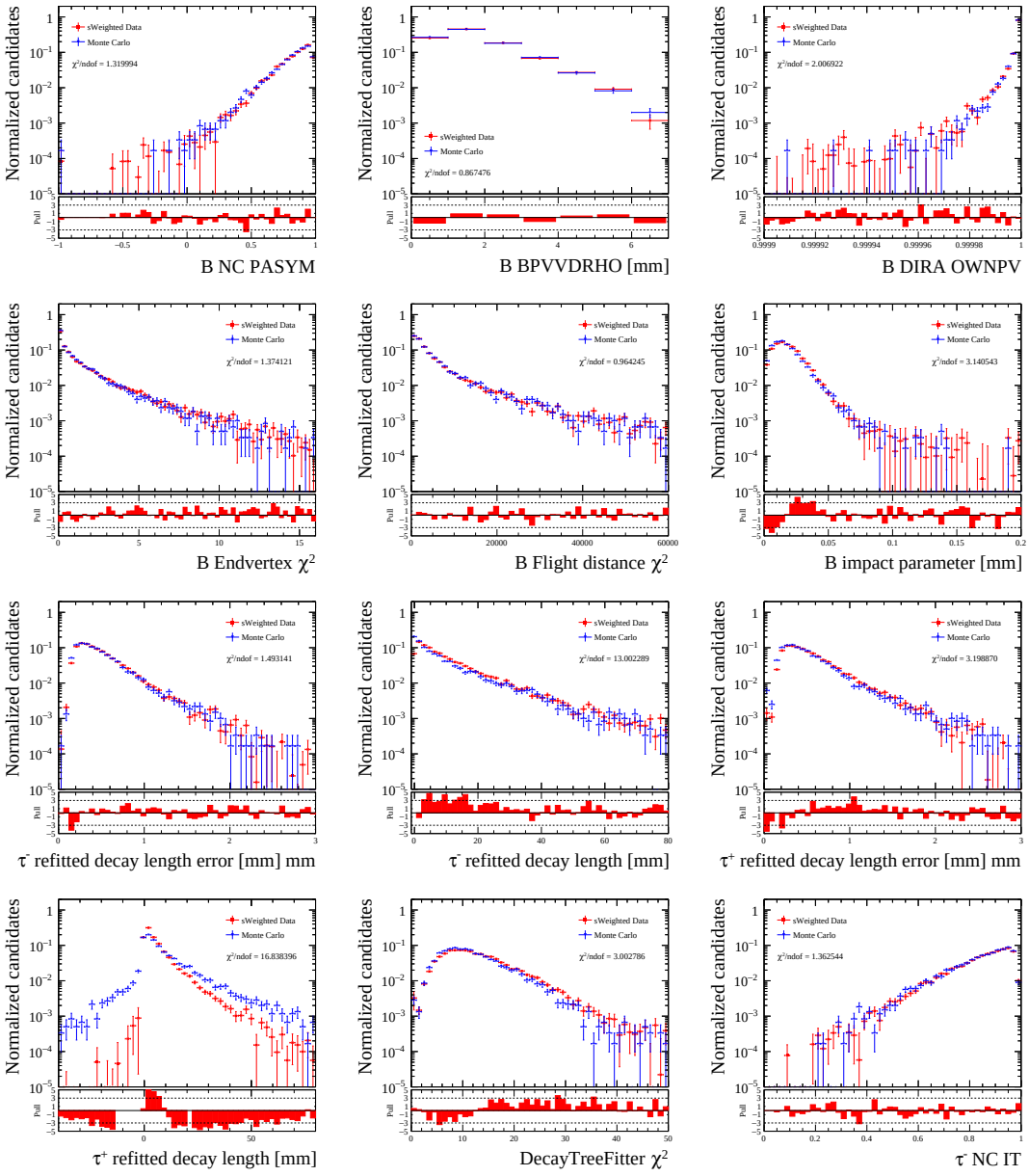


Figure C.19: Data-MC comparison for the 2016 dataset using the $B^0 \to D^- D_s^+$ channel (part 1).

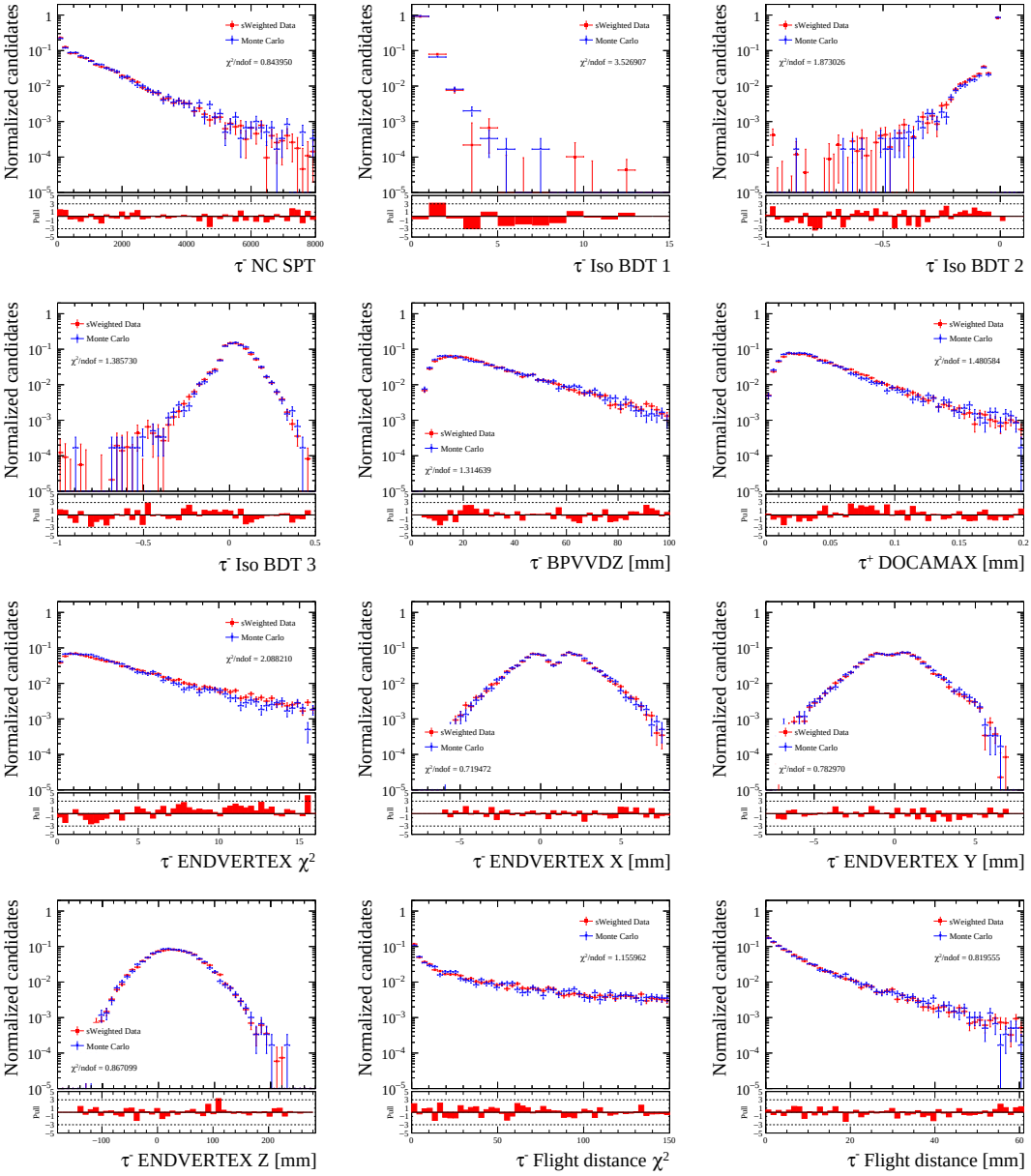


Figure C.20: Data-MC comparison for the 2016 dataset using the $B^0 \to D^- D_s^+$ channel (part 2).

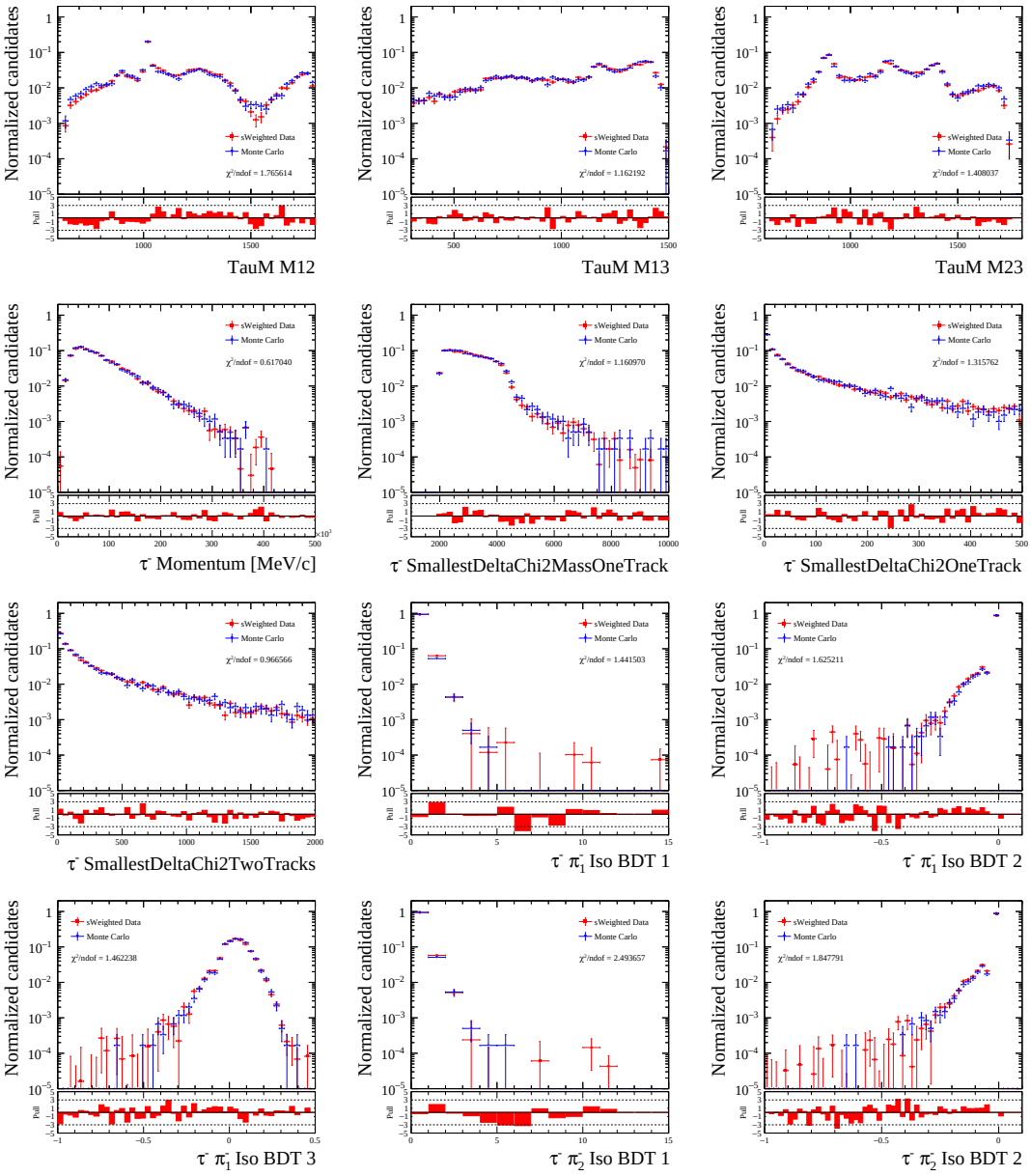


Figure C.21: Data-MC comparison for the 2016 dataset using the $B^0 \to D^- D_s^+$ channel (part 3).

