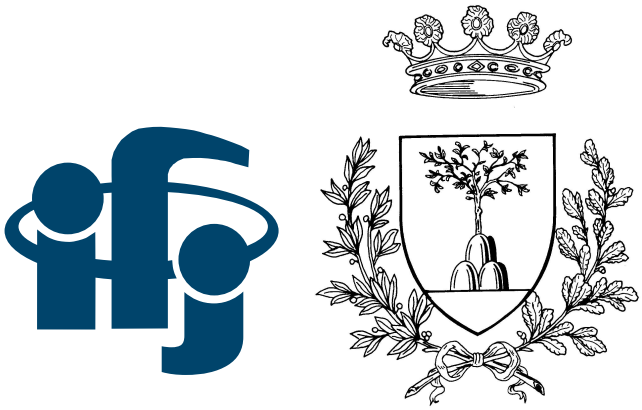


The Henryk Niewodniczański Institute of Nuclear Physics
Polish Academy of Sciences
University of Ferrara

Doctoral dissertation

defended by
Małgorzata Maria Pikies

Searches for New Physics in $b \to s\ell^+\ell^-$ transitions at the LHCb experiment

Thesis Supervisors:
dr hab. Marcin Chrząszcz, prof. IFJ PAN
prof. Roberto Calabrese
Auxiliary Supervisor:
dr Jihyun Bhom

Kraków, July 2018

Contents

List of Figures

List of Tables

Abstract

This thesis describes a search for new physics through the lepton flavour violating decay channel - $B^+ \to K^+ \mu^\pm e^\mp$. The analysis was performed with a data sample corresponding to an integrated luminosity of 3 fb^{-1} gathered by the LHCb experiment from proton - proton collisions in the LHC accelerator. The collisions were performed at a centre-of-mass energy of 7 TeV (8 TeV) in 2011 (2012). No signal was observed, thus the upper limits were set to be:

- $\mathscr{B}(B^+ \to K^+ \mu^+ e^-) < 8.3 \times 10^{-9}$,
- $\mathscr{B}(B^+ \to K^+ \mu^- e^+) < 6.1 \times 10^{-9}$.

Streszczenie

Praca doktorska opisuje poszukiwania nowej fizyki poprzez rozpad łamiący zasadę zachowania zapachu leptonów - $B^+ \to K^+ \mu^\pm e^\mp$. Analiza została przeprowadzona na próbce danych odpowiadającej 3 fb^{-1} scałkowanej świetlności, zebranej przez eksperyment LHCb ze zderzeń proton-proton w akceleratorze LHC. Zderzenia przeprowadzono dla układu środka masy wynoszącej 7 TeV (8 TeV) w latach 2011 (2012). Nie zaobserwowano sygnału, tak więc górny limit został oszacowany:

- $\mathscr{B}(B^+ \to K^+ \mu^+ e^-) < 8.3 \times 10^{-9}$,
- $\mathscr{B}(B^+ \to K^+ \mu^- e^+) < 6.1 \times 10^{-9}$.

Acknowledgments

I would like to acknowledge my supervisor dr hab. Marcin Chrząszcz for his support and for sharing his skills with me. I would like to thank prof. dr hab. Tadeusz Lesiak who drew me into the Institute of Nuclear Physics and made my PhD possible.

I would also like to thank all the founding agencies: KNOW for the PhD scholarship and the National Science Center for my grant.

Last but not least, I would like to thank my parents for supporting all my life choices which have led me to where I am now.

CHAPTER 1

Introduction

The Standard Model (SM) of elementary particles is currently the best model describing building blocks of our Universe and their interactions. In physics, there are four known fundamental forces: strong, weak, electromagnetic, and gravity. The SM describes the first three of them. The strong force is mediated by gluons, and is formulated in terms of Quantum Chromodynamics (QCD). The weak force is mediated by $W^{\pm}$ and Z bosons, and is responsible for radioactive decays. The last quantum force is electromagnetic, mediated by the photon (γ) and responsible for interactions between electrically charged particles . The theory describing electromagnetic interactions is quantum electrodynamics (QED). Weak and electromagnetic forces are unified into a single theory of electroweak interactions (1979 Nobel Prize in Physics for S. Glashow, A. Salam, and S. Weinberg [1]), which together with the QCD create the SM. Even though the SM is considered to be completed since its last missing particle (the Higgs boson) has been discovered (2013 Nobel Prize in Physics for F. Englert and P. W. Higgs) [2, 3], it is still not satisfactory enough to describe the Universe around us, as it fails to explain the following experimental observations:

- the observed mater-antimatter asymmetry exceeds its SM predictions;
- only 5% of the Universe is built from known matter particles; the rest of the Universe is composed of dark energy (70%) and dark matter particles (20%), which do not interact with any known force;
- the observation of the neutrino oscillations indicates that they have masses, which is the opposite of the initial SM assumption. This problem can be solved for example by adding *ad hoc* dimension-5 operators. The resulting effective field theory is quite often refereed to as the SM.

In order to resolve the flaws of the SM many extensions have been proposed, but so far no convincing experimental evidence of beyond the SM (BSM) physics has been observed. There are two main approaches to search for BSM physics in high energy collisions, with the first one being a direct search for non-SM particles, and the other - an indirect search. The latter can infer the existence of BSM physics by precise measurements of processes which are very well described in the SM. This approach allows us to test the SM predictions potentially caused by particles heavier than the energy available in the centre-mass frame. For the latter very promising are the Flavour Changing Neutral Current (FCNC) processes, such as $b \to s\ell^+\ell^-$ transitions. Those decays are highly suppressed in the SM as they can occur only through electroweak penguin and box diagrams. As they are very rare, they are sensitive to new physics contributions.

The LHCb experiment has made several measurements of B meson decays, showing discrepancies with the SM predictions. Among many, there are the lepton flavour universality (LFU) tests such as R_K, R_{K^*}, and R_{D^*}, which showed tensions within 2-4 standard deviations from the SM.

This thesis describes the search for the charged lepton flavour violating (cLFV) decay $B^+ \to K^+\mu^\pm e^\mp$ (charge conjugation is included in this monograph). The SM branching fraction of this decay occurs via neutrino oscillations, thus the measurement of this decay is beyond any experimental reach. The observation of the lepton flavour violating transition would be a clear sign of BSM physics.

This thesis is organised in the following way: Chapter 2 describes the theoretical description of the SM together with recently seen anomalies and some selected BSM models. Chapter 3 describes the experimental apparatus used to collect high quality data necessary for this search. Chapter 4 describes the search for the $B^+ \to K^+\mu^\pm e^\mp$ decay and its results. The work performed in this thesis is summarised in Chapter 5.

CHAPTER 2

Theoretical description

This chapter describes the theoretical overview of and the motivation behind the study of the $b \to s\ell\bar{\ell}$ transitions. The first section contains a short summary of the currently best theoretical description of known elementary particles. The second section briefly summarises popular BSM models related to the subject of this thesis. The third section shows interesting tensions within the SM predictions in recent measurements, which became a motivator for this analysis.

2.1 The Standard Model

2.1.1 The field content of the Standard Model

The Standard Model is a Quantum Field Theory (QFT) description of fields with spins 0, 1/2, and 1. Half-integer fields are called fermions, and they can be divided into quarks and leptons. Only left-handed fermions form weak isospin (T_3) doublets, where the handedness of a particle defines whether its spin projection onto the momentum vector is positive (R, for right-handed particles) or negative (L, for left-handed particles). The weak isospin describes how a particle transforms under the SU(2) group. Its possible values are +1/2 (u, c, t quarks and neutrinos) and -1/2 (d, s, b quarks, and charged leptons). The left-handed lepton field components ($L_{\ell=e,\mu,\tau}$) can be expressed as:

$$L_\ell = \frac{1-\gamma_5}{2}\begin{pmatrix} \psi_{\nu_\ell} \\ \psi_\ell \end{pmatrix}, \tag{2.1}$$

where the upper component corresponds to neutral neutrinos, and the lower one corresponds to charged leptons. The right-handed ($R_{\ell=e,\mu,\tau}$) components form only singlets

(there is no right-handed neutrino component), as defined below:

$$R_\ell = \frac{1+\gamma_5}{2}\psi_\ell. \tag{2.2}$$

Quarks come in colour triplets and are bound together by the strong force into states called hadrons. A very interesting aspect of the strong force is the fact that its strength grows with distance. Quarks are confined within hadrons, so it is impossible to measure a single quark. All observed hadrons have a colour singlet 1 (they are colourless). Quark - anti-quark ($q\bar{q}$) pairs constitute mesons, three quarks (qqq) form baryons, and higher quark combinations create exotic hadron states like tetraquarks ($q\bar{q}q\bar{q}$) and pentaquarks ($qqqq\bar{q}$). All quarks have fractional charge 2/3 (u, c, t quarks) or -1/3 (d, s, b quarks) [4–6].

All particles can be characterised by quantum numbers. Mathematically, a quantum number is an eigenvalue of an operator, which is acting on a specific state. Multiplicative quantum numbers are correlated with discrete symmetries (e.g. C, P, CP, T symmetries), while additive quantum numbers are associated with the continuous symmetry group (e.g. electric charge). Lepton number (+1 for all leptons, and -1 for their anti-leptons) is conserved without exceptions. Moreover, processes like a decay of a muon into an electron and a photon have never been observed not without a reason. Even tough $\mu^+ \to e^+ + \gamma$ conserve a lepton number and a charge, they do not fulfil yet another conservation law. That law is the lepton flavour conservation, according to which, the lepton flavour (e, μ, τ) is supposed to be conserved in particle interactions. It was thought to be conserved without exceptions, up to the discovery of neutrino oscillations, which led to the lepton flavour violation in the neutral lepton sector (the flavour violation will be discussed in detail in Sec. 2.1.3 and 2.1.4). The electron, muon, and taon all have a negative electric charge -1, and the corresponding neutrinos are electrically neutral. Fermions properties are summarized in Tab. 2.1 and 2.2 below.

Quark	Mass [MeV/c^2]	Charge [e^-]	T_3
u	$2.2^{+0.6}_{-0.4}$	$+2/3$	$+1/2$
d	$4.7^{+0.5}_{-0.4}$	$-1/3$	$-1/2$
c	1280 ± 3	$+2/3$	$+1/2$
s	96 ± 5	$-1/3$	$-1/2$
t	$(160^{+5}_{-4}) \times 10^3$	$+2/3$	$+1/2$
b	$4180^{+0.04}_{-0.03}$	$-1/3$	$-1/2$

Table 2.1: Properties of the SM quarks [7], T_3 - weak isospin.

The SM fundamental interactions, are generated by the gauge symmetry:

$$G_{SM} = SU(3)_C \times SU(2)_W \times U(1)_Y, \tag{2.3}$$

where the subscript C denotes the colour, the subscript W denotes the weak isospin, and Y denotes the hypercharge. The $SU(3)_C$ group is tied with strong interactions, while the $SU(2)_W \times U(1)_Y$ is tied with electroweak interactions.

Lepton	Mass [MeV/c^2]	Charge [e^-]	L	L_e	L_μ	L_τ
ν_e	< 2	0	1	1	0	0
e	0.5109989461	-1	1	1	0	0
ν_μ	< 0.19	0	1	0	1	0
μ	0.1134289257	-1	1	0	1	0
ν_τ	< 18.2	0	1	0	0	1
τ	1776.86 ± 0.12	-1	1	0	0	1

Table 2.2: Properties of the SM leptons [7], L - lepton number, L_ℓ - lepton flavours.

We can distinguish massless gluons, which are responsible for strong interactions among quarks G^A_μ, where $A = 1, ..., 8$ stand for 8 kinds of a gluon, depending on their colour combinations, and μ=0,1,2,3 stands for the timelike (0) and spacelike (1,2,3) components of the gluon field. The remaining spin 1 fields are massless photon γ responsible for electromagnetic interactions, and massive $W^\pm$ and Z^0 bosons, which mediate weak interactions. Gauge bosons and their masses are listed in Tab. 2.3 below. Last but not least there is the spin-0 field, which spontaneously breaks the $SU(2)_W \times U(1)_Y$ symmetry [8, 9].

Boson	Mass [GeV/c^2]	Interaction
γ	0	electromagnetic
g	0	strong
$W^\pm$	80.385 ± 0.015	weak
Z	91.1876 ± 0.0021	weak

Table 2.3: Gauge bosons of the SM responsible for fundamental interactions and masses [7].

The SM Lagrangian in terms of the aforementioned fields is given with the following formula:

$$\mathscr{L} = \mathscr{L}_{gauge} + \mathscr{L}_{Higgs} + \mathscr{L}_{Yukawa}, \tag{2.4}$$

where the first $\mathscr{L}_{gauge}$ term contains the above-mentioned gauge interactions, the $\mathscr{L}_{Higgs}$ describes the Higgs interaction, and $\mathscr{L}_{Yukawa}$ couples the Higgs field to fermions.

2.1.2 Weak interactions and the electroweak theory

Weak interactions occur among all fermions. One of the first observed weak decays is the β-decay in atoms:

$$^{60}_{27}\mathrm{Co} \rightarrow ^{60}_{28}\mathrm{Ni}\ \mathrm{e}^-\ \overline{\nu}_\mathrm{e}\ 2\gamma, \tag{2.5}$$

which was used in the Wu experiment [10] to show that weak interactions did not conserve parity P. Anti-neutrinos always have their spin pointed in the direction parallel to their velocity (they are right-handed). The lack of the ν_R and $\overline{\nu}_L$ also implies a charge C conjugation violation, with the conservation of the CP operation. The comparison of the decay rates of another β-decay, $^{14}\mathrm{O} \rightarrow ^{14}\mathrm{N}^{*}\ \mathrm{e}^{+}\nu_{\mathrm{e}}$, with the purely leptonic muon decay $\mu^{-} \rightarrow e^{-}\overline{\nu}_{e}\nu_{\mu}$ leads to the conclusion that both decays have the same physics origin. Their decay rates are proportional to the (weak coupling) Fermi constant G_F, which enters the current-current weak interactions Lagrangian :

$$\mathscr{L}_W = -\frac{G_F}{2}\sum_{\alpha=0}^{\alpha=3} J^{\alpha}(x)^{\dagger} J_{\alpha}(x), \tag{2.6}$$

with currents given by:

$$\begin{aligned} J_{\alpha}(x) &= \sum_{\ell=e,\mu,\tau} \overline{\nu}_{\ell}(x)\gamma_{\alpha}(1-\gamma_5)\ell(x), \\ &= \sum_{\ell=e,\mu,\tau} \overline{\nu}_{\ell}(x)\gamma_{\alpha}\ell(x) - \sum_{\ell=e,\mu,\tau} \overline{\nu}_{\ell}(x)\gamma_{\alpha}\gamma_5\ell(x), \\ &= V_{\alpha}(x) - A_{\alpha}(x). \end{aligned} \tag{2.7}$$

The position four-vector is $x = (t,\vec{x})$, the γ_{α} are the Dirac gamma matrices. The charge-lowering current (corresponding to the W^{-} boson) is the $J^{\alpha}(x)^{\dagger}$ contribution, and the $J_{\alpha}(x)$ is the charge-raising contribution (corresponding to the W^{+} boson). The $\gamma_{\alpha}(1-\gamma_5)$ structure of the weak current gives raise to the aforementioned parity violation, according to which the weak current J_{α} decomposes into the parity conserving vector (V_{α}) and parity violating axial (A_{α}) terms.

The weak Lagrangian gives a precise and experimentally confirmed description of many weak processes, such as the previously mentioned muon decay $\mu^{-} \rightarrow e^{-}\overline{\nu}_{e}\nu_{\mu}$. However, there are numerous measurements that imply the existence of the neutral weak current. With the new contribution from neutral currents (they correspond to the Z^0 and γ bosons), the weak interaction phenomenology became a renormalizable electroweak theory. The new weak neutral current $J^3_{\mu}(x)$, which was built analogically to already known charged weak currents, was missing a right-handed component. The electromagnetic neutral current consists of both right- and left-handed components. It is given by:

$$j^{em}_{\mu} = -\bar{L}_{\ell}\gamma_{\mu}QL_{\ell}, \tag{2.8}$$

where L_{ℓ} is the isospin doublet from Eq. 2.1, and Q is the charge operator. The electromagnetic current can be given by the combination of two new currents:

$$j^{em}_{\mu} = J^3_{\mu} + \frac{1}{2}j^{Y}_{\mu}, \tag{2.9}$$

where J^3_μ is the neutral current, and j^Y_μ is the weak hypercharge current. The weak hypercharge is defined with:

$$Q = T^3 + \frac{Y}{2}, \tag{2.10}$$

where T^3 is the previously mentioned (third component) weak isospin. The weak isospin and hypercharge quantum numbers of leptons are summarised in Tab. 2.4.

	T^3	Q	Y
Leptons			
$\nu_{\ell\,L}$	1/2	0	-1
ℓ_L	-1/2	-1	-1
ℓ_R	0	-1	-2

Table 2.4: Electroweak quantum numbers for leptons.

2.1.3 Quark mixing

Originally, quarks were considered to change their flavour only within their generation. An observation of the decay $K^+ \to \mu\nu_\mu$ was a clear sign of the quark mixing, because K^+ is made of u and $\bar{s}$ quarks. The generalized $N \times N$ $(N = 3)$ matrix that connects three flavour eigenstates of the down-type quarks (d', s', b') with the corresponding three mass eigenstates (d, s, b) is called the Cabibbo-Kobayashi-Maskawa (CKM) matrix.

The relation between the flavour and mass eigenstates is:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = V_{CKM} \begin{pmatrix} d \\ s \\ b \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \tag{2.11}$$

The currently best known moduli values of the V_{CKM} matrix parameters are [7]:

$$V_{CKM} = \begin{pmatrix} 0.97417 \pm 0.00021 & 0.2248 \pm 0.0006 & (4.09 \pm 0.39) \times 10^{-3} \\ 0.220 \pm 0.005 & 0.995 \pm 0.016 & (40.5 \pm 1.5) \times 10^{-3} \\ (8.2 \pm 0.6) \times 10^{-3} & (40.0 \pm 2.7) \times 10^{-3} & 1.009 \pm 0.031 \end{pmatrix}. \tag{2.12}$$

The transition probability between the i and j type quarks is given by the $|V_{ij}|^2$. The hierarchical structure of the matrix can be seen with the Wolfenstein parametrisation:

$$V_{CKM} = \begin{pmatrix} 1 - \lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \lambda^2/2 & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4), \tag{2.13}$$

with:

$$\begin{aligned} \lambda &= \sin(\theta_{12}), \\ A\lambda^2 &= \sin(\theta_{23}), \\ A\lambda^3(\rho - i\eta) &= \sin(\theta_{13})e^{i\delta}. \end{aligned} \tag{2.14}$$

The angles θ_{12}, θ_{23}, and θ_{13} are quark mixing matrices, and δ is the CP violating phase.

2.1.4 Lepton flavour violation

At the time the SM was being created, neutrinos were considered to be massless particles; however, an observation of neutrino oscillations [11] suggests that neutrinos have masses. Analogically to the quark mixing matrix, Pontecorvo, Maki, Nakagawa, and Sakata created the PMNS neutrino mixing matrix:

$$V_{PMNS} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{i\delta} \\ -s_{13}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & c_{23}c_{13} \\ s_{12}s_{23} - c_{12}s_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix}, \tag{2.15}$$

where $s_{mn} = \sin\theta_{mn}$ and $c_{mn} = \cos\theta_{mn}$ are mixing angles, and δ is the CP violating phase. In contrast to the CKM matrix, the PMNS matrix does not show a hierarchical structure [12]:

$$V_{PMNS} = \begin{pmatrix} 0.800 \to 0.844 & 0.515 \to 0.581 & 0.139 \to 0.155 \\ 0.229 \to 0.516 & 0.438 \to 0.699 & 0.614 \to 0.790 \\ 0.249 \to 0.528 & 0.462 \to 0.715 & 0.595 \to 0.776 \end{pmatrix}. \tag{2.16}$$

The above intervals correspond to three standard deviations from the central values. The difference between the structures of CKM and PMNS matrices is visible but its origin is unknown. The observation of neutrino oscillations was a clear sign of lepton flavour violation in the neutral lepton sector. Lepton flavour violation in the charged lepton sector has not yet been observed.

2.1.5 Effective field theory

In the Operator Product Expansion (OPE) method, the electroweak scale and above are integrated out, and the effective weak Hamiltonian has the following structure:

$$\mathscr{H}_{eff} = \frac{G_F}{\sqrt{2}} \sum_i V^i_{CKM} C_i(\mu) Q_i, \tag{2.17}$$

where G_F is the previously mentioned Fermi constant and V^i_{CKM} is an adequate CKM matrix element. Q_i is relevant for the decay local operator, which represents the effective point-like vertices, and the corresponding C_i Willson coefficient can be interpreted as

an effective coupling constant. Physics contributions are separated accordingly to the energy scale μ into short-distance (C_i) and long-distance (Q_i) contributions. Fig. 2.1 below shows quark-level diagrams of the β-decay in the full (left) and effective (right) theory.

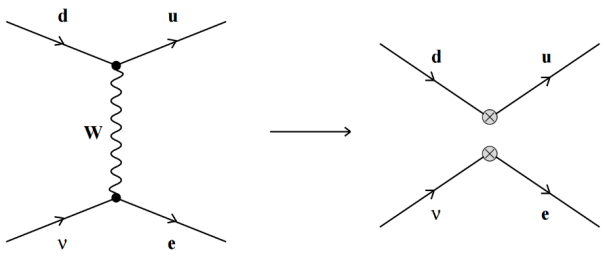


Figure 2.1: β-decay at the quark level. In the effective theory the W propagator is replaced by an effective vertex [13].

Title flavour changing neutral current (FCNC) $b \to s\ell^+\ell^-$ processes can be described within the effective field theory (EFT). Their electroweak penguin (EWP) diagram is showed in Fig. 2.2, where the b quark passes into the s through the t state, emitting two leptons. The transition is governed by electroweak W, γ, and Z bosons. Within the EFT

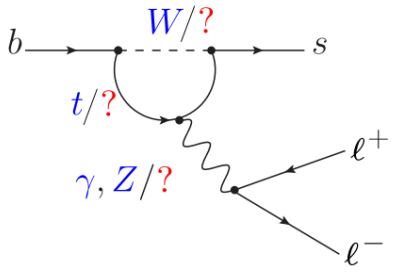


Figure 2.2: Electroweak penguin diagram for the FCNC $b \to s\ell^+\ell^-$ transition. Possible new particles can enter this diagram.

approach, the effective Hamiltonian is given by the formula:

$$\mathscr{H}_{eff} = -\frac{4G_F}{\sqrt{2}} V_{tb}V_{ts}^* \sum_i [C_i(\mu)O_i(\mu) + C_i'(\mu)O_i'(\mu)], \tag{2.18}$$

$$
\begin{aligned}
&i = 1,2 && \text{Tree,} \\
&i = 3-6,8 && \text{Gluon penguin,} \\
&i = 7 && \text{Photon penguin,} \\
&i = 9,10 && \text{Electroweak penguin,} \\
&i = S && \text{Higgs (scalar) penguin,} \\
&i = P && \text{Pseudoscalar penguin,}
\end{aligned}
\tag{2.19}
$$

where different Willson coefficients correspond to different physics contributions.

Most heavy flavour physics measurements are interpreted within the EFT. For recent results see Sec. 2.2.1.

2.2 Beyond the Standard Model

2.2.1 Recent anomalies

In the past few years experiments dedicated to heavy flavour physics have reported many anomalies, some of them deviating even at the level of 3-4 σ from the Standard Model predictions. New physics searches can be segregated into lepton flavour universality tests, differential branching fraction measurements, and angular observables measurements.

Lepton universality requires the same behaviour from leptons in the cases where the lepton mass can be neglected. As mentioned in Sec. 1, the LHCb experiment has performed several LFU tests. The mean way to do that is a theoretically very clean type of an observable, the branching fraction ratio:

$$
R_K = \frac{\mathscr{B}(B^+ \to K^+ \mu^+ \mu^-)}{\mathscr{B}(B^+ \to K^+ e^+ e^-)}, R_{K^*} = \frac{\mathscr{B}(B^0 \to K^{*0} \mu^+ \mu^-)}{\mathscr{B}(B^0 \to K^{*0} e^+ e^-)}, \qquad R_{D^*} = \frac{\mathscr{B}(\bar{B}^0 \to D^{*+} \tau^- \overline{\nu}_\tau)}{\mathscr{B}(\bar{B}^0 \to D^{*+} \ell \overline{\nu}_\ell)}, \ell = \mu, e. \tag{2.20}
$$

According to the SM predictions, ratios involving muons and electrons should be equal to unity (excluding the low-q^2 region) [14–16]. In the case of taons, their mass cannot be neglected, thus the ratio lies below unity [17]. Measured by the LHCb experiment, the values of those ratios are within 2-3 σ from their predicted SM values [18–20]:

$$
R_K = 0.745^{+0.090}_{-0.074} \pm 0.036 \quad (1 < q^2 < 6)\ \text{GeV}^2/c^4, \tag{2.21}
$$

$$
R_{K^*} = \begin{cases} 0.66^{+0.11}_{-0.07} \pm 0.03 & (0.045 < q^2 < 1.1)\ \text{GeV}^2/c^4, \\ 0.069^{+0.11}_{-0.07} \pm 0.05 & (1.1 < q^2 < 6.0)\ \text{GeV}^2/c^4, \end{cases} \tag{2.22}
$$

$$
R_{D^*} = 0.336 \pm 0.027(stat) \pm 0.030(syst). \tag{2.23}
$$

The LHCb collaboration has made a series of important EWP measurements, such as differential branching fraction measurements and angular analyses. Among many

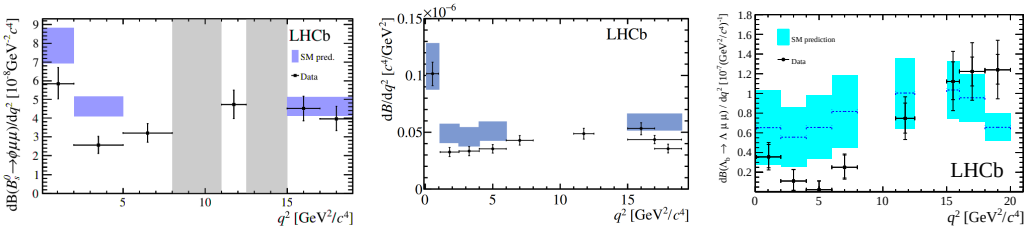


Figure 2.3: Differential branching fractions distributions for $B_s^0 \to \phi\mu\mu$ (left), [21] $B^0 \to K^{*0}\mu^+\mu^-$ (middle) [22], and $\Lambda_b \to \Lambda\mu\mu$ (right) [23].

interesting results, the branching fraction measurements of $B^0 \to K^{*0}\mu^+\mu^-$, $B_s^0 \to \phi\mu\mu$, and $\Lambda_b \to \Lambda\mu\mu$ decay channels have showed deviations from the SM predictions in the low di-lepton mass squared (q^2) region, as can be seen in Fig. 2.3.

