Department of Physics and Astronomy
University of Heidelberg

Master thesis

in Physics

submitted by

Miguel Ruiz Díaz

born in Toledo, Spain

2022

Study of $B^0 \to K^{*0}e^+e^-$ decays for a new test of lepton flavor universality in the high di-lepton invariant mass region at the LHCb experiment

This Master thesis has been carried out by Miguel Ruiz Díaz

at the

Physikalisches Institut

under the supervision of

Prof. Dr. Stephanie Hansmann-Menzemer

and

Dr. Martino Borsato

Abstract:

Lepton flavor universality tests using rare B meson decays are amongst the most sensitive probes of the Standard Model's flavor structure. LHCb has made significant contributions to the field in recent years and it is aiming at a new measurement of the ratio between the branching fractions of B^0 meson decays to muons $B^0 \to K^{*0}\mu^+\mu^-$ and electrons $B^0 \to K^{*0}e^+e^-$ in the high di-lepton invariant mass region, also known as high-q^2 region. The experimental uncertainties on the measurement are dominated by the poorer mass resolution in the electron channel, which suffers from a much larger background contribution and a smaller signal efficiency. Constraining these backgrounds constitutes one of the main challenges of the analysis and the main goal of this thesis.

Data recorded at the LHCb experiment from 2011 to 2018 is used together with Monte Carlo simulated data to build a model that describes the shape of the $B^0 \to K^{0*}ee$ reconstructed invariant B^0 mass distribution. This model will be used to perform the mass fit from which the signal yield will be extracted. A data driven method is developed to describe the shape of the combinatorial background mass distribution. This method is validated in real data using the region of the q^2 spectrum dominated by the contribution of the $\psi(2S)$ resonance. The yields of background components leaking into the high-q^2 region from the $\psi(2S)$ region are also constrained from mass fits performed in the $\psi(2S)$ region. The strategy is validated in several cross-checks and disagreements on the $O(10\%)$ level are found amongst them. Possible improvements in the analysis to remove these disagreements are discussed. The strategy to select the high-q^2 region is evaluated and the expected amount of over-reconstructed background coming from $B^+ \to K^+e^+e^-$ decays is estimated and found to be on the per-cent level.

Zussamenfassung:

Lepton-Flavor-Universalitätstests mit seltenen B-Meson-Zerfällen gehören zu den empfindlichsten Methoden, die Flavor-Struktur des Standardmodells zu untersuchen. LHCb hat in den letzten Jahren bedeutende Beiträge zu diesem Bereich geleistet und zielt auf eine neue Messung des Verhältnisses zwischen den Verzweigungsanteilen der B^0-Meson-Zerfälle zu Myonen $B^0 \to K^{*0}\mu^+\mu^-$ und Elektronen $B^0 \to K^{*0}e^+e^-$ in der Region mit hoher invarianter Di-Lepton-Masse, auch bekannt als Region mit hohem q^2. Die experimentellen Unsicherheiten bei der Messung werden von der schlechteren Massenauflösung im Elektronenkanal dominiert, der unter einem viel größeren Hintergrundbeitrag und einer geringeren Signaleffizienz leidet. Die Eingrenzung dieser Hintergründe ist eine der größten Herausforderungen bei der Analyse und das Hauptziel dieser Arbeit.

Die am LHCb-Experiment von 2011 bis 2018 aufgezeichneten Daten werden zusammen mit Monte-Carlo-simulierten Daten verwendet, um ein Modell zu erstellen, das die Form der $B^0 \to K^{0*}ee$ rekonstruierten invarianten B^0-Massenverteilung beschreibt. Dieses Modell wird verwendet, um die Massenanpassung durchzuführen, aus der die Anzahl der Signalkandidaten extrahiert wird. Es wird eine datengesteuerte Methode entwickelt, um die Form der kombinatorischen Hintergrundmassenverteilung zu beschreiben. Diese Methode wird an realen Daten validiert, wobei der Bereich des q^2-Spektrums verwendet wird, der durch den Beitrag der $\psi(2S)$-Resonanz dominiert wird. Die Anzahl der Hintergrundkandidaten, die aus der $\psi(2S)$-Region in die hohe q^2-Region abwandern, wird ebenfalls durch Massenanpassungen in der $\psi(2S)$-Region eingeschränkt. Die Strategie wird in mehreren Quervergleichen validiert und es werden Unstimmigkeiten auf dem $O(10\%)$-Niveau gefunden. Mögliche Verbesserungen in der Analyse zur Beseitigung dieser Unstimmigkeiten werden diskutiert. Die Strategie zur Auswahl der hohen q^2-Region wird evaluiert und die erwartete Menge an überrekonstruiertem Hintergrund aus $B^+ \to K^+e^+e^-$-Zerfällen wird abgeschätzt und liegt im Prozentbereich.

Erklärung:

Ich versichere, dass ich diese Arbeit selbstständig verfasst habe und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Heidelberg, den 15.11.2022 Miguel

Contents

Chapter 1

Introduction

The Standard Model of particle physics (SM) was completed with the discovery of the Higgs boson in 2012 by CMS [2] and ATLAS [1]. It is considered one of the most successful theories in physics having made some of the most accurate predictions so far. However, despite its great achievements, there is an ever growing amount of physical phenomena that the SM is not able to explain. How does gravity behave inside a black hole? What is the origin of neutrino masses? What is the nature of dark matter and dark energy? Why is there so much matter compared to antimatter? These are all questions the SM can not answer. Addressing these questions will probably require the extension of the SM by including new mechanisms or fields with masses at energy scales way beyond the scope of current or even foreseeable high energy particle colliders. Thankfully, the formalism of QFT gives us some hope by introducing the concept of *virtual particles*. They appear as internal propagators in Feynman diagrams and can not be directly seen in experiments. If a massive particle contributes to a "low energy" process as a virtual particle it could generate measurable effects. Thus, the existence of new massive particles could be indirectly inferred by making very precise measurements of processes which are allowed in the SM and comparing the results with theory predictions. These indirect searches of new physics phenomena at low energy processes become more sensitive the more precise the SM predictions for the processes are. This usually involves looking at processes which are largely suppressed in the SM since small deviations caused by new physics phenomena might result in visible contributions. Decays of hadrons containing b quarks provide an excellent tool to look for physics beyond the SM since they are heavy enough to display a large amount of decay channels and are copiously produced in pp collisions and detected at the LHCb experiment.

This thesis is focused on *rare* B meson decays, which involve $b \to sll$ transitions at quark level. A transition from a b quark to a strange quark s is only possible in the SM at Next to Leading Order (NLO) in processes involving internal Electro-Weak (EW) loops. They are called Flavor Changing Neutral Current (FCNC) transitions since they change the quark flavor but not the electric charge of the quarks involved (both b and s quarks have an electric charge of $-\frac{1}{3}q_e$, where q_e is the electron charge). FCNCs provide an excellent tool to test the SM structure since they are highly suppressed, presenting branching fractions below 10^{-6} [4], and are very sensible to contributions form new heavy particles or new interactions with very small couplings. Rare B decays have been extensively studied in the past using different observables that test different aspects of the SM and a pattern of deviations with respect to the SM predictions has steadily started to emerge. These deviations, together with some others found in D-meson and Λ_b-baryon decays have been collectively named *flavor anomalies* [9,10]. They seem to follow a consistent pattern and have started to get the attention of the whole particle physics community. Many of these anomalies appear in $b \to sll$ transitions. Particularly interesting are tests where the decay ratios to leptons of different flavor are compared since they provide a direct measurement of the relative ratio of their couplings. In the SM, couplings involving different lepton generations have all the same size. This property of the SM is known as Lepton Flavor Universality (LFU) and has been extensively tested in a large number of experiments [5-8]. It could however be violated in many New Physics (NP) scenarios and recent LHCb results have found hints of a possible violation in rare B meson decays [11,12]. This thesis is focused on the $R_{K^{*0}}$ observable, defined as:

$$R_{K^{0*}} = \frac{\int_{q^2_{bin}} \frac{d\Gamma(B^0 \to K^{0*}\mu\mu)}{dq^2} dq^2}{\int_{q^2_{bin}} \frac{d\Gamma(B^0 \to K^{0*}ee)}{dq^2} dq^2} \quad \text{with} \quad q^2 \equiv m(ll)^2. \tag{1.1}$$

In the SM, $R_{K^{*0}} \simeq 1$ and the only distinction between the electron and the muon channels comes form small phase space differences due to small differences between their masses. This ratio was first measured at LHCb in 2017 and deviations of 2.1-2.5σ were found in two different q^2 intervals [11]. In this thesis we focus on the high-q^2 region, unexplored by LHCb so far. The electron channel process $B^0 \to K^{*0}e^+e^-$ is studied in detail as it contains the main sources of experimental uncertainties for the determination of the $R_{K^{*0}}$ ratio and it is therefore very important to understand it properly. The main challenges arising in the high-q^2 region are related to the large background rates which are especially problematic due to the small signal efficiency, and the distortion of the B^0 mass shape, affected by the cut on the q^2-value. The main goal of this thesis is to develop a strategy to control the background shapes and yields before moving on to the $R_{K^{*0}}$ measurement.

In this thesis, an analysis of the main backgrounds populating the high-q^2 region in the $B^0 \to K^{*0}e^+e^-$ process is presented. In chapter 2, the theoretical background needed to understand the measurement is introduced. The main ingredients of the SM are described, with a focus on the features relevant to understand the physics of FCNC processes. The LHCb experiment, its sub-detectors, and its trigger system are introduced in chapter 3. After the general introduction, the analysis strategy is explained in detail in chapter 4. The main experimental challenges are discussed and the general strategy to tackle them is layered out. Data and simulation samples used in this analysis are introduced in chapter 5. The full set of selection cuts is explained, as well as the truth-matching strategy applied to simulated samples. Chapter 6 describes the techniques applied in this thesis to suppress known background sources. Results are presented in chapter 6 and chapter 7 contains the conclusions together with an outlook for the next steps and future prospects.

Chapter 2

Lepton Flavor Universality in the Standard Model

In this chapter the basic theoretical background required to understand the physics behind the processes analysed in this thesis is described. In section 2.1 a short summary of the elementary ingredients that lead to the Standard Model of particle physics (SM) as we know it today is presented. The Electroweak (EW) model, central to understand the physics of $b \to sl^+l^-$ transitions, is explained in detail in section 2.2. Finally in section 2.3 the Effective Field Theory (EFT) formalism is introduced. We will explain how it can be applied to search for new physics in $b \to sl^+l^-$ decays.

2.1 The Standard Model of particle physics

The SM of particle physics is a Quantum Field Theory (QFT) which was developed during the second half of the 20th century to describe the physics of strong and EW interactions. Its current form was completed during the 70s and since then it has made a large number of predictions, experimentally confirmed with unprecedented precision. One of the basic building blocks of the SM is the concept of *gauge theory* and *gauge interactions*. In a gauge theory, interactions between particles are described as the consequence of the exchange of massless vector bosons, acting as "force carriers". Interactions are usually depicted with the use of *Feynman diagrams*, introduced by Richard Feynman in the 1940s as a way to visualize the mathematical expressions describing the processes of interaction between particles. Particles are depicted as lines that meet each other in interaction vertices where they exchange momentum.

The use of Feynman diagrams highlights one of the most important tools used in model particle physics: *perturbation theory*. Each of the interaction vertices in a Feynman diagram carries a factor of the relevant coupling. Perturbation theory can be applied in processes where the couplings -which measure the strength of the interaction- are smaller than one, therefore, the amount of vertices in a Feynman diagram indicates the order in perturbation theory at which the diagram contribution becomes relevant. If interactions are perturbative, Feynman diagrams with many vertices are suppressed with respect to simpler diagrams with less vertices and can be neglected. In the SM, electroweak interactions can be treated perturbatively at low energies[1], this is not the case of strong interactions, which only become perturbative at high energies and therefore are much harder to treat.

[1]Here, low energies refer to energies accessible in current particle physics experiments

The gauge structure of the SM follows from the concept of *gauge symmetry*, a symmetry under local transformations given by elements of the SM symmetry group:

$$SU(3)_c \times SU(2)_L \times U(1)_Y.$$

In this group, the $SU(3)_c$ part describes strong interactions and the sub index c refers to *color*, which is the name given to the conserved charge under local $SU(3)_c$ transformations. Particles coupling to the strong interaction are called *quarks*. They transform under the fundamental representation of the symmetry group and carry one of the three types of color charges. The gauge bosons mediating the interaction are called gluons and transform under the 8-dimensional adjoint representation. The name color was coined in analogy with the three elementary colors which neutralize each other when mixed, giving white. In the same way, three quarks with different color charges bound together by the strong interaction form a colorless structure (strongly neutral). The theory that describes strong interactions is accordingly called *Quantum Chromodynamics* (QCD) and has made a large number of predictions confirmed later in experiment and survived all the experimental test it has been subject to. QCD was the first example of a successful *non-abelian gauge theory*. This is because the $SU(3)$ group is a non abelian group, meaning that the group generators do not commute with each other. One of the physical consequences of this non-abelian nature is that gluons are charged under the strong interaction so they can interact with themselves. This feature is key to understand the properties of the strong interaction such as *asymptotic freedom* and *color confinement* which respectively happen at high and low energies[2]. They explain the existence and properties of hadrons.

The $SU(2)_L \times U(1)_Y$ part describes the EW part of the SM. It follows from a combination of two symmetry groups and therefore it has 4 vector bosons associated. From the $SU(2)_L$ part we get three bosons, transforming under adjoint representation of $SU(2)$, and from the $U(1)_Y$ part we get the remaining one. Linear combinations of these 4 vector bosons give rise to the three gauge bosons of the weak interaction, $W^{\pm}$ and Z^0, and the photon of Quantum Electrodynamics (QED) after its *Spontaneous Symmetry Breaking* (SSB). In this symmetry group, the sub-index L indicates that the $SU(2)_L$ vector bosons couple to left-handed chiral particles only, this is not true for the weak interactions -the low energy phase after symmetry breaking- as it is shown in section 2.2. The quantum number Y is the hyper-charge, and is the conserved charge under $U(1)_Y$ transformations. The $SU(2)_L$ subgroup also has a conserved charge called *weak isospin* in analogy to spin, which is also the conserved quantity under the $SU(2)$ group that describes spatial rotations. Particles coupling to EW interactions form left handed chiral electro-weak doublets when they couple to $SU(2)_L$ and right-handed chiral electro-weak singlets if they only couple to $U(1)_Y$. The rich phenomenology of EW interactions is explained in detail in section 2.2.

Notably, what is probably the interaction we most often experience in our daily lives is not included in the SM. The microscopic nature of gravity and its quantum mechanical behavior remains as one of the big mysteries in modern physics and discovering it is amongst the biggest challenges of fundamental physics in the 21st century.

Now that we have seen what the main ingredients of the SM are, it is a good time to look at its particle content[3]. Particles that make up matter are *Dirac fermions*. They follow Dirac's equation -neutrinos might be an exception to this- and have spin 1/2. They can be further subdivided into

[2]Here the relevant energy scale is given by the QCD confinement scale $\Lambda_{QCD} \sim 200MeV$ which roughly marks the threshold energy above which QCD becomes perturbative.

[3]It would probably be more accurate to speak about field content since quantum fields are the actual building blocks of the SM model.

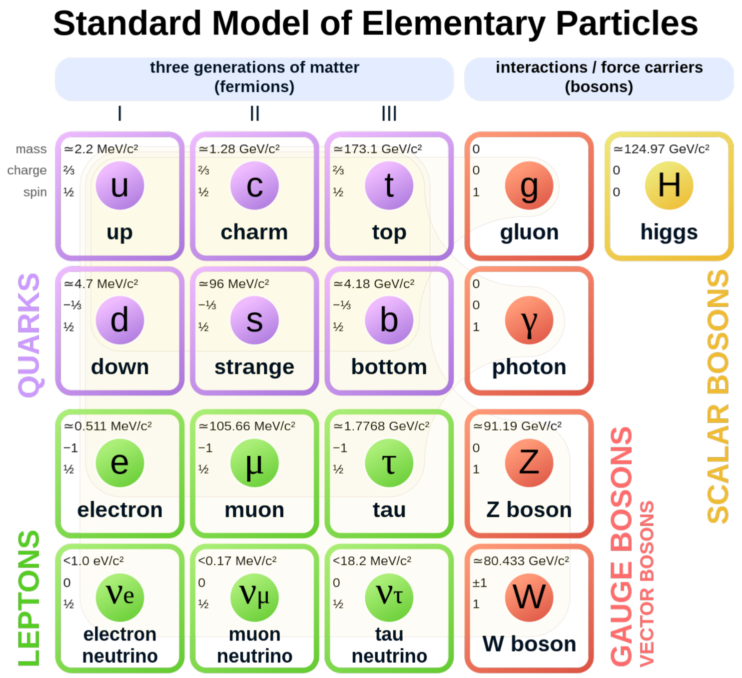


Figure 2.1: Summary of the elementary particles and their properties in the SM. Taken from [13].

two groups, those which couple to the strong interaction are called quarks and those that do not are called leptons. All fermions couple to the weak interaction and all of them are electrically charged except the neutrinos. The so called *force carriers* are the vector bosons mentioned before and the last and newest member of the SM is the Higgs boson, which is a scalar boson. Its associated field explains the origin of the masses of elementary particles as well as the emergence of the weak and electromagnetic interaction as low energy phases of the unified EW interaction. The table in figure 2.1 summarizes the properties of the SM particles.

The reported quark masses are the *current masses*. They differ from the so called *constituent masses*, which are typically much larger -especially for the lighter quarks- and can be regarded as effective masses accounting for the interaction of the quarks with the QCD potential inside the hadron. Quark masses can not be measured directly as they are always confined in bound states[4]. They need to be be extracted from measurements of hadronic properties and their definition depends on the particular theoretical framework used [4]. Only upper limits for the neutrino masses are shown in figure 2.1, on top of those limits there is also a much tighter limit on the sum of neutrino masses coming from cosmological measurements[4]:

$$\sum_i m(\nu_i) < 0.13 \text{ eV}/c^2. \tag{2.1}$$

It is worth noting tough that this limit has some dependence on the cosmological model used. As of today, the origin of neutrino masses constitutes one of the biggest mysteries in particle physics.

Finally, it is important to mention that each of the particles in the SM has an associated *antiparticle* which has exactly the same quantum numbers but opposite charges.

[4]The top quark is an exception to this. Its lifetime is shorter than the hadronization time so that, when produced in a high energy collision, it decays before hadronising.

2.1.1 The Standard Model Lagrangian

All the information about the particle content and interactions in the SM can be encapsulated in a relatively compact object, this object is the Lagrangian[5]. Employing the tools provided by QFT together with the lagrangian, one can compute observables for every process where particles included in the SM are involved. The SM Lagrangian can be divided into three components:

$$\mathcal{L} = \mathcal{L}_{Higgs} + \mathcal{L}_{Gauge} + \mathcal{L}_{Yukawa}. \tag{2.2}$$

The first component describes the dynamics of the Higgs boson:

$$\mathcal{L}_{Higgs} = [D_\mu \phi]^\dagger [D^\mu \phi] + \mu^2 \phi^\dagger \phi - \frac{\lambda}{4}[\phi^\dagger \phi]^2, \tag{2.3}$$

where D denotes the covariant derivative, ϕ is the Higgs doublet -by definition, it transforms as a doublet under EW interactions- and the parameters μ and λ respectively control the Higgs mass and self-interaction. Whenever an index appears repeated in an equation, summation is implied. The Higgs doublet is defined as follows:

$$\phi = \begin{pmatrix} h^+ \\ v + h^0 \end{pmatrix} \quad Y_\phi = 1/2, \tag{2.4}$$

where v is the vacuum expectation value, conventionally extracted from the neutral component. On the other hand, the covariant derivative is defined as:

$$D^\mu = \partial^\mu + i g_s G^\mu_a T^a + i g W^\mu_b \tau^b + i g' B^\mu Y. \tag{2.5}$$

Here g_s, g and g' are the coupling constants of the strong, and EW interactions, respectively. T^a, τ^b and Y are the generators of $SU(3)_c$, $SU(2)_L$ and $U(1)_Y$ and G^μ_a, W^μ_b and B^μ are the gluon and electroweak boson fields. The presence of the generators ensures that only fields transforming in non trivial ways under the symmetry groups will couple to the corresponding vector bosons.

The gauge part of the lagrangian describes the dynamics of the gauge bosons and fermions:

$$\mathcal{L}_{Gauge} = \sum_F i\overline{\psi_F}\gamma_\mu D^\mu \psi_F + G^a_{\mu\nu} G^{\mu\nu}_a + W^a_{\mu\nu} W^{\mu\nu}_a + B\mu\nu B^{\mu\nu}. \tag{2.6}$$

Here the index F runs over all the fermion doublets, the lepton doublets:

$$\begin{pmatrix} \nu_e \\ e^- \end{pmatrix} \quad \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix} \quad \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}, \tag{2.7}$$

And the quark doublets:

$$\begin{pmatrix} u \\ d \end{pmatrix} \quad \begin{pmatrix} c \\ s \end{pmatrix} \quad \begin{pmatrix} t \\ b \end{pmatrix}. \tag{2.8}$$

The tensors contain information about how the gauge bosons behave and interact among them. They can be regarded as generalizations of Maxwell's equations of electromagnetism. The simplest one is $B_{\mu\nu}$:

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu, \tag{2.9}$$

[5]Even tough in this case it would be more accurate to call it Lagrange density.

which is identical to the celebrated electromagnetic tensor. It tells us that the corresponding boson is neutral under the interaction associated to its own symmetry group as it does not contain interaction terms. The other two tensors are identical in form, only the gluon tensor is given here:

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + g_s f^{abc} G^b_\mu G^c_\nu. \tag{2.10}$$

It contains an extra term describing the interaction between gluons: in a non abelian gauge theory gauge bosons are charged with respect to the interaction they mediate. This generates a much richer phenomenology.

Finally, we have the Yukawa term, which describes the interaction between the Higgs doublet and the fermions, generating mass terms for the latter:

$$\mathcal{L}_{Yukawa} = -\overline{Q^i_L} Y^{ij}_U \varepsilon \phi^* U^j_R - \overline{Q^i_L} Y^{ij}_D \phi D^j_R + \overline{L^i_L} Y^{ij}_E \phi E^j_R + h.c. \quad . \tag{2.11}$$

Here the latin indices run over the three fermion generations. Y_U, Y_D and Y_E are the Yukawa matrices containing the mass terms for u-type quarks, d-type quarks and charged leptons, respectively, and term $h.c.$ denotes the hermitian conjugate of everything else. The matrix ε is a 2×2 matrix defined in weak-isospin space to ensure the absence of max mixing couplings of the type $\overline{d_L} u_R$ after SSB, not seen in nature:

$$\varepsilon = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{2.12}$$

2.1.2 Lepton flavor universality

Equation 2.6 encodes the principle of Lepton Flavor Universality (LFU). If we look at the first term we see that all the terms in the sum are completely identical, there is a priory no difference between different fermion families. There is however a trick, the covariant derivative defined in equation 2.5 does not act in the same way on all fermions, the second term associated to the strong interaction vanishes when acting on leptons since we required leptons to be colorless. This establishes a clear difference between quarks and leptons. However, if we focus on the leptons only, all the other terms are completely equivalent as long as we require all lepton families to have the same hyper-charge and the two electroweak couplings to be universal. This is not the most general case but it is what experiments tell us. In order to find a difference between leptons we have to go to equation 2.11, there we find the Yukawa matrices whose elements are different for each of the lepton families. This is required to incorporate in the model the experimental observation of different leptons having different masses. Lepton universality follows from here: μ and τ leptons are ”heavier copies” of the electron, they have identical properties and can only be differentiated by the strength of their couplings to the Higgs boson, which generates different masses once the SSB mechanism is active.

2.2 The Electroweak sector

In this section we take a closer look at the part of the SM that describes weak and electromagnetic interactions. The electroweak model, also known as *Glashow-Weinberg-Salam* model (GSW) was put forward during the late 60s on an attempt to unify both interactions into a single theory [14-17]. Attempts to build a gauge theory describing weak interactions had already been made prior to the development of the EW model [18], however, there were a number of problems arising from the nature of the mass terms present in the Lagrangian. First of all, massive gauge bosons were needed in order to explain the short range nature of the weak interaction, which are not allowed by gauge symmetry. Similar problems arose when trying to explain the origin of fermion masses while imposing gauge symmetry in the lagrangian. By that time it was clear that in a QFT model where the weak interaction would emerge as the consequence of a gauge symmetry, that symmetry would have to be broken. In the Standard Model the breaking of the electroweak symmetry is achieved *spontaneously* due to a phase transition in the field configuration of a scalar field [19-23]. This mechanism breaks the electroweak symmetry leaving only the known $U(1)$ symmetry of electromagnetism

$$SU(2)_L \times U(1)_Y \to U(1)_Q.$$

The corresponding scalar field is the Higgs field. One of the physical consequences of the Spontaneous Symmetry Breaking (SSB) of the electroweak symmetry is the emergence of three massive gauge bosons acting as mediators of the weak interaction which can be expressed as linear combinations of the gauge bosons introduced in section 2.1. The charged W bosons are defined as:

$$W^{\pm} = \frac{1}{\sqrt{2}}[W^1 \mp iW^2], \tag{2.13}$$

and the two neutral bosons:

$$\begin{pmatrix} Z^0 \\ A \end{pmatrix} = \begin{pmatrix} \cos\theta_W & -\sin\theta_W \\ \sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} W^3 \\ B^0 \end{pmatrix}. \tag{2.14}$$

The mixing parameter θ_W is the *weak mixing angle*, also known as *Weinberg angle*, and is a free parameter to be determined experimentally.

The prediction of a third weak boson, neutral, and mediator of the so far undiscovered weak neutral current is considered to be one of the first successes of the model. Not only was the existence of the Z^0 boson anticipated, but the relationship between its mass and the W mass was predicted accurately. The properties of the new neutral current interaction predicted by the model were also confirmed in future experiments. The EW model also establishes relationships between the couplings in the EW sector, which have been experimentally verified:

$$\cos\theta_W = \frac{M_W}{M_Z}, \quad e = g\sin\theta_W, \quad g' = g\tan\theta_W. \tag{2.15}$$

After SSB we are left with the interaction vertices for the weak charged currents (CC) and the neutral current (NC) shown in figure 2.2.

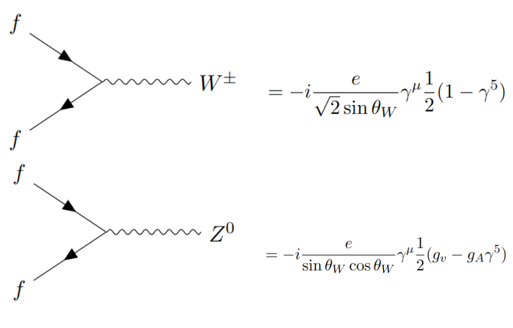


Figure 2.2: Interaction vertices for the weak charged currents (top) and the neutral current (bottom).

The weak CC has a *Vector-Axial* (V-A) structure, which follows from the fact that it only couples to left-handed chiral fermions since the combination $\frac{1}{2}(1-\gamma^5)$ is nothing but the fermionic projector onto the left-handed chiral subspace. In this equation γ^5 can expressed as a product of Dirac gamma matrices and the name of the currents follows from the behavior of each of the combinations of gamma matrices under parity transformations: γ^μ behaves as a 4-vector whereas the combination $\gamma^\mu\gamma^5$ behaves as an axial vector. The weak NC on the other side has a somewhat more complicated structure, the vector and axial-vector coefficients are different for every fermion type and are given in the SM by:

$$g_v = T^3 - 2Q\sin^2\theta_W, \quad g_A = T^3, \tag{2.16}$$

where T^3 is the eigenvalue of the third generator of the $SU(2)_L$ group, it has a value of $1/2$ for neutrinos and u-type quarks and a value of $-1/2$ for d-type quarks and charged leptons. The parameter Q is the electric charge of the fermion. The combination of both vector and axial-vector currents is the origin of the C and P violation in weak processes.

The form of the interaction vertices follows from the expression of the CC and NC interactions after SSB. The CC of the weak lagrangian has the following form[6]:

$$\mathcal{L}_{CC} \supset \overline{u^i_L}\gamma^\mu W^+_\mu d^i_L + \overline{e^i_L}\gamma^\mu W^+_\mu \nu^i_L + h.c. \quad , \tag{2.17}$$

where the index i runs over the three generations and we have used the definition of the chiral projection operator $\frac{1}{2}(1-\gamma^5)$ to simplify the currents writing the left-handed chiral components only, the right handed components vanish. For the NC we have:

$$\mathcal{L}_{NC} \supset \overline{Q^i} Z^0_\mu \gamma^\mu (g_v - g_a\gamma^5) Q^i + \overline{L^i} Z^0_\mu \gamma^\mu (g_v - g_a\gamma^5) L^i. \tag{2.18}$$

Here we have used the letters Q and L to denote the quark and lepton doublets (they contain both left and right handed components). We can see that NCs are diagonal in flavor space i.e. they do not mix different flavors. This is the reason why FCNC do not appear in the SM at tree level. They can only be induced by loop quantum corrections in processes involving virtual W bosons.

2.2.1 Quark mixing and the CKM matrix

The Yukawa matrices have been introduced in equation 2.11, defined in flavor space. These matrices do not necessarily have to be diagonal and, in fact, they are not diagonal in the SM. We can generalize the formalism by introducing unitary matrices relating the mass and flavor states:

$$u = \mathbf{U_u}\hat{u} \quad d = \mathbf{U_d}\hat{d}. \tag{2.19}$$

We have denoted the mass states with a hat to differentiate them from the flavor states, defined as the eigenstates of the weak interaction. This has consequences on the expression of the CC interactions which mix different flavors:

$$\mathcal{L}_{CC} \supset \overline{\hat{u}_L}\mathbf{U_u}^\dagger\mathbf{U_d}\gamma^\mu W^+_\mu \hat{d}_L + \overline{\hat{d}_L}\mathbf{U_d}^\dagger\mathbf{U_u}\gamma^\mu W^-_\mu \hat{u}_L + \text{leptons}. \tag{2.20}$$

For simplicity, we have removed the index i from equation 2.17 so that now $\hat{u}$ and $\hat{d}$ represent 3-component vectors in flavor space. The matrix connecting different quark flavors in CC interactions is known as the CKM matrix, named after Cabbibo, who was the first to propose a similar quark

[6]The equation given here assumes the quark states to be flavor states, if we work in the mass basis we would need to include the corresponding CKM matrix elements, see following section.

mixing mechanism [24], and Kobayashi and Maskawa, who extended Cabbibo's model to include the 6 known quarks [25]:

$$\mathbf{V_{CKM}} \equiv \mathbf{U_u}^{\dagger}\mathbf{U_d}. \tag{2.21}$$

Using the CKM matrix we can rewrite the the expression for the CC interaction:

$$\mathcal{L}_{CC} \supset \overline{\hat{u}_L}\mathbf{V_{CKM}}\gamma^{\mu}W_{\mu}^{+}\hat{d}_L + \overline{\hat{d}_L}\mathbf{V_{CKM}}^{\dagger}\gamma^{\mu}W_{\mu}^{-}\hat{u}_L + \text{leptons}. \tag{2.22}$$

Thus, the Higgs mechanism allows for mixing between quark generations through CC interactions as long as the quarks are massive and the masses of all u-type and d-type quarks are different among themselves. If the masses were the same the mass and flavor basis would be equivalent and there would be no mixing between quark generations.

It is worth noting however that NC are also diagonal in the mass basis as they do not mix different flavors. From the unitarity of the introduced matrices it follows that:

$$\mathbf{U_u}^{\dagger}\mathbf{U_u} = \mathbf{U_d}^{\dagger}\mathbf{U_d} = \mathbf{1}. \tag{2.23}$$

It can be shown that the CKM matrix introduces 4 new parameters to the SM, which have to be experimentally determined. Three of them can be expressed as real mixing angles and the fourth one will be a complex phase, responsible for CP violation in the weak sector. The CKM matrix presents a hierarchical structure suppressing mixing between generations:

$$|\mathbf{V_{CKM}}| = \begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix} \approx \begin{pmatrix} 0.974 & 0.224 & 0.004 \\ 0.221 & 0.975 & 0.041 \\ 0.009 & 0.0041 & 1.014 \end{pmatrix}. \tag{2.24}$$

The values are obtained from a global fit by the *CKMfitter* collaboration, combining all available measurements [4]. Note that from the unitarity of the CKM matrix it follows that all its elements must have a magnitude smaller than one, the experimental error in the $|V_{tb}|$ measurement is large and when taken into account, the value is still compatible with being smaller than 1. The origin of this hierarchical structure is a mystery of the SM.

2.2.2 Flavor changing neutral currents

Quark mixing allows the existence of FCNC which can be induced in the SM through the exchange of W bosons in Next to Leading Order (NLO) processes in QFT. We use a $b \to s\gamma$ transition to illustrate the physics. The Feynman diagram for this process is shown in figure 2.3.

From figure 2.3 it is clear that FCNCs involve transitions between different quark generations in at least two interaction vertices. Thus, without quark mixing, FCNC processes would be prohibited

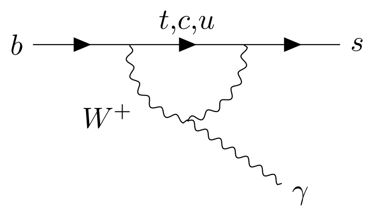


Figure 2.3: Feynman diagram describing a $b \to s\gamma$ transition at NLO.

even at loop level. Using the *Feynman rules* we can obtain the matrix element for the process, ignoring factors which are irrelevant in our discussion:

$$\mathcal{M} \propto e\frac{g^2}{16\pi^2}V_{tb}V_{ts}^*F\left(\frac{m_t^2}{m_W^2}\right). \tag{2.25}$$

In order to obtain this expression one needs to use the unitarity of the CKM matrix together with the fact that $m_t \gg m_c, m_u$. The function F comes out as the result of integrating the momenta of the virtual particles over the loop. It can be shown to be a function of the ratio between the masses of the involved quarks squared and m_W^2. The large mass of the t quark compared to the masses of the lighter quark flavors allow us to keep the term involving the t quark only. From equation 2.25 we can single out two main suppression factors that make these type of decays very rare in the SM:

- **Loop suppression:** since both the couplings o the weak and the electromagnetic interaction are smaller than 1, the larger the amount of vertices in a Feynman diagrams the smaller its contribution is. For this reason, diagrams with loops are suppressed with respect to tree level diagrams, with only two interaction vertices.

- **CKM suppression:** the hierarchical structure of the CKM matrix shown in the previous section indicates that transitions from a t quark to anything else but a b quark have a very low probability. From equation 2.24 we can estimate a suppression of $O(10^{-4})$ coming from the CKM matrix elements only.

Note that in $b \to sl^+l^-$ transitions there is still an additional suppression factor coming from the extra vertex where the virtual photon or Z^0 boson decays into a pair of leptons. All this makes $b \to sl^+l^-$ transitions an extraordinary tool to look for new physics.

2.3 Effective field theories and $b \to sl^+l^-$ transitions

The formalism of effective Field Theory (EFT) is a very powerful tool developed within the framework of QFT, especially useful for the analysis of processes where different energy scales are involved. The main advantage is that it provides a framework to study such processes in a model independent way. The only needed inputs are the operators describing the different types of interaction, which will depend on the process under consideration, and the relevant energy scales involved. It thus can be used to probe our theory predictions without having to rely on specific models, making it the perfect tool to look for new physics in the broadest sense.

The idea behind the concept of EFT is simple, consider a QFT with a large fundamental scale μ, this scale could be the mass of a particle contributing to a ”low energy” process with energy E such that $E \ll \mu$. It then seems natural to look for an expansion in powers of E/μ [27]. This expansion can formally be done within the functional integral formulation of QFT. The idea is similar to the idea of *renormalization*, we set a high energy cut-off scale and integrate out in the functional integral the field modes corresponding to the microscopical degrees of freedom, which are only relevant at energies above the chosen cut-off. This removes the *Ultraviolet UV* divergences from the QFT and enables us to make predictions with it. The details of the calculation can be found, for example, in [27]. The final and relevant result for our discussion is that it is possible to define an object, called the *effective Hamiltonian*, which is expanded as the sum of all possible contributing operators multiplied by their respective coupling strength coefficients:

$$\mathcal{H}_{eff} = \sum_i C_i(\mu) O_i(\mu). \tag{2.26}$$

The coefficients C_i are the *Wilson coefficients* [29]. They contain information about the *short-distance* perturbative effects which are integrated out to get the expansion and are process independent. They are specific to the local operators O_i defined by their Lorentz structure [27]. These local operators contain information about the *long-distance* non-perturbative effects present in the theory. There is one for every Lorentz-invariant quantity that can be built out of the initial and final particles participating in the process. The operators appearing in this expansion are local in the sense that they describe *contact-like* interactions, happening at a single space-time point. The non-local effects coming from the propagators of virtual particles that could contribute to the process have been integrated out and are contained in the corresponding Wilson coefficients. Since the Wilson coefficients are process independent they can be calculated in the SM by matching the corresponding local operators to the SM ones. For a process from an initial state $|i\rangle$ to a final state $|f\rangle$ we have:

$$\mathcal{M}(i \to f)_{SM} = \sum_i C_i(\mu) \langle f| O_i(\mu) |i\rangle . \tag{2.27}$$

Both coefficients and operators are evaluated at a given energy scale μ where matching to SM predictions is reliable. Then one can derive the *Renormalization Group Equations* (RGE) for the EFT operators and their respective Wilson coefficients in order to evolve them to the energy scale relevant to our process. The RGE were first derived within the context of renormalization theory. They predict that couplings in the SM are scale dependent rather than just having constant values. This scale dependence absorbs the UV divergences found when computing quantum corrections to tree level processes in QFT and has been experimentally confirmed from measurements of the SM couplings at different scales. As an example of this scale dependence, figure 2.4 shows the most recent results confirming the running of the strong coupling constant.

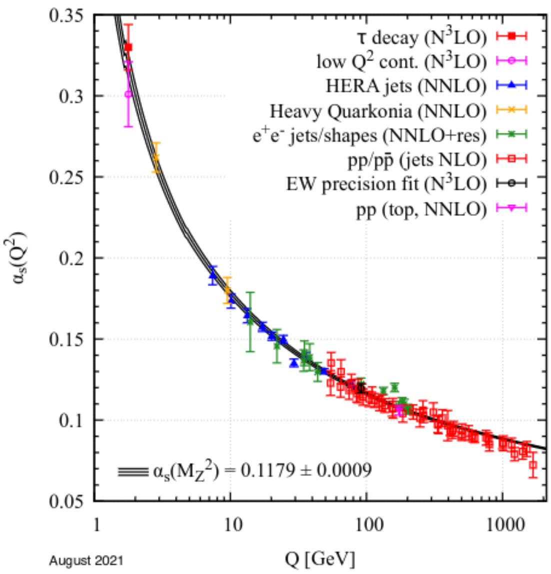


Figure 2.4: Summary of the latest measurements of α_s as a function of the energy scale Q^2. The different processes used to measure it together with the order in perturbation theory of the QCD calculations used to extract its value are indicated in the legend. Taken from [4].

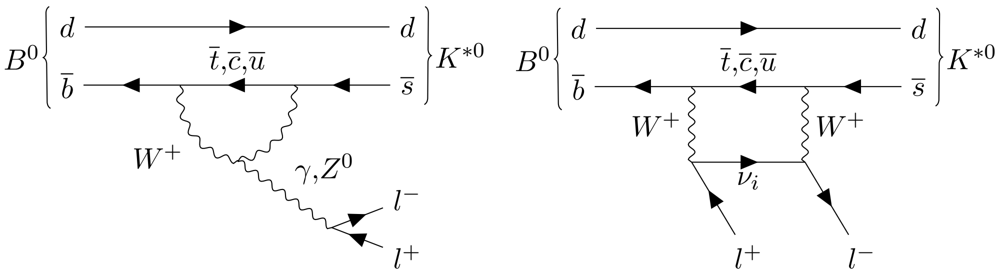


Figure 2.5: Dominant SM Feynman diagrams for the $B^0 \to K^{*0} l^+ l^-$ decay.

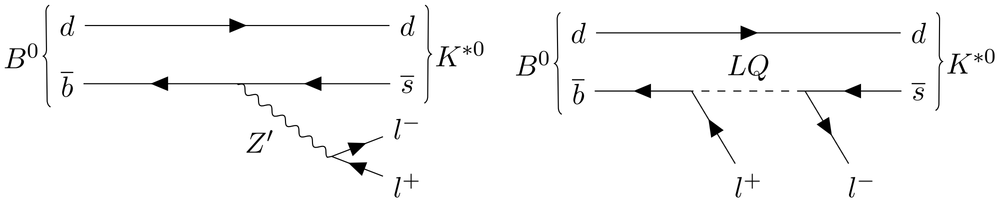


Figure 2.6: NP contributions to the $B^0 \to K^{*0} e^+ e^-$ decay from a model with a $Z^{0'}$ vector boson (left) and a lepto-quark (right).

Once we have the values of the Wilson coefficients computed within the SM, we can compare them with the ones extracted from measurements of processes where the corresponding local operators participate. It is customary to split the Wilson coefficients into SM and NP contributions when expressing the experimental results:

$$C_i \equiv C_i^{SM} + C_i^{NP}. \tag{2.28}$$

The coefficients C_i^{NP} contain all possible contributions from any NP effect not accounted for in the SM.

2.3.1 Effective operators in $B^0 \to K^{*0} l^+ l^-$ decays

Using the formalism of EFT we can get valuable information about the properties of any new effects that might be contributing to $B^0 \to K^{*0} l^+ l^-$ decays. The two main SM contributions for this decay come from NLO Feynman diagrams involving internal loops where W bosons are exchanged. They are depicted in figure[7] 2.5.

As it was pointed out in section 2.1, the couplings between the leptons and the W, Z^0 and γ bosons participating in the two diagrams do not distinguish the flavor of the final leptons. However, many NP models predict the existence of non-universal lepton couplings that could generate sizeable contribution to the branching fractions. Among the most popular ones we have models predicting the existence of lepto-quarks[31, 32], bosonic particles that would introduce new vertices between quarks and leptons, or models with $Z^{0'}$ vector bosons[33-35] usually built within the context of dark matter. These NP processes can happen at tree level, avoiding the loop suppression present in SM diagrams. Feynman diagrams for these models are shown in figure 2.6. Contributions from diagrams like the ones represented in figure 2.6 could modify the magnitude of some of the Wilson coefficients that participate in the process.

[7] There is actually a third diagram contributing to the same order in perturbation theory, equivalent to the left one but with the internal quarks and W^+ lines exchanged.

The EFT describing the physics of $b \to sl^+l^-$ decays is the *Weak Effective Field Theory* (WET), which can be regarded as the modern generalization of the Fermi theory of weak interactions [39]. The Fermi theory is an approximation to the SM theory of weak interactions, valid when $E \ll m_W$, which introduces a contact 4-particle interaction to describe processes like nuclear β decay. In such theory, short distance contributions associated to the exchange of W bosons are absorbed in the Fermi constant G_F. In the same way, WET can be applied to processes where $E \ll m_W$, a condition that is fulfilled in $b \to sl^+l^-$ transitions, where the relevant energy scale is given by the b quark mass $m_b \ll m_W$. The effective hamiltonian [40] for FCNC $b \to sl^+l^-$ transitions takes the form:

$$\mathcal{H}_{eff} = \frac{-4G_F}{\sqrt{2}} V_{tb}V_{ts}^* \sum_i C_i(\mu_{m_b})O_i(\mu_{m_b}). \tag{2.29}$$

Here we have made the same approximation as in equation 2.25, neglecting the contributions from u and c quarks, suppressed by their smaller masses within the loops.

The relevant local operators describing FCNC processes are [38]:

$$O_7 = \frac{e}{16\pi^2} m_b(\bar{s}\sigma_{\mu\nu}P_R b)F^{\mu\nu}, \tag{2.30}$$

$$O_{9l} = \frac{e}{16\pi^2} m_b(\bar{s}\gamma_\mu P_L b)(\bar{l}\gamma^\mu l), \tag{2.31}$$

$$O_{10l} = \frac{e}{16\pi^2} m_b(\bar{s}\gamma_\mu P_L b)(\bar{l}\gamma^\mu \gamma_5 l), \tag{2.32}$$

where, by definition, $\sigma_{\mu\nu} = \frac{i}{2}[\gamma_\mu, \gamma_\nu]$ and $F^{\mu\nu}$ is Maxwell's electromagnetic tensor, analogous to the one defined in equation 2.9. $P_{L,R}$ are the left and right-chiral projection operators. On top of these three we would also have the operators O_7', O_{9l}' and O_{10l}' where the P_R are switched by P_L and vice-versa. These operators are suppressed in the SM by a factor $O_i' \sim \frac{m_s}{m_b}O_i$ but could be enhanced in the presence of NP [41].

The processes representing the contributions from each of the local operators are represented in figure 2.7.

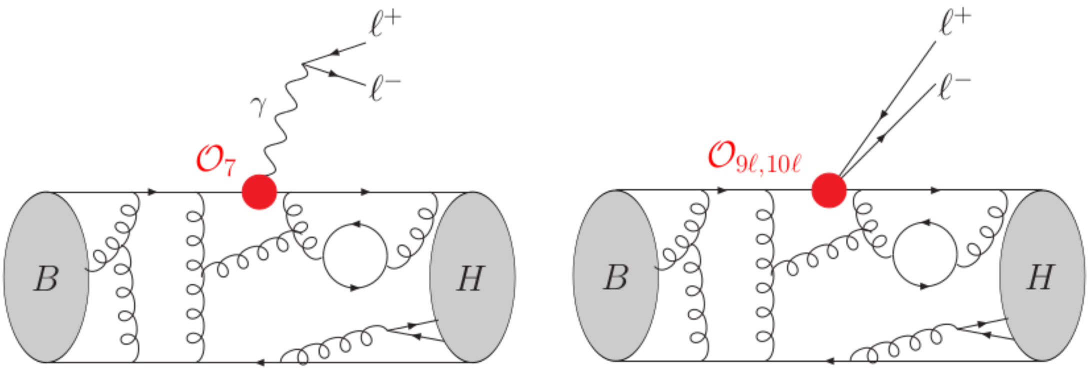


Figure 2.7: Representation of the two main contributions to $B^0 \to K^{*0}l^+l^-$ decay processes in the WET approach. The red dots represent the local operators. Taken from [38]

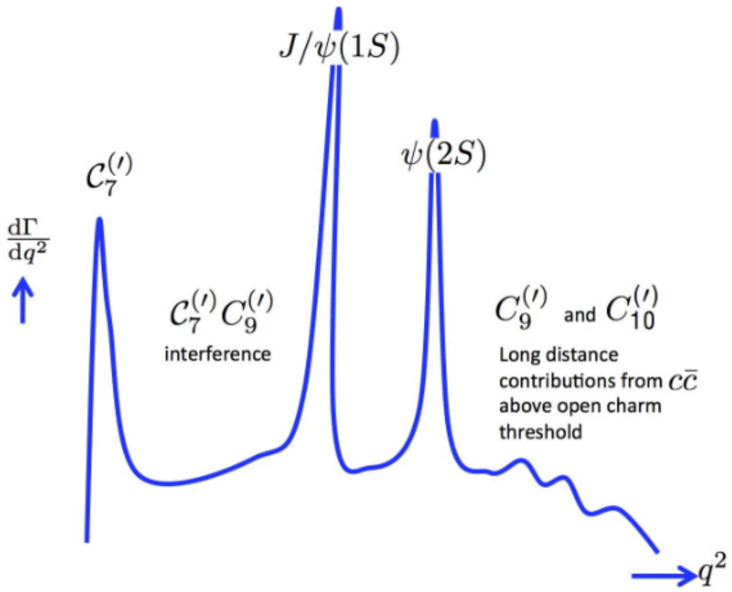


Figure 2.8: Sketch of the q^2 spectrum of the $B^0 \to K^{*0}l^+l^-$ process. The main Wilson coefficients contribution to each region are indicated. The height of the charmonium resonance peaks is not to scale. Adapted from [38].

We can see that soft gluon exchange plays an important role in the calculation of the branching ratios for $B^0 \to K^{*0}l^+l^-$ decays. All the associated non-perturbative QCD effects, which are the main sources of theory uncertainties are contained in the *form factors*, defined as the hadronic matrix elements $\langle K^{*0}|\, O_i(\mu)\, |B^0\rangle$. In this case the involved operators are the hadronic effective operators, describing the hadronic part of the decay, which can also be treated within the EFT formalism. They are of course different form the ones introduced in equations 2.30-2.32 to describe the leptonic part of the decay.

We can also use the EFT formalism to understand the phenomenology of the observed di-lepton squared invariant mass q^2 spectrum in $B^0 \to K^{*0}l^+l^-$ decays. It can help us to understand the physics behind these decays as well the discovery potential of testing LFU in the high-q^2 region. A sketch with the expected shape of the q^2 spectrum is shown in figure 2.8. We can see a pole at low energies associated to the Wilson coefficient C_7 that represents a photon exchange in the SM. The q^2 interval $1-6$ GeV$^2/c^4$ is conventionally referred to as the central q^2 region. In this region, effects coming from the interference between C_7 and C_{9l} dominate. This makes the central q^2 region sensible to new physics effects, mainly coming from vector currents contributing to $C_{9l}^{(\prime)}$, which is suppressed in the SM. The q^2 region $7-15$ GeV$^2/c^4$ contains the two largest structures found in the whole spectrum corresponding to J/ψ and $\psi(2S)$ charmonium $c\bar{c}$ resonances. These contributions are associated to tree level decays of the B^0 meson and have much larger branching fractions of $O(10^{-3}) - O(10^{-4})$. These large resonance peaks can not be easily vetoed in this region and the corresponding q^2 interval is therefore usually not used in LFU tests. Charmonium resonances can however be useful as control channels since they decay electromagnetically to the final lepton pairs and they are known to respect LFU. They can be used to cross-check the analysis strategy followed for the rare modes as well as in the definition of double ratios, reducing the experimental systematic uncertainties. Finally we have the high-q^2 region, defined above the $\psi(2S)$ resonance. This region is the main focus of this thesis. It is dominated by $C_{9l}^{(\prime)}$ and $C_{10l}^{(\prime)}$ contributions, which in the SM are associated to a Z^0 exchange, taking over the photon contribution at high energies. The high-q^2 region is also polluted by broad peaks from long distance contributions coming from excited charmonium resonances. As their mass is above the open charm threshold they mostly decay to D mesons and their contribution to $B^0 \to K^{*0}l^+l^-$ becomes less important. These charmonium contributions are ignored in the analysis presented in this thesis as they appear as wiggles that tend to cancel when integrated over the whole q^2 bin [38]. On top of that, their contributions are also expected to be LFU since they decay in the same way as J/ψ and $\psi(2S)$ so that neglecting them does not bias the results[8].

It is important to note that in presence of NP, both central and high-q^2 regions are expected to be affected in a similar way since the only difference between $C_{9l}^{(\prime)}$ and $C_{10l}^{(\prime)}$ are the nature of the couplings to leptons being the first coefficient associated to a vector coupling and the second one to an axial-vector coupling. The magnitude of NP contributions mostly depends on the energy scale at which NP effects become relevant and this energy scale is expected to be much larger than the mass of the B^0 meson, making the difference in energy between the high and central q^2 regions negligible.

2.3.2 Probing new physics with EFT

The fact that the Wilson coefficients are process independent allows the combination of results coming from different observables in a quite straightforward way. This approach can strengthen the significance of small deviations when they follow a coherent picture, showing deviations in the same set of Wilson coefficients coming from different measurements. Such a pattern has started to emerge during the last years with many results coming from rare B meson decays showing small deviations from the SM expectations. These are the *flavor anomalies* mentioned in chapter 1. A summary of all the results is provided in figure 2.9. In the figure, the latest LFU measurements are shown using different decays with the numbers in brackets indicating the q^2 range used in the analysis. The parameter P_5' is an angular observable that can be measured in decays like $B^0 \to K^{*0}l^+l^-$ and describes the angular distribution of the final state particles [43].

Besides presenting a coherent deviation pattern, all these observables are sensitive to the same set of Wilson coefficients. Hence, a combination of the results based on the EFT framework can increase the overall significance of the discrepancies in a model independent way. Such analysis can be performed using the `flavio` software package [44]. The result of the combination of some of the measurements is shown in figure 2.10.

The figure shows related observables grouped together independently. The blue region contains the results for theoretically clean observables, including LFU observables like the one studied in this thesis and the branching fraction of the fully leptonic $B_s^0 \to \mu^+\mu^-$ decay. These observables are mostly free from QCD uncertainties making theoretical predictions more robust. The region represented in orange groups together results form measurements of branching fractions and angular observables in semileptonic B meson decays. They contain larger theoretical uncertainties coming from the QCD part and results are therefore subject to the sensitivity of the theory predictions. One can see that significant deviations from the SM predictions can be found when combining all the measurements. However, care must be taken when interpreting this type of global fits since they just summarize a pattern of deviations into a common parametrization. This does not necessarily mean that the deviations can be explained by a single physics model.

[8]Indeed they would wash out any potential LFU violation coming from the $B^0 \to K^{*0}e^+e^-$ signal decay rather than introducing a positive bias in the final result.

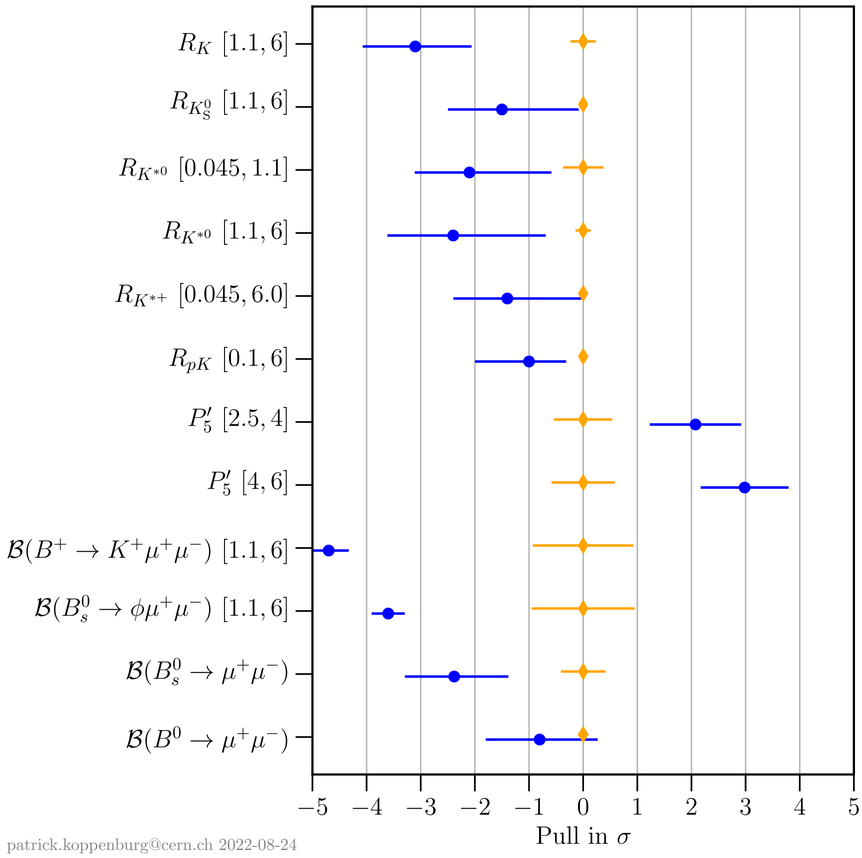


Figure 2.9: Summary of flavor anomalies found in rare B meson decays. Blue dots represent the measured values for the observables shown in the y axis, shifted such that the SM prediction is set to 0 for all of them to allow for an easier comparison. Taken from [42].

