Searching for New Physics in $B^0 \to K^{*0}\mu^+\mu^-$ decays

Malte Hecker

High Energy Physics
Blackett Laboratory
Imperial College London

A thesis submitted to Imperial College London
for the degree of Doctor of Philosophy.

September 2020

Abstract

In this thesis, an angular analysis of $B^0 \to K^{*0}\mu^+\mu^-$ decays is presented. The angluar and the invariant dimuon mass squared (q^2) distributions are parameterised in an unbinned maximum likelihood fit.
The signal model is based on the decay amplitudes of $B^0 \to K^{*0}\mu^+\mu^-$ decays, including the penguin amplitudes and the amplitudes involving hadronic resonances, i.e. $B^0 \to V(\to \mu^+\mu^-)K^{*0}$ decays, where V are $c\bar{c}$ vector resonances, such as the J/ψ.
Fitting for the decay amplitudes directly allows for the determination of the vector and pseudo-vector coupling constants $\mathcal{C}_9$ and $\mathcal{C}_{10}$ from data. These can be compared to Standard Model (SM) calculations in order to search for New Physics (NP) beyond the SM.
Since both the penguin amplitudes and the hadronic amplitudes are included in the model, the interference between them can be determined directly from data. The level of interference between the resonant amplitudes and the penguin amplitudes in $B^0 \to K^{*0}\mu^+\mu^-$ decays has never been measured before and cannot be predicted from first principles in the SM. Due to that lack of knowledge, the SM predictions of $B^0 \to K^{*0}\mu^+\mu^-$ decays are currently not fully trusted in the particle physics community. This holds back the interpretation of the discrepancies between the measured q^2-binned angular observables of $B^0 \to K^{*0}\mu^+\mu^-$ decays [1] and the SM predictions, as clear indication of NP. The fit presented in this thesis is intended to help to solve this issue.

The parameterisation of the background events is complex for this analysis due to a kinematic constraint applied to all events, which causes a distortion of the background. A novel background fitting procedure and new correction techniques are presented in this thesis.

While the full dataset, which will be used for the planned publication, is not available yet, a preliminary fit to the 2011 and 2012 data, corresponding to approximately $3\,\text{fb}^{-1}$ of integrated luminosity, is presented in this thesis to demonstrate the ability of the model to describe the data. The fit results also provide a first indication, that the discrepancy between experiment and SM still persists in the angular and q^2 distributions of $B^0 \to K^{*0}\mu^+\mu^-$ decays, despite a sizeable contribution from the interference of the hadronic amplitudes with the penguin amplitudes.

Declaration of originality

The analysis presented in this thesis is the result of the research I performed between March 2017 and September 2020, with the support of the Imperial College High Energy Physics group, the Imperial College LHCb group, and an analysis team of members of the LHCb Collaboration. The analysis work presented in this thesis was performed by myself with some exceptions.
The signal model presented in chapter 3 was developed and implemented by the authors of Ref. [2], but I contributed by performing several crosschecks of the model using data and improved the computational performance of the fit.
The event selection as well as the acceptance correction and the resolution model presented in chapter 5, were carried out and developed by other members of the LHCb collaboration. I contributed with determining the resolution parameters in the J/ψ region from data (presented in section 5.3.2) as that was not part of the original analysis.
My main contribution to the analysis is the development and implementation of the background parameterisation presented in chapter 6. I also performed the full fit to Run 1 data, presented in chapter 7. However, the systematic uncertainties not related to the background (sections 7.2.2 - 7.2.6) were determined by other members of the analysis team. Also, the calculations of the angular observables and the P-Wave decay rate with the signal model, presented in sections 7.3.1 and 7.3.2, have been implemented by other members of the analysis team.
This thesis has not been submitted for any other qualification.

Malte Hecker, September 2020

Copyright declaration

Acknowledgements

First of all I would like to thank Ulrik Egede for his excellent supervision and mentoring, for providing me with the opportunity to work on this exciting analysis, and giving me freedom to pursue my own ideas while constantly guiding me with his great experience and insight. I would also like to thank Kostas Petridis for his excellent support and mentoring, for hosting me in Bristol, and for countless coding and falafel sessions. Also, I would like to thank Patrick Owen and Gabriela Pomery for the great teamwork and many helpful discussions. Furthermore, I would like to thank Mitesh Patel for his excellent support and guidance. I would also like to thank Mike McCann for many helpful discussions and his great help with running flavio for the comparison of the result to previous measurements. I would also like to thank Paula Álvarez Cartell, Mark Smith, and the Imperial College LHCb Group for many helpful discussions. I thank the Imperial HEP group as well as the LHCb collaboration.

I would also like to thank the Imperial College President's PhD Scholarship for funding my PhD.

I am grateful to Felix for the great cooperation and companionship at every stage of our PhDs. I am also grateful to Giacomo, Vukashin, Thibaud, and Sally for many great years of friendship. I owe lot of gratitude to my parents and my sister for their great support. Finally, I thank Quynhi for her unwavering support and for giving me strength and confidence.

Contents

List of Figures

List of Tables

1. Introduction

The Standard Model (SM) of particle physics is currently the most accurate theory describing the fundamental constituents of matter and how they interact. Despite that, the SM is not the ideal fundamental theory to describe our universe. For example, it does not incorporate gravity, it does not describe Dark Matter [3], and it cannot explain the observed matter-antimatter asymmetry in our universe [4]. This motivates searches for New Physics (NP) in the form of new fundamental particles beyond the Standard Model. Two types of searches for NP are typically conducted at particle accelerators. Direct searches aim at producing NP in high energy particle collisions and detecting the NP by measuring a peak in the invariant mass spectrum of the decay products. Direct searches are limited by the centre of mass energy of the collisions. Indirect Searches on the other hand, work by performing precision measurements of known decays and comparing them to SM predictions, in order to find a discrepancy between experiment and theory. Any discrepancy, also referred to as an anomaly, would point towards virtual NP contributions to the decay. Indirect searches are therefore not limited by the beam energy of the particle accelerator. However, they are dependent on the understanding of the SM predictions which often rely on approximation techniques.

In recent years, several anomalies have been found in decays involving the rare $b \to s\ell^+\ell^-$ process. These include branching fraction measurements [5–8], angular observables [1, 6, 9–12], and ratios of branching fractions between $b \to s\ell^+\ell^-$ decays with different flavours of leptons [13–15]. The anomalies could be explained by introducing new vector or axial-vector particles such as a Z' [16] or leptoquarks [17].

However, a large contribution to the overall discrepancy between $b \to s\ell^+\ell^-$ measurements and predictions comes from the angular observables of $B^0 \to K^{*0}\mu^+\mu^-$ decays. The calculation of the angular observables involves an Effective Field Theory (EFT) which largely relies on the assumption that the calculation of the quark level $b \to s\ell^+\ell^-$ process factorises with calculation of the hadron level $B^0 \to K^{*0}$ transition. This assumption breaks down for decays involving hadronic resonances in the from of $B^0 \to V(\to \mu^+\mu^-)K^{*0}$ decays, where V are $q\overline{q}$ (mainly $c\overline{c}$) vector resonances such as the J/ψ. The $q\overline{q}$ loop occurring in these decays can exchange gluons with the spectator quarks of the $B^0 \to K^{*0}$ transition such that the factorisation assumption is not valid.

In previous angular analyses of $B^0 \to K^{*0}\mu^+\mu^-$ decays, the $m_{\mu\mu}$ regions dominated by the resonant decays (e.g. the region around $m_{\mu\mu} \approx m_{J/\psi}$, where $m_{\mu\mu}$ is the reconstructed

invariant mass of the muon pair and $m_{J/\psi}$ is the true mass of the J/ψ meson) are omitted to limit the effect of the resonant decay modes. However, due to the large magnitudes of the tree level $B^0 \to V(\to \mu^+\mu^-)K^{*0}$ decays relative to the penguin decay, interference of the hadronic amplitudes with the penguin amplitudes can lead to so-called non-local contributions far away from the pole masses of the resonances. These non-local contributions alter the angular distributions and since the resonances are vector particles, the non-local contributions can mimic the presence of a virtual NP vector particle.
The level of interference is unknown and cannot be predicted from first principles in the SM calculations. Therefore, the SM calculations are currently not fully trusted, which prohibits the interpretation of the observed anomalies of the angular distributions of $B^0 \to K^{*0}\mu^+\mu^-$ decays as clear indication of the presence of NP.

The analysis presented in this thesis aims at fitting the angular and $m^2_{\mu\mu}$ (q^2) distributions of $B^0 \to K^{*0}\mu^+\mu^-$ decays including the resonant regions, using data collected with the LHCb detector. The empirical model (presented in chapter 3), which is used for this fit, is based on the decay amplitudes, including the resonant amplitudes. This allows the direct measurement of the Wilson Coefficients $\mathcal{C}_9$ and $\mathcal{C}_{10,}$ which encapsulate the vector and axial vector coupling strengths in $b \to s\ell^+\ell^-$ processes, as well as the measurement of the magnitudes and phases of the resonant amplitudes relative to the penguin amplitudes. Therefore, the level of interference is determined directly from data.

Several experimental effects need to be taken into account for this analysis, including the q^2 resolution of the detector (discussed in chapter 5) which causes the reconstructed peaks of some of the resonances to be much wider than their respective natural widths. To improve the q^2 resolution, a kinematic fit is performed, varying the reconstructed four momenta of the final state particles to best match the invariant mass of the final state particles ($m^{\text{rec}}_{B^0}$) to the true mass of the B^0 meson. For the combinatorial background events, the kinematic fit has the opposite effect, creating a $m^{\text{rec}}_{B^0}$-dependent smearing of the q^2 distribution. This complicates the background parameterisation which is presented in chapter 6.
Another effect which necessitates a novel solution is caused by one of the vetos used in the event selection, where background events of $B^+ \to K^+\mu^+\mu^-$ decays are removed by cutting on the invariant mass of the $K\mu\mu$ system. This creates a sculpting of the $\cos\theta_K$, q^2, and $m^{\text{rec}}_{B^0}$ distributions of the combinatorial background events in the upper mass side-band. This is taken into account when determining the background parameterisation as discussed in section 6.4.

The full dataset of Run 1 and Run 2 events, which is planned to be used for this analysis, is not available yet. A first fit to the Run 1 data is presented in chapter 7. It demonstrates that the empirical model used in this thesis can describe the data well. Since the empirical model uses the information contained in the q^2 dependence of the angular observables, a better experimental precision on the Wilson Coefficients $\mathcal{C}_9$ and $\mathcal{C}_{10}$ is achieved than in previous fits to $B^0 \to K^{*0}\mu^+\mu^-$ decays in Run 1. The fit gives a first preliminary indication that the discrepancy between the measured value for $\mathcal{C}_9$ and the SM prediction still persists, even when accounting for the interference between the hadronic amplitudes and the penguin amplitudes.

2. Theoretical overview of $B^0 \to K^{*0}\mu^+\mu^-$ decays

2.1 Introduction to the Standard Model of Particle Physics

The aim of particle physics is to explain the fundamental constituents of matter in the universe. The current best theory describing the fundamental particles as well as their interactions is the Standard Model (SM)[1]. The SM is a quantum field theory and interprets particles as quantum oscillations of fields. The Lagrangian, which describes the dynamics of the fields and which can be used to calculate measurable quantities, can be split into two parts

$$\mathcal{L}^{\mathrm{SM}} = \mathcal{L}^{\mathrm{gauge}} + \mathcal{L}^{\mathrm{Higgs}}. \tag{2.1}$$

$\mathcal{L}^{\mathrm{gauge}}$ describes the electromagnetic (EM), weak and strong interaction and follows the local gauge symmetry group SU(3) $\times$ SU(2) $\times$ U(1). Specifically the strong interaction obeys the SU(3) group and the electromagnetic and weak force obey the SU(2) $\times$ U(1) group. Each group has a specific number of generators (8 for SU(3) and 4 for SU(2) $\times$ U(1)) and each generator corresponds to a gauge boson. Gauge bosons, which all have spin 1, are responsible for carrying the three fundamental forces: The electromagnetic force is carried by the massless photon γ. The weak force is mediated by the massive Z, W^+ and W^- bosons. The strong force is carried by the eight massless gluons.

The particles that constitute matter are called fermions and have spin 1/2. There are two types of fermions: quarks and leptons. For every fermion there is an anti-fermion with opposite EM charge. Quarks come in six flavours which can be grouped into three generations with rising masses: The first generation quarks are the up u, down d; the second generation quarks are the charm c, strange s; the third generation quarks are the top t and beauty b. Quarks combine to hadrons which can be baryons consisting of three quarks or three antiquarks or mesons consisting of a quark and an anti-quark. More recently exotic hadronic states consisting of four or five quarks or antiquarks, referred to as Tetraquarks and Pentaquarks, have been observed [20]. There are three flavours of

[1] This section is written in reference to [18, 19]. Additional references are given in the text.

leptons with a charged lepton (e, μ, τ) and a neutrino (ν_e, ν_μ ν_τ) for each flavour. Every fermion can have either left or right handed chirality. In the massless limit, left handed chirality means the spin is parallel to the momentum of the particle, and right handed chirality means the spin is anti-parallel to the momentum.

The gauge bosons mediate the forces between particles depending on which particles they couple to based on the particles charges: The photon interacts with all EM charged particles, i.e. all quarks, the charged leptons, and the W^+ and W^- bosons. The Z, W^+, and W^- bosons interact with all fermions. Gluons interact with particles carrying color charge which are the quarks as well as the gluons themselves. The strengths of the interactions are set by the coupling constants (α for the electromagnetic interaction, α_W for the weak interaction, and α_s for the strong interaction), which are the free parameters of the gauge sector of the SM. Due to the self coupling of the gluons and the large value of α_s, processes which involve gluon exchanges cannot be calculated perturbatively to arbitrary order and therefore accuracy. Several approximation techniques are used to make predictions of hadron decays as explained in sections 2.4 and 2.5.

The Higgs sector of the SM ($\mathcal{L}^{\text{Higgs}}$) describes a scalar field which is responsible for giving mass to the Z, W^+ and W^- bosons as well as the fermions through spontaneous symmetry breaking [21–24]. A further consequence of the Higgs field is the existence of the Higgs Boson.

Expressed in the flavour basis, the Lagrangian for weak neutral current interactions of quarks can be written as

$$\mathcal{L}_{\text{NC}} = ig_W[\overline{u}_{Lj} Z_\mu \gamma^\mu u_{Lj}] \tag{2.2}$$

where Z_μ is the neutral current, g_W is the weak coupling strength and u_{Lj} denotes a left handed quark with j being the flavour index. Similarly, the Lagrangian for weak charged current interactions of quarks can be written as

$$\mathcal{L}_{\text{CC}} = \frac{ig_W}{\sqrt{2}}[W^+_\mu \overline{u}_{Lj} \gamma^\mu d_{Lj} + W^+_\mu \overline{d}_{Lj} \gamma^\mu u_{Lj}], \tag{2.3}$$

where W_μ denotes the charged current and u_{Lj} stands for a u type quark with generation index j and d_{Lj} for a d type quark.

The Lagrangians can be rotated into the physical mass basis using two unitary matrices: U^d_{ij} acting on d_L and U^u_{ij} acting on u_L. Applying this to equation 2.2 gives

$$\mathcal{L}_{\text{NC}} = ig_W[\overline{u}_{L\alpha}(U^u_L)_{\alpha j}(U^{u\dagger}_L)_{j\beta} Z_\mu \gamma^\mu u_{L\beta}] = ig_W[\overline{u}_{L\alpha}\delta_{\alpha\beta}] Z_\mu \gamma^\mu u_{L\beta}. \tag{2.4}$$

Due to the unitary matrices cancelling in equation 2.4, the Lagrangian does not allow any off diagonal transitions that would mix quark flavours. In other words, tree-level[2] flavour changing neutral currents (FCNC) are forbidden in the SM. However, when applying the same rotations to equation 2.3 we get

$$\mathcal{L}_{\mathrm{CC}} = \frac{ig_W}{\sqrt{2}}[W^+_\mu \overline{u}_{L\alpha}(U^u_L)_{\alpha j}(U^{d\dagger}_L)_{j\beta}\gamma^\mu d_{Lj} + W^+_\mu \overline{d}_{Lj}(U^d_L)_{\alpha j}(U^{u\dagger}_L)_{j\beta}\gamma^\mu u_{Lj}]. \tag{2.5}$$

Thus, for charged weak currents the matrices do not cancel and the Lagrangian does allow flavour changing transitions between generations α and β. The strengths of these transitions are free parameters in the SM an are summarised in the CKM matrix [25, 26] $V_{\alpha\beta}$ which is commonly expressed in the Wolfenstein parameterisation [27]:

$$V_{\alpha\beta} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} = \begin{pmatrix} 1-\lambda^2/2 & \lambda & 0 \\ -\lambda & 1-\lambda^2/2 & A\lambda^2 \\ 0 & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^3), \tag{2.6}$$

with $\lambda \approx 0.22$ and $A \approx 1.25$. In the third order terms, the CKM matrix allows for *CP* violation, which is the difference of a physical system after charge conjugation (changing all particles for anti-particles and changing all anti-particles for particles) and parity transformation (reversing the sign of all spacial coordinates). The CKM matrix is the only source for *CP* violation in the SM.

As stated above, Lagrangians can be used to calculate observable quantities, such as particles' lifetimes or differential decay rates. To this end, the modulus squared of the amplitude $\mathcal{A}$ of a decay is calculated with

$$|\mathcal{A}|^2 = |\langle f \,|\, \mathcal{O} \,|\, i\rangle|^2 \,, \tag{2.7}$$

where $|i\rangle$ and $|f\rangle$ are the initial and initial states of the fields. For example $|i\rangle = |b\rangle$ and $|s\ell^+\ell^-\rangle$ for a b quark decaying into an s quark and a lepton pair, which is discussed in section 2.3. The operator $\mathcal{O}$ is a function of the interaction Lagrangian which describes the interaction through which the transition from $|i\rangle$ to $|f\rangle$ occurs. Generally, $\mathcal{A}$ cannot be computed fully analytically in the SM and several approximation methods are used. The Lagrangian used for calculating $b \to s\ell^+\ell^-$ decays is discussed in section 2.4.

[2] i.e. to first order

2.2 Indirect searches for New Physics

Despite its successes, the SM is far from being the ideal fundamental theory to describe our universe. The shortcomings of the SM include:

- It does not incorporate gravity. A complete model of nature should be able to describe all fundamental forces.
- The particles described by the SM make up only 4.9% of the universe [3, 28].
- The amount of *CP* violation allowed by the CKM matrix is about ten orders of magnitude too low in order to explain the matter-antimatter asymmetry observed in our universe [4].

These facts motivate searching for new physics (NP) beyond the SM. There are two different approaches for this: The first approach is a *direct search* which attempts to detect a new fundamental particle by creating it in a high energy collision and reconstructing it via its decay products. The advantage of this approach is that discovering NP is free of theoretical uncertainties. The disadvantage is that the search is limited by the collision energy. The second approach for finding NP is the *indirect search*. It works by performing precision measurements of known processes that can be predicted by the SM and looking for a discrepancy between experiment and theory. NP can modify the processes virtually, i.e. without the NP particle being produced on-shell. Thus, indirect searches can be sensitive to NP scales far greater than the collision energy. The disadvantage of indirect searches is that they require a good understanding of the SM prediction in order to correctly interpret any difference between experiment and theory as a discovery of NP.

The measurable observable used for an indirect search can be expressed as a sum of the amplitude describing the SM interaction A_{SM} and the amplitude describing the NP interaction A_{NP}

$$Q \sim |A_{\mathrm{SM}} + A_{\mathrm{NP}}|^2. \tag{2.8}$$

Since the sensitivity of Q to A_{NP} is larger when A_{SM} is small, decays which are heavily suppressed in the SM are ideal for searching for NP.

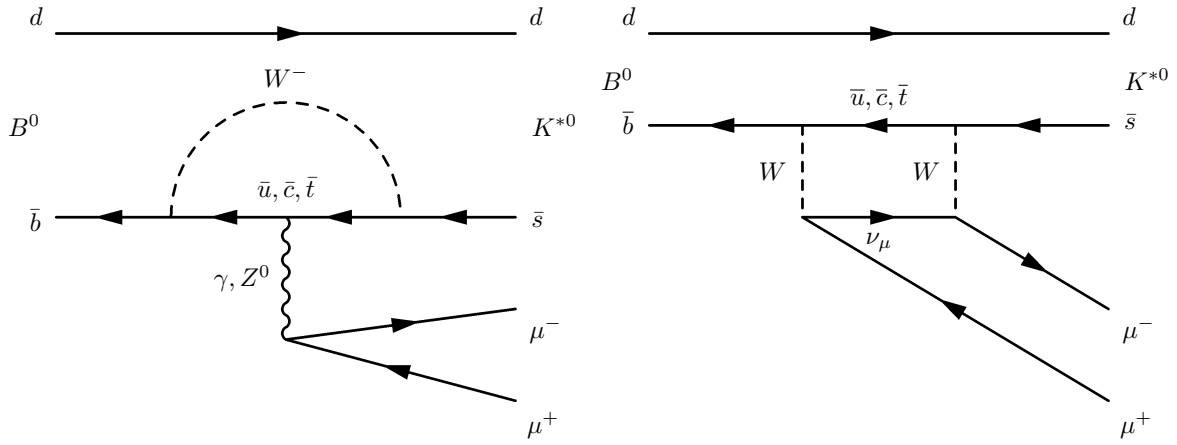


Figure 2.1: Feynman diagrams describing the $B^0 \to K^{*0}\mu^+\mu^-$ decay in the lowest order in the SM. The penguin loop diagram (left) involves a W boson and either a γ or Z boson. The box diagram involves two W bosons. Figures from Ref. [29].

2.3 Flavour changing neutral currents of B mesons

As discussed in section 2.1, in the SM, FCNC transitions are forbidden at tree-level. However, FCNC can still occur in the SM via loop-level processes. One example for this is the $B^0 \to K^{*0}\mu^+\mu^-$ decay which involves a $\overline{b} \to \overline{s}\ell^+\ell^-$ transition[3]. The corresponding Feynman diagrams for the electroweak penguin process and the box process are shown in figure 2.1. Due to the additional weak couplings in the loop, the $B^0 \to K^{*0}\mu^+\mu^-$ decay is heavily suppressed and has a measured branching fraction of $(1.06 \pm 0.10) \times 10^{-6}$ [30]. This makes the $B^0 \to K^{*0}\mu^+\mu^-$ decay an excellent candidate for indirect searches for NP, because NP interactions which allow $b \to s\ell^+\ell^-$ transitions can have a sizeable impact on the differential decay rate. Furthermore, the $B^0 \to K^{*0}\mu^+\mu^-$ decay, which has leptons in the final state is preferred over rare B^0 decays with fully hadronic final states, because the latter are much harder to predict since the assumption of the factorisation of the quark level transition and the hadronic transition (see section 2.5) does not hold to the same degree. Experimentally, muons are preferred in the final state over electrons, since muons produce less bremstrahlung and can therefore be measured with a higher resolution. Furthermore, muons are easier to trigger on and to identify as explained in section 4.6.

In order to compare experimental data to the SM prediction, a framework for calculating $b \to s\ell^+\ell^-$ processes is needed, which is explained in the next section.

[3]Throughout the thesis charge conjugation is implied unless specific *CP* states are explicitly discussed.

2.4 Effective Field theory for $b \to s\ell^+\ell^-$ decays

In order to calculate $b \to s\ell^+\ell^-$ processes in a model-independent framework that also allows sensitivity to NP, an Effective Field Theory (EFT) is devised using the Operator Product Expansion (OPE) method [31]. It works by separating physical effects based on their energy scales: All heavy degrees of freedom above a chosen energy scale μ, such as the mass of the W-boson m_W, are separated from the lighter degrees of freedom which are smaller than μ. Since the heavy degrees of freedom are characterised by a small distance scale of $1/m_W \approx 10^{-2}\,\text{GeV}^{-1}$, they are referred to as short-distance. Correspondingly, the lighter degrees of freedom which are of the order $1/\Lambda_{\text{QCD}} \approx 5\,\text{GeV}^{-1}$ (assuming $\Lambda_{\text{QCD}} \approx 200\,\text{MeV}$) are referred to as long distance.

Decoupling the long distance and short distance effects allows writing the matrix element of an effective Hamiltonian $\mathcal{H}_{\text{eff}}$ at a given scale as

$$\langle f|\,\mathcal{H}_{\text{eff}}\,|i\rangle = \sum_i \mathcal{C}_i(\mu)\,\langle f|\,\mathcal{O}_i\,|i\rangle\,|_\mu \tag{2.9}$$

where $\mathcal{C}_i$ - referred to as Wilson Coefficients - are complex numbers that encapsulate the short distance contributions, whereas $\mathcal{O}_i$ are operators describing the long distance contributions.

By construction, the effect of heavy NP is encoded into the Wilsons coefficients. Therefore, by calculating the values of the Wilson Coefficients in the SM ($\mathcal{C}_i^{\text{SM}}$) and comparing to the values of the Wilson coefficients that best describe experimental data $\mathcal{C}_i$, the contributions from NP can be determined with

$$\mathcal{C}_i^{\text{NP}} = \mathcal{C}_i - \mathcal{C}_i^{\text{SM}}. \tag{2.10}$$

For $b \to s\ell^+\ell^-$ decays the complete effective Hamiltonian can then be written as [32,33]

$$\mathcal{H}_\varepsilon = -\frac{4\,G_F}{\sqrt{2}}\left(V_{tb}V_{ts}^*\mathcal{H}_\varepsilon^{(t)} + V_{ub}V_{us}^*\mathcal{H}_\varepsilon^{(u)}\right) \tag{2.11}$$

with

$$\begin{aligned}
\mathcal{H}_\varepsilon^{(t)} &= C_1\mathcal{O}_1^c + C_2\mathcal{O}_2^c + \sum_{i=3}^{6} C_i\mathcal{O}_i + \sum_{i=7,8,9,10,P,S} \left(C_i\mathcal{O}_i + C_i'\mathcal{O}_i'\right), \\
\mathcal{H}_\varepsilon^{(u)} &= C_1(\mathcal{O}_1^c - \mathcal{O}_1^u) + C_2(\mathcal{O}_2^c - \mathcal{O}_2^u).
\end{aligned}$$

Due to the small size of the relevant CKM elements, contributions from $\mathcal{H}_\varepsilon^{(u)}$ are doubly CKM-suppressed and usually neglected [34]. Also the scalar and pseudoscalar operators ($\mathcal{O}_S$ and $\mathcal{O}_P$) are highly suppressed and usually ignored [32].

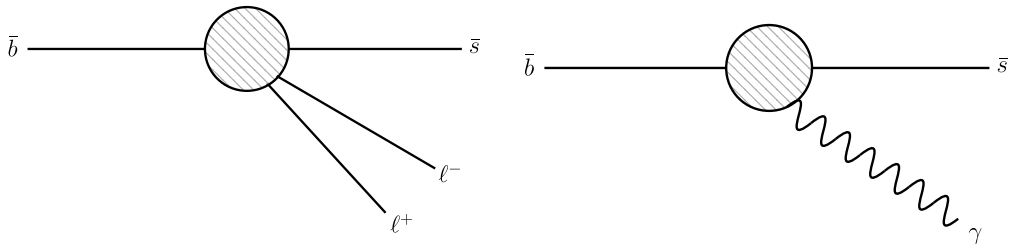


Figure 2.2: Diagrams illustrating the EFT operators that contribute to $b \to s\ell^+\ell^-$ decays. Left describes the semileptonic operators $\mathcal{O}_9$ and $\mathcal{O}_{10}$ involving a vector current. Right depicts the electromagnetic operator $\mathcal{O}_7$ involving a photon current.

In the SM, the largest contributions to $b \to s\ell^+\ell^-$ transitions come from the three operators $\mathcal{O}_{7,9,10}$ which can be expressed as [32]:

$$\mathcal{O}_7 = \frac{e}{g^2} m_b (\bar{s}\sigma_{\mu\nu} P_R b) F^{\mu\nu}, \qquad \mathcal{O}'_7 = \frac{e}{g^2} m_b (\bar{s}\sigma_{\mu\nu} P_L b) F^{\mu\nu}, \tag{2.12}$$

$$\mathcal{O}_9 = \frac{e^2}{g^2} (\bar{s}\gamma_\mu P_L b)(\bar{\mu}\gamma^\mu \mu), \qquad \mathcal{O}'_9 = \frac{e^2}{g^2} (\bar{s}\gamma_\mu P_R b)(\bar{\mu}\gamma^\mu \mu), \tag{2.13}$$

$$\mathcal{O}_{10} = \frac{e^2}{g^2} (\bar{s}\gamma_\mu P_L b)(\bar{\mu}\gamma^\mu \gamma_5 \mu), \qquad \mathcal{O}'_{10} = \frac{e^2}{g^2} (\bar{s}\gamma_\mu P_R b)(\bar{\mu}\gamma^\mu \gamma_5 \mu), \tag{2.14}$$

where g is the strong coupling constant, $P_{L,R} = (1 \pm \gamma_5)/2$ are the left/right chirality projections, $F^{\mu\nu}$ is the electromagnetic field tensor and $\sigma_{\mu\nu}$ are the Pauli-spin matrices. The primed operators $\mathcal{O}'_{7,9,10}$ denote the operators with opposite chirality which vanish or are highly suppressed in the SM [32].

$\mathcal{O}_7$ is the electromagnetic operator corresponding to the radiation of a photon, whereas $\mathcal{O}_9$ and $\mathcal{O}_{10}$ are the semileptonic vectors which correspond to a vector and axial-vector current respectively. With respect to the SM diagrams (figure 2.1), $\mathcal{O}_7$ corresponds to the photon penguin diagram, whereas $\mathcal{O}_9$ and $\mathcal{O}_{10}$ correspond to the Z and W box diagrams. Within the EFT, $\mathcal{O}_7$, $\mathcal{O}_9$ and $\mathcal{O}_{10}$ can be illustrated with simple diagrams where all heavy particles are encapsulated into effective couplings as shown in figure 2.2. This also shows the similarity of the EFT for $b \to s\ell^+\ell^-$ decays to Fermi's effective theory of weak interactions, which in the limit $q^2 \ll m_W^2$ was able to describe the beta decay accurately with a four-fermion vertex before electroweak theory was devised and before the W was discovered [18].

The Wilson coefficients can be expanded perturbatively as [32, 33]

$$C_i = C_i^{(0)} + \frac{\alpha_s}{4\pi} C_i^{(1)} + \left(\frac{\alpha_s}{4\pi}\right)^2 C_i^{(2)} + O(\alpha_s^3)\,, \tag{2.15}$$

where $C_i^{(n)}$ refers to the n-loop level contribution. The values for $C_i^{(n)}$ are calculated by matching them to the full SM processes at the scale $\mu = m_W$ as described in Ref. [33]. Since tree level FCNC are forbidden in the SM this means C_i^0 becomes zero[4]. The Wilson coefficients are calculated up to two loop contributions i.e. with next-to-next-leading order (NNLO) precision. Subsequently, the Wilson Coefficients are evolved from $\mu = m_W$ to $\mu = m_b$ using a Renormalization Group Equation [33]. Due to the NNLO terms the Wilson Coefficients from equation 2.11 end up mixing, i.e. one cannot necessarily assume a specific contribution to a given Wilson Coefficient. Therefore, it is common to express the *effective* Wilson Coefficients which for $C_{7,9,10}^{\text{eff}}$ are given by:

$$\begin{aligned} C_7^{\text{eff}} &= \frac{4\pi}{\alpha_s} C_7 - \frac{1}{3} C_3 - \frac{4}{9} C_4 - \frac{20}{3} C_5 - \frac{80}{9} C_6 \,, \\ C_9^{\text{eff}} &= \frac{4\pi}{\alpha_s} C_9 + Y(q^2) \,, \\ C_{10}^{\text{eff}} &= \frac{4\pi}{\alpha_s} C_{10} \,, \qquad C_{7,8,9,10}^{\prime,\text{eff}} = \frac{4\pi}{\alpha_s} C'_{7,8,9,10} \,, \end{aligned} \tag{2.16}$$

where $Y(q^2)$ is a large collection of terms including $C_{1,2,3,4,5,6}$ and can be found in Ref. [32].

The values of the Wilson coefficients depend on the choice of the exact values of m_W and m_b for setting the energy scales in the OPE, as well as m_Z and m_t when integrating out the heavy particles. This introduces a phenomenological uncertainty in the EFT predictions of $b \to s\ell^+\ell^-$ processes, as investigated in Ref. [33]. The values for the Wilson coefficients obtained in Ref. [32] are C_9^{eff} - $Y(q^2)$ = 4.211; C_{10}^{eff} = -4.103; C_7^{eff} = -0.304; $C_7'^{\text{eff}}$ = -0.0061.

The relative contributions of the Wilson Coefficients C_7^{eff}, C_9^{eff} and C_{10}^{eff} to the decay rate of $B^0 \to K^{*0}\mu^+\mu^-$ vary as a function of the invariant mass of the muon pair squared (q^2). C_7^{eff} dominates the region of $q^2 \leq 1\text{GeV}^2/c^2$, as it is more likely to radiate a virtual photon at low q^2, thus causing a photon pole in that kinematic region. With rising q^2, contributions from C_9^{eff} and C_{10}^{eff} dominate the spectrum. Thus, the best experimental sensitivity to all three dominant Wilson coefficients can be obtained by fitting data across the full q^2 spectrum.

Experimentally, the quark level $b \to s\ell^+\ell^-$ transition is not observed directly but within the hadronic decay of $B^0 \to K^{*0}\mu^+\mu^-$. Therefore, in order to parameterise experimental data and extract values for the Wilson coefficients, a model is required which describes the hadronic $B^0 \to K^{*0}$ transition, as explained in the following section.

[4]except for $\mathcal{O}_2$ [32].

2.5 Form Factors

When calculating $B^0 \to K^{*0}\mu^+\mu^-$ decays, the quark level $b \to s\ell^+\ell^-$ transitions and the $B^0 \to K^{*0}$ transition are assumed to factorise, which is known as the naive factorisation assumption [32].

The matrix elements which describe the hadronic $B^0 \to K^{*0}$ transition are referred to as form factors (FFs). They can only be calculated non-perturbatively. Two different methods are commonly used to calculate the FFs: The Light Cone Sum Rules (LCSR) method which is only valid at low q^2 and Lattice QCD which is valid at high q^2. For the analysis presented in this thesis, a continuous parameterisation of the form factors across the full q^2 spectrum is needed. This has been achieved in Ref. [35], where the six FFs needed to describe $B^0 \to K^{*0}\mu^+\mu^-$ decays (V, A_1, A_{12}, T_1, T_2, T_{23}) have been calculated using LCSR and Lattice QCD and subsequently parameterised as a function of q^2 using the expression [35]:

$$F_i(q^2) = \frac{1}{1 - q^2/m_{R,i}^2} \sum_{k=0}^{2} \alpha_k^i \left[z(q^2) - z(0) \right]^k , \tag{2.17}$$

where

$$z(t) = \frac{\sqrt{t_+ - t} - \sqrt{t_+ - t_0}}{\sqrt{t_+ - t} + \sqrt{t_+ - t_0}} , \tag{2.18}$$

with $t_\pm \equiv (m_B \pm m_{K^*})^2$ and $t_0 \equiv t_+(1 - \sqrt{1 - t_-/t_+})$. $m_{R,i}$ denotes the masses of the excited B^0 states as given in Ref. [35]. α_k^i are the FF parameters. Equation 2.17 is truncated after the 3rd order (quadratic in z) which was found to describe the FFs well [35]. Thus there are three free parameters for each of the six form factors.

The parameters are obtained in a combined fit to the calculated FFs while taking into account all correlations between the FFs. Figure 2.3 shows the calculated FFs in bins of q^2 obtained from LCSR (red points) and Lattice QCD (blue points), as well as the parameterised FFs as a function of q^2 (grey band). The authors of [35] published the best fit values for all α_k^i as well as the full covariance matrix. These are used in the fit to LHCb data described in this thesis to constrain the FFs.

The actual decay being observed experimentally is not $B^0 \to K^{*0}\mu^+\mu^-$ but $B^0 \to K^{*0}(\to K^+\pi^-)\mu^+\mu^-$. In order to be able to use the $B^0 \to K^{*0}$ FFs, the K^{*0} is treated as a stable particle in the decay of the B^0 meson and the decay of the K^{*0} is then treated as an independent process [32]. The impact of this approximation is a currently debated issue and is explored in Ref. [36].

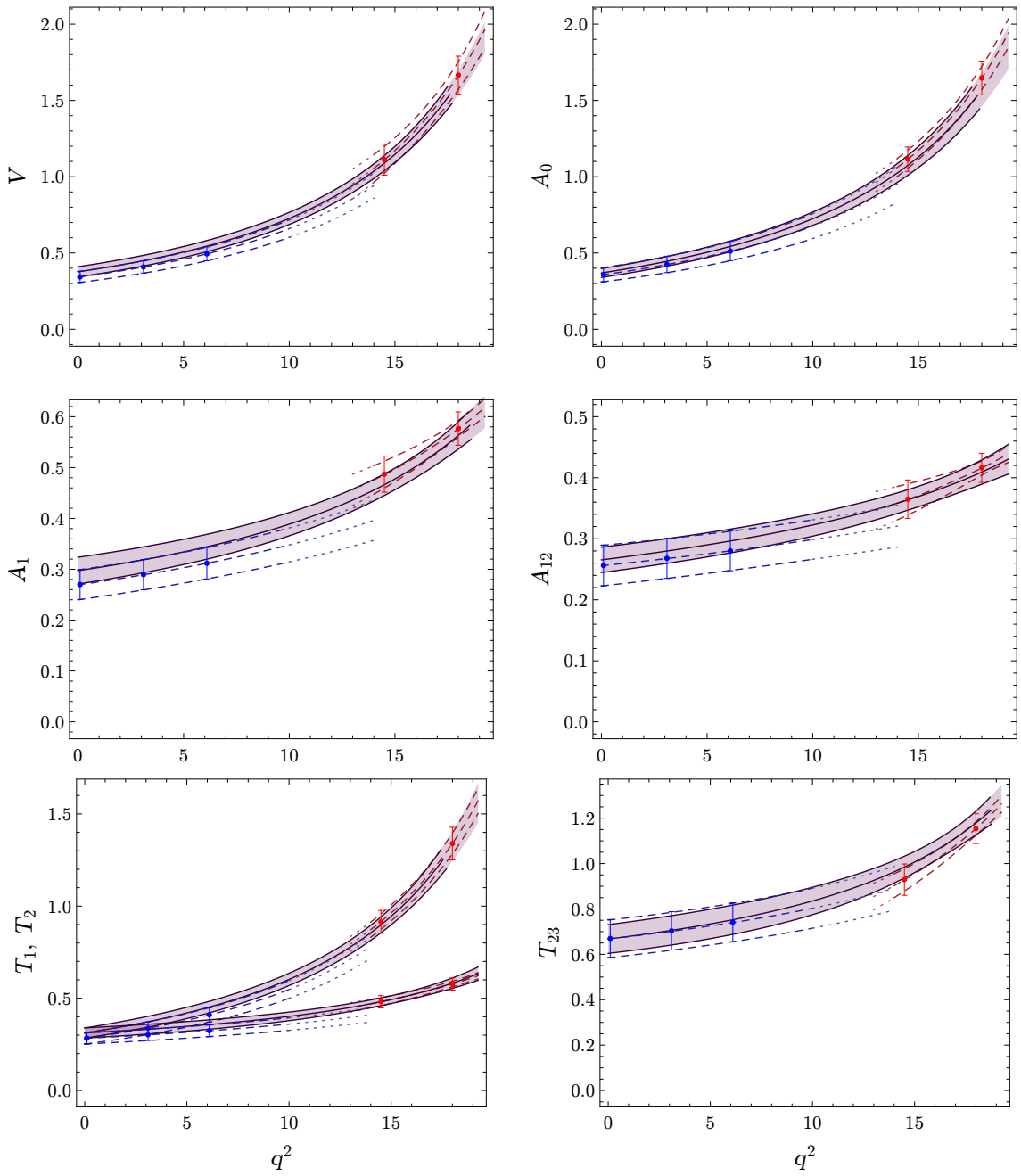


Figure 2.3: Form factors obtained from LCSR (red points) and Lattice QCD (blue points). The grey band shows the combined z-expansion fit. Figure from Ref. [35].

2.6 Angular definitions

Considering only the P-wave configuration (angular momentum of 1) of the $K^+\pi^-$ system[5], the decay $B^0 \to K^{*0}(\to K^+\pi^-)\mu^+\mu^-$ can be fully described by three decay angles and q^2, which as previously stated is the squared invariant mass of the dimuon system. There are several conventions for the definition of the angles. In this thesis, the LHCb convention is used which is defined in Ref. [8] and also used in Refs. [1, 9, 10]. Under this convention, the angular distributions of the $B^0 \to K^{*0}\mu^+\mu^-$ and $\overline{B}^0 \to \overline{K}^{*0}\mu^+\mu^-$ decays are the same, assuming there is no CP violation.[6].

$\cos\theta_\ell$ is defined as the cosine of the angle θ_ℓ between the direction of the μ^+ (μ^-) in the rest frame of the dimuon system and the direction of the dimuon in the B^0 ($\overline{B}^0$) rest frame. The explicit definition of $\cos\theta_\ell$ for $B^0 \to K^{*0}\mu^+\mu^-$ decays is given by

$$\cos\theta_\ell = \left(\hat{p}_{\mu^+}^{(\mu^+\mu^-)}\right) \cdot \left(\hat{p}_{\mu^+\mu^-}^{(B^0)}\right) = \left(\hat{p}_{\mu^+}^{(\mu^+\mu^-)}\right) \cdot \left(-\hat{p}_{B^0}^{(\mu^+\mu^-)}\right) \ , \tag{2.19}$$

where $\hat{p}_X^{(Y)}$ are unit vectors for the direction of particle X in the rest frame of the system Y. The definition of $\cos\theta_\ell$ for $\overline{B}^0 \to \overline{K}^{*0}\mu^+\mu^-$ is:

$$\cos\theta_\ell = \left(\hat{p}_{\mu^-}^{(\mu^+\mu^-)}\right) \cdot \left(\hat{p}_{\mu^+\mu^-}^{(\overline{B}^0)}\right) = \left(\hat{p}_{\mu^-}^{(\mu^+\mu^-)}\right) \cdot \left(-\hat{p}_{\overline{B}^0}^{(\mu^+\mu^-)}\right) \ . \tag{2.20}$$

Similarly, $\cos\theta_K$ is defined as the cosine of the angle θ_K between the direction of the K^+ (K^-) in the K^{*0} ($\overline{K}^{*0}$) rest frame and the direction of the K^{*0} ($\overline{K}^{*0}$) in the rest frame of the B^0 ($\overline{B}^0$). Explicitly, $\cos\theta_K$ for $B^0 \to K^{*0}\mu^+\mu^-$ decays is defined as

$$\cos\theta_K = \left(\hat{p}_{K^+}^{(K^{*0})}\right) \cdot \left(\hat{p}_{K^{*0}}^{(B^0)}\right) = \left(\hat{p}_{K^+}^{(K^{*0})}\right) \cdot \left(-\hat{p}_{B^0}^{(K^{*0})}\right) . \tag{2.21}$$

For $\overline{B}^0 \to \overline{K}^{*0}\mu^+\mu^-$ decays the definition of $\cos\theta_K$ is:

$$\cos\theta_K = \left(\hat{p}_{K^-}^{(K^{*0})}\right) \cdot \left(\hat{p}_{K^{*0}}^{(\overline{B}^0)}\right) = \left(\hat{p}_{K^-}^{(K^{*0})}\right) \cdot \left(-\hat{p}_{\overline{B}^0}^{(K^{*0})}\right) \tag{2.22}$$

The angle ϕ is defined as the angle between the plane containing the μ^+ and μ^- and the plane containing the kaon and pion from the K^{*0} ($\overline{K}^{*0}$). Explicitly for $B^0 \to K^{*0}\mu^+\mu^-$ decays ϕ is defined with

$$\cos\phi = \left(\hat{p}_{\mu^-}^{(\overline{B}^0)} \times \hat{p}_{\mu^+}^{(\overline{B}^0)}\right) \cdot \left(\hat{p}_{K^-}^{(\overline{B}^0)} \times \hat{p}_{\pi^+}^{(\overline{B}^0)}\right) \ , \tag{2.23}$$

[5] See section 2.9 for a discussion of the S-wave contribution.

[6] For a definition of the 'theorists convention' and how to convert between conventions see [37].

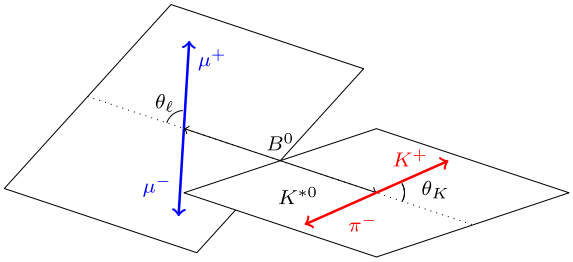
$$\sin\phi = -\left[\left(\hat{p}_{\mu^-}^{(\overline{B}^0)} \times \hat{p}_{\mu^+}^{(\overline{B}^0)}\right) \times \left(\hat{p}_{K^-}^{(\overline{B}^0)} \times \hat{p}_{\pi^+}^{(\overline{B}^0)}\right)\right] \cdot \hat{p}_{\overline{K}^{*0}}^{(\overline{B}^0)} \tag{2.24}$$

whereas for $\overline{B}^0 \to \overline{K}^{*0}\mu^+\mu^-$ decays ϕ is defined with

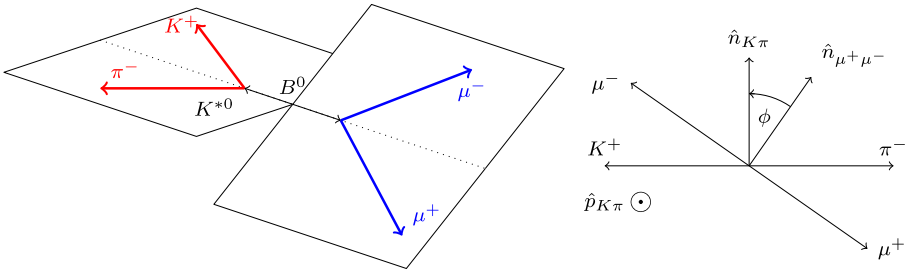
$$\cos\phi = \left(\hat{p}_{\mu^+}^{(B^0)} \times \hat{p}_{\mu^-}^{(B^0)}\right) \cdot \left(\hat{p}_{K^+}^{(B^0)} \times \hat{p}_{\pi^-}^{(B^0)}\right) \ , \tag{2.25}$$

$$\sin\phi = \left[\left(\hat{p}_{\mu^+}^{(B^0)} \times \hat{p}_{\mu^-}^{(B^0)}\right) \times \left(\hat{p}_{K^+}^{(B^0)} \times \hat{p}_{\pi^-}^{(B^0)}\right)\right] \cdot \hat{p}_{K^{*0}}^{(B^0)} \tag{2.26}$$

Figure 2.4 shows an illustration of the definition of the angles for $B^0 \to K^{*0}\mu^+\mu^-$ decays. The differential decay rate of $B^0 \to K^{*0}\mu^+\mu^-$ and $\overline{B}^0 \to \overline{K}^{*0}\mu^+\mu^-$ decays can be expressed using the angles defined above as shown in the next section.



(a) θ_K and θ_ℓ definitions for the B^0 decay



(b) ϕ definition for the B^0 decay

Figure 2.4: Illustration of the definition of the angles θ_L, θ_K and ϕ for $B^0 \to K^{*0}\mu^+\mu^-$ decays. Figure from Ref. [8].

2.7 Differential decay rate

The differential decay rate of $B^0 \to K^{*0}\mu^+\mu^-$ decays can be calculated by squaring the matrix element for $B^0 \to K^{*0}\mu^+\mu^-$ decays from the EFT (see section 2.4) and summing

over the spins. When considering only the P-wave configuration of the $K^+\pi^-$ system and also ignoring scalar contributions to the dimuon system, the differential decay rate can be expressed in terms of the decay angles, defined in the previous section, and q^2 [38]:

$$\begin{aligned}\frac{\mathrm{d}^4\Gamma[B^0\to K^{*0}\mu^+\mu^-]}{\mathrm{d}q^2\mathrm{d}\vec{\Omega}} =& \frac{9}{32\pi}\sum_i J_i(q^2) f_i(\cos\theta_l, \cos\theta_K, \phi) \\ =& \frac{9}{32\pi}\Big[J_1^s \sin^2\theta_\mathrm{K} + J_1^c \cos^2\theta_\mathrm{K} \\ &+ J_2^s \sin^2\theta_\mathrm{K}\cos 2\theta_\mathrm{l} + J_2^c \cos^2\theta_\mathrm{K}\cos 2\theta_\mathrm{l} \\ &+ J_3 \sin^2\theta_\mathrm{K}\sin^2\theta_\mathrm{l}\cos 2\phi + J_4 \sin 2\theta_\mathrm{K}\sin 2\theta_\mathrm{l}\cos\phi \\ &+ J_5 \sin 2\theta_\mathrm{K}\sin\theta_\mathrm{l}\cos\phi + J_6^s \sin^2\theta_\mathrm{K}\cos\theta_\mathrm{l} \\ &+ J_7 \sin 2\theta_\mathrm{K}\sin\theta_\mathrm{l}\sin\phi + J_8 \sin 2\theta_\mathrm{K}\sin 2\theta_\mathrm{l}\sin\phi \\ &+ J_9 \sin^2\theta_\mathrm{K}\sin^2\theta_\mathrm{l}\sin 2\phi\Big] \,.\end{aligned} \tag{2.27}$$

The ($\cos\theta_\ell$, $\cos\theta_K$, ϕ) dependence of the decay rate is introduced by spherical harmonics which multiply each of the q^2 dependent angular observables $J_i(q^2)$. The angular observables $J_i(q^2)$ encapsulate the complete information that can be extracted from measurements of $B^0 \to K^{*0}\mu^+\mu^-$ [32]. They can be expressed as bi-linear combinations of the 6 complex transversity amplitudes $\mathcal{A}_0^{\mathrm{L,R}}$, $\mathcal{A}_\parallel^{\mathrm{L,R}}$, $\mathcal{A}_\perp^{\mathrm{L,R}}$ which are presented in the following section (section 2.8).

Explicitly, the angular observables are given by [38]:

$$
\begin{aligned}
J_{1s} &= \frac{(2+\beta_\ell^2)}{4}\left[|A_\perp^L|^2+|A_\parallel^L|^2+|A_\perp^R|^2+|A_\parallel^R|^2\right]+\frac{4m_\ell^2}{q^2}\mathrm{Re}\left(A_\perp^L A_\perp^{R*}+A_\parallel^L A_\parallel^{R*}\right),\\
J_{1c} &= |A_0^L|^2+|A_0^R|^2+\frac{4m_\ell^2}{q^2}\left[|A_t|^2+2\mathrm{Re}(A_0^L A_0^{R*})\right]+\beta_\ell^2|A_S|^2,\\
J_{2s} &= \frac{\beta_\ell^2}{4}\left[|A_\perp^L|^2+|A_\parallel^L|^2+|A_\perp^R|^2+|A_\parallel^R|^2\right],\\
J_{2c} &= -\beta_\ell^2\left[|A_0^L|^2+|A_0^R|^2\right],\\
J_3 &= \frac{1}{2}\beta_\ell^2\left[|A_\perp^L|^2-|A_\parallel^L|^2+|A_\perp^R|^2-|A_\parallel^R|^2\right],\\
J_4 &= \frac{1}{\sqrt{2}}\beta_\ell^2\left[\mathrm{Re}(A_0^L A_\parallel^{L*}+A_0^R A_\parallel^{R*})\right],\\
J_5 &= \sqrt{2}\beta_\ell\left[\mathrm{Re}(A_0^L A_\perp^{L*}-A_0^R A_\perp^{R*})-\frac{m_\ell}{\sqrt{q^2}}\mathrm{Re}(A_\parallel^L A_S^*+A_\parallel^{R*}A_S)\right],\\
J_{6s} &= 2\beta_\ell\left[\mathrm{Re}(A_\parallel^L A_\perp^{L*}-A_\parallel^R A_\perp^{R*})\right],\\
J_7 &= \sqrt{2}\beta_\ell\left[\mathrm{Im}(A_0^L A_\parallel^{L*}-A_0^R A_\parallel^{R*})+\frac{m_\ell}{\sqrt{q^2}}\mathrm{Im}(A_\perp^L A_S^*-A_\perp^{R*}A_S))\right],\\
J_8 &= \frac{1}{\sqrt{2}}\beta_\ell^2\left[\mathrm{Im}(A_0^L A_\perp^{L*}+A_0^R A_\perp^{R*})\right],\\
J_9 &= \beta_\ell^2\left[\mathrm{Im}(A_\parallel^{L*}A_\perp^L+A_\parallel^{R*}A_\perp^R)\right],
\end{aligned}
\tag{2.28}
$$

where the parameter β_ℓ is given by

$$
\beta_\ell=\sqrt{1-\frac{4m_\ell^2}{q^2}}. \tag{2.29}
$$

Due to the choice of the angular basis in the LHCb convention, the differential decay rate for the *CP* transformed $\overline{B}^0\to\overline{K}^{*0}\mu^+\mu^-$ decay,

$$
\frac{\mathrm{d}^4\bar{\Gamma}[B^0\to K^{*0}\mu^+\mu^-]}{\mathrm{d}q^2\mathrm{d}\vec{\Omega}}=\frac{9}{32\pi}\sum_i\bar{J}_i(q^2)f_i(\cos\theta_l,\cos\theta_K,\phi), \tag{2.30}
$$

results in the same angular distribution as the $B^0\to K^{*0}\mu^+\mu^-$ decay. Therefore, one can sum over the decay rates to determine a set of *CP*-averaged observables S_i:

$$
S_i=\left(J_i+\bar{J}_i\right)\Big/\left(\frac{\mathrm{d}\Gamma}{\mathrm{d}q^2}+\frac{\mathrm{d}\bar{\Gamma}}{\mathrm{d}q^2}\right). \tag{2.31}
$$

This allows merging the $B^0 \to K^{*0}\mu^+\mu^-$ and $\overline{B}^0 \to \overline{K}^{*0}\mu^+\mu^-$ data sets, yielding twice the number of events available for fitting.[7]

The observable S_{6s} is a related to the forwared-backward asymmetry of the dimuon system and is therefore commonly replaced by $A_{\rm FB} = \frac{3}{4}S_{6s}$. The observable S_{1c} is a measure of the fraction of the longitudinal polarisation of the K^{*0} and is commonly referred to as $F_L = S_{1c}$. It is also common to define *optimized* angular observables for which the $B^0 \to K^{*0}$ FF uncertainties cancel to first order and thus allow more precise SM predictions. The convention for the optimized observables used by LHCb is [1, 10]:

$$\begin{aligned} P_1 &= \frac{2\,S_3}{(1-F_{\rm L})} = A_{\rm T}^{(2)}\,, \\ P_2 &= \frac{2}{3}\frac{A_{\rm FB}}{(1-F_{\rm L})}\,, \\ P_3 &= \frac{-S_9}{(1-F_{\rm L})}\,, \\ P'_{4,5,8} &= \frac{S_{4,5,8}}{\sqrt{F_{\rm L}(1-F_{\rm L})}}\,, \\ P'_6 &= \frac{S_7}{\sqrt{F_{\rm L}(1-F_{\rm L})}}\,. \end{aligned} \tag{2.32}$$

It should be noted that in some theoretical predictions the definition of the optimized observables are different [39].

2.8 Transversity amplitudes

The K^{*0} in the $B^0 \to K^{*0}\mu^+\mu^-$ decay has three possible polarisation states: longitudinal (denoted with '0'), parallel (denoted with '$\|$') or perpendicular (denoted with '$\perp$'). The decay with each polarisation state is described by a corresponding amplitude. Furthermore, the muon system can either have left or right handed chirality (denoted L and R). Thus, there are six complex amplitudes $\mathcal{A}_0^{\rm L,R}$, $\mathcal{A}_\|^{\rm L,R}$, $\mathcal{A}_\perp^{\rm L,R}$ that can fully describe the $B^0 \to K^{*0}\mu^+\mu^-$ decay rate in the P-wave configuration. The amplitudes contain the Wilson coefficients C_7, C_9 and C_{10} as well as the FFs described in section 2.5 and can be written as [2]:

$$\mathcal{A}_0^{\rm L,R}(q^2) = -8N\frac{m_B m_{K^*}}{\sqrt{q^2}}\left\{(C_9 \mp C_{10})A_{12}(q^2) + \frac{m_b}{m_B + m_{K^*}}C_7 T_{23}(q^2) + \mathcal{G}_0(q^2)\right\}, \tag{2.33}$$

[7] *CP* violating effects are negligible in the SM and are so far also not favoured as a NP effect.

$$\mathcal{A}_{\parallel}^{\mathrm{L,R}}(q^2) = -N\sqrt{2}(m_B^2 - m_{K^*}^2)\left\{(C_9 \mp C_{10})\frac{A_1(q^2)}{m_B - m_{K^*}} + \frac{2m_b}{q^2}C_7 T_2(q^2) + \mathcal{G}_{\parallel}(q^2)\right\}, \tag{2.34}$$

$$\mathcal{A}_{\perp}^{\mathrm{L,R}}(q^2) = N\sqrt{2\lambda}\left\{(C_9 \mp C_{10})\frac{V(q^2)}{m_B + m_{K^*}} + \frac{2m_b}{q^2}C_7 T_1(q^2) + \mathcal{G}_{\perp}(q^2)\right\}, \tag{2.35}$$

with

$$\lambda = m_B^4 + m_{K^*}^4 + q^4 - 2(m_B^2 m_{K^*}^2 + m_{K^*}^2 q^2 + m_B^2 q^2),\ \beta_\ell = \sqrt{1 - 4m_\ell^2/q^2} \tag{2.36}$$

and

$$N = V_{tb}V_{ts}^*\sqrt{\frac{G_F^2\alpha^2}{3 \times 2^{10}\pi^5 m_B^3}q^2\lambda^{1/2}\beta_\mu}\,. \tag{2.37}$$

The functions $\mathcal{G}_0(q^2)$, $\mathcal{G}_{\parallel}(q^2)$, and $\mathcal{G}_{\perp}(q^2)$ in equations 2.33, 2.34, and 2.35 describe the non-local contributions to the $B^0 \to K^{*0}\mu^+\mu^-$ decay which are discussed in section 2.11. An empirical model to parameterise $\mathcal{G}_0(q^2)$, $\mathcal{G}_{\parallel}(q^2)$, and $\mathcal{G}_{\perp}(q^2)$ is presented in chapter 3.

Since $C_9 \approx -C_{10}$, the right handed amplitudes are highly suppressed in the SM.

2.9 S-wave contribution

Apart from the P-wave configuration of the K^{*0}, the $B^0 \to K^{*0}\mu^+\mu^-$ decay can also occur via kaon resonances with angular momentum 0 (S-wave configuration) such as the ${K^{*0}}_0(800)$. The S-Wave decays and make up a significant contribution of the observed $B^0 \to K^{*0}\mu^+\mu^-$ decays [40]. Since the S-Wave decays have different angular distributions than the P-Wave decays, their presence would alter the angular observables if ignored in a fit to data. The S-wave contribution is therefore included in the differential decay rate by including 6 additional angular terms with the corresponding angular observables J'_{1c}, J'_{2c}, J'_4, J'_5, J'_7, J'_8, which are described by the S-wave amplitudes and the interference between the P-wave and S-wave amplitudes. The differential decay rate then becomes [41]:

$$
\begin{aligned}
\frac{\mathrm{d}^4\Gamma[B^0\to K^{*0}\mu^+\mu^-]}{\mathrm{d}q^2\mathrm{d}\vec{\Omega}} =& \frac{9}{32\pi}\sum_i J_i(q^2) f_i(\cos\theta_l,\cos\theta_K,\phi) \\
=& \frac{9}{32\pi}\Big[J_1^s \sin^2\theta_\mathrm{K} + J_1^c\cos^2\theta_\mathrm{K} + J'_{1c}(1-\cos^2\theta_\mathrm{l}) \\
&+ J_2^s\sin^2\theta_\mathrm{K}\cos 2\theta_\mathrm{l} + J_2^c\cos^2\theta_\mathrm{K}\cos 2\theta_\mathrm{l} + J'_{2c}\cos^2\theta_\mathrm{K}(1-\cos 2\theta_\mathrm{l}) \\
&+ J_3\sin^2\theta_\mathrm{K}\sin^2\theta_\mathrm{l}\cos 2\phi + J_4\sin 2\theta_\mathrm{K}\sin 2\theta_\mathrm{l}\cos\phi + J'_4\sin 2\theta_\mathrm{l}\sin\theta_\mathrm{K}\cos\phi \\
&+ J_5\sin 2\theta_\mathrm{K}\sin\theta_\mathrm{l}\cos\phi + J'_5\sin\theta_\mathrm{K}\sin\theta_\mathrm{l}\cos\phi + J_6^s\sin^2\theta_\mathrm{K}\cos\theta_\mathrm{l} \\
&+ J_7\sin 2\theta_\mathrm{K}\sin\theta_\mathrm{l}\sin\phi + J'_7\sin\theta_\mathrm{K}\sin\theta_\mathrm{l}\sin\phi + J_8\sin 2\theta_\mathrm{K}\sin 2\theta_\mathrm{l}\sin\phi \\
&+ J'_8\sin 2\theta_\mathrm{l}\sin\theta_\mathrm{K}\sin\phi + J_9\sin^2\theta_\mathrm{K}\sin^2\theta_\mathrm{l}\sin 2\phi\Big] \,.
\end{aligned}
\tag{2.38}
$$

The angular observables J'_{1c}, J'_{2c}, J'_4, J'_5, J'_7, J'_8 are functions of the S-wave amplitudes $\mathcal{A}_{00}^{\mathrm{L,R}}$ as well as the P-wave amplitudes $\mathcal{A}_0^{\mathrm{L,R}}$, $\mathcal{A}_\parallel^{\mathrm{L,R}}$, $\mathcal{A}_\perp^{\mathrm{L,R}}$ (see section 2.8) and are given by

$$
\begin{aligned}
J'_{1c} &= \frac{1}{3}|\mathcal{A}_{00}^{\mathrm{L}}|^2 + \frac{1}{3}|\mathcal{A}_{00}^{\mathrm{R}}|^2 \,, \\
J'_{2c} &= \frac{2}{\sqrt{3}}\left[\mathrm{Re}(\mathcal{A}_{00}^{\mathrm{L}}\mathcal{A}_0^{\mathrm{L}*}) + (L\to R)\right] \,, \\
J'_4 &= \sqrt{\frac{2}{3}}\left[\mathrm{Re}(\mathcal{A}_{00}^{\mathrm{L}}\mathcal{A}_\parallel^{\mathrm{L}*}) + (L\to R)\right] \,, \\
J'_5 &= 2\sqrt{\frac{2}{3}}\left[\mathrm{Re}(\mathcal{A}_{00}^{\mathrm{L}}\mathcal{A}_\perp^{\mathrm{L}*}) - (L\to R)\right] \,, \\
J'_7 &= 2\sqrt{\frac{2}{3}}\left[\mathrm{Re}(\mathcal{A}_{00}^{\mathrm{L}}\mathcal{A}_\parallel^{\mathrm{L}*}) - (L\to R)\right] \,, \\
J'_8 &= \sqrt{\frac{2}{3}}\left[\mathrm{Re}(\mathcal{A}_{00}^{\mathrm{L}}\mathcal{A}_\perp^{\mathrm{L}*}) + (L\to R)\right] \,,
\end{aligned}
\tag{2.39}
$$

The S-wave amplitudes $\mathcal{A}_{00}^{\mathrm{L,R}}$ are given by

$$
\mathcal{A}_{00}^{\mathrm{L,R}}(q^2) = -N\frac{\lambda_{K_0^{*0}}}{\sqrt{q^2}}\left\{(C_9 \mp C_{10})F_1(q^2) + \frac{2m_b}{m_B+m_{K^*}}C_7 F_T(q^2) + \mathcal{G}_{00}(q^2)\right\}, \tag{2.40}
$$

with

$$
\lambda_{K_0^{*0}} = m_B^4 + m_{K_0^*}^4 + q^4 - 2(m_B^2 m_{K_0^*}^2 + m_{K_0^*}^2 q^2 + m_B^2 q^2),\ \beta_\ell = \sqrt{1-4m_\ell^2/q^2} \tag{2.41}
$$

	$F(0)$	α_F	b_F
$F_1(q^2)^{K_0^*(800)}$	0.27	-2.1	1.2
$F_T(q^2)^{K_0^*(800)}$	0.3	-2.2	1.2

Table 2.1: Coefficients for the S-wave FFs taken from Ref. [43].

and

$$N = V_{tb}V_{ts}^* \sqrt{\frac{G_F^2\alpha^2}{3 \times 2^{10}\pi^5 m_B^3} q^2 \sqrt{\lambda_{K_0^{*0}}}\beta_\mu}\,. \tag{2.42}$$

The functions $F_1(q^2)$ and $F_T(q^2)$ in equation 2.40 are the FFs associated to the S-wave $B^0 \to K^{*0}\mu^+\mu^-$ decay. These FFs have not been studied extensively yet. There is a calculation using a perturbative QCD approach which expresses the FFs with the following parameterisation [42, 43]

$$F_{l,T}(q^2) = \frac{F(0)}{1 + \alpha_F q^2/m_B^2 + b_F(q^2/m_B^2)^2}, \tag{2.43}$$

where $F(0)$, α_F, b_F are the individual FF coefficients. The FF coefficients for the FFs of the $B^0 \to K^{*0}\mu^+\mu^-$ decay in the $K^{*0}{}_0(800)$ configuration calculated in Ref. [43] are given in table 2.1. The paper does not give any uncertainties or correlations of these coefficients. This poses an issue for performing a fit to data while taking the S-wave contribution into account. The way this is handled in this thesis is described in section 3.3 and the systematic uncertainty due to the poor understanding of the S-wave FFs is discussed in section 7.2.3.

2.10 Anomalies of $b \to s\ell^+\ell^-$ decays

There are several measurements of $b \to s\ell^+\ell^-$ processes that show tensions with the SM. They can be grouped into three categories: Branching fraction measurements, angular analyses, and lepton flavor universality (LFU) tests.

LHCb has measured the branching fractions of several $b \to s\ell^+\ell^-$ decays to be lower than the SM predictions: $B^0 \to K^{*0}\mu^+\mu^-$ [5, 8], $B_s^0 \to \phi\mu^+\mu^-$ [6], and $\Lambda_b^0 \to \Lambda\mu^+\mu^-$ [7].

In angular analyses, fits to the angular distributions, such as the distribution of the angles defined in section 2.6, are performed to obtain a set of angular observables (such as the observables given in eq. 2.32). These fits allow the extraction of the full information

encoded in the decays which make them ideal for finding NP. Measurements of angular observables that show discrepancies with the SM, performed by LHCb include [1, 6, 9–12].

Both the branching fraction measurements and the measurements of angular observables suffer from a common disadvantage: The theoretical predictions depend on the understanding of the hadronic interaction of the decay. Even for optimized angular observables (equation 2.32), in which the FF uncertainties cancel to first order, the theoretical uncertainties are significantly large when comparing experimental results with SM predictions. This dependence on the theoretical understanding of the hadronic interactions is removed in LFU tests, which are measurements of ratios of branching fractions such as $\mathrm{BF}(B^0 \to K^{*0}\mu^+\mu^-)/\mathrm{BF}(B^0 \to K^{*0}e^+e^-)$. Since the SM couplings are lepton flavour universal, the ratios are easily predicted with high precision within the SM and any effects of the hadronic interactions cancel out. Measured LFU ratios that are discrepant from the SM include $B^0 \to K^{*0}\ell^+\ell^-$ decays [13], and $B^+ \to K^+\ell^+\ell^-$ decays [14, 15].

Despite the LFU ratios being the theoretically 'cleanest' observables, one cannot solely rely on them for searching for NP in $b \to s\ell^+\ell^-$ decays. Firstly, the most significant discrepancy with the SM is currently observed in the angular observables of $B^0 \to K^{*0}\mu^+\mu^-$ decays [1]. Secondly, LFU ratios will only reveal NP if the NP does in fact violate LFU, which may not be the case.

The various $b \to s\ell^+\ell^-$ anomalies can be interpreted in a coherent way by performing global fits determining the values of the Wilson Coefficients with which the predictions best describe the available measurements of $b \to s\ell^+\ell^-$ processes [16, 44]. Different hypotheses regarding the nature of the NP are tested such as 1D hypotheses allowing $\mathcal{C}_9^{\mathrm{NP}}$ only, or $\mathcal{C}_9^{\mathrm{NP}} = -\mathcal{C}_{10}^{\mathrm{NP}}$ or 2D hypotheses such as allowing ($\mathcal{C}_9^{\mathrm{NP}}$, $\mathcal{C}_{10}^{\mathrm{NP}}$). The left plot in figure 2.5 shows the allowed regions for the ($\mathcal{C}_9^{\mathrm{NP}}$, $\mathcal{C}_{10}^{\mathrm{NP}}$) hypothesis from fits to ATLAS (green), Belle (violet), CMS (yellow) and LHCb (red) as well as the combined (blue) data. The best fit values are (-0.91, 0.18) with a pull of 5.6σ with respect to the SM [15]. This indicates a significant shift in the vector coupling $\mathcal{C}_9$. However, there is an ongoing debate in the community about the presence of so called non-local effects which might mimic NP by shifting the observed value of $\mathcal{C}_9$, as discussed in the following section.

To further illustrate the importance of branching fraction measurements and measurements of angular observables, the right side of figure 2.5 shows the fit to the LFU ratios only, which yields a much less clear picture.

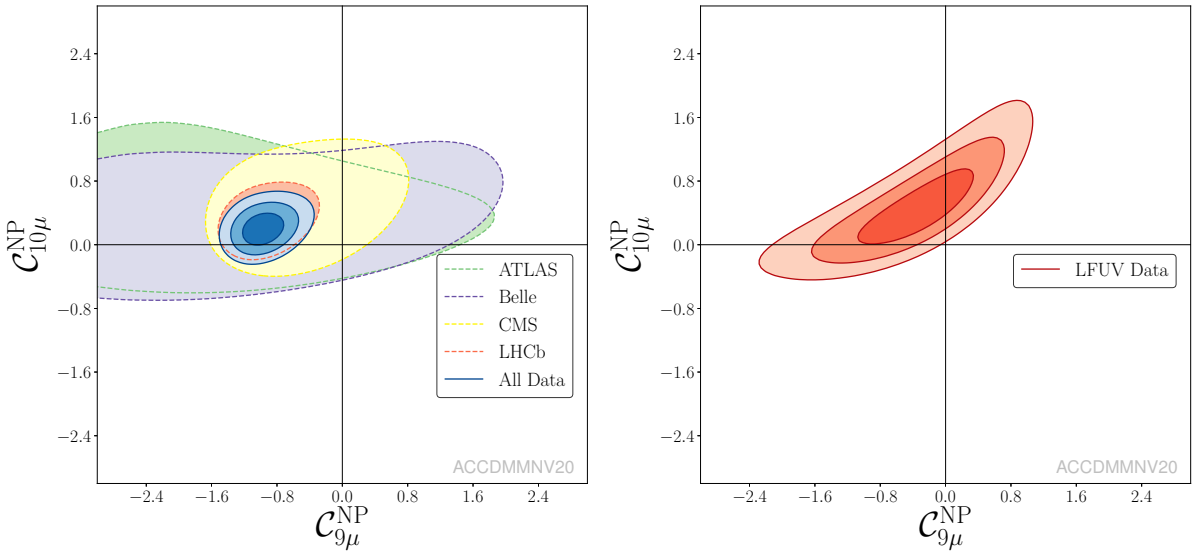


Figure 2.5: Contours of global fits under the $(\mathcal{C}_9^{\rm NP}, \mathcal{C}_{10}^{\rm NP})$ hypothesis using all available $b \to s\ell^+\ell^-$ observables (left) and LFU ratios only (right). Figure from Ref. [16].

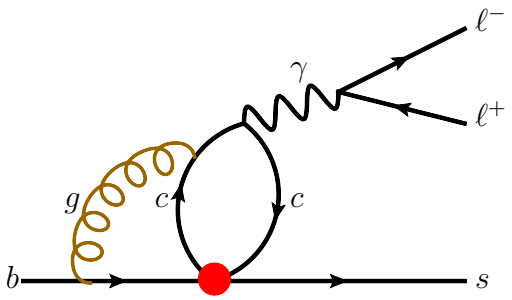


Figure 2.6: Feynman diagram for a $b \to sc\bar{c}$ process in the EFT. Figure from Ref. [45].

2.11 Effect of non-local contributions

As described in section 2.3, $b \to s\ell^+\ell^-$ transitions occur only at the loop level in the SM. The same final state can be obtained through $b \to sq\bar{q}$ processes if the $q\bar{q}$ pair decays into a lepton pair, as shown in figure 2.6. The $q\bar{q}$ pair is a bound state which can make up vector meson resonances such as $\rho(770)$, $\phi(1020)$, J/ψ, $\psi(2S)$, $\psi(3770)$, $\psi(4040)$ and $\psi(4160)$. Since decays such as $B^0 \to J/\psi K^{*0}$ can also occur on the tree level, their amplitudes dominate the total $B^0 \to K^{*0}\mu^+\mu^-$ decay rate at $q^2 = m_{J/\psi}^2$ (or the pole mass of the respective resonance). Figure 2.7 shows the total decay rate of $B^0 \to K^{*0}\mu^+\mu^-$ including

all $b \to sq\overline{q}$ transitions and the peaks of the resonances are clearly visible.

Since the $q\overline{q}$ can exchange gluons with the b quark, as shown in figure 2.6, the factorisation assumption (i.e. describing the quark level $b \to s\ell^+\ell^-$ transition and the formation of the K^{*0} meson separately) does not hold for $b \to sq\overline{q}$ processes. Consequently, the decay amplitudes written in terms of the Wilson Coefficients and the FFs (see equations 2.33, 2.34, 2.35) do not properly describe the q^2 dependence of the processes involving $q\overline{q}$ loops. Since the resonances are vector particles, their presence effectively changes the value of $\mathcal{C}_9$ in a q^2 dependent way. This is expressed by the functions $\mathcal{G}_0(q^2)$, $\mathcal{G}_\parallel(q^2)$, and $\mathcal{G}_\perp(q^2)$ in equations 2.33, 2.34, 2.35. The change of $\mathcal{C}_9$ becomes largest at the pole masses of the resonances.
In analyses that aim to be sensitive to NP by measuring observables related to the penguin amplitudes, the resonant q^2 regions are discarded from analysis (see for example Ref. [1]). However, the $b \to sq\overline{q}$ processes can have effects on the $B^0 \to K^{*0}\mu^+\mu^-$ decay rate, far away from the pole masses of the resonances, through interference of the $b \to sq\overline{q}$ amplitudes with the $b \to s\ell^+\ell^-$ penguin amplitudes. Therefore, the effects of the hadronic resonances are referred to as *non-local contributions*. Since the phases of the resonant amplitudes relative to the penguin amplitudes cannot be calculated from first principles and have never been measured before, the level of interference, i.e. the size of the non-local contributions is unknown.

There is an ongoing debate among theorists how to take non-local contributions into account when calculating $B^0 \to K^{*0}\mu^+\mu^-$ decays. One approach, which claims validity only in the heavy quark and large energy limit (corresponding to $q^2 < 6\,\mathrm{GeV}^2/c^4$), uses QCD factorisation methods [32, 34, 46]. Moreover, there is a model to determine the level of charm loop interference through analyticity [47], but it is also only valid away from the pole masses of the resonances. The effect of the higher ψ resonances is studied in Ref. [48].

In chapter 3, an empirical model is presented which aims to describe the full q^2 spectrum by parameterising the q^2 dependence of the resonances with relativistic Breit Wigner functions and assigning a magnitude and phase for each resonance relative to the penguin amplitude. This allows determining the Wilson Coefficients directly from data while also measuring the phases of the resonant amplitudes relative to the penguin amplitude. Thus, the effect of the non-local contributions can be assessed experimentally.

LHCb has published a similar measurement of the phase difference of the resonant and penguin amplitudes in the $B^+ \to K^+\mu^+\mu^-$ decay, where the level of interference was found to be small [49]. However, the conclusions cannot be transferred to the $B^0 \to K^{*0}\mu^+\mu^-$

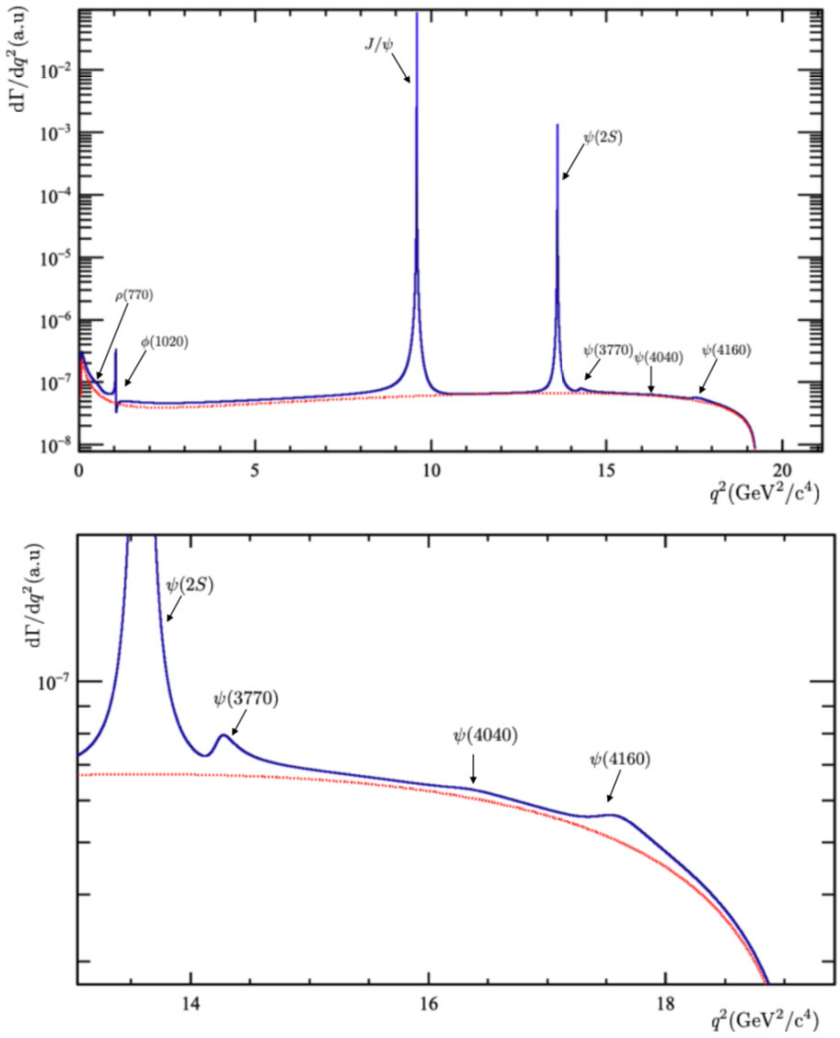


Figure 2.7: Decay rate of $B^0 \to K^{*0}\mu^+\mu^-$ as a function of q^2 including all relevant $b \to sq\overline{q}$ transitions (blue) and using the penguin amplitudes only (red). Based on the model described in chapter 3. Figure from Ref. [50].

decay, since its amplitude structure is very different. The $B^+ \to K^+\mu^+\mu^-$ decay is described by two amplitudes rather than the six amplitudes of $B^0 \to K^{*0}\mu^+\mu^-$ (see section 2.8) and the final state of $B^+ \to K^+\mu^+\mu^-$ is described by a single angle and q^2, whereas $B^0 \to K^{*0}\mu^+\mu^-$ requires three angles and q^2 (see section 2.6).