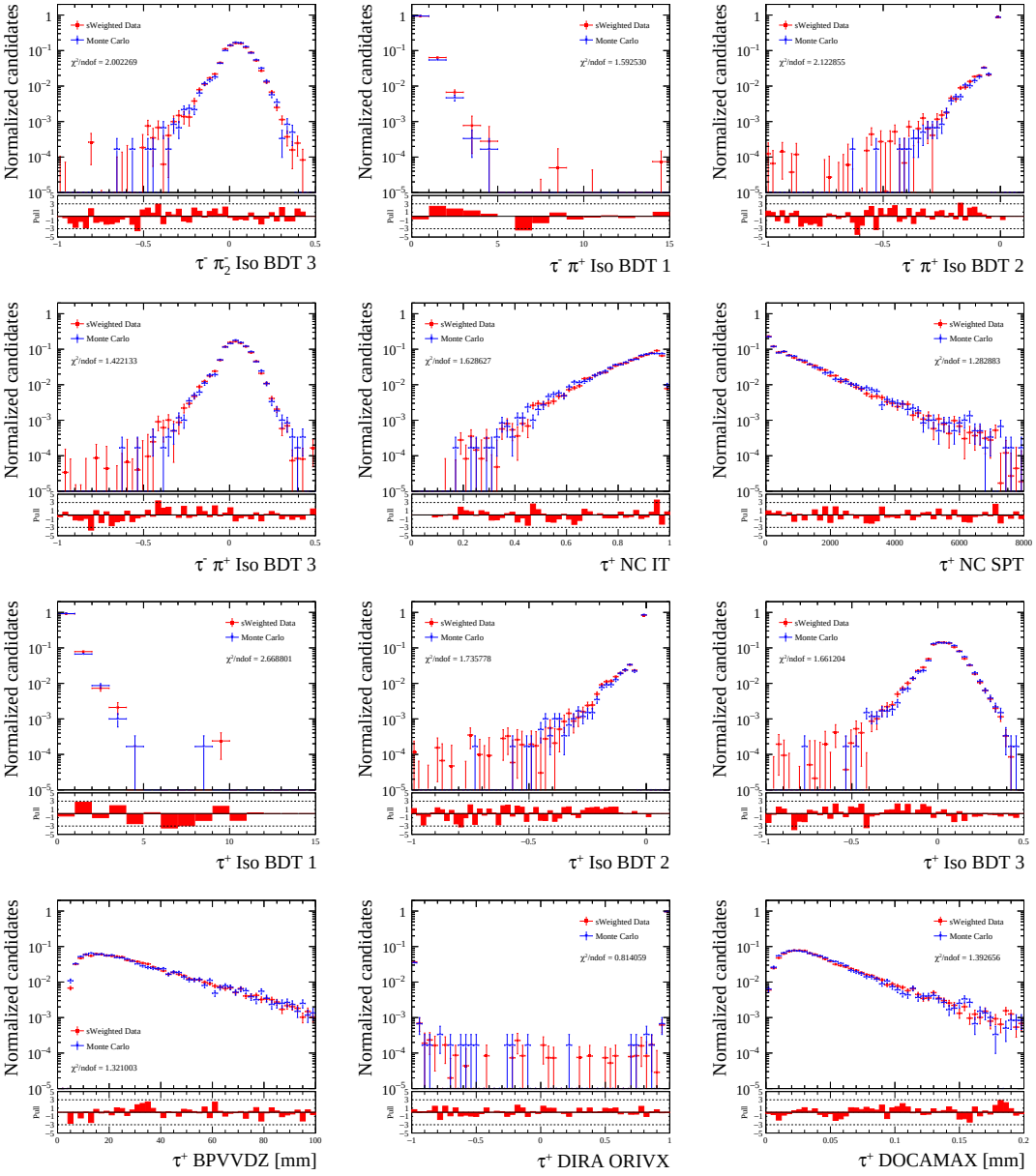


Figure C.22: Data-MC comparison for the 2016 dataset using the $B^0 \to D^- D_s^+$ channel (part 4).

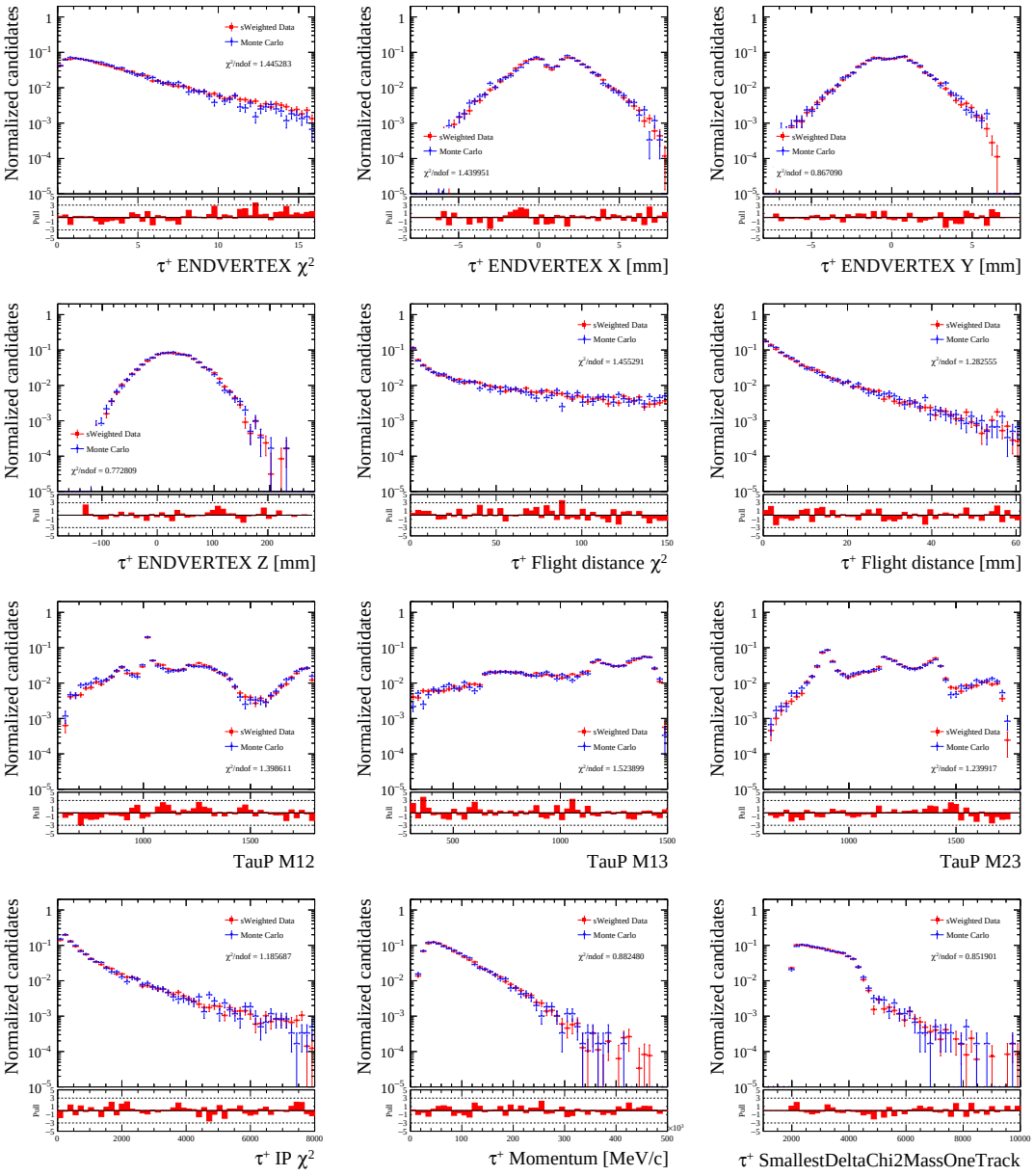


Figure C.23: Data-MC comparison for the 2016 dataset using the $B^0 \to D^- D_s^+$ channel (part 5).

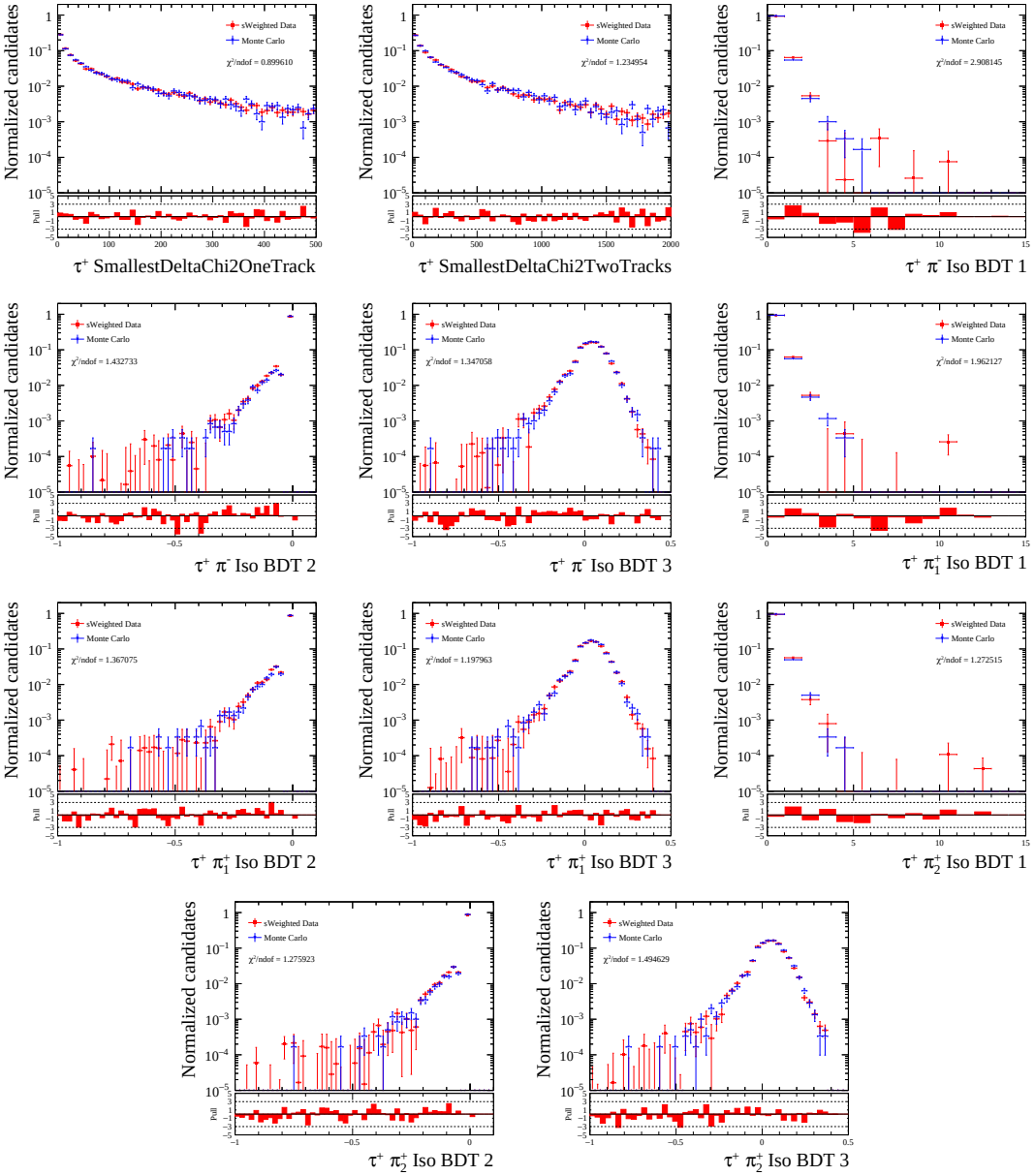


Figure C.24: Data-MC comparison for the 2016 dataset using the $B^0 \to D^- D_s^+$ channel (part 6).

