Imperial College London

Amplitude ansatz analysis of the decay $B^0 \to K^{*0}\mu^+\mu^-$

Matthew Christopher Birch

Department of High Energy Physics

Blackett Laboratory

Imperial College London

A thesis submitted at Imperial College London

for the degree of Doctor of Philosophy

September 2023

Abstract

The LHCb experiment at CERN has produced many intriguing results in $b \to s\ell\ell$ and $b \to c\ell\nu$ decays, which point to potential contributions from New Physics. This thesis describes an analysis of $B^0 \to K^{*0}\mu^+\mu^-$where the decay amplitudes are measured as functions of q^2, the dimuon invariant mass squared. Measuring the angular distribution of this decay results in a plethora of observables that can be used to constrain the type of New Physics contributions. Owing to the number of parameterisations of the Standard Model and other models which can describe New Physics effects, Legendre polynomial ansatzes are used to describe the variation with q^2 in order to be as model-independent as possible. A selection strategy for $B^0 \to K^{*0}\mu^+\mu^-$is outlined, resulting in low levels of misidentified and combinatorial background. Pseudoexperiment studies are performed in order to develop the analysis strategy and further understand the symmetries of the angular distribution and the fit. Blinded results from 9 fb^{-1} of data collected by the LHCb detector at the LHC at CERN are shown, where the fit is performed in the region $1.25 < q^2 < 8$ GeV$^2/c^4$. The fit is performed with four-parameter ansatzes for the P-wave and one-parameter ansatzes for the S-wave. Pseudoexperiment studies are performed from the data fit. The data fit quality is determined to be good, with p-value = 70%. Where the Hessian is not a good description of the uncertainties, fits to the log-likelihood profiles are performed with bifurcated parabolas to extract the uncertainties, resulting in good-quality statistical coverage.

Statement of Originality

The work presented in this thesis is the result of research I performed with the High Energy Physics group at Imperial College London between February 2020 and September 2023. This is with support from members of the Imperial College LHCb group, and members of the LHCb collaboration. I performed the work presented in this thesis, with the exceptions stated in the relevant chapters, such as producing the simulation and data samples, and the $B^+ \to K^{*+}\mu^+\mu^-$ BDT. Any plots which are not produced by myself are appropriately referenced.

This thesis has not been submitted for any other qualification.

Matthew Birch

September 2023

Copyright Declaration

Acknowledgements

Firstly, I would like to express my sincere gratitude to my supervisor, Mike McCann, for his exceptional mentorship and guidance throughout my PhD. Your expert advice and unwavering encouragement were instrumental in shaping my research and driving me to achieve my best work. I feel incredibly honoured to have had the opportunity to work with you.

I am also deeply grateful to the amplitude ansatz team: Kostas Petridis, Rui Wang, and Paula Álvarez-Cartelle, for the invaluable help, guidance, and many insightful discussions. I extend my thanks to Mitesh Patel for his excellent support and guidance, and to Mark Smith for the interesting and helpful discussions. I thank the LHCb Imperial group for creating such a wonderful and supportive community that I am proud to be part of.

To my friends, I would like to extend my heartfelt gratitude for your constant support and motivation that has made my PhD journey an unforgettable experience. Lastly, I express my profound appreciation to my family, including my parents Jane and Rich, and my sister Rachel, for their unwavering support and encouragement.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In recent years, experiments have measured quantities exhibiting discrepancies with respect to the Standard Model (SM). These include branching fractions, ratios of branching fractions, and angular coefficients. Examples of ratios of branching fractions are tests of Lepton Flavour Universality (LFU). Ignoring lepton masses, the SM predictions for the branching fractions of $B \to X\ell\ell$ are the same between the three lepton generations. There are theoretical uncertainties of these branching fractions due to low energy QCD effects, such as hadronic form-factors. However by measuring ratios of branching fractions, such as $R_{D^{(*)}} \equiv \frac{\mathcal{B}(B\to D^{(*)}\tau\nu_\mu)}{\mathcal{B}(B\to D^{(*)}\mu\nu_\tau)}$, these uncertainties cancel to first order. The most recent combined R_D and R_{D^*} LHCb measurement sees a 1.9σ deviation with respect to the SM, and global fits show a 3.2σ deviation with the SM, as indicated by Figure 1.1 [1].

Deviations are also seen in branching fractions. An example is the branching fraction of $B^0 \to K^{*0}\mu^+\mu^-$, measured with Run 1 data at LHCb, as shown by Figure 1.2 [2], where q^2 is the square of the dimuon invariant mass. This measurement is currently limited by systematic uncertainties. Nevertheless, it has been observed that in most q^2 bins the measured differential branching fraction is lower than the theoretical prediction. In addition the decay $B_s^0 \to \phi\mu^+\mu^-$, measured with the Run 1 and Run 2 datasets, shows a 3.2σ deviation with respect the SM, as seen in Figure 1.3a [3]. Also the differential branching fraction of decays such as $B^+ \to K^+\mu^+\mu^-$ show deviations with respect to the SM (e.g. Figure 1.3b [4]). What's intriguing about these plots is all of these measurements show a coherent undershooting with respect to the SM.

With regards to angular observables, these have been studied in decays such as

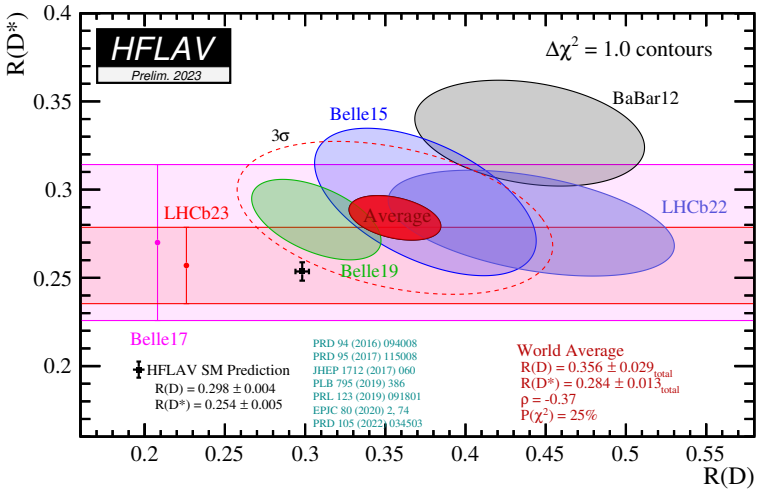


Figure 1.1: R_D and R_{D^*} measurements as of 2023 [1], where $R_{D^{(*)}} = \frac{\mathcal{B}(B\to D^{(*)}\tau\nu_\mu)}{\mathcal{B}(B\to D^{(*)}\mu\nu_\tau)}$. The world average of these measurements is shown by the red contour.

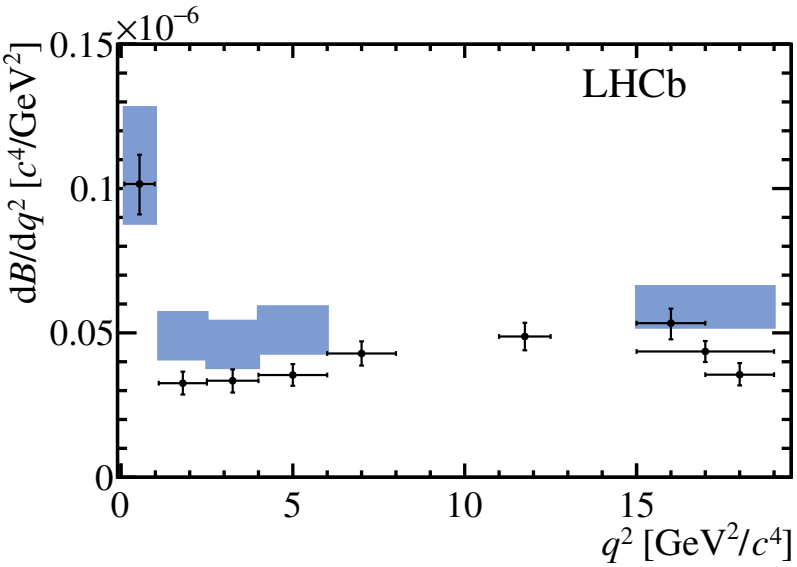
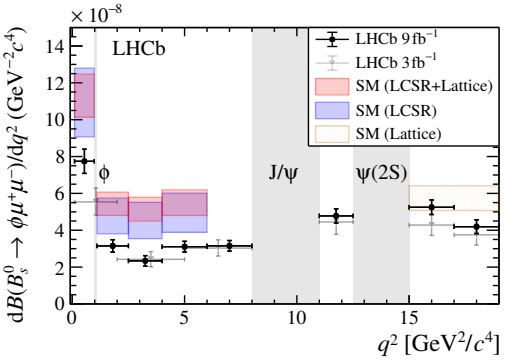
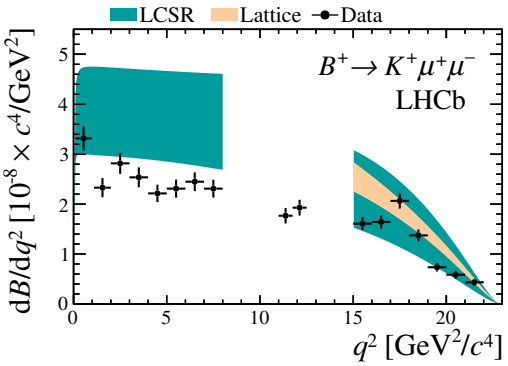


Figure 1.2: Branching fraction of $B^0 \to K^{*0}\mu^+\mu^-$, measured with Run 1 data at LHCb [2]. The data are shown by the black points and the blue bands corresponds to the Standard Model predictions from Refs. [5] [6].

$B^0 \to K^{*0}\mu^+\mu^-$. The latest published analysis of $B^0 \to K^{*0}\mu^+\mu^-$, which measures the angular coefficients of this decay, uses Run 1 and 2016 data from LHCb [7]. Local discrepancies are seen in observables, for example the observable P_5', as shown by Figure 1.4. The P observables are defined in Chapter 2. As described by [7], fits to the angular observable reveal discrepancies of 3σ in the real part of the value that parameterises vector currents, Re(C_9) (discussed further in Chapter 2).



(a) Branching fraction of $B_s^0 \to \phi\mu^+\mu^-$, measured with Run 1+Run 2 data at LHCb. The data are overlaid with Standard Model predictions using light cone sum rules (LCSR) and lattice QCD [3].



(b) Branching fraction of $B^+ \to K^+\mu^+\mu^-$, measured with Run 1 data at LHCb. The data are overlaid with Standard Model predictions using light cone sum rules (LCSR) and lattice QCD [4].

Figure 1.3: Branching fractions of decays involving $b \to s\ell\ell$ transitions measured at LHCb.

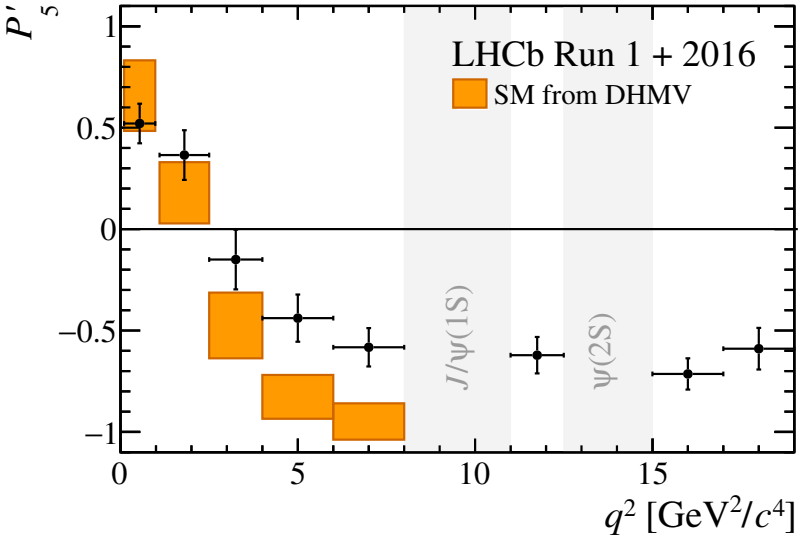


Figure 1.4: An angular observable from the $B^0 \to K^{*0}\mu^+\mu^-$ analysis at LHCb which uses Run 1 and 2016 data [7]. The data are shown by the black points and the orange bands show theoretical predictions from Ref. [8].

These Flavour Anomalies can be described by contributions from New Physics (NP). Global fits claim the tension to be $> 5\sigma$. Examples of global fits are described in [9]. A scenario discussed in [9] is where a NP coupling $C_{9\mu}$ is introduced, where the subscript μ indicates these contributions couple to muons. A $C_{9'\mu} = -C_{10'\mu}$ NP contribution is also introduced, where C_{10} is a value that parameterises axial-vector contributions and the $'$ indicates right-handed quark couplings. This fit, with a p-value goodness-of-fit of 31.1%, shows a pull of 7.1σ with respect to the SM, as shown by Figure 1.5. Contours from major

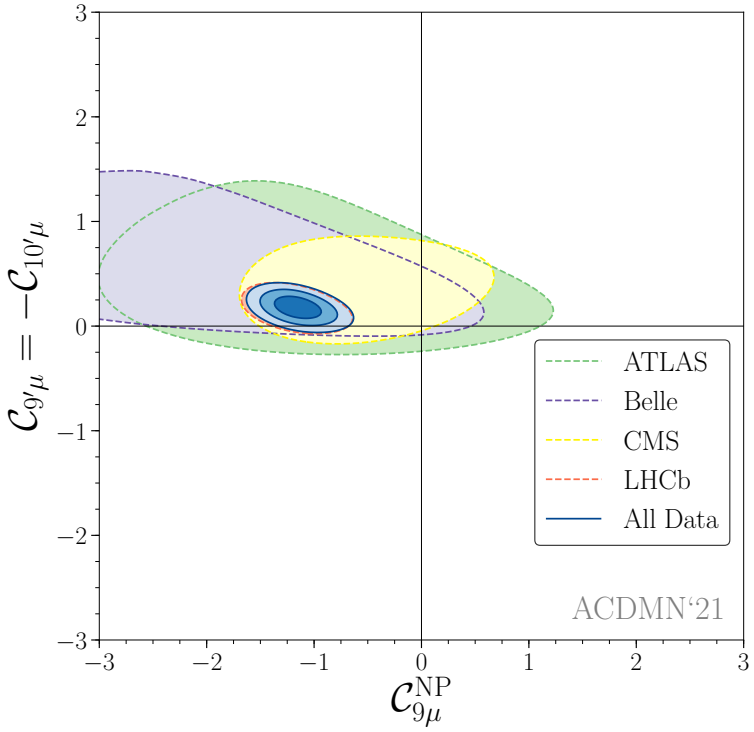


Figure 1.5: Global fit of $b \to s\ell\ell$ anomalies,where a New Physics coupling $C_{9\mu}$ is introduced, in addition to a $C_{9'\mu} = -C_{10'\mu}$ New Physics contribution [9]. The dashed lines represent 3σ regions and the solid lines correspond to 1σ, 2σ, and 3σ regions.

experiments are shown, in addition to the combination.

The analysis described in this thesis measures the decay amplitudes of $B^0 \to K^{*0}\mu^+\mu^-$ using Run 1 and Run 2 data from LHCb, which corresponds to 9 fb^{-1} of data. As described by Ref. [10], different models (i.e. the SM and NP models) result in different shapes of the angular observables as a function of q^2. The decay amplitudes are measured unbinned in q^2 (where q^2 is the dimuon invariant mass squared) in order to increase sensitivity to NP, since the shape of the amplitudes and hence the angular observables in q^2 can be exploited. There are many possible ways to theoretically describe this decay. In order to be as model-independent as possible, ansatzes are used to describe the decay amplitudes, as initially motivated by [11]. Providing the amplitude components and covariance matrix allows one to generate synthetic datasets from the fit, and perform fits to these synthetic datasets with any choice of model. A similar approach approach was performed as part of the Run 1 analysis of $B^0 \to K^{*0}\mu^+\mu^-$ at LHCb [12], however only the zero crossing points (values of q^2 where the angular observables cross zero) were published since the log-likelihood surface was found to be non-parabolic. These zero crossing points were

computed via bootstrapping [13].

This analysis is in collaboration with members from the Imperial, Bristol and Cambridge groups in LHCb. Unless otherwise specified, the work described in this thesis is my own.

Chapter 2 provides a theoretical overview of this decay and Chapter 3 describes the LHCb experiment and data flow. Chapter 4 presents an overall analysis flow and Chapter 5 describes the data selection used in this analysis. The acceptance, which is used to account for the angular and q^2 structures introduced by the selection, is described in Chapter 6. The fitting strategy is described in Chapter 7 and studies with pseudoexperiments are described in Chapters 8 and 9. The data results are shown in Chapters 10 and 11 for the control mode and the rare mode respectively. Conclusions are drawn in Chapter 12.

Chapter 2

Theory

This chapter describes the theoretical aspects related to this analysis. Section 2.1 provides an overview of the Standard Model, and problems with the Standard Model are discussed in Section 2.2. Section 2.3 describes the decay $B^0 \to K^{*0}\mu^+\mu^-$, which is the subject of the analysis described in this thesis. Effective field theories, the basis of the theoretical framework describing this decay, are discussed in Section 2.4, and hadronic effects are described in Section 2.5. The angles used to parameterise the decay $B^0 \to K^{*0}\mu^+\mu^-$ are defined in Section 2.6 and the differential decay rate is described in Section 2.7. The CP-averages and asymmetries are defined in Section 2.8. The differential decay rate as a function of the mass of the kaon-pion system, $m_{K\pi}$, is described in Section 2.9 and the symmetries of the decay rates are discussed in Section 2.10.

2.1 The Standard Model

The Standard Model (SM) is a quantum field theory describing three of the four known interactions of nature (the electromagnetic, strong, and weak interactions). Fundamental particles are excitations of their respective quantum fields, and are presented in Table 2.1 [14].

The SM can be written in a Lagrangian formalism. Certain gauge transformations applied to terms in the Lagrangian leaves the Lagrangian invariant. Groups can be used to describe these transformations. The electroweak interaction, which corresponds to the electromagnetic interaction unified with the weak interaction, obeys symmetries described

Fermions			**Bosons**	
Up (u)	**Charm** (c)	**Top** (t)	**Gluon** (g)	**Higgs** (H)
Mass = 2.2 MeV/c^2	1.275 GeV/c^2	172.8 GeV/c^2	0	125.3 GeV/c^2
Charge = 2/3	2/3	2/3	0	0
Spin = 1/2	1/2	1/2	1	0
Down (d)	**Strange** (s)	**Beauty** (b)	**Photon** (γ)	
Mass = 4.7 MeV/c^2	95 MeV/c^2	4.18 GeV/c^2	0	
Charge = −1/3	−1/3	−1/3	0	
Spin = 1/2	1/2	1/2	1	
Electron (e)	**Muon** (μ)	**Tau** (τ)	**W boson** (W)	
Mass = 511 keV/c^2	105.66 MeV/c^2	1.7769 GeV/c^2	80.433 GeV/c^2	
Charge = −1	−1	−1	1	
Spin = 1/2	1/2	1/2	1	
Electron neutrino (ν_e)	**Muon neutrino** (ν_μ)	**Tau neutrino** (ν_τ)	**Z boson** (Z)	
Mass < 0.8 eV/c^2	< 0.19 eV/c^2	< 18.2 eV/c^2	91.19 GeV/c^2	
Charge = 0	0	0	0	
Spin = 1/2	1/2	1/2	1	

Table 2.1: The Standard Model of particle physics.

by the SU(2)×U(1) groups [15] [16] [17].

The SM is based on the symmetry group $\mathrm{SU(3)}_C \times \mathrm{SU(2)}_L \times \mathrm{U(1)}_Y$ where C corresponds to colour, L represents left-handed fermions, and the hypercharge Y is given by

$$Y = 2(Q - I_3), \tag{2.1}$$

where Q is the electric charge and I_3 is the third component of the weak isospin. Left-handed quarks form a doublet under $SU(2)_L$, with hypercharge $Y = 1/3$ (for up-type quarks, $Y = 2(2/3 - 1/2)$; for down type quarks, $Y = 2(-1/3 + 1/2)$). The right-handed quarks are singlets under $SU(2)_L$, where $Y = 4/3$ for up-type right-handed quarks, and $Y = -2/3$ for down-type right-handed quarks [18].

In the SM, quark masses are generated through interaction with Higgs doublet Φ. Expanding the field around the ground state, in the 'unitrary gauge', $\Phi = \begin{pmatrix} (v+h)/\sqrt{2} \\ 0 \end{pmatrix}$ where h is a real scalar field and v is the Higgs field vacuum expectation value [19] [20] [21]. The Yukawa Lagrangian describes interaction between left-handed and right-handed quarks through the Higgs and is given by

$$\mathcal{L}_{\text{Yukawa}} = -G_D \bar{Q}_L \Phi d_R - G_U \bar{Q}_L \Phi^c u_R \tag{2.2}$$

where

$$Q_L = \begin{bmatrix} u_L \\ d_L \end{bmatrix} \tag{2.3}$$

is a $SU(2)_L$ left-handed quark doublet, u_R and d_R are the the $SU(2)_L$ right-handed quark singlets, and G_D and G_U are constants. Extending this to three generations, these constants become flavour matrices. However this results in the mass matrices after symmetry breaking becoming non-diagonal.

In order to make the mass matrices diagonal, the left- and right- handed quarks are redefined as

$$\begin{aligned} u_{X,i} &\rightarrow M_{u_X,ij} u_{X,j} \\ d_{X,i} &\rightarrow M_{d_X,ij} u_{X,j}, \end{aligned} \tag{2.4}$$

where M are orthogonal matrices, i and j correspond to the quark generations, and $X = \{L, R\}$. Applying these transformations results in a diagonal mass matrix, in addition to the kinetic, electromagnetic and neutral current terms. The charge-current becomes of the form

$$\overline{u}_L^f (M_{u_L}^\dagger M_{d_L})_{fg} d_L^g, \tag{2.5}$$

which manifests as mixing between the quark generations. $(M_{u_L}^\dagger M_{d_L})_{fg}$ can be associated to the CKM matrix [22] [23]. Explicitly, the charge-current term becomes

$$\frac{-v}{\sqrt{2}} \begin{pmatrix} \overline{u}_L & \overline{c}_L & \overline{t}_L \end{pmatrix} \gamma^\mu W_\mu^+ V_{CKM} \begin{pmatrix} d_L \\ s_L \\ b_L \end{pmatrix}, \tag{2.6}$$

where $V_{CKM} \equiv M_{u_L}^\dagger M_{d_L} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}$.

2.2 Problems with the Standard Model

Whilst the SM is the most successful theory of particle physics, there are several problems with the SM, most notably observed phenomena which are unexplained. A striking example is the absence of gravity in the SM. The most successful theory describing gravity to date is

General Relativity (GR) [24], which is a classical theory that describes gravity in terms of the curvature of spacetime. The SM and GR have completely different theoretical frameworks and attempts to unify the SM and GR have been extremely challenging. An example challenge when attempting to introduce gravity in the SM is the non-renormalisability of GR. Nevertheless, there are several theories which can describe quantum gravity, such as string theory and loop quantum gravity, however these theories have not been experimentally verified.

There is a plethora of evidence for Dark Matter (DM) however the SM has no DM candidate. Evidence for DM arises from the cosmic microwave background, gravitational lensing, galactic clusters and velocity profiles of galaxies [25] [26] [27]. From cosmological observations, DM is predicted to be approximately 27% of the mass-energy content of the universe. There are several candidates for DM, such as Weakly Interacting Massive Particles [28]. Recent limits exclude cross sections above 9.2×10^{-48} at WIMP mass of 36 GeV/c^2 [29]. Another candidate for dark matter is axions, which was proposed [30] [31] to solve the strong CP problem in quantum chromodynamics [32].

The SM cannot explain a large portion of the matter-antimatter asymmetry we observe in the universe. The Sakharov conditions for baryogenesis are baryon number violation, CP violation, and interactions out of thermal equilibrium [33]. In principle, the SM allows for baryon number violation through non-perturbative processes [34], however this is not enough to account for the matter-antimatter asymmetry in the universe. No baryon number violating process has been observed. In addition, the CP violation measured in the quark sector so far is not enough to explain the matter-antimatter asymmetry in the universe. Current limits on electric dipole moments such as Ref. [35] also provide constraints on possible sources of CP violation. Recent measurements from T2K indicate CP violation in the neutrino sector [36], with a preference near maximal CP violation, however more data is required to determine whether the CP violation here is enough to explain the matter-antimatter asymmetry in the universe.

There is evidence for dark energy such as from supernovae [37] and the cosmic microwave background, however the SM has no dark energy candidate. The cosmological constant in GR can account for dark energy, however the vacuum energy density predicted in the SM is many orders of magnitude larger than the observed value of dark energy. Various

proposals have been put forward to explain dark energy, such as quintessence [38] [39] or chameleons [40], as well as black hole accretion [41].

There are also many theoretical problems in the SM, such as the hierarchy problem. The Higgs mass in the SM has very large quantum corrections, which are orders of magnitude larger than the corrected mass of the Higgs. Naively one would expect the corrected Higgs mass to be of the same order of magnitude as the very large quantum corrections, i.e. towards the Planck mass. Thus there a high amount of fine-tuning in the cancellation of the bare Higgs mass and the quantum corrections. There is the question of why gravity is so much weaker than the other interactions. There are several proposals to address this, such as supersymmetry or extra dimensions.

In the SM, neutrinos are massless. However, since neutrinos are observed to oscillate [42] [43], they are therefore required to have mass. It is currently not known how neutrinos acquire their mass and whether neutrinos are Dirac or Majorana particles. In addition, recent results [44] present an upper limit of the effective electron anti-neutrino mass of 0.8 eV/c^2 at a 90% confidence level. This is much smaller than the masses of other SM massive particles. It is not known why the neutrinos mass is so small, as well as whether the neutrino masses are arranged in the 'normal hierarchy' or the 'inverted hierarchy'. In addition, future measurements will determine whether CP violation in the neutrino sector is larger than that in the quark sector [36].

There are a large number of free parameters in the SM such as the fermion masses, Higgs mass, and elements of the CKM matrix. In addition there is no explanation for why there are three generations of matter. To summarise, there are phenomena observed in nature but not explained by the SM, in addition to theoretical questions regarding the structure of the SM, with the hope of a more fundamental theory which can explain some of these mysteries.

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

As discussed in Section 2.2, there are deficits with the SM, for example it cannot explain the matter-antimatter asymmetry in our universe, and it also does not have a dark matter candidate. One of the most effective ways to search for Beyond the Standard Model (BSM)

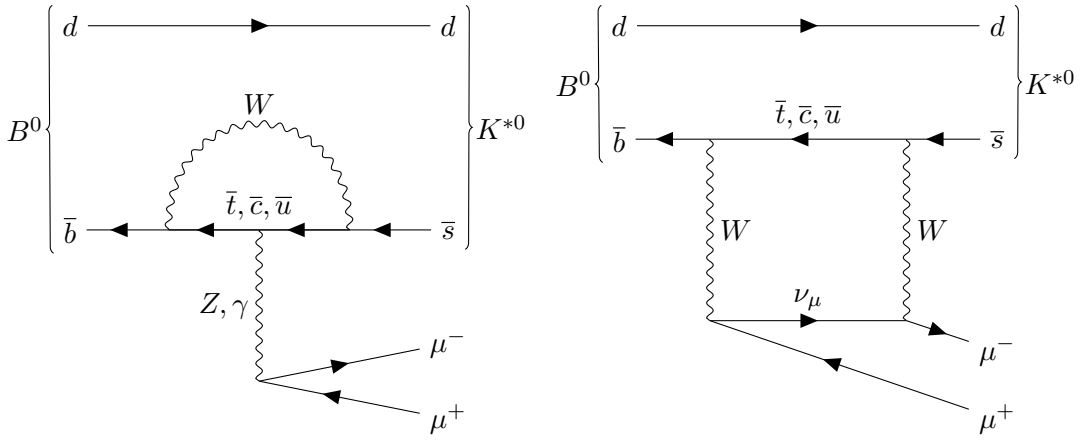


Figure 2.1: Example Feynman diagrams describing the decay $B^0 \to K^{*0}\mu^+\mu^-$.

physics is by measuring rare decays such as $B^0 \to K^{*0}\mu^+\mu^-$. Searching for NP through rare decays can probe BSM physics at energies beyond those accessible by direct searches since the BSM particle can be produced virtually. Quantities such as branching fractions and angular observables can be measured and compared to the SM. As discussed in Chapter 1, recent measurements in $b \to s\ell\ell$ and $b \to c\ell\nu$ transitions, such as $B^0 \to K^{*0}\mu^+\mu^-$, have revealed discrepancies with respect to the SM. By performing an angular analysis of $B^0 \to K^{*0}\mu^+\mu^-$, not only one can determine whether there is NP, but one can also deduce the nature of the NP model.

The decay $B^0 \to K^{*0}\mu^+\mu^-$ occurs through the quark-level transition $b \to s\ell\ell$. It is a Flavour Changing Neutral Current (FCNC) thus it can only occur through loop-level Feynman diagrams in the SM, as indicated by Figure 2.1. Since this process can only occur at loop-level and involves at least one off-diagonal CKM element (i.e. V_{ts}), it is highly suppressed in the SM. Studying this decay (and other rare decay processes) allows one to search for contributions from BSM physics without large SM backgrounds. These NP contributions could occur through tree-level Feynman diagrams and potentially be large compared to the SM. In addition, measuring these processes provides a test of the flavour structure in the SM and helps constrain possible NP models.

2.4 Effective field theories

Processes such as $B^0 \to K^{*0}\mu^+\mu^-$ have contributions with different associated energy scales: $\Lambda_{QCD} \sim 0.2$ GeV, which is the strength of Quantum Chromodynamics (QCD) interactions between the quarks, $m_b \sim 4$ GeV/c^2, which is the mass of the b quark, and $m_W \sim 80$ GeV/c^2, which is the mass of the W boson involved in the FCNC. This results in calculations when summing over Feynman diagrams becoming extremely difficult to perform due to large logarithms which do not get smaller as the summation order increases, thus requiring summation to all orders. This problem is solved by computing these processes in an Effective Field Theory (EFT) framework via an Operator Product Expansion (OPE), where a product of local quantum fields is expanded as a summation of those fields [45] [46].

EFTs allow one to separate high and low-energy effects. Suppose we have a particle with mass M and a process which occurs at energy E, and $E < \mu < M$, where μ is some scale below which the dynamics of M is unimportant. The quantum fields can be decomposed into high and low energy modes, ψ_L and ψ_H, where they describe effects below and above μ respectively. Analogously to Fermi theory, the fields at high energies ψ_H are integrated out. This results in the EFT Lagrangian

$$\mathcal{L}_{EFT} = \sum_i \mathcal{C}_i(\mu)\mathcal{O}_i(\mu) \tag{2.7}$$

where $\mathcal{C}_i(\mu)$ are Wilson coefficients, which describe effects at high energies. The operators $\mathcal{O}_i(\mu)$ describe effects at low energies. Thus the high and low energy effects factorise. However the observables should not depend on the scale μ, thus the Wilson coefficients $\mathcal{C}_i(\mu)$ obey the Renormalisation Group Equation (RGE) [46]

$$\mu \frac{d\mathcal{C}_i(\mu)}{d\mu} = \gamma_{ij}\mathcal{C}_j(\mu). \tag{2.8}$$

This equation provides a relationship between the Wilson coefficients and the scale μ. Hence predictions are generated at low energies by matching the theory at high energy, i.e. at M_W, where perturbation theory can be used, and evolving the predictions from M_W to m_b. Thus one has $\mathcal{C}_i(m_b) = K_{ij}\mathcal{C}_j(M_W)$, where K_{ij} is some matrix.

2.4.1 The effective Hamiltonian for the decay $B^0 \to K^{*0}\mu^+\mu^-$

The effective Hamiltonian, after performing an Operator Product Expansion in $\frac{1}{M_W^2}$, is given by

$$\mathcal{H}_{eff} = -\frac{4G_F}{\sqrt{2}} V_{tb}V_{ts}^* \sum_i \mathcal{C}_i(\mu)\mathcal{O}_i(\mu), \tag{2.9}$$

where G_F is the Fermi constant, $\mathcal{C}_i(\mu)$ are Wilson coefficients and $\mathcal{O}_i(\mu)$ are operators. The operators are $\mathcal{O}_i$ are defined as

$$\begin{aligned}
\mathcal{O}_{1u} &= [\bar{s}\gamma_\mu T^a P_L u][\bar{u}\gamma^\mu T^a P_L b] & \mathcal{O}_5 &= [\bar{s}\gamma_\mu\gamma_\nu\gamma_\rho P_L b]\sum_q [\bar{q}\gamma^\mu\gamma^\nu\gamma^\rho P_L q] \\
\mathcal{O}_{1c} &= [\bar{s}\gamma_\mu T^a P_L c][\bar{c}\gamma^\mu T^a P_L b] & \mathcal{O}_6 &= [\bar{s}\gamma_\mu\gamma_\nu\gamma_\rho T^a P_L b]\sum_q [\bar{q}\gamma^\mu\gamma^\nu\gamma^\rho T^a P_L q] \\
\mathcal{O}_{2u} &= [\bar{s}\gamma_\mu P_L u][\bar{u}\gamma^\mu P_L b] & \mathcal{O}_7 &= \frac{e}{(4\pi)^2} m_b(\bar{s}\sigma^{\mu\nu} P_R b)F_{\mu\nu} \\
\mathcal{O}_{2c} &= [\bar{s}\gamma_\mu P_L c][\bar{c}\gamma^\mu P_L b] & \mathcal{O}_8 &= \frac{e}{(4\pi)^2} m_b(\bar{s}\sigma^{\mu\nu} P_R T^a b)G^a_{\mu\nu} \\
\mathcal{O}_3 &= [\bar{s}\gamma_\mu P_L b]\sum_q [\bar{q}\gamma^\mu P_L q] & \mathcal{O}_9 &= \frac{e^2}{(4\pi)^2}(\bar{s}\gamma_\mu P_L b)(\bar{\ell}\gamma^\mu \ell) \\
\mathcal{O}_4 &= [\bar{s}\gamma_\mu T^a P_L b]\sum_q [\bar{q}\gamma^\mu T^a P_L q] & \mathcal{O}_{10} &= \frac{e^2}{(4\pi)^2}(\bar{s}\gamma_\mu P_L b)(\bar{\ell}\gamma^\mu\gamma^5 \ell)
\end{aligned} \tag{2.10}$$

where T^a are the $\mathrm{SU}(3)_C$ generators, $P_{L/R} = (1 \mp \gamma_5)/2$ are the chiral projection operators, $\sigma^{\mu\nu}$ are the Pauli spin matrices, $F_{\mu\nu}$ is the electromagnetic field tensor, and $G^a_{\mu\nu}$ is the gluon field tensor. One can also define the primed operators $\mathcal{O}'$ by applying the transformation $P_L \leftrightarrow P_R$.

The dominant contributions to the process $B^0 \to K^{*0}\mu^+\mu^-$ in the SM are from $\mathcal{O}_9$, $\mathcal{O}_{10}$, and $\mathcal{O}_7$. $\mathcal{O}_9$ and $\mathcal{O}_{10}$ are the semi-leptonic operators which describe vector and axial-vector currents respectively. The electromagnetic dipole operator is $\mathcal{O}_7$. The operators $\mathcal{O}_{1u}$, $\mathcal{O}_{1c}$, $\mathcal{O}_{2u}$, and $\mathcal{O}_{2c}$ are the charge-current operators, also known as the 4-quark operators, and describe processes where $q\bar{q}$ pairs are produced. These also contribute to the process $B^0 \to K^{*0}\mu^+\mu^-$ since the quark pair can subsequently decay to two muons. The gluon dipole operator $\mathcal{O}_8$ and the QCD penguin operators $\mathcal{O}_3$, $\mathcal{O}_4$, $\mathcal{O}_5$, and and $\mathcal{O}_6$ contribute, albeit at a lower level compared to the charge-current operators.

The dimuon invariant mass squared (q^2) spectrum of the decay $B^0 \to K^{*0}\mu^+\mu^-$ is shown in Figure 2.2. Below $q^2 = (2m_\mu)^2$, the decay rate is exactly zero. In the decay

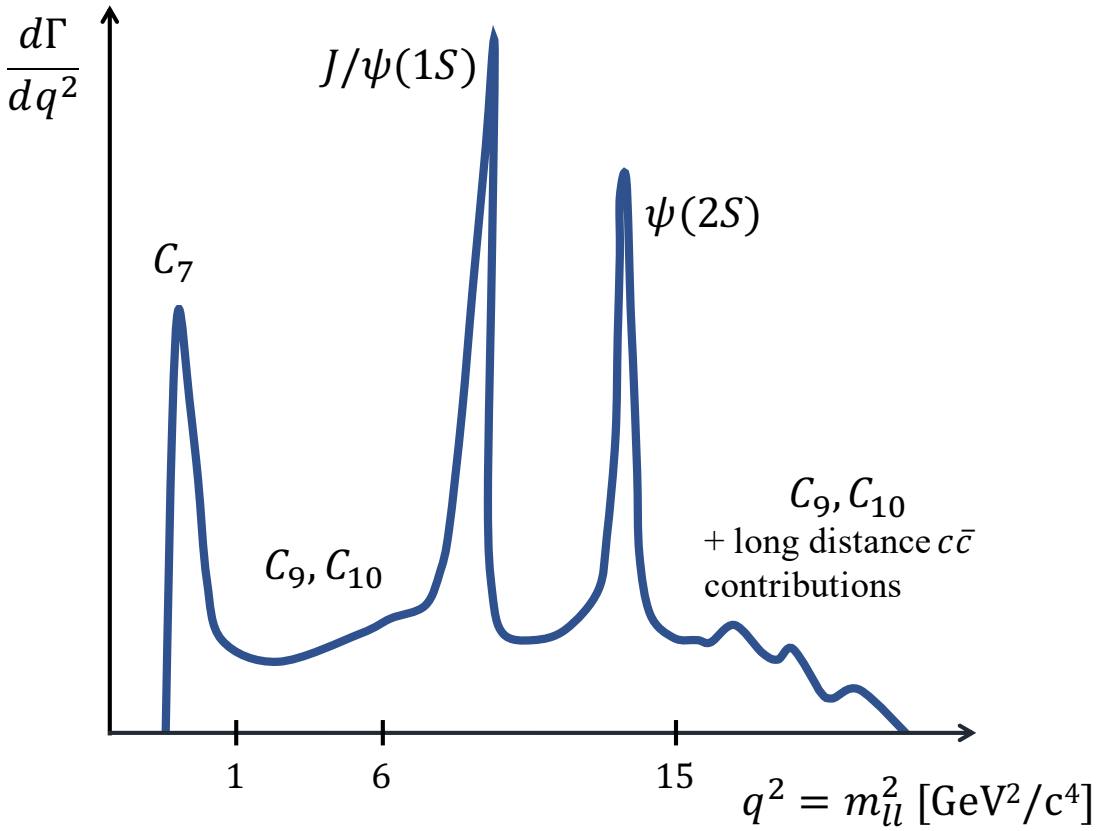


Figure 2.2: Dimuon invariant mass squared (q^2) spectrum of the decay $B^0 \to K^{*0}\mu^+\mu^-$. There are two narrow states, the $J/\psi(1S)$ meson, with mass 3096 MeV, and the $\psi(2S)$ meson, with mass 3686 MeV. Most of the regions dominated by the rare mode are dominated by the Wilson coefficients $\mathcal{C}_9$ and $\mathcal{C}_{10}$ and the low q^2 region is dominated by $\mathcal{C}_7$. The q^2 region the analysis described in thesis is the region $1.25 < q^2 < 8$ GeV$^2/c^4$. The J/ψ region is used for validating the analysis.

$B^0 \to K^* X(c\bar{c})$, where $X(c\bar{c}) \to \mu^+\mu^-$, there are two narrow states, the $J/\psi(1S)$ meson, with mass 3096 MeV, and the $\psi(2S)$ meson, with mass 3686 MeV. There are broad $c\bar{c}$ resonances at higher mass. In addition to the 1-particle contributions, there are 2-particle contributions such as $D\bar{D}$ and $D^*\bar{D}$ and a $\tau\tau$ contribution. Most of the regions dominated by the rare mode are dominated by the Wilson coefficients $\mathcal{C}_9$ and $\mathcal{C}_{10}$ and the low q^2 region is dominated by $\mathcal{C}_7$. The analysis described in this thesis is performed in the q^2 regions $1.25 < q^2 < 8$ GeV$^2/c^4$ and $11 < q^2 < 12.5$ GeV$^2/c^4$. The J/ψ region is used for validating the analysis. This analysis makes an assumption that the muons are massless, as adopted by previous analyses in $B^0 \to K^{*0}\mu^+\mu^-$ [7]. This is a valid assumption since in the q^2 regions this analysis is performed in, $q^2 \gg (2m_\mu)^2$, where m_μ is the muon mass.

2.5 Hadronic effects

The amplitude, describing the transition from the initial state to the final state, is given by

$$\mathcal{A}_{EFT} = \sum_i \mathcal{C}_i(m_B)\langle K^{*0}\mu^+\mu^-|\mathcal{O}_i(m_B)|B^0\rangle. \tag{2.11}$$

Whilst calculating $\mathcal{C}_i(m_B)$ is not problematic, calculating the matrix elements $\langle K^{*0}\mu^+\mu^-|\mathcal{O}_i(m_B)|B^0\rangle$ is much more complicated since QCD is non-perturbative at low energies. These can be split into local and non-local hadronic effects. The amplitudes are sensitive to form-factors which encapsulate local hadronic effects, such as QCD interactions corresponding to the bound meson state and the vacuum. These are described in more detail in Section 2.5.1. In addition, the amplitudes also have a contribution from non-local effects, which are described in Section 2.5.2.

As described by Ref. [47], the coefficients $\mathcal{C}_7$ and $\mathcal{C}_{10}$ appear with contributions from other Wilson coefficients $\mathcal{C}_i$ in the matrix elements. It is thus convenient to define the effective Wilson Coefficients $\mathcal{C}_7^{\text{eff}}$, $\mathcal{C}_9^{\text{eff}}$, and $\mathcal{C}_{10}^{\text{eff}}$, which are given by [47]

$$\begin{aligned}
C_7^{\text{eff}} &= \frac{4\pi}{\alpha_s}C_7 - \frac{1}{3}C_3 - \frac{4}{9}C_4 - \frac{20}{3}C_5 - \frac{80}{9}C_6 \\
C_9^{\text{eff}} &= \frac{4\pi}{\alpha_s}C_9 + Y(q^2) \\
C_{10}^{\text{eff}} &= \frac{4\pi}{\alpha_s}C_{10} \\
C_7^{\text{eff}\prime} &= \frac{4\pi}{\alpha_s}C'_7 \\
C_9^{\text{eff}\prime} &= \frac{4\pi}{\alpha_s}C'_9 \\
C_{10}^{\text{eff}\prime} &= \frac{4\pi}{\alpha_s}C'_{10}
\end{aligned} \tag{2.12}$$

where

$$\begin{aligned}
Y(q^2) =& h(q^2, m_c)\left(\frac{4}{3}C_1 + C_2 + 6C_3 + 60C_5\right) \\
&- \frac{1}{2}h(q^2, m_b)\left(7C_3 + \frac{4}{3}C_4 + 76C_5 + \frac{64}{3}C_6\right) \\
&- \frac{1}{2}h(q^2, 0)\left(C_3 + \frac{4}{3}C_4 + 16C_5 + \frac{64}{3}C_6\right) \\
&+ \frac{4}{3}C_3 + \frac{64}{9}C_5 + \frac{64}{27}C_6
\end{aligned} \tag{2.13}$$

with

$$h(q^2, m_q) = -\frac{4}{9}\left(\ln\frac{m_q^2}{\mu^2} - \frac{2}{3} - z\right) - \frac{4}{9}(2-z)\sqrt{|z-1|} \times \begin{cases} \arctan\frac{1}{\sqrt{z-1}} & z > 1 \\ \ln\frac{1+\sqrt{1+z}}{\sqrt{z}} - \frac{i\pi}{2} & z \leq 1 \end{cases} \tag{2.14}$$

and $z = 4m_q^2/q^2$. μ is the renormalisation scale and m_b and m_c are the masses of the b and c quarks respectively.

Transversity amplitudes $\mathcal{A}$ [48] are commonly used to describe the decay $B^0 \to K^{*0}\mu^+\mu^-$. These have a definite parity and can be written in terms of the Helicity amplitudes [49]. As described by Ref. [47], the transversity amplitudes are written in terms of the effective Wilson coeffients and form-factors. There are three polarisations of the K^{*0}: parallel, transverse and longitudinal, represented by the subscripts $\|$, $\perp$, and 0 respectively. The decay involves a muon-antimuon pair, so the superscripts L and R are used to represent left-handed and right-handed muonic currents respectively. There is also an addition amplitude $\mathcal{A}_t$, which corresponds to where the gauge boson (the virtual Z or photon) polarisation vector is time-like in the gauge boson rest frame. Assuming massless muons, this amplitude vanishes [47]. The amplitudes written in terms of the effective Wilson coefficients and form-factors are given as

$$\begin{aligned}
\mathcal{A}_\|^{L,R} =& -N\sqrt{2}(m_B^2 - m_{K^*}^2)\left[\left[(\mathcal{C}_9^{\text{eff}} - \mathcal{C}_9^{\text{eff}'}) \mp (\mathcal{C}_{10}^{\text{eff}} - \mathcal{C}_{10}^{\text{eff}'})\right]\frac{A_1(q^2)}{m_B - m_{K^*}}\right. \\
&\left. + \frac{2m_b}{q^2}(\mathcal{C}_7^{\text{eff}} - \mathcal{C}_7^{\text{eff}'})T_2(q^2)\right] \\
\mathcal{A}_\perp^{L,R} =& N\sqrt{2\lambda}\left[\left[(\mathcal{C}_9^{\text{eff}} + \mathcal{C}_9^{\text{eff}'}) \mp (\mathcal{C}_{10}^{\text{eff}} + \mathcal{C}_{10}^{\text{eff}'})\right]\frac{V(q^2)}{m_B + m_{K^*}} + \frac{2m_b}{q^2}(\mathcal{C}_7^{\text{eff}} + \mathcal{C}_7^{\text{eff}'})T_1(q^2)\right] \\
\mathcal{A}_0^{L,R} =& -\frac{N}{2m_K^*\sqrt{q^2}}\left\{(\mathcal{C}_9^{\text{eff}} - \mathcal{C}_9^{\text{eff}'}) \mp (\mathcal{C}_{10}^{\text{eff}} - \mathcal{C}_{10}^{\text{eff}'})\right. \\
&\times\left[(m_B^2 - m_{K^*}^2 - q^2)(m_B + m_K^*)A_1(q^2) - \lambda\frac{A_2(q^2)}{m_B + m_K^*}\right] \\
&\left. + 2m_B(\mathcal{C}_7^{\text{eff}} - \mathcal{C}_7^{\text{eff}'})\left[(m_B^2 + 3m_{K^*}^2 - q^2)T_2(q^2) - \frac{\lambda}{m_B^2 - m_{K^*}^2}T_3(q^2)\right]\right\}
\end{aligned} \tag{2.15}$$

where m_B is the B^0 mass, m_{K^*} is the mass of the K^{*0}, and $N = V_{tb}V_{ts}^*\left[\frac{G_F^2\alpha^2}{3\times 2^{10}\pi^5 m_B^3}q^2\lambda^{1/2}\right]^{1/2}$ with $\lambda = m_B^4 + m_{K^*}^4 + q^4 - 2(m_B m_{K^*} + m_{K^*}q^2 + m_B^2 q^2)$. T_1, T_2 and T_3 are dipole form-

factors and V, A_1, and A_2 are vector and axial-vector form-factors.

2.5.1 Local hadronic effects

The hadronic matrix elements cannot be computed perturbatively due to low energy QCD effects. They are parameterised as summations of Lorentz structures and functions of kinematic variables [50]. These functions are called 'form-factors'.

There are two categories of techniques in which form factors are computed by the theoretical community. One, Lattice QCD, is generally computed at low hadronic recoil, i.e. high q^2. Whilst in principle Lattice QCD can be computed at any value of q^2, the predictions are generally more precise at high q^2, i.e. $q^2 \gtrsim 10$ GeV$^2/c^4$ which corresponds to low hadronic recoil. Alternatively, one can use a method such as Light Cone Sum Rules [5] [51]. Here the predictions are computed at negative q^2 where the theory is most precise and extrapolated into the $q^2 > 0$ region. The uncertainty due to form-factors contributes significantly to the overall uncertainty of the theoretical predictions.

2.5.2 Non-local hadronic effects

Even when removing the q^2 regions corresponding to the resonances in order to measure the rare mode, there is still a possibility of interference between the resonances and the rare mode. The effective Wilson coefficients $\mathcal{C}_7^{\text{eff}}$, $\mathcal{C}_9^{\text{eff}}$, and $\mathcal{C}_{10}^{\text{eff}}$ include contributions from the 4-quark operators, thus there is a need to compute predictions for $b \to c\bar{c}s$ processes. The next-to-leading order non-local contributions are difficult to compute, in particular the interference between the non-local contributions and the rare mode.

The predictions for these non-local effects can be estimated through Light Cone Sum Rules [52] and extrapolating to higher q^2, or through data-driven methods, such as measuring the resonances from the data.

2.6 Angles used to parameterise the decay $B^0 \to K^{*0}\mu^+\mu^-$

In order to measure the decay amplitudes or angular coefficients of $B^0 \to K^{*0}\mu^+\mu^-$, where $K^{*0} \to K^+\pi^-$, the angular distribution of the decay products (K^+, π^-, μ^+, and μ^-) is measured. The three angles θ_ℓ, θ_K, and ϕ are used to parameterise the decay

$B^0 \to K^{*0}\mu^+\mu^-$, providing a complete description of the angular distribution of the decay products in $\cos\theta_\ell$, $\cos\theta_K$, and ϕ. A schematic of the angular definitions is given by Figure 2.3 [53]. The notation n_{ab} corresponds to the direction normal to the plane containing particles a and b in the B^0 (or $\bar{B}^0$) rest frame. The angular basis is defined in [53].

The angle θ_ℓ is defined as the angle between the direction of the $\mu^+(\mu^-)$ in the dimuon rest frame and the direction of the dimuon system in the B^0 ($\bar{B}^0$) rest frame. The angle θ_K is the angle between the direction of the kaon in the K^{*0} ($\bar{K}^{*0}$) rest frame and the direction of the K^{*0} ($\bar{K}^{*0}$) in the B^0 ($\bar{B}^0$) rest frame. Explicitly, $\cos\theta_\ell$ and $\cos\theta_K$ are defined as

$$\cos\theta_\ell = \left(\hat{p}_{\mu^+}^{(\mu^+\mu^-)}\right)\cdot\left(\hat{p}_{\mu^+\mu^-}^{(B^0)}\right) = \left(\hat{p}_{\mu^+}^{(\mu^+\mu^-)}\right)\cdot\left(-\hat{p}_{B^0}^{(\mu^+\mu^-)}\right), \tag{2.16}$$

$$\cos\theta_K = \left(\hat{p}_{K^+}^{(K^{*0})}\right)\cdot\left(\hat{p}_{K^{*0}}^{(B^0)}\right) = \left(\hat{p}_{K^+}^{(K^{*0})}\right)\cdot\left(-\hat{p}_{B^0}^{(K^{*0})}\right) \tag{2.17}$$

for the B^0 and

$$\cos\theta_\ell = \left(\hat{p}_{\mu^-}^{(\mu^+\mu^-)}\right)\cdot\left(\hat{p}_{\mu^+\mu^-}^{(\bar{B}^0)}\right) = \left(\hat{p}_{\mu^-}^{(\mu^+\mu^-)}\right)\cdot\left(-\hat{p}_{\bar{B}^0}^{(\mu^+\mu^-)}\right), \tag{2.18}$$

$$\cos\theta_K = \left(\hat{p}_{K^-}^{(\bar{K}^{*0})}\right)\cdot\left(\hat{p}_{\bar{K}^{*0}}^{(\bar{B}^0)}\right) = \left(\hat{p}_{K^-}^{(\bar{K}^{*0})}\right)\cdot\left(-\hat{p}_{\bar{B}^0}^{(\bar{K}^{*0})}\right) \tag{2.19}$$

for the $\bar{B}^0$ decay.

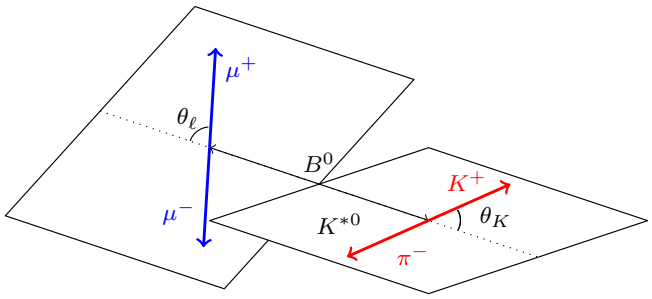
The angle ϕ is the angle between the plane containing the μ^+ and the μ^- and the plane containing the kaon and pion from the K^{*0}.

The definition of the angle ϕ is given by

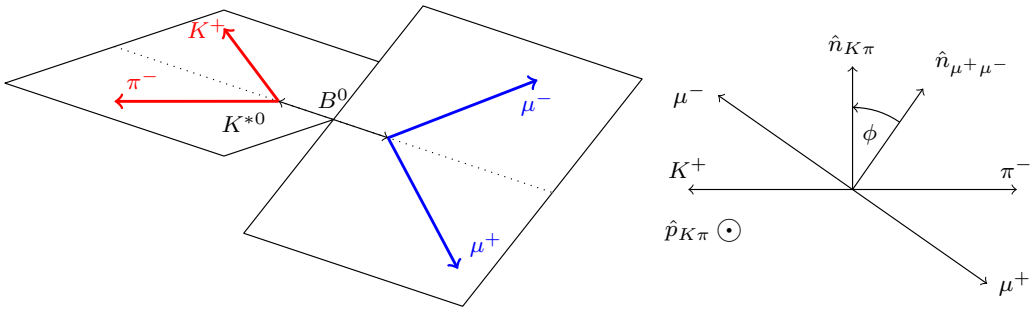
$$\cos\phi = \left(\hat{p}_{\mu^+}^{B^0}\times\hat{p}_{\mu^-}^{B^0}\right)\cdot\left(\hat{p}_{K^+}^{B^0}\times\hat{p}_{\pi^-}^{B^0}\right), \tag{2.20}$$

$$\sin\phi = \left[\left(\hat{p}_{\mu^+}^{B^0}\times\hat{p}_{\mu^-}^{B^0}\right)\cdot\left(\hat{p}_{K^+}^{B^0}\times\hat{p}_{\pi^-}^{B^0}\right)\right]\cdot\hat{p}_{K^{*0}}^{B^0} \tag{2.21}$$

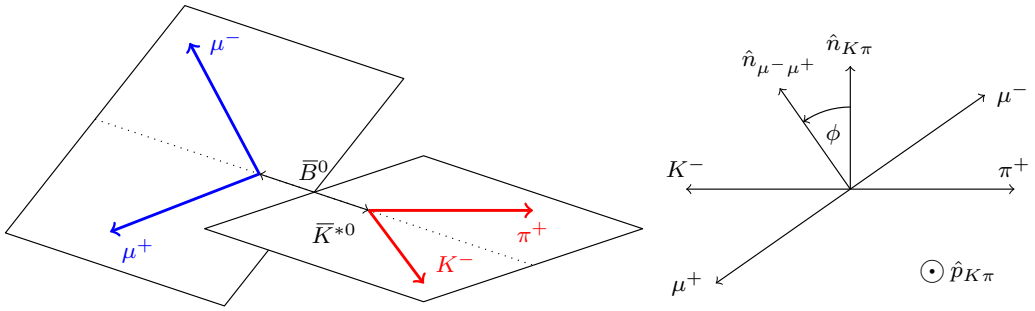
for the B^0 and



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



(b) ϕ definition for the B^0 decay



(c) ϕ definition for the $\overline{B}^0$ decay

Figure 2.3: Schematic of the angular basis used to parameterise $B^0 \to K^{*0}\mu^+\mu^-$ and $\bar{B}^0 \to \bar{K}^{*0}\mu^+\mu^-$ decays [53]. The notation n_{ab} is used to represent the direction normal to the plane containing particles a and b in the B^0 (or $\bar{B}^0$) rest frame.

$$\cos\phi = \left(\hat{p}^{\bar{B}^0}_{\mu^-} \times \hat{p}^{\bar{B}^0}_{\mu^+}\right) \cdot \left(\hat{p}^{\bar{B}^0}_{K^-} \times \hat{p}^{\bar{B}^0}_{\pi^+}\right), \tag{2.22}$$

$$\sin\phi = -\left[\left(\hat{p}^{\bar{B}^0}_{\mu^-} \times \hat{p}^{\bar{B}^0}_{\mu^+}\right) \cdot \left(\hat{p}^{\bar{B}^0}_{K^-} \times \hat{p}^{\bar{B}^0}_{\pi^+}\right)\right] \cdot \hat{p}^{\bar{B}^0}_{K^{*0}} \tag{2.23}$$

for the $\bar{B}^0$ decay. The $\hat{p}^{(Y)}_X$ are unit vectors describing the direction of a particle X in the rest frame of the system Y. In this angular basis, the angular definition for the $\bar{B}^0$ is a CP transformation of that for the B^0 decay.

2.7 Differential decay rate of $B^0 \to K^{*0}\mu^+\mu^-$

The decay $B^0 \to K^{*0}\mu^+\mu^-$, where $K^{*0} \to K^+\pi^-$, is written as a function of the angles $\cos\theta_\ell$, $\cos\theta_K$, and ϕ. The decay rate of $B^0 \to K^{*0}\mu^+\mu^-$ is given by $\Gamma_P(B^0 \to K^{*0}\mu^+\mu^-)$. The differential decay rate can be expressed in terms of the angles and q^2 and is given by $\frac{\mathrm{d}\Gamma_P(B^0\to K^{*0}\mu^+\mu^-)}{\mathrm{d}\cos\theta_\ell \mathrm{d}\cos\theta_K \mathrm{d}\phi \mathrm{d}q^2}$. This can be interpreted as a Probability Density Function (PDF) of $B^0 \to K^{*0}\mu^+\mu^-$ in the angles and q^2.

Since the presence of the K^{*0} is inferred from measuring the $K^+\pi^-$ state, states other than the $K^{*0}(892)$ contribute to the $K^+\pi^-$ spectrum. Within the $m_{K\pi}$ window used in this analysis, which is 100 MeV/c^2 about the K^{*0} pole mass, the state where the $K^+\pi^-$ system has angular momentum of 1, referred to as the P-wave, is dominant. The state where the $K^+\pi^-$ system has angular momentum of 0, referred to as the S-wave, also contributes at a non-negligible level [2] and is therefore also considered in this analysis. Higher order waves are neglected since they are dominant in regions of $m_{K\pi}$ outside the $m_{K\pi}$ window used in this analysis [14]. In addition, as part of Ref. [2], the effect of neglecting a D-wave contribution was computed and found to be negligible. Thus the differential decay rates for the S-wave, P-wave, and interference between the S-wave and P-wave, are measured.

Thus the differential decay rate for $B^0 \to K^{*0}\mu^+\mu^-$ where the $K\pi$ system is in a P-wave

configuration, can be written as [54]

$$\begin{aligned}\frac{\mathrm{d}\Gamma_P(B^0 \to K^{*0}\mu^+\mu^-)}{\mathrm{d}\cos\theta_\ell \mathrm{d}\cos\theta_K \mathrm{d}\phi \mathrm{d}q^2} = \frac{9}{32\pi}\big[&J_{1s}\sin^2\theta_K + J_{1c}\cos^2\theta_\ell + J_{2s}\sin^2\theta_K\cos 2\theta_\ell + J_{2c}\cos^2\theta_K\cos 2\theta_\ell \\ &+J_3\sin^2\theta_K\sin^2\theta_\ell\cos 2\phi + J_4\sin 2\theta_K\sin 2\theta_\ell\cos\phi \\ &+J_5\sin 2\theta_K\sin\theta_\ell\cos\phi + J_{6s}\sin^2\theta_K\cos\theta_\ell \\ &+J_7\sin 2\theta_K\sin\theta_\ell\sin\phi + J_8\sin 2\theta_K\sin 2\theta_\ell\sin\phi \\ &+J_9\sin^2\theta_K\sin^2\theta_\ell\sin 2\phi\big],\end{aligned} \tag{2.24}$$

where $J_i = J_i(q^2)$ are quantities constructed from bilinear combinations of the decay amplitudes.

Likewise, the differential decay rate for the decay $B^0 \to K^{*0}\mu^+\mu^-$, where the $K\pi$ system is in an S-wave configuration, as well as the interference terms, is written as

$$\begin{aligned}\frac{\mathrm{d}\Gamma_S(B^0 \to K^{*0}\mu^+\mu^-)}{\mathrm{d}\cos\theta_\ell \mathrm{d}\cos\theta_K \mathrm{d}\phi \mathrm{d}q^2} = \frac{1}{4\pi}\big[&(\tilde{J}^c_{1a} + \tilde{J}^c_{2a}\cos 2\theta_\ell) \\ &+\tilde{J}^c_{1b}\cos\theta_K + \tilde{J}^c_{2b}\cos 2\theta_\ell\cos\theta_K \\ &+\tilde{J}_4\sin 2\theta_\ell\sin\theta_K\cos\phi + \tilde{J}_5\sin\theta_\ell\sin\theta_K\cos\phi \\ &+\tilde{J}_7\sin\theta_\ell\sin\theta_K\sin\phi + \tilde{J}_8\sin 2\theta_\ell\sin\theta_K\sin\phi\big].\end{aligned} \tag{2.25}$$

The $B^0 \to K^{*0}\mu^+\mu^-$ decay amplitudes are described in terms their real and imaginary components, as well as the three transversity states of the K^{*0} ($\|$, $\perp$ and 0 denoting parallel, transverse and longitudinal polarisations respectively) and the chiralities of the dimuon system (L and R corresponding the dimuon system in a left-handed and right-handed configuration). Following reference [54], in terms of these decay amplitudes, assuming massless muons, and no scalar or tensor contributions (as in the SM), the P-wave J terms are written as

$$
\begin{aligned}
J_{1s} &= \frac{3}{4}\left[|\mathcal{A}_{\parallel}^{\mathrm{L}}|^2 + |\mathcal{A}_{\perp}^{\mathrm{L}}|^2 + |\mathcal{A}_{\parallel}^{\mathrm{R}}|^2 + |\mathcal{A}_{\perp}^{\mathrm{R}}|^2\right] \\
J_{1c} &= |\mathcal{A}_{0}^{\mathrm{L}}|^2 + |\mathcal{A}_{0}^{\mathrm{R}}|^2 \\
J_{2s} &= \frac{1}{4}\left[|\mathcal{A}_{\parallel}^{\mathrm{L}}|^2 + |\mathcal{A}_{\perp}^{\mathrm{L}}|^2 + |\mathcal{A}_{\parallel}^{\mathrm{R}}|^2 + |\mathcal{A}_{\perp}^{\mathrm{R}}|^2\right] \\
J_{2c} &= -|\mathcal{A}_{0}^{\mathrm{L}}|^2 - |\mathcal{A}_{0}^{\mathrm{R}}|^2 \\
J_{3} &= \frac{1}{2}\left[|\mathcal{A}_{\perp}^{\mathrm{L}}|^2 - |\mathcal{A}_{\parallel}^{\mathrm{L}}|^2 + |\mathcal{A}_{\perp}^{\mathrm{R}}|^2 - |\mathcal{A}_{\parallel}^{\mathrm{R}}|^2\right]
\end{aligned}
\qquad
\begin{aligned}
J_{4} &= \sqrt{\frac{1}{2}}\mathrm{Re}(\mathcal{A}_{0}^{\mathrm{L}}\mathcal{A}_{\parallel}^{\mathrm{L}*} + \mathcal{A}_{0}^{\mathrm{R}}\mathcal{A}_{\parallel}^{\mathrm{R}*}) \\
J_{5} &= \sqrt{2}\mathrm{Re}(\mathcal{A}_{0}^{\mathrm{L}}\mathcal{A}_{\perp}^{\mathrm{L}*} - \mathcal{A}_{0}^{\mathrm{R}}\mathcal{A}_{\perp}^{\mathrm{R}*}) \\
J_{6s} &= 2\mathrm{Re}(\mathcal{A}_{\parallel}^{\mathrm{L}}\mathcal{A}_{\perp}^{\mathrm{L}*} - \mathcal{A}_{\parallel}^{\mathrm{R}}\mathcal{A}_{\perp}^{\mathrm{R}*}) \\
J_{7} &= \sqrt{2}\mathrm{Im}(\mathcal{A}_{0}^{\mathrm{L}}\mathcal{A}_{\parallel}^{\mathrm{L}*} - \mathcal{A}_{0}^{\mathrm{R}}\mathcal{A}_{\parallel}^{\mathrm{R}*}) \\
J_{8} &= \sqrt{\frac{1}{2}}\mathrm{Im}(\mathcal{A}_{0}^{\mathrm{L}}\mathcal{A}_{\perp}^{\mathrm{L}*} + \mathcal{A}_{0}^{\mathrm{R}}\mathcal{A}_{\perp}^{\mathrm{R}*}) \\
J_{9} &= \mathrm{Im}(\mathcal{A}_{\parallel}^{\mathrm{L}*}\mathcal{A}_{\perp}^{\mathrm{L}} + \mathcal{A}_{\parallel}^{\mathrm{R}*}\mathcal{A}_{\perp}^{\mathrm{R}}) \quad .
\end{aligned}
\tag{2.26}
$$

One can also write the S-wave amplitudes (denoted by the subscript 00) in terms of their real and imaginary components and the chiralities of the dimuon system. In terms of the amplitudes, the S-wave and interference J terms are written as

$$
\begin{aligned}
\tilde{J}_{1a}^{c} &= \frac{3}{8}\left[|\mathcal{A}_{00}^{\mathrm{L}}|^2 + |\mathcal{A}_{00}^{\mathrm{R}}|^2\right] \\
\tilde{J}_{2a}^{c} &= -\frac{3}{8}\left[|\mathcal{A}_{00}^{\mathrm{L}}|^2 + |\mathcal{A}_{00}^{\mathrm{R}}|^2\right] \\
\tilde{J}_{1b}^{c} &= \sqrt{\frac{27}{16}}\mathrm{Re}(\mathcal{A}_{00}^{\mathrm{L}}\mathcal{A}_{0}^{\mathrm{L}*} + \mathcal{A}_{00}^{\mathrm{R}}\mathcal{A}_{0}^{\mathrm{R}*}) \\
\tilde{J}_{2b}^{c} &= -\sqrt{\frac{27}{16}}\mathrm{Re}(\mathcal{A}_{00}^{\mathrm{L}}\mathcal{A}_{0}^{\mathrm{L}*} + \mathcal{A}_{00}^{\mathrm{R}}\mathcal{A}_{0}^{\mathrm{R}*})
\end{aligned}
\qquad
\begin{aligned}
\tilde{J}_{4} &= \sqrt{\frac{27}{32}}\mathrm{Re}(\mathcal{A}_{00}^{\mathrm{L}}\mathcal{A}_{\parallel}^{\mathrm{L}*} + \mathcal{A}_{00}^{\mathrm{R}}\mathcal{A}_{\parallel}^{\mathrm{R}*}) \\
\tilde{J}_{5} &= \sqrt{\frac{27}{8}}\mathrm{Re}(\mathcal{A}_{00}^{\mathrm{L}}\mathcal{A}_{\perp}^{\mathrm{L}*} - \mathcal{A}_{00}^{\mathrm{R}}\mathcal{A}_{\perp}^{\mathrm{R}*}) \\
\tilde{J}_{7} &= \sqrt{\frac{27}{8}}\mathrm{Im}(\mathcal{A}_{00}^{\mathrm{L}}\mathcal{A}_{\parallel}^{\mathrm{L}*} - \mathcal{A}_{00}^{\mathrm{R}}\mathcal{A}_{\parallel}^{\mathrm{R}*}) \\
\tilde{J}_{8} &= \sqrt{\frac{27}{32}}\mathrm{Im}(\mathcal{A}_{00}^{\mathrm{L}}\mathcal{A}_{\perp}^{\mathrm{L}*} + \mathcal{A}_{00}^{\mathrm{R}}\mathcal{A}_{\perp}^{\mathrm{R}*}).
\end{aligned}
\tag{2.27}
$$

Thus, combining the P-wave, S-wave, and interference, the differential decay rate of $B^0 \to K^{*0}\mu^+\mu^-$ is given by

$$
\frac{\mathrm{d}^4\Gamma(B^0 \to K^{*0}\mu^+\mu^-)}{\mathrm{d}\cos\theta_\ell \mathrm{d}\cos\theta_K \mathrm{d}\phi \mathrm{d}q^2} = \frac{\mathrm{d}\Gamma_P(B^0 \to K^{*0}\mu^+\mu^-)}{\mathrm{d}\cos\theta_\ell \mathrm{d}\cos\theta_K \mathrm{d}\phi \mathrm{d}q^2} + \frac{\mathrm{d}\Gamma_S(B^0 \to K^{*0}\mu^+\mu^-)}{\mathrm{d}\cos\theta_\ell \mathrm{d}\cos\theta_K \mathrm{d}\phi \mathrm{d}q^2}. \tag{2.28}
$$

2.8 CP-averages and asymmetries

For this analysis, the amplitudes $\mathcal{A}$ are measured. However one can define a set of observables [54] corresponding to the CP-averages for the P-wave and the S-wave by

$$
\begin{aligned}
S_i &= \frac{\left(J_i + \bar{J}_i\right)}{\left(\frac{\mathrm{d}\Gamma_P}{\mathrm{d}q^2} + \frac{\mathrm{d}\bar{\Gamma}_P}{\mathrm{d}q^2}\right)}, \text{ the P-wave only observables} \\
F_S &= \frac{\left(\frac{\mathrm{d}\Gamma_S}{\mathrm{d}q^2} + \frac{\mathrm{d}\bar{\Gamma}_S}{\mathrm{d}q^2}\right)}{\left(\frac{\mathrm{d}\Gamma_S}{\mathrm{d}q^2} + \frac{\mathrm{d}\Gamma_P}{\mathrm{d}q^2} + \frac{\mathrm{d}\bar{\Gamma}_S}{\mathrm{d}q^2} + \frac{\mathrm{d}\bar{\Gamma}_P}{\mathrm{d}q^2}\right)} \\
&= \frac{8}{3}\frac{\tilde{J}^c_{1a} + \bar{\tilde{J}}^c_{1a}}{\left(\frac{\mathrm{d}\Gamma_S}{\mathrm{d}q^2} + \frac{\mathrm{d}\Gamma_P}{\mathrm{d}q^2} + \frac{\mathrm{d}\bar{\Gamma}_S}{\mathrm{d}q^2} + \frac{\mathrm{d}\bar{\Gamma}_P}{\mathrm{d}q^2}\right)}, \text{ the fraction of S-wave.}
\end{aligned}
\tag{2.29}
$$

Here $\frac{\mathrm{d}\Gamma_{S/P}}{\mathrm{d}q^2}$ are the decay rates for the S-wave and P-wave, differential in q^2, i.e.

$$
\begin{aligned}
\frac{\mathrm{d}\Gamma_S}{\mathrm{d}q^2} &= \int_{-\pi}^{\pi}\int_{-1}^{+1}\int_{-1}^{+1} \frac{\mathrm{d}\Gamma_S(B^0 \to K^{*0}\mu^+\mu^-)}{\mathrm{d}\cos\theta_\ell \mathrm{d}\cos\theta_K \mathrm{d}\phi \mathrm{d}q^2}\ \mathrm{d}\cos\theta_\ell\ \mathrm{d}\cos\theta_K\ \mathrm{d}\phi \\
\frac{\mathrm{d}\bar{\Gamma}_S}{\mathrm{d}q^2} &= \int_{-\pi}^{\pi}\int_{-1}^{+1}\int_{-1}^{+1} \frac{\mathrm{d}\Gamma_S(\bar{B}^0 \to \bar{K}^{*0}\mu^+\mu^-)}{\mathrm{d}\cos\theta_\ell \mathrm{d}\cos\theta_K \mathrm{d}\phi \mathrm{d}q^2}\ \mathrm{d}\cos\theta_\ell\ \mathrm{d}\cos\theta_K\ \mathrm{d}\phi \\
\frac{\mathrm{d}\Gamma_P}{\mathrm{d}q^2} &= \int_{-\pi}^{\pi}\int_{-1}^{+1}\int_{-1}^{+1} \frac{\mathrm{d}\Gamma_P(B^0 \to K^{*0}\mu^+\mu^-)}{\mathrm{d}\cos\theta_\ell \mathrm{d}\cos\theta_K \mathrm{d}\phi \mathrm{d}q^2}\ \mathrm{d}\cos\theta_\ell\ \mathrm{d}\cos\theta_K\ \mathrm{d}\phi \\
\frac{\mathrm{d}\bar{\Gamma}_P}{\mathrm{d}q^2} &= \int_{-\pi}^{\pi}\int_{-1}^{+1}\int_{-1}^{+1} \frac{\mathrm{d}\Gamma_P(\bar{B}^0 \to \bar{K}^{*0}\mu^+\mu^-)}{\mathrm{d}\cos\theta_\ell \mathrm{d}\cos\theta_K \mathrm{d}\phi \mathrm{d}q^2}\ \mathrm{d}\cos\theta_\ell\ \mathrm{d}\cos\theta_K\ \mathrm{d}\phi.
\end{aligned}
\tag{2.30}
$$

The CP-averaged interference terms are defined as

$$
\begin{aligned}
S_{S1} &= \frac{8}{3}\frac{\tilde{J}^c_{1b} + \bar{\tilde{J}}^c_{1b}}{\left(\frac{\mathrm{d}\Gamma_S}{\mathrm{d}q^2} + \frac{\mathrm{d}\Gamma_P}{\mathrm{d}q^2} + \frac{\mathrm{d}\bar{\Gamma}_S}{\mathrm{d}q^2} + \frac{\mathrm{d}\bar{\Gamma}_P}{\mathrm{d}q^2}\right)} \\
S_{S2} &= \frac{4}{3}\frac{\tilde{J}_4 + \bar{\tilde{J}}_4}{\left(\frac{\mathrm{d}\Gamma_S}{\mathrm{d}q^2} + \frac{\mathrm{d}\Gamma_P}{\mathrm{d}q^2} + \frac{\mathrm{d}\bar{\Gamma}_S}{\mathrm{d}q^2} + \frac{\mathrm{d}\bar{\Gamma}_P}{\mathrm{d}q^2}\right)} \\
S_{S3} &= \frac{4}{3}\frac{\tilde{J}_5 + \bar{\tilde{J}}_5}{\left(\frac{\mathrm{d}\Gamma_S}{\mathrm{d}q^2} + \frac{\mathrm{d}\Gamma_P}{\mathrm{d}q^2} + \frac{\mathrm{d}\bar{\Gamma}_S}{\mathrm{d}q^2} + \frac{\mathrm{d}\bar{\Gamma}_P}{\mathrm{d}q^2}\right)} \\
S_{S4} &= \frac{4}{3}\frac{\tilde{J}_7 + \bar{\tilde{J}}_7}{\left(\frac{\mathrm{d}\Gamma_S}{\mathrm{d}q^2} + \frac{\mathrm{d}\Gamma_P}{\mathrm{d}q^2} + \frac{\mathrm{d}\bar{\Gamma}_S}{\mathrm{d}q^2} + \frac{\mathrm{d}\bar{\Gamma}_P}{\mathrm{d}q^2}\right)} \\
S_{S5} &= \frac{4}{3}\frac{\tilde{J}_8 + \bar{\tilde{J}}_8}{\left(\frac{\mathrm{d}\Gamma_S}{\mathrm{d}q^2} + \frac{\mathrm{d}\Gamma_P}{\mathrm{d}q^2} + \frac{\mathrm{d}\bar{\Gamma}_S}{\mathrm{d}q^2} + \frac{\mathrm{d}\bar{\Gamma}_P}{\mathrm{d}q^2}\right)}.
\end{aligned}
\tag{2.31}
$$

The fraction of longitudinal polarisation of the K^{*0}, $F_L = S_{1c}$. The forward-backward asymmetry of the dimuon system is defined as $A_{FB} = \frac{3}{4}S_{6s}$. From Equation 2.26, $S_{2c} = -F_L$, and $S_{2s} = 3S_{1s} = \frac{1}{4}(1 - F_L)$. Thus the set of P-wave observables obtained from the amplitudes, without any degeneracies, is $\{F_L, S_3, S_4, S_5, A_{FB}, S_7, S_8, S_9\}$.

One can define 'form-factor independent' observables, where theoretical uncertainties

due to local contributions (form-factors) cancel to first order. Following e.g. [53], [55], and [56], these are defined as

$$\begin{aligned} P_1 &= \frac{2S_3}{1-F_L} & P_4' &= \frac{S_4}{\sqrt{F_L(1-F_L)}} \\ P_2 &= \frac{2A_{FB}}{3(1-F_L)} & P_5' &= \frac{S_5}{\sqrt{F_L(1-F_L)}} \\ P_3 &= \frac{-S_9}{1-F_L} & P_6' &= \frac{S_7}{\sqrt{F_L(1-F_L)}} \\ & & P_8' &= \frac{S_8}{\sqrt{F_L(1-F_L)}} \end{aligned} \tag{2.32}$$

The CP-asymmetries can be analogously defined by

$$\begin{aligned} A_i &= \frac{\left(J_i - \bar{J}_i\right)}{\left(\frac{\mathrm{d}\Gamma_P}{\mathrm{d}q^2} + \frac{\mathrm{d}\bar{\Gamma}_P}{\mathrm{d}q^2}\right)} & A_{S2} &= \frac{4}{3}\frac{\tilde{J}_4 - \bar{\tilde{J}}_4}{\left(\frac{\mathrm{d}\Gamma_S}{\mathrm{d}q^2} + \frac{\mathrm{d}\Gamma_P}{\mathrm{d}q^2} + \frac{\mathrm{d}\bar{\Gamma}_S}{\mathrm{d}q^2} + \frac{\mathrm{d}\bar{\Gamma}_P}{\mathrm{d}q^2}\right)} \\ AF_S &= \frac{\left(\frac{\mathrm{d}\Gamma_S}{\mathrm{d}q^2} - \frac{\mathrm{d}\bar{\Gamma}_S}{\mathrm{d}q^2}\right)}{\left(\frac{\mathrm{d}\Gamma_S}{\mathrm{d}q^2} + \frac{\mathrm{d}\Gamma_P}{\mathrm{d}q^2} + \frac{\mathrm{d}\bar{\Gamma}_S}{\mathrm{d}q^2} + \frac{\mathrm{d}\bar{\Gamma}_P}{\mathrm{d}q^2}\right)} & A_{S3} &= \frac{4}{3}\frac{\tilde{J}_5 - \bar{\tilde{J}}_5}{\left(\frac{\mathrm{d}\Gamma_S}{\mathrm{d}q^2} + \frac{\mathrm{d}\Gamma_P}{\mathrm{d}q^2} + \frac{\mathrm{d}\bar{\Gamma}_S}{\mathrm{d}q^2} + \frac{\mathrm{d}\bar{\Gamma}_P}{\mathrm{d}q^2}\right)} \\ &= \frac{8}{3}\frac{\tilde{J}^c_{1a} - \bar{\tilde{J}}^c_{1a}}{\left(\frac{\mathrm{d}\Gamma_S}{\mathrm{d}q^2} + \frac{\mathrm{d}\Gamma_P}{\mathrm{d}q^2} + \frac{\mathrm{d}\bar{\Gamma}_S}{\mathrm{d}q^2} + \frac{\mathrm{d}\bar{\Gamma}_P}{\mathrm{d}q^2}\right)} & A_{S4} &= \frac{4}{3}\frac{\tilde{J}_7 - \bar{\tilde{J}}_7}{\left(\frac{\mathrm{d}\Gamma_S}{\mathrm{d}q^2} + \frac{\mathrm{d}\Gamma_P}{\mathrm{d}q^2} + \frac{\mathrm{d}\bar{\Gamma}_S}{\mathrm{d}q^2} + \frac{\mathrm{d}\bar{\Gamma}_P}{\mathrm{d}q^2}\right)} \\ A_{S1} &= \frac{8}{3}\frac{\tilde{J}^c_{1b} - \bar{\tilde{J}}^c_{1b}}{\left(\frac{\mathrm{d}\Gamma_S}{\mathrm{d}q^2} + \frac{\mathrm{d}\Gamma_P}{\mathrm{d}q^2} + \frac{\mathrm{d}\bar{\Gamma}_S}{\mathrm{d}q^2} + \frac{\mathrm{d}\bar{\Gamma}_P}{\mathrm{d}q^2}\right)} & A_{S5} &= \frac{4}{3}\frac{\tilde{J}_8 - \bar{\tilde{J}}_8}{\left(\frac{\mathrm{d}\Gamma_S}{\mathrm{d}q^2} + \frac{\mathrm{d}\Gamma_P}{\mathrm{d}q^2} + \frac{\mathrm{d}\bar{\Gamma}_S}{\mathrm{d}q^2} + \frac{\mathrm{d}\bar{\Gamma}_P}{\mathrm{d}q^2}\right)}. \end{aligned} \tag{2.33}$$

2.9 Including $m_{K\pi}$

The above treatment of the amplitudes assumes the K^{*0} is infinitely narrow. In reality the K^{*0} has a non-zero width and so the angular distribution of $B^0 \to K^{*0}\mu^+\mu^-$ is a function of $m_{K\pi}$. Thus the differential decay distribution as described in Equation 2.28 does not describe the genuine distribution since it does not include the $m_{K\pi}$ dependence. Therefore the decay amplitudes depend on $m_{K\pi}$ as well as q^2, i.e. $\mathcal{A} = \mathcal{A}(m_{K\pi}, q^2)$. This subsection describes the models used to describe the $m_{K\pi}$ lineshape. Following e.g. [54], factorisation between $m_{K\pi}$ and q^2 is assumed, i.e.

$$\mathcal{A}(m_{K\pi}, q^2) = \mathcal{A}(m_{K\pi})\mathcal{A}(q^2). \tag{2.34}$$

The differential decay distribution is rewritten to explicitly contain the $m_{K\pi}$ by applying the transformation

$$\begin{aligned} \mathcal{A}_P &\to \mathcal{A}_P \mathcal{A}_{BW} \\ \mathcal{A}_S &\to \mathcal{A}_S \mathcal{A}_{LASS} \end{aligned} \tag{2.35}$$

to the P-wave amplitudes $\mathcal{A}_P$, and the S-wave amplitudes $\mathcal{A}_S$. Here, the P-wave and S-wave lineshapes are parameterised by the Breit-Wigner ($\mathcal{A}_{BW}$) and LASS ($\mathcal{A}_{LASS}$) parameterisations respectively. Thus the 4D differential decay rate as in Equation 2.28 becomes a 5D differential decay rate, i.e.

$$\frac{\mathrm{d}^4\Gamma(B^0 \to K^{*0}\mu^+\mu^-)}{\mathrm{d}\cos\theta_\ell \mathrm{d}\cos\theta_K \mathrm{d}\phi \mathrm{d}q^2} \to \frac{\mathrm{d}^5\Gamma(B^0 \to K^{*0}\mu^+\mu^-)}{\mathrm{d}\cos\theta_\ell \mathrm{d}\cos\theta_K \mathrm{d}\phi \mathrm{d}q^2 \mathrm{d}m_{K\pi}}. \tag{2.36}$$

The P-wave $m_{K\pi}$ lineshape is given by

$$\mathcal{A}_{\mathrm{BW}}(m_{K\pi}, q^2) = \frac{\sqrt{ph}(\frac{p}{p_0})^{L_{B^0}}(\frac{h}{h_0})^{L_{K\pi}} B'_{L_B}(p, p_0) B'_{L_{K\pi}}(h, h_0)}{m^2_{K^{*0}} - m^2_{K\pi} - i m_{K^{*0}} \Gamma(m_{K\pi})} \tag{2.37}$$

where p is the momentum of the $K\pi$ system in the B^0 rest frame and h is the momentum of the kaon in the $K\pi$ rest frame. p_0 and h_0 are the equivalent quantities, evaluated at the resonance peak. The spins of the $K\pi$ system and the B^0 are given by $L_{K\pi} = 1$ and $L_{B^0} = 0$ respectively. The pole mass of the K^{*0} is given by $m_{K^{*0}}$.

