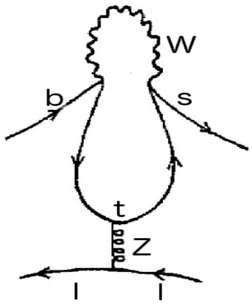
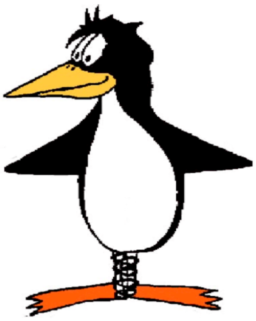




Investigating the flavour anomalies through the rare beauty decay $B^0 \to K^{*0}\mu^+\mu^-$

Maria Carolina Feliciano Faria

Thesis to obtain the Master of Science Degree in

Engineering Physics

Supervisors: Prof. Dr. Nuno Teotónio Viegas Guerreiro Leonardo
Dr. Alessio Boletti

Examination Committee

Chairperson: Prof. Dr. Mário João Martins Pimenta
Supervisor: Prof. Dr. Nuno Teotónio Viegas Guerreiro Leonardo
Members of the Committee: Dr. Sara Fiorendi
Dr. Jonathan Jason Hollar

October 2021

Acknowledgments

I would like to thank my supervisors Prof. Nuno Leornardo and Dr. Alessio Boletti for all the support, ideas and dedication they have given me during the course of these months. I am lucky to have had you both as supervisors. A special thanks to Nuno for all the things he has taught me ever since I firstly did my summer internship in LIP in 2018 and for having read every detail of this thesis. I thank the members of the LIP CMS group for having listened to me every Wednesday morning as I gave them the weekly update of my work. I am also thankful to our collaborators in the $B^0 \to K^{*0}\mu^+\mu^-$ angular analysis, among them Dr. Alessio Boletti, Dr. Sara Fiorendi and Dr. Paolo Dini, for their nice comments and suggestions and for having paved the road for the work of this thesis.

I would like to thank my parents, without whom I would not have been able to do this thesis. I thank them for all the equally important things they have taught me, such as the nobility of hard work.

Finally, I would like to thank my friends. Thank you Ana, for always being by my side ever since the 1st grade and for always being there for me when I needed your help. Thank you Vladlen, for having shared my joys and worries over the course of these two years in MEFT. I thank also my friends in Porto, who helped me to overcome the harder times of those three years in physics and to have fun outside the university. A special thanks to my lab partner João.

Thank you!

Resumo

Medimos a probabilidade de ocurrência do decaimento $B^0 \to K^{*0}\mu^+\mu^-$ (*branching fraction*, BF) em função do quadrado da massa invariante do par de muões, usando o canal $B^0 \to J/\psi K^{*0}$ como normalização. Usamos dados de colisões protão-protão a $\sqrt{s} = 13\,$TeV, que foram adquiridos com a experiência Compact Muon Solenoid (CMS) no Large Hadron Collider (LHC) durante os anos 2016, 2017 e 2018. A análise é validada, comparando a razão das BFs dos dois canais ressonantes, $B^0 \to J/\psi K^{*0}$ e $B^0 \to \psi(2S)K^{*0}$, com a atual média mundial. O nosso resultado é consistente com este valor dentro de $0.2\,\sigma$. Os resultados para a BF do decaimento $B^0 \to K^{*0}\mu^+\mu^-$ são também consistentes com as medições anteriores realizadas com dados da Run 1 do LHC e com as previsões do modelo padrão. Estes resultados preliminares são os mais precisos até ao momento.

O decaimento $B^0 \to K^{*0}\mu^+\mu^-$ é um processo belo e raro que no modelo padrão só pode proceder através de diagramas de Feynman de ordem superior, o que o torna muito sensível a nova física. Tensões com o modelo padrão foram já reportadas num dos parâmetros que entra na expressão da sua taxa de decaimento, o P_5', que pertence a uma classe de discrepâncias no setor do sabor comumente referidas como anomalias de sabor. A BF do decaimento $B^0 \to K^{*0}\mu^+\mu^-$ pode entrar em ajustes globais que permitem restringir cenários de nova física e investigar a origem destas aliciantes anomalias de sabor.

Palavras-Chave

Anomalias de sabor, decaimentos raros, experiência CMS, nova física, física do mesão B, LHC

Abstract

We measure the $B^0 \to K^{*0}\mu^+\mu^-$ decay branching fraction as a function of the di-muon invariant mass squared, using the $B^0 \to J/\psi K^{*0}$ decay as normalisation channel. We use proton-proton collision data at $\sqrt{s} = 13$ TeV, collected by the Compact Muon Solenoid (CMS) experiment at the Large Hadron Collider (LHC) during the years 2016, 2017 and 2018. The analysis is validated by comparing the ratio between the branching fractions of the two resonant channels, $B^0 \to J/\psi K^{*0}$ and $B^0 \to \psi(2S)K^{*0}$, with the current world average. Our result is consistent with this value within $0.2\,\sigma$. The $B^0 \to K^{*0}\mu^+\mu^-$ branching fraction results are also consistent with previous LHC Run 1 measurements and with the Standard Model (SM) predictions. These preliminary results are the most precise to date.

The $B^0 \to K^{*0}\mu^+\mu^-$ decay is a rare beauty process which can only proceed at loop order in the SM, making it very sensitive to new physics (NP). Tensions with the SM have already been reported in one of the parameters appearing in its decay rate, P_5', belonging to the class of discrepancies in the flavour sector commonly referred to as flavour anomalies. The $B^0 \to K^{*0}\mu^+\mu^-$ branching fraction can enter in global fits to help constraining NP scenarios and investigating the source of these alluring flavour anomalies.

Keywords

Flavour anomalies, rare decays, CMS experiment, new physics, B physics, LHC

Contents

List of Figures

List of Tables

Acronyms

BDT	Boosted Decision Tree
CKM	Cabibbo–Kobayashi–Maskawa
C	charge conjugation
CP	charge parity
CTF	Combinatorial Track Finder
CMS	Compact Muon Solenoid
DCA	distance of closest approach
EFT	Effective Field Theories
ECAL	Electromagnetic Calorimeter
EMTF	Endcap Muon Track Finder
FCNC	Flavour Changing Neutral Current
FCC	Future Circular Collider
HCAL	Hadronic Calorimeter
HEP	High Energy Physics
HIP	High Ionising Particles
HLT	High Level Trigger
HL	High Luminosity
LEP	Large Electron-Positron
LHC	Large Hadron Collider

LFU	Lepton Flavour Universality
LFUV	Lepton Flavour Universality Violation
LQs	leptoquarks
L1	level 1
LCSRs	Light Cone Sum Rules
MC	Monte Carlo
NP	new physics
P	parity
PDG	Particle Data Group
PDF	Probability Density Function
pp	proton-proton
QCDF	QCD Factorisation
QCD	Quantum Chromodynamics
QED	Quantum Electrodynamics
RT	right-tagged
SSB	Spontaneous Symmetry Breaking
SM	Standard Model
VEV	Vacuum Expectaction Value
WC	Wilson Coefficients
WT	wrong-tagged

Chapter 1

Introduction

The Standard Model (SM) describes the elementary particles as well as the electromagnetic, weak and strong interactions between them in terms of fields. Ever since its development in the latter half of the 20th century, it has met great successes. Among them are the prediction of many new particles, such as the W and Z bosons, the gluon and several quark flavours, which culminated with the discovery of the Higgs boson in 2012. Another incredible success is the precision with which the magnetic moment of the electron is known: the experimental measurement agrees with the theoretical prediction up to 10 significant figures; making it the most accurately verified prediction in the history of physics.

Notwithstanding its many accomplishments, the SM has also several shortcomings. The theory does not accommodate gravity, one of the four fundamental interactions of nature. It does not provide a mechanism to explain the existence of neutrino masses, which we now know is the case due to neutrino oscillations [17]. There is no symmetry in the SM that protects the Higgs mass from being much smaller or larger than its experimental value of 125 GeV (hierarchy problem). It does not explain the origin of the matter-antimatter asymmetry, which could be connected to the charge parity (CP) violation present in weak interactions. And it only describes around 5% of the energy and matter content of the universe.

As a consequence, searches for physics beyond the SM are well motivated and are one of the main driving forces for the development of particle accelerators, such as the Large Hadron Collider (LHC) at CERN and the Future Circular Collider (FCC). Searches for new physics (NP) fall into two complementary approaches: direct and indirect searches. The former occurs at the high-energy frontier, where NP particles are detected as excesses in distributions (e.g. a bump in the mass spectrum). The latter occurs at the high-precision frontier, where NP particles are detected by their effect in several particle properties, such as cross-sections, decay rates and branching fractions.

At colliders, the high-energy frontier is reached by increasing the centre-of-mass energy of the collisions, whereas the high-precision frontier is reached by increasing the luminosity. The LHC has started operating in 2011 and has since then undergone major upgrades envisaging the increase of both energy and luminosity. Since the start of its operation, it has already had two data-taking periods: Run 1 (2011-2012) and Run 2 (2015-2018). During Run 1, it achieved centre-of-mass energies of 7 and 8 TeV and in Run 2, it increased this energy up to 13 TeV with significant increases also in luminosity. Run 3 is foreseen to start in 2022, and to arrive at energies of 14 TeV, being the last planned data-taking period of the current LHC. The High Luminosity (HL)-LHC is projected to start its operation in 2027, aiming to arrive at an ultimate integrated luminosity of up to 4000 fb^{-1}. The FCC, a proposed particle accelerator which would integrate both hadron collisions like LHC and electron-positron collisions like the former Large Electron-Positron (LEP) collider, aims to push even further the energy and luminosity frontiers, with the goal of reaching collision energies of 100 TeV.

Thus far, no clear sign of NP has emerged from collider data. Nonetheless, a class of discrepancies with the flavour sector of the SM, pertaining to the category of indirect searches, has been slowly making its stand. These are the so-called **flavour anomalies** and have been detected in two sets of quark level transitions: $b \to sll$ (beauty to strange quark plus pair of charged leptons) and $b \to cl\bar{\nu}$ (beauty to charm quark plus charged lepton and neutrino). The work of this thesis will focus on the former.

1.1 Flavour anomalies in $b \to sll$ transitions and beyond

The Flavour Changing Neutral Current (FCNC) transitions, $b \to sll$, are realised in several rare B meson decays, both leptonic, e.g. $B_s^0 \to \mu^+\mu^-$, and semileptonic, e.g. $B \to X\mu^+\mu^-$, where X stands for a strange hadron (K, K^{*0}, ϕ...). Deviations from the SM expectations have been found in several observables of these decays, namely on branching fractions, such as $\mathcal{B}(B_s^0 \to \phi\mu^+\mu^-)$, angular parameters appearing in the decay rate of $B^0 \to K^{*0}\mu^+\mu^-$, such as P_5', and on ratios of branching fractions, such as $R_{K^{(*)}} = \frac{\mathcal{B}(B^0 \to K^{(*)}\mu^+\mu^-)}{\mathcal{B}(B^0 \to K^{(*)}e^+e^-)}$. The latest results on two of these observables can be seen in Fig. 1.1. In Fig. 1.1 (left), the LHCb result [1] on the differential branching fraction of the $B_s^0 \to \phi\mu^+\mu^-$ decay is shown to have a 3.6 σ deviation with the SM prediction in the q^2 region 1.1-6.0 GeV^2. In Fig. 1.1 (right), the LHCb result [2], superimposed with R_K measurements from Belle [18] and BaBar [19], is shown to have a discrepancy of 3.1 σ with its predicted value of 1.

Even though observables like $\mathcal{B}(B_s^0 \to \phi\mu^+\mu^-)$ and P_5' are prone to theoretical uncertainties coming from Quantum Chromodynamics (QCD) form factor calculations, variables such as $R_{K^{(*)}}$ are very clean, since these uncertainties cancel when doing the ratio, because leptons do not interact via the strong force. This makes the latest LHCb result [2] a very strong hint of NP and, if proven true, it would imply the violation of one of the assumptions of the SM: that the electroweak gauge bosons W and Z have identical couplings to all the three lepton flavours. A state of affairs referred to as Lepton Flavour Universality Violation (LFUV), which would mean that the rates of the decays $B^+ \to K^+\mu^+\mu^-$ and $B^+ \to K^+e^+e^-$ should be equal and their ratio equal to 1, as displayed by the dashed vertical line in Fig. 1.1 (right).

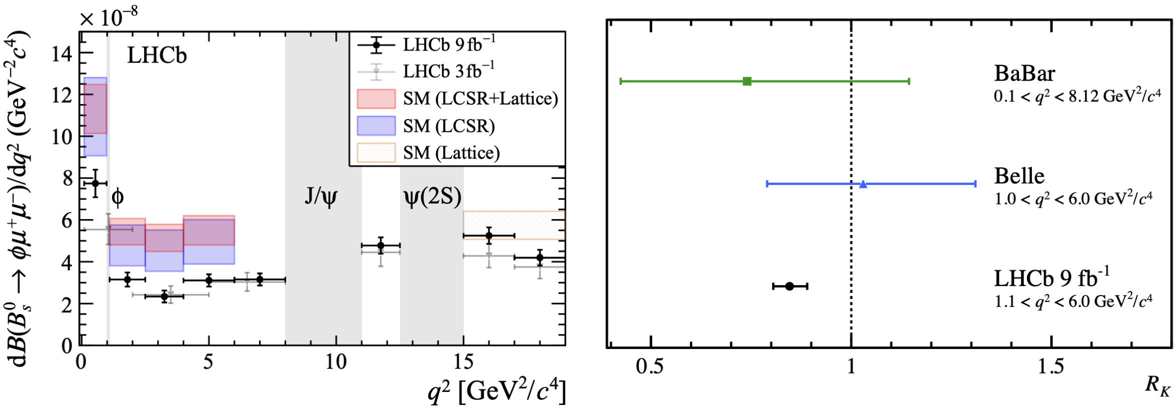


Figure 1.1: LHCb result on the differential branching fraction of the $B_s^0 \to \phi\mu^+\mu^-$ decay. In the q^2 region 1.1-6.0 GeV^2, the measurement is found to be 3.6 σ below the SM prediction. Taken from Ref. [1] (left). LHCb result on R_K superimposed with results from Belle and BaBar. The measured R_K has a 3.1 σ discrepancy with the SM. Taken from Ref. [2] (right).

There is possibly a relation between the flavour anomalies found on the neutral $b \to sll$ and on the charged $b \to cl\bar{\nu}$ transitions with other anomalies in the flavour sector: they may be originating from a common Lepton Flavour Universality (LFU) violating interaction. Besides the $R_{K^{(*)}}$ results, measurements on $R_{D^{(*)}}$ [20–22] and on the anomalous magnetic moments $(g-2)_l$ of charged leptons also signal departures from the SM prediction of LFU. Here there is the long-standing discrepancy in $(g-2)_\mu$, whose most recent result [23] signals a tension of 4.2 σ. A collection of measurements (in blue) which see deviations from the theoretical expectation (in orange) can be seen in Fig. 1.2. [1,6] intervals refer to the di-lepton invariant mass squared (q^2) ranges in GeV^2. All these results are connected with the muon sector and some authors [24] start to refer to them jointly as *muon anomalies*.

Although the individual discrepancies detected in these measurements do not yet display sufficient levels of statistical significance, when taken together in global fits they do, and they can be used to constrain possible NP models, as will be explained in Sec. 2.4. In Ref. [25], a naive combination of R_K and R_{K^*} measurements is reported to have a tension with the SM above 4 σ. In Ref. [26], the combined significance of the deviations found on R_{K^*} and $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ is shown to reach 4.7 σ. And in Ref. [27], global fits to several $b \to sll$ observables pass over the 5 σ barrier. Therefore, a clarification of the flavour anomalies is a current priority in High Energy Physics (HEP).

In this thesis, we will study a decay which is a particular realisation of the aforementioned $b \to sll$ transitions, $B^0 \to K^{*0}\mu^+\mu^-$, with the Compact Muon Solenoid (CMS) experiment at the LHC. This decay has been thoroughly studied in the literature. The previous analyses performed by the CMS collaboration [4, 28, 29] all used Run 1 proton-proton (pp) collision data and measured several angular parameters as well as the differential branching fraction. The results were found to be consistent with the SM predictions and with the ones obtained with previous measurements [14, 30–32].

The first complete angular analysis was performed by the LHCb collaboration [33] with 2011 pp data

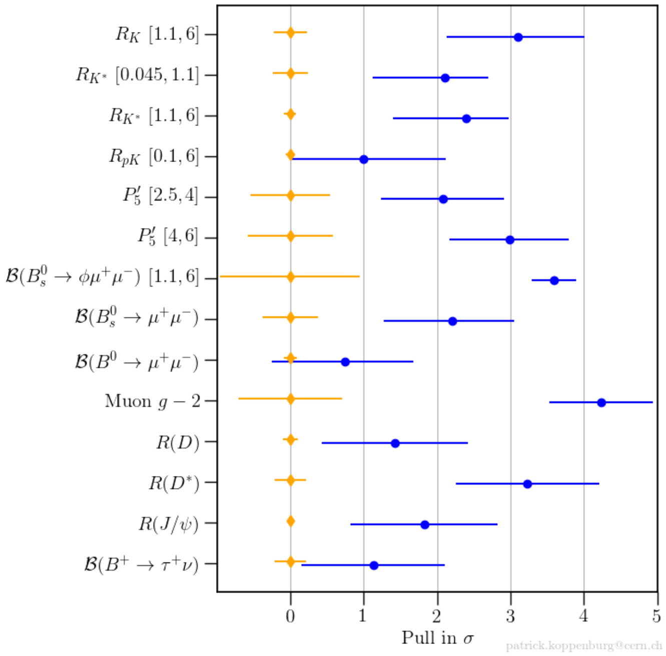


Figure 1.2: Collection of experimental measurements (in blue) and their shift with respect to the theoretical prediction (in orange), which is set at zero. [1,6] intervals refer to the di-lepton invariant mass squared (q^2) ranges in GeV^2. Taken from `http://www.scholarpedia.org/article/Rare_decays_of_b_hadrons`.

at $\sqrt{s} = 7$ TeV. A 3.7 σ discrepancy was detected in one parameter that enters in the angular decay rate: P_5'. Later, Belle independently reported a similar tension [34]. However, ATLAS measurements [35] were consistent with the SM. After these results, LHCb made another two publications [36, 37] and the tensions persisted with the new data. The previous differential branching fraction measurements by Belle [30], LHCb [14] and CMS [4, 28] were all in good agreement with the SM prediction, but in Ref. [14] they favoured lower values. In Fig. 1.3 (left), the latest LHCb result [3] on the P_5' parameter can be seen. A local discrepancy of 2.5 and 2.9 σ in the ranges 4.0 $< q^2 <$ 6.0 and 6.0 $< q^2 <$ 8.0 GeV^2 is reported, being slightly lower the previous tensions [36] (2.8 and 3.0 σ). However, from the global fit of several angular observables, the discrepancy with the SM is shown to increase from 3.0 to 3.3 σ. In Fig. 1.3 (right), the previous CMS measurement [4] of the $B^0 \to K^{*0}\mu^+\mu^-$ differential branching fraction can be seen. The results are consistent with the SM.

Though some anomalies may originate from statistical fluctuations, underestimated systematic uncertainties or even a deficient SM prediction or measurement, the persistence of the flavour anomalies makes them very alluring, to both theorists and experimentalists. In the theoretical community, people have been developing models that attempt to explain the anomalies along with other tensions in the flavour sector, such as the anomalous magnetic moment of the muon $(g-2)_\mu$. Among these, the most favoured ones are models with extra gauge bosons (Z') or leptoquarks (LQs), which will be explained with more detail in Ch. 2.

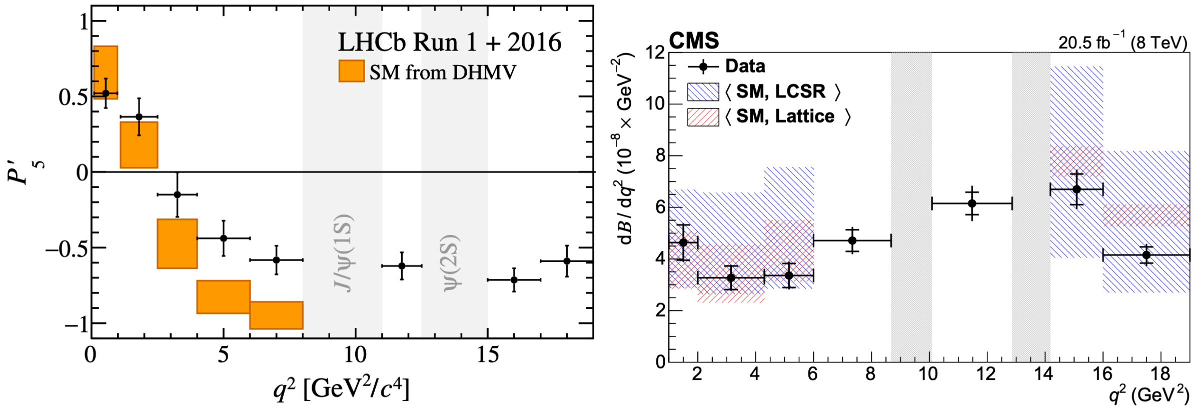


Figure 1.3: Latest LHCb result [3] on the P_5' parameter, where a tension with the SM of 2.5 and 2.9 σ is reported in the ranges 4.0 $< q^2 <$ 6.0 and 6.0 $< q^2 <$ 8.0 GeV^2, respectively (left). Latest CMS result [4] on the differential branching fraction of the $B^0 \to K^{*0}\mu^+\mu^-$ decay (right), which is consistent with the SM.

From the experimental side, many tensions with the SM have been reported independently by LHCb, Belle and BaBar throughout the years. The results from larger experiments, like CMS and ATLAS, were however reported to be consistent with the SM predictions but were mainly dominated by statistical uncertainties. It is now of paramount importance that more precision measurements are performed so that we can gauge the role of these anomalies in the current physics paradigm. With the available Run 2 data, where much higher luminosities were achieved, if we can push the significance of the results over the 5 σ barrier, they will count as an observation of flavour anomalies, making them the first established NP seen at the LHC.

1.2 Analysis introduction

In this thesis, the differential branching fraction of the $B^0 \to K^{*0}\mu^+\mu^-$ will be measured as a function of the di-muon invariant mass squared (q^2). The data used was collected by the CMS detector in pp collisions at $\sqrt{s}$ =13 TeV, during the data-taking years 2016, 2017 and 2018 of Run 2, corresponding to a total integrated luminosity of 139.5 fb^{-1}. The work is inserted in a larger analysis [13] which also has the goal of measuring several angular parameters appearing in the decay rate.

The $B^0 \to K^{*0}\mu^+\mu^-$ decay has a fully charged final state $K^+\pi^-\mu^+\mu^-$ composed of a kaon and a pion that result from the K^{*0} and two opposite charge muons. Its branching fraction $\mathcal{B}(B^0 \to K^{*0}\mu^+\mu^-)$ measures the number of B^0 mesons that decay into the particular final state $K^{*0}\mu^+\mu^-$ with respect to the total number of B^0 mesons that decay via all possible modes. It is then proportional to the number of signal events that we measure (Y_S). Since Y_S is not the total number of B^0 that decay to the final state $K^{*0}\mu^+\mu^-$ but actually the number of them that do and also pass our trigger and selection criteria (see Ch. 4), we need to correct Y_S by the detector efficiency ϵ_S, which measures the fraction of events that pass our selections.

The final ingredient that is missing for the $\mathcal{B}(B^0 \to K^{*0}\mu^+\mu^-)$ calculation is the number of B^0 mesons that decay via any possible mode. This quantity is not experimentally accessible at the LHC. In order to cope with this, what is actually measured is a normalised branching fraction so that this quantity cancels when doing the ratio. Moreover, due to the presence of large uncertainties regarding trigger and production efficiencies at the LHC, only ratios of branching fractions can be measured precisely.

The channel used as normalisation needs to have the same final state as the signal decay of interest, its production dynamics and decay kinematics should be the same as the $B^0 \to K^{*0}\mu^+\mu^-$ decay and its absolute branching fraction should be known with an uncertainty negligible when compared with other sources. In our study, we use the $B^0 \to J/\psi K^{*0}$ resonance as our normalisation channel, where $J/\psi \to \mu^+\mu^-$ and $\mathcal{B}(B^0 \to J/\psi K^{*0}) \times \mathcal{B}(J/\psi \to \mu^+\mu^-) =$ 0.127 % × 5.961 % [38].

The analysis is performed in bins of q^2, as defined in Tab. 1.1. The choice of q^2 bins is similar to the one found on previous analyses and is motivated by theoretical considerations. As will be explained in Ch. 2, the $B^0 \to K^{*0}\mu^+\mu^-$ effective Hamiltonian can be written in terms of Wilson Coefficients (WC), whose values can be modified in the presence of NP and each q^2 region is sensitive to different coefficients C_i, as can be seen in Fig. 1.4. The theoretical calculations are performed with good control in the low q^2 region 1-6 GeV^2. Above 6 GeV^2, at the charmonium resonances, QCD Factorisation (QCDF) methods, necessary to calculate some non-factorisable terms in the decay rate, fail and lattice QCD calculations are necessary. The region below 1 GeV^2 is dominated by the photon pole ($b \to s\gamma$).

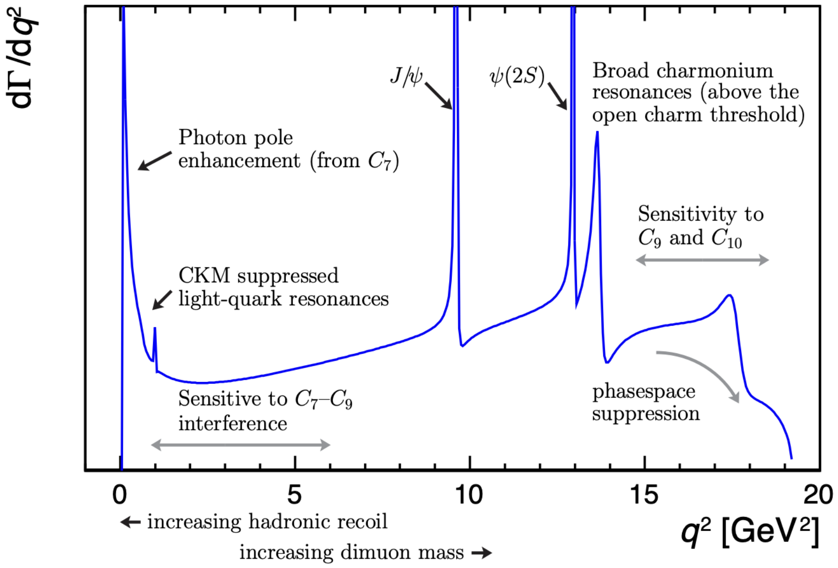


Figure 1.4: Differential decay rate as a function of q^2 for $B^0 \to K^{*0}l^+l^-$ processes. Each region of q^2 is sensitive to different WCs C_i. Taken from Ref. [5].

The q^2 ranges $8.68 < q^2 < 10.09$ GeV^2 and $12.86 < q^2 < 14.18$ GeV^2 correspond, respectively, to the resonant channels $B^0 \to J/\psi K^{*0}$ and $B^0 \to \psi(2S)K^{*0}$, which occur much more frequently in the collisions. These decays have the same final state as the non-resonant channel of interest $K^+\pi^-\mu^+\mu^-$, where $\mu^+\mu^-$ result from the J/ψ and $\psi(2S)$ decays, and they are both used as control channels. The former is additionally used to normalise the branching fraction measurement

$$\frac{d\mathcal{B}(B^0 \to K^{*0}\mu^+\mu^-)}{dq^2} = \frac{Y_S}{Y_N}\frac{\epsilon_N}{\epsilon_S}\frac{\mathcal{B}(B^0 \to J/\psi K^{*0}) \times \mathcal{B}(J/\psi \to \mu^+\mu^-)}{\Delta q_i^2}, \tag{1.1}$$

where Y_S and Y_N are the yields of signal and normalisation channels, respectively, ϵ_S and ϵ_N are the detector efficiencies for signal and normalisation channels, respectively, and i runs over the non-resonant q^2 bins defined in Tab. 1.1.

Table 1.1: q^2 bins used in the analysis.

bin index	q^2 range [GeV^2]
0	1-2
1	2-4.3
2	4.3-6
3	6-8.68
4	8.68 - 10.09 (J/ψ region)
5	10.09-12.86
6	12.86-14.18 ($\psi(2S)$ region)
7	14.18-16

Despite the fact that the previous branching fraction results were all consistent with the SM, they suffered both from large theoretical uncertainties and large statistical uncertainties which dominated Run 1 results. Furthermore, this observable can enter in global fits which help constraining NP scenarios and it is also an ingredient of R_{K^*}, whose theoretical uncertainties are much lower and in which several tensions have already been reported. This stresses the importance of this study in light of the current flavour anomalies. Compared with the CMS Run 1 analysis [4], the current work will benefit both from a centre-of-mass energy increase (from 8 TeV to 13 TeV), which will augment the B^0 cross-section by a factor of about 2, and from a luminosity increase (from 20.5 fb^{-1} to 139.5 fb^{-1}). Both factors will allow to improve the precision of the measurements.

The structure of this thesis is as follows. We begin in Ch. 2 with a theoretical introduction of the $b \to sll$ transitions and how they can be studied in the context of Effective Field Theories (EFT). A particular emphasis will be put on the $B^0 \to K^{*0}\mu^+\mu^-$ decay, where a description of its angular decay rate will be provided. We continue on Ch. 3 with a description of the CMS detector at the LHC, highlighting the components more relevant to the analysis. We also give an overview of the CMS trigger system, describing the triggers used in the analysis. In Ch. 4, we describe the datasets and simulation samples used as well as the offline selections. A study of the possible background sources contaminating the decay channels and what cuts are used to reduce them is also presented. The method used to extract the yields Y_S and Y_N in Eq. (1.1) is provided in Ch. 5 and the efficiency computation (ϵ_S and ϵ_N) can be found in Ch. 6. The Monte Carlo (MC) validation is described in Ch. 7 and a study of systematic uncertainties is presented in Ch. 8. We finally give in Ch. 9 the results with a discussion and in Ch. 10 the conclusions.

Chapter 2

Theoretical framework

In this chapter, we begin by giving an overview of the electroweak sector of the SM in Sec. 2.1 and in Sec. 2.2 we present the theoretical machinery of EFTs and how it can be used to search for NP in $b \to sll$ transitions. Using this formalism, the $B^0 \to K^{*0}\mu^+\mu^-$ decay rate will be given as a function of angular parameters in Sec. 2.3. This decay rate is then used in Ch. 6 in the efficiency computation. Finally, in Sec. 2.4 we explain how global fits to the flavour anomalies can be used to constrain NP scenarios and we refer two classes of NP models that have been more favoured in the theoretical community: models with an extra gauge boson (Z') and with LQs.

2.1 Electroweak sector of the Standard Model

The electroweak sector of the SM unifies the electromagnetic interaction with the weak nuclear force. The electromagnetic interaction has gauge group $U(1)_Q$ where Q is the electric charge, it has a massless and neutral gauge boson (the photon) and has an infinite range. The weak interaction is described by the group $SU(2)$, it has three gauge bosons ($W^\pm$ and Z^0) and its range is confined to the atomic nucleus (0.1-0.01 fm). The SM electroweak sector has gauge group $SU(2) \times U(1)_Y$, where $SU(2)$ has three generators T_1, T_2 and T_3 (weak isospin) and $U(1)_Y$ has one generator Y (weak hypercharge). In nature, only the $U(1)_Q$ symmetry is observed and results from the Spontaneous Symmetry Breaking (SSB) of $SU(2) \times U(1)_Y$ (see Sec. 2.1.1), which also gives origin to the $W^\pm$ and Z^0 masses.

Besides the photon and the $W^\pm$ and Z^0 bosons, the SM particle content also includes fermions, which can be separated into two categories: leptons and quarks. Leptons do not interact via the strong force and they come in three different flavours: electron, muon and tau ($Q = -1$); and the corresponding neutrinos ($Q = 0$), as can be seen in Tab. 2.1. Flavour is the analogous of the electric charge for the weak nuclear force. As indicated in Tab. 2.1, fermions can be organised in families (columns): particles in the same row have the same electric charge but a different flavour charge. Unlike leptons, quarks interact via the strong force and they are confined inside hadrons: mesons ($q\bar{q}$ pair) and baryons ($q_1q_2q_3$), among other more exotic states. They also come in three flavours: up (u), charm (c) and top (t) with $Q = +2/3$ and down (d), strange (s) and beauty (b) with $Q = -1/3$.

Table 2.1: Leptons (first two rows) and quarks (last two rows).

1st family	2nd family	3rd family	Q
ν_e	ν_μ	ν_τ	0
e^-	μ^-	τ^-	-1
u	c	t	$+2/3$
d	s	b	$-1/3$

The weak interaction violates parity, meaning that it attributes different quantum numbers to the left- and right-handed chirality components of leptons and quarks (the SM is a chiral gauge theory). Since we are interested in the quark-level $b \to sll$ transitions, we will be focusing on quarks hereafter. The left-handed components of quarks are doublets of $SU(2)$, $\begin{pmatrix} U_L \\ D_L \end{pmatrix}$, with $\psi_L = P_L\psi$, $P_L = \frac{1-\gamma_5}{2}$, $U = u, c, t$ and $D = d, s, b$ and the right-handed components are singlets of $SU(2)$, U_R and D_R, with $\psi_R = P_R\psi$ and $P_R = \frac{1+\gamma_5}{2}$. γ_5 is written in terms of the gamma matrices: $\gamma_5 = i\gamma^0\gamma^1\gamma^2\gamma^3$.

Another interesting property of the weak interaction is that it violates CP, the combined symmetry of parity (P) and charge conjugation (C), which transforms a particle into its anti-particle counterpart. This has profound implications in cosmology, being one of Sakharov's criteria for baryogenesis, one of the current proposed explanations for the observed matter-antimatter asymmetry in the universe.

In order to better understand the weak interaction, we need to dive into the scalar sector of the SM, which will be explained in the next section.

2.1.1 Scalar sector of the Standard Model

As previously mentioned, what we observe in nature is not the gauge symmetry of the SM, $SU(2) \times U(1)_Y$, but the $U(1)_Q$ symmetry of electromagnetism. This is a result of SSB, which occurs by the presence of a Vacuum Expectaction Value (VEV) of a scalar field. This scalar field needs to be both in a doublet of $SU(2)$ and neutral in order to break $SU(2)$ and preserve $U(1)_Q$, respectively. If we denote the scalar $SU(2)$ doublet by ϕ we can write

$$\phi = \begin{pmatrix} \Phi^+ \\ \Phi^0 \end{pmatrix}$$

where Φ^+ and Φ^0 are complex fields. The scalar potential V must be invariant under $SU(2) \times U(1)_Y$ and lead to a renormalisable theory[1] $V = \mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2$, where λ is dimensionless and μ has mass dimension.

---

[1] In order for a theory to be renormalisable, its Lagrangian cannot have fields with mass dimension greater than 4.

The minimnum v of V is the VEV of the neutral scalar field Φ^0, $v = \sqrt{-\frac{\mu^2}{2\lambda}}$. If we expand Φ^0 around its VEV by writing $\Phi^0 = v + \frac{H+i\chi}{\sqrt{2}}$, where H and χ are real fields and substitute this expression in V we conclude that χ and $\Phi^\pm$ are massless and H has mass $m_H = \sqrt{-2\mu^2}$. The χ and $\Phi^\pm$ are therefore the three Goldstone bosons corresponding to the three broken generators of $SU(2)$[2]. H is the Higgs boson, a particle discoverd at the LHC in 2012. Its mass is measured to be 125.10 ± 0.14 GeV [38] but it is not a prediction of the SM. In fact, at loop order, it has a quadratically divergent contribution δm_H^2 which needs to be cutoff at an energy scale Λ. The larger Λ is, the smaller μ needs to be so that $m_H =$ 125 GeV. Imposing $\delta m_H^2 < m_H^2$ (naturalness), implies that $\Lambda <$ 1 TeV, meaning that the SM, as a natural EFT, is only valid up to the 1 TeV energy scale. This is the source of the hierarchy problem mentioned in Ch. 1. If we further replace the expression for ϕ in the SM Lagrangian, we conclude that the photon remains massless and the $W^\pm$ and Z^0 bosons acquire masses which are related by the experimentally well verified relation $m_W = m_Z \cos\theta_w$, where θ_w is the Weinberg angle ($M_W =$ 80.379 ± 0.012 GeV and $M_Z =$ 91.1876 ± 0.0021 GeV [38]).

2.1.2 Yukawa interactions of the quarks

The quark Yukawa Lagrangian describes the interactions between quarks and the scalar field ϕ defined in Sec. 2.1.1 and can be written as

$$\mathcal{L}_{Yukawa} = -\sum_{i,j=1}^{3} Y_{ij}(\bar{U}_{L_i}\ \bar{D}_{L_i})D_{R_j}(\Phi^+\ \Phi^0)^\dagger - \sum_{i,j=1}^{3} Y'_{ij}(\bar{U}_{L_i}\ \bar{D}_{L_i})U_{R_j}(\Phi^{0*}\ \ -\Phi^-)^\dagger + \text{H.c.}, \tag{2.1}$$

where i, j run over the three quark families and Y and Y' are 3×3 matrices of Yukawa couplings and H.c. stands for hermitian conjugate.

The quark mass matrices $M_D = vY$ and $M_U = vY'$ do not commute and so there is no basis in flavour space where both matrices are diagonal. However, there are 3×3 unitary matrices $B^{U,D}_{L,R}$ such that the two can be bi-diagonalised

$$B_L^{U\dagger} M_U B_R^U = M_u = diag(m_u, m_c, m_t) \tag{2.2}$$

$$B_L^{D\dagger} M_D B_R^D = M_d = diag(m_d, m_s, m_b). \tag{2.3}$$

If we re-define the quark fields by $\mathcal{U}_L = B_L^{U\dagger} U_L$, $\mathcal{U}_R = B_R^{U\dagger} U_R$, $\mathcal{D}_L = B_L^{D\dagger} D_L$ and $\mathcal{D}_R = B_R^{D\dagger} D_R$, the Yukawa Lagrangian becomes

$$\mathcal{L}_{Yukawa} = \bar{\mathcal{U}}(M_u V P_L - V M_d P_R)\mathcal{D}\frac{\Phi^+}{v} + \bar{\mathcal{D}}(V^\dagger M_u P_R - M_d V^\dagger P_L)\mathcal{U}\frac{\Phi^-}{v} + \ldots, \tag{2.4}$$

---

[2] The Goldstone theorem states that for each spontaneously broken gauge symmetry generator there is a corresponding massless scalar, a Goldstone boson.

where $\mathcal{D} = \mathcal{D}_L + \mathcal{D}_R, \mathcal{U} = \mathcal{U}_L + \mathcal{U}_R$ and a notation has been adopted in order to get rid of the i, j indices. $V = B_L^{U\dagger} B_L^D$ is the Cabibbo–Kobayashi–Maskawa (CKM) matrix and it represents the relative rotation between the two eigenbases of the up and down mass matrices.