Special attention should be put on the angular analysis of the $B^0 \to K^{*0}\mu^+\mu^-$. This four body decay gives a full angular description and access to three helicity angles ($\cos\theta_\ell$, $\cos\theta_K$, ϕ). Its decay rate can be given in terms of those angles and a q^2:

$$\begin{aligned}\frac{d^4\Gamma}{dq^2\, d\cos\theta_l\, d\cos\theta_K\, d\phi} = \frac{9}{32\pi}\Big[&J_1^s \sin^2\theta_K + \\ &J_1^c \cos^2\theta_K + J_2^s \sin^2\theta_K \cos 2\theta_l + \\ &J_2^c \cos^2\theta_K \cos 2\theta_l + J_3 \sin^2\theta_K \sin^2\theta_l \cos 2\phi + \\ &J_4 \sin 2\theta_K \sin 2\theta_l \cos\phi + J_5 \sin 2\theta_K \sin\theta_l \cos\phi + \\ &J_6 \sin^2\theta_K \cos\theta_l + J_7 \sin 2\theta_K \sin\theta_l \sin\phi + \\ &J_8 \sin 2\theta_K \sin 2\theta_l \sin\phi + J_9 sin^2\theta_K \sin^2\theta_l \sin 2\phi\Big].\end{aligned} \tag{2.24}$$

The angular observables J_i can be combined into CP-averages S_j and CP-asymmetries A_j:

$$S_j = (J_j + \bar{J}_j)/\Big(\frac{d\Gamma}{dq^2} + \frac{d\bar{\Gamma}}{dq^2}\Big),\;\; A_j = (J_j - \bar{J}_j)/\Big(\frac{d\Gamma}{dq^2} + \frac{d\bar{\Gamma}}{dq^2}\Big). \tag{2.25}$$

Very clean observables, free from theoretical form-factor uncertainties, can be extracted by dividing them by the $\sqrt{F_L(1-F_L)}$ [24]:

$$P'_{4,5} = S_{4,5}/\sqrt{F_L(1-F_L)}, \tag{2.26}$$

where F_L is the longitudinal polarization fraction of the K^{*0}. The distribution of the P'_5 shows a behaviour in the low-q^2 region (see Fig. 2.4) similar to what was observed in the differential branching fraction measurements.

All aforementioned decays are from the family of the FCNC $b \to s\ell^+\ell^-$ processes. They are highly suppressed in the SM and can only occur through EWP loop and box diagrams, which makes them sensitive to possible non-SM effects. The low-q^2

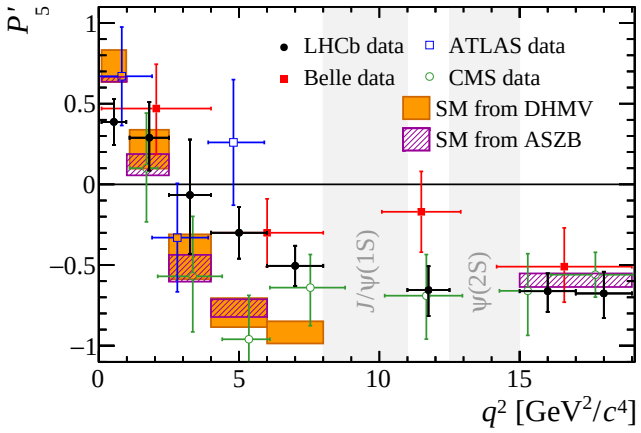


Figure 2.4: A P'_5 distribution in the $B^0 \to K^{*0}\mu^+\mu^-$ decay channel [25–28].

region is mostly sensitive to C_7 and C_9 Wilson coefficients (see Feg. 2.5). What is interesting is the fact that the same corrections to Wilson coefficients make up for the existing discrepancies. The SM central value for $Re(C_9)$ is 4.27. In the case of the $B^0 \to K^{*0}\mu^+\mu^-$, the best fit-point corresponds to the $\Delta Re(C_9) = -1.04 \pm 0.25$, which corresponds to 3.4 standard deviations [25].

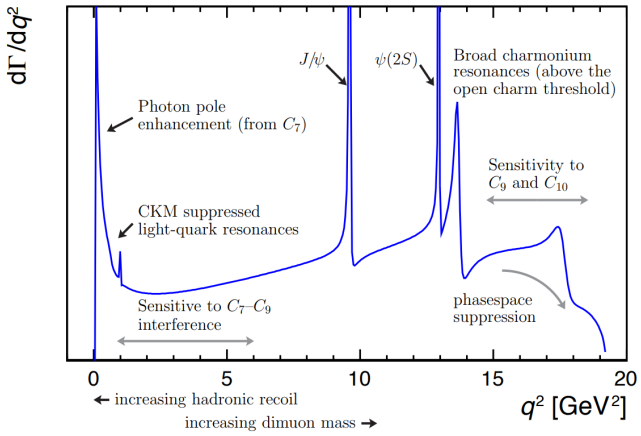


Figure 2.5: Wilson Coefficients sensitivity as a function of q^2 [29].

The above-mentioned results may be explained by a few BSM models, in which modifications to coupling to muons are favoured. Particularly, a change in couplings with vector or axial-vector operators models can explain the afore-mentioned measurements. Among those models are leptoquark models and new heavy bosons, which are briefly described in the next sections.

2.2.2 Leptoquarks

Conforming to all leptoquark models, new hypothetical particles called leptoquarks, which carry a lepton and baryon number, may explain the currently seen discrepancies within the SM predictions. Those particles allow interactions between quarks and leptons as shown in Fig. 2.6.

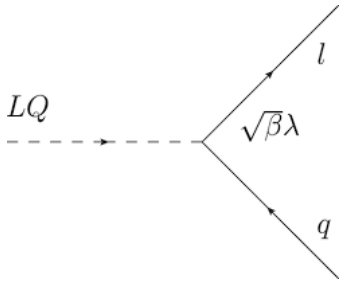


Figure 2.6: An example of leptoquark LQ interaction with a lepton ℓ and a quark q.

Leptoquark models can explain results like $R_K^{exp} < R_K^{SM}$, $R_{K^*}^{exp} < R_{K^*}^{SM}$, and $R_{D^*}^{exp} > R_{D^*}^{SM}$. Possible scalar leptoquarks are listed in Table 2.5, while vector leptoquarks are omitted since their models are not renormalizable. Within these models only the $(\bar{3},3)_{1/3}$ leptoquark can explain both R_K and R_{K^*} results on its own. Yet another possibility is the amended $(3,2)_{7/6}$ leptoquark model, which forbids tree-level contributions, thus only a loop-level can cause a shift in the C_9 Wilson coefficient.

$(SU(3)_c, SU(2)_L)_{U(1)_Y}$	Couplings	R_K/R_K^{SM}
$(\bar{3},3)_{1/3}$	$C_9 = -C_{10}$	< 1
$(\bar{3},1)_{4/3}$	$(C_9)' = (C_{10})'$	≈ 1
$(3,2)_{7/6}$	$C_9 = C_{10}$	> 1
$(3,2)_{1/6}$	$(C_9)' = -(C_{10})'$	< 1

Table 2.5: Scalar LQ states that modify $\mathscr{B}$ $(B^+ \to K^+\mu^+\mu^-)$, the corresponding Wilson coefficients, and R_K predictions [30].

2.2.3 Heavy bosons

Another new physics scenario important for this thesis is the possible existence of new electroweak charged and neutral heavy bosons called $W^{\pm}$' and Z' respectively. Vector bosons, like those particles, could be the cause of the observed anomalies in rare decays or they could give signals to hadron colliders if they are not too heavy. An example of a Z' boson decaying to a pair of lepton - anti-lepton is shown in Fig. 2.7.

Local gauge invariances of the SM give rise to already known strong and electroweak bosons. Models with new heavy bosons are obtained by extending the symmetry group.

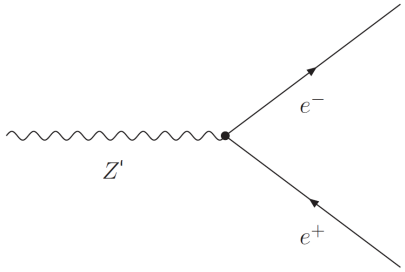


Figure 2.7: A Z' decaying to an electron and a positron.

As an example, a heavy neutral boson with mass $M_{Z'} = 1\,\text{TeV}$ would imply the new physics contribution to the C_9 Wilson coefficient to be ~ -1.5 [31]. In contrast to leptoquark models, Z'models cannot explain the R_{D^*} anomaly.

2.3 Analysis motivation

As mentioned in Sec. 1, the recent results point towards the lepton non-universality, which could be linked with the lepton flavour violation.

There are several new physics which allow the searched $B^+ \to K^+\mu^\pm e^\mp$ decay to occur. In one scenario, the mixing matrix from neutrino oscillations was used on charged leptons. Two possible scenarios of the leptonic 3×3 matrix $C_L = K\tilde{C_L}K^\dagger$ were investigated [32]:

$$\tilde{C_L} = diag(0,\varepsilon_A,1),\ \tilde{C_L} = diag(\varepsilon_B,0,1), \tag{2.27}$$

where K is the leptonic mixing matrix obtained from the neutrino oscillations measurements, and $\varepsilon \ll 1$. The small parameter was given in terms of the mixing angles:

$$\varepsilon_A \sim -\frac{\tan^2\theta_{13}}{\sin^2\theta_{12}},\ \varepsilon_B \sim -\frac{\tan^2\theta_{13}}{\cos^2\theta_{12}}. \tag{2.28}$$

Branching fractions for $B \to K\ell_{\,i}\,\ell_{\,j}$ (i and j denote different flavours of leptons) were estimated to be:

$$\begin{aligned}
\mathscr{B}(B \to Ke^\pm\mu^\mp) &\sim 10^{-10},\\
\mathscr{B}(B \to Ke^\pm\tau^\mp) &\sim 10^{-9},\\
\mathscr{B}(B \to K\mu^\pm\tau^\mp) &\sim 10^{-8}.
\end{aligned} \tag{2.29}$$

The $B \to K\mu^\pm\tau^{\,\mp}$ decay has the highest branching fraction, but τ leptons are experimentally very challenging at LHCb, hence the selection of the $B^+ \to K^+\mu^\pm e^\mp$ decay channel. For its branching fractions distribution see Fig. 2.8.

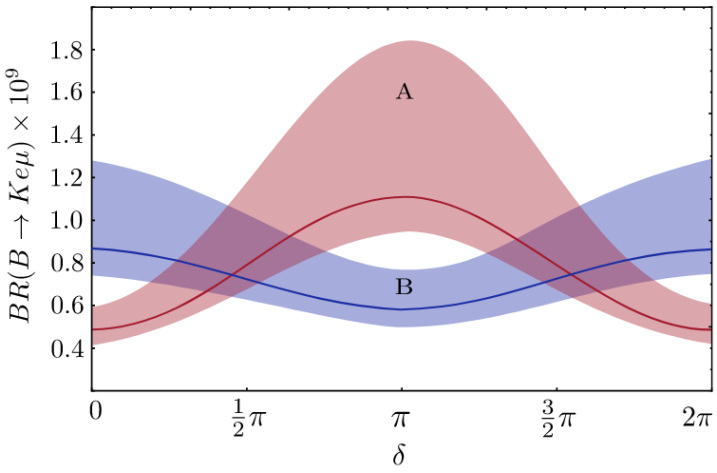


Figure 2.8: $B^+ \to K^+\mu^\pm e^\mp$ branching fraction distribution for both leptonic mixing matrix scenarios, as a function of the CP violating phase δ [32].

In the vector leptoquark scenario [33] with the discrete non-abelian symmetry A_4 the lepton mixing is obtained. The Higgs vacuum expectations values (VEV) for u- and d- quarks were similar. The parametric suppression of leptoquark youkawas is given:

$$(Y_{AB})_{ij} \sim \lambda^{(q(A_i)+q(B_j)}, \ (Y_{\bar{A}B})_{ij} \sim \lambda^{(-q(A_i)+q(B_j)}, \qquad A = Q,U,D, \ B = L,E, \tag{2.30}$$

where $\lambda \sim 0.2$, A stands for the quark Q, lepton L doublets, or U quark singlet, and B can stand for the quark singlet U, or charged leptons E. In the above equation, an interference between the quark and lepton charges is clear. The branching fraction of the searched decay can be enhanced to even higher values in the specific leptoquark scenario:

$$\mathscr{B}(B \to Ke^\pm\mu^\mp) \sim 3 \times 10^{-8} \left(\frac{1-R_K}{0.23}\right)^2. \tag{2.31}$$

The above estimation with the recent R_K measurement gives the value which is within the LHCb experimental reach in the next years of data-taking.

The branching fraction of the searched decay channel $B^+ \to K^+\mu^\pm e^\mp$ is highly suppressed in the SM, thus it is beyond any experimental reach. Any observation would be a clear signal of the physics beyond the SM.

CHAPTER 3

Experimental setup

The analysis described in this dissertation was performed using data from high energy proton - proton collisions at the Large Hadron Collider (LHC), collected by the Large Hadron Collider beauty (LHCb) experiment. In this section, a brief overview of the aforementioned experimental apparatus together with the data collection specifics is given.

3.1 The Large Hadron Collider

The LHC is a 27 km long accelerator, which makes it the biggest particle collider in the world. It is located at the European Organization for Nuclear Research (CERN) near Geneva, Switzerland. The LHC accelerator complex is schematically shown in Fig. 3.1. A hydrogen gas, stripped from electrons by the electric field, serves as a proton source. The protons first enter the linear accelerator called Linac 2, which accelerates them to energies of 50 MeV. Later the protons enter the BOOSTER, after which their energy reaches 1.4 GeV. Afterwards, they are boosted to 25 GeV in the Proton Synchrotron (PS). Next, the protons are transferred to the Super Proton Synchrotron (SPS) and accelerated to the energy of 450 GeV. The last machine is the LHC accelerator, in which the protons fill two beam pipes and are sent in opposite directions. In the LHC the protons gain their maximum energy of 6.5 TeV, thus the total energy during collisions is 13 TeV in the centre-of-mass system. The beams collide inside four detectors: CMS [34], ATLAS [35], LHCb [36], and ALICE [37].

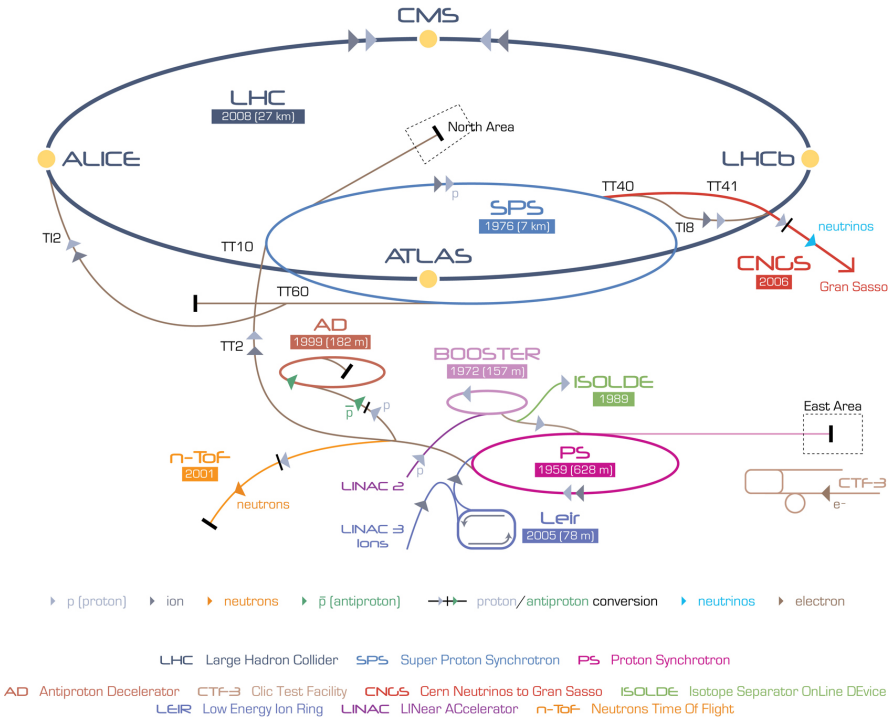


Figure 3.1: The LHC accelerator complex [38].

3.2 The Large Hadron Collider beauty

LHCb is an experiment dedicated to heavy flavour physics. It is a forward spectrometer covering a pseudorapidity region of $2 < \eta < 5$, where $\eta = -\ln(\tan\frac{\theta}{2})$ and θ is the angle between the beam and the particle's direction. It consists of two major subsystems: particle identification and tracking systems. The first subsystem consists of Cherenkov light detectors (RICH1/2), calorimeters (ECAL,HCAL,SPD/PS), and muon stations (M1-M5). The second one is composed of the Vertex Locator (VELO), the Tracker Turicensis (TT) detector, and T stations (T1-T3). A schematic view of the LHCb detector is given in Fig. 3.2. The tracking and particle identification detectors together with the system used for triggering particles are described in more detail in the following sections.

3.2.1 The tracking system

The first tracking detector from the interaction point is the VELO detector. It consists of 21 stations, each made of two halves containing two types of silicon modules with microstrip sensors. The R sensors give information about the radial distance, and the ϕ sensors give information about the azimuthal angle. During the initial stage of the beam injection, the detector modules retract to a position where they are at a distance of 6 cm from each other (the so-called VELO open position, see Fig. 3.3). When the beam

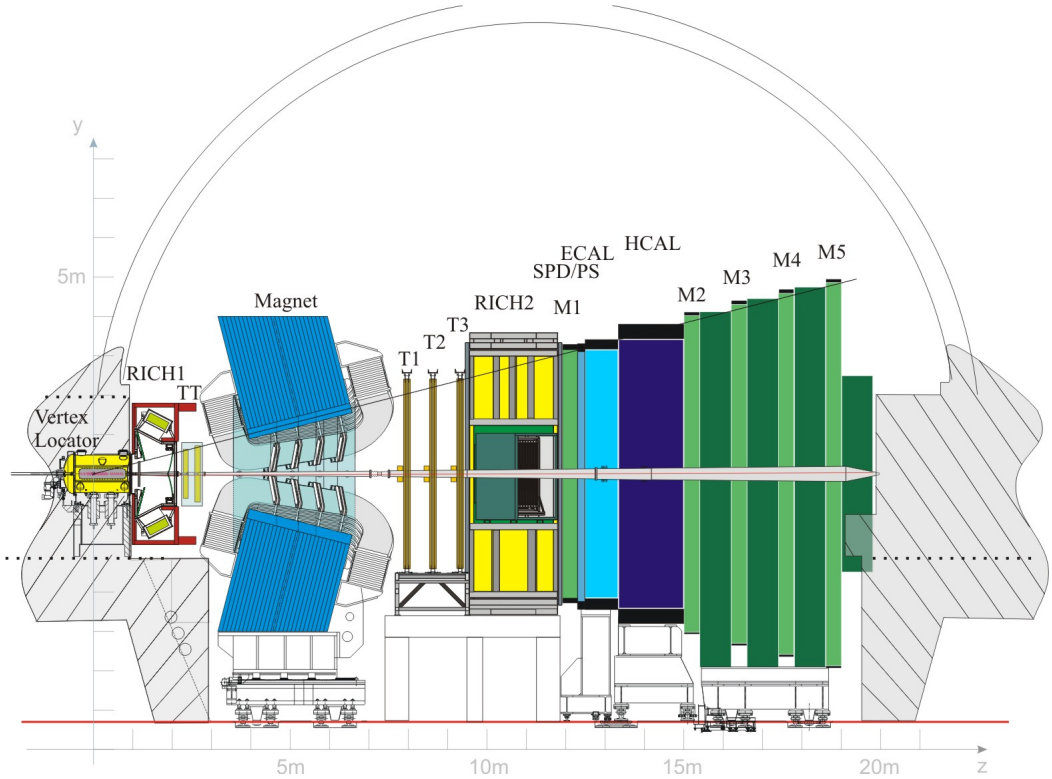


Figure 3.2: A schematic view of the LHCb experiment [39].

reaches stability, the VELO modules are closed, and as a result they are only 8 mm away from the interaction point (the so-called VELO fully closed position, see Fig. 3.3). The VELO detector is kept in a vacuum to avoid multiple scatterings, thus ensuring the best possible momentum resolution.

Next come the tracking stations (TT and T1-T3). The TT is located before the magnet, and it improves with the track momentum resolution. It consists of four layers called *xuvx*. The *x* planes are in an arrangement with vertical strips, and the planes *u* and *v* are rotated by $\pm 5^\circ$ with respect to the *x* planes. The rest of the T stations are located between the magnet and the RICH2. Since in the central part of the detector the occupancy is much higher, those stations are made of two tracker types. The inner tracker (IT) is made of similar silicon microstrip sensors and the outer tracker (OT) is made of straw tubes. Both, IT and OT detectors are made of four layers rotated in the same scheme as the TT detector.

The magnet, which is a dipole magnet with an integrated strength of 4 Tm, bends charged particles, making it possible to infer about their momentum and charge. The magnet's polarity can be switched between the so called `MagUp` and `MagDown`.

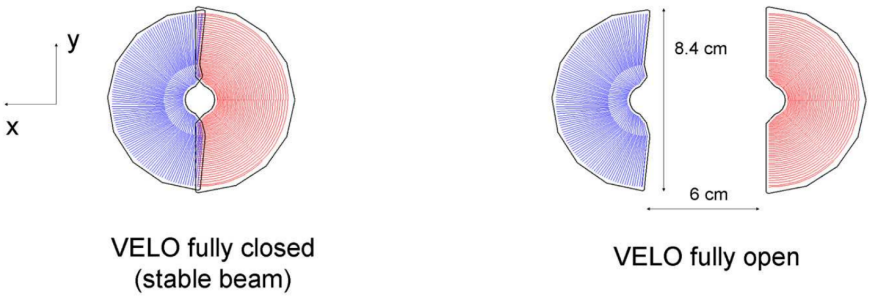


Figure 3.3: The front face of the VELO silicon modules in the fully closed and fully open position [39].

3.2.2 The particle identification system

The first two detectors in the particle identification system are the Ring Imaging Cherenkov (RICH) detectors. Charged particles transversing an interacting medium (radiator), in which their velocity is greater than the speed of light, can emit Cherenkov light. The characteristic angle in which the light is emitted by the particle depends on its velocity and the refractive index n of the medium:

$$\sin\alpha = \frac{v_f}{v} = \frac{c}{nv}. \tag{3.1}$$

The particle identification is possible by calculating its mass from the combined information of the measured Cherenkov angle and the measured momentum of the particle.

The RICH detectors, which take advantage of the above described phenomena, are shown in Fig. 3.4. The emitted photons are collected by Photon Detectors (inside RICH1) and by Hybrid Photon Detectors (HPD, inside RICH2). The possible momentum range of the measured particles depends on a radiator. The radiators used in the RICH1 are aerogel and C_4F_{10} gas, with refractive indices of 1.03 and 1.005 respectively. CF_4 gas with a refractive index of 1.0014 is used as a radiator in the RICH2. The combination of those three radiators gives the total momentum range of 5 - 100 GeV/c. Low momentum particles are distinguished in the RICH1 detector, and higher momentum particles are distinguished in the RICH2 detector.

The calorimeter system consists of four subdetectors: Pre-Shower detector (PS), Scintillating Pad Detector (SPD), Electromagnetic Calorimeter (ECAL), and Hadron Calorimeter (HCAL). Its purpose, next to the particle identification, is to measure the particles' energies. All calorimeters are segmented, thus the particle tracks can be matched with the energy deposits. The calorimeters are composed of layers of absorbers in which an interaction occurs, and active materials in which a detection occurs. Next, scintillation photons are detected in Photon Multiplier Tubes (PMTs). The PS and SPD detectors read information from the PMTs outside the LHCb acceptance, and ECAL and HCAL are located in the downstream region. ECAL is composed of layers of plastic scintillator and lead, and HCAL is composed of scintillator layers and iron.

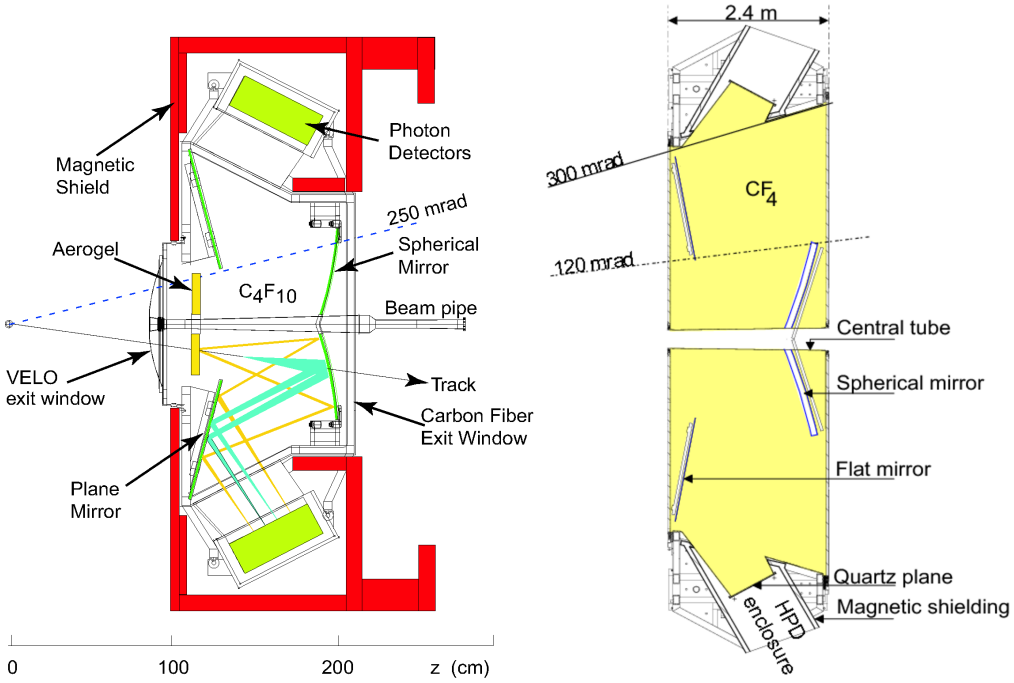


Figure 3.4: Cherenkov light detectors, RICH1 (left) and RICH2 (right) [39].

Electrons pose the biggest problem in the LHCb particle identification and reconstruction. The bremsstrahlung radiation energy is proportional to the inverse of the particle mass raised to the power of 4, thus it strongly affects electron reconstruction. Figure 3.5 shows how in LHCb the radiated photons are looked for, and how γ clusters in ECAL are matched with the electron tracks. The electron momentum is then corrected by the information from the cluster. If the bremsstrahlung photon is emitted after the magnet, it hits the same cell inside ECAL as the electron, thus the energy is properly measured. In the opposite case, the electron is bent and the photon does not reach the same cell in ECAL. In the latter case the bremsstrahlung recovery algorithm is applied, to look for neutral ECAL clusters that match the electron trajectories.

The last in the particle identification systems is the muon system. It consists of five muon stations (see M1-M5 in Fig. 3.2). The first muon station M1 lies before the calorimeters. Its purpose is to help with the transverse momentum (p_{T}) measurement for the trigger. It is made of Gas-Electron Multipliers (GEM), which have very good ageing properties, thus it can bear the high radiation level in that region. Stations M2-M5 are located after the calorimeter system. They are made of Multi-Wire Proportional Chambers (MWPCs) and lead absorbers. Each muon station is followed by an iron wall to stop other particles. As a result of this design, only muons can reach the last muon station. Thus, the number of the stations penetrated by muons as a function of the particle momentum (see Tab. 3.1) serves as the criterium for high efficiency loose binary selection (the so-called isMuon requirement). The muon efficiency distribution for muons of different momenta is shown in Fig. 3.6. The smaller efficiency for lower

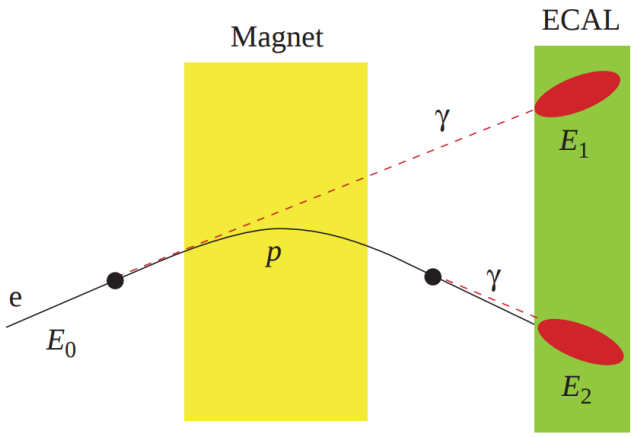


Figure 3.5: The bremsstrahlung recovery scheme [40].

Table 3.1: Muon stations required for the isMuon requirement, depending on the muon's momentum [41].

Momentum range	Muon stations
$3\,\text{GeV}/c < p < 6\,\text{GeV}/c$	M2 and M3
$6\,\text{GeV}/c < p < 10\,\text{GeV}/c$	M2, M3, and M4 or M5
$p > 10\,\text{GeV}/c$	M2, M3, M4, and M5

momentum particles is the result of the fact that some of them can fall outside the acceptance [41].