The Blatt-Weisskopf barrier factors, B'_L, are used to factor in the production and decay of the resonance. These are given by

$$B'_0(h, h_0) = 1 \qquad\qquad B'_1(h, h_0) = \sqrt{\frac{1 + (h_0 d)^2}{1 + (hd)^2}}, \tag{2.38}$$

where $d = 1.6$ GeV^{-1} as used by [57]. Finally the decay width $\Gamma(m_{K\pi})$ is given by

$$\Gamma(m_{K\pi}) = \Gamma(m_{K^{*0}}) \left(\frac{h}{h_0}\right)^{(2L_{K\pi}+1)} \frac{m_{K^{*0}}}{m_{K\pi}} \left(B'_{L_{K\pi}}(h, h_0)\right)^2. \tag{2.39}$$

The S-wave is described by the LASS parameterisation [58], $\mathcal{A}_{\mathrm{LASS}}(m_{K\pi}, q^2)$, given by

$$\begin{aligned} \mathcal{A}_{\mathrm{LASS}}(m_{K\pi}, q^2) = &\left[\sqrt{ph}\left(\frac{p}{p_0}\right)^{L_B^0}\left(\frac{h}{h_0}\right)^{L_{K\pi}} B'_{L_B}(p, p_0) B'_{L_{K\pi}}(h, h_0)\right] \times \\ &\left(\frac{m_{K\pi}}{\cot\delta_B - i} + \frac{e^{2i\delta_B}}{\cot\delta_R - i}\right) \end{aligned} \tag{2.40}$$

where $L_{B^0} = 1$ and $L_{K\pi} = 0$. The non-resonant component of the LASS parameterisastion is described by $\cot_{\delta_B}$, which is defined as

$$\cot\delta_B = \frac{1}{ah} + \frac{rh}{2}. \tag{2.41}$$

Here r and a are fixed to 1.7 GeV^{-1} and 2.48 GeV^{-1} respectively [59] [60].

The $m_{K\pi}$ window chosen for this analysis is $0.796 < m_{K\pi} < 0.996$ GeV/c^2. Introducing the $m_{K\pi}$ dependence in the four-differential distribution, then integrating over $m_{K\pi}$, one has

$$\begin{aligned}\frac{\mathrm{d}^4\Gamma(B^0\to K^{*0}\mu^+\mu^-)}{\mathrm{d}\cos\theta_\ell\mathrm{d}\cos\theta_K\mathrm{d}\phi\mathrm{d}q^2} =& \frac{\mathrm{d}\Gamma_P(B^0\to K^{*0}\mu^+\mu^-)}{\mathrm{d}\cos\theta_\ell\mathrm{d}\cos\theta_K\mathrm{d}\phi\mathrm{d}q^2}\int_{0.796}^{0.996}|\mathcal{A}_{\mathrm{BW}}(m_{K\pi},q^2)|^2\mathrm{d}m_{K\pi}\\ &+\frac{1}{4\pi}\left[(\tilde{J}^c_{1a}+\tilde{J}^c_{2a}\cos 2\theta_\ell)\int_{0.796}^{0.996}|\mathcal{A}_{\mathrm{LASS}}(m_{K\pi},q^2)|^2\mathrm{d}m_{K\pi}\right.\\ &+\sqrt{\frac{27}{16}}\mathrm{Re}\Big((\mathcal{A}^{\mathrm{L}}_{00}\mathcal{A}^{\mathrm{L}*}_0+\mathcal{A}^{\mathrm{R}}_{00}\mathcal{A}^{\mathrm{R}*}_0)\int_{0.796}^{0.996}\mathcal{A}_{\mathrm{LASS}}(m_{K\pi},q^2)\mathcal{A}^*_{\mathrm{BW}}(m_{K\pi},q^2)\mathrm{d}m_{K\pi}\Big)\cos\theta_K\\ &+\text{ other interference terms.}\end{aligned} \tag{2.42}$$

Regarding the normalisation of the $m_{K\pi}$ lineshapes, consider integrating Equation 2.42 over $\cos\theta_\ell$, $\cos\theta_K$, and ϕ. The only terms which remain are J_{1s}, J_{1c}, J_{2s}, J_{2c}, $\tilde{J}^c_{1a}$, and $\tilde{J}^c_{2a}$. Thus after simplifying, one has

$$\begin{aligned}\frac{\mathrm{d}\Gamma(B^0\to K^{*0}\mu^+\mu^-)}{\mathrm{d}q^2} =& \Big(|\mathcal{A}_\parallel|^2+|\mathcal{A}_\perp|^2+|\mathcal{A}_0|^2\Big)\int_{0.796}^{0.996}|\mathcal{A}_{\mathrm{BW}}(m_{K\pi},q^2)|^2\mathrm{d}m_{K\pi}\\ &+|\mathcal{A}_{00}|^2\int_{0.796}^{0.996}|\mathcal{A}_{\mathrm{LASS}}(m_{K\pi},q^2)|^2\mathrm{d}m_{K\pi}.\end{aligned} \tag{2.43}$$

However, it is required that

$$\begin{aligned}\frac{\mathrm{d}\Gamma(B^0\to K^{*0}\mu^+\mu^-)}{\mathrm{d}q^2} &= \frac{\mathrm{d}\Gamma_P(B^0\to K^{*0}\mu^+\mu^-)}{\mathrm{d}q^2}+\frac{\mathrm{d}\Gamma_S(B^0\to K^{*0}\mu^+\mu^-)}{\mathrm{d}q^2}\\ &= |\mathcal{A}_\parallel|^2+|\mathcal{A}_\perp|^2+|\mathcal{A}_0|^2+|\mathcal{A}_{00}|^2.\end{aligned} \tag{2.44}$$

In order for this to be satisfied, the $m_{K\pi}$ lineshapes are thus normalised to the $m_{K\pi}$ window

used in this analysis, in other words

$$\begin{aligned}\int_{0.796}^{0.996} |\mathcal{A}_{\rm BW}(m_{K\pi}, q^2)|^2 {\rm d}m_{K\pi} = 1 \\ \int_{0.796}^{0.996} |\mathcal{A}_{\rm LASS}(m_{K\pi}, q^2)|^2 {\rm d}m_{K\pi} = 1.\end{aligned} \tag{2.45}$$

This is enforced by redefining the $m_{K\pi}$ lineshapes normalised to the square root of the integrals of the lineshapes squared over the $m_{K\pi}$ window,

$$\begin{aligned}\mathcal{A}_{\rm BW}(m_{K\pi}, q^2) \to \frac{\mathcal{A}_{\rm BW}(m_{K\pi}, q^2)}{\sqrt{\int_{0.796}^{0.996} |\mathcal{A}_{\rm BW}(m_{K\pi}, q^2)|^2 {\rm d}m_{K\pi}}} \\ \mathcal{A}_{\rm LASS}(m_{K\pi}, q^2) \to \frac{\mathcal{A}_{\rm LASS}(m_{K\pi}, q^2)}{\sqrt{\int_{0.796}^{0.996} |\mathcal{A}_{\rm LASS}(m_{K\pi}, q^2)|^2 {\rm d}m_{K\pi}}}.\end{aligned} \tag{2.46}$$

2.10 Symmetries of the angular distribution

As described by [61], the eight complex decay amplitudes ($\mathcal{A}_\parallel^{\rm L}$, $\mathcal{A}_\parallel^{\rm R}$, $\mathcal{A}_\perp^{\rm L}$, $\mathcal{A}_\perp^{\rm R}$, $\mathcal{A}_0^{\rm L}$, $\mathcal{A}_0^{\rm R}$, $\mathcal{A}_{00}^{\rm L}$, $\mathcal{A}_{00}^{\rm R}$) can be arranged into four complex vectors

$$n_\parallel = \begin{pmatrix} \mathcal{A}_\parallel^{\rm L} \\ \mathcal{A}_\parallel^{\rm R*} \end{pmatrix}, \quad n_\perp = \begin{pmatrix} \mathcal{A}_\perp^{\rm L} \\ -\mathcal{A}_\perp^{\rm R*} \end{pmatrix}, \quad n_0 = \begin{pmatrix} \mathcal{A}_0^{\rm L} \\ \mathcal{A}_0^{\rm R*} \end{pmatrix}, \quad n_{00} = \begin{pmatrix} \mathcal{A}_{00}^{\rm L} \\ \mathcal{A}_{00}^{\rm R*} \end{pmatrix}. \tag{2.47}$$

Certain transformations can be applied to these amplitude vectors leaving the differential decay rate invariant. These transformations can be written as $n_i \to U_{ij} n_j \equiv n'_j$.

The differential decay rate as written as Equation 2.28 obeys four symmetries. The symmetry transformations can be written as

$$n'_i = U n_i = \begin{pmatrix} e^{i\phi_L} & 0 \\ 0 & e^{-i\phi_R} \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cosh i\eta & -\sinh i\eta \\ -\sinh i\eta & \cosh i\eta \end{pmatrix} n_i, \tag{2.48}$$

where U is a matrix which defines the symmetries. These are two global phases, ϕ_L and ϕ_R, a rotation θ about the real and imaginary components of the amplitudes, and a rotation η which mixes the real and imaginary components of the amplitudes [61]. Thus one can rotate the amplitude vectors by four angles as defined in Equation 2.48 with the angular

distribution (Equation 2.28) invariant under this rotation. Thus the amplitude basis is not unique.

Introducing $m_{K\pi}$ into the differential decay rate results in a breaking of two symmetries. This occurs in the interference terms, such as $\tilde{J}^c_{1b} \propto \mathrm{Re}(\mathcal{A}^{\mathrm{L}}_{00}\mathcal{A}^{\mathrm{L}*}_{0} + \mathcal{A}^{\mathrm{R}}_{00}\mathcal{A}^{\mathrm{R}*}_{0})$. Here this J-term obeys the symmetries described in Equation 2.48. After applying the transformations in Equation 2.35, the term $\mathrm{Im}(\mathcal{A}^{\mathrm{L}}_{00}\mathcal{A}^{\mathrm{L}*}_{0} + \mathcal{A}^{\mathrm{R}}_{00}\mathcal{A}^{\mathrm{R}*}_{0})$ appears, which violates two of the symmetries described in Equation 2.48 and is only invariant under the rotations about the phases ϕ_L for the left-handed amplitudes and $-\phi_R$ for the right-handed amplitudes. Thus the differential decay rate as a function of $m_{K\pi}$ obeys the symmetries described by

$$n'_i = Un_i = \begin{pmatrix} e^{i\phi_L} & 0 \\ 0 & e^{-i\phi_R} \end{pmatrix} n_i. \tag{2.49}$$

Considering the differential decay rate integrated over $m_{K\pi}$, if the K^{*0} has a narrow width, then the P-wave and S-wave lineshapes will be delta-functions, thus the $m_{K\pi}$ integrals are exactly equal to one. However, because the $m_{K\pi}$ lineshapes for the P-wave and the S-wave are not narrow, there is an interference between these lineshapes. In other words, the interference integral

$$\int_{0.796}^{0.996} \mathcal{A}_{\mathrm{LASS}}(m_{K\pi}, q^2)\mathcal{A}^*_{\mathrm{BW}}(m_{K\pi}, q^2)\mathrm{d}m_{K\pi} \tag{2.50}$$

is not equal to one. This technically results in the same symmetry-breaking terms in the differential decay rate being introduced as in the case where the differential decay rate is a function of $m_{K\pi}$. The effect of the symmetry breaking on the P-wave observables after integrating over $m_{K\pi}$ due to this complex-valued integral was studied on pseudoexperiments and found to be negligible.

With regards to discrete symmetries of the differential decay rate, there is a symmetry where all amplitudes can be rotated by a phase of π, i.e.

$$\mathcal{A} \to -\mathcal{A}, \tag{2.51}$$

In the case of the differential decay rate which assumes massless muons and no scalars

or tensors, there are two discrete symmetries since there are no terms of the form $\mathcal{A}_L\mathcal{A}_R$. Thus the left handed and/or the right handed amplitudes can be rotated by a phase π and the differential decay rate invariant under this sign flip, so there are two discrete symmetries in the differential decay rate. There are therefore four sets of amplitudes which result in identical differential decay rates based on these discrete symmetries: $(\mathcal{A}_L, \mathcal{A}_R)$, $(\mathcal{A}_L, -\mathcal{A}_R)$, $(-\mathcal{A}_L, \mathcal{A}_R)$, and $(-\mathcal{A}_L, -\mathcal{A}_R)$. As these symmetries are not physically meaningful, and the sets of amplitudes are identical apart from sign flips, it is sufficient to parameterise only one set in order to describe the differential decay rate.

Chapter 3

The LHCb experiment

This chapter describes the detector, trigger, and data flow of the LHCb experiment. Section 3.1 describes CERN and the Large Hadron Collider (LHC). LHCb, an experiment based at the LHC, is described in Section 3.2. The LHCb trigger system is described in Section 3.3. The LHCb upgrade during Long Shutdown 2 is described in Section 3.4 and the data flow is discussed in Section 3.5.

3.1 The Large Hadron Collider

The European Organization for Nuclear Research (CERN) ia a particle physics laboratory based in Geneva, Switzerland. It runs the Large Hadron Collider (the LHC), the largest and highest energy particle accelerator in the world.

As described in [62], the LHC is a hadron accelerator and collider with 26.7 km circumference. The tunnel used to house the LHC, formerly used for the Large Electron Positron collider (LEP), has a depth which varies between 45 m and 170 m. It contains two beam pipes, with bunches of hadrons being accelerated in opposite directions. These interact at four points, where the four large experiments (ATLAS, CMS, ALICE, LHCb) are located. Most of the LHC physics programme is dedicated to proton-proton collisions. There are also heavy ion collisions, which are not discussed here.

The LHC uses superconducting magnets utilising NbTi and cooled down using superfluid helium to 1.9 K. Dipole magnets with peak field 8.33 T are used for bending and quadrupole magnets are used for focusing the beam. Sextupoles and octupoles are used for corrections

such as the beam chromaticity and Landau damping [63].

There have been two main Run periods of the LHC, Run 1 (2009-2013) and Run 2 (2015-2018). Between Run 1 and Run 2 the LHC and experiments were upgraded. During 2019-2021 the LHC and experiments were shut down for further upgrades. The latest Run period (Run 3) started in 2022. The analysis described in this thesis uses data collected in 2011, 2012, 2016, 2017, and 2018.

Each proton beam consists of 2080 bunches, each with spacing of 25 ns (apart from 50 ns in Run 1), thus the collision rate is 40 MHz (20 MHz in Run 1). The maximum intensity of the beam is 1.15×10^{11} protons per bunch. The centre of mass energy $\sqrt{s}$ of the proton-proton collisions was 7 TeV for 2011, 8 TeV for 2012 and 13 TeV for Run 2.

3.2 The LHCb experiment

At high energies beauty hadrons are primarily produced in the forward and backward directions. This is shown in Figure 3.1 [64], which shows the $b\bar{b}$ production cross-sections as a function of angles. Here, proton-proton interactions are simulated via $\bar{q}q \to \bar{b}b$, $gg \to \bar{b}b$, $\bar{q}q \to \bar{b}bg$ (where $q \neq b$), $\bar{b}b \to \bar{b}bg$, and $gg \to \bar{b}bg$ at collision energy $\sqrt{s} = 14$ TeV. Most $b\bar{b}$ pairs are produced along the beam direction.

The LHCb experiment is primarily designed to measure properties associated to beauty hadrons. It is a forward-arm spectrometer based at Point 8 of the LHC at CERN. Referring back to Figure 3.1, the detector geometry of LHCb is chosen such that it can capture a large number of $b\bar{b}$ pairs (the LHCb acceptance is shown in red).

A schematic of the detector is shown in Figure 3.2 [65]. LHCb employs a right-handed coordinate system, where the origin is located in the Vertex Locator. The z axis traverses along beam and the y axis is along the vertical direction.

The proton-proton interactions occur inside the VErtex LOcator (VELO) [66] [67]. The VELO is a silicon microstrip detector, consisting of 84 modules. These modules are in two halves (42 modules each) and are arranged perpendicular to the beamline. During physics data-taking, the modules are 8 mm away from the beamline. The main role of the VELO is to measure and distinguish between primary vertices (proton-proton interaction points) and secondary vertices (locations of the decay vertices). These secondary verticies

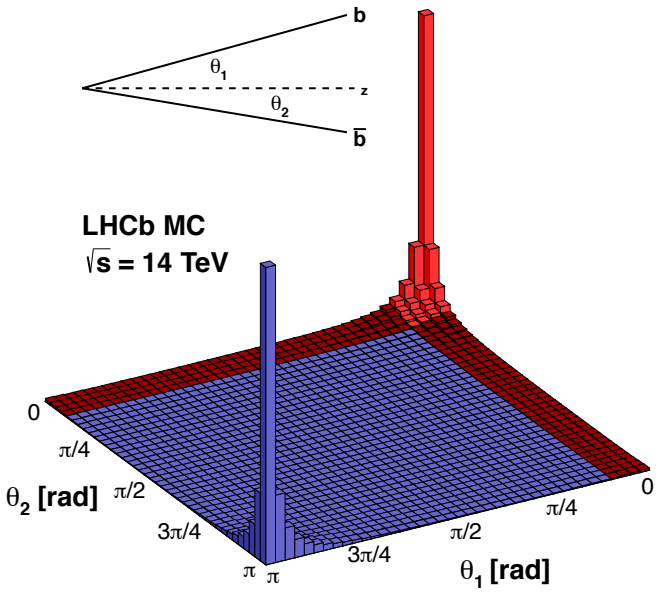


Figure 3.1: Plot of the $b\bar{b}$ production cross-sections at the LHC, as a function of angle between the b and beam direction in the lab frame (θ_1), and angle between the $\bar{b}$ and beam direction in the lab frame (θ_2) [64]. Most $b\bar{b}$ pairs are produced along the beam direction. The area shaded red is the region which is captured by LHCb.

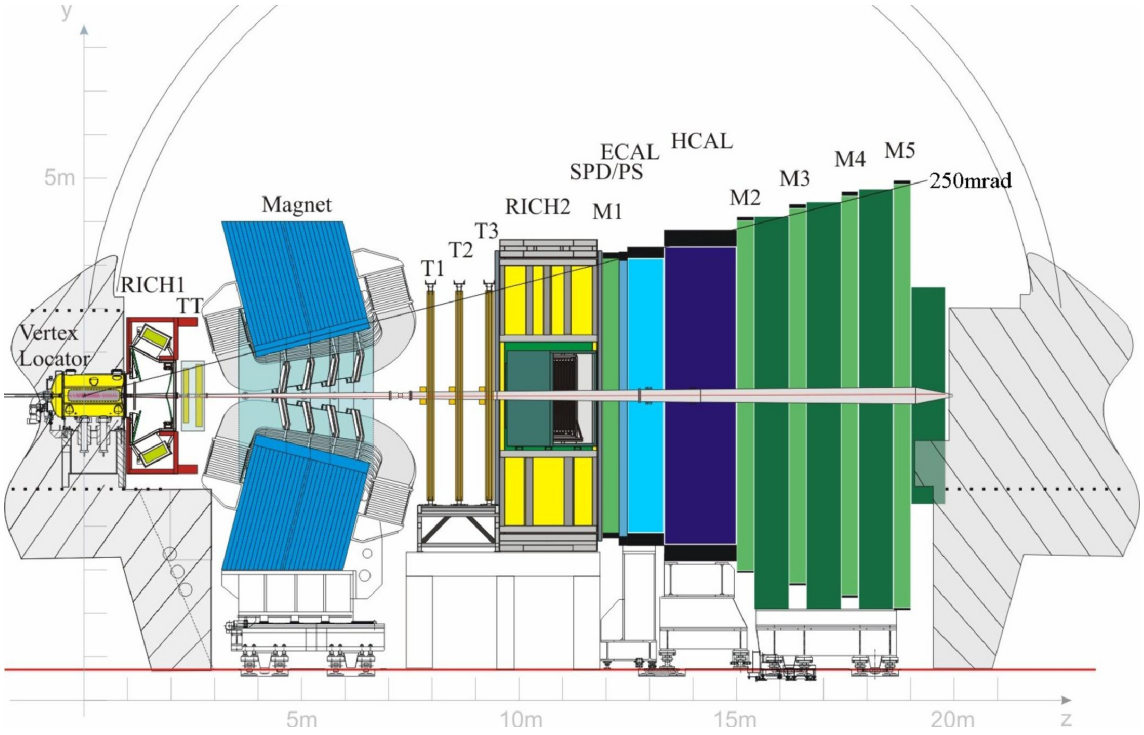


Figure 3.2: Schematic of the LHCb detector [65].

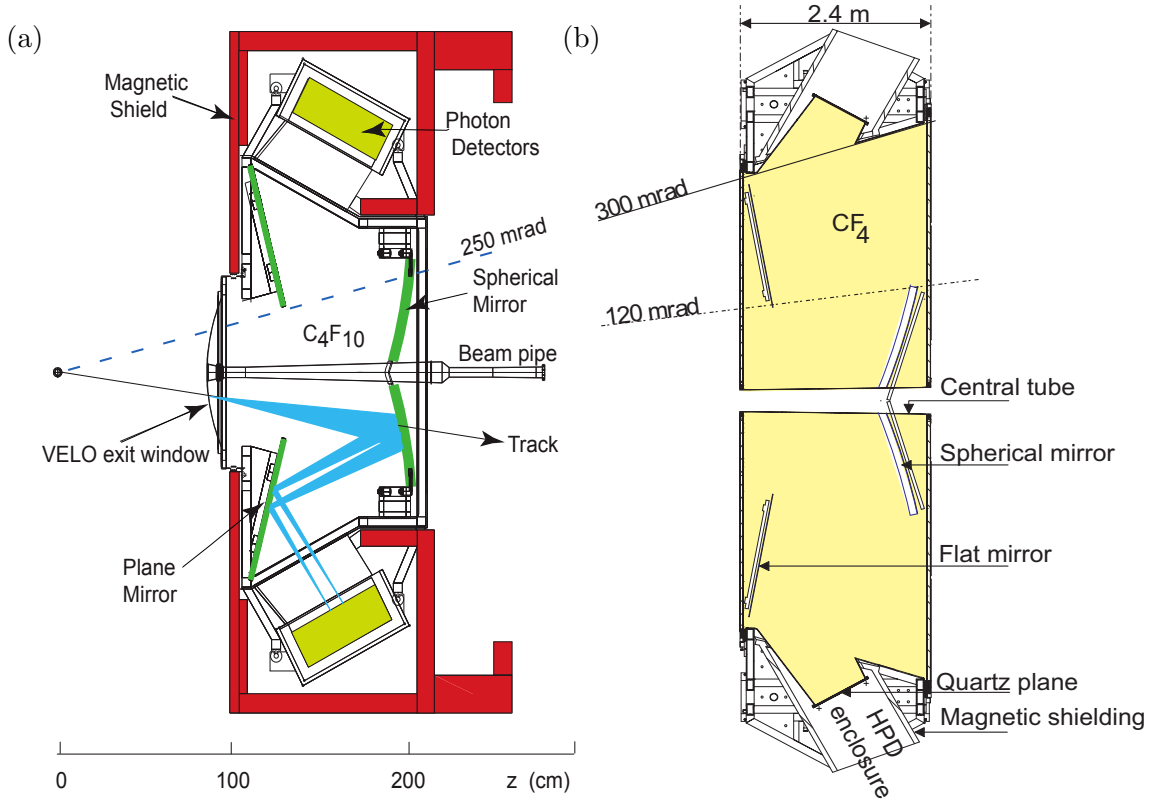


Figure 3.3: Schematic of RICH1 (a) and RICH2 (b) [70].

are generally of heavy flavour hadrons, such as B mesons. The VELO performs extremely well, with over 98% track reconstruction efficiency. For vertices with 25 tracks the primary vertex resolution is 13 μm in the xy plane and 71 μm along the beam axis.

Long-lived charged hadrons are identified using the Ring Imaging Cherenkov (RICH) detectors [68] [69]. There are two RICH detectors, RICH1 and RICH2, designed to measure particles with low ($2 - 40$ GeV/c) and high ($15 - 100$ GeV/c) momenta respectively. Cherenkov photons are emitted when charged particles move faster than light through a radiator. For RICH1 the radiator used is C_4F_{10} (along with aerogel in Run 1) and for RICH2 CF_4 is used, which has a refractive index lower than that of C_4F_{10}. As shown in Figure 3.3, mirrors are used to reflect the Cherenkov light into Hybrid Photon Detectors (HPDs), where the photoelectrons produced are read out as hits. These photon detectors are separated from the radiator volume by quartz windows. Each track is assigned a variable which is the difference of log-likelihoods between the hypothesis of the particle being of type X and type Y from the RICH system, denoted DLL_{XY}

Downstream of the RICH1 detector is a dipole magnet with bending power 4 Tm. Each

magnetic coil consists of 15 layers, where each layer is 15 turns of aluminium. Data are collected with the direction of the magnetic field either in the positive or the negative y-axis directions, known as the 'MagUp' and 'MagDown' configurations. This is in order to reduce uncertainties pertaining to detector asymmetries since periodically changing the magnetic field direction allows different charges to experience both halves of the detector.

The LHCb tracking system consists of the Tracker Turicensis (TT) tracker between RICH1 and the magnet, and three tracking stations T1-3, between the magnet and RICH2. The TT employs 4 silicon microstrip layers and the each of the tracking stations T1-3 have 4 planes consisting of silicon microstrip layers and straw tubes.

The calorimetry system consists of the Scintillating Pad Detector (SPD), Preshower (PS), Electromagnetic Calorimeter (ECAL), and Hadron Calorimeter (HCAL). The PS and SPD are scintillator planes, with a lead wall between them. The ECAL, designed for electron and photon identification, consists of alternating layers of lead plates and scintillator tiles with total depth 25 times the radiation length. The HCAL, designed for hadron identification, consists of layers of iron and scintillator.

For muon identification, the muon station M1-5 are used. These consist of multiwire proportional chambers with the exception of the innermost chambers of the first station M1, where gas electron multiplier detectors are used. The muon stations provide momentum and binary information based upon the number of stations where a hit is found within a region around the track extrapolated from the tracking system. This binary information results in the probability of hadrons to be misidentified as muons at around 1%, with a muon efficiency in the range 95-98% [71]. The muon stations are important for the trigger system as explained in Section 3.3.

There are PID probability variables constructed from combining the information from the sub-detectors into Neural Networks (NNs). These are known as the ProbNNx variables, which corresponds to a score from the NN increasing from 0 to 1 for how likely each track is of species x.

3.3 LHCb Trigger System

The LHCb trigger system used in Run 1 and Run 2 consists of one hardware stage and two software stages [72]. As mentioned in Section 3.1, the bunch crossing rate is 40 MHz. The `L0` hardware trigger has a readout rate of 1 MHz where the trigger is implemented in Field-Programmable Gate Arrays. This trigger fires based on high transverse energy / momentum signatures in the calorimeters or muon stations. The analysis described in this thesis requires the `L0` trigger to have fired in the muon stations.

There are two stages of the software trigger, `HLT1` and `HLT2`. The `HLT1` trigger is based on a selection of events which require minimal reconstruction, such as muon tracks with large impact parameters. For the HLT2 stage a full event reconstruction is performed, where analysis-type specific selections are performed, such as a trigger designed to select B meson decays. The software triggers reduce the readout rate to 5 kHz in Run 1 and 12.5 kHz in Run 2 in order for the events to be saved. A schematic of the LHCb trigger for Run 2 is shown as Figure 3.4 [73].

3.4 LHCb Upgrade (2019-2022)

This section briefly describes the LHCb upgrade during Long Shutdown 2 (LS2) of the LHC [74]. From Run 3, LHCb collects five times the instantaneous luminosity than that of Run 2.

The VELO is upgraded, with more advanced technology where the silicon microstrips are replaced with pixel sensors resulting in improved hit resolution. The VELO will be much closer to the beam than previously (5 mm, compared to 8 mm). This will result in a 40% improvement in vertex resolution.

The RICH detectors have undergone a large upgrade, with the optical system being able to deal with the higher expected occupancy. In addition, new photon detector are being used, where the HPDs are being replaced with Multianode Photomultiplier Tubes (MaPMTs) (Figure 3.5) which allow data to be read out at a rate of 40 MHz. These MaPMTs are assembled on columns which include the electronics. Some images of the RICH upgrade installation are shown in Figure 3.6.

Downstream of RICH1, in replacement of the TT, is the Upstream Tracker (UT), using

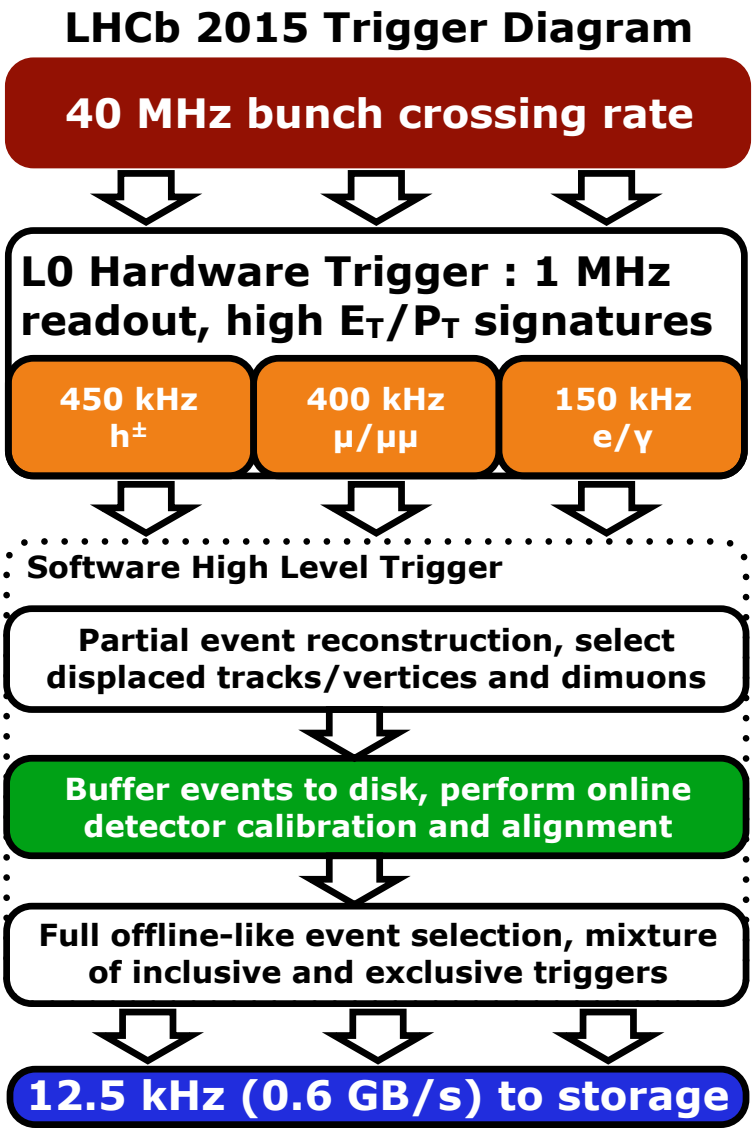


Figure 3.4: Illustration of the LHCb trigger used in Run 2 [73].

silicon strip technology and able to handle the higher expected occupancy. The Scintillating Fiber (SciFi) tracker is downstream of the magnet, in replacement of the T1-3 tracking stations. The SciFi uses scintillating fibres, with high expected hit efficiency and resolution better than 100 μm in the bending plane, read out at 40 MHz.

The trigger is also upgraded so it can run at the increased luminosity. The hardware trigger is removed, thus the LHCb triggers from Run 3 are entirely software-based. The HLT will perform a reconstruction at the rate of 30 MHz. These events are buffered to disk where offline particle identification and track information are used to perform the trigger selection. Events are then stored at a rate 2-5 GB/s [75].

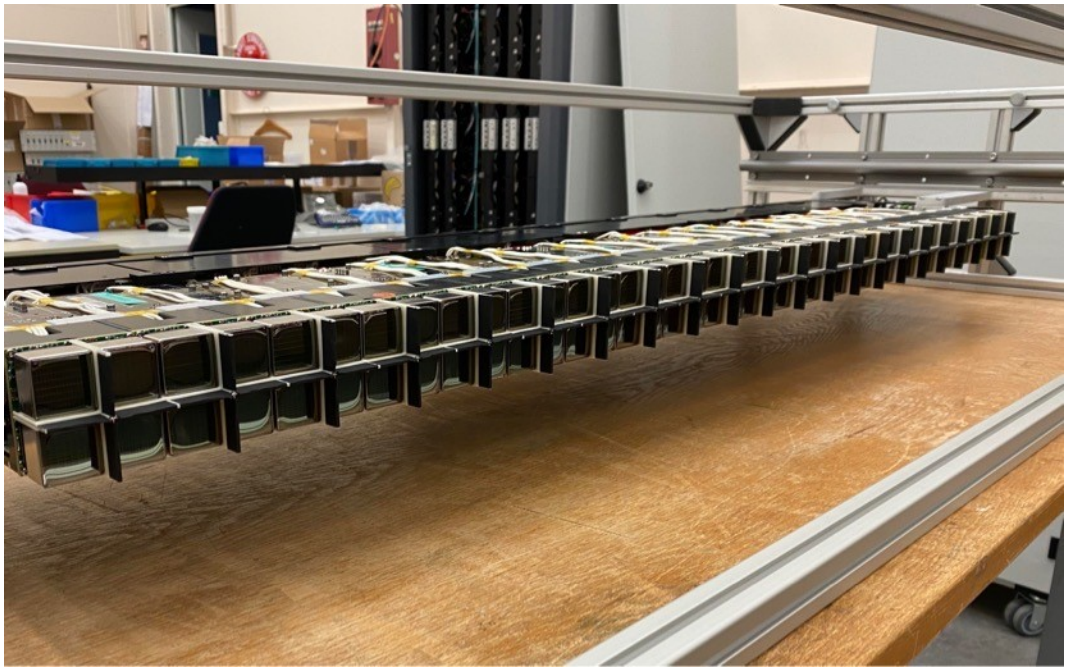


Figure 3.5: MaPMTs used for the RICH upgrade.

Figure 3.6: Installation of the columns in the RICH1 detector.

3.5 Data flow of the LHCb experiment

This analysis relies on simulation. Unless otherwise specified, the simulation used for this analysis was generated centrally by members of the LHCb collaboration. The Gauss simulation framework [76] is responsible for the event and detector simulation used in this analysis. Specifically, proton-proton interactions up to and including hadronisation are simulated using Pythia. The decays of particles, such as B mesons, are simulated using EvtGen. PHOTOS is used for final state radiation. GEANT4 is used to simulate the propagation and interaction of the particles through the detector. The digitisation of the detector hits is simulated by the Boole application [77]. The trigger uses the Moore software framework, and reconstruction is performed using the Brunel software [77].

Offline selections are applied centrally to the stored data, known as 'stripping', using the software DaVinci. Each type of analysis has a set of stripping requirements. The data pre-stripping cannot be accessed directly due to computing restrictions. The analysis presented in this thesis uses data filtered through the `B2XMuMu` stripping line, as described in Section 5.2.

3.6 Using the LHCb experiment to measure the decay $B^0 \to K^{*0}\mu^+\mu^-$

There are several features of the LHCb detector which makes it a suitable experiment to measure the decay $B^0 \to K^{*0}\mu^+\mu^-$. An example is the VELO, which is designed to measure primary and secondary vertices. The B^0 is produced at the primary vertex and decays at the secondary vertex. The B^0 mean lifetime is 1.52×10^{-12} s [14], and due to its Lorentz boost the B^0 travels a few mm before it decays. A variable which is powerful in separating signal B^0 decays from background events is the Impact Parameter (IP). This is the shortest distance between the direction of momentum of a track and the primary vertex. Since the B^0 decays into four charged particles ($B^0 \to K^{*0}\mu^+\mu^-$, and $K^{*0} \to K^+\pi^-$), the IPs of the K, π, and muons can be examined. For combinatorial background events (background events where at least one track is not from the signal), the IP is typically small compared to the signal IP since most combinatorial charged tracks originate from the

primary vertex. Thus background events can be rejected by examining the IPs of the K, π, and muon tracks, resulting in a higher signal purity. In addition, quantities such as the B^0 end vertex χ^2, which corresponds to the quality of the vertex fit of the secondary vertex, can be used to distinguish signal from background. This is because for combinatorial background not all tracks originate from the B^0 thus resulting in a poor vertex fit.

Another feature of the LHCb detector are the RICH detectors which are vital for particle identification, such as distinguishing kaons from pions. In particular for $B^0 \to K^{*0}\mu^+\mu^-$, there are background events which correspond to $\pi\pi$ and KK states misidentified as $K\pi$. The RICH detectors, in addition to the muon system and calorimeters, provide particle identification variables such as $DLL_{K\pi}$ which is the difference of log-likelihoods of a track being a kaon and a pion. As discussed by [78], a loose selection on the this variable results in a high signal efficiency, and a more stringent selection on this variable results in a good background rejection. Thus the information from the RICH detectors helps to remove combinatorial events and also events from B^0 decays which appear signal-like, such as $B_s^0 \to \phi\mu^+\mu^-$, where $\phi \to K^+K^-$, and one of the kaons is reconstructed as a pion.

Chapter 4

Analysis flow of the amplitude ansatz analysis of $B^0 \to K^{*0}\mu^+\mu^-$

This chapter describes the analysis flow of the amplitude ansatz analysis the decay $B^0 \to K^{*0}\mu^+\mu^-$ at LHCb. The objective of this analysis is to measure the decay amplitudes as functions of dimuon invariant mass squared, q^2, aiming to be as model-independent as possible. The amplitudes are modelled using summations of Legendre polynomials in q^2, multiplied by coefficients which vary freely in the fit to data.

A selection is applied to the data collected by LHCb in order to select signal candidates. This process is described in Chapter 5. A set of triggers are used to decide which events to save whilst LHCb is recording data. The first stage of offline selection, known as 'stripping', is performed centrally by the LHCb collaboration and is applied before the data is made available for analysis. Other selections, for example requirements on the Particle IDentification (PID) variables of the hadrons, are applied. Boosted Decision Trees (BDTs) are used to remove background processes which have been misidentified as signal, as well as background processes where at least one track is not from the signal decay.

Due to detector effects, selection, and the reconstruction, the angular and q^2 distributions are warped. This is corrected for by an acceptance function, an efficiency which depends on the angles and q^2. The determination of this function from simulation is described in Chapter 6. A goodness-of-fit method using BDTs has been developed and applied to the acceptance function fit to simulation in order to ensure the acceptance is a

good fit of the simulation.

With the selection applied and acceptance determined, the data can be fitted. A five dimensional unbinned maximum likelihood fit is performed to the data with the selection applied. The fit is performed in the B^0 mass, q^2, $\cos\theta_\ell$, $\cos\theta_K$, and ϕ. The decay distribution of the decay $B^0 \to K^{*0}\mu^+\mu^-$ is written in terms of q^2, $\cos\theta_\ell$, $\cos\theta_K$, and ϕ. This decay distribution is multiplied by the acceptance function in order to obtain the observed decay distribution. This corresponds to the angular signal PDF. Since the data consists of signal and remaining background after the selection is applied, the B^0 mass is included in the fit since the signal and background have very different shapes — the signal lineshape is the sum of two Crystal Ball functions and the background is described by an exponential. In the angles and q^2, the background is described by summations of Chebyshev polynomials multiplied by coefficients. In the B^0 mass signal description, a component accounting for the decay $B_s^0 \to \bar{K}^{*0}\mu^+\mu^-$ is included. More information regarding the fit strategy is given in Chapter 7. Studies pertaining to the ansatz model and whether to include the kaon-pion invariant mass in the fit are described in Chapter 8.

The data fit regions are $1.25 < q^2 < 8$ GeV$^2/c^4$ and $11 < q^2 < 12.5$ GeV$^2/c^4$. These fits are referred to as the 'rare mode' fits. The blinded results for the fit in the $1.25 < q^2 < 8$ GeV$^2/c^4$ region are presented in Chapter 11. The decay $B^0 \to K^{*0}J/\psi(\to \mu^+\mu^-)$ is dominant in the q^2 region $9.223 < q^2 < 9.966$ GeV$^2/c^4$. This region is used as a control mode, where fits are performed in this region in order to validate the implementation of the fitter and to verify agreement between data taking periods. Results from this region are given in Chapter 10.

The free parameters in the fit for both the control mode and rare mode are the amplitude coefficients, background angular and q^2 coefficients, number of background events, mean and widths of the signal B^0 mass lineshapes, and the exponential background parameter. The P-wave amplitude coefficients are the parameters of interest and the remainder of parameters are treated as nuisance parameters.

For the control mode fit, the tails of the signal lineshapes are also floated. These are used to fix the signal lineshape tail parameters in the rare mode fits. In both fits, the difference between the B^0 and B_s^0 masses and the fraction of $B_s^0 \to \bar{K}^{*0}\mu^+\mu^-$ decays relative to $B^0 \to K^{*0}\mu^+\mu^-$ decays are fixed.

Chapter 5

Selection of $B^0 \to K^{*0}\mu^+\mu^-$ candidates

This chapter outlines the selection used in the $B^0 \to K^{*0}\mu^+\mu^-$ amplitude ansatz analysis. The B^0 is produced and decays inside the VELO, resulting in four charged tracks corresponding to K, π, μ^+, and μ^-. There are several types background process in the data which need to be removed. A class of background which requires addressing is combinatorial background, which correspond to background processes where at least one reconstructed track is not from the signal. There are also misidentified backgrounds, where a decay is incorrectly reconstructed as signal due to misidentifying the particles in the final state. These tend to peak in the B^0 mass and are thus known as 'peaking backgrounds'. The selection strategy described in this chapter is designed to ensure the quality of the primary and secondary vertices, where the B^0 is produced and decays. The Particle IDentification (PID) information of the hadrons and muons is used to ensure the particle of the correct species is reconstructed.

Sections 5.1 and 5.2 describe the trigger and the first stage of offline selection used in this analysis. The selections applied directly after stripping are described in Section 5.3. The removal of peaking backgrounds are described in Sections 5.4 and 5.5. Section 5.6 describes the boosted decision tree (BDT) used to remove combinatorial events and Section 5.7 describes a BDT specifically targeted to remove backgrounds such as $B^+ \to K^{*+}\mu^+\mu^-$. The validation of this selection on a background sample, where the muons have the same

Table 5.1: Trigger requirements in $B^0 \to K^{*0}\mu^+\mu^-$ analyses. There are three trigger stages: Level 0 (L0), High Level Trigger 1 (HLT1), and High Level Trigger 2 (HLT2).

	2011+2012	2016	2017+2018
L0	Muon	Muon	Muon
	DiMuon	DiMuon	DiMuon
HLT1	TrackAllL0	TrackMVA	TrackMVA
	TrackMuon	TrackMuon	TrackMuon
		TwoTrackMVA	TwoTrackMVA
HLT2	Topo[2,3,4]BodyBBDT	Topo[2,3,4]Body	Topo[2,3,4]Body
	TopoMu[2,3,4]BodyBBDT	TopoMu[2,3,4]Body	TopoMu[2,3,4]Body
	DiMuonDetached		DiMuonDetached
	SingleMuon		

sign, is described in Section 5.8 and the performance of the selection is described in Section 5.9.

5.1 Trigger

As described in Section 3.3, the LHCb trigger system consists of three stages: the hardware trigger stage Level 0 (L0), and the two software stages High Level Triggers 1 and 2 (HLT1 and HLT2). The trigger requirements are common amongst $B^0 \to K^{*0}\mu^+\mu^-$ analyses, written by the LHCb collaboration and are summarised in Table 5.1.

In order for an event to pass the trigger, the candidate has to fire at least one trigger per trigger level. All triggers are required to be Triggered On Signal (TOS), i.e. the trigger to fire on the signal candidate, not on another process in the event. This allows for efficiencies for the TOS events to be calculated from the TOS and the Trigger Independent of Signal (where the trigger fires from a process which is not due to the signal) samples.

Since the B^0 is heavy and has a long lifetime, the final state particles (K, π, μ^+, μ^-) typically have high transverse momenta. The final state particles are also produced displaced from the primary vertex as described in Section 3.6. Thus the triggers used in this analysis use variables such as the transverse momentum and the chi-squared of the impact parameter with respect to the primary vertex for the final state particles.

For the Level 0 trigger (L0), the Muon and DiMuon triggers require the candidate to pass a threshold of muon transverse momentum p_T, or product of muon transverse momenta respectively. These thresholds change between data-taking periods. The muon transverse

momentum threshold for the `Muon` trigger is in the region $1 - 2$ GeV/c and the product of muon transverse momenta for the `DiMuon` trigger is typically a few GeV$^2/c^2$.

At the `HLT1` trigger level, a partial event reconstruction is performed. Here tracks in the VELO are reconstructed and matched with hits in the muon chambers, as well as hits in the tracking stations. For Run 1, tracks with at least one track with a large enough p_T and χ^2_{IP} (chi-squared of the Impact Parameter (IP)) with respect to the primary vertices are selected by the `TrackAllL0` trigger. The `TrackMuon` trigger is specifically designed to select decays involving muon tracks with large χ^2_{IP} with respect to the primary vertices [79], thus selecting muons whose tracks are displaced from the primary vertex. For Run 2, the `TrackAllL0` trigger was replaced with `TrackMVA` and `TwoTrackMVA`, where machine learning techniques are employed to select tracks based on the p_T and χ^2_{IP}.

A full event reconstruction is performed at the `HLT2` level, including information from other sub-detectors, such as the RICH and the Calorimeters (CALO). In this analysis the `Topo[N]BodyBBDT` and `Topo[N]Body` triggers are are applied, which are designed to select events involving N-body B meson decays. The `SingleMuon` and `DiMuonDetached` triggers require a displaced muon or dimuon system respectively.

The main impact the trigger has on the acceptance shape is the rise from low q^2 to the central q^2 region, seen in Figure 6.2. At low q^2, the acceptance at high and low $\cos\theta_\ell$ moves much more rapidly than at higher q^2. This is due to the `L0` triggers since requiring thresholds on the muon transverse momenta warps q^2 and $\cos\theta_\ell$ in particular. However the parameterisation used to describe the acceptance is sufficient to describe this, validated by a goodness-of-fit described in Chapter 6.

5.2 Stripping

Due to computing restrictions, the first stage of the offline selection of the data is performed centrally by the LHCb collaboration before it is made available to analysts. This stage is known as 'stripping', as described in Section 3.5. Analyses studying the decay $B^0 \to K^{*0}\mu^+\mu^-$ using the full Run 1 and Run 2 datasets use the latest version of the `B2XMuMu` stripping line, which is designed to select events involving a B meson decaying into at least two muons. The selections which form part of the `B2XMuMu` stripping line are

Table 5.2: Requirements in the `B2XMuMu` stripping line.

Candidate	Requirement
B^0	Measured mass $\in [4600, 7100]$ MeV/c^2
	Mass $\in [4700, 7000]$ MeV/c^2
	End vertex χ^2 per DOF < 8
	DIRA > 0.9999
	Flight distance $\chi^2 > 64$
	Maximum daughter $\chi^2_{IP} > 9$
K^{*0}	Measured mass $\in [0, 6200]$ MeV/c^2
	Mass $\in [0, 6200]$ MeV/c^2
	End vertex χ^2 per DOF < 12
	DIRA > -0.9
	Flight distance $\chi^2 > 16$
	Minimum $\chi^2_{IP} > 0$
	Maximum daughter $\chi^2_{IP} > 6$
$\mu^+\mu^-$	Measured mass < 7100 MeV/c^2
	End vertex χ^2 per DOF < 12
	DIRA > -0.9
	Flight distance $\chi^2 > 9$
	Maximum χ^2_{IP} of either μ^+ or $\mu^- > 6$
K^+, π^-	Minimum $\chi^2_{IP} > 6$
	Track registered in the RICH subsystem
μ^+, μ^-	Minimum $\chi^2_{IP} > 6$
	$DLL_{\mu\pi} > -3$
	Positive `isMuon` detected, based on muon chamber hits
Tracks	Ghost probability < 0.5
Event	Maximum hits in Scintillating Pad Detector < 600

shown in Table 5.2.

Figure 5.1 is a schematic showing the primary vertex, where a proton-proton collision occurs, and the decay vertex of a B^0 meson. The decay vertex of the B^0 is also known as the 'secondary vertex', or the 'end vertex'. The distance between the primary vertex and the decay (secondary) vertex is the flight distance. The 'measured mass' is the invariant mass of the system based on the 4-momenta. The 'mass' corresponds to the invariant mass obtained from the vertex fit. The DIRA is the cosine of the angle between the reconstructed momentum and reconstructed line of flight of the candidate. Thus from Table 5.2, the primary and secondary vertices are required to be of good quality. For each track, the variable DLL_{XY} is assigned which corresponds to the difference of log-likelihoods for the hypotheses of the particle being of type X and type Y. The variable `isMuon` corresponds to hits in the muon system which are consistent with muons. As part of the reconstruction a

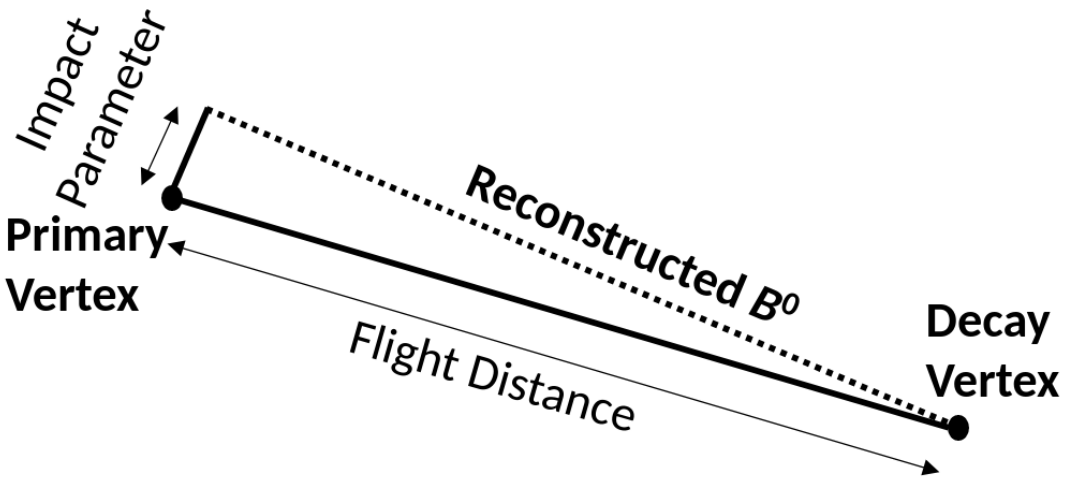


Figure 5.1: Schematic showing the primary vertex, decay (secondary) vertex, flight distance and impact parameter. The DIRA is the cosine of the angle between the reconstructed momentum and reconstructed line of flight of the candidate.

"ghost probability", based on fit quality, the number of missing hits, and energy deposited in the tracker, is computed. Selecting low ghost probability tracks effectively removes tracks that do not correspond to real particles. The cut on the maximum number of hits in the Scintillating Pad Detector is applied in order to remove events which are too crowded.

5.3 Cleaning cuts

Some cleaning cuts follow the same selections as the ongoing binned angular analysis, as shown in Table 5.3. Here, pairs of tracks are required to be well separated in order to reduce the number of clones (tracks which share a large percentage of hits from the VELO and T-stations). The requirement on the B^0 end vertex χ^2 per DOF is tightened in order to select B mesons with a good fit quality. As described by the CP-asymmetries analysis of the decays $B^0 \to K^{*0}\mu^+\mu^-$ and $B^+ \to K^+\mu^+\mu^-$ [80], tracks which are close the edges of the detector can be swept out by the magnetic field, resulting in highly asymmetric decay rates. Thus the selection based on the kaon momenta, motivated by Ref. [80], aims to reduce raw detector asymmetries, which are largest at the edges of the detector.

The general Particle IDentification (PID) requirements are shown in Table 5.4. These selections are based on ProbNN variables and remove a large amount of mis-ID background. In particular, these help to remove KK and $\pi\pi$ backgrounds, in addition to mis-ID

Table 5.3: Cleaning cuts which are identical to those in the ongoing binned angular analysis.

Candidate	**Requirement**
Track pairs	Angular separation $\theta > 1$ mrad
B^0	End vertex χ^2/DOF < 6
K^+	$p_Z(K^+) > 2000$ MeV/c $\lvert p_X(K^+)\rvert < 0.33(p_Z(K^+) - 2000)$

backgrounds such as $B^0 \to K^{*0} J/\psi$ where $\mu \to \pi$, $\pi \to \mu$. These typically peak in the B^0 mass and is thus vital to remove them. The selections based on the PID requirements help with the overall reduction of these backgrounds and allow the further selection stages to be more efficient in removing more signal-like backgrounds.

Table 5.4: Hadron general PID requirements

Candidate	**Requirement**
K	ProbNNk > 0.2
π	ProbNN$\pi > 0.2$

Additionally, the selection $5170 < m_{K\pi\mu^+\mu^-} < 5700$ MeV/c^2 is applied. The $m_{K\pi}$ window chosen is $796 < m_{K\pi} < 996$ MeV/c^2. The q^2 window chosen for analysis in the rare mode is $1.25 < q^2 < 8$ GeV$^2/c^4$. The region $11 < q^2 < 12.5$ GeV$^2/c^4$ is also included, however studies in this region are not presented in this thesis. The lower q^2 boundary at 1.25 GeV$^2/c^4$ is due to low mass resonances such as the $\phi(1020)$ resulting in fast-moving behaviour in the amplitudes. It was found that a lower bound at 1.25 GeV$^2/c^4$ results in a better description of the amplitudes since the model used does not need to describe the $\phi(1020)$ resonance. The choice of the upper q^2 bound 8 GeV$^2/c^4$ in the low q^2 is due to the radiative tail of the J/ψ entering the B^0 mass region, and that the amplitudes are more fast-moving as q^2 becomes closer to the J/ψ. For the J/ψ region, the q^2 region corresponds to $9.223 < q^2 < 9.966$ GeV$^2/c^4$. The purpose of measuring the J/ψ resonance is to validate the fitter and the acceptance.

5.4 Peaking background vetoes

There are several decays which could be incorrectly reconstructed as the signal and may cause fit biases. These are often decays with a reconstructed $K\pi\mu\mu$ final state, where at least one particle has been identified as an incorrect species. These decays generally peak

in the B mass and have angular structures which do not factorise. Table 5.5 shows a list of some incorrectly reconstructed backgrounds which may cause fit biases. Details regarding the mis-ID for each background are shown.

Table 5.5: A summary of the peaking backgrounds. The type of background and mis-ID is also specified.

Mode	**Type**
$B^0 \to K^{*0}\mu^+\mu^-$, where $K \leftrightarrow \pi$	Double hadron mis-ID
$B_s^0 \to \phi\mu^+\mu^-$, where $K \to \pi$	Single hadron mis-ID
$B^0 \to K^{*0}J/\psi$, where $\mu \leftrightarrow K$	Hadron-muon swap
$B^0 \to K^{*0}J/\psi$, where $\mu \leftrightarrow \pi$	Hadron-muon swap
$B^0 \to K^{*0}\psi(2S)$, where $\mu \leftrightarrow K$	Hadron-muon swap
$B^0 \to K^{*0}\psi(2S)$, where $\mu \leftrightarrow \pi$	Hadron-muon swap
$\Lambda_b \to pK\mu^+\mu^-$, where $p \to \pi$	Single hadron mis-ID
$\Lambda_b \to pK\mu^+\mu^-$, where $p \to K$ and $K \to \pi$	Double hadron mis-ID

Variables which are used to distinguish these decays are the PID variables and two-body or four-body masses under the swapped hypothesis. Therefore a cut needs to be applied in these variables in order to remove these backgrounds. An option is to apply a 'rectangular' cut, for example removing all events with a mass under the swapped hypothesis in a certain range, where the PID is above or below some threshold. However this is not the optimal cut. A more efficient cut can be determined by using Boosted Decision Trees (BDTs), which was initially developed for the binned angular analysis. Here, BDTs are used to classify each event by providing a score, allowing for signal / peaking background separation. A cut is applied based on the BDT scores.

Decision Trees (DTs) are structures consisting of nodes and leaves which use a sequence of cuts to bisect the data, with the goal of each cut to maximise the purity of the two samples. Each subsample is treated independently with further bisecting cuts to form a tree-like structure until a stopping condition is reached. However training a DT could result in biases and/or variances in the results which is undesirable. Boosting is therefore used, where weak learners are combined into strong learners. A BDT consists of a forest of trees, with each tree of a shallow depth. Whilst each individual tree may not separate the samples effectively, if better than a random guess, the result of boosting results in the ensemble of weak classifiers becoming a strong classifier.

For each of these background modes a sample of simulated events is obtained with the

relevant mis-identification hypothesis enforced. The trigger, stripping, cleaning cuts and invariant mass cuts are applied (B mass, $K\pi$ mass, rare mode q^2). The remaining samples are used to train a Boosted Decision Tree (BDT) for each background. For the signal sample, $B^0 \to K^{*0}\mu^+\mu^-$ simulation is used, with the same selection applied. The BDT is trained using the `xgboost` algorithm [81], a type of BDT based upon a gradient-descent framework.

For most of the BDTs the cut point is chosen such that the signal efficiency is equal to that of the vetoes used in the previous binned angular analysis [7]. Sometimes harsher BDT cut points are used in order to get the overall estimated level of peaking backgrounds to a reasonable level (i.e below 1% of the signal after the full selection is applied). Due to the small samples that pass the initial selection requirements, k-fold cross validation is used in order to make sure the BDT is statistically independent. This is a procedure where the dataset is split into k subsets, or 'folds', and for each fold the BDT is trained on the other folds and tested for that fold. Here 10 folds are used, and one fold is chosen for use in the analysis. The BDTs use PID variables and masses under the swapped hypothesis in the training. The DLL variables are chosen instead of ProbNN to reduce the possibility of correlations with kinematic variables which are poorly modelled in simulation. Since the cleaning cuts are applied, in particular the hadron PID selection, the BDTs are optimised to remove peaking backgrounds which appear signal-like.

Table 5.6 shows the training variables used in these BDTs. These include invariant masses under the swapped hypotheses, such as $m_{swap}(K\pi)$, where the $K\pi$ invariant mass is computed with the kaon and pion masses swapped. The DLL variables are also used.

Table 5.7 presents figures of the BDT responses and signal and background rejection efficiencies. Note these efficiencies are conditional. Previous analyses of $B^0 \to K^{*0}\mu^+\mu^-$ did not have hadron PID requirements as strong as shown in Table 5.4 thus these peaking background BDTs are optimised to remove mis-ID backgrounds which appear particularly signal-like. For BDTs involving true events from the J/ψ, only events with the swapped dimuon mass hypothesis in the J/ψ region ($3050 < m(\mu^+\mu^-) < 3150$ MeV/c^2) are considered in the training. Likewise for BDTs involving true events from the $\psi(2S)$, only events with the swapped dimuon mass hypothesis in the $\psi(2S)$ region ($3626 < m(\mu^+\mu^-) < 3746$ MeV/c^2) are considered. The efficiencies shown are still for the full samples.

Table 5.6: Training variables used in the peaking background BDTs.

Background	mis-ID	Training variables
$B^0 \to K^{*0}\mu^+\mu^-$	$K \leftrightarrow \pi$ mis-ID	dihadron mass with the background hypothesis, $DLL_{K\pi}$ for the hadrons
$B^0_s \to \phi\mu^+\mu^-$	$K \to \pi$ mis-ID	dihadron mass with the background hypothesis, $DLL_{K\pi}$ for the pion
$B^0 \to K^{*0}J/\psi$	$\mu \leftrightarrow K$ swap	dimuon mass with the background hypothesis, $DLL_{K\pi}$ and $DLL_{\mu\pi}$ for the kaon and muon
$B^0 \to K^{*0}J/\psi$	$\mu \leftrightarrow \pi$ swap	dimuon mass with the background hypothesis, $DLL_{\mu\pi}$ for the pion and muon
$B^0 \to K^{*0}\psi(2S)$	$\mu \leftrightarrow K$ swap	dimuon mass with the background hypothesis, $DLL_{K\pi}$ and $DLL_{\mu\pi}$ for the kaon and muon
$B^0 \to K^{*0}\psi(2S)$	$\mu \leftrightarrow \pi$ swap	dimuon mass with the background hypothesis, $DLL_{\mu\pi}$ for the pion and muon
$\Lambda_b \to pK\mu^+\mu^-$	$p \to \pi$ mis-ID	dihadron mass with the background hypothesis, B mass with the background hypothesis, $DLL_{p\pi}$ for the pion
$\Lambda_b \to pK\mu^+\mu^-$	$p \to K$, $K \to \pi$ mis-ID	dihadron mass with the background hypothesis, B mass with the background hypothesis, $DLL_{K\pi}$ for the pion, $DLL_{p\pi}$ for the kaon

Table 5.7: Signal efficiency and background rejection efficiency for each background, with figures showing the BDT response for each background indicated.

Background	**mis-ID**	**Figure**	**Signal efficiency**	**Background rejection efficiency**
$B^0 \to K^{*0}\mu^+\mu^-$	$K \leftrightarrow \pi$ mis-ID	5.2a	$99.24 \pm 0.01\%$	$89.3 \pm 0.4\%$
$B^0_s \to \phi\mu^+\mu^-$	$K \to \pi$ mis-ID	5.2b	99.896 ± 0.004 %	90.2 ± 0.2 %
$B^0 \to K^{*0}J/\psi$	$\mu \leftrightarrow K$ swap	5.2c	$99.23 \pm 0.01\%$	$96 \pm 2\%$
$B^0 \to K^{*0}J/\psi$	$\mu \leftrightarrow \pi$ swap	5.2d	$99.634 \pm 0.007\%$	$97.5 \pm 0.7\%$
$B^0 \to K^{*0}\psi(2S)$	$\mu \leftrightarrow K$ swap	5.2e	$99.557 \pm 0.008\%$	$96 \pm 2\%$
$B^0 \to K^{*0}\psi(2S)$	$\mu \leftrightarrow \pi$ swap	5.2f	$99.915 \pm 0.008\%$	$98 \pm 1\%$
$\Lambda_b \to pK\mu^+\mu^-$	$p \to \pi$ mis-ID	5.2g	$99.304 \pm 0.008\%$	$89.6 \pm 0.5\%$
$\Lambda_b \to pK\mu^+\mu^-$	$p \to K$, $K \to \pi$ mis-ID	5.2h	$98.948 \pm 0.008\%$	$83.1 \pm 0.8\%$

All of these peaking background BDTs are applied to data and simulation, with the selection requiring events to pass the chosen cut points for all of the BDTS.

5.5 $B^+ \to K^+\mu^+\mu^-$ veto

There is expected to be a background due to the decay products of $B^+ \to K^+\mu^+\mu^-$ being paired with a pion from elsewhere in the event. Here the same veto as is in other $B^0 \to K^{*0}\mu^+\mu^-$ analyses is used, as proposed in [82]. For

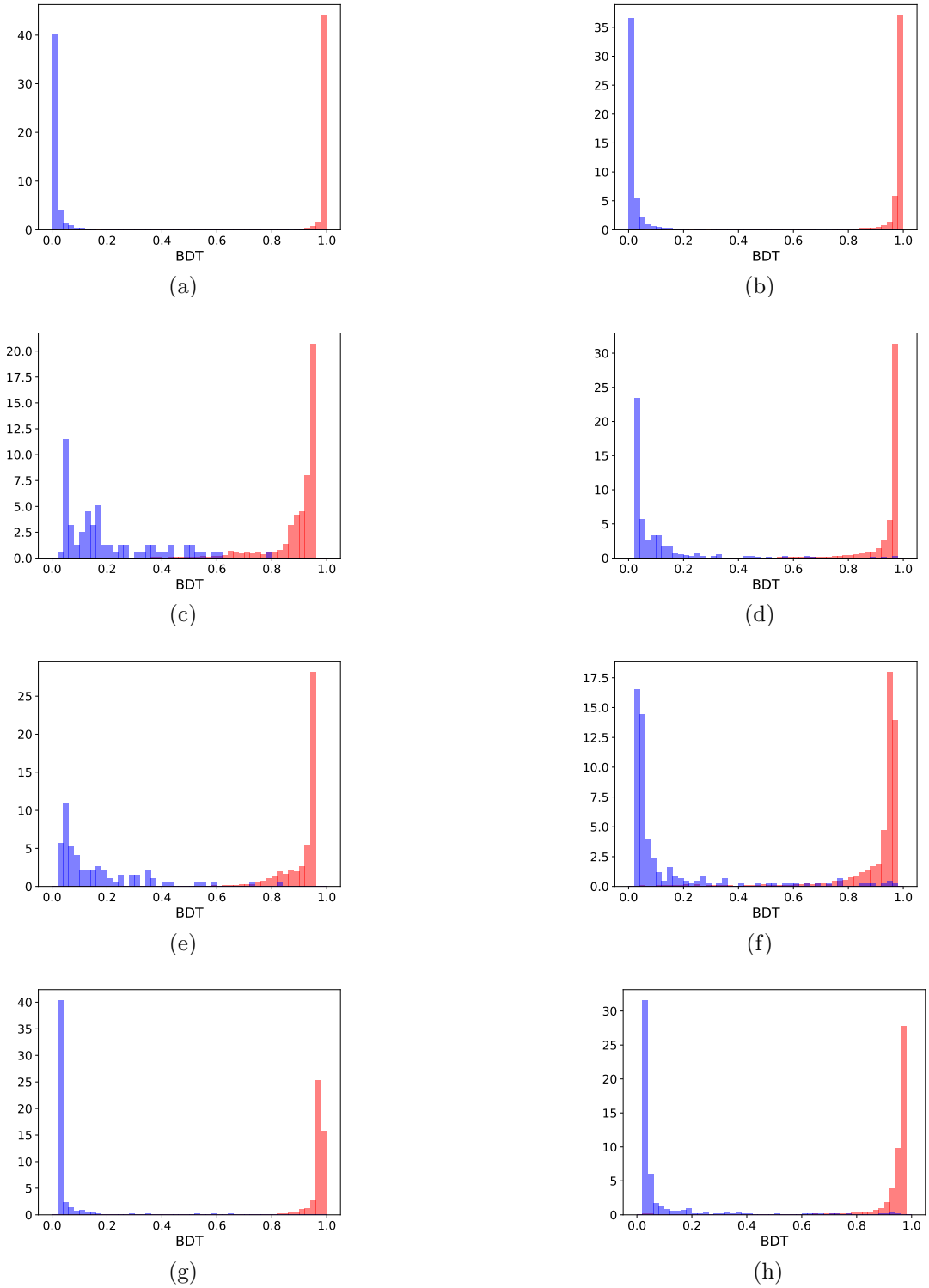


Figure 5.2: Results of the training of various BDTs on signal simulation (red) against background simulation (blue). The data is kFolded, and the response shown here is for the testing sample. For the background the BDTs shown here are trained on (a) double hadron mis-ID from $B^0 \to K^{*0}\mu^+\mu^-$, (b) $B_s^0 \to \phi\mu^+\mu^-$, (c) $\mu \leftrightarrow K$ swap from the J/ψ, (d) $\mu \leftrightarrow \pi$ swap from the J/ψ, (e) $\mu \leftrightarrow K$ swap from the $\psi(2S)$, (f) $\mu \leftrightarrow \pi$ swap from the $\psi(2S)$, (g) $\Lambda_b \to pK\mu^+\mu^-$, where $p \to \pi$, and (h) $\Lambda_b \to pK\mu^+\mu^-$, where $p \to K$ and $K \to \pi$.

those event with $m_B > 5380$ MeV/c^2 (the B^0 mass plus a pion) the region with $5220 < m(K\mu^+\mu^-) < 5340$ MeV/c^2 is excluded, corresponding to a 60 MeV/c^2 window around the B^+ mass. This selection causes a hole in the space of m_B, $\cos\theta_K$, and q^2 above the B^0 mass which is accounted for in the background PDF as described in Section 7.2.1.

5.6 Combinatorial BDT

After removing backgrounds arising from particle mis-identification, there remains a large amount of background arising from combinatorial events. These are where at least one particle does not originate from the B^0. Since these backgrounds in general are in the full B^0 mass range and not centred in a particular mass region, they cannot be removed as easily as $B^+ \to K^+\mu^+\mu^-$. In addition, given the different centre-of-mass energies of the pp collisions between Run 1 and Run 2, the combinatorial background is expected to be different such that treating these periods separately results in a more efficient removal of these backgrounds. Two BDTs (one for each run period) are therefore designed to remove these combinatorial backgrounds, one each for Run 1 and Run 2, using the `xgboost` algorithm [81]. k-fold cross validation is used with 10 folds. Separate BDTs were not considered for the peaking backgrounds in Table 5.5 since these backgrounds are controlled at a reasonable level, as shown by Section 5.9. The estimated peaking background yields relative to the signal after the peaking background BDTs were applied were also found to not vary much between data taking periods.

Since this BDT is designed to remove any combinatorial event rather than a specific background, data is used as a proxy for the background. The BDT is trained on the upper B^0 mass sideband data (defined as the region $5350 < m_B < 5700$ MeV/c^2) for the background. This region is the selected (trigger, stripping, cleaning cuts, mass selection, peaking background vetoes) region of the data sample, using the Run 1 sample for the Run 1 BDT and the Run 2 sample for the Run 2 BDT.

For the signal, the BDT is trained on the $B^0 \to K^{*0}\mu^+\mu^-$ simulation, with all of the selections including mass selections applied. The set of BDT training variables are

- B^0 lifetime, p_T (transverse momentum), p (momentum), end vertex χ^2, DIRA
- $DLL_{\mu\pi}(\mu^+)$, $DLL_{\mu\pi}(\mu^-)$, $DLL_{K\pi}(K)$, $DLL_{K\pi}(\pi)$

- Muon and hadron isolation variables
- IP χ^2 of the pion, kaon, μ^+, μ^-, B^0
- Flight distance χ^2 of the K^{*0}, B^0
- End vertex χ^2 of the K^{*0}.

The isolation variable for a reconstructed hadron or muon track corresponds to the number of extra tracks (i.e. tracks which are not from the signal candidate) that can form a vertex with the reconstructed track. Thus the isolation variables are discrete.

The choice of variables and training samples are such that the combinatorial BDT is trained to remove many event topologies, such as combinatorial events where the hadrons and muons are displaced from the primary vertex. Of these variables, the IP χ^2, flight distance χ^2 and end vertex χ^2 variables are found to be useful since they are effective at removing different kinds of combinatorial backgrounds (e.g. $B^+ \to \bar{D}^0 \mu^+ \nu_\mu$, where $\bar{D}^0 \to K^+ \mu^- \bar{\nu_\mu}$ with a combinatorial pion).

The simulation has per-candidate weights applied in order to account for mismodelling in the simulation. These were developed by the ongoing binned angular analysis, and consist of trigger weights, track weights, and kinematic weights. The trigger weights account for mismodelling of the `L0` trigger in the simulation. The efficiency of the trigger is computed using $B^+ \to J/\psi K^+$ data and simulation, and is parametrised in two dimensions by the p_T of the two decay product muons. The TISTOS method [83] is used where the numbers of candidates which are Triggered Independent of Signal (TIS), Triggered On Signal (TOS) and simultaneously TIS and TOS (TISTOS) are used to extract an efficiency of the trigger. The ratio of the TISTOS efficiencies in data to simulation is applied as a correction for the trigger efficiency. The track weights account for mismodelling of the tracking efficiency, with the correction applied based on the p_T and η variables of each of the decay products since these variables are correlated with the tracking efficiency. These corrections are obtained by simulating a decay channel which is experimentally clean and with large yields in data, such as $J/\psi \to \mu^+ \mu^-$ and comparing this with data in order to extract a correction. Finally, the kinematic corrections are used to correct the B^0 p and p_T spectra, where $B^+ \to J/\psi K^+$ is used for these corrections. Here, per-event weights are obtained via training BDTs on $B^+ \to J/\psi K^+$ simulation and data.

The distribution of the training variables for the signal and background samples for the Run 1 BDT are shown as Figures 5.3, 5.4, and 5.5. Similarly, the distribution of the training variables for the signal and background samples for the Run 2 BDT are shown as Figures 5.6, 5.7, and 5.8. The variables which were found to have highest variable importance (number of times a variable is used to split the data) are the B^0 end vertex χ^2, the B^0 DIRA, and the flight distance χ^2 of the B^0. These are important variables since they correspond to the quality of the vertex fit and reconstructed B^0. In addition, the hadron isolation variables provide the BDT with greater separation power since, compared to the signal, tracks from combinatorial background often have more tracks which are nearby.

The agreement between simulation and data of the input variables has been examined by comparing the input variables in the control mode. The input variables in $B^0 \to K^{*0}J/\psi$ simulation are compared with $B^0 \to K^{*0}J/\psi$ data which has been background subtracted using the sWeight technique [84]. These are shown as Figure 5.9, 5.10, and 5.11 for 2017. For the rest of the data periods, these are shown in Appendix A. As seen in these plots, the agreement between simulation and data in the control mode appears to be reasonable, but not perfect. This is most likely due to a number of features in the data. Firstly, given the backgrounds are non-negligible in the control mode, and these do not necessarily factorise with the B^0 mass, the sWeighting procedure is not perfect. In addition, even if there is no issue with the sWeighting, the signal data sample is not 100% P-wave $B^0 \to K^{*0}J/\psi$ as in the simulation. There is an S-wave contribution as well as contributions from exotic states, i.e. $B^0 \to Z(4430)^- K^+$, where $Z(4430)^- \to J/\psi\pi^-$. This is evident in Figure 5.12 which shows $\cos\theta_K$ in the $B^0 \to K^{*0}J/\psi$ simulation and sWeighted $B^0 \to K^{*0}J/\psi$ data.

Given the angles are different from the outset, comparisons of the BDT efficiency between simulation and data were made. Figure 5.13 shows the BDT efficiency as a function of $\cos\theta_\ell$, $\cos\theta_K$, and ϕ for the $B^0 \to K^{*0}J/\psi$ simulation and data. The efficiency is fairly flat in $\cos\theta_\ell$ and ϕ, and mostly flat in $\cos\theta_K$ with a drop at high $\cos\theta_K$. Comparing simulation to data, the BDT angular efficiency appears well modelled apart from at high $\cos\theta_K$, where the difference is most likely due to backgrounds which have not been sWeighted properly in data, as well as the exotic states and S-wave of the J/ψ data. In order to investigate this, the efficiencies were examined in a narrow $m_{K\pi}$ window around

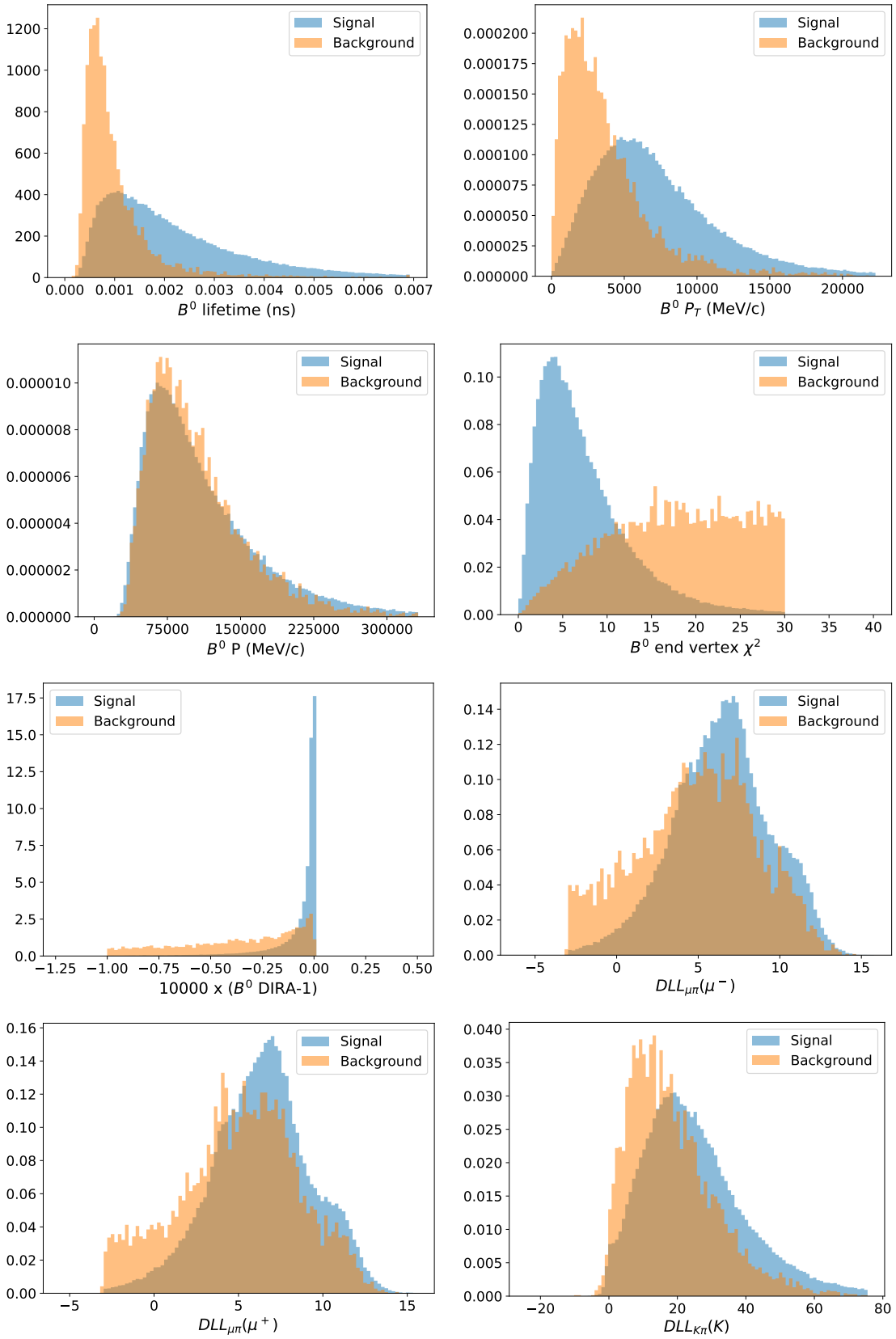


Figure 5.3: Training variables for the Run 1 combinatorial BDT. Signal is shown in blue and background is shown in orange. The distributions are normalised to unit area.