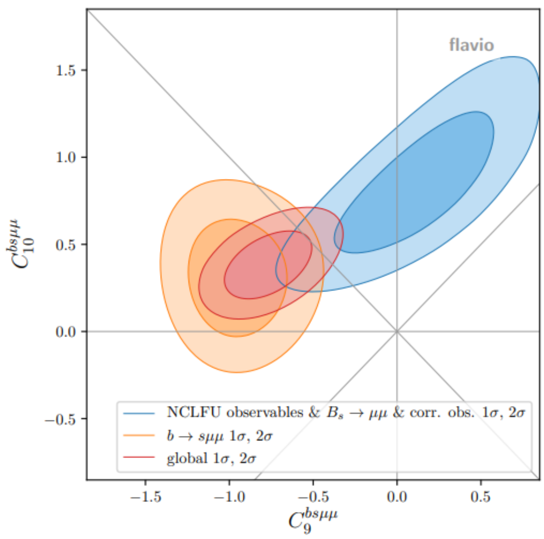


Figure 2.10: Global fit of the Wilson coefficients $C_9^{bs\mu\mu}$ and $C_{10}^{bs\mu\mu}$ obtained from measurements of processes mediated by the FCNC transition $b \to s\mu^+\mu^-$. Only NP contributions are shown, the SM contributions are set to 0. Related observables have been grouped together and then combined into a global fit which is represented in red. Contour lines represent the corresponding 1σ and 2σ limits. Taken from [37].

Chapter 3

The LHCb experiment

The theoretical background required to understand the physics of $B^0 \to K^{*0}l^+l^-$ decays was introduced in chapter 2. The next step is to understand how these decays are reconstructed at LHCb. In this chapter the LHCb detector is described, including the subdetectors and trigger system. In section 3.1 the LHCb detector and its main features are introduced. The different parts of the LHCb experiment are described in more detail in section 3.2. In section 3.3 some details about the tracking and particle identification performance are given. The trigger system is introduced in section 3.4, highlighting the differences between Run1 and Run2 and amongst different years of data-taking.

3.1 The LHCb detector at the LHC

The LHCb experiment is one of the four main experiments held at the *Large Hadron Collider* (LHC), the largest particle accelerator currently existing and central part of the *European Organization for nuclear Research* (CERN) accelerator complex. Located near Geneva inside a 27km long circular tunnel crossing the border between France and Switzerland, it accelerates bunches of protons (and sometimes also heavy ions) in opposite directions making them collide at four interaction points where the detectors are held.

The CERN accelerator complex contains a series of smaller accelerators, some of them recycled from past colliders, which sequentially accelerate charged particles until they become energetic enough to be injected in the LHC. There they are guided by superconducting magnets and receive the last boost reaching a final energy of 3.5 TeV in 2010-2011, 4TeV in 2012-2013 and 6.5TeV in 2015-2018 before they collide. A schematic representation of the full accelerator complex with the different experiments currently running at CERN is shown in figure 3.1.

The LHCb detector is located at point 8 in the LHC tunnel, inside of an underground cavern shared with the MoEDAL experiment [48]. It is an asymmetric single arm forward spectrometer specialised in the detection and study of decays of hadrons containing b and c quarks. The geometry of the detector makes it unique among the other LHC experiments. It was chosen to maximize the acceptance in the high pseudo-rapidity region where most of the $b\bar{b}$ and $c\bar{c}$ pairs are produced. Pseudo-rapidity is defined as:

$$|\eta| = |\ln(\tan(\theta/2))|, \tag{3.1}$$

where θ is the polar angle with respect to the proton momentum. This means that b and c hadrons are boosted along the forward -and backward- direction. This feature is a consequence of how they are produced at the LHC. The main production mechanism of heavy quark pairs is gluon-gluon fusion. Given the distribution of gluons inside of the proton, most of them carrying a small fraction of the proton's momentum, the most likely scenario is a parton level process where a low x gluon (x is the momentum fraction carried by the parton) interacts with another gluon that carries a sufficiently large amount of momentum to make up the energy necessary for producing the quark pair. The distribution of $b\bar{b}$ pairs produced at the LHC can be seen in figure [3.2]. In numbers, about 25% of produced $b\bar{b}$ pairs fall inside the LHCb acceptance, which only covers roughly a 3% of the full solid angle.

In the following sections we describe the overall experimental layout of the LHCb detector for the 2010-2018 data-taking period to which the data used in this analysis correspond. After 2018, LHCb underwent a big upgrade where most of the components were renovated in order to cope with the more challenging experimental conditions for Run 3, specially the larger instantaneous luminosity delivered by the LHC. Updated information about the new LHCb detector and its detector subsystems can be found in the LHCb website [50].

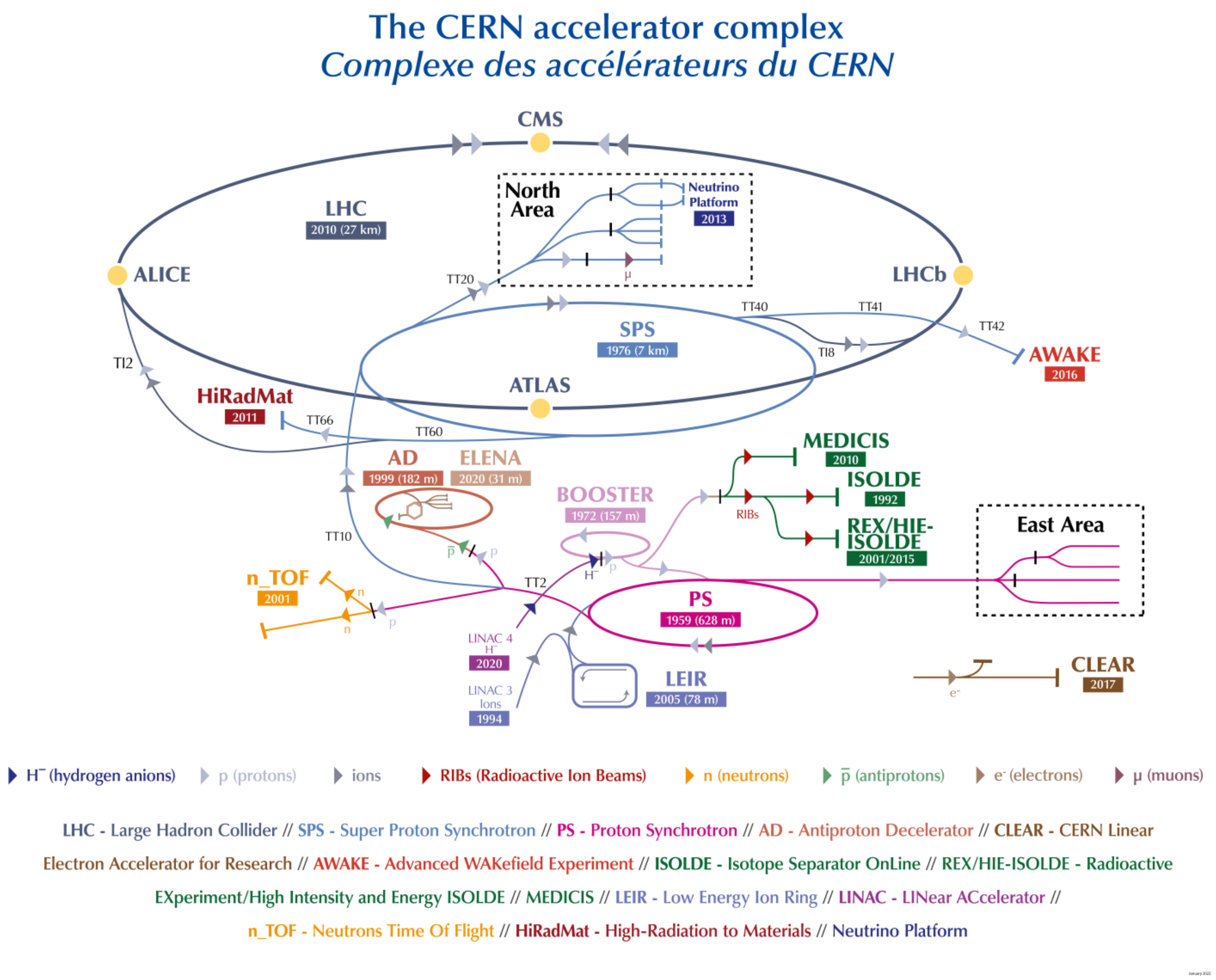


Figure 3.1: Schematic layout of the CERN accelerator complex [47].

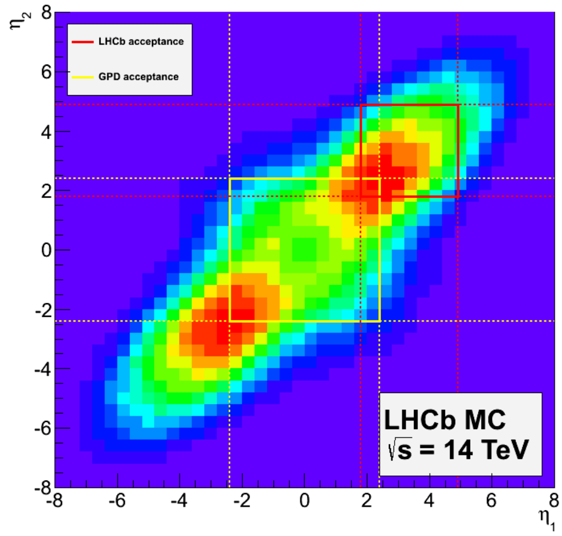


Figure 3.2: Pseudo-rapidity distribution of $b\bar{b}$ pairs produced at the LHC. The LHCb acceptance (red box) is compared to the acceptance of other General Purpose Detectors (GPD) at the LHC (yellow box). The plot is generated under the Run2 (2015-2018) experimental conditions using MC simulations where contributions from different parton level processes are included [49].

3.1.1 Experimental operation and setup

The main design physics goals of the LHCb detector are the search of NP in CP violation and rare decays of beauty and charm hadrons. The operating conditions and experimental setup of the LHCb detector are optimised to match the requirements needed to study these processes. One of the most challenging experimental signatures of these decays is the presence of secondary vertices which can be displaced by only a few mm from the primary pp interaction vertex so that their reconstruction require an excellent vertex resolution. Vertex resolution is affected by pile-up μ_{vis}, defined as the average number of visible pp interactions per bunch crossing. A large amount of pile up increases the background contribution and reduces the efficiency for reconstructing secondary vertices. The amount of pile-up grows with the instantaneous luminosity, which is proportional to the event rate. In order to keep a low pile-up rate a *luminosity levelling* procedure is introduced at LHCb. Proton beams are defocused at the LHCb interaction point reducing the overlap between the proton bunches at the beginning of a fill[1], when the instantaneous luminosity is larger. Then the transverse overlap between the beams is gradually increased in order to compensate the decrease in luminosity towards the end of the fill. This allows to keep the instantaneous luminosity recorded by LHCb stable within about 5% as can be seen in figure [3.3].

Keeping the instantaneous luminosity constant is important at LHCb in order to maintain the experimental conditions as stable as possible during the fill. In particular, it is crucial to have a constant track multiplicity -which grows with μ_{vis}- in order to achieve a good vertex resolution.

A very good mass resolution down to very low particle momenta is also required to resolve the small mass differences between B meson flavors. This is essential for measurements of rare decays like the ones studied in this thesis. Momentum measurements at LHCb are provided by the spec-

[1]In high energy physics, a fill consists of a set of proton bunches which circulate through the beamline and collide at the interaction points. A typical LHC fill lasts for about 12 hours.

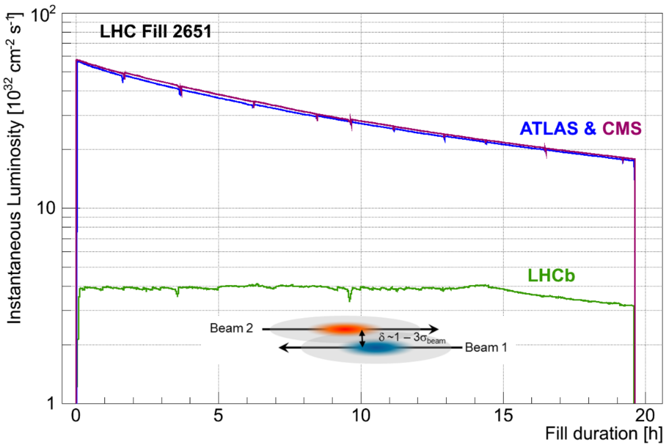


Figure 3.3: Evolution of the instantaneous luminosity during LHC fill 2651 at Run 1. The luminosity recorded by ATLAS and CMS are also shown for comparison. Taken from [45].

trometer, consisting on a warm dipole magnet which delivers an integrated field of about 4Tm, and the tracking system. The tracking system consists of the Vertex Locator (VELO) and the Trigger Tracker (TT) located upstream of the magnet, and the four tracking stations downstream of the magnet. Energy measurements are made available by the calorimeter system, described in section 3.2, which also provides useful information for particle identification. This information is complementary to the one provided by the Ring Imaging Cherenkov (RICH) detectors for charged particles, located upstream and downstream of the magnet to cover the low and high momentum ranges.

The layout of the LHCb experiment is shown in figure 3.4. LHCb uses a right-handed coordinate system with the z axis along the beamline directed towards the detector, the y axis along the vertical direction and the x axis perpendicular and pointing towards the center of the LHC circumference. The origin is located at the interaction point. Cylindrical polar coordinates (r, ϕ, z) are also used when appropriate. Charged particles are bent in the horizontal plane by the magnet and the LHCb detector has an angular coverage from approximately 15 mrad to 300 (250) mrad in the horizontal (vertical) plane [46]. The lower limit is given by the size of the beampipe.

Most detector components are assembled in two halves referred to as the detector A- and C-sides which can be moved horizontally allowing the access to the beampipe and facilitating maintenance. One of the advantages of the LHCb detector design is that it allows to keeps most of the read-out electronics outside of the acceptance so that they can be accessed more easily and their interference with the performance of the detector is reduced [46]. This is not possible in detectors with a closed geometry such as CMS or ATLAS.

3.2 The detector subsystems

Each of the detectors comprising the LHCb experiment are specialised on different tasks providing complementary information about the recorded events which is then used both online in the trigger and offline in the final analysis. In the following, a more detailed description of the different parts of the LHCb detector is given, highlighting the features that are relevant for the analysis presented in this thesis.

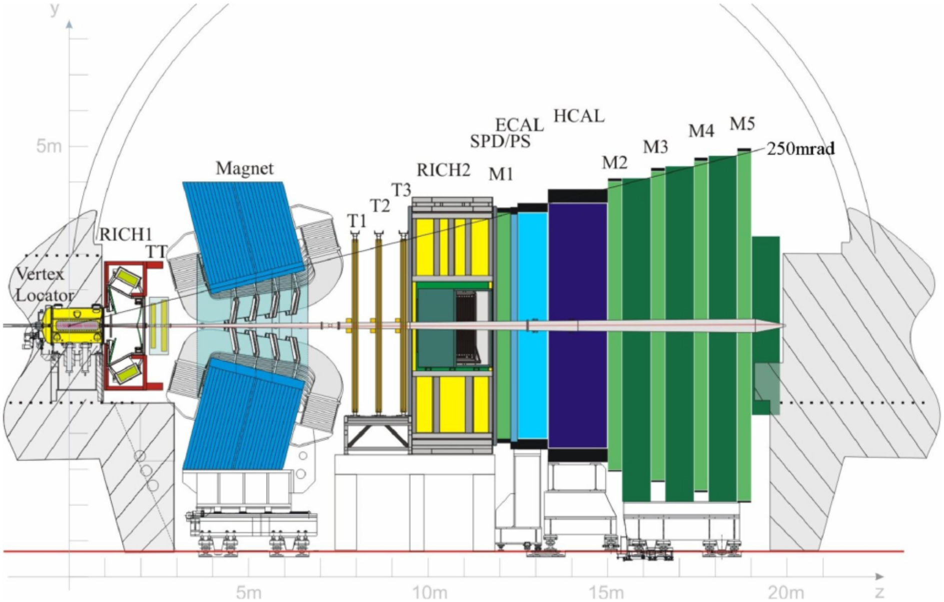


Figure 3.4: Side view of the LHCb detector. Taken from [45]

Figure 3.5: Photo of the VELO modules mounted in line. Taken from [51]

3.2.1 The Vertex Locator

The VELO detector provides the required vertex resolution for analyses performed at LHCb and plays an important role in the LHCb trigger, selecting events that contain signatures with displaced secondary vertices. It consists of a series of 42 silicon modules arranged along the beamline, very close to the interaction point to identify displaced secondary vertices like those of B meson decays. The layout of the VELO detector is shown in figure 3.5. It covers the pseudorapidity range $1.6 < \eta < 4.9$ pseudorapidity range, providing a polar angle coverage down to 15 mrad for particles emerging at $z = 10.6$ cm downstream of the nominal interaction point [46]. Each VELO module consists of two sensors that measure the r and ϕ coordinates of the particle hits. These are respectively called R and Φ sensors. They are placed at a radial distance from the LHC beams which is smaller than the beamline aperture required during injection. This means that the modules have to be retracted by 29 mm every time beams are injected. Retraction is done automatically by the VELO stepper motors whose movement is monitored to calculate the corresponding alignment

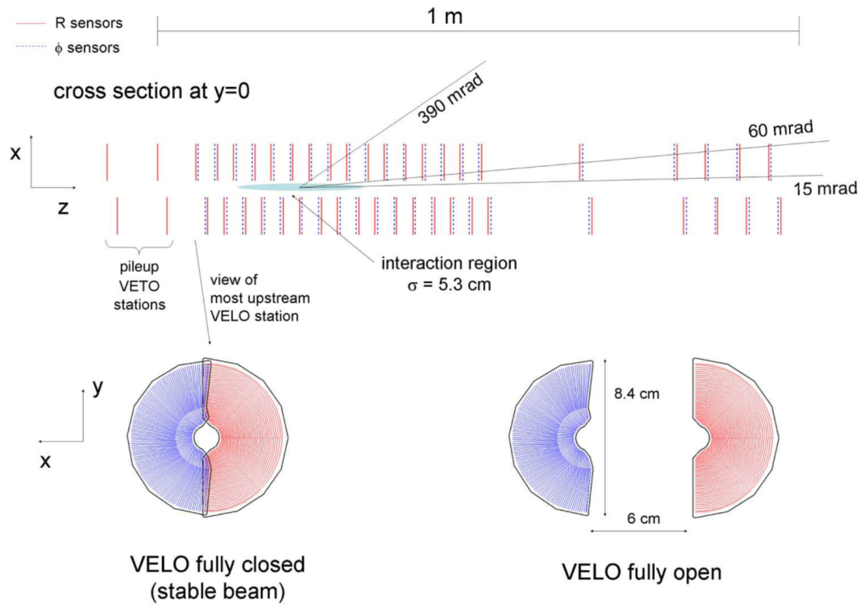


Figure 3.6: Cross section of the VELO modules in the xz plane showing the geometry and the position of the sensors. The bottom inset shows the front face of one of the modules in both the closed and open positions. Red and blue lines represent the orientation of the silicon strips in each sensor. Taken from [46]

corrections needed to compensate any possible misalignment caused in the process. The sensors are mounted inside of a vacuum vessel separated from the machine vacuum by a thin wall of a corrugated aluminum sheet to reduce the amount of material traversed by charged particles, these are the *RF-foils*. In order to cover the full azimuthal acceptance the two detector halves are required to overlap. Figure 3.6 shows the arrangement of the VELO modules along the beamline.

In the R sensors the diode strips are implanted along concentric semicircles around the beamline. In the Φ sensors the strips run almost radially, with a skew introduced at a radius of 17.25 mm in order to improve pattern recognition. The distribution of the strips in the two sensors is also shown in figure 3.6. The pitch of the strips within a module varies from 38 μm at the inner radius of 8.2 mm, increasing linearly to 102 μm at the outer radius of 42 mm.

3.2.2 The Trigger Tracker

The Trigger Tracker is the second component of the LHCb tracker after the VELO. It is located just before the magnet, downstream the RICH1 detector, and it is made up of silicon microstrip sensors with an homogeneous strip pitch of 198 μm. It is a 150 cm wide and 130 cm high planar station covering the full LHCb acceptance with a total active area of around 8.4 m^2 composed of four detection layers in a special $x - u - v - x$ arrangement with vertical strips in the x layers and strips tilted by -5° and +5° in the u and v layers respectively [46]. The design was chosen in order to improve the spatial resolution in the y component. Figure 3.7 shows the layout of the four TT layers.

As can be seen in figure 3.7, the layers are divided into separated read-out sectors with increasing read-out strip length as one goes farther away from the beamline. Such an arrangement allows to keep the maximum read-out strip occupancy at the level of a few percent while minimizing the number of read-out channels reducing the overall cost of the detector [46].

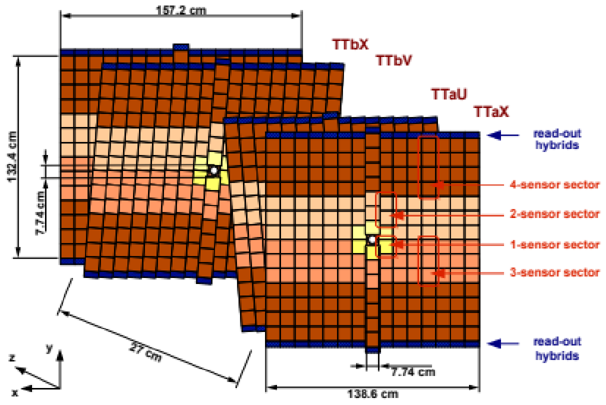


Figure 3.7: Layout of the four TT detection layers. The four different read-out sectors are indicated in different colors. Taken from [52].

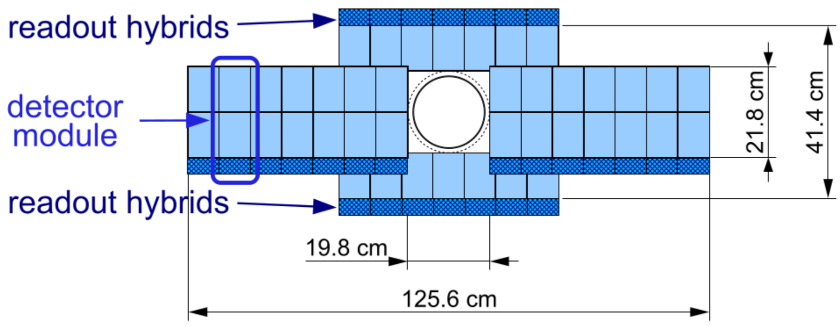


Figure 3.8: Layout of one of the IT detection layers around the beampipe. Taken from [46].

3.2.3 The Tracking Stations

The last part of the tracking system, located downstream the magnet, are the three tracking stations T1-T3. Each station contains two subdetectors with different detector technologies. the region closer to the beamline uses silicon microstrip sensors and it is called the Inner Tracker (IT). The outer region uses straw tubes and it is called the Outer Tracker (OT). The choice of the different detector technologies is motivated by the different expected particle densities and rates in the two detector regions. In the region closer to the beampipe a more granular detector is needed in order to keep a manageable detector occupancy whereas in the outer region the use of straw tubes helps reducing the overall cost and material budget of the detector. As they are located behind the magnet, charged particles are required to carry a minimum momentum of at least 1.5 GeV/c to reach the tracking stations.

The IT uses the same detector technology as the TT, it covers a 120 cm wide and 40 cm high cross-shaped region in the center of the tracking stations. Each IT station contains 4 detector layers disposed in the same way as in the TT. The layout of the IT modules is shown in figure 3.8.

The OT is a drift-time detector, charged particles traversing the straw tubes will ionise the gas and their trajectories will be reconstructed from the drift-times of the ionisation electrons moving towards the wires, measured with respect to the beam crossing signal. The maximum drift time in the straw tubes is about 35 ns but to account for variations in the drift-time and the signal

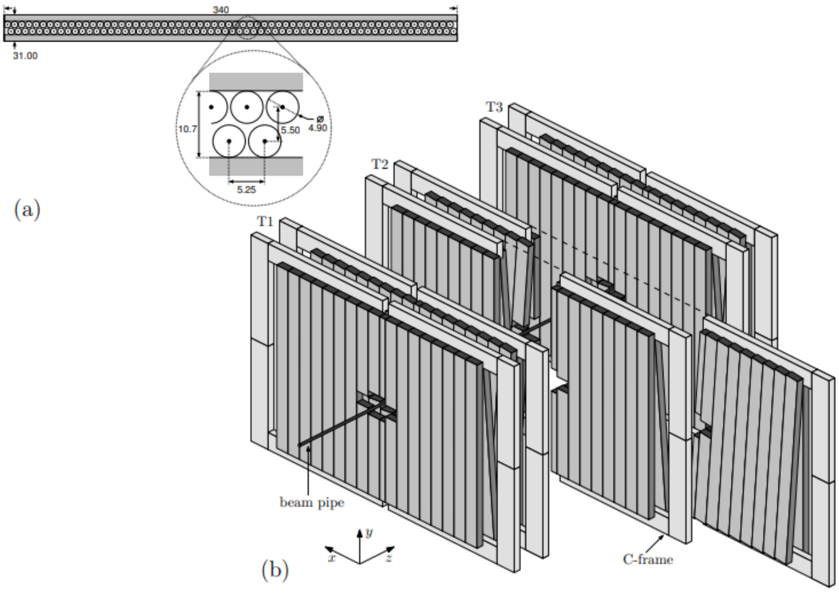


Figure 3.9: (a) Cross-section of one of the modules with distances in mm. (b) Arrangement of the modules in retractable stations. The cross-shaped hole in the center is occupied by the IT modules. Taken from [53].

propagation time through the wire the detector is read-out every 75 ns, corresponding to three bunch crossings given the 25 ns LHC rate. As a consequence of this, straw tubes show an average occupancy of about 1.8 overlapping events in 2012 [45].

The arrangement of the OT modules is shown in figure 3.9. Each of the modules contains two staggered layers of drift tubes covering an angular acceptance of 300 mrad in the horizontal (bending) plane and 250 in the vertical (non-bending) plane.

3.2.4 The RICH detectors

Particle identification is a central part in almost all the studies performed at LHCb and it will also play a very important role in the presented analysis. One of the experimental ways to get information about the identity of a particle is combining measurements of its velocity and momentum. This allows us to calculate its mass from:

$$p = \gamma m \beta c \quad \text{with} \quad \gamma = \frac{1}{\sqrt{1-\beta^2}} \quad , \quad \beta = \frac{v}{c}. \tag{3.2}$$

The approach taken in LHCb is to use the Cherenkov light emitted by charged particles when they travel faster than light in a given medium. Particles travelling inside a medium with refractive index n at a speed $\beta > 1/n$ will emit photons within an angle

$$\cos\theta_C = \frac{1}{n\beta}, \tag{3.3}$$

where θ_C is measured with respect to the particle's direction of motion. The experimental signature is then a photon ring reconstructed in the RICH detectors. The radius of the ring provides information about the particle's velocity.

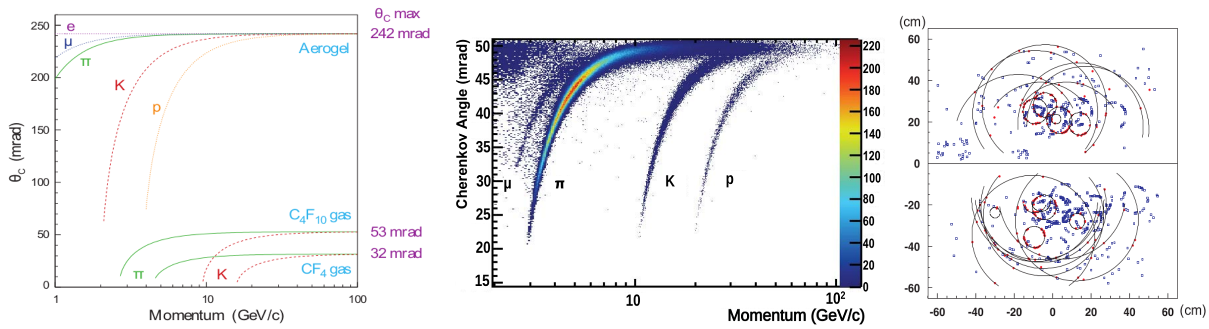


Figure 3.10: (left) Cherenkov angle versus particle momentum for the RICH1 and RICH2 radiator materials [46]. (center) Reconstructed Cherenkov angle as a function of momentum using the C_4F_{10} RICH1 radiator [45]. (right) Event display in RICH1. The smaller rings are created by the C_4F_{10} radiator whereas the largest rings are generated by light emitted when the charged particles traverse the aerogel radiator [46].

Charged hadron identification is achieved in LHCb by the use of two Ring Imaging Cherenkov (RICH1 and RICH2) detectors located upstream and downstream of the magnet and read-out by Hybrid Photon Detectors (HPDs). Together, they cover the 2-100 GeV/c momentum range allowing for the identification of protons, pions and kaons (mostly) which are the final state particles of most B and D meson decays. The correct identification of π and K mesons is particularly relevant in this thesis since the K^{*0} from the $B^0 \to K^{*0}l^+l^-$ decay is reconstructed using its hadronic decay $K^{*0} \to K^+\pi^-$. The RICH1 detector covers the low momentum charged particle range from about ~1-60 GeV/c using aerogel and C_4F_{10} as radiators[2]. It is located upstream of the magnet covering the full LHCb acceptance. Downstream of the magnet we find the RICH2 detector, covering the high momentum range ~15-100 GeV/c and using a single CF_4 radiator covering a limited acceptance of $\sim \pm 15$ mrad to ± 120 mrad in the horizontal plane and ± 100 mrad in the vertical plane. The selection of the radiation materials is motivated by the momentum range aimed at by the two detectors and it is determined by their refraction index n. Figure 3.10 (left) shows a comparison between the performance of different radiator materials. The lower bound in the momentum coverage is given by the requirement $\beta > 1/n$, in order to emit Cherenkov light. the upper bound appears due to the loss of resolution when the angle approaches its maximum value $\theta_{max} = \arccos(1/n)$ as shown in figure 3.10 (center).

In both RICH detectors, Cherenkov light is focused with a combination of spherical and flat mirrors to guide the light towards the HPDs. Figure 3.11 shows the layout of the RICH1 detector, the layout of the RICH2 detector is similar but it is mounted in an horizontal arrangement [46].

3.2.5 The calorimeter system

The LHCb calorimeter system has four components, each of them designed for a specific task. They are listed below in order, from the closest to the furthest away from the interaction point (IP):

- **Scintillating Pad Detector (SPD):** first calorimeter layer located downstream of the RICH2 detector and the first muon station. It improves the separation between electrons and photons and plays an important role in the first stage of the LHCb trigger, which rejects high multiplicity events by imposing a cut on the number of SPD hits.

[2]In a Cherenkov detector the radiator is the material where the charged particles emit their Cherenkov light.

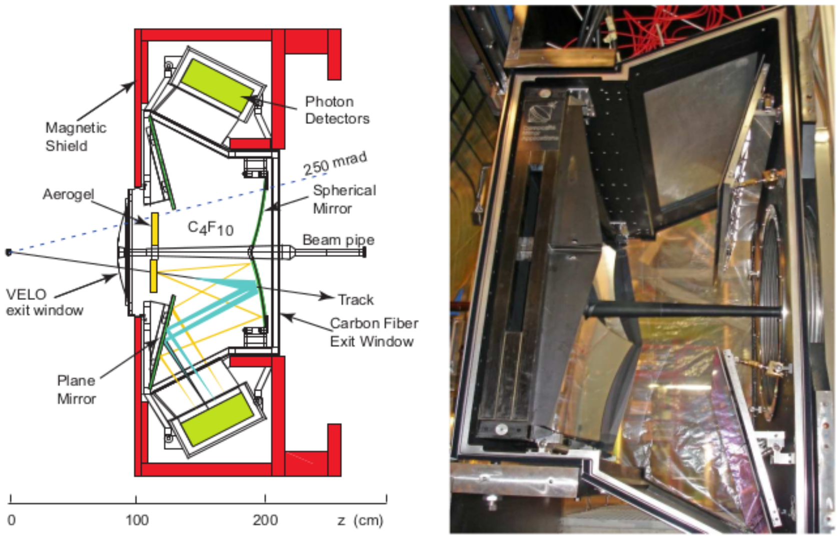


Figure 3.11: (left) Side view of the RICH1 detector layout. The detector is surrounded by an iron shield to protect it from the magnetic field created by the spectrometer magnet. (right) Photo of the detector with the interaction point in the right hand side. Taken from [46].

- **Preshower (PS):** it is located after a 15mm thick lead converter downstream of the SPD and it has the main role of suppressing the background from charged pions. It does it by measuring the longitudinal partitioning of the electromagnetic showers in the PS and the ECAL which has a different topology for pions and electrons or photons.

- **Electromagnetic Calorimeter (ECAL):** a classical sampling calorimeter with shashlik arrangement using lead as the absorber and a plastic scintillator as the detector material. It provides a measurement of the electron and photon energies, which are usually fully absorbed in the ECAL volume. This energy measurement is then used to separate electrons and pions as part of the particle identification algorithm and in the hardware trigger, see sections 3.3 and 3.4.

- **Hadronic calorimeter (HCAL):** a sampling calorimeter made from iron and scintillating tiles as absorber and active material. Its main purpose is to provide energy measurements and identification of hadrons.

Figure 3.12 shows a sketch with the four calorimeter layers overlaid with drawings representing the typical energy deposition patterns of different particles.

All calorimeters are arranged following a pojective geometry, meaning that their transverse dimensions scale with the distance from the interaction point. They adopt a variable lateral segmentation to match the variation in hit density, which drops by about two orders of magnitude from the inner part towards the edge of the calorimeters surfaces. The produced scintillation light is read-out by wavelength-shifting optical fibres and transmitted to photomultipliers. The dimensions of the different calorimeter components are chosen to match the required length for the longitudinal and lateral development of showers generated by different particles as shown in figure 3.12. The length of the HCAL is limited by the space constraints in the LHCb cavern. [46]

3.2.6 The muon stations

The muon system provides fast signals containing information used in the muon trigger as well as muon identification. It is composed of five rectangular shaped stations (M1-M5) placed downstream the calorimeters and equipped with Multi Wire Proportional Chambers (MWPC), sandwiched between 80 cm thick iron absorbers. They use rectangular pads to read-out the wires providing space-point information of the muon tracks. The pads have variable size, being largest in the outermost regions given the lower channel occupancy and the limited spatial resolution due to multiple scattering which dominates at large angles and low momenta. Stations M1-M3 have a smaller lateral segmentation and they are used to define the track direction and calculate the transverse momentum, p_T, of muons used as part of the hardware trigger.

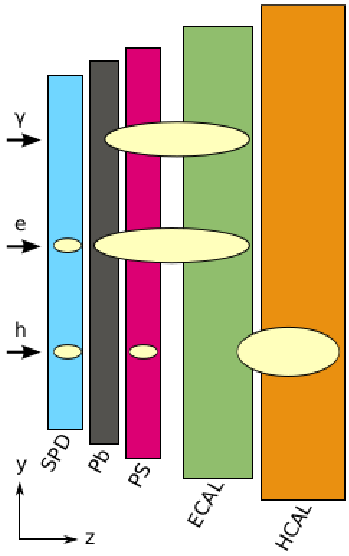


Figure 3.12: Side view of the calorimeters showing the different particle signatures. Taken from [54].

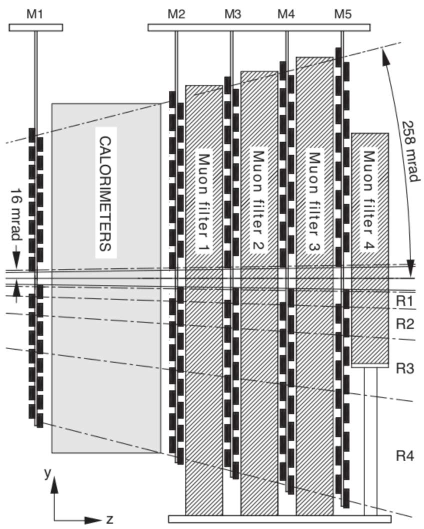


Figure 3.13: Side view of the muon system. Taken from [46].

The main purpose of stations M4 and M5 is the identification of penetrating high energy particles, their spatial resolution is therefore more limited. Station M1 is located upstream of the calorimeter system. It employs a Gas Electron Multiplier (GEM) detector technology, faster than the MWPC, to cope with the higher particle rates. It improves the p_T resolution for the trigger. All stations follow the projective geometry of the calorimeters. The minimum momentum a muon must have to be able to transverse the five stations is about 6 GeV/c. A layout of the muon system is shown in figure 3.13.

3.3 Detector performance

A detailed description of the LHCb detector components was presented in the previous section. In this section, we summarize the tracking and particle identification performance achieved by the detector.

3.3.1 Tracking

Tracking performance can be evaluated in data and MC simulations making use of decays with large branching fractions and known signatures in order to keep a low background rate. Depending on the reconstructed trajectories four different track categories are defined. They are shown in figure 3.14.

- **Long tracks:** they transverse the full tracking system. These are the highest quality tracks and the ones used in this thesis.
- **Upstream tracks:** they only have hits on the VELO and TT stations. They are mostly associated to low momentum particles which are unable to reach the T stations.
- **Downstream tracks:** they only pass through the TT and T stations. They typically correspond to the decay products of long-lived neutral particles such as K_s^0, which decay after the VELO.
- **VELO tracks:** they only leave hits on the VELO and are typically large angle or backward tracks. They can be used to improve the primary vertex reconstruction.
- **T tracks:** they are only seen in the T stations and are usually the outcome of secondary interactions.

The two results shown in this section are the mass and momentum resolution of the LHCb spectrometer as a function of the particle's energy. Both of them are evaluated using muon decays since they leave clean signatures in all the tracking stations. For the estimation of the momentum resolution $J/\psi \to \mu^+\mu^-$ decays are used. The momentum resolution as a function of the particle's momentum is shown in figure 3.15. It is on the per-cent level up to 300 GeV/c. The low momentum resolution is limited by multiple scattering whereas at large momenta the resolution decreases due to the larger uncertainty on the bending angle. The mass resolution is driven by the momentum resolution, it is evaluated from reconstructed decays of J/ψ, $\psi(2S)$, $\Upsilon(1S)$, $\Upsilon(2S)$, $\Upsilon(3S)$ and the Z^0 boson into muons. Results are also shown in figure 3.15. Again, the resolution is found to be at the per-cent level in the mass range relevant for this thesis.

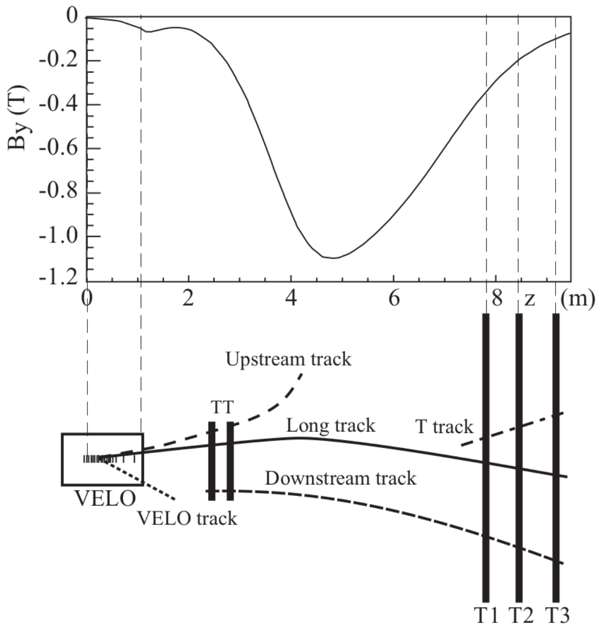


Figure 3.14: Illustration of the different track types. The B_y field component is also plotted for reference. Taken from [45].

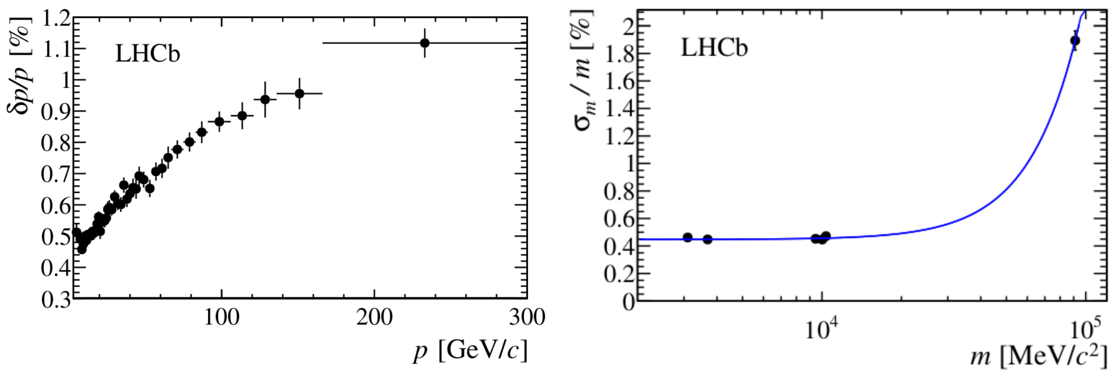


Figure 3.15: (left) Relative momentum resolution as a function of momentum for long tracks using $J/\psi \rightarrow \mu^+\mu^-$ decays. (right) Relative mass resolution as a function of the mass obtained from mass fits of dimuon resonances indicated in the text. The curve is an empirical power-law fit to the data points. Both results are obtained using LHCb Run1 data. Taken from [45].

3.3.2 Particle identification

Particle identification at LHCb is provided by the combination of information extracted from the calorimeter system, the RICH detectors, and the muon stations. Relevant to this analysis, the main role of the calorimeters is to separate the signatures of electrons and charged hadrons. This can be done by building likelihoods for the electron and hadron hypotheses based on the different shower development patterns described in section 3.2 and using tracking information combined with energy measurements from the calorimeters. The general procedure is to construct signal and background likelihood distributions using information from the ECAL, the PS and the HCAL. One of the most discriminating variables is the ratio E^{ECAL}/pc, being close to 1 for electrons which loose most of their energy in the ECAL and relatively small for hadrons, which loose most of their energy in the HCAL. The distributions of this variable for electrons and hadrons used to build the

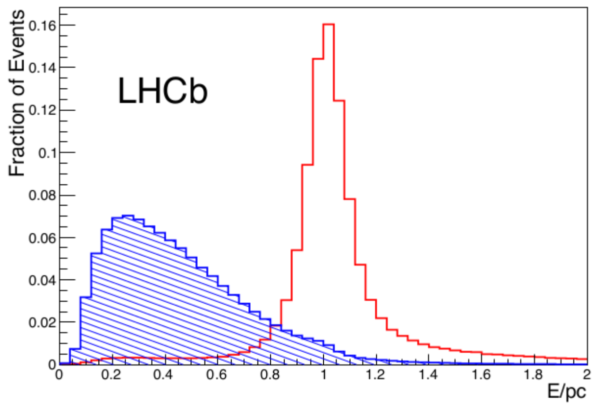


Figure 3.16: Distribution of E/pc in the ECAL for electrons (red) and hadrons (blue), obtained from data recorded in 2011. Taken from [45].

two hypotheses are shown in figure 3.16. The electron distribution is smeared due to the finite detector resolution.

The muon stations provide muon identification associating hits in the muon chambers to trajectories reconstructed by the tracking system. It also provides information about highly energetic hadrons that survive after the HCAL.

Finally, the two RICH detectors provide crucial information for the correct identification of different charged hadron species, most notably protons, kaons and pions, which are all relevant in our analysis.

Information from all subdetectors is combined into a single set of more powerful variables. The classical approach is to build individual likelihoods for the signal and background hypothesis for every subdetector. The logarithm of these likelihoods is subtracted and they are all added linearly to form a set of combined likelihoods for each signal and background pair. Charged pions are by far the largest background in almost every process studied at LHCb and $B^0 \to K^{*0}l^+l^-$ are no exception. For this reason the pion hypothesis is used as a reference point and we define the Delta Log-Likelihood (`DLL`) variables as:

$$\texttt{DLL}_{x-\pi} \equiv \Delta\log\mathcal{L}_{comb}(x-\pi). \tag{3.4}$$

They are proportional to the probability that the particle identified as x is actually x and not a pion.

A second more sophisticated approach makes use of multivariate techniques to combine information from all subdetectors into single probability values for each particle hypothesis. They define variables denoted as $\texttt{X\_ProbNN}_{\texttt{y}}$ representing the probability that a particle identified as X is y. The performance of the two PID variables is shown in figure 3.17. The improvement achieved by the multivariate classifier is clear, both of them are used to select data i this analysis.

3.4 The trigger system

A central part of the LHCb experiment is its trigger system. The trigger allows the selection of events with interesting physical signatures which are written to storage. The goal of the trigger

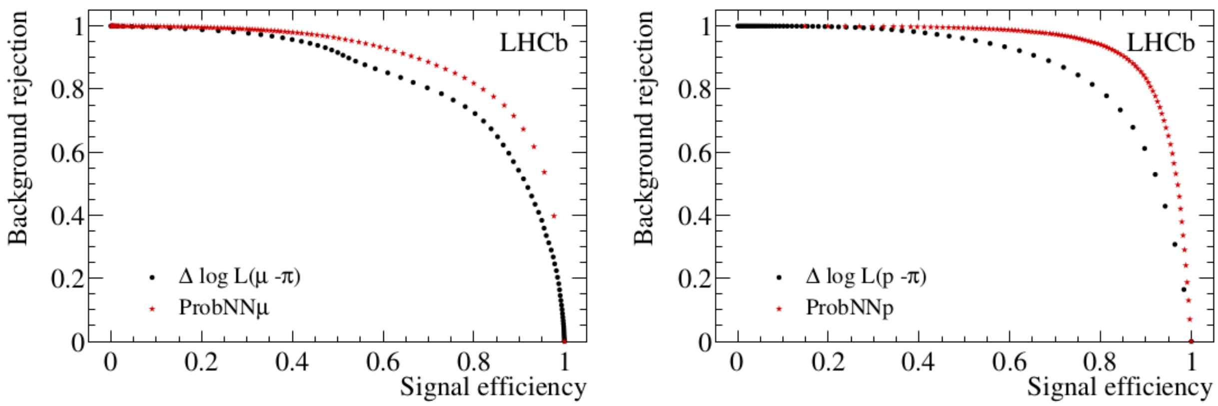


Figure 3.17: Background rejection versus muon (left) and proton (right) identification using both `PID` and `ProbNN` variables. The pion hypothesis has been taken in both cases as representative of the background. Efficiencies are measured using $\Sigma^+ \to p\mu^+\mu^-$ decays. Taken from [45].

is to provide the highest possible efficiency for events selected in the offline analysis, rejecting the large uninteresting backgrounds.

During Run 1 and Run 2, the LHCb trigger uses a combination of a hardware and a software stage in two levels: Level-0 (`L0`) trigger and High Level Trigger (HLT). The L0 trigger is implemented in hardware and employs information form the calorimeters and muon system to select events with interesting signatures. It reduces the event rate from 40MHz down to 1MHz. Events accepted by the `L0` trigger are passed to the HLT, which is a C++ application running on the Event Filter Farm (EFF), a large computer farm with a 10PB capacity, able to simultaneously run $\approx$ 50000 single threaded processes. The HLT structure and performance as well as the and `L0` thresholds are different in Run 1 and Run 2. This motivates the splitting of the data-sets into different data-taking periods with different trigger selections and potentially different background compositions.

The HLT used in Run 1 executed a simplified version of the full offline event reconstruction, making decisions based on a limited amount of information and notably lacking important inputs for particle identification coming from the RICH detectors. The main caveats were the lack of low momentum charged particles at the first stage and resolution differences between the online and offline reconstruction. This makes it more difficult to precisely compute absolute trigger efficiencies [56]. The trigger system was redesigned for Run 2 to enable a real time detector alignment and calibration and a real-time analysis in order to allow the HLT to perform an offline level full event reconstruction. A comparison between the Run 1 an Run 2 trigger schemes is shown in figure 3.18.

In both cases, the `L0` trigger employs a set of *trigger lines*[3] to select data according to different thresholds applied to information coming from the calorimeter and muon system. The `L0`-calorimeter trigger uses information from the SPD, PS, ECAL and HCAL to make a decision based on both the maximum energy deposited in 2×2 calorimeter cell clusters and the amount of hits seen in the SPD. The latter requirement reduces the event complexity to enable a faster reconstruction in the HLT. Muon triggers are also defined selecting events based on the p_T of the observed muons measured in the muon stations. Table 3.1 shows some of the `L0` trigger lines used in Run 1 and Run 2.

[3]Trigger lines are binary classification algorithms consisting on a series of cuts on selected variables of interest. A trigger consists of a set of trigger lines.

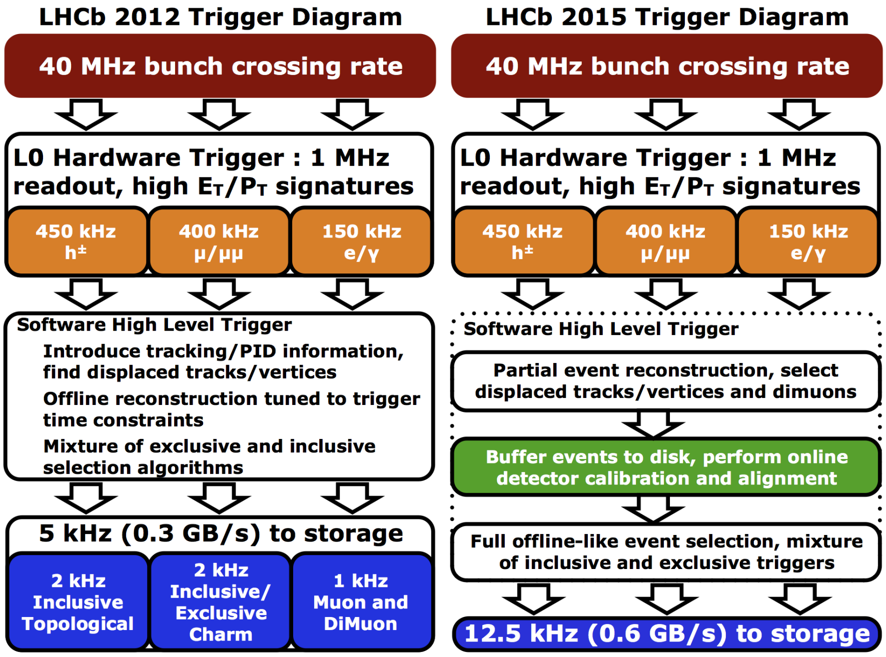


Figure 3.18: Comparison between the Run 1 (left) and the Run 2 (right) trigger schemes. Both triggers will be used in this analysis for the corresponding data-taking periods. Taken from [55].

L0 trigger	E_T or p_T threshold (lower)					SPD threshold	
	2011	2012	2015	2016	2017	Run 1	Run 2
L0Hadron	3.5 GeV	3.7 GeV	3.6 GeV	3.7 GeV	3.46 GeV	600	450
L0Photon	2.5 GeV	3.0 GeV	2.7 GeV	2.78 GeV	2.47 GeV	600	450
L0Electron	2.5 GeV	3.0 GeV	2.7 GeV	2.4 GeV	2.11 GeV	600	450
L0Muon	1.48 GeV	1.76 GeV	2.8 GeV	1.8 GeV	1.35 GeV	600	450
L0Dimuon	$(1.3\text{ GeV})^2$	$(1.6\text{ GeV})^2$	$(1.3\text{ GeV})^2$	$(1.5\text{ GeV})^2$	$(1.3\text{ GeV})^2$	900	900

Table 3.1: Some of the most commonly used L0 trigger lines showing the thresholds defined for each year. The L0Hadron, L0Photon and L0Electron lines select events based on the largest reconstructed E_T from calorimeter clusters. the L0Muon and L0Dimuon lines select events based on the largest and two largest p_T tracks in the muon stations respectively. Adapted from [56] and [45].

The HLT is divided into two stages: HLT1 and HLT2. The first one performs an inclusive selection based on information from the VELO[4] looking for displaced secondary vertices and trying to match their tracks with hits in the TT and tracking stations. It employs a forward reconstruction algorithm (faster that the backward reconstruction used offline but less efficient) and selects events based on their track signatures. Events passing the selection are buffered to disk to be further processed in the HLT2 stage. In Run 2 a real-time alignment and calibration is done before passing the events to HLT2 where a full event reconstruction is performed. In Run 1, events passing the HLT1 are directly sent to the HLT2 where an improved but still simplified reconstruction is performed. At this stage, more specific trigger lines based on the decay topology of the processes of interest are written and applied independently for each analysis.

[4]The read-out of the VELO is fast enough to be run on all events passing the L0 trigger.

Chapter 4

Analysis strategy

The final goal of the analysis presented in this thesis is to perform a measurement of $R_{K^{*0}}$ in the high-q^2 region above the two charmonium resonances, J/ψ and $\psi(2S)$, that dominate the q^2 spectrum in $B^0 \to K^{*0}l^+l^-$ decays (see figure 2.8). The general strategy is similar to the one followed in previous $R_{K^{*0}}$ analyses, performed in the q^2 region below the two resonances [11]. Data and simulation samples are the same as the ones employed in the R_X analysis, an effort obtain a simultaneous measurement of both R_K and $R_{K^{*0}}$ in the low $0.1\ \text{GeV}^2/c^4 < q^2 < 1.1\ \text{GeV}^2/c^4$ and central $1.1\ \text{GeV}^2/c^4 < q^2 < 6.0\ \text{GeV}^2/c^4$ dilepton squared invariant mass regions. We employ the R_X analysis framework and tuples and many of the data selection requirements (explained in chapter 5) and corrections applied to simulated data are inspired on the ones used in that analysis.

In the signal $B^0 \to K^{*0}l^+l^-$ decays, the K^{*0} is reconstructed from its decay $K^{*0} \to K^+\pi^-$. The Branching Fraction (BF) of this decay is close to 100% and both the K^+ and the π^- experimental signatures are well identified in LHCb. In the whole analysis no distinction is made between CP conjugates; everything said about $B^0 \to K^{*0}l^+l^-$ applies in the same way to the CP conjugated process $\overline{B}^0 \to \overline{K}^{*0}l^-l^+$. Decays where the e^+e^- pair comes from a charmonium resonance (J/ψ or $\psi(2S)$) are extensively be used through this thesis. They are sometimes referred to as *resonant channels*. They provide a way to cross-check the analysis strategy in q^2 regions where the amount of statistics in increased by up to 3 orders of magnitude due to their much larger branching fractions and can be used as normalization channels in the final result. Using the J/ψ resonant channel as normalization the $R_{K^{*0}}$ observable can be expressed as a double ratio in the following way:

$$R_{K^{*0}} = \frac{\mathcal{N}_{B^0\to K^{*0}\mu^+\mu^-}}{\mathcal{N}_{B^0\to K^{*0}J/\psi(\to\mu^+\mu^-)}} \cdot \frac{\mathcal{N}_{B^0\to K^{*0}J/\psi(\to e^+e^-)}}{\mathcal{N}_{B^0\to K^{*0}e^+e^-}} \cdot \frac{\varepsilon_{B^0\to K^{*0}J/\psi(\to\mu^+\mu^-)}}{\varepsilon_{B^0\to K^{*0}\mu^+\mu^-}} \cdot \frac{\varepsilon_{B^0\to K^{*0}e^+e^-}}{\varepsilon_{B^0\to K^{*0}J/\psi(\to e^+e^-)}}. \tag{4.1}$$

This way of expressing $R_{K^{*0}}$ is experimentally more robust than just defining it as the ratio between the muon and electron yields and efficiencies. The reason is that many sources of systematic uncertainties, mostly coming from the computation of the efficiencies, cancel out. Equation 4.1 is obtained by dividing the muon to electron ratio by the analogous ratio in the J/ψ resonant region. This is possible because the corresponding ratios for the two charmonium resonances, conventionally defined as:

$$r_{J/\psi}^{K^{*0}} = \frac{\mathcal{N}_{B^0\to K^{*0}J/\psi(\to\mu^+\mu^-)}}{\mathcal{N}_{B^0\to K^{*0}J/\psi(\to e^+e^-)}} \cdot \frac{\varepsilon_{B^0\to K^{*0}J/\psi(\to e^+e^-)}}{\varepsilon_{B^0\to K^{*0}J/\psi(\to\mu^+\mu^-)}}, \tag{4.2}$$

$$r_{\psi(2S)}^{K^{*0}} = \frac{\mathcal{N}_{B^0\to K^{*0}\psi(2S)(\to\mu^+\mu^-)}}{\mathcal{N}_{B^0\to K^{*0}\psi(2S)(\to e^+e^-)}} \cdot \frac{\varepsilon_{B^0\to K^{*0}\psi(2S)(\to e^+e^-)}}{\varepsilon_{B^0\to K^{*0}\psi(2S)(\to\mu^+\mu^-)}}, \tag{4.3}$$

are known to respect LFU and have been measured to be very close to 1 in the past [11]. The usage of J/ψ instead of $\psi(2S)$ as a normalization channel follows the convention from previous analyses. However, it still has to be decided which one to use for the ratios defined in the high-q^2 region since the $\psi(2S)$ region is closer in terms of kinematics and has a more similar background composition so that the cancellation of systematic uncertainties could work better. The main caveats are the larger background contamination and lower statistics as compared to the J/ψ region.