Measuring the non-local contributions directly from data will help our understanding of whether the observed anomalies in the angular observables of $B^0 \to K^{*0}\mu^+\mu^-$ decays are the result of underestimated interference of $c\bar{c}$-loops with the penguin.

3. Empirical model for searching for New Physics in $B^0 \to K^{*0}\mu^+\mu^-$

As explained in the previous chapter, measurements of angular observables in $B^0 \to K^{*0}\mu^+\mu^-$ decays have shown discrepancies to SM predictions, indicating a shift of the vector Wilson Coefficient $\mathcal{C}_9$ with respect to the SM. However, this cannot clearly be interpreted as an indication for NP because interference from hadronic resonances with the short distance amplitudes can also cause a shift of $\mathcal{C}_9$. Therefore, these non-local effects can mimic NP if they are not taken into account correctly in the SM predictions.
The purpose of this thesis is to perform a novel fit to the full q^2 spectrum in the range $(0.1 \leq q^2 \leq 18.0)$ GeV$^2/c^4$, parameterising both the penguin amplitudes as well as the hadronic resonance amplitudes and the interference between them.

The model containing both the penguin and hadronic amplitudes is presented in section 3.1. Its prediction of the non-local effects is compared to predictions from other models in section 3.2. The inclusion of the S-wave component in the model, which is necessary to describe real data, is described in section 3.3.

3.1 Modelling non-local contributions

The model which is used in this thesis to parameterise the full $B^0 \to K^{*0}\mu^+\mu^-$ spectrum is based on Ref. [2][1].
The signal model is based on the full differential decay rate (see equation 2.38), where the angular observables J_i and J_i' are expressed in terms of the decay amplitudes (see equations 2.28 and 2.39). The effect of the non-local contributions is introduced on the amplitude level. Recalling from section 2.8, the P-wave amplitudes can be written as

$$\mathcal{A}_0^{\mathrm{L,R}}(q^2) = -8N\frac{m_B m_{K^*}}{\sqrt{q^2}}\left\{(C_9 \mp C_{10})A_{12}(q^2) + \frac{m_b}{m_B + m_{K^*}}C_7 T_{23}(q^2) + \mathcal{G}_0(q^2)\right\}, \quad (3.1)$$

$$\mathcal{A}_\parallel^{\mathrm{L,R}}(q^2) = -N\sqrt{2}(m_B^2 - m_{K^*}^2)\left\{(C_9 \mp C_{10})\frac{A_1(q^2)}{m_B - m_{K^*}} + \frac{2m_b}{q^2}C_7 T_2(q^2) + \mathcal{G}_\parallel(q^2)\right\}, \quad (3.2)$$

[1] This entire section is written in reference to Ref. [2].

$$\mathcal{A}_{\perp}^{\mathrm{L,R}}(q^2) = N\sqrt{2\lambda}\left\{(C_9 \mp C_{10})\frac{V(q^2)}{m_B + m_{K^*}} + \frac{2m_b}{q^2}C_7 T_1(q^2) + \mathcal{G}_{\perp}(q^2)\right\}, \tag{3.3}$$

with

$$\lambda = m_B^4 + m_{K^*}^4 + q^4 - 2(m_B^2 m_{K^*}^2 + m_{K^*}^2 q^2 + m_B^2 q^2),\ \beta_\ell = \sqrt{1 - 4m_\ell^2/q^2} \tag{3.4}$$

and

$$N = V_{tb}V_{ts}^*\sqrt{\frac{G_F^2\alpha^2}{3\times 2^{10}\pi^5 m_B^3}q^2\lambda^{1/2}\beta_\mu}\,, \tag{3.5}$$

where the functions in $\mathcal{G}_0(q^2)$, $\mathcal{G}_\parallel(q^2)$, and $\mathcal{G}_\perp(q^2)$ describe the non-local hadronic contributions.

In this empirical model, inspired by Refs. [48, 49, 51], the hadronic contributions are parameterised with

$$\mathcal{G}_0(q^2) = \frac{m_b}{m_B + m_{K^*}}T_{23}(q^2)\zeta^0 e^{i\omega^0} + A_{12}(q^2)\sum_j \eta_j^0 e^{i\theta_j^0}A_j^{\mathrm{res}}(q^2), \tag{3.6}$$

$$\mathcal{G}_\parallel(q^2) = \frac{2m_b}{q^2}T_2(q^2)\zeta^\parallel e^{i\omega^\parallel} + \frac{A_1(q^2)}{m_B - m_{K^*}}\sum_j \eta_j^\parallel e^{i\theta_j^\parallel}A_j^{\mathrm{res}}(q^2), \tag{3.7}$$

$$\mathcal{G}_\perp(q^2) = \frac{2m_b}{q^2}T_1(q^2)\zeta^\perp e^{i\omega^\perp} + \frac{V(q^2)}{m_B + m_{K^*}}\sum_j \eta_j^\perp e^{i\theta_j^\perp}A_j^{\mathrm{res}}(q^2). \tag{3.8}$$

The first terms are the non-local contributions to the $\mathcal{C}_7$ terms. The respective contribution to each transversity amplitude has a magnitude ζ^λ and phase ω^λ which can be determined in a fit to data. There is a degeneracy of ζ^λ with the first order parameter of each respective form factor, i.e. with $\alpha_0^{T_{23}}$, $\alpha_0^{T_2}$, $\alpha_0^{T_1}$. Therefore, the first order FF parameters are fixed to the values determined in Ref. [35], whereas the 2nd and 3rd order parameters are floated and constrained to the covariance matrix (also published in Ref. [35]).
The second terms in equations 3.6, 3.7, and 3.8 are the non-local contributions to the $\mathcal{C}_9$ terms. The sums indicate sums over the considered vector resonances $j \in\{\rho^0,\ \phi(1020),\ J/\psi\,,\ \psi(2S),\ \psi(3770),\ \psi(4040),\ \psi(4160)\}$. Each resonance amplitude is multiplied with a complex number where η_j^0, $\eta_j^\parallel$, $\eta_j^\perp$ are the magnitudes and are θ_j^0, $\theta_j^\perp$, $\theta_j^\parallel$ the phases of each resonant amplitude relative to C_9. These phases and magnitudes can also be determined in a fit to data.

The q^2 dependence of each resonance is modelled by a relativistic Breit-Wigner $A_j^{\rm res}(q^2)$, similarly to the model used in Ref. [49], given by

$$A_j^{\rm res}(q^2) = \frac{m_{{\rm res}\,j}\Gamma_{{\rm res}\,j}}{(m_{{\rm res}\,j}^2 - q^2) - im_{{\rm res}\,j}\Gamma_j(q^2)}, \tag{3.9}$$

where $m_{{\rm res}\,j}$ is the pole mass and $\Gamma_{{\rm res}\,j}$ the natural width of resonance j. The respective values are taken from [30]. The function $\Gamma_j(q^2)$ represents the running width which is given by

$$\Gamma_j(q^2) = \frac{p}{p_{{\rm res}\,j}}\frac{m_{{\rm res}\,j}}{q}\Gamma_{{\rm res}\,j}, \tag{3.10}$$

where p is the momentum of the muons in the dimuon rest frame evaluated at q, and $p_{{\rm res}\,j}$ is the momentum evaluated at the mass of the resonance.

Since there are three magnitudes and phases for each resonance, it is common to define the phases $\theta_j^{\perp}$, $\theta_j^{\parallel}$ to be relative to θ_j^0. Thus, θ_j^0 is defined to be the overall phase between the resonance j and the penguin amplitudes.

Existing measurements of $B^0 \to VK^{*0}$ decays, where V denotes any vector meson resonance appearing in the $B^0 \to K^{*0}\mu^+\mu^-$ spectrum, are sensitive to the relative phases $\theta_j^{\perp}$, $\theta_j^{\parallel}$ between the resonant amplitudes. However, the phases θ_j^0 between the resonant and the penguin amplitudes, and thus the level of interference between the resonances and the penguin, are completely unknown. This is because there has never been an analysis of $B^0 \to K^{*0}\mu^+\mu^-$ data which parameterised both the penguin and resonant amplitudes simultaneously, which is exactly what the analysis presented in this thesis is aiming to accomplish by fitting LHCb data with the model explained above.

3.2 Comparison to other models

In order to make predictions with the model described in the previous section, available measurements of the relative phases $\theta_j^{\perp}$, $\theta_j^{\parallel}$ and magnitudes η_j^0, $\eta_j^{\perp}$, $\eta_j^{\parallel}$ can be used[2].

For the decays $B^0 \to J/\psi K^{*0}$ and $B^0 \to \psi(2S)K^{*0}$, measurements of the relative phases and magnitudes of the transversity amplitudes are available from the LHCb, Babar and Belle collaborations [52–54]. Also, the branching fractions have been measured by the Belle collaboration [53, 55]. Furthermore, the relative phases and magnitudes and branching fractions of the $B^0 \to \phi K^{*0}$ have been measured [56–58]. Using these inputs, the magnitudes η_j^0, $\eta_j^{\parallel}$, $\eta_j^{\perp}$ of each resonant amplitude, normalised to the respective FFs,

[2]This section is written in reference to Ref. [2].

are calculated with

$$|\eta_j^0|^2 = \frac{f_j^0 \mathcal{B}(B^0 \to VK^{*0}) \times \mathcal{B}(V \to \mu^+\mu^-)}{\tau_B \int \left| 8N \frac{m_B m_{K^*}}{\sqrt{q^2}} A_j^{\rm res}(q^2) A_{12}(q^2) \right|^2 dq^2},$$

$$|\eta_j^{\parallel}|^2 = \frac{f_j^{\parallel} \mathcal{B}(B^0 \to VK^{*0}) \times \mathcal{B}(V \to \mu^+\mu^-)}{\tau_B \int \left| N\sqrt{2}(m_B^2 - m_{K^*}^2) A_j^{\rm res}(q^2) \frac{A_1(q^2)}{m_B - m_{K^*}} \right|^2 dq^2}, \tag{3.11}$$

$$|\eta_j^{\perp}|^2 = \frac{f_j^{\perp} \mathcal{B}(B^0 \to VK^{*0}) \times \mathcal{B}(V \to \mu^+\mu^-)}{\tau_B \int \left| N\sqrt{2\lambda} A_j^{\rm res}(q^2) \frac{V(q^2)}{m_B + m_{K^*}} \right|^2 dq^2},$$

where f_j^0, $f_j^{\parallel}$, and $f_j^{\perp}$ are the measured polarisation fractions of the $B^0 \to VK^{*0}$ decays and τ_B denotes the lifetime of the B^0 meson. Calculating the magnitudes with the above expressions ensures that the integral of the sum of squared magnitudes of a given resonance yields the correct experimental branching fraction.

For the $B^0 \to \rho^0 K^{*0}$ decay, the total decay amplitude is set using the world average branching fraction and the relative phases and magnitudes are set to those of the $B^0 \to \phi K^{*0}$ decay. For the decays $B^0 \to V_\psi K^{*0}$, with $V_\psi \in \{\psi(3770), \psi(4040), \text{and}\, \psi(4160)\}$, there are no available measurements. Therefore, for the purpose of making predictions to compare to other models, the relative phases and magnitudes of the decays are assumed to be the same as $B^0 \to J/\psi K^{*0}$ and the branching fractions are approximated by scaling the known branching fractions of $B^0 \to \psi(2S)K^{*0}$ with $\psi(2S) \to \mu^+\mu^-$ by the ratio of the known branching fractions of $B^+ \to \psi(2S)K^+$ and $B^+ \to V_\psi K^+$ with $V_\psi \to \mu^+\mu^-$.
As discussed above, the phase θ_j^0 relative to the penguin is unknown for each resonance. Therefore, these phases are varied and the effect of different values are studied when comparing to different models.

The first model to which the empirical model, described in section 3.1, is compared to is the prediction of non-local charm loop contributions presented in Ref. [59]. It uses light cone sum rules calculations of the $B^0 \to K^{*0}$ matrix elements including contributions from $c\bar{c}$ loops for $q^2 \ll 4\,m_c^4$ and a hadronic dispersion relation to extrapolate to higher q^2. The extrapolation uses experimental input for the amplitude structure of the $B^0 \to J/\psi\, K^{*0}$ and $B^0 \to \psi(2S)K^{*0}$ decays. The effect of the charm loops are represented as $\Delta\mathcal{C}_9(q^2)$,

a q^2 dependent correction to the Wilson Coefficient $\mathcal{C}_9$. The model in Ref. [59] uses the invariant amplitude base and the authors also provide the relation of this basis to the transversity basis which is used in Ref. [2] and in this thesis.
Figure 3.1 shows the real part of $\Delta\mathcal{C}_9(q^2)$ predicted in Ref. [59] (magenta band) as well as the predictions made with the model described in section 3.1, considering only the hadronic contributions from $B^0 \to J/\psi K^{*0}$ and $B^0 \to \psi(2S)K^{*0}$ decays, with different values for the phases. The black dashed line shows the prediction obtained when setting all phases to zero, thereby ignoring the measured relative phases between the hadronic amplitudes. The cyan line shows the prediction obtained when setting the relative phases between the hadronic amplitudes to the measured values but setting the overall phases between the resonances and the penguin amplitudes $\theta^0_{J/\psi,\psi(2S)}$ to zero. The dashed cyan line shows the prediction obtained with the relative phases set to the measured values and setting $\theta^0_{J/\psi,\psi(2S)} = \pi$. When ignoring the relative phases of the hadronic amplitudes, the prediction of $\Delta\mathcal{C}_9(q^2)$ made with the model described in section 3.1 is consistent with the prediction from Ref. [59]. However, when taking the relative phases of the hadronic amplitudes into account, the predictions are different. The size of the difference depends on the choice of the free phases $\theta^0_{J/\psi,\psi(2S)}$.

The second model to which the empirical model, described in section 3.1, is compared is the prediction of the non-local charm loop contributions presented in Ref. [47]. The non-local contributions are calculated at $q^2 < 0$ to next-to-leading order in α_s and then parameterised as a function of q^2 using a z-expansion, truncated after the second order, similar to equation 2.17. Figure 3.2 shows the real and imaginary parts of the q^2 dependent shifts of $\mathcal{C}_9$ in the transversity basis predicted by [47] (magenta points). Also shown are the corresponding predictions obtained with the model described in section 3.1, when considering hadronic contributions only from the J/ψ and $\psi(2S)$, with three different choices for the phases $\theta^0_{J/\psi,\psi(2S)}$ (cyan lines). The empirical model shows good agreement with the predictions from Ref. [47] when setting $\theta^0_{J/\psi,\psi(2S)} = \pi/8$. The differences of the models in the imaginary parts of $\Delta\mathcal{C}_9(q^2)$ is due to the choice for the $\omega^\lambda = \pi$ which is discussed in more detail in Ref. [2]. Smaller values for ω^λ have been found to give a better agreement with Ref. [47] but bigger values for ω^λ give better agreement with Ref. [59]. [3]

[3]It should be noted that the same model (described in this section) is used for both figure 3.1 and 3.2 with the respective phase configurations explained in the figures and in the text. The difference in line-shapes of $\Delta\mathcal{C}_9(q^2)$ between the figures is due to the fact that completely different amplitude bases are used. No trivial direct comparison can be made between them.

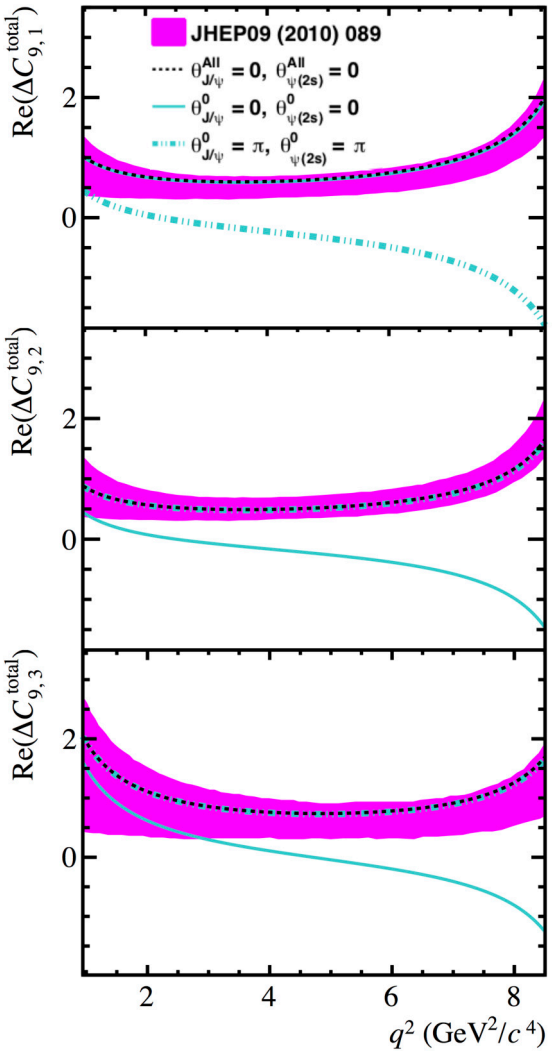


Figure 3.1: q^2 dependent shift of $\mathcal{C}_9$ due to the effect of the non-local contributions to the $B^0 \to K^{*0}\mu^+\mu^-$ invariant amplitudes. The prediction from Ref. [59] is shown with the magenta band. The prediction using the model described in section 3.1 taking into account only the hadronic contributions from J/ψ and $\psi(2S)$ is also shown for different choices for the phases. The phases $\theta^0_{J/\psi,\psi(2S)}$ are set to 0 (solid cyan line) and π (dashed cyan line). Also shown is the prediction where all phases including the relative phases between the resonant amplitudes are set to 0 (dashed black line). Figure from Ref. [2].

In summary, the empirical model, presented in section 3.1 and in Ref. [2], shows good agreement with existing models which predict the non-local contributions to $B^0 \to K^{*0}\mu^+\mu^-$ decays, depending on the choice of the values of the phases. Contrary to previous models, the empirical model can naturally include contributions from any vector resonance

appearing in the q^2 spectrum of $B^0 \to K^{*0}\mu^+\mu^-$ decays. These include the light quark resonances such as the $\phi(1020)$ and ρ^0, as well as the heavy resonances appearing at $q^2 > 4m_D^2$, where m_D is the mass of the lowest mass D meson, such as the $\psi(3770)$, $\psi(4040)$, and $\psi(4160)$.

However, the caveat of the empirical model is that it cannot be generalised easily as it relies on relativistic BW functions and therefore does not conserve unitarity. Furthermore, it does not necessarily predict the non-local contributions to vanish at the kinematic endpoint. Nonetheless, these considerations do not pose an issue for fitting experimental data, as the phase space of real events vanishes at the kinematic end point, so there is no data that needs to be described. Furthermore, in a fit to experimental data the model does not need to be able to describe negative q^2. The BW model can describe the data in the considered kinematic region well. Thus, it can be concluded that while the model is not a *true* model in the strictest theoretical sense, it is a reasonable approximation of a true model which does agree well in its predictions with more theoretically sound models (shown above), and can be used to describe and learn from experimental data.

3.3 Inclusion of the S-wave component in the model

In sections 3.1 and 3.2 only the P-wave contributions to the $B^0 \to K^{*0}\mu^+\mu^-$ decay were considered. In order to describe data, also the S-wave contributions have to be included in the model. As discussed in section 2.9, the form factors $F_1(q^2)$ and $F_T(q^2)$ associated to the S-wave amplitudes are not very well understood. To avoid decreasing the sensitivity to the Wilson Coefficients, the S-wave amplitudes are decoupled from the rest of the model by replacing the Wilson coefficients in the S-wave amplitudes with

$$\begin{aligned} (\mathcal{C}_9 - \mathcal{C}_{10}) &\to \mathcal{C}_{V1}^S \\ (\mathcal{C}_9 + \mathcal{C}_{10}) &\to \mathcal{C}_{V2}^S \\ \mathcal{C}_7 &\to \mathcal{C}_T^{\mathrm{S}} \end{aligned} \tag{3.12}$$

This way the overall normalisation of the S-wave form factors is effectively free in the fit and the actual Wilson coefficients are solely determined in the P-wave amplitudes. The S-wave amplitudes then become

$$\mathcal{A}_{00}^{\mathrm{L,R}}(q^2) = -N\frac{\lambda_{K_0^{*0}}}{\sqrt{q^2}}\left\{C_{V1,V2}^S F_1(q^2) + \frac{2m_b}{m_B + m_{K^*}} C_T^S F_T(q^2) + \mathcal{G}_{00}(q^2)\right\}. \tag{3.13}$$

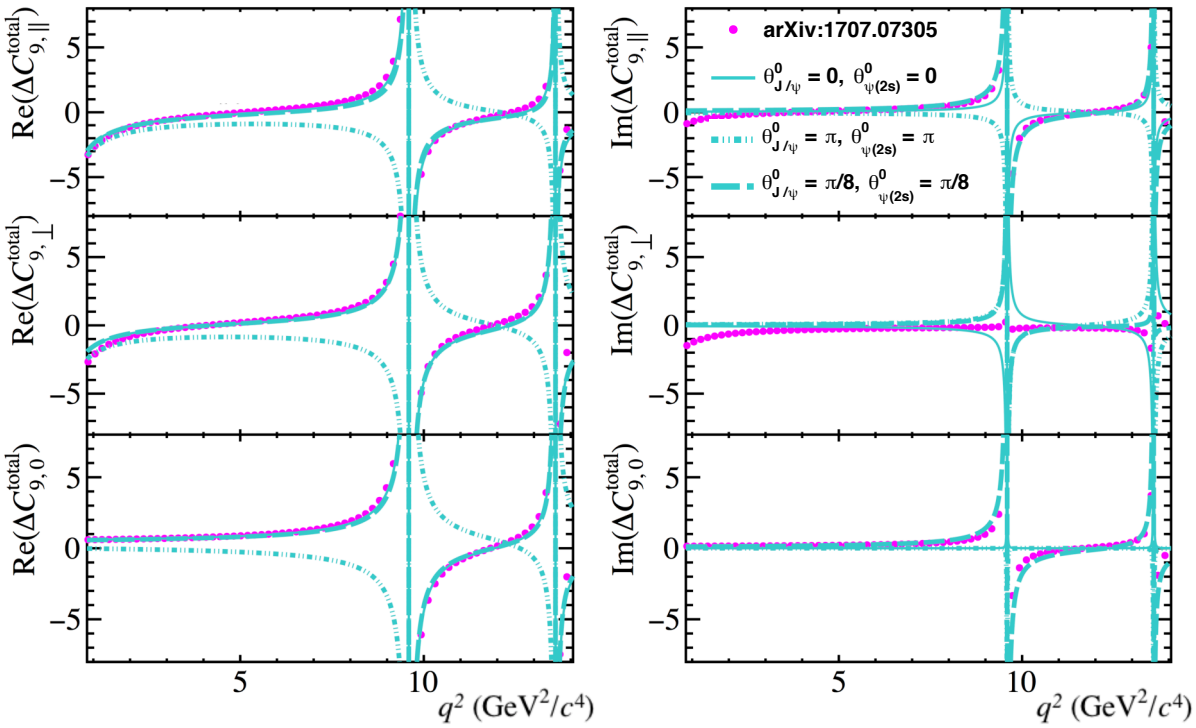


Figure 3.2: Non-local contributions to the $B^0 \to K^{*0}\mu^+\mu^-$ amplitudes in the transversity basis as a function of q^2 predicted in Ref. [47] (magenta points) and predicted with the model described in section 3.1 when considering hadronic contributions only for the J/ψ and $\psi(2S)$, for different values of the phases $\theta^0_{J/\psi\,,\psi(2S)}$ (cyan bands). Figure from Ref. [2].

Non-local contributions are also included for the S-wave amplitudes, similarly to how they are included for the P-wave amplitudes:

$$\mathcal{G}_{00}(q^2) = F_1(q^2) \sum_{j=J/\psi\,,\psi(2S)} \eta^0_j e^{i\theta^0_j} A^{\mathrm{res}}_j(q^2). \tag{3.14}$$

As indicated in the sum, non-local S-wave contributions are only included for the J/ψ and the $\psi(2S)$, because they are the dominant resonances. Due to the much smaller branching fractions of the $B^0 \to VK^{*0}$ decays involving the other resonances, compounded with the small fraction of S-wave decays overall (roughly 8%), contributions from the other resonances can safely be neglected. Furthermore, any non-local contribution to the $\mathcal{C}_7$ terms in the S-wave amplitudes would be too small for a fit to be sensitive to them. Thus they are not included explicitly in $\mathcal{G}_{00}(q^2)$. However, due to the reparameterisation $C_7 \to C_T^{\mathrm{S}}$, the factor C_T^{S} can effectively capture the real part of non-local S-wave contributions to $\mathcal{C}_7$.

When considering both the P-wave and S-wave contributions, the mass of the $K^+\pi^-$

system $m_{K\pi}$, is needed to fully describe $B^0 \to K^{*0}\mu^+\mu^-$ decays. The $m_{K\pi}$ distribution helps to distinguish between the P-wave and S-wave contributions since the P-wave amplitude has much more narrow $m_{K\pi}$ distribution. In the model used in this thesis, the $m_{K\pi}$ dependence is integrated out in order to reduce the number of dimensions in the fit from five ($\cos\theta_\ell$, $\cos\theta_K$, ϕ, q^2, $m_{K\pi}$) to four. To this end, the angular observables in the differential decay rate (eq. 2.38) which contain the P-wave amplitudes (J_{1s}, J_{1c}, J_{2s}, J_{2c}, J_3, J_4, J_5, J_{6s}, J_{6c}, J_{6c}, J_7, J_8, J_9) are multiplied with the integral of the P-wave $m_{K\pi}$ line shape, which is calculated with

$$\mathcal{I}_P^{m_{K\pi}} = \int_{796MeV/c^2}^{996MeV/c^2} |g(m_{K\pi})|^2 dm_{K\pi}, \tag{3.15}$$

where $g(m_{K\pi})$ is a relativistic Breit-Wigner function with the mass and width of the $K^{*0}(892)$. The integral is performed in the 100 MeV/c^2 window around the $K^{*0}(892)$ pole mass.

Similarly, the angular observable made up of the S-wave amplitudes (J'_{1c}) is multiplied with the integral of the S-wave $m_{K\pi}$ line shape, calculated with

$$\mathcal{I}_S^{m_{K\pi}} = \int_{796MeV/c^2}^{996MeV/c^2} |f(m_{K\pi})|^2 dm_{K\pi}, \tag{3.16}$$

where the function $f(m_{K\pi})$ is given by the LASS model [60].

The interference terms (J'_{2c}, J'_4, J'_5, J'_7, J'_8) which are made up of P-wave and S-wave amplitudes get multiplied with $\mathcal{I}_{P-S}^{m_{K\pi}}$, given by

$$\mathcal{I}_{P-S}^{m_{K\pi}} = \int_{796MeV/c^2}^{996MeV/c^2} f(m_{K\pi}) g(m_{K\pi})^* dm_{K\pi}. \tag{3.17}$$

3.4 Summary of the parameters of the signal model

The parameters of the signal model used in this thesis are summarised in table 3.1. The table indicates the treatment of each parameter during the fit to real data - i.e. whether it is left floating freely, constrained via an external constraint, or fixed to an externally measured or calculated value.

The fit was found to be more stable when expressing the amplitudes of the higher ψ resonances and the non local effects on $\Delta\mathcal{C}_7$ in terms of the real and imaginary parts rather than the respective magnitudes and phases.

For the fit to Run 1 data (shown in chapter 7), the relative magnitudes and phases of the $\rho(770)$ and the $\phi(1020)$ are fixed but may be floated in future fits to the merged Run 1 and Run 2 data. The resolution parameters in the mid and high q^2 region are determined from data in separate fits. Then, during the full fit of the signal model, the slope parameters α^2 and α^3 are floated again. This is discussed in section 5.3.

The constraints of the FF parameters are discussed in section 2.5 and 3.3. Furthermore, the fit uses a constraint on $\Delta\mathcal{C}_7$ taken from Ref. [61].

Wilson Coeffs	
$\lvert\mathcal{C}_9\rvert$	free
$\theta_{\mathcal{C}_9}$	fixed
$\lvert\mathcal{C}_{10}\rvert$	free
$\theta_{\mathcal{C}_{10}}$	fixed
$\lvert\mathcal{C}_7\rvert$	fixed
$\theta_{\mathcal{C}_7}$	fixed
Resonances	
$\lvert A_{\parallel}^{J/\psi}\rvert$	free
$\theta_{\parallel}^{J/\psi}$	free
$\lvert A_{\perp}^{J/\psi}\rvert$	free
$\theta_{\perp}^{J/\psi}$	free
$\lvert A_{0}^{J/\psi}\rvert$	fixed
$\theta_{0}^{J/\psi}$	free
$\lvert A_{\parallel}^{\psi(2S)}\rvert$	free
$\theta_{\parallel}^{\psi(2S)}$	free
$\lvert A_{\perp}^{\psi(2S)}\rvert$	free
$\theta_{\perp}^{\psi(2S)}$	free
$\lvert A_{0}^{\psi(2S)}\rvert$	free
$\theta_{0}^{\psi(2S)}$	free
$\lvert A_{\parallel}^{\rho(770)}\rvert$	fixed
$\theta_{\parallel}^{\rho(770)}$	fixed
$\lvert A_{\perp}^{\rho(770)}\rvert$	fixed
$\theta_{\perp}^{\rho(770)}$	fixed
$\lvert A_{0}^{\rho(770)})\rvert$	free
$\theta_{0}^{\rho(770)}$	free
$\lvert A_{\parallel}^{\phi(1020)}\rvert$	fixed
$\theta_{\parallel}^{\phi(1020)}$	fixed
$\lvert A_{\perp}^{\phi(1020)}\rvert$	fixed
$\theta_{\perp}^{\phi(1020)}$	fixed
$\lvert A_{0}^{\phi(1020)})\rvert$	fixed
$\theta_{0}^{\phi(1020)}$	free
$\mathrm{Re}(A_{\parallel}^{\psi(3770)})$	free
$\mathrm{Im}(A_{\parallel}^{\psi(3770)})$	free
$\mathrm{Re}(A_{\perp}^{\psi(3770)})$	free
$\mathrm{Im}(A_{\perp}^{\psi(3770)})$	free
$\mathrm{Re}(A_{0}^{\psi(3770)})$	free
$\mathrm{Im}(A_{0}^{\psi(3770)})$	free

Resonances	
$\mathrm{Re}(A_{\parallel}^{\psi(4040)})$	free
$\mathrm{Im}(A_{\parallel}^{\psi(4040)})$	free
$\mathrm{Re}(A_{\perp}^{\psi(4040)})$	free
$\mathrm{Im}(A_{\perp}^{\psi(4040)})$	free
$\mathrm{Re}(A_{0}^{\psi(4040)})$	free
$\mathrm{Im}(A_{0}^{\psi(4040)})$	free
$\mathrm{Re}(A_{\parallel}^{\psi(4160)})$	free
$\mathrm{Im}(A_{\parallel}^{\psi(4160)})$	free
$\mathrm{Re}(A_{\perp}^{\psi(4160)})$	free
$\mathrm{Im}((A_{\perp}^{\psi(4160)})$	free
$\mathrm{Re}(A_{0}^{\psi(4160)})$	free
$\mathrm{Im}(A_{0}^{\psi(4160)})$	free
FFs	[35]
α_1^{A0}	constr.
α_2^{A0}	constr.
α_0^{A1}	constr.
α_1^{A1}	constr.
α_2^{A1}	constr.
α_0^{A12}	constr.
α_1^{A12}	constr.
α_2^{A12}	constr.
α_0^{V}	constr.
α_1^{V}	constr.
α_2^{V}	constr.
α_0^{T1}	fixed
α_1^{T1}	constr.
α_2^{T1}	constr.
α_0^{T2}	fixed
α_1^{T2}	constr.
α_2^{T2}	constr.
α_0^{T23}	fixed
α_1^{T23}	constr.
α_2^{T23}	constr.

S-Wave	
$\lvert A_{00}^{J/\psi}\rvert$	free
$\theta_{00}^{J/\psi}$	free
$\lvert A_{00}^{\psi(2S)}\rvert$	free
$\theta_{00}^{\psi(2S)}$	free
α_0^{F1}	fixed [43]
α_1^{F1}	fixed [43]
α_2^{F1}	fixed [43]
α_0^{FT}	fixed [43]
α_1^{FT}	fixed [43]
α_2^{FT}	fixed [43]
$\mathcal{C}_{V1}^{S}/(\mathcal{C}_9-\mathcal{C}_{10})$	free
$\mathcal{C}_{V1}^{S}/(\mathcal{C}_9+\mathcal{C}_{10})$	free
$\mathcal{C}_{T}^{S}/\mathcal{C}_7$	free
$\boldsymbol{\Delta\mathcal{C}_7}$	[61]
$\mathrm{Re}(\zeta^{\parallel}e^{i\omega\parallel})$	constr.
$\mathrm{Im}(\zeta^{\parallel}e^{i\omega\parallel})$	constr.
$\mathrm{Re}(\zeta^{\perp}e^{i\omega\perp})$	constr.
$\mathrm{Im}(\zeta^{\perp}e^{i\omega\perp})$	constr.
$\mathrm{Re}(\zeta^{0}e^{i\omega 0})$	constr.
$\mathrm{Re}(\zeta^{0}e^{i\omega 0})$	constr.
Resolution	(sec. 5.3)
α^2	free
N_u^2	fixed
N_l^2	fixed
σ_{gauss}^2	fixed
$\sigma_{CB,}^2$	fixed
f^2	fixed
α^3	free
N_u^3	fixed
N_l^3	fixed
σ_{gauss}^3	fixed
σ_{CB}^3	fixed
f^3	fixed

Table 3.1: All parameters of the signal model. The treatment of each parameter (free floating, fixed, or constrained using external constraints) in the fit to data is indicated.

4. Experimental Setup

4.1 The Large Hadron Collider

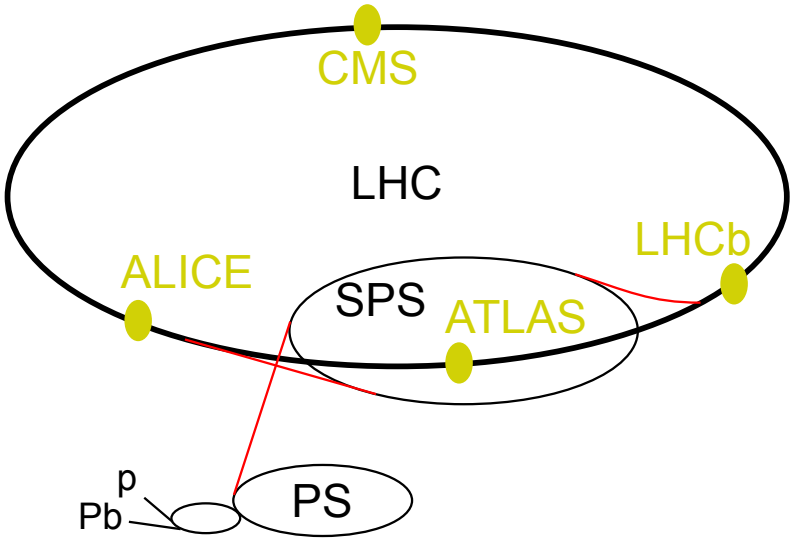


Figure 4.1: The LHC experiments and the pre-accelerators. From Ref. [62].

In order to study the nature of matter at the smallest accessible scale, the European Organisation for Nuclear research (CERN) in Geneva in Switzerland has built the Large Hadron Collider (LHC) [63]. Measuring 26.7 km in circumference, it is the worlds largest particle collider. It is also the worlds most powerful collider, capable of accelerating protons to energies of up to 6.5 TeV. The acceleration is achieved by 16 radio-frequency cavities placed along the beam line, while 1232 superconducting dipole magnets provide a strong magnetic field of up to 7.7 T, bending the proton beams along the ring.
Two proton beams are accelerated simultaneously in opposing direction and collided with centre of mass energies ($\sqrt{s}$) of up to 13 TeV at four intersections of the beam pipes. At each intersection, a particle detector system measures and records the outgoing particles produced in the collisions. Figure 4.1 shows a schematic of the LHC and its pre-acceletators, the Proton Synchroton (PS) and the Super Proton Synchroton (SPS), as well as the positions of the four particle detector systems (also referred to as experiments) at the LHC: ATLAS [64] and CMS [65] are general purpose experiments with a large focus on searching for new fundamental heavy particles such as the Higgs Boson which was

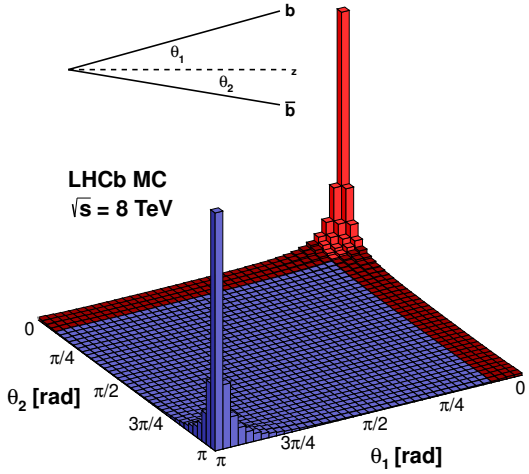


Figure 4.2: Angular distribution of simulated $pp \to b\bar{b}X$ production at $\sqrt{s} = 8\,\text{TeV}$ using PYTHIA 8 [71]. The bins marked red indicate the angular acceptance of the LHCb detector. Figure from Ref. [72].

discovered in 2012 [66, 67]. ALICE [68] studies the Quark Gluon Plasma [69] created in Pb-Pb collisions[1]. LHCb [70] is dedicated to heavy flavour physics with the main focus of finding indirect evidence for New Physics in CP violation and rare decays of hadrons containing b and c quarks.

4.2 LHCb detector layout

At the LHC, b-hadrons are produced through the creation of $b\bar{b}$ quark pairs in the pp-collisions, and the subsequent hadronisation of the b and $\bar{b}$ quarks. The predominant process for $pp \to b\bar{b}$ is gluon fusion, where each gluon carries a certain fraction of its proton's momentum as determined by the proton's Parton Distribution Function. Any asymmetry between the momenta of the gluons causes the $b\bar{b}$ system to be boosted along the beam line with respect to the rest frame of the pp system. Figure 4.2 shows the distribution of the angles θ_1 (θ_2) between the momentum of the b-quark ($\bar{b}$-quark) and the beam line for simulated $pp \to b\bar{b}X$ production. The simulation shows that the majority of $b\bar{b}$ quark pairs collimate in a narrow cone around the beam line on either side of the collision point. In order to study the decay of b-hadrons at the LHC, it is therefore preferential to build a

[1]The LHC is also capable of accelerating Pb ions to energies of up to 2.76 TeV per nucleon.

detector that covers the forward (around $\theta \approx 0$) and/or backward (around $\theta \approx \pi$) region rather than the transverse plane. The angular acceptance of LHCb covers the forward region of approximately 10 mrad to 300 mrad as shown by the red bins in figure 4.2. The angular acceptance is often expressed in terms of pseudorapidity which is defined as:

$$\eta = -\log\left(\tan\left(\frac{\theta}{2}\right)\right). \tag{4.1}$$

The angular acceptance of LHCb is $2 < \eta < 5$.

The choice to build a one sided forward detector rather than a dual sided detector, and therefore loosing half of the $b\bar{b}$ pairs, is justified by the space constraints of the cavern in which LHCb was built. It is preferential to build a single detector arm and making it as long as possible. This allows to fit in more detector systems, as well as to achieve the highest possible precision in the measurement of the momenta of the particles due to a longer leaver arm.

The production cross section for $pp \to b\bar{b}X$ within the LHCb acceptance has been measured as $\sigma(pp \to b\bar{b}X) = 72.0 \pm 0.3 \pm 6.8\,\mu\text{b}$ at $\sqrt{s} = 7\,\text{TeV}$ and $\sigma(pp \to b\bar{b}X) = 144 \pm 1 \pm 21\,\mu\text{b}$ at $\sqrt{s} = 13\,\text{TeV}$ [73]. Therefore, an equal amount of integrated recorded luminosity, which is a measure for the number of recorded pp collisions, at $\sqrt{s} = 13\text{TeV}$ contains about twice as many $b\bar{b}$ pairs as at $\sqrt{s} = 7\text{TeV}$. Table 4.1 summarises the centre of mass energy, the $b\bar{b}$ production cross section as well as the integrated recorded luminosity for the data which is used in this thesis ordered by the year of data taking. The combined periods of 2011 and 2012 are commonly referred to as Run 1 while the combined periods of 2016 and 2017 are referred to in this thesis as Run 2.[2]

Figure 4.3 shows a schematic slice of the LHCb detector in the y-z plane. The z-axis is defined to be parallel to the beam line and the x-y plane is perpendicular to the z axis. The collision point at $x = y = z = 0$ is surrounded by the Vertex Locator (VELO) which tracks charged particles and is responsible for determining the position of the pp collision vertex as well as secondary vertices of decaying particles. Further down the beam line ($z > 0$) is a magnet providing an integrated magnetic field of 4 Tm which bends the path of charged particles. Additional tracking of charged particles, and the determination of the particle's momenta, is performed by the tracking stations consisting of the TT (before the magnet) and T1, T2 and T3 (behind the magnet). Particle identification is achieved

[2]There was also data taken in 2015 and 2018 which are often included in what is referred to as Run 2. However, in this thesis the data from 2015 is omitted for data quality reasons and the data from 2018 is omitted due to a lack of finalised processing of the data by the LHCb collaboration.

Year	$\sqrt{s}$ [TeV]	$\sigma(pp \to b\bar{b}X)$ [μb]	integrated recorded luminosity [fb^{-1}]
2011	7	∼72	1.1
2012	8	∼82	2.1
2016	13	∼144	1.7
2017	13	∼144	1.7

Table 4.1: Centre of mass energy, $b\bar{b}$ production cross section and integrated recorded luminosity at LHCb for the data periods used in this thesis.

using two Ring Imaging Cerenkov detectors RICH1 and RICH2. The calorimeter system, comprised of the scintillating pad detector (SPD) and pre-shower (PS), followed by an electromagnetic (ECAL) and a hadronic calorimeter(HCAL), allows photons, electrons and hadrons to be identified and measures their energy. The muon system (M1-M5) identifies and measures the momentum of muons and also allows triggering on muons.

In the following, the detector systems which are most important for this analysis are explained in more detail.

4.3 Vertexing

Due to their long lifetime, hadrons containing b and c quarks typically fly on the order of several mm before decaying. Therefore, events containing b or c-hadrons, can be identified by finding secondary decay vertices away from the primary pp-vertex (PV). Charged particles produced in a vertex traverse through the VELO [74] depositing small doses of energy, so called *hits*, in the VELO's silicon sensors. Tracking algorithms reconstruct the trajectories, referred to as *tracks*, of the particles. By extrapolating the tracks into the interaction region, the vertices can be reconstructed. Since it is an extrapolation, the further away the reconstructed hits are from the vertex, the larger the uncertainty on the vertex position. Therefore, the VELO is built as close to the beam line as possible: during operation it is 8 mm away from the interaction point which is the closer than any other LHC detector.

The VELO is made of 21 half-moon shaped modules placed on either side of the beam line as shown in figure 4.4. Each module consists of two sets of silicon strip sensors, one

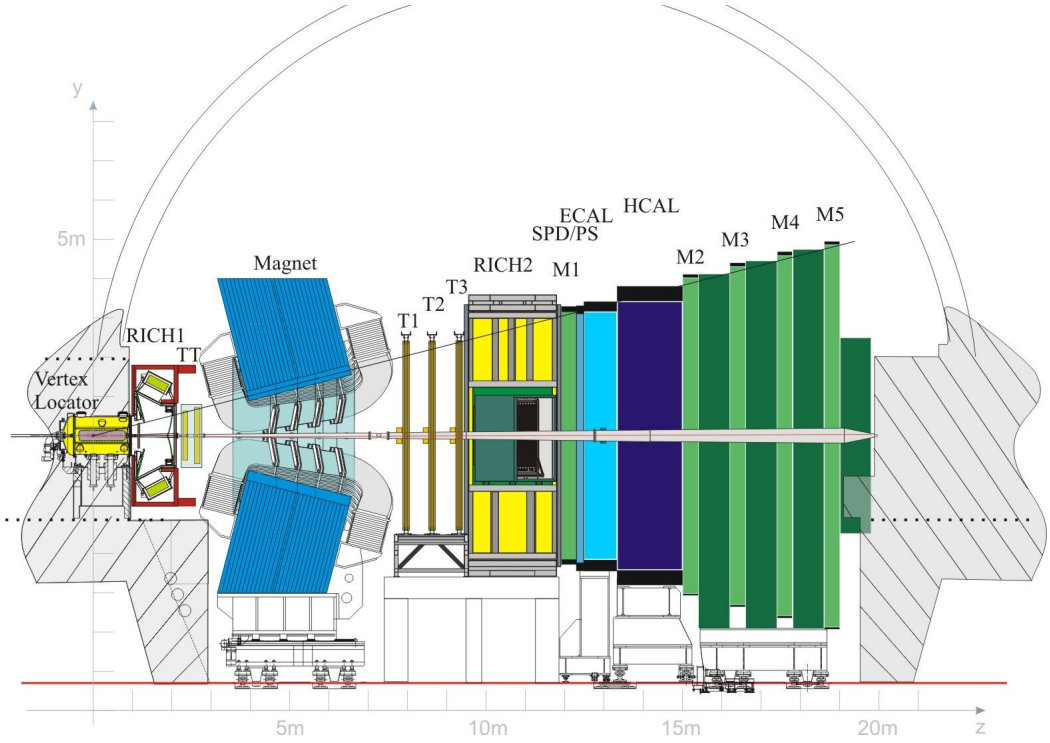


Figure 4.3: Schematic of the LHCb detector. The z-axis is defined to be parallel to the beam line and the x-y plane is perpendicular to the z axis. The origin (0, 0, 0) is defined to be roughly at the collision point. Figure from Ref. [70].

with the strips lined in radial direction to measure the polar angle of hits (ϕ sensors, shown in blue in figure 4.4) and one with strips lined circularly around the module to measure the axial distance of the hits (R sensors, shown in red in figure 4.4)).

The hit position resolution is determined by the width of the strips (referred to as pitch) and by the fact that each module has two perpendicular layers of strips. The pitch varies from $\sim$ 40 µm close to the beam to $\sim$ 100 µm away from the beam, achieving a maximal resolution of $\sim$ 4 µm on the hit position.
The precision on the vertex position depends, besides the hit resolution of the modules and the distance of the modules to the vertex, on the number of tracks coming from each vertex. The resolution on the x (z) coordinate of a vertex is $\sim$ 35 µm ($\sim$ 260 µm) for a vertex with five tracks while the precision is $\sim$ 12 µm ($\sim$ 80 µm) for a vertex with 30 tracks [75].

Besides determining the position of primary and secondary vertices, the VELO is also used to determine the impact parameter (IP) of tracks. IP is defined as the closest distance

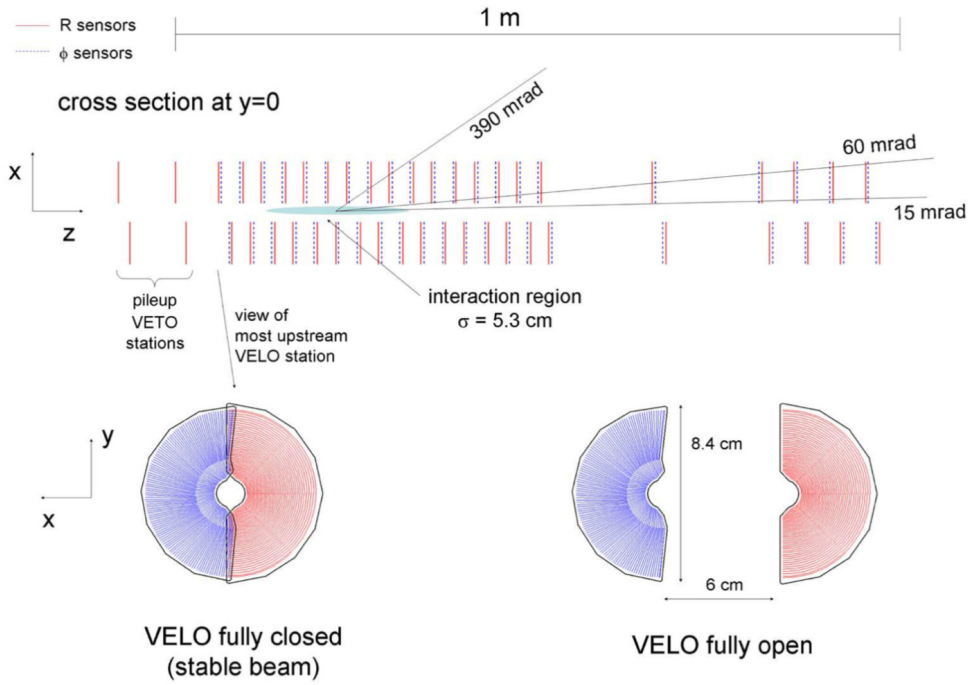


Figure 4.4: Schematic of the VELO detector. Taken from Ref. [70].

of a track to a vertex. Events that contain b- or c-hadrons are characterised by tracks with large IP with respect to the primary vertex. The precision on IP is $\sim$ 20 µm for high momentum tracks.

Both sides of the VELO can be moved in and out of the interaction zone. The narrow distance of 8 mm between the VELO modules and the beam line is only kept while the proton beams are stable. Whenever the proton beams are not stable, for example during ramp up of the beam energy, the VELO models are moved outward by 3 cm to protect them from damage from stray beam particles.

4.4 Tracking

Besides the VELO, several more tracking detectors are installed in LHCb further downstream the beam line. The main purpose of these is to determine the momentum of charged particles from the curvature of the tracks in the known magnetic field. The first tracker is the Tracker Turicensis (TT), positioned right before the magnet (see figure 4.3), which has four layers of silicon strip detectors. Behind the magnet are three tracking detectors T1, T2, T3 with four layers each. Each layer is split into two different detector

technologies. The sections close to the beam-pipe, referred to as Inner Tracker (IT) [76], are subject to a higher occupancy of particles traversing through it and therefore consist of high resolution silicon trackers. The outer sections away from the beam, referred to as Outer Tracker (OT) [77], have a lower occupancy and therefore consist of gas based straw tube detectors, which are cheaper than silicons strips but have a lower resolution.

The momentum resolution of the tracking system varies from $\delta p/p = 0.4\%$ at $p = 5\,\mathrm{GeV}/c$ to $\delta p/p = 0.6\%$ at $p = 100\,\mathrm{GeV}/c$.

4.5 Particle Identification using Ring Imaging Cherenkov dectectors

Many measurements performed by LHCb rely heavily on correctly identifying particle species to reconstruct specific decays. For the reconstruction of the $B^0 \to K^{*0}(\to K^+\pi^-)\mu^+\mu^-$ decay the correct identification of kaons and pions is essential.

In LHCb, the main system responsible for particle identification (PID) of hadrons is the RICH system, consisting of two Ring imaging Cherenkov detectors: RICH1 and RICH2 [78]. They work based on the fact that a particle traversing through a medium faster than the speed of light in that medium radiates a cone of light in a process called Cherenkov radiation. The RICH detectors collect this light using mirrors and photo-detectors and determine the opening angle θ_c of the light cone. The angle θ_c is related to the mass m and momentum p of the particle as well as the refractive index n of the medium:

$$\cos(\theta_c) = \frac{\sqrt{m^2\,c^2 + p^2}}{pn}. \tag{4.2}$$

The left hand side of figure 4.5 shows the reconstructed Cherenkov angles in RICH1 as a function of particle momentum for different particle species. Clear distinct bands for the different particle species are visible, however with increasing momentum they all tend towards the same asymptotic limit and loose all separation. This is because for large momenta $\cos(\theta_c) \to 1/n$, i.e. all particles produce the same radius only dependent on the refractive index of the medium. Therefore, the PID separation power of the RICH decreases with increasing particle momentum.

The two RICH detectors use different mediums, with different refractive indices, optimized for resolution at different momentum ranges. RICH1, which is placed before the magnet and uses C_4F_{10} ($n = 1.0014$), provides PID for kaons and pions with momenta

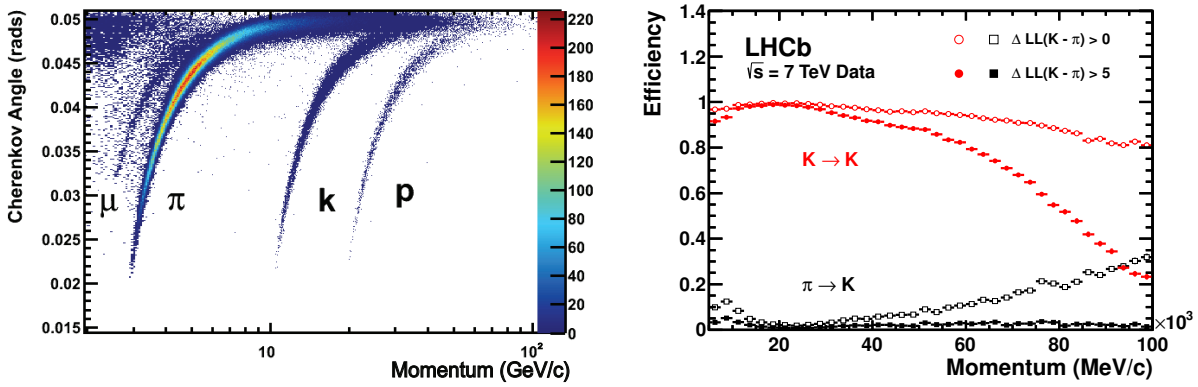


Figure 4.5: Left: Reconstructed Cherenkov angles as a function of reconstructed particle momentum in calibration data with the C_4F_{10} radiator of the RICH1. Distinct bands are visible for different particle species. The separation between the lines decreases with increasing momentum. Right: Kaon identification efficiency (red) and pion misidentification rate (black) in the RICH system using two different cuts on the $\mathrm{DLL}_{K\pi}$ variable: $\mathrm{DLL}_{K\pi} > 0$: (open points) and $\mathrm{DLL}_{K\pi} > 5$ (closed points). Figures from Ref. [79].

of $2 - 40\,\mathrm{GeV}/c$ in the full LHCb angular acceptance of $25 - 300\,\mathrm{mrad}$ [79]. RICH2 uses CF_4 ($n = 1.0005$) and provides PID for kaons and pions with momenta of $15 - 100$ GeV/c, within an angular acceptance more focused on the region closer to the beam line of $15 - 120\,\mathrm{mrad}$ [79]. The choice of angular coverage is driven by the fact that higher momentum particles are more likely to traverse the detector at smaller polar angles.

Using the momentum information obtained by the tracking, as well as the Cherenkov angles reconstructed via the photons measured in the RICH, equation 4.2 is used to fit different mass hypotheses to each track. The difference in Log-Likelihoods (DLL) of two different mass hypotheses can then be used to discriminate between two particle species for a given track, for example:

$$\mathrm{DLL}_{K\pi} = \log \mathcal{L}(\theta_c, p|K) - \log \mathcal{L}(\theta_c, p|\pi) \tag{4.3}$$

The kaon identification efficiency as well as the rate of pions being misidentified as kaons using the $\mathrm{DLL}_{K\pi}$ variable is shown in the right hand side of figure 4.5. Two different cuts on $\mathrm{DLL}_{K\pi}$ are compared. A stricter cut of $\mathrm{DLL}_{K\pi} > 5$ keeps the pion misidentification rate below 5%, even at high momenta. However, it also causes the kaon identification efficiency rate to drop significantly to less then 50% for $p > 80$ MeV/c. On the other hand, a looser cut of $\mathrm{DLL}_{K\pi}$ keeps the kaon identification efficiency above 80% for $p < 80$ MeV/c at the cost of an increased pion misidentification rate which reaches ∼30% at $p = 100$ MeV/c.

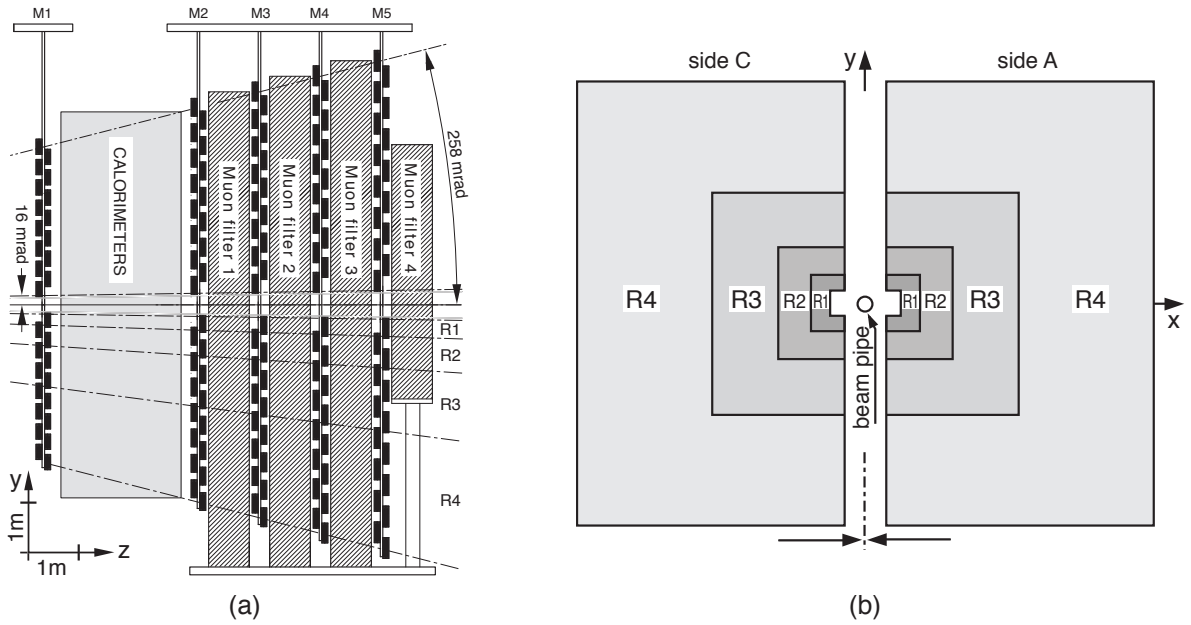


Figure 4.6: Schematic of the muon detector in the y-z plane (left) and x-y plane (right). Figure from Ref. [80].

The left hand side of figure 4.5 shows that the separation of muons and pions is rather poor in the RICH. Therefore, an additional detector system is used for identifying muons as explained in the next section.

4.6 Muon Identification

Just like kaons and pions, the identification muons is crucial for the analysis presented in this thesis.

Experimentally muons have the advantage of penetrating material much better than most other particles, while also being long lived enough to travel through the full length of the detector. The LHCb muon system [81] takes advantage of this fact by being placed at the very end of the LHCb apparatus as shown in figure 4.3.

The muon system consists of five rectangular stations M1-M5 with M1 being placed in front of the calorimeters and M2-M5 being placed behind the calorimeters. Sections M2-M5 are interlaced with 80 cm thick iron shields - labelled as muon filters in figure 4.6 - which absorb all particles except for muons with $p > 6$ GeV/c.

Each muon system station has 276 multi-wire proportional chambers (MWPCs) which detect hits of charged particles passing through them. In case of M1, the sections closest

momentum range	Required hits in muon stations
$3\,\mathrm{GeV}/c < p < 6\,\mathrm{GeV}/c$	M2 and M3
$6\,\mathrm{GeV}/c < p < 10\,\mathrm{GeV}/c$	M2, M3 and either M4 or M5
$p > 10\,\mathrm{GeV}/c$	M2, M3, M4 and M5

Table 4.2: The conditions for the isMuon variable. From Ref. [82].

to the interaction region, which are subject to higher levels of radiation, are made up of 12 gas electron multiplier (GEM) detectors instead of MWPCs. Each station is segmented into 4 regions (marked R1-R4 in the right hand side of figure 4.6), in which the size of the MWPCs is increasing as a function of the distance from the beam-pipe, ensuring consistent occupancy across the regions.

Since all other particles are stopped in the calorimeters or the iron shields, muon tracks can be identified simply by requiring hits in the muon stations. Whether or not a track is labelled as a muon is expressed by the binary variable isMuon. The conditions for isMuon are given in table 4.2. The isMuon variable has an identification efficiency of about 95% (for tracks with $0.8\,\mathrm{GeV}/c < p_\mathrm{T} < 1.7\,\mathrm{GeV}/c$)[3] - 100% for tracks with $p_\mathrm{T} > 5\,\mathrm{GeV}/c$ [82]. The misidentification probability is below $\sim 2\%$ for protons, kaons and pions [82].

Another PID variable based on the muon system, is a likelihood determined by matching hits in the muon stations with the extrapolated track of a charged particle determined by the tracker.

The muon system is also used in the trigger as explained in the following section.

4.7 Trigger

The LHC provides pp-collisions at a rate of 40 MHz. This is many orders of magnitude higher than the rate at which data can be processed and stored. Also, only a fraction of the pp-collisions produce $b(c)$-hadrons, and since the amount of storage is limited, it is preferential to discard all events that likely do not contain any $b(c)$-hadrons. This is achieved by the trigger system, which decides when to read out the detector and which events get written to long term storage.

When filtering the events that were written to storage to obtain an event sample

[3] p_T refers to the transverse momentum - i.e. the component of the momentum transverse to the beam line.

containing a specific signal decay such as $B^0 \to K^{*0}\mu^+\mu^-$, the events can be classified with two categories. Events are classified as TOS (Trigger on Signal), if the trigger objects of the signal decay (for example the muons of the $B^0 \to K^{*0}\mu^+\mu^-$ decay) have sufficient energy to trigger the event. Events are classified as TIS (Trigger independent of Signal), if the events contain particles not associated to the signal decay with high enough energy to cause the event to be triggered. These definitions allow the determination of the TOS trigger efficiency ϵ^{TOS}, using background subtracted data, with

$$\epsilon^{\text{TOS}} = \frac{N(\text{TOS and TIS})}{N(\text{TIS})}, \tag{4.4}$$

where $N(\text{TIS})$ refers to the number of TIS events in the sample, and $N(\text{TOS and TIS})$ is the number of events which meet both the TOS and TIS definitions.

The trigger is split into two stages: The low level trigger (L0) which is implemented in specialised hardware, and the high level trigger (HLT) which is implemented in software. Both stages are explained in the following.

4.7.1 Low Level Trigger

The main goal of the L0 is to decide when to read out the detector. It uses information from the calorimeters and the muon system, both of which can be read out quickly, to accept or reject events.

The L0 calorimeter trigger and L0 muon trigger work independently. Events triggered by the L0 calorimeter trigger are not used in this analysis and therefore this section focuses on the L0 muon trigger.

Since events which contain heavy flavor hadrons are often characterised by particles with high transverse momentum p_{T}, the L0 muon trigger accepts events containing either a single high-p_{T} muon (`L0Muon`) or a pair of muons with high p_{T} (`L0DiMuon`).

The L0 muon trigger works by reconstructing straight-line tracks from aligned hits in the five muon stations. An estimate of the p_{T} is determined under the assumption that each track is from a muon which originated from the interaction region and traversed the known integrated magnetic field. The p_{T} resolution is $\sim 25\%$.
The trigger algorithm then identifies the largest and second largest p_{T} in each event and accepts the event if either the largest p_{T} is above the `L0Muon` threshold, or the product of the largest and the second largest p_{T} is above the `L0DiMuon` threshold. The thresholds used for taking the data, which is analysed in this thesis, is given in table 4.3. The thresholds are optimized for maximised signal efficiency under the different LHC running conditions.

	L0Muon threshold	L0DiMuon threshold
	$p_{\mathrm{T}}^{\mathrm{largest}}$	$p_{\mathrm{T}}^{\mathrm{largest}} \times p_{\mathrm{T}}^{2^{\mathrm{nd}}\ \mathrm{largest}}$
2011 [84]	> 1.48 GeV	$> (1.296\ \mathrm{GeV})^2$
2012 [84]	> 1.76 GeV	$> (1.6\ \mathrm{GeV})^2$
2016 [83]	> 1.8 GeV	$> 2.25\ \mathrm{GeV}^2$
2017 [83]	> 1.35 GeV	$> 1.69\ \mathrm{GeV}^2$

Table 4.3: L0Muon and L0DiMuon thresholds used to take the majority of the data in each year.

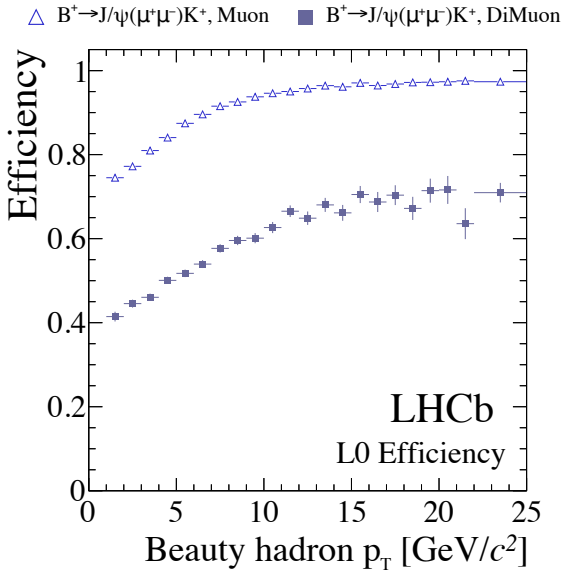


Figure 4.7: Efficiencies of the L0Muon and L0DiMuon trigger in Run 2 data for $B^+ \to J/\psi(\mu^+\mu^-)K^+$ decays as a function of the p_{T} of the B^+. The efficiencies were determined using samples of trigger unbiased signal events. Based on a Figure from Ref. [83].

Furthermore, the L0 trigger includes a requirement on the maximum number of hits in the SPD (SPD multiplicty) to reduce the complexity of the events and therefore speed up the reconstruction. For L0Muon, the SPD multiplicity is required to be smaller than 450, while for L0DiMuon, which accepts a much lower event rate, a looser cut on the SPD multiplicity of < 900 is used. The use of the SPD multiplicty does not cause a significant loss in signal efficiency [83].

Figure 4.7 shows the trigger efficiencies (defined in equation 4.4) of L0Muon and L0DiMuon in Run 2 data for $B^+ \to J/\psi(\mu^+\mu^-)K^+$ decay as a function of the p_{T} of the

B^+. In terms of the trigger the $B^+ \to J/\psi(\mu^+\mu^-)K^+$ decay is very similar to the decay analysed in this thesis, which therefore has a comparable trigger efficiency.

The output bandwidth of the L0 trigger is 1 MHz, of which 400 kHz are reserved for the muon triggers `L0Muon` and `L0DiMuon`.

4.7.2 High Level Trigger

Events that passed the L0 trigger are then processed by the HLT. The HLT is implemented in software and runs on large computing farm of 27000 physical cores, capable of running $\sim$ 50000 threads using hyper-threading. The HLT processes the events in two stages: HLT1 and HLT2.

High Level Trigger 1

In HLT1 a partial event reconstruction is performed. The information from the full tracking system is used to reconstruct tracks of charged particles that deposited hits in all tracking stations. These tracks are referred to as long tracks. The quality of each track is determined by calculating the χ^2 of the fitted track with respect to the associated hits. Furthermore, tracks in the VELO are used to reconstruct the PV. Due to timing constraints of HLT1, the only PID that is performed is the identification of muons, which is achieved by matching long tracks with hits in the muon stations. This is only attempted for tracks in events which were triggered by `L0Muon`.

The decision whether or not an event is accepted or rejected by HLT1 is based on several independent algorithms called trigger lines. Besides specialised trigger lines used for collecting calibration data or for collecting low multiplicity events, there are two different inclusive trigger lines as well as muon trigger lines. In the analysis presented in this thesis, events selected by the inclusive trigger lines and the muon trigger lines are used.

The inclusive trigger lines have the goal of selecting events which contain b- or c-hadrons by looking for either a single track with large IP or a two track vertex with large displacement with respect to the PV. The two track trigger line was only introduced in 2015 onward. The tracks must meet quality requirements based on their χ^2.

The muon trigger lines only run on events selected by `L0Muon`. They aim at selecting events containing decays of b- or c- hadrons with muons in the final state. There are three major muon lines: The first looks for a single muon with large IP with respect to the PV and with large p_{T}; the second muon line selects dimuon pairs with a dimuon mass matching with the mass of charmonia or bottomonium resonances; the third muon line

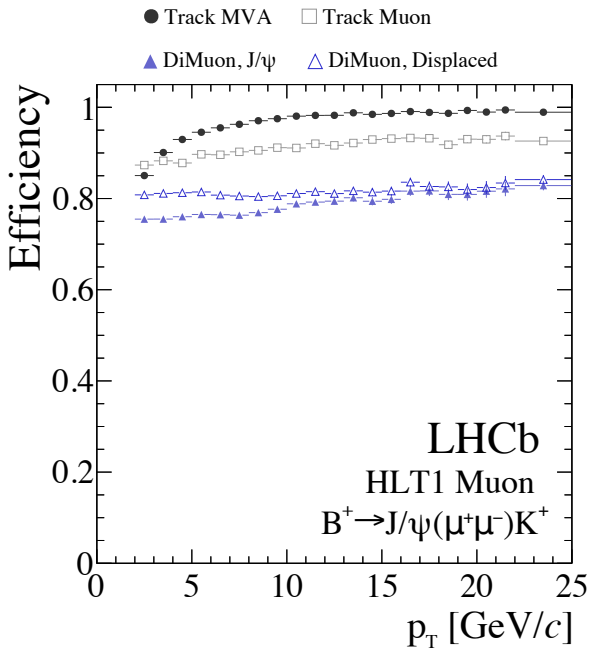


Figure 4.8: Efficiency of the HLT1 muon trigger lines and the HLT1 inclusive single track trigger line (Track MVA) in Run 2 data for $B^+ \to J/\psi(\mu^+\mu^-)K^+$ decays as a function of the p_{T} of the B^+. Figure from Ref. [83].

selects dimuons from a displaced secondary vertex. The efficiency of the muon lines and the efficiency of the single track inclusive line are shown in figure 4.8. Since the muon lines are only applied to events triggered by `L0Muon` and `L0DiMuon`, their absolute efficiencies are lower than the efficiency of the inclusive line, which runs on all L0 events.

The HLT1 trigger stage reduces the event rate of $\sim$ 1 MHz, which it receives from L0, to $\sim$ 110 kHz which gets passed to the second stage HLT2.

High Level Trigger 2

In HLT2 a full event reconstruction is performed. Besides long tracks, also tracks that only deposited hits in the T stations (T tracks) as well as tracks that deposited hits in the TT and the T stations (downstream tracks) are reconstructed. HLT2 also performs full PID using the muon system, the RICH detectors, and the calorimeter system.

In Run 1 the detector was not fully calibrated during the reconstruction in HLT2, causing the event variables to slightly differ from the variables obtained during offline

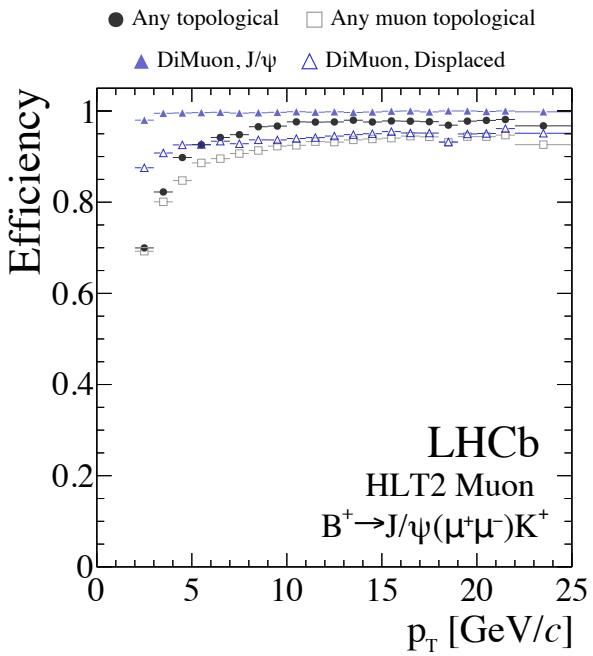


Figure 4.9: Efficiency of select HLT2 muon trigger lines, the inclusive topological trigger ("any topological") lines and the topological triggers lines where one track is required to be identified as a muon ("any muon topological") in Run 2 data for $B^+ \to J/\psi(\mu^+\mu^-)K^+$ decays as a function of the p_{T} of the B^+. Figure from Ref. [83].

reconstruction[4]. During Run 2, due to improvements in computing power and optimisation of algorithms, it was possible to use the full calibration data in HLT2, achieving the highest possible resolution of the variables. This allowed the HLT2 trigger decisions to be optimized in Run 2.

Similarly to HLT1, the decision whether an event is accepted or rejected by HLT2 is based on several independent trigger lines. The HLT2 lines which are relevant for this analysis are presented in the following.

Inclusive b-hadron trigger lines search for a two-, three-, or four-track vertex that is displaced from the PV and has sizeable p_{T}. These trigger lines are also referred to as topological trigger lines as they use the typical topology of b-hadron decays to identify them. They are implemented with multivariate classifiers which are trained to optimize the selection criteria to select b-hadrons which can be fully reconstructed inside the LHCb acceptance and to reject c-hadrons which are the biggest contribution to background.

The HLT2 muon trigger lines select a wide range of muonic signals, similar to the

[4]Offline reconstruction refers to the event reconstruction which is performed on the stored data.

HLT1 trigger lines. Important for this analysis are the muon and dimuon trigger lines which require large displacement from the PV. Figure 4.9 shows the efficiency (defined in equation 4.4) of HLT2 muon trigger lines, the inclusive topological trigger ("any topological") lines and the topological triggers lines where one track is required to be identified as a muon ("any muon topological"). Compared to the any topological trigger, the any muon topological lines have a lower efficiency because they only process events selected by the HLT1 single-muon trigger line.

Due to improvements in various parts of the data acquisition (DAQ), allowing faster storage of events, the output rate of HLT2 trigger changed over the course of the data taking: 3 kHz in 2011 [85], 5 kHz in 2012 [84], and 12.5 kHz from 2015 onward [83].

4.8 Generation of simulated events

As in any particle physics experiment, the distributions of the observables measured by the LHCb detector - in case of this analysis the helicty angles and q^2 of $B^0 \to K^{*0}\mu^+\mu^-$ - are different to their true distributions in nature. This is due to the effect of the detector acceptance, detector response, the trigger and the subsequent steps in the event selection[5], which may favor certain values of the observables over others. In order to correct the measured distributions for these effects, so called full Monte Carlo Simulations (MC) are produced. In this context 'full' refers to the fact that the full sequence of events from the initial pp collision until the reconstruction of the final state particles is simulated. Simulations are produced in several stages, where each stage is handled by a different software package.

The pp collisions and the primary particles produced within them - including b-hadrons - are simulated using Pythia [71, 86]. The subsequent decay of unstable particles (e.g. b-hadrons) into secondary particles is handled by EvtGen [87]. The interaction of the produced particles with the LHCb detector and the detector response is then simulated using GEANT4 [88]. The output from the simulated detector response is then digitized by the BOOLE package [89] and reconstructed in the same way as data from the real detector using BRUNEL [89].

The resulting simulations can be analysed just like data but also carry the additional information of the true variables (e.g. momentum, PID, decay chain information) for each particle. Comparing true and reconstructed distributions of MC events, allows the

[5]The event selection will be presented in section 5.1.

quantification of detector/trigger/selection effects as well as the optimization of signal selection procedures. For this analysis, this includes the parameterisation of the acceptance (see section 5.2) as well as part of the parameterisation of the resolution (see section 5.3).

4.8.1 Correcting differences between data and simulation

In order to rely on the simulations to model the acceptance and resolution, it is important to ensure that the simulated detector replicates all effects of the real detector correctly. Any systematic difference in the distributions of simulated events and data events are corrected for. Differences in the PID variables are corrected using so-called *PID resampling* and several other differences are corrected by reweighting the simulated events to match the distributions in data.

PID resampling

Since the PID variables are used to suppress several background contributions, as described in section 5.1, it is important that the PID variables agree well between data and simulation. However, due to the presence of a large number of low energy photons in the RICH detector, as well as due to other factors, the simulation does not reproduce the PID variables well. In order to improve the agreement between simulation and data, the PID variables in the simulations are resampled using clean high-statistics control samples from data as input. The data control samples, given in table 4.4, are selected using only tracking quantities. The control samples are used to produce calibration histograms of each PID variable in bins of track pseudorapidity η, the number of tracks per event (`nTracks`), and track p_{T}. Each track in the simulation is then matched to a bin in η, `nTracks`, and p_{T} and each PID variable is randomly sampled from the respective calibration histogram.

To validate the resampling method, the PID variables are compared in $B^0 \to J/\psi\, K^{*0}$ data and $B^0 \to J/\psi\, K^{*0}$ simulation. Since this data also includes background events, the

Particle	Sample
K	$D^{*+} \to D^0(\to K^+\pi^-)\pi^+$
π	$D^{*+} \to D^0(\to K^+\pi^-)\pi^+$
μ	$J/\psi \to \mu^+\mu^-$

Table 4.4: High statistics data samples used for the PID resampling.