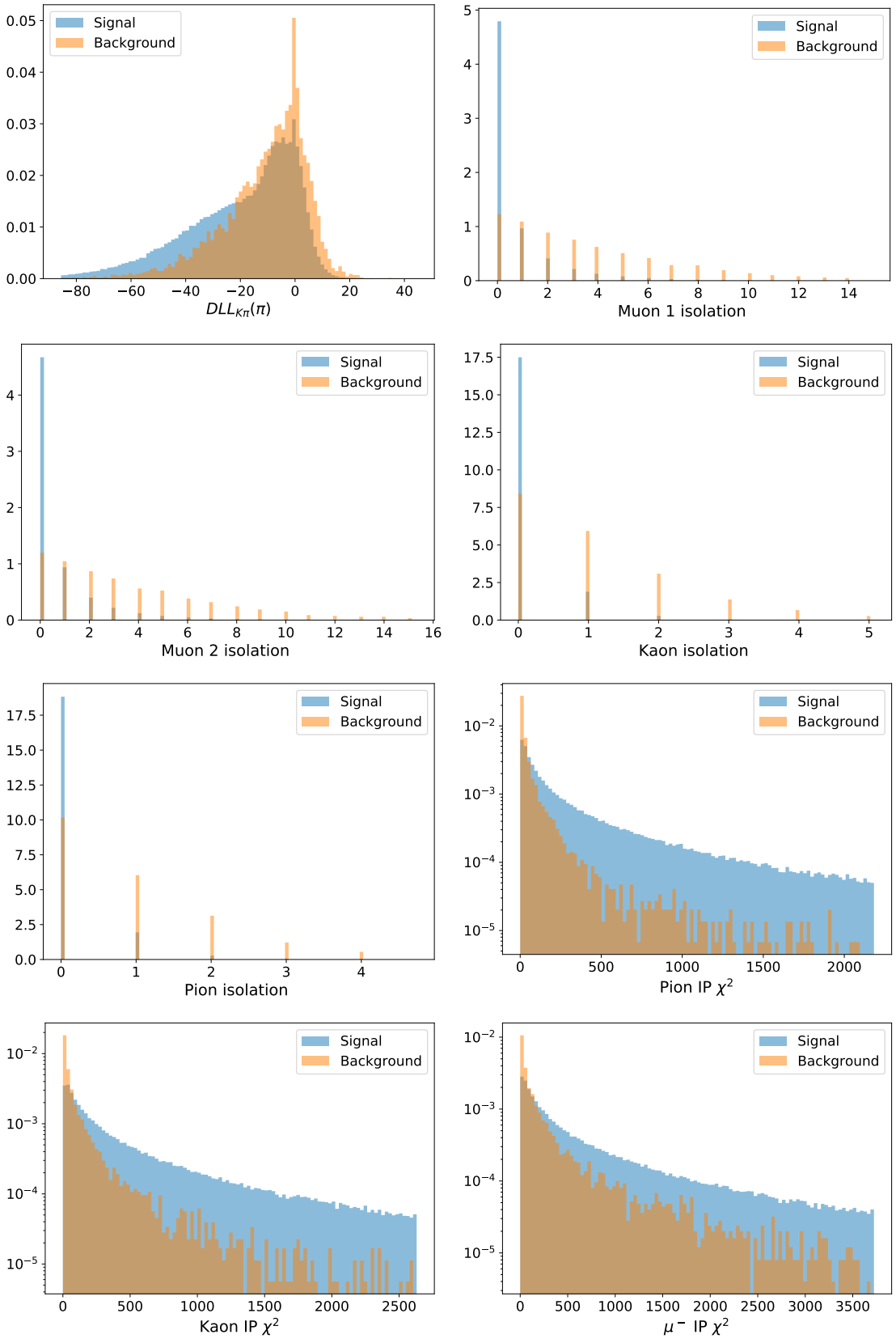


Figure 5.4: Training variables for the Run 1 combinatorial BDT. Signal is shown in blue and background is shown in orange. The distributions are normalised to unit area.

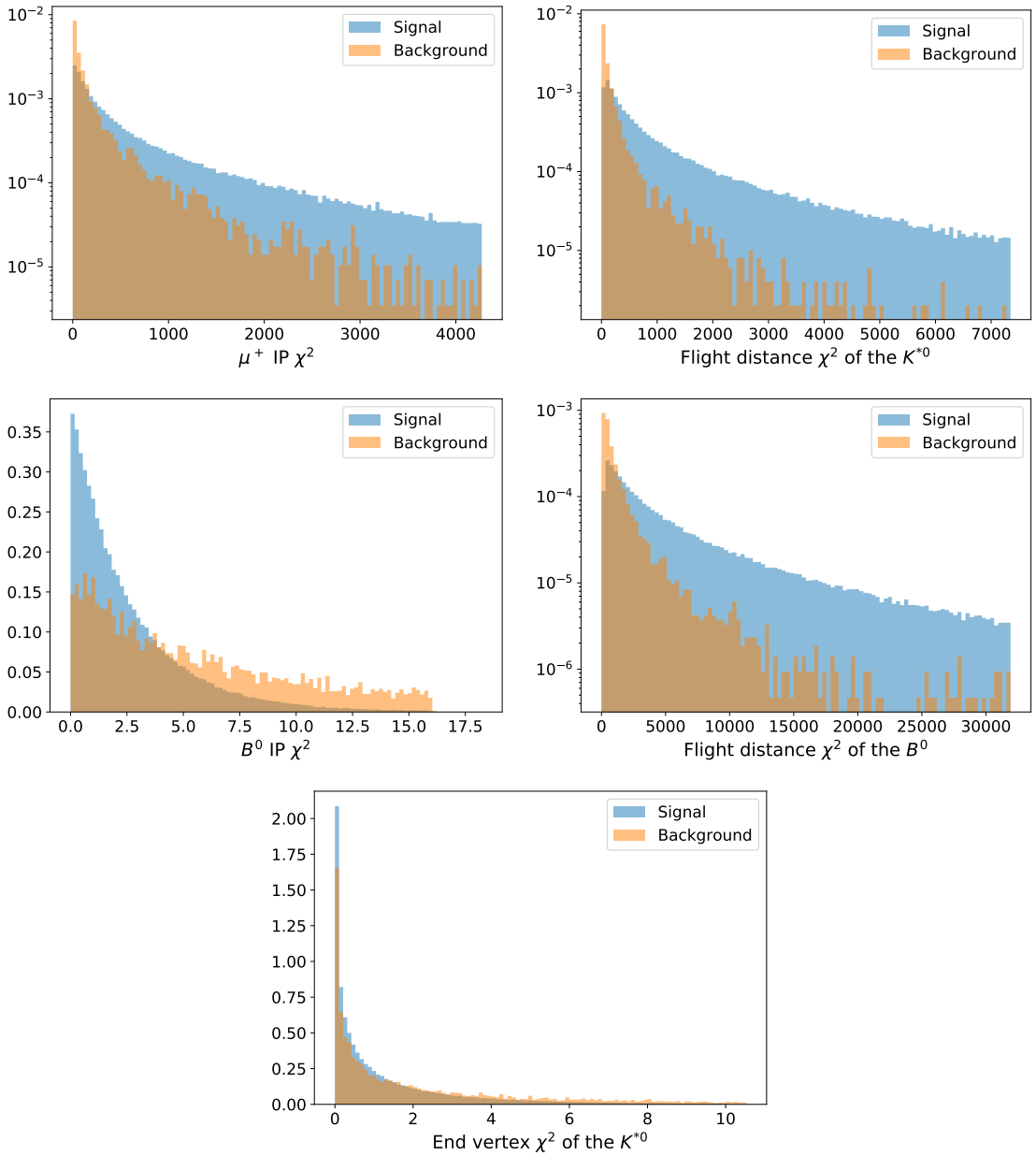


Figure 5.5: Training variables for the Run 1 combinatorial BDT. Signal is shown in blue and background is shown in orange. The distributions are normalised to unit area.

the K^{*0} pole mass. A selection $876 < m_{K\pi} < 916$ MeV/c^2 was applied to both data and simulation to ensure a sample which has a high purity of P-wave J/ψ. The efficiencies for 2017 in this $m_{K\pi}$ region are shown in Figure 5.14. As seen in the high $\cos\theta_K$ region, the discrepancy between data and simulation reduces as the data becomes more P-wave dominated. The same behaviour is seen in the other data periods, as shown in Appendix B.

The results of the training for Run 1 and Run 2 can be seen in Figure 5.15. The BDT

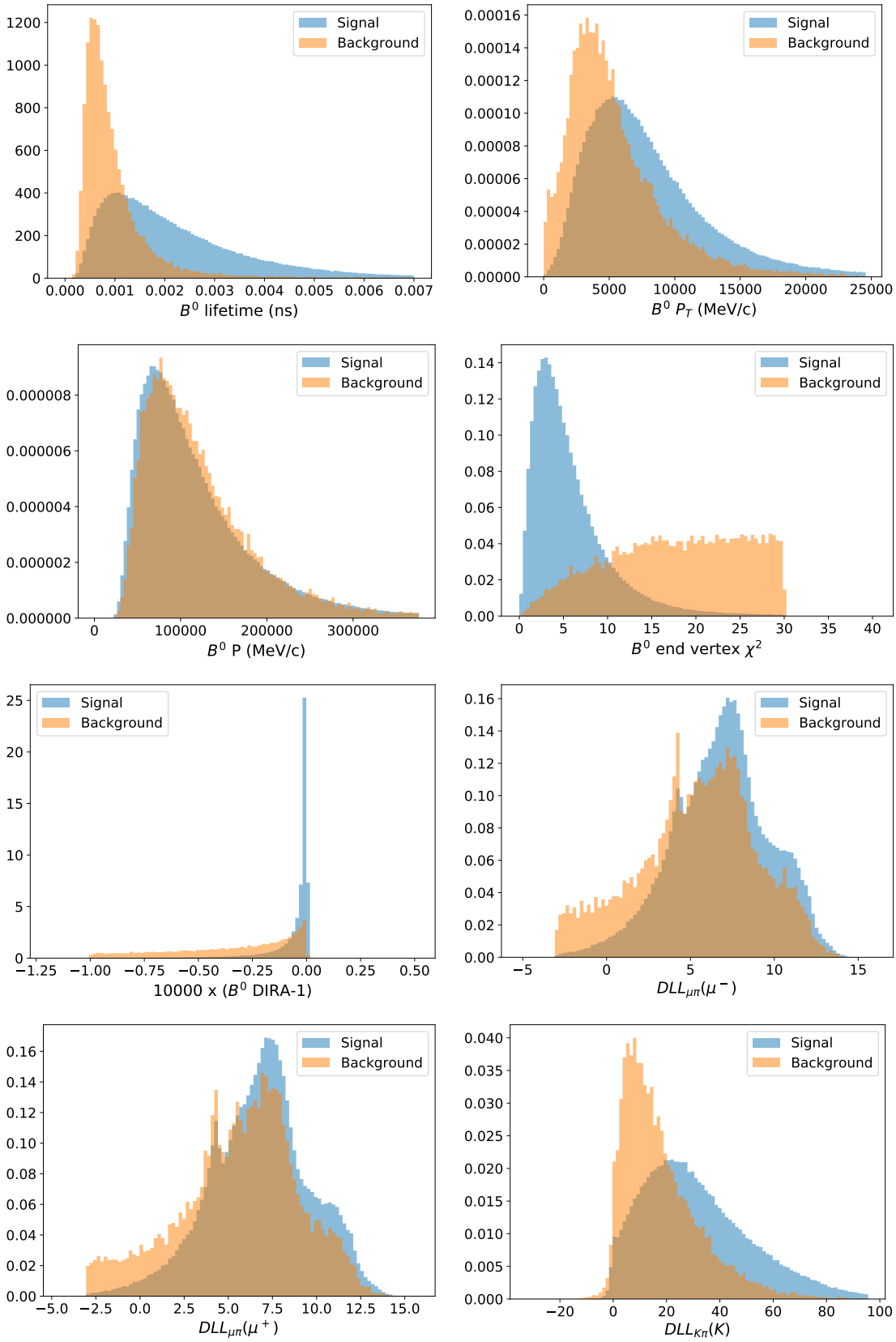


Figure 5.6: Training variables for the Run 2 combinatorial BDT. Signal is shown in blue and background is shown in orange. The distributions are normalised to unit area.

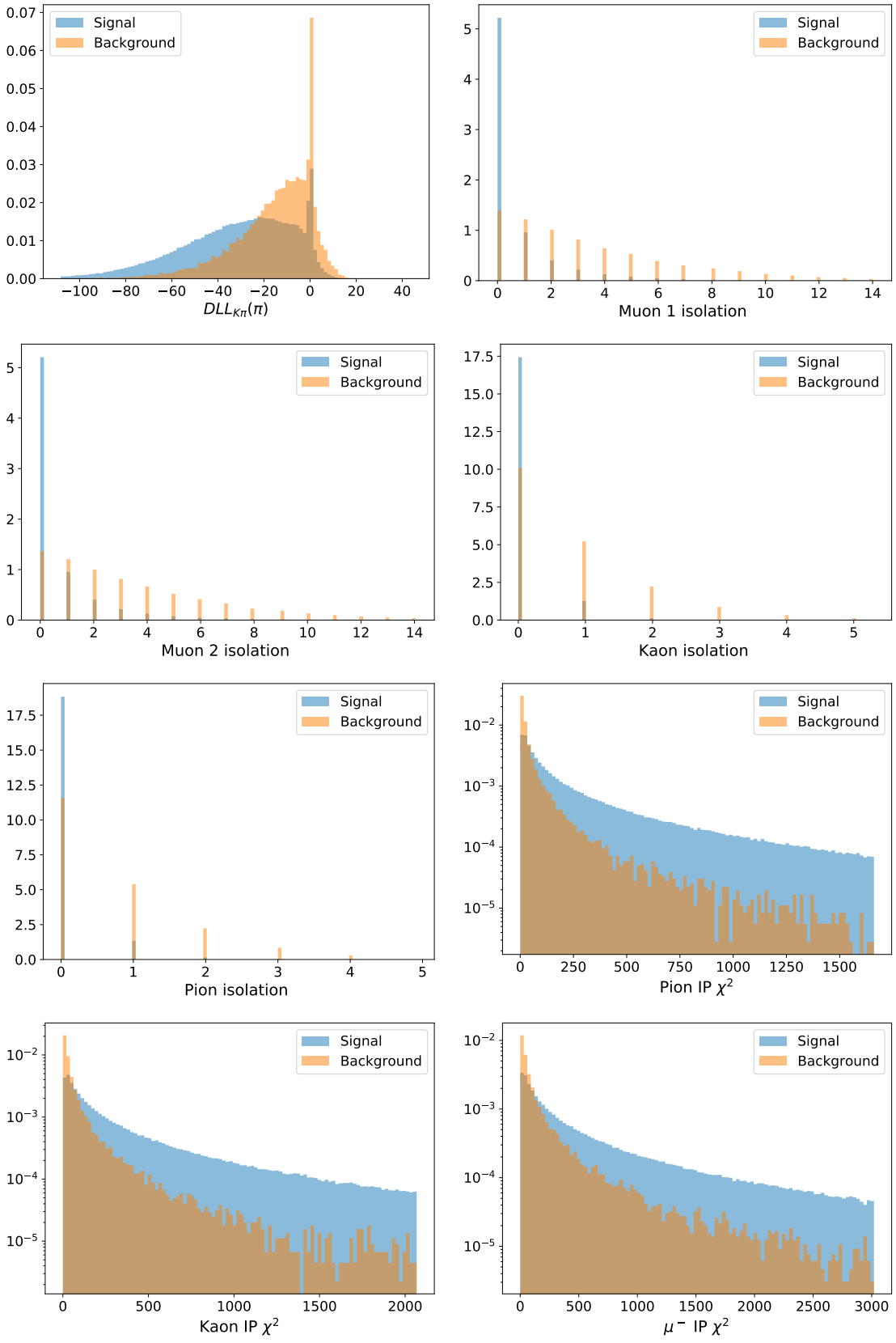


Figure 5.7: Training variables for the Run 2 combinatorial BDT. Signal is shown in blue and background is shown in orange. The distributions are normalised to unit area.

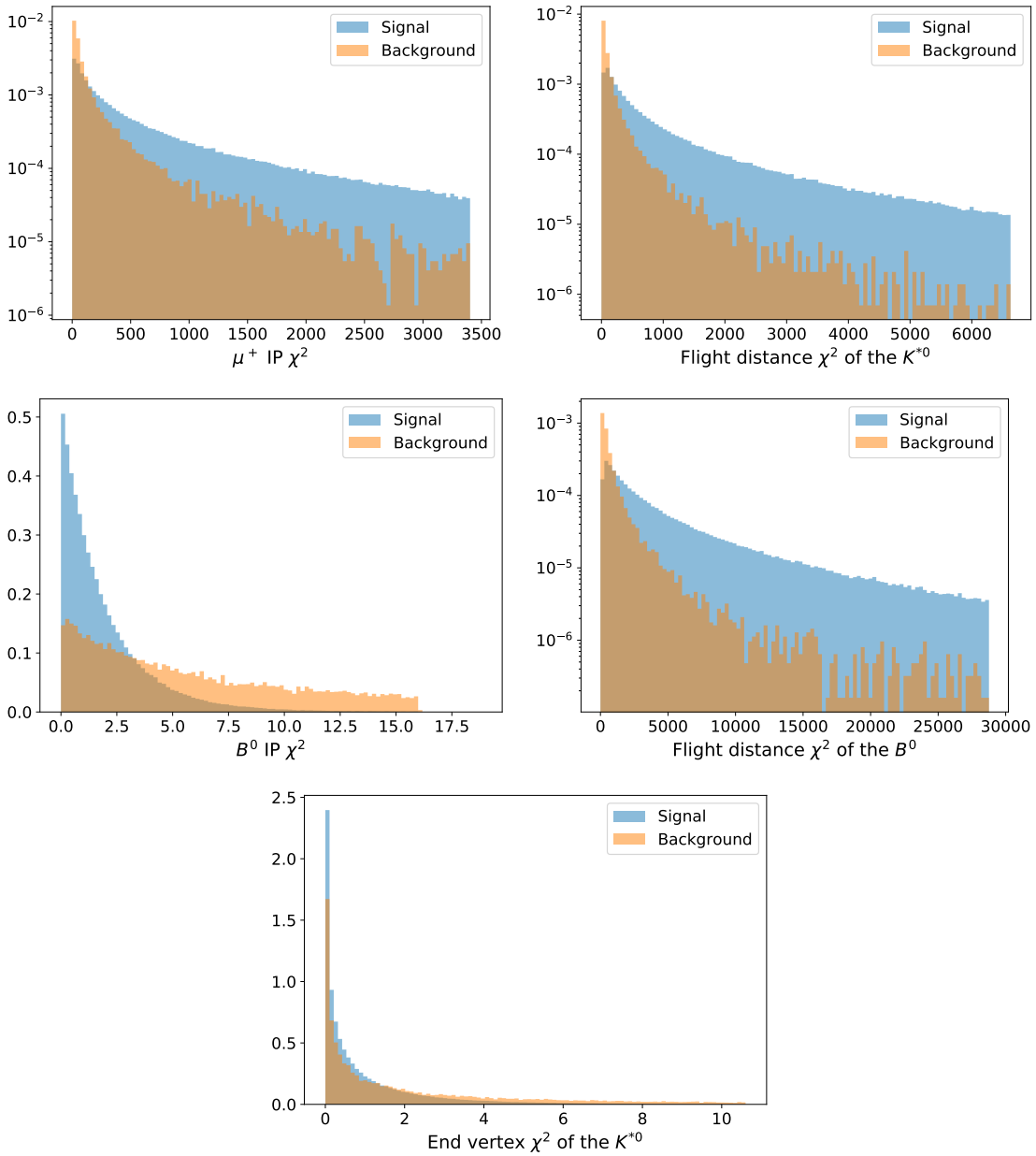


Figure 5.8: Training variables for the Run 2 combinatorial BDT. Signal is shown in blue and background is shown in orange. The distributions are normalised to unit area.

clearly performs well. The Receiver Operating Characteristic (ROC) curves for Run 1 and Run 2 are shown as 5.16. The high quality performance of the BDTs is due to the use of `xgboost`, a training sample which is representative of the signal (i.e. $B^0 \to K^{*0}\mu^+\mu^-$ simulation), and input variables which allow good discrimination between signal and background events.

The optimum cut points are chosen such that the statistical significance $S/\sqrt{S+B}$,

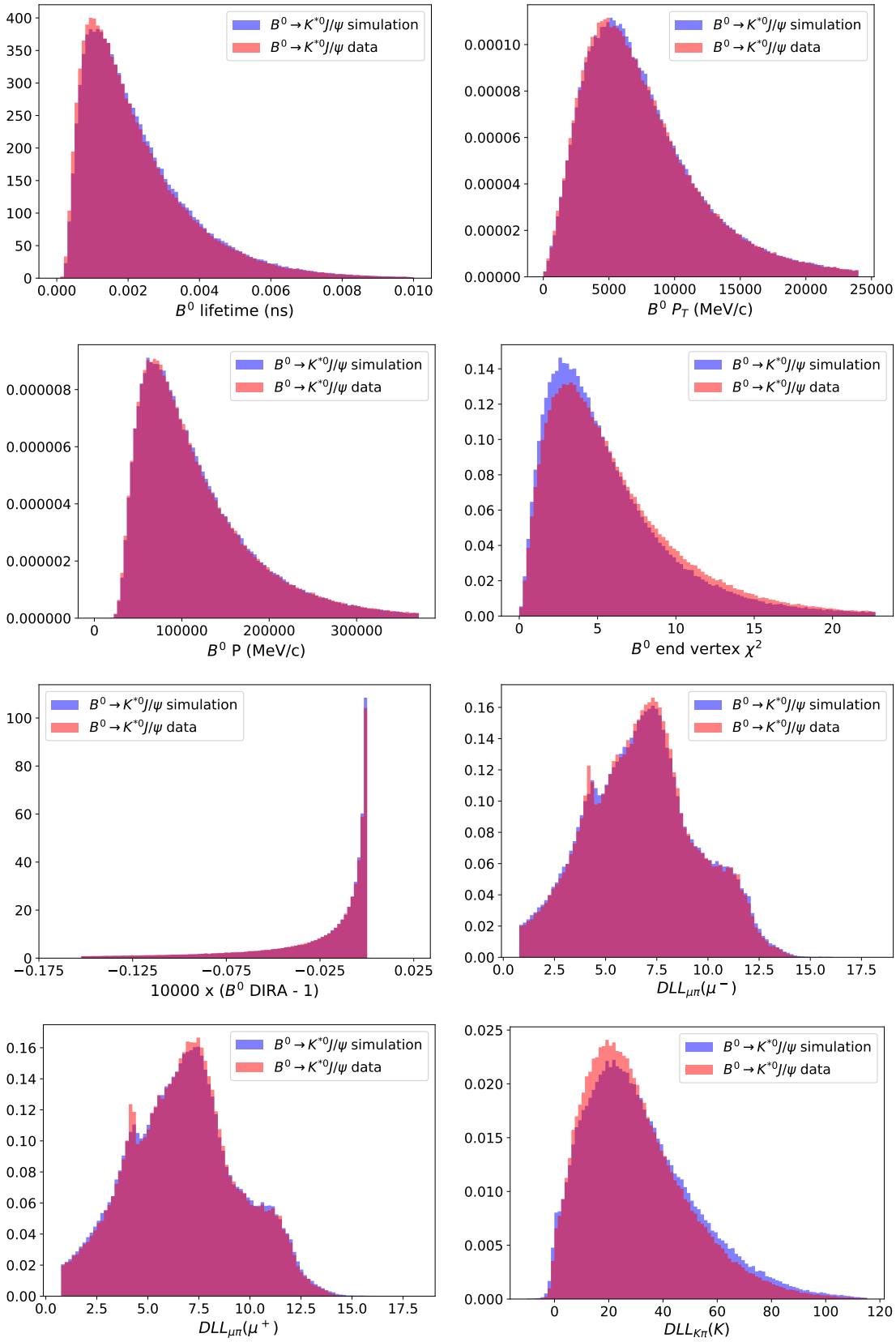


Figure 5.9: Comparison between $B^0 \to K^{*0}J/\psi$ simulation and sWeighted $B^0 \to K^{*0}J/\psi$ data for the BDT training variables for 2017. The distributions are normalised to unit area.

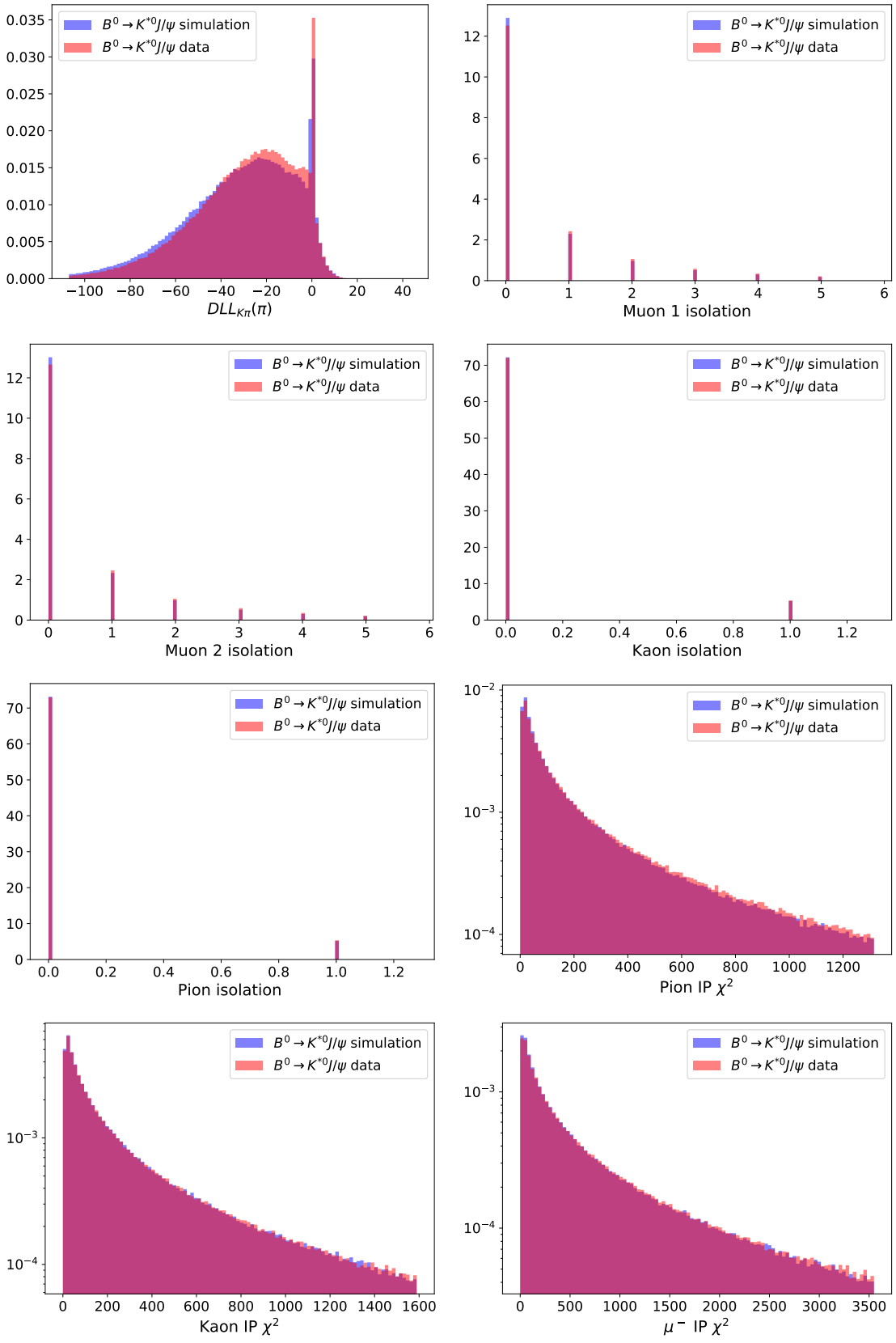


Figure 5.10: Comparison between $B^0 \to K^{*0} J/\psi$ simulation and sWeighted $B^0 \to K^{*0} J/\psi$ data for the BDT training variables for 2017. The distributions are normalised to unit area.

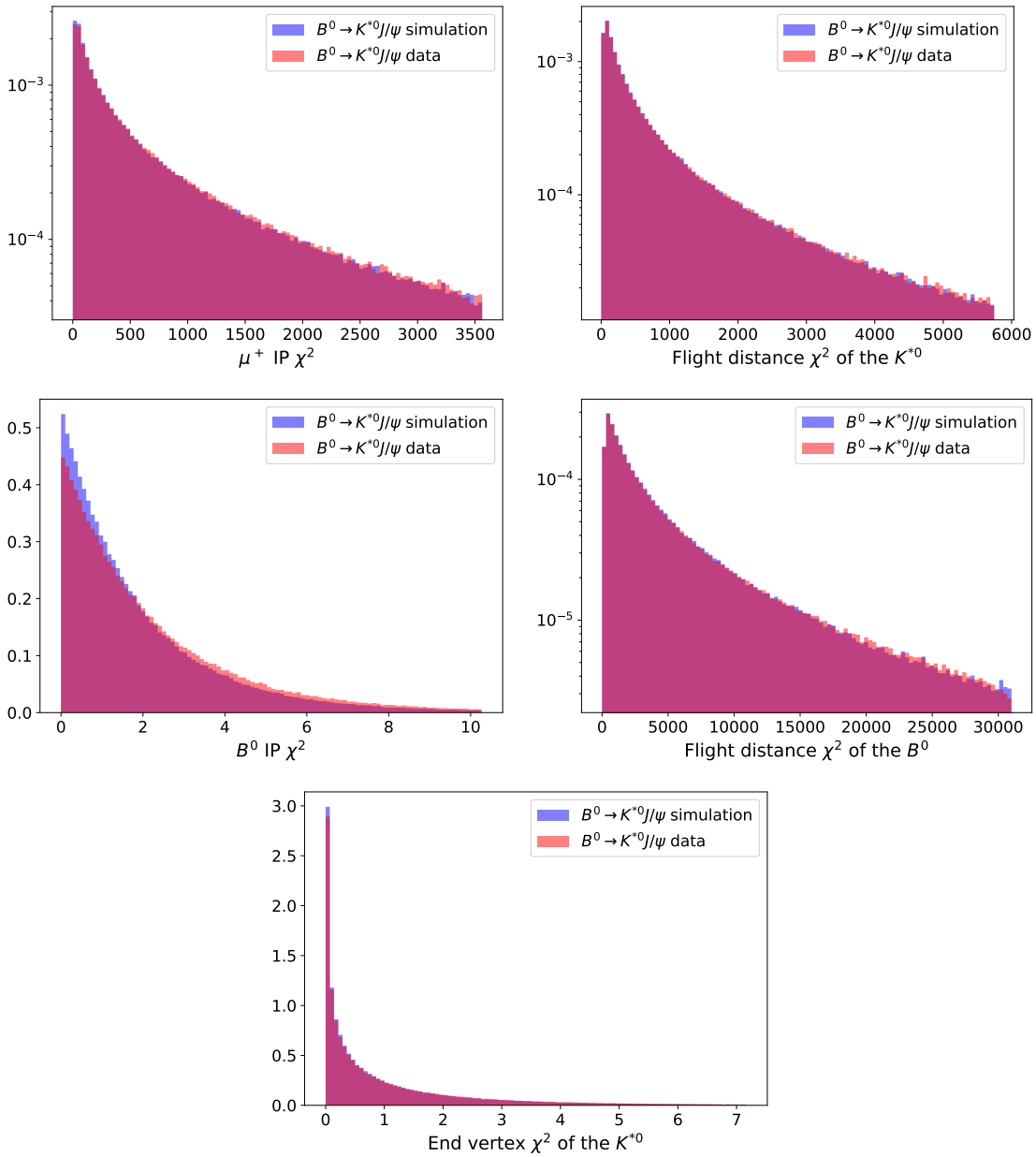


Figure 5.11: Comparison between $B^0 \to K^{*0} J/\psi$ simulation and sWeighted $B^0 \to K^{*0} J/\psi$ data for the BDT training variables for 2017. The distributions are normalised to unit area.

where S is the estimated signal yield, and B is the expected background yield, is maximised. The yield of the background is estimated directly from the upper B^0 mass sideband samples. They are fit with an exponential function and the yield is extrapolated into the full fit region ($5170 < m_B < 5700$ MeV/c^2). The signal yields are estimated by examining the simulation efficiencies and the $B^0 \to K^{*0} J/\psi$ data fits.

The significance $S/\sqrt{S+B}$ is maximised for the Run 1 and Run 2 BDTs. For Run 1,

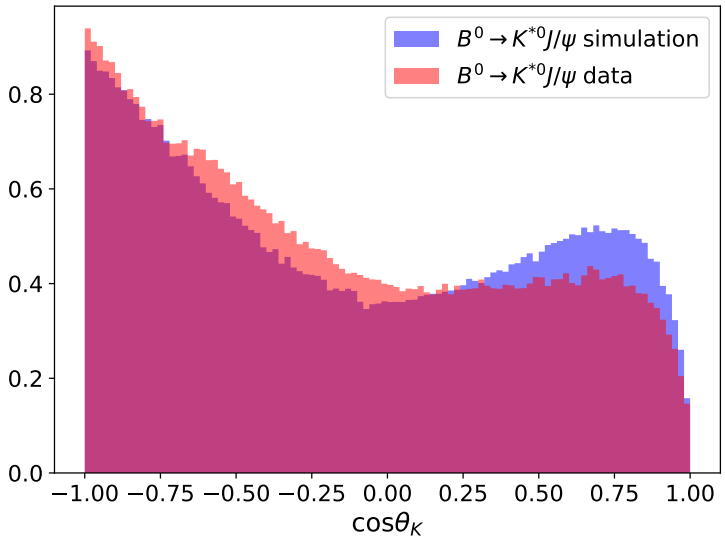


Figure 5.12: Comparison between $B^0 \to K^{*0} J/\psi$ simulation and sWeighted $B^0 \to K^{*0} J/\psi$ data for $\cos\theta_K$ for 2017. The deviations are caused by the presence of exotic states in the data.

Figure 5.17 shows the signal and background efficiencies, and significance as a function of BDT cut point. The same is shown for Run 2 in Figure 5.18. Table 5.8 shows the optimum BDT working points and efficiencies.

Table 5.8: Optimum combinatorial BDT working points and efficiencies.

Run Period	**BDT cut**	**Significance**	**Signal efficiency**	**Background rejection efficiency**
1	0.97672	30.4	0.924	0.984
2	0.97983	60.8	0.917	0.987

5.7 $B^+ \to K^{*+}\mu^+\mu^-$ BDT

Another BDT was trained, which was developed for the ongoing binned analysis, to further reduce background levels, in particular $B^+ \to K^{*+}\mu^+\mu^-$, where $K^{*+} \to K^+\pi^0$ or $K^{*+} \to K^0_S\pi^+$. Such backgrounds have angular structures which do not factorise with the B^0 mass. There are two BDTs, one for each decay mode, using `xgboost`. This BDT is trained on variables such as the χ^2_{IP} of the tracks, rapidities, difference between rapidity of the hadrons, and the K^{*0} and B^0 vertex χ^2. The cut points chosen for the BDT are such that $> 90\%$ of signal is retained.

Figure 5.19 shows $\cos\theta_\ell$ and $\cos\theta_K$ in the lower mass sideband, defined as the region

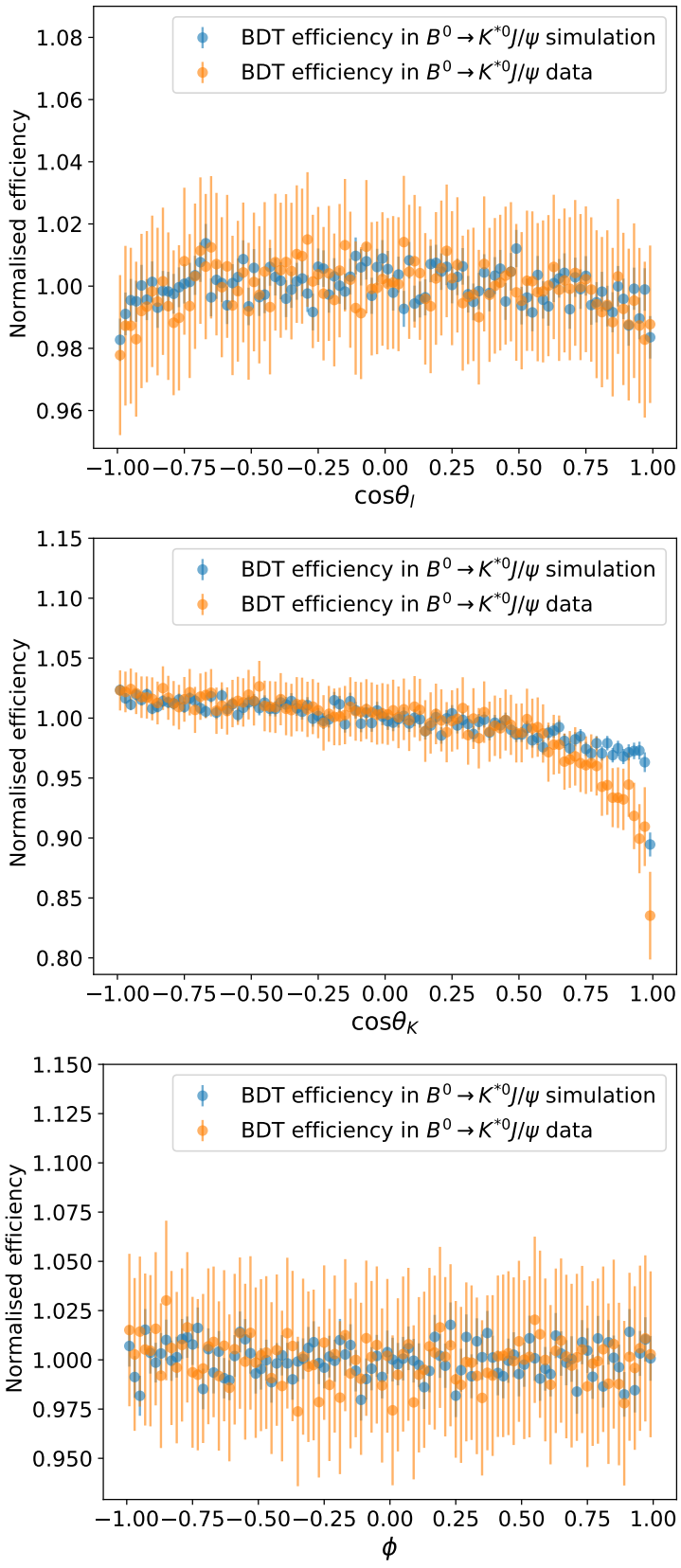


Figure 5.13: Combinatorial BDT efficiency as a function of $\cos\theta_\ell$, $\cos\theta_K$, and ϕ for 2017 $B^0 \to K^{*0} J/\psi$ simulation and 2017 $B^0 \to K^{*0} J/\psi$ data, where the uncertainties are statistical.

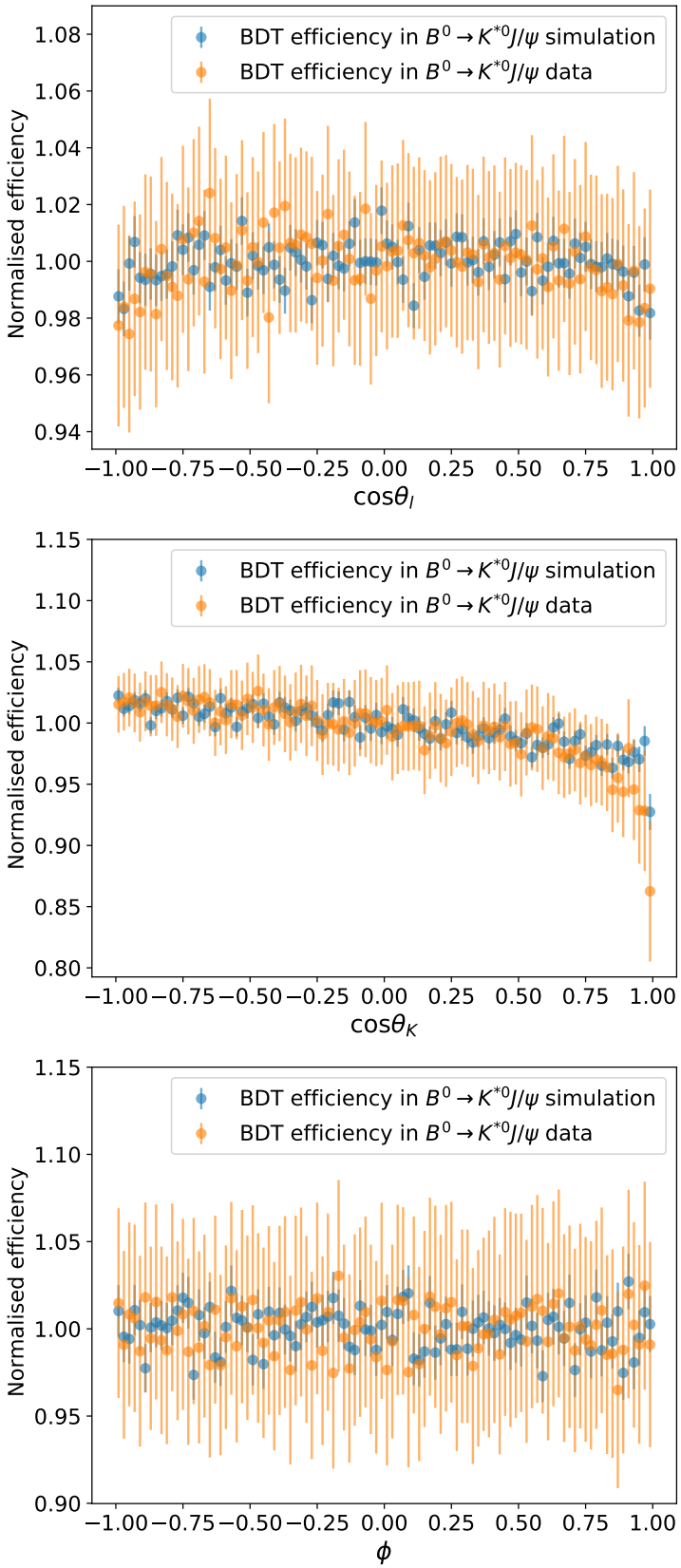


Figure 5.14: Combinatorial BDT efficiency as a function of $\cos\theta_\ell$, $\cos\theta_K$, and ϕ for 2017 $B^0 \to K^{*0} J/\psi$ simulation and 2017 $B^0 \to K^{*0} J/\psi$ data, in the region $876 < m_{K\pi} < 916$ MeV/c^2, where the uncertainties are statistical. This better aligns the data with the simulated P-wave state.

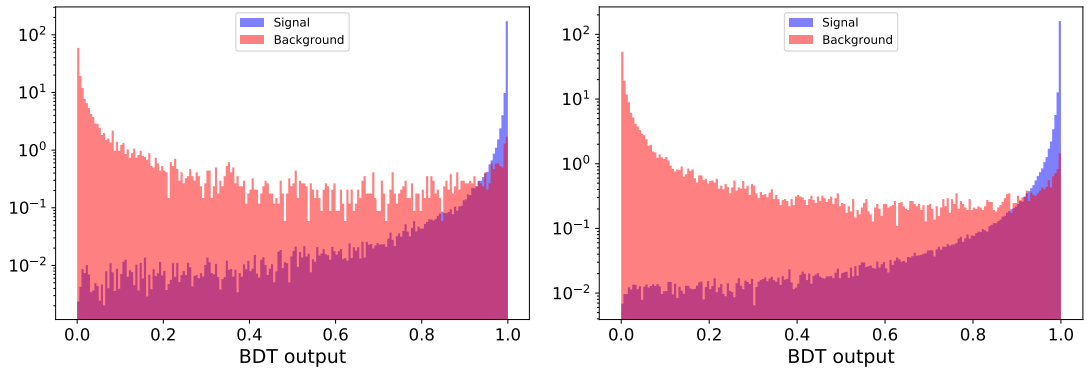


Figure 5.15: Result of the training of the combinatorial BDT for (left) Run 1 and (right) Run 2.

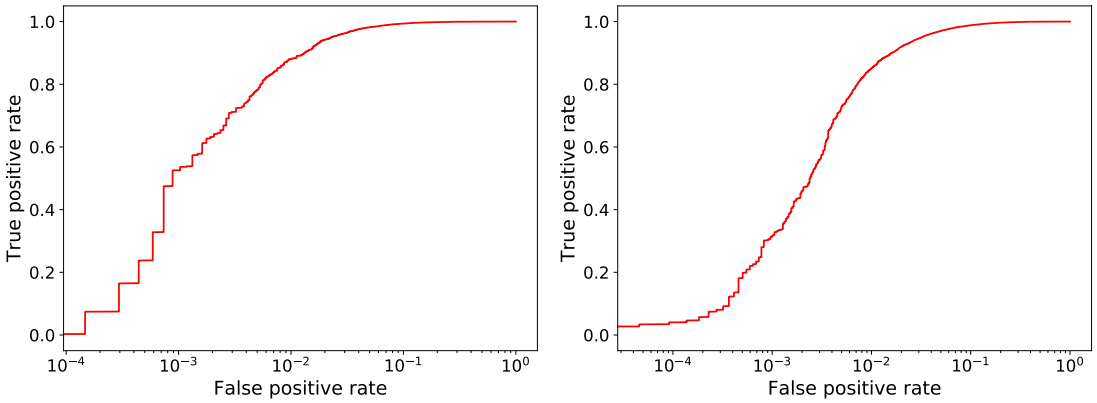


Figure 5.16: ROC curves from the training of the combinatorial BDT for (left) Run 1 and (right) Run 2. Note the x-axis uses a log scale.

$5170 < m_B < 5220$ MeV/c^2. Figure 5.20 shows $\cos\theta_\ell$ and $\cos\theta_K$ in the upper mass sideband, defined as the region $5350 < m_B < 5700$ MeV/c^2. Green and blue show the angles after applying the combinatorial and subsequently applying the $B^+ \to K^{*+}\mu^+\mu^-$ BDT, respectively.

Applied to the ansatz analysis selection, this BDT reduces peaking structures such as the one seen at high $\cos\theta_K$ in the lower mass sideband. This feature is predominantly from $B^+ \to K^{*+}\mu^+\mu^-$, where $K^{*+} \to K^+\pi^0$ or $K^{*+} \to K^0_S\pi^+$ and does not factorise with the B^0 mass. It is critical that the angles must factorise with each other and the B^0 mass since this is assumed by the background fit model. The $B^+ \to K^{*+}\mu^+\mu^-$ BDT significantly reduces this structure, thus improving the fit quality and reducing the possibility of a poor goodness-of-fit.

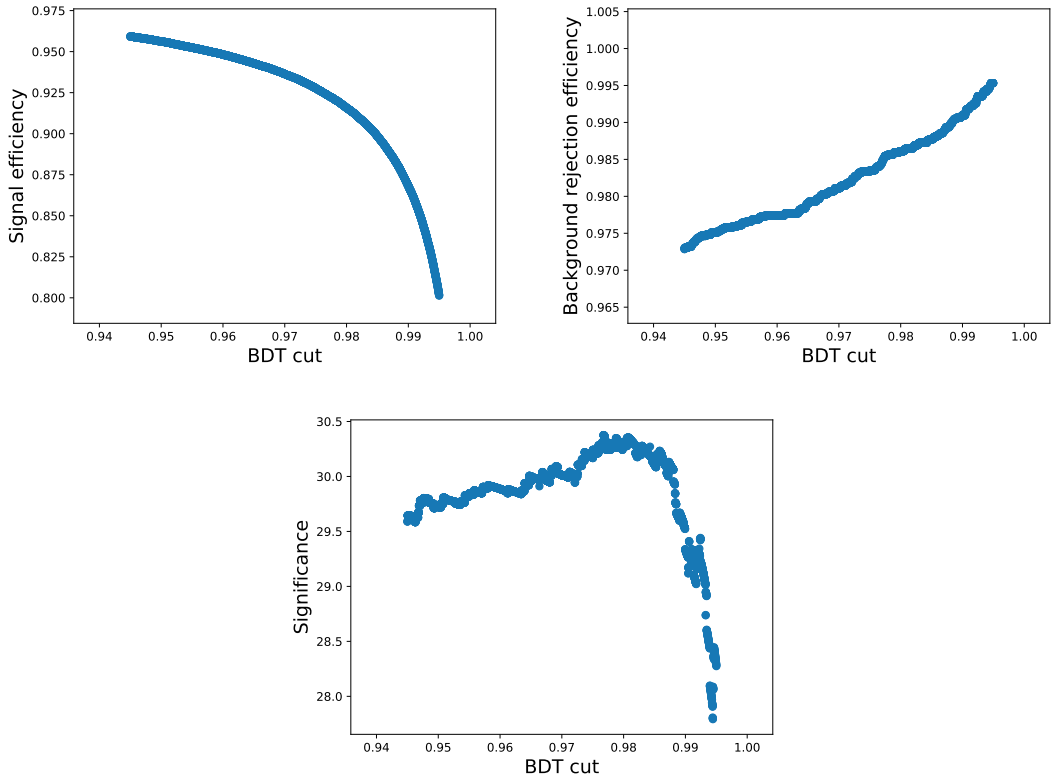


Figure 5.17: Signal efficiency, background rejection efficiency, and significance for the Run 1 BDT as a function of BDT cut point. The BDT cut point is chosen such that the significance is at a maximum.

5.8 Validation of the selection on the same-sign muon sample

The `B2XMuMu` stripping line also contains a sample of reconstructed decays where both muons have the same sign. This sample contains no signal and is an excellent proxy for combinatorial background in the full B^0 mass range, since the reconstructed decay is topologically similar to $B^0 \to K^{*0}\mu^+\mu^-$ and the same triggers and stripping are applied. This sample also provides a way to examine how the selection performs in the full B^0 mass range. This same-sign muon sample was used to examine the combinatorial BDT efficiency across the full B^0 mass range and study the factorisability between the angles and the B^0 mass.

Run 1 + Run 2 datasets were used from the `B2XMuXMu` stripping line, the same stripping line as the opposite sign sample. The triggers are also the same. Starting with $\sim$ 10 million events, the mass cuts (m_B, $m_{K\pi}$, and q^2) remove 93% of events. The peaking background vetoes, cleaning cuts and PID requirements remove a further 97% of events.

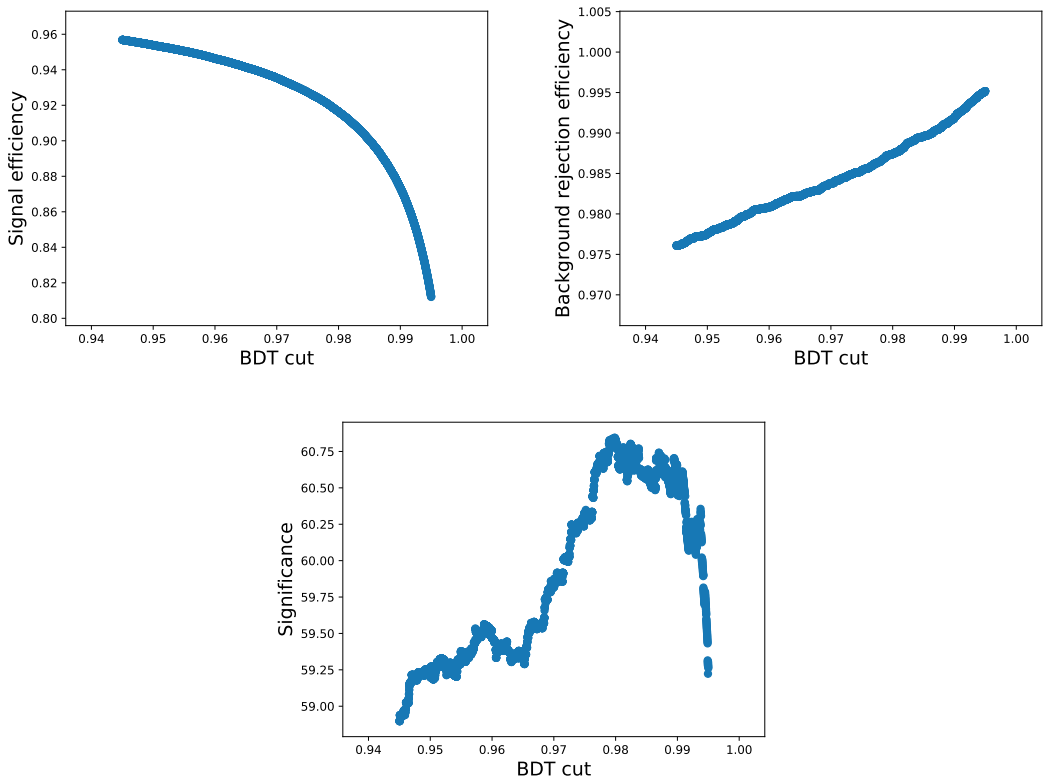


Figure 5.18: Signal efficiency, background rejection efficiency, and significance for the Run 2 BDT as a function of BDT cut point. The BDT cut point is chosen such that the significance is at a maximum.

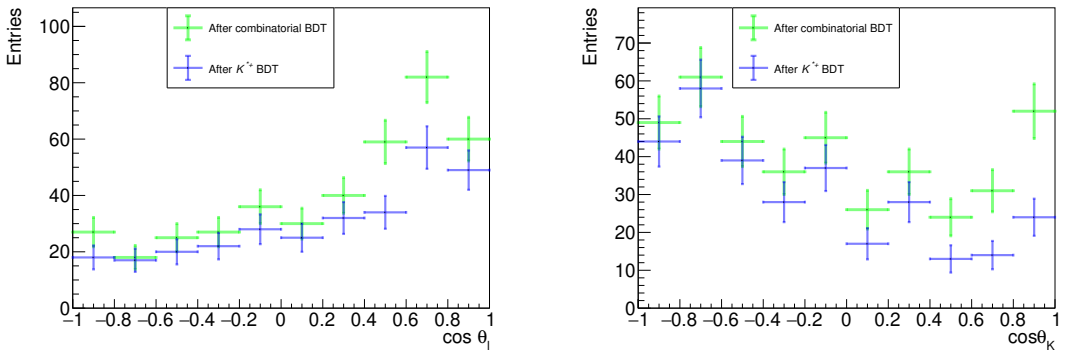


Figure 5.19: $\cos\theta_\ell$ and $\cos\theta_K$ in the lower B^0 mass sideband for all run periods, in the regions $1.25 < q^2 < 8$ GeV$^2/c^4$ and $11 < q^2 < 12.5$ GeV$^2/c^4$. Green and blue show the angles after applying the combinatorial and subsequently applying the $B^+ \to K^{*+}\mu^+\mu^-$ BDT, respectively. The $B^+ \to K^{*+}\mu^+\mu^-$ BDT reduces the overall amount of combinatorial background and helps to reduce peaking structures such as the one seen at high $\cos\theta_K$.

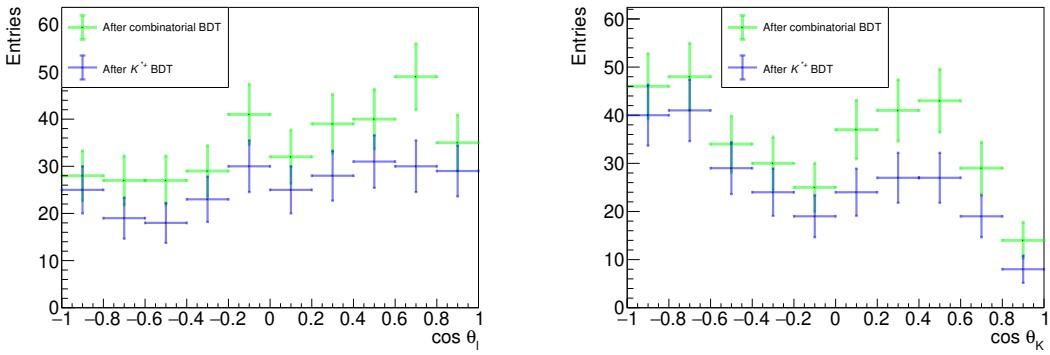


Figure 5.20: $\cos\theta_\ell$ and $\cos\theta_K$ in the upper mass B^0 sideband for all run periods, in the regions $1.25 < q^2 < 8$ GeV$^2/c^4$ and $11 < q^2 < 12.5$ GeV$^2/c^4$. Green and blue show the angles after applying the combinatorial and subsequently applying the $B^+ \to K^{*+}\mu^+\mu^-$ BDT, respectively. The $B^+ \to K^{*+}\mu^+\mu^-$ BDT reduces the overall amount of combinatorial background.

The combinatorial BDT removes 98% of the remaining data resulting in 391 events.

5.8.1 BDT efficiency

The BDT efficiency was studied across the full B^0 mass range in the same-sign muon sample. Figure 5.21 shows the B^0 mass before and after applying the combinatorial BDT (blue and orange respectively), in the range $4900 < m_B < 6100$ MeV/c^2. The efficiency of the combinatorial BDT is found by dividing these two histograms, as shown by Figure 5.22. Within statistical uncertainties, the BDT efficiency is a smooth function.

The correlation between the BDT score and the B^0 mass was studied. The histograms of BDT score against B^0 mass are shown as Figure 5.23. A correlation of -0.04247 was measured between BDT score and B^0 mass in the region $5170 < m_B < 5700$ MeV/c^2. This is evident from looking at the BDT efficiency. Given no significant peaks are seen in the B^0 mass, the main check which needs to be performed is whether the B^0 mass distribution can be modelled with an exponential function, which is the nominal background description in the B^0 mass in the fit to opposite-sign data.

The $B^+ \to K^{*+}\mu^+\mu^-$ BDT was also applied to the same-sign sample and B^0 mass distributions were plotted in the two fit regions, $1.25 < q^2 < 8$ GeV$^2/c^4$ and $11 < q^2 < 12.5$ GeV$^2/c^4$. These are shown by the blue histograms in Figures 5.24 and 5.25 respectively. These distributions were fitted with exponential functions and chi-square tests were performed. The chi-square values for the lower and upper q^2 regions are 4.11 and 10.26.

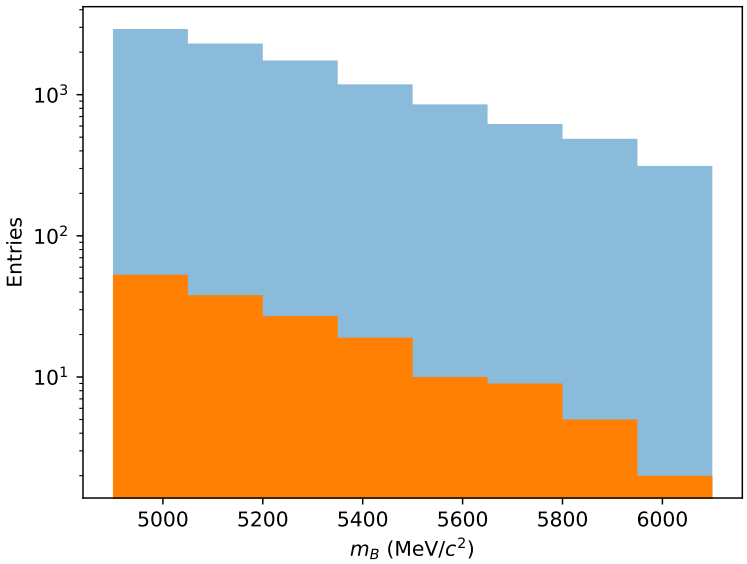


Figure 5.21: B^0 mass distribution in the same-sign muon sample before (blue) and after (orange) applying the combinatorial BDT in the region $4900 < m_B < 6100$ MeV/c^2.

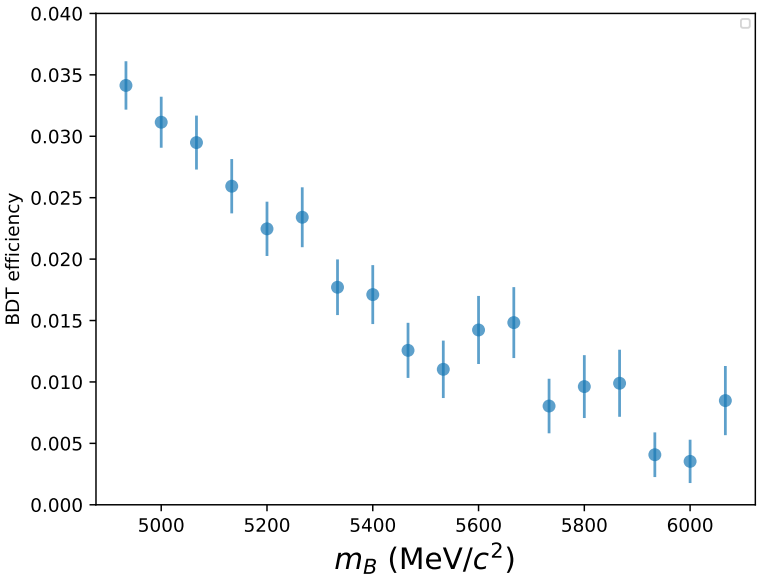


Figure 5.22: Combinatorial BDT efficiency on the same-sign muon sample in the region $4900 < m_B < 6100$ MeV/c^2.

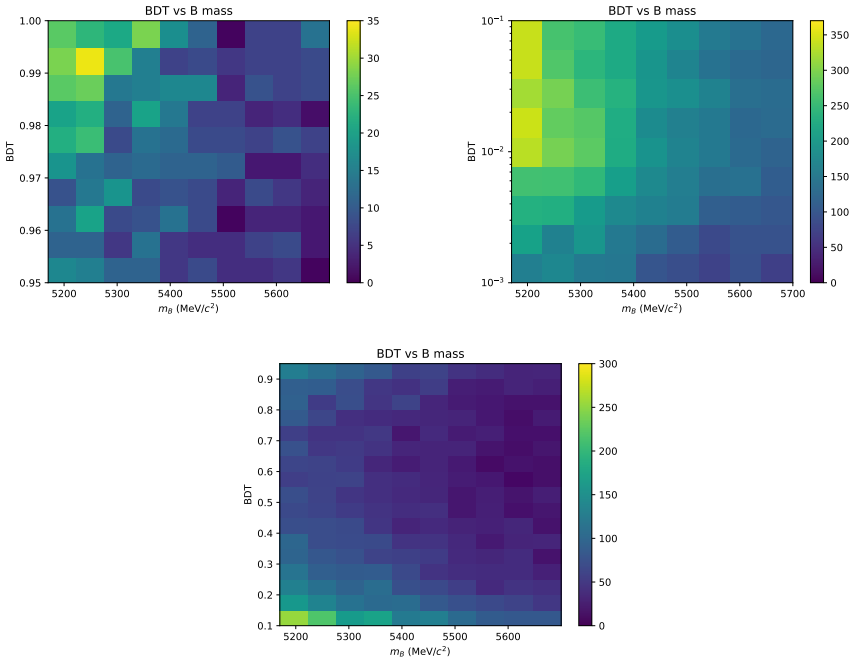


Figure 5.23: Correlation between BDT and B^0 mass in the same-sign muon sample in $BDT > 0.95$ (top left), $0.001 < BDT < 0.1$ (top right) and $0.1 < BDT < 0.95$ (bottom).

With 6 degrees of freedom, this results in p-values of 66.15% and 11.42% respectively. Thus the B^0 mass distributions of the same-sign muon sample after applying the full selection can be described by exponential functions. Given this and the good-quality goodness-of-fit presented in Chapter 11, the background component of the B^0 mass in the opposite-sign data is describable using an exponential function. For the main analysis which uses the opposite-sign data, the decay constant in the exponential function is a free parameters in the fit.

5.8.2 Background factorisability

In the fit, the background is parameterised as factorising between the B^0 mass, each angle, and q^2, i.e. $P_{bkg} = P(m_B)P(\cos\theta_\ell)P(\cos\theta_K)P(\phi)P(q^2)$. In particular, the B^0 mass factorises with the angular PDF, i.e. $P_{bkg} = P(m_B) \times P(\Omega, q^2)$.

Given that there are backgrounds which do factorise in the B^0 mass, it is necessary to examine the factorisation of the combinatorial, which can be done with the same-sign muon sample. If they do not factorise, one has to improve the background description in order to incorporate non-factorisable effects, or improve the selection. Again, the dataset

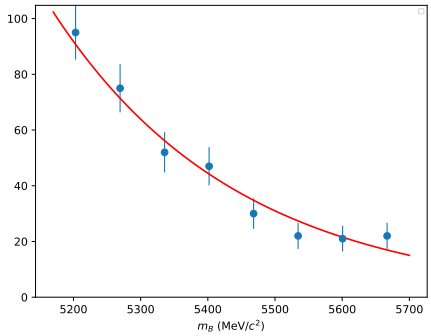


Figure 5.24: B^0 mass distribution in the same-sign muon sample with the full selection applied, in the region $1.25 < q^2 < 8$ GeV$^2/c^4$ (blue). From fitting an exponential function to this distribution (red), a p-value of 66.15% is determined from a chi-square test.

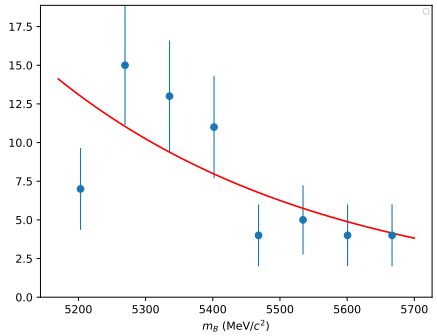


Figure 5.25: B^0 mass distribution in the same-sign muon sample with the full selection applied, in the region $11 < q^2 < 12.5$ GeV$^2/c^4$ (blue). From fitting an exponential function to this distribution (red), a p-value of 11.42% is determined from a chi-square test.

was split into the two q^2 regions ($1.25 < q^2 < 8$ GeV$^2/c^4$ and $11 < q^2 < 12.5$ GeV$^2/c^4$) and treated entirely separately.

Legendre polynomials are fitted to the angular distributions in the full B^0 mass range. For the lower q^2 region, two parameters were fitted for each angular variable $\cos\theta_\ell$, $\cos\theta_K$, and ϕ. The parameterisation is of the form $1 + aL_1 + bL_2$ where a, b are fit parameters and L_i are Legendre polynomials of order i. The angle ϕ is scaled to the region $-1 \leq x \leq 1$ when evaluating the polynomial. Owing to the low amount of data in the interresonance region ($11 < q^2 < 12.5$ GeV$^2/c^4$), only one parameter was fitted, i.e. a parameterisation of the form $1 + aL_1$.

The B^0 mass region was also split into two bins of B^0 mass, and the angular fits were performed. Each B^0 mass bin contains the same number of events. The coefficients can be compared between B^0 mass bins for the lower q^2 region or the interresonance region. The same study was also performed where the B^0 mass was split into three bins of equal size.

For the lower q^2 region, the results for the a parameter for the angles are shown as Figure 5.26. The results for the b parameter for the angles are shown as Figure 5.27. Within the data sample statistics, the coefficients are consistent.

For the interresonance region, the results for the a parameter for the angles are shown as Figure 5.28. The coefficients are consistent within each other given the data statistics.

The angles and fits in the full B^0 mass range and the three bins in B^0 mass are plotted.

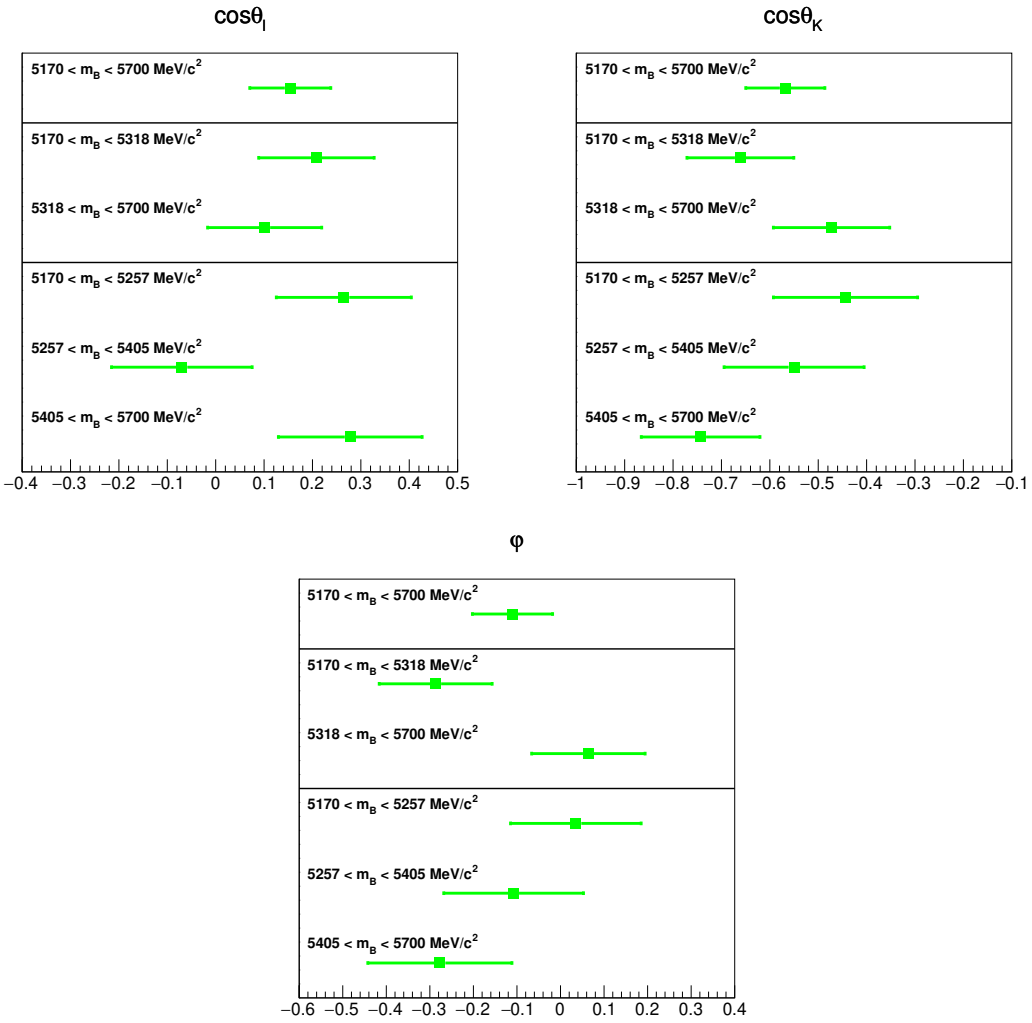


Figure 5.26: Results from fits to the angles in the low q^2 region in the selected same-sign muon sample. This plot shows the first Legendre coefficient for the three angles. The fit was performed in the full B^0 mass range, two B^0 mass regions of equal size, and three B^0 mass regions of equal size.

This is in order to visually examine the quality of the fits and the shape of the backgrounds in different B mass regions. These are shown as Figure 5.29 for $\cos\theta_\ell$, $\cos\theta_K$ and ϕ. Likewise, for the interresonance region, these are shown as Figure 5.30. Within the data statistics, the backgrounds indeed appear to factorise in B^0 mass.

Chi-square goodness of fit tests were performed on the fits where the B^0 mass is split into three bins. The fit in each B^0 mass region is compared to the data in the other two regions in order to examine the compatibility between the angular shapes between different B^0 mass regions. The reason for not comparing the fit to the data in all three regions is to avoid correlations between the fit and data used to perform the fit. The p-values are

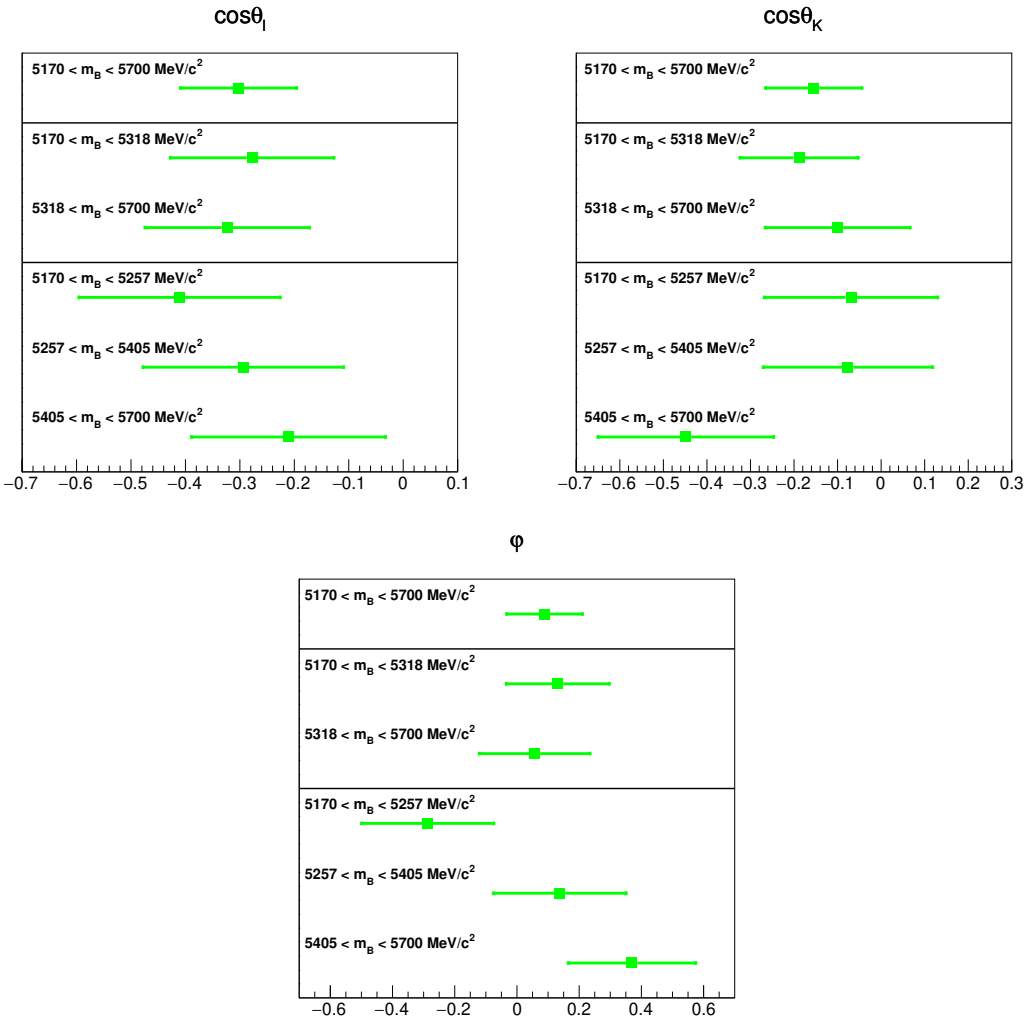


Figure 5.27: Results from fits to the angles in the low q^2 region in the selected same-sign muon sample. This plot shows the second Legendre coefficient for the three angles. The fit was performed in the full B^0 mass range, two B^0 mass regions of equal size, and three B^0 mass regions of equal size.

shown in Table 5.9.

5.9 Performance of the selection

The performance of the selection is examined by computing expected peaking background yields relative to the signal. This was examined at various points in the selection. Figure 5.31 shows the expected peaking background yields in the two signal regions after the trigger, stripping and all cleaning cuts apart from the hadron PID has been applied. The sum of all peaking backgrounds relative to the signal is large, at 85% in the low q^2 region

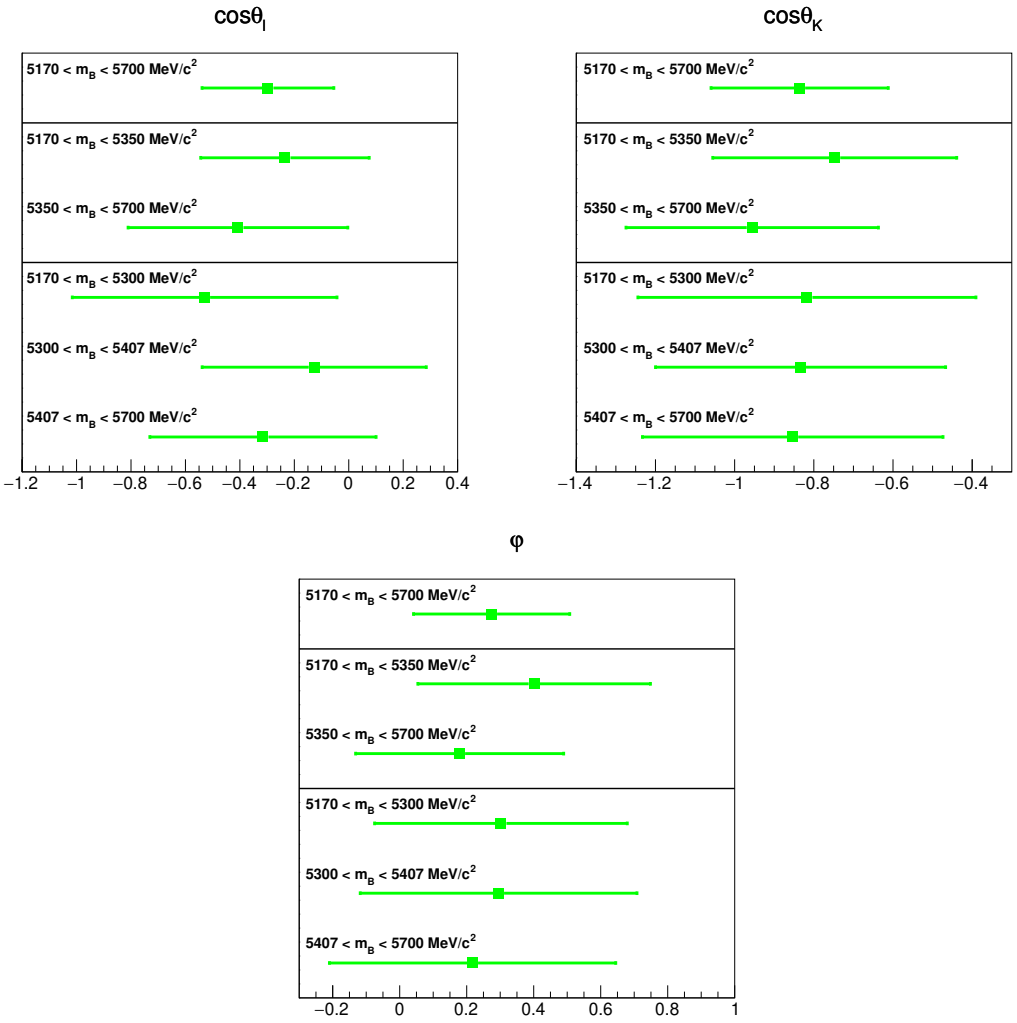


Figure 5.28: Results from fits to the angles in the interresonance region in the selected same-sign muon sample. This plot shows the first Legendre coefficient for the three angles. The fit was performed in the full B^0 mass range, two B^0 mass regions of equal size, and three B^0 mass regions of equal size.

and 70% in the interresonance region. The largest background is the K and π mis-ID in $B^0 \to K^{*0}\mu^+\mu^-$. Figure 5.32 shows the same after the hadron PID selection is applied. Applying the hadron PID requirements reduces the overall levels of peaking backgrounds in both q^2 regions, towards 12% in the low q^2 region and 5% in the interresonance region. The largest backgrounds in the two regions are now the μ-π swap from $B^0 \to K^{*0}J/\psi$ and $K \to \pi$ mis-ID in $B_s^0 \to \phi\mu^+\mu^-$ respectively. These are reduced significantly by the peaking background veto BDTs discussed in Section 5.4. Figure 5.33 shows the expected peaking background yields relative to the signal after the peaking background veto BDTs have been applied. These have reduced overall levels of background to around 1.7% of the

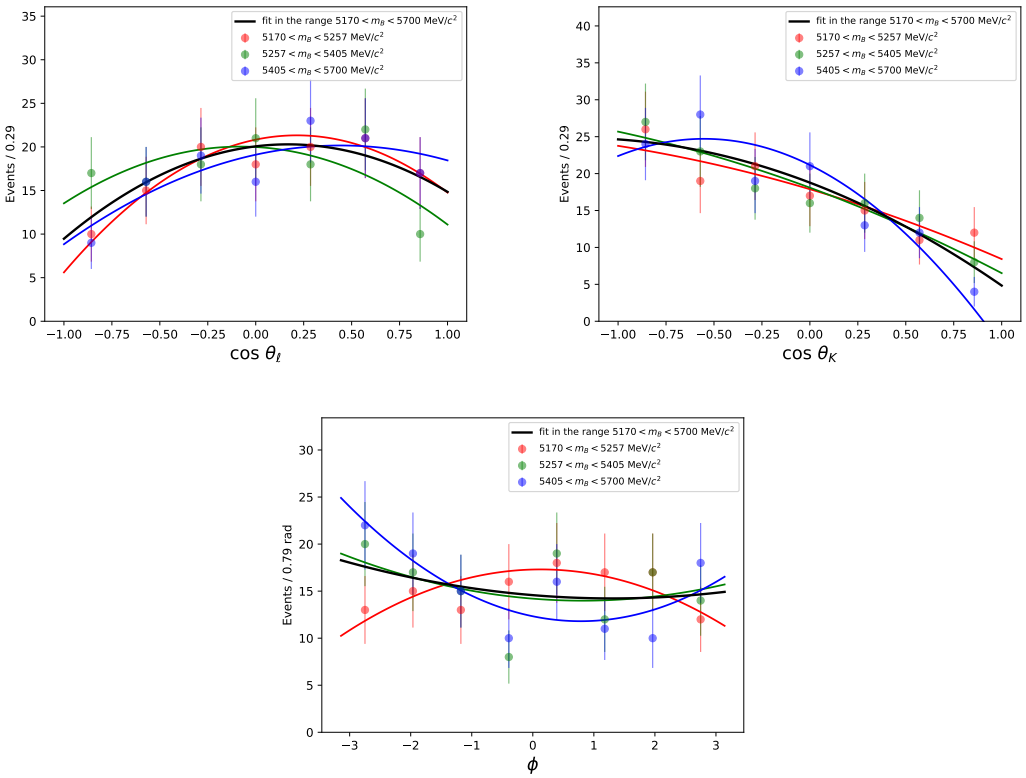


Figure 5.29: $\cos\theta_\ell$, $\cos\theta_K$ and ϕ in different B^0 mass regions in the same-sign muon sample in the low q^2 region with fits to second order Legendre polynomials.