Due to differences in the trigger configuration and collision energies, data is splitted into three data-taking periods, referred to as: Run1 (2011-2011), Run2p1 (2015-2016) and Run2p2 (2017-2018). The same is done with MC samples when they are available. This allows to cross-check the results as well as to have a more reliable description of the different components from MC simulations, which are produced independently to describe the different experimental conditions and trigger configuration of each data-taking period. The difference in collision energy and the configuration of the HLT between data taken during Run1 and Run2 can have an impact on the background composition of the data samples. Performing a separate analysis can help to handle these backgrounds better. This is extremely important in the high-q^2 region.

The main challenge of this analysis is to be able to control the backgrounds polluting the high-q^2 region in the electron channel. Notably, the high-q^2 region contains a large background contribution leaking from the $\psi(2S)$ region which is not present neither in the region below the resonances nor in the muon channel. Some of these backgrounds can be specially problematic since they are known to peak below the signal in the reconstructed B^0 mass distribution. If they are not properly modelled they could bias the result on the electron signal yield, extracted from a fit to the reconstructed B^0 mass in data selected in the high-q^2 region. The high-q^2 region and the $\psi(2S)$ region have a lot of common features which are exploited throughout this thesis. They are used to constrain the background yields as well as to model the shape of components employed in the fits to the B^0 mass.

4.1 Experimental challenges

One could think that a measurement of $R_{K^{*0}}$ in the high-q^2 region would follow closely the work done in the past in the region below the charmonium resonances. Some of the challenges are indeed common, however, the $\psi(2S)$ resonance being so close to the signal region and the phase space limit of the decay make the the electron fits considerably more complex. A summary of the main difficulties is listed below:

- Shapes of signal and background components are distorted by the q^2 selection, making them more difficult to model.

- The electron selection efficiency is considerably lower then the muon efficiency. This is especially problematic in the high-q^2 region as the amount of signal is scarce.

- Bremsstrahlung photons emitted by electrons are recovered by a dedicated algorithm that sometimes overestimates the electron energy. This results in $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ and $B^0 \to X\psi(2S)(\to e^+e^-)$ candidates leaking into the high-q^2 region from the $\psi(2S)$ region.

The difference in the reconstruction efficiencies is a consequence of the smaller trigger efficiency for electron candidates due to the tighter `L0` trigger requirements needed to select events with

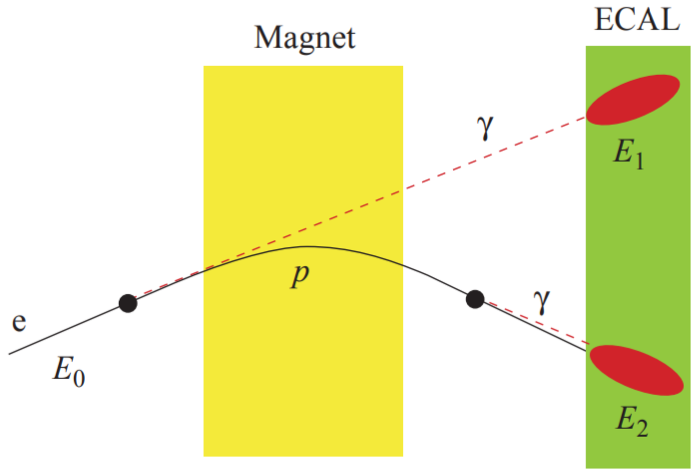


Figure 4.1: Sketch of the Bremsstrahlung recovery algorithm. Taken from [46].

electrons. As mentioned in section 3.4, electron trigger decisions are based on information coming from the calorimeters which suffer from a much larger occupancy than the muon stations. For this reason tighter trigger requirements are needed in order to keep a manageable event rate for the HLT, reducing the trigger efficiency for electrons. The trigger lines used to select data are described in chapter 5. The emission of Bremsstrahlung photons by final state electrons comprises the main difference between the reconstruction of the muon and electron channels at LHCb. We know from electromagnetism that charged particles emit radiation when accelerated by an electromagnetic field. In the LHCb detector this mostly happens when electrons interact with the material in the VELO and in the tracking stations. The emission of Bremsstrahlung photons by electrons can lead to two very different scenarios described in figure 4.1.

Bremsstrahlung photons are mostly emitted along the electron's trajectory. If a photon is emitted after the magnet it will hit the same calorimeter cells as the electron and its energy will be automatically added to the electron energy. However, if a photon is emitted before the magnet it will end up in a different calorimeter sector since, contrary to the electron, its trajectory is not deviated by the magnetic field. If the energy from these photons would not be recovered we would end up with a long tail towards lower q^2 values due to the missing energy, drastically reducing the energy resolution and selection efficiency. To avoid this, a Bremsstrahlung recovery algorithm is designed. This algorithm looks for photon clusters not associated to tracks in a calorimeter region corresponding to the extrapolated electron trajectory before the magnet. If such clusters are found, their energy is added to the reconstructed electron energy. This procedure is limited in a number of ways:

- Only clusters with $E_T > 75$ MeV are considered. This is necessary to reduce the probability of assigning random photon clusters to the electrons.

- The extrapolated electron track trajectory might fall outside of he calorimeter acceptance. If this is the case, no Bremsstrahlung photons are recovered.

- Wrongly interpreted Bremsstrahlung clusters might be added to the electrons. This can lead to an overestimation of the electron's energy.

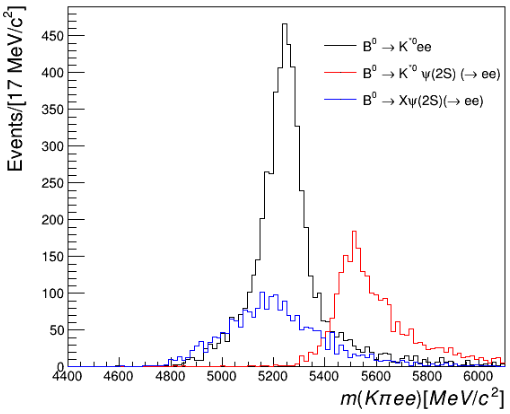


Figure 4.2: Reconstructed B^0 mass distribution for the signal and leaking background candidates selected in the high-q^2 region from MC simulated data corresponding to the Run2p1 period. The different components are indicated in the legend. Events are selected with$q^2 > 15$ GeV$^2/c^4$. Note that MC samples are not weighted so that the relative amounts of the different components do not represent the expected amounts in real data.

The first two issues limit the performance of the procedure, reducing the mass resolution int he electron channel with respect to the muon channel. The latter is the most problematic one in the high-q^2 region as it causes candidates from the $\psi(2S)$ region to leak into the high-q^2 region. This problem is enhanced by the fact that we need to cut as close as possible to the $\psi(2S)$ region in order to have enough signal owing to the reduced efficiency we have in the electron channel and the limited amount of phase space left above the $\psi(2S)$ resonance. Muons are exempt from this issue since the probability that a charged particle emits a Bremsstrahlung photon falls with the particle's mass as $P_{brem} \sim m^{-4}$. Bremsstrahlung emission is therefore neglected for the muon channel.

The reconstructed B^0 mass distribution in the high-q^2 region is shown in figure 4.2 for the signal and the two background components leaking form the $\psi(2S)$ region. It is obtained from MC simulated data and candidates are selected with $q^2 > 15$ GeV$^2/c^4$. We can see a clear peak at large masses coming from $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ events. As the energy of the electrons is overestimated by the Bremsstrahlung recovery procedure the reconstructed B^0-meson mass is also overestimated by roughly the same amount. This results in a background component peaking above the B^0 mass. More problematic are partially reconstructed $B^0 \to X\psi(2S)(\to e^+e^-)$ candidates, where X represents some excited hadronic state decaying into a $K\pi$ pair and "something else": $X \to YK\pi$. The $K\pi$ pair can either come from a K^{*0} or not and Y represents some missing particle or particles not reconstructed from the decay chain. In most cases it is a π^0 [1]. Here we have an interplay between the overestimated energy of the electrons and the missing energy from the not reconstructed particles which makes this component peak near the B^0 mass right below the signal. This is the most dangerous background in the high-q^2 region since it can very easily be absorbed by the signal component in the fits to the reconstructed B^0 mass performed to extract the signal yield. On the other hand it is an inclusive process and therefore it is less well modelled in MC simulations.

A careful strategy is developed to suppress the contribution of backgrounds leaking from the $\psi(2S)$ region and constrain their yields, it is explained in the next section.

[1] It can also contain processes where a highly excited charmonium resonance decays into a $\psi(2S)$ ad the remaining part of the decay is missed. However, these processes are highly suppressed for the $\psi(2S)$ resonance since all excited charmonium states above $\psi(2S)$ lay above the open charm threshold and their decay widths are dominated by channels with D mesons in the final state.

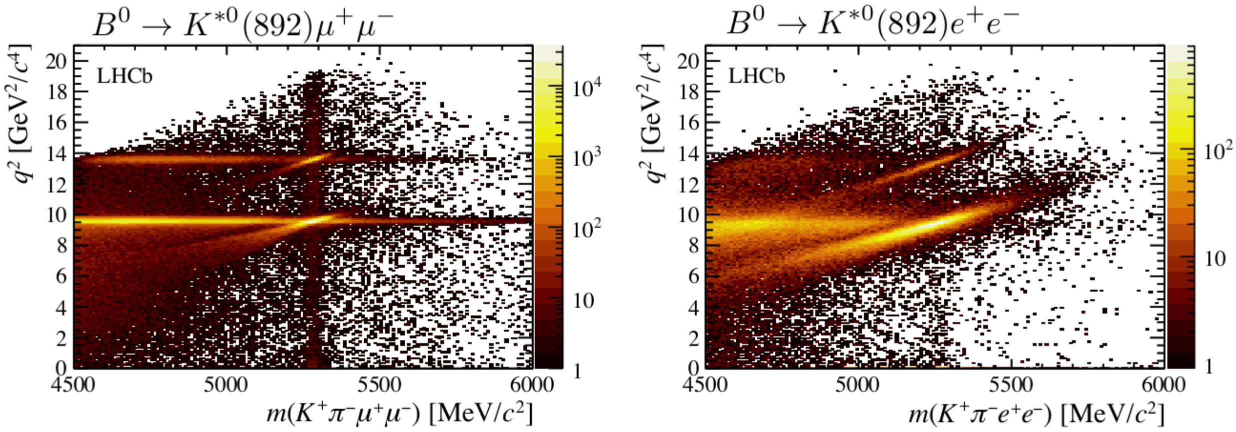


Figure 4.3: Number of $B^0 \to K^{*0}l^+l^-$ final states for the (left) muon and (right) electron channels as a function of q^2 and the reconstructed B^0 mass in Run1 data. Taken from [11]

The last issue is related to the q^2 cut needed to select the high q^2 region. It can be better understood from the plot shown in figure 4.3. There we can see a 2D distribution of candidates in the q^2 vs B^0 reconstructed invariant mass plane. The upper left triangle in both figures corresponds to the unphysical region where the reconstructed $m(K\pi l^+l^-)$ mass would be larger than the mass of the B^0 meson due to the large lepton q^2. In both electron and muon processes we can clearly see the contribution from the two resonances, with J/ψ peaking at roughly $q^2 \approx 9.59$ GeV$^2/c^4$ and $\psi(2S)$ at $q^2 \approx 13.59$ GeV$^2/c^4$. Candidates at lower masses are associated to combinatorial and partially reconstructed backgrounds. The plot highlights the differences between the muon and the electron channels. The number of events is larger in the muon channel and the resolution is better. In particular we can see a clear vertical band centered at the B^0 mass distributed roughly uniformly in the whole q^2 spectrum in the muon channel. These are signal candidates and their contribution is washed out in the electron channel due to the lower mass resolution.

The presence of the empty triangle at high-q^2 values distorts the shape of backgrounds populating the low mass region, as well as the shape signal tail on the left hand side. This is not a problem for the analysis in the q^2 region below the charmonium resonances since there the unphysical region starts way below the mass region of interest. In the high-q^2 region electron mass fits need to be extended down to, at least, 4600 MeV/c^2 in order to include all the signal due to the limited resolution so that the edge of the physical spectrum plays a more significant role. This mostly affects affects the shape of the combinatorial background mass distribution, for which a dedicated treatment will be needed (see section 4.2). Again, this effect is not relevant for muons which do not require such a large mass range to accommodate all the signal.

4.2 Strategy for the fits to the reconstructed B^0 mass

This section summarizes the common features of the strategy followed to perform the fits to the reconstructed B^0 mass presented in this thesis. Two advanced techniques extensively used in many of the fits are also explained in detail.

First of all, the shapes of the mass distributions for all the background components are either fixed from fits performed using MC simulated data or constrained, when possible, for the combinatorial background component. The shape of the combinatorial background mass distribution is derived in a data driven way, using data selected by requiring the two final electrons to have the same charge. We refer to it as *Same Sign* (SS) data, more details about it are given in chapter 5. When enough statistics in the SS data sample are available, the shape parameters of the function employed to model the mass distribution of the combinatorial background component are constrained from a mass fit performed in SS data. This type of parameter constraints are extensively used in all fits performed in this thesis. They allow us to improve the fit algorithm by adding extra information about the parameter of interest while keeping the fit flexible enough to accommodate the data. They are extremely useful for fits having a large number of parameters, improving the convergence of the likelihood minimization and stabilizing the fit algorithm. A parameter is constrained by multiplying the likelihood by a penalty function. The fitting algorithm will try to maximize the likelihood function[2] finding the optimal values of the parameters until convergence is reached. The penalty function reduces the value of the likelihood when the value of the parameter is far away from the constrained value. In this thesis we employ gaussian constrains defining the penalty function as a gaussian distribution centered at the constrained value and with a standard deviation given by the uncertainty of the constrained parameter. For the fits performed in the J/ψ and $\psi(2S)$ resonant regions the yields of backgrounds coming from known decays are also constrained with respect to the yield fitted for the signal component using the following formula:

$$\mathcal{N}_B = \mathcal{N}_S \cdot \frac{f_B \mathcal{B}(B)}{f_S \mathcal{B}(S)} \cdot \frac{\varepsilon_B}{\varepsilon_S}. \tag{4.4}$$

These type of backgrounds are known as *exclusive backgrounds*. In equation 4.4 B represents the exclusive background component and S represents the signal in the fit, either $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ or $B^0 \to K^{*0}J/\psi(\to e^+e^-)$, depending on the q^2 region of interest. The first factor is the signal yield result from the fit, only the central value is taken for the constrain. The second factor is the ratio of hadronization fractions, it is only present when signal and background processes are associated to the decay of different hadron species. The third factor is the ratio between the branching fractions of the two decays and the last factor is the efficiency ratio. Hadronization fractions and branching fractions are taken form the *Particle Data Group* (PDG) [4] whereas the efficiencies are computed using fully corrected MC simulations. The fit parameters are the hadronization fraction, the ratio between the branching fractions and the efficiency ratio. Therefore, the yield of the background component is replaced in the fit by up to three constrained parameters. The uncertainties on these parameters come from the experimental uncertainties in the measured hadronization fractions and branching fractions and from the uncertainties in the calculation of the efficiencies from MC simulations. This can only be done for the fits performed in the resonant regions. For the rare signal mode using the signal branching fraction to constrain the background yields would bias the result since any NP effect contributing to the signal process would not be included in the constraint. We are only allowed to do it here because we know that resonant processes do not contain any NP (at least within the current experimental sensitivity) and their branching fractions are known to a good precision.

The yield of background components from candidates leaking from other q^2 regions into the region of interest is constrained in a similar way:

$$\mathcal{N}_{fit} = \mathcal{N}_{leak} \cdot \frac{\varepsilon_{fit}}{\varepsilon_{leak}}. \tag{4.5}$$

[2] What it actually does is to minimize the negative log-likelihood function since minimization algorithms are computationally more efficient, but that does not affect the discussion.

This is the case of $B^0 \to K^{*0}J/\psi(\to e^+e^-)$ and $B \to XJ/\psi(\to e^+e^-)$ candidates leaking into the $\psi(2S)$ region from the J/ψ region and $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ and $B^0 \to X\psi(2S)(\to e^+e^-)$ candidates leaking into the high-q^2 region from the $\psi(2S)$ region. In equation 4.5 *fit* refers to the q^2 region where the candidates are leaking into, this is the region where the mass fits are performed. The sub-index *leak* refers to the q^2 region where the candidates are coming from.

In all the fits, the shape of the signal mass distribution is modelled from mass fits performed using MC simulated data. Data is splitted into three Bremsstrahlung categories:

- **Brem 0G:** no Bremsstrahlung photons added to the electrons.
- **Brem 1G:** one photon added to either of the electrons.
- **Brem 2G:** two or more Bremsstrahlung photons added.

This approach has proven to improve the modelling of the signal since each of the Bremsstrahlung categories contains different effects associated to the imperfect Bremsstrahlung recovery algorithm. In the fits to data the three fit functions are added in a weighted sum. The weights are calculated using fully corrected MC simulations.

Most of the fits presented in this thesis are performed using the R_X analysis framework, based on CERN's `ROOT` software, and the package *RooFit* to perform unbinned maximum likelihood fits.

4.2.1 Constrained kinematic fits

The mass resolution can be improved by taking into account the known topology of the decays as well as energy and momentum conservation in each of the decay vertices to "correct" the kinematics of the final particles using information from the decay tree. This is done by the `DecayTreeFitter` algorithm [57], originally developed for the BaBar experiment and used in many LHCb analyses. The algorithm takes the complete decay chain and parameterises it in terms of vertex positions, decay lengths and momenta of the particles. Then it performs a simultaneous fit constraining all these parameters by imposing energy and momentum conservation to recalculate the reconstructed invariant mass. Figure 4.4 shows the effect of applying the kinematic constraint to the reconstructed B^0 mass using the decay $B^0 \to K^{*0}J/\psi(\to e^+e^-)$ as a reference.

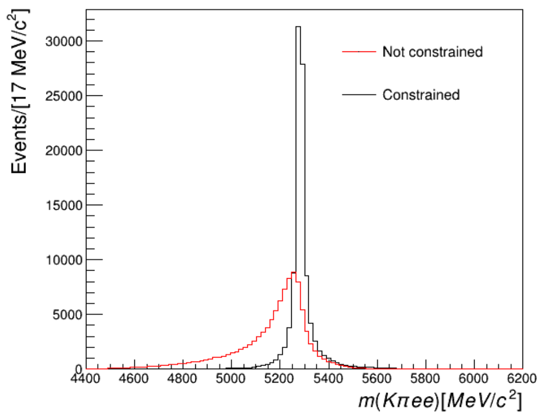


Figure 4.4: Reconstructed invariant mass of the B^0 meson from simulated $B^0 \to K^{*0}J/\psi(\to e^+e^-)$ candidates before and after applying the kinematic constraint.

In this case the kinematic constraint corrects the trajectories of the final state particles so that they match the reconstructed decay vertex of the B^0 meson. Similarly, the energy and momentum of the electrons are corrected by requiring the invariant mass of the electron pair to match the J/ψ mass. The main correction comes from the modification of the invariant mass of the e^+e^- pair as this procedure allows to partially mitigate the effects of the imperfect Bremsstrahlung recovery. Constraining the kinematics of the decay we get a clear improvement in the mass resolution and the separation of the background and signal components. Fits to the kinematically constrained B^0 mass are therefore more robust and and the results are more reliable.

Unfortunately, the effectiveness of the kinematic constrain is smaller when there is no intermediate resonance involved as the mass of this resonance can not be used to constrain the reconstructed invariant mass of the electrons. This is the case for the high-q^2 region.

4.2.2 Kernel Density Estimation

Very often along this thesis we have to deal with fit components that present a complicated shape, difficult to model using a conventional *Probability Distribution Function* (PDF). They would require the addition of many parameters to the fit function making the algorithm potentially unstable or non converging at all. In many cases, these components are modelled from MC datasets with low statistics due to the data selection, making it even harder to fit a conventional function. For these scenarios, the *Kernel Density Estimation* (KDE) [74] technique comes in handy allowing us to approximate the PDF to describe the dataset.

A KDE is a non-parametric method which approximates the shape of the underlying probability density of a finite dataset using information from the distribution of measured values. The two main ingredients of the algorithm are the *kernel* and the *bandwidth*. The kernel is the basic function used to approximate the distribution, once the kernel is given, the algorithm models the PDF of the dataset as a superposition of kernels calculated for each of the data points. Different kernels are better suited to different applications depending on the expected shape of the distribution. The most flexible one and the one used in this thesis is the gaussian kernel. The bandwidth parameter controls the width of the kernel function. Larger bandwidths generally favour smoothness over detail preservation and usually work better for smaller datasets where large statistical fluctuations which do not necessarily reflect the behavior of the underlying PDF can occur. On the contrary,

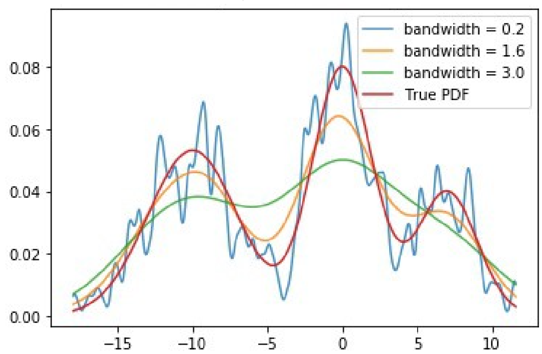


Figure 4.5: Example of the application of the KDE method to a finite dataset drawn from a known distribution. The underlying distribution is shown in red. Approximations using different bandwidths with a gaussian kernel are represented in different colors. Taken from [77].

smaller bandwidths can provide a more faithful estimation when datasets are large enough. Figure 4.5 shows a comparison of the results achieved with different bandwidths used to fit randomly generated data out of a known distribution. We can see how smaller bandwidths can generate artificial ”wiggles” reflecting statistical fluctuations on the dataset whereas too large bandwidths can miss details of the distribution.

The algorithm works by applying weights to each of the kernel estimates, calculated from the neighbouring points which are included within the kernel curve. This is done for every point in the dataset and weights are used in the linear superposition.

This method is used in the fits to the reconstructed B^0 to model partially reconstructed backgrounds which can have a complicated shape, as well as backgrounds with small contributions for which possible mismodellings of the underlying distribution have a lower on the results.

4.3 Blinding strategy

One of the main differences between this and past $R_{K^{*0}}$ analyses is the blinding strategy, motivated by the difficulties of modelling the backgrounds in the high-q^2 region. The fit to the reconstructed B^0 mass in the electron channel is blinded in a window around the B^0 mass: $[5100, 5400]$ MeV$/c^2$, where the mass of the B^0 meson is $m(B^0) = 5259.66 \pm 0.12 MeV/c^2$ [4]. In order to avoid any bias coming from the expected interplay between signal and background yields in the mass fit, this window will be kept blinded until all backgrounds are under control and the complete strategy for the analysis is agreed. The decision to blind only a relatively tight region around the B^0 mass instead of fully blinding the mass fit is motivated by the fact that the side-bands of the fit are needed in order to check the modelling of the backgrounds. These backgrounds include $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ candidates leaking from the $\psi(2S)$ region, populating the right hand side of the mass peak, the partially reconstructed $B^0 \to Xe^+e^-$ background, populating the left side-band, and the combinatorial background, which extends over the full mass range. This blinding is done in order to minimise any possible bias in the final result of the signal yield since most of the signal is contained inside of the blinded window.

Some of the decisions taken as part of the analysis strategy are driven by the blinding of the B^0 mass region in the fits performed at high-q^2 values.

Chapter 5

Data and simulation samples

This chapter is dedicated to describe in detail the data and MC simulated samples used in the analysis. In section 5.1 the strategy followed to select data is explained including the set of trigger lines, stripping cuts and the selection of same sign data. Simulated samples are described in section 5.2, the truth matching strategy is also explained in detail. Finally, in section 5.3, the three q^2 bins used in the analysis are defined with a special focus on the strategy followed to select the high-q^2 region.

5.1 Data selection

We use the full available LHCb dataset from *pp* collisions recorded in the period from 2011 to 2018, corresponding to a total integrated luminosity of about 9 fb^{-1}. Table 5.1 summarizes the integrated luminosities and center-of-mass energies corresponding to each year of data-taking.

Data is selected in three steps where selection requirements are sequentially tightened. The first step consists of choosing a trigger configuration to select the signal. After a general `L0` selection based on features of the signal candidates, a more specific set of HLT lines are designed to select events based on the kinematics and topology of the signal decay. Events passing the trigger selection are required to be triggered purely by features corresponding to the signal candidate, these are called TOS events, standing for *Trigger On Signal*. This choice makes it simpler to emulate the trigger response in MC simulations.

The next step is the stripping selection. It consists of a set of *stripping lines* which tighten the requirements imposed by the *trigger lines*. They also introduce PID information which is not available at trigger level in the Run1 configuration. The main difference is that stripping uses information from the full event reconstruction allowing to apply a more specific signal selection.

The last stage corresponds to the *offline selection* where a set of tighter requirements are set, including more specific PID criteria to suppress miss-identified backgrounds.

Year	**2011**	**2012**	**2015**	**2016**	**2017**	**2018**
$\mathcal{L}$ [fb^{-1}]	1.1	2.1	0.3	1.7	1.7	2.2
$\sqrt{s}$ [TeV]	7	8	13	13	13	13

Table 5.1: Description of dataset.

Both HLT and stripping selections are designed to rely on features of the signal candidate which are very similar to the offline selection in order to simplify the calculation of efficiencies. This is unfortunately not possible for the `L0` selection, which is not always described well in simulation. A set of corrections designed to align the `L0` trigger response in simulation to model the reconstructed data are therefore needed in order to properly compute the efficiencies.

5.1.1 Trigger selection

The `L0` trigger selection chosen for this analysis makes use of the following trigger categories:

- **L0I:** `L0Hadron_TIS` (B) || `L0Muon_TIS` (B) || `L0Electron_TIS` (B).
- **L0M:** `L0Muon` (μ_1, μ_2) && !L0I(B).
- **L0E:** `L0Electron` (e_1, e_2) && !L0I(B).

The L0I category selects events passing the trigger due to features independent of the signal. They are referred to as TIS events, standing for *Trigger Independent of Signal*. It is therefore an inclusive trigger category. The second one is the L0L category, where L denotes the flavor of the final leptons, either E for electrons or M for muons, as indicated above. This category selects events triggered by electrons or muons that were later reconstructed as part of the signal. The definition of the two trigger categories is then only possible after the full events are reconstructed and full tracking and PID information is available as only then we have the information to decide whether the particles triggering the events are part of the signal or not. The definition of the trigger lines is given in table 3.1, the cuts for 2018 are the same as those for 2017. It should be noted that even tough the L0I category selects TIS events, it does not exclude TOS. Events passing the L0I trigger selection will be either TIS or TIS&TOS where the second category refers to events which were triggered by particles belonging to the signal and independent of the signal. On the contrary, the L0L categories are exclusive categories selecting events triggered by either electron or muon signal features. They remove events that could have been triggered by features independent of the signal and ,therefore, it only contains TOS events.

In this analysis only events belonging to the L0L category will be used. The main advantage of the L0I category is that, as events are always triggered by features not belonging to the signal, its trigger efficiency is more similar between the electron and muon channels than the L0L efficiency, reducing systematic uncertainties in the final result. However, difficulties modelling backgrounds in the high-q^2 region motivate the removal of the L0I category from this analysis. As it provides an inclusive selection, it is more sensible to variations in the experimental and trigger conditions, which are different in different data-taking periods. This affects the background composition making it harder to develop a strategy to control them. On top of that, most events in the high-q^2 region belong to the L0L category so we do not give up too much statistics by removing the L0I category from the selection.

An important background source in this analysis is the mis-identified background. It includes candidates where the identities of some of the particles belonging to the B^0 decay are swapped (mostly the kaon and the pion), as well as candidates reconstructed as signal due to the wrong identification of some particle. Many of the processes contribution to this background are suppressed by the vetoes introduced in section 6.1. The choice of the L0L trigger category also improves the understanding of the efficiencies. As events are only triggered by the final leptons and mis-identified

<table>
<tr><th colspan="2"></th><th>Muon channel</th><th>Electron channel</th></tr>
<tr><td rowspan="3">Run 1</td><td>HLT1</td><td colspan="2"><code>Hlt1TrackAllL0</code></td></tr>
<tr><td rowspan="2">HLT2</td><td colspan="2"><code>Hlt2Topo[2,3]BodyBBDT</code></td></tr>
<tr><td><code>Hlt2TopoMu[2,3]BodyBBDT</code></td><td><code>Hlt2TopoE[2,3]BodyBBDT</code></td></tr>
<tr><td rowspan="3">2015</td><td>HLT1</td><td colspan="2"><code>Hlt1TrackMVA</code></td></tr>
<tr><td rowspan="2">HLT2</td><td colspan="2"><code>Hlt2Topo[2,3]Body</code></td></tr>
<tr><td><code>Hlt2TopoMu[2,3]Body</code></td><td>-</td></tr>
<tr><td rowspan="4">2016-2018</td><td>HLT1</td><td colspan="2"><code>Hlt1TrackMVA</code></td></tr>
<tr><td rowspan="3">HLT2</td><td colspan="2"><code>Hlt2Topo[2,3]Body</code></td></tr>
<tr><td><code>Hlt2TopoMu[2,3]Body</code></td><td><code>Hlt2TopoE[2,3]Body</code></td></tr>
<tr><td><code>Hlt2TopoMuMu[2,3]Body</code></td><td><code>Hlt2TopoEE[2,3]Body</code></td></tr>
</table>

Table 5.2: List of HLT lines used in the analysis.

backgrounds always involve the hadronic part, trigger and mis-identified background efficiencies can be decoupled and calculated separately.

After the first `L0` trigger selection events are passed to the HLT algorithm where further trigger requirements are applied. Table 5.2 lists the set of HLT lines used in this analysis. For data taken in Run1 the HLT1 line `Hlt1TrackAllL0` requires the existence of well reconstructed tracks and imposes a threshold on their Impact Parameter[1] (IP) with respect to any Primary Vertex (PV). The `Hlt1TrackMVA` lines used in Run2 also look for well reconstructed tracks with large IP but using a more sophisticated MVA algorithm.

In the HLT2 level, data is selected using information about the topology of the decay. The lines `Hlt2Topo[2,3]Body` and `Hlt2Topo[2,3]BodyBBDT` look for two- and three-body decay signatures originating from a displaced vertex whereas the more specific lines `Hlt2TopoMu(E)[2,3]Body` and `Hlt2TopoMu(E)[2,3]BodyBBDT` tighten the previous selection by requiring one of the tracks to be identified as a muon (electron). The more restrictive lines are `Hlt2TopoMuMu(EE)[2,3]Body`, which specifically look for two muon (electron) tracks coming from a displaced vertex in a two- or three-body decay. These are only used in data taken from 2016 to 2018.

5.1.2 Stripping and offline selection

Events passing the trigger selection are further required to fulfill a set of stripping requirements designed to reduce the background pollution of the offline data samples. The stripping lines `Bu2LLKmmLine` and `Bu2LLKeeLine2` are applied to the muon and electron modes respectively. Their content is summarized in table 5.3.

After a global selection based on the number of hits in the SPD to remove high multiplicity events, dedicated cuts on kinematic variables describing the topology of the decay together with PID cuts are applied to suppress contribution from backgrounds with similar decay kinematics that managed to pass the trigger selection.

First, the reconstructed four-body invariant mass of the B^0 meson is constrained to be within a relatively wide window around its known mass, the same is done for the reconstructed mass of the

[1] The impact parameter of a track with respect to a decay vertex is defined as the shortest distance from the extrapolated track to the vertex position.

Applied to	Requirement
Global	$\texttt{nSPDHits} < 600$
B^0	$\lvert m - m_{B^0}^{PDG}\rvert < 1500\text{MeV}/c^2$ $\texttt{DIRA} > 0.9995$ $\chi^2_{IP}(primary) < 25$ $\chi^2_{vtx}/ndf < 9$ $\chi^2_{FD} > 100$
K^{*0}	$\lvert m - m_{K^{*0}}^{PDG}\rvert < 300\text{MeV}/c^2$ $p_T > 500\text{MeV}/c$ $\chi^2_{vtx}/ndf < 25$
K	$\texttt{PID}_{K\pi} > -5$ (only data) $\chi^2_{IP}(primary) > 9$
π	$\chi^2_{IP}(primary) > 9$
ll	$m < 5500\text{MeV}/c^2$ $\chi^2_{vtx}/ndf < 9$ $\chi^2_{FD} > 16$
μ	$\chi^2_{IP}(primary) > 9$ $p_T > 300\text{MeV}/c^2$ `hasMuon == 1` (only data) `isMuon == 1` (only data)
e	$\texttt{DLL}_{e\pi} > 0$ (only data) $p_T > 300\text{MeV}/c^2$ $\chi^2_{IP}(primary) > 9$

Table 5.3: Summary of the `Bu2LLKeeLine` and `Bu2LLKmmLine` stripping requirements. The variables used for the cuts are introduced in the text. For the full set of selection cuts see documentation [58,59].

K^{*0} meson. Cuts on the transverse momentum of the K^{*0} and leptons are also defined to reduce backgrounds which tend to populate the low p_T regions.

Kinematic variables describing the decay topology are also exploited and requirements on them are applied to every particle that belongs to the process. The variable χ^2_{IP} represents the difference between the vertex fit χ^2 of a given reconstructed vertex when reconstructed with and without taking into account the selected particle's track. It is related to the IP of the particle with respect to the vertex. An upper limit on this variable is required for the reconstructed B^0 meson in order to ensure that it is produced in a primary pp interaction. A lower threshold is required for all other particles from the decay since they should be produced in the decay of the B^0 meson and not in a primary vertex. A very discriminating variable against combinatorial background is the `DIRA`, defined as the cosine of the angle between the reconstructed momentum of the B^0 meson and the vector connecting the reconstructed PV and its decay vertex. In signal events these two vectors should be completely aligned so that a tight cut on the `DIRA` provides a strong background suppression with a very high signal efficiency. Another kinematic variable of interest is χ^2_{vtx}/ndf which quantifies the quality of the fit to reconstruct the vertex. The smaller the value of this variable the better the fit result. It is applied to the vertex where the particles are produced ensuring that only events with well reconstructed vertices are kept. Last, χ^2_{FD} grows with the flight distance of a particle from its origin vertex, a tighter cut is applied to the B^0 meson because of the larger particle density close to the PV where the B^0 meson is produced.

Type	Applied to	Requirement
Global	Multiplicity	`nSPDHits` $< 600(450)$ Run1(Run2)
Fiducial calibration	All tracks	$\chi^2_{track}/ndf < 3$ $\chi^2_{IP} > 9$ `hasRich == 1` `GhostProb` < 0.4
	K, π	$p_T > 250\text{MeV}/c$ $2\text{GeV}/c < p < 200\text{GeV}/c$ `InAccMuon == 1`
	e	$p_T > 500\text{MeV}/c$ $3\text{GeV}/c < p < 200\text{GeV}/c$ `hasCalo == 1`
	μ	$p_T > 800\text{MeV}/c$ $3\text{GeV}/c < p < 200\text{GeV}/c$ `InAccMuon == 1`
Fiducial acceptance	e	$!(x_{ECAL} < 363.6\text{mm}\&\&y_{ECAL} < 283.6\text{mm})$ $\texttt{region}_{ECAL} \geq 0$ $(ee) - ECAL_{Distance} > 100$
Kinematic	K^{*0}	$\lvert m - m^{PDG}_{K^{*0}}\rvert < 100\text{MeV}/c^2$ $p_T > 500\text{MeV}/c$
Clone tracks	All tracks	$\theta(l_{1,2}, h) > 0.500\text{mrad}$ $\theta(l_1, l_2) > 0.500\text{mrad}$
PID	K	$\texttt{DLL}_{K\pi} > 0$ $\texttt{ProbNNk} \cdot (1 - \texttt{ProbNNp}) > 0.05$
	π	$\texttt{ProbNNpi} \cdot (1 - \texttt{ProbNNK}) \cdot (1 - \texttt{ProbNNp}) > 0.1$
	μ	$\texttt{ProbNNmu} > 0.2$
	e	$\texttt{DLL}_{e\pi} > 2$ $\texttt{ProbNNe} > 0.2$

Table 5.4: Offline selection requirements.

Finally, PID cuts are also included in the stripping requirements to suppress mis-identified backgrounds before the offline selection. These PID selections make use of the `DLL` variables introduced in section 3.3 for electrons and kaons. The muon PID selection is based on two binary indicators defined as `isMuon` and `hasMuon`. The `isMuon` flag indicates that the track identified as a muon has been matched to hits in the muon stations whereas `hasMuon` indicates the presence of tracks in the muon stations for the selected event. These cuts are only applied to data. Simulated events are selected using the alternative `FilterBu2LLKNoPID` stripping line, using the same selection cuts but removing the PID requirements. The PID response is not well modelled in MC simulations. They are calibrated using the `PIDCalib` package and further corrected to align the PID response so that data is properly described.

The final selection stage is the *offline selection* where the main objective is to further suppress mis-identified backgrounds by applying tighter PID cuts than the ones included in the stripping lines. The list of generic offline cuts is given in table 5.4. The cuts are classified into different categories depending on their purpose.

The *Fiducial calibration* cuts are introduced to restrict the data used in the analysis to the phase space covered by the simulated samples. Cuts on momentum and transverse momentum match the kinematic acceptance of the `PIDCalib` package. Additionally, all tracks are required to leave hits in the RICH detectors (`hasRich`), muon and hadron tracks are required to be within the acceptance of the muon stations (`InAccMuon`) and electron tracks are required to leave hits in the calorimeters (`hasCalo`). These requirements allow to separate the calculation of acceptance and selection efficiencies. The χ^2_{track} variable is a measure of the fit quality of the track reconstruction; an upper threshold ensures that only events with well reconstructed tracks are kept. Finally, the variable `GhostProb` is used to reduce the number of so-called "ghost tracks". These are fake tracks created by the combination of random hits. They are a source of combinatorial background.

The *Fiducial acceptance cuts* require electron candidates to be fully contained in the acceptance of the calorimeters and reconstructed electron pairs to have well separated clusters. These cuts ensure that the electron energy is well measured and removes acceptance effects from the reconstruction efficiency.

Some further *Kinematic* cuts on the reconstructed K^{*0} meson are applied to reduce hadronic partially reconstructed backgrounds. The K^{*0} meson is reconstructed by matching a $K^+\pi^-$ pair to the decay vertex with the invariant mass of the $K^+\pi^-$ system required to be within a window of the known K^{*0} mass. Unfortunately, this mass window also includes events where the $K^+\pi^-$ pair is produced as the result of a decay chain where some particles -neutral pions in most cases- are not reconstructed. The side-bands of the K^{*0} distribution are mostly populated by these events. Selecting a tighter window removes many of these background events without having a big impact on the signal efficiency. The effect of this selection is shown in figure 5.1, which compares the reconstructed K^{*0} mass in simulated data between the $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ and $B^0 \to X\psi(2S)(\to e^+e^-)$ samples after full offline selection. Similar arguments apply to the p_T selection as most backgrounds populate the low p_T region.

The offline selection also contains cuts to remove *Clone tracks*, defined as pairs of tracks which share at least 70% of their total hits in the tracking detectors. Most of them are removed by the `Clone Killer` algorithm, run as part of the tracking. In order to further suppress residual contributions, tracks separated by small angles are removed since most of them correspond to clone tracks. The last part of the offline selection consists of a set of PID cuts to reduce mis-identified backgrounds and background due to swaps where the identity of particles belonging to the signal process is interchanged (this mostly happens for K and π). This is achieved with the use of a combination of cuts on both `DLL` and `ProbNN` variables.

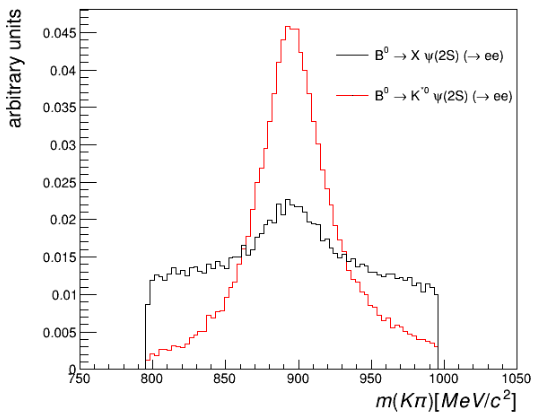


Figure 5.1: Reconstructed K^{*0} mass in simulated events after full offline selection. Samples are indicated in the legend.

5.1.3 Same sign data

Two special striping lines are employed in order to select same sign data, used in this analysis to model the shape of the combinatorial background B^0 mass distribution in the electron channel. Combinatorial background comprises candidates where at least some of the reconstructed particles attached to them are not coming from a B^0 decay. These randomly built candidates can mimic the signature of the signal passing all the trigger, stripping and offline requirements. In this analysis, the main source of combinatorial background involves random electrons and and pions which are copiously produced in the LHC, being the final states of most hadronic decay chains. SS data is selected using the `Bu2LLKmmSSLine` [60] and `Bu2KeeSSLine2` [61] stripping lines. These stripping lines are identical to the nominal ones used in data selection but require the two final leptons and/or hadrons to have the same charge. They select $B^0 \to K^+\pi^-l^+l^+$, $B^0 \to K^+\pi^+l^+l^+$ and $B^0 \to K^+\pi^+l^+l^-$ events along with their CP conjugates. Such events can only be of random nature since they all violate the conservation of electric charge. They are therefore expected to follow the same distribution as combinatorial background. The validity of this assumption is checked in chapter 7.

When examining the distribution of SS data in the J/ψ region a sharp peak in the in the mass distribution of B^0 candidates reconstructed after applying the kinematic constraint is found, located right at the B^0 mass. A similar structure is also found when the kinematic constraint is released, this time having the shape of a broad peaking structure right below the B^0 mass as can be seen in figure 5.2.

Further investigations have shown that these events originate from genuine $B^0 \to K^{*0}J/\psi(\to e^+e^-)$ decays when a hard Bremsstrahlung photon is radiated by one of the final state electrons before reaching the tracking stations. The same peaking structure is found when restripping the $B^0 \to K^{*0}J/\psi(\to e^+e^-)$ MC samples with the `Bu2KeeSSLine2` line. This hard Bremsstrahlung photon is converted into a e^+e^- pair and one of these electrons is reconstructed as part of the signal candidate. About half of the times this electron will have the same charge as the electron reconstructed from the genuine J/ψ decay and will therefore be selected by the `Bu2KeeSSLine2` stripping line. This explains why the broad peak in the right side of figure 5.2 is located below the B^0 mass as the two electrons from the reconstructed candidate are missing. The missing energy from these electrons is "recovered" by the kinematic constraint which matches the invariant mass

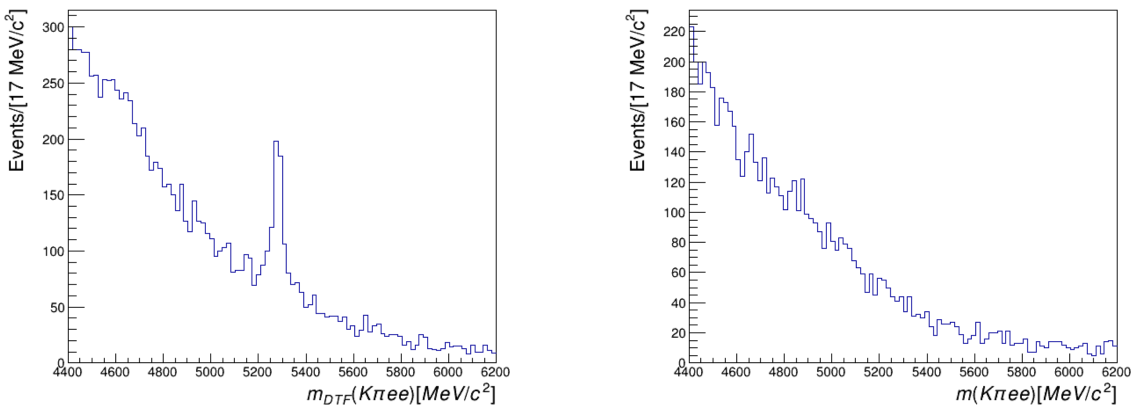


Figure 5.2: Reconstructed B^0 mass distribution of R2p2 SS data candidates selected with the `Bu2KeeSSLine2` stripping line after full offline selection in the J/ψ region. Left (right) figure shows the results with (without) the kinematic constraint.

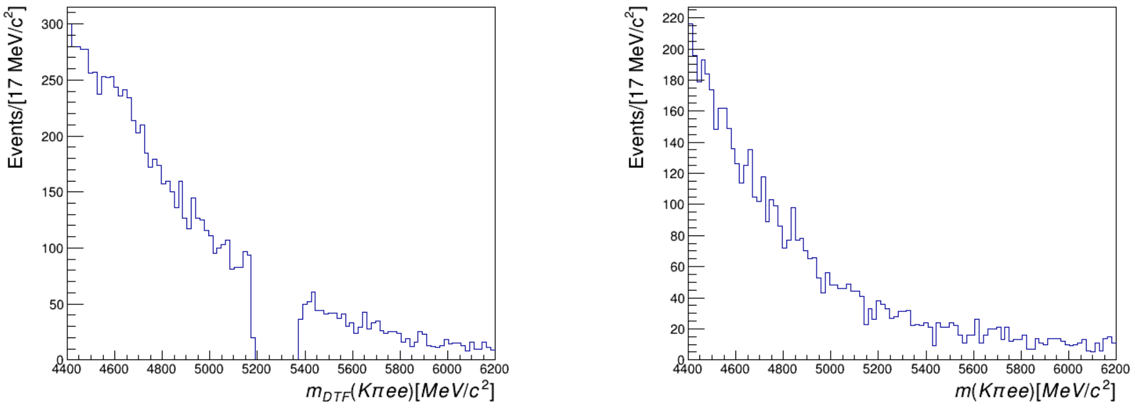


Figure 5.3: Reconstructed B^0 invariant mass distribution of R2p2 SS data candidates selected with the `Bu2KeeSSLine2` stripping line after full offline selection in the J/ψ region, including the SS veto. Left (right) figure shows the results with (without) the kinematic constrain.

of the reconstructed electrons to the J/ψ mass with the resulting peak in the constrained mass distribution. This hypothesis is also supported by the fact that such a peak is not found in the muon channel where Bremsstrahlung photon emission happens at negligible rates. The same effect is seen in the $\psi(2S)$ region although there the peak is less pronounced due to lower statistics.

In order to be able to use SS data as a proxy for combinatorial background, these genuinely signal events are removed by applying a veto on the kinematically constrained four-body B^0 invariant mass in the two resonance regions:, this is especially important when performing constrained kinematic fits:

$$|m_{B^0}^{\mathrm{DTF}} - m_{B^0}^{PDG}| > 100MeV/c^2, \tag{5.1}$$

where DTF refers to the `DecayTreeFitter` tool used to perform the kinematic constrain. Figure 5.3 shows the effect of the veto on the B^0 mass distribution for the SS sample.

No veto can be applied to remove the analogous contributions coming from rare modes. However, the effect is expected to be totally negligible is not considered in this analysis.

5.2 Simulated samples

Monte Carlo simulations are essential to obtain the shapes of the Probability Density Functions (PDF) for data fits as well as to compute efficiencies and train MVA classifiers. In LHCb, simulated samples are generated using the `GAUSS` [68] software framework. Primary pp collisions are generated using `PYTHIA8` [62] with a specific LHCb configuration. Hadronic decays are simulated with `EVTGEN`[63] which takes all available information about decay channels and branching fractions from `DecFiles` [70]. Final state radiation is generated with `PHOTOS` [66]. Finally, the propagation of particles through the detector is implemented in `GEANT4` which simulates the detector response. The full set of simulated samples used in this analysis is listed in table 5.5, including information about the available years of data-taking.

Sample	Years	Comments
$B^0 \to K^{*0}e^+e^-$	2011, 2012, 2015, 2016, 2017, 2018	
$B^+ \to K^+e^+e^-$	2011, 2012, 2015, 2016	Samples for other years are not processed with the $R_{K^{*0}}$ stripping lines from table 5.3
$B^0 \to K^{*0}J/\psi(\to e^+e^-)$	2011, 2012, 2015, 2016, 2017, 2018	
$B^0 \to XJ/\psi(\to e^+e^-)$	2017, 2018	Samples for other years are available but they were generated with a older sim version so they are not used
$B^+ \to XJ/\psi(\to e^+e^-)$	2017, 2018	Samples for other years are available but they were generated with a older sim version so they are not used
$B_s^0 \to XJ/\psi(\to e^+e^-)$	2017, 2018	Samples for other years are available but they were generated with a older sim version so they are not used
$\Lambda_b^0 \to pKJ/\psi(\to e^+e^-)$	2011, 2012, 2016, 2017, 2018	p identified as π
$B_s^0 \to \phi(\to K^+K^-)J/\psi(\to e^+e^-)$	2011, 2012, 2015, 2016, 2017, 2018	K identified as π
$B^0 \to K^{*0}(K \leftrightarrow \pi)J/\psi(\to e^+e^-)$	2011, 2012, 2015, 2016, 2017, 2018	$K \leftrightarrow \pi$ swap
$B^0 \to K^{*0}\psi(2S)(\to \pi\pi J/\psi(\to e^+e^-))$	2011, 2012, 2015, 2016, 2017	The two pions are not reconstructed
$B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$	2011, 2012, 2015, 2016, 2017, 2018	
$B^+ \to K^+\psi(2S)(\to e^+e^-)$	2011, 2012, 2015, 2016, 2017, 2018	
$B^0 \to X\psi(2S)(\to e^+e^-)$	2011, 2012, 2015, 2016, 2017, 2018	
$\Lambda_b^0 \to pK\psi(2S)(\to e^+e^-)$	2011, 2012, 2016, 2017	p identified as π
$B_s^0 \to \phi(\to K^+K^-)\psi(2S)(\to e^+e^-)$	2011, 2012, 2015, 2016, 2017	K identified as π
$B^0 \to K^{*0}(K \leftrightarrow \pi)\psi(2S)(\to e^+e^-)$	2011, 2012, 2015, 2016, 2017, 2018	$K \leftrightarrow \pi$ swap

Table 5.5: List of simulated samples used in the analysis.

The precise details about the HLT configurations are stored in *Trigger Configuration Keys* (TCK) which encode information about the configuration of trigger lines used to select the events. They tell us how often trigger lines are executed and are very useful to know what lines are responsible for the selection of events, something we need to know to properly compute the trigger efficiencies. In order to simplify MC simulations, a single TCK value is used to simulate trigger decisions on MC events being it the one for which data has the larger fraction in every year. This introduces differences between the trigger response in data and MC simulations which are corrected with a series of trigger alignment weights for the efficiency calculations.

5.2.1 Truth matching

In order to be able to use the MC samples one has to make sure that their content actually represents the process we want to simulate with them. For that purpose we can use the information about the particle identities at *generator level* before they are passed to the `GEANT4` algorithm simulating the detector response.

Signal candidates are truth-matched using the `TupleToolMCBackgroundInfo` tool which defines a set of *Background Categories* (`BKGCAT`) that classify events using detailed information about how they are generated. The tool goes through the decay chain of the simulated candidate and employs information about the `TRUEID`[2] of the particles to classify candidates as signal or background according to different predefined categories. For the truth-matching of the exclusive background samples a more traditional approach is taken. A dedicated truth-matching is applied to each of them, matching the IDs of all the particles in the decay chain to the correct ones. These IDs can be traced back through the decay chain by using variables containing information about the `TRUEID` of the mothers and grandmothers of final state particles. An exception to this applies to leaking background components corresponding to signal events from a different q^2 region, they are truth-matched as signal samples.

Partially reconstructed samples

Special care must be taken when truth-matching partially reconstructed background samples. These samples are inclusive so that we must remove components already included in other samples in order to avoid modelling the same process twice.

Starting with the $B^0 \to X\psi(2S)(\to e^+e^-)$ sample (only used for electrons), the signal contribution from $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ is removed and the electrons are required to come from a $\psi(2S)$, removing any contribution coming from the J/ψ resonance which are already modelled in the J/ψ samples. In this analysis, the $B^0 \to X\psi(2S)(\to e^+e^-)$ sample models the full partially reconstructed background contribution in the $\psi(2S)$ resonance region. It absorbs the $B^+ \to X\psi(2S)(\to e^+e^-)$ component, which is expected to follow a similar mass distribution, and the small $B_s^0 \to X\psi(2S)(\to e^+e^-)$ contribution.

For the $B \to XJ/\psi(\to e^+e^-)$ samples, leptons are required to come from a J/ψ. Depending on the flavor of the B meson responsible for the decay different exclusive processes are removed:

- $\boldsymbol{B^0}$: we remove signal $B^0 \to K^{*0}J/\psi(\to e^+e^-)$ candidates as well as the contribution from $B^0 \to K^{*0}\psi(2S)(\to XJ/\psi(\to e^+e^-))$ processes which are modeled using the dedicated $B^0 \to K^{*0}\psi(2S)(\to \pi\pi J/\psi(\to e^+e^-))$ sample. It should be noted that the removed $B^0 \to K^{*0}\psi(2S)(\to XJ/\psi(\to e^+e^-))$ candidates not only include the decay modelled in the dedicated sample. However, the $B^0 \to K^{*0}\psi(2S)(\to \pi\pi J/\psi(\to e^+e^-))$ component amounts to about 86% of the full inclusive $B^0 \to K^{*0}\psi(2S)(\to XJ/\psi(\to e^+e^-))$ set of decays. Missing components can be considered negligible and are nevertheless expected to follow a similar mass distribution.

[2] This is the ID of the particle at generator level.

- $\boldsymbol{B_s^0}$: we remove the exclusive $\overline{B}_s^0 \to K^{*0}J/\psi(\to e^+e^-)$ decay. This process is modelled independently in the mass fits, see chapter 7.

No exclusive processes are removed from the $B^+ \to XJ/\psi(\to e^+e^-)$ sample since only partially reconstructed processes can pass the full event selection.

Furthermore, a dedicated treatment is applied to the three $B \to XJ/\psi(\to e^+e^-)$ samples depending on the q^2 region of interest:

- $\boldsymbol{J/\psi}$ **region**: samples are splitted into *leptonic* and *hadronic* partially reconstructed components in order to improve the modelling of partially reconstructed backgrounds in the mass fits. The *hadronic* partially reconstructed sample is built by requiring the leptonic part to be well reconstructed, i.e. the electrons are required to come from a J/ψ and the J/ψ from the corresponding B meson, either B^0, B_s or B^+ depending on the sample. The *leptonic* component is selected by inverting the previous cut so that it contains both events with missing particles from the leptonic part but with the hadronic part being well reconstructed and processes where both the leptonic and hadronic parts are partially reconstructed.

- $\boldsymbol{\psi(2S)}$ **region**: the lower statistics does not allow for the splitting into leptonic and hadronic components and the full samples with only the general truth-matching requirements to remove exclusive processes are kept.

After truth-matching, the largest contribution to the hadronic part of the partially reconstructed decays comes form $B^0 \to K_1(1270)^0(\to K\rho)J/\psi(\to e^+e^-)$ candidates, where the ρ meson decays into a pair of pions: $\rho \to \pi\pi^0$. This decay corresponds to about 59% of the candidates included in the sample. The leptonic component is dominated by $B^0 \to K^{*0}\chi_{c1}(1P)^0(\to \gamma J/\psi)$ decays, amounting to about 47% of it. The $B^0 \to X\psi(2S)(\to e^+e^-)$ sample contains hadronic partially reconstructed processes only. Leptonic partially reconstructed candidates come from cascade decays of excited charmonium resonances ending up in $\psi(2S)$. Given that $c\bar{c}$ resonances which are more massive than $\psi(2S)$ are already above the open charm threshold, their branching fractions for decays into a $\psi(2S)$ are small and not very well known so that they are neglected in the generation of the sample.