The entries of V are measured experimentally [38] and the matrix is found to be almost diagonal

$$V = \begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix} = \begin{pmatrix} 0.97370 \pm 0.00014 & 0.2245 \pm 0.0008 & 0.00382 \pm 0.00024 \\ 0.221 \pm 0.004 & 0.987 \pm 0.011 & 0.0410 \pm 0.0014 \\ 0.0080 \pm 0.0003 & 0.0388 \pm 0.0011 & 1.013 \pm 0.039 \end{pmatrix}.$$

The charged current transitions depend on the entries of the CKM matrix V_{ij}, which give the relative strength for a certain $q_i \to q_j$ transition to occur

$$\mathcal{L}_{W^\pm} = \frac{g}{\sqrt{2}}(\bar{\mathcal{U}}\gamma^\mu V P_L \mathcal{D} W_\mu^+ + \bar{\mathcal{D}} V^\dagger \gamma^\mu P_L \mathcal{U} W_\mu^-). \tag{2.5}$$

Therefore, charged currents change flavour. Transitions between quarks of the same family are less suppressed, because the diagonal elements of V are close to 1. Transitions between the 1st and 2nd or between the 2nd and 3rd families are more suppressed and transitions between the 1st and 3rd families are even more suppressed, since $|V_{td}|$ and $|V_{ub}|$ are very close to 0.

Neutral current transitions, on the other hand, do not depend on V_{ij}

$$\mathcal{L}_\gamma = eA_\mu\Big(-\frac{2}{3}\bar{\mathcal{U}}\gamma^\mu \mathcal{U} + \frac{1}{3}\bar{\mathcal{D}}\gamma^\mu \mathcal{D}\Big), \tag{2.6}$$

where A_μ represents the photon and

$$\mathcal{L}_{Z^0} = \frac{g}{\cos\theta_w} Z_\mu \Big[\bar{\mathcal{U}}\gamma^\mu\Big(\frac{P_L}{2} - \frac{2\sin\theta_w{}^2}{3}\Big)\mathcal{U} + \bar{\mathcal{D}}\gamma^\mu\Big(\frac{\sin\theta_w{}^2}{3} - \frac{P_L}{2}\Big)\mathcal{D}\Big]. \tag{2.7}$$

As a consequence, there are no FCNC decays at tree level in the SM.

The $b \to sll$ transitions are highly suppressed. On the one hand, because they correspond to a FCNC transition between a beauty quark from the 3rd family and a strange quark from the 2nd family. On the other hand, because the leading order loop diagram is a penguin loop with a t-quark mediator (see Fig. 2.1 a)) which is Cabibbo suppressed since it is proportional to the CKM elements $|V_{ts}V_{tb}| \approx 10^{-2}$. Other SM loop contributions proceed via box diagrams, as can be seen in Fig. 2.1 b).

The fact that these transitions are so suppressed makes decays realising them very rare and sensitive to NP, which could appear as yet undiscovered fundamental particles, such as LQs or heavier gauge bosons (Z') (see Fig. 2.1 c) and 2.1 d)), which would allow the existence of these FCNC transitions at tree level. These NP contributions could produce sizeable effects in the decay rates and branching fractions which would appear as flavour anomalies. These indirect searches for NP are done with the help of the formalism of EFTs, as will be explained in the next section.

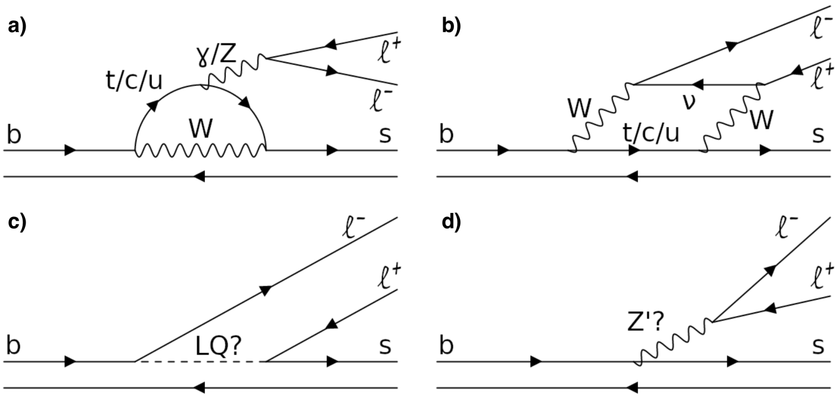


Figure 2.1: Loop diagrams contributing to the $b \to s\, l^+l^-$ transitions in the SM (top) and in possible NP scenarios (bottom). a) and b) correspond to a penguin loop and a box diagram, respectively. c) and d) correspond to tree level diagrams mediated by LQs and a heavy gauge boson (Z'), respectively.

2.2 Effective field theories

When describing a physical system, one can normally focus on the degrees of freedom that are relevant at the distance scales under consideration. This means that all particles whose Compton wavelength $\lambda = h/mc$ is smaller than those distance scales are irrelevant and can be integrated out. In cases where very disparate mass scales appear, it is advantageous to construct an EFT where the degrees of freedom that become relevant at much lower distance scales (much higher energies) do not appear explicitly. An example would be a NP particle whose mass could be hundreds of TeV and that cannot be produced directly in the collisions at the LHC, at an energy scale smaller than its mass. Consequently, the EFT Lagrangian, valid at the energy scales of interest, does not contain this degree of freedom.

The fact that this is possible is ensured by the decoupling theorem [39], which states that the heavy degrees of freedom decouple at energy scales much lower than their mass. Mathematically, this means that we can perform a functional integral over these fields i.e. we can write the partition function $Z[j]$, with external source $j(x)$, as

$$\begin{aligned} Z[j] &= \int [d\phi][d\Phi] \exp\left(\int d^4x \big[\mathcal{L}(\phi,\Phi) + j\phi\big]\right) \\ &= \int [d\phi] \exp\left(\int d^4x \big[\mathcal{L}_{eff}(\phi) + j\phi\big]\right) \end{aligned} \tag{2.8}$$

where ϕ and Φ correspond to the light (μ) and heavy (Λ) mass fields, respectively, and

$$\exp\left(\int d^4x \mathcal{L}_{eff}(\phi)\right) = \int [d\Phi] \exp\left(\int d^4x \mathcal{L}(\phi,\Phi)\right). \tag{2.9}$$

For length scales $x \gg 1/\Lambda$, all interactions due to Φ become local and $\mathcal{L}_{eff}$ is a local effective Lagrangian. When we consider decay processes containing b-hadrons, the relevant scale of such transitions is the b-quark mass, $m_b \approx 5\,\mathrm{GeV}$. The SM contains particles whose masses are much larger than this energy scale: the top quark, the weak bosons and the Higgs masses ($m \approx 100\,\mathrm{GeV}$). Therefore, in an EFT of the SM, these particles do not appear explicitly.

The starting point for the construction of an EFT is the presence of a large energy scale Λ, which could be the W boson mass or even a predicted NP scale. Then one separates the long- and short-distance contributions to the matrix elements of the effective Hamiltonian $\mathcal{H}_{eff}$ denoting a transition from an initial state $|i\rangle$ to a final state $|f\rangle$, whose energies $E_{i,f} < \Lambda$, by means of the operator-product expansion

$$\langle f|\, \mathcal{H}_{eff}\, |i\rangle = \sum_k C_k(\Lambda)\, \langle f|\, \mathcal{O}_k\, |i\rangle, \tag{2.10}$$

where $C_k(\Lambda)$ are the WCs and contain physics information above Λ, being sensitive to NP, and $\langle f|\, \mathcal{O}_k\, |i\rangle$ are the matrix elements of the local operators $\mathcal{O}_k$, which contain physics information below Λ.

2.2.1 Fermi's theory of the weak interactions

In 1934, Fermi formulated a theory of β-decay as a phenomenological modification of Quantum Electrodynamics (QED), before the SM had been developed. We now look at Fermi's theory of weak interactions as an EFT of the SM, which is very useful to describe processes occurring at energy scales much lower than the $W^{\pm}$ and Z^0 masses.

The charged current interactions described in Eq. (2.5) can be written in terms of the currents

$$J^{\mu}_{\pm} = \frac{J^{\mu}_1 \mp i J^{\mu}_2}{\sqrt{2}}, \tag{2.11}$$

with

$$J^{\mu}_a = g \sum_{\Psi} \bar{\Psi} \gamma^{\mu} P_L \frac{\tau_a}{2} \Psi, \tag{2.12}$$

where Ψ runs over the lepton and quark fields and τ_a are the Pauli matrices. Whereas the neutral current interactions described in Eqs. (2.7) and (2.6) can be written as

$$J^{\mu}_0 = \frac{g}{\cos\theta_w} (J^{\mu}_3 - \sin\theta_w{}^2 J^{\mu}_{EM}), \tag{2.13}$$

where J^{μ}_{EM} is the electromagnetic current

$$J^{\mu}_{EM} = \sum_{\Psi} eQ \bar{\Psi} \gamma^{\mu} \Psi. \tag{2.14}$$

For low energy weak processes, i.e. processes mediated by a $W^{\pm}$ or Z^0 boson whose exchanged momentum p is much lower than the boson mass M, we can Taylor expand the gauge bosons propagators in terms of p^2/M^2 and keep only the constant leading order term

$$\frac{1}{p^2 - M^2} = -\frac{1}{M^2} + \mathcal{O}\left(\frac{p^2}{M^2}\right). \tag{2.15}$$

Since the Fourier transform of a constant is a Dirac δ-function, the process with the exchange of a W/Z boson in Fig. 2.2 (a) is in Fermi's theory a point-like interaction (see Fig. 2.2 (b)) with coupling constant G_F

$$\mathcal{L}_{eff}^{Fermi} = \frac{8G_F}{\sqrt{2}}(J_\mu^+ J_{-\mu} + \frac{1}{2}J_0^\mu J_{0\mu}), \tag{2.16}$$

where $G_F = 1.166 \times 10^{-5}$ GeV^2 is the analogous of the WCs in the Fermi theory.

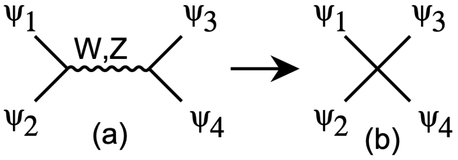


Figure 2.2: Tree level interaction between 4 fermions, mediated by the W/Z bosons in the SM (a) and point-like interaction between 4 fermions in Fermi's theory. Taken from Ref. [6].

2.2.2 Effective field theory of $b \to sl^+l^-$ transitions

The $b \to sl^+l^-$ are quark-level transitions. What makes quarks more complicated than leptons is that they interact via the strong force and that the strong coupling α_s diverges below an energy scale $\Lambda_{QCD} \approx$ 0.2 GeV [40], whereas the QED coupling α_e is approximately constant ($\alpha_e \approx$ 1/137) at large momenta. This means that at very large momenta α_s becomes small, quarks are asymptotically free and perturbation theory methods can be applied. However, at low momenta α_s becomes large, quarks become confined inside hadrons and perturbation theory fails.

This has direct consequences to the definition of current-quark masses. Even though we can safely compare the theoretical prediction of e.g. the electron mass with the experimental measurement, because α_e is approximately constant at common collision energies, we cannot do the same for quarks. In this case, we need to define a renormalisation scale μ at which this comparison can be performed. Moreover, current-quark masses also depend on the way divergences are removed from the QCD Lagrangian, something referred to as renormalisation scheme. The Particle Data Group (PDG) [38] quotes the current-quark masses in the $\overline{\text{MS}}$ scheme at a renormalisation scale $\mu = 2$ GeV. The fact that the choice of renormalisation scale is arbitrary is included by imposing that QCD must be invariant under the renormalisation group, meaning that no physical observable can depend on μ. The evolution of QCD parameters in terms of μ is governed by the renormalisation group equations.

Another consequence of quark confinement is that they cannot be probed directly in experiments. However, their dynamics can be studied by scattering leptons off hadrons or by scattering hadrons on other hadrons. From these interactions the observables that can be extracted are form factors that encode hadrons' momentum-dependent interactions with photons, W and Z bosons.

Applying a similar reasoning to what is done in Fermi's theory, we can describe the $b \to sl^+l^-$ transitions which occur at scales much lower than the W/Z, top-quark and NP scales with an EFT of the SM. In this way, the interactions depicted in Fig. 2.1 reduce to point-like interactions with couplings given by the corresponding WCs.

The u-quark loop contribution in Fig. 2.1 a) is doubly Cabibbo suppressed since $V_{ub}V_{us}^* \approx 0.0007 \ll V_{tb}V_{ts}^* \approx 0.04$ and can be neglected. With this approximation and using the unitarity of the CKM matrix, $V_{ub}V_{us}^* + V_{cb}V_{cs}^* + V_{tb}V_{ts}^* = 0$, the t-quark and c-quark contributions are related by $V_{tb}V_{ts}^* = -V_{cb}V_{cs}^*$. The effective Hamiltonian can then be written as

$$\mathcal{H}_{eff} = -\frac{4G_F}{\sqrt{2}} V_{ts}V_{tb}^* \Big[C_1\mathcal{O}_1^c + C_2\mathcal{O}_2^c + \sum_{i=3}^{6} C_i\mathcal{O}_i + \sum_{i=7,8,9,10,P,S} (C_i\mathcal{O}_i + C_i'\mathcal{O}_i') \Big]. \tag{2.17}$$

The operators $\mathcal{O}_{i\leq 6}$ are identical to the P_i given in Ref. [41], with $\mathcal{O}_1^c$ and $\mathcal{O}_2^c$ describing point like interactions between the b- / s-quarks and the c-quarks, which cannot be integrated out at scales $\mu \approx m_b$. The remaining ones are given by

$$\mathcal{O}_7 = \frac{e}{g^2} m_b (\bar{s}\sigma_{\mu\nu}P_R b)F^{\mu\nu}, \qquad \mathcal{O}_7' = \frac{e}{g^2} m_b (\bar{s}\sigma_{\mu\nu}P_L b)F^{\mu\nu}, \tag{2.18}$$

$$\mathcal{O}_8 = \frac{1}{g} m_b (\bar{s}\sigma_{\mu\nu}\tau^a P_R b)G^{\mu\nu a}, \qquad \mathcal{O}_8' = \frac{1}{g} m_b (\bar{s}\sigma_{\mu\nu}\tau^a P_L b)G^{\mu\nu a}, \tag{2.19}$$

$$\mathcal{O}_9 = \frac{e^2}{g^2} (\bar{s}\gamma_\mu P_L b)(\bar{\mu}\gamma^\mu \mu), \qquad \mathcal{O}_9' = \frac{e^2}{g^2} (\bar{s}\gamma_\mu P_R b)(\bar{\mu}\gamma^\mu \mu), \tag{2.20}$$

$$\mathcal{O}_{10} = \frac{e^2}{g^2} (\bar{s}\gamma_\mu P_L b)(\bar{l}\gamma^\mu \gamma_5 l), \qquad \mathcal{O}_{10}' = \frac{e^2}{g^2} (\bar{s}\gamma_\mu P_R b)(\bar{l}\gamma^\mu \gamma_5 l), \tag{2.21}$$

$$\mathcal{O}_S = \frac{e^2}{16\pi^2} m_b (\bar{s}P_R b)(\bar{l}l), \qquad \mathcal{O}_S' = \frac{e^2}{16\pi^2} m_b (\bar{s}P_L b)(\bar{l}l), \tag{2.22}$$

$$\mathcal{O}_P = \frac{e^2}{16\pi^2} m_b (\bar{s}P_R b)(\bar{l}\gamma_5 l), \qquad \mathcal{O}_S' = \frac{e^2}{16\pi^2} m_b (\bar{s}P_L b)(\bar{l}\gamma_5 l), \tag{2.23}$$

where $g = \sqrt{4\pi\alpha_s}$, m_b is the b-quark mass in the $\overline{\text{MS}}$ scheme, $F^{\mu\nu}$ is the photon field strength tensor and $G^{\mu\nu}$ is the gluon field strength tensor. $\mathcal{O}_i$ and $\mathcal{O}_i'$ have opposite chirality.

The WCs C_i encode short-distance physics and possibly NP. They are calculated at the matching scale $\mu = M_W$ in a perturbative expansion in powers of $\alpha_s(M_W)$

$$C_i = C_i^{(0)} + \frac{\alpha_s}{4\pi} C_i^{(1)} + \left(\frac{\alpha_s}{4\pi}\right)^2 C_i^{(2)} + \mathcal{O}(\alpha_s^3), \tag{2.24}$$

where $C_i^{(0)}$ and $C_i^{(n)}$ are tree level and n-loop contributions, respectively. They are then evolved down to scales $\mu \approx m_b$ by means of the renormalisation group equation. Any NP effect enters in $C_i(M_W)$, while the evolution to lower scales is determined by the SM.

2.3 The $B^0 \to K^{*0}\mu^+\mu^-$ decay rate

The $B^0 \to K^{*0}(K^+\pi^-)\mu^+\mu^-$ decay has a fully charged final state composed of a kaon, a pion and two opposite charge muons, which are easily identifiable in the CMS detector. The charge of the hadrons in the final state determines the CP-state of the decay, i.e. whether it is $B^0 \to K^{*0}\mu^+\mu^- \to K^+\pi^-\mu^+\mu^-$ or $\bar{B}^0 \to \bar{K}^{*0}\mu^+\mu^- \to K^-\pi^+\mu^+\mu^-$. The four-particle final state can be fully described by a set of four kinematic variables: q^2, θ_K, θ_l and ϕ. q^2 is the di-muon invariant mass squared and θ_K, θ_l and ϕ are the three angles depicted in Fig. 2.3. θ_l is the angle between the momentum of the μ^+ (μ^-) and the direction opposite to the B^0 ($\bar{B}^0$) momentum in the di-muon rest frame. θ_K is the angle between the direction of the kaon and the direction opposite to the B^0 ($\bar{B}^0$) momentum in the K^{*0} ($\bar{K}^{*0}$) rest frame. And ϕ is the angle between the plane containing the μ^+ and μ^- momenta and the plane containing the kaon and pion momenta, in the B^0 ($\bar{B}^0$) rest frame.

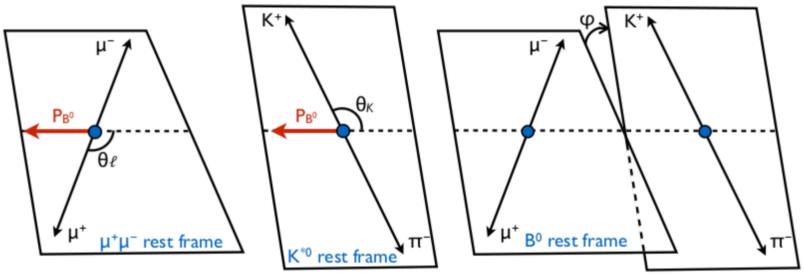


Figure 2.3: Graphical illustration of the three angular variables θ_l, $\theta_K \in [0,\pi]$ and $\phi \in [-\pi,\pi]$ used to describe the decay $B^0 \to K^{*0}(K^+\pi^-)\mu^+\mu^-$. Taken from Ref. [7].

The $B^0 \to K^{*0}\mu^+\mu^-$ decay is a realisation of the $b \to sl^+l^-$ transitions and its effective Hamiltonian has the form described in Eq. (2.17). The most relevant operators for this decay are $\mathcal{O}_7^{(\prime)}$, $\mathcal{O}_9^{(\prime)}$ and $\mathcal{O}_{10}^{(\prime)}$. $\mathcal{O}_7^{(\prime)}$ describe the electromagnetic penguin and $\mathcal{O}_9^{(\prime)}$ and $\mathcal{O}_{10}^{(\prime)}$ describe a penguin loop with a t-quark mediator in which the muon pair originates from a photon and from a Z boson, respectively. While the operators $\mathcal{O}_7$, $\mathcal{O}_9$ and $\mathcal{O}_{10}$ exist in the SM effective Hamiltonian, the chirality flipped $\mathcal{O}_7'$, $\mathcal{O}_9'$ and $\mathcal{O}_{10}'$ are introduced by NP.

In order to compute the decay rate, the amplitude $\mathcal{M} = \langle K\pi | \mathcal{H}_{eff} | \bar{B} \rangle$ needs to be obtained. The

$B \to K^{*0}$ matrix elements of the operators $\mathcal{O}^{(\prime)}_{7,9,10,P,S}$ can be expressed in terms of seven form factors which depend on the momentum transfer p^2 between the B and the K^{*0} mesons. Besides the terms proportional to form factors, the $B^0 \to K^{*0}\mu^+\mu^-$ amplitude also contains non-factorisable terms, which correspond to the matrix elements of the $\mathcal{O}_{1-6}$ and $\mathcal{O}_8$. In the heavy quark and large energy limit, they can be calculated using QCDF [42], which is only valid below the charmonium resonances (6 GeV^2).

The form factors are calculated using non-perturbative methods. Below 6 GeV^2, they are computed using Light Cone Sum Rules (LCSRs) [15]. These calculations are only valid in the phase space region where the K^{*0} final state has a large recoil momentum in the B^0 rest frame (low q^2), where physics can be described within the light-cone formalism. Above 6 GeV^2, lattice QCD calculations [16] are needed. They are currently restricted to a phase space in which the K^{*0} final state has a small recoil momentum in the B^0 rest frame (high q^2) since at low q^2 there are large discretisation and statistical errors [38]. The LCSRs and lattice QCD calculations are therefore complementary.

The $B^0 \to K^{*0}\mu^+\mu^-$ decay amplitude is given by

$$\begin{aligned}\mathcal{M} =&\frac{G_F\alpha_e}{\sqrt{2}\pi}\Bigg[\Bigg(\langle K\pi|\,\bar{s}\gamma^{\mu}(C_9^{\text{eff}}P_L + C_9'^{\text{eff}}P_R)b\,|\bar{B}\rangle - \frac{2m_b}{q^2}\langle K\pi|\,\bar{s}i\sigma^{\mu\nu}q_{\nu}(C_7^{\text{eff}}P_R + C_7'^{\text{eff}}P_L)b\,|\bar{B}\rangle\Bigg)(\bar{\mu}\gamma_{\mu}\mu)\\ &+\langle K\pi|\,\bar{s}\gamma^{\mu}(C_{10}^{\text{eff}}P_L + C_{10}'^{\text{eff}}P_R)b\,|\bar{B}\rangle\,(\bar{\mu}\gamma_{\mu}\gamma_5\mu) + \langle K\pi|\,\bar{s}(C_SP_R + C_S'P_L)b\,|\bar{B}\rangle\,(\bar{\mu}\mu)\\ &+\langle K\pi|\,\bar{s}(C_PP_R + C +_P P_L)b\,|\bar{B}\rangle\,(\bar{\mu}\gamma_5\mu)\Bigg],\end{aligned} \tag{2.25}$$

where $C^{\text{eff}}_{7,9,10}$ are combinations of WCs, as defined in Ref. [43] and $C'^{\text{eff}}_{7,9,10} = \frac{4\pi}{\alpha_s}C'_{7,9,10}$. Squaring $\mathcal{M}$ and summing over the spin of the final state particles, we obtain the decay rate of $\bar{B}^0 \to \bar{K}^{*0}\mu^+\mu^-$

$$\frac{d^4\Gamma}{dq^2d\cos\theta_l d\cos\theta_K d\phi} = \frac{9}{32\pi}I(q^2,\theta_l,\theta_K,\phi), \tag{2.26}$$

where

$$\begin{aligned}I(q^2,\theta_l,\theta_K,\phi) =& I_1^s\sin{\theta_K}^2 + I_1^c\cos{\theta_K}^2 + (I_2^s\sin{\theta_K}^2 + I_2^c\cos{\theta_K}^2)\cos\theta_l\\ &+ I_3\sin{\theta_K}^2\sin{\theta_l}^2\cos 2\phi + I_4\sin 2\theta_K\sin 2\theta_l\cos\phi\\ &+ I_5\sin 2\theta_K\sin\theta_l\cos\phi\\ &+ (I_6^s\sin{\theta_K}^2 + I_6^c\cos{\theta_K}^2)\cos\theta_l + I_7\sin 2\theta_K\sin\theta_l\sin\phi\\ &+ I_8\sin 2\theta_K\sin 2\theta_l\sin\phi + I_9\sin{\theta_K}^2\sin{\theta_l}^2\sin 2\phi.\end{aligned} \tag{2.27}$$

The corresponding expression for the $B^0 \to K^{*0}\mu^+\mu^-$ decay is given by

$$\frac{d^4\bar{\Gamma}}{dq^2d\cos\theta_l d\cos\theta_K d\phi} = \frac{9}{32\pi}\bar{I}(q^2,\theta_l,\theta_K,\phi), \tag{2.28}$$

where $\bar{I}(q^2,\theta_l,\theta_K,\phi)$ is obtained from Eq. (2.27) doing $I^{(n)}_{1,2,3,4,7} \to \bar{I}^{(n)}_{1,2,3,4,7}$ and $I^{(a)}_{5,6,8,9} \to -\bar{I}^{(a)}_{5,6,8,9}$, since after CP conjugation $\theta_l \to -\theta_l$, $\theta_K \to -\theta_K$ and $\phi \to \phi$. Therefore, the $B^0 \to K^{*0}\mu^+\mu^-$ decay is completely described by 12 angular parameters $\bar{I}_i^{(a)}$, that depend only on q^2.

In order to separate CP-conserving from CP-violating NP effects, one defines 12 CP averaged angular coefficients

$$S_i^{(a)} = \frac{I_i^{(a)} + \bar{I}_i^{(a)}}{\frac{d(\Gamma + \bar{\Gamma})}{dq^2}}, \tag{2.29}$$

as well as 12 CP asymmetries

$$A_i^{(a)} = \frac{I_i^{(a)} - \bar{I}_i^{(a)}}{\frac{d(\Gamma + \bar{\Gamma})}{dq^2}}. \tag{2.30}$$

Taking the CP average means that CP violating effects in $S_i^{(a)}$ are removed which results in a cleaner observable. Taking the CP asymmetry, on the other hand, means that any CP violation effect can be easily identifiable. In this study we will take the former approach.

We can define a set of parameters in terms of the $S_i^{(a)}$. The forward-backward asymmetry of the di-muon pair $A_{FB} = \frac{3}{8}(2S_6^s + S_6^c)$ as well as the transverse and longitudinal polarisation of the K^{*0} meson, $F_T = 4S_2^s$ and $F_L = -S_2^c$, respectively. These parameters depend on form factors and are prone to hadronic uncertainties coming from LCSRs and lattice QCD calculations. In order to reduce these theoretical uncertainties, a set of parameters $P_i^{(\prime)}$, that are independent of form factors at leading order, was constructed [44]: $P_1 = \frac{2S_3}{F_T}$, $P_2 = \frac{S_6^s}{2F_T}$, $P_3 = -\frac{S_9}{F_T}$, $P_4' = \frac{2S_4}{\sqrt{F_T F_L}}$, $P_5' = \frac{S_5}{\sqrt{F_T F_L}}$, $P_6' = -\frac{S_7}{\sqrt{F_T F_L}}$ and $P_8' = -\frac{2S_8}{\sqrt{F_T F_L}}$. In terms of these parameters, the CP-averaged $B^0 \to K^{*0}\mu^+\mu^-$ decay rate can be written as

$$\begin{aligned}
\frac{1}{d\Gamma/dq^2}\frac{d^4\Gamma}{dq^2 d\cos\theta_l d\cos\theta_K d\phi} =& \frac{9}{32\pi}\Bigg[\frac{3}{4}F_T \sin\theta_K{}^2 + F_L \cos\theta_K{}^2 \\
&+ (\frac{1}{4}F_T \sin\theta_K{}^2 - F_L \cos\theta_K{}^2)\cos 2\theta_l \\
&+ \frac{1}{2}P_1 F_T \sin\theta_K{}^2 \sin\theta_l{}^2 \cos 2\phi \\
&+ \sqrt{F_T F_L}(\frac{1}{2}P_4' \sin 2\theta_K \sin 2\theta_l \cos\phi + P_5' \sin 2\theta_K \sin\theta_l \cos\phi) \\
&\sqrt{F_T F_L}(P_6' \sin 2\theta_K \sin\theta_l \sin\phi - \frac{1}{2}P_8' \sin 2\theta_K \sin 2\theta_l \sin\phi) \\
&+ 2P_2 F_T \sin\theta_K{}^2 \cos\theta_l - P_3 F_T \sin\theta_K{}^2 \sin\theta_l{}^2 \sin 2\phi\Bigg].
\end{aligned} \tag{2.31}$$

This decay rate is used in the computation of the detector efficiencies, as will be explained in Ch. 6.

2.4 Global fits to the flavour anomalies and new physics

As explained in Ch. 1, the current individual tensions with the SM do not yet display sufficient levels of statistical significance. Nonetheless, when one combines the information from several independent measurements in global fits, the significance of the results increases and we can use this information to constrain NP scenarios. Regarding the FCNC transitions $b \to sll$, the tensions detected in the LFU observable $R_{K^{(*)}}$ seem to indicate the presence of NP in particular in the muon sector $b \to s\mu^+\mu^-$. NP can either be manifest by a modification of the WCs or by the introduction of new operators $\mathcal{O}_i$. The most favoured Lorentz structures for the operators are vector (V) and axial-vector (A), which in the effective Hamiltonian in Eq. (2.17) are described by the operators $\mathcal{O}_9^{(')}$ and $\mathcal{O}_{10}^{(')}$. $\mathcal{O}_9$ and $\mathcal{O}_{10}$ exist in the SM Hamiltonian but their WCs can be modified by NP: $C_9 = C_9^{SM} + C_9^{NP}$ and $C_{10} = C_{10}^{SM} + C_{10}^{NP}$. The chirality flipped operators $\mathcal{O}_9'$ and $\mathcal{O}_{10}'$ do not exist, however, in the SM but can be provided by NP.

For the global fits, several measurements are taken into account, both the ones where tensions with the SM have been reported and the ones which are consistent with it. In Ref. [8] some of the results used are: the angular analysis of $B^0 \to K^{*0}\mu^+\mu^-$ [3] and of $B^\pm \to K^{*\pm}\mu^+\mu^-$ [45]; the branching fraction of the $B_s^0 \to \mu^+\mu^-$ and $B^0 \to K^{*0}\mu^+\mu^-$ decays [46, 47]; the recent R_K measurement by LHCb [2] and the $B_s^0 \to \mu^+\mu^-$ lifetime (τ_{eff}) [46]. A distinction is made between theoretically clean observables, such as $R_K^{(*)}$ and $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ ("LFU,$B_s \to \mu\mu$") and the ones suffering from theoretical uncertianties, such as the differential branching fractions of the decays $B^\pm \to K^{*\pm}\mu^+\mu^-$, $B^0 \to K^{*0}\mu^+\mu^-$ and $B_s^0 \to \phi\mu^+\mu^-$ and their CP averaged angular parameters ("$b \to s\mu\mu$ observables").

The most economic solutions to the anomalies are the so-called "1D scenarios", where only one WC contributes or the values of two new WCs are related, so that there is only one extra parameter. Scenarios with only C_9^{NP}, $C_9^{NP} = -C_{10}^{NP}$ or $C_9^{NP} = -C_9'$ fit the data much better than the SM [48]. Another class of solutions are the "2D scenarios", where NP contributes to a pair of WCs and they are expected to give much better fits with respect to the SM and the 1D scenarios.

In Ref. [8], the best 1D scenarios are gauged according to their significance in σ defined as $\sqrt{\Delta\chi^2}$, where $\Delta\chi^2$ is the difference between the χ^2 value of the fit and the SM prediction. When considering only the "$b \to s\mu\mu$ observables", which could be prone to theoretical uncertainties, the best 1D fits have $C_9^{NP} \approx$ -0.87 or the left-handed combination $C_9^{NP} = -C_{10}^{NP} \approx$ -0.60. In these scenarios, the agreement between theory and data is improved by more than 4 σ with respect to the SM. When taking into account only the "LFU,$B_s \to \mu\mu$" observables, whose discrepancies cannot be explained by hadronic uncertainties, the scenarios $C_{10}^{NP} \approx$ 0.60 and $C_9^{NP} = -C_{10}^{NP} \approx$ -0.35 have a significance of 4.7 σ and 4.6 σ, respectively. Finally, when considering all the observables together, the significance increases to 5.7 σ and 5.9 σ for the $C_9^{NP} \approx$ -0.80 and $C_9^{NP} = -C_{10}^{NP} \approx$ 0.41 scenarios, respectively.

In Fig. 2.4, the two-parameter scenarios (C_9^{NP}, C_{10}^{NP}) and (C_9^{NP}, C_9') are presented in the left and right plots, respectively. In the left plot, the yellow band represents the 1 σ experimental limits coming from the $B_s^0 \to \mu^+\mu^-$ branching fraction. In the right plot, the pink bands represent the 1 σ and 2 σ limits from R_{K^*}. In both plots, the blue bands represent the 1 σ and 2 σ bands for a combination of the clean R_K and R_{K^*} observables, the orange bands represent the 1 σ and 2 σ constraints from the "$b \to s\mu\mu$ observables" and the red bands represent the result of the global fit. The SM prediction is at (0,0).

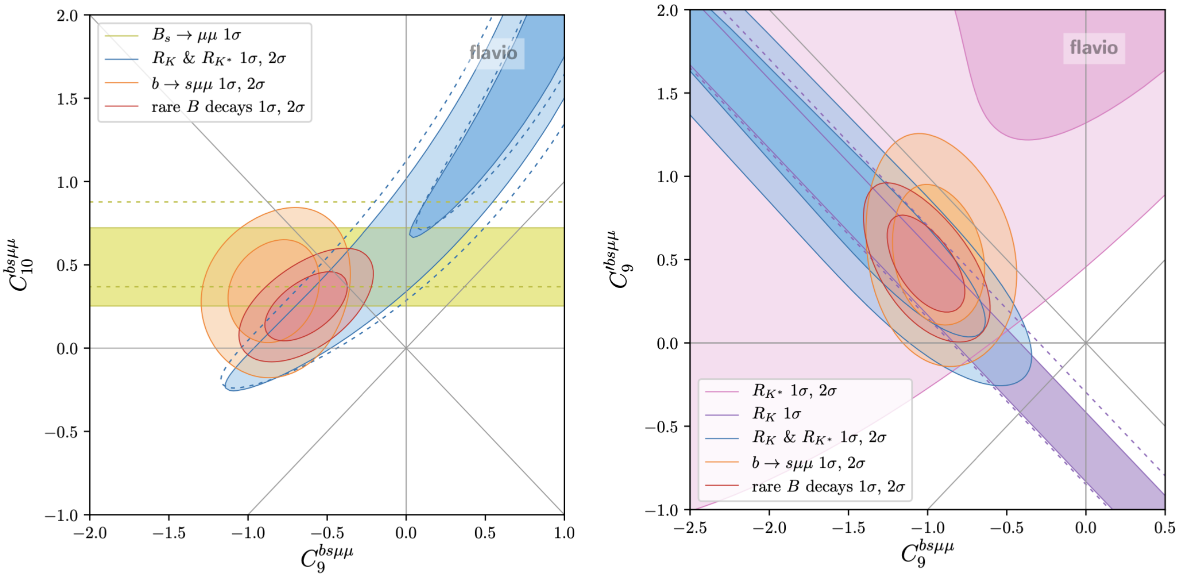


Figure 2.4: Constraints on the WCs in the C_9^{NP} vs C_{10}^{NP} (left) and in the C_9^{NP} vs C_9' (right) plane . The dashed (solid) lines represent the experimental limits before (after) Moriond 2021. The SM prediction is at the point (0,0). In red, the 1 σ and 2 σ contours of the global fit are depicted. Taken from Ref. [8].

The dashed (solid) lines represent the results before (after) the updates presented in Moriond 2021. In the (C_9^{NP}, C_{10}^{NP}) scenario, the $B_s^0 \to \mu^+\mu^-$ branching fraction prefers a slightly positive C_{10}^{NP}, whereas the "$b \to s\mu\mu$ observables" prefer a negative C_9^{NP}. The best fit point corresponds to $(C_9^{NP}, C_{10}^{NP}) \approx$ (-0.63,0.25) at 5.7 σ. In the (C_9^{NP}, C_9') scenario, the best fit point is $(C_9^{NP}, C_9') \approx$ (-1.01,0.47) at 5.9 σ.

Among the new physics models that try to explain some of the observed flavour anomalies, the most favoured ones include LQs or an heavier gauge boson Z'. LQs are hypothetical particles with both lepton and baryon quantum numbers that couple to leptons and quarks, therefore allowing for the FCNC $b \to sll$ transitions to occur at tree level. In light of the $R_{K^{(*)}}$ anomalies, LQs with different couplings to electrons, muons and/or taus have been theorised in order to accommodate LFU-violating interactions. The most favoured LQs are the scalar S_3 and vector V_1 and V_3, described in Refs. [25, 49], whose masses could be tens of TeV. Both S_3, V_1 and V_3 can induce the 1D scenario $C_9^{NP} = -C_{10}^{NP}$, where the WCs are proportional to the couplings of the LQs to leptons and quarks and inversely proportional to the LQs mass. The direct searches for these particles are being pursued at the LHC [50, 51]. The future FCC, with the envisaged collision energy of 100 TeV, will help to push these searches even further.

Another famous class of theories enlarges the SM gauge group with a $U(1)_X$ symmetry, with gauge boson Z' and different $U(1)_X$ charges for leptons and muons, therefore allowing for LFUV, as explained in Refs. [52, 53]. The Z' boson acquires mass by the SSB of $U(1)_X$ at TeV scales. In Ref. [53], Z' mediates the $b \to sll$ transitions via $\mathcal{O}_9$. The C_9^{NP} is proportional to the $U(1)_X$ coupling constant and inversely proportional to the Z' mass and from global fits to this WC the parameters of the model can be constrained. In Ref. [52], the Z' gauge boson can contribute to the WCs C_9^{NP} and C_{10}^{NP} and the parameters of the model can be constrained from global fits in the C_9^{NP} vs C_{10}^{NP} plane. Direct searches for these heavy gauge bosons are also being performed at the LHC [54].

We conclude this chapter with the mindset that there is still a lot about the world that we do not know and it is with the combined efforts from the experimental and theoretical communities that we can push our understanding further. As Feynman himself said ”Nature’s imagination is so much greater than man’s, she will never going to let us relax”.

Chapter 3

CMS detector and online selections

In this chapter, we start by describing how particles are collided at the LHC in Sec. 3.1. We then focus on one of the LHC experiments, CMS, and give an account of the main features of its sub-detectors in Sec. 3.2. We also present a description of its trigger system in Sec. 3.3, describing the trigger (online) selections used in the analysis.

3.1 The Large Hadron Collider

The LHC is a particle accelerator at CERN that pushes protons or heavy ions to near the speed of light. It consists of a 27 km ring of superconducting magnets with accelerating structures that increase the energy of the particles along the way.

The CERN accelerator complex consists of a succession of machines with increasing higher energies. Each of them accelerates a beam of particles to a given energy before injecting them to the next one. The LHC is the last machine in the chain, where particles achieve their highest energies. Inside the LHC, the two particle beams travel in opposite directions in separate beam pipes and are made to collide in the center of the particle detectors. They are guided around the accelerator ring by a strong magnetic field maintained by superconducting electromagnets. These magnets are kept at -271.3° C, at a temperature colder than outer space, so that they can be in a superconducting state, offering no resistance to the passage of electric current. The accelerator is connected to a vast distribution of liquid helium which cools the magnets.

There are currently 8 experiments at the LHC which use detectors to analyse the particles produced by the collisions. The 9th experiment, SND@LHC, has been approved in March 2021 and is being comissioned, having the goal of detecting and studying neutrinos. The proton beams are collected in each detector in packets called bunches. The collision rate is proportional to the instantaneous luminosity of the accelerator, which, under the assumption of gaussian beams, is defined as:

$$\mathcal{L} = \frac{f N_b n_p^2}{4\pi\sigma_x\sigma_y}, \tag{3.1}$$

where f is the revolution frequency of the bunches, N_b is the number of bunches, n_p is the number of protons in each bunch and σ_x, σ_y are the proton beam's transverse dispersion along the x and y axis. The total integrated luminosity is defined as $L = \int \mathcal{L}dt$. The integrated luminosity delivered by CMS in pp collisions over the Run 2 data-taking years can be seen in Fig. 3.1 (left): L =4.3 fb^{-1} (2015), L =41.6 fb^{-1} (2016), L =49.8 fb^{-1} (2017) and L =67.9 fb^{-1} (2018). In our work, we use data collected during the years 2016, 2017 and 2018, corresponding to a total integrated luminosity of L =139.5 fb^{-1}.

