DISSERTATION

submitted to the

Combined Faculty of Mathematics, Engineering and Natural Sciences

of Heidelberg University, Germany

for the degree of

Doctor of Natural Sciences

Put forward by

M.Sc. Daniel Joachim Unverzagt

born in Frankfurt, Germany

Oral examination: 20.06.2024

Search for Lepton Flavour Violation in four-body charm decays at LHCb

Referees:
Prof. Dr. Ulrich Uwer
Prof. Dr. Klaus Reygers

Abstract

This dissertation presents a search for physics beyond the standard model by the study of lepton flavour violation in four-body charm decays. A search for the decays $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$, $D^0 \to K^-\pi^+\mu^\pm e^\mp$ and $D^0 \to K^-K^+\mu^\pm e^\mp$ using data collected by the LHCb experiment is performed. These decays provide a unique opportunity to search for new physics in the up-type quark sector, similar to multiple searches performed in the down-type quark sector in other works. The decays are studied using proton-proton collision data corresponding to 6 fb^{-1} of integrated luminosity recorded at a centre-of-mass energy of 13 TeV during the years 2015-2018. An expected upper limit on the branching fraction for these decays is estimated at the order of 10^{-8}. Furthermore, separate expected limits for decays with different relative charge constellations in the final dilepton state in respect to the flavour of the D^0 meson are reported, providing increased sensitivity to new physics scenarios. It is shown that the current best limit on the branching ratios can be improved by two orders of magnitude with the LHCb dataset. This analysis is performed on a blind dataset with possible $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$, $D^0 \to K^-\pi^+\mu^\pm e^\mp$ and $D^0 \to K^-K^+\mu^\pm e^\mp$ candidates removed from the dataset. The results are therefore validated by performing an updated measurement of the branching fraction for $D^0 \to \pi^-\pi^+\mu^-\mu^+$ and $D^0 \to K^-K^+\mu^-\mu^+$ decays.

Zusammenfassung

Diese Dissertation befasst sich mit einer Suche nach Physik jenseits des Standardmodells durch die Untersuchung von Lepton-Flavour-Verletzung in Vierkörper -Charm-Zerfällen. Es wird eine Suche nach den Zerfällen $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$, $D^0 \to K^-\pi^+\mu^\pm e^\mp$ und $D^0 \to K^-K^+\mu^\pm e^\mp$ unter Verwendung der vom LHCb-Experiment gesammelten Daten durchgeführt. Diese Zerfälle bieten eine einzigartige Gelegenheit, nach neuer Physik im Up-Typ-Quark-Sektor zu suchen, ähnlich wie bei mehreren Suchen im Down-Typ-Quark-Sektor in der Vergangenheit. Die Zerfälle werden anhand von Proton-Proton-Kollisionsdaten untersucht, die einer integrierten Luminosität von 6 fb^{-1} entsprechen und bei einer Schwerpunktsenergie von 13 TeV in den Jahren 2015-2018 aufgezeichnet wurden. Eine erwartete Obergrenze für den Verzweigungsverältnisse für diese Zerfälle wird bestimmt und liegt in der Größenordnung von 10^{-8}. Darüber hinaus werden separate erwartete Grenzwerte für Zerfälle mit unterschiedlichen relativen Ladungskonstellationen im finalen Zwei-Leptonystem im Vergleich zum D^0-Meson angegeben, die eine erhöhte Empfindlichkeit für Szenarien der neuen Physik bieten. Es wird gezeigt, dass die derzeit beste Grenze für die Verzweigungsverhältnisse mit dem LHCb-Datensatz um zwei Größenordnungen verbessert werden kann. Für diese Analyse eine Blindstudie ist durchgeführt, wo mögliche $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$, $D^0 \to K^-\pi^+\mu^\pm e^\mp$ und $D^0 \to K^-K^+\mu^\pm e^\mp$ aus dem Datensatz entfernt werden. Die Ergebnisse sind daher durch eine aktualisierte Messung der Verzweigungsverhältnisse für $D^0 \to \pi^-\pi^+\mu^-\mu^+$ und $D^0 \to K^-K^+\mu^-\mu^+$ Zerfälle validiert.

Preface

This analysis was performed with a small analysis team within the LHCb collaboration. The LHCb collaboration is an international association of more than 1500 scientists and engineers from 99 institutes in 22 countries (April 2024). This implies the usage of common software to analyse the data collected by the LHCb experiment. The collection of the data and the development of common software used to analyse the data result from the efforts of many current and former collaboration members.

Contents

Chapter 1

Introduction

The universe, by today's understanding, is in physical terms described by two prevailing theories, General Relativity and the Standard Model (SM) of particle physics. General Relativity is used to describe gravitational effects, which typically manifest at macroscopic scales. The SM describes the three other fundamental forces, the electromagnetic, the weak and the strong force, including their interaction with the basic building blocks of matter. The SM itself was confirmed and completed with the discovery of the Higgs boson 2012 [1, 2]. To date, the SM describes experimental results of high-energy physics with great accuracy. However, the SM can not explain the large matter-antimatter asymmetry seen in the universe [3]. Other examples of phenomena unexplained by the SM are the non-zero neutrino mass or the presence of dark matter and dark energy observed at large scales. One of the fundamental goals in physics is to understand and describe nature through a universal theory, which the SM does not. Most high-energy experiments aim, therefore, to find its limits, for example by the search for forbidden processes.

Proposed by Sakharov [4], three so-called "Sakharov conditions" have to be fulfilled in order to explain the large matter-antimatter asymmetry in the universe:

1. Interactions out of thermal equilibrium
2. Baryon number violation
3. Charge conjugation Parity violation (CP violation)

Baryon number violation refers to a hypothetical physical process that took place in the universe's early phases, introducing a large baryonic asymmetry, also often referred to as baryogenesis. Similar to this, leptogenesis refers to a hypothetical physical process in the universe's early phases that introduces a lepton-antilepton

asymmetry. Leptogenesis is not strictly necessary to explain the large matter-antimatter asymmetry. However, charge conservation suggests that lepton- and baryon-asymmetry should be of the same order of magnitude [5]. In addition, the baryon-lepton number difference is a precisely conserved property of the SM. Consequently, Leptogenesis without Baryogenesis is not possible within the SM, implying Physics Beyond the SM [6]. Therefore, many measurements focus on the search for a lepton number asymmetry.

The LHCb collaboration's goal is to study flavoured hadrons containing c- and b-quarks. For this purpose, the LHCb detector is built at the Large Hadron Collider (LHC), a circular collider situated at the European Organization for Nuclear Research (CERN). The detector is built at one of four proton-proton collision points, providing many b- and c-hadrons to study[1]. This allows the study of Lepton Number (LN) and Baryon Number (BN) violations. Instead of just the LN, the lepton flavour number, the generation-dependent lepton number, is often studied. In studies of rare b-hadron decays sensitive to transitions of the form $b \to s l^+ l^-$, multiple hints for violation of lepton universality (LU) have been observed [8–12], some but not all could be refuted [13, 14]. These hints further increase the interest in searches of lepton number violation as many New Physics (NP) models naturally link LU and LN violation.

Searches of lepton flavour number violation (LFV) in c-hadron transitions are less widely studied than in b-hadron decays, even though c-hadron decays are also sensitive for searches for LFV [15]. With b-hadrons, the so-called down-type sector, named after the quark with the same charge in the first generation, is tested. Complementary searches for the up-type sector can be performed when studying c-hadrons. New Physics does not necessarily couple equally to down-type quarks (b-decays) and up-type quarks (charm decays). These studies are crucial to understanding the flavour pattern of possible SM extensions. The LHCb collaboration is playing a leading role in studying rare and forbidden decays of charm hadrons, which might be sensitive to effects beyond the Standard Model. As such, two-body, $D^0 \to \mu^\pm e^\mp$ decays [16] and three-body, $D^+_{(s)} \to h^+ l^+ l^-$ decays [17], decays were already studied at LHCb. The h stands for either a pion or kaon[2].

The BaBar collaboration published a recent search for four-body $D^0 \to h^- h^{(\prime)+} \mu^\pm e^\mp$

[1] Around 10^{13} $c\bar{c}$ quark pairs are produced within the LHCb detectors acceptance between 2015 and 2018 [7].

[2] This abbreviation will be used throughout this work.

decays with an upper limit on the branching ratio at around 10^{-6} at the 90% confidence level [18]. Up until now, these decays have never been studied within the LHCb collaboration. This doctoral work presents the search for lepton flavour violation in four-body charm decays using LHCb data collected between 2015-2018, the world's largest dataset of charm decays to date.

This thesis aims to update the limits set by the BaBar collaboration and further explore lepton flavour violation/conservation in charm decays. With the work already done for this thesis, it is possible to set up an expected limit in the range of $10^{-8} - 10^{-7}$, increasing the search sensitivity for this decay by two orders of magnitude. In addition, a branching fraction measurement for $D^0 \to h^- h^+ \mu^- \mu^+$ decays is performed, which is used to validate and crosscheck the still blinded search for LFV. The analysis will only be unblinded after an internal review and approval from the collaboration.

The structure of the thesis is as follows: chapter 2 introduces the concept of lepton flavour violation. The LHC accelerator and the LHCb detector with its sub-detectors are described in chapter 3. In chapter 4, measurement strategy and important methods are outlined and described. In chapter 5, the possible decay candidate reconstruction and selection are discussed. From these selected decays, a branching ratio will be calculated. Two ingredients are needed for this: the number of observed candidates and the efficiency ratio. The efficiency ratio calculation can be found in chapter 6. The number of observed candidates and their significance are evaluated by a fit explained in chapter 7. Afterwards, with the help of pseudo experiments and data-driven methods, all dominant systematic uncertainties are estimated as explained in chapter 8, followed by a brief summary in chapter 10. The $D^0 \to h^{(\prime)-} h^+ \mu^\pm e^\mp$ decays and $D^0 \to h^- h^+ \mu^- \mu^+$ decays are studied and reported alongside each other. The main purpose of this work's study of $D^0 \to h^- h^+ \mu^- \mu^+$ decays is to crosscheck the main analysis steps; therefore, no detailed study about all systematic uncertainties is performed.

Chapter 2

Phenomenology of LFV in charm decays

This chapter covers the theoretical background needed to study rare and forbidden four-body decays of hadrons containing charm quarks. The chapter is split into four parts. First, a slightly more general overview of the current theoretical framework, provided by the Standard Model, to describe charm decays is given. Second, the phenomenological formalism used to describe four-body charm decays is introduced. Third, LFV is explained as a concept to search for physics beyond the Standard Model. Lastly, the branching fraction is introduced as an observable to measure the decay rate.

The Standard Model (SM) of particle physics was formulated to describe the interaction of elementary particles with each other. It has undergone successful validation through numerous experiments in the last 50 years. Nevertheless, the SM as such is an incomplete theory, and many unresolved questions remain. One example of its incompleteness is the absence of gravity within its formalism. Also, none of the particles accounted for in the SM appear as plausible candidates for dark matter, and the theory fails to explain the dark matter observed in astrophysical experiments [19–21]. Furthermore, symmetry violations, like charge-parity violation, present in the SM fail to explain the significant matter-antimatter asymmetry in the universe [22].

Two decay modes are studied in this thesis. First, $D^0 \to h^- h^+ \mu^- \mu^+$ decays, which will also serve as a reference to the primarily studied decays in this thesis. These decays were already studied in a previous work [23]. Second, $D^0 \to h^{(\prime)-} h^+ \mu^\pm e^\mp$ decays are studied, with the measurement of their branching ratio being the primary goal of this thesis. $D^0 \to h^{(\prime)-} h^+ \mu^\pm e^\mp$ decays themselves are forbidden within the SM and thus cannot be

described by it. However, an often applied assumption is that if new physics phenomena exist, they behave similarly to the SM and can be described by minimal extensions. This may be true or not. With the help of the existing effective field theory and its operators, it is possible to introduce a generic parameterisation for possible new physics effects.

2.1 Standard Model

The Standard Model (SM) of particle physics, in its current formulation, describes all interactions in particle physics. The current formulation of the SM dates back to the mid-1970s [24–26]. For a more detailed and general description, see Ref. [27]. For a more in-depth reading about $D^0 \to h^- h^+ \mu^- \mu^+$ decays see Ref. [23].

The SM is a renormalisable quantum field theory. It combines Lorentz invariance (special relativity) and quantum theory. The model describes twelve matter particles with spin 1/2, and interactions are described by force-carrying particles (gauge bosons) with spin 1. The underlying symmetry, defining the three forces described by the SM, is given by the special unitarity group:

$$G_{SM} = SU(3)_C \times SU(2)_L \times U(1)_Y \tag{2.1.1}$$

where $SU(3)_C$ invariance leads to the strong interaction and $SU(2)_L \times U(1)_Y$ to the electromagnetic and weak interaction. Each continuous symmetry of the system, indicated by the index, leads to a conserved charge, according to Noether's theorem [28]. The interactions are mediated by gauge bosons[1], coupling to particle fields that carry the associated charge under the respective symmetry transformation. Like in a classical field theory, the equation of motion can be determined by minimisation of the action computed for the Lagrangian. The Lagrangian contains all possible renormalisable terms invariant under the underlying symmetry group. Summarised in figure 2.1 is the particle content of the SM.

Matter particles, called fermions, are categorised into two groups, depending on their gauge-boson interactions: quarks and leptons. Up-type quarks (u, c, t) have an electric charge of two-thirds, given in units of the electron charge $e \approx 1.6 \times 10^{-19}\,C$, and down-type quarks have an electric charge of $-1/3$. In the lepton group, it is possible to separate between charged leptons (e^-, μ^-, τ^-) with a charge of minus one and uncharged

[1]The integer spin property arises from the gauge invariance.

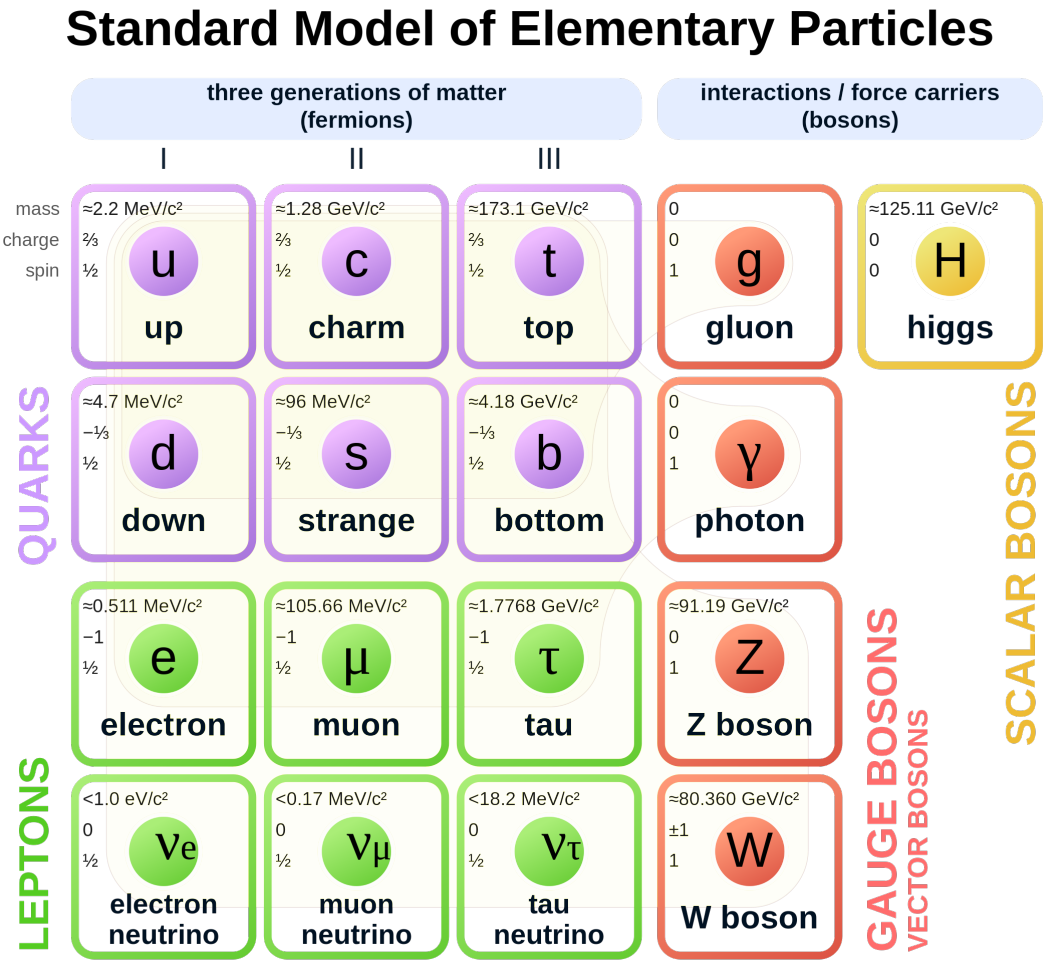


Figure 2.1: Standard model of elementary particles grouped by family, generation and flavour [29].

neutrinos (ν_e, ν_μ, ν_τ). For each particle, an antiparticle exists with inverted charges. As depicted in figure 2.1, these particles can be divided into three so-called generations.

The strong force binds quarks together, preventing the observation of isolated "free" quarks. This force is described by eight massless gluons, which interact exclusively with objects carrying colour charge, which are quarks and gluons themselves. Due to gluon self-interactions, the strong force's coupling strength, α_s, increases at large distances, leading to confinement and prohibiting freely propagating quarks. The typical length scale of the strong force is 10^{-15} metres. Quarks combine to form colour-neutral bound states, called hadrons, including mesons (quark-antiquark pairs) and baryons (combinations of three (anti-)quarks). Recent experimental findings have confirmed more intricate hadron structures with four or five quarks, called tetraquarks and pentaquarks, respectively [30,31].

The theoretical framework governing the strong force is Quantum Chromodynamics (QCD). Perturbation theory is applicable for predictions when α_s is not too large, with an energy scale $\Lambda_{QCD} \approx 200\,\text{MeV}$ defining the boundary above which perturbative methods can be utilised. Below Λ_{QCD}, where the coupling strength becomes too large, non-perturbative techniques like lattice QCD are employed. However, in scenarios like charm physics, where the mass of the c-quark is marginally above Λ_{QCD}, the reliance on perturbation theory introduces uncertainties.

The electroweak force is a unified description of the electromagnetic and weak forces, and its comprehension relies on the concept of spontaneous symmetry breaking, a phenomenon enacted by the Higgs mechanism within the SM. In this framework, an additional scalar field, known as the Higgs field, possesses a ground state that does not adhere to the gauge symmetry of the SM. This results in a mixing of three $SU(2)_L$ gauge fields and a single $U(1)_Y$ gauge field into two electrically charged gauge bosons, W^+ and W^-, a neutral particle Z^0 and the massless boson of the electromagnetic force, the photon (γ). The $W^\pm$ and Z^0 particles acquire significant masses, approximately $80\,\text{GeV}/c^2$ and $91\,\text{GeV}/c^2$, respectively, becoming massive mediators of the weak force. This high mass limits the range of the weak interaction to approximately 10^{-18} metres. In contrast, the electromagnetic force carried by the massless photon lacks such limitations on its range.

Flavour-changing currents in the SM occur only via weakly charged interactions, mediated by $W^\pm$. These interactions are usually in the form $q \to W^+ q'$, where q and q' are quarks of different types, either up- or down-type. Due to a mismatch between the quark electroweak flavour states and mass eigenstates, the non-diagonal Yukawa coupling term to the mass introduces a quark mixing matrix called Cabibbo–Kobayashi–Maskawa (CKM) matrix [32, 33]. The matrix elements of the CKM matrix give the strength of the flavour-changing weak interaction. This can best be seen when writing the CKM matrix in the Wolfenstein parameterisation [34]:

$$V_{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} = \begin{pmatrix} 1-\lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1-\lambda^2/2 & A\lambda^2 \\ A\lambda^3(1-\rho-i\eta) & A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4) \tag{2.1.2}$$

with $\lambda \approx 0.23$, $A \approx 0.83$, $\bar{\rho} \approx 0.16$ and $\bar{\eta} \approx 0.35$ [35]. The indices denote the corresponding transition, $V_{qq'}$. Diagonal elements, V_{ud}, V_{cs} and V_{tb}, govern transitions within a generation and are of the order $\mathcal{O}(1)$, while transitions between differ-

ent generations experience suppression by powers of λ. The SM does not provide predictions for the individual parameters; the values need to be determined experimentally.

2.1.1 Flavour within the SM and beyond

Baryon and Lepton Number conservation are both *accidental symmetries* of the SM, meaning that no fundamental theoretical reason imposed by a SM symmetry exists for them to be conserved. This thesis studies Lepton Number Violation; therefore, the following discussion will be limited to leptons; however, in principle, the same string of argumentation is also valid for Baryon Numbers[2].

When studying a decay, the *Lepton Number* can be defined as $L = N_{l^-} - N_{l^+}$, where $N_{l^-}(N_{l^+})$ is the number of (anti-)leptons. Lepton Number conservation means that L is the same for a decay's initial and final state. This conservation also applies to the different lepton species and their flavour numbers (L_e, L_μ, L_τ), respectively. The index indicates the respective lepton and its associated neutrino. Contrary to what was initially stated in the SM, the observation of neutrino oscillations leads to Lepton Flavour Violation (LFV) and shows that neutrinos have small yet non-zero masses [36]. Nevertheless, due to the small neutrino mass, LFV due to neutrinos can be neglected for most high-energy collision experiments, including the measurement presented in this thesis, as the effects are multiple orders of magnitudes below the typical experimental sensitivity.

2.2 Searches for New Physics

Two different methodologies are applied in the search for Physics Beyond the Standard Model, *direct* and *indirect searches*. Direct searches involve the exploration of interactions between new particles, such as potential dark matter candidates and conventional baryonic matter. These direct searches extend to the direct production of new particles in the high-energy environment of particle colliders, exemplified by detectors such as XENONnT [37, 38] or future experiments like DARWIN [39]. These detectors aim to identify rare interactions between dark matter candidates and the nuclei of a target material, typically a noble gas *e.g.* xenon. A typical example of a direct search at particle colliders is the search and discovery of the Higgs boson by ATLAS and CMS [1, 2].

[2]To explain the large matter-antimatter discrepancies in the universe both baryon and lepton number violation would be required.

Although successful in the past, the sensitivity of these searches is inherently constrained by the available centre-of-mass energy in the collisions.

Indirect searches for New Physics are, for example, done by Muon g-2, Belle, LHCb or in experiments measuring the electron electric dipole (EDM). According to Heisenberg's uncertainty principle, quantum loops allow the participation of heavy, as-yet-unobserved particles in quantum mechanical processes, even when the energy is insufficient for direct production. It is, therefore, possible to probe physics beyond the SM that goes beyond the limitations of direct collider searches. With this approach, new physics may manifest itself in processes forbidden by the SM or in modifications to existing SM processes through interference with SM contributions. A typical example of this was the observation of neutral kaon mixing and the following unexpected suppression of *flavour-changing neutral currents* (FCNC) [40]. This motivated the formulation of the GIM mechanism [41] and the consequent prediction of a fourth quark. This prediction was confirmed by the direct discovery of J/ψ, a bound $c\bar{c}$-state [42][3]. This also highlights one of the weaknesses of indirect searches. A direct observation is often needed to confirm the prediction unambiguously.

The study of FCNC transitions with leptons in the final state, typically referred to as semileptonic decays, provides one such way to search for NP. These are, for example, $b \to s l^- l^+$ and $b \to d l^- l^+$ transitions in B-decays, $s \to d l^- l^+$ transitions in kaon decays, and $c \to u l^- l^+$ transitions in charm decays, where $l^- l^+$ is a pair of oppositely charged leptons. Of these transitions, those transitions of the form $c \to u l^- l^+$ play a unique role as they are the only transitions accessing the bound up-type sector.

One of the main challenges in studying charm decays in general and semileptonic charm decays specifically is the large uncertainties in predicting QCD effects. To circumvent this problem, theoretical "clean" observables are desired, which provide a smaller theoretical uncertainty than experimentally expected. Angular observables or searches in phase-space regions where QCD effects are suppressed provide good opportunities to search for new physics contributions. For example, after the first observation of $D^0 \to h^- h^+ \mu^- \mu^+$ decays [44], one goal was to measure the relative branching fraction in regions away from phase-space regions dominated by intermediate resonances[4] [44]. In figure 2.2,

[3]Three months earlier, a possible first observation of a charm decay in cosmic ray showers was published [43].

[4]Intermediate resonances refer to decays via intermediate particle state for example via $\eta((u\bar{u} + d\bar{d} - 2s\bar{s})/\sqrt{6})$, $\eta'((u\bar{u} + d\bar{d} + s\bar{s})/\sqrt{3})$, $\rho^0((d\bar{d} + u\bar{u})/\sqrt{2}; \text{Isospin} = 1)$, $\omega((d\bar{d} + u\bar{u})/\sqrt{2}; \text{Isospin} = 0)$ or $\phi(s\bar{s})$ mesons.

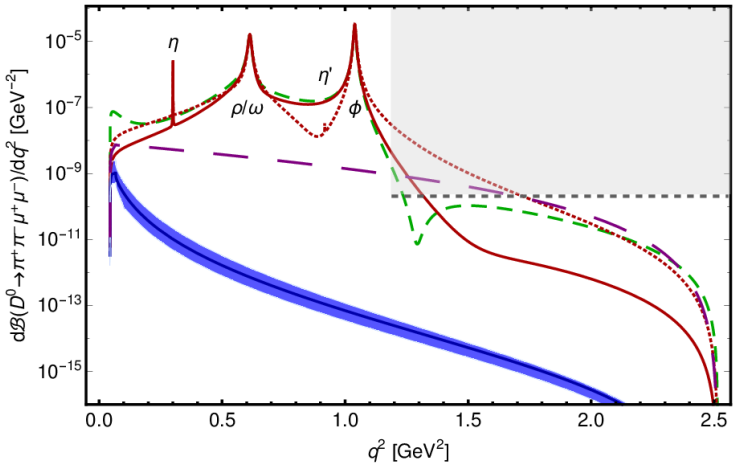


Figure 2.2: Theoretical predicted and experimental measured differential branching fraction $dD^0 \to \pi^-\pi^+\mu^-\mu^+/dq^2$. The SM-predicted differential branching fractions via intermediate resonances are shown in red and green. The solid red and dashed red lines represent different assumptions on the strong phase for the same model [46], while the green line assumes a different theoretical model [48]. For the purpose of this illustration, the three lines (two red and one green) illustrate uncertainties of the theoretical prediction related to the strong phase. The blue bands represent non-resonant contributions, including theoretical uncertainties of hadronic form factors. The grey dashed line shows the experimental upper limit on the branching fraction in the grey shaded dimuon mass region [44]. Possible NP scenarios are illustrated in purple [46]. The figure is taken from Ref. [45]

predictions and the current experimental limit on the differential branching fraction $dD^0 \to \pi^-\pi^+\mu^-\mu^+/dq^2$ are displayed [45]. q^2 is the squared four-momentum of the dimuon system. The non-resonant (in blue) contribution is orders of magnitude smaller than the resonant predictions (in red and green), making them non-accessible for experiments. For high q^2, away from the resonances, the sensitivity for New Physics scenarios (in purple) is enhanced [46]. However, the uncertainty on the theoretical predictions would make it hard to interpret the result if the decay would be observed in the high-q^2 region. In a second step, angular observables were defined, and an angular analysis was performed [47].

Another opportunity is to study decays forbidden in the SM[5], like LFV decays. Next to various decays studied in the b-sector, LHCb did also study $D^0 \to \mu^\pm e^\mp$ decays [16] and $D^+_{(s)} \to h^+l^+l^-$ decays [17]. While $D^0 \to \mu^\pm e^\mp$ decays provide an experimentally clean signature, they are strongly helicity suppressed because of no hadrons in the final state, reducing the decay rate significantly. The expected branching fraction for the in

[5]Simplifying the theoretical prediction significantly.

the SM allowed $D^0 \to \mu^-\mu^+$ decay is of order 10^{-11} [49, 50], which is well below the current experimental sensitivity [51]. Another possible decay which was never studied is $D^0 \to \tau^\mp e^\pm$, the only D^0 decay which kinematically allows to study the tau system. The helicity suppression is loosened in $D^+_{(s)} \to h^+l^+l^-$ decays due to the additional hadron in the final state. Next to LFV decays also, lepton number violating decays of the form $D^+_{(s)} \to h^+l^+l^-$ were studied. However, currently, for $D^+_{(s)} \to h^+l^+l^-$ decays, the Run 2 dataset is only partially studied [17].

This work covers the study of LFV in four-body charm decays, $D^0 \to h^{(\prime)-}h^+\mu^\pm e^\mp$ decays. The most stringent limits to date are given by the BaBar collaboration [18]. The decays were not yet studied at LHCb. Noteworthy, LFV in the down-type quark sector does not necessarily introduce LFV in the up-type quark sector and vice versa. Therefore, the search in the up-type quark sector provides a complementary search for LFV to the already existing searches in the down-type quark sector.

In the future, similar to searches in the down-type quark sector [13, 14], also the study of Lepton Flavour Universality (LFU) will be of interest:

$$R^c_{hh} = \frac{\int_{q^2_{min}}^{q^2_{max}} \frac{d\mathcal{B}(D^0 \to h^+h^-\mu^+\mu^-)}{dq^2}\, dq^2}{\int_{q^2_{min}}^{q^2_{max}} \frac{d\mathcal{B}(D^0 \to h^+h^-e^+e^-)}{dq^2}\, dq^2} \tag{2.2.1}$$

Assuming Lepton Flavour conservation, this ratio R^c_{hh}, phase space corrected, is expected to equal unity. The advantage of studying LFU is that from a theoretical point of view, precise predictions are possible, as most hadronic effects cancel in the ratio. Experimentally, a good understanding of electron and muon efficiencies is crucial. Electrons interact more with the detector material, making predictions less precise and more complicated for the trigger system, resulting in a lower trigger efficiency. This is also a reason why $D^0 \to h^{(\prime)-}h^+e^-e^+$ decays are not yet observed. However, they are potentially within experimental reach using the full Run 2 dataset [52]. In addition, as both $D^0 \to h^{(\prime)-}h^+e^-e^+$ and $D^0 \to h^-h^+\mu^-\mu^+$ decays are rare decays, a sufficiently large data sample will be needed[6]. LHCb published recently a similar combined measurement of LFU in the b-sector for $B^+ \to K^+l^+l^-$ decays ($\mathrm{R_K}$) and $B^0 \to K^*l^+l^-$ decays ($\mathrm{R_{K^*}}$) ($\mathrm{R_X}$) [13, 14]. The results are compatible with the SM.

A comprehensive and well-explained overview of rare charm decays' experimental and theoretical status can be found in Ref. [45].

[6]Given the expected sensitivity and planned luminosity, this may be the case for Run 3 or Run 4.

2.3 Phenomenology of rare and forbidden charm decays

The argumentation and formalism in this chapter follow Ref. [15]. A good introduction to effective field theories can be found in Ref. [53] and Ref. [54].

A model-independent effective field theory (EFT) is usually formulated to describe processes in particle physics phenomenologically. EFT allows the simplification of practical calculations in field theory, which often makes such calculations feasible, and it provides a systematic formalism for the analysis of multi-scale problems. The basic idea is to restrict oneself to the energy scale Λ of the physical phenomena[7] while ignoring substructures and degrees of freedom at higher energies. Intuitively, one integrates all heavy degrees of freedom ($m > \Lambda$) out and absorbs them in some factor C_i. Lighter quark fields ($m < \Lambda$) are expressed by local operators P_i and are considered as massless[8]. The Lagrangian of these effective fields is written as a sum of local fields based on the operator product expansion (OPE) [55]:

$$\mathcal{L}_{\text{eff}} \sim \sum_i C_i P_i, \tag{2.3.1}$$

where the strength C_i of the fields described by the operators P_i are called Wilson coefficients. The SM operators up to dimension six for FCNC charm decays are the following [56–58]:

$$P_1^{(q)} = (\bar{u}_L \gamma_{\mu_1} T^\alpha q_L)(\bar{q}_L \gamma^{\mu_1} T^\alpha c_L),$$
$$P_2^{(q)} = (\bar{u}_L \gamma_{\mu_1} q_L)(\bar{q}_L \gamma^{\mu_1} c_L),$$

the so-called current-current operators;

$$P_3 = (\bar{u}_L \gamma_{\mu_1} c_L) \sum_{\{q:m_q<\Lambda\}} (\bar{q} \gamma^{\mu_1} q),$$
$$P_4 = (\bar{u}_L \gamma_{\mu_1} T^\alpha c_L) \sum_{\{q:m_q<\Lambda\}} (\bar{q} \gamma^{\mu_1} T^\alpha q),$$
$$P_5 = (\bar{u}_L \gamma_{\mu_1} \gamma_{\mu_2} \gamma_{\mu_3} c_L) \sum_{\{q:m_q<\Lambda\}} (\bar{q} \gamma^{\mu_1} \gamma^{\mu_2} \gamma^{\mu_3} q),$$
$$P_6 = (\bar{u}_L \gamma_{\mu_1} \gamma_{\mu_2} \gamma_{\mu_3} T^\alpha c_L) \sum_{\{q:m_q<\Lambda\}} (\bar{q} \gamma^{\mu_1} \gamma^{\mu_2} \gamma^{\mu_3} T^\alpha q),$$

[7] Typically, the mass scale of the $W^\pm$ boson or the c quark for charm decays.

[8] The operator product expansion is only possible because of the assumption $m \ll \Lambda$.

the QCD and electroweak penguin operators;

$$P_7 = \frac{e}{g_s^2} m_c (\bar{u}_L \sigma_{\mu_1\mu_2} c_R) F_{\mu_1\mu_2},$$
$$P_8 = \frac{e}{g_s} m_c (\bar{u}_L \sigma_{\mu_1\mu_2} T^\alpha c_R) G^\alpha_{\mu_1\mu_2},$$

the electromagnetic and chromo-magnetic dipole operators and

$$P_9 = \frac{e^2}{g_s^2} (\bar{u}_L \gamma_{\mu_1} c_L)(\bar{l}\gamma^{\mu_1} l),$$
$$P_{10} = \frac{e^2}{g_s^2} (\bar{u}_L \gamma_{\mu_1} c_L)(\bar{l}\gamma^{\mu_1}\gamma_5 l),$$

the leptonic operators. In the operators, $q = d, s, b$ and γ^μ represent the Dirac matrices. Left and right-handed quarks are donated by $q_{L,R} = P_{L,R} q$, with the chirality operator $P_{L,R} = \frac{1}{2}(1 \mp \gamma^5)$. The electromagnetic field strength tensor is denoted by $F^{\mu\nu}$, and T^a represents the Gell-Mann matrices, which describe internal colour rotations of gluon fields in QCD. $G^\alpha_{\mu_1\mu_2}$ is the chromomagnetic field strength tensor. In charm decays, the dominant operators are the current-current operators P_1 and P_2 at the scale m_W because the CKM matrix leads to an exact cancellation for massless light fields[9]. The effective weak Lagrangian can be written down as follows:

$$\mathcal{L}_{\text{eff}}^{\text{weak}} \sim \sum_{q \in d,s} V_{cq}^* V_{uq} \left(\sum_{i=1}^{2} C_i^{(q)} P_i + \sum_{i=3}^{10} C_i^{(q)} P_i \right) \tag{2.3.2}$$

Given the large uncertainties in theory prediction by QCD effects, one usually tries to define so-called null-tests. An example is the exploitation of the absence of axial vector currents[10], corresponding to $P_{10} = 0$. This allows to test specific angular distributions [46, 47].

The model-independent representation via operators allows to parameterise NP effects. Within the same formalism, the operator basis can be extended by lepton flavour-violating currents:

$$P_9^{(ll')} = (\bar{u}\gamma_{\mu_1} c_L)(\bar{l}\gamma^{\mu_1} l'), \qquad P_9^{(ll')\prime} = (\bar{u}\gamma_{\mu_1} c_R)(\bar{l}\gamma^{\mu_1} l') \tag{2.3.3}$$
$$P_{10}^{(ll')} = (\bar{u}\gamma_{\mu_1} c_L)(\bar{l}\gamma^{\mu_1}\gamma_5 l'), \qquad P_{10}^{(ll')\prime} = (\bar{u}\gamma_{\mu_1} c_R)(\bar{l}\gamma^{\mu_1}\gamma_5 l') \tag{2.3.4}$$
$$P_S^{(ll')} = (\bar{u} c_L)(\bar{l} l'), \qquad P_S^{(ll')\prime} = (\bar{u} c_R)(\bar{l} l') \tag{2.3.5}$$
$$P_P^{(ll')} = (\bar{u}\gamma_{\mu_1} c_L)(\bar{l}\gamma_5 l'), \qquad P_P^{(ll')\prime} = (\bar{u}\gamma_{\mu_1} c_R)(\bar{l}\gamma_5 l') \tag{2.3.6}$$
$$P_T^{(ll')} = \frac{1}{2}(\bar{u}\sigma_{\mu_1\mu_2} c_L)(\bar{l}\sigma_{\mu_1\mu_2} l'), \qquad P_T^{(ll')\prime} = \frac{1}{2}(\bar{u}\sigma^{\mu_1\mu_2} c_L)(\bar{l}\sigma_{\mu_1\mu_2}\gamma_5 l') \tag{2.3.7}$$

[9] Normally referred to as GIM mechanism.

[10] This is due to the GIM mechanism.

Resulting in the effective $qq'll'$ Lagrangian[11] of the form:

$$\mathcal{L}_{\text{eff}}^{\text{LQ}} \sim \lambda_{c\mu}^{*}\lambda_{ue}\left(\sum_i C_i P_i^{(\prime)}\right) + \lambda_{ce}^{*}\lambda_{u\mu}\left(\sum_i C_i P_i^{(\prime)}\right), \quad i \in (9, 10, S, P, T) \tag{2.3.8}$$

where λ_{ql}, q is the quark and l is the lepton at the corresponding vertex. This gives the strength of the lepton-flavour violating coupling, similar to the CKM matrix elements for flavour-changing weak interactions. The exact form and contributing terms depend on the structure of the $qq'll'$ coupling, for example if it is a vector or scalar mediator, or its charge. A listing of different Leptoquark scenarios can be found in Ref. [15]. By applying experimental constraints on the Wilson coefficients, the approach is similar to the one in Ref. [15]. It is possible to work out up to which branching fractions NP scenarios are not yet excluded. The following limits are reported [46]:

$$\mathcal{B}(D^0 \to \pi^-\pi^+\mu^\pm e^\mp) \lesssim 10^{-7}, \quad \mathcal{B}(D^0 \to K^-K^+\mu^\pm e^\mp) \lesssim 10^{-9}. \tag{2.3.9}$$

The goal of this analysis is to improve and validate these limits, if possible.

Following the above discussion, and different to the previous measurement by the BaBar collaboration [18], also an individual search for the differently charged $D^0 \to h^{(\prime)-}h^+\mu^-e^+$ (a) and $D^0 \to h^{(\prime)-}h^+\mu^+e^-$ (b) decays is performed. This allows to test an increased number of NP scenarios as a different set of couplings, λ_{ql} matrix elements, are tested. This is illustrated in figure 2.3a and b. The limits by the BaBar collaboration at 90% CL are given by [18]:

$$\mathcal{B}(D^0 \to \pi^-\pi^+\mu^\pm e^\mp) \lesssim 1.7 \times 10^{-6}, \tag{2.3.10}$$

$$\mathcal{B}(D^0 \to K^-\pi^+\mu^\pm e^\mp) \lesssim 1.9 \times 10^{-6}, \tag{2.3.11}$$

$$\mathcal{B}(D^0 \to K^-K^+\mu^\pm e^\mp) \lesssim 1.0 \times 10^{-6}. \tag{2.3.12}$$

Both limits for $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ decays and $D^0 \to K^-K^+\mu^\pm e^\mp$ decays are within the for NP already excluded region.

The in the EFT formalism excluded regions are worked out on the basis of FCNC transitions, which can only occur in loop processes. This is the case for $D^0 \to \pi^-\pi^+\mu^-\mu^+$ and $D^0 \to K^-K^+\mu^-\mu^+$ decays. These are then extended to $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ and $D^0 \to K^-K^+\mu^\pm e^\mp$ decays. However, a limit for $D^0 \to K^-\pi^+\mu^\pm e^\mp$ decays is also provided as it is unknown in which ways NP may manifest[12]. Experimentally, the difference between all three decays is small, allowing the simultaneous study of all three decays.

[11] In literature this is often referred to as Leptoquark-Lagrangian.

[12] Technically it would also be possible to study $D^0 \to K^+\pi^-\mu^\pm e^\mp$ decays. However, similar to

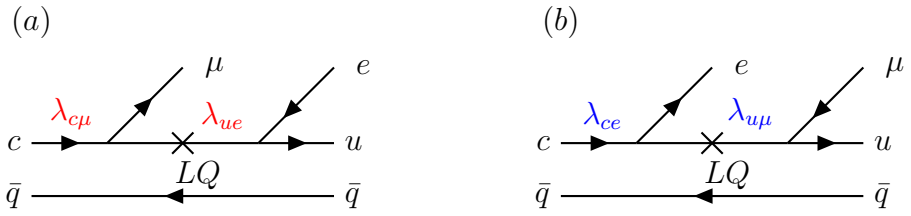


Figure 2.3: Example diagrams for possible Leptoquark contributions.

2.4 Branching fraction

The branching fraction is the fraction of particles decaying by an individual decay mode, here $D^0 \to h^{(\prime)-}h^+\mu^\pm e^\mp$ decays. This number can be calculated by dividing the number of $D^0 \to h^{(\prime)-}h^+\mu^\pm e^\mp$ decays by the number of expected D^0 mesons, scaled by the absolute efficiency of detecting $D^0 \to h^{(\prime)-}h^+\mu^\pm e^\mp$. To estimate the number of D^0 mesons produced in pp collisions, one needs to know the integrated luminosity of the detector and the probability of an D^0 to be produced in a proton-proton collision, the *cross-section* $\sigma(pp \to D^0)$. The instantaneous luminosity is defined in such a way that:

$$\mathcal{L} = \frac{1}{\sigma}\frac{d\mathrm{N}}{dt} \tag{2.4.1}$$

where N is the number of detected particles. Knowing the geometry and number of particles in each bunch, $N_{1,2}$, the luminosity can be calculated by:

$$\mathcal{L} = f_{coll}\frac{N_1 N_2}{4\pi\sigma_x^*\sigma_y^*}\mathcal{F}, \tag{2.4.2}$$

where f_{coll} is the frequency of collisions, σ_x^*, y is the root mean square of the transverse beam size in horizontal and vertical directions at the interaction point, and $\mathcal{F}$ is a factor to account for inefficient geometric overlapping of the two beams due to finite bunch length and dynamic effects. The integrated luminosity,

$$\mathcal{L}^{int} = \int \mathcal{L}\, dt, \tag{2.4.3}$$

is combined with the D^0 meson cross-section. The branching fraction can then be calculated as follows:

$$\mathcal{B}(D^0 \to h^{(\prime)-}h^+l^{(\prime)-}l^+) = \frac{N(D^0 \to h^{(\prime)-}h^+l^{(\prime)-}l^+)}{\sigma(pp \to D^0)\ \mathcal{L}^{int}} \times \frac{1}{\epsilon(D^0 \to h^{(\prime)-}h^+l^{(\prime)-}l^+)}, \tag{2.4.4}$$

$D^0 \to K^+\pi^-\pi^-\pi^+$ and $D^0 \to K^-\pi^+\pi^-\pi^+$ decays, with a branching fraction $\mathcal{O}(10^{-4})$ and $\mathcal{O}(10^{-2})$, respectively, these decays are expected to be additionally suppressed by two orders of magnitude because of the CKM matrix. For the Cabibbo favoured $D^0 \to K^-\pi^+\pi^-\pi^+$ decay $\lambda_{cs}\lambda_{ud} \sim 1$ contributes, while for the double Cabibbo suppressed decays $\lambda_{cd}\lambda_{us} \sim 10^{-2}$ enters the calculation.

where $\sigma(pp \to D^0)\ \mathcal{L}^{int}$ corresponds to the number of produced D^0 mesons at LHCb. Both quantities can only be calculated to a limited precision around 8% for the cross-section [7] and $1-3\%$ for the luminosity [59, 60]. To circumvent this problem, the branching fraction relative to a normalisation channel is calculated. For this, the measurement $D^0 \to K^-\pi^+[\mu^+\mu^-]_{\rho^0/\omega}$ was chosen, as decay is well understood and has a similar decay topology to the signal decay. The similar topology of the final state particles allows the cancellation of possible biases which may affect the measurement. The relative branching fraction can then be written as:

$$\frac{\mathcal{B}(D^0 \to h^{(\prime)-}h^+l^{(\prime)-}l^+)}{\mathcal{B}(D^0 \to K^-\pi^+\mu^-\mu^+)} = \frac{N(D^0 \to h^{(\prime)-}h^+l^{(\prime)-}l^+)}{N(D^0 \to K^-\pi^+\mu^-\mu^+)} \cdot \frac{\epsilon(D^0 \to K^-\pi^+\mu^-\mu^+)}{\epsilon(D^0 \to h^{(\prime)-}h^+l^{(\prime)-}l^+)} \quad (2.4.5)$$

where ϵ donates the efficiency to detect the corresponding decay, N is the number and $\mathcal{B}$ the branching fraction of the associated decay.

It is only necessary to estimate the *yield* and *efficiency ratio* between $D^0 \to h^{(\prime)-}h^+\mu^\pm e^\mp$ and $D^0 \to K^-\pi^+[\mu^+\mu^-]_{\rho^0/\omega}$, cancelling most systematic effects. For $\mathcal{B}(D^0 \to K^-\pi^+\mu^-\mu^+)$, the previously measured value will be used [61].

The second advantage of performing a relative branching fraction measurement is that possible systematic effects, for example an over- or underestimation of the efficiency due to a wrongly assumed material budget of the detector. Ideally the effect on the efficiency is multiplicative and will cancel.

2.4.1 Blinding data

Typically, when a measurement is made, the data is blinded to not bias the measurement unconsciously. An example of such bias would be the so-called *confirmation bias*, which is the tendency to favour one's prior belief. The critical value, the value one wants to measure in this measurement and the value that needs to be blinded, is the branching ratio and its significance. For this purpose, a unknown blind b is introduced into equation 2.4.5:

$$\frac{\mathcal{B}(D^0 \to h^{(\prime)-}h^+e^\mp\mu^\pm) + b}{\mathcal{B}(D^0 \to K^-\pi^+\mu^-\mu^+)} = \frac{N(D^0 \to h^{(\prime)-}h^+e^\mp\mu^\pm)}{N(D^0 \to K^-\pi^+\mu^-\mu^+)} \cdot \frac{\epsilon(D^0 \to K^-\pi^+\mu^-\mu^+)}{\epsilon(D^0 \to h^{(\prime)-}h^+e^\mp\mu^\pm)} \quad (2.4.6)$$

the blind b corresponds to a random number between 0 and 10^{-6}. The blind makes one blind to the actual measured branching fraction while not affecting the uncertainty. The uncertainty of the branching ratio is the value one wants to minimise in such a measurement. The result can later, as soon as the measurement is reviewed and approved

by the collaboration, be unblinded by setting the value to zero. To not negate the blind, it is not possible to look into the actual recorded data, as the existence of the decay is also unknown and blind. However, this makes it nearly impossible to validate and crosscheck this measurement. Therefore, a second unblinded measurement of the branching fraction of two decays $D^0 \to K^-K^+\mu^-\mu^+$ and $D^0 \to \pi^-\pi^+\mu^-\mu^+$ is performed. Both decays, in principle, can be treated the same as the LFV decay modes. The main difference between the two decays is the electron in the final state. Both decays were already previously observed, which allows to compare the results with the previous measurement [44].

Chapter 3

The LHCb experiment

This chapter briefly describes the LHC and, in more detail, the LHCb detector and its subsystems. The measurement performed uses data collected between the years 2015 and 2018. Therefore, only the detector's design during this period is discussed in this thesis. Thereafter, particle interactions with the detector, with a special focus on the electron reconstruction, are explained. Lastly, the trigger system is detailed, and the procedure to recover recorded events and create simulated events for this analysis is summarized.

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [62] is situated at the European Organization for Nuclear Research (CERN). The LHC is a hadron collider inside a 27 km tunnel across the French and Swiss border, designed to collide two proton beams at a centre of mass energy of around 13 TeV. This high energy can only be reached with the help of a complex accelerator system, which provides particles to the LHC at 450 GeV. The proton beams are then further accelerated by superconducting dipole magnets up to 6.5 TeV each, providing a centre of mass energy of up to 13 TeV. The beams are kept on a circular orbit by superconducting magnets. Along the circumference of the LHC, the particles are collided at four collision points at a collision rate of up to 40 MHz[1]. Four large experiments, CMS, ATLAS, ALICE and LHCb, are located at these points. ATLAS and CMS are so-called general-purpose detectors designed to cover a broad spectrum of high-energy physics. ALICE is specialised in the study of heavy ion collisions, increasing the understanding of QCD. The LHCb experiment is a dedicated flavour physics detector designed to understand the properties of c- and b-hadrons and their decays. Both LHCb

[1]This corresponds to one collision every 25 ns.

and ALICE are operated with a lower luminosity[2] than CMS and ATLAS.

The first protons started colliding in the LHC end of 2009. Between 2010 and 2012, the first data-taking period, called Run 1, took place. In 2010 and 2011, a centre of mass energy of $\sqrt{s} = 7\,\text{TeV}$ was reached. This was increased to $\sqrt{s} = 8\,\text{TeV}$ in 2012. Run 1 followed a long shutdown period of two years, during which modifications to the LHC dipole magnets were performed to allow for a higher centre of mass energy. Between 2015 and 2018, the next data-taking period followed, called Run 2, which ended with a second shutdown lasting until 2022. In Run 2, a centre of mass energy of 13 TeV was reached.

3.2 The LHCb detector

The LHCb detector is designed to study heavy b- and c-flavoured hadrons, *i.e.* heavy mesons containing a b-quark or c-quark. A more detailed description of the experiment can be found in Ref. [63]. Due to their lifetime and their large boost, b-quarks decay, by design, within the first $1 - 2\,\text{cm}$ of the detector. With the help of these data samples, the LHCb collaboration has made significant contributions to the field of flavour physics. At the energies provided by the LHC b- and c-quarks, typically produced as $q\overline{q}$ pairs, are produced mainly in the forward or backward direction. This is illustrated in figure 3.1 when looking at simulated $b\bar{b}$ quark pairs. Around 10^{13} $c\bar{c}$ [7] and around 10^{12} $b\bar{b}$ [64] pairs are produced within the LHCb acceptance during Run 2. To measure b- and c-hadrons, the LHCb detector is built as a forward spectrometer[3]. The design is contrary to the general-purpose detectors like ATLAS or CMS.

To be able to perform high-precision flavour physics measurements, the decay topology needs to be accurately measured with an excellent momentum resolution. Additionally, the final state particles need to be identified precisely. The LHCb detector provides all these requirements. The sub-detectors can be categorized into two categories. First, the tracking system to provide topological and momentum information. Second, the particle identification systems to identify the particle species. The LHCb detector with its various sub-systems is shown in figure 3.2. A right-handed cartesian coordinate system is used. The z-axis points parallel to the beam. The y-axis points upwards.

[2] In the case of ALICE also with a lower collision rate.

[3] Forward represents here both directions along the beam pipe. The LHCb detector is equipped only in one direction for practical and financial reasons.

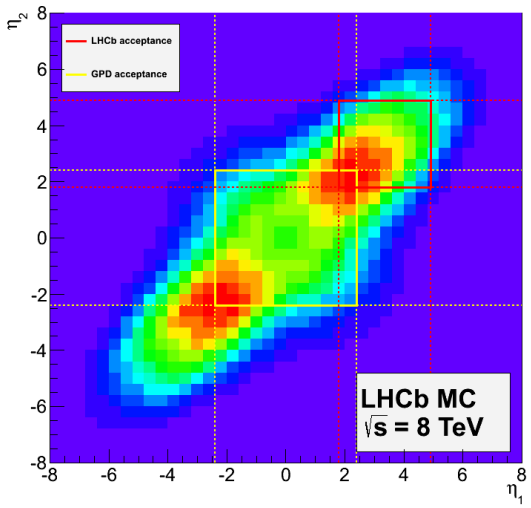


Figure 3.1: The heat map shown in these plots were created using PYTHIA8 [65] and CTEQ6 NLO [66]. The LHCb acceptance covers $1.8 < \eta < 4.9$. The figure is taken from Ref. [67]. The red square marking the LHCb acceptance contains 24% of the produced $b\bar{b}$-pairs.

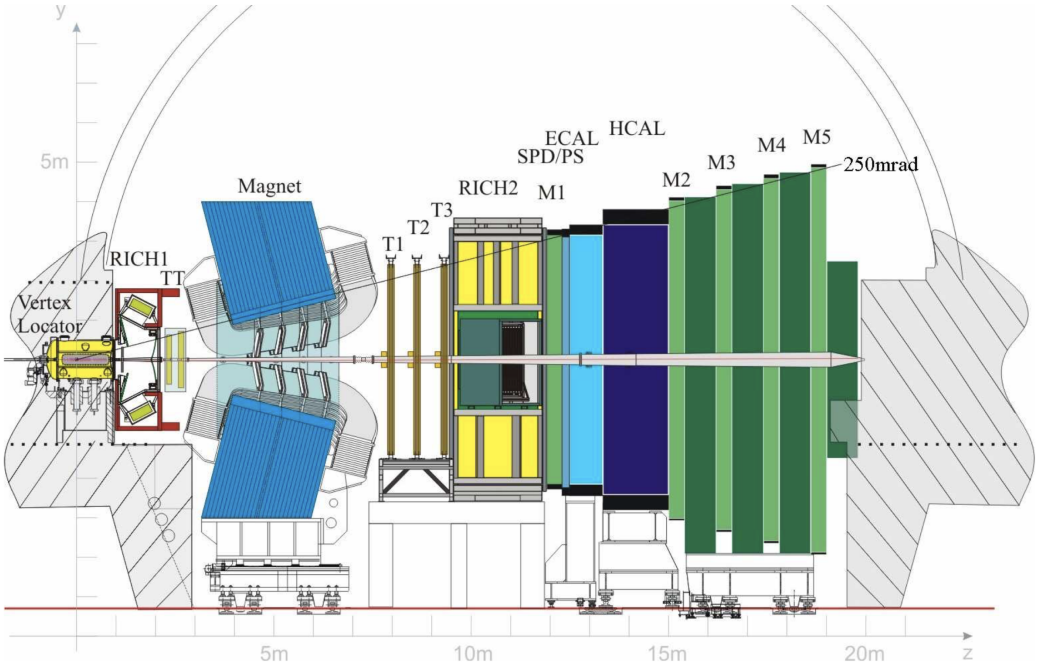


Figure 3.2: LHCb detector with its sub-systems [68].

3.2.1 Tracking system

A precise track reconstruction and momentum estimate is crucial to identify the decays studied in this thesis. A track is defined as the trajectory of a charged particle. The tracking system consists of three sub-detectors: The VErtex LOcator, VELO, the Tracker

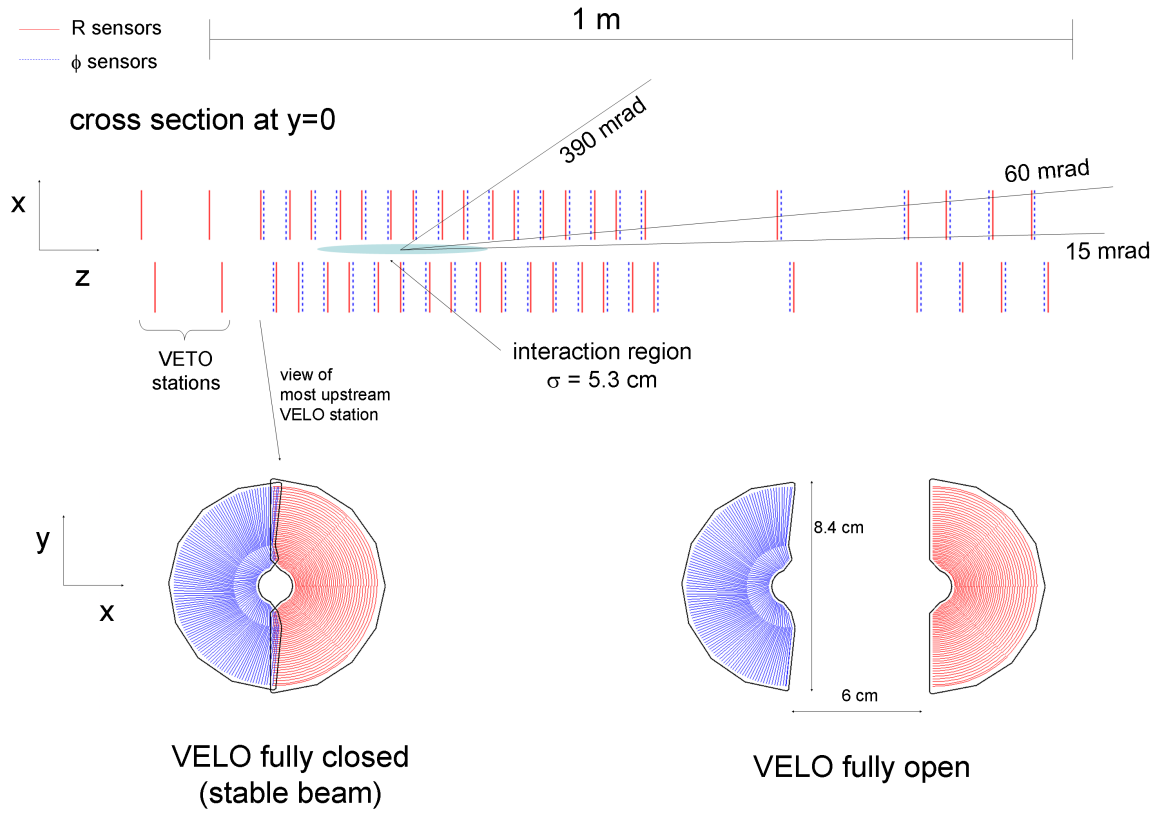


Figure 3.3: Schematic view of the Vertex Locator. At the top is the overall arrangement of the stations along the z-axis. At the bottom, the two halves of one station [69].

Turicensis, TT, and the Tracking Stations, T1-3. The tracking stations themselves consist of the Inner Tracker, IT, and the Outer Tracker, OT.

The first component of the LHCb tracking system is the VELO. The VELO is used to reconstruct the decay of heavy-flavoured hadrons precisely. The typical b- and c-hadrons lifetime of approximately 1 ps and their high boost allow a travel distance of about 1 cm before decaying. Therefore, they decay within the VELO close to the primary vertex. The detector is as close as 8 mm to the collision point and consists of 21 stations aligned in the z-direction with two different types of silicon-strip sensors. The R sensor measures a particle's radial distance, and the ϕ sensor measures the azimuthal coordinate. The position of the stations measures the z component, see figure 3.3, allowing a precise measurement of the decay vertex location. In order for a track to be identified, at least three hits are necessary.

The primary purpose of the Track Turicensis (TT) is to allow the reconstruction of low-momentum particles ($< 2\,\mathrm{GeV}/c$), which are bent by the magnetic field out of the acceptance of the tracking stations downstream, and long-lived particles, which decay outside of the VELO. The TT is a micro-strip silicon detector comprising of four sensor layers in two stations. These stations have a distance of 30 cm to each other along the beam axis. The first and last layers are aligned vertically in the y-direction. The middle layers are rotated by an angle of $+5°$ and $-5°$ to provide the necessary resolution in y-direction.

Crucial for the momentum measurement, a dipole magnet is placed between the TT and the main tracking stations. The magnet has an integrated B-field strength of about 4 Tm. The main component of the magnetic field is along the y-axis. Therefore, the (x, z)-plane is, in good approximation, the bending plane. The bending measured by the VELO and the tracking stations after the magnet allows for precise, single-particle momentum measurement.[4] A special feature of LHCb is the ability to change the magnet's polarity. This allows any charge asymmetry introduced by the detector to be cancelled out to a large extent if the collected data samples with the two different polarities (*MagUp* and *MagDown*) have approximately the same size. The magnet polarity is reversed approximately every two weeks.

The main tracker consists of three tracking stations (T1, T2, T3) located after the magnet, which completes the tracking system. The IT, silicon-strip detectors, cover the inner area close to the beamline. The large outer area at more central rapidity is covered by the Outer Tracker (OT) straw-tube detector. The three tracking stations consist of 24 layers, each split into four double layers. Each layer consists of two rows of straw tubes filled with gas. In figure 3.4 the cross section of a single layer and the arrangement of all layers within the detector is shown.

3.2.2 Particle identification system

The particle species of the quasi-stable final state particles are identified with the help of the particle identification system. The particle identification system consists of a Cherenkov detector system, the calorimeter system, and the muon stations. Typically, the information of multiple subsystems is combined into one classifier, which is used

[4] This is in stark contrast to CMS and ATLAS, which have a more compact design and rely heavily on their calorimeters.

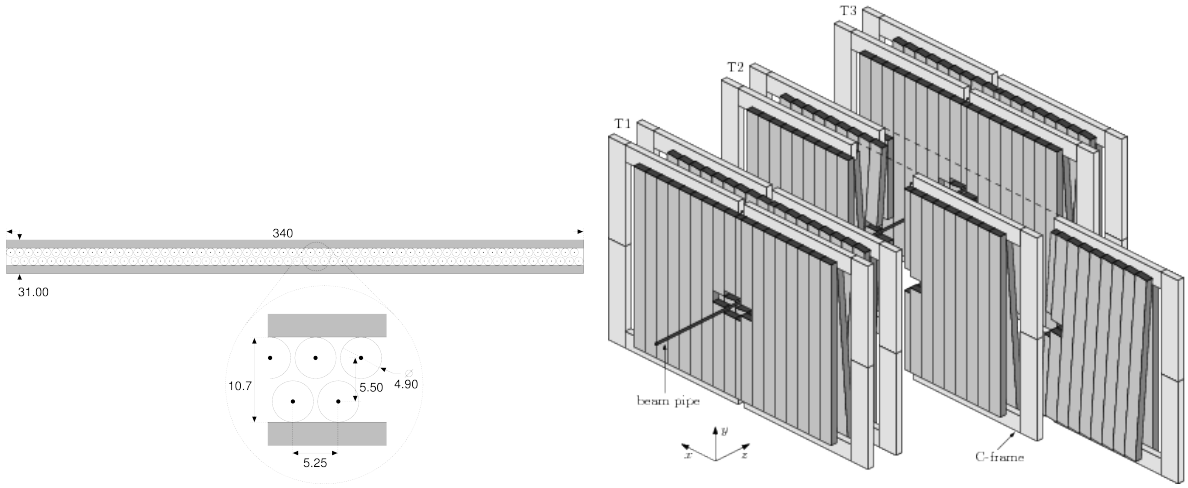


Figure 3.4: On the left schematically the cross-section of one double layer of the OT with the straw tubes is shown. On the right, the overall arrangement of the different tracking stations T1-3 is shown. [70]

to assign a particle type. In general, particles are categorised into photons, long-lived[5] charged leptons, electrons and muons, and long-lived charged hadrons, pions[6] and kaons. Short-lived particles are then reconstructed by combining the decay products.

The first of the two Ring Imaging CHerenkov detectors, RICH 1, is directly behind the VELO. The second RICH detector, RICH 2, is located between the tracking stations and the calorimeter system. The two RICH detectors exploit the Cherenkov radiation emitted by relativistic charged particles. When a charged particle traverses a medium with a refractive index $n > 1$ at a velocity v exceeding the local phase velocity of light, it emits Cherenkov light. The emitted photons follow the Cherenkov angle, θ, in relation to the direction of the particle's momentum.

$$\theta = \frac{c}{nv} \tag{3.2.1}$$

Different mass hypotheses for the particle can be tested with the candidate's measured Cherenkov angle (velocity) and momentum, as illustrated in figure 3.5. RICH 1 is used to identify particles with low momentum of 1-60 GeV/c, whereas RICH 2 is used for particles with higher momentum of 15-100 GeV/c.

Calorimeters measure a particle's energy deposits either in its entirety if stopped within the detector material, or parts of its energy while traversing[7]. The complete arrangement

[5]Particle with a lifetime long enough to pass through the magnet.

[6]By default, all particles are assumed to be pions until categorised differently.

[7]If the particle reacts with the detector material at all.

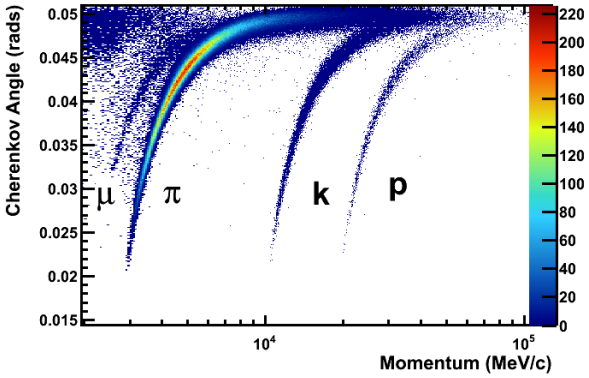


Figure 3.5: Reconstructed Cherenkov angle as a function of track momentum in the C_4F_{10} radiator [71].

of the calorimeter system is shown in figure 3.6 on the left.

The first station of the calorimeter system is the Scintillating Pad Detector (SPD). It detects charged particles which ionize the detector material. The detector material then scintillates, and photomultipliers detect the released photons. The SPD separates photons and electrons, as only the latter are detected.

Directly behind the SPD is the PreShower detector (PS), which measures the energy deposited in the lead.

Afterwards, the electron and photons' energy deposit is measured in the Electromagnetic CALorimeter (ECAL). The ECAL is a shashlik calorimeter and consists of 2 mm thick lead and 4 mm thick scintillating material installed alternately.

Downstream of the ECAL, the Hadronic CALorimeter (HCAL) is located, which is used to measure the energy of hadrons. The HCAL is, in principle, similarly built to the ECAL, with larger cells consisting of iron between the scintillators.
Behind the calorimeters the muon stations (M2-M5) are located. The muons produced at LHCb penetrate through all detector subsystems, including calorimeters. The muon stations are separated by 80 cm thick iron plates called the Muon filter. This ensures that only muons can reach the last station as the filter absorbs all other particles, particularly hadrons. The muon stations play an essential role in the trigger system[8]. The arrangement of the muon stations in the detector is illustrated in figure 3.6 on the right.

[8]The measurement presented in this thesis heavily relies on the muon trigger.

The RICH 1 and RICH 2 dominantly contribute to identifying charged hadrons, SPD and ECAL electrons and photons, and the muon system, with the help of the HCAL to veto other particles, muons.

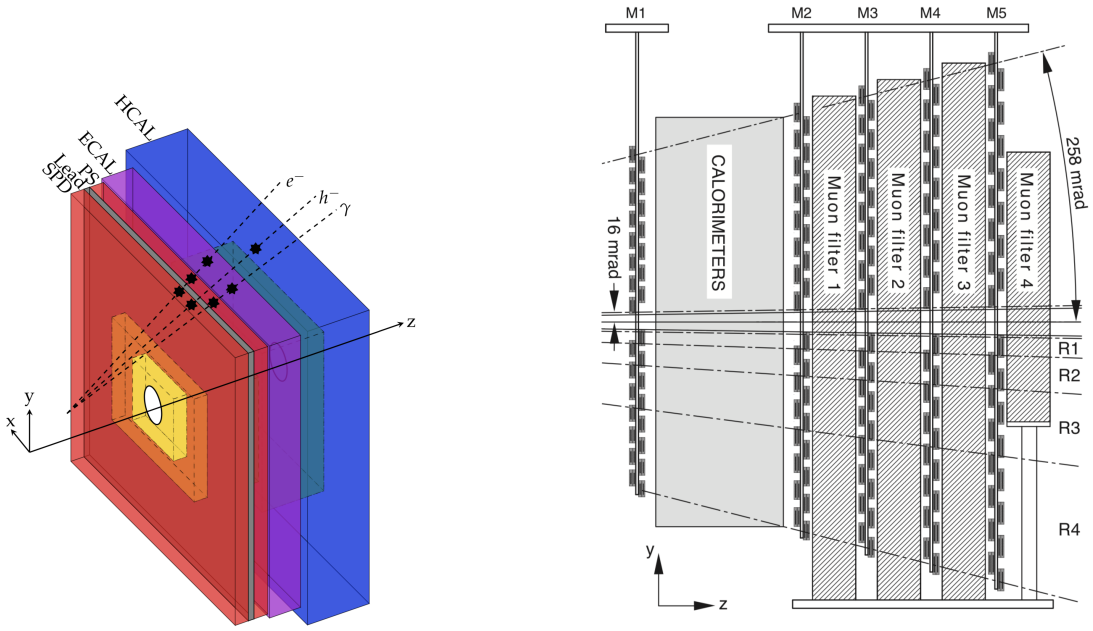


Figure 3.6: Schematic view of the calorimeter system on the left and on the right the location of the muon stations, including the position of the calorimeter system. The figures are taken from Ref. [72, 73].

3.3 Interactions of charged particles within the detector

To be able to measure the LFV decay $D^0 \to h^{(\prime)-}h^+\mu^\pm e^\mp$, it is crucial to understand how the final state particles used to reconstruct the decay react within the detector. For a more in-depth reading, see Ref. [74].

The D^{*+} meson decays via the strong interaction immediately after production, also called prompt decay, because of the short decay time. The consequent D^0 meson is expected to fly up to a few millimetres before it decays into two hadrons, an electron and a muon[9]. To be able to reconstruct the decays, a well-displaced secondary vertex is needed, introducing a minimum requirement on the flight distance and consequently

[9]As the decay itself is forbidden; this is just an assumption and must not necessarily hold true.

decay-time of the D^0. The four possible final state particles, electron, muon and the two hadrons, pion and kaon, leave the VELO area and traverse further through the detector. A sketch of the decay topology underlaid by the LHCb detector is shown in figure 3.7.