Table 5.9: p-values when comparing fits to angles in a B^0 mass region to the distributions in the other B^0 mass regions.

Angle	q^2 **region** ($\mathbf{GeV}^2/c^4$)	B^0 **mass region** ($\mathbf{MeV}/c^2$)	**p-value (%)**
$\cos\theta_\ell$	$1.25 < q^2 < 8$	$5170 < m_B < 5257$	84.7
$\cos\theta_\ell$	$1.25 < q^2 < 8$	$5257 < m_B < 5405$	58.9
$\cos\theta_\ell$	$1.25 < q^2 < 8$	$5405 < m_B < 5700$	41.3
$\cos\theta_K$	$1.25 < q^2 < 8$	$5170 < m_B < 5257$	41.3
$\cos\theta_K$	$1.25 < q^2 < 8$	$5257 < m_B < 5405$	72.6
$\cos\theta_K$	$1.25 < q^2 < 8$	$5405 < m_B < 5700$	23.0
ϕ	$1.25 < q^2 < 8$	$5170 < m_B < 5257$	19.8
ϕ	$1.25 < q^2 < 8$	$5257 < m_B < 5405$	19.3
ϕ	$1.25 < q^2 < 8$	$5405 < m_B < 5700$	8.6
$\cos\theta_\ell$	$11 < q^2 < 12.5$	$5170 < m_B < 5300$	23.1
$\cos\theta_\ell$	$11 < q^2 < 12.5$	$5300 < m_B < 5407$	50.9
$\cos\theta_\ell$	$11 < q^2 < 12.5$	$5407 < m_B < 5700$	46.8
$\cos\theta_K$	$11 < q^2 < 12.5$	$5170 < m_B < 5300$	6.2
$\cos\theta_K$	$11 < q^2 < 12.5$	$5300 < m_B < 5407$	39.6
$\cos\theta_K$	$11 < q^2 < 12.5$	$5407 < m_B < 5700$	48.2
ϕ	$11 < q^2 < 12.5$	$5170 < m_B < 5300$	18.2
ϕ	$11 < q^2 < 12.5$	$5300 < m_B < 5407$	75.8
ϕ	$11 < q^2 < 12.5$	$5407 < m_B < 5700$	13.3

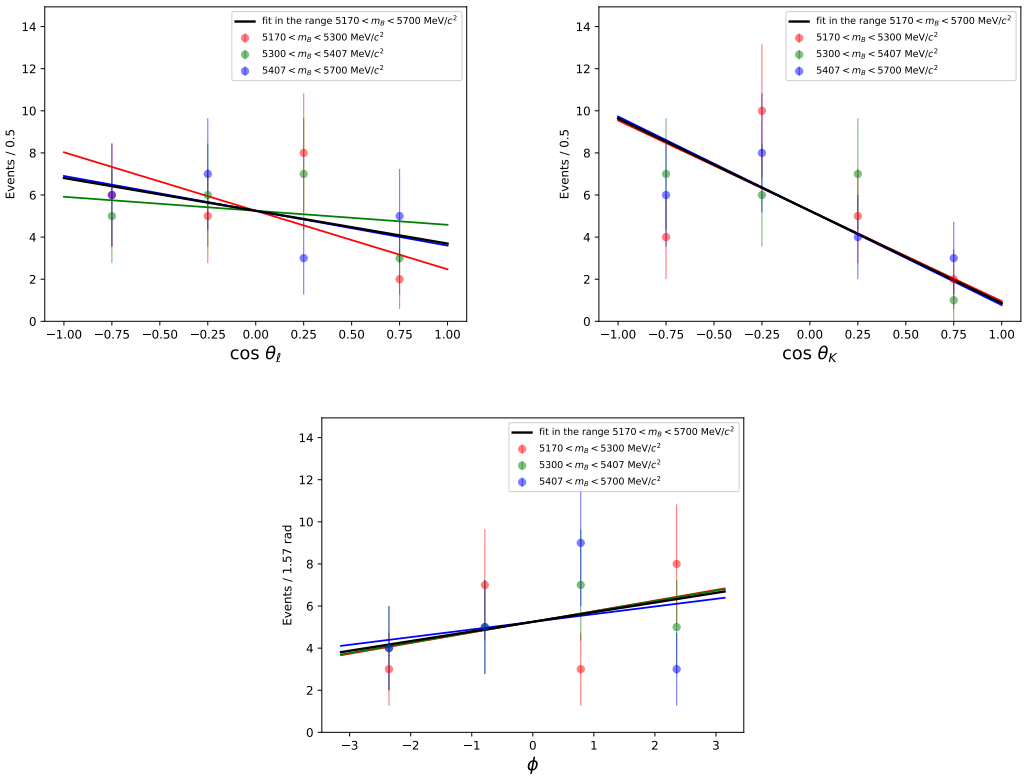


Figure 5.30: $\cos\theta_\ell$, $\cos\theta_K$ and ϕ in different B^0 mass regions in the same-sign muon sample in the interresonance region with fits to first order Legendre polynomials.

signal. Now $B^+ \to K^{*+}\mu^+\mu^-$ is largest in both regions. Figures 5.34 and 5.35 show the expected peaking background yields relative to the signal after applying the combinatorial BDT and subsequently applying the K^{*+} BDT respectively. The resulting overall levels of peaking background are low (0.9% and 0.4% of the signal in the two q^2 regions) thus overall levels of backgrounds will have a negligible effect on the signal. In addition there is no single peaking background considered here which is greater than 0.3% of the signal.

Figure 5.36 shows the B^0 mass distributions for the signal simulation and the background simulation. The background simulation corresponds to the total background, obtained by adding together the individual background components scaled to the expected yields relative to the signal. The top plots show the low q^2 region and the bottom plots show the interresonance region. The left plots show the B^0 mass after the trigger, stripping, and cleaning cuts (including hadron PID) are applied and the right plots show the B^0 mass after the full selection is applied. The same is shown for $\cos\theta_\ell$, $\cos\theta_K$ and ϕ in Figures 5.37, 5.38, and 5.39. As seen in these plots, the background is peaking in certain angular and/or B^0

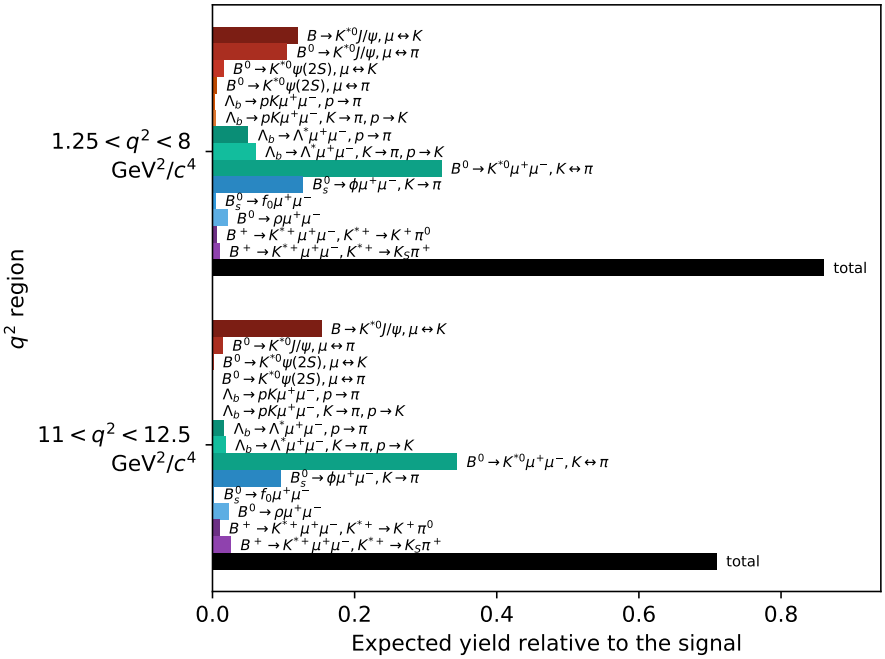


Figure 5.31: Expected peaking background yields relative to the signal in the two q^2 regions of interest, for the full dataset. This is after the trigger, stripping and cleaning cuts apart from the hadron PID are applied. The sum of all peaking backgrounds relative to the signal is 85% in the low q^2 region and 70% in the interresonance region.

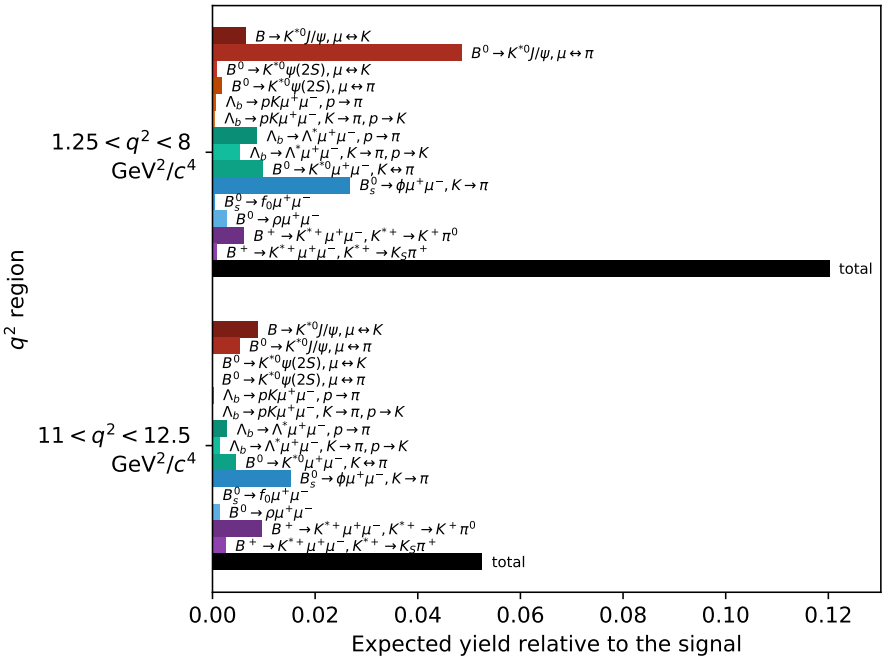


Figure 5.32: Expected peaking background yields relative to the signal in the two q^2 regions of interest, for the full dataset. This is after the trigger, stripping and cleaning cuts (including hadron PID) are applied. The sum of all peaking backgrounds relative to the signal is 12% in the low q^2 region and 5% in the interresonance region.

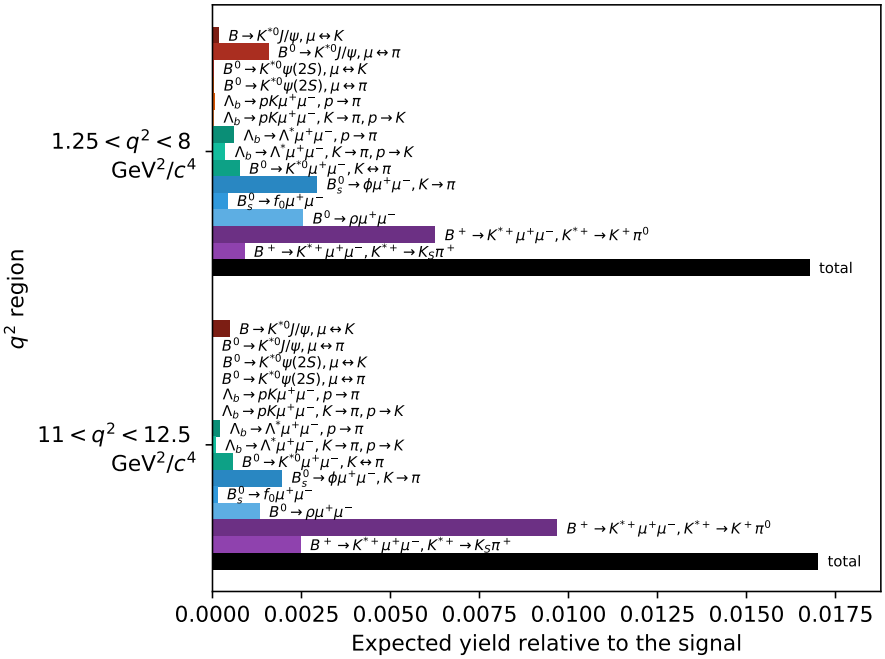


Figure 5.33: Expected peaking background yields relative to the signal in the two q^2 regions of interest, for the full dataset. This is after the trigger, stripping, cleaning cuts, and peaking background veto BDTs are applied. The sum of all peaking backgrounds relative to the signal is 1.7% in both q^2 regions.

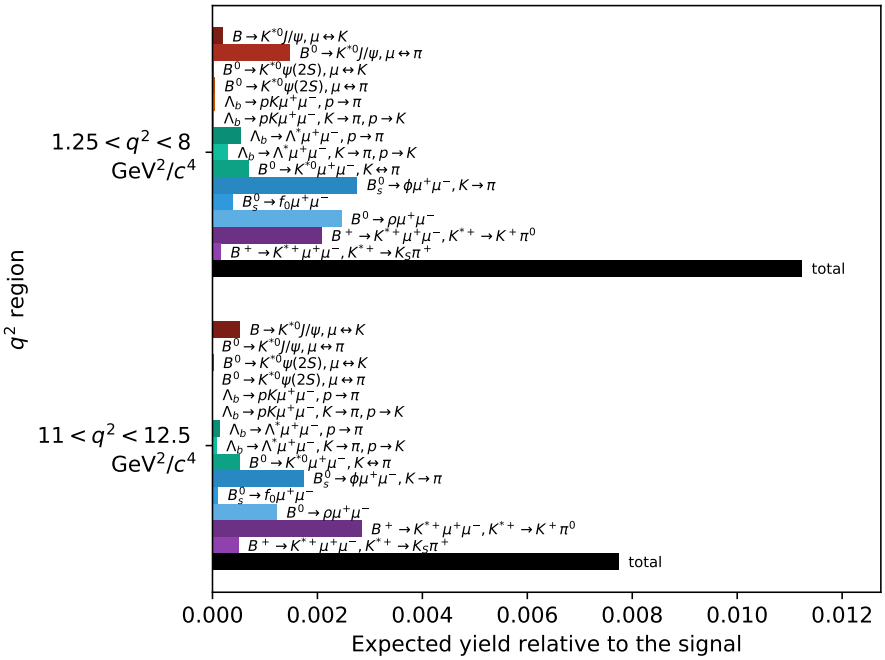


Figure 5.34: Expected peaking background yields relative to the signal in the two q^2 regions of interest, for the full dataset. This is after the trigger, stripping, cleaning cuts, peaking background veto BDTs and combinatorial BDT are applied. The sum of all peaking backgrounds relative to the signal is 1.1% in the low q^2 region and 0.8% in the interresonance region.

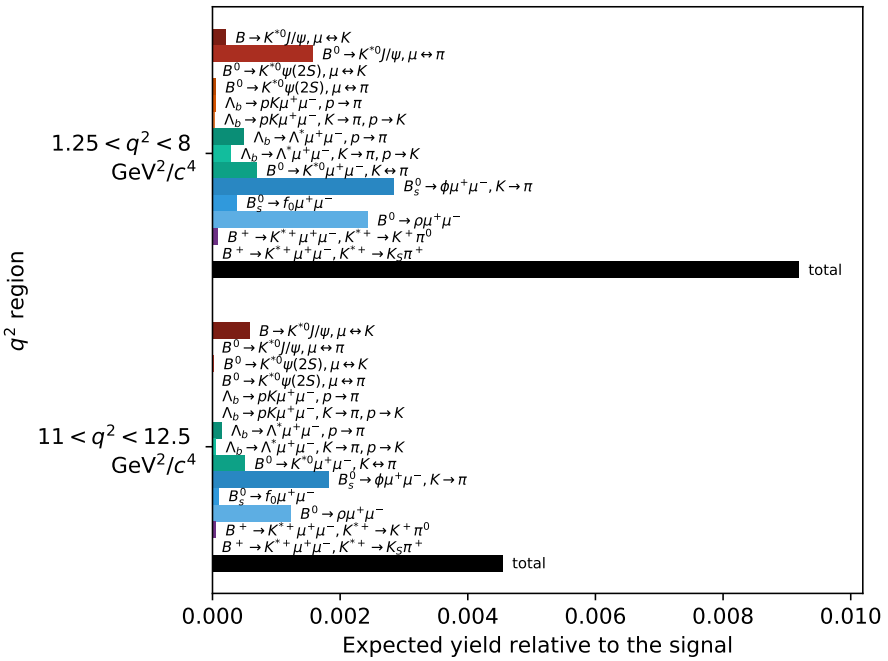


Figure 5.35: Expected peaking background yields relative to the signal in the two q^2 regions of interest, for the full dataset. This is with all of the selection applied. The sum of all peaking backgrounds relative to the signal is 0.9% in the low q^2 region and 0.4% in the interresonance region.

mass regions. Thus it is important to remove these backgrounds. The selection described in this chapter is very powerful at removing these peaking backgrounds as well as overall levels of combinatorial background resulting in a clean sample which can be used for the amplitude ansatz analysis of $B^0 \to K^{*0}\mu^+\mu^-$. The expected yields are estimated. After the selection is applied, in the $1.25 < q^2 < 8$ GeV$^2/c^4$ region, the expected yields are 4167 signal events, 726 combinatorial background events, and 38 combined peaking background events.

Since the overall level of peaking backgrounds is low compared to the signal, and that the signal data fit shows a good-quality goodness-of-fit, as shown in Section 11.3.1, the effect of the remaining peaking backgrounds on the signal is deemed to be negligible. With regards to the combinatorial background, since it can be described by an exponential function in the B^0 mass, unlike the signal which is described by the sum of two Crystal Ball functions, the background is thus uncorrelated with the signal.

Table 5.10 presents a summary of the applied selections with efficiencies. The efficiencies were determined from $B^0 \to K^{*0}\mu^+\mu^-$ simulation. The efficiencies shown for each selection

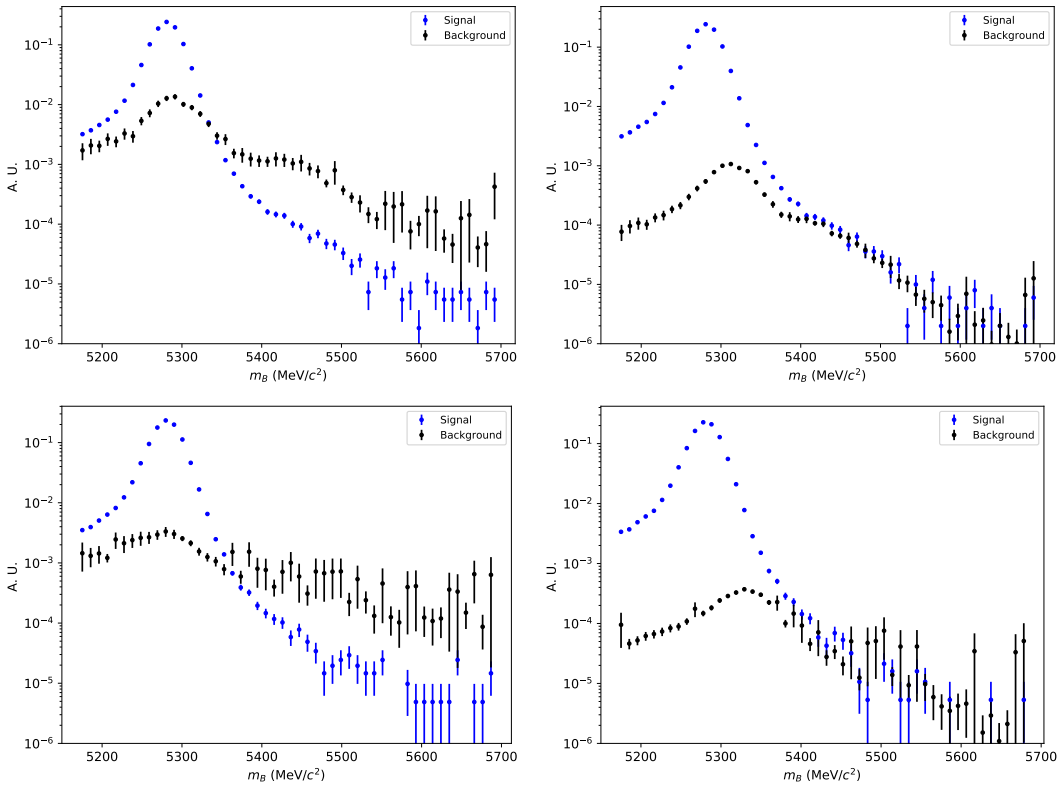


Figure 5.36: B^0 mass distributions for the signal and total peaking background simulation. The background simulation corresponds to the total background, obtained by adding together the individual background components scaled to the expected yields relative to the signal. The top plots show the low q^2 region and the bottom plots show the interresonance region. The left plots are after applying the trigger, stripping, and cleaning cuts (including hadron PID) and the right plots are after the full selection is applied. Note a log scale is used for the y-axis.

level are conditional with respect to the previous selection level. The detector acceptance efficiency refers to the proportion of $B^0 \to K^{*0}\mu^+\mu^-$ decays where the reconstructed final state particle tracks (K, π, μ^+, μ^-) are within the LHCb detector acceptance. The combined efficiency of the trigger, stripping, and cleaning cuts is 8.17%, of which the stripping is the most stringent. The q^2 selection refers to the selection of events which are in the q^2 windows $1.25 < q^2 < 8$ GeV/c^4 and $11 < q^2 < 12.5$ GeV/c^4. The mass cuts refer to the kaon-pion invariant mass ($m_{K\pi}$) and B mass (m_B) selections. The peaking background veto BDTs and $K\mu\mu$ veto are described in Sections 5.4 and 5.5 respectively. The combinatorial and K^{*+} BDTs are described in Sections 5.6 and 5.7 respectively. The total efficiency, which is a combination of all of the efficiencies, is also shown.

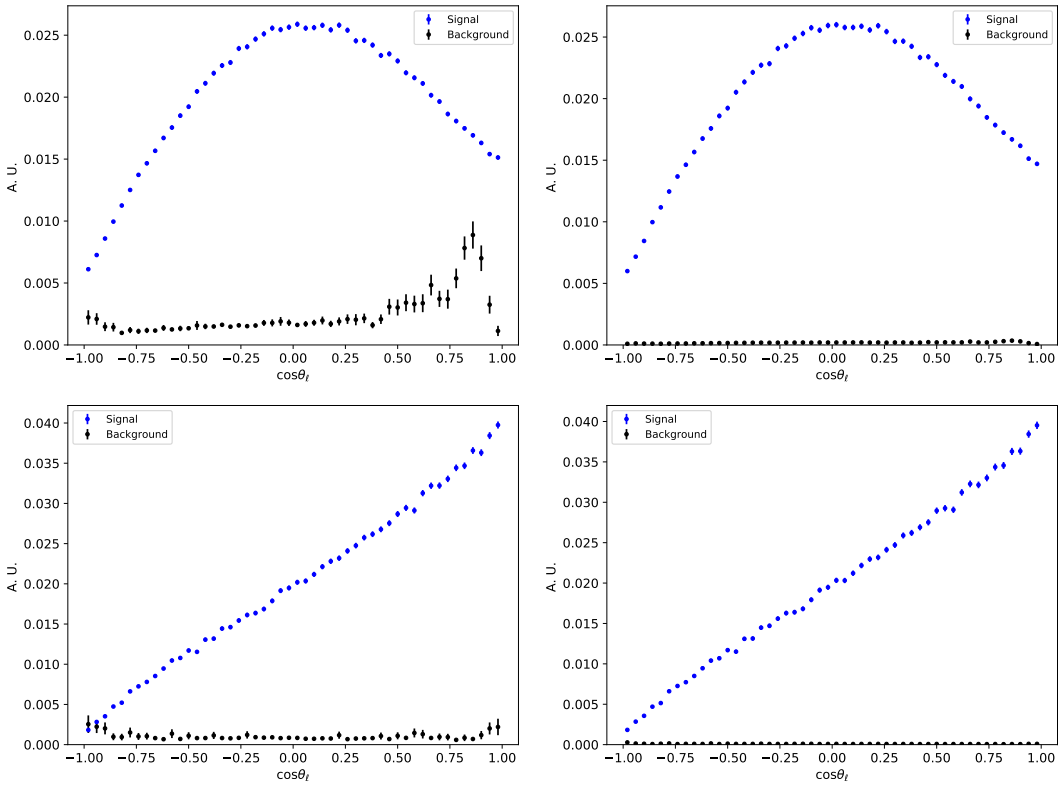


Figure 5.37: $\cos\theta_\ell$ distributions for the signal and total peaking background simulation. The background simulation corresponds to the total background, obtained by adding together the individual background components scaled to the expected yields relative to the signal. The top plots show the low q^2 region and the bottom plots show the interresonance region. The left plots are after applying the trigger, stripping, and cleaning cuts (including hadron PID) and the right plots are after the full selection is applied.

Table 5.10: Summary of applied selections and efficiencies. For each selection stage the efficiency shown is conditional with respect to the previous selection stage. The total efficiency is also shown.

Selection	**Efficiency**
Detector acceptance	16.79%
Trigger (Table 5.1), stripping (Table 5.2), cleaning cuts (Tables 5.3 and 5.4)	8.17%
q^2 ($1.25 < q^2 < 8$ GeV$^2/c^4$, $11 < q^2 < 12.5$ GeV$^2/c^4$)	37.16%
Mass selections ($796 < m_{K\pi} < 996$ MeV/c^2, $5170 < m_B < 5700$ MeV/c^2)	92.53%
Peaking background veto BDTs (Table 5.7)	95.79%
$K\mu\mu$ veto (Exclude events with $m_B > 5380$ MeV/c^2 and $5220 < m(K\mu^+\mu^-) < 5340$ MeV/c^2)	99.99%
Combinatorial BDT (Table 5.8)	91.85%
K^{*+} BDT ($K^{*+} \to K^+\pi^0$ BDT score > 0.65, $K^{*+} \to K^0_S\pi^+$ BDT score > 0.68)	92.72%
Total	**0.38%**

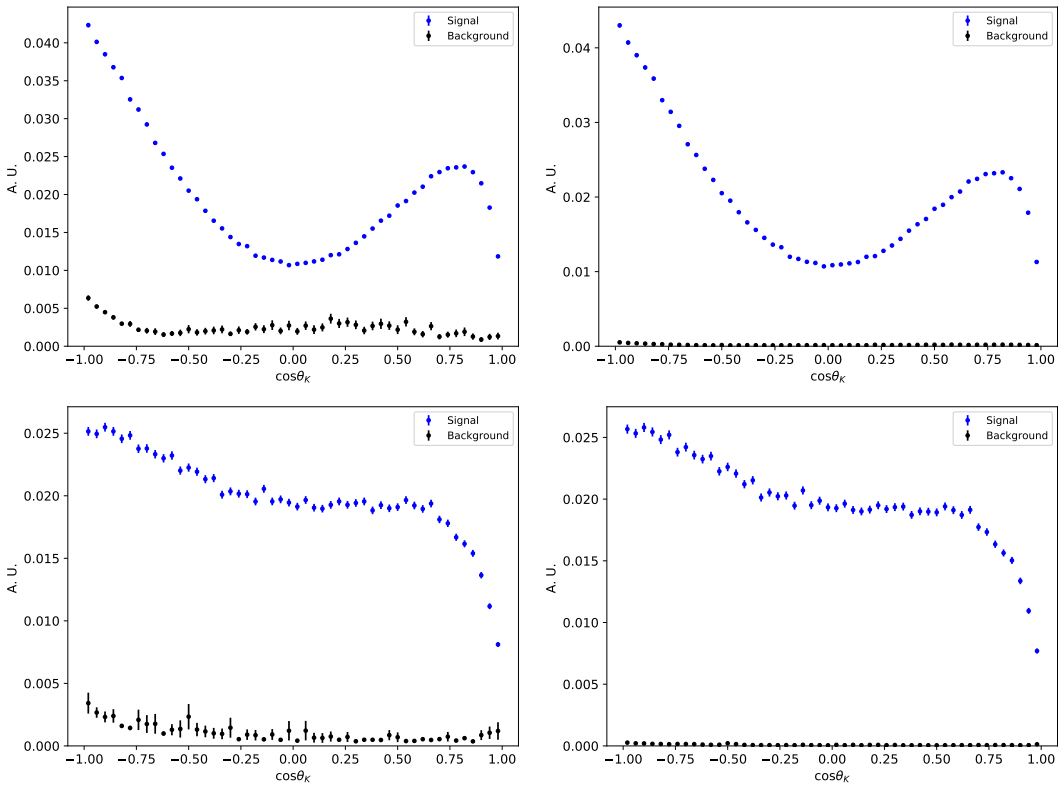


Figure 5.38: $\cos\theta_K$ distributions for the signal and peaking background simulation. The background simulation corresponds to the total background, obtained by adding together the individual background components scaled to the expected yields relative to the signal. The top plots show the low q^2 region and the bottom plots show the interresonance region. The left plots are after applying the trigger, stripping, and cleaning cuts (including hadron PID) and the right plots are after the full selection is applied.

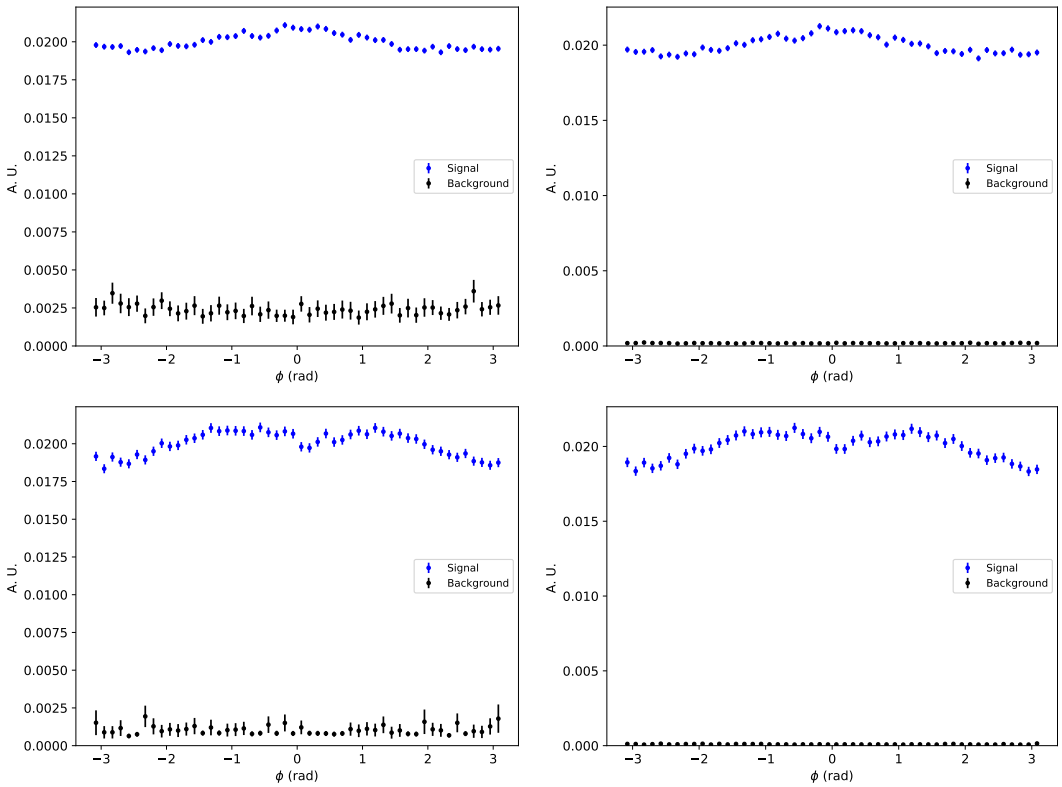


Figure 5.39: ϕ distributions for the signal and peaking background simulation. The background simulation corresponds to the total background, obtained by adding together the individual background components scaled to the expected yields relative to the signal. The top plots show the low q^2 region and the bottom plots show the interresonance region. The left plots are after applying the trigger, stripping, and cleaning cuts (including hadron PID) and the right plots are after the full selection is applied.

Chapter 6

Acceptance

Detector effects, selection cuts and reconstruction distort the true angular and q^2 distributions resulting in the measured distributions. To account for this effect, an acceptance function, ε, is used. A simulation is generated flat in the angles and q^2. This simulation is propagated through the detector and the selection is applied. This is used to ascertain the acceptance. The acceptance is parameterised in $\cos\theta_l$, $\cos\theta_K$, ϕ and q^2. In the amplitude ansatz fit, this enters as

$$\left.\frac{\mathrm{d}\Gamma}{\mathrm{d}\Omega \mathrm{d}q^2}\right|_{\text{experiment}} = \varepsilon(\cos\theta_l, \cos\theta_K, \phi, q^2) \left.\frac{\mathrm{d}\Gamma}{\mathrm{d}\Omega \mathrm{d}q^2}\right|_{\text{true}}, \tag{6.1}$$

where the acceptance function is used to transform from the measured angular distribution of the decay products to the underlying angular distribution which depends only on the underlying physics. Similarly to other $B^0 \to K^{*0}\mu^+\mu^-$analyses, the acceptance is modelled as the polynomial

$$\varepsilon(\cos\theta_l, \cos\theta_K, \phi, q^2) = \sum_{ijmn} c_{ijmn} L_i(\cos\theta_l) L_j(\cos\theta_K) L_m(\phi') L_n(q^{2\prime}), \tag{6.2}$$

where L_a denotes Legendre polynomials of order a. The quantities ϕ' and $q^{2\prime}$ correspond to ϕ and q^2 scaled to the range $-1 \leq x \leq 1$. The method of computing this function follows other $B^0 \to K^{*0}\mu^+\mu^-$analyses, where the coefficients c_{ijmn} are obtained by utilising the orthonormality of the Legendre polynomials, namely

$$\int_{-1}^{+1} L_a(x) L_b(x)\,\mathrm{d}x \;=\; \frac{2}{2a+1}\delta_{ab}\;, \tag{6.3}$$

where δ_{ab} is the Kronecker delta. Hence, explicitly, the coefficients c_{ijmn} are obtained via method of moments and are given by

$$\begin{aligned} c_{ijmn} = \frac{1}{\sum_{a=1}^{N} w_a} \sum_{a=1}^{N} w_a \Big[\Big(\frac{2i+1}{2}\Big)\Big(\frac{2j+1}{2}\Big)\Big(\frac{2m+1}{2}\Big)\Big(\frac{2n+1}{2}\Big) \\ \times L_i(\cos\theta_l) L_j(\cos\theta_K) L_m(\phi') L_n(q^{2\prime}) \Big]\;, \end{aligned} \tag{6.4}$$

where w_a is the weight corresponding to event a and N is the number of events in the simulated sample. The per-candidate weight consists of weights to account for mismodelling in the simulation, as described in Chapter 5.

Acceptance functions are ascertained using $B^0 \to K^{*0}\mu^+\mu^-$ simulation which is generated flat in q^2 and the angles. It is generated with a Briet-Wigner parameterisation of the $m_{K\pi}$ lineshape. The acceptance is computed integrated over the $m_{K\pi}$ region $0.796 < m_{K\pi} < 0.996$ GeV/c^2, in the q^2 region $1 < q^2 < 15$ GeV$^2/c^4$.

The acceptance for the B^0 dataset is different from the acceptance for the $\bar{B}^0$ dataset, thus the acceptance is also computed split into B^0 and $\bar{B}^0$ flavours and treated separately. This is in order to avoid the potential Punzi effect [85]. This effect describes a potential bias resulting from using the same template when fitting two classes of events. In particular, a bias is induced if the template depends on the event class. For this analysis, since the B^0 and $\bar{B}^0$ have potentially different acceptance shapes, using a single acceptance and applying it to both the B^0 and $\bar{B}^0$ Probability Density Functions (PDFs) can induce biases in the fit results, thus this must be avoided. Hence in the fit where the B^0 and $\bar{B}^0$ have separate PDFs, each PDF has its own acceptance function. Thus there are two acceptance functions in this case, one corresponding to the B^0 dataset, and another corresponding to the $\bar{B}^0$ dataset.

6.1 Goodness-of-fit

Beyond the accuracy of the simulation there are two sources of uncertainty related to the acceptance function. The first is the limited size of the simulation samples. The second is the choice of the orders of the polynomials used to fit the simulation. If the orders are too low the shape will not be accurately captured however if they are too high then statistical variations are modelled. Thus the maximum required orders is correlated to the size of the simulation sample available so one must determine the orders of the acceptance function in order to obtain a function which does not fluctuate with statistical variations and is still a good description of the simulation.

One way to ascertain whether the acceptance function can model the simulation is by visually inspecting 1D and 2D projections and comparing to simulation, in slices of q^2. However this is an insufficient way of determining the goodness-of-fit of the acceptance since there could be mis-modellings in the multidimensional space which are not visible in the 1D or 2D projections. In addition, since the acceptance is in four dimensions, χ^2 test-statistics are not sufficient to determine if a good fit of the acceptance simulation has been achieved since variations are averaged out when binning and chi-square statistics becomes increasingly complex as the dimensionality increases. Thus BDTs, which are designed for multivariate analyses, are utilised.

This multivariate analysis-based goodness-of-fit test is performed by training a BDT to discriminate between the simulation and a psuedoexperiment generated from the model used to fit the simulation (in this case, the acceptance function). This is compared to BDTs trained on statistically compatible distributions, which correspond to pseudoexperiments generated from the model. If a BDT cannot distinguish between the simulation and a pseudoexperiment generated from the model better than a pseudoexperimet generated from the model from other pseudoexperiments then the model is a good description of the simulation.

From an acceptance function, a pseudoexperiment with a large number of events (greater than 5x the simulation size) is generated in $\cos\theta_\ell$, $\cos\theta_K$, ϕ, and q^2, which are the variables used to parameterise the acceptance. This pseudoexperiment is known as the 'benchmark pseudoexperiment' since all BDTs in this goodness-of-fit method are

trained to distinguish this benchmark pseudoexperiment from another sample (either the simulation or other pseudoexperiments). More pseudoexperiments are generated but with the same number of events as the simulation used to derive the acceptance. The reason for generating the benchmark pseudoexperiment with high yields is to avoid double counting the expected statistical variation of the acceptance. BDTs are then trained to discriminate these pseudoexperiments from the benchmark pseudoexperiment. The training variables are $\cos\theta_\ell$, $\cos\theta_K$, ϕ, and q^2.

A metric for the performance of the BDT is the area under the Receiver Operating Characteristic curve (ROC AUC) and is a quantity which describes how well the BDT can discriminate between two samples. The ROC curve is the true positive rate against the false positive rate so if the ROC AUC is 0.5 the BDT is no better than a random guess. If the ROC AUC is 1 then the BDT is a perfect algorithm. The distribution of the BDTs trained to separate the benchmark pseudoexperiment from other psuedoexperiments is plotted. Since these pseudoexperiments are generated from the same model, the only differences between them are due to statistics.

Finally another BDT is trained to discriminate betwee the benchmark pseudoexperiment and the simulation used to derive the acceptance. The ROC AUC is obtained, compared with the distribution for the BDTs comparing the benchmark pseudoexperiment with pseudoexperiments and a p-value is calculated. If the fit is good, the BDT should not be able to differentiate between the benchmark pseudoexperiment and simulation better than it can differentiate statistical variations (distribution when comparing benchmark pseudoexperiment with pseudoexperiments). If the fit is bad, the ROC AUC of the BDT trained to discriminate between the benchmark psuedoexperiment and simulation will be larger than the distribution of BDTs trained to discriminate between the benchmark pseudoexperiment and pseudoexperiments.

Figure 6.1 shows distributions of the goodness-of-fit metric, for two different sets of polynomial orders. Here, the 2016 simulation sample is used. On the left is the result for one set of polynomials and the right, alternative polynomials where the maximum orders are higher. The blue distribution is that of the ROC AUC of the BDTs used to separate the benchmark pseudoexperiment from the other pseudoexperiments. This distribution is a distribution of figure-of-merits resulting from goodness-of-fits to datasets where the only

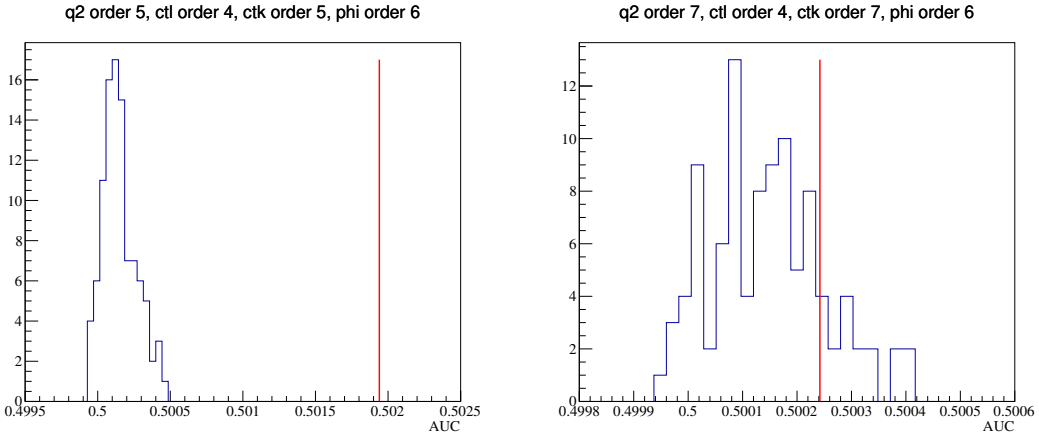


Figure 6.1: Example goodness-of-fit plots. The 2016 simulation sample is used. On the left is the result for one set of polynomials and the right, alternative polynomials where the maximum orders are higher. The blue distribution is that of the ROC AUC of the BDTs to separate the benchmark pseudoexperiment from the other pseudoexperiments; the red line is that for the BDT to separate benchmark pseudoexperiment from the simulation.

differences are due to statistical fluctuations. The red line shows the figure-of-merit (i.e. the ROC AUC) corresponding to the BDT which is trained to separate the benchmark pseudoexperiment from the simulation. If the figure-of-merit corresponding to this BDT is much larger than the figure-of-merits arising from statistical fluctuations, for example the left plot, the BDT is much better at separating a pseudoexperiment from the simulation than it is from separating the pseudoexperiment from other pseudoexperiments. Thus the acceptance function in this case is a bad fit of the simulation. However, if the figure-of-merit from the BDT trained to separate a pseudoexperiment from the simulation is comparable to the distribution of figure-of-merits from BDTs trained to separate the pseudoexperiment from other pseudoexperiments, for example the right plot, then the acceptance here is a good description of the simulation.

6.2 Choice of acceptances

Having found high enough orders to describe the simulation, care must be taken to avoid overfitting. Therefore the polynomial orders are decreased until the fit is no longer acceptable. The lowest set of orders that gives a sufficient quality fit is used in the analysis. If the p-value is greater than 5% the choice of orders is considered to result in an acceptance function which is a good description of the simulation. Where B^0 and $\bar{B}^0$ are treated

Year	Number of acceptance events	J/ψ data signal yield	Weight
Run 1	147244	152390	1.035
2016	135687	158710	1.170
2017	130394	170034	1.304
2018	125025	209108	1.673

Table 6.1: J/ψ signal data yields, number of acceptance simulation events which pass the selection in the J/ψ window, and weight used for each sample when computing the B^0 acceptance.

Year	Number of acceptance events	J/ψ data signal yield	Weight
Run 1	148141	148782	1.004
2016	134143	154544	1.152
2017	129402	166172	1.284
2018	124967	203765	1.631

Table 6.2: J/ψ signal data yields, number of acceptance simulation events which pass the selection in the J/ψ window, and weight used for each sample when computing the $\bar{B}^0$ acceptance.

separately, the maximum orders in each of the four dimensions are chosen to be the same. The lowest set of orders that gives a good description for both B^0 and $\bar{B}^0$ is chosen.

Since the fit combines the Run 1 and Run 2 datasets, the combined Run 1 + Run 2 acceptances are computed. In order to do this, for each data period (Run 1, 2016, 2017, 2018) the simulation is normalised such that it represents the relative fraction of that year's data in the total dataset. This is by applying a per-year weight which is the J/ψ data fit yield divided by the amount of simulation in the J/ψ window. This is shown in Table 6.1 for the B^0 dataset and Table 6.2 for the $\bar{B}^0$ dataset. For the acceptances used for the CP-averaged fit, again the simulation used to compute the acceptance is normalised to the amount of simulation in the J/ψ window and scaled according to the J/ψ data fits. This is shown in Table 6.3.

Year	Number of acceptance events	J/ψ data signal yield	Weight
Run 1	295385	303920	1.029
2016	269830	313483	1.162
2017	259796	336601	1.296
2018	249992	414027	1.656

Table 6.3: J/ψ signal data yields, number of acceptance simulation events which pass the selection in the J/ψ window, and weight used for each sample when computing the combined B^0 and $\bar{B}^0$ acceptance.

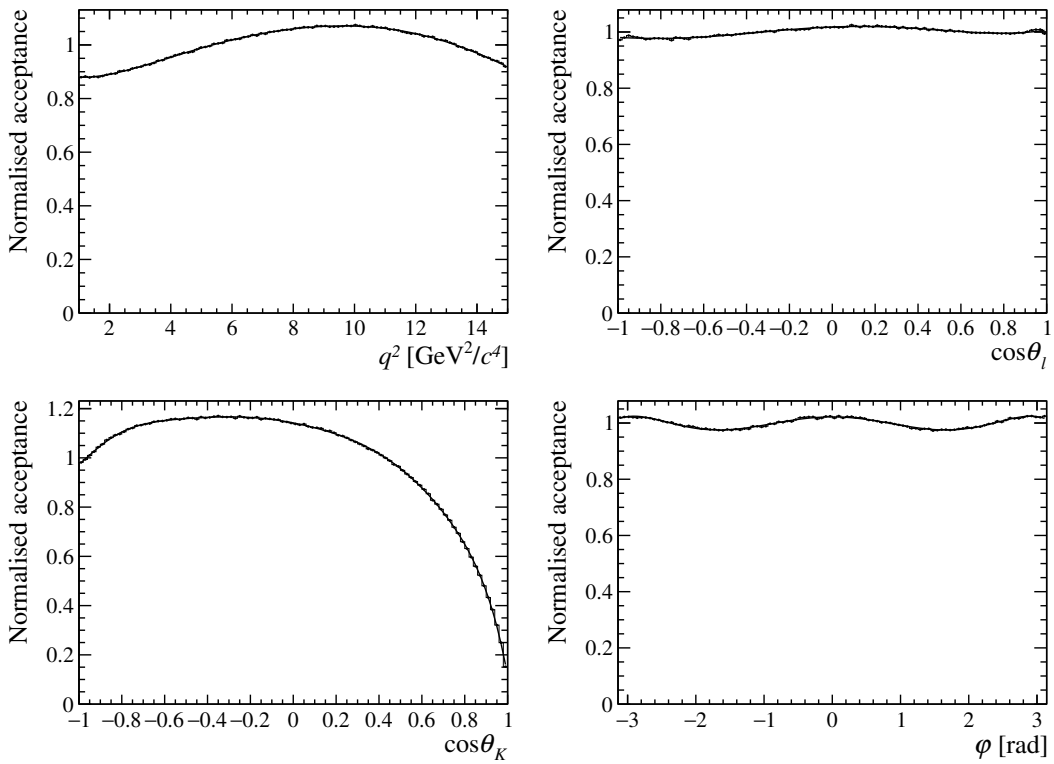


Figure 6.2: Projections of the combined B^0 and $\bar{B}^0$ acceptance in q^2, $\cos\theta_\ell$, $\cos\theta_K$, and ϕ.

For theB^0 and $\bar{B}^0$ acceptances, the optimum maximum orders are 6 ($\cos\theta_\ell$), 8 ($\cos\theta_K$), 6 (ϕ), and 5 (q^2). For the combined B^0 and $\bar{B}^0$ acceptance, the optimum maximum orders are 6 ($\cos\theta_\ell$), 9 ($\cos\theta_K$), 6 (ϕ), and 6 (q^2).

Figure 6.2 shows the projections of the combined B^0 and $\bar{B}^0$ acceptance in q^2, $\cos\theta_\ell$, $\cos\theta_K$, and ϕ.

Chapter 7

Fitting Strategy

A fitting framework is implemented in C++, using the TMinuit [86] library in ROOT. In this fitting framework, the PDFs are defined, the data are loaded and the fit is performed. This chapter describes the PDFs and the fits that are performed.

The fitting framework is written such that an extended maximum likelihood fit is performed to the data. A total PDF is formed which contains the signal (B^0), a B_s^0 component, and the combinatorial background. The angular part of the signal PDF is formed from the amplitudes, which are parameterised with q^2-dependent ansatzes.

7.1 Signal PDF

7.1.1 Angular signal PDF

The angular signal PDF is described in Chapter 2.7. The J-terms are written as bilinear combinations of the decay amplitudes. The PDF $\frac{\mathrm{d}^4\Gamma(B^0\to K^{*0}\mu^+\mu^-)}{\mathrm{d}\cos\theta_\ell \mathrm{d}\cos\theta_K \mathrm{d}\phi \mathrm{d}q^2}$ is fitted for the signal 4D fit. 5D fits, where the PDF is a function of $m_{K\pi}$, were also studied, where the PDF is given by $\frac{\mathrm{d}^5\Gamma(B^0\to K^{*0}\mu^+\mu^-)}{\mathrm{d}\cos\theta_\ell \mathrm{d}\cos\theta_K \mathrm{d}\phi \mathrm{d}q^2 \mathrm{d}m_{K\pi}}$. As shown in Section 8.1, the performance of the 5D fit is insufficient with the existing statistics, and so a 4D fit nominally is performed for this analysis. For the CP-average fit, where the $B^0 \to K^{*0}\mu^+\mu^-$and $\bar{B}^0 \to \bar{K}^{*0}\mu^+\mu^-$ decay amplitudes are set to be identical, there is just one PDF. For the asymmetries fit there are two PDFs, one for the B^0 amplitudes, and one for the $\bar{B}^0$ amplitudes.

Ansatz model

The ansatz model used for this analysis is

$$\mathcal{A} = \sum_i \alpha_i L_i \tag{7.1}$$

for the amplitudes. L_i are Legendre polynomials in q^2 and a_i are complex coefficients. When evaluating the Legendre polynomials, q^2 is scaled to the range $-1 \leq x \leq 1$. An alternative amplitude ansatz was also considered. This is described by the parameterisation

$$\mathcal{A} = \alpha + \beta q^2 + \frac{\gamma}{q^2} \tag{7.2}$$

as originally proposed in [11]. This ansatz was not adopted due to the very large correlations between coefficients and presence of local minima seen through pseudoexperiment studies. More about the motivation for the Legendre ansatz is given in Section 8.2.

Amplitude basis

As described in Section 2.10, there are four continuous symmetries of the PDF. Thus the effective number of amplitudes (8 complex amplitudes $\times$ (real, imaginary) = 16) is not equal to the number of degrees-of-freedom in the PDF (12). This results in an infinite number of sets of amplitudes which result in the same PDF. In order to define an amplitude basis an effective rotation is performed that leaves four of the amplitudes at arbitrary constants, which do not need to float in the fit. As initially motivated by [11], the amplitude basis chosen is where $\mathrm{Im}(\mathcal{A}_\perp^R) = \mathrm{Im}(\mathcal{A}_0^L) = \mathrm{Re}(\mathcal{A}_0^R) = \mathrm{Im}(\mathcal{A}_0^R) = 0$. All other amplitudes float in the fit.

Amplitude ansatz description of the theoretical predictions

The Legendre polynomial ansatz with four parameters can describe a variety of models as shown in Figures 7.1, 7.2, and 7.3. These show the P-wave amplitudes in the nominal amplitude basis. The predictions shown in black are for two SM models (Figures 7.1 and 7.2) where the differences between them are the phases between the $\phi(1020)$ and the J/ψ and the rare mode, and a NP model (Figure 7.3), where there is a shift $\Delta C_9 = -1$ with

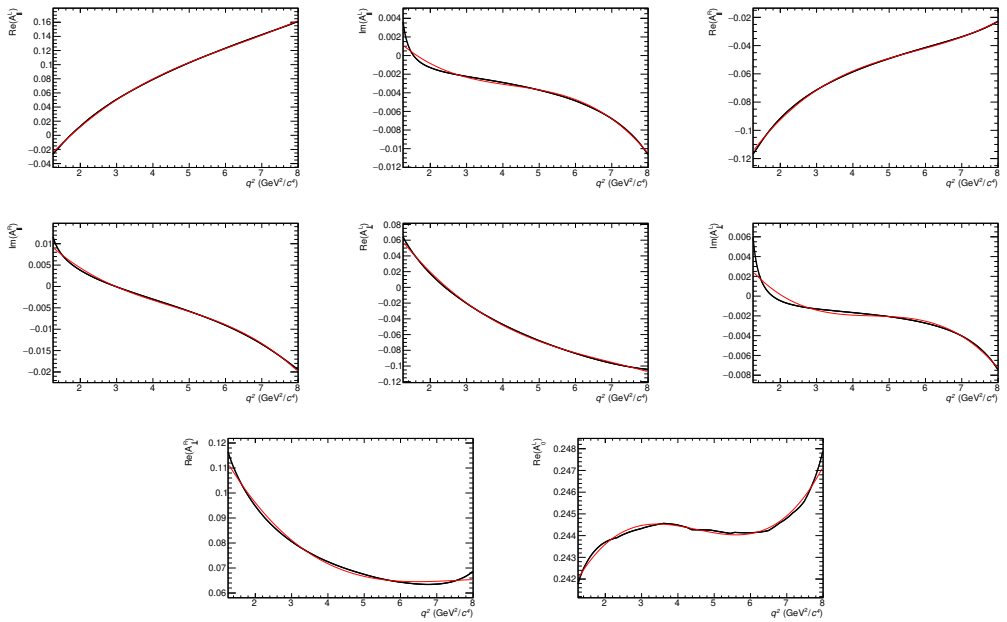


Figure 7.1: Theoretical predictions of the P-wave amplitudes in the transformed basis (black) and fits to these prediction using four-parameter Legendre polynomial amplitudes ansatzes (red). The theoretical predictions shown are for the SM, where the phases between the rare mode and the $\phi(1020)$ and J/ψ resonances are both $\frac{\pi}{2}$.

respect to the SM. These predictions were generated by another member of the analysis. Shown in red are fits to these predictions with four-parameter Legendre polynomial ansatzes. As shown in these plots, the predictions vary between the models and the ansatz chosen can describe a variety of models. This is also verified by performing a goodness-of-fit to the data, as shown in Section 11.3.1.

7.1.2 B^0 mass lineshape

The B^0 mass m_B is factorised out from the angular PDF. It is described by a sum of two Crystal Ball functions [87], with tails on opposite sides. The B^0 mass signal parameterisation is

$$P_B = f_{\text{core}} P_{CB}(m_B, \mu, \sigma_1, \alpha_1, n_1) + (1 - f_{\text{core}}) P_{CB}(m_B, \mu, \sigma_2, \alpha_2, n_2) \tag{7.3}$$

where

$$P_{CB}(m_B, \mu, \sigma_i, \alpha_i, n_i) = N e^{-\frac{(m_B - \mu)^2}{2\sigma_i^2}} \tag{7.4}$$

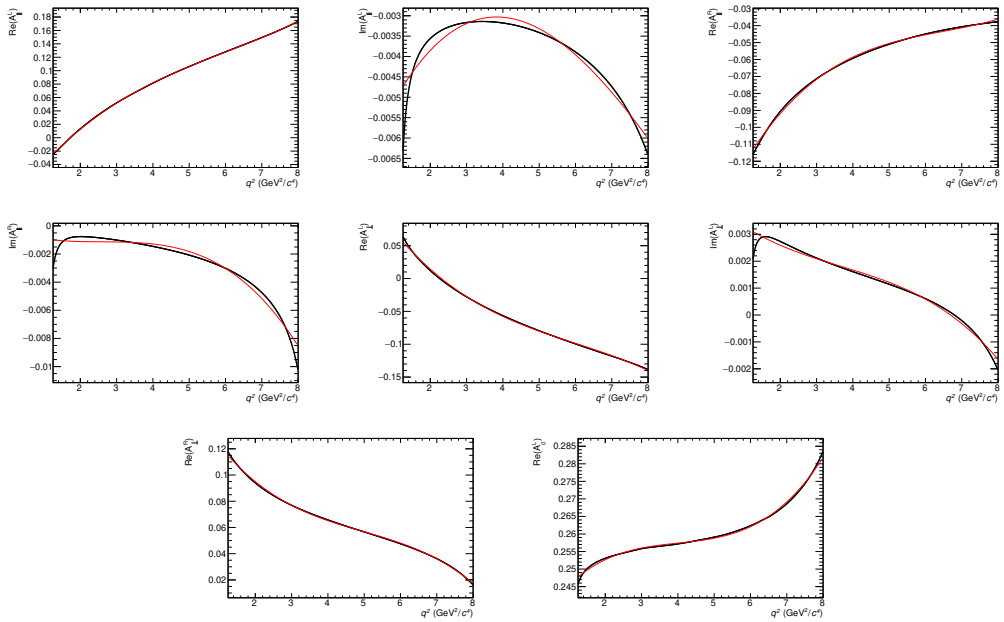


Figure 7.2: Theoretical predictions of the P-wave amplitudes in the transformed basis (black) and fits to these prediction using four-parameter Legendre polynomial amplitudes ansatzes (red). The theoretical predictions shown are for the SM, where the phases between the rare mode and the $\phi(1020)$ and J/ψ resonances are both 0.

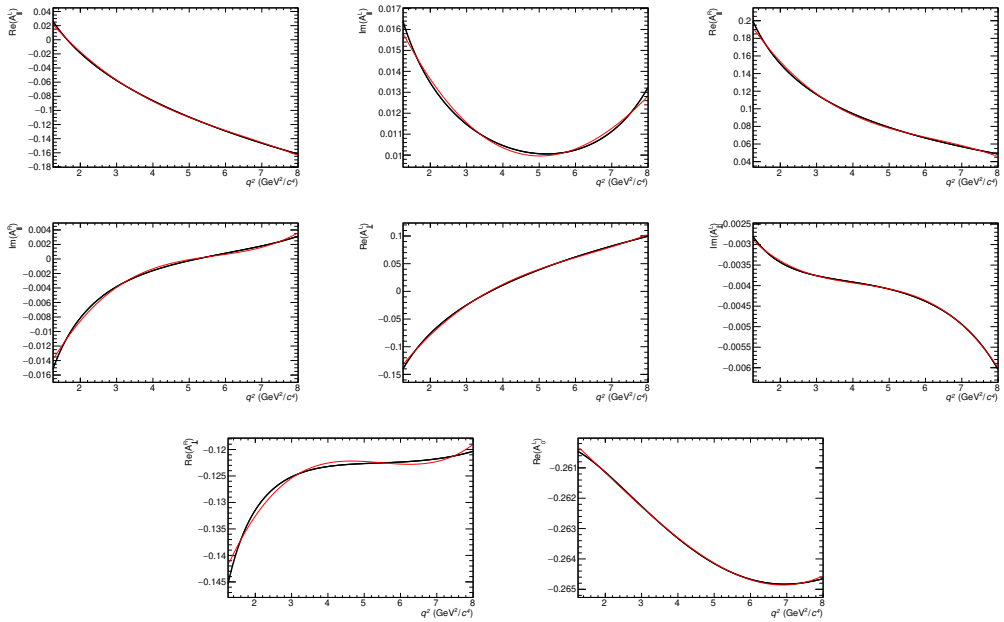


Figure 7.3: Theoretical predictions of the P-wave amplitudes in the transformed basis (black) and fits to these prediction using four-parameter Legendre polynomial amplitudes ansatzes (red). The theoretical predictions shown are for a NP model with $\Delta C_9 = -1$.

for $\frac{m_B-\mu}{\sigma_i} > -\alpha_i$ and

$$P_{CB}(m_B, \mu, \sigma_i, \alpha_i, n_i) = N \left(\frac{n_i}{|a_i|}\right)^{n_i} e^{-\frac{\alpha_i^2}{2}} \left(\frac{n_i}{|a_i|} - |a_i| - \frac{m_B - \mu}{\sigma_i}\right)^{-n_i} \tag{7.5}$$

where $\frac{m_B-\mu}{\sigma_i} \leq -\alpha_i$, and N is a normalisation factor. The parameters α_i, n_i and f_{core} are extracted from the control mode and fixed in the rare mode fits.

7.1.3 Contribution from $B_s^0 \to \bar{K}^{*0}\mu^+\mu^-$

The contribution from $\bar{B}_s^0 \to K^{*0}\mu^+\mu^-$ needs to be accounted for in order to have a good-quality B^0 fit. Therefore $N_{Sig}P_{Sig}$ for the CP-average fit, where N_{Sig,B^0} is the number of B^0 events and P_{Sig} is the signal PDF, is modified to be

$$N_{Sig,B^0}P_{Sig} \to N_{Sig,B^0}P_{Sig,B^0} + N_{Sig,B_s^0}P_{Sig,B_s^0} = N_{Sig,B^0}(P_{Sig,B^0} + f_{B_s^0}P_{Sig,B_s^0}) \tag{7.6}$$

where $f_{B_s^0} = \frac{N_{Sig,B_s^0}}{N_{Sig,B^0}} = \frac{\mathcal{B}(B_s^0 \to \bar{K}^{*0}\mu^+\mu^-)}{\mathcal{B}(B^0 \to K^{*0}\mu^+\mu^-)} \frac{f_s}{f_d}$, with f_s and f_d as the B_s^0 and B^0 fragmentation fractions.

The B^0 and $\bar{B}^0$ datasets are split based on the charge of the kaon, since $B^0 \to K^{*0}\mu^+\mu^-$ and $\bar{B}^0 \to \bar{K}^{*0}\mu^+\mu^-$, where $K^{*0} \to K^+\pi^-$ and $\bar{K}^{*0} \to K^-\pi^+$. Since $\bar{B}_s^0 \to K^{*0}\mu^+\mu^-$ and $B_s^0 \to \bar{K}^{*0}\mu^+\mu^-$, the K^+ dataset contains the B^0 and $\bar{B}_s^0$ and the K^- dataset contains the $\bar{B}^0$ and B_s^0. Thus in the asymmetries fit the equivalent of Equation 7.6 for the K^+ dataset is

$$N_{Sig,B^0}P_{Sig,B^0} + N_{Sig,\bar{B}_s^0}P_{Sig,\bar{B}_s^0} = N_{Sig,B^0}(P_{Sig,B^0} + f_{\bar{B}_s^0}P_{Sig,\bar{B}_s^0}) \tag{7.7}$$

and for the K^- dataset the equivalent is

$$N_{Sig,\bar{B}^0}P_{Sig,\bar{B}^0} + N_{Sig,B_s^0}P_{Sig,B_s^0} = N_{Sig,\bar{B}^0}(P_{Sig,\bar{B}^0} + f_{B_s^0}P_{Sig,B_s^0}) \tag{7.8}$$

where $f_{\bar{B}_s^0} = \frac{N_{Sig,\bar{B}_s^0}}{N_{Sig,B^0}} = \frac{\mathcal{B}(\bar{B}_s^0 \to K^{*0}\mu^+\mu^-)}{\mathcal{B}(B^0 \to K^{*0}\mu^+\mu^-)} \frac{f_s}{f_d}$ and $f_{B_s^0} = \frac{N_{Sig,B_s^0}}{N_{Sig,\bar{B}^0}} = \frac{\mathcal{B}(B_s^0 \to \bar{K}^{*0}\mu^+\mu^-)}{\mathcal{B}(\bar{B}^0 \to \bar{K}^{*0}\mu^+\mu^-)} \frac{f_s}{f_d}$. The ratio of the B_s^0 to B^0 yields are fixed in both the control mode fit and the rare mode fits. The values of these ratios are fixed to the values recorded in the PDG [14], assuming no CP asymmetry. The hadronisation fraction is taken from a recent measurement at LHCb [88].

Therefore for all fits $f_{B_s^0} = f_{\bar{B}_s^0} = 0.0077577$.

The B_s^0 lineshape has identical parameters to that of the B^0, but with a shift $\Delta M = 87.26$ MeV/c^2 of the mean μ. There are also differences in the angular PDF. The LHCb angular convention is chosen such that the PDFs have the same form for B^0 and $\bar{B}^0$. However, as the B_s^0 decays to a K^- and not a K^+, a CP-transformation has to be applied to the final state. This corresponds to a transformation in ϕ, $\phi \to 2\pi - \phi$. This can therefore be resolved by flipping the signs of J_7, J_8, J_9, $\tilde{J}_7$ and $\tilde{J}_8$.

Thus the B_s^0 PDF has the same angular PDF as the B^0 but with the signs of J_7, J_8, J_9, $\tilde{J}_7$ and $\tilde{J}_8$ flipped, as described in Figure 4 of [49]. The amplitude coefficients are shared in the fit between P_{Sig,B^0} and P_{Sig,B_s^0} for the CP-averages fit. For the CP-asymmetries fit, the parameters are shared between P_{Sig,B^0} and P_{Sig,B_s^0}, and $P_{Sig,\bar{B}^0}$ and $P_{Sig,\bar{B}_s^0}$. This approximation is considered to be acceptable since the contributions from the B_s^0 processes are small compared to those from the main B^0 processes.

Acceptance for $B_s^0 \to \bar{K}^{*0}\mu^+\mu^-$

The acceptance (i.e. the q^2 and angular efficiencies) for the B_s^0 was studied using $B_s^0 \to \bar{K}^{*0}\mu^+\mu^-$ simulation. Since differences could potentially arise due to the different decay kinematics, i.e. that the B_s^0 mass is different from the B^0, there could potentially be differences between the $B_s^0 \to \bar{K}^{*0}\mu^+\mu^-$ acceptance and the $B^0 \to K^{*0}\mu^+\mu^-$ acceptance.

The $B_s^0 \to \bar{K}^{*0}\mu^+\mu^-$ acceptance was determined by simulating $B_s^0 \to \bar{K}^{*0}\mu^+\mu^-$ events and applying the full selection. Weights were applied in order to make the variables q^2, $\cos\theta_\ell$, $\cos\theta_K$, and ϕ flat if no selection is applied. Thus the weighted selected simulation corresponds to the acceptance.

The $B_s^0 \to \bar{K}^{*0}\mu^+\mu^-$ acceptance was compared to the $B^0 \to K^{*0}\mu^+\mu^-$acceptance. Figure 7.4 shows the q^2, $\cos\theta_\ell$, $\cos\theta_K$, and ϕ projections of acceptances for the B_s^0 (red) and B^0 (blue) decay channels, using 2017 simulation. As seen in these plots, the acceptances are the same within statistical fluctuations. Thus given the similarity of the acceptances for the B^0 and B_s^0 decays, and that the expected number of B_s^0 is a small fraction of the expected number of B^0 events, the acceptances for the B^0 and B_s^0 are treated to be the same, namely the acceptance of the B^0 decay discussed in Chapter 6.

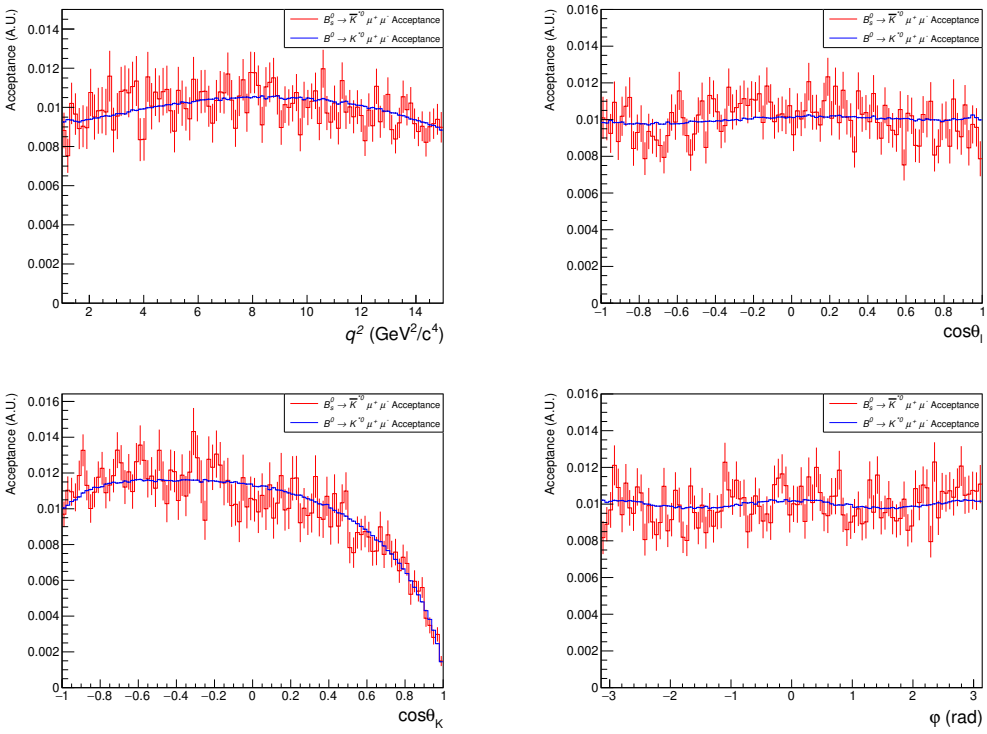


Figure 7.4: Comparison between the B^0 acceptance (blue) and the B_s^0 acceptance (red) in q^2, $\cos\theta_\ell$, $\cos\theta_K$, and ϕ.

7.2 Background PDF

The background is described by a five dimensional PDF

$$P_{Bkg}\left(q^2, \cos\theta_\ell, \cos\theta_K, \phi, m_{B^0}\right) = P\left(q^2\right) P\left(\cos\theta_\ell\right) P\left(\cos\theta_K\right) P(\phi) P\left(m_{B^0}\right) \tag{7.9}$$

where the functions $P\left(q^2\right)$, $P\left(\cos\theta_\ell\right)$, $P\left(\cos\theta_K\right)$, and $P(\phi)$ are second order Chebyshev polynomials with two parameters describing each function. When evaluating the polynomials, ϕ and q^2 are both scaled to the range $-1 \leq x \leq 1$. The background in the B^0 mass $P(m_{B^0})$ is an exponential function. There are therefore nine free background PDF parameters in total. The number of background events is also floated. For the CP-averages fit, this is the combined number of B^0 and $\bar{B}^0$ background events. For the CP-asymmetries fit, the number of background events for the B^0 and $\bar{B}^0$ datasets float separately. For the CP-asymmetries fit the background PDFs are identical between B^0 and $\bar{B}^0$ apart from the overall yields. This was shown to be a reasonable approximation in studies performed as part of the binned analysis.

7.2.1 $B^+ \to K^+\mu^+\mu^-$ veto

There is a selection veto for removing $B^+ \to K^+\mu^+\mu^-$ events. All events with $5220 < m_{K\mu\mu} < 5340$ MeV/c^2 and $m_{B^0} > 5380$ MeV/c^2 are removed which removes the majority of $B^+ \to K^+\mu^+\mu^-$ events, where $m_{K\mu\mu}$ is the invariant mass of the $K\mu\mu$ system. However this distorts the four-dimensional space of $\cos\theta_K$, $m_{B^0}, m_{K\pi}$ and q^2, above the signal B^0 mass. Since the nominal fit is performed after integrating over $m_{K\pi}$, only effects in $\cos\theta_K$, m_{B^0} and q^2 are taken into account. Following the procedure described in [82], a histogram describing this cut out region is created by mapping out simulated events which were cut out by the veto in regions of $\cos\theta_K$, m_{B^0}, $m_{K\pi}$ and q^2, and integrating over $m_{K\pi}$. This histogram was generated by another member of the amplitude ansatz analysis. Events that fall into this region are set to 0 and the background PDF is renormalised correspondingly.

The correction of the $K^+\mu^+\mu^-$ veto is only applied to the background PDF since this veto takes effect far from the B^0 signal peak in the B^0 mass distribution (fewer than 0.01% of events are removed). It is closer to the B_s^0 signal peak but the effect is expected to be small since the B_s^0 signal peak is already small.

7.3 Unbinned extended maximum likelihood fit

An unbinned extended maximum likelihood fit is performed, where the likelihood

$$\begin{aligned}&\frac{(N_{Sig,B^0}+N_{Sig,B_s^0}+N_{Bkg})^{N_{Dat}}e^{-(N_{Sig,B^0}+N_{Sig,B_s^0}+N_{Bkg})}}{N_{Dat}!}\times\\&\prod_{i=0}^{N_{Dat}}\left(\frac{1}{(N_{Sig,B^0}+N_{Sig,B_s^0}+N_{Bkg})}\Big(N_{Sig,B^0}P_{Sig,B^0}(x_i,\eta)+\right.\\&\left.N_{Sig,B_s^0}(x_i,\eta)P_{Sig,B_s^0}(x_i,\eta)+N_{Bkg}P_{Bkg}(x_i,\eta)\Big)\right)\end{aligned} \tag{7.10}$$

is maximised. Here N_{Sig,B^0}, $N_{Sig,B_s^0} \equiv f_{B_s^0}N_{Sig,B^0}$ and N_{Bkg} are the means of Poissons which correspond to the expected number of B^0, B_s^0 and background events. N_{Dat} is the total number of events in the dataset, x_i corresponds to the angles and mass variables at event i, and η are the fit parameters.

The signal PDF is not sensitive to the global scaling. If all of the amplitudes are

multiplied by a constant, it would cancel when normalising the PDF. Thus the overall scale of the PDF needs to be fixed. Since the observables are computed normalised to the rate, such as in Equation 2.29, the overall scale of the amplitudes is meaningless. In order to set the overall scaling, an extended maximum likelihood fit is performed, where

$$N_{\text{Sig },B^0} = \alpha \int P_{\text{Sig },B^0}\Big|_{\text{Fit}} \; \mathrm{d}\cos\theta_\ell \; \mathrm{d}\cos\theta_K \; \mathrm{d}\phi \; \mathrm{d}q^2 \; \mathrm{d}m_B \tag{7.11}$$

where α is a constant. This is nominally set as 50,000 for the rare mode fits and 2,000,000 for the J/ψ fit. These are arbitrary values — the angular PDF consists of bilinear combinations of the decay amplitudes, thus if all of the amplitudes are scaled by a constant, this constant can be factored out of the PDF. This constant cancels when normalising the PDF. The overall scale of the amplitudes is thus chosen by choosing an arbitrary value of α, such that $N_{\text{Sig, }B^0}$ always corresponds to the observed number of signal events. In other words, the amplitudes in P_{Sig,B^0} and P_{Sig,B^0_s} will scale such that N_{Sig,B^0} corresponds to the observed number of signal events, and this arbitrary scale is chosen by setting α to an abritrary constant.

For the CP-asymmetries fit, there are two scalings, α_{B^0} and $\alpha_{\bar{B}^0}$, where

$$N_{\text{Sig },B^0} = \alpha_{B^0} \int P_{\text{Sig },B^0}\Big|_{\text{Fit}} \; \mathrm{d}\cos\theta_\ell \; \mathrm{d}\cos\theta_K \; \mathrm{d}\phi \; \mathrm{d}q^2 \; \mathrm{d}m_B \tag{7.12}$$

and

$$N_{\text{Sig},\bar{B}^0} = \alpha_{\bar{B}^0} \int P_{\text{Sig },\bar{B}^0}\Big|_{\text{Fit}} \; \mathrm{d}\cos\theta_\ell \; \mathrm{d}\cos\theta_K \; \mathrm{d}\phi \; \mathrm{d}q^2 \; \mathrm{d}m_B. \tag{7.13}$$

In order to allow for the B^0 amplitudes to be directly comparable to the $\bar{B}^0$ amplitudes, the two constants α_{B^0} and $\alpha_{\bar{B}^0}$ can be set to be the same. However, this does not factor out detection and production asymmetries between B^0 and $\bar{B}^0$. In order to incorporate this effect, the detection and production asymmetries are measured from the $B^0 \to K^{*0}J/\psi$ data fits by forming the ratio $C = \frac{N_{B^0}}{N_{\bar{B}^0}}$ from the signal yields. This is the ratio between the number of $\bar{B}^0$ events to the B^0 events in the J/ψ mode. Since the genuine CP-asymmetry A_{CP} of the J/ψ is assumed to be 0 [89], this ratio corresponds to the amount of production and detection asymmetry. Thus by setting $\alpha_{\bar{B}^0} = C\alpha_{B^0}$, the production and detection asymmetries are factored out.

The negative log-likelihood function, which is minimised, is written as

$$-\log\mathcal{L} = -\sum_{N_{\text{Dat}}} \log\left(N_{Sig,B^0}P_{Sig,B^0} + N_{Sig,B_s^0}P_{Sig,B_s^0} + N_{Bkg}P_{Bkg}\right) + \left(N_{Sig,B^0} + N_{Sig,B_s^0} + N_{Bkg}\right) \tag{7.14}$$

for the CP-averages fit and

$$\begin{aligned}-\log\mathcal{L} = &-\sum_{N_{\text{Dat,K}^+}} \log\left(N_{Sig,B^0}P_{Sig,B^0} + N_{Sig,\bar{B}_s^0}P_{Sig,\bar{B}_s^0} + N_{Bkg,B^0}P_{Bkg,B^0}\right) - \\ &\sum_{N_{\text{Dat,K}^-}} \log\left(N_{Sig,\bar{B}^0}P_{Sig,\bar{B}^0} + N_{Sig,B_s^0}P_{Sig,B_s^0} + N_{Bkg,\bar{B}^0}P_{Bkg,\bar{B}^0}\right) + \\ &\left(N_{Sig,B^0} + N_{Sig,B_s^0} + N_{Bkg,B^0} + N_{Sig,\bar{B}^0} + N_{Sig,\bar{B}_s^0} + N_{Bkg,\bar{B}^0}\right)\end{aligned} \tag{7.15}$$

for the CP-asymmetries fit. Here, N_{Sig} denotes the signal yield as shown in Equations 7.11, 7.12 and 7.13, and N_{Bkg} is the background yield.

7.4 Fit strategy

The fitter measures the amplitude coefficients, B^0 mass lineshape parameters, background parameters, and the background yields, returning the fit parameters. There are some P-wave amplitude parameters where the uncertainties from the Hessian error matrix cannot be used since they do not provide good descriptions of the 1D log-likelihood surfaces. Therefore fits to these 1D log-likelihood surfaces are performed with bifurcated parabolas in order to extract the uncertainties. The correlation matrix is computed from the Hessian covariance matrix.

The amplitudes and observables are then computed by sampling from the fit covaraince matrix. There are several fit configurations for the data fits as described below.

- Fit the control mode ($B^0 \to K^{*0}J/\psi$) samples. Here the number of parameters in the amplitude ansatz is set to be one, i.e. constants, to remove the q^2 dependencies. This fit is performed in a 60 MeV/c^2 window about the J/ψ mass $m_{J/\psi}$, i.e. $(m_{J/\psi} - 60\ \text{MeV}/c^2) < q^2 < (m_{J/\psi} + 60\ \text{MeV}/c^2)$. The acceptance function is evaluated at the J/ψ mass. The fit in the CP-averages or the CP-asymmetries configuration is performed. The results of this fit are in Chapter 10.

- Fit in the range $1.25 < q^2 < 8.0\ \text{GeV}^2/c^4$. The number of parameters for the

amplitude ansatz is set to be 4, and nominally the CP-averages fit is performed, with results in Chapter 11. The motivation for the number of parameters for the ansatz and why a CP-asymmetries fit is not nominally performed for this analysis is also given in Chapter 11.

- Fit in the inter-resonance region $11 < q^2 < 12.5$ GeV$^2/c^4$. Results in this region do not feature in this thesis since the pseudoexperiment studies were primarily performed by another member of the amplitude ansatz analysis.

In all cases, the S-wave is set to be flat in q^2. The combined Run 1 + Run 2 datasets are fit for each q^2 region, with the exception of the control mode where each data period is fitted separately.

Chapter 8

Pseudoexperiment studies

The pseudoexperiment studies described in this Chapter are based on Standard Model pseudoexperiements with no acceptance or background. These studies were performed in order to understand the analysis in more detail, such as the study of apparent biases and symmetries of the PDF, as well as answering questions pertaining to the analysis strategy, such as whether to measure the PDF as a function of the kaon-pion invariant mass ($m_{K\pi}$) or which amplitude ansatz to use.

8.1 Study of the 5D fit

Whilst this analysis is performed on the decay $B^0 \to K^{*0}\mu^+\mu^-$, in reality it is $B^0 \to K^+\pi^-\mu^+\mu^-$. In the signal data, the dominant contribution in the $K^+\pi^-$ system corresponds to the $K^{*0}(892)$, where the kaon and pion are in a P-wave configuration. However there is also a contribution where the $K^+\pi^-$ system is in an S-wave configuration, for example the $K_0^*(700)$, also known as the κ, and the $K_0^*(1430)$ [14]. There are more contributions however they are negligble in the $K\pi$ mass window $0.796 < m_{K\pi} < 0.996$ MeV/c^2, where $m_{K\pi}$ is the kaon-pion invariant mass. Whilst the P-wave and S-wave contributions can be distinguished since they have separate amplitudes, measuring $m_{K\pi}$ provides additional information regarding how likely a particular event is P-wave or S-wave.