With the increase of luminosity during Run 2, there was a consequent increase of pileup, i.e. the number of unwanted extra collisions that overlap in the detector and hide rare processes that are worthwhile studying. The CMS average pileup in pp collisions $< \mu >$ during the Run 2 data-taking years can be seen in Fig. 3.1 (right): $< \mu >$=13 (2015), $< \mu >$=27 (2016), $< \mu >$=38 (2017) and $< \mu >$=37 (2018). In order to cope with this pileup increase, several upgrades in the pixel detector and trigger system were performed [55, 56] (see Secs. 3.2.3 and 3.3).

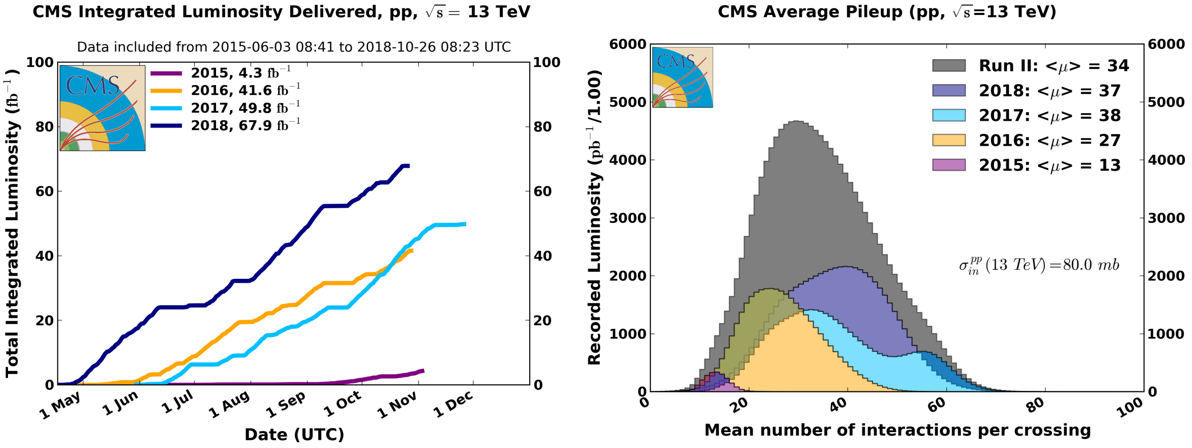


Figure 3.1: CMS integrated luminosity (left) and average pileup (right) in pp collisions during Run 2. Taken from Ref. [9].

3.2 The Compact Muon Solenoid detector

CMS is a general purpose detector situated at interaction point 5 of the LHC, designed and built to study physics at the TeV scale. A schematic representation of the detector can be seen in Fig. 3.2.

It has a cylindrical shape and is built around a 6 m diameter, 13 m long superconducting magnet that generates a magnetic field of $\approx$ 4 T, around 100.000 times stronger than the magnetic field of the Earth.

The solenoid is large enough to accommodate the tracking and calorimeter systems inside. Outside of it there are the muon stations. The complete detector is 21 m long, 15 m wide and 15 m high.

The different CMS sub-detectors will be described in the next sections. A more complete description can be found in Ref. [57]. For this analysis, the main sub-detectors used are the silicon tracker and the muon detector, described in Secs. 3.2.3 and 3.2.6, respectively. These detectors are divided into a barrel and two endcap sections.

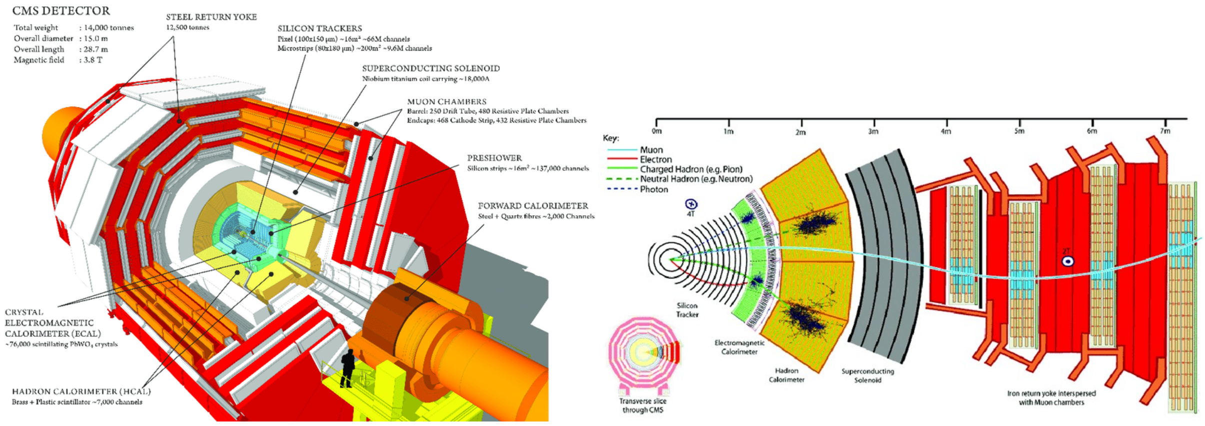


Figure 3.2: Schematic representation of the CMS detector identifying dimensions and different constituents (left) and transverse view identifying the layers where each type of particle is detected (right).

3.2.1 Coordinate system

The CMS experiment uses a right-handed coordinate system with origin at the collision point. The x axis points to the centre of the LHC ring, the y axis points vertically upwards, perpendicular to the LHC plane, and the z axis points along the anti-clockwise beam direction. The polar angle θ is measured from the positive z axis, the azimuthal angle ϕ is measured from the positive x axis in the x-y plane and the radius r measures the distance from the z axis. The pseudorapidity η is related with the polar angle θ by the relation $\eta = -\ln\left[\tan\left(\theta/2\right)\right]$. For $\theta = 90°$, the particle is moving in the transverse plane and $|\eta| = 0$. The closest to zero θ is, the largest will $|\eta|$ be and the particle will move closer and closer to the beam direction (z axis). This region with large $|\eta|$ values ($|\eta| > 1$) is called the forward region.

Other variables of interest are the particle's transverse energy (E_T) and momentum (p_T). The reason why we work with the transverse energy/momentum is because we do not know what is the fraction of longitudinal momentum that quarks take from the colliding protons. We know, however, that the sum of the total transverse momentum of the particles in the final state should be zero, due to momentum conservation. This helps us to detect missing momentum in the events, which could be due to neutrinos or a new weakly interacting particle. The rapidity, defined as $y = \frac{1}{2}\ln\left(\frac{E+p_z}{E-p_z}\right)$, where E is the particle energy and p_z the component of its momentum along the beam direction, is also often used.

3.2.2 Magnet system

CMS has a single large superconducting solenoid, which consists of a cylindrical coil of wire that, in the presence of an electric current, generates an axial magnetic field of $B =$ 3.8 T. To achieve the high magnetic field, a current of 20 kA is necessary. The wires are made of superconducting strands coated with copper. Each strand can carry more than 2000 A when cooled to 4.4 K using liquid helium.

As the superconducting solenoid needs to operate at liquid helium temperature, it needs to be maintained within a vacuum vessel to isolate it from the exterior. The solenoid also requires a return yoke to control the field outside of it, which is represented in red in Fig. 3.2, and it constitutes the main support ("skeleton") of the CMS detector. The field strength in the return yoke is $B \approx$ 2 T.

3.2.3 Silicon tracker

The silicon tracker detects charged particles within $|\eta| < 2.5$, which leave a trail in this sub-detector referred to as a track. The pixel detector is situated in the innermost region of the CMS detector and is composed of four layers in the barrel region at radii 29, 68, 109 and 160 mm and three disks located in the forward regions at radii 291, 396 and 516 mm from the centre of the detector. It is built from 1856 silicon sensor modules with 160×416 pixels connected to 16 readout chips. Further from the interaction region is a microstrip detector, composed of 10 barrel layers and 3 inner and 9 outer disks on either end of the detector. In total, the microstrip detector contains around 10 million strips and, together with the pixel detector, has an impact parameter resolution of about 15μm [58].

When pp bunches collide, they produce a luminous region in the detector called the beamspot. Several primary vertices are also created, which represent the location of a pp collision. In the case of b hadrons, since they are relatively long-lived ($\tau \approx 10^{-12}$ s) they travel a distance of a few millimetres in the detector (see Fig. 3.3). As a consequence, their primary vertex is displaced from their secondary vertex, i.e. the region in space where they decay, which facilitates their reconstruction.

The tracks ($h^{\pm}$) are reconstructed from the hits detected in the pixel and microstrip detectors. CMS uses a Combinatorial Track Finder (CTF) algorithm [59] to determine their trajectories[1]. The pixel detector is close to the interaction point to measure the position of the vertices accurately and to seed the tracks. The track seeds serve as the initial direction for the track reconstruction algorithm, which then searches for additional hits that can be associated with the track candidate, until the track reaches the end of the tracker, contains too many missing hits or if its p_T falls below a certain value.

---

[1] Given that the magnetic field inside the tracker is approximately constant, charged particles have an helitical trajectory, which is defined by 5 parameters: d_0, z_0, ϕ, $\cot\theta$ and p_T; where (x_0, y_0, z_0) are the coordinates of the collision point, $d_0 = -y_0 \cos\phi + x_0 \sin\phi$ is the distance of closest approach (DCA) of the track to the beam axis and θ and ϕ are the angles defined in Sec. 3.2.1.

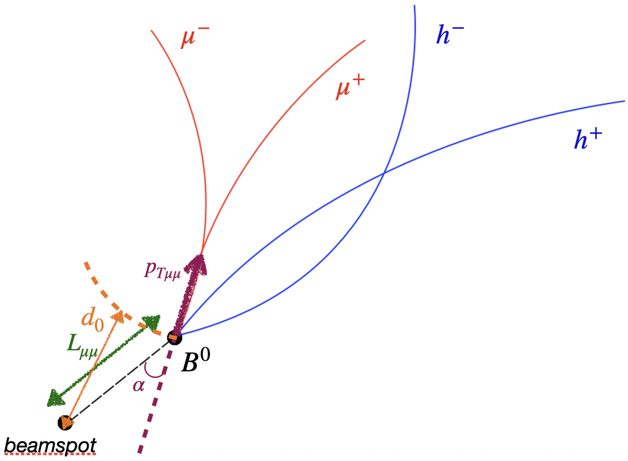


Figure 3.3: Illustration of the B^0 decay and some of the variables that are used in the selections.

The transverse momentum of the track candidate is obtained with the knowledge of the magnetic field inside the solenoid B and the radius of curavature of its trajectory R, $p_T = qBR$, where q is the electric charge of the track. As a consequence of the high granularity of the silicon tracker and the strong and homogeneous magnetic field, a transverse momentum resolution of about 1.5 % is achieved for muons. The uncertainty in the track reconstruction efficiency for charged hadrons is estimated to be 2.3 % in Run 2 [60]. The branching fraction normalisation in Eq. (1.1), allows the absence of this uncertainty in the measurement, since the error in the ratio ϵ_N/ϵ_S is negligible.

The vertices are reconstructed in two steps. The first one involves grouping several tracks into vertex candidates (vertex finding), whereas the second one involves determining the best estimate for the vertex parameters for a given set of tracks (vertex fitting). The probability of the tracks coming from the same interaction vertex is quantified by means of the χ^2 test (see Ch. 5).

The pixel detector at work during Run 2 was installed in 2016/17 as part of the CMS Phase I upgrade [55], which allowed the performance of the pixel detector to be maintained or even increased in the new Run 2 high event rate environment. A comparison between the tracking efficiency (left) and the number of reconstructed tracks that do not correspond to a charged particle (fake tracks) (right), between the track reconstruction with the original (blue) and upgraded (orange) pixel detector used in years 2016 and 2017, respectively, can be seen in Fig. 3.4. A considerable increase in the tracking efficiency is observed, particularly in the forward regions. A reduction in the fake rate was observed in all $|\eta|$ regions.

3.2.4 Electromagnetic calorimeter

The Electromagnetic Calorimeter (ECAL) measures the energy of incident electrons/positrons and photons. It performs measurements of the incident position and tries to distinguish single photons from pairs of closely spaced photons coming from decays of neutral pions (π^0 rejection). The ECAL is made of a

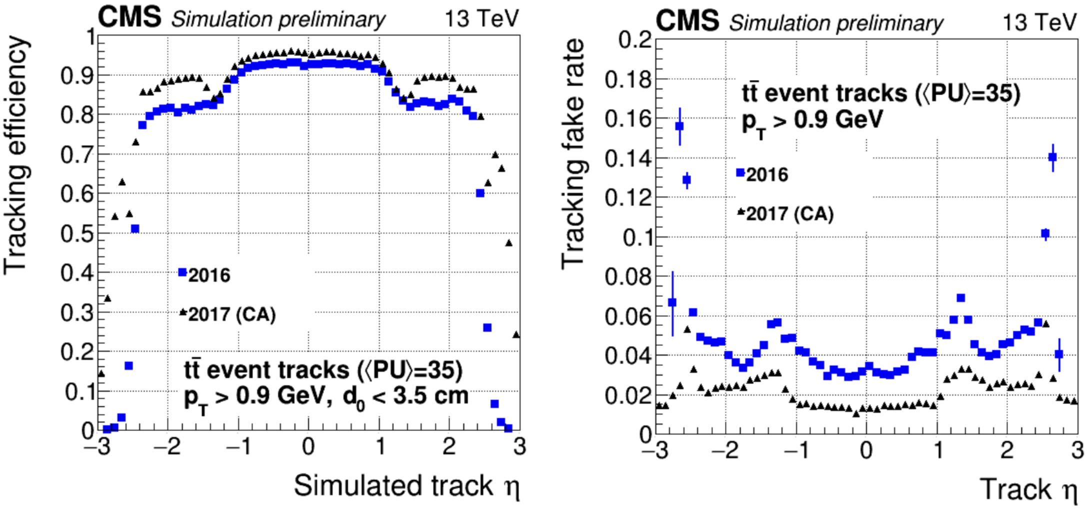


Figure 3.4: Tracking efficiency (left) and fake rate (right) of the original pixel detector (blue) and the upgraded one (black), used in years 2016 and 2917, respectively. Taken from Ref. [10].

crystal volume which consists on a truncated pyramid of lead tungstate (PbWO$_4$). The incident electrons and photons deposit almost all their energy in the crystals producing scintillating light, whose amount is a linear function of the particle's energy. Photomultipliers convert this light intro electrical signals.

In addition to the crystals there is a detector in the endcaps, called endcap preshower, whose primary function is π^0 rejection by measuring the transverse profile of electromagnetic showers[2]. It is a fine granularity detector in order to be able to have enough resolution to distinguish the two photons coming from the π^0 decay, whose separation is a few millimeters. The preshower is a sampling calorimeter[3] with two thin layers of lead (to generate the showers) followed by a layer of silicon strip sensors (to produce and detect the emitted light).

3.2.5 Hadronic calorimeter

The Hadronic Calorimeter (HCAL) surrounds the ECAL and its main purpose is the identification of quarks, gluons and neutrinos by measuring the energy and direction of jets and of missing transverse energy in the events, which is also a crucial signature of NP particles. For a good missing energy resolution, the calorimeter needs to be as hermetic as possible. The CMS HCAL is hermetic up to $|\eta| = 5$. The HCAL aids in identifying electrons, photons and muons in conjunction with other sub-detectors.

The barrel and endcap HCAL are sampling calorimeters made from thick brass plates interconnected with a thick plastic scintillator. The scintillators emit blue-violet light, whose amount is proportional to the energy of the incident hadron. The two forward HCALs are also sampling calorimeters, made from a

---

[2] Electromagnetic showers are a cascade of e^+e^- pairs and photons. It starts with a high energetic photon or electron. The photons create e^+e^- pairs via pair production and the electrons emit photons via bremsstrahlung radiation.

[3] In a sampling calorimeter, the material that produces the particle shower is different from the one that detects the emitted light.

large steel absorber (to generate the showers) embedded with quartz fibres. The large absorber is necessary given the high radiation present at high $|\eta|$. The energy of jets is measured from Cerenkov light[4] produced as charged particles traverse the quartz fibres.

3.2.6 Muon system

The muon system is designed to detect muons within $|\eta| < 2.4$, with four muon stations interleaved with the iron return yoke, to make full use of the magnetic field in this region ($B = 2$ T). Each station consists of several layers of drift tubes and cathode strip chambers in the regions $|\eta| < 1.2$ and $0.9 < |\eta| < 2.4$, respectively. They are complemented with resistive plate chambers covering the range $|\eta| < 1.6$.

The muons are reconstructed independently in the silicon tracker (muon tracker track) and by combining hits in the three muon sub-detectors (standalone muon). They are subsequently combined with tracks found in the silicon tracker to form global muons. For a global muon, a standalone muon is matched to a track by comparing their trajectory parameters after propagation to a common surface at the innermost muon station of the reconstructed standalone muon track.

3.3 The CMS Trigger system

Events are selected using a two-level trigger system composed by a level 1 (L1) trigger and a High Level Trigger (HLT), which reduce the event rate to around 100 kHz and 1 kHz, respectively, before data storage. The L1 trigger is made from custom electronics (hardware trigger) and uses coarse information from the calorimeters and from the muon detectors to try to select the most interesting events, e.g. events with high p_T muons. The information from the sub-detectors is held in pipeline memory buffers on front-end electronics.

The HLT relies on $\approx$ 100 commercial processors (software trigger) and is the first trigger level that has access to the complete silicon tracker information, allowing more precise selections to be made online. It performs the readout of the front-end electronics after the L1 trigger and executes physics selection algorithms on the read events, in order to accept the ones with the most interesting physics content. The HLT contains many paths (HLT paths), each applying a different set of selection strategies that are more suitable to a certain physics analysis. A more complete description of the CMS trigger system can be found in Ref. [61].

During Run 2, the LHC reached a peak instantaneous luminosity of $\approx 2 \times 10^{34}$cm^{-2}s^{-1} with an average pileup of $<\mu>= 34$. Under these conditions, the online event selection is a very challenging task. In order to cope with this, the L1 trigger was completely upgraded [56]. The upgraded system as a lower muon trigger rate while maintaining the same efficiency. In Fig. 3.5 (left), the L1 muon trigger

---

[4] Cerenkov light is emitted whenever a charged particle travels faster than the speed of light inside a given material.

efficiency as a function of the reconstructed p_T can be seen for different $|\eta|$ ranges, using 2017 events. In Fig. 3.5 (right), the ratio between the trigger rates as a function of the L1 muon $|\eta|$, between the upgraded trigger (2017) and the old trigger (2015) can be seen. The upgraded trigger has rates 20-80 % lower across the entire η range.

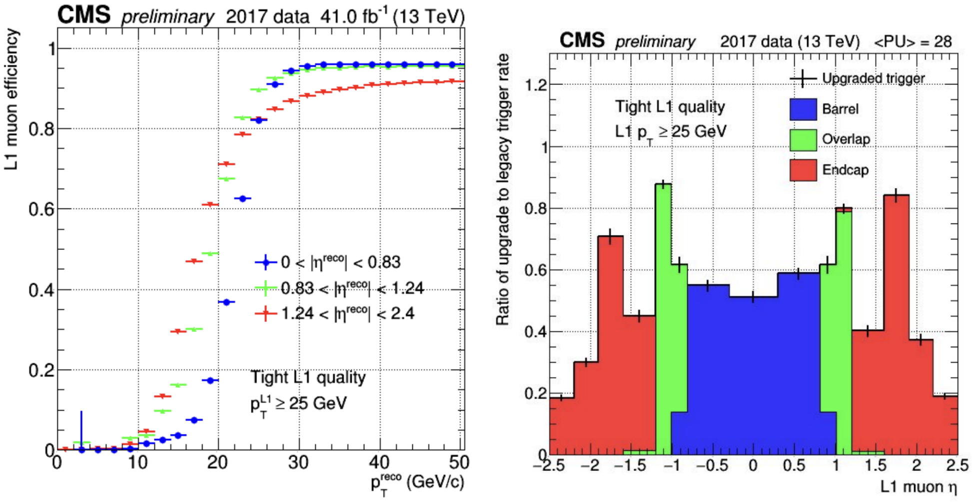


Figure 3.5: L1 muon trigger efficiency as a function of the transverse momentum of the reconstructed muon candidate (left) for different $|\eta|$ regions (2017). Ratio between the trigger rates between the upgraded (2017) and old (2015) trigger as a function of the L1 muon η (right). Taken from Refs. [11, 12]

In our analysis, we use three HLT paths with the same requirements but covering different ranges of the di-muon invariant mass: [2.9-3.3] GeV (J/ψ); [3.3-4.05] GeV ($\psi(2S)$) and [1-2.9] $\cup$ [4.0-4.8] GeV (non-resonant channel). This set of triggers has been active during the full Run 2 and underwent small modifications during the three years. A mistake in the code of one of the di-muon filters prevented one of the paths to select the events in the high di-muon invariant mass region [4.0-4.8] GeV. This implies the absence of these events in this region for the majority of 2017 and the whole of 2018 data samples. In the current version of the analysis we are not, therefore, measuring the differential branching fraction in this mass region.

The events used in the analysis were collected with triggers requiring that two opposite charge muons and an additional track are found in the event and form a displaced 3-body vertex. The following requirements are applied at the HLT:

- single muon $p_T >$ 4 GeV and $|\eta| <$ 2.5;
- di-muon $p_T >$ 6.9 GeV;
- di-muon vertex confidence level $>$ 0.1;
- di-muon invariant mass in the window specified in the text;

- DCA between the two muon momenta < 0.5 cm;
- DCA between the muon momentum and the beamspot in the transverse plane < 2 cm;
- $|\vec{L}_{\mu\mu}|/\sigma_{\mu\mu} > 3$, where $|\vec{L}_{\mu\mu}|$ is the distance in the transverse plane between the di-muon vertex and the beamspot and $\sigma_{\mu\mu}$ is the corresponding uncertainty (see Fig. 3.3);
- $\cos\alpha > 0.9$, where α is the pointing angle, in the transverse plane, between the di-muon momentum vector and $\vec{L}_{\mu\mu}$ (see Fig. 3.3);
- p_T of the additional track > 0.8 GeV and $|\eta| < 2.5$;
- di-muon plus track vertex normalised $\chi^2 < 10$.

In 2017 and 2018, the p_T of the track was required to be greater than 1.2 GeV and a requirement on the significance of the track DCA from the beamspot was introduced (DCA significance > 2).

Chapter 4

Datasets and offline selections

In this chapter, we start by defining the data and simulation samples used in the analysis in Sec. 4.1. In Sec. 4.2, we present the baseline selection of the analysis and finally, in Sec. 4.3, we present the sources of background studied and the optimised selections that are used to reject them.

4.1 Data and simulation samples

In this analysis we use data from pp collisions at the LHC at $\sqrt{s}$ = 13 TeV, collected by the CMS detector during the years 2016, 2017 and 2018, corresponding to a total integrated luminosity of L = 139.5 fb^{-1}. Besides data samples, the analysis also uses Monte Carlo (MC) simulated samples. The MC event generator is a program which simulates particle physics events with the same probability as they occur in nature. There are several such event generators, among them `PYTHIA` v.8 [62].

The collision event simulation starts by identifying a hard process[1] and generating the kinematics of the initial and final state particles. It then evolves the final state, including the decays of heavy quarks, the evolution of the parton shower and the hadronisation of partons to form final state hadrons. In our analysis, decays of particles containing b or c quarks are simulated with the `EVTGEN` package [63]. Creation of charged particles during the decays of B mesons can cause some energy to be radiated through photons, which is included via the `PHOTOS` package [64].

The response of the CMS detector, including the interaction of particles with matter and the different sub-detectors described in Ch. 3, is simulated with `GEANT4` [65]. Pileup events are added to represent the number of multiple vertices per event as seen on data (pileup weight). The simulated samples are used to estimate background contributions, as will be explained in Sec. 4.3, to compute the detector efficiency and perform studies of systematic uncertainties, as will be explained in Chs. 6 and 8.

---

[1] An hard process is an interaction with a high momentum transfer.

4.1.1 Data-taking conditions

During the three years, a few important changes in the data-taking conditions occurred. In the first part of the 2016 data-taking period, the Endcap Muon Track Finder (EMTF) was misconfigured and it was only assigning the measured p_T to the highest-quality muon track in a 60° sector in ϕ. Consequently, even if two muons were found in the same endcap and within 60° in ϕ from each other, they could not fire a double muon $L1$ seed, because one of them always had $p_T = 0$. This introduces a q^2 dependence on the measured event yield which affects the branching fraction result. Also during 2016, the strip tracker was operated with suboptimal settings, causing the tracking efficiency to decrease with increasing instantaneous luminosity. This effect has been found to be caused by High Ionising Particles (HIP) that saturate the hit strips, which become fully blinded in the next bunch crossings and then only partially until full recovery. Events from these problematic periods of 2016 are not, therefore, used in this analysis. The effect of this removal in the precision of the final results is studied in Sec. 5.5.

As explained in the previous chapter, in 2017 a new pixel detector was installed. In October 2017, a series of DC-DC converters powering the pixel detector broke when modules were power-cycled [66].

4.1.2 Fiducial region

The fiducial region of the analysis i.e. the region of phase space that is covered by the analysis, is defined by the presence of a B^0 candidate with $|\eta_B| < 3$, where η_B is the pseudorapidity of the B^0 meson. For $|\eta_B| > 3$, the probability of detecting the four final state particles that result from the B^0 decay is very low and therefore a filter to these events is used in the MC simulations in order to save computing time and disk space. This cut does not introduce any q^2 dependence on the efficiencies computed in Ch. 6. The 2D scatter plot of the B^0 transverse momentum (p_T) vs pseudorapidity (η) for the generated events lying in the fiducial region can be seen in Fig. 4.1.

4.1.3 Acceptance cuts

The detector acceptance is defined by the region of phase space within which the final state particles are potentially detectable. Particles with too large pseudorapidities, move very close to the beam axis and escape the CMS detector without leaving a signal in the sub-detectors. Moreover, they need to have a minimum p_T in order to be able to leave a significant amount of hits in the silicon tracker and, in the case of muons, in the muon stations so that they can be correctly reconstructed.

Therefore, a set of requirements on the final state particles is applied in the generation level MC in order to reject events falling outside the acceptance of the detector

$$|\eta(\mu^+,\mu^-,K^\pm,\pi^\mp)| < 2.5,\ p_T(\mu^+,\mu^-) > 2.5\,\text{GeV},\ \ p_T(K^\pm,\pi^\mp) > 0.4\,\text{GeV}. \tag{4.1}$$

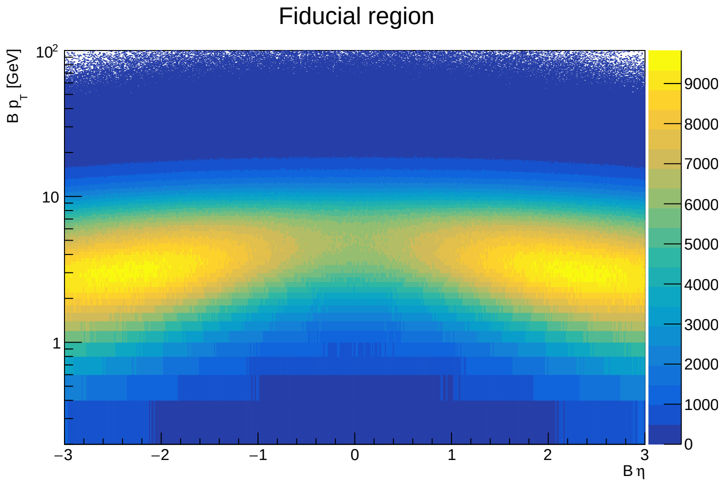


Figure 4.1: 2D scatter plot of the B^0 transverse momentum (p_T) vs pseudorapidity (η) for events in the fiducial region of the analysis.

4.2 B reconstruction and baseline selection

The signal and control channels $B^0 \to K^{*0}\mu^+\mu^-$, $B^0 \to J/\psi K^{*0}$ and $B^0 \to \psi(2S)K^{*0}$ are reconstructed through their decay into the fully charged final state $K^+\pi^-\mu^+\mu^-$. Only events passing the triggers described in Sec. 3.3 are considered. B^0 candidates are formed by combining two opposite charge muons with a $K^{*0}(892)$ meson. The two muons and at least one track should be matched to the objects firing the HLT. The selections applied at HLT are re-applied in the offline analysis.

Muons are required to pass the following criteria:

- single muon $p_T >$ 4 GeV and $|\eta| <$ 2.5;
- muon tracker track matched with at least one muon segment in both x and y coordinates;
- number of track layers with measurements $>$ 5;
- number of pixel layers with measurements $>$ 0;
- muon track passing the high-purity flag [59] in order to reject fake tracks;
- transverse (longitudinal) impact parameter $<$ 0.3 (20) cm with respect to the primary vertex.

The two hadron tracks are required to satisfy the following selections:

- $p_T >$ 0.8 GeV and $|\eta| <$ 2.4; in 2017 and 2018, the p_T of the offline track matched to the one firing the trigger is required to be $>$ 1.2 GeV;
- pass the high-purity flag [59] in order to reject fake tracks;

- DCA of the track with respect to the beamspot in the transverse plane (d_0) significance greater than 0.8; in 2017 and 2018, the DCA significance of the track matched to the one firing the trigger is required to be $>$ 2 (see Fig. 3.3);
- tracks reconstructed as global muons which have at least 1 pixel layer with measurement and 6 tracker layers with measurements are discarded;
- tracks which are also reconstructed as muon tracker tracks are discarded.

We apply the following requirements on the K^{*0} meson:

- the two tracks must successfully fit to a common vertex;
- the K^{*0} invariant mass is computed for both $K^+\pi^-$ and $K^-\pi^+$ mass hypotheses, and at least one of the two combinations is required to lie in a 3 σ mass window from the K^{*0} nominal mass [38], where σ is its natural width (50 MeV);
- $m_{KK} >$ 1.035 GeV, where m_{KK} is the invariant mass of the two hadron tracks with kaon mass assigned. This cut was studied in the Run 1 analysis [29] and is useful to reject $B_s^0 \to J/\psi(\mu^+\mu^-)$ $\phi(K^+K^-)$ events.

The two muon and two hadron candidates are fitted to a common vertex and a minimum χ^2 probability of 1 % is required.

4.2.1 Flavour-tag assignment and mistag fraction

The decay $B^0 \to K^{*0}(K^+\pi^-)\mu^+\mu^-$ and its CP-conjugate $\bar{B}^0 \to \bar{K}^{*0}(K^-\pi^+)\mu^+\mu^-$ both occur in the collisions and they leave the same signature in the CMS detector: two opposite charge tracks $(h^\pm)$ and two opposite charge muons $(\mu^\pm)$ (see Fig. 3.3). Therefore, a flavour-tag assignment on the two opposite charge tracks composing the final state particles is necessary to distinguish between the two combinations: $(h^+, h^-) = (K^+, \pi^-)$ or $(h^+, h^-) = (\pi^+, K^-)$. As described in Sec. 4.2, the invariant masses of both the $K^+\pi^-$ and $K^-\pi^+$ combinations is computed and saved. The flavour-tag assignment consists on selecting the combination whose invariant mass is closest to the K^{*0} nominal mass [38], thereby determining the CP state of the decay.

This assignment does not always result in the correct flavour for the tracks and, consequently, there is a fraction of mis-tagged events in our signal. Henceforth, events with the correct flavour will be referred to as right-tagged (RT) events and the ones with mis-tagged flavour will be referred to as wrong-tagged (WT) events. The mis-tag fraction (f^M) is defined as the ratio between the number of WT events and the total number of signal events and is determined from simulation. The mis-tag fractions are around 12-13%, depending on the q^2 bin and year.

4.3 Background studies and optimised selections

In this section, we report the selections that have been applied in order to reduce or completely remove background decays contaminating our signal. A multivariate analysis [13] was used to optimise the background rejection power, as will be described in Sec. 4.3.1.

A source of background referred to as combinatorial background is present in every physics analysis. It describes events in which the four particles in the final state do not all come from the same vertex. Consequently, the 4-body invariant mass can have a random value. This constitutes the main source of background in our analysis and needs to be modelled accordingly, as explained in Ch. 5.

4.3.1 Multivariate analysis

A multivariate analysis [13] was performed in order to reduce the amount of background, using a Boosted Decision Tree (BDT) discriminator to separate signal from background events. A decision tree has several branches and on each branch binary decisions are done. The training proceeds iteratively until some figure of merit is optimized and one has a signal and a background sample. The optimised figure of merit was the Punzi pseudo-significance [67].

The training was performed on signal MC and on background events from the data[2] sidebands. Sidebands are defined as events with m_{B^0} in the range $[m_{B^0}^{PDG} - 7\sigma, m_{B^0}^{PDG} - 3\sigma]$ (left) and $[m_{B^0}^{PDG} + 3\sigma, m_{B^0}^{PDG} + 7\sigma]$ (right), where σ is the B^0 signal width measured on MC, and are regions dominated by background events. Data and MC samples were split into 11 sub-samples, 10 for training and testing and 1 to use as analysis data, where the BDT score is generated. This process is repeated 11 times and each sub-sample is used as the analysis data exactly once. The BDT response on the signal and background samples for year 2018 can be seen in Fig. 4.2 (left).

The BDT score cut to be used in the analysis was chosen by maximising the expected $S/\sqrt{S+B}$ in the signal region, where S is the signal yield extracted from a fit to the signal MC sub-sample and B is the background yield extracted from a fit to the left and right data sub-sample sidebands, extrapolated to the signal region. In Fig. 4.2 (right), the $S/\sqrt{S+B}$ values as a function of the BDT score, evaluated on the 11 sub-samples and their average can be seen for year 2018. The working points giving the best $S/\sqrt{S+B}$ values are: BDT score $>$ 0.99 (2016), $>$ 0.97 (2017) and $>$ 0.975 (2018).

4.3.2 Peaking background contributions

Peaking backgrounds are decays that do not correspond to our signal and can result from e.g. other b-hadron decays. They are not of combinatorial nature and can give rise to structure in the mass spectrum. The resonant $B^0 \to J/\psi K^{*0}$ channel was used to evaluate residual contamination from this type of

---

[2] Events from the resonant channels are not used in the optimisation process.

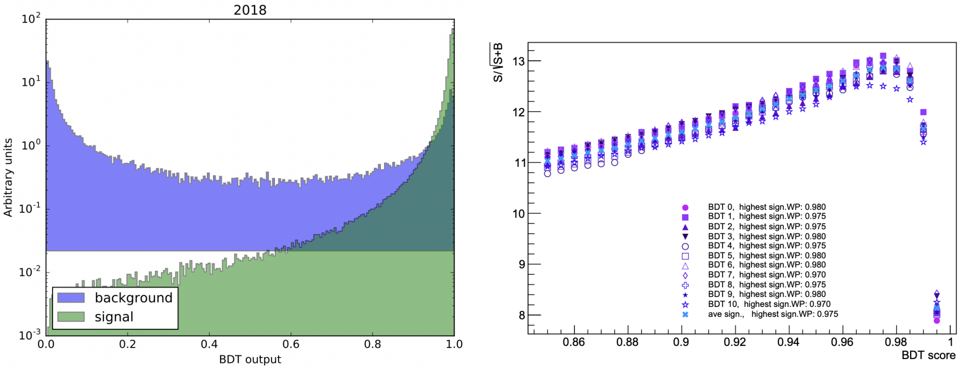


Figure 4.2: BDT response on signal and background samples for year 2018 (left). $S/\sqrt{S+B}$ as a function of the BDT score calculated in the 11 sub-samples and their average for year 2018 (right). Taken from Ref. [13].

decays. Simulated samples of $B^+ \to J/\psi K^+$ and $B^0_s \to J/\psi\phi$ decays were used. The $K^+\pi^-\mu^+\mu^-$ invariant mass from the simulated samples and the data is shown in Fig. 4.3 (left).

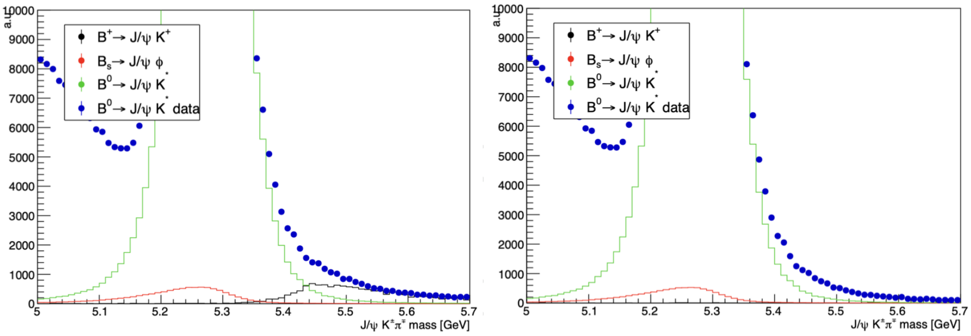


Figure 4.3: $K^+\pi^-\mu^+\mu^-$ invariant mass distribution after the event selections (left) and after the additional cut on $m(\mu^+\mu^-K^+)$ and $m(\mu^+\mu^-)$ (right) are applied, for events in q^2 bin 4, for data (blue points), $B^0 \to J/\psi K^{*0}$ MC (green), $B^+ \to J/\psi K^+$ MC (black) and $B^0_s \to J/\psi\phi$ MC (red). MC distributions are normalised to data luminosity. Taken from Ref. [13].

A small contamination from B^0_s decays is present in the signal region; however, it is very small compared to the B^0 signal and is considered to be negligible. On the other hand, $B^+ \to J/\psi K^+$ events significantly pollute the right sideband, while being absent in the signal region. The removal of these events is necessary in the angular analysis [13] in order to guarantee that the distribution of events lying in the sidebands correctly models the background under the signal.

The rejection of $B^+ \to J/\psi K^+$ events is achieved through a 2D requirement on the invariant mass of the $\mu^+\mu^-K^+$ and $\mu^+\mu^-K^-$ combinations, where K^+ and K^- are the two tracks forming the K^{*0} meson, having the kaon mass assigned. The 2D distributions of $m(\mu^+\mu^-K^+)$ vs $m(\mu^+\mu^-K^-)$ from the $B^+ \to J/\psi K^+$ and $B^0 \to J/\psi K^{*0}$ MC samples are shown in Fig. 4.4. The 1D $\mu^+\mu^-K^+$ and $\mu^+\mu^-K^-$ distributions for $B^+ \to J/\psi K^+$ events after the event selections are fitted with a gaussian, whose width,

σ_{fit}, is used to define the mass cut. Events are rejected if

$$|m(\mu^+\mu^-K^+) - m_{B^+}^{PDG}| < 3\sigma_{fit} \vee |m(\mu^+\mu^-K^-) - m_{B^+}^{PDG}| < 3\sigma_{fit}. \tag{4.2}$$

This additional selection allows to completely remove $B^+ \to J/\psi K^+$ events, as shown in Fig. 4.3 (right), while retaining 99.2% (98.5%) of $B^0 \to J/\psi K^{*0}$ ($B^0 \to K^{*0}\mu^+\mu^-$) signal.