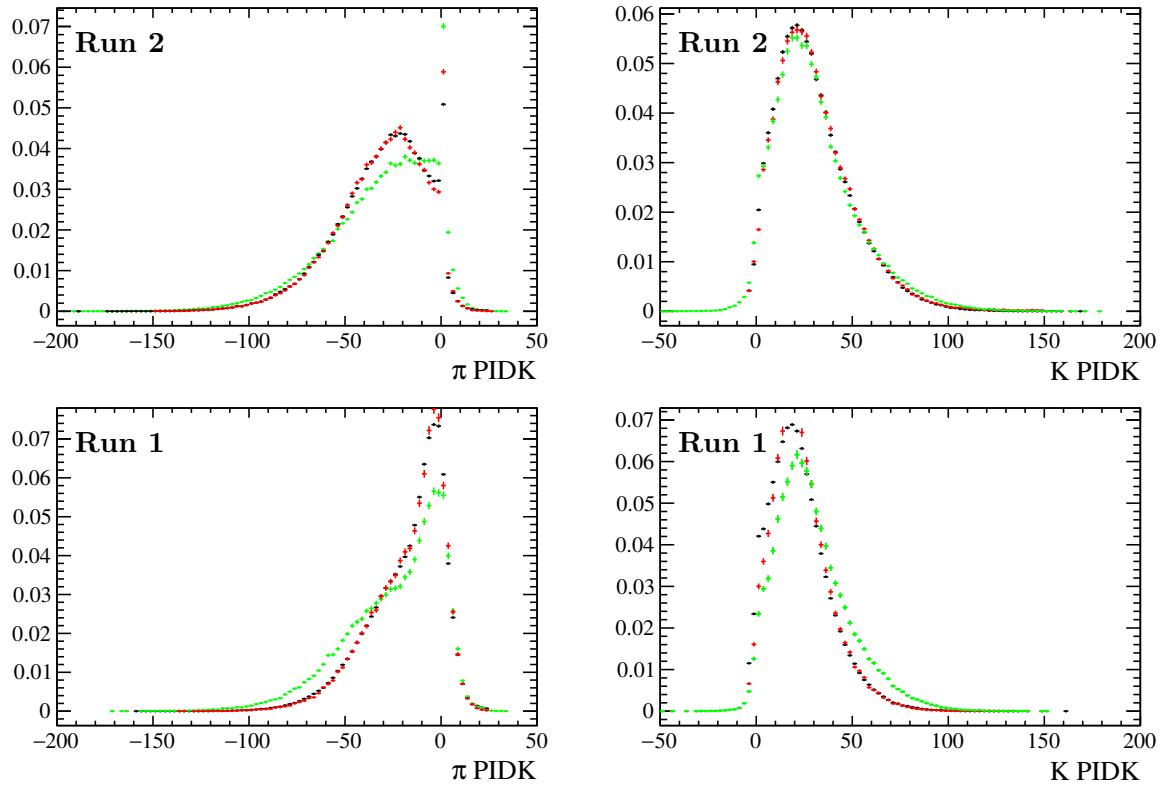


Figure 4.10: PID distributions of the hadrons in *sWeighted* $B^0 \to J/\psi K^{*0}$ data and simulation. The top figures show the comparison of *sWeighted* data (black), non-resampled simulation (green) and resampled simulation (red) for Run 2. The same is shown in the bottom plots but for Run 1.

sPlot technique described in Ref. [90] is used. The *sPlot* technique works by calculating a weight for each event, proportional to the probability of the event being a signal event. These so-called *sWeights* are determined by using a control variable, in this case the $m^{\text{rec}}_{B^0}$ distribution. Applying the *sWeights* to the events in the $B^0 \to J/\psi K^{*0}$ data results signal-like distributions in each of the variables of interest.

Figure 4.10 shows the distribution of the PID variables of the hadrons in *sWeighted* $B^0 \to J/\psi K^{*0}$ data (black points) and in non-resampled simulation (green) and resampled simulation (red). The top plots show the Run 2 data and simulation while the bottom plots show the Run 1 data and simulation. The resampled simulation shows good agreement with the *sWeighted* data. The spike at 0 in the Run 2 pion PID distribution is caused by events which are below the RICH momentum threshold. The threshold was higher during Run 2 than during Run 1 because of hardware modifications of the RICH detector.

Kinematic reweighting

The distributions of the number of tracks per event `nTracks`, the vertex fit quality of the signal candidate (χ^2_{Vtx}), and the B^0 candidate p_{T} in the simualation show discrepencies to the respective distributions in data and therefore need to be corrected. This is done by determining weights for each simulated event such that the simulation matches the data. The weights are calculated by comparing the `nTracks`, χ^2_{Vtx}, $p_{\text{T}}(B^0)$ distributions in *sWeighted* $B^0 \to J/\psi\, K^{*0}$ data to the distributions in the simulation. The weights are derived sequentially, whereby first a weight is derived to correct the `nTracks` distribution, which is then applied to the simulated events before determining the weight for $p_{\text{T}}(B^0)$. Finally, the product of the first two weights are applied before deriving the weight for χ^2_{Vtx}. These weights are then applied to all simulated events based on the `nTracks`, χ^2_{Vtx}, and $p_{\text{T}}(B^0)$ of each event. The effect of the reweighing is shown for Run 2 simulations in figure 4.11. The distributions of the reweighed simulation agree well with the *sWeighted* data.

Several checks have been performed to further validate the reweighting [91]. It has been confirmed that quantities being reweighed are not correlated to the angles $\cos\theta_K$, $\cos\theta_\ell$, and ϕ. Also the agreement of data and simulation has been confirmed for several other variables. Furthermore, it has been shown that the weights can be applied to correct the simulation across the full q^2 range.

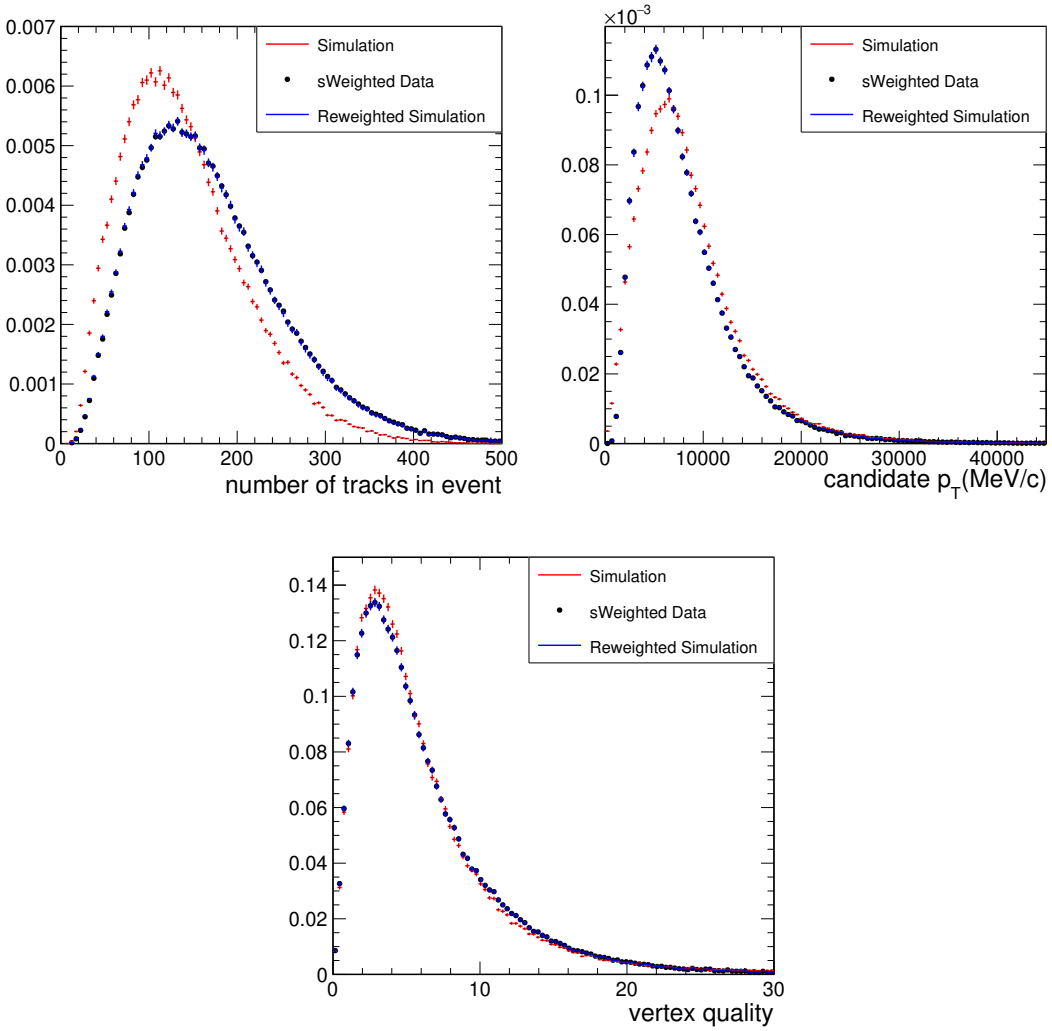


Figure 4.11: Distributions of `nTracks`, χ^2_{Vtx}, and $p_{\mathrm{T}}(B^0)$ of *sWeighted* data (black) and the corresponding simulation before (red) and after (blue) reweighing.

5. Event selection and corrections of experimental effects

As stated in section 4.2, the data analysed in this thesis was taken in 2011 and 2012 (Run 1) as well as 2016 and 2017 (Run 2). Ultimately, it is planned to fit the merged datasets of Run 1 and Run 2. However, in order to fit the signal model to the data, an acceptance model is required (see section 5.2). Currently, only separate acceptance parameterisations for Run 1 and Run 2 data are available, but in order to fit the merged dataset an averaged acceptance parametersation needs to be determined. This is beyond the scope of this thesis. Since the parameterisation of the background (see chapter 6) does not require an acceptance parameterisation, the background studies are presented based on the merged dataset. Furthermore, a fit with the full model including the signal and background components is performed using Run 1 data only.

The goal of the analysis presented in this thesis is to simultaneously determine the Wilson Coefficients as well as the resonant contributions by fitting the full q^2 spectrum of $B^0 \to K^{*0}\mu^+\mu^-$ events. The procedure for selecting $B^0 \to K^{*0}\mu^+\mu^-$ candidate events from the events recorded by LHCb is described in section 5.1. The particle reconstruction and the event selection procedure can influence the shape of the angular and q^2 distributions of the reconstructed $B^0 \to K^{*0}\mu^+\mu^-$ decays. These so called acceptance effects are accounted for during the fit by multiplying the signal model with an acceptance function as described in section 5.2.
Due to the limited resolution of the detector, the reconstructed peaks of the narrow resonances $\phi(1020)$, J/ψ, and $\psi(2S)$ are much wider than their natural widths. In order to properly describe the data with the model described in chapter 3, the model is convolved with a resolution function as described in section 5.3.

5.1 Selection

The events analysed in this thesis were recorded by LHCb in 2011 and 2012 (Run 1) as well as 2016 and 2017 (Run 2). The purpose of the selection procedure is to identify the events which most likely contain signal candidates. Signal candidates are reconstructed by correctly identifying and combining the tracks of the four final state particles of

$B^0 \to K^{*0}(\to K^+\pi^-)\mu^+\mu^-$ decays including all resonant modes. The final state particles are identified using the mass and charge hypothesis from the event reconstruction described in chapter 4. Any event that does not contain a correctly reconstructed signal candidate is referred to as a background event.

The fit presented in this thesis is using the same event sample as the published measurement of the q^2-binned angular observables [1]. Additionally, the 2017 data is analysed in this thesis using the identical selection algorithms as for the 2016 data. The selection procedure is therefore presented briefly, omitting any discussion of the optimisation of the different cut values, which would be beyond the scope of this thesis. For a more detailed discussion see Refs. [91–93].

The selection is performed in four main stages. It starts with the trigger, which was explained in section 4.7. The sample of the triggered events is then passed through the *stripping* where various criteria are applied to reduce the size of the event sample by selecting only well reconstructed events that are likely to contain signal candidates. The stripped data is then passed through the so called pre-selection, a series of cuts aimed at reducing the number of combinatorial background events, while retaining as many signal events as possible. Combinatorial background events are events where the candidate is formed by random combinations of tracks not belonging to a signal decay. The stripping as well as the pre-selection is presented in section 5.1.1.

Background contributions from misidentified decays, referred to as peaking backgrounds, are removed with specific cuts as described in section 5.1.2. In the last step of the selection process, a multivariate analysis (MVA) classifier, which is trained to identify signal events, is used to further reduce background contributions as explained in section 5.1.3. The resulting number of candidate events in the different kinematic regions of interest are summarised in section 5.1.4.

5.1.1 Stripping and Preselection

The triggers used to collect events containing decays of B mesons with muons in the final state are explained in section 4.7. After the event reconstruction is complete and potential signal decays are identified, only TOS events are selected for analysis, i.e. only events where the trigger objects associated to the signal decay were responsible for triggering the event.

The triggered events are then passed through a stripping algorithm, also referred to as stripping line. The conditions of the stripping line `B2XMuMu` used to select $B^0 \to K^{*0}\mu^+\mu^-$

Candidate	Stripping Run 1	Run 2
B meson	IP $\chi^2 < 16$ (best PV)	IP $\chi^2 < 16$ (best PV)
B meson	$4800\,\text{MeV}/c^2 < m^{\text{rec}}_{B^0} < 7100\,\text{MeV}/c^2$	$4800\,\text{MeV}/c^2 < m^{\text{rec}}_{B^0} < 7100\,\text{MeV}/c^2$
B meson	DIRA angle $< 14\,\text{mrad}$	DIRA angle $< 14\,\text{mrad}$
B meson	flight distance $\chi^2 > 121$	flight distance $\chi^2 > 121$
B meson	vertex $\chi^2/\text{ndf} < 8$	vertex $\chi^2/\text{ndf} < 8$
$\mu^+\mu^-$	$m(\mu^+\mu^-) < 7100\,\text{MeV}/c^2$	$m(\mu^+\mu^-) < 7100\,\text{MeV}/c^2$
$\mu^+\mu^-$	vertex $\chi^2/\text{ndf} < 9$	vertex $\chi^2/\text{ndf} < 9$
K^{*0}	$m(K^+\pi^-) < 6200\,\text{MeV}/c^2$	$m(K^+\pi^-) < 6200\,\text{MeV}/c^2$
K^{*0}	vertex $\chi^2/\text{ndf} < 9$	vertex $\chi^2/\text{ndf} < 8$
K^{*0}	flight distance $\chi^2 > 9$	flight distance $\chi^2 > 16$
tracks	ghost Prob < 0.4	ghost Prob < 0.5
hadron	min IP $\chi^2 > 9$	min IP $\chi^2 > 6$
muon	min IP $\chi^2 > 9$	min IP $\chi^2 > 9$
muon	`IsMuon`	`IsMuon`
muon	$\text{DLL}_{\mu\pi} > -3$	$\text{DLL}_{\mu\pi} > -3$
GEC	SPD Mult. < 600	SPD Mult. < 600

Table 5.1: Stripping selection criteria of the `B2XMuMu` line for Run 1 and Run 2.

Candidates	Selection
Track	$0 < \theta < 400$ mrad
Track Pairs	$\theta_{pair} > 1$ mrad
K	$\text{DLL}_{K\pi} >$ -5
π	$\text{DLL}_{K\pi} < 25$
PV	$\lvert X - \langle X\rangle\rvert < 5\,\text{mm}$
PV	$\lvert Y - \langle Y\rangle\rvert < 5\,\text{mm}$
PV	$\lvert Z - \langle Z\rangle\rvert < 200\,\text{mm}$

Table 5.2: Pre-selection cuts applied to the stripped event sample. In this table θ refers to the angle of a track relative to the beamline and θ_{pair} is the opening angle between two track pairs.

events, including any resonant modes such as $B^0 \to J/\psi\, K^{*0}$, are summarised in table 5.1 for Run 1 and Run 2. The differences between the cut values for Run 1 and Run 2 are due to changes in the running condition of LHCb and are discussed in Ref. [91].

The first series of cuts of the stripping are related to the B^0 meson candidate. The B^0 is required to be likely to originate from the PV and a broad cut is placed on $m_{K\pi\mu\mu}$, i.e. the reconstructed invariant mass of the B^0 ($m_{K\pi\mu\mu}$ is also denoted as $m^{\text{rec}}_{B^0}$ hereafter). The DIRA angle is defined as the angle between a line drawn from the PV vertex to the decay vertex (DV) of the B^0 and the sum of the 4-momenta of its decay products. Requiring a small DIRA angle increases the probability of having a well reconstructed B^0 meson. The B^0 candidate is also required to have travelled a significant distance from the PV before decaying. Furthermore, a reasonable quality of the DV of the B^0 is required.
To ensure the two muons originate from the same vertex, a cut is placed on the dimuon vertex quality. Also a broad limit is placed on the dimuon mass to reject events with dimuon masses much larger than the expected range in $B^0 \to K^{*0}\mu^+\mu^-$ decays[1].
The reconstructed K^{*0} is also required to have a good vertex quality. Furthermore, the flight distance of the K^{*0} relative to the PV is required to be sufficiently large to ensure the K^{*0} vertex is not compatible with being part of the PV.

For each final state track, a cut is placed on the probability of being a ghost track, defined as a track that is falsely reconstructed from hits in the detector that are not associated to a given particle. Furthermore, it is required that the final state particles are well separated from the PV by requiring a minimum IP. For each muon the PID variable `IsMuon` (explained in section 4.6) is required to be true and a minimum requirement is placed on the probability that the muon is a muon rather than a pion.
Lastly, in order to reject events with a large number of background tracks, which are less likely to be reconstructed well, a cut is placed on the SPD multiplicity (see also section 4.7.1).

After the stripping, the pre-selection cuts are applied which are summarised in table 5.2. Identical cuts are used for Run 1 and Run 2 data. All final state tracks are required to have angles relative to the beamline such that the track is within the acceptance of the LHCb detector. Also, a minimum requirement is placed on the opening angle between track pairs to ensure that tracks can be well separated in the reconstruction. The PID variable $\text{DLL}_{K\pi}$ is used to ensure a high probability that both the pion and the kaon are identified correctly. Finally, in order to reject events containing tracks from beam-gas interactions, cuts are placed on the distance of the PV position from the average PV position.

After the pre-selection, an additional cut is placed on the $K\pi$ invariant mass, $m_{K\pi}$,

[1]Ultimately, the fit is performed in the range $(0.18 \leq q^2 \leq 18)$ GeV$^2/c^4$

choosing a window of $\pm 100\,\mathrm{MeV}/c^2$ around the nominal K^{*0} mass:

$$795.9 < m_{K\pi} < 995.9\,\mathrm{MeV}/c^2. \tag{5.1}$$

Furthermore, for events which contain multiple reconstructed $B^0 \to K^{*0}\mu^+\mu^-$ candidates ($< 1\%$ of all events after selection) one candidate is chosen randomly and all tracks associated to the other candidates are considered to be background tracks.

5.1.2 Peaking backgrounds

Several decays can be misidentified as $B^0 \to K^{*0}(\to K^+\pi^-)\mu^+\mu^-$ decays and therefore end up in the selected event sample as so called peaking backgrounds. Furthermore, true $B^0 \to K^{*0}(\to K^+\pi^-)\mu^+\mu^-$ decays can be reconstructed wrongly (for example by swapping the K and π) which also contributes to the peaking backgrounds. To reject peaking backgrounds, specific cuts are used as outlined below.

Misreconstructed $B^0 \to K^{*0}(\to K^+\pi^-)\mu^+\mu^-$ decays where the K and π were swapped are rejected by requiring

$$\mathrm{DLL}_{K\pi}(K) > \mathrm{DLL}_{K\pi}(\pi). \tag{5.2}$$

Two vetos are used to reject $\Lambda_b^0 \to pK^-\mu^+\mu^-$ decays. They can mimic signal decays if the proton is misidentified as a pion. These are vetoed by rejecting events with

$$\begin{aligned} &\mathrm{DLL}_{p\pi}(\pi) > 0 \text{ and} \\ &m_{K(\pi\to p)\mu\mu} \in [5575, 5665]\,\mathrm{MeV}/c^2, \end{aligned} \tag{5.3}$$

where $m_{K(\pi\to p)\mu\mu}$ denotes the invariant mass of the reconstructed B^0 candidate when swapping the mass of the K for the mass of a proton. $\Lambda_b^0 \to pK^-\mu^+\mu^-$ decays can also mimic signal decays if the proton is misidentified as a kaon and the kaon as pion. To remove this background source, events with

$$\begin{aligned} &\mathrm{DLL}_{K\pi}(\pi) > 0 \text{ and} \\ &m_{(K\to p)(\pi\to K)\mu\mu} \in [5575, 5665]\,\mathrm{MeV}/c^2 \end{aligned} \tag{5.4}$$

are rejected.

The decays $B^0 \to J/\psi(\to \mu^+\mu^-)K^{*0}(\to K^+\pi^-)$ and $B^0 \to \psi(2S)(\to \mu^+\mu^-)K^{*0}(\to K^+\pi^-)$ can contribute to the peaking backgrounds if the final state particles are swapped

through misidentification: $\mu^- \leftrightarrow \pi^-$ or $K^+ \leftrightarrow \mu^+$. These swapped backgrounds from $B^0 \to J/\psi\, K^{*0}$ are vetoed by rejecting events with

$$\begin{aligned} m_{(\pi\to\mu)\mu} \in [3036, 3156]\,\mathrm{MeV}/c^2 \text{ and} \\ \mathrm{DLL}_{\mu\pi}(\pi) > 5.0 \text{ or } \texttt{isMuon}(\pi), \end{aligned} \tag{5.5}$$

as well as rejecting events with

$$\begin{aligned} m_{(K\to\mu)\mu} \in [3036, 3156]\,\mathrm{MeV}/c^2 \text{ and} \\ \mathrm{DLL}_{\mu\pi}(K) > 5.0 \text{ or } \texttt{isMuon}(K). \end{aligned} \tag{5.6}$$

Analogously, the swapped backgrounds from $B^0 \to \psi(2S)K^{*0}$ decays are vetoed by rejecting events with

$$\begin{aligned} m_{(\pi\to\mu)\mu} \in [3626, 3746]\,\mathrm{MeV}/c^2 \text{ and} \\ \mathrm{DLL}_{\mu\pi}(\pi) > 5.0 \text{ or } \texttt{isMuon}(\pi) \end{aligned} \tag{5.7}$$

and rejecting events with

$$\begin{aligned} m_{(K\to\mu)\mu} \in [3626, 3746]\,\mathrm{MeV}/c^2 \text{ and} \\ \mathrm{DLL}_{\mu\pi}(K) > 5.0 \text{ or } \texttt{isMuon}(K). \end{aligned} \tag{5.8}$$

.

The rare decay $B_s^0 \to \phi(\to K^+K^-)\mu^+\mu^-$ can mimic a signal event if the K^- is misidentified as a π^-. This background is vetoed by rejecting events with

$$m_{K(\pi\to K)\mu\mu} \in [5321, 5411]\,\mathrm{MeV}/c^2 \tag{5.9}$$

and rejecting events with

$$\begin{aligned} m_{K(\pi\to K)} \in [1010, 1030]\,\mathrm{MeV}/c^2 \text{ and } \mathrm{DLL}_{K\pi}(\pi) > -10 \text{ or} \\ m_{K(\pi\to K)} \in [1030, 1075]\,\mathrm{MeV}/c^2 \text{ and } \mathrm{DLL}_{K\pi}(\pi) > +10. \end{aligned} \tag{5.10}$$

Finally, the rare decay $B^+ \to K^+\mu^+\mu^-$ can contribute to the peaking backgrounds if the final state is combined with a random pion from the rest of the event. This contribution is only present in the upper mass side-band of the $m_{B^0}^{\mathrm{rec}}$ spectrum and is removed by rejecting events with

$$\begin{aligned} m_{K\pi\mu\mu} > 5380\,\mathrm{MeV}/c^2 \text{ and} \\ m_{K\mu\mu} \in [5220, 5340]\,\mathrm{MeV}/c^2. \end{aligned} \tag{5.11}$$

This veto also removes combinatorial background events in the upper mass side-band, causing a sculpting of the background distributions. This has to be taken into account when using the upper mass side-band to parameterise the combinatorial background, as described in chapter 6.

5.1.3 Multivariate Selection

In the final step of the selection procedure a multivariate analysis technique is used to further suppress background events. This involves training a multivariate classifier, in this case a Boosted Decision Tree (BDT) [94], to classify signal and background events using their distinct characteristics in data. The input variables for the BDT are:

- B^0 lifetime;
- B^0 momentum and p_{T};
- B^0 direction angle (DIRA);
- $K^+\pi^-\mu^+\mu^-$ vertex χ^2;
- $\mathrm{DLL}_{K\pi}$ of the kaon and pion;
- $\mathrm{DLL}_{\mu\pi}$ of the muons;
- isolation of the four final state particles.

The isolation is defined as the number of background tracks that can form a vertex with a given final state track.
The BDT is trained using data. The events in the upper mass side-band ($(5350 \leq m^{\mathrm{rec}}_{B^0} \leq 7000)$ MeV/c^2) are labelled as background events. The sample of events labelled as signal events is obtained from $B^0 \to J/\psi\, K^{*0}$ data. Since this data also includes background events, the *sPlot* technique, already mentioned in section 4.8, is used. Applying the per-event *sWeights* to the $B^0 \to J/\psi\, K^{*0}$ data yields signal-like distributions in each of the input variables.

In order to maximise the performance of the BDT, the *k-folding* technique is employed. The dataset is divided into 10 sub-samples and the BDT is trained using 9 of the sub-samples as input and evaluated on the remaining sub-sample. Then, the training step is repeated while including the previously excluded sub-sample and excluding a different sub-sample. This is repeated until 10 different BDTs have been trained. When applying

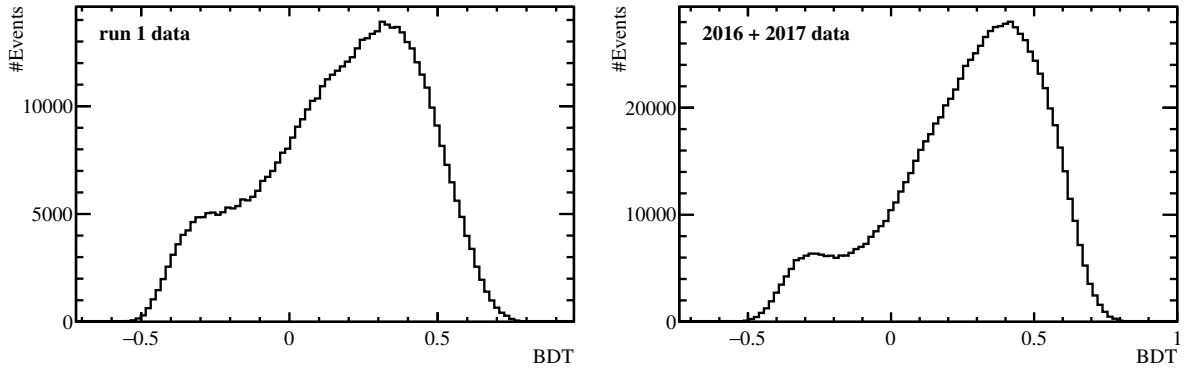


Figure 5.1: Distribution of the BDT variable of the events in the kinematic region considered in this thesis ($(5220 \leq m^{\text{rec}}_{B^0} \leq 5840)$ MeV/c^2, $(0.18 \leq q^2 \leq 18.0)$ GeV$^2/c^4$, $(795.9 \leq m_{K\pi} \leq 995.9)$ MeV/c^2) in Run 1 data (left) and Run 2 data (right). All events with BDT > 0.1 are used for the analysis.

the multivariate classifier to the event sample after stripping and pre-selection, the output of these 10 BDTs is averaged to obtain a single BDT variable for each event – except for the events in the J/ψ region, where for each given event only the single BDT is used that was trained on the 9 sub-samples which did not include the given event.

The distribution of the BDT variable in Run 1 data (left) and Run 2 data (right) is shown in figure 5.1.

The cut value for selecting the event sample for analysis is chosen such that signal significance defined as

$$s = \frac{N_{\text{sig}}}{\sqrt{N_{\text{sig}} + N_{\text{bkg}}}} \tag{5.12}$$

is maximised. This is given for BDT > 0.1 for both Run 1 and Run 2 data as shown in Ref. [91][2]. The signal efficiency for this BDT cut value is 77% in Run 1 and 87% in Run 2, whereas the background rejection efficiency is 95% in Run 1 and 97% in Run 2 [50].

[2]The actual maximum for s is found at BDT ≈ 0 for both Run 1 and Run 2 data. However in order to be conservative and to avoid being affected by fluctuations which may shift the observed maximum, a slightly higher cut is used for the analysis.

5.1.4 Number of candidate events after selection

The number of candidate events per q^2-region after the selection in Run 1 only data and in the merged Run 1 and Run 2 data are given in table 5.3 for the signal region and the upper mass side-band.

	$m_{B^0}^{\text{rec}}$ region	low q^2	mid q^2	high q^2	total
Run 1	signal region	450	260772	17547	278769
	side-band	64	3156	419	3639
Run 1 + Run 2	signal region	1556	868624	57504	927684
	side-band	169	7637	1018	8824

Table 5.3: Number of candidate events per q^2 region (see table 5.4 for the definition of the q^2 regions) in the signal region ($(5239.58 \leq m_{B^0}^{\text{rec}} \leq 5319.58)$ MeV/c^2) and in the upper mass side-band ($(5440 \leq m_{B^0}^{\text{rec}} \leq 5840)$ MeV/c^2) in Run 1 only and merged Run 1 and Run 2 data.

5.2 Acceptance

The angular and q^2 distributions of the final event sample are not the same as the true distributions of $B^0 \to K^{*0}\mu^+\mu^-$ events, due to so called acceptance effects. These include the limited coverage and limited efficiency of the detector as well potential biases and inefficiencies in the selection procedure. To take the acceptance effects into account in the fit to data, the signal model (described in chapter 3) is multiplied with an acceptance function.

To determine the acceptance function, simulated $B^0 \to K^{*0}\mu^+\mu^-$ events, which are generated according to a phase space model, are analysed. The phase space model yields flat $\cos\theta_\ell$, $\cos\theta_K$, and ϕ distributions of the generated events. However, the q^2 distribution is not flat as it is less likely to generate an event with large q^2. Therefore, the events are reweighed to obtain a flat q^2 distribution at generator level. The generated events are passed through the full detector simulation, reconstruction and selection procedures (see section 4.8). Since the distributions are flat at generator level, any non-flat shape in the $\cos\theta_\ell$, $\cos\theta_K$, ϕ, and q^2 distributions of the events after reconstruction and selection is due

to the acceptance effects. Therefore, the acceptance function is obtained by parameterising the $\cos\theta_\ell$, $\cos\theta_K$, ϕ and q^2 distributions of the events after reconstruction and selection with

$$\varepsilon(\cos\theta_\ell, \cos\theta_K, \phi, q^2) = \sum_{k,l,m,n} c_{klmn} P(\cos\theta_\ell, k) P(\cos\theta_K, l) P(\phi, m) P(q^2, n), \tag{5.13}$$

where $P(x, m)$ are Legendre polynomials as a function of x and of order m. The observables ϕ and q^2 are re-scaled to the range $-1 \leq x \leq 1$ when evaluating the polynomial. The factors c_{klmn} are the acceptance coefficients. Using the orthogonality of the Legendre polynomials, the acceptance coefficients can be calculated with

$$\begin{aligned} c_{klmn} = \frac{1}{N'} \sum_{i=1}^{N} w_i \Bigg[&\left(\frac{2k+1}{2}\right)\left(\frac{2l+1}{2}\right)\left(\frac{2m+1}{2}\right)\left(\frac{2n+1}{2}\right) \\ &\times P(\cos\theta_\ell, k) P(\cos\theta_K, l) P(\phi, m) P(q^2, n) \Bigg], \end{aligned} \tag{5.14}$$

where the sum indicates a sum over the simulated events. The factors w_i include the weights per event which are applied to obtain a flat q^2 distribution as well as the weights used to correct for differences in the simulation and data (described in section 4.8.1). N' is the normalisation associated with the weights, given by $N' = \sum_{i=1}^{N} w_i$.

The orders of the Legendre polynomials are chosen as the set of lowest orders with which the acceptance parameterisation can describe the acceptance effect well: $k = 4$, $l = 5$, $m = 6$, $n = 5$ [1]. Thus, there are 720 coefficients in total.
In this thesis, the same acceptance coefficients are used as in the published measurement of the angular observables in $B^0 \to K^{*0}\mu^+\mu^-$ decays [1]. The acceptance is parameterised for Run 1 and Run 2 separately, since the detector condition as well as some of the selection criteria are different for the two periods. Since, only Run 1 data is analysed with the signal model (see introduction of this chapter), the acceptance is only shown for Run 1. Figures 5.2 shows the q^2, $\cos\theta_\ell$, $\cos\theta_K$, and ϕ distributions of the phase-space simulated $B^0 \to K^{*0}\mu^+\mu^-$ events after reconstruction and selection as well as the one-dimensional projections of the acceptance parameterisation for Run 1.

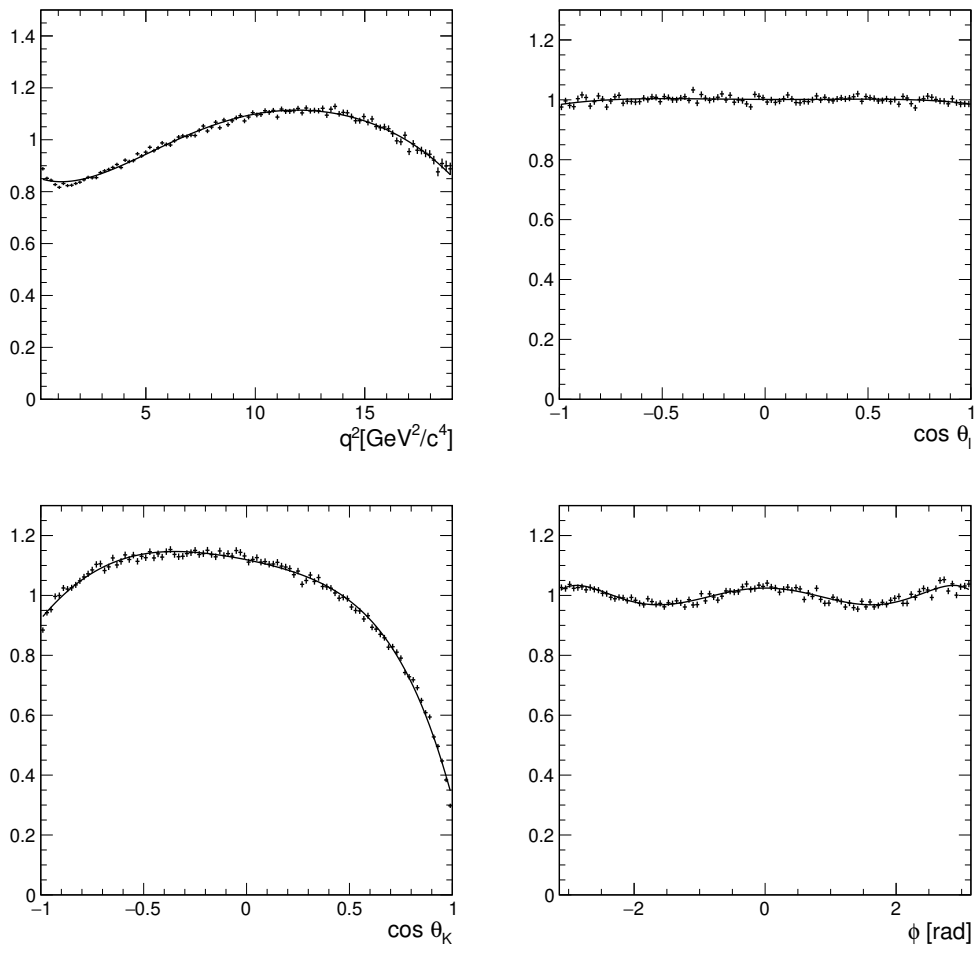


Figure 5.2: One-dimensional projections of the four-dimensional acceptance parametrisation on q^2, $\cos\theta_\ell$, $\cos\theta_K$, and ϕ for Run 1. The points are the distributions of $B^0 \to K^{*0}\mu^+\mu^-$ events generated with a phase space model and passed though the full reconstruction and event selection procedure.

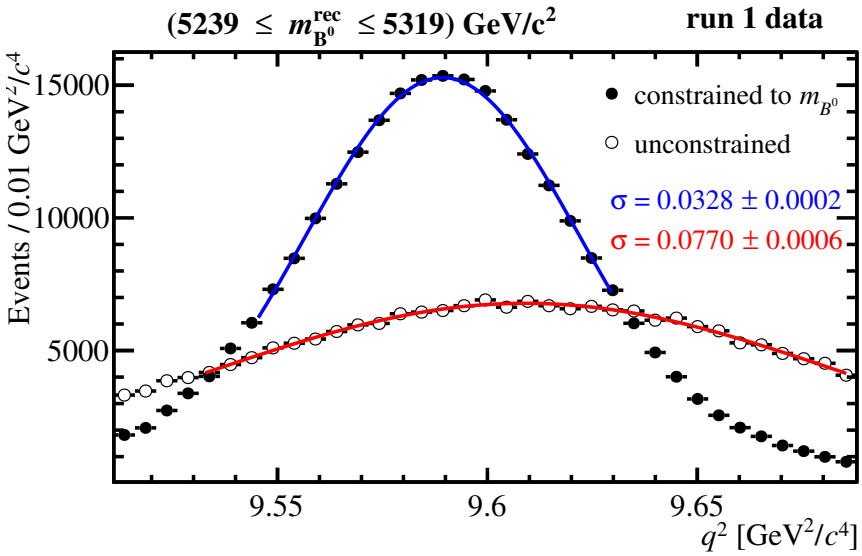


Figure 5.3: Unconstrained-q^2 (open points) and constrained-q^2 (closed points) distribution of the Run 1 events in the narrow core region of the J/ψ peak. Each distribution is parameterised with a Gaussian (blue line and red line) and the resulting width parameters (σ) are given on the figures. The mass constraint improves the q^2 resolution by a factor of ~ 2.3.

5.3 Resolution

5.3.1 B^0 mass constraint

The model described in chapter 3 contains the narrow resonances ϕ, J/ψ, and $\psi(2S)$ which have natural widths much narrower than the q^2 resolution in the measurement. In order to maximize the sensitivity to the interference between the resonances and the penguin amplitudes, improving the q^2 resolution is crucial. Therefore, a mass constraint is used, which works by varying the reconstructed 4-momenta of the final state particles within their measured uncertainties such that the invariant mass $m_{K\pi\mu\mu}$ best matches the known mass of the B^0 meson ($m_{B^0} = 5279.58\,\mathrm{MeV}/c^2$ [30]). Since the constraint adds additional information into the measurement, the resolution in q^2 is improved with this method.[3]

At the peak position of the J/ψ, the mass constraint improves the q^2 resolution by a factor of ~ 2.3 as shown in figure 5.3, where the unconstrained q^2 distribution (open points) and the constrained q^2 distribution (closed points) of the events in the core of

[3]The mass constraint also improves the angular resolution, but since the angular distributions have no sharp peaks, the angular resolution is much less important than the q^2 resolution. Thus, the unconstrained angles are used in the fit.

the J/ψ peak in Run 1 data are shown. Also shown are Gaussian fits (red line) to each distribution and the resulting width parameters (σ).
For the remainder of the thesis q^2 refers to the mass constrained q^2, unless specified otherwise.

5.3.2 Resolution model

Despite the improvement of the q^2 resolution, discussed in the previous section, the reconstructed peaks of the $\phi(1020)$, J/ψ, and $\psi(2S)$ resonances are still much wider than the respective natural widths. Thus, in order to fit the signal model (described in section 3) to the data, the model is convolved with a resolution function, following the approach used in Ref. [49]. The other resonances besides the $\phi(1020)$, J/ψ, and $\psi(2S)$ are much wider than the q^2 resolution and are therefore not significantly influenced by the resolution. However, the three narrow resonances can only be described when modelling the resolution correctly.

The q^2 resolution changes as a function of q^2, but it would not be computationally feasible to model the resolution with a function that continuously changes as a function of q^2.[4] Therefore, three independent resolution regions are defined such that each region contains one of the narrow resonances. The q^2 regions are given in table 5.4. The resolution parameters are constant within each q^2 region. The fit with the signal model is then performed as a simultaneous fit[5] to these regions, where all signal parameters are shared across the regions but the resolution parameters in each region are independent.

The resolution in each region is modelled with a sum of a Gaussian G and a double sided crystal ball C with common mean parameter μ, i.e.

$$R(q^2, \vec{\lambda}) = fG(q^2, \mu, \sigma_G) + (1-f)C(q^2, \mu, \sigma_C, \alpha_l, \alpha_u, \eta_l, \eta_u), \tag{5.15}$$

where f is the relative fraction of the Gaussian with respect to the double sided Crystal Ball function, and σ_G and σ_C are the widths of the Gaussian and the double sided Crystal

[4]The convolution is implemented via a Fast Fourier Transform, which would not be possible if the resolution parameters depend on q^2. During the fit, the convolution needs to be calculated for each call of the likelihood sum. Therefore, an efficient algorithm for calculating the convolution is crucial in order to maximise the computational efficiency of the fit.

[5]'Simultaneous fit' refers to a maximum likelihood fit where the event sample is split into sub-samples according to a defined segmentation of the data (e.g the three q^2 regions). A separate likelihood sum is calculated for each region, where the PDF depends on the region (e.g. different resolution parameters). Parameters that enter the PDF in several regions (e.g. all signal parameters) are referred to as shared parameters.

	q^2 range ($\mathrm{GeV}^2/\mathrm{c}^4$)
low q^2	[0.18, 3.24]
mid q^2	[3.24, 11.56]
high q^2	[11.56, 18.0]

Table 5.4: Definition of q^2 regions used throughout the thesis.

Ball function respectively. The double sided Crystal ball function is given by

$$C(q^2, \mu, \sigma_C, \alpha_l, \alpha_u, \eta_l, \eta_u) = \begin{cases} A_l(B_l - \delta)^{\eta_l} & \text{if } \delta < \alpha_l \\ e^{-\frac{1}{2}\delta^2} & \text{if } \alpha_l < \delta < \alpha_u \\ A_u(B_u - \delta)^{\eta_u} & \text{if } \delta > \alpha_u \end{cases} \tag{5.16}$$

with,

$$\begin{aligned} \delta &= (q^2 - \mu)\sigma_C, \\ A_{l,u} &= \left(\frac{\eta_{l,u}}{|\alpha_{l,u}|}\right)^{\eta_{l,u}} e^{-\frac{1}{2}\cdot|\alpha_{l,u}|^2}, \\ B_{l,u} &= \frac{\eta_{l,u}}{|\alpha_{l,u}|} - |\alpha_{l,u}|, \end{aligned} \tag{5.17}$$

where $\alpha_{l,u}$ and $\eta_{l,u}$ are the parameters that set the slope and onset of the lower (denoted with index l) and upper (denoted with index u) exponential tail.

The strategy for determining the parameters depends on the q^2 region. In the low q^2 region the number of resonant $B^0 \to \phi K^{*0}$ events in the data is too small to determine the resolution from data reliably. Hence, in this region the resolution parameters are determined using $B^0 \to K^{*0}\mu^+\mu^-$ simulations. In the mid and high q^2 region, the number of resonant events is sufficient for determining the resolution parameters directly from data.

The resolution parameters for Run 1 and Run 2 data have been compared and were found to be compatible. Therefore, it is feasible to determine a resolution parameterisation for the merged dataset using the same functional form as given in equation 5.15. However, as discussed in the introduction of this chapter, in this thesis only the Run 1 data is fitted with the signal model, and therefore only the Run 1 resolution parameterisation is shown in this section.

Determination of the resolution parameters in the low q^2 region

In order to determine the resolution parameters in the low q^2 region, simulated $B^0 \to K^{*0}\mu^+\mu^-$ events are used, which contain both the information of the reconstructed particles as well as the generator level 'true' values for all kinematic observables. For each event, the difference of true and reconstructed q^2 is calculated. The distribution of q^2_{true} - q^2_{rec} of all events within $(0.18 \leq q^2_{\text{rec}} \leq 3.24)$ GeV$^2/c^4$ is then parameterised with the resolution model (equation 5.15) as shown in figure 5.4. The resulting resolution parameters are given in table 5.5. In the final fit to data (described in chapter 7), the resolution parameters in the low q^2 region are fixed to these values.

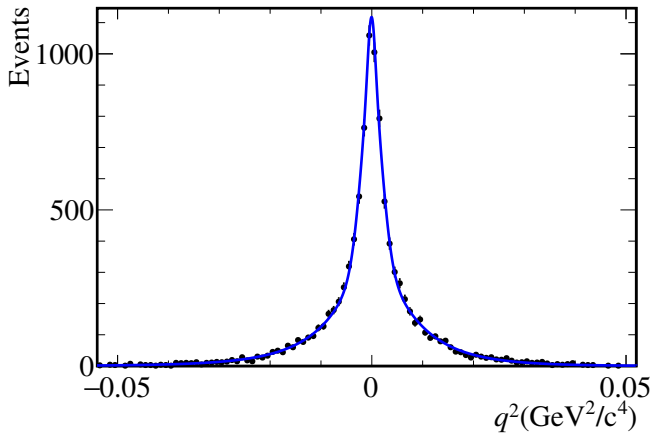


Figure 5.4: Fit of the resolution model to the q^2_{true} - q^2_{rec} distribution of simulated $B^0 \to K^{*0}\mu^+\mu^-$ decays with $(0.18 \leq q^2_{\text{rec}} \leq 3.24)$ GeV$^2/c^4$ for Run 1. Figure from Ref. [50].

Parameter	Value
σ_C	$2.02\times10^{-3} \pm 2.76\times10^{-4}$
σ_G	$7.21\times10^{-2} \pm 2.76\times10^{-4}$
f	$3.33\times10^{-2} \pm 2.47\times10^{-2}$
α_l	$-2.80\times10^{-1} \pm 2.21\times10^{-2}$
α_u	$2.93\times10^{-1} \pm 2.33\times10^{-2}$
η_l	18.84 ± 7.95
η_u	9.31 ± 1.71

Table 5.5: Resolution parameters in the low q^2 region. The values are obtained by fitting q^2_{true} - q^2_{rec} of simulated $B^0 \to K^{*0}\mu^+\mu^-$ events with $(0.18 \leq q^2_{\text{rec}} \leq 3.24)$ GeV$^2/c^4$ with the resolution model given in equation 5.15.

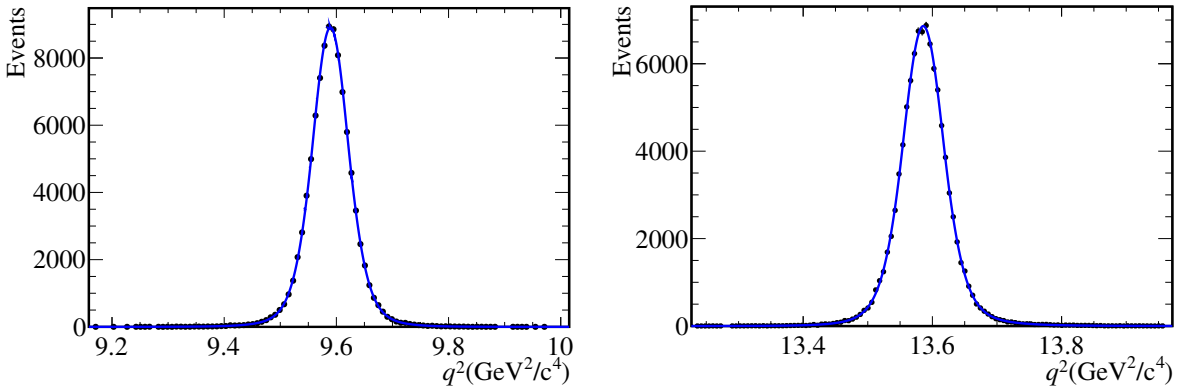


Figure 5.5: Fits of relativistic Breit Wigner functions convolved with the resolution model to the q^2 distribution of simulated $B^0 \to J/\psi K^{*0}$ (left) and $B^0 \to \psi(2S)K^{*0}$ (right) decays for Run1. Figures from Ref. [50].

Determination of the resolution parameters in the mid and high q^2 region

The values of the resolution parameters in the mid and high q^2 regions are determined in several steps. Firstly, simulated $B^0 \to J/\psi K^{*0}$ and $B^0 \to \psi(2S)K^{*0}$ events are used to validate the resolution function. A relativistic Breit-Wigner function is convolved with the resolution model and fitted to the q^2 distribution of simulated $B^0 \to J/\psi K^{*0}$ or $B^0 \to \psi(2S)K^{*0}$ events as shown in figure 5.5. It is found that relativistic Breit-Wigner functions convolved with the resolution model are good descriptions of the J/ψ and $\psi(2S)$ peaks. The simulation studies also show that the peaks can be well described with a symmetric slope parameter $\alpha_l = -\alpha_u$. Therefore, the resolution model is simplified by replacing α_l and $-\alpha_u$ with a single α parameter.

Secondly, the pole mass of the J/ψ amplitude in the signal model, described in chapter 3, is set to the measured peak position of the J/ψ peak in data: $m_{J/\psi} = 3096.63 \pm 0.01\,\text{MeV}/c^2$. Therefore, the μ parameter in the resolution model is set to zero since the resolution function does not need to account for any shift of the J/ψ peak.

Thirdly, the data events in the core region of the J/ψ peak ($(9.44 \leq q^2_{\text{rec}} \leq 9.74)\ \text{GeV}^2/c^4$) are fitted with the full signal model convolved with the resolution model. The J/ψ amplitude is dominating the considered q^2 region, such that effects from the penguin amplitudes are negligible. In this fit, the tail parameters (α, $\eta_{l,u}$) are fixed to the best fit values from the fit to the simulation and only the widths σ_G and σ_C and the fraction f are allowed to float.

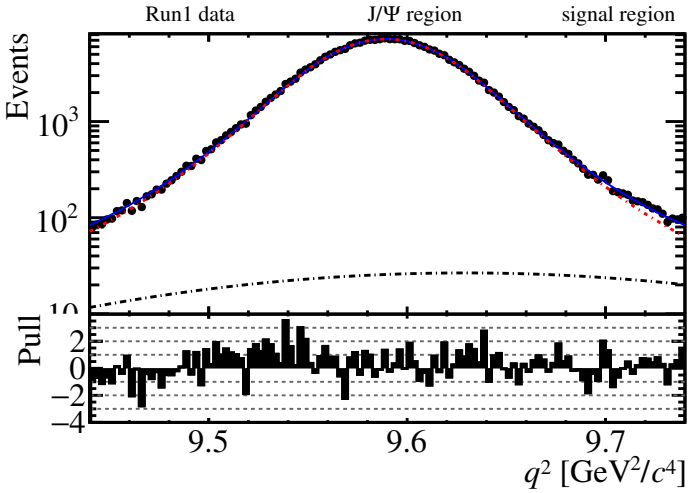


Figure 5.6: Projection on q^2 of the 4 dimensional fit to the events in the core region of the J/ψ peak in Run 1 data using the full signal and background model to determine the core parameters of the resolution model. The dotted red line indicates the signal PDF, the dashed black line indicates the background PDF (see chapter 6), and the solid blue line shows the sum of the signal and background PDFs. The pull is defined as the difference of the value of data and the value of the projection of the PDF at the center of the respective bin, divided by the uncertainty of the data.

Figure 5.6 shows the data in the core region of the J/ψ peak as a function of q^2 as well as the q^2 projection of the fit PDF. The dotted red line shows the signal PDF, the dashed black line shows the background PDF (explained in detail in chapter 6), and the solid blue line shows the sum of the signal and background PDFs.

The determination of the resolution parameters in the high q^2 region is done similarly by fitting the $\psi(2S)$ peak. In the last step the core of the $\psi(2S)$ in data is fitted with a single Breit Wigner function convolved with the resolution model to determine the core parameters of the resolution model as shown in figure 5.7. In the signal model, the pole mass of the $\psi(2S)$ amplitude is set to the measured peak position of the $\psi(2S)$ peak in data: $m_{\psi(2S)} = 3685.66 \pm 0.05\,\mathrm{MeV}/c^2$.

In the final simultaneous fit to all three q^2 regions (described in chapter 7), the α parameter of the resolution functions in the mid and high q^2 regions are floated again, to minimize the dependence of the resolution model on the simulation[6].

The distortion of the angular distributions, due to the limited resolution of the detector, is considered to be negligible compared to the variations in the angular spectra, which

[6]The parameters η_l and η_u have been found to be largely redundant with the α parameter and are fixed to the best fit values obtained in the fit to simulation.

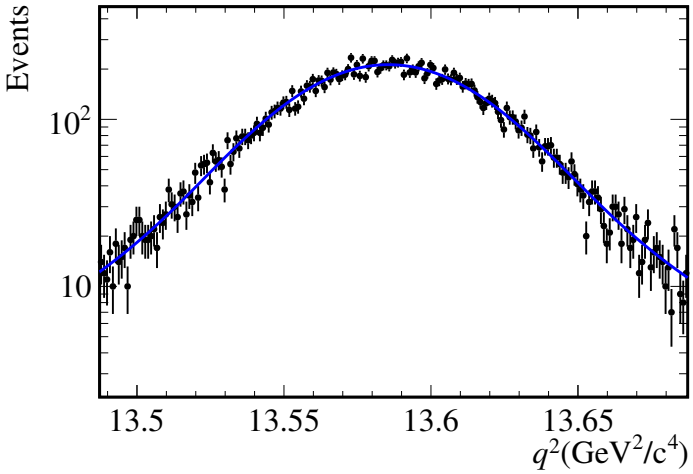


Figure 5.7: Fit to the core region of the $\psi(2S)$ peak in Run 1 data using a relativitic Breit Wigner convolved with the resolution model to determine the core parameters of the resolution model. Figure from Ref. [50].

contain no narrow peaks. The systematic uncertainty associated to the angular resolution is described in section 7.2.4.

Parameter	Value
σ_C	$4.40\times10^{-2} \pm 1.84\times10^{-3}$
σ_G	$2.77\times10^{-2} \pm 4.98\times10^{-4}$
f	$4.47\times10^{-1} \pm 4.70\times10^{-2}$
α_l	$-9.98\times10^{-1} \pm 1.75\times10^{-1}$
α_u	$1.04 \pm 1.76\times10^{-1}$
η_l	17.10 ± 7.74
η_u	11.07 ± 3.03

Table 5.6: The resolution parameters for the mid q^2 region for Run 1 data. The tail parameters are obtained from a fit to simulated $B^0 \to J/\psi K^{*0}$ events and the core parameters are obtained from fitting the core of the J/ψ peak in data.

Parameter	Value
σ_C	$5.73\times10^{-2} \pm 1.54\times10^{-2}$
σ_G	$3.03\times10^{-2} \pm 1.67\times10^{-3}$
f	$6.50\times10^{-1} \pm 1.29\times10^{-1}$
α_l	$-1.10 \pm 6.66\times10^{-2}$
α_u	$1.10 \pm 6.73\times10^{-2}$
η_l	$6.24 \pm 5.75\times10^{-1}$
η_u	11.94 ± 2.12

Table 5.7: The resolution parameters for the high q^2 region for Run 1 data. The tail parameters are obtained from a fit to simulated $B^0 \to \psi(2S)K^{*0}$ events and the core parameters are obtained from fitting the core of the $\psi(2S)$ peak in data.

6. Background

6.1 Introduction

A major experimental challenge for the analysis presented in this thesis is the modelling of the background. The final selection of $B^0 \to K^{*0}\mu^+\mu^-$ candidates contains about 10% combinatorial background events in the q^2 regions dominated by penguin $B^0 \to K^{*0}\mu^+\mu^-$ decays and about 2.5% in the regions dominated by $B^0 \to J/\psi\, K^{*0}$ and $B^0 \to \psi(2S)K^{*0}$. Thus, in order to extract information about the signal decays from data, it is crucial to correctly determine the background shapes in all dimensions as well as the background yield. The functional form of the background PDF is chosen empirically (see section 6.3 for the explicit expressions used) and the parameters are determined directly from data.

Modelling the background for the fit presented in this thesis is complex for several reasons:

(1) The high dimensionality of the $B^0 \to K^{*0}\mu^+\mu^-$ system: The fit is performed in the three helicity angles $\cos\theta_\ell$, $\cos\theta_K$, ϕ (see section 2.6), q^2, as well as $m^{\rm rec}_{B^0}$.

(2) The B^0 mass constraint of the reconstructed final state particles (see section 5.3.1): While greatly improving the q^2 resolution for signal events, it has the opposite effect for background events. It distorts their q^2 distribution and creates a strong correlation between q^2 and $m^{\rm rec}_{B^0}$. Thus, it is impossible to simply fit all events across the full $m^{\rm rec}_{B^0}$ range to determine the background shape. The solution for this is presented in section 6.2.

(3) Since the events across the full q^2 range, including the resonant regions, are used in the fit, there are three types of combinatorial background rather than one: $K^+\pi^-\mu^+\mu^-$, $J/\psi\,(\to \mu^+\mu^-)K^+\pi^-$, and $\psi(2S)(\to \mu^+\mu^-)K^+\pi^-$.

(4) The cut on $m_{K\mu\mu}$, used to remove background events from $B^+ \to K^+\mu^+\mu^-$ decays (see equation 5.11 in section 5.1.2), creates a gap in the phase space of the upper mass side-band, which – when ignored – causes a sizeable bias in the background parameterisation. This effect is discussed in detail in section 6.4 and a novel procedure for resolving the issue is presented.

There are other angular analyses of $B^0 \to K^{*0}\mu^+\mu^-$ decays [95], [96] and $B^+ \to K^+\mu^+\mu^-$ decays [49] where one or two out of these conditions exist, but never all four.[1]

Since the background parameters are extrapolated from the upper mass side-band into the signal region, potential correlations of the angles with $m^{\rm rec}_{B^0}$ are investigated and accounted for as described in section 6.7.

All the issues outlined above are dealt with in further detail in the rest of this chapter.

A feasibility study for including the lower mass side-band ($(5130 \leq m^{\rm rec}_{B^0} \leq 5210)$ MeV/c^2) in the background fit has been performed. The precision of the background parameters could potentially be improved by including the lower mass side-band, since it would turn the extrapolation of the background parameters along $m^{\rm rec}_{B^0}$ into an interpolation, and would also increase the number of background events used for determining the background shape. However, the lower mass side-band contains approximately 50% signal events, which have a q^2 distribution that is highly distorted in a non-trivial way. The distortion is due to the fact that the events are mass constrained to the center of the lower mass side-band (see section 6.2). Consequently, the systematic uncertainty introduced by modelling the distorted signal contribution outweighs the improvements in the statistical precision of the background parameters.

6.2 Strategy

The fit of the signal model[2] is performed in the range $(5239.58 \leq m^{\rm rec}_{B^0} \leq 5319.58)$ MeV/c^2 indicated with the green area in figure 6.1, which shows the $m^{\rm rec}_{B^0}$ distribution of all selected candidates in the merged Run 1 and Run 2 data. The signal region is chosen such that contributions from $B^0_s \to J/\psi\, K^{*0}$ and $B^0_s \to \psi(2S)K^{*0}$ decays, visible as a hump at around $m^{\rm rec}_{B^0} \approx 5375$ MeV/c^2, and partially reconstructed B^0 decays at $m^{\rm rec}_{B^0} \lesssim 5130$ MeV/c^2 are fully avoided. About 90% of the $B^0 \to K^{*0}\mu^+\mu^-$ signal events are contained in this region.

The upper mass side-band (SB) starting at $m^{\rm rec}_{B^0} \geq 5440$ MeV/c^2 is used to determine a parameterisation of the background events in the signal region. The upper mass side-band is made up purely of combinatorial background events which can be separated into three different contributions: Fully combinatorial $K^+\pi^-\mu^+\mu^-$ events, and resonant combinatorial

[1] Strictly speaking the $B^+ \to K^+\mu^+\mu^-$ veto was used in both analyses but did not cause a sizeable bias due to different fit strategies and smaller datasets.

[2] The signal model is described in chapter 3 and the fit of the full signal and background model to Run 1 data is shown in chapter 7.

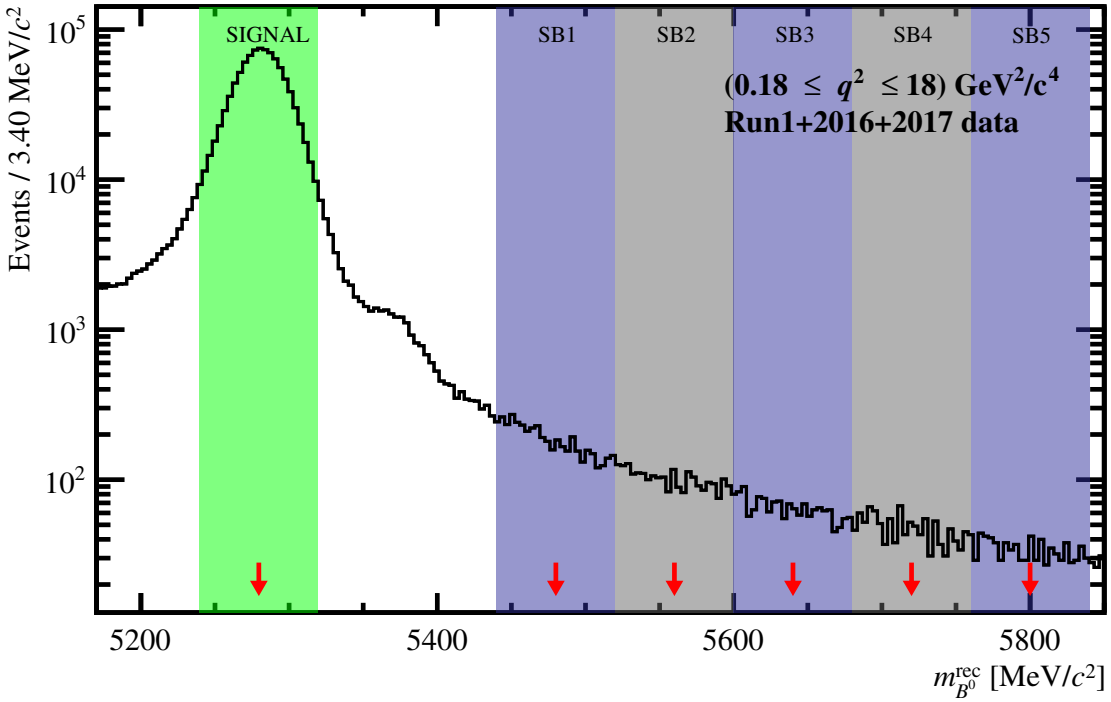


Figure 6.1: Invariant mass distribution of the $B^0 \to K^{*0}\mu^+\mu^-$ candidates in the merged Run 1 and Run 2 data. The green area indicates the signal region. The blue and grey areas indicate the side-band regions. The red arrows show the mass constraint points for each region. The $B_s^0 \to J/\psi K^{*0}$ and $B_s^0 \to \psi(2S)K^{*0}$ decays are visible as a hump at $m_{B^0}^{\text{rec}} \approx 5375$ MeV/c^2 and are fully avoided in the signal region as well as in the side-band regions.

background events (henceforth referred to as resonant background) containing charmonia resonances paired with random kaons and pions: $J/\psi(\to \mu^+\mu^-)K^+\pi^-$ and $\psi(2S)(\to \mu^+\mu^-)K^+\pi^-$.

The three background contributions have independent shapes in all dimensions and can be most easily distinguished via the q^2 distribution (see figure 6.2) where the two resonant backgrounds peak at the J/ψ and $\psi(2S)$ mass respectively, while the fully combinatorial background spans the full q^2 range.

Effect of the mass constraint

As described in section 5.3.1, the 4-momenta of the final state particles of all $B^0 \to K^{*0}\mu^+\mu^-$ candidates are mass constrained to the B^0 mass in order to improve the q^2 resolution of the signal events. However, for background events this has the opposite effect. The q^2 distribution of the background is distorted since the mass constraint shifts the momenta

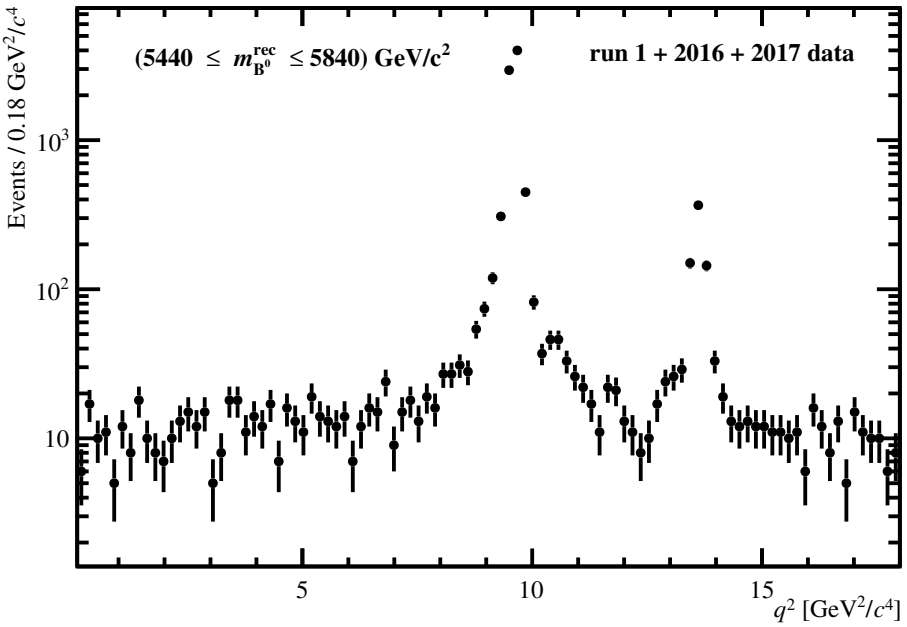


Figure 6.2: Unconstrained q^2 distribution of the background events in the upper mass side-band $(5440 \leq m_{B^0}^{\text{rec}} \leq 5480)$ MeV/c^2 in the merged Run 1 and Run 2 dataset.

of the final state particles to best match the invariant mass with the B^0 mass, despite the fact that the final state particles are in fact the not decay products of a B^0. This effect is illustrated in figure 6.3 where the q^2 distributions of the events in several sub-regions of the upper mass side-band are shown without mass constraint (black points) and after constraining the events to the B^0 mass (red points). The large charmonia peaks are widened and shifted by the mass constraint. The distortion becomes stronger as a function of $m_{B^0}^{\text{rec}}$, creating a correlation of the q^2 shape of the background with $m_{B^0}^{\text{rec}}$. Due to this effect, it is impossible to parameterise the background events across the full $m_{B^0}^{\text{rec}}$ range in a simple fit, and it is not straight forward to use events in the upper mass side-band to determine the shape of the background in the signal region.

To solve the issue created by the mass constraint, the upper mass side-band is segmented into five sub-regions (named SB1 - SB5), indicated with the blue and grey areas in figure 6.1. The $m_{B^0}^{\text{rec}}$ ranges of SB1 - SB5 are given in table 6.1. Each SB region has the same width of 80 MeV/c^2 as the signal region. The events in each SB region are mass constrained to the central $m_{B^0}^{\text{rec}}$-value of the respective region (marked by red arrows in figure 6.1). Since the SB regions have the same width as the signal region, the SB constraints yield the same level of distortion of the q^2 shape of the background as the B^0-mass constraint

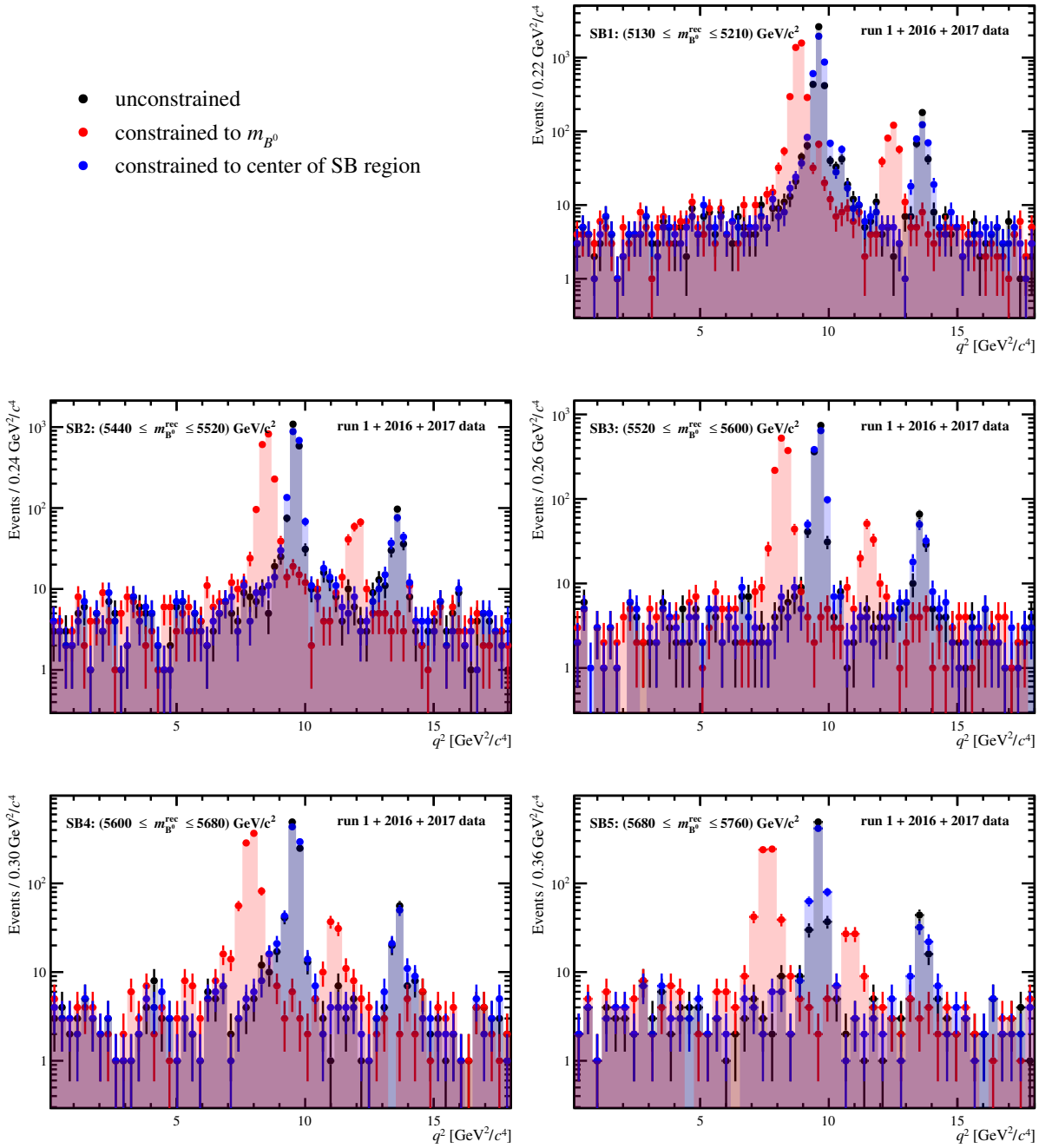


Figure 6.3: The q^2 distributions of the events in all upper mass SB regions without mass constraint (black), constrained to m_{B^0} (red), and constrained to the center of the respective SB region (green).

in the signal region. The q^2 distribution of the SB constrained events is shown with the blue points in figure 6.3. The charmonia peaks are widened compared to the peaks in unconstrained q^2. In the context of the background fit, q^2 refers to the SB constrained q^2

	$m_{B^0}^{\text{rec}}$ range (MeV/c^2)	constraint point (MeV/c^2)	scaled distance to m_{B^0}
Signal	[5239.58, 5319.58]	5279.58	0
SB1	[5440, 5520]	5480	∼0.3851
SB2	[5520, 5600]	5560	∼0.5388
SB3	[5600, 5680]	5640	∼0.6926
SB4	[5680, 5760]	5720	∼0.8463
SB5	[5760, 5840]	5800	1.0

Table 6.1: Definition of the $m_{B^0}^{\text{rec}}$ regions used throughout this thesis. The second column contains the $m_{B^0}^{\text{rec}}$ values to which the events in each region are mass constrained to. The third column gives the distance of the centre of each region to the true mass of the B^0, scaled by the furthest distance.

(using the centre of the respective SB region) hereafter, unless specified otherwise.
The background parameters are determined by simultaneously fitting the five SB regions and extrapolating the background parameters into the signal region.

6.3 Background parameterisation

Since the three background contributions (fully combinatorial background, J/ψ-background, and $\psi(2S)$-background) have different kinematic behaviours, it cannot be assumed that the shapes of their distributions are the same. Therefore, each of the three background contributions is parameterised by an independent PDF. The functional form in each dimension of each PDF is chosen empirically.

The fully combinatorial background is parameterised as

$$\mathcal{P}_{\text{com}}(\vec{\Omega}, q^2, m_{B^0}^{\text{rec}}) = \mathcal{C}_{\cos\theta_\ell}(\cos\theta_\ell) \times \mathcal{C}_{\cos\theta_K}(\cos\theta_K) \times \mathcal{C}_{\phi}(\phi) \times \mathcal{W}(q^2) \times \mathcal{M}(m_{B^0}^{\text{rec}}), \quad (6.1)$$

where $\mathcal{C}_i$ are Chebychev polynomials of second order. $\mathcal{W}(q^2)$ is a two parameter Weibull function given by [97]

$$\mathcal{W}(q^2) = a \cdot b \cdot (b \cdot q^2)^{a-1} \cdot \mathrm{e}^{-(b\cdot q^2)^a}, \quad (6.2)$$

where a and b are the parameters of the function which are hereafter referred to as $a_{\text{comb}}^{q^2}$ and $b_{\text{comb}}^{q^2}$. Finally, $\mathcal{M}(m_{B^0}^{\text{rec}})$ is a falling exponential.

The resonant backgrounds are parameterised using the same functional forms (but with fully independent parameters) except for the q^2 dimension, which is are described by a Crystal Ball function $\mathcal{CB}(q^2)$ [98]. Thus, the PDFs for the resonant backgrounds are

$$\mathcal{P}_{J/\psi\,/\psi(2S)}(\vec{\Omega}, q^2, m) = \mathcal{C}_{\cos\theta_\ell}(\cos\theta_\ell) \times \mathcal{C}_{\cos\theta_K}(\cos\theta_K) \times \mathcal{C}_\phi(\phi) \times \mathcal{CB}(q^2) \times \mathcal{M}(m). \quad (6.3)$$

As described in section 5.3, the signal fit is performed simultaneously in three regions in q^2. Therefore, the background is also parameterised in those q^2 regions. The low q^2 region contains only fully combinatorial events, whereas the mid and the high q^2 regions also contain contributions from J/ψ and $\psi(2S)$ background events respectively. Therefore, the total PDF describing the background is given by:

$$\mathcal{P}_{\text{bkg}}(\vec{\Omega}, q^2, m) = \begin{cases} \mathcal{P}_{\text{com}}(\vec{\Omega}, q^2, m) & \text{in low } q^2 \\ f_{J/\psi}\mathcal{P}_{J/\psi}(\vec{\Omega}, q^2, m) + (1 - f_{J/\psi})\mathcal{P}_{\text{com}}(\vec{\Omega}, q^2, m) & \text{in mid } q^2 \\ f_{\psi(2S)}\mathcal{P}_{\psi(2S)}(\vec{\Omega}, q^2, m) + (1 - f_{\psi(2S)})\mathcal{P}_{\text{com}}(\vec{\Omega}, q^2, m) & \text{in high } q^2, \end{cases} \quad (6.4)$$

where $f_{J/\psi}$ and $f_{\psi(2S)}$ are the fractions of resonant over fully combinatorial background in the mid and high q^2 region respectively. The most power for determining these fractions comes from the q^2 dimensions.

Each Chebychev polynomial has two parameters, the mass dimension is described by one parameter for each component, the Weibull function has two parameters and each Crystal Ball function has four parameters. Also, there are the two background fractions $f_{J/\psi}$ and $f_{\psi 2s}$. Thus, the background events are described by 33 parameters in total.
The background parameters are determined by simultaneously fitting the three q^2 regions and five $m^{\text{rec}}_{B^0}$ SB regions. Thus, there are 15 normalisation regions for the background fit. All parameters are shared across the SB regions. The parameters of $\mathcal{P}_{\text{com}}$ are also shared across the q^2 regions, whereas the parameters of $\mathcal{P}_{\text{J}/\psi}$ and $\mathcal{P}_{\psi 2\text{s}}$ are only determined in the mid and high q^2 region respectively. Table 6.2 shows all background parameters, ordered by background component and dimension, indicating in which normalisation region each parameter is determined as well as which parameters are fixed in the fit.
The parameter $b^{q^2}_{\text{comb}}$ is fixed to 0.0001, since the fit favours very small values for this parameter, but becomes unstable at the lower limit of 0. The parameters tail parameters $\eta^{q^2}_{J/\psi}$ and $\eta^{q^2}_{\psi(2S)}$ are fixed because they are highly correlated with $\alpha^{q^2}_{J/\psi}$ and $\alpha^{q^2}_{\psi(2S)}$ respectively, causing the fit to be unstable. The values $\eta^{q^2}_{J/\psi} = 0.8$ and $\eta^{q^2}_{\psi(2S)} = 50$ have been found to yield good fits to the data[3]. Furthermore, the parameter $\alpha^{q^2}_{\psi(2S)}$ is fixed to 0.6, since the fit

[3]Due to the correlation with the α parameters the exact values of $\eta^{q^2}_{J/\psi}$ and $\eta^{q^2}_{\psi(2S)}$ are not important

has very little sensitivity to it and floating it causes an unstable fit.[4]

The background PDF, described above, assumes factorisation of all dimensions. This holds true for the angles and q^2 as shown in section 6.6. However, there are dependencies of some of the background parameters on $m^{\text{rec}}_{B^0}$, which are taken into account as discussed in section 6.7.

[4]Due to the large value of $\eta^{q^2}_{\psi(2S)}$, the Crystal Ball function of the $\psi(2S)$ effectively becomes a Gaussian.

component	dim.	name	fixed	q^2 region		
				low	mid	high
fully comb.	$\cos\theta_\ell$	$a^{\mathrm{L}}_{\mathrm{comb}}$		x	x	x
		$b^{\mathrm{L}}_{\mathrm{comb}}$				
	$\cos\theta_K$	$a^{\mathrm{K}}_{\mathrm{comb}}$				
		$b^{\mathrm{K}}_{\mathrm{comb}}$				
	ϕ	a^{ϕ}_{comb}				
		b^{ϕ}_{comb}				
	q^2	$a^{q^2}_{\mathrm{comb}}$				
		$b^{q^2}_{\mathrm{comb}}$	x			
	$m^{\mathrm{rec}}_{B^0}$	s^{m}_{comb}				
J/ψ	$\cos\theta_\ell$	$a^{\mathrm{L}}_{J/\psi}$			x	
		$b^{\mathrm{L}}_{J/\psi}$				
	$\cos\theta_K$	$a^{\mathrm{K}}_{J/\psi}$				
		$b^{\mathrm{K}}_{J/\psi}$				
	ϕ	$a^{\phi}_{J/\psi}$				
		$b^{\phi}_{J/\psi}$				
	q^2	$\mu^{q^2}_{J/\psi}$				
		$\sigma^{q^2}_{J/\psi}$				
		$\alpha^{q^2}_{J/\psi}$				
		$\eta^{q^2}_{J/\psi}$	x			
	$m^{\mathrm{rec}}_{B^0}$	$s^{m}_{J/\psi}$				
$\psi(2S)$	$\cos\theta_\ell$	$a^{\mathrm{L}}_{\psi(2S)}$				x
		$b^{\mathrm{L}}_{\psi(2S)}$				
	$\cos\theta_K$	$a^{\mathrm{K}}_{\psi(2S)}$				
		$b^{\mathrm{K}}_{\psi(2S)}$				
	ϕ	$a^{\phi}_{\psi(2S)}$				
		$b^{\phi}_{\psi(2S)}$				
	q^2	$\mu^{q^2}_{\psi(2S)}$				
		$\sigma^{q^2}_{\psi(2S)}$				
		$\alpha^{q^2}_{\psi(2S)}$	x			
		$\eta^{q^2}_{\psi(2S)}$	x			
	$m^{\mathrm{rec}}_{B^0}$	$s^{m}_{\psi(2S)}$				
fractions		$f_{J/\psi}$			x	
		$f_{\psi(2S)}$				x

Table 6.2: The parameters of the background PDF, ordered by background component and dimension. Also given are the q^2 regions in which each parameter is determined. All parameters are shared across all SB windows. Parameters that are permanently fixed in the side-band fit are marked in the fourth column.