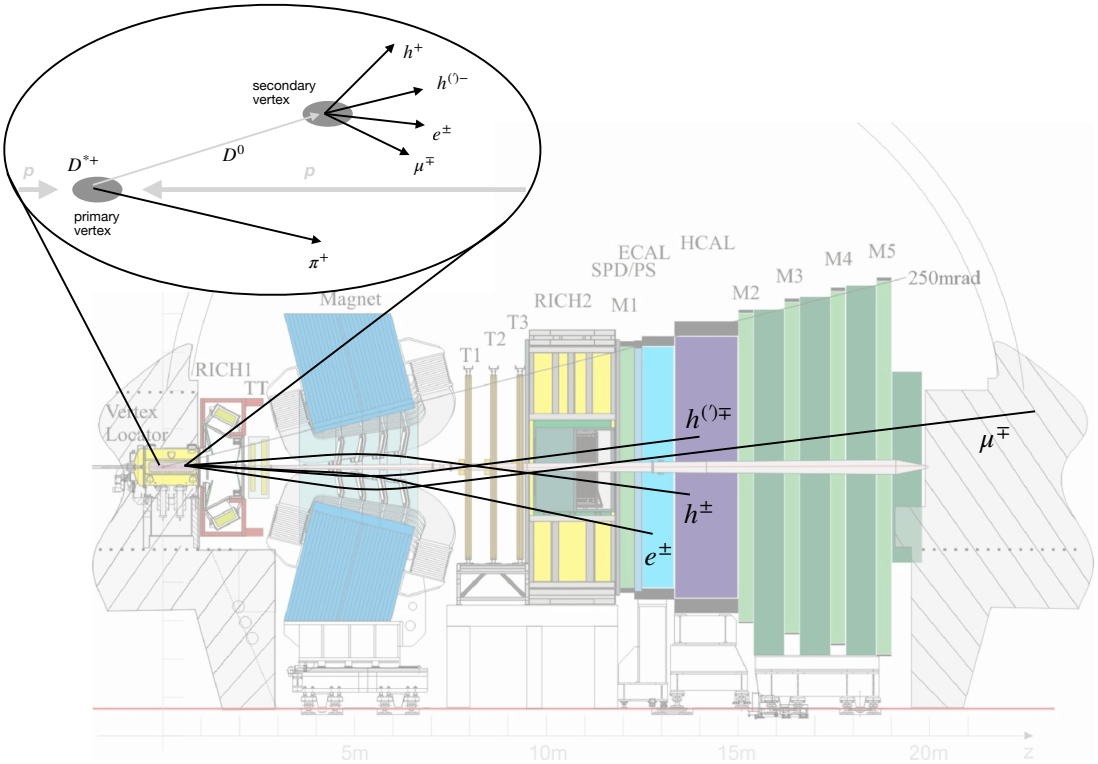


Figure 3.7: Sketch of the $D^0 \to h^{(\prime)-}h^+\mu^\pm e^\mp$ decay topology with the magnet bending plan rotated by $\pi/2$ around the z-axis and underlaid by a projection of the detector. The relative scale, including the angles, is arbitrary and not realistic. The figure of the detector is taken from Ref. [68].

Electron energy loss is characterised by two types of interaction: ionisation and bremsstrahlung. At low energies, the energy loss of electrons is dominated by ionisation. The Bethe-Bloch formula describes the energy loss dE for charged particles in matter due to ionisation. The formula depends on the mean excitation energy, E, the particle mass, m, in the case of the electron $m = m_e$, and the nuclear charge of the medium it traverses:

$$\left(\frac{dE}{dx}\right)_{ion} \sim Z \cdot \ln\left(\frac{E}{m}\right) \tag{3.3.1}$$

Highly energetic, multiple GeV, electrons like the ones measured at LHCb undergo bremsstrahlung emission as a consequence of their interaction with matter. This mechanism

provides a second source of energy loss for charged particles. The electron energy loss due to bremsstrahlung is proportional to the energy of the electron:

$$\left(\frac{dE}{dx}\right)_{brem} = -\frac{E}{x_0} \tag{3.3.2}$$

resulting in

$$E(x) = e^{-\frac{x}{x_0}} \tag{3.3.3}$$

where x_0 is the so-called radiation length. The average distance an electron's energy to be reduced by a factor $1/e$. The radiation length is given by the medium and scales approximately by $1/Z^2$.

$$\left(\frac{dE}{dx}\right)_{brem} \sim Z^2 \cdot \frac{E}{m^2} \tag{3.3.4}$$

To see at which energies the energy loss is dominated by which process, the so-called critical energy, E_c, is calculated. The critical energy is the energy where the two energy loss mechanisms are the same:

$$\left(\frac{dE}{dx}\right)_{brem} = \left(\frac{dE}{dx}\right)_{ion}. \tag{3.3.5}$$

The resulting critical energy for electrons [74] is given by

$$E_c \approx \frac{800}{Z}\,\text{MeV} \tag{3.3.6}$$

A similar calculation can also be performed for example for the muon. The resulting rate of energy loss for muons is suppressed by a factor $m_e^2/m_\mu^2 \approx 2.5 \times 10^{-5}$ compared to electrons.

3.3.1 Reconstruction of electrons

To properly measure the electron energy at production to reconstruct the D^0 meson, the bremsstrahlung emitted within the detector needs to be measured and matched to the electron. Bremsstrahlung is predominantly emitted when electrons pass through the detector material. If this happens after the magnet, it does not influence the momentum estimate, which is calculated by the bending of the electron track in the magnetic field, as illustrated in figure 3.8a. However, around two-thirds of the electrons emit bremsstrahlung before the magnet. If bremsstrahlung is emitted before the magnet, the electron loses energy and velocity, consequently changing its momentum. This leads to an underestimation of the momentum.

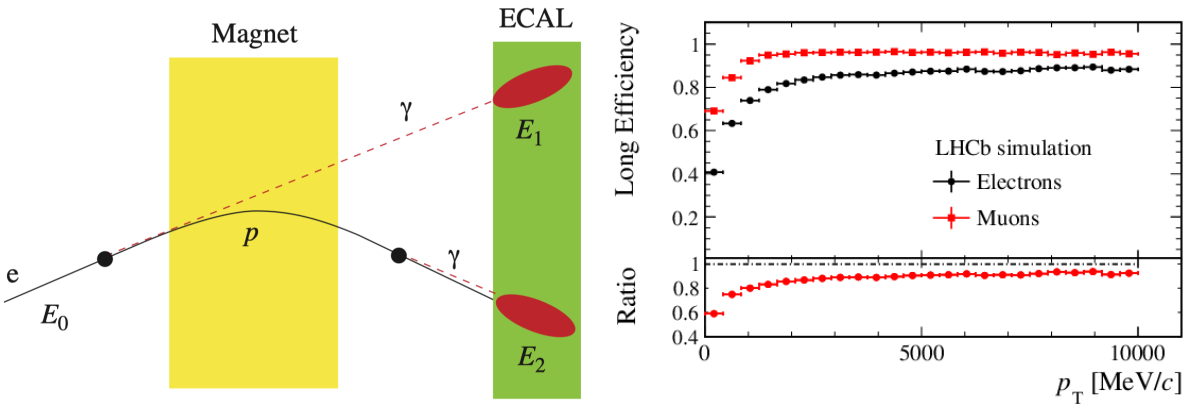


Figure 3.8: a) Schematic illustration of bremsstrahlung correction [76]. b) Track-reconstruction efficiency for muons and electrons as a function of the lepton's transverse momentum [75]

The photon also does not interact with the magnetic field (and the tracking stations), making a recovery of the lost momentum by the tracking system impossible. To correct for this loss in energy, the possible bremsstrahlung is matched to the respective energy cluster in the ECAL. This is done at the cost of a decreased resolution due to the added photon detected in the ECAL. A second consequence of the energy loss is that the particle's trajectory can be altered such that the electron no longer traverses all tracking stations, leading to a loss in the track reconstruction, or no or a wrong energy cluster is found. The efficiency for long tracks, tracks reconstructed using the VELO, TT and OT, significantly differs between electrons and muons, as shown in figure 3.8b[10], while the track reconstruction efficiency within the VELO, using only VELO information, is similar for muon and electrons [75].

3.4 Trigger

To be able to process the collisions at a rate of 40 MHz, a trigger system is used. The trigger allows to discard randomly detected particles and events that are not of interest. The trigger is built up in three successively executed steps called levels:

- Level-0 (L0): The Level-0 trigger is implemented on hardware level. Detector information is processed on dedicated hardware. The trigger decision is made based on either a high energy disposition E_T in the calorimeters or a high transverse momentum p_T[11] signature in the muon stations. In this step, the rate is reduced to

[10] Most, $\sim 68\%(95\%)$, of the electrons and muons studied in this analysis have a transverse momentum below 2000(5000) MeV/c.

[11] p_T is the momentum component in the plane perpendicular to the beam axis.

1 MHz.

- High-Level Trigger 1 (HLT1): The first software-based trigger is used to trigger on loose topological requirements, reducing the rate to 110 kHz. At this trigger level, the event is only partially reconstructed.
- High-Level Trigger 2 (HLT2): The second level software trigger is used to trigger on the fully reconstructed decay chain, which is reconstructed during the trigger process. After this, the events passing the trigger are written to an offline storage at a rate of 12.5 kHz.

An illustration of the trigger system can be found in figure 3.9.

The first trigger level, L0, differentiates between hadrons, electrons, photons and muons. The hadron, electron and photon trigger decisions called `L0Hadron`, `L0Electron`, and `L0Photon` are made by the energy disposition in the calorimeters. The thresholds were adjusted multiple times during Run 2, but typical values for the minimal E_T of the hadrons were $E_T \gtrsim 3.5\,\mathrm{GeV}/c^2$ and $E_T \gtrsim 3.0\,\mathrm{GeV}/c^2$ for electrons and photons. For the muon decision, called `L0Muon`, a hit in the last muon station and typically a value of $p_T \gtrsim 1.7\,\mathrm{GeV}/c$ is required.

At HLT1, simplified tracking and vertexing algorithms are performed using information from the tracking stations and the VELO. High (transverse) momentum tracks, typically $p_T \gtrsim 1.6\,\mathrm{GeV}/c$, originating from a secondary displaced vertex are searched and selected at this stage. The criteria are called `Hlt1Track` or, in the case of the two-track implementation, `Hlt1TwoTrack`. In the case that `L0Muon` was returned, a separate algorithm with a lower p_T is executed, $p_T \gtrsim 1\,\mathrm{GeV}/c$, called `Hlt1TrackMuon`. Instead of the static p_T and displacement criteria, also a selection based on multivariate analysis (MVA) techniques is implemented for the above-mentioned trigger decisions, notated by an additional MVA supplement, *i.e.* `Hlt1TrackMVA`. Also here, the specific requirements were adjusted during the data-taking period.

At HLT2, a full event reconstruction takes place, which includes information from all detector subsystems. At this stage, two categories of trigger decisions are defined: inclusive and exclusive trigger decisions. The inclusive selection requirements are typically based on topological properties like the number of displaced vertices and the number of final state particles. Also, loose kinematic properties are enforced. Especially in the case of rare (or forbidden) decays and charm decays, because of the large production

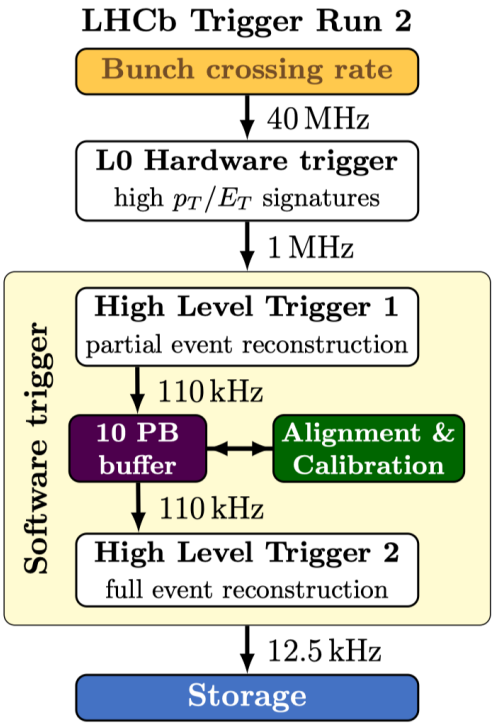


Figure 3.9: Overview of the LHCb trigger system [77].

cross section, typically exclusive trigger requirements are enforced. Decays passing the exclusive trigger decision are fully reconstructed with specific final states enforced by particle identification requirements. Specific exclusive lines are implemented for all the decays studied in this work. The detailed requirements are explained and detailed in chapter 5.

When talking about a trigger decision in context to one specific signal decay chain, one can differentiate between decays/events triggered by the signal candidates, `TOS`, and decays triggered independently of the signal candidates, `TIS`. A trigger decision on a signal candidate can be both `TIS` and `TOS`, simultaneous[12].

3.5 Recorded data samples and simulation

In Run 2, data from proton-proton collisions from 2015-2018 at a centre-of-mass energy of 13 TeV corresponding to an integrated luminosity of $6\,\text{fb}^{-1}$ was collected. The typical

[12]Depending on the trigger decision, a third possibility exists. It is also possible that a triggered event is neither `TIS` nor `TOS`. This is the case if neither the presence of the signal alone nor the rest of the event alone are sufficient to generate a positive trigger decision, but both are necessary.

Run 2 data flow at LHCb starts with the trigger selection, first hardware (L0) and then software (HLT). For each trigger level, all events passing any of the available trigger decisions are saved. After this, a so-called stripping selection is applied to the complete dataset[13]. The stripping selection consists of various topological, kinematic and reconstruction quality requirements tuned for the desired decay one wants to study. Particle candidates are selected and assigned to their respective hypotheses during this step. All necessary variables needed for further analysis, including trigger decision criteria, are assigned, calculated and provided in a standardised way for further analysis. After this, the desired trigger selection is reapplied to the dataset, and additional selection steps can be applied.

In this measurement, simulated samples are used for efficiency estimation and the extraction of line shapes when the mass distributions in data are fitted. Simulated samples for each year of data taking are generated. The simulation samples are processed in the same way and under the same conditions as the data. `Pythia` [65, 78] is used to generate events containing B- and D-mesons within the LHCb simulation software framework `GAUSS` [79]. The `EvtGen` [80] package then simulates the decay of the mesons. For the forbidden decays, no intermediate resonances are assumed in the simulation. For the allowed $D^0 \to h^- h^+ \mu^- \mu^+$ and $D^0 \to K^- \pi^+ \mu^- \mu^+$ decays, a sum of resonant and non-resonant contributions without interference is assumed. The different assumed intermediate resonances and the fraction to the total simulation sample are listed in table 3.1. Particle interactions with the detector are simulated with the `GEANT4` package [81]. The digitalization of the detector response is done with the `BOOLE` package [82]. Afterwards, the events need to pass through the `MOORE` package [83] emulating the trigger.

After the trigger, or `MOORE`, the recorded and simulated data samples are passed to the Brunel application [84] for offline reconstruction. Within the DaVinci framework [85] so, so-called stripping lines, which are similar to HLT2 trigger lines, are applied on the reconstructed events. The output is then stored in so-called ROOT files, which allow further offline analysis steps.

[13]Alternatively, also a so-called Turbo selection was already applied to some decays in Run 2, saving immediately the output of the HLT2 step. This was made standard for Run 3.

Channel	Component	Fraction
$D^0 \to K^-K^+\mu^-\mu^+$	$\phi(1020)(\to K^-K^+)\rho^0(770)(\to \mu^-\mu^+)$	0.25
	$\phi(1020)(\to K^-K^+)\mu^-\mu^+$	0.25
	$K^-K^+\rho^0(770)(\to \mu^-\mu^+)$	0.25
	$K^-K^+\mu^-\mu^+$	0.25
$D^0 \to K^-\pi^+\mu^-\mu^+$	$K^{*0}(892)(\to K^-\pi^+)\rho^0(770)(\to \mu^-\mu^+)$	$0.16\bar{6}$
	$K^{*0}(892)(\to K^-\pi^+)\omega(782)(\to \mu^-\mu^+)$	$0.16\bar{6}$
	$K^{*0}(892)(\to K^-\pi^+)\mu^-\mu^+$	$0.16\bar{6}$
	$K^-\pi^+\rho^0(770)(\to \mu^-\mu^+)$	$0.16\bar{6}$
	$K^-\pi^+\omega(782)(\to \mu^-\mu^+)$	$0.16\bar{6}$
	$K^-\pi^+\mu^-\mu^+$	$0.16\bar{6}$
$D^0 \to \pi^-\pi^+\mu^-\mu^+$	$\rho^0(770)(\to \pi^-\pi^+)\rho^0(770)(\to \mu^-\mu^+)$	0.125
	$\rho^0(770)(\to \pi^-\pi^+)\omega(782))(\to \mu^-\mu^+)$	0.125
	$\rho^0(770)(\to \pi^-\pi^+)\phi(1020)(\to \mu^-\mu^+)$	0.125
	$\rho^0(770)(\to \pi^-\pi^+)\mu^-\mu^+$	0.125
	$\pi^-\pi^+)\rho^0(770)(\to \mu^-\mu^+)$	0.125
	$\pi^-\pi^+)\omega(782)(\to \mu^-\mu^+)$	0.125
	$\pi^-\pi^+)\phi(1020)(\to \mu^-\mu^+)$	0.125
	$\pi^-\pi^+\mu^-\mu^+$	0.125

Table 3.1: Simulated composition in terms of dimuon and dihadron intermediate resonances, modified from Ref. [23].

Chapter 4

Analysis Strategy and Statistical Methods

This chapter provides an overview of the tools and more complex statistical methods used in this analysis. Methods to calculate the uncertainties of efficiencies, optimise the event selections and evaluate limits are reviewed. For the event selection, this includes the principles of a multivariate classifier and the calculation and data-driven correction in simulations of the particle identification variables. The Punzi figure of merit [86], a measure of the statistical sensitivity of the selection procedure, is introduced. The limit-setting procedure is explained, and lastly, an overview of the analysis strategy is provided.

4.1 Bayesian error estimation

Usually, uncertainties of efficiencies are calculated using a binomial error estimate. However, this description of the uncertainties is not necessarily true. Efficiencies are only well defined between zero and one, inclusive. Therefore, a prior is introduced in addition to the binomial error estimation. This is motivated by the argumentation in Ref. [87] and leads to a more intuitive uncertainty estimate as the probability is set to zero for efficiency values above and below one. Using the Bayesian theorem, the following ansatz is made:

$$P(\epsilon; k, n) = \frac{P(k; \epsilon, n) P(\epsilon; n)}{C} \tag{4.1.1}$$

where C is the overall normalisation constant,

$$C = \frac{1}{n+1}, \tag{4.1.2}$$

$P(\epsilon; n)$ is the prior,

$$P(\epsilon; n) = \begin{cases} 1 & \text{if } 0 \leq \epsilon \leq 1 \\ 0 & \text{otherwise} \end{cases}, \tag{4.1.3}$$

and $P(\epsilon; k, n)$ is the final efficiency probability function, given by a Binomial distribution,

$$P(\epsilon; k, n) = (n+1)\binom{n}{k}\epsilon^k(1-\epsilon)^{n-k}. \tag{4.1.4}$$

n is the number of events before the selection step, k after the selection step, and ϵ is the efficiency of the selection step. The factor $(n+1)$ is introduced by the normalisation. This now allows calculating the following variables:

- mean: $\bar{\epsilon} = \frac{k+1}{n+2}$
- mode: $\text{mode}(\epsilon) = \frac{k}{n}$ (most probable value; the efficiency)
- variance: $V(\epsilon) = \frac{(k+1)(k+2)}{(n+2)(n+3)} - \frac{(k+1)^2}{(n+2)^2}$

The square root of the variance gives the error in the efficiency.

4.2 Multivariate Analysis

A multivariate classification model is used to discriminate between signal decays and various background sources. This allows the combination of multiple features and observables into a single discriminating variable between signal and background candidates. In this analysis, a Gradient Boosted Decision Tree (BDT) classifier is used, available in the Toolkit for Multivariate Data Analysis in short TMVA [88, 89].

A decision tree is a binary tree-structured classifier, similar to the one sketched in figure 4.1. Multiple decision nodes, which use a single discriminating variable, categorise events as signal-like or background-like. These nodes are structured in a tree-like fashion, starting with a so-called root node and ending with leaf nodes. A single event is then categorised as background or signal depending on the classification of the leaf node it ends up during the training. The maximal depth of a tree is defined by the longest path in the tree from the tree root to a leaf. The root node has a depth of zero.

The tree is trained using a training sample. The training sample typically contains labelled background and signal events, created either by simulations or extracted from recorded

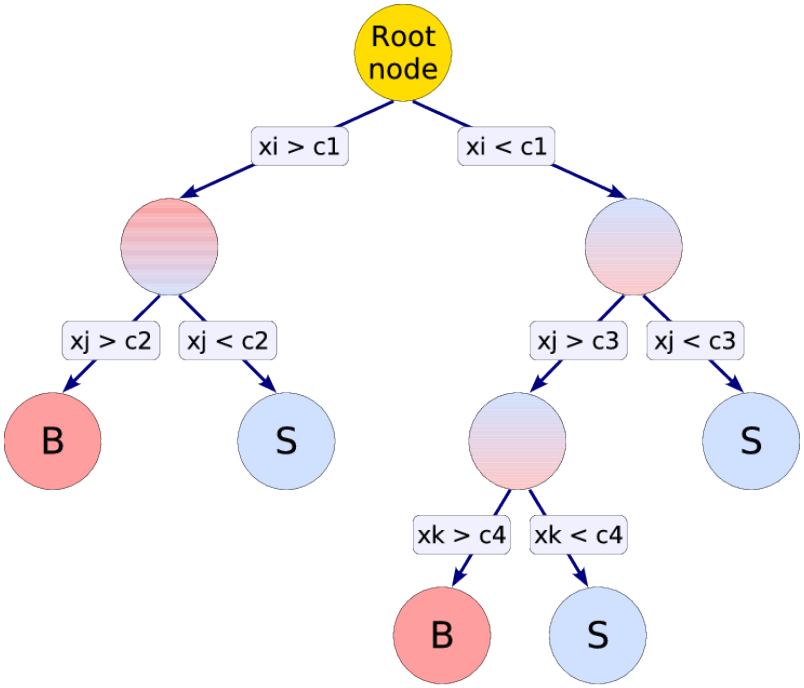


Figure 4.1: Sketch of a decision tree [89]. The variables x are split into two branches at each node by the cut value c_i. Signal and background leaves are marked by S and B.

events. Starting with the root node, the best variable and selection value to separate signal from background in the training sample is selected. The sample is then divided into two sub-samples by the selected cut criteria, and the process is repeated. This is done until a certain maximum depth of the tree is reached or a minimum number of events is reached at one of the nodes[1]. In the case of this measurement, until a depth of 2 or 2.5% of events from the training sample are left. The last created set of nodes, leave nodes, are then categorised as signal or background depending on the dominant training event type in the node.

A single decision tree can be easily interpreted and trained but is also sensitive to statistical fluctuations in the training sample. A so-called boosted decision tree is used to address this problem. For this, multiple decision trees are trained by means of a stochastic resampling of the initial training sample[2]. The parameters for each tree are determined by minimising the mis-classification rate via a binomial-likelihood loss function, or cross-entropy loss function [90]; because of this, the name Gradient Boosted Decision Tree.

[1]The purity of the sample is also commonly used.

[2]Another type of boosting commonly used, instead of resampling, is to re-weight the training sample based on the results of the previously trained decision tree.

4.3 Particle identification in data and simulations

The information from the particle identification systems is combined in two ways [91]. The first method calculates a log-likelihood difference, called DLL. A likelihood $\mathcal{L}_{\text{det}}$ for each particle for each sub-detector under different particle hypotheses is computed. The log-likelihoods are then added linearly to a combined log-likelihood $\log\mathcal{L}$. The log-likelihood difference of the tested particle hypothesis X is then calculated relative to the pion hypothesis, the most abundant particle in the detector:

$$\text{DLL}_X = \Delta\log\mathcal{L}(X - \pi), \tag{4.3.1}$$

where X can be the electron, muon, kaon, or proton hypothesis.

The second method uses a neural network, a multivariate analysis technique, to combine the information of all sub-detectors. This method also considers correlations between the information of the different subsystems. The neural network classifier, defined between zero and one, for each particle hypothesis x is referred to as `ProbNN`x. The more simple DLL classifier is typically used in the HLT 2 trigger stage, while the more complex, better-performing `ProbNN` classifier is used for offline analysis.

In general, in the LHCb simulation, the more complex `ProbNN` variables are very difficult to describe correctly. To enhance the agreement with data, a software package, `PIDGen2` [92, 93], exists, which discards the simulated PID variables and resamples them based on the observed PID variable distributions in calibration decays[3]. For this purpose is the PID variable distribution parameterised as a function of the number of tracks in the event, `nTracks`, the momentum, p, and the transverse momentum, p_{T}[4]. Instead of resampling the PID variable, x, according to the data distribution,

$$P_{exp}(x_{exp}|p_T, \eta, \texttt{nTracks}) = \int_{-\infty}^{x} p_{exp}(y|p_T, \eta, \texttt{nTracks})dy \tag{4.3.2}$$

in bins of `nTracks`, p and p_T the PID variable is corrected. For this the PID distribution, which is typical unknown in its exact representation, a kernel density estimation is used [94]. The *kernel density estimator* for the PID distribution is given by:

$$P_{KDE}(x_{exp}|p_T, \eta, \texttt{nTracks}) = \frac{1}{N\sigma}\sum_{i}^{N} K\left(\frac{x_{exp} - x_i}{\sigma}|p_T, \eta, \texttt{nTracks}\right), \tag{4.3.3}$$

[3] $Dz \to K^-\pi^+$ for kaons and pions, $J/\psi \to \mu^+\mu^-(e^+e^-)$ for muons (electrons).

[4] Any two variables of p,p_{T} and η do work for the binning.

with the kernel $K(x)$, given by gaussian distribution, the kernel width, or size, σ, number of events N and corrected PID variable $x_i\,(i = 1...N)$. The advantage of the kernel estimate, by construction, is that no knowledge about the exact model behind the process is not needed while still ensuring a good quality of the PDF description. Using a gaussian kernel also ensure a smooth distribution, different to the discreteness of a histogram. In the limit of large N with fixed kernel size σ, P_{KDE} corresponds to a convolution of the true PDF P_{exp} with the kernel $K(x)$. Asymptotically, for a large sample size and an infinitesimal small kernel size, P_{KDE} matches P_{exp}. However, for a limited sample size, the kernel size is balanced in such a way that the kernel is as narrow as possible while still allowing enough statistics within the kernel for a precise estimate of the PID variable. The kernel density estimation procedure is performed using the `Meerkat` library [94].

4.4 Punzi figure of merit

A measure to quantify the statistical sensitivity of the selection is needed. The Punzi figure of merit provides a definition of the said sensitivity that is unique and well-defined [86]. The definition is based on a frequentist approach, avoiding the choice of a prior distribution for the searched signal. The full derivation of the Punzi figure of merit can be found in Ref. [86]. The Punzi figure of merit of a given selection is calculated as follows:

$$FOM = \frac{\epsilon_{sig}}{\frac{a}{2} + \sqrt{N_{bkg}}}, \tag{4.4.1}$$

where ϵ_{sig} is the signal efficiency and N_{bkg} is the number of expected background candidates in the studied signal region. The tuneable character a corresponds to the significance in standard deviations. This measurement aims to find hints for new physics; in particle physics, this corresponds to a signal significance of 3σ; therefore, in this analysis, $a = 3$ is used.

4.5 Limit setting

This section reviews, based on Ref [95] and Ref [96], the limit-setting procedure used for this measurement. In the search for LFV in rare four-body charm decays, two scenarios need to be considered. Either the LFV decays are observed, in which case the signal branching fraction will be measured, or, in the case of no observation, an upper limit on the branching fraction will be established using the so-called $\mathrm{CL_s}$ method. The quoted limit corresponds to the upper limit of the observable's confidence interval. A confidence interval of a parameter in its frequentist definition

is called an interval containing the true value of the parameter at a given fraction, called confidence level (CL). The upper limit for a given CL is defined such that the fraction of limits exceeding the true value of the parameter of interest equals the chosen CL.

Numerous approaches have been suggested and are in use for establishing confidence intervals for rare signal searches. The extent to which the fraction of confidence intervals containing the true value of the parameter of interest matches the chosen CL for a particular method is called coverage. Methods were this fraction matches the CL are considered to have coverage, whereas those with a lower fraction have under-coverage. Methods with a fraction higher than the CL have over-coverage. An optimal method achieves coverage, but in practice, this is often reached only asymptotically with higher statistics of the measurement. The CL_s method is one such method.

4.5.1 Statistical tests

The general procedure of excluding or measuring a new signal is to define a null hypothesis H_0, which includes all known processes but not the sought-after signal, called background-only hypothesis, and an alternative hypothesis, against which is tested, H_1, which includes the sought-after signal as well as the background, called background + signal hypothesis. When setting a limit, this is reversed. The model with signal plus background hypothesis plays the role of H_0, which is tested against the background-only hypothesis H_1.

The agreement of the observed data with a given hypothesis H is usually quantified by computing a p-value. The p-value corresponds to the probability of finding data of equal or greater compatibility with the predictions of H, assuming H is true. A hypothesis is regarded as excluded if its p-value is observed below a specific threshold. This p-value is usually converted into an equivalent significance Z in particle physics. Z is defined such that a Gaussian distributed variable's upper tail probability equals p at Z standard deviations above its mean. The typical threshold in the particle physics community to claim a discovery is $Z = 5$, $p = 2.87 \times 10^{-7}$. Meanwhile, a signal hypothesis is treated as excluded at a threshold of $Z = 1.64$, $p = 0.05$. In terms of CL, this corresponds to 95%.

The result itself is still blind in the case of this analysis; therefore, only the sensitivity for an expected significance, Z, will be reported in this thesis. The expectation value of Z is calculated by testing the background-only hypothesis H_0 under the assumption of the signal plus background hypothesis, H_1. Because of the nonlinear behaviour between Z and

p, instead of the mean, the median of Z will be used to express the "expected significance"[5].

The confidence interval is calculated with the help of a test statistic. A commonly used test statistic, t_μ, is the Profile Likelihood Ratio

$$t_\mu = -2\ln\left(\frac{\mathcal{L}(\mu|\hat{\hat{\boldsymbol{\theta}}})}{\mathcal{L}(\hat{\mu}|\hat{\boldsymbol{\theta}})}\right) \tag{4.5.1}$$

where the likelihood of the data, for a hypothetical value μ, the branching ratio here, is divided by the likelihood obtained from the best fit to data. The quantity $\boldsymbol{\theta} = (\theta_1, \theta_2, ...)$ denotes the nuisance parameter of the probability density functions (PDF) from which the likelihood is constructed. The parameters maximising the likelihood for a specified μ are given by $\hat{\hat{\boldsymbol{\theta}}}$ and for best fit to data value $\hat{\mu}$ are given by $\hat{\boldsymbol{\theta}}$. The data itself might be distributed to a different true value μ'. The PDF $f(t_\mu|\mu')$ of the test statistic, assuming a large sample and parameters far from boundary conditions [97, 98], can then be derived to be:

$$f(t_\mu|\mu') = \frac{1}{2\sqrt{t_\mu}}\frac{1}{\sqrt{2\pi}}\left[\exp -\frac{1}{2}\left(\sqrt{t_\mu} + \frac{\mu-\mu'}{\sigma}\right) + \exp -\frac{1}{2}\left(\sqrt{t_\mu} - \frac{\mu-\mu'}{\sigma}\right)\right], \tag{4.5.2}$$

σ denotes the standard deviation of $\hat{\mu}$. Under null hypothesis, $\mu = \mu'$, the distribution simplifies to a χ^2-distribution. The p-value for an observed value μ_{obs} is

$$p_\mu = \int_{t_{\mu,\text{obs}}}^{\infty} f(t_\mu|\mu)dt_\mu. \tag{4.5.3}$$

If p_μ lies below the threshold $1 - \text{CL}$, the hypothesis is rejected.

4.5.2 CL_s method

The CL_s method was introduced by physicists at the LEP experiment at CERN [99] when providing exclusion limits from the Higgs search. The CL_s method is a modification of the p-value method to account for the fact that the data may not be sensitive to the potential signal, for example, because of background. The introduced tests allow setting a limit on the observed data with the signal plus background hypothesis. From a physical point of view, one wants to set a limit only on the signal rate; for this, the CL_s method was developed. In the CL_s method, the p-value of the signal plus background hypothesis is normalised to the one for the background-only hypothesis:

$$\text{CL}_\text{s} = \frac{\text{CL}_\text{s+b}}{\text{CL}_\text{b}} \tag{4.5.4}$$

[5]The relation between Z and p can be expressed as $Z = \phi^{-1}(1-p)$ with ϕ^{-1} being the quantile (inverse of the distribution) of the standard Gaussian. The equation is satisfied by the median of Z and p but not by the mean.

By design, the value CL_{s+b} has coverage in the parameter of interest, resulting in an over-coverage for CL_s, as $0 \leq CL_b \leq 1$. In the limit of a clear separation between signal plus background and background-only hypotheses, CL_b will be equal to unity, yielding $CL_s \approx CL_{s+b}$.

4.5.3 Treatment of systematic uncertainties

The branching fraction calculation and limit setting do not depend only on the signal fit. Instead, external nuisance parameters, such as the efficiency ratio or the branching ratio for the normalisation mode, with their respective uncertainties, need to be considered. Each parameter θ, usually with its own uncertainty and described by a Gaussian PDF, adds an additional degree of freedom described by an additional dimension of confidence region. However, no perfect method exists to project this multidimensional confidence region, under consideration of all correlations, onto the parameter of interest. The approach used in this thesis, provided and implemented in the `GammaCombo` framework [96], is to profile the likelihood. This means that all nuisance parameters are refitted at a given parameter of interest, and the likelihood is evaluated. In general, the test statistic can no longer be derived analytically for this approach and has, therefore, been determined by pseudo experiments. In the next step, when determining the test statistics distribution at a given value of the observable via the generation of pseudo experiments, the nuisance parameters are set to the values obtained from the profiled fit in the generation. The approach is commonly called the *Plugin* method. It can be shown that this method has approximately coverage [100, 101].

4.6 Analysis strategy

The analysis follows the following steps:

1. *Data selection.* A selection optimised for $D^0 \to h^{(\prime)-}h^+\mu^\pm e^\mp$ decays is applied. Two main sources of background are focused in the selection background from randomly combined particle tracks and wrongly identified decays with the same decay topology as $D^0 \to h^{(\prime)-}h^+\mu^\pm e^\mp$ decays. The selection is optimised for the best signal sensitivity using the Punzi figure of merit.

2. *Efficiency estimation.* Following the data selection steps, the efficiency ratio is estimated by simulations and, if needed, corrected by data-driven methods. This is done individually for each selection step.

3. *Signal significance estimation.* Leftover background candidates are determined by a maximum likelihood fit. Simultaneously, the significance of the signal yield, together with its observed value, is estimated. For this, a sophisticated model to describe the data distributions is evaluated. At this step also, the statistical sensitivity of the recorded data samples and branching fractions for $D^0 \to h^-h^+\mu^-\mu^+$ decays is calculated.

4. *Evaluation of systematic uncertainties.* Uncertainties introduced in the efficiency correction process are studied and evaluated. In addition, uncertainties due to the size of the simulation data sample are calculated with the help of a bootstrapping algorithm, and uncertainties regarding wrongly identified particles are evaluated using pseudo-experiments.

5. *Limit setting or branching ratio estimation.* The limit-setting procedure is set up for the blind data, and using the statistical sensitivity and the nuisance parameters with their respective uncertainties, an expected limit is calculated using the $\mathrm{CL_s}$ method.

The measurement itself and the procedure development are performed blind, without looking at the data. To be able to validate all steps of the branching ratio calculation, $D^0 \to h^-h^+\mu^-\mu^+$ decays are studied. All of the above steps, apart from the systematic uncertainties and the limit calculation, are performed for the reference decay $D^0 \to h^-h^+\mu^-\mu^+$. The resulting branching ratio can then be compared with a previous measurement [44].

Chapter 5

Reconstructing and selecting charm decays at LHCb

This chapter discusses the data reconstruction and selection. The discussion of trigger and stripping requirements are merged in the following to avoid redundancies and provide a better overview of the underlying selection requirements. After the trigger and stripping, further selection requirements are enforced to reduce the remaining background as efficiently as possible. Following these selection steps, a gradient-boosted decision tree is trained to remove remaining background events. The decision tree is then tuned for the best signal sensitivity in combination with PID requirements using a Punzi figure of merit [86].

All selected $D^0 \to h^{(\prime)-}h^+l^{(\prime)-}l^+$ decay candidates are required to arise from the decay chain $D^{*+} \to D^0(\to h^{(\prime)-}h^+l^{(\prime)-}l^+)\pi^+$, where the D^{*+} meson is produced in the primary pp interaction. The D^{*+} decays instantaneously in the primary interaction point via the strong interaction into a D^0 and a charged pion. Selecting D^0 mesons produced from D^{*+} helps to greatly reduce background contamination compared to selecting D^0 produced in the primary interaction. An additional advantage of this is that the pion identifies (tags) the flavour of the D^0 meson at the time of its production. The flavour of the D^0 allows separating unambiguously $D^{*+} \to D^0(\to h^{(\prime)-}h^+\mu^-e^+)\pi^+$ decays from $D^{*-} \to \bar{D}^0(\to h^{(\prime)-}h^+\mu^-e^+)\pi^-$ decays, with the help of the charge of the accompanying pion. By knowing the flavour of the D^0 it is possible to test different couplings λ_{ql}. This separation is done in the last step during the calculation of the branching ratio or its limit. The pion produced by the D^{*+} has, on average, a much smaller momentum ($\sim 5\,\mathrm{GeV}/c$) compared to the D^0 momentum ($\sim 100\,\mathrm{GeV}/c$). This is because of the small mass difference of the D^{*+} and its decay products. To avoid confusion, the pion is in the

following referred to as slow (or soft) pion π_s^+ to not be mixed up with pions originating from the D^0 meson.

The $D^0 \to h^- h^+ \mu^- \mu^+$ decay candidates and the normalisation mode $D^0 \to K^- \pi^+ \mu^+ \mu^-$ decay candidates are selected and reconstructed the same as $D^0 \to h^{(\prime)-} h^+ \mu^\pm e^\mp$ decays, apart from the requirements on the electron. The muon requirements are instead applied to both lepton candidates. Additionally, the normalisation mode is limited to muons originating from a ρ^0/ω resonance, notated $D^0 \to K^- \pi^+ [\mu^+ \mu^-]_{\rho^0/\omega}$. This is done by a cut in the reconstructed invariant dilepton mass limiting the range to the region dominated by the ρ^0/ω resonance, given by $675\,\mathrm{MeV}/c^2 < q < 875\,\mathrm{MeV}/c^2$, where q is the square-root of the squared four-momentum of the dimuon system.

One of the main background sources studied throughout this thesis are *misidentified hadronic decays* of the form $D^0 \to h^{(\prime)-} h^+ \pi^- \pi^+$. This background persists due to its large production rate and similar decay topology to the searched LFV decays. The branching fraction for $D^0 \to K^- \pi^+ \pi^+ \pi^-$ decays is around $\mathcal{O}(10^{-2})$ and the branching fractions for $D^0 \to K^- K^+ \pi^- \pi^+$ and $D^0 \to \pi^- \pi^+ \pi^- \pi^+$ decays are both around $\mathcal{O}(10^{-3})$ [102]. Separate simulation samples are produced to study these decays. Due to the high inefficiency of the selection for these decays, a slightly altered set of particle identification requirements is applied to these decays, detailed in section 5.4. Apart from this, the hadronic modes are processed in the same way as the $D^0 \to h^{(\prime)-} h^+ \mu^\pm e^\mp$ decays. This background is suppressed with the help of PID requirements.

The second main background source are decay candidates made of *randomly combined tracks*. These tracks are most often random pion tracks, as pions are the most abundant particle species produced in pp collisions. Another source of these tracks are so-called ghost tracks. These ghost tracks are produced when unrelated hits in the different tracking stations are combined and used to reconstruct a fake (ghost) particle. This background is suppressed by constraining the decay topology and its kinematic. Specifically for ghost tracks also, the `ProbNN` variable `TRACKGhostProb` to identify ghost candidates exists, which allows to reduce the background from ghost candidates to a negligible level.

For simplicity reasons the following nomenclature is used throughout this thesis for the different decay channels:

- *LFV signal channel*: $D^0 \to K^- K^+ \mu^\pm e^\mp$, $D^0 \to K^- \pi^+ \mu^\pm e^\mp$ and $D^0 \to \pi^- \pi^+ \mu^\pm e^\mp$ decays

- *Reference or dimuon decay channel*: $D^0 \to K^-K^+\mu^-\mu^+$ and $D^0 \to \pi^-\pi^+\mu^-\mu^+$ decays
- *Normalisation channel*: $D^0 \to K^-\pi^+[\mu^+\mu^-]_{\rho^0/\omega}$ decay

For the two main background sources, the following nomenclature is used:

- *(Fully) hadronic* or *misID* background: $D^0 \to K^-K^+\pi^-\pi^+$, $D^0 \to K^-\pi^+\pi^+\pi^-$ and $D^0 \to \pi^-\pi^+\pi^-\pi^+$ decays, where two oppositely charged pions are misidentified as leptons, either as $\mu^-\mu^+$ or $\mu^\pm e^\mp$.
- *Combinatorial* background: background from randomly combined tracks.

A discussion of other, at this step neglected background sources can be found in section 8.1. For completeness, the selection requirements reported combined in this chapter can be found separated for stripping, HLT2 lines and additional applied cuts in appendix A.

5.1 Trigger requirements

This measurement uses two L0 trigger requirements: `L0Muon` and `L0Electron`. `L0Muon` requires a straight muon track with a minimum transverse momentum, p_T, with respect to the beam pipe[1]. `L0Electron` requires a minimum of deposited transverse energy in a reconstructed calorimeter cluster[2]. A cluster consists of 2×2 calorimeter cells. In addition, a "global event cut" is applied which requires fewer than 450 hits in the SPD, rejecting very busy events in favour for more simple and faster-to-reconstruct events. This does not result in a significant loss of absolute signal efficiency compared to not applying the multiplicity requirement [77]. The signal candidates selected by the stripping selection are required to fulfil at least one of the following hardware-trigger requirements:

- The muon candidate triggers `L0Muon` or
- independently of the signal signature, `L0Muon` or `L0Electron` is triggered by the rest of the event.

[1] If not otherwise defined the transverse momentum is always relative to the beam pipe. The minimum transverse momentum is 2.8 GeV/c in 2015, 1.8 GeV/c in 2016 and 1.35 GeV/c in 2017 and 2018

[2] The minimum transverse momentum is 2.7 GeV in 2015, 2.4 GeV in 2016 and 2.11 GeV in 2017 and 2018.

At HLT1 the D^0 or at least one of its daughter tracks needs to be reconstructed[3]. Due to the large number of charm decays produced, dedicated HLT2 lines are designed for each decay mode. The D^{*+} (for 2017 and 2018) or the D^0 (for 2015 and 2016) needs to trigger their dedicated exclusive line. The HLT2 trigger line used in this analysis combines two leptons into a dilepton and two hadrons into a dihadron object, which is further combined into a D^0. With the start of 2017, the HLT2 line has been modified such that a D^{*+} candidate is formed at the trigger level, while only a D^0 was formed before.

The selection steps between signal and normalisation modes are aligned whenever possible. In this way, most systematic uncertainties or possible biases will be largely cancelled when calculating the ratio of efficiencies. The same L0 and HLT1 requirements are chosen for the normalisation channel. Since the normalisation mode contains two final-state muons, a randomly selected one is considered in the L0 selection. The other muon is not considered in the trigger selection, similar to the electron in the signal modes.

5.2 Decay reconstruction and first selection steps

The aim of the reconstruction and selection is to build events with the desired topology and particle composition. During this process, obvious background and poorly reconstructed candidates are also removed. Candidates are reconstructed in a standardised bottom-up approach. The five final-state particles are combined to build the complete decay chain, $D^{*+} \to D^0(\to h^{(\prime)-}h^+\mu^\pm e^\mp)\pi_s^+$. Subsequently, the entire decay chain is refitted using a `DecayTreeFitter` (DTF). The DTF simultaneously fits the full decay chain under the constraint that the D^0 and π_s^+ originate from the same point (D^{*+}), called primary vertex (PV), using a Kalman filter [103]. Energies are calculated from the measured momentum of the particle under a given mass hypothesis, *i.e.* $E = \sqrt{m^2 + \mathbf{p}^2}$. The mass m is the known PDG mass of the assumed particle [104]. Unlike other large particle detectors, for example ATLAS, the energy deposits measured in the calorimeter system are not taken into account[4]. Due to the better resolution than that of the energy deposited in the calorimeter system, the momentum components p_x, p_y and p_z measured in the tracking stations are used. This does not hold true in the case of bremsstrahlung photons, for which the measured energy deposit in the calorimeter is added to the electron's energy. Consequently, the masses of the combined objects are computed from the sum of the four

[3]The D^0 or one of the daughters has to trigger `Hlt1TrackMuon` or `Hlt1TrackMVA` or `Hlt1TwoTrackMVA`.
[4]This information is mainly used for L0 trigger decisions.

momenta $p = (E, p_x, p_y, p_z)$ of the final-state particles.

A minimum transverse momentum condition is imposed on the final-state tracks and combined objects. This is done to remove large amounts of combinatorial background, dominant in these regions, and due to bandwidth limitations of the reconstruction algorithm.

Starting with the four final state particles, all four tracks should provide a minimum track quality (Track $\chi^2/\text{dof} < 3$) and originate from a secondary vertex. This is done with the impact parameter (IP) $\Delta\chi^2_{\text{IP}}$ variable, which provides a measure of the compatibility of a particle coming from the primary vertex. A sketch of the decay topology can be found in figure 5.1. $\Delta\chi^2_{\text{IP}}$ is the difference in χ^2 obtained when fitting the PV with and without the considered particle, therefore taking into account the uncertainties on the track parameters and the position of the primary vertex. A value of $\Delta\chi^2_{\text{IP}} > 3$ is required.

One final state particle needs to have the `isMuon` tag, meaning that a reconstructed track in the tracking station was matched to hits in the muon station [105].

Another final state particle needs to pass a loose electron PID requirement, $\text{DLL}_e > -2$. Stronger PID requirements for both leptons are chosen in a later stage in the selection.

The kaon candidate is selected by requiring a loose kaon PID requirement, $\text{DLL}_K > -5$. Both pion and kaon candidates are then required to pass a loose PID selection, ProbNNpi > 0.5 and ProbNNK > 0.5, respectively.

The artificial ghost tracks created by the reconstruction software are reduced by selecting tracks that satisfy the ghost probability criteria (ProbNNghost < 0.3). The L0 trigger decision relies strongly on the muon. Later, when the trigger efficiency correction is performed, the efficiency is described as a single-particle efficiency. Therefore, only muons that do not share hits in the muon chambers with other muons are selected.

To reduce the bandwidth used by the HLT2 line, first the secondary vertex is located, and then the four particles are combined. Given a large number of particles, mainly pions, produced in the primary collision, combining four particles immediately and then identifying if the vertex is displaced is extremely resource-demanding and, therefore, not done. To identify the secondary vertex, first two lepton candidates are combined

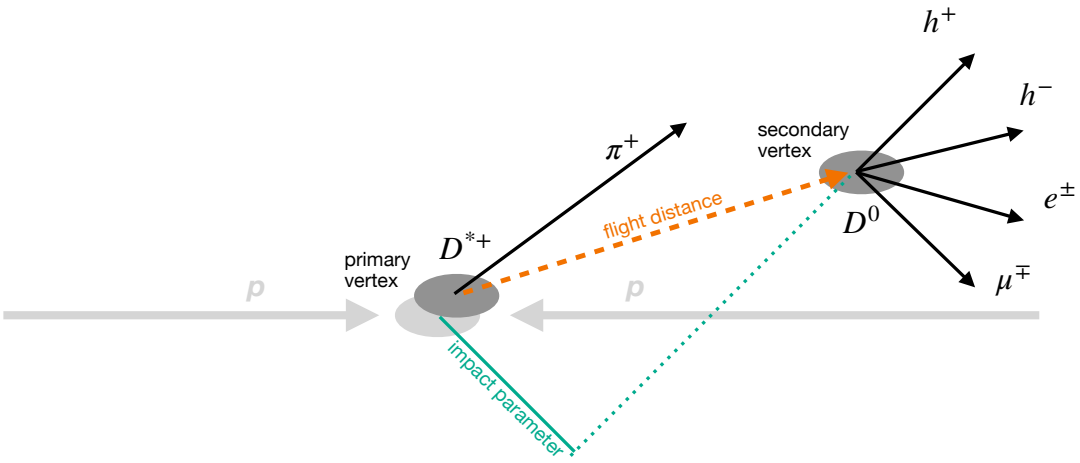


Figure 5.1: Schematic view of the decay topology of $D^0 \to h^{(\prime)-}h^+\mu^\pm e^\mp$ decays arising from $D^{*+} \to D^0\pi^+$ decays. The relative scale, including the angles, is arbitrary and not realistic. The flight distance for the D^0 and the impact parameter of the h^+ are indicated.

into a simpler dilepton object whose decay vertex then needs to be displaced from the primary vertex. To combine both leptons into a dilepton object, both candidates must be consistent with coming from a common vertex set by a cut on the distance of the closest approach (DOCA) between the two particles. In addition, an upper limit is set on the combined invariant mass of the two leptons, and the sum of transverse momenta must be positive. Both cuts are chosen such that all (correctly reconstructed) signal events pass them and only obvious background candidates are removed. The idea at this stage is to define simple and quickly executable selection steps to keep the complexity low. The two leptons are then combined into a dilepton object, with a physical, not negative, flight distance, the distance between the primary and secondary vertex. Due to resolution effects, a negative flight distance is possible, but this also means that the secondary and primary vertex are not distinguishable. The displacement of the secondary vertex is then enforced by the flight distance $\Delta\chi^2_{\text{FD}}$ ($\Delta\chi^2_{\text{FD}} > 20$) similar to $\Delta\chi^2_{\text{IP}}$, the variable combines the information of the flight distance, FD, and the uncertainties on the fit variables.

If all four tracks are consistent in coming from the same secondary vertex and fulfil loose p_T requirements to remove fake events, they are combined into a D^0 candidate. By requiring that at least one of the daughter particles is incompatible with the primary vertex, it is ensured that the daughters originate from a secondary vertex. A D^0 decay vertex is fitted to tracks that meet all the requirements mentioned above and the additional requirement on the quality of the vertex $\chi^2/\text{dof} < 15$. Here again,

the secondary D^0 vertex needs to be significantly displaced from the primary vertex, $\Delta\chi^2_{\mathrm{FP}} > 49$. This time, $\Delta\chi^2_{\mathrm{FP}}$ is calculated using the vertex constraint by all four daughter particle tracks, which, in principle, supersedes the previous $\Delta\chi^2_{\mathrm{FP}}$ cut. However, this is only partially true because of technical reasons[5]. The cosine of the angle between the reconstructed momentum vector of the D^0 and the vector connecting the primary, $D^{*+} \to D^0\pi^+$, and secondary, $D^0 \to h^{(\prime)-}h^+\mu^\pm e^\mp$, vertex is required to be larger than 0.9999. This ensures that the flight direction and reconstructed momentum match within their resolution, *i.e.* the cosine of their enclosed angle is close to unity. This is not necessarily true for combinatorial background. The D^0 originates from the primary vertex (D^{*+}). A requirement on the impact parameter $\Delta\chi^2_{\mathrm{IP}}(< 25)$ suppresses the background originating from secondary decay vertices. The sum over the square root of the different daughter $\Delta\chi^2_{\mathrm{IP}}$ values, given by $\sum\sqrt{\Delta\chi^2_{\mathrm{IP}}}$, has to be larger than 12, ensuring again a displaced secondary vertex. The reconstructed D^0 mass is limited to the range $1800 - 1940\,\mathrm{MeV}/c^2$.

Lastly, a low-momentum pion with $p > 120\,\mathrm{MeV}/c$ is added to form the D^{*+} decay vertex. A minimum vertex quality is required (Vertex $\chi^2/\mathrm{dof} < 25$). A window in the mass difference $\Delta m := m(D^{*+} - D^0)$, is defined such that $3\,\sigma = 99.73\%$ of signal events pass the selection requirement. The number is evaluated from simulation by calculating the standard deviation σ of the mass resolution with the help of a Gaussian fit to the simulation dataset. This cut removes around 90% of the background candidates. The cut is best illustrated in $D^0 \to K^-\pi^+[\mu^+\mu^-]_{\rho^0/\omega}$ decays, see figure 5.2. The signal decays and misID background decays both peak in the Δm distribution. In the left figure, the Δm mass window requirement is shown. In the right figure, the reconstructed D^0 mass $m(D^0)$ distribution under signal hypothesis[6] before and after the Δm cut is reported. In the $m(D^0)$ distribution, two peaks close to each other can be observed. The smaller peak at the D^0 mass corresponds to the signal, the larger peak left of the D^0 mass is due to misidentified $D^0 \to K^-\pi^+\pi^+\pi^-$ decays. The same is also observed for the two reference decays, $D^0 \to \pi^-\pi^+\mu^-\mu^+$ and $D^0 \to K^-K^+\mu^-\mu^+$, shown in figure 5.3. Note here that in the two pion case, $D^0 \to \pi^-\pi^+\mu^-\mu^+$ decays, significantly more combinatorial background can be observed. This is because most of the combinatorial background is from random pions, which are vetoed in the reconstruction for $D^0 \to K^-K^+\mu^-\mu^+$ and

[5]From a technical point of view, as the dilepton object is only constructed for the trigger decision and then discarded, it is impossible to recover the dilepton object easily. However, it is possible to access all other variables by exclusion. The $\chi^2_{\mathrm{FD}} > 49$ requirement of the dilepton object does not supersede the requirements on the positive flight distance and $\Delta\chi^2_{\mathrm{FD}} > 20$.

[6]If not specifically mentioned, the mass distributions are always reconstructed under the respective signal hypothesis.

$D^0 \to K^-\pi^+[\mu^+\mu^-]_{\rho^0/\omega}$ decays by the DLL requirement on the kaon(s).

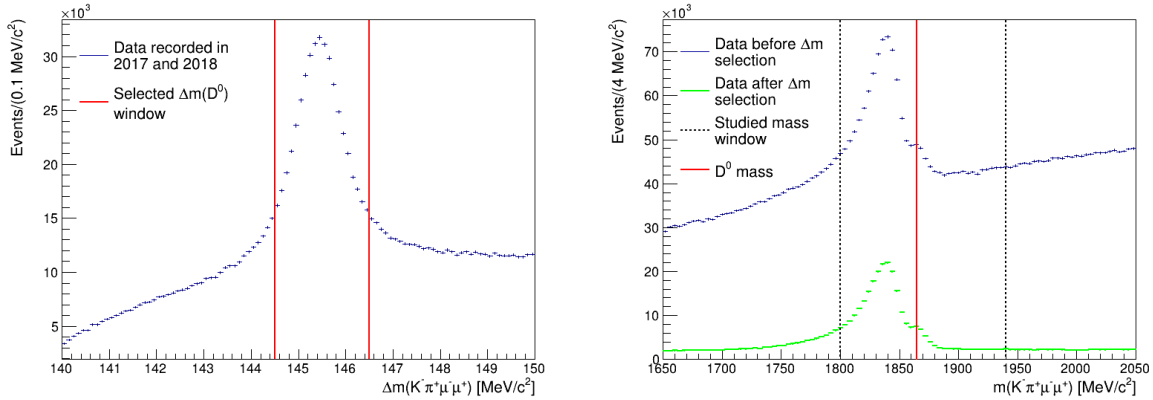


Figure 5.2: The under $D^0 \to K^-\pi^+[\mu^+\mu^-]_{\rho^0/\omega}$ hypothesis reconstructed $\Delta m := m(D^{*+} - D^0)$ (left) and D^0 (right) mass distribution for data recorded in 2017 and 2018 passing the $D^0 \to K^-\pi^+[\mu^+\mu^-]_{\rho^0/\omega}$ selection requirements. Left: The selected Δm mass window is marked by the red lines. Right: In blue and green is the data distribution before and after the Δm cut shown, respectively. The D^0 mass [35] is marked by a red line. Signal events are expected to be distributed around the D^0 mass while the large observed peak left of the D^0 mass is due to misidentified background events.

All the selection requirements mentioned above are summarised in table 5.1.

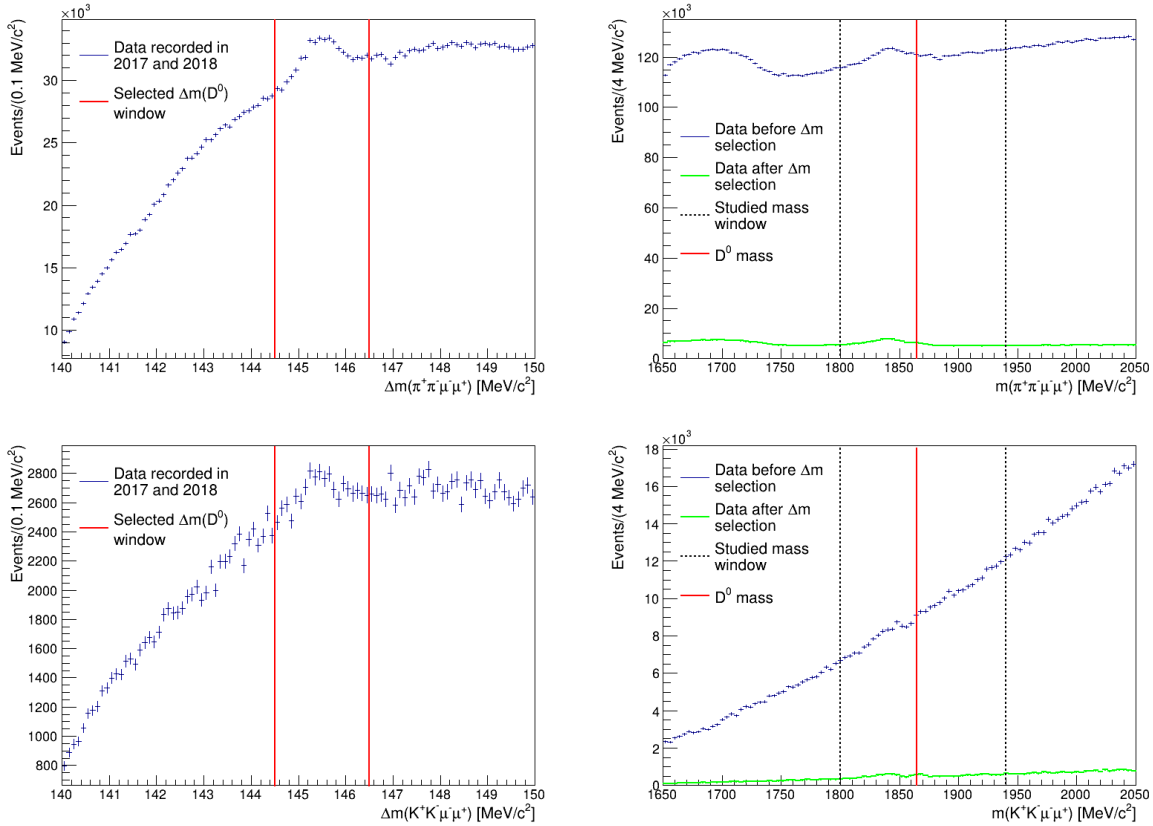


Figure 5.3: The under $D^0 \to \pi^-\pi^+\mu^-\mu^+$ (up) and $D^0 \to K^-K^+\mu^-\mu^+$ (down) hypothesis reconstructed $\Delta m := m(D^{*+} - D^0)$ (left) and D^0 (right) mass distribution for data recorded in 2017 and 2018 passing the $D^0 \to \pi^-\pi^+\mu^-\mu^+$ and $D^0 \to K^-K^+\mu^-\mu^+$ selection requirements, respectively. Left: The selected Δm mass window is marked by the red lines. Right: In blue and green is the data distribution before and after the Δm cut shown, respectively. The D^0 mass [35] is marked by a red line. Signal candidates are expected to be distributed around the D^0 mass, while the large observed peak left of the D^0 mass is due to misidentified background events. The second peak, the furthest left peak in the $D^0 \to \pi^-\pi^+\mu^-\mu^+$ distribution, is due to wrongly identified $D^0 \to K^-\pi^+\pi^+\pi^-$ events.

Particle	Variable		signal modes (normalisation)	
			2015-2016	2017-2018
h, μ, e	p	$>$	3 GeV/c	3 GeV/c
	p_T	$>$	300 MeV/c	300 MeV/c
	Track χ^2/dof	$<$	3	3
	`TRACKGhostProb`	$<$	0.3	0.3
μ	$\Delta\chi^2_{\text{IP}}$	$>$	3	3
	`isMuon`		true	true
	`NShared`	$=$	0	0
e	$\Delta\chi^2_{\text{IP}}$	$>$	3	3
	DLL_e	$>$	-2	-2
K	DLL_K	$>$	-5	-5
	`ProbNNK`	$>$	0.5	0.5
π	`ProbNNpi`	$>$	0.5	0.5
$(\ell\ell)$	m$(\ell\ell)$	$<$	2100 MeV/c^2	2100 MeV/c^2
	$\sum p_T$	$>$	0. MeV/c	0. MeV/c
	distance of closest approach	$<$	0.1 mm	0.1 mm
Dilepton object	$\Delta\chi^2_{\text{FD}}$	$>$	20(9)	20(9)
	flight distance	$>$	0 mm	0 mm
$(hh\ell\ell)$	max p_T	$>$	0. MeV/c	0. MeV/c
	$\sum p_T$	$>$	3000 MeV/c	3000 MeV/c
	minimal distance of closest approach	$<$	0.1 mm	0.2 mm
	maximal distance of closest approach	$<$	0.2 mm	0.3 mm
	$\Delta\chi^2_{\text{IP}}$ of at least one daughter particle	$>$	9	9
D^0	m(D^0)	$>$	1800 MeV/c^2	1700 MeV/c^2
		$<$	1950 MeV/c^2	2050 MeV/c^2
	Vertex χ^2/dof	$<$	15	15
	$\Delta\chi^2_{\text{FD}}$	$>$	49	49
	cosine angle of direction of flight	$>$	0.9999	0.9999
	$\Delta\chi^2_{\text{IP}}$	$<$	25	25
	$\sum\sqrt{\Delta\chi^2_{\text{IP}}}$	$>$	12	8
π_s	p_T	$>$	–	120 MeV/c
D^{*+}	Δm	$>$	144.5 MeV/c^2	144.5 MeV/c^2
		$<$	146.5 MeV/c^2	146.5 MeV/c^2
	Vertex χ^2/dof	$<$	–	25

Table 5.1: Summary of all selection requirements for all signal and normalisation candidates in the trigger and the following selection. Due to the added D^{*+} reconstruction in the trigger selection between 2015-2016 and 2017-2018, the selection requirements are separately listed for the two data-taking periods.

5.3 TMVA and PID selection

After the basic selection covering the decay topology, kinematics, and vertex quality, two types of background remain. These are the background from misidentified particles and combinatorial background. The background due to misidentified particles consists mainly of wrongly identified $D^0 \to h^{(\prime)-}h^+\pi^-\pi^+$ decays. This background is also observed when studying the reconstructed D^0 mass distribution for $D^0 \to h^-h^+\mu^-\mu^+$ decays and the normalisation decay $D^0 \to K^-\pi^+[\mu^+\mu^-]_{\rho^0/\omega}$ in figure 5.4. The large visible peak left to the expected D^0 mass [35], which is marked by the red line, are expected to arise from wrongly reconstructed $D^0 \to h^{(\prime)-}h^+\pi^-\pi^+$ decays[7]. This is because of the missing energy in the reconstructed mass, as the muon mass is wrongly assumed for the pion in the calculation, shifting the distribution to the left (lower mass direction) of the D^0 mass. Wrongly identified hadronic decays are suppressed by stronger particle identification requirements. Both selection decisions are optimised simultaneously to increase the significance of the signal. The second observed source of background is flatly distributed in the reconstructed invariant D^0 mass distribution. This background is labelled as combinatorial background. The combinational background is suppressed with the help of a trained multivariate classifier (TMVA).

5.3.1 TMVA

The combinatorial background is reduced with the help of a TMVA using the TMVA framework [88]. For this purpose, a boosted decision tree, BDT, is trained. The BDT is trained against a signal and background data sample. As background proxy, events in with no recovered bremsstrahlung in the upper D^0 sideband, $1900\,\mathrm{MeV}/c^2 < m(D^0) < 2050\,\mathrm{MeV}/c^2$, are used. This upper sideband is chosen to avoid any signal contamination due to the ECAL resolution or wrongly added bremsstrahlung. This is illustrated in figure 5.5, where the D^0 mass resolution for simulated $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ decays is shown for decay candidates with and without recovered bremsstrahlung. The distribution showing events with no reconstructed bremsstrahlung is distorted to lower masses, featuring a long tail, because of lost bremsstrahlung photons, which leads to an underestimation of the momentum and, therefore, the reconstructed invariant D^0 mass. The relatively sharp drop

[7]The second large peak in the D^0 distribution for $D^0 \to \pi^-\pi^+\mu^-\mu^+$ decays outside of the studied mass window is due to wrongly identified $D^0 \to K^-\pi^+\pi^+\pi^-$ decays, where additionally the kaon is misidentified as pion. The slope visible over the complete reconstructed mass spectrum in the D^0 distribution for $D^0 \to K^-K^+\mu^-\mu^+$ decays is also expected to arise from misidentified $D^0 \to K^-\pi^+\pi^+\pi^-$ decays, with the third pion being additionally misidentified as kaon.

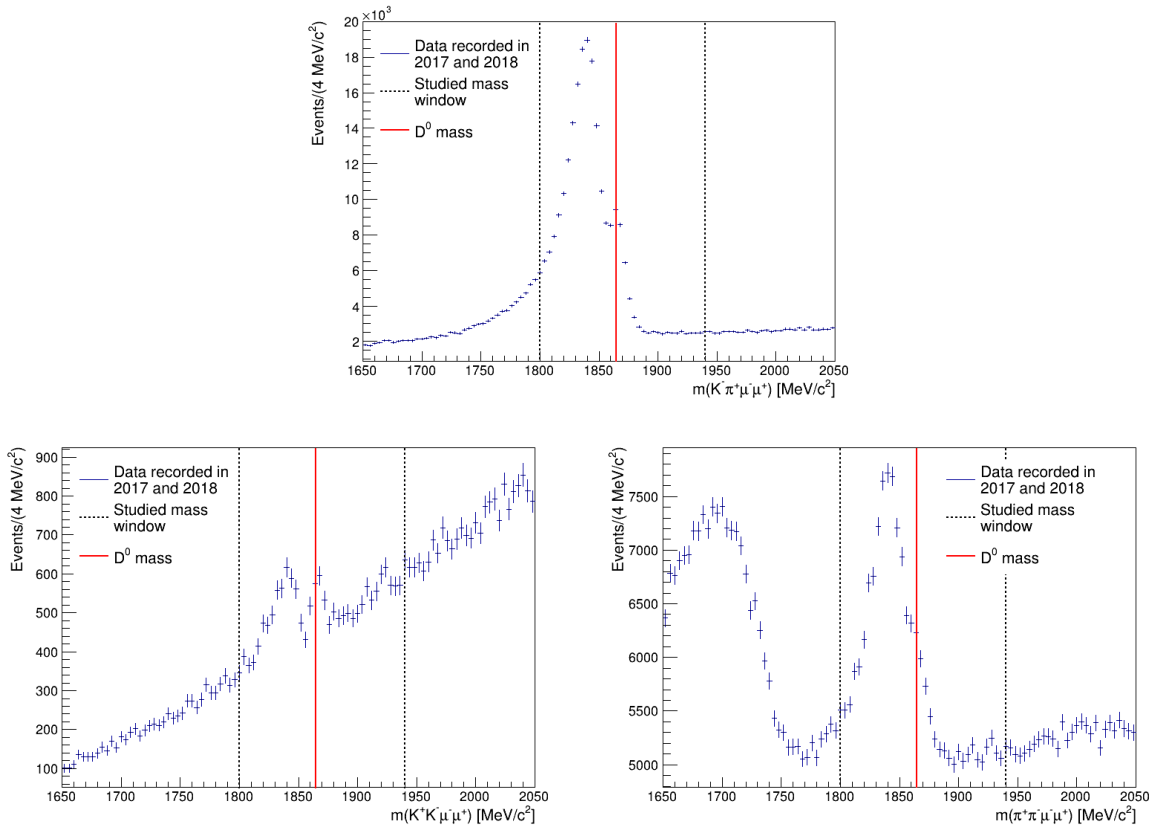


Figure 5.4: Reconstructed invariant D^0 mass distribution for recorded events in the years 2017 and 2018 passing the selection requirements for the normalisation decay $D^0 \to K^-\pi^+[\mu^+\mu^-]_{\rho^0/\omega}$ (upper row) and the reference channel, $D^0 \to K^-K^+\mu^-\mu^+$ decays (lower left) and $D^0 \to \pi^-\pi^+\mu^-\mu^+$ decays (lower right). The red line marks the D^0 mass [35]. The dotted lines mark the selection cuts on the D^0 mass.

for events above the D^0 mass is due to the momentum resolution of the tracking stations. The distribution for events with reconstructed bremsstrahlung is more symmetric but also wider, which is due to the ECAL resolution. In addition, sometimes too much energy is added either because the wrong calorimeter cluster is added or multiple particles deposited energy in the same calorimeter cluster, leading to an overestimation of the reconstructed mass. Simulated signal candidates as shown in figure 5.5 are used as a signal proxy.