5.3 Definition of the q^2 bins

The definition of the q^2 regions follows the strategy from previous $R_{K^{*0}}$ analyses. It is shown in table 5.6 for the two resonant channels.
Selections in the muon channel can be made tighter than in the electron channel due to the better q^2 resolution, as pointed out in section 4.1. In the electron channel, the $\psi(2S)$ region is selected in two different ways. A more conventional q^2 selection is used when performing fits to the kinematically constrained B^0 mass whereas for fits to the B^0 mass without the kinematic constraint a selection

Region	Muon channel	Electron channel
J/ψ	$\lvert m(\mu^+\mu^-) - m^{PDG}_{J/\psi}\rvert < 100$ MeV$/c^2$	$6 < q^2 < 11$ GeV$^2/c^4$
$\psi(2S)$	$\lvert m(\mu^+\mu^-) - m^{PDG}_{\psi(2S)}\rvert < 100$ MeV$/c^2$	$11 < q^2 < 15$ GeV$^2/c^4$ q2BDT$_{J/\psi} > 0.95$ && q2BDT$_{\psi(2S)} < 0.95$

Table 5.6: Definition of the q^2 selection bins for the resonant channels.

based on two MVA *Boosted Decision Tree* (BDT) variables is applied. The meaning and definition of the q2BDT variables will be given in the next section.

5.3.1 The q2BDT

As explained in section 4.1, the high-q^2 region is contaminated by a large leaking background component coming from the $\psi(2S)$ region which can be problematic in the mass fits. In an attempt to reduce it, a BDT is trained to select signal events in the high-q^2 region and reject events coming from $\psi(2S)$ decays with the goal of improving the signal-to-background ratio.

The classifier is trained using the `CATBOOST` package developed by Yandex [72] in conjunction with the *Reproducible Experiment Platform* [73]. Both signal and background training samples are fully selected and truth matched following the procedure presented in sections 5.1 and 5.2 and including the exclusive background vetoes explained in chapter 6. Signal is modelled using $B^0 \to K^{*0}e^+e^-$ simulated candidates selected with $q^2_{TRUE} > m^2(\psi(2S))$, where q^2_{TRUE} is defined at generator level and it represents the true q^2 of the electrons coming from the B^0 decays before any detector reconstruction effects. For the background $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ candidates we use MC simulated data selected with a looser $q^2_{TRUE} > 13.2$ GeV$^2/c^4$ in order to make sure all $\psi(2S)$ events are included. Figure 5.4 shows the q^2_{TRUE} distribution of signal and background samples employed in the training. As expected, the generator level q^2 distribution for the $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ process shows a very narrow peak at $m^2(\psi(2S))$, only affected by final state radiation.

Training data was split into Run1, Run2p1 and Run2p2 data-taking periods in order to account for the different trigger configurations and collision energies that could have an impact in the q2BDT performance. The training is performed following a conventional *k-fold* approach with 10 folds so that 10 different classifiers are iteratively trained for each data-taking period. Each classifiers uses 9/10ths of the events for the training and the remaining 1/10th to test the performance.

In order to keep the q2BDT as simple and robust as possible only three training features are used: the electron q^2 with and without Bremsstrahlung recovery (q^2 and q^2_{TRACK}) and the number of Bremsstrahlung photons added to the electrons (n_{brem}). The q^2 without Bremsstrahlung (q^2_{TRACK}) is calculated using information from the tracking stations only, before the Bremsstrahlung recovery algorithm is applied. The possibility of using variables related to the hadronic part of the decay

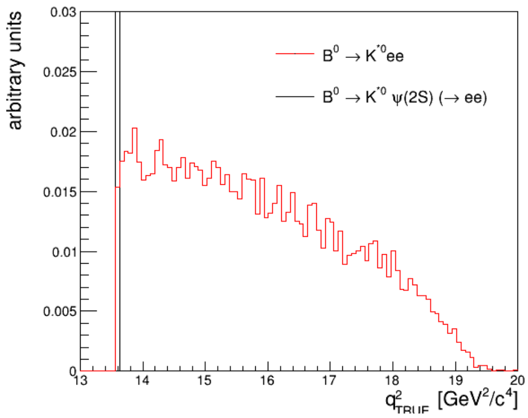


Figure 5.4: Generator level q^2 distribution for selected simulated $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ and $B^0 \to K^{*0}e^+e^-$ events used for the MVA training.

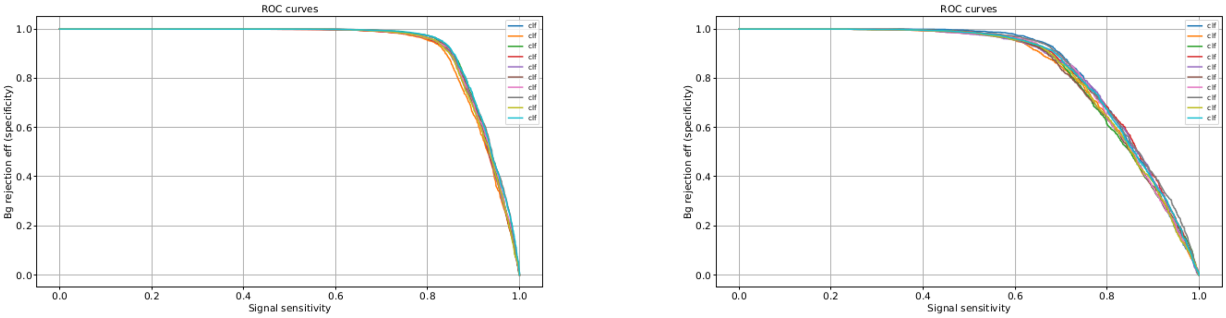


Figure 5.5: Performance of the two trained q2BDTs using Run2p2 MC samples for the validation q2-BDT$_{J/\psi}$ (left) and the nominal q2-BDT$_{\psi(2S)}$ (right). Results are shown for all 10 folds indicated with different colors. Taken from [65].

has also been investigated, however, it is discouraged by the notably different response between $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ and $B^0 \to X\psi(2S)(\to e^+e^-)$ samples, making it more difficult to understand and more sensible to mis-modellings in the inclusive $B^0 \to X\psi(2S)(\to e^+e^-)$ sample. Having it depending on electron features only, a similar performance for the two background components leaking into the high q^2 region form the $\psi(2S)$ region is expected, this is investigated in chapter 7.

In order to cross-check the q2BDT response on data, a second MVA is trained to reject J/ψ leaking backgrounds form the $\psi(2S)$ region. The training was kept as close as possible, using the same features and algorithm. Simulated $B^0 \to K^{*0}e^+e^-$ decays selected with $q^2_{TRUE} > m^2(J/\psi)$ are used as signal proxy and $B^0 \to K^{*0}J/\psi(\to e^+e^-)$ candidates with $q^2_{TRUE} > 9.2\ \text{GeV}^2/c^4$ as background. The performance achieved by the two q2BDTs is shown in figure 5.5 where the *Receiver Operating Characteristic* (ROC) curves are shown for each of the 10 folds trained for the R2p2 period. They compare the background rejection efficiency, defined as $1-\varepsilon_{bkg}$, with the signal efficiency ε_{sig}. The results from different folds are consistent with each other. A better performance is achieved by the validation q2BDT$_{J/\psi}$. This is expected since the J/ψ resonance has a narrower peak and there is much more phase space above it to separate it from the signal.

It is also important to check that there is no overtraining signs in the performance, we can do it by checking the response of the classifier in data not used for training. Overtraining plots are shown in figure 5.6 for one of the folds using Run2p2 MC simulated data. An overall agreement between the response for the training and test samples is seen with no signs of overtraining. The reason why wee need this second q2BDT classifier to test the performance of the nominal one is explained in section 7.1. In that section, the validation rsults are also presented.

In order to show the improvements achieved with this new approach for selecting data in the high-q^2 region we can plot the background against signal efficiency curve for the q2BDT and compare it with the performances of the more conventional q^2 and q^2_{TRACK} selections. Results are shown in figure 5.7 with dots indicating reference values for signal and background efficiencies used for comparison. The red dot corresponds to the efficiencies achieved with the current preliminary working point used to select the high-q^2 region: q2BDT > 0.95. Blue and green dots represent the signal and background efficiencies achieved with the q^2 and q^2_{TRACK} selections for the same background and signal efficiencies achieved with the q2BDT. The q2BDT cut is yet not optimised. The optimization will require not only maximizing the signal-to-background ratio, which would be the basic figure of merit, but also to look at how background shapes at high-q^2 are affected by the

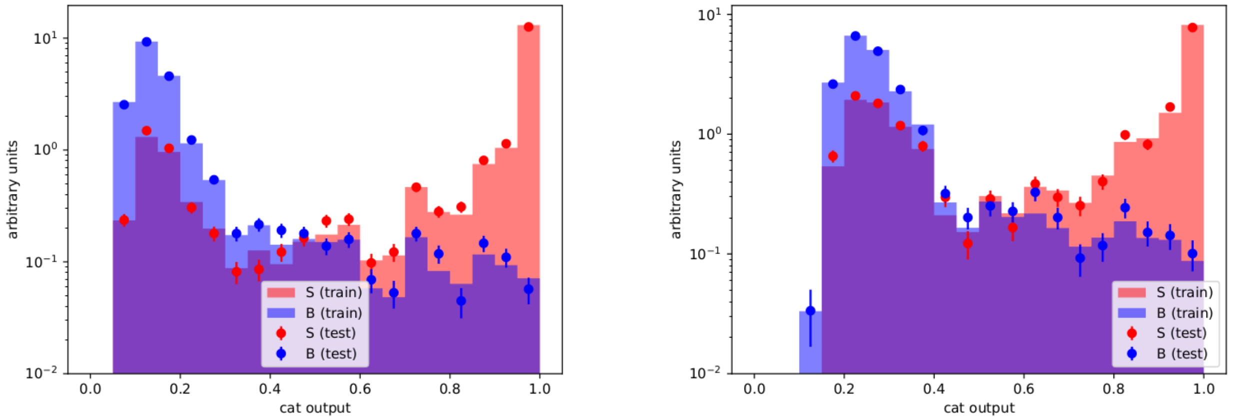


Figure 5.6: Overtraining plots for one of the folds in the Run2p2 period for the validation q2BDT$_{J/\psi}$ (left) and the nominal q2BDT$_{\psi(2S)}$ (right). Histograms represent the q2BDT distributions used for training, corresponding to 9/10ths of the data samples. Points represent the 1/10th of the data not used in that fold. Taken from [65].

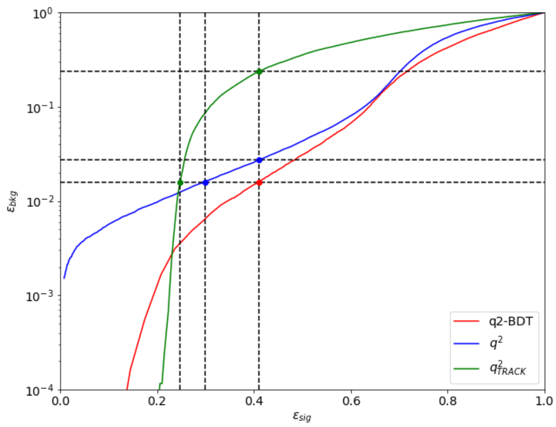
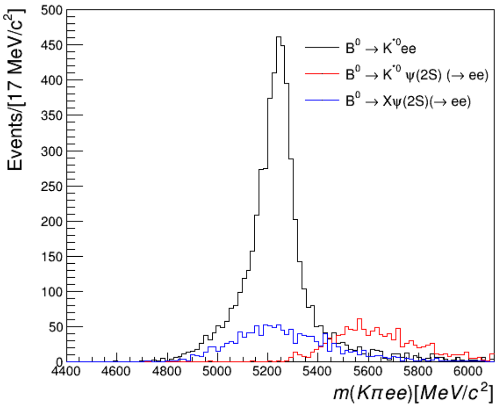


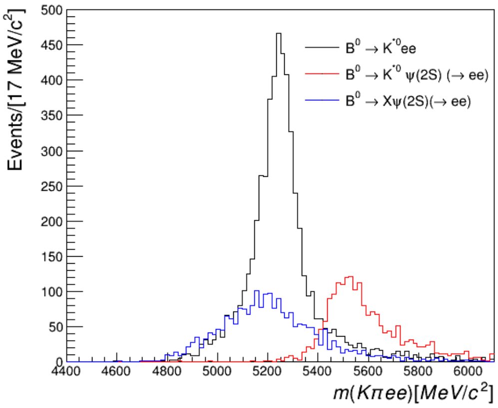
Figure 5.7: Background versus signal efficiency in the high-q^2 region for the three different q^2 selections tested. The training samples $B^0 \to K^{*0}e^+e^-$ and $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ selected with the q^2_{TRUE} cuts indicated in the text respectively represent the signal and background components. The green dot represents the current q2BDT > 0.95 working point which is compared with the q^2 and q^2_{TRACK} efficiencies for the same signal and background performance in blue and green, respectively. A logarithmic scale is used in the y axis for better comparison.

cut, which could potentially become harder to model. All the results presented in this thesis are general enough to be independent of the q2BDT cut choice. Once the cut is optimised q^2 selections should be modified accordingly but conclusions are expected to hold. With the current q2BDT working point we achieve a signal efficiency of $\sim 41.0\%$ with an efficiency for leaking $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ background of only $\sim 1.6\%$. In order to achieve the same signal efficiency using q^2- and q^2_{TRACK}-based selections we would have to allow background efficiencies of $\sim 2.7\%$ and $\sim 23.6\%$ respectively. On the other hand, for the same background rejection we would only be keeping $\sim 29.9\%$ and $\sim 24.7\%$ of the signal.

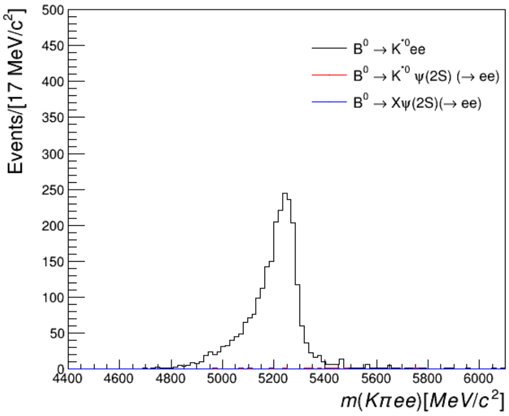
In order to compare the different approaches to select the high-q^2 region, reconstructed B^0 invariant



(a) q2BDT > 0.95



(b) $q^2 > 15\ \mathrm{GeV}^2/c^4$



(c) $q^2_{TRACK} > 14\ \mathrm{GeV}^2/c^4$

Figure 5.8: Reconstructed B^0 mass distribution in the high-q^2 from simulated data corresponding to the Run2p1 period after full selection. Different selections are indicated in the captions. Note that MC samples are not weighted so that the relative amount of each component for a particular selection might differ in real data.

mass distributions are shown in figure 5.8 using different cuts on q^2, q^2_{TRACK} and q2BDT. The cut $q^2_{TRACK} > 14\mathrm{GeV}^2/c^4$ suppresses backgrounds leaking from the $\psi(2S)$ region to a negligible level but at the cost of also reducing the signal yield by $\sim 53\%$ with respect to the MVA based selection. On the other hand, if we compare the distributions after the q^2 and q2BDT cuts we can see that both $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ and $B^0 \to X\psi(2S)(\to e^+e^-)$ leaking background components are suppressed for a similar signal efficiency if we use the q2BDT selection.

Chapter 6

Background suppression

In chapter 5 the general data selection strategy was introduced. It is based on a set of trigger, stripping and offline cuts which employ known signal features to remove backgrounds based on the known decay topology and applying PID requirements. Backgrounds surviving this selection can be suppressed by employing more specific techniques designed to tackle each of them independently. Section 6.1 introduces the different strategies followed in this analysis to suppress background contributions coming from exclusive decays which are known to pass the general selection criteria. These backgrounds are called exclusive backgrounds in the following. Combinatorial and partially reconstructed backgrounds are harder to handle since they contain contributions from many different processes. Dedicated MVAs are trained to reduce them, they are presented in section 6.2.

The strategy to suppress backgrounds is fully inherited from the R_X analysis, we use the same exclusive background vetoes as well as the MVAs trained by the R_X team. Even tough they were optimized for the low and central q^2 regions introduced in chapter 3 they show a comparably good performance in the high-q^2 region studied in this thesis.

6.1 Exclusive background vetoes

Exclusive backgrounds passing the global selection can be suppressed by applying dedicated vetoes on certain kinematic quantities for which they behave different than signal. Backgrounds considered here populate the mass region used in the analysis and have sufficiently large branching fractions to give non negligible contributions. Most of them are suppressed to negligible levels after the vetoes are applied, some still need to be considered in the fits to the reconstructed B^0 mass performed to extract the signal yield and some others will be absorbed by other fit components. In one way or another, reducing their contribution helps having a better control of them making the fits to the reconstructed B^0 mass more stable and robust.

Known exclusive backgrounds appear in three types:

- **Mis-identified backgrounds**: as mentioned in section 5.1 they are generated by a wrong identification of at least one of the final particles of the signal decay. They include candidates with a similar but different final state which can be wrongly reconstructed as signal candidates if some of the final particles are mis-identified. They can be very dangerous due to their similar decay kinematics. They populate regions close to the B^0 mass and sometimes even present peaking structures that can be easily absorbed by the signal component in the fits to the reconstructed B^0 mass if not taken into account. They are specially problematic in the rare modes as some of the candidates generating mis-identified backgrounds have branching ratios which are several orders of magnitude larger than the branching ratio of the signal decay.

- **Partially reconstructed backgrounds**: they appear as the result of decays with many particles in the final state, a subset of which are the same as in the signal decay. They can be reconstructed as signal if some of the particles which are part of the decay are not reconstructed. In principle they are less dangerous than mis-identified background since the missing energy from the lost particles makes them populate low reconstructed B^0 mass regions. They are however problematic in the electron mode due to the larger mass range needed to include all the signal candidates from the $B^0 \to K^{*0}e^+e^-$ decay.

- **Over-reconstructed background**: they are a type of combinatorial background originating when random particles flying within the detector are attached to decays in such a way that the reconstructed final state coincides with the signal final state. They are naturally suppressed by the dedicated combinatorial MVA selection introduced in section 6.2.

In the following, dedicated vetoes to suppress known background processes belonging to either of the three mentioned categories are introduced in detail.

6.1.1 Mis-identified backgrounds

Very often it is possible to veto these backgrounds modifying the mass hypothesis of particles that might have been mis-identified. This reverts the mis-identification allowing the reconstruction of intermediate resonances in the decay chain that can be removed by cuts. This is usually combined with the tightening of the PID requirements of particles participating in the miss-identification. In the following we go through the five possible types of mis-identification which lead to several known exclusive background components.

$K \to \pi$ mis-identification

The relevant processes are either resonant or non-resonant $B_s^0 \to \phi(\to K^+K^-)l^+l^-$ candidates. Due to the mass difference between the K and the π these events, which would otherwise peak at the B_s^0 mass, can peak close to the B^0 mass below the signal. This makes this background potentially dangerous. Fortunately, it is largely suppressed by the K^{*0} mass window applied in offline selection. Remaining events can be efficiently vetoed with a 2D selection in the (`Pi_ProbNNPi`, $m(KK_{\to\pi})$) plane by rejecting the red region shown in figure 6.1.

The `Pi_ProbNNpi` variable represents the probability that a π is correctly identified whereas $m(KK_{\to\pi})$ is the reconstructed $K\pi$ invariant using the K mass hypothesis for the π. The threshold in $m(KK_{\to\pi})$ rejects events for which the reconstructed invariant mass with the correct mass hypothesis is lower than $m(\phi) + 20$ MeV$/c^2$.

This component can be safely neglected in the high-q^2 region die to the low statistics. In the resonant regions it is included as an exclusive background component in the fits to the reconstructed B^0 mass. The yield of this background component is constrained with respect to the signal yield in the fits as explained in section 4.2. As shown in chapter 7, it has a residual contribution.

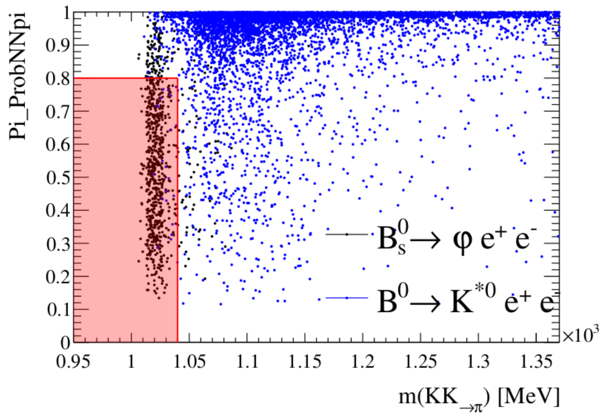


Figure 6.1: Veto against $B_s^0 \to \phi(\to K^+K^-)l^+l^-$ events. Black and blue points represent background and signal candidates from MC simulated samples fully selected in the central q^2 region. The meaning of the variables in explained in the text. Taken from [64].

$p \to \pi$ mis-identification

The known contribution to the studied decays comes from $\Lambda_b^0 \to pKl^+l^-$ candidates. It presents a broad mass peak under the signal. This background is suppressed by the offline π PID requirement and its branching fraction is small compared to the signal branching fraction. In the J/ψ region we have:

$$\mathcal{B}(\Lambda_b^0 \to pKJ/\psi(\to e^+e^-))/\mathcal{B}(B^0 \to K^{*0}J/\psi(\to e^+e^-)) \sim 0.4, \tag{6.1}$$

with similar numbers for the muon channel. This ratio is expected to be even smaller for the rare channel in the high-q^2 region. Removing events in a window around the Λ_b^0 mass after correcting the identity of the π meson does not work here since such a veto would also remove a large signal fraction so. Therefore the component is kept in the fits to the reconstructed B^0 mass. As before, it is only considered in the mass fits performed in the two resonant regions and its yield is constrained to a very small contribution.

$K \leftrightarrow \pi$ swap

This background originates from signal processes for which the identities of the K and the π are swapped. It requires a double mis-identification and it is therefore very suppressed by the PID requirements. Again, there is no efficient way to veto it and it is included in the B^0 mass fits performed int the resonant regions.

$\pi \to l$ mis-identification

Backgrounds of this type can be generated from semileptonic cascade decay candidates such as $B^0 \to \pi^-\overline{D}^0(\to K^+\pi^-)l^+\nu_l$, $B^0 \to D^{*-}(\to \pi^-\overline{D}^0(\to K^+\pi^-))l^+\nu_l$ and $B^0 \to D^-(\to \pi^-K^{*0})l^+\nu_l$. Despite being suppressed by the vertex requirements from the topological trigger and stripping lines due to the displaced D meson decay vertex and the electron PID requirements, these decays can still yield important contributions to data. This is due to their large branching fractions (up to 5 times larger than $\mathcal{B}(B^0 \to K^{*0}J/\psi(\to e^+e^-))$). Following the same strategy as for the $B_s^0 \to \phi(\to K^+K^-)l^+l^-$ candidates, semileptonic backgrounds can be suppressed by removing events within a 30 MeV/c^2 window around the D^0/D^- mass after correcting the mass hypothesis

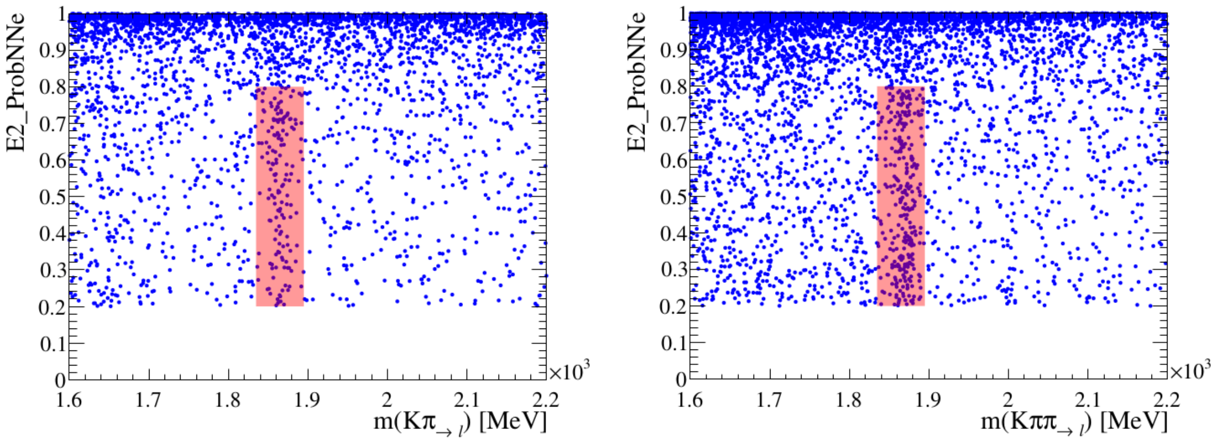


Figure 6.2: Veto against $B^0 \to D^{*-}(\to \pi^-\overline{D}^0(\to K^+\pi^-))l^+\nu_l$ and $B^0 \to \pi^-\overline{D}^0(\to K^+\pi^-)l^+\nu_l$ (left) and $B^0 \to D^-(\to \pi^-K^{*0})l^+\nu_l$ (right). Points represent 2012 data selected in the central q^2 region. Variables are explained in the text. Taken from [64].

of the leptons as shown in figure 6.2.

The `E2_ProbNNe` variable is the probability that the electron labelled as `E2` in the data tuples is correctly identified as an electron. The same distribution is observed when looking at the other electron and the veto is applied to both of them separately. As before, the variable $m(K\pi_{\to l})$ is the reconstructed Kl invariant mass where the pion mass hypothesis has been used for the lepton (an electron in this case).

This veto is applied using real data as a reference instead of MC simulations and a larger contribution from $B^0 \to D^-(\to \pi^-K^{*0})l^+\nu_l$ can be seen in the right plot as compared to the $B^0 \to \pi^-\overline{D}^0(\to K^+\pi^-)l^+\nu_l$ and $B^0 \to D^{*-}(\to \pi^-\overline{D}^0(\to K^+\pi^-))l^+\nu_l$ which both contribute to the left plot. The reason is the much larger suppression of the last two processes coming form the K^{*0} mass window which affects less the former given that it also has a K^{*0} in the final state.

In the electron channel the masses are calculated using the tracking information only, without the Bremsstrahlung photons recovered in the calorimeters, in order to avoid any biases. After the vetoes, all three backgrounds are suppressed to negligible levels and are not considered in any of the mass fits performed in chapter 7. They nevertheless populate the low reconstructed B^0 mass region due to the missing energy of the neutrinos so that they are not as dangerous as the previous ones. Any residual contribution is absorbed by the partially reconstructed and combinatorial background components.

Other known exclusive candidates also considered as part of this category are the fully hadronic $B^0 \to K^{*0}(\to K^+\pi^-)\rho(\to \pi^+\pi^-)$ and $B^0 \to \pi^+D^-(\to \pi^-K^{*0})$ processes. However, they require a double mis-identification being strongly suppressed by PID requirements and neglected in the analysis.

$h \leftrightarrow l$ swap

These backgrounds are signal processes where the identities of one of the hadrons and leptons are swapped. The K^{*0} mass window together with PID requirements efficiently suppress them. The

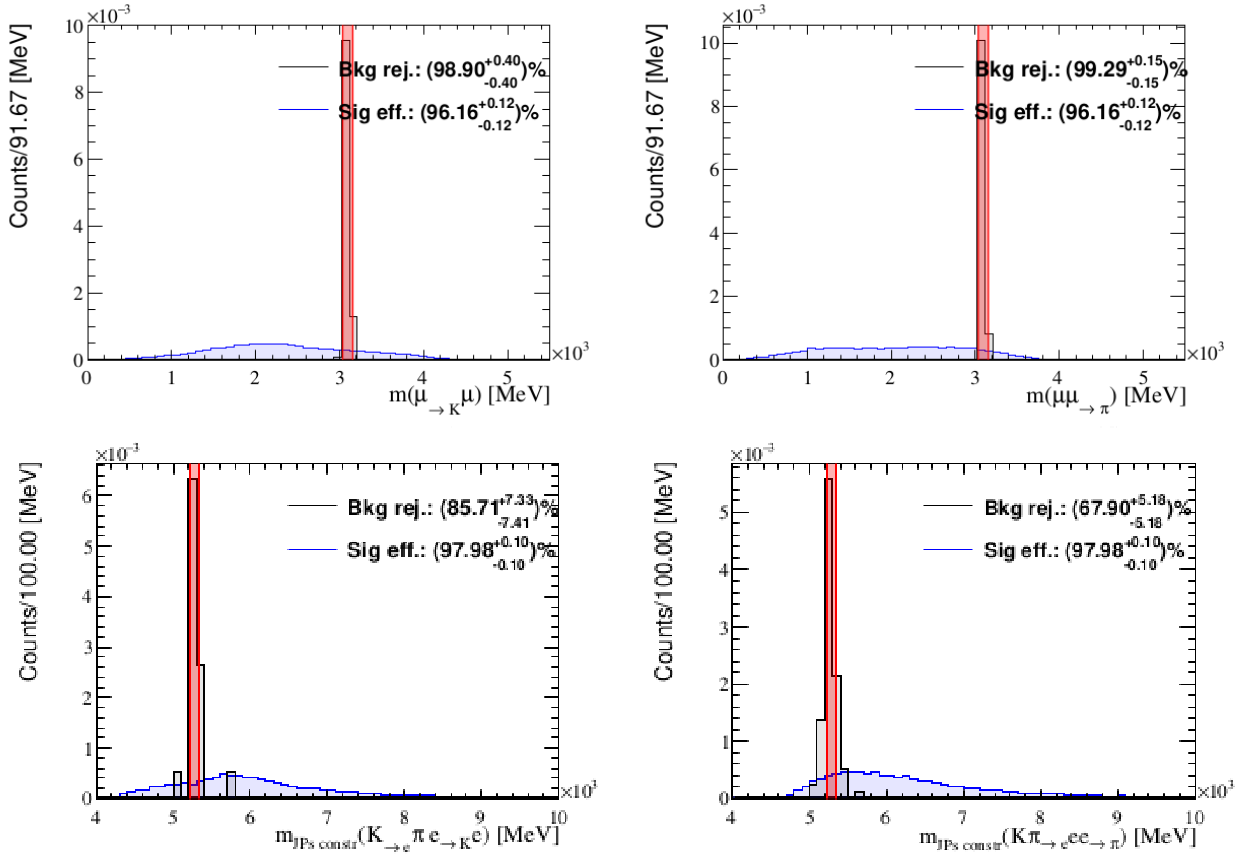


Figure 6.3: Veto against $h \leftrightarrow \mu$ (top) and $h \leftrightarrow e$ (bottom) mis-identified backgrounds. The left (right) panels show the two considered K (π) swap scenarios. Black and blue distributions represent background and signal MC simulated samples, respectively. Signal selection and background rejection efficiencies are shown for each case, calculated after all previous selection cuts. Taken from [64].

strategy to veto these backgrounds differs between the electron and muon modes. For muons a 60 MeV$/c^2$ wide mass window around the J/ψ and $\psi(2S)$ masses is vetoed after correcting the mass hypotheses of either the K and the π. Both swaps contribute and both are vetoed separately. In the electron channel this approach becomes inefficient due to the larger width of the two resonances. In this case the strategy is to veto the resonances in the kinematic constrained four-body invariant mass after after swapping the mass hypotheses of the electrons and the hadrons. The same 60 MeV$/c^2$ window is chosen for consistency. Vetoes applied to the muon and electron candidates are shown in figure 6.3. The figure shows the results for the J/ψ region and considering both the $K \leftrightarrow l$ and $\pi \leftrightarrow l$ swap scenarios.

On top of the mass vetoes, tighter e and μ PID requirements are applied within the vetoed window so that backgrounds are suppressed to negligible levels and do not need to be considered in the analysis.

Unfortunately, it is not possible to veto these backgrounds in the rare channel due to the absence of an intermediate resonance that can be used to separate them. Their contribution is expected to be small enough to neglect them and they are not considered in this thesis. A dedicated treatment might be needed in the future to confirm this and derive the corresponding systematic uncertainty for the determination of the signal yield.

Background	Requirement
$B_s^0 \to \phi(\to K^+K^-)l^+l^-$	$!(m(KK_{\to\pi}) < 1040\mathrm{MeV}/c^2\&\&\texttt{Pi\_ProbNNpi} < 0.8)$
$B^0 \to \pi^- \overline{D}^0(\to K^+\pi^-)l^+\nu_l$	$!(\lvert m(K^+\pi^-_{\to l^-}) - m_{D^0}^{PDG}\rvert < 30\mathrm{MeV}/c^2\&\&\texttt{L\_ProbNNl} < 0.8)$
$B^0 \to D^{*-}(\to \pi^- \overline{D}^0(\to K^+\pi^-))l^+\nu_l$	
$B^0 \to D^-(\to \pi^- K^{*0})l^+\nu_l$	$!(\lvert m(K^+\pi^+\pi^-_{\to l^-}) - m_{D^-}^{PDG}\rvert < 30\mathrm{MeV}/c^2\&\&\texttt{L\_ProbNNl} < 0.8)$
$h \leftrightarrow l$ swap	$!(\lvert m(\mu\mu_{\to h}) - m_{J/\psi,\psi(2S)}^{PDG}\rvert < 60\mathrm{MeV}/c^2\&\&\texttt{M\_ProbNNmu} < 0.8)$
	$!(\lvert m(hh_{\to e}ee_{\to h})^{J/\psi,\psi(2S)-DTF} - m_{B^0}^{PDG}\rvert < 60\mathrm{MeV}/c^2\&\&\texttt{E\_ProbNNe} < 0.8)$

Table 6.1: Summary of mis-identified background vetoes.

Summary

Table 6.1 presents a summary of all the vetoes applied in this analysis to suppress mis-identified backgrounds.

6.1.2 Partially reconstructed backgrounds

Despite of vertex and topological cuts specifically designed to reduce the partially reconstructed background rate, a large amount of partially reconstructed candidates manage to pass all the requirements, contaminating the dataset. Many of the contributing decays have large branching fractions. On top of that, there is a large number of decays which, not being important on their own, can pile up to make a large partially reconstructed background component. Different strategies have been designed in order to suppress this background which, as mentioned in section 4.1, is specially dangerous in the high-q^2 region. There, the $B^0 \to X\psi(2S)(\to e^+e^-)$ partially reconstructed background component peaks at the B^0 mass below the signal in the electron mode due to an unfortunate interplay between the missing energy of the non reconstructed particles and the overestimated energy of the electrons.

In this section we describe the two main contributions for which dedicated studies have been performed. Inclusive MC samples accounting also for many other contributions known to be less relevant are used to model the shape of partially reconstructed backgrounds in the signal mass fits.

Double semileptonic background

The main partially reconstructed background contribution comes from the double semileptonic decay $B^0 \to D^-(\to K^{*0}l^-\overline{\nu}_l)l^+\nu_l$ with the exact same final state as the signal besides the two missing neutrinos. This process has a huge branching fraction:

$$\mathcal{B}(B^0 \to D^-(\to K^{*0}l^-\overline{\nu}_l)l^+\nu_l)/\mathcal{B}(B^0 \to XJ/\psi(\to l^+l^-)) \sim 16, \tag{6.2}$$

and it is not suppressed by neither the K^{*0} mass cut nor the PID requirements. Thus, it can make a significant contribution to the signal mass fits. Fortunately, it is possible to design a very efficient veto to suppress these events if we apply a cut on the reconstructed $m(K\pi e^-)$ as shown in figure 6.4.

The veto requires $m(K\pi e^-) > m(D^-)$, removing most of the background contribution. Remaining events contribute to a very small amount and populate the low reconstructed B^0 mass sideband. They are absorbed in the signal mass fits performed for the electron mode by the combinatorial and partially reconstructed background components. Their contribution to the ass fits in the muon mode is negligible due to the narrower mass window employed there.

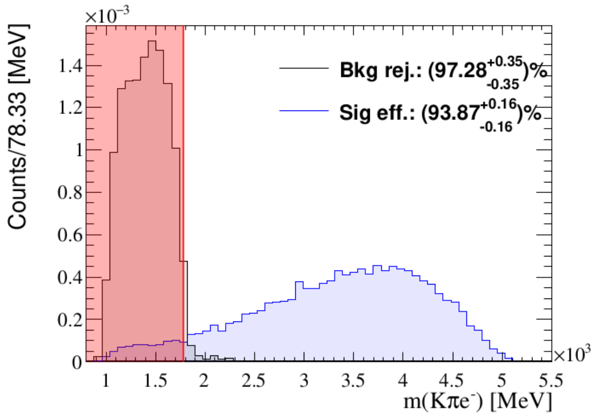


Figure 6.4: Veto against $B^0 \to D^-(\to K^{*0}l^-\overline{\nu}_l)l^+\nu_l$ background. Signal (background) MC samples selected in the J/ψ region are shown in blue (black). Signal selection and background rejection efficiencies are indicated in the legend. Taken from [64].

Partially reconstructed background from B^+

Resonant and non resonant $B^+ \to K^+\pi^+\pi^-l^+l^-$ processes, where the $K\pi\pi$ system can be either produced directly or via an intermediate state, can survive all the selection criteria if one of the two pions is not reconstructed. There is no efficient way to veto it. It is used as partially reconstructed background proxy to train the MVA classifier introduced in section 6.2 and included in the signal mass fits as part of inclusive partially reconstructed background samles.

6.1.3 Over-reconstructed background

The most relevant known source of over-reconstructed background is generated by resonant and non-resonant $B^+ \to K^+e^+e^-$ events to which a soft π is attached. It can be vetoed by requiring:

$$\max((m(Kl^+l^-), m(K_{\to\pi}l^+l^-)) < 5100 \text{ MeV}/c^2 \quad \text{with} \quad m(B^0) - m(\pi) \approx 5100 \text{ MeV}/c^2, \tag{6.3}$$

which also accounts for the possibility of a $\pi \to K$ mis-identification. This veto has been applied in previous analyses where it reduced the over reconstructed background efficiency to negligible levels with very little impact on the signal. It is however not applied in this analysis for two main reasons:

- The veto reduces a larger amount of signal in the high-q^2 region as compared with lower q^2 regions in the electron channel: in the central-q^2 region the efficiency of the veto on signal is $\sim 98\%$ whereas in the high-q^2 region it is reduced to $\sim 86\%$. A reduction on the signal efficiency is expected since in the high-q^2 region the electrons are more energetic and it becomes easier for the Ke^+e^- system to fall within the vetoed window. This effect is further enhanced by the Bremsstrahlung recovery procedure. Given the scarce amount of signal in the high-q^2 region we can not afford to lose $\sim 12\%$ of the signal candidates unless it is strictly necessary.

- The shape of the combinatorial background B^0 mass distribution is distorted by the cut. This can be evaluated using SS data as shown in figure 6.5. The over-reconstructed background veto distorts the right hand side of the combinatorial background mass distribution, making it peak at the B^0 mass and be much harder to constrain. This is specially problematic due to the blinded window around the B^0 mass. Would the veto be kept, it would not be possible to use the right sideband to constrain the shape of the combinatorial background.

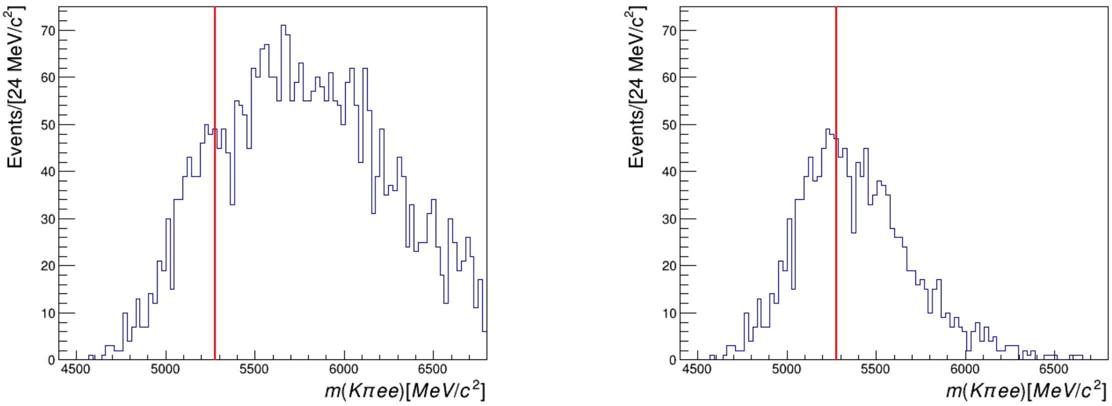


Figure 6.5: Effect of the over-reconstructed background veto on the shape of the combinatorial background B^0 mass distribution. SS data from the Run2p2 period is selected in the high-q^2 region with q2BDT > 0.95. Data is shown after full selection but without the MVA cuts in order to have more statistics. Left (right) panel shows the distribution before (after) the veto is applied. The position of the B^0 mass is indicated by the red vertical line.

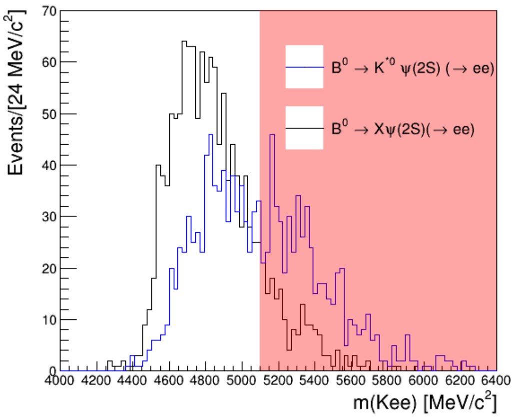


Figure 6.6: Effect of the over reconstructed background veto on the two leakage backgrounds selected in the high-q^2 region. Simulated data after full selection.

A third unwanted effect is the reduction of the $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ background component when the veto is applied. Even tough this might seem like an advantage, the most dangerous background leaking from the $\psi(2S)$ region is in fact the partially reconstructed $B^0 \to X\psi(2S)(\to e^+e^-)$ component. As shown in figure 6.6, this background is less affected by the veto due to the missing energy. Having a larger $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ background component in the right sideband of the B^0 mass might be helpful to constrain it better in the signal mass fit. This can help having a better control on the $B^0 \to X\psi(2S)(\to e^+e^-)$ partially reconstructed component peaking underneath the signal before unblinding the fit.

6.2 Multivariate classifiers

Remaining combinatorial and partially reconstructed background components can be suppressed by training two dedicated MVA classifiers which rely on different features of the decay that can be used to separate these components form the signal. The training approach follows exactly the same algorithm used to train the q2BDT explained in section 5.3. The final outcome is an MVA classifier dedicated to suppress the combinatorial background, referred to as $\mathrm{MVA_{Comb}}$, and a second classifier which suppresses partially reconstructed backgrounds, known as $\mathrm{MVA_{PR}}$. Both classifiers are independently trained for each of the data-taking periods considered in this analysis. Their selection cuts are also optimized independently for each period and q^2 region. The combinatorial $\mathrm{MVA_{Comb}}$ classifier is trained and applied independently to the electron and muon channels whereas $\mathrm{MVA_{PR}}$ is only used for the electron mode.

Samples used to train the classifiers have the full event selection applied. In order to reduce biases, an approach is taken to ensure that the same amount of signal and background events were used in every training, this is especially important for the SS data sample used as a proxy for combinatorial background, which is the most limited on statistics. For each of the training periods the statistics of either the signal or background sample is limited to match the number of events in the smaller sample. For the $\mathrm{MVA_{Comb}}$ classifier the limiting sample is always the combinatorial background sample. When only part of a sample is used, selected events are randomly picked combining years and magnet polarities in order to minimize biases in the training procedure.

In the following, MVAs are described and their performance in real data is shown. A detailed explanation about the cut optimization is also given.

The training and optimization of the two MVAs has been performed by the R_X team using the central- and low-q^2 regions, their performance has been checked in the high-q^2 region where they show similar results.

6.2.1 Combinatorial MVA

As a signal proxy, the combinatorial MVA uses fully selected and truth-matched $B^0 \to K^{*0} l^+ l^-$ simulated events for both muon and electron modes. The combinatorial background proxy was taken from real data selected in the central-q^2 region and populating the upper mass sideband in the B^0 mass. In this region the upper B^0 mass sideband is populated almost exclusively by combinatorial background. This choice avoids any possible bias introduced by using, for example, same sign data as combinatorial background proxy for the training. The definition of the upper sideband differs for electrons and muons due to the difference in resolution. For the electron mode, combinatorial background is selected with $m(B^0) > 5600\mathrm{MeV}/c^2$ whereas for the muon channel the selection window is expanded down to $m(B^0) > 5400\mathrm{MeV}/c^2$. In order to enhance the statistics in the electron channel the K^{*0} invariant mass window used to select events introduced in table 5.4 is widened to a window of $\pm 200\mathrm{MeV}/c^2$. No appreciable bias in the $\mathrm{MVA_{Comb}}$ performance is seen due to this decision.

The classifier is trained on a set of 23 different variables which contain information about the kinematic and geometric features of final state particles as well as the signal candidate vertex and vertex fit quality. These variable can be used to separate combinatorial background processes and signal candidates. The same training features are used for the electron and muon mode in order to keep the two training as close as possible. The most discriminating variables found after the training are χ^2_{DTF}/ndf, which quantifies the quality of the kinematic fit to the B^0 meson decay performed with the `DecayTreeFitter` tool, and the B^0 meson p_T.

Figure 6.7 shows the ROC curves for each of the 10 folds trained in the Run2p2 period. We can see

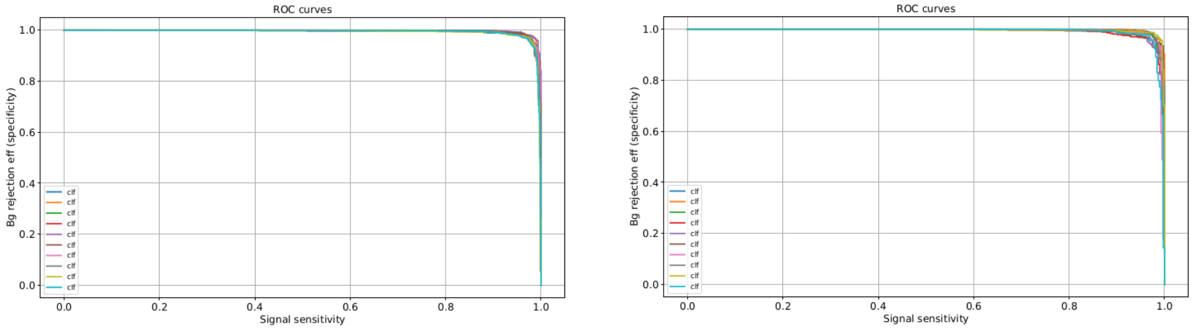


Figure 6.7: Background rejection against signal selection efficiency achieved with the 10 folds trained on muon (left) and electron (right) R2p2 data. Each fold is represented with a different color indicated in the legend. Taken from [64].

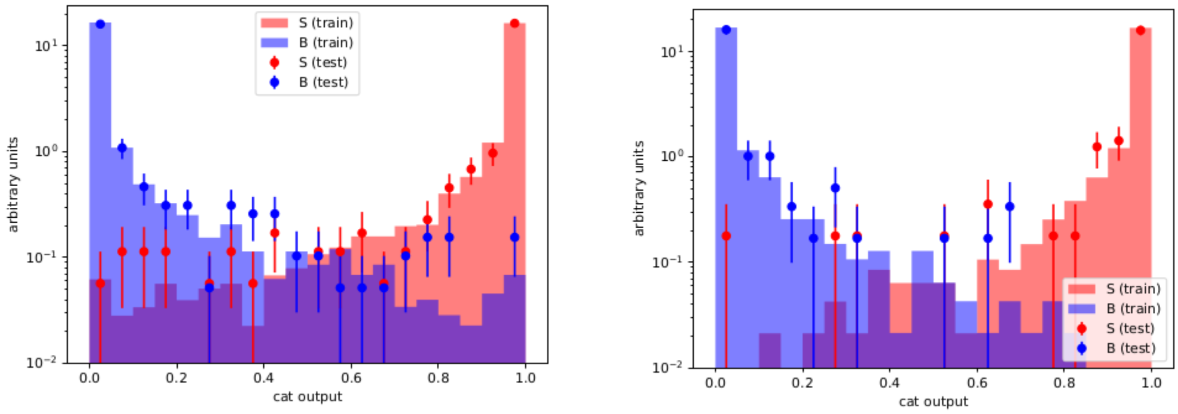


Figure 6.8: Overtraining plots for one of the Run1 folds for the muon (left) and electron (right) channels. Histograms represent the $\mathrm{MVA_{Comb}}$ distributions of data used for training whereas points represent the MVA outcome for testing data. Taken from [64].

a consistent behavior between different folds and an overall very good performance for both electron and muon modes. Overtraining is also checked for both muon and electron channels following the same procedure explained in section 5.3. Results for the Run1 period are shown in figure 6.8 with no signs of overtraining found in any of the performed checks.

6.2.2 Partially reconstructed MVA

A similar approach is taken to suppress partially reconstructed backgrounds. Signal is modelled using fully selected and truth-matched $B^0 \to K^{*0}e^+e^-$ candidates in the low and central q^2 regions. Partially reconstructed background is modelled using a dedicated $B^+ \to K^+\pi^+\pi^-l^+l^-$ simulated sample as as proxy, also fully selected and truth-matched. Even tough this sample is only modelling one of the partially reconstructed background components, the MVA classifier has a good performance suppressing any type of partially reconstructed background (see chapter 7). This is because the variables used for training are not specific to this particular decay and any other type of partially reconstructed background component is expected to show a similar behavior.

A total of 14 variables are used for the training. They represent kinematic and geometric features of the selected candidates and include some specifically designed cone isolation variables that

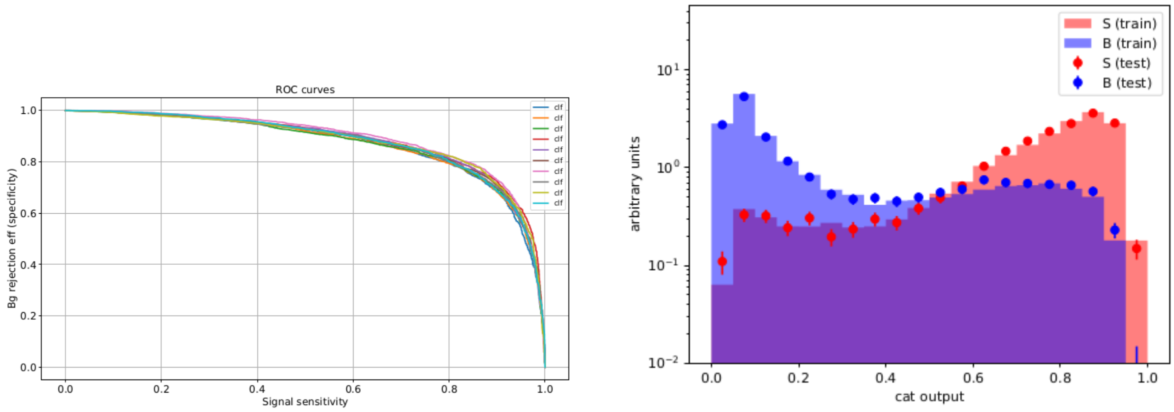


Figure 6.9: (Left) Background rejection against signal selection efficiency for the trainings in the Run2p2 periods using 10 different folds represented with different colors. (Right) Overtraining plot for the Run1 period comparing the $\mathrm{MVA_{PR}}$ distribution of training data (histograms) and testing data (right) Taken from [64].

quantify the isolation of final-state particle tracks in both space and momentum. These cone isolation variables are defined in a 0.5 mrad cone around the particle of interest. Several quantities are computed for which a different behavior between signal and partially reconstructed event candidates is expected. Two of these variables turned out to be the most discriminating ones: the *vertex isolation one-track-mass* and the *vertex isolation one-track-χ^2*. They are computed with the `VertexIsolation` tool from the `RelatedInfoTools`. An algorithm is implemented that sequentially adds tracks of the event to the reconstructed vertex of the decaying particle (the B^0 meson in this case). A new vertex is built after each step. The vertex isolation one-track-mass is defined as the reconstructed invariant mass of the resulting candidate with the smallest χ^2_{vtx}. This variable shows a sharp peak at the B^0 mass for signal events whereas a broader distribution is expected for partially reconstructed backgrounds. Similarly, the vertex isolation one-track-χ^2 is the smaller χ^2_{vtx} value achieved in the algorithm. It is close to zero for signal candidates and presents larger values for partially reconstructed backgrounds.

The performance is shown in figure 6.9 for the Run2p2 period. It is worse than the performance of the $\mathrm{MVA_{Comb}}$ classifier since partially reconstructed background features are more similar to signal features. As before, a overall agreement between different folds is seen. Figure 6.9 also shows the overtraining plots demonstrating good agreement between training and testing samples.

6.2.3 Cut optimisation

The optimisation of the MVA classifiers is performed using the signal significance as the figure of merit. It is defined as $N_S/\sqrt{N_S + N_B}$ where N_S and N_B are the expected number of signal and background events. It is done independently for electron and muon modes and for every data-taking period and q^2 region. Different approaches are taken to estimate the number of signal and background events for the electron and muon channels and in different q^2 regions.

For electrons, the expected amount of signal is calculated within a signal region defined as $5150 < m(B^0) < 5350\ \mathrm{MeV}/c^2$. The expected amount of signal events is then calculated using the following formula:

$$N_S = N_{B^0\to K^{*0}J/\psi(\to e^+e^-)} \cdot \frac{\mathcal{B}(S)}{\mathcal{B}(B^0 \to K^{*0}J/\psi(\to e^+e^-))} \cdot \frac{\varepsilon(S)}{\varepsilon(B^0 \to K^{*0}J/\psi(\to e^+e^-))} \cdot \frac{N_{MC|\mathrm{MVA_{Comb}}}}{N_{MC}} \tag{6.4}$$

The $B^0 \to K^{*0}J/\psi(\to e^+e^-)$ process is used as the control mode and its yield is extracted from constrained kinematic fits to the reconstructed B^0 mass. It is multiplied by the branching fraction ratio and the efficiency ratio between the signal mode of interest and the control mode. Efficiencies are computed using fully corrected MC simulated samples. The last factor can be interpreted as the signal efficiency of the MVA selection. For the rare modes, the branching fraction is calculated for the considered q^2 bin and can be computed using the *flavio* python package [44], which performs SM calculations of flavor physics processes. Needless to say, possible NP contributions are not accounted for in the calculation but, since they are expected to be small, they can be neglected in this case without the fear of introducing any appreciable bias. The number of events computed from MC simulations before and after applying the MVA cut is obtained using fully selected MC data for the process under consideration and taking into account the signal region defined before.

The extraction of the number of expected background events is done differently for the rare and resonant channels. For the resonant channels, fits to the reconstructed B^0 mass are performed in the upper and lower sidebands of the signal region. The signal region is excluded from the fits. The upper sideband of the B^0 mass is used to constrain the shape of the combinatorial background whereas the lower sideband is employed to extract the partially reconstructed background yield. The value of N_B is obtained separately for each component and the results are extrapolated to the signal region of interest. For the rare channels it much more complicated to independently extract the partially reconstructed background and combinatorial background yields from the sidebands since there is much less data and both components are much more correlated in the mass fits. In this case the full mass region is used for the fits and the combined combinatorial and partially reconstructed background yield is extracted. This means that the optimization for the rare channels is done simultaneously for both MVAs whereas in the control channels each of them is optimized independently.

For the muon mode a tighter mass window of $\pm 50\mathrm{MeV}/c^2$ around the B^0 mass is defined. The procedure for extracting the expected number of signal and background events is completely identical to the approach followed for the resonant electron channels since here the amount of data is larger and the resolution and separation of background components is much better.

All the fits in the electron and muon modes are performed independently for each data-taking period, iterating over different cuts in the MVA variables. After that, the value of the MVA cuts maximizing the significance is found. The optimized MVA cuts are shown in table 6.2.