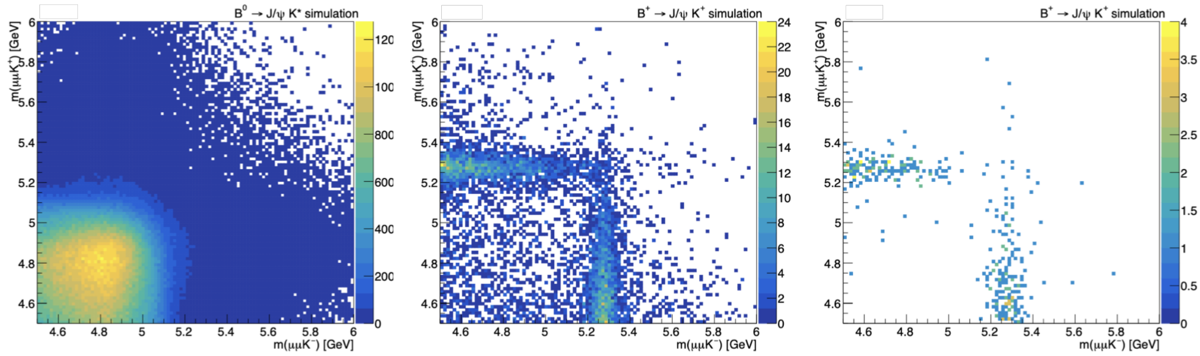


Figure 4.4: Distribution of $m(\mu^+\mu^-K^+)$ vs $m(\mu^+\mu^-K^-)$ for the $B^0 \to J/\psi K^{*0}$ (left) and $B^+ \to J/\psi K^+$ (center) MC samples. The right plot shows $m(\mu^+\mu^-K^+)$ vs $m(\mu^+\mu^-K^-)$ from the $B^+ \to K^+ J/\psi$ sample for events surviving the cut in Eq. 4.2. Taken from Ref. [13].

4.3.3 Feed-through from resonant channels

The resonant channels $B^0 \to J/\psi K^{*0}$ and $B^0 \to \psi(2S)K^{*0}$ corresponding to q^2 bins 4 and 6, respectively, can leak into the adjacent q^2 bins, contaminating the non-resonant channel $B^0 \to K^{*0}\mu^+\mu^-$. The discrimination between signal, $B^0 \to K^{*0}\mu^+\mu^-$, and control channels, $B^0 \to J/\psi K^{*0}$ and $B^0 \to \psi(2S)K^{*0}$, is based on the di-muon invariant mass. The signal sample is required to pass the selection

- $|m(\mu^+\mu^-) - m_{\psi}^{PDG}| > 3\sigma_{\mu\mu}$,

where m_{ψ}^{PDG} is the PDG mass of the J/ψ and $\psi(2S)$ mesons and $\sigma_{\mu\mu}$ is the uncertainty on the di-muon invariant mass, calculated per each candidate through the propagation of the track uncertainties.

As already done in the Run 1 analysis [29], a selection cut (called "B0 & Psi mass cut" henceforth) acting jointly on $m(K^{\pm}\pi^{\mp}\mu^+\mu^-)$ and $m(\mu^+\mu^-)$ is applied in order to reject the control channel events which spill in the neighbour q^2 bins. In particular, the cut is such that events are rejected if

- $|(m(K^{\pm}\pi^{\mp}\mu^+\mu^-) - m_{B^0}^{PDG}) - (m(\mu^+\mu^-) - m_{\psi}^{PDG})| < \Delta m.$

The value of Δm is tuned independently for the three di-muon mass regions: below the J/ψ (bin 3), between the $\psi(2S)$ and J/ψ (bin 5) and above the $\psi(2S)$ (bin 7). The optimisation criteria is the reduction of the feed-through from the resonant channels to a negligible level, below the 10% of the total background on data within the $m(K^{\pm}\pi^{\mp}\mu^+\mu^-)$ mass window $|m(K^{\pm}\pi^{\mp}\mu^+\mu^-) - m_{B^0}^{PDG}| < 0.28$ GeV for bin 3 and 5, while for bin 7 the resonant background level is required to be less than 1%. The background from the control channels (B_{peak}) and the overall background (B_{tot}) are evaluated as:

- B_{peak}: number of feed-through events from the control channel simulations;
- B_{tot}: difference between number of events in data and in the signal (non-resonant) simulation.

The optimisation procedure results in the following Δm cuts:

- bin 3: $\Delta m_{J/\psi} =$ 180 MeV;
- bin 5: $\Delta m_{J/\psi} =$ 80 MeV and $\Delta m_{\psi(2S)} =$ 90 MeV;
- bin 7: bin 7: $\Delta m_{\psi(2S)} =$ 80 MeV.

The scatter plots of the di-muon invariant mass squared vs the $m(K^{\pm}\pi^{\mp}\mu^{+}\mu^{-})$ for data events before (left) and after (right) the J/ψ and $\psi(2S)$ rejection cuts as well as the B0 & Psi mass cuts are applied are shown in Fig. 4.5, for year 2018. The vertical lines depict the q^2 bins defined in Tab. 1.1.

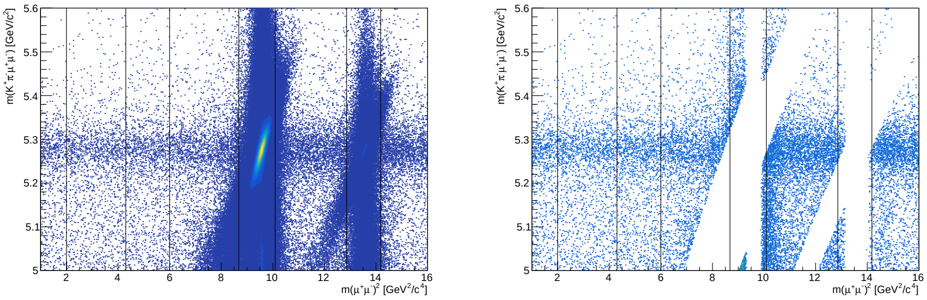


Figure 4.5: Scatter plots of the di-muon invariant mass squared vs the $m(K^{\pm}\pi^{\mp}\mu^{+}\mu^{-})$ for data events, before (left) and after (right) the J/psi and $\psi(2S)$ rejection and the B0 & Psi mass cuts are applied for year 2018.

4.3.4 Partially reconstructed backgrounds

Partially reconstructed backgrounds are formed by decays similar to our signal where an additional particle in the final state is missed in the reconstruction. The presence of these decays is studied in the $B^0 \to J/\psi K^{*0}$ channel. A clear example of this kind of contamination are events from the $B^+ \to \psi(2S)K^+$ decay, with $\psi(2S) \to J/\psi\pi^+\pi^-$, in which the softest pion is missed. These events populate the left $K^+\pi^-\mu^+\mu^-$ sideband, showing up as a wide shape in the mass spectrum and as a peak in the $\cos\theta_K$ distribution, as can be seen in Fig. 4.6.

An additional selection to remove this background was studied on simulated events of $B^+ \to \psi(2S)K^+$ and on both MC and data $B^0 \to J/\psi K^{*0}$ events from 2018 samples. The rejection is achieved using cuts on the following pairs of invariant masses: $(m_{\mu\mu\pi}, m_{\mu\mu K})$ and $(m_{B^+}^{PDG} - m_{\mu\mu K\pi}^{wt}, m_{K\pi}^{wt} - m_{K^{*0}}^{wt})$; where in $m^{wt}_{\mu\mu K\pi}$ and $m^{wt}_{K\pi}$, the pion and kaon mass assignments are swapped and $m^{PDG}_{B^+}$ and $m^{PDG}_{K^{*0}}$ are the nominal B^+ and K^{*0} masses from the PDG [38].

For each of the planes defined by the pair of variables, the partially reconstructed events are identified and removed. To reduce the impact of the selection criteria, only events lying in the affected region in both $(m_{\mu\mu\pi}, m_{\mu\mu K})$ and $(m^{PDG}_{B^+} - m^{wt}_{\mu\mu K\pi}, m^{wt}_{K\pi} - m^{wt}_{K^{*0}})$ planes are removed. Moreover, in the studies performed on MC, in the partially reconstructed $B^+ \to \psi(2S)K^+$ events, the kaon mass is always assigned to the leading track[3] during the flavour-tagging procedure. Events are rejected if:

- $(m^{wt}_{K\pi} - m^{PDG}_{K^{*0}} + 0.4) > -2.5 \times (m^{PDG}_{B^+} - m^{wt}_{K\pi\mu\mu} - 0.3) \wedge (m^{wt}_{K\pi} - m^{PDG}_{K^{*0}}) > 0$;
- $3.2 < m_{\mu\mu\pi} < 3.6$ GeV $\wedge$ $4.7 < m_{\mu\mu K} < 4.9$ GeV $\wedge$ $0.6 \times (m_{\mu\mu K} - 3.8) > (m_{\mu\mu\pi} - 3)$;
- $p^K_T > m^{\pi}_T$.

In Fig. 4.6 the mass spectrum (left) and the $\cos\theta_K$ angular distribution (right) of the events with mass in the range [5.0-5.15] GeV are shown, before and after the application of the rejection criteria for $B^+ \to \psi(2S)K^+$ events. The effect of the cut on the $B^0 \to J/\psi K^{*0}$ signal is negligible. The rejection cut is effective in removing the excess of events in the $\cos\theta_K$ distribution.

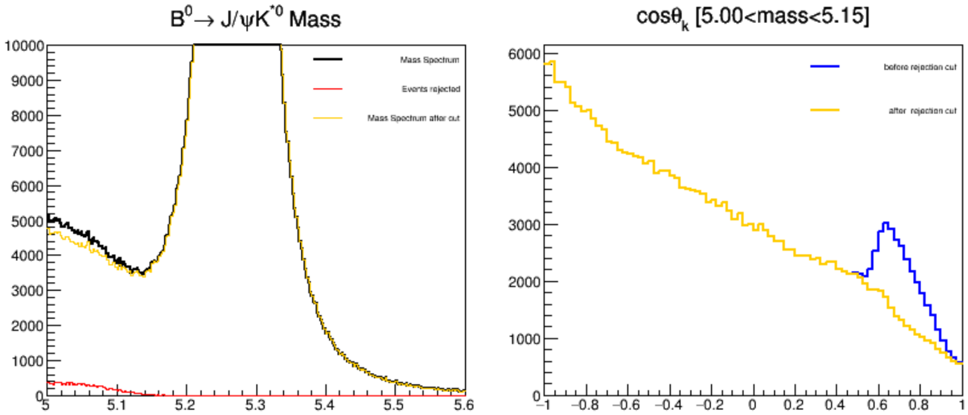


Figure 4.6: $B^0 \to J/\psi K^{*0}$ mass spectrum (left) with (yellow) and without (black) the application of the rejection cut for the $B^+ \to \psi(2S)K^+$ events (red). $\cos\theta_K$ angular distribution (right) for events with masses in the range [5.0-5.15] GeV, with (yellow) and without (blue) the rejection cut. Taken from Ref. [13].

---

[3] The leading track is the one with the highest momentum.

Chapter 5

Yields extraction

We present in Sec. 5.1 and 5.2 the methodology used to obtain the yields Y_S and Y_N of both signal $B^0 \to K^{*0}\mu^+\mu^-$ and normalisation $B^0 \to J/\psi K^{*0}$ channels which are necessary to compute the branching fraction in Eq. (1.1). We then present in Sec. 5.3 the results of the fit as well as its validation in Sec. 5.4. Finally, in Sec. 5.5 we study the impact of removing events coming from the problematic periods of 2016, as described in Sec. 4.1.1, in the precision of the final results.

5.1 Probability Density Function

Each event reconstructed as a B^0 candidate and passing the trigger and offline selections described in Chs. 3 and 4 has an associate $K^+\pi^-\mu^+\mu^-$ invariant mass m_i. The distribution of invariant masses for all considered events m follows a Probability Density Function (PDF) $\mathcal{P}(m;\vec{\lambda})$, where $\vec{\lambda}$ is a finite set of parameters and $\mathcal{P}(m;\vec{\lambda})\ dm$ gives the probability to find an event with $m_i \in [m + dm]$. In our analysis, the PDF model contains two signal components, corresponding to RT and WT events from the $B^0 \to K^{*0}\mu^+\mu^-$, $B^0 \to J/\psi K^{*0}$ and $B^0 \to \psi(2S)K^{*0}$ decays and a combinatorial background component. In the case of the J/ψ channel, there is an additional background component describing events from unremoved partially reconstructed decays, as will be explained in the following.

The total PDF is defined as

$$\mathcal{P}(m;\vec{\lambda}) = Y_{sig} \times [(1 - f^M)\mathcal{P}_{RT}(m;\vec{\lambda}_{RT}) + f^M \mathcal{P}_{WT}(m;\vec{\lambda}_{WT})] + Y_{bkg} \times \mathcal{P}_{bkg}(m;\vec{\lambda}_{bkg}), \tag{5.1}$$

where $\mathcal{P}_{RT}$, $\mathcal{P}_{WT}$ and $\mathcal{P}_{bkg}$ are the PDFs of RT signal, WT signal and combinatorial background, respectively, and $\vec{\lambda}_{RT}$, $\vec{\lambda}_{WT}$ and $\vec{\lambda}_{bkg}$ the corresponding parameters. For the J/ψ channel, there is an additional term $f_{erf} \times Y_{sig} \times \mathcal{P}_{erf}(m)$, where f_{erf} is the fraction of partially reconstructed events. f^M is the fraction of WT events, as described in Sec. 4.2.1. Y_{sig} and Y_{bkg} are the yields of signal and combinatorial background components, respectively. For the non-resonant q^2 bins $Y_{sig} = Y_S$ and for the $B^0 \to J/\psi K^{*0}$ q^2 bin $Y_{sig} = Y_N$, where Y_S and Y_N are the yields appearing in Eq. (1.1).

The RT signal PDF is parametrised as

$$\mathcal{P}_{RT}(m; \vec{\lambda}_{RT}) = \begin{cases} \text{DoubleCB}(m; \bar{m}, \sigma, \alpha_1, \alpha_2, n_1, n_2), & \text{if } q^2 \text{ bin} = 0, 1, 2, 3 \\ (1-f) \times \text{CB}(m; \bar{m}, \sigma_1, \alpha_1, n_1) + f \times \text{CB}(m; \bar{m}, \sigma_2, \alpha_2, n_2), & \text{if } q^2 \text{ bin} = 4, 5, 6 \\ (1-f) \times \text{CB}(m; \bar{m}, \sigma_1, \alpha_1, n_1) + f \times \text{Gauss}(m; \bar{m}, \sigma_2), & \text{if } q^2 \text{ bin} = 7 \end{cases}$$

where DoubleCB$(m; \bar{m}, \sigma, \alpha_1, \alpha_2, n_1, n_2)$ is a double crystal-ball function

$$\text{DoubleCB}(m; \bar{m}, \sigma, \alpha_1, \alpha_2, n_1, n_2) = \begin{cases} e^{-\frac{t^2}{2}}, & \text{if } -\alpha_1 < t < \alpha_2 \\ e^{-\frac{\alpha_1^2}{2}} \left[1 - \frac{\alpha_1}{n_1}(\alpha_1 + t)\right]^{-n_1}, & \text{if } t \leq -\alpha_1 \\ e^{-\frac{\alpha_2^2}{2}} \left[1 - \frac{\alpha_2}{n_2}(\alpha_2 - t)\right]^{-n_2}, & \text{if } -t \geq \alpha_2 \end{cases}$$

with $t = \frac{m - \bar{m}}{\sigma}$, CB$(m; \bar{m}, \sigma, \alpha, n)$ is a single crystal-ball function

$$\text{CB}(x; \bar{x}, \sigma, \alpha, n) = \begin{cases} e^{-\frac{t^2}{2}}, & \text{if } t > -\alpha \\ \left(\frac{n}{|\alpha|}\right)^n e^{-\frac{|\alpha|^2}{2}} \left[\frac{n}{|\alpha|} - |\alpha| - t\right]^{-n}, & \text{if } t \leq -\alpha \end{cases}$$

and Gauss$(m; \bar{m}, \sigma)$ is a Gaussian function

$$\text{Gauss}(m; \bar{m}, \sigma) = e^{-\frac{(m-\bar{m})^2}{2\sigma^2}}.$$

For all q^2 bins, the WT signal PDF is described by a double crystal-ball function $\mathcal{P}_{WT}(m; \vec{\lambda}_{WT}) =$ DoubleCB$(m; \bar{m}, \sigma, \alpha_1, \alpha_2, n_1, n_2)$ and the combinatorial background is described by an exponential function $\mathcal{P}(m; \vec{\lambda}_{bkg}) = e^{\lambda m}$. All PDFs are normalised to the mass range used in the fits $m \in$ [5.0,5.6] GeV.

The shape of $\mathcal{P}_{erf}$ is taken from MC. Possible contributions from unremoved backgrounds were studied using the 2018 inclusive MC containing $B \to J/\psi\ X$ decays, where X is a hadron. In Fig. 5.1 (left), the signal distribution (blue) as well as the combinatorial background (cyan), peaking backgrounds (green) and partially reconstructed backgrounds (red) distributions can be seen. The contribution from peaking background events is negligible and can be considered to be part of the combinatorial background. However, there is a significant presence of partially reconstructed events in the left sideband.

A fit to this distribution was performed using the sum of an exponential and a complementary error function and the result of the fit can be seen in Fig. 5.1 (right). When fitting data events from the J/ψ channel, $\mathcal{P}_{erf}$ is fixed to the shape obtained in MC. The normalisation f_{erf} is not, however, fixed to MC, where its value (≈ 31%) is much larger than the one found in data. The 2018 data fit to the J/ψ q^2 bin is firstly performed with f_{erf} free to float, where one obtains the result $f_{erf} = 0.0672 \pm 0.0018$. The fits for all years are then performed with the normalisation fixed to this value.

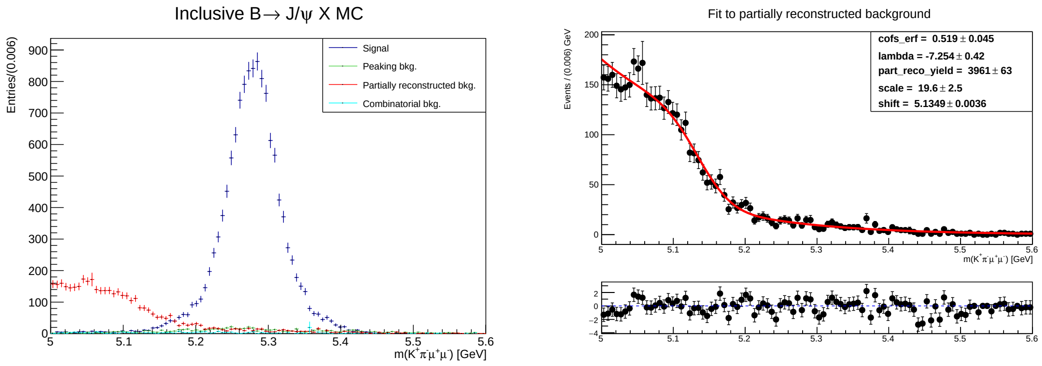


Figure 5.1: Different backgrounds and signal distributions in the 2018 $B \to J/\psi\ X$ inclusive MC (left) and result of the fit to the distribution of partially reconstructed events (right).

5.2 Maximum likelihood estimation

The maximum likelihood estimation [38] is a method used to determine the estimators $\hat{\lambda}_i$ for the unknown parameters λ_i of our model. The maximum likelihood theorem states that the best estimators are the ones that maximise the likelihood function, i.e. the ones that maximise the likelihood that the process described by our model produced the data that we are observing. The likelihood function $\mathcal{L}(\vec{\lambda})$ corresponds to the model PDF when regarded as a function of the parameters $\mathcal{P}(m;\vec{\lambda})$. When the sample consists of Y_{obs} statistically independent events, each following the same PDF, the joint likelihood factorises and can be written as

$$\mathcal{L}(\vec{\lambda}) = \prod_{i=1}^{Y_{obs}} \mathcal{P}(m_i;\vec{\lambda}), \tag{5.2}$$

in which Y_{obs} is regarded as fixed.

When the probability to observe Y_{obs} events depends on $\vec{\lambda}$ the likelihood function is extended

$$\mathcal{L}(\vec{\lambda}) = \prod_{i=1}^{Y_{obs}} \mathcal{P}(m_i;\vec{\lambda}) \times \frac{e^{-Y} Y^{Y_{obs}}}{Y_{obs}!}, \tag{5.3}$$

by the factor $\frac{e^{-Y} Y^{Y_{obs}}}{Y_{obs}!}$, which constrains the sum of signal and background yields Y to follow a Poisson distribution. From the parameters obtained with this procedure, Y_{sig} is our parameter of interest.

The maximisation of the likelihood can be performed analytically for the simplest cases but a numerical approach is necessary in most realistic problems. In many situations, it is convenient to use the the negative of the logarithm $-\ln\mathcal{L}$ instead of $\mathcal{L}$, because the product of the terms is transformed in the sum of logarithms, which is advantageous in the computation, e.g. when dealing with the product of large exponentials. In the analysis, the minimisation of $-\ln\mathcal{L}$ is performed with the `Minuit` [68] package.

The estimator $\hat{\lambda}_i$ for the parameter λ_i has an associate error, which is estimated by analysing the region around the minimum of $-\ln \mathcal{L}$. We can Taylor expand $-\ln \mathcal{L}$ around its minimum $\lambda = \hat{\lambda}$ by writing

$$-\ln \mathcal{L}(\lambda) = -\ln \mathcal{L}(\hat{\lambda}) - \frac{d \ln \mathcal{L}}{d\lambda}\Big|_{\lambda=\hat{\lambda}}(\lambda - \hat{\lambda}) - \frac{1}{2}\frac{d^2 \ln \mathcal{L}}{d\lambda^2}\Big|_{\lambda=\hat{\lambda}}(\lambda - \hat{\lambda})^2 + \mathcal{O}((\lambda - \hat{\lambda})^3). \tag{5.4}$$

Since $\hat{\lambda}$ is a minimum of $-\ln \mathcal{L}$, the first derivative term is equal to zero. Moreover, in the limit of a large number of events Y_{obs}, the estimator $\hat{\lambda}$ converges to λ and we can neglect higher order terms in the difference $\lambda - \hat{\lambda}$. We can then write

$$-\ln \mathcal{L}(\lambda) \approx -\ln \mathcal{L}(\hat{\lambda}) - \frac{1}{2}\frac{d^2 \ln \mathcal{L}}{d\lambda^2}\Big|_{\lambda=\hat{\lambda}}(\lambda - \hat{\lambda})^2, \tag{5.5}$$

or, equivalently,

$$\mathcal{L}(\lambda) \approx C \times e^{\frac{1}{2}\frac{d^2 \ln \mathcal{L}}{d\lambda^2}(\lambda - \hat{\lambda})^2}. \tag{5.6}$$

Therefore, in the limit of large Y_{obs} the likelihood function is Gaussian distributed with $(\sigma^{-1})^2 = -\frac{d^2 \ln \mathcal{L}}{d\lambda^2}\Big|_{\lambda=\hat{\lambda}}$, where σ is the error of the estimator (1 σ interval). It corresponds to the range around the minimum for which $-\ln \mathcal{L}$ increases by +0.5. The estimator is then quoted as $\hat{\lambda} \pm \sigma$. If Y_{obs} is not sufficiently large, $\mathcal{L}$ is not exactly Gaussian distributed and the estimator will have asymmetric uncertainties (σ_+ and σ_-).

When more than one parameter is simultaneously extracted from the fit, the σ_i errors of the estimators $\hat{\lambda}_i$ are defined as the square root of the diagonal elements of the covariance matrix, whose inverse is given by

$$(V^{-1})_{ij} = -\frac{\partial^2 \ln \mathcal{L}}{\partial\lambda_i \partial\lambda_j}\Big|_{\hat{\vec{\lambda}}}, \tag{5.7}$$

evaluated at the minmum of $-\ln \mathcal{L}$. V_{ij} is an invertible square symmetric matrix, whose entries correspond to the covariances between parameters λ_i and λ_j, $V_{ij} = cov(\lambda_i, \lambda_j)$, measuring how much the two parameters vary together: $cov(\lambda_i, \lambda_j) = \frac{1}{Y_{obs}-1}\sum_{n=1}^{Y_{obs}} E[(\lambda_i^n - \hat{\lambda}_i)(\lambda_j^n - \hat{\lambda}_j)]$; where E denotes the expectation value. The diagonal elements of V_{ij} correspond to the covariance of one parameter with each other, i.e. its variance, $cov(\lambda_i, \lambda_i) = var(\lambda_i) = \frac{1}{Y_{obs}-1}\sum_{n=1}^{Y_{obs}} E[(\lambda_i^n - \hat{\lambda}_i)^2]$. The statistical error of the estimator $\hat{\lambda}_i$ is then $\sigma_i = \sqrt{V_{ii}}$. In the case of one dimension, V^{-1} is a 1×1 matrix given by $V^{-1} = -\frac{d^2 \ln \mathcal{L}}{d\lambda^2}\Big|_{\lambda=\hat{\lambda}}$ and we recover the result of the last paragraph.

5.3 Nominal fit results

We perform a one-dimensional extended unbinned[1] maximum likelihood fit to the $K^+\pi^-\mu^+\mu^-$ invariant mass distribution. Since in data there is not enough information to distinguish RT from WT events, we adopted the following fitting strategy. We started by fitting the RT and WT events separately using dedicated MC simulations with the same selections as in data. From these fits we extracted the estimators

---

[1] Bins of the $K^+\pi^-\mu^+\mu^-$ invariant mass are not used in the likelihood fit since the low statistics q^2 bins do not allow for an adequate binning.

$\hat{\lambda}'_i \pm \sigma'_i$ with which we defined Gaussian constraints centred in $\hat{\lambda}'_i$ with width given by σ'_i. When fitting the data, Gaussian constraints on all parameters, except on the RT mean ($\bar{m}_{RT}$) and yields, are applied.

A re-parametrisation of the original PDF defined in Eq. (5.1) is made, moving from ($\bar{m}_{RT}$, $\bar{m}_{WT}$) to ($\bar{m}_{RT}$, $\bar{m}_{diff}$), where $\bar{m}_{diff} = \bar{m}_{WT} - \bar{m}_{RT}$ is the difference between the RT and WT means. A Gaussian constraint on this parameter is also applied, centred in the difference between the means obtained in MC and with width equal to the quadratic sum of their uncertainties. The RT and WT PDFs are multiplied by the Gaussian constraints on the RT and WT parameters, respectively.

Finally, the total PDF is multiplied by a Gaussian constraint on the fraction of WT events f^M. Its central value is given by the ratio $\frac{Y_{WT}}{Y_{WT}+Y_{RT}}$, where Y_{RT} and Y_{WT} are the RT and WT yields obtained from the MC fits. Its error is obtained by measuring the mis-tag fraction on sub-samples of MC events with data-like statistics and fitting the distribution with a Gaussian function. The resulting Gaussian width is used as the uncertainty on the f^M constraint when fitting the data. The obtained f^M values range from 12% to 14%, depending on the q^2 bin and the data-taking year.

The results of the fit to the RT and WT MC can be seen in Figs. 5.2 and 5.3, respectively, for q^2 bins 0, 4 and 7 in year 2018. In each plot, the result of the fit with the data points superimposed can be seen. The signal yield estimate is shown in the top left region of each plot, with the respective statistical uncertainty. On the bottom panel, the pull defined as $\text{Pull}_i = \frac{Y_i^{data} - Y_i^{fit}}{\sigma_i^{data}}$ for each $m(K^+\pi^-\mu^+\mu^-)$ bin i is shown, where Y_i^{data} is the yield found on data with the respective error σ_i^{data} and Y_i^{fit} is the yield given by the fit. The pull gives us a way to quantify the deviation between the fit result and the data points and should ideally be distributed around zero within a few standard deviations.

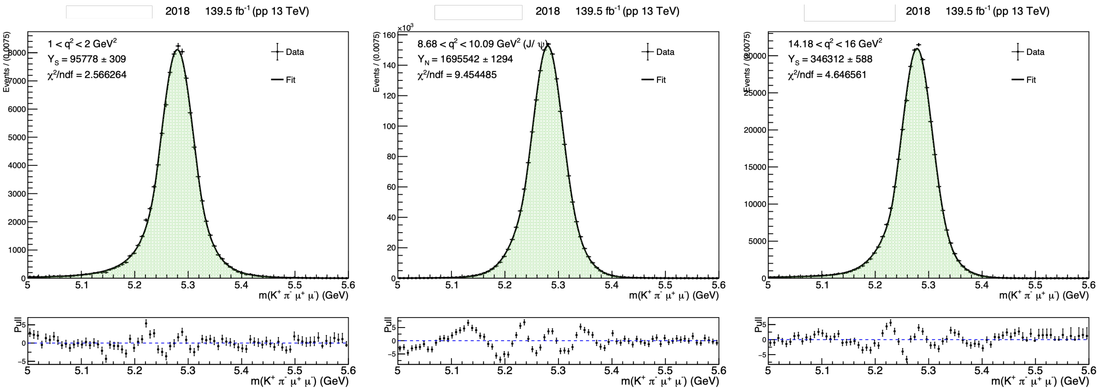


Figure 5.2: Results of the fit to the RT MC in year 2018, for q^2 bins 0 (left), 4 (centre) and 7 (right).

Another way to ascertain the quality of the fit is by means of the χ^2 test. The χ^2 is defined as $\chi^2 = \sum_{i=1}^{n_{bins}} \text{Pull}_i^2$, with the number of $m(K^+\pi^-\mu^+\mu^-)$ bins $n_{bins} = 80$ and the Pull_i defined in the previous paragraph. If the fit results for the yields Y_i^{fit} are normally distributed around the data points

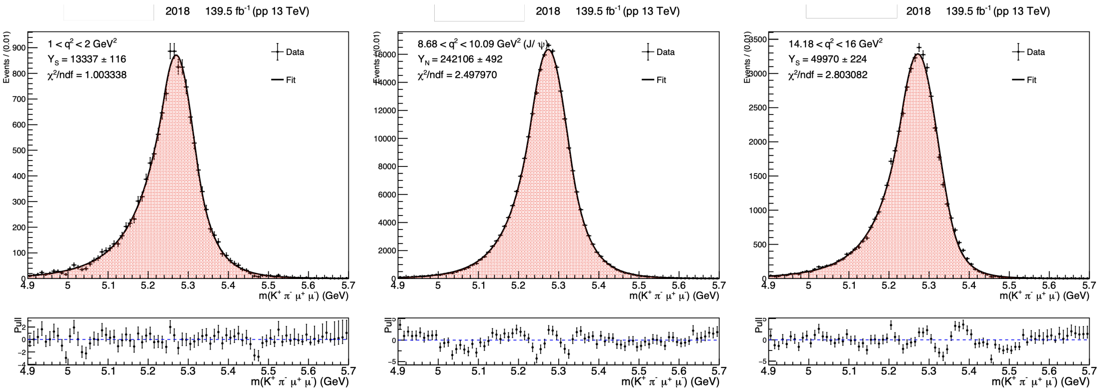


Figure 5.3: Results of the fit to the WT MC in year 2018, for q^2 bins 0 (left), 4 (centre) and 7 (right).

with central value equal to Y_i^{data} and width σ_i^{data}, then the χ^2 variable follows the χ^2 distribution

$$\mathcal{P}(\chi^2) = \frac{(\chi^2/2)^{\frac{ndf}{2}-1}\exp\left\{-\frac{\chi^2}{2}\right\}}{2\Gamma(ndf/2)}, \tag{5.8}$$

where $\Gamma(x)$ is the Gamma function. The expectation value of χ^2 is the number of degrees of freedom, that in this case, corresponds to the number of bins minus the number of fitted parameters. Therefore, for a good fit quality, we expect $\chi^2 \approx ndf$ or, equivalently, $\chi^2/ndf \approx 1$.

The data fit results in year 2018 are shown in Fig. 5.4. The data fit results for the remaining data-taking years can be found in Appendix A. The total fit (in black) can be seen superimposed with the data points. The RT and WT components are painted in green and red, respectively. The exponential curve of the combinatorial background is shown in blue and the component describing the partially reconstructed events in the J/ψ channel is shown in cyan. The q^2 ranges covered by each q^2 bin can be seen in the top left region of each plot.

The signal yields for all q^2 bins and data-taking years are collected in Tab. 5.1 with the corresponding statistical errors. The χ^2/ndf value of each fit is also shown. The lowest statistical error is 0.13% for the J/ψ channel in year 2018, whereas the largest statistical error is 11.1% for q^2 bin 0 in year 2016.

5.4 Fit validation

Besides checking the quality of the fit by means of pull plots and the χ^2 values, we also want to know whether our estimate for the signal yield is unbiased and whether we can trust its statistical error. In order to do that, a toy MC study is performed using the `RooMCStudy` class. 5000 toy samples are generated and fitted using the same model PDF used to fit the data. The fit parameters as well as the corresponding errors are saved for each toy MC.

We are particularly interested in validating the signal yield result, since it is the parameter we use to

Table 5.1: Signal yields with respective statistical uncertainties obtained from the data fits to all q^2 bins and data-taking years. The χ^2/ndf of each fit is shown in the last column.

q^2 bin	Year	Nominal yield	Stat. error (%)	χ^2/ndf
0	2016	118 ± 13	11.1	0.83
	2017	307 ± 21	6.9	0.68
	2018	501 ± 27	5.4	1.21
1	2016	232 ± 20	6.4	0.73
	2017	581 ± 30	5.1	0.78
	2018	982 ± 39	4.0	0.65
2	2016	181 ± 18	9.9	0.79
	2017	495 ± 28	5.6	0.87
	2018	822 ± 36	4.4	0.82
3	2016	398 ± 25	6.3	1.17
	2017	1013 ± 39	3.9	1.42
	2018	1609 ± 50	3.1	1.79
4	2016	292609 ± 693	0.24	6.52
	2017	734519 ± 1152	0.16	14.12
	2018	1532024 ± 1985	0.13	11.70
5	2016	634 ± 32	5.0	1.06
	2017	1363 ± 46	3.4	1.43
	2018	3022 ± 68	2.2	1.84
6	2016	20900 ± 181	0.9	2.05
	2017	48108 ± 321	0.7	1.38
	2018	97752 ± 435	0.4	3.93
7	2016	359 ± 23	6.4	0.84
	2017	836 ± 33	4.0	0.92
	2018	1837 ± 49	2.7	0.65

compute the branching fraction. We analyse the pull distribution

$$\text{Pull}(Y) = \frac{Y_i^{toy} - Y^{data}}{\sigma_i^{toy}}, \tag{5.9}$$

where Y_i^{toy} is the signal yield obtained from the fit to the toy MC sample with index i, σ_i^{toy} is the corresponding statistical error and Y^{data} is the signal yield obtained from the fit to data.

Due to the central limit theorem[2], the pull distribution should be normally distributed for a sufficiently large number of toy MC samples. Moreover, if the estimator for the signal yield Y^{data} is unbiased, we expect to obtain statistically compatible estimates in each repeated measurement. The distribution of pulls for an unbiased fit should follow a Gaussian distribution with unit width, centred upon zero. Shifts in the mean can suggest a bias in the fitting procedure, while deviations from a unit width can suggest that the statistical uncertainty returned by the fit is not reliable.

In Fig. 5.5, the signal yield pull distributions in year 2018 can be seen. The pull distributions for the remaining data-taking years can be seen in Appendix B. The values of the means and widths of the

---

[2] The central limit theorem states that if one has a population with mean μ and standard deviation σ and take a sufficiently large number of random samples from the population, then the distribution of the sample means will be approximately Gaussian.

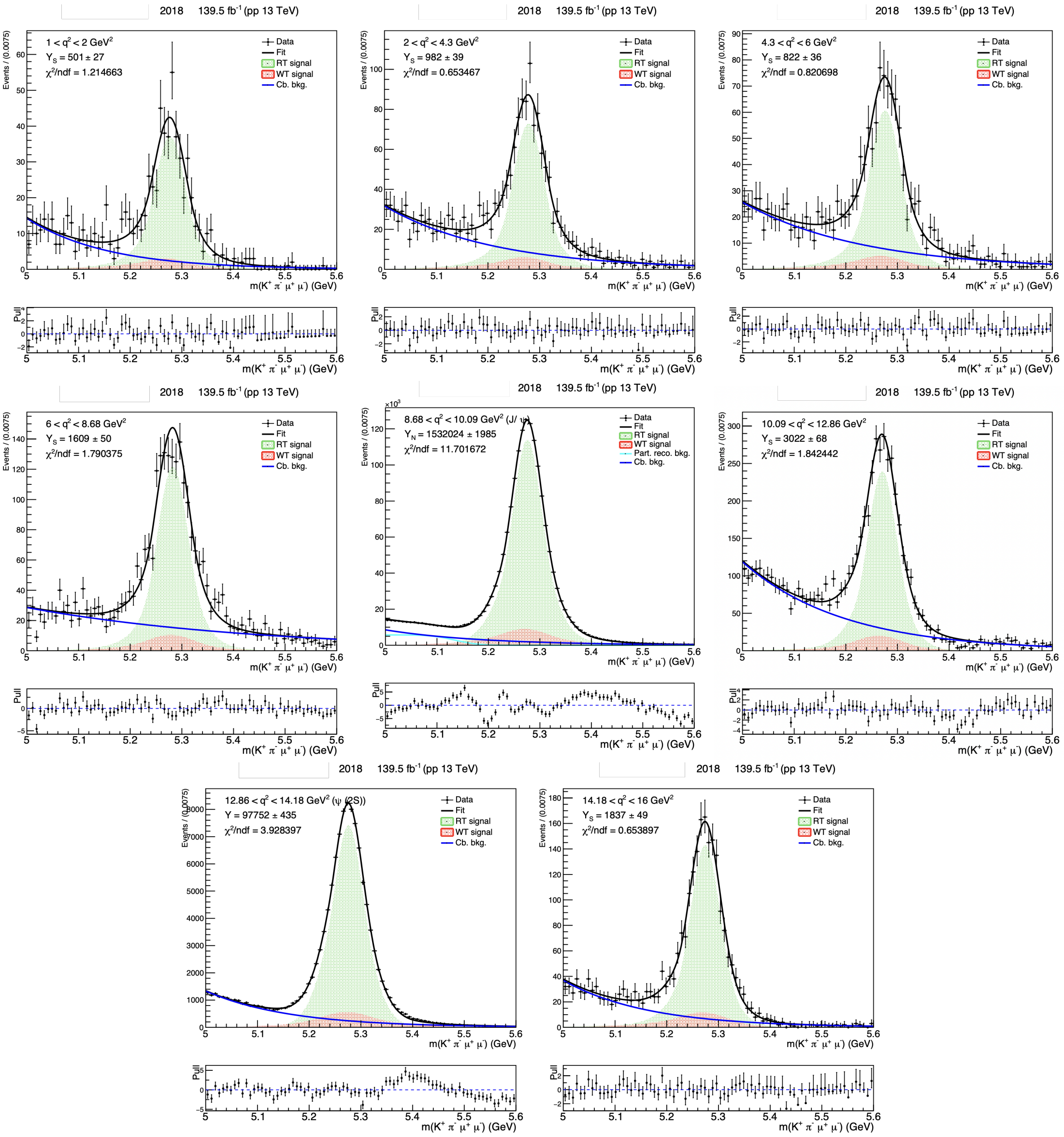


Figure 5.4: Data fit results in year 2018. The q^2 ranges covered by each q^2 bin can be seen in the top left region of each plot. The RT and WT components are shown in green and red, respectively. The combinatorial background is shown in blue and the partially reconstructed background in the J/ψ channel in cyan.

signal yield pull distributions for all q^2 bins and years are collected in Tab. 5.2. The fit bias is taken as a systematic error in the yields, as will be explained in Ch. 8.