The feasibility regarding whether to fit $m_{K\pi}$ was investigated. This was tested in the q^2 region $1.25 < q^2 < 7$ GeV$^2/c^4$ since at the time of performing this study, the nominal q^2 region for the rare mode fit was $1.25 < q^2 < 7$ GeV$^2/c^4$. There are several advantages of

fitting $m_{K\pi}$, most notably it allows one to better to distinguish the P-wave contribution, which peaks in $m_{K\pi}$, and the S-wave contribution, which is relatively flat in $m_{K\pi}$. This allows one to make precise measurements of the S-wave parameters. Measuring $m_{K\pi}$ thus allows a much better precision of the measurement of the fraction of S-wave and the interference observables.

As described in Section 2.9, the 5D angular distribution only exhibits two symmetries: a phase each for the left-handed and right-handed amplitudes. As a reminder, four amplitudes are set to zero in the basis-fixing condition in the 4D fit. For the 5D fit, only two amplitudes are fixed to zero to fix the amplitude basis, since there are two continuous symmetries of the PDF. The choice $\mathrm{Im}(\mathcal{A}_\perp^R) = \mathrm{Im}(\mathcal{A}_0^L) = 0$ was made. Pseudoexperiments were generated from the amplitude predictions, with the expected Run 1 + Run 2 sample yields. The precision of the two fits were compared by performing a 5D fit (where $m_{K\pi}$ is fitted, and two amplitudes are fixed to zero) and a 4D fit (where $m_{K\pi}$ is not fitted and four amplitudes are fixed to zero).

The change in precision of the amplitudes which float in both fits was examined at points in q^2. The ratio between the 5D and 4D uncertainties (σ_{5D} and σ_{4D} respectively) for the amplitudes was plotted. Therefore if $\frac{\langle\sigma_{5D}\rangle}{\langle\sigma_{4D}\rangle}$ is less than one, there is a precision increase from the 5D fit, and if this is greater than one, there is a precision decrease. Figure 8.1 shows the change in precision for the amplitudes at $q^2 = 4\ \mathrm{GeV}^2/c^4$. In order to make a like-for-like comparison, the fitted amplitudes are transformed back into the untransformed amplitude basis. Comparing the amplitudes which float in both the 4D and 5D fits (all apart from $\mathrm{Im}(\mathcal{A}_\perp^R)$, $\mathrm{Im}(\mathcal{A}_0^L)$, $\mathrm{Re}(\mathcal{A}_0^R)$ and $\mathrm{Im}(\mathcal{A}_0^R)$), there is an increase in precision in the S-wave amplitudes, however there is a considerable decrease in precision for some of the P-wave amplitudes. This was seen across other points in q^2.

From the amplitudes, the observables are computed. The change in precision was also studied for the observables. These are shown in Figure 8.2 for $q^2 = 4\ \mathrm{GeV}^2/c^4$. As seen from this plot, when the amplitudes combine to form the observables, no precision is lost. The precision of the P-wave observables remains unchanged, whereas the precision of the S-wave and interference observables improve. The same behaviour is seen at other points in q^2.

The correlation matrices for the ensembles of 4D and 5D fits are compared and are

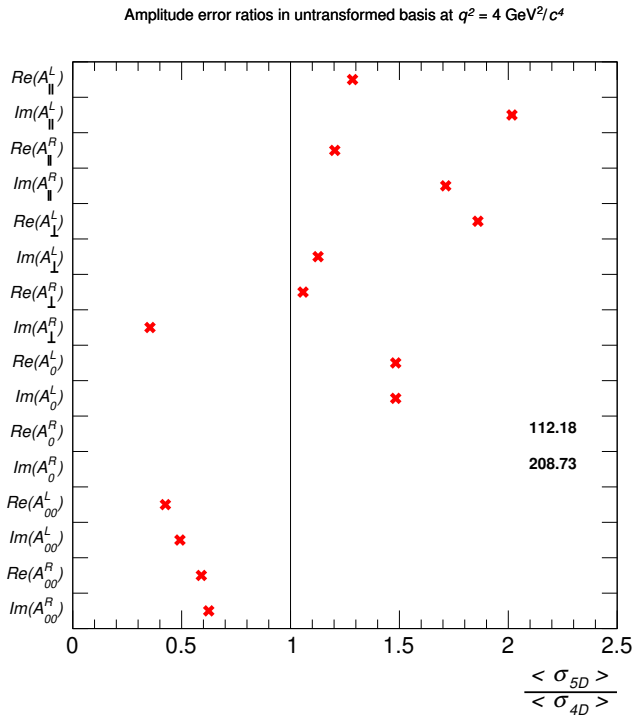


Figure 8.1: Change in amplitude precision when comparing the 5D fit to the 4D fit at $q^2 = 4\ \mathrm{GeV}^2/c^4$, where the fitted amplitudes are transformed back into the original untransformed amplitude basis. The S-wave amplitudes ($\mathrm{Re}(\mathcal{A}_{00}^L)$, $\mathrm{Im}(\mathcal{A}_{00}^L)$, $\mathrm{Re}(\mathcal{A}_{00}^R)$, and $\mathrm{Im}(\mathcal{A}_{00}^R)$) result in an increase in precision in the 5D fits compared to the 4D fits, however the P-wave amplitudes generally have a decrease in precision. Where the points cannot be plotted because they are outside the x axis range, the values of $\frac{\langle\sigma_{5D}\rangle}{\langle\sigma_{4D}\rangle}$ are quoted.

shown as Figures 8.3 and 8.4 respectively. The amplitude ansatz proposed in [11] was used for both ensembles of fits. Whilst the S-wave sees an improvement in the 5D fits compared to the 4D fits, there are larger correlations between amplitudes in the 5D fits.

In order to investigate the large correlations and why the P-wave amplitudes lose precision in the 5D fits, fits were performed where the amplitudes are parameterised as constants in q^2 in order to simplify the fit. The average correlation matrix for an ensemble of pseudoexperiments is shown as Figure 8.5. There are some very large correlations, for example $\mathrm{Corr}(\mathrm{Im}(\mathcal{A}_{\parallel}^L), \mathrm{Im}(A_0^R)) = 0.78$ and $\mathrm{Corr}(\mathrm{Re}(A_{\perp}^L), \mathrm{Re}(A_0^R)) = 0.81$. Such high correlations between P-wave amplitudes are not present in the 4D fits. The 5D fit appears to behave as though there are too many free parameters.

A study was performed examining how much of the symmetry in the 4D PDF gets broken by allowing the PDF to be a function of $m_{K\pi}$. Note that when comparing the 5D fit to the 4D fit, two more amplitudes are measured. This is because $m_{K\pi}$ breaks two

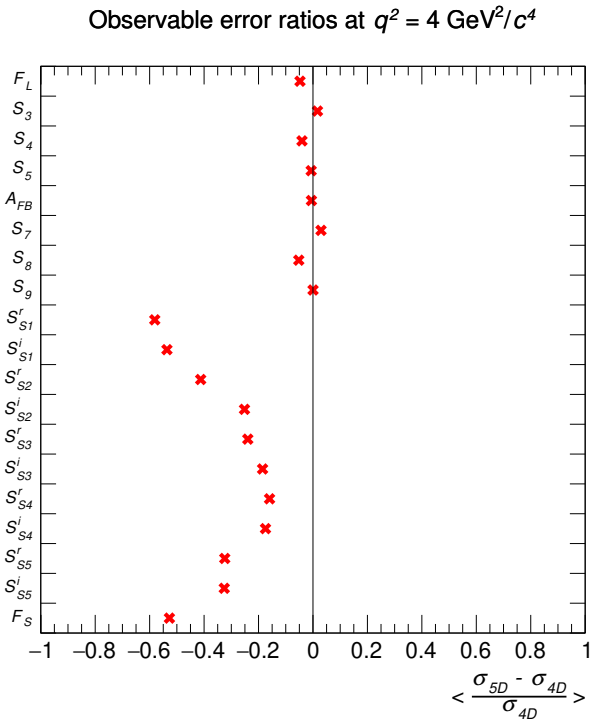


Figure 8.2: Change in observable precision when comparing the 5D fit to the 4D fit at $q^2 = 4\,\mathrm{GeV}^2/c^4$

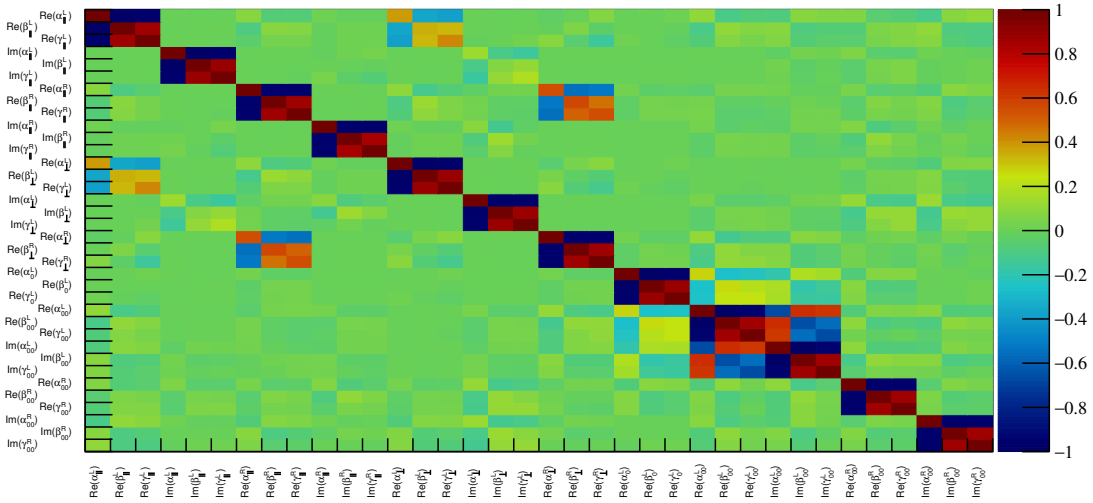


Figure 8.3: Correlation matrix from ensembles of 4D fits to pseudoexperiments. The ansatz proposed in [11] was used.

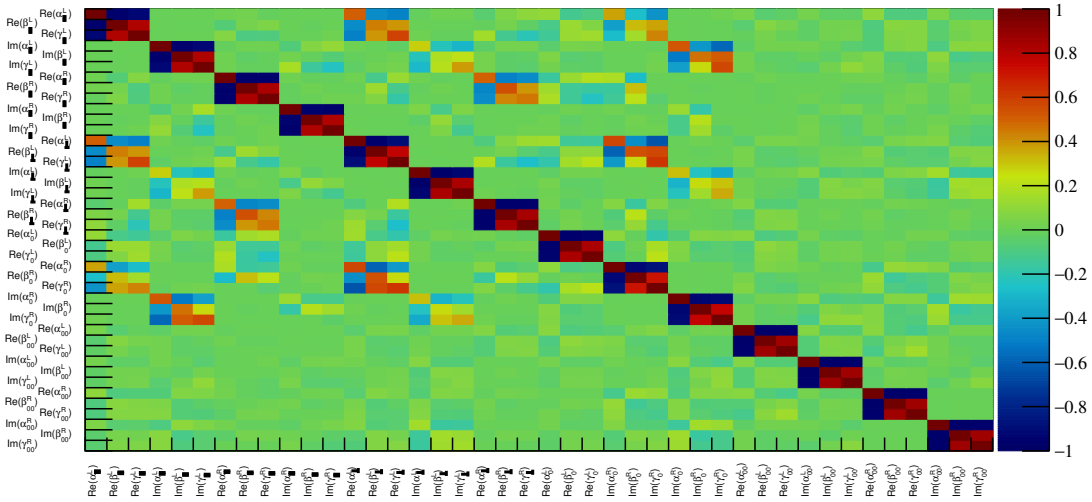


Figure 8.4: Correlation matrix from ensembles of 5D fits to pseudoexperiments. The ansatz proposed in [11] was used.

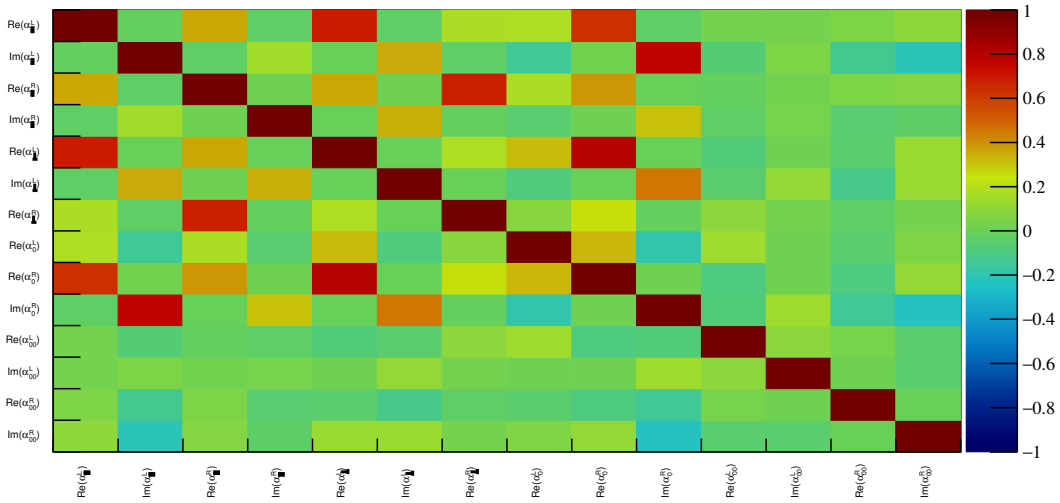


Figure 8.5: Average correlation matrix from 5D fits, where the amplitudes are flat in q^2.

symmetries in the angular distribution. The observables where this breaking of symmetry happens (the interference observables) are investigated in order to understand the fit behaviour. If there are no large differences between the observables in the 4D PDF and the 5D PDF, this would be because the symmetries of the 4D PDF which get broken in the 5D PDF only result in only small deviations. This would explain why the fitter behaves as though there are too many free parameters. The pure P-wave and S-wave terms have no symmetry breaking when transforming from a 4D PDF to a 5D PDF since the absolute

value of the lineshape squared can always be factored out from the observable. Here, the observable S_{S1} is investigated. Explicitly, in the 4D PDF, this is simply

$$S_{S1} \propto Re(\mathcal{A}_{00}^{L}\mathcal{A}_{0}^{L*} + \mathcal{A}_{00}^{R}\mathcal{A}_{0}^{R*}) + CP. \tag{8.1}$$

For the 5D PDF, the P-wave and S-wave lineshapes $\mathcal{A}_P$ and $\mathcal{A}_S$ are included, so

$$S_{S1} \to Re(\mathcal{A}_{00}^{L}\mathcal{A}_{S}\mathcal{A}_{0}^{L*}\mathcal{A}_{P}^{*} + \mathcal{A}_{00}^{R}\mathcal{A}_{S}\mathcal{A}_{0}^{R*}\mathcal{A}_{P}^{*}) + CP. \tag{8.2}$$

Thus

$$S_{S1} \propto Re(\mathcal{A}_{00}^{L}\mathcal{A}_{0}^{L*} + \mathcal{A}_{00}^{R}\mathcal{A}_{0}^{R*})Re(\mathcal{A}_{S}\mathcal{A}_{P}^{*}) - Im(\mathcal{A}_{00}^{L}\mathcal{A}_{0}^{L*} + \mathcal{A}_{00}^{R}\mathcal{A}_{0}^{R*})Im(\mathcal{A}_{S}\mathcal{A}_{P}^{*}) + CP. \tag{8.3}$$

Returning to the q^2-dependent fits, Figure 8.6 shows S_{S1} as a function of $m_{K\pi}$, as explicitly written in Equation 8.3. The 1σ and 2σ error bands from the 5D fits are shown in green and yellow respectively. The true value is shown by the black dotted line and the median is shown by the magenta line. The blue line is the observable S_{S1} computed from the true value of amplitudes in the nominal 4D amplitude basis. This 4D basis has a different basis fixing condition than the 5D basis since the number of symmetries are different. Therefore examining the difference between this line and the results of the 5D ensemble of fits provides a measure of the amount of breaking of symmetry when transforming from a 4D PDF to a 5D PDF.

This plot is made at the q^2 value where the difference between the 4D basis and true observable (i.e. in the 5D basis) is the largest. As mentioned, this difference can be interpreted as the amount of symmetry breaking in this observable. In other words, the fact the blue line is different from the black line is the reason why two more amplitudes are required to be floated in the 5D fits compared to the 4D fits. As seen in this plot, compared to the statistical uncertainty in the 5D fits, this difference is tiny. Thus the fit is not sensitive the symmetry breaking in this observable. This is a decrease in precision in the P-wave amplitudes is seen in the 5D fits, and the large correlations between results in observables which do not lose precision. The equivalent plots for S_{S2}, S_{S3}, S_{S4}, and S_{S5} were made, since these are the other observables where the symmetry breaking occurs.

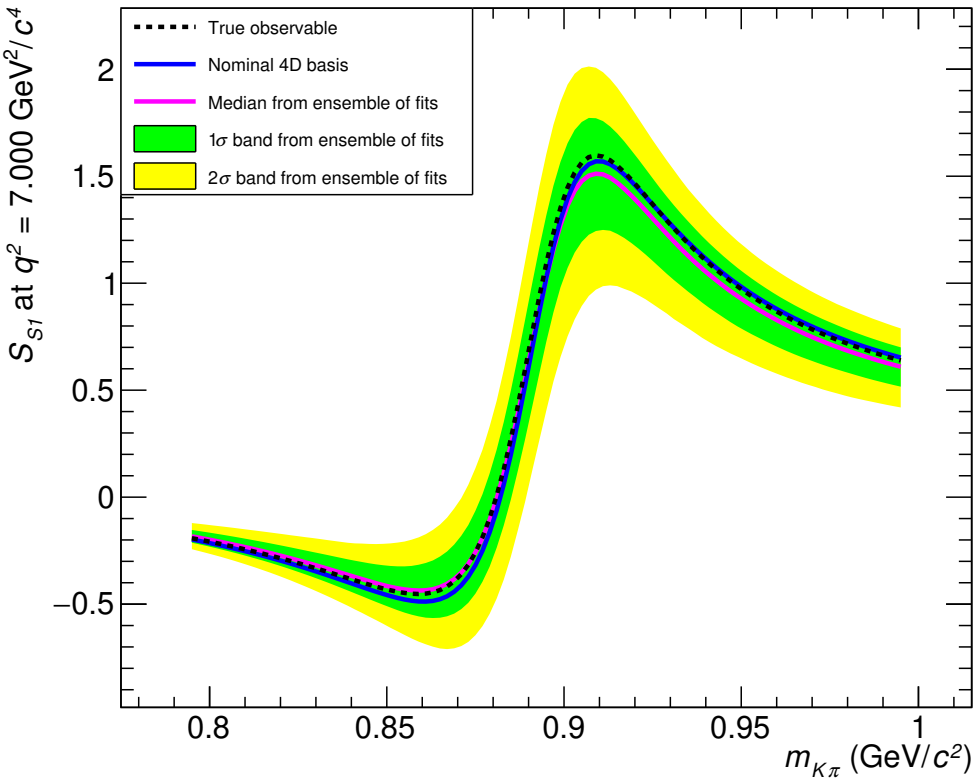


Figure 8.6: The observable S_{S1} as a function of $m_{K\pi}$ at $q^2 = 7$ GeV$^2/c^4$. The median, 1σ and 2σ error bands from ensembles of 5D fits are shown in magenta, green, and yellow respectively. The true value is shown in black and the true values in the 4D amplitude basis are shown in blue.

These are shown as Figure 8.7.

An investigation was performed investigating whether there is a better basis-fixing condition, since another basis could result in a larger difference between the symmetry-breaking observable computed in the 4D basis and the observable computed in the 5D basis. As a reminder for the reader, there are four angles which are used to set the basis, as defined in Section 2.10. They are the two global phases ($\phi_{L,R}$), θ, and η. The global phases are ignored in this study since they correspond to symmetries in both the 4D and 5D PDFs. Nominally in the 4D PDF, the basis fixing condition arises from selecting particular angles which rotate the amplitudes such that the condition is satisfied. However there could be another basis-fixing condition (i.e. $\mathrm{Im}(\mathcal{A}_\perp^R) = C_1$, $\mathrm{Im}(\mathcal{A}_0^L) = C_2$, $\mathrm{Re}(\mathcal{A}_0^R) = C_3$, $\mathrm{Im}(\mathcal{A}_0^R) = C_4$, where C_i are constants), such that the 4D PDF has a larger breaking of

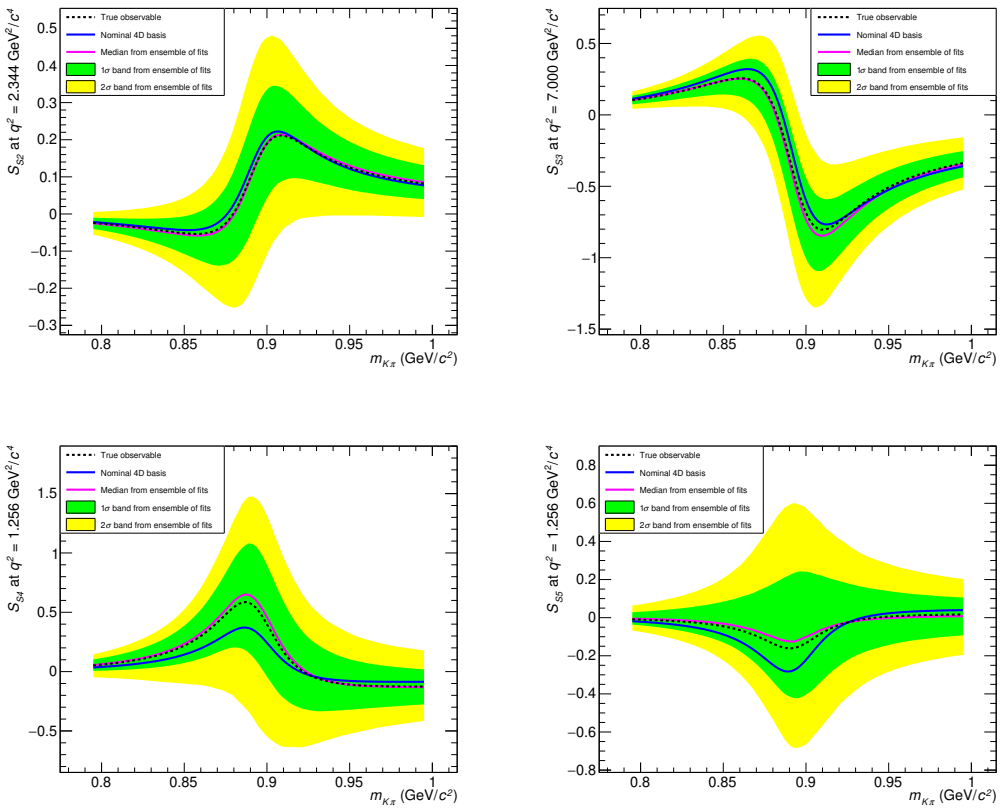


Figure 8.7: Observables S_{S2}, S_{S3}, S_{S4}, and S_{S5} as functions of $m_{K\pi}$ at $q^2 = 7$ GeV$^2/c^4$. The median, 1σ and 2σ error bands from ensembles of 5D fits are shown in magenta, green, and yellow respectively. The true values are shown in black and the true values in the 4D amplitude basis are shown in blue.

symmetry. In other words, another 4D, or crucially another 5D amplitude basis could be used, where the symmetry breaking observables are more sensitive in this basis than the previous basis. This was studied by scanning values of S_{S1} at different values of θ and η and maximising the difference between the observable in 4D basis and the true value of the observable (i.e. in the 5D basis).

Example plots of one-dimensional scans of the angles at $q^2 = 6$ GeV$^2/c^4$ are shown in Figure 8.8. Shown are the angles θ (left) and η (right) as functions of $m_{K\pi}$, where all of the other angles are set to those used in the nominal 4D basis. As seen in these plots, there are certain values of these angles which results in a large difference of S_{S1}, larger than the difference seen in Figure 8.6. Therefore there exists amplitude bases where there is a large breaking of symmetry when transforming from a 4D PDF to a 5D PDF. Thus a 5D fit could be performed (as long as the amplitude ansatz can describe the amplitudes)

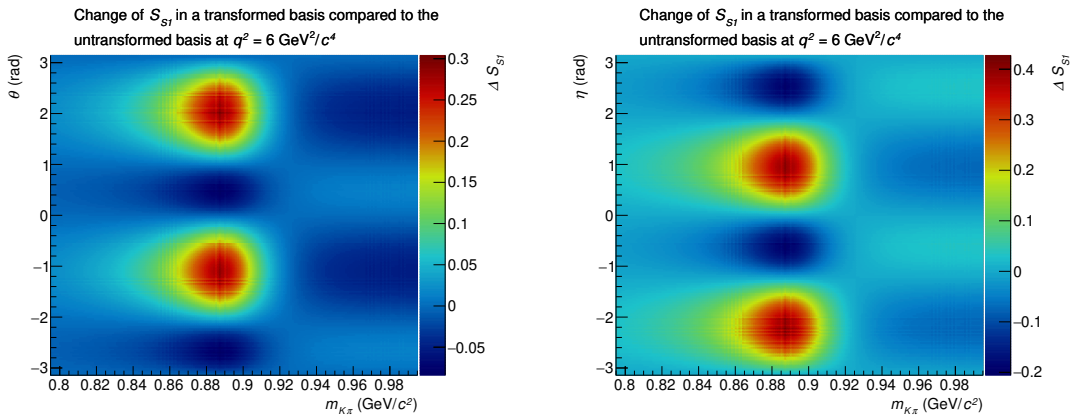


Figure 8.8: The difference between the observable S_{S1} in the 4D basis where the angle is θ (left) or η (right), and all other angles are as used in the nominal 4D basis, and S_{S1} in the 5D basis. This is at $q^2 = 6$ GeV$^2/c^4$.

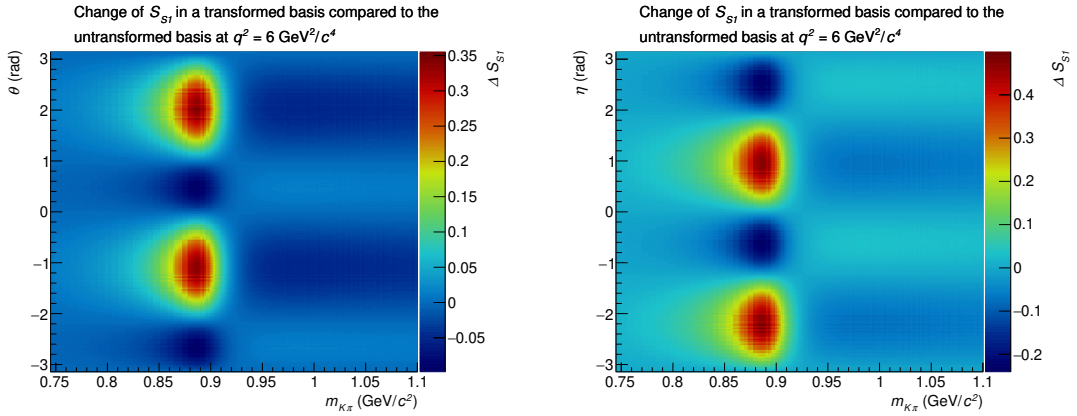


Figure 8.9: The difference between the observable S_{S1} in the 4D basis where the angle is θ (left) or η (right), and all other angles are as used in the nominal 4D basis, and S_{S1} in the 5D basis, in the $m_{K\pi}$ window $0.745 < m_{K\pi} < 1.1$ GeV/c^2. This is at $q^2 = 6$ GeV$^2/c^4$.

without seeing such large correlations between the P-wave amplitudes.

Whether to fit in a wider $m_{K\pi}$ window was considered. If there is a large amount of breaking of symmetry in a region outside the nominal $m_{K\pi}$ window, then the fitter would behave much better. Figure 8.9 shows the same as Figure 8.8 but in a wider $m_{K\pi}$ window. The same plots as Figure 8.9 but at $q^2 = 4$ GeV$^2/c^4$ and $q^2 = 2$ GeV$^2/c^4$ are shown as Figures 8.10 and 8.11 respectively. The breaking of symmetry is maximal at around the K^{*0} P-wave peak. Thus there is not much motivation with the regards to the fit behaviour for increasing the $m_{K\pi}$ window. In addition, a wide $m_{K\pi}$ window results in larger levels of combinatorial and peaking backgrounds, which would need to be dealt with.

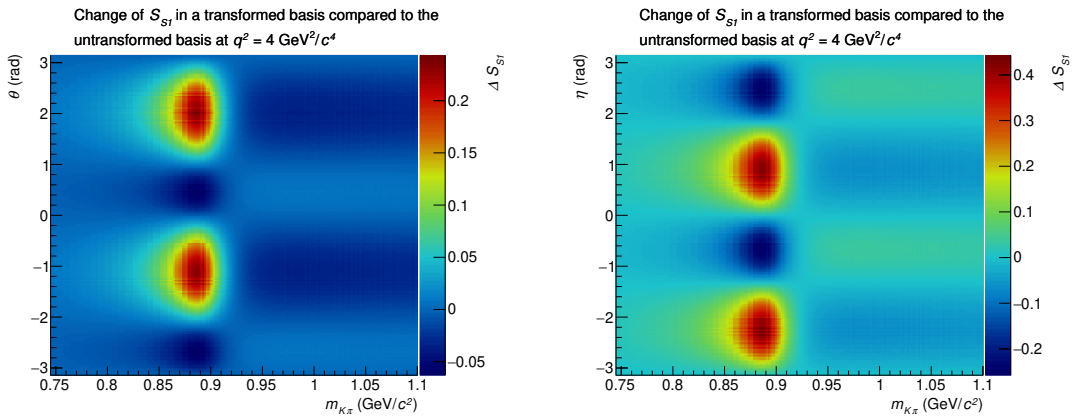


Figure 8.10: The difference between the observable S_{S1} in the 4D basis where the angle is θ (left) or η (right), and all other angles are as used in the nominal 4D basis, and S_{S1} in the 5D basis, in the $m_{K\pi}$ window $0.745 < m_{K\pi} < 1.1$ GeV/c^2. This is at $q^2 = 4$ GeV$^2/c^4$.

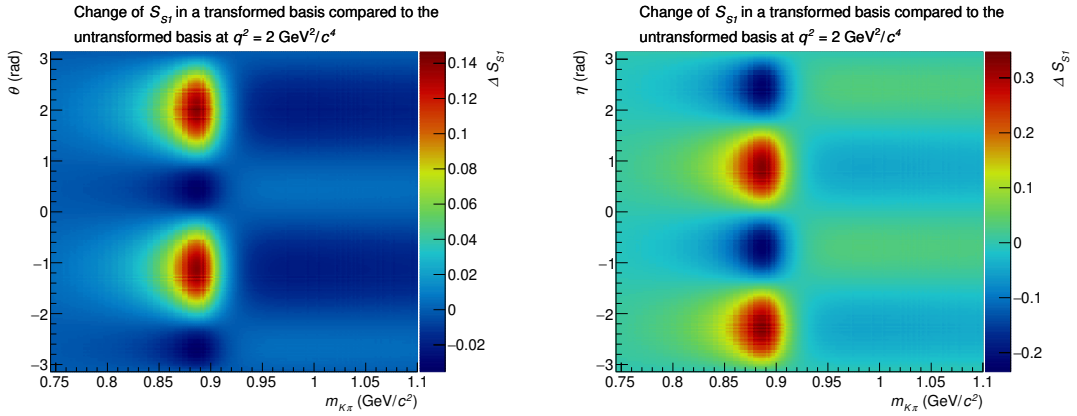


Figure 8.11: The difference between the observable S_{S1} in the 4D basis where the angle is θ (left) or η (right), and all other angles are as used in the nominal 4D basis, and S_{S1} in the 5D basis, in the $m_{K\pi}$ window $0.745 < m_{K\pi} < 1.1$ GeV/c^2. This is at $q^2 = 2$ GeV$^2/c^4$.

Two-dimensional scans of the angles θ and $m_{K\pi}$ were performed to find the angles and q^2 such that the 4D basis with maximal breaking of symmetry. Figure 8.12 shows the same as Figures 8.6 and 8.7, with the basis with maximal breaking of symmetry shown in red. Therefore, there is indeed a 4D basis which is sensitive to a large amount of breaking of symmetry in the PDF when transforming to a 5D basis. One can consider performing the fit in one of these alternative bases — the 5D basis would not be the nominal 5D basis, but an alternative basis where two amplitudes are fixed to some non-zero values. However there are drawbacks with this. The terms where the symmetries get broken tend to involve right-handed amplitudes which are usually small. If these are rotated such that

they become large, or a large amplitude is rotated to become small, this can result in fast-moving or even discontinuous behaviour in some other amplitudes. This makes those amplitudes extremely difficult to be described by a simple ansatz. In addition, since the choice of basis is very model-specific, and the studies performed were based on a particular parameterisation of the Standard Model (even the Standard Model predictions can vary due to different parameters for the non-local contributions), this basis is not necessarily the best for the data. In addition, the S-wave currently has no theoretical predictions, and global fits only use the P-wave observables. Thus it is decided not to fit the PDF as a function of $m_{K\pi}$. It is possible that the PDF as a function of $m_{K\pi}$ will be fitted in the future, given the increase in data and potential interest from theorists in the S-wave and interference parts of the PDF.

8.2 Ansatz choice

Initially, the ansatz proposed in [11] was studied. The ansatz considered was

$$\mathcal{A} = \alpha + \beta q^2 + \frac{\gamma}{q^2}, \tag{8.4}$$

where α, β, and γ are amplitude coefficients. However it was noticed through pseudoexperiment studies the correlations between the amplitude coefficients are extremely high. As shown in Figure 8.3, the absolute values of correlations between amplitude components within an amplitude are greater than 0.95.

Issues were also noted in the behaviour of the log-likelihood profiles. Figure 8.13 shows some log-likelihood profiles from an example pseudoexperiment, where the ansatz proposed in [11] was used. In these plots, for each amplitude parameter of interest, the parameter is fixed to various values in the fit whilst all other parameters float. The fit is performed 200 times for each value, where the free parameters' initial values are randomised for each fit. The log-likelihood is recorded, therefore mapping out the minima. Each point may correspond to different log-likelihoods since some fits may land in the solution corresponding to the global minimum, others may land in local minima. From Figure 8.13, the global minimum and the other global minimum where $\mathcal{A} \to -\mathcal{A}$ are seen. Therefore the global minimum which is quoted is arbitrarily chosen. The red line shows the parabola from

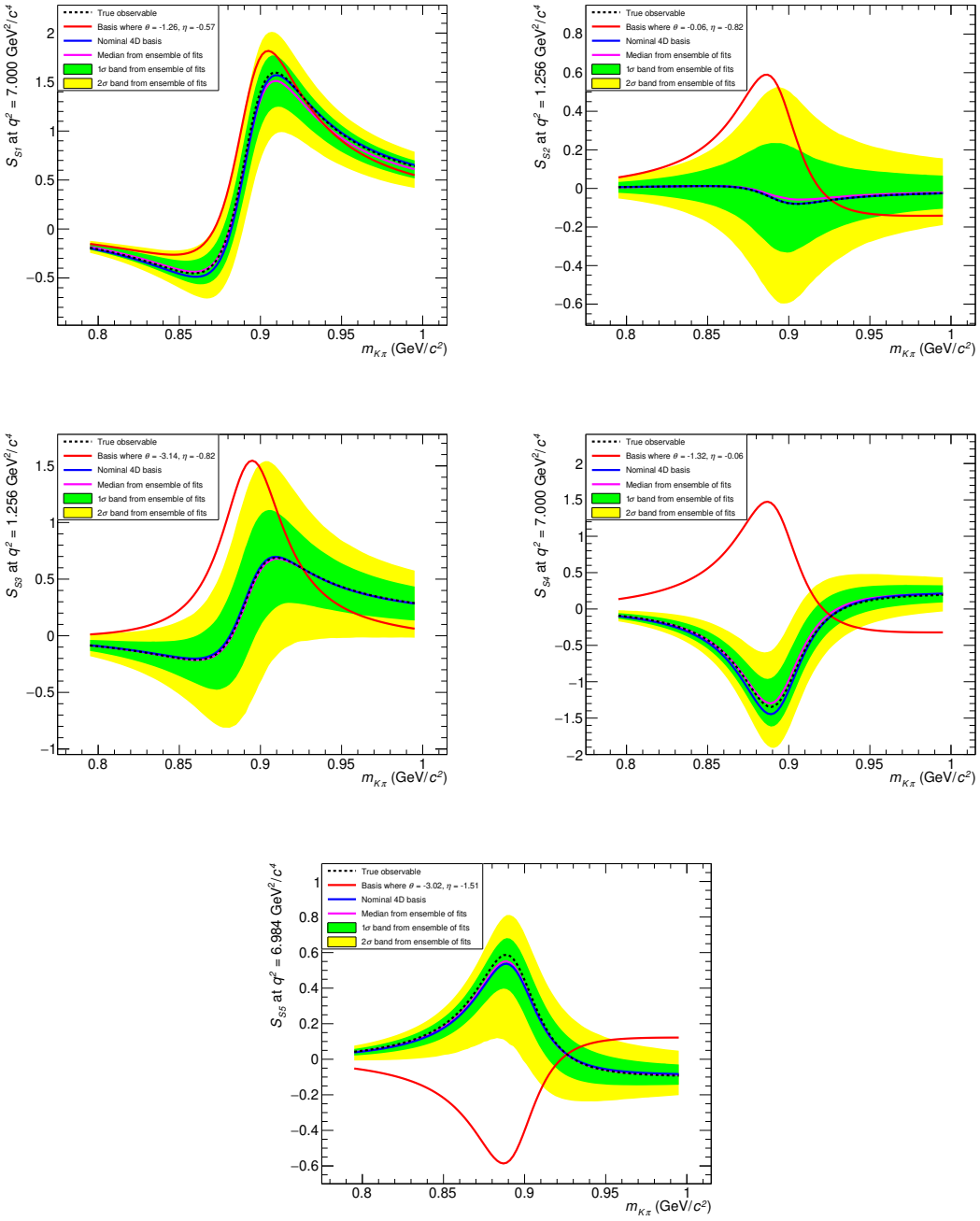


Figure 8.12: Observables S_{S1}, S_{S2}, S_{S3}, S_{S4}, and S_{S5} as functions of $m_{K\pi}$ at the q^2 value with maximal breaking of symmetry. The median, 1σ and 2σ error bands from ensembles of 5D fits are shown in magenta, green, and yellow respectively. The true values are shown in black and the true values in the 4D amplitude basis are shown in blue. Shown in red is the basis with maximal breaking of symmetry.

the Hessian error matrix. There are local minima around $\Delta LLH < 4.5$ away from the global minima. In addition, as seen in the right handed amplitudes, there are local minima very close to the global minimum. Given this presence of local minima, it is extremely challenging to parameterise this log-likelihood surface thus the ansatz proposed in [11] is not adopted.

The amplitude ansatz $\mathcal{A} = \sum \alpha_i L_i(q^{2\prime})$ was considered, where α_i corresponds to the amplitude coefficients, and $L_i(q^{2\prime})$ are Legendre polynomials in $q^{2\prime}$, where $q^{2\prime}$ corresponds to q^2 scaled to the range $-1 \leq x \leq 1$. Ansatzes with three parameters were considered, i.e. $\mathcal{A} = \alpha L_0(q^{2\prime}) + \beta L_1(q^{2\prime}) + \gamma L_2(q^{2\prime})$, where α, β, and γ are amplitude components. Pseudoexperiment studies were performed with four parameters for the ansatz. Due to the orthogonality of Legendre polynomials, no large correlations were seen between amplitude ansatz components, as shown in Figure 8.14. Profile log-likelihoods were plotted for an example pseudoexperiment, where the fit is performed with a 3-parameter Legendre ansatz. Example of these are shown in Figure 8.15. Here at each point the parameter of interest is fixed and all other parameters vary freely in the fit. The fit is performed 200 times for each value, where the free parameters' initial values are randomised.

Given there are no local minima seen here, it is thus much easier to parameterise the profile log-likelihoods. This provides a strong motivation for using Legendre polynomials for the amplitude ansatz. The profile log-likelihoods were studied on data and in cases where the Hessian is not a good parameterisation of the log-likelihood surface, fits to bifurcated parabola provided a good description. The number of parameters for P-wave will be set to four since they can describe a variety of models as shown in Section 7.1.1 and provide a good fit to data, as shown in Chapter 11.

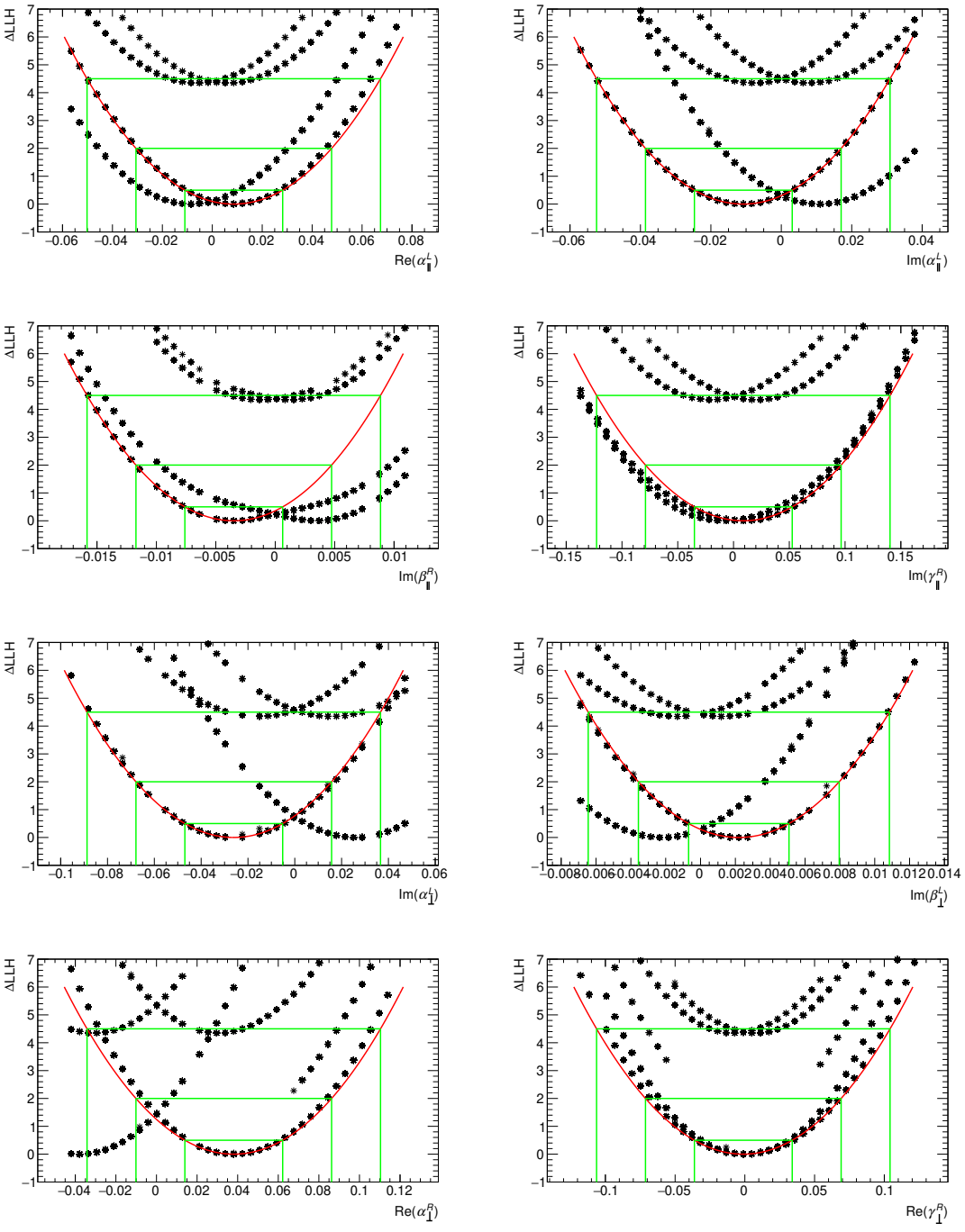


Figure 8.13: Profile log-likelihoods from an example pseudoexperiement, where the fit is performed with the ansatz proposed by [11]. At each point the parameter of interest is fixed and all other parameters vary freely in the fit. The fit is performed 200 times for each value, where the free parameters' initial values are randomised. The two global minima are seen, as well as some local minima. Shown in red is the log likelihood surface from the Hessian error matrix. The green lines indicate 1σ, 2σ, and 3σ from the Hessian error matrix.

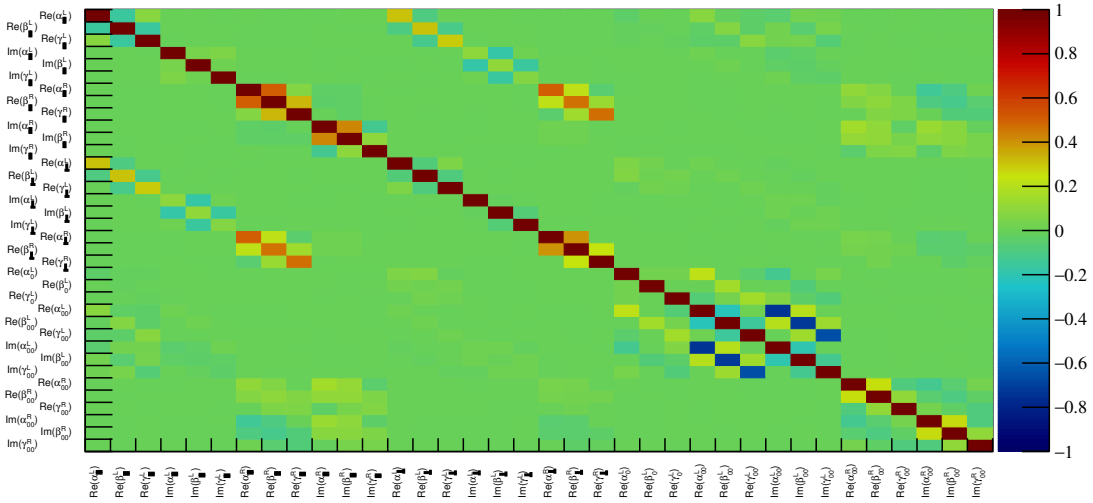


Figure 8.14: Correlation matrix from ensembles of 4D fits to pseudoexperiments. Here the ansatz $\mathcal{A} = \sum \alpha_i L_i(q^{2\prime})$ was used, where α_i corresponds to the amplitude coefficients, and $L_i(q^{2\prime})$ are Legendre polynomials in q^2, where q^2 is scaled to the range $-1 \leq x \leq 1$.

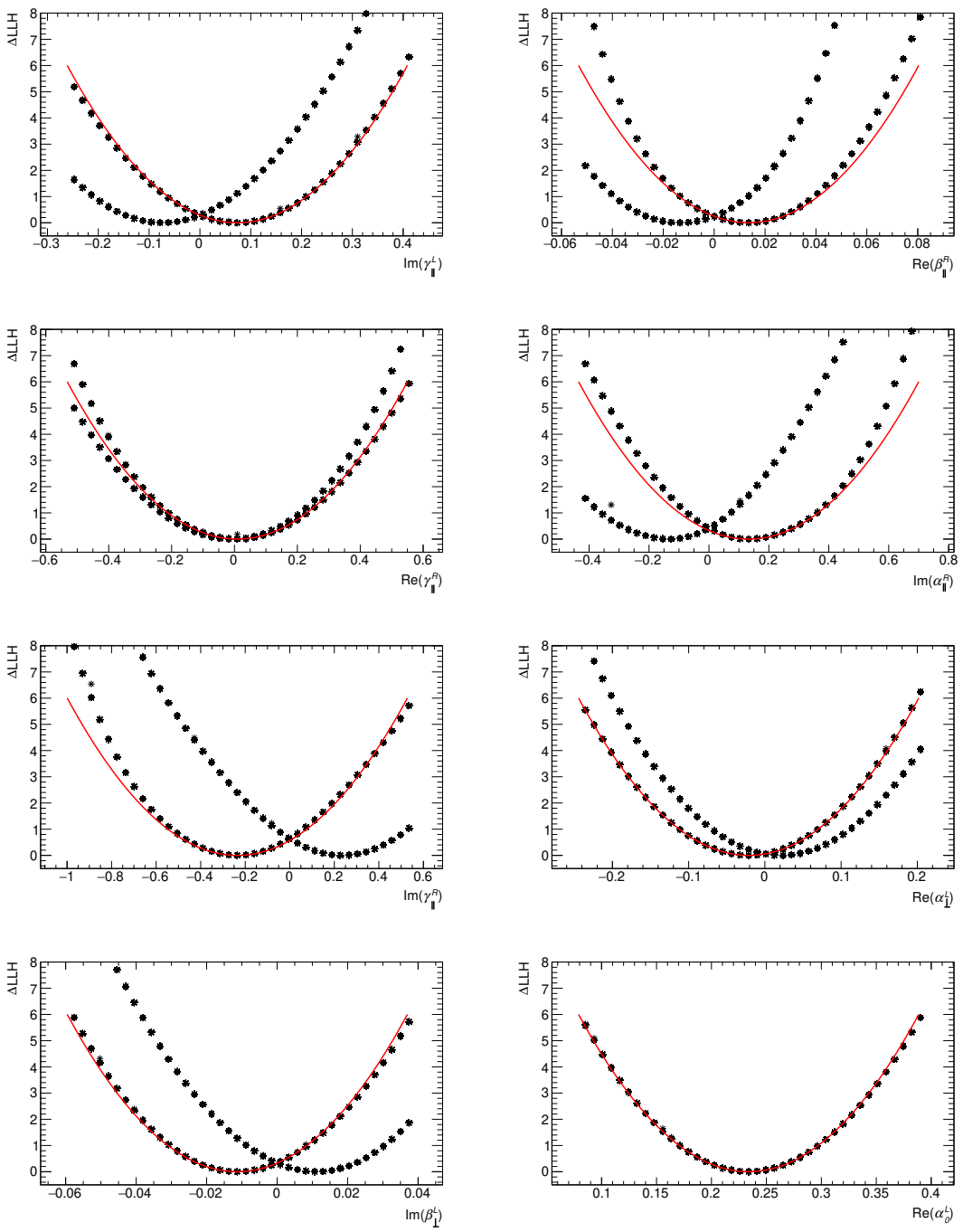


Figure 8.15: Profile log-likelihoods from an example pseudoexperiement, where the fit is performed with a three-parameter Legendre polynomial ansatz. At each point the parameter of interest is fixed and all other parameters vary freely in the fit. The fit is performed 200 times for each value, where the free parameters' initial values are randomised. The two global minima are seen. There are no local minima present. Shown in red is the log likelihood surface from the Hessian error matrix.

8.3 Study of apparent fit biases

In order to validate the fitter to ensure it is working before applying it to data, pseudoexperiments are generated from the amplitude ansatz predictions and fit back based on the four parameter Legendre polynomial ansatz in the region $1.25 < q^2 < 7$ GeV$^2/c^4$.

The q^2-dependent amplitudes are computed from the fitted amplitude components. The distributions of the amplitudes from ensembles of fits are shown as Figure 8.16. The black and magenta lines show the true value and median respectively, and the green and yellow bands shows the 1σ and 2σ error bands. From the amplitudes, observables are computed, which are shown in Figures 8.17 and 8.18.

As seen in Figures 8.16, 8.17 and 8.18 there appears to be some fit biases. These apparent biases are largest in $\mathrm{Re}(\mathcal{A}_\parallel^L)$, S_4 and P_4'. The distribution of fits for $\mathrm{Re}(\mathcal{A}_\parallel^L)$ is shown as Figure 8.19. Therefore the most likely value of $\mathrm{Re}(\mathcal{A}_\parallel^L)$ indeed appears to be biased. This was also seen in S_4, and P_4'.

Pseudoexperiment studies were performed at high sample yields (10× nominal sample yields) to check whether the the median gets closer to the true value as the sample yields is increased. Figures 8.20, 8.21 show some amplitudes and observables where the pseduoexperiments are generated with 1× (left), and 10× (right) nominal sample yields. The difference between the median and true values decrease as the sample yields is increased. Thus the fitter framework is correctly working. Additionally, the distribution of fits were plotted at $q^2 = 6$ GeV$^2/c^4$ for $\mathrm{Re}(\mathcal{A}_\parallel^L)$, S_4, and P_4', where the pseudoexperiments are generated with 1× and 10× nominal sample yields, as shown in Figure 8.22. The true value is shown in black and the median is shown in magenta. The apparent biases reduce as sample yields is increased.

Studies were performed investigating these apparent biases, and it was found that there are no genuine biases. The apparent biases in the observables are in fact artefacts of projecting a physical boundary visible in the S_3/S_4 plane onto a single observable. There are physical boundaries in the PDF since the PDF cannot be less than zero and the observables are combinations of amplitudes normalised to the total P-wave rate. Figure 8.23 shows the physical region of S_3 and S_4. The red point shows the the median and statistical uncertainty from an ensemble of fits at $q^2 = 6$ GeV$^2/c^4$. The SM prediction and thus the

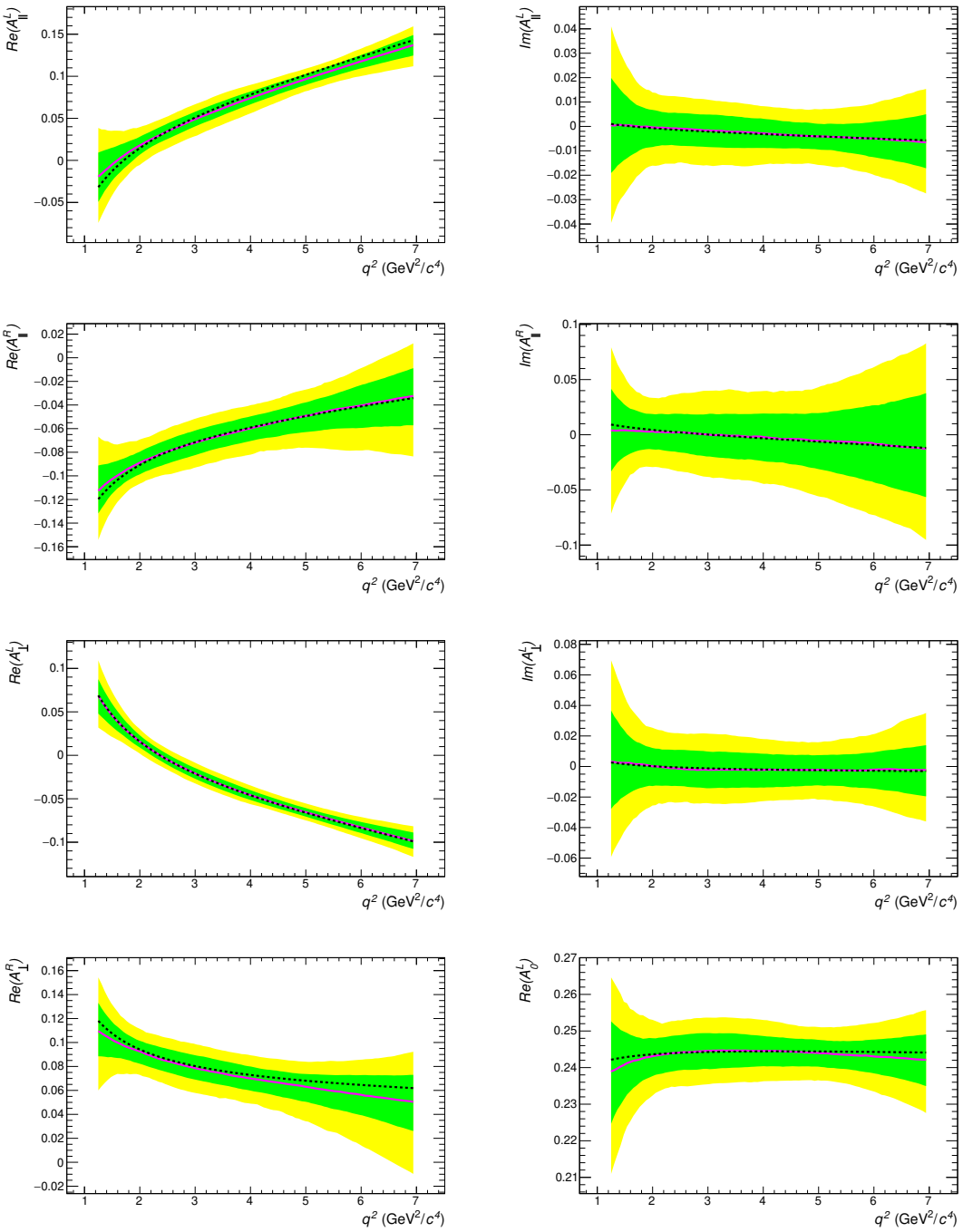


Figure 8.16: P-wave amplitudes from ensembles of fits in the $1.25 < q^2 < 7$ GeV$^2/c^4$ region. The black and magenta lines show the true value and median respectively, and the green and yellow bands shows the 1σ and 2σ error bands.

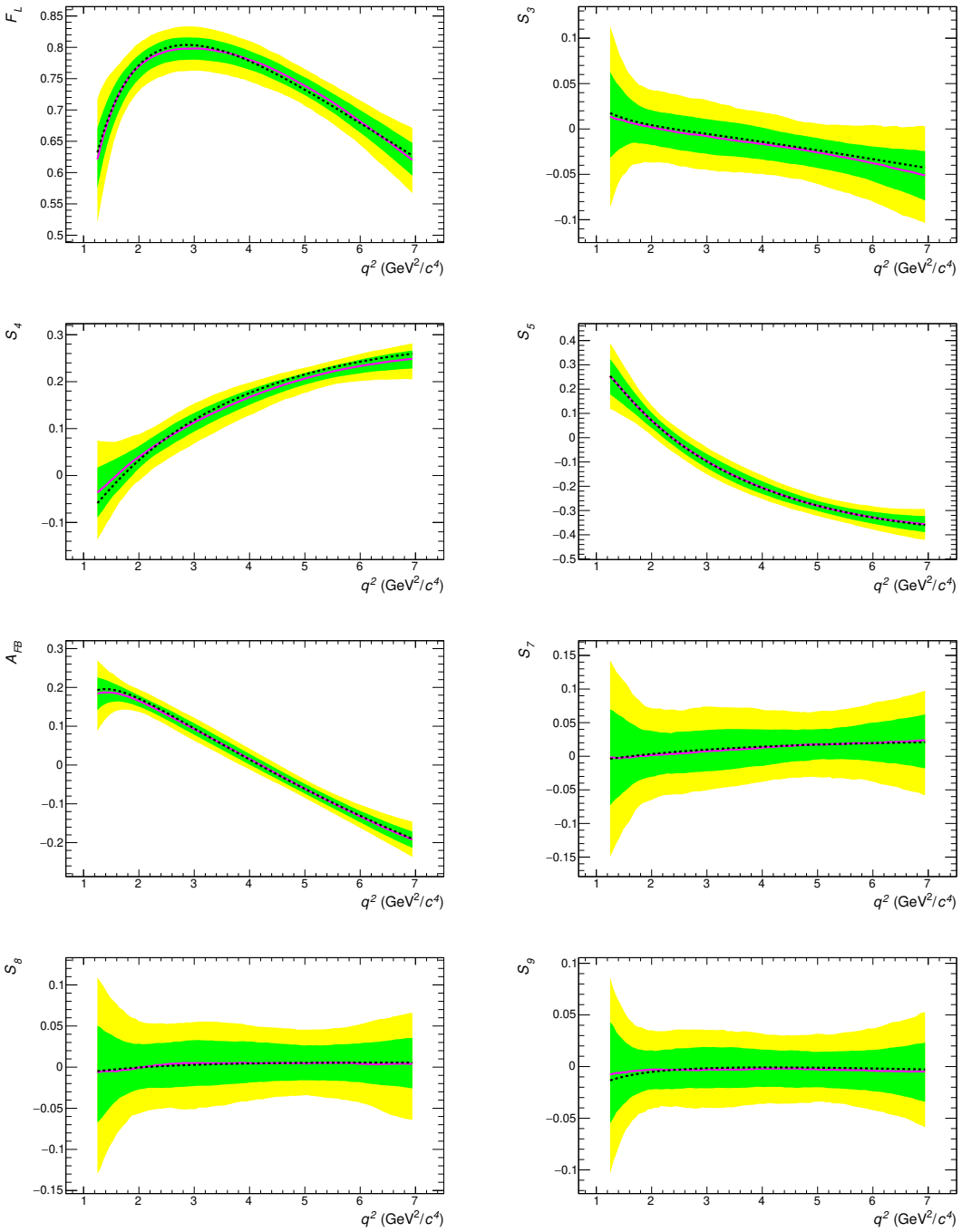


Figure 8.17: Observables (Ss) from ensembles of fits in the $1.25 < q^2 < 7$ GeV$^2/c^4$ region. The black and magenta lines show the true value and median respectively, and the green and yellow bands shows the 1σ and 2σ error bands.

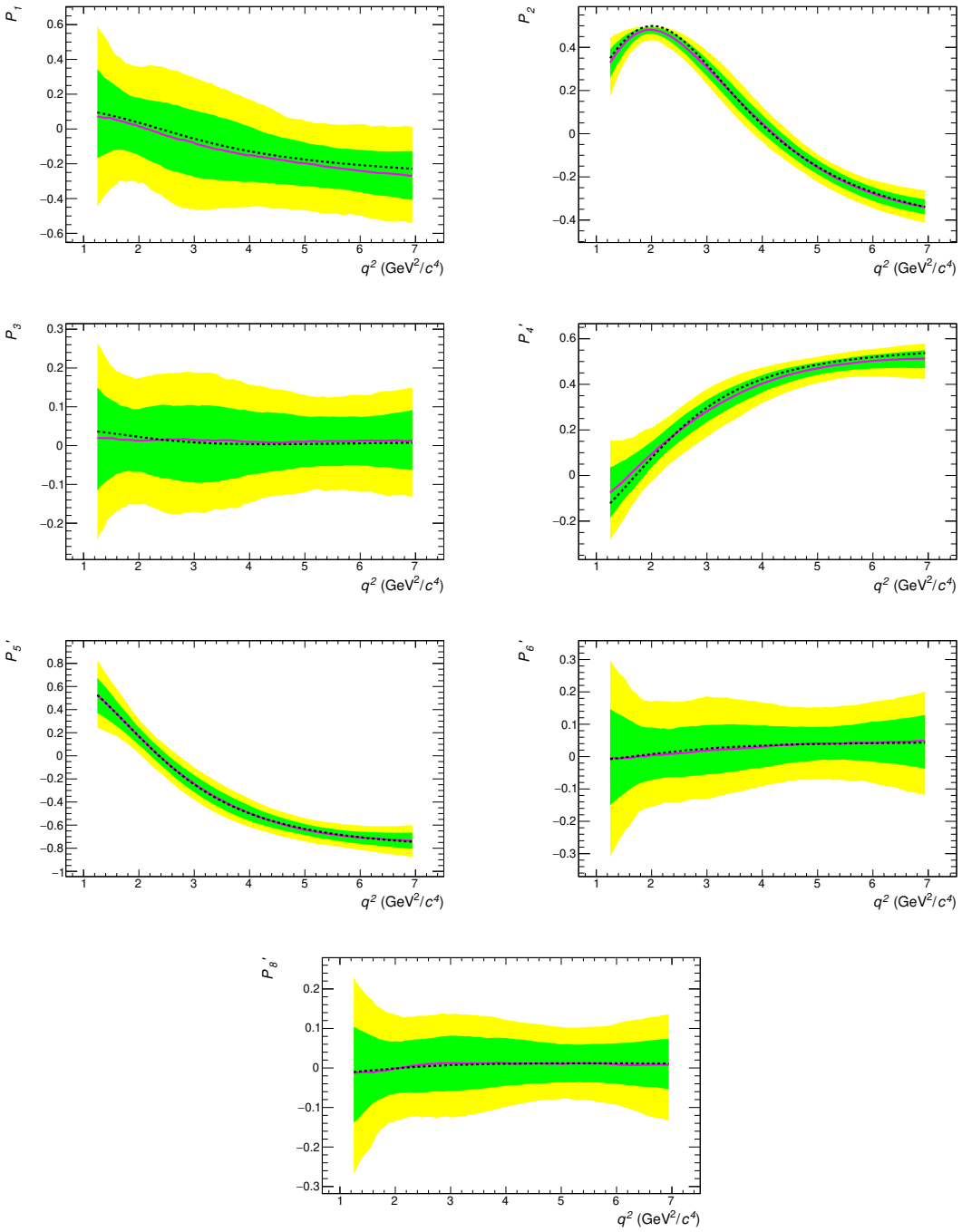


Figure 8.18: Observables (Ps) from ensembles of fits in the $1.25 < q^2 < 7$ GeV$^2/c^4$ region. The black and magenta lines show the true value and median respectively, and the green and yellow bands shows the 1σ and 2σ error bands.

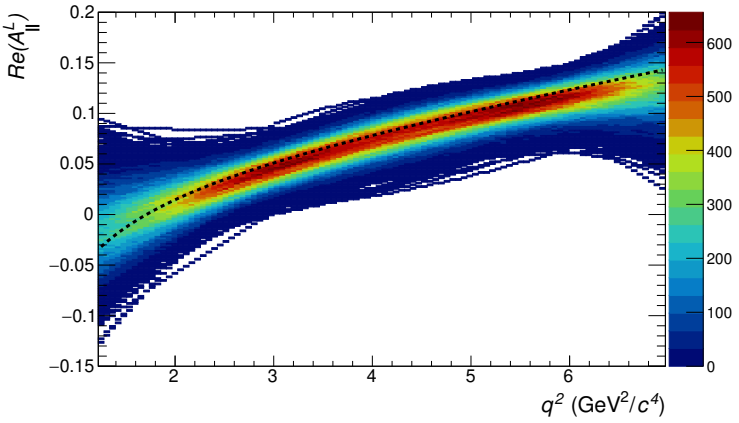


Figure 8.19: Distribution of fits of Re($\mathcal{A}_{\parallel}^{L}$) of pseudoexperiments generated from the SM amplitude predictions. The black line shows the true value.

results from the fits are close to the physical boundary.

In order to simplify the dimensionality of the problem, pseudoexperiments were generated at $q^2 = 6$ GeV$^2/c^4$ and fits were performed where only Re($\mathcal{A}_{\parallel}^{L}$) and the right-handed amplitudes are floated. The bias in Re($\mathcal{A}_{\parallel}^{L}$) is still seen. Figure 8.24 shows Im($\mathcal{A}_{\parallel}^{R}$) and Re($\mathcal{A}_{\parallel}^{L}$) (left), and S3 and S4 (right) at $q^2 = 6$ GeV$^2/c^4$. The true values are shown in black. In addition, Figure 8.25 shows the same but where only Im($\mathcal{A}_{\parallel}^{R}$) and Re($\mathcal{A}_{\parallel}^{L}$) are floated in the fit. Here the best fit point is unbiased and projecting this 2D histogram onto 1D results in an apparent bias. It was also found that applying a harsh cut in S_3, where all fits with $S_3 > -0.02$ are removed, results in no bias in Re($\mathcal{A}_{\parallel}^{L}$), S_4, and P_4'.

The bias in S_4 is an artefact from projecting the physical boundary onto one dimension. Thus S_4 may appear biased even though the most likely value when taking into account all of the observables is unbiased. Re($\mathcal{A}_{\parallel}^{L}$) is highly correlated with S_4, since $S_4 \propto$ Re($\mathcal{A}_{\parallel}^{L}$)Re($\mathcal{A}_{0}^{L}$) and Re($\mathcal{A}_{0}^{L}$) is a large amplitude which is relatively flat in q^2. Since the amplitudes are used to construct the observables, the effect from the boundary is also seen in the amplitudes. This was checked by generating amplitudes based on their true SM values, and sampling from the fit covariance matrix. Figure 8.26 shows this for Re($\mathcal{A}_{\parallel}^{L}$) (left) and Re($\mathcal{A}_{0}^{L}$) (right). From the amplitudes, observables are computed. Figure 8.27 shows the distribution of S_4 and P_4', where apparent biases are seen. The apparent bias shown in the observables is seen, so projecting the boundary onto 1D has an effect. This study was applied to the full q^2 range. Here, events are sampled according to the true amplitudes and fit covariance

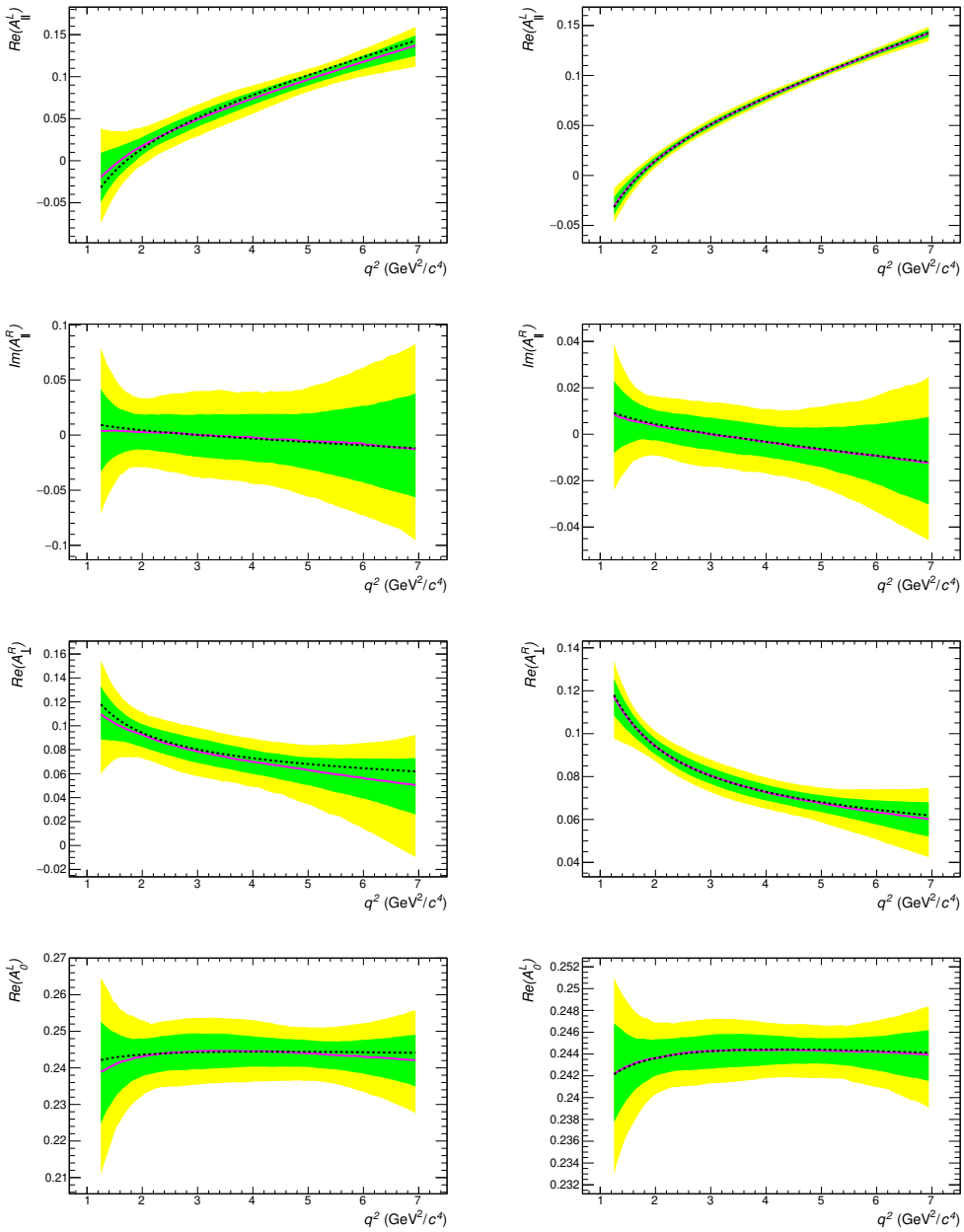


Figure 8.20: Some P-wave amplitudes from ensembles of fits in the $1.25 < q^2 < 7$ GeV$^2/c^4$ region, where the pseudoexperiments are generated with nominal sample yields (left) and 10× nominal sample yields (right). The black and magenta lines show the true value and median respectively, and the green and yellow bands shows the 1σ and 2σ error bands.

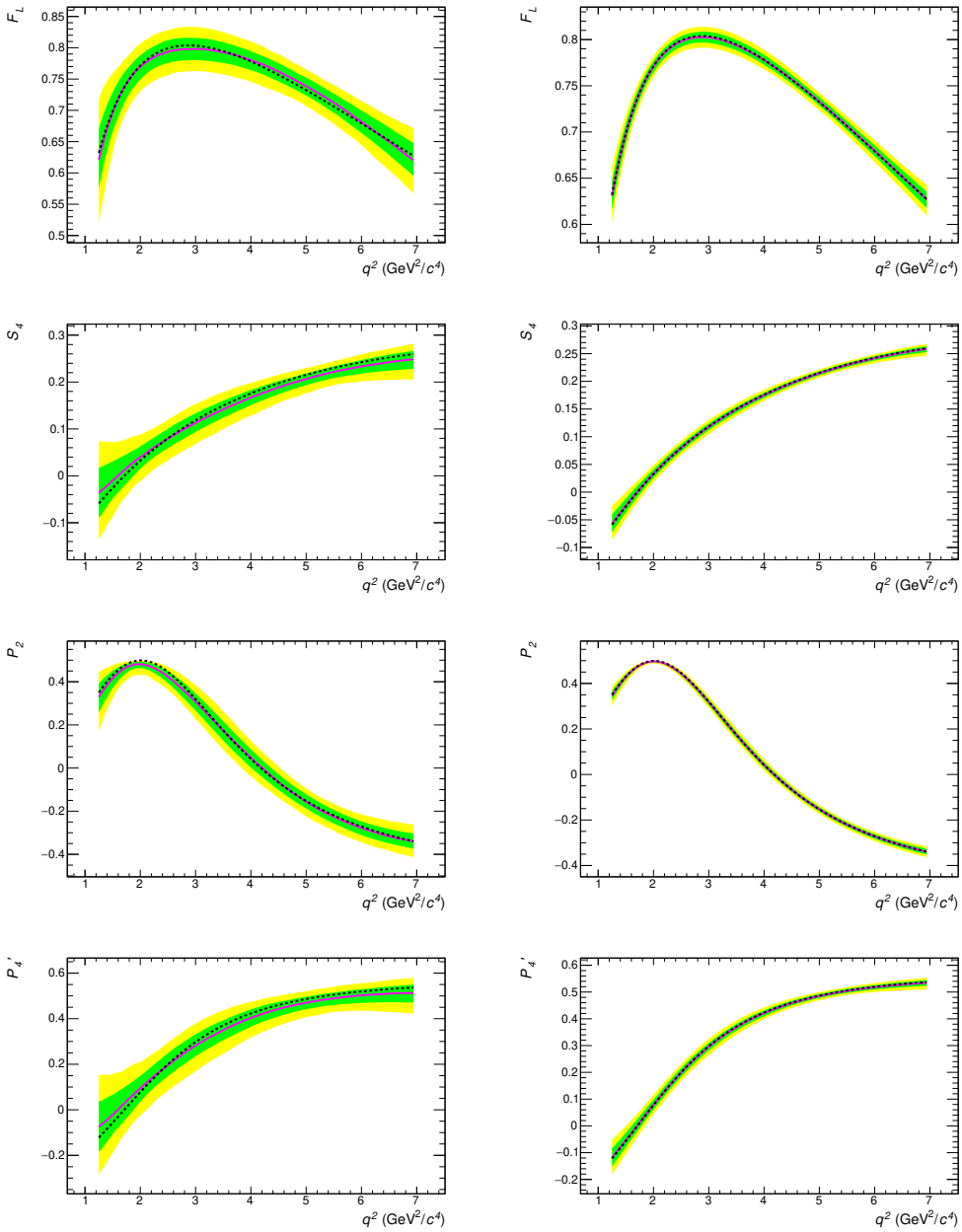


Figure 8.21: Some P-wave observables from ensembles of fits in the $1.25 < q^2 < 7$ GeV$^2/c^4$ region, where the pseudoexperiments are generated with nominal sample yields (left) and 10× nominal sample yields (right). The black and magenta lines show the true value and median respectively, and the green and yellow bands shows the 1σ and 2σ error bands.

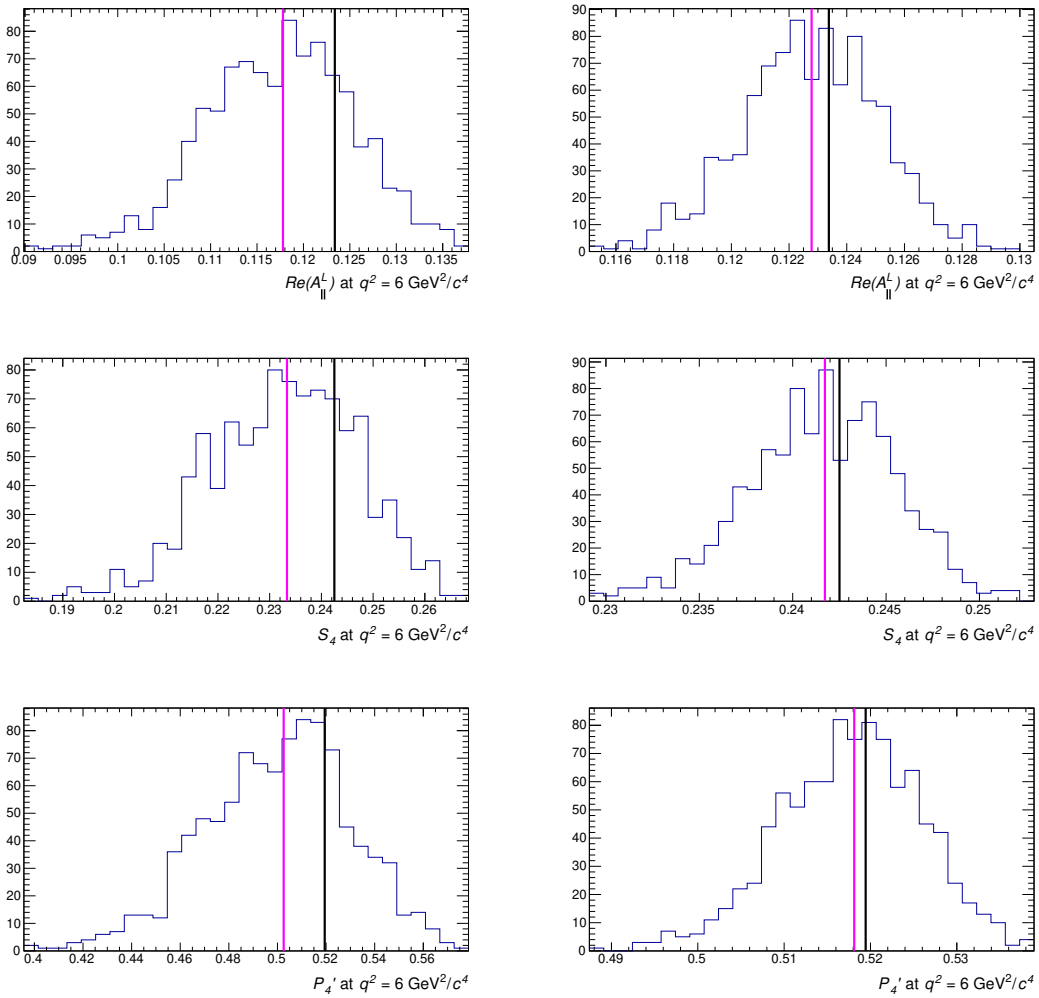


Figure 8.22: Distribution of fits to pseudoexperiments generated at 1× (left) and 10× (right) nominal sample yields. Shown are $\mathrm{Re}(\mathcal{A}_\parallel^L)$, S_4, and P_4' at $q^2 = 6$ GeV$^2/c^4$. The black and magenta lines show the true value and median respectively.

matrix, as shown in Figure 8.28 for $\mathrm{Re}(\mathcal{A}_\parallel^L)$ (left) and $\mathrm{Re}(\mathcal{A}_0^L)$ (right). By construction, there are no biases in this test. Figure 8.29 shows the distribution of S_4 (left) and P_4' (right), where the apparent biases are seen.

Generating pseudoexperiments close to and away from the S3/S4 boundary results in fits with the parameter $\mathrm{Re}(\mathcal{A}_\parallel^L)$ appearing biased and unbiased respectively. Figures 8.30 and 8.31 show this. Therefore the physical boundary has a profound effect on whether the observable appears biased in 1D. However this does not conclude whether there is a genuine bias in $\mathrm{Re}(\mathcal{A}_\parallel^L)$.