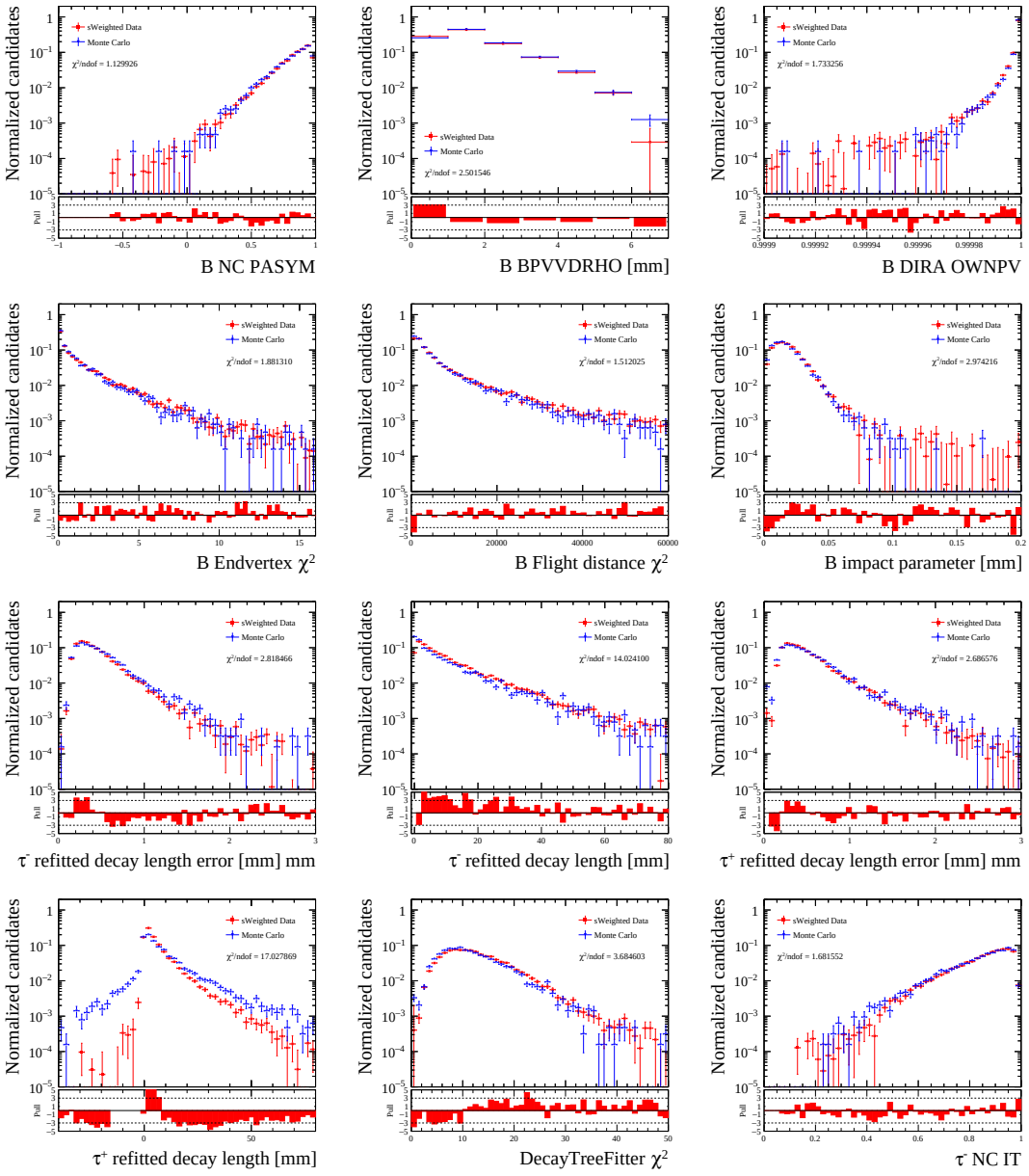


Figure C.25: Data-MC comparison for the 2017 dataset using the $B^0 \to D^- D_s^+$ channel (part 1).

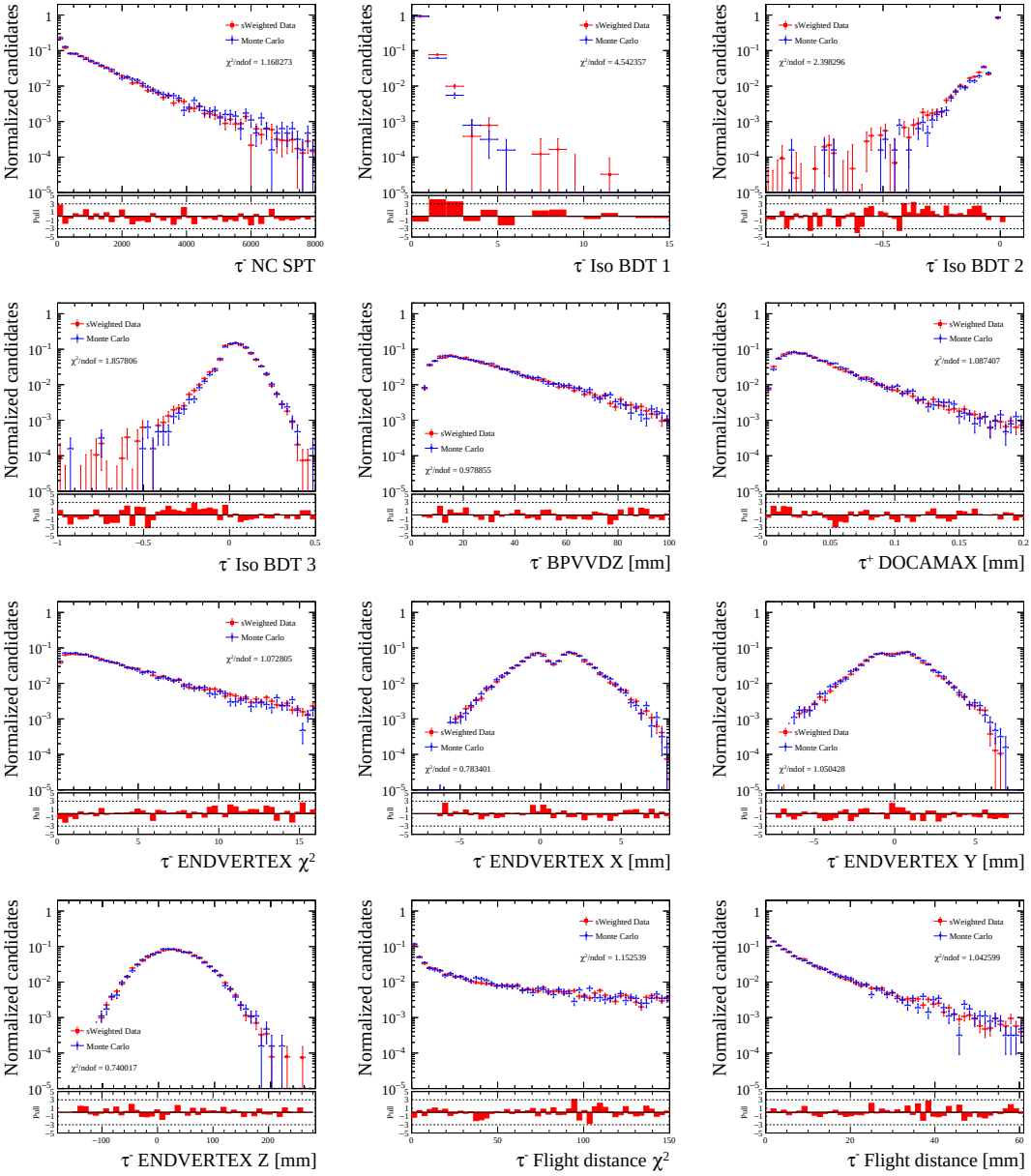


Figure C.26: Data-MC comparison for the 2017 dataset using the $B^0 \to D^- D_s^+$ channel (part 2).

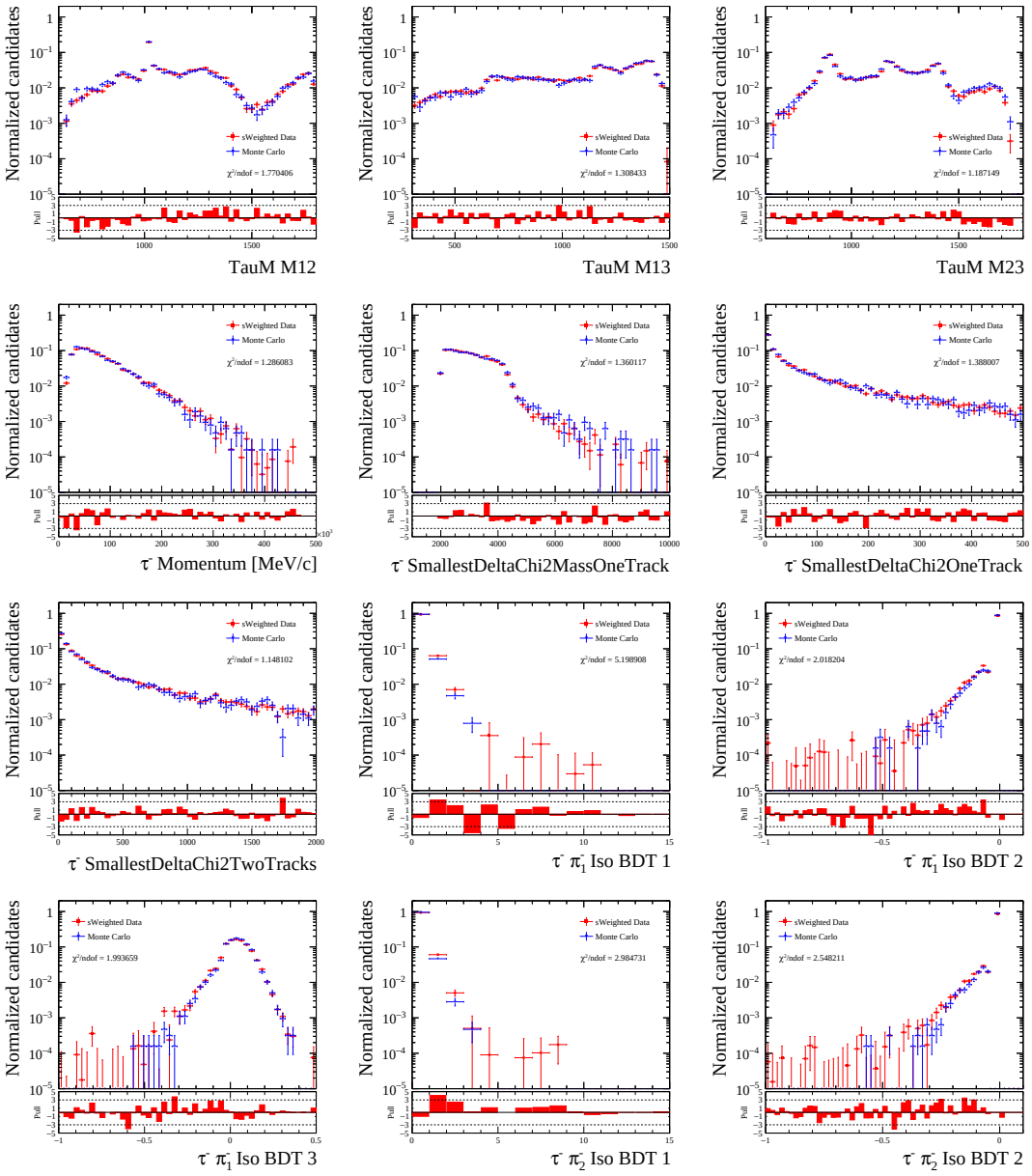


Figure C.27: Data-MC comparison for the 2017 dataset using the $B^0 \to D^- D_s^+$ channel (part 3).