5.5 Impact of removing the problematic periods of 2016

As explained in Sec. 4.1.1, in this analysis we do not consider the events from 2016 corresponding to the periods in which the EMTF was misconfigured (”EMFT L1” period) and the hit strip tracker was being saturated by HIP (”APV” period). In the ”APVL1” period both problems were present simultaneously. In

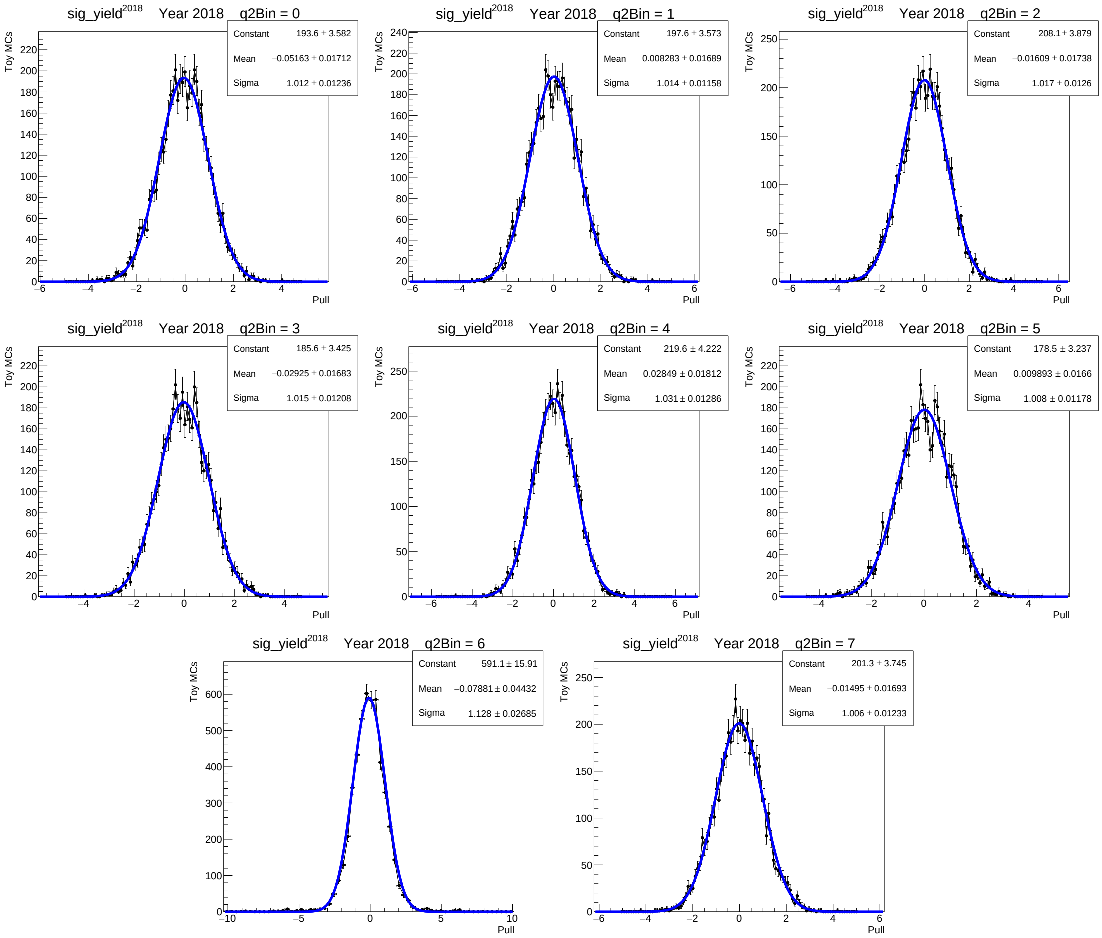


Figure 5.5: Signal yield pull distributions resulting from the toy MC study in year 2018. The mean and width values of the Gaussian distributions are shown in the top right panel.

this section, we study the impact of this removal in the precision of the final result and its effect in the yield ratio Y_S/Y_N, which appears in Eq. (1.1).

In Fig. 5.6 (left), the relative statistical errors in years 2016 (blue) before (dashed line) and after (solid line) the removal, 2017 (orange) and 2018 (green) can be seen. We observe a decrease of a factor of about 1.3-1.4, depending on the q^2 bin, in the precision of the 2016 results after the removal. In the same figure, the relative statistical error of the 3 years combined result is also shown in black, before (dashed line) and after (solid line) the removal. The combination of the three years is performed as explained in Sec. 9.2.2. Here we see a decrease of a factor of about 1.1-1.2, depending on the q^2 bin, and we conclude that this removal does not affect significantly the precision of the final result. Still in the same plot, the statistical errors of the previous CMS Run 1 [4] analysis are also shown in red. We see an increase of a factor of about 3 in the statistical precision of the branching fraction measurement with respect to the previous Run 1 analysis. The same comparison is performed in Ch. 9, taking also into account the systematic error of the results.

Table 5.2: Values of the means and widths of the Gaussian pull distributions for the signal yield.

q^2 bin	Year	Mean	Width
0	2016	-0.048 ± 0.017	0.997 ± 0.012
	2017	-0.036 ± 0.017	1.023 ± 0.012
	2018	-0.052 ± 0.017	1.012 ± 0.012
1	2016	-0.011 ± 0.017	1.000 ± 0.012
	2017	$+0.026 \pm 0.016$	1.001 ± 0.012
	2018	$+0.008 \pm 0.017$	1.014 ± 0.012
2	2016	$+0.002 \pm 0.017$	1.012 ± 0.012
	2017	-0.022 ± 0.017	1.021 ± 0.012
	2018	-0.0016 ± 0.017	1.017 ± 0.013
3	2016	-0.0006 ± 0.017	1.018 ± 0.013
	2017	-0.014 ± 0.017	1.025 ± 0.012
	2018	-0.0029 ± 0.017	1.015 ± 0.012
4	2016	-0.153 ± 0.017	1.018 ± 0.012
	2017	$+0.062 \pm 0.017$	0.999 ± 0.012
	2018	$+0.028 \pm 0.018$	1.031 ± 0.013
5	2016	$+0.001 \pm 0.017$	0.995 ± 0.012
	2017	-0.022 ± 0.016	0.995 ± 0.012
	2018	-0.010 ± 0.017	1.008 ± 0.012
6	2016	-0.111 ± 0.037	1.032 ± 0.021
	2017	-0.046 ± 0.035	1.072 ± 0.021
	2018	-0.079 ± 0.044	1.128 ± 0.027
7	2016	-0.020 ± 0.017	1.006 ± 0.012
	2017	-0.010 ± 0.016	1.001 ± 0.012
	2018	-0.015 ± 0.017	1.006 ± 0.012

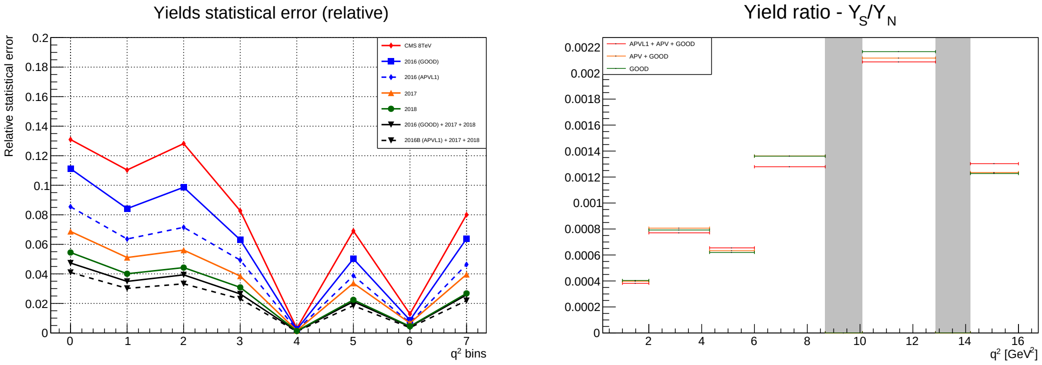


Figure 5.6: In the left, the relative statistical errors in years 2016 (blue) before (dashed line) and after (solid line) the removal, 2917 (orange) and 2018 (green). The 3 year combined results (black) before (dashed line) and after (solid line) the removal can also be seen, as well as the previous CMS Run 1 [4] results (red). In the right, the yield ration Y_S/Y_N obtained with the total 2016 sample (red), after removing the APVL1 period (orange) and after removing both periods (green).

In Fig. 5.6 (right), the yield ratio Y_S/Y_N obtained with the total 2016 sample (APVL1+APV+GOOD) in red, the 2016 sample after removing the APVL1 period (APV+GOOD) in orange and the 2016 sample after removing both periods (GOOD) in green can be seen. The removal of the APV period does not change notably the yield ratio, except in q^2 bin 5 where a larger difference is seen. The effect of removing the APVL1 period is more significant in the highest q^2 bins.

Chapter 6

Detector efficiency

In this chapter, we present the methodology used to compute the efficiencies ϵ_S and ϵ_N of both signal $B^0 \to K^{*0}\mu^+\mu^-$ and normalisation $B^0 \to J/\psi K^{*0}$ channels appearing in Eq. (1.1). We begin in Sec. 6.1 by presenting the efficiency definition. In Sec. 6.2, we present a simplified method to compute the efficiencies which relies solely on the MC simulations. Finally, in Sec. 6.3, we present a correction to this simple method which makes use of the $B^0 \to K^{*0}\mu^+\mu^-$ decay rate in Eq. (2.31). A comparison between the two methods is also provided.

6.1 Efficiency definition

The efficiency is defined as the ratio between the number of signal candidates that are accepted, reconstructed and selected over the total number of signal events produced in the collisions, within the fiducial region of the analysis. The accepted candidates are the ones passing the selections defined in Eq. (4.1). The efficiencies (ϵ_S and ϵ_N) are used to correct the measured signal yields (Y_S and Y_N) obtained in Ch. 5 in the branching fraction measurement, so that the experimental results can be compared with the theory predictions.

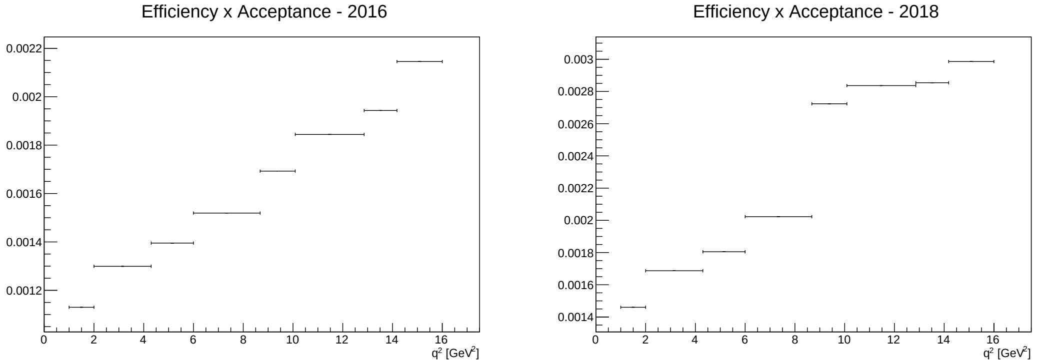
Since we want to take into account the effect of the pileup re-weighting mentioned in Sec. 4.1, the total efficiency ϵ is defined as the product between two terms: the selection efficiency (N_{sel}/D_{sel}) and the acceptance (N_{acc}/D_{acc}) as

$$\epsilon = \frac{N_{acc}}{D_{acc}} \times \frac{N_{sel}}{D_{sel}}. \tag{6.1}$$

The selection efficiency gives the fraction of signal candidates that are reconstructed and pass all the selections described in Ch. 4 over the total number of candidates that are accepted, whereas the detector acceptance gives the fraction of signal candidates that are accepted over the total number of signal events that are produced, within the fiducial region. The terms N_{sel} and D_{sel} have the pileup weight, whereas the terms N_{acc} and D_{acc} do not.

6.2 MC based efficiency

The efficiency defined in Eq. (6.1) can be obtained simply by taking the ratio between the number of events in the numerators and denominators. Separate histograms binned in q^2 are firstly filled with simulated events passing the respective selections. The ratio between the histograms corresponding to $N_{acc} \times N_{sel}$ and $D_{acc} \times D_{sel}$ is posteriorly performed using the `TEfficiency` class, bin by bin. Considering that k events pass the mentioned selections, out of a sample containing N events, the efficiency estimate is $\hat{\epsilon} = \frac{k}{N}$. The results of this procedure are shown in Fig. 6.1 for each q^2 bin in year 2016 (left) and 2018 (right). The result for 2017 can be seen Appendix C.


Efficiency x Acceptance - 2016

Efficiency x Acceptance - 2018

Figure 6.1: Detector efficiency for each q^2 bin, using solely the MC simulations, in year 2016 (left) and 2018 (right).

The process of having k successful events out of a total of N follows the binomial distribution

$$\mathcal{B}(k;\epsilon,N) = \frac{N!}{k!(N-k)!}\epsilon^k(1-\epsilon)^{N-k}, \tag{6.2}$$

in which the probability of having k successes (ϵ) is regarded as a parameter and the number of successes (k) as the variable. If instead we regard the number of successes (k) as a fixed parameter and the probability of success (ϵ) as the variable, then ϵ follows the β-distribution

$$\beta(\epsilon;\alpha,\beta) = \frac{1}{B(\alpha,\beta)}\epsilon^{\alpha-1}(1-\epsilon)^{\beta-1}, \tag{6.3}$$

where the β-function $B(\alpha,\beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt$ ensures that the β-distribution is normalised to unity. As in the case of the binomial distribution, we can interpret $\alpha - 1$ as the number of events passing the selections and $\beta - 1$ as the number of events that are rejected.

Using the latter interpretation, the statistical error of the efficiency can be obtained using the Clopper-Pearson method [69], representing the finite size of the MC. This method is regarded as exact, since it makes use of the β-distribution rather than an approximation, and its use is recommended by the PDG [38]. In the Clopper-Pearson method, the coverage probability of the interval, $[\sigma_\epsilon^-, \sigma_\epsilon^+]$, is fixed in advance as $100 \times (1-\alpha)\%$, where $1-\alpha = 0.68$ for a 1σ interval. Then the lower limit (σ_ϵ^-) of the interval is obtained by solving the equation

$$\mathcal{C}(x) = P(x \geq k; N) = \frac{\alpha}{2}, \tag{6.4}$$

where $P(x \geq k; N)$ gives the probability that the number of events (x) is greater or equal to the number of events passing the selections (k). $\mathcal{C}(x)$ is the cumulative distribution function of the β-distribution defined in Eq. (6.3). We can write

$$\begin{aligned}
P(x \geq k; N) =& 1 - P(x \leq k-1; N) \\
=& 1 - \int_0^{1-\epsilon} \mathcal{B}(t, \alpha = k, \beta = N-k+1) dt \\
=& 1 - \frac{1}{C}\int_0^{1-\epsilon} t^{N-k}(1-t)^{k-1} dt \\
=& 1 - \frac{1}{C}\int_\epsilon^{1} t^{k-1}(1-t)^{N-k} dt \\
=& \frac{1}{C}\int_0^{\epsilon} t^{k-1}(1-t)^{N-k} dt \\
=& I_\epsilon(\alpha = k, \beta = N-k+1),
\end{aligned} \tag{6.5}$$

where $I_\epsilon(\alpha,\beta) = \frac{B(\epsilon;\alpha,\beta)}{B(\alpha,\beta)}$ is the regularised incomplete β-function, C is a normalisation factor and $1-t = \frac{x}{N}$ is the probability of having x events passing the selections. The lower bound (σ_ϵ^-) of the Clopper-Pearson interval corresponds, therefore, to the $\frac{\alpha}{2}$ quantile[1] of the β-distribution.

Similarly, the upper limit (σ_ϵ^+) of the Clopper-Pearson interval is obtained by solving the equation

$$\mathcal{C}(x) = P(x \leq k; N) = 1 - \frac{\alpha}{2}, \tag{6.6}$$

with $P(x \leq k; N)$ giving the probability that the number of events (x) is less or equal to the number of events passing the selections (k). Analogously, it corresponds to the $1 - \frac{\alpha}{2}$ quantile of the β-distribution.

The efficiency statistical uncertainties obtained with the Clopper-Pearson method are collected in Tab. 6.1, for each q^2 bin and data-taking year.

6.3 Angular correction

The MC simulations are generated using values for the angular parameters appearing in the $B^0 \to K^{*0}\mu^+\mu^-$ decay rate in Eq. (2.31) that are close to the SM predictions. Since probing the flavour anoma-

---

[1] The quantile determines how many values of a distribution are above or below a certain limit.

Table 6.1: Relative statistical errors of the efficiency computed with the Clopper-Pearson method for each data-taking year and q^2 bin. The upper (+) and lower (-) limits are shown in %.

q^2 bin	0	1	2	3	4	5	6	7
2016 (+)	7×10^{-4} %	3×10^{-4} %	4×10^{-4} %	3×10^{-4} %	-2×10^{-3} %	7×10^{-5} %	3×10^{-4} %	2×10^{-4} %
2016 (-)	7×10^{-4} %	2×10^{-4} %	4×10^{-4} %	3×10^{-4} %	-8×10^{-3} %	2×10^{-4} %	2×10^{-4} %	1×10^{-4} %
2017 (+)	6×10^{-4} %	3×10^{-4} %	4×10^{-4} %	2×10^{-5} %	-4×10^{-3} %	-5×10^{-3} %	-1×10^{-4} %	2×10^{-4} %
2017 (-)	6×10^{-4} %	3×10^{-4} %	4×10^{-4} %	2×10^{-4} %	2×10^{-4} %	1×10^{-4} %	2×10^{-4} %	1×10^{-4} %
2018 (+)	6×10^{-4} %	2×10^{-4} %	3×10^{-4} %	2×10^{-10} %	-6×10^{-3} %	1×10^{-4} %	-9×10^{-3} %	2×10^{-4} %
2018 (-)	6×10^{-4} %	9×10^{-5} %	3×10^{-4} %	-7×10^{-3} %	-1×10^{-4} %	6×10^{-3} %	1×10^{-4} %	2×10^{-4} %

lies is one of the goals of our study, the dependence of the efficiencies on the angular variables, $\cos\theta_K$, $\cos\theta_l$ and ϕ defined in Sec. 2.3 needs to be considered: $\epsilon(\cos\theta_K, \cos\theta_l, \phi)$. To enter in the branching fraction calculation, these angular efficiency functions need to be integrated over the angular variables, weighted by the decay rate.

The angular parameters are extracted [13] from an extended unbinned maximum likelihood fit to the four kinematic variables describing the decay: the $K^+\pi^-\mu^+\mu^-$ invariant mass, m, and the three angular variables. The PDF model used in this fit is defined as

$$\begin{aligned}\mathcal{P}(m, \theta_K, \theta_l, \phi) = & Y_{sig}\Big[(1 - f^M)\mathcal{P}_{RT}(m)\mathcal{D}(\cos\theta_K, \cos\theta_l, \phi)\epsilon^{RT}(\cos\theta_K, \cos\theta_l, \phi) \\ & + f^M\mathcal{P}_{WT}(m)\mathcal{D}(-\cos\theta_K, -\cos\theta_l, \phi)\epsilon^{WT}(\cos\theta_K, \cos\theta_l, \phi)\Big] \\ & + Y_{bkg}\mathcal{P}_{bkg}(m)\mathcal{P}_{bkg}(\cos\theta_K, \cos\theta_l, \phi).\end{aligned} \tag{6.7}$$

The first, second and third terms describe, respectively, the RT signal, the WT signal and the background events. $\mathcal{P}_{RT}(m)$, $\mathcal{P}_{WT}(m)$ and $\mathcal{P}_{bkg}(m)$ are the corresponding mass PDFs. $\mathcal{D}(\cos\theta_K, \cos\theta_l, \phi)$[2] is the angular decay rate given by Eq. (2.31) and includes the pure-physics information of the decay. This theoretical description needs to be adapted to the experimental conditions with the angular efficiency functions $\epsilon^{RT}(\cos\theta_K, \cos\theta_l, \phi)$ and $\epsilon^{WT}(\cos\theta_K, \cos\theta_l, \phi)$ for RT and WT events, respectively. Finally, $\mathcal{P}_{bkg}(\cos\theta_K,\ cos\theta_l, \phi)$ describes the angular distribution of the background events and is given by a multivariate Bernstein polynomial. The extraction of the angular parameters was not performed in the context of the work of this thesis. However, the angular parameters obtained from this fit are used to compute the averaged efficiencies presented in this section.

In the angular analysis, a blinding procedure is used in order to avoid any bias introduced by human decisions in the setup of the analysis. During the blinded period, which includes the steps of event selection, fit strategy and validation and efficiency measurement, only the MC samples and data from the mass sidebands have been used, while the distribution of events in the data signal region has not been considered. Since at the time this thesis was written, the angular analysis has not been unblinded yet, we used the signal angular parameters from LHCb [3].

---

[2] Since under CP: $\theta_K \to \theta_K - \pi$, $\theta_l \to \theta_l - \pi$ and $\phi \to \phi$; the angular decay rate as a function of the reconstructed variables to be used for the WT events is $\mathcal{D}(-\cos\theta_K, -\cos\theta_l, \phi)$. The efficiency functions need not be corrected since they are computed using the reconstructed variables.

The averaged efficiencies $\epsilon(\vec{p})$ are obtained using the values for the angular parameters $\vec{p} = (F_L, P_1, P_2, P_3, P_4', P_5', P_6', P_8')$ that are obtained with the four-dimensional maximum likelihood fit to data and MC, which can be found in Appendix C.

We denote, generally, a numerator and a denominator by $N(\vec{p})^{data}$ and $D(\vec{p})^{data}$, respectively, such that the efficiency in Eq. (6.1) is given by $\epsilon(\vec{p}) = \frac{N(\vec{p})^{data}}{D(\vec{p})^{data}}$. We can obtain the numerator doing the following integration

$$N(\vec{p})^{data} = \int \epsilon(\vec{\theta}) D(\vec{p}, \vec{\theta})^{data} d\vec{\theta}, \tag{6.8}$$

where $\vec{\theta} = (\cos\theta_K, \cos\theta_l, \phi)$ and $\epsilon(\vec{\theta})$ is the angular efficiency function. Since $\epsilon(\vec{\theta})$ should be the same in data and MC, we can write

$$N(\vec{p})^{data} = \int \frac{N(\vec{p}', \vec{\theta})^{MC}}{D(\vec{p}', \vec{\theta})^{MC}} \times D(\vec{p}, \vec{\theta})^{data} d\vec{\theta}, \tag{6.9}$$

where $\vec{p}'$ are the MC angular parameters. We can, therefore, obtain $N(\vec{p})^{data}$ by summing over the MC numerator events, re-weighted with the ratio $\frac{D(\vec{p},\vec{\theta})^{data}}{D(\vec{p}',\vec{\theta})^{MC}}$. $D(\vec{p}, \vec{\theta})^{data}$ and $D(\vec{p}', \vec{\theta})^{MC}$ are obtained substituting in the $B^0 \to K^{*0}\mu^+\mu^-$ decay rate the angular parameters obtained from the fit to data and MC, respectively, and the values of the angular variables associated to each event in the sum. Similarly, we can obtain $D(\vec{p})^{data}$ by summing over the MC denominator events, re-weighted by the ratio $\frac{N(\vec{p},\vec{\theta})^{data}}{N(\vec{p}',\vec{\theta})^{MC}}$.

Practically, the averaged efficiencies are computed by, firstly, filling four weighted histograms and then using the `TEfficiency` class to divide the $N_{acc} \times N_{sel}$ and $D_{acc} \times D_{sel}$ histograms, bin by bin. The results of this procedure can be seen in Fig. 6.2 for each q^2 bin in year 2016 (left) and 2018 (right). The efficiency result for year 2017 can be found in Appendix C.

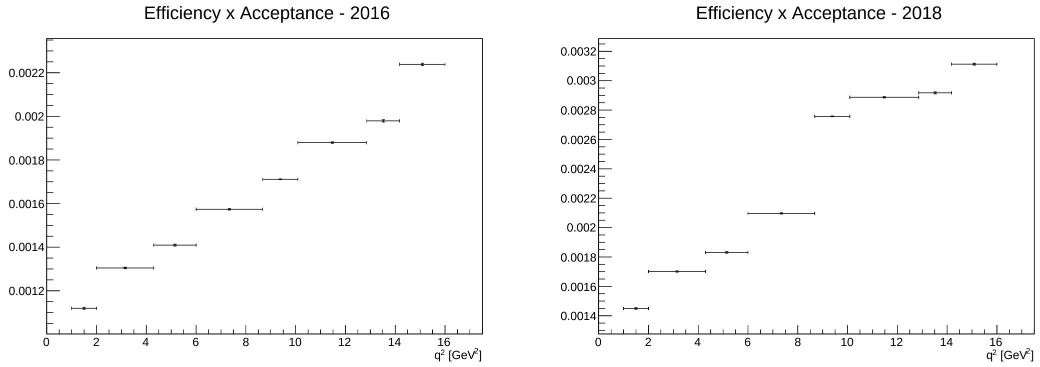


Figure 6.2: Averaged efficiencies for each q^2 bin in year 2016 (left) and 2018 (right).

Since in this case the histograms are filled with weights, the efficiency statistical errors cannot be estimated using the exact Clopper-Pearson method. They are obtained assuming that the efficiency follows a Gaussian distribution with standard deviation

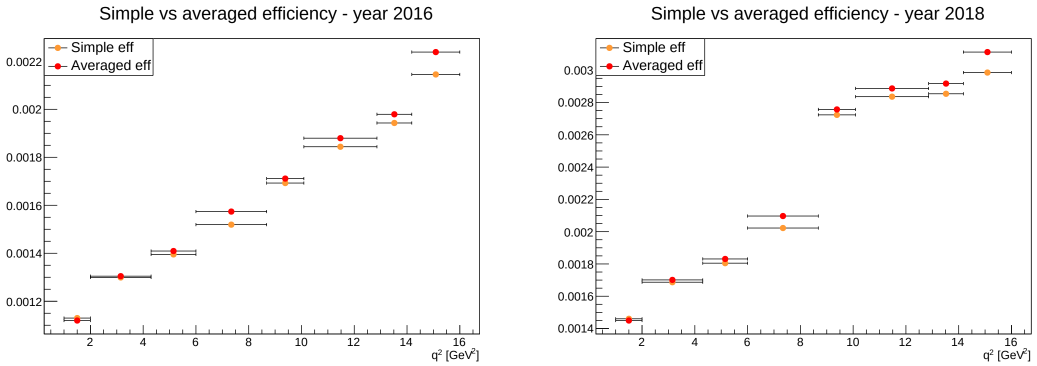
$$\sigma_\epsilon = \sqrt{\frac{\hat{\epsilon}(1-\hat{\epsilon})}{N}}, \tag{6.10}$$

where $\hat{\epsilon}$ is the estimator for the efficiency. This is only a good approximation in the limit of large N, as a consequence of the central limit theorem. The efficiency relative statistical errors in % for each q^2 bin and data-taking years are collected in Tab. 6.2. These errors are taken as a systematic uncertainty in the final branching fraction result, reflecting the finite size of the MC sample (see Ch. 8).

Table 6.2: Relative statistical errors in % of the averaged efficiency, computed using the normal approximation, for each data-taking year and q^2 bin.

q^2 bin	0	1	2	3	4	5	6	7
2016	0.42 %	0.27 %	0.31 %	0.22 %	0.095 %	0.19 %	0.27 %	0.23 %
2017	0.38 %	0.25 %	0.29 %	0.21 %	0.092 %	0.17 %	0.24 %	0.22 %
2018	0.39 %	0.25 %	0.29 %	0.20 %	0.090 %	0.16 %	0.23 %	0.20 %

In Fig. 6.3, a comparison between the simple efficiency calculation described in Sec. 6.2 and the averaged efficiencies obtained in this section is presented, for years 2016 (left) and 2018 (right). The comparison in year 2017 can be found in Appendix C. The largest differences are found at high q^2.


Simple vs averaged efficiency - year 2016

Simple vs averaged efficiency - year 2018

Figure 6.3: Comparison between the simple efficiency calculation described in Sec. 6.2 and the averaged efficiencies, for years 2016 (left) and 2018 (right).

Chapter 7

Monte Carlo validation

As explained in the previous chapter, the detector efficiency computation relies on the MC simulations and, as a consequence, we need to validate the MC by comparing it with data. Since in MC there is only the signal component, before this comparison can be made the signal and background events in data need to be separated. This is achieved by making use of the ${}_s\mathcal{P}lot$ method [70] described in Sec. 7.1. The comparison between data and MC is performed in Sec. 7.2. The ratios between the data and MC signal distributions obtained in this way are used to estimate a systematic uncertainty, which quantifies the data vs MC disagreements, as will be explained in Ch. 8.

7.1 The ${}_s\mathcal{P}lot$ method

When our data sample is composed of more than two species, in our case, signal (RT + WT) and background events, we can use the ${}_s\mathcal{P}lot$ method to obtain the distribution of a set of control variables for events belonging to each species. A control variable x can be, for instance, the transverse momentum or the pseudorapidity of the B^0 meson. It needs to be uncorrelated with the discriminating variable, i.e. the variable appearing in the likelihood expression in Eq. (5.3): the $K^+\pi^-\mu^+\mu^-$ invariant mass m. The goal is then to obtain the distribution of x for the signal events, $M_s(x)$.

The first step is to perform an extended maximum likelihood fit to the $K^+\pi^-\mu^+\mu^-$ invariant mass m and extract the signal and background yields, Y_{sig} and Y_{bkg}, as explained in Ch. 5. The sWeights are then defined as

$$
{}_s\mathcal{P}_n(m_i) = \frac{\sum_{j=s,b} V_{nj} f_j(m_i)}{\sum_{k=s,b} Y_k f_k(m_i)}, \tag{7.1}
$$

where the sums run over the different species (signal s and background b), $f_n(m_i)$ represents the corresponding mass PDFs and V_{nj} is the covariance matrix, whose inverse is given by

$$V_{nj}^{-1} = -\frac{\partial^2 \ln \mathcal{L}}{\partial Y_n \partial Y_j} = \sum_{i=1}^{Y} \frac{f_n(m_i) f_j(m_i)}{\left(\sum_{k=s,b} Y_k f_k(m_i) \right)^2}. \tag{7.2}$$

When other parameters are fitted simultaneously with the yields, as is the case in our analysis, a second fit needs to be performed in which these parameters are fixed, so that the covariance matrix defined in Eq. (7.2) is correctly determined.

The second step is to fill histograms of the x variable weighted by the sWeights in Eq. (7.1)

$$Y_n \, {}_s\tilde{M}_n(\bar{x})\delta x = \sum_{i \in \delta x} {}_s\mathcal{P}_n(m_i), \tag{7.3}$$

where the sum runs over the events in the x-bin centred in $\bar{x}$ with width δx. The sum of their entries is equal to the yields Y_n given by the fit and the quadratic sum of the error bars per bin, defined as

$$\sigma[Y_n \, {}_s\tilde{M}_n(\bar{x})\delta x] = \sqrt{\sum_{i \in \delta x} ({}_s\mathcal{P}_n)^2}, \tag{7.4}$$

is equal to the yields statistical uncertainties, $\sqrt{V_{ii}}$.

If we go to the continuous limit, by averaging the left- and right-hand sides of Eq. (7.3), the average of the sum is transformed in the integral $\int dm \int dx \sum_{j=s,b} Y_j f_j(m,x) \delta(x - \bar{x}) \delta x$, which corresponds to total number of events in an infinitesimal region of phase space. Because the control variable x is independent of the discriminating variable m, $f_j(m,x) = f_j(m) M_j(x)$. Then, using the sWeight definition in Eq. (7.1), Eq. (7.3) becomes

$$< Y_n \, {}_s\tilde{M}_n(\bar{x}) >= \sum_{j=s,b} Y_j M_j(\bar{x}) \int dm f_j(m) \frac{\sum_{l=s,b} V_{nl} f_l(m)}{\sum_{k=s,b} Y_k f_k(m)}, \tag{7.5}$$

where the integral of the Dirac δ-function was performed. In the continuous limit, the inverse of the covariance matrix takes the form (see Ref. [70])

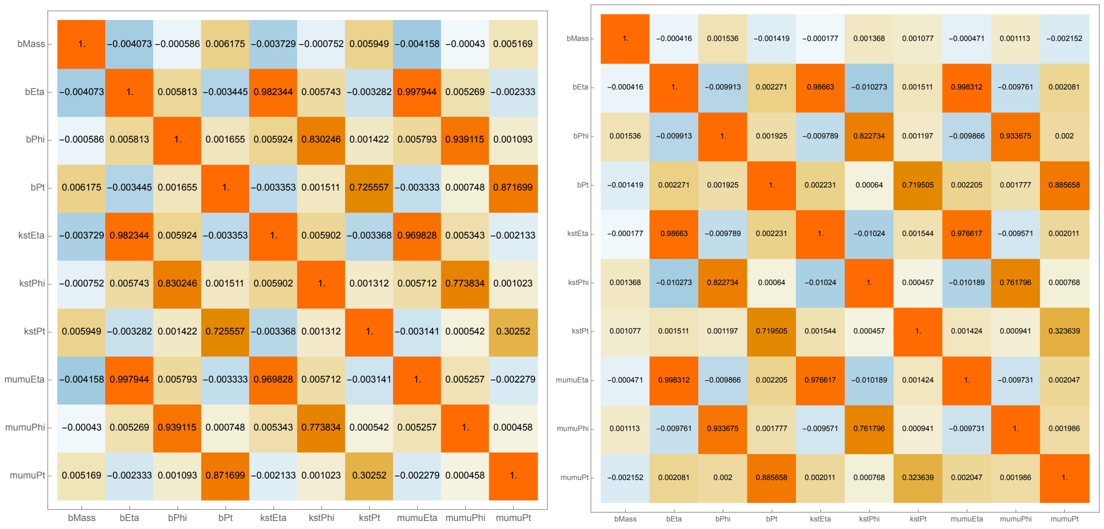
$$V_{nj}^{-1} = \int dm \frac{f_n(m) f_j(m)}{\sum_{k=s,b} Y_k f_k(m)}. \tag{7.6}$$

Using this expression, Eq. (7.5) can be written as

$$
\begin{aligned}
< Y_n \, {}_s\tilde{M}_n(\bar{x}) > &= \sum_{j=s,b} Y_j M_j(\bar{x}) \sum_{l=s,b} V_{nl} V_{jl}^{-1} \\
&= \sum_{j=s,b} Y_j M_j(\bar{x}) \delta_{nj} \\
&= Y_n M_n(\bar{x}).
\end{aligned}
\tag{7.7}
$$

Therefore, on average, the ${}_s\mathcal{P}lot$ method reproduces the true x distributions, for signal and background events, within the statistical uncertainties defined in Eq. (7.4), using the covariance matrix and the yields obtained from the likelihood fit. The resulting $M_n(\bar{x})$ distributions are only valid if x is uncorrelated with the discriminating variable m otherwise the step $f_j(m,x) = f_j(m)M_j(x)$ could not be used. A way to quantify the correlation among the control variables and m is by means of the correlation factor $r = cov(x,y)/\sigma_x\sigma_y$, where the covariance among two variables x and y was defined in Ch. 5 and $\sigma_x = \sqrt{1/Y_{obs} \sum_{i=1}^{Y_{obs}} (x_i - \bar{x})^2}$ is the standard deviation of x, with the sum running over the total number of observed events. If $r = 1$, x and y are totally correlated and if $x = 0$, x and y are totally uncorrelated.

In Fig. 7.1, the matrix plots show the correlation factor for a set of control variables for q^2 bin 4 in year 2016 (left) and 2018 (right). The selected control variables are: ”bEta”, ”bPhi”, ”bPt”, ”kstEta”, ”kstPhi”, ”kstPt”, ”mumuEta”, ”mumuPhi” and ”mumuPt”; corresponding, respectively, to the pseudorapidity, angle ϕ and transverse momentum of the B^0 meson, the K^{*0} meson and the di-muon system. Darker shades of blue correspond to r values closer to 0 whereas darker shades of orange correspond to r values closer to 1. The matrix plot is symmetric since $cov(x,y) = cov(y.x)$. In the first row/column, the correlation factor between the selected control variables and m (”bMass”) is shown and we can conclude that they are uncorrelated. The correlation plot for year 2017 is similar to the ones found for years 2016 and 2018.



	bMass	bEta	bPhi	bPt	kstEta	kstPhi	kstPt	mumuEta	mumuPhi	mumuPt
bMass	1.	−0.004073	−0.000586	0.006175	−0.003729	−0.000752	0.005949	−0.004158	−0.00043	0.005169
bEta	−0.004073	1.	0.005813	−0.003445	0.982344	0.005743	−0.003282	0.997944	0.005269	−0.002333
bPhi	−0.000586	0.005813	1.	0.001655	0.005924	0.830246	0.001422	0.005793	0.939115	0.001093
bPt	0.006175	−0.003445	0.001655	1.	−0.003353	0.001511	0.725557	−0.003333	0.000748	0.871699
kstEta	−0.003729	0.982344	0.005924	−0.003353	1.	0.005902	−0.003368	0.969828	0.005343	−0.002133
kstPhi	−0.000752	0.005743	0.830246	0.001511	0.005902	1.	0.001312	0.005712	0.773834	0.001023
kstPt	0.005949	−0.003282	0.001422	0.725557	−0.003368	0.001312	1.	−0.003141	0.000542	0.30252
mumuEta	−0.004158	0.997944	0.005793	−0.003333	0.969828	0.005712	−0.003141	1.	0.005257	−0.002279
mumuPhi	−0.00043	0.005269	0.939115	0.000748	0.005343	0.773834	0.000542	0.005257	1.	0.000458
mumuPt	0.005169	−0.002333	0.001093	0.871699	−0.002133	0.001023	0.30252	−0.002279	0.000458	1.

	bMass	bEta	bPhi	bPt	kstEta	kstPhi	kstPt	mumuEta	mumuPhi	mumuPt
bMass	1.	−0.000416	0.001536	−0.001419	−0.000177	0.001368	0.001077	−0.000471	0.001113	−0.002152
bEta	−0.000416	1.	−0.009913	0.002271	0.98663	−0.010273	0.001511	0.998312	−0.009761	0.002081
bPhi	0.001536	−0.009913	1.	0.001925	−0.009789	0.822734	0.001197	−0.009866	0.933675	0.002
bPt	−0.001419	0.002271	0.001925	1.	0.002231	0.00064	0.719505	0.002205	0.001777	0.885658
kstEta	−0.000177	0.98663	−0.009789	0.002231	1.	−0.01024	0.001544	0.976617	−0.009571	0.002011
kstPhi	0.001368	−0.010273	0.822734	0.00064	−0.01024	1.	0.000457	−0.010189	0.761796	0.000768
kstPt	0.001077	0.001511	0.001197	0.719505	0.001544	0.000457	1.	0.001424	0.000941	0.323639
mumuEta	−0.000471	0.998312	−0.009866	0.002205	0.976617	−0.010189	0.001424	1.	−0.009731	0.002047
mumuPhi	0.001113	−0.009761	0.933675	0.001777	−0.009571	0.761796	0.000941	−0.009731	1.	0.001986
mumuPt	−0.002152	0.002081	0.002	0.885658	0.002011	0.000768	0.323639	0.002047	0.001986	1.

Figure 7.1: Matrix plots showing the correlation factor between the set of selected control variables and the discriminating variable m (”bMass”) for q^2 bin 4 in year 2016 (left) and 2018 (right).