Mode	$q^2 region$	Run1	Run2p1	Run2p2
Electron	J/ψ	$\mathrm{MVA_{Comb}} > 0.2\&\&\mathrm{MVA_{PR}} > 0.05$		
	$\psi(2S)$	$\mathrm{MVA_{Comb}} > 0.3$	$\mathrm{MVA_{Comb}} > 0.55$	$\mathrm{MVA_{Comb}} > 0.67$
	high	$\mathrm{MVA_{Comb}} > 0.9\&\&\mathrm{MVA_{PR}} > 0.4$		
Muon	J/ψ	$\mathrm{MVA_{Comb}} > 0.05$		
	$\psi(2S)$	$\mathrm{MVA_{Comb}} > 0.05$		
	high	$\mathrm{MVA_{Comb}} > 0.63$	$\mathrm{MVA_{Comb}} > 0.77$	$\mathrm{MVA_{Comb}} > 0.64$

Table 6.2: Summary of the MVA cuts used in this analysis.

A few points should be noted about the optimization:

- The MVA cuts defined for the high-q^2 region are not optimized following the described procedure. In this thesis we are using the same cuts optimized for the central-q^2 region. Further studies might be needed to re-optimize the cuts in the high-q^2 region given the different kinematics and background composition. For the electron mode in particular it might be worth to loosen the combinatorial MVA cut in order to allow a larger amount of combinatorial background. This can help to constrain its shape more easily in the mass fit, reducing its correlation with other components and making the fit more stable.

- In the electron mode J/ψ region and in both muon mode control channels, unified cuts are applied to all data-taking periods. In these modes, the signal efficiency is found to be relatively flat as a function of the MVA cuts and the behavior of the background efficiency is also consistent within periods. The approach taken is then to select a value of the cut corresponding to the point for which the background rejection efficiencies start to reach plateau.

- Unified cuts are also employed for the electron rare mode. This is a consequence of the optimization procedure for the central-q^2 region. There, a optimization combining the three run periods is found to be compatible with the optimized values for each period individually due to the smaller statistics. Therefore, the optimized values from the combined analysis are taken.

Chapter 7

Results

This chapter is dedicated to present the main results obtained during the course of the work leading to this thesis. It is structured as a set of sections, each of them containing a different task performed as part of the analysis. In section 7.1 the studies done to evaluate the response of the q2BDT introduced in section 5.3 are presented. We compare the response for the two leaking background components that contaminate the high-q^2 region. Section 7.2 presents the results of the studies performed to validate the usage of SS data as a proxy to model the B^0 mass distribution of combinatorial background in the high-q^2 region. A model to describe the signal in the high-q^2 region is also built and its performance is shown in section 7.3. In section 7.4 we present the strategy designed to constrain the backgrounds leaking into the high-q^2 region. Finally, in section 7.5, we estimate the expected amount of over-reconstructed (OR) background polluting the reconstructed B^0 mass fits in the high-q^2 region.

7.1 Evaluation of the q2BDT response

The high-q^2 region is selected with the aid of a dedicated BDT classifier trained to reduce the yield of the background components leaking from the $\psi(2S)$ region. In order to evaluate its performance using the $\psi(2S)$ region as a benchmark, a second classifier is trained to reject backgrounds leaking from the J/ψ region and polluting the $\psi(2S)$ region. The two classifiers are introduced in section 5.3. The classifiers are trained using $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ and $B^0 \to K^{*0}J/\psi(\to e^+e^-)$ simulated samples as background proxies for the q2BDT$_{\psi(2S)}$ and the q2BDT$_{J/\psi}$ respectively. A similar performance is expected for leaking partially reconstructed background components, which are the most problematic. The choice of $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ instead of $B^0 \to X\psi(2S)(\to e^+e^-)$ as a background proxy is motivated by the fact that the simulation of partially reconstructed backgrounds is generally less reliable. This is especially a problem in the $\psi(2S)$ region where branching fractions are known to a much lower precision. Hence, results obtained using $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ as background proxy are expected to be more robust. Nevertheless, in order to keep the response as close as possible for the two background components, the features chosen for the training rely on electron properties only, which are expected to have a very similar behavior for the two of them. This is because the two backgrounds only differ in the reconstruction of the hadronic part [1].

The study shown here compares the relative efficiency between the selection of the two relevant $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ and $B^0 \to X\psi(2S)(\to e^+e^-)$ components in the high-q^2 region and in

[1] This is not entirely true for the $B^0 \to XJ/\psi(\to e^+e^-)$ background polluting the $\psi(2S)$,which should be rejected by the q2BDT$_{J/\psi}$ classifier, since we know it carries a non negligible partially reconstructed leptonic component. For processes in the $\psi(2S)$ region this component is expected to be small and is not included in the simulated sample.

the $\psi(2S)$ region selected using both q2BDT classifiers. We want to check if the relative response of the BDT classifiers is the same. This is something we would expect given that they were both trained using the same features and signal proxy. The high-q^2 and $\psi(2S)$ regions are selected in the following way:

- **High q^2 region:** q2BDT$_{\psi(2S)} > 0.95$
- **$\psi(2S)$ region:** q2BDT$_{J/\psi} > 0.95$ && q2BDT$_{\psi(2S)} < 0.95$

The selection of the $\psi(2S)$ region makes use of the two q2BDT classifiers. The inverted selection on q2BDT$_{\psi(2S)}$ relative to the selection of the high-q^2 region ensures that there is no overlap between the two samples. The cut on q2BDT$_{\psi(2S)}$ has however very little impact in the selection of the $\psi(2S)$ region since only 11.6 % of the $\psi(2S)$ candidates leak into the high-q^2 region. This way of selecting the high-q^2 and $\psi(2S)$ regions will be kept for most of the remaining results in this thesis. We will be testing the following observable:

$$\varepsilon_r = \frac{\left(\frac{\mathcal{N}_{B^0\to X\psi(2S)(\to e^+e^-)}}{\mathcal{N}_{B^0\to K^{*0}\psi(2S)(\to e^+e^-)}}\right)^{\text{high-}q^2}}{\left(\frac{\mathcal{N}_{B^0\to X\psi(2S)(\to e^+e^-)}}{\mathcal{N}_{B^0\to K^{*0}\psi(2S)(\to e^+e^-)}}\right)^{\psi(2S)}}. \tag{7.1}$$

This observable is sensitive to the q^2 dependence of the of the relative $B^0 \to X\psi(2S)(\to e^+e^-)$ to $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ yield fraction. It is expected to be very close to 1 given the absence of leptonic partially reconstructed processes affecting the inclusive sample. Any deviation from unity should be explainable in terms of small differences in the q^2 spectra or any of the other variables used to train the classifiers. This relative efficiency is determined using fully selected simulated samples, including all background vetoes presented in section 6.1 but excluding the MVA selection in order to remove any effect induced by the MVA cuts. In this way, ε_r singles out the relative efficiencies of the two q^2 selections, as any other contribution to the efficiencies cancels in the ratio. Samples are splitted into the three usual data-taking periods and results are shown in figure 7.1.

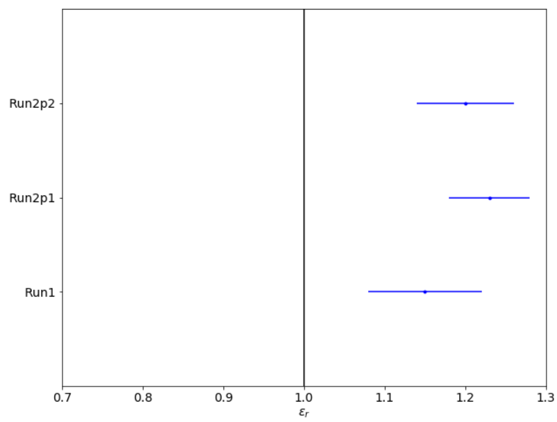


Figure 7.1: Relative efficiency for selecting $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ and $B^0 \to X\psi(2S)(\to e^+e^-)$ candidates in the high-q^2 and $\psi(2S)$ regions computed using fully selected and truth-matched MC simulated samples. Results are splitted into data-taking periods. The variable ε_r is defined in the text.

Results indicate that there is a larger fraction of $B^0 \to X\psi(2S)(\to e^+e^-)$ candidates relative to $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ in the high-q^2 region compared to the $\psi(2S)$ region. Consistency between different years implies that deviations from one are likely due to differences in the physical behavior of the two components, rather than problems with the q2BDT performance. This is because independent classifiers were trained for each data-taking period. Uncertainties are calculated treating the number of events counted in each region as Poisson distributed variables and assuming no correlations between them.

Differences are found to be primarily caused by the mismatching of the q^2 distributions between the two samples. This is shown in figure 7.2. The relative amount of the partially reconstructed component is larger in the right hand side tail, corresponding to larger q^2 values. Investigations show that this is related to differences in some kinematic variables of the B^0 decay. This effect is enhanced by the over-reconstructed background veto which, as explained in section 6.1, is more efficient for $B^0 \to X\psi(2S)(\to e^+e^-)$ partially reconstructed candidates than for $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ candidates in the high-q^2 region. Results shown here do not include the veto.

A lot of effort has been dedicated to try to bring the relative efficiency as close to one as possible. If ε_r is close to unity it would be easier to control the amount of $B^0 \to X\psi(2S)(\to e^+e^-)$ background in the high-q^2 region. One of the main goals of this analysis is to constrain the amount of partially reconstructed background leaking from the $\psi(2S)$ region into the high-q^2 region. This is a very dangerous background since it peaks underneath the signal (see figure 4.2). The simplest way to constrain the two leakage yields in the signal mass fits performed in the high-q^2 region is to extrapolate the results from measurements of these yields performed in the $\psi(2S)$ region. Yield results can be extrapolated using the efficiency ratios computed from MC simulations with the formulas:

$$\mathcal{N}^{\text{high-}q^2}_{B^0\to K^{*0}\psi(2S)(\to e^+e^-)} = \mathcal{N}^{\psi(2S)}_{B^0\to K^{*0}\psi(2S)(\to e^+e^-)} \cdot \frac{\varepsilon^{\text{high-}q^2}_{B^0\to K^{*0}\psi(2S)(\to e^+e^-)}}{\varepsilon^{\psi(2S)}_{B^0\to K^{*0}\psi(2S)(\to e^+e^-)}}, \tag{7.2}$$

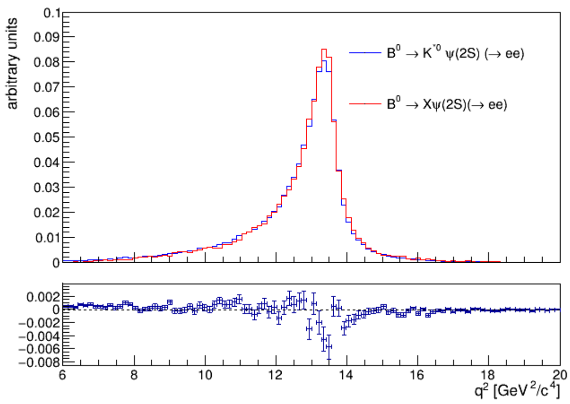


Figure 7.2: The upper panel shows the q^2 distribution of the two relevant $\psi(2S)$ components indicated in the legend taken from fully selected simulated samples in the R2p2 period. The histograms are normalized for a better comparison. The lower panel shows the absolute differences in each of the bins with errors calculated assuming Poisson statistics.

$$\mathcal{N}^{\text{high-}q^2}_{B^0\to X\psi(2S)(\to e^+e^-)} = \mathcal{N}^{\psi(2S)}_{B^0\to X\psi(2S)(\to e^+e^-)} \cdot \frac{\varepsilon^{\text{high-}q^2}_{B^0\to X\psi(2S)(\to e^+e^-)}}{\varepsilon^{\psi(2S)}_{B^0\to X\psi(2S)(\to e^+e^-)}}. \tag{7.3}$$

The disadvantage of this approach is that efficiencies for the partially reconstructed background component computed from MC simulated data are not very reliable due to possible mismodellings in the simulation. The constraint on the $B^0 \to X\psi(2S)(\to e^+e^-)$ leaking background yield is therefore sensitive to these mismodellings. There is an alternative way to constrain the yield of the $B^0 \to X\psi(2S)(\to e^+e^-)$ component. We can reduce the impact of systematic uncertainties coming from the simulation of the inclusive sample by expressing the $B^0 \to X\psi(2S)(\to e^+e^-)$ yield in terms of the $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ yield, both measured in the high-q^2 region. This approach is more robust as long as their relative efficiency ratio is close to unity. Dividing equations 7.2 and 7.3 and rearranging the result we find:

$$\mathcal{N}^{\text{high-}q^2}_{B^0\to X\psi(2S)(\to e^+e^-)} = \mathcal{N}^{\text{high-}q^2}_{B^0\to K^{*0}\psi(2S)(\to e^+e^-)} \cdot \varepsilon_r \cdot \frac{\mathcal{N}^{\psi(2S)}_{B^0\to X\psi(2S)(\to e^+e^-)}}{\mathcal{N}^{\psi(2S)}_{B^0\to K^{*0}\psi(2S)(\to e^+e^-)}}, \tag{7.4}$$

where the ratio of efficiencies ε_r is defined in equation 7.1. If ε_r is close to unity, the constrain on the $B^0 \to X\psi(2S)(\to e^+e^-)$ background yield becomes almost independent of the MC modelling of the inclusive sample. Unfortunately, it is not possible to bring ε_r closer to one without negatively affecting the signal efficiency in the high-q^2 region.

An interesting outcome of these studies is the finding that the value of ε_r strongly depends on the MVA cuts applied to suppress combinatorial and partially reconstructed backgrounds. Figure 7.3 shows the evolution of ε_r in bins of both MVA_{Comb} and MVA_{PR}. The same MVA cuts are applied in the two q^2 regions. A clear tendency is seen where the discrepancies increase as the MVA cuts are tightened. Results are shown for the R2p2 period, other periods show consistent results.

The MVA classifiers mostly reject background candidates with low reconstructed B^0 masses since they are more easily distinguishable from signal candidates. Leaking partially reconstructed

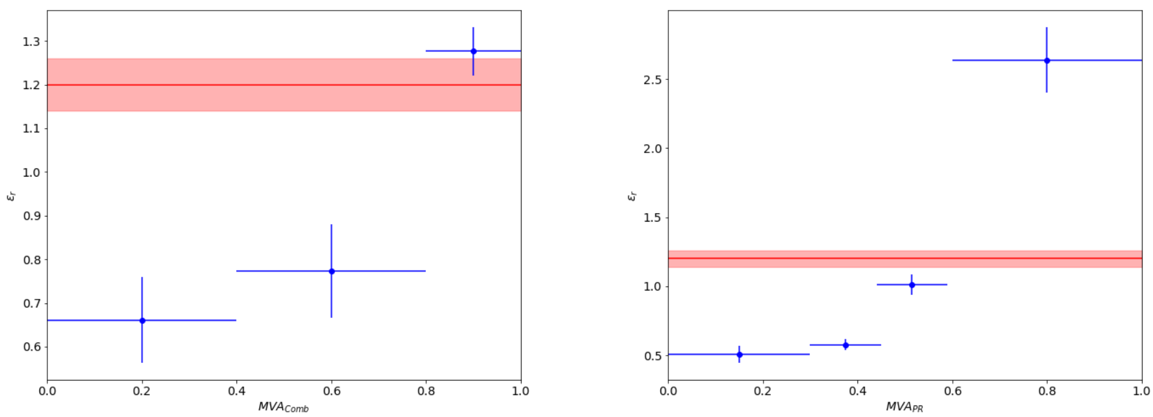


Figure 7.3: Evolution of ε_r ,defined in the text, as a function of MVA_{Comb} (left) and MVA_{PR} (right). Results are calculated using R2p2 simulated samples and uncertainties assume a Poisson behavior of the number of events. The red band represents the average value obtained without MVA cuts. Bin widths are selected ensuring similar statistics in each of them.

candidates are located closer to the B^0 mass and are therefore more similar to true signal candidates and difficult to efficiently reject. This explains why after applying tighter MVA cuts the relative fraction of partially reconstructed $B^0 \to X\psi(2S)(\to e^+e^-)$ selected candidates with respect to fully reconstructed $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ candidates increases in the high-q^2 region relative to the $\psi(2S)$ region. When we invert the MVA selection (lower bins in figure 7.3) we see the opposite effect. This is because inverting the MVA cuts imply rejecting signal and selecting a background enriched sample where leaking partially reconstructed $B^0 \to X\psi(2S)(\to e^+e^-)$ candidates are rejected together with signal candidates, keeping mostly candidates populating the low reconstructed B^0 mass region.

This behavior also motivates loosening the MVA cuts in the high-q^2 region, adding up to the reasons exposed in section 6.2.

7.2 Validation of the usage of same sign data to model the mass distribution of combinatorial background

Anther important step in the analysis is to check whether we can rely on SS data to model the reconstructed B^0 mass distribution followed by combinatorial background in the high-q^2 region. In particular, it is important to check if the turn-on effect seen in the high-q^2 region as a consequence of the q^2 selection is well described in SS data. In order to check this we can use the $\psi(2S)$ region as a benchmark since, going low enough in the reconstructed B^0 mass, we see the exact same effect as in the high-q^2 region. This also allows us to check in real data how does the q2BDT behaves once it has been confirmed its rejection power against leaking partially reconstructed backgrounds.

The validation is done in the following steps:

- A fit to the reconstructed B^0 mass is performed using SS data selected in the $\psi(2S)$ region.
- We select a combinatorial background enriched sample and perform the same fit as in SS data. In order to check if the shape of the mass distribution is well described in SS data, the shape of the PDF used to model the combinatorial background component in this fit is constrained taking the parameters from the result of the fit performed with SS data.
- The fit results are compared. In particular, the combinatorial background yield extracted from the second fit should be of roughly the same order of magnitude as the SS data yield.

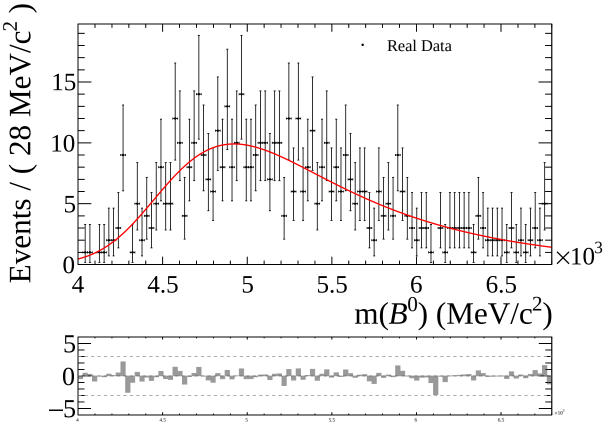
Results are presented in the following, explaining the details of the strategy to perform the fits.

7.2.1 Fit to the reconstructed B^0 mass in SS data

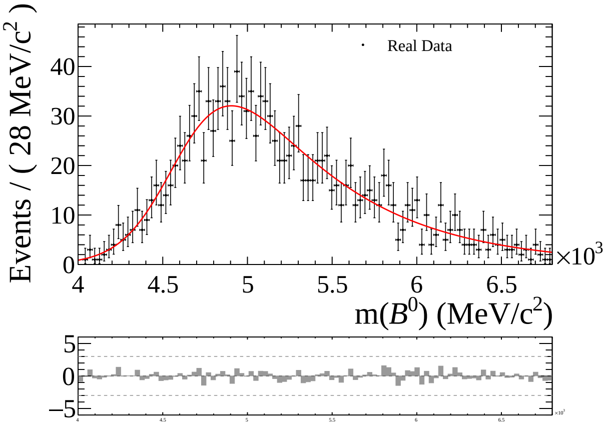
For the fit to the reconstructed B^0 mass performed in SS data we use a fully selected SS data sample, including the exclusive background vetoes introduced in section 6.1. It is selected in the $\psi(2S)$ region using the two q2BDTs as indicated in table 5.6 of section 5.3 and with a tight inverted $\mathrm{MVA_{Comb}}$ cut. This $\mathrm{MVA_{Comb}}$ cut selects the combinatorial background enriched sample we want to compare it with:

$$\mathrm{MVA_{Comb}} < 0.05.$$

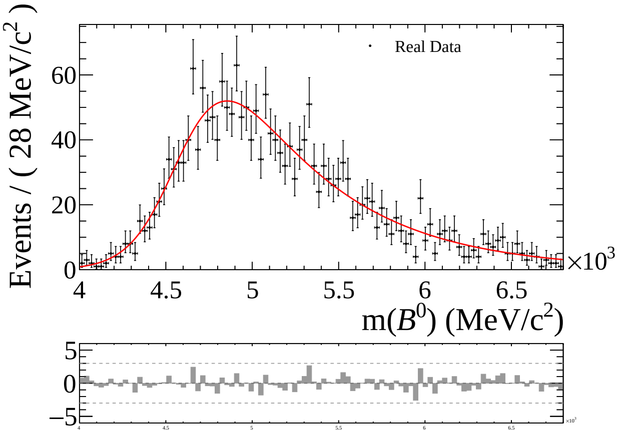
This selection suppresses most of the signal and background components from the dataset. The



(a) Run1.



(b) Run2p1.



(c) Run2p2.

Figure 7.4: Fits to the reconstructed B^0 mass in SS data selected in the in the $\psi(2S)$ region. Results for each period of data-taking are shown and indicated in the captions. The red line in the plots represents the fitted function. The lower panels show the pull histograms of the fits.

veto to remove the peak from signal events in the SS data sample is also applied (see section 5.1).

The PDF chosen to model the SS data mass distribution is an exponential function with a turn-on term. The exponential models the usual exponentially falling behavior of combinatorial background whereas the turn-on factor accounts for the distortion effect caused by the q^2 cut at low reconstructed B^0 masses. It is a 4-parameter PDF with the following functional form:

$$f(x;\alpha,\lambda,\beta,\sigma) = \frac{e^{\lambda(x-\alpha)}}{\left[1+(2^{\beta}-1)e^{-\frac{x-\alpha}{\sigma}}\right]^{\frac{1}{\beta}}}. \tag{7.5}$$

The λ parameter is the slope of the exponential function. The width of the distribution and the position of the maximum are controlled by σ and α, respectively. The parameter β is used to tune the shape. All of the parameters are positive by definition. To cross-check the results and also to account for differences in the trigger selection and experimental conditions, data is splitted into the usual three data-taking periods and fitted independently in each case. The fit is performed in the following mass range:

$$4000 < m(B^0) < 6800 \text{ MeV}/c^2. \tag{7.6}$$

The extended lower limit ensures that the turn-on effect is included in the mass range employed for the fit. The mass window is extended up to 6800 MeV/c^2 in order to have enough statistics for the exponential slope. Results are shown in figure 7.4 with the fit function superimposed on data. The shape of the SS data mass distribution is very well described in the model for the three data-taking periods. The Run1 data distribution looks noticeably flatter. This might be due to differences in the HLT strategy affecting the combinatorial background shape or, perhaps, it is just a consequence of the lower statistics.

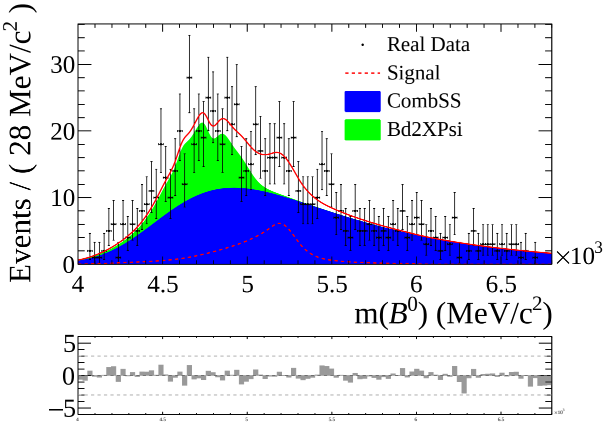
7.2.2 Constraining the shape of combinatorial background

A combinatorial background enriched sample can be obtained by applying a tight inverted MVA$_{\text{Comb}}$ cut as explained in the previous section. After the selection only some partially reconstructed $B^0 \to X\psi(2S)(\to e^+e^-)$ background and a small signal component are remaining, besides the combinatorial background itself. The shape of the combinatorial background mass distribution is modelled using the same function employed to model SS data introduced in equation 7.5. The parameters are gauss-constrained from the results of the SS data fits as explained in section 4.2. The model developed in section 7.3 is used for the signal component, details about it are deferred to that section since they are not relevant here. The partially reconstructed background mass distribution is modelled using a *RooKeysPdf*, which is the `RooFit` implementation of a KDE model. It uses a gaussian kernel with an adaptive method to modify the gaussian bandwidth for different regions of the dataset so that larger bandwidths are used in regions with lower event density and vice-versa.

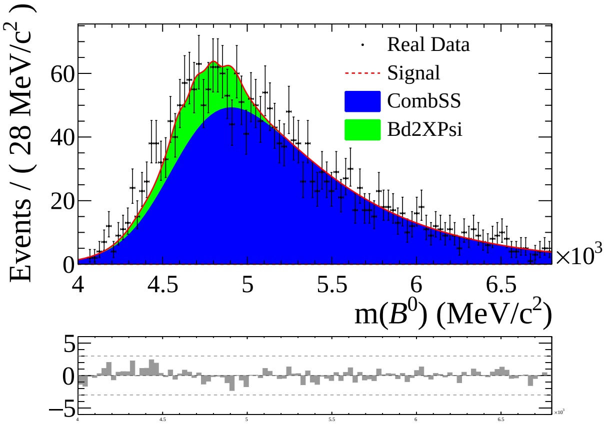
All fit settings are kept the same as in the SS data fit to simplify the comparison. In particular, the same q^2 selection and mass range are employed. Data is fully selected including the exclusive background vetoes. Results are shown in figure 7.5. A good agreement is seen between the results in SS data shown in the previous section and the fit results using the combinatorial background enriched sample. In particular, we see the same turn-on effect in both cases, which is well modeled in the PDF used to describe the shape of the combinatorial background and constrained form the SS data mass fits.

The signal yield is fitted to be negligible in the two Run2 data samples whereas a small contribution is needed in order to describe data from Run1, which also requires a larger partially reconstructed background $B^0 \to X\psi(2S)(\to e^+e^-)$ component. This is a consequence of the anomalous effect seen in SS data for the Run1 period. We can numerically compare the SS data and *Opposite Sign* (OS) combinatorial data yields. Numbers do not necessarily have to match or be compatible due to the differences in the stripping selection which can lead to different efficiencies and possible contamination of SS data from events that do not have a combinatorial nature like the peaks associated to signal events explained in section 5.1. However, the yields are expected to have the same order of magnitude and a very large difference between the selected amount of SS data and the fitted amount of combinatorial background might point at mismodelings in the combinatorial background yield estimation. Results are compared in table 7.1 where uncertainties in both yields are included. For the SS data, uncertainties are computed assuming the yield to follow a Poisson distribution whereas the uncertainties in the combinatorial OS data yields are directly coming from the fit results.

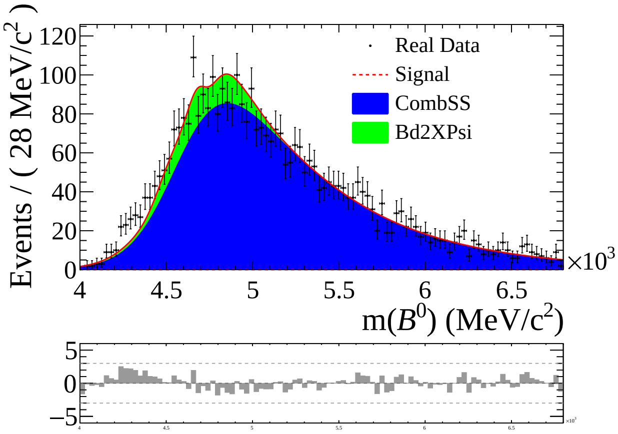
The ratio between the OS ad SS data yields in the Run1 period is significantly smaller compared to the ratios obtained from Run2 results. This is probably coming from either a mismodelling in the



(a) Run1.



(b) Run2p1.



(c) Run2p2.

Figure 7.5: Fits to the reconstructed B^0 mass using a combinatorial background enriched data sample selected in the $\psi(2S)$ region. Selection cuts are explained in the text. Different fit components are indicated in the legend. Backgrounds are represented stacked on top of each other whereas the signal component is shown as a dashed line. The red solid line is the full fitted function. Results for different data-taking periods are shown as indicated in the captions. The lower panels show the pull histograms of the fits.

Component	**Run1**	**Run2p1**	**Run2p2**
Combinatorial OS	601 ± 42	2157 ± 63	3473 ± 84
Combinatorial SS	519 ± 23	1403 ± 37	2089 ± 46
Ratio(OS/SS)	1.16	1.54	1.66

Table 7.1: Comparison between the SS and *Opposite Sign* (OS) data yields.

combinatorial background shape due to the lower statistics in the SS data sample used to constrain it, or from differences in the trigger selection. Despite that, we can see that results are consistent amongst run periods. The yield of combinatorial background is consistently larger than the SS data yield but always within the same order of magnitude. This validates the usage of SS data to model the shape of the combinatorial background mass distribution in the high-q^2 region.

7.3 Building a model for the signal

The mass distribution of signal $B^0 \to K^{*0}e^+e^-$ candidates selected in the high-q^2 region is also distorted by the q^2 selection (see section 4.1). A special treatment is therefore required in order to build a model to describe it. Having a model accurately describing the signal mass shape is specially important to correctly estimate the signal yield in the high-q^2 region and reduce systematic uncertainties in the final result.

In this section, the model developed to describe the signal mass shape is explained in detail. The model is built using $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ simulated data selected in the $\psi(2S)$ region and then tested using a $B^0 \to K^{*0}e^+e^-$ dataset selected in the high-q^2 region, also from MC simulations. In order to account for differences between Bremsstrahlung categories, the model is built independently for each of them. Validation in data requires control of the backgrounds present in the $\psi(2S)$ region. This is the task section 7.4 is dedicated to so we have to wait until then to see the model at work .

7.3.1 Mass fits to simulated data selected in the $\psi(2S)$ region

The development of the model is put forward following a sequence of steps, designed to sequentially incorporate the effects of the q^2 cuts into the signal PDF. They are summarized below:

- First we perform a fit to the reconstructed B^0 mass using $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ simulated candidates without applying any q^2 selection but keeping all other selection requirements. This allows to model the core of the signal describing the mass peak.

- Then we compute the q^2 selection efficiency profile as a function of the reconstructed B^0 mass for each Bremsstrahlung category. By doing this we single out the effects of the q^2 cuts on each dataset.

- The q^2 selection effects are finally included on the signal model using the information drawn from the efficiency profiles and the model is applied to fit the fully selected simulated data, including now the q^2 selection as well.

In the following full procedure followed to develop the signal model is shown. The signal core is modelled using a *Double Sided Crystal Ball* (DSCB) function. It is a generalized version of the *Crystal Ball* (CB) function introduced to model the asymmetric shape of resonances seen in the Crystal Ball experiment, carried out during the 80s at the *Stanford Linear Accelerator Centre* (SLAC) [75]. The CB function is very often used in high energy physics to model the shape of resonances reconstructed from final states which contain undetected particles leading to energy losses which are reflected in the lineshape of the resonance fit as long asymmetric tails towards low reconstructed masses. It consists of a gaussian core modelling the body of the resonance attached to a falling power law tail. The Bremsstrahlung recovery algorithm implemented at LHCb to recover the energy lost by the electrons and improve the reconstructed mass resolution results in an additional tail to the right hand side of resonance peak. This tail comes from events where photon energy clusters are wrongly associated to electrons, overestimating their energy. To model this second tail, the CB function needs to be generalized by including a second falling power law tail in the right hand side of the distribution. The DSCB function is a 6 parameter distribution with a central core modelled with a gaussian function and two identical power law tails at the two sides of it:

$$f_{DSCB}(x;m,\sigma,\alpha_L,n_L,\alpha_R,n_R) = \begin{cases} A_L\left(\frac{n_l}{|\alpha_L|} - |\alpha_L| - \frac{x-m}{\sigma}\right)^{-n_L} & \text{for } \frac{x-m}{\sigma} \leq -\alpha_L \\ G(x;m,\sigma) & \text{for } -\alpha_L < \frac{x-m}{\sigma} < \alpha_R \\ A_R\left(\frac{n_R}{|\alpha_R|} - |\alpha_R| + \frac{x-m}{\sigma}\right)^{-n_R} & \text{for } \frac{x-m}{\sigma} \geq \alpha_R, \end{cases} \tag{7.7}$$

where $G(x;m,\sigma)$ is a gaussian distribution with mean m and width σ. The factors A_L and A_R connect the gaussian core to the power law tails in a smooth and differentiable way:

$$A_{L,R} = \left(\frac{n_{L,R}}{|\alpha_{L,R}|}\right)^{n_{L,R}} e^{-\frac{|\alpha_{L,R}|^2}{2}}. \tag{7.8}$$

Hence, the parameters m and σ are the mean and standard deviation of the gaussian core, α_L and α_R delimit the interval within which the gaussian function is defined and $n_{L,R}$ control the behavior of the power law tails. All parameters are defined to be positive.

Fully selected $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ simulated events are used to model the signal mass distribution. The selection includes the exclusive background vetoes and MVA cuts optimized for the $\psi(2S)$ region introduced in sections 6.1 and 6.2. Candidates are fully truth-matched as explained in section 5.2. Results are shown in figure 7.6 for each of the Bremsstrahlung categories for the Run1 period. Results in other data-taking periods are consistent.

In the 0G category, both CB and DSCB functions have been tested since the right hand side tail is not really needed here. Results are compatible and the DSCB shape is kept for consistency with the models employed in the other Bremsstrahlung categories. The effect of the Bremsstrahlung recovery procedure in the lineshape of the signal mass distribution is evident. The shapes of the tails are perfectly modelled in all the fits, with no relevant distortions found due to the MVA cuts and vetoes employed to suppress exclusive backgrounds. The fact that the shape of the mass distribution is different in each of the Bremsstrahlung categories is what motivates the splitting of the dataset to model the reconstructed B^0 mass shape of signal candidates.

Once the core of the signal model is fixed the next thing to do is to find a way to describe the effects associated to the q^2 selection cuts affecting the tails of the distributions. We can see how the cuts modify the shape of the reconstructed B^0 mass distribution of $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ candidates by looking at the efficiency profile of the q^2 selection as a function of the reconstructed B^0 mass. Results are shown in figure 7.7 using the R2p2 period as a reference.

The cuts on q^2 introduce a step-like behavior in the two sides of the distribution. In order to see the effect of the q2BDT selection results are compared with the ones obtained using the traditional approach of selecting a window in q^2, shown in the right side of figure 7.7. When we select data using the q2BDT approach the two steps to the sides of the signal core are sifted towards the right hand side as the number of recovered Bremsstrahlung photons increases. This is in contrast to the behavior observed when using the normal q^2 cuts, for which they are always located in the same place. This difference is what makes the q2BDT selection more efficient. When no Bremsstrahlung photons are recovered, the cuts on q^2 are shifted to the left with respect to the nominal ones given the absence of backgrounds from leaking candidates coming from the J/ψ region. This increases the signal efficiency. As more Bremsstrahlung photons are added, the cuts get shifted to the right to keep the amount of leaking backgrounds low. It is interesting to note that the size of the selection window is kept approximately constant for the three Bremsstrahlung categories in the $\psi(2S)$

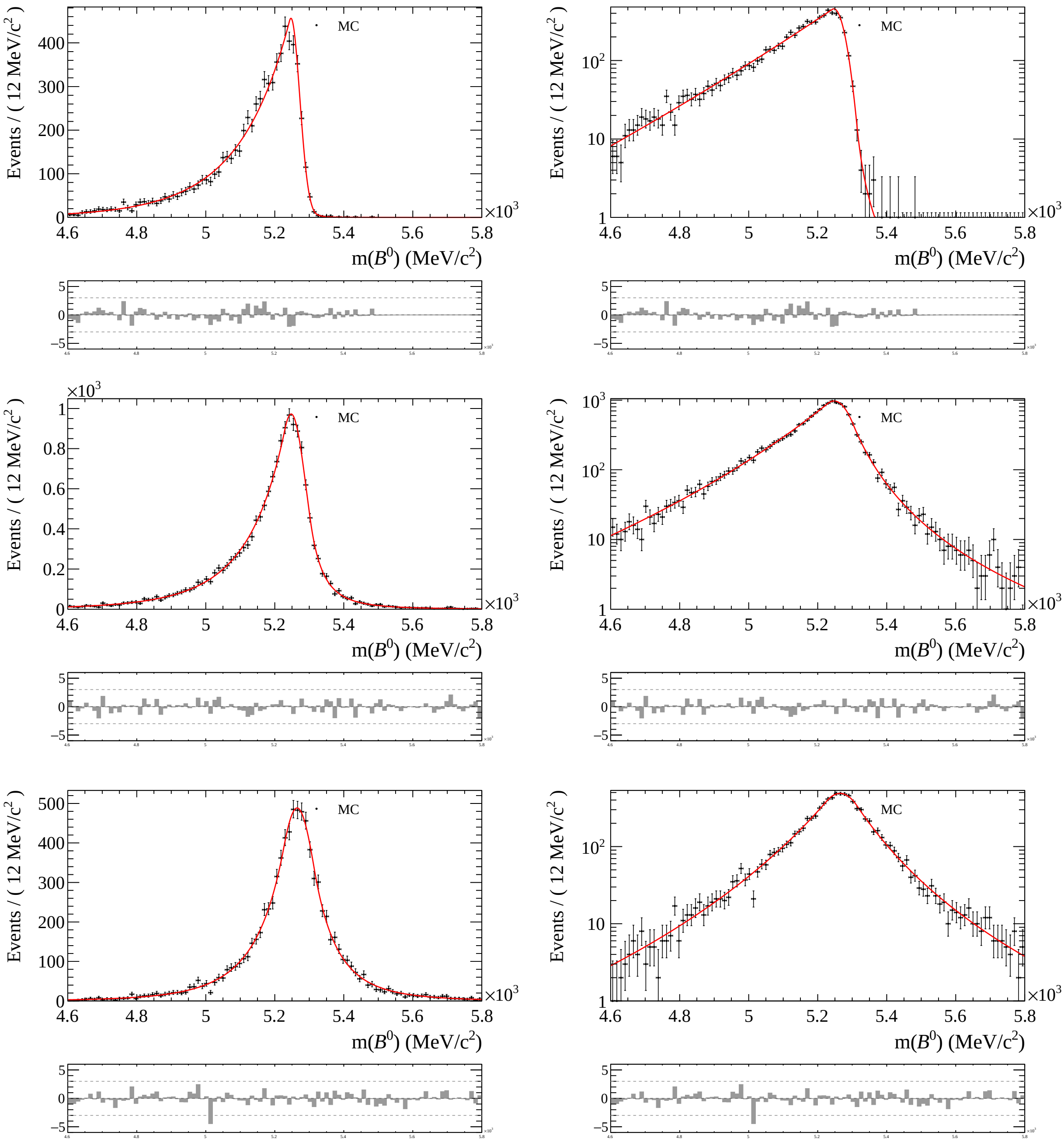


Figure 7.6: Fits to the reconstructed B^0 mass in the $\psi(2S)$ region using a MC simulated $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ sample. The first, second and third rows show the fits to the OG, 1G and 2G Bremsstrahlung categories, respectively. In each row the same fit is shown twice, using a linear scale for the y axis (left) and a logarithmic scale (right). The lower panels show the pull histograms of the fits.

region. The q2BDT selection is extremely efficient here performing even better than in the high-q^2 region. When the high-q^2 region is selected using the q2BDT the size of the q^2 selection window gets reduced as the number of Bremsstrahlung photons added to the candidates increases. This is due to the shift of the single step to the left hand side (corresponding to the right hand side step here). This observation explains the worse performance achieved by the q2BDT$_{\psi(2S)}$ classifier compared to the q2BDT$_{J/\psi}$ classifier seen in section 5.3.

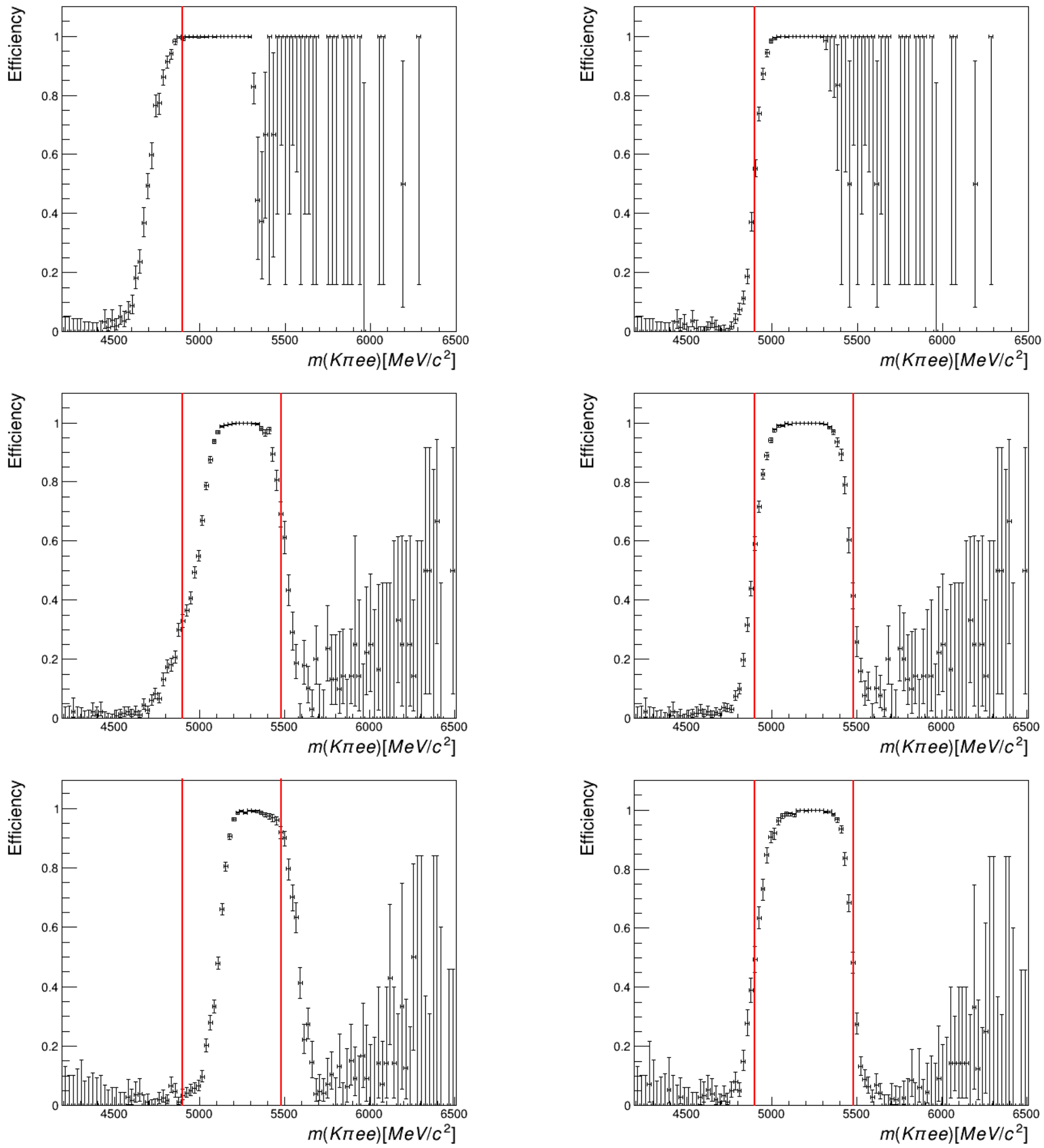


Figure 7.7: Efficiency of the q^2 selection as a function of the reconstructed B^0 mass using simulated $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ data from the Run2p2 period, selected in the $\psi(2S)$ region. From the upper to the lower rows, results for the 0G, 1G and 2G categories are shown. The left column shows the efficiencies of the q2BDT selection whereas the right shows the efficiencies obtained with the conventional q^2 window. The red vertical lines show the approximate position of the steps associated to the more conventional q^2 selection.

In light of the results, to model the effect of the q^2 cuts in the $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ mass distribution two smooth step functions are attached to the tails of the DSCB function for the 1G and 2G Bremsstrahlung categories and only one to the left tail for the 0G category. This is because in the 0G category only the low q^2 cut makes an impact. Different analytical functions for the steps have been tried. The ones giving the best results are the following:

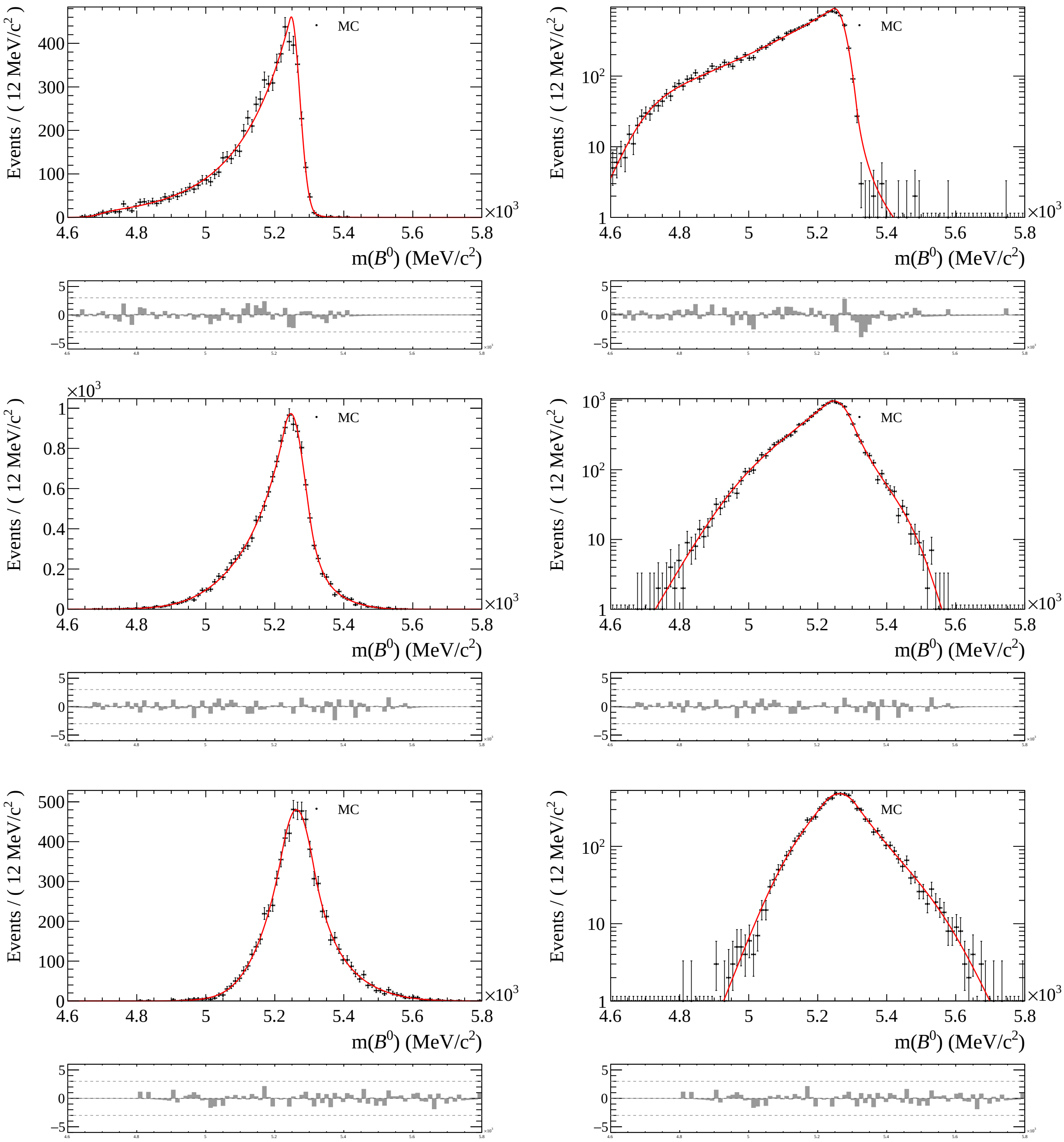


Figure 7.8: Final fits to the reconstructed B^0 mass using simulated Run1 $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ data selected in the $\psi(2S)$ region with the aid of the two q2BDTS introduced in section 5.3. From top to bottom we find the results for the 0G, 1G and 2G categories. Left and right columns show the same fits in linear and logarithmic scale, respectively. The lower panels show the pull histograms of the fits.

$$f_L(x; m_l, \sigma_l) = \frac{1}{1 + e^{-\frac{x-m_l}{\sigma_l}}}, \tag{7.9}$$

$$f_R(x; m_r, \sigma_r) = \frac{1}{1 + e^{\frac{x-m_r}{\sigma_r}}}. \tag{7.10}$$

Both functions depend on two parameters only. The parameters $m_{l,r}$ control the position of the steps and $\sigma_{l,r}$ their width. After the inclusion of these two step functions to the sides of the DSCB function we are left with a PDF depending on 10 independent parameters for the 1G and 2G Bremsstrahlung categories and 8 for the 0G category. However, not all of the parameters are allowed to float in the final mass fits. Parameters controlling the shapes of the tails in the DSCB function are fixed from the results of the mass fits performed without the q^2 selection. In the end, the only free parameters are the mean and width of the DSCB gaussian core and the parameters controlling the position and width of the steps.

The model is tested using simulated $B^0 \to K^{*0}e^+e^-$ data, selected in the same way as before but now including the q2BDT selection. Results are shown in figure 7.8 for the Run1 period. They are seen to be consistent for other data-taking periods. The model also shows good performance against modifications in the MVA and PID cuts (see section 7.4) as well as when releasing some of the exclusive background vetoes, confirming its robustness and flexibility.

7.3.2 Mass fits to simulated data selected in the high-q^2 region

The model developed in previous section is tested int he high-q^2 region using MC simulated $B^0 \to K^{*0}e^+e^-$ data to which full event selection is applied. This includes the exclusive background vetoes and the tighter MVA cuts introduced in section 6.2. Data is selected in the high-q^2 region using the nominal q2BDT variable with the cut defined in section 5.3 (q2BDT$_{\psi(2S)} > 0.95$) and fitted following the strategy developed for the $\psi(2S)$ region:

- Perform a fit to the reconstructed B^0 mass using simulated data without the q^2 selection. From the results we fix the parameters of the DSCB tails.

- Include the smooth-step functions into the model and perform a mass fit to simulated selected in the high-q^2 region.

The only difference is that here we only need a step to the the left hand side of the DSCB function. This is because now there is only a q^2 cut that needs to be modelled. Results are shown in figure 7.9 for the Run2p2 period.

As before, the model has been tested under different conditions and data-taking periods and it always manages to describe data properly. Similarities between the signal mass shape in the high-q^2 region and in the $\psi(2S)$ region are evident.

7.4 Constraining the yields of leaking backgrounds

A central part of this thesis deals with the development of a strategy to control the two leakage background contributions coming from the $\psi(2S)$ resonance that populate the high-q^2 region. The yields of these components need to be constrained before unblinding the mass fits in order to make them stable. Especially important is the handle on the partially reconstructed $B^0 \to X\psi(2S)(\to e^+e^-)$ yield since it populates the region below the signal and correlations between their yields can very easily occur.

As in the studies presented in the previous sections, the $\psi(2S)$ resonance will play a very important role here, not only because it contains the leaking components that we want to constrain in the mass fits performed in the high-q^2 region but also because it suffers from a similar contamination

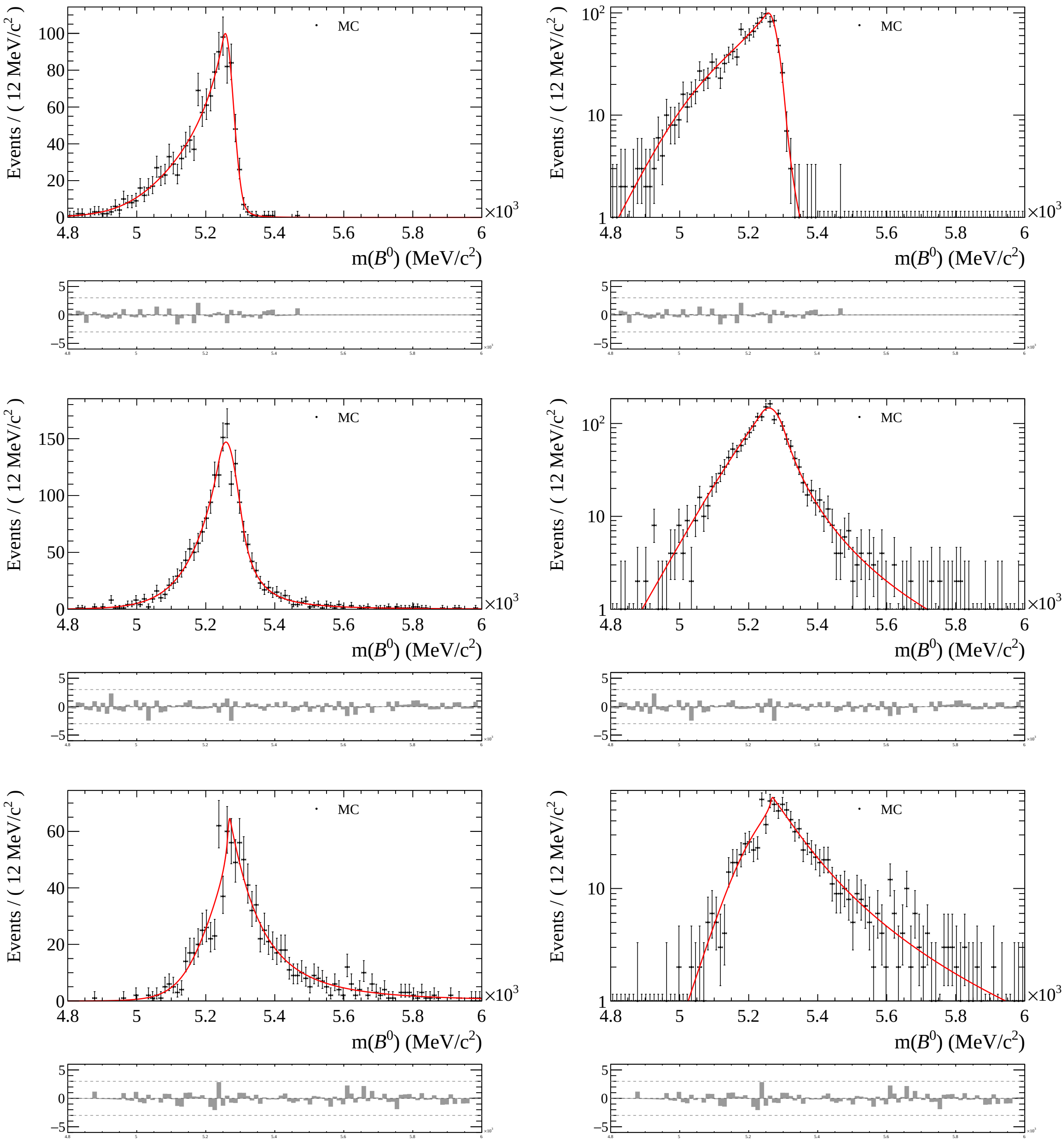


Figure 7.9: Fits to the reconstructed B^0 mass using simulated Run2p2 $B^0 \to K^{*0}e^+e^-$ data in the high-q^2 region selected with the q2BDT as explained in the text. From top to bottom we find the results for the 0G, 1G and 2G categories. Left and right columns show the same fits in linear and logarithmic scale, respectively. The lower panels show the pull histograms of the fits.

from leaking background candidates coming from the J/ψ region. It should be noted however that the analogy between the $\psi(2S)$ and high-q^2 regions is limited in a number of ways. First and most importantly, statistics in the reconstructed B^0 mass fits performed in the $\psi(2S)$ resonance region are much larger than in the fits to the rare channel at high-q^2 values. In particular, the signal branching fraction is much larger, which translates into a larger signal to background ratio. This has the direct implication that any possible mismodelling in the background components potentially leading to an incorrect estimation of the amount of background has a significantly lower impact on the signal yield. All this contributes to making the fits done in the $\psi(2S)$ region considerably simpler.