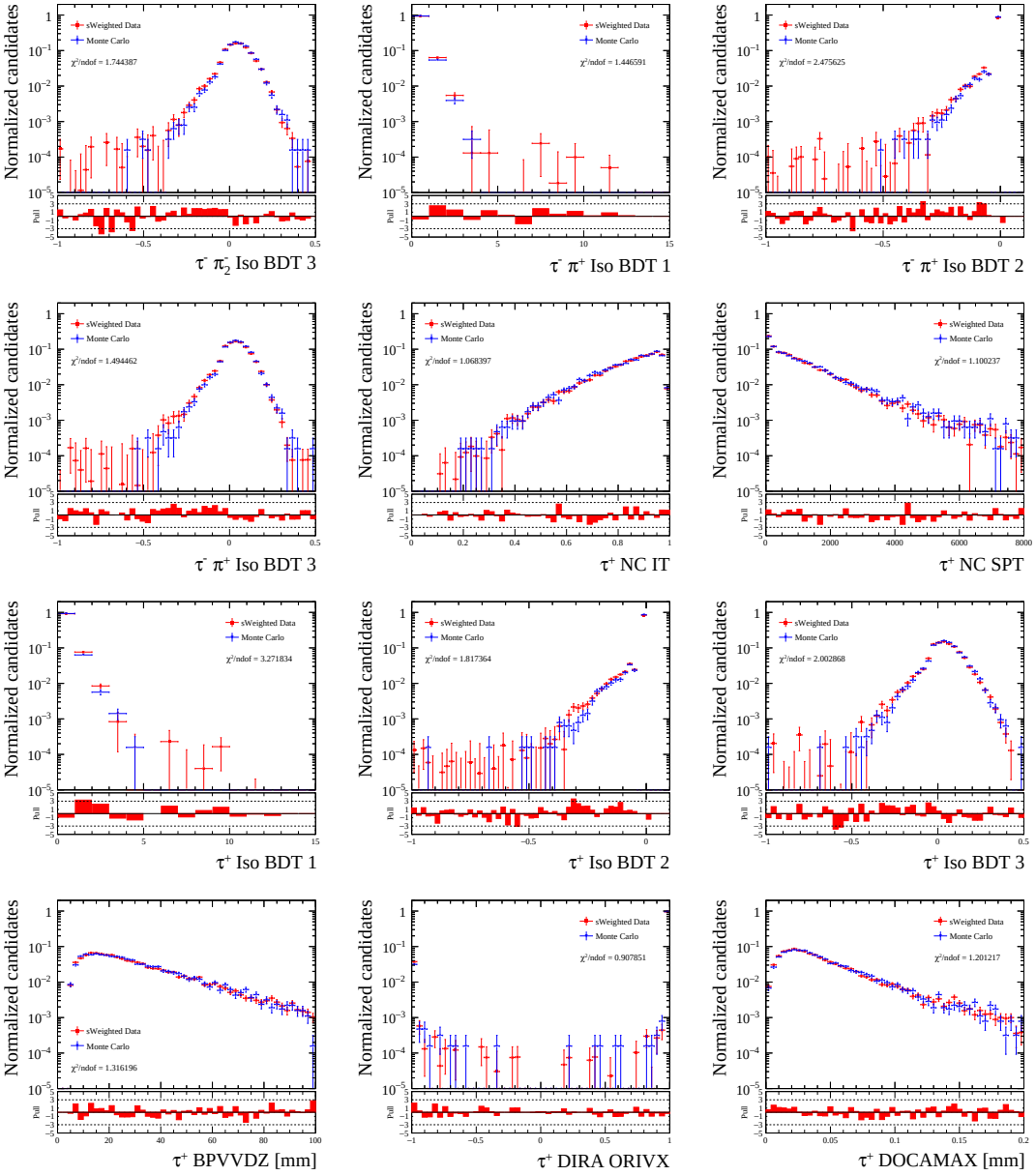


Figure C.28: Data-MC comparison for the 2017 dataset using the $B^0 \to D^- D_s^+$ channel (part 4).

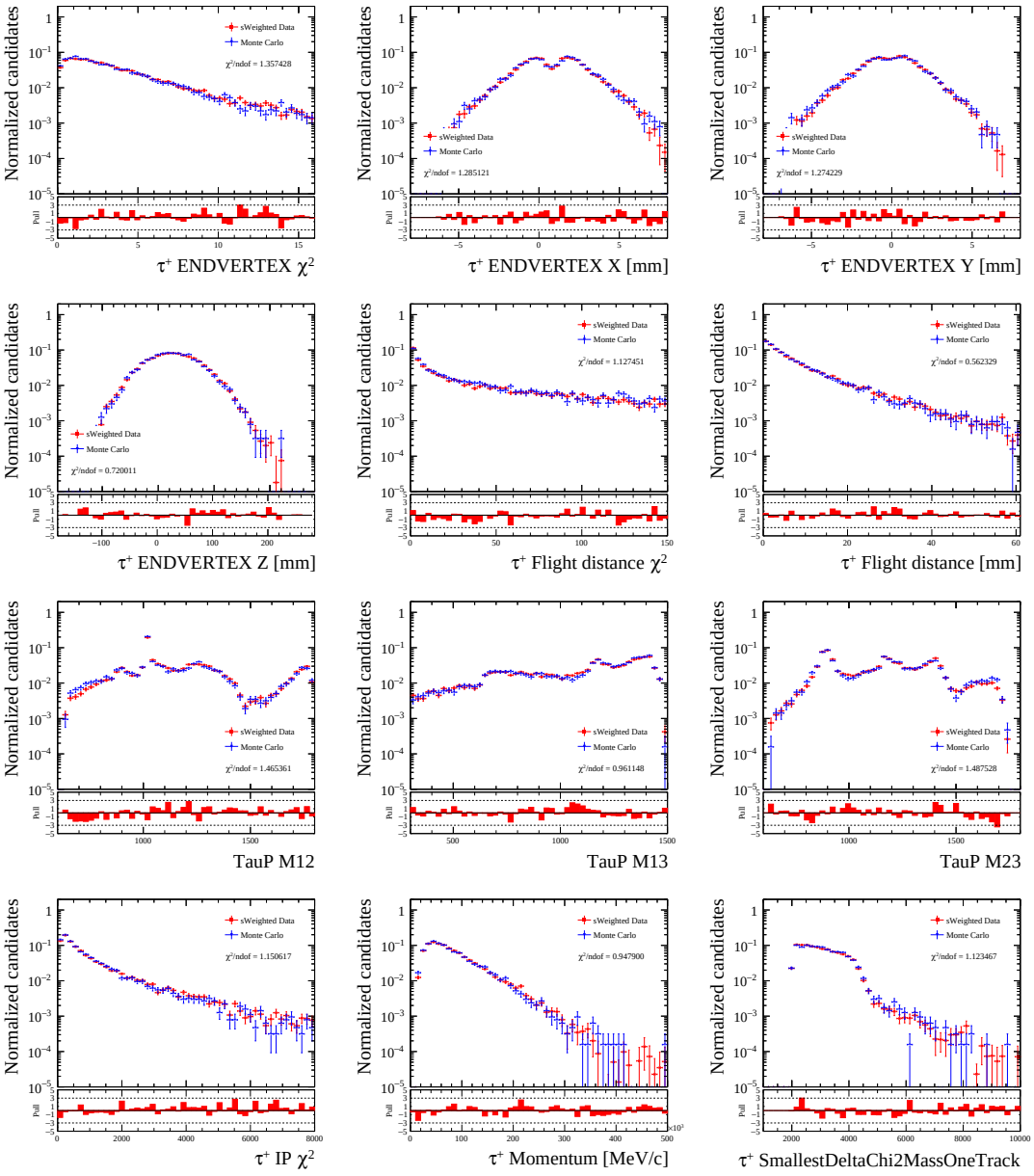


Figure C.29: Data-MC comparison for the 2017 dataset using the $B^0 \to D^- D_s^+$ channel (part 5).

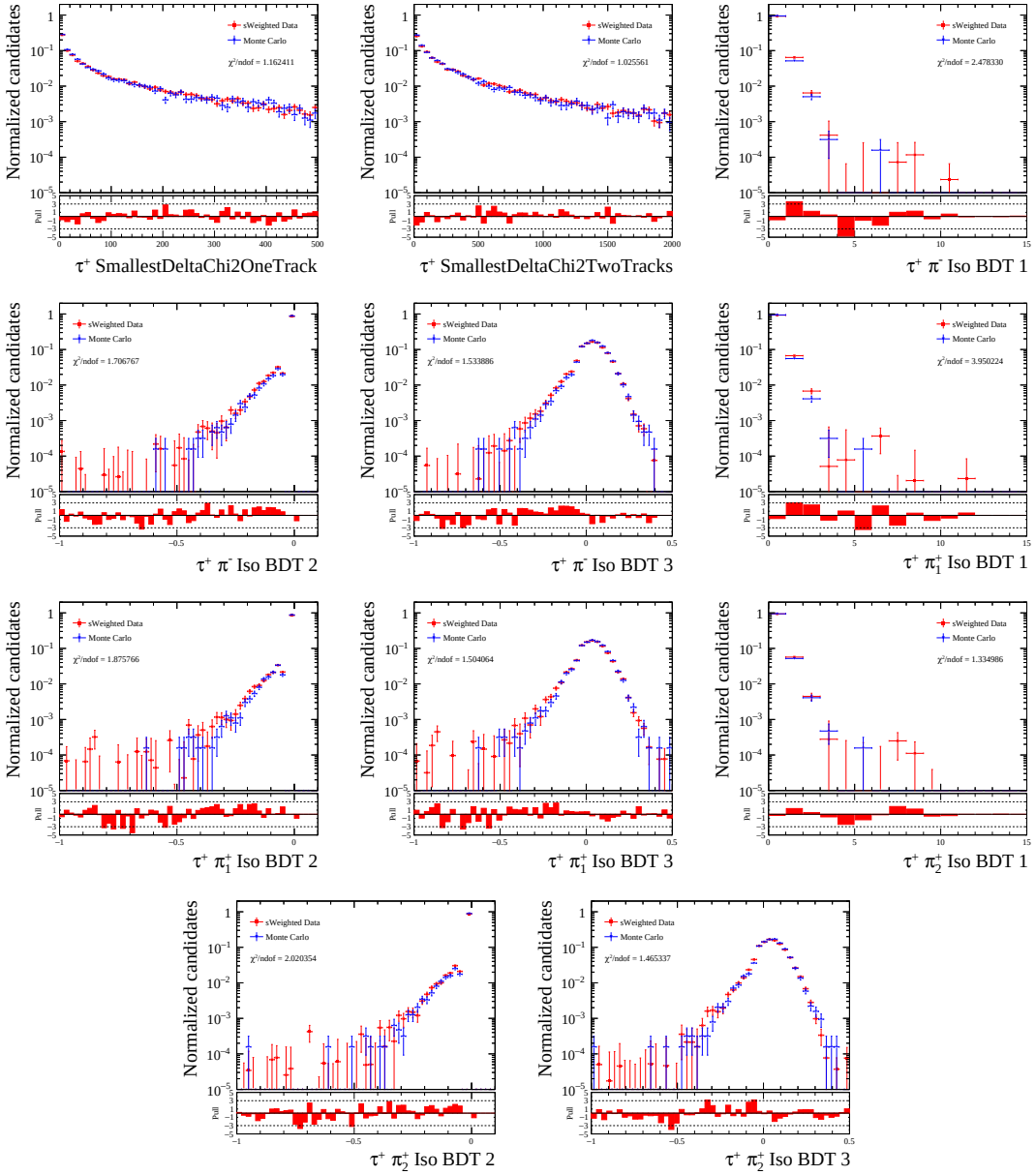


Figure C.30: Data-MC comparison for the 2017 dataset using the $B^0 \to D^- D_s^+$ channel (part 6).

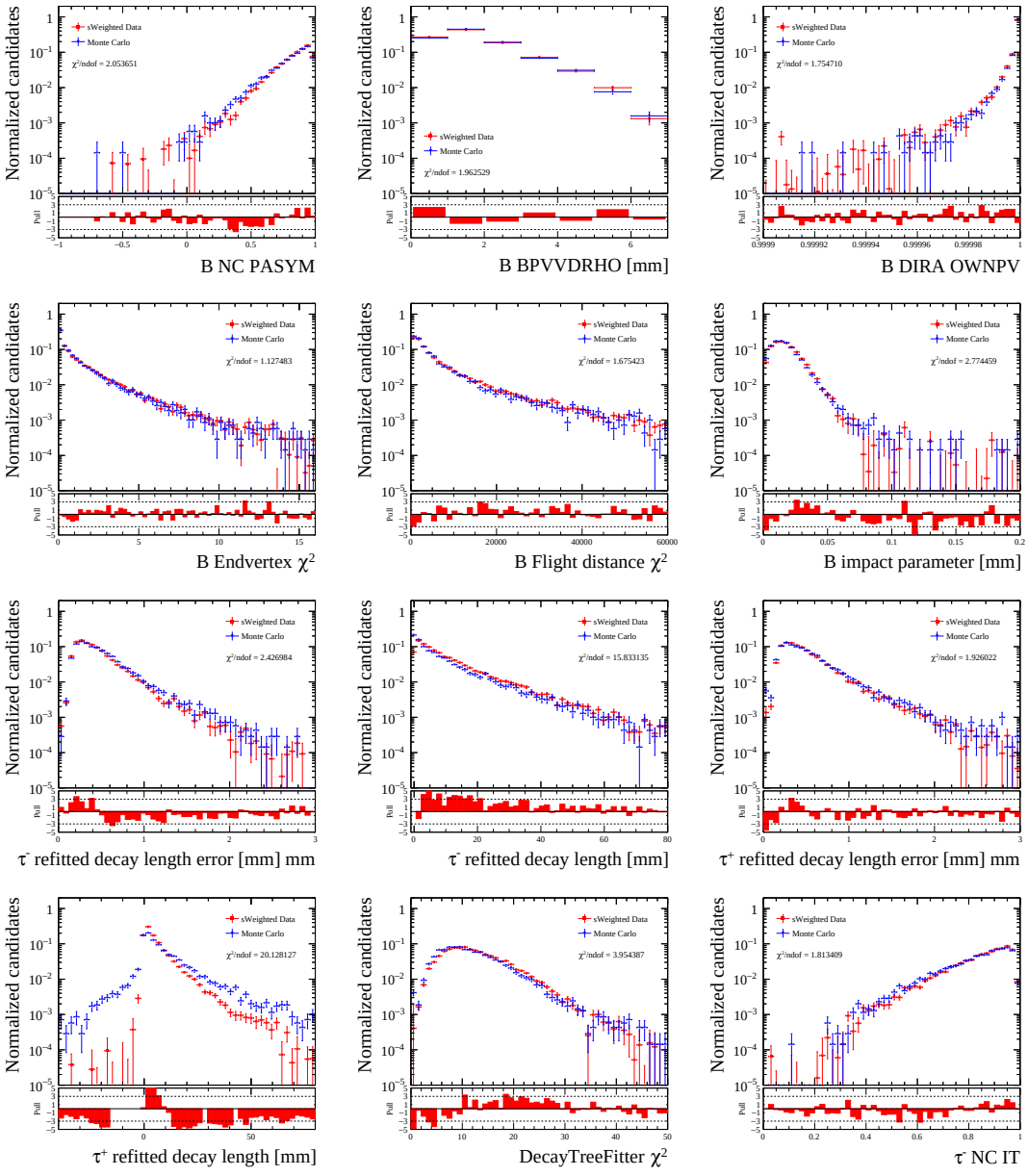


Figure C.31: Data-MC comparison for the 2018 dataset using the $B^0 \to D^- D_s^+$ channel (part 1).