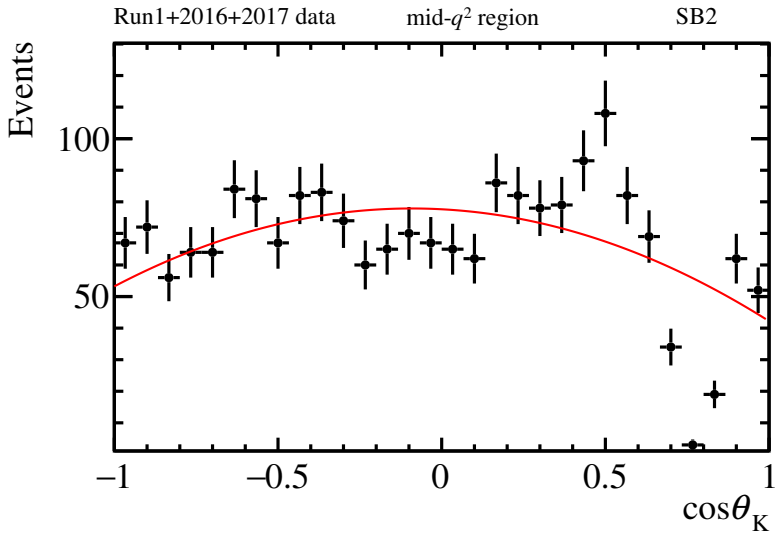


Figure 6.4: $\cos\theta_K$ distribution of the events in the mid q^2 and SB2 region in Run 1+Run 2 data. The red line shows a second order polynomial fit which cannot describe the data due the drop-off of the data at $\cos\theta_K \geq 0.5$ which is caused by the $K^+\mu^+\mu^-$ veto.

6.4 Effect of the $B^+ \to K^+\mu^+\mu^-$ veto on the upper mass side-band

As described in section 5.1, the process of selecting $B^0 \to K^{*0}\mu^+\mu^-$ events includes a veto where all events with

$$\begin{aligned} m_{K\pi\mu\mu} &> 5380\,\text{MeV}/c^2 \text{ and} \\ m_{K\mu\mu} &\in [5220, 5340]\,\text{MeV}/c^2 \end{aligned} \tag{6.5}$$

are removed from the analysis. This $K\mu\mu$-veto fully suppresses background events from $B^+ \to K^+\mu^+\mu^-$ decays which – when paired with random pions – would otherwise populate the upper mass side-band and greatly complicate the parameterisation of the background.

However, the veto also removes all combinatorial background events in a certain region of the final state phase space, which effectively changes the shape of the $\cos\theta_K$, q^2, and $m^{\text{rec}}_{B^0}$ distributions as well as the relative yields of the fully combinatorial and resonant backgrounds in the upper mass side-band.

Figure 6.4 shows the $\cos\theta_K$ distribution of the events in the mid q^2 range in SB2 as well as the projection of a second order Chebychev polynomial fit. The strong drop-off in the distribution at $\cos\theta_K \geq 0.5$ is caused by the veto. The drop-off cannot be described

by the 2nd order polynomial and causes the parametrisation to be wrong with respect to the true underlying distribution. Furthermore, this effect is not present in the signal region and the extrapolation of the background parameters from the upper side-band into the signal region would thus be invalid.

To solve this, the background PDF is adjusted such that the fit effectively "ignores" the affected phase space, as described in the following.

6.4.1 Phase space Monte Carlo simulation

Since $m_{K\mu\mu}$ is not a variable that is included in the fit, the definition of the $K^+\mu^+\mu^-$ veto needs to be translated from a cut in $m_{K\mu\mu}$ into a cut in ($\vec{\Omega}$, q^2, $m^{\text{rec}}_{B^0}$)-space, in order to adjust the PDF for the effect of the veto.

A phase-space Monte Carlo simulation is used to this end. $B^0 \to K^+\pi^-\mu^+\mu^-$ decays are simulated according to a phase-space model for varying masses of the B^0 meson: Two million events are generated for each of 50 steps in a range of $(5440 \leq m^{\text{rec}}_{B^0} \leq 5860)$ MeV/c^2. Events are then selected within the $m_{K\pi}$ window which is used throughout the analysis (equation 5.1). Subsequently, events are selected using the reverse $K\mu\mu$-veto condition (equation 6.5) - i.e. only the events that are usually vetoed are kept.

Figure 6.5 shows the $\cos\theta_\ell$, $\cos\theta_K$, ϕ, q^2, $m^{\text{rec}}_{B^0}$, and $m_{K\pi}$ projections of the ratio of the events selected by the veto divided by all events within the $m_{K\pi}$ window. The events selected by the $K\mu\mu$-veto are evenly distributed along the $\cos\theta_\ell$ and ϕ dimensions, indicating that the veto has no effect in these dimensions. However, there is a clear dependence of the number of selected events on $\cos\theta_K$, q^2, $m^{\text{rec}}_{B^0}$ and $m_{K\pi}$.

The strongest dependence of the number of events selected by the veto is on $\cos\theta_K$. Most of the $K\mu\mu$-veto selected events concentrate in a narrow window in $\cos\theta_K$ of about $0.5 < \cos\theta_K < 1.0$, which creates the clear drop-off in the $\cos\theta_K$ distribution observed in data. The effect on the q^2 distribution is smaller and over a wide q^2 range. With rising q^2, more and more events are vetoed, which when ignored in a fit would effectively pull the PDF down at high q^2 with respect to the true distribution. Only the fully combinatorial background is affected by this since the resonant backgrounds are concentrated in narrow q^2 windows. The effect of the veto on the $m^{\text{rec}}_{B^0}$ distribution is the same for all three background components, and causes a pull on the slope parameters of the background PDF starting at around $m^{\text{rec}}_{B^0} \approx 5.5$ GeV/c^2. The effect on the $m_{K\pi}$ shape is small but may still be relevant for analyses that fit for $m_{K\pi}$. In this work $m_{K\pi}$ is not fitted for, and thus the 4 dimensional veto-affected region can be projected down into a 3 dimensional

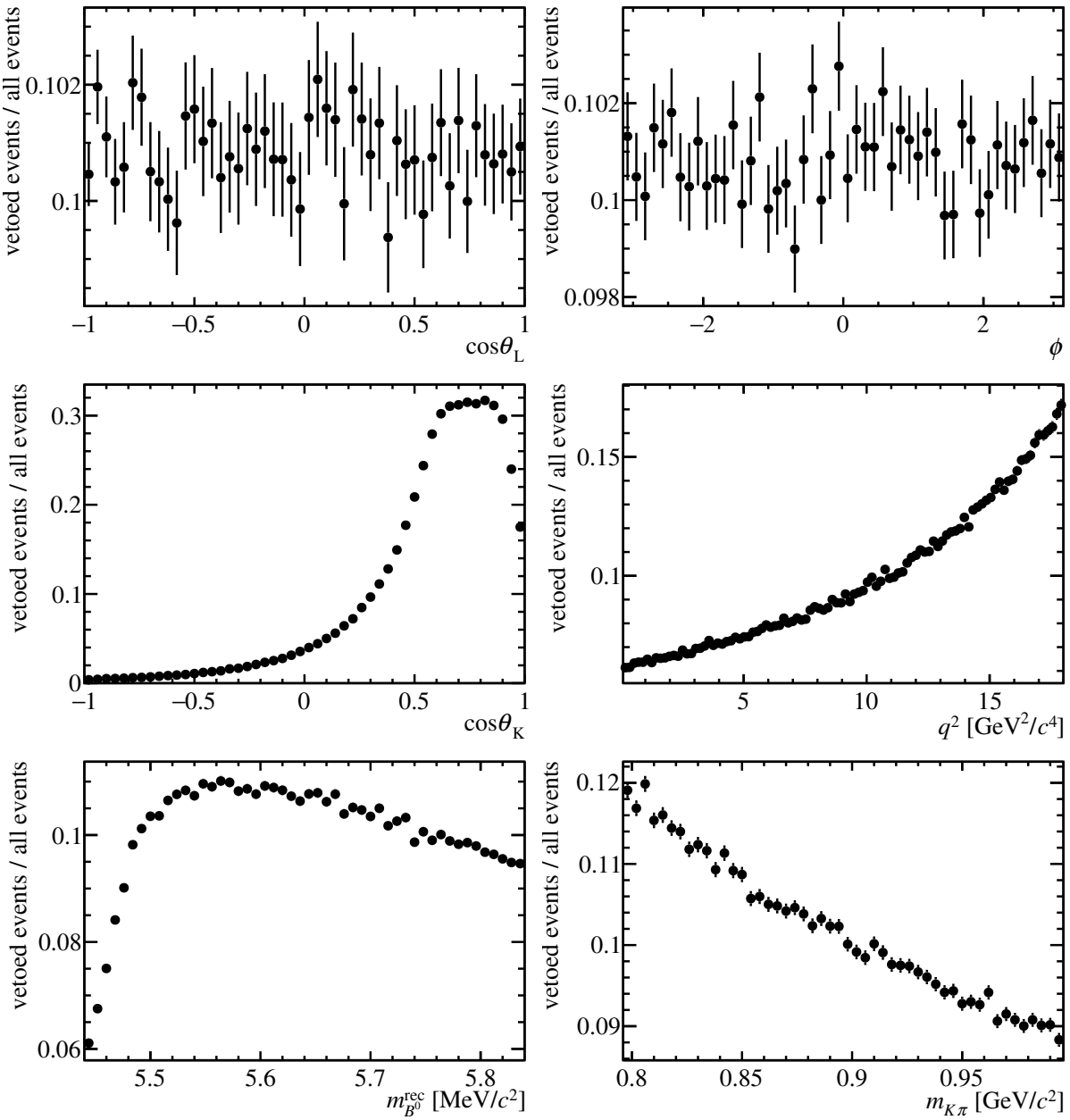


Figure 6.5: Ratios of the distributions of phase space toy events selected by the $K\mu\mu$-veto condition divided by all phase space toy events. The graphs are zero-suppressed on the y-axis except for the graph of $\cos\theta_K$.

affected region in ($\cos\theta_K$, q^2, $m^{\text{rec}}_{B^0}$).

The correlations between $\cos\theta_K$, q^2, and $m^{\text{rec}}_{B^0}$ lead to a non-trivial shape of the affected ($\cos\theta_K$, q^2, $m^{\text{rec}}_{B^0}$)-phase space. One way to illustrate this is to think of the veto as a cut on $\cos\theta_K$, where the removed $\cos\theta_K$ range depends on q^2 and $m^{\text{rec}}_{B^0}$. Figure 6.6 shows the $m^{\text{rec}}_{B^0}$

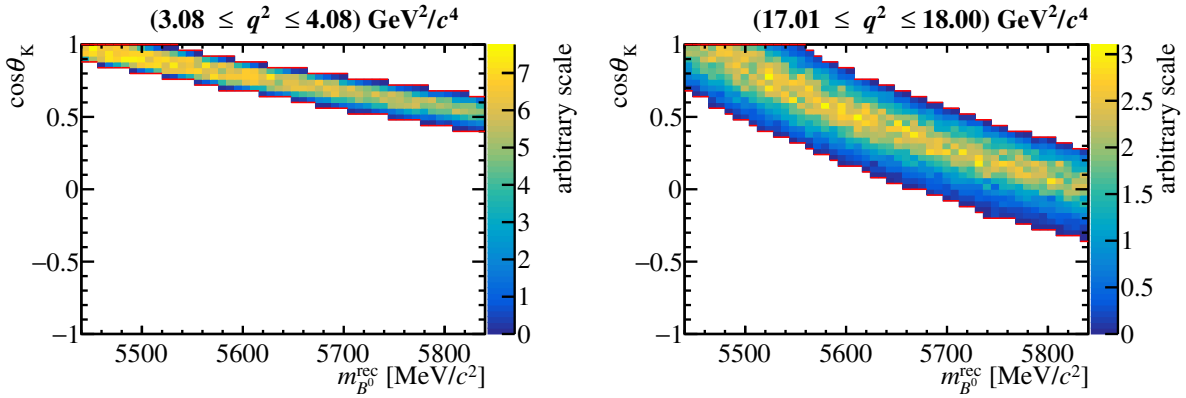


Figure 6.6: $(m^{\text{rec}}_{B^0}, \cos\theta_K)$-projections of the $K\mu\mu$-veto affected phase space in different bins in q^2. The red lines indicate the low and high $\cos\theta_K$ values of the $\cos\theta_K$-window which is removed by the veto. The full set of plots is shown in appendix A.1

vs $\cos\theta_K$ distribution of all phase-space toy events selected by the veto within two different q^2 ranges: $(3.08 \leq q^2 \leq 4.08)$ GeV$^2/c^4$ (left) and $(17.01 \leq q^2 \leq 18)$ GeV$^2/c^4$ (right). The red lines indicate the $\cos\theta_K$-window which is affected by the veto in the given $m^{\text{rec}}_{B^0}$ bin for the given q^2 range. The size and shape of the veto-affected area in $(m^{\text{rec}}_{B^0}, \cos\theta_K)$ greatly depends on q^2. The complete set of plots showing the affected $(m^{\text{rec}}_{B^0}, \cos\theta_K)$ areas in bins of q^2 as well as the affected $(m^{\text{rec}}_{B^0}, q^2)$-area in bins of $\cos\theta_K$ is shown in appendix A.1.

In order to obtain a well defined description of the veto-affected $(\cos\theta_K,\ q^2,\ m^{\text{rec}}_{B^0})$-volume, a three dimensional histogram is filled with all phase space events that are selected by the veto. This histogram is shown in figure 6.7. The size of each box is proportional to the number of entries. The phase space of each bin that has at least one entry is considered affected by the veto. This non parametric definition of the veto-affected region in $(\cos\theta_K,\ q^2,\ m^{\text{rec}}_{B^0})$-space is used in the side-band fit to adjust the PDF for the effect of the veto as described in section 6.4.2.

Parameter choices for the Monte Carlo Method and smoothing algorithm

Using a Monte Carlo simulation to translate the $m_{K\mu\mu}$ cut into a cut in $(\cos\theta_K,\ q^2,\ m^{\text{rec}}_{B^0})$, and then using a non parametric description of the veto-affected $(\cos\theta_K,\ q^2,\ m^{\text{rec}}_{B^0})$-region, comes with a few details to consider:

A finer binning of the $(\cos\theta_K,\ q^2,\ m^{\text{rec}}_{B^0})$-histogram used to describe the affected phase space (see figure 6.7) yields a more detailed approximation of the affected volume. However,

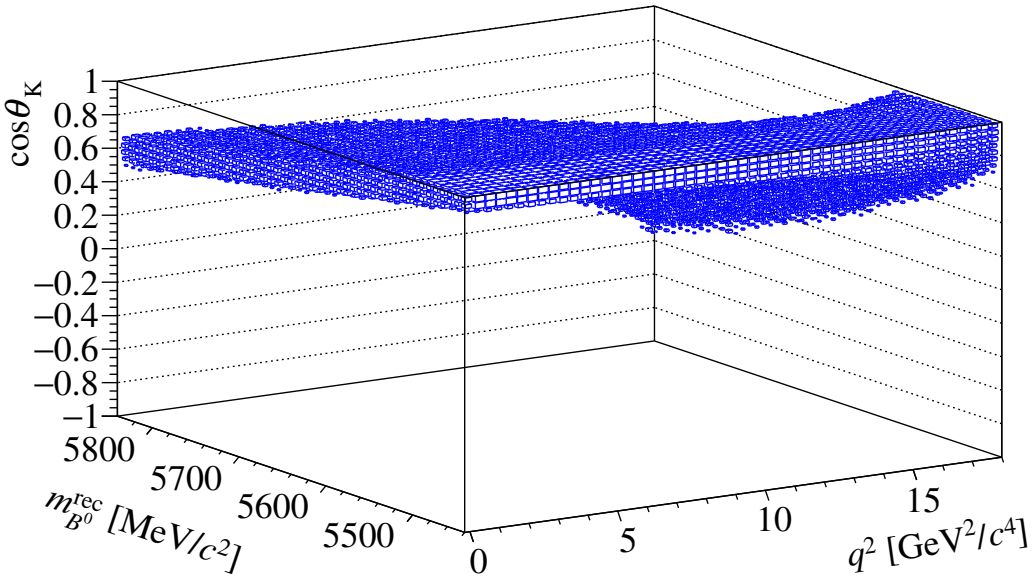


Figure 6.7: $(\cos\theta_K,\ q^2,\ m^{\rm rec}_{B^0})$ distribution of the phase space simulated events selected by the $B^+ \to K^+\mu^+\mu^-$ veto. The size of each box is proportional to the number of entries. The phase space of each bin that has at least one entry is considered affected by the veto.

a finer binning of the cut histogram increases the computing time of the side-band fit, since the calculation of the normalisation of the PDF scales with the number of affected bins (see section 6.4.2). Furthermore, a finer binning of the cut histogram also increases the probability for any bin contained in the affected volume to have zero entries - which would falsely classify that bin as unaffected. This can be offset by generating more phase space events for every step in $m^{\rm rec}_{B^0}$. However, due to the fact that a single entry in a phase space bin is enough for the bin to be considered as affected by the veto, a higher number of events also leads to an effective widening of the cut volume. This is demonstrated in table 6.3 where the fraction of phase space considered as 'affected by veto' is given for different numbers of events generated per step in $m^{\rm rec}_{B^0}$.

A too large cut-volume leads to additional events being unnecessarily removed from the upper mass side-band which were not removed by the actual $K^+\mu^+\mu^-$-veto, thereby decreasing the statistical precision of the side-band fit. On the other hand, in order to be sure that no affected region in phase space is missed, it is actually preferential to slightly overestimate the affected region.

Taking all of the above into account, a binning of $50 \times 50 \times 50$ is chosen for the $(\cos\theta_K,$

number of events generated per step in $m^{\text{rec}}_{B^0}$	fraction of phase space considered as 'affected by veto'
0.1 million	13.55%
0.5 million	15.01%
1 million	15.47%
2 million	15.77%
3 million	15.98%
4 million	16.1%

Table 6.3: Dependence of the size of the ($\cos\theta_K$, q^2, $m^{\text{rec}}_{B^0}$)-cut on the number of events generated in the phase space simulation.

q^2, $m^{\text{rec}}_{B^0}$) histogram and 2 million $B^0 \to K^+\pi^-\mu^+\mu^-$ phase space events are generated for each bin in $m^{\text{rec}}_{B^0}$. Furthermore, an algorithm has been developed which scans the 3D histogram and identifies the outer edges of the out-most affected bins in every dimension (the red lines in figure 6.6 show the outer edges in $\cos\theta_K$). Bins that have no entry but are contained between bins that are affected are filled to ensure they meet the 'affected by veto' condition. The algorithm also ensures that the dimensions are unanimous in their classification of every bin. This means that, if a bin is between the lowest and highest affected bin in $\cos\theta_K$, it must also between the lowest and highest bin along the q^2 and $m^{\text{rec}}_{B^0}$ direction. Due to the curved shape of the affected volume this is not automatically the case for every bin.

In order to demonstrate that the ($\cos\theta_K$, q^2, $m^{\text{rec}}_{B^0}$)-region determined via the method described above does indeed model the region affected by the veto, figure 6.8 compares the ($m^{\text{rec}}_{B^0}$, $\cos\theta_K$)-distribution of events which were rejected by the $K\mu\mu$-veto in the range $(9 \leq q^2 \leq 10)$ GeV$^2/c^4$ in Run 1 data (colored histogram) with the ($m^{\text{rec}}_{B^0}$, $\cos\theta_K$)-area which is marked as affected by the phase space simulation (red boxes). All events rejected by the veto are within the area which is marked as affected. The area that is considered as affected is indeed a bit larger than the area occupied by the vetoed events but as stated earlier this is preferred to ensure no affected phase space is missed.

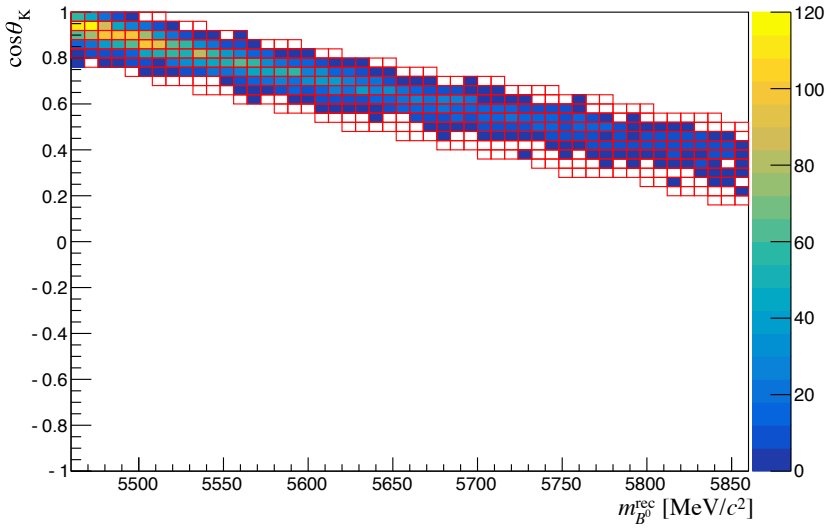


Figure 6.8: $(m^{\text{rec}}_{B^0}, \cos\theta_K)$ distribution of the events rejected by the veto in the range $(9 \leq q^2 \leq 10)$ GeV$^2/c^4$ in Run 1 data. Red boxes: $(m^{\text{rec}}_{B^0}, \cos\theta_K)$-area marked as affected by the $K^+\mu^+\mu^-$-veto via the phase space simulation in same q^2 range.

6.4.2 Correcting for the veto in the side-band fit

The 3D histogram shown in figure 6.7 is used to place an explicit cut, removing all events in the veto-affected ($\cos\theta_K$, q^2, $m^{\text{rec}}_{B^0}$)-volume. Hereafter, this cut is referred to as ($\cos\theta_K$, q^2, $m^{\text{rec}}_{B^0}$)-cut. About 6% of the side-band events that passed all selection criteria including the original $K^+\mu^+\mu^-$-veto are removed by this cut.
Furthermore, the ($\cos\theta_K$, q^2, m)-histogram is used to adjust the background PDF by setting the PDF to zero in the cut volume:

$$\mathcal{P}^{\text{with-cut}}_{\text{bkg}}(\vec{\Omega}, q^2, m^{\text{rec}}_{B^0}) = \begin{cases} 0 & \text{for}(\cos\theta_K,\ q^2,\ \text{and}\, m^{\text{rec}}_{B^0})\ \text{in cut volume} \\ \mathcal{P}_{\text{bkg}}(\vec{\Omega}, q^2, m^{\text{rec}}_{B^0}) & \text{otherwise,} \end{cases} \tag{6.6}$$

where $\mathcal{P}_{\text{bkg}}(\vec{\Omega}, q^2, m^{\text{rec}}_{B^0})$ is given by equation 6.4. Setting the PDF to zero in the cut volume, ensures that the fit does not 'expect' any events in this volume and is thus not affected by the absence of events there.

When calculating the normalisation of $\mathcal{P}^{\text{with-cut}}_{\text{bkg}}$, the fact that some of the phase space is missing needs to be taken into account. Due to the binned definition of the cut volume, the normalisation fo the background PDF in each q^2 region (low, mid, and high q^2) and

each SB region of the simultaneous side-band fit (SB1 - SB5) can be calculated with

$$\int \mathcal{P}_{\text{bkg}}^{\text{with-cut}} = \int_{-1}^{1}\int_{-1}^{1}\int_{-\pi}^{\pi}\int_{q^2_{\text{min}}}^{q^2_{\text{max}}}\int_{m_{\text{min}}}^{m_{\text{max}}} \mathcal{P}_{\text{bkg}}^{\text{without-cut}}(\vec{\Omega}, q^2, m)\,\mathrm{d}m\mathrm{d}q^2\mathrm{dcos}\theta_K\mathrm{d}\phi\mathrm{dcos}\theta_l$$
$$- \sum_i \int_{-1}^{1}\int_{-1}^{1}\int_{\cos\theta_K{}^i_{\text{min}}}^{\cos\theta_K{}^i_{\text{max}}}\int_{q^{2,i}_{\text{min}}}^{q^{2,i}_{\text{max}}}\int_{m^i_{\text{min}}}^{m^i_{\text{max}}} \mathcal{P}_{\text{bkg}}^{\text{without-cut}}(\vec{\Omega}, q^2, m)\,\mathrm{d}m\mathrm{d}q^2\mathrm{dcos}\theta_K\mathrm{d}\phi\mathrm{dcos}\theta_l, \tag{6.7}$$

where $[q^2_{\text{min}}, q^2_{\text{max}}]$ and $[m_{\text{min}}, m_{\text{max}}]$ are the fit ranges in q^2 and $m^{\text{rec}}_{B^0}$ of the given q^2 and SB region. The sum indicates a sum over all bins i which are affected by the veto, while $[\cos\theta_K{}^i_{\text{min}}, \cos\theta_K{}^i_{\text{max}}]$, $[q^{2,i}_{\text{min}}, q^{2,i}_{\text{max}}]$, and $[m^i_{\text{min}}, m^i_{\text{max}}]$ are the $\cos\theta_K$, q^2, and $m^{\text{rec}}_{B^0}$ ranges of bin i.

Correction of the background fractions

With the PDF adjusted as described above, the fit recovers the correct shapes of the background distributions unaffected by the veto as demonstrated in the following section (section 6.4.3). However, the best fit values for the PDF fractions $f_{J/\psi}$ and $f_{\psi 2s}$ (see equation 6.4) would still be the relative yields after the veto, which are different to the true relative yields before the veto. They are different, because the resonant and the fully combinatorial backgrounds lose different relative amounts of events to the veto, since the shapes of their distributions are different - most notably in q^2 where the resonant backgrounds peak at the J/ψ and $\psi(2S)$ masses squared, whereas the fully combinatorial background is distributed much more evenly across q^2.
The effect of the veto on the background fractions is corrected for during the fit by calculating the fraction of lost events for each background contribution:

$$R_\lambda = \frac{\int \mathcal{P}_\lambda^{without-cut}}{\int \mathcal{P}_\lambda^{with-cut}}, \tag{6.8}$$

where λ stands for fully combinatorial (com), J/ψ or $\psi(2S)$ and the integrals denote 5D integrals over $(\vec{\Omega}, q^2, m^{\text{rec}}_{B^0})$. Using equation 6.8, $f_{J/\psi}$ and $f_{\psi 2s}$ in equation 6.4 are replaced with

$$f^a_\alpha = \frac{f^b_\alpha R_{\text{com}}}{f^b_\alpha R_{\text{com}} + (1 - f^b_\alpha) R_\alpha}. \tag{6.9}$$

Here, α stands for either J/ψ or $\psi(2S)$. f^a_α is the fraction after cut and f^b_α is the fraction before cut. When replacing $f_{J/\psi}$ and $f_{\psi 2s}$ in equation 6.4 with $f^a_{J/\psi}$ and $f^a_{\psi(2S)}$, the total PDF describes the events after the veto, but the fit determines

the floating parameters $f^b_{J/\psi}$ and $f^b_{\psi(2S)}$, i.e. the fractions before the veto. The fractions before the veto are then extrapolated into the signal region as described in section 6.7.

6.4.3 Demonstration of the efficacy of the veto correction

The efficacy of the method described above is demonstrated with a toy study. Events are generated according to $\mathcal{P}^{\text{without-cut}}_{\text{bkg}}$ (i.e. equation 6.4). The resulting $\cos\theta_K$, q^2, and $m^{\text{rec}}_{B^0}$ distributions are shown by the red points in figure 6.9. All events in the veto affected ($\cos\theta_K$, q^2, $m^{\text{rec}}_{B^0}$)-volume are then removed from the toy sample, yielding the distributions shown by the blue points.
The sample after the cut is then fitted with $\mathcal{P}^{\text{with-cut}}_{\text{bkg}}$, as described above. Projections of $\mathcal{P}^{\text{with-cut}}_{\text{bkg}}$ are calculated in the bins of the ($\cos\theta_K$, q^2, $m^{\text{rec}}_{B^0}$)-cut histogram and added together while scaling each projection with the number of events in the respective bin. The resulting projections are shown by the black lines in figure 6.9 which show good agreement with the distributions of toy events after the cut.
The parameters obtained from this fit are then used to draw projections of $\mathcal{P}^{\text{without-cut}}_{\text{bkg}}$ scaled with the total number of events before the cut as shown by the red line. These projections are in very good agreement with the distributions of the toy sample before cut, demonstrating that the correct parameters can be obtained even when fitting the events after the cut.

Pull study

Furthermore, a pull-study is conducted to test the correction of the veto. The pull of a fit parameter is defined as

$$\text{pull}(a) = \frac{a - a_{\text{true}}}{\sigma(a)}, \tag{6.10}$$

where a is the value returned by the fit, a_{true} is the value with which the toy sample was generated, and $\sigma(a)$ is the uncertainty returned by the fit. If the fit has no bias for parameter a and the uncertainty $\sigma(a)$ is a 1σ standard deviation, then pull(a) is distributed according to a normal distribution with $\mu(\text{pull}(a)) = 0$ and $\sigma(\text{pull}(a)) = 1$. Thus, by generating many event samples, fitting each one and calculating the pull for every parameter in every sample it can be tested whether the fit returns the correct parameters without any biases.

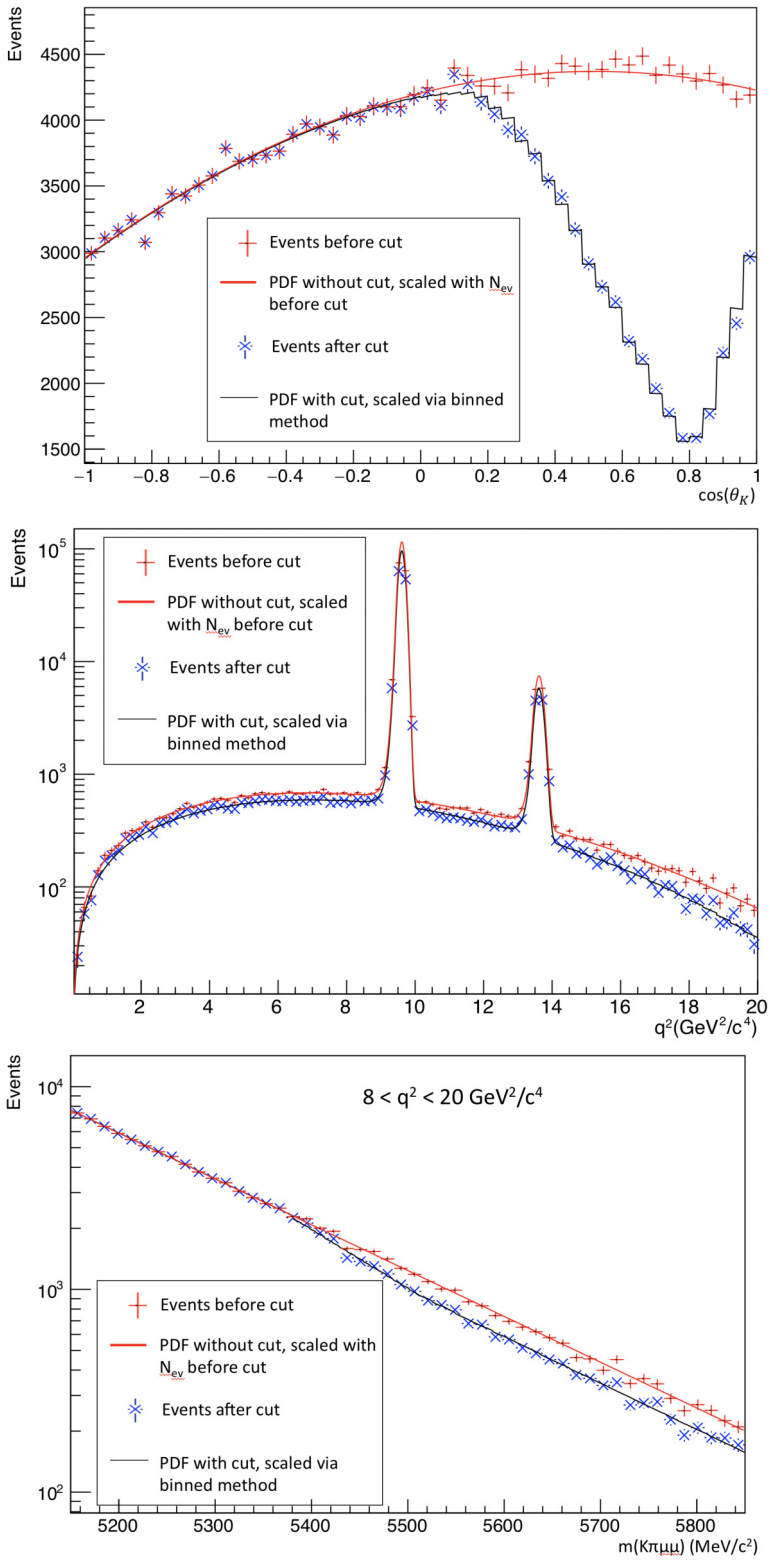


Figure 6.9: $\cos\theta_K$, q^2, $m^{\text{rec}}_{B^0}$ distributions of a toy MC before and after the ($\cos\theta_K$, q^2, m)-cut. The sample of events is fitted with the PDF which has been adjusted for the missing phase space. The projections of that fit are shown with the black line. Using the parameters from that fit, projections of the PDF without adjusting for the missing phase space are shown with the red line which shows good agreement with the distribution of the events before the cut. This demonstrates that the method is capable of determining the correct background shapes even when fitting the event sample after cut.

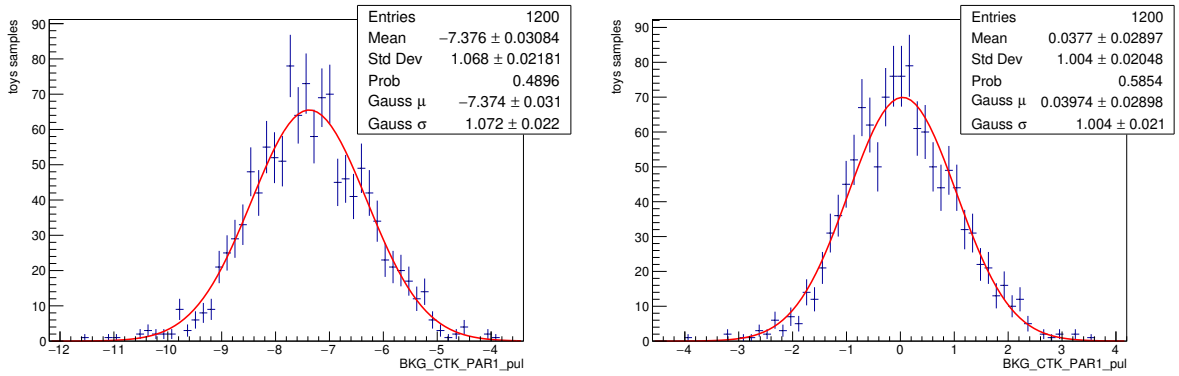


Figure 6.10: Distribution of pull($a^{\mathrm{K}}_{\mathrm{comb}}$) in fits to 1200 toy event samples with approximately 80000 background events each. The ($\cos\theta_K$, q^2, m)-cut was applied to the each toy sample. The events were then fitted without correcting for the veto (left figure), and with the correction for the veto (right figure). Also displayed are the mean and standard deviation of the histograms as well as a Gaussian fit to each histogram (red line) and the resulting μ and σ parameters. There is a clear bias on $a^{\mathrm{K}}_{\mathrm{comb}}$ when ignoring the missing phase space but when adjusting the PDF accordingly the bias is removed.

To test the correction of the veto, 1200 background toy event samples with approximately the same number of background events as in the merged Run 1 and Run 2 data are generated.[5] The ($\cos\theta_K$, q^2, $m^{\mathrm{rec}}_{B^0}$)-cut is applied to each toy sample, and each sample is then fitted with and without correcting for the missing phase space. Figure 6.10 shows the distribution of the pull of $a^{\mathrm{K}}_{\mathrm{comb}}$ for the fits ignoring the gap (left) and the fits correcting for it (right). The boxes in each figure display the mean and standard deviation of the pull values as well as results from Gaussian fits (red lines) to the pull distributions. The background fits ignoring the effect of the veto have a strong bias of about $-7.4\sigma(a^{\mathrm{K}}_{\mathrm{comb}})$ i.e. the fits return too small values of $a^{\mathrm{K}}_{\mathrm{comb}}$, whereas the background fits, which use the correction for the missing phase space, return the correct value for $a^{\mathrm{K}}_{\mathrm{comb}}$ without any bias.

The μ_{pull} values resulting from Gaussian fits to the pull distributions of all background parameters are given in table 6.4. The biases due to the ignored veto are largest for the $\cos\theta_K$ parameters of the J/ψ background and second and third largest for the $\cos\theta_K$ parameters of the $\psi(2S)$ and combinatorial background. This is due to the fact that there are more J/ψ background events than $\psi(2S)$ background events (and more $\psi(2S)$ background events than fully combinatorial events). Thus, the test is most sensitive to the mismodeling of the J/ψ shape and hence gives a bigger bias for the J/ψ parameters. The

[5]The number of events for each toy sample is sampled from a Poisson distribution with a mean of the number of background events in the merged Run 1 and Run 2 data.

centre column of table 6.4 shows that the fits which include the correction for the veto can recover all parameters correctly apart from a few small biases. These biases are due to the low number of events and the high number of floating parameters. To show that the fits have no biases when the number of events is higher, the pull study is repeated with twice the number of background events per toy sample and the resulting μ_{pull} values are given in the right column in table 6.4.

	Run 1+Run 2 stats no correction	Run 1+Run 2 stats with correction	2×(Run 1+Run 2 stats) with correction
	μ_{pull}	μ_{pull}	μ_{pull}
s^{m}_{comb}	0.510 ± 0.028	-0.038 ± 0.028	0.019 ± 0.029
$s^{m}_{J/\psi}$	1.050 ± 0.028	-0.066 ± 0.028	0.020 ± 0.029
$s^{m}_{\psi(2S)}$	0.340 ± 0.029	-0.036 ± 0.029	0.020 ± 0.029
$a^{\text{L}}_{\text{comb}}$	-0.025 ± 0.029	-0.025 ± 0.029	-0.021 ± 0.029
$b^{\text{L}}_{\text{comb}}$	0.010 ± 0.029	0.003 ± 0.029	-0.051 ± 0.029
$a^{\text{K}}_{\text{comb}}$	-7.374 ± 0.031	0.040 ± 0.029	-0.030 ± 0.029
$b^{\text{K}}_{\text{comb}}$	-3.897 ± 0.026	0.082 ± 0.029	-0.010 ± 0.029
a^{ϕ}_{comb}	-0.014 ± 0.030	-0.014 ± 0.030	-0.008 ± 0.030
b^{ϕ}_{comb}	0.042 ± 0.029	0.041 ± 0.029	0.021 ± 0.030
$a^{q^2}_{\text{comb}}$	-0.478 ± 0.029	0.001 ± 0.029	0.001 ± 0.029
$a^{\text{L}}_{J/\psi}$	0.050 ± 0.028	0.050 ± 0.028	-0.012 ± 0.028
$b^{\text{L}}_{J/\psi}$	-0.032 ± 0.029	-0.028 ± 0.029	-0.000 ± 0.029
$a^{\text{K}}_{J/\psi}$	-24.337 ± 0.030	0.038 ± 0.029	-0.007 ± 0.029
$b^{\text{K}}_{J/\psi}$	-14.157 ± 0.026	-0.007 ± 0.029	0.017 ± 0.029
$a^{\phi}_{J/\psi}$	0.019 ± 0.028	0.019 ± 0.028	-0.016 ± 0.029
$b^{\phi}_{J/\psi}$	0.023 ± 0.028	0.023 ± 0.028	0.016 ± 0.029
$\mu^{q^2}_{J/\psi}$	-0.130 ± 0.029	-0.007 ± 0.029	0.020 ± 0.029
$\sigma^{q^2}_{J/\psi}$	-0.097 ± 0.029	-0.064 ± 0.029	-0.015 ± 0.030
$\alpha^{q^2}_{J/\psi}$	-0.108 ± 0.029	-0.032 ± 0.029	-0.022 ± 0.030
$a^{\text{L}}_{\psi(2S)}$	0.028 ± 0.029	0.026 ± 0.029	-0.017 ± 0.030
$b^{\text{L}}_{\psi(2S)}$	-0.023 ± 0.029	-0.055 ± 0.029	0.026 ± 0.030
$a^{\text{K}}_{\psi(2S)}$	-8.815 ± 0.031	0.045 ± 0.029	0.001 ± 0.028
$b^{\text{K}}_{\psi(2S)}$	-3.822 ± 0.025	0.070 ± 0.029	0.010 ± 0.029
$a^{\phi}_{\psi(2S)}$	0.027 ± 0.028	0.025 ± 0.028	0.017 ± 0.030
$b^{\phi}_{\psi(2S)}$	0.006 ± 0.030	0.000 ± 0.030	0.015 ± 0.030
$\mu^{q^2}_{\psi(2S)}$	-0.109 ± 0.029	0.026 ± 0.029	0.017 ± 0.029
$\sigma^{q^2}_{\psi(2S)}$	-0.008 ± 0.030	-0.080 ± 0.030	-0.031 ± 0.029
$f_{J/\psi}$	0.432 ± 0.030	0.056 ± 0.029	0.035 ± 0.029
$f_{\psi(2S)}$	0.093 ± 0.030	0.056 ± 0.030	0.049 ± 0.030

Table 6.4: Results of Gaussian fits to the pull distributions of all floating background parameters in 1200 fits to toy sample swith the same (doubled) number of background events as in the merged Run 1 and Run 2 dataset. The uncertainties given are the statistical uncertainties from the Gaussian fits. The ($\cos\theta_K$, q^2, m)-cut was applied to each toy sample. The events were then fitted with and without for the gap in the phase space. Ignoring the gap causes biases for $\cos\theta_K$, q^2 and $m^{\text{rec}}_{B^0}$ parameters. Even with the correction there are small biases for some of the parameters which are due to low statistics. The right column shows the results from a pull study with double the number of events per toy sample which shows that the biases are removed.

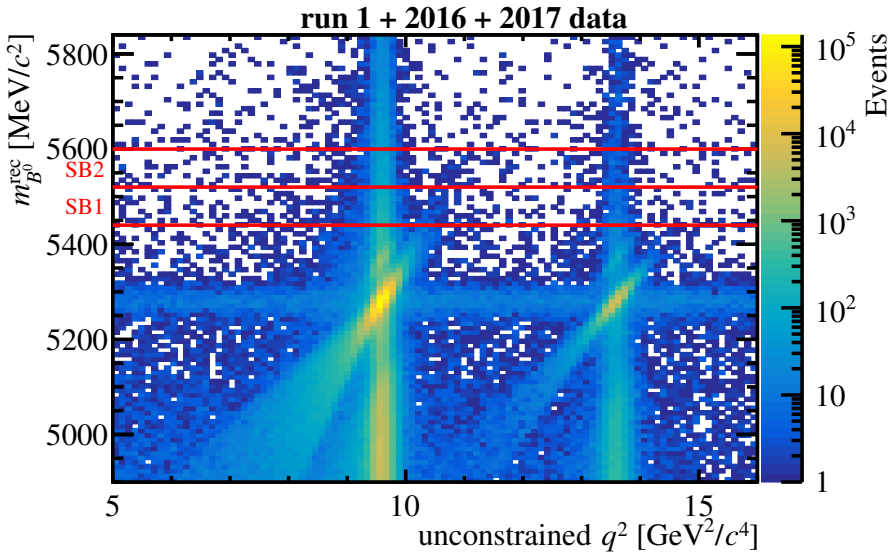


Figure 6.11: q^2 vs $m^{\mathrm{rec}}_{B^0}$ distribution of the events in Run 1+Run 2 data. The $m^{\mathrm{rec}}_{B^0}$ regions of the side-band regions SB1 and SB2 are indicated with the horizontal red lines. The diagonal band crossing $q^2 = 9.8$ GeV$^2/c^4$ is made up of signal $B^0 \to J/\psi K^{*0}$ events which leak into SB1 and SB2.

6.5 Signal in side-band veto

There is a contribution of $B^0 \to J/\psi K^{*0}$ signal decays in the upper mass side-band which – if ignored – can cause a bias for the background parameterisation.

Figure 6.11 shows the (unconstrained-q^2, $m^{\mathrm{rec}}_{B^0}$) distribution of the events in the merged Run 1 and Run 2 dataset. There is a diagonal band crossing $q^2 = 9.8\,\mathrm{GeV}^2/c^4$ which is made up of $B^0 \to J/\psi K^{*0}$ decays. The events in the lower and higher end of the diagonal band are $B^0 \to J/\psi K^{*0}$ decays where the J/ψ mass was either reconstructed too low or too high, which causes the correlation between q^2 and $m^{\mathrm{rec}}_{B^0}$.

These events leak into the first two side-band regions SB1 and SB2 (indicated with red lines in figure 6.11) and become a significant contribution in these regions despite being so far from the true mass of the B^0 meson. They are significant because they are shifted so far in q^2 that they are separated from the background J/ψ peak. Thus, they are populating an area in q^2 where only the fully combinatorial background is present.[6] This

[6]There are also $B^0 \to J/\psi K^{*0}$ events in SB1 and SB2, which have high $m^{\mathrm{rec}}_{B^0}$ because the K^{*0} was reconstructed too high instead of the J/ψ. These events are not shifted in q^2 and therefore populate the q^2 region dominated by the background J/ψ events. The level of signal events under the background J/ψ peak in SB1 is on the level of 2% and can safely be ignored.

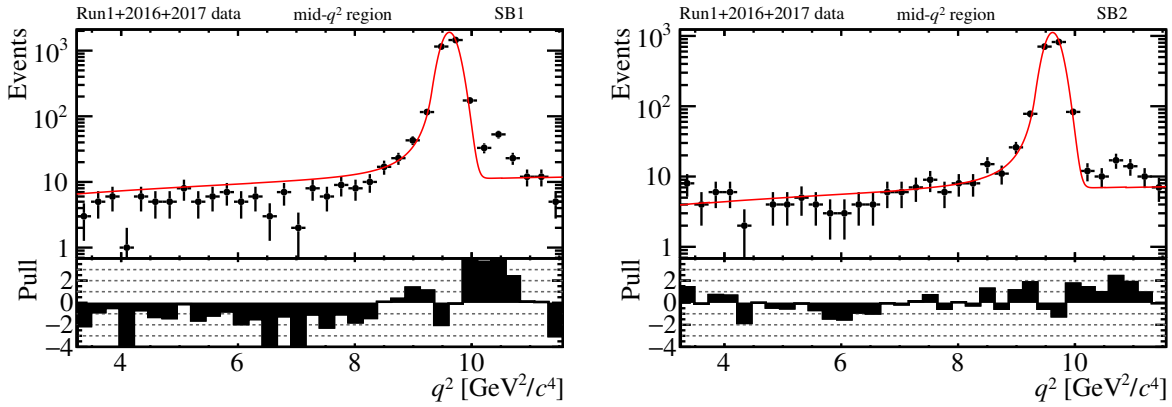


Figure 6.12: q^2 distributions of the events in the mid-q^2 region in SB1 (left) and SB2 (right). Also shown is the q^2 projection of the background PDF. The small peak above the background J/ψ peak is from $B^0 \to J/\psi K^{*0}$ signal events. Since the signal peak is not described by the background model, the presence of the peak causes the combinatorial background to be overestimated by the fit PDF as seen most clearly in the left figure for $q^2 < 8\,\mathrm{GeV}^2/c^2$.

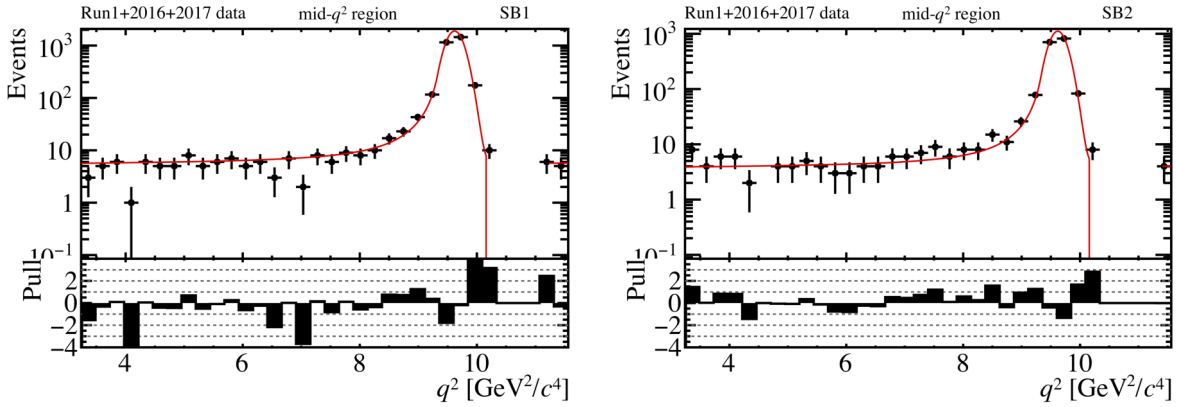


Figure 6.13: q^2 distributions of the events in the mid-q^2 region in SB1 (left) and SB2 (right) after applying the signal veto to remove contributions from $B^0 \to J/\psi K^{*0}$ signal events. Also shown is the q^2 projection of the background PDF which has been fitted to the data while accounting for missing phase space due to the signal veto.

is demonstrated in figure 6.12 which shows the q^2 distributions of the events in the mid-q^2 region in SB1 and SB2. The $B^0 \to J/\psi K^{*0}$ signal events are visible as small peaks above the background J/ψ peak. When fitting the background without taking this contribution into account, the signal peaks cause the level of fully combinatorial background to be overestimated as shown by the projection of the fitted background PDF in figure 6.13.

In order to remove the signal contribution in the side-band, a similar technique is used as for the $K\mu\mu$-veto described in the previous section. The q^2 ranges populated

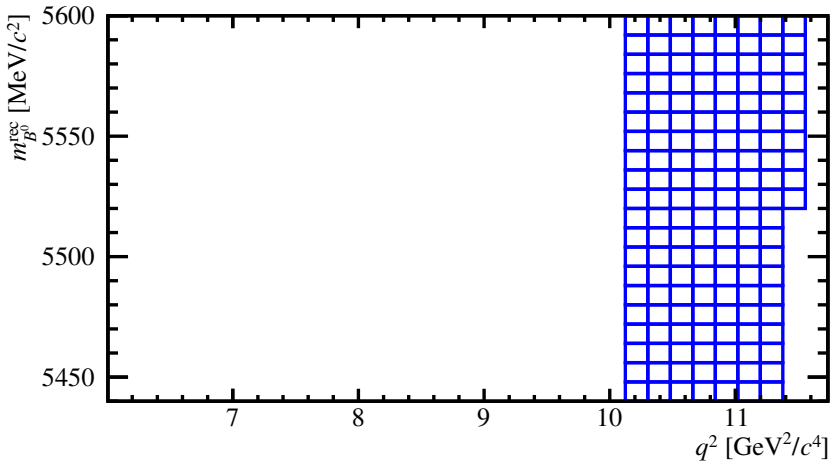


Figure 6.14: Rejected (q^2, $m_{B^0}^{\mathrm{rec}}$) region of the signal veto. Each box represents a bin in phase space in which all events are rejected and in which the background PDF is set to zero.

by $B^0 \to J/\psi K^{*0}$ signal events in SB1 and SB2 are removed from the analysis and the background PDF is adjusted accordingly, analogously to the method described in section 6.4.2. Figure 6.14 shows the (q^2,$m_{B^0}^{\mathrm{rec}}$)-histogram indicating the rejected phase space, where each blue box represents a phase space bin which is removed from the analysis. The q^2 ranges are chosen such that 100% of the $B^0 \to J/\psi K^{*0}$ signal events are removed.

Figure 6.13 shows the q^2 distribution in SB1 and SB2 as well as the q^2 projection of the background fit, after rejecting the events in the signal affected phase space and adjusting the PDF accordingly. The PDF now describes the combinatorial background well.

As can be seen in figure 6.11, the contribution from $B^0 \to \psi(2S)K^{*0}$ decays, which is the second largest signal contribution, is too small to be significant in the upper mass side-band. Therefore, a similar veto for $B^0 \to \psi(2S)K^{*0}$ decays is not necessary.

6.6 Factorisation of angles and q^2

The PDF used to describe the background assumes factorisation in all dimensions, which greatly simplifies the expression. To justify this assumption, it is necessary to check the data for correlations in any of the dimensions. In this section, it is shown that it is a reasonable approximation that the three angles and q^2 are uncorrelated. The potential dependence of the angular and q^2 shapes on $m^{\text{rec}}_{B^0}$ and the treatment of this dependence is separately discussed in section 6.7.

6.6.1 Factorisation of each angle with q^2

Investigating the factorisation of the angles with q^2 is only necessary for the fully combinatorial background as it spreads across the full q^2 range. The resonant background events contain real J/ψ and $\psi(2S)$ mesons which have very narrow decay widths such that the angular distributions factorise with q^2 naturally.

Figure 6.15 shows the $\cos\theta_\ell$, $\cos\theta_K$, and ϕ distributions of the fully combinatorial background in the low, mid and high q^2 region. For this study, all events in the sub-regions containing the resonant backgrounds, i.e. $(8 \leq q^2 \leq 11.56)$ GeV$^2/c^4$ and $(12.9 \leq q^2 \leq 14.5)$ GeV$^2/c^4$, are removed, since they would otherwise dominate the mid and high q^2 event sample, thus making it impossible to study the q^2 dependence of the angular distributions of the fully combinatorial background. The bottom of each figure shows the ratios of the angular distributions in the mid and high q^2 region over the distributions in the low q^2 region. No significant dependence of the shape of the angular distribution on q^2 is observed, except in $\cos\theta_K$ where the high and mid q^2 distribution falls more clearly at high $\cos\theta_K$ than in low q^2 region. However, this is due to the effect of the $K\mu\mu$-veto which becomes stronger with rising q^2 and is corrected for in the side-band fit as explained in section 6.4.

Since no significant dependence of the shape of the angular distribution on q^2 is observed, a PDF that factorises in each angle with q^2 is a good description of the data. In future measurements with more data, this should be reviewed again as the measurements might become sensitive to potential correlations.

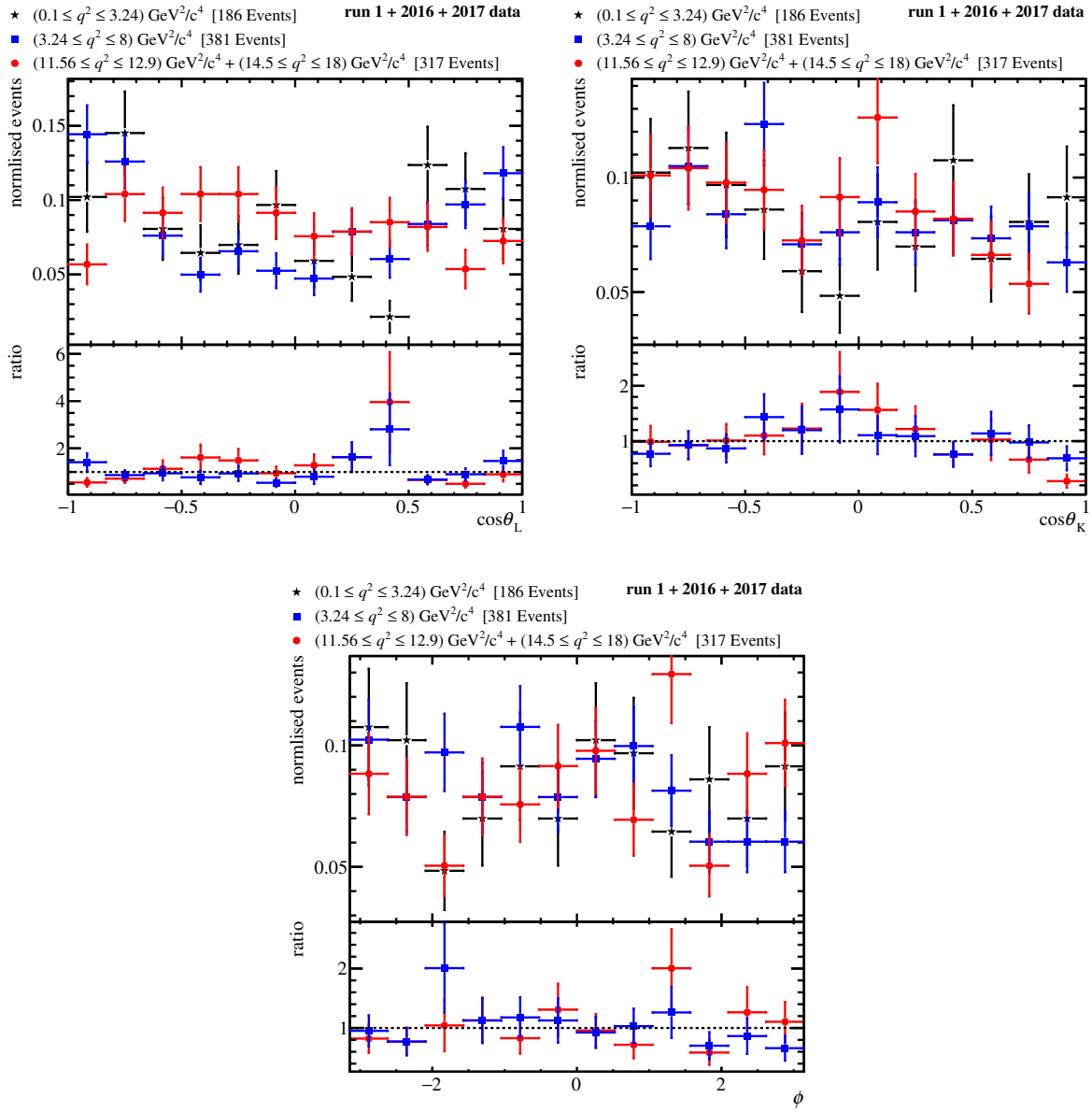


Figure 6.15: $\cos\theta_\ell$, $\cos\theta_K$, and ϕ distributions of the fully combinatorial background in the low q^2 region (black stars), mid q^2 region (blue squares), and high q^2 region (red circles). Events from the resonant sub-regions are not included. Also shown are the ratios of the angular distributions in the mid and high q^2 region over the angular distributions in the low q^2 region.

6.6.2 Factorisation of the three angles

To test the assumption that the angluar distributions of the background can be described with factorising Chebychev polynomials, a goodness of fit (g.o.f.) test is used. After performing the simultaneous side-band fit (as described in this chapter) to the merged

Run 1 and Run 2 data, a three dimensional g.o.f. test is carried out in the dimensions $\cos\theta_\ell$,$\cos\theta_K$, and ϕ. If the fit cannot be rejected, it can be assumed that using a PDF which factorises in the three angles is a reasonable approximation.

Goodness of fit test: Mixed sample method

To determine the level of agreement between the data and the fit PDF, a multivariate unbinned g.o.f. test is implemented. The test is based on the mixed sample method described in Ref. [99].

If two event samples share the same parent distributions, then they *mix* perfectly when combined into a single event sample.[7] This can be used to judge the g.o.f. of a fit to data by simulating toy events from the fit PDF and determining how well the toy events mix with the data events.

To quantify how well two event samples mix, the n_k nearest neighbours of any given event in the combined dataset must be identified. For this purpose, the normalised Euclidean distance of two events $\vec{x_i}$ and $\vec{x_j}$ in the three dimensional space is defined as

$$\|\vec{x_i} - \vec{x_j}\|^2 = \sum_{v=1}^{3} \left(\frac{x_i^v - x_j^v}{w_v} \right)^2 , \tag{6.11}$$

with $v \in \{\cos\theta_\ell, \cos\theta_K, \phi\}$ and where w_v denotes the weights for the different dimensions. The weights are chosen as the largest possible distance in each dimension given by the respective ranges, i.e. $w_{\cos\theta_\ell} = w_{\cos\theta_K} = 2$ and $w_\phi = \pi$. In the case of $v = \phi$, the distance $x_i^v - x_j^v$ is determined as the shortest distance between the two angles.
Following Ref. [99], the test statistic that quantifies the quality of the mixing of the toy MC event sample with the dataset is defined as

$$T = \frac{1}{n_k(n_d + n_m)} \sum_{i=1}^{n_d+n_m} \sum_{k=1}^{n_k} I(i,k), \tag{6.12}$$

where n_d and n_m are the number of events in the dataset and the toy MC sample respectively, n_k is the number of nearest neighbours considered for each event, and $I(i,k)$ is defined with

$$I(i,k) = \begin{cases} 1 & \text{if event } i \text{ and its } k^{th} \text{ neighbour are both data events or both toy events} \\ 0 & \text{otherwise.} \end{cases} \tag{6.13}$$

[7] The notion of *mixing* in this context will become clear after the explanation of the method in this section.

If $n_d = n_m$ and the toy sample and the data sample mix perfectly, the expectation value for T is 1/2. In other words, on average one would expect half of the nearest neighbours of each event to be from the same sample as the event itself. Worse mixing i.e. less agreement between the two samples yields higher values for T.

The pull of T is defined as

$$\text{pull}(T) = \frac{(T - \mu_T)}{\sigma_T}, \tag{6.14}$$

where μ_T and σ_T are the expectation value and standard deviation of T. The expectation value is given by

$$\mu_T = \frac{n_d(n_d - 1) + n_m(n_m - 1)}{n(n-1)} \tag{6.15}$$

with $n = n_d + n_m$. The standard deviation of T depends on the shape of the PDF and can only be approximated:

$$\lim_{n,n_k,D\to\infty} \sigma_T^2 = \frac{1}{nn_k}\left(\frac{n_d n_m}{n^2} + 4\frac{n_d^2 n_m^2}{n^4}\right), \tag{6.16}$$

where D is the number of dimensions of the g.o.f. test. A detailed discussion of this approximation can be found in Ref. [99] where it was found to hold well for $n_m = 10 \times n_d$ and $n_k = 10$ in a two dimensional test. If the approximation of σ_T is correct and the data sample and the toy sample follow the same parent distribution – i.e. if the fit PDF models the data perfectly – then the pull is expected to have a limiting standard normal distribution (mean of zero and standard deviation of one). Since larger values for T are obtained if the distributions do not agree well, the rejection of the hypotheses that the two distributions have the same parent distribution – i.e. that the fit describes the data well – is a one-sided cut at high pull values. For example, fits to data that yield a pull ≥ 1.64 are rejected at 95% confidence level.

In order to ensure that the ($\cos\theta_K$,q^2,$m^{\text{rec}}_{B^0}$)-cut used to remove the phase-space which is affected by the $K\mu\mu$-veto (see section 6.4) and the signal in side-band veto (see section 6.5) do not influence the g.o.f. test, both cuts are applied to the side-band data as well as to every toy simulation sample.

Since the g.o.f. test is performed in three dimensions and the approximation of σ_T (equation 6.16) may depend on the shape of the PDF, the g.o.f. method is tested first to determine the expected pull distribution. This is done by performing g.o.f. tests for perfect 'fits' where it is a priori known that the PDF can describe the events. To this end, an ensemble of 200 toy event samples is generated from the background PDF with the

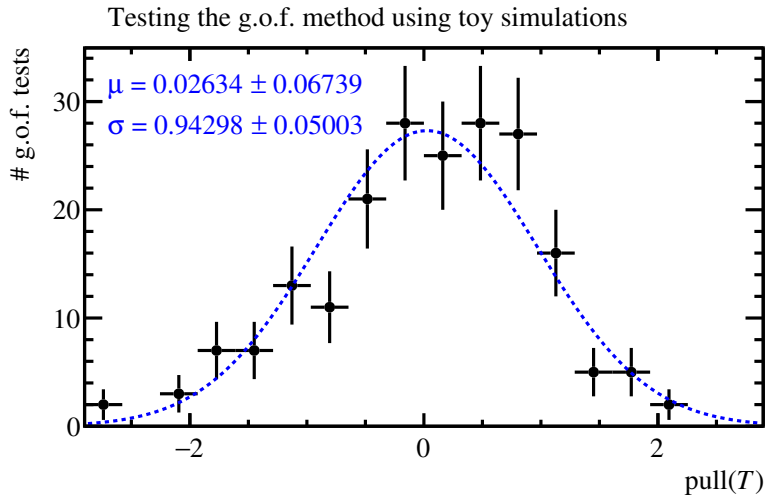


Figure 6.16: Pull distributions obtained from performing three dimensional g.o.f. tests for 200 simulated toy samples. The mean and standard deviation of the pull distribution, determined by performing a Gaussian fit (dotted blue line) are given in the plot.

same number of events per sample as in the upper mass side-band data ($n_d = 8824$). In this test of the g.o.f. method, these *small* toy event samples represent the data sample in the real g.o.f. test. Each small toy sample is then compared with a larger toy event sample of $n_m = 10 \times n_d = 88240$ events, generated from the same background PDF. The g.o.f. tests are performed with $n_k = 10$. The resulting pull distribution is shown in Figure 6.16. The mean value obtained with a Gaussian fit to the pull distribution is compatible with zero as expected. Moreover, the width of the pull distribution roughly agrees with unity. Therefore, it can be concluded that the approximation for the standard deviation (equation 6.16) holds well for the three dimensional g.o.f. test with the background PDF obtained in the fit to the Run 1 and Run 2 data.

The mixed sample method is then used to determine the ($\cos\theta_\ell$,$\cos\theta_K$,ϕ)-g.o.f. of the upper mass side-band fit to the merged Run 1 and Run 2 data. Since the g.o.f. test is influenced by statistical fluctuations in the toy sample, the test is repeated 200 times and the average pull value is determined. Therefore, an ensemble of 200 toy MC samples with $n_m = 10 \times n_d = 88240$ events each is generated from the fit PDF. Each toy sample is compared to the upper side-band data to determine a pull value. The pull distribution of the 200 tests is shown in figure 6.17. The mean pull value, determined with a Gaussian fit, is -0.049 ± 0.057. This indicates that the fit cannot be rejected, or in other words the $\cos\theta_\ell$,$\cos\theta_K$, and ϕ distributions of the events in the upper mass side-band are in good agreement with the background PDF.

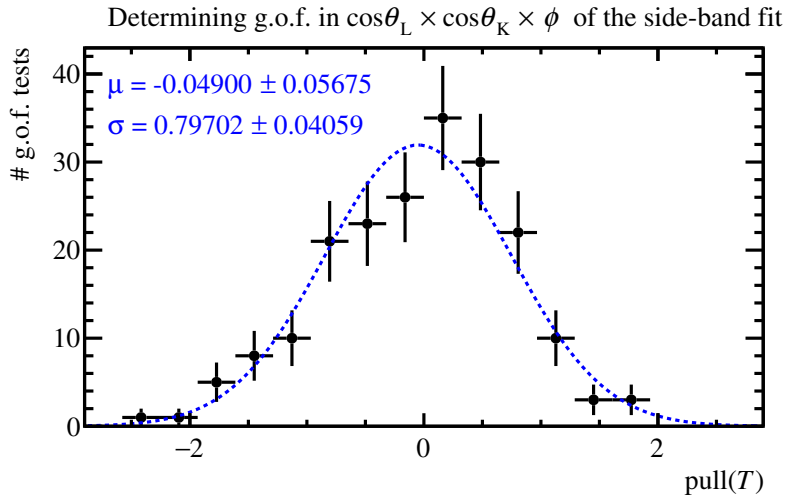


Figure 6.17: Results of 200 g.o.f. tests performed in ($\cos\theta_\ell$, $\cos\theta_K$, ϕ) for the upper mass side-band fit. Each test is done by simulating a large toy event sample (with 10 times the number of events as the data sample) from the background PDF and comparing the toy event sample to the upper mass side-band data, using the mixed event method to determine a pull value. The mean pull value, determined with the Gaussian fit shown with the dashed blue line, is compatible with zero, indicating that the fit PDF is compatible with the data.

Note that the width of the pull distribution shown in figure 6.17 cannot be interpreted in the same way as the width of the distribution in figure 6.16. Figure 6.17 shows the results from repeating the *same* g.o.f. test 200 times using the same data sample each time. The width of the pull distribution is determined by the statistical fluctuations in the toy samples. These fluctuations are proportional to $\sqrt{n_m}$. In contrast, figure 6.16 shows the results from 200 individual g.o.f. tests. In each test an independent *small* toy event sample (representing the data sample of the real g.o.f. test) is compared to a corresponding large toy MC sample. Only in this case the expected value for σ_T is unity.

Since the background PDF factorises in $\cos\theta_\ell$,$\cos\theta_K$ and ϕ, it can also be concluded that the assumption of the factorisation of the angles holds. This test should be repeated in upcoming analyses which use more data as the fit may become sensitive to potential correlations.

A complementary test of the factorisation assumption has been performed for the measurement of the q^2-binned angular observables [1], via a method of moments analysis on the upper mass side-band of the data. The angular distributions of the upper mass side-band data were parameterised with fully correlated Legendre polynomials and were also parameterised with factorised Legendre polynomials. It was found that all diagonal

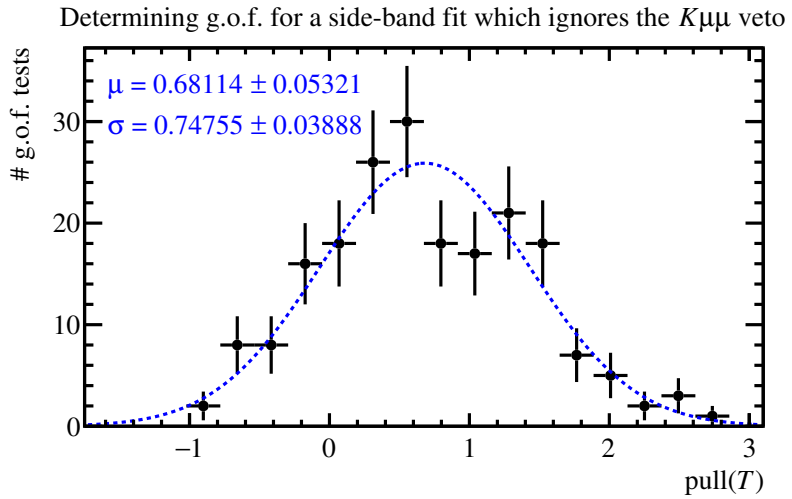


Figure 6.18: Results of 200 g.o.f. tests performed in ($\cos\theta_\ell$, $\cos\theta_K$, ϕ) for a bad fit to the upper mass side-band which ignores the effect of the $K\mu\mu$ veto. Each test is done by simulating a large toy event sample (with 10 times the number of events as the data sample) from the background PDF and comparing the toy event sample to the upper mass side-band data, using the mixed event method to determine a pull value. The mean pull value indicates that the fit PDF can be rejected at the 75% confidence level.

coefficients in the correlated Legendre polynomials are compatible with the corresponding coefficients in the factorised Legendre polynomials. In other words no significant correlations have been found [91].

Example of a rejected bad fit

To demonstrate the ability of the mixed sample g.o.f. test to reject bad fits, the side-band fit is repeated but without applying the ($\cos\theta_K$,q^2,$m^{\rm rec}_{B^0}$)-cut and without the corresponding adjustment of the PDF (see section 6.4). An exemplary projection of this fit is shown in figure 6.4. The g.o.f. test is carried out 200 times, as described above but the ($\cos\theta_K$,q^2,$m^{\rm rec}_{B^0}$)-cut is not applied to the data or the toy simulations. The resulting pull(T) distribution is shown in figure 6.18. The mean pull value of -0.68 ± 0.05 indicates that the fit can be rejected at the 75% confidence level.

6.7 Mass dependence of the background parameters

Since the background parameters are extrapolated along $m^{\rm rec}_{B^0}$, any dependence of the background parameters on $m^{\rm rec}_{B^0}$ has to be considered in order to ensure an accurate description of the background in the signal region.

To this end, the fact that the side-band regions are fitted simultaneously is taken advantage of. In each side-band region SB_i, every background parameter $p^x_{\mathrm{SB}_i}$ for $x \in (\cos\theta_\ell,\ \cos\theta_K,\ \phi,\ q^2)$ (i.e. all parameters in table 6.2 except for the $m^{\mathrm{rec}}_{B^0}$ parameters and the fractions) is reparameterised with:

$$p^x_{\mathrm{SB}_i} = p^x_{\mathrm{a}} + p^x_{\mathrm{md}} \times m^d_{\mathrm{SB}_i} \tag{6.17}$$

where p^x_{a} and p^x_{md} are fit parameters and m^d_i is the distance of the center of region SB_i to the B^0 mass. In order to ensure that p_{a} and p_{md} have comparable scales, m^d_i is scaled by the distance between the highest SB region (SB5) and the B^0 mass. The values of $m^d_{\mathrm{SB}_i}$ can be found the right column of table 6.1. This means, within each SB region, factorisation of the angles and q^2 with $m^{\mathrm{rec}}_{B^0}$ is still assumed but the fit allows for a linear dependence of every background parameter from region to region.
The benefit of this reparameterisation of the background parameters is that it allows taking a mass dependence into account while still giving the minimal number of background parameters in the signal region (since $m^d_{\mathrm{Signal}} = 0$) which simplifies the combined signal and background fit.

When reparameterising all background angular and q^2 parameters with equation 6.17, the total number of floating parameter becomes 52. In this case, due to the high number of floating parameters and the large correlations between most p^x_{a} and p^x_{md} parameters, the uncertainties of the background parameters in the signal region increase significantly (up to a factor 3). It is therefore preferential not to unnecessarily give additinoal freedom to the fit model if the model with some of the p^x_{md} parameters fixed to zero is also a good description of the data.

The most significant dependence of the angular shape on the SB region is observed for the $\cos\theta_K$ dimension[8], while the $\cos\theta_\ell$, ϕ and q^2 shapes are consistent between SB regions. Therefore, the nominal model allows mass dependence only in the $\cos\theta_K$ dimension. The alternative of parameterising the $m^{\mathrm{rec}}_{B^0}$ dependence of all background parameters, is considered as a systematic uncertainty as described in section 7.2.1.

Mass dependence of the background fractions

Due to the different $m^{\mathrm{rec}}_{B^0}$ slopes of the three background components, the background fractions $f_{J/\psi}$ and $f_{\psi(2S)}$ are different in each mass region. To take this into account, the

[8]Even after correcting for the effect of the $K\mu\mu$-veto.

fraction in each side-band region $f_{J/\psi}^{\mathrm{SB}_i}$ ($f_{\psi 2s}^{\mathrm{SB}_i}$) is calculated as a function of the fraction in the signal region $f_{J/\psi}^{\mathrm{S}}$ ($f_{\psi(2S)}^{\mathrm{S}}$) :

$$f_{J/\psi}^{\mathrm{SB_i}} = \frac{f_{J/\psi}^{\mathrm{S}} \times \frac{\int_{\mathrm{SB_i}} \mathcal{P}_{J/\psi}}{\int_{\mathrm{S}} \mathcal{P}_{J/\psi}}}{f_{J/\psi}^{\mathrm{S}} \times \frac{\int_{\mathrm{SB}_i} \mathcal{P}_{J/\psi}}{\int_{\mathrm{S}} \mathcal{P}_{J/\psi}} + (1 - f_{J/\psi}^{\mathrm{S}}) \times \frac{\int_{\mathrm{SB}_i} \mathcal{P}_{\mathrm{com}}}{\int_{\mathrm{S}} \mathcal{P}_{\mathrm{com}}}} \tag{6.18}$$

where $\int_{\mathrm{S}}$ and $\int_{\mathrm{SB}_i}$ denote the 5-dimenstional integrals of the PDF over the signal region and the side-band region SB_i respectively. The expression for the $\psi(2S)$ fraction $f_{\psi 2s}^{\mathrm{SB}_i}$ is analogous.

By replacing the background fractions $f_{J/\psi}^{b}$ and $f_{\psi(2S)}^{b}$ in equation 6.9 with the above expressions for $f_{J/\psi}^{\mathrm{SB}_i}$ and $f_{\psi(2S)}^{\mathrm{SB}_i}$ respectively, the background fractions in the signal region are direct fit parameters in the side-band fit, while equation 6.18 gives the background fraction in each side-band region such that the background PDF can describe the side-band data.

6.8 Determination of the signal fraction

The total PDF for the combined signal and background fit in the signal region ($(5239.58 \leq m_{B^0}^{\mathrm{rec}} \leq 5319.58)$ MeV/c^2) is given by

$$\mathcal{P}_{\mathrm{bkg}}^{i}(\vec{\Omega}, q^2) = f_{\mathrm{sig}}^{i} \cdot \mathcal{P}_{\mathrm{sig}}(\vec{\Omega}, q^2) + (1 - f_{\mathrm{sig}}^{i}) \cdot \mathcal{P}_{\mathrm{bkg}}(\vec{\Omega}, q^2), \tag{6.19}$$

where i indicates the q^2-region i.e. low, mid, and high q^2 (see table 5.4). $\mathcal{P}_{\mathrm{sig}}(\vec{\Omega}, q^2)$ denotes the signal PDF which is given by the model described in chapter 3, convolved with the resolution model (described in section 5.3) and multiplied with the acceptance model (described in section 5.2). $\mathcal{P}_{\mathrm{bkg}}(\vec{\Omega}, q^2)$ denotes the background PDF described throughout this chapter. Note however, that in the fit to the signal region, $m_{B^0}^{\mathrm{rec}}$ is not included in the fit. The differences between the signal PDFs in the three q^2 regions are the resolution parameters (see section 5.3). The differences between the background PDFs in the three q^2 regions are the different background components present in each region (see section 6.2 and equation 6.4).

The parameter f_{sig}^{i} in equation 6.19 denotes the fraction of signal events in the signal region in q^2 region i ($i \in \{$low q^2, mid q^2, high q^2 $\}$). These fractions need to be determined from data. They can be calculated with

$$f_{\mathrm{sig}}^{i} = 1 - \frac{N_{\mathrm{bkg}}^{\text{signal region},i}}{N_{\mathrm{total}}^{\text{signal region},i}}, \tag{6.20}$$

Here $N_{\text{bkg}}^{\text{signal region, } i}$ denotes the number of background events and $N_{\text{total}}^{\text{signal region } i}$ the sum of signal events and background events in q^2 region i in the signal region. While the total number of events in each q^2 region in the signal region is a priori known for the given dataset, the number of background events in each q^2 region in the signal region needs to be estimated from a fit.

There are two different approaches for estimating the number of background events in the signal region.:

A) Using the slope parameters (s^m_{comb}, $s^m_{J/\psi}$, $s^m_{\psi(2S)}$) and the background fractions ($f_{J/\psi}$, $f_{\psi(2S)}$), obtained in the simultaneous 5D side-band fit, to extrapolate the yield of each background component from the upper mass side-band into the signal region.

B) Performing separate fits to $m^{\text{rec}}_{B^0}$ in the full range ($5220 \leq m^{\text{rec}}_{B^0} \leq 5840$) MeV/$c^2$, parameterising both the signal and background contributions and determining $f^i_{\text{full range}}$. The number of background events in the full $m^{\text{rec}}_{B^0}$-range can then be calculated in each q^2 region using $f^i_{\text{full range}}$. These background yields are then interpolated into the signal region using the fitted background slopes.