Signal and background proxies are separated randomly into two disjoint equally sized subsamples. One subsample is used to train the BDT, and one to test. The following variables are used to train the BDT:

- Maximum distance of closest approach between all D^0 meson daughter particles, notated "`DO MAXDOCA`"

- Pseudorapitity η of the D^0 track, also notated "`D0 ETA`"
- Asymmetry of p_T from tracks in a cone around the D^0 track with an opening angle of 1.1 rad, defined by:

$$\texttt{CONEPTASYM} = \frac{p_T(D^0) - p_T(\text{cone})}{p_T(D^0) + p_T(\text{cone})},$$

  where $p_T(D^0)$ is the transverse momentum of the D^0 and $p_T(\text{cone})$ the transverse momentum of all non-signal parameters in a cone around the D^0 track, notated "`D0 CONEPTASYM`"
- χ^2_{IP} of the interaction points for both hadrons, $h0$ and $h1$, also notated "`h0 IPCHI2`" and "`h1 IPCHI2`"
- D^0 vertex reconstruction quality χ^2/ndof, also notated "`D0 VERTEX_CHI2NDOF`"
- Cosine of the direction angle of the D^0 meson track, also notated "`D0 DIRA`"
- Transverse momentum p_T of the D^{*+}, also notated "`Dstarplus PT`"
- p_T and p of the slow tag pion originating from D^{*+}, also notated "`Slow Pion PT`" and "`Slow Pion P`"

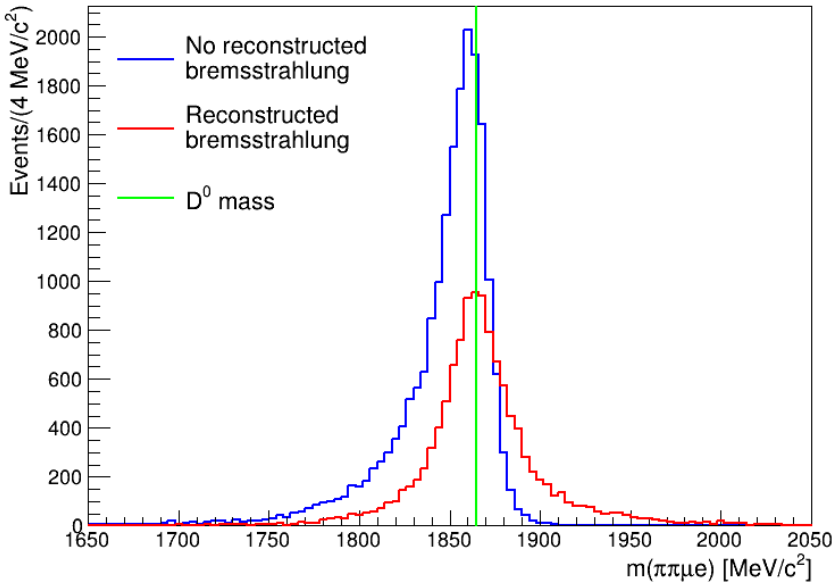


Figure 5.5: D^0 mass resolution for simulated $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ decays for decay candidates with and without recovered bremsstrahlung. The green line marks the D^0 mass. [35].

The variables are selected by their discrimination power in an iterative process. For this, the BDT is trained with a larger set of variables containing the same variables as are used in the trigger and stripping selection apart from PID variables and the reconstructed D^0 mass. Training the BDT directly on the reconstructed D^0 mass would make a statistical separation by a fit between the leftover background difficult, if not impossible, as the BDT could sculpt the reconstructed D^0 mass distribution. The variables that have no or minimal impact on the BDT classifier are then removed iteratively by retraining the BDT at each step with a smaller set of variables. A minimal set of discriminating variables, while providing the same BDT performance as the full variable set, was selected to prevent overtraining.

The `CONEPTASYM` variable has a special role. This variable sums up the p_T of all particle candidates in a cone around the D^0 track. A larger value of this so-called isolation variable hints to a more isolated track, which is favoured by signal decays. Typically, the hadronic background given mainly by pions is less isolated than signal candidates. An isolated track is also easier to reconstruct correctly, avoiding backgrounds from wrongly reconstructed decays, or ghost tracks. The `CONEPTASYM` variable requires information not related directly to the decay, which was only made available in a processable way after the data-taking period of Run 2 was concluded. It is, therefore, not included in the trigger or stripping selection. It is nevertheless added to the BDT as it provides additional information to the signal background separation.

A separate BDT is trained for each LFV decay channel ($D^0 \to K^-K^+\mu^\pm e^\mp$, $D^0 \to K^-\pi^+\mu^\pm e^\mp$ and $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$). Note that only the χ^2_{IP} variable depends directly on the daughter particles of the D^0 meson. However, the DLL selection of the kaon may affect the combinatorial background distributions. This effect is also partially observed when comparing the individual BDT input variable distributions for the background proxy. Therefore, three BDTs are trained. For the two reference decays, $D^0 \to K^-K^+\mu^-\mu^+$ and $D^0 \to \pi^-\pi^+\mu^-\mu^+$, no separate BDT is trained BDT. Instead, the same BDT as for $D^0 \to K^-K^+\mu^\pm e^\mp$ and $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ decays, respectively, is used.

In figure 5.6, as an example for $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ decays, the normalised input variable distributions for the signal and the background proxy are shown. The distributions for $D^0 \to K^-K^+\mu^\pm e^\mp$ and $D^0 \to K^-\pi^+\mu^\pm e^\mp$ decays are reported in appendix B. The variables with the highest separation power, as expected by comparing the distributions, are `D0 MAXDOCA`, `D0 VERTEX_CHI2NDOF`, `D0 CONEPTASYM` and `D0 DIRA`.

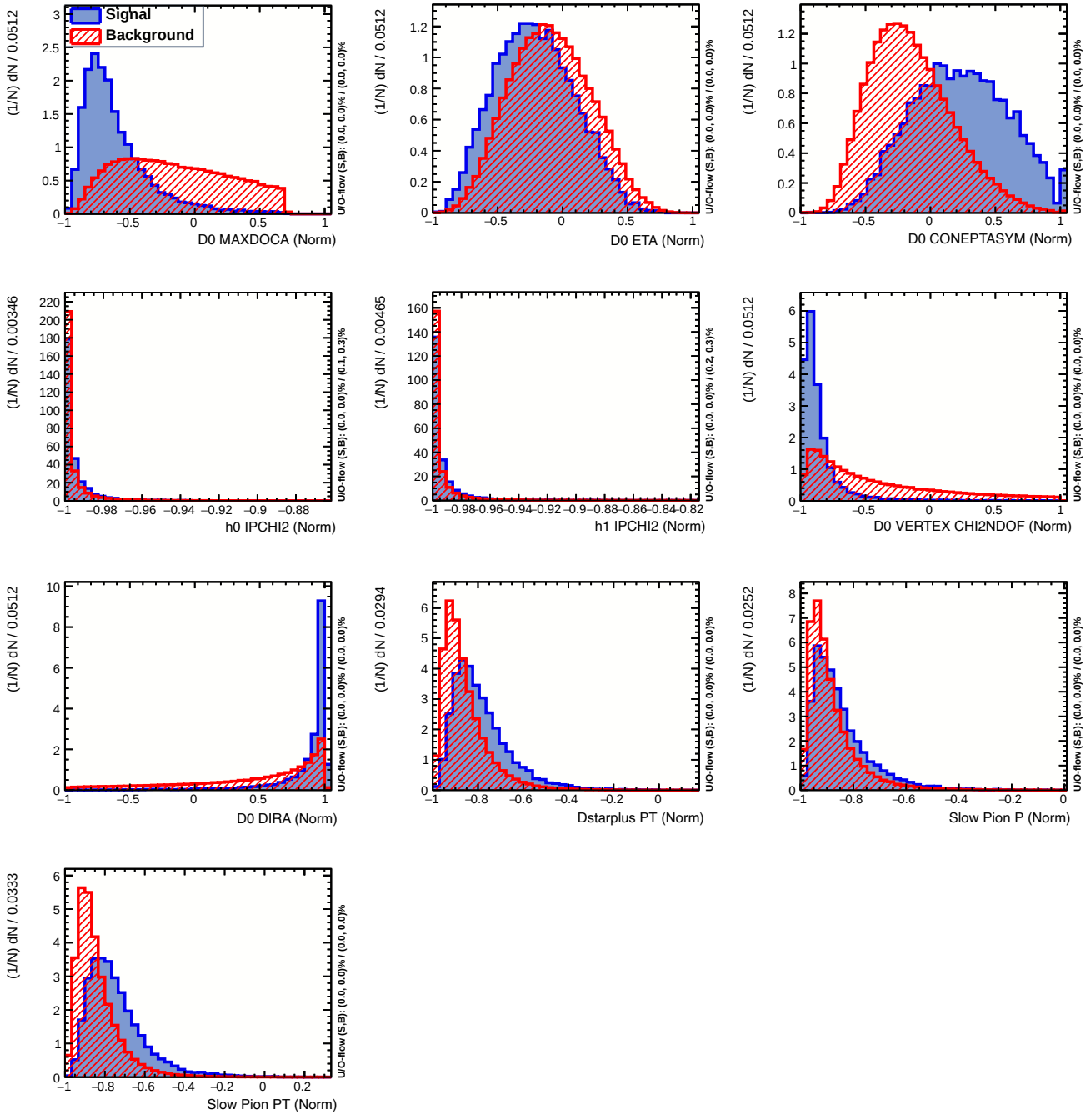


Figure 5.6: Normalised BDT input variable distributions for signal, $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ decays, (blue) and background proxy (red) overlaid.

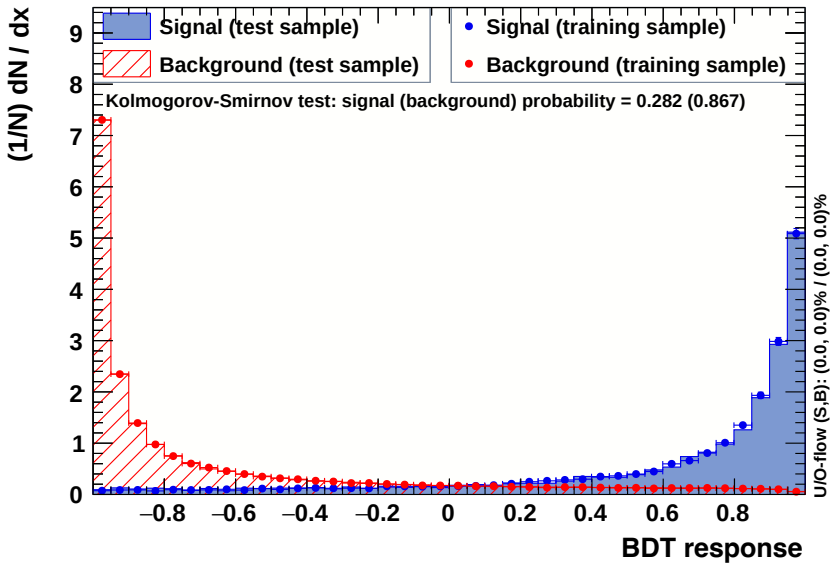


Figure 5.7: Distribution of the BDT output, the `BDT response`. Both test and training samples normalised and overlaid by each other are shown. Both distributions match, indicating no overtraining, which is also confirmed by the reported Kolmogorov-Smirnov test result [106], which gives a measure of the compatibility of the two distributions.

Due to different trigger requirements for the 2015/16 data samples, only a smaller sideband window is available. The BDT is trained using only 2017 and 2018 data, as these are the data samples with the highest statistics.

An example distribution of the resulting `BDT response` can be found in figure 5.7. To verify that the BDT is not overtrained, the `BDT response` for test and training sample are compared, and a Kolmogorov-Smirnov test [106] is performed, also shown in figure 5.7. No overtraining, which would be seen in the difference between the distribution of test and training samples, is observed for the three different decay channels. The choice of the `BDT response` value to cut on is discussed in section 5.3.3.

5.3.2 Particle identification

The second main source of background are misidentified $D \to h^-h^+\pi^-\pi^+$ decays ($h = \pi^\mp, K^\mp$) which have a branching fraction of $O(10^{-2} - 10^{-3})$. With a typical misidentification rate of 1% for pions to be misidentified as leptons, a measurable component is expected after the final selection. This fraction can be reduced using the discriminating PID variable `ProbNN`. In addition, a fit will be performed to determine the number of

leftover misID and the number of signal candidates, which is detailed in chapter 7.

5.3.3 Optimisation

To avoid removing too many signal candidates, both selection steps are simultaneously optimised for the signal sensitivity. For this, the Punzi figure of merit [86] is used, which is introduced in section 4.4. The three optimised discriminators are the two lepton PID variables, `ProbNN` for both leptons, and `BDT response`. The Punzi figure of merit, which value is maximised, is calculated as follows:

$$FOM = \frac{\epsilon_{sig}}{\frac{3}{2} + \sqrt{N_{bkg}}} \tag{5.3.1}$$

where ϵ_{sig} is the signal efficiency and N_{bkg} is the number of expected background candidates. The number of expected background candidates is evaluated for the region where most signal candidates are expected, $m(D^0) \in [1800\,\mathrm{MeV}/c^2, 1900\,\mathrm{MeV}/c^2]$[8]. N_{bkg} is the sum of combinatorial background candidates, N_{comb}, and peaking misidentified background candidates, N_{misID}.

Signal efficiency

The signal efficiency for PID and BDT selection is estimated from simulations by counting the events before and after the BDT and lepton PID selection cuts. The PID variables are corrected using the `PIDGen` software package [92, 93]. Details of the resampling procedure are explained in section 4.3. Exemplary, the efficiencies for PID and BDT selection for $D^0 \to K^-K^+\mu^\pm e^\mp$ are shown figure 5.8. As expected, the efficiency is lower for stricter selection requirements and drops faster for stricter PID cuts on the electron compared to cuts applied to the muon. All signal efficiencies as a function of BDT and PID selection are reported in appendix C.

Expected combinatorial background yield

To avoid unblinding the signal region, the expected combinatorial background is extrapolated from the upper D^0 mass sideband ($m(D^0) > 1900\,\mathrm{MeV}/c^2$). Furthermore, to avoid any signal contribution, the Δm selection is inverted. Only candidates outside the Δm

[8] Signal candidates above $1900\,\mathrm{MeV}/c^2$ are possible due to ECAL resolution effects. These are, therefore, excluded from the background proxy for the training of the BDT. However, realistically, no signal candidates are expected above $1900\,\mathrm{MeV}/c^2$ considering current limits.

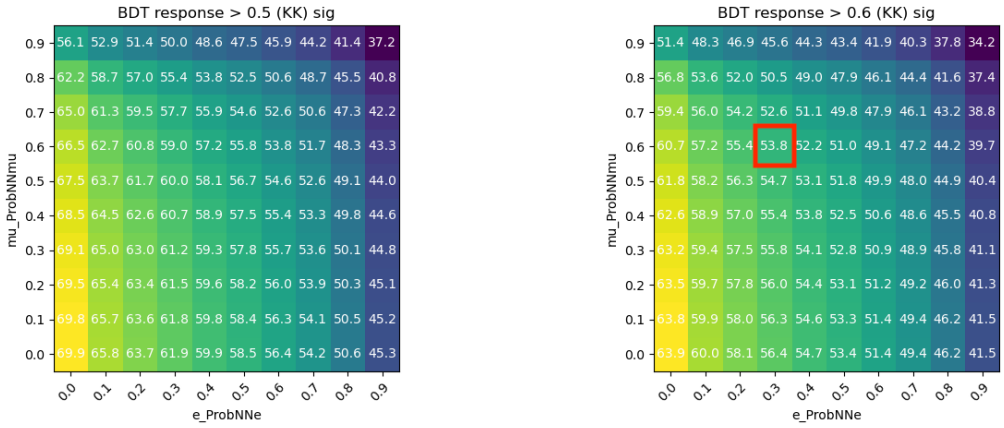


Figure 5.8: PID and BDT signal efficiency (in %) for $D^0 \to K^-K^+\mu^\pm e^\mp$ decays evaluated from simulations in bins of `ProbNNmu`, `ProbNNe` and the `BDT response`. Only two `BDT response` bins, `BDT response` > 0.5 (left) and `BDT response` > 0.6 (right), are shown for simplicity reasons. Marked in red is the efficiency value for the optimal selection values, which maximises the Punzi figure of merit.

window are selected $(\,]144.5\,\mathrm{MeV}/c^2, 146.5\,\mathrm{MeV}/c^2[\,)$, removing 99.7 % of signal candidates. As an illustration, the selected combinatorial background candidates correspond to the in the upper sideband ($m(D^0) > 1900\,\mathrm{MeV}/c^2$) in figure 5.2, page 52, and figure 5.3, page 53, removed background candidates. Due to this choice, two scaling factors are needed:

1. A factor that scales the number of background candidates in the upper $m(D^0)$ sideband to the number of background candidates in the signal window.

2. A factor which scales the background candidates in the signal window outside the Δm window to background candidates within the Δm window.

The first scaling factor is calculated with the help of a linear fit to the D^0 mass window. For this, an inverted PID selection is applied to remove all signal events. The inverted PID selection requires `ProbNNmu` < 0.1 for the muon and `ProbNNe` < 0.1 for the electron candidate such that everything except the desired leptons is selected. The mass distribution with the overlaid fit for the different decay channels is shown in figure 5.9. The factor is defined by the fraction of the integral in the upper sideband and the signal window.

The second scaling factor is defined by the ratio of background events in the signal window outside the Δm window and within the Δm window. The factor is calculated with the help of a fit to the Δm distribution with the same inverted PID selection applied. The fit, shown in figure 5.10, contains two components: a polynomial function to describe the combinatorial background (green) and a Gaussian distribution (red) to describe other

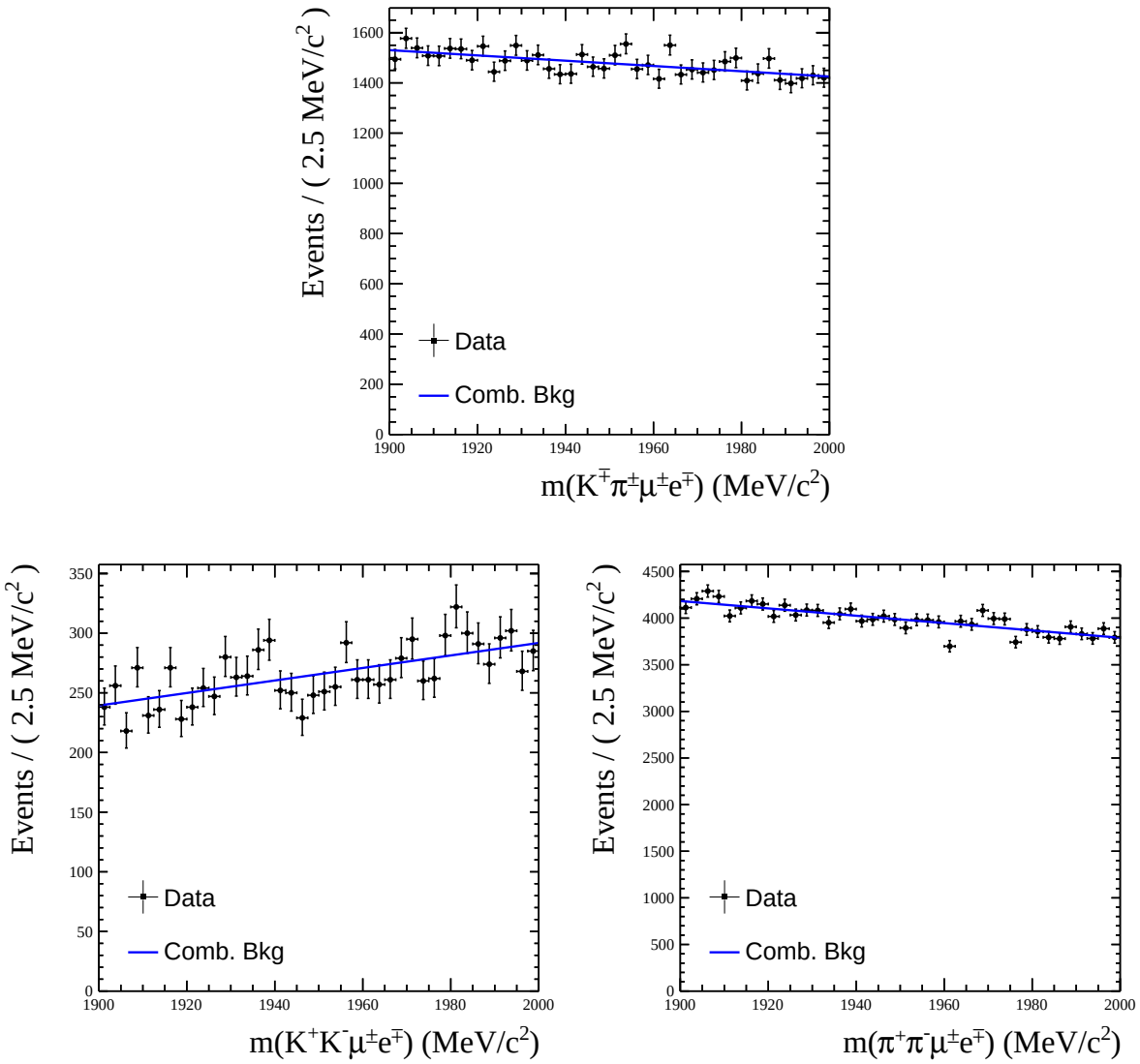


Figure 5.9: The lower left distributions show data reconstructed as $D^0 \to K^-K^+\mu^{\pm}e^{\mp}$ decays, the upper as $D^0 \to K^-\pi^+\mu^{\pm}e^{\mp}$ decays and the lower right as $D^0 \to \pi^-\pi^+\mu^{\pm}e^{\mp}$ decays with inverted PID selection.

possible D^0 decays, for example, decays of the form $D^0 \to h^{(\prime)-}h^+\pi^-\pi^+$. The ratio is then extracted by integrating the polynomial function within the in the selection applied Δm region marked in grey and the complete available Δm region. The ratio of the two integrations can then be used to correct the yield of combinatorial background candidates outside the grey-marked area to the number of combinatorial background candidates within. It is visible that the fit, figure 5.10, does not describe the mass distribution perfectly; nevertheless, for the purpose of this extrapolation, the agreement is sufficient. The calculated number of background candidates matches the in data observed number

of background candidates within their uncertainties. This is shown in section 5.3.4. Exemplary, the expected number of combinatorial background candidates in dependence of the PID and BDT selection for $D^0 \to K^-K^+\mu^\pm e^\mp$ are shown figure 5.11.

The scaling factors, which depend on the shape of the distribution of the combinatorial background, are assumed to be at least approximately independent of PID and BDT selection. The results from section 5.3.4 support this assumption. The same scaling factors are used for each PID and BDT requirement throughout the optimisation, while the to-be scaled background yield is measured in each `ProbNNmu/ProbNNe/BDT response` bin. The results of this calculation can be found in appendix C.

Expected misidentified background yield

The observed misID background in $D^0 \to K^-K^+\mu^-\mu^+$, $D^0 \to K^-\pi^+\mu^+\mu^-$ and $D^0 \to \pi^-\pi^+\mu^-\mu^+$ decays is scaled to match the number of misID background candidates in the signal mode. For this, it is necessary to account for the different selections between dimuon and LFV decay modes, which are given by a different set of PID, kinematic and topological requirements. A special role here has the more complex `ProbNN` PID classifier, as the description of this classifier is difficult in simulations. The selection difference, therefore, is evaluated with data. In the following, the `ProbNN` PID selection is therefore treated separately. A second effect that needs to be accounted for is that only the number of misID background candidates within the previously defined signal window, $m(D^0) \in [1800, 1900]$ under $D^0 \to h^{(\prime)-}h^+\mu^\pm e^\mp$ mass hypothesis, are of interest.

In the following, three scaling factors are defined: A scaling factor for the tested `ProbNN` PID selection, given by the ratio r_{misID}. A scaling factor to account for different kinematic cuts during the trigger stage[9], given by the ratio r_{sele} and a third scaling factor r_{shape}, which accounts for the different mass hypotheses in the reconstruction. The then observed yield of misID background candidates in $D^0 \to K^-K^+\mu^-\mu^+$, $D^0 \to K^-\pi^+\mu^+\mu^-$ and $D^0 \to \pi^-\pi^+\mu^-\mu^+$ decays can be scaled by the ratio r, given by

$$r = r_{misID} \times r_{sele} \times r_{shape}, \tag{5.3.2}$$

to the number of misID background events for the LFV decay modes.

[9]This factor is needed mainly due to the not recoverable dilepton object introduced for computational reasons during the HLT2 step.

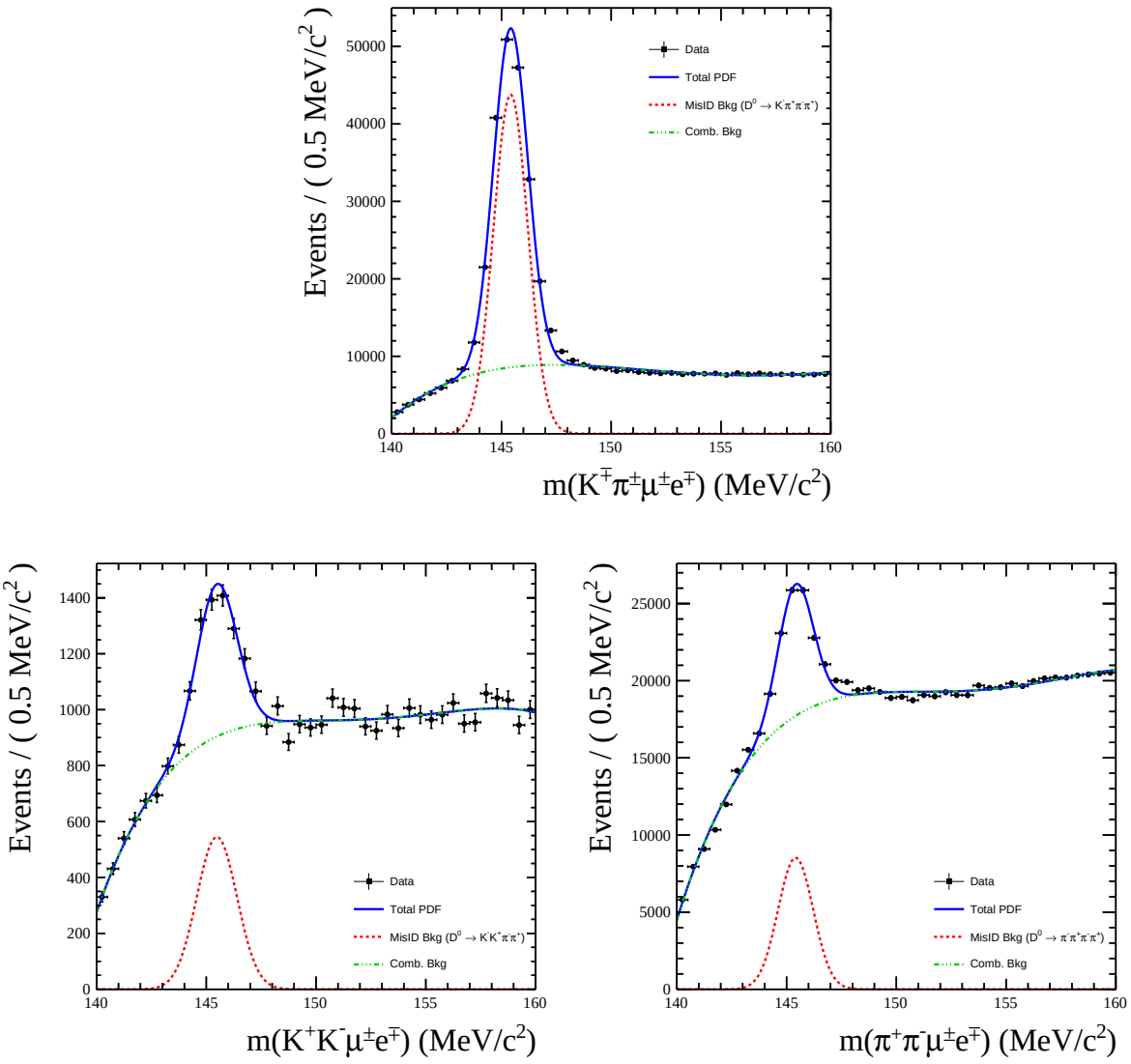


Figure 5.10: The reconstructed Δm distribution from data recorded in 2017 and 2018 reconstructed under $D^0 \to K^-K^+\mu^{\pm}e^{\mp}$ decay hypotheses (lower left), $D^0 \to K^-\pi^+\mu^{\pm}e^{\mp}$ decay hypothesis (up) and $D^0 \to \pi^-\pi^+\mu^{\pm}e^{\mp}$ decay hypotheses (lower right). The signal is blinded by applying an inverted PID selection. The reconstructed Δm distribution is overlaid with a fit projection (blue) of a polynomial function (green) and a Gaussian distribution (red). The Gaussian distribution is used to describe possible misID candidates, while the polynomial background is used to describe the combinatorial background.

The ratio r_{misID} is defined by:

$$r_{misID} = \frac{p(\pi|e)}{p(\pi|\mu)}, \tag{5.3.3}$$

where $p(\pi|e)(p(\pi|\mu))$ is the misidentification rate of a pion into an electron (muon). The misidentification rate is evaluated using the PIDCalib software package [92], a

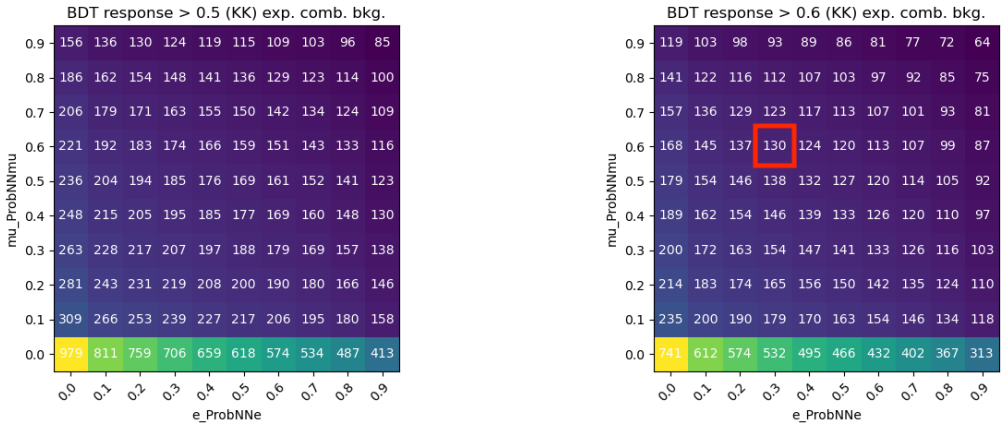


Figure 5.11: Expected number of combinatorial background candidates for $D^0 \to K^-K^+\mu^\pm e^\mp$ decay candidates evaluated from simulations in bins of `ProbNNmu`, `ProbNNe` and the `BDT response`. Only two `BDT response` bins, `BDT response` > 0.5 (left) and `BDT response` > 0.6 (right), are shown for simplicity reasons. Marked in red is the expected number of combinatorial background candidates for the optimal selection values, which maximises the Punzi figure of merit.

tool widely employed within the LHCb Collaboration. The misidentification rate is calculated by measuring the efficiency of the tested `ProbNN` requirement on a statistically large calibration sample, $D^0 \to K^-\pi^+$ decays. The large number of events and high purity of the sample allow for a precise efficiency estimate. Secondly, it is assumed that the PID variables can be fully parameterised by a set of known variables, here the momenta p, the pseudorapidity η, and the number of tracks in the event, `nTracks`. The sample is then partitioned into subsamples in which the PID variable distribution does not vary significantly. In each bin, the number of pions passing the tested `ProbNN` requirement is evaluated and compared to the original number of pions in the respective bin, allowing the evaluation of the misidentification rate. The misidentification rate, in bins of η, p, and `nTracks`, is then matched to the pions in $D^0 \to h^{(\prime)-}h^+\pi^-\pi^+$ decays. This ratio is calculated in each `BDT response` bin for a fixed `ProbNN` selection on the to-be scaled dimuon mode. For this, the central bin requiring `ProbNNmu` > 0.5 for both muon candidates was chosen. The advantage of this choice is that with a single evaluation via a fit of the misID background in this PID bin, the misID background in all other PID bins for the same `BDT response` bin can be evaluated.

The ratio r_{sele} is calculated by processing the same simulation sample with the different stripping and trigger lines for signal and normalisation channels, leading to:

$$r_{sele} = \frac{N(D^0 \to h^{(\prime)-}h^+\pi^-\pi^+ | D^0 \to h^{(\prime)-}h^+\mu^\pm e^\mp)}{N(D^0 \to h^{(\prime)-}h^+\pi^-\pi^+ | D^0 \to h^{(\prime)-}h^+\mu^-\mu^+)} \tag{5.3.4}$$

$N(X|Y)$ stands for the number of events of decay X passing the selection for decay Y. By multiplying the observed number of misID candidates in $D^0 \to h^{(\prime)-}h^+\mu^-\mu^+$ decays with r_{sele} and r_{misID}, the expected number of misID candidates is calculated.

The ratio r_{shape} is the ratio of the integral of the two PDFs describing the misID backgrounds obtained by fitting the D^0 mass shape using simulation data under $D^0 \to h^{(\prime)-}h^+\mu^\pm e^\mp$ and $D^0 \to h^-h^+\mu^-\mu^+$ mass hypotheses within the signal window. Both PDFs are normalised over the complete reconstructed mass range.

After calculating the correction factors for each Punzi figure of merit bin relative to the central, the number of misID background candidates in the central bin is evaluated via a fit to data. The number of misID background candidates is then scaled by r to the number of expected misID background candidates for the LFV decay modes. The fitted distribution for the optimal `BDT response` in the central PID bin, which has a `ProbNN` selection of larger than 0.5 for both leptons, is shown in figure 5.12. It is also visible that the applied BDT selection already suppresses the combinatorial background significantly compared to the distributions shown in figure 5.4. Here, it is observed that $D^0 \to \pi^-\pi^+\mu^-\mu^+$ candidates contain a larger contribution of misID background candidates than $D^0 \to K^-\pi^+\mu^+\mu^-$ or $D^0 \to K^-K^+\mu^-\mu^+$ candidates do. The same is true for the expected misID background candidates in $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ candidates compared to $D^0 \to K^-\pi^+\mu^\pm e^\mp$ or $D^0 \to K^-K^+\mu^\pm e^\mp$ candidates. This kind of hierarchy is also observed for the optimal selection cut values of the `ProbNN` variable, with $D^0 \to K^-K^+\mu^-\mu^+$ candidates having relatively loose PID requirements, $D^0 \to K^-\pi^+\mu^\pm e^\mp$ candidates having more strict PID requirements, and $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ candidates having the strictest PID requirements. Details of the applied fit model can be found in chapter 7. Exemplary, the expected number of misID background candiadates in dependence of the PID and BDT selection for $D^0 \to K^-K^+\mu^\pm e^\mp$ are shown figure 5.13.

Punzi figure of merit

Combining the expected signal sensitivity and expected background under the signal window by using equation 5.3.1 yields the final figure of merit, which is maximised. The Punzi figure of merit for $D^0 \to K^-K^+\mu^\pm e^\mp$, $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ and $D^0 \to K^-\pi^+\mu^\pm e^\mp$ decay candidates are also reported in figure 5.14. The corresponding selection to achieve the maximum value in the Punzi figure of merit can be found in table 5.2. If multiple

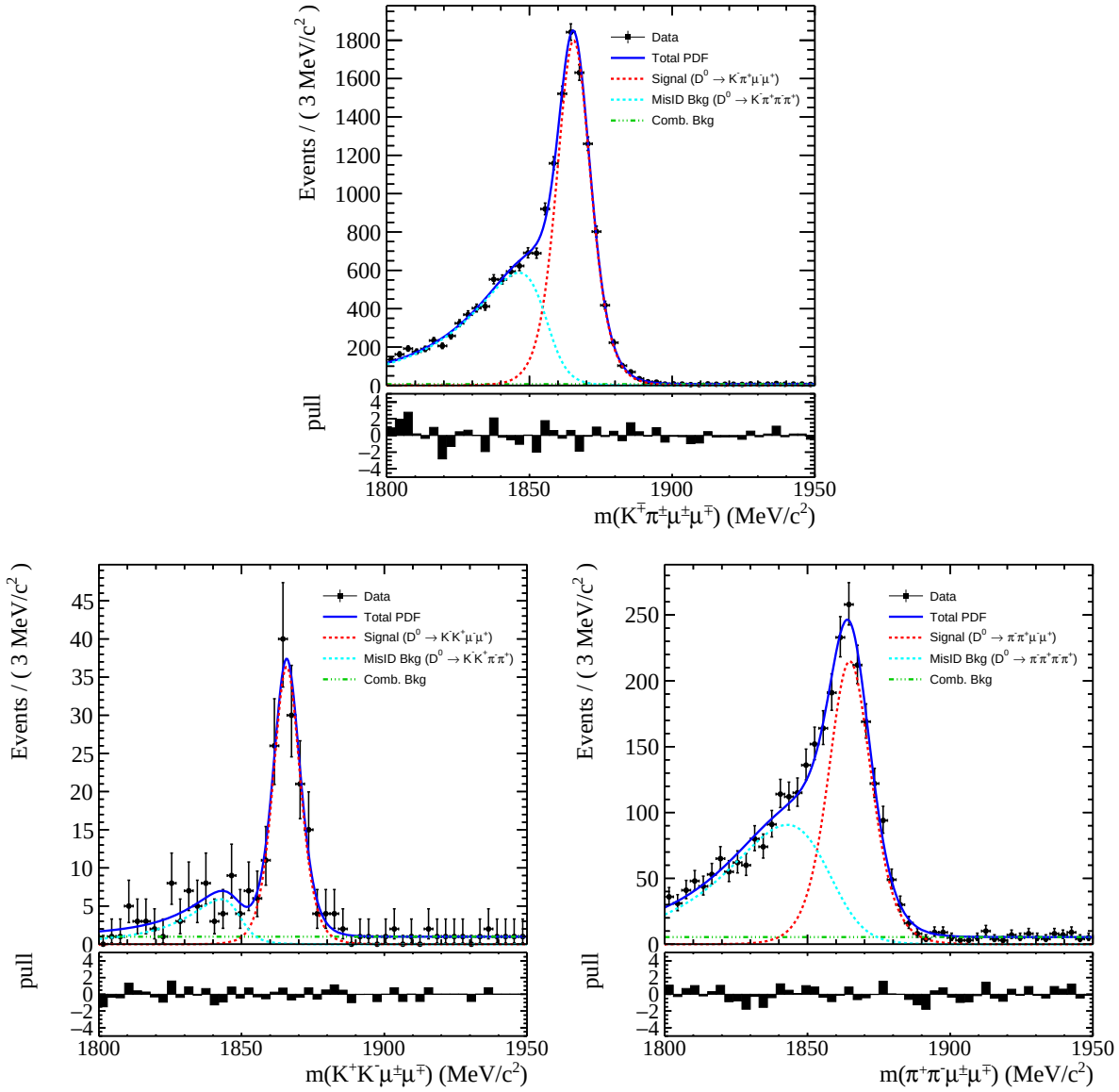


Figure 5.12: The invariant reconstructed D^0 mass distribution for $D^0 \to K^-K^+\mu^-\mu^+$ (lower left), $D^0 \to K^-\pi^+[\mu^+\mu^-]_{\rho^0/\omega}$ (up) and $D^0 \to \pi^-\pi^+\mu^-\mu^+$ (lower right) decays. The distribution is overlaid by the fit projection (blue) used to calculate the number of misID background candidates (cyan). Below the reconstructed mass, the pull distribution, showing the difference between data and fitted distribution normalised by its error for each bin, is reported.

bins have the same maximum value, the bin with the most background candidates is chosen[10]. This is done for technical reasons, as the fit has a higher chance to converge with more events. Especially for $D^0 \to K^-K^+\mu^\pm e^\mp$ candidates, stricter cuts would potentially remove all combinatorial background events, leading to an overparameterisation of the fit.

[10]Technically, the bin with the highest signal plus background statistic is chosen, but the expected number of LFV signal candidates is always zero.

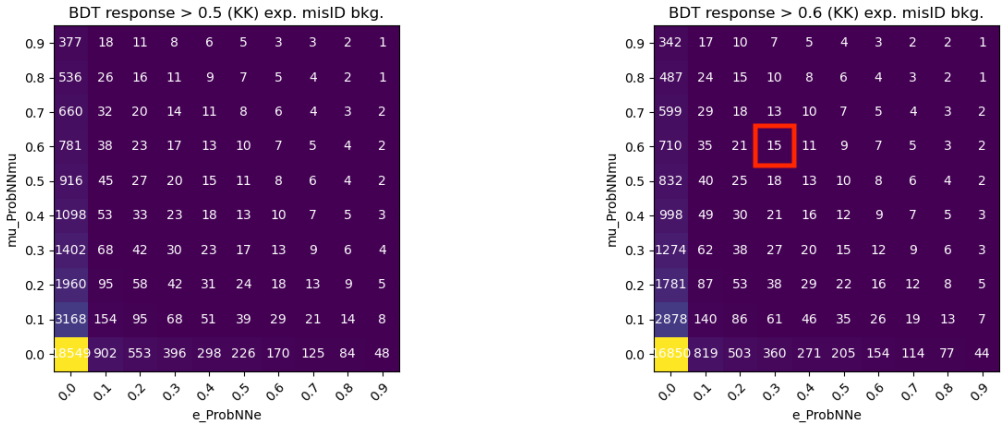


Figure 5.13: Expected number of misID background events for $D^0 \to K^-K^+\mu^\pm e^\mp$ decay candidates evaluated from simulations in bins of `ProbNNmu`, `ProbNNe` and the `BDT response`. Only two `BDT response` bins, `BDT response` > 0.5 (left) and `BDT response` > 0.6 (right), are shown for simplicity reasons. Marked in red is the expected number of misID background events for the optimal selection values, which maximises the Punzi figure of merit.

This would require a different statistical treatment of this decay channel. However, given the similar values for the Punzi figure of merit, representing the signal sensitivity, this would just complicate the measurement without improving the result.

Variable	$K^-K^+\mu^\pm e^\mp$	$K^-\pi^+\mu^\pm e^\mp$	$\pi^-\pi^+\mu^\pm e^\mp$
`ProbNNmu`	0.60	0.90	0.90
`ProbNNe`	0.30	0.90	0.60
`BDT response`	0.60	0.60	0.80

Table 5.2: Optimal cut values for the three different LFV signal decays found by maximising the Punzi figure of merit.

Similar to the in figure 5.4, page 56, reported distributions, many combinatorial background events are expected in the data sample processed by the $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ selection. In the data samples passing the selection for $D^0 \to K^-\pi^+\mu^\pm e^\mp$ and $D^0 \to K^-K^+\mu^\pm e^\mp$ decays, combinatorial background events are expected to be suppressed because of the DLL requirement on the kaon candidate. Data processed by the $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ and $D^0 \to K^-\pi^+\mu^\pm e^\mp$ selection still contain many misidentified decay candidates compared to candidates passing the $D^0 \to K^-K^+\mu^\pm e^\mp$ selection. The same is expected for the LFV decay modes with the respectively matching hadron composition. This observation also matches the optimal found cut values. The loosest cut values are found for the $D^0 \to K^-K^+\mu^\pm e^\mp$ selection. While stricter values for the

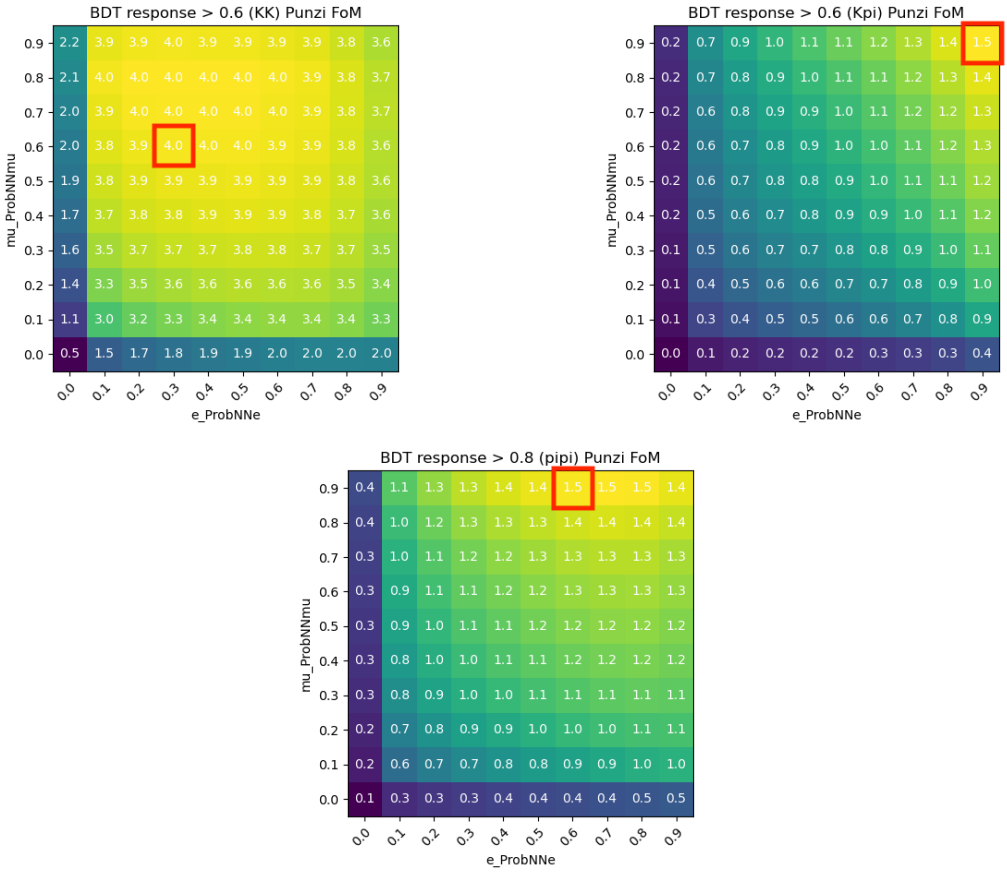


Figure 5.14: Punzi figure of merit for $D^0 \to K^-\pi^+\mu^\pm e^\mp$ (up), $D^0 \to K^-K^+\mu^\pm e^\mp$ (down left) and $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ (down right) in bins of `ProbNNmu` and `ProbNNe` in the optimal `BDT response` bin. Left shows the Punzi figure of merit for $D^0 \to K^-K^+\mu^\pm e^\mp$, middle $D^0 \to K^-\pi^+\mu^\pm e^\mp$ and right $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$.

$D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ selection are found, including a stricter value for the `BDT response`, which is used to suppress combinatorial background. For the by misidentified background candidates dominated $D^0 \to K^-\pi^+\mu^\pm e^\mp$ selected data set, strict PID cut values are found.

For completeness, all plots showing signal efficiency, expected combinatorial background, and expected misidentified background in each `BDT response` bin for the three different LFV signal channels are reported in appendix C.

5.3.4 Crosscheck: Optimisation

The optimisation procedure is cross-checked and validated by comparing the extrapolated background yields for misID and combinatorial background obtained during the optimisation with the misID and combinatorial background observed in recorded data

	$D^0 \to K^- K^+ \mu^\pm e^\mp$	$D^0 \to K^- \pi^+ \mu^\pm e^\mp$	$D^0 \to \pi^- \pi^+ \mu^\pm e^\mp$
$N_{\text{Comb;data}}$	195 ± 18	638 ± 33	813 ± 35
$N_{\text{Comb;exp}}$	214 ± 5	671 ± 11	835 ± 12
$N_{\text{misID;data}}$	30 ± 11	254 ± 22	269 ± 22
$N_{\text{misID;exp}}$	17 ± 3	247 ± 4	286 ± 9

Table 5.3: Comparison between expected, notated by exp, and measured, notated by data, background yields for the LFV signal modes.

for the LFV decay channels. For this, a fit similar to the one in figure 5.12 has to be performed. However, as the signal window is the reconstructed D^0 mass is blind, a blinded fit has to be performed, which is detailed in chapter 7[11].

The observed and calculated expected events for combinatorial and misID backgrounds are summarised in table 5.3. The quoted errors for the observed backgrounds are of a statistical nature and are directly extracted during the maximum likelihood fit. The uncertainties associated with the expected yields consist of statistical errors derived from the fits and are subsequently scaled by the scaling factors. It is important to note that these uncertainties do not account for uncertainties introduced by the scaling factors. Consequently, the resulting uncertainties on the extrapolated values are underestimated. Nevertheless, the observed and calculated expected events match within their uncertainties, validating the previously introduced calculation procedure.

5.4 Concluding remarks

The LFV signal samples are further split into candidates with reconstructed bremsstrahlung and no reconstructed bremsstrahlung. In addition, the data is divided into two data-taking periods, 2015/2016 and 2017/2018. Different efficiencies are expected for the two data-taking periods because of changes in the electron reconstruction and the added D^{*+} meson at HLT2 in 2017/2018. In addition, this separation allows to check the internal consistency of this measurement. The same TMVA classifier and `BDT response`

[11]Technically, the reconstructed D^0 mass range studied in chapter 7 and the for the optimisation defined signal region do not match. The fit region, given by $m(D^0) \in \left[1800\,\text{MeV}/c^2, 1920\,\text{MeV}/c^2\right]$, also includes events from the upper sideband $m(D^0) > 1900\,\text{MeV}/c^2$. This was done so that the background is better described by the fit. Therefore, are all values recalculated for this comparison in the larger fit region of the reconstructed D^0 mass.

requirement for signal and normalisation decay are enforced to reduce the impact of possible systematic effects introduced in the optimisation. Consequently, the same selection to its normalisation is applied for each signal decay. This is also true for the `ProbNNmu` threshold value, which is randomly assigned to one of the muon candidates. For simplicity reasons, the other muon candidate has the `ProbNNe` threshold value applied to its `ProbNNmu` classifier. The recorded normalisation samples are then indexed by the signal channel associated with its selection. To keep track, six data categories for each measured branching ratio are introduced. Four data categories for the LFV decay modes, denoted (2015/2016 `noBrem`, 2015/2016 `Brem`, 2017/2018 `noBrem` and 2017/2018 `Brem`), and two for the normalisation mode, $D^0 \to K^-\pi^+[\mu^+\mu^-]_{\rho^0/\omega}$, denoted (`norm` 2015/2016, `norm` 2017/2018), are defined.

The same selection as applied to $D^0 \to K^-K^+\mu^\pm e^\mp$ and $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ decays is applied to the reference decays $D^0 \to K^-K^+\mu^-\mu^+$ and $D^0 \to \pi^-\pi^+\mu^-\mu^+$, respectively. The electron PID requirements are replaced by muon PID requirements. At the stripping level, $\text{DLL}_e > -2$ is replaced by `isMuon` for the second lepton. The `ProbNNe` PID variable is adjusted to the muon case, `ProbNNmu`. The main purpose for measuring the reference channels is to validate the data pipeline. Therefore, the same `ProbNN` values as for $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ and $D^0 \to K^-K^+\mu^\pm e^\mp$ decays are enforced instead of adjusting the optimisation procedure. Also, the same BDT as for the $D^0 \to K^-K^+\mu^\pm e^\mp$ and $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ decay modes is used for the reference modes, $D^0 \to K^-K^+\mu^-\mu^+$ and $D^0 \to \pi^-\pi^+\mu^-\mu^+$. Due to the blinding of the LFV decay modes the reference modes provide an important possibility to crosscheck the measurement procedure. However, to properly optimise the selection, the expected background sources need to be independent of the to-be-studied data sample. Using directly the data sample for $D^0 \to K^-K^+\mu^-\mu^+$ and $D^0 \to \pi^-\pi^+\mu^-\mu^+$ decays to optimise the selection of the same data sample would introduce a sensitivity to statistical fluctuations of the sample, which is not desired. Instead of six categories, the data samples associated with the individual branching ratio measurements are categorised into four categories. Therefore, using the same approach to calculate the number of misID background events would not be possible. To be able to validate and crosscheck the data pipeline, no separate optimisation is performed[12].

[12]The BDT selection should, in principle, be equally good for the reference modes. The largest difference is introduced by the DLL decision of the electron candidate. The PID selection cuts, especially the one evaluated for the electron candidate in the LFV decay modes, are probably too strong given that the muon `ProbNN` variable performs better in general. This can also be observed when comparing signal efficiency and the expected number of misID background events simultaneously. Just to give an example, picking the bin [`mu_ProbNNmu` > 0.8, `e_ProbNNe` > 0.0, `BDT response` > 0.5], `e_ProbNNe` > 0.0 corresponds to no

This is also the reason why it is not advisable to optimise the selection using directly the in data observed background yield. Also, here, a different selection for the two different reference decay modes, $D^0 \to K^-K^+\mu^\pm e^\mp$ and $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$, is applied to the normalisation mode, $D^0 \to K^-\pi^+[\mu^+\mu^-]_{\rho^0/\omega}$. No bremsstrahlung separation is needed. Here four data categories are defined, two for the dimuon signal decay modes, $D^0 \to K^-K^+\mu^-\mu^+$ and $D^0 \to \pi^-\pi^+\mu^-\mu^+$, denoted (2015/2016 `ref` and 2017/2018 `ref`), and two for the normalisation mode, $D^0 \to K^-\pi^+[\mu^+\mu^-]_{\rho^0/\omega}$, denoted (`norm` 2015/2016, `norm` 2017/2018).

The hadronic simulation samples are processed the same way as the signal lines, apart from the lepton PID requirements. The DLL requirement is removed altogether, while `ProbNNmu` > 0.5 and `isMuon` is enforced for one misidentified lepton candidate randomly.

`e_ProbNNe` selection, has a signal efficiency of 62% with 516 expected events while the same `mu_ProbNNmu` value, this time bin [`mu_ProbNNmu` > 0.0, `e_ProbNNe` > 0.8, `BDT response` > 0.5] has a signal efficiency of 51% and 84 expected background events. Though the signal efficiency is lower, significantly fewer background events are expected.

Chapter 6

Efficiency ratio estimation

This chapter details the calculation of the efficiency ratio. The structure of this chapter follows the event reconstruction procedure apart from the trigger efficiency calculation, which is performed last. The LFV signal and dimuon reference mode efficiency ratios are summarised in figure 6.6, page 90.

The efficiency ratio for the five different signal decays, which are the dimuon decays, $D^0 \to \pi^-\pi^+\mu^-\mu^+$ and $D^0 \to K^-K^+\mu^-\mu^+$, and the LFV decays, $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$, $D^0 \to K^-\pi^+\mu^\pm e^\mp$ and $D^0 \to K^-K^+\mu^\pm e^\mp$, is next to the yield ratio and normalisation mode branching ratio needed to calculate the corresponding signal branching ratio. All efficiencies for the dimuon reference mode are calculated and corrected using the same procedure as the LFV signal decays. The normalisation mode is selected according to the associated signal mode. Therefore, in the following chapter, for each selection step, if the efficiencies deviate, three efficiency values according to the applied selection are reported for the normalisation mode. The normalisation mode selection for the dimoun decays $D^0 \to \pi^-\pi^+\mu^-\mu^+$ and $D^0 \to K^-K^+\mu^-\mu^+$ is the same as for the two LFV decays $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ and $D^0 \to K^-K^+\mu^\pm e^\mp$, respectively. In a perfect case, the simulated data would mirror all detector and physics effects, which would allow a precise efficiency estimate from the simulated samples. However, this is not necessarily true. Therefore, the efficiency estimate is crosschecked and, if needed, corrected by data.

The efficiency for the signal can be split into five different parts:

$$\epsilon = \epsilon_{\text{acc}} \cdot \epsilon_{\text{rec}} \cdot \epsilon_{\text{PID}} \cdot \epsilon_{\text{BDT}} \cdot \epsilon_{\text{trig}}, \tag{6.0.1}$$

where the individual partial efficiencies correct for the detector acceptance, decay reconstruction, PID efficiency, BDT efficiency and trigger efficiency and are defined as

follows:

$$\begin{aligned}
\epsilon_{\text{acc}} &= N_{\text{in acceptance}}/N_{\text{produced}}, \\
\epsilon_{\text{rec}} &= N_{\text{reconstructed, stripped and selected}}/N_{\text{in acceptance}}, \\
\epsilon_{\text{PID}} &= N_{\text{passing PID selection}}/N_{\text{reconstructed, stripped and selected}}, \\
\epsilon_{\text{BDT}} &= N_{\text{passing BDT selection}}/N_{\text{passing PID selection}}, \\
\epsilon_{\text{trig}} &= N_{\text{passing trigger requirements}}/N_{\text{passing BDT selection}}.
\end{aligned}$$

$N_{x/y}$ is the number of events before and after the selection step, indicated by the subscripts x and y, respectively. For the calculation of the branching ratio, only the ratio between the signal efficiency and normalisation efficiency

$$r^{c,y} = \frac{\epsilon^{c,y}}{\epsilon^{\texttt{norm},y}} \tag{6.0.2}$$

is needed, with y donating the data taking periods and c the signal category, $c \in \{\texttt{noBrem}, \texttt{Brem}\}$ for the LFV decays modes, $D^0 \to K^-K^+\mu^\pm e^\mp$, $D^0 \to K^-\pi^+\mu^\pm e^\mp$ and $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$, and $c = \texttt{ref}$ for the dimuon reference modes, $D^0 \to \pi^-\pi^+\mu^-\mu^+$ and $D^0 \to K^-K^+\mu^-\mu^+$. Therefore, the ratios are quoted alongside the efficiencies in this chapter. Apart from a different trigger reconstruction between the two different data-taking periods, leading to a different expected efficiency estimate, studying these efficiencies separately is an important cross-check for the consistency of this measurement. In addition to the efficiencies reported in the following sections, the same efficiencies for each selection step and data category are tabulated and can be found in appendix D.

6.1 Acceptance

The LHCb detector covers only a limited region of the solid angle in the forward direction. The acceptance efficiency is defined by the geometrical region of the detector, which is $1.8 < \eta < 4.9$, where η is the pseudorapidity. The acceptance efficiencies are estimated during the production of the simulation data set by GEANT4 [81]. The tables are produced centrally and are made available to the complete collaboration [107]. The acceptance efficiency covers the geometrical region of the detector and is independent of the polarity or data-taking period. One combined value for all data-taking periods and data categories is used for the acceptance efficiency and ratio.

The centrally produced tables only provide the phase-space integrated efficiencies. An additional simulation sample is produced to calculate the acceptance efficiency for the

normalisation mode, where the measurement is limited to a narrow dimuon mass range. The acceptance efficiency for the normalisation mode is $\epsilon_{acc}^{\texttt{norm}} = 19.79\% \pm 0.30\%$. The uncertainty quoted is due to the simulation sample size. The signal mode efficiencies and associated ratios for the LFV and dimuon decay modes are reported in table 6.1.

	$D^0 \to K^-K^+\mu^\pm e^\mp$	$D^0 \to K^-\pi^+\mu^\pm e^\mp$	$D^0 \to \pi^-\pi^+\mu^\pm e^\mp$
$\epsilon_{acc}\,[\%]$	$19.93\,\% \pm 0.04\,\%$	$18.87\,\% \pm 0.04\,\%$	$18.00\,\% \pm 0.04\,\%$
r_{acc}	1.007 ± 0.015	0.954 ± 0.015	0.909 ± 0.014

	$D^0 \to K^-K^+\mu^-\mu^+$	$D^0 \to \pi^-\pi^+\mu^-\mu^+$
$\epsilon_{acc}\,[\%]$	$20.69\,\% \pm 0.09\,\%$	$18.31\,\% \pm 0.08\,\%$
r_{acc}	1.045 ± 0.017	0.925 ± 0.015

Table 6.1: The efficiency ratio $r_{acc} = \epsilon_{acc}/\epsilon_{acc}^{\texttt{norm}}$ for the detector acceptance efficiency ϵ_{acc}. The uncertainties quoted are due to the limited simulation sample size.

The kinematics of the daughter particles, especially the opening angle between the individual daughter particles, depend on the available phase space, leading to a ratio that deviates significantly from one. The same effect is observed when comparing dimuon, $D^0 \to K^-K^+\mu^-\mu^+$ and $D^0 \to \pi^-\pi^+\mu^-\mu^+$, and LFV decay modes, $D^0 \to K^-K^+\mu^\pm e^\mp$ and $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$, respectively. Due to the heavier muon, compared to the electron, a larger opening angle between the daughter particles is expected, resulting in fewer particles close to the beam pipe and outside the acceptance ($\eta > 4.9$). The same argumentation chain also applies to the different hadron compositions of the final state particles.

6.2 Decay reconstruction

The efficiencies of the kinematic and vertex quality requirements in the reconstruction and selection are calculated using simulated data. To correct discrepancies between data and simulations introduced during the tracking by the tracking algorithm, dedicated tracking correction factors exist.

These correction factors are calculated using a so-called *tag and probe* method. For the *tag and probe* probe method a *tag* particle is selected to remove background contaminations and an associated *probe* particle is used to determine the efficiency in bins

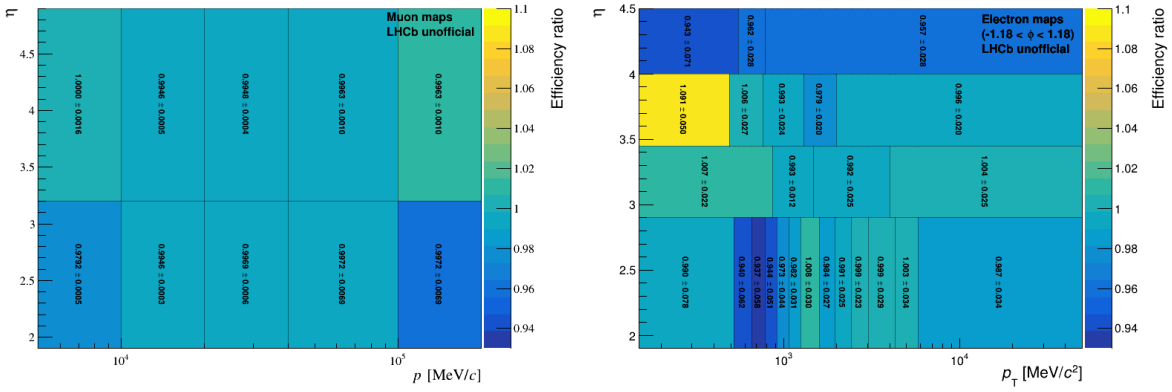


Figure 6.1: Muon (left) and electron(right) tracking data-simulation correction map, exemplary for 2018, in kinematic variables of the muon and electron, respectively. The values for the electron correction map are taken from the R_X analysis [13, 14]. The values for the muon correction map are provided centrally to the LHCb collaboration.

of the pseudorapitity η and the momentum p of the *probe* particle[1]. A high purity is advantageous for this calculation. Therefore, a high statistics sample of $J/\psi \to \mu^+\mu^-$ decays, with a tight selection on the *probe* particle, is used. In this case, one muon is used as *tag* particle, the other as *probe* particle. With this method, the tracking efficiency can be calculated from data. The resulting efficiency is then compared to the from simulations obtained efficiency. Comparing both efficiencies then allows the calculation of data-simulation correction maps in bins of η and p. An example of such a map for the muon can be found in figure 6.1. A value of one corresponds to a matching efficiency between data and simulation in the corresponding bin. The same procedure is repeated for the electron by using $J/\psi \to e^+e^-$ decays from $B^+ \to J/\psi K^+$ decays. Here in addition to η and p_T the correction factors are also binned in ϕ, which is the rotation angle around the beam pipe. The muon maps are centrally provided to the LHCb collaboration, while the electron correction maps are taken from the R_X analysis [13, 14]. The maps are then applied to the muon and electron candidates. Exemplary in figure 6.1 one such data-correction map for the muon and electron is reported. The complete set of data-simulation correction values can be found in appendix E. The irregular binning for the electron is done to ensure a minimum number of events in each bin. Most values are close to unity, showing a general good agreement between data and simulations.

For the hadrons, no such data-simulation correction maps exist. Hadrons, in addition,

[1]Alternatively to p (or η) also the transverse momentum p_T defined relative to the beam pipe can be used. This will also be the case in the following for the electron.

also react with the detector material. Those hadrons are partially lost for the tracking algorithm, which affects the tracking efficiency. Therefore, two effects need to be considered: the track reconstruction and material interactions. The first effect, the track reconstruction, is addressed by applying the muons maps to the hadrons, as both muon and hadron tracks are charged tracks[2]. The second effect, lost tracks due to matrial interactions, are already included in simulations. However due to the uncertainties of the material budget of the detector, a systematic uncertainty will be assigned.

The ratios and efficiency ratio in the data categories are reported in table 6.2. The separation into the different bremsstrahlung categories is factored into the efficiency. The main difference between normalisation and LFV signal modes arises due to the lower electron reconstruction efficiency [75]. This is also observed when comparing the efficiencies for $D^0 \to K^-K^+\mu^-\mu^+$ and $D^0 \to \pi^-\pi^+\mu^-\mu^+$ decays with the efficiency of $D^0 \to K^-K^+\mu^\pm e^\mp$ and $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ decays respectively. The second expected and observed difference in the efficiency is due to the smaller available phase space when comparing decays with heavier and lighter particles in the final state, as observed when, for example, comparing $D^0 \to K^-K^+\mu^-\mu^+$ and $D^0 \to \pi^-\pi^+\mu^-\mu^+$ decays.

6.3 Particle Identification and BDT selection

The more complicated simulated `ProbNN` PID variables are transformed using data. For the technical details, see section 4.3. For this, three different high-statistic calibration samples are used:

- $J/\psi \to \mu^+\mu^-$ decays for the determination of the PID efficiency for muons
- $J/\psi \to e^-e^+$ decays from $B^+ \to J/\psi K^+$ decays for the PID efficiency for electrons
- $D^0 \to K^-\pi^+$decays from $D^{*0} \to D^0\pi^+$ decays for the PID efficiency for pions and kaons

The corrected efficiencies and ratios are reported in table 6.3. The impact of the transformation of the muon PID variable is negligible. The efficiency is affected at a sub-percent level, leading to only a sub-percent effect on the efficiency ratio. However, the electron efficiency is corrected by $\sim 2\%$, the pion efficiencies by up to $\sim 7\%$ and the kaon efficiency by about $\sim 1\%$. Due to the different final state particles in the efficiency ratio, this leads to a non-neglectable correction.

[2]Without produced bremsstrahlung, which would additionally effect the tracking algorithm.

		$D^0 \to K^-K^+\mu^\pm e^\mp$	$D^0 \to K^-\pi^+\mu^\pm e^\mp$	$D^0 \to \pi^-\pi^+\mu^\pm e^\mp$
$\epsilon_{rec}^{\texttt{norm},1516}$	$[10^{-4}]$		113.58 ± 0.73	
$\epsilon_{rec}^{\texttt{norm},1718}$	$[10^{-4}]$		168.3 ± 0.29	
$\epsilon_{rec}^{\texttt{noBrem},1516}$	$[10^{-4}]$	41.64 ± 0.57	46.74 ± 0.43	55.72 ± 0.47
$\epsilon_{rec}^{\texttt{noBrem},1718}$	$[10^{-4}]$	55.95 ± 0.48	62.86 ± 0.36	78.14 ± 0.4
$\epsilon_{rec}^{\texttt{Brem},1516}$	$[10^{-4}]$	19.05 ± 0.39	26.48 ± 0.32	34.49 ± 0.38
$\epsilon_{rec}^{\texttt{Brem},1718}$	$[10^{-4}]$	25.05 ± 0.32	35.13 ± 0.27	47.52 ± 0.31
$r_{rec}^{\texttt{noBrem},1516}$		0.367 ± 0.006	0.412 ± 0.005	0.491 ± 0.005
$r_{rec}^{\texttt{noBrem},1718}$		0.332 ± 0.003	0.374 ± 0.003	0.464 ± 0.003
$r_{rec}^{\texttt{Brem},1516}$		0.168 ± 0.004	0.233 ± 0.003	0.304 ± 0.004
$r_{rec}^{\texttt{Brem},1718}$		0.149 ± 0.002	0.209 ± 0.002	0.282 ± 0.002

		$D^0 \to K^-K^+\mu^-\mu^+$	$D^0 \to \pi^-\pi^+\mu^-\mu^+$
$\epsilon_{rec}^{\texttt{ref},1516}$	$[10^{-4}]$	107.70 ± 0.79	116.46 ± 0.76
$\epsilon_{rec}^{\texttt{ref},1718}$	$[10^{-4}]$	157.71 ± 0.69	173.51 ± 0.71
$r_{rec}^{\texttt{ref},1516}$		0.708 ± 0.010	1.176 ± 0.013
$r_{rec}^{\texttt{ref},1718}$		0.703 ± 0.005	1.189 ± 0.007

Table 6.2: Decay reconstruction efficiency and efficiency ratio for the different signal modes split by the different data categories and years. The efficiencies for the normalisation mode is the same for all five signal decays.

It is possible, with the help of the previously applied PID selection, to compare the `BDT response` between data and simulations. This is done by applying a fit to the reconstructed D^0 mass (m_{D^0}) distribution and calculating for each event i so-called sWeights using the ${}_s\mathcal{P}lot$ formalism [108]. The weights are defined by:

$$\mathcal{P}_n(m_{D^0}(i)) = \frac{\sum_{j=1}^{N_S} \boldsymbol{V}_{nj} f_j(m_{D^0}(i))}{\sum_{k=1}^{N_S} N_k f_k(m_{D^0}(i))}, \tag{6.3.1}$$

where $\boldsymbol{V}_{ij}$ is the covariance matrix, $m_{D^0}(i)$ gives the m_{D^0} for event i, N_S is the number of event categories, signal (sig) and background (bkg), and $n \in \{\text{sig}, \text{bkg}\}$ denotes the category for which the weight is calculated. The sWeights effectively allow the subtraction of background candidates from the studied data distributions. For this, the studied variable distribution has to be independent of the distribution used to calculate the sWeights. The normalised `BDT response` for s-weighted data and simulations for the dimuon modes are reported in figure 6.2. Additionally, the residual distribution normalised

		$D^0 \to K^-K^+\mu^\pm e^\mp$	$D^0 \to K^-\pi^+\mu^\pm e^\mp$	$D^0 \to \pi^-\pi^+\mu^\pm e^\mp$
$\epsilon_{PID}^{\texttt{norm},1516}$	[%]	71.38 ± 0.19	35.21 ± 0.20	47.78 ± 0.20
$\epsilon_{PID}^{\texttt{norm},1718}$	[%]	72.65 ± 0.16	37.06 ± 0.18	49.69 ± 0.18
$\epsilon_{PID}^{\texttt{noBrem},1516}$	[%]	62.27 ± 0.63	35.65 ± 0.41	46.41 ± 0.40
$\epsilon_{PID}^{\texttt{noBrem},1718}$	[%]	59.66 ± 0.42	33.25 ± 0.26	45.29 ± 0.26
$\epsilon_{PID}^{\texttt{Brem},1516}$	[%]	64.09 ± 0.95	38.93 ± 0.56	48.89 ± 0.52
$\epsilon_{PID}^{\texttt{Brem},1718}$	[%]	60.61 ± 0.64	35.28 ± 0.36	47.48 ± 0.34
$r_{PID}^{\texttt{noBrem},1516}$		0.872 ± 0.009	1.012 ± 0.013	0.971 ± 0.009
$r_{PID}^{\texttt{noBrem},1718}$		0.821 ± 0.006	0.897 ± 0.008	0.911 ± 0.006
$r_{PID}^{\texttt{Brem},1516}$		0.898 ± 0.014	1.105 ± 0.017	1.023 ± 0.012
$r_{PID}^{\texttt{Brem},1718}$		0.834 ± 0.009	0.952 ± 0.011	0.956 ± 0.008

		$D^0 \to K^-K^+\mu^-\mu^+$	$D^0 \to \pi^-\pi^+\mu^-\mu^+$
$\epsilon_{PID}^{\texttt{ref},1516}$	[%]	67.77 ± 0.31	51.14 ± 0.24
$\epsilon_{PID}^{\texttt{ref},1718}$	[%]	68.64 ± 0.19	54.59 ± 0.16
$r_{PID}^{\texttt{ref},1516}$		0.951 ± 0.005	1.072 ± 0.007
$r_{PID}^{\texttt{ref},1718}$		0.946 ± 0.003	1.100 ± 0.005

Table 6.3: PID efficiency and efficiency ratio for the different signal modes split by the different data categories and years. The efficiencies for the normalisation mode selection used for $D^0 \to K^-K^+\mu^-\mu^+$ ($D^0 \to \pi^-\pi^+\mu^-\mu^+$) decays and $D^0 \to K^-K^+\mu^\pm e^\mp$ ($D^0 \to \pi^-\pi^+\mu^\pm e^\mp$) decays is the same, due to the same selection applied to the normalisation mode.

by its uncertainty, the so-called pull, is reported for each bin. Because of the agreement between simulation and data no further corrections must be applied to the simulated events.