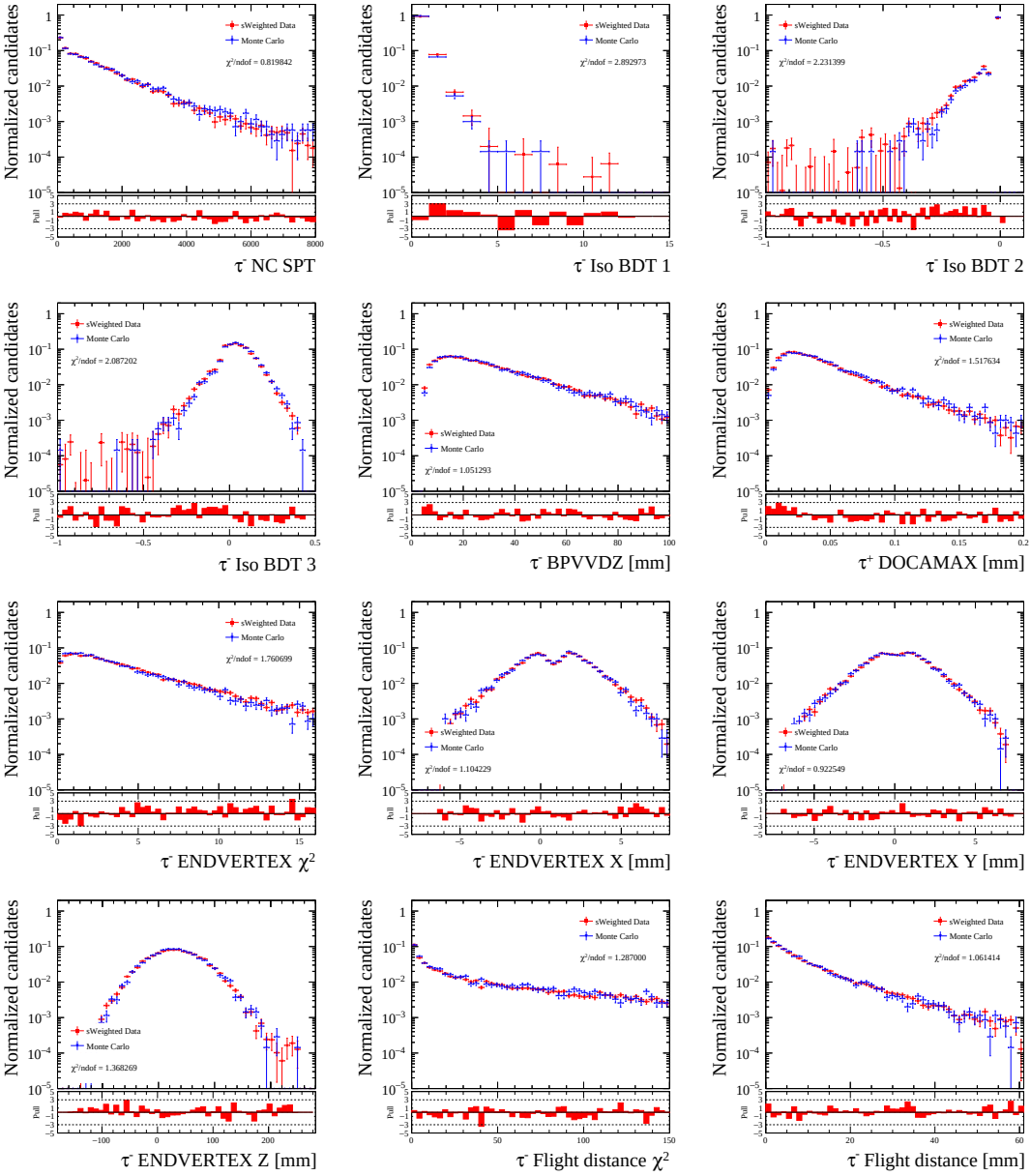


Figure C.32: Data-MC comparison for the 2018 dataset using the $B^0 \to D^- D_s^+$ channel (part 2).

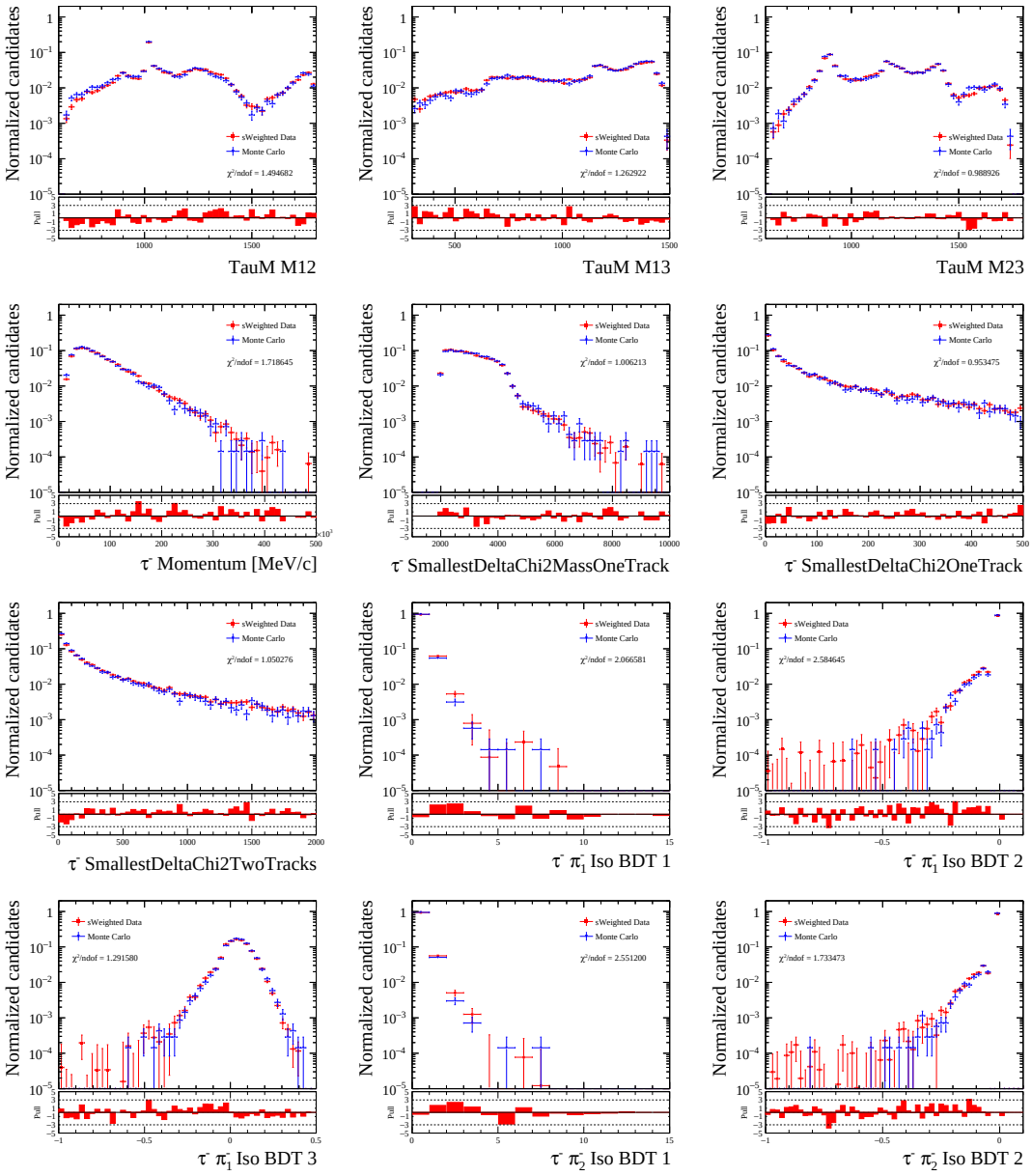


Figure C.33: Data-MC comparison for the 2018 dataset using the $B^0 \to D^- D_s^+$ channel (part 3).

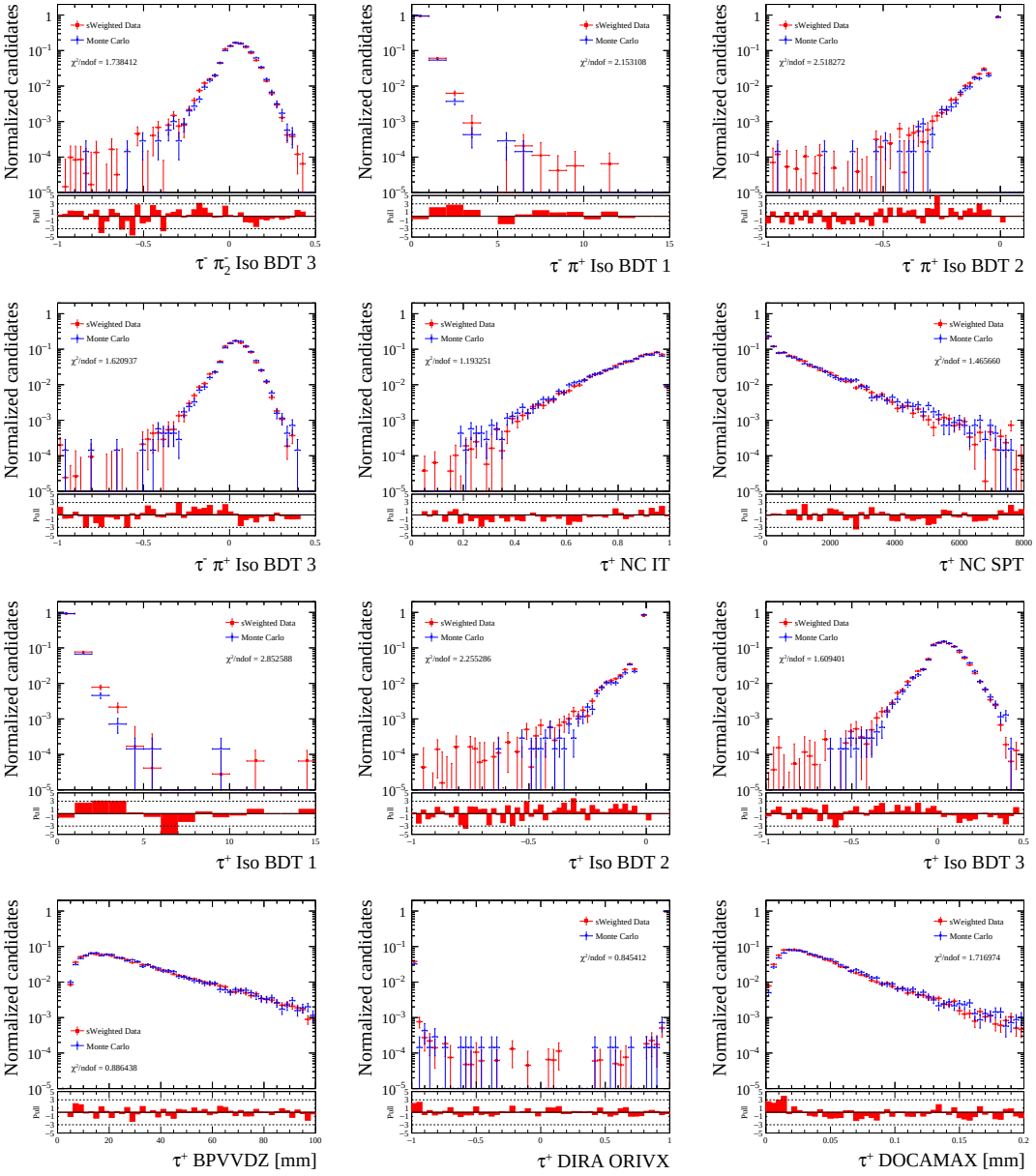


Figure C.34: Data-MC comparison for the 2018 dataset using the $B^0 \to D^- D_s^+$ channel (part 4).