Final particle identity information is given in two types of variables. One variable is the difference in the log-likelihood of the two mass hypothesis $\text{DLL}_{\text{X}\pi}$, where X (π) denotes that a particle is identified under the X (π) mass hypothesis. The other variable gets the information from the neutral network classifier trained on Monte Carlo samples, and it is called ProbNN [42].

3.2.3 The trigger system

In order to reduce a very high data rate, a set of hardware and software stages called the trigger is used. In LHCb the hardware Level-zero (L0) trigger uses information from the calorimeters and muon stations. It requires particles to have high transverse momentum in the muon stations, and high transverse energy in the calorimeter system. As a result, the rate is decreased from the initial 40 MHz to $\sim$ 1 MHz. The L0 trigger consists of three independent sub-triggers: L0-Calorimeter, L0-Muon, and L0-PileUp. The first L0 trigger computes the transverse energy E_T deposited in the SPD, PS, ECAL, and HCAL detectors. Its value is then used to select the following types of candidates:

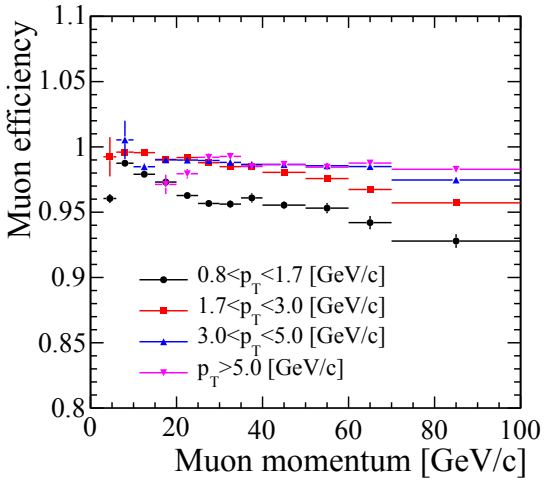


Figure 3.6: The muon selection efficiency as a function of the muon momentum [40].

- `L0Hadron` - the most energetic clusters from HCAL, includes the corresponding ECAL energy;
- `L0Photon` - the most energetic clusters from ECAL, with 1-2 hits in PS and no SPD hits in the corresponding cells;
- `L0Electron` - the most energetic clusters from ECAL, with 1-2 hits in PS and at least one SPD hit in the corresponding cells.

In addition, the events with too many SPD hits are rejected. The second L0 trigger looks for `L0Muon` candidates with a high p_T. The L0 trigger cuts relevant for this analysis are summarised in Tab. 3.2.

Table 3.2: Transverse momentum/energy L0 requirements [39].

	p_T or E_T		SPD hits
	2011	2012	2011 and 2012
single muon	1.48 GeV/c	1.76 GeV/c	600
electron	2.50 GeV	3.00 GeV	600

The next stage is the software High-level-trigger (HLT), which runs on $\sim$ 29000 logical CPU cores in the Event Filter Farm (EFF). It consists of two stages called HLT1 and HLT2, and takes the entire event information. The HLT1 uses the VELO information for the impact parameter (IP) selection, where tracks displaced from the PV are accepted. Also a p_T selection is applied, for example the minimum $p_T > 0.5 - 1.25$ GeV/c was required for non-muon tracks, where the tracks with corresponding muon

hits are triggered when their momentum is greater than a value between 3 and 6 GeV/c (depending on the run settings). The HLT1 trigger level reduces the data rate to 30 kHz, which makes it manageable for the HLT2. The HLT2 topological trigger uses more advanced methods, such as multivariate algorithms, to perform full event reconstruction. It looks for the tracks with $p > 3$ GeV/c and $p_T > 0.3$ GeV/c. The generic beauty HLT2 trigger lines look for the tracks displaced from the PV with at least two charged daughters. Within the muon triggers, single muon candidates have to pass the transverse momentum requirement $p_T > 10$ GeV/c or $p_T > 1.3$ GeV/c if they do not originate from the PV. There is no mass requirement for the tracks displaced from the PV. The muons with $p_T > 2$ GeV/c are accepted. As a result of the HLT trigger system, the data rate is reduced to ~ 5 kHz.

CHAPTER 4

Search for the $B^+ \to K^+\mu^\pm e^\mp$ decay

4.1 Analysis strategy

The discussed analysis is performed using the data collected by the LHCb experiment at the Large Hadron Collider at CERN. The data sample corresponds to an integrated luminosity of 1.0 fb^{-1} of proton-proton collisions at a centre-of-mass energy of 7 TeV and 2.0 fb^{-1} at 8 TeV. In this data we have selected three types of decays: $B^+ \to K^+ J/\psi(\to \mu^+\mu^-)$, $B^+ \to K^+ J/\psi(\to e^+e^-)$, and $B^+ \to K^+\mu^\pm e^\mp$ candidates with specific selection criteria. The first decay is used to normalise the number of observed signal candidates with the following equation:

$$\begin{aligned}\mathscr{B}(B^+ \to K^+\mu^\pm e^\mp) &= \mathscr{B}(B^+ \to K^+ J/\psi(\to \mu^+\mu^-)) \\ &\times \frac{N(B^+ \to K^+\mu^\pm e^\mp)}{\varepsilon(B^+ \to K^+\mu^\pm e^\mp)} \frac{\varepsilon(B^+ \to K^+ J/\psi(\to \mu^+\mu^-))}{N(B^+ \to K^+ J/\psi(\to \mu^+\mu^-))} \\ &= \alpha(B^+ \to K^+\mu^\pm e^\mp) \times N(B^+ \to K^+\mu^\pm e^\mp), \end{aligned} \tag{4.1}$$

where $\mathscr{B}(B^+ \to K^+ J/\psi(\to \mu^+\mu^-)) = (6.116 \pm 0.1879) \times 10^{-5}$ [7] is the branching fraction of the normalisation channel, $\varepsilon(B^+ \to K^+\mu^\pm e^\mp)$ and $\varepsilon(B^+ \to K^+ J/\psi(\to \mu^+\mu^-))$ are efficiencies of the selection for the signal channel and the normalization channel respectively. The $N(B^+ \to K^+\mu^\pm e^\mp)$ and $N(B^+ \to K^+ J/\psi(\to \mu^+\mu^-))$ are the observed numbers of the candidates, respectively. All the values, excluding the number of signal events, are combined into the so-called normalisation factor, denoted as α. The decays $B^+ \to K^+ J/\psi(\to e^+e^-)$ and $B^+ \to K^+ J/\psi(\to \mu^+\mu^-)$ are used to predict the mass shape distribution of the searched decay $B^+ \to K^+\mu^\pm e^\mp$. The selection of the $B^+ \to K^+\mu^\pm e^\mp$ is chosen to be as similar as possible to the aforementioned charmonium

normalisation decay channel. For this reason its systematic uncertainties to large extend cancel with the normalisation channel uncertainties in Eq. 4.1. Further discrimination between the signal and the background is obtained with multivariate classifiers (see Sec. 4.6) trained on MC signal samples and the data. One of the classifiers is specialised to reduce the combinatorial background while the other one - to eliminate partially reconstructed background events. The decay of interest is examined separately in two charge combinations $B^+ \to K^+\mu^-e^+$ and $B^+ \to K^+\mu^+e^-$, where the kaon and electron and kaon and muon have the same charges, respectively. This analysis is performed blindly, thus the signal region in the data is excluded from the analysis until the final selection is obtained.

4.2 Data sets and Monte Carlo samples

The event reconstruction is performed by the BRUNEL [43] package, which provides the vertices information and momenta of particles. Next, the first pre-selection is performed using `Stripping21r{0,1}p1` (for more details see Sec. 4.3) within the DAVINCI [44] package.

In addition to the gathered dataset, a number of simulated MC events are generated using the GAUSS [45] package. This process is separated into two stages: particle generation and decays simulation. The first one is performed by PYTHIA 8 [46], and the second one - by EVTGEN [47]. The MC simulation includes not only signal and control channels events, but also possible background sources (see background studies in Sec. 4.8). During the MC generation, only the particles that fall into the detectors' acceptance are saved by requiring a pseudorapidity to be within $0.05 < \eta < 0.4$. This is later referred to as the acceptance requirement. The prepared MC samples are processed with the same reconstruction and stripping requirements (see Sec. 4.3) as the collected data. The 2011 and 2012 samples are listed in Tab. 4.1.

The models used in the MC generation of the aforementioned samples are [47]:

- BTOSLLBALL - a model for a scalar particle (B-meson) decaying into three particles, where two of them are leptons;
- FLATSQDALITZ - a model for a particle decaying with flat squared Dalitz distributions;
- PHSP - a model of decays to n-bodies, with the matrix element equal to unity;
- PHOTOS ISGW2 - an updated Isgur-Scora-Grinstein-Wise model for the semileptonic meson decays, with final state radiation using the `PHOTOS` package [48];
- PHOTOS VLL - a model for a vector meson decaying to a pair of charged leptons, with final state radiation using the `PHOTOS` package;
- SVS - a model for a scalar particle decaying to a vector meson and a scalar particle.

Table 4.1: List of Monte Carlo samples simulated for 2011 and 2012 data-taking conditions.

Decay channel	Number of events 2011	2012	Model
$B^+ \to K^+ e^+ \mu^-$	508068	1 016 442	PHSP
$B^+ \to K^+ J/\psi(\to \mu^+\mu^-)$	2 018 494	4 020 672	SVS, PHOTOS VLL
$B^+ \to K^+ J/\psi(\to e^+ e^-)$	1 002 688	2 056 775	SVS, PHOTOS VLL
$B^+ \to K^+ \mu^+ \mu^-$	527 247	2 022 527	BTOSLLBALL
$B^+ \to K^+ e^+ e^-$	-	7 992 163	PHSP
$\Lambda_b^0 \to p K^- J/\psi(\to e^+ e^-)$	319 557	607 045	PHSP, PHOTOS VLL
$\Lambda_b^0 \to p K^- \mu^+ \mu^-$	4 021 922	2 019 539	PHSP, PHOTOS VLL
$\Lambda_b^0 \to p K^- J/\psi(\to \mu^+\mu^-)$	284 804	632 210	PHSP, PHOTOS VLL
$\Lambda_b^0 \to p K^- e^+ e^-$	-	1 054 532	PHSP
$B^+ \to \overline{D}^0(\to K^+ e^- \overline{\nu}_e) e^+ \nu_e$	-	2 011 907	PHOTOS ISGW2
$B^- \to D^0(\to K^- \mu^+ \nu_\mu) \overline{\nu}_\mu \mu^-$	-	12 400 491	PHOTOS ISGW2
$B^+ \to \overline{D}^0(\to K^+ \pi^-) e^+ \nu_e$	-	2 013 106	PHOTOS ISGW2, PHSP
$B^+ \to K^+ \pi^+ \pi^-$	514 500	1 020 995	FLATSQDALITZ
$B^+ \to K^- \pi^+ \pi^+$	516 000	1 022 997	PHSP

The desired model for the non-SM decays is the PHSP model, while the prepared MC signal samples were generated with the BTOSLLBALL model, so the BTOSLLBALL-PHSP weights are calculated.

4.3 Stripping selection

Loose selection requirements based on the topologies and kinematics of the aforementioned decays are collected in the so-called stripping lines. The stripping selection reduces the event rate and makes the data analysis faster. Each member of the LHCb experiment can compose his/her own set of loose selection criteria for a chosen decay in a specific stripping line. The gathered data, after being processed with stripping lines, is later stored in and is available to all users. Three different stripping lines are chosen to collect signal and control channel candidates. `StrippingBu2LLK_meLine` is used for the signal channel $B^+ \to K^+ \mu^\pm e^\mp$, and `StrippingBu2LLK_mmLine` and `StrippingBu2LLK_eeLine2` are used for $B^+ \to K^+ J/\psi(\to \mu^+\mu^-)$ and $B^+ \to K^+ J/\psi(\to e^+ e^-)$ control channels, respectively. The same stripping selection is used on the corresponding MC samples. All background samples are processed with the same stripping line as the signal channel.

The requirements present in the chosen stripping lines are summarised in Tab. 4.2. They are based on the following variables:

- the transverse momentum p_{T} of a particle;
- the invariant mass $m(\ell\ell)$ of a dilepton system;
- the difference of the reconstructed mother particle's mass (m) from its PDG [7] value (m_{PDG}): $|m - m_{\mathrm{PDG}}|$;
- the cosine of the `DIRA` angle, which is the angle between the mother particle's momentum direction and the flight direction from its primary vertex (PV) to the secondary vertex (SV);
- the `isMuon` boolean value, which gives the information if the particle's track fulfills the muon system hit requirements (see Sec. 3.2.2);
- the `hasMuon` boolean value, which gives the information if the track has hits in the muon system;
- the difference $\mathrm{DLL}_{X\pi}$ in log-likelihood of the two mass hypothesis of a particle X and a pion;
- the χ^2 of the fit to the given vertex - χ^2_{vtx} [49];
- the χ^2 of the fit to the given flight distance - χ^2_{FD} [49];
- the impact parameter's χ^2_{IP}, defined as the difference between χ^2 of the fit to the PV, including all the particles present in the decay, and the χ^2 of the fit to the SV;
- the number of hits (n_{SPD}) in the SPD detector.

The particle identification (PID) variables, such as $DLL_{X\pi}$, are highly efficient in discriminating signal events from background sources. Because of significant data - MC differences in PID variables, a data driven calibration is used in further steps of the analysis (for more details see Sec. 4.5.3) to correct for these discrepancies.

4.4 Preseletion and trigger requirements

Those first loose requirements on the signal channel are introduced to reduce the combinatorial background and to get rid of some background sources, such as decays occurring through charmonium resonances or charm decays. The decays originating from partially reconstructed D^0 decays are effectively vetoed (as can be seen in Fig. 4.1) with the requirement on the invariant mass of the kaon and lepton pair $m_{K^\pm\ell^\mp} < 1885\,\mathrm{MeV}/c^2$. J/ψ and $\psi(2S)$ resonances efficiently rejected by the following requirements on mass - swap hypotheses:

Table 4.2: Summary of stripping requirements from the `Bu2LLK` lines.

Particle or event	Variable	Requirement
Event	n_{SPD}	< 600
B	$\lvert m - m_{\text{PDG}}\rvert$	$< 1500\,\text{MeV}/c^2$
	χ^2_{vtx}	< 9
	χ^2_{FD}	> 100
	χ^2_{IP}	< 25
	DIRA	> 0.9995
K	p_T	$> 400\,\text{MeV}/c$
	χ^2_{IP}	> 9
e	p_T	$> 300\,\text{MeV}/c$
	χ^2_{IP}	> 9
μ	p_T	$> 300\,\text{MeV}/c$
	χ^2_{IP}	> 9
	isMuon	True
	hasMuon	True
e/μ	$m(e\mu)$	$> 100\,\text{MeV}/c^2$
	$\chi^2_{\text{vtx}}(e\mu)$	< 9
Dimuon	p_T	$> 0\,\text{MeV}/c$
	m	$< 5500\,\text{MeV}/c^2$
	χ^2_{vtx}	< 9
	χ^2_{FD}	> 16
	χ^2_{IP}	> 0
Dielectron	p_T	$> 0\,\text{MeV}/c$
	m	$< 5500\,\text{MeV}/c^2$
	χ^2_{vtx}	< 9
	χ^2_{FD}	> 16
	χ^2_{IP}	> 0

- kaon with muon mass: 3000 (2950) MeV/c^2 $< m_{K^\mp\mu^\pm(K^\pm e^\mp)} <$ 3200 MeV/c^2 and 3630 MeV/c^2 $< m_{K^\mp\mu^\pm} <$ 3740 MeV/c^2;
- kaon with electron mass: 3000 MeV/c^2 $< m_{K^\pm e^\mp} <$ 3200 MeV/c^2 and 3630 MeV/c^2 $< m_{K^\pm e^\mp} <$ 3740 MeV/c^2;
- electron with muon mass: 2950 MeV/c^2 $< m_{e^\mp\mu^\pm} <$ 3200 MeV/c^2 and 3630 MeV/c^2 $< m_{e^\mp\mu^\pm} <$ 3740 MeV/c^2;
- muon with electron mass: 3000 MeV/c^2 $< m_{\mu^\pm e^\mp} <$ 3200 MeV/c^2 and 3630 MeV/c^2 $< m_{\mu^\pm e^\mp} <$ 3740 MeV/c^2.

As can be seen in Fig. 4.2, there is some contribution left from the $B^+ \to K^+ J/\psi(\to e^+e^-)$ decays, but further selection removes this background completely (see Sec. 4.8). The contribution from the $B^+ \to K^+ J/\psi(\to \mu^+\mu^-)$ is reduced with the aforementioned charmonium vetoes.

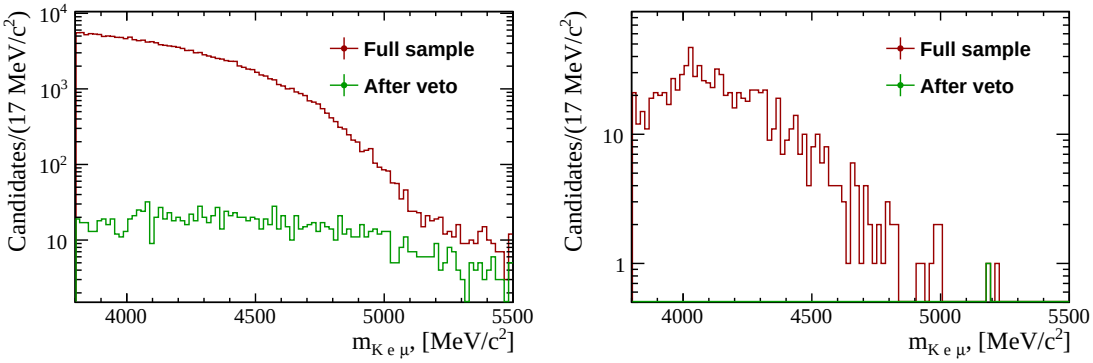


Figure 4.1: The D^0 veto on $B^- \to D^0(\to K^-\mu^+\nu_\mu)\mu^-\overline{\nu}_\mu$ (left) and $B^- \to D^0(\to K^-e^+\nu_e)e^-\overline{\nu}_e$ MC samples. The events before the veto are marked in red colour, the events after the veto are marked in green. The y-axis scale is logarithmic.

The backgrounds originating from decays with misidentified particles are easily reduced with requirements on PID variables:

- the neural-net based probability `ProbNN` for the muon to be a muon and the kaon to be a kaon has to be greater than 0.2;
- the difference in log-likelihoods $\texttt{DLL}_{e\pi}$ for the electron has to be greater than 3.0.

In addition to the aforementioned requirements, all daughter tracks are required to have hits in the RICH detector subsystem (the so-called `hasRich` requirement). The electrons are required to have hits in the calorimeter system (the so-called `hasCalo` requirement) but not to fulfil the `isMuon` requirement. The kaons are required to reach the muon chambers geometrical acceptance (the `InAccMuon` requirement) and not to fulfil loose muon requirements (the `isMuonLoose` accepts the muon candidates

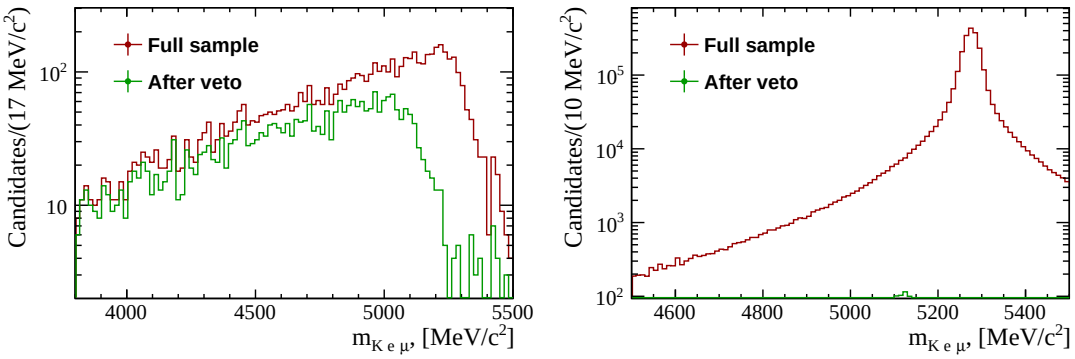


Figure 4.2: All charmonium vetoes applied to $B^+ \to K^+ J/\psi\,(\to e^+e^-)$ (left) and $B^+ \to K^+ J/\psi\,(\to \mu^+\mu^-)$ (right) MC samples. The events before the veto are marked in red colour, the events after the veto are marked in green. The y-axis scale is logarithmic.

with the momentum between 3 and 6 GeV/c that have hits in at least M1-M4 muon stations and the candidates with the momentum greater than 6 GeV/c that have hits in all muon stations). The detector's requirements are summarised in Tab. 4.3. The

Table 4.3: Preselection requirements applied to the signal and control channels.

Particle	Cut
K	ProbNNk > 0.2
	InAccMuon $= 1$
	isMuonLoose $= 0$
	hasRich $= 1$
μ	ProbNNmu > 0.2
	hasRich $= 1$
e	PIDe > 3.0
	hasRich $= 1$
	hasCalo $= 1$
	isMuon $= 0$

preselection of the control channels is chosen to be as similar as possible to the signal channel's. The same D^0 veto, PID requirements, and subdetectors requirements are applied. To further reduce the pollution from the combinatorial background, a set of trigger requirements is applied. In the case of the signal channel, the muons have to pass the `L0Muon` trigger. The electrons and muons from the control channels are required to pass the `L0Electron` and `L0Muon` triggers, respectively. From the software HLT1 trigger requirements, the searched decay and the muon control channel are required to fulfil the `TrackMuonDecision` requirement, and also all decay channels have to pass the `TrackAllL0Decision` trigger. From the software HLT2 trigger requirements,

all decay channels are required to fulfil the `Topo[2,3]BodyBBDTDecision`, and the decays involving the muons have to pass the `SingleMuon(LowPT)Decision` and `TopoMu[2,3]BodyBBDTDecision` triggers [50].

4.5 Data - Monte Carlo agreement

The best possible agreement between the data and MC distributions is needed for the analysis to reduce possible systematics and biases. Unfortunately, the simulated MC samples do not perfectly reflect all the distributions, so they have to be corrected for the known differences. Kinematic distributions were corrected by the data-MC re-weighting. PID distributions were corrected using the PID re-sampling procedure. The clear data sample was extracted using the *sPlot* technique [51].

4.5.1 The *sPlot* technique and control channel fits

The *sPlot* is a statistical technique used to unfold distributions from mixed samples. With this method, the searched distribution of the so-called control variable (x) can be reconstructed using the known probability density function (PDF) distribution of the so-called discriminating variables (y). Both variable types are assumed to be uncorrelated, thus the PDF of the whole sample factorizes into:

$$f_i(x,y) = M_i(x) f_i(y), \tag{4.2}$$

where i stands for the i'th component of the sample, $M_i(x)$ is its searched distribution, and $f_i(y)$ is the known PDF distribution. Thus the first step of the method is the maximum Likelihood fit to obtain $f_i(y)$. In this case, the data samples consist of two components: the combinatorial background distribution and the control channel distribution. The $B^+ \to K^+ J/\psi\,(\to \mu^+\mu^-)$ and $B^+ \to K^+ J/\psi\,(\to e^+e^-)$ B candidate's mass shapes are well described by the double Crystal Ball (CB) function [52]. The Crystal Ball function given by:

$$f_{CB}(m;\mu,\sigma,\alpha,n) = N_{CB} \times \begin{cases} e^{-\frac{1}{2}\left(\frac{m-\mu}{\sigma}\right)^2} & \frac{m-\mu}{\sigma} > \alpha \\ \left(\frac{n}{|\alpha|}\right)^n e^{-\frac{\alpha^2}{2}} \times \left(\frac{n}{|\alpha|} - |\alpha| - \frac{(m-\mu)}{\sigma}\right)^{-n} & \frac{m-\mu}{\sigma} < \alpha \end{cases}, \tag{4.3}$$

i.e. it consists of a Gaussian core (above $\alpha\sigma$) and a power-law low-end tail (below $\alpha\sigma$), where m is the invariant mass of the $K\ell\ell$ system, N_{CB} is the normalisation factor, μ is the average value of the B mass, and n and α are the tail's parameters. The combinatorial background is fitted with an exponential function. The total PDF is given with:

$$\begin{aligned} PDF_{total} =& N \times \\ & (n_{sig}[c\; f_{CB}(m;\mu,\sigma_1,\alpha,n) + (1-c) f_{CB}(m;\mu,\sigma_2,\alpha,n)] + n_{bkg} e^{\lambda m}), \end{aligned} \tag{4.4}$$

where n_{sig} is the number of fitted signal events, n_{bkg} is the number of combinatorial background events, c is the fraction of the first Crystal Ball function, and λ is the exponent's tail parameter.

The fit to the 2011 and 2012 $B^+ \to K^+ J/\psi\,(\to \mu^+\mu^-)$ data samples is shown in Fig. 4.3, and the corresponding fit parameters are given in Tab. 4.4. The fitting mass range is restricted to $5180\,\mathrm{MeV}/c^2 < m_{K\mu\mu} < 5700\,\mathrm{MeV}/c^2$.

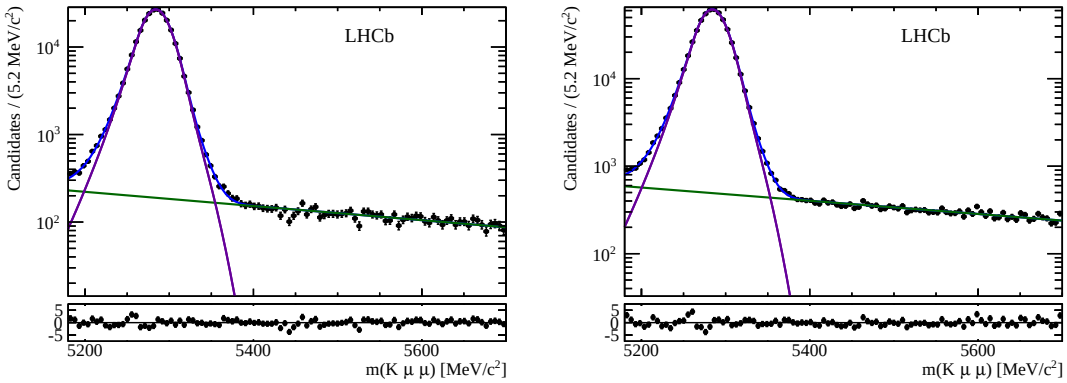


Figure 4.3: Fit to the 2011 (left) and 2012 (right) data sample. The purple line is for the signal component, the green line - for the combinatorial background, and the blue line - for the total PDF. The y-axis scale is logarithmic.

Table 4.4: Results of the $m_{K\mu\mu}$ fit for the 2011 and 2012 $B^+ \to K^+ J/\psi\,(\to \mu^+\mu^-)$ data sample.

Fit parameter	2011	2012
$\mu\,[\mathrm{MeV}/c^2]$	5283.9 ± 0.05	5284.10 ± 0.03
$\sigma\,[\mathrm{MeV}/c^2]$	16.27 ± 0.11	16.11 ± 0.08
$\sigma_1\,[\mathrm{MeV}/c^2]$	27.96 ± 0.49	27.10 ± 0.28
c	0.77 ± 0.014	0.732 ± 0.010
n	26.85 ± 7.97	17.88 ± 28.92
α	1.41274 ± 0.02027	1.45064 ± 0.01531
$\lambda\,[\mathrm{MeV}/c^2]$	-0.001865 ± 0.000107	-0.001739 ± 0.000656
n_{sig}	239200 ± 584	552154 ± 816.44
n_{bkg}	14715.3 ± 247.026	38835.1 ± 391.61

The mass range of the m_{ee} is chosen to be $^{+200}_{-500}$ MeV/c^2 around the known J/ψ mass. The fit parameters and the corresponding plots are given below. The fitting mass range is restricted to $5000\,\mathrm{MeV}/c^2 < m_{K\mu\mu} < 5700\,\mathrm{MeV}/c^2$.