The signal is blind. Therefore, the data-simulation comparison is only done for the normalisation and reference channels. In figure 6.3, the normalised `BDT response` trained for the different signal channels for s-weighted normalisation mode data and simulations is shown. Given the good agreement in the normalisation and reference mode, reported in **??**, no further corrections must be applied to the simulated events.

The efficiency ratio and single efficiencies per category are reported in table 6.4. The efficiencies are consistent for all decay modes.

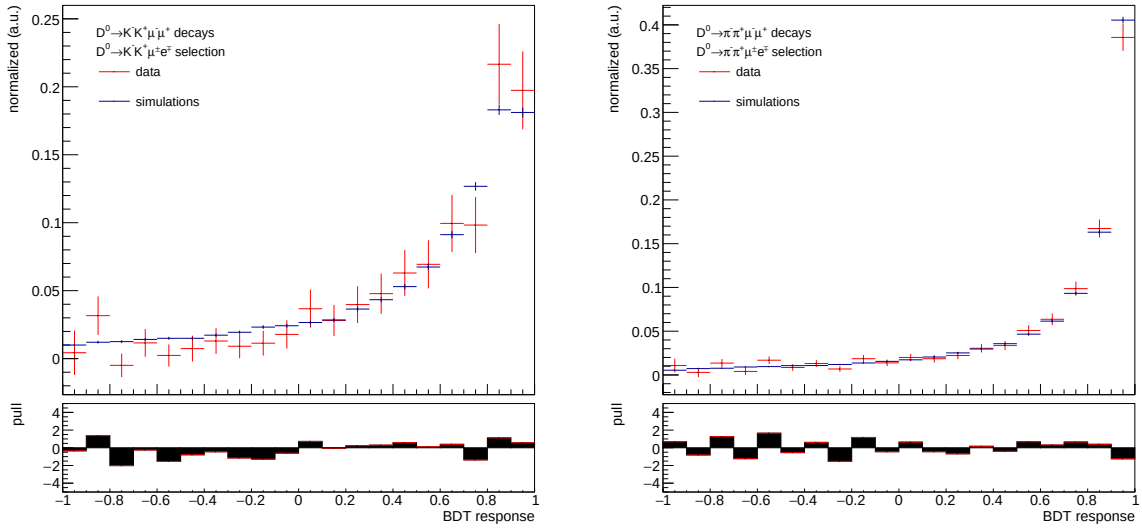


Figure 6.2: Comparison of the `BDT response` between recorded s-weighted [108] data and simulations. The BDT is trained for $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ and $D^0 \to K^-K^+\mu^\pm e^\mp$ decays and applied to $D^0 \to K^-K^+\mu^-\mu^+$ (left) and $D^0 \to \pi^-\pi^+\mu^-\mu^+$ (right) decays, respectively. Below each distribution, the pull, the difference between the two distributions normalised by its uncertainty, is reported.

6.4 Trigger

The trigger efficiency is separately estimated for the high-level trigger, ϵ_{HLT} and the low-level hardware trigger, ϵ_{L0}. The software trigger efficiencies are directly calculated from the simulated samples. The hardware trigger, which is more difficult to describe in simulations precisely, is evaluated separately in the two disjoint trigger categories, $\epsilon_{TIS\backslash TOS}$ and ϵ_{TOS}:

- ϵ_{TOS}: Trigger efficiency of events passing `L0Muon TOS`
- $\epsilon_{TIS\backslash TOS}$: Trigger efficiency of events passing `L0Muon` and `L0Electron TIS` but not `L0Muon TOS`

The naive approach would be to compute trigger efficiency by the ratio of signal events accepted by the trigger and all events within the acceptance of the detector. However, this is impossible, as the number of signal events in the acceptance is unknown for the recorded data samples. One might use, in this case, the simulated samples to estimate trigger efficiency for data, but especially for the hardware trigger, potential disparities between simulation and real data remain unaccounted for. To correct possible disparities, similarly as was done for the tracking efficiency, data-simulation weights are calculated

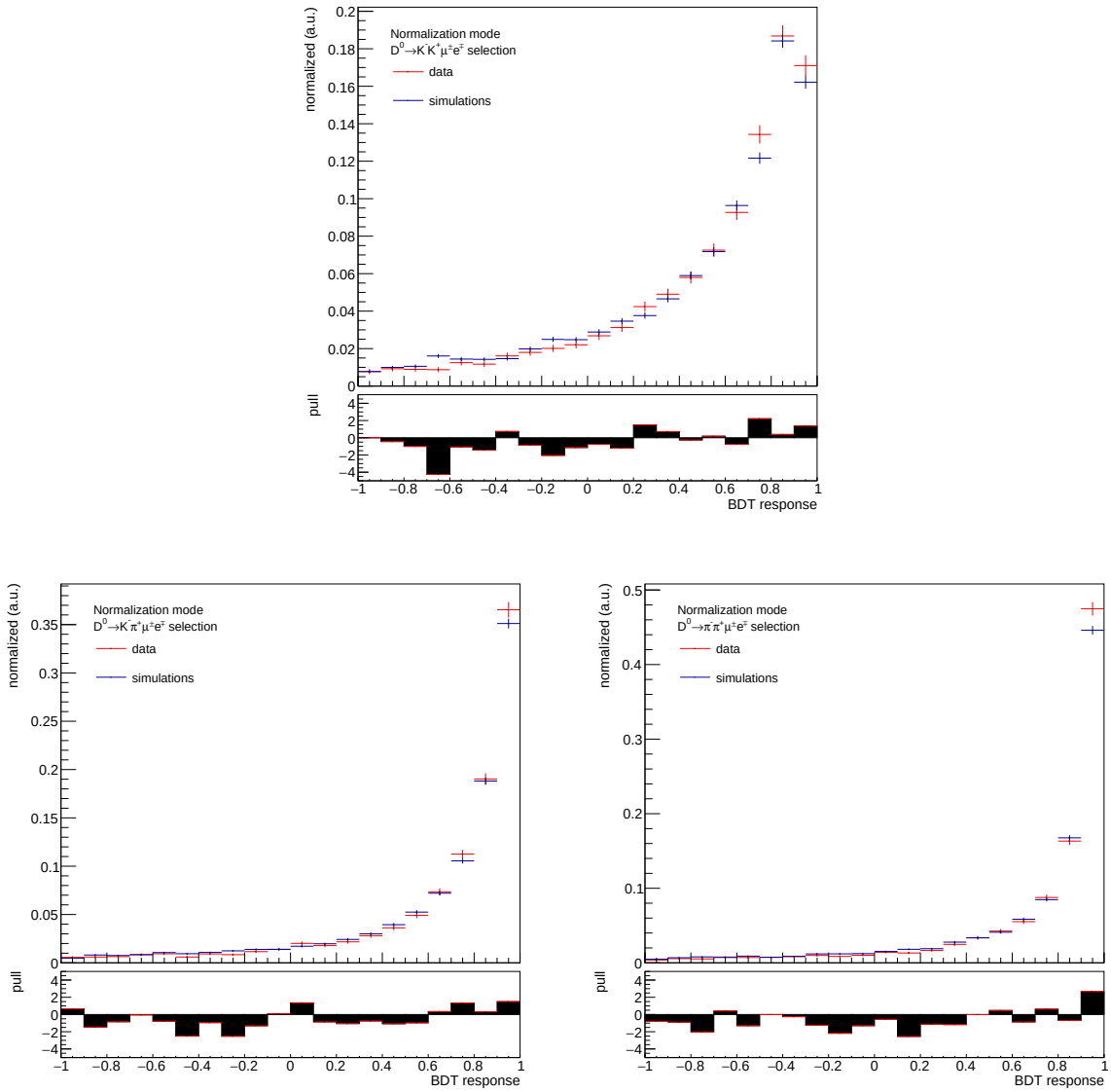


Figure 6.3: Comparison of the `BDT response` for the three trained BDTs, for $D^0 \to K^-\pi^+\mu^\pm e^\mp$ (up), $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ (lower left) and $D^0 \to K^-K^+\mu^\pm e^\mp$ (lower right) decays, applied to the normalisation decay mode, $D^0 \to K^-\pi^+[\mu^+\mu^-]_{\rho^0/\omega}$, between s-weighted [108] recorded data and simulations. Below each distribution, the pull, the difference between the two distributions normalised by its uncertainty, is reported.

with the help of a calibration sample, here $B^0 \to K^{*0}(\to K^+\pi^-)J/\psi(\to \mu^+\mu^-)$ decays. For this the trigger efficiency for the default trigger selection needs to be evaluated in data and simulations for the calibration sample. The same method with a different trigger selection, but the same calibration sample was originally developed for the measurement of R_X [13, 14].

		$D^0 \to K^-K^+\mu^\pm e^\mp$	$D^0 \to K^-\pi^+\mu^\pm e^\mp$	$D^0 \to \pi^-\pi^+\mu^\pm e^\mp$
$\epsilon_{BDT}^{\texttt{norm},1516}$	[%]	48.25 ± 0.24	67.22 ± 0.32	54.46 ± 0.30
$\epsilon_{BDT}^{\texttt{norm},1718}$	[%]	45.49 ± 0.22	64.81 ± 0.29	51.07 ± 0.26
$\epsilon_{BDT}^{\texttt{noBrem},1516}$	[%]	50.74 ± 0.83	62.4 ± 0.69	47.55 ± 0.59
$\epsilon_{BDT}^{\texttt{noBrem},1718}$	[%]	50.77 ± 0.56	61.84 ± 0.47	49.05 ± 0.39
$\epsilon_{BDT}^{\texttt{Brem},1516}$	[%]	55.96 ± 1.23	68.02 ± 0.86	54.84 ± 0.74
$\epsilon_{BDT}^{\texttt{Brem},1718}$	[%]	56.65 ± 0.84	68.38 ± 0.6	56.19 ± 0.49
$r_{BDT}^{\texttt{noBrem},1516}$		1.052 ± 0.018	0.928 ± 0.011	0.873 ± 0.012
$r_{BDT}^{\texttt{noBrem},1718}$		1.116 ± 0.013	0.954 ± 0.008	0.960 ± 0.009
$r_{BDT}^{\texttt{Brem},1516}$		1.160 ± 0.026	1.012 ± 0.014	1.007 ± 0.015
$r_{BDT}^{\texttt{Brem},1718}$		1.245 ± 0.019	1.055 ± 0.01	1.100 ± 0.011

		$D^0 \to K^-K^+\mu^-\mu^+$	$D^0 \to \pi^-\pi^+\mu^-\mu^+$
$\epsilon_{BDT}^{\texttt{ref},1516}$	[%]	52.37 ± 0.40	50.48 ± 0.33
$\epsilon_{BDT}^{\texttt{ref},1718}$	[%]	49.08 ± 0.25	48.01 ± 0.21
$r_{BDT}^{\texttt{ref},1516}$		1.085 ± 0.010	0.927 ± 0.008
$r_{BDT}^{\texttt{ref},1718}$		1.079 ± 0.007	0.94 ± 0.006

Table 6.4: BDT efficiency and efficiency ratio for the different signal modes split by the different data categories and years. The efficiencies for the normalisation mode selection used for $D^0 \to K^-K^+\mu^-\mu^+$ ($D^0 \to \pi^-\pi^+\mu^-\mu^+$) decays and $D^0 \to K^-K^+\mu^\pm e^\mp$ ($D^0 \to \pi^-\pi^+\mu^\pm e^\mp$) decays is the same, due to the same selection applied to the normalisation mode.

To evaluate the efficiencies in the calibration sample the so-called `TISTOS` method is used [109]. The principal idea of the `TISTOS` method is to calculate the efficiency on an independent, triggered subsample. If the subsample is independent of the trigger decisions, *i.e.* decays are drawn randomly, the subsample trigger efficiency corresponds, within its uncertainty, to the efficiency of the complete sample. In the case of the `L0Muon` trigger decision, the efficiency is calculated in the following way:

$$\epsilon_{TOS} = \frac{N_{TIS\&TOS}}{N_{TIS}}, \tag{6.4.1}$$

where N_{TIS} corresponds to the number of `TIS` events by the ECAL, `L0Electron TIS`, and Muon stations `L0Muon TIS` and $N_{TIS\&TOS}$ is the number of events which trigger `L0Muon TOS` in the `TIS` selected sample.

Analogously to this, $\epsilon_{TIS\backslash TOS}$ is calculated by:

$$\epsilon_{TIS\backslash TOS} = \frac{N_{TIS\&TOS}}{N_{TOS}}, \tag{6.4.2}$$

where N_{TOS} corresponds to the number of events after selecting `L0Muon` and `L0Hadron TOS` on any of the daughter particles and $N_{TIS\&TOS}$ is the number of events triggered by `L0Muon` and `L0Electron TIS` in this subsample. The high occupancy linked with relatively large cluster sizes does not allow the assumption that `L0Electron TOS` and `L0Electron TIS` are independent of each other. The problem here is that due to the ECAL resolution, it is not possible to ensure that the `TIS` and `TOS` tags are properly assigned. Thus, trigger information from the ECAL is omitted.

To evaluate the efficiency of data with the `TISTOS` method, a high-purity sample is required to identify the desired decay events independent of their decay trigger selections[3]. The samples are corrected for simulation data differences in the `TISTOS` method to ensure an unbiased trigger efficiency measurement. These trigger efficiency correction weights are estimated using $B^0 \to K^{*0}(\to K^+\pi^-)J/\psi(\to \mu^+\mu^-)$ decays. An independent selection leading to a high purity is applied to these decays, the same as for the R_X analysis, by selecting a narrow J/ψ mass window [13, 14].

The data-simulation correction weight w_{L0} is then defined by:

$$w_{L0} = \frac{\epsilon_{L0}^{data}}{\epsilon_{L0}^{MC}} \tag{6.4.3}$$

where both efficiencies ϵ_{L0}^{data} and ϵ_{L0}^{MC} are determined using the `TISTOS` method. The weights are calculated and applied individually for both L0 trigger categories. The `L0Muon TOS` weights depend on the (transverse) momentum of the muon track and are applied to the simulation sample in bins of p_T and p with the following bin edges:

- p: $[0, 15000, 10000000]\,\mathrm{MeV}/c$
- p_T: $[0, 550, 750, 1000, 1250, 1600, 2000, 2800, 20000]\,\mathrm{MeV}/c$

The weights in their corresponding bins are reported in figure 6.4.

The efficiency in the $\epsilon_{TIS\backslash TOS}$ category depends mainly on the detector occupancy and potentially on the position of the particle tracks within the detector. For this purpose the weights are applied in bins of `nTracks`, which is the number of reconstructed tracks, and

[3]This is the reason why the trigger efficiency is calculated last.

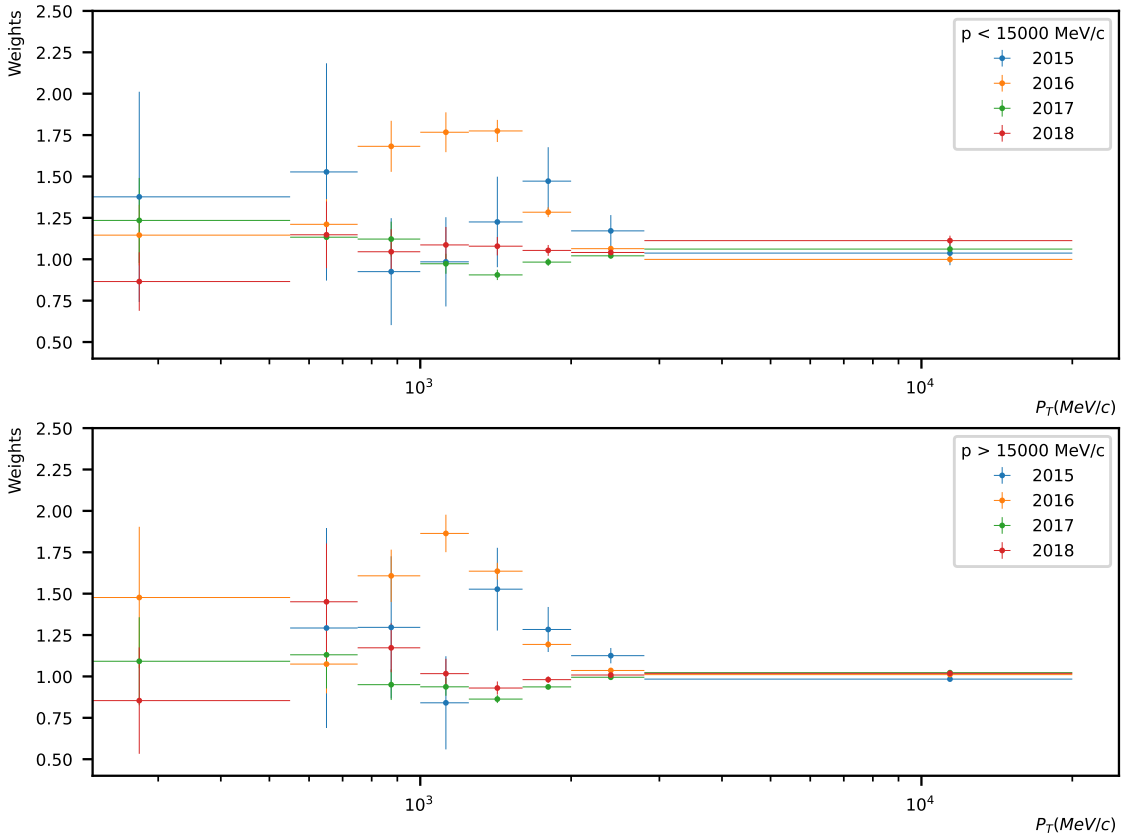


Figure 6.4: Muon data-simulation trigger correction weights in bins of p_T and p, split by the different years.

p_T of the mother particle, the B^0 candidate for the calibration sample or for signal the D^{*+} candidate. The bin edges are defined as follows:

- `nTracks`: $[0, 80, 120, 170, 700]$
- p_T: $[0, 4000, 6000, 8000, 10000, 12000, 13000, 15000, 17000, 19000, 50000]$ MeV/c

The weights in their corresponding bins are reported in figure 6.5.

The individual correction weights are mostly consistent over the different years. However, some deviations between the muon correction weights can be observed. It is important to note that for all years, the correction weights have relatively large uncertainties. These uncertainties are studied in detail in section 8.7, and a systematic uncertainty has to be assigned. The final corrected trigger efficiencies are reported in table 6.5. The `L0Muon`

trigger decision strongly depends on the momentum of the muon it was triggered by. This can also partially be observed for the LFV decay modes. The momentum of the muon scales with the available phase-space for the dilepton object. However, a direct interpretation of the trigger efficiencies at this step is difficult due to the pre-existing selection applied to the data samples, such as the different PID requirements on the lepton candidates.

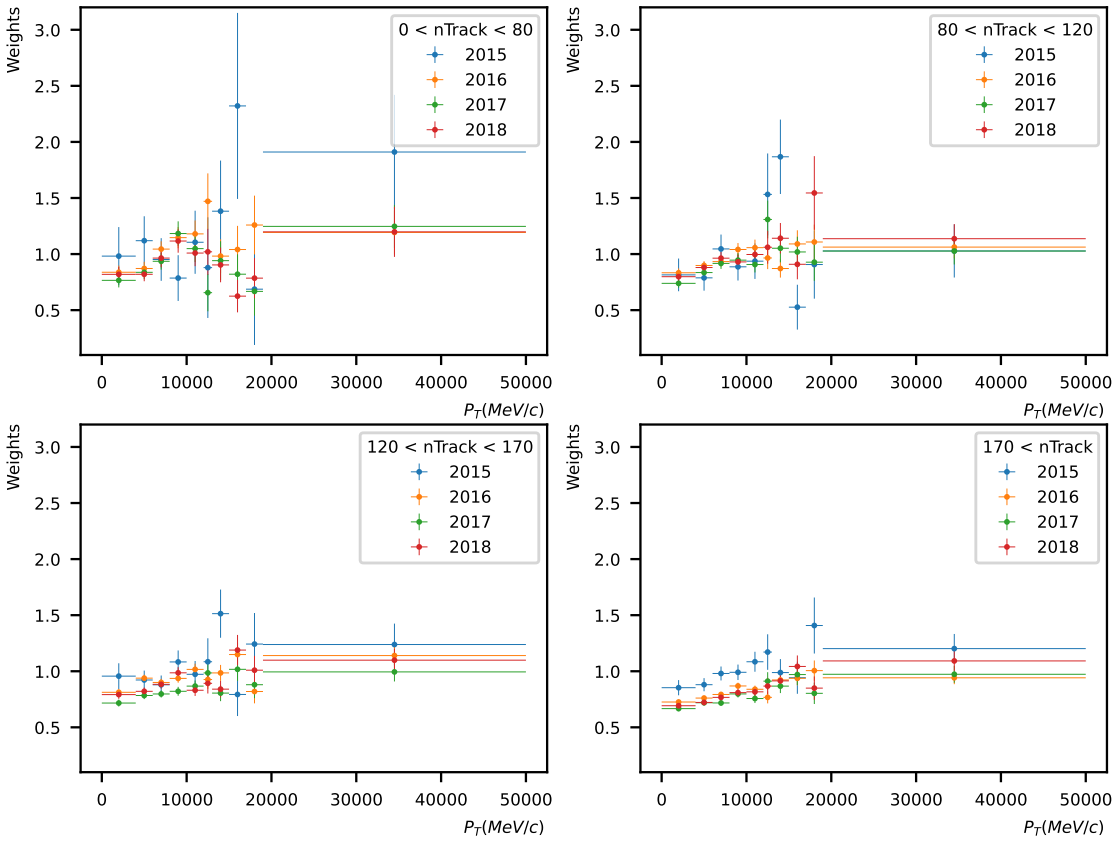


Figure 6.5: Electron data-simulation trigger correction weights in bins of p_T and p, split by the different years.

		$D^0 \to K^-K^+\mu^\pm e^\mp$	$D^0 \to K^-\pi^+\mu^\pm e^\mp$	$D^0 \to \pi^-\pi^+\mu^\pm e^\mp$
$\epsilon_{trig}^{\texttt{norm},1516}$	[%]	11.04 ± 0.01	14.99 ± 0.01	15.41 ± 0.01
$\epsilon_{trig}^{\texttt{norm},17168}$	[%]	28.90 ± 0.0	36.18 ± 0.01	36.94 ± 0.01
$\epsilon_{trig}^{\texttt{noBrem},1516}$	[%]	5.30 ± 0.02	10.16 ± 0.01	13.59 ± 0.01
$\epsilon_{trig}^{\texttt{noBrem},1718}$	[%]	15.60 ± 0.01	27.28 ± 0.01	33.44 ± 0.01
$\epsilon_{trig}^{\texttt{Brem},1516}$	[%]	2.10 ± 0.02	5.03 ± 0.02	3.94 ± 0.02
$\epsilon_{trig}^{\texttt{Brem},1718}$	[%]	12.43 ± 0.01	24.94 ± 0.01	31.83 ± 0.01
$r_{trig}^{\texttt{noBrem},1516}$		0.480 ± 0.046	0.678 ± 0.036	0.882 ± 0.042
$r_{trig}^{\texttt{noBrem},1718}$		0.540 ± 0.012	0.754 ± 0.011	0.905 ± 0.010
$r_{trig}^{\texttt{Brem},1516}$		0.190 ± 0.030	0.336 ± 0.033	0.256 ± 0.025
$r_{trig}^{\texttt{Brem},1718}$		0.430 ± 0.014	0.689 ± 0.012	0.862 ± 0.011

		$D^0 \to K^-K^+\mu^-\mu^+$	$D^0 \to \pi^-\pi^+\mu^-\mu^+$
$\epsilon_{trig}^{\texttt{ref},1516}$	[%]	6.87 ± 0.37	16.33 ± 0.31
$\epsilon_{trig}^{\texttt{ref},1718}$	[%]	21.43 ± 0.15	40.81 ± 0.15
$r_{trig}^{\texttt{ref},1516}$		0.629 ± 0.037	1.060 ± 0.031
$r_{trig}^{\texttt{ref},1718}$		0.741 ± 0.007	1.106 ± 0.007

Table 6.5: The efficiency ratio for the trigger efficiency for the different signal modes split by the different data categories used in the simultaneous fit.

6.5 Concluding remarks

All efficiency ratios, as well as the final efficiency ratio

$$r = r_{acc} \times r_{rec} \times r_{PID} \times r_{BDT} \times r_{trig}, \tag{6.5.1}$$

for the dimuon and LFV decays are summarised in figure 6.6. The efficiency ratio is larger for dimuon decays than for decays including electrons. This is mainly due to a worse tracking efficiency for electrons compared to muons. This can be seen when directly comparing r_{rec} between reference and signal decay modes in figure 6.6. Note that the normalisation mode efficiencies for $D^0 \to \pi^-\pi^+\mu^-\mu^+$ and $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ decays and $D^0 \to K^-K^+\mu^-\mu^+$ and $D^0 \to K^-K^+\mu^\pm e^\mp$ decays are the same, allowing a direct comparison[4]. The second source leading to a lower efficiency is the inefficient trigger in the case of bremsstrahlung during the first two years of data taking.

Comparing the different decay modes it can be observed that the ratio for decays with lighter hadrons in the final state is larger. This is expected because of an increased or reduced available phase space for all daughter particles due to the kaon mass. The increased phase space leads to higher momenta of the daughter particles, thus increasing the efficiency and vice versa.

In general, it can be observed that the trigger and reconstruction efficiency are the limiting factors when studying electron decays, at least compared to muon decays, at the LHCb detector. No unexpected behaviour is observed in the efficiency ratio. With the obtained efficiency ratio, only the signal to normalisation yield ratio is missing to calculate the branching fraction for the signal decays.

[4]The separation in bremsstrahlung categories is factored into the reconstruction efficiency. To compare the values directly, the bremsstrahlung categories need to be added for r_{rec}.

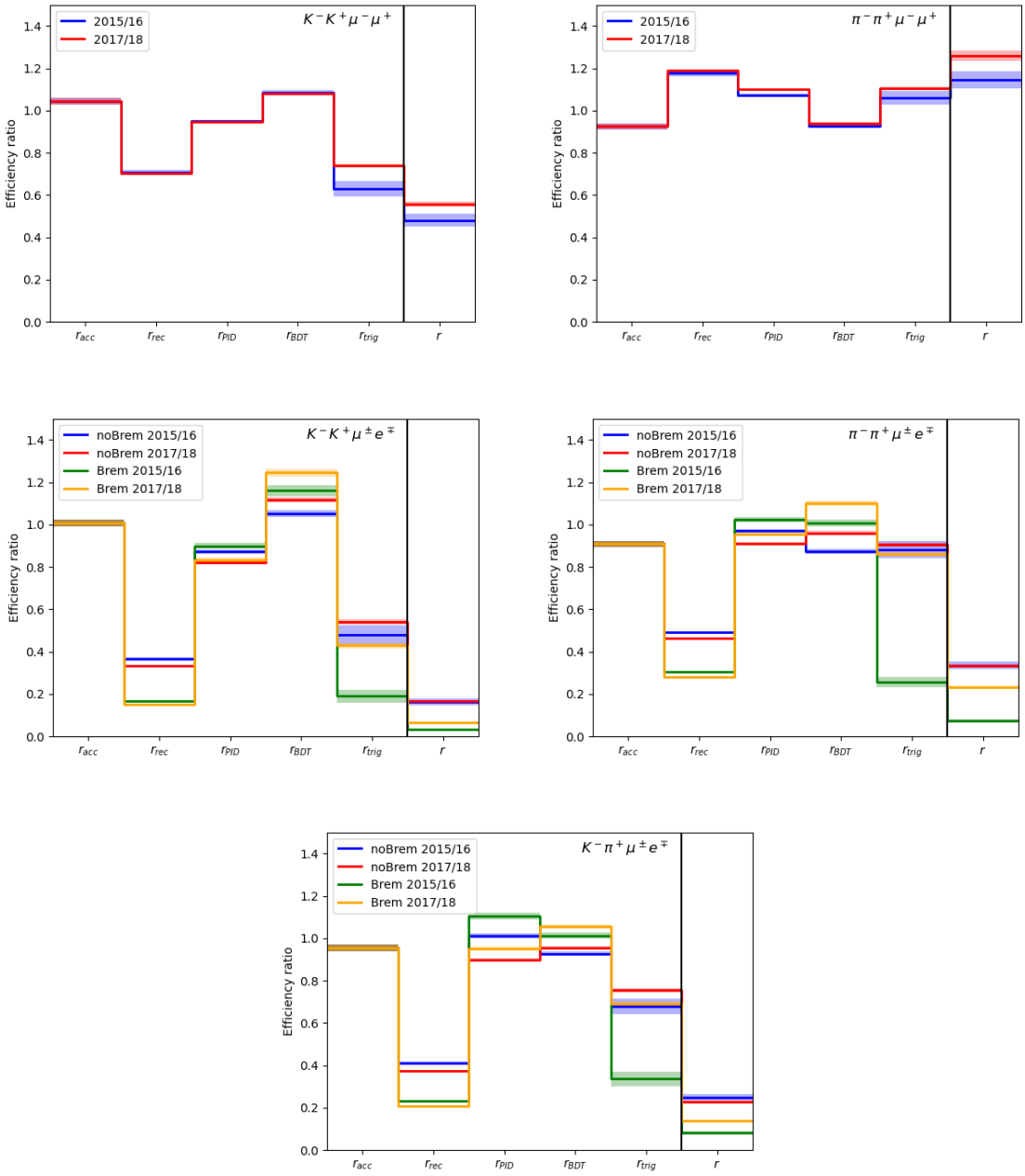


Figure 6.6: Summary of the efficiency ratios for the two reference decay modes, $D^0 \to K^-K^+\mu^-\mu^+$ and $D^0 \to \pi^-\pi^+\mu^-\mu^+$ (upper row) and the LFV decay modes, $D^0 \to K^-K^+\mu^\pm e^\mp$, $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ (both middle row) and $D^0 \to K^-\pi^+\mu^\pm e^\mp$ (lower row). The efficiency ratios are reported separately for each data category, time period and selection step. Additionally also the overall efficiency ratio $r = r_{acc} \times r_{rec} \times r_{PID} \times r_{BDT} \times r_{trig}$ is reported.

Chapter 7

Determination of signal yield

In this chapter, the estimation of the yield ratio is detailed. The number of left-over background and signal events after the selection process is calculated by a fit to the data sample. First, distributions to describe signal and background are defined, which are combined into a simultaneous fit. Using the efficiency and normalisation ratio as external input, the fit is used to evaluate the branching fraction directly. In the next step, pseudo-experiments are generated and measured to validate the fit. Lastly, the resulting statistical sensitivity for the LFV decays and the branching fraction for $D^0 \to h^- h^+ \mu^- \mu^+$ decays is reported.

After optimising the signal for its significance in observing LFV decays, the data sample is not background-free. To describe the data sample composition, three categories are defined:

- *Signal*: possible candidates of the decay of interest, $D^0 \to h^{(\prime)-} h^+ \mu^\pm e^\mp$ decays, $D^0 \to h^- h^+ \mu^- \mu^+$ decays and the normalisation decay $D^0 \to K^- \pi^+ [\mu^+ \mu^-]_{\rho^0/\omega}$. These are expected to peak in the reconstructed invariant mass at the D^0 mass.
- *Peaking background*: candidates of the form $D^0 \to h^{(\prime)-} h^+ \pi^- \pi^+$ where two pions are misidentified as muon and electron.
- *Non-peaking background*: a collection of all candidates who passed the selection, detailed in chapter 5, not associated with a particular decay. This includes single candidates from partially or wrongly reconstructed decays, or random combinations of tracks or particles that meet all selection criteria.

Other in the fit model neglected background sources are studied in section 8.1. To separate and count the number of potential signal candidates or evaluate the significance

in case of no observed signal, the invariant D^0 mass under the signal mass hypothesis is studied.

All three data components behave differently in the reconstructed invariant D^0 mass distribution. The signal is expected to be distributed around the D^0 mass. In contrast, the fully hadronic decays are expected to be majorly distributed in the lower mass spectrum due to the wrong mass hypothesis applied in the calculation of the D^0 mass. The combinatorial background should be smoothly distributed in the studied mass window around the D^0 mass.

The distributions themselves are extracted either from data or are fixed by simulation. This is done to avoid any ambiguity in the fit and properly separate signal and background candidates. The individual components are combined into one fit model, allowing direct access to the signal component's yield. A simultaneous fit to all decay categories is performed to simplify the calculation and directly estimate the significance of the model compared to the background-only hypothesis. This allows the direct calculation of the branching fraction. As mentioned before, the data samples used to study the LFV decays are divided into three categories, `noBrem` LFV decay candidates where no bremsstrahlung is reconstructed, `Brem` LFV decay candidates where bremsstrahlung is reconstructed and `norm` which consists of candidates passing the selection for the normalisation mode which matches the signal mode[1]. For the dimoun mode two data categories are considered `ref`, which contains the signal, and `norm`, which contains the normalisation mode. These categories are then further split into two data taking periods, one for 2015 and 2016 and another for 2017 and 2018.

LFV decays

For the signal decays, the shape of the signal distribution and possible background sources depend strongly on the recovery of the bremsstrahlung photon. If a photon is added to the decay chain, due to the limited ECAL resolution[2] a wider but mostly symmetric mass distribution around the D^0 is expected, provided the correct photon is added during recovery. In contrast, if no photon is added to the decay chain, it often means that the photon was lost during the reconstruction and not necessarily that no photon was emitted.

[1]The main difference here is that three independent BDTs were trained for the three LFV decay modes. To avoid biases introduced by the BDT, three different selections were applied to the normalisation mode. For simplicity reasons also the PID selection steps are kept as similar as possible to the signal modes.

[2]In comparison to the momentum resolution of the tracking system.

An asymmetric distribution with a lower tail and a relatively sharp edge at the D^0 mass is expected for decays without added photons. Due to energy and momentum conservation, no signal candidates with a reconstructed D^0 mass larger than the real D^0 mass are expected within their resolution. The edge is defined by the momentum resolution of the tracking system. The second big difference concerning possible background contamination is that pions do not emit bremsstrahlung, allowing the use of the photon as a powerful tag to separate decays containing electrons and fully hadronic decays.

$D^0 \to h^-h^+\mu^-\mu^+$ and $D^0 \to K^-\pi^+[\mu^+\mu^-]_{\rho^0/\omega}$ decays

The same distributions as for the signal mode are used for the fit to the reference and normalisation mode. The reference mode candidates are expected to be more symmetrically distributed around the D^0 mass. Due to the smaller mass difference, misID candidates are expected to be distributed closer to the D^0 mass. Signal, misID and combinatorial background event distributions use the same parameterisation as used for the signal mode.

7.1 Signal

The signal distribution is described by a Johnson's S_U distribution [110]. The Johnson's S_U distribution is a four-parametric function resulting from a variable transformation of a normal distribution to allow for asymmetric tails. It has the following form:

$$f(m;\mu,\lambda,\gamma,\delta) = \frac{\delta}{\lambda\sqrt{2\pi}} \frac{1}{\sqrt{1+\left(\frac{m-\mu}{\lambda}\right)^2}} \exp\left[-\frac{1}{2}\left(\gamma+\delta\sinh^{-1}\left(\frac{m-\mu}{\lambda}\right)\right)^2\right] \quad (7.1.1)$$

μ is the location parameter of the Gaussian component, λ the width parameter of the Gaussian component, γ the shape parameter that distorts distribution to the left or right and δ the shape parameter that determines the strength of the Gaussian-like component.

The asymmetric Johnson's S_U distribution is chosen to parameterise the upper or lower mass distribution tail while keeping the free parameters at a minimum, compared, for example, to a double-sided Crystal-Ball distribution [111]. In section 8.2, alternative signal parameterisations and their systematic influence are studied. Exemplary in figure 7.1 a fit of the Johnson's S_U distribution to the normalisation channel[3] and to $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ decays in the two data categories `Brem` and `noBrem` is shown.

[3]Applying the selection normalisation channel selection for the $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ normalisation.

The signal shape in the `Brem` category has, as expected, a more pronounced tail for higher masses, contrary to the distribution in the `noBrem` category with a more pronounced tail for lower masses. Because of the momentum dependence of the efficiency, a slight asymmetry for $D^0 \to K^-\pi^+[\mu^+\mu^-]_{\rho^0/\omega}$ decays, under $D^0 \to K^-\pi^+[\mu^+\mu^-]_{\rho^0/\omega}$ mass hypothesis, is expected, which is also observed in figure 7.1. The same is true for the reference modes. All parameters describing the signal distributions, as well as the number of simulated events used to obtain the shape parameters, are reported in table 7.1 for

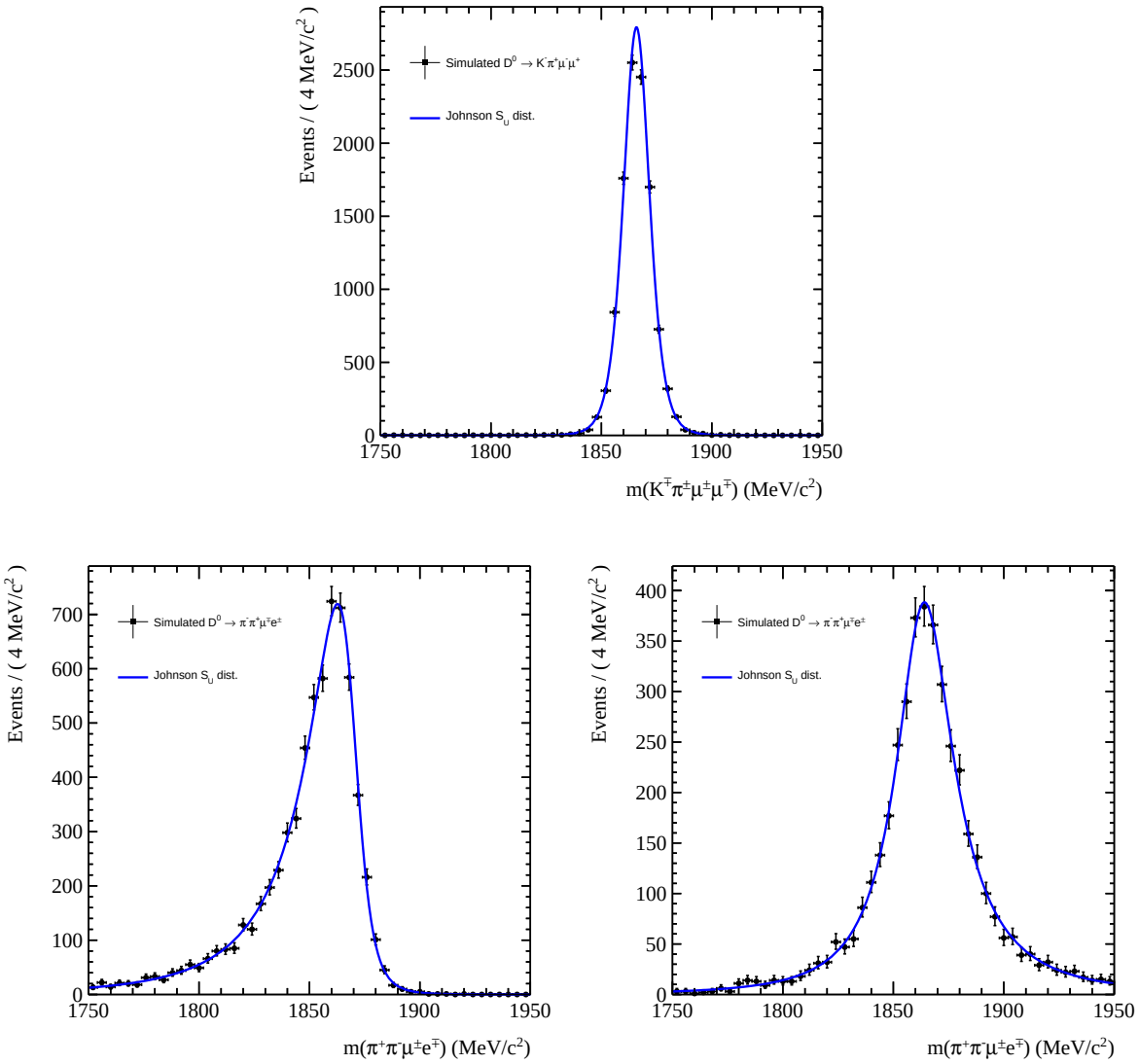


Figure 7.1: Example shape of the signal distribution used to estimate the number of normalisation candidates (up) and $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ candidates in the data categories `noBrem` (lower left) and `Brem` (lower right). The signal shapes are projected on the simulation samples for the data taking period between 2017 and 2018, which were used to extract the shape.

data taken in 2015 and 2016 and in table 7.2 for data taken in 2017 and 2018. Generally, the shape parameters are compatible within their uncertainties between the two data taking periods. All signal shape distributions for the LFV decays modes are reported in appendix F.

Parameter	$D^0 \to K^-K^+\mu^\pm e^\mp$	$D^0 \to K^-\pi^+\mu^\pm e^\mp$	$D^0 \to \pi^-\pi^+\mu^\pm e^\mp$
$N_{sig}^{1516,\texttt{norm}}$	3493 ± 59	3116 ± 56	3351 ± 58
$\delta_{sig}^{1516,\texttt{norm}}$	1.79 ± 0.096	1.92 ± 0.12	1.85 ± 0.11
$\gamma_{sig}^{1516,\texttt{norm}}$	0.05 ± 0.068	0.011 ± 0.08	0.052 ± 0.073
$\mu_{sig}^{1516,\texttt{norm}}$	1866.13 ± 0.49	1865.89 ± 0.57	1866.21 ± 0.54
$\lambda_{sig}^{1516,\texttt{norm}}$	11.69 ± 0.76	12.59 ± 0.94	12.34 ± 0.86
$N_{sig}^{1516,\texttt{Brem}}$	73 ± 9	197 ± 14	225 ± 15
$\delta_{sig}^{1516,\texttt{Brem}}$	0.9 ± 0.27	0.76 ± 0.16	1.09 ± 0.19
$\gamma_{sig}^{1516,\texttt{Brem}}$	-0.79 ± 0.32	-0.41 ± 0.14	-0.41 ± 0.17
$\mu_{sig}^{1516,\texttt{Brem}}$	1858.1 ± 4.5	1861.9 ± 2.0	1862.1 ± 2.4
$\lambda_{sig}^{1516,\texttt{Brem}}$	10.9 ± 4.1	10.4 ± 2.9	14.2 ± 3.3
$N_{sig}^{1516,\texttt{noBrem}}$	288 ± 17	610 ± 25	949 ± 31
$\delta_{sig}^{1516,\texttt{noBrem}}$	1.42 ± 0.25	1.25 ± 0.14	1.4 ± 0.13
$\gamma_{sig}^{1516,\texttt{noBrem}}$	1.43 ± 0.43	1.51 ± 0.23	1.24 ± 0.15
$\mu_{sig}^{1516,\texttt{noBrem}}$	1870.5 ± 3.4	1872.4 ± 2.0	1870.7 ± 1.7
$\lambda_{sig}^{1516,\texttt{noBrem}}$	10.6 ± 1.9	10.2 ± 1.4	15.5 ± 1.9

Parameter	$D^0 \to K^-K^+\mu^-\mu^+$	$D^0 \to \pi^-\pi^+\mu^-\mu^+$
N_{sig}^{1516}	8136 ± 90	16360 ± 130
δ_{sig}^{1516}	2.29 ± 0.32	2.12 ± 0.18
γ_{sig}^{1516}	-0.42 ± 0.22	0.08 ± 0.1
μ_{sig}^{1516}	1863.5 ± 1.2	1866.73 ± 0.9
λ_{sig}^{1516}	11.8 ± 1.9	17.2 ± 1.7

Table 7.1: Summary of the signal fit parameters for the LFV signal decays and correspondingly selected normalisation modes obtained by a fit to the simulation data samples for 2015 and 2016. The number of simulated events passing the selection procedure, as well as the shape parameters of the Johnson S_U distribution, are reported.

Parameter	$D^0 \to K^-K^+\mu^\pm e^\mp$	$D^0 \to K^-\pi^+\mu^\pm e^\mp$	$D^0 \to \pi^-\pi^+\mu^\pm e^\mp$
$N_{sig}^{1718,\texttt{norm}}$	12240 ± 110	10600 ± 100	11060 ± 110
$\delta_{sig}^{1718,\texttt{norm}}$	1.813 ± 0.052	1.815 ± 0.056	1.815 ± 0.054
$\gamma_{sig}^{1718,\texttt{norm}}$	0.012 ± 0.037	0.007 ± 0.04	0.009 ± 0.039
$\mu_{sig}^{1718,\texttt{norm}}$	1865.99 ± 0.26	1865.9 ± 0.27	1865.92 ± 0.27
$\lambda_{sig}^{1718,\texttt{norm}}$	11.51 ± 0.4	11.39 ± 0.43	11.47 ± 0.42
$N_{sig}^{1718,\texttt{Brem}}$	911 ± 30	2229 ± 47	4167 ± 65
$\delta_{sig}^{1718,\texttt{Brem}}$	0.664 ± 0.06	0.8 ± 0.049	0.803 ± 0.037
$\gamma_{sig}^{1718,\texttt{Brem}}$	-0.428 ± 0.065	-0.179 ± 0.044	-0.135 ± 0.033
$\mu_{sig}^{1718,\texttt{Brem}}$	1861.36 ± 0.88	1862.5 ± 0.76	1863.13 ± 0.6
$\lambda_{sig}^{1718,\texttt{Brem}}$	9.3 ± 1.1	13.4 ± 1.1	14.23 ± 0.84
$N_{sig}^{1718,\texttt{noBrem}}$	2074 ± 46	3968 ± 63	6537 ± 81
$\delta_{sig}^{1718,\texttt{noBrem}}$	0.919 ± 0.044	0.995 ± 0.038	1.045 ± 0.034
$\gamma_{sig}^{1718,\texttt{noBrem}}$	1.21 ± 0.075	1.297 ± 0.062	1.231 ± 0.05
$\mu_{sig}^{1718,\texttt{noBrem}}$	1868.03 ± 0.59	1869.42 ± 0.58	1870.36 ± 0.57
$\lambda_{sig}^{1718,\texttt{noBrem}}$	6.5 ± 0.43	8.51 ± 0.43	10.93 ± 0.45

Parameter	$D^0 \to K^-K^+\mu^-\mu^+$	$D^0 \to \pi^-\pi^+\mu^-\mu^+$
N_{sig}^{1718}	8136 ± 90	16360 ± 130
δ_{sig}^{1718}	1.906 ± 0.078	2.062 ± 0.065
γ_{sig}^{1718}	-0.035 ± 0.048	-0.015 ± 0.039
μ_{sig}^{1718}	1865.52 ± 0.26	1865.9 ± 0.33
λ_{sig}^{1718}	9.62 ± 0.47	16.41 ± 0.6

Table 7.2: Summary of the signal fit parameters for the LFV signal decays and correspondingly selected normalisation modes obtained by a fit to the simulation data samples for 2017 and 2018. The number of simulated events passing the selection procedure, as well as the shape parameters of the Johnson S_U distribution, are reported.

7.2 Peaking background

Misidentified $D^0 \to h^{(\prime)-}h^+\pi^-\pi^+$ decays, where the pions are misidentified as muon and electron, are separated from the signal by fitting the four-body invariant mass spectrum. While passing the same selection steps as for the signal candidates, the candidates are expected to peak to the left of the D^0 mass because of the wrong mass hypothesis. The two pions are reconstructed as muon-electron (muon-muon) candidates with a too-low mass. The same procedure was used in previous studies of $D^0 \to h^-h^+\mu^-\mu^+$ decays. In this analysis, the mass difference between the misidentified particle mass and the reconstructed mass assumption is even greater, leading to better signal-background separation. This advantage is reduced by the tail in the distribution introduced by missing bremsstrahlung. Bremsstrahlung recovery provides an additional opportunity, as heavier particles produce no bremsstrahlung, allowing for good signal-background separation[4].

A Johnson's S_U distribution is chosen to describe this component, defined in equation 7.1. Simulated samples are used to determine the shape of the distribution and the fraction of misID candidates between candidates with and without reconstructed bremsstrahlung. Due to the high inefficiency for background candidates, a looser selection is applied to the simulation samples. Also, the shapes for both data-taking periods are fixed using the larger simulation samples produced for 2017 and 2018. For the background of the LFV signal modes, the default selection without explicitly reapplying the HLT2 line and without PID requirements, except `ProbNNmu` > 0.5 on the muon candidate, is applied. For the bremsstrahlung recovery, the pion must be reconstructed as an electron, therefore also $DLL_e > -2$ is applied. The shape and bremsstrahlung candidate fraction are fixed to values obtained from simulations. All misID background shape parameters obtained for the LFV decay modes, including the normalisation mode with the corresponding selection, are reported in table 7.3. As an example, the misID distributions obtained from simulations for the $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ decay mode and the correspondingly selected normalisation mode are reported in figure 7.2. Note that even though the misID contribution is suppressed strongly if bremsstrahlung is reconstructed, the distribution is moved under the signal distribution because of wrongly added bremsstrahlung[5].

The same procedure as for the LFV signal mode without PID requirements except `ProbNNmu` > 0.5 on one randomly assigned muon is applied for the normalisation, $D^0 \to K^-\pi^+[\mu^+\mu^-]_{\rho^0/\omega}$, and reference, $D^0 \to K^-K^+\mu^-\mu^+$ and $D^0 \to \pi^-\pi^+\mu^-\mu^+$, decay

[4] At the end, both bremsstrahlung categories contribute to equal amounts to the limit.

[5] Pions do not emit bremsstrahlung.

channels.

The shape of the mass distribution depends on the PID distribution. Therefore, a systematic uncertainty will be assigned, which is detailed in section 8.4. All fit results, including the values for the shape parameters, are reported in table 7.3. All peaking background shape distributions for the LFV decays modes are reported in appendix F.

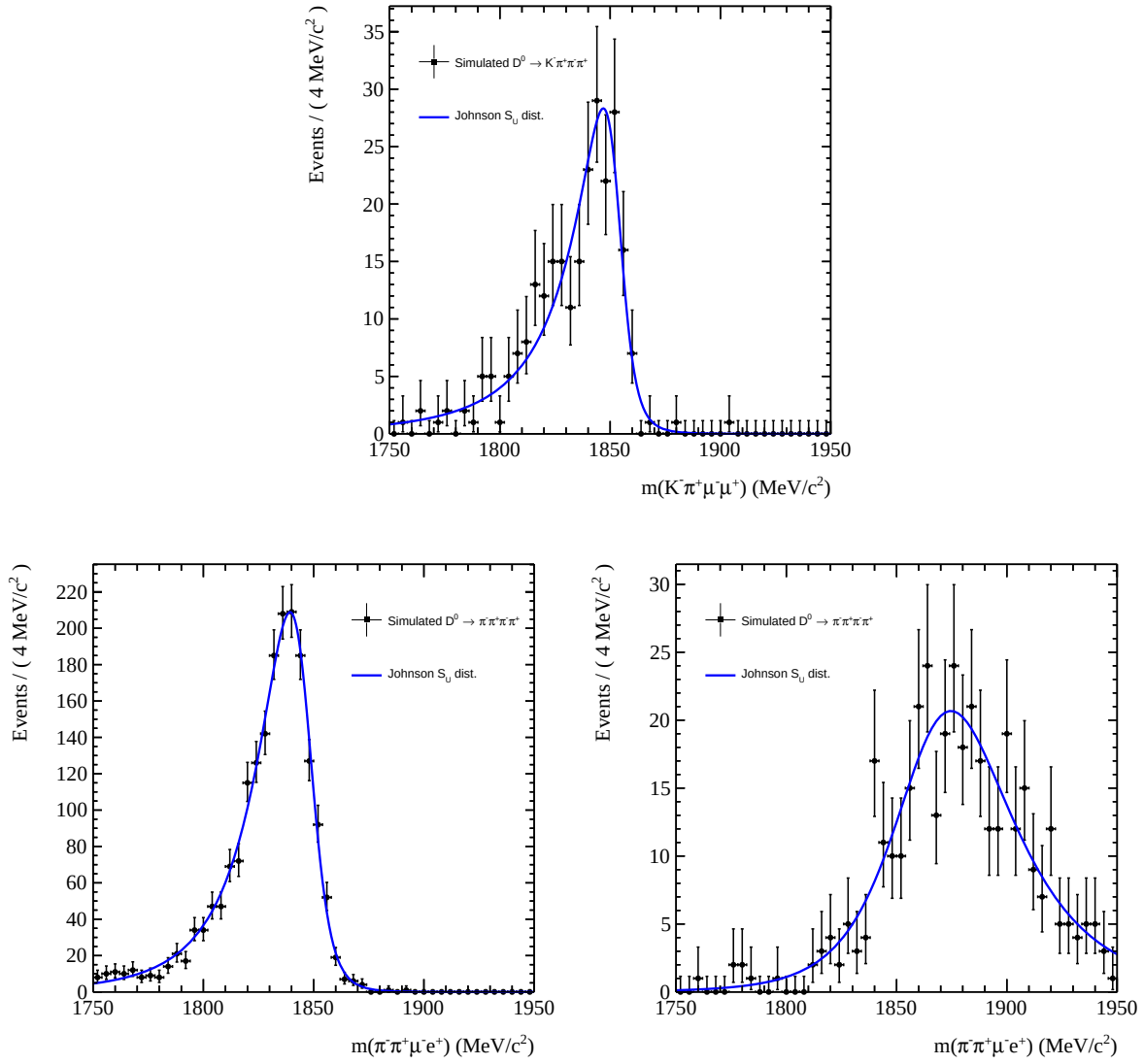


Figure 7.2: Example misID shape fit to simulated $D^0 \to K^-\pi^+\pi^+\pi^-$ (up) and $D^0 \to \pi^-\pi^+\pi^-\pi^+$ (down) decays. The distribution in the upper plot shows the reconstructed mass under $D^0 \to K^-\pi^+[\mu^+\mu^-]_{\rho^0/\omega}$ mass hypothesis. The lower distributions, $D^0 \to \pi^-\pi^+\pi^-\pi^+$ decays, are reconstructed under $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ mass hypothesis and applied bremsstrahlung recovery. Events where no bremsstrahlung (lower left) is added and events with added bremsstrahlung (lower right) are shown. All three plots are for the data-taking period 2017/2018.

Parameter	$D^0 \to K^-K^+\mu^\pm e^\mp$	$D^0 \to K^-\pi^+\mu^\pm e^\mp$	$D^0 \to \pi^-\pi^+\mu^\pm e^\mp$
$N_{misID}^{1718,\texttt{norm}}$	412 ± 20	310 ± 18	249 ± 16
$\delta_{misID}^{1718,\texttt{norm}}$	1.28 ± 0.22	1.24 ± 0.23	1.0 ± 0.16
$\gamma_{misID}^{1718,\texttt{norm}}$	2.06 ± 0.48	1.62 ± 0.33	1.29 ± 0.2
$\mu_{misID}^{1718,\texttt{norm}}$	1859.8 ± 3.9	1857.5 ± 3.2	1854.2 ± 2.0
$\lambda_{misID}^{1718,\texttt{norm}}$	10.4 ± 1.9	11.2 ± 2.7	9.7 ± 2.4
$N_{misID}^{1718,\texttt{Brem}}$	122 ± 11	166 ± 13	376 ± 19
$\delta_{misID}^{1718,\texttt{Brem}}$	0.34 ± 0.24	0.35 ± 0.29	1.35 ± 0.4
$\gamma_{misID}^{1718,\texttt{Brem}}$	0.27 ± 0.41	-0.86 ± 0.65	-0.44 ± 0.25
$\mu_{misID}^{1718,\texttt{Brem}}$	1824.7 ± 4.6	1859.0 ± 5.1	1866.0 ± 7.1
$\lambda_{misID}^{1718,\texttt{Brem}}$	10.4 ± 4.3	15.8 ± 6.7	40.0 ± 13.0
$N_{misID}^{1718,\texttt{noBrem}}$	142 ± 12	572 ± 24	1910 ± 44
$\delta_{misID}^{1718,\texttt{noBrem}}$	1.4 ± 0.46	1.07 ± 0.12	1.223 ± 0.084
$\gamma_{misID}^{1718,\texttt{noBrem}}$	2.08 ± 0.87	1.12 ± 0.15	1.19 ± 0.11
$\mu_{misID}^{1718,\texttt{noBrem}}$	1846.6 ± 8.0	1843.4 ± 1.7	1848.3 ± 1.3
$\lambda_{misID}^{1718,\texttt{noBrem}}$	13.0 ± 4.4	11.0 ± 1.7	13.8 ± 1.1

Parameter	$D^0 \to K^-K^+\mu^-\mu^+$	$D^0 \to \pi^-\pi^+\mu^-\mu^+$
N_{misID}^{1718}	113 ± 11	367 ± 19
δ_{misID}^{1718}	1.49 ± 0.22	1.05 ± 0.12
γ_{misID}^{1718}	6.5 ± 4.0	1.45 ± 0.22
μ_{misID}^{1718}	1860.4 ± 2.6	1854.4 ± 2.4
λ_{misID}^{1718}	0.7 ± 1.9	10.9 ± 1.8

Table 7.3: Summary of misID background fit parameters for LFV and dimuon decays obtained by a fit to the simulation data samples. The corresponding normalisation mode parameters are also reported. The column labels mark the applied selection. The normalisation misID background shapes are calculated from simulated $D^0 \to K^-\pi^+\pi^+\pi^-$ decays under $D^0 \to K^-\pi^+[\mu^+\mu^-]_{\rho^0/\omega}$ mass hypotheses. The mass distribution of the misID backgrounds, $D^0 \to \pi^-\pi^+\pi^-\pi^+$, $D^0 \to K^-\pi^+\pi^+\pi^-$ and $D^0 \to K^-K^+\pi^-\pi^+$ decays, are reconstructed under the corresponding signal mass hypothesis, $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ ($D^0 \to \pi^-\pi^+\mu^-\mu^+$), $D^0 \to K^-\pi^+\mu^\pm e^\mp$ and $D^0 \to K^-K^+\mu^\pm e^\mp$ ($D^0 \to \pi^-\pi^+\mu^-\mu^+$), respectively. The misID background parameters for the normalisation modes are the same for the dimuon signal modes, $D^0 \to \pi^-\pi^+\mu^-\mu^+$ and $D^0 \to K^-K^+\mu^-\mu^+$, and $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ and $D^0 \to K^-K^+\mu^\pm e^\mp$ decays, respectively. The number of simulated events passing the selection procedure, as well as the shape parameters of the Johnson S_U distribution, are reported.

7.3 Non-peaking background

A third fit component is introduced to describe the leftover background events grouped together as *combinatorial background*. As no significant peaking background sources are suspected under the signal, the linear background is approximated by a linear function. For this purpose, a Chebyshev polynomial, 1st order, defined by:

$$\mathcal{T}(m|c_1) = \frac{1}{\mathcal{N}_L}(1 + c_1 m), \tag{7.3.1}$$

where $\mathcal{N}_L$ is the normalisation and c_1 the slope of the function. The slope parameter is fixed by the same fit from the upper mass sideband $m(D^0) > 1900\,\text{MeV}/c^2$. The fit distribution and fit projection can be found in figure 5.10 and is the same as used during the optimisation of the selection in section 5.3.3. The same selection as for the data sample is applied, but without a cut on the `BDT response` and an inverted PID selection. The inverted PID selection requires that the PID variable `ProbNN` for the leptons is smaller than 0.1. This ensures that no signal candidates are contaminating the upper sideband. The impact of the chosen parameterisation and the choice of the fit function is studied in section 8.3.

Parameter	$D^0 \to K^-K^+\mu^\pm e^\mp$	$D^0 \to K^-\pi^+\mu^\pm e^\mp$	$D^0 \to \pi^-\pi^+\mu^\pm e^\mp$
$c_1^{1516,\texttt{noBrem}}$	-0.009 ± 0.014	-0.127 ± 0.068	-0.116 ± 0.049
$c_1^{1718,\texttt{noBrem}}$	-0.141 ± 0.048	-0.101 ± 0.023	-0.021 ± 0.015
$c_1^{1516,\texttt{Brem}}$	0.0^*	0.0^*	-0.031 ± 0.129
$c_1^{1718,\texttt{Brem}}$	0.046 ± 0.139	-0.683 ± 0.026	-0.178 ± 0.033

Parameter	$D^0 \to K^-K^+\mu^-\mu^+$	$D^0 \to K^-\pi^+[\mu^+\mu^-]_{\rho^0/\omega}$	$D^0 \to \pi^-\pi^+\mu^-\mu^+$
c_1^{1516}	0.0^*	-0.11 ± 0.20	0.15 ± 0.21
c_1^{1718}	-0.30 ± 0.21	-0.11 ± 0.20	-0.11 ± 0.06

Table 7.4: Slope of the Chebychev polynomial used in the data fit obtained from the upper mass sideband $m(D^0) > 1900\,\text{MeV}/c^2$. The upper table shows the resulting slope for the LFV decay modes in the different data categories. The lower table shows the slope for the dimuon reference decays and the normalisation mode. The slope is assumed to be the same for the differently selected normalisation modes. *No events pass the selection, and the slope is set to 0.0.

7.4 Simultaneous Fit

The three fit components, signal, peaking background (misID) and combinatorial background, are combined into one fit function. With the shape parameters fixed from the simulation, there are only three free parameters left, the three yields. By solving equation 2.4.5 for the signal branching fraction,

$$\mathcal{B}(D^0 \to h^{(\prime)-}h^+l^{(\prime)-}l^+) = \frac{\mathrm{N_{sig}}}{\mathrm{N_{norm}}} \cdot \frac{\epsilon_{norm}}{\epsilon_{sig}} \cdot \mathcal{B}(D^0 \to K^-\pi^+[\mu^+\mu^-]_{\rho^0/\omega}), \tag{7.4.1}$$

a simultaneous fit between normalisation and signal mode is set up. $\mathcal{B}(D^0 \to h^{(\prime)-}h^+l^{(\prime)-}l^+)$ is the signal branching fraction and $\mathrm{N_{sig}}$ and $\mathrm{N_{norm}}$ the measured yields in signal and normalisation sample. $\mathcal{B}(D^0 \to K^-\pi^+[\mu^+\mu^-]_{\rho^0/\omega})$ is the branching fraction for the normalisation decay. For $\mathcal{B}(D^0 \to K^-\pi^+[\mu^+\mu^-]_{\rho^0/\omega})$ the PDG value is taken [102]. $\epsilon_{sig}/\epsilon_{norm}$ is the efficiency ratio. The efficiency ratio and the normalisation mode's branching fraction are provided as external input.

The signal yield is parameterised as a function of the branching fraction. This parameterisation allows to directly estimate the branching ratio without performing a second calculation, for example by an additional fit, to estimate the branching ratio. The statistical uncertainty can be accessed by the likelihood function. Following equation 7.4.1, the four signal yields are parameterised in the following way:

$$\mathrm{N}_{\mathrm{sig}}^{y\,(,c)} = \frac{\mathcal{B}(D^0 \to h^{(\prime)-}h^+l^{(\prime)-}l^+) + b}{\mathcal{B}(D^0 \to K^-\pi^+\mu^-\mu^+)} \cdot \mathrm{N}_{\mathrm{norm}}^{y} \cdot \frac{\epsilon_{sig}^{y\,(,c)}}{\epsilon_{norm}^{y}} \tag{7.4.2}$$

The upper index y donates the year of the data taking period, 2015 and 2016 (1516) and 2017 and 2018 (1718), and c the bremsstrahlung categories, `noBrem` and `Brem`. The blind b is zero for the dimuon modes and a random unknown number for the LFV decay modes. A separation in bremsstrahlung categories is not needed for the reference modes. A shift parameter ν_{sig} for the signal component is introduced for the reference modes. The shift accounts for possible differences in the peak position between simulations and data. The normalisation yield is free, with no dependencies between the two categories. Additionally, a shift parameter ν_{norm} is introduced to account for possible data-simulation differences. To increase the stability of the fit, the ratio between the misID background in the two categories is fixed to simulations. The uncertainty on the ratio is added as systematic uncertainty during the limit setting.

To summarise for the LFV signal (dimuon reference) mode, all shape and peak positions are fixed to simulations, apart from the two (four) shift parameters ν. The twelve

Parameter	$D^0 \to K^-K^+\mu^\pm e^\mp$	$D^0 \to K^-\pi^+\mu^\pm e^\mp$	$D^0 \to \pi^-\pi^+\mu^\pm e^\mp$
$\mathcal{B}(D^0 \to h^{(\prime)-}h^+\mu^\pm e^\mp)[10^{-8}]$	x.x $\pm$ 2.6	x.x $\pm$ 5.0	x.x $\pm$ 3.4
$N_{\text{Comb}}^{1516,\texttt{Brem}}$	5.7 ± 2.5	33.5 ± 6.1	18.8 ± 4.8
$N_{\text{Comb}}^{1516,\texttt{noBrem}}$	11.4 ± 4.7	59.1 ± 10.0	67.0 ± 11.0
$N_{\text{Comb}}^{1516,\texttt{norm}}$	10.7 ± 7.0	3.8 ± 3.9	10.2 ± 6.5
$N_{\text{Comb}}^{1718,\texttt{Brem}}$	70.4 ± 9.1	324.0 ± 20.0	455.0 ± 24.0
$N_{\text{Comb}}^{1718,\texttt{noBrem}}$	125.0 ± 15.0	315.0 ± 26.0	357.0 ± 26.0
$N_{\text{Comb}}^{1718,\texttt{norm}}$	82.0 ± 19.0	42.0 ± 13.0	32.0 ± 13.0
$N_{\texttt{norm}}^{1516}$	1004.0 ± 30.0	763.0 ± 26.0	866.0 ± 28.0
$N_{\texttt{norm}}^{1718}$	4745.0 ± 66.0	3685.0 ± 57.0	4215.0 ± 61.0
$N_{\text{misID}}^{1516,\text{norm}}$	583.0 ± 26.0	152.0 ± 14.0	232.0 ± 17.0
$N_{\text{misID}}^{1718,\text{norm}}$	2914.0 ± 57.0	777.0 ± 32.0	1325.0 ± 40.0
N_{misID}^{1516}	10.1 ± 3.9	41.4 ± 8.3	63.3 ± 9.9
N_{misID}^{1718}	30.0 ± 11.0	254.0 ± 22.0	269.0 ± 22.0
ν_{norm}^{1516}	-0.18 ± 0.24	-0.86 ± 0.26	-0.73 ± 0.25
ν_{norm}^{1718}	-0.45 ± 0.1	-0.78 ± 0.11	-0.61 ± 0.11

Table 7.5: Summary of fit parameters, the yields and position parameters ν, for the signal decays obtained by the simultaneous fit. The position parameters ν for the normalisation mode are defined relative to the position, which is given by the peak position μ of the signal contribution of the normalisation mode obtained from simulations. The quoted branching fraction is blind.

(eight) background yields, two per category, and the branching fraction are free to float. The signal mode background parameters are reduced by two degrees of freedom by fixing the ratio between the bremsstrahlung categories, resulting in 13 degrees of freedom for the signal mode and 12 for the reference mode. The parameters obtained by the fit are reported in table 7.5 for the signal modes and table 7.6 for the reference modes.

The goodness of the fit can be partially judged by the pull distribution, defined as the difference between the fit value and data per bin normalised to the error of the data point in the respective bin. In section 7.5, it is checked that the fit procedure itself does not bias the measured branching ratio. The reference mode fit results and pull distributions are shown in figure 7.3 for $D^0 \to K^-K^+\mu^-\mu^+$ and figure 7.4 for $D^0 \to \pi^-\pi^+\mu^-\mu^+$. The signal mode results can be found in figure 7.5 for $D^0 \to K^-K^+\mu^\pm e^\mp$, figure 7.6 for $D^0 \to K^-\pi^+\mu^\pm e^\mp$ and figure 7.7 for $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$. The signal modes are blind. Therefore, only the pull distribution is reported in the blind area. In general, a good

Parameter	$D^0 \to K^-K^+\mu^-\mu^+$	$D^0 \to \pi^-\pi^+\mu^-\mu^+$
$\mathcal{B}(D^0 \to h^-h^+\mu^-\mu^+)[10^{-8}]$	20.9 ± 1.6	83.5 ± 2.7
$\mathrm{N}^{1516}_{\mathrm{Comb}}$	6.8 ± 3.8	30.0 ± 9.1
$\mathrm{N}^{1516,\mathtt{norm}}_{\mathrm{Comb}}$	9.2 ± 6.0	8.7 ± 5.6
$\mathrm{N}^{1718}_{\mathrm{Comb}}$	52.0 ± 11.0	137.0 ± 18.0
$\mathrm{N}^{1718,\mathtt{norm}}_{\mathrm{Comb}}$	52.0 ± 13.0	20.7 ± 8.7
$\mathrm{N}^{1516}_{\mathrm{norm}}$	947.0 ± 28.0	826.0 ± 25.0
$\mathrm{N}^{1718}_{\mathrm{norm}}$	4469.0 ± 62.0	3942.0 ± 57.0
$\mathrm{N}^{1516}_{\mathrm{misID}}$	15.3 ± 4.5	121.0 ± 16.0
$\mathrm{N}^{1516,\mathtt{norm}}_{\mathrm{misID}}$	540.0 ± 28.0	226.0 ± 19.0
$\mathrm{N}^{1718}_{\mathrm{misID}}$	57.2 ± 9.4	631.0 ± 36.0
$\mathrm{N}^{1718,\mathtt{norm}}_{\mathrm{misID}}$	2616.0 ± 60.0	1229.0 ± 43.0
$\nu^{1516}_{\mathrm{sig}}$	-1.3 ± 1.0	-1.73 ± 0.69
$\nu^{1516}_{\mathrm{norm}}$	-0.2 ± 0.24	-0.71 ± 0.25
$\nu^{1718}_{\mathrm{sig}}$	0.02 ± 0.49	-1.52 ± 0.31
$\nu^{1718}_{\mathrm{norm}}$	-0.49 ± 0.1	-0.62 ± 0.11

Table 7.6: Summary of fit parameters, the yields and position parameters ν, for the reference decays obtained by the simultaneous fit. The position parameters ν are defined relative to the position, which is given by the peak position μ of the signal distribution obtained from simulations.

agreement between the fit model and data is observed.