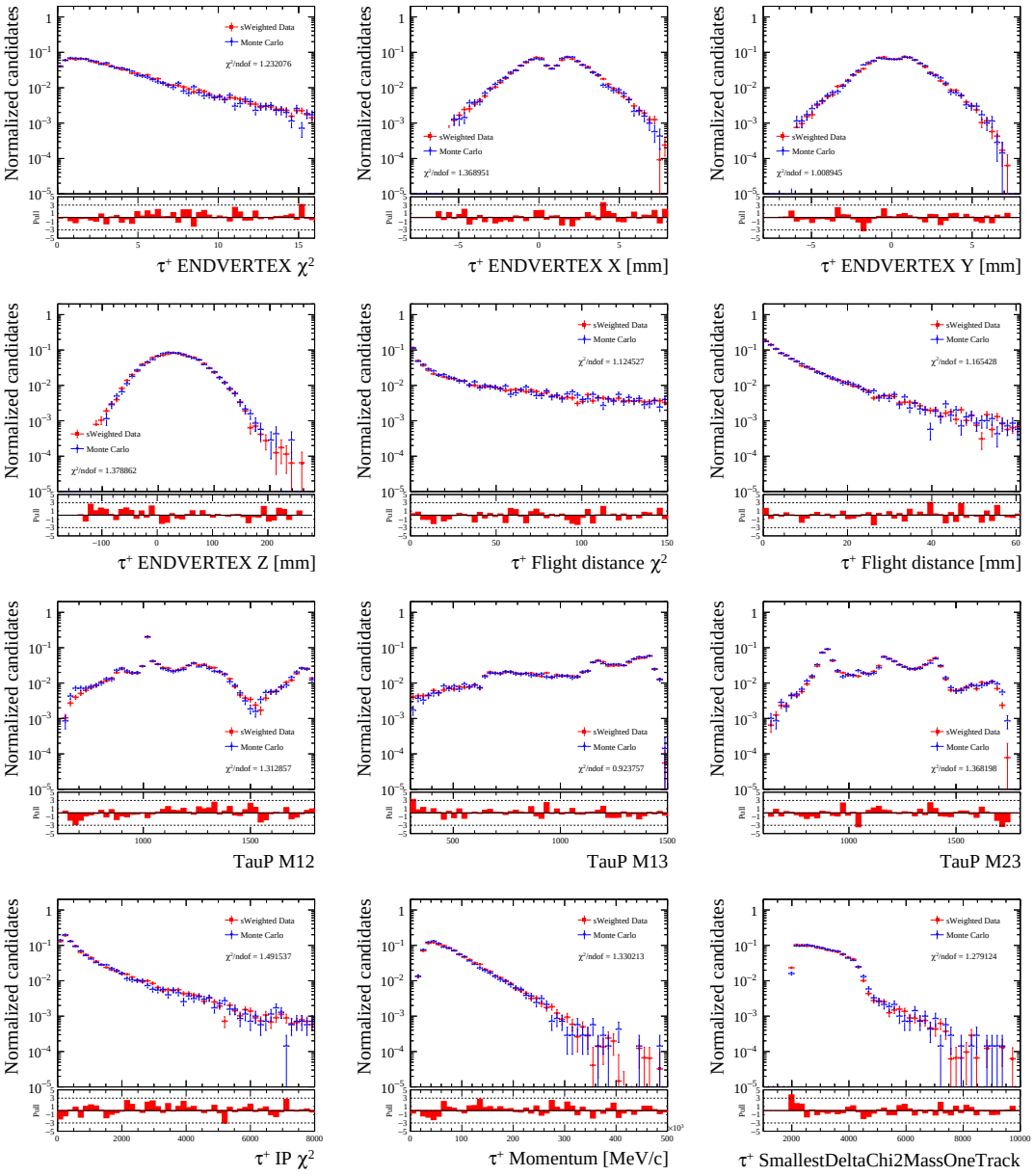


Figure C.35: Data-MC comparison for the 2018 dataset using the $B^0 \to D^- D_s^+$ channel (part 5).

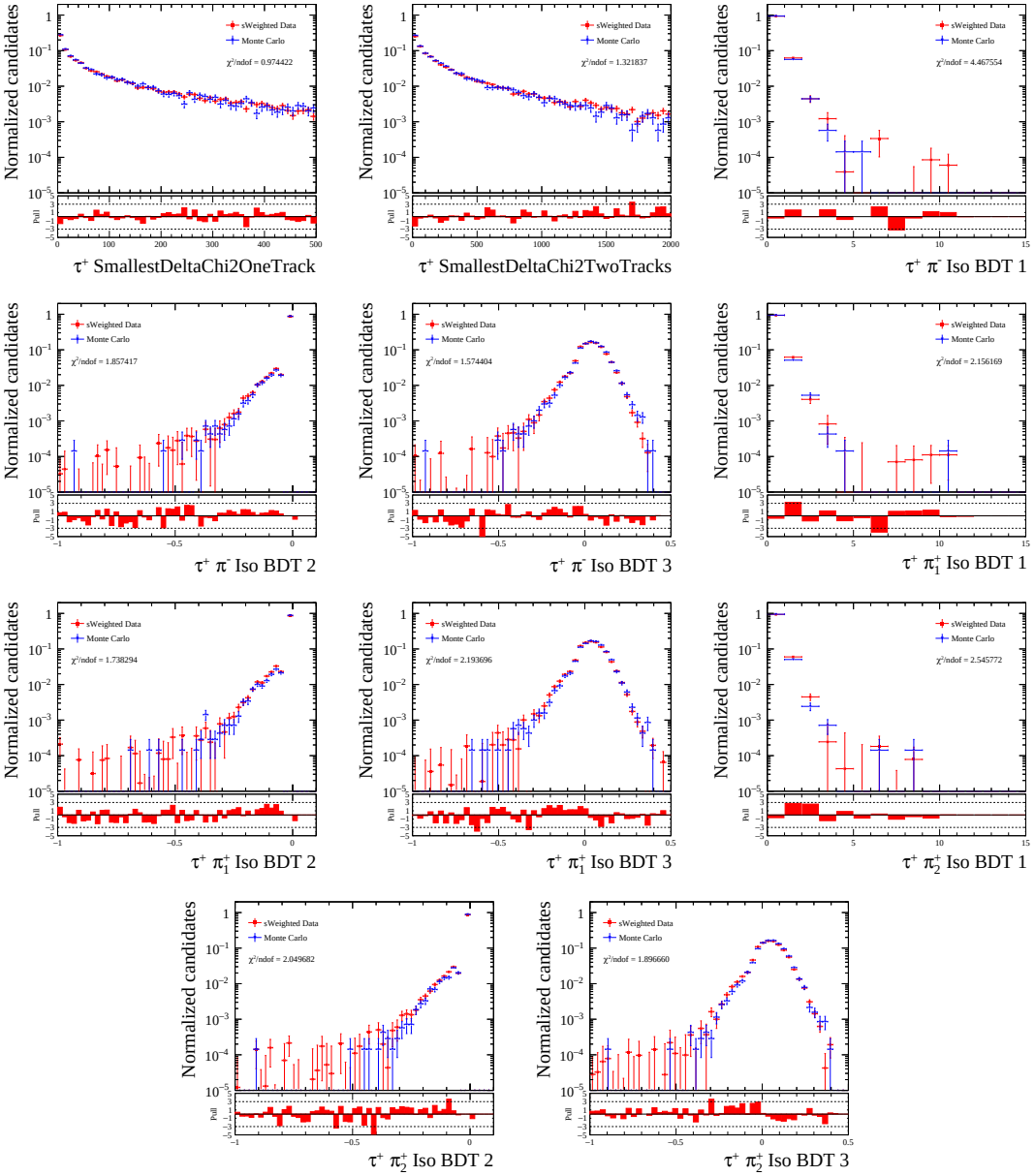


Figure C.36: Data-MC comparison for the 2018 dataset using the $B^0 \to D^- D_s^+$ channel (part 6).

Bibliography

[1] CERN. *https://home.cern/about/physics/dark-matter*. Accessed 17 January 2021 (cit. on pp. v, 5).

[2] CERN. *https://home.cern/science/physics/matter-antimatter-asymmetry-problem*. Accessed 17 January 2021 (cit. on pp. v, 5).

[3] Y. Fukuda et al. "Evidence for oscillation of atmospheric neutrinos". *Phys. Rev. Lett.* 81 (1998), pp. 1562–1567. DOI: 10.1103/PhysRevLett.81.1562. arXiv: hep-ex/9807003 (cit. on pp. v, 5, 19).

[4] N. Agafonova et al. "Final Results of the OPERA Experiment on ν_τ Appearance in the CNGS Neutrino Beam". *Phys. Rev. Lett.* 120.21 (2018). [Erratum: Phys.Rev.Lett. 121, 139901 (2018)], p. 211801. DOI: 10.1103/PhysRevLett.120.211801. arXiv: 1804.04912 [hep-ex] (cit. on pp. v, 5, 19).

[5] R. Aaij et al. "Test of lepton universality with $B^0 \to K^{*0}\ell^+\ell^-$ decays". *JHEP* 08 (2017), p. 055. DOI: 10.1007/JHEP08(2017)055. arXiv: 1705.05802 [hep-ex] (cit. on pp. v, 27).

[6] Roel Aaij et al. "Test of lepton universality in beauty-quark decays" (Mar. 2021). arXiv: 2103.11769 [hep-ex] (cit. on pp. v, 27).

[7] Roel Aaij et al. "Measurement of CP-Averaged Observables in the $B^0 \to K^{*0}\mu^+\mu^-$ Decay". *Phys. Rev. Lett.* 125.1 (2020), p. 011802. DOI: 10.1103/PhysRevLett.125.011802. arXiv: 2003.04831 [hep-ex] (cit. on pp. v, 28).

[8] Roel Aaij et al. "Angular analysis of the $B^+ \to K^{*+}\mu^+\mu^-$ decay" (Dec. 2020). arXiv: 2012.13241 [hep-ex] (cit. on pp. v, 28).

[9] R. Aaij et al. "Measurement of the ratio of the $B^0 \to D^{*-}\tau^+\nu_\tau$ and $B^0 \to D^{*-}\mu^+\nu_\mu$ branching fractions using three-prong τ-lepton decays". *Phys. Rev. Lett.* 120.17 (2018), p. 171802. DOI: 10.1103/PhysRevLett.120.171802. arXiv: 1708.08856 [hep-ex] (cit. on pp. vi, 29).

[10] A. Abdesselam et al. "Measurement of $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ with a semileptonic tagging method" (Apr. 2019). arXiv: 1904.08794 [hep-ex] (cit. on pp. vi, 29).

[11] J. P. Lees et al. "Measurement of an Excess of $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ Decays and Implications for Charged Higgs Bosons". *Phys. Rev. D* 88.7 (2013), p. 072012. DOI: 10.1103/PhysRevD.88.072012. arXiv: 1303.0571 [hep-ex] (cit. on pp. vi, 29).

[12] Roel Aaij et al. "Search for the decays $B_s^0 \to \tau^+\tau^-$ and $B^0 \to \tau^+\tau^-$". *Phys. Rev. Lett.* 118.25 (2017), p. 251802. DOI: 10.1103/PhysRevLett.118.251802. arXiv: 1703.02508 [hep-ex] (cit. on pp. vi, 26).

[13] J. P. Lees et al. "Search for $B^+ \to K^+\tau^+\tau^-$ at the BaBar experiment". *Phys. Rev. Lett.* 118.3 (2017), p. 031802. DOI: 10.1103/PhysRevLett.118.031802. arXiv: 1605.09637 [hep-ex] (cit. on pp. vi, 31).

[14] T. V. Dong et al. "Search for the decay $B^0 \to K^{*0}\tau^+\tau^-$ at the Belle experiment" (Oct. 2021). arXiv: 2110.03871 [hep-ex] (cit. on pp. vi, 31).