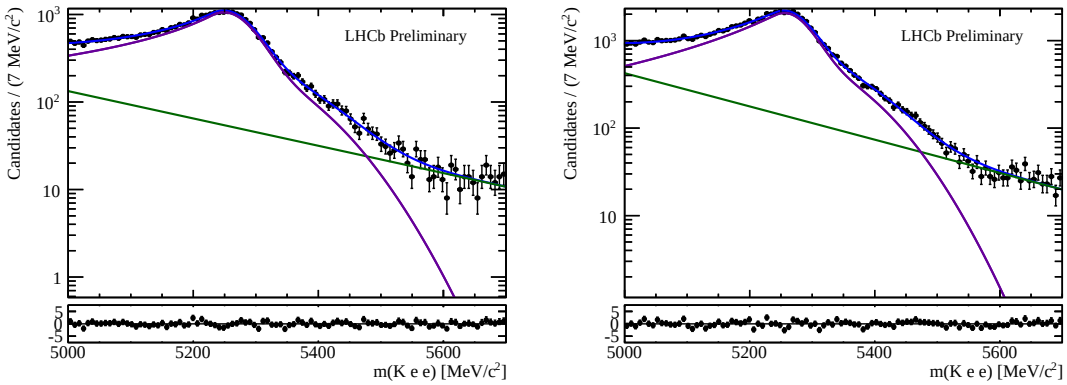


Figure 4.4: Fit to the 2011 (left) and 2012 (right) data sample to the invariant mass of the B candidate from the $B^+ \to K^+ J/\psi\,(\to e^+e^-)$ decay. The purple line is for the signal, the green line - for the combinatorial background, and the blue line - for the total PDF. The y-axis scale is logarithmic.

Table 4.5: Results of the m_{Kee} fit for the 2011 and 2012 $B^+ \to K^+ J/\psi\,(\to e^+e^-)$ data candidates.

Fit parameter	2011	2012
$\mu\,[\mathrm{MeV}/c^2]$	5248.99 ± 1.58	5255.43 ± 1.14
$\sigma\,[\mathrm{MeV}/c^2]$	106.29 ± 3.97	99.91 ± 2.29
$\sigma_1\,[\mathrm{MeV}/c^2]$	42.05 ± 1.67	36.75 ± 1.25
c	0.340 ± 0.023	0.431 ± 0.017
n	1 ± 0.13	2.68 ± 1.029
α	0.43601 ± 0.03912	0.37111 ± 0.02667
$\lambda\,[\mathrm{MeV}/c^2]$	-0.003595 ± 0.000358	-0.004374 ± 0.00129
n_{sig}	31451 ± 736.4	58660.2 ± 1169.99
n_{bkg}	4863.35 ± 718.13	13215.4 ± 1150.7

With control channel *PDF* distributions, the so-called *sWeights* are calculated [51], which are used to clear the data sample from the combinatorial background.

4.5.2 Re-weighting of kinematic variables

At this stage the MC samples are corrected for the observed data - MC kinematic differences. For this purpose a control channel with a high purity is chosen. The sample is extracted from the $B^+ \to K^+ J/\psi\,(\to \mu^+\mu^-)$ candidates with *sWeights* obtained from the *sPlot* technique, as described above. Subsequently, the distributions of n_{Tracks}, p_{T} of B^+ candidates and its χ^2_{vtx} are compared with the corresponding distributions in the $B^+ \to K^+ J/\psi\,(\to \mu^+\mu^-)$ MC sample. Both the data and MC samples are processed

with the pre-selection and trigger requirements detailed in Sec. 4.4. Additionally, a requirement around J/ψ mass (± 60 MeV/c^2) is applied to the dimuon invariant mass. The data-MC weights are obtained by dividing the normalised data distributions by the normalised MC distributions. Since the distributions are not significantly correlated, the sequential re-weighting is applied. The data-MC weights are applied one after another, where every next sequence uses the weight derived from the previous step.

The calculated weights are applied to the $B^+ \to K^+ J/\psi\,(\to \mu^+\mu^-)$, $B^+ \to K^+ J/\psi\,(\to e^+e^-)$, $B^+ \to K^+\mu^+e^-$, and exclusive background MC samples. The comparison of the $B^+ \to K^+ J/\psi\,(\to \mu^+\mu^-)$ simulated events with the data before and after the re-weighting procedure for the detector track occupancy n_{Tracks}, B meson's transverse momentum p_T and its vertex χ^2 is given in Fig. 4.5. The remaining distributions used for the training of multivariate classifier can be seen in Appendix A. After the procedure, there are still some small data - MC discrepancies. They are accounted for in the systematic uncertainty calculation.

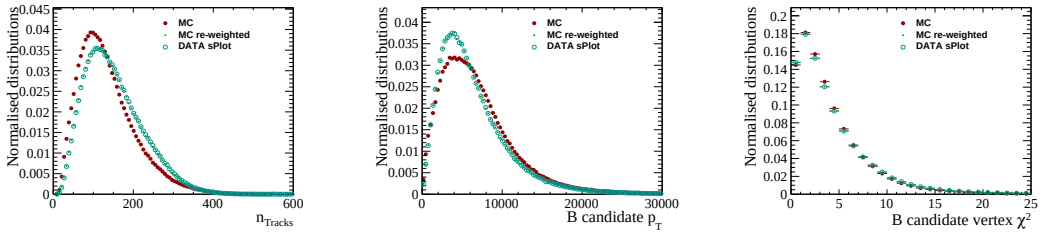


Figure 4.5: Data - MC differences in the $B^+ \to K^+ J/\psi\,(\to \mu^+\mu^-)$ decay channel before (MC) and after the re-weighting procedure (MC re-weighted). The distributions are normalised and compared with the clear signal sample taken from the data (DATA sPlot).

4.5.3 PID calibration

A simulation of PID variables is challenging because they depend not only on the kinematics but also on the occupancy of the PID system detectors (RICH, muon system, and calorimeters) and on experimental conditions such as temperature, alignment, or gas pressure. Thus, their values differ significantly between the data and simulation. In this step of the analysis, the `PidCalib` package [53] is used to re-sample PID variables. The PID resampling approach uses *sWeights* and high purity datasets of $D^{*+} \to D^0(\to K^-\pi^+)\pi^+$, $J/\psi \to e^+e^-$, and $J/\psi \to \mu^+\mu^-$ decays, provided by the PID working group. Moreover, the approach requires that no previous PID requirements have been applied. The calibration sample is split in bins of p, η, and n_{Tracks}. The population of each bin is then used as a PDF function to randomly pick a correct PID value for the corresponding bin in the simulated sample for each particle. The effect of the resampling procedure can be easily seen in the $B^+ \to K^+ J/\psi\,(\to \mu^+\mu^-)$ and $B^+ \to K^+ J/\psi\,(\to e^+e^-)$ control channels. Fig. 4.6 and 4.7 show the PID distributions from the MC samples before and after resampling, together with *sWeighted* data distributions for

$B^+ \to K^+ J/\psi\,(\to \mu^+\mu^-)$ and $B^+ \to K^+ J/\psi\,(\to e^+e^-)$, respectively. The distributions are zoomed in the region of interest, and the figures in the full range can be seen in Appendix B.

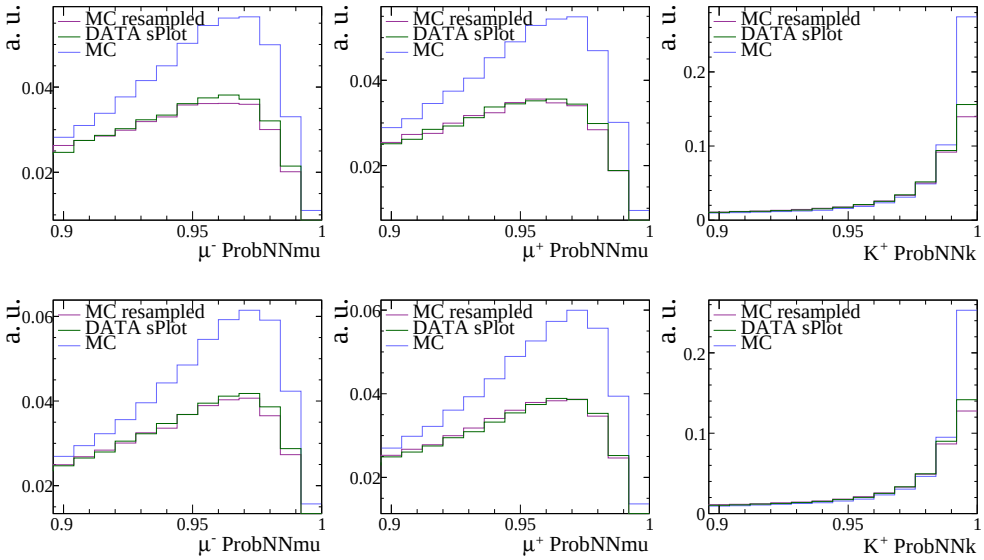


Figure 4.6: A comparison of the PID distributions before and after resampling procedure with their corresponding *sWeighted* data distributions for 2011 (top) and 2012 (bottom) $B^+ \to K^+ J/\psi\,(\to \mu^+\mu^-)$ samples.

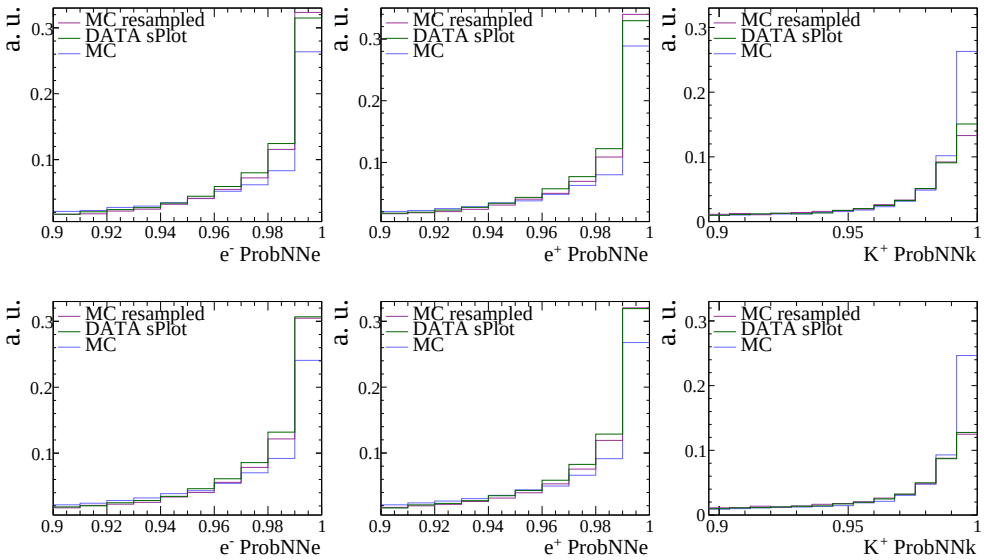


Figure 4.7: A comparison of the PID distributions before and after resampling procedure with their corresponding *sWeighted* data distributions for 2011 (top) and 2012 (bottom) $B^+ \to K^+ J/\psi\,(\to e^+e^-)$ samples.

4.6 Multivariate strategy

The purpose of Machine Learning [54] methods is to predict the behaviour of the examined data sample. The learning process uses multiple (N) input variables and considers their correlations. One event corresponds to a point in the N-dimensional space, which can be more signal- or background-like. Using signal and background samples, each event is assigned an output variable (the classifier response). At this stage, two multivariate (MVA) classifiers are trained to further suppress the background present in the data.

4.6.1 Training samples

Simulated $B^+ \to K^+\mu^\pm e^\mp$ signal events are used as a signal proxy and data sidebands are used as background proxies in the multivariate training. Two types of classifiers are trained: one classifier is trained with the upper sideband data ($5385\,\mathrm{MeV}/c^2 \leq m_{Ke\mu} \leq 6000\,\mathrm{MeV}/c^2$) to discriminate against the combinatorial background. The other classifier is trained to discriminate against the partially reconstructed background, thus the lower sideband data ($4550\,\mathrm{MeV}/c^2 \leq m_{Ke\mu} \leq 4985\,\mathrm{MeV}/c^2$) is used for the training. Both sidebands can be seen in Fig. 4.8. All used samples share the same stripping, pre-selection, and trigger requirements. The MC samples are also corrected for their data - MC differences (as described in the previous section).

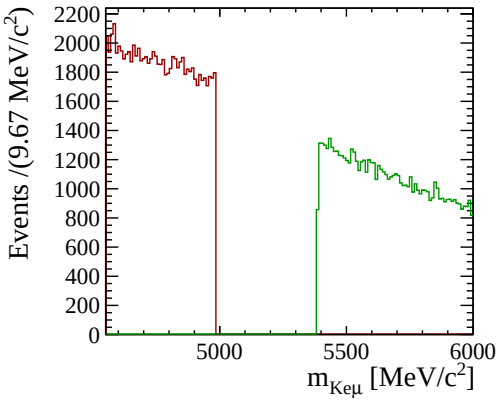


Figure 4.8: The blinded $K\mu e$ invariant mass distribution with stripping, pre-selection, and trigger requirements applied. The lower sideband data is marked with the red line, and the upper sideband data is marked with the green line.

4.6.2 k-Folding technique

For the most efficient handling of the data, the k-Folding technique [55] is used. According to the technique, the data sample is mixed, randomised, and split into k equal subsets. Next, k multivariate classifiers are trained, each one omitting i'th subset ($i \in [1,k]$). The

classifiers are applied to the corresponding subsets omitted during their training. The procedure is schematically shown in Fig. 4.9. With this approach, the training sample is increased when compared to the standard MVA training method, where the sample is split 50:50 (one half is used for the training and testing phase, and then the trained classifier is applied to the other half). Moreover the classifiers have $\frac{8}{9}$ common events, thus their responses are very similar to each other and so the related systematic error is decreased.

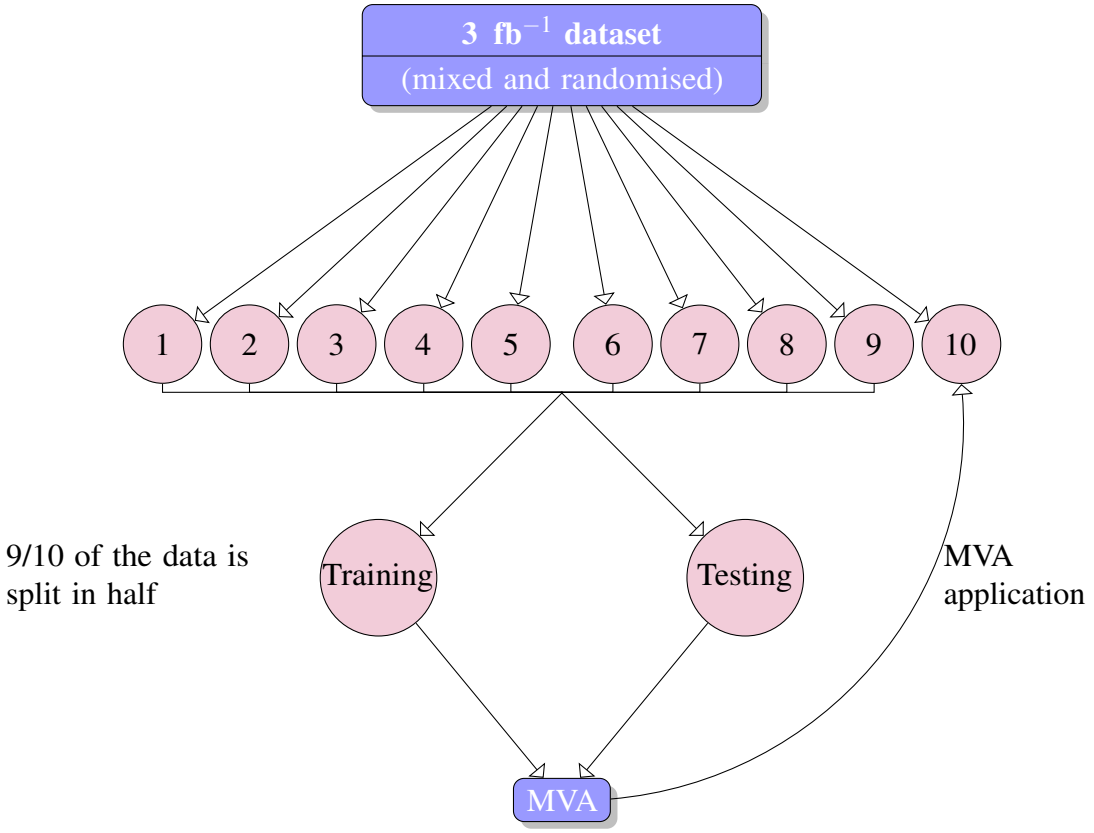


Figure 4.9: A schematic diagram of the k-Folding technique, with $k = 10$.

The procedure is performed on the full dataset (2011 and 2012 real data samples for both magnet polarities) and on the corresponding corrected MC signal samples. The samples are split into ten equal subsets, thus ten classifiers are trained.

4.6.3 The Boosted Decision Tree method

The classifier's type used is a boosted decision tree (BDT) [56] from the *Toolkit for Multivariate Analysis* (TMVA) software [57]. This type of a classifier shows the best proportion of the signal with the highest background rejection. A decision tree is a set of binary questions for which the most discriminating variable is used to separate the

sample. The sequence of questions continues until the final node is reached, as can be seen in Fig. 4.10. Depending on the majority of events in each set, the sub-sample is marked as background- or signal- like.

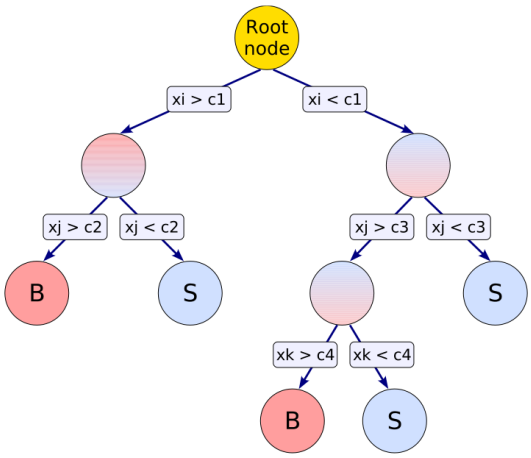


Figure 4.10: An example of a decision tree [57]. Starting from the root node, the sample's phase space is separated into signal (S) and background (B) regions.

A boosted classifier is constructed from a forest of decision trees. The boosting procedure strengthens the classifier's response against fluctuations and enhances its response with respect to a single tree. Different boosting algorithms are available within the TMVA software. In this analysis the *Gradient Boost* algorithm is used, in which a binomial log-likelihood loss function $L(F(x);y) = ln(1+e^{-2F(x)y})$ is used for the classification, where $F(x)$ is the model's response, and y is the classifier obtained from the training. The function is minimised by calculating the current gradient in each iteration.

4.6.4 Discrimination against the combinatorial background

A set of the most discriminating variables against the combinatorial background is chosen:

- the transverse momentum p_{T} of the B candidate,
- the momentum p of the B candidate,
- the impact parameter χ^2 (χ^2_{IP}), of the B candidate,
- the direction angle (DIRA) of the B candidate,
- the quality of the $Ke\mu$ vertex χ^2,
- the B flight distance χ^2,

- the impact parameter χ^2 (χ^2_{IP}) of the kaon,
- the minimum and maximum of electron and muon IP candidates,
- isolation variables from $B_s \to \mu\mu$ analysis [58].

There are two types of isolations: track and cone. The B meson candidate's track isolations (two isolations, each corresponding to one of the final state leptons) are defined as $I = |\vec{p}_T^{\,\mu\mu}|/(\sum_i p_T^i + |\vec{p}_T^{\,\mu\mu}|)$, where i stands for tracks with $\sqrt{\Delta\eta^2 + \Delta\phi^2} \leq 1$, where $\Delta\eta$ and $\Delta\phi$ are the pseudorapidity and the azimuthal angle of the track with respect to $|\vec{p}_T^{\,\mu\mu}|$. This isolation variable gives the information if any of the two leptons can create a good vertex with another track in the event. The cone isolation variables give information on how well the B candidate's decay vertex is separated from the other tracks in the decay.

The chosen topological, kinematic, and isolation variables show a good agreement between the data and the simulation (see Appendix A). Fig. 4.11 shows two examples of discriminating variables: it is clearly seen that the upper data sideband distributions differ significantly from the MC signal distributions. Hence, the choice of those variables for the efficient separation between the signal and background events. The comparisons of all the chosen variables can be seen in Appendix C.

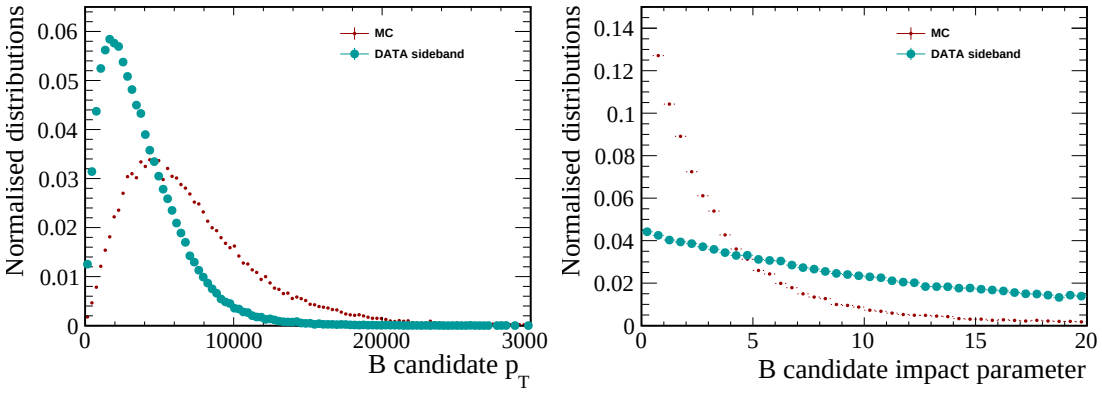


Figure 4.11: Comparison of BDTG training variables for data-taking conditions in 2011 and 2012. MC simulated signal distributions are illustrated as red dots whereas background distributions are shown in blue.

4.6.5 Discrimination against the partially reconstructed background

A classifier trained to discriminate from the partially reconstructed background uses the same variables as the ones mentioned above, as well as the so-called HOP variable. The HOP can be explained with a B-meson decay of the $B \to Y_h X_e$ type shown schematically in Fig. 4.12. The Y_h stands for hadrons of the final state and the X_e stands for electrons in the final state. With this kinematic description, the sum of transverse momentums of

the decay should be equal to zero, where P_t is the transverse momentum with respect to the direction of flight.

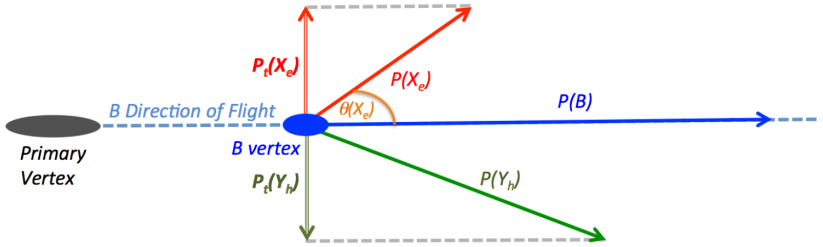


Figure 4.12: Kinematic description of a B meson decay involving electrons (e) and hadrons (h).

The ratio HOP is defined as:

$$\mathrm{HOP} = \frac{\mathrm{P_t(Y_h)}}{\mathrm{P_t(X_e)}}, \tag{4.5}$$

and in the ideal case it should be equal to 1. In experimental conditions, however, it is not perfectly equal to unity because of the inefficiency of the electron's bremsstrahlung recovery. However, in the case of the partially reconstructed decays, where a part of the $P_t(Y_h)$ is missing by at least one missing particle, the ratio HOP should significantly differ from unity. Thus, the HOP information has a discriminating power against the partially reconstructed background. The comparison of the HOP distribution with the simulated signal events with the lower sideband data can be seen in Fig. 4.13. The simulated signal and lower sideband data distributions of all variables used for the BDTGHOP training can be seen in Appendix D.

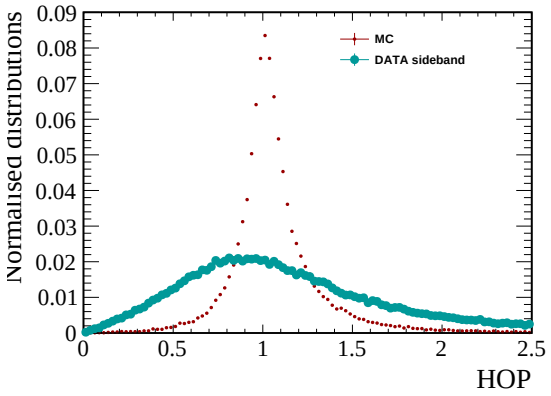


Figure 4.13: The HOP distribution for data-taking conditions in 2011 and 2012. MC simulated signal distributions are illustrated as red dots whereas background distributions are shown in blue.

4.6.6 Classifiers performance

The performance of the trained classifiers has been shown with the Receiver Operating Characteristics (ROC) curves. A ROC curve visualises the dependence of a background rejection from a signal efficiency for different possible cutpoints on the MVA requirement. The highest background rejection with the best possible signal efficiency is desired. As can be seen in Fig. 4.14, the performance of the classifier trained against the combinatorial background (BDT) is much better than that of the classifier trained against the partially reconstructed background (BDTHOP). The reason for that is a bigger data sample used for the BDT training which is significant for the performance, and the complicated structure of this type of background.

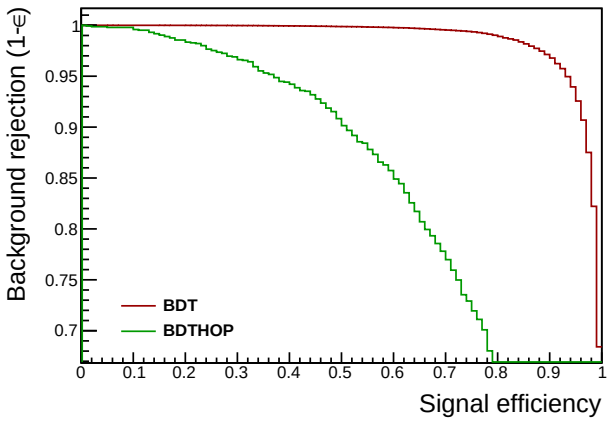


Figure 4.14: ROC curves for BDT and BDTHOP for fold $k = 0$. The other folds show similar features.

An important feature to avoid is a classifier's over-training. When this happens, the classifier's performance is high only on the data used for training and does not generalise to other samples. A test has been performed where the training and testing samples are superimposed and their classifier's responses are compared. Since no significant difference can be seen between those two distributions (see Fig. 4.15), the classifiers are considered not to be over-trained.

A possible mass correlation would be an undesired feature, because the number of expected background events is estimated by looking at the invariant $Ke\mu$ mass distribution. The trained MVAs are proven not to be correlated with $m_{Ke\mu}$ (see Fig. 4.16). The distributions are examined on the upper sideband data.

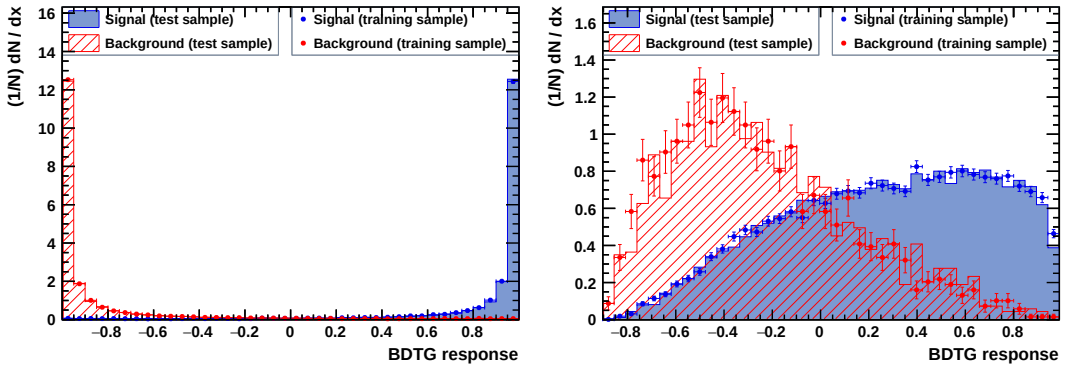


Figure 4.15: Overtraining test for BDT (left) and BDTHOP (right) classifiers for fold $k = 0$.