A method to check if there is a genuine bias in the multidimensional space has been developed. This is by selecting the region around the most likely value of a parameter

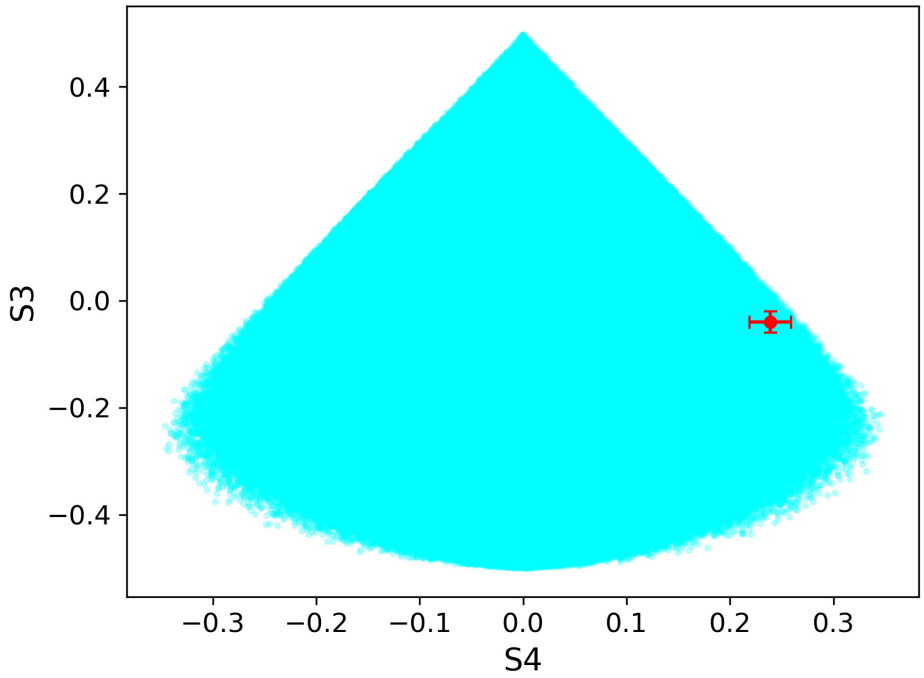


Figure 8.23: Physical region of S_3 and S_4. The red point shows the median and statistical uncertainty from ensembles of fits at $q^2 = 6$ GeV$^2/c^4$.

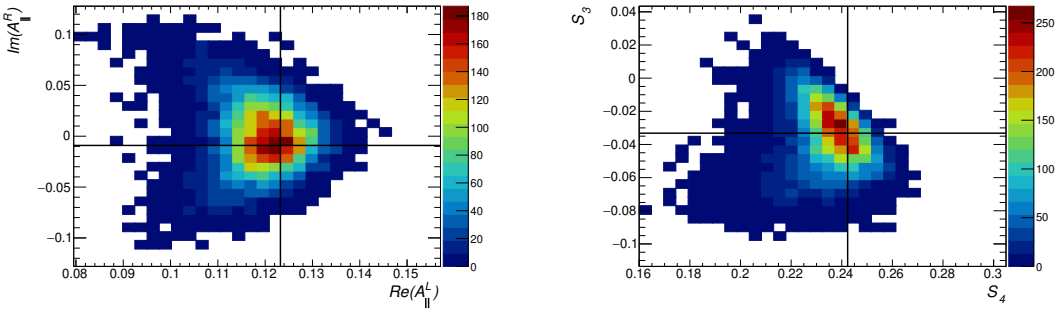


Figure 8.24: Im($\mathcal{A}_\parallel^R$) and Re($\mathcal{A}_\parallel^L$) (left), and S3 and S4 (right) at $q^2 = 6$ GeV$^2/c^4$, where only Re($\mathcal{A}_\parallel^L$) and the right-handed amplitudes are floated. The true values are shown in black.

of interest and checking whether the bias reduces as the region selected gets smaller. By decreasing the size of the region selected around the most likely value, the effect of the physical boundary decreases. Here, the parameter of interest is Re($\mathcal{A}_\parallel^L$) at $q^2 = 6$ GeV$^2/c^4$. The lower and upper X% of fits of all of the other amplitudes and the amplitude Re($\mathcal{A}_\parallel^L$) is examined. X is increased to see whether the bias decreases.

Figure 8.32 shows the distribution of Re($\mathcal{A}_\parallel^L$) at $q^2 = 6$ GeV$^2/c^4$ where: none; upper

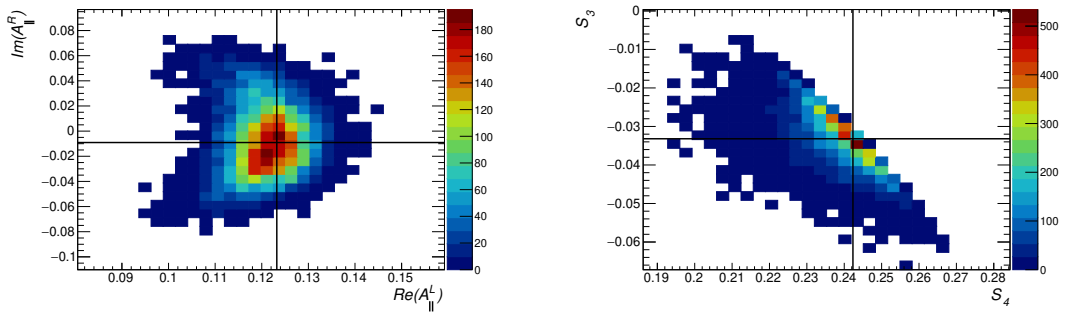


Figure 8.25: $\mathrm{Im}(\mathcal{A}_\parallel^R)$ and $\mathrm{Re}(\mathcal{A}_\parallel^L)$ (left), and S3 and S4 (right) at $q^2 = 6$ GeV$^2/c^4$, where only $\mathrm{Re}(\mathcal{A}_\parallel^L)$ and $\mathrm{Im}(\mathcal{A}_\parallel^R)$ are floated. The true values are shown in black.

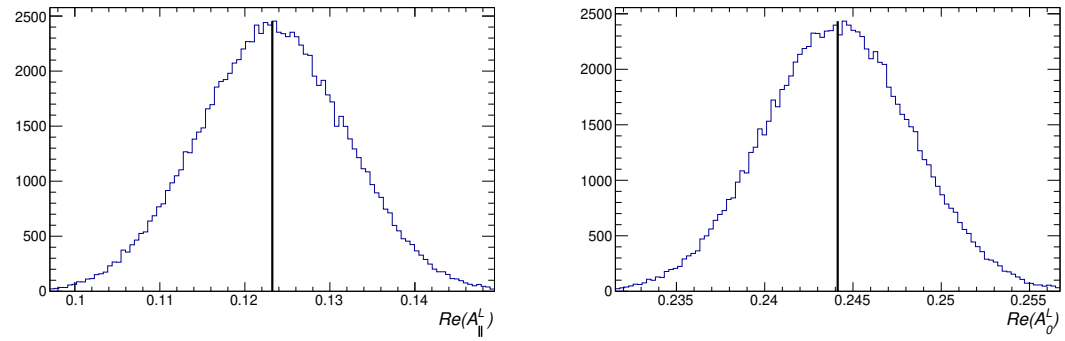


Figure 8.26: Distribution of pseudoevents when sampling from the SM amplitude predictions at $q^2 = 6$ GeV$^2/c^4$ and fit covariance matrix (where constant amplitudes are fitted) for $\mathrm{Re}(\mathcal{A}_\parallel^L)$ (left) and $\mathrm{Re}(\mathcal{A}_0^L)$ (right). The black line shows the true values of the amplitudes.

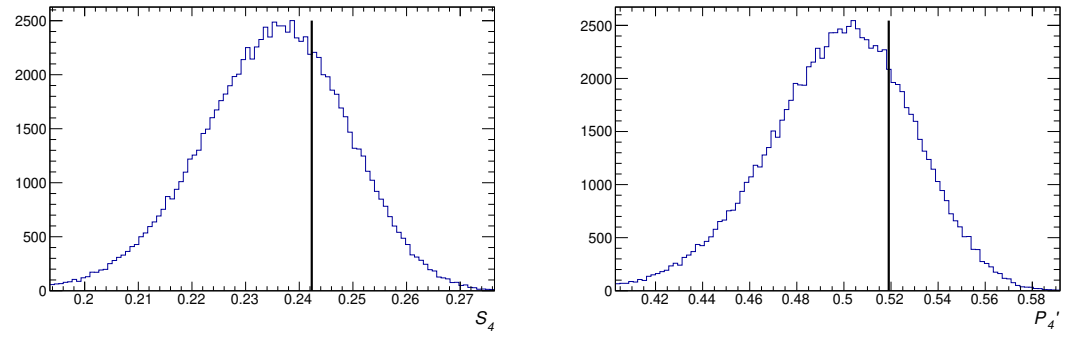


Figure 8.27: Distribution of pseudoevents when sampling from the true SM amplitude predictions at $q^2 = 6$ GeV$^2/c^4$ and fit covariance matrix (where constant amplitudes are fitted) for S_4 (left) and P_4' (right). The black line shows the true values of the amplitudes.

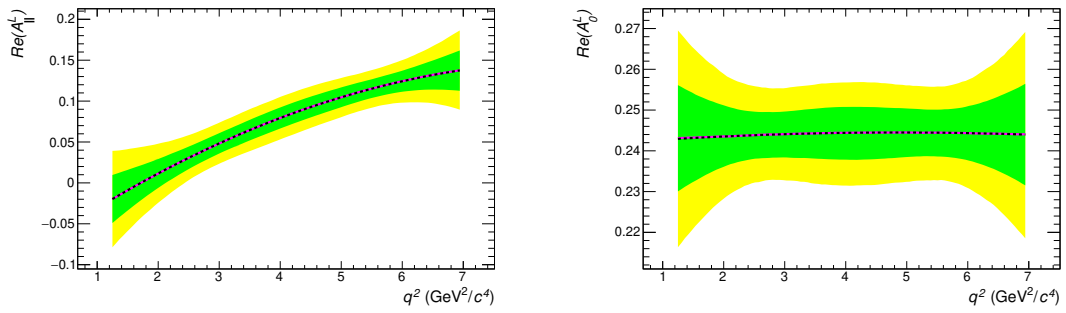


Figure 8.28: Distribution of pseudoevents sampled from the amplitudes and fit covariance matrix for $\mathrm{Re}(\mathcal{A}_{\parallel}^{L})$ (left) and $\mathrm{Re}(\mathcal{A}_{0}^{L})$ (right). The black line shows the true values of the amplitudes. The magenta line shows the median and green and yellow bands show the 1σ and 2σ regions respectively.

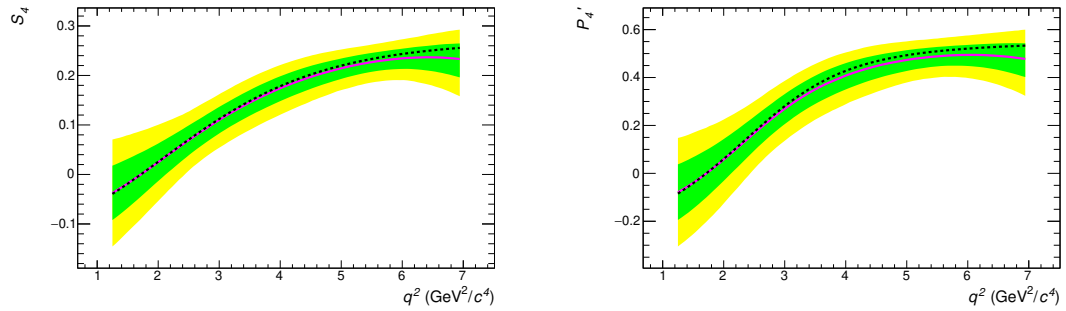


Figure 8.29: Distribution of pseudoevents sampled from the amplitudes and fit covariance matrix for S_4 (left) and P_4' (right). The black line shows the true values of the observables. The magenta line shows the median and green and yellow bands show the 1σ and 2σ regions respectively.

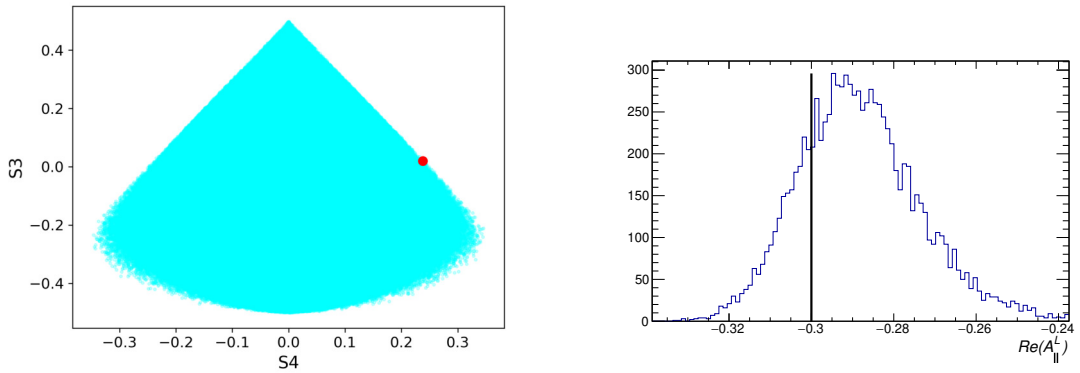


Figure 8.30: Pseudoexperiments are generated close to the S3/S4 boundary (from the red point on the left plot) resulting in distribution of fits to $\mathrm{Re}(\mathcal{A}_{\parallel}^{L})$ (right). The true value is shown by the black line. There is an apparent bias in this amplitude.

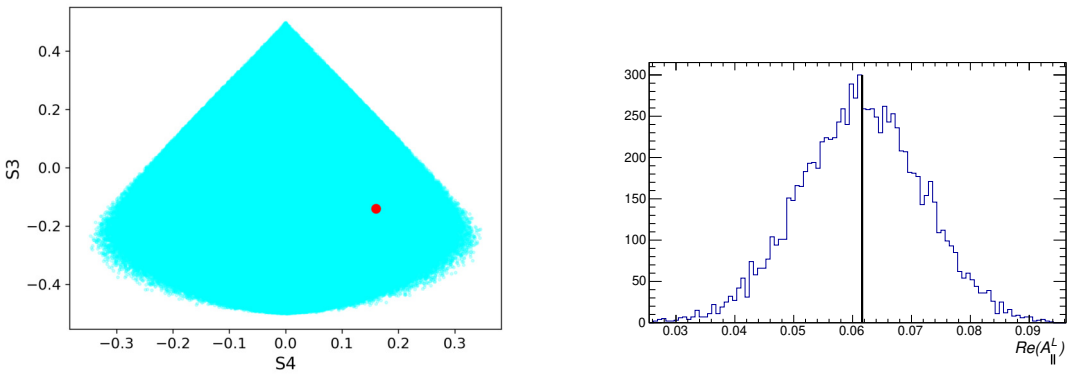


Figure 8.31: Pseudoexperiments are generated away from the S3/S4 boundary (from the red point on the left plot) resulting in distribution of fits to Re($\mathcal{A}_{\parallel}^{L}$) (right). The true value is shown by the black line. There is no apparent bias in this amplitude.

and lower 5%; 10%; 15%; and 20% of fits in the other amplitudes are removed. As a region which includes the most likely value gets smaller, the apparent bias seen in Re($\mathcal{A}_{\parallel}^{L}$) reduces. The same plot at $q^2 = 4$ GeV$^2/c^4$ is shown as Figure 8.33. The bias seen in this amplitude is seen to reduce as the region containing the most likely value is decreased.

Thus it is concluded that this apparent bias is not a genuine bias. Therefore no systematic uncertainty is needed to account for fit biases since the most likely value of all of the amplitudes and the observables combined is unbiased. It should be noted that one should not use the result from one observable alone, due to the apparent bias which is an artefact of the physical boundary. All of the P-wave observables must be used together.

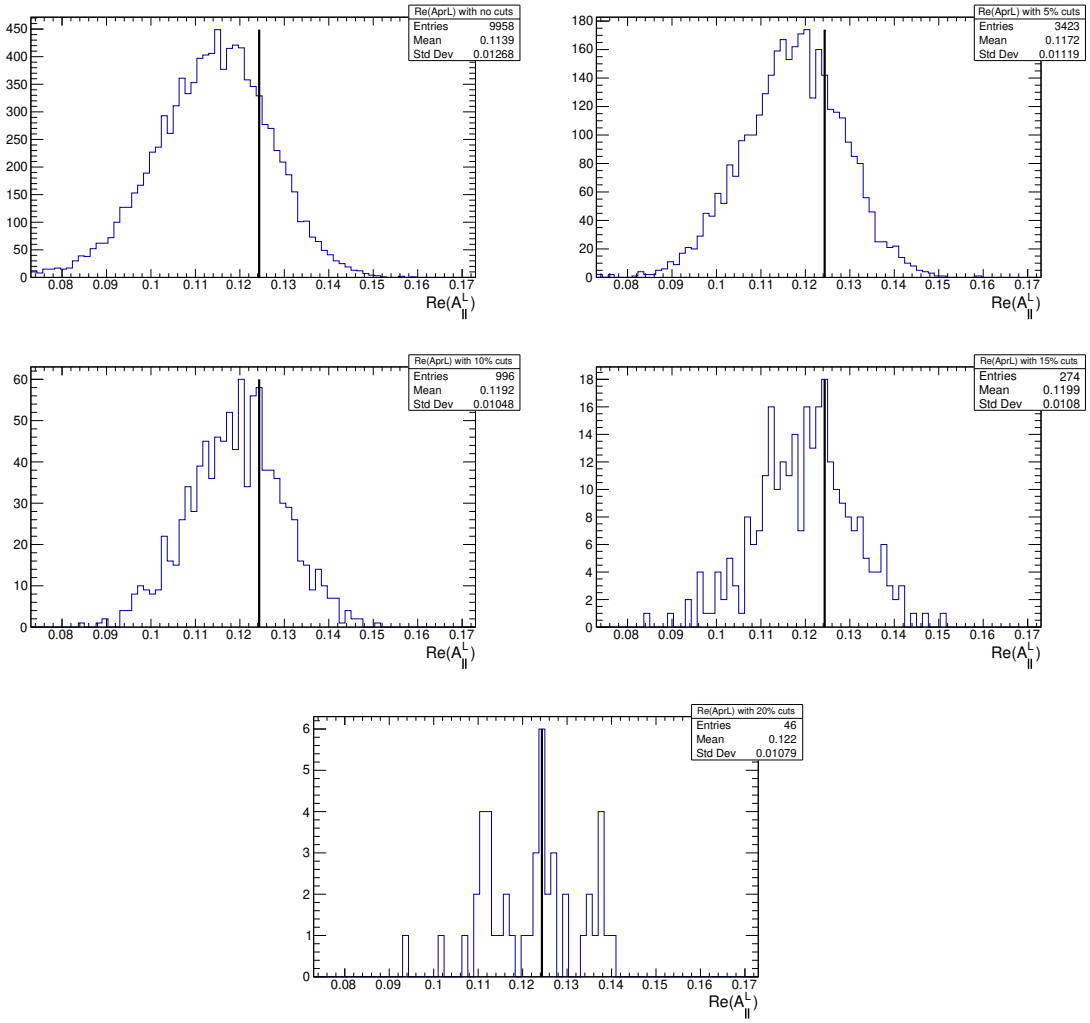


Figure 8.32: Distribution of $\mathrm{Re}(\mathcal{A}_{\parallel}^{L})$ at $q^2 = 6\ \mathrm{GeV}^2/c^4$ from an ensemble of fits where no, upper and lower 5%, 10%, 15%, and 20% of fits in the other amplitudes are removed. The true value is shown in black.

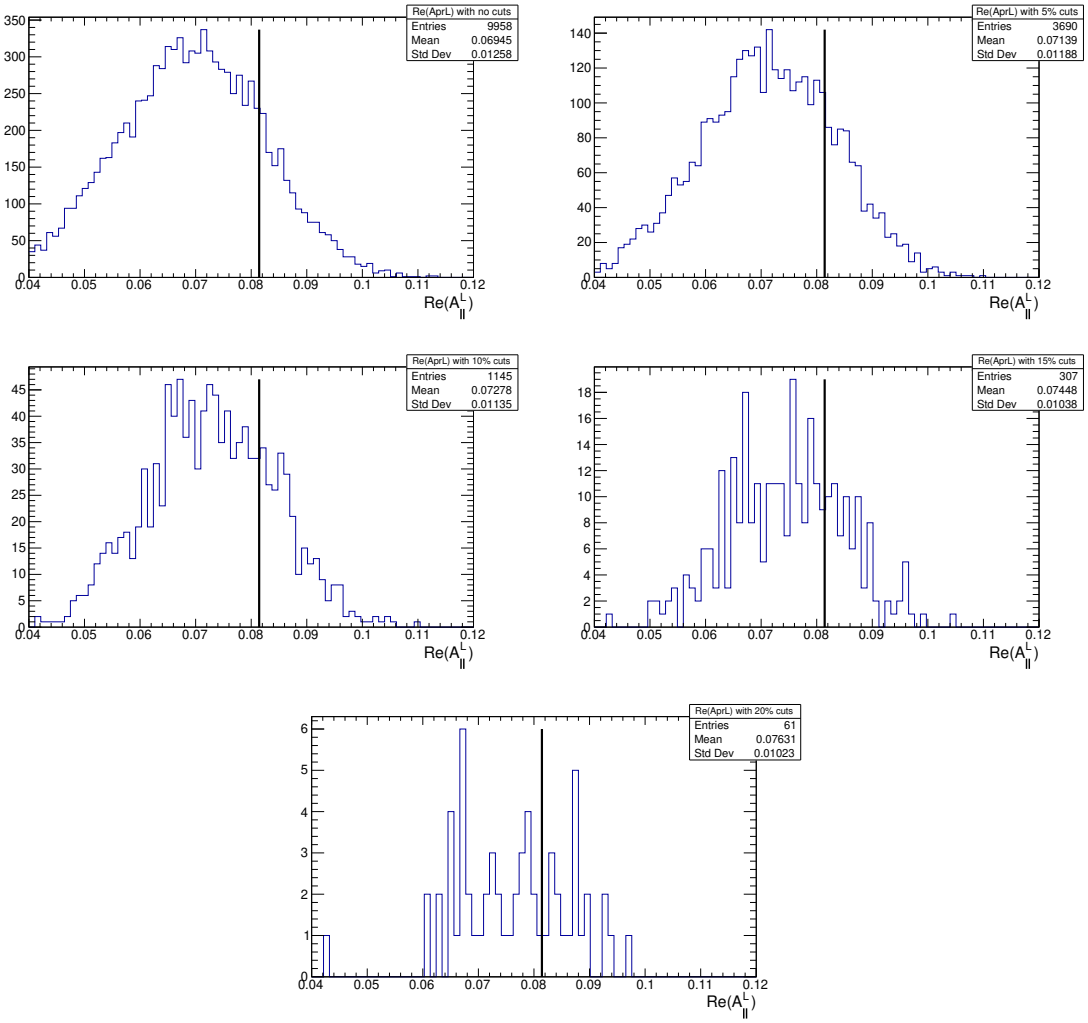


Figure 8.33: Distribution of $\mathrm{Re}(\mathcal{A}_{\parallel}^{L})$ at $q^2 = 4$ GeV$^2/c^4$ from an ensemble of fits where no, upper and lower 5%, 10%, 15%, and 20% of fits in the other amplitudes are removed. The true value is shown in black.

Chapter 9

Pseudoexperiments generated from the data fit

This section describes the pseudoexperiment studies in the region $1.25 < q^2 < 8.0$ GeV$^2/c^4$, where the pseudoexperiments are generated from the best fit point to data. All plots in this section are blinded. The default fit configuration in this region is the CP-symmetries fit, where the P-wave amplitudes have 4 parameters. Explicitly, the P-wave amplitudes $\mathcal{A}$ are described by the ansatz

$$\mathcal{A}(q^2) = \alpha L_0(q^{2\prime}) + \beta L_1(q^{2\prime}) + \gamma L_2(q^{2\prime}) + \delta L_3(q^{2\prime}), \tag{9.1}$$

where $L_i(q^{2\prime})$ are Legendre polynomials, and $q^{2\prime}$ corresponds to q^2 scaled to the range $-1 \leq x \leq 1$. The amplitude components are α, β, γ, and δ.

For the studies performed in this chapter, 1090 pseudoexperiments are generated from the fit result to data. These are fully representative of the data fit — the acceptance is included as well as the background and the $K\mu\mu$ veto. Example projections from a pseudoexperiment are shown as Figure 9.1.

Results from ensembles of fits were studied. The residuals of the parallel amplitude parameters are shown as Figures 9.2 and 9.3. The results for the left-handed transverse amplitude parameters are shown in Figure 9.4. Figure 9.5 shows the residuals for the right-handed amplitude parameters and the longitudinal amplitude parameters. All other P-wave amplitude parameters are fixed to zero. The black lines show the values from the

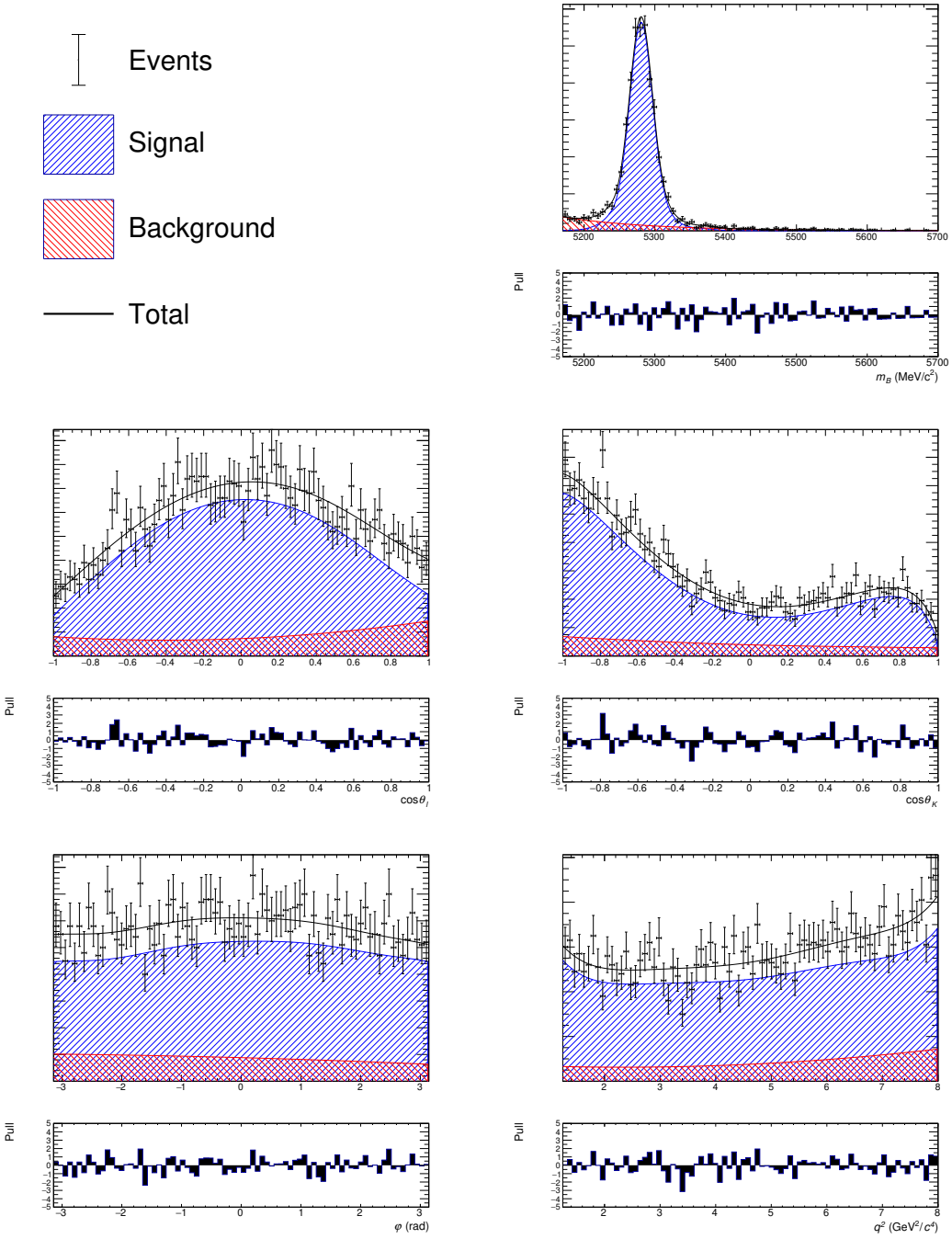


Figure 9.1: Projections from fitting a pseudoexperiment generated in the $1.25 < q^2 < 8$ GeV$^2/c^4$ region. The value of the parameters used for the generation are from the data fit. The y-axes have been blinded. Shown is the B^0 mass (top right), $\cos\theta_\ell$ (middle left), $\cos\theta_K$ (middle right), ϕ (bottom left), and q^2 (bottom right). The signal is shown in blue and the background is shown in red. As seen from the pulls, the fitter is reasonably behaved.

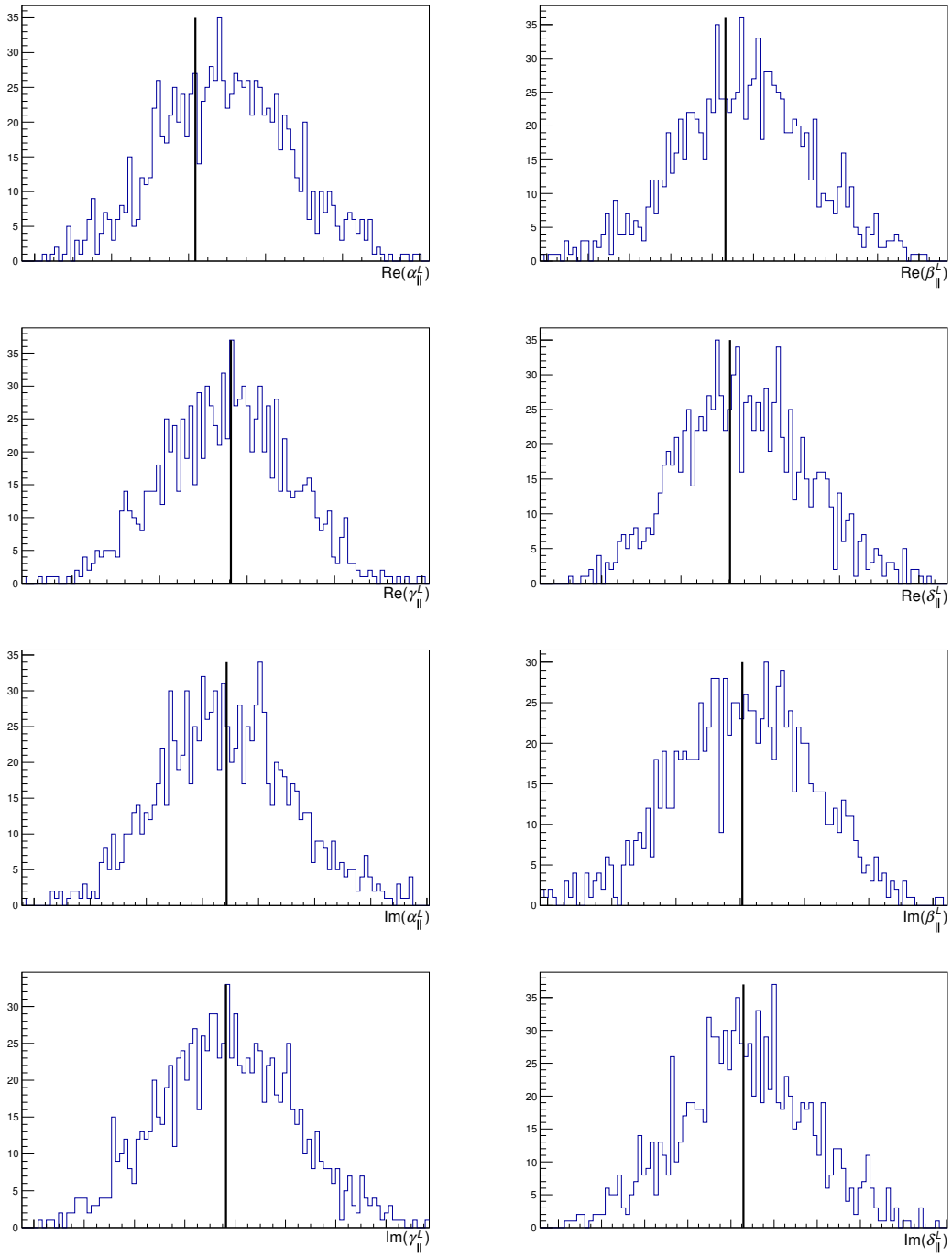


Figure 9.2: Residuals of the P-wave parameters (left-handed parallel amplitude components) from ensembles of fits in the $1.25 < q^2 < 8$ GeV$^2/c^4$ region. The value of the parameters used for the generation are from the data fit. The experimental values of the parameters are shown by the black lines. The x-axes have been blinded.

data fit used to generate the pseudoexperiments, i.e. the experimental values. As seen from these plots there are no large biases.

The distribution of fits of the amplitudes was also examined. The amplitude is computed

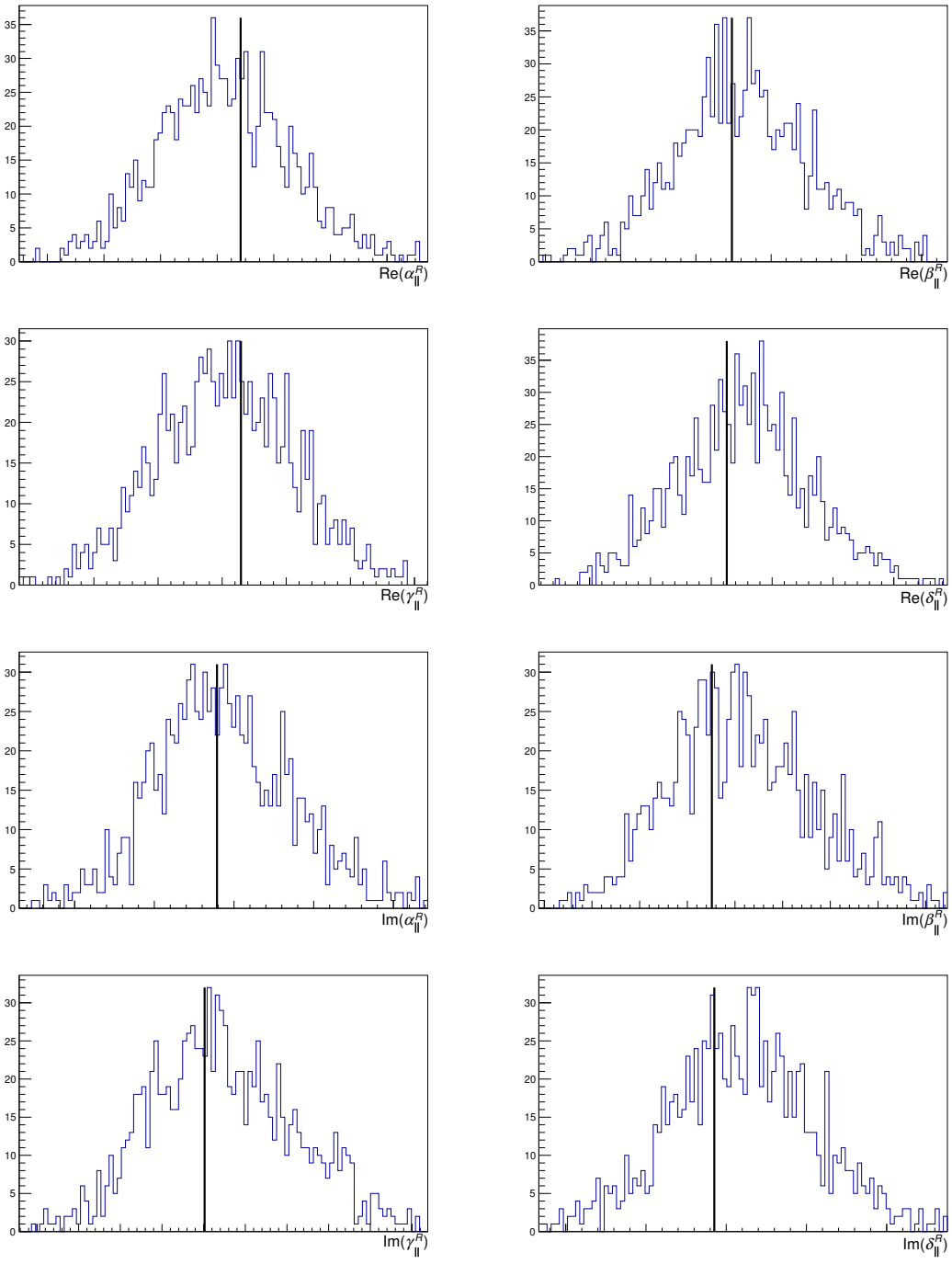


Figure 9.3: Residuals of the P-wave parameters (right-handed parallel amplitude components) from ensembles of fits in the $1.25 < q^2 < 8$ GeV$^2/c^4$ region. The value of the parameters used for the generation are from the data fit. The experimental values of the parameters are shown by the black lines. The x-axes have been blinded.

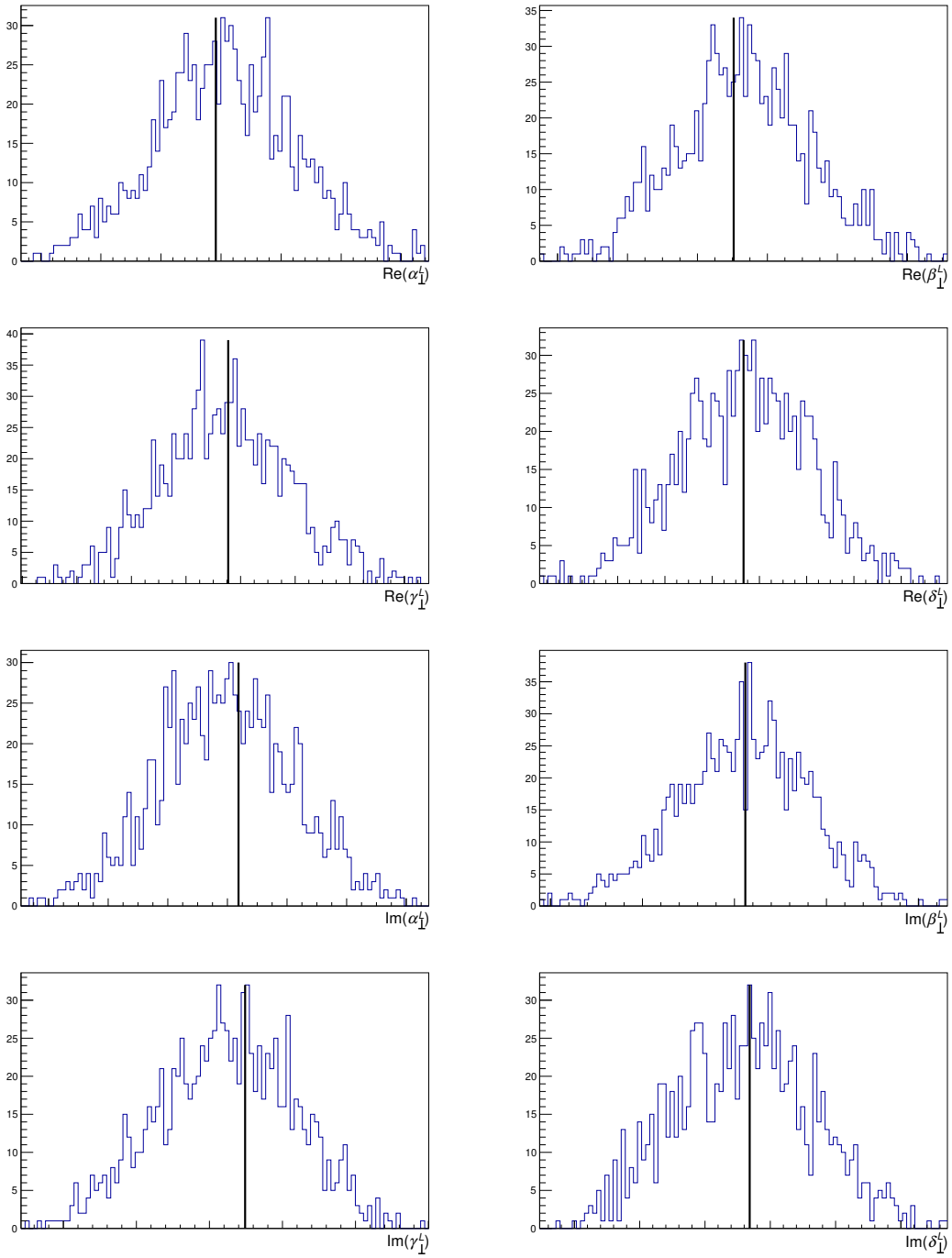


Figure 9.4: Residuals of the P-wave parameters (left-handed transverse amplitude components) from ensembles of fits in the $1.25 < q^2 < 8$ GeV$^2/c^4$ region. The value of the parameters used for the generation are from the data fit. The experimental values of the parameters are shown by the black lines. The x-axes have been blinded.

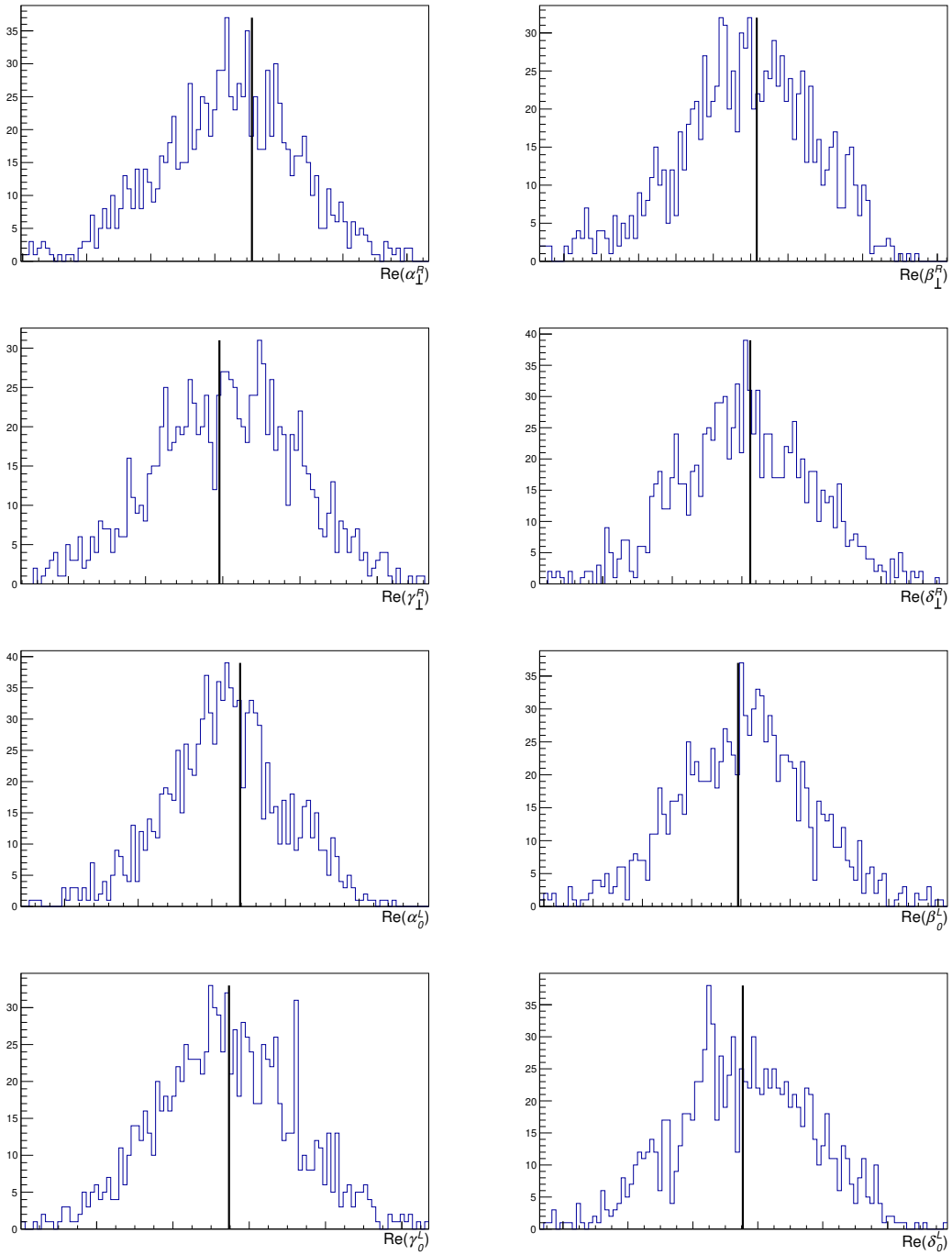


Figure 9.5: Residuals of the P-wave parameters (right-handed transverse and longitudinal amplitude components) from ensembles of fits in the $1.25 < q^2 < 8\,\mathrm{GeV}^2/c^4$ region. The value of the parameters used for the generation are from the data fit. The experimental values of the parameters are shown by the black lines. The x-axes have been blinded.

from the amplitude components. Here the amplitudes are shifted by the experimental value at each q^2 point for the blinding. Thus by definition the experimental value is at zero. These plots are shown as Figure 9.6 for the P-wave amplitudes. The experimental value is shown by the black line. There are no large biases seen here, thus the fit is performing well. However the largest apparent bias seen is in the amplitude $Re(\mathcal{A}_{\parallel}^{L})$, which is discussed in more detail in Chapter 8.3.

The coverage of the pseudoexperiments was also studied. The fraction of fits where experimental value is between the best fit point μ minus a shift x of the Hessian uncertainty σ_{HESSE} and μ plus a shift x of σ_{HESSE} of the fits were studied. This is compared to the absolute value of a cumulative distribution of a Normal distribution (i.e. a Gaussian with mean = 0 and $\sigma = 1$). These are shown as Figures 9.7, 9.8, 9.9, and 9.10. There are some components where incorrect coverage is seen, others where the coverage is correct. This is also compared to the fraction of fits where the experimental value is between $\mu - x\sigma_{new}$ and $\mu + x\sigma_{new}$ where σ_{new} corresponds to the uncertainties when fitting the profile log-likelihood from the pseudoexperiments with bifurcated parabolas. As shown in these figures, the uncertainties when fitting the profile log-likelihood from the pseudoexperiments with bifurcated parabolas overall gives correct coverage compared to the uncertainties from the Hessian error matrix, which sometimes results in incorrect coverage. Chapter 11 shows the profile log-likelihoods from the fits to data.

To conclude, pseudoexperients generated from the data best-fit point suggest the fit is well behaved and there are no large biases seen. From examining the coverage, the quoted uncertainties are found to be trustworthy. In most cases, the Hessian error matrix can be used. Where the Hessian cannot be used, the uncertainties are extracted by fitting the 1D log-likelihood profile to bifurcated parabolas. Given the fit behaves very well and the uncertainties can be trusted, this will enable one to generate synthetic datasets with accurate coverage from the fit results.

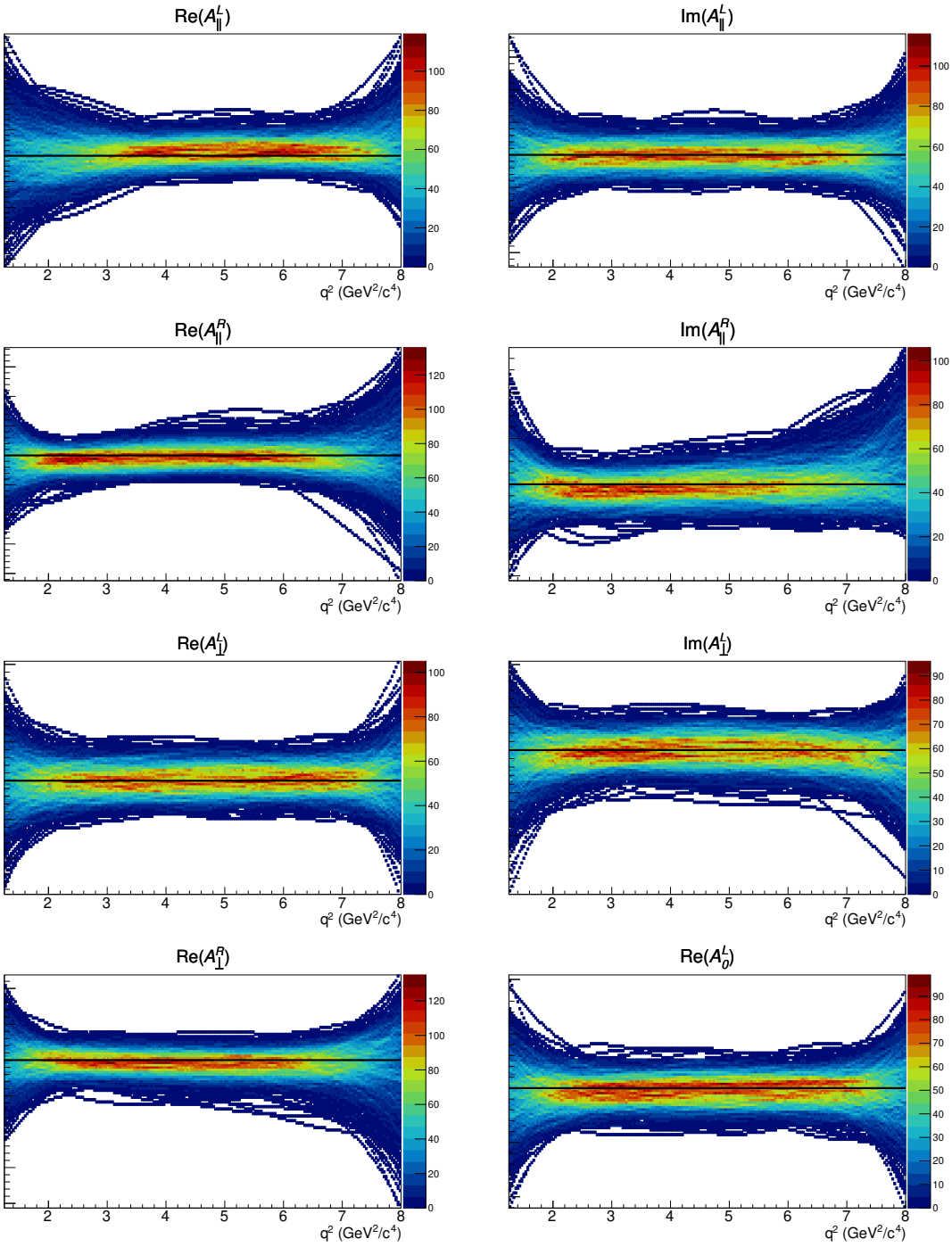


Figure 9.6: Distribution of fits of the P-wave amplitudes from ensembles of fits in the $1.25 < q^2 < 8$ GeV$^2/c^4$ region. The value of the parameters used for the generation are from the data fit. The experimental values of the parameters are shown by the black lines. All fits are shifted by the experimental value. The y-axes have been blinded.

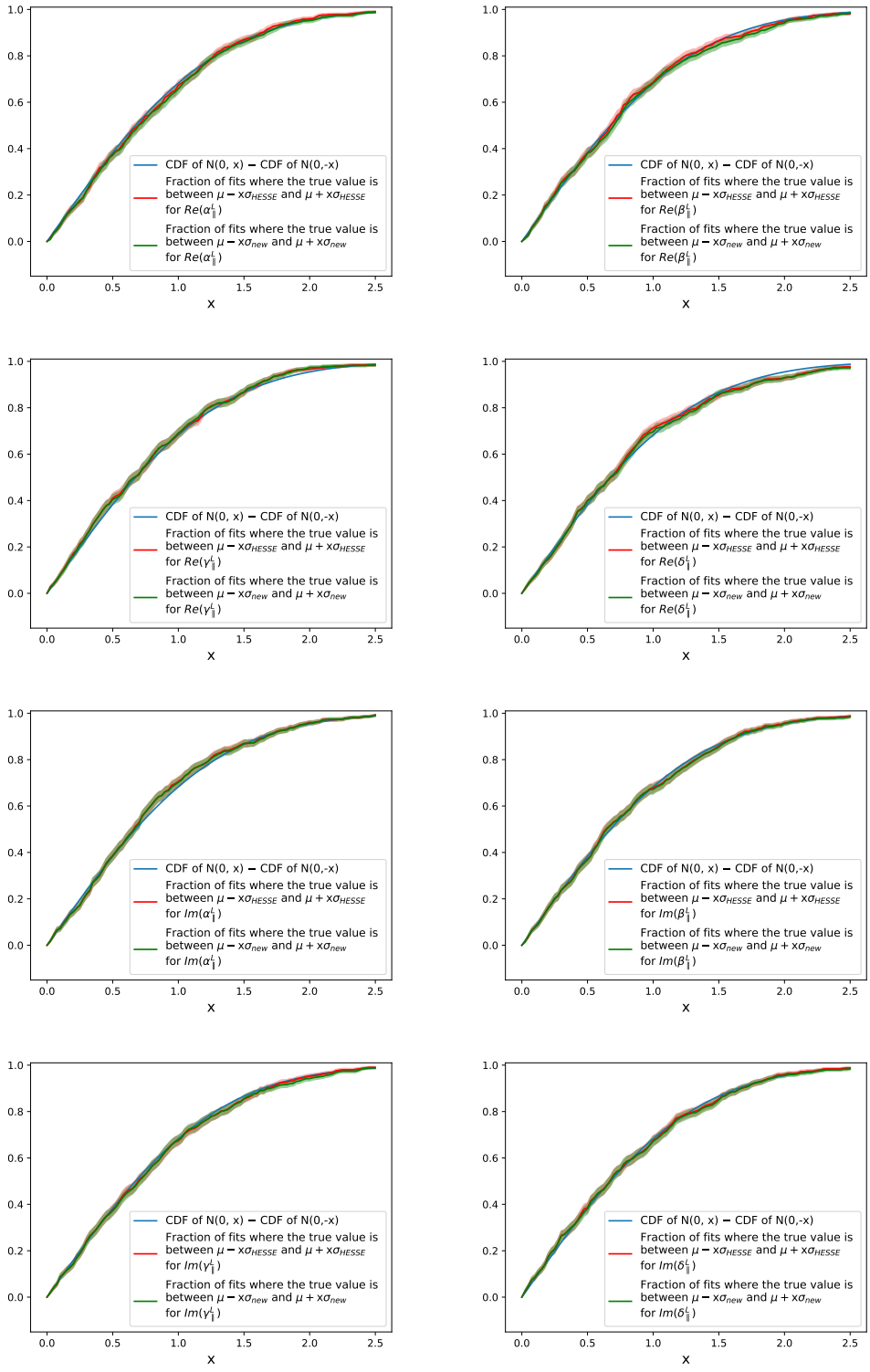


Figure 9.7: Coverage of the left-handed parallel amplitude components from ensembles of fits in the $1.25 < q^2 < 8$ GeV$^2/c^4$ region. The value of the parameters used for the generation are from the data fit. Shown are the fraction of fits where the experimental value is between $\mu - x\sigma_{HESSE}$ and $\mu + x\sigma_{HESSE}$ where μ is the best fit point σ_{HESSE} is the Hessian uncertainty, with error bands shown owing to the finite number of pseudoexperiments. Also shown is the same but for the uncertainties extracted from fitting bifurcated parabola to the profile log-likelihoods for each toy. These are compared the absolute value of the cumulative distribution of a Normal distribution.

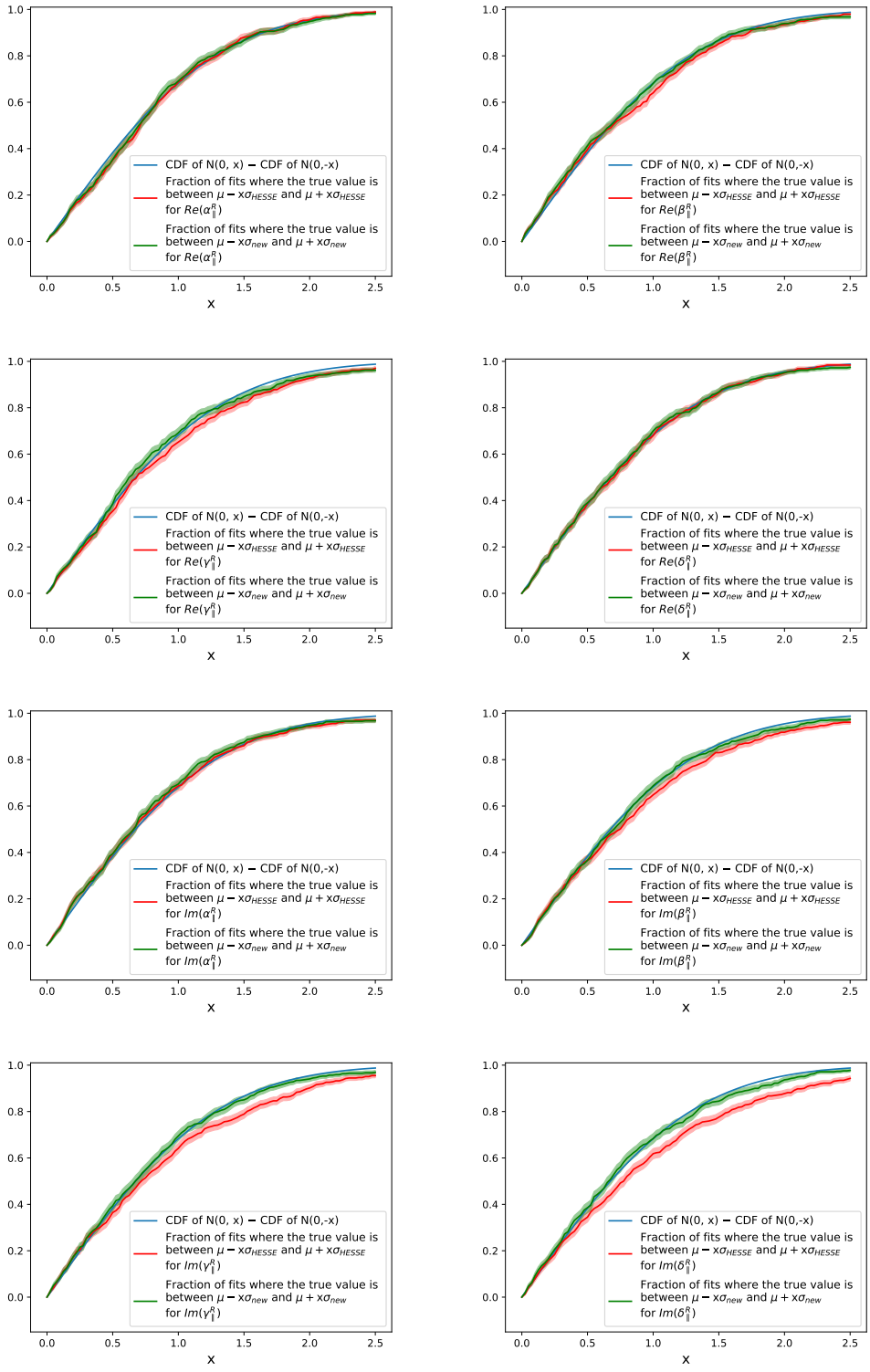


Figure 9.8: Coverage of the right-handed parallel amplitude components from ensembles of fits in the $1.25 < q^2 < 8$ GeV$^2/c^4$ region. The value of the parameters used for the generation are from the data fit. Shown are the fraction of fits where the experimental value is between $\mu - x\sigma_{HESSE}$ and $\mu + x\sigma_{HESSE}$ where μ is the best fit point σ_{HESSE} is the Hessian uncertainty, with error bands shown owing to the finite number of pseudoexperiments. Also shown is the same but for the uncertainties extracted from fitting bifurcated parabola to the profile log-likelihoods for each toy. These are compared the absolute value of the cumulative distribution of a Normal distribution.

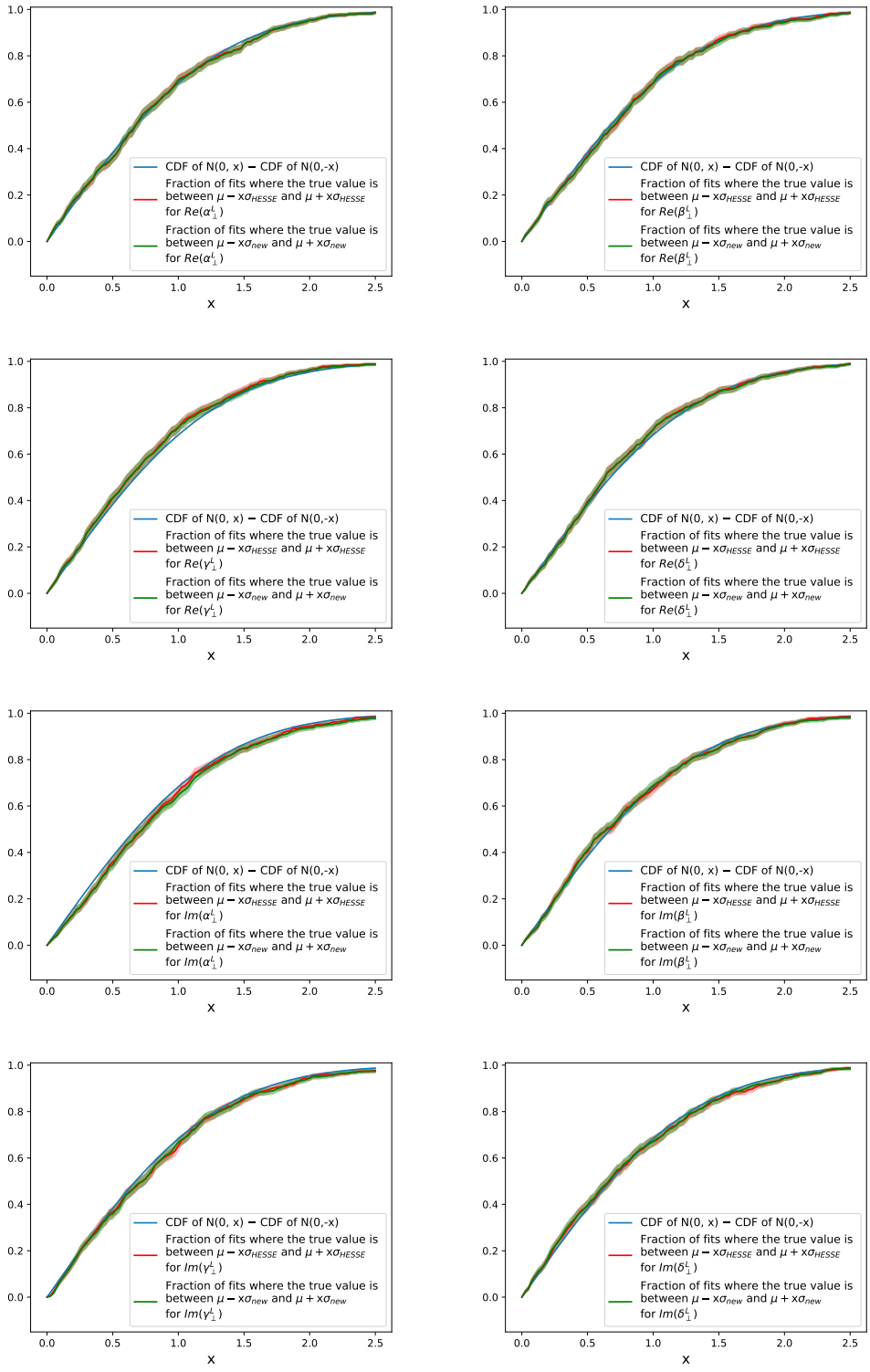


Figure 9.9: Coverage of the left-handed transverse amplitude components from ensembles of fits in the $1.25 < q^2 < 8$ GeV$^2/c^4$ region. The value of the parameters used for the generation are from the data fit. Shown are the fraction of fits where the experimental value is between $\mu - x\sigma_{HESSE}$ and $\mu + x\sigma_{HESSE}$ where μ is the best fit point σ_{HESSE} is the Hessian uncertainty, with error bands shown owing to the finite number of pseudoexperiments. Also shown is the same but for the uncertainties extracted from fitting bifurcated parabola to the profile log-likelihoods for each toy. These are compared the absolute value of the cumulative distribution of a Normal distribution.

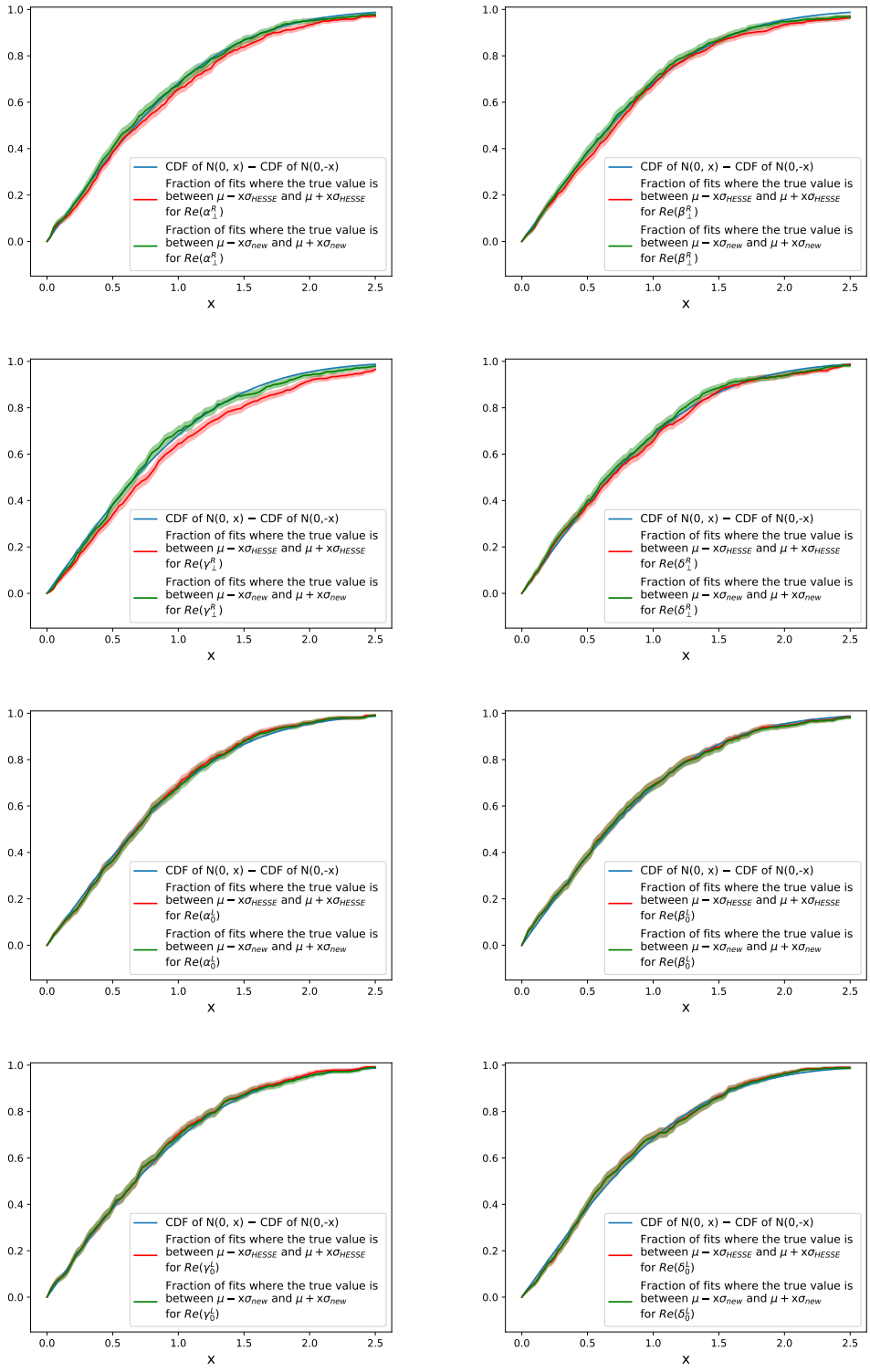


Figure 9.10: Coverage of the right-handed transverse and longitudinal amplitude components from ensembles of fits in the $1.25 < q^2 < 8$ GeV$^2/c^4$ region. The value of the parameters used for the generation are from the data fit. Shown are the fraction of fits where the experimental value is between $\mu - x\sigma_{HESSE}$ and $\mu + x\sigma_{HESSE}$ where μ is the best fit point σ_{HESSE} is the Hessian uncertainty, with error bands shown owing to the finite number of pseudoexperiments. Also shown is the same but for the uncertainties extracted from fitting bifurcated parabola to the profile log-likelihoods for each toy. These are compared the absolute value of the cumulative distribution of a Normal distribution.

Chapter 10

Results from the control mode fits

The control mode $B^0 \to K^{*0}J/\psi$ is fitted in order to verify agreement between data periods. The J/ψ fits shown in this chapter were performed in the q^2 region $9.223 < q^2 < 9.966$ GeV$^2/c^4$. The fits were performed with amplitudes flat in q^2, i.e. $\mathcal{A} = \alpha$ for each amplitude, where α is complex. The observables were computed by sampling from the amplitudes covariance matrix. The B_s^0 yield relative to the B^0 is fixed to 0.0077075 and mass difference between the B_s^0 and B^0 is fixed to 87.26 MeV/c^2. The tails of the signal Crystal Ball functions are free parameters. The acceptances with the same orders of polynomials used to describe the distributions in the angles as in the full Run 1 and Run 2 fits were used. The acceptance is evaluated at the midpoint of the q^2 window $9.223 < q^2 < 9.966$ GeV$^2/c^4$. There is no q^2 dependence in this fit since the amplitudes are very fast-moving and thus extremely difficult to describe with Legendre polynomial ansatzes. Thus amplitude fits binned in q^2 are performed.

10.1 CP-averages fit

The fits in the CP-average configuration were performed. For these fits all of the B^0 mass parameters are floated. The plots of the Run 1 fit are shown as Figure 10.1. The 2016, 2017, and 2018 fits are shown as Figures 10.2 and 10.3, and 10.4 respectively. The large pulls in $\cos\theta_K$ are due to exotic states decaying to $\pi J/\psi$, e.g. $B^0 \to Z(4430)^-(\to \pi^- J/\psi)K^+$, where $Z(4430)^-$ is a $c\bar{c}d\bar{u}$ state. These are consistent with pulls seen in pseudoexperiments where the exotic states are added in the pseudoexperiments based on [90].

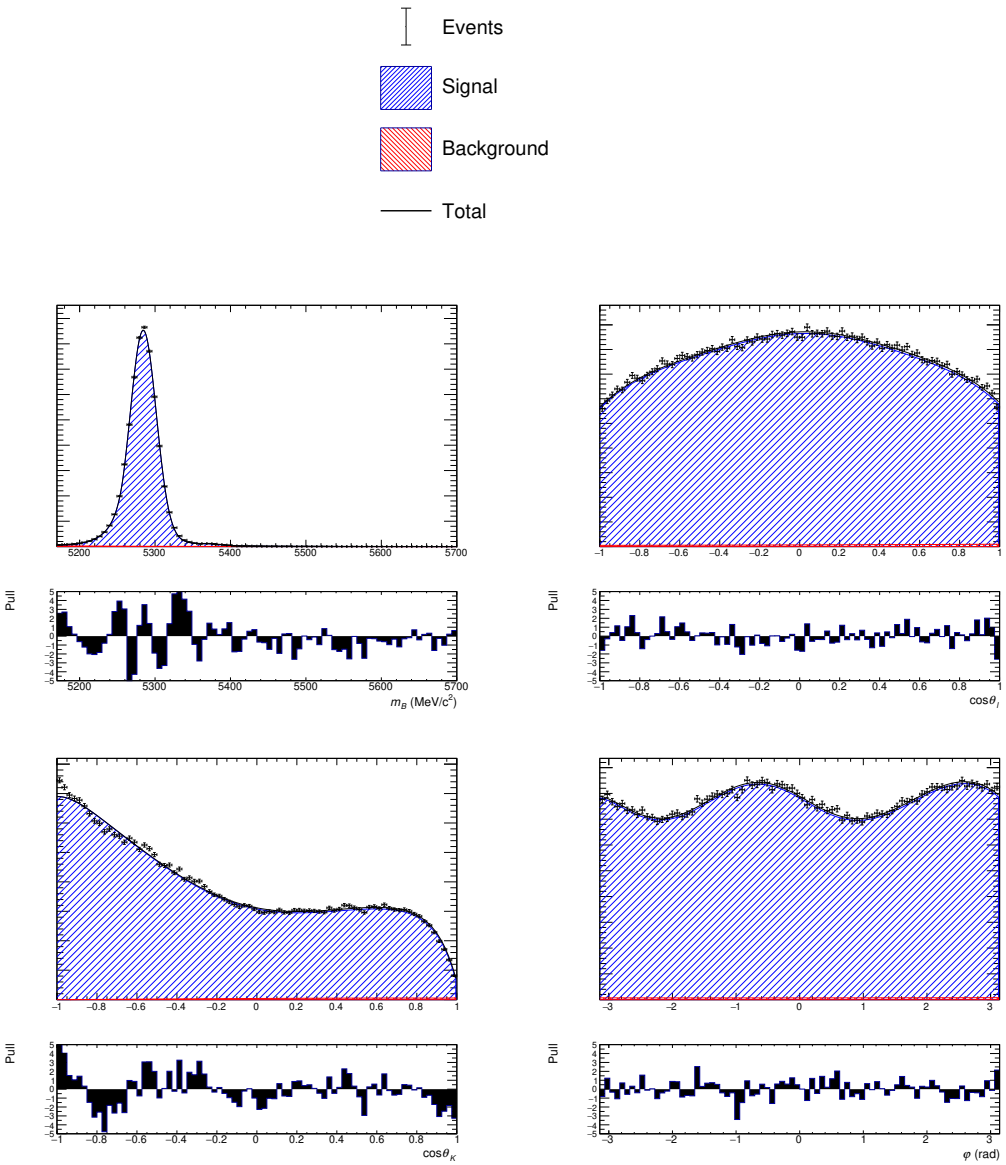


Figure 10.1: Plots of the $B^0 \to K^{*0} J/\psi$ CP-averages fit to Run 1 data. The signal is shown in blue and the background is shown in red. Shown are the B^0 mass, $\cos\theta_\ell$, $\cos\theta_K$, and ϕ.

The observables in each data period were computed. These are shown as Table 10.1 for Run 1 and 2016, and Table 10.2 for 2017 and 2018. Note the uncertainties shown are just statistical. The comparisons between each data period were made and are shown as Table 10.3. Since these comparisons were made with just the statistical uncertainties and J/ψ mode is expected to be systematically limited, the agreement between the data periods is good.

Table 10.1: Results for the obsvervables from the Run 1 and 2016 $B^0 \to K^{*0} J/\psi$ fits, where the fits are performed in the CP-averages configuration.

	Run 1	2016
FL	0.55865 ± 0.00117	0.56369 ± 0.00141
S3	-0.00581 ± 0.00115	-0.00341 ± 0.00166
S4	-0.24556 ± 0.00065	-0.24438 ± 0.00101
S5	-0.0023 ± 0.0014	0.00176 ± 0.00204
AFB	-0.00142 ± 0.00113	-0.00025 ± 0.00125
S7	-0.0022 ± 0.00181	-0.00276 ± 0.00212
S8	-0.05103 ± 0.00137	-0.05636 ± 0.002
S9	-0.09147 ± 0.00148	-0.09308 ± 0.00175
FS	0.0834 ± 0.00157	0.08166 ± 0.00245

Table 10.2: Results for the obsvervables from the 2017 and 2018 $B^0 \to K^{*0} J/\psi$ fits, where the fits are performed in the CP-averages configuration.

	2017	2018
FL	0.55975 ± 0.00141	0.56131 ± 0.00128
S3	-0.00555 ± 0.00163	-0.00402 ± 0.00149
S4	-0.24605 ± 0.00099	-0.24453 ± 0.0009
S5	-0.00013 ± 0.002	0.00465 ± 0.00179
AFB	-0.00142 ± 0.00123	-0.00275 ± 0.00109
S7	0.00252 ± 0.00207	-0.0019 ± 0.00187
S8	-0.05497 ± 0.00202	-0.05317 ± 0.00181
S9	-0.09185 ± 0.00176	-0.09223 ± 0.00163
FS	0.08847 ± 0.00251	0.08198 ± 0.00225

Table 10.3: Comparisons of obsvervables (in numbers of σ) for the $B^0 \to K^{*0} J/\psi$ fits between each data period, where the fits are performed in the CP-averages configuration.

	Run 1 - 16	Run 1 - 17	Run 1 - 18	16 - 17	16 - 18	17 - 18
FL	-2.75	-0.6	-1.53	1.98	1.25	-0.82
S3	-1.19	-0.13	-0.95	0.92	0.28	-0.69
S4	-0.98	0.41	-0.93	1.18	0.11	-1.14
S5	-1.63	-0.89	-3.06	0.66	-1.06	-1.78
AFB	-0.69	0.0	0.85	0.67	1.51	0.81
S7	0.2	-1.71	-0.12	-1.78	-0.31	1.58
S8	2.2	1.62	0.94	-0.49	-1.18	-0.66
S9	0.7	0.17	0.35	-0.49	-0.35	0.16
FS	0.6	-1.71	0.52	-1.94	-0.1	1.93

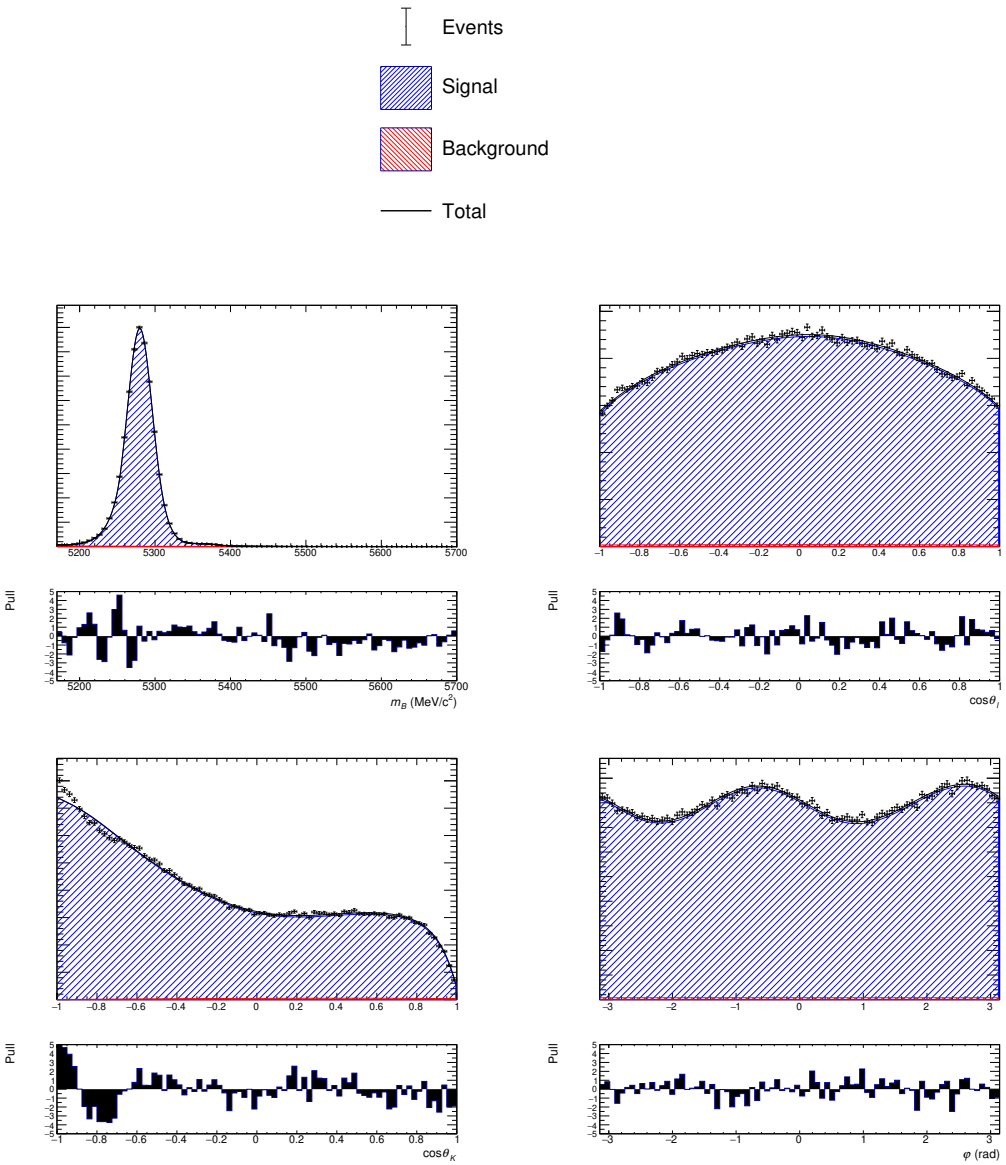


Figure 10.2: Plots of the $B^0 \to K^{*0} J/\psi$ CP-averages fit to 2016 data. The signal is shown in blue and the background is shown in red. Shown are the B^0 mass, $\cos\theta_\ell$, $\cos\theta_K$, and ϕ.

10.2 CP-asymmetries fit

The fits in the asymmetries configuration were performed, with identical extended terms, i.e. assuming no production or detection asymmetry. The production and detection asymmetries here are therefore absorbed in the amplitudes. From the measured J/ψ yields, the detection and production asymmetries are computed in order to modify the extended term for the rare mode CP-asymmetries fit. The B^0 mass tail parameters are fixed to those measured in $B^0 \to K^{*0} J/\psi$ simulation.