Both methods have been implemented and tested and are described below. However, due to the superior statistical power obtained with method B, this method is chosen for the analysis. Therefore method A is only outlined briefly.

6.8.1 Method A - using the slopes and fractions from the side-band fit

The simultaneous side-band fit provides best fit values for the background slopes s^m_{comb}, $s^m_{J/\psi}$, $s^m_{\psi(2S)}$), all angular and q^2 parameters, as well as the background fractions $f^{\text{before,cut}}_{J/\psi}$ and $f^{\text{before,cut}}_{\psi 2s}$, where 'before cut' refers to both the $K\mu\mu$-veto (see equation 6.9) and the signal veto (see section 6.5.) Furthermore, in each q^2 region, the number of background events after the $K\mu\mu$-veto and signal-veto ($N^{\text{aftercut}}_{SB_i,tot}$) is known in each side-band window.

In the mid q^2 region the number of background events in the signal region can therefore

be calculated with (for shortness $m = m^{\mathrm{rec}}_{B^0}$)

$$N^{\text{signal region, mid } q^2}_{\text{bkg}} = \sum_{\mathrm{SB}_i=1}^{5} \left[f^{\mathrm{SB_i}}_{J/\psi,\text{after cut}} \cdot N^{\text{after cut}}_{i,tot} \cdot \frac{\int \mathcal{P}^{\text{beforecut}}_{i,J/\psi}}{\int \mathcal{P}^{\text{after cut}}_{i,J/\psi}} \right] \cdot \frac{\int_{m=s1}^{m=s2} \mathcal{M}_{J/\psi}(m)dm}{\int_{m=b1}^{m=b2} \mathcal{M}_{J/\psi}(m)dm} + \sum_{\mathrm{SB}_i=1}^{5} \left[(1 - f^{\mathrm{SB_i}}_{J/\psi}) * N^{aftercut}_{i,tot} \cdot \frac{\int \mathcal{P}^{\text{before cut}}_{i,c}}{\int \mathcal{P}^{\text{after cut}}_{i,c}} \right] \cdot \frac{\int_{m=s1}^{m=s2} \mathcal{M}_{c}(m)dm}{\int_{m=b1}^{m=b2} \mathcal{M}_{c}(m)dm} \tag{6.21}$$

where the first term calculates the number of J/ψ background events in the signal region and the second term calculates the number of fully combinatorial background events in the signal region. The sum indicates the sum over the five SB regions (SB_1,..., SB_5). The expression in the first square brackets calculates the number of background J/ψ events in SB_i by multiplying the total number of background events in SB_i with the value of the J/ψ fraction in SB_i (see equation 6.18) and then correcting this yield for the effect of the $K\mu\mu$ veto and signal veto by multiplying the yield with the ratio of the integrals of the background J/ψ PDF without and with adjustment for the missing phase-space. The sum of veto-corrected J/ψ yields is then extrapolated into the signal region using the ratio of the integral of $\mathcal{M}_{J/\psi}$ (which is simply an exponential with slope parameter $s^m_{J/\psi}$) in the signal region ($s_1 = 5239.58$ MeV/c^2, $s_2 = 5319.58$ MeV/c^2) over the integral of $\mathcal{M}_{J/\psi}$ in the full side-band ($b_1 = 5440$ MeV/c^2, $b_2 = 5840$ MeV/c^2). The second term *for the fully combinatorial background) is analogous.

The calculation of the number of background events in the signal region in the high q^2 region ($N^{\text{signal region, high } q^2}_{\text{bkg}}$) is done analogously to equation 6.21. The calculation of $N^{\text{signal region, low } q^2}_{\text{bkg}}$ is simpler, since there is only one background contribution and thus there is only the second term containing the fully combinatorial background PDF and there is no background fraction needed.

The advantage of this method is that it uses a correction for the effect of the $K\mu\mu$-veto and signal veto and cleanly separates the individual background components, before extrapolating the yield of each component using the corresponding slope. The disadvantage is that the method only uses information from the upper mass side-band, and furthermore uses an extrapolation down into the signal region along a steeply falling exponential.

The method is tested using toy simulations. To this end, 1000 toy samples with ~ 1 million events each (approximately equivalent to the number of events in the merged Run 1 and Run 2 dataset) are generated using the full signal (including acceptance and resolution effects) and background model. Each toy is fitted with the simultaneous side-band fit described in this chapter, and the signal fraction in each q^2 region is calculated

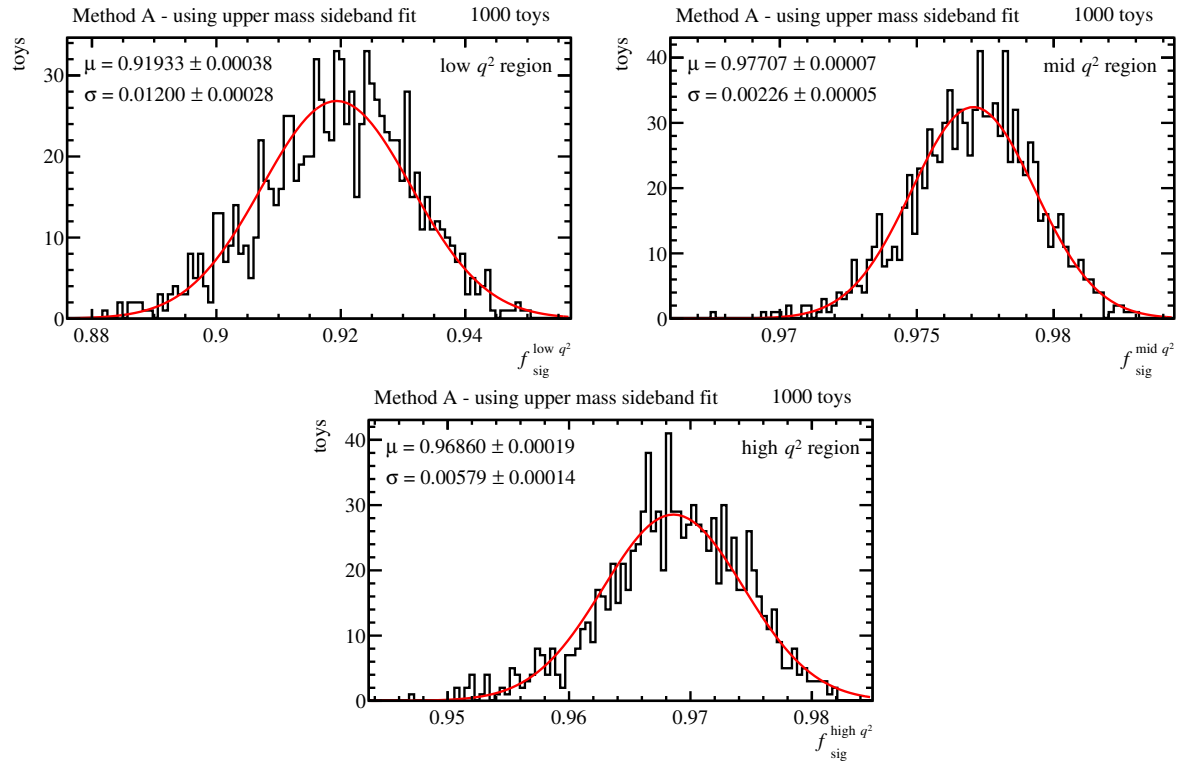


Figure 6.19: Distribution of the signal fraction per q^2 region determined via method A – using the information from the simultaneous side-band fit – in fits to 1000 toy simulations of ~ 1 million events each. Also shown are Gaussian fits to the distributions of the signal fractions.

for each fit. The resulting values for the signal fraction per q^2 region are shown in figure 6.19 along with Gaussian fits to the f_{sig}^{i} distributions. Based on the standard deviation obtained via the Gaussian fits, the precision on the signal fractions for the merged Run 1+Run 2 dataset is 1.3% for $f_{\text{sig}}^{\text{low}\,q^2}$, 0.02% for $f_{\text{sig}}^{\text{mid}\,q^2}$, and 0.6% for $f_{\text{sig}}^{\text{high}\,q^2}$. However, the distributions of f_{sig}^{i} are slightly asymmetric, most notably for $f_{\text{sig}}^{\text{low}\,q^2}$, which would make the uncertainties of the signal parameters asymmetric as well and therefore complicate the interpretation of the results. Furthermore, the statistical power of method A is inferior to the alternative method described in the following section.

6.8.2 Method B - Performing separate $m_{B^0}^{\text{rec}}$ fits

In this method, one dimensional fits to $m_{B^0}^{\text{rec}}$ are performed across the full range of $(5220 \leq m_{B^0}^{\text{rec}} \leq 5840)$ MeV/c^2 in each q^2 region, while modeling both the signal and background components. The resulting background slopes and signal fractions are then used to estimate the number of background events in the signal region.

region	short notation	q^2 range [GeV^2/c^4]
low-q^2	QR1	$0.18 \leq q^2 \leq 3.24$
resonant mid-q^2	$\mathrm{QR2_{res}}$	$8.20 \leq q^2 \leq 10.6$
fully combinatorial mid-q^2	$\mathrm{QR2_{comb}}$	$3.24 \leq q^2 \leq 8.20$ and $10.6 \leq q^2 \leq 11.56$
resonant high-q^2	$\mathrm{QR3_{res}}$	$12.4 \leq q^2 \leq 14.4$
fully combinatorial high-q^2	$\mathrm{QR3_{comb}}$	$11.56 \leq q^2 \leq 12.4$ and $14.4 \leq q^2 \leq 18$

Table 6.5: q^2 regions for the simultaneous $m^{\mathrm{rec}}_{B^0}$-fit. Unconstrained q^2 is used here.

Isolating the three background components

Since the three background components can have different $m^{\mathrm{rec}}_{B^0}$ slopes, the mid-q^2 and high-q^2 regions are split up into sub regions to isolate each component. The q^2 windows of QR1, $\mathrm{QR2_{comb}}$, $\mathrm{QR2_{res}}$, $\mathrm{QR3_{comb}}$, and $\mathrm{QR3_{res}}$ are given in table 6.5.

Due to the effect of the mass constraint on the q^2 distribution of the background, described in section 6.2, the unconstrained q^2 dimension is used for splitting up the data into these regions. Using unconstrained q^2 allows choosing the q^2 ranges such that all events from signal $B^0 \to J/\psi K^{*0}$ ($B^0 \to \psi(2S)K^{*0}$) events as well as all background J/ψ ($\psi(2S)$) events are included in the resonant regions, while making the resonant regions as narrow as possible. Since there are also fully combinatorial events (which may have a different slope) contained in the resonant regions, there is a small inherent bias on the background slopes with this method, which is taken into account as a systematic uncertainty as described in section 7.2.1.[9]

[9]If the B^0-mass constrained q^2 would be used to split the data into q^2 regions, the position and widths of the resonant background peaks would be correlated with $m^{\mathrm{rec}}_{B^0}$. This would cause the background peaks in the upper mass side-band to be shifted and to be wider than the signal $B^0 \to J/\psi K^{*0}$ and $B^0 \to \psi(2S)K^{*0}$ peaks in the signal region. This would make it necessary to choose a much wider resonant q^2 window to include both components fully, which would in turn dilute the sample of resonant events with even more fully combinatorial events. In addition to decreasing the statistical precision of s^m_{comb} this would also increase the systematic uncertainty on $s^m_{J/\psi}$ and $s^m_{\psi(2S)}$.

Simultaneous $m^{\text{rec}}_{B^0}$ fit

The $m^{\text{rec}}_{B^0}$ fit is performed simultaneously to the five q^2 regions defined in table 6.5. The total PDF in each q^2 region j ($j \in \{\text{QR1}, \text{QR2}_{\text{res}}, \text{QR2}_{\text{comb}}, \text{QR3}_{\text{res}}, \text{QR3}_{\text{comb}}\}$) is given by

$$\mathcal{P}^j_{\text{total}}(m^{\text{rec}}_{B^0}) = f^{\text{full region},j}_{\text{sig}} \cdot \mathcal{P}^j_{\text{signal}}(m^{\text{rec}}_{B^0}) + (1 - f^{\text{full region, }j}_{\text{sig}}) \cdot \mathcal{P}^j_{\text{bkg}}(m^{\text{rec}}_{B^0}), \tag{6.22}$$

where the signal component is parameterised by

$$\mathcal{P}^j_{\text{signal}}(m|\vec{\lambda}) = f_{\text{core}}\mathcal{P}_{\text{CB}}(m|\mu, \sigma_1, \alpha, n) + (1 - f_{\text{core}})\mathcal{P}_{\text{CB}}(m|\mu, \sigma_2, \alpha, n), \tag{6.23}$$

i.e. a sum of two Crystal Ball functions with shared mean (μ) and tail parameters (α, n) but different widths (σ_1 and σ_2). The background $\mathcal{P}^j_{\text{bkg}}(m^{\text{rec}}_{B^0})$ is modelled with a falling exponential with slope s^j.

All parameters of the simultaneous $m^{\text{rec}}_{B^0}$ fit are summarised in table 6.6. In the resonant mid-q^2 and resonant high-q^2 regions, the fit is sensitive to events from $B^0_s \to J/\psi\, K^{*0}$ and $B^0_s \to \psi(2S)K^{*0}$ decays respectively. Thus, in these regions, a second signal component is included which uses the same parameterisation as given in equation 6.23 with a mean μ shifted by $\Delta m = m(B^0_s) - m(B^0) = 87.19\,\text{MeV}/c^2$ [30]. The fraction f_{B^0/B_s} of B^0 decays over B^0_s decays is floated in the fit and shared between the regions QR2_{res} and QR3_{res}. The contribution from rare $B^0_s \to K^{*0}\mu^+\mu^-$ decays is approximately 1% of the $B^0 \to K^{*0}\mu^+\mu^-$ signal decays and can be neglected [100]. Therefore, no contributions from B^0_s decays need to be included in the regions QR1, QR2_{comb}, and QR3_{comb}.

As part of the analysis for Ref. [96], the q^2 dependence of the $m^{\text{rec}}_{B^0}$ signal shape was investigated using Monte Carlo simulated samples of $B^0 \to K^{*0}\mu^+\mu^-$ and $B^0 \to J/\psi\, K^{*0}$ decays. It was found that all parameters except for the widths σ_1 and σ_2 are independent of q^2. Based on the MC simulations, the widths are approximately 1.038 times wider in the high-q^2 region of ($11.56 \leq q^2 \leq 18$) GeV^2/c^4 than in the lower q^2 regions [91]. Thus, during the simultaneous $m^{\text{rec}}_{B^0}$-fit in this analysis, all signal parameters are shared across the five q^2 regions and the width parameters (σ and σ_2) are multiplied with 1.038 in QR3_{comb} and QR3_{res}. The signal fractions $f^{\text{full region},j}_{\text{sig}}$ are independently floating in all five q^2 regions.

The background slopes in the regions QR1, QR2_{comb}, and QR3_{comb} have been found to be statistically compatible and therefore the parameters s^{QR1}, $s^{\text{QR2}_{\text{comb}}}$, $s^{\text{QR3}_{\text{comb}}}$ are replaced with a single parameter which is shared across the three regions. Figure 6.20 shows the $m^{\text{rec}}_{B^0}$ distributions in the five q^2 regions as well as the projection of the PDFs from the simultaneous fit to the data.

name	QR1	$\mathrm{QR2_{comb}}$	$\mathrm{QR2_{res}}$	$\mathrm{QR3_{comb}}$	$\mathrm{QR3_{res}}$
s^{QR1}	x				
$s^{\mathrm{QR2_{comb}}}$		x			
$s^{\mathrm{QR2_{res}}}$			x		
$s^{\mathrm{QR3_{comb}}}$				x	
$s^{\mathrm{QR3_{res}}}$					x
f_s^{QR1}	x				
$f_s^{\mathrm{QR2\_COMB}}$		x			
$f_s^{\mathrm{QR2\_RES}}$			x		
$f_s^{\mathrm{QR3\_COMB}}$				x	
$f_s^{\mathrm{QR3\_RES}}$					x
$\mu^{m,\mathrm{signal}}$	x	x	x	x	x
$\sigma^{m,\mathrm{signal}}$	x	x	x	x ($\times$1.038)	x ($\times$1.038)
$\sigma_2^{m,\mathrm{signal}}$	x	x	x	x ($\times$1.038)	x ($\times$1.038)
$\alpha^{m,\mathrm{signal}}$	x	x	x	x	x
f_{core}	x	x	x	x	x
f_{B^0/B_s}			x		x

Table 6.6: Fit parameters of the simultaneous $m_{B^0}^{\mathrm{rec}}$ fit. The region(s) in which each parameter is/are determined is indicated. Parameters that are determined in several regions are shared parameters in the fit. The sigma parameters are multiplied with 1.038 in the high q^2 regions to account for the q^2 dependence of the B^0 peak as discussed in the text.

Calculation of the background yield in the signal region

The number of background events $N_{\mathrm{total}}^{\mathrm{signal\ region},j}$ in the signal region in each of the five q^2 regions ($j \in \{\mathrm{QR1}, \mathrm{QR2_{res}}, \mathrm{QR2_{comb}}, \mathrm{QR3_{res}}, \mathrm{QR3_{comb}}\}$) is then calculated by interpolating the background yield in the full $m_{B^0}^{\mathrm{rec}}$ range into the signal region with

$$N_{\mathrm{bkg}}^{\mathrm{signal\ region},j} = (1 - f_{\mathrm{sig}}^{\mathrm{full\ region},j}) \cdot N_{\mathrm{total}}^{\mathrm{full\ region},j} \cdot \frac{\int_{m_{B^0}^{\mathrm{rec}}=5239.58\ \mathrm{MeV}/c}^{m_{B^0}^{\mathrm{rec}}=5319.58\ \mathrm{MeV}/c^2} \mathcal{P}_{\mathrm{bkg}}^{j}(m_{B^0}^{\mathrm{rec}})}{\int_{m_{B^0}^{\mathrm{rec}}=5220\ \mathrm{MeV}/c^2}^{m_{B^0}^{\mathrm{rec}}=5840\ \mathrm{MeV}/c^2} \mathcal{P}_{\mathrm{bkg}}^{j}(m_{B^0}^{\mathrm{rec}})} \tag{6.24}$$

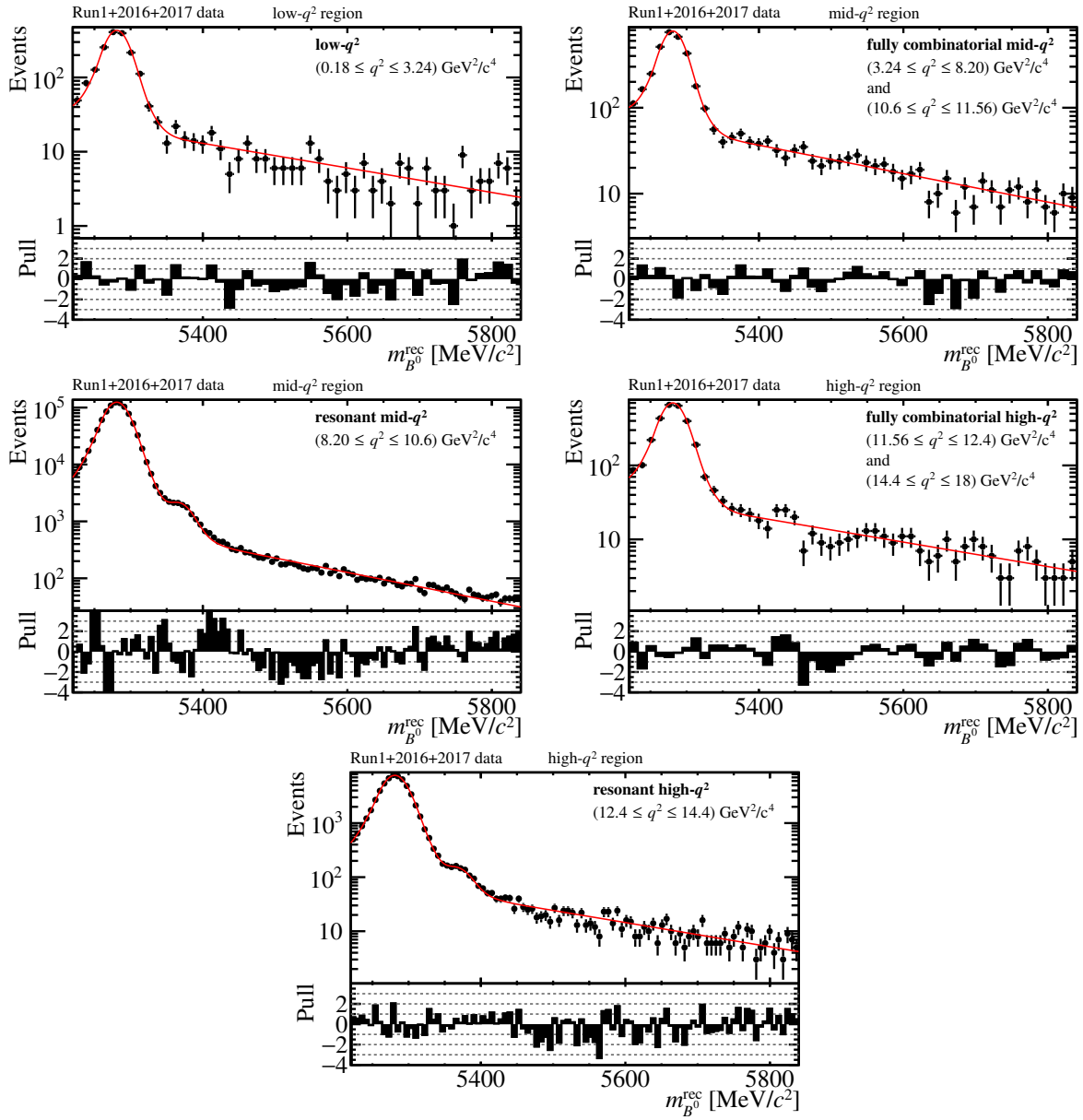


Figure 6.20: $m^{\rm rec}_{B^0}$ distribution in the five q^2 regions of the simultaneous $m^{\rm rec}_{B^0}$ fit of the merged Run 1 and Run 2 data. Also shown are the projections of the simultaneous $m^{\rm rec}_{B^0}$-only fit to those regions. The signal parameters are shared across all q^2 regions and the width parameters (σ and σ_2) are multiplied with a factor of 1.0385 in the high-q^2 regions to account for the q^2 dependence of the $m^{\rm rec}_{B^0}$ resolution. The slope of the fully combinatorial background is shared across low-, mid- and high-q^2, whereas the slopes in the resonant regions are independent.

where $N^{\text{full region},j}_{\text{total}}$ is the total number of events within the full $m^{\rm rec}_{B^0}$ range. $\mathcal{P}^j_{\rm bkg}(m^{\rm rec}_{B^0})$ is a falling exponential with slope s^j.

Finally, the number of background events in the signal region in the low q^2 region is given by

$$N_{\text{bkg}}^{\text{signal region, low } q^2} = N_{\text{total}}^{\text{signal region, QR1}}, \tag{6.25}$$

while the number of background events in the signal region in the mid q^2 region is calculated with

$$N_{\text{bkg}}^{\text{signal region, mid } q^2} = N_{\text{total}}^{\text{signal region, QR2}_{\text{res}}} + N_{\text{total}}^{\text{signal region, QR2}_{\text{comb}}}. \tag{6.26}$$

Analogously, the number of background events in the signal region in the high q^2 region is calculated with

$$N_{\text{bkg}}^{\text{signal region, high } q^2} = N_{\text{total}}^{\text{signal region, QR3}_{\text{res}}} + N_{\text{total}}^{\text{signal region, QR3}_{\text{comb}}}. \tag{6.27}$$

These estimated numbers of background events in the signal region are then used to calculate the signal fraction in the signal region using equation 6.20.
Figure 6.21 shows the signal fraction per q^2 region determined in fits to 1000 toy samples with $\sim$ 1 million events each. Also shown are Gaussian fits to the f_{sig}^i distributions to determine the mean and standard deviation of the signal fraction obtained with the method described in this section. The mean value of each f_{sig}^i is in good agreement with the *true* values of f_{sig}^i given by

$$\begin{aligned} \text{true}(f_{\text{sig}}^{\text{low } q^2}) &= 0.8955 \pm 0.0052 \\ \text{true}(f_{\text{sig}}^{\text{mid } q^2}) &= 0.97393 \pm 0.0034 \\ \text{true}(f_{\text{sig}}^{\text{high } q^2}) &= 0.9626 \pm 0.0063. \end{aligned} \tag{6.28}$$

The true values were obtained by generating a single toy with 10 million events and counting the number of signal and background events in the signal region. The uncertainties given in equation 6.28 are the propagated Poisson uncertainties.
Based on the standard deviation obtained in the Gaussian fits in figure 6.21, the statistical precision on the signal fractions for the merged Run 1+Run 2 dataset is 0.04% for $f_{\text{sig}}^{\text{low } q^2}$, 0.003% for $f_{\text{sig}}^{\text{mid } q^2}$, and 0.01% for $f_{\text{sig}}^{\text{high } q^2}$.

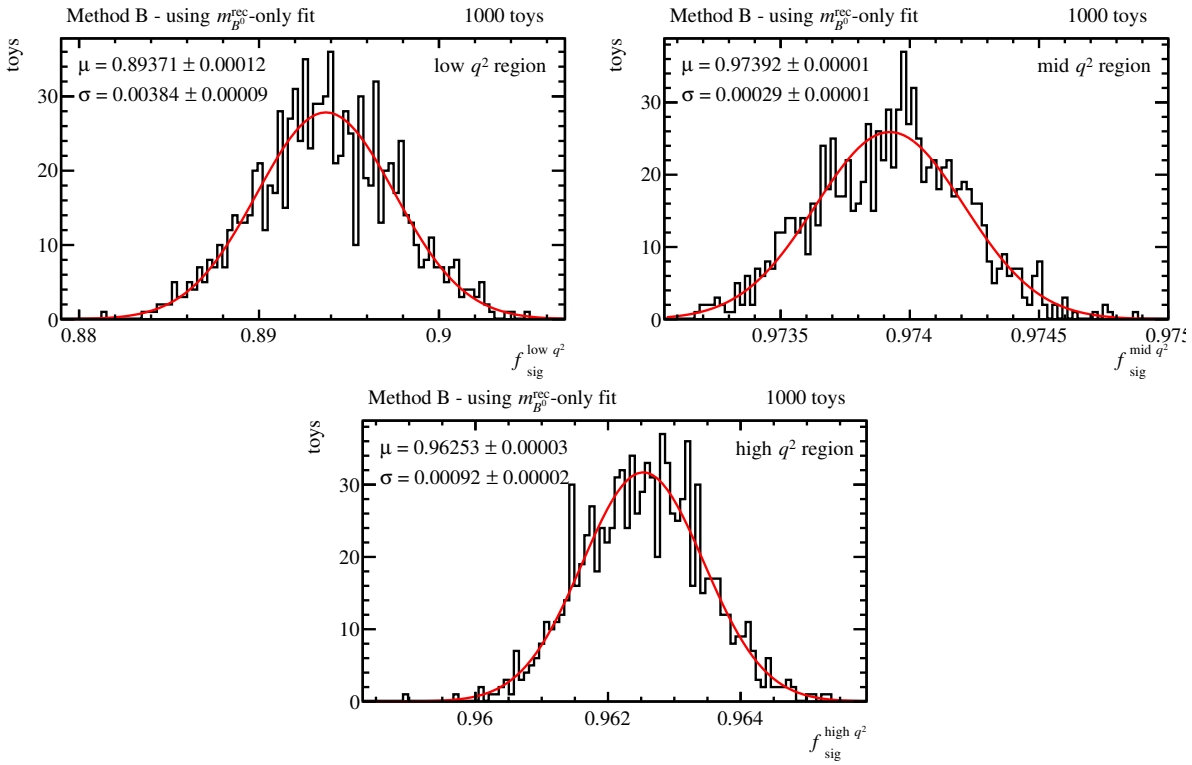


Figure 6.21: Distribution of signal fraction per q^2 region determined via method B – using the information from the simultaneous $m^{\text{rec}}_{B^0}$ fit – in fits to 1000 toy simulations of ~ 1 million events each. Also shown are Gaussian fits to the f^i_{sig} distributions and the resulting mean μ and σ values.

6.9 Summary of the fitting strategy

In this section, the fit strategy used in this thesis is briefly summarised.

(**step 1**) $m_{B^0}^{\mathrm{rec}}$ fit

(**step 2**) side-band fit

(**step 3**) signal-region fit.

In **step 1**, the $m_{B^0}^{\mathrm{rec}}$ fit is performed in five simultaneous q^2 regions as described in section 6.8.2.

In **step 2**, the 5D ($\cos\theta_\ell$,$\cos\theta_K$,ϕ,q^2,$m_{B^0}^{\mathrm{rec}}$) side-band fit is performed simultaneously in five SB regions and three q^2 regions, while applying the correction for the effect of the $K\mu\mu$-veto (see section 6.4) and the signal-veto (see section 6.5), and while allowing for a mass dependence of the $\cos\theta_K$ parameters (see section 6.7).
During the fit in **step 2**, the $m_{B^0}^{\mathrm{rec}}$-slope parameters of the three background components are constrained to the slopes determined in **step 1**. All other parameters from **step 1** are also floated again and constrained using the full covariance matrix (14×14) determined in **step 1**. The constraint is implemented by adding the constraint term c to the log-Likelihood, where c is given by

$$c = -0.5 \cdot \vec{\lambda}_{\mathrm{div}}(K_{\lambda\lambda}^{-1}\vec{\lambda}_{\mathrm{div}}), \tag{6.29}$$

where $K_{\lambda\lambda}$ is the covariance matrix from the simultaneous $m_{B^0}^{\mathrm{rec}}$ fit and $\vec{\lambda}_{\mathrm{div}}$ is given by

$$\vec{\lambda}_{\mathrm{div}} = \begin{pmatrix} s^m_{\mathrm{comb}} \\ s^m_{J/\psi} \\ s^m_{\psi(2S)} \\ f^{\text{full region},j}_{\mathrm{sig}} \\ \vec{\lambda}^{m_{B^0}^{\mathrm{rec}}}_{\text{signal model}} \end{pmatrix} - \begin{pmatrix} s^{\mathrm{QR1}} \\ s^{\mathrm{QR2_{res}}} \\ s^{\mathrm{Q2_{res}}} \\ f^{\text{full region},j}_{\mathrm{sig}} \\ \vec{\lambda}^{m_{B^0}^{\mathrm{rec}}}_{\text{signal model}} \end{pmatrix}_{\text{best fit values from step 1}} . \tag{6.30}$$

with $j \in \{\mathrm{QR1}, \mathrm{QR2_{res}}, \mathrm{QR2_{comb}}, \mathrm{QR3_{res}}, \mathrm{QR3_{comb}}\}$. Note that s^{QR1} represents the single slope parameter that is determined as a shared parameter in QR1, $\mathrm{QR2_{comb}}$, and $\mathrm{QR3_{comb}}$. $\vec{\lambda}^{m_{B^0}^{\mathrm{rec}}}_{\text{signal model}}$ denotes the signal parameters from **step 1**.

Except for the slope parameters, the parameters from the simultaneous $m_{B^0}^{\mathrm{rec}}$ fit have no influence on the background parameters and are purely constrained from the covariance

matrix. The reason for floating them again in **step 2**, is that the simultaneous side-band fit then yields a single covariance matrix (46×46) containing all parameters from both **step 1** and **step 2**. This way all potential correlations are taken into account.

In **step 3** the signal and background ($\cos\theta_\ell$, $\cos\theta_K$, ϕ, q^2) fit is performed in the signal region ($(5239.58 \leq m^{\text{rec}}_{B^0} \leq 5319.58)$ MeV/c^2). All parameters contained in the (46×46) covariance matrix from **step 2** are floated in this fit and constrained to the best fit values from **step 2** using the (46×46) covariance matrix. The information obtained from fitting the side-band is therefore used via the constraint, while the background events in the signal region are adding further information about the background shape. The signal fraction is calculated for each likelihood call using the expressions in equations 6.20 and 6.24.

Since **step 3** requires the full fit with the signal model, it cannot be carried out for the merged Run 1+Run 2 dataset as explained in chapter 5. However, it is carried out for the fit to Run 1 data presented in chapter 7.

6.10 Test of the background fit using toy simulations

To test the stability of the fits described in this chapter, a pull study is performed. For this study, 1000 toy samples with ~ 1 million events each (approximately equivalent to the number of events in merged the Run 1 and Run 2 dataset) are generated using the full signal (including acceptance and resolution effects) and background model. Each toy sample is fitted with the simultaneous $m^{\text{rec}}_{B^0}$ fit and the simultaneous side-band fit, i.e. **step 1** and **step 2** described in section 6.9.
The ($\cos\theta_K$, q^2, $m^{\text{rec}}_{B^0}$)-cut which mimics the $K\mu\mu$-veto and the signal-veto are applied to the simulated toys and the background PDF is adjusted accordingly.

The pull of each parameter is calculated for each of the 1000 fits using equation 6.10. The pull distribution of each parameter is fitted with a Gaussian to determine the mean μ_{pull} and standard deviation σ_{pull}. Table 6.7 shows μ_{pull} and σ_{pull} for all background parameters which were floating in the fits to the toy samples. The width of the pull distribution of all parameters is compatible with unity, indicating that for each parameter the statistical uncertainty obtained in the fit, agrees well with one standard deviation.
For an unbiased parameter one expects $\mu_{\text{pull}} = 0$. Most parameters are unbiased. The largest biases are observed for the slope parameters s^m_{comb}, $s^m_{J/\psi}$, $s^m_{\psi(2S)}$ which are due to the fact that they are constrained to the slopes determined in the $m^{\text{rec}}_{B^0}$ fit which are biased

by construction (see section 6.8.2). Note however, that the statistical uncertainties of the slope parameters are on the order of 1-4%.[10]

The biases of the background fractions $f_{J/\psi}$ and $f_{\psi(2S)}$ are due to the fact that the value of each background fraction in a given SB window depends on the slope parameters (see equation 6.18) and therefore the biases in the slopes lead to biases in the background fractions. The small biases observed for some of the angular and q^2 parameters are due to correlations of those parameters with the slopes and fractions.

The effect of the biases of the background parameters on the signal parameters is taken into account as a systematic uncertainty as described in section 7.2.1.

[10]Since the pull is the difference of the best fit value and the true value divided by the statistical uncertainty, a bias of $\mu_{\text{pull}} \approx 2$ equates to a 2-8% shift of the slope parameter values.

	μ_{pull}	σ_{pull}
s^{m}_{comb}	1.923 ± 0.033	1.038 ± 0.024
$s^{m}_{J/\psi}$	-2.418 ± 0.032	0.993 ± 0.024
$s^{m}_{\psi(2S)}$	-2.520 ± 0.034	1.034 ± 0.026
$a^{\text{L}}_{\text{comb}}$	0.016 ± 0.030	0.943 ± 0.021
$b^{\text{L}}_{\text{comb}}$	-0.005 ± 0.031	0.994 ± 0.022
$a^{\text{K}}_{\text{comb}}$	0.205 ± 0.032	1.016 ± 0.023
$b^{\text{K}}_{\text{comb}}$	0.399 ± 0.031	0.981 ± 0.022
a^{ϕ}_{comb}	0.033 ± 0.032	1.019 ± 0.023
b^{ϕ}_{comb}	0.004 ± 0.032	1.024 ± 0.023
$a^{q^2}_{\text{comb}}$	0.035 ± 0.032	1.013 ± 0.023
$a^{\text{L}}_{J/\psi}$	-0.003 ± 0.031	0.975 ± 0.022
$b^{\text{L}}_{J/\psi}$	0.073 ± 0.032	1.000 ± 0.022
$a^{\text{K}}_{J/\psi}$	-0.070 ± 0.031	0.993 ± 0.022
$b^{\text{K}}_{J/\psi}$	-0.118 ± 0.032	1.001 ± 0.022
$a^{\phi}_{J/\psi}$	-0.036 ± 0.032	1.026 ± 0.023
$b^{\phi}_{J/\psi}$	-0.005 ± 0.032	0.997 ± 0.022
$\mu^{q^2}_{J/\psi}$	0.014 ± 0.032	1.008 ± 0.023
$\sigma^{q^2}_{J/\psi}$	-0.102 ± 0.031	0.983 ± 0.022
$\alpha^{q^2}_{J/\psi}$	0.008 ± 0.032	0.997 ± 0.022

	μ_{pull}	σ_{pull}
$a^{\text{L}}_{\psi(2S)}$	-0.008 ± 0.032	1.010 ± 0.023
$b^{\text{L}}_{\psi(2S)}$	-0.013 ± 0.032	1.020 ± 0.023
$a^{\text{K}}_{\psi(2S)}$	-0.107 ± 0.032	1.005 ± 0.022
$b^{\text{K}}_{\psi(2S)}$	-0.161 ± 0.031	0.968 ± 0.022
$a^{\phi}_{\psi(2S)}$	-0.010 ± 0.033	1.030 ± 0.023
$b^{\phi}_{\psi(2S)}$	0.027 ± 0.033	1.036 ± 0.023
$\mu^{q^2}_{\psi(2S)}$	-0.050 ± 0.033	1.037 ± 0.023
$\sigma^{q^2}_{\psi(2S)}$	-0.084 ± 0.032	1.016 ± 0.023
$f_{J/\psi}$	-0.989 ± 0.026	0.837 ± 0.019
$f_{\psi(2S)}$	-1.159 ± 0.026	0.812 ± 0.018
$a^{\text{K}}_{\text{md,comb}}$	-0.193 ± 0.031	0.992 ± 0.022
$b^{\text{K}}_{\text{md,comb}}$	-0.374 ± 0.031	0.981 ± 0.022
$a^{\text{K}}_{\text{md},J/\psi}$	0.090 ± 0.032	1.004 ± 0.022
$b^{\text{K}}_{\text{md},J/\psi}$	0.120 ± 0.032	1.000 ± 0.022
$a^{\text{K}}_{\text{md},\psi(2S)}$	0.113 ± 0.032	1.000 ± 0.022
$b^{\text{K}}_{\text{md},\psi(2S)}$	0.188 ± 0.030	0.957 ± 0.021

Table 6.7: Fit results of Gaussian fits to the pull distributions of all background parameters in fits to 1000 toy simulations. For an unbiased parameter one expects $\mu_{\text{pull}} = 0$. The observed biases are discussed in the text. All values for σ_{pull} are compatible with 1, demonstrating that for each parameter the uncertainty obtained by the fits agrees well with one standard deviation.

6.11 Background fit to the merged Run 1 and Run 2 data

In this section the results of the background fit to the merged Run 1 and Run 2 dataset are presented. As explained in section 6.9 the simultaneous $m^{\text{rec}}_{B^0}$ fit is performed first and the projections of this fit is shown in figure 6.20. The results of all parameters floating during the 5D simultaneous side-band fit are given in table 6.8.

Figures 6.22, 6.23, and 6.24 show some exemplary distributions of the angular and q^2 distributions in some of the ($m^{\text{rec}}_{B^0}$, q^2) regions as well as the corresponding projections of the background PDF. The full set of distributions and fit projections for all five SB regions and all three q^2 regions are shown in appendix A.3. The background PDF describes the data well in all kinematic windows. The $m^{\text{rec}}_{B^0}$ and q^2 dependent gap in the $\cos\theta_K$ distribution caused by the $K\mu\mu$ veto is well described by the PDF. Furthermore, the $m^{\text{rec}}_{B^0}$-region dependence of the shape of the $\cos\theta_K$ distributions is well described.

The background PDF does not seem to capture the slope of the data in the $m^{\text{rec}}_{B^0}$ distribution at $m^{\text{rec}}_{B^0} < 5580\,\text{MeV}$ in the mid q^2 region shown in figure A.19. The local change in slope is most likely due to $B^0 \to J/\psi\, K^{*0}$ signal events leaking into the upper mass side-band (see section 6.5) which is not well described by the double Crystal Ball function used to parameterise the signal peak. However, the $m^{\text{rec}}_{B^0}$ fit over the full $m^{\text{rec}}_{B^0}$ range, shown in figure 6.20, demonstrates that this is a local issue and does not affect the general description of the $m^{\text{rec}}_{B^0}$ distribution of the signal and background.

name	result	name	fit result
s^m_{comb}	$(3.71 \pm 0.16)\times10^{-3}$	$a^{\phi}_{\psi(2S)}$	-0.141 ± 0.070
$s^m_{J/\psi}$	$(5.847 \pm 0.090)\times10^{-3}$	$b^{\phi}_{\psi(2S)}$	0.055 ± 0.067
$s^m_{\psi(2S)}$	$(5.20 \pm 0.21)\times10^{-3}$	$\mu^{q^2}_{\psi(2S)}$	13.6147 ± 0.0092
$a^{\text{L}}_{\text{comb}}$	0.074 ± 0.057	$\sigma^{q^2}_{\psi(2S)}$	0.2073 ± 0.0078
$b^{\text{L}}_{\text{comb}}$	0.223 ± 0.051	$f_{J/\psi}$	0.9528 ± 0.0043
$a^{\text{K}}_{\text{comb}}$	0.25 ± 0.20	$f_{\psi(2S)}$	0.731 ± 0.026
$b^{\text{K}}_{\text{comb}}$	0.21 ± 0.18	$a^{\text{K}}_{\text{md,comb}}$	-0.22 ± 0.28
a^{ϕ}_{comb}	-0.090 ± 0.059	$b^{\text{K}}_{\text{md,comb}}$	-0.11 ± 0.26
b^{ϕ}_{comb}	-0.005 ± 0.056	$a^{\text{K}}_{\text{md},J/\psi}$	-0.05 ± 0.10
$a^{q^2}_{\text{comb}}$	1.080 ± 0.090	$b^{\text{K}}_{\text{md},J/\psi}$	-0.83 ± 0.11
$a^{\text{L}}_{J/\psi}$	-0.028 ± 0.021	$a^{\text{K}}_{\text{md},\psi(2S)}$	-0.16 ± 0.38
$b^{\text{L}}_{J/\psi}$	-0.500 ± 0.023	$b^{\text{K}}_{\text{md},\psi(2S)}$	-0.40 ± 0.41
$a^{\text{K}}_{J/\psi}$	0.242 ± 0.066	f_s^{QR1}	0.755 ± 0.013
$b^{\text{K}}_{J/\psi}$	0.448 ± 0.067	$f_s^{\text{QR2\_COMB}}$	0.96685 ± 0.00060
$a^{\phi}_{J/\psi}$	0.026 ± 0.021	$f_s^{\text{QR3\_COMB}}$	0.658 ± 0.011
$b^{\phi}_{J/\psi}$	-0.022 ± 0.020	$f_s^{\text{QR2\_RES}}$	0.9494 ± 0.0020
$\mu^{q^2}_{J/\psi}$	9.6184 ± 0.0019	$f_s^{\text{QR3\_RES}}$	0.774 ± 0.010
$\sigma^{q^2}_{J/\psi}$	0.1467 ± 0.0016	$\mu^{m,\text{signal}}$	5280.956 ± 0.024
$\alpha^{q^2}_{J/\psi}$	0.965 ± 0.039	$\sigma^{m,\text{signal}}$	15.738 ± 0.083
$a^{\text{L}}_{\psi(2S)}$	-0.118 ± 0.071	$\sigma_2^{m,\text{signal}}$	26.78 ± 0.32
$b^{\text{L}}_{\psi(2S)}$	-0.502 ± 0.080	$\alpha^{m,\text{signal}}$	1.5589 ± 0.0080
$a^{\text{K}}_{\psi(2S)}$	0.34 ± 0.25	f_{core}	0.721 ± 0.012
$b^{\text{K}}_{\psi(2S)}$	0.03 ± 0.27	f_{B^0/B_s}	0.98767 ± 0.00021

Table 6.8: Fit results of the background parameters and the parameters describing the $m^{\text{rec}}_{B^0}$ distribution of the signal, for the merged Run 1 and Run 2 data. The uncertainties are statistical uncertainties.

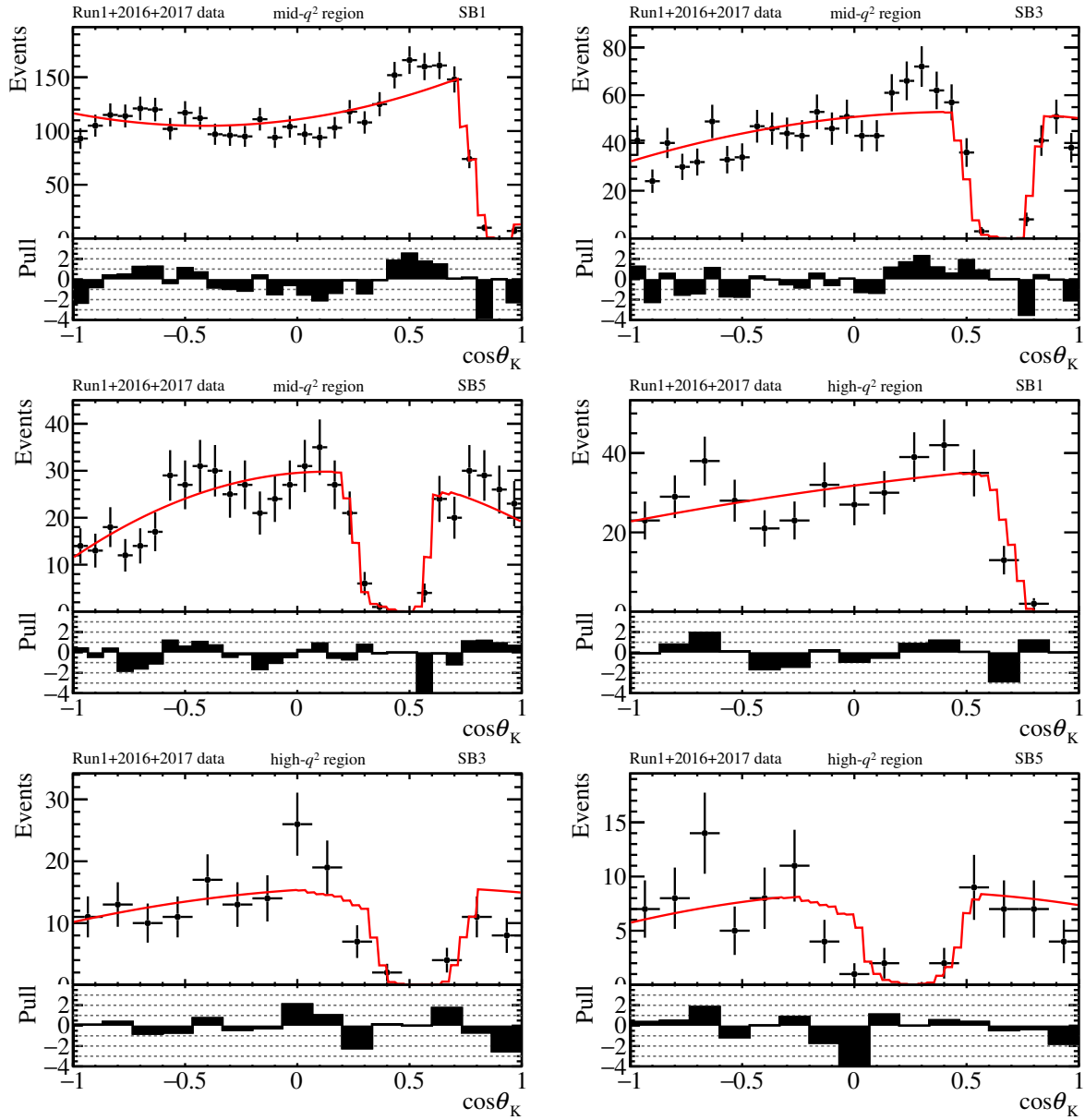


Figure 6.22: $\cos\theta_K$ distributions in the mid, and high-q^2 regions in SB1, SB3, and SB5 in the merged Run 1 and Run 2 data. Also shown are the projections of the simultaneous side-band fit. The $m^{\text{rec}}_{B^0}$-dependent gap in the $\cos\theta_K$ distribution caused by the $K\mu\mu$ veto is well described by the PDF which is adjusted for the missing phase space (see section 6.5. See appendix A.3 for the full set of plots.

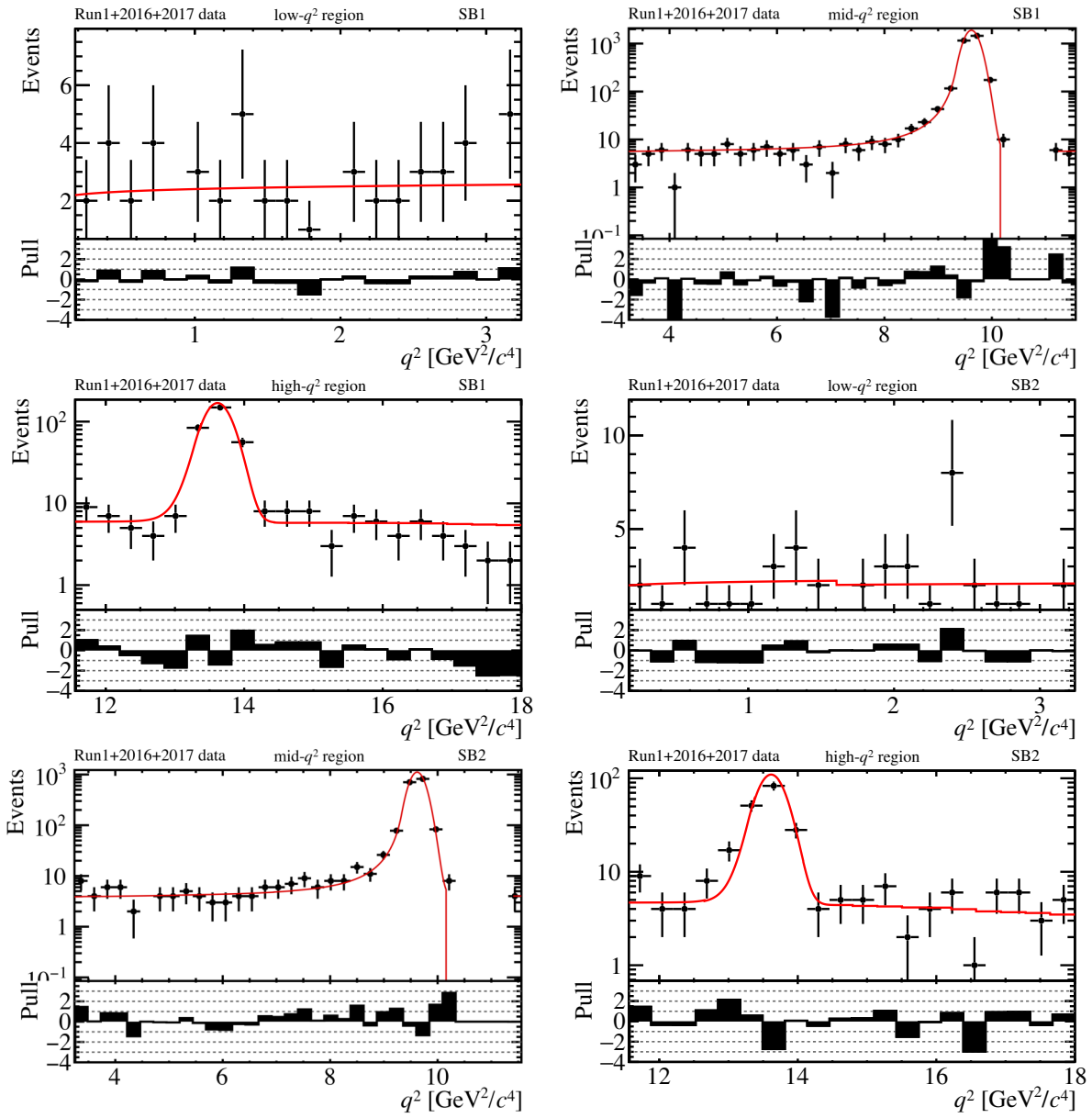


Figure 6.23: q^2 distributions in the low, mid, and high-q^2 regions in SB1 and SB2 Also shown are the projections of the simultaneous side-band fit. See appendix A.3 for the full set of plots. The gaps in the q^2 distribution in the mid-q^2 region are caused by the signal veto explained in section 6.5

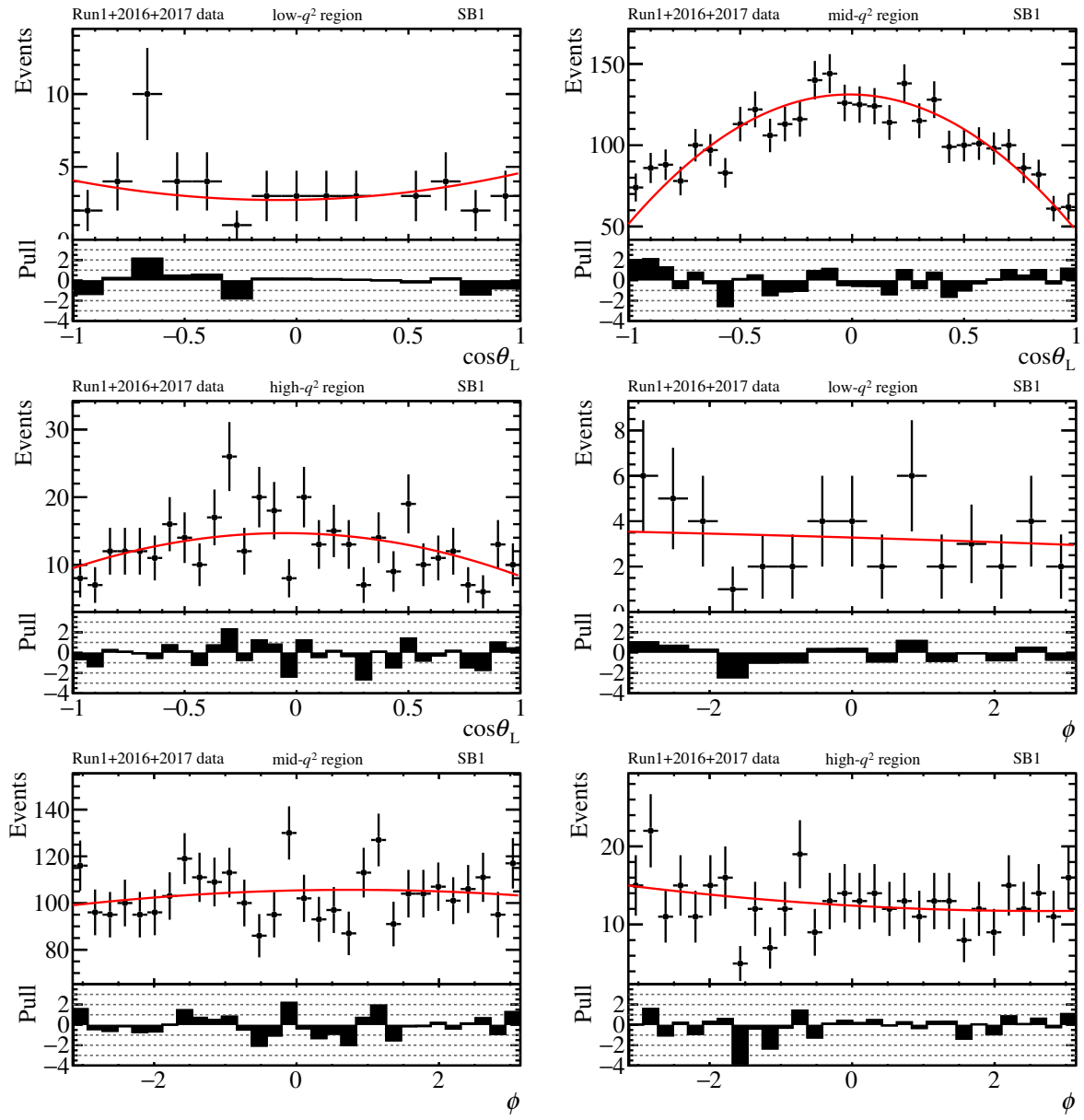


Figure 6.24: $\cos\theta_\ell$ and ϕ distributions in the low, mid, and high-q^2 regions in SB1 in the merged Run 1 and Run 2 data. Also shown are the projections of the simultaneous side-band fit. See appendix A.3 for the full set of plots.

7. Fit to Run 1 data

As described in chapter 5, the full fit with the signal model can currently only be carried out for Run 1 data[1]. The fit to the Run 1 data presented in this chapter, is to be seen as a proof of concept to demonstrate the ability of the model to describe the data. Any physics interpretations are preliminary and in order to draw real conclusions, the published result which will use merged Run 1 and Run 2 data should be awaited.

The signal model, described in chapter 3, is convolved with the resolution model (described in section 5.3.2) and multiplied with the acceptance model (described in section 5.2). The fit procedure is presented in section 6.9. For the fit to Run 1 data all three steps of the fit procedure are carried out. The results of the background fit to the Run 1 data are presented briefly in section 7.1. The systematic uncertainties are discussed in section 7.2 and the results of the full fit to Run 1 data are presented in section 7.3.

7.1 Background fit to Run 1 data

The fit procedure for the fit to the Run 1 data follows the steps outlined in section 6.9. The projections of the simultaneous $m^{\text{rec}}_{B^0}$ fit (step 1) are shown in appendix A.4.
Following the simultaneous $m^{\text{rec}}_{B^0}$ fit, the simultaneous side-band fit is performed (step 2), while floating the parameters of the simultaneous $m^{\text{rec}}_{B^0}$ fit again and constraining them to the results from step 1. The background parameters as well as the parameters of the simultaneous $m^{\text{rec}}_{B^0}$ fit are then floated again in the signal region fit (step 3) and constrained to the best fit values from step 2.
The background events in the signal region therefore add an additional constraint to the background parameters. This additional constraint is significant, since there are more background events in the signal region than in the upper mass side-band. For example, in the mid q^2 region there are ~ 4450 background events in the signal region and ~ 3150 events in the upper mass side-band. However, due to the large number of signal events in the signal region (~ 256300 in the mid q^2 region) the fit is still less sensitive to the background shapes in the signal region than in the upper mass side-band.

In order to study the effect of floating the background parameters in the signal region, the background parameterisation obtained in the signal region fit (step 3) is compared to

[1]The averaged acceptance model for the merged Run 1 and Run 2 has yet to be determined which is beyond the scope of this thesis.

the background parameterisation obtained purely from side-band data (step 2).
Figure 7.1 shows the $\cos\theta_K$, $\cos\theta_\ell$, and ϕ distributions of the Run 1 data in SB1 and SB2 in the mid-q^2 region (the full set of figures for all dimensions and regions can be found in appendix A.5). The projections of the background PDF after the side-band fit (blue) and after the signal region fit (red) are shown as dashed lines. The blue and red shaded areas indicate the 68% confidence interval of each PDF, which was obtained by fluctuating all parameters using the respective covariance matrix.
The background PDFs (after SB fit and after signal region fit) are in good general agreement with some exceptions. The biggest discrepancies are observed for the $\cos\theta_\ell$ and $\cos\theta_K$ distributions in SB1 and SB2 in the mid-q^2 region, indicating a systematic uncertainty related to the extrapolation of the background parameters from the side-band into the signal region as discussed below. The projections in the ϕ dimension fully agree in all regions.
Figure 7.2 shows the q^2 distribution in SB1 in all three q^2 regions as well as the projections and confidence intervals of the background PDF before and after the signal region fit. The only minor disagreement between the two PDFs can be found in the lower tail of the J/ψ peak in the mid-q^2 region.

The comparison of the background PDFs before and after the signal region fit also constitutes a systematic check of the extrapolation of the background parameters along the $m^{\text{rec}}_{B^0}$ dimension (see section 6.7). As the background PDFs before and after the signal region fits are not completely statistically compatible, a systematic uncertainty related to the extrapolation of the background parameters is determined in section 7.2.1.

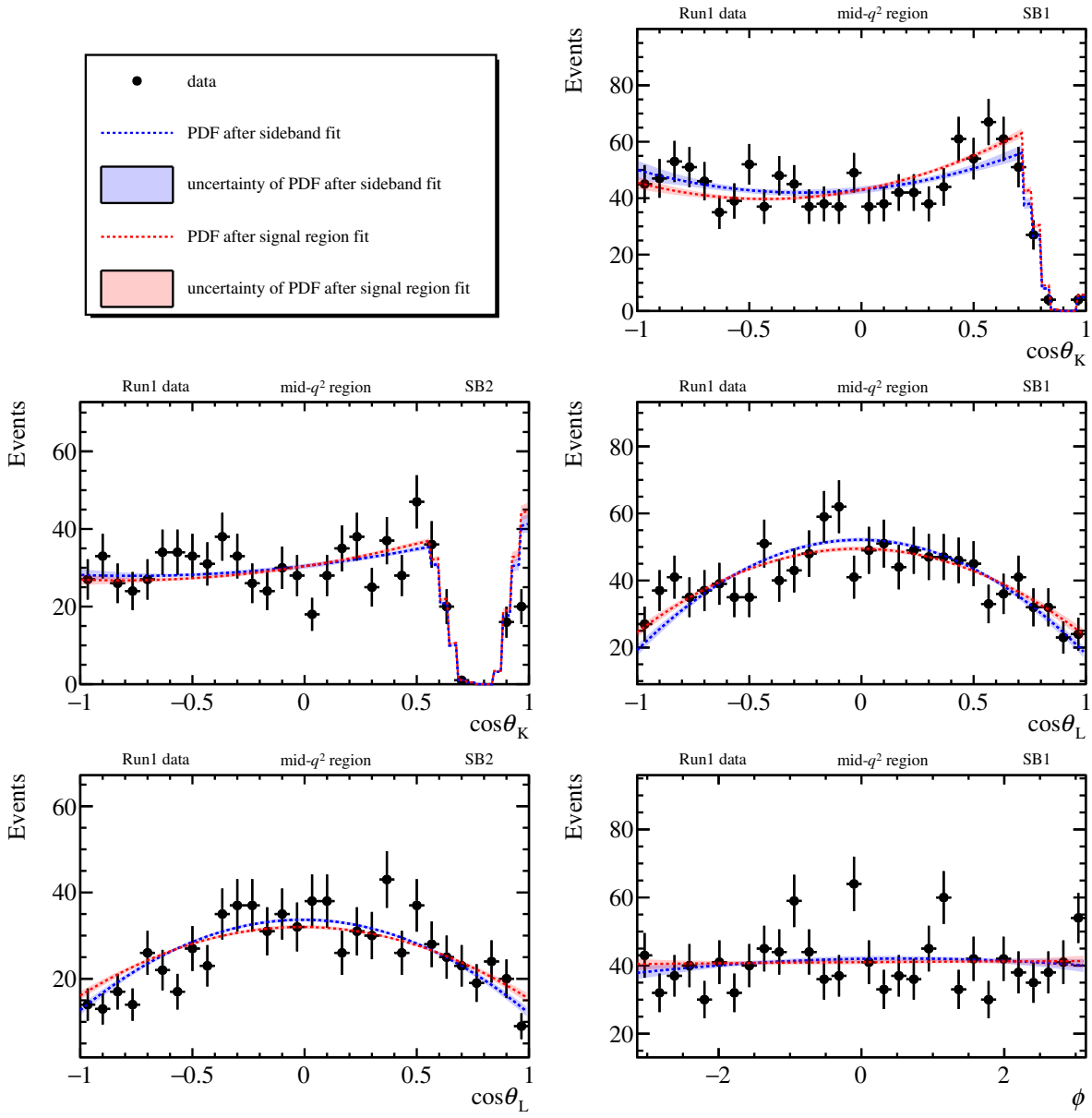


Figure 7.1: $\cos\theta_K$, $\cos\theta_\ell$, and ϕ distributions in the first two upper side-band regions (SB1 and SB2) in the mid-q^2 region in Run 1 data. Also shown are the projections of the background PDF after the simultaneous side-band fit (blue dashed line) as well as the PDFs' 68% confidence interval (blue shaded area), determined by varying all parameters using the covariance matrix obtained in the fit. Furthermore, the background PDF after the signal region fit (red dashed line) and the respective 68% confidence interval (red shaded area) are shown. Note that the background events in the signal region (not shown here) add additional constraints on the background parameters in the signal region fit.

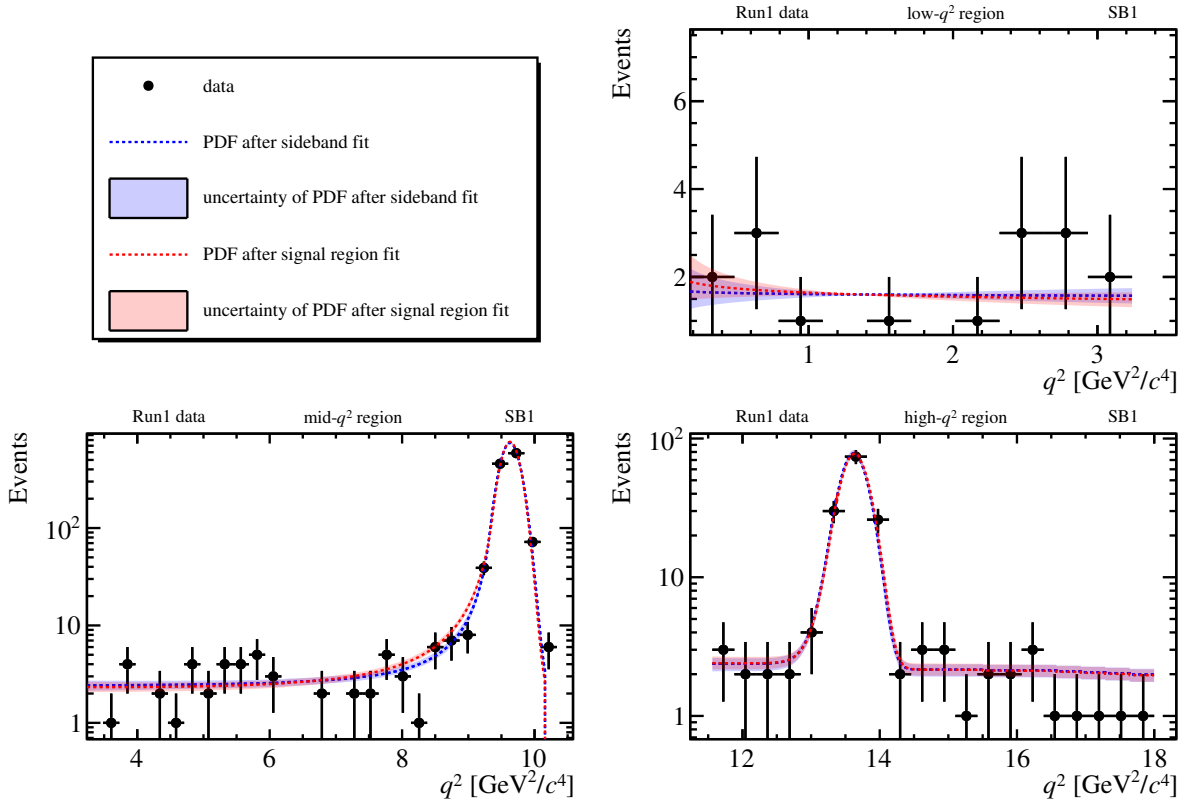


Figure 7.2: q^2 distribution in SB1 in all three q^2 regions in Run 1 data. Also shown are the projections of the background PDF after the simultaneous side-band fit (blue dashed line) as well as the PDFs' 68% confidence interval (blue shaded area) determined by varying all parameters using the covariance matrix obtained in the fit. Furthermore, the background PDF after the signal region fit (red dashed line) and the respective 68% confidence interval (red shaded area) are shown. Note that the background events in the signal region (not shown here) add additional constraints on the background parameters in the signal region fit.

7.2 Systematic uncertainties

7.2.1 Backround parameterisation

There are two major sources of systematic uncertainties of the signal parameters caused by the background parameterisation. Firstly, the fit procedure for determining the background parameterisation results in small biases for some of the background parameters which then can cause biases for the signal parameters. Secondly, the extrapolation of the background parameters along $m_{B^0}^{\mathrm{rec}}$ causes a systematic uncertainty.
Other sources of systematic uncertainties such as the choice of the parameterisation of the background components were considered, for example higher orders of the Chebychev polynomials, used to parameterise the angular distributions of the background. However, the order chosen in this thesis gives the best agreement with the data. Fits with higher orders are susceptible to over-fitting the data, exhibit large fluctuations of the PDF at the borders of the distributions, and are therefore no reasonable variations of the fit strategy.

Biases from the fit procedure

The biases of the background parameters caused by the background fit procedure (see chapter 6) can cause biases of the signal parameters. These are treated as a source of systematic uncertainty. To determine the systematic uncertainties caused by the background fit procedure, a toy study is performed. One thousand toy samples of three million events each are generated with the full signal and background model, including resolution and acceptance effects. The signal and background events are generated separately using the relative yields observed in data, and each event is labelled either signal or background. Each toy sample is then fitted twice:

(A) The signal events are fitted with the signal only model.

(B) All events (signal and background) are fitted with the full three-step fit procedure described in section 6.9.

Then, the difference of each signal parameter obtained in fit (A) and fit (B) is calculated for each toy sample, e.g. $\Delta\mathcal{C}_9 = \mathcal{C}_9^{\mathrm{Fit\,A}} - \mathcal{C}_9^{\mathrm{Fit\,B}}$.
Figure 7.3 shows the distributions of $\Delta\mathcal{C}_9$ and $\Delta\mathcal{C}_{10}$ as well as Gaussian fits to determine the mean μ_Δ and width σ_Δ of the distributions. The absolute value of the mean shift i.e. $|\mu_\Delta|$,

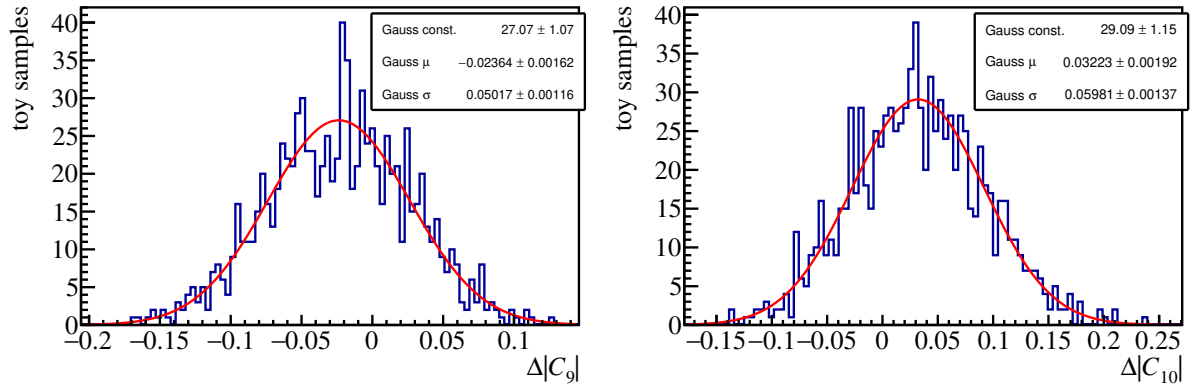


Figure 7.3: $\Delta\mathcal{C}_9$ and $\Delta\mathcal{C}_{10}$ distributions (where $\Delta\mathcal{C}_9 = \mathcal{C}_9^{\text{Fit A}} - \mathcal{C}_9^{\text{Fit B}}$ and analogously for $\Delta\mathcal{C}_{10}$). Each distribution is fitted with a Gaussian to determine the μ_Δ which is taken as the systematic uncertainty related to the background parameterisation.

is taken as the systematic uncertainty caused by the background parameterisation.[2] This is done for all signal parameters. The resulting systematic uncertainties and the systematic uncertainties as a fraction of the respective statistical uncertainty from the fit to Run 1 data are given in table 7.1.[3] Only the parameters with systematic uncertainties of at least 5% of the statistical uncertainty are shown. The systematic uncertainties for $\mathcal{C}_9$ and $\mathcal{C}_{10}$ are 9% and 16% of the statistical uncertainty respectively. The real part[4] of the non-local contribution to $\mathcal{C}_7$ in the A_0 amplitude has a systematic uncertainty of 6% of the statistical uncertainty. Also one of the FF parameters, α_2^{A1}, is affected at the level of 11% of the statistical uncertainties, as are the resolution parameters in the mid-q^2 region (α^2) and in the high q^2 region (α^3) at 14% and 13% of the statistical uncertainty.

[2]The widths of the $\Delta\mathcal{C}_{9,10}$ distributions are correlated to the statistical uncertainty added by the background, since no background events are included in fit (A).

[3]Technically, the "statistical" uncertainties here are the fit uncertainties which include the statistical uncertainties as well as the theoretical uncertainties from the FF constrain.

[4]The fit was found to be more stable when fitting for the real and imaginary parts of the non-local contribution to $\mathcal{C}_7$ instead of the corresponding magnitudes $\zeta^{\perp,\parallel,0}$ and phases $\omega^{\perp,\parallel,0}$.

	μ_Δ	$\mu_\Delta/\sigma_{\text{statistical}}$
$\lvert A_{\parallel}^{\psi(2S)}\rvert$	(-40.94 ± 2.13)$\times 10^{-7}$	0.09
$\lvert A_{\perp}^{\psi(2S)}\rvert$	(-47.71 ± 2.39)$\times 10^{-7}$	0.10
$\lvert A_{0}^{\psi(2S)}\rvert$	(-71.815 ± 3.044)$\times 10^{-7}$	0.12
$\mathrm{Im}(A_{0}^{\psi(4160)})$	(-19.79 ± 1.31)$\times 10^{-7}$	0.05
FF α_2^{A1}	0.0547 ± 0.0035	0.11
$\lvert A_{00}^{\psi(2S)}\rvert$	(-26.49 ± 1.64)$\times 10^{-7}$	0.05
$\lvert C_9\rvert$	-0.0236 ± 0.0016	0.09
$\lvert C_{10}\rvert$	0.0322 ± 0.0019	0.16
$\mathrm{Re}(\Delta C_7^{\mathrm{A0}})$	(5.13 ± 0.77)$\times 10^{-3}$	0.06
resolution α^2	(1.050 ± 0.064)$\times 10^{-3}$	0.14
resolution α^3	(8.96 ± 0.34)$\times 10^{-3}$	0.13

Table 7.1: Systematic uncertainties of the signal parameters due to the background fitting procedure. The systematic uncertainties are determined by generating 1000 high statistics toy samples and comparing the fit results of a signal only fit to the results from the full signal and background fit. The absolute value of the mean difference of the fit parameters ($|\mu_\Delta|$, obtained with Gaussian fits) is taken as systematic uncertainty. In the right column, the systematic uncertainties are given as a fraction of the statistical uncertainty of the respective parameter. Only the parameters which have systematic uncertainties of at least 5% of the statistical uncertainty are shown.

Extrapolation of the background parameters

Since the background PDFs before and after the signal region fit are not fully statistically compatible, as shown in figure 7.1 and figure 7.2, the systematic uncertainties related to the extrapolation of the background parameters is determined. The systematic uncertainties are estimated by using varied fit strategies to fit the data:

(A) The nominal fit as explained in section 6.9.

(B) The background parameters and the signal fractions are fixed to the best fit results of the simultaneous $m_{B^0}^{\text{rec}}$ fit and the simultaneous side-band fit.

(C) The same fit strategy as (A) except that all background parameters are reparameterised with $p^x_{\text{SB}_i} = p^x_{\text{a}} + p^x_{\text{md}} \times m^d_{\text{SB}_i}$ instead of just the $\cos\theta_K$ parameters (see section 6.7).

Both variations constitute tests for the $m_{B^0}^{\text{rec}}$ dependence of the background parameterisation and the systematic uncertainties determined with the two methods are therefore correlated. However, for the Run 1 only data, the number of events are too small to ensure a stable fit with variation (C) which uses 52 free parameters to describe the background (also see table 5.3 for the number of events). Therefore, this method is omitted for the Run 1 only fit, but could be used for the fit to the merged Run 1 and Run 2 data.

For this thesis, only the variation (B) is considered and each signal parameter from fit (B) is compared to the respective value obtained in the nominal fit (A). The absolute difference is taken as systematic uncertainty. The projections of the background PDF of fit (A) and (B) can be compared in figures 7.1 and 7.2 as well as appendix A.5.
The systematic uncertainties resulting from the extrapolation of the background parameters along $m_{B^0}^{\text{rec}}$ are given in table 7.2 along with the systematic uncertainties as a fraction of the statistical uncertainties in the nominal fit.
The systematic uncertainties are quite large, most notably for $\mathcal{C}_9$ and $\mathcal{C}_{10}$ where the systematic uncertainties are 20% and 65% of the statistical uncertainties respectively. The extrapolation of the background parameters along $m_{B^0}^{\text{rec}}$ is therefore the dominant systematic uncertainty for $\mathcal{C}_{10}$.

However, the nominal fit uses background events in the signal region[5], and is therefore less affected by the extrapolation of the background parameters than the study presented

[5] As stated before in the mid q^2 region, there are ~ 5215 background events in the signal region and 3156 background events in the upper mass side-band

in this section suggests. In other words, it could be argued that the fit with the fixed background is not an ideal variation of the fit strategy for determining a systematic uncertainty. The variation is overestimating the 'misunderstanding' of the background, since the varied fit simply ignores some of the information available in the data. Therefore, the systematic uncertainties given in this section are likely overestimated. However, in order to stay conservative, the systematic uncertainties are used as given in table 7.2. This method should be revisited for the fit to the merged Run 1 and Run 2 data.

	$\sigma_{\rm syst}$	$\sigma_{\rm syst}/\sigma_{\rm stat}$
$\lvert A_{\parallel}^{J/\psi}\rvert$	0.000006	0.26
$\theta_{\parallel}^{J/\psi}$	0.003202	0.26
$\lvert A_{\perp}^{J/\psi}\rvert$	0.000002	0.09
$\theta_{0}^{J/\psi}$	0.093489	0.93
$\lvert A_{\parallel}^{\psi(2S)}\rvert$	0.000008	0.18
$\lvert A_{\perp}^{\psi(2S)}\rvert$	0.000007	0.15
$\theta_{0}^{\psi(2S)}$	0.126169	0.40
$\lvert A_{0}^{\rho(770)})\rvert$	0.000004	0.11
$\theta_{0}^{\rho(770)}$	0.000003	0.06
$\mathrm{Im}(A_{0}^{\phi(1020)})$	0.000003	0.20
$\mathrm{Re}(A_{\parallel}^{\psi(3770)})$	0.000010	0.39
$\mathrm{Im}(A_{\parallel}^{\psi(3770)})$	0.000002	0.10
$\mathrm{Re}(A_{\perp}^{\psi(3770)})$	0.000001	0.06
$\mathrm{Im}(A_{\perp}^{\psi(3770)})$	0.000006	0.27
$\mathrm{Re}(A_{0}^{\psi(3770)})$	0.000004	0.13
$\mathrm{Im}(A_{0}^{\psi(3770)})$	0.000005	0.13
$\mathrm{Re}(A_{\parallel}^{\psi(4040)})$	0.000003	0.08
$\mathrm{Re}(A_{\perp}^{\psi(4040)})$	0.000002	0.07
$\mathrm{Im}(A_{\perp}^{\psi(4040)})$	0.000003	0.12
$\mathrm{Re}(A_{0}^{\psi(4040)})$	0.000002	0.05
$\mathrm{Im}(A_{0}^{\psi(4040)})$	0.000004	0.09
$\mathrm{Im}(A_{\parallel}^{\psi(4160)})$	0.000002	0.05
$\mathrm{Re}(A_{\perp}^{\psi(4160)})$	0.000004	0.12
$\mathrm{Re}(A_{0}^{\psi(4160)})$	0.000005	0.13
$\mathrm{Im}(A_{0}^{\psi(4160)})$	0.000006	0.16

	$\sigma_{\rm syst}$	$\sigma_{\rm syst}/\sigma_{\rm stat}$
α_{0}^{A12}	0.000773	0.08
α_{2}^{A12}	0.020461	0.05
$\lvert A_{00}^{J/\psi}\rvert$	0.000034	0.66
$\theta_{00}^{J/\psi}$	0.083515	0.83
$\lvert A_{00}^{\psi(2S)}\rvert$	0.000013	0.25
$\theta_{00}^{\psi(2S)}$	0.141741	0.45
$\lvert C_{9}\rvert$	0.050106	0.20
$\lvert C_{10}\rvert$	0.130912	0.65
Scale Swave V1	0.955726	0.41
Scale Swave V2	1.025880	0.45
Scale Swave T	0.412042	0.35
$\mathrm{Re}(\Delta C_{7}^{\mathrm{AP}})$	0.002409	0.12
$\mathrm{Re}(\Delta C_{7}^{\mathrm{AT}})$	0.015854	0.41
$\mathrm{Im}(\Delta C_{7}^{\mathrm{AT}})$	0.004485	0.19
$\mathrm{Re}(\Delta C_{7}^{\mathrm{A0}})$	0.005057	0.06
$\mathrm{Im}(\Delta C_{7}^{\mathrm{A0}})$	0.002005	0.08
α^{2}	0.001761	0.23
α^{3}	0.038108	0.54

Table 7.2: Systematic uncertainties related to the extrapolation of the background parameters along $m_{B^0}^{\rm rec}$. Also shown are the systematic uncertainties as a fraction of the statistical uncertainty. Only the parameters with systematic uncertainties which are at least 5% of the statistical uncertainty are shown.