[15] Bernat Capdevila et al. "Searching for New Physics with $b \to s\tau^+\tau^-$ processes". *Phys. Rev. Lett.* 120.18 (2018), p. 181802. DOI: 10.1103/PhysRevLett.120.181802. arXiv: 1712.01919 [hep-ph] (cit. on pp. vi, 1, 31, 32).

[16] P. A. Zyla et al. "Review of particle physics". *To be published in Prog. Theor. Exp. Phys.* 6 (2020), p. 083C01 (cit. on pp. viii, 17, 18, 56).

[17] A L Read. "Presentation of search results: The CL_s technique". *J. Phys.* G28 (2002), p. 2693. DOI: 10.1088/0954-3899/28/10/313 (cit. on pp. xii, 58, 174).

[18] Franz Mandl and Graham Shaw. *Quantum Field Theory*. John Wiley & Sons, 2010. ISBN: 9780471496830 (cit. on p. 5).

[19] Wikipedia. *https://en.wikipedia.org/wiki/Standard_Model*. Accessed 17 January 2021 (cit. on p. 8).

[20] Murray Gell-Mann. "A Schematic Model of Baryons and Mesons". *Phys. Lett.* 8 (1964), pp. 214–215. DOI: 10.1016/S0031-9163(64)92001-3 (cit. on p. 9).

[21] Roel Aaij et al. "Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \to J/\psi K^- p$ Decays". *Phys. Rev. Lett.* 115 (2015), p. 072001. DOI: 10.1103/PhysRevLett.115.072001. arXiv: 1507.03414 [hep-ex] (cit. on p. 9).

[22] Roel Aaij et al. "Observation of a narrow pentaquark state, $P_c(4312)^+$, and of two-peak structure of the $P_c(4450)^+$". *Phys. Rev. Lett.* 122.22 (2019), p. 222001. DOI: 10.1103/PhysRevLett.122.222001. arXiv: 1904.03947 [hep-ex] (cit. on p. 9).

[23] C. S. Wu et al. "Experimental Test of Parity Conservation in β Decay". *Phys. Rev.* 105 (1957), pp. 1413–1414. DOI: 10.1103/PhysRev.105.1413 (cit. on p. 12).

[24] J. H. Christenson et al. "Evidence for the 2π Decay of the K_2^0 Meson". *Phys. Rev. Lett.* 13 (1964), pp. 138–140. DOI: 10.1103/PhysRevLett.13.138 (cit. on p. 12).

[25] Gerhart Lüders. "Proof of the TCP theorem". *Annals of Physics* 2.1 (1957), pp. 1–15. ISSN: 0003-4916. DOI: https://doi.org/10.1016/0003-4916(57)90032-5. URL: http://www.sciencedirect.com/science/article/pii/0003491657900325 (cit. on p. 12).

[26] Werner Heisenberg. "Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik". *Zeitschrift für Physik* 43 (), pp. 172–198 (cit. on p. 12).

[27] Peter W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". *Phys. Rev. Lett.* 13 (16 Oct. 1964). https://link.aps.org/doi/10.1103/PhysRevLett.13.508, pp. 508–509. DOI: 10.1103/PhysRevLett.13.508 (cit. on p. 12).

[28] Wim de Boer. "The Discovery of the Higgs Boson with the CMS Detector and its Implications for Supersymmetry and Cosmology". *Time and Matter*. Sept. 2013. arXiv: 1309.0721 [hep-ph] (cit. on p. 13).

[29] J. Goldstone. "Field theories with 'Superconductor' solutions". *Il Nuovo Cimento (1955-1965)* 19.1 (Jan. 1961). https://doi.org/10.1007/BF02812722, pp. 154–164. ISSN: 1827-6121. DOI: 10.1007/BF02812722 (cit. on p. 13).

[30] Nicola Cabibbo. "Unitary Symmetry and Leptonic Decays". *Phys. Rev. Lett.* 10 (12 June 1963). https://link.aps.org/doi/10.1103/PhysRevLett.10.531, pp. 531–533. DOI: 10.1103/PhysRevLett.10.531 (cit. on p. 16).

[31] S. L. Glashow, J. Iliopoulos, and L. Maiani. "Weak Interactions with Lepton-Hadron Symmetry". *Phys. Rev. D* 2 (7 Oct. 1970). https://link.aps.org/doi/10.1103/PhysRevD.2.1285, pp. 1285–1292. DOI: 10.1103/PhysRevD.2.1285 (cit. on p. 16).

[32] Salvatore Mele. "The Measurement of the Number of Light Neutrino Species at LEP". *Adv. Ser. Dir. High Energy Phys.* 23 (2015). http://cds.cern.ch/record/2103251, pp. 89–106 (cit. on p. 17).

[33] Lincoln Wolfenstein. "Parametrization of the Kobayashi-Maskawa Matrix". *Phys. Rev. Lett.* 51 (1983), p. 1945. DOI: 10.1103/PhysRevLett.51.1945 (cit. on p. 17).

[34] CKMfitter Group (J. Charles et al.) *Constraints in the $(\bar{\rho},\bar{\eta})$ plane*. Eur. Phys. J. C41, 1-131 (2005) [hep-ph/0406184]. http://ckmfitter.in2p3.fr (cit. on p. 19).

[35] Giovanni Villadoro. *Axion Dark Matter*. Accessed 14 April 2021. URL: https://indico.cern.ch/event/1016498/attachments/2226129/3770809/2021-CERN-gv.pdf (cit. on p. 20).

[36] M. Kamionkowski. "WIMP and axion dark matter". *ICTP Summer School in High-Energy Physics and Cosmology*. June 1997. arXiv: hep-ph/9710467 (cit. on p. 20).

[37] CERN. *https://home.cern/science/physics/supersymmetry*. Accessed 14 April 2021 (cit. on p. 20).

[38] Thomas Blake, Gaia Lanfranchi, and David M. Straub. "Rare B Decays as Tests of the Standard Model". *Prog. Part. Nucl. Phys.* 92 (2017), pp. 50–91. DOI: 10.1016/j.ppnp.2016.10.001. arXiv: 1606.00916 [hep-ph] (cit. on p. 21).

[39] Kristof De Bruyn et al. "Branching Ratio Measurements of B_s Decays". *Phys. Rev. D* 86 (2012), p. 014027. DOI: 10.1103/PhysRevD.86.014027. arXiv: 1204.1735 [hep-ph] (cit. on p. 23).

[40] G. D'Ambrosio et al. "Minimal flavor violation: An Effective field theory approach". *Nucl. Phys. B* 645 (2002), pp. 155–187. DOI: 10.1016/S0550-3213(02)00836-2. arXiv: hep-ph/0207036 (cit. on p. 25).

[41] A. J. Buras et al. "Universal unitarity triangle and physics beyond the standard model". *Phys. Lett. B* 500 (2001), pp. 161–167. DOI: 10.1016/S0370-2693(01)00061-2. arXiv: hep-ph/0007085 (cit. on p. 25).

[42] Rodrigo Alonso, Benjamin Grinstein, and Jorge Martin Camalich. "$SU(2) \times U(1)$ gauge invariance and the shape of new physics in rare B decays". *Phys. Rev. Lett.* 113 (2014), p. 241802. DOI: 10.1103/PhysRevLett.113.241802. arXiv: 1407.7044 [hep-ph] (cit. on p. 25).

[43] Christoph Bobeth et al. "$B_{s,d} \to l^+l^-$ in the Standard Model with Reduced Theoretical Uncertainty". *Phys. Rev. Lett.* 112 (2014), p. 101801. DOI: 10.1103/PhysRevLett.112.101801. arXiv: 1311.0903 [hep-ph] (cit. on p. 25).

[44] Roel Aaij et al. "Search for the Rare Decays $B_s^0 \to e^+e^-$ and $B^0 \to e^+e^-$". *Phys. Rev. Lett.* 124.21 (2020), p. 211802. DOI: 10.1103/PhysRevLett.124.211802. arXiv: 2003.03999 [hep-ex] (cit. on p. 25).

[45] Vardan Khachatryan et al. "Observation of the rare $B_s^0 \to \mu^+\mu^-$ decay from the combined analysis of CMS and LHCb data". *Nature* 522 (2015), pp. 68–72. DOI: 10.1038/nature14474. arXiv: 1411.4413 [hep-ex] (cit. on p. 25).

[46] Roel Aaij et al. "Analysis of neutral B-meson decays into two muons" (Aug. 2021). arXiv: 2108.09284 [hep-ex] (cit. on p. 26).

[47] Bernard Aubert et al. "A search for the rare decay $B^0 \to \tau^+\tau^-$ at BABAR". *Phys. Rev. Lett.* 96 (2006), p. 241802. DOI: 10.1103/PhysRevLett.96.241802. arXiv: hep-ex/0511015 (cit. on p. 26).

[48] Roel Aaij et al. "Search for the lepton-flavour-violating decays $B_s^0 \to \tau^\pm\mu^\mp$ and $B^0 \to \tau^\pm\mu^\mp$". *Phys. Rev. Lett.* 123.21 (2019), p. 211801. DOI: 10.1103/PhysRevLett.123.211801. arXiv: 1905.06614 [hep-ex] (cit. on p. 26).

[49] Roel Aaij et al. "Search for the lepton-flavour violating decays $B^0_{(s)} \to e^\pm\mu^\mp$". *JHEP* 03 (2018), p. 078. DOI: 10.1007/JHEP03(2018)078. arXiv: 1710.04111 [hep-ex] (cit. on p. 26).

[50] Roel Aaij et al. "Search for the lepton flavour violating decay $B^+ \to K^+\mu^-\tau^+$ using B^{*0}_{s2} decays". *JHEP* 06 (2020), p. 129. DOI: 10.1007/JHEP06(2020)129. arXiv: 2003.04352 [hep-ex] (cit. on p. 26).

[51] 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". *JHEP* 01 (2021), p. 163. DOI: 10.1007/JHEP01(2021)163. arXiv: 2007.05658 [hep-ex] (cit. on p. 26).

[52] Roel Aaij et al. "Search for Lepton-Flavor Violating Decays $B^+ \to K^+\mu^\pm e^\mp$". *Phys. Rev. Lett.* 123.24 (2019), p. 241802. DOI: 10.1103/PhysRevLett.123.241802. arXiv: 1909.01010 [hep-ex] (cit. on p. 26).

[53] Marzia Bordone, Gino Isidori, and Andrea Pattori. "On the Standard Model predictions for R_K and R_{K^*}". *Eur. Phys. J. C* 76.8 (2016), p. 440. DOI: 10.1140/epjc/s10052-016-4274-7. arXiv: 1605.07633 [hep-ph] (cit. on p. 27).

[54] Roel Aaij et al. "Angular analysis of the $B^0 \to K^{*0}\mu^+\mu^-$ decay using 3 fb^{-1} of integrated luminosity". *JHEP* 02 (2016), p. 104. DOI: 10.1007/JHEP02(2016)104. arXiv: 1512.04442 [hep-ex] (cit. on p. 28).

[55] Joaquim Matias et al. "Complete Anatomy of $\bar{B}_d \to \bar{K}^{*0}(\to K\pi)l^+l^-$ and its angular distribution". *JHEP* 04 (2012), p. 104. DOI: 10.1007/JHEP04(2012)104. arXiv: 1202.4266 [hep-ph] (cit. on p. 28).

[56] Sébastien Descotes-Genon et al. "On the impact of power corrections in the prediction of $B \to K^*\mu^+\mu^-$ observables". *JHEP* 12 (2014), p. 125. DOI: 10.1007/JHEP12(2014)125. arXiv: 1407.8526 [hep-ph] (cit. on p. 28).

[57] Sébastien Descotes-Genon et al. "Global analysis of $b \to s\ell\ell$ anomalies". *JHEP* 06 (2016), p. 092. DOI: 10.1007/JHEP06(2016)092. arXiv: 1510.04239 [hep-ph] (cit. on p. 28).

[58] Yasmine Sara Amhis et al. "Averages of b-hadron, c-hadron, and τ-lepton properties as of 2018" (2019). Updated results and plots available here. arXiv: 1909.12524 [hep-ex] (cit. on p. 29).

[59] R. Aaij et al. "Measurement of the ratio of branching fractions $\mathcal{B}(B_c^+ \to J/\psi\tau^+\nu_\tau)/\mathcal{B}(B_c^+ \to J/\psi\mu^+\nu_\mu)$". *Phys. Rev. Lett.* 120.12 (2018), p. 121801. DOI: 10.1103/PhysRevLett.120.121801. arXiv: 1711.05623 [hep-ex] (cit. on p. 29).