The background yields observed in the LFV data samples are compatible with the predicted background yields during the optimisation; for details, see section 5.3.4. For the normalisation mode and $D^0 \to \pi^-\pi^+\mu^-\mu^+$ decays a mismatch between the in simulations and data obtained signal position can be observed. Due to the complex selection process, it is difficult to estimate the reason for this. However, a clear separation between signal and misID background is nevertheless possible.

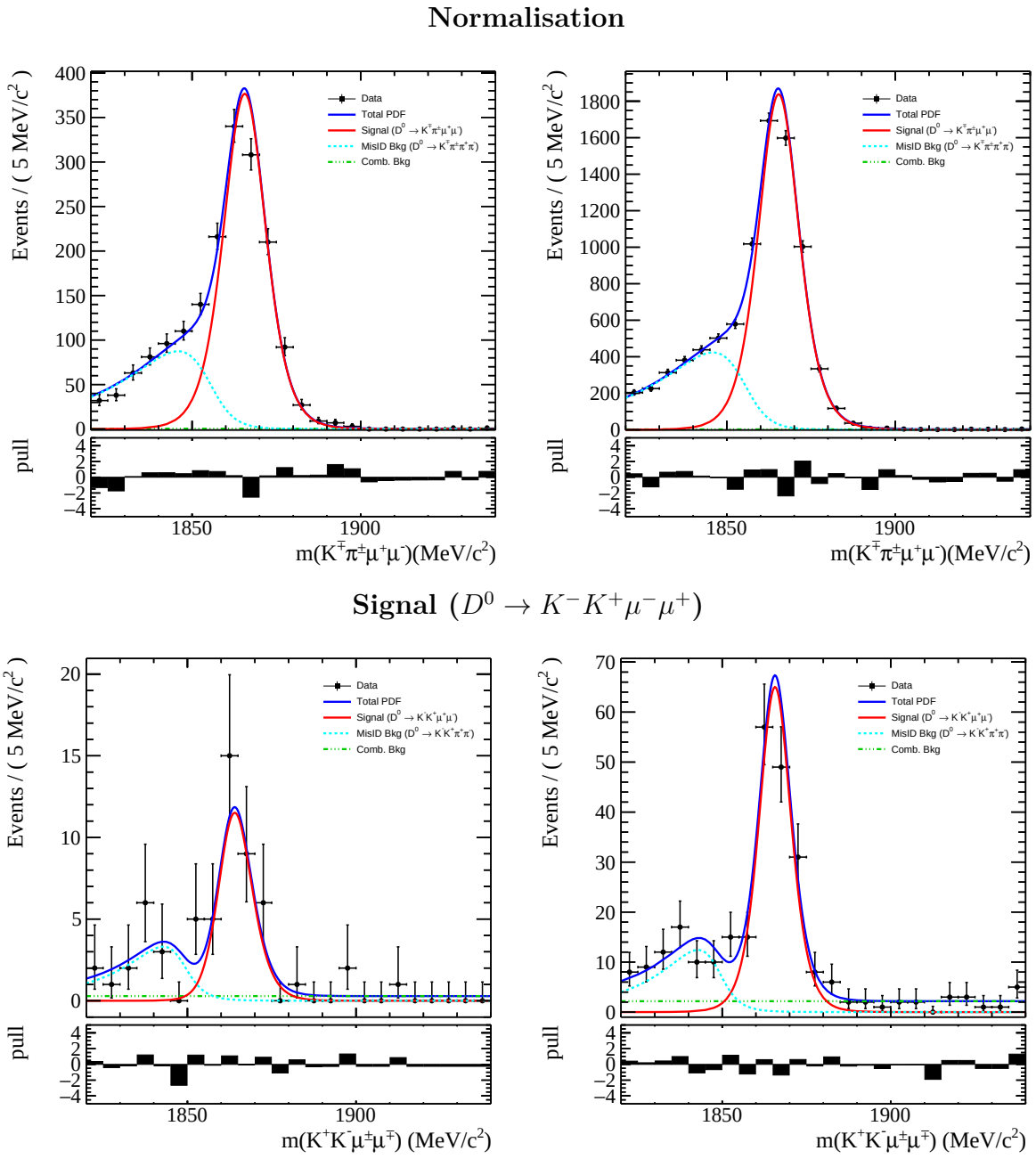


Figure 7.3: Simultaneous fit to the reconstructed invariant D^0 mass for $D^0 \to K^-\pi^+[\mu^+\mu^-]_{\rho^0/\omega}$ (up) and $D^0 \to K^-K^+\mu^-\mu^+$ (down) for data recorded in 2015/2016 (left) and 2017/2018 (right).

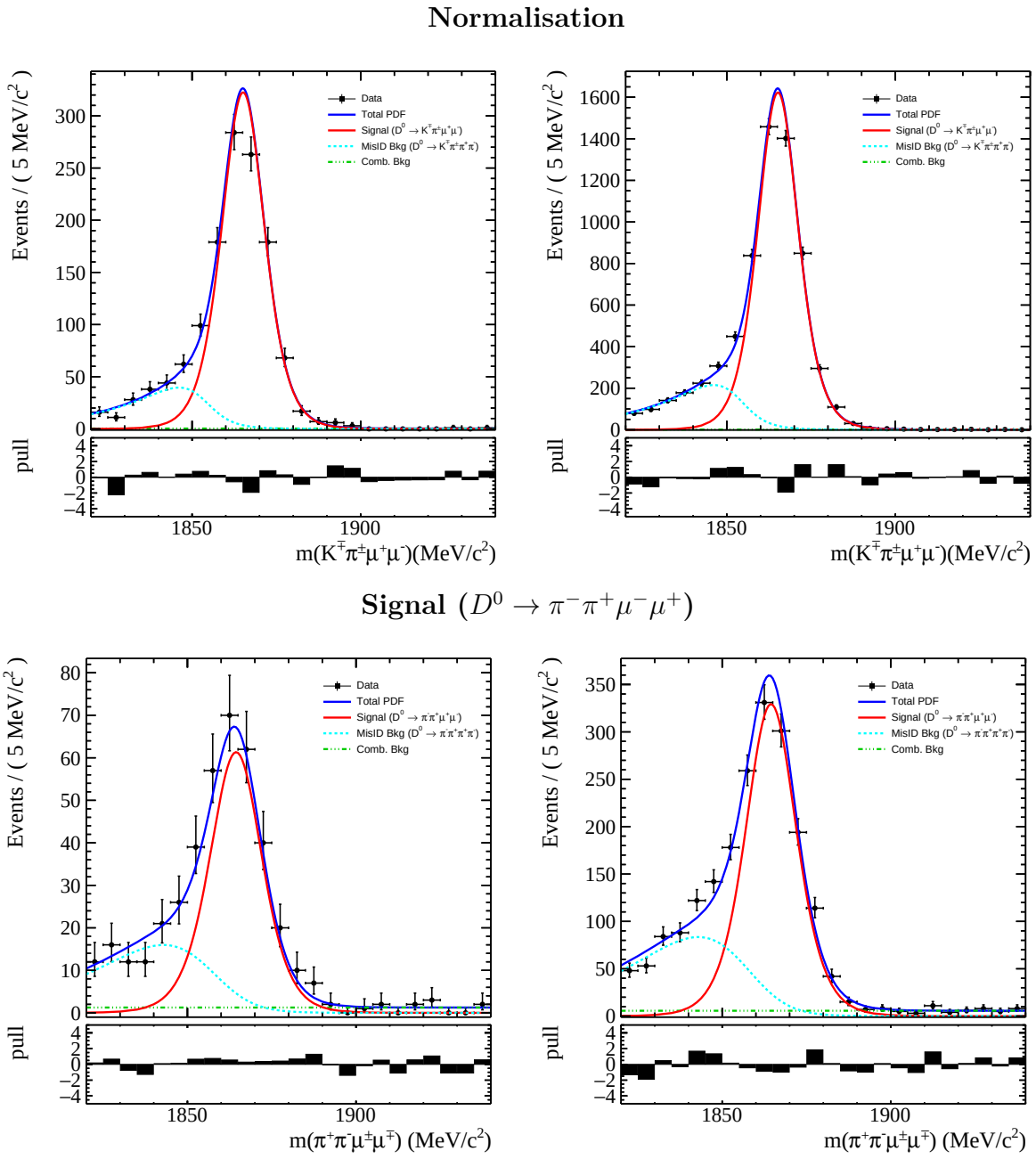
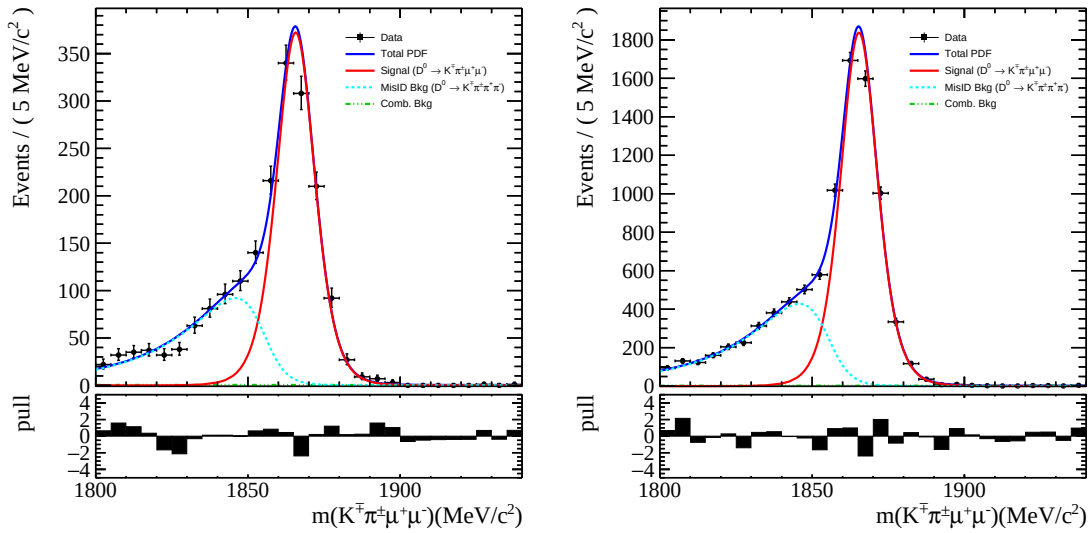


Figure 7.4: Simultaneous fit to the reconstructed invariant D^0 mass for $D^0 \to K^-\pi^+[\mu^+\mu^-]_{\rho^0/\omega}$ (up) and $D^0 \to \pi^-\pi^+\mu^-\mu^+$ (down) for data recorded in 2015/2016 (left) and 2017/2018 (right).



Signal ($D^0 \to K^-K^+\mu^\pm e^\mp$)

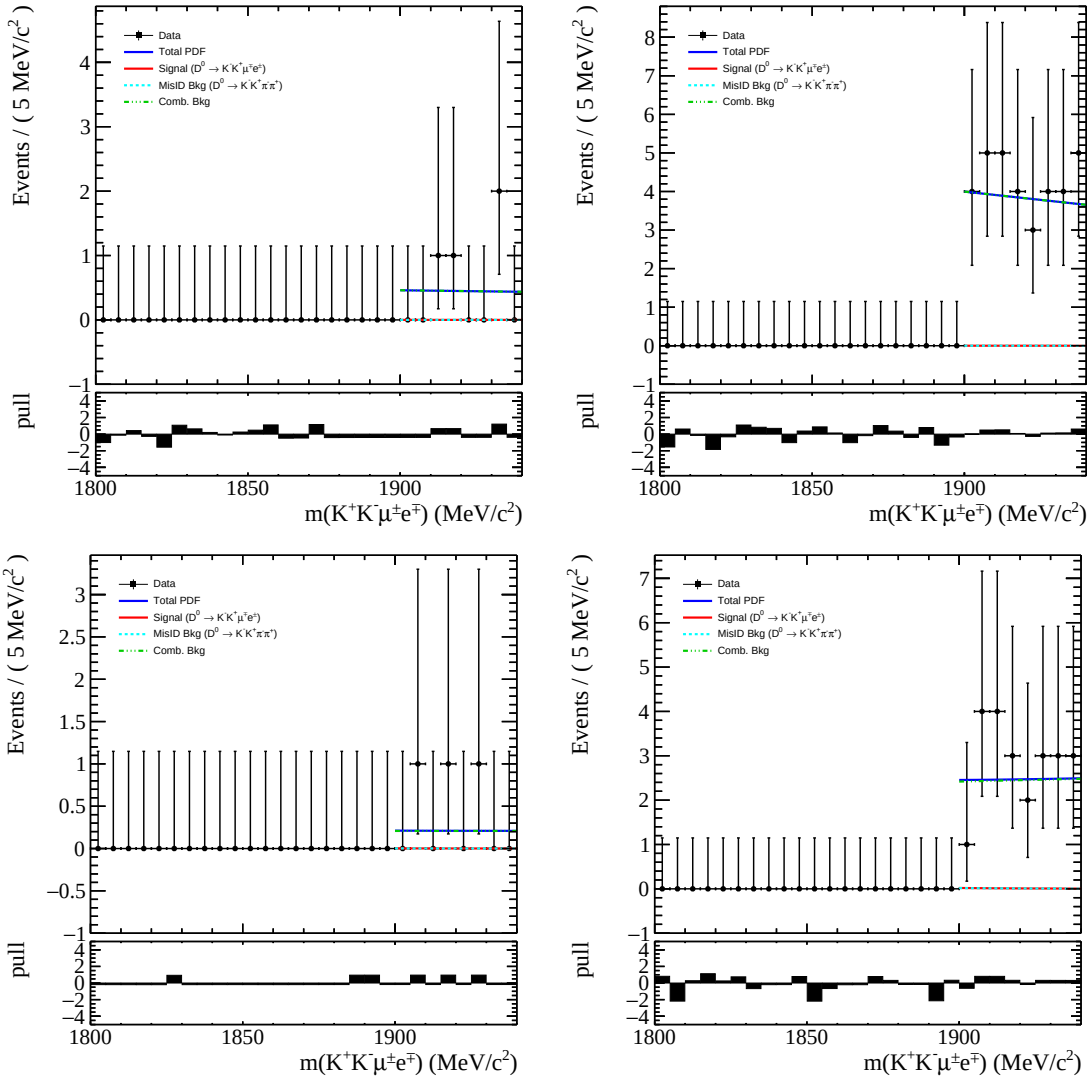


Figure 7.5: Blinded simultaneous fit to the reconstructed invariant D^0 mass for $D^0 \to K^-\pi^+[\mu^+\mu^-]_{\rho^0/\omega}$ (up) and $D^0 \to K^-K^+\mu^\pm e^\mp$ in `noBrem` (middle) and `Brem` (down) for data recorded in 2015/2016 (left) and 2017/2018 (right).

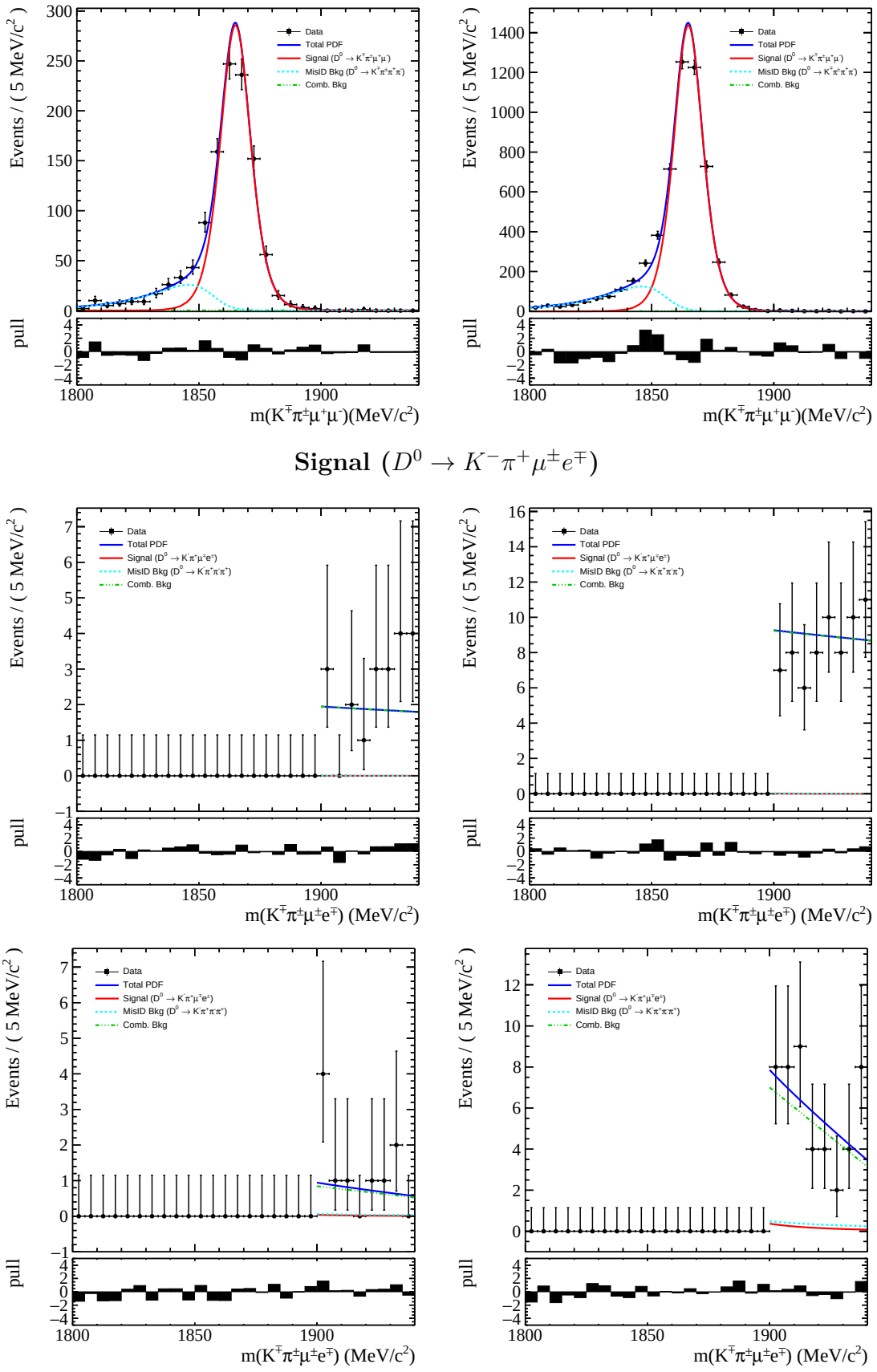
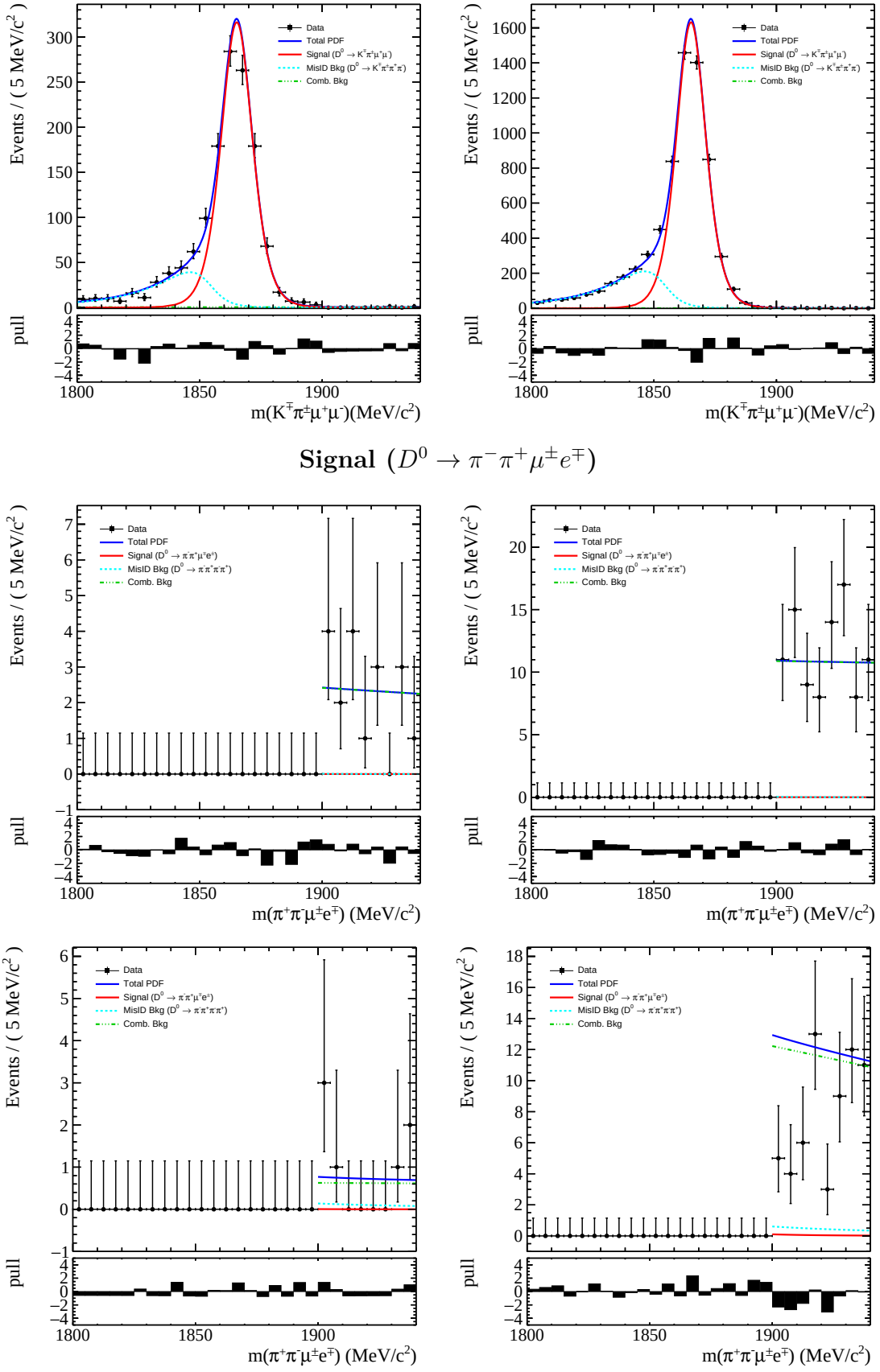


Figure 7.6: Blinded simultaneous fit to the reconstructed invariant D^0 mass for $D^0 \to K^-\pi^+[\mu^+\mu^-]_{\rho^0/\omega}$ (up) and $D^0 \to K^-\pi^+\mu^\pm e^\mp$ in `noBrem` (middle) and `Brem` (down) for data recorded in 2015/2016 (left) and 2017/2018 (right).



Signal ($D^0 \to \pi^-\pi^+\mu^\pm e^\mp$)

Figure 7.7: Blinded simultaneous fit to the reconstructed invariant D^0 mass for $D^0 \to K^-\pi^+[\mu^+\mu^-]_{\rho^0/\omega}$ (up) and $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ in `noBrem` (middle) and `Brem` (down) for data recorded in 2015/2016 (left) and 2017/2018 (right).

7.5 Validation of the fit procedure

Large samples of pseudo-experiments (1000) are generated for various branching fractions to test that the fit procedure returns an unbiased result of the true branching fraction and a correct uncertainty estimate on the measured value. The pseudo-experiments are created using the model used in the simultaneous fit and fixing all non-fixed parameters apart from the yield to the ones observed area in data. The signal channel, background and normalisation yields are taken from the simultaneous fit. The signal yield is calculated from the generated branching fraction using equation 7.4.1. All other parameters are fixed. During the generation process, the total number of events in the normalisation and signal channel, k, is drawn from a Poisson distribution:

$$f(k; N_{tot}) = \frac{N_{tot}^{k} e^{-N_{tot}}}{k!}, \tag{7.5.1}$$

where N_{tot} is the sum of the calculated signal yield and the background yields observed in data. The simultaneous fit is then applied to the pseudo datasets. To check whether the fit works, the pull distribution for the branching fraction, defined as

$$\texttt{pull} = \frac{\mathcal{B}(D^0 \to h^{(\prime)-} h^+ l^{(\prime)-} l^+)^{fit} - \mathcal{B}(D^0 \to h^{(\prime)-} h^+ l^{(\prime)-} l^+)^{Gen}}{\sigma^{fit}}, \tag{7.5.2}$$

is used. A mean of zero is expected for an unbiased estimate of the branching fraction. A width of one of the pull distribution is expected for the correct error estimation. The pull mean and width of the 1000 pseudo-experiments are estimated by a fit of a normal distribution to the distribution of obtained branching ratios. In figure 7.8, the pull distributions for the reference mode corresponding to the measured branching ratios, $\mathcal{B}(D^0 \to \pi^-\pi^+\mu^-\mu^+)^{Gen} = 8.35 \times 10^{-7}$ and $\mathcal{B}(D^0 \to K^-K^+\mu^-\mu^+)^{Gen} = 2.09 \times 10^{-7}$, overlaid with a normal distribution's fit projection are shown.

The widths and means are compatible with one and zero, respectively, for the two reference modes $D^0 \to K^-K^+\mu^-\mu^+$ and $D^0 \to \pi^-\pi^+\mu^-\mu^+$, confirming that the fit procedure itself does not bias the measurement and the errors represent the statistical fluctuation of the sample. For the signal modes, a scan for various branching fractions is performed. The range of the studied branching fractions is chosen from no observation up to the order of magnitude not yet excluded by BaBar. In figure 7.9, the means and widths from the pull distributions for the signal mods for different generated branching fractions are compared. The widths and means are compatible with one and zero, respectively, validating that the fitting procedure works, is unbiased, and yields the expected error estimate.

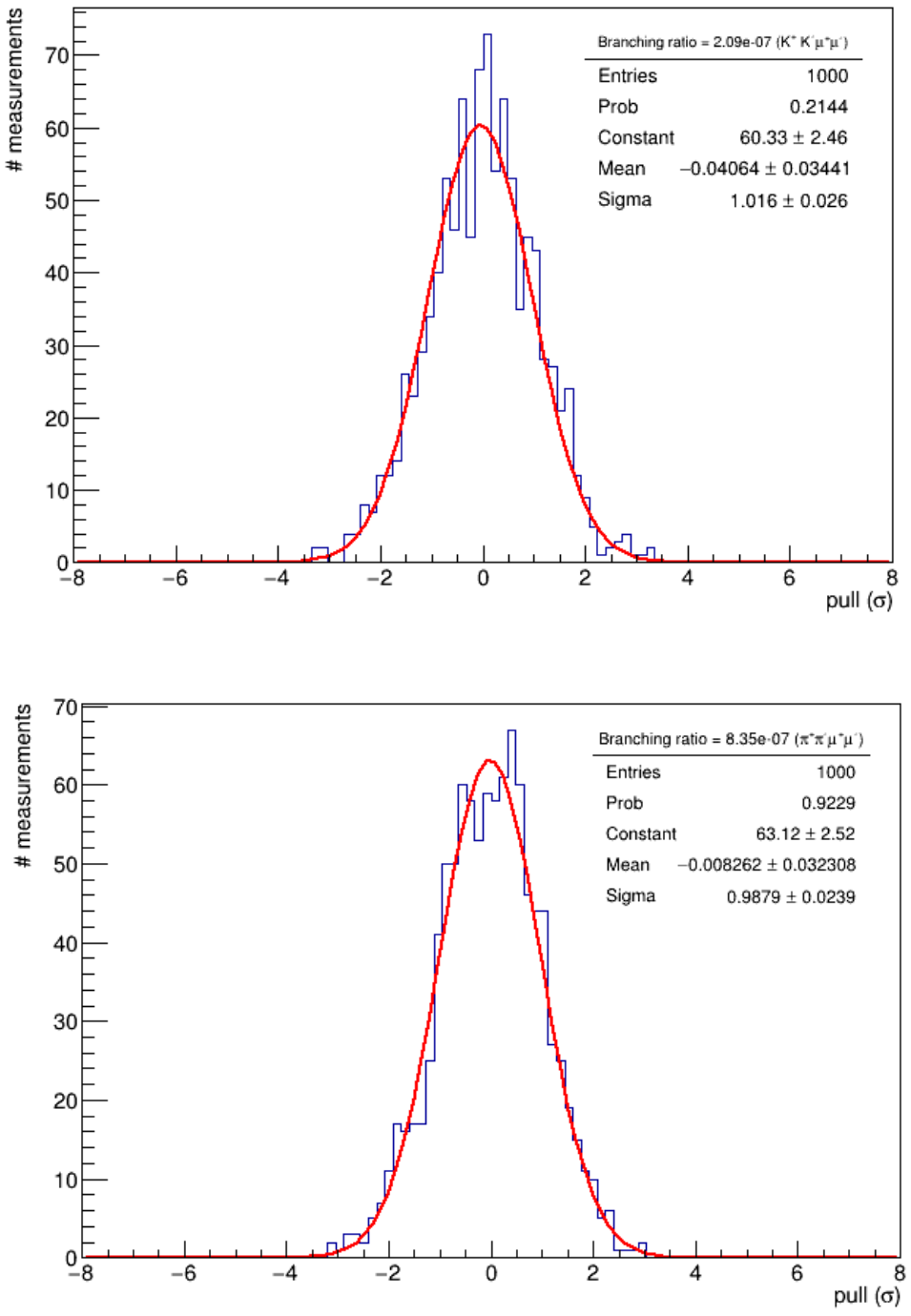


Figure 7.8: Pull distribution of the branching fraction (blue) for the two reference modes, $D^0 \to K^-K^+\mu^-\mu^+$ (up) and $D^0 \to \pi^-\pi^+\mu^-\mu^+$ (down), generated with the observed branching ratio. The pull distribution is overlaid with the fit projection of a normal distribution (red).

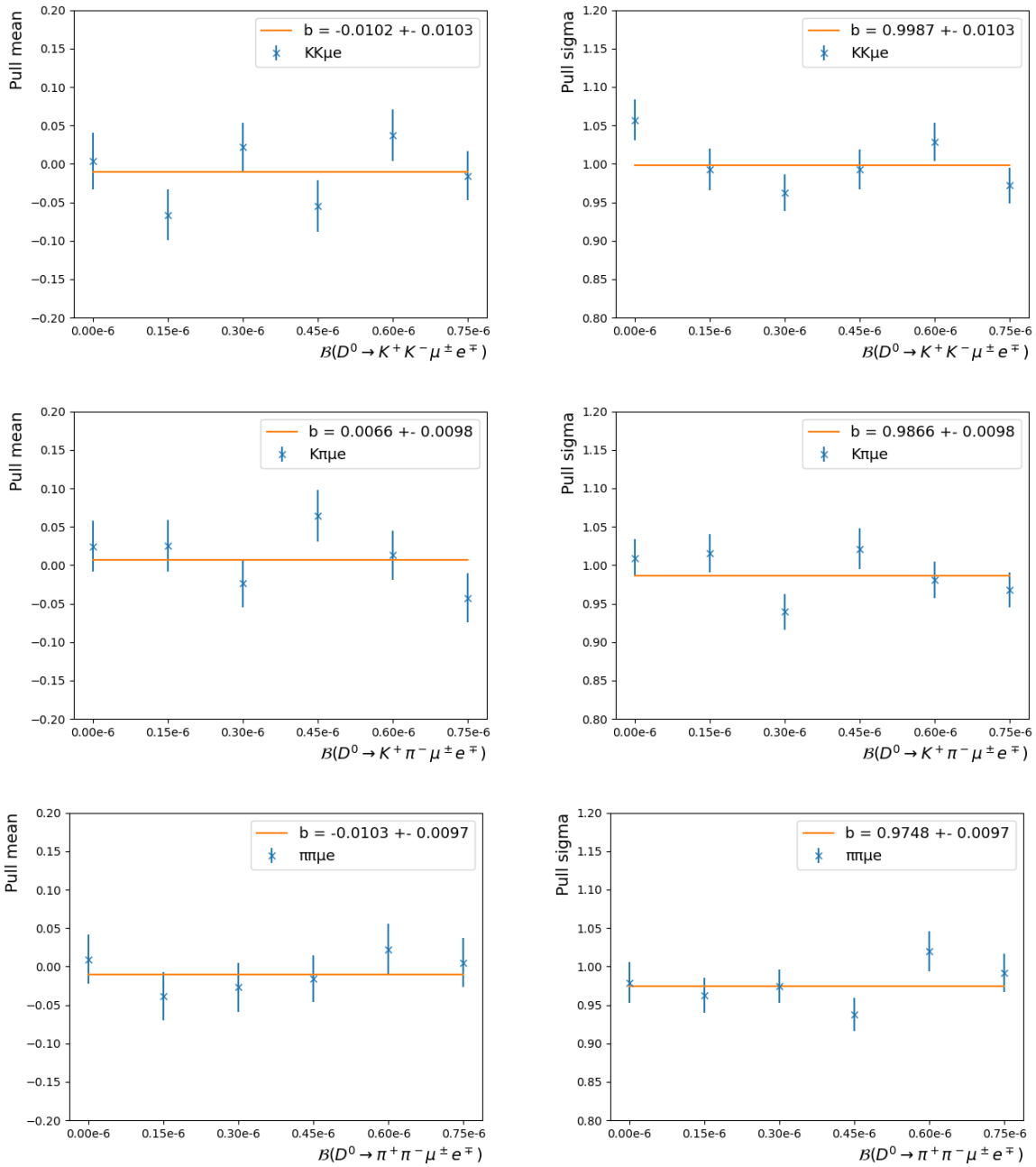


Figure 7.9: Mean of the pull distribution for different branching fractions overlaid with a linear fit in orange (left column). The standard deviation of the Gaussian fit to the pull distribution overlaid with a linear fit in orange (right column). The scan results for $D^0 \to K^-K^+\mu^\pm e^\mp$ (up), $D^0 \to K^-\pi^+\mu^\pm e^\mp$ (middle) and $D^0 \to K^-K^+\mu^\pm e^\mp$ (down) are shown.

7.6 Conclusion

The following branching ratios are measured for the reference decays:

$$\mathcal{B}(D^0 \to K^-K^+\mu^-\mu^+) = (2.09 \pm 0.16) \times 10^{-7} \tag{7.6.1}$$

$$\mathcal{B}(D^0 \to \pi^-\pi^+\mu^-\mu^+) = (8.35 \pm 0.27) \times 10^{-7} \tag{7.6.2}$$

The quoted uncertainty is only related to the fit accounting for the statistics of the data sample. The results for the same branching ratios measured with the Run 1 dataset [44] are:

$$\mathcal{B}(D^0 \to K^-K^+\mu^-\mu^+) = (1.54 \pm 0.27 \pm 0.09 \pm 0.16) \times 10^{-7} \tag{7.6.3}$$

$$\mathcal{B}(D^0 \to \pi^-\pi^+\mu^-\mu^+) = (9.64 \pm 0.48 \pm 0.51 \pm 0.97) \times 10^{-7} \tag{7.6.4}$$

The first quoted uncertainty corresponds to the statistical uncertainty on the signal dataset. The second covers all systematic uncertainty apart from the uncertainty of the normalisation branching fraction, which is quoted as the third uncertainty. The statistical uncertainty on the branching ratio introduced by the dataset size of the normalisation mode of $0.25 \times 10^{-7} (0.04 \times 10^{-7})$ [44] for $D^0 \to \pi^-\pi^+\mu^-\mu^+$ ($D^0 \to K^-K^+\mu^-\mu^+$) decays is reported as part of the systematic uncertainty. To compare the statistical uncertainties for both measurements introduced by the size of the data sample, the statistical uncertainty of the signal and normalisation dataset are combined. Both uncertainties are added in quadrature. The following branching ratio with their statistical uncertainties can be quoted for Run 1:

$$\mathcal{B}(D^0 \to K^-K^+\mu^-\mu^+) = (1.54 \pm 0.27) \times 10^{-7} \tag{7.6.5}$$

$$\mathcal{B}(D^0 \to \pi^-\pi^+\mu^-\mu^+) = (9.64 \pm 0.54) \times 10^{-7} \tag{7.6.6}$$

Comparing these values with the Run 2 results shows that the statistical uncertainties could be reduced by approximately a factor of 2.0 for $D^0 \to \pi^-\pi^+\mu^-\mu^+$ decays and a factor of 1.7 for $D^0 \to K^-K^+\mu^-\mu^+$ decays. By scaling the luminosity, which does not account for changes in the performance of the detector, a factor of 1.7 is expected.

For the LFV signal decays only the statistical uncertainties of the branching ratio can be given, due to blinding:

$$\mathcal{B}(D^0 \to K^-K^+\mu^\pm e^\mp) = (x.xx \pm 2.5) \times 10^{-8} \tag{7.6.7}$$

$$\mathcal{B}(D^0 \to K^-\pi^+\mu^\pm e^\mp) = (x.xx \pm 4.9) \times 10^{-8} \tag{7.6.8}$$

$$\mathcal{B}(D^0 \to \pi^-\pi^+\mu^\pm e^\mp) = (x.xx \pm 3.3) \times 10^{-8} \tag{7.6.9}$$

The central value itself would require unblinding the measurement[6]. Compared to the result by BaBar, this corresponds to an improvement by two orders of magnitude. However, uncertainties introduced by the measurement procedure are not covered at the moment and will be detailed in chapter 8. Noticeable is also that the statistical uncertainties, mainly because of the electron, are only slightly worse, even though the selection is less efficient for the LFV signal decays. Given that the selection was tuned with a main focus on the LFV decay modes, this is not necessarily unexpected.

To be able to test a wider range of new physics scenarios, the data is split into two samples. The samples are split by the charge constellation between the muon and slow pion. This way, the charge of the muon and electron are correlated to the charge of the pion and consequent flavour of the D^0. Given that the charge conjugated decay products are also taken into account, the efficiency is assumed to be the same. The fits and yield parameters are given in appendix F for completeness. Apart from splitting the data sample in two within their uncertainties similar sized subsamples, no differences to the original fits are observed. The statistical uncertainties of the branching fractions for the sample where the muon and slow pion have the same charge[7] are given by:

$$\mathcal{B}(D^0 \to K^- K^+ \mu^+ e^-) = (x.xx \pm 2.2) \times 10^{-8} \tag{7.6.10}$$

$$\mathcal{B}(D^0 \to K^- \pi^+ \mu^+ e^-) = (x.xx \pm 4.2) \times 10^{-8} \tag{7.6.11}$$

$$\mathcal{B}(D^0 \to \pi^- \pi^+ \mu^+ e^-) = (x.xx \pm 2.3) \times 10^{-8} \tag{7.6.12}$$

The statistical uncertainties of the branching fractions for the sample where the muon and slow pion have the opposite charge are given by:

$$\mathcal{B}(D^0 \to K^- K^+ \mu^- e^+) = (x.xx \pm 1.3) \times 10^{-8} \tag{7.6.13}$$

$$\mathcal{B}(D^0 \to K^- \pi^+ \mu^- e^+) = (x.xx \pm 2.5) \times 10^{-8} \tag{7.6.14}$$

$$\mathcal{B}(D^0 \to \pi^- \pi^+ \mu^- e^+) = (x.xx \pm 2.3) \times 10^{-8} \tag{7.6.15}$$

Due to the decrease of events in the subsamples the absolute uncertainty decreases. However, the relative uncertainty increases, assuming that the signal branching fraction should be halved in the subsamples. When the statistical uncertainties from both subsamples are combined in quadrature, disregarding the effects of rounding, the same uncertainties that were derived from the fit to the combined data sample are obtained.

[6]For this, prior approval by the collaboration is needed.

[7]The D^0 meson and the muon have opposite flavour.

Chapter 8

Systematic uncertainties

This chapter covers the study of systematic effects possibly affecting the branching fraction measurement. The systematic uncertainties are separated into two categories: those associated with determining the yield and those associated with estimating the efficiency. Each systematic uncertainty is covered in a dedicated section in this chapter. The leading order systematic uncertainties are summarised in table 8.9, page 139, for the LFV decay modes and table 8.10, page 140, for the dimuon decay modes.

The following sources of systematic uncertainties are studied in the context of the yield determination:

- Parameterisation of the signal component
- Assumptions of the combinatorial background shape
- Parameterisation of the misID background

All three systematic uncertainties are associated with different components of the signal fit and are treated individually. These systematic uncertainties are estimated by the generation of toy experiments to test the impact of underlying assumptions in the parameterisation of the individual fit component. Possible neglected additional background sources are discussed in section 8.1.

The following systematic uncertainties are the dominant uncertainties connected to the efficiency estimation:

- Tracking efficiency
- Particle Identification

- Trigger efficiency
- Model assumptions for simulated samples
- Limited statistics of the simulation samples

All uncertainties connected to the sample size of the simulation or data sample are evaluated using a bootstrapping approach. Bootstrapping refers to any test or metric using random sampling with replacement. This means, that in the resampling process random candidates from the original sample are drawn while allowing multiple occurrences of the same elements. The measurement is then repeated multiple times with the resampled datasets, allowing the estimation of the observable's statistical distribution [112]. This is equivalent to an analytic approach apart from the fact that it is easier to account for the various binning schemes used in this analysis.

All leading systematic uncertainties considered for the limit setting are summarised in table 8.9. Efficiency uncertainties are considered separately for each data category, similar to the individual efficiency ratios in the branching ratio calculation. In the summary table, only the uncertainty of the integrated data sample is quoted. The individual values, if calculated, are reported in their corresponding subsection.

8.1 Discussion of neglected background sources

The strong electron and muon PID selection suppresses other possible, in the SM allowed, peaking background sources in the mass signal window. In the upper row of figure 8.1, the normalised mass shapes for the listed background events reconstructed under the $D^0 \to h^{(\prime)-}h^+\mu^\pm e^\mp$ mass hypothesis are shown as obtained from simplified simulations with RapidSim [113]. The mass hypothesis is always applied in the same pattern as the LFV decay given in the legend. For completeness, the double misidentified decays included in the fit, $\pi^+\pi^- \to \mu^+e^-$ misidentification, are included. In the lower row of figure 8.1, decays are shown with a single muon (electron) being misidentified as an electron (muon). Exemplary, the distribution for $D^0 \to K^-K^+\mu^\pm e^\mp$ and $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ decays are shown. The distributions for $D^0 \to K^-K^+\mu^-\mu^+$ and $D^0 \to \pi^-\pi^+\mu^-\mu^+$ decays are expected to be similar, while the distribution for $D^0 \to K^-\pi^+\mu^\pm e^\mp$ decays is a mixture between the two distribution with misidentified decays distributed in the lower and upper reconstructed mass in the figure relative to the signal peak. Important to note in the figures is the closeness and overlap between the double misidentified hadronic decays, $\pi^\mp\pi^\pm \to \mu^\mp e^\pm$ misidentification, with the LFV signal distribution. The same is expected

and during the yield determination observed for the dimuon signal modes. Secondly, the distribution for the single misidentified dielectron and dimuon decays, $e^{\mp} \to \mu^{\mp}$ or $e^{\mp} \to \mu^{\mp}$ misidentification, is directly below the reconstructed mass distribution of the LFV decay with the matching hadron species composition. Also, here the same is true for the dimuon reference decays, with the difference that dielectron decays would need to be double misidentified, and the single misidentified LFV decay modes are not allowed within the SM. These background decays could lead to a fake signal during the yield determination and are, therefore, specifically targeted in this study.

For this study, a conservative misidentification rate of 5% is assumed for pions to be misidentified as kaons and vice versa, and a rate of 3% for pions to be misidentified as leptons. The electron muon misidentification rate is assumed to be less than 1%. In the following, possible decays expected to peak in the studied reconstructed D^0 mass window are listed with their expected yields:

- $D^0 \to h^{(\prime)-}h^+\pi^-\pi^+$: (included in the fit)
  $\mathcal{B}(D^0 \to h^{(\prime)-}h^+\pi^-\pi^+) \sim \mathcal{O}(10^{-3} - 10^{-2})$ [35]$\to N_{\text{exp}} \sim \mathcal{O}(100 - 1000)$
- $D^0 \to h^-h^+\mu^-\mu^+$:
  $\mathcal{B}(D^0 \to h^-h^+\mu^-\mu^+) \sim \mathcal{O}(10^{-8} - 10^{-7})$ [35]$\to N_{\text{exp}} \sim \mathcal{O}(0.1 - 1)$
- $D^0 \to h^{(\prime)-}h^+e^-e^+$:
  $\mathcal{B}(D^0 \to h^{(\prime)-}h^+e^-e^+) \sim \mathcal{O}(10^{-8} - 10^{-7})$ [35, 52]$\to N_{\text{exp}} \sim \mathcal{O}(0.1 - 1)$

The triple misidentified $D^0 \to h^{(\prime)-}h^+\pi^-\pi^+$ decays, with an additional pion-kaon misidentification, are expected to be further suppressed by at least two orders of magnitude, compared to the double misidentified $D^0 \to h^{(\prime)-}h^+\pi^-\pi^+$ decays, and peak outside the signal window. Contributions from both $D^0 \to h^{(\prime)-}h^+e^-e^+$ and $D^0 \to h^-h^+\mu^-\mu^+$ decays are at least one order of magnitude below the expected sensitivity of this measurement, $\mathcal{O}(10^{-8} - 10^{-7})$. BaBar did the first observation of $D^0 \to K^-\pi^+[e^+e^-]_{\rho^0/\omega}$ decays, while $D^0 \to \pi^+\pi^-e^+e^-$ and $D^0 \to K^+K^-e^+e^-$ are not observed yet. Therefore, the expected values are quoted [52][1].

For the reference decay channel, $D^0 \to K^-K^+\mu^-\mu^+$ and $D^0 \to \pi^-\pi^+\mu^-\mu^+$, the above list is reduced to the misidentified hadronic decays of the form $D^0 \to h^{(\prime)-}h^+\pi^-\pi^+$. The LFV decay modes are forbidden within the SM, and decays of the form $D^0 \to h^{(\prime)-}h^+e^-e^+$

[1] Under the SM assumption of lepton flavour universality, the branching fractions for $D^0 \to h^{(\prime)-}h^+e^-e^+$ and $D^0 \to h^-h^+\mu^-\mu^+$ are expected to be similar in magnitude.

would require that both electrons be misidentified as muons, reducing the expected background yield by an additional order of magnitude.

The number of expected background candidates for $D^0 \to h^{(\prime)-}h^+\pi^-\pi^+$ decays is also

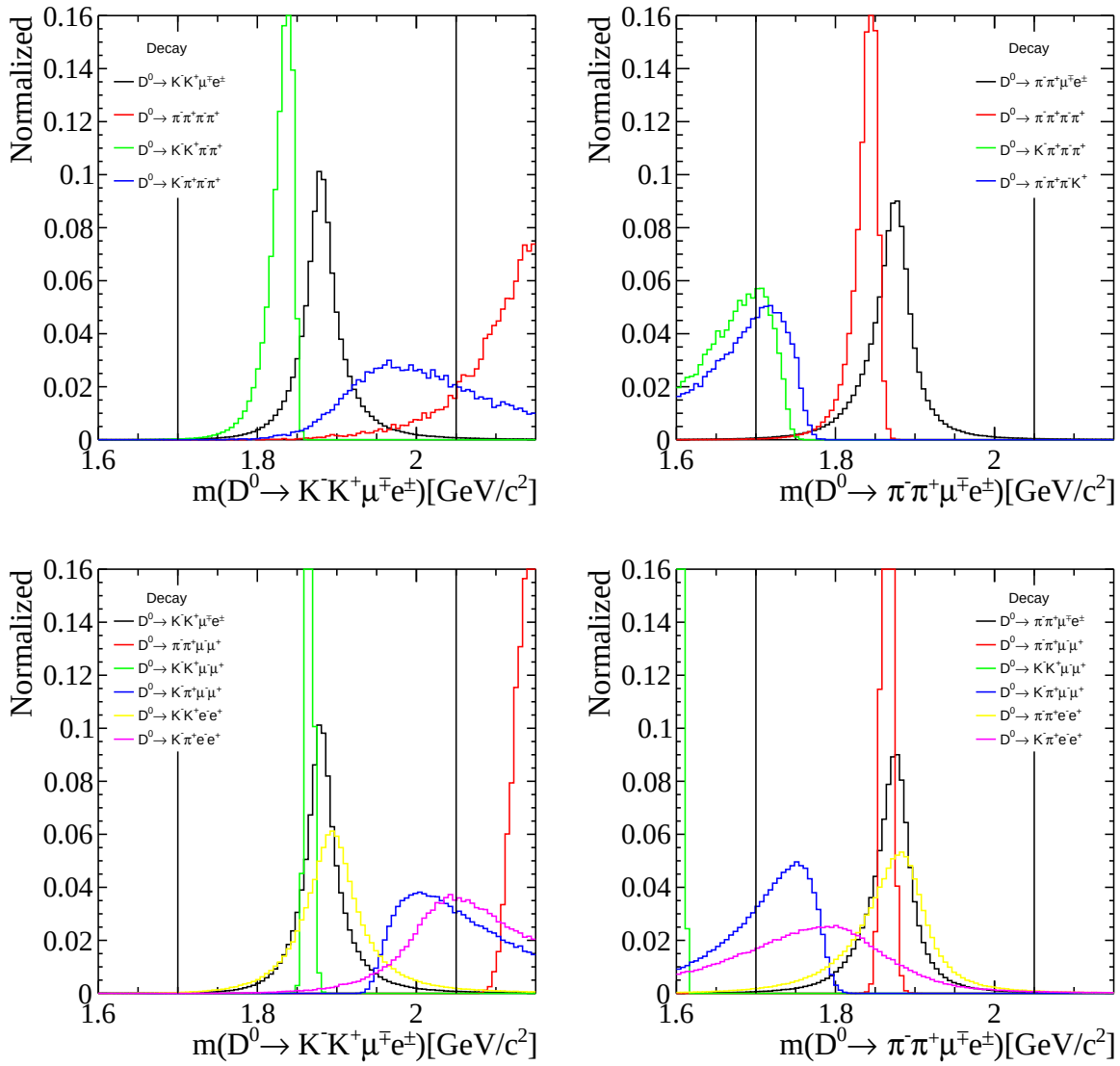


Figure 8.1: Normalised potential peaking background mass shapes estimated with RapidSim [113] for $D^0 \to K^+K^-\mu^\pm e^\mp$ (left) and $D^0 \to \pi^+\pi^-\mu^\pm e^\mp$ candidates (right). In addition, the LFV signal is shown. The upper row shows decays where hadrons are misidentified as leptons. The lower row shows decays where a muon is misidentified as an electron and vice versa from $D^0 \to h^{(\prime)-}h^+e^-e^+$ and $D^0 \to h^-h^+\mu^-\mu^+$. All candidates are reconstructed using the signal mass hypothesis. The masses are assigned to the background decay daughters according to the order of the signal decay daughters. The black lines mark the signal window applied during the trigger stage. The signal window studied in the fit is defined between 1.8 GeV/c^2 and 1.95 GeV/c^2.

observed in data. For details on the yield determination, see chapter 7. All other possible background sources are expected to either not peak within the studied reconstructed D^0 mass window and are therefore described by the linear combinatorial background function in the fit, or if they peak within the reconstructed D^0 mass window, the expected yield is too small to influence the yield determination. If an candidate would pass the selection, this candidate would be absorbed into the description of the combinatorial background. The uncertainty on the signal yield in the fit is approximately ± 10 signal candidates, making the fit insensitive to single candidates. The applied fit is, therefore, sufficient to describe all background sources. No systematic uncertainty needs to be assigned.

8.2 Parameterisation of the signal

The true shape of the LFV signal and its parameterisation are unknown. Therefore, multiple pseudo-experiments with differently parameterised signal shapes are created. The largest uncertainty is expected to originate from the signal distribution overlapping with the misID background distribution. This makes it difficult to separate misID background and signal candidates in this region. The effect is studied in detail in section 8.4. Similar to the study in section 8.4 where the misID background description is varied, also here mistakenly signal candidates are absorbed by the misID distribution (or vice versa) and are consequently miscounted. This effect is also later observed in this section, in figure 8.6, seen by the data points significantly deviating from zero.

The parameterisation of the signal shape is studied by creating pseudo-experiments with a double-sided crystal ball [111] and a Bukin [114] distribution. In the case of no observation, the parameterisation itself is not expected to contribute to the systematic uncertainty. This may change in the case of an observation due to the overlap and consequent possible miscounting. The double-sided crystal ball function consists of a Gauss distribution with individual widths for the left and right side of the peak position, which goes over into a power law, individually for both sides, in the tails of the distribution. The distribution is

defined as follows:

$$f(m; m_0, \sigma, \alpha_L, n_L, \alpha_R, n_R) = \begin{cases} A_L \cdot (B_L - \frac{m-m_0}{\sigma_L})^{-n_L}, & \text{for } \frac{m-m_0}{\sigma_L} < -\alpha_L \\ \exp\left(-\frac{1}{2} \cdot \left[\frac{m-m_0}{\sigma_L}\right]^2\right), & \text{for } \frac{m-m_0}{\sigma_L} \leq 0 \\ \exp\left(-\frac{1}{2} \cdot \left[\frac{m-m_0}{\sigma_R}\right]^2\right), & \text{for } \frac{m-m_0}{\sigma_R} \leq \alpha_R \\ A_R \cdot (B_R + \frac{m-m_0}{\sigma_R})^{-n_R}, & \text{for } \frac{m-m_0}{\sigma_R} > \alpha_R \end{cases} \tag{8.2.1}$$

where m_0 defines the location and $\sigma_{L,R}$ the width of the Gaussian component. $\alpha_{L,R}$ defines the location of transition to a power law in units of the standard deviation and $n_{L,R}$ is the exponent of the power-law tail. The shape is individually defined for the left and right side of the Gaussian peak position m_0 notated by the indices L and R. The normalisation factors, also introduced to ensure a continuous distribution, are noted as $A_{L,R}$ and $B_{L,R}$ and are given by:

$$A_{L,R} = \left(\frac{n_{L,R}}{|\alpha_{L,R}|}\right)^{n_{L,R}} \cdot \exp\left(-\frac{|\alpha_{L,R}|^2}{2}\right),$$
$$B_{L,R} = \frac{n_{L,R}}{|\alpha_{L,R}|} - |\alpha_{L,R}|\,.$$

The second function used is a Bukin distribution, which is derived by the convolution of a Gauss distribution with an exponential function. By this convolution, an ambiguity arises as a compact distribution (Gauss distribution) is convoluted with a diverging function (exponential function). This problem is avoided by the use of an asymptotic expansion leading to a slightly more lengthy expression of the then distribution function [114]. The Bukin distribution is defined as follows, omitting normalisation terms for better readability:

$$f(m; m_0, \sigma, \xi, \rho_L, \rho_R) = \begin{cases} \exp\left[\rho_L \frac{(x-x_1)^2}{(m_0-x_1)^2} - c_1 + \frac{\xi\gamma(\xi)(x-x_1)c_0}{2\sigma(\gamma(\xi)-\xi)^2 \ln(\gamma(\xi)+\xi)}\right], & \text{for } x < x_1 \\ \exp\left[-\frac{c_1 \ln\left(1+\frac{4\xi\gamma(\xi)(x-m_0)}{(c_0\sigma)}\right)^2}{\ln(1+2\xi\,(\xi-\gamma(\xi)))^2}\right], & \text{for } x < x_2 \\ \exp\left[\rho_R \frac{(x-x_2)^2}{(m_0-x_2)^2} - c_1 + \frac{\xi\gamma(\xi)(x-x_2)c_0}{2\sigma(\gamma(\xi)+\xi)^2 \ln(\gamma(\xi)+\xi)}\right] & \text{otherwise} \end{cases} \tag{8.2.2}$$

where m_0 is the peak position, σ the peak width as FWHM divided by $c_0 = 2\sqrt{2\ln 2} \approx 2.35$, ξ a value describing the peak asymmetry, usually smaller than one, and $\rho_{L,R}$ the left and

right tail parameters.

$$
\begin{aligned}
c_1 &= \ln 2 \\
x_{1,2} &= m_0 + \frac{1}{2}\,\sigma c_1 \left(\frac{\xi}{\sqrt{\xi+1}} \mp 1 \right) \\
\gamma(\xi) &= \sqrt{\xi^2+1}
\end{aligned}
$$

The parameters of both distributions are fitted and fixed to the distributions in the simulation samples. An example of both distributions, obtained from $D^0 \to K^- K^+ \mu^\pm e^\mp$ decay simulation data with corresponding error bands can be found in figure 8.2.

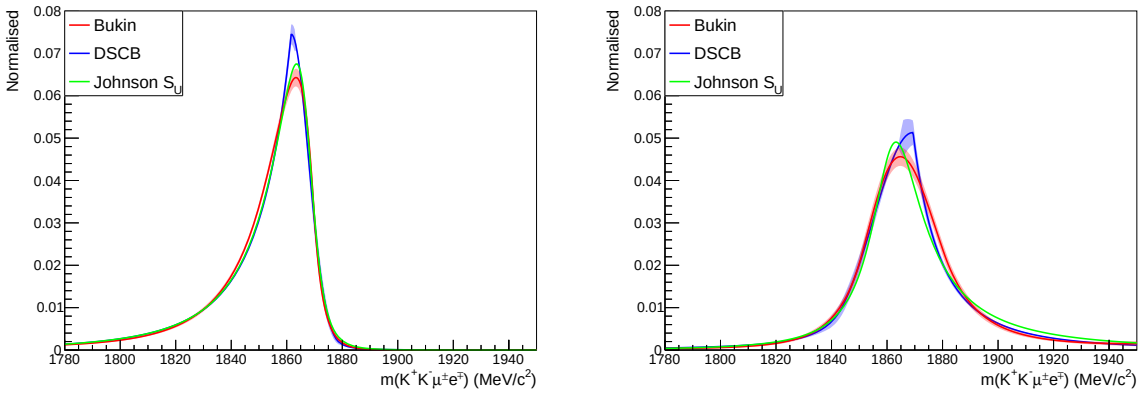


Figure 8.2: Exemplary the different signal distribution, given by a Bukin distribution (red) and a double-sided crystal ball function (DSCB; blue) shapes used in the generation of the pseudo-experiments compared to the default fit function, given by a Johnson S_U distribution (green), for $D^0 \to K^- K^+ \mu^\pm e^\mp$ signal candidates. All distributions are obtained from simulation. The left (right) plot shows signal candidates without (with) added bremstrahlung during the reconstruction.

The crystal ball function has a more pronounced peak, while the Bukin distribution emphasises the tails of the distribution, especially the ECAL resolution component in the `Brem` category. The pseudo-data is created by drawing random candidates with a given signal branching ratio assumption. The resulting distribution is then fitted by the default fit with the signal described by a Johnson S_U. From this fit, the branching ratio is then estimated. Repeating this a thousand times allows the calculation of the pull distribution, introduced in section 7.5. The pull is defined by

$$
\texttt{pull} = \frac{\mathcal{B}_{meas.}(D^0 \to h^{(\prime)-} h^+ l^{(\prime)-} l^+) - \mathcal{B}_{gen.}(D^0 \to h^{(\prime)-} h^+ l^{(\prime)-} l^+)}{\sigma_{meas.}}. \tag{8.2.3}
$$

The pull distributions under the assumption of lepton flavour conservation, $\mathcal{B}(D^0 \to h^{(\prime)-} h^+ \mu^\pm e^\mp)_{gen.} = 0$, are reported in figure 8.3, separately for the Bukin distribution and

crystal ball function. The mean is compatible with zero. No shift or bias is observed in the distribution. Therefore, no systematic uncertainty will be assigned in the case of no observed signal. This was also expected, as the statistical error obtained from the branching ratio fit should sufficiently describe all statistical fluctuations.

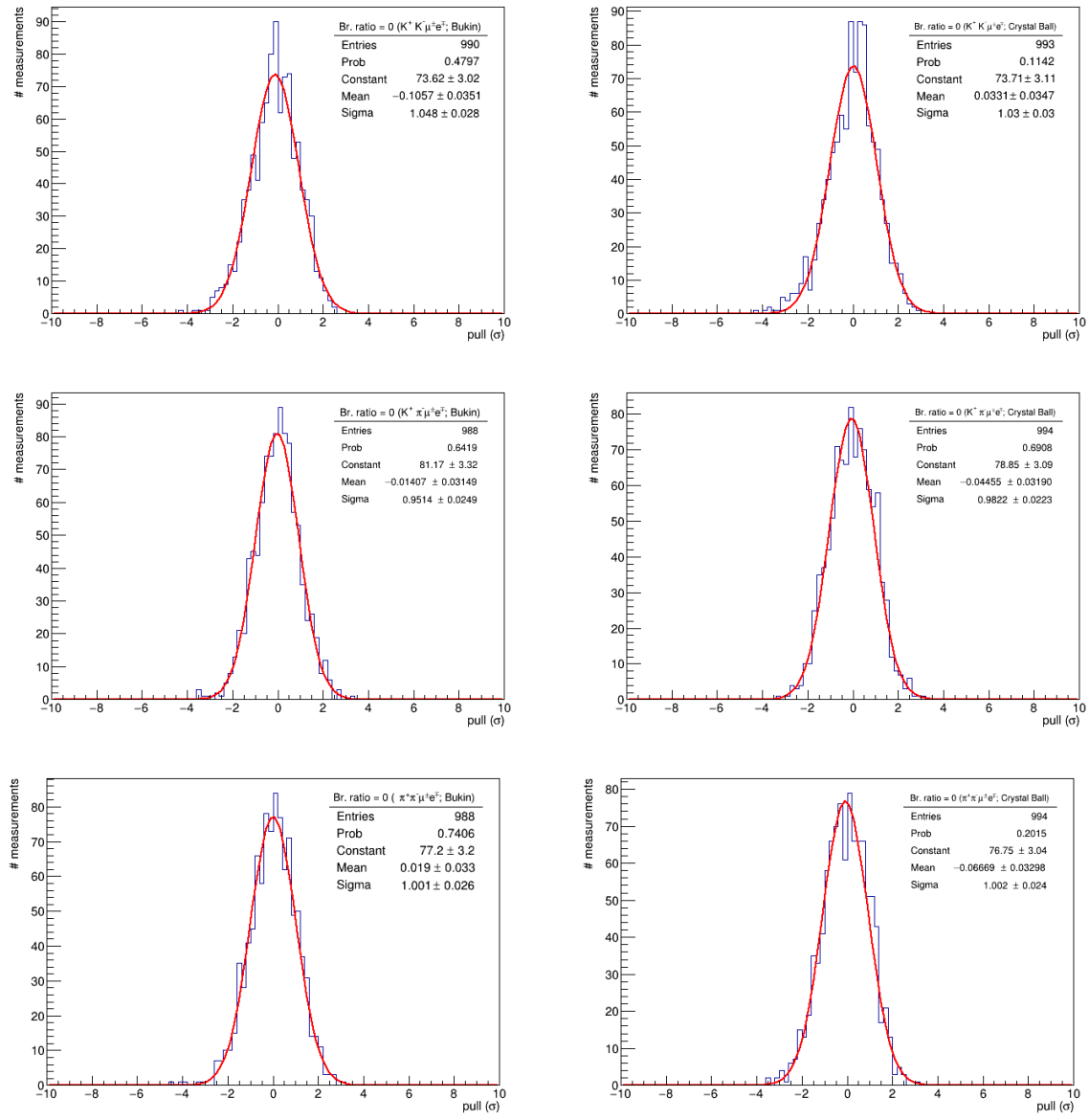


Figure 8.3: Pull distribution for the three different LFV decay modes, $D^0 \to K^-K^+\mu^\pm e^\mp$ (upper row), $D^0 \to K^-\pi^+\mu^\pm e^\mp$ (middle row) and $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ (lower row), evaluated by the default fit procedure, using a Johnson S_U, for different generated signal shapes, a Bukin distribution (left) and a crystal ball function (right). The distributions are overlaid by a Gaussian fit projection in red. For the signal, a branching ratio of zero is assumed for all three decay modes.

The same procedure is repeated for the dimuon signal decays. Instead of two asymmetric distributions, a symmetric and an asymmetric distribution are used. For the symmetric distribution, a Gaussian distribution is used. For the asymmetric distribution, a Double-sided crystal ball function is used. Both distributions, as well as the default Johnson S_U distribution used in the fit, exemplary for $D^0 \to K^-K^+\mu^-\mu^+$ decays, are shown in figure 8.4. Important to note here is the narrowness of the distribution compared to the signal distribution of the LFV decay mode shown in figure 8.2. As the default distribution is already nearly symmetric the variations to the default signal distributions are relatively small. In figure 8.5, the pull distributions for the two parameterisations are shown. The same signal yield as in data is assumed during the generation. No bias, seen by a mean which is compatible with zero, is observed. Therefore, no systematic uncertainty is assigned to the dimuon modes.

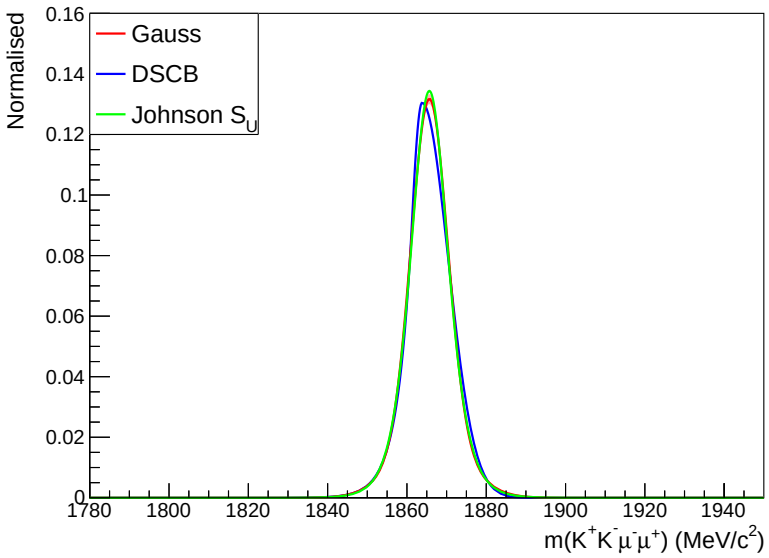


Figure 8.4: Exemplary the different signal distribution, given by a Gaussian distribution (red) and a double-sided crystal ball function (DSCB, blue), shapes used in the generation of the pseudo-experiments compared to the default fit function, given by a Johnson S_U distribution (green), for $D^0 \to K^-K^+\mu^-\mu^+$ signal candidates. All distributions are obtained from simulation.

This procedure is repeated for different branching ratio assumptions. The resulting means of the different pull distributions in dependence on the branching ratio used in the generation are reported in figure 8.6. The observed shift originates from a wrong description of the tail, leading to an overestimation of the branching ratio. For the $D^0 \to K^-K^+\mu^\pm e^\mp$ decay, nearly no misID background is expected and observed, meaning

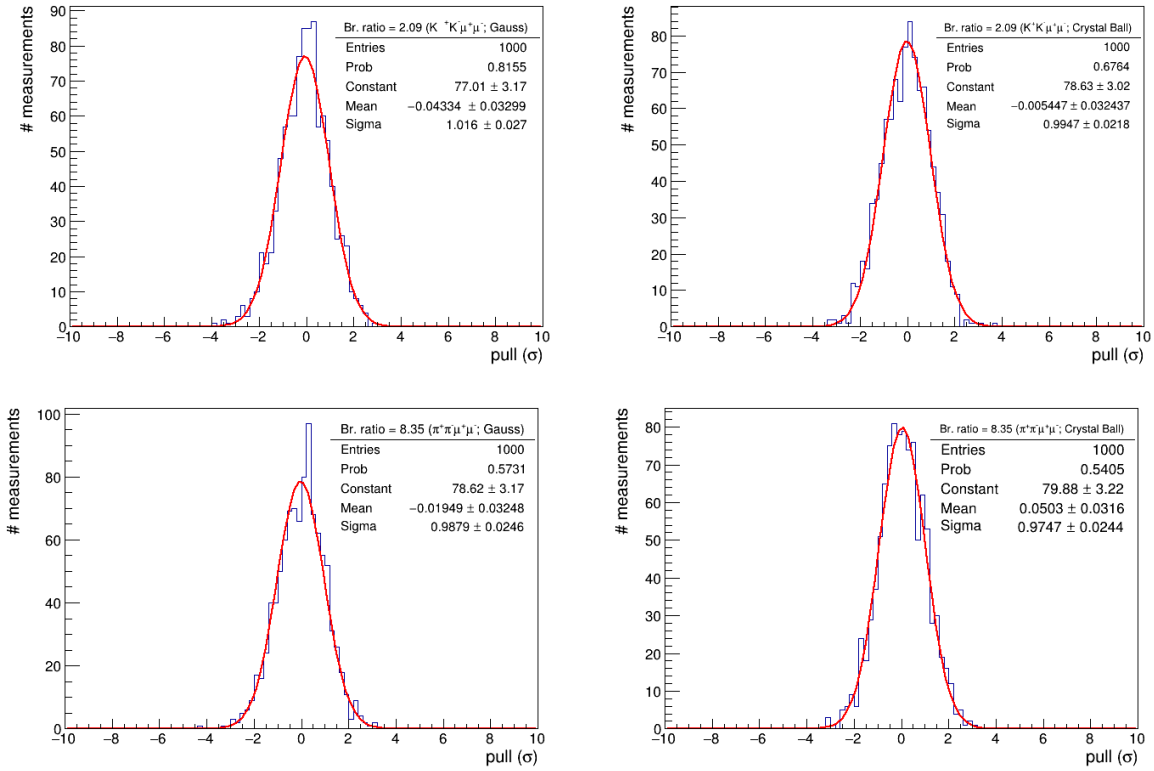


Figure 8.5: Pull distribution for the two different dimuon decay chains, $D^0 \to K^-K^+\mu^-\mu^+$ (up) and $D^0 \to \pi^-\pi^+\mu^-\mu^+$ (down), evaluated by the default fit procedure, using a Johnson S_U, for different generated signal shapes, a Gaussian distribution (left) and a crystal ball function (right). The distributions are overlaid by a Gaussian fit projection in red. For the signal yield, a branching ratio of 2.09×10^{-7} for $D^0 \to K^-K^+\mu^-\mu^+$ decays and 8.35×10^{-7} for $D^0 \to \pi^-\pi^+\mu^-\mu^+$ decays is assumed.

no prominent misID peak is generated. Here, the slightly more flat Bukin distribution leads to an underestimated branching ratio as signal decays are absorbed into the combinatorial background[2].