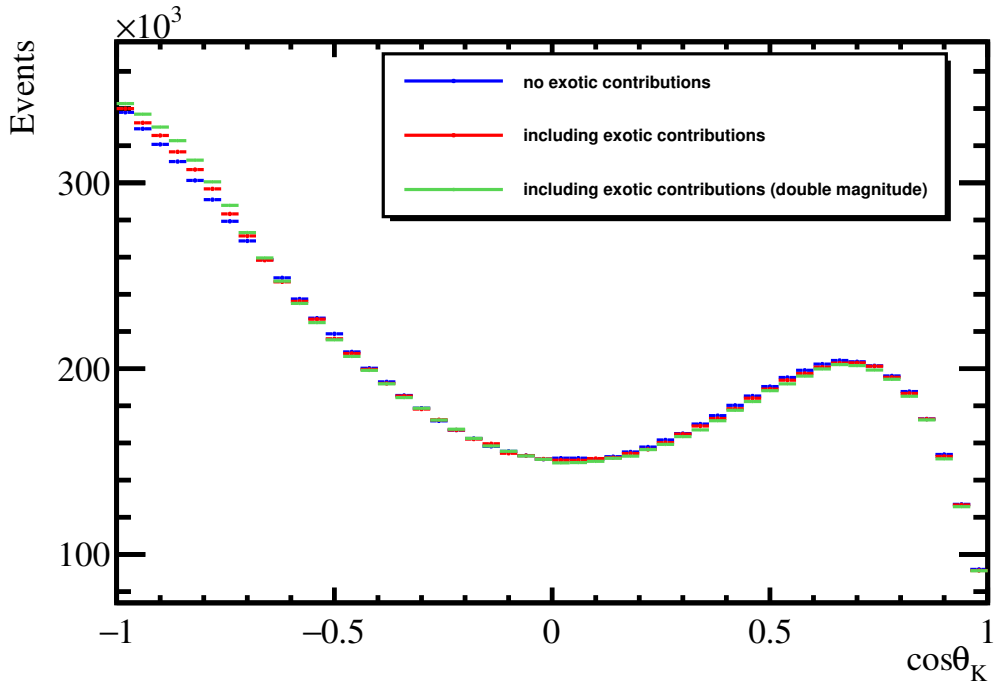


Figure 7.4: $\cos\theta_K$ distribution of toy events generated with the model transformed into the helicity basis including the exotic contributions (red), including the exotic contributions with doubled magnitudes (green), and without exotic contributions (blue). Figure from Ref. [50].

7.2.2 Exotic charmonium-like states

Several exotic charmonium-like states with a quark content of $c\bar{c}ud$ have been observed in decays of B^0 mesons. The first observation was made by the Belle experiment by measuring the $\mathrm{Z}(4430)^-$ state in the $\psi(2S)\pi^+$ invariant mass spectrum of the $\overline{B}^0 \to \psi(2S)K^-\pi^+$ decay [101, 102]. The existence of the $\mathrm{Z}(4430)^-$ was also confirmed by the LHCb collaboration [103]. The Belle collaboration performed a full angular analysis of the $\mathrm{Z}(4430)^- \to \psi(2S)\pi^+$ decay [104]. Two more $c\bar{c}ud$ states have been observed by the Belle collaboration in the $J/\psi\pi^\pm$ spectrum of the $\overline{B}^0 \to J/\psi K^-\pi^+$ decay: $\mathrm{Z}(4330)^+$ and the $\mathrm{Z}(4200)^+$ [105].

The presence of these exotic states in the $J/\psi\pi^\pm$ spectrum of $B^0 \to J/\psi K^-\pi^+$ decays can have an impact on the angular distributions as shown in Ref. [106], particularly adding a peak in the $\cos\theta_K$ distribution at $\cos\theta_K < -0.5$. Since the exotic charmonium-like states are not included in the empirical model used in this thesis, a corresponding systematic uncertainty is assigned. To this end, toy samples are simulated with an extended model which includes the amplitudes corresponding to the $\mathrm{Z}(4430)^\pm$ and $\mathrm{Z}(4200)^+$ states, using

the magnitudes and phases measured by the Belle collaboration [104, 105]. Since the measurements were performed using the helicity amplitude basis ($H^{0,+,-}$), the model described in chapter 3 is transformed into the helicity formalism using the relations given in Ref. [107], in order to allow the inclusion of the exotic contributions into the model. Using the extended model, toy samples are generated while also including acceptance and resolution effects (see sections 5.2 and 5.3). These toy samples are labelled "exotic" toys. Then, using the same model, but setting all exotic amplitudes to zero,[6] the same number of toy samples, labelled "non exotic" toys, are generated. Figure 7.4 shows the $\cos\theta_K$ distribution of one of the exotic toy samples (red points) and of the corresponding non-exotic toy sample (blue points). The effect of the presence of the exotic states is clearly visible at $\cos\theta_K < -0.5$.

Both the exotic and non-exotic toys are fitted with the nominal model described in chapter 3 and the mean difference of the fit parameters is used as a systematic uncertainty. The systematic uncertainties are given in table 7.3 for the parameters most affected by the presence of the exotic states. Also given are the systematic uncertainties divided by the statistical uncertainties from the fit to Run 1 data. For the Wilson coefficients $\mathcal{C}_9$ and $\mathcal{C}_{10,}$ the systematic uncertainty is on the order of 8% of the statistical uncertainty. The systematic uncertainty of the phase of the J/ψ amplitudes relative to the penguin amplitudes is approximately 6% of the statistical uncertainty.

Since the measured magnitudes of the $\mathrm{Z}(4430)^{\pm}$ and $\mathrm{Z}(4200)^{+}$ amplitudes have large uncertainties, a second toy study is performed where the magnitudes of the exotic amplitudes are doubled. Figure 7.4 shows the $\cos\theta_K$ distribution of a toy sample with doubled magnitudes of the exotic amplitudes (green points). The resulting systematic uncertainties, shown in table 7.4, are used for the final result presented in section 7.3.

[6] All other parameters remain unchanged and the same list of starting seeds are used for generating the non-exotic toys as were used for generating the exotic toys.

Parameter	Systematic uncertainty	Systematic uncertainty/statistical uncertainty
$\theta^0_{J/\psi}$	5.92×10^{-3}	0.06
$\|C_9\|$	2.07×10^{-2}	0.08
$\|C_{10}\|$	1.61×10^{-2}	0.09

Table 7.3: The systematic uncertainties due to ignoring the presence of the charmonium-like exotic states $Z(4430)^{\pm}$ and $Z(4200)^{+}$ in the signal model. The systematic uncertainties as a fraction of the expected statistical uncertainties from fit to Run 1 data are given in the right coloumn. The uncertainties are given for the C_9, C_{10} and the phase $\theta^0_{J/\psi}$. The effect of the exotic states is found to be negligible for the other parameters.

Parameter	Systematic uncertainty	Systematic uncertainty/statistical uncertainty
$\theta^0_{J/\psi}$	1.27×10^{-2}	0.13
$\|C_9\|$	4.88×10^{-2}	0.19
$\|C_{10}\|$	5.21×10^{-2}	0.26

Table 7.4: The systematic uncertainties due to ignoring the presence of the charmonium-like exotic states $Z(4430)^{\pm}$ and $Z(4200)^{+}$ in the signal model when doubling the magnitudes of the amplitudes of the exotic states. The systematic uncertainties as a fraction of the statistical uncertainties from the fit to Run 1 data are given in the right column. The uncertainties are given for the C_9, C_{10} and the phase $\theta^0_{J/\psi}$. The effect of the exotic states is found to be negligible for the other parameters.

7.2.3 S-Wave form factors

As discussed in sections 2.9 and 3.3, the signal model used in this thesis includes contributions from $B^0 \to K^{*0}\mu^+\mu^-$ decays in the S-Wave configuration. The S-Wave amplitudes are expressed in terms of the two from factors $F_1(q^2)$ and $F_T(q^2)$ as shown in equation 2.40. The S-Wave form factors have not been studied very extensively and therefore have large uncertainties. The nominal model in this thesis uses the values of the S-Wave form factors coefficients $F(0)^{1,T}$, $\alpha_F^{1,T}$, and $b_F^{1,T}$ calculated in Ref. [43]. The authors did not publish any uncertainties for these coefficients.

Therefore, a conservative estimate of the systematic uncertainties, caused by the lack of understanding of the S-Wave form factors, is determined for the signal parameters. To this end, toy simulations with 10 millions events each are produced using the nominal signal model, including resolution and acceptance effects. Each toy sample is then fitted twice, once with the nominal signal model, and once with a modified model where the S-Wave from factors $F_{1,T}$ are replaced with the form factors that enter into the longitudinal P-wave amplitude $\mathcal{A}_0^{\mathrm{L,R}}$ (see equation 2.33). Explicitly, F_1 is replaced with A_{12} and F_T is replaced with T_{23}. The motivation behind using the $\mathcal{A}_0^{\mathrm{L,R}}$ FFs instead of the S-Wave FFs as a systematic variation, is that the K^{*0} in the S-Wave configuration is longitudinally polarised just like the $\mathcal{A}_0^{\mathrm{L,R}}$ amplitude.

The mean difference of each fit parameter in the nominal fit and the fit with swapped FFs is taken as a systematic uncertainty. The resulting systematic uncertainties as well as the systematic uncertainties as a fraction of the statistical uncertainties from the fit to Run 1 data are given in table 7.5. Only the parameters for which the systematic uncertainty is at least 5% of the statistical uncertainty are given. The systematic uncertainties, are significant for several parameters, for example for $\mathcal{C}_9$ where the systematic uncertainty is $\sim 46\%$ of the expected statistical uncertainty.

However, as shown in figure 7.5, the S-Wave FFs (blue line) and P-Wave FFs (dashed black line and red area) are vastly different, indicating that the systematic uncertainties obtained from swapping the S-Wave FFs for the P-Wave FFs are likely an overestimation.

Furthermore, figure 7.6 shows the $B^+ \to K^+$ form factors f_0, f_+, and f_T, calculated using lattice QCD [108]. The notation is such that f_+ corresponds to F_1 and f_T corresponds to F_T. f_0 denotes the scalar form factor which is not included in the model used in this thesis. The similarity between the $B^+ \to K^+$ form factors and the S-Wave $B^0 \to K^{*0}$ form factors, and the fact that the $B^+ \to K^+\mu^+\mu^-$ decay is also in an S-Wave configuration, suggest that the $B^+ \to K^+$ form factors could be used as a systematic variation in order

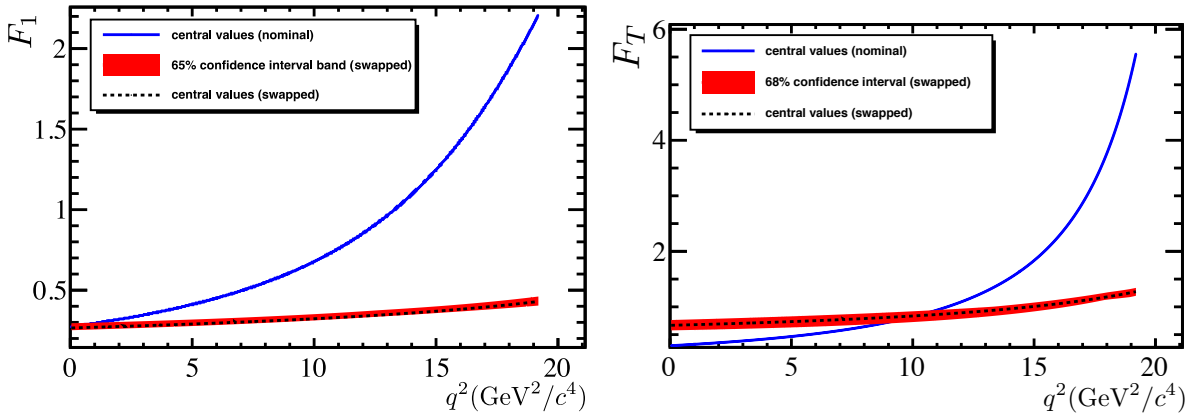


Figure 7.5: The S-Wave form factors F_1 and F_T as a function of q^2 using the nominal coefficients from Ref. [43] (blue line). Also shown are the form factors of the P-Wave amplitude $\mathcal{A}_0$: A_{12} (left) and T_{23} (right) – shown with the dashed black line and the red area indicating the 68% confidence interval – with which F_1 and F_T are replaced in order to estimate a systematic uncertainty associated to the S-Wave form factors. Figures from Ref. [50].

to estimate the systematic uncertainties.

However, in order to avoid underestimating the systematic uncertainties, the more conservative study outlined above is used in this thesis. Further theoretical calculations of the S-Wave $B^0 \to K^{*0}$ FFs could improve the precision of the fit with the model described in this thesis.

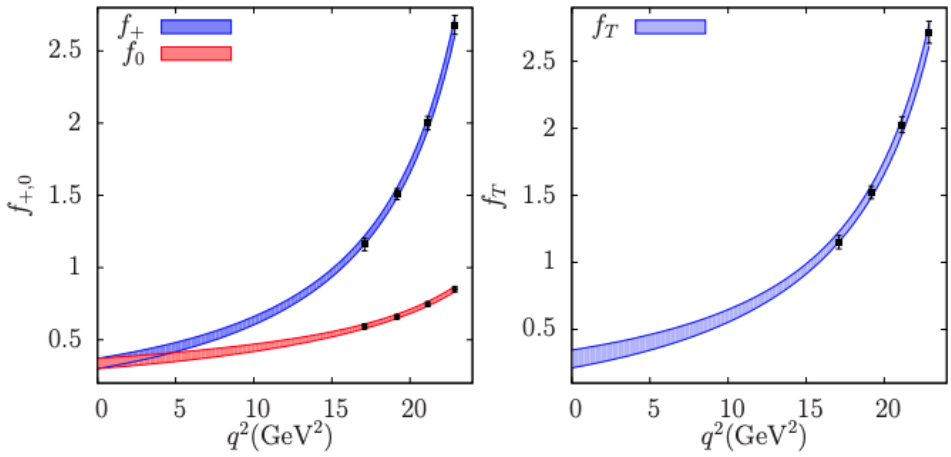


Figure 7.6: The $B^+ \to K^+$ form factors f_0, f_+, f_T as a function of q^2. Relating these to the $B^0 \to K^{*0}$ form factors, f_+ corresponds to F_1 and f_T corresponds to F_T. f_0 is the scalar form factor which is not included in the model used in this thesis. Figures from Ref. [108].

Parameter	Systematic uncertainty	Systematic uncertainty/Statistical uncertainty
$\|C_9\|$	1.15×10^{-1}	0.46
$\|C_{10}\|$	2.12×10^{-2}	0.11
$\theta^0_{J/\psi}$	2.86×10^{-4}	0.29
$\eta^{\parallel}_{\psi(2S)}$	2.02×10^{-5}	0.45
$\eta^{\perp}_{\psi(2S)}$	2.41×10^{-5}	0.52
$\eta^{0}_{\psi(2S)}$	3.01×10^{-5}	0.50
$\eta^{00}_{\psi(2S)}$	9.97×10^{-6}	0.20
$\mathrm{Re}(\mathrm{A}^0_{\rho^0})$	6.35×10^{-6}	0.21
Im $(\mathrm{A}^0_{\rho^0})$	2.90×10^{-6}	0.07
$\mathrm{Re}(\mathrm{A}^{\parallel}_{\psi(3770)})$	1.61×10^{-6}	0.06
$\mathrm{Im}(\mathrm{A}^{\perp}_{\psi(3770)})$	4.87×10^{-6}	0.22
$\mathrm{Re}(\mathrm{A}^{\parallel}_{\psi(4040)})$	2.82×10^{-6}	0.09
$\mathrm{Im}(\mathrm{A}^{\parallel}_{\psi(4040)})$	6.07×10^{-6}	0.15
$\mathrm{Re}(\mathrm{A}^{\perp}_{\psi(4040)})$	9.89×10^{-6}	0.39
$\mathrm{Im}(\mathrm{A}^{\perp}_{\psi(4040)})$	5.56×10^{-6}	0.22
$\mathrm{Re}(\mathrm{A}^{0}_{\psi(4040)})$	1.71×10^{-6}	0.05
$\mathrm{Im}(\mathrm{A}^{0}_{\psi(4040)})$	4.05×10^{-6}	0.10
$\mathrm{Re}(\mathrm{A}^{\parallel}_{\psi(4160)})$	3.55×10^{-6}	0.09
$\mathrm{Im}(\mathrm{A}^{\parallel}_{\psi(4160)})$	6.69×10^{-6}	0.155
$\mathrm{Re}(\mathrm{A}^{\perp}_{\psi(4160)})$	1.04×10^{-5}	0.325
$\mathrm{Im}(\mathrm{A}^{\perp}_{\psi(4160)})$	7.38×10^{-6}	0.27
$\mathrm{Im}(\mathrm{A}^{0}_{\psi(4160)})$	5.55×10^{-6}	0.14
$\mathrm{Re}(\zeta^{\parallel} e^{i\omega\parallel})$	7.76×10^{-3}	0.20
$\mathrm{Re}(\zeta^{\perp} e^{i\omega\perp})$	1.27×10^{-2}	0.33
$\mathrm{Im}(\zeta^{\perp} e^{i\omega\perp})$	6.16×10^{-3}	0.27
$\mathrm{Re}(\zeta^{0} e^{i\omega 0})$	1.03×10^{-1}	1.18

Table 7.5: The systematic uncertainties associated to the poor understanding of the S-Wave form factors. In the right column, the systematic uncertainties as a fraction of the statistical uncertainties from the fit to the Run 1 data are given. Only parameters with a systematic uncertainty which is at least 5% of the statistical uncertainty are given.

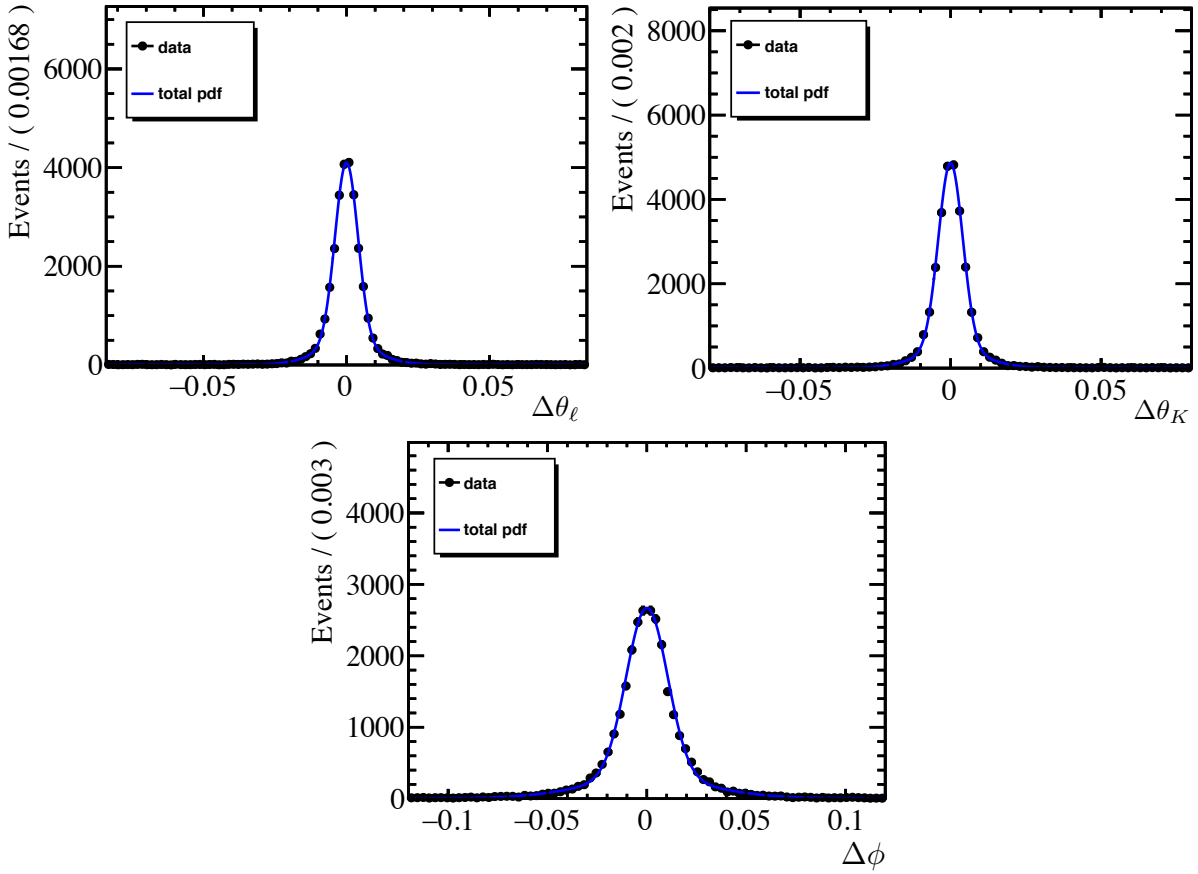


Figure 7.7: Distributions of $\Delta\theta_\ell$, $\Delta\theta_K$ and $\Delta\phi$ defined as the differences between the angles of the true and reconstructed decay products in $B^0 \to K^{*0}\mu^+\mu^-$ simulations. Also shown are fits with triple Gaussian PDFs used to parameterise the angular resolution. Figure from Ref. [50].

7.2.4 Angular Resolution

As explained in section 5.3, the signal model used in this thesis is convolved with a resolution model in q^2, but ignores resolution effects in the angles. This is due to the fact that the q^2 distribution includes extremely narrow peaks which are much narrower than the q^2 resolution but the angular distributions do not contain any such peaks and are varying slower than the angular resolution. Nevertheless, ignoring the limited angular resolution is a source of systematic uncertainty. The estimation of this systematic uncertainty is presented in this section.

Firstly, simulated $B^0 \to K^{*0}\mu^+\mu^-$ events from the full MC simulation (see section 4.8) in the ranges ($5239.58 \leq m^{\rm rec}_{B^0} \leq 5319.58$) MeV/$c^2$ and ($795.9 \leq m_{K\pi} \leq 995.9$) MeV/$c^2$ are used to a obtain a parameterisation of the angular resolution. Using the true (i.e.

generator level) and the reconstructed four momenta of the final state particles, the true and reconstructed angles are calculated. The distributions of the differences $\Delta x = x^{rec} - x^{tru}$ for $x \in \{\cos\theta_K, \cos\theta_\ell, \phi\}$ are then parameterised using triple Gaussian PDFs. The triple Gaussian is defined as a sum of three Gaussians with independent means $\mu_{1,2,3}$, independent widths $\sigma_{1,2,3}$ and two fractions $f_{1,2}$. The distributions of $\Delta\cos\theta_\ell$, $\Delta\cos\theta_K$, and $\Delta\phi$ along with the respective triple Gaussian fit are shown in figure 7.7. The triple Gaussian PDFs describe the distributions well. The resulting resolution parameters for the resolution in $\cos\theta_\ell$, $\cos\theta_K$, and ϕ are given in tables 7.6, 7.7, and 7.8 respectively.

The angular resolution parameterisations are then used in a toy study designed to determine a systematic uncertainty. Using the nominal signal model, including q^2-resolution and acceptance effects, toy simulation samples are generated with 10 million events each. Each sample is then fitted twice. First, the events are fitted with the nominal model. Then, the angles in the toy samples are smeared using the resolution parameterisations shown in figure 7.7 and then fitted with the nominal model again. The mean difference of the fit parameters in the two fits are taken as the systematic uncertainty related to ignoring the angular resolution.

The resulting systematic uncertainties are given in table 7.9. The largest systematic uncertainty is observed for the real part of the non-local contribution to the $\mathcal{C}_7$ in the longitudinal amplitude ($\mathrm{Re}(\zeta^0 e^{i\omega 0})$), with a systematic uncertainty of 14% of the statistical uncertainty. The systematic uncertainties of the Wilson Coefficients are small, with $\mathcal{C}_{10}$ having the larger systematic uncertainty of 8% of the statistical uncertainty. With more data, these systematic uncertainties will become more significant and therefore it may be advisable to include the angular resolution model in the fit in future measurements which include Run 3 data.

Parameter	Value
σ_1	$3.99 \times 10^{-3} \pm 4.69 \times 10^{-5}$
σ_2	$1.02 \times 10^{-2} \pm 2.57 \times 10^{-4}$
σ_3	$1.28 \times 10^{+0} \pm 8.21 \times 10^{-3}$
μ_1	$-1.56 \times 10^{-2} \pm 3.43 \times 10^{-5}$
μ_2	$3.08 \times 10^{-5} \pm 1.64 \times 10^{-4}$
μ_3	$-2.34 \times 10^{-2} \pm 1.12 \times 10^{-2}$
f_1	$7.84 \times 10^{-1} \pm 1.18 \times 10^{-2}$
f_2	$6.81 \times 10^{-1} \pm 2.34 \times 10^{-3}$

Table 7.6: Result of the fit parameters from fitting a triple Gaussian to the distribution of $\Delta\cos\theta_\ell$ in order to parameterise the $\cos\theta_\ell$ resolution.

Parameter	Value
σ_1	$4.05 \times 10^{-3} \pm 4.51 \times 10^{-5}$
σ_2	$1.07 \times 10^{-2} \pm 2.75 \times 10^{-4}$
σ_3	$3.69 \times 10^{+2} \pm 1.84 \times 10^{2}$
μ_1	$-1.62 \times 10^{-2} \pm 3.36 \times 10^{-5}$
μ_2	$4.01 \times 10^{-5} \pm 1.77 \times 10^{-4}$
μ_3	$2.35 \times 10^{-1} \pm 1.57 \times 10^{-1}$
f_1	$7.98 \times 10^{-1} \pm 1.07 \times 10^{-2}$
f_2	$6.83 \times 10^{-1} \pm 2.33 \times 10^{-3}$

Table 7.7: Result of the fit parameters from fitting a triple Gaussian to the distribution of $\Delta\cos\theta_K$ in order to parameterise the $\cos\theta_K$ resolution.

Parameter	Value
σ_1	$1.04 \times 10^{-2} \pm 1.29 \times 10^{-4}$
σ_2	$3.08 \times 10^{-2} \pm 6.85 \times 10^{-4}$
σ_3	$5.00 \times 10^{+2} \pm 3.77 \times 10^{2}$
μ_1	$6.99 \times 10^{-5} \pm 9.26 \times 10^{-5}$
μ_2	$-1.12 \times 10^{-4} \pm 4.43 \times 10^{-4}$
μ_3	$6.58 \times 10^{-2} \pm 1.29 \times 10^{-1}$
f_1	$7.43 \times 10^{-1} \pm 1.02 \times 10^{-2}$
f_2	$6.75 \times 10^{-1} \pm 2.38 \times 10^{-3}$

Table 7.8: Result of the fit parameters from fitting a triple Gaussian to the distribution of $\Delta\cos\theta_K$ in $B^0 \to K^{*0}\mu^+\mu^-$ simulations in order to parameterise the ϕ resolution.

Parameter	Systematic uncertainty	Systematic uncertainty/Statistical uncertainty
$\|C_9\|$	8.84×10^{-3}	0.04
$\|C_{10}\|$	1.59×10^{-2}	0.08
$\eta^{\parallel}_{J/\psi}$	1.45×10^{-6}	0.06
$\eta^{\perp}_{J/\psi}$	1.36×10^{-6}	0.06
$\eta^{\perp}_{\psi(2S)}$	3.69×10^{-6}	0.08
$\eta^{0}_{\psi(2S)}$	3.31×10^{-6}	0.05
$\mathrm{Re}(\zeta^{\perp}e^{i\omega\perp})$	2.74×10^{-3}	0.07
$\mathrm{Re}(\zeta^{0}e^{i\omega 0})$	1.24×10^{-2}	0.14

Table 7.9: The systematic uncertainties related to ignoring the angluar resolution in the model, and the systematic uncertainties divided by the respective statistical uncertainties from the fit to Run 1 data. Only parameters that have a systematic uncertainty which is at least 5% of the expected statistical uncertainty are given.

7.2.5 Acceptance

The acceptance parameterisation described in section 5.2 is dependent on the choice of the orders of the Legendre polynomials. In order to avoid over-fitting and oscillations at the edges of the distributions, the lowest set of orders is chosen with which the acceptance parameterisation can still describe the acceptance effect well. In order to ascertain the systematic uncertainty associated to the orders of the Legendre polynomials, an alternative higher order acceptance parameterisation is determined, where each order has been increased by 3. Toy event samples are generated from the signal model using the higher order acceptance model and fitted with the nominal model and with the higher order model. No significant difference between the two fits is observed for any of the signal parameters, indicating that the systematic uncertainty associated to the choice of orders of the acceptance function is negligible.

Furthermore, the limited number of simulated events used for determining the acceptance coefficients can be a cause of systematic uncertainty. This has been investigated extensively for the measurement of the q^2-binned angular observables, which uses the same acceptance parameterisation [1, 91]. It was found that the systematic uncertainty associated to the limited number of simulated events used for determining the acceptance parameterisation is negligible. To crosscheck these findings with the model used in this thesis, variations of the acceptance coefficients are created using their covariance matrix. Then, toy simulations are generated using the nominal acceptance model and fitted with the nominal model and with the varied acceptance. The differences of the signal parameters caused by the varied acceptance is found to be negligible compared to the statistical uncertainties of the signal parameters.

A further source of systematic uncertainties related to the acceptance model are due to the fact that the full MC simulations, used to determine the acceptance parameterisation, may not perfectly reproduce the data. The data driven corrections for the differences between data and simulation are described in section 4.8.1. The PID resampling significantly improves the agreement of the PID variables between data and simulation. In order to determine a corresponding systematic uncertainty, the small residual differences between data and simulation are corrected by applying weights. The weights are calculated based on *sWeighted* $B^0 \to J/\psi K^{*0}$ data and the PID-resampled simulation. An alternative acceptance parameterisation is determined based on the additionally weighted simulation. Then, toy simulations are generated with the signal model using the nominal acceptance model and fitted back with the nominal acceptance and the alternative acceptance. The

differences of the signal parameters caused by the alternative acceptance model are negligible compared to the statistical uncertainties.
The systematic uncertainty related to the kinematic reweighting is asssed by determining three alternative acceptance parameterisations, where one of the three weights (used for correcting the `nTracks`, χ^2_{Vtx}, and $p_{\mathrm{T}}^{B^0}$ distributions respectively) is omitted in each one. Then, toy simulation samples are generated with the signal model and each of the three alternative acceptance functions. Each toy sample is fitted with the nominal acceptance and the respective alternative acceptance model. The resulting systematic uncertainties are negligible compared to the statistical uncertainties.

7.2.6 Residual Peaking Backgrounds

Not all of the vetos used to reject peaking backgrounds described in section 5.1.2 are 100% efficient, i.e. a small number of peaking background events are still contained in the final event selection. In order to determine the systematic uncertainties related to these residual peaking backgrounds, the ($\cos\theta_\ell$, $\cos\theta_K$, ϕ, q^2, $m^{\mathrm{rec}}_{B^0}$) distribution of each peaking background component is modelled using simulations. Subsequently, toy event samples are generated for each peaking background contribution and injected into signal toy event samples, using the expected rate of each peaking background contribution. The resulting peaking-background-enriched toy samples are then fitted using the nominal signal model. No significant bias for any of the signal parameters was found, which is due to the extremely small rate of the peaking backgrounds. Therefore, the systematic uncertainty associated to ignoring the residual peaking backgrounds is considered to be negligible.

7.2.7 Summary of the systematic uncertainties

A summary of the systematic uncertainties of the most important signal parameters, i.e. the magnitudes of the Wilson Coefficients $\mathcal{C}_9$ and $\mathcal{C}_{10,}$ and the magnitudes and phases of the J/ψ and $\psi(2S)$ P-Wave amplitudes, is given in table 7.10. The largest systematic uncertainty of each parameter is printed in bold. The background extrapolation is the dominant systematic uncertainty for most of the parameters shown in the table, except for the magnitude of $\mathcal{C}_9$ and the magnitudes of the $\psi(2S)$ amplitudes for which the systematic uncertainties related to the S-Wave FFs are dominant. As discussed in sections 7.2.1 and 7.2.3, both the systematic uncertainties from the background extrapolation and the systematic uncertainties related to the S-Wave FFs are conservative estimates, and should be revisited for the planned fit of the merged Run 1 and Run 2 data.

	bkg fit	bkg extrapolation	exotics	S-wave FF	ang. resolution
$\lvert C_9 \rvert$	2.36×10^{-2}	5.01×10^{-2}	4.88×10^{-2}	$\mathbf{1.15 \times 10^{-1}}$	8.84×10^{-3}
$\lvert C_{10} \rvert$	3.22×10^{-2}	$\mathbf{1.31 \times 10^{-1}}$	5.21×10^{-2}	2.12×10^{-2}	1.59×10^{-2}
$\theta_{\parallel}^{J/\psi}$	1.15×10^{-5}	$\mathbf{3.2 \times 10^{-3}}$		2.86×10^{-4}	
$\theta_{\perp}^{J/\psi}$	$\mathbf{5.22 \times 10^{-5}}$	4.2×10^{-5}			
$\theta_{0}^{J/\psi}$	5.92×10^{-3}	$\mathbf{9.35 \times 10^{-2}}$	1.27×10^{-2}		8.92×10^{-4}
$\theta_{\parallel}^{\psi(2S)}$	3.95×10^{-4}	$\mathbf{1.08 \times 10^{-3}}$			
$\theta_{\perp}^{\psi(2S)}$	7.17×10^{-4}	$\mathbf{1.43 \times 10^{-3}}$			
$\theta_{0}^{\psi(2S)}$	6.9×10^{-3}	$\mathbf{1.26 \times 10^{-1}}$	5.84×10^{-3}		
$\lvert A_{\parallel}^{J/\psi} \rvert$	1.94×10^{-7}	$\mathbf{6 \times 10^{-6}}$			1.45×10^{-6}
$\lvert A_{\perp}^{J/\psi} \rvert$	4.12×10^{-8}	$\mathbf{2 \times 10^{-6}}$			1.36×10^{-6}
$\lvert A_{\parallel}^{\psi(2S)} \rvert$	4.09×10^{-6}	8×10^{-6}		$\mathbf{2.02 \times 10^{-5}}$	1.39×10^{-6}
$\lvert A_{\perp}^{\psi(2S)} \rvert$	4.77×10^{-6}	7×10^{-6}		$\mathbf{2.41 \times 10^{-5}}$	3.69×10^{-6}
$\lvert A_{0}^{\psi(2S)} \rvert$	7.19×10^{-6}	2×10^{-6}		$\mathbf{3.01 \times 10^{-5}}$	3.31×10^{-6}

Table 7.10: Summary of the systematic uncertainties of the magnitudes of the Wilson Coefficients $\mathcal{C}_9$ and $\mathcal{C}_{10,}$ and the magnitudes and phases of the J/ψ and $\psi(2S)$ P-Wave amplitudes. Only systematic uncertainties which are at least 5% of the respective statistical uncertainty from the fit to Run 1 data are included in the table. The largest systematic uncertainty of each parameter is printed in bold.

7.3 Results of the fit to Run 1 data

The fit procedure used for the fit to the Run 1 data is described in section 6.9.
The best fit values for the signal fractions are:

$$\begin{aligned} f_{\text{sig}}^{\text{lowq}^2} &= 0.8898 \pm 0.0006 \\ f_{\text{sig}}^{\text{midq}^2} &= 0.98292 \pm 0.00005 \\ f_{\text{sig}}^{\text{highq}^2} &= 0.9722 \pm 0.0002, \end{aligned} \tag{7.1}$$

where the uncertainties are estimated statistical uncertainties based on toy simulations, since the signal fractions are not direct fit parameters in the fit (see section 6.8).

The $\cos\theta_K$ and $\cos\theta_\ell$ distributions in the three q^2 regions (low-, mid-, and high-q^2) in the signal region ($(5239.58 \leq m_{B^0}^{\text{rec}} \leq 5319.58)$ MeV/c^2) are shown in figure 7.8. Also shown are the projections of the total fit PDF (blue line) and the projections of the signal component (dashed red line) and the background component (dotted and dashed black line). The pull distributions are shown on the bottom of each plot, where the pull is defined as the difference between each data point and the value of the projection of the total PDF at the centre of the respective bin, divided by the uncertainty of the data. The total PDF agrees well with the data, except for $\cos\theta_K < -0.5$. This discrepancy is due to the presence of exotic charmonium-like states, such as the Z(4430), in the data, which are ignored in the model. The corresponding systematic uncertainty is discussed in section 7.2.2.

Figure 7.9 shows the ϕ and q^2 distributions in the three q^2 regions in the signal region along with the PDF projections. The total PDF describes the data well including the q^2 regions of the higher ψ resonances visible as broad peaks above the $\psi(2S)$ peak. Also the ρ^0 and $\phi(1020)$ resonances are visible in the low-q^2 region and are well described by the PDF. The large pulls at the J/ψ and $\psi(2S)$ peaks are due to the coarse binning of the data in the plot and the rapidly changing shape of the distribution. The model does describe these peaks well, as demonstrated with the fit to the narrow J/ψ region shown with much finer binning in figure 5.6.

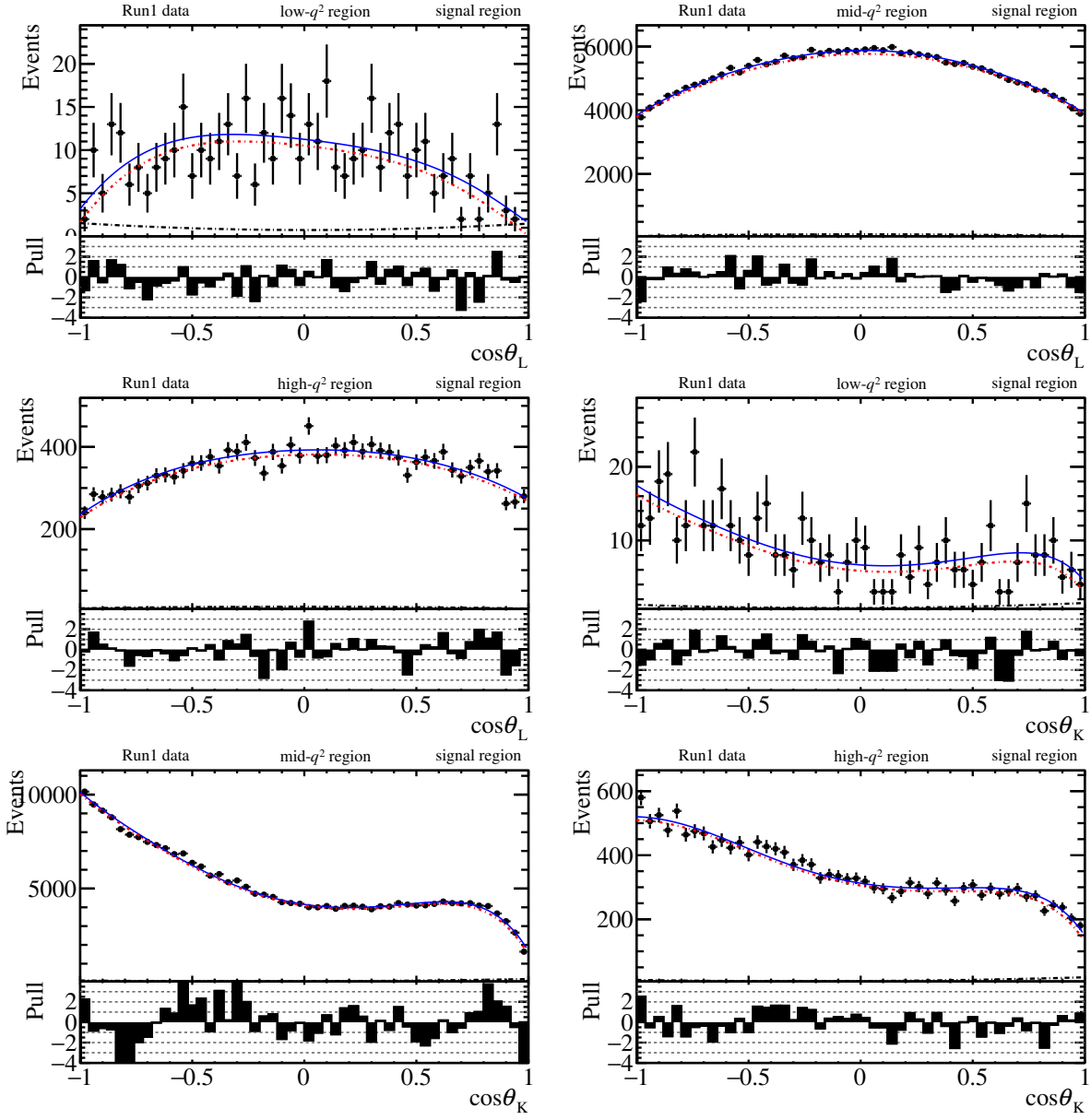


Figure 7.8: $\cos\theta_K$ and $\cos\theta_\ell$ and distributions in the three q^2 regions in the signal region ($(5239.58 \leq m^{\text{rec}}_{B^0} \leq 5319.58)$ MeV/c^2) in Run 1 data. Also shown are the projections of the total fit PDF (blue line) and the projections of the signal component (dashed red line) and the background component (dotted and dashed black line). The pull is defined as the difference between the data and the value of the projection of the total PDF at the centre of the respective bin divided by the uncertainty of the data.

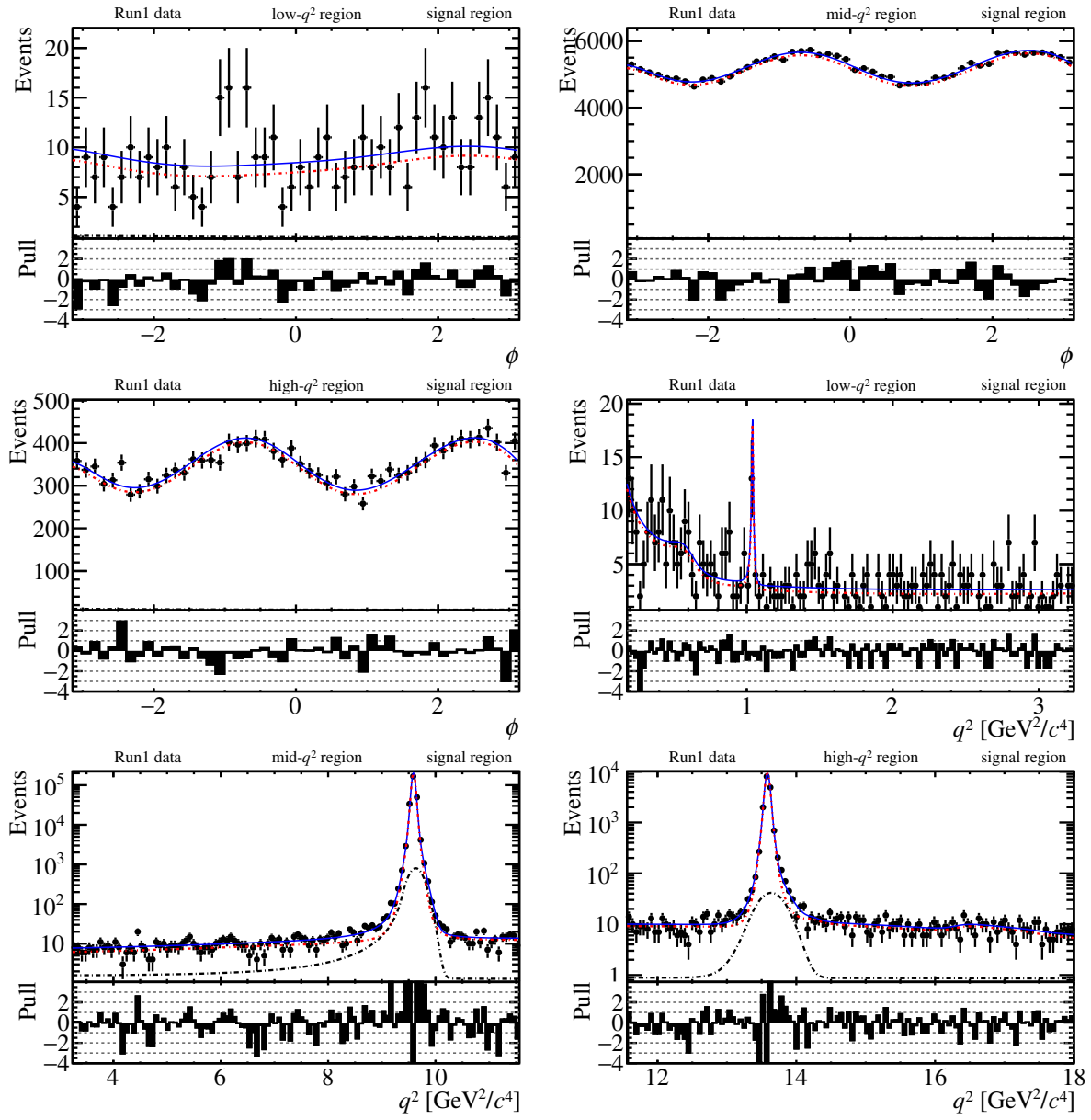


Figure 7.9: ϕ and q^2 and distributions in the three q^2 regions in the signal region ($(5239.58 \leq m^{\text{rec}}_{B^0} \leq 5319.58)$ MeV/c^2) in Run 1 data. Also shown are the projections of the total fit PDF (blue line) and the projections of the signal component (dashed red line) and the background component (dotted and dashed black line). The pull is defined as the difference between the data and the value of the projection of the total PDF at the centre of the respective bin divided by the uncertainty of the data. The large pulls observed at the J/ψ and $\psi(2S)$ peak are due to the coarse binning of the data and the steep slopes of the peaks. The PDF does describe the data well as demonstrated in the fit to the narrow J/ψ region shown in figure 5.6.

	results $\pm\sigma_{\text{fit}} \pm \sigma_{\text{syst}}$	$\sigma_{\text{syst}}/\sigma_{\text{fit}}$	precision
$\lvert C_9 \rvert$	3.64 ± 0.25 ± 0.14	0.55	0.078
$\lvert C_{10} \rvert$	4.42 ± 0.20 ± 0.15	0.73	0.057
$\theta_{\parallel}^{J/\psi}$	0.207 ± 0.012 ± 0.0032	0.26	0.062
$\theta_{\perp}^{J/\psi}$	-0.2411 ± 0.0095 ± 0.000067	0.01	0.039
$\theta_{0}^{J/\psi}$	-1.63 ± 0.10 ± 0.095	0.94	0.085
$\theta_{\parallel}^{\psi(2S)}$	-0.674 ± 0.048 ± 0.0011	0.02	0.071
$\theta_{\perp}^{\psi(2S)}$	-2.6431 ± 0.045 ± 0.0016	0.04	0.017
$\theta_{0}^{\psi(2S)}$	-1.72 ± 0.32 ± 0.13	0.40	0.198
$\lvert A_{\parallel}^{J/\psi} \rvert$	0.004344 ± 0.000023 ± 0.0000062	0.27	0.005
$\lvert A_{\perp}^{J/\psi} \rvert$	0.004178 ± 0.000024 ± 0.0000024	0.10	0.006
$\lvert A_{\parallel}^{\psi(2S)} \rvert$	0.001073 ± 0.000045 ± 0.000022	0.50	0.046
$\lvert A_{\perp}^{\psi(2S)} \rvert$	0.001065 ± 0.000046 ± 0.000026	0.56	0.050
$\lvert A_{0}^{\psi(2S)} \rvert$	0.001596 ± 0.000060 ± 0.000031	0.52	0.042

Table 7.11: Fit results of the magnitudes of the Wilson Coefficients $\mathcal{C}_9$ and $\mathcal{C}_{10}$ and the magnitudes and phases of the J/ψ and $\psi(2S)$ P-Wave amplitudes from the fit to Run 1 data. The uncertainties are the fit uncertainties (including statistical uncertainties and theoretical uncertainties from the P-Wave FF constraint) and systematic uncertainties. The ratio of the systematic uncertainties and the fit uncertainties are given in the centre column. The precision, defined as the total uncertainty (fit uncertainties and systematic uncertainties added in quadrature) divided by the best fit value, is given in the right column. The results for the full set of floating signal parameters are given in table A.1.

The results from the fit to Run 1 data are given in table 7.11 including the fit uncertainties – consisting of the statistical uncertainties and the theoretical uncertainties of the P-Wave FF constraint (see section 2.5) – and the combined systematic uncertainties. The results are given for the magnitudes of the Wilson coefficients $\mathcal{C}_9$ and $\mathcal{C}_{10,}$ and the magnitudes and phases of the J/ψ and $\psi(2S)$ P-Wave amplitudes. The results for the full set of signal parameters are given in appendix A.2.

The precision on the phase of the J/ψ ($\psi(2S)$) amplitudes relative to the penguin amplitudes i.e. $\theta_0^{J/\psi}$ ($\theta_0^{\psi(2S)}$) is on the level of 8.5% (19.8%) which is precise enough to determine the level of the non-local contributions as shown in section 7.3.2.

The result for $\mathcal{C}_9$ can be compared to the value of $\mathcal{C}_9$ obtained from the measurement of the q^2-binned angular observables using the same Run 1 dataset [10]. In the paper the authors present a χ^2 fit to the angular observables using the EOS software package [109], to

determine a best fit value for the real part of $\mathcal{C}_9$. They report the best fit value for $\mathrm{Re}(\mathcal{C}_9)$ to be shifted by $\Delta\mathrm{Re}(\mathcal{C}_9) = -1.04 \pm 0.25$ from the SM central value of $\mathrm{Re}(\mathcal{C}_9) = 4.27$ [110], corresponding to a 3.4 standard deviation discrepancy to the SM based on the difference in χ^2 of the best fit point to the SM point. However, a more recent fit of the q^2-binned angular observables from Run 1 using the FLAVIO package [111] with up to date SM nuisance parameters gives a tension of 3.0σ [1]. Therefore, the uncertainty of $\Delta\mathrm{Re}(\mathcal{C}_9)$ reported in Ref. [10] is likely underestimated.

The best fit value obtained in the fit presented in this thesis, corresponds to $\Delta\mathrm{Re}(\mathcal{C}_9) = -0.63 \pm 0.29$, where the uncertainty is the combined fit uncertainty and systematic uncertainty. This value is closer to the SM model than the one obtained from the binned angular observables, but still only agrees with the SM at 2.2 standard deviations. Based on the SM value of $\mathrm{Re}(\mathcal{C}_{10}) = -4.17$ [110], the best fit value for $\mathcal{C}_{10}$ obtained in the fit presented in this thesis corresponds to $\Delta\mathrm{Re}(\mathcal{C}_{10}) = -0.25 \pm 0.25$, which is compatible with the SM at one standard deviation.

Figure 7.10 shows the best fit point from this thesis (blue cross) in the $\mathcal{C}_9^{\mathrm{NP}}$ - $\mathcal{C}_{10}^{\mathrm{NP}}$ plane as well as the 1σ, 2σ and 3σ contours (blue areas) determined using the covariance matrix from the fit and the systematic uncertainties. The SM values ($\mathcal{C}_9^{\mathrm{NP}} = \mathcal{C}_{10}^{\mathrm{NP}} = 0$) are indicated with dashed lines. The blue contours visualise the discrepancy of the measurement from this thesis with the SM at 2.2σ for $\mathcal{C}_9$ and 1σ for $\mathcal{C}_{10}$. Also, the hypothesis of $\mathcal{C}_9^{\mathrm{NP}} = -\mathcal{C}_{10}^{\mathrm{NP}}$, which is commonly considered in global fits to $b \to s\ell^+\ell^-$ measurements [16, 112], is likely rejected by the fit presented in this thesis, since the results are only compatible with this hypothesis at $\sim 2.5\sigma$ to 3σ.

In order to compare the result from this thesis to previous $B^0 \to K^{*0}\mu^+\mu^-$ measurements made with LHCb Run 1 data, the FLAVIO package is used to determine the allowed region in the $\mathcal{C}_9^{\mathrm{NP}} - \mathcal{C}_{10}^{\mathrm{NP}}$ plane based on the q^2-binned angular observables of $B^0 \to K^{*0}\mu^+\mu^-$ decays [10] and the measurement of the $B^0 \to K^{*0}\mu^+\mu^-$ branching fraction [40]. The branching fraction measurement is included to allow better comparability with the measurement in this thesis. The signal model used in this thesis is sensitive to the relative magnitudes of the penguin amplitudes and resonant amplitudes and therefore uses branching fraction information from data. The 1σ band obtained from the Run 1 q^2-binned angular observables and branching fraction measurement is shown with the orange area in figure 7.10.

The measurement from this thesis achieves a better precision for both $\mathcal{C}_9$ and $\mathcal{C}_{10}$. This is because the analysis presented in this thesis uses more events than the previous analyses,

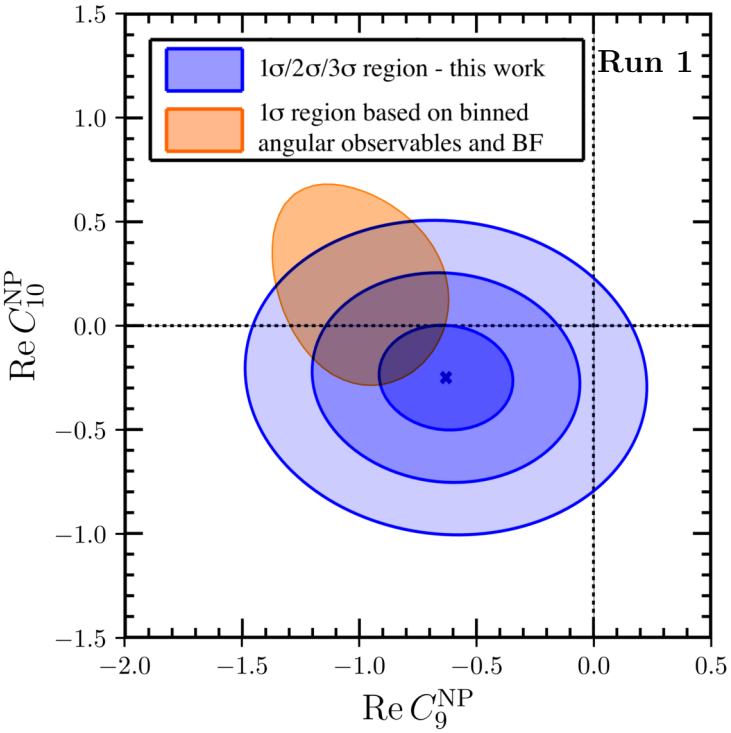


Figure 7.10: Allowed region in the $\mathcal{C}_9^{\rm NP}$ - $\mathcal{C}_{10}^{\rm NP}$ plane obtained using the best fit values, the covariance matrix, systematic uncertainties in the fit to Run 1 data presented in this thesis (blue areas). Shown are the 1σ, 2σ and 3σ regions. The best fit values for $(\mathcal{C}_9^{\rm NP}, \mathcal{C}_{10}^{\rm NP})$ are indicated with a blue cross. Also shown is the 1σ region (orange area) obtained with the FLAVIO package [111] using the q^2-binned angular observables of $B^0 \to K^{*0}\mu^+\mu^-$ decays from Run 1 [10] and the measurement of the $B^0 \to K^{*0}\mu^+\mu^-$ branching fraction [40] as input.

by fitting the full q^2 range and not omitting the resonant regions. Also, the analysis presented in this thesis is the first analysis that uses the full information encapsulated in the correlation of the events across q^2 by not binning the data. Furthermore, the non-local contributions are determined from data such that no theoretical uncertainty associated to these previously unknown contributions need to taken into account.

The latter point could also explain the fact that the central value of $(\mathcal{C}_9^{\rm NP}, \mathcal{C}_{10}^{\rm NP})$ obtained in this thesis is shifted with respect to the previous analyses. The SM calculations, which are used in $b \to s\ell^+\ell^-$ global fit packages such as EOS and FLAVIO, take the unknown non-local contributions into account only by allowing for a systematic variation of the angular observables, not by shifting the central values of the angular observables. Therefore, when fitting the q^2-binned angular observables with FLAVIO, the resulting central values of the Wilson coefficients do not take potential non-local effects into account. In contrast

to that, the fit presented in this thesis explicitly determines the relative contributions of the penguin amplitudes and the resonant amplitudes from data such that the resulting Wilson Coefficients are unaffected by the non-local contributions.[7]

7.3.1 Comparing angular observables to q^2-binned measurement

The model described in this thesis can be used to calculate the *CP*-averaged angular observables S_i (see equation 2.31 and equation 2.28 in section 2.7) as a function of q^2. This allows the comparison of the results from this thesis to the published q^2-binned angular observables measured in Run 1 data [10].

Figure 7.11 shows the full set of *CP*-averaged angular observables A_{FB}, F_{L}, S_3, S_4, S_5, S_7, S_8, and S_9 as a function of q^2 calculated with the model used in this thesis, after fitting the model to Run 1 data. The blue shaded area depicts the 68% confidence level of the model, determined by varying the parameters according to the covariance matrix obtained in the fit. Also shown with the black data points are the measured values of the q^2-binned angular observables from Ref. [10]. Note, that this is not a fit of the unbinned model to the binned observables, but rather it is a comparison of the results of two independent fits to the same data – with the exception that the events in the resonant regions, containing contain the $\phi(1020)$, J/ψ and $\psi(2S)$ resonances, are omitted in the q^2-binned fit, but are included in the fit presented in this thesis.

The angular observables calculated with the fitted model from this thesis, are in good general agreement with the directly measured binned angular observables. This is further confirmation that the model used in this thesis is a good description of the data. The fit from this thesis determines the angular observables with better precision than the q^2-binned measurement, as expected from the comparison discussed in the previous section. The discrepancies observed in the first q^2 bin are due to the fact that the model used for the q^2-binned fit assumes massless leptons, which creates a redundancy between the angular terms J_1^s and J_1^c and therefore allows less information to be extracted from the angular fit [113].

The SM predictions given in Ref. [112] (using FF input from Ref. [35]) are also shown in figure 7.11 with green boxes. Since the systematic uncertainties are not included in the confidence intervals for the measurement from this thesis in figure 7.11, no quantitative

[7] However, as discussed in section 3.2 the fit in this thesis does come with the caveat of being model dependent.

conclusions should be drawn from these comparisons.[8] Still, the comparison allows qualitative conclusions based on the central values.

The largest discrepancies between the q^2-binned measurement and SM predictions have been found in the third and fourth bin in S_5 [10]. Even though the value of S_5 at the center of the fourth bin, calculated with the model from this thesis using the best fit values from the fit to Run 1 data, is slightly closer to the SM than the q^2-binned measurement, the discrepancies remain. In the second bin of S_5 the discrepancy with the SM is increased compared to the q^2-binned measurement. Furthermore, the small discrepancies observed for A_{FB} remain. In terms of F_{L} the trend of a discrepancy between measurement and SM prediction is clearer with the measurement from this thesis as the extremely high value of F_{L} in the third bin, obtained in the q^2-binned measurement, is likely not confirmed.

[8] However, quantitative conclusions can be drawn from the direct measurement of the Wilson Coefficients $\mathcal{C}_9$ and $\mathcal{C}_{10}$ as described in the previous section.

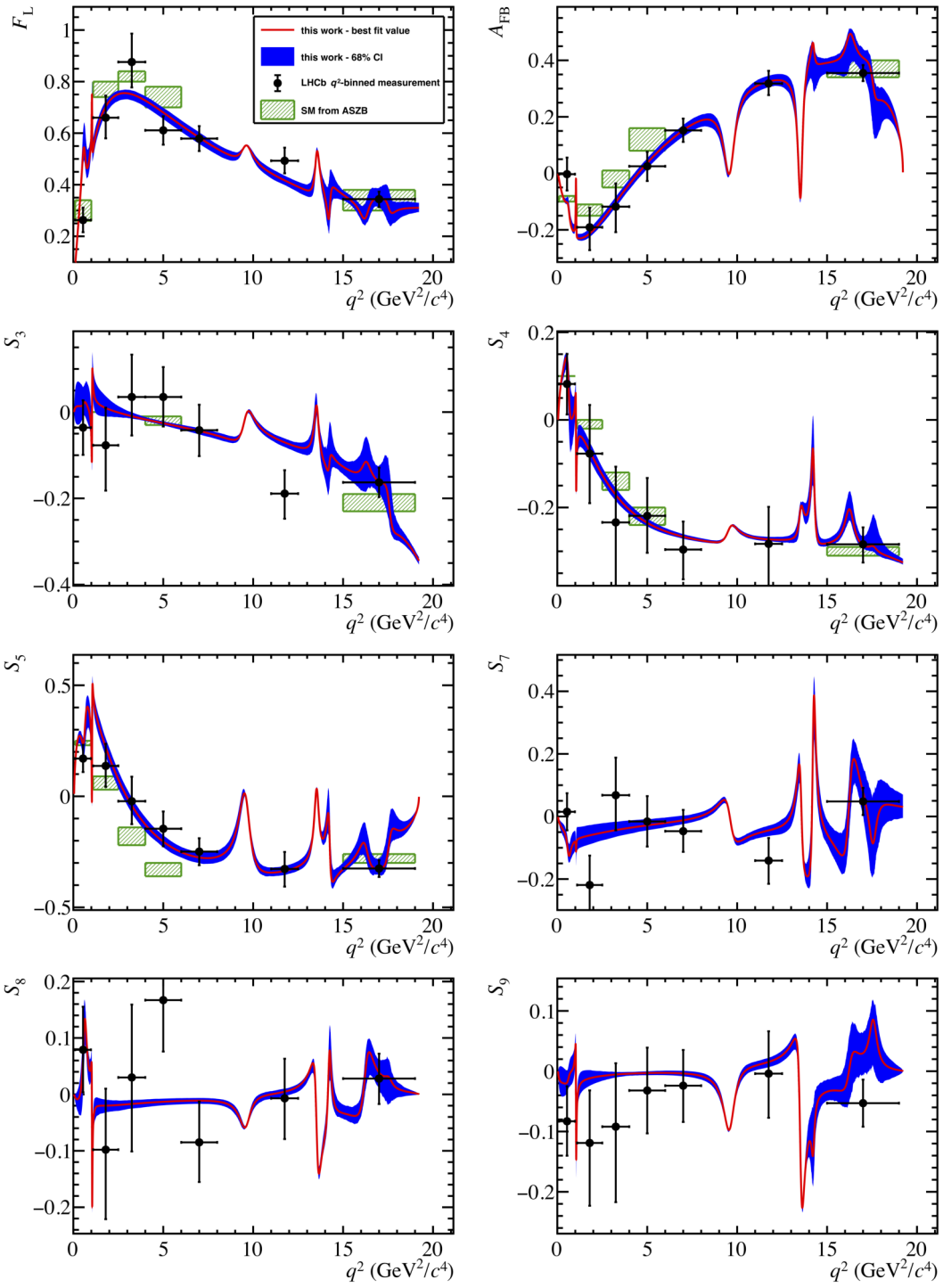


Figure 7.11: Comparison of the *CP*-averaged angular observables measured in bins of q^2 (black points) [10] to the angular observables calculated with the model described in this thesis using the best fit values from the fit to Run 1 data (red line). The blue shaded area indicates the 68% confidence interval of the model obtained by fluctuating the parameters according to the covariance matrix from the fit. Note that this is a comparison of the results of two independent fits to the same data – with the exception that the q^2-binned fit does not use the events in the resonant regions. The green boxes show the SM predictions given in Ref. [112] (using FF input from Ref. [35]).

7.3.2 Effect of the non-local contributions

To illustrate the effect of the non-local contributions on the angular observables, figure 7.12 shows $S_5(q^2)$ calculated with three variations of the model from in this thesis (with the parameters obtained in the fit to Run 1 data), as well as the measured values of S_5 in bins of q^2 from Ref. [10]. In the top left plot, only the penguin amplitudes are used to calculate S_5. The penguin amplitudes alone cannot describe the binned measurements of S_5, most notably in the range of $(3 \leq q^2 \leq 8)$ GeV$^2/c^4$. In the top right plot, the penguin amplitudes as well as the resonant amplitudes are used to calculate S_5 but any non-local effect on $\mathcal{C}_7$ are ignored. The effect of the resonances and their interference with the penguin amplitudes is clearly visible. The inclusion of the resonant amplitudes in the model improves the agreement between the un-binned and binned measurement. In the bottom plot in figure 7.12, the full model is used to calculate S_5 giving the best agreement with the binned measurement. This indicates that both the hadronic resonance contributions to $\mathcal{C}_{9,}$ as well as the non-local contributions to $\mathcal{C}_7$ are relevant in order to describe $B^0 \to K^{*0}\mu^+\mu^-$ decays, even in the range of $(3 \leq q^2 \leq 8)$ GeV$^2/c^4$, where the penguin decays make up the largest contribution to the decay rate.

To demonstrate the effect of the non-local contributions on the decay rate of $B^0 \to K^{*0}\mu^+\mu^-$ decays, figure 7.13 shows the differential decay rate of the P-Wave amplitudes calculated from the model described in this thesis after fitting it to the Run 1 data. The cyan band shows the decay rate obtained when including only the penguin amplitudes in the model. The red band shows the decay rate obtained when including the penguin amplitudes as well as resonance amplitudes in the model but omitting non-local contributions to C_7. The blue band shows the decay rate calculated with the full model, i.e. including the penguin and resonance amplitudes as well as the non-local contributions to C_7. The resonances cause destructive interference in the regions $(1 \leq q^2 \leq 8)$ GeV$^2/c^4$ and $(11 \leq q^2 \leq 13)$ GeV$^2/c^4$ while they cause constructive interference in the high q^2 region at $(16 \leq q^2 \leq 19)$ GeV$^2/c^4$. The non-local contributions to C_7 cause a decrease of the decay rate at the photon pole at $q^2 \leq 1$ GeV$^2/c^4$ but a slight overall increase at $q^2 \geq 1$ GeV$^2/c^4$.

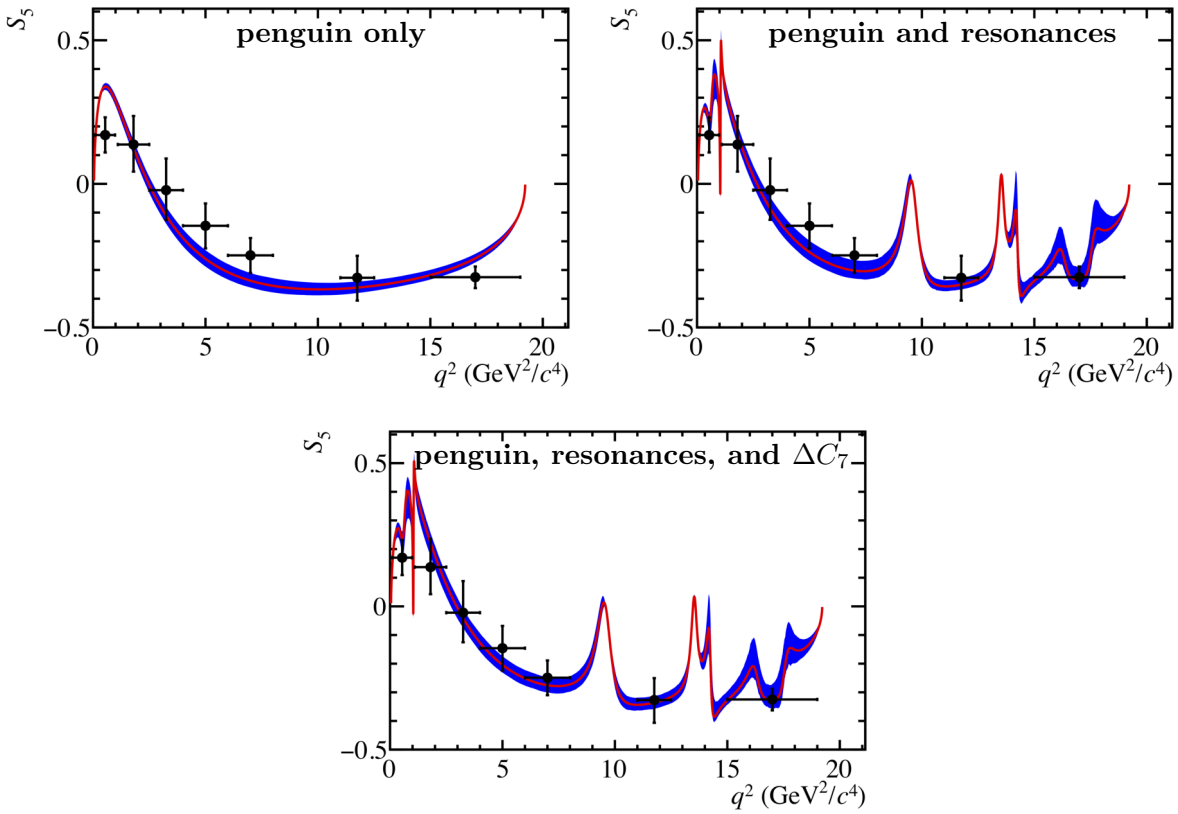


Figure 7.12: Study of the effect of the non-local contributions on the angular observable $S_5(q^2)$. Shown are the measured values of S_5 in bins of q^2 (black points) from Ref. [10] as well as the calculation of S_5 as a function of q^2 made with the model described in this thesis using the best fit values from the fit to Run 1 data (red lines). In the top left plot, the calculations are made including only the penguin amplitudes. In the top right plot, the calculations are made including the penguin amplitudes as well as all resonance amplitudes but omitting any non local contribution to C_7. In the bottom plot the calculations are made using the full model i.e. including the penguin and resonance amplitudes as well as the non-local contributions to C_7.

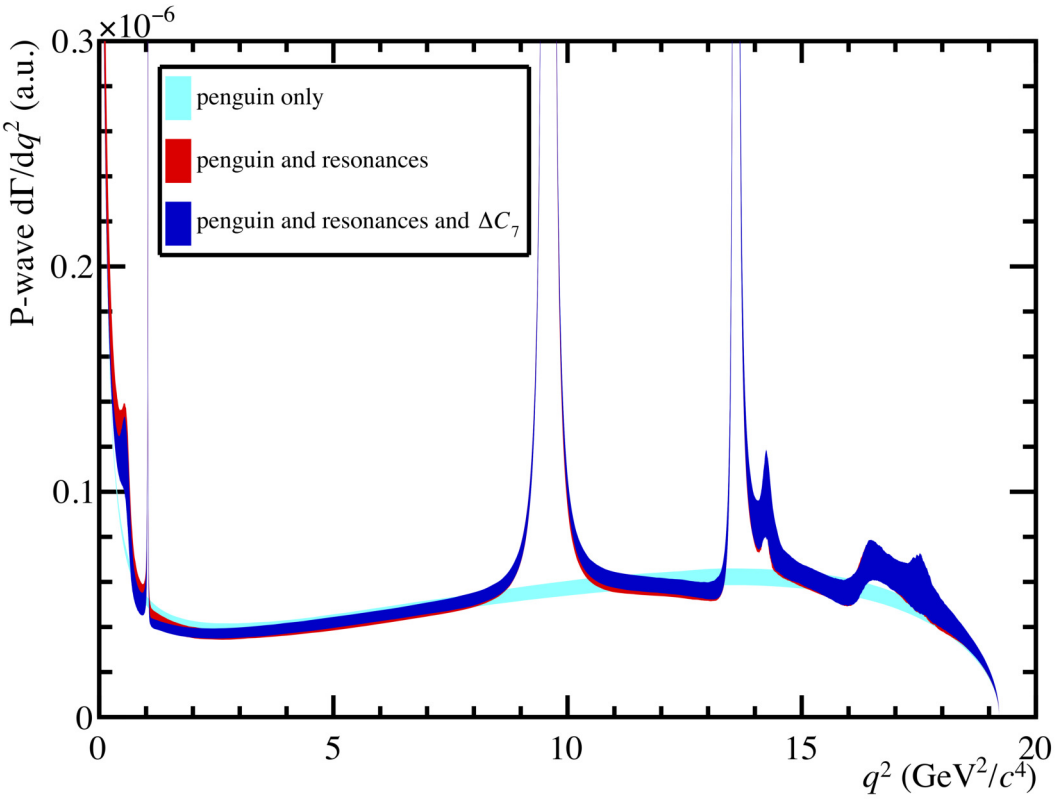


Figure 7.13: Differential decay rate of P-Wave $B^0 \to K^{*0}\mu^+\mu^-$ decays as a function of q^2, calculated with the model described in this thesis after fitting it to the Run 1 data. The bands are obtained by varying all parameters within the covariance matrix from the fit. The cyan band shows the decay rate obtained when including only the penguin amplitudes in the model. The red band shows the decay rate obtained when including the penguin amplitudes as well as resonance amplitudes in the model. The blue band shows the decay rate calculated with the full model, i.e. including the penguin and resonance amplitudes as well as the non-local contributions to C_7.

8. Conclusions

This thesis presents a novel four dimensional fit to the angular and q^2 distributions of $B^0 \to K^{*0}\mu^+\mu^-$ decays, including the resonant modes such as $B^0 \to J/\psi K^{*0}$, in Run 1 data recorded by the LHCb detector. The empirical model used for the fit is based on the decay amplitudes. This allows for the determination of the Wilson Coefficients $\mathcal{C}_9$ and $\mathcal{C}_{10}$ as well as the relative magnitudes and phases of the resonant amplitudes relative to the penguin amplitudes directly from data.
The best fit value for the magnitude of the Wilson coefficient $\mathcal{C}_9$ is $|\mathcal{C}_9| = 3.64 \pm 0.25 \pm 0.14$, where the first uncertainty includes the statistical uncertainty and the theoretical uncertainties (P-Wave form factors) and the second uncertainty is the systematic uncertainty. The result corresponds to a shift of $\Delta\mathrm{Re}(\mathcal{C}_9) = -0.63 \pm 0.29$ (where the uncertainty is the combined statistical, theoretical, and systematic uncertainty) with respect to the SM central value of $\mathrm{Re}(\mathcal{C}_9) = 4.27$. Therefore, the measurement agrees with the SM only at 2.2 standard deviations. This indicates that there remains a discrepancy between the measured and predicted angular distributions of $B^0 \to K^{*0}\mu^+\mu^-$ decays even when taking the non-local contributions into account.
The empirical model is capable of determining the Wilson Coefficients $\mathcal{C}_9$ and $\mathcal{C}_{10}$ with better higher precision than previous analyses of $B^0 \to K^{*0}\mu^+\mu^-$, since it uses the full information contained in the q^2 dependence and correlations of the angular observables.
The precision on the phases of the J/ψ and $\psi(2S)$ amplitudes relative to the penguin amplitudes are on the level of 8.5% and 20%, which is precise enough to ascertain the level of the non-local contributions. In terms of the P-Wave decay rate, the resonances lead to mild destructive interference in the regions $(1 \leq q^2 \leq 8)$ GeV$^2/c^4$ and $(10 \leq q^2 \leq 13)$ GeV$^2/c^4$ and cause constructive interference in the high q^2 region at $(16 \leq q^2 \leq 19)$ GeV$^2/c^4$. Comparing the angular observables, calculated with the model from this thesis, to the q^2 binned measurement of the angular observables [10] reveals that the penguin amplitudes alone cannot describe the angular observables. Both the hadronic contribution to $\mathcal{C}_9$ as well as the non-local contributions to $\mathcal{C}_7$ are necessary to reproduce the q^2 binned measurement.

The merged Run 1 and Run 2 dataset contains approximately 3 times more events than Run 1 only. Assuming that the statistical uncertainties scale with $\sim 1/\sqrt{3}$, while the systematic uncertainties and the central values remain the same, the expected precision for $\mathcal{C}_9$ is $\sim 5\%$ and the precision for $\mathcal{C}_{10}$ is $\sim 4\%$ for the fit to the merged Run 1 and Run 2 data.

Bibliography

[1] LHCb, R. Aaij *et al.*, *Measurement of CP-averaged observables in the $B^0 \to K^{*0}\mu^+\mu^-$ decay*, Phys. Rev. Lett. **125** (2020) 011802, arXiv:2003.04831.

[2] T. Blake, U. Egede, P. Owen, K. A. Petridis, and G. Pomery, *An empirical model to determine the hadronic resonance contributions to $\overline{B}^0 \to \overline{K}^{*0}\mu^+\mu^-$ transitions*, Eur. Phys. J. **C78** (2018) 453, arXiv:1709.03921.

[3] M. Schumann, *Dark Matter 2014*, EPJ Web Conf. **96** (2015) 01027, arXiv:1501.01200.

[4] A. D. Sakharov, *Violation ofCPin variance,casymmetry, and baryon asymmetry of the universe*, Soviet Physics Uspekhi **34** (1991) 392.

[5] LHCb, R. Aaij *et al.*, *Differential branching fractions and isospin asymmetries of $B \to K^{(*)}\mu^+\mu^-$ decays*, JHEP **06** (2014) 133, arXiv:1403.8044.

[6] LHCb, R. Aaij *et al.*, *Angular analysis and differential branching fraction of the decay $B_s^0 \to \phi\mu^+\mu^-$*, JHEP **09** (2015) 179, arXiv:1506.08777.

[7] LHCb, R. Aaij *et al.*, *Differential branching fraction and angular analysis of $\Lambda_b^0 \to \Lambda\mu^+\mu^-$ decays*, JHEP **06** (2015) 115, arXiv:1503.07138, [Erratum: JHEP 09, 145 (2018)].

[8] LHCb, R. Aaij *et al.*, *Differential branching fraction and angular analysis of the decay $B^0 \to K^{*0}\mu^+\mu^-$*, JHEP **08** (2013) 131, arXiv:1304.6325.