A second difference has to do with the impact of the distortion effects coming form the q^2 selection which are less problematic in the $\psi(2S)$ region since they only become noticeable at lower reconstructed B^0 masses as compared with the high-q^2 region. This has already been shown in section 7.2. It is necessary to go down to very low masses in order to see the turn-on effect in the combinatorial background mass shape in the $\psi(2S)$ region. For a direct comparison we can look at the reconstructed B^0 mass shape of combinatorial background expected from SS data in the high-q^2 region shown in figure 6.5 when discussing the effect of the over reconstructed background cut. There, we see the turn-on effect appearing very close to the B^0 mass. As explained in section 4.1, this is caused by the edge of the physical region shown in figure 4.3 being further away from the B^0 mass in the $\psi(2S)$ region. As a result, the shape of fit components and specially of the combinatorial background become easier to model in the $\psi(2S)$ region than in the high-q^2 region.

One last difference has to do with the location of the leaking backgrounds in the mass fit. The $\psi(2S)$ region is further away from the J/ψ region in the q^2 spectrum than the high-q^2 region is from the $\psi(2S)$ resonance. This has the direct consequence that leaking backgrounds coming from the J/ψ region are more displaced towards larger B^0 masses. Hence, they are better separated from the signal, making them easier to constrain using the right sideband of the B^0 mass which contains a smaller signal component (see results in following sections).

All these differences largely simplify the fits done in the $\psi(2S)$ region with respect to the high-q^2 region. It is therefore a perfect benchmark region to test and validate the strategy designed for the fits performed in the high-q^2 region allowing for a large number of cross-checks to be made. The similarities with the high-q^2 region are still large enough to let us easily extrapolate results from the $\psi(2S)$ region as it has already been demonstrated for the signal model presented in section 7.3.

In this section we will explain the full strategy to constrain the leaking backgrounds in the high-q^2 region. This will require a number of steps which are outlined below:

- First of all we perform fits to the kinematically constrained reconstructed B^0 mass in the J/ψ region. We want to use the signal and partially reconstructed background yield results to constrain the yields of the leaking backgrounds in the $\psi(2S)$ region.

- Then we make use of fully corrected and truth-matched MC simulated samples to constrain the yields of $B^0 \to K^{*0} J/\psi(\to e^+e^-)$ and $B \to X J/\psi(\to e^+e^-)$ candidates in the $\psi(2S)$ region.

- After that we perform fits to the reconstructed B^0 invariant mass in the $\psi(2S)$ region using the kinematically constrained B^0 mass variable and constraining the yields of the leaking backgrounds using the results of the previous step.

- Then the yield of $B^0 \to X\psi(2S)(\to e^+e^-)$ background candidates is constrained in a second fit to the reconstructed invariant mass of the B^0 meson in the $\psi(2S)$ region using the results of the fit to the kinematically constrained mass. Corrections from MC simulations are needed in order to translate the results from one fit to another since background efficiencies are different due to the different mass range and selections used. This step allows us to check if the measured $B^0 \to X\psi(2S)(\to e^+e^-)$ yield is compatible with a different fit setting for the same data, providing a powerful test of the result.

- The next step is to perform the new mass fit in the $\psi(2S)$ region without the kinematic constrain on the B^0 mass. This will require a re-calculation of the constraints on leaking backgrounds coming from the J/ψ region.

- Finally we compare the yield of $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ signal candidates measured using the two different fit methods. This is a very stringent test of the whole procedure since, given that both measurements are of the exact same observable and are performed using the same dataset but arranged in a different way[2], not only should they be compatible with each other but there exists a large correlation between the variables, increasing the significance of any possible disagreement. This is in complete analogy to what happens to the $R_{K^{*0}}$ observable, whose theoretical uncertainty is very small due to the large correlation between the QCD uncertainties affecting the muon and electron channels.

In all the steps, data is splitted into the usual three data-taking periods which are fitted independently. This serves as an internal validation of the fit performance allowing us to quickly spot any possible instabilities in the results.

As a cross-check, the same mass fits are performed using three different settings where the MVA and PID requirements are tightened with respect to the working point values. First, the electron PID cuts are tightened with respect to the ones introduced in the stripping requirements. The tighter electron PID cuts defined for this cross-check are:

$$\texttt{DLL}_{e\pi} > 3 \quad \& \quad \texttt{EProbNNe} > 0.4. \tag{7.11}$$

Performing the mass fits with tighter electron PID cuts tests the impact of any possible extra misidentified background not suppressed by the vetoes introduced in section 6.1. The second cross-check tightens the MVA cuts to the ones optimized for the rare modes. The aim of this second cross-check is two-fold. First, it aligns the MVA cuts in both J/ψ and $\psi(2S)$ regions and run periods, reducing the impact of mismodellings of the MVA response in MC simulated data employed to compute the efficiencies needed to constrain the yields of the leaking backgrounds populating the $\psi(2S)$ region. This allows to check for the presence of any possible bias and brings the fit configuration in the two resonant regions closer to the preliminarily configuration designed for the fits performed the high-q^2 region. On the other hand, it tests the stability of the results in terms of the amount of combinatorial and partially reconstructed backgrounds. This is very important in this analysis since one of the goals is to be able to constrain the amount of $B^0 \to X\psi(2S)(\to e^+e^-)$ partially reconstructed background leaking into the high-q^2 region which requires a reliable measurement of its yield. The whole fit procedure is repeated for every cross-check, including the efficiency calculations using MC simulated samples. In all the mass fits, the yields of the fit components are forced to be positive.

7.4.1 Mass fit in the J/ψ region

Fits to the reconstructed B^0 invariant mass in the J/ψ region are presented in this section. We use fully selected data recorded at LHCb for the mass fits and MC simulated data to get the mass distribution of the fit components. Simulated samples are truth-matched as explained in section

[2]The overlap between the datasets selected for the two fits is close to 100%.

5.2. Both data and MC simulated data are required to pass the exclusive background vetoes and MVA cuts introduced in chapter 6. The J/ψ region is selected as indicated in table 5.6 of section 5.3.

The procedure to get the final results is presented step by step, highlighting the most relevant details.

Fit settings

Data selected in the J/ψ region contain contributions from a large number of processes from a very different nature. The PDFs describing their mass distributions are either fixed or constrained in the final mass fits and the yields of known exclusive background components which survive the vetoes are constrained with respect to the signal yield using the formula given in equation 4.4. In the end, mass fits only contain 4 free floating parameters which are the yields of the hadronic and leptonic partially reconstructed background components, the combinatorial background yield and, of course, the signal $B^0 \to K^{*0}J/\psi(\to e^+e^-)$ yield.

Background components with either low statistics or complicated mass distributions are modeled with a *RooKeysPdf* fitted to MC simulated data. They include both hadronic and leptonic partially reconstructed background components as well as backgrounds coming from the exclusive $\Lambda_b^0 \to pKJ/\psi(\to e^+e^-)$, $B_s^0 \to \phi(\to K^+K^-)J/\psi(\to e^+e^-)$, $B^0 \to K^{*0}(K \leftrightarrow \pi)J/\psi(\to e^+e^-)$ and $B^0 \to K^{*0}\psi(2S)(\to \pi\pi J/\psi(\to e^+e^-))$ decays. The shape of the PDFs describing their mass distribution are taken from the fits to MC simulated data. When necessary, these fits are performed in an extended mass range in order to remove biases affecting the shape of their mass distribution near the edges on the fit range. The combinatorial background mass distribution is constrained from mass fits to SS data, extending the fit range with respect to the one chosen for the data fits in order to increase the amount of statistics.

The model chosen to describe the signal mass distribution consists of the sum of a CB and an *Ipatia2* function, which is a generalised version of the *Hypatia* function described in [76]. The *Hypatia* function is a generalization of the CB function, suited to scenarios where the core of the resonance peak does not follow a gaussian distribution. It deals with situations where signal candidates have non uniform uncertainties on the mass resolution. This can happen when performing fits to a kinematically constrained reconstructed mass since the effect of the kinematic constrain distorts the shape of the resonance peak and there is no reason to expect it to be well described by a gaussian function. The *Ipatia2* function adds a second tail to the *Hyptia* function making it a generalization of the BSCB distribution. It is a 9 parameter distribution with a hyperbolic central core describing the body of the resonance and two identical power law tails to the sides of it:

$$f_{Ipatia2}(x;m,\sigma,\lambda,\zeta,\beta,\alpha_L,n_l,\alpha_R,n_R) = \begin{cases} \frac{G(m-\alpha_R\sigma,m,\sigma,\lambda,\zeta,\beta)}{\left(1-\frac{x}{n_L G(\ldots)/G'(\ldots)-\alpha_L\sigma}\right)^{n_L}} & \text{for } \frac{x-m}{\sigma} \leq -\alpha_L \\ G(x;m,\sigma,\lambda,\zeta,\beta) & \text{for } -\alpha_L < \frac{x-m}{\sigma} < \alpha_R \\ \frac{G(m+\alpha_R\sigma,m,\sigma,\lambda,\zeta,\beta)}{\left(1+\frac{x}{n_R G(\ldots)/G'(\ldots)+\alpha_R\sigma}\right)^{n_R}} & \text{for } \frac{x-m}{\sigma} \geq \alpha_R. \end{cases} \tag{7.12}$$

Here, G is the the core function and G' is its derivative:

$$G(x;m,\sigma,\lambda,\zeta,\beta) = ((x-m)^2+\sigma^2A_\lambda^2(\zeta))^{\frac{1}{2}\lambda-\frac{1}{4}}e^{\beta(x-m)}K_{\lambda-\frac{1}{2}}\left(\zeta\sqrt{1+\left(\frac{x-m}{\sigma A_\lambda(\zeta)}\right)^2}\right), \tag{7.13}$$

$$A^2_\lambda(\zeta) = \frac{\zeta K_\lambda(\zeta)}{K_{\lambda+1}(\zeta)}. \tag{7.14}$$

The functions K_λ are the spherical Bessel functions of the third kind. As before, m and σ represent the mean and standard deviation of the core distribution and the parameters $\alpha_{L,R}$ and $n_{L,R}$ are analogous to the ones included in the definition of the DSCB function. The three new parameters are β, which controls the asymmetry of the core two parameters; and λ and ζ, which introduce a larger flexibility in the shape of the central peak. The parameter β will be fixed to 0, corresponding to the symmetric case and ζ, which showed to have very little impact on the final shape is set to $\zeta = 0.005$. This parameter has shown to improve the fit convergence and stability when taking values close to 0. All the parameters are positive with the exception of λ, which can take either positive or negative values.

The CB function is added to the model on top of the *Ipatia2* function in order to account for distortions caused by the q^2 selection. These distortions do not show the simple step-like behavior seen in section 7.3 for the $\psi(2S)$ region and are therefore harder to model. Different models to describe the signal mass distribution were tested but the CB+*Ipatia2* combination turned out to give the best and most stable results within the mass range employed for the fits. The relative fraction of the two components is left as a free parameter in the mass fits to simulated data and both CB and *Ipatia2* functions are forced to have the same mean. The overall PDF has therefore a total of 11 floating parameters to describe the mass shape of the $B^0 \to K^{*0} J/\psi(\to e^+e^-)$ signal, 7 of these parameters come from the *Ipatia2* function, 3 of them are the width and tail parameters of the CB and the last one is the relative fraction of the two components. Mass fits are performed independently for the three Bremsstrahlung categories. Results are then combined using fully corrected and truth-matched MC simulations to compute the relative fractions. All parameters are fixed from the fits to MC simulated data and only some freedom is left in the mean and width of the signal PDF which respectively include a shift and a scale factor that are allowed to float.

The resulting PDF after the mass fit for the $B^0 \to K^{*0} J/\psi(\to e^+e^-)$ signal is also used to model the shape of the $\overline{B}{}^0_s \to K^{*0} J/\psi(\to e^+e^-)$ background component. The whole mass distribution is sifted to the right by the mass difference between the B^0 and B^0_s mesons, taken from PDG. This can be done because the final state is identical for the two processes so that the same detector resolution effects affecting their mass shape are expected.

The mass range is chosen to allow a large enough amount of statistics to measure the partially reconstructed background yields. The right sideband is used to constrain the mass shape of the combinatorial background. The lower limit rejects background contributions leaking from the $\psi(2S)$ region, which populate the low reconstructed invariant mass region of data selected in the J/ψ region. This is due to inefficiencies in the Bremsstrahlung recovery procedure leading to candidates where the energy of the electrons has not been fully recovered. The chosen range is the following:

$$4800 < m(K\pi ee)^{DTF-J/\psi} < 6200 \text{ MeV}/c^2. \tag{7.15}$$

Fit components

In the following, the results of the mass fits performed to get the mass distribution of the components used to describe data in the J/ψ region are presented. All results are shown using the

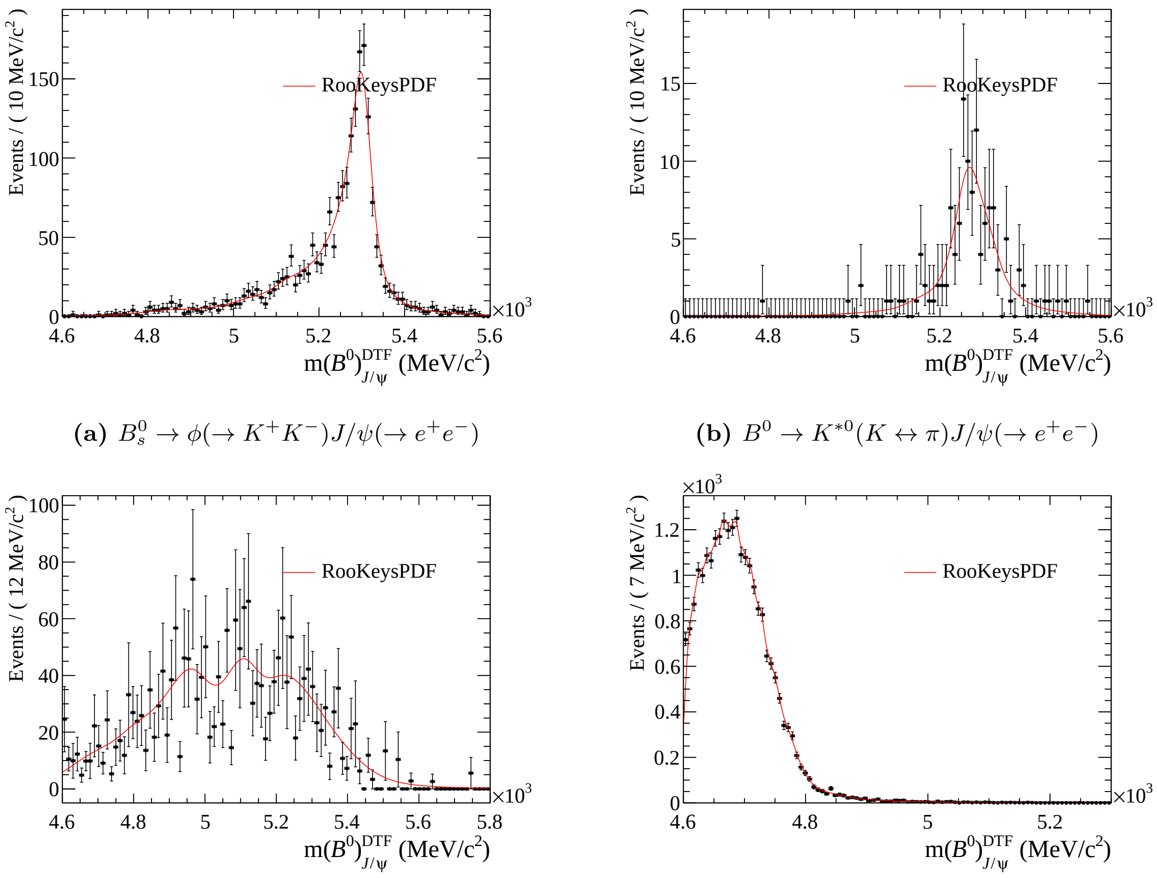
Run2p2 period as a reference, results from other data-taking periods are fully compatible.

Mass fits to MC simulated data describing the exclusive background components are shown in figure 7.10. The partially reconstructed background components are built by combining the $B^0 \to XJ/\psi(\to e^+e^-)$, $B^+ \to XJ/\psi(\to e^+e^-)$ and $B_s^0 \to XJ/\psi(\to e^+e^-)$ inclusive MC samples selected following the truth-matching procedure described in section 5.2. Mass fits are performed independently for each of the samples and then combined into an overall PDF describing the mass distribution of the fully inclusive partially reconstructed background component . It is defined as the weighted sum of the PDFs fitted for each component, with the weights calculated using the following formula:

$$w_i \propto \frac{f_i}{f_u} \cdot \frac{\mathcal{B}(\text{Exclusive decay})}{f_{DEC}(\text{Exclusive decay})} \cdot \frac{N_{MC|sel}}{N_{MC|gen}}. \tag{7.16}$$

The normalisation of the weights is computed imposing the following requirement:

$$\sum_i w_i = 1. \tag{7.17}$$



(a) $B_s^0 \to \phi(\to K^+K^-)J/\psi(\to e^+e^-)$

(b) $B^0 \to K^{*0}(K \leftrightarrow \pi)J/\psi(\to e^+e^-)$

(c) $\Lambda_b^0 \to pKJ/\psi(\to e^+e^-)$

(d) $B^0 \to K^{*0}\psi(2S)(\to \pi\pi J/\psi(\to e^+e^-))$

Figure 7.10: Fits to the kinematically constrained B^0 mass in MC simulated data for the exclusive backgrounds considered in the fits performed in the J/ψ region. The background components are indicated in the labels.

Sample	Decay	$\mathcal{B}$	f_{DEC}	$\frac{f}{f_u}$	$\frac{N_{sel}}{N_{gen}} \times 10^{-5}$	w
$B^0 \to XJ/\psi$	$B^0 \to K^{*0}J/\psi$	1.27×10^{-3}	0.1885	1	26.2	0.57
$B^+ \to XJ/\psi$	$B^+ \to K^+J/\psi$	1.01×10^{-3}	0.1596	1	20.5	0.42
$B_s^0 \to XJ/\psi$	$B_s^0 \to \phi(\to K^+K^-)J/\psi$	0.53×10^{-3}	0.1077	0.244	1.2	0.01

Table 7.2: Calculated weights for the components in the hadronic partially reconstructed PDF using Run2p2 simulated data.

Sample	Decay	$\mathcal{B}$	f_{DEC}	$\frac{f}{f_u}$	$\frac{N_{sel}}{N_{gen}} \times 10^{-5}$	w
$B^0 \to XJ/\psi$	$B^0 \to K^{*0}J/\psi$	1.27×10^{-3}	0.1885	1	4.69	0.99
$B^+ \to XJ/\psi$	$B^+ \to K^+J/\psi$	1.01×10^{-3}	0.1596	1	0.01	~ 0
$B_s^0 \to XJ/\psi$	$B_s^0 \to \phi(\to K^+K^-)J/\psi$	0.53×10^{-3}	0.1077	0.244	0.36	0.01

Table 7.3: Calculated weights for the components in the leptonic partially reconstructed PDF using Run2p2 simulated data.

The first factor in equation 7.16 is the ratio of hadronization fractions of the B meson flavors responsible for each of the decays. The fraction in the denominator is the one of the B^+ meson (an $u\bar{d}$ flavor state) which is by convention used as the normalization channel. This factor is necessary because the hadronization process is not accounted for in the generation of the decay samples. The second factor takes the branching fraction of an exclusive decay contained in the sample and normalizes it by the relative fraction of that exclusive decay in the inclusive simulated cocktail. The latter is taken from the `DecFiles` passed to the `EVTGEN` software to generate the sample. The decay with the largest branching fraction is used in each case for the calculation. The last factor is the selection efficiency, defined as the total number of events contained in the sample within the fit range after full selection divided by the number of generated events, also taken from the `DecFiles`. Tables 7.2 and 7.3 show the weights calculated for the hadronic and leptonic components in the Run2p2 period, selected with the nominal PID requirements and MVA cuts optimized for that q^2 region and data-taking period. Similar results are found in other periods. The major difference is found in the production fraction of B_s^0 with respect to B^0 and B^+ which is slightly smaller in Run2 due to the larger collision energy. The hadronization fractions of B^0 and B^+ are approximately equal in all runs as a consequence of *isospin symmetry* and are assumed to be identical for this calculation. Decays used for the normalization correspond to processes which are removed from the inclusive samples in the truth-matching. However, they can still be used as control channels to compute the expected relative contribution of each partially reconstructed background type.

For the hadronic part both $B^+ \to XJ/\psi(\to e^+e^-)$ and $B^0 \to XJ/\psi(\to e^+e^-)$ contributions are roughly of the same order whereas the leptonic sample is clearly dominated by the $B^0 \to XJ/\psi(\to e^+e^-)$ contribution. This can be understood as a consequence of isospin symmetry. For a B^+ decay candidate to pass the event selection and be reconstructed as a signal event, the hadronic part needs to have at least one missing non reconstructed particle since the hadronic part of any final state of a B^+ decay needs to have a positive electric charge. Hence, leptonic partially reconstructed processes originating from a B^+ decay will contain missing particles from both the leptonic and hadronic part, making the component peak at very low masses. This is not the case for B^0 and B_s^0 decays which can generate partially reconstructed processes passing the general event selection where only the leptonic part has missing particles, populating regions closer to the B^0 mass and therefore generating a larger contribution to the mass fits.

Mass fits for the leptonic and hadronic partially reconstructed backgrounds are shown in figure 7.11.

Figure 7.12 shows the final PDF after combining all the components with the corresponding weights.

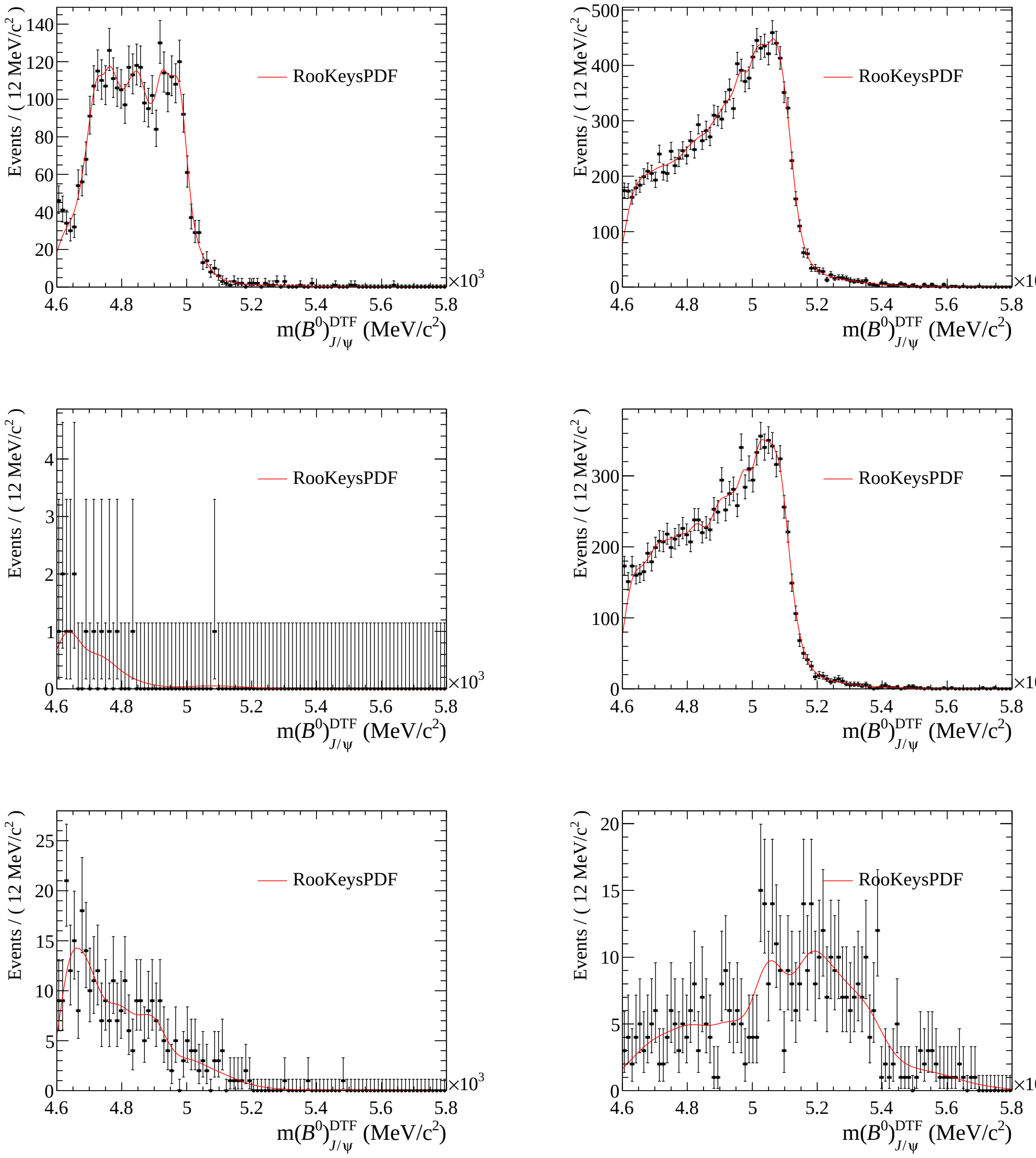


Figure 7.11: Fits to the kinematically constrained B^0 mass in MC simulated data for the partially reconstructed background samples selected in the J/ψ region. From the first to the third row we find the fits for the $B^0 \to XJ/\psi(\to e^+e^-)$, $B^+ \to XJ/\psi(\to e^+e^-)$ and $B^0_s \to XJ/\psi(\to e^+e^-)$ background components. Partially reconstructed leptonic (hadronic) components are shown in the left (right) column.

The mass fit to Run2p2 SS data used to constrain the shape of the combinatorial background mass distribution is shown in figure 7.13. The points within the B^0 mass window removed by the veto are not included in the fits. We can see that data is very well described by the exponential model in this fit range since the effect of the q^2 selection is only relevant at very low reconstructed B^0

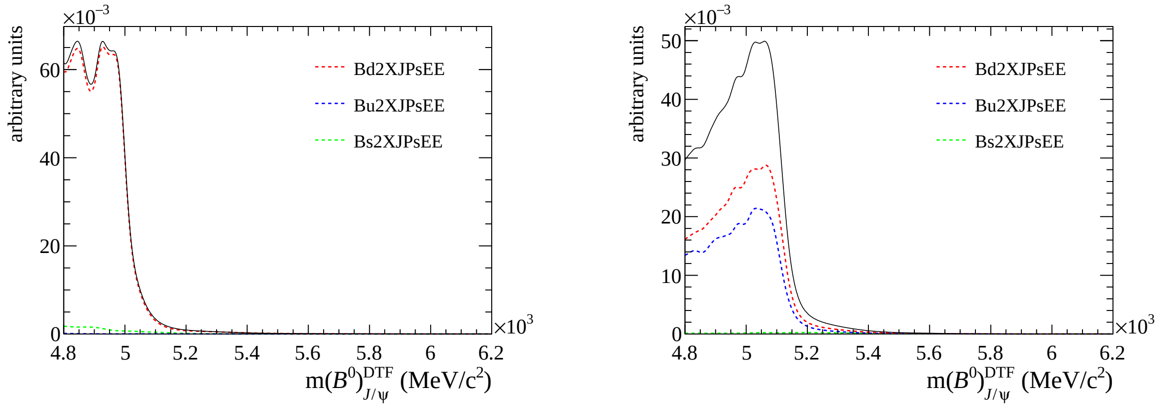


Figure 7.12: Combined PDF to describing the mass distribution of partially reconstructed backgrounds. The leptonic component is shown in the left hand side, the plot to the right shows the hadronic component. Contributions to each PDF are indicated in the legend.

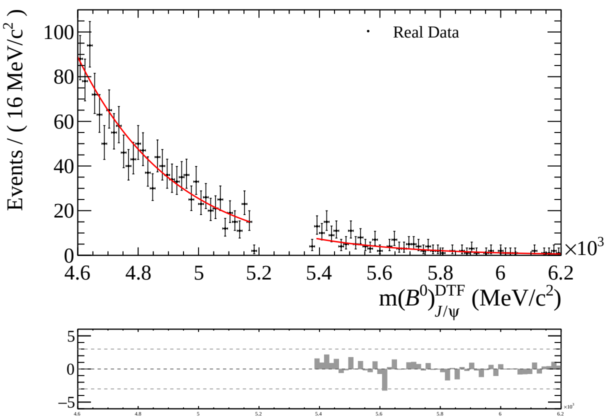


Figure 7.13: Fit to the kinematically constrained B^0 mass using SS data selected in the Run2p2 period. The vetoed points are excluded from the fit. The lower panel shows the pull histogram of the fit.

mass in the J/ψ region.

Finally, $B^0 \to K^{*0} J/\psi(\to e^+e^-)$ signal mass fits to MC simulated data are shown in figure 7.14 for the tree separated Bremsstrahlung categories. Simulated data is well described by the model in all the fits with only small fluctuations near the tails, which are harder to model, but with negligible impact on the results. Effects from Bremsstrahlung recovery seen in the previous section are washed out by the kinematic constrain on the B^0 mass.

Fit to data

Once we have the PDFs to describe the mass distribution of all the fit components we can bring everything together and perform the mass fits in real data. Table 7.4 presents a summary of the fit components used, including a short explanation about how each of them are implemented and the labels used in the legends to refer to them.

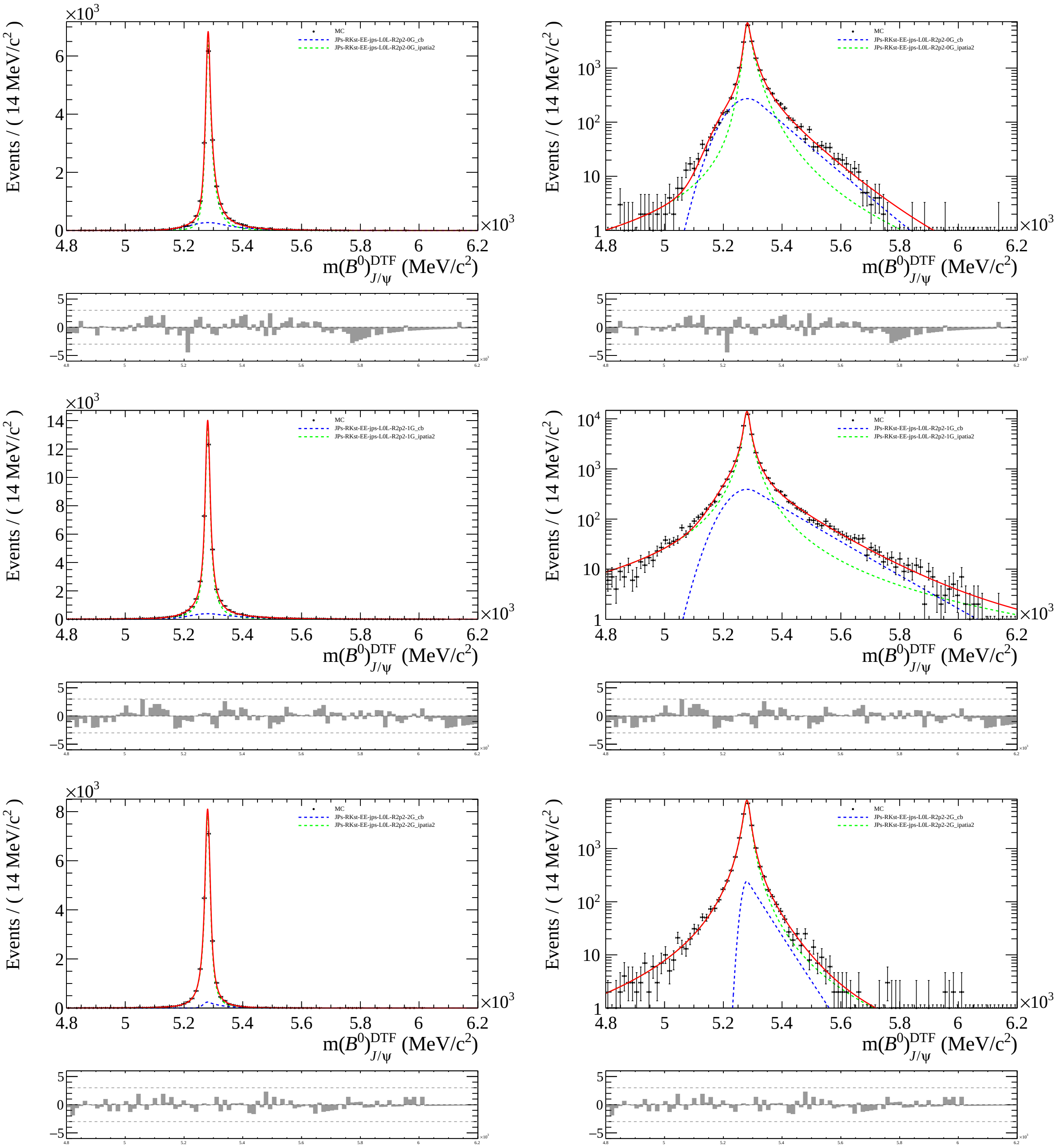


Figure 7.14: Fits to the kinematically constrained B^0 mass in simulated Run2p2 $B^0 \to K^{*0} J/\psi(\to e^+e^-)$ data selected in the J/ψ region. From top to bottom we find the 0G, 1G and 2G results. Each fit is shown twice, in linear scale in the left column and in logarithmic scale in the right column. components are indicated in the legend with the CB function represented by the blue dashed line and the *Ipatia2* function by the green dashed line. The lower panels show the pull histograms of the fits.

Fits to data are shown in figure 7.15 using the nominal PID and MVA cuts. Data is well described by the model and results are consistent among data-taking periods. In spite of this, a large correlation is found between the combinatorial and partially reconstructed background yields. It does not have a big impact on the signal yield but could lead to a wrong estimation of the amount of partially reconstructed background. This could be a problem since we want to use it to constrain

Label	Component	Model	Implementation
Signal	$B^0 \to K^{*0} J/\psi(\to e^+e^-)$	*Ipatia2* + CB fitted to MC simulated data splitted into the three Bremsstrahlung categories. The three categories are combined into a single PDF weighted by the Bremsstrahlung fraction, computed in simulated data.	Shape fixed from a fit to MC simulated data with a shift on the mean and a scale factor on the width which are floating in the data fits. The yield is free.
CombSS	Combinatorial background	Exponential function fitted to SS data.	Exponential slope constrained from the fit to SS data. The yield is free.
PartRecoL	$B \to XJ/\psi(\to e^+e^-)$-leptonic	KDE fitted independently to MC simulated data for the B^0, B_s^0 and B^+ components. Fit functions are combined using weights as explained in the text.	Shape taken from the fits to simulated data. The yield is free.
PartRecoH	$B \to XJ/\psi(\to e^+e^-)$-hadronic	KDE fitted independently to MC simulated data for the B^0, B_s^0 and B^+ components. Fit functions are combined using weights as explained in the text.	Shape taken from the fits to simulated data. The yield is free.
Lb	$\Lambda_b^0 \to pKJ/\psi(\to e^+e^-)$	KDE fitted to MC simulated data.	Shape taken from the fits to simulated data. The yield is constrained with respect to the signal yield.
Bs2Phi	$B_s^0 \to \phi(\to K^+K^-)J/\psi(\to e^+e^-)$	KDE fitted to MC simulated data.	Shape taken from the fits to simulated data. The yield is constrained with respect to the signal yield.
HadSwap	$B^0 \to K^{*0}(K \leftrightarrow \pi)J/\psi(\to e^+e^-)$	KDE fitted to simulated data.	Shape taken from the fits to simulated data. The yield is constrained with respect to the signal yield.
Psi2JPsX	$B^0 \to K^{*0}\psi(2S)(\to \pi\pi J/\psi(\to e^+e^-))$	KDE fitted to simulated data.	Shape taken from the fits to simulated data. The yield is constrained with respect to the signal yield.
Bs	$\overline{B}_s^0 \to K^{*0} J/\psi(\to e^+e^-)$	Copied from the signal model.	Signal PDF shifted by the mass difference between B^0 and B_s^0 and. The yield is constrained with respect to the signal yield.

Table 7.4: Summary of the fit components.

the yield partially reconstructed background $B \to XJ/\psi(\to e^+e^-)$ component leaking into the $\psi(2S)$ region. A lot of effort has been devoted to try to reduce this correlation combining the two partially reconstructed background components into a single one, modifying the mass range used in the fit or modelling the combinatorial background using a free exponential PDF. The final settings have been chosen to maximize stability within data-taking periods.

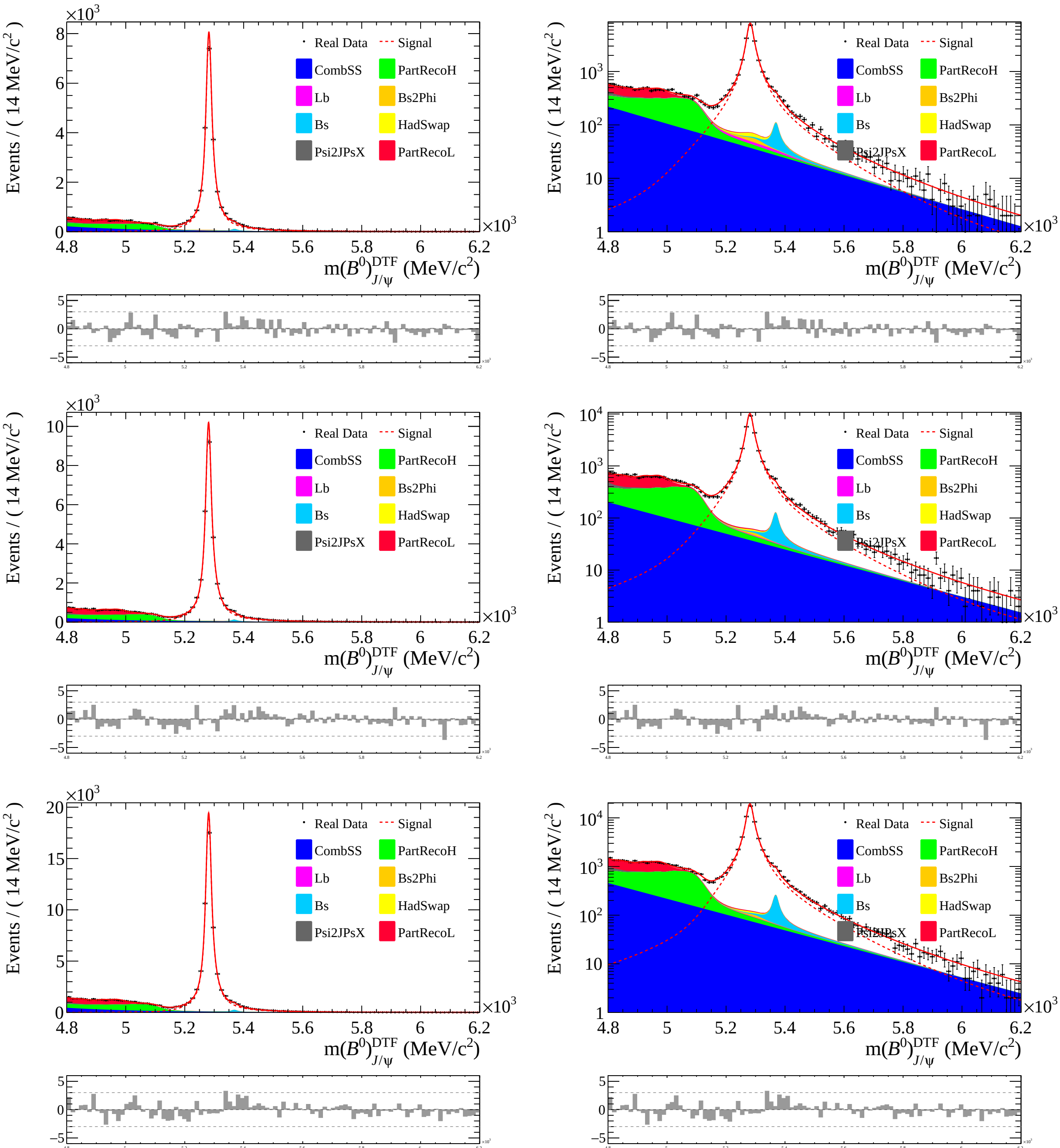


Figure 7.15: Fits to the kinematically constrained B^0 mass using fully selected real data in the J/ψ region. From up to bottom fits are shown for Run1, Run2p1 and Run2p2 data. In each row the same fit is shown twice, using a linear scale for the y axis (left) and a logarithmic scale (right). Components are indicated in the legend with backgrounds stacked on top of each other for a better visualization. The signal component is presented as a non stacked red dotted line. The lower panels show the pull histograms of the fits.

The amount of combinatorial background extracted from the fits is consistently larger than the amount of SS data selected under the same conditions in all the data-taking periods. Taking the Run2p2 period as a reference, 5698 ± 572 event candidates were fitted as part of the combinatorial background. This is about 6 times larger than the yield of SS data, measured to be 895 ± 30. This observation together with the large correlation between the combinatorial background and partially reconstructed background yields motivate the cross-check of the results using tighter MVA cuts in order to better separate the two components. When applying the tighter cuts the difference between the yields of combinatorial background and SS data is reduced to roughly 2.5 more combinatorial background than SS data in the Run2p2 period. The main difference is that by applying the tighter MVA cuts we remove most of the combinatorial background. As a consequence, the large correlation between the combinatorial background yield and the yield of partially reconstructed backgrounds is reduced, improving the estimation of the yields.

7.4.2 Fit to the kinematically constrained B^0 mass in the $\psi(2S)$ region

Following the strategy presented before, the next step is to fit the kinematically constrained B^0 mass in the $\psi(2S)$ region, constraining the yields of backgrounds leaking from the J/ψ region. As before, data is fully selected including the exclusive background vetoes and MVA requirements and applying the q^2 selection presented in table 5.6.

Fit settings

The basic fit settings are kept as close as possible to the ones for the fits performed in the J/ψ region. The mass distributions of background components are modelled in a very similar way and the strategy to constrain their yields is also completely analogous to the strategy followed in the J/ψ region. We have now only 3 free floating parameters: the combinatorial background yield, the partially reconstructed $B^0 \to X\psi(2S)(\to e^+e^-)$ background yield and the signal yield.

The shape of the combinatorial background mass distribution is constrained from a fit to SS data. The fit is done in an extended mass region using the exponential function with a turn-on term introduced in equation 7.5. The mass range is chosen to cover the full SS data mass distribution in order to be able to properly model the turn-on effect.

The model employed to describe the signal component combines features from the mass fits in the J/ψ region and the signal model introduced in section 7.3, developed to describe the signal mass shape in the $\psi(2S)$ region when the kinematic constrain in the reconstructed B^0 mass is not used. The core of the PDF is described by an *Ipatia2* function with step functions added to the tails in order to model the effects generated by the q^2 selection, which are stronger here than in the J/ψ region. Again, independent mass fits are performed in each Bremsstrahlung category and then combined to get the signal PDF. In the mass fit to data, the mean of the resulting distribution includes a shift and the width is multiplied by a scale factor which are both allowed to float in order to have more flexibility.

The main complication of the fits performed in the $\psi(2S)$ region with respect to the J/ψ region are the two J/ψ leaking background components. Their shapes are fixed from mass fits to MC simulated data, truth-matched according to the procedure presented in section 5.2 and fitted using a *RooKeysPdf*. The partially reconstructed background leaking component is in this case not splitted into its leptonic and hadronic parts as it was done in the J/ψ region since the lower statistics on the samples here does not allow for reliable mass fits if they are performed separately. In order to constrain the yield of the partially reconstructed leaking background we first need to calculate the total amount of partially reconstructed background in the J/ψ region for each data-taking period. We do it by combining the leptonic and hadronic partially reconstructed background yields resulting from the fits performed there, taking into account their correlation for the calculation of the uncertainty. In order to define the constraint we need to compute the efficiencies in both q^2 regions, taking into account the mass ranges used in the fits from where the yields are extracted. The efficiencies are defined as follows:

$$\varepsilon = \frac{N_{MC|sel}}{N_{MC|gen}}. \tag{7.18}$$

In this equation the numerator contains the number of selected events form the MC simulated samples after all requirements are applied. The number in the denominator is the amount of *generator level* events, taken from the corresponding `DecFiles`. It is the total amount of generated events in the simulated sample. Leaking $B^0 \to K^{*0} J/\psi(\to e^+e^-)$ candidates are constrained in the same way using the signal yield from the mass fits performed in the J/ψ region. Relative efficiencies and yields are shown in table 7.5 for $B^0 \to K^{*0} J/\psi(\to e^+e^-)$ and in table 7.6 for the partially reconstructed $B \to XJ/\psi(\to e^+e^-)$ component.

Relative efficiencies of partially reconstructed background components are almost identical in different run periods for two reasons. First, as indicated in section 5.2, MC samples used to model partially reconstructed backgrounds coming from the J/ψ region are the same in all periods (they are the ones generated for the Run2p2 period). Secondly, the main difference between runs that could affect the relative efficiency are the different MVA cuts applied in the $\psi(2S)$ region. However, as explained in section 7.1, MVA cuts are less efficient against leaking partially reconstructed backgrounds and therefore their effect on them is minimized.

The mass range used in the fit is chosen to allow a large enough amount of partially reconstructed background on the left B^0-mass sideband as well as leaking $B^0 \to K^{*0} J/\psi(\to e^+e^-)$ background on the right sideband. Having a large enough amount of partially reconstructed background is necessary in order to be able to use the yield extracted from the fit to constrain its yield in the high-q^2 region. The $B^0 \to K^{*0} J/\psi(\to e^+e^-)$ background located to the right of the B^0 mass allows us to check the performance of our strategy to constrain leaking background components. We use the following mass range:

$$4900 < m(K\pi ee)^{DTF-\psi(2S)} < 6200 \text{ MeV}/c^2. \tag{7.19}$$

Period	$\mathcal{N}^{J/\psi}$	$\frac{\varepsilon^{\psi(2S)}}{\varepsilon^{J/\psi}} \times 10^{-2}$	$\mathcal{N}^{\psi(2S)}$
Run1	15482 ± 147	1.139 ± 0.032	176.3 ± 5.2
Run2p1	28564 ± 195	1.278 ± 0.030	365.1 ± 9.0
Run2p2	53931 ± 270	1.344 ± 0.043	725 ± 24

Table 7.5: Constrained values of the yield of $B^0 \to K^{*0} J/\psi(\to e^+e^-)$ candidates for the fits performed using the kinematically constrained B^0 mass in the $\psi(2S)$ region.

Period	$\mathcal{N}^{J/\psi}_{lep}$	$\mathcal{N}^{J/\psi}_{had}$	$\mathcal{N}^{J/\psi}_{total}$	$\frac{\varepsilon^{\psi(2S)}}{\varepsilon^{J/\psi}} \times 10^{-2}$	$\mathcal{N}^{\psi(2S)}_{total}$
Run1	1603 ± 185	2779 ± 154	4382 ± 263	5.21 ± 0.94	226 ± 43
Run2p1	3055 ± 231	5368 ± 203	8423 ± 327	5.23 ± 0.95	434 ± 80
Run2p2	6125 ± 333	11362 ± 288	17488 ± 472	5.22 ± 0.95	901 ± 165

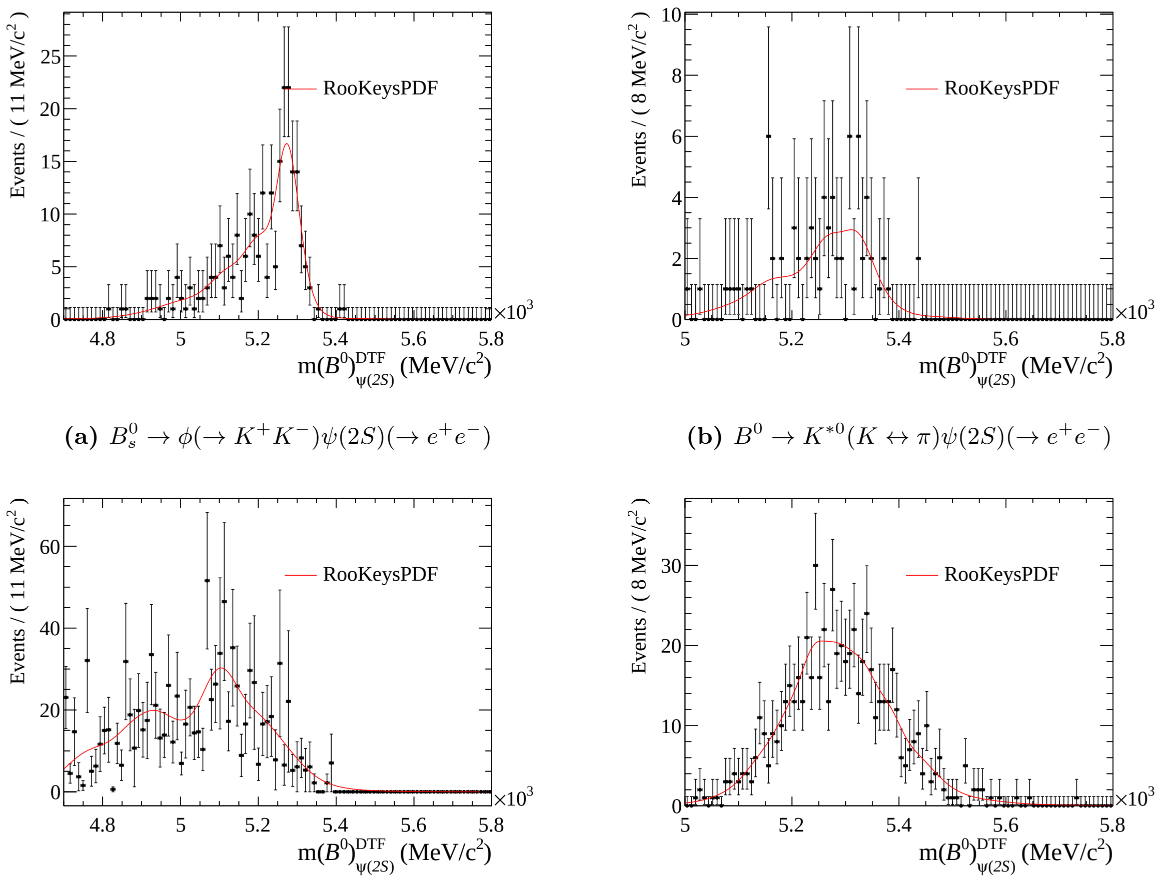
Table 7.6: Constrained values of the yield of $B^0 \to XJ/\psi(\to e^+e^-)$ candidates for the fits performed using the kinematically constrained B^0 mass in the $\psi(2S)$ region.

Fit components

All results shown in this section are obtained using either data or simulated data selected in the $\psi(2S)$ region under the Run2p2 data-taking conditions. Full event selection is applied including the background vetoes and the nominal MVA cuts optimized for the Run2p2 period in the $\psi(2S)$ q^2 bin. Figure 7.16 shows the mass fits performed to get the mass distribution of the exclusive background components. The mass fits to simulated data used to get the leaking $B^0 \to K^{*0}J/\psi(\to e^+e^-)$ background and partially reconstructed background $B^0 \to X\psi(2S)(\to e^+e^-)$ PDFs are shown in figure 7.17.

We are preliminarily only using the B^0 component to model partially reconstructed $\psi(2S)$ background candidates. As only the hadronic component is relevant in processes involving a $\psi(2S)$ and this component is very similar in both B^+ and B^0 samples we can absorb both components into the B^0 contribution. The similarity between the two components has been checked by comparing the $B^0 \to X\psi(2S)(\to e^+e^-)$ and $B^+ \to K^+\pi^+\pi^-\psi(2S)(\to e^+e^-)$ mass distributions. The contribution of the B_s^0 component is expected to be small and can be neglected in a first approximation.

The strategy to model the partially reconstructed background components leaking from the J/ψ



(a) $B_s^0 \to \phi(\to K^+K^-)\psi(2S)(\to e^+e^-)$

(b) $B^0 \to K^{*0}(K \leftrightarrow \pi)\psi(2S)(\to e^+e^-)$

(c) $\Lambda_b^0 \to pK\psi(2S)(\to e^+e^-)$

(d) $B^0 \to K^{*0}\psi(2S)(\to \pi\pi J/\psi(\to e^+e^-))$

Figure 7.16: Fits to the kinematically constrained B^0 mass using MC simulated data for the exclusive backgrounds considered in the $\psi(2S)$ region. Components are indicated in the captions.

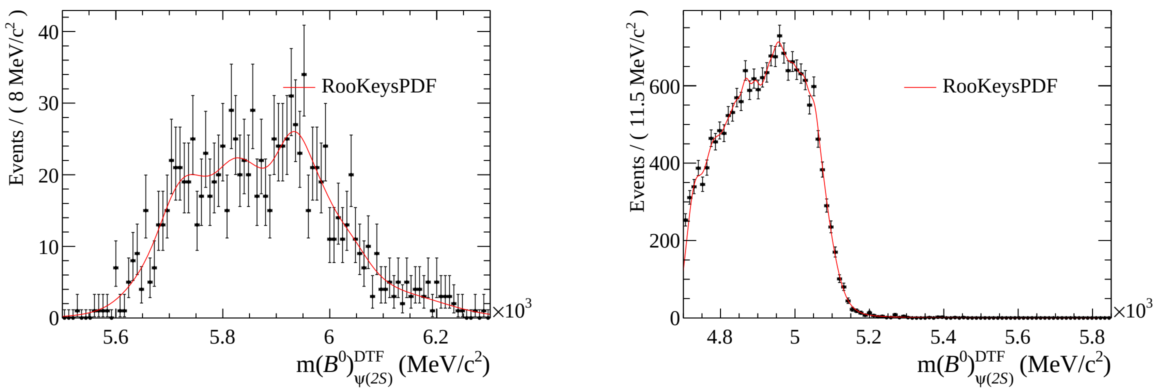


Figure 7.17: Fits to the kinematically constrained B^0 mass using MC simulated data for the $B^0 \to K^{*0}J/\psi(\to e^+e^-)$ (left) and $B^0 \to X\psi(2S)(\to e^+e^-)$ (right) backgrounds in the $\psi(2S)$ region.

Sample	Exclusive decay	$\mathcal{B}$	f_{DEC}	$\frac{f}{f_u}$	$\frac{N_{sel}}{N_{gen}} \times 10^{-5}$	w
$B^0 \to XJ/\psi$	$B^0 \to K^{*0}J/\psi$	1.27×10^{-3}	0.1885	1	0.463	0.63
$B^+ \to XJ/\psi$	$B^+ \to K^+J/\psi$	1.01×10^{-3}	0.1596	1	0.281	0.36
$B_s^0 \to XJ/\psi$	$B_s^0 \to \phi(\to K^+K^-)J/\psi$	0.53×10^{-3}	0.1077	0.244	0.0427	0.01

Table 7.7: Calculated weights for the components in the leaking partially reconstructed background PDF using R2p2 simulated data for the fits to the kinematically constrained mass in the $\psi(2S)$ region.

region is identical to the one followed in the J/ψ region. Individual components are weighted using the formula from equation 7.16 and then combined to form the total PDF. Results are shown in table 7.7. The mass fits to get the shape of each component and their weighted sum are shown in figure 7.18.

The combinatorial background mass shape is taken from a mass fit to SS data, excluding a window of ± 100 MeV/c^2 around the B^0 mass. The fit is shown in figure 7.19. The distortion effect caused by the q^2 selection is evident also in the constrained kinematic variable. It is perfectly described by the model.

Finally, the mass fits used to build the signal PDF are presented in figure 7.20 for each Bremsstrahlung category. The model is able to satisfactory describe data in all the fits, including the distortion effects in the tails of the distribution generated by the q^2 selection.

Fit to data

Table 7.8 summarizes the set of components employed in the data fit to the reconstructed kinematically constrained B^0 mass performed in the $\psi(2S)$ region. It includes information about the implementation of the components in the fit and the labels employed in the legend.