Noticeable is also that for the dimuon decay modes, no systematic uncertainty is assigned, as a clean separation between signal and misID background is possible. However, due to the lower momentum resolution, which leads to a wider reconstructed mass distribution, this is not possible for the LFV decay channel. Even though the reconstructed mass distribution of the misID background is expected to peak further away from the D^0 mass, this is negated by a long tail for the signal distribution for reconstructed masses below the D^0 mass due to missing bremsstrahlung. A separation between misID background

[2]This is also true for the measured number of misID candidates if generated.

and signal candidates for decays with reconstructed bremsstrahlung is only possible as the relative misID yield between decay candidates with and without reconstructed bremsstrahlung is fixed from simulations.

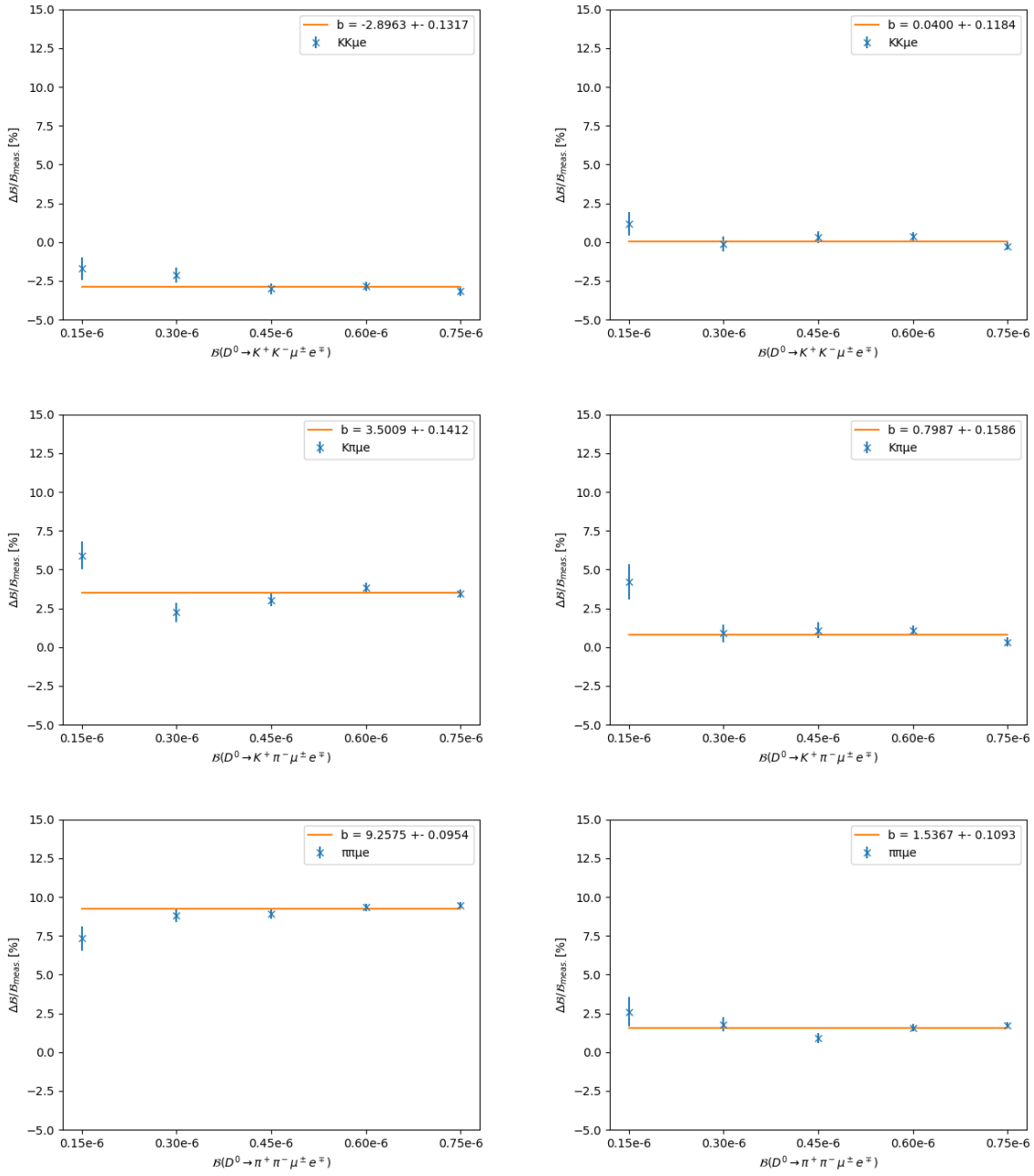


Figure 8.6: The deviation between measured and generated signal branching fraction $\Delta\mathcal{B} := \mathcal{B}_{meas.} - \mathcal{B}_{gen.}$ relative to the measured branching fraction in dependence of the assumed generated signal branching ratio for the three different LFV decay chains, $D^0 \to K^-K^+\mu^\pm e^\mp$ (upper row), $D^0 \to K^-\pi^+\mu^\pm e^\mp$ (middle row) and $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ (lower row), evaluated by the default fit procedure, using a Johnson S_U, for different generated signal shapes, a gaussian distribution (left) and a crystal ball function (right).

8.3 Combinatorial background assumptions

In section 7.3, the fit used to parameterise the combinatorial background is described. The impact of the parameterisation is estimated by generating toy samples with a varied slope. For this purpose, the slope is varied by a standard deviation. Exemplary in figure 8.7, the pull distribution for mass distributions with an added (subtracted) standard deviation to the slope for the LFV decay modes is shown. The pull is defined as the difference between the generated and measured branching ratios normalised by its uncertainty. As the branching ratio is unknown, signal assuming various different branching ratios was generated. The pull is shown in dependence of the generated signal branching fraction. No significant deviations that are not covered by the statistical uncertainties between generated and measured branching ratios are observed. No systematic uncertainty is assigned for the dimuon and LFV decay modes.

8.4 Parameterisation of the misID background

The strong PID selection applied on possible signal candidates directly affects the shape of the misID background. As the misID peak lies below the expected signal distribution, it is one of the main limiting factors in this search. In figure 8.8, multiple shapes extracted, exemplary, from simulated $D^0 \to \pi^-\pi^+\pi^-\pi^+$ decays, under $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ mass hypothesis, selected with different PID requirements are drawn. The shapes can be loosely categorised into three scenarios. The red shape, the furthest curve to the left, corresponds to `isMuon` on the muon candidate being the loosest PID requirement. As soon as any `ProbNN` requirements are set, the shape is shifted to higher masses under the observed signal. The blue, green, cyan and orange curves are extracted by requiring `ProbNN` requirements on both muons. However, the three shapes with stricter PID requirements on both muons, green, cyan, and orange, are unreliable due to the low number of leftover candidates for the shape evaluation. For $D^0 \to K^-K^+\pi^-\pi^+$ and $D^0 \to K^-\pi^+\pi^+\pi^-$ decays, no candidates pass these selection steps at all and producing a sufficiently high number of simulated background candidates which exactly mimic the selection is not possible, as it would be computationally too expensive. However, to still study the impact of the misID parameterisation, different shapes are produced, by requiring different single muon PID requirements. This effectively moves the distribution closer to the D^0 mass for stricter PID requirements. The misID background is expected to behave similarly in the dimuon modes. Mainly, the mass hypothesis during the reconstruction is changed for one of the pions from an electron to a muon, moving the

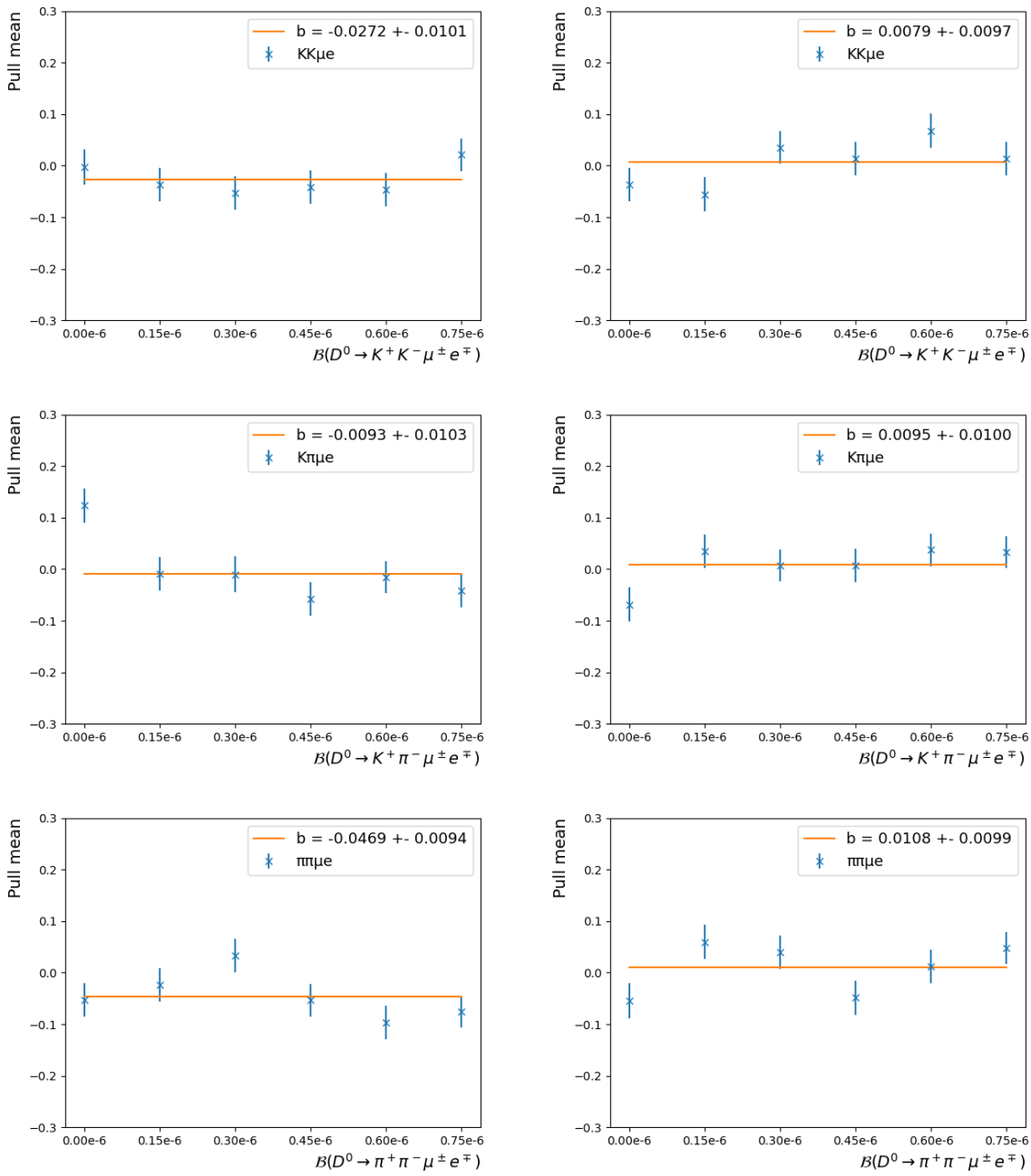


Figure 8.7: The mean of the pull distribution for the three different LFV decay chains, $D^0 \to K^-K^+\mu^\pm e^\mp$ (upper row), $D^0 \to K^-\pi^+\mu^\pm e^\mp$ (middle row) and $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ (lower row), under different linear background hypotheses.

distribution closer to the D^0 mass. `ProbNNmu` has, for the same cut value, a higher background redemption rate, making it with the current simulation samples impossible to show the curves with `ProbNNmu` on both muons. Given the low number of misID background candidates, in the case of added bremsstrahlung photons, no different shapes

are tested for candidates with added bremsstrahlung photons.

For each shape and decay mode, a thousand pseudo experiments are performed. For this, data distributions using the background yields observed in data are produced. The combinatorial background is described by the same linear distribution as was used for the data fit. For the misID background, the different shapes obtained from simulations, by scanning in 0.1 steps over the `ProbNNmu` value are used.

For the dimuon signal, the number of signal candidates is calculated by the efficiency and the branching ratio observed in data[3]. All yields are varied by a Poisson distribution. The difference between the generated and measured mean value of the branching ratio for the dimuon decay modes, $D^0 \to K^-K^+\mu^-\mu^+$ and $D^0 \to \pi^-\pi^+\mu^-\mu^+$ are reported in figure 8.9 in dependence of the `ProbNNmu` value. The maximum absolute deviation of the mean value of the branching ratio, $max(|\Delta\mathcal{B}|)$, is assigned as a systematic uncertainty.

[3]This is equal to the signal yield observed in data.

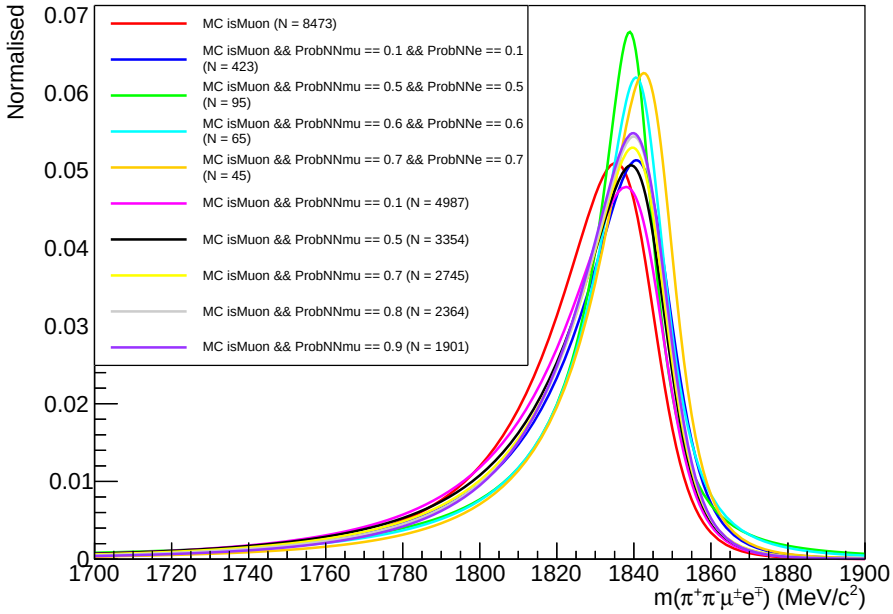


Figure 8.8: Example misID shapes obtained by different PID requirements, on the pions which have the lepton hypothesis assigned, from simulated $D^0 \to \pi^-\pi^+\pi^-\pi^+$ decays under $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ mass hypothesis. Only candidates without wrongly added bremsstrahlung are shown.

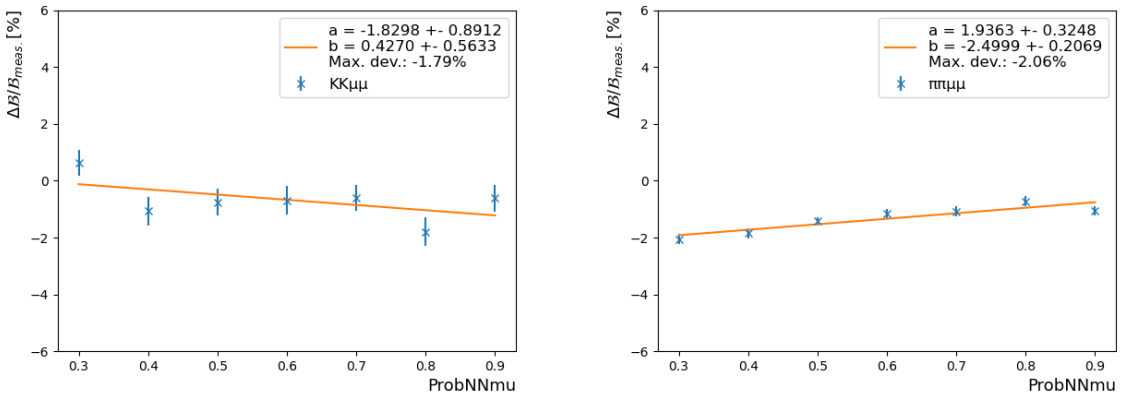


Figure 8.9: From left to right the mean of the pull distributions for $D^0 \to K^-K^+\mu^\pm e^\mp$, $D^0 \to K^-\pi^+\mu^\pm e^\mp$ and $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ pseudo-experiments measurements. Right column: the mean in dependence of the PID cut on the generated shape, assuming a branching ratio of zero. Left column: maximal deviation of the mean from zero for different branching ratio assumptions.

For the blinded LFV decay modes, $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$, $D^0 \to K^-\pi^+\mu^\pm e^\mp$ and $D^0 \to K^-K^+\mu^\pm e^\mp$ decays, it is not possible to generate data with the branching ratio observed in data, as this value is blinded. Therefore, for the case of no observation, the same scan as for the dimuon decay modes is repeated for the LFV decay modes. For the signal yield, zero candidates are assumed, which is done by setting the branching ratio to zero during the generation. The maximum absolute deviation of the mean value of the branching ratio, $max(|\Delta\mathcal{B}|)$, is assigned as a systematic uncertainty in case no signal is observed. To estimate the uncertainty for the case of an observation of LFV decays, the same scan is repeated for different signal branching fractions. The range of this scan is chosen such that the region not yet experimentally excluded is covered. In figure 8.10 the scan results for the different shape assumption are reported, noted by the corresponding `ProbNN` value. Also in figure 8.10, the maximum absolute deviation of the mean value of the branching ratio, $max(|\Delta\mathcal{B}|)$, normalised by the generated branching ratio in dependence of the generated branching fraction are reported. Studying $max(|\Delta\mathcal{B}|)/\mathcal{B}_{meas.}$ shows, that the relative deviation decreases for larger branching ratios. This is also expected as the size of the uncertainty is expected to depend on the number of misidentified background candidates, which is constant apart from statistical fluctuations by the Poisson distribution, as the difference in the generated and measured branching fractions arises due to miscounting signal candidates as misID background candidates or vice versa. Because of the blinding, only a range for the systematic uncertainty due to the misID parameterisation can be given. In case that signal would be observed, it is possible

to repeat the scan for the observed signal branching ratio or alternatively evaluate the value from the distributions in the right column.

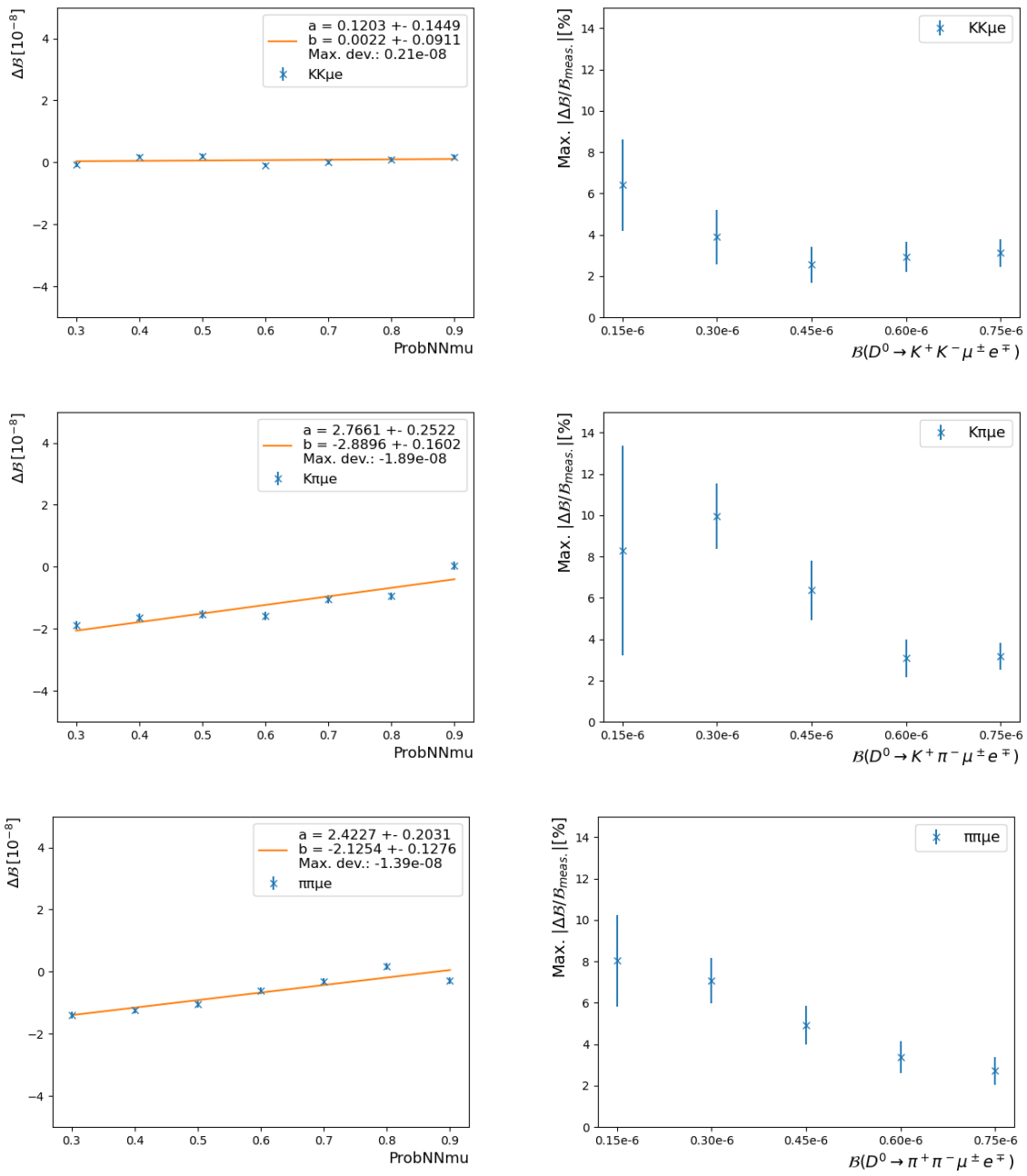


Figure 8.10: From the top row to the bottom row, the mean of the pull distributions for $D^0 \to K^-K^+\mu^\pm e^\mp$, $D^0 \to K^-\pi^+\mu^\pm e^\mp$ and $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$. Left column: the mean in dependence of the PID cut on the generated shape, assuming a branching ratio of zero. Right column: maximal deviation of the mean from zero for different branching ratio assumptions.

All uncertainties quoted in this section, including the uncertainties given by the range for the LFV decay modes, are smaller, though similar in magnitude, than the statistical uncertainties obtained in the previous chapter, chapter 7, which are of order $\mathcal{O}(0.3 \times 10^{-7})$.

8.5 Tracking efficiency

The tracking efficiency is corrected with the help of data-simulation correction maps, as detailed in section 6.2. The associated map values are only known to a certain precision, which needs to be considered. Additionally, until now, uncertainties introduced during the detector simulation due to hadronic interactions with the detector material have not been covered.

An advantage when studying systematic uncertainties is the similar kinematic distribution. Applying the same data-simulation correction map to a track in the denominator and nominator of the efficiency ratio leads to a strong correlation that cancels in the final ratio. Due to the different tracking map used for the electron, the relative systematic uncertainty from the tracking maps does not cancel between the electron and muon. Because of the high statistics in the simulation and data samples, statistical uncertainties due to sample size can be neglected. Nevertheless, an uncertainty is assigned because of possible data and simulation discrepancies. Therefore, the simulation samples are re-weighted in different parameters, such as the number of primary vertices or the number of hits and tracks in different subdetectors. The largest difference introduced by this procedure in the tracking maps is taken as systematic uncertainty [115]. A relative uncertainty of 0.8% is assigned for both the muon and the electron. The square root of the squared sum of the individual systematic uncertainties gives the uncertainty on the efficiency ratio.

A second uncertainty is assigned to the efficiency ratio due to limited knowledge of hadronic interactions within the LHCb detector. From simulations of $B^0 \to J/\psi K^{*0}$, it is estimatd that around 11% of the kaons and 14% of the pions cannot be reconstructed due to hadronic interactions with the detector before the last tracking station. The uncertainty on the material budget used to simulate hadronic interactions within the detector is 10%. Therefore, a relative systematic uncertainty of $11\% \times 10\% = 1.1\%$ for the kaon and $14\% \times 10\% = 1.4\%$ for the pion is assigned. More details can be found in Ref. [115].

For equal particle species in the nominator and denominator, the relative uncertainties in the efficiency ratio are assumed to cancel. Otherwise, the uncertainties between the two particles are treated as uncorrelated. This allows a conservative estimate for an upper limit on the uncertainty. The results for the LFV decay channels are summarised in table 8.1. For the dimuon reference channel, only the difference in the hadron species composition needs to be considered, leading to an assigned systematic uncertainty of 1.8% for both decay channels respectively.

	$\frac{\epsilon(D^0\to K^-K^+\mu^\pm e^\mp)}{\epsilon(D^0\to K^-\pi^+\mu^-\mu^+)}$	$\frac{\epsilon(D^0\to K^-\pi^+\mu^\pm e^\mp)}{\epsilon(D^0\to K^-\pi^+\mu^-\mu^+)}$	$\frac{\epsilon(D^0\to \pi^-\pi^+\mu^\pm e^\mp)}{\epsilon(D^0\to K^-\pi^+\mu^-\mu^+)}$
h_0	-	-	1.8%
h_1	1.8%	-	-
μ	-	-	-
e	1.1%	1.1%	1.1%
total	2.1%	1.1%	2.1%

Table 8.1: Relative systematic uncertainties assigned to the efficiency ratio because of the tracking procedure. Candidates marked with - coincide between the signal and normalisation channel and are assumed to be 100% correlated, therefore cancelling systematic uncertainties in the ratio.

8.6 Particle identification

The efficiency of the PID selection is estimated by correcting the PID variables with data as described in section 4.3. Two sources of uncertainties are considered:

- Uncertainties due to limited statistics in the calibration samples. Alternative PID templates with different seeds of the random generator are used to test and assign a systematic uncertainty.
- Uncertainties due to the parameterised kernel width[4] of the PID control sample. For this purpose, different kernel widths are tested.

Given the large sample size of the calibration modes used in the resampling process, $D^0 \to K^-\pi^+$ for the hadrons and $J\psi \to \mu^+\mu^-(e^+e^-)$ for the leptons, no significant

[4]The kernel used is a Gaussian distribution, and the kernel width corresponds to the standard deviation of said distribution.

Category	Year	$\frac{\epsilon(D^0\to K^-K^+\mu^\pm e^\mp)}{\epsilon(D^0\to K^-\pi^+\mu^-\mu^+)}$	$\frac{\epsilon(D^0\to K^-\pi^+\mu^\pm e^\mp)}{\epsilon(D^0\to K^-\pi^+\mu^-\mu^+)}$	$\frac{\epsilon(D^0\to \pi^-\pi^+\mu^\pm e^\mp)}{\epsilon(D^0\to K^-\pi^+\mu^-\mu^+)}$
`noBrem`	2015/2016	1.5%	1.5%	0.8%
	2017/2018	1.8%	2.3%	2.4%
`Brem`	2015/2016	1.6%	0.8%	0.9%
	2017/2018	0.1%	1.2%	0.8%

Table 8.2: Uncertainties on the PID efficiency ratio due to the kernel size.

	$\frac{\epsilon(D^0\to K^-K^+\mu^-\mu^+)}{\epsilon(D^0\to K^-\pi^+\mu^-\mu^+)}$	$\frac{\epsilon(D^0\to \pi^-\pi^+\mu^-\mu^+)}{\epsilon(D^0\to K^-\pi^+\mu^-\mu^+)}$
2015/2016	0.4	0.1
2017/2018	1.0	0.3

Table 8.3: Relative error due to different kernel sizes in the efficiency estimation for the different branching ratios, in %.

impact is expected. This is checked by resampling the PID variables with five different random seeds. The fluctuation due to random seeds is negligible.

The calibration sample is binned with 100 bins each in the variables p_T, η, `nTracks` and the desired PID response. Then, these variables are smeared in four dimensions with a particular kernel width size. The systematic bias given by the finite binning is tested by changing the kernel width size. For this purpose, the kernel widths are varied around their nominal values. Each kernel size is generated with three different seeds. The largest average variation for different kernel sizes is assigned as systematic uncertainty. The uncertainties for the LFV decay modes are summarised in table 8.2 and for the dimuon decay modes in table 8.3.

8.7 Trigger efficiency

When calculating the trigger correction maps, the limited data and simulation sample size leads to relatively large uncertainties of the applied trigger weights. However, enforcing the same trigger requirements for signal and normalisation mode and a similar expected kinematic distribution for both decays leads to a strong correlation between the correction weights applied on the normalisation mode and the signal mode. This leads to a strong

cancellation of the uncertainties in the ratio. A bootstrapping approach is chosen to account for possible effects due to the applied binning scheme and the correlation introduced by applying the same weights to signal and normalisation. The calibration samples are resampled by randomly drawing events from the original sample until the same number of events fluctuated by a Poisson distribution is reached. These new samples are then used to calculate new correction weights and apply those to the data. This is repeated a thousand times. Finally, the standard deviation of the resulting trigger efficiency ratio distribution is evaluated. The relative systematic uncertainties for the LFV decays related to the trigger weights are reported in table 8.4 and for the dimuon reference mode in table 8.5. The uncertainties on the trigger selection are usually one of the main systematic uncertainties in branching ratio measurements at LHCb. However, due to the same applied trigger selection and similar kinematic distribution between signal and normalisation mode, the calculated trigger uncertainties are relatively small compared to the other systematic uncertainties of this measurement. The uncertainties are treated as uncorrelated with other uncertainties.

Category	Year	$\frac{\epsilon(D^0\to K^-K^+\mu^\pm e^\mp)}{\epsilon(D^0\to K^-\pi^+\mu^-\mu^+)}$	$\frac{\epsilon(D^0\to K^-\pi^+\mu^\pm e^\mp)}{\epsilon(D^0\to K^-\pi^+\mu^-\mu^+)}$	$\frac{\epsilon(D^0\to \pi^-\pi^+\mu^\pm e^\mp)}{\epsilon(D^0\to K^-\pi^+\mu^-\mu^+)}$
`noBrem`	2015/2016	0.4%	0.4%	0.3%
	2017/2018	0.4%	0.3%	0.3%
`Brem`	2015/2016	1.5%	0.4%	0.6%
	2017/2018	0.5%	0.3%	0.2%

Table 8.4: Uncertainty on the efficiency ratio due to trigger weights for the LFV decay modes.

Year	$\frac{\epsilon(D^0\to K^-K^+\mu^-\mu^+)}{\epsilon(D^0\to K^-\pi^+\mu^-\mu^+)}$	$\frac{\epsilon(D^0\to \pi^-\pi^+\mu^-\mu^+)}{\epsilon(D^0\to K^-\pi^+\mu^-\mu^+)}$
2015/2016	0.4%	0.2%
2017/2018	0.5%	0.3%

Table 8.5: Uncertainty on the efficiency ratio due to trigger weights for the dimuon reference mode.

8.8 Model assumptions for simulated samples

The selection efficiency depends on the kinematic and geometric distributions of the studied decay, which are derived from simulations. A wrong modelling of the dihadron or dimuon spectrum can, therefore, bias the efficiency. As the precise decay model for the reference dimuon and the normalisation mode is unknown, the potential bias introduced by the decay model is evaluated by decreasing and increasing the relative contributions of the dihadron and dimuon resonances by 50% relative to the default model, reported in table 3.1 on page 33. Afterwards, the efficiency is recomputed for the varied model, including all data-simulation correction weights for the trigger and tracking efficiencies. The resulting efficiency is then compared to the result of the default model. The RMS of the relative deviations is assigned as a systematic uncertainty. This calculation is performed separately for the dimuon and dihadron assumptions, which are then added in quadrature. The results for the two reference decay modes and the normalisation channel under the three different applied selections can be found in table 8.6.

Norm. decays	$D^0 \to K^-K^+\mu^\pm e^\mp$	$D^0 \to K^-\pi^+\mu^\pm e^\mp$	$D^0 \to \pi^-\pi^+\mu^\pm e^\mp$
dimuon	1.0%	1.1%	1.0%
dihadron	2.9%	3.1%	3.1%
Syst. unc. eff.	3.1%	3.3%	3.3%

Dimuon decays	$D^0 \to K^-K^+\mu^\pm\mu^\mp$	$D^0 \to \pi^-\pi^+\mu^\pm\mu^\mp$
dimuon	1.4%	0.9%
dihadron	2.7%	3.1%
Syst. unc. eff.	3.0%	3.2%

Table 8.6: The upper table shows the uncertainties introduced by varying the dihadron and dimuon spectrum as well as the combined uncertainty on the efficiency for the normalisation mode selected for the decays specified in the header. The normalisation mode selection for $D^0 \to K^-K^+\mu^-\mu^+$ and $D^0 \to K^-K^+\mu^-\mu^+$ decays is the same as for $D^0 \to K^-K^+\mu^\pm e^\mp$ and $D^0 \to K^-K^+\mu^\pm e^\mp$ decays. The lower table shows the uncertainties for the dimuon signal decays.

The observed variation depends on the momentum of the daughter particles, as a higher momentum results in a higher efficiency, while for a lower momentum the efficiency decreases[5]. For the dimuon decays, the square root of the squared sum of the relative

[5]For the muon this is well illustrated in figure 3.8 on page 29, which shows the tracking efficiency

uncertainties on the efficiency is assigned as systematic uncertainty to the efficiency ratio and, consequently, the branching ratio. For the LFV decays, only the relative uncertainty of the normalisation efficiency is assigned as systematic uncertainty. Overall, the systematic uncertainties are similar in size to the uncertainties in the previously published measurement of the branching fractions for $D^0 \to \pi^-\pi^+\mu^-\mu^+$ and $D^0 \to K^-K^+\mu^-\mu^+$ of about 3.4% [44]. This measurement was performed in different dimuon bins of the reconstructed mass in which, using a similar approach, the decay model was varied. As the dimuon reconstructed and dihadron reconstructed mass are strongly correlated, a direct comparison of the evaluated systematic uncertainties is difficult. The effect on the decay model due to the normalisation mode was not considered. However, the systematic uncertainties on the decay model for the normalisation mode can be compared to the uncertainty of the decay model obtained for the first observation of the normalisation decay of 2.9% [61]. The same model assumptions and procedure to evaluate the systematic uncertainties are used in this measurement.

8.9 Limited simulation statistic

A bootstrapping algorithm is used to account for statistical fluctuations of the simulation samples. The numerical approach given by the bootstrapping algorithm was chosen to account for the trigger and tracking correction weights and their respective signal-normalisation mode correlation in the efficiency ratio. For this, a combined weight ($w_{trig} \times w_{track}$) is calculated with every candidate that passes the selection. All other candidates get assigned a weight of zero. Then, a bootstrapping algorithm is applied to the data set to evaluate the uncertainty. The relative uncertainty divided by year and decay mode on the efficiency ratio can be found in table 8.7 for the LFV decay modes. The uncertainties for the reference decay modes, $D^0 \to K^-K^+\mu^-\mu^+$ and $D^0 \to \pi^-\pi^+\mu^-\mu^+$, are reported in table 8.8.

Most weights are close to unity. The influence of the weights in the uncertainty calculation is, therefore, expected to be small. This can be confirmed by analytically calculating the uncertainties, neglecting the weights altogether using the Bayesian binomial errors. The resulting uncertainties are quoted in the brackets in table 8.7. No correlation between signal and normalisation is assumed during the calculation. Only a small difference between analytical and numerical values is observed as expected.

in dependence of the muon(electron) momentum. As a reminder, the typical momentum of muons for decays studied in this thesis is below 5000 MeV/c.

Category	Year	$\frac{\epsilon(D^0\to K^-K^+\mu^\pm e^\mp)}{\epsilon(D^0\to K^-\pi^+\mu^-\mu^+)}$	$\frac{\epsilon(D^0\to K^-\pi^+\mu^\pm e^\mp)}{\epsilon(D^0\to K^-\pi^+\mu^-\mu^+)}$	$\frac{\epsilon(D^0\to \pi^-\pi^+\mu^\pm e^\mp)}{\epsilon(D^0\to K^-\pi^+\mu^-\mu^+)}$
`noBrem`	2015/2016	8.0%(7.9%)	5.1%(5.1%)	4.2%(4.1%)
	2017/2018	3.1%(3.1%)	2.2%(2.2%)	1.7%(1.7%)
`Brem`	2015/2016	15.0%(14.7%)	8.6%(8.6%)	7.6%(7.5%)
	2017/2018	4.5%(4.5%)	2.9%(2.9%)	2.1%(2.1%)

Table 8.7: Relative uncertainties due to limited simulation sample size in the efficiency ratio estimation for the LFV decay modes. The values in the brackets are analytically calculated using the Bayesian binomial error, neglecting the weights altogether, to illustrate the impact of the bootstrapping algorithm.

Year	$\frac{\epsilon(D^0\to K^-K^+\mu^-\mu^+)}{\epsilon(D^0\to K^-\pi^+\mu^-\mu^+)}$	$\frac{\epsilon(D^0\to \pi^-\pi^+\mu^-\mu^+)}{\epsilon(D^0\to K^-\pi^+\mu^-\mu^+)}$
2015/2016	4.1%(4.0%)	2.9%(3.0%)
2017/2018	1.6%(1.6%)	1.4%(1.4%)

Table 8.8: Relative uncertainties due to limited simulation sample size in the efficiency ratio estimation for the dimuon decay modes. The values in the brackets are analytically calculated using the Bayesian binomial error, neglecting the weights altogether, to illustrate the impact of the bootstrapping algorithm.

The above calculation, using a bootstrapping algorithm, is repeated with all data categories and years combined. The different years are scaled according to the recorded luminosity, while the ratio between candidates with and without reconstructed bremsstrahlung is evaluated from simulations. Note that the uncertainties in the data categories with most of the expected candidates are the smallest. The resulting uncertainties for the LFV decays are reported in table 8.9 and for the dimuon decay modes in table 8.10.

8.10 Summary & Concluding Remarks

All non-negligible systematic uncertainties on the branching ratio are combined and summarised in table 8.9 for the LFV decays and in table 8.10 for the dimuon decays.

Decay	$D^0 \to K^-K^+\mu^\pm e^\mp$	$D^0 \to K^-\pi^+\mu^\pm e^\mp$	$D^0 \to \pi^-\pi^+\mu^\pm e^\mp$
Yield ratio			
Par. of the signal	2.9%(−)	3.5%(−)	9.3%(−)
Par. of the peaking bkg	$2.5\% - 6.4\%(0.21 \times 10^{-8})$	$3.1\% - 10.0\%(1.89 \times 10^{-8})$	$2.7\% - 8.0\%(1.39 \times 10^{-8})$
Systematic unc.	$3.8\% - 7.0\%(0.21 \times 10^{-8})$	$4.7\% - 10.6\%(1.89 \times 10^{-8})$	$9.7\% - 12.3\%(1.39 \times 10^{-8})$
Efficiency ratio			
Tracking maps	1.1%	1.1%	1.1%
Had. material interactions	2.1%	1.1%	2.1%
PID var. transformation	0.7%	0.8%	0.7%
Trigger correction	0.3%	0.2%	0.2%
Simulated decay models	2.9%	3.1%	3.1%
Limited simulation statistic	3.0%	1.9%	1.8%
Systematic unc.	4.9%	4.0%	4.4%
Total sys. unc.	$5.4\% - 8.0\%(0.21 \times 10^{-8})$	$5.4\% - 10.9\%(1.89 \times 10^{-8})$	$10.2\% - 12.7\%(1.39 \times 10^{-8})$

Table 8.9: Summary of all systematic uncertainties for the LFV decay modes. The systematic uncertainties in the determination of the yield ratio are given relative to the measured branching fraction. In brackets, the uncertainties for the case of no observation are quoted. The efficiency ratio uncertainties are given relative to the measured branching fraction.

One of the largest systematic uncertainties of 10% arises from the limited knowledge of the normalisation mode branching ratio, which will be quoted as additional uncertainty.

The main systematic uncertainty assigned to the yield ratio is due to the parameterisation of the misID background. This systematic uncertainty is introduced because of the fixed parametrisation obtained from simulations during the fit procedure. This uncertainty is assigned to the LFV decay modes and the dimuon decay mode. Due to the wider reconstructed mass distribution for LFV decays, both distributions have a large overlap, resulting in a larger probability of wrongly identifying (and miscounting) signal and background candidates. This is also observed during the study of the systematic uncertainties, resulting in a larger uncertainty due to the peaking background for the LFV decays than the dimuon decays.

The three largest systematic uncertainties affecting the efficiency ratio are introduced because of the unknown decay model of the normalisation mode and the consequent assumptions made during the generation of the simulation samples, the size of the simulation samples, and the limited knowledge of the detector's material budget.

The easiest systematic uncertainty to reduce is the uncertainty due to the simulation sample size. However, as the uncertainty scales by the square root of the sample size, a significantly larger simulation sample and significantly more computing resources would be required. Fast simulation techniques, for example implemented in the RapidSim framework [116], will play an important role in the future, especially when studying rare or forbidden decays. In this context also, a better decay model for the dimuon and dihadron resonances would help to improve the quality of the simulation samples. For B-meson decays, such a model already exists. For charm decays, due to larger theoretical uncertainties, the formulation of such a model is more challenging and would require improved theory predictions. From an experimental point of view, it is, therefore, also important to measure or provide bounds for specific phase space regions to crosscheck and improve the theoretical predictions, as was done with the measurement using Run 1 data. Already performing this analysis in different phase space bins of the dihadron and dimuon mass would allow to increase the handle on this uncertainty.

Decay	$D^0 \to K^-K^+\mu^-\mu^+$	$D^0 \to \pi^-\pi^+\mu^-\mu^+$
Yield ratio		
Par. of the peaking bkg	1.8%	2.1%
Systematic unc.	1.8%	2.1%
Efficiency ratio		
Tracking maps	1.1%	1.1%
Had. material interactions	1.8%	1.8%
PID var. transformation	0.7%	0.3%
Trigger correction	0.4%	0.2%
Simulated decay models	4.3%	4.6%
Limited simulation statistic	1.7%	1.3%
Systematic unc.	5.1%	5.2%
Total sys. unc.	5.4%	5.6%

Table 8.10: Summary of all systematic uncertainties for the dimuon decay modes. The systematic uncertainties in the determination of the yield ratio are given relative to the measured branching fraction. The efficiency ratio uncertainties are given relative to the measured branching fraction.

In Run 2, the material budget per track has an uncertainty of 10% assigned [117]. Ideally, one would like to crosscheck if this estimate is correct and, if possible, improve it. The improvement or validation of this estimate for Run 3 is currently studied within the LHCb collaboration. This is done by directly measuring the impact of the material interactions by comparing $D^0 \to K^-\pi^+$and $D^0 \to K^-\pi^+\pi^+\pi^-$ decays.

Dimuon decays

The following branching fractions are measured for the reference mode:

$$\mathcal{B}(D^0 \to K^-K^+\mu^-\mu^+) = (2.09 \pm 0.16 \pm 0.11 \pm 0.21) \times 10^{-7} \tag{8.10.1}$$

$$\mathcal{B}(D^0 \to \pi^-\pi^+\mu^-\mu^+) = (8.35 \pm 0.27 \pm 0.47 \pm 0.84) \times 10^{-7} \tag{8.10.2}$$

The first quoted uncertainty corresponds to the statistical uncertainty. The second covers all systematic uncertainties apart from the uncertainty of the normalisation branching fraction, which is quoted as the third uncertainty.

The results for the same branching ratios measured with the Run 1 dataset [44] are:

$$\mathcal{B}(D^0 \to K^-K^+\mu^-\mu^+) = (1.54 \pm 0.27 \pm 0.09 \pm 0.16) \times 10^{-7} \tag{8.10.3}$$

$$\mathcal{B}(D^0 \to \pi^-\pi^+\mu^-\mu^+) = (9.64 \pm 0.48 \pm 0.51 \pm 0.97) \times 10^{-7} \tag{8.10.4}$$

The first quoted uncertainty corresponds to the statistical uncertainty on the **signal** dataset[6]. The second uncertainty covers all systematic uncertainty apart from the uncertainty of the normalisation branching fraction, which is quoted as the third uncertainty. The following relative systematic uncertainties were considered for the Run 1 measurement [23, 44]:

- Uncertainty due to the parameterisation of the peaking background: 1.4%
- Statistical uncertainty of the normalisation mode for $D^0 \to \pi^-\pi^+\mu^-\mu^+$ ($D^0 \to K^-K^+\mu^-\mu^+$): 2.6% (2.7%)
- Uncertainty due to the fit model used to estimate the simulation yields: 1.0%
- Uncertainty because of the unknown signal decay models in simulations: 3.4%

[6]The statistical uncertainty due to the normalisation mode dataset is included in the systematic uncertainty.

- Uncertainty evaluated by variations of the binning schemes during the PID efficiency calculation: 0.8%
- Uncertainty because of the L0 trigger efficiency correction: 1.3%
- Uncertainty due to observed differences in the BDT efficiency between data and simulations: 1.3%

Most uncertainties are similar in size for the Run 1 and Run 2 measurements. One should note, that the shape of the peaking background in Run 1 was evaluated from separately recorded $D^0 \to h^{(\prime)-}h^+\pi^-\pi^+$ decays contrary to Run 2 where simulated events were used. Usually, the approach using data is preferable. However, to keep the analysis similar for the LFV and dimuon decay modes, simulated events are used, as no bremsstrahlung recovery is available for the recorded hadronic decay modes.

Additionally, while the uncertainties are similar in size, different corrections were applied for the Run 1 measurement than for the Run 2 measurement. For the PID and Trigger correction, more advanced methods are available nowadays, for example by using kernel density estimates instead of a fixed binning scheme. This can also be seen in the smaller systematic uncertainties assigned to the respective methods.

The correction factors used for the tracking algorithm were not yet evaluated for the Run 1 measurement and are left unaccounted for. The same applies to the uncertainty due to hadronic interactions with the detector material.

No uncertainty was in the Run 1 measurement assigned to the decay model used to produce the simulated events, which were used to calculate the normalisation mode efficiencies. However, using the Run 2 created simulation samples, which use the same underlying decay models, it is possible to reproduce the Run 1 results and apply the same procedure to the simulated normalisation mode samples. This results in a larger systematic uncertainty because of the unknown dihadron and dimuon spectrum for Run 2.

For Run 1, no direct uncertainty is quoted due to the simulated sample size.

For the Run 2 measurement, a similar relative systematic uncertainty is assigned to the branching fraction of $D^0 \to K^-K^+\mu^-\mu^+$ decays. For the branching fraction of $D^0 \to \pi^-\pi^+\mu^-\mu^+$ decays, a larger systematic uncertainty, 5.4% compared to 4.8%, is assigned. While the individual systematic uncertainties related to the efficiency improved,

some previously not considered uncertainties are now included.

To compare the Run 1 and Run 2 results, only uncorrelated uncertainties have to be considered. Between Run 1 and Run 2, only minor changes occurred to the detector. However, it is important to keep in mind that for the Run 2 measurements, more, often small, corrections to the efficiency ratio are applied. In Run 1 the differences between data and simulation introduced, for example, by the tracking algorithm were not corrected. Also, no systematic uncertainty was assigned. To provide a conservative comparison, only uncertainties of a statistical nature will be considered:

- Statistical uncertainty on the signal yield
- Statistical uncertainty on the normalisation yield
- The uncertainty related to the simulation sample size (Run 2)
- The uncertainty by the fit model to estimate the simulation yields (Run 1)

The measurement of $\mathcal{B}(D^0 \to K^-K^+\mu^-\mu^+)$ is compatible within 1.75 standard deviations and the $D^0 \to \pi^-\pi^+\mu^-\mu^+$ measurement within 2.07 standard deviations with the result of Run 1, taking just statistical uncertainties into account. The measured branching fraction for Run 2, quoted only with uncertainties of statistical nature, are:

$$\mathcal{B}(D^0 \to K^-K^+\mu^-\mu^+) = (2.09 \pm 0.16) \times 10^{-7} \tag{8.10.5}$$

$$\mathcal{B}(D^0 \to \pi^-\pi^+\mu^-\mu^+) = (8.35 \pm 0.29) \times 10^{-7} \tag{8.10.6}$$

The measured branching fraction for Run 1, quoted only with uncertainties of statistical nature, are:

$$\mathcal{B}(D^0 \to K^-K^+\mu^-\mu^+) = (1.54 \pm 0.27) \times 10^{-7} \tag{8.10.7}$$

$$\mathcal{B}(D^0 \to \pi^-\pi^+\mu^-\mu^+) = (9.64 \pm 0.55) \times 10^{-7} \tag{8.10.8}$$

The agreement of approximately two standard deviations between the two measurements suggests that the measurement procedure provides a reasonable result of the branching fractions.

LFV decays

For the LFV decay modes, two cases are considered: The case that no LFV is observed, which will lead to a limit on the branching fraction, and the case that LFV is observed,

in which case the branching fraction will be measured and reported. This is relevant, as the systematic uncertainty of the peaking background depends non-linearly on the measured branching fraction. As this dependence is not proportional to the measured branching fraction (or yield ratio) like in the case of the efficiency ratio, it is also not possible to quote a relative uncertainty. Because of this, only a range for the systematic uncertainty in case of observation is given, while for the case of no observation, $\mathcal{B}(D^0 \to h^{(\prime)-}h^+\mu^\pm e^\mp) = 0$, an absolute uncertainty will be quoted, which is provided in the bracket in table 8.9. For the limit calculation, the efficiency ratio and the yield ratio are treated as nuisance parameters with their own uncertainties.

In the simplified case that a branching fraction of zero is measured, the total uncertainty reported in brackets in table 8.9 is taken as systematic uncertainty:

$$\mathcal{B}(D^0 \to K^-K^+\mu^\pm e^\mp) = (0.00 \pm 2.54 \pm 0.21) \times 10^{-8} \tag{8.10.9}$$

$$\mathcal{B}(D^0 \to K^-\pi^+\mu^\pm e^\mp) = (0.00 \pm 4.90 \pm 1.89) \times 10^{-8} \tag{8.10.10}$$

$$\mathcal{B}(D^0 \to \pi^-\pi^+\mu^\pm e^\mp) = (0.00 \pm 3.26 \pm 1.39) \times 10^{-8} \tag{8.10.11}$$

where the first uncertainty corresponds to the statistical uncertainty of normalisation and signal dataset, and the second is the systematic uncertainty. However, typically, in searches like this, a branching fraction unequal to but compatible with zero is observed, which would require a recalculation of the systematic uncertainty as the systematic uncertainty depends on the observed central value of the branching fraction. For this case, a relative total systematic uncertainty is quoted in table 8.9.

A special role has the uncertainty due to the peaking background. This uncertainty scales approximately inversely by the observed branching fraction, as it is directly linked to the misID background yield. This was kept constant, apart from statistical fluctuations mimicked by varying the yield according to a Poisson distribution. One should add that due to this behaviour of scaling inversely with the branching fraction, it is not expected that this uncertainty exceeds the statistical uncertainty[7]. The uncertainty arises due to the overlap of signal and misID background, as signal candidates are moved due to lost bremsstrahlung photons under the misID component of the fit. This uncertainty could be removed by neglecting candidates with no bremsstrahlung recovery, as hadronic decays do not produce bremsstrahlung. However, this would remove half the statistical power of this measurement, increasing the statistical uncertainty (and the expected limit)

[7]This can be seen in figure 8.10. The systematic uncertainty will reach for all three LFV decay modes a maximum around 2.3×10^{-8}.

approximately by a factor $\sim \sqrt{2} \approx 1.41$. This is also observed when performing the measurement just within one bremsstrahlung data category.

Chapter 9

Limit setting

In this chapter, the limit calculation and its results for the LFV decay modes are reported. For the two observed reference decay modes, $D^0 \to \pi^-\pi^+\mu^-\mu^+$ and $D^0 \to K^-K^+\mu^-\mu^+$, no limit is calculated.

Without unblinding the result, it is still possible to provide the expected experimental limit under the hypothesis that no signal is observed. The observed limit is expected to be compatible, lying within the error bands, with the expected limit in case no signal is observed. The expected limits are calculated by the `GammaCombo` framework [96]. For this, the likelihood of the branching ratio is profiled. The yield ratio, efficiency ratio and normalisation branching ratio are treated as nuisance parameters with their respective uncertainties. The normalisation branching ratio uncertainty is 10% [61]. The uncertainty on the efficiency ratio is fully described by its systematic uncertainty, reported in table 8.9, page 139. For the uncertainty on the yield ratio, the systematic uncertainty on the yield is numerically combined during the generation of the pseudo-experiments with the likelihood obtained by the fit. The profiling itself is done via pseudo-experiments. The for the LFV signal modes at 95%(90%) CL evaluated limits are:

$$\mathcal{B}(D^0 \to K^-K^+\mu^\pm e^\mp) \leq 6.5(5.5) \times 10^{-8} \tag{9.0.1}$$

$$\mathcal{B}(D^0 \to K^-\pi^+\mu^\pm e^\mp) \leq 11.2(9.5) \times 10^{-8} \tag{9.0.2}$$

$$\mathcal{B}(D^0 \to \pi^-\pi^+\mu^\pm e^\mp) \leq 7.8(6.6) \times 10^{-8} \tag{9.0.3}$$

The corresponding confidence contour using the $\mathrm{CL_s}$ p-value can be found in figure 9.1. The red curve marks the expected limit at a certain $\mathrm{CL_s}$ value. The blue and light blue areas mark the error bands of the expected limit evaluated under the hypothesis that no signal is observed. The same scan as for the expected limit, after unblinding, will be repeated with the observed central value, which allows the calculation of the observed

limits at 95%(90%) CL. The curve of the observed limit is expected to lie within the blue error bands of the expected limit.

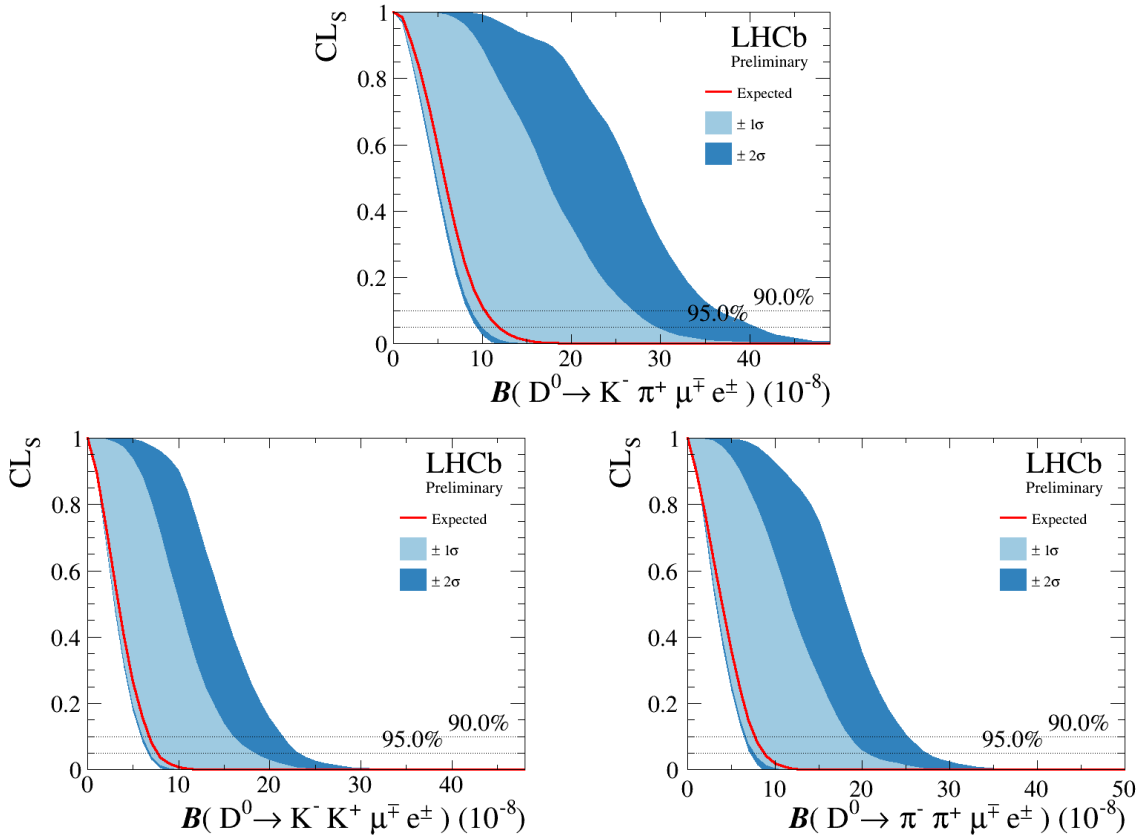


Figure 9.1: Confidence contour obtained by the CL_s method and the expected limits for the background-only hypothesis. The dotted lines marked the CL. In blue are the error bands of the expected limit shown. The upper figure shows the confidence contour for $D^0 \to K^-\pi^+\mu^\pm e^\mp$ decays, the lower left figure for $D^0 \to K^-K^+\mu^\pm e^\mp$ decays and the lower right figure for $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ decays.

To be able to set more precise limits for new physics theories, e.g. leptoquark theories, the limit is also quoted separately for signal candidates with the muon having the same and opposite charge of the slow pion of the D^{*+} decay, tagging the flavour of the D^0 meson. For this the data samples are split by the relative charges in the final state particle composition. By this the charge of the muon and electron are correlated to the charge of the pion. As the charged conjugated decay products are considered too, the efficiency is assumed to be the same. For example, in the two pion case, $D^{*+} \to D^0(\to \pi^+\pi^-\mu^-e^+)\pi^+$ decays and $D^{*-} \to \bar{D}^0(\to \pi^-\pi^+\mu^+e^-)\pi^-$ decays are both considered together, the slow pion and muon have the opposite charge, similar to $D^{*+} \to D^0(\to \pi^+\pi^-\mu^\mp e^\pm)\pi^+$ and

$D^{*+} \to D^0(\to \pi^+\pi^-\mu^\mp e^\pm)\pi^+$ no charge is favoured for a specific final decay particle[1]. Using the CL_s method, the following limits at 95%(90%) CL are set for muons with the same charge as the accompanying slow tag pion:

$$\mathcal{B}(D^0 \to K^+K^-\mu^+e^-) \leq 5.5(4.8) \times 10^{-8} \tag{9.0.4}$$

$$\mathcal{B}(D^0 \to K^-\pi^+\mu^+e^-) \leq 9.0(8.1) \times 10^{-8} \tag{9.0.5}$$

$$\mathcal{B}(D^0 \to \pi^+\pi^-\mu^+e^-) \leq 5.7(4.9) \times 10^{-8} \tag{9.0.6}$$

and muons with the opposite charge as the accompanying slow tag pion:

$$\mathcal{B}(D^0 \to K^+K^-\mu^-e^+) \leq 4.5(3.9) \times 10^{-8} \tag{9.0.7}$$

$$\mathcal{B}(D^0 \to K^-\pi^+\mu^-e^+) \leq 6.7(6.0) \times 10^{-8} \tag{9.0.8}$$

$$\mathcal{B}(D^0 \to \pi^+\pi^-\mu^-e^+) \leq 6.4(5.8) \times 10^{-8} \tag{9.0.9}$$

The corresponding contour plots can be found in figure 9.2.

The expected limits show that in the case of no observed signal, the observed limits for the LFV decay channel are expected to improve the current best (observed) limits by the BaBar collaboration by two orders of magnitude. The expected limits for the separately charged final states are smaller than the expected limit for the flavour-combined LFV signal decays. This is also expected as the efficiency is assumed to be the same, while the data sample size is halved, reducing the absolute uncertainties on the yields[2] and branching ratio.

[1]This is contrary to CP violation measurements in $\bar{D}^0 - D^0$ mixing where the charge of the slow pion which tags the flavour is crucial. In those measurements, the slow pion efficiency is evaluated separately for π^- and $\pi+$ as both interact differently with the detector material.

[2]The relative uncertainties on the yields increase as expected by the smaller data sample size.

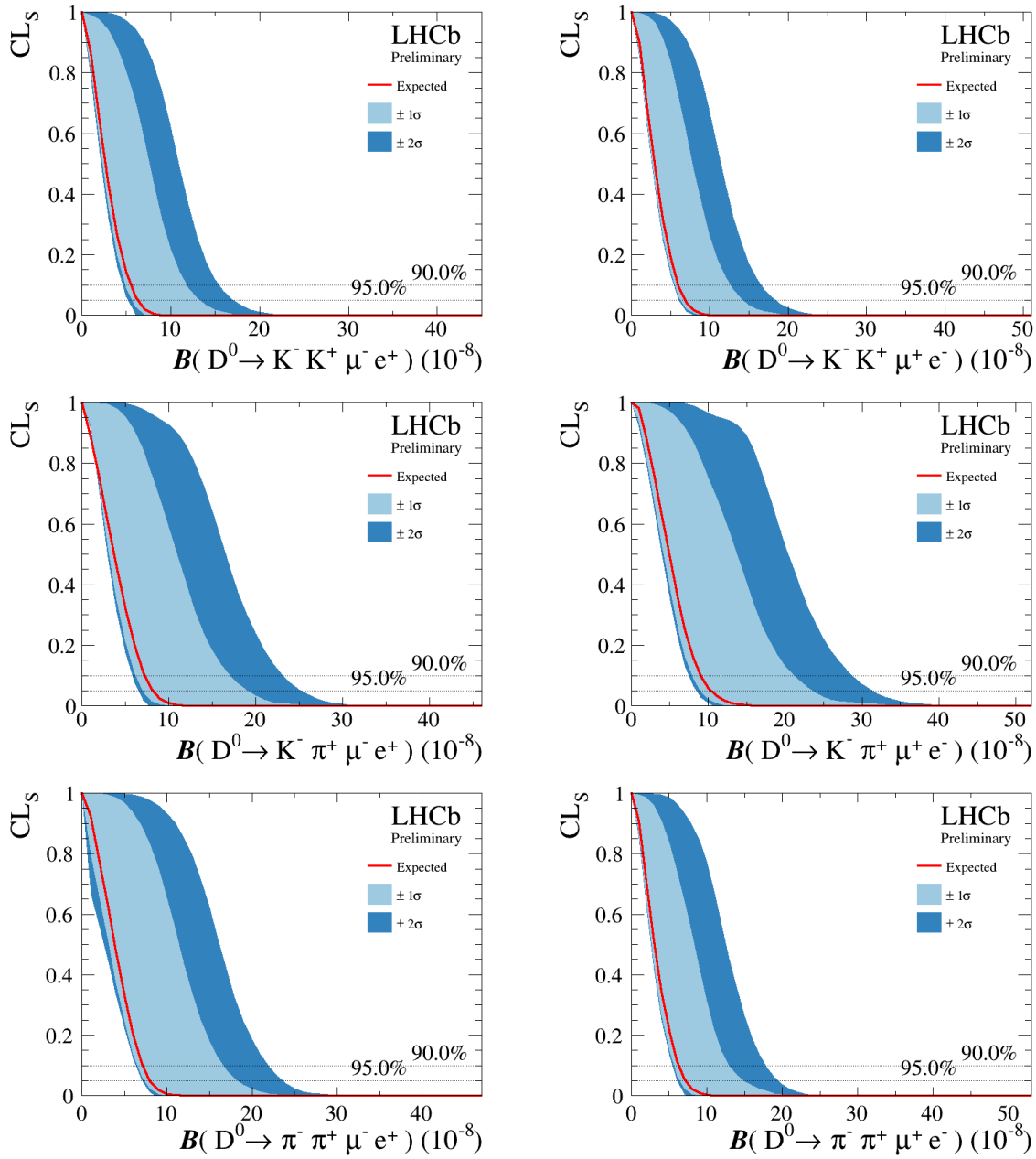


Figure 9.2: Confidence contour obtained by the CL_s method and the expected limits for the background-only hypothesis. The dotted lines marked the CL. In blue are the error bands of the expected limit shown. The left(right) column shows the confidence contours for decays with oppositely(equally) charged muon and accompanying slow tag pion. The upper row shows the results for $D^0 \to K^-K^+\mu^\pm e^\mp$ decays, the middle row for $D^0 \to K^-\pi^+\mu^\pm e^\mp$ decays and the lower row for $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ decays.

Chapter 10

Conclusion & Outlook

The first search for LFV in four-body D^0 meson decays at LHCb is presented. This search uses pp collision data collected with the LHCb detector between 2015 and 2018, corresponding to an integrated luminosity of $6\,fb^{-1}$. This measurement is still blinded; therefore, the expected sensitivity and upper limits, using the $\mathrm{CL_s}$ *method, are reported. Using the LHCb dataset, it is shown that it is possible to improve the current best limits, set by BaBar, by two orders of magnitude.*

The aim of this search of the three LFV decays, $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$, $D^0 \to K^-\pi^+\mu^\pm e^\mp$ and $D^0 \to K^-K^+\mu^\pm e^\mp$, is, in case of no significant signal, to set an upper limit on the branching ratio with the help of the $\mathrm{CL_s}$ method, or, if a signal would be observed, to perform a measurement of the branching fraction. The limits themselves help to constrain possible theory extension to the SM by probing not yet experimentally excluded regions in which new physics scenarios could manifest.

The complete selection sequence to extract possible $D^0 \to h^{(\prime)-}h^+\mu^\pm e^\mp$ decay candidates is presented. After enforcing topological and reconstruction quality requirements, the selection is optimised for the signal significance to reduce the remaining background sources. Particle identification requirements obtained from a neural network are optimised for wrongly identified particles. Simultaneously, a multivariate classifier is trained and optimised to suppress background originating from tracks wrongly combined as signal candidates. As the last step in the selection process, a fit model is constructed to separate the remaining background from the signal candidates statistically. To avoid biases in the optimisation and selection, the signal region in data is blinded.

The efficiency is evaluated on simulated events and, if needed, corrected using data.

Efficiency and observed candidate yields are combined into a branching fraction measurement with which the significance and the limit in case no signal is observed will be evaluated.

To test and validate the measurement procedure, a reference measurement of the branching fraction for $D^0 \to K^-K^+\mu^-\mu^+$ and $D^0 \to \pi^-\pi^+\mu^-\mu^+$ decays, using the same methods as for the study of the LFV decay modes, is performed. A branching fraction compatible with previous measurements is measured:

$$\mathcal{B}(D^0 \to K^-K^+\mu^-\mu^+) = (2.09 \pm 0.16 \pm 0.11 \pm 0.21) \times 10^{-7}$$
$$\mathcal{B}(D^0 \to \pi^-\pi^+\mu^-\mu^+) = (8.35 \pm 0.27 \pm 0.47 \pm 0.84) \times 10^{-7}$$

The first quoted uncertainty corresponds to the statistical uncertainty on the dataset. The second evaluated uncertainty covers all systematic uncertainties apart from the uncertainty of the normalisation branching ratio, which is quoted as the third uncertainty. The results for the same branching ratios measured with the Run 1 dataset [44] are:

$$\mathcal{B}(D^0 \to K^-K^+\mu^-\mu^+) = (1.54 \pm 0.27 \pm 0.09 \pm 0.16) \times 10^{-7}$$
$$\mathcal{B}(D^0 \to \pi^-\pi^+\mu^-\mu^+) = (9.64 \pm 0.48 \pm 0.51 \pm 0.97) \times 10^{-7}$$

The measured statistical uncertainties are a factor 1.7 smaller with respect to the published measurements. The systematic uncertainties are of similar size.

The measurement of $\mathcal{B}(D^0 \to K^-K^+\mu^-\mu^+)(\mathcal{B}(D^0 \to \pi^-\pi^+\mu^-\mu^+))$ is compatible within $-1.75(+2.07)$ standard deviations with the results of Run 1. For this comparison, only uncertainties of statistical nature are taken into account. All other uncertainties are assumed to be strongly correlated. The agreement by around two standard deviations between the two measurements suggests that the measurement procedure provides an unbiased measurement of the branching fractions. One would have hoped for a better agreement of less than one standard deviation between the two measurements. In an attempt to better understand the source of this deviation, the Run 2 measurement selection was adjusted to match the data selection applied for the Run 1 measurement. However, both the efficiency ratio and the yield ratio deviate from the Run 1 results coherently, enhancing the effect of the deviation. This makes it difficult to narrow the source of the deviation down. It is, therefore, not possible to identify an individual source of this deviation. Contrarily, this supports the hypotheses that the deviation is caused by statistical fluctuations. Possible next steps to confirm that the deviation is only of a statistical nature would be to reanalyse the Run 1 data sample using the same procedure

as was used for the Run 2 measurement.

While the branching fraction for $D^0 \to K^-K^+\mu^-\mu^+$ decays is still statistically limited, the branching fraction for $D^0 \to \pi^-\pi^+\mu^-\mu^+$ decays is systematically limited. The largest systematic is due to the unknown decay model used to generate simulated events for the efficiency estimate. This uncertainty can be reduced by studying the decay in individual phase space regions, dihadron and dilepton reconstructed mass bins, where the contribution due to the different intermediate resonances is known. Fluctuating the relative contribution of the different decay models by 50%, as was done for the Run 1 measurement, most likely over-evaluates the effect and only provides an upper limit on the uncertainty. A more sophisticated study by scaling the resonant and non-resonant contributions to the already measured branching ratios in the reconstructed dimuon mass should allow to correct the decay model partially and reduce the uncertainty. The best solution to this problem would be the implementation of a dedicated decay model of this decay in simulations, like it is already done, for example for $D^0 \to K^-\pi^+\pi^+\pi^-$ decays.

The systematic uncertainty assigned to the yield ratio is similar for this measurement with respect to the Run 1 measurement. In general, it was possible to reduce the size of most systematic uncertainties introduced by the efficiency ratio. New systematic uncertainties which were not considered in the Run 1 measurement were added, the overall systematic uncertainty did slightly increase (stayed the same) for $D^0 \to \pi^-\pi^+\mu^-\mu^+$ ($D^0 \to K^-K^+\mu^-\mu^+$) decays. One such uncertainty, for example, is the uncertainty due to hadronic interactions with the detector material, which was not considered for the Run 1 measurement. The largest systematic uncertainty is due to the limited knowledge of the normalisation mode branching ratio. To reduce this uncertainty, a new measurement of the branching ratio would be required.