[9] LHCb, R. Aaij *et al.*, *Measurement of Form-Factor-Independent Observables in the Decay $B^0 \to K^{*0}\mu^+\mu^-$*, Phys. Rev. Lett. **111** (2013) 191801, arXiv:1308.1707.

[10] LHCb, R. Aaij *et al.*, *Angular analysis of the $B^0 \to K^{*0}\mu^+\mu^-$ decay using 3 fb^{-1} of integrated luminosity*, JHEP **02** (2016) 104, arXiv:1512.04442.

[11] LHCb, R. Aaij *et al.*, *Angular moments of the decay $\Lambda_b^0 \to \Lambda\mu^+\mu^-$ at low hadronic recoil*, JHEP **09** (2018) 146, arXiv:1808.00264.

[12] LHCb, R. Aaij *et al.*, *Angular analysis of the $B^0 \to K^{*0}e^+e^-$ decay in the low-q^2 region*, JHEP **04** (2015) 064, arXiv:1501.03038.

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

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

[15] LHCb, R. Aaij *et al.*, *Search for lepton-universality violation in $B^+ \to K^+\ell^+\ell^-$ decays*, Phys. Rev. Lett. **122** (2019) 191801, arXiv:1903.09252.

[16] M. Algueró *et al.*, *Emerging patterns of New Physics with and without Lepton Flavour Universal contributions*, Eur. Phys. J. C **79** (2019) 714, arXiv:1903.09578, [Addendum: Eur.Phys.J.C 80, 511 (2020)].

[17] B. Gripaios, M. Nardecchia, and S. A. Renner, *Composite leptoquarks and anomalies in B-meson decays*, JHEP **05** (2015) 006, arXiv:1412.1791.

[18] M. Thomson, *Modern particle physics*, Cambridge University Press, New York, 2013.

[19] M. D. Schwartz, *Quantum Field Theory and the Standard Model*, Cambridge University Press, 2014.

[20] G. Cowan and T. Gershon, *Tetraquarks and Pentaquarks*, 2399-2891, IOP Publishing, 2018.

[21] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) 508.

[22] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (1964) 321.

[23] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Global conservation laws and massless particles*, Phys. Rev. Lett. **13** (1964) 585.

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

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

[26] M. Kobayashi and T. Maskawa, *CP-Violation in the renormalizable theory of weak interaction*, Prog. Theor. Phys. **49** (1973) 652.

[27] L. Wolfenstein, *Parametrization of the kobayashi-maskawa matrix*, Phys. Rev. Lett. **51** (1983) 1945.

[28] Planck, P. A. R. Ade *et al.*, *Planck 2013 results. XXII. Constraints on inflation*, Astron. Astrophys. **571** (2014) A22, arXiv:1303.5082.

[29] *Feynman Diagram Library*, https://www.physik.uzh.ch/~che/FeynDiag/index.php. Accessed: 2020-07-08.

[30] Particle Data Group, M. Tanabashi *et al.*, *Review of particle physics*, Phys. Rev. **D98** (2018) 030001.

[31] K. G. Wilson, *Non-lagrangian models of current algebra*, Phys. Rev. **179** (1969) 1499.

[32] W. Altmannshofer *et al.*, *Symmetries and Asymmetries of $B \to K^* \mu^+ \mu^-$ Decays in the Standard Model and Beyond*, JHEP **01** (2009) 019, arXiv:0811.1214.

[33] C. Bobeth, M. Misiak, and J. Urban, *Photonic penguins at two loops and m_t dependence of $BR[B \to X_s l^+ l^-]$*, Nucl. Phys. B **574** (2000) 291, arXiv:hep-ph/9910220.

[34] M. Beneke, T. Feldmann, and D. Seidel, *Exclusive radiative and electroweak $b \to d$ and $b \to s$ penguin decays at NLO*, Eur. Phys. J. C **41** (2005) 173, arXiv:hep-ph/0412400.

[35] A. Bharucha, D. M. Straub, and R. Zwicky, *$B \to V \ell^+ \ell^-$ in the Standard Model from light-cone sum rules*, JHEP **08** (2016) 098, arXiv:1503.05534.

[36] S. Descotes-Genon, A. Khodjamirian, and J. Virto, *Light-cone sum rules for $B \to K\pi$ form factors and applications to rare decays*, JHEP **12** (2019) 083, arXiv:1908.02267.

[37] J. Gratrex, M. Hopfer, and R. Zwicky, *Generalised helicity formalism, higher moments and the $B \to K_{J_K}(\to K\pi)\bar{\ell}_1 \ell_2$ angular distributions*, Phys. Rev. **D93** (2016) 054008, arXiv:1506.03970.

[38] J. Matias, F. Mescia, M. Ramon, and J. Virto, *Complete Anatomy of $\bar{B}_d -> \bar{K}^{*0}(-> K\pi) l^+ l^-$ and its angular distribution*, JHEP **04** (2012) 104, arXiv:1202.4266.

[39] S. Descotes-Genon, J. Matias, M. Ramon, and J. Virto, *Implications from clean observables for the binned analysis of $B- > K*\mu^+\mu^-$ at large recoil*, JHEP **01** (2013) 048, arXiv:1207.2753.

[40] LHCb, R. Aaij *et al.*, *Measurements of the S-wave fraction in $B^0 \to K^+\pi^-\mu^+\mu^-$ decays and the $B^0 \to K^*(892)^0\mu^+\mu^-$ differential branching fraction*, JHEP **11** (2016) 047, arXiv:1606.04731, [Erratum: JHEP 04, 142 (2017)].

[41] D. Becirevic and A. Tayduganov, *Impact of $B \to K_0^*\ell^+\ell^-$ on the New Physics search in $B \to K^*\ell^+\ell^-$ decay*, Nucl. Phys. B **868** (2013) 368, arXiv:1207.4004.

[42] R.-H. Li, C.-D. Lu, W. Wang, and X.-X. Wang, *$B \to S$ Transition Form Factors in the PQCD approach*, Phys. Rev. D **79** (2009) 014013, arXiv:0811.2648.

[43] M. Döring, U.-G. Meißner, and W. Wang, *Chiral Dynamics and S-wave Contributions in Semileptonic B decays*, JHEP **10** (2013) 011, arXiv:1307.0947.

[44] J. Aebischer *et al.*, *B-decay discrepancies after Moriond 2019*, Eur. Phys. J. C **80** (2020) 252, arXiv:1903.10434.

[45] T. Humair, *Testing lepton universality in penguin decays of beauty mesons using the LHCb detector*, May, 2019. Presented 02 May 2019.

[46] M. Beneke and T. Feldmann, *Symmetry breaking corrections to heavy to light B meson form-factors at large recoil*, Nucl. Phys. B **592** (2001) 3, arXiv:hep-ph/0008255.

[47] C. Bobeth, M. Chrzaszcz, D. van Dyk, and J. Virto, *Long-distance effects in $B \to K^*\ell\ell$ from analyticity*, Eur. Phys. J. C **78** (2018) 451, arXiv:1707.07305.

[48] J. Lyon and R. Zwicky, *Resonances gone topsy turvy - the charm of QCD or new physics in $b \to s\ell^+\ell^-$?*, arXiv:1406.0566.

[49] LHCb collaboration, R. Aaij *et al.*, *Measurement of the phase difference between short- and long-distance amplitudes in the $B^+ \to K^+\mu^+\mu^-$ decay*, Eur. Phys. J. **C77** (2017) 161, arXiv:1612.06764.

[50] G. J. Pomery, *Understanding the impact of non-local contributions to $\bar{B}^0 \to \bar{K}^{*0}\mu^+\mu^-$ transitions*, Sep, 2019. Presented 21 Nov 2019.

[51] S. Braß, G. Hiller, and I. Nisandzic, *Zooming in on $B \to K^*\ell\ell$ decays at low recoil*, Eur. Phys. J. C **77** (2017) 16, arXiv:1606.00775.

[52] LHCb, R. Aaij *et al.*, *Measurement of the polarization amplitudes in $B^0 \to J/\psi K^*(892)^0$ decays*, Phys. Rev. D **88** (2013) 052002, arXiv:1307.2782.

[53] Belle, K. Chilikin *et al.*, *Experimental constraints on the spin and parity of the $Z(4430)^+$*, Phys. Rev. **D88** (2013) 074026, arXiv:1306.4894.

[54] BaBar, B. Aubert *et al.*, *Measurement of decay amplitudes of $B \to J/\psi K^*, \psi(2S)K^*$, and $\chi_{c1}K^*$ with an angular analysis*, Phys. Rev. D **76** (2007) 031102, arXiv:0704.0522.

[55] Belle, K. Chilikin *et al.*, *Observation of a new charged charmoniumlike state in $\bar{B}^0 \to J/\psi K^-\pi^+$ decays*, Phys. Rev. D **90** (2014) 112009, arXiv:1408.6457.

[56] LHCb collaboration, R. Aaij *et al.*, *Measurement of polarization amplitudes and CP asymmetries in $B^0 \to \phi K^*(892)^0$*, JHEP **05** (2014) 069, arXiv:1403.2888.

[57] Belle, M. Prim *et al.*, *Angular analysis of $B^0 \to \phi K^*$ decays and search for CP violation at Belle*, Phys. Rev. D **88** (2013) 072004, arXiv:1308.1830.

[58] BaBar, B. Aubert *et al.*, *Time-Dependent and Time-Integrated Angular Analysis of $B \to \phi K_s \pi^0$ and $B \to \phi K^+\pi^-$*, Phys. Rev. D **78** (2008) 092008, arXiv:0808.3586.

[59] A. Khodjamirian, T. Mannel, A. A. Pivovarov, and Y.-M. Wang, *Charm-loop effect in $B \to K^{(*)}\ell^+\ell^-$ and $B \to K^*\gamma$*, JHEP **09** (2010) 089, arXiv:1006.4945.

[60] D. Aston *et al.*, *A study of $K^-\pi^+$ scattering in the reaction $K^-p \to K^-\pi^+$ at 11 GeV/c*, Nuclear Physics B **296** (1988) 493 .

[61] A. Paul and D. M. Straub, *Constraints on new physics from radiative b decays*, Journal of High Energy Physics **2017** (2017) 27.

[62] A. Horvath, *"The LHC experiments and the preaccelerators"*, licensed under Creative Commons Attribution- Share Alike 2.5 Generic license, Wikimedia Commons, 2006, https://commons.wikimedia.org/wiki/File:LHC.svg.

[63] O. S. Bruning *et al.*, *LHC Design Report Vol.1: The LHC Main Ring*, 2004. CERN-2004-003-V1.

[64] ATLAS collaboration, G. Aad *et al.*, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[65] CMS collaboration, S. Chatrchyan *et al.*, *The CMS Experiment at the CERN LHC*, JINST **3** (2008) S08004.

[66] 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.

[67] 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.

[68] ALICE collaboration, K. Aamodt *et al.*, *The ALICE experiment at the CERN LHC*, JINST **3** (2008) S08002.

[69] D. M. Jamil and P. J. T. Rhee, *Quark–gluon plasma, by K. Yagi, T. Hatsuda and Y. Miake*, Contemporary Physics **50** (2009) 665, https://doi.org/10.1080/00107510902978246.

[70] LHCb collaboration, A. A. Alves Jr *et al.*, *The LHCb detector at the LHC*, JINST **3** (2008) S08005.

[71] T. Sjöstrand, S. Mrenna, and P. Skands, *A brief introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, arXiv:0710.3820.

[72] LHCb collaboration, C. Elsässer, *$\bar{b}b$ production angle plots*, https://lhcb.web.cern.ch/lhcb/speakersbureau/html/bb_ProductionAngles.html, Accessed: 2020-05-05.

[73] LHCb collaboration, R. Aaij *et al.*, *Measurement of the b-quark production cross-section in 7 and 13 TeV pp collisions*, Phys. Rev. Lett. **118** (2017) 052002, arXiv:1612.05140, [Erratum: Phys. Rev. Lett.119,no.16,169901(2017)].

[74] LHCb collaboration, *LHCb VELO (VErtex LOcator): Technical Design Report*, CERN-LHCC-2001-011, LHCb-TDR-005.

[75] R. Aaij *et al.*, *Performance of the LHCb Vertex Locator*, JINST **9** (2014) P09007, arXiv:1405.7808.

[76] LHCb collaboration, *LHCb inner tracker: Technical Design Report*, CERN-LHCC-2002-029, LHCb-TDR-008.

[77] LHCb collaboration, *LHCb outer tracker: Technical Design Report*, CERN-LHCC-2001-024, LHCb-TDR-006.

[78] LHCb collaboration, *LHCb RICH: Technical Design Report*, CERN-LHCC-2000-037, LHCb-TDR-003.

[79] LHCb RICH Group, M. Adinolfi *et al.*, *Performance of the LHCb RICH detector at the LHC*, Eur. Phys. J. C **73** (2013) 2431, arXiv:1211.6759.

[80] J. Alves, A.Ã. *et al.*, *Performance of the LHCb muon system*, JINST **8** (2013) P02022, arXiv:1211.1346.

[81] LHCb collaboration, *LHCb muon system: Technical Design Report*, CERN-LHCC-2001-010, LHCb-TDR-004.

[82] F. Archilli *et al.*, *Performance of the Muon Identification at LHCb*, JINST **8** (2013) P10020, arXiv:1306.0249.

[83] LHCb, 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.

[84] LHCb HLT project, J. Albrecht, V. V. Gligorov, G. Raven, and S. Tolk, *Performance of the LHCb High Level Trigger in 2012*, J. Phys. Conf. Ser. **513** (2014) 012001, arXiv:1310.8544.

[85] R. Aaij *et al.*, *The LHCb Trigger and its Performance in 2011*, JINST **8** (2013) P04022, arXiv:1211.3055.

[86] T. Sjöstrand, S. Mrenna, and P. Skands, *PYTHIA 6.4 physics and manual*, JHEP **05** (2006) 026, arXiv:hep-ph/0603175.

[87] D. J. Lange, *The evtgen particle decay simulation package*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **462** (2001) 152 , BEAUTY2000, Proceedings of the 7th Int. Conf. on B-Physics at Hadron Machines.

[88] S. Agostinelli *et al.*, *Geant4—a simulation toolkit*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **506** (2003) 250 .

[89] G. Corti, *Software for the LHCb experiment*, , Proceedings published in IEEE Transactions on Nuclear Science, Vol.53, Issue 3(3), pp.1323-1328, 2006.

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

[91] P. Alvarez Cartelle *et al.*, *Angular analysis of $B^0 \to K^{*0}\mu^+\mu^-$ decays using Run 1 and 2016 data*, LHCb-ANA-2017-055 (2020), LHCb internal.

[92] C. Parkinson, *The angular analysis of the $B^0 \to K^{*0}\mu^+\mu^-$ decay at LHCb*, Jun, 2013. Presented 26 Jul 2013.

[93] S. Coquereau *et al.*, *The $B^0 \to K^{*0}\mu^+\mu^-$ selection for 3fb^{-1} of LHCb data*, LHCb-INT-2013-058 (2013), LHCb internal.

[94] B. P. Roe *et al.*, *Boosted decision trees, an alternative to artificial neural networks*, Nucl. Instrum. Meth. A **543** (2005) 577, arXiv:physics/0408124.

[95] LHCb collaboration, R. Aaij *et al.*, *Angular analysis of the $B^0 \to K^{*0}\mu^+\mu^-$ decay*, arXiv:1512.04442, submitted to JHEP.

[96] LHCb, R. Aaij *et al.*, *Measurement of CP-averaged observables in the $B^0 \to K^{*0}\mu^+\mu^-$ decay*, arXiv:2003.04831.

[97] S. U. Papoulis, Athanasios Papoulis; Pillai, *Probability, Random Variables, and Stochastic Processes (4th ed.)*, McGraw-Hill, Boston, 2002.

[98] T. Skwarnicki, *A study of the radiative cascade transitions between the Upsilon-prime and Upsilon resonances*, PhD thesis, Institute of Nuclear Physics, Krakow, 1986, DESY-F31-86-02.

[99] M. Williams, *How good are your fits? Unbinned multivariate goodness-of-fit tests in high energy physics*, JINST **5** (2010) P09004, arXiv:1006.3019.

[100] LHCb, R. Aaij *et al.*, *Evidence for the decay $B^0_S \to \overline{K}^{*0}\mu^+\mu^-$*, JHEP **07** (2018) 020, arXiv:1804.07167.

[101] Belle, S. K. Choi *et al.*, *Observation of a resonance-like structure in the $\pi^\pm\psi'$ mass distribution in exclusive $B \to K\pi^\pm\psi'$ decays*, Phys. Rev. Lett. **100** (2008) 142001, arXiv:0708.1790.

[102] Belle, R. Mizuk *et al.*, *Dalitz analysis of $B \to K\pi^+\psi'$ decays and the Z(4430)+*, Phys. Rev. **D80** (2009) 031104, arXiv:0905.2869.

[103] LHCb, R. Aaij *et al.*, *Observation of the resonant character of the $Z(4430)^-$ state*, Phys. Rev. Lett. **112** (2014) 222002, arXiv:1404.1903.

[104] Belle, K. Chilikin *et al.*, *Experimental constraints on the spin and parity of the $Z(4430)^+$*, Phys. Rev. **D88** (2013) 074026, arXiv:1306.4894.

[105] Belle, K. Chilikin *et al.*, *Observation of a new charged charmoniumlike state in $\overline{B}^0 \to J/\psi K^-\pi^+$ decays*, Phys. Rev. **D90** (2014) 112009, arXiv:1408.6457.

[106] LHCb, R. Aaij *et al.*, *Measurements of the S-wave fraction in $B^0 \to K^+\pi^-\mu^+\mu^-$ decays and the $B^0 \to K^*(892)^0\mu^+\mu^-$ differential branching fraction*, JHEP **11** (2016) 047, arXiv:1606.04731, [Erratum: JHEP04,142(2017)].

[107] J. Gratrex, M. Hopfer, and R. Zwicky, *Generalised helicity formalism, higher moments and the $B \to K_{J_K}(\to K\pi)\bar{\ell}_1\ell_2$ angular distributions*, Phys. Rev. **D93** (2016) 054008, arXiv:1506.03970.

[108] J. A. Bailey *et al.*, *$B \to Kl^+l^-$ Decay Form Factors from Three-Flavor Lattice QCD*, Phys. Rev. **D93** (2016) 025026, arXiv:1509.06235.

[109] C. Bobeth, G. Hiller, and D. van Dyk, *The Benefits of $\bar{B} -> \bar{K}^*l^+l^-$ Decays at Low Recoil*, JHEP **07** (2010) 098, arXiv:1006.5013.

[110] F. Beaujean, C. Bobeth, and D. van Dyk, *Comprehensive Bayesian analysis of rare (semi)leptonic and radiative B decays*, Eur. Phys. J. C **74** (2014) 2897, arXiv:1310.2478, [Erratum: Eur.Phys.J.C 74, 3179 (2014)].

[111] D. M. Straub, *flavio: a Python package for flavour and precision phenomenology in the Standard Model and beyond*, arXiv:1810.08132.

[112] W. Altmannshofer and D. M. Straub, *New physics in $b \to s$ transitions after LHC run 1*, Eur. Phys. J. C **75** (2015) 382, arXiv:1411.3161.

[113] U. Egede, T. Hurth, J. Matias, M. Ramon, and W. Reece, *New observables in the decay mode $\bar{B}_d \to \bar{K}^{*0}l^+l^-$*, JHEP **11** (2008) 032, arXiv:0807.2589.

A. Appendix

A.1 2D projections of the phase space affected by the $B^+ \to K^+\mu^+\mu^-$ veto

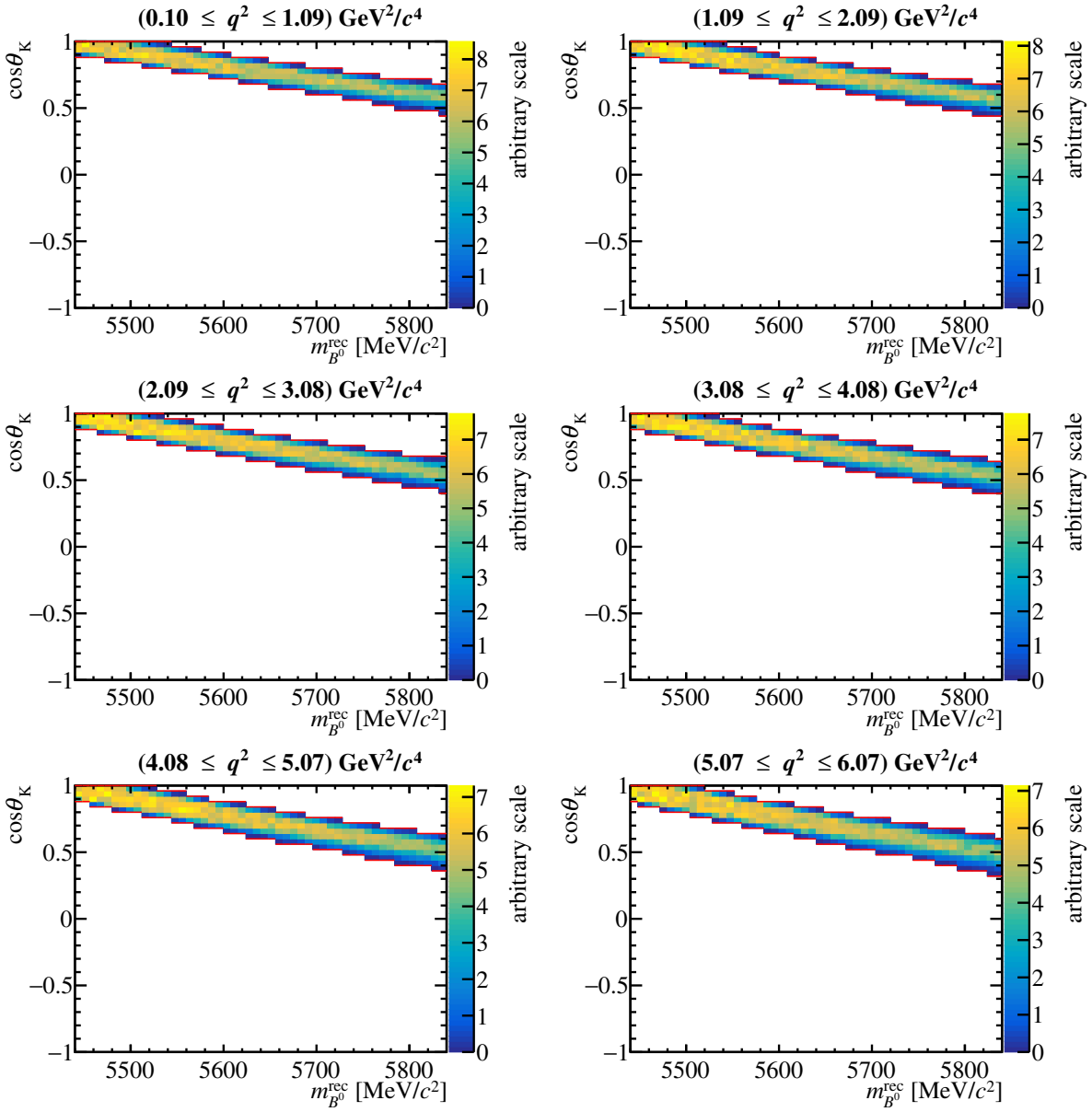


Figure A.1: $(m^{\text{rec}}_{B^0}, \cos\theta_K)$-projections of the $K^+\mu^+\mu^-$ veto affected phase space in several bins in q^2.

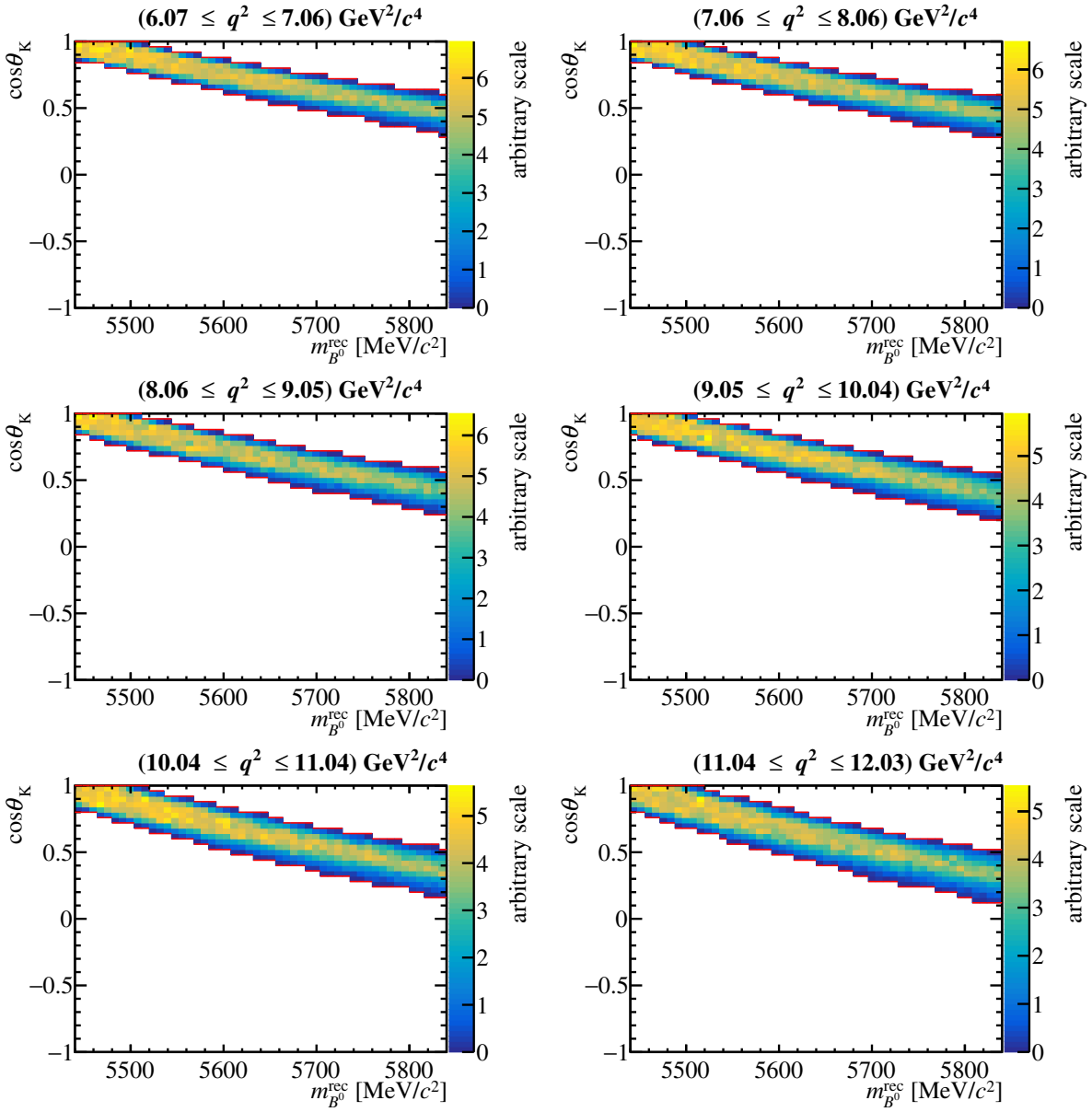


Figure A.2: $(m^{\text{rec}}_{B^0}, \cos\theta_K)$-projections of the $K^+\mu^+\mu^-$ veto affected phase space in several bins in q^2.

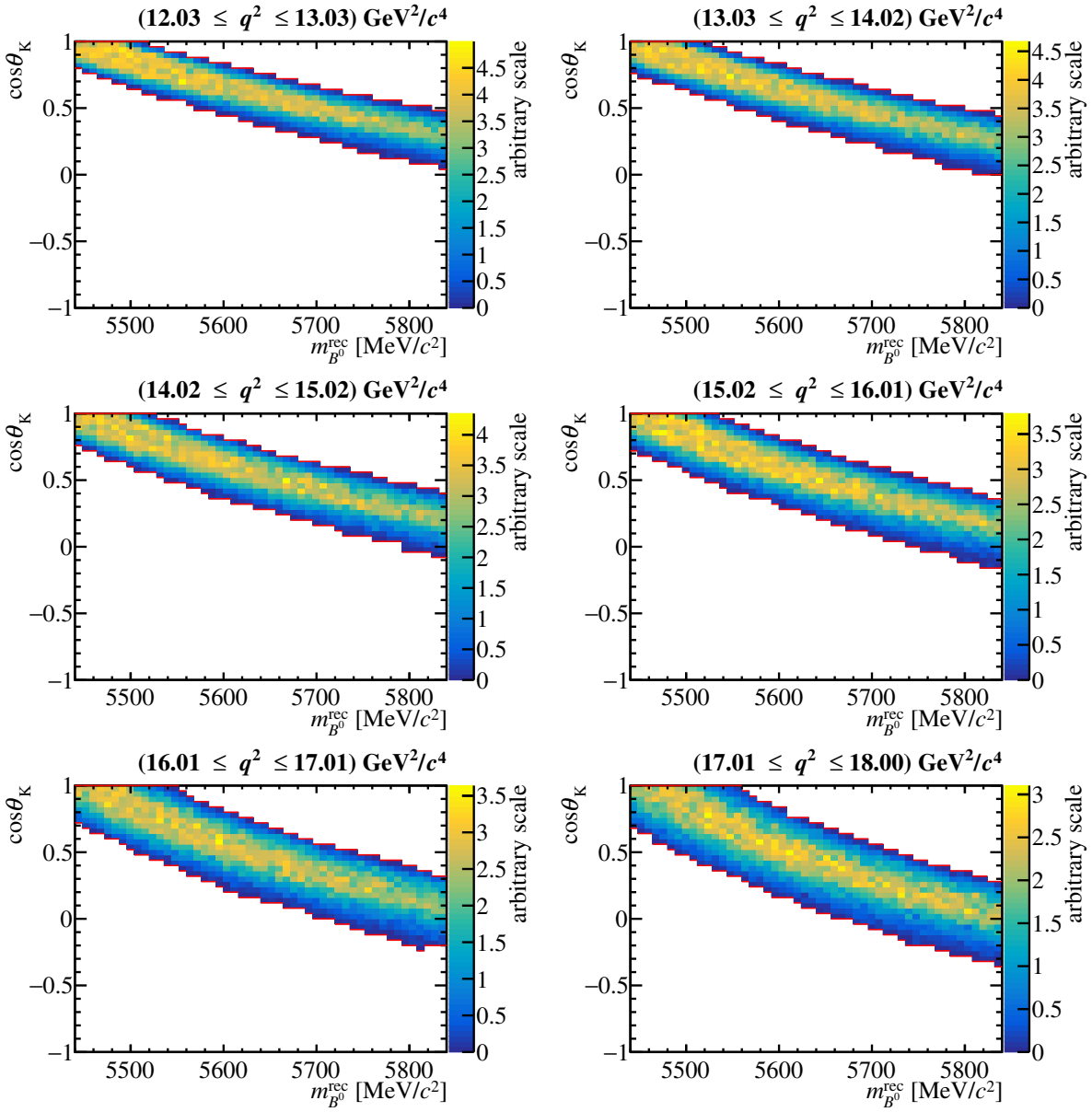


Figure A.3: $(m^{\text{rec}}_{B^0}, \cos\theta_K)$-projections of the $K^+\mu^+\mu^-$ veto affected phase space in several bins in q^2.

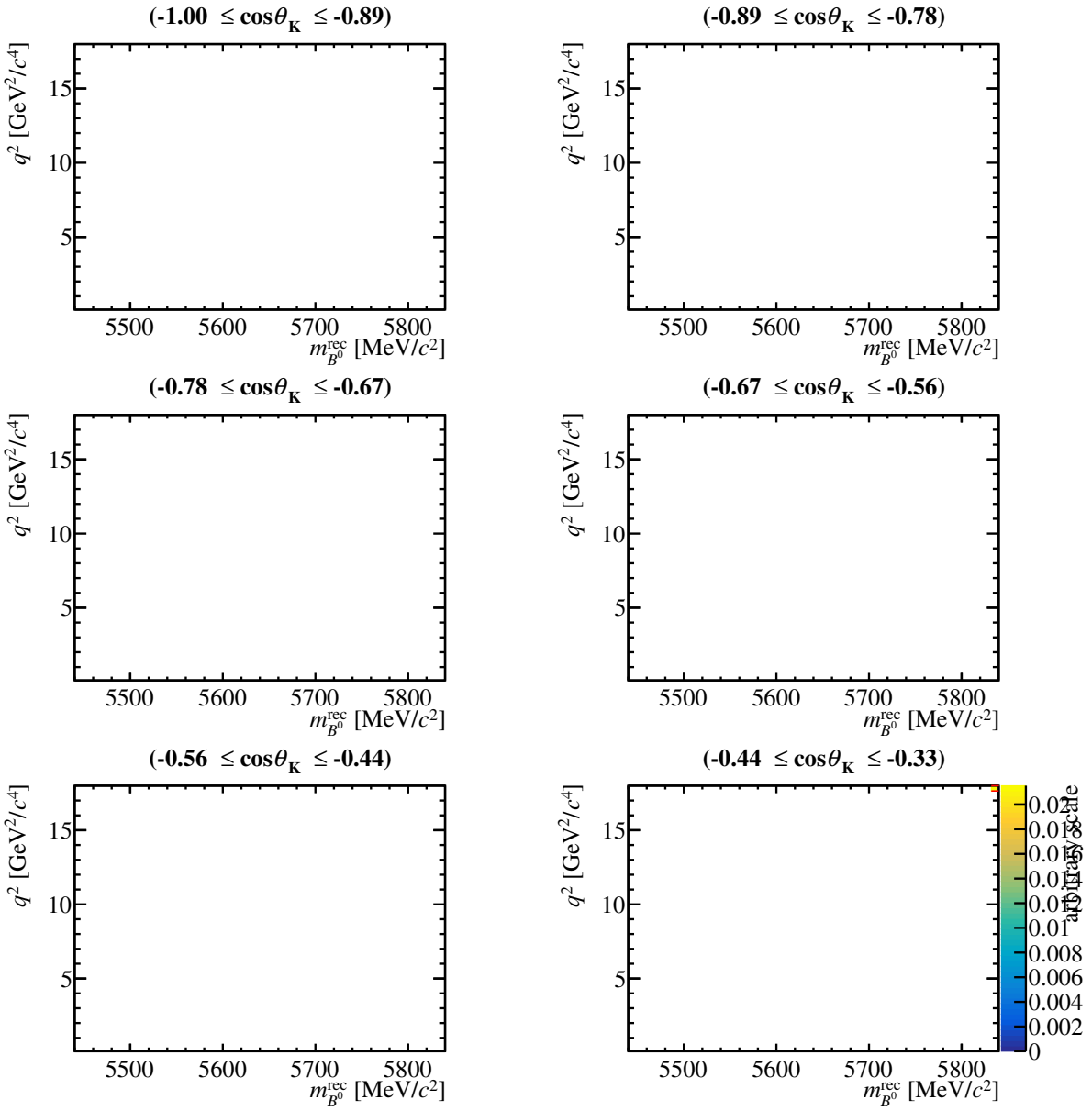


Figure A.4: $(m^{\text{rec}}_{B^0}, q^2)$-projections of the $K^+\mu^+\mu^-$ veto affected phase space in several bins in $\cos\theta_K$.

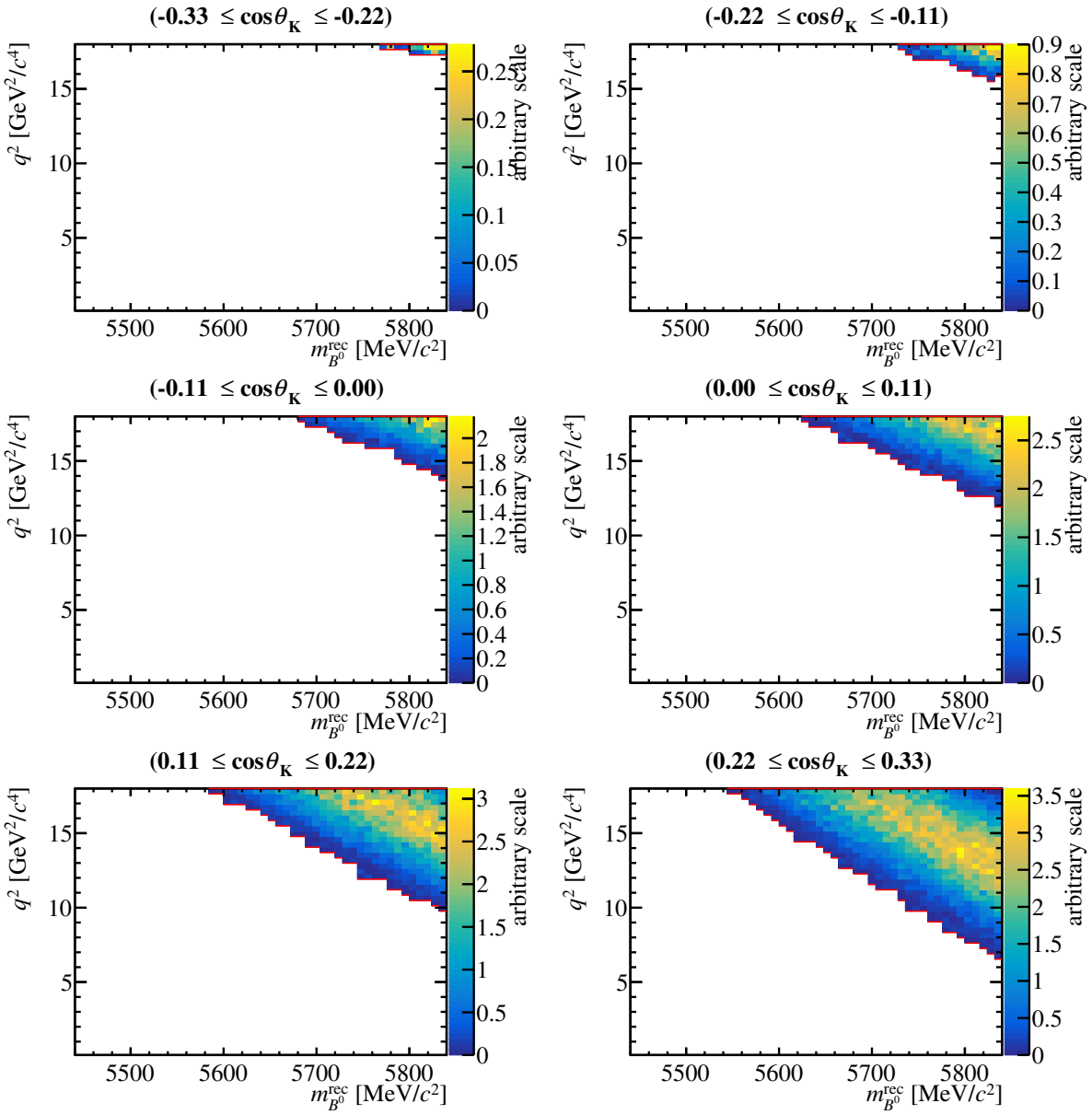


Figure A.5: $(m^{\text{rec}}_{B^0}, q^2)$-projections of the $K^+\mu^+\mu^-$ veto affected phase space in several bins in $\cos\theta_K$.

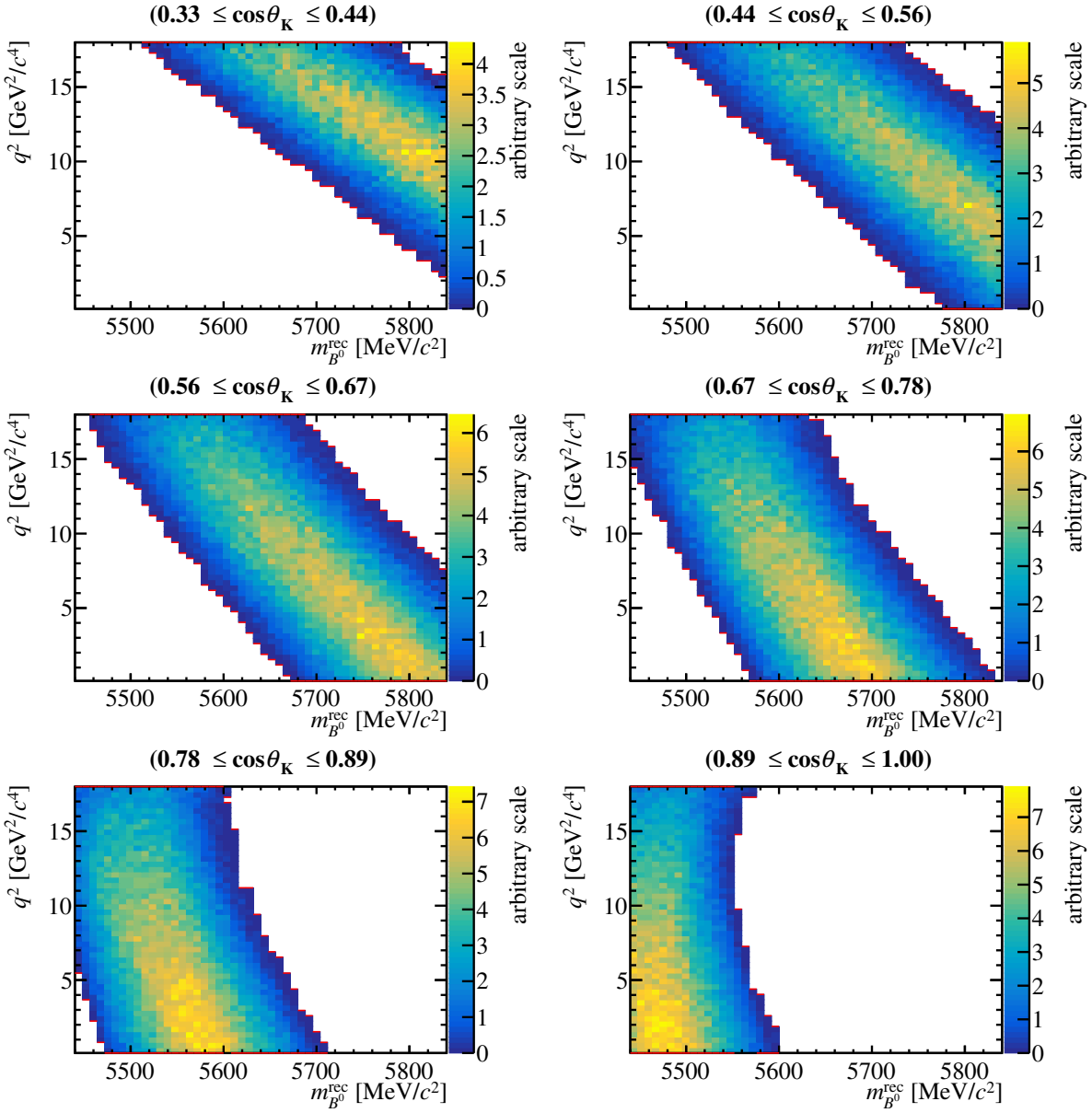


Figure A.6: $(m^{\text{rec}}_{B^0}, q^2)$-projections of the $K^+\mu^+\mu^-$ veto affected phase space in several bins in $\cos\theta_K$.

A.2 Full fit results of the fit to Run 1 data

	results $\pm\sigma_{\text{fit}}$		result $\pm\sigma_{\text{fit}}$
$\lvert A_{\parallel}^{J/\psi}\rvert$	$(4.344 \pm 0.023)\times10^{-3}$	$Im(A_{\perp}^{\psi(4160)})$	$(6.982 \pm 27.141)\times10^{-6}$
$\theta_{\parallel}^{J/\psi}$	0.207 ± 0.012	$Re(A_{0}^{\psi(4160)})$	$(2.186 \pm 4.037)\times10^{-5}$
$\lvert A_{\perp}^{J/\psi}\rvert$	$(4.178 \pm 0.024)\times10^{-3}$	$Im(A_{0}^{\psi(4160)})$	$(1.56 \pm 3.93)\times10^{-5}$
$\theta_{\perp}^{J/\psi}$	-6.5243 ± 0.0095	α_0^{A1}	0.2963 ± 0.0099
$\theta_{0}^{J/\psi}$	-1.63 ± 0.10	α_1^{A1}	0.396 ± 0.057
$\lvert A_{\parallel}^{\psi(2S)}\rvert$	$(1.073 \pm 0.045)\times10^{-3}$	α_2^{A1}	1.30 ± 0.50
$\theta_{\parallel}^{\psi(2S)}$	-0.674 ± 0.048	α_0^{A12}	0.2682 ± 0.0098
$\lvert A_{\perp}^{\psi(2S)}\rvert$	$(1.065 \pm 0.046)\times10^{-3}$	α_1^{A12}	0.536 ± 0.068
$\theta_{\perp}^{\psi(2S)}$	3.640 ± 0.045	α_2^{A12}	0.47 ± 0.39
$\lvert A_{0}^{\psi(2S)}\rvert$	$(1.596 \pm 0.060)\times10^{-3}$	α_0^{V}	0.375 ± 0.013
$\theta_{0}^{\psi(2S)}$	-1.72 ± 0.32	α_1^{V}	-1.166 ± 0.074
$\lvert A_{0}^{\rho(770)})\rvert$	$(8.04 \pm 3.29)\times10^{-5}$	α_2^{V}	2.5 ± 1.3
$\theta_{0}^{\rho(770)}$	$(-16.953 \pm 41.249)\times10^{-6}$	$\lvert A_{00}^{J/\psi}\rvert$	$(2.990 \pm 0.051)\times10^{-3}$
$Im(A_{0}^{\phi(1020)})$	$(6.82 \pm 1.60)\times10^{-5}$	$\theta_{00}^{J/\psi}$	-4.80 ± 0.10
$Re(A_{\parallel}^{\psi(3770)})$	$(2.12 \pm 2.57)\times10^{-5}$	$\lvert A_{00}^{\psi(2S)}\rvert$	$(1.060 \pm 0.052)\times10^{-3}$
$Im(A_{\parallel}^{\psi(3770)})$	$(-13.777 \pm 24.890)\times10^{-6}$	$\theta_{00}^{\psi(2S)}$	-2.58 ± 0.32
$Re(A_{\perp}^{\psi(3770)})$	$(-68.5287 \pm 217.2680)\times10^{-7}$	$\lvert C_9\rvert$	3.64 ± 0.25
$Im(A_{\perp}^{\psi(3770)})$	$(-17.578 \pm 21.653)\times10^{-6}$	$\lvert C_{10}\rvert$	4.42 ± 0.20
$Re(A_{0}^{\psi(3770)})$	$(-81.584 \pm 34.212)\times10^{-6}$	$\mathcal{C}_{V1}^{S}/(\mathcal{C}_9 - \mathcal{C}_{10})$	0.8 ± 2.3
$Im(A_{0}^{\psi(3770)})$	$(6.032 \pm 4.048)\times10^{-5}$	$\mathcal{C}_{V1}^{S}/(\mathcal{C}_9 + \mathcal{C}_{10})$	4.7 ± 2.3
$Re(A_{\parallel}^{\psi(4040)})$	$(-34.833 \pm 33.120)\times10^{-6}$	$\mathcal{C}_{T}^{S}/\mathcal{C}_7$	-1.7 ± 1.2
$Im(A_{\parallel}^{\psi(4040)})$	$(3.127 \pm 41.185)\times10^{-6}$	$Re(\Delta C_7^{\text{AP}})$	-0.009 ± 0.038
$Re(A_{\perp}^{\psi(4040)})$	$(-17.250 \pm 26.129)\times10^{-6}$	$Im(\Delta C_7^{\text{AP}})$	-0.041 ± 0.020
$Im(A_{\perp}^{\psi(4040)})$	$(-32.273 \pm 25.469)\times10^{-6}$	$Re(\Delta C_7^{\text{AP}})$	-0.026 ± 0.039
$Re(A_{0}^{\psi(4040)})$	$(-70.6644 \pm 32.0690)\times10^{-6}$	$Im(\Delta C_7^{\text{AP}})$	-0.024 ± 0.023
$Im(A_{0}^{\psi(4040)})$	$(8.66 \pm 4.14)\times10^{-5}$	$Re(\Delta C_7^{\text{AP}})$	-0.102 ± 0.087
$Re(A_{\parallel}^{\psi(4160)})$	$(-17.200 \pm 39.992)\times10^{-6}$	$Im(\Delta C_7^{\text{AP}})$	-0.017 ± 0.024
$Im(A_{\parallel}^{\psi(4160)})$	$(7.594 \pm 43.102)\times10^{-6}$	$\alpha^2_{\text{resolution}}$	1.1656 ± 0.0076
$Re(A_{\perp}^{\psi(4160)})$	$(3.23 \pm 3.20)\times10^{-5}$	$\alpha^3_{\text{resolution}}$	1.380 ± 0.071

Table A.1: Fit results of all floating signal parameters from the fit to Run 1 data. The uncertainties are the fit uncertainties, which include statistical uncertainties and theoretical uncertainties from the P-Wave FF constraint and the ΔC_7 constraint.

A.3 Projections of the upper mass side-band fit in merged Run 1 and Run 2 data

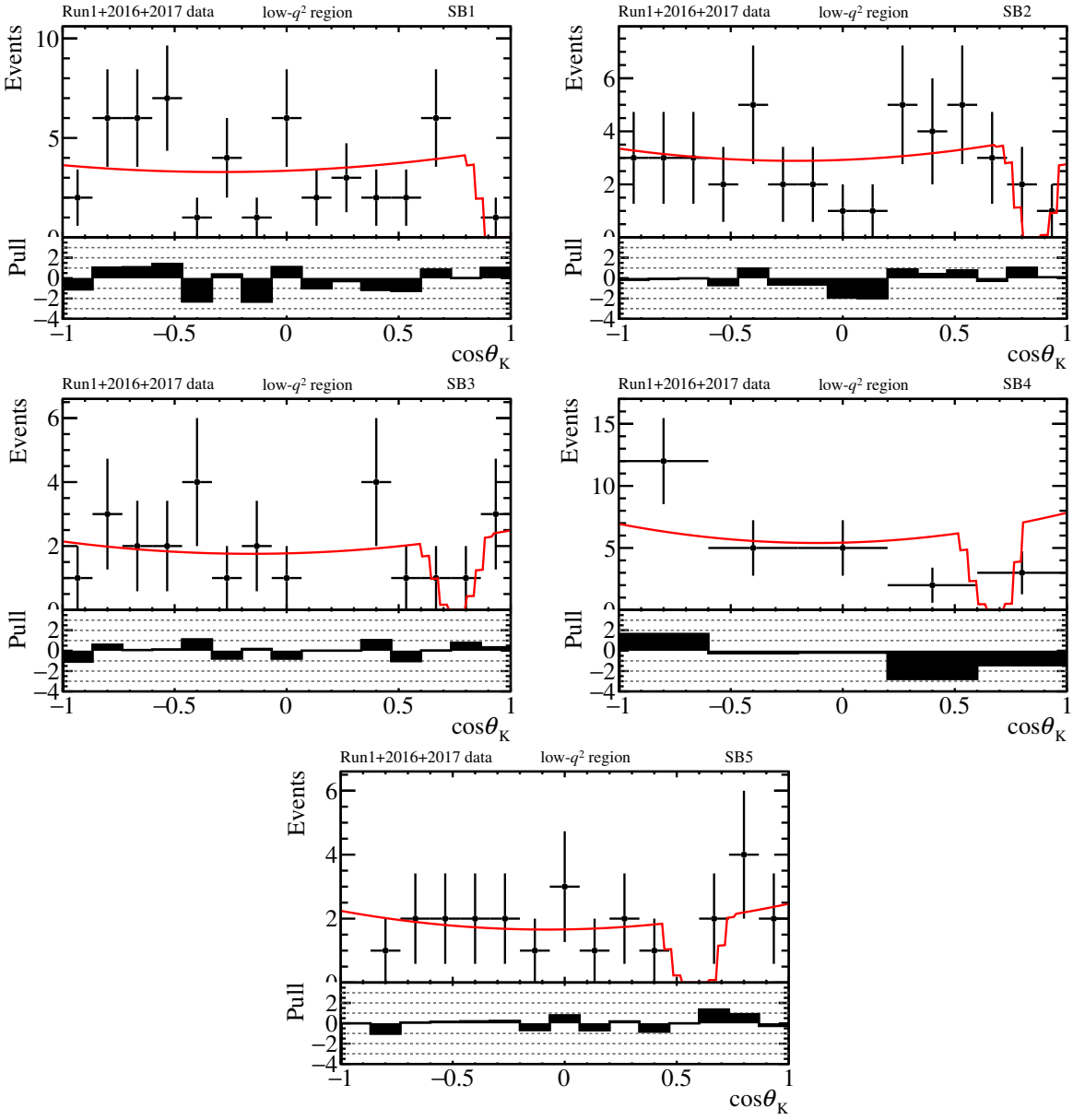


Figure A.7: $\cos\theta_K$ distributions in each mass side-band region in the low-q^2 region. Also shown are the projections of the simultaneous side-band fit described in chapter 6

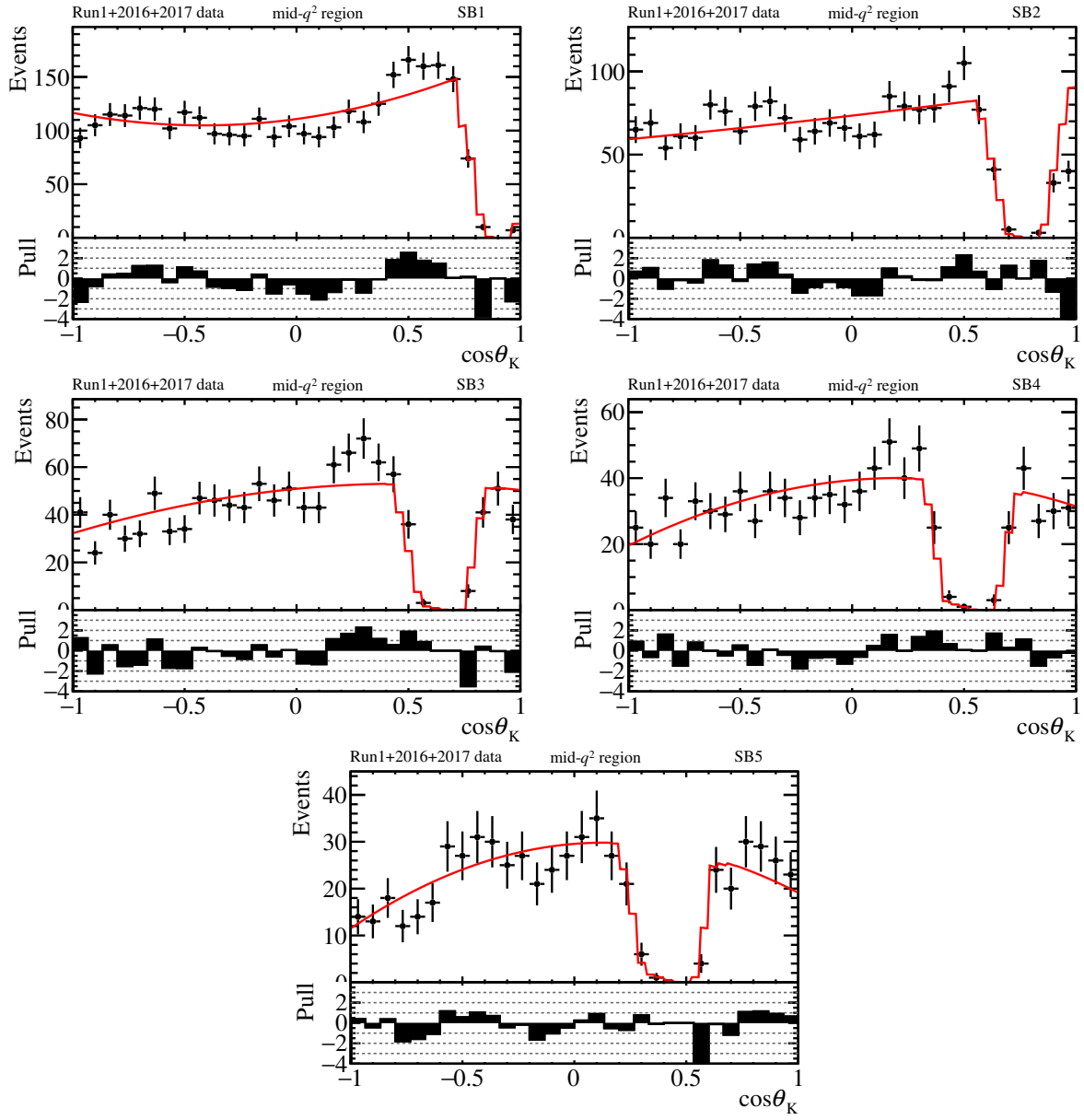


Figure A.8: $\cos\theta_K$ distributions in each mass side-band region in the mid-q^2 region. Also shown are the projections of the simultaneous side-band fit described in chapter 6

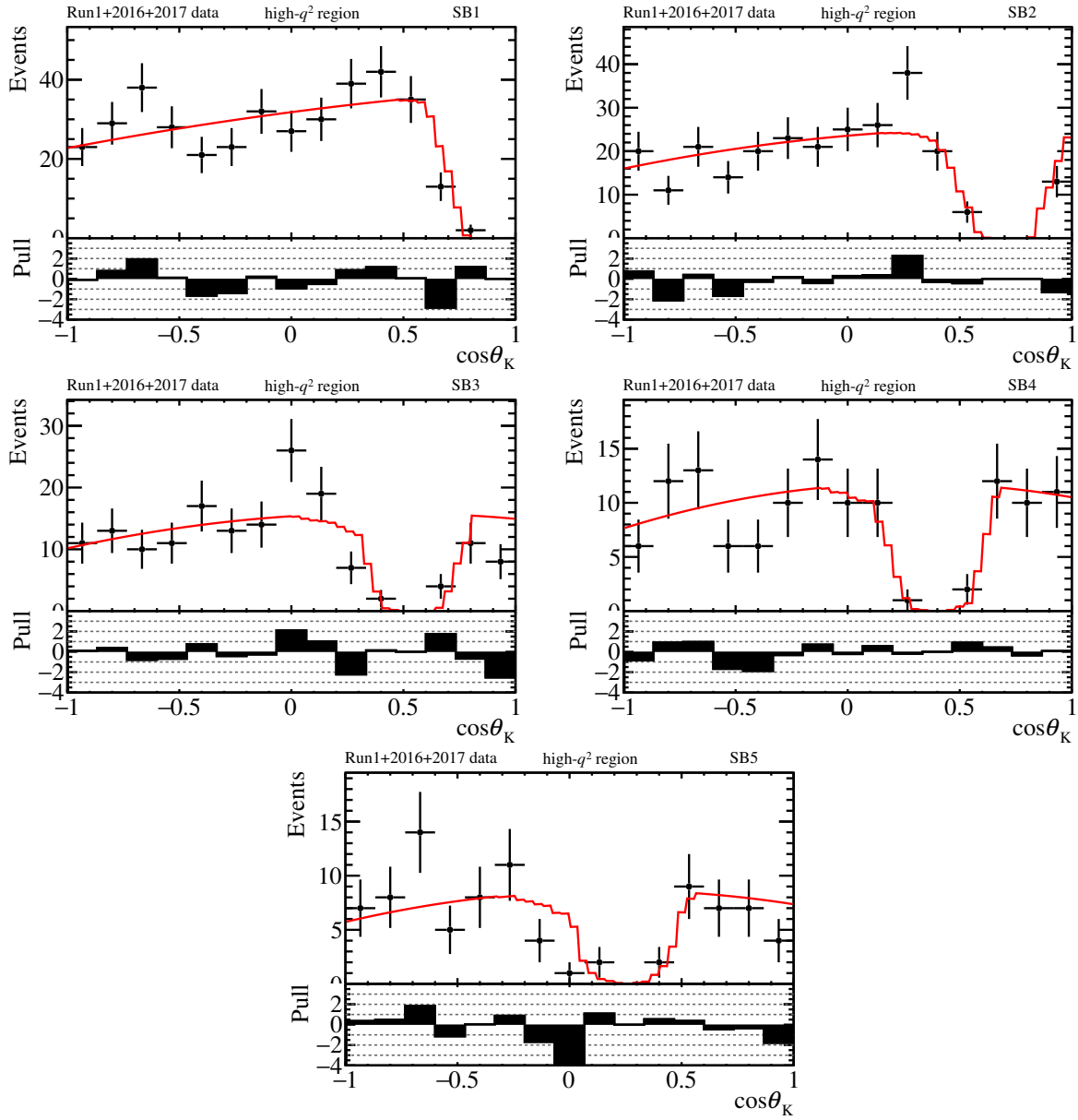


Figure A.9: $\cos\theta_K$ distributions in each mass side-band region in the high-q^2 region. Also shown are the projections of the simultaneous side-band fit described in chapter 6

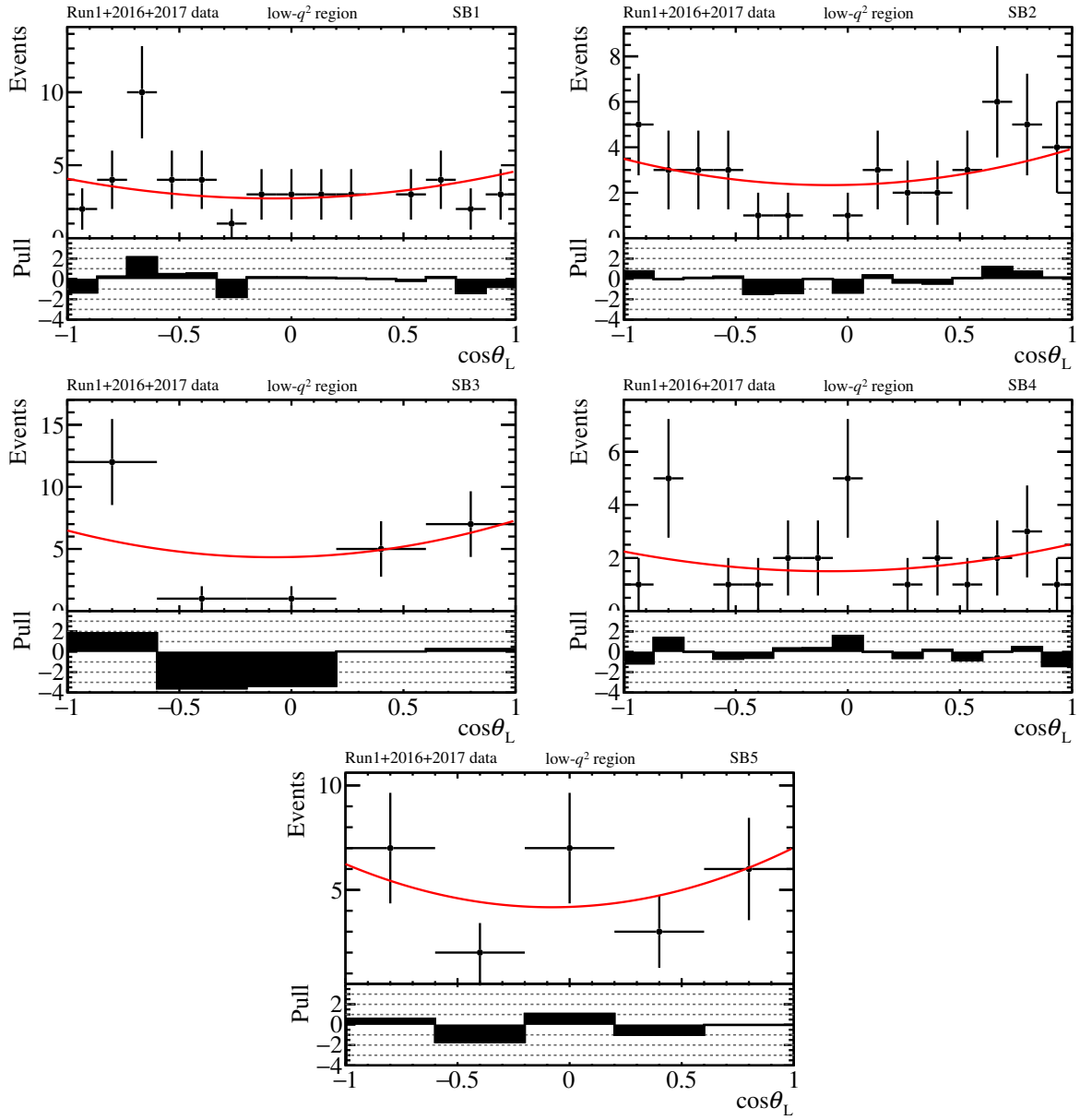


Figure A.10: $\cos\theta_\ell$ distributions in each mass side-band region in the low-q^2 region. Also shown are the projections of the simultaneous side-band fit described in chapter 6

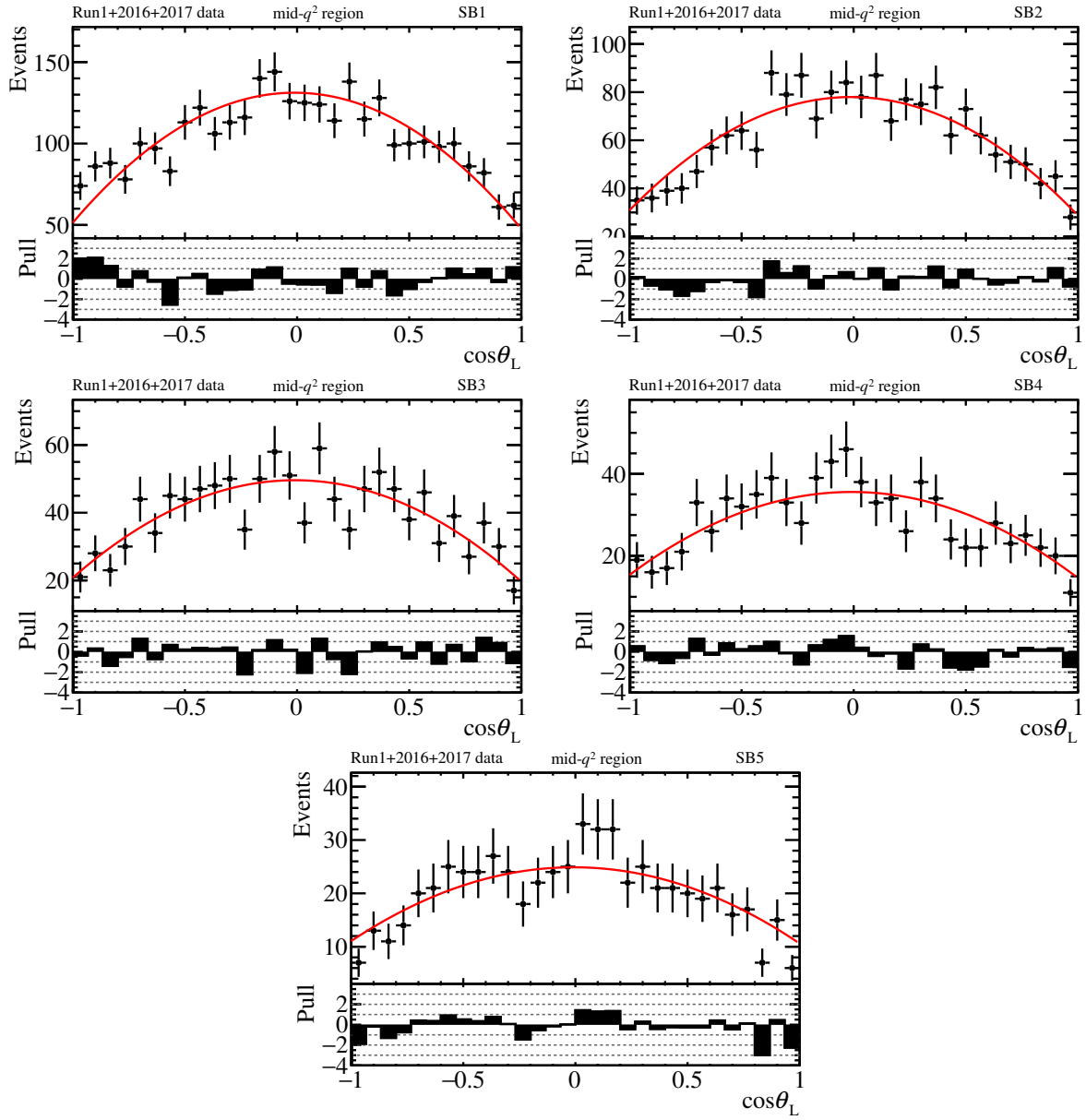


Figure A.11: $\cos\theta_\ell$ distributions in each mass side-band region in the mid-q^2 region. Also shown are the projections of the simultaneous side-band fit described in chapter 6

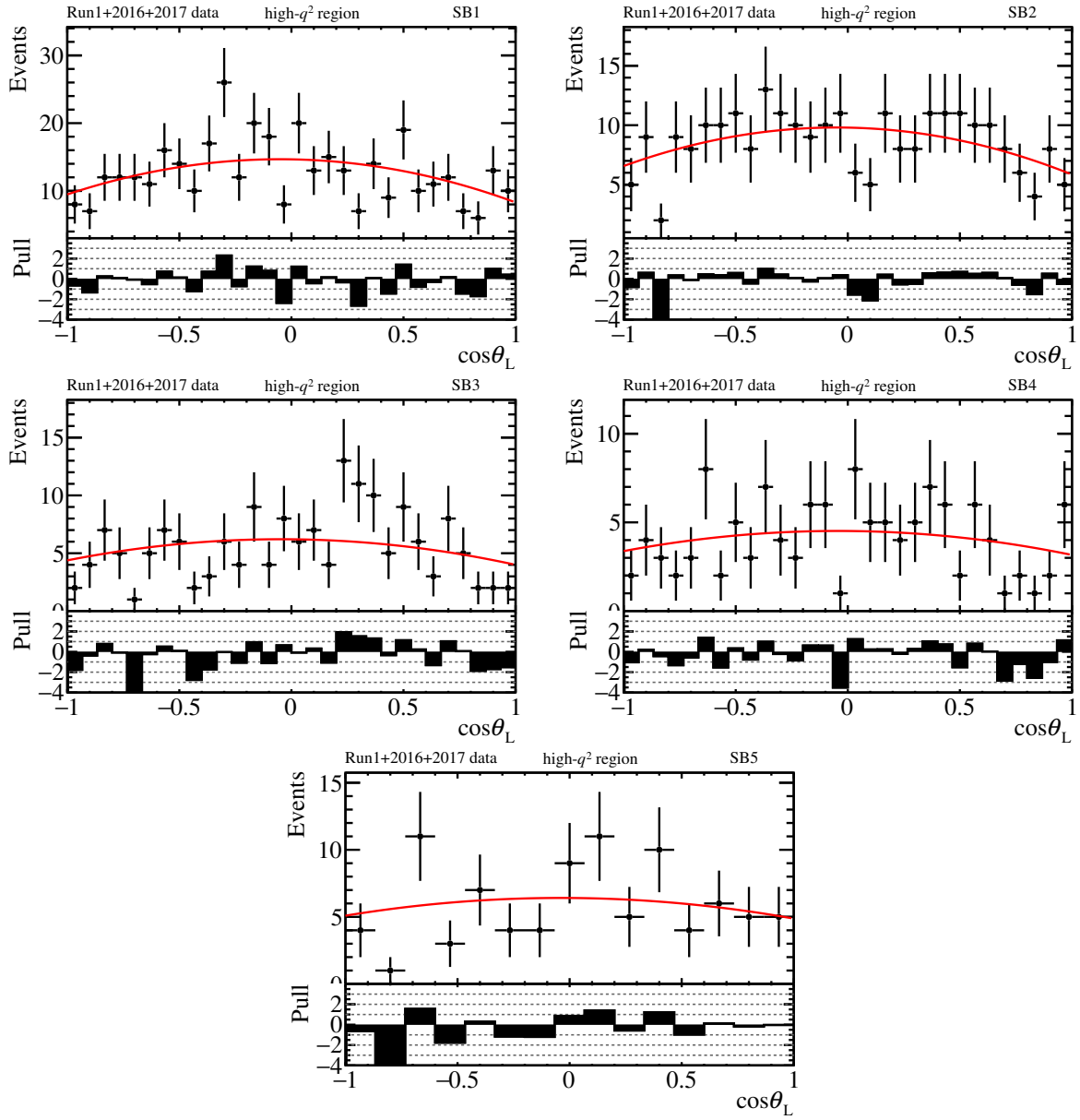


Figure A.12: $\cos\theta_\ell$ distributions in each mass side-band region in the high-q^2 region. Also shown are the projections of the simultaneous side-band fit described in chapter 6

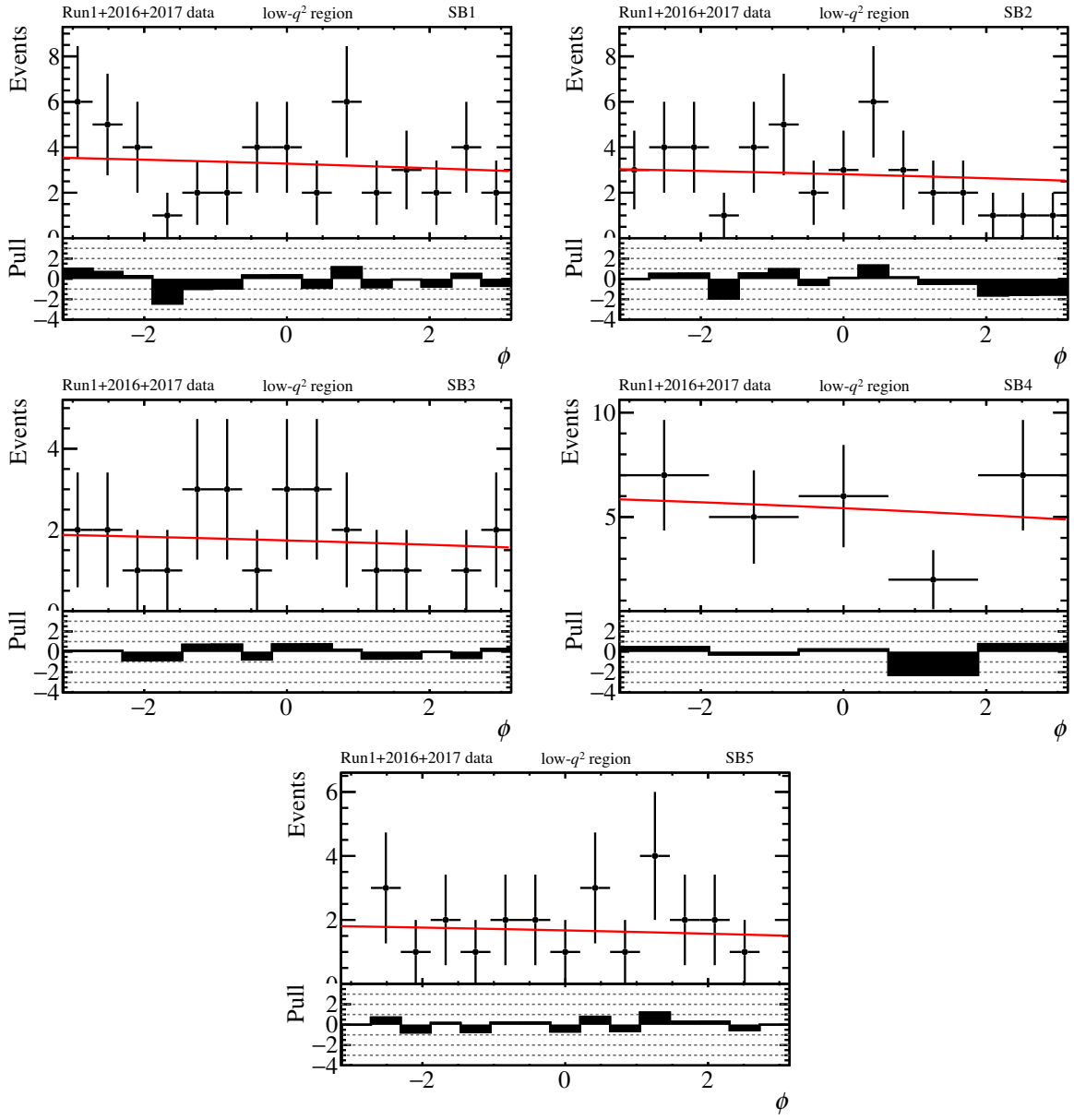


Figure A.13: ϕ distribution in each mass side-band region in the low-q^2 region. Also shown are the projections of the simultaneous side-band fit described in chapter 6

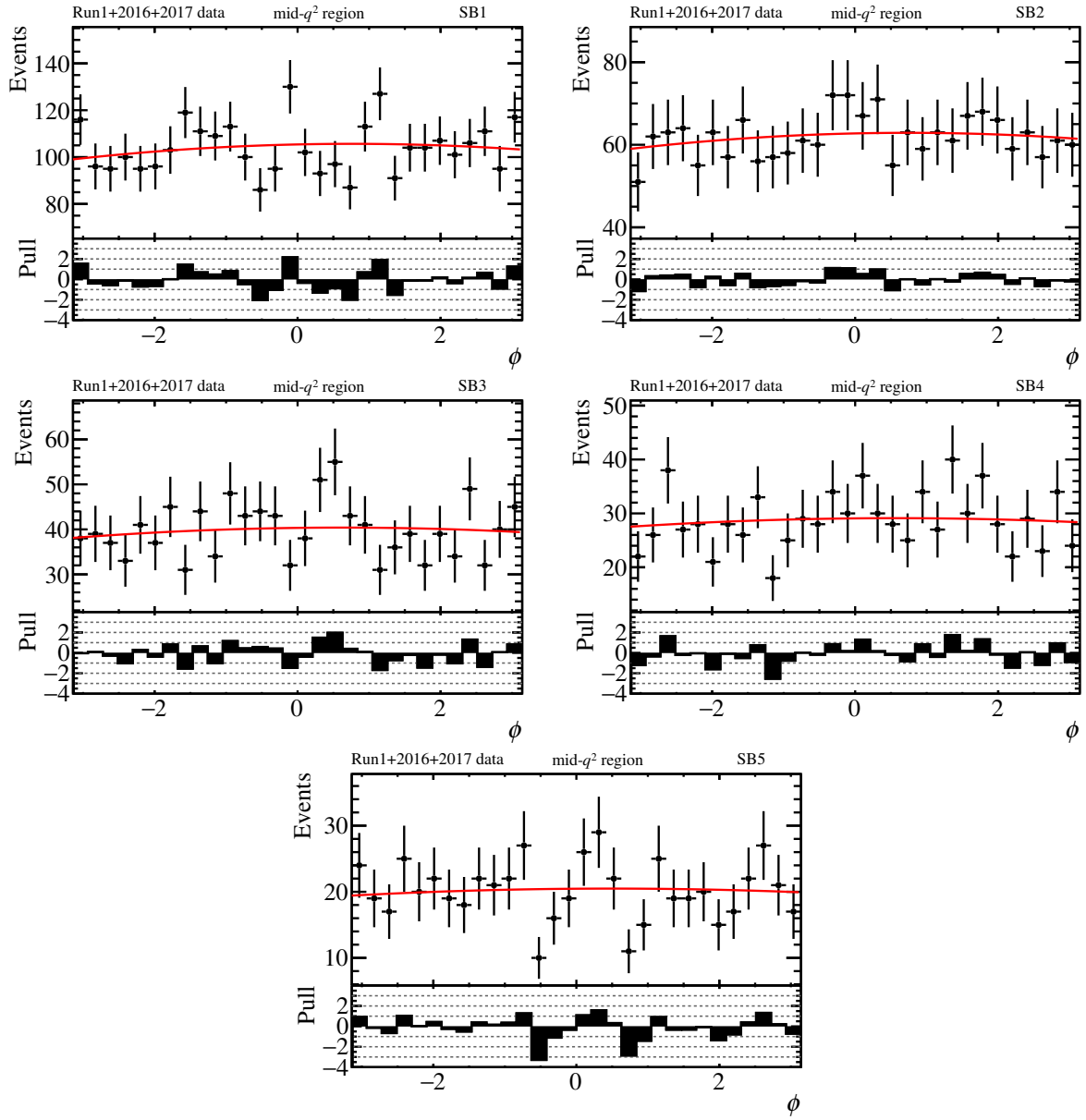


Figure A.14: ϕ distribution in each mass side-band region in the mid-q^2 region. Also shown are the projections of the simultaneous side-band fit described in chapter 6