Mass fits to data recorded by LHCb and fully selected for the analysis using the nominal PID and MVA cuts are shown in figure 7.21. In general, data is well described in all data-taking periods, including the leaking backgrounds from the J/ψ region which populate the right hand side of the signal peak. Nevertheless, there is a clear discrepancy between the fitted background composition in the Run2p2 period as compared to the Run1 and Run2p1 data fits. The exponential slope of

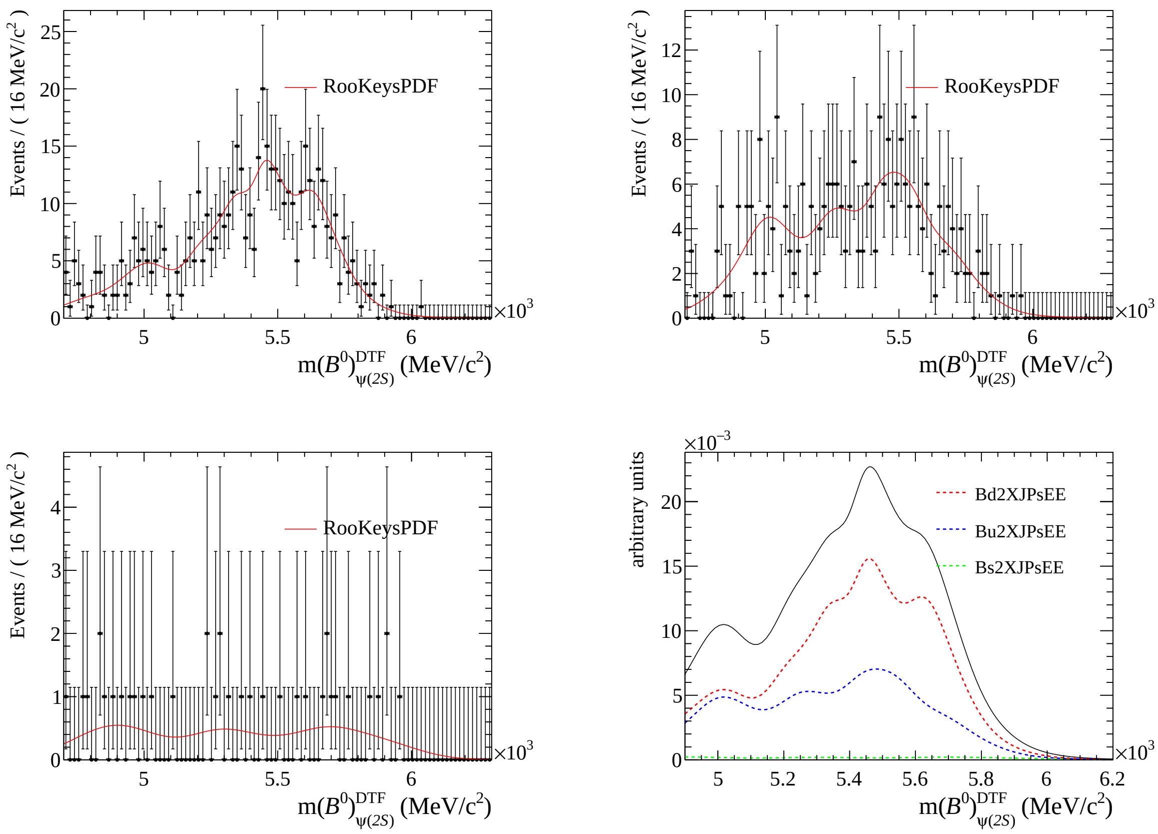


Figure 7.18: Fits to the kinematically constrained B^0 mass using simulated data from the Run2p2 period selected in the $\psi(2S)$ region. In the top row we find $B^0 \to XJ/\psi(\to e^+e^-)$ (left) and $B^+ \to XJ/\psi(\to e^+e^-)$ (right). The bottom row shows $B^0_s \to XJ/\psi(\to e^+e^-)$ (left) and the weighted sum of the three components (right).

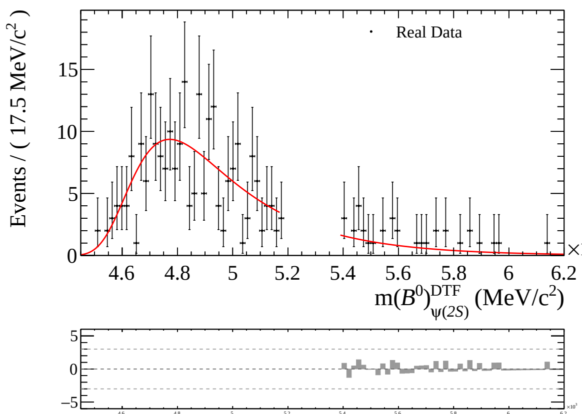


Figure 7.19: Fit to the kinematically constrained B^0 mass using fully selected SS data from the Run2p2 period selected in the $\psi(2S)$ region. Excluded candidates are not used in the fit. The lower panel shows the pull histogram of the fit.

the PDF used to describe the mass shape of the combinatorial background in the Run2p2 period is smaller, leading to a flatter distribution. This has an impact on the $B^0 \to X\psi(2S)(\to e^+e^-)$

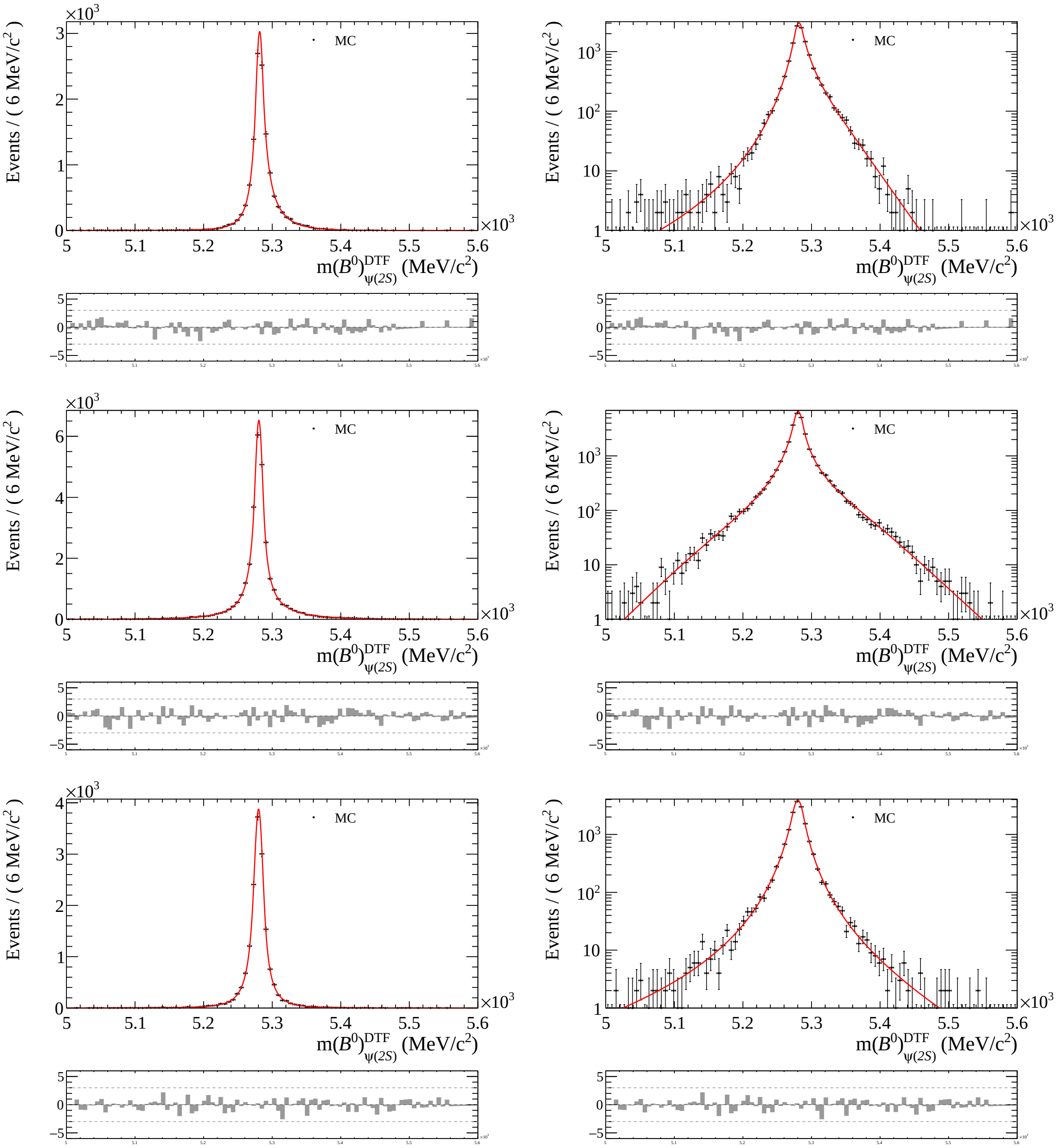


Figure 7.20: Fits to the kinematically constrained B^0 mass using fully selected simulated $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ data in the $\psi(2S)$ region corresponding to the Run2p2 period. From top to bottom we can see the fits in the 0G, 1G and 2G categories in linear scale in the left hand side and using a logarithmic scale for the y axis in the right hand side. The lower panels show the pull histograms of the fits.

partially reconstructed background yield, which becomes larger in the fit performed in the Run2p2 period. At the same time this reduces the signal yield. The fact that the amount of combinatorial background is wrongly estimated in the mass fits is clear when comparing the values extracted from the fits with the measured amounts of SS data. Values for the comparison are given in table 7.9. The combinatorial background yield is way too large in all data taking periods. It might seem a bit shocking that the selected amount of SS data candidates in the Run1 period is larger than

Label	Component	Model	Implementation
Signal	$B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$	*Ipatia2* with step functions attached to the tails. Fitted to MC data splitted into the three Bremsstrahlung categories. The tree fit results are combined computing the Bremsstrahlung fractions from fully corrected MC samples.	Shape fixed from the fit to simulated data with a shift on the mean and a scale factor on the width, both floating in the final mass fits. The yield is free.
CombSS	Combinatorial background	Exponential function with turn-on from equation 7.5 obtained from a mass fits to SS data.	Parameters of the PDF constrained from the SS data fit. The yield is free.
Bd2XPsi	$B^0 \to X\psi(2S)(\to e^+e^-)$	KDE fitted to simulated data.	Shape taken from the fit to simulated data. The yield is free.
PartReco	$B \to XJ/\psi(\to e^+e^-)$	KDE fitted to simulated data independently for the B^0, B^0_s and B^+ components. Fit functions are combined using the weights efined in the text.	Shape taken from the fits to simulated data. The yield constrained from the results extracted from the fit performed in the J/ψ region.
Lb	$\Lambda^0_b \to pK\psi(2S)(\to e^+e^-)$	KDE fitted to simulated data.	Shape taken from the fit to simulated data. The yield is constrained with respect to the signal yield.
Bs2Phi	$B^0_s \to \phi(\to K^+K^-)\psi(2S)(\to e^+e^-)$	KDE fitted to simulated data.	Shape taken from the fit to simulated data. The yield is constrained with respect to the signal yield.
HadSwap	$B^0 \to K^{*0}(K \leftrightarrow \pi)\psi(2S)(\to e^+e^-)$	KDE fitted to simulated data.	Shape taken from the fit to simulated data. The yield is constrained with respect to the signal yield.
Psi2JPsX	$B^0 \to K^{*0}\psi(2S)(\to \pi\pi J/\psi(\to e^+e^-))$	KDE fitted to simulated data.	Shape taken from the fit to simulated data. The yield is constrained with respect to the signal yield.
Bs	$\overline{B}^0_s \to K^{*0}\psi(2S)(\to e^+e^-)$	Copied from the signal fit model	Signal PDF shifted by the mass difference between B^0 and B^0_s. The yield is constrained with respect to the signal yield.
Leakage	$B^0 \to K^{*0}J/\psi(\to e^+e^-)$	KDE fitted to MC simulated data.	Shape taken from the fit to simulated data. The yield constrained from the results extracted from the fit performed in the J/ψ region.

Table 7.8: Summary of the fit components.

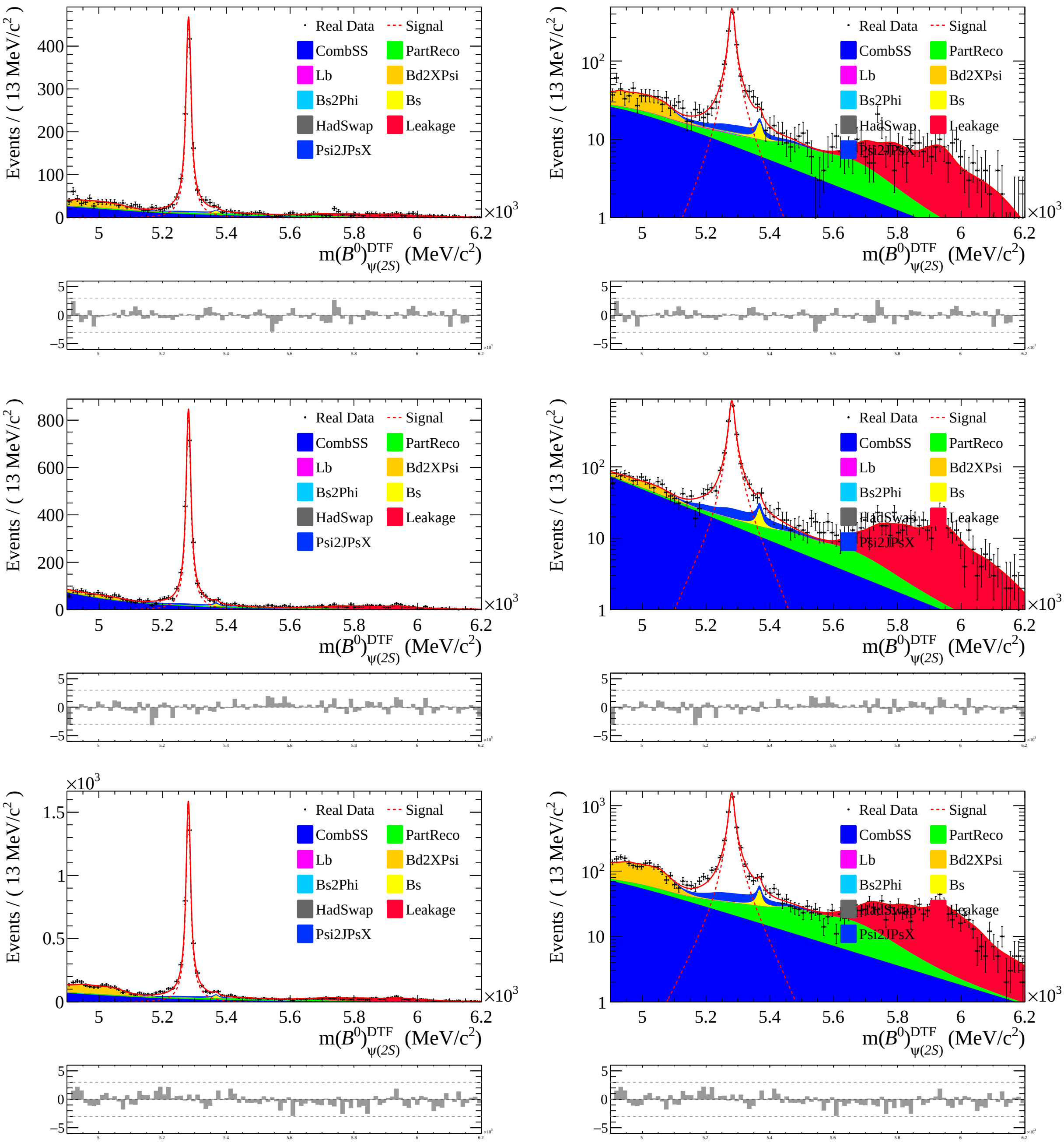


Figure 7.21: Fits to the kinematically constrained reconstructed B^0 mass in the $\psi(2S)$ region using fully selected data. Results for Run1, Run2p1 and Run2p2 data are shown in the first, second and third row respectively. The left columns shows the results in linear scale, a logarithmic scale in the y axis is used for the results in the right column. Different fit components are represented with different colors and indicated in the legend. Backgrounds are stacked on top of each other for a better visualization. The signal component is not stacked. The lower panels show the pull histograms of the fits.

in the Run2p1 period. This is however a consequence of the different MVA selection. As shown in table 6.2, the MVA_{Comb} cut applied to the $\psi(2S)$ region in the Run2p1 period is considerably larger than in the Run1 period. This explains the observed result.

Component	Run1	Run2p1	Run2p2
Combinatorial OS	488 ± 69	865 ± 100	1745 ± 124
Combinatorial SS	105 ± 10	86 ± 9	156 ± 12
Ratio(OS/SS)	4.6	10.1	11.2

Table 7.9: Comparison between combinatorial background yield extracted from fits to the kinematically constrained B^0 mass in the $\psi(2S)$ region and the amount of SS data selected under the same conditions.

Observed differences indicate the presence of instabilities in the fit model which need to be fixed before applying the constraints on the leaking backgrounds in the high-q^2 region and unblinding the B^0 mass window. Cross-check results tightening the MVA and electron PID cuts still suffer from the same issue. This points to either a systematic mismodelling of the $B^0 \to X\psi(2S)(\to e^+e^-)$ partially reconstructed background composition or to the necessity to include some other backgrounds in the mass fit to properly describe data. Possible ways to improve the fit stability are discussed at the end of this section.

Despite the issue with the stability within data-taking periods, one of the main goals of this section is to show that we can control the yields of leaking backgrounds by constraining them using the results from the mass fits performed in the J/ψ region. By selecting the q^2 region in the "traditional" way without using the q2BDT variables we allow a larger amount of leakage background candidates from the J/ψ region in order to have more statistics to check their modelling in the mass fits. On top of that, the kinematic constrain applied to the reconstructed B^0 mass provides a better separation between the leaking backgrounds and other components contributing to the data distribution so that we can get rid of correlations with other fit components. Results shown in figure 7.21 provide a strong validation of the strategy to constrain leaking background components.

7.4.3 Mass fit in the $\psi(2S)$ region without the kinematic constraint

Results from the previous section demonstrate the robustness of the procedure designed to constrain leaking backgrounds, however, they also show the existence of instabilities on the mass fits when results from different data-taking periods are compared. This section presents a second method to fit the reconstructed B^0 mass distribution, this time without applying the kinematic constrain on the B^0 mass. Having a second method to extract the yields allows to test the impact of the observed disagreements on the signal and partially reconstructed $B^0 \to X\psi(2S)(\to e^+e^-)$ yields.

Data is now selected using the two q2BDT variables as explained in section 5.3. All other selection cuts are identical to the ones employed for the previous mass fits in order to remove any bias coming from the calculation of efficiencies using simulated data that could affect the comparison of the results. Results presented in this section also serve as a validation in real data of the model to describe the signal mass distribution developed in section 7.3.

Fit settings

The main difference between this and the previous method to fit the invariant B^0 mass distribution in the $\psi(2S)$ region is the removal of the kinematic constraint on the B^0 mass and the use of the q2BDT variables to reduce the impact of the leaking backgrounds from the J/ψ region. This fit strategy is much closer to the one that will be used in the high-q^2 region since there it is not possible to constrain the kinematics of the electrons due to the lack of an intermediate resonance.

One of the main goals of this thesis is to be able to constrain the partially reconstructed background component leaking into the high-q^2 region from the $\psi(2S)$ region. In order to do this in a reliable way it is necessary to check that the fitted $B^0 \to X\psi(2S)(\to e^+e^-)$ yield in the $\psi(2S)$ region is correctly estimated. This can be done by performing a second mass fit constraining the partially reconstructed background yield from the previous one in order to check whether the estimation of the constrain is correct. If this would be the case, fits should be stable and the extracted signal yields should match.

The constraint on the yield of the $B^0 \to X\psi(2S)(\to e^+e^-)$ component is computed by efficiency-correcting the yield value extracted form the fits to the kinematically constrained B^0 mass. For this purpose, we use the same formula used to constrain the leaking components coming from the J/ψ region:

$$\mathcal{N}_{B^0\to X\psi(2S)(\to e^+e^-)} = \mathcal{N}^{DTF}_{B^0\to X\psi(2S)(\to e^+e^-)} \cdot \frac{\varepsilon_{B^0\to X\psi(2S)(\to e^+e^-)}}{\varepsilon^{DTF}_{B^0\to X\psi(2S)(\to e^+e^-)}}. \tag{7.20}$$

The variables with the DTF superscript refer to the kinematically constrained B^0 mass fit. The efficiencies are calculated using the formula from equation 7.18. Efficiencies and yields are shown in table 7.10. The efficiency-corrected yields are larger due to the larger mass window used in these new settings.

Yields of leaking components from the J/ψ region are constrained in the same way as in the previous fit. Constrained values are shown in table 7.11 for the $B^0 \to K^{*0}J/\psi(\to e^+e^-)$ background component and in table 7.12 for the $B^0 \to XJ/\psi(\to e^+e^-)$ partially reconstructed background.

If we compare the efficiencies and constrained yields of tables 7.11 and 7.12 with those of tables 7.5 and 7.6 we can see that the amount of leaking backgrounds is reduced in this new settings. This is a consequence of having used the BDT classifiers in order to select the $\psi(2S)$ region instead of the more conventional q^2 window.

One big difference with respect to the previous fit is that in this case it is not possible to constrain the shape of the combinatorial background mass distribution from SS data fits to the reconstructed B^0 mass. If the mass distribution would be constrained the fits would either not converge or not provide an accurate description of data. This is most probably due to a wrong estimation of the $B^0 \to X\psi(2S)(\to e^+e^-)$ yield in the fits to the kinematically constrained B^0 mass due to the large correlation between this component and the combinatorial background leading to the observed instabilities. The modelling of combinatorial background therefore relies on a exponential function with floating slope included in the final data fits. The combinatorial background mass distribution would be more accurately described by the exponential function with turn-on introduced in equation 7.5, however, having a function with four free floating parameters makes the fit unstable. The

Period	$\mathcal{N}^{DTF}$	$\frac{\varepsilon}{\varepsilon^{DTF}}$	$\mathcal{N}$
Run1	292 ± 45	1.104 ± 0.018	323 ± 50
Run2p1	421 ± 68	1.121 ± 0.012	472 ± 77
Run2p2	947 ± 80	1.109 ± 0.016	1051 ± 90

Table 7.10: Constrained values of the yield of $B^0 \to X\psi(2S)(\to e^+e^-)$ candidates for the mass fits performed in the $\psi(2S)$ region. The constraint comes form the yield extracted from the fits to the kinematically constrained B^0 mass in the same q^2 region.

Period	$\mathcal{N}^{J/\psi}$	$\frac{\varepsilon^{\psi(2S)}}{\varepsilon^{J/\psi}} \times 10^{-2}$	$\mathcal{N}^{\psi(2S)}$
Run1	15482 ± 147	0.793 ± 0.027	122.7 ± 4.3
Run2p1	28564 ± 195	0.746 ± 0.023	213.2 ± 6.8
Run2p2	53931 ± 270	0.865 ± 0.035	466 ± 19

Table 7.11: Constrained values of the yield of $B^0 \to K^{*0}J/\psi(\to e^+e^-)$ candidates for the mass fits performed in the $\psi(2S)$ region.

Period	$\mathcal{N}^{J/\psi}_{lep}$	$\mathcal{N}^{J/\psi}_{had}$	$\mathcal{N}^{J/\psi}_{total}$	$\frac{\varepsilon^{\psi(2S)}}{\varepsilon^{J/\psi}} \times 10^{-2}$	$\mathcal{N}^{\psi(2S)}_{total}$
Run1	1603 ± 185	2779 ± 154	4382 ± 263	2.94 ± 0.70	129 ± 32
Run2p1	3055 ± 231	5368 ± 203	8423 ± 327	2.90 ± 0.69	244 ± 59
Run2p2	6125 ± 333	11362 ± 288	17488 ± 472	2.77 ± 0.68	485 ± 120

Table 7.12: Constrained values of the yield of $B^0 \to XJ/\psi(\to e^+e^-)$ candidates for the mass fits performed in the $\psi(2S)$ region.

shape of the combinatorial background mass distribution is approximately exponential within the mass rage employed in the fits so that using an exponential function to describe it is still a good approximation. Besides the exponential slope, the other two free parameters in the mass fits are the signal and combinatorial background yields. All other background yields are either constrained with respect to the signal yield or from external fit results.

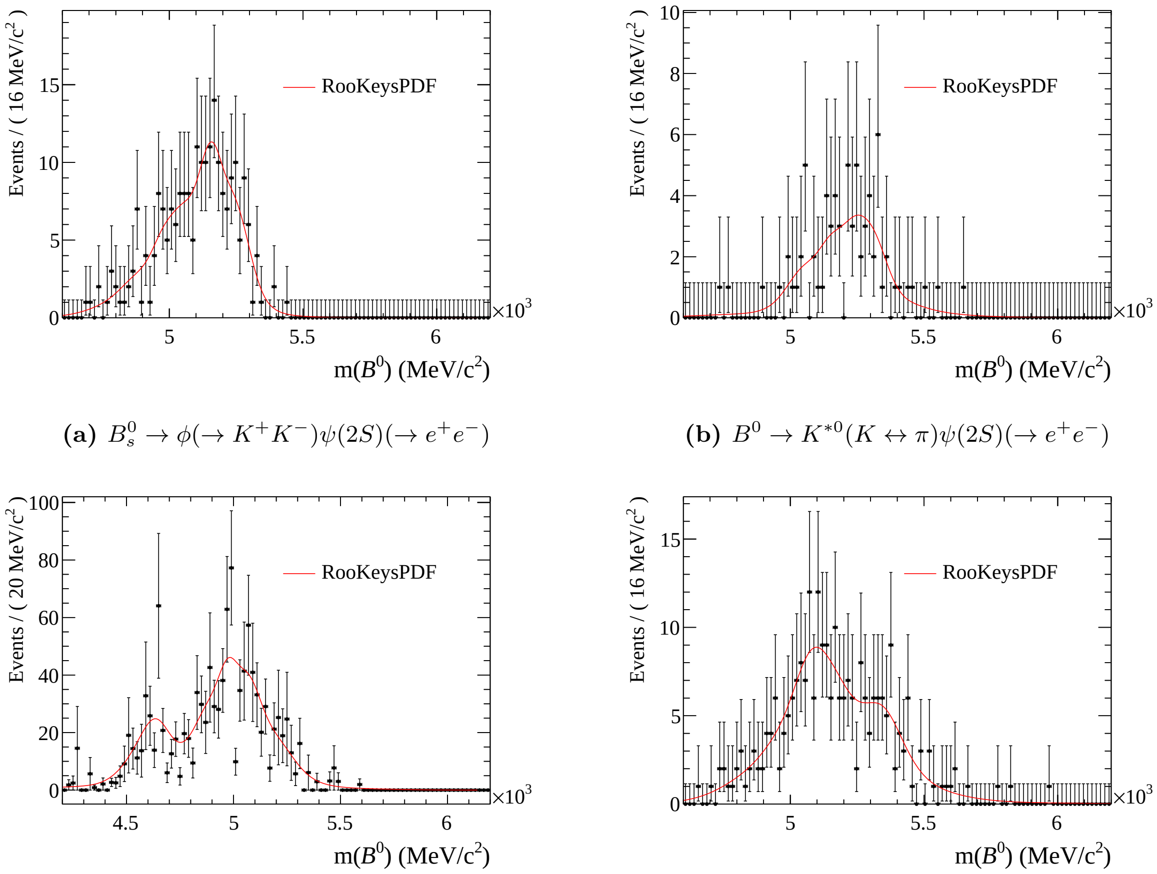
The mass range used for these new fit settings has been extended with respect to the mass range employed in the constrained kinematic fits of previous section in order to include a similar amount of signal events in both fits, simplifying the comparison of the signal yields. The kinematic constrain applied to the B^0 mass in previous fits improved the signal resolution, allowing us to narrow the fit range. Without the kinematic constrain, signal candidates present a mass distribution with a long tail to the left due to energy taken away by Bremsstrahlung photons which were not recovered. This can bee observed in the mass fits shown in section 7.3. The mass range used for the new fits is the following:

$$4800 < m(K\pi ee) < 6200 \text{ MeV}/c^2. \tag{7.21}$$

Fit components

The procedure to get the mass fit is completely identical to the previous one for the fits to the kinematically constrained B^0 mass. Mass fits in simulated data are shown for the Run2p2 period after applying full event selection. Figure 7.22 shows the mass fits for the exclusive backgrounds. Fits for the $B^0 \to X\psi(2S)(\to e^+e^-)$ partially reconstructed background and $B^0 \to K^{*0}J/\psi(\to e^+e^-)$ leaking background are shown in figure 7.23.

The mass distribution of leaking partially reconstructed backgrounds is modelled in the same way as in the previous fits. The PDFs are independently obtained for each of the contributing processes and they are added in the end weighting each component using the branching ratios of known decays. The computed weights are shown in table 7.13. Individual mass fits for the $B \to XJ/\psi(\to e^+e^-)$ components and their weighted sum are shown in figure 7.24. The mass distribution of signal candidates is described using the model developed in section 7.3.



(a) $B_s^0 \to \phi(\to K^+K^-)\psi(2S)(\to e^+e^-)$

(b) $B^0 \to K^{*0}(K \leftrightarrow \pi)\psi(2S)(\to e^+e^-)$

(c) $\Lambda_b^0 \to pK\psi(2S)(\to e^+e^-)$

(d) $B^0 \to K^{*0}\psi(2S)(\to \pi\pi J/\psi(\to e^+e^-))$

Figure 7.22: Fits to the reconstructed B^0 mass using simulated data corresponding to the Run2p2 period and selected in the $\psi(2S)$ region to model the shapes of exclusive backgrounds. Components are indicated in the captions.

Fit to data

Data mass fits are performed splitting data into the usual data-taking periods. Full selection is applied in every case. The set of fit components is the same as for the fits to the kinematically constrained B^0 invariant mass. Results corresponding to the nominal MVA and PID cuts are shown in figure 7.25.

Instabilities seen in the fits to the kinematically constrained invariant mass are also seen here. This is no surprise since the applied constraint to the partially reconstructed $B^0 \to X\psi(2S)(\to e^+e^-)$ background yield introduces a correlation between the fit results so that any mismodelling from the previous fit would automatically be introduced here. A very clear indication that the partially reconstructed background $B^0 \to X\psi(2S)(\to e^+e^-)$ yield extracted from the previous fit is not correct follows from the fact that it has been necessary to model the combinatorial background with a free exponential PDF in order to properly describe data. Despite this issue, data seems well described in the full mass range and in all data-taking periods. Small discrepancies are most probably inherited from the mismodelling of the partially reconstructed background yield. In particular, both leaking components and their yields seem to perfectly match data. On the other

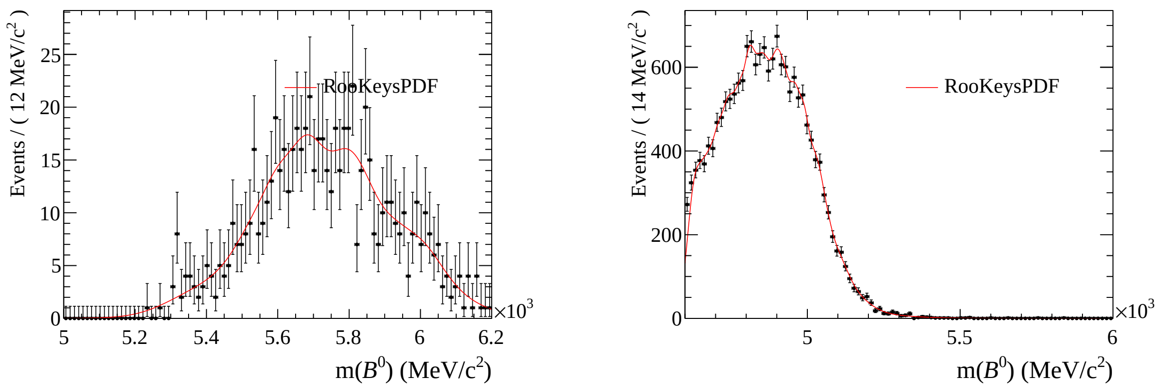


Figure 7.23: Fits to the reconstructed B^0 mass using simulated data corresponding to the Run2p2 period and selected in the $\psi(2S)$ region for the leaking $B^0 \to K^{*0} J/\psi(\to e^+e^-)$ background (left) and the partially reconstructed $B^0 \to X\psi(2S)(\to e^+e^-)$ background (right).

Sample	Exclusive decay	$\mathcal{B}$	f_{DEC}	$\frac{f}{f_u}$	$\frac{N_{sel}}{Ngen} \times 10^{-5}$	w
$B^0 \to XJ/\psi$	$B^0 \to K^{*0}J/\psi$	1.27×10^{-3}	0.1885	1	1.16	0.65
$B^+ \to XJ/\psi$	$B^+ \to K^+J/\psi$	1.01×10^{-3}	0.1596	1	0.639	0.34
$B_s^0 \to XJ/\psi$	$B_s^0 \to \phi(\to K^+K^-)J/\psi$	0.53×10^{-3}	0.1077	0.244	0.0803	0.01

Table 7.13: Calculated weights for the components in the leaking partially reconstructed background PDF using R2p2 simulated data for the mass fits performed in the $\psi(2S)$ region.

hand, the signal model developed in section 7.3 also provides a good description of the B^0 mass peak.

7.4.4 Cross-checks and discussion of the results

As a last validation step of the strategy to constrain the leaking background yields in the high-q^2 region we can compare the yields of the $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ signal components extracted from the mass fits performed in the $\psi(2S)$ region using the two different fitting strategies presented in the previous sections. This is an important cross-check in light of the observed instabilities in the background composition of the mass fits and the seemingly underestimated amount of $B^0 \to X\psi(2S)(\to e^+e^-)$ background. We want to check the impact of the instabilities found in the mass fits on the measured signal yields.

The cross-check is performed by computing the ratio of the signal yields independently measured from fits to the reconstructed B^0 mass with and without applying the kinematic constraints. In order to compare the results, the yield extracted from the kinematically constrained mass fit is corrected by the efficiency ratio between the two different fit settings:

$$\mathcal{N}^{DTF-corr}_{B^0\to K^{*0}\psi(2S)(\to e^+e^-)} = \mathcal{N}^{DTF}_{B^0\to K^{*0}\psi(2S)(\to e^+e^-)} \cdot \frac{\varepsilon_{B^0\to K^{*0}\psi(2S)(\to e^+e^-)}}{\varepsilon^{DTF}_{B^0\to K^{*0}\psi(2S)(\to e^+e^-)}}. \tag{7.22}$$

The ratio between the two yields is then defined as:

$$R \equiv \frac{\mathcal{N}^{DTF-corr}_{B^0\to K^{*0}\psi(2S)(\to e^+e^-)}}{\mathcal{N}_{B^0\to K^{*0}\psi(2S)(\to e^+e^-)}}. \tag{7.23}$$

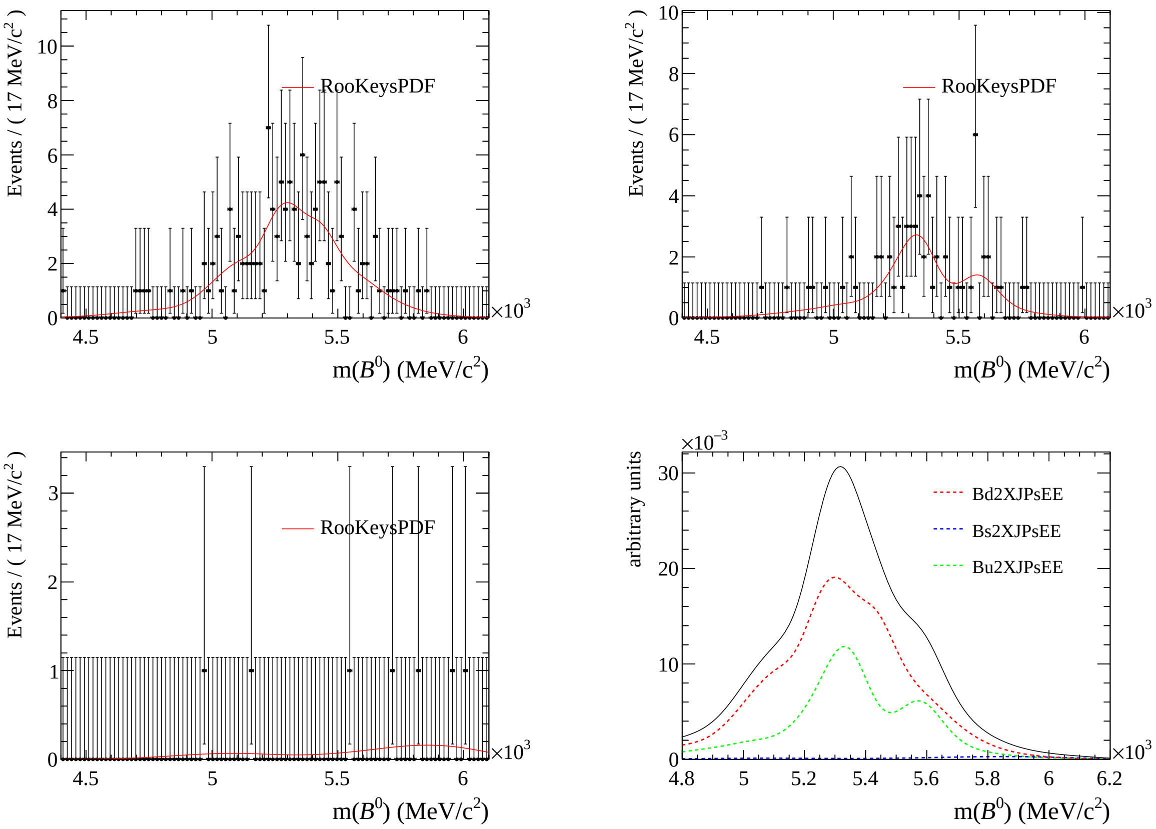


Figure 7.24: Reconstructed B^0 mass fits to simulated Run2p2 data in the $\psi(2S)$ region. In the top row we find $B^0 \to XJ/\psi(\to e^+e^-)$ (left) and $B^+ \to XJ/\psi(\to e^+e^-)$ (right). The bottom row shows $B^0_s \to XJ/\psi(\to e^+e^-)$ (left) and the weighted sum of the three components (right).

In order to compute the uncertainty, the two yields are assumed to be fully correlated. This is only approximately true since there is some level of interplay between the two results coming form the constrain of the $B^0 \to X\psi(2S)(\to e^+e^-)$ background yield introduced in the mass fit without the kinematic constrain on the B^0 mass. The correlation coming from it is not accounted for. In any case, the correlation is expected to be very large and results and conclusions are independent on the exact assumption about it. Table 7.14 shows the comparison between the two yields for the results using the nominal PID and MVA requirements. Ratios are shown in figure 7.26, including the results with the tighter PID and MVA requirements applied to perform the cross-checks.

Relative differences between the yield values are always below the 10% level and ratios are close to unity. However, given the large correlation between the yields, the uncertainties on R are

Period	$\mathcal{N}^{DTF-corr}$	$\mathcal{N}$	R	Relative difference (%)
Run1	1187 ± 39	1184 ± 64	1.003 ± 0.021	0.3
Run2p1	2209 ± 53	2055 ± 75	1.075 ± 0.014	7.5
Run2p2	4047 ± 74	4143 ± 106	0.9767 ± 0.0071	2.3

Table 7.14: Comparison between the $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ yields measured using the two different methods explained in the text. Results are given for the nominal PID and MVA cuts.

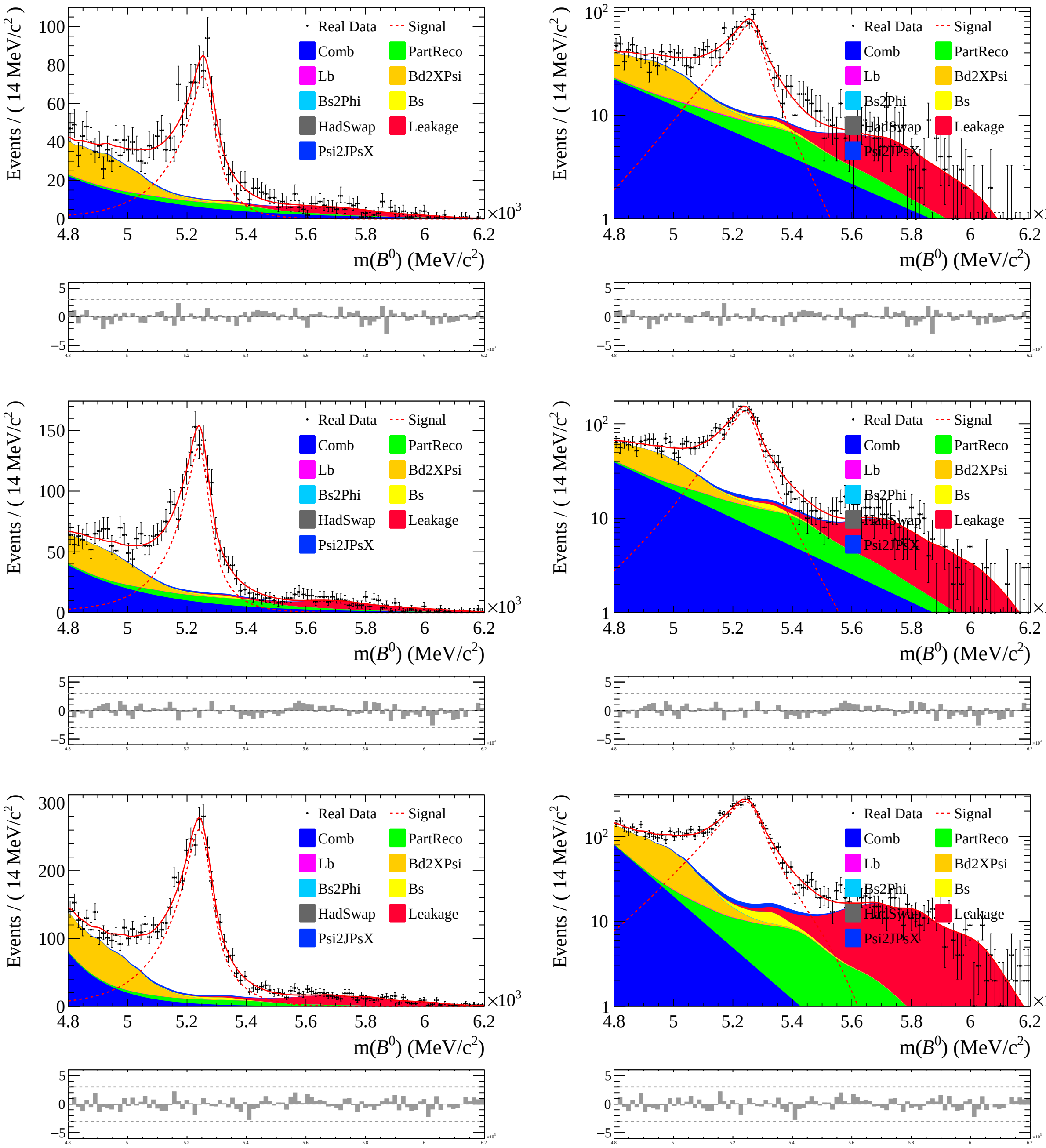
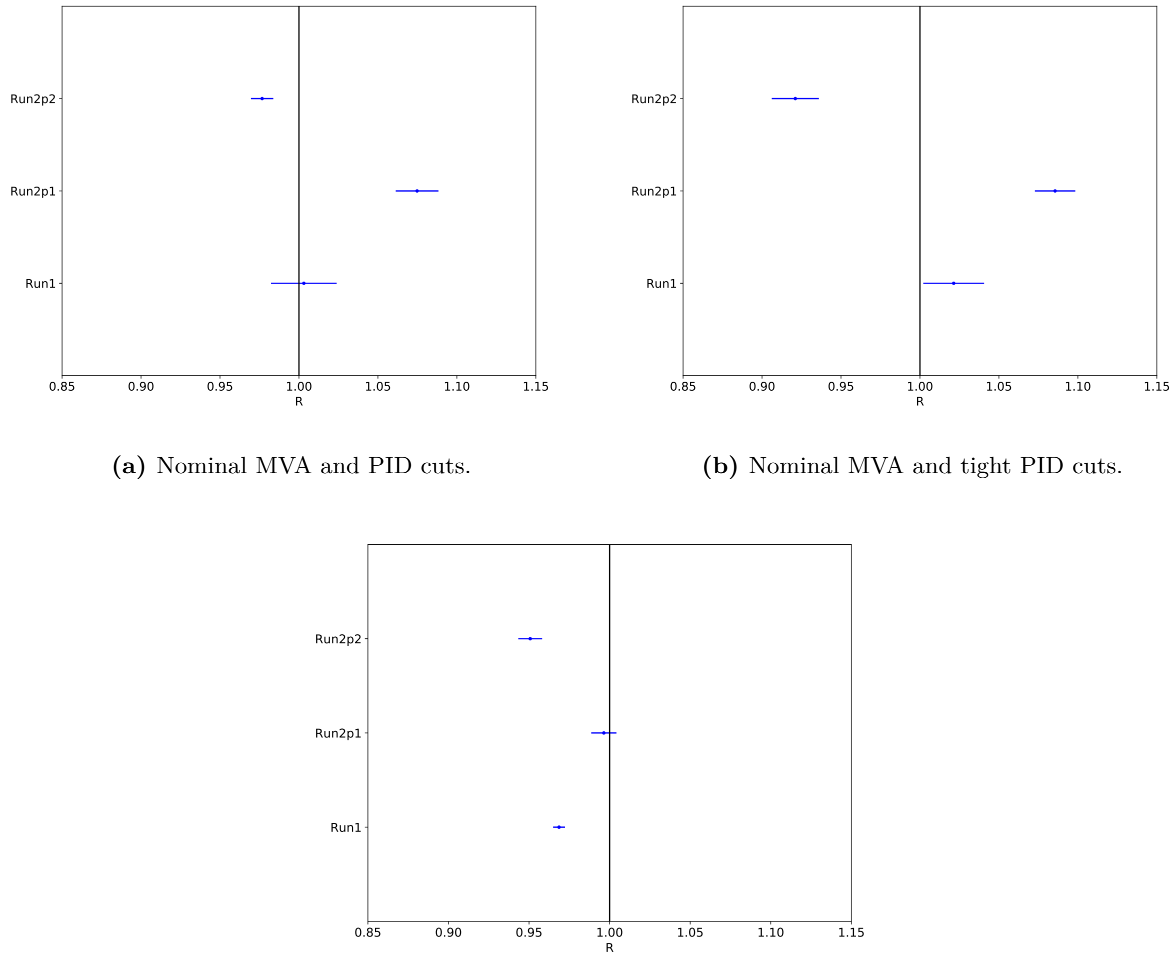


Figure 7.25: Fits to the reconstructed B^0 mass selected using real data selected in the $\psi(2S)$ region. Results for Run1, Run2p1 and Run2p2 data are shown in the first, second and third row, respectively. The left column shows the results in linear scale, a logarithmic scale in the y axis is used for the results in the right column. Different fit components are represented with different colors and indicated in the legend. Backgrounds are stacked on top of each other for a better visualization. The signal component is not stacked. The lower panels show the pull histograms of the fits.

small and deviations from 1 and among data-taking periods are significant. From figure 7.26 it is evident that the instabilities found in the mass fits translate into instabilities in the measured yields. It is interesting to note how results seem to follow a common deviation pattern with $R(\text{Run2p1}) > R(\text{Run1}) > R(\text{Run2p2})$ in all the performed cross-checks. This points to a system-



(a) Nominal MVA and PID cuts.

(b) Nominal MVA and tight PID cuts.

(c) Tight MVA and nominal PID cuts.

Figure 7.26: Comparison between the $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ yields measured using the two different methods explained in the text. Results are shown for different data-taking periods and different MVA and PID cut settings indicated in the captions. The values in the x axis correspond to the ratio between the yields defined the text.

atic origin for the observed deviations.

Deviations become smaller and results between run periods are more aligned after tightening the MVA cuts. This is expected since tightening the MVA cuts reduces the impact of combinatorial and partially reconstructed backgrounds in the mass fit as well as the effect of mismodellings associated to them.

Possible solutions involve the inclusion of so far not considered background components that might have contributions to data which are not properly described by any of the background included in the mass fits. One of these backgrounds is the double-semileptonic background. The veto to suppress it does not completely remove the background contribution to the mass fits due to its large branching ratio. It is assumed that the small amount of leftover events can be absorbed by the combinatorial background component without having a large impact in the final results.

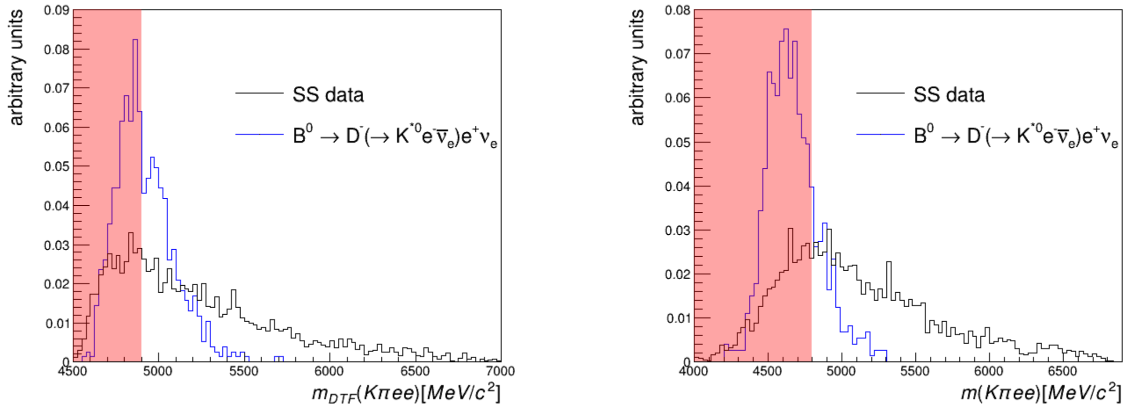


Figure 7.27: Comparison between the mass distributions of combinatorial background modelled from SS data and $B^0 \to D^-(\to K^{*0}l^-\overline{\nu}_l)l^+\nu_l$ background candidates. Mass distributions with (without) the kinematic constrain on the B^0 mass are shown in the plots in the left (right) hand side. The red shaded region corresponds to the region not included in the mass fits.

While this is expected to hold very well for the mass fits without the kinematic constrain, in the kinematic constrained B^0 mass the $B^0 \to D^-(\to K^{*0}l^-\overline{\nu}_l)l^+\nu_l$ background presents a visible peak near the edge of the mass range used in the fits to extract the yields. This might have a non negligible impact on the combinatorial background and partially reconstructed background modelling. Figure 7.27 shows a comparison between the mass distribution of combinatorial background candidates taken from SS data and $B^0 \to D^-(\to K^{*0}l^-\overline{\nu}_l)l^+\nu_l$ background candidates taken from a dedicated simulated sample. For a better visualization, both MVA cuts and the veto against $B^0 \to D^-(\to K^{*0}l^-\overline{\nu}_l)l^+\nu_l$ background have been removed. Data is represented for the Run2p2 period. Histograms are normalized for a better comparison.

The inclusion of a dedicated component to model the $B^0 \to D^-(\to K^{*0}l^-\overline{\nu}_l)l^+\nu_l$ contribution to the mass fits would require the generation of larger MC simulated samples in order to have enough statistics to model its mass distribution. An advantage of including it is that its yield can be constrained with respect to the signal yield for the fits performed in the resonant channels using equation 4.4. This could help stabilise the mass fits.

A second approach to improve the mass fit stability is to increase the amount of SS data used to constrain the shape of the combinatorial background mass distribution. This can be achieved by loosening the MVA_{Comb} cut when performing the mass fit to SS data. If we do this, care must be taken not to introduce any bias in the final result since the shape of the mass distribution can also be affected by the MVA_{Comb} cut. Figure 7.28 shows the efficiency profile of the MVA_{Comb} selection represented as a function of the reconstructed B^0 mass in the $\psi(2S)$ region. Results are obtained using SS data from the Run2p2 data taking period. A cut on $\text{MVA}_{\text{Comb}} > 0.67$ corresponding to the optimized value for the Run2p2 period in the $\psi(2S)$ region is used for reference.

A clear trend is seen, being the classifier more efficient for larger B^0 masses both with and without the kinematic constrain. This makes sense because the probability to find a signal candidate at large reconstructed B^0 mass becomes small. This makes it easier to classify events as background.

Results shown in figure 7.28 seem to discourage this second approach. It could however be feasible to constrain the turn-on parameters only, leaving the slope parameter free in the final mass

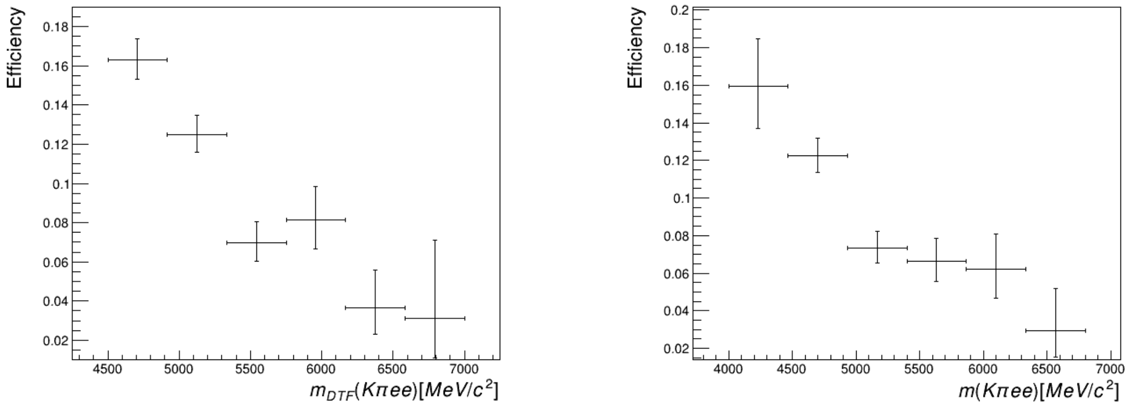


Figure 7.28: Efficiency profile of the MVA_{Comb} selection as a function of the reconstructed B^0 mass. Results are obtained using SS data from the Run2p2 period selected in the $\psi(2S)$ region. The efficiency results correspond to a cut of $\text{MVA}_{\text{Comb}} > 0.67$. Left (right) plots show results for the B^0 mass with (without) the kinematic constrain.

fits. Further studies would be needed in order to quantify the amount of bias introduced by doing it.

Finally, a better stability on the mass fits could be achieved by improving the modelling of the partially reconstructed $B^0 \to X\psi(2S)(\to e^+e^-)$ mass distribution. The generation of the missing simulated samples modelling the B^+ and B_s^0 contributions is planned for the future. This is very important for the measurement of $R_{K^{*0}}$ since this background leaks into the high-q^2 region and populates the mass region below the signal peak.

7.5 Constraining the over-reconstructed background yield

As explained in section 6.1, the over-reconstructed (OR) background veto has been removed in this analysis in order to keep the signal efficiency as large as possible and facilitate the description of the combinatorial background B^0 mass distribution in the high-q^2 region. The last part of this thesis presents the studies performed to estimate the expected amount of OR background in the high-q^2 region. This is necessary in order to compute the corresponding systematic uncertainty in the final result since this background component will not be considered in the fits to the reconstructed B^0 mass in the high-q^2 region as there is no simple way to model it.

The amount of OR background can be measured in data using one of the two resonant channels (either the J/ψ or the $\psi(2S)$ region). This background originates from $B^+ \to K^+e^+e^-$ candidates to which a random pion from the underlying event is attached so that the final four-body system is reconstructed as a $B^0 \to K^{*0}e^+e^-$. Hence, the invariant mass distribution of the $K^+e^+e^-$ system peaks at the B^+ mass for these candidates. If the two electrons are produced via an intermediate charmonium resonance we can apply a kinematic constrain to the reconstructed 4-body $m(K\pi ee)$ invariant mass. This improves the mass resolution of the OR background peak in the $m(Kee)$ distribution so that its yield can be easily extracted from fits to this variable. This technique could also be used to apply a veto against the OR background in the resonant J/ψ and $\psi(2S)$ regions, more efficient on signal than the one introduced in equation 6.3, but it is unfortunately not possible to apply it in the high-q^2 region due to the lack of an intermediate resonance.