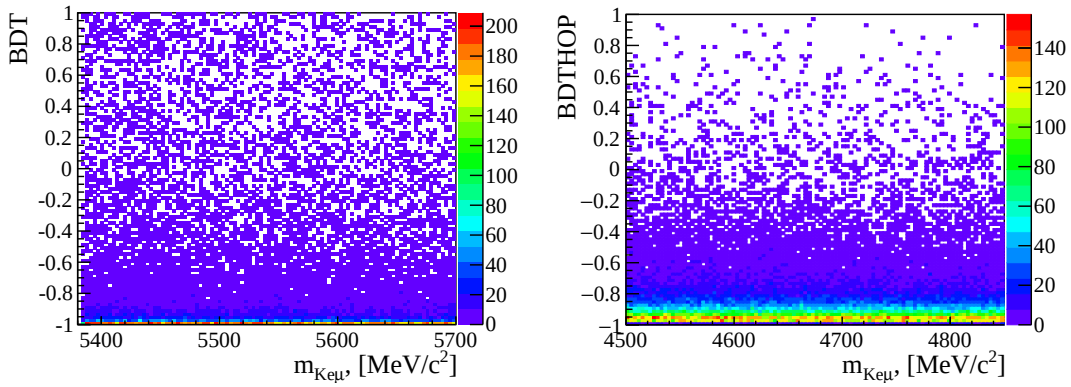


Figure 4.16: Correlation of BDT (left) and BDTHOP (right) with respect to the $Ke\mu$ invariant mass.

As mentioned in Sec 4.6.2, 10 classifiers are trained (for each classifier type), each omitting one fold. The trained classifiers are in a very good agreement among each other (see Fig. 4.17).

4.7 Optimisation

To get the highest possible number of signal events with the lowest possible background contamination, the optimisation of the MVA requirements is performed together with the optimisation of highly background-discriminating PID variables. The optimisation of the classifiers' requirements is performed using data sidebands. In the procedure the whole dataset is split in two parts, according to the charge combination of the final state particles ($B^+ \to K^+\mu^+e^-$ and $B^+ \to K^+\mu^-e^+$). The optimisation performed on one sample is later applied to the other one. A sequential optimisation is used, where firstly the BDT optimal working point is found and applied, then the BDTHOP requirement

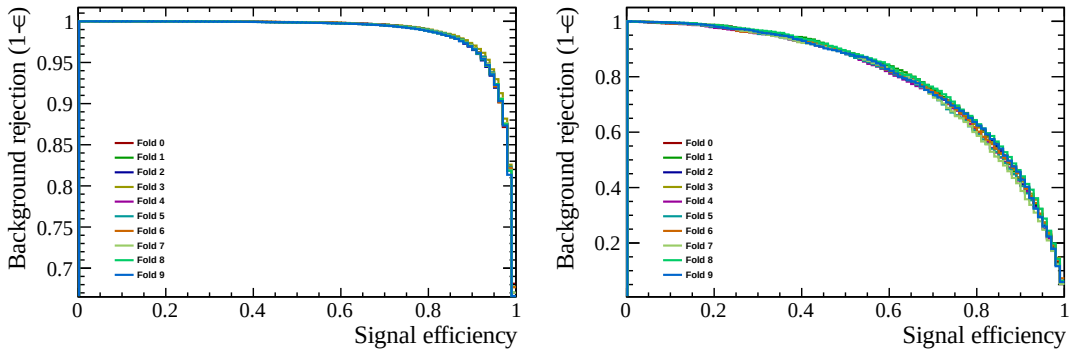


Figure 4.17: ROC curves for all 10 folds, for BDT (left) and BDTHOP (right). Each color corresponds to a different fold.

is optimised and applied, and in the last step the PID requirements are chosen. The sequential method is chosen because the amount of background events left after the previous requirements is not sufficient for the full 5 dimensional optimisation (two classifier's responses, and three PID variables). The whole procedure is performed using the CL_s method [59] (described in the next section), and the Punzi figure of merit [60] is used as a cross-check for the BDTHOP optimisation.

4.7.1 The CL_s method

The CL_s method can be used to set a conservative limit on the signal hypothesis, especially in the cases where the data sample is very limited, as it is in this analysis. The confidence of the chosen test-statistic q of the hypothesis, which assumes the existence of signal (s) and background (b) events in the sample, is defined as:

$$CL_{s+b} = \int_{-\infty}^{Q} \frac{dP_{s+b}}{dQ} dQ. \tag{4.6}$$

where $\frac{dP_{s+b}}{dQ}$ is the probability distribution function (PDF) and Q stands for the observed value of the test-statistic. Small values of CL_{s+b} favour the background hypothesis. In the situation of little background events, their fluctuation may lead to too strong signal exclusion. The confidence for the background only hypothesis is equal to:

$$CL_b = \int_{-\infty}^{Q} \frac{dP_b}{dQ} dQ. \tag{4.7}$$

where $\frac{dP_b}{dQ}$ is the probability distribution function for the background hypothesis only. An upper confidence limit for an exclusion is defined as a value of a population parameter (for example a particle's mass), for which the exclusion confidence is less than the specified confidence level. One can estimate the confidence in the signal hypothesis

with :

$$CL_s = \frac{CL_{s+b}}{CL_b}. \tag{4.8}$$

The exclusion of the signal hypothesis at the confidence level CL occurs when

$$1 - CL_s \leq CL. \tag{4.9}$$

4.7.2 BDT optimisation

Since this classifier is used to discriminate against the combinatorial background, only the upper sideband data is used, which is composed only of that type of background. A scan is performed over the BDT classifier's response. For each possible requirement the upper sideband data distribution is fitted with an exponential function (see Fig. 4.18). The extrapolation of the fitted function into the signal region is used to estimate its background pollution. Next, the expected upper limit is calculated with the CL_s method

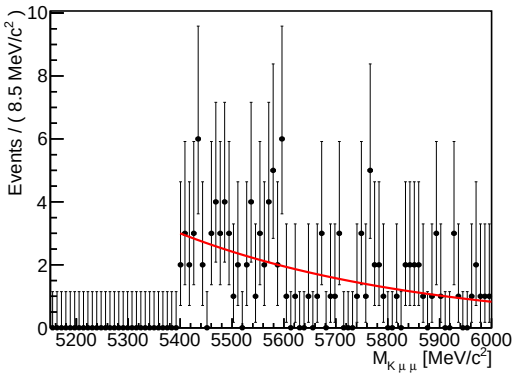


Figure 4.18: An exponential fit to the upper sideband data.

(described in the previous section). The optimal point for the BDT selection requirement is chosen to correspond to the lowest upper limit. For the optimisation procedure, the normalisation factor (see Eq. 4.1) is assumed to be equal to 10^{-10}, which does not have any impact on the optimal point. As mentioned before, the optimisation is performed on $B^+ \to K^+\mu^+e^-$ and $B^+ \to K^+\mu^-e^+$ samples separately. The obtained CL_s distributions are shown in Fig. 4.19. The optimal BDT cut value is set to be BDT > 0.98 for both charge combinations.

4.7.3 BDTHOP optimisation

With the optimal BDT cut applied, the BDTHOP requirement is optimised with a sample from the lower and the upper sideband data. The lower sideband data here is used because this classifier is trained to discriminate against the partially reconstructed

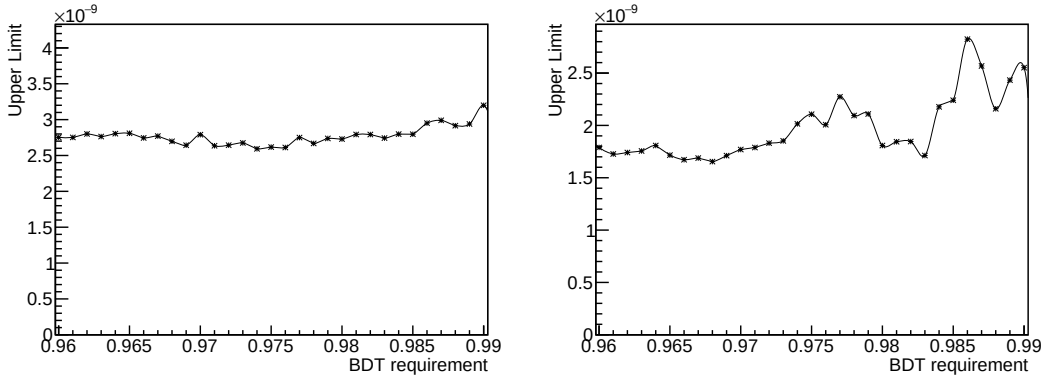


Figure 4.19: Expected upper limit calculated for different BDT cuts for $B^+ \to K^+\mu^+e^-$ (left) and $B^+ \to K^+\mu^-e^+$ (right) charge combinations.

backgrounds, which populate the left side of the examined mass range. The upper sideband data is used to help with the stability of the fit. The fit to the invariant mass distribution in the blinded data sample is shown in Fig. 4.20. The result of the BDTHOP

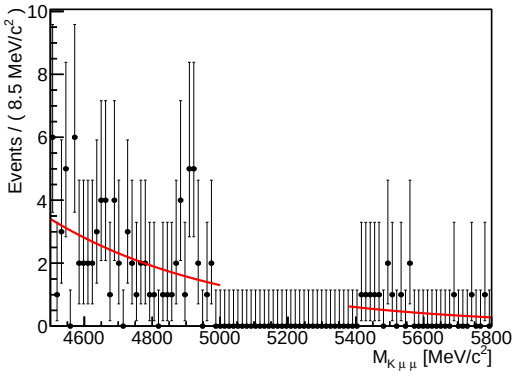


Figure 4.20: An exponential fit to upper and lower data sidebands.

optimisation can be seen in Fig. 4.21. The optimisation is performed after the BDT requirement is applied. With a decrease in the data sample the statistical fluctuations become more visible. The optimal BDTHOP selection requirement of BDTHOP>0.5 is chosen for both charge combinations of the signal decay.

Since the BDTHOP optimal requirement has a big statistical uncertainty, the same optimisation procedure is repeated with the Punzi Figure of Merit [60]:

$$\text{FOM} = \frac{\varepsilon}{a/2+\sqrt{b}}, \tag{4.10}$$

where ε is the signal efficiency, a ($a=3$ in this analysis) is a factor necessary to protect its value from vanishing background counts, and b is the number of background events. The result of this optimization is shown in Fig. 4.22. The same broad maximum is visible, which confirms the result of the original optimisation.

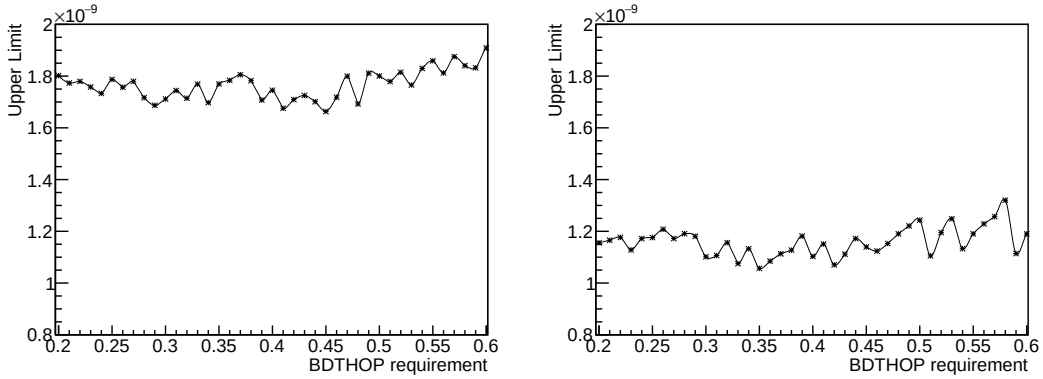


Figure 4.21: Expected upper limit calculated for different BDTHOP cuts for $B^+ \to K^+\mu^+e^-$ (left) and $B^+ \to K^+\mu^-e^+$ (right) charge combinations.

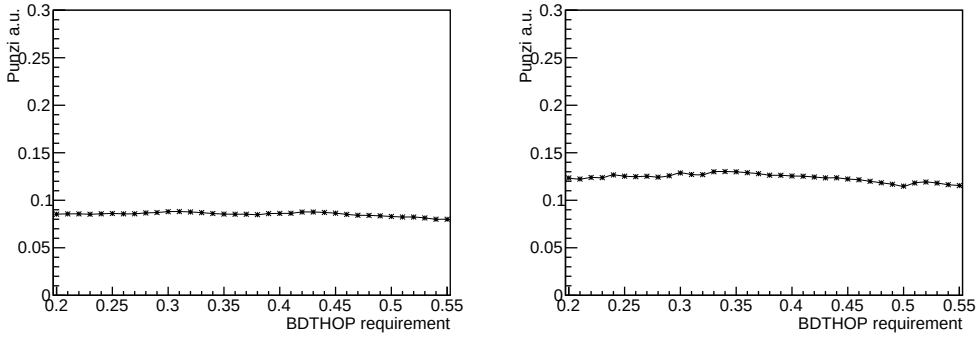


Figure 4.22: Results of the BDTHOP optimisation using the so-called Punzi FoM: $B^+ \to K^+\mu^+e^-$ (left), $B^+ \to K^+\mu^-e^+$ (right).

4.7.4 PID

The PID requirements are chosen after the BDT and the BDTHOP optimised requirements are applied. After such an effective selection not enough background events are left to perform a robust optimisation for all the three PID variables. A set of requirements based on a different LHCb analysis is chosen to further clean the sample:

$$\text{ProbNNmu} > 0.70, \tag{4.11}$$

$$\text{ProbNNe} > 0.65, \tag{4.12}$$

$$\text{ProbNNk} > 0.65. \tag{4.13}$$

A fit to the data sidebands is used to evaluate the expected background yields in the signal region. The calculated yields are reported in Tab. 4.6. The difference in yields can already be observed in the samples after the stripping selection, thus this is not an effect caused by the selection.

Decay channel	Expected background yield
$B^+ \to K^+\mu^+e^-$	4.26 ± 1.16
$B^+ \to K^+\mu^-e^+$	0.98 ± 0.68

Table 4.6: The expected number of background events in the signal region after the selection.

4.8 Exclusive background studies

In this section two types of possible background sources are studied:

- peaking backgrounds - fully reconstructed decays, in which one particle or more is misidentified, thus the final state particles are reconstructed as $Ke\mu$,
- partially reconstructed backgrounds - at least 4-body decays in which one particle or more is lost.

Various decays with misidentified and/or not reconstructed particles in the final state could mimic the $B^+ \to K^+\mu^\pm e^\mp$ signal decay. The most dangerous background sources are examined below using simulated MC samples and their yields are estimated. As for the signal MC, the background samples are re-weighted (using the same data-MC weights as described in Sec. 4.5.2) and the PID variables are re-sampled (see Sec. 4.5.3), and they are also processed with the same reconstruction and stripping requirements as the signal channel.

4.8.1 Peaking backgrounds

The candidates coming from B^+ decays with K^+ $\ell^+\ell^-$ in the final state, where one lepton (electron or muon) is misidentified, are reduced by PID requirements applied in the selection. Additionally, decays through the charmonium resonance ($B^+ \to J/\psi\,(\to \ell^+\ell^-)K^+$) are also reduced by the charmonium veto. The B^+ decays contributions to two pions with double misidentification are reduced by PID requirements.

4.8.2 Partially reconstructed backgrounds

Another group of backgrounds comes from B^+ decays to states with a charmed meson, lepton, and neutrino ($\bar{D}^0\ell^+\nu_\ell$). The charmed particles can later decay into states such as $K^+\ell^-\overline{\nu}_\ell$ or $K^+\pi^-$ (with misidentification). Those B^+ decays always have some amount of energy missing, which is due to the presence of at least one neutrino. Those background sources are easily reduced with the charm veto. Also, the latter decay with $K^+\pi^-$ in the final state is reduced by PID requirements applied in the selection. Λ_b^0 decays with pK^- $\ell^+\ell^-$ in the final state can be dangerous when one lepton is lost and

either the proton is misidentified as a lepton, or the proton is misidentified as a kaon and the true kaon is misidentified as a lepton. Those contributions are reduced by the PID requirements applied in the selection. The decays that occur through the charmonium resonance are additionally reduced with the charmonium veto, which was described in Sec. 4.4.

4.8.3 Estimation of background contributions

The background events are estimated with:

$$N_{\text{bkg}} = N_{\text{control}} \cdot \frac{\varepsilon^{\text{tot}}_{\text{bkg}}}{\varepsilon^{\text{tot}}_{\text{control}}} \cdot \frac{\mathscr{B}^{\text{bkg}}}{\mathscr{B}^{\text{control}}} \cdot \frac{f_{\text{bkg}}}{f_u}, \tag{4.14}$$

where the control stands for the $B^+ \to K^+ J/\psi\,(\to \mu^+\mu^-)$ control channel decay. $\varepsilon^{\text{tot}}_{\text{bkg}}$ stands for the total efficiency of the background decay. The value is estimated with simulated MC samples:

$$\varepsilon^{\text{tot}}_{\text{bkg}} = \frac{N^{\text{sel}}_{\text{bkg}}}{N^{\text{gen}}_{\text{bkg}}}, \tag{4.15}$$

where $N^{\text{sel}}_{\text{bkg}}$ is the number of events surviving the total selection, and $N^{\text{gen}}_{\text{bkg}}$ is the generated number of MC events. The number of generated events is calculated with the number of simulated events N_{sim} (listed in Tab. 4.1): $N^{\text{gen}}_{\text{bkg}} = N^{\text{sim}}_{\text{bkg}}/\varepsilon^{\text{acc}}_{bkg}$. The $\varepsilon^{\text{tot}}_{\text{control}}$ stands for the total efficiency of the control channel (see Sec. 4.10 for the evaluation). $\mathscr{B}^{\text{bkg}}$ stands for the background's branching fractions, and the $\mathscr{B}^{\text{control}}$ is the well known branching fraction of the control channel. The backgrounds branching fractions are listed in Tab. 4.7. They are taken from the PDG [7], with an exception of $\Lambda_b^0 \to pK^-\ell^+\ell^-$ decays, for which it was estimated with:

$$\frac{\mathscr{B}(\Lambda_b^0 \to pK^-\ell^+\ell^-)}{\mathscr{B}(B^+ \to K^+\mu^+\mu^-)} = \frac{\mathscr{B}(\Lambda_b^0 \to pK^- J/\psi\,(\to \ell^+\ell^-))}{\mathscr{B}(B^+ \to K^+ J/\psi\,(\to \mu^+\mu^-))}. \tag{4.16}$$

The ratio f_{bkg}/f_u corresponds to the production fraction of the decaying b hadron. For almost all background decays this value is equal to 1. The only exception are decays originating from Λ_b^0 for which the value $f_{\Lambda_b^0}/f_u$ is taken from [61] to be 0.240 ± 0.022.

Some of the background samples do not survive the selection, thus Eq. 4.14 does not fully apply. In that case, an upper limit for the efficiency of the selection is calculated, using a Bayesian approach. Then, the number of background events is multiplied by the factor F to estimate the pollution of the signal window. F is defined as a fraction of background events falling into the signal region. It is evaluated after the preselection and trigger requirements, with a few exceptions. A conservative factor of 0.91 is assigned to peaking backgrounds, such as $B^+ \to K^+e^+e^-$, $B^+ \to K^+\mu^+\mu^-$, and $B^+ \to K^+\pi^+\pi^-$. It is the highest possible factor as this value is taken from the efficiency of the signal channel and the background efficiency is not expected to be higher than the signal's.

Table 4.7: Branching ratios of the possible background sources [7].

Decay channel	Branching ratio
$B^+ \to K^+\mu^+\mu^-$	$(4.43 \pm 2.40) \times 10^{-7}$
$B^+ \to K^+e^+e^-$	$(5.50 \pm 0.70) \times 10^{-7}$
$B^+ \to K^+J/\psi(\to \mu^+\mu^-)$	$(6.12 \pm 0.19) \times 10^{-5}$
$B^+ \to K^+J/\psi(\to e^+e^-)$	$(6.13 \pm 0.19) \times 10^{-5}$
$\Lambda_b^0 \to pK^-\mu^+\mu^-$	$(1.38 \pm 0.27) \times 10^{-7}$
$\Lambda_b^0 \to pK^-e^+e^-$	$(1.38 \pm 0.39) \times 10^{-7}$
$\Lambda_b^0 \to pK^-J/\psi(\to \mu^+\mu^-)$	$(1.91 \pm 0.36) \times 10^{-5}$
$\Lambda_b^0 \to pK^-J/\psi(\to e^+e^-)$	$(1.91 \pm 0.36) \times 10^{-5}$
$B^- \to D^0(\to K^-\mu^+\nu_\mu)\overline{\nu}_\mu\mu^-$	$(7.51 \pm 0.47) \times 10^{-4}$
$B^+ \to \bar{D}^0(\to K^+e^-\overline{\nu}_e)e^+\nu_e$	$(8.01 \pm 0.40) \times 10^{-4}$
$B^+ \to \bar{D}^0(\to K^+\pi^-)e^+\nu_e$	$(8.83 \pm 0.44) \times 10^{-4}$
$B^+ \to K^+\pi^+\pi^-$	$(5.10 \pm 0.29) \times 10^{-5}$

The peaking nature of $B^+ \to K^+J/\psi(\to \mu^+\mu^-)$ and $B^+ \to K^+J/\psi(\to e^+e^-)$ decays is effectively reduced by charmonium vetos, thus the conservative approach is not needed. Moreover, the factor for charmonium Λ_b^0 decays is estimated by scaling the factor of non-resonant Λ_b^0 decay by the ratio of factors $B^+ \to K^+J/\psi(\to \mu^+\mu^-)$ and $B^+ \to K^+\mu^+\mu^-$ or $B^+ \to K^+J/\psi(\to e^+e^-)$ and $B^+ \to K^+e^+e^-$. The estimated F for $B^- \to D^0(\to K^-\mu^+\nu_\mu)\overline{\nu}_\mu\mu^-$ is equal to 0.3 and is applied to all charm decays, because its determination for the other D^0 decays is poorer. The factors are reported in Tab. 4.8 and the F^* stands for a factor chosen for a specific decay. For this calculation, the effect of PID cuts is neglected, hence the additional factor (F_{PID}) of 0.3 is applied to $B^+ \to K^+\pi^+\pi^-$.

The yields are estimated in the signal region ($5100\,\text{MeV}/c^2 < m_{K\mu e} < 5370\,\text{MeV}/c^2$) and are reported in the last column of Tab. 4.8. In summary, the selection efficiently reduces the pollution from exclusive backgrounds to a negligible level.

4.9 Invariant mass shape of the $Ke\mu$ system

This section describes the prediction of the invariant mass distribution of the $Ke\mu$ system. The invariant mass shape of the searched decay is necessary for the mass requirement efficiency calculation and for the yield extraction. Both these values are used later in the branching fraction calculation. Unfortunately, simulated signal events do not precisely show the proper shape. Hence, the shape is estimated the data driven way by fitting the simulated MC samples of $B^+ \to K^+J/\psi(\to \mu^+\mu^-)$, $B^+ \to K^+J/\psi(\to e^+e^-)$ and $B^+ \to K^+\mu^\pm e^\mp$ channels and the control channels data samples. The shapes of all the distributions are modelled by a double same sided Crystal Ball function (described in Sec. 4.5.1). All the parameters of the function are floated besides the tail parameter n,

Table 4.8: Number of signal (assuming a branching fraction of 10^{-8}) and background expected yields in the signal region after selection, estimated combining 2011 and 2012 MC samples whenever possible.

MC sample	Full mass region	F	F^*	$F_{\rm PID}$	Signal region
$B^+ \to K^+\mu e$	5.45 ± 1.0	0.908 ± 0.004	0.91	1	4.96 ± 0.91
$B^+ \to K^+\mu^+\mu^-$	< 0.0099	0.84 ± 0.1	0.91	1	< 0.009
$B^+ \to K^+e^+e^-$	0.0029 ± 0.0050	0.36 ± 0.1	0.91	1	0.0026 ± 0.0045
$B^+ \to K^+J/\psi(\to \mu^+\mu^-)$	< 0.57	0.10 ± 0.1	0.10	1	< 0.057
$B^+ \to K^+J/\psi(\to e^+e^-)$	< 1.13	0.10 ± 0.1	0.10	1	< 0.11
$\Lambda_b^0 \to pK^-\mu^+\mu^-$	0.0034 ± 0.0017	0.36 ± 0.10	0.4	1	0.0014 ± 0.0008
$\Lambda_b^0 \to pK^-e^+e^-$	< 0.0013	0.03 ± 0.02	0.4	1	< 0.00052
$\Lambda_b^0 \to pK^-J/\psi(\to \mu^+\mu^-)$	< 0.28	0.03 ± 0.02	0.05	1	< 0.014
$\Lambda_b^0 \to pK^-J/\psi(\to e^+e^-)$	< 0.28	< 0.03	0.05	1	< 0.014
$B^- \to D^0(\to K^-\mu^+\nu_\mu)\overline{\nu}_\mu\mu^-$	< 2.7	< 0.03	0.03	1	< 0.081
$B^+ \to \bar{D}^0(\to K^+e^-\overline{\nu}_e)e^+\nu_e$	< 0.54	< 0.01	0.03	1	< 0.0162
$B^+ \to \bar{D}^0(\to K^+\pi^-)e^+\nu_e$	< 3.0	< 0.14	0.03	1	< 0.09
$B^+ \to K^+\pi^+\pi^-$	< 2.0	0.73 ± 0.2	0.91	0.03	< 0.055

which is taken from the fit to the $B^+ \to K^+J/\psi(\to \mu^+\mu^-)$ data sample (see the right plot in Fig. 4.23c). The fits are performed on the samples after the full selection with additional requirements:

- $B^+ \to K^+J/\psi(\to \mu^+\mu^-)$: the di-lepton mass is required to be ± 60 MeV/c^2 around the known mass of the J/ψ resonance in both MC and data samples; the fitting range of the data sample is chosen to be $m_{K\mu\mu} \in [5180, 5700]$, and for the MC sample is restricted to $m_{K\mu\mu} \in [5180, 5370]$ MeV/c^2,

- $B^+ \to K^+J/\psi(\to e^+e^-)$: the di-lepton mass is required to be $^{+200}_{-500}$ MeV/c^2 around the known mass of the J/ψ resonance in both MC and data samples; the fitting range of the data sample is $m_{Kee} \in [4800, 5700]$ MeV/c^2, which is wider compared to the muon mode because of the existence of the visible tail on the lower side of the peak, the MC sample fitting range is restricted to $m_{Kee} \in [4900, 5500]$ MeV/c^2; the constraint on the J/ψ mass $m_{Kee} > 5150$ MeV/c^2 is applied to clear the electron data sample from additional background pollution,

- $B^+ \to K^+\mu^\pm e^\mp$: the di-lepton mass is required to be $^{+200}_{-500}$ MeV/c^2 around the known mass of the J/ψ resonance in both MC and data samples; the fitting range is restricted to $m_{Ke\mu} \in [4900, 5500]$ MeV/c^2.

A possible missing energy due to radiated photons in decays involving electrons can significantly change the reconstructed B^+ mass distribution. Thus, the signal shape prediction is performed separately for two bremsstrahlung categories: `HasBremAdded=0` is the name of the subsample where there is no bremsstrahlung photon associated with any of the electrons, and `HasBremAdded=1` is the name of the subsample where there is one bremsstrahlung photon recovered for just one electron.