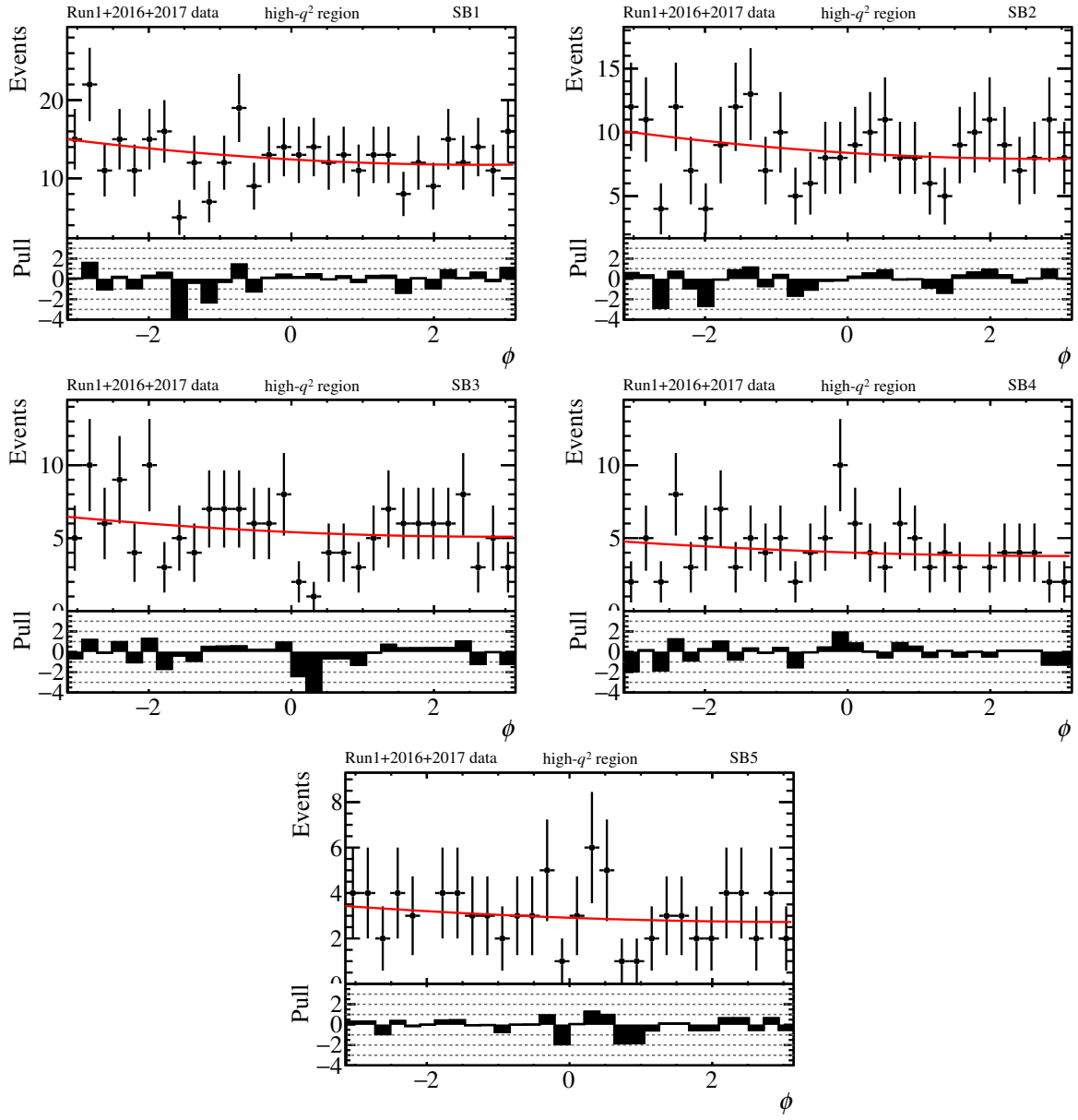


Figure A.15: ϕ distribution in each mass side-band region in the high-q^2 region. Also shown are the projections of the simultaneous side-band fit described in chapter 6

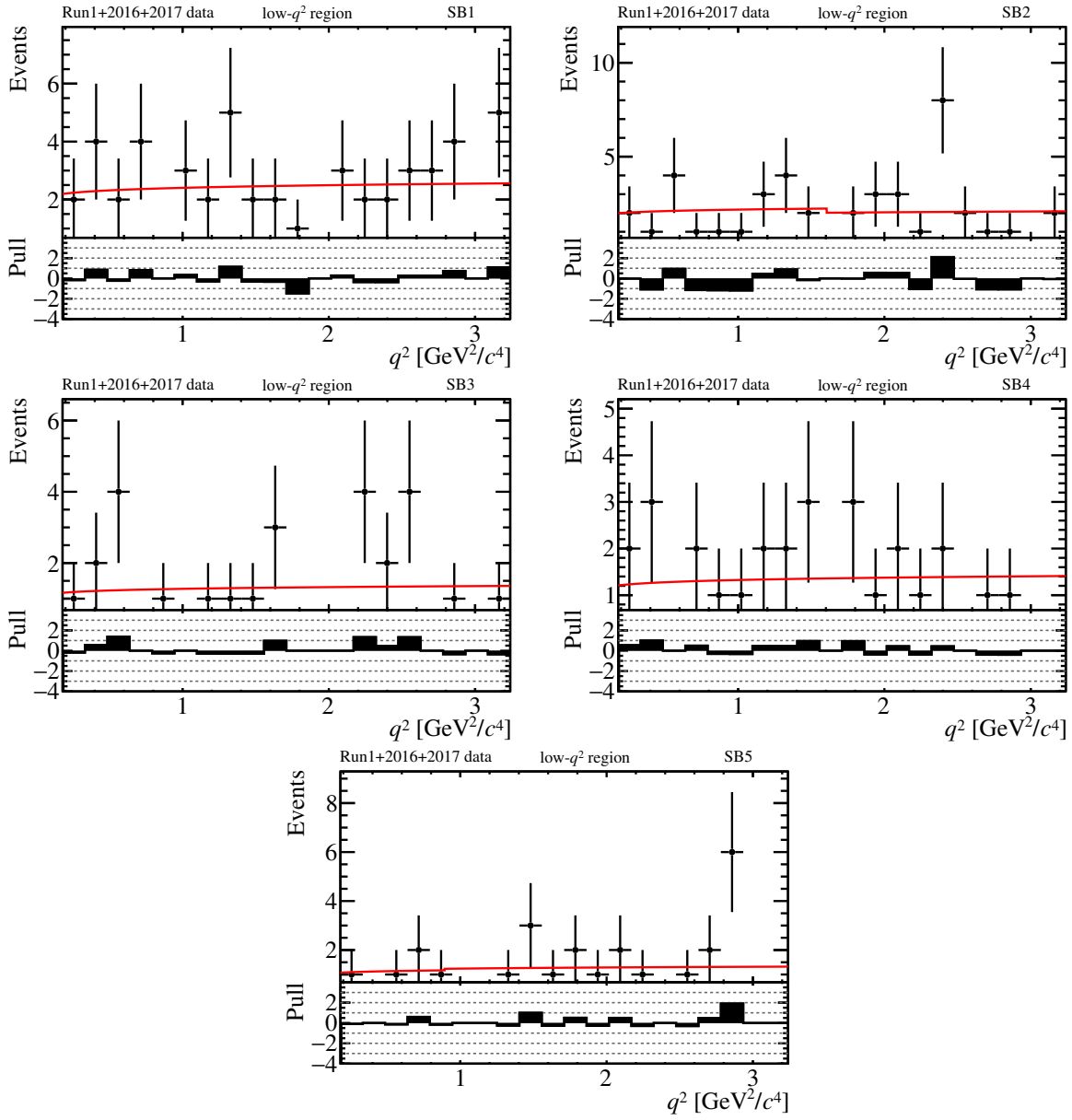


Figure A.16: q^2 distribution in each mass side-band region in the low-q^2 region. Also shown are the projections of the simultaneous side-band fit described in chapter 6

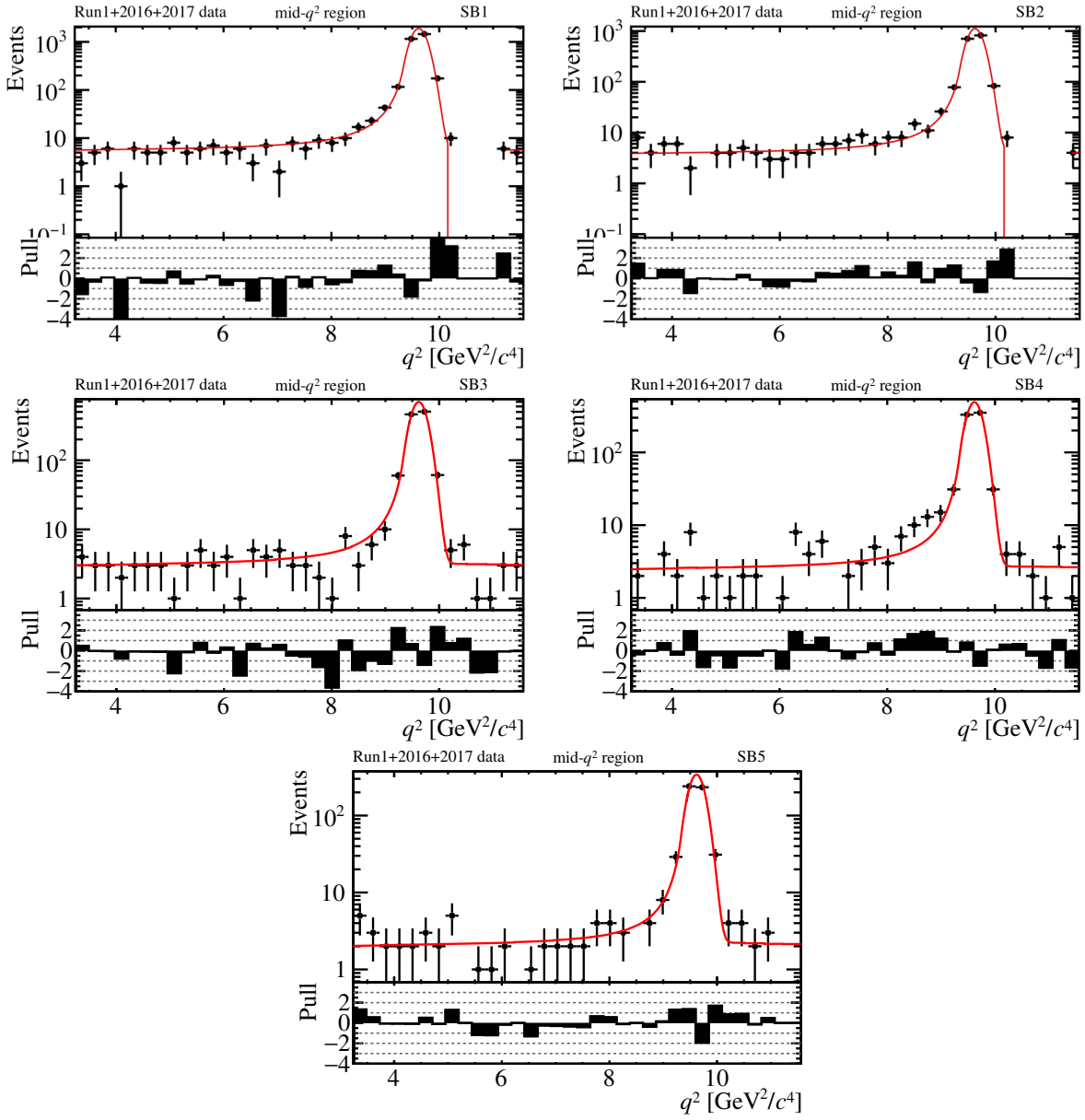


Figure A.17: q^2 distribution in each mass side-band region in the mid-q^2 region. Also shown are the projections of the simultaneous side-band fit described in chapter 6

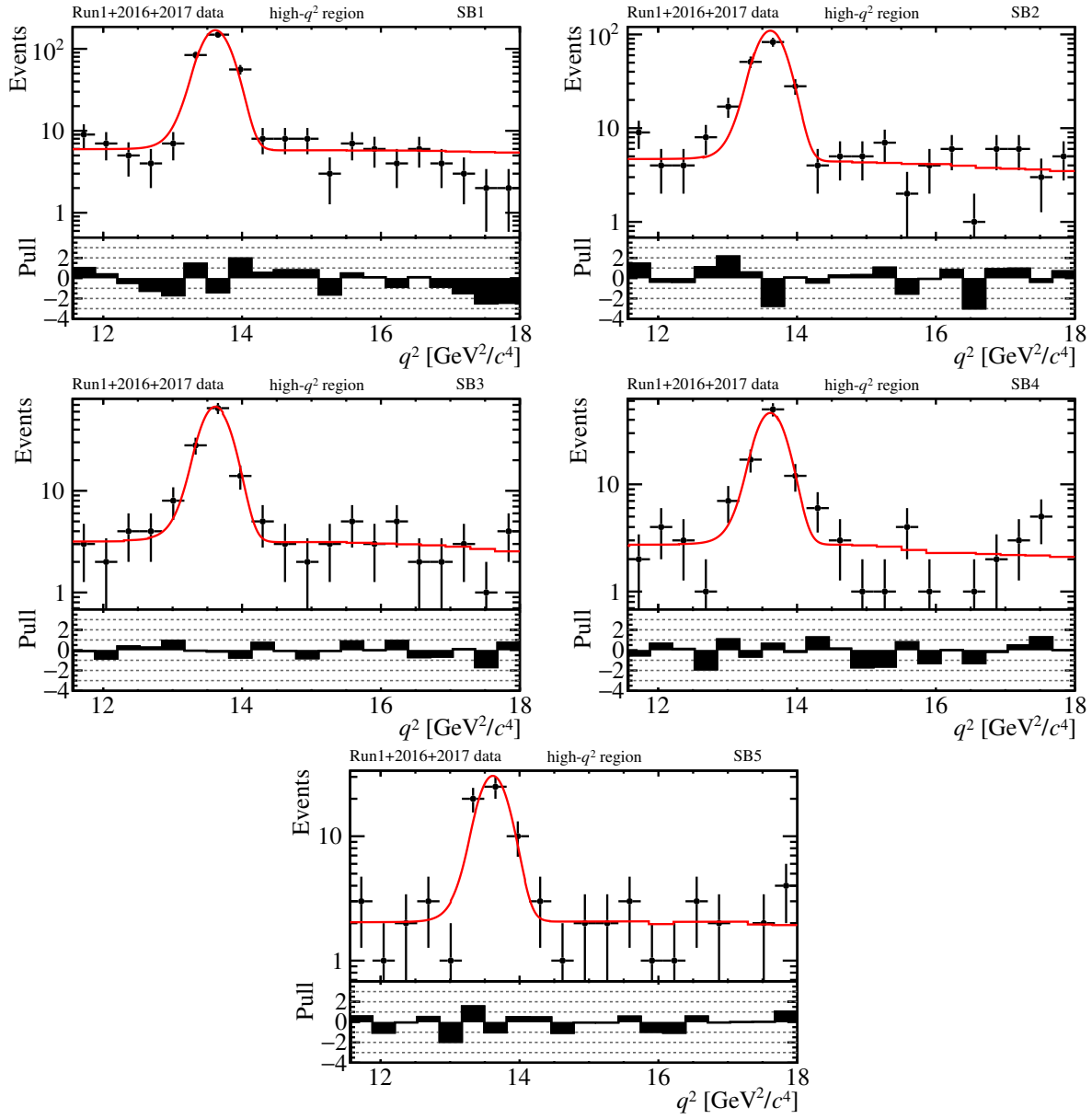


Figure A.18: q^2 distribution in each mass side-band region in the high-q^2 region. Also shown are the projections of the simultaneous side-band fit described in chapter 6

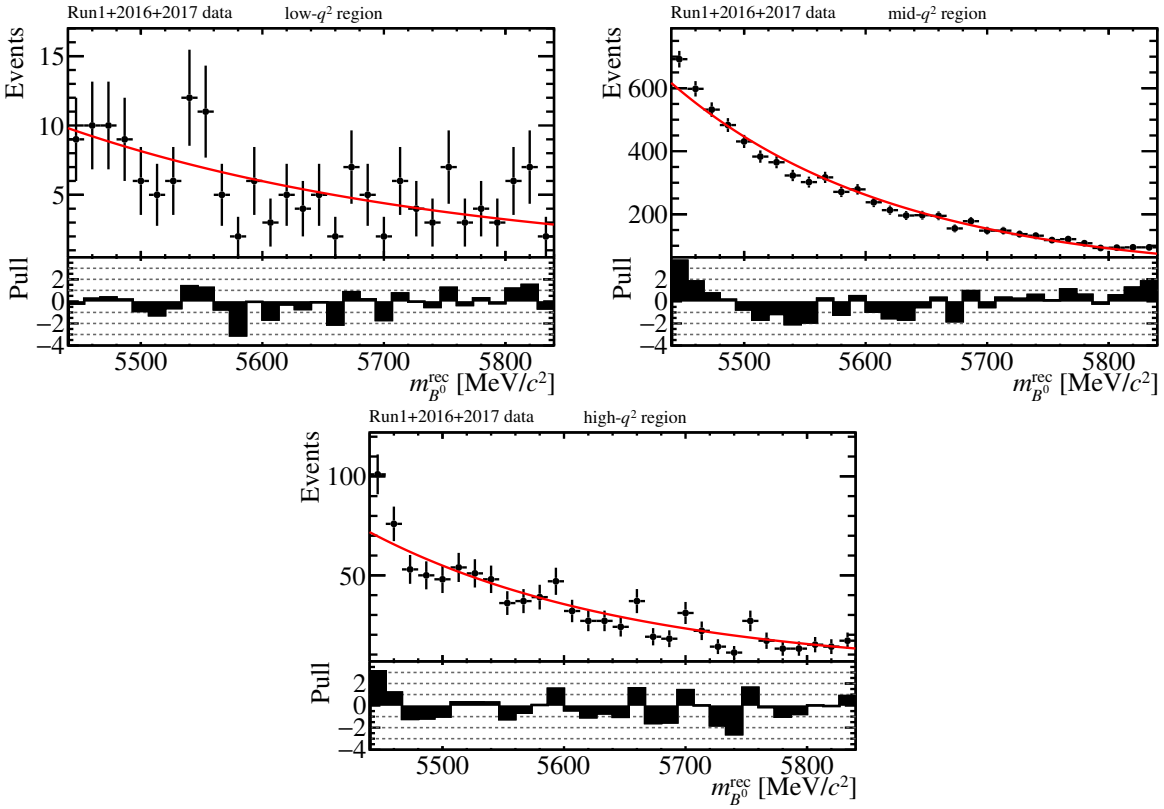


Figure A.19: $m^{\rm rec}_{B^0}$ distribution in each q^2 region. Also shown are the projections of the simultaneous side-band fit described in chapter 6

A.4 Projections of the simultaneous $m^{\mathrm{rec}}_{B^0}$ fit in Run 1 data

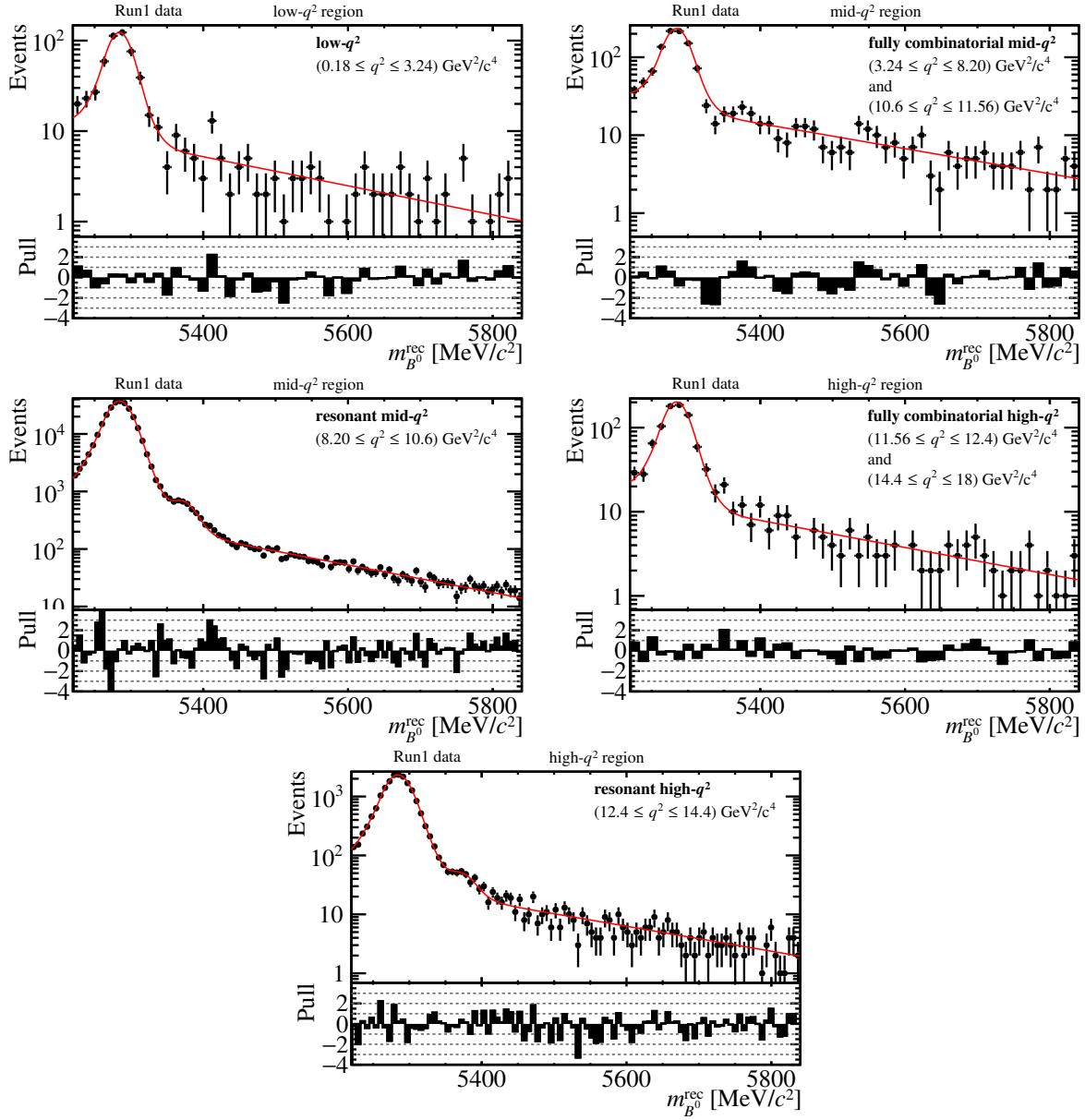


Figure A.20: $m_{B^0}^{\rm rec}$ distribution in the five q^2 regions of the simulatenous $m_{B^0}^{\rm rec}$ fit in Run 1 data. Also shown are the projections of the simultaneous $m_{B^0}^{\rm rec}$-only fit to those regions. The signal parameters are shared across all q^2 regions and the width parameters (σ and σ_2) are multiplied with a factor of 1.0385 in the high-q^2 regions to account for the q^2 dependence of the $m_{B^0}^{\rm rec}$ resolution. The slope of the fully combinatorial background is shared across low-, mid- and high-q^2, whereas the slopes in the resonant regions are independent.

A.5 Projections of the upper mass side-band fit in Run 1 data

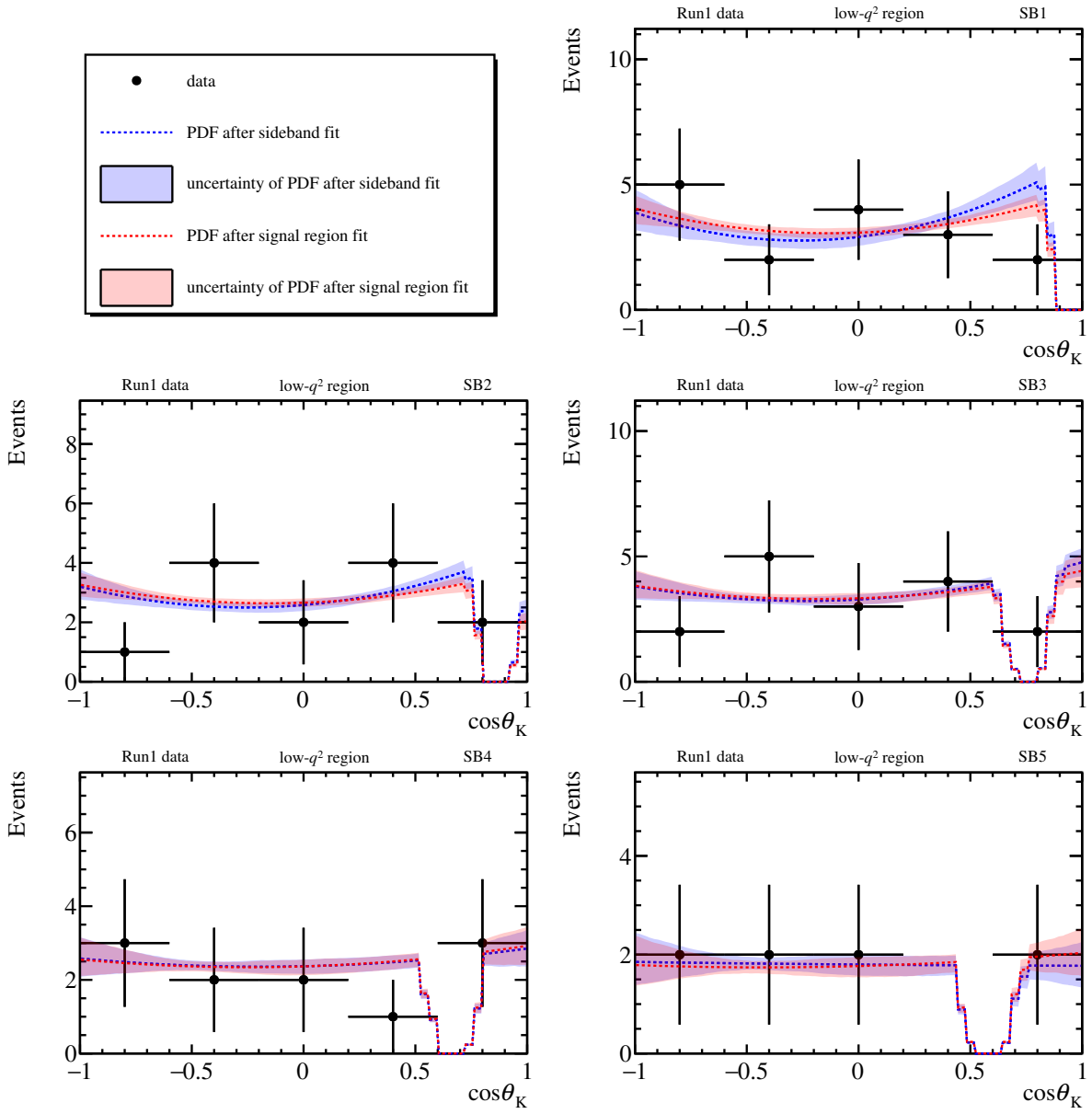


Figure A.21: $\cos\theta_K$ distribution in each mass side-band region in the low-q^2 region in Run 1 data. Also shown are the projections of the background PDF after the simultaneous side-band fit (blue dashed line) as well as the PDFs' 68% confidence interval (blue shaded area) obtained by varying the parameters using the covariance matrix obtained in the fit. Furthermore, the background PDF after the signal region fit (red dashed line) and the respective 68% confidence interval (red shaded area) are shown. Note that the background events in the signal region (not shown here) add additional constraints on the background parameters in the signal region fit.

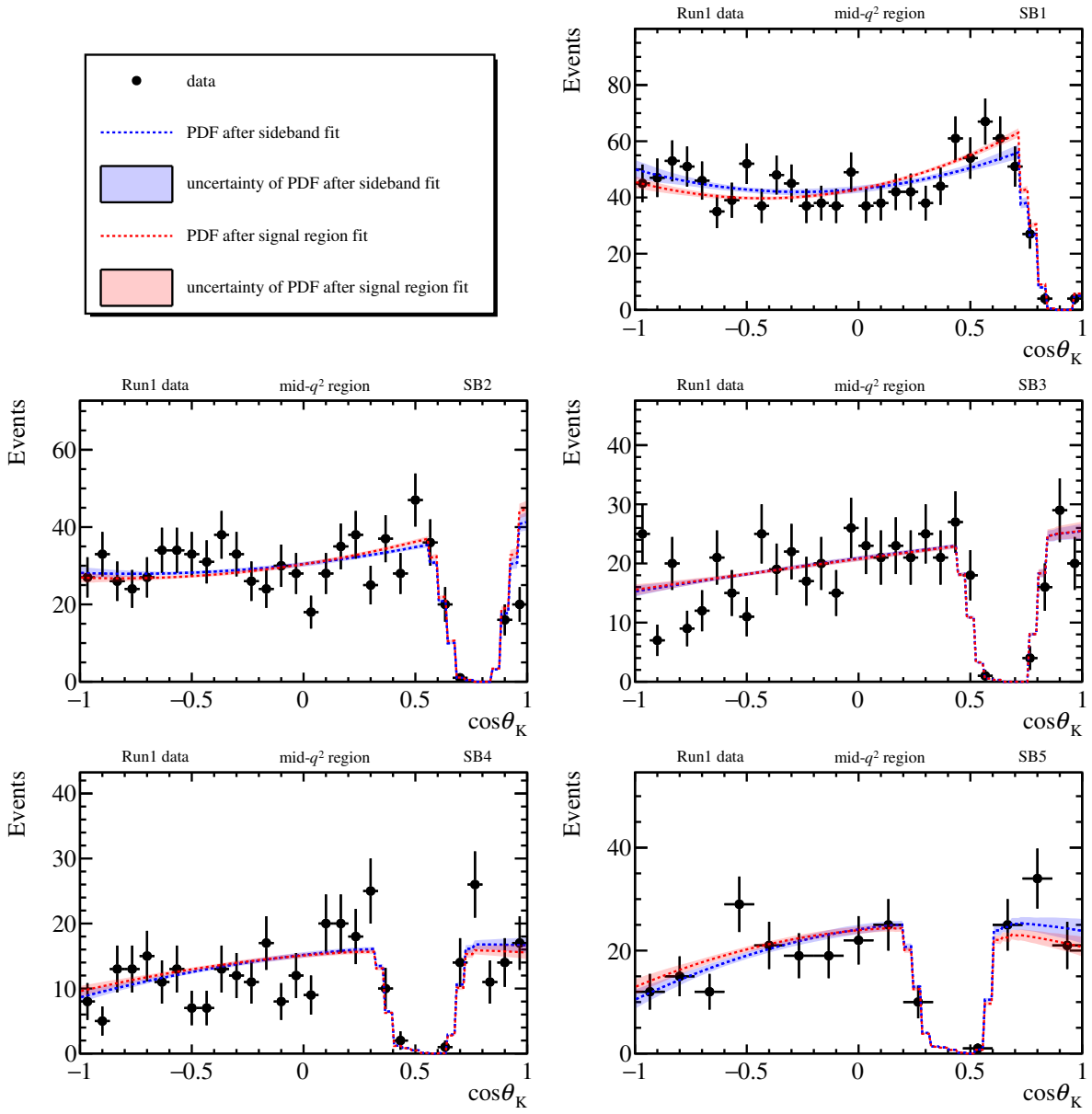


Figure A.22: $\cos\theta_K$ distribution in each mass side-band region in the mid-q^2 region in Run 1 data. Also shown are the projections of the background PDF after the simultaneous side-band fit (blue dashed line) as well as the PDFs' 68% confidence interval (blue shaded area) obtained by varying the parameters using the covariance matrix obtained in the fit. Furthermore, the background PDF after the signal region fit (red dashed line) and the respective 68% confidence interval (red shaded area) are shown. Note that the background events in the signal region (not shown here) add additional constraints on the background parameters in the signal region fit.

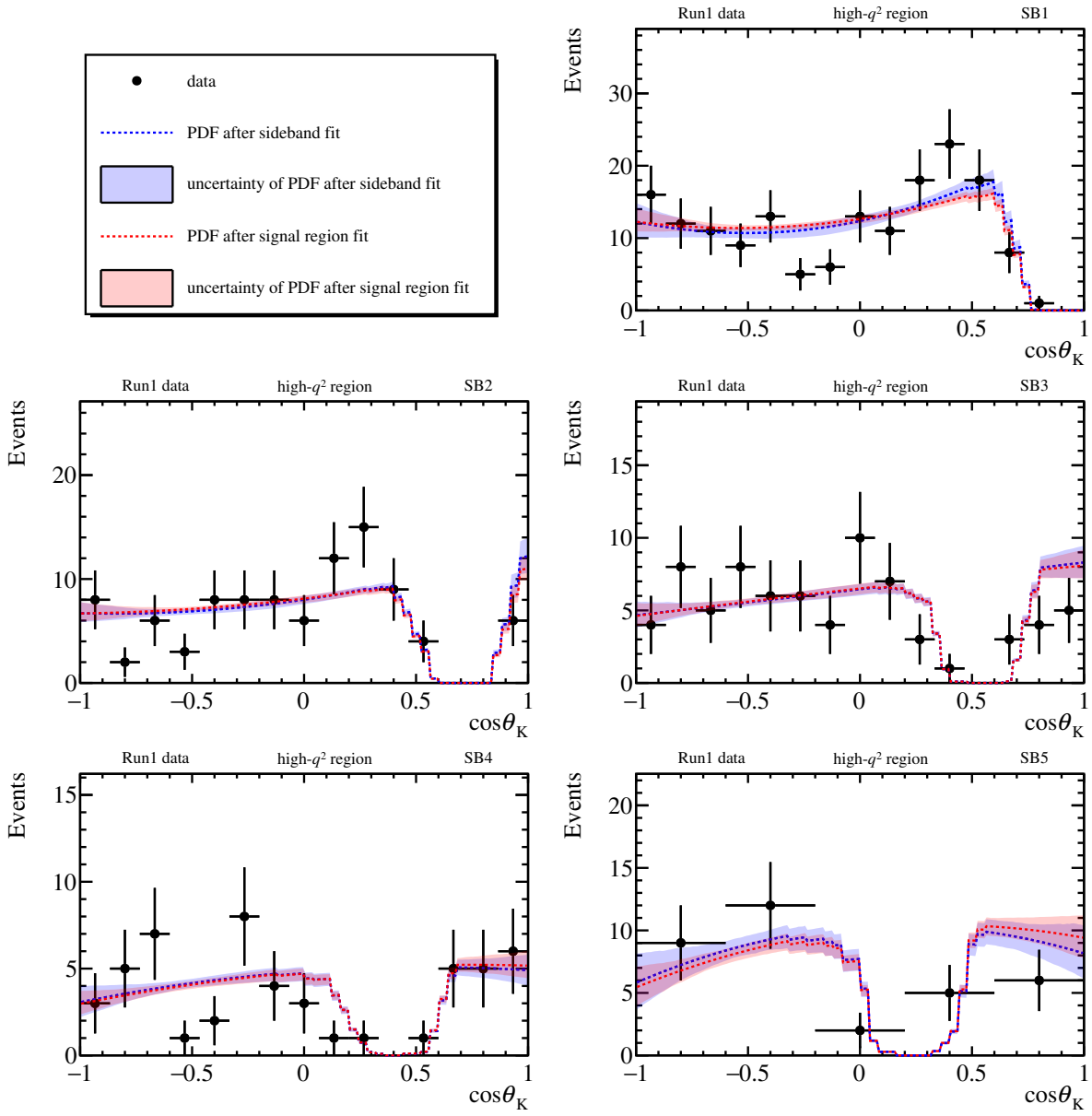


Figure A.23: $\cos\theta_K$ distribution in each mass side-band region in the high-q^2 region in Run 1 data. Also shown are the projections of the background PDF after the simultaneous side-band fit (blue dashed line) as well as the PDFs' 68% confidence interval (blue shaded area), obtained by varying the parameters using the covariance matrix obtained in the fit. Furthermore, the background PDF after the signal region fit (red dashed line) and the respective 68% confidence interval (red shaded area) are shown. Note that the background events in the signal region (not shown here) add additional constraints on the background parameters in the signal region fit.

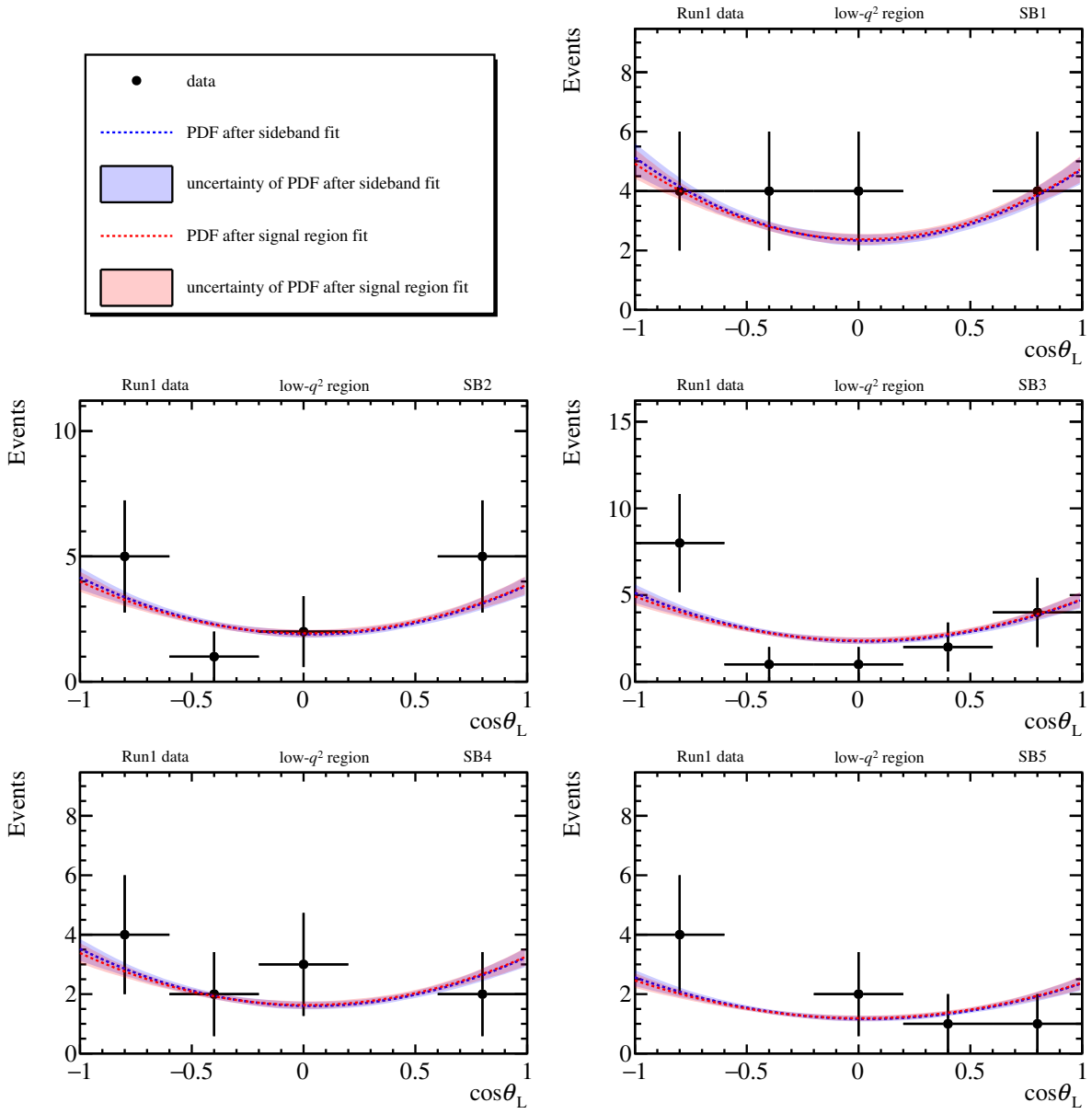


Figure A.24: $\cos\theta_\ell$ distribution in each mass side-band region in the low-q^2 region in Run 1 data. Also shown are the projections of the background PDF after the simultaneous side-band fit (blue dashed line) as well as the PDFs' 68% confidence interval (blue shaded area) obtained by varying the parameters using the covariance matrix obtained in the fit. Furthermore, the background PDF after the signal region fit (red dashed line) and the respective 68% confidence interval (red shaded area) are shown. Note that the background events in the signal region (not shown here) add additional constraints on the background parameters in the signal region fit.

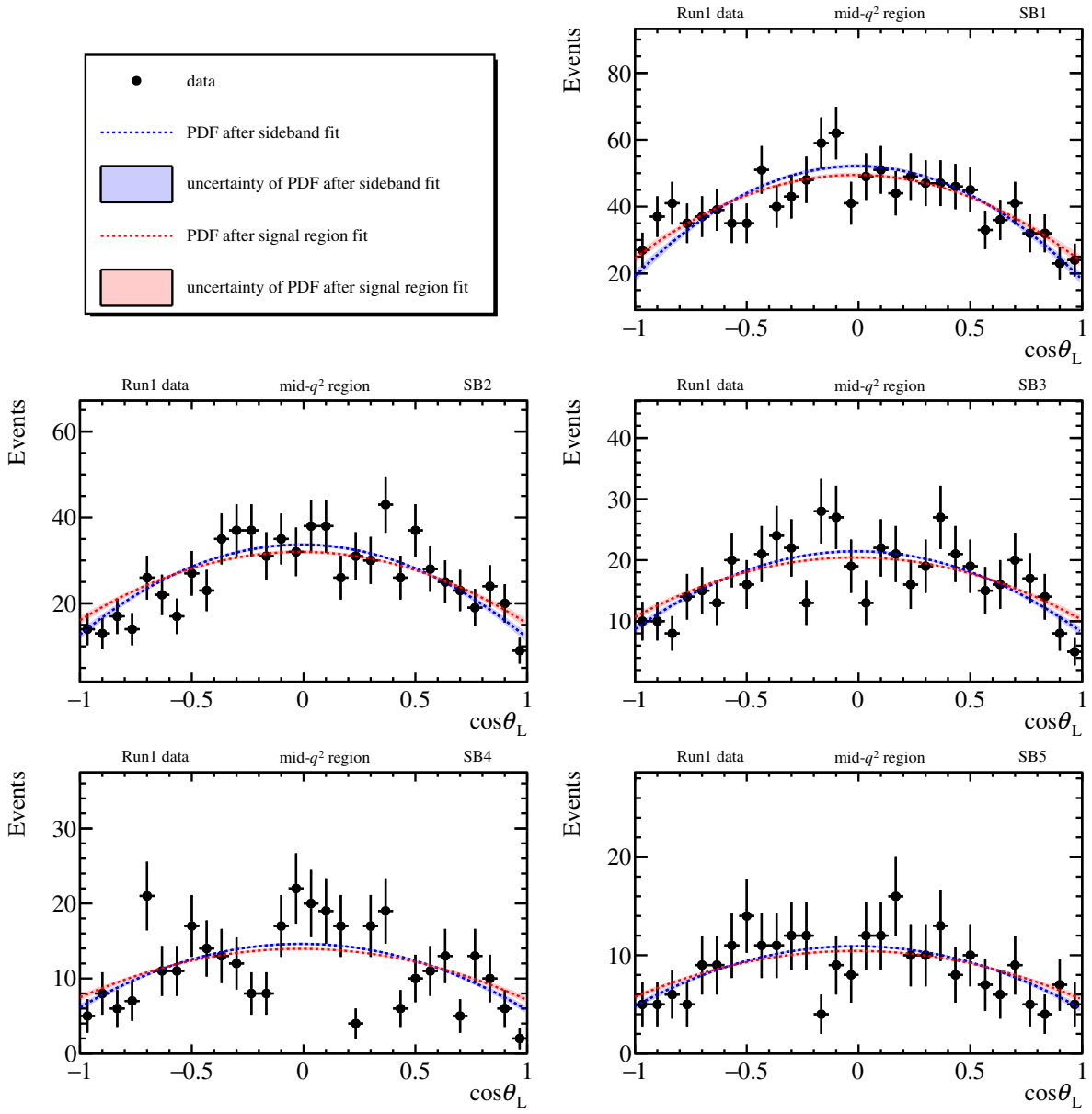


Figure A.25: $\cos\theta_\ell$ distribution in each mass side-band region in the mid-q^2 region in Run 1 data. Also shown are the projections of the background PDF after the simultaneous side-band fit (blue dashed line) as well as the PDFs' 68% confidence interval (blue shaded area) obtained by varying the parameters using the covariance matrix obtained in the fit. Furthermore, the background PDF after the signal region fit (red dashed line) and the respective 68% confidence interval (red shaded area) are shown. Note that the background events in the signal region (not shown here) add additional constraints on the background parameters in the signal region fit.

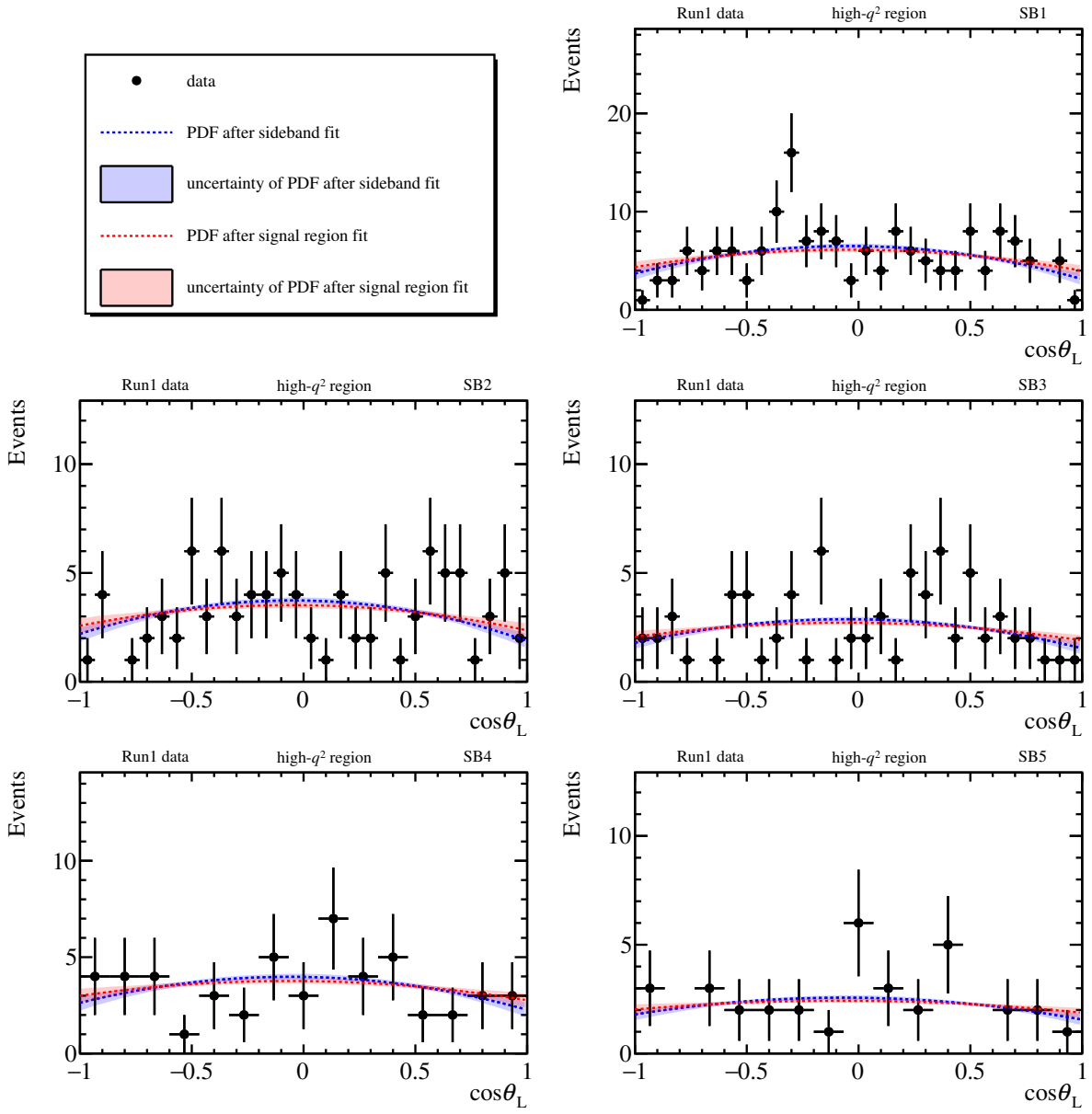


Figure A.26: $\cos\theta_\ell$ distribution in each mass side-band region in the high-q^2 region in Run 1 data. Also shown are the projections of the background PDF after the simultaneous side-band fit (blue dashed line) as well as the PDFs' 68% confidence interval (blue shaded area) obtained by varying the parameters using the covariance matrix obtained in the fit. Furthermore, the background PDF after the signal region fit (red dashed line) and the respective 68% confidence interval (red shaded area) are shown. Note that the background events in the signal region (not shown here) add additional constraints on the background parameters in the signal region fit.

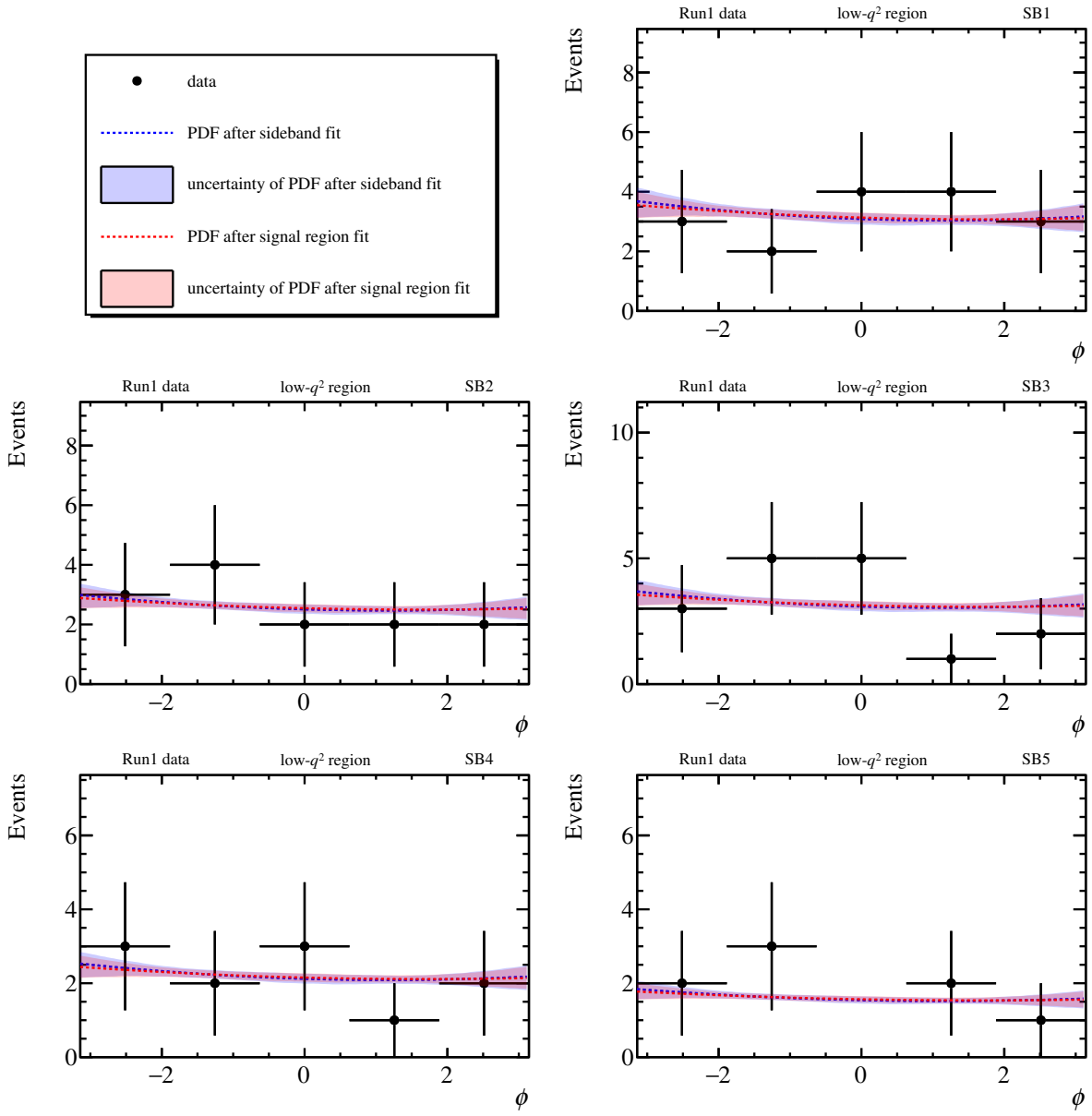


Figure A.27: ϕ distribution in each mass side-band region in the low-q^2 region in Run 1 data. Also shown are the projections of the background PDF after the simultaneous side-band fit (blue dashed line) as well as the PDFs' 68% confidence interval (blue shaded area) obtained by varying the parameters using the covariance matrix obtained in the fit. Furthermore, the background PDF after the signal region fit (red dashed line) and the respective 68% confidence interval (red shaded area) are shown. Note that the background events in the signal region (not shown here) add additional constraints on the background parameters in the signal region fit.

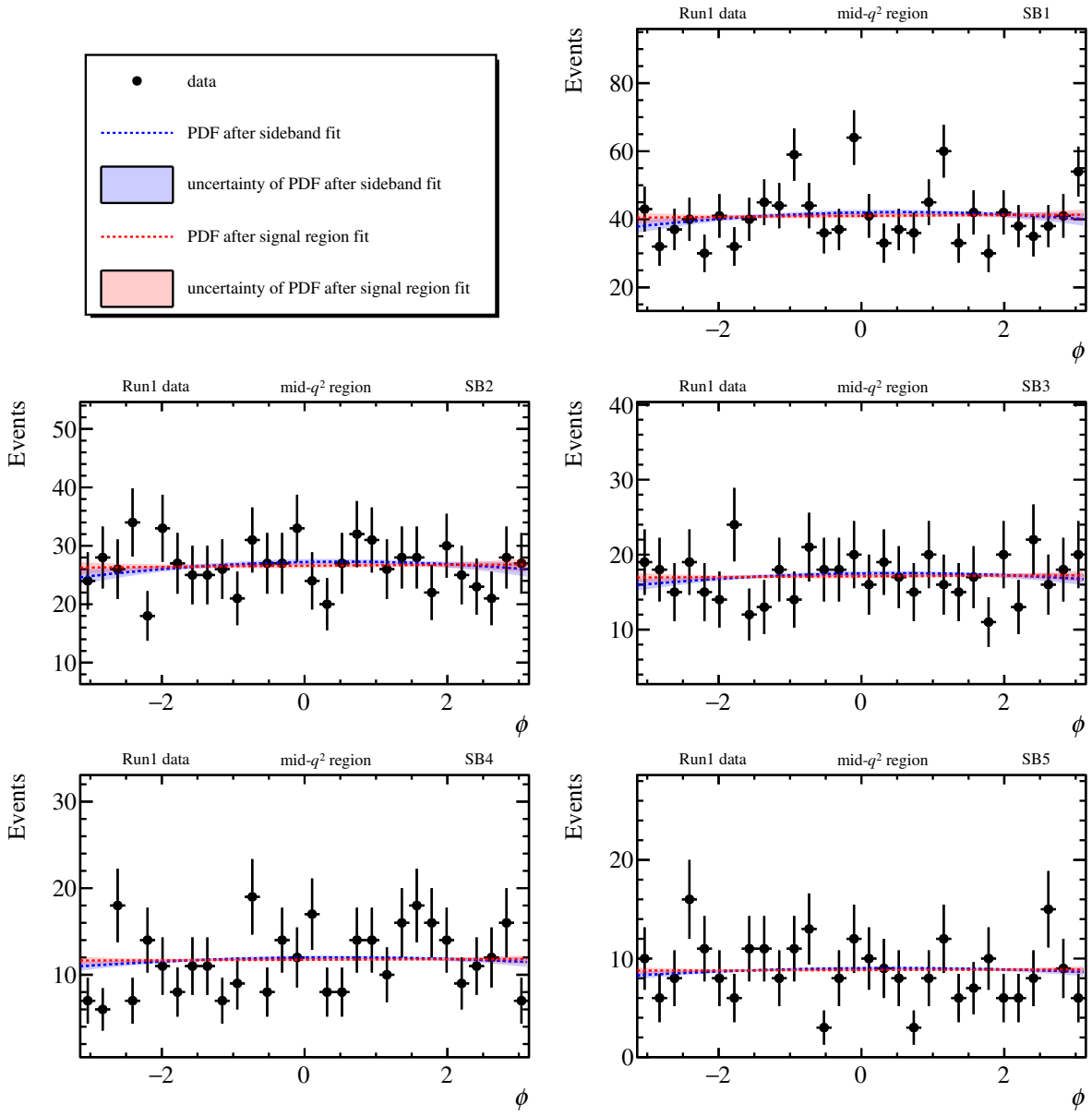


Figure A.28: ϕ distribution in each mass side-band region in the mid-q^2 region in Run 1 data. Also shown are the projections of the background PDF after the simultaneous side-band fit (blue dashed line) as well as the PDFs' 68% confidence interval (blue shaded area) obtained by varying the parameters using the covariance matrix obtained in the fit. Furthermore, the background PDF after the signal region fit (red dashed line) and the respective 68% confidence interval (red shaded area) are shown. Note that the background events in the signal region (not shown here) add additional constraints on the background parameters in the signal region fit.

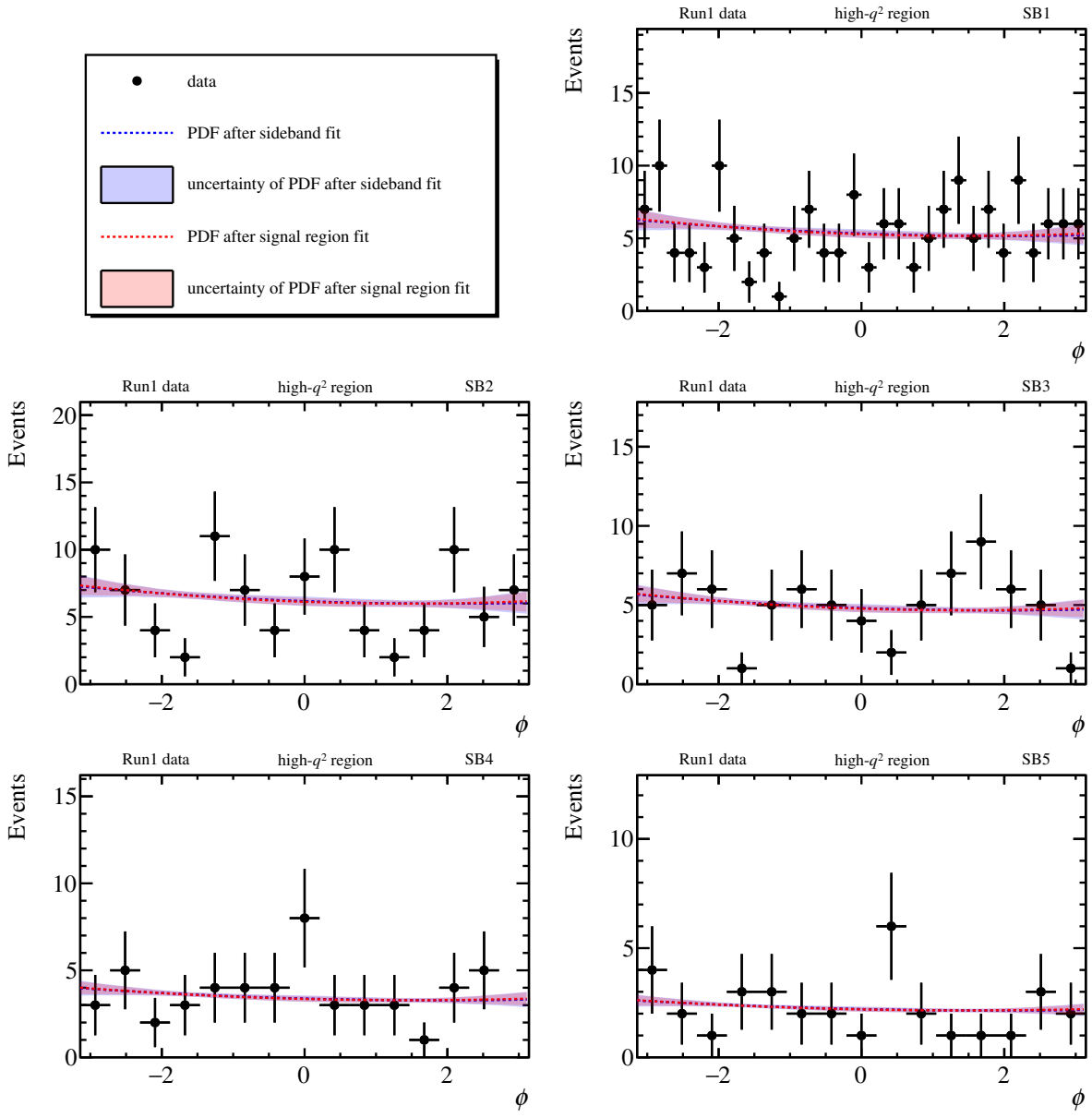


Figure A.29: ϕ distribution in each mass side-band region in the high-q^2 region in Run 1 data. Also shown are the projections of the background PDF after the simultaneous side-band fit (blue dashed line) as well as the PDFs' 68% confidence interval (blue shaded area) obtained by varying the parameters using the covariance matrix obtained in the fit. Furthermore, the background PDF after the signal region fit (red dashed line) and the respective 68% confidence interval (red shaded area) are shown. Note that the background events in the signal region (not shown here) add additional constraints on the background parameters in the signal region fit.

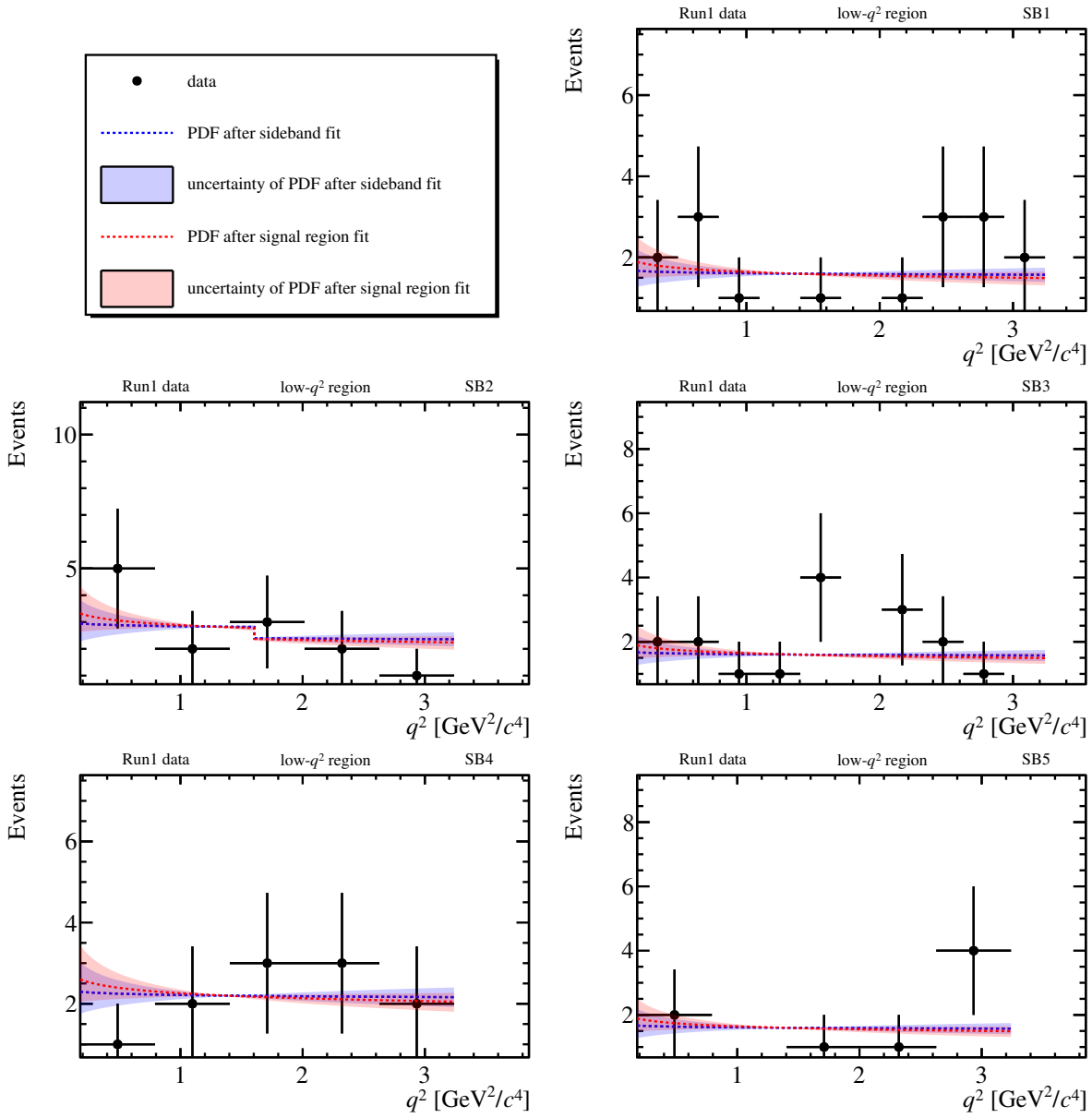


Figure A.30: q^2 distribution in each mass side-band region in the low-q^2 region in Run 1 data. Also shown are the projections of the background PDF after the simultaneous side-band fit (blue dashed line) as well as the PDFs' 68% confidence interval (blue shaded area) obtained by varying the parameters using the covariance matrix obtained in the fit. Furthermore, the background PDF after the signal region fit (red dashed line) and the respective 68% confidence interval (red shaded area) are shown. Note that the background events in the signal region (not shown here) add additional constraints on the background parameters in the signal region fit.

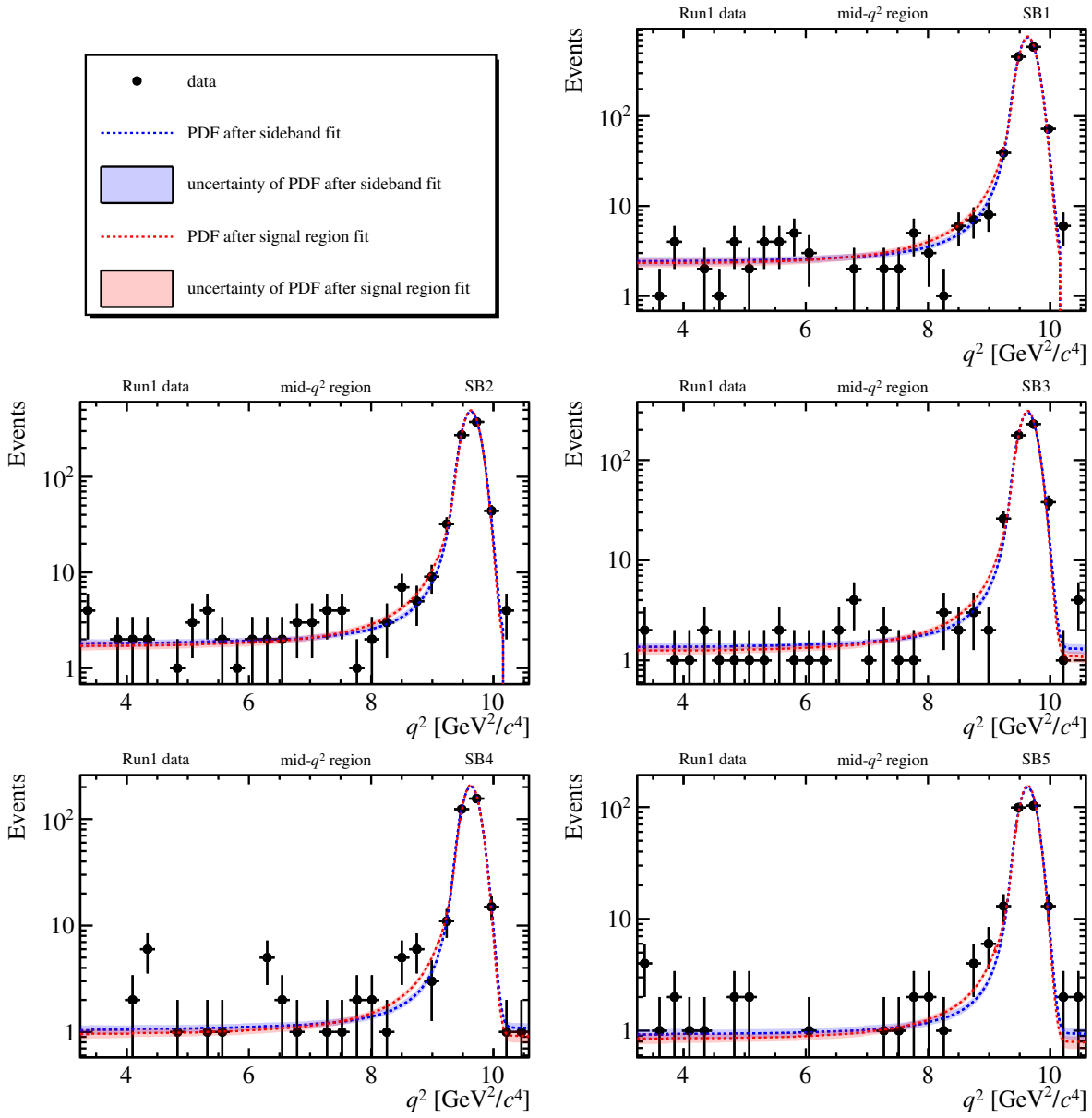


Figure A.31: q^2 distribution in each mass side-band region in the mid-q^2 region in Run 1 data. Also shown are the projections of the background PDF after the simultaneous side-band fit (blue dashed line) as well as the PDFs' 68% confidence interval (blue shaded area) obtained by varying the parameters using the covariance matrix obtained in the fit. Furthermore, the background PDF after the signal region fit (red dashed line) and the respective 68% confidence interval (red shaded area) are shown. Note that the background events in the signal region (not shown here) add additional constraints on the background parameters in the signal region fit.

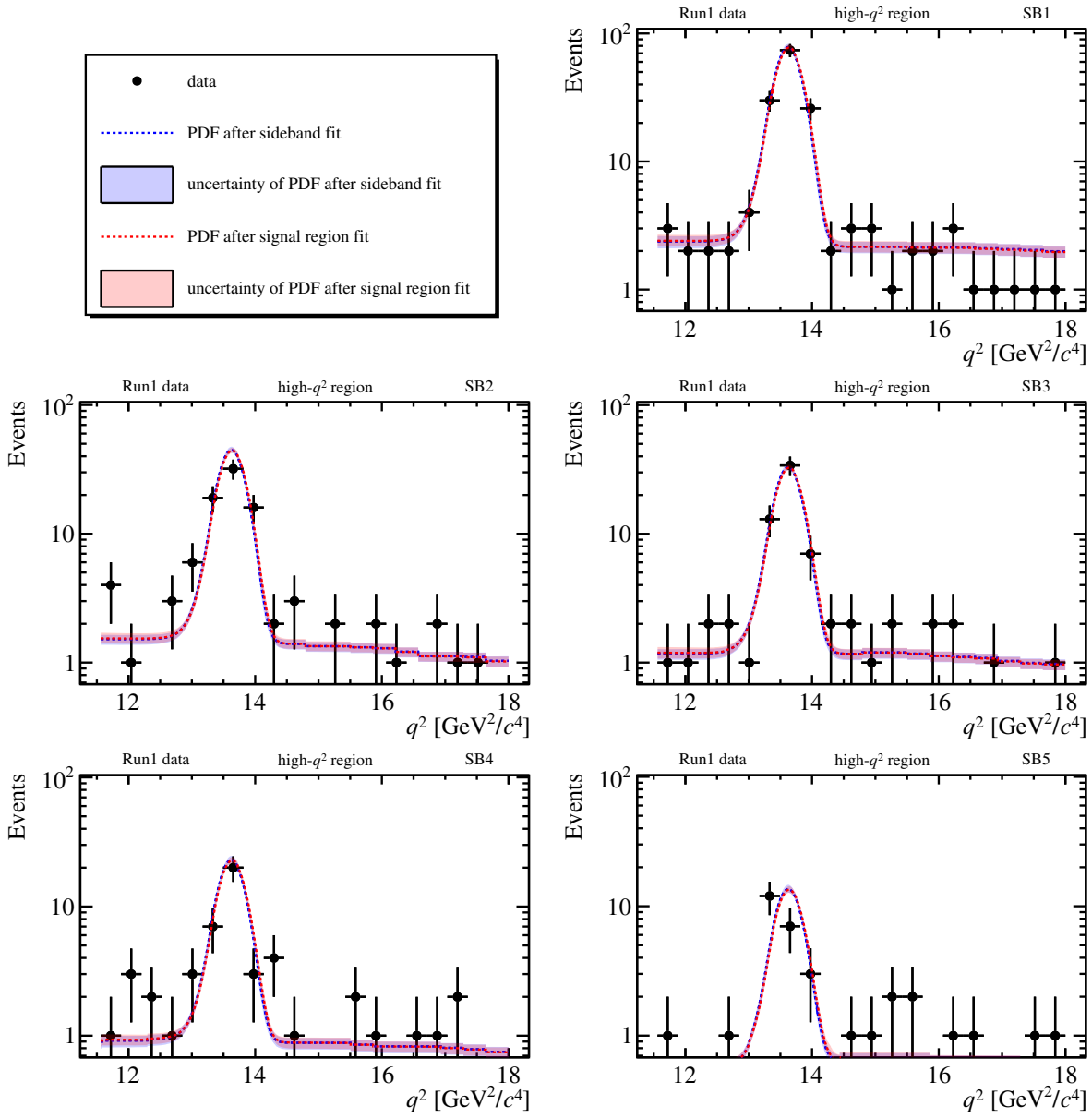


Figure A.32: q^2 distribution in each mass side-band region in the high-q^2 region in Run 1 data. Also shown are the projections of the background PDF after the simultaneous side-band fit (blue dashed line) as well as the PDFs' 68% confidence interval (blue shaded area) obtained by varying the parameters using the covariance matrix obtained in the fit. Furthermore, the background PDF after the signal region fit (red dashed line) and the respective 68% confidence interval (red shaded area) are shown. Note that the background events in the signal region (not shown here) add additional constraints on the background parameters in the signal region fit.

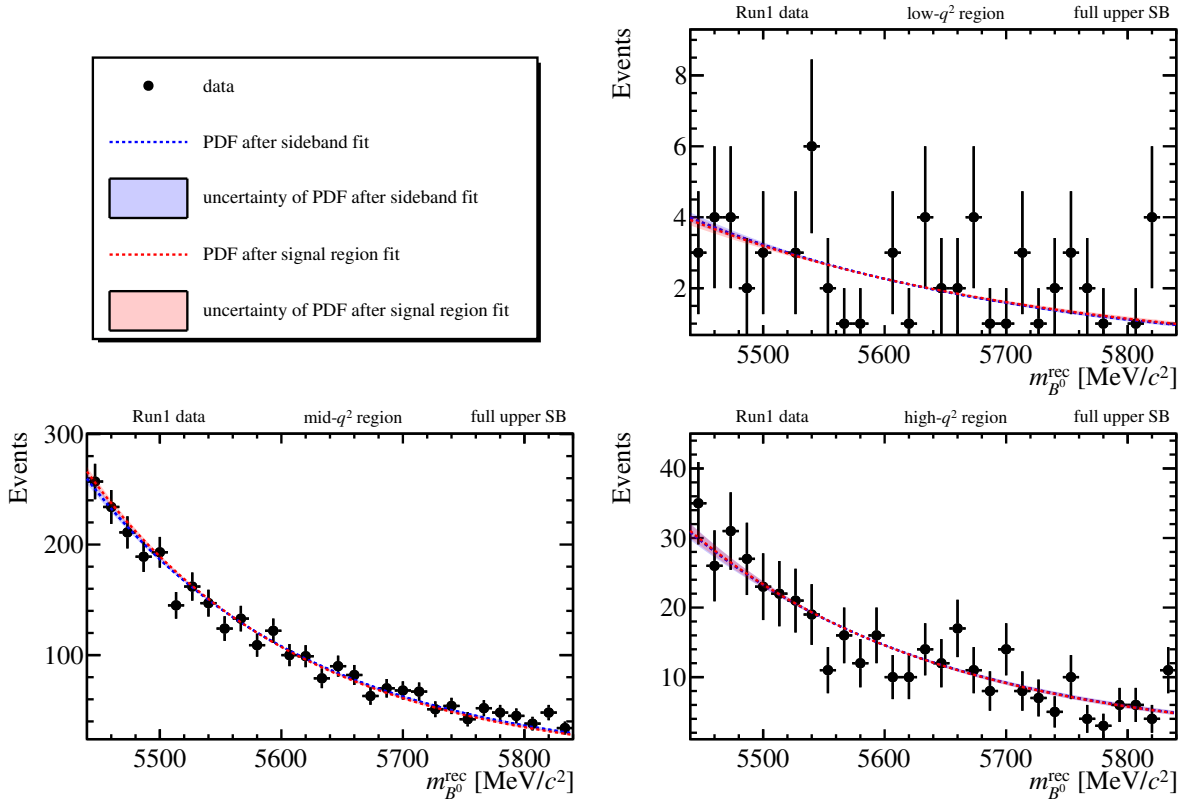


Figure A.33: $m_{B^0}^{\mathrm{rec}}$ distribution in each q^2 region in Run 1 data. Also shown are the projections of the background PDF after the simultaneous side-band fit (blue dashed line) as well as the PDFs' 68% confidence interval (blue shaded area) obtained by varying the parameters using the covariance matrix obtained in the fit. Furthermore, the background PDF after the signal region fit (red dashed line) and the respective 68% confidence interval (red shaded area) are shown. Note that the background events in the signal region (not shown here) add additional constraints on the background parameters in the signal region fit.