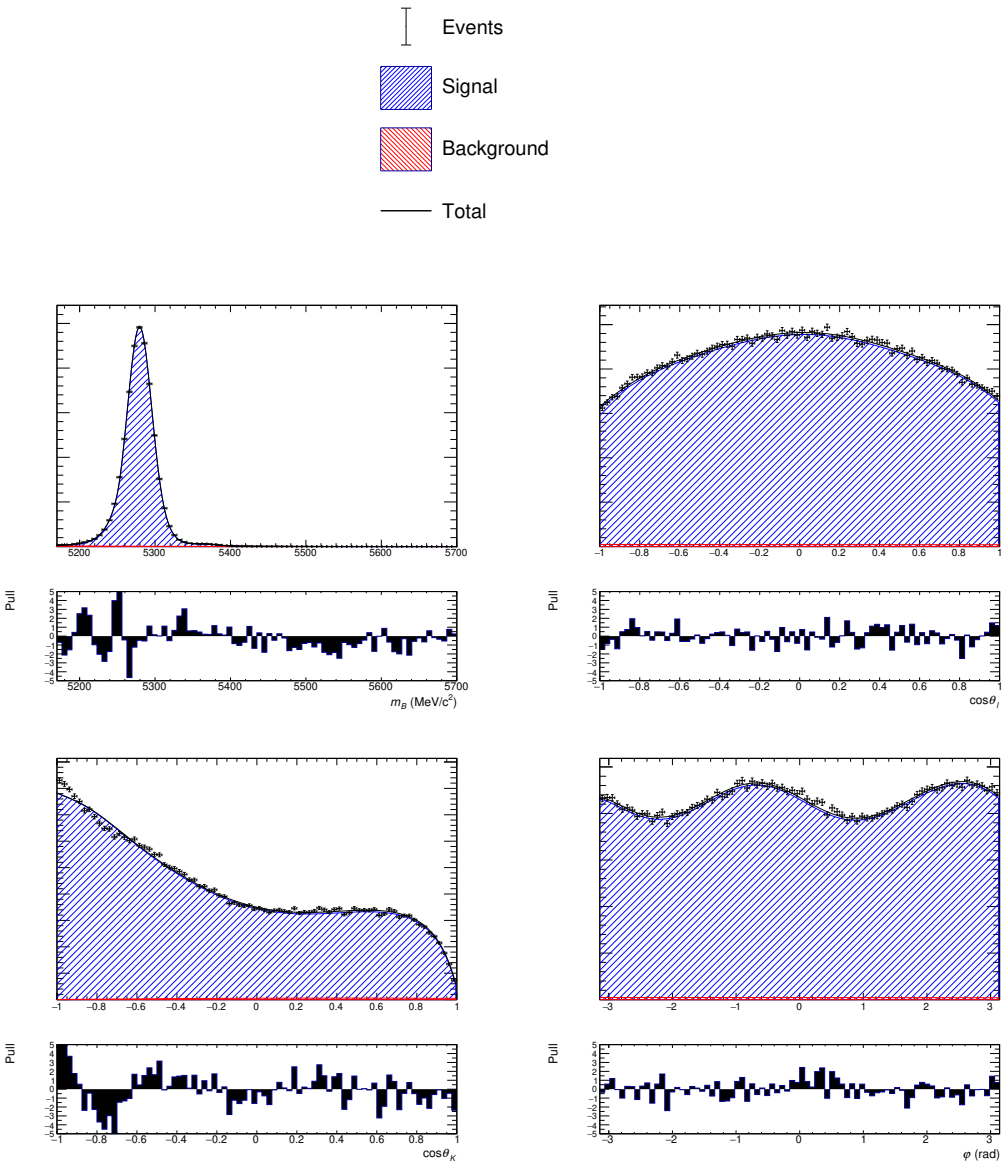


Figure 10.3: Plots of the $B^0 \to K^{*0} J/\psi$ CP-averages fit to 2017 data. The signal is shown in blue and the background is shown in red. Shown are the B^0 mass, $\cos\theta_\ell$, $\cos\theta_K$, and ϕ.

The observables in each data period were computed. These are shown as Table 10.4 for Run 1 and 2016, and Table 10.5 for 2017 and 2018. The comparisons between each data period are shown as Table 10.6. There are no observables more than 3σ apart and thus the agreement between data periods is very good since only the statistical uncertainties are used in this comparison.

Table 10.4: Results for the obsvervables from the Run 1 and 2016 $B^0 \to K^{*0} J/\psi$ fits, where the fits are performed in the CP-asymmetries configuration.

	Run 1	2016
FL	0.55895 ± 0.00153	0.56449 ± 0.00144
S3	-0.00679 ± 0.00176	-0.00378 ± 0.00166
S4	-0.24618 ± 0.00107	-0.24455 ± 0.00103
S5	-0.00236 ± 0.00208	0.00224 ± 0.00207
AFB	-0.0001 ± 0.00131	-0.00037 ± 0.00125
S7	-0.00271 ± 0.00218	-0.00275 ± 0.0021
S8	-0.05166 ± 0.00206	-0.05673 ± 0.00202
S9	-0.09153 ± 0.00189	-0.09307 ± 0.00177
FS	0.0811 ± 0.00254	0.08154 ± 0.00251
AFL	-0.0009 ± 0.00109	0.00247 ± 0.00104
A3	-0.00067 ± 0.00132	-0.00016 ± 0.00122
A4	-0.00038 ± 0.00081	0.00039 ± 0.00075
A5	0.0003 ± 0.0016	-0.00092 ± 0.00151
A6S	0.00231 ± 0.0013	-0.00046 ± 0.00122
A7	0.0019 ± 0.00165	-0.00289 ± 0.00155
A8	-0.00017 ± 0.00411	0.00132 ± 0.00406
A9	-0.00128 ± 0.0037	-0.00102 ± 0.00357
AFS	-0.00839 ± 0.00596	-0.01295 ± 0.00588

Table 10.5: Results for the obsvervables from the 2017 and 2018 $B^0 \to K^{*0} J/\psi$ fits, where the fits are performed in the CP-asymmetries configuration.

	2017	2018
FL	0.56038 ± 0.0014	0.56262 ± 0.00125
S3	-0.0062 ± 0.00164	-0.00475 ± 0.00144
S4	-0.24638 ± 0.00098	-0.24496 ± 0.0009
S5	-0.00051 ± 0.002	0.00418 ± 0.00179
AFB	-0.00125 ± 0.00124	-0.00271 ± 0.0011
S7	0.00263 ± 0.0021	-0.00191 ± 0.00183
S8	-0.05531 ± 0.00198	-0.0537 ± 0.00176
S9	-0.09197 ± 0.00173	-0.09201 ± 0.00153
FS	0.08785 ± 0.00249	0.08125 ± 0.00219
AFL	-0.00115 ± 0.00094	0.00074 ± 0.00068
A3	0.00072 ± 0.00111	-0.00083 ± 0.00082
A4	0.00024 ± 0.00067	-0.00029 ± 0.0005
A5	0.0007 ± 0.00134	0.0009 ± 0.00099
A6S	-0.00146 ± 0.00108	-0.00041 ± 0.0008
A7	-0.0012 ± 0.00139	-0.00018 ± 0.00103
A8	-0.00018 ± 0.00385	0.00154 ± 0.00362
A9	0.00031 ± 0.00348	-0.00016 ± 0.00313
AFS	-0.00291 ± 0.00599	-0.00155 ± 0.00511

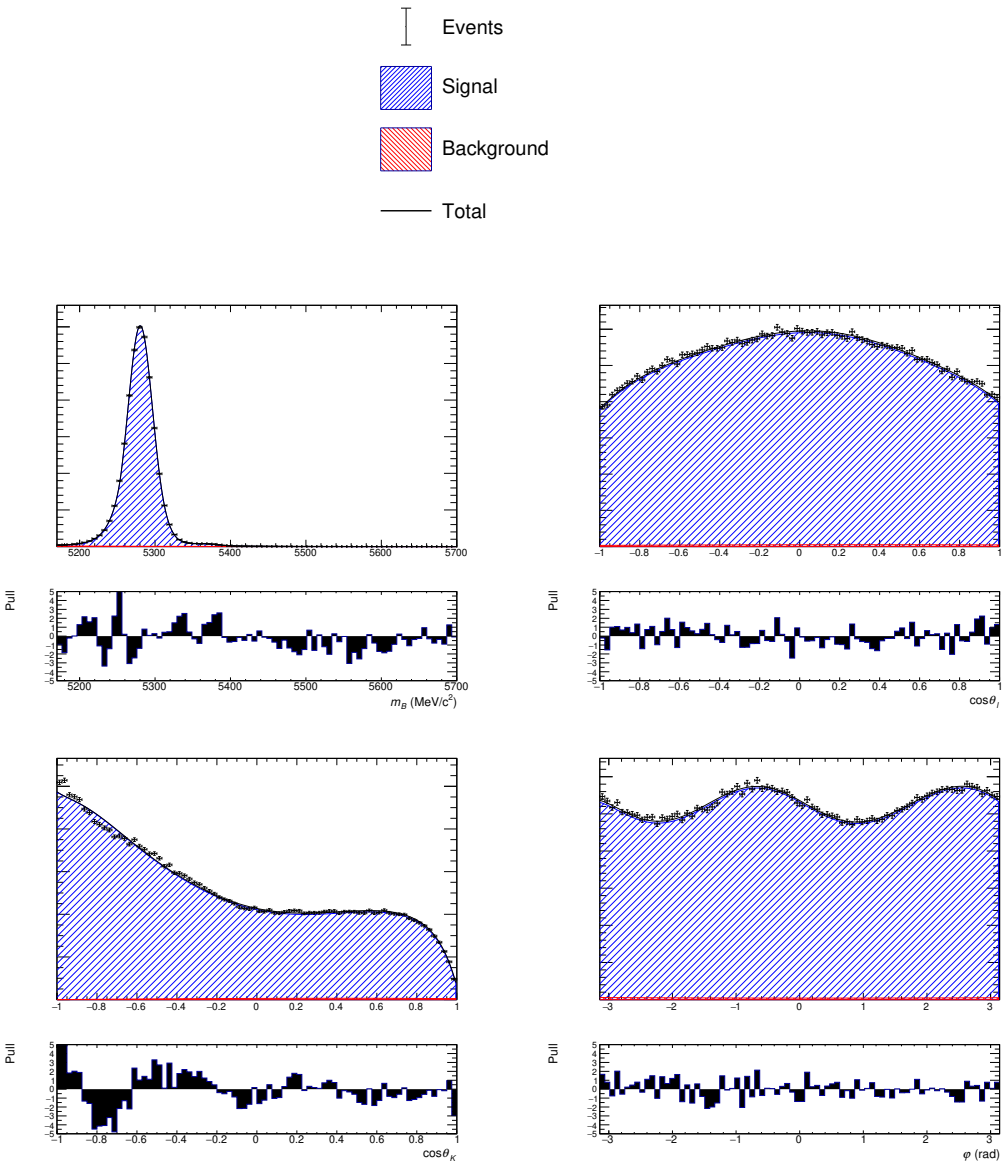


Figure 10.4: Plots of the $B^0 \to K^{*0} J/\psi$ CP-averages fit to 2018 data. The signal is shown in blue and the background is shown in red. Shown are the B^0 mass, $\cos\theta_\ell$, $\cos\theta_K$, and ϕ.

10.3 Using $B^0 \to K^{*0} J/\psi$ to verify the acceptance

The quantity q^2 is a derived quantity from quantities which the acceptance highly depends on, such as the opening angle and magnitudes of lepton momenta. It is possible this is not modelled well in the simulation used to derive the acceptance. The acceptance simulation cannot be compared directly to data since the acceptance simulation uses a model which is flat in all of the angles and q^2 and of course the data is model-dependent. Thus in order to verify the acceptance description, notably that applying cuts to the

Table 10.6: Comparisons of obsvervables (in numbers of σ) for the $B^0 \to K^{*0}J/\psi$ fits between each data period, where the fits are performed in the CP-asymmetries configuration.

	Run 1 - 2016	Run 1 - 2017	Run 1 - 2018	2016 - 2017	2016 - 2018	2017 - 2018
FL	−2.64	−0.69	−1.86	2.04	0.98	−1.19
S3	−1.24	−0.24	−0.9	1.04	0.44	−0.66
S4	−1.1	0.14	−0.87	1.3	0.3	−1.07
S5	−1.57	−0.64	−2.38	0.96	−0.71	−1.75
AFB	0.15	0.64	1.53	0.5	1.4	0.88
S7	0.01	−1.76	−0.28	−1.81	−0.3	1.63
S8	1.76	1.28	0.75	−0.5	−1.13	−0.61
S9	0.6	0.17	0.2	−0.45	−0.45	0.02
FS	−0.12	−1.9	−0.04	−1.78	0.09	1.99
AFL	−2.24	0.17	−1.28	2.58	1.4	−1.63
A3	−0.29	−0.81	0.1	−0.53	0.46	1.12
A4	−0.7	−0.59	−0.1	0.15	0.75	0.63
A5	0.55	−0.19	−0.32	−0.8	−1.01	−0.12
A6S	1.55	2.23	1.78	0.61	−0.04	−0.78
A7	2.12	1.44	1.07	−0.81	−1.46	−0.59
A8	−0.26	0.0	−0.31	0.27	−0.04	−0.33
A9	−0.05	−0.31	−0.23	−0.27	−0.18	0.1
AFS	0.54	−0.65	−0.87	−1.2	−1.46	−0.17

muon momenta and muon opening angle in the lab frame is accurately described by the acceptance, $B^0 \to K^{*0}J/\psi$ fits are performed with various selections to the muon momenta and opening angles. This is the first time this check has been performed in the decay $B^0 \to K^{*0}\mu^+\mu^-$. Figure 10.5 shows the angle between the two muons in the lab frame (α) against $\log_{10}$ of the maximum muon momentum. This is shown for 2016 simulation in the J/ψ region (blue) and the rare mode region (red).

This region is divided into 16 bins such that each bin has roughly the same number of events from simulation. The Run 1 and Run 2 bin boundaries are slightly different owing to slightly different muon momentum distributions. For each data period, $B^0 \to K^{*0}J/\psi$ fits were performed in each bin as defined by Figure 10.5 and compared to the fits in the full muon momenta region. The acceptance is recomputed in each bin. For Run 1, the pulls ((fit result in bin i - fit result in the full momenta region)/uncertainty from the bin i fit) are shown in Figures 10.6 and 10.7 for the CP-averaged observables and CP-asymmetries respectively.

For 2016, these are shown in Figures 10.8 and 10.9 for the CP-averages and CP-asymmetries respectively. Likewise, for 2017 and 2018, they are shown in Figures 10.10

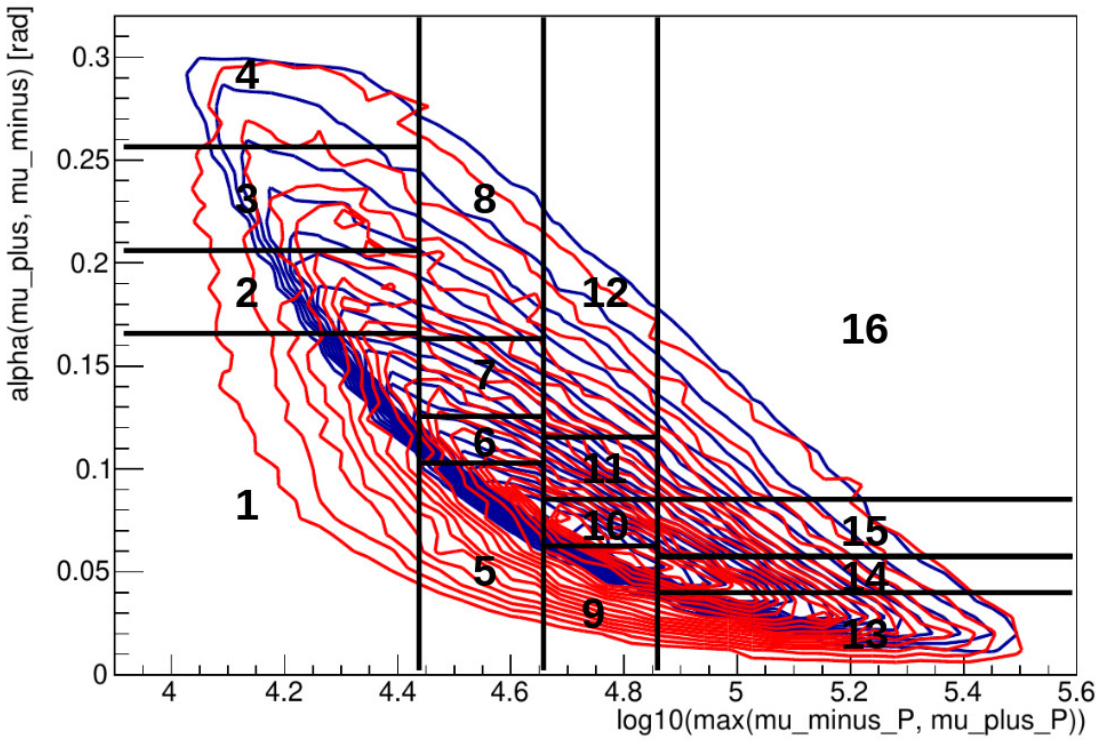


Figure 10.5: Angle between the two muons in the lab frame (α) against $\log_{10}$ of the maximum muon momentum. This is shown for 2016 simulation in the J/ψ region (blue) and the rare mode region (red). Each contour line corresponds to a set of points which are in regions with the same density.

and 10.11, and Figures 10.12 and 10.13.

From looking at these plots, it appears the observables in each bin are consistent with the combined fit. Thus selecting a region of α, $\log_{10}(max(p_\mu^+, p_\mu^-))$ results in an acceptance function which can describe this selection. In order to quantify the goodness-of-fit for each bin, chi-squared tests were performed assuming the uncertainty of the combined fit is exactly zero. Each bin has 18 observables (nine CP-averaged observables and nine CP-asymmetries) thus the number of degrees of freedom is 17. The p-values are shown in Table 10.7. As seen in this table, most p-values are above 5%. There is a presence of exotic contributions which are unaccounted for which explains why more bins than expected have p-values less than 5%. It is thus concluded that the acceptance is a reasonable description of the true angular efficiency as seen in data.

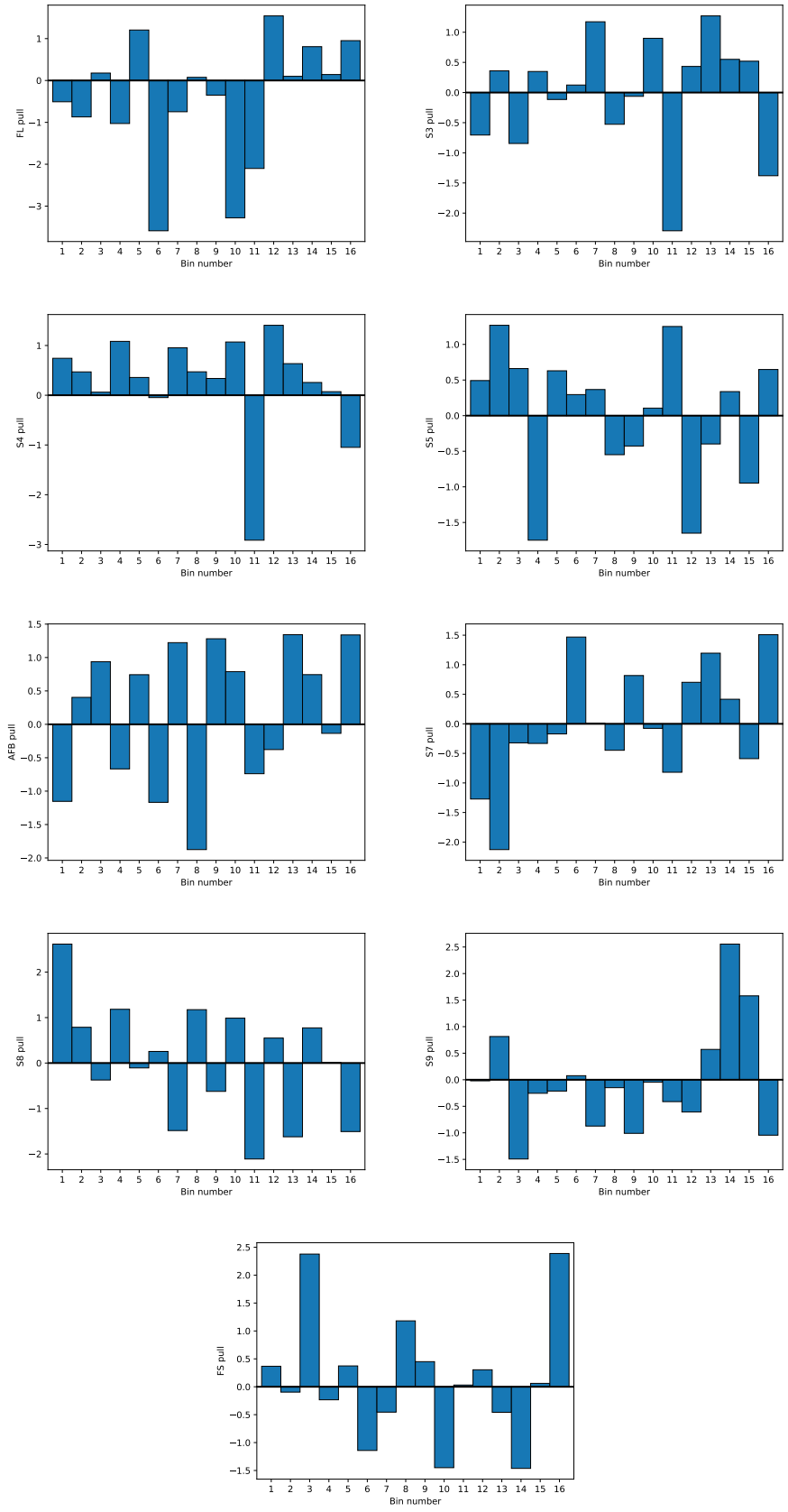


Figure 10.6: Pulls from comparing the CP-averaged observables from the Run 1 $B^0 \to K^{*0} J/\psi$ fit and the fits in the bins defined in Figure 10.5. For each observable the pull is defined as the difference between the fit result in bin i and the fit result in the full momenta region, divided by the uncertainty from the bin i fit.

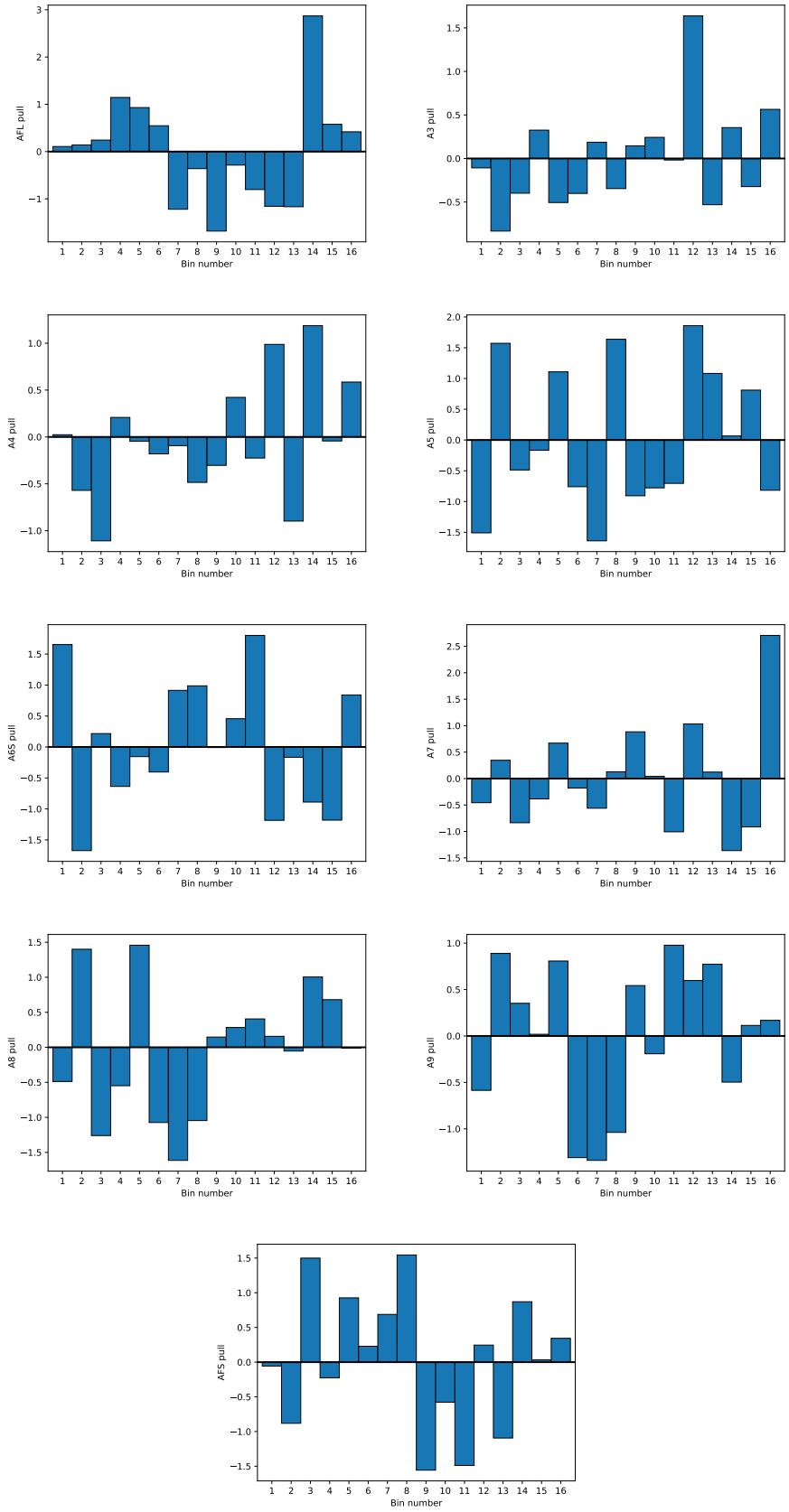


Figure 10.7: Pulls from comparing the CP-asymmetries from the Run 1 $B^0 \to K^{*0}J/\psi$ fit and the fits in the bins defined in Figure 10.5. For each observable the pull is defined as the difference between the fit result in bin i and the fit result in the full momenta region, divided by the uncertainty from the bin i fit.

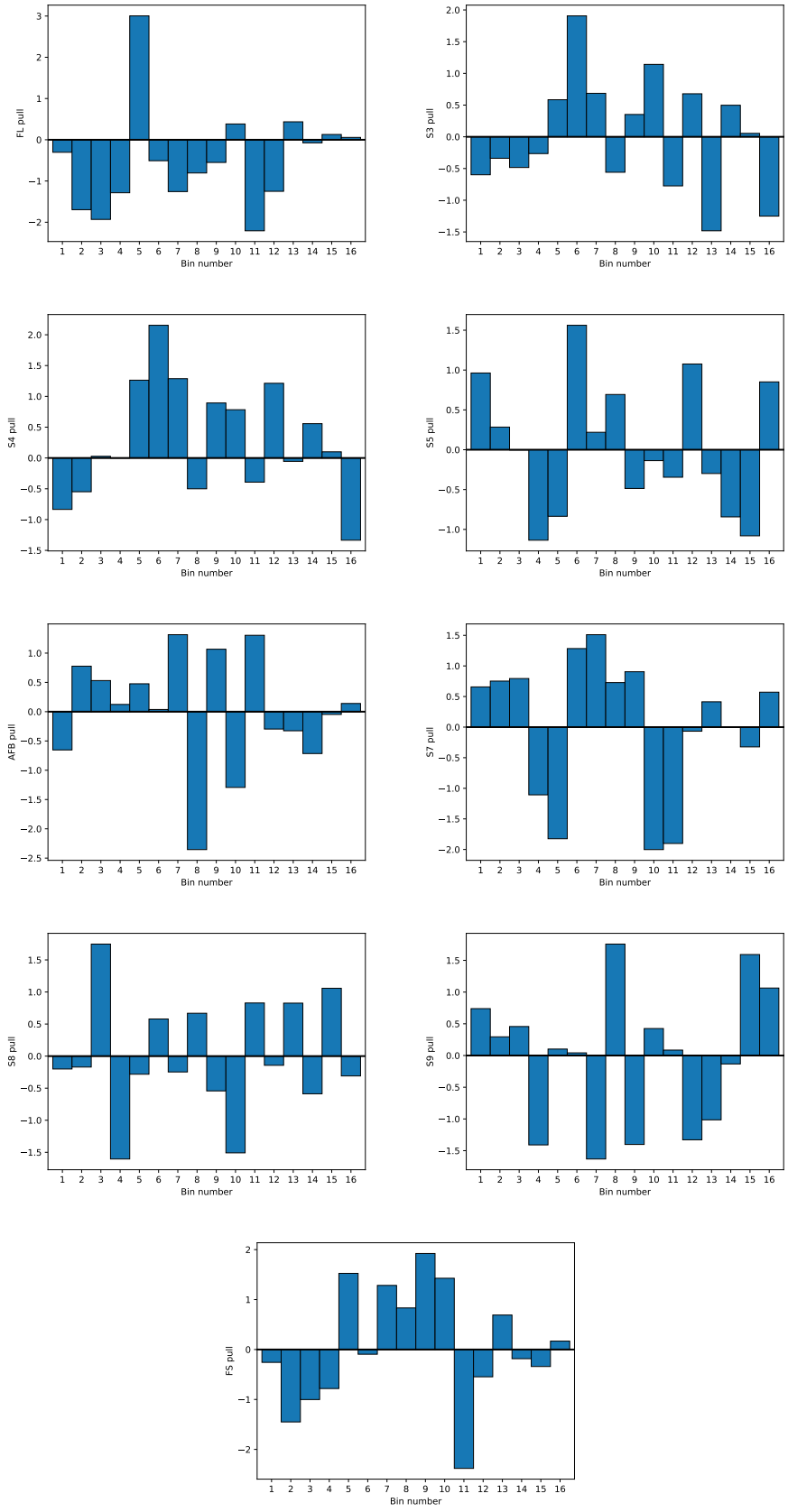


Figure 10.8: Pulls from comparing the CP-averaged observables from the 2016 $B^0 \to K^{*0}J/\psi$ fit and the fits in the bins defined in Figure 10.5. For each observable the pull is defined as the difference between the fit result in bin i and the fit result in the full momenta region, divided by the uncertainty from the bin i fit.

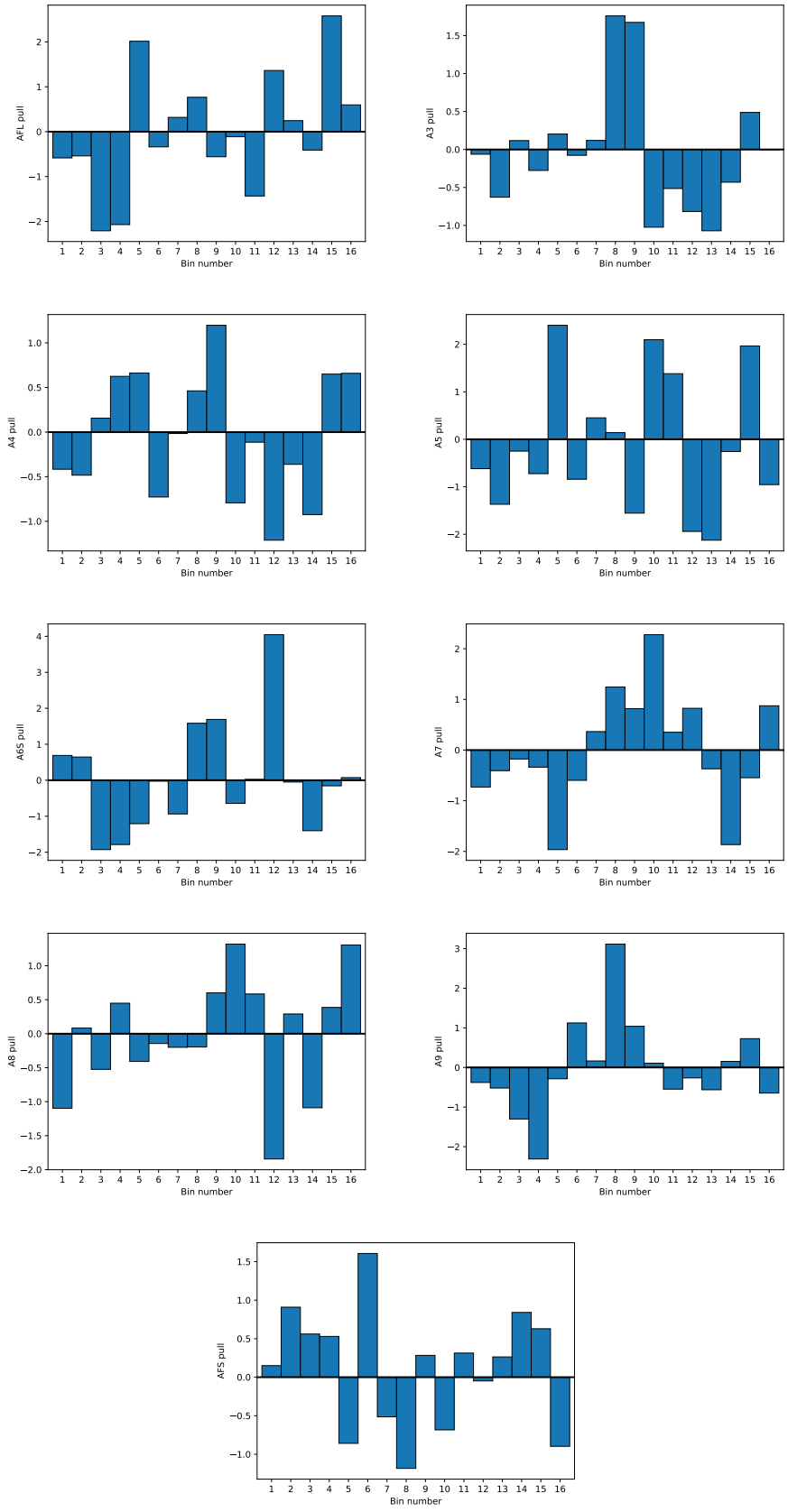


Figure 10.9: Pulls from comparing the CP-asymmetries from the 2016 $B^0 \to K^{*0} J/\psi$ fit and the fits in the bins defined in Figure 10.5. For each observable the pull is defined as the difference between the fit result in bin i and the fit result in the full momenta region, divided by the uncertainty from the bin i fit.

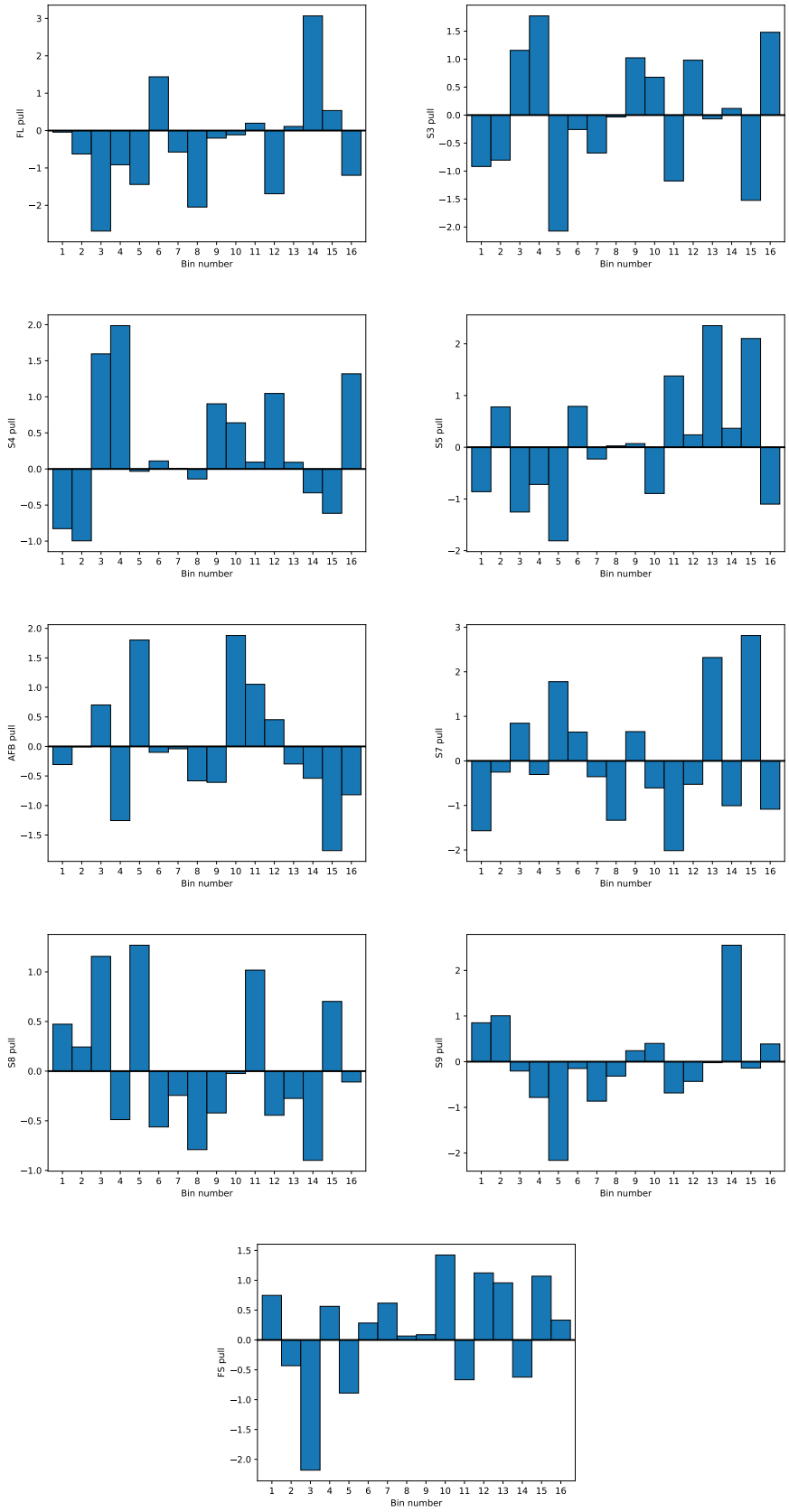


Figure 10.10: Pulls from comparing the CP-averaged observables from the 2017 $B^0 \to K^{*0} J/\psi$ fit and the fits in the bins defined in Figure 10.5. For each observable the pull is defined as the difference between the fit result in bin i and the fit result in the full momenta region, divided by the uncertainty from the bin i fit.

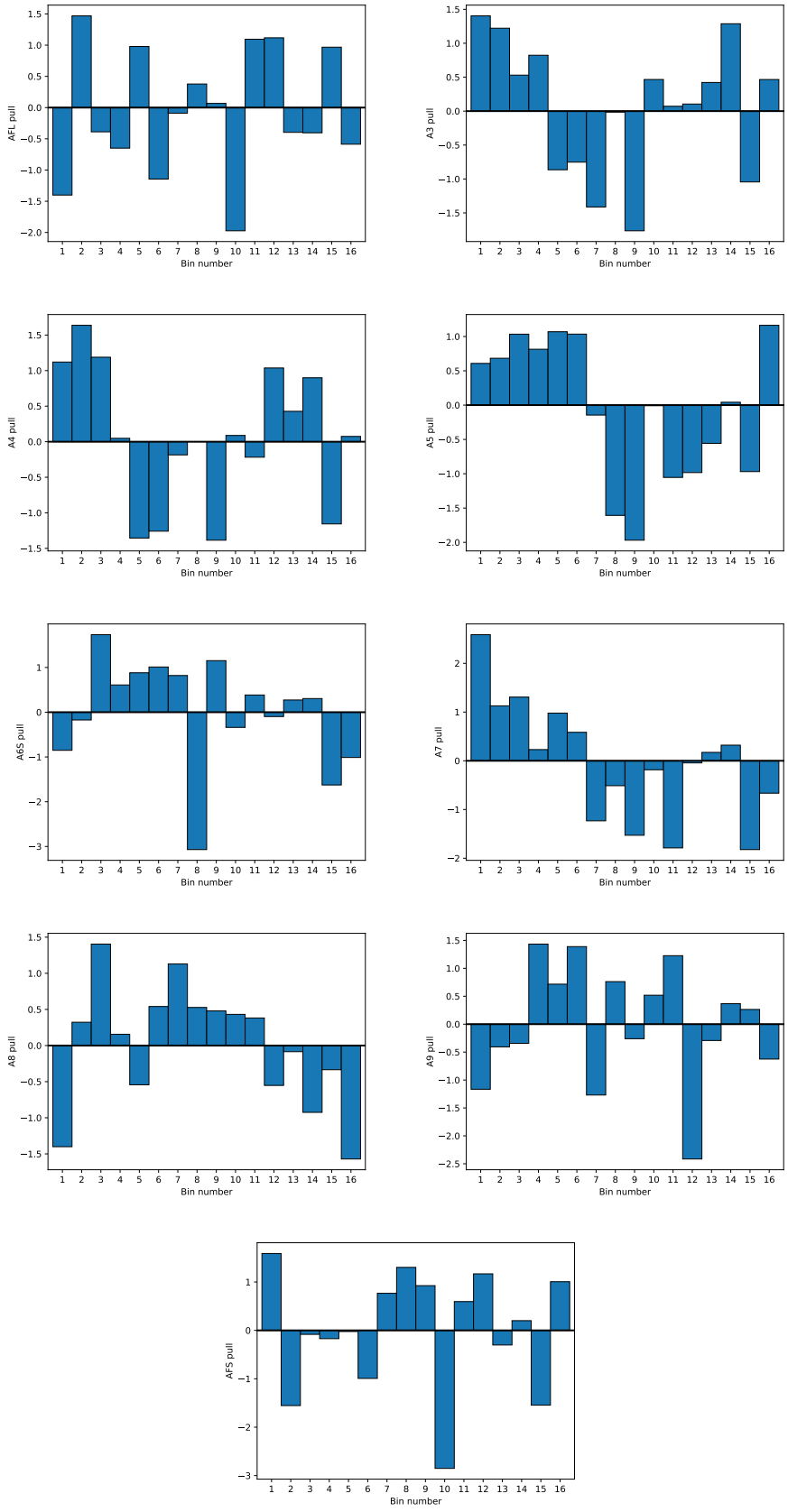


Figure 10.11: Pulls from comparing the CP-asymmetries from the 2017 $B^0 \to K^{*0} J/\psi$ fit and the fits in the bins defined in Figure 10.5. For each observable the pull is defined as the difference between the fit result in bin i and the fit result in the full momenta region, divided by the uncertainty from the bin i fit.

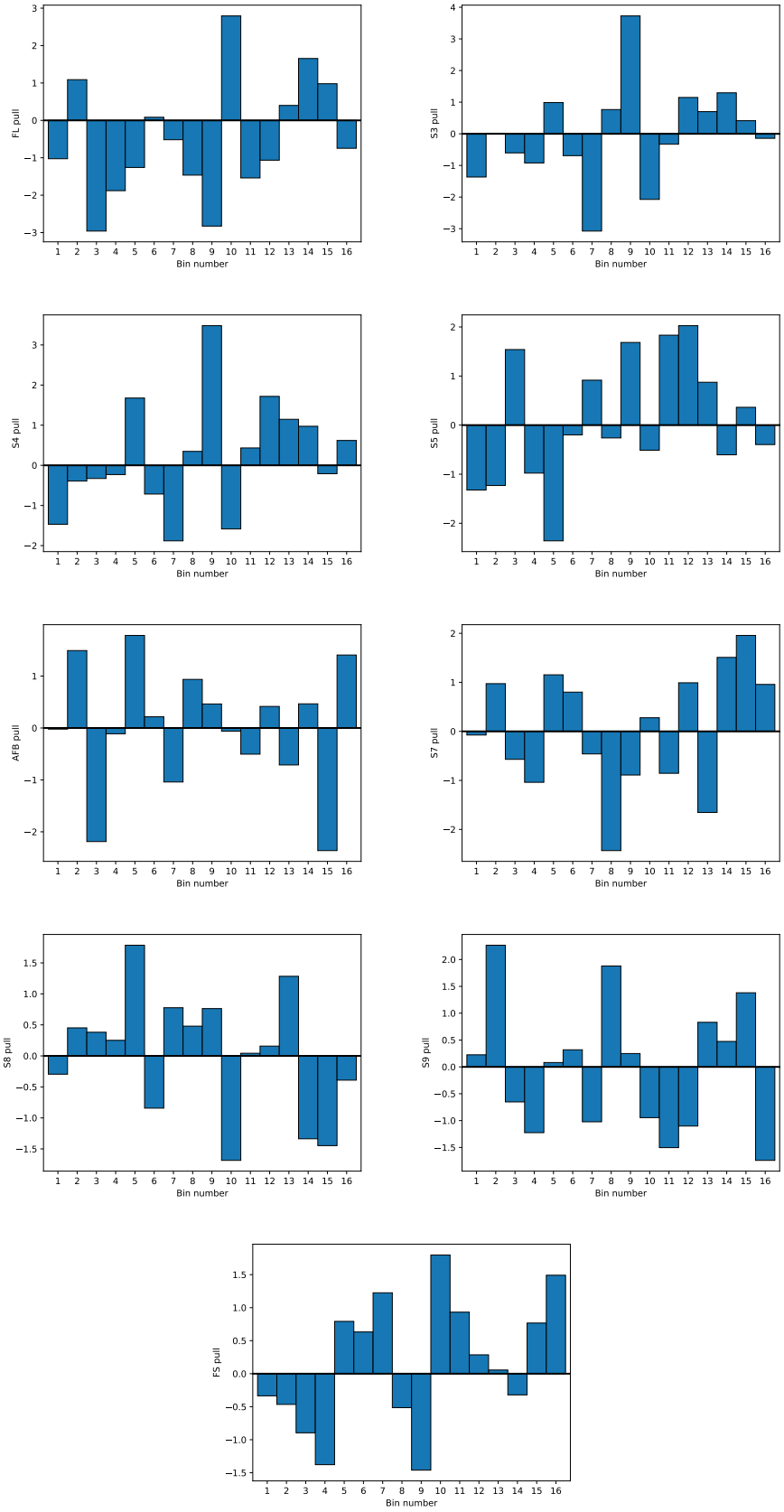


Figure 10.12: Pulls from comparing the CP-averaged observables from the 2018 $B^0 \to K^{*0} J/\psi$ fit and the fits in the bins defined in Figure 10.5. For each observable the pull is defined as the difference between the fit result in bin i and the fit result in the full momenta region, divided by the uncertainty from the bin i fit.

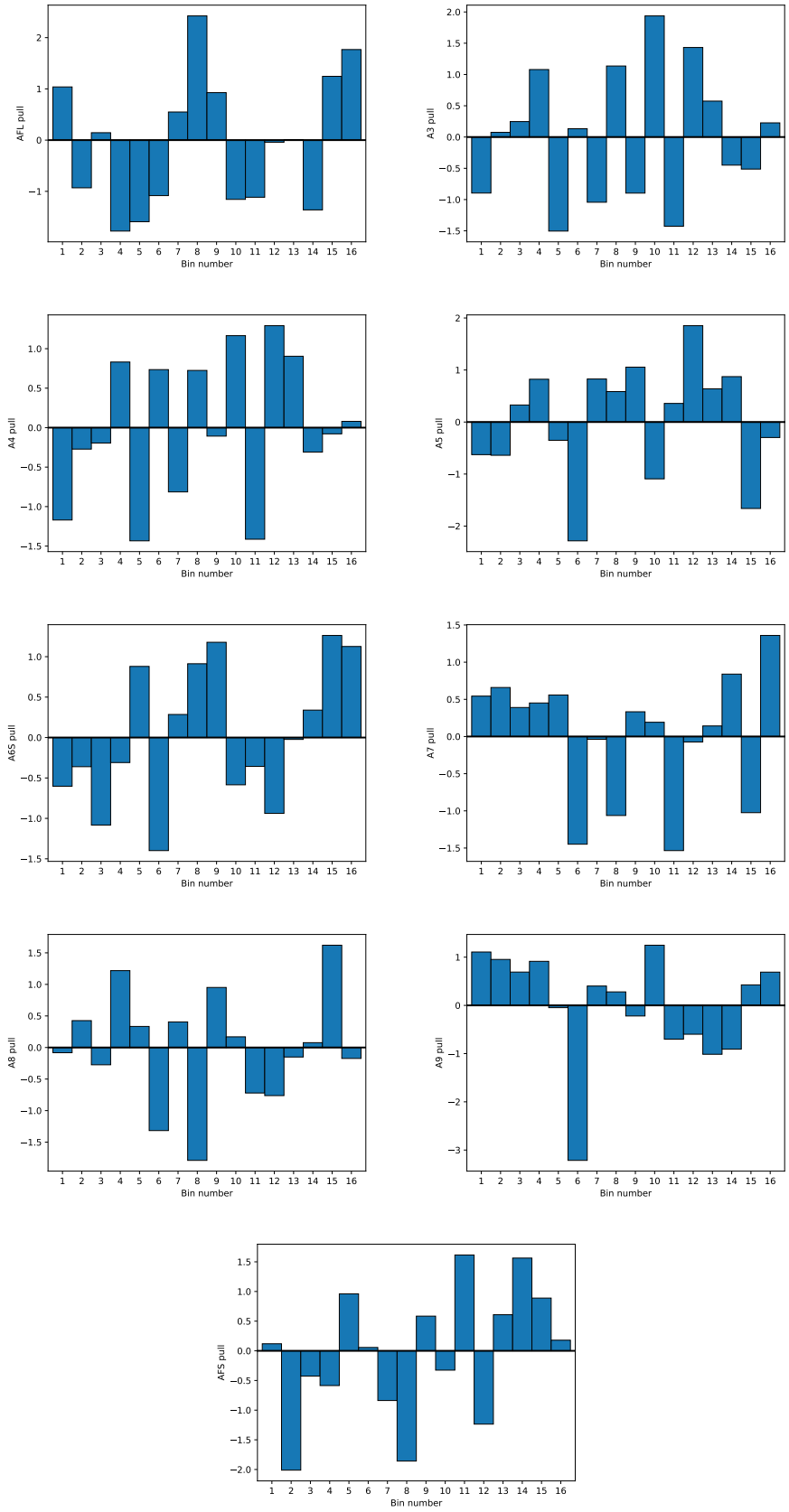


Figure 10.13: Pulls from comparing the CP-asymmetries from the 2018 $B^0 \to K^{*0} J/\psi$ fit and the fits in the bins defined in Figure 10.5. For each observable the pull is defined as the difference between the fit result in bin i and the fit result in the full momenta region, divided by the uncertainty from the bin i fit.

Table 10.7: p-values (in %) from chi-squared goodness-of-fit test when comparing fits in bins of (α, $\log_{10}(max(p_\mu^+, p_\mu^-))$) to the combined fit in each data period. In these tests, the uncertainty of the combined fit is assumed to be zero and the number of degrees of freedom is 17.

	Run 1	2016	2017	2018
Bin 1	49.6	98.6	13.7	85.3
Bin 2	38.8	81.5	73.3	36.5
Bin 3	67.7	29.7	7.2	30.8
Bin 4	92.8	11.1	73.7	37.3
Bin 5	87.1	1.0	0.1	8.3
Bin 6	20.7	55.6	83.1	7.2
Bin 7	48.3	78.5	84.8	44.1
Bin 8	53.9	0.4	11.2	0.5
Bin 9	85.5	29.5	67.3	1.4
Bin 10	45.7	6.7	38.6	11.6
Bin 11	2.9	13.9	30.0	41.6
Bin 12	30.2	1.9	50.9	34.6
Bin 13	68.7	66.9	70.4	88.8
Bin 14	3.5	90.9	3.5	45.7
Bin 15	96.9	36.1	1.7	1.0
Bin 16	23.4	72.3	67.6	64.3

Chapter 11

Results in the region

$1.25 < q^2 < 8$ GeV$^2/c^4$

The fit in the region $1.25 < q^2 < 8$ GeV$^2/c^4$ is performed with the full Run 1 + 2 dataset. The default fit performed in this amplitude ansatz analysis is the CP-average fit with a 4-parameter ansatz, with results shown in Section 11.3. Alternative fits including the asymmetries and with high numbers of parameters were tested, but yield insufficiently well behaved log-likelihood profiles. These are given in Sections 11.1 and 11.2 for the CP-asymmetries fit and 5-parameter CP-average fit respectively. As in the control mode fits, the B_s^0 yield relative to the B^0 and the mass difference between the B_s^0 and B^0 are fixed. Unlike the control mode fits, the signal Crystal Ball tail parameters are fixed in this fit configuration. Note all of the plots in this chapter are blinded. As a reminder, the amplitude ansatz for the P-wave is

$$\mathcal{A} = \alpha L_0(q^{2\prime}) + \beta L_1(q^{2\prime}) + \gamma L_2(q^{2\prime}) + \delta L_3(q^{2\prime}) + \varepsilon L_4(q^{2\prime}) \tag{11.1}$$

where L_i are Legendre polynomials of order i and $q^{2\prime}$ corresponds to q^2 scaled to the range $-1 \leq x \leq 1$. For the 4-parameter fits the values ε are fixed to zero.

11.1 Fit in the CP-asymmetries configuration with a 4-parameter ansatz

This section describes the CP-asymmetries fit. This fit is performed with a 4-parameter ansatz describing the P-wave amplitudes and flat S-wave. Profile likelihoods were examined for all of the P-wave amplitude components. The values of the amplitudes are blinded.

Figure 11.1 shows example profile log-likelihoods. The red line shows the expected shape of the log-likelihood surface from the Hessian error matrix and the green lines show the values of the log-likelihood at 1, 2, and 3 sigma as given by the Hessian error matrix. As seen in these plots, the Hessian error matrix is not always a good description of the log-likelihood surface. If there is a an asymmetry, this can be dealt with by fitting the log-likelihood with a bifurcated parabola. However the local minima seen are extremely difficult to deal with. Since the full likelihood surface including the local minima cannot be described by a parabola, the fit in the CP-asymmetries configuration will not be the default fit.

11.2 Fit in the CP-averages configuration with a 5-parameter ansatz

Since the fit with the asymmetries configuration yields log-likelihood profiles which exhibit local minima thus cannot be used, the fit in the CP-averages configuration with a 5-parameter ansatz is performed. This analysis aims for the amplitude model to be as generic as possible, so the aim is to parameterise the amplitudes with as many parameters as possible, as long as a good-quality fit is achieved and the description of the uncertainties is trustworthy. The log-likelihood profiles for this fit were studied.

Example profiles are shown in Figure 11.2. There are hints of local minima. Given the presence of local minima seen in these profiles, the fit with five parameters for the amplitude ansatz will not be the default fit.

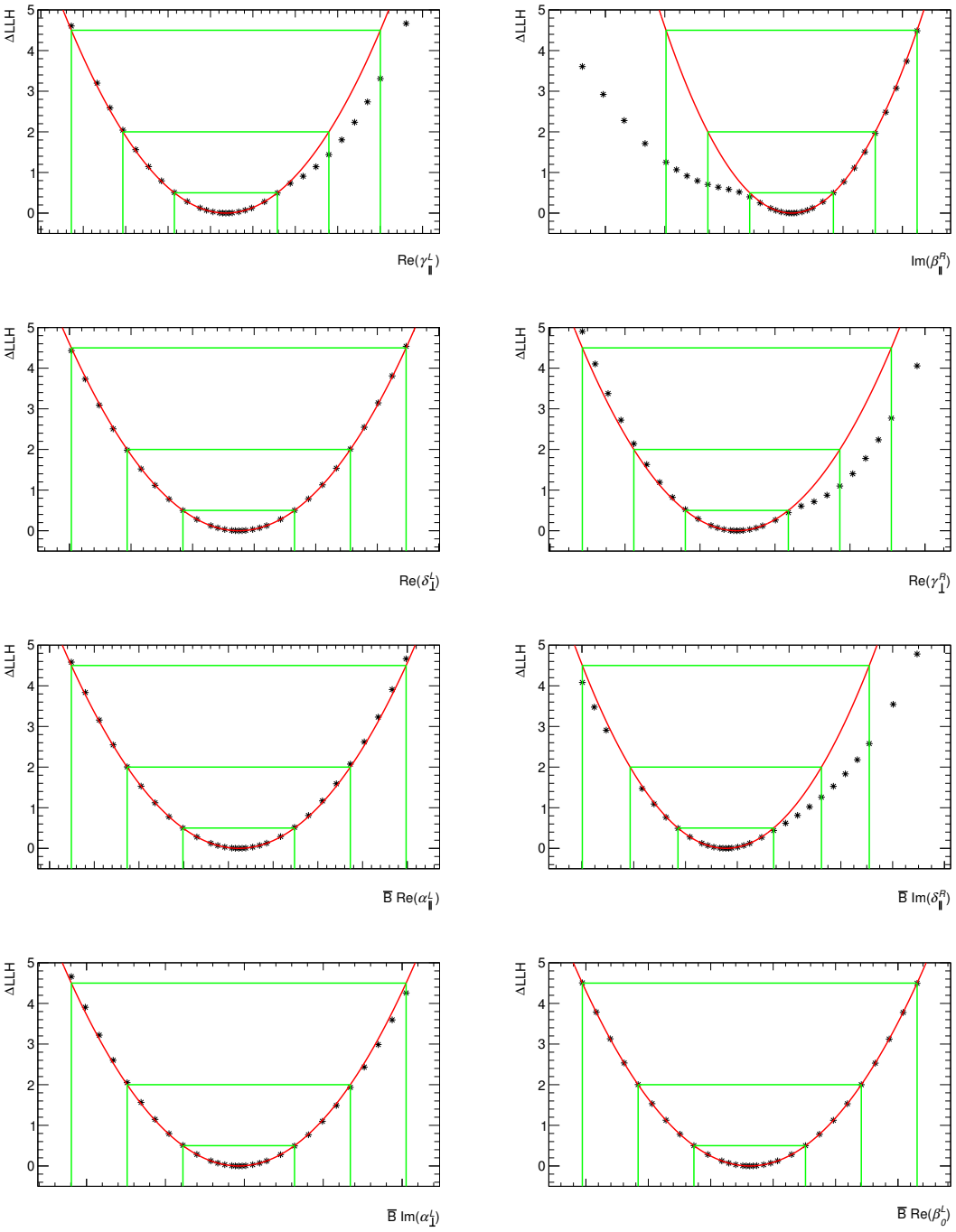


Figure 11.1: Example log-likelihood profiles from the combined Run 1 + Run 2 data fit in the region $1.25 < q^2 < 8$ GeV$^2/c^4$. The fit is performed in the CP-asymmetries configuration with four parameters for the P-wave amplitudes. The green lines corresponds to 1, 2, and 3 sigma from the Hessian error matrix. The red parabola corresponds to the Hessian error matrix.

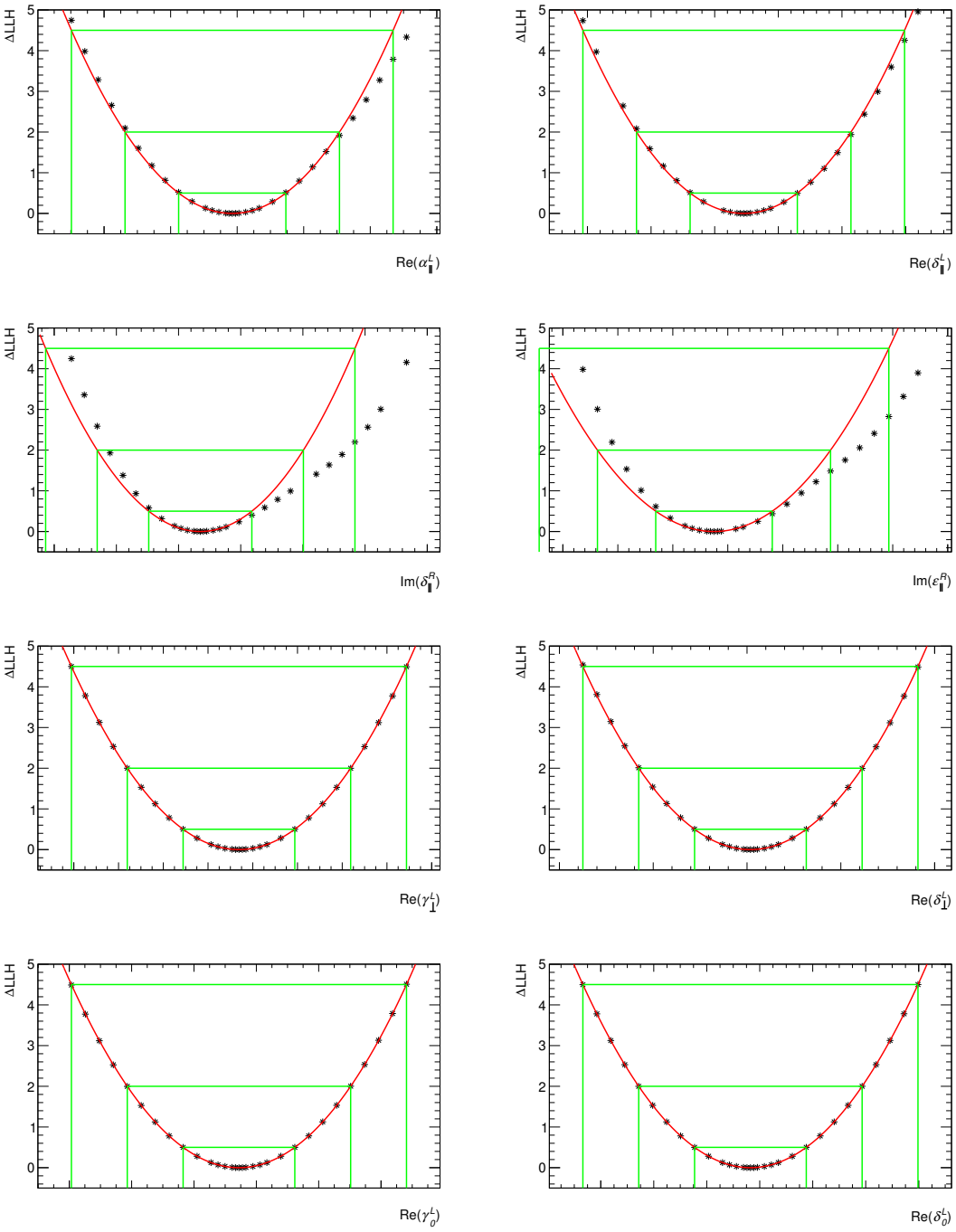


Figure 11.2: Example log-likelihood profiles from the combined Run 1 + Run 2 data fit in the region $1.25 < q^2 < 8$ GeV$^2/c^4$. The fit is performed in the CP-averages configuration with five parameters for the P-wave amplitudes. The green lines corresponds to 1, 2, and 3 sigma from the Hessian error matrix. The red parabola corresponds to the Hessian error matrix.

11.3 Fit in the CP-averages configuration with a 4-parameter ansatz

In this section, the fit is performed in the CP-averages configuration, where each P-wave amplitude floated has four parameters. The fit projections are shown in Figure 11.3.

The profile likelihoods of the coefficients are examined and are shown in Figures 11.4, 11.5, 11.6, and 11.7. Whilst the Hessian error matrix cannot describe some of the amplitudes perfectly (shown in red), it can be seen that the profiles can be well described by bifurcated parabolas (blue). The correlation matrix for this fit for the parameters of interest is shown as Figure 11.8. Most correlations have a magnitude smaller than 0.1 and there are no correlations with magnitudes greater than 0.6. The largest correlations are between the amplitude components of $\mathrm{Im}(\mathcal{A}_{\parallel}^{R})$ and also the correlation between $\mathrm{Im}(\mathcal{A}_{\parallel}^{R})$ and $\mathrm{Im}(\mathcal{A}_{\perp}^{L})$.

11.3.1 Goodness-of-fit test

In order to test the robustness of the amplitude ansatz, a goodness-of-fit test is performed. This follows the same method as the acceptance goodness-of-fit. The goodness-of-fit plot is shown in Figure 11.9. The blue histogram corresponds to the distribution of BDT figures of merit when training toys against a toy. This is the expected distribution from statistical fluctuations. The red line corresponds to training a BDT between the data and a toy. A p-value is computed, which is 70%. Thus the fit is a good description of the data.

11.4 Systematic uncertainties

Robust checks have been performed to confirm the 4-parameter fit in the CP-averages configuration should be the nominal fit in the Run 1 + Run 2 amplitude ansatz analysis of $B^0 \to K^{*0}\mu^+\mu^-$, in the q^2 region $1.25 < q^2 < 8$ GeV$^2/c^4$. Before unblinding the results of the amplitudes and observables, and thus the Wilson coefficients $\mathcal{C}_9$ and $\mathcal{C}_{10}$, systematic uncertainties will be computed. These include uncertainties due to the simulation corrections applied in the acceptance, such as corrections for the B^0 kinematics, trigger corrections, and tracking corrections. The systematic uncertainty due to acceptance simulation statistics will also be computed, as well as model systematic uncertainties such as the m_B model

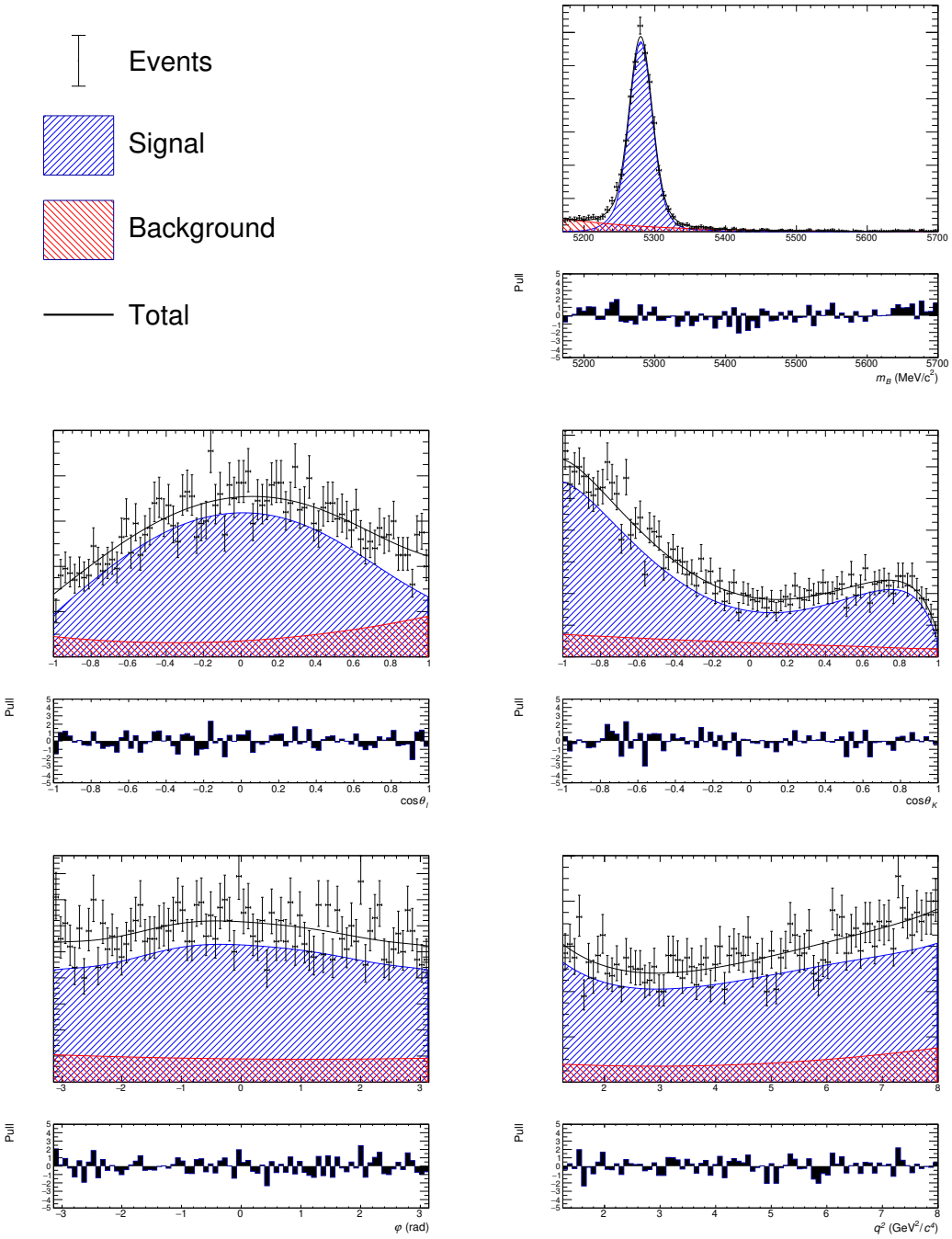


Figure 11.3: Projections for the combined Run 1 + Run 2 data fit in the region $1.25 < q^2 < 8$ GeV$^2/c^4$. The fit is performed in the CP-averages configuration, where each P-wave amplitude is parameterised with four parameters. The signal is shown in blue and the background is shown in red.

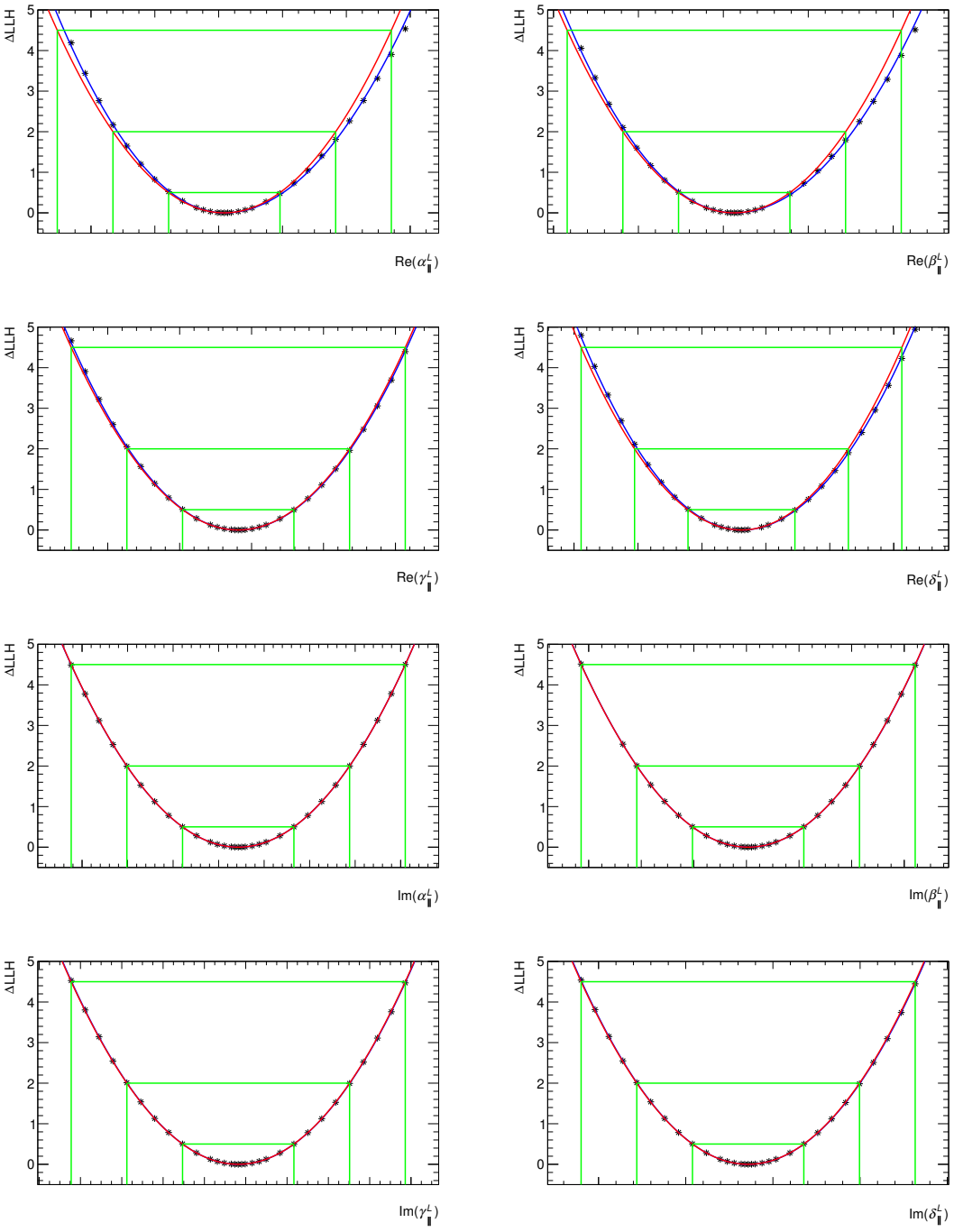


Figure 11.4: Profile log-likelihood from the combined Run 1 + Run 2 data fit in the region $1.25 < q^2 < 8$ GeV$^2/c^4$ for the left-handed parallel amplitudes. The fit is performed in the CP-averages configuration with four parameters for the P-wave amplitudes. The green lines corresponds to 1, 2, and 3 sigma from the Hessian error matrix. The red parabola corresponds to the Hessian error matrix and the blue parabolas are fits to these profiles with bifurcated parabolas.

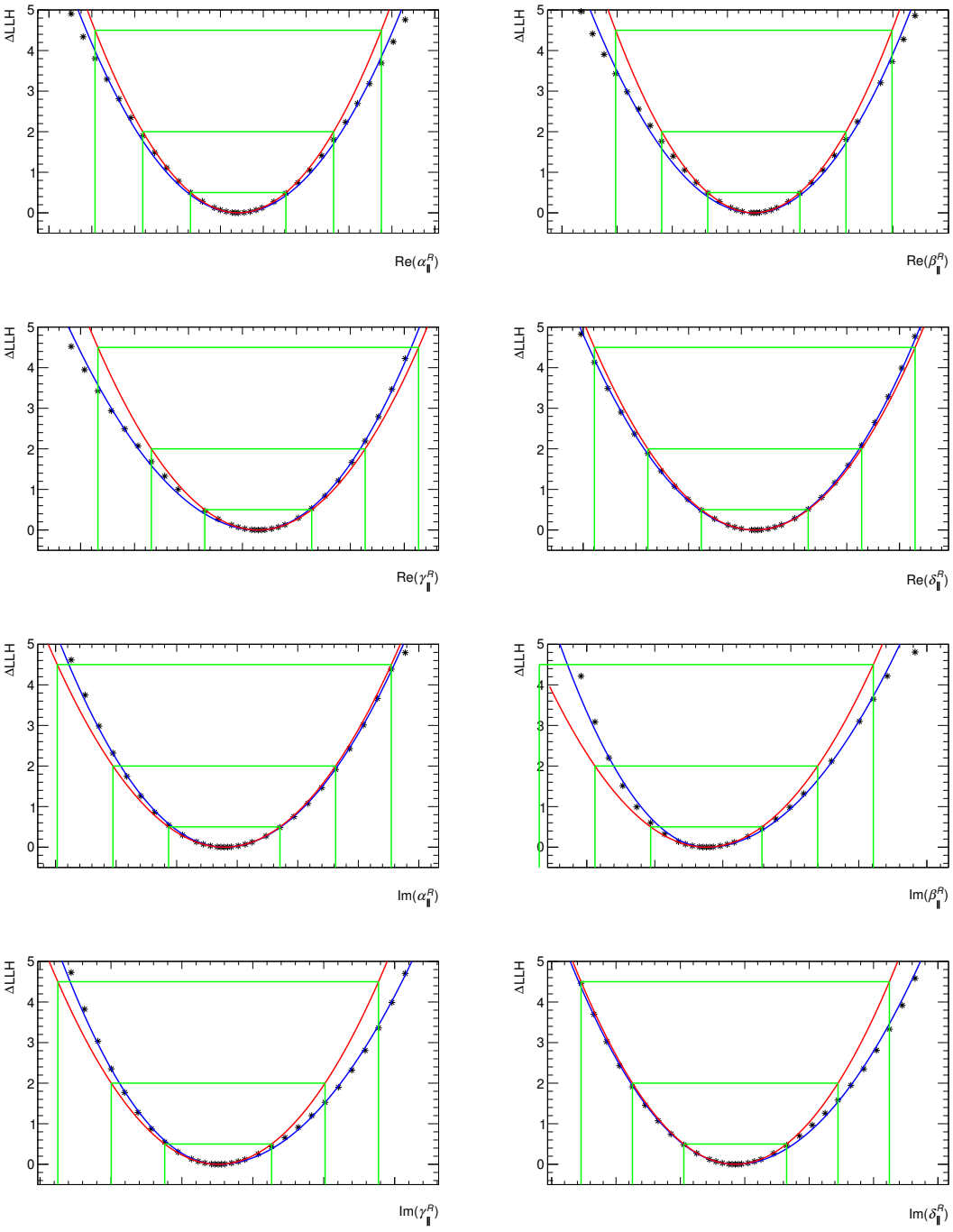


Figure 11.5: Profile log-likelihood from the combined Run 1 + Run 2 data fit in the region $1.25 < q^2 < 8$ GeV$^2/c^4$ for the right-handed parallel amplitudes. The fit is performed in the CP-averages configuration with four parameters for the P-wave amplitudes. The green lines corresponds to 1, 2, and 3 sigma from the Hessian error matrix. The red parabola corresponds to the Hessian error matrix and the blue parabolas are fits to these profiles with bifurcated parabolas.

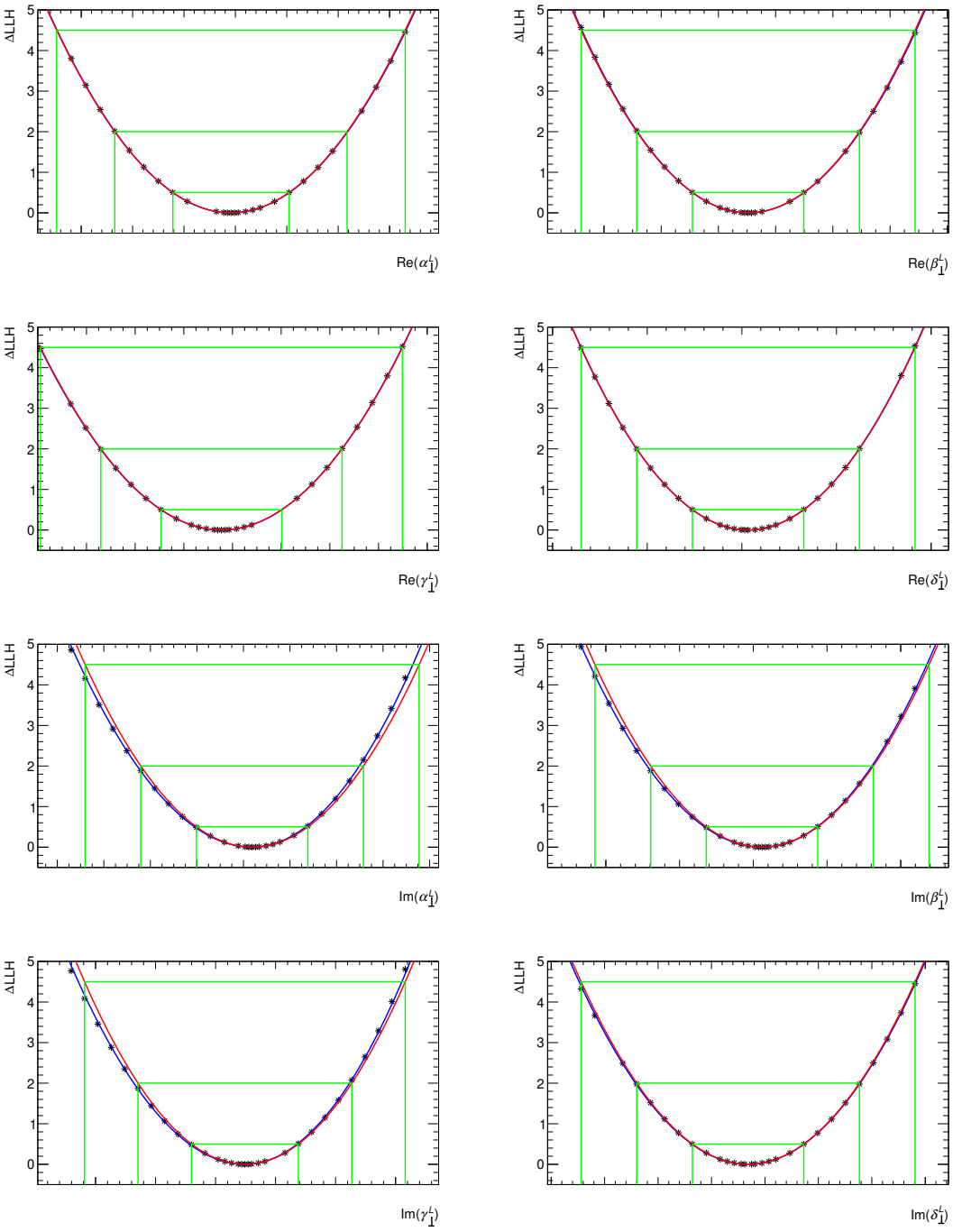


Figure 11.6: Profile log-likelihood from the combined Run 1 + Run 2 data fit in the region $1.25 < q^2 < 8$ GeV$^2/c^4$ for the left-handed transverse amplitudes. The fit is performed in the CP-averages configuration with four parameters for the P-wave amplitudes. The green lines corresponds to 1, 2, and 3 sigma from the Hessian error matrix. The red parabola corresponds to the Hessian error matrix and the blue parabolas are fits to these profiles with bifurcated parabolas.