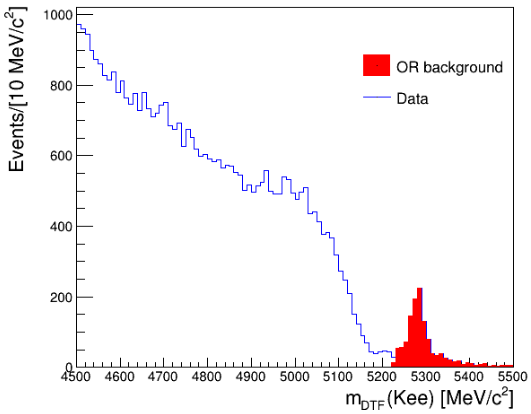


Figure 7.29: Reconstructed 3-body $m(Kee)$ invariant mass distribution using data from the Run2p2 period. Data is fully selected in the J/ψ region and with a kinematic constrain applied to the 4-body $m(K\pi ee)$ mass. The OR background peak is represented as a red filled histogram for a better visualization.

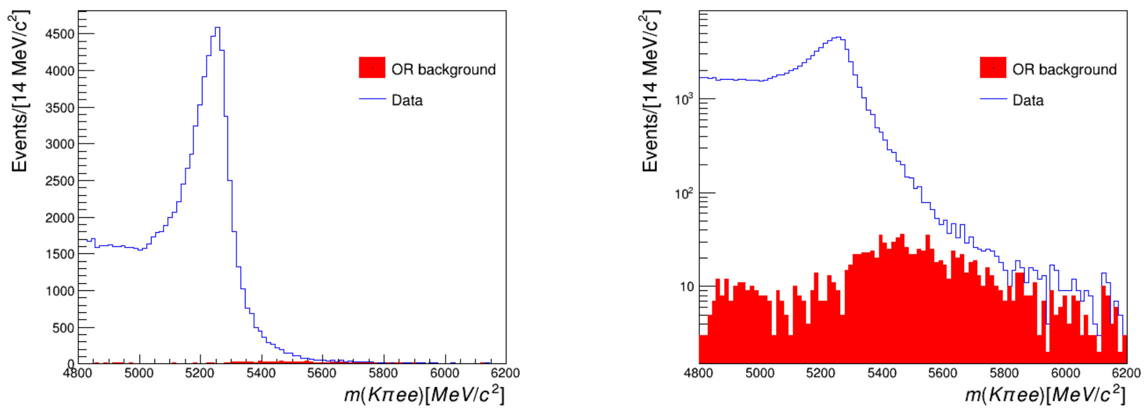


Figure 7.30: Reconstructed invariant $m(K\pi ee)$ distribution using data from the Run2p2 period. Data is selected in the J/ψ region. The OR background is selected from figure in the same way as in figure 7.29, applying a cut on the $m(Kee)$ variable, and represented in red. Data is plotted in linear scale in the left plot and using a logarithmic scale for the y axis in the plot to the right.

Figure 7.29 shows the $m(Kee)$ invariant mass distribution after the kinematic constrain is applied to the reconstructed 4-body B^0 invariant mass. Daata collected in the $Ru2p2$ data-taking period is used after applying full event selection in the J/ψ region. The OR background peak is clearly seen at the upper edge of the spectrum, right at the B^0 mass[3]. The rest of the data corresponds to B^0 candidates, either signal or background, as well as combinatorial background. They populate the region below the B^0 mass due to the missing pion, not included in the $m(Kee)$ distribution.

The OR background presents a relatively flat distribution in the reconstructed 4-body B^0 mass variable. It is therefore not expected to be a problem for the mass fits since it can be easily absorbed by other background components. This is shown in figure 7.30 for data from the Run2p2 period selected in the J/ψ region. The OR background component is selected as in figure 7.29, by cutting on the $m(Kee)$ distribution after the kinematic constraint is applied to the reconstructed B^0 mass. As it can be seen, it has an almost negligible contribution to the selected dataset.

[3]The B^0 and B^+ masses are roughly the same and we will not make any distinction between them here.

We can estimate the expected amount of OR background relative to the signal yield in the high-q^2 region by measuring the ratio between the yields of OR $B^+ \to K^+J/\psi(\to e^+e^-)$ and $B^0 \to K^{*0}J/\psi(\to e^+e^-)$ signal candidates in the J/ψ region and extrapolating the result to the high-q^2 region using MC simulated data. The J/ψ region is chosen to measure the yields due to the larger amount of statistics and the simpler background composition compared to the $\psi(2S)$ region.

7.5.1 Measuring the over reconstructed background yield in the J/ψ region

The first thing we need to do is to measure the amount of $B^0 \to K^{*0}J/\psi(\to e^+e^-)$ candidates in the J/ψ region. This is done by performing a fit to the reconstructed and kinematically constrained B^0 invariant using fully selected data. The fit settings are completely identical to the ones in the previous section. The only difference is that the mass range is narrowed in order to remove the $B^0 \to K^{*0}\psi(2S)(\to \pi\pi J/\psi(\to e^+e^-))$ background component from the left sideband. This is done in order to simplify the fits since in this section we only care about the signal yield and getting the yields of partially reconstructed backgrounds right is not as important as it was in the previous section. The mass range used in the fits is:

$$4900 < m(K\pi ee)^{DTF-J/\psi} < 6200 \text{ MeV}/c^2. \tag{7.24}$$

In order to cross-check the results, the fits are done using two different sets of MVA cuts: the ones optimized for the J/ψ region and the tight ones used in the rare channels. They are both indicated in table 6.2 and they are the same ones used for the cross checks performed in the previous section. Results with the tighter MVA cuts are shown in figure 7.31 for the three data-taking periods. When using the nominal MVA cuts optimized for the J/ψ region the results are very similar to the ones presented in section 7.4. The large suppression of the combinatorial background component is clear if we compare these results with the ones shown in figure 7.15.

In order to measure the OR background yield a window around the B^0 mass is selected in the 3-body $m(Kee)$ invariant mass after applying a kinematic constrain to the 4-body $m(K\pi ee)$ invariant mass as shown in figure 7.29. The following window is chosen:

$$5150 < m(Kee)^{DTF-J/\psi} < 5800 \text{ MeV}/c^2. \tag{7.25}$$

The window is extended up to 5800 MeV/c^2 in order to increase the amount of statistics as much as possible as well as to constrain the shape of the combinatorial background mass distribution. The lower edge is chosen so that the mass range contains most of the OR signal without compromising the fit stability by introducing a lot of background coming from $B^0 \to K^{*0}J/\psi(\to e^+e^-)$ candidates.

Only two background components are included in the fits, all other considered background show negligible contributions in the selected mass region. The main background component is the background coming from $B^0 \to K^{*0}J/\psi(\to e^+e^-)$ processes, mostly populating the lower edge of the mass spectrum. Its yield is constrained using the results obtained in the fits to the $m(K\pi ee)$ mass, corrected by the efficiency ratio using an equation analogous to 4.5. Its shape is modelled using MC simulated data fitted with a *RooKeysPdf*. The second background is the combinatorial background which is expected to have a very small contribution in the chosen mass range since it is close to the edge of the spectrum. Its shape is modelled using a free exponential function in the data mass fit since the amount of SS data within the fit region is not enough to reliably perform a mass fit to constrain its shape. As in previous fits, the combinatorial background yield is left as a free parameter.

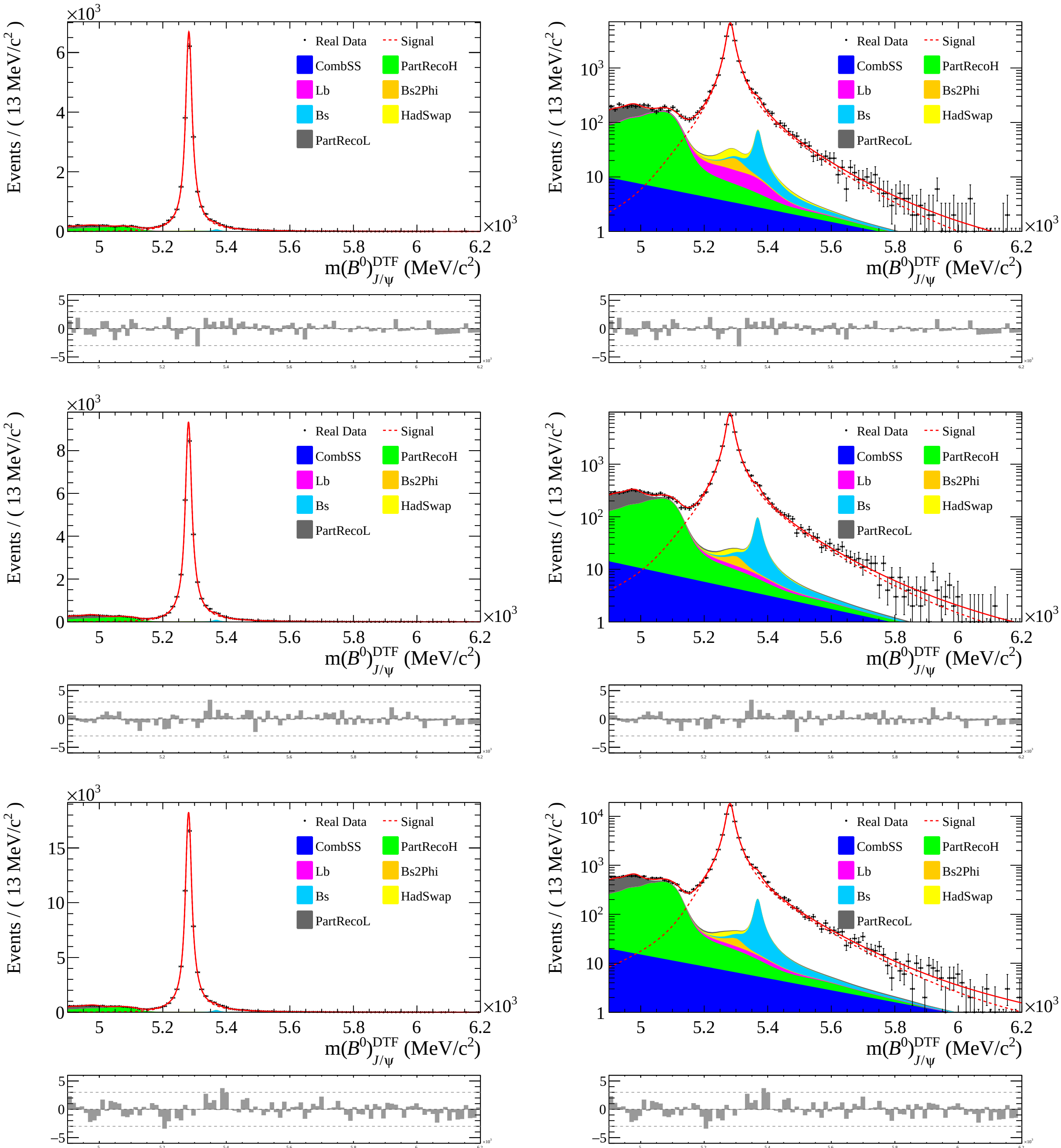


Figure 7.31: Fits to the kinematically constrained B^0 mass using fully selected real data in the J/ψ region. From up to bottom fits are shown for Run1, Run2p1 and Run2p2 data. In each row the same fit is shown twice, using a linear scale for the y axis (left) and a logarithmic scale (right). Components are indicated in the legend with backgrounds stacked on top of each other for a better visualization. The signal component is presented as a non stacked red dotted line. Data is selected with the tight MVA cuts optimized for the rare channels defined in table 6.2. The lower panels show the pull histograms of the fits.

The modelling of the OR background component relies on the fact that the shape of the OR background peak in the reconstructed $m(Kee)$ mass distribution is expected to be the same as the shape of the $B^0 \to K^{*0} J/\psi(\to e^+e^-)$ signal in the fits to the kinematically constrained B^0 mass. We can therefore fit MC simulated data selected in the J/ψ region to the reconstructed 4-body

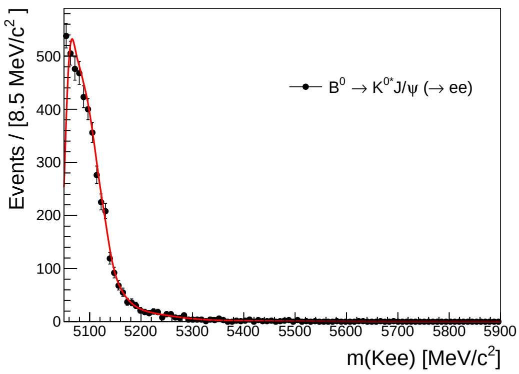


Figure 7.32: Fit to the reconstructed 3-body $m(Kee)$ invariant mass in simulated $B^0 \to K^{*0}J/\psi(\to e^+e^-)$ data, fully selected and truth-matched in the J/ψ region.

$m(K\pi ee)$ invariant mass with the kinematic constraint applied in order to model the shape of the OR background peak.

The model to describe the signal peak is simplified in this case and the shape of the mass distribution is approximated using a DSCB function whose shape parameters are fixed from mass fits using a $B^0 \to K^{*0}J/\psi(\to e^+e^-)$ MC simulated sample. The reason is that the model defined for the nominal mass fits in the J/ψ region, using the sum of an *Ipatia2* and a CB function, is too complex to fit the much smaller amount of data that we have in the OR background peak. This is true even if the shape parameters are fixed. Several approaches have been tried and a simple model using a DSCB function turns out to give the best and most stable results. For the shake of simplicity, MC simulated data is not splitted into the usual three Bremsstrahlung categories either.

The fit to the reconstructed $m(Kee)$ mass used to model the shape of the $B^0 \to K^{*0}J/\psi(\to e^+e^-)$ background is shown in figure 7.32 for the Run1 data-taking period. Results in other periods are consistent. An extended window is chosen in order to ensure that data close to the edges of the mass range is correctly modelled.

Fits to data are presented in figure 7.33. Results are shown for the looser and tighter MVA cuts and using data from the three data taking periods. It can be seen how the size of the OR background peak relative to the background component coming from $B^0 \to K^{*0}J/\psi(\to e^+e^-)$ candidates becomes smaller when tightening the MVA cuts. This is expected and it confirms the combinatorial nature of the OR background peak.

The combinatorial background component is very small in all the mass fits. In the fits with tighter MVA cuts it becomes completely negligible.

7.5.2 Extrapolation to the high q^2 region

In order to extrapolate the results to the high-q^2 region we need to compute the relative efficiencies for both $B^0 \to K^{*0}J/\psi(\to e^+e^-)$ and $B^+ \to K^+J/\psi(\to e^+e^-)$ processes as well as for the non-resonant channels $B^0 \to K^{*0}e^+e^-$ and $B^+ \to K^+e^+e^-$. The main issue in this calculation comes from the OR background modelling in MC simulated samples. Trigger, stripping and offline requirements introduced in section 5.1 for the $R_{K^{*0}}$ measurement select $B^0 \to K^{*0}e^+e^-$ candidates

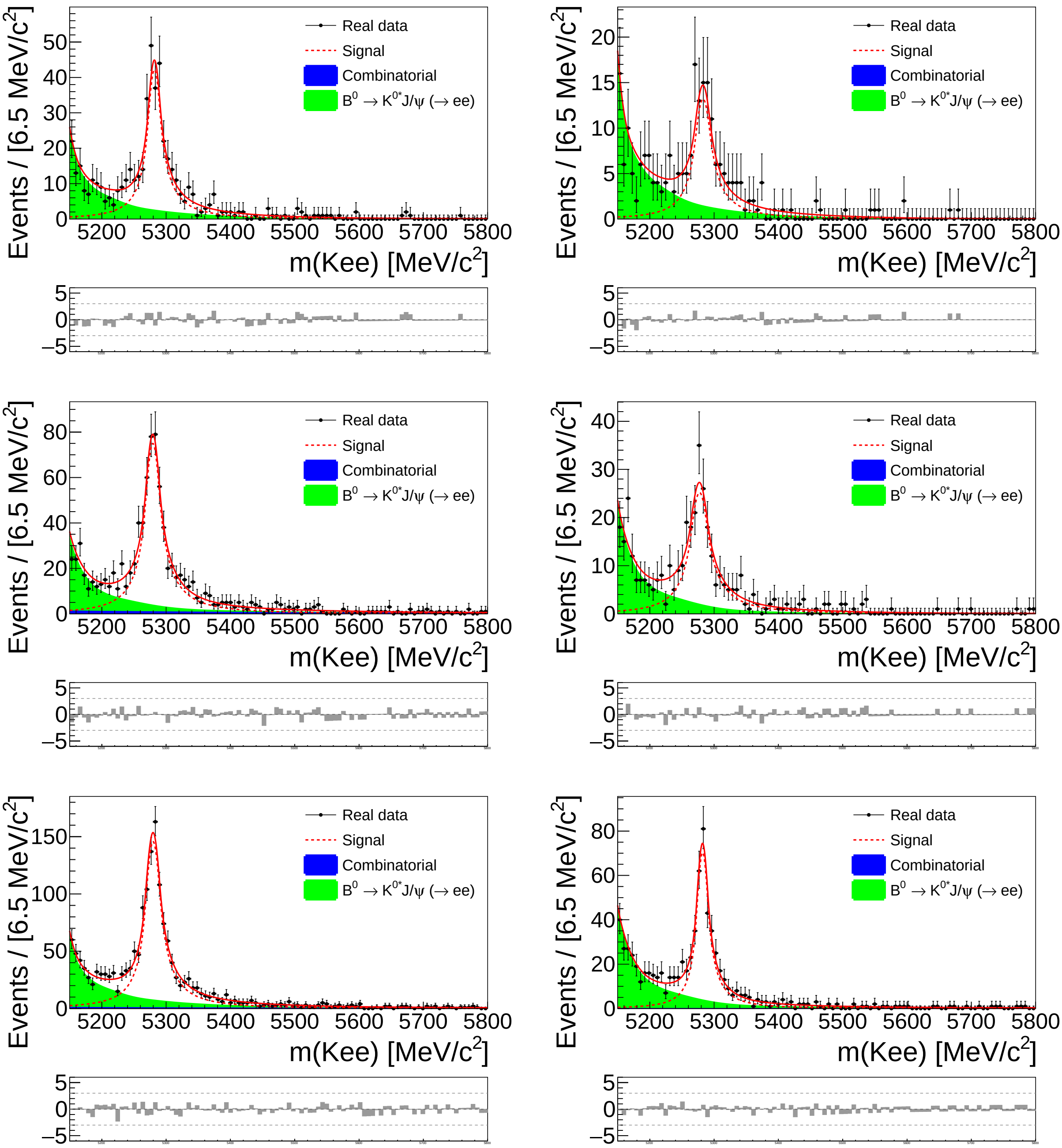


Figure 7.33: Fits to the reconstructed 3-body $m(Kee)$ invariant mass using fully selected data in the J/ψ region from the Run1 (top), Run2p1 (middle) and Run2p2 (bottom) periods. The plots in the column on the left hand side show the fit results with the looser MVA cuts optimized for the J/ψ region whereas the plots in the right hands side column show the results with the tighter MVA cuts used in the rare channels. Fit components are indicated in the legend. Background are stacked. The signal is not stacked and it is represented as a dashed red line. The lower panels show the pull histograms of the fits.

based on information from the decay topology of the 4-body final state and apply kinematic cuts and PID requirements on all particles. This includes the pions coming from the K^{*0} decay, which are not part of the final state of $B^+ \to K^+e^+e^-$ simulated candidates.

	Sample	J/ψ q^2 cuts	High q^2 cuts
Fully reconstructed	$B^+ \to K^+e^+e^-$	$6 < q^2 < 11$ GeV$^2/c^4$ $6 < q^2_{TRUE} < 11$ GeV$^2/c^4$	q2BDT > 0.95 $q^2_{TRUE} > 15$ GeV$^2/c^4$
	$B^0 \to K^{*0}e^+e^-$	$6 < q^2 < 11$ GeV$^2/c^4$ $6 < q^2_{TRUE} < 11$ GeV$^2/c^4$	q2BDT > 0.95 $q^2_{TRUE} > 15$ GeV$^2/c^4$
Generator level	$B^+ \to K^+e^+e^-$	$6 < q^2_{TRUE} < 11$ GeV$^2/c^4$	$q^2_{TRUE} > 15$ GeV$^2/c^4$
	$B^0 \to K^{*0}e^+e^-$	$6 < q^2_{TRUE} < 11$ GeV$^2/c^4$	$q^2_{TRUE} > 15$ GeV$^2/c^4$

Table 7.15: List of q^2 selection cuts employed to compute the selection efficiencies for the MC simulated $B^0 \to K^{*0}e^+e^-$ and $B^+ \to K^+e^+e^-$ samples.

In this thesis we are using the tuples processed within the R_X framework. The tuples containing the simulated samples for the $B^+ \to K^+e^+e^-$ decay are generally processed using the trigger and stripping lines for the R_K measurement, which looks for Kee final states. Thus, these tuples need to be re-processed for the $R_{K^{*0}}$ analysis so that they simulate OR background candidates including in the final state random pions from the underlying pp event, simulated with `PYTHIA8`. So far this has only been done for the MC samples simulating rare $B^+ \to K^+e^+e^-$ candidates. One of the assumptions made for the extrapolation of the results to the high-q^2 region is that the reconstruction efficiencies are the same for $B^+ \to K^+J/\psi(\to e^+e^-)$ and $B^+ \to K^+e^+e^-$ candidates. This assumption is based on the fact that, from an experimental point of view, both decay candidates have the same final state and the only way to differentiate the two processes is by looking at their q^2 distribution. Therefore, $B^+ \to K^+e^+e^-$ simulated MC samples are used in this thesis to compute the reconstruction efficiencies in both the J/ψ and high-q^2 regions. For consistency, the same is done for the $B^0 \to K^{*0}e^+e^-$ process.

The computation of the efficiencies in the J/ψ region takes into account both the q^2 selection and the mass ranges employed to perform the fits to measure the $B^0 \to K^{*0}J/\psi(\to e^+e^-)$ and OR background yields. For the high-q^2 region a more conservative approach is taken. Only the q^2 selection is taken into account for the efficiency computation since the B^0 mass range employed to perform the fit in the high-q^2 region has not yet been fixed. A further selection requirement using a reconstructed B^0 mass window would only reduce the expected values for the OR background yield relative to the signal yield. The efficiencies are calculated as the ratio between the number of selected candidates and the number of generator level candidates. In order to minimize biases, the number of generated events for the efficiencies computed in the J/ψ and high-q^2 regions is calculated in each case by selecting a window on q^2_{TRUE} encompassing the two regions. The same selection window is applied to the fully reconstructed samples for consistency and it is kept as large as possible to have enough statistics for the computation, specially for the $B^+ \to K^+e^+e^-$ samples. Table 7.15 summarizes the q^2 selection cuts applied to the MC simulated samples employed for the calculation of the efficiencies. On top of these cuts we have the two different sets of MVA cuts commented before, the full offline selection requirements applied to the fully reconstructed samples and the mass windows employed to fit the data and extract the yields in the J/ψ region.

In order to estimate the expected amount of OR background candidates relative to signal candidates in the high-q^2 region we employ a formula similar to the one defined in equation 7.4. The relative amount of OR background with respect to signal in the high-q^2 region can be expressed as:

$$\frac{\mathcal{N}^{\text{High } q^2}_{B^+\to K^+e^+e^-}}{\mathcal{N}^{\text{High } q^2}_{B^0\to K^{*0}e^+e^-}} = \frac{\mathcal{B}(B^+ \to K^+e^+e^-)}{\mathcal{B}(B^0 \to K^{*0}e^+e^-)} \cdot \frac{\varepsilon^{\text{High } q^2}_{B^+\to K^+e^+e^-}}{\varepsilon^{\text{High } q^2}_{B^0\to K^{*0}e^+e^-}}. \tag{7.26}$$

And a similar expression is found in the J/ψ region:

$$\frac{\mathcal{N}^{J/\psi}_{B^+\to K^+J/\psi(\to e^+e^-)}}{\mathcal{N}^{J/\psi}_{B^0\to K^{*0}J/\psi(\to e^+e^-)}} = \frac{\mathcal{B}(B^+\to K^+J/\psi(\to e^+e^-))}{\mathcal{B}(B^0\to K^{*0}J/\psi(\to e^+e^-))}\cdot\frac{\varepsilon^{J/\psi}_{B^+\to K^+J/\psi(\to e^+e^-)}}{\varepsilon^{J/\psi}_{B^0\to K^{*0}J/\psi(\to e^+e^-)}}. \tag{7.27}$$

Using the approximations $\varepsilon^{J/\psi}_{B^0\to K^{*0}J/\psi(\to e^+e^-)} \approx \varepsilon^{J/\psi}_{B^0\to K^{*0}e^+e^-}$ and $\varepsilon^{J/\psi}_{B^+\to K^+J/\psi(\to e^+e^-)} \approx \varepsilon^{J/\psi}_{B^+\to K^+e^+e^-}$ and dividing equation 7.26 by 7.27 we find:

$$\begin{aligned}\frac{\mathcal{N}^{\text{High } q^2}_{B^+\to K^+e^+e^-}}{\mathcal{N}^{\text{High } q^2}_{B^0\to K^{*0}e^+e^-}} =& \frac{\mathcal{B}(B^+\to K^+e^+e^-)}{\mathcal{B}(B^0\to K^{*0}e^+e^-)}\cdot\frac{\mathcal{B}(B^0\to K^{*0}J/\psi(\to e^+e^-))}{\mathcal{B}(B^+\to K^+J/\psi(\to e^+e^-))}\cdot \\ &\frac{\varepsilon^{\text{High } q^2}_{B^+\to K^+e^+e^-}}{\varepsilon^{J/\psi}_{B^+\to K^+e^+e^-}}\cdot\frac{\varepsilon^{J/\psi}_{B^0\to K^{*0}e^+e^-}}{\varepsilon^{\text{High } q^2}_{B^0\to K^{*0}e^+e^-}}\cdot\frac{\mathcal{N}^{J/\psi}_{B^+\to K^+J/\psi(\to e^+e^-)}}{\mathcal{N}^{J/\psi}_{B^0\to K^{*0}J/\psi(\to e^+e^-)}}.\end{aligned} \tag{7.28}$$

The branching fractions for the resonant processes in the J/ψ region are taken from PDG. Unfortunately, no measurements are yet available for the branching fractions of the $B^0\to K^{*0}e^+e^-$ and $B^+\to K^+e^+e^-$ decays in the high-q^2 region. They are calculated using the `flavio` package which computes SM theoretical predictions for flavor physics observables. This calculation does not account for possible NP contributions. Nevertheless, any NP contribution is expected to approximately cancel in the ratio of branching fractions since it would contribute similarly to both $B^0\to K^{*0}e^+e^-$ and $B^+\to K^+e^+e^-$ processes. Hence, this way of estimating the amount of OR background in the high-q^2 region does not introduce any bias in the $R_{K^{*0}}$ measurement.

Results of the calculation are presented in tables 7.16, 7.17 and 7.18 for the Run1, Run2p1 and Run2p2 data-taking periods, respectively, and using the looser and tighter MVA cuts. Simulated $B^+\to K^+e^+e^-$ MC samples are not available for the Run2p2 period so the Run2p1 simulated samples are used instead for the calculation of the efficiencies. No correlations between observables have been taken into account in the calculation. They are present in the theoretical calculation of the ratio of branching fractions in the rare channel due to the correlation between their QCD uncertainties. QCD uncertainties are similar in both decays, which are related by an isospin transformation, so that one would expect a positive correlation. The ratio of measured yields in the J/ψ region also presents some level of correlation due to the interplay between the two components in both mass fits. Nevertheless, the effect of these correlations would only modify the uncertainties of the branching fraction ratio of the rare processes and the ratio between the yields in the J/ψ region. These uncertainties do not have a large impact in the final uncertainty, dominated by the large errors in the calculation of the $B^+\to K^+e^+e^-$ efficiencies due to the limited statistics available after full event selection.

Figure 7.34 compares the measured OR background fractions in the J/ψ region with the values extrapolated to the high-q^2 region. The expected relative amount of OR background with respect to signal is generally smaller in the high-q^2 region than in the J/ψ region. The effect is larger when the values are compared after applying the tighter MVA cuts. This is expected since the MVA$_{\text{Comb}}$ classifier provides a larger combinatorial background suppression in the high-q^2 region, affecting the OR background in a similar way.

The relative amount of OR background with respect to the signal yield is smaller in the Run1 data-taking period whereas results for the two Run2 periods are similar between them. This is consistent in the two fit settings and independent of the MVA cut values. Differences between Run1

MVA	$\frac{\mathcal{N}^{J/\psi}_{Kee}}{\mathcal{N}^{J/\psi}_{K^{*0}ee}}$ [%]	$\frac{\mathcal{B}(Kee)}{\mathcal{B}(K^{*0}ee)}$	$\frac{\mathcal{B}(K^{*0}J/\psi)}{\mathcal{B}(KJ/\psi)}$	$\frac{\varepsilon^{\text{High }q^2}_{Kee}}{\varepsilon^{J/\psi}_{Kee}}$	$\frac{\varepsilon^{J/\psi}_{K^{*0}ee}}{\varepsilon^{\text{High }q^2}_{K^{*0}ee}}$	$\frac{\mathcal{N}^{\text{High }q^2}_{Kee}}{\mathcal{N}^{\text{High }q^2}_{K^{*0}ee}}$ [%]
Loose	1.156 ± 0.078	0.386 ± 0.059	1.257 ± 0.061	1.12 ± 0.35	1.170 ± 0.024	0.73 ± 0.26
Tight	0.556 ± 0.060	0.386 ± 0.059	1.257 ± 0.061	0.61 ± 0.38	1.202 ± 0.028	0.19 ± 0.13

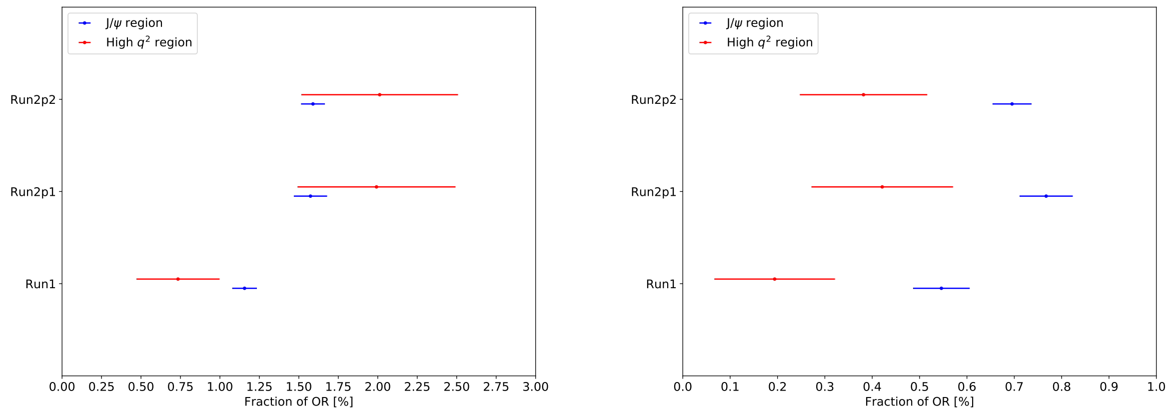
Table 7.16: Calculation of the relative amount of OR to signal yields in the high-q^2 region for the Run1 data-taking period. A shorthand notation has been used with Kee representing $B^+ \to K^+e^+e^-$ candidates and $K^{*0}ee$ representing $B^0 \to K^{*0}e^+e^-$ candidates and similarly for KJ/ψ and $K^{*0}J/\psi$.

MVA	$\frac{\mathcal{N}^{J/\psi}_{Kee}}{\mathcal{N}^{J/\psi}_{K^{*0}ee}}$ [%]	$\frac{\mathcal{B}(Kee)}{\mathcal{B}(K^{*0}ee)}$	$\frac{\mathcal{B}(K^{*0}J/\psi)}{\mathcal{B}(KJ/\psi)}$	$\frac{\varepsilon^{\text{High }q^2}_{Kee}}{\varepsilon^{J/\psi}_{Kee}}$	$\frac{\varepsilon^{J/\psi}_{K^{*0}ee}}{\varepsilon^{\text{High }q^2}_{K^{*0}ee}}$	$\frac{\mathcal{N}^{\text{High }q^2}_{Kee}}{\mathcal{N}^{\text{High }q^2}_{K^{*0}ee}}$ [%]
Loose	1.57 ± 0.11	0.386 ± 0.059	1.257 ± 0.061	2.04 ± 0.37	1.275 ± 0.022	1.99 ± 0.50
Tight	0.767 ± 0.056	0.386 ± 0.059	1.257 ± 0.061	0.84 ± 0.26	1.337 ± 0.026	0.42 ± 0.15

Table 7.17: Calculation of the relative amount of OR to signal yields in the high-q^2 region for the Run2p1 data-taking period. A shorthand notation has been used with Kee representing $B^+ \to K^+e^+e^-$ candidates and $K^{*0}ee$ representing $B^0 \to K^{*0}e^+e^-$ candidates and similarly for KJ/ψ and $K^{*0}J/\psi$.

MVA	$\frac{\mathcal{N}^{J/\psi}_{Kee}}{\mathcal{N}^{J/\psi}_{K^{*0}ee}}$ [%]	$\frac{\mathcal{B}(Kee)}{\mathcal{B}(K^{*0}ee)}$	$\frac{\mathcal{B}(K^{*0}J/\psi)}{\mathcal{B}(KJ/\psi)}$	$\frac{\varepsilon^{\text{High }q^2}_{Kee}}{\varepsilon^{J/\psi}_{Kee}}$	$\frac{\varepsilon^{J/\psi}_{K^{*0}ee}}{\varepsilon^{\text{High }q^2}_{K^{*0}ee}}$	$\frac{\mathcal{N}^{\text{High }q^2}_{Kee}}{\mathcal{N}^{\text{High }q^2}_{K^{*0}ee}}$ [%]
Loose	1.590 ± 0.076	0.386 ± 0.059	1.257 ± 0.061	2.04 ± 0.37	1.250 ± 0.018	1.97 ± 0.49
Tight	0.695 ± 0.041	0.386 ± 0.059	1.257 ± 0.061	0.84 ± 0.26	1.307 ± 0.022	0.37 ± 0.13

Table 7.18: Calculation of the relative amount of OR to signal yields in the high-q^2 region for the Run2p2 data-taking period. A shorthand notation has been used with Kee representing $B^+ \to K^+e^+e^-$ candidates and $K^{*0}ee$ representing $B^0 \to K^{*0}e^+e^-$ candidates and similarly for KJ/ψ and $K^{*0}J/\psi$.



(a) Loose MVA cuts

(b) Tight MVA cuts

Figure 7.34: Comparison between the fraction of OR background candidates relative to signal candidates in the J/ψ and high-q^2 regions. Blue points with error bars represent the measured values in the J/ψ region and red points the extrapolated values to the high-q^2 region.

and Run2 data are expected due to the different data taking conditions and trigger configuration. In particular, the track multiplicity (number of reconstructed track per event) is substantially larger in Run2 data due to the larger collision energy. This increases the average number of soft pions that can be attached to $B^+ \to K^+e^+e^-$ candidates resulting in a relative increase in the amount of OR background.

From the results we conclude that the expected amount of OR background in the high-q^2 region is of order $\sim 1-2\%$ with respect to the signal yield. This justifies neglecting the OR background component when performing the fits to the reconstructed B^0 mass in the high-q^2 region, accounting for its contribution within the final systematic uncertainty.

Chapter 8

Conclusions

A study of the experimental features of $B^0 \to K^{*0}e^+e^-$ decays at the LHCb experiment in the high di-lepton invariant mass region has been presented in this thesis. The dataset employed in the analysis corresponds to an integrated luminosity of ~ 9 fb^{-1} of data recorded in the LHCb experiment during the 2011-2018 data-taking period. Alongside, Monte Carlo simulated data describing the relevant decay processes contributing to the studied dataset has been thoroughly used for the computation of efficiencies and to model the expected invariant mass distribution of signal and background components in almost every part of the analysis. Simulated datasets were produced independently for each data-taking period and configuration of the magnet polarity to describe as precisely as possible the response of the LHCb detector in all experimental conditions.

The $B^0 \to K^{*0}e^+e^-$ process constitutes the main source of experimental uncertainty in the measurement of $R_{K^{*0}}$, which is the final goal of this analysis. This is primarily due to the poorer mass resolution caused by the emission of Bremsstrahlung radiation from the electrons, which is not always properly recovered and associated to the B^0 meson decay. This poorer resolution increases the impact of backgrounds generated by a large number of processes which can mimic the experimental signatures of the signal decay. Characterising these backgrounds and precisely describing the properties of signal candidates is therefore extremely important in order to perform a reliable measurement of the expected signal yield in the high-q^2 region. Different techniques have been developed to mitigate the impact of these background processes . The q^2 spectra of the $B^0 \to K^{*0}e^+e^-$ decay is conventionally divided into different regions which contain different physical contributions. Besides the high-q^2 region, the two resonant regions dominated by the contribution of decays coming from the charmonium $c\bar{c}$ resonances J/ψ and $\psi(2S)$ play a very important role in this thesis. Their contribution peaks within the q^2 intervals $[6, 11]$ GeV$^2/c^4$ and $[11, 15]$ GeV$^2/c^4$ respectively an they correspond to tree level processes with much larger decay rates, providing an ideal tool to test the strategy designed for the analysis of the high-q^2 region.

A large background contribution is associated to $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ and $B^0 \to X\psi(2S)(\to e^+e^-)$ partially reconstructed candidates leaking from the $\psi(2S)$ region into the high-q^2 region. This background is generated by misassociation of calorimeter photon clusters to signal candidates in the procedure designed for the LHCb experiment to recover the energy lost by the electrons due to Bremsstrahlung photon emission. This effect can generate an overestimation of the invariant mass of the final electrons making candidates from the $\psi(2S)$ region migrate to the high-q^2 region. In order to mitigate this effect, a multivariate BDT classifier was trained. It selects signal events in the high-q^2 region and rejects background events associated to the $\psi(2S)$ resonance improving the performance in terms of signal efficiency and background rejection compared with the more traditional selection approach based on a simple cut on the q^2 variable. This classifier uses simulated $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ and $B^0 \to K^{*0}e^+e^-$ candidates as background and signal proxies respectively and is trained on electron features in order to achieve a similar suppression of the partially reconstructed $B^0 \to X\psi(2S)(\to e^+e^-)$ background, which also leaks into the high-q^2 region. The performance is evaluated using a second classifier trained on the same features to reject J/ψ leaking events which also populate the $\psi(2S)$ region. The relative efficiency for the two types of $\psi(2S)$ processes that leak into the high-q^2 region is evaluated comparing the $\psi(2S)$ and high-q^2 regions making use of the two trained classifiers. A similar performance is observed for the two leaking background components with only small differences between the two q^2 regions which can be explained as a consequence of physical differences in the q^2 spectra. This verifies the good performance of the BDT selection for the two background processes.

The mass distribution of signal $B^0 \to K^{*0}e^+e^-$ and combinatorial background candidates is distorted by the q^2 selection in the high-q^2 region, making them difficult to model. A data driven method has been developed to model the shape of the combinatorial background B^0 mass distribution using *Same Sign* data as a proxy. This method has been validated in the $\psi(2S)$ region, which suffers from a similar distortion effect at low reconstructed B^0 masses. A good agreement is seen between the SS data and the combinatorial background distributions studied by selecting a combinatorial background enriched data sample splitted into the different data-taking periods. A model to describe the shape of the signal mass distribution has also been built. It consists of a DSCB core with steps attached to the tails modelling the effect of the q^2 cuts. It has been tested in MC simulated data in both the $\psi(2S)$ and high-q^2 regions and validated in real data in the $\psi(2S)$ regions. The model is able to accurately describe all the features of the signal mass distribution, showing consistent results for all data-taking periods and Bremsstrahlung categories.

A central part of this thesis deals with the development of a strategy to constrain the yields of the leaking background components in the high-q^2 region. The plan is to apply the constraints to the B^0 mass fits after measuring the yield of the two leaking $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ and $B^0 \to X\psi(2S)(\to e^+e^-)$ components in the $\psi(2S)$ region, extrapolating the yield results to the high-q^2 region using relative efficiencies computed from MC simulated data. This strategy is tested in the $\psi(2S)$ region, which also contains background components leaking form the J/ψ region. In order to cross-check the results, the fits to the reconstructed B^0 mass in the $\psi(2S)$ region are performed in two different ways: with and without applying a kinematic constraint to the reconstructed B^0 mass variable. The consistency between the results is checked by comparing the $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ yields extracted from the two fits, which are repeated applying tighter cuts on the electron PID variables and the combinatorial MVA, trained to suppress combinatorial background, to test their stability. Although the results are similar, discrepancies on the level of $\sim 10\%$ are observed when comparing the two methods and some instabilities are found between data-taking periods and when comparing results with different MVA and PID cuts. Further work will be required to understand the origin of the discrepancies. This includes testing the impact of $B^0 \to D^-(\to K^{*0}l^-\overline{\nu}_l)l^+\nu_l$ decays on the mass fits performed in the $\psi(2S)$ region, which have been so far neglected, and improving the modelling of the partially reconstructed background $B^0 \to X\psi(2S)(\to e^+e^-)$ component in MC simulated data.

Finally, the expected amount of over-reconstructed background candidates relative to signal candidates in the high-q^2 is estimated from a measurement performed in the J/ψ region. The amount of OR background is measured in the J/ψ region from a fit to the 3-body $m(Kee)$ invariant mass, where the OR background shows a peak located at the B^0 mass after applying a kinematic constrain to the reconstructed 4-body B^0 mass variable. The measured yield is then extrapolated to the high-q^2 region using efficiencies computed from MC simulated samples and ratios of branching fractions. Results are compared between data-taking periods and repeated after applying tighter combinatorial MVA cuts and are shown to be consistent. An average amount of $\sim 1-2\%$ OR background contribution with respect to the signal yield is expected in the high-q^2 region with variations depending on the data-taking period and MVA cuts applied. This result will be used to estimate the systematic uncertainty component associated to the OR background.

The immediate next step in the analysis consists of understanding and fixing the discrepancies found between the results obtained for the $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ yields from the mass fits performed in the $\psi(2S)$ region. In order to validate the result obtained for the $B^0 \to X\psi(2S)(\to e^+e^-)$ yield, a constraint is applied in the fits to the reconstructed B^0 mass performed in the $\psi(2S)$ region without the kinematic constraint. The constraint is derived from the result obtained from the fits to the kinematically constrained B^0 mass variable, for which the separation between signal and background components is better. Once the two fits show stable results and yields for the $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ component agree, the next step would be to release the constraint on the $B^0 \to X\psi(2S)(\to e^+e^-)$ component in the mass fit without the kinematic constrain on the B^0 mass variable. This would allow to check the stability of the fit results when the $B^0 \to X\psi(2S)(\to e^+e^-)$ yield is left as a free component. This check is important because the $B^0 \to Xe^+e^-$ partially reconstructed background component can not be constrained in the fit to the B^0 mass performed in the high-q^2 region. If results are positive, the constraints derived from the mass fits performed in the $\psi(2S)$ region will be applied to the mass fits in the high-q^2 region and a fit to to the sidebands of the blinded B^0 mass window will be performed. This will allow to check the stability of the mass fits before unblinding the B^0 mass window, which contains most of the signal. Besides that, further studies are also required in order to establish a working point to optimize the cuts applied on both the combinatorial MVA and q2BDT classifiers in the high-q^2 region. The results presented in this thesis are expected to hold regardless o the optimized cut values for the classifiers.

Bibliography

[1] ATLAS collaboration, G. Aad et al., Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC, Phys. Lett. B716 (2012) 1, arXiv:1207.7214.

[2] CMS collaboration, S. Chatrchyan et al., Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC, Phys. Lett. B716 (2012) 30, arXiv:1207.7235

[3] Heisenberg, W. The physical content of quantum kinematics and mechanics. In Wheeler, J. A. & Zurek, W. H. (eds.) Quantum Theory and Measurement, 62–84 (Princeton UP, Princeton, NJ, 9 1983). [Originally published: Z. Phys., 43, 172-98 (1927)].

[4] R.L. Workman et al. (Particle Data Group), Prog. Theor. Exp. Phys. 2022, 083C01 (2022)

[5] ATLAS Collaboration. Test of the universality of τ and μ lepton couplings in W-boson decays with the ATLAS detector. Nat. Phys. 17, 813–818 (2021). https://doi.org/10.1038/s41567-021-01236-w

[6] The ALEPH, DELPHI, L3, OPAL, SLD Collaborations, the LEP Electroweak Working Group, the SLD Electroweak and Heavy Flavour Groups, Precision Electroweak Measurements on the Z Resonance, Phys. Rept. 427 (2006) 257, arXiv:hep-ex/0509008.

[7] DELPHI, OPAL, LEP Electroweak, ALEPH, L3, S. Schael et al., Electroweak Measurements in Electron-Positron Collisions at W-Boson-Pair Energies at LEP, Phys. Rept. 532 (2013) 119, arXiv:1302.3415.

[8] A. Pich, Precision Tau Physics, Prog. Part. Nucl. Phys. 75 (2014) 41, arXiv:1310.7922.

[9] B Flavor Anomalies: 2021 Theoretical Status Report David London and Joaquim Matias Annual Review of Nuclear and Particle Science 2022 72:1, 37-68

[10] Elena Graverini and on behalf of theATLAS, CMS and LHCb collaborations 2019 J. Phys.: Conf. Ser. 1137 012025

[11] The LHCb collaboration., Aaij, R., Adeva, B. et al. Test of lepton universality with $B^0 \rightarrow K^{*0}l^+l^-$ decays. J. High Energ. Phys. 2017, 55 (2017). https://-doi.org/10.1007/JHEP08(2017)055

[12] LHCb collaboration. Test of lepton universality in beauty-quark decays. Nat. Phys. 18, 277–282 (2022). https://doi.org/10.1038/s41567-021-01478-8

[13] Wikipedia, https://en.wikipedia.org/wiki/Standard_Model

[14] S.L. Glashow, Nucl. Phys. 22 (1961) 579

[15] S. Weinberg, Phys. Rev. Lett. 19 (1967) 1264

[16] S.L. Glashow, J. Iliopoulos and L. Maiani, Phys. Rev. D2 (1970) 1285.

[17] A. Salam, in Elementary Particle Theory, ed. N. Svartholm (Almquist and Wiksells, Stockholm, 1969), p. 367.

[18] J. S. Schwinger, A Theory of the Fundamental Interactions, Annals Phys. **2** (1957), 407-434 doi:10.1016/0003-4916(57)90015-5

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

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

[21] P. W. Higgs, Broken symmetries and the masses of gauge bosons, Phys. Rev. Lett. 13 (1964) 508.

[22] P. W. Higgs, Spontaneous symmetry breakdown without massless bosons, Phys. Rev. 145 (1966) 1156.

[23] T. W. B. Kibble, Symmetry breaking in non-Abelian gauge theories, Phys. Rev. 155 (1967) 1554.

[24] N. Cabibbo, Unitary symmetry and leptonic decays. Phys. Rev. Lett. 10, 531–533 (1963)

[25] M. Kobayashi, T. Maskawa, CP violation in the renormalizable theory of weak interaction. Prog. Theor. Phys. 49, 652–657 (1973)

[26] S. L. Glashow, J. Iliopoulos, and L. Maiani, Weak interactions with lepton-hadron symmetry, Phys. Rev. D2 (1970) 1285.

[27] M. Neubert, Effective field theory and heavy quark physics, In Theoretical Advanced Study Institute in Elementary Particle Physics: Physics in D 4, doi:10.1142/9789812773579 0004 (2005), arXiv:hep-ph/0512222

[28] T. Blake, T. Gershon, and G. Hiller. "Rare b hadron decays at the LHC". In: Ann. Rev. Nucl. Part. Sci. 65 (2015), pp. 113–143. doi: 10.1146/annurev-nucl102014-022231. arXiv: 1501.03309 [hep-ex].

[29] K. Wilson, W. Zimmermann, Operator product expansions and composite field operators in the general framework of quantum field theory, Commun. Math. Phys. 24 (1972) 87–106. doi:10.1007/BF01878448.

[30] G. Buchalla, A. J. Buras, M. E. Lautenbacher, Weak decays beyond leading logarithms, Rev. Mod. Phys. 68 (1996) 1125–1144. arXiv:hep-ph/9512380, doi:10.1103/RevModPhys.68.1125.

[31] D. Bečirević, S. Fajfer, N. Košnik, and O. Sumensari, Leptoquark model to explain the B-physics anomalies,R_K and R_D , Phys. Rev. D94 (2016) 115021, arXiv:1608.08501.

[32] A. Crivellin, D. Müller, and T. Ota, Simultaneous explanation of R(D(∗)) and $b \to s^{+-}$: the last scalar leptoquarks standing, arXiv:1703.09226.

[33] S. Descotes-Genon, J. Matias, and J. Virto, Understanding the $B \to K^{*}\mu^{+}\mu^{-}$ anomaly, Phys. Rev. D88 (2013) 074002, arXiv:1307.5683.

[34] R. Gauld, F. Goertz, and U. Haisch, An explicit Z' -boson explanation of the $B \to K^{*}\mu^{+}\mu^{-}$ anomaly, JHEP 01 (2014) 069, arXiv:1310.1082.

[35] W. Altmannshofer, S. Gori, S. Profumo, and F. S. Queiroz, Explaining dark matter and B decay anomalies with an $L_\mu - L_\tau$ model, JHEP 12 (2016) 106, arXiv:1609.04026.

[36] https://cds.cern.ch/record/2826122/files/20220902-marin-CIPANP-puzzles%20(2).pdf

[37] https://agenda.infn.it/event/25022/contributions/130214/attachments/79891/104059/Stangl_21-03_LaThuile.pdf

[38] S.Bifani, S.Descotes-Genon, A.Romero Vidal and M.H.Schune, Review of Lepton Universality tests in B decays, J. Phys. G **46** (2019) no.2, 023001 doi:10.1088/1361-6471/aaf5de [arXiv:1809.06229 [hep-ex]].

[39] E. Fermi, Tentativo di una Teoria dell'emissione dei raggi , La Ricerca Scientifica 4 (1933) 491.

[40] K.G. Chetyrkin, M. Misiak, M. Munz, Weak radiative B meson decay beyond leading logarithms. Phys. Lett. B 400, 206–219 (1997), arXiv:hep-ph/9612313

[41] L. Pescatore, Searching for new physics in $b \to sl^{+}l^{-}$ transitions at the LHCb experiment, doi:10.1007/978-3-319-66423-1.

[42] https://www.nikhef.nl/~pkoppenb/anomalies//AnomaliesPlot-dimuons-signed.pdf

[43] LHCb collaboration, R. Aaij et al., Measurement of CP-averaged observables in the $B^0K^0\mu+$ $\mu-$decay, Phys. Rev. Lett. 125 (2020) 011802, arXiv:2003.04831.
[44] D. M. Straub, flavio: A python package for flavour and precision phenomenology in the Standard Model and beyond, arXiv:1810.08132.
[45] LHCb collaboration, R. Aaij et al., LHCb detector performance, Int. J. Mod. Phys. A30 (2015) 1530022, arXiv:1412.6352.
[46] LHCb collaboration, A. A. Alves Jr. et al., The LHCb detector at the LHC, JINST 3 (2008) S08005.
[47] https://cds.cern.ch/record/2800984/files/CCC-v2022.png?subformat=icon-1440
[48] https://home.cern/science/experiments/moedal-mapp
[49] https://lhcb.web.cern.ch/lhcb/speakersbureau/html/bb/_ProductionAngles.html
[50] https://lhcbproject.web.cern.ch/lhcbproject/Publications/LHCbProjectPublic/Summary_DP.html
[51] R. Aaij et al., Performance of the LHCb Vertex Locator, JINST 9 (2014) P09007, arXiv:1405.7808.
[52] Abellan Beteta et al., Monitoring radiation damage in the LHCb Tracker Turicensis, JINST 15 (2020) no.08, P08016, arXiv:1809.05063.
[53] P. d'Argent et al., Improved performance of the LHCb Outer Tracker in LHC Run 2, 2017 JINST 12 P11016, arXiv:1708.00819.
[54] D. Gerick, Search for physics beyond the Standard Model in an angular analysis of the $B^+ \to$ K^{*++-} decay at LHCb, doi:10.11588/heidok.00029883
[55] LHCb collaboration, Trigger Schemes, https://lhcb.web.cern.ch/speakersbureau/html/TriggerScheme.html
[56] LHCb Collaboration, R.Aaij et al., Design and performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC, JINST 14 (2019) P04013, arXiv:1812.10790
[57] W.D. Hulsbergen, Decay chain fitting with a Kalman filter, Nucl. Instrum. Meth. A 552, 566–575(2005), arXiv:physics/0503191
[58] Patrick Koppenburg et al. *StrippingBu2LLK_eeLine2* http://lhcbdoc.web.cern.ch/lhcbdoc/stripping/config/stripping28r2/leptonic/strippingbu2llk_eeline2.html
[59] Patrick Koppenburg et al. *StrippingBu2LLK_mmLine* http://lhcbdoc.web.cern.ch/lhcbdoc/stripping/config/stripping28r2/leptonic/strippingbu2llk_mmline.html
[60] Patrick Koppenburg et al. *StrippingBu2LLK_mmSSLine* http://lhcbdoc.web.cern.ch/lhcbdoc/stripping/config/stripping28r2/leptonic/strippingbu2llk_mmssline.html
[61] Patrick Koppenburg et al. *StrippingBu2LLK_eeSSLine2* http://lhcbdoc.web.cern.ch/lhcbdoc/stripping/config/stripping28r2/leptonic/strippingbu2llk_eessline2.html
[62] T. Sjostrand, S. Mrenna, P.Z. Skands, A. Brief, Introduction to PYTHIA 8.1. Comput. Phys. Commun. 178, 852–867 (2008), arXiv:0710.3820
[63] D.J. Lange, The EvtGen particle decay simulation package. Nucl. Instrum. Meth. A 462, 152–155 (2001)
[64] R_X analysis note. Not public at the time of writing this thesis.
[65] High-q^2 $R_{K^{*0}}$ analysis note. Not public at the time of writing this thesis.
[66] P. Golonka, Z. Was, PHOTOS Monte Carlo: a precision tool for QED corrections in Z and W decays. Eur. Phys. J. C 45, 97–107 (2006), arXiv:hep-ph/0506026v2
[67] Geant4 Collaboration, J. Allison, K. Amako, J. Apostolakis, H. Araujo, P. Dubois, et al., Geant4 developments and applications, IEEE Trans. Nucl. Sci. 53, 270 (2006)
[68] M. Clemencic et al., The LHCB simulation application, Gauss: design, evolution and experience. J. Phys. Conf. Ser. 331, 032023 (2011)
[69] R. Brun, F. Rademakers, S. Panacek, ROOT, an object oriented data analysis framework. Conf. Proc. C 000917, 11–42 (2000)

[70] The DecFiles package, http://lhcbdoc.web.cern.ch/lhcbdoc/decfiles/ http://lhcb-doc.web.cern.ch/lhcbdoc/decfiles/

[71] https://twiki.cern.ch/twiki/bin/view/LHCb/TupleToolMCBackgroundInfo

[72] L. Prokhorenkova et al., Catboost: Unbiased boosting with categorical features, in Proceedings of the 32Nd International Conference on Neural Information Processing Systems, NIPS'18, (USA), pp. 6639–6649, Curran Associates Inc., 2018.

[73] T. Likhomanenko et al., Reproducible Experiment Platform, J. Phys. Conf. Ser. 664 (2015), no. 5 052022, arXiv:1510.00624.

[74] Cranmer KS, Kernel Estimation in High-Energy Physics. Computer Physics Communications 136:198-207,2001, arXiv:hep-ex/0011057

[75] M. Oreglia. A Study of the Reactions $\psi' \to \gamma\gamma\psi$. PhD thesis, SLAC, 1980.

[76] D. Martínez Santos, F. Dupertuis, Mass distributions marginalized over per-event errors. Nucl. Instrum. Methods A 764, 150 (2014), arXiv:1312.5000

[77] https://towardsdatascience.com/modality-tests-and-kernel-density-estimations-3f349bb9e595