For the LFV decays the signal mass window and thus the measured central value of the branching fraction is blinded. Most of the systematic uncertainties depend on the measured central value. Assuming "no observation", *i.e.* a branching fraction of zero allows to report the following uncertainties:

$$\mathcal{B}(D^0 \to K^-K^+\mu^\pm e^\mp) = (0.00 \pm 2.5 \pm 0.2) \times 10^{-8}$$
$$\mathcal{B}(D^0 \to K^-\pi^+\mu^\pm e^\mp) = (0.00 \pm 4.9 \pm 1.9) \times 10^{-8}$$
$$\mathcal{B}(D^0 \to \pi^-\pi^+\mu^\pm e^\mp) = (0.00 \pm 3.3 \pm 1.4) \times 10^{-8}$$

The first quoted uncertainty corresponds to the statistical uncertainty on the dataset.

The second uncertainty covers all systematic uncertainties. In case of an observation, the systematic uncertainties would need to be reevaluated for the observed branching ratio and would increase the uncertainty. In the following, uncertainties for the branching fraction are reported:

$$\mathcal{B}(D^0 \to K^- K^+ \mu^\pm e^\mp) = (x \pm 2.5 \pm (0.2 \textbf{ or } (2.5 - 6.4)\%) \pm 5.7\% \pm 10\%) \times 10^{-8}$$
$$\mathcal{B}(D^0 \to K^- \pi^+ \mu^\pm e^\mp) = (x \pm 4.9 \pm (1.9 \textbf{ or } (3.1 - 10.0)\%) \pm 5.3\% \pm 10\%) \times 10^{-8}$$
$$\mathcal{B}(D^0 \to \pi^- \pi^+ \mu^\pm e^\mp) = (x \pm 3.3 \pm (1.4 \textbf{ or } (2.7 - 8.0)\%) \pm 10.3\% \pm 10\%) \times 10^{-8}$$

The first quoted uncertainty corresponds to the statistical uncertainty of the dataset. For the second uncertainty, two values are provided: an absolute uncertainty, providing a lower limit for the uncertainty, and a relative uncertainty in per cent. The absolute value of the relative uncertainty is calculated by multiplying the relative uncertainty with the measured branching fraction. The "**or**" notates here that whichever absolute value of the two uncertainties is larger will be assigned as uncertainty. This uncertainty covers the systematic uncertainty due to the parameterisation of the misID background. The third uncertainty, given relative to the measured branching ratio, covers all other systematic uncertainties, apart from the misID parameterisation and the uncertainty of the normalisation branching ratio. The fourth relative uncertainty is the uncertainty due to the uncertainty of the normalisation branching fraction.

Using the CL_s method, the following expected limits at 95%(90%) CL are evaluated:

$$\mathcal{B}(D^0 \to K^- K^+ \mu^\pm e^\mp) \leq 6.5(5.5) \times 10^{-8}$$
$$\mathcal{B}(D^0 \to K^- \pi^+ \mu^\pm e^\mp) \leq 11.2(9.5) \times 10^{-8}$$
$$\mathcal{B}(D^0 \to \pi^- \pi^+ \mu^\pm e^\mp) \leq 7.8(6.6) \times 10^{-8}$$

With the expected limit obtained from the LHCb dataset for a non observation, the current best limits, set by BaBar, will be improved by two orders of magnitude.

To be able to set more precise limits for new physics theories, the data samples are split depending on the relative charges in the final state particle composition. In this way, the flavour of the D^0 meson is directly linked to the lepton flavour. Using the CL_s method, the following limits are expected at 95%(90%) CL for muons with the same charge as the

accompanying slow tag pion (the D^0 and muon have the opposite flavour):

$$\mathcal{B}(D^0 \to K^+K^-\mu^+e^-) \leq 5.5(4.8) \times 10^{-8}$$
$$\mathcal{B}(D^0 \to K^-\pi^+\mu^+e^-) \leq 9.0(8.1) \times 10^{-8}$$
$$\mathcal{B}(D^0 \to \pi^+\pi^-\mu^+e^-) \leq 5.7(4.9) \times 10^{-8}$$

The expected 95%(90%) CL limits for muons with the opposite charge as the accompanying slow tag pion (the D^0 and muon have the same flavour) are:

$$\mathcal{B}(D^0 \to K^+K^-\mu^-e^+) \leq 4.5(3.9) \times 10^{-8}$$
$$\mathcal{B}(D^0 \to K^-\pi^+\mu^-e^+) \leq 6.7(6.0) \times 10^{-8}$$
$$\mathcal{B}(D^0 \to \pi^+\pi^-\mu^-e^+) \leq 6.4(5.8) \times 10^{-8}$$

No previous measurement for the differently charged final states exists.

According to theoretical calculations [46], new physics is not yet excluded for:

$$\mathcal{B}(D^0 \to \pi^-\pi^+\mu^\pm e^\mp) \lesssim 10^{-7}$$
$$\mathcal{B}(D^0 \to K^-K^+\mu^\pm e^\mp) \lesssim 10^{-9}.$$

The above calculated expected limit for $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ decays shows that it is possible to further restrict possible new physics scenarios with the LHCb Run 2 dataset. $D^0 \to K^-K^+\mu^\pm e^\mp$ decays above 10^{-9} are already excluded with the help of auxiliary measurements. Nevertheless, this measurement will provide an important direct input. For $D^0 \to K^-\pi^+\mu^\pm e^\mp$ decays, no such calculation was performed.

The largest systematic uncertainties are introduced by misidentified hadronic decays of the form $D^0 \to h^{(\prime)-}h^+\pi^-\pi^+$. The reason for this is that candidates where a bremsstrahlung photon is lost are difficult to separate from the misidentified hadronic decays. However, the understanding of the misID background scales similar to the understanding of the signal decays with the recorded data sample. It will, therefore, with today's understanding, not be a limiting factor for future measurements.

Two of the three largest uncertainties introduced through the efficiency ratio are due to the size and quality of the simulation samples used to estimate the efficiency. The quality of the simulation sample can be improved by the implementation of a dedicated decay model using data of already existing measurements [44, 61]. The sample size of the simulated sample can be increased by generating more events. However, a full detector

simulation was used in this thesis, and no fast simulation methods were used. One would need to change this for the generation of significantly larger simulation samples. The uncertainty on the decay model is introduced by the normalisation mode, while the uncertainty due to the limited statistic is driven by the signal modes. The third large uncertainty introduced by the efficiency ratio is the uncertainty due to hadronic interaction with the detector material. During the evaluation of this uncertainty, it is assumed that the material per track is known to a precision of 10%. Ideally, one would like to directly measure the effect of the hadronic interactions instead of performing an approximation.

If, after unblinding, a signal would be observed by this measurement, it would be crucial to perform a second independent measurement, which could be performed with the upgraded, mostly new LHCb detector using data recorded in Run 3. If no signal is observed, redoing the measurement with a larger dataset, without external input indicating differently, would only be sensible if the limit could be increased by another order of magnitude. This would require a by a factor hundred larger dataset, which is not achievable in the near future. However, a feasible next step would be to use the understanding gained here for four body decays containing muon and electrons to perform a lepton flavour universality (LFU) measurement between $D^0 \to h^+h^-e^+e^-$ and $D^0 \to h^+h^-\mu^+\mu^-$ decays:

$$R^c_{hh} = \frac{\int_{q^2_{min}}^{q^2_{max}} \frac{d\mathcal{B}(D^0 \to h^+h^-\mu^+\mu^-)}{dq^2}\, dq^2}{\int_{q^2_{min}}^{q^2_{max}} \frac{d\mathcal{B}(D^0 \to h^+h^-e^+e^-)}{dq^2}\, dq^2}$$

In case of Lepton Flavour conservation, the ratio R^c_{hh} is expected to be equal to one. This would complement the already performed searches in the down-type quark sector [13, 14]. The advantage of performing a LFU measurement is that most theoretical uncertainties cancel in the ratio. Based on the results in this thesis, especially the relative efficiency estimates between muon and electrons, combined with the already published measurement of the observed $D^0 \to K^-K^+\mu^-\mu^+$ and $D^0 \to \pi^-\pi^+\mu^-\mu^+$ decays [44] and the expected results from the ongoing search for $D^0 \to K^+K^-e^+e^-$ and $D^0 \to \pi^+\pi^-e^+e^-$ decays [52] a measurement of LFU in four body charm decays could be feasible in the near future by using the larger Run 3 dataset. It is worth noting that the LFV decay modes, similar to the dimuon decay modes, were measured relative to $D^0 \to K^-\pi^+[\mu^+\mu^-]_{\rho^0/\omega}$ decays. The dielectron modes are measured relative to $D^0 \to K^-\pi^+[e^+e^-]_{\rho^0/\omega}$. In this thesis, muon and electron candidates were studied relative to each other, which provides the linking component.

More directly, with this thesis it was possible to show that with the LHCb dataset, the current best limits on the branching fraction of $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$, $D^0 \to K^-\pi^+\mu^\pm e^\mp$ and $D^0 \to K^-K^+\mu^\pm e^\mp$ decays set by BaBar will be improved by two orders of magnitude. In addition, an updated measurement of the branching ratio for $D^0 \to K^-K^+\mu^-\mu^+$ and $D^0 \to \pi^-\pi^+\mu^-\mu^+$ decays was performed, which provides an important validation for the still-blind search. Lastly, additional limits considering the flavour composition of the LFV decays are provided to further limit new physics scenarios.

Bibliography

[1] ATLAS, G. 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) 1, arXiv:1207.7214.

[2] CMS, S. 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) 30, arXiv:1207.7235.

[3] G. R. Farrar and M. E. Shaposhnikov, *Baryon asymmetry of the universe in the minimal standard model*, Phys. Rev. Lett. **70** (1993) 2833.

[4] A. D. Sakharov, *Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe*, Pisma Zh. Eksp. Teor. Fiz. **5** (1967) 32, [Usp. Fiz. Nauk161,no.5,61(1991)].

[5] V. Simha and G. Steigman, *Constraining The Universal Lepton Asymmetry*, JCAP **08** (2008) 011, arXiv:0806.0179.

[6] V. A. Kuzmin, V. A. Rubakov, and M. E. Shaposhnikov, *On the Anomalous Electroweak Baryon Number Nonconservation in the Early Universe*, Phys. Lett. B **155** (1985) 36.

[7] LHCb, R. Aaij *et al.*, *Measurements of prompt charm production cross-sections in pp collisions at* $\sqrt{s} = 13$ *TeV*, JHEP **03** (2016) 159, arXiv:1510.01707, [Erratum: JHEP 09, 013 (2016), Erratum: JHEP 05, 074 (2017)].

[8] LHCb collaboration, R. Aaij *et al.*, *Test of lepton universality using* $B^+ \to K^+\ell^+\ell^-$ *decays*, Phys. Rev. Lett. **113** (2014) 151601, arXiv:1406.6482.

[9] LHCb collaboration, R. Aaij *et al.*, *Test of lepton universality with* $B^0 \to K^{*0}\ell^+\ell^-$ *decays*, JHEP **08** (2017) 055, arXiv:1705.05802.

[10] LHCb, R. Aaij *et al.*, *Test of lepton universality with* $\Lambda_b^0 \to pK^-\ell^+\ell^-$ *decays*, JHEP **05** (2020) 040, arXiv:1912.08139.

[11] LHCb, R. Aaij *et al.*, *Tests of lepton universality using $B^0 \to K^0_S \ell^+\ell^-$ and $B^+ \to K^{*+}\ell^+\ell^-$ decays*, Phys. Rev. Lett. **128** (2022) 191802, arXiv:2110.09501.

[12] LHCb, R. Aaij *et al.*, *Test of lepton universality in beauty-quark decays*, Nature Phys. **18** (2022) 277, arXiv:2103.11769, [Addendum: Nature Phys. 19, (2023)].

[13] LHCb, R. Aaij *et al.*, *Measurement of lepton universality parameters in $B^+ \to K^+\ell^+\ell^-$ and $B^0 \to K^{*0}\ell^+\ell^-$ decays*, Phys. Rev. D **108** (2023) 032002, arXiv:2212.09153, All figures and tables, along with any supplementary material and additional information, are available at https://cern.ch/lhcbproject/Publications/p/LHCb-PAPER-2022-045.html (LHCb public pages).

[14] LHCb, R. Aaij *et al.*, *Test of lepton universality in $b \to s\ell^+\ell^-$ decays*, Phys. Rev. Lett. **131** (2023) 051803, arXiv:2212.09152, All figures and tables, along with any supplementary material and additional information, are available at https://cern.ch/lhcbproject/Publications/p/LHCb-PAPER-2022-046.html (LHCb public pages).

[15] S. de Boer and G. Hiller, *Flavor and new physics opportunities with rare charm decays into leptons*, Phys. Rev. D **93** (2016) 074001, arXiv:1510.00311.

[16] LHCb collaboration, R. Aaij *et al.*, *Search for the lepton-flavour violating decay $D^0 \to e^\pm\mu^\mp$*, Phys. Lett. **B754** (2016) 167, arXiv:1512.00322.

[17] LHCb, R. Aaij *et al.*, *Searches for 25 rare and forbidden decays of D^+ and D^+_s mesons*, JHEP **06** (2021) 044, arXiv:2011.00217.

[18] BaBar, J. P. Lees *et al.*, *Search for Rare or Forbidden Decays of the D^0 Meson*, Phys. Rev. Lett. **124** (2020) 071802, arXiv:1905.00608.

[19] V. C. Rubin, N. Thonnard, and W. K. Ford, Jr. *Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 /R = 4kpc/ to UGC 2885 /R = 122 kpc/*, Astrophys. J. **238** (1980) 471.

[20] K. G. Begeman, A. H. Broeils, and R. H. Sanders, *Extended rotation curves of spiral galaxies: Dark haloes and modified dynamics*, Mon. Not. Roy. Astron. Soc. **249** (1991) 523.

[21] M. Bartelmann and P. Schneider, *Weak gravitational lensing*, Phys. Rept. **340** (2001) 291, arXiv:astro-ph/9912508.

[22] G. Steigman, *Observational tests of antimatter cosmologies*, Annual Review of Astronomy and Astrophysics **14** (1976) 339.

[23] D. S. Mitzel, *Search for new physics in rare four-body charm decays at LHCb*, 2019. Presented 19 Dec 2018.

[24] S. L. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys. **22** (1961) 579.

[25] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264.

[26] A. Salam, *Weak and Electromagnetic Interactions*, Conf. Proc. **C680519** (1968) 367.

[27] M. E. Peskin and D. V. Schroeder, *An Introduction to quantum field theory*, Addison-Wesley, Reading, USA, 1995.

[28] E. Noether, *Invariant Variation Problems*, Gott. Nachr. **1918** (1918) 235, arXiv:physics/0503066, [Transp. Theory Statist. Phys.1,186(1971)].

[29] Wikimedia Commons, Cush, *Standard model of elementary particles*, .

[30] LHCb collaboration, R. Aaij *et al.*, *Observation of a narrow $P_c(4312)^+$ state and of two-peak structure of the $P_c(4450)^+$ in $\Lambda_b^0 \to J/\psi p K^-$ decays*, LHCb-PAPER-2019-014 CERN-EP-2019-058, in preparation.

[31] LHCb, R. Aaij *et al.*, *Observation of an exotic narrow doubly charmed tetraquark*, Nature Phys. **18** (2022) 751, arXiv:2109.01038.

[32] N. Cabibbo, *Unitary symmetry and leptonic decays*, Phys. Rev. Lett. **10** (1963) 531.

[33] M. Kobayashi and T. Maskawa, *CP-Violation in the Renormalizable Theory of Weak Interaction*, Progress of Theoretical Physics **49** (1973) 652, arXiv:https://academic.oup.com/ptp/article-pdf/49/2/652/5257692/49-2-652.pdf.

[34] L. Wolfenstein, *Parametrization of the Kobayashi-Maskawa Matrix*, Phys. Rev. Lett. **51** (1983) 1945.

[35] Particle Data Group, R. L. Workman and Others, *Review of Particle Physics*, PTEP **2022** (2022) 083C01.

[36] Super-Kamiokande, Y. Fukuda *et al.*, *Evidence for oscillation of atmospheric neutrinos*, Phys. Rev. Lett. **81** (1998) 1562, arXiv:hep-ex/9807003.

[37] XENON, E. Aprile *et al.*, *The XENON1T Dark Matter Experiment*, Eur. Phys. J. C **77** (2017) 881, arXiv:1708.07051.

[38] XENON, E. Aprile *et al.*, *Physics reach of the XENON1T dark matter experiment*, JCAP **04** (2016) 027, arXiv:1512.07501.

[39] DARWIN, J. Aalbers *et al.*, *DARWIN: towards the ultimate dark matter detector*, JCAP **11** (2016) 017, arXiv:1606.07001.

[40] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay, *Evidence for the* 2π *Decay of the* K_2^0 *Meson*, Phys. Rev. Lett. **13** (1964) 138.

[41] S. L. Glashow, J. Iliopoulos, and L. Maiani, *Weak Interactions with Lepton-Hadron Symmetry*, Phys. Rev. D **2** (1970) 1285.

[42] SLAC-SP-017, J. E. Augustin *et al.*, *Discovery of a Narrow Resonance in* e^+e^- *Annihilation*, Phys. Rev. Lett. **33** (1974) 1406.

[43] E. G. Cazzoli *et al.*, *Evidence for Delta S = - Delta Q Currents or Charmed Baryon Production by Neutrinos*, Phys. Rev. Lett. **34** (1975) 1125.

[44] LHCb collaboration, R. Aaij *et al.*, *Observation of* D^0 *meson decays to* $\pi^+\pi^-\mu^+\mu^-$ *and* $K^+K^-\mu^+\mu^-$ *final states*, Phys. Rev. Lett. **119** (2017) 181805, arXiv:1707.08377.

[45] H. Gisbert, M. Golz, and D. S. Mitzel, *Theoretical and experimental status of rare charm decays*, Mod. Phys. Lett. A **36** (2021) 2130002, arXiv:2011.09478.

[46] S. De Boer and G. Hiller, *Null tests from angular distributions in* $D \to P_1P_2l^+l^-$, $l = e, \mu$ *decays on and off peak*, Phys. Rev. D **98** (2018) 035041, arXiv:1805.08516.

[47] LHCb, R. Aaij *et al.*, *Angular Analysis of* $D^0 \to \pi^+\pi^-\mu^+\mu^-$ *and* $D^0 \to K^+K^-\mu^+\mu^-$ *Decays and Search for CP Violation*, Phys. Rev. Lett. **128** (2022) 221801, arXiv:2111.03327.

[48] L. Cappiello, O. Catà, and G. D'Ambrosio, *Standard model prediction and new physics tests for* $D^0 \to h_1^+h_2^-\ell^+\ell^-$ *(*$h = \pi, K$*;* $\ell = e, \mu$*)*, JHEP **04** (2013) 135, arXiv:1209.4235.

[49] BaBar, J. P. Lees *et al.*, *Search for the Decay* $D^0- > \gamma\gamma$ *and Measurement of the Branching Fraction for* $D^0- > \pi^0\pi^0$, Phys. Rev. D **85** (2012) 091107, arXiv:1110.6480.

[50] G. Burdman, E. Golowich, J. L. Hewett, and S. Pakvasa, *Rare charm decays in the standard model and beyond*, Phys. Rev. D **66** (2002) 014009, arXiv:hep-ph/0112235.

[51] LHCb collaboration, R. Aaij *et al.*, *Search for the rare decay $D^0 \to \mu^+\mu^-$*, Phys. Lett. **B725** (2013) 15, arXiv:1305.5059.

[52] A. Scarabotto, *Search for rare four-body charm decays with electrons in the final state and long track reconstruction for the LHCb trigger*, 2023. Presented 01 Sep 2023.

[53] M. Neubert, *Effective field theory and heavy quark physics*, in *Physics in $D \geq 4$. Proceedings, Theoretical Advanced Study Institute in elementary particle physics, TASI 2004, Boulder, USA, June 6-July 2, 2004*, 149–194, 2005, arXiv:hep-ph/0512222.

[54] J. Polchinski, *Effective field theory and the Fermi surface*, in *Theoretical Advanced Study Institute (TASI 92): From Black Holes and Strings to Particles*, 0235–276, 1992, arXiv:hep-th/9210046.

[55] K. G. Wilson, *Nonlagrangian models of current algebra*, Phys. Rev. **179** (1969) 1499.

[56] K. G. Chetyrkin, M. Misiak, and M. Munz, *Weak radiative B meson decay beyond leading logarithms*, Phys. Lett. **B400** (1997) 206, arXiv:hep-ph/9612313, [Erratum: Phys. Lett.B425,414(1998)].

[57] C. Bobeth, M. Misiak, and J. Urban, *Photonic penguins at two loops and m_t dependence of $BR[B \to X_s\ell^+\ell^-]$*, Nucl. Phys. **B574** (2000) 291, arXiv:hep-ph/9910220.

[58] P. Gambino, M. Gorbahn, and U. Haisch, *Anomalous dimension matrix for radiative and rare semileptonic B decays up to three loops*, Nucl. Phys. **B673** (2003) 238, arXiv:hep-ph/0306079.

[59] LHCb, R. Aaij *et al.*, *Absolute luminosity measurements with the LHCb detector at the LHC*, JINST **7** (2012) P01010, arXiv:1110.2866.

[60] LHCb, R. Aaij *et al.*, *Precision luminosity measurements at LHCb*, JINST **9** (2014) P12005, arXiv:1410.0149.

[61] LHCb collaboration, R. Aaij *et al.*, *First observation of the decay $D^0 \to K^-\pi^+\mu^+\mu^-$ in the $\rho^0 - \omega$ region of the dimuon mass spectrum*, Phys. Lett. **B757** (2016) 558, arXiv:1510.08367.

[62] L. R. Evans and P. Bryant, *LHC Machine*, JINST **3** (2008) S08001. 164 p, This report is an abridged version of the LHC Design Report (CERN-2004-003).

[63] LHCb collaboration, *LHCb reoptimized detector design and performance: Technical Design Report*, CERN-LHCC-2003-030, 2003.

[64] LHCb collaboration, R. Aaij *et al.*, *Measurement of the b-quark production cross-section in 7 and 13* TeV *pp collisions*, Phys. Rev. Lett. **118** (2017) 052002, Erratum ibid. **119** (2017) 169901, arXiv:1612.05140.

[65] T. Sjöstrand *et al.*, *An introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191** (2015) 159, arXiv:1410.3012.

[66] H. L. Lai *et al.*, *Improved parton distributions from global analysis of recent deep inelastic scattering and inclusive jet data*, Phys. Rev. D **55** (1997) 1280, arXiv:hep-ph/9606399.

[67] LHCb collaboration, C. Elsasser, *$\bar{b}b$ production angle plots*, .

[68] LHCb Collaboration, A. A. Alves *et al.*, *The LHCb Detector at the LHC*, JINST **3** (2008) S08005, Also published by CERN Geneva in 2010.

[69] LHCb, *Lhcb webpage*, .

[70] LHCb Outer Tracker Group, R. Arink *et al.*, *Performance of the LHCb Outer Tracker*, JINST **9** (2014) P01002, arXiv:1311.3893.

[71] M. Adinolfi *et al.*, *Performance of the LHCb RICH detector at the LHC*, Eur. Phys. J. **C73** (2013) 2431, arXiv:1211.6759.

[72] LHCb, A. A. Alves, Jr. *et al.*, *The LHCb Detector at the LHC*, JINST **3** (2008) S08005.

[73] J. R. Harrison, *Radiation damage studies in the LHCb VELO detector and searches for lepton flavour and baryon number violating tau decays*, PhD thesis, Manchester U., 2014.

[74] M. Thomson, *Modern particle physics*, Cambridge University Press, New York, 2013.

[75] LHCb collaboration, R. Aaij *et al.*, *Measurement of the electron reconstruction efficiency at LHCb*, JINST **14** (2019) P11023, arXiv:1909.02957.

[76] LHCb collaboration, A. A. Alves Jr. *et al.*, *The LHCb detector at the LHC*, JINST **3** (2008) S08005.

[77] R. Aaij *et al.*, *Performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC*, JINST **14** (2019) P04013, arXiv:1812.10790.

[78] T. Sjostrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **05** (2006) 026, arXiv:hep-ph/0603175.

[79] LHCb collaboration, *The Gauss project*, http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/gauss/.

[80] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. A **462** (2001) 152.

[81] GEANT4, S. Agostinelli *et al.*, *GEANT4–a simulation toolkit*, Nucl. Instrum. Meth. A **506** (2003) 250.

[82] LHCb collaboration, *The Boole project*, http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/boole/.

[83] LHCb collaboration, *The Moore project*, http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/moore/.

[84] LHCb collaboration, *The Brunel project*, http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/brunel/.

[85] LHCb collaboration, *The daVinci project*, http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/davinci/.

[86] G. Punzi, *Sensitivity of searches for new signals and its optimization*, eConf **C030908** (2003) MODT002, arXiv:physics/0308063.

[87] M. Paterno, *Calculating efficiencies and their uncertainties*, .

[88] Höcker, Andreas and Speckmayer, Peter and Stelzer, Jörg and Tegenfeldt, Fredrik and Voss, Helge, *TMVA: The Toolkit for Multivariate Data Analysis with ROOT*, .

[89] TMVA, A. Hocker *et al.*, *TMVA - Toolkit for Multivariate Data Analysis*, arXiv:physics/0703039.

[90] J. H. Friedman, *Greedy function approximation: A gradient boosting machine.*, Annals Statist. **29** (2001) 1189.

[91] LHCb collaboration, R. Aaij *et al.*, *LHCb detector performance*, Int. J. Mod. Phys. **A30** (2015) 1530022, arXiv:1412.6352.

[92] L. Anderlini *et al.*, *The PIDCalib package* , CERN, Geneva, 2016.

[93] A. Poluektov, *Correction of simulated particle identification response in LHCb using kernel density estimation.* , CERN, Geneva, 2017.

[94] A. Poluektov, *Kernel density estimation of a multidimensional efficiency profile*, JINST **10** (2015) P02011, arXiv:1411.5528.

[95] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, The European Physical Journal C **71** (2011) .

[96] T. Mombächer, *Beautiful leptons - setting limits to New Physics with the LHCb experiment*, PhD thesis, Tech. U., Dortmund (main), Tech. U., Dortmund (main), 2020, doi: 10.17877/DE290R-21750.

[97] S. S. Wilks, *The large-sample distribution of the likelihood ratio for testing composite hypotheses*, Ann. Math. Stat. **9** (1938) 60.

[98] A. Wald, *Tests of statistical hypotheses concerning several parameters when the number of observations is large*, Transactions of the American Mathematical Society 54 no. 3 426 (1943).

[99] A. L. Read, *Presentation of search results: The CL_s technique*, J. Phys. G **28** (2002) 2693.

[100] C.-S. Chuang and T. L. Lai, *Hybrid Resampling Methods for Confidence Intervals*, Statist. Sinica **10** (2000) 1.

[101] S. Bodhisattva, M. Walker, and M. Woodroofe, *On the Unified Method with Nuisance Parameters*, Statist. Sinica **19** (2009) 301.

[102] Particle Data Group, M. Tanabashi *et al.*, *Review of particle physics*, Phys. Rev. **D98** (2018) 030001.

[103] W. D. Hulsbergen, *Decay chain fitting with a Kalman filter*, Nucl. Instrum. Meth. **A552** (2005) 566, arXiv:physics/0503191.

[104] Particle Data Group, C. Patrignani *et al.*, *Review of particle physics*, Chin. Phys. **C40** (2016) 100001.

[105] F. Archilli *et al.*, *Performance of the muon identification at LHCb*, JINST **8** (2013) P10020, arXiv:1306.0249.

[106] L. Chakravarti and Roy, *Handbook of Methods of Applied Statistics, Volume I*, John Wiley and Sons, 1967.

[107] LHCb collaboration, *Statistics table*, .

[108] M. Pivk and F. R. Le Diberder, *sPlot: A statistical tool to unfold data distributions*, Nucl. Instrum. Meth. **A555** (2005) 356, arXiv:physics/0402083.

[109] S. Tolk, J. Albrecht, F. Dettori, and A. Pellegrino, *Data driven trigger efficiency determination at LHCb* , CERN, Geneva, 2014.

[110] N. L. Johnson, *Systems of frequency curves generated by methods of translation*, Biometrika **36** (1949) 149.

[111] T. Skwarnicki, *A study of the radiative cascade transitions between the Upsilon-prime and Upsilon resonances*, PhD thesis, Institute of Nuclear Physics, Krakow, 1986, DESY-F31-86-02.

[112] B. Efron, *Bootstrap Methods: Another Look at the Jackknife*, Annals Statist. **7** (1979) 1.

[113] G. A. Cowan, D. C. Craik, and M. D. Needham, *RapidSim: an application for the fast simulation of heavy-quark hadron decays*, Comput. Phys. Commun. **214** (2017) 239, arXiv:1612.07489.

[114] A. D. Bukin, *Fitting function for asymmetric peaks*, 2007.

[115] M. De Cian *et al.*, *Measurement of the track finding efficiency* , CERN, Geneva, 2012.

[116] G. A. Cowan, D. C. Craik, and M. D. Needham, *RapidSim: an application for the fast simulation of heavy-quark hadron decays*, Comput. Phys. Commun. **214** (2017) 239, arXiv:1612.07489.

[117] LHCb collaboration, R. Aaij *et al.*, *Measurement of the track reconstruction efficiency at LHCb*, JINST **10** (2015) P02007, arXiv:1408.1251.

Appendix A

Detailed selection steps

Detailed tables on the selection explained in chapter 5 are provided, split up into HLT2 trigger requirements, table A.2, stripping requirements, table A.1, and additional selection steps, table A.3.

Particle	Variable		signal modes		normalisation	
			2015-2016 (S24r1,S28r1)	2017-2018 (S29r2,S34)	2015-2016	2017-2018
h, μ, e	p	$>$	$3\,\text{GeV}/c$	$3\,\text{GeV}/c$	$3\,\text{GeV}/c$	$3\,\text{GeV}/c$
	p_T	$>$	$300\,\text{MeV}/c$	$300\,\text{MeV}/c$	$300\,\text{MeV}/c$	$300\,\text{MeV}/c$
	Track χ^2/dof	$<$	3	3	3	3
μ	IP χ^2	$>$	3	3	3	3
	`isMuon`		true	true	true	true
e	IP χ^2	$>$	3	3	3	3
	DLL_e	$>$	-2	-2		
K	DLL_K	$>$	-5	-5	-5	-5
$(hh\ell\ell)$	m$(hh\ell\ell)$	$>$	$m_{\text{PDG}} - 120\,\text{MeV}/c^2$	$m_{\text{PDG}} - 240\,\text{MeV}/c^2$	$m_{\text{PDG}} - 120\,\text{MeV}/c^2$	$m_{\text{PDG}} - 240\,\text{MeV}/c^2$
		$<$	$m_{\text{PDG}} + 120\,\text{MeV}/c^2$	$m_{\text{PDG}} + 240\,\text{MeV}/c^2$	$m_{\text{PDG}} + 120\,\text{MeV}/c^2$	$m_{\text{PDG}} + 240\,\text{MeV}/c^2$
	max DOCA	$<$	0.3 mm	0.3 mm	0.3 mm	0.3 mm
	max IP χ^2	$>$	9	9	9	9
D^0	m(D^0)	$>$	$m_{\text{PDG}} - 100\,\text{MeV}/c^2$	$m_{\text{PDG}} - 220\,\text{MeV}/c^2$	$m_{\text{PDG}} - 100\,\text{MeV}/c^2$	$m_{\text{PDG}} - 220\,\text{MeV}/c^2$
		$<$	$m_{\text{PDG}} + 100\,\text{MeV}/c^2$	$m_{\text{PDG}} + 220\,\text{MeV}/c^2$	$m_{\text{PDG}} + 100\,\text{MeV}/c^2$	$m_{\text{PDG}} + 220\,\text{MeV}/c^2$
	Vertex χ^2/dof	$<$	20	20	20	20
	FD χ^2	$>$	30	16	30	16
	DIRA	$>$	0.9998	0.9999	0.9998	0.9999
	IP χ^2	$<$	36	36	36	36
	p_T	$>$	$2\,\text{GeV}/c$	$2\,\text{GeV}/c$	$2\,\text{GeV}/c$	$2\,\text{GeV}/c$
	p	$>$	$3\,\text{GeV}/c$	$3\,\text{GeV}/c$	$3\,\text{GeV}/c$	$3\,\text{GeV}/c$
π_s	p_T	$>$	$120\,\text{MeV}/c$	$120\,\text{MeV}/c$	$120\,\text{MeV}/c$	$120\,\text{MeV}/c$
$D^0\pi_s^+$	Δm	$>$	$136.4\,\text{MeV}/c^2$	$125.4\,\text{MeV}/c^2$	$136.4\,\text{MeV}/c^2$	$125.4\,\text{MeV}/c^2$
		$<$	$165.4\,\text{MeV}/c^2$	$185.4\,\text{MeV}/c^2$	$165.4\,\text{MeV}/c^2$	$185.4\,\text{MeV}/c^2$
	p_T	$>$	$2\,\text{GeV}/c$	$2\,\text{GeV}/c$	$2\,\text{GeV}/c$	$2\,\text{GeV}/c$
	max DOCA	$<$	0.3 mm	0.3 mm	0.3 mm	0.3 mm
D^{*+}	Δm	$>$	$137.4\,\text{MeV}/c^2$	$129.4\,\text{MeV}/c^2$	$137.4\,\text{MeV}/c^2$	$129.4\,\text{MeV}/c^2$
		$<$	$163.4\,\text{MeV}/c^2$	$181.4\,\text{MeV}/c^2$	$163.4\,\text{MeV}/c^2$	$181.4\,\text{MeV}/c^2$
	Vertex χ^2/dof	$<$	20	20	20	20

Table A.1: Stripping selection requirements.

Particle	Variable		signal modes		normalisation	
			2015-2016	2017-2018	2015-2016	2017-2018
h, μ, e	p	$>$	3 GeV/c	3 GeV/c	3 GeV/c	3 GeV/c
	p_T	$>$	300 MeV/c	300 MeV/c	300 MeV/c	300 MeV/c
	Track χ^2/dof	$<$	4	4	4	4
μ	IP χ^2	$>$	2	2	2	2
e	IP χ^2	$>$	3	3	3	3
$(\ell\ell)$	m$(\ell\ell)$	$<$	2100 MeV/c^2	2100 MeV/c^2	2100 MeV/c^2	2100 MeV/c^2
	$\sum p_T$	$>$	0. MeV/c	0. MeV/c	0. MeV/c	0. MeV/c
	DOCA	$<$	0.1 mm	0.1 mm	0.1 mm	0.1 mm
Dilepton object	FD χ^2	$>$	20	20	9	9
	FD	$>$	0 mm	0 mm	0 mm	0 mm
	$M_{corrected}$	$<$	3500 MeV/c^2	3500 MeV/c^2	3500 MeV/c^2	3500 MeV/c^2
$(hh\ell\ell)$	m$(hh\ell\ell)$	$>$	–	1550 MeV/c^2	–	1550 MeV/c^2
		$<$	2100 MeV/c^2	2200 MeV/c^2	2100 MeV/c^2	2200 MeV/c^2
	max p_T	$>$	0. MeV/c	0. MeV/c	0. MeV/c	0. MeV/c
	$\sum p_T$	$>$	3000 MeV/c	3000 MeV/c	3000 MeV/c	3000 MeV/c
	min DOCA	$<$	0.1 mm	0.2 mm	0.1 mm	0.2 mm
	max DOCA	$<$	0.2 mm	0.3 mm	0.2 mm	0.3 mm
	max IP χ^2	$>$	9	9	9	9
D^0	m(D^0)	$>$	1800 MeV/c^2	1700 MeV/c^2	1800 MeV/c^2	1700 MeV/c^2
		$<$	1950 MeV/c^2	2050 MeV/c^2	1950 MeV/c^2	2050 MeV/c^2
	Vertex χ^2/dof	$<$	15	15	15	15
	$M_{corrected}$	$<$	3500 MeV/c^2	3500 MeV/c^2	3500 MeV/c^2	3500 MeV/c^2
	FD χ^2	$>$	49	49	36	16
	DIRA	$>$	0.9999	0.9999	0.9999	0.9999
	IP χ^2	$<$	25	25	25	25
	$\sum \sqrt{IP\chi^2}$	$>$	12	8	12	8
π_s	p_T	$>$	–	120 MeV/c	–	120 MeV/c
$(D^0\pi_s)$	Q	$>$	–	130 MeV/c^2-m_π	–	130 MeV/c^2-m_π
		$<$	–	180 MeV/c^2-m_π	–	180 MeV/c^2-m_π
D^{*+}	Q	$>$	–	130 MeV/c^2-m_π	–	130 MeV/c^2-m_π
		$<$	–	170 MeV/c^2-m_π	–	170 MeV/c^2-m_π
	Vertex χ^2/dof	$<$	–	25	–	25

Table A.2: Requirements for D^0 (2015/2016) and D^{*+} (2017/2018) candidates of the HLT2 selection.

particle	Variable	$D^0 \to h^+h^-\mu^\pm e^\mp$	$D^0 \to K^-\pi^+\mu^+\mu^-$
K, π, μ	`TRACKGhostProb`	< 0.3	< 0.3
μ	`NShared`	$== 0$	$== 0$
π	`ProbNNpi`	> 0.5	> 0.5
K	`ProbNNK`	> 0.5	> 0.5
D^0	IPχ^2	< 9	< 9
D^0	`FD_CHI2`	> 49	> 49
D^{*+}	Δm	$> 144.5\,\mathrm{MeV}/c^2$	$> 144.5\,\mathrm{MeV}/c^2$
		$< 146.5\,\mathrm{MeV}/c^2$	$< 146.5\,\mathrm{MeV}/c^2$

Table A.3: Additional selection requirements.

Appendix B

Additional training variable distributions of the BDT

In this section, all variable distributions used to train the BDT for signal and background proxy, normalised and overlaid, are shown. For a detailed description, see chapter 5.

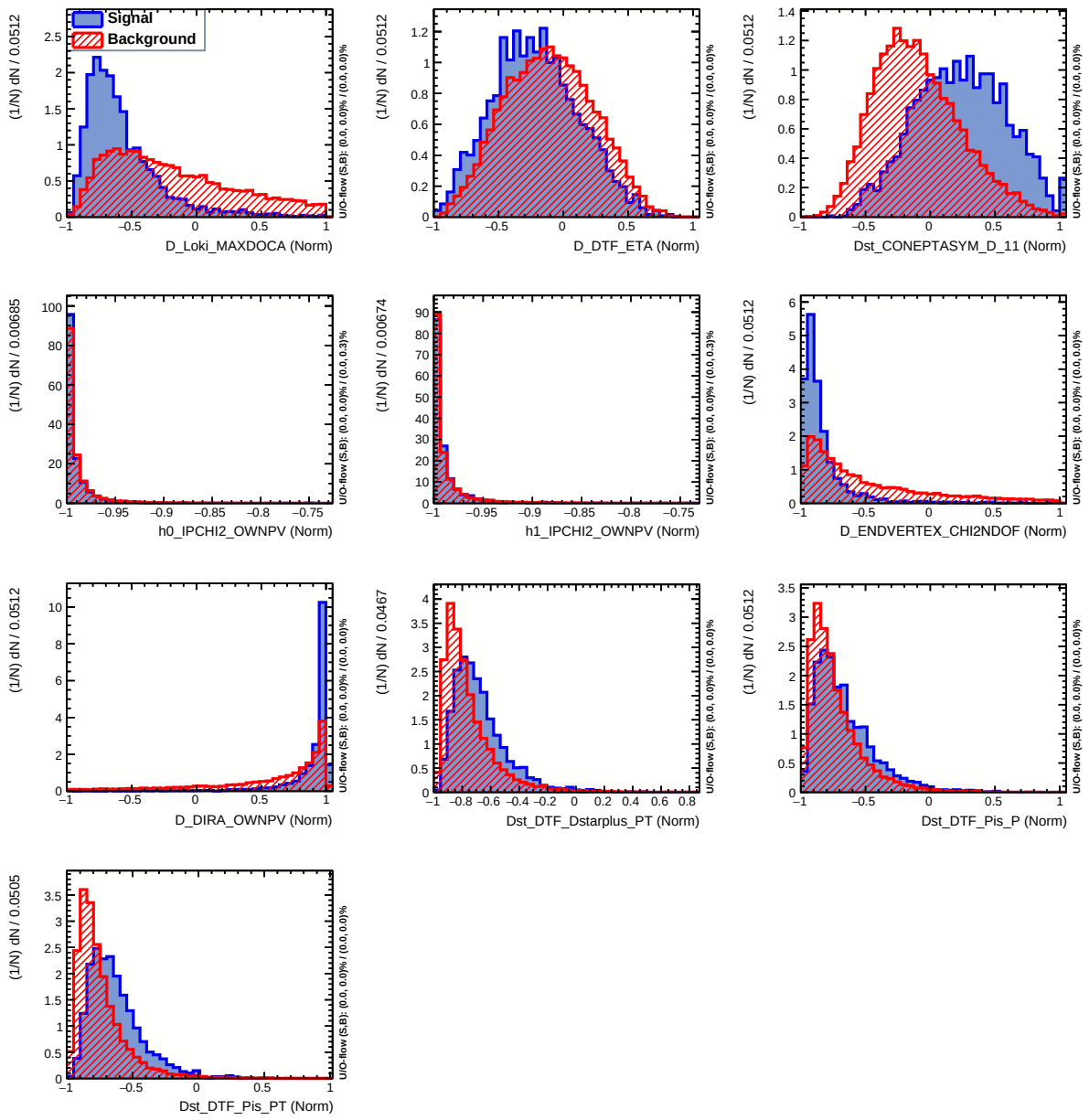


Figure B.1: Normalised BDT input variable distributions for signal, $D^0 \to K^-K^+\mu^\pm e^\mp$ decays, (blue) and background proxy (red) overlaid.

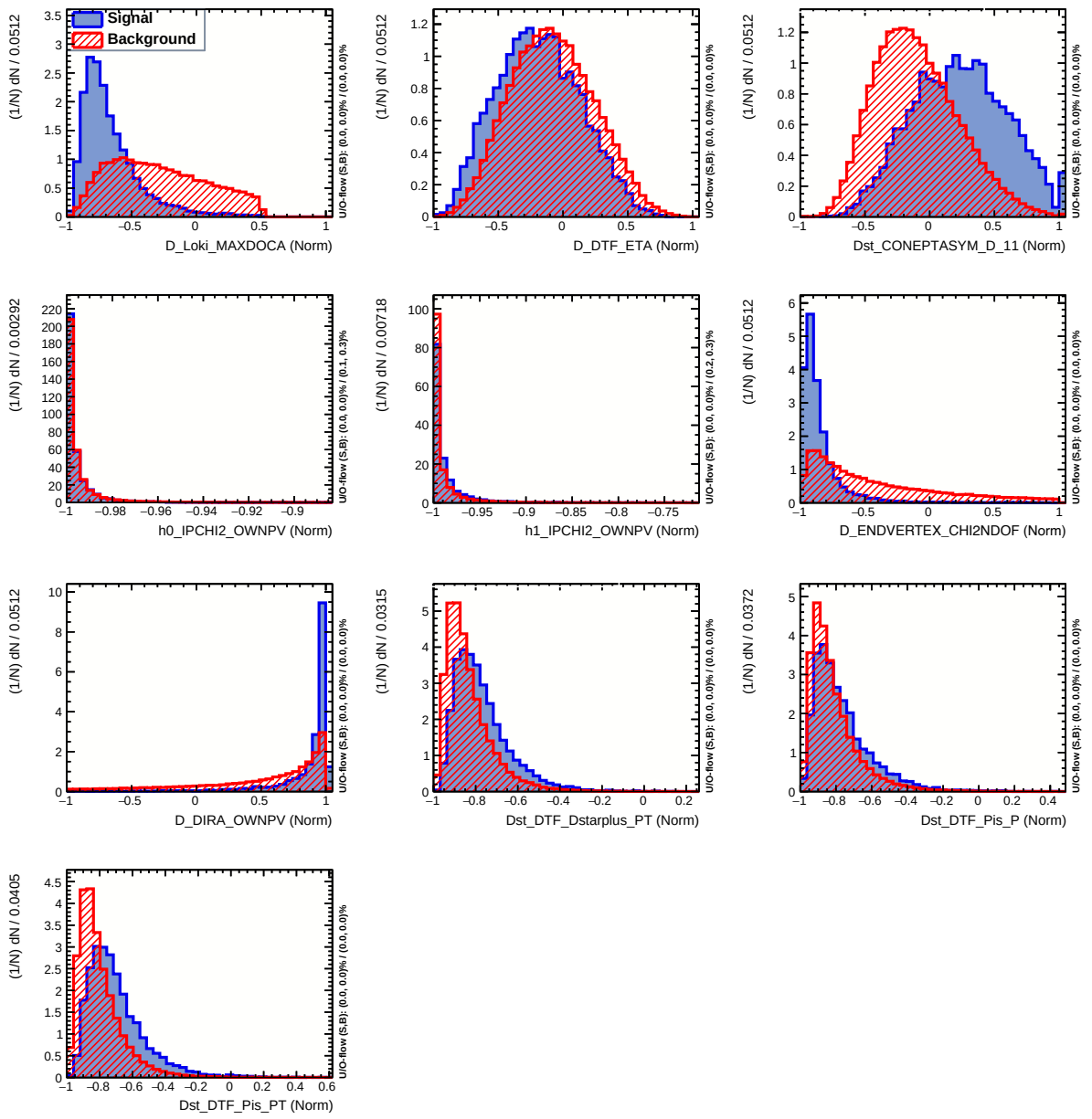


Figure B.2: Normalised BDT input variable distributions for signal, $D^0 \to K^-\pi^+\mu^\pm e^\mp$ decays, (blue) and background proxy (red) overlaid.

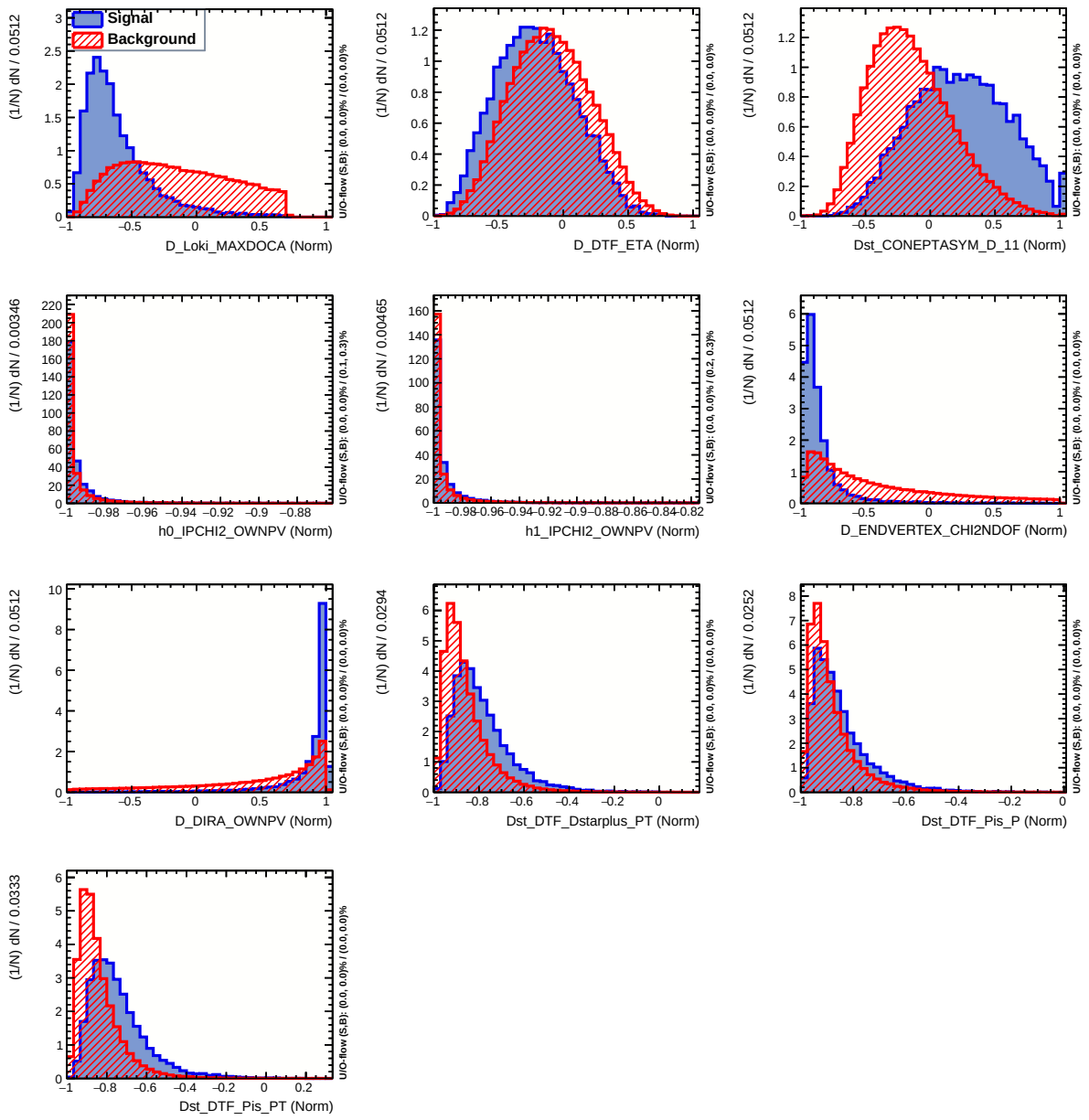


Figure B.3: Normalised BDT input variable distributions for signal, $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ decays, and background proxy overlaid.

Appendix C

Additional material on the optimisation

In the following all signal efficiencies and expected background yields separately for combinatorial and misID background in BDT response and PID bins are reported.

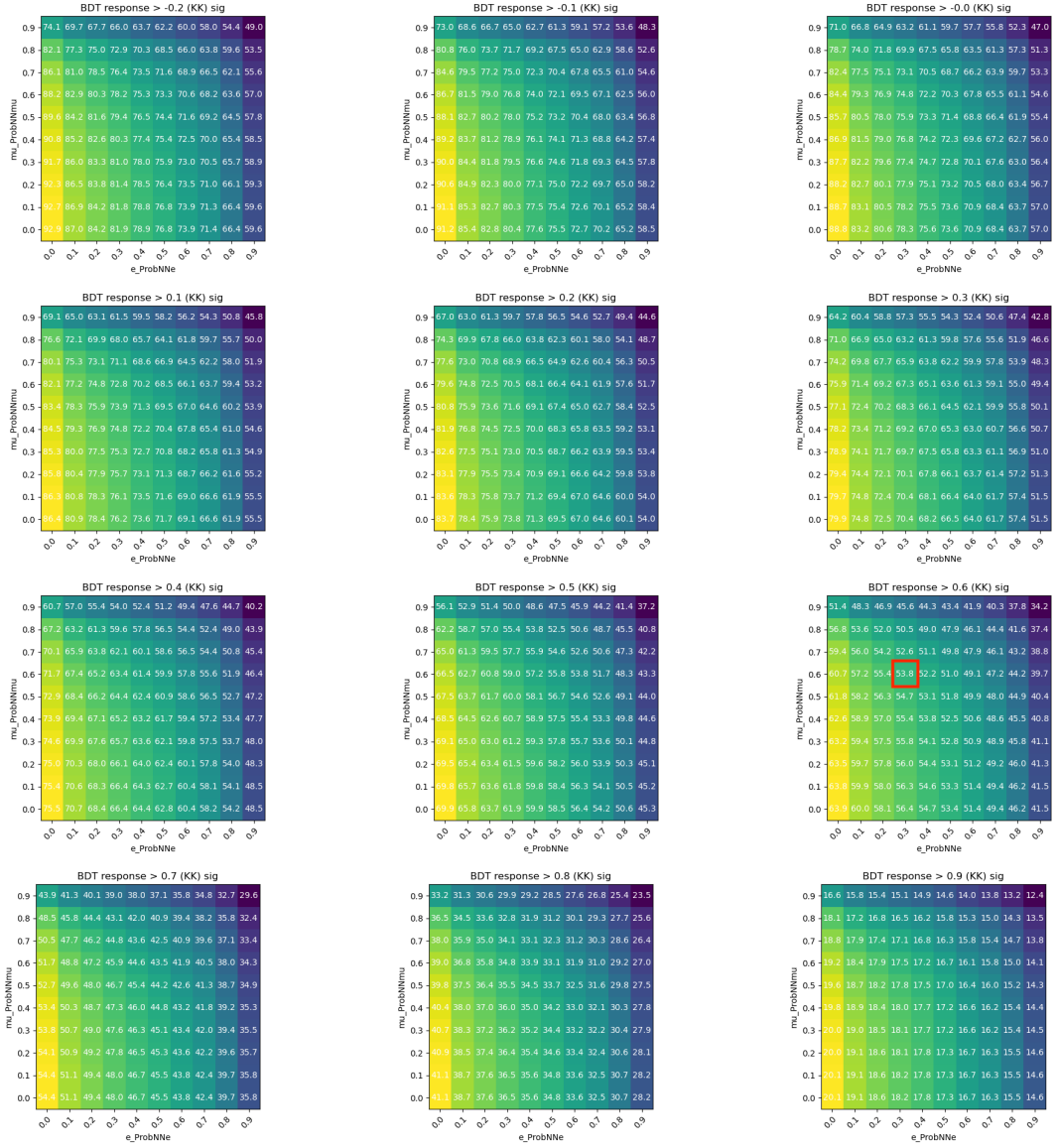


Figure C.1: $D^0 \to K^- K^+ \mu^\pm e^\mp$ decay signal efficiencies.

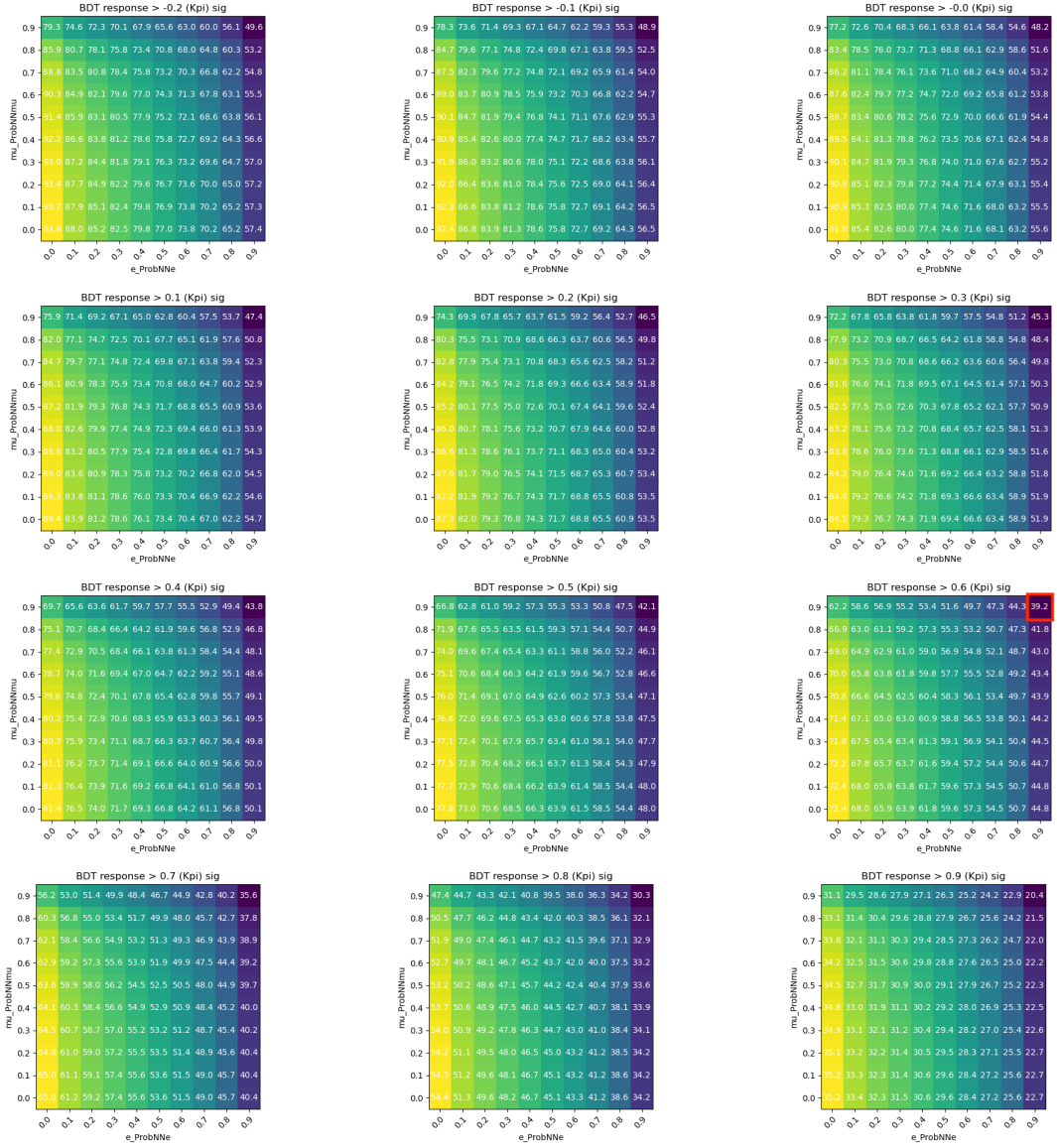


Figure C.2: $D^0 \to K^-\pi^+\mu^\pm e^\mp$ decay signal efficiencies.

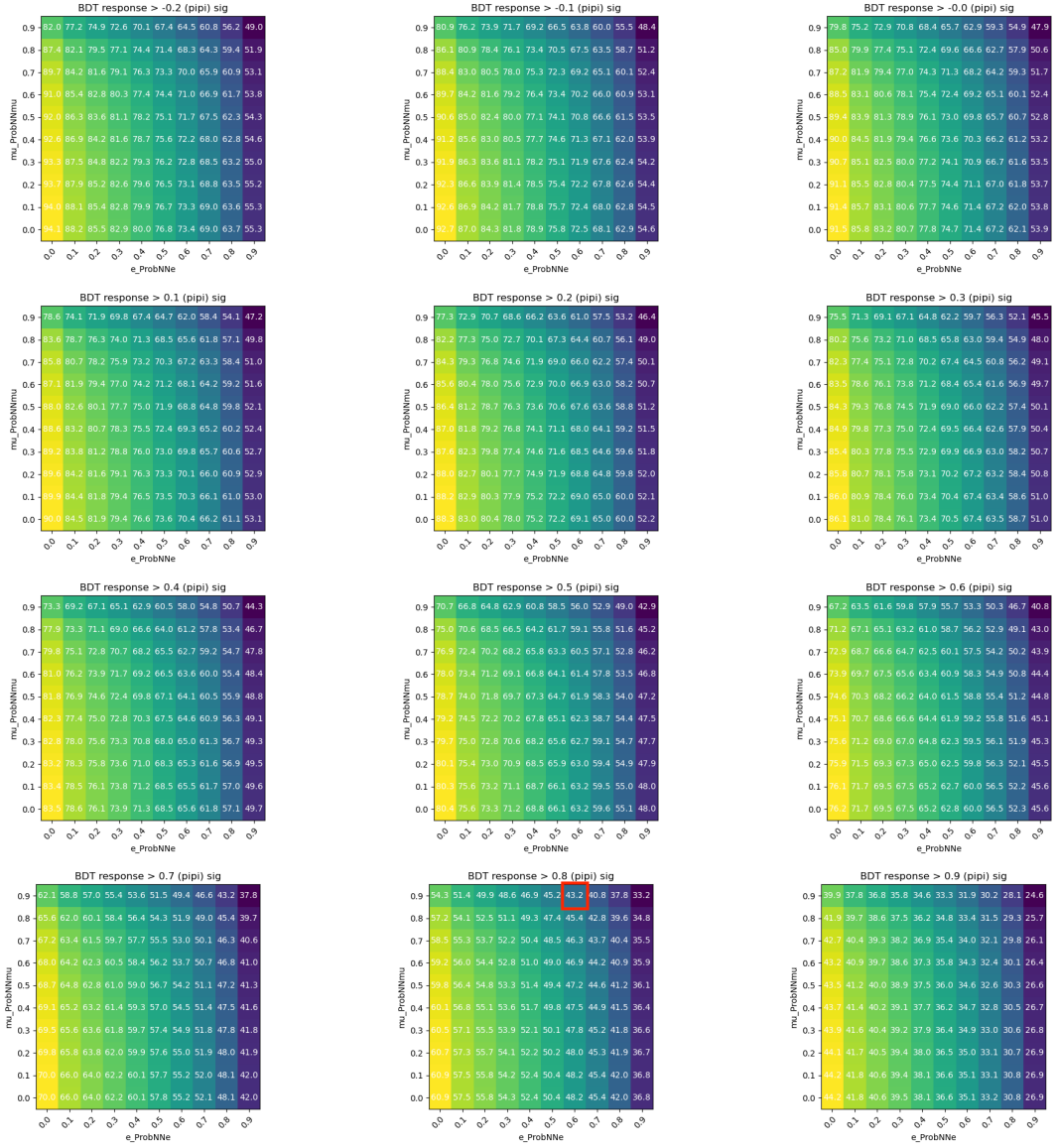


Figure C.3: $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ decay signal efficiencies.

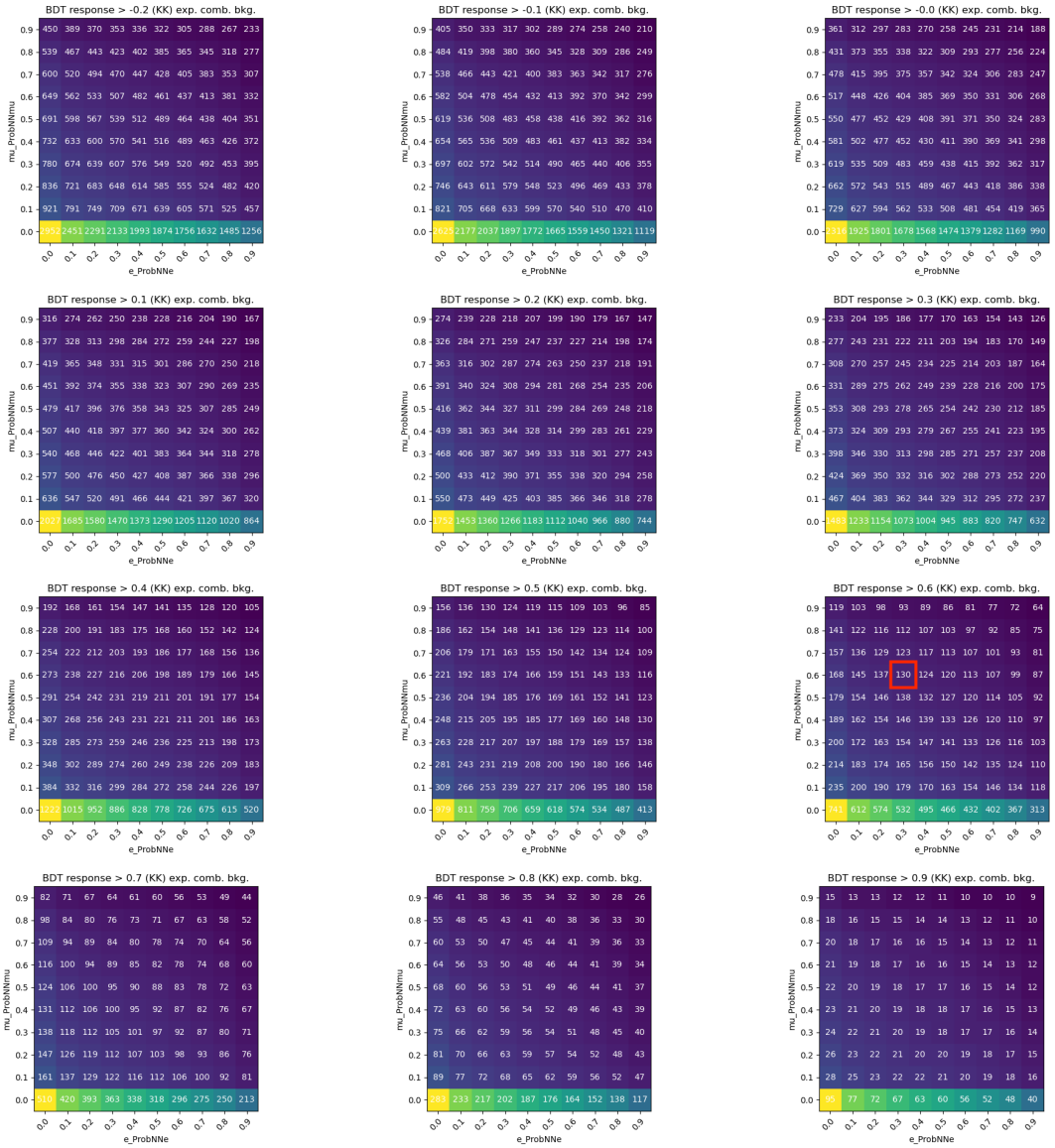


Figure C.4: Expected number of combinatorial background candidates for $D^0 \to K^-K^+\mu^\pm e^\mp$ decay selection.

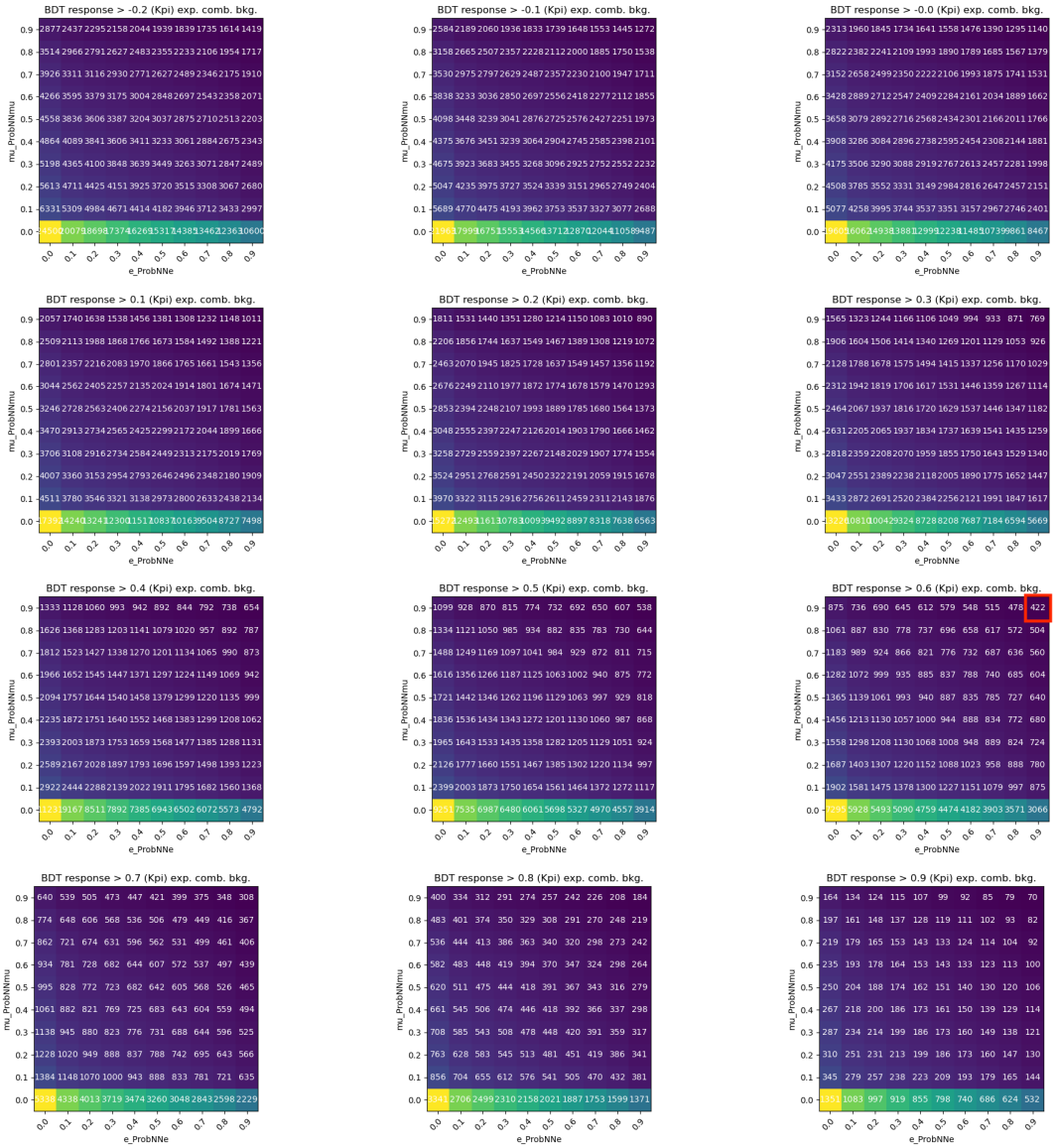


Figure C.5: Expected number of combinatorial background candidates for $D^0 \to K^-\pi^+\mu^\pm e^\mp$ decay selection.