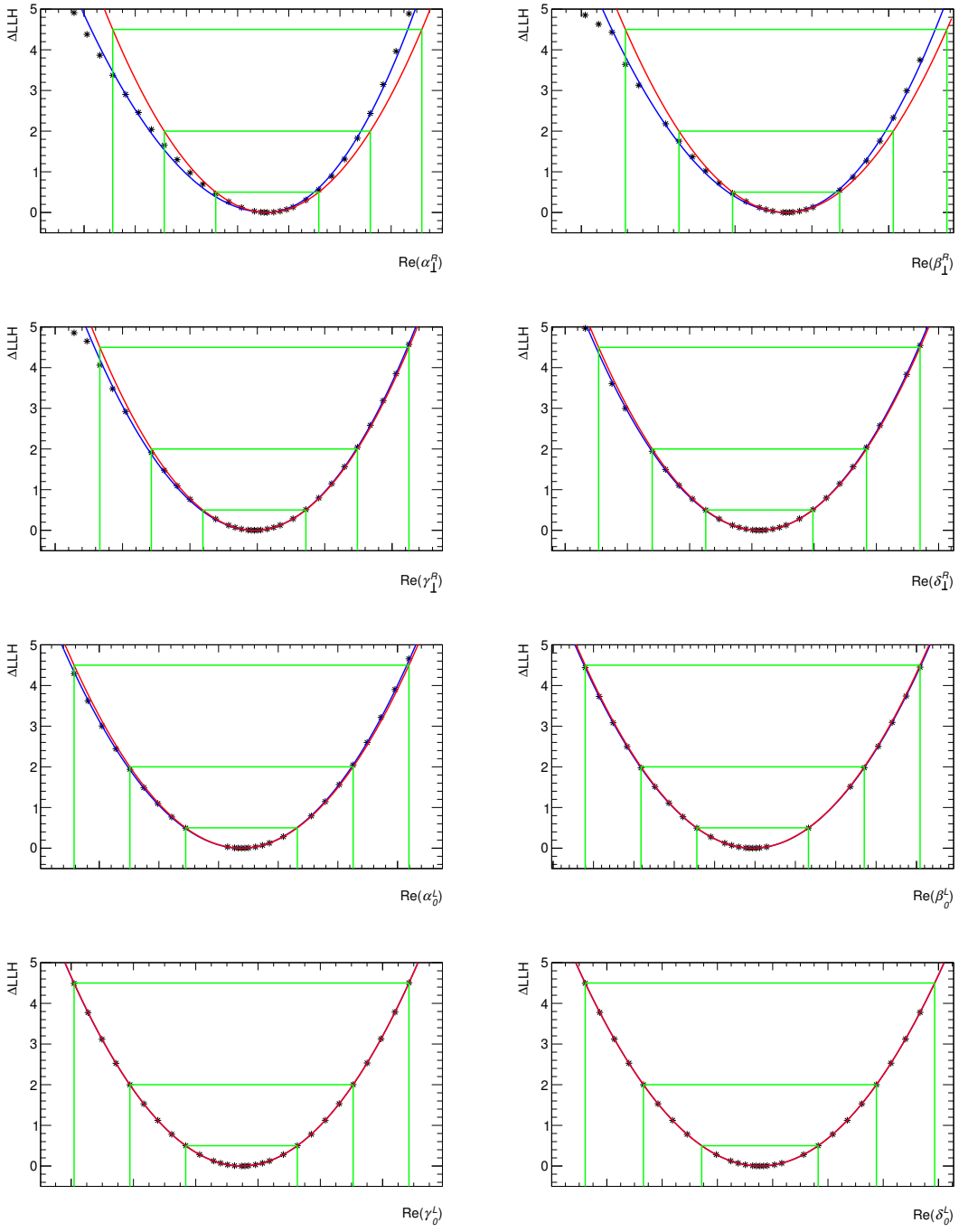


Figure 11.7: Profile log-likelihood from the combined Run 1 + Run 2 data fit in the region $1.25 < q^2 < 8$ GeV$^2/c^4$ for the right-handed transverse and longitudinal amplitudes. The fit is performed in the CP-averages configuration with four parameters for the P-wave amplitudes. The green lines corresponds to 1, 2, and 3 sigma from the Hessian error matrix. The red parabola corresponds to the Hessian error matrix and the blue parabolas are fits to these profiles with bifurcated parabolas.

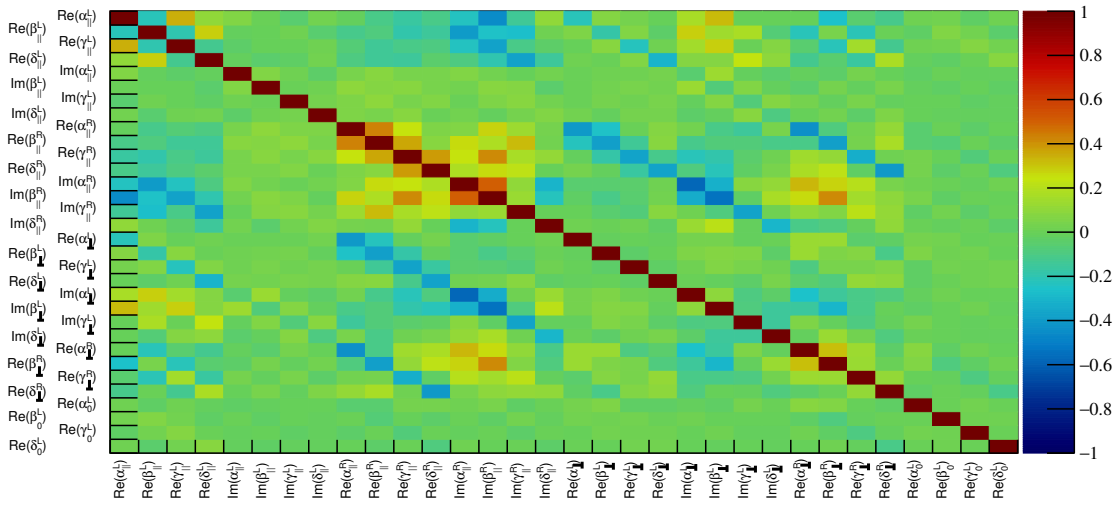


Figure 11.8: Correlation matrix for the amplitude ansatz fit to data in the $1.25 < q^2 < 8$ GeV$^2/c^4$ region for the parameters of interest.

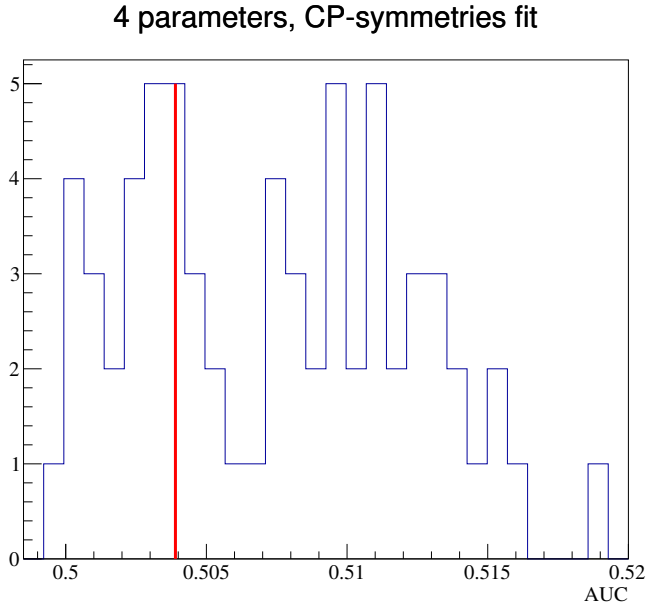


Figure 11.9: Goodness-of-fit plot corresponding to the fit in the $1.25 < q^2 < 8$ GeV$^2/c^4$ region. The blue distribution corresponds to the ROC AUCs of the BDTs to separate the benchmark pseudoexperiment from the other pseudoexperiments generated from the data best fit point. The red line is that for the BDT to separate benchmark pseudoexperiment from the data. The corresponding p-value is 70%.

and the $m_{K\pi}$ model. Finally the q^2 resolution systematic uncertainty will be computed. The systematic uncertainties overall are expected to be small compared to the statistical uncertainties.

As with the binned analyses, one of the largest expected systematic uncertainty is the uncertainty from the corrections to the `L0` trigger efficiency. Nominally for this systematic uncertainty, an alternative set of corrections are examined, i.e. $B^0 \to J/\psi K^{*0}$ rather than $B^+ \to J/\psi K^+$ for the trigger weights. However a conservative estimate of this systematic uncertainty is to not apply any corrections for the trigger and examine how the fit parameters change when the nominal trigger corrections are applied. This systematic uncertainty is computed by generating pseudoexperiments at 100 times the observed data yields. These pseudoexperiments are generated with acceptances without the nominal trigger corrections applied. Two fits are performed to each pseudoexperiment — one where the alternative (no corrections applied) acceptance is used and another where the nominal acceptance is used. For each fit parameter, the difference between the two fits is examined. The systematic uncertainty calculated is the mean and width of these differences, added up in quadrature. Table 11.1 presents the `L0` systematic uncertainty for each parameter as a ratio of the smaller of the upper or lower statistical uncertainty. As seen in the table, this conservative `L0` systematic uncertainty is small compared to the statistical uncertainty for the amplitude components. This means the uncertainties in this measurement will be completely dominated by the statistical uncertainties evaluated in Section 11.3.

Table 11.1: Ratio between the L0 systematic uncertainty and the smaller of the upper or lower statistical uncertainty for each fit parameter.

Parameter	$\frac{\texttt{L0 systematic uncertainty}}{\textbf{statistical uncertainty}}$
$\mathrm{Re}(\alpha_{\parallel}^{L})$	0.0255
$\mathrm{Re}(\beta_{\parallel}^{L})$	0.0111
$\mathrm{Re}(\gamma_{\parallel}^{L})$	0.0062
$\mathrm{Re}(\delta_{\parallel}^{L})$	0.0032
$\mathrm{Im}(\alpha_{\parallel}^{L})$	0.0041
$\mathrm{Im}(\beta_{\parallel}^{L})$	0.0012
$\mathrm{Im}(\gamma_{\parallel}^{L})$	0.0017
$\mathrm{Im}(\delta_{\parallel}^{L})$	0.0013
$\mathrm{Re}(\alpha_{\parallel}^{R})$	0.0169
$\mathrm{Re}(\beta_{\parallel}^{R})$	0.0076
$\mathrm{Re}(\gamma_{\parallel}^{R})$	0.0032
$\mathrm{Re}(\delta_{\parallel}^{R})$	0.0022
$\mathrm{Im}(\alpha_{\parallel}^{R})$	0.0034
$\mathrm{Im}(\beta_{\parallel}^{R})$	0.0064
$\mathrm{Im}(\gamma_{\parallel}^{R})$	0.0040
$\mathrm{Im}(\delta_{\parallel}^{R})$	0.0008
$\mathrm{Re}(\alpha_{\perp}^{L})$	0.0138
$\mathrm{Re}(\beta_{\perp}^{L})$	0.0121
$\mathrm{Re}(\gamma_{\perp}^{L})$	0.0036
$\mathrm{Re}(\delta_{\perp}^{L})$	0.0015
$\mathrm{Im}(\alpha_{\perp}^{L})$	0.0061
$\mathrm{Im}(\beta_{\perp}^{L})$	0.0026
$\mathrm{Im}(\gamma_{\perp}^{L})$	0.0032
$\mathrm{Im}(\delta_{\perp}^{L})$	0.0016
$\mathrm{Re}(\alpha_{\perp}^{R})$	0.0188
$\mathrm{Re}(\beta_{\perp}^{R})$	0.0077
$\mathrm{Re}(\gamma_{\perp}^{R})$	0.0058
$\mathrm{Re}(\delta_{\perp}^{R})$	0.0021
$\mathrm{Re}(\alpha_{0}^{L})$	0.0886
$\mathrm{Re}(\beta_{0}^{L})$	0.0279
$\mathrm{Re}(\gamma_{0}^{L})$	0.0043
$\mathrm{Re}(\delta_{0}^{L})$	0.0024

Chapter 12

Conclusions

As discussed in Chapter 1, now is a very exciting time in flavour physics. There is a growing collection of measurements taken by the LHCb experiment involving $b \to s\ell\ell$ and $b \to c\ell\nu$ showing discrepancies with respect to the SM. An example is the angular observables in the decay $B^0 \to K^{*0}\mu^+\mu^-$, which involves a $b \to s\ell\ell$ transition, where discrepancies with respect to the Standard Model are seen in the latest published analysis [7].

This thesis describes an amplitude analysis $B^0 \to K^{*0}\mu^+\mu^-$. Measuring the amplitudes provides a complete description of the angular distribution of $B^0 \to K^{*0}\mu^+\mu^-$. The analysis is also performed unbinned in q^2 in order to become senstivie to the shapes of the amplitudes and thus the observables in q^2. Since there are theoretical uncertainties and thus various parameterisations of the local and non-local hadronic effects, this amplitude analysis aims to be as model-independent as possible, whilst still being unbinned in q^2.

A selection strategy is described in Chapter 5 resulting in low levels of background. The combinatorial BDT is able to remove a large portion of the background. The selection results in estimated 4167 signal events, 726 combinatorial background events, and 38 combined peaking background events in the region $1.25 < q^2 < 8$ GeV$^2/c^4$ for the Run 1+2 LHCb datasets. A novel goodness-of-fit method is described in Chapter 6 where a goodness-of-fit test can be performed to a multi-dimensional function using machine learning techniques. The acceptance, which is used to account for the angular and q^2 efficiency is also described in this chapter.

The fit strategy is described in Chapter 7. Extensive work was performed validating the fitter and demonstrating the fit is unbiased in multidimensions, in addition to investigating

fundamental aspects of the analysis strategy, such as whether to fit $m_{K\pi}$ or which amplitude ansatz to use, as described in Chapter 8. Pseudoexperiments from the fit to data as described in Chapter 9 in the $1.25 < q^2 < 8$ GeV$^2/c^4$ region have been performed and behave well. The results from the control mode and in $1.25 < q^2 < 8$ GeV$^2/c^4$ are described in Chapters 10 and 11 respectively. Due to the presence of local minima in other fit configurations, the default strategy is to fit the data using 4-parameter ansatzes for the P-wave, and perform the fit in the CP-averages configuration. This fit behaves reasonably well and no local minima are seen $< 3\sigma$ away from the best fit point.

Future work entails computing systematic uncertainties (which are expected to be small compared to the statistical uncertainties) and unblinding the results. The results from this analysis will be of high interest to the flavour physics community since the amplitudes and thus the observables are unbinned in q^2, while aiming to be as model-independent as possible. In addition, datasets representative of the signal-only acceptance-corrected LHCb dataset can be generated from the fit results, allowing one to perform fits to this synthetic dataset with any choice of model and further study theoretical parameterisations of the decay $B^0 \to K^{*0}\mu^+\mu^-$.

Bibliography

[1] HFLAV. *Preliminary average of R(D) and R(D*) for Winter 2023*. https://hflav-eos.web.cern.ch/hflav-eos/semi/winter23_prel/html/RDsDsstar/RDRDs.html. [Online; accessed April 25, 2023]. 2023.

[2] LHCb Collaboration. "Measurements of the S-wave fraction in $B^0 \to K^+\pi^-\mu^+\mu^-$ decays and the $B^0 \to K^{*0}\mu^+\mu^-$ differential branching fraction". In: *JHEP* 11 (2016). [Erratum: JHEP04,142(2017)], p. 047. DOI: 10.1007/JHEP04(2017)142. arXiv: 1606.04731 [hep-ex].

[3] LHCb Collaboration. "Branching Fraction Measurements of the Rare $B_s^0 \to \phi\mu^+\mu^-$ and $B_s^0 \to f_2'(1525)\mu^+\mu^-$ Decays". In: *Phys. Rev. Lett.* 127 (2021), p. 151801. DOI: 10.1103/PhysRevLett.127.151801.

[4] LHCb Collaboration. "Differential branching fractions and isospin asymmetries of $B \to K^*\mu^+\mu^-$ decays". In: *JHEP* 06 (2014), p. 133. DOI: 10.1007/JHEP06(2014)133.

[5] Aoife Bharucha, David M. Straub, and Roman Zwicky. "$B \to V\ell^+\ell^-$ in the Standard Model from light-cone sum rules". In: *Journal of High Energy Physics* 2016.8 (2016). DOI: 10.1007/jhep08(2016)098.

[6] Ronald R. Horgan et al. "Lattice QCD calculation of form factors describing the rare decays $B \to K^*\ell^+\ell^-$ and $B_s \to \phi\ell^+\ell^-$". In: *Physical Review D* 89.9 (2014). DOI: 10.1103/physrevd.89.094501. URL: https://doi.org/10.1103%2Fphysrevd.89.094501.

[7] LHCb Collaboration. "Measurement of CP-Averaged Observables in the $B^0 \to K^{*0}\mu^+\mu^-$ Decay". In: *Phys. Rev. Lett.* 125 (2020), p. 011802. DOI: 10.1103/PhysRevLett.125.011802.

[8] Sébastien Descotes-Genon et al. "On the impact of power corrections in the prediction of $B \to K^* \mu^+ \mu^-$ observables". In: *Journal of High Energy Physics* 2014.12 (Dec. 2014). DOI: 10.1007/jhep12(2014)125. URL: https://doi.org/10.1007%2Fjhep12%282014%29125.

[9] Marcel Algueró et al. "$b \to s\ell\ell$ global fits after R_{K_S} and $R_{K^{*+}}$". In: *The European Physical Journal C* 82.4 (2022). DOI: 10.1140/epjc/s10052-022-10231-1.

[10] Bernat Capdevila et al. "Patterns of New Physics in $b \to s\ell^+\ell^-$ transitions in the light of recent data". In: *Journal of High Energy Physics* 2018.1 (Jan. 2018). DOI: 10.1007/jhep01(2018)093. URL: https://doi.org/10.1007%2Fjhep01%282018%29093.

[11] Ulrik Egede, Mitesh Patel, and Konstantinos A. Petridis. "Method for an unbinned measurement of the q^2 dependent decay amplitudes of $\bar{B}^0 \to \bar{K}^{*0} \mu^+ \mu^-$ decays". In: *JHEP* 06 (2015), p. 084. DOI: 10.1007/JHEP06(2015)084.

[12] R. Aaij et al. "Angular analysis of the $B^0 \to K^{*0} \mu^+ \mu^-$ decay using 3 fb^{-1} of integrated luminosity". In: *Journal of High Energy Physics* 2016.2 (Feb. 2016). DOI: 10.1007/jhep02(2016)104. URL: https://doi.org/10.1007%2Fjhep02%282016%29104.

[13] B. Efron. "Bootstrap Methods: Another Look at the Jackknife". In: *The Annals of Statistics* 7.1 (1979), pp. 1–26. DOI: 10.1214/aos/1176344552. URL: https://doi.org/10.1214/aos/1176344552.

[14] Particle Data Group. "Review of Particle Physics". In: *Phys. Rev. D* 98 (2018), p. 030001. DOI: 10.1103/PhysRevD.98.030001.

[15] Sheldon L. Glashow. "Partial-symmetries of weak interactions". In: *Nuclear Physics* 22.4 (1961), pp. 579–588. ISSN: 0029-5582. DOI: https://doi.org/10.1016/0029-5582(61)90469-2.

[16] Abdus Salam and J. C. Ward. "Gauge Theory of Elementary Interactions". In: *Phys. Rev.* 136 (3B 1964), B763–B768. DOI: 10.1103/PhysRev.136.B763.

[17] Steven Weinberg. "A Model of Leptons". In: *Phys. Rev. Lett.* 19 (21 1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264.

[18] Mark Thomson. *Modern particle physics*. New York: Cambridge University Press, 2013. ISBN: 978-1-107-03426-6. DOI: 10.1017/CBO9781139525367.

[19] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". In: *Physics Review Letters* 13.9 (1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321.

[20] Peter W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Physics Review Letters* 13.16 (1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508.

[21] G. S. Guralnik, C. R. Hagen, and T. W. Kibble. "Global Conservation Laws and Massless Particles". In: *Physics Review Letters* 13.20 (1964), pp. 585–587. DOI: 10.1103/PhysRevLett.13.585.

[22] Nicola Cabibbo. "Unitary Symmetry and Leptonic Decays". In: *Physics Review Letters* 10.12 (1963), pp. 531–533. DOI: 10.1103/PhysRevLett.10.531.

[23] M. Kobayashi and T. Maskawa. "CP-Violation in the Renormalizable Theory of Weak Interaction". In: *Progress of Theoretical Physics* 49.2 (1973), pp. 652–657. DOI: 10.1143/PTP.49.652.

[24] Charles W. Misner, Kip S. Thorne, and John Archibald Wheeler. *Gravitation*. W. H. Freeman, 1973. ISBN: 978-0-7167-0344-0.

[25] Vera C. Rubin, W. Kent Ford Jr., and Norbert Thonnard. "Extended rotation curves of high-luminosity spiral galaxies. IV. Systematic dynamical properties, Sa through Sc". In: *Astrophys. J. Lett.* 225 (1978), pp. L107–L111. DOI: 10.1086/182804.

[26] Douglas Clowe et al. "A Direct Empirical Proof of the Existence of Dark Matter". In: *The Astrophysical Journal* 648.2 (2006), pp. L109–L113. DOI: 10.1086/508162.

[27] Planck Collaboration. "Planck 2018 results". In: *Astronomy I& Astrophysics* 641 (2020), A6. DOI: 10.1051/0004-6361/201833910.

[28] Katherine Garrett and Gintaras Duda. "Dark Matter: A Primer". In: *Advances in Astronomy* 2011 (2011), pp. 1–22. DOI: 10.1155/2011/968283.

[29] LUX-ZEPLIN Collaboration. "First Dark Matter Search Results from the LUX-ZEPLIN (LZ) Experiment". In: *Phys. Rev. Lett.* 131 (2023), p. 041002. DOI: 10.1103/PhysRevLett.131.041002.

[30] F. Wilczek. “Problem of Strong P and T Invariance in the Presence of Instantons”. In: *Phys. Rev. Lett.* 40 (1978), pp. 279–282. DOI: 10.1103/PhysRevLett.40.279.

[31] Steven Weinberg. “A New Light Boson?” In: *Phys. Rev. Lett.* 40 (1978), pp. 223–226. DOI: 10.1103/PhysRevLett.40.223.

[32] R. D. Peccei and Helen R. Quinn. “CP Conservation in the Presence of Pseudoparticles”. In: *Phys. Rev. Lett.* 38 (1977), pp. 1440–1443. DOI: 10.1103/PhysRevLett.38.1440.

[33] A. D. Sakharov. “Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe”. In: *Pisma Zh. Eksp. Teor. Fiz.* 5 (1967), pp. 32–35. DOI: 10.1070/PU1991v034n05ABEH002497.

[34] G. ’t Hooft. “Symmetry Breaking through Bell-Jackiw Anomalies”. In: *Phys. Rev. Lett.* 37 (1976), pp. 8–11. DOI: 10.1103/PhysRevLett.37.8.

[35] C. Abel et al. “Measurement of the Permanent Electric Dipole Moment of the Neutron”. In: *Phys. Rev. Lett.* 124 (2020), p. 081803. DOI: 10.1103/PhysRevLett.124.081803.

[36] The T2K Collaboration. “Constraint on the matter-antimatter symmetry-violating phase in neutrino oscillations”. In: *Nature* 580.7803 (2020), pp. 339–344. DOI: 10.1038/s41586-020-2177-0.

[37] Adam G. Riess et al. “Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant”. In: *The Astronomical Journal* 116.3 (1998), pp. 1009–1038. DOI: 10.1086/300499.

[38] Bharat Ratra and P. J. E. Peebles. “Cosmological consequences of a rolling homogeneous scalar field”. In: *Phys. Rev. D* 37 (1988), pp. 3406–3427. DOI: 10.1103/PhysRevD.37.3406.

[39] R. R. Caldwell, Rahul Dave, and Paul J. Steinhardt. “Cosmological Imprint of an Energy Component with General Equation of State”. In: *Phys. Rev. Lett.* 80 (1998), pp. 1582–1585. DOI: 10.1103/PhysRevLett.80.1582.

[40] Justin Khoury and Amanda Weltman. “Chameleon cosmology”. In: *Physical Review D* 69.4, 044026 (2004), p. 044026. DOI: 10.1103/PhysRevD.69.044026.

[41] Duncan Farrah et al. “A Preferential Growth Channel for Supermassive Black Holes in Elliptical Galaxies at z $<$ 2”. In: *The Astrophysical Journal* 943.2 (2023), p. 133. DOI: 10.3847/1538-4357/acac2e.

[42] Raymond Davis, Don S. Harmer, and Kenneth C. Hoffman. “Search for Neutrinos from the Sun”. In: *Phys. Rev. Lett.* 20 (1968), pp. 1205–1209. DOI: 10.1103/PhysRevLett.20.1205.

[43] Super-Kamiokande Collaboration. “Evidence for Oscillation of Atmospheric Neutrinos”. In: *Phys. Rev. Lett.* 81 (1998), pp. 1562–1567. DOI: 10.1103/PhysRevLett.81.1562.

[44] KATRIN Collaboration. “Direct neutrino-mass measurement with sub-electronvolt sensitivity”. In: *Nature Phys.* 18.2 (2022), pp. 160–166. DOI: 10.1038/s41567-021-01463-1.

[45] Kenneth G. Wilson. “Non-Lagrangian Models of Current Algebra”. In: *Phys. Rev.* 179 (1969), pp. 1499–1512. DOI: 10.1103/PhysRev.179.1499.

[46] Andrzej J. Buras. *Weak Hamiltonian, CP Violation and Rare Decays*. 1998. arXiv: hep-ph/9806471 [hep-ph].

[47] Wolfgang Altmannshofer et al. “Symmetries and asymmetries of $B \to K^* \mu^+ \mu^-$-decays in the Standard Model and beyond”. In: *Journal of High Energy Physics* 2009.01 (2009), p. 019. DOI: 10.1088/1126-6708/2009/01/019.

[48] Elliot Leader. *Spin in Particle Physics*. Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology. Cambridge University Press, 2001. DOI: 10.1017/CBO9780511524455.

[49] James Gratrex, Markus Hopfer, and Roman Zwicky. “Generalised helicity formalism, higher moments and the $B \to K_{J_K}(\to K\pi)\bar{\ell}_1\ell_2$ angular distributions”. In: *Phys. Rev. D* 93.5 (2016), p. 054008. DOI: 10.1103/PhysRevD.93.054008.

[50] Sébastien Descotes-Genon et al. “Light-cone sum rules for S-wave $B \to K\pi$ form factors”. In: *Journal of High Energy Physics* 2023.6 (2023). DOI: 10.1007/jhep06(2023)034.

[51] Sébastien Descotes-Genon, Alexander Khodjamirian, and Javier Virto. "Light-cone sum rules for B → K π form factors and applications to rare decays". In: *Journal of High Energy Physics* 2019.12 (2019). DOI: 10.1007/jhep12(2019)083.

[52] Nico Gubernari, Danny van Dyk, and Javier Virto. "Non-local matrix elements in $B(s) \to \{K^*, \phi\}\ell^+\ell^-$". In: *Journal of High Energy Physics* 2021.2 (2021). DOI: 10.1007/jhep02(2021)088.

[53] LHCb Collaboration. "Differential branching fraction and angular analysis of the decay $B^0 \to K^{*0}\mu^+\mu^-$". In: *Journal of High Energy Physics* (2013). DOI: 10.1007/jhep08(2013)131.

[54] Marcel Algueró et al. "A complete description of P- and S-wave contributions to the $B^0 \to K^+\pi^-\ell^+\ell^-$ decay". In: *Journal of High Energy Physics* 2021.12 (2021). DOI: 10.1007/jhep12(2021)085.

[55] Sébastien Descotes-Genon et al. "Implications from clean observables for the binned analysis of $B \to K^*\mu^+\mu^-$ - at large recoil". In: *Journal of High Energy Physics* 2013.1 (2013). DOI: 10.1007/jhep01(2013)048.

[56] LHCb Collaboration. "Measurement of Form-Factor-Independent Observables in the Decay $B^0 \to K^{*0}\mu^+\mu^-$". In: *Physical Review Letters* 111.19 (2013). DOI: 10.1103/physrevlett.111.191801.

[57] Belle Collaboration. "Experimental constraints on the spin and parity of the $Z(4430)^+$". In: *Physical Review D* 88.7 (2013). DOI: 10.1103/physrevd.88.074026.

[58] D. Aston et al. "A study of $K^-\pi^+$ scattering in the reaction $K^-p \to K^-\pi^+n$ at 11 GeV/c". In: *Nuclear Physics B* 296.3 (1988), pp. 493–526. ISSN: 0550-3213. DOI: https://doi.org/10.1016/0550-3213(88)90028-4.

[59] LHCb Collaboration. "Observation of the resonant character of the $Z(4430)^-$ state". In: *Physical Review Letters* 112.22 (2014). DOI: 10.1103/physrevlett.112.222002.

[60] Joel Bressieux et al. "Evidence for the resonant character of the $Z(4430)^- \to \psi(2S)\pi^-$ mass peak observed in $B^0 \to \psi(2S)K^+\pi^-$ decays, and determination of the $Z(4430)^-$ spin-parity". In: (2013). URL: https://cds.cern.ch/record/1553766.

[61] J. Matias et al. "Complete anatomy of $\bar{B} \to \bar{K}^{*0}\ell\ell$ and its angular distribution". In: *Journal of High Energy Physics* 2012.4 (2012). DOI: 10.1007/jhep04(2012)104.

[62] Lyndon Evans and Philip Bryant. "LHC Machine". In: *Journal of Instrumentation* 3.08 (2008), S08001. DOI: 10.1088/1748-0221/3/08/S08001.

[63] Oliver Sim Brüning et al. *LHC Design Report*. CERN Yellow Reports: Monographs. Geneva: CERN, 2004. URL: http://cds.cern.ch/record/782076.

[64] Christian Elsässer. *$\bar{b}b$ production angle plots*. 2022. URL: https://lhcb.web.cern.ch/speakersbureau/html/bb_productionangles.html.

[65] "LHCb detector performance". In: *International Journal of Modern Physics A* 30.07 (2015), p. 1530022. DOI: 10.1142/s0217751x15300227.

[66] The LHCb VELO Group. "Performance of the LHCb Vertex Locator". In: *Journal of Instrumentation* 9.09 (2014), P09007–P09007. DOI: 10.1088/1748-0221/9/09/p09007.

[67] The LHCb VELO Group. "Measurement of thermal properties of the LHCb VELO detector using track-based software alignment". In: (2022). DOI: 10.48550/ARXIV.2205.13477.

[68] The LHCb RICH Collaboration. "Performance of the LHCb RICH detector at the LHC". In: *The European Physical Journal C* 73.5 (2013). DOI: 10.1140/epjc/s10052-013-2431-9.

[69] R. Calabrese et al. "Performance of the LHCb RICH detectors during LHC Run 2". In: *Journal of Instrumentation* 17.07 (2022), P07013. DOI: 10.1088/1748-0221/17/07/p07013.

[70] LHCb Collaboration. *LHCb reoptimized detector design and performance: Technical Design Report, CERN, CERN-LHCC-2003-030*. Geneva, 2003. URL: https://cds.cern.ch/record/630827?ln=en.

[71] F Archilli et al. "Performance of the Muon Identification at LHCb". In: *Journal of Instrumentation* 8.10 (2013), P10020–P10020. DOI: 10.1088/1748-0221/8/10/p10020.

[72] R. Aaij et al. "Design and performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC". In: *Journal of Instrumentation* 14.04 (2019), P04013–P04013. DOI: 10.1088/1748-0221/14/04/p04013.

[73] LHCb Collaboration. *Trigger Schemes*. 2023. URL: http://lhcb.web.cern.ch/lhcb/speakersbureau/html/TriggerScheme.html.

[74] Alessio Piucci. "The LHCb Upgrade". In: *Journal of Physics: Conference Series* 878.1 (2017), p. 012012. DOI: 10.1088/1742-6596/878/1/012012.

[75] Tomasz Szumlak. "Real time analysis with the upgraded LHCb trigger in Run III". In: *Journal of Physics: Conference Series* 898.3 (2017), p. 032051. DOI: 10.1088/1742-6596/898/3/032051.

[76] M Clemencic et al. "The LHCb Simulation Application, Gauss: Design, Evolution and Experience". In: *Journal of Physics: Conference Series* 331.3 (2011), p. 032023. DOI: 10.1088/1742-6596/331/3/032023.

[77] G. Corti et al. "Software for the LHCb experiment". In: *IEEE Symposium Conference Record Nuclear Science 2004.* Vol. 4. 2004, 2048–2052 Vol. 4. DOI: 10.1109/NSSMIC.2004.1462666.

[78] R. Calabrese et al. "Performance of the LHCb RICH detectors during LHC Run 2". In: *Journal of Instrumentation* 17.07 (2022), P07013. DOI: 10.1088/1748-0221/17/07/p07013. URL: https://doi.org/10.1088%2F1748-0221%2F17%2F07%2Fp07013.

[79] R Aaij et al. "The LHCb trigger and its performance in 2011". In: *Journal of Instrumentation* 8.04 (2013), P04022–P04022. DOI: 10.1088/1748-0221/8/04/p04022.

[80] The LHCb Collaboration. "Measurement of CP asymmetries in the decays $B^0 \to K^{*0}\mu^+\mu^-$and $B^+ \to K^+\mu^+\mu^-$". In: *Journal of High Energy Physics* 2014.177 (2014). DOI: https://doi.org/10.1007/JHEP09(2014)177.

[81] Tianqi Chen and Carlos Guestrin. "XGBoost: A Scalable Tree Boosting System". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16. San Francisco, California, USA: ACM, 2016, pp. 785–794. ISBN: 978-1-4503-4232-2. DOI: 10.1145/2939672.2939785.

[82] Malte Hecker. "Searching for new physics in $B^0 \to K^{*0}\mu^+\mu^-$decays". PhD thesis. Imperial College London, 2021. URL: https://spiral.imperial.ac.uk/handle/10044/1/92031.

[83] S. Tolk, J. Albrecht, F. Dettori, A. Pellegrino. "Data driven trigger efficiency determination at LHCb". In: (2014). URL: https://cds.cern.ch/record/1701134.

[84] M. Pivk and F.R. Le Diberder. "A statistical tool to unfold data distributions". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 555.1-2 (2005), pp. 356–369. DOI: 10.1016/j.nima.2005.08.106.

[85] Giovanni Punzi. "Comments on Likelihood fits with variable resolution". In: (2004). DOI: 10.48550/ARXIV.PHYSICS/0401045.

[86] F. James. *Function Minimization and Error Analysis Reference Manual*. 2023. URL: http://cdssls.cern.ch/record/2296388/files/minuit.pdf.

[87] John Erthal Gaiser. "Charmonium Spectroscopy From Radiative Decays of the J/ψ and ψ'". Other thesis. 1982. URL: https://inspirehep.net/files/3a6ffc353519cd9cbac9b439b523ef6d.

[88] The LHCb Collaboration. "Precise measurement of the f_s/f_d ratio of fragmentation fractions and of B^0_s decay branching fractions". In: *Physical Review D* 104.3 (2021). DOI: 10.1103/physrevd.104.032005.

[89] The BABAR Collaboration and B. Aubert. *Measurement of Decay Amplitudes of $B \to (c\bar{c})K^{*0}$ with an angular analysis, for $(c\bar{c}) = J/\psi$, $\psi(2S)$ and χ_{c1}*. 2006. arXiv: hep-ex/0607081 [hep-ex].

[90] K. Chilikin et al. "Observation of a new charged charmonium like state in $\bar{B}^0 \to J/\psi K^-\pi^+$ decays". In: *Physical Review D* 90.11 (2014). DOI: 10.1103/physrevd.90.112009.

Appendix A

Comparison of the input variables in $B^0 \to K^{*0} J/\psi$ simulation and sWeighted $B^0 \to K^{*0} J/\psi$ data for the combinatorial BDT in for Run 1, 2016, and 2018

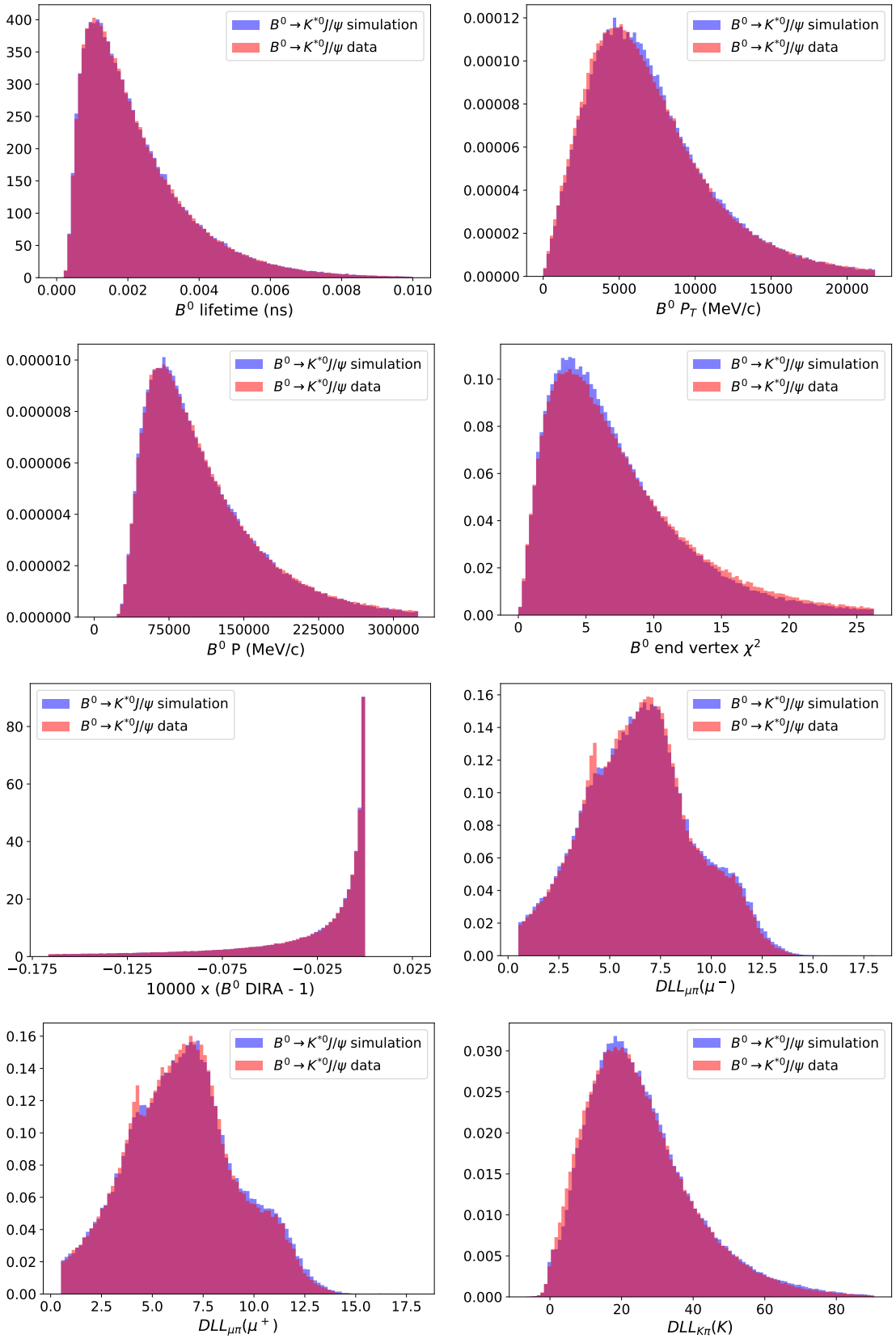


Figure A.1: Comparison between $B^0 \to K^{*0}J/\psi$ simulation and sWeighted $B^0 \to K^{*0}J/\psi$ data for the BDT training variables for Run 1. The distributions are normalised to unit area.

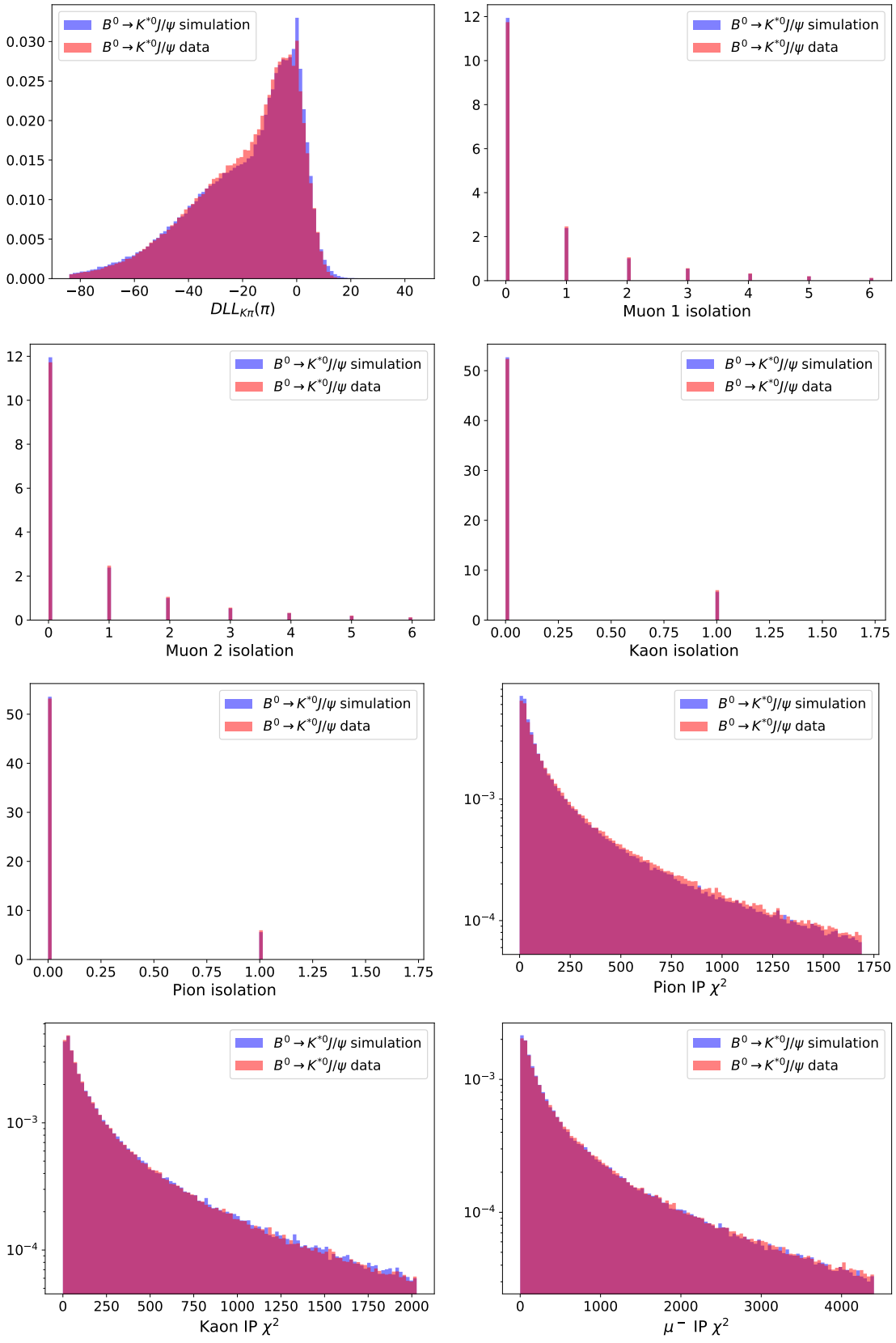


Figure A.2: Comparison between $B^0 \to K^{*0} J/\psi$ simulation and sWeighted $B^0 \to K^{*0} J/\psi$ data for the BDT training variables for Run 1. The distributions are normalised to unit area.

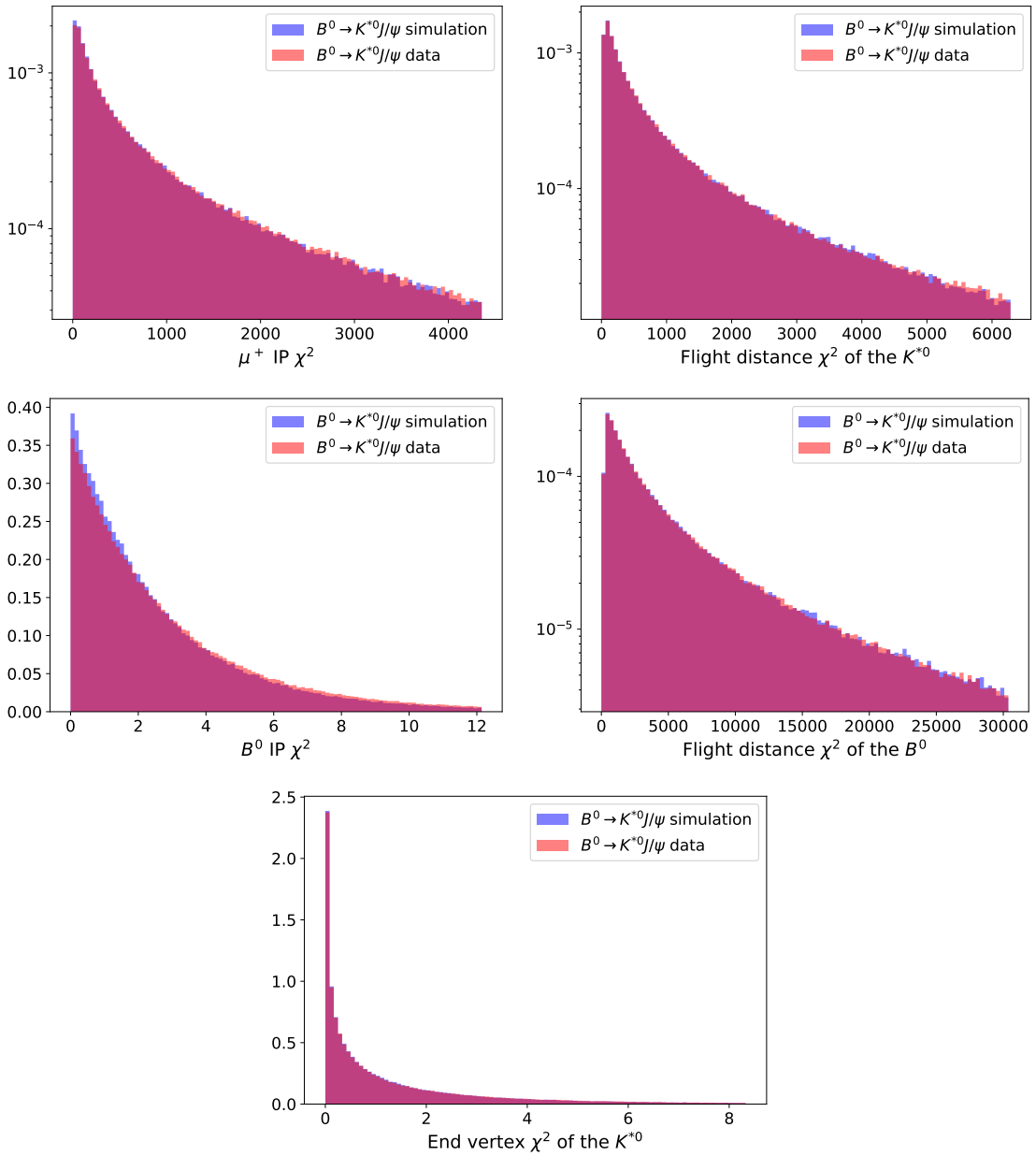


Figure A.3: Comparison between $B^0 \to K^{*0} J/\psi$ simulation and sWeighted $B^0 \to K^{*0} J/\psi$ data for the BDT training variables for Run 1. The distributions are normalised to unit area.

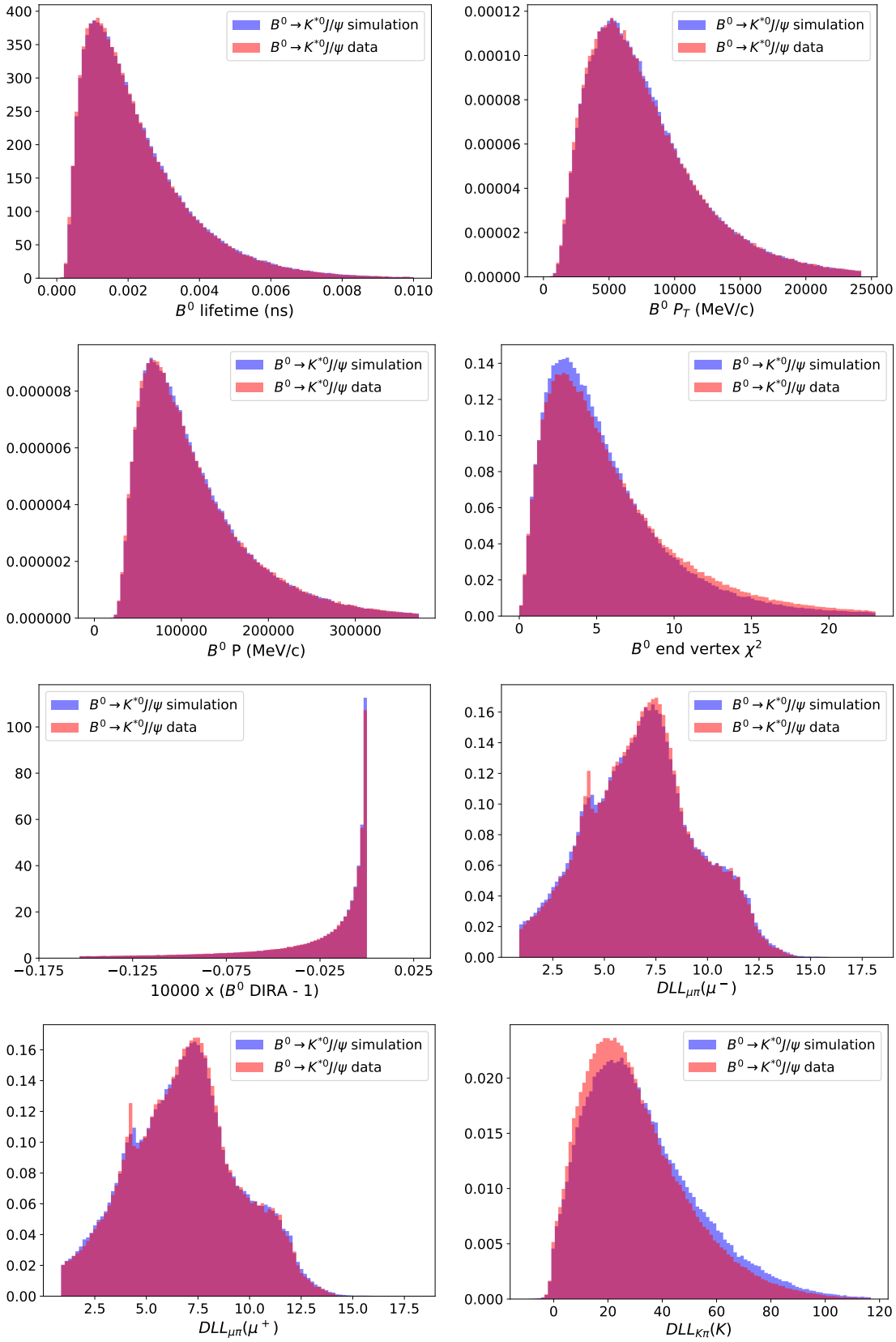


Figure A.4: Comparison between $B^0 \to K^{*0} J/\psi$ simulation and sWeighted $B^0 \to K^{*0} J/\psi$ data for the BDT training variables for 2016. The distributions are normalised to unit area.

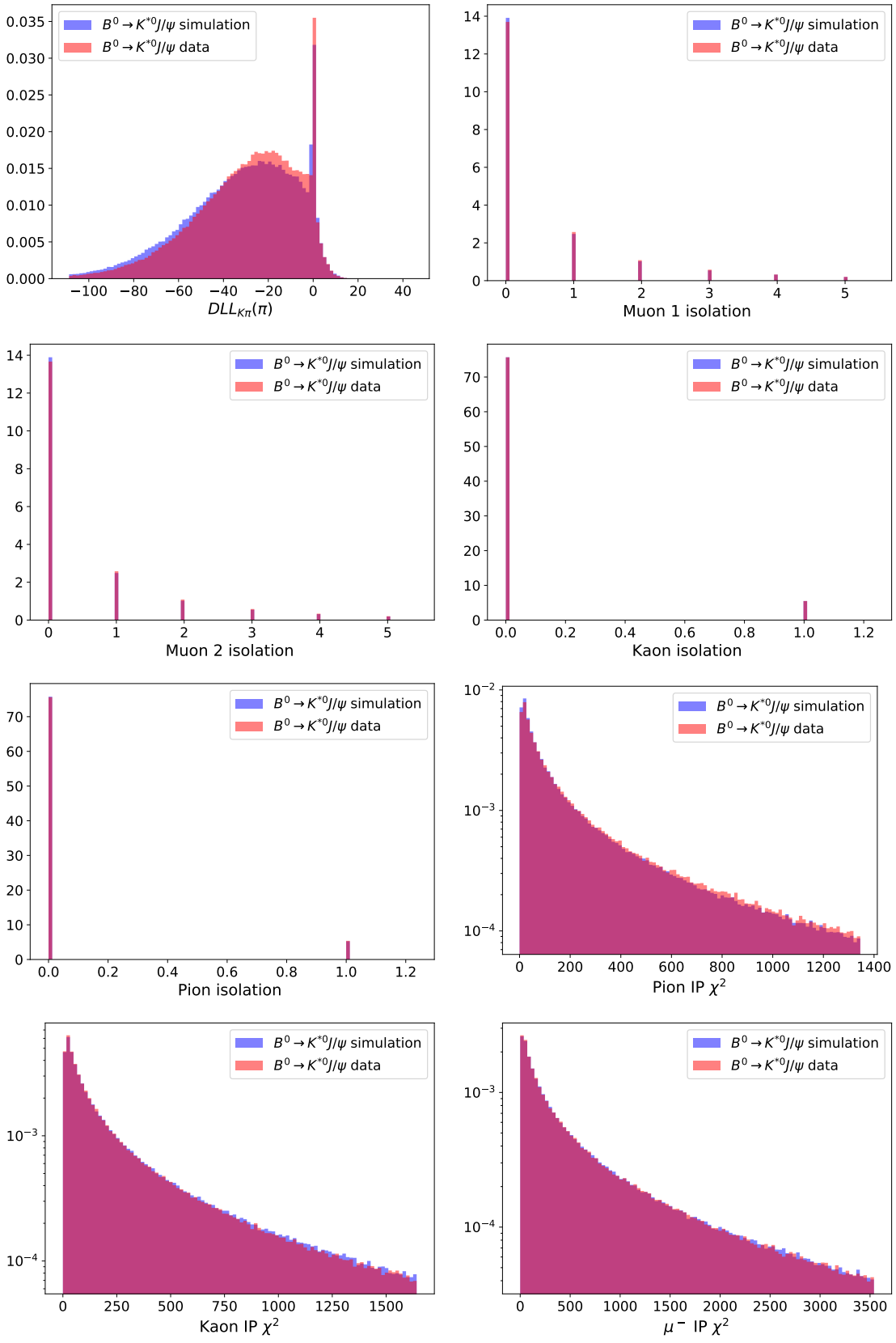


Figure A.5: Comparison between $B^0 \to K^{*0} J/\psi$ simulation and sWeighted $B^0 \to K^{*0} J/\psi$ data for the BDT training variables for 2016. The distributions are normalised to unit area.

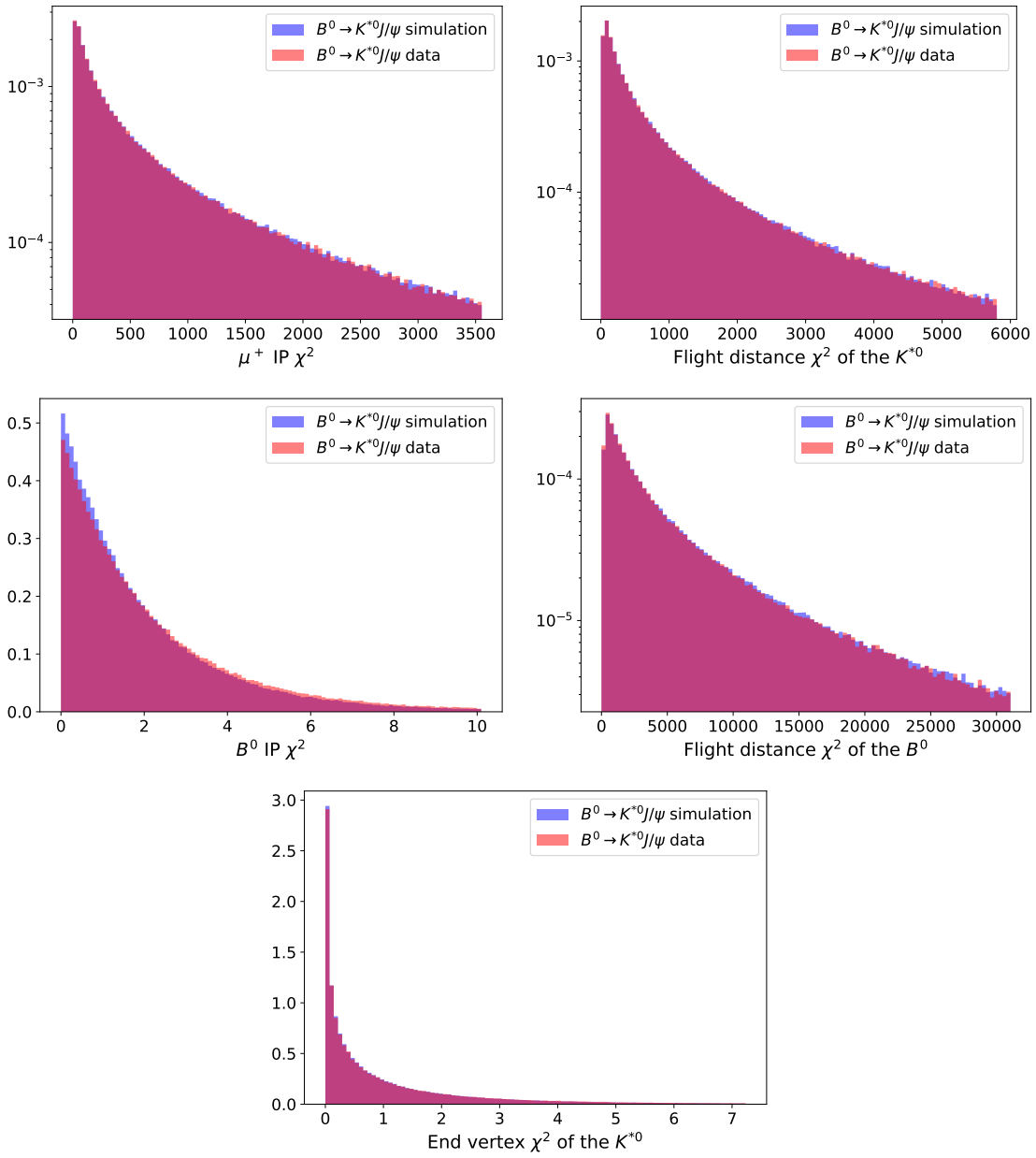


Figure A.6: Comparison between $B^0 \to K^{*0} J/\psi$ simulation and sWeighted $B^0 \to K^{*0} J/\psi$ data for the BDT training variables for 2016. The distributions are normalised to unit area.

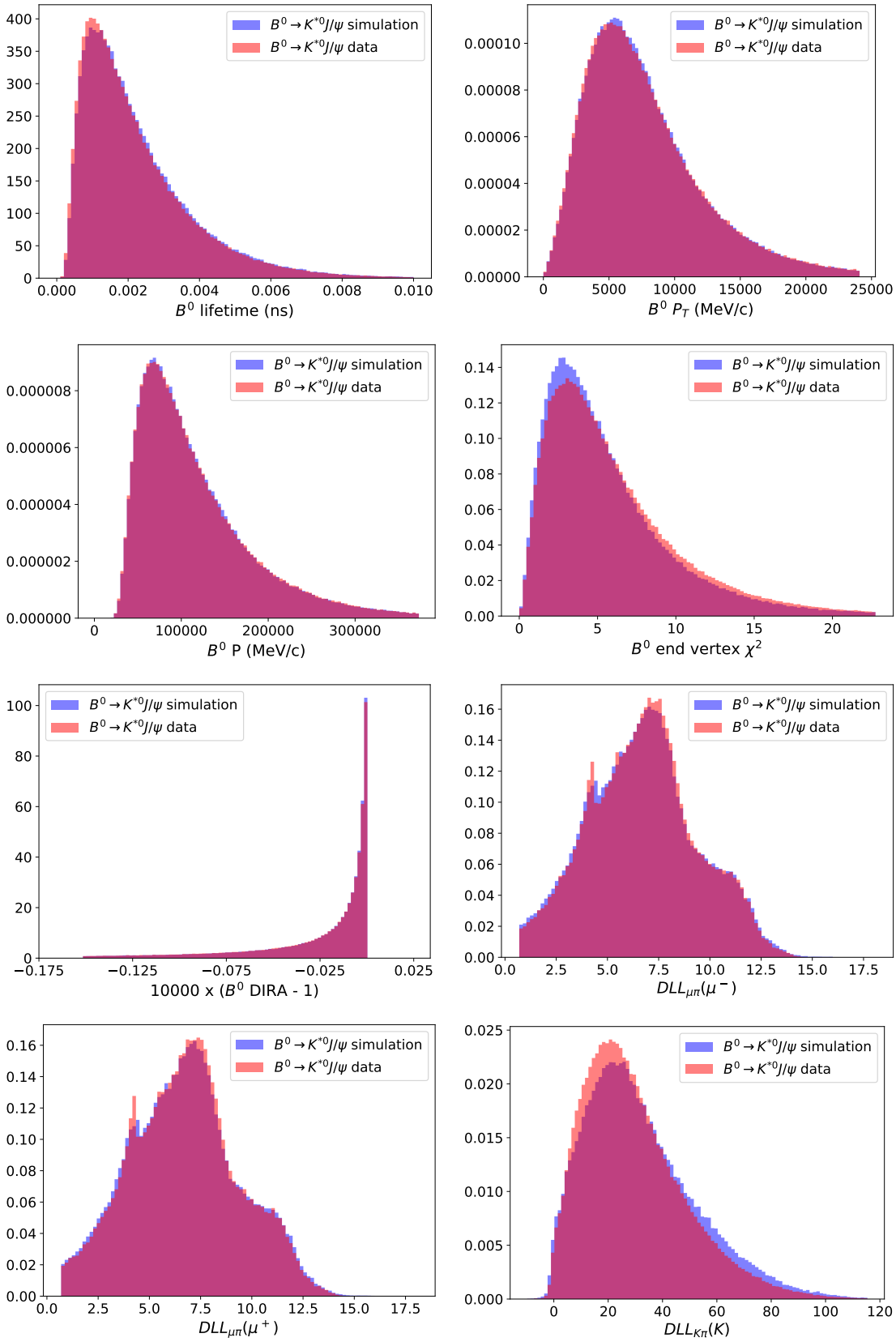


Figure A.7: Comparison between $B^0 \to K^{*0} J/\psi$ simulation and sWeighted $B^0 \to K^{*0} J/\psi$ data for the BDT training variables for 2018. The distributions are normalised to unit area.

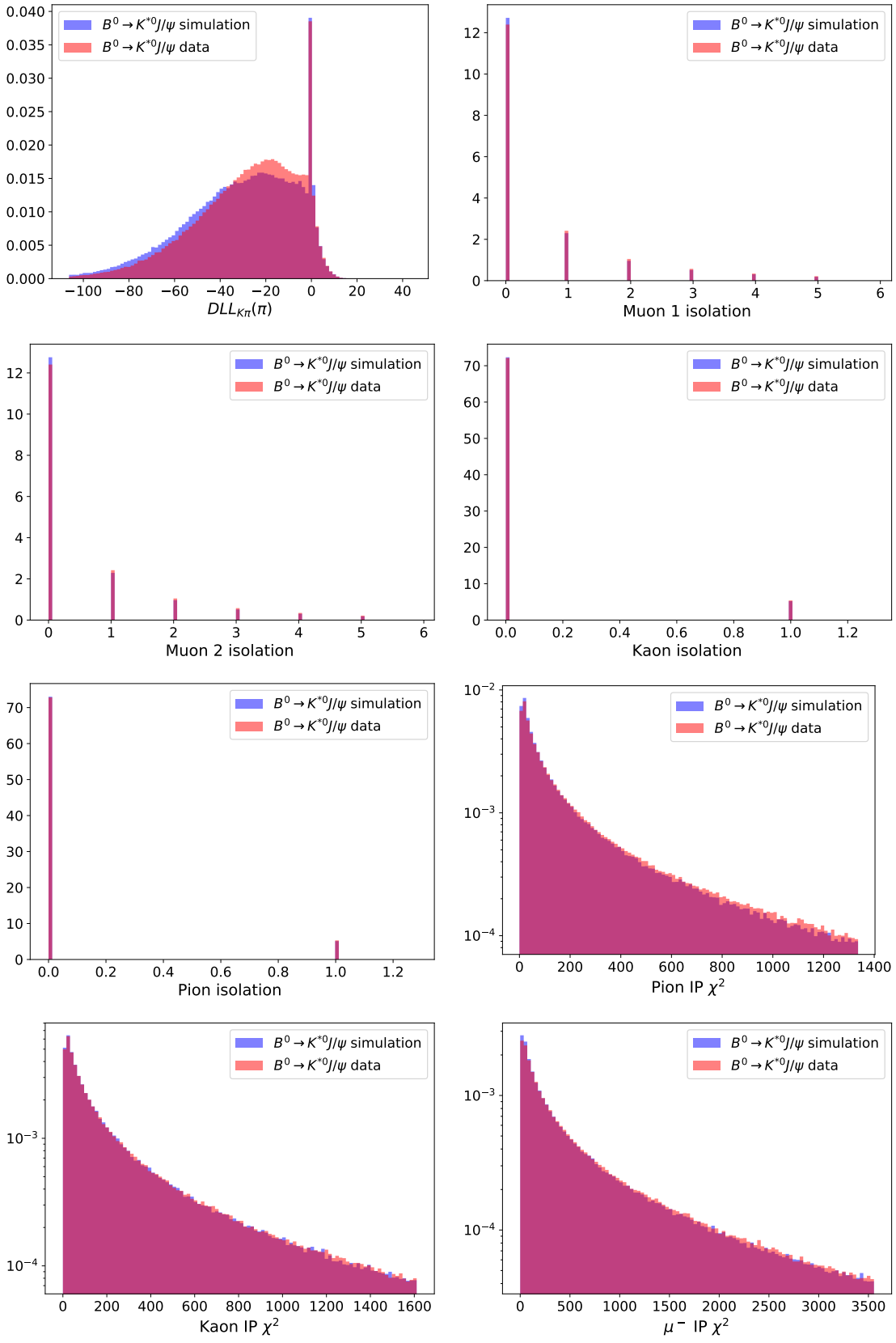


Figure A.8: Comparison between $B^0 \to K^{*0} J/\psi$ simulation and sWeighted $B^0 \to K^{*0} J/\psi$ data for the BDT training variables for 2018. The distributions are normalised to unit area.

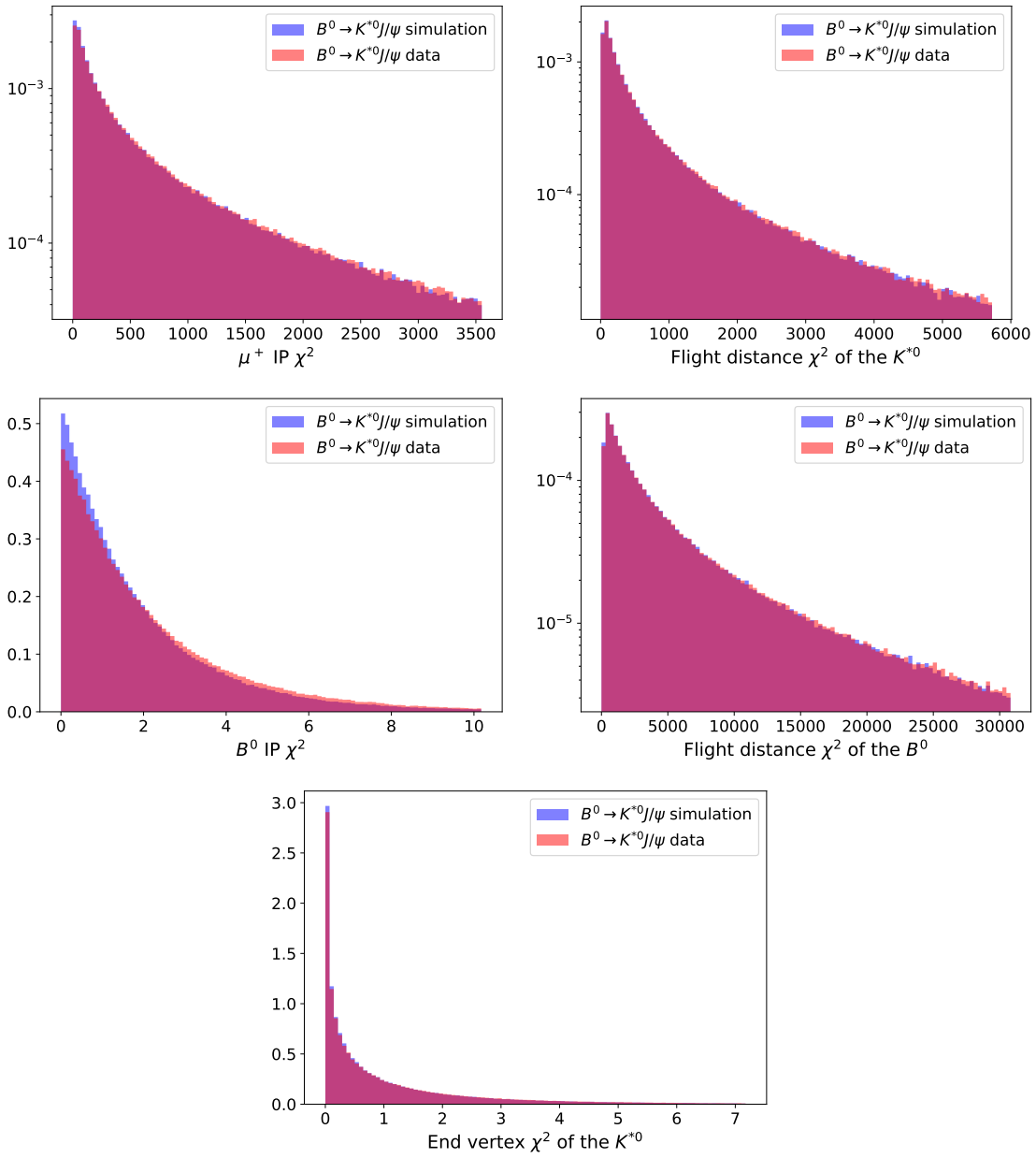


Figure A.9: Comparison between $B^0 \to K^{*0} J/\psi$ simulation and sWeighted $B^0 \to K^{*0} J/\psi$ data for the BDT training variables for 2018. The distributions are normalised to unit area.

Appendix B

Combinatorial BDT efficiency as a function of $\cos\theta_\ell$, $\cos\theta_K$, and ϕ for $B^0 \to K^{*0} J/\psi$ simulation and $B^0 \to K^{*0} J/\psi$ data, in Run 1, 2016, and 2018, in the nominal and narrow $m_{K\pi}$ windows

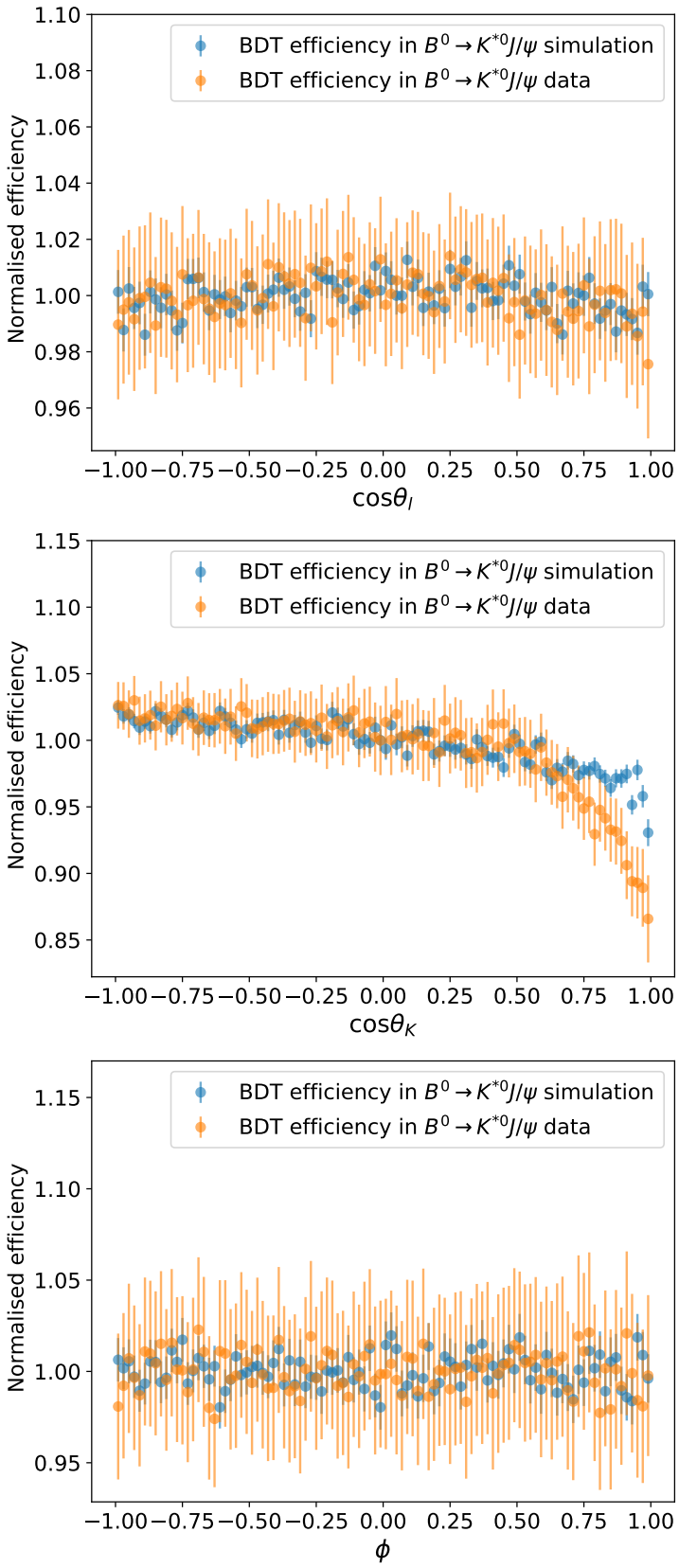


Figure B.1: Combinatorial BDT efficiency as a function of $\cos\theta_\ell$, $\cos\theta_K$, and ϕ for Run 1 $B^0 \to K^{*0} J/\psi$ simulation and Run 1 $B^0 \to K^{*0} J/\psi$ data, where the uncertainties are statistical.

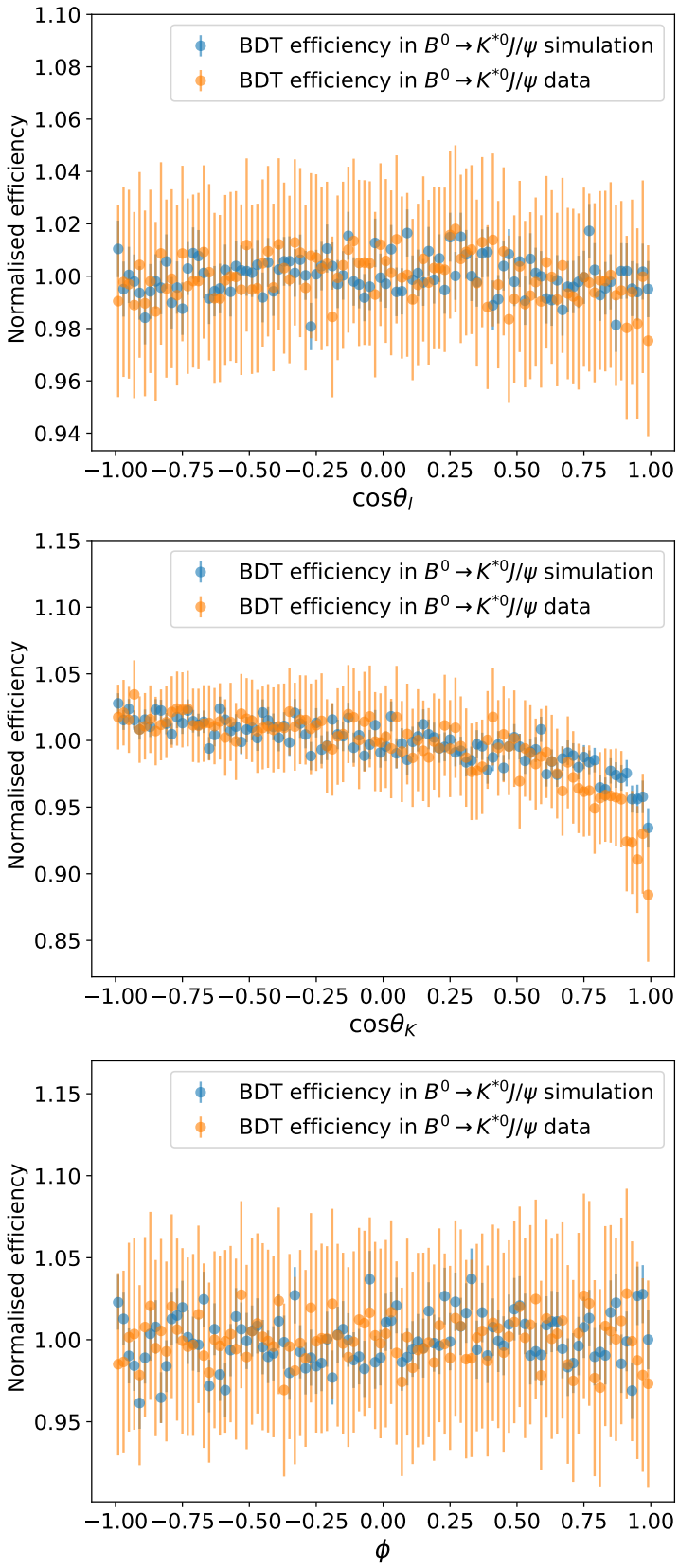


Figure B.2: Combinatorial BDT efficiency as a function of $\cos\theta_\ell$, $\cos\theta_K$, and ϕ for Run 1 $B^0 \to K^{*0} J/\psi$ simulation and Run 1 $B^0 \to K^{*0} J/\psi$ data, in the region $876 < m_{K\pi} < 916$ MeV/c^2, where the uncertainties are statistical. This better aligns the data with the simulated P-wave state.

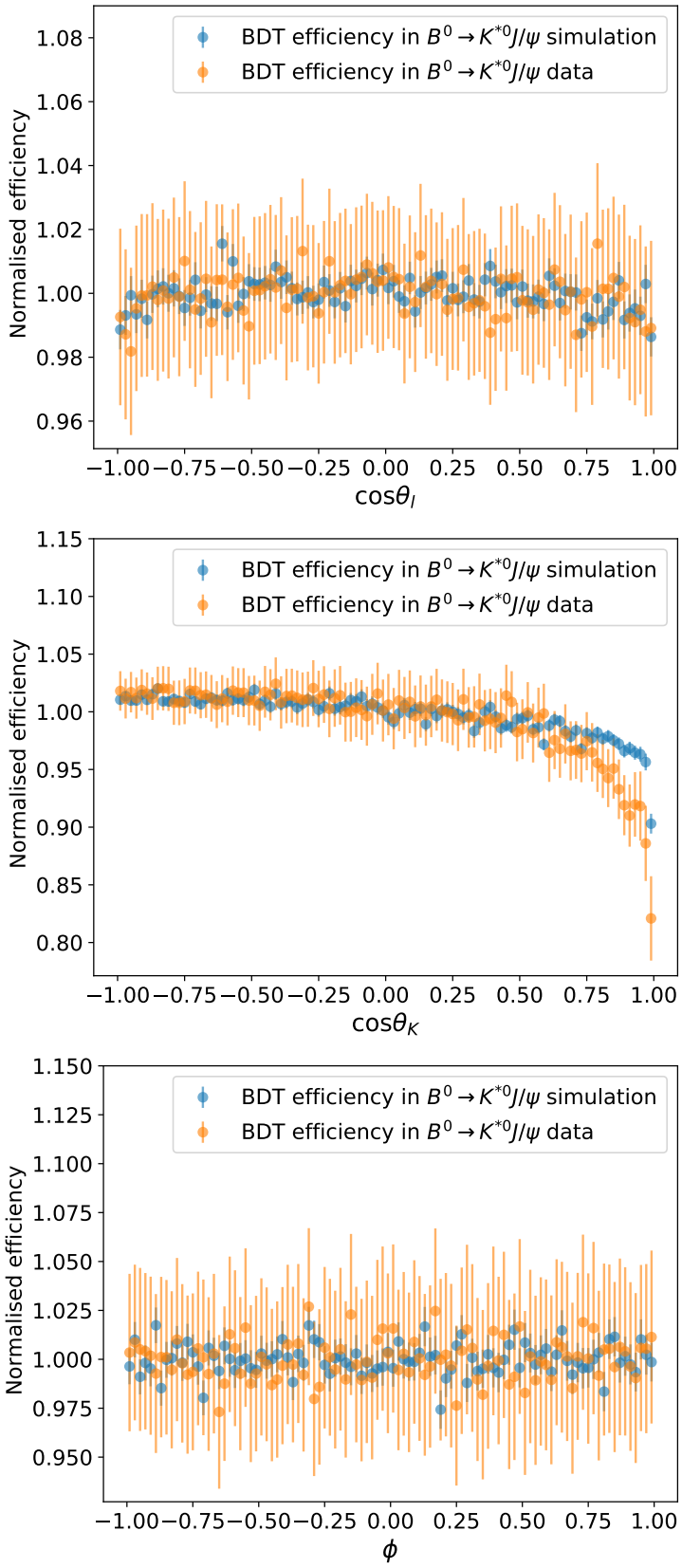