The ${}_s\mathcal{P}lot$ method is implemented with the `RooStats::SPlot` class. The results obtained are shown in Fig. 7.2, for q^2 bin 4 in year 2018, for the selected control variables. The signal and background distributions, $M_s(\bar{x})$ and $M_b(\bar{x})$, are shown in red and blue, respectively. The two distributions are normalised over the range of each control variable. In the y-axis, the normalised number of events divided by the bin width is shown.

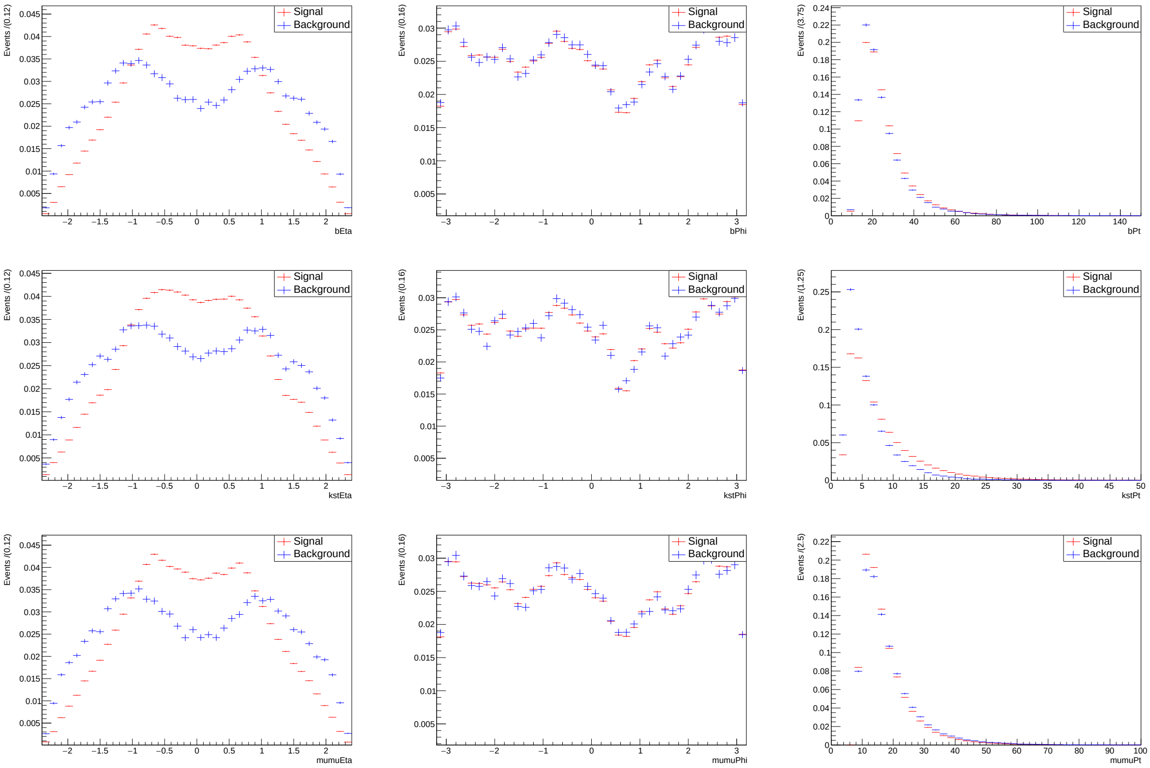


Figure 7.2: Signal (in red) and background (in blue) distributions obtained with the ${}_s\mathcal{P}lot$ method, in year 2018 for q^2 bin 4, for different control variables.

7.2 Data vs Monte Carlo comparison

We can now compare the signal distributions obtained with the ${}_s\mathcal{P}lot$ method described in the previous section with the MC. The results of this comparison for q^2 bin 4 can be found in Fig. 7.3 for the selected control variables in year 2018. The comparisons for the remaining data-taking years can be found in Appendix D. The signal data and MC distributions are shown in red and green, respectively, and they are both normalised to the range of the control variable. In the bottom panel, the bin by bin ratio between the two distributions can be seen. This ratio is computed using the `TRatioPlot` class, using the ”divsym” option. The ratio error bars per bin are given by the quadratic sum of the errors in the data and MC bins.

The largest discrepancies between data and MC are found for the pseudorapidity control variables and they are quantified in a systematic uncertainty on the final branching fraction result, as will be explained in Ch. 8. For this systematic study, the ratios between data and MC for the pseudorapidity of the B^0 meson in the J/ψ channel are used.

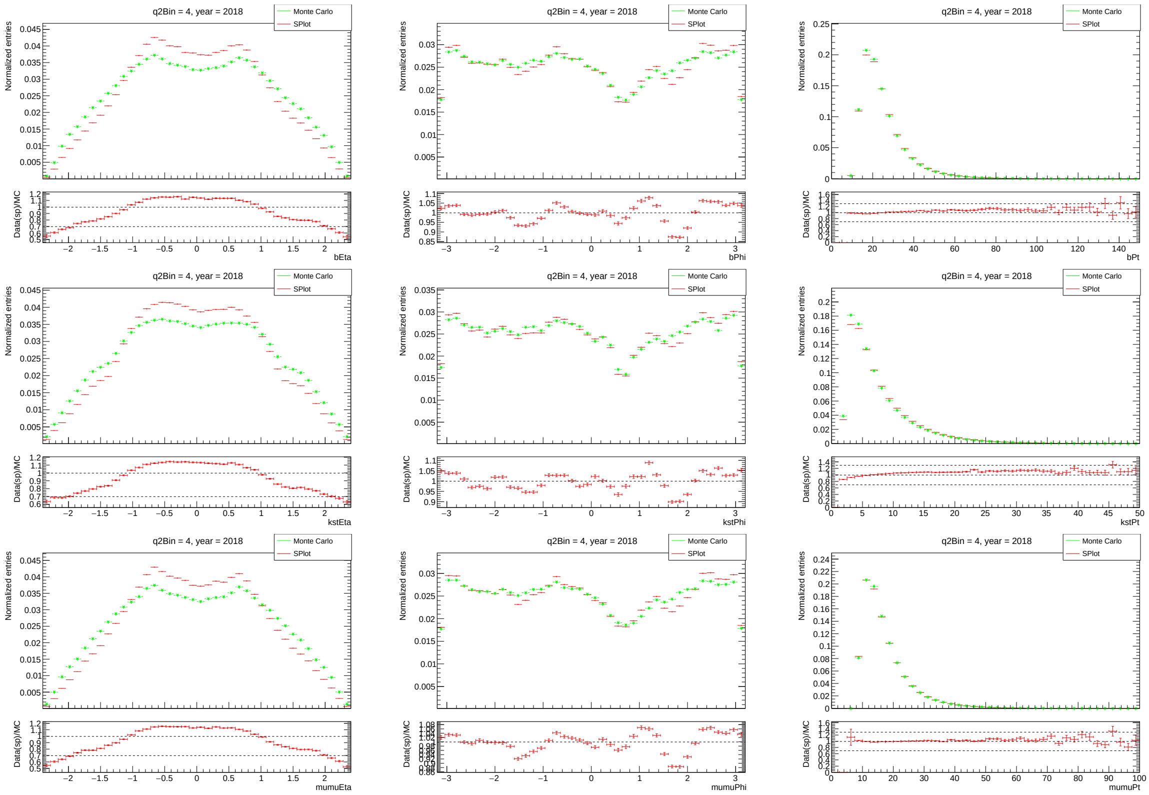


Figure 7.3: Comparison between data signal distribution (in red) obtained with the ${}_s\mathcal{P}lot$ method and the MC (in green), for different control variables in year 2018 for q^2 bin 4.

Chapter 8

Systematic uncertainties

In this chapter, we describe the studies of the different sources of systematic uncertainties in the analysis, which originate from the assumptions and methods used to extract the yields (Ch. 5) and compute the efficiencies (Ch. 6) necessary to obtain the $B^0 \to K^{*0}\mu^+\mu^-$ branching fraction defined in Eq. (1.1). We start in Sec. 8.1 to talk about the uncertainties associated with the model used to fit the $K^+\pi^-\mu^+\mu^-$ invariant mass and the fit bias (see Sec. 5.4). In Sec. 8.2, we explain how the discrepancies found between data and MC in Ch. 7 are quantified in a branching fraction systematic error. Since the method used to compute the averaged efficiencies in Sec. 6.3 makes use of the angular parameters obtained in Refs. [3, 13], the definition of a systematic uncertainty related with their statistical errors is explained in Sec. 8.3. The efficiency statistical uncertainties presented in Tab. 6.2 are propagated to the branching fraction systematic error. Finally, the uncertainties of the input branching fractions $\mathcal{B}(B^0 \to J/\psi K^{*0})$ and $\mathcal{B}(J/\psi \to \mu^+\mu^-)$ are also considered as part of the systematic error, being 3.94% and 0.55% [38], respectively.

8.1 Yield systematics

The systematic uncertainties associated with the yields arise from biases in the choice of the nominal fit and on how well the chosen model fits the observed data distribution. The former is estimated by using different fit configurations, other than the nominal presented in Ch. 5, to extract the yields Y_S and Y_N appearing in Eq. (1.1). The latter quantifies how reliable the yields obtained with the nominal fit are. The different fit variations used will be presented in Sec. 8.1.1. The total yields systematic uncertainties are presented in Sec. 8.1.2.

8.1.1 Fit variations

We use the following fit configurations

- **Scale factor**: all the fit parameters, except the RT mean and the yields, are fixed to the values found in MC. A "scale factor" parameter is multiplied by all the widths to allow for possible differences in the mass resolution between data and MC. This parameter is left free to float in the two resonant q^2 bins, J/ψ (bin 4) and $\psi(2S)$ (bin 6), and is fixed to the values obtained in the J/ψ channel for the signal q^2 bins. The scale factor fit results for 2018 can be seen in Fig. 8.1 and the scale factor parameters obtained in the two resonant channels for all the data-taking years can be found in Tab. 8.1. The largest value for the scale factor is found in q^2 bin 4, in year 2017, where the data prefers a shape 10% wider than in MC.

- **Mistag fraction**: nominal fit in which the fraction of WT events is fixed to the values obtained from a fit to the MC containing both RT and WT events, applying the same Gaussian constraints used to fit the data. It allows for possible differences in the amount of WT events in data and MC. The mistag fraction values obtained from the MC fit can be seen in Tab. 8.2 for all q^2 bins and data-taking years. The results of the fit to data with the mistag fraction fixed to these values can be seen in Fig. 8.2 for year 2018.

- **No erf**: the component describing the partially reconstructed events in the J/ψ channel is removed. The fit result in year 2018 can be seen in Fig. 8.3 (left).

- **Largest/lowest f_{erf}**: nominal fit in which the fraction of partially reconstructed events with respect to signal (f_{erf}) is fixed to its estimated largest and lowest limits, respectively. In order to estimate these limits, several fit configurations were performed in which f_{erf} was left free to float. The variation giving the largest values for f_{erf} is the one that has the largest number of degrees of freedom: the RT parameters are unconstrained. The variation giving the lowest values for f_{erf} is the one that has the lowest number of degrees of freedom: the scale factor variation. The f_{erf} values obtained with these two variations are collected in Tab. 8.3 for all data-taking years and the fit results for year 2018 can be seen in Fig. 8.3 (centre) and (right).

The fit results for the remaining data-taking years can be seen in Appendix E.

Table 8.1: Scale factor parameters resulting from the J/ψ and $\psi(2S)$ resonant channels fits in all data-taking years.

Year	J/ψ	$\psi(2S)$
2016	1.0694 ± 0.0026	1.0554 ± 0.0096
2017	1.1028 ± 0.0017	1.0793 ± 0.0062
2018	1.0786 ± 0.0011	1.0577 ± 0.0043

The yield uncertainties estimated with the fit variations are estimated in the following way. The difference between the signal yields obtained with each variation and the nominal fit is computed. Only the maximum difference between the variations "no erf", "largest f_{erf}" and "lowest f_{erf}" and the nominal fit

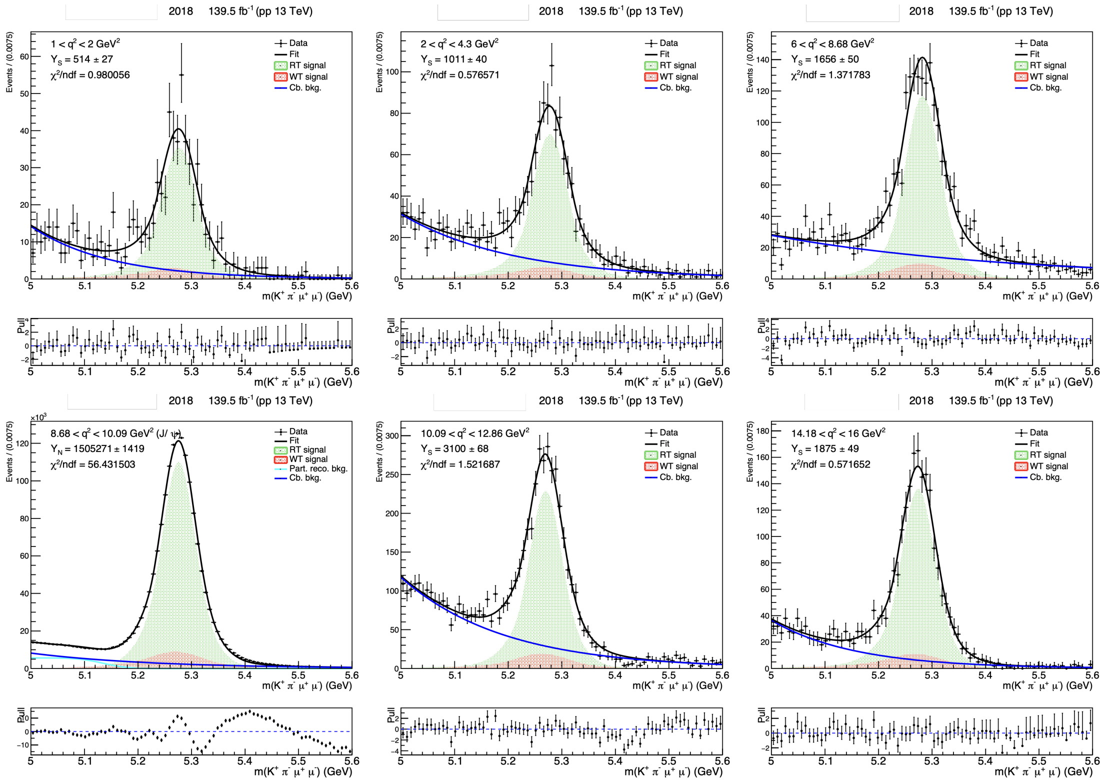


Figure 8.1: Scale factor fit results in year 2018. The results for q^2 bin 2 and 6 can be seen in Appendix E.

is considered in the J/ψ channel, since they all have to due with the fit component describing the partially reconstructed events. The systematic error is then given by the quadratic sum of each independent variation.

The relative signal yield differences for each variation can be found in Tab. 8.4, where the last column shows the total contribution of the fit variations to the yield systematics. In the signal q^2 bins, the largest contribution to the systematic error comes from the scale factor variation, whereas in the normalisation channel, it comes from the ”no erf” variation, in which only a single exponential is used to describe the background. The scale factor uncertainty is larger in year 2017, where the scale factor parameter

Table 8.2: Fraction of WT events obtained in the fit to the MC containing WT and RT events applying the same Gaussian constraints used to fit the data.

q^2 bin / Year	2016	2017	2018
0	0.1130 $\pm$ 0.0155	0.1098 $\pm$ 0.0123	0.1122 $\pm$ 0.0123
1	0.1171 $\pm$ 0.0107	0.1172 $\pm$ 0.0087	0.1170 $\pm$ 0.0080
2	0.1213 $\pm$ 0.0115	0.1157 $\pm$ 0.0102	0.1191 $\pm$ 0.0091
3	0.1191 $\pm$ 0.0091	0.1150 $\pm$ 0.0071	0.1200 $\pm$ 0.0065
4	0.1304 $\pm$ 0.0003	0.1266 $\pm$ 0.0003	0.1250 $\pm$ 0.0002
5	0.1403 $\pm$ 0.0078	0.1347 $\pm$ 0.0061	0.1301 $\pm$ 0.0049
6	0.1367 $\pm$ 0.0009	0.1270 $\pm$ 0.0007	0.1280 $\pm$ 0.0007
7	0.1313 $\pm$ 0.0085	0.1286 $\pm$ 0.0071	0.1243 $\pm$ 0.0061

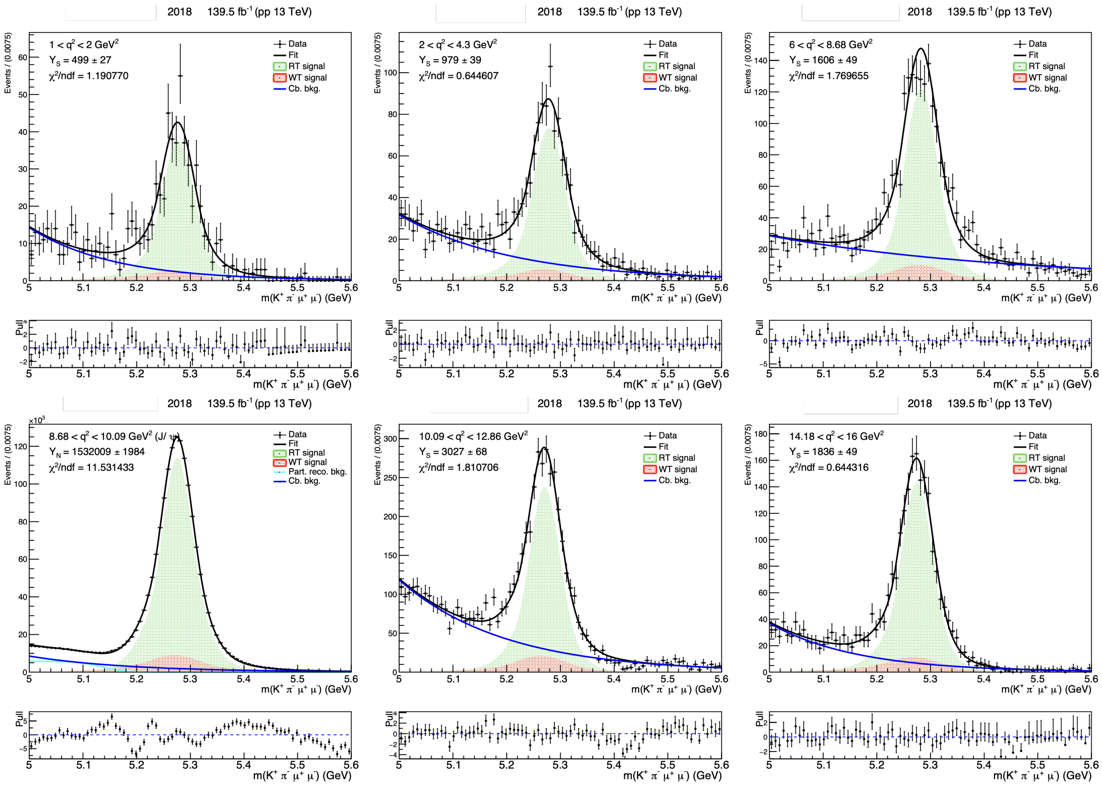


Figure 8.2: Mistag fraction fit results in year 2018. The results for q^2 bin 2 and 6 can be seen in Appendix E.

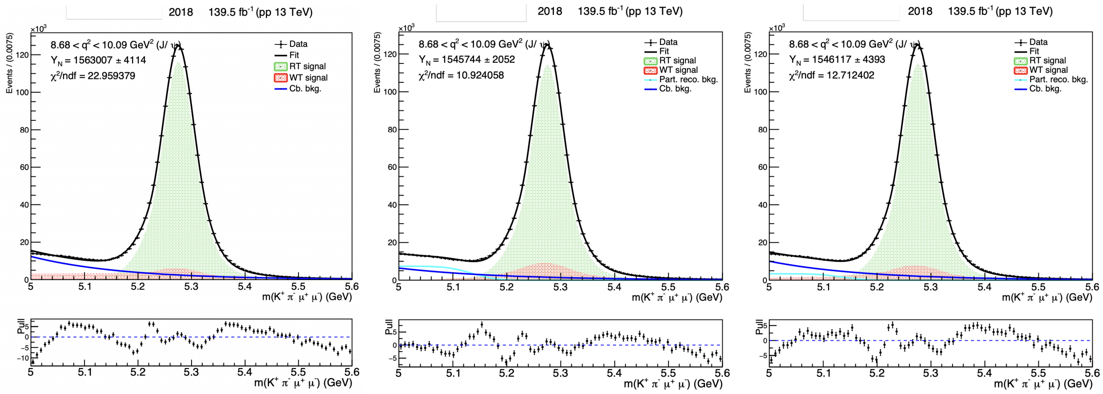


Figure 8.3: No erf (left), largest f_{erf} (centre) and lowest f_{erf} (right) fit results in year 2018.

Table 8.3: Largest and lowest values for f_{erf} obtained in the J/ψ channel, for all data-taking years.

Year	Largest f_{erf}	Lowest f_{erf}
2016	0.0750 ± 0.0032	0.0381 ± 0.0039
2017	0.0832 ± 0.0034	0.0386 ± 0.0023
2018	0.0859 ± 0.0025	0.0396 ± 0.0016

obtained in the J/ψ channel in Tab. 8.1 reaches its largest value. The largest error of 3.99 % is found in q^2 bin 3 in year 2017 and the lowest error of 0.62 % is found in q^2 bin 6 ($\psi(2S)$) in year 2016.

Table 8.4: Yield systematic uncertainties from the fit variations in %.

q^2 bin	Year	Mistag fraction (%)	Scale factor (%)	No erf (%)	Largest f_{erf} (%)	Lowest f_{erf} (%)	Total (%)
0	2016	0.32	2.55	—	—	—	2.57
	2017	0.39	3.30	—	—	—	3.33
	2018	0.40	2.71	—	—	—	2.74
1	2016	0.23	2.63	—	—	—	2.64
	2017	0.15	3.53	—	—	—	3.53
	2018	0.26	2.92	—	—	—	2.93
2	2016	0.20	2.80	—	—	—	2.81
	2017	0.32	3.77	—	—	—	3.78
	2018	0.25	3.01	—	—	—	3.02
3	2016	0.19	2.84	—	—	—	2.85
	2017	0.35	3.97	—	—	—	3.99
	2018	0.14	2.94	—	—	—	2.95
4	2016	0.00	0.09	2.33	0.69	0.97	2.33
	2017	0.01	0.52	2.52	0.78	1.07	2.57
	2018	0.00	1.75	2.02	0.90	0.92	2.67
5	2016	0.26	2.53	—	—	—	2.55
	2017	0.25	3.49	—	—	—	3.50
	2018	0.16	2.57	—	—	—	2.58
6	2016	0.01	0.62	—	—	—	0.62
	2017	0.00	2.14	—	—	—	2.14
	2018	0.02	0.86	—	—	—	0.86
7	2016	0.18	2.03	—	—	—	2.04
	2017	0.01	2.63	—	—	—	2.63
	2018	0.08	2.04	—	—	—	2.04

8.1.2 Total yield systematic uncertainty

The total yield systematic uncertainty is given by the quadratic sum of the two considered sources: fit variations and fit bias. The former is computed as explained in Sec. 8.1.1, whereas the latter is given by the product between the mean of the signal yield pull distributions presented in Sec. 5.4 and the yield statistical uncertainty. The fit variations and the fit bias uncertainties are collected in Tab. 8.5 as well as the total error, which can be seen in the last column. This error is propagated to the branching fraction.

8.2 Data vs MC discrepancies

The ratios between the B^0 pseudorapidity signal distribution from data (${}_s\mathcal{P}lot$) and MC in the J/ψ channel presented in Ch. 7 are used to compute a weighted efficiency

$$\epsilon^{wei} = \frac{N_{acc}}{D_{acc}} \times \left(\frac{N_{sel}}{D_{sel}}\right)^{wei} . \tag{8.1}$$

ϵ^{wei} is computed in the same way as the efficiency ϵ defined in Sec. 6.1, with the difference that the $\frac{D(\vec{p},\vec{\theta})^{data}}{D(\vec{p}',\vec{\theta})^{MC}}$ weights in Eq. (6.9) for the numerator N_{sel} and $\frac{N(\vec{p},\vec{\theta})^{data}}{N(\vec{p}',\vec{\theta})^{MC}}$ for the denominator D_{sel} are multiplied by the data/MC ratios. In Fig. 8.4, the averaged efficiencies ϵ (black) and the weighted efficiencies ϵ^{wei} (red) can be seen for years 2016 (left) and 2018 (right).

The systematic error associated with these discrepancies is estimated by computing the $B^0 \to K^{*0}\mu^+\mu^-$ branching fraction with ϵ and ϵ^{wei} to avoid correlations between the two, and the error is taken as the difference. The values of these systematic errors will be presented in Ch. 9.

Table 8.5: Yield systematic uncertainties in %.

q^2 bin	Year	Fit variations	Fit bias	Total
0	2016	2.57	0.53	2.62
	2017	3.33	0.25	3.34
	2018	2.74	0.28	2.75
1	2016	2.64	0.09	2.64
	2017	3.53	0.13	3.53
	2018	2.93	0.03	2.93
2	2016	2.81	0.02	2.81
	2017	3.78	0.12	3.79
	2018	3.02	0.07	3.02
3	2016	2.85	0.00	2.85
	2017	3.99	0.05	3.99
	2018	2.95	0.09	2.95
4	2016	2.33	0.03	2.33
	2017	2.57	0.01	2.57
	2018	2.67	0.00	2.67
5	2016	2.55	0.00	2.55
	2017	3.50	0.07	3.50
	2018	2.58	0.02	2.58
6	2016	0.62	0.10	0.63
	2017	2.14	0.03	2.14
	2018	0.86	0.04	0.86
7	2016	2.04	0.13	2.04
	2017	2.63	0.04	2.63
	2018	2.04	0.04	2.04

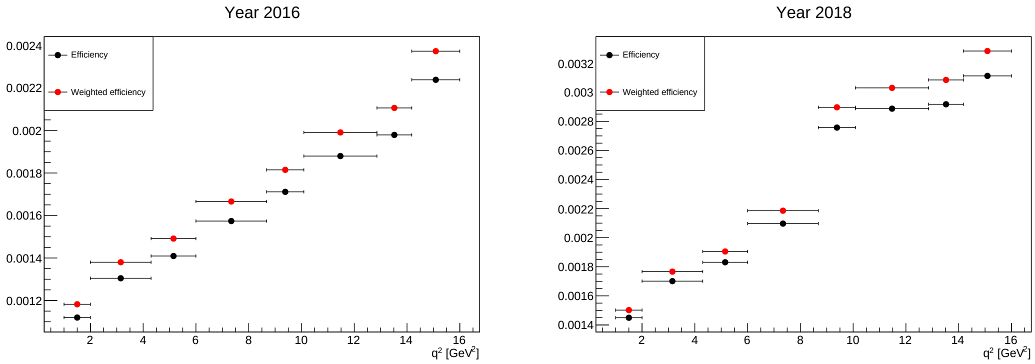


Figure 8.4: Averaged efficiencies computed as explained in Sec. 6.3 (black) and averaged efficiencies weighted by the data/MC ratios presented in Ch. 7 (red) for years 2016 (left) and 2018 (right).

8.3 Angular parameters statistical uncertainties

The angular parameters obtained from the data and MC fits, which are taken from Refs. [3, 13] are an input to the averaged efficiencies computation explained in Sec. 6.3. As a consequence, their statistical uncertainties need to be accounted for in a systematic uncertainty.

For this purpose, two additional ”efficiencies” ϵ^{+} and ϵ^{-} are computed. The former, uses the nominal values plus statistical uncertainties of all the angular parameters simultaneously, whereas the latter uses their nominal values minus statistical uncertainties. In Fig. 8.5, the ϵ (black), ϵ^+ (red) and ϵ^- (blue) efficiencies can be seen for years 2016 (left) and 2018 (right).

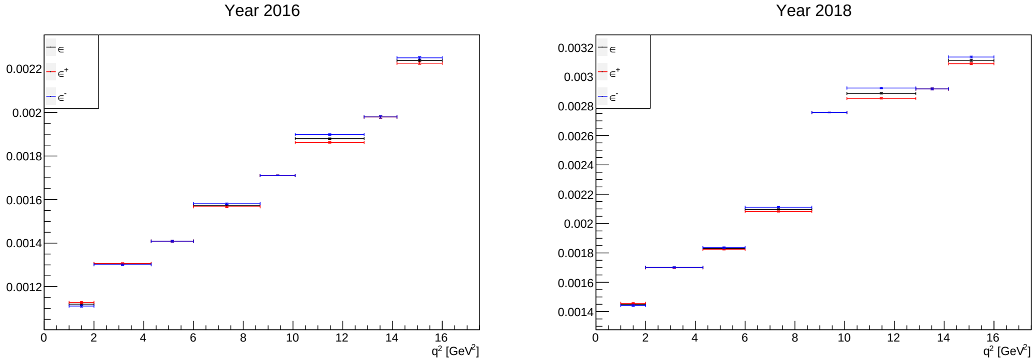


Figure 8.5: Averaged efficiencies ϵ computed as explained in Sec. 6.3 (black) and the efficiencies computed with the nominal values plus statistical uncertainties of all the angular parameters simultaneously ϵ^+ (red) and nominal values minus statistical uncertainties ϵ^- (blue) for years 2016 (left) and 2018 (right).

Similarly to what was explained in Sec. 8.2, two additional branching fractions are calculated using ϵ^+ and ϵ^- to avoid correlations between ϵ, ϵ^+ and ϵ^-. The difference between the two and the nominal branching fraction is computed and the systematic error is defined as the maximum difference with respect to the nominal value. The values of these errors will be presented in Ch. 9.

Chapter 9

Results and discussion

In this chapter, we present the main result of the analysis, the $B^0 \to K^{*0}\mu^+\mu^-$ differential branching fraction as a function of the di-muon invariant mass squared (q^2). We start in Sec. 9.1 by validating the analysis, comparing the obtained ratio between the $B^0 \to \psi(2S)K^{*0}$ and $B^0 \to J/\psi K^{*0}$ branching fractions with the current world average [38]. We then proceed in Sec. 9.2 by presenting the $B^0 \to K^{*0}\mu^+\mu^-$ differential branching fraction results and comparing them with the previous CMS [4] and LHCb [14] calculations and the SM predictions [15, 16]. In Sec. 9.2.1 we present the single year results, whereas in Sec. 9.2.2 we present the three years weighted average. Finally, in Sec. 9.2.3, we study the effect that the efficiency angular correction described in Sec. 6.3 has in the branching fraction measurement.

9.1 Analysis validation

The analysis is validated by comparing the ratio between the branching fractions of the two resonant channels, $B^0 \to \psi(2S)K^{*0}$ and $B^0 \to J/\psi K^{*0}$, with the current world average [38]. The PDG value is given by

$$R_{PDG} = \frac{\mathcal{B}(B^0 \to \psi(2S)K^{*0}) \times \mathcal{B}(\psi(2S) \to \mu^+\mu^-)}{\mathcal{B}(B^0 \to J/\psi K^{*0}) \times \mathcal{B}(J/\psi \to \mu^+\mu^-)}, \tag{9.1}$$

where $\mathcal{B}(B^0 \to \psi(2S)K^{*0}) = (5.9 \pm 0.4) \times 10^{-4}$, $\mathcal{B}(\psi(2S) \to \mu^+\mu^-) = (80 \pm 6) \times 10^{-4}$, $\mathcal{B}(B^0 \to J/\psi K^{*0}) = (1.27 \pm 0.05) \times 10^{-3}$ and $\mathcal{B}(J/\psi \to \mu^+\mu^-) = (5.961 \pm 0.033) \times 10^{-2}$. We compute this ratio as

$$R = \frac{Y}{Y_N}\frac{\epsilon_N}{\epsilon}, \tag{9.2}$$

where Y, Y_N are the $B^0 \to \psi(2S)K^{*0}$ and $B^0 \to J/\psi K^{*0}$ signal yields, respectively, obtained as explained in Ch. 5 and ϵ, ϵ_N are the $B^0 \to \psi(2S)K^{*0}$ and $B^0 \to J/\psi K^{*0}$ averaged efficiencies, respectively, computed as described in Sec. 6.3.

In Fig. 9.1, the obtained results for this ratio as well as the PDG value can be seen with the respective statistical (black) and systematic (red) uncertainties. The PDG value is shown in the bottom and the dashed blue line indicates its nominal value for reference. The single year results as well as the three years combined result (see Sec. 9.2.2) in the top can be seen. In Tab. 9.1, the nominal values for the ratio as well as the relative statistical and systematic uncertainties can be seen. The three years combined result is in agreement with the PDG value within 0.2 σ, therefore validating the analysis.

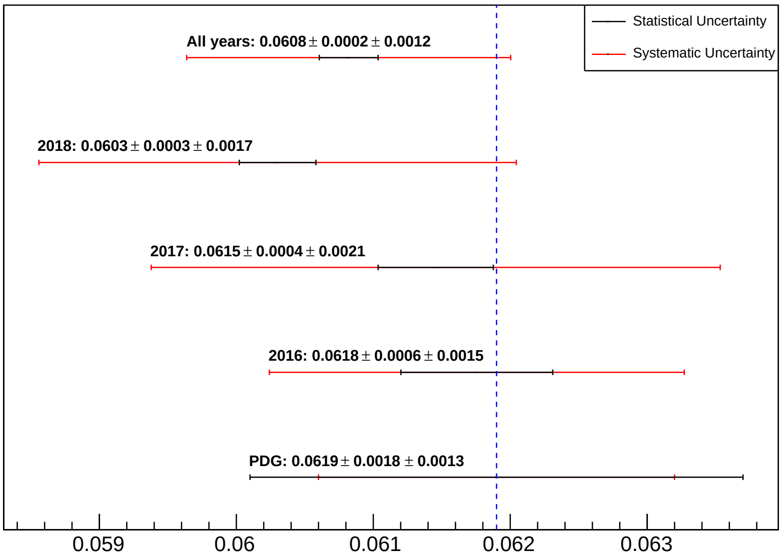


Figure 9.1: Single year and three year average (top) results for the ratio between the branching fractions of the two resonant channels, $B^0 \to \psi(2S)K^{*0}$ and $B^0 \to J/\psi K^{*0}$, with the respective statistical (black) and systematic (red) uncertainties. On the bottom, the PDG value can be seen with the dashed blue line indicating its nominal value for reference.

Table 9.1: Resonant channels branching fraction ratio with relative uncertainties in %.

Year	BF ratio	Stat. error (%)	Syst. error (%)
2016	0.0618 ± 0.0006 ± 0.0015	0.90	2.45
2017	0.0615 ± 0.0004 ± 0.0021	0.68	3.38
2018	0.0603 ± 0.0003 ± 0.0017	0.46	2.89

9.2 $B^0 \to K^{*0}\mu^+\mu^-$ differential branching fraction

Gathering the results of Ch. 5 and Ch. 6 for the determination of the yields and efficiencies, respectively, we present in this section the main result of the analysis, the $B^0 \to K^{*0}\mu^+\mu^-$ differential branching fraction. The results presented in Sec. 9.2.1 and Sec. 9.2.2 are obtained with the angular corrected averaged efficiencies described in Sec. 6.3. In Sec. 9.2.3, we study the impact of this correction by comparing these results with the ones obtained with the MC based efficiencies described in Sec. 6.2.

9.2.1 Single year results

In Fig. 9.2 the $B^0 \to K^{*0}\mu^+\mu^-$ differential branching fraction in years 2016 (top left), 2017 (top right) and 2018 (bottom) can be seen. Our results in black are superimposed with the previous CMS Run 1 measurement in lilac. In the plots, both statistical and systematic uncertainties are shown. The former has a horizontal tick in the end whereas the latter does not. The uncertainties of the input $\mathcal{B}(B^0 \to J/\psi K^{*0})$ and $\mathcal{B}(J/\psi \to \mu^+\mu^-)$ branching fractions are not included in the systematic error bars. The two grey bands indicate the q^2 ranges of the two resonant channels, $B^0 \to J/\psi K^{*0}$ and $B^0 \to \psi(2S)K^{*0}$.

From the plots we can see that the single year results are all consistent with the previous CMS Run 1 calculations [4]. The largest discrepancies are found at the highest q^2 bins, in particular in q^2 bin 7. They are however all within 3 σ. In terms of the precision of the results, while in 2016 the size of the error bars is comparable with the Run 1 analysis, wee see a significant reduction in 2017, culminating in 2018. In Tab. 9.2, the factor of increase in the precision of the results in comparison with the previous CMS Run 1 calculations can be seen.

In Tabs. 9.3, 9.4 and 9.5 the relative branching fraction uncertainties (in %) for years 2016, 2017 and 2018, respectively, can be seen. In the first two rows, the total statistical and systematic error of the results is shown. In the other rows, the contribution of the different sources of systematic uncertainty described in Ch. 8 is shown in %. The sources with the largest contribution to the branching fraction systematic error come from the yield uncertainties (fit variations) and from the $\mathcal{B}(B^0 \to J/\psi K^{*0})$ PDG uncertainty. The systematic uncertainties are comparable with the statistical uncertainties in most q^2 bins. In 2016, the statistical uncertainties dominate for the three first signal q^2 bins. In year 2018, the systematic uncertainties dominate in the last three signal q^2 bins.

Table 9.2: Factor of increase in the precision of the branching fraction measurement between our results and the previous CMS Run 1 analysis [4].