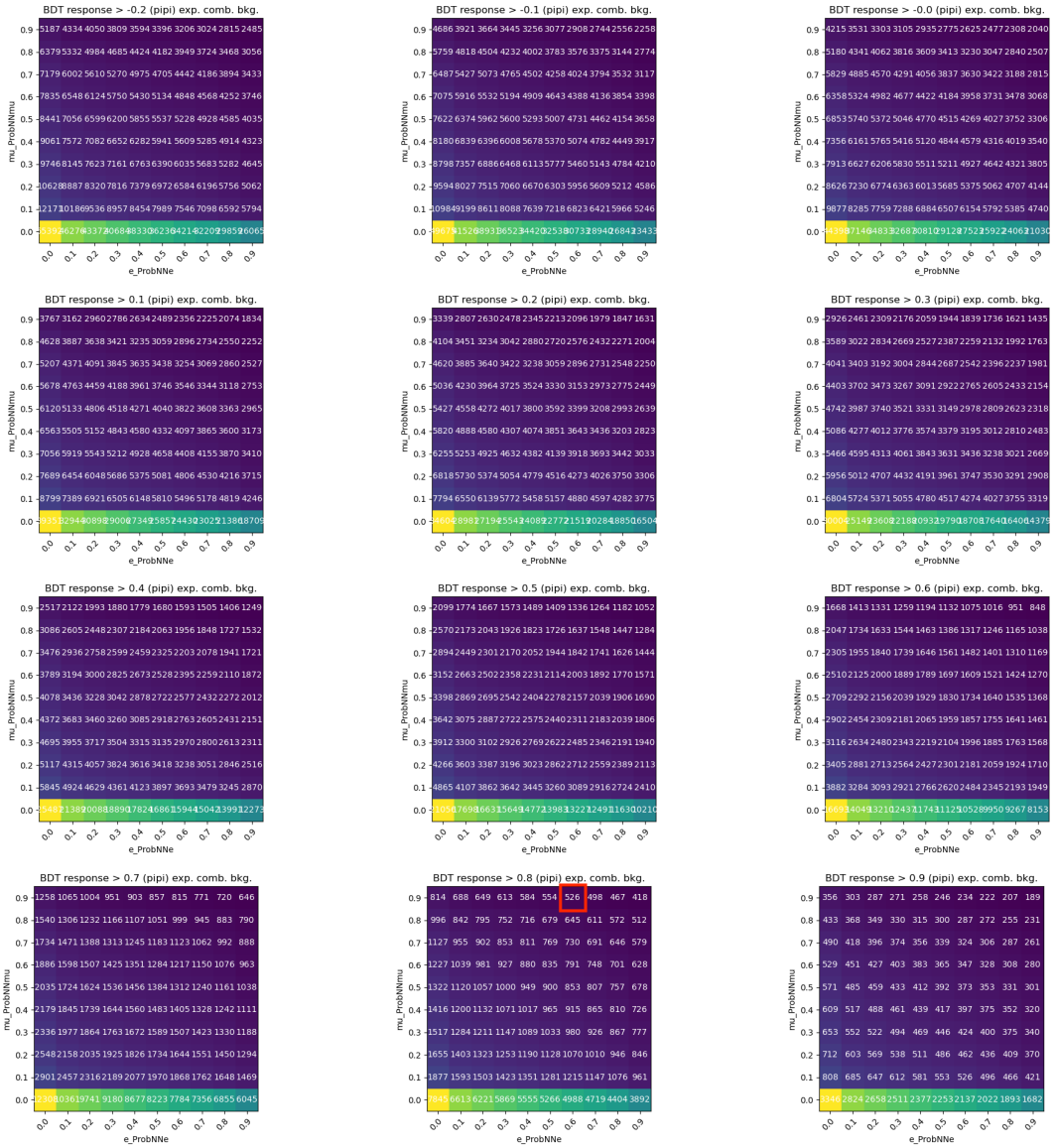


Figure C.6: Expected number of combinatorial background candidates for $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ decay selection.

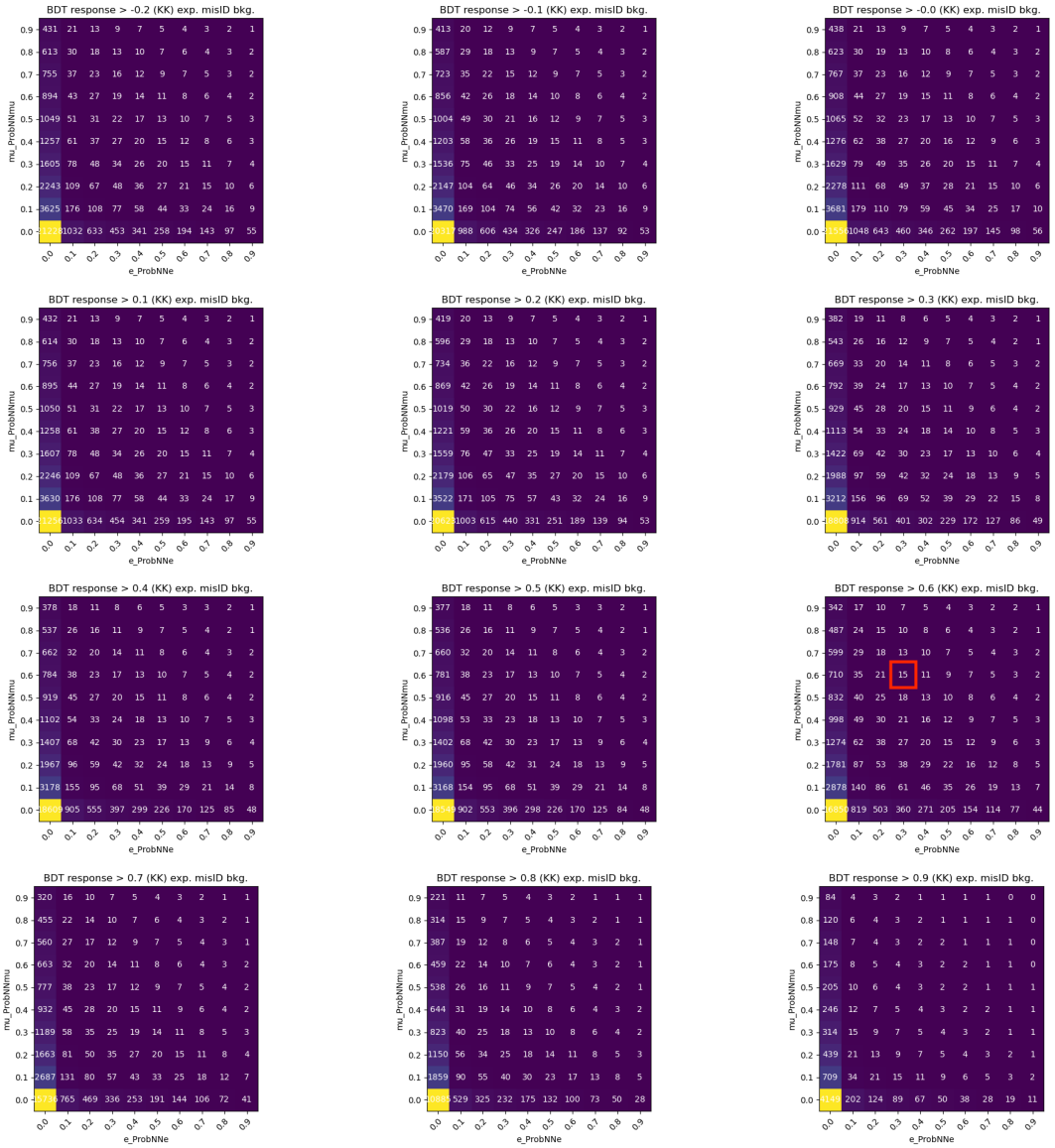


Figure C.7: Expected number of misID background candidates for $D^0 \to K^-K^+\mu^\pm e^\mp$ decay selection.

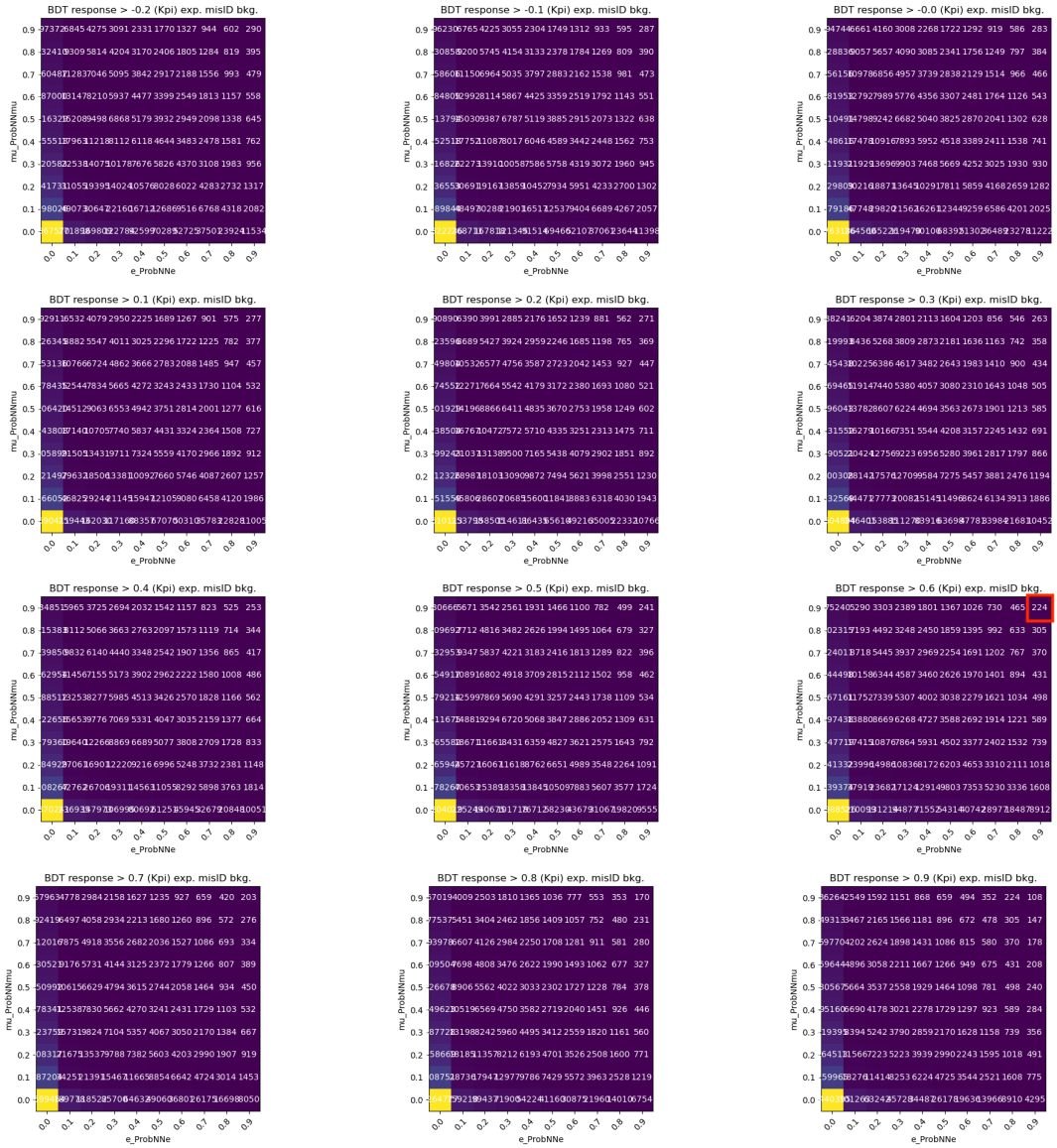


Figure C.8: Expected number of misID background candidates for $D^0 \to K^-\pi^+\mu^\pm e^\mp$ decay selection.

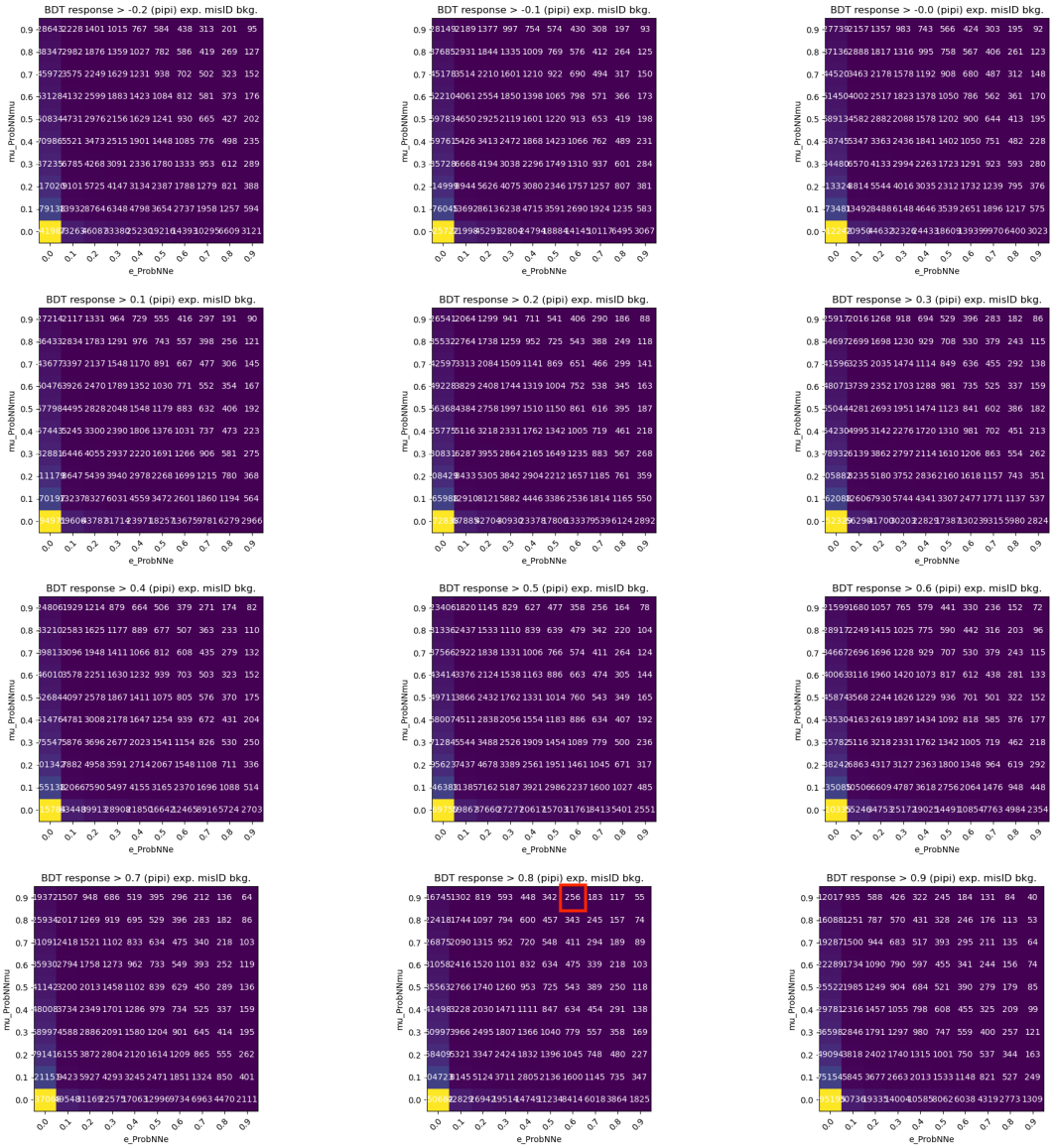


Figure C.9: Expected number of misID background candidates for $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ decay selection.

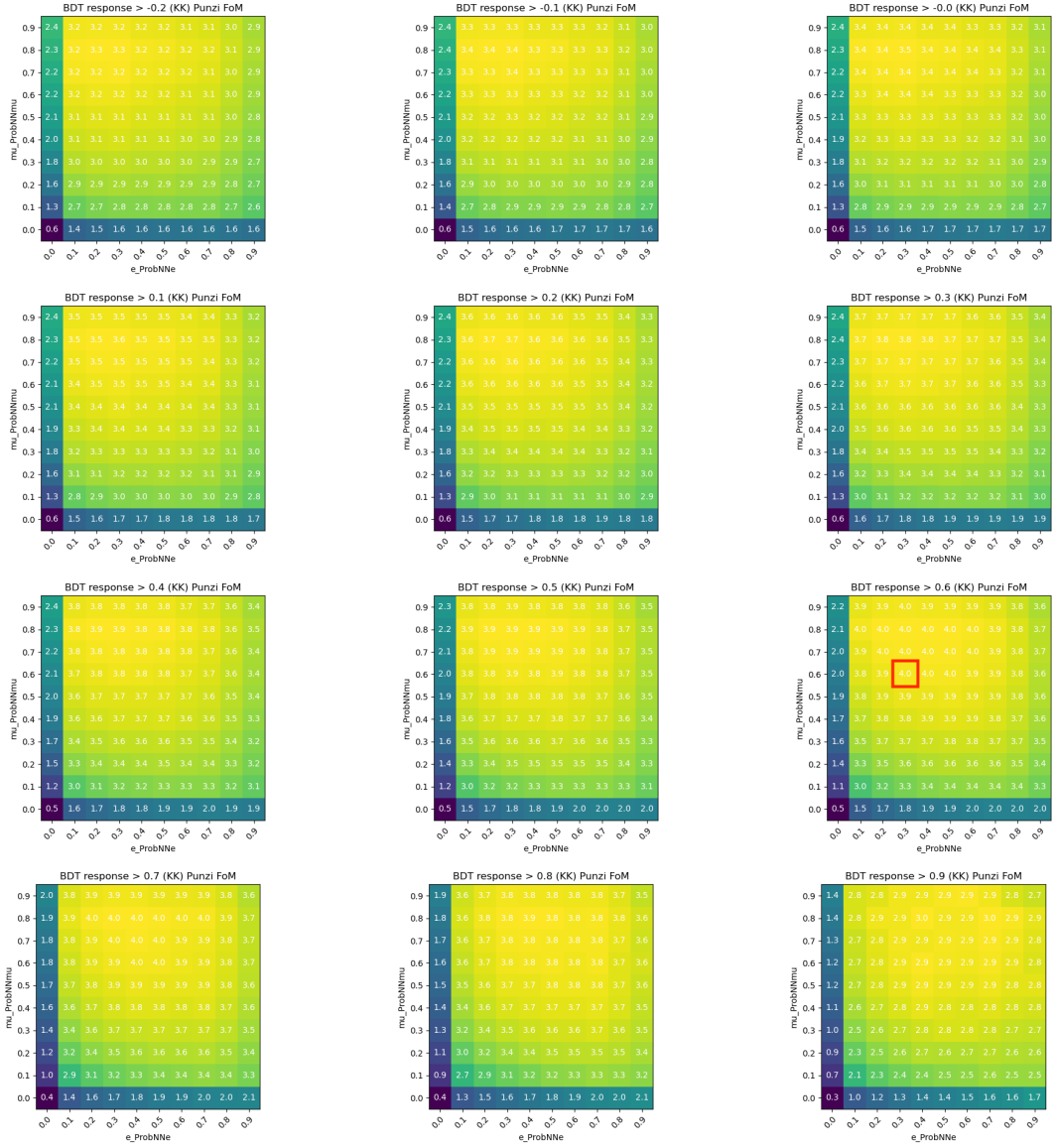


Figure C.10: Punzi figure of merit for $D^0 \to K^-K^+\mu^\pm e^\mp$ decays.

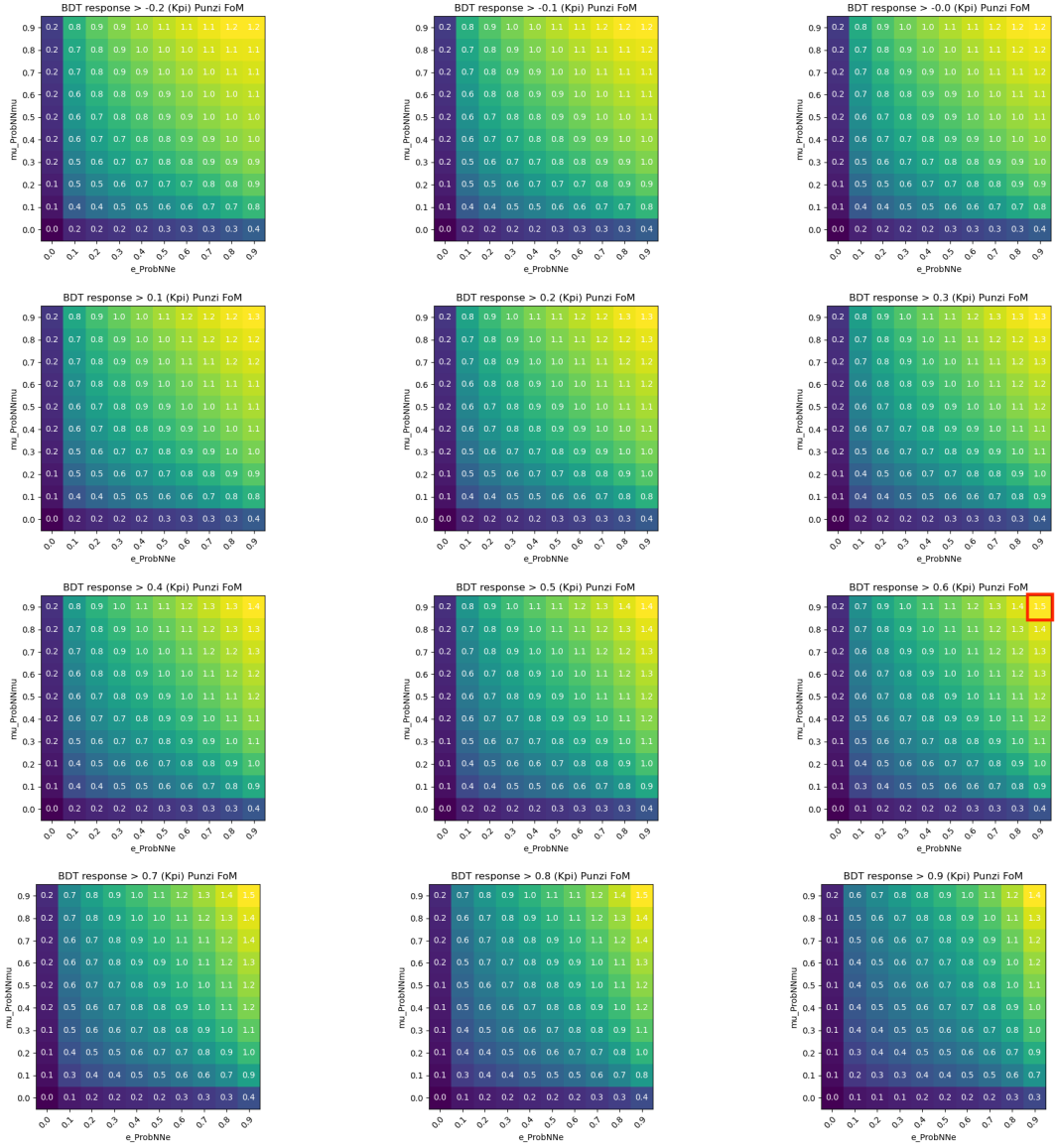


Figure C.11: Punzi figure of merit for $D^0 \to K^-\pi^+\mu^\pm e^\mp$ decays.

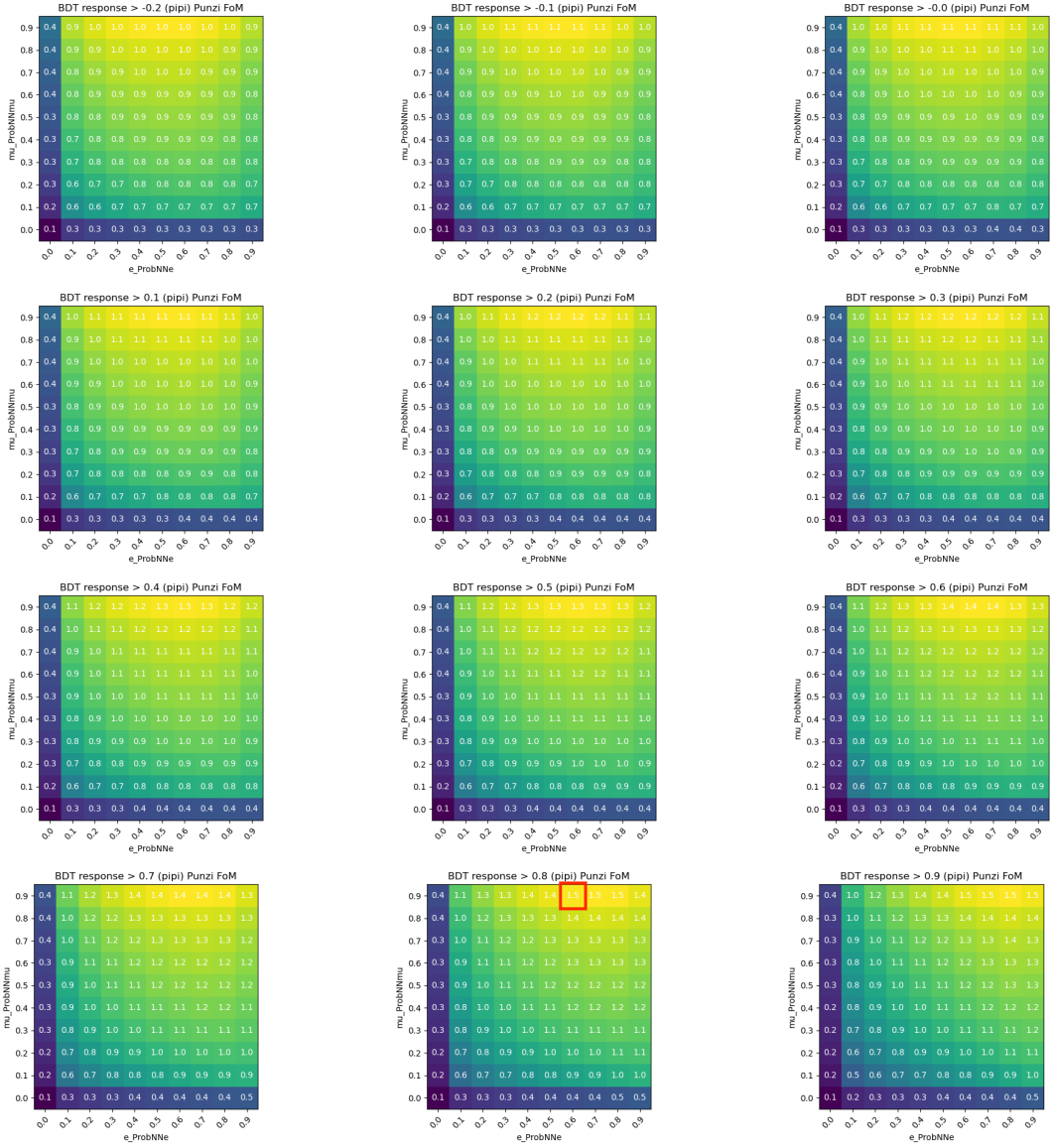


Figure C.12: Punzi figure of merit for $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ decays.

Appendix D

Efficiency Tables

Summary of the efficiencies for the selection explained in chapter 5.

D.1 Reference decays

All efficiency ratios for the reference modes are reported in table D.1.

	$\frac{\epsilon(D^0\to K^-K^+\mu^-\mu^+)}{\epsilon(D^0\to K^-\pi^+\mu^-\mu^+)}$	$\frac{\epsilon(D^0\to \pi^-\pi^+\mu^-\mu^+)}{\epsilon(D^0\to K^-\pi^+\mu^-\mu^+)}$
r_{acc}	1.045 ± 0.017	0.925 ± 0.015
r_{rec}	0.747 ± 0.01	1.146 ± 0.012
r_{sele}	0.948 ± 0.005	1.025 ± 0.004
r_{PID}	0.951 ± 0.005	1.072 ± 0.007
r_{BDT}	1.085 ± 0.01	0.927 ± 0.008
r_{Trig}	0.629 ± 0.037	1.060 ± 0.031
r	0.481 ± 0.031	1.145 ± 0.042

(a) 2015/2016

	$\frac{\epsilon(D^0\to K^-K^+\mu^-\mu^+)}{\epsilon(D^0\to K^-\pi^+\mu^-\mu^+)}$	$\frac{\epsilon(D^0\to \pi^-\pi^+\mu^-\mu^+)}{\epsilon(D^0\to K^-\pi^+\mu^-\mu^+)}$
r_{acc}	1.045 ± 0.017	0.925 ± 0.015
r_{rec}	0.751 ± 0.004	1.154 ± 0.006
r_{sele}	0.937 ± 0.003	1.031 ± 0.002
r_{PID}	0.946 ± 0.003	1.100 ± 0.005
r_{BDT}	1.079 ± 0.007	0.940 ± 0.006
r_{Trig}	0.741 ± 0.007	1.106 ± 0.007
r	0.556 ± 0.012	1.259 ± 0.025

(b) 2017/2018

Table D.1: Summary table for all efficiency ratios acceptance, reconstruction, PID selection, BDT selection and trigger selection for the different reference modes.

D.2 Signal decays

Efficiencies for each selection step and fit category are reported in table D.2, table D.3 and table D.4 for the different signal modes.

	$\texttt{noBrem}_{15/16}$	$\texttt{noBrem}_{17/18}$	$\texttt{Brem}_{15/16}$	$\texttt{Brem}_{17/18}$	$\texttt{norm}_{15/16}$	$\texttt{norm}_{17/18}$
$\epsilon_{acc}[\%]$	19.927 ± 0.044	19.927 ± 0.044	19.927 ± 0.044	19.927 ± 0.044	19.79 ± 0.3	19.79 ± 0.3
$\epsilon_{rec}[\%]$	41.641 ± 0.567	55.951 ± 0.479	19.046 ± 0.389	25.049 ± 0.32	113.577 ± 0.726	168.302 ± 0.706
$\epsilon_{PID}[\%]$	62.268 ± 0.632	59.657 ± 0.425	64.094 ± 0.951	60.613 ± 0.641	71.378 ± 0.185	72.649 ± 0.164
$\epsilon_{BDT}[\%]$	50.737 ± 0.826	50.766 ± 0.56	55.958 ± 1.229	56.652 ± 0.835	48.252 ± 0.242	45.49 ± 0.215
$\epsilon_{trig}[\%]$	5.295 ± 0.467	15.598 ± 0.342	2.098 ± 0.322	12.429 ± 0.409	11.039 ± 0.392	28.902 ± 0.146
$\epsilon[\%]$	0.139 ± 0.013	0.527 ± 0.014	0.029 ± 0.004	0.213 ± 0.008	0.855 ± 0.034	3.181 ± 0.055

Table D.2: Efficiencies for $D^0 \to K^- K^+ \mu^\pm e^\mp$ split by the different fit categories used in the simultaneous fit.

	$\texttt{noBrem}_{15/16}$	$\texttt{noBrem}_{17/18}$	$\texttt{Brem}_{15/16}$	$\texttt{Brem}_{17/18}$	$\texttt{norm}_{15/16}$	$\texttt{norm}_{17/18}$
$\epsilon_{acc}[\%]$	18.872 ± 0.041	18.872 ± 0.041	18.872 ± 0.041	18.872 ± 0.041	19.79 ± 0.3	19.79 ± 0.3
$\epsilon_{rec}[\%]$	46.743 ± 0.427	62.863 ± 0.355	26.48 ± 0.323	35.128 ± 0.267	113.577 ± 0.726	168.302 ± 0.706
$\epsilon_{PID}[\%]$	35.649 ± 0.408	33.25 ± 0.264	38.926 ± 0.563	35.278 ± 0.364	35.215 ± 0.196	37.06 ± 0.178
$\epsilon_{BDT}[\%]$	62.396 ± 0.691	61.837 ± 0.471	68.016 ± 0.864	68.382 ± 0.596	67.223 ± 0.324	64.813 ± 0.289
$\epsilon_{trig}[\%]$	10.163 ± 0.46	27.277 ± 0.355	5.032 ± 0.482	24.943 ± 0.427	14.994 ± 0.397	36.175 ± 0.187
$\epsilon[\%]$	0.199 ± 0.01	0.665 ± 0.012	0.067 ± 0.007	0.399 ± 0.009	0.798 ± 0.026	2.894 ± 0.052

Table D.3: Efficiencies for $D^0 \to K^- \pi^+ \mu^\pm e^\mp$ split by the different fit categories used in the simultaneous fit.

	$\texttt{noBrem}_{15/16}$	$\texttt{noBrem}_{17/18}$	$\texttt{Brem}_{15/16}$	$\texttt{Brem}_{17/18}$	$\texttt{norm}_{15/16}$	$\texttt{norm}_{17/18}$
$\epsilon_{acc}[\%]$	17.996 ± 0.04	17.996 ± 0.04	17.996 ± 0.04	17.996 ± 0.04	19.79 ± 0.3	19.79 ± 0.3
$\epsilon_{rec}[\%]$	55.715 ± 0.467	78.136 ± 0.404	34.493 ± 0.376	47.515 ± 0.313	113.577 ± 0.726	168.302 ± 0.706
$\epsilon_{PID}[\%]$	46.412 ± 0.402	45.286 ± 0.264	48.887 ± 0.522	47.485 ± 0.341	47.778 ± 0.205	49.687 ± 0.184
$\epsilon_{BDT}[\%]$	47.551 ± 0.591	49.049 ± 0.395	54.842 ± 0.743	56.187 ± 0.492	54.464 ± 0.295	51.073 ± 0.261
$\epsilon_{trig}[\%]$	13.595 ± 0.507	33.44 ± 0.342	3.943 ± 0.359	31.829 ± 0.391	15.407 ± 0.466	36.941 ± 0.148
$\epsilon[\%]$	0.301 ± 0.012	1.044 ± 0.016	0.066 ± 0.006	0.726 ± 0.013	0.901 ± 0.032	3.122 ± 0.054

Table D.4: Efficiencies for $D^0 \to \pi^- \pi^+ \mu^\pm e^\mp$ split by the different fit categories used in the simultaneous fit.

Appendix E

Tracking Maps

In figure E.1 the data simulation correction maps for the muon reconstruction efficiency ratio, for details see section 6.2, are reported. The electron maps are reported in figure E.2 (2015), figure E.3 (2016), figure E.4 (2017) and figure E.5 (2018).

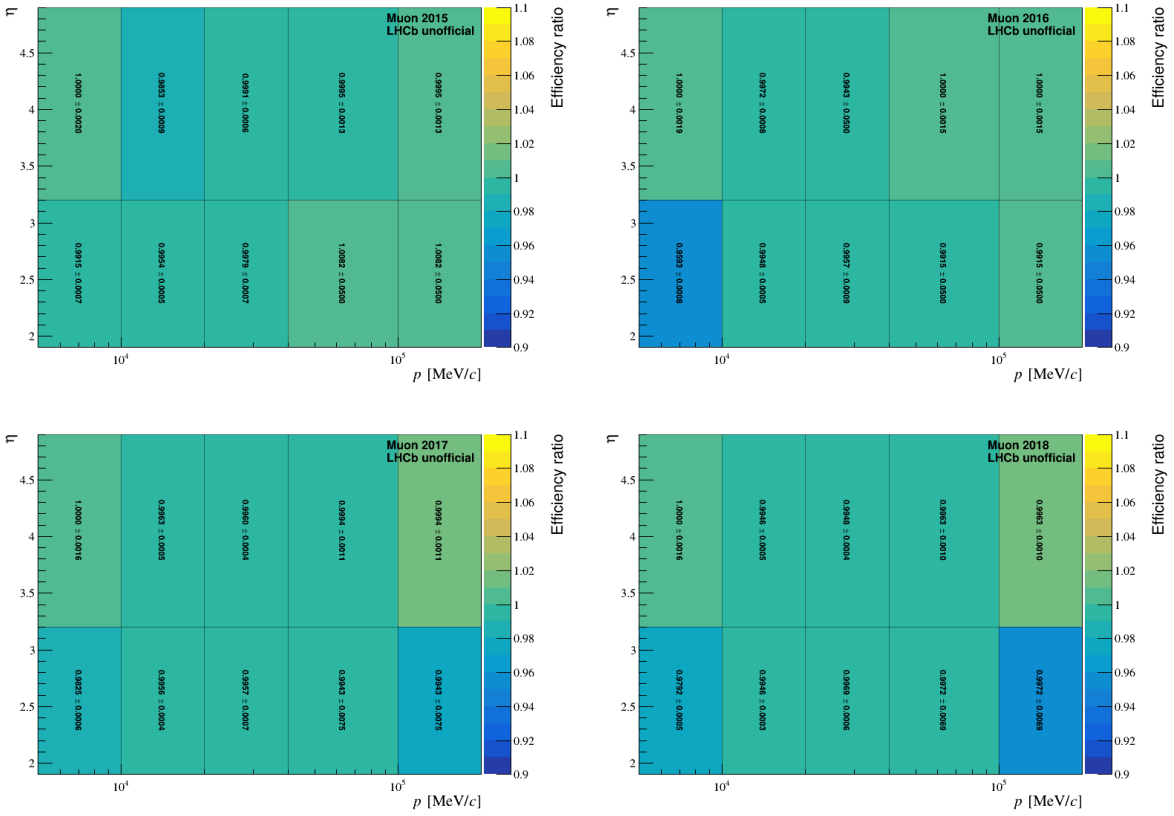


Figure E.1: Muon tracking data-simulation correction map in kinematic variables of the muon. The values for the muon correction map are provided centrally to the LHCb collaboration.

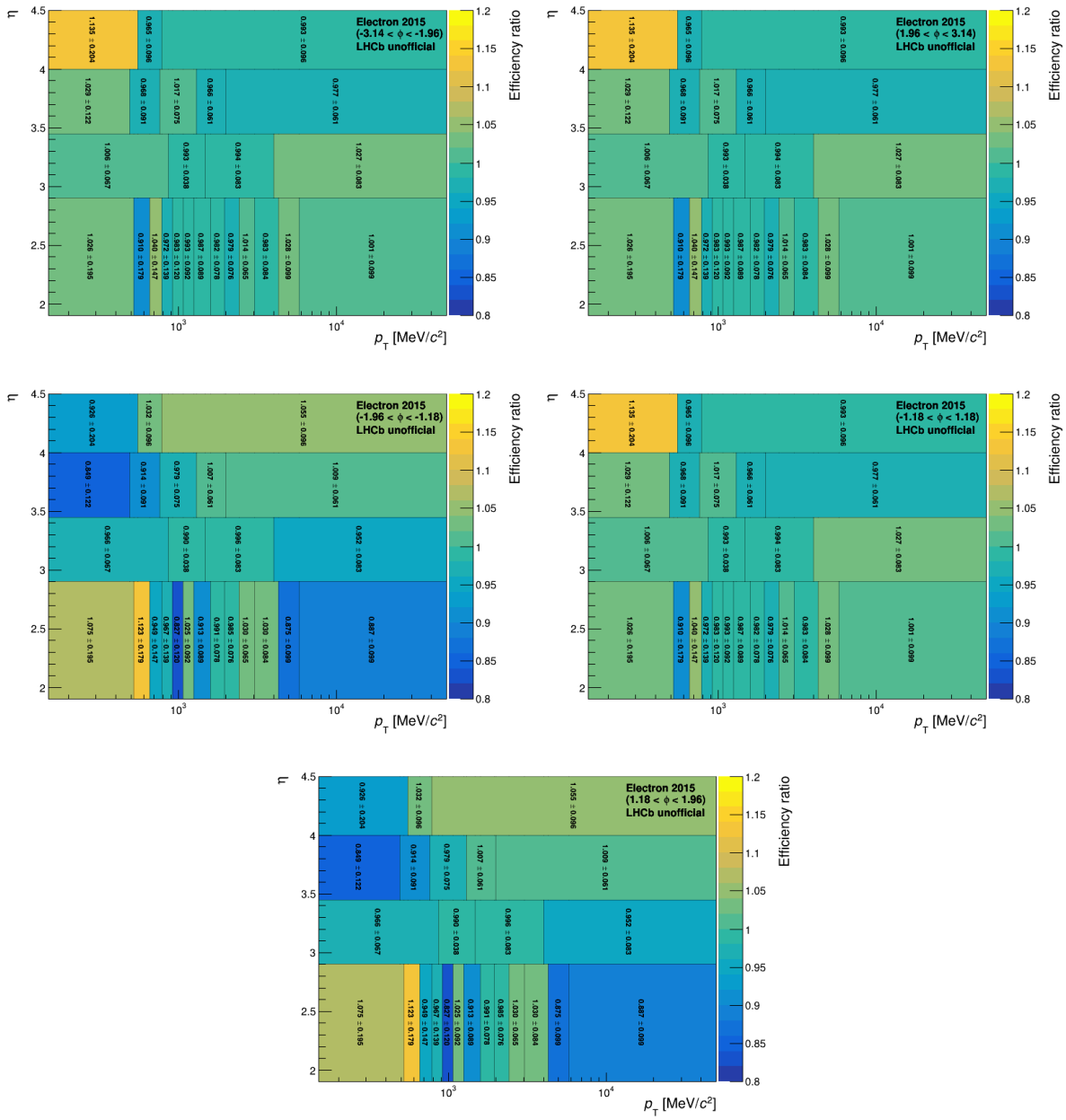


Figure E.2: Electron tracking data-simulation correction map in kinematic variables of the electron for 2015. The values for the electron correction map are taken from the R_X analysis [13, 14].

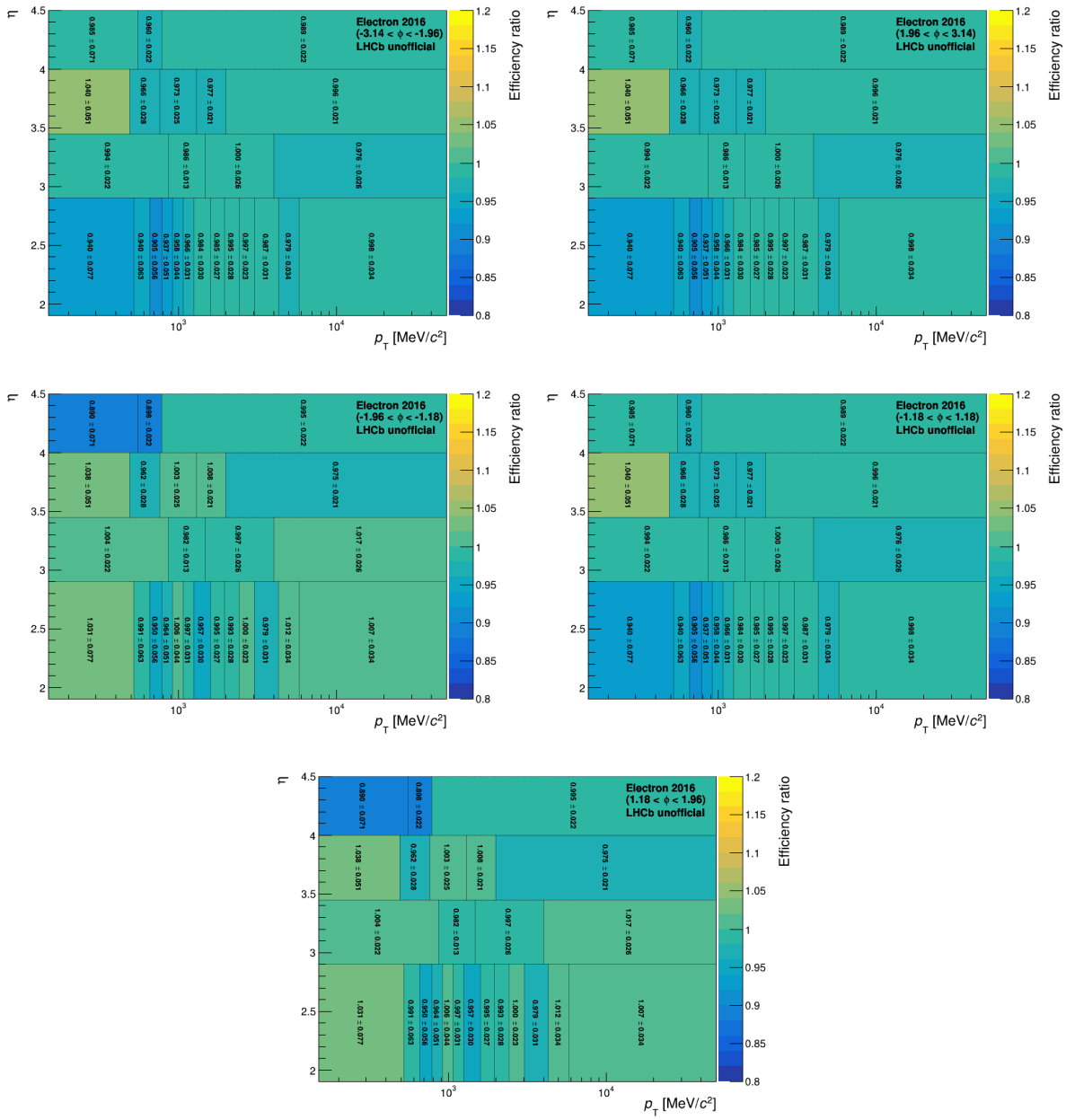


Figure E.3: Electron tracking data-simulation correction map in kinematic variables of the electron for 2016. The values for the electron correction map are taken from the R_X analysis [13, 14].

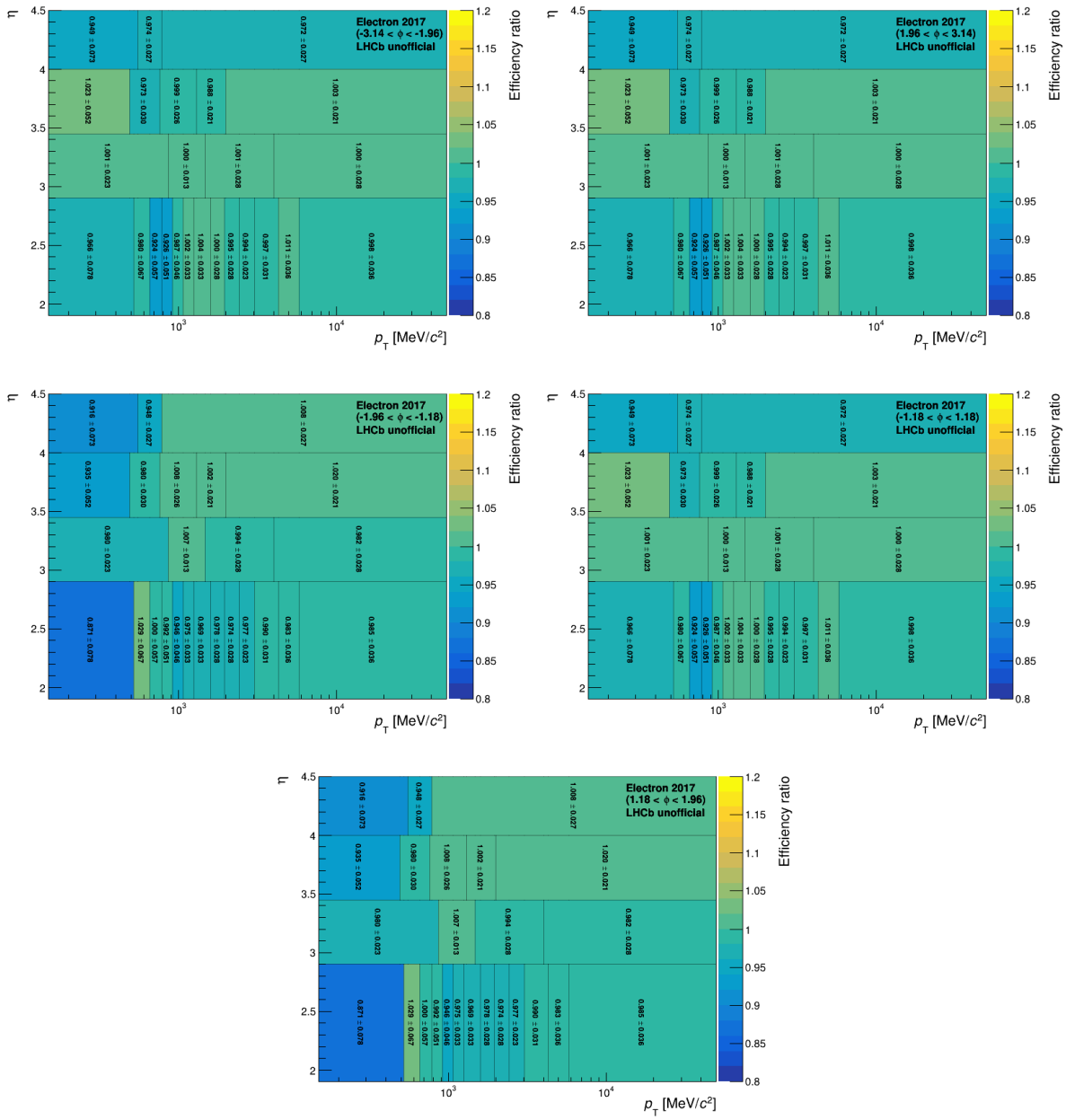


Figure E.4: Electron tracking data-simulation correction map in kinematic variables of the electron for 2017. The values for the electron correction map are taken from the R_X analysis [13, 14].

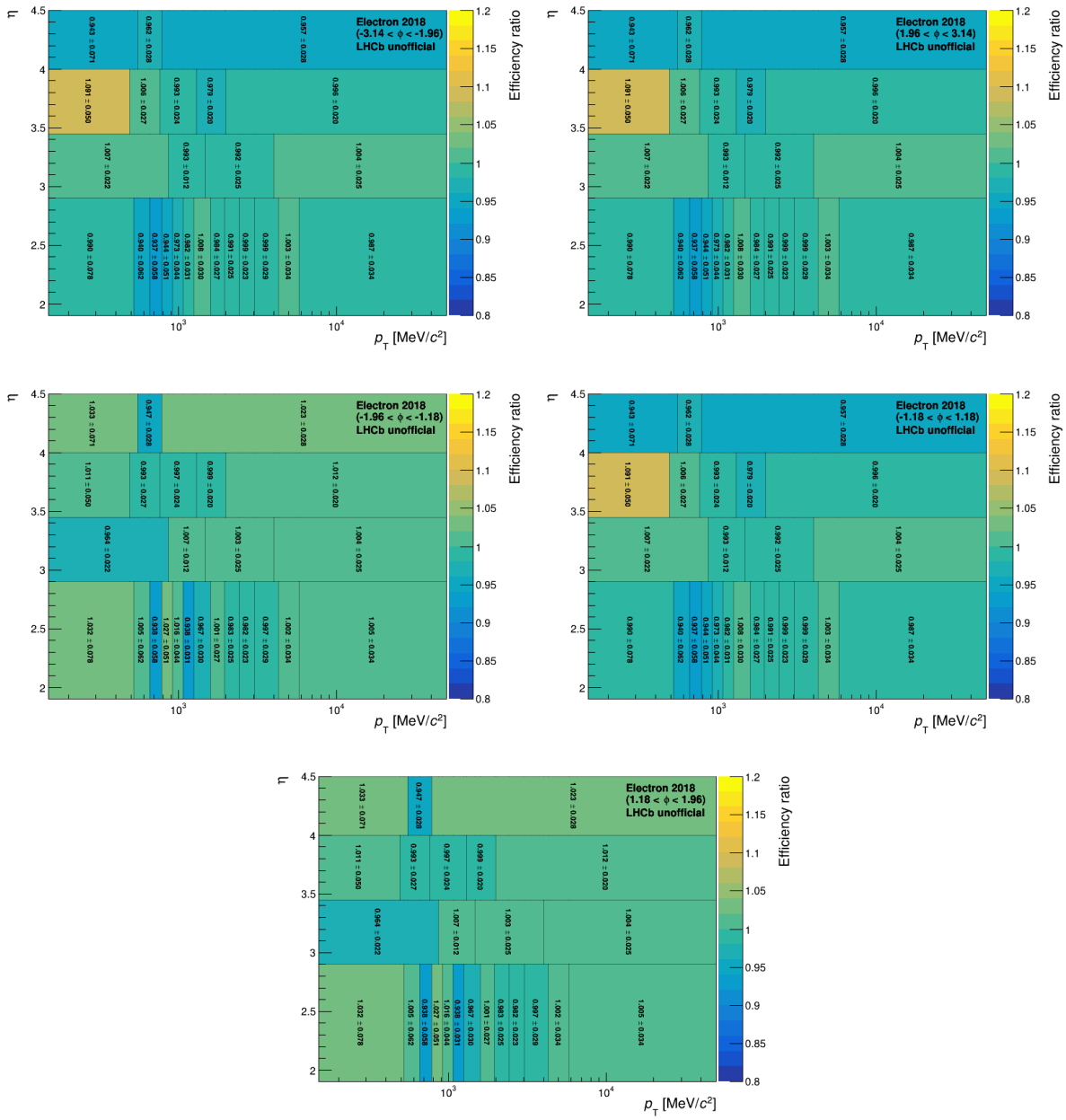


Figure E.5: Electron tracking data-simulation correction map in kinematic variables of the electron for 2018. The values for the electron correction map are taken from the R_X analysis [13, 14].

Appendix F

Fit results

In the following, all fits and their results to the simulation samples to fix the shape of the signal and misID component in the simultaneous fit are summarised.

F.1 Signal shape fit

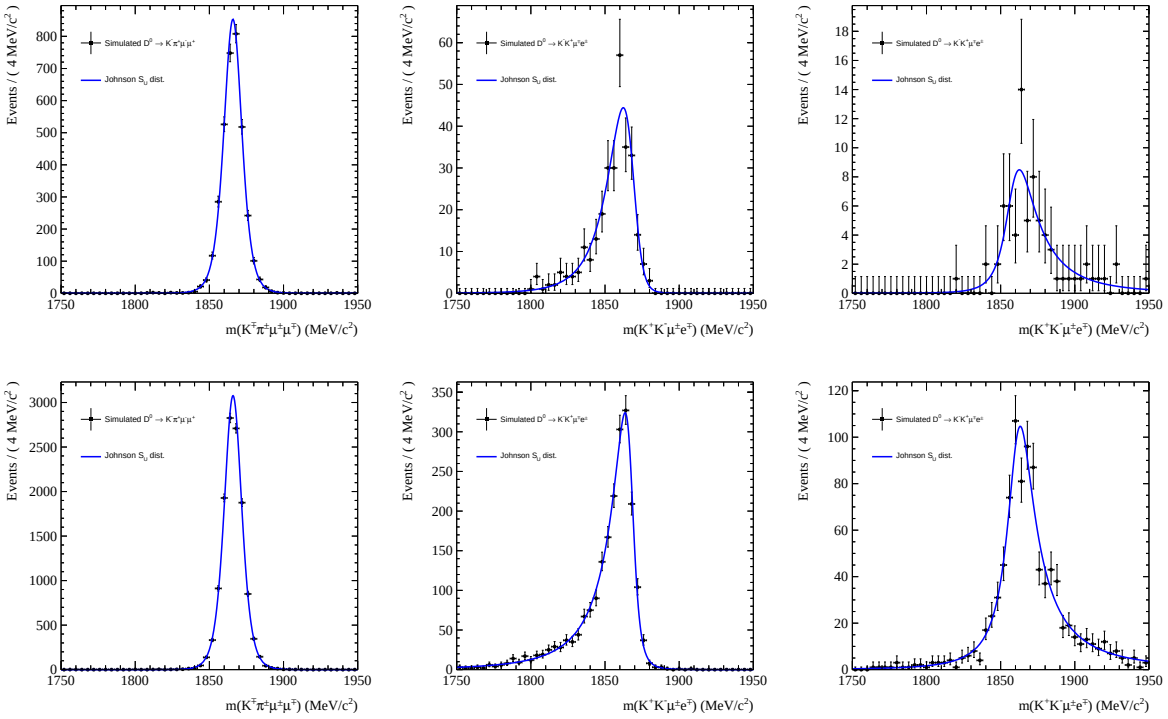


Figure F.1: MC fits to extract the shapes for the simultaneous fit for $D^0 \to K^-K^+\mu^\pm e^\mp$. Left is the Fit of the normalisation channel (with $D^0 \to K^-K^+\mu^\pm e^\mp$ BDT selection). In the middle (right), the fit to the reconstructed invariant D^0 mass in category `noBrem` (`Brem`). The upper row shows the fits for 2015/2016, and the lower row for 2017/2018.

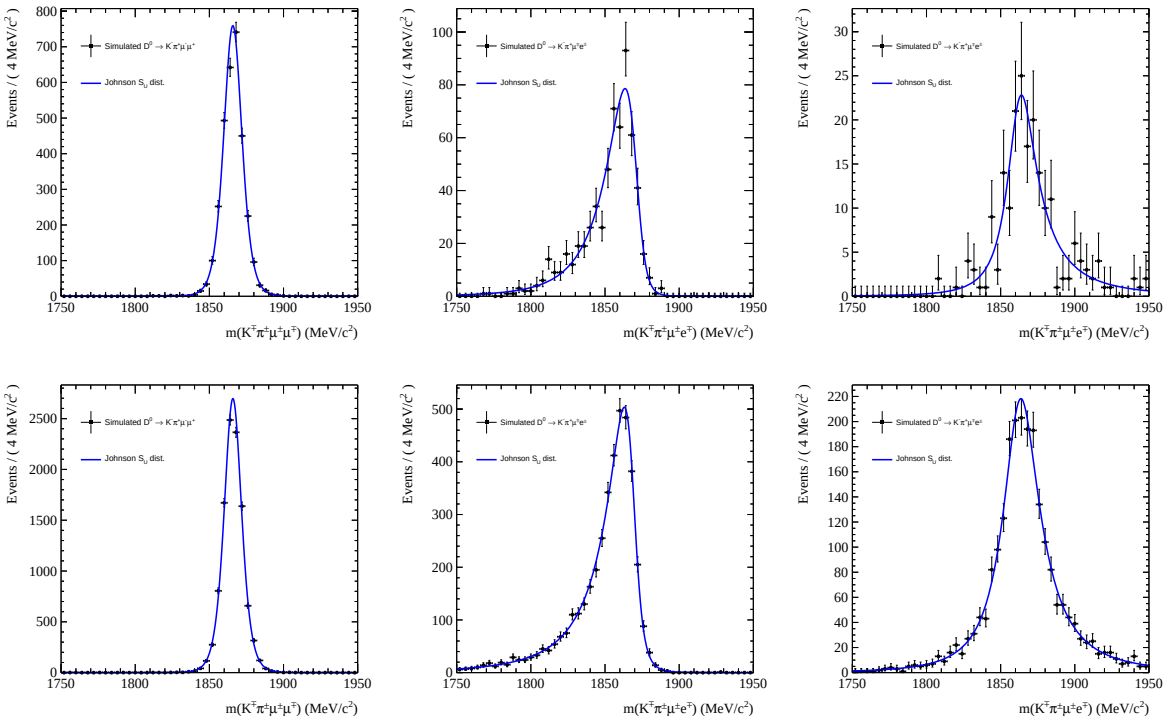


Figure F.2: MC fits to extract the shapes for the simultaneous fit for $D^0 \to K^-\pi^+\mu^\pm e^\mp$. Left is the fit of the normalisation channel (with $D^0 \to K^-\pi^+\mu^\pm e^\mp$ BDT selection). In the middle (right), the fit to the reconstructed invariant D^0 mass in category `noBrem` (`Brem`). The upper row shows the fits for 2015/2016, and the lower row for 2017/2018.

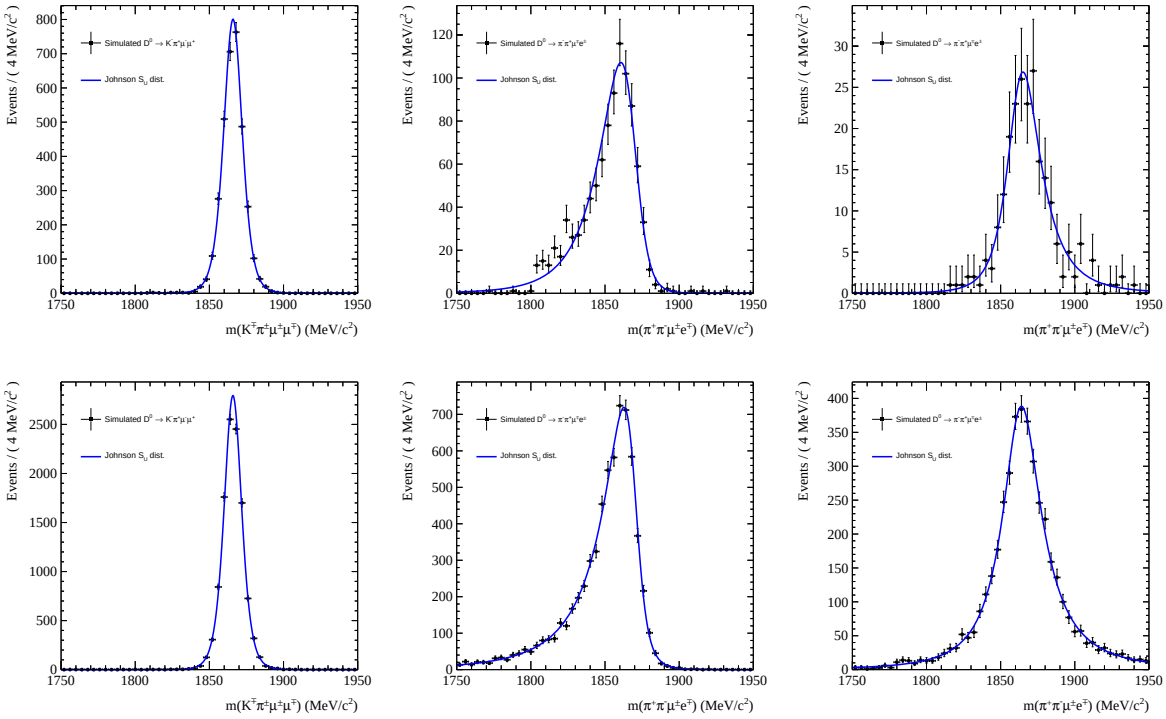


Figure F.3: MC fits to extract the shapes for the simultaneous fit for $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$. Left is the fit of the normalisation channel (with $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ BDT selection). In the middle (right), the fit to the reconstructed invariant D^0 mass in category `noBrem` (`Brem`). The upper row shows the fits for 2015/2016, and the lower row for 2017/2018.

F.2 Peaking background shape fit

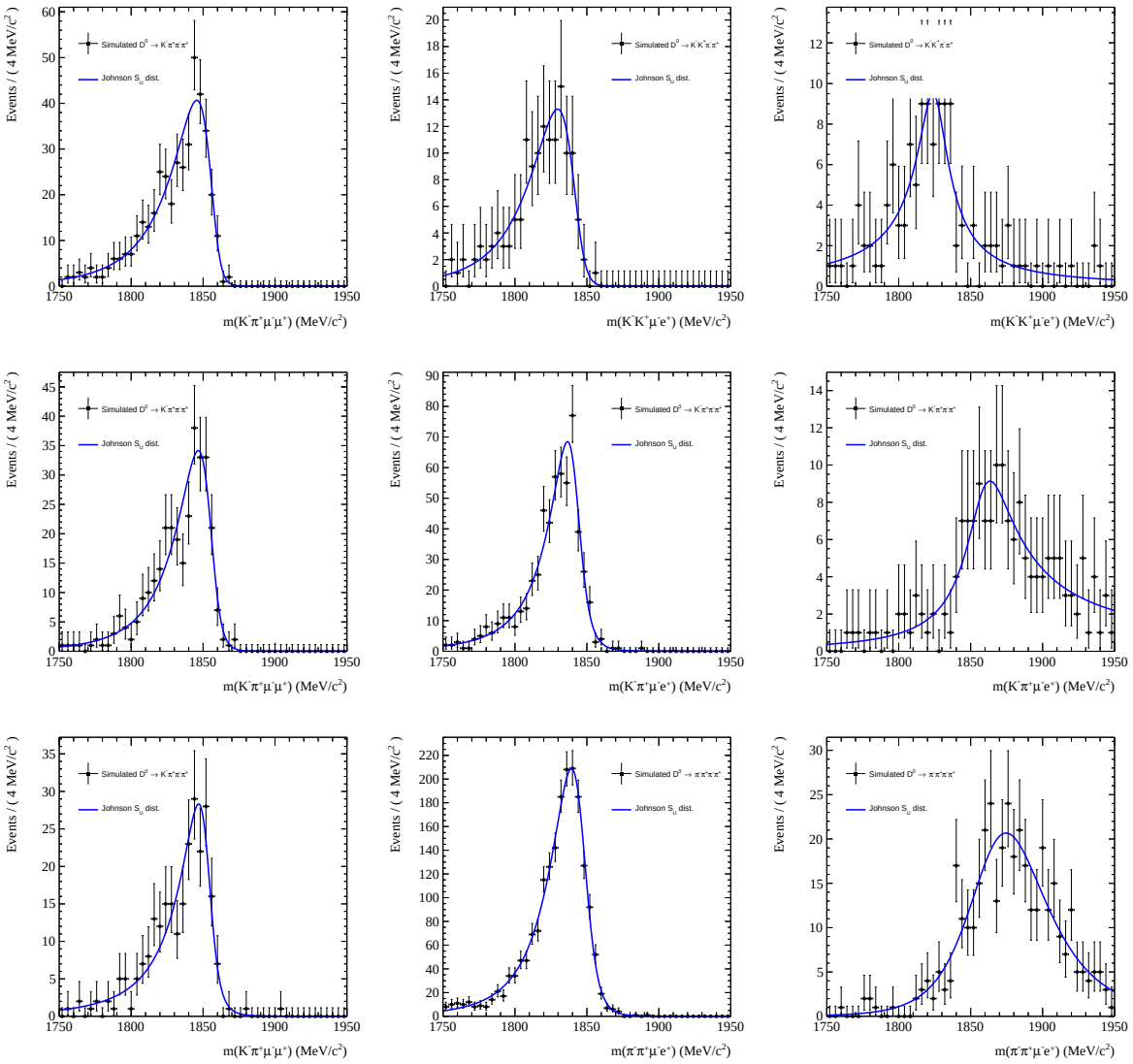


Figure F.4: MC fits to extract the shapes for the simultaneous fit for $D^0 \to K^-K^+\mu^\pm e^\mp$ (top), $D^0 \to K^-\pi^+\mu^\pm e^\mp$ (middle) and $D^0 \to \pi^-\pi^+\mu^\pm e^\mp$ (bottom). Left is the fit to the normalisation channel (the difference in the distribution is due to a different BDT selection). In the middle (right), the fit to the reconstructed invariant D^0 mass in category `noBrem` (`Brem`).

F.3 Fit results

In table F.1, the fit results for the data sample with opposite charge and in table F.2 with the same charge between the slow pion and muon are reported.

Parameter	$D^0 \to K^-K^+\mu^-e^+$	$D^0 \to K^-\pi^+\mu^-e^+$	$D^0 \to \pi^-\pi^+\mu^-e^+$
branchingRatio	$x.x \pm 1.3$	$x.x \pm 2.5$	$x.x \pm 2.3$
$\mathrm{N}^{1516}_{\mathrm{Comb,Brem}}$	1.8 ± 1.4	11.2 ± 3.5	6.5 ± 2.9
$\mathrm{N}^{1516}_{\mathrm{Comb,noBrem}}$	4.0 ± 3.0	23.5 ± 6.2	40.2 ± 8.2
$\mathrm{N}^{1516}_{\mathrm{Comb,norm}}$	11.4 ± 7.5	3.9 ± 4.1	10.7 ± 6.9
$\mathrm{N}^{1718}_{\mathrm{Comb,Brem}}$	26.3 ± 5.5	81.0 ± 10.0	246.0 ± 18.0
$\mathrm{N}^{1718}_{\mathrm{Comb,noBrem}}$	48.6 ± 9.5	119.0 ± 15.0	183.0 ± 19.0
$\mathrm{N}^{1718}_{\mathrm{Comb,norm}}$	92.0 ± 21.0	45.0 ± 14.0	35.0 ± 14.0
$\mathrm{N}^{1516}_{\mathrm{sig}}$	1244.0 ± 38.0	946.0 ± 32.0	1073.0 ± 35.0
$\mathrm{N}^{1718}_{\mathrm{sig}}$	5883.0 ± 82.0	4569.0 ± 71.0	5226.0 ± 76.0
$\mathrm{N}^{1516}_{\mathrm{misID,norm}}$	668.0 ± 29.0	181.0 ± 17.0	268.0 ± 20.0
$\mathrm{N}^{1718}_{\mathrm{misID,norm}}$	3335.0 ± 66.0	921.0 ± 38.0	1531.0 ± 47.0
$\mathrm{N}^{1516}_{\mathrm{misID}}$	7.8 ± 3.4	16.5 ± 5.7	29.8 ± 7.8
$\mathrm{N}^{1718}_{\mathrm{misID}}$	24.8 ± 8.2	117.0 ± 15.0	137.0 ± 18.0
$\nu^{1516}_{\mathrm{norm}}$	-0.18 ± 0.24	-0.86 ± 0.26	-0.72 ± 0.25
$\nu^{1718}_{\mathrm{norm}}$	-0.45 ± 0.1	-0.79 ± 0.11	-0.61 ± 0.11

Table F.1: Fit results of the simultaneous fit. The quoted branching ratio is blind.

Parameter	$D^0 \to K^-K^+\mu^+e^-$	$D^0 \to K^-\pi^+\mu^+e^-$	$D^0 \to \pi^-\pi^+\mu^+e^-$
branchingRatio	$x.x \pm 2.2$	$x.x \pm 4.2$	$x.x \pm 2.3$
$\mathrm{N}^{1516}_{\mathrm{Comb,Brem}}$	3.8 ± 2.0	21.2 ± 5.0	10.4 ± 3.8
$\mathrm{N}^{1516}_{\mathrm{Comb,noBrem}}$	8.1 ± 3.9	35.7 ± 7.9	28.1 ± 7.0
$\mathrm{N}^{1516}_{\mathrm{Comb,norm}}$	11.4 ± 7.5	3.9 ± 4.1	10.7 ± 6.9
$\mathrm{N}^{1718}_{\mathrm{Comb,Brem}}$	45.1 ± 7.3	234.0 ± 18.0	208.0 ± 16.0
$\mathrm{N}^{1718}_{\mathrm{Comb,noBrem}}$	83.0 ± 11.0	203.0 ± 21.0	176.0 ± 18.0
$\mathrm{N}^{1718}_{\mathrm{Comb,norm}}$	92.0 ± 21.0	45.0 ± 14.0	35.0 ± 14.0
$\mathrm{N}^{1516}_{\mathrm{sig}}$	1245.0 ± 38.0	946.0 ± 32.0	1074.0 ± 35.0
$\mathrm{N}^{1718}_{\mathrm{sig}}$	5882.0 ± 82.0	4569.0 ± 71.0	5225.0 ± 76.0
$\mathrm{N}^{1516}_{\mathrm{misID,norm}}$	668.0 ± 29.0	181.0 ± 17.0	268.0 ± 20.0
$\mathrm{N}^{1718}_{\mathrm{misID,norm}}$	3335.0 ± 66.0	921.0 ± 38.0	1531.0 ± 46.0
$\mathrm{N}^{1516}_{\mathrm{misID}}$	3.3 ± 2.7	30.0 ± 7.5	43.2 ± 8.4
$\mathrm{N}^{1718}_{\mathrm{misID}}$	0.4 ± 2.2	162.0 ± 20.0	176.0 ± 19.0
$\nu^{1516}_{\mathrm{norm}}$	-0.18 ± 0.24	-0.86 ± 0.26	-0.73 ± 0.25
$\nu^{1718}_{\mathrm{norm}}$	-0.45 ± 0.1	-0.79 ± 0.11	-0.61 ± 0.11

Table F.2: Fit results of the simultaneous fit. The quoted branching ratio is blind.

Eigenständigkeitserklärung

Hiermit bestätige ich, dass ich die vorliegende Arbeit selbständig verfasst habe und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Unterschrift .....