[60] Marcel Algueró et al. "$b \to s\ell\ell$ global fits after Moriond 2021 results". *55th Rencontres de Moriond on QCD and High Energy Interactions*. Apr. 2021. arXiv: 2104.08921 [hep-ph] (cit. on p. 30).

[61] Wolfgang Altmannshofer and Peter Stangl. "New physics in rare B decays after Moriond 2021". *Eur. Phys. J. C* 81.10 (2021), p. 952. DOI: 10.1140/epjc/s10052-021-09725-1. arXiv: 2103.13370 [hep-ph] (cit. on p. 30).

[62] Marco Ciuchini et al. "Lessons from the $B^{0,+} \to K^{*0,+}\mu^+\mu^-$ angular analyses". *Phys. Rev. D* 103.1 (2021), p. 015030. DOI: 10.1103/PhysRevD.103.015030. arXiv: 2011.01212 [hep-ph] (cit. on p. 30).

[63] T. Hurth et al. "More Indications for Lepton Nonuniversality in $b \to s\ell^+\ell^-$" (Apr. 2021). arXiv: 2104.10058 [hep-ph] (cit. on p. 30).

[64] A. Augusto Alves Jr. et al. "The LHCb Detector at the LHC". *JINST* 3 (2008), S08005. DOI: 10.1088/1748-0221/3/08/S08005 (cit. on p. 35).

[65] Roel Aaij et al. "LHCb Detector Performance". *Int. J. Mod. Phys. A* 30.07 (2015), p. 1530022. DOI: 10.1142/S0217751X15300227. arXiv: 1412.6352 [hep-ex] (cit. on p. 35).

[66] "LHC Machine". *JINST* 3 (2008). Ed. by Lyndon Evans and Philip Bryant, S08001. DOI: 10.1088/1748-0221/3/08/S08001 (cit. on p. 35).

[67] A. Abashian et al. "The Belle Detector". *Nucl. Instrum. Meth. A* 479 (2002), pp. 117–232. DOI: 10.1016/S0168-9002(01)02013-7 (cit. on p. 36).

[68] T. Abe et al. "Belle II Technical Design Report" (Nov. 2010). arXiv: 1011.0352 [physics.ins-det] (cit. on p. 36).

[69] Bernard Aubert et al. "The BaBar detector". *Nucl. Instrum. Meth. A* 479 (2002), pp. 1–116. DOI: 10.1016/S0168-9002(01)02012-5. arXiv: hep-ex/0105044 (cit. on p. 36).

[70] Roel Aaij et al. "Measurement of the b-quark production cross-section in 7 and 13 TeV pp collisions". *Phys. Rev. Lett.* 118.5 (2017). [Erratum: Phys.Rev.Lett. 119, 169901 (2017)], p. 052002. DOI: 10.1103/PhysRevLett.118.052002. arXiv: 1612.05140 [hep-ex] (cit. on p. 36).

[71] S. Abachi et al. "The D0 Detector". *Nucl. Instrum. Meth. A* 338 (1994), pp. 185–253. DOI: 10.1016/0168-9002(94)91312-9 (cit. on p. 36).

[72] F. Abe et al. "The CDF Detector: An Overview". *Nucl. Instrum. Meth. A* 271 (1988), pp. 387–403. DOI: 10.1016/0168-9002(88)90298-7 (cit. on p. 36).

[73] The ATLAS Collaboration. "The ATLAS Experiment at the CERN Large Hadron Collider". *Journal of Instrumentation* 3.08 (Aug. 2008), S08003–S08003. DOI: 10.1088/1748-0221/3/08/s08003. URL: https://doi.org/10.1088/1748-0221/3/08/s08003 (cit. on p. 37).

[74] The CMS Collaboration. "The CMS experiment at the CERN LHC". *Journal of Instrumentation* 3.08 (Aug. 2008), S08004–S08004. DOI: 10.1088/1748-0221/3/08/s08004. URL: https://doi.org/10.1088/1748-0221/3/08/s08004 (cit. on p. 37).

[75] Georges Aad et al. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". *Phys. Lett. B* 716 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex] (cit. on p. 37).

[76] Serguei Chatrchyan et al. "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC". *Phys. Lett. B* 716 (2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex] (cit. on p. 37).

[77] The ALICE Collaboration. "The ALICE experiment at the CERN LHC". *Journal of Instrumentation* 3.08 (Aug. 2008), S08002–S08002. DOI: 10.1088/1748-0221/3/08/s08002. URL: https://doi.org/10.1088/1748-0221/3/08/s08002 (cit. on p. 37).

[78] Esma Mobs. "The CERN accelerator complex - August 2018. Complexe des accélérateurs du CERN - Août 2018" (Aug. 2018). General Photo. URL: http://cds.cern.ch/record/2636343 (cit. on p. 38).

[79] Roel Aaij et al. "LHCb Detector Performance". *Int. J. Mod. Phys. A* 30.07 (2015), p. 1530022. DOI: 10.1142/S0217751X15300227. arXiv: 1412.6352 [hep-ex] (cit. on p. 48).

[80] Johannes Albrecht et al. "Performance of the LHCb High Level Trigger in 2012". *J. Phys. Conf. Ser.* 513 (2014). Ed. by D. L. Groep and D. Bonacorsi, p. 012001. DOI: 10.1088/1742-6596/513/1/012001. arXiv: 1310.8544 [hep-ex] (cit. on p. 49).

[81] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. "A brief introduction to PYTHIA 8.1". *Computer Physics Communications* 178.11 (2008). http://www.sciencedirect.com/science/article/pii/S0010465508000441, pp. 852–867. ISSN: 0010-4655. DOI: https://doi.org/10.1016/j.cpc.2008.01.036 (cit. on p. 51).

[82] D. J. Lange. "The EvtGen particle decay simulation package". *Nucl. Instrum. Meth. A* 462 (2001). Ed. by S. Erhan, P. Schlein, and Y. Rozen, pp. 152–155. DOI: 10.1016/S0168-9002(01)00089-4 (cit. on p. 52).

[83] Piotr Golonka and Zbigniew Was. "PHOTOS Monte Carlo: A Precision tool for QED corrections in Z and W decays". *Eur. Phys. J. C* 45 (2006), pp. 97–107. DOI: 10.1140/epjc/s2005-02396-4. arXiv: hep-ph/0506026 (cit. on p. 52).

[84] S. Agostinelli et al. "GEANT4–a simulation toolkit". *Nucl. Instrum. Meth. A* 506 (2003), pp. 250–303. DOI: 10.1016/S0168-9002(03)01368-8 (cit. on p. 52).

[85] R. Aaij et al. "First observations of $\overline{B}{}^0_s \to D^+D^-$, $D^+_s D^-$ and $D^0\overline{D}{}^0$ decays". *Phys. Rev.* D87 (2013), p. 092007. DOI: 10.1103/PhysRevD.87.092007. arXiv: 1302.5854 [hep-ex] (cit. on pp. 58, 103).

[86] I. M. Nugent et al. "Resonance chiral Lagrangian currents and experimental data for $\tau^- \to \pi^-\pi^-\pi^+\nu_\tau$". *Phys. Rev. D* 88 (2013), p. 093012. DOI: 10.1103/PhysRevD.88.093012. arXiv: 1310.1053 [hep-ph] (cit. on p. 62).

[87] V. Gligorov. *https://twiki.cern.ch/twiki/pub/LHCb/Particle2MC/BkgCat_211013.key.pdf*. Accessed 01 February 2021 (cit. on p. 64).

[88] Andreas Hocker et al. "TMVA - Toolkit for Multivariate Data Analysis" (Mar. 2007). arXiv: physics/0703039 (cit. on pp. 68, 92).

[89] R. Aaij et al. "Search for the decays $B^0_s \to \tau^+\tau^-$ and $B^0 \to \tau^+\tau^-$". *Phys. Rev. Lett.* 118 (2017), p. 251802. DOI: 10.1103/PhysRevLett.118.251802. arXiv: 1703.02508 [hep-ex] (cit. on p. 69).

[90] Michel De Cian et al. *Fast neural-net based fake track rejection in the LHCb reconstruction*. Tech. rep. Geneva: CERN, Mar. 2017. URL: https://cds.cern.ch/record/2255039 (cit. on p. 72).

[91] A. Abulencia et al. "Search for $B_s \to \mu^+\mu^-$ and $B_d \to \mu^+\mu^-$ decays in $p\bar{p}$ collisions with CDF II". *Phys. Rev. Lett.* 95 (2005). [Erratum: Phys.Rev.Lett. 95, 249905 (2005)], p. 221805. DOI: 10.1103/PhysRevLett.95.221805. arXiv: hep-ex/0508036 (cit. on p. 73).

[92] Wouter D. Hulsbergen. "Decay chain fitting with a Kalman filter". *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 552.3 (Nov. 2005), pp. 566–575. ISSN: 0168-9002. DOI: 10.1016/j.nima.2005.06.078. URL: http://dx.doi.org/10.1016/j.nima.2005.06.078 (cit. on p. 77).

[93] Diego Martinez Santos and Frederic Dupertuis. "Mass distributions marginalized over per-event errors". *Nucl. Instrum. Meth.* A764 (2014), pp. 150–155. DOI: 10.1016/j.nima.2014.06.081. arXiv: 1312.5000 [hep-ex] (cit. on p. 103).

[94] Lucio Anderlini et al. *The PIDCalib package*. Tech. rep. Geneva: CERN, July 2016. URL: https://cds.cern.ch/record/2202412 (cit. on p. 123).

[95] R. Aaij et al. "Selection and processing of calibration samples to measure the particle identification performance of the LHCb experiment in Run 2". *Eur. Phys. J. Tech. Instr.* 6 (2018), p. 1. DOI: 10.1140/epjti/s40485-019-0050-z. arXiv: 1803.00824 [hep-ex] (cit. on p. 123).

[96] Muriel Pivk and Francois R. Le Diberder. "sPlot: A statistical tool to unfold data distributions". *Nucl. Instrum. Meth.* A555 (2005), pp. 356–369. DOI: 10.1016/j.nima.2005.08.106. arXiv: physics/0402083 [physics.data-an] (cit. on p. 126).

[97] Kyle Cranmer et al. "HistFactory: A tool for creating statistical models for use with RooFit and RooStats" (June 2012) (cit. on p. 147).

[98] Wouter Verkerke and David P. Kirkby. "The RooFit toolkit for data modeling". *eConf* C0303241 (2003). Ed. by L. Lyons and Muge Karagoz, MOLT007. arXiv: physics/0306116 (cit. on p. 147).

[99] Alexander L. Read. "Presentation of search results: The CL(s) technique". *J. Phys. G* 28 (2002). Ed. by M. R. Whalley and L. Lyons, pp. 2693–2704. DOI: 10.1088/0954-3899/28/10/313 (cit. on p. 147).

[100] Lorenzo Moneta et al. *The RooStats Project*. 2011. arXiv: 1009.1003 [physics.data-an] (cit. on p. 147).

[101] Glen Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". *Eur. Phys. J. C* 71 (2011). [Erratum: Eur.Phys.J.C 73, 2501 (2013)], p. 1554. DOI: 10.1140/epjc/s10052-011-1554-0. arXiv: 1007.1727 [physics.data-an] (cit. on p. 174).

[102] Roger J. Barlow and Christine Beeston. "Fitting using finite Monte Carlo samples". *Comput. Phys. Commun.* 77 (1993), pp. 219–228. DOI: 10.1016/0010-4655(93)90005-W (cit. on p. 174).

[103] Cédric Méaux. "A phenomenological study of $D^+ \to \pi^+\ell^+\ell^-$ and the search for $B^0_{(s)} \to \tau^+\tau^-$ at LHCb". PhD thesis. Marseille, CPPM, 2019 (cit. on p. 182).

[104] Giovanni Punzi. "Sensitivity of searches for new signals and its optimization". *eConf* C030908 (2003), MODT002. arXiv: physics/0308063 [physics] (cit. on p. 182).

[105] A. Abada et al. "FCC Physics Opportunities: Future Circular Collider Conceptual Design Report Volume 1". *Eur. Phys. J. C* 79.6 (2019), p. 474. DOI: 10.1140/epjc/s10052-019-6904-3 (cit. on p. 199).

[106] "The Compact Linear e^+e^- Collider (CLIC): Physics Potential" (Dec. 2018). Ed. by P. Roloff et al. arXiv: 1812.07986 [hep-ex] (cit. on p. 199).