Year / q^2 bin	0	1	2	3	5	7
2016	1.39	1.72	1.65	1.63	1.92	2.81
2017	2.03	2.40	2.23	1.92	2.38	3.48
2018	2.37	2.73	2.67	2.49	2.75	4.03

Table 9.3: Differential branching fraction uncertainties - year 2016

	0	1	2	3	5	7	Normalization
Stat. error ($\%$)	11.12	8.42	9.87	6.31	5.02	6.38	—
Syst. error ($\%$)	5.40	5.33	5.41	5.44	5.36	5.08	—
Yield syst.	2.62	2.64	2.81	2.85	2.55	2.04	2.33
Data-MC disc.	0.44	0.25	0.22	0.19	0.12	0.03	—
MC finite size	0.42	0.27	0.31	0.22	0.19	0.23	0.10
Angular param.	0.79	0.30	0.03	0.44	0.99	0.55	—
$B(B^0 \to J/\psi K^{*0})$	3.94	3.94	3.94	3.94	3.94	3.94	—
$B(J/\psi \to J/\mu^+\mu^-)$	0.55	0.55	0.55	0.55	0.55	0.55	—

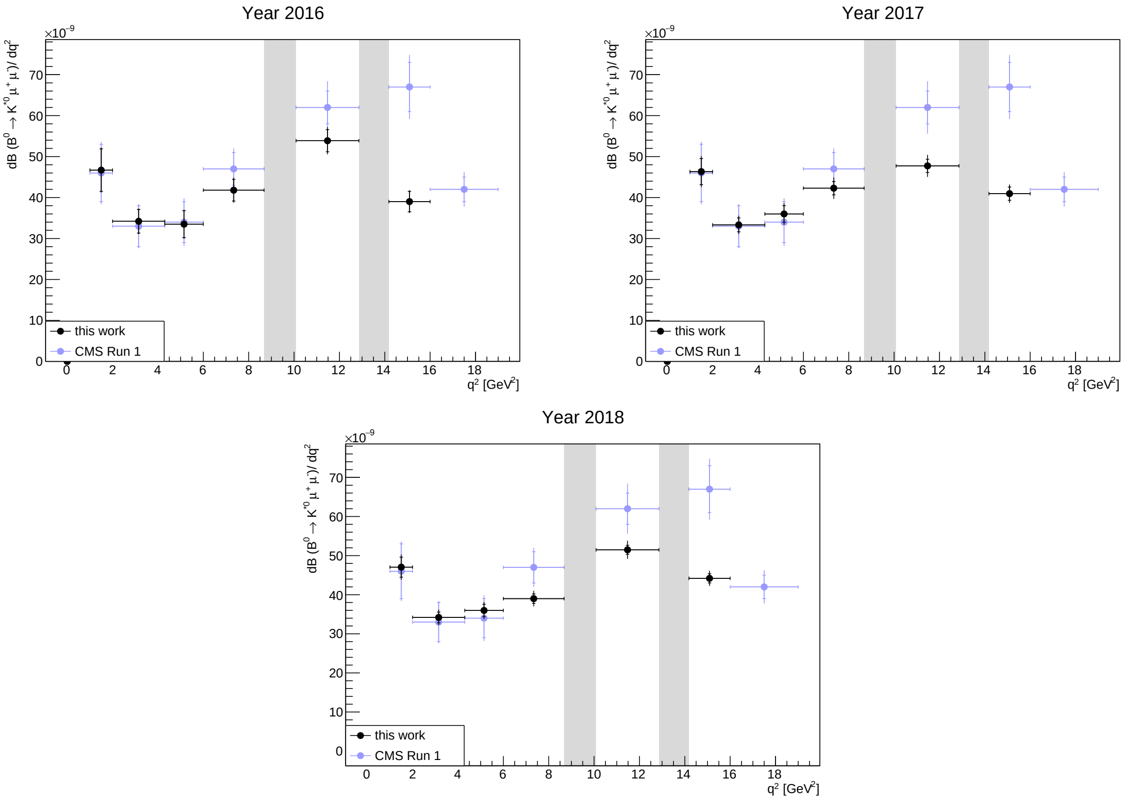


Figure 9.2: $B^0 \to K^{*0}\mu^+\mu^-$ differential branching fraction single year results: 2016 (top left), 2017 (top right) and 2018 (bottom). Our results in black can be seen superimposed with the CMS Run 1 results in lilac.

Table 9.4: Differential branching fraction uncertainties - year 2017

	0	1	2	3	5	7	Normalization
Stat. error (%)	6.87	5.11	5.61	3.85	3.37	3.96	—
Syst. error (%)	5.87	5.92	6.08	6.24	6.03	5.49	—
Yield syst.	3.34	3.53	3.79	3.99	3.50	2.63	2.57
Data-MC disc.	0.60	0.34	0.21	0.19	0.12	0.12	—
MC finite size	0.38	0.25	0.29	0.21	0.17	0.22	0.09
Angular param.	0.61	0.11	0.35	0.76	1.26	0.89	—
$B(B^0 \to J/\psi K^{*0})$	3.94	3.94	3.94	3.94	3.94	3.94	—
$B(J/\psi \to J/\mu^+\mu^-)$	0.55	0.55	0.55	0.55	0.55	0.55	—

9.2.2 Three years averaged result

The combination of the branching fraction results from the three years presented in Sec. 9.2.1 is performed using the weighted average

$$\bar{B} = \frac{\sum_{i=1}^{3} w_i B_i}{\sum_{i=1}^{3} w_i}, \tag{9.3}$$

where B_i is the value of each measurement with statistical uncertainty σ_i.The weights w_i are defined as $w_i = 1/\sigma_i$. The weighted average in Eq. (9.3) is the CMS statistical committee recommendation [71] to

Table 9.5: Differential branching fraction uncertainties - year 2018

	0	1	2	3	5	7	Normalization
Stat. error (%)	5.45	4.01	4.42	3.09	2.24	2.68	—
Syst. error (%)	5.74	5.74	5.76	5.73	5.58	5.28	—
Yield syst.	2.75	2.93	3.02	2.95	2.58	2.04	2.67
Data-MC disc.	1.40	1.16	0.97	0.80	0.11	0.41	—
MC finite size	0.39	0.25	0.29	0.20	0.16	0.20	0.09
Angular param.	0.54	0.15	0.31	0.69	1.24	0.73	—
$B(B^0 \to J/\psi K^{*0})$	3.94	3.94	3.94	3.94	3.94	3.94	—
$B(J/\psi \to J/\mu^+\mu^-)$	0.55	0.55	0.55	0.55	0.55	0.55	—

combine the results of three uncorrelated measurements.

In Fig. 9.3, the three years averaged result can be seen in black. Both the statistical and systematic error bars are shown in the plot. The former has a horizontal tick in the end, whereas the latter does not. The two grey bands indicate the q^2 ranges of the two resonant channels. Superimposed in the plot are the previous CMS [4] (lilac) and LHCb [14] (grey) Run 1 results as well as the SM predictions [15,16] (pink shaded region). At low q^2, below the charmonium resonances, the $B^0 \to K^{*0}$ form factor calculations come from LCSRs [15] whereas at high q^2, they come from lattice QCD [16].

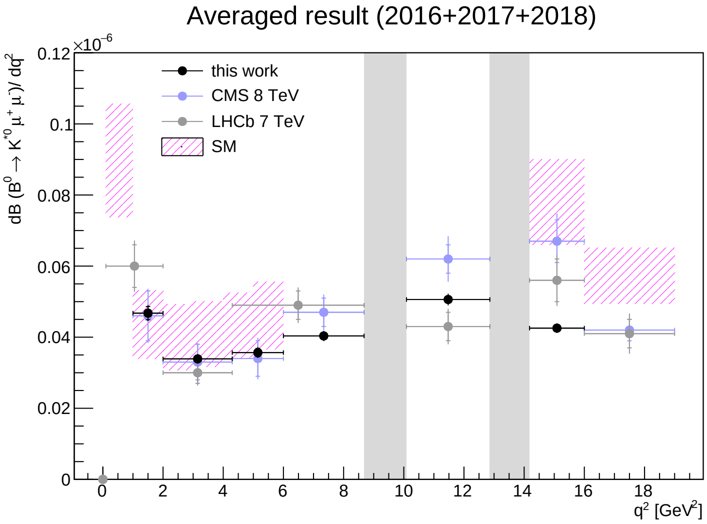


Figure 9.3: Three years average result (black) superimposed with the previous CMS [4] (lilac) and LHCb [14] (grey) Run 1 results as well as the SM predictions [15,16] (pink shaded region). The two grey bands represent the q^2 ranges of the two resonant channels.

From Fig. 9.3, we can see that our results are consistent with the previous CMS [4] and LHCb [14] Run 1 measurements while having an increased precision. The factor of increase in the precision of the results in comparison with the previous CMS analysis can be seen in Tab. 9.6. Our results are also consistent with the SM. In Tab. 9.7, the 3-years averaged results for all q^2 bins can be seen with the respective relative statistical and systematic uncertainties. The systematic uncertainties are, in general,

Table 9.6: Factor of increase in the precision of the branching fraction measurement (three years averaged result) in comparison with the previous CMS Run 1 calculation [4].

q^2 bin	0	1	2	3	5	7
Factor	3.39	3.96	3.80	2.80	3.06	4.77

Table 9.7: Differential branching fraction - weighted average

q^2	dB/dq^2 $[10^{-7}$ GeV$^{-2}]$	Stat. error (%)	Syst. error (%)
0	0.47 ± 0.02 ± 0.02	3.99	3.71
1	0.34 ± 0.01 ± 0.01	2.95	3.75
2	0.36 ± 0.01 ± 0.01	3.28	3.80
3	0.40 ± 0.01 ± 0.02	2.25	3.81
5	0.51 ± 0.01 ± 0.02	1.75	3.82
7	0.43 ± 0.01 ± 0.02	2.10	3.56

comparable with the statistical uncertainties.

9.2.3 Effect of the angular correction

In Fig. 9.4, the branching fraction calculated with the angular corrected efficiencies described in Sec. 6.3 (blue) and the MC based efficiencies described in Sec. 6.2 (black) can be seen. The error bars contain both statistical and systematic uncertainties. The former has a horizontal tick in the end whereas the latter does not. In the case of the MC based results, both the higher and lower limits of the efficiency statistical uncertainties presented in Tab. 6.1 are considered. Superimposed in the plots are also the SM predictions [15, 16] (pink shaded region). The angular correction does not change significantly the nominal values of the $B^0 \to K^{*0}\mu^+\mu^-$ branching fraction.

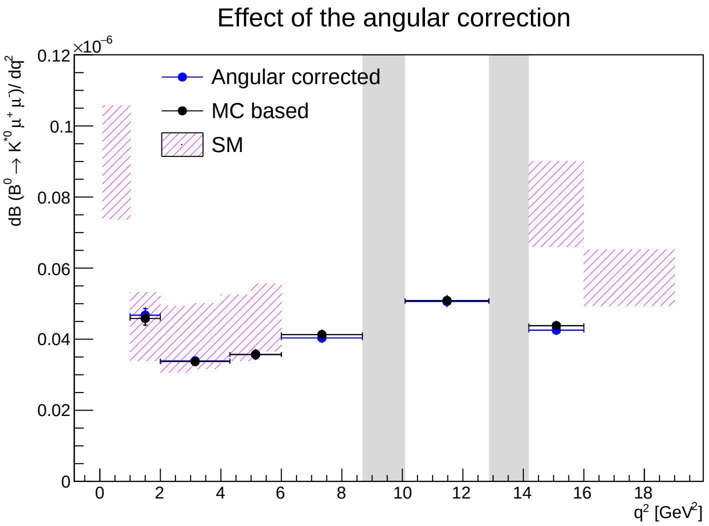


Figure 9.4: Comparison between the branching fraction results obtained with the angular corrected efficiencies described in Sec. 6.3 (blue) and the MC based efficiencies described in Sec. 6.2 (black). Superimposed are the SM predictions [15, 16] (pink shaded region). The two grey bands represent the q^2 ranges of the two resonant channels.

Chapter 10

Conclusions

In this thesis, we measured the differential branching fraction of the $B^0 \to K^{*0}\mu^+\mu^-$ decay as a function of the di-muon invariant mass squared (q^2), using the $B^0 \to J/\psi K^{*0}$ resonance as normalisation channel. The work of the thesis was inserted in a larger analysis [13] which also has the goal of measuring several angular parameters appearing in the $B^0 \to K^{*0}\mu^+\mu^-$ decay rate. We used pp collision data at $\sqrt{s} = 13\,\text{TeV}$ collected by the CMS experiment in LHC Run 2. The analysis was performed independently for each data-taking year and q^2 bin.

The analysis was validated by computing the ratio of branching fractions of the resonant $B^0 \to J/\psi K^{*0}$ and $B^0 \to \psi(2S)K^{*0}$ channels and comparing it with the current world average [38]. Our result was consistent with that value within $0.2\,\sigma$, therefore validating our analysis. The $B^0 \to K^{*0}\mu^+\mu^-$ branching fraction results were also found to be consistent with the previous CMS [4] and LHCb [14] measurements in Run 1 and with the SM predictions [15, 16], and are the most precise to date.

The $B^0 \to K^{*0}\mu^+\mu^-$ decay is a rare beauty process that can only proceed at loop order in the SM, making it very sensitive to NP. Tensions with the SM have been reported in one of the angular parameters appearing in its decay rate, P_5'. Other tensions in the flavour sector have also been reported in other decays and observables, being commonly referred to as flavour anomalies. The $B^0 \to K^{*0}\mu^+\mu^-$ branching fraction can enter in global fits, helping to constrain NP scenarios and to investigate the source of these alluring flavour anomalies.

For the measurement of the $B^0 \to K^{*0}\mu^+\mu^-$ differential branching fraction two quantities were needed: the measured signal yield and the detector efficiency. The former was obtained by performing a one-dimensional extended unbinned maximum likelihood fit to the $K^+\pi^-\mu^+\mu^-$ invariant mass. The latter was obtained in two ways. Firstly, the efficiencies were calculated by simply determining the fraction of simulated events passing the selections described in Chs. 3 and 4 (MC based efficiency). Since in the analysis we want to probe the flavour anomalies and the angular parameters in the MC simulations are fixed to values close to the SM predictions, a second method was employed, which consisted on making an average of the efficiencies over the angular variables, θ_K, θ_l and ϕ, weighted by the $B^0 \to K^{*0}\mu^+\mu^-$ decay rate. For this, the angular parameters obtained in Refs. [3, 13] were used. We compared the branching fractions obtained with these two methods and concluded that the angular correction does not affect significantly the nominal values.

The work presented in this thesis is part of the ongoing effort to achieve a better understanding of the flavour anomalies. Future endeavours are also being advanced by both experimental and theoretical communities. The accumulation of larger datasets (LHC Run 3, BelleII) will result in a reduction of statistical uncertainties. The complementary exploration made by different experiments will increase the robustness of the results by creating orthogonal systematic effects. The theoretical uncertainties can be reduced by exploring additional observables, with a smaller form-factor dependence. In the future, if the statistical significance of the anomalies crosses the $5\,\sigma$ barrier, they will be the first established NP seen at the LHC.

Bibliography

[1] LHCb, “Branching fraction measurements of the rare $B_s^0 \to \phi\mu^+\mu^-$ and $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays,” 2021, arxiv:2105.14007 [hep-ex].

[2] LHCb, “Test of lepton universality in beauty-quark decays,” 2021, arxiv:2103.11769 [hep-ex].

[3] LHCb, “Measurement of CP-averaged observables in the $B^0 \to K^{*0}\mu^+\mu^-$ decay,” 2020, arxiv:2003.04831 [hep-ex]. [Online]. Available: https://doi.org/10.1103/PhysRevLett.125.011802

[4] CMS, “Angular analysis of the decay $B^0 \to K^{*0}\mu^+\mu^-$ from pp collisions at $\sqrt{s} = 8$ TeV,” 2016, arxiv:1507.08126 [hep-ex]. [Online]. Available: https://doi.org/10.1016/j.physletb.2015.12.020

[5] T. Blake, G. Lanfranchi, and D. M. Straub, “Rare B decays as tests of the standard model,” 2017, arxiv:1606.00916 [hep-ph]. [Online]. Available: https://doi.org/10.1016/j.ppnp.2016.10.001

[6] D. B. Kaplan, “Five lectures on effective field theory,” 2005, arxiv:nucl-th/0510023.

[7] A. Boletti, “Measurement of angular parameters from the decay $B^0 \to K^{*0}\mu^+\mu^-$ in proton-proton collisions at $\sqrt{s} = 8$ TeV,” Ph.D. dissertation, Università degli Studi di Padova, 2018.

[8] W. Altmannshofer and P. Stangl, “New physics in rare B decays after Moriond 2021,” 2021, arxiv:2103.13370 [hep-ph].

[9] CMS, “CMS luminosity,” accessed on October 2nd 2021. [Online]. Available: https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults

[10] CMS, “CMS track reconstruction performance during Run 2 and developments for Run 3,” 2021, arxiv:2012.07035 [physics.ins-det]. [Online]. Available: https://doi.org/10.22323/1.390.0733

[11] CMS, “Level-1 muon trigger performance in 2017 data and comparison with the legacy muon trigger system,” 2017, DP-2017/041. [Online]. Available: https://cds.cern.ch/record/2286327/

[12] CMS, “Level-1 muon trigger performance with the full 2017 dataset,” 2018, DP-2018/002. [Online]. Available: https://cds.cern.ch/record/2306376/

[13] CMS, "Full angular analysis of decay $B^0 \to K^{*0}\mu^+\mu^-$ at 13 TeV," 2021, CMS-AN-18-138.

[14] LHCb, "Differential branching fraction and angular analysis of the decay $B^0 \to K^{*0}\mu^+\mu^-$," 2013, arxiv:1304.6325 [hep-ex]. [Online]. Available: https://doi.org/10.1007/JHEP08(2013)131

[15] A. Bharucha, D. M. Straub, and R. Zwicky, "$B \to V\ell^+\ell^-$ in the standard model from light-cone sum rules," 2016, arxiv:1503.05534 [hep-ph]. [Online]. Available: https://doi.org/10.1007/JHEP08(2016)098

[16] R. R. Horgan, Z. Liu, S. Meinel, and M. Wingate, "Rare B decays using lattice QCD form factors," 2015, arxiv:1501.00367 [hep-lat]. [Online]. Available: https://doi.org/10.22323/1.214.0372

[17] Y. Fukuda *et al.*, "Measurement of the flux and zenith-angle distribution of upward throughgoing muons by super-kamiokande," *Phys. Rev. Lett.*, vol. 82, pp. 2644–2648, 1999. [Online]. Available: https://doi.org/10.1103/PhysRevLett.82.2644

[18] Belle, "Test of lepton flavor universality and search for lepton flavor violation in $B \to K\ell\ell$ decays," 2021, arxiv:1908.01848 [hep-ex]. [Online]. Available: https://doi.org/10.1007/JHEP03(2021)105

[19] BaBar, "Measurement of branching fractions and rate asymmetries in the rare decays $B \to K^{(*)}l^+l^-$," 2012, arxiv:1204.3933 [hep-ex]. [Online]. Available: https://doi.org/10.1103/PhysRevD.86.032012

[20] BaBar, "Evidence for an excess of $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ decays," 2012, arxiv:1205.5442 [hep-ex]. [Online]. Available: https://doi.org/10.1103/PhysRevLett.109.101802

[21] LHCb, "Test of lepton flavor universality by the measurement of the $B^0 \to D^{*-}\tau^+\nu_\tau$ branching fraction using three-prong τ decays," 2018, arxiv:1711.02505 [hep-ex]. [Online]. Available: https://doi.org/10.1103/PhysRevD.97.072013

[22] Belle, "Measurement of $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ with a semileptonic tagging method," 2019, arxiv:1904.08794 [hep-ex].

[23] Muon g-2, "Measurement of the positive muon anomalous magnetic moment to 0.46 ppm," 2021, arxiv:2104.03281 [hep-ex]. [Online]. Available: https://doi.org/10.1103/PhysRevLett.126.141801

[24] A. Greljo, P. Stangl, and A. E. Thomsen, "A model of muon anomalies," 2021, arxiv:2103.13991 [hep-ph]. [Online]. Available: https://doi.org/10.1016/j.physletb.2021.136554

[25] G. Hiller, D. Loose, and I. Nišandžić, "Flavorful leptoquarks at the LHC and beyond: spin 1," 2021, arxiv:2103.12724 [hep-ph]. [Online]. Available: https://doi.org/10.1007/JHEP06(2021)080

[26] D. Marzocca and S. Trifinopoulos, "A minimal explanation of flavour anomalies: B-meson decays, muon magnetic moment, and the Cabibbo angle," 2021, arxiv:2104.05730 [hep-ph]. [Online]. Available: https://doi.org/10.1103/PhysRevLett.127.061803

[27] W.-F. Chang, "One colorful resolution to the neutrino mass generation, three lepton flavor universality anomalies, and the Cabibbo angle anomaly," 2021, arxiv:2105.06917 [hep-ph]. [Online]. Available: https://doi.org/10.1007/JHEP09(2021)043

[28] CMS, "Angular analysis and branching fraction measurement of the decay $B^0 \to K^{*0}\mu^+\mu^-$," 2013, arxiv:1308.3409 [hep-ex]. [Online]. Available: https://doi.org/10.1016/j.physletb.2013.10.017

[29] CMS, "Measurement of angular parameters from the decay $\mathrm{B}^0 \to \mathrm{K}^{*0}\mu^+\mu^-$ in proton-proton collisions at $\sqrt{s} = 8$ TeV," 2018, arxiv:1710.02846 [hep-ex]. [Online]. Available: https://doi.org/10.1016/j.physletb.2018.04.030

[30] Belle, "Measurement of the differential branching fraction and forward-backword asymmetry for $B \to K^{(*)}\ell^+\ell^-$," 2009, arxiv:0904.0770 [hep-ex]. [Online]. Available: https://doi.org/10.1103/PhysRevLett.103.171801

[31] BaBar, "Angular distributions in the decays $B^0 \to K^* l^+ l^-$," 2009, arxiv:0804.4412 [hep-ex]. [Online]. Available: https://doi.org/10.1103/PhysRevD.79.031102

[32] CDF, "Measurements of the angular distributions in the decays $B \to K^{(*)}\mu^+\mu^-$ at CDF," 2012. [Online]. Available: https://doi.org/10.1103/PhysRevLett.108.081807

[33] LHCb, "Measurement of form-factor-independent observables in the decay $B^0 \to K^{*0}\mu^+\mu^-$," *Phys. Rev. Lett.*, 2013. [Online]. Available: https://doi.org/10.1103/PhysRevLett.111.191801

[34] Belle, "Lepton-flavor-dependent angular analysis of $B \to K^*\ell^+\ell^-$," *Phys. Rev. Lett.*, 2017. [Online]. Available: https://doi.org/10.1103/PhysRevLett.118.111801

[35] ATLAS, "Angular analysis of $B_d^0 \to K^*\mu^+\mu^-$ decays in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector," 2018, arxiv:1805.04000 [hep-ex]. [Online]. Available: https://doi.org/10.1007/JHEP10(2018)047

[36] LHCb, "Angular analysis of the $B^0 \to K^{*0}\mu^+\mu^-$ decay using 3 fb^{-1} of integrated luminosity," 2016, arxiv:1512.04442 [hep-ex]. [Online]. Available: https://doi.org/10.1007/JHEP02(2016)104

[37] LHCb, "Measurement of CP-averaged observables in the $B^0 \to K^{*0}\mu^+\mu^-$ decay," *Phys. Rev. Lett.*, 2020. [Online]. Available: https://doi.org/10.1103/PhysRevLett.125.011802

[38] PDG, "Review of particle physics," *Progress of Theoretical and Experimental Physics*, vol. 2020, no. 8, 2020, 083C01. [Online]. Available: https://doi.org/10.1093/ptep/ptaa104

[39] T. Appelquist and J. Carazzone, “Infrared singularities and massive fields,” *Phys. Rev. D*, vol. 11, pp. 2856–2861, 1975. [Online]. Available: https://doi.org/10.1103/PhysRevD.11.2856

[40] J. M. Campbell, E. W. N. Glover, and C. J. Maxwell, “Determination of the QCD parameter $\Lambda^{(5)}_{\bar{MS}}$ from the measured energy dependence of the average value 1-Thrust,” 1998, arxiv:hep-ph/9803254. [Online]. Available: https://doi.org/10.1103/PhysRevLett.81.1568

[41] A. Khodjamirian, “QCD sum rules for heavy flavor physics,” 2001, arxiv:hep-ph/0108205. [Online]. Available: https://doi.org/10.1063/1.1435927

[42] M. Beneke and M. Neubert, “QCD factorization for B→PP and B→PV decays,” 2003, arxiv:0308039 [hep-ph]. [Online]. Available: https://doi.org/10.1016/j.nuclphysb.2003.09.026

[43] W. Altmannshofer *et al.*, “Symmetries and asymmetries of $B \to K^{*}\mu^{+}\mu^{-}$ decays in the standard model and beyond,” 2009, arxiv:0811.1214 [hep-ph]. [Online]. Available: https://doi.org/10.1088/1126-6708/2009/01/019

[44] S. Descotes-Genon *et al.*, “Optimizing the basis of $B \to K^{*}ll$ observables in the full kinematic range,” 2013, arxiv:1303.5794 [hep-ph]. [Online]. Available: https://doi.org/10.1007/JHEP05(2013)137

[45] LHCb, “Angular analysis of the $b^{+} \to k^{*+}\mu^{+}\mu^{-}$ decay,” 2021. [Online]. Available: https://doi.org/10.1103/PhysRevLett.126.161802

[46] ATLAS, “Study of the rare decays of B^0_s and B^0 mesons into muon pairs using data collected during 2015 and 2016 with the ATLAS detector,” 2019, arxiv:1812.03017 [hep-ex]. [Online]. Available: https://doi.org/10.1007/JHEP04(2019)098

[47] CMS, “Measurement of properties of $\mathrm{B}^0_\mathrm{s} \to \mu^{+}\mu^{-}$ decays and search for $\mathrm{B}^0 \to \mu^{+}\mu^{-}$ with the CMS experiment,” 2020, arxiv:1910.12127 [hep-ex]. [Online]. Available: https://doi.org/10.1007/JHEP04(2020)188

[48] B. Capdevila *et al.*, “Patterns of new physics in $b \to s\ell^{+}\ell^{-}$ transitions in the light of recent data,” 2018, arxiv:1704.05340 [hep-ph]. [Online]. Available: https://doi.org/10.1007/JHEP01(2018)093

[49] A. Angelescu *et al.*, “On the single leptoquark solutions to the B-physics anomalies,” 2021, arxiv:2103.12504 [hep-ph]. [Online]. Available: https://doi.org/10.1103/PhysRevD.104.055017

[50] CMS, “Leptoquark searches in CMS,” in *53rd Rencontres de Moriond on Electroweak Interactions and Unified Theories*, 2018, arxiv:1901.03570 [hep-ex].

[51] ATLAS, “Searching for leptoquarks with the ATLAS detector,” in *17th Conference on Flavor Physics and CP Violation*, 2019, arxiv:1906.08983 [hep-ex].

[52] W. Altmannshofer, J. Davighi, and M. Nardecchia, “Gauging the accidental symmetries of the standard model, and implications for the flavor anomalies,” 2020, arxiv:1909.02021 [hep-ph]. [Online]. Available: https://doi.org/10.1103/PhysRevD.101.015004

[53] J. Davighi, “Anomalous Z' bosons for anomalous B decays,” 2021, arxiv:2105.06918 [hep-ph]. [Online]. Available: https://doi.org/10.1007/JHEP08(2021)101

[54] CMS, “Search for an $L_\mu - L_\tau$ gauge boson using Z$\rightarrow 4\mu$ events in proton-proton collisions at $\sqrt{s} = 13$ TeV,” 2019, arxiv:1808.03684 [hep-ex]. [Online]. Available: https://doi.org/10.1016/j.physletb.2019.01.072

[55] CMS, “The CMS phase-1 pixel detector upgrade,” 2021, arxiv:2012.14304 [physics.ins-det]. [Online]. Available: https://doi.org/10.1088/1748-0221/16/02/P02027

[56] CMS, “The CMS trigger in Run 2,” 2017. [Online]. Available: https://doi.org/10.22323/1.314.0523

[57] CMS, “The CMS experiment at the CERN LHC,” 2008. [Online]. Available: https://doi.org/10.1088/1748-0221/3/08/S08004

[58] V. Khachatryan *et al.*, “CMS tracking performance results from early LHC operation,” 2010, arxiv:1007.1988 [physics.ins-det]. [Online]. Available: https://doi.org/10.1140/epjc/s10052-010-1491-3

[59] CMS, “Description and performance of track and primary-vertex reconstruction with the CMS tracker,” 2014, arxiv:1405.6569 [physics.ins-det]. [Online]. Available: https://doi.org/10.1088/1748-0221/9/10/P10009

[60] V. Khachatryan *et al.*, “Tracking POG results for pion efficiency with the D^* meson using data from 2016 and 2017,” 2018, cMS-DP-2018-050. [Online]. Available: http://cds.cern.ch/record/2634940/files/DP2018_050.pdf?version=1

[61] CMS, “The CMS trigger system,” 2017, arxiv:1609.02366 [physics.ins-det]. [Online]. Available: https://doi.org/10.1088/1748-0221/12/01/P01020

[62] T. Sjöstrand *et al.*, “An introduction to PYTHIA 8.2,” 2015, arxiv:1410.3012 [hep-ph]. [Online]. Available: https://doi.org/10.1016/j.cpc.2015.01.024

[63] D. J. Lange, “The EvtGen particle decay simulation package,” *Nucl. Instrum. Meth. A*, vol. 462, pp. 152–155, 2001. [Online]. Available: https://doi.org/10.1016/S0168-9002(01)00089-4

[64] N. Davidson, T. Przedzinski, and Z. Was, “PHOTOS interface in C++: technical and physics documentation,” 2016, arxiv:1011.0937 [hep-ph]. [Online]. Available: https://doi.org/10.1016/j.cpc.2015.09.013

[65] S. Agostinelli *et al.*, “GEANT4–a simulation toolkit,” *Nucl. Instrum. Meth. A*, vol. 506, pp. 250–303, 2003. [Online]. Available: https://doi.org/10.1016/S0168-9002(03)01368-8

[66] A. Modak, “CMS phase-1 pixel detector: operational experience, performance and lessons learned,” *PoS*, vol. Vertex2019, p. 009, 2020. [Online]. Available: https://doi.org/10.22323/1.373.0009

[67] G. Punzi, “Sensitivity of searches for new signals and its optimization,” *eConf*, vol. C030908, p. MODT002, 2003, arxiv:physics/0308063.

[68] F. James, “MINUIT function minimization and error analysis: reference manual version 94.1,” 1994, CERN-D-506, CERN-D506.

[69] C. J. Clopper and E. S. Pearson, “The Use of confidence or fiducial limits illustrated in the case of the binomial.” *Biometrika*, vol. 26, no. 4, p. 404–413, 1994.

[70] M. Pivk and F. R. Le Diberder, “SPlot: a statistical tool to unfold data distributions,” 2005, arxiv:physics/0402083. [Online]. Available: https://doi.org/10.1016/j.nima.2005.08.106

[71] CMS, “CMS statistics committee recommendations,” accessed on October 2nd 2021. [Online]. Available: https://twiki.cern.ch/twiki/bin/viewauth/CMS/StatComCombination

Appendix A

Nominal data fit results (2016 and 2017)

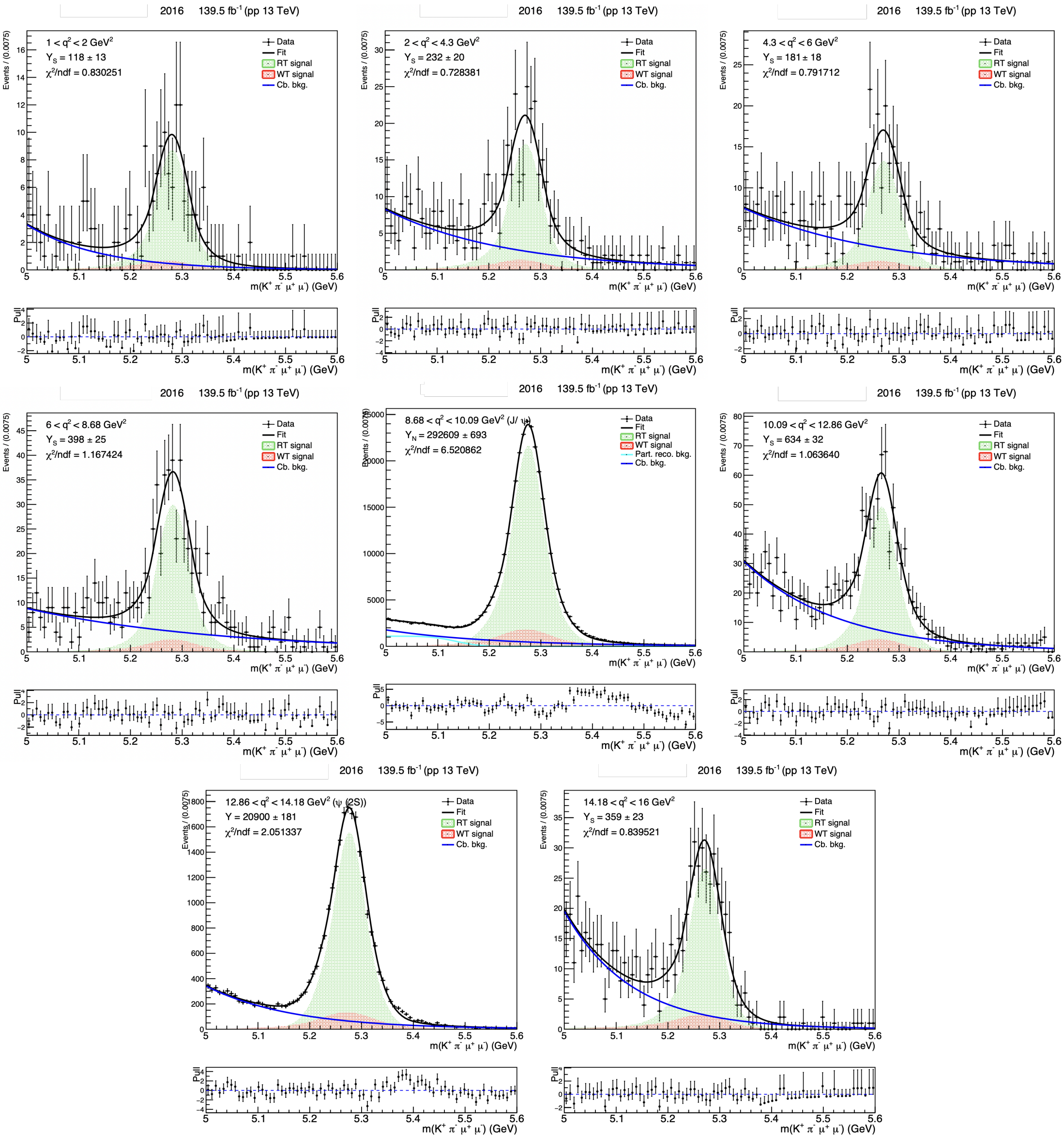


Figure A.1: Data fit results in year 2016. The q^2 ranges covered by each q^2 bin can be seen in the top left region of each plot. The RT and WT components are shown in green and red, respectively. The combinatorial background is shown in blue and the partially reconstructed background in cyan.

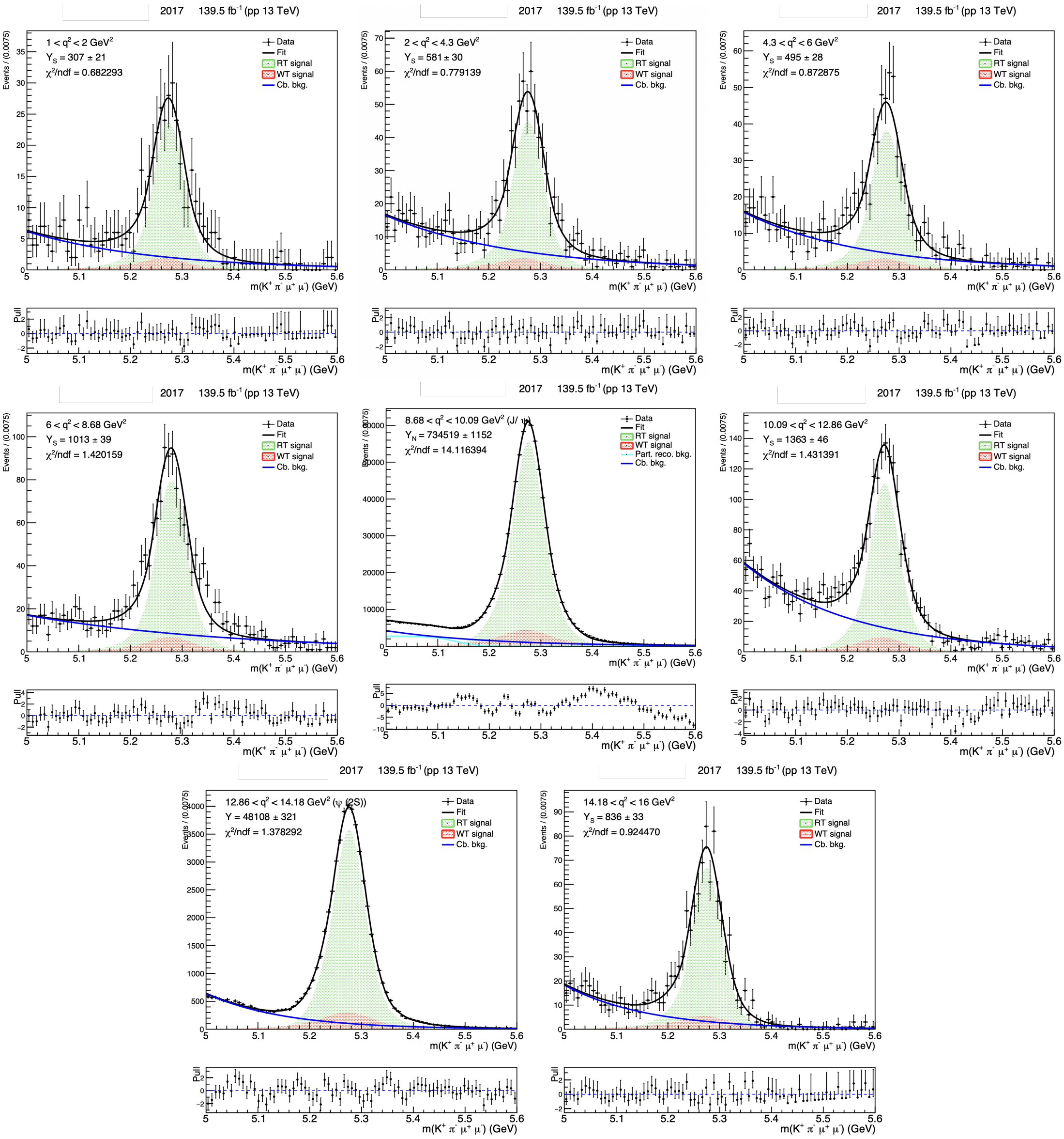


Figure A.2: Data fit results in year 2017. The q^2 ranges covered by each q^2 bin can be seen in the top left region of each plot. The RT and WT components are shown in green and red, respectively. The combinatorial background is shown in blue and the partially reconstructed background in cyan.

Appendix B

Signal yield pull distributions (2016 and 2017)

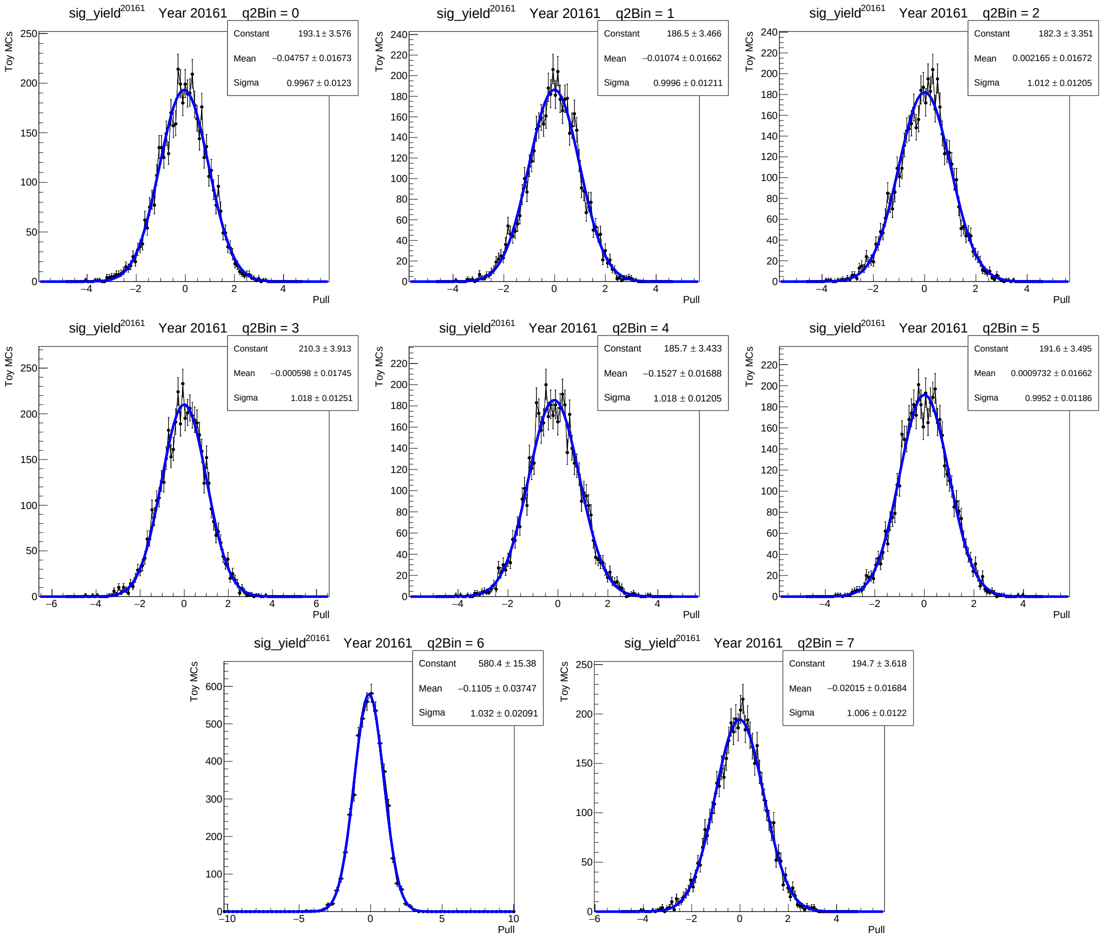


Figure B.1: Signal yield pull distributions resulting from the toy MC study in year 2016. The mean and width values of the Gaussian distributions are shown in the top right panel.