4.9.1 Estimation of the $B^+ \to K^+\mu^\pm e^\mp$ shape

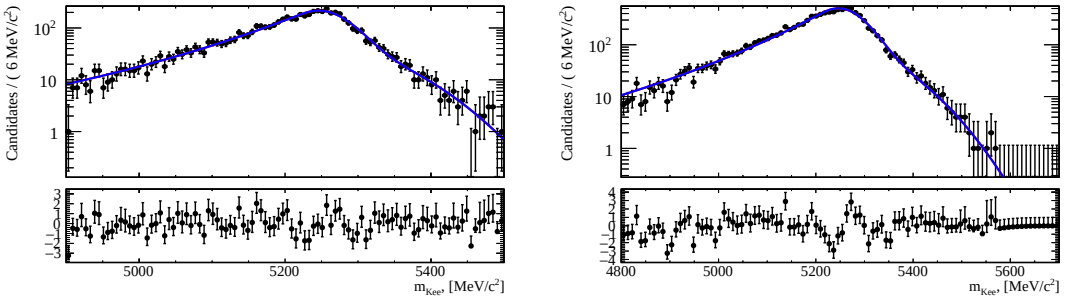
The shape is estimated by correcting fit parameters of simulated $B^+ \to K^+\mu^\pm e^\mp$ events by the parameters difference from fits to electron and muon channels:

$$P_{\mu e}^{\text{Predicted}} = P_{ee}^{\text{Data}} + (P_{\mu e}^{\text{MC}} - P_{ee}^{\text{MC}}) \cdot \frac{P_{ee}^{\text{Data}} - P_{\mu\mu}^{\text{Data}}}{P_{ee}^{\text{MC}} - P_{\mu\mu}^{\text{MC}}}, \tag{4.17}$$

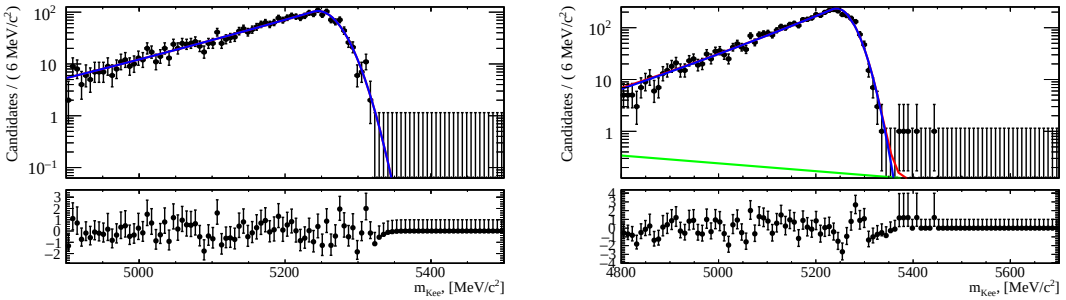
where P_{ee}^{MC} and P_{ee}^{Data} stands for parameters from fits to $B^+ \to K^+ J/\psi(\to e^+e^-)$ MC and data samples, respectively. $P_{\mu\mu}^{\text{MC}}$ and $P_{\mu\mu}^{\text{Data}}$ stand for the $B^+ \to K^+ J/\psi(\to \mu^+\mu^-)$ fit parameters to the MC and data samples, respectively. The $P_{\mu e}^{\text{MC}}$ stands for parameters from the fit to the signal MC sample. The values obtained from fits to control channels are reported in Tab. 4.9 and the corresponding distributions can be seen in Fig. 4.23 with their pull distributions. It can be clearly seen that the chosen model describes their shapes correctly.

Table 4.9: Fit parameters for the control channel on simulated events (MC), on data sample (Data), and their predicted values. The parameter n is fixed.

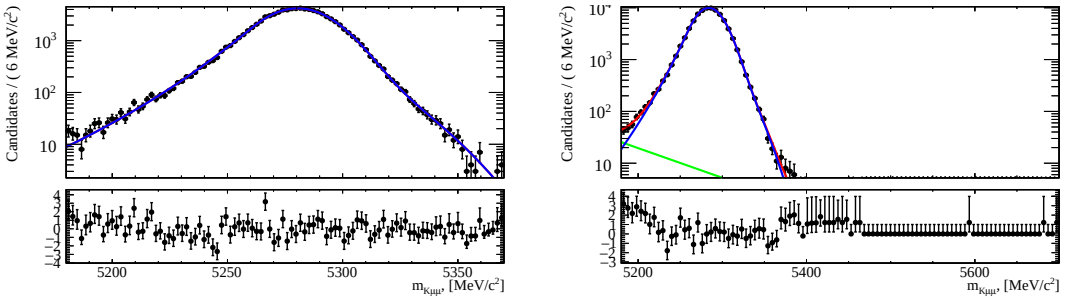
Parameter	MC	Data
$B^+ \to K^+ J/\psi(\to e^+e^-)$, `HasBremAdded=0`		
$\mu\,[\text{MeV}/c^2]$	5245.06 ± 1.67	5242.32 ± 1.41
$\sigma\,[\text{MeV}/c^2]$	26.18 ± 1.06	30.28 ± 0.79
$\sigma_1\,[\text{MeV}/c^2]$	26.18 ± 6.21	30.28 ± 1.25
n	20.36 ± 0.0	20.36 ± 0.0
α	0.25 ± 0.01	0.27 ± 0.01
c	1.0 ± 0.06	0.999 ± 0.297
$B^+ \to K^+ J/\psi(\to e^+e^-)$, `HasBremAdded=1`		
$\mu\,[\text{MeV}/c^2]$	5245.7 ± 1.53	5246.04 ± 1.16
$\sigma\,[\text{MeV}/c^2]$	37.65 ± 2.16	47.07 ± 1.33
$\sigma_1\,[\text{MeV}/c^2]$	89.26 ± 5.15	101.24 ± 4.07
n	20.36 ± 0.0	20.36 ± 0.0
α	0.55 ± 0.02	0.6 ± 0.02
c	0.668 ± 0.052	0.731 ± 0.033
$B^+ \to K^+ J/\psi(\to \mu^+\mu^-)$		
$\mu\,[\text{MeV}/c^2]$	5280.47 ± 0.08	5284.0 ± 0.07
$\sigma\,[\text{MeV}/c^2]$	14.51 ± 0.14	16.44 ± 0.19
$\sigma_1\,[\text{MeV}/c^2]$	27.03 ± 0.95	27.36 ± 1.12
n	20.36 ± 0.0	20.36 ± 6.36
α	1.36 ± 0.02	1.51 ± 0.03
c	0.868 ± 0.02	0.847 ± 0.029



(a) Fits to $B^+ \to K^+ J/\psi\,(\to e^+e^-)$ channel, for the `HasBremAdded=1`



(b) Fits to $B^+ \to K^+ J/\psi\,(\to e^+e^-)$ channel, for the `HasBremAdded=0`



(c) Fits to $B^+ \to K^+ J/\psi\,(\to \mu^+\mu^-)$ channel.

Figure 4.23: Fits to control channels MC (left) and data (right) samples. Pull distributions can be seen at the bottom of the plots. The y-axis scale is logarithmic. The blue line corresponds to the total PDF distribution, combinatorial background is green, and the searched double CB distribution is marked in red. The same convention is applied in all fit plots in this chapter.

As expected, the widths of the Gaussian core are smaller for the muon control channel than for the electron control channel. The same values for both widths in the `HasBremAdded=0` are caused by the vanishing contribution from one of the Crystall Balls (see values of the c parameter). In the electron channels, the difference between the data and MC parameters is clear, especially for the `HasBremAdded=1` category. Hence these corrections are necessary.

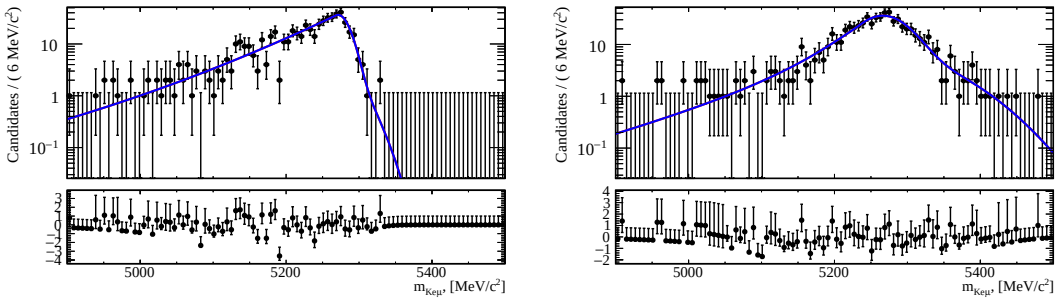


Figure 4.24: Fits to $B^+ \to K^+\mu^\pm e^\mp$ simulated sample to the `HasBremAdded=0` category (left) and to the `HasBremAdded=1` category (right). Pull distributions can be seen at the bottom of the plots. The y-axis scale is logarithmic.

MC fit parameters to the signal sample are reported in Tab. 4.10 together with its predictions based on the aforementioned fits. The fits to both bremsstrahlung categories, which can be seen in Fig. 4.24, show that the model is successful in describing their shape.

4.9.2 Cross-check on the $B^+ \to K^+ J/\psi\,(\to e^+e^-)$ shape

The invariant mass of the Kee is estimated by the calibration of its simulated samples, with information from the $B^+ \to K^+ J/\psi\,(\to \mu^+\mu^-)$ decay channel. The formula used for each parameter is given by:

$$P_{ee} = P_{ee}^{\mathrm{MC}} \cdot \frac{P_{\mu\mu}^{\mathrm{MC}}}{P_{\mu\mu}^{\mathrm{Data}}}, \tag{4.18}$$

where P_{ee}^{MC} stands for parameters from the fit to the $B^+ \to K^+ J/\psi\,(\to e^+e^-)$ MC sample. $P_{\mu\mu}^{\mathrm{MC}}$ and $P_{\mu\mu}^{\mathrm{Data}}$ stand for the $B^+ \to K^+ J/\psi\,(\to \mu^+\mu^-)$ fit parameters to MC and data samples, respectively. The same fit parameters are used as in the previous subsection. As mentioned previously, the prediction is performed in two bremsstrahlung categories. As can be seen in Fig. 4.25, the simulation does not describe the electron channel's mass shape perfectly, yet the calibration procedure brings the distributions closer. The resulting difference is not relevant considering the statistical precision of this analysis.

Table 4.10: Fit parameters for the $B^+ \to K^+\mu^+e^-$ simulated sample (MC), and their predicted values, n is fixed.

Parameter	MC	Predicted
`HasBremAdded=0`		
$\mu\,[\mathrm{MeV}/c^2]$	5271.32 ± 2.81	5273.23 ± 3.37
$\sigma\,[\mathrm{MeV}/c^2]$	13.87 ± 1.75	15.69 ± 2.09
$\sigma_1\,[\mathrm{MeV}/c^2]$	28.72 ± 2.21	21.55 ± 44.56
n	20.36 ± 0.0	20.36 ± 0.0
α	0.22 ± 0.02	0.24 ± 0.02
c	0.882 ± 0.112	0.863 ± 0.136
`HasBremAdded=1`		
$\mu\,[\mathrm{MeV}/c^2]$	5266.33 ± 1.93	5268.56 ± 2.26
$\sigma\,[\mathrm{MeV}/c^2]$	31.56 ± 3.2	39.01 ± 4.83
$\sigma_1\,[\mathrm{MeV}/c^2]$	79.7 ± 6.26	89.88 ± 9.7
n	20.36 ± 0.0	20.36 ± 0.0
α	0.9 ± 0.06	0.99 ± 0.07
c	0.671 ± 0.109	0.732 ± 0.078

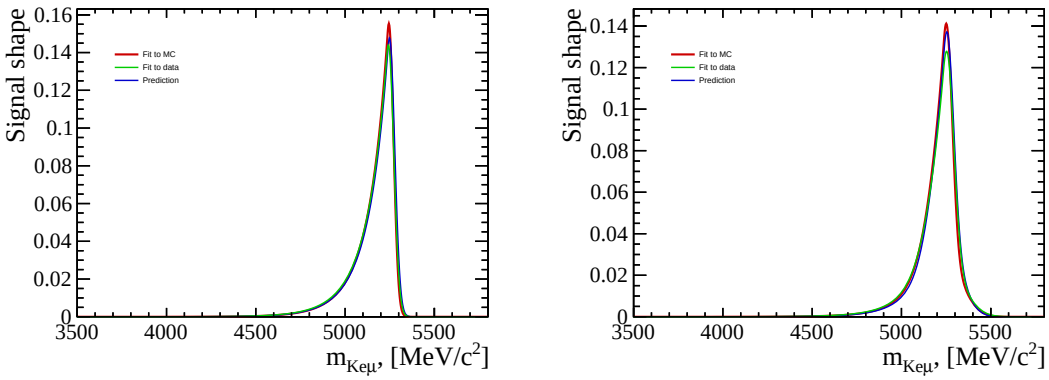


Figure 4.25: Prediction of a shape for the $B \to KJ/\psi(\to ee)$ channel. The red colour is a shape fitted to MC samples, green to data, and blue is our prediction. The plot on the left corresponds to the `HasBremAdded=0` category, and the one on the right - to the `HasBremAdded=1` category.

4.10 Normalisation of the signal yield

The normalisation constant α is taken from the already defined branching fraction formula 4.1 and it is given with:

$$\alpha(B^+ \to K^+\mu^\pm e^\mp) = \frac{\mathscr{B}(B^+ \to K^+ J/\psi(\to \mu^+\mu^-))}{N(B^+ \to K^+ J/\psi(\to \mu^+\mu^-))} \times \frac{\varepsilon(B^+ \to K^+ J/\psi(\to \mu^+\mu^-))}{\varepsilon(B^+ \to K^+\mu^\pm e^\mp)}. \tag{4.19}$$

As mentioned in Sec. 4.1, ε stands for the total selection efficiency. It can be decomposed into its components:

$$\begin{aligned}\varepsilon = \varepsilon_{\text{PID|BDTHOP}} \times \varepsilon_{\text{BDTHOP|BDT}} \times \varepsilon_{\text{BDT|PIDs}} \times \varepsilon_{\text{PIDs|trig}} \\ \times \varepsilon_{\text{trig|presel}} \times \varepsilon_{\text{presel|rec\&strip}} \times \varepsilon_{\text{rec\&strip|acc}} \times \varepsilon_{\text{acc}},\end{aligned} \tag{4.20}$$

where the following ingredients stand for:

- ε_{acc} stands for the geometric acceptance,
- $\varepsilon_{\text{rec\&strip|acc}}$ is the reconstruction stripping selection efficiency (excluding PID requirements),
- $\varepsilon_{\text{presel|rec}}$ is the preselection requirements efficiency (excluding PID requirements),
- $\varepsilon_{\text{trig|presel}}$ is the trigger requirements efficiency,
- $\varepsilon_{\text{PIDs|trig}}$ is the efficiency of the PID requirements excluded from the previous calculations,
- $\varepsilon_{\text{BDT|PIDs}}$ is the BDT requirement efficiency,
- $\varepsilon_{\text{BDTHOP|BDT}}$ is the BDTHOP requirement efficiency,
- $\varepsilon_{\text{PID|BDTHOP}}$ is the efficiency of the final PID requirements.

The efficiencies are determined separately for 2011 and 2012 data taking conditions and separately for two charge combinations $B^+ \to K^+\mu^+e^-$ and $B^+ \to K^+\mu^-e^+$ of the signal channel. In the case of the signal channel, the normalisation factor is calculated also in two $K\mu e$ invariant mass ranges (a full mass range, and a signal region). The latter is obtained by a small change in the efficiency formula:

$$\varepsilon = \varepsilon \times \varepsilon_{\text{mass}}, \tag{4.21}$$

where $\varepsilon_{\text{mass}}$ stands for the signal mass region requirement ($5100\,\text{MeV}/c^2 < m_{K\mu e} < 5370\,\text{MeV}/c^2$). In the case of the control channel, an extra requirement was applied to the invariant mass of the $K\mu\mu$ system around the known J/ψ mass. Thus the control channel efficiency is given by:

$$\varepsilon = \varepsilon \times \varepsilon_{\text{J}/\psi}. \tag{4.22}$$

4.10.1 Geometric acceptance efficiency

This requirement is applied during the MC generation. As already mentioned, this criterion requires that the particles lie within the pseudorapidity range of $0.05 < \eta < 0.4$, and the events outside that region are not simulated. The acceptance efficiency (ε_{acc}) is provided in [62]. The efficiency for the signal channel is equal to (16.69 ± 0.03) % and (16.93 ± 0.031) % for 2011 and 2012 data taking conditions, respectively. The same value is taken for both charge combinations of the signal channel, as they are expected to be the same. The efficiency for the $B^+ \to K^+ J/\psi\,(\to \mu^+\mu^-)$ control channel is equal to (16.37 ± 0.021) % and (16.66 ± 0.048) % for 2011 and 2012 data taking conditions, respectively. These efficiencies are similar to each other, and this fact has been observed in other LHCb analyses.

4.10.2 Reconstruction and stripping efficiencies

The efficiencies of the reconstruction and stripping are combined. They are calculated using information about the fraction of the simulated events passing the reconstruction and stripping requirements. As mentioned before, the PID requirements are removed while processing the simulated events, and they are taken into account separately in one of the subsequent steps of the efficiency calculation. The calculation is made separately for both charge combinations and its results are also merged into one value. The efficiencies for the signal channels and the normalisation channel are given in Tab. 4.11.

Table 4.11: Reconstruction and stripping efficiency, excluding PID requirements. Given for the searched decay and its control channel, separately for 2011 and 2012 data taking conditions.

Decay channel	$\varepsilon_{\text{rec\&strip}\|\text{acc}}$ [%]	
	2011	2012
$B^+ \to K^+\mu^-e^+$	17.719 ± 0.076	16.219 ± 0.052
$B^+ \to K^+\mu^+e^-$	17.377 ± 0.075	16.429 ± 0.052
$B^+ \to K^+\mu^\pm e^\mp$	17.546 ± 0.053	16.324 ± 0.037
$B^+ \to K^+ J/\psi\,(\to \mu^+\mu^-)$	26.22 ± 0.061	23.99 ± 0.044

The 2011 efficiencies are slightly higher compared to the 2012 efficiencies and as expected, the di-muon $B^+ \to K^+ J/\psi\,(\to \mu^+\mu^-)$ efficiency is greater than the rare mode's efficiency.

4.10.3 Preselection efficiency

These calculations are performed as in the previous calculation, and they are carried out using simulated samples. Likewise, the PID requirements are removed while processing

the simulated events, and the results are reported for the 2011 and 2012 data taking conditions separately. The results are summarised in Tab. 4.12.

Table 4.12: Preselection efficiency of signal channels, excluding PID requirements. Given for the searched decay and its control channel, separately for 2011 and 2012 data taking conditions.

Decay channel	$\varepsilon_{\text{presel\|rec\&strip}}$ [%]	
	2011	2012
$B^+ \to K^+\mu^-e^+$	40.21 ± 0.23	41.35 ± 0.17
$B^+ \to K^+\mu^+e^-$	56.52 ± 0.23	55.76 ± 0.17
$B^+ \to K^+\mu^\pm e^\mp$	48.35 ± 0.16	48.57 ± 0.12
$B^+ \to K^+ J/\psi(\to \mu^+\mu^-)$	72.56 ± 0.061	72.53 ± 0.05

The difference between both charge combinations of the signal mode is clear. The main dissimilarity comes from the D^0 veto's efficiency (see Sec. 4.4). That contrast is caused by the asymmetry of Dalitz distributions in the charge combinations (which was visible even at the level of the generation $B^+ \to K^+\mu^-e^+$ and $B^+ \to K^+\mu^+e^-$ events). Again, the efficiency of the normalisation is greater than the rare modes efficiencies. These efficiencies are comparable between the 2011 and 2012 data taking conditions.

4.10.4 Trigger efficiency

Trigger requirements efficiencies are calculated similarly to the previous calculations, using information about the events passing the trigger requirements and the events passing the previous requirements. The obtained values, for both data taking conditions are given in Tab. 4.13.

Table 4.13: Trigger efficiency of signal channels, excluding PID requirements. Given for the searched decay and its control channel, separately for 2011 and 2012 data taking conditions.

Decay channel	$\varepsilon_{\text{trig\|presel}}$ [%]	
	2011	2012
$B^+ \to K^+\mu^-e^+$	48.01 ± 0.34	50.22 ± 0.25
$B^+ \to K^+\mu^+e^-$	50.74 ± 0.34	49.97 ± 0.25
$B^+ \to K^+\mu^\pm e^\mp$	49.6 ± 0.24	50.07 ± 0.18
$B^+ \to K^+ J/\psi(\to \mu^+\mu^-)$	74.67 ± 0.07	75.45 ± 0.05

The difference between the signal charge combinations is small (smaller than relatively 6%), and as previously, the efficiency of the normalisation mode is the highest. The efficiencies between the data-taking conditions do not vary significantly.

4.10.5 PID efficiency - stripping and preselection

After the MC samples are preselected and triggered, a PID resampling (described in Sec. 4.5.3) is performed. With these corrections it is possible to properly calculate the efficiency of the PID requirements, which were removed from stripping and preselection. In addition, at this step, all the MC samples are corrected for data-MC differences (described in 4.5.2). Thus this efficiency is calculated as a ratio of the sum of data-MC weights from the simulated events passing these requirements over the overall sum of weights. The calculated efficiencies are summarised in Tab. 4.14.

Table 4.14: PID stripping and preselection requirements efficiency of signal channels, excluding PID requirements. Given for the searched decay and its control channel, separately for 2011 and 2012 data taking conditions.

Decay channel	$\varepsilon_{\text{PIDs\|trig}}$ [%]	
	2011	2012
$B^+ \to K^+\mu^-e^+$	83.25 ± 0.34	84.82 ± 0.24
$B^+ \to K^+\mu^+e^-$	84.47 ± 0.34	84.79 ± 0.25
$B^+ \to K^+\mu^\pm e^\mp$	83.98 ± 0.24	84.8 ± 0.18
$B^+ \to K^+ J/\psi(\to \mu^+\mu^-)$	87.89 ± 0.06	88.49 ± 0.04

This efficiencies do not significantly differ among the signal channels. Again, the efficiency of the $B^+ \to K^+ J/\psi(\to \mu^+\mu^-)$ is greater than $B^+ \to K^+\mu^\pm e^\mp$. Since these PID requirements are very loose, the efficiencies are substantial.

4.10.6 Multivariate classifier efficiency

In this section the efficiencies of trained multivariate classifiers (BDT and BDTHOP separately) are reported (see Tab. 4.15) for both data taking conditions (2011 and 2012). The BDT requirement efficiency is calculated on top of the previous requirements, and the BDTHOP requirement efficiency is calculated with the BDT requirement applied. These efficiencies are obtained using the data-MC weight, as for the preceding PID requirements.

The differences among the signal channels in the simulated samples for 2011 data-taking conditions are greater than in those simulated for 2012. The highest one is relatively ~ 10 %. Multivariate classifiers are trained on simulated $B^+ \to K^+\mu^\pm e^\mp$ events and optimised to get the best limit for signal channels, so in this case the efficiency for the $B^+ \to K^+ J/\psi(\to \mu^+\mu^-)$ decay is not necessarily higher (see the BDT efficiency). The BDTHOP requirement efficiency is greater for the control channel as this decay's lower sideband is cleaner.

Table 4.15: BDT efficiency of signal channels, excluding PID requirements. Given for the searched decay and its control channel, separately for 2011 and 2012 data taking conditions.

Decay channel	2011	2012
	$\varepsilon_{\text{BDT}\|\text{PIDs}}$ [%]	
$B^+ \to K^+\mu^-e^+$	53.17 ± 0.49	49.63 ± 0.37
$B^+ \to K^+\mu^+e^-$	48.17 ± 0.51	47.41 ± 0.38
$B^+ \to K^+\mu^\pm e^\mp$	50.14 ± 0.35	48.35 ± 0.26
$B^+ \to K^+ J/\psi(\to \mu^+\mu^-)$	46.84 ± 0.1	43.71 ± 0.07
	$\varepsilon_{\text{BDTHOP}\|\text{BDT}}$ [%]	
$B^+ \to K^+\mu^-e^+$	53.05 ± 0.68	49.28 ± 0.52
$B^+ \to K^+\mu^+e^-$	52.27 ± 0.7	51.04 ± 0.54
$B^+ \to K^+\mu^\pm e^\mp$	52.59 ± 0.49	50.27 ± 0.37
$B^+ \to K^+ J/\psi(\to \mu^+\mu^-)$	58.53 ± 0.14	55.83 ± 0.11

4.10.7 PID efficiency

The final PID requirements are calculated on top of the multivariate selection. As in the previous step, the efficiency is calculated as the sum of weights for both signal and control channels, for 2011 and 2012 data taking conditions separately. The efficiencies are summarised in Tab. 4.16.

Table 4.16: PID efficiency of signal channels, excluding PID requirements. Given for the searched decay and its control channel, separately for 2011 and 2012 data taking conditions.

Decay channel	$\varepsilon_{\text{PID}\|\text{BDTHOP}}$ [%]	
	2011	2012
$B^+ \to K^+\mu^-e^+$	54.59 ± 0.92	51.41 ± 0.72
$B^+ \to K^+\mu^+e^-$	53.14 ± 0.95	53.85 ± 0.73
$B^+ \to K^+\mu^\pm e^\mp$	53.75 ± 0.66	52.81 ± 0.52
$B^+ \to K^+ J/\psi(\to \mu^+\mu^-)$	42.03 ± 0.18	42.79 ± 0.14

The differences between the data-taking conditions are small, as well as the difference among the signal channels. The PID requirement efficiency for the control channel is smaller because this requirement is optimised for the signal channels.

4.10.8 The J/ψ mass window requirement

The final requirement on the $B^+ \to K^+ J/\psi(\to \mu^+\mu^-)$ control sample is the J/ψ mass requirement, which is applied when calculating the signal yield. The requirement on

the $B^+ \to K^+ J/\psi(\to \mu^+\mu^-)$ sample is chosen to be $3037 < m_{J/\psi} < 4057\,\text{MeV}$. The efficiencies of this requirement are $96.99 \pm 0.09\%$ and $97.09 \pm 0.07\%$, for 2011 and 2012 data taking conditions, respectively.

4.10.9 The signal mass window efficiency

The last efficiency calculated is the signal mass window requirement efficiency. As mentioned before, the signal window is chosen to be $5100\,\text{MeV}/c^2 < m_{K\mu e} < 5370\,\text{MeV}/c^2$. Its efficiency is taken from the PDF of the predicted invariant $K\mu e$ mass shape (see Sec. 4.9 for more details). The estimation is performed separately for `HasBremAdded=0` and `HasBremAdded=1` categories, and the results are also merged into one value for both samples. The efficiency is taken as a fraction of the integrated PDF in the signal region, over the PDF integrated in the full mass range. The obtained results are given in Tab. 4.17. The combined efficiency is calculated using fractions of the `HasBremAdded=0`, and the `HasBremAdded=1` events:

$$\varepsilon_{mass} = \varepsilon_{mass}^{brem0} f_0 + \varepsilon_{mass}^{brem1} f_1, \tag{4.23}$$

where $\varepsilon_{mass}^{brem0}$ ($\varepsilon_{mass}^{brem1}$) is the mass requirement efficiency of the `HasBremAdded=0` (`HasBremAdded=1`) category, and the fractions are $f_0 = 0.43$ and $f_1 = 0.57$.

Table 4.17: Mass requirement efficiency for the $B^+ \to K^+\mu^\pm e^\mp$ channel for `HasBremAdded=0`, and `HasBremAdded=1`.

	ε_{mass} [%]
`HasBremAdded=0`	91.87 ± 2.57
`HasBremAdded=1`	88.61 ± 2.02
Combined	90.001 ± 1.64

4.10.10 Normalisation summary

The above outcomes are used to calculate the total efficiencies, according to Eq. 4.21. Then, the efficiencies are used to compute the normalisation factor for both data taking conditions separately. All ingredients for the normalisation factors and their values are summarised in Tab 4.18 and 4.19 for 2011 and 2012 data-taking conditions respectively. The results are also given for the signal window. The normalisation factors for 2012 data-taking conditions are better as the data sample is greater.

In the end, the obtained normalisation factors are merged to get one normalisation factor for the whole data sample. It as achieved with the following formula:

$$\frac{1}{\alpha} = \frac{1}{\alpha_{2011}} + \frac{1}{\alpha_{2012}}.$$

The results are given in Table 4.20 below.

Table 4.18: Normalization summary for 2011 datasets, with the full mass range and including signal window efficiency for signal samples.

Normalization term	Full mass range	Signal window
$\varepsilon(B^+ \to K^+\mu^- e^+)$	0.000732 ± 0.000022	0.000659 ± 0.000021
$\varepsilon(B^+ \to K^+\mu^+ e^-)$	$0.00094 \pm 0.00002.5$	0.000846 ± 0.000024
$\varepsilon(B^+ \to K^+\mu^\pm e^\mp)$	0.000836 ± 0.000017	0.000752 ± 0.000016
$\varepsilon(B^+ \to K^+ J/\psi(\to \mu^+\mu^-))$	0.002286 ± 0.000014	
$\mathscr{B}(B^+ \to K^+ J/\psi(\to \mu^+\mu^-))$	$(6.11 \pm 0.19) \times 10^{-5}$	
$N(B^+ \to K^+ J/\psi(\to \mu^+\mu^-))$	26920.5 ± 159.318	
$\alpha(B^+ \to K^+\mu^- e^+)$	$(7.0832 \pm 0.3109) \times 10^{-9}$	$(7.8679 \pm 0.3554) \times 10^{-9}$
$\alpha(B^+ \to K^+\mu^+ e^-)$	$(5.5159 \pm 0.2293) \times 10^{-9}$	$(6.1288 \pm 0.2619) \times 10^{-9}$

Table 4.19: Normalization summary for 2012 datasets, with the full mass range and including signal window efficiency for signal samples.

Normalization term	Full mass range	Signal window
$\varepsilon(B^+ \to K^+\mu^- e^+)$	0.000608 ± 0.000014	0.000547 ± 0.000013
$\varepsilon(B^+ \to K^+\mu^+ e^-)$	0.000856 ± 0.000017	0.000771 ± 0.000016
$\varepsilon(B^+ \to K^+\mu^\pm e^\mp)$	0.000732 ± 0.000011	0.000658 ± 0.00001
$\varepsilon(B^+ \to K^+ J/\psi(\to \mu^+\mu^-))$	0.001962 ± 0.000009	
$\mathscr{B}(B^+ \to K^+ J/\psi(\to \mu^+\mu^-))$	$(6.11 \pm 0.19) \times 10^{-5}$	
$N(B^+ \to K^+ J/\psi(\to \mu^+\mu^-))$	59165.4 ± 248.133	
$\alpha(B^+ \to K^+\mu^- e^+)$	$(3.3325 \pm 0.1297) \times 10^{-9}$	$(3.7041 \pm 0.1457) \times 10^{-9}$
$\alpha(B^+ \to K^+\mu^+ e^-)$	$(2.3670 \pm 0.0879) \times 10^{-9}$	$(2.6279 \pm 0.0989) \times 10^{-9}$

Table 4.20: The total normalization factor for 2011 and 2012 datasets combined.

Normalization term	Full mass range	Signal window
$B^+ \to K^+\mu^- e^+$	$(2.27 \pm 0.16) \times 10^{-9}$	$(2.52 \pm 0.18) \times 10^{-9}$
$B^+ \to K^+\mu^+ e^-$	$(1.66 \pm 0.12) \times 10^{-9}$	$(1.84 \pm 0.14) \times 10^{-9}$
$B^+ \to K^+\mu^\pm e^\mp$	$(1.92 \pm 0.12) \times 10^{-9}$	$(2.13 \pm 0.13) \times 10^{-9}$

4.11 Systematic uncertainty calculation

Most of the external systematic uncertainty sources cancel, which is due to the fact that this is a normalised measurement. Yet, several systematic uncertainty sources are studied in the following sections. The results are summarised in Sec. 4.11.7.

4.11.1 Kinematic re-weighting uncertainty

After the re-weighting procedure there are still some data-MC discrepancies left. Their impact is estimated by re-weighting samples in the B impact parameter and HOP, after the original re-weighting procedure is performed. Afterwards, the BDT and BDTHOP efficiency is recalculated with the new weights. The systematic uncertainty is estimated with the formula:

$$\Delta\varepsilon_{\mathrm{IP}\setminus\mathrm{HOP}} = \frac{|\varepsilon_{\mathrm{IP}\setminus\mathrm{HOP}} - \varepsilon_{\mathrm{nominal}}|}{\varepsilon_{\mathrm{nominal}}}, \tag{4.24}$$

where $\varepsilon_{\mathrm{nominal}}$ stands for the BDT and BDTHOP efficiency calculated with the original data-MC weights, and the $\varepsilon_{\mathrm{IP}\setminus\mathrm{HOP}}$ stands for the efficiency calculated with the aforementioned weights. This is reported in Tab. 4.21. A conservative systematic of 1 % is assigned.

Table 4.21: Systematic uncertainty, based on BDT and BDTHOP requirements efficiencies, calculated separately for 2011 and 2012 data taking conditions for $B^+ \to K^+\mu^\pm e^\mp$.

Systematic uncertainty [%]	
$\Delta\varepsilon_{IP\ 2011}$	0.54
$\Delta\varepsilon_{IP\ 2012}$	0.59
$\Delta\varepsilon_{HOP\ 2011}$	0.45
$\Delta\varepsilon_{HOP\ 2012}$	0.48

4.11.2 PID resampling uncertainty

As mentioned previously, for the PID resampling procedure the calibration sample is split in bins of p, η, and n_{Tracks}. The systematic uncertainty is estimated by calculating PID efficiencies using the old binning scheme and the samples with slightly modified PID corrections using coarser and finer binning schemes, where the number of initial bins is varied by roughly a factor of $1/2$ and 2, respectively, with respect to the default binning scheme. The difference in the efficiencies is assigned as a systematic uncertainty. These efficiencies correspond not only to PID$_s$ and PID requirements but also to BDT and BDTHOP requirements, because their efficiencies can also be affected by the choice of binning. The efficiencies are calculated for the signal and normalisation channels

since all of them enter the normalisation factor α. The total efficiency is given by:

$$\varepsilon = \varepsilon_{\text{PID|BDTHOP}} \cdot \varepsilon_{\text{BDTHOP|BDT}} \cdot \varepsilon_{\text{BDT|PIDs}} \cdot \varepsilon_{\text{PIDs|trig}} \cdot \varepsilon_{\text{x}}, \tag{4.25}$$

$$\varepsilon^{\text{F/C}} = \varepsilon^{\text{F/C}}_{\text{PID|BDTHOP}} \cdot \varepsilon^{\text{F/C}}_{\text{BDTHOP|BDT}} \cdot \varepsilon^{\text{F/C}}_{\text{BDT|PIDs}} \cdot \varepsilon^{\text{F/C}}_{\text{PIDs|trig}} \cdot \varepsilon_{\text{x}}, \tag{4.26}$$

where all the efficiencies prior to the PID resampling are gathered in ε_{x}. As a result, the total efficiency associated with the Fine (F) and Coarse (C) binning scheme is:

$$\varepsilon^{\text{F/C}} = \varepsilon \cdot \frac{\varepsilon^{\text{F/C}}_{\text{PID|BDTHOP}} \cdot \varepsilon^{\text{F/C}}_{\text{BDTHOP|BDT}} \cdot \varepsilon^{\text{F/C}}_{\text{BDT|PIDs}} \cdot \varepsilon^{\text{F/C}}_{\text{PIDs|trig}}}{\varepsilon_{\text{PID|BDTHOP}} \cdot \varepsilon_{\text{BDTHOP|BDT}} \cdot \varepsilon_{\text{BDT|PIDs}} \cdot \varepsilon_{\text{PIDs|trig}}} \tag{4.27}$$

$$= \varepsilon \cdot f^{\text{F/C}}_{\text{PID}} \cdot f^{\text{F/C}}_{\text{BDTHOP}} \cdot f^{\text{F/C}}_{\text{BDT}} \cdot f^{\text{F/C}}_{\text{PIDs}}, \tag{4.28}$$

where $f^{\text{F/C}}_{\text{req.}}$ is the ratio of the requirement's efficiency with Fine or Coarse binning over its original value. The change in the normalisation factor is estimated with the formula:

$$\Delta\alpha^{\text{F/C}} = \frac{|\alpha^{\text{F/C}} - \alpha|}{\alpha}. \tag{4.29}$$

The calculated normalisation factors for the full dataset are given in Tab. 4.22. The highest change (6.6 % for $B^+ \to K^+\mu^-e^+$ and 4.8 % for $B^+ \to K^+\mu^+e^-$) in the normalisation factors is assigned as a systematic uncertainty.

Table 4.22: Normalisation factors for the original and modified binning schemes and the corresponding systematic uncertainty. The values are calculated for the full dataset and extracted from simulations.

	$B^+ \to K^+\mu^-e^+$	$B^+ \to K^+\mu^+e^-$
$\alpha \times 10^9$	2.27 ± 0.16	1.66 ± 0.12
$\alpha^{\text{F}} \times 10^9$	2.42 ± 0.19	1.64 ± 0.11
$\alpha^{\text{C}} \times 10^9$	2.35 ± 0.18	1.74 ± 0.13
Systematic uncertainty [%]		
$\Delta\alpha^{\text{F}}$	6.6	1.2
$\Delta\alpha^{\text{C}}$	3.5	4.8

A systematic uncertainty of 0.1% due to the sWeighting of the kaon and muon calibration samples is assigned as suggested by the PID Calib group.

4.11.3 Fitting model uncertainty

Using the bootstrapping method [63] another signal parametrization is predicted. With this method new samples are obtained, each produced by fluctuating every event in

the original sample by a Poisson distribution with ($\mu = 1$). New Crystal Ball fits are obtained based on these samples. The difference in the mass requirements efficiencies between this parametrisation and the original fit is a systematic uncertainty.

The mass requirement efficiencies calculated with the fit results on the bootstrapped data and MC are given in Tab. 4.23. The mass requirement efficiency from the original fit is repeated here for convenience. The corresponding systematic uncertainty is equal to 0.21 %.

Table 4.23: Mass requirement efficiencies and corresponding uncertainty.

	Brem0	Brem1	Combined
Bootstraped	0.9067 ± 0.0107	0.8918 ± 0.0132	0.8981 ± 0.0089
Original	0.9187 ± 0.0257	0.8861 ± 0.0202	0.90001 ± 0.01595

4.11.4 Normalisation uncertainty

The systematic uncertainty on the normalisation constant comprises effects of the efficiencies, yields and the normalisation branching fraction, as given in Tables 4.18 and 4.19. For the 2011 and 2012 datasets, the control channel branching fraction gives the highest systematic uncertainty $\mathscr{B}(B^+ \to K^+ J/\psi(\to \mu^+\mu^-)) = (6.11 \pm 0.19) \times 10^{-5}$, and its relative value is 3.11%. The total normalisation uncertainties are 7.6 % and 7.1 % for $B^+ \to K^+\mu^+e^-$ and $B^+ \to K^+\mu^-e^+$, respectively.

4.11.5 Background uncertainty

As the yields of exclusive background sources have been found to be negligible (see Sec. 4.8), they are not taken into account in the limit evaluation. The limit evaluation is performed by extrapolating the exponential background distribution from the data sidebands into the signal region. The fit takes into account partially reconstructed backgrounds (if present, they would extend into the sidebands). The systematic uncertainty on this approach is evaluated by comparing the difference in the expected number of events taken from the original background fit with the expected number of events calculated by fitting the data sidebands with a first order Chebychev polynomial. The corresponding systematic is estimated to be 1.67 and 0.68, for $B^+ \to K^+\mu^+e^-$ and $B^+ \to K^+\mu^-e^+$ respectively.

4.11.6 PHSP model uncertainty

PHSP is used as a model for the simulation, because the possible underlying NP models are not known. Instead of the systematic uncertainty, the efficiency maps are provided (see Fig. 4.26). Thanks to that, the interpretations in terms of other models are possible.

As can be seen, there is no visible pattern in these distributions. These maps are mostly flat, with the small exception in the lower values of the $m_{K^+e^+}$.

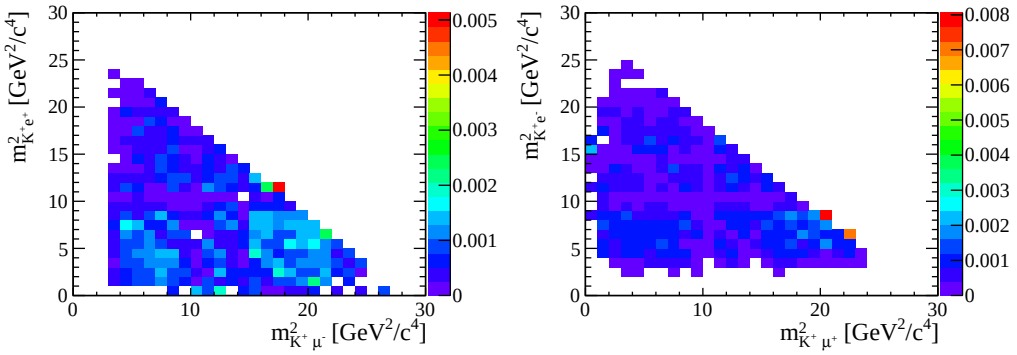


Figure 4.26: Dalitz plot of total efficiencies for the $B^+ \to K^+\mu^-e^+$ (left) and $B^+ \to K^+\mu^+e^-$ (right).

4.11.7 Systematic uncertainty summary

The total systematic uncertainty and all contributions are listed in Table 4.24.

Table 4.24: Systematics summary. These values are in % unless stated otherwise.

Effect	$B^+ \to K^+\mu^+e^-$	$B^+ \to K^+\mu^-e^+$
Data-MC corrections	1.0	1.0
PID resampling - binning	4.8	6.6
PID resampling - sWeighting	0.1	0.1
Fitting model	0.21	0.21
Normalisation	7.6	7.1
Background (not in %)	1.7	0.7

The systematic uncertainty of this analysis is not comparable in size with the statistical uncertainties of the expected number of background events / signal yield.

4.12 Results

Special permission has been given by the LHCb collaboration to unblind the signal region for the purpose of this thesis. As a result, no significant excess of events is observed. The upper limits on signal branching fractions are calculated using the statistical CL_s method (described in Sec. 4.7.1). Since the CL_s value is calculated as $CL_s = CL_{s+b}/CL_b$, the knowledge of the expected number of events for the signal s and the background b is necessary, as well as the observed number of events.

4.12.1 Expected upper limits

The expected CL_s values as a function of possible branching fractions are illustrated with dashed lines in Fig. 4.28. The computation is performed under the hypothesis of the observation of background only events. Prior to the unblinding, the observed number of events is unknown. The expected number of background events is taken from the sideband fit to the $m_{Ke\mu}$. The number of signal events depends on the assumed branching fraction (the X-axis) and on the calculated normalisation factors (see Tab. 4.20). The expected upper limit, prior to unblinding, at 90 % C.L. is expected to be:

- $\mathscr{B}(B^+ \to K^+\mu^+e^-) < 11.8 \times 10^{-9}$,
- $\mathscr{B}(B^+ \to K^+\mu^-e^+) < 6.1 \times 10^{-9}$.

4.12.2 Observed upper limits

The estimation of the expected signal events has been performed in Sec. 4.8.3. After unblinding, the upper limits are recalculated using systematic uncertainties and the number of events inside the signal region. The unblinded mass distributions of signal candidates can be seen in Fig. 4.27.

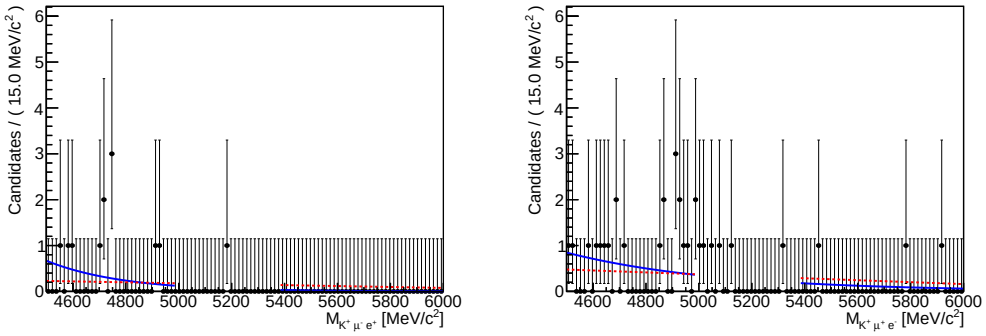


Figure 4.27: Distributions of $m_{Ke\mu}$ after the selection, for Run1 data. The exponential (purple line) is performed on data sidebands. The red line corresponds to the Chebychev polynomial used for the background uncertainty calculation.

No excess of signal events has been observed. The observed and expected (repeated here for the convenience) numbers of events are reported in Tab. 4.25.

Table 4.25: Expected and observed numbers of events in the signal region.

Decay channel	Expected	Observed
$B^+ \to K^+\mu^+e^-$	4.26 ± 1.16	2
$B^+ \to K^+\mu^-e^+$	0.98 ± 0.68	1

The upper limits on branching fractions of both signal channels at the 90 % C.L. are found to be:

- $\mathscr{B}(B^+ \to K^+\mu^+e^-) < 8.3 \times 10^{-9}$,
- $\mathscr{B}(B^+ \to K^+\mu^-e^+) < 6.1 \times 10^{-9}$.

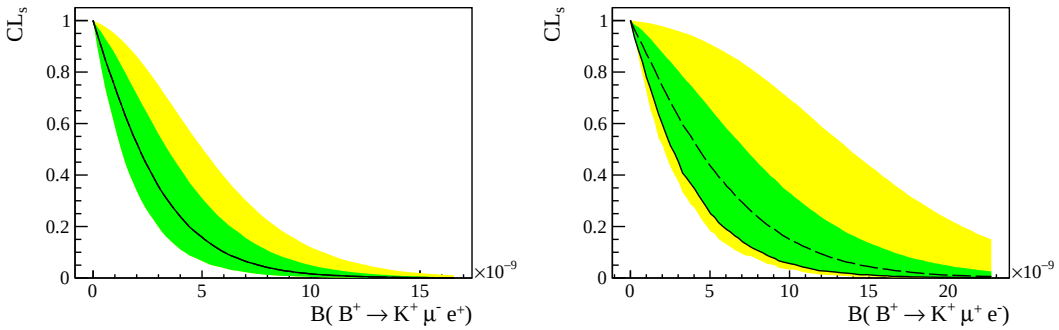


Figure 4.28: Expected (dashed line) and observed (solid line) CL_s curves with 68% (yellow) and 90 % (green) containment bands. The left plot corresponds to the $B^+ \to K^+\mu^-e^+$ signal channel and the right one corresponds to the $B^+ \to K^+\mu^+e^-$ signal channel.

These limits are calculated using MC samples, which were simulated using the PHSP model.

CHAPTER 5

Summary

The presented thesis is focused on the searches for the lepton flavour violating decay channels $B^+ \to K^+\mu^-e^+$ and $B^+ \to K^+\mu^+e^-$. The studies were performed using the dataset of proton - proton LHC collisions gathered by the LHCb experiment in the years 2011 and 2012. Searches for very rare decay channels are one of the main tools to search for the New Physics. No statistically significant signal has been observed for both searched channels. As a result, the upper limits on their branching fractions were set to be:

- $\mathscr{B}(B^+ \to K^+\mu^+e^-) < 8.3 \times 10^{-9}$,
- $\mathscr{B}(B^+ \to K^+\mu^-e^+) < 6.1 \times 10^{-9}$.

at 90 % CL.

The signal selection process was prepared and implemented, as well as for the control channel's $B^+ \to K^+J/\psi\,(\to \mu^+\mu^-)$ and $B^+ \to K^+J/\psi\,(\to e^+e^-)$. The strongest selection requirement was based on two multivariate classifiers. The first classifier was trained to discriminate against the combinatorial background, and the second one was specialised in the partially reconstructed background. The selection process efficiently rejected all possible exclusive background sources, which has been proved by dedicated studies with their MC samples.

Bibliography

[1] S.Weinberg, *Conceptual foundations of the unified theory of weak and electromagnetic interactions*, Rev. Mod. Phys. 52, 515 (1980).

[2] The CMS Collaboration. *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:hep-ex/1207.7235.

[3] The ATLAS Collaboration. *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) 1-29, arXiv:hep-ex/1207.7214.

[4] D. Strottman, *Multi-quark baryons and the MIT bag model*, Phys. Rev. D20 (1979) 748.

[5] Hogaasen, H. and Sorba, P. , *The systematics of possibly narrow quark states with baryon number one*, Nucl. Phys. B145 (1978) 119.

[6] R. L. Jaffe, *Multiquark hadrons. I. Phenomenology of $Q^2\bar{Q}^2$ mesons*, Phys. Rev. D15 (1977) 267.

[7] C. Patrignani et al. (Particle Data Group), *The Review of Particle Physics (2017)*, Chin. Phys. C, 40, 100001 (2016) and 2017 update.

[8] F. Englert and R. Brout, *Broken symmetry and the mass of gauge vector mesons*, Phys. Rev. Lett. 13 (1964) 321.

[9] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, Phys. Lett. 12 (1964) 132.

[10] Wu, C. S.; Ambler, E.; Hayward, R. W.; Hoppes, D. D.; Hudson, R. P. *Experimental Test of Parity Conservation in Beta Decay*, Physical Review. 105 (4): 1413–1415 (1957).

[11] Boger et al., *The Sudbury Neutrino Observatory*, Nucl. Instrum. Methods A449, 172 (2000), arXiv:nucl-ex/9910016.

[12] I. Esteban, M. C. Gonzalez-Garcia, M. Maltoni, I. Martinez-Soler, T. Schwetz, *Updated fit to three neutrino mixing: exploring the accelerator-reactor complementarity*, JHEP 01 (2017) 087, arXiv:hep-ph/1611.01514.

[13] A. J. Buras, *Weak Hamiltonian, CP violation and rare decays, in Probing the standard model of particle interactions. Proceedings, Summer School in Theoretical Physics, NATO Advanced Study Institute, 68th session, Les Houches, France, July 28-September 5*, 1997. Pt. 1, 2, pp. 281–539, 1998.

[14] C. Bobeth, G. Hiller, G. Piranishvili *Angular Distributions of $B \to K\, ll$ Decays*, JHEP 12 (2007) 040, arXiv:hep-ph/0709.4174.

[15] G. Hiller, F. Kruger, *More model-independent analysis of $b \to s$ processes*, Phys. Rev. D69 (2004) 074020, arXiv:hep-ph/0310219.

[16] C. Bouchard et al., *Standard Model predictions for $B \to K\ell^+\ell^-$ with form factors from lattice QCD*, Phys. Rev. Lett. 111 (2013) 162002, arXiv:hep-ph/1306.0434.

[17] S. Fajfer, J. F. Kamenik, I. Nisandzic, *On the B to D* tau nu Sensitivity to New Physics*, Phys. Rev. D 85 (2012) 094025, arXiv:hep-ph/1203.2654 .

[18] R. Aaij *et al.* (LHCb Collaboration), *Test of Lepton Universality $B^+ \to K^+\ell\ell$ Using Decays*, Phys. Rev. Lett. 113 (2014) 151601, arXiv:hep-ex/1406.6482 .

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

[20] R. Aaij *et al.* (LHCb Collaboration), *Measurement of the ratio of branching fractions $\mathscr{B}(\bar{B}^0 \to D^{*+}\tau^-\nu_\tau)/\mathscr{B}(\bar{B}^0 \to D^{*+}\mu^-\nu_\mu)$*, Phys. Rev. Lett. 115, 111803 (2015), arXiv:hep-ex/1506.08614.

[21] LHCb collaboration, *Angular analysis and differential branching fraction of the decay $B_s^0 \to \phi\mu^+\mu^-$*, JHEP 09 (2015) 179, arXiv:hep-ex/1506.08777.

[22] LHCb collaboration, *Measurements of the S-wave fraction in $B^0 \to K^+\pi^-\mu^+\mu^-$ decays and the $B^0 \to K^{*0}(892)\mu^+\mu^-$ differential branching fraction*, JHEP 11 (2016) 047, JHEP 04 (2017) 142, arXiv:hep-ex/1606.04731.

[23] LHCb collaboration, *Differential branching fraction and angular analysis of $\Lambda_b^0 \to \Lambda\mu\mu$*, JHEP 06 (2015) 115, arXiv:hep-ex/1503.07138.

[24] J. Matias, F. Mescia, M. Ramon, J. Virto, *Complete Anatomy of $B \to K^* ll$ and its angular distribution*, JHEP 04 (2012) 104, arXiv:hep-ph/1202.4266.

[25] LHCb collaboration, *Angular analysis of the $B^0 \to K^{*0}\mu^+\mu^-$ decay using 3 fb^{-1} of integrated luminosity*, JHEP 02 (2016) 104, arXiv:hep-ex/1512.04442 .

[26] Belle collaboration, *Lepton-Flavor-Dependent Angular Analysis of $B \to K^*\ell^+\ell^-$*, Phys. Rev. Lett 118, 111801 (2017), arXiv:hep-ex/1612.05014.

[27] CMS collaboration, *Measurement of angular parameters from the decay $B^0 \to K^{*0}\mu^+\mu^-$ in proton-proton collisions at $\sqrt{s}$ = 8* TeV, Phys. Lett. B 781 (2018) 517, arXiv:hep-ex/1710.02846.

[28] ATLAS collaboration, *Angular analysis of $B^0 \to K^{*0}\mu^+\mu^-$ decays in pp collisions at $\sqrt{s}$ = 8* TeV *with the ATLAS detector*, CERN-EP-2017-161, arXiv:hep-ex/1805.04000.

[29] T. Blake, G. Lanfranchi, D. M. Straub *Decays as Tests of the Standard Model*, Progress in Particle and Nuclear Physics, Volume 92 (2017), 50-91, arXiv:hep-ph/1606.00916.

[30] O Sumensari, *Leptoquark models for the B-physics anomalies*, arXiv:hep-ph/1705.07591.

[31] S. Descotes-Genon, J. Matias, J. Virto, *Understanding the $B^0 \to K^{*0}\mu^+\mu^-$ Anomaly*, Phys. Rev. D 88, 074002 (2013), arXiv:hep-ph/1307.5683.

[32] S. M. Boucenna, J. W.F. Valle, A. Vicente, *Are the B decay anomalies related to neutrino oscillations?*, Phys. Let. B (2015) 09 040, arXiv:hep-ph/1503.07099.

[33] G. Hiller, D. Loose, K. Schönwald, *Leptoquark Flavor Patterns & B Decay Anomalies*, JHEP 12 (2016) 027, arXiv:hep-ph/1609.08895.

[34] CMS. cms.cern/detector

[35] ATLAS. atlas.ch/

[36] LHCb. lhcb-public.web.cern.ch/lhcb-public/

[37] ALICE. aliceinfo.cern.ch/Public/Welcome.html

[38] http://te-epc-lpc.web.cern.ch/te-epc-lpc/machines/lhc/general.stm

[39] R. Aaij *et al.* (LHCb Collaboration), *LHCb Detector Performance*, Int. J. Mod. Phys. A 30, 1530022 (2015), arXiv:hep-ex/1412.6352.

[40] R. Aaij *et al.* (LHCb Collaboration), *The LHCb Detector at the LHC*, JINST 3 (2008) S08005.

[41] F. Archilli, W. Baldini, G. Bencivenni, N. Bondar, W. Bonivento, S. Cadeddu, P. Campana, A. Cardini, P. Ciambrone, X. Cid Vidal, C. Deplano, P. De Simone, A. Falabella, M. Frosini, S. Furcas, E. Furfaro, M. Gandelman, J.A. Hernando Morata, G. Graziani, A. Lai, G. Lanfranchi, J.H. Lopes, O. Maev, G. Manca, G. Martellotti, A. Massafferri, D. Milanes, R. Oldeman, M. Palutan, G. Passaleva, D. Pinci, E. Polycarpo, R. Santacesaria, E. Santovetti, A. Sarti, A. Satta, B. Schmidt, B. Sciascia, F. Soomro, A. Sciubba, S. Vecchi, *Performance of the Muon Identification at LHCb*, JINST 8 (2013) P10020, arXiv:physics.ins-det/1306.0249.

[42] G. Passaleva, *A recurrent neural network for track reconstruction in the LHCb Muon System*, Nuclear Science Symposium Conference Record NSS '08. IEEE. (2008).