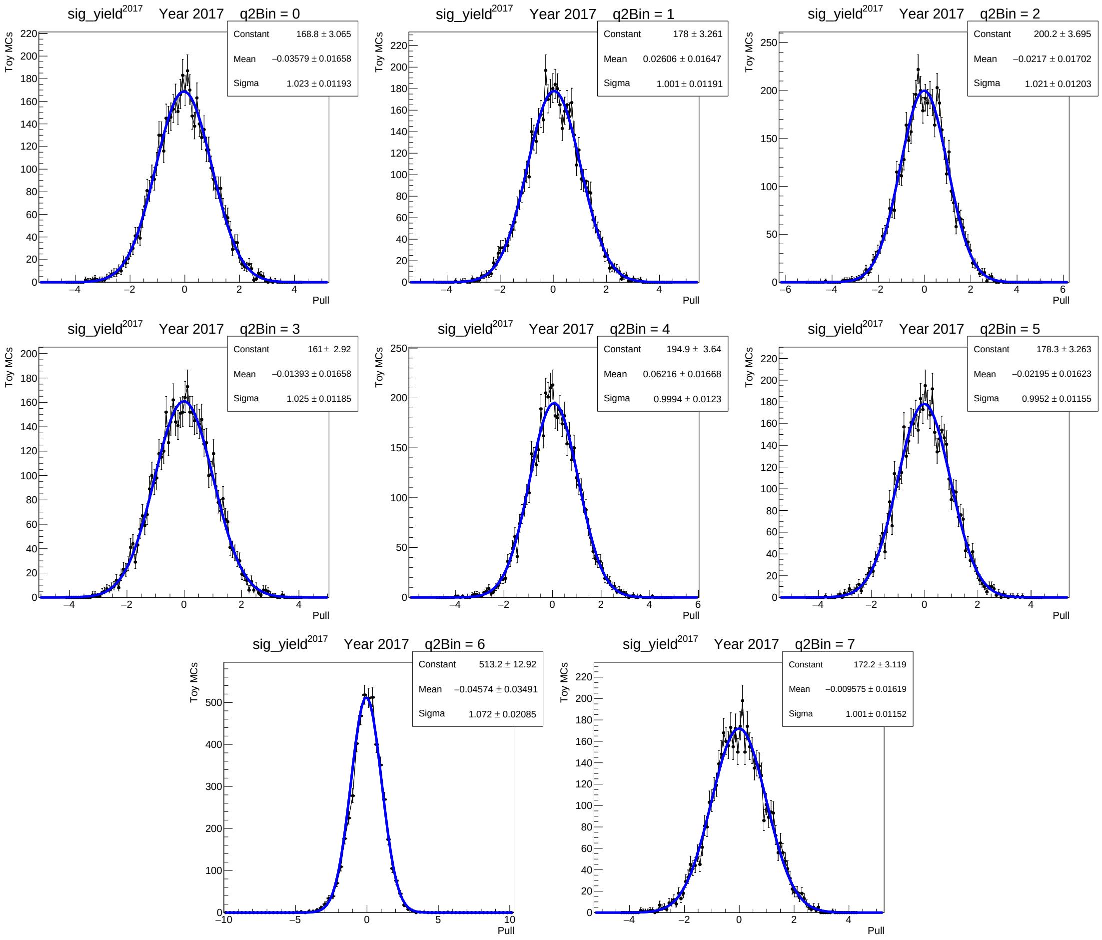


Figure B.2: Signal yield pull distributions resulting from the toy MC study in year 2017. The mean and width values of the Gaussian distributions are shown in the top right panel.

Appendix C

Detector efficiency (2017) and angular parameters

The results of the MC based efficiency computation presented in Sec. 6.2 and the averaged efficiency computation presented in Sec. 6.3 for year 2017 are presented in Fig. C.1 (top left) and (top right) plots, respectively. The comparison between the two methods is also presented in Fig. C.1 (bottom).

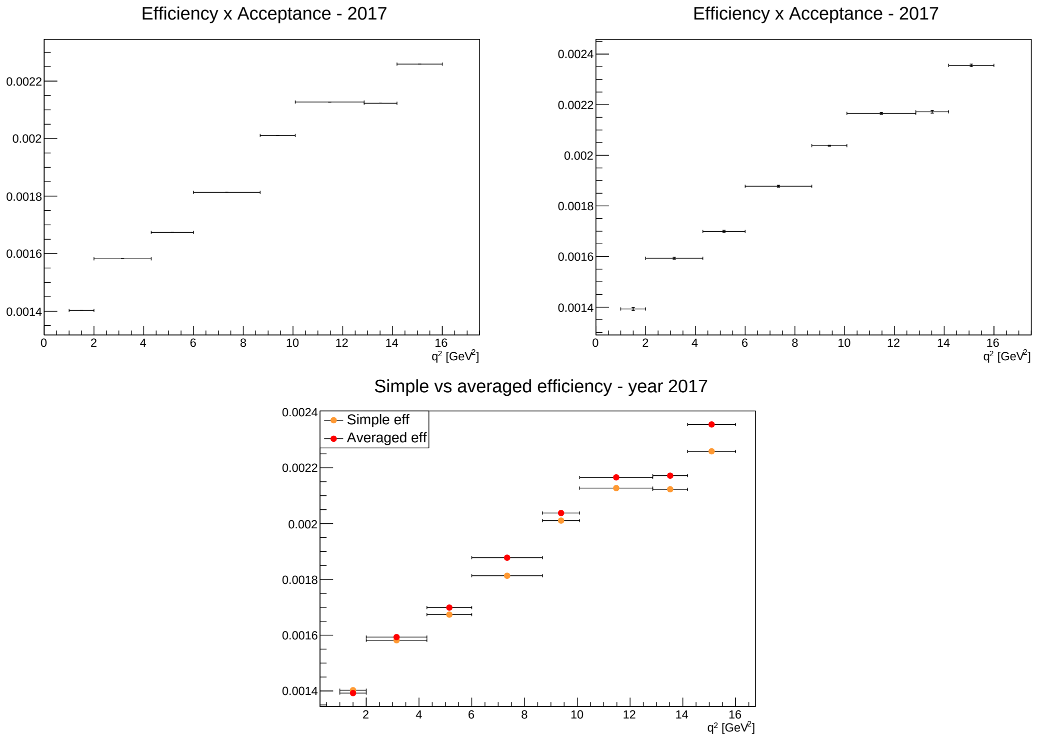


Figure C.1: Results of the MC based efficiency computation (top left) and the averaged efficiency computation (top right) for year 2017 and comparison between the two methods (bottom).

The angular parameters obtained in Ref. [13] and used in the computation of the averaged efficiencies in Sec. 6.3 can be found in Tab. for MC and data, respectively.

Table C.1: Angular parameters appearing in the $B^0 \to K^{*0}\mu^+\mu^-$ decay rate in Eq. (2.31) resulting from the fit to the MC in Ref. [13].

q^2 bin	F_L	P_1	P_2	P_3
0	0.714 ± 0.0005	0.004 ± 0.004	-0.388 ± 0.0008	0.00242 ± 0.00214
1	0.813 ± 0.0003	-0.052 ± 0.004	-0.258 ± 0.0009	0.00088 ± 0.00221
2	0.740 ± 0.0004	-0.113 ± 0.004	0.205 ± 0.0009	0.00325 ± 0.00196
3	0.618 ± 0.0003	-0.147 ± 0.002	0.418 ± 0.0004	0.00198 ± 0.00098
4	0.6000 ± 0.0005	-0.198 ± 0.0031	-0.0003 ± 0.0009	0.0004 ± 0.0025
5	0.456 ± 0.0003	-0.236 ± 0.001	0.461 ± 0.0002	0.00115 ± 0.00066
6	0.6000 ± 0.0021	-0.199 ± 0.0126	-0.0004 ± 0.0034	0.0007 ± 0.0099
7	0.370 ± 0.0004	-0.402 ± 0.001	0.445 ± 0.0003	0.00080 ± 0.00077

q^2 bin	P'_4	P'_5	P'_6	P'_8
0	0.163 ± 0.003	0.345 ± 0.002	-0.00237 ± 0.00181	0.00226 ± 0.00343
1	-0.522 ± 0.003	-0.310 ± 0.001	-0.00095 ± 0.00151	0.00051 ± 0.00289
2	-0.917 ± 0.003	-0.723 ± 0.001	-0.00074 ± 0.00153	-0.00032 ± 0.00298
3	-1.029 ± 0.001	-0.861 ± 0.001	0.00002 ± 0.00100	-0.00143 ± 0.00191
4	-0.870 ± 0.0033	0.0012 ± 0.0019	-0.0003 ± 0.0016	0.0004 ± 0.0040
5	-1.098 ± 0.001	-0.852 ± 0.001	-0.00091 ± 0.00089	0.00037 ± 0.00167
6	-0.816 ± 0.0123	0.0005 ± 0.0078	0.0000 ± 0.0065	0.0002 ± 0.0158
7	-1.178 ± 0.001	-0.761 ± 0.001	-0.00114 ± 0.00118	0.00324 ± 0.00226

Table C.2: Angular parameters appearing in the $B^0 \to K^{*0}\mu^+\mu^-$ decay rate in Eq. (2.31) resulting from the fit to data. The values for the resonant channels in q^2 bins 4 and 6 are taken from Ref. [13], whereas the parameters in the signal q^2 bins are taken from Ref. [3].

q^2 bin	F_L	P_1	P_2	P_3
0	0.655 ± 0.046	-0.617 ± 0.296	-0.443 ± 0.100	0.324 ± 0.147
1	0.756 ± 0.047	0.168 ± 0.371	-0.191 ± 0.116	0.049 ± 0.195
2	0.684 ± 0.035	0.088 ± 0.235	0.105 ± 0.068	-0.090 ± 0.139
3	0.645 ± 0.030	-0.071 ± 0.211	0.207 ± 0.048	-0.068 ± 0.104
4	0.55501 ± 0.000624	-0.015484 ± 0.00312	-0.0012001 ± 0.000958	0.23983 ± 0.00255
5	0.461 ± 0.031	-0.460 ± 0.132	0.411 ± 0.033	-0.078 ± 0.077
6	0.52136 ± 0.00218	-0.039087 ± 0.0126	0.0059624 ± 0.00338	0.46670 ± 0.0103
7	0.352 ± 0.026	-0.511 ± 0.096	0.396 ± 0.022	0.000 ± 0.056

q^2 bin	P'_4	P'_5	P'_6	P'_8
0	-0.080 ± 0.142	0.365 ± 0.122	-0.226 ± 0.128	-0.366 ± 0.158
1	-0.435 ± 0.169	-0.150 ± 0.144	-0.155 ± 0.148	0.037 ± 0.169
2	-0.312 ± 0.115	-0.439 ± 00.111	-0.293 ± 0.117	0.166 ± 0.127
3	-0.574 ± 0.091	-0.583 ± 0.090	-0.155 ± 0.098	-0.129 ± 0.098
4	-0.95271 ± 0.00329	-0.0063346 ± 0.00193	0.0018248 ± 0.00157	-0.21960 ± 0.00398
5	-0.491 ± 0.095	-0.622 ± 0.088	-0.193 ± 0.100	0.018 ± 0.099
6	-0.77781 ± 0.0123	0.0022117 ± 0.00778	0.010996 ± 0.00645	-0.35748 ± 0.0159
7	-0.626 ± 0.069	-0.714 ± 0.074	0.061 ± 0.085	0.007 ± 0.086

Appendix D

Monte Carlo validation (2016 and 2017)

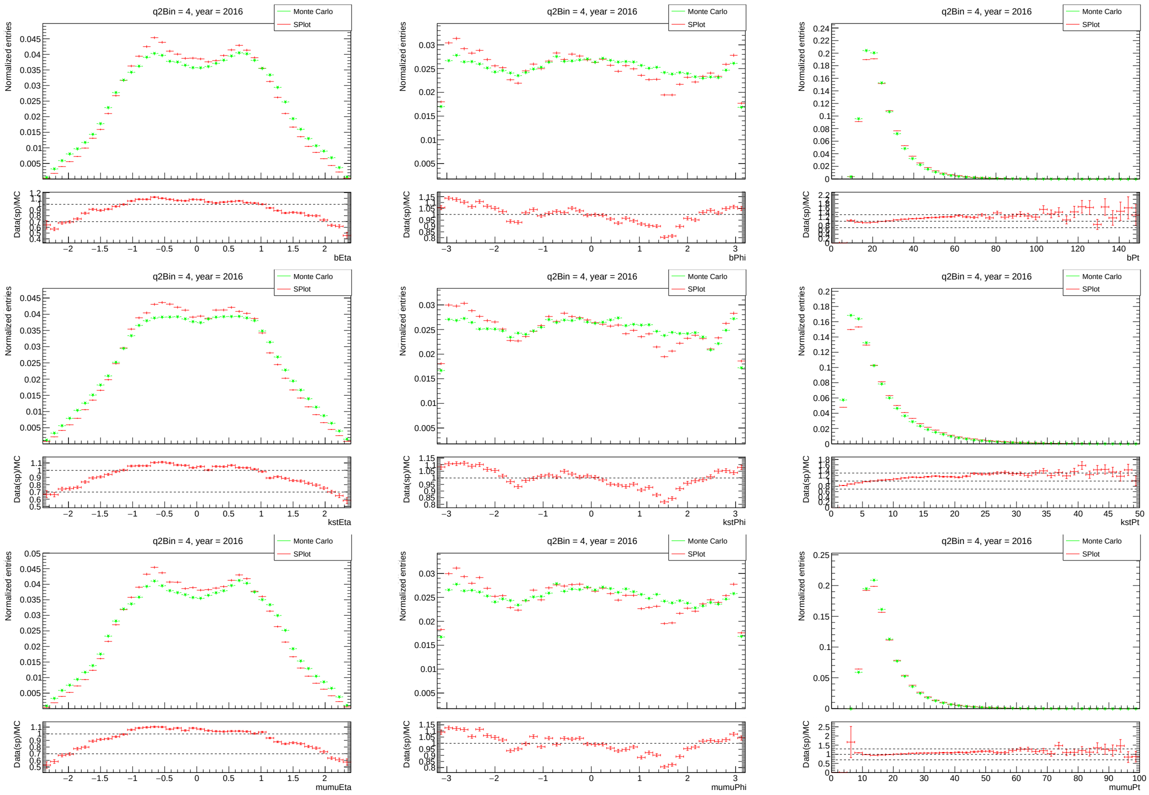


Figure D.1: Comparison between data signal distribution (in red) obtained with he $_s\mathcal{P}lot$ method and the MC (in green), for different control variables in year 2016 for q^2 bin 4.

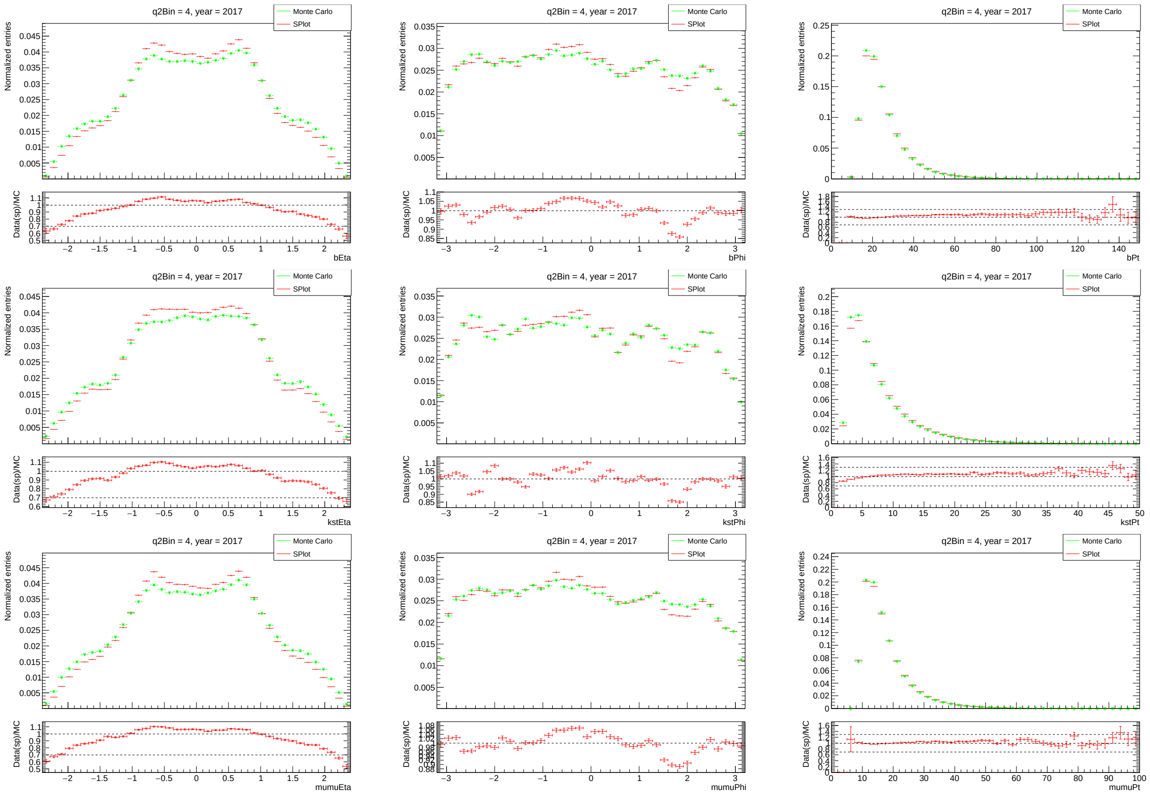


Figure D.2: Comparison between data signal distribution (in red) obtained with he $_s\mathcal{P}lot$ method and the MC (in green), for different control variables in year 2017 for q^2 bin 4.

Appendix E

Fit variations

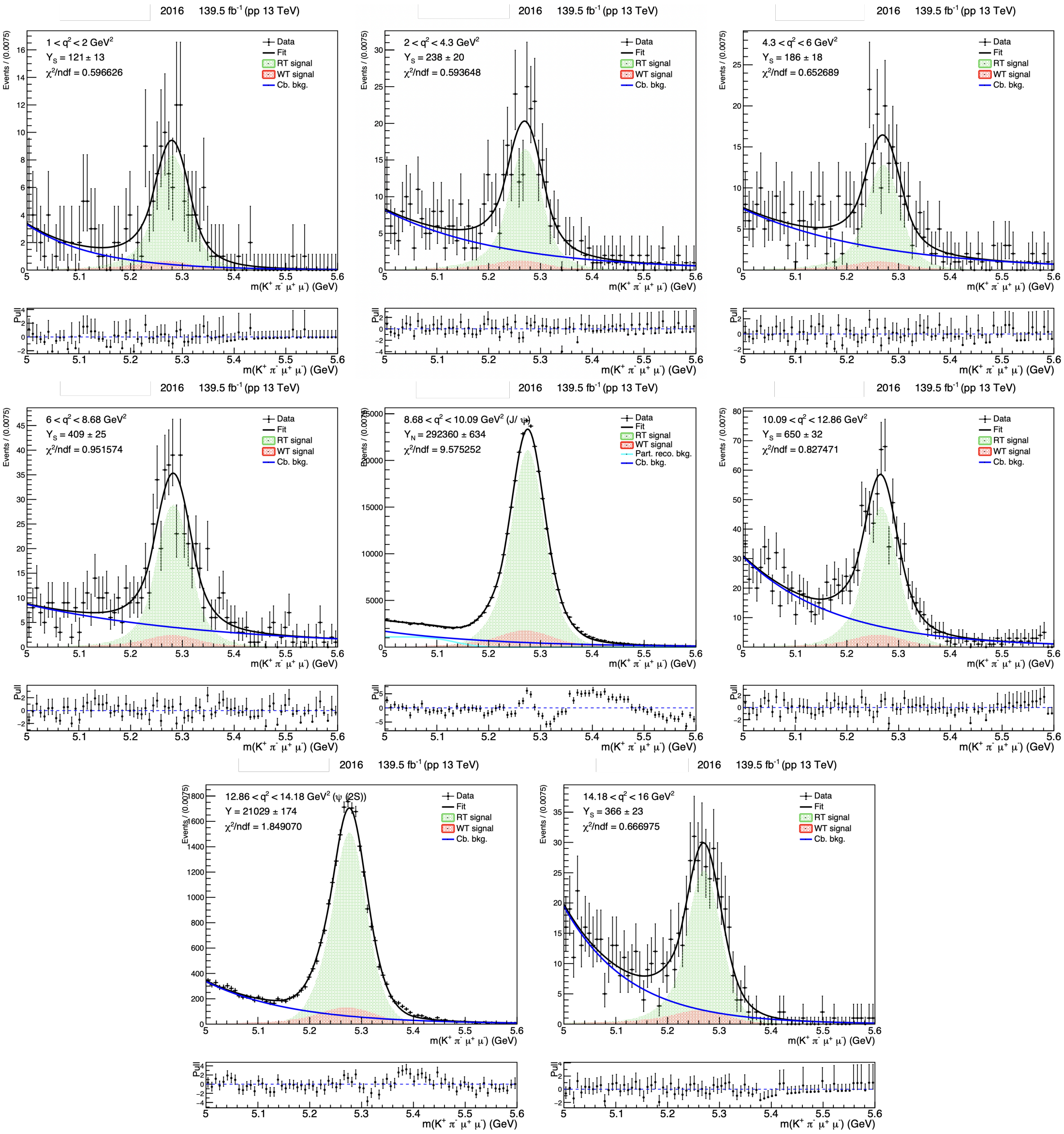


Figure E.1: Scale factor fit results in year 2016 for all q^2 bins.

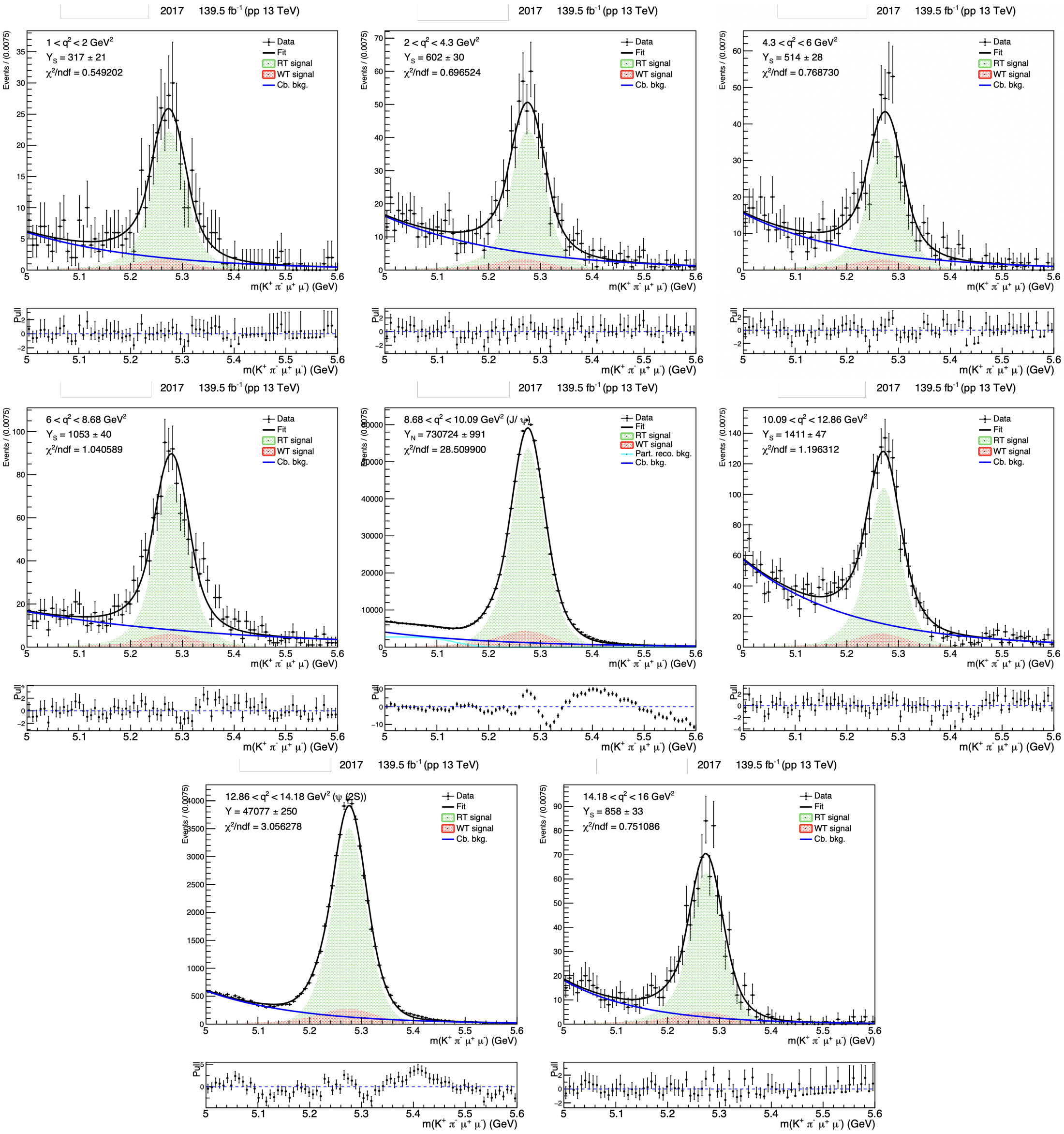


Figure E.2: Scale factor fit results in year 2017 for all q^2 bins.

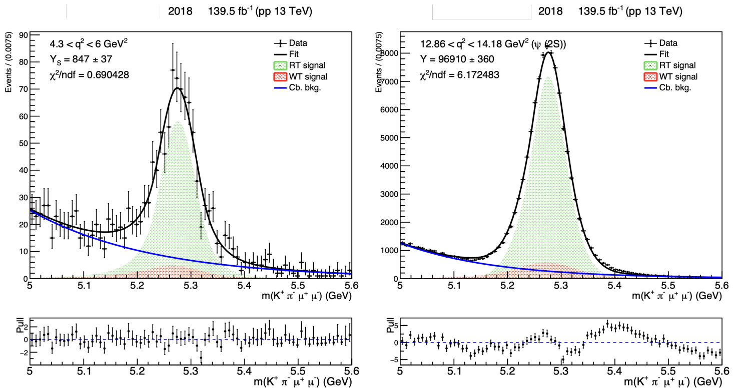


Figure E.3: Scale factor fit results in year 2018 for q^2 bins 2 (left) and 6 (right).

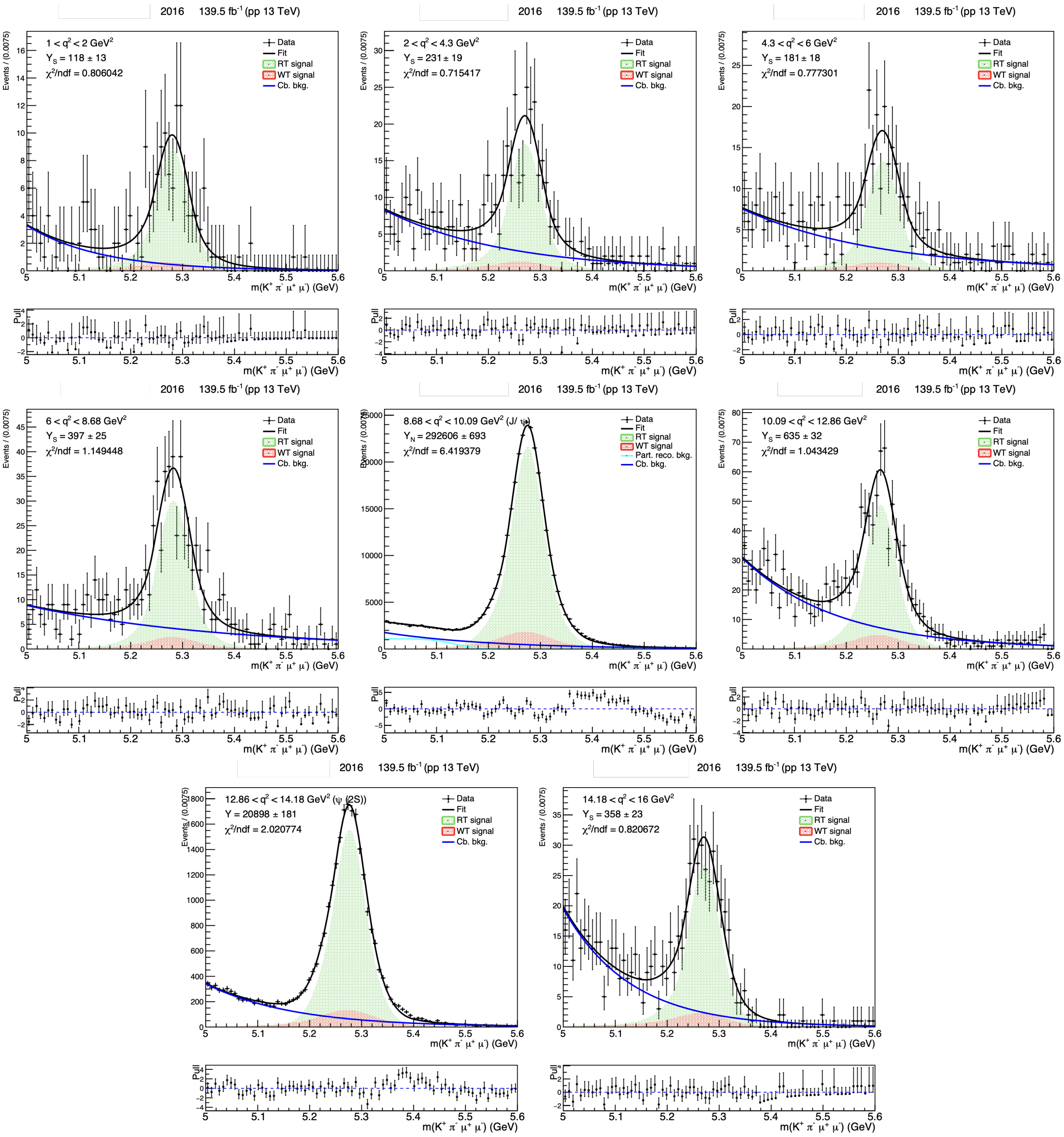


Figure E.4: Mistag fraction fit results in year 2016 for all q^2 bins.

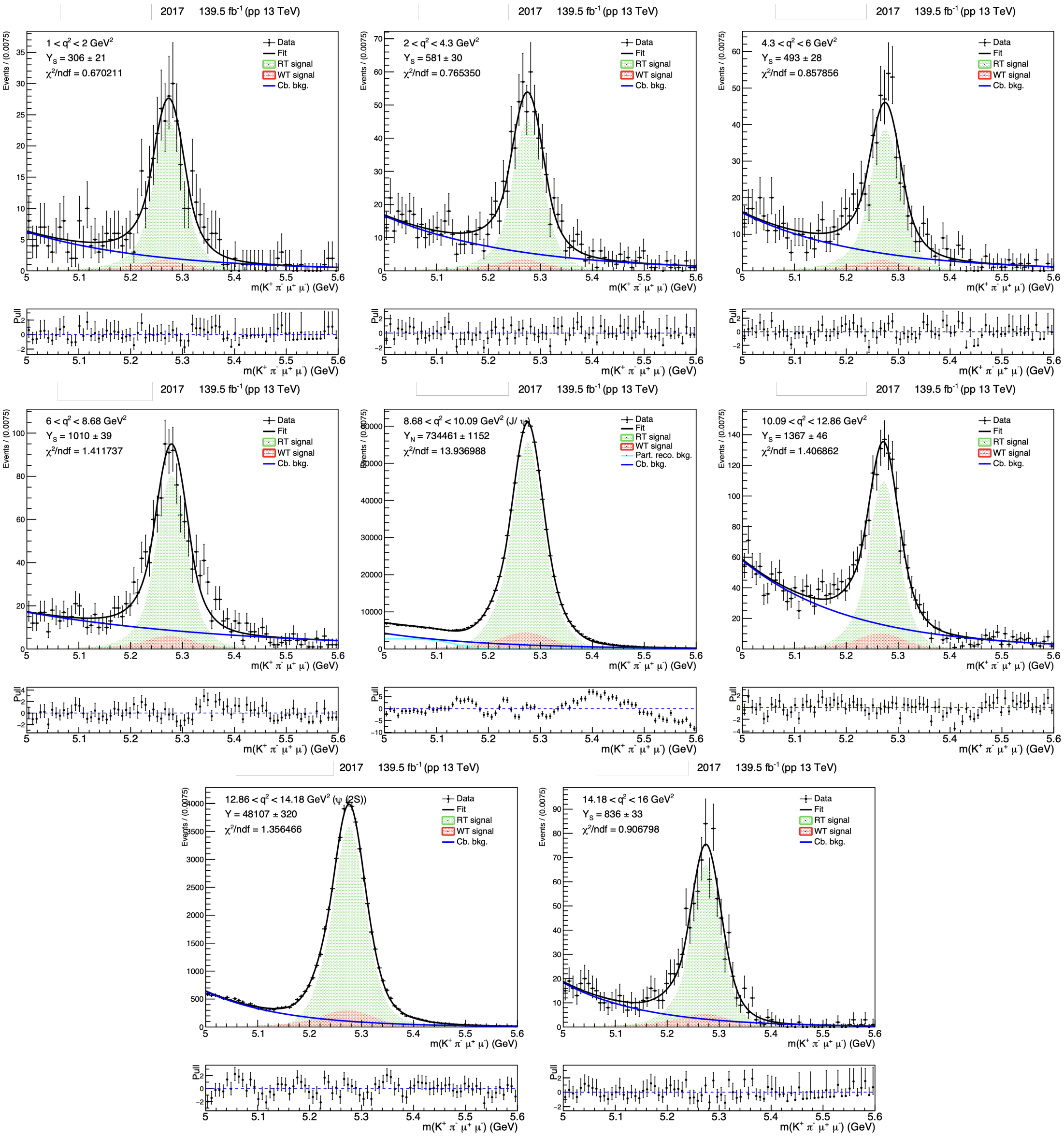


Figure E.5: Mistag fraction fit results in year 2017 for all q^2 bins.

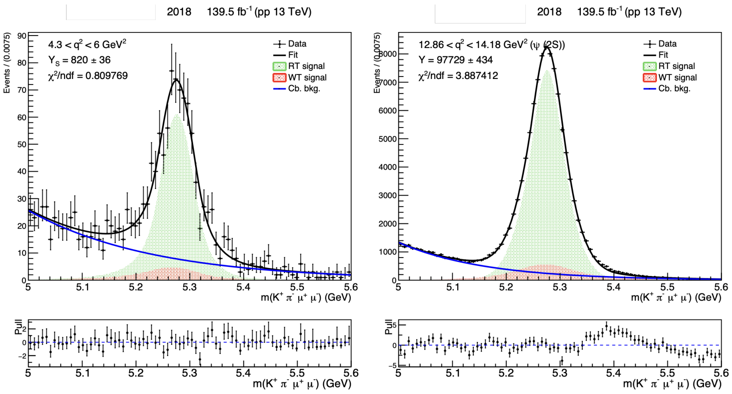


Figure E.6: Mistag fraction fit results in year 2018 for q^2 bins 2 (left) and 6 (right).

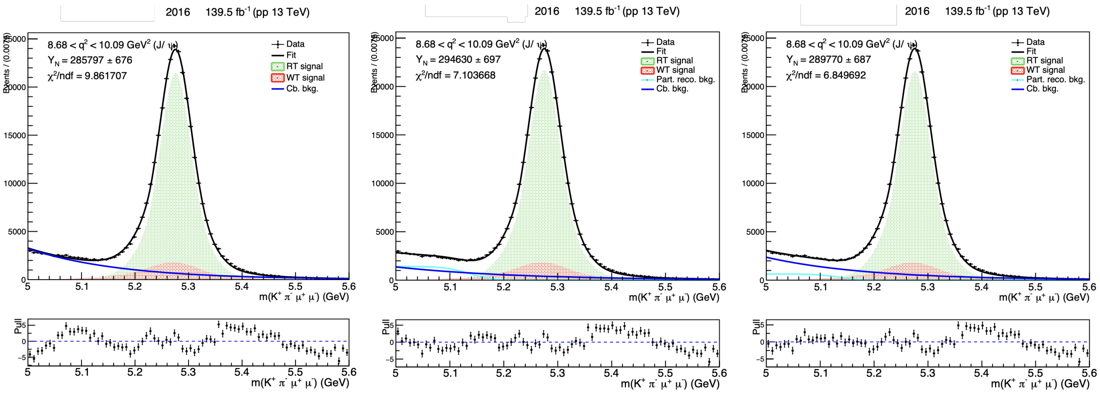


Figure E.7: No erf (left), largest f_{erf} (centre) and lowest f_{erf} (right) fit results in year 2016.

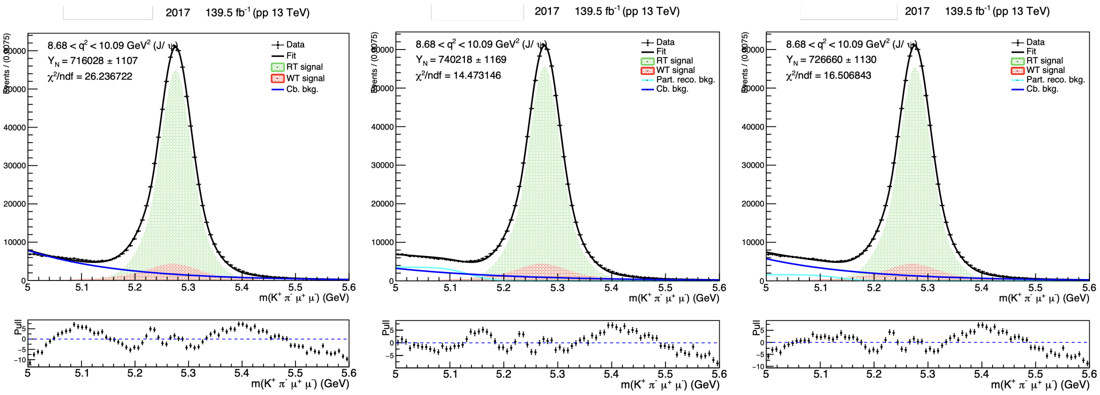


Figure E.8: No erf (left), largest f_{erf} (centre) and lowest f_{erf} (right) fit results in year 2017.