[43] BRUNEL, lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/brunel

[44] DAVINCI, lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/davinci

[45] GAUSS, lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/gauss/

[46] T. Sjostrand, S. Mrenna, P. Skands, *A Brief Introduction to* PYTHIA *8.1*, Comput.Phys.Commun.178:852-867,2008

[47] A. Ryd, D. Lange, N. Kuznetsova, S. Versille, M. Rotondo, D. P. Kirkby, F. K. Wuerthwein, A. Ishikawa *EvtGen: A Monte Carlo Generator for B-Physics*, EVTGEN-V00-11-07.

[48] P. Golonka, Z. Was, *PHOTOS Monte Carlo: a precision tool for QED corrections in Z and W decays*, Eur.Phys.J.C45:97-107 (2006), arXiv:hep-ph/0506026.

[49] https://twiki.cern.ch/twiki/bin/view/LHCb/VertexFitters

[50] T. Head, *The LHCb trigger system*, JINST 9 (2014) C09015.

[51] M. Pivk, F. R. Le Diberder, *sPlot: a statistical tool to unfold data distributions*, Nucl. Instrum. Meth. A555: 356-369 (2005), arXiv:physics/0402083.

[52] T. Skwarnicki, *A study of the radiative CASCADE transitions between the Upsilon-Prime and Upsilon resonances*, DESY-F31-86-02, DESY-F-31-86-02.

[53] L. Anderlini et al., *The PIDCalib package*, LHCb-PUB-2016-021, CERN, Geneva, Jul, 2016.

[54] C. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[55] A. Blum, A. Kalai, and J. Langford, *Beating the hold-out: Bounds for k-fold and progressive cross-validation*, Proceedings of the Twelfth Annual Conference on Computational Learning Theory, COLT '99, (New York, NY, USA), pp. 203–208, ACM, 1999.

[56] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and regression trees, Wadsworth international group*, Belmont, California, USA, 1984.

[57] Hocker, A. , Stelzer, J. , Tegenfeldt, F. , Voss, H. , Voss, K. and others, *TMVA - Toolkit for Multivariate Data Analysis*, PoS ACAT (2007) 040.

[58] LHCb collaboration, *First evidence for the decay* $B_s^0 \to \mu^+\mu^-$, Phys. Rev. Lett. 110, 021801 (2013), arXiv:hep-ex/1211.2674.

[59] A. L. Read, *Modified frequentist analysis of search results (the* CL_s *method)*, 10.5170/CERN-2000-005.81.

[60] G. Punzi, *Sensitivity of searches for new signals and its optimization*, arXiv:physics/0308063.

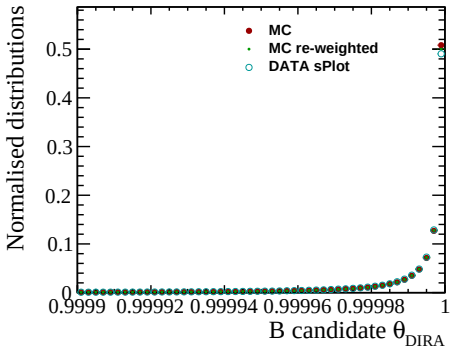
[61] Y. Amhis et al., *Averages of b-hadron, c-hadron, and* τ *-lepton properties as of summer 2016*, Eur. Phys. J. C77 (2017) 895, arXiv:1612.07233.

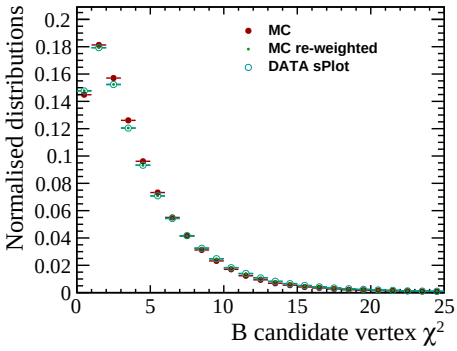
[62] Simulation Statistics, http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/STATISTICS/

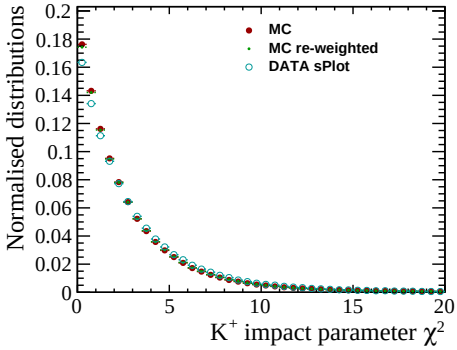
[63] B. Efron, *Bootstrap methods: Another look at the jackknife*, Ann. Statist. 7 (1979) 1.

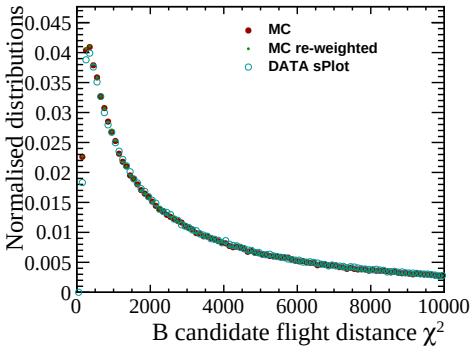
APPENDIX A

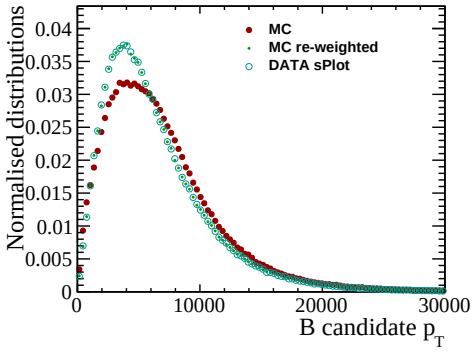
Data-MC comparison

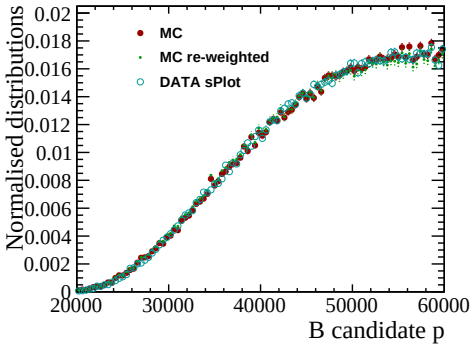


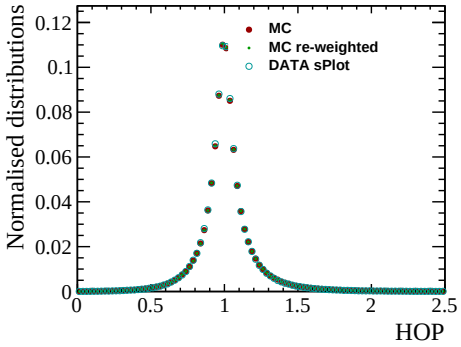


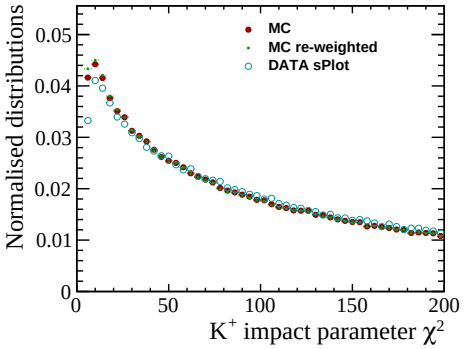


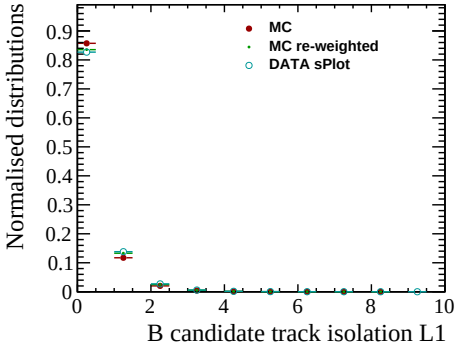


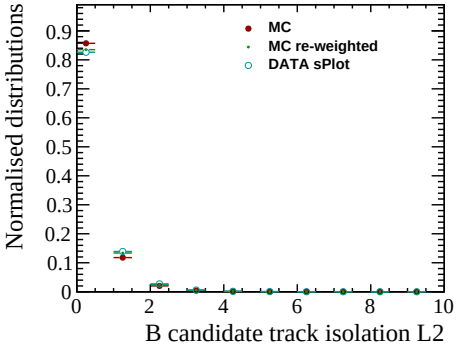

MC
MC re-weighted
DATA sPlot
Normalised distributions
B candidate p_T

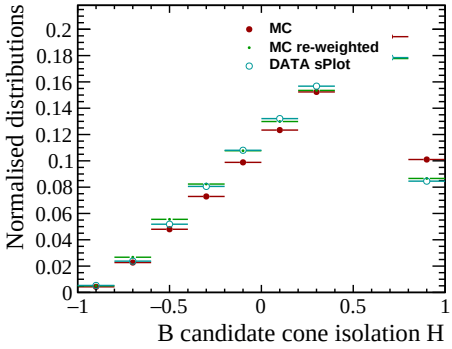

MC
MC re-weighted
DATA sPlot
Normalised distributions
B candidate p


MC
MC re-weighted
DATA sPlot
Normalised distributions
HOP


MC
MC re-weighted
DATA sPlot
Normalised distributions
K^+ impact parameter χ^2


MC
MC re-weighted
DATA sPlot
Normalised distributions
B candidate track isolation L1

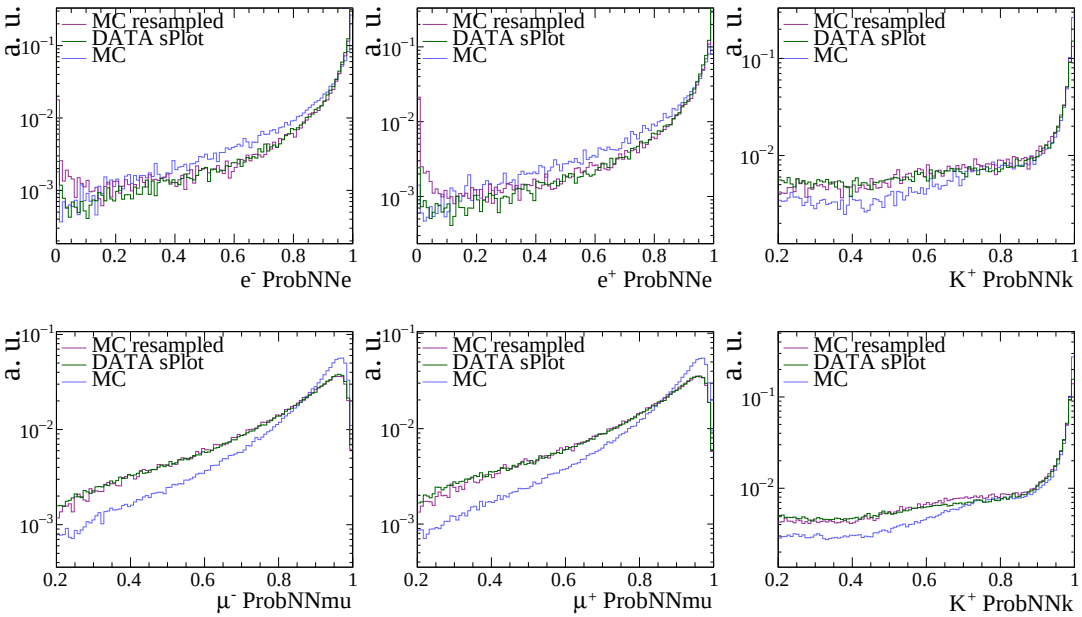

MC
MC re-weighted
DATA sPlot
Normalised distributions
B candidate track isolation L2


MC
MC re-weighted
DATA sPlot
Normalised distributions
B candidate cone isolation H

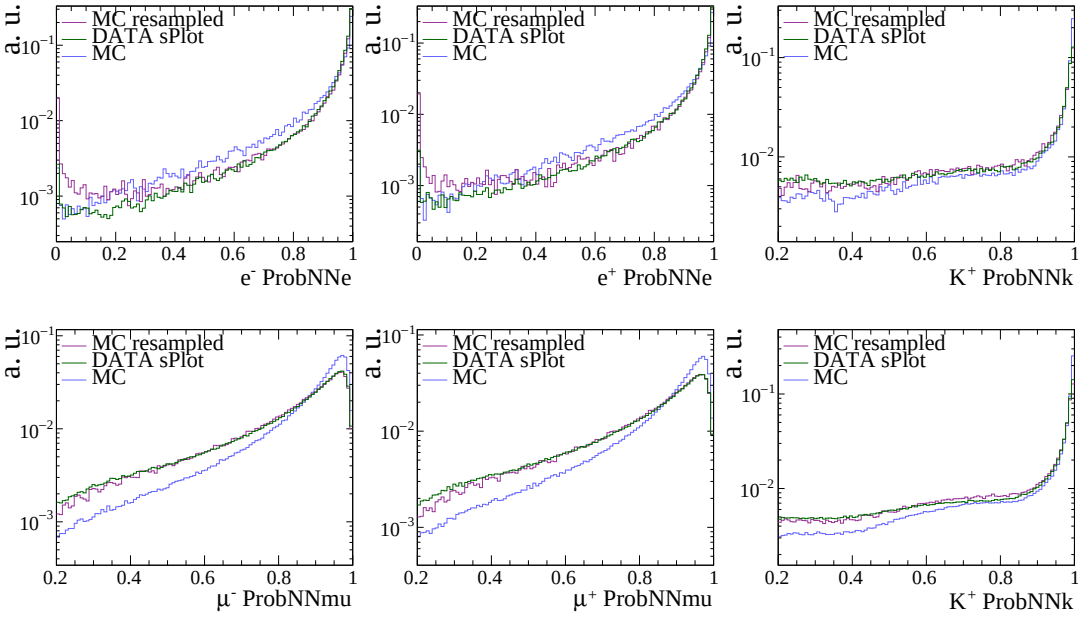
APPENDIX B

Resampled PID distributions

2011 samples:

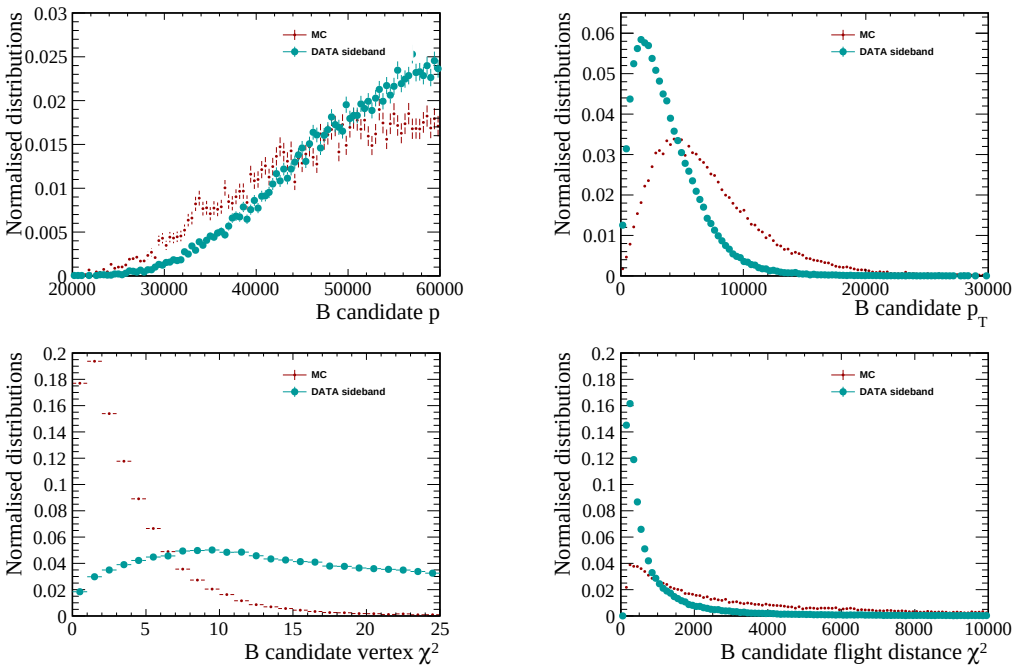


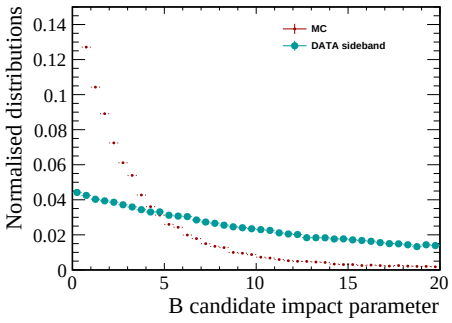
2012 samples:

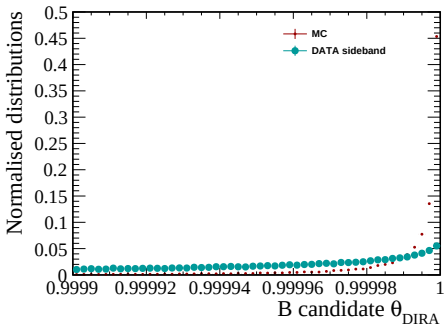


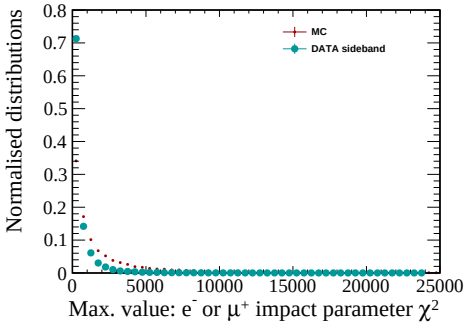
APPENDIX C

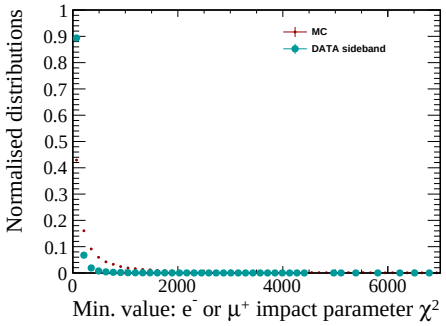
BDT training inputs

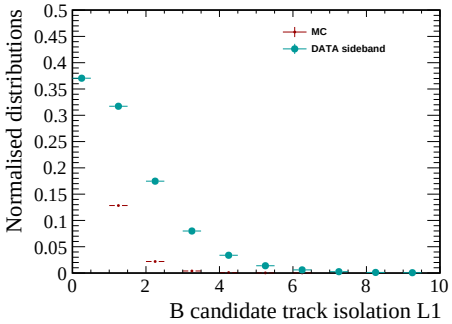


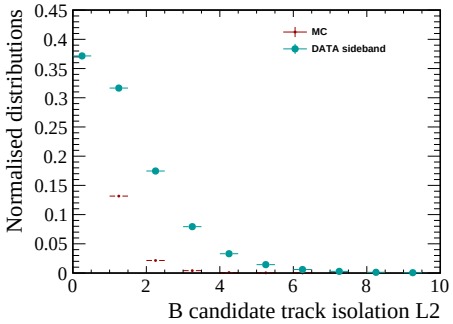

Normalised distributions
MC
DATA sideband
B candidate impact parameter

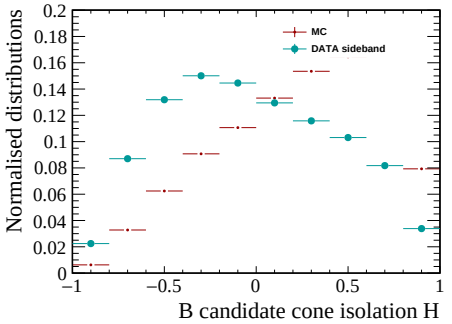

Normalised distributions
MC
DATA sideband
B candidate θDIRA

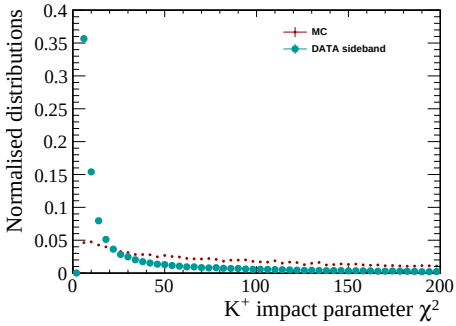

Normalised distributions
MC
DATA sideband
Max. value: e- or μ+ impact parameter χ2


Normalised distributions
MC
DATA sideband
Min. value: e- or μ+ impact parameter χ2


Normalised distributions
MC
DATA sideband
B candidate track isolation L1

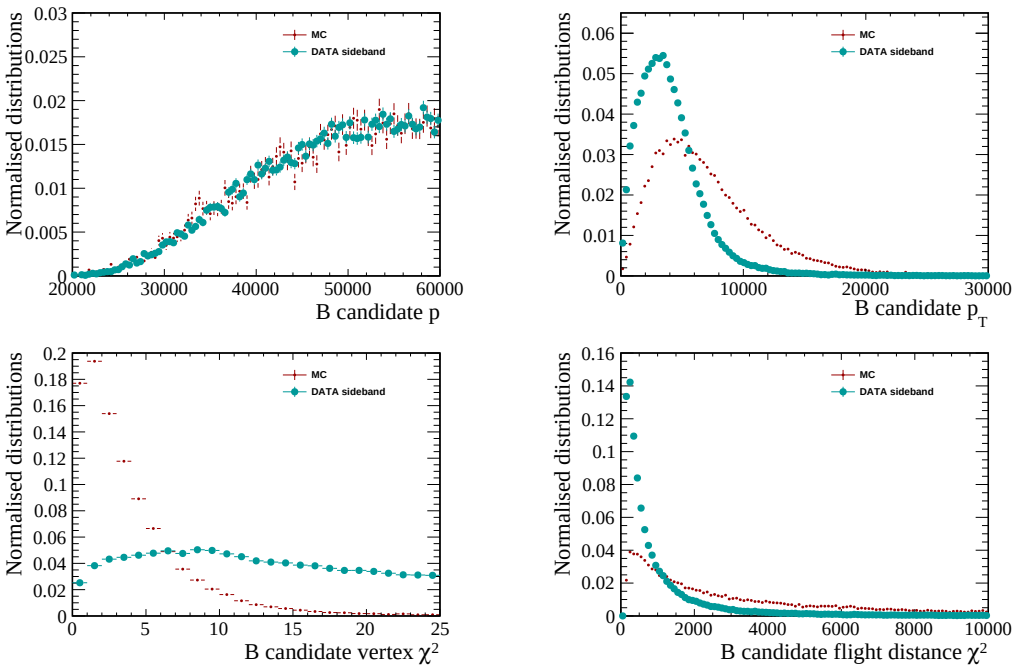

Normalised distributions
MC
DATA sideband
B candidate track isolation L2

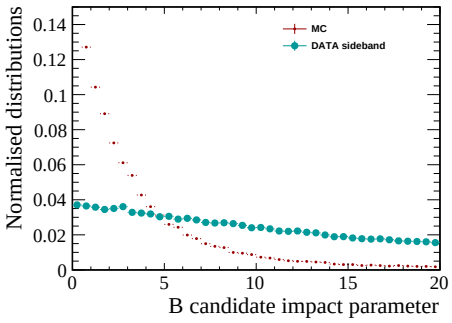

Normalised distributions
MC
DATA sideband
B candidate cone isolation H

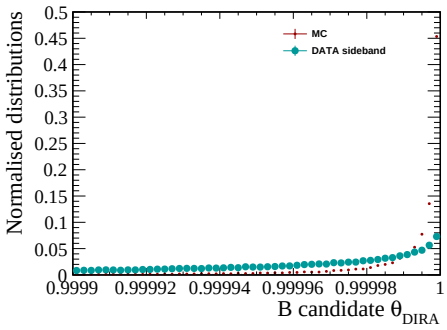

Normalised distributions
MC
DATA sideband
K+ impact parameter χ2

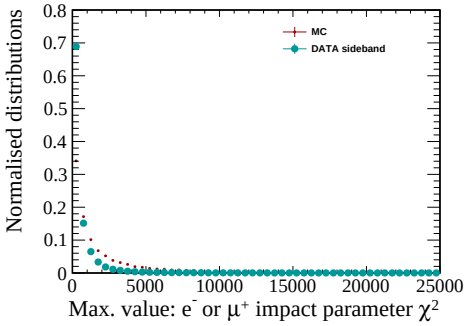
APPENDIX D

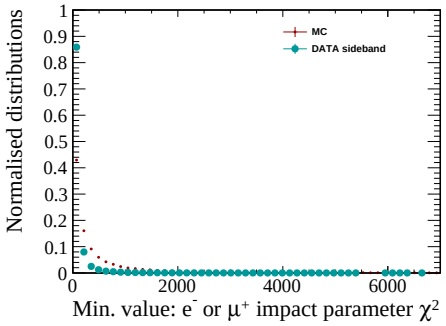
BDTHOP training inputs

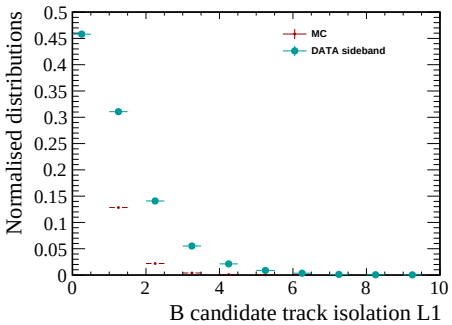


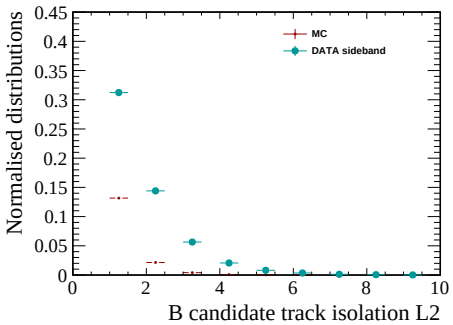

Normalised distributions
MC
DATA sideband
B candidate impact parameter

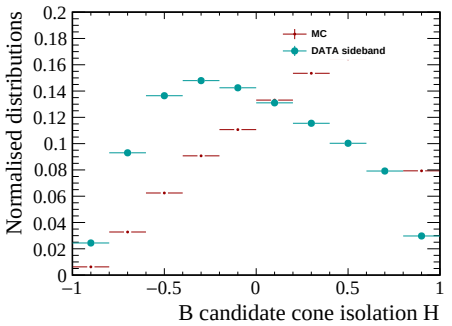

Normalised distributions
MC
DATA sideband
B candidate θDIRA

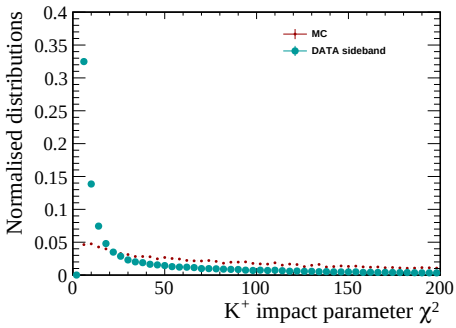

Normalised distributions
MC
DATA sideband
Max. value: e- or μ+ impact parameter χ2

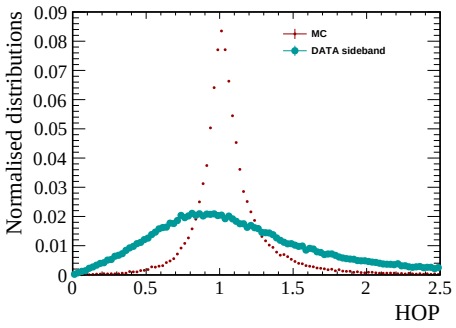

Normalised distributions
MC
DATA sideband
Min. value: e- or μ+ impact parameter χ2


Normalised distributions
MC
DATA sideband
B candidate track isolation L1


Normalised distributions
MC
DATA sideband
B candidate track isolation L2


Normalised distributions
MC
DATA sideband
B candidate cone isolation H


Normalised distributions
MC
DATA sideband
K+ impact parameter χ2


Normalised distributions
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
0
0.5
1
1.5
2
2.5
HOP
MC
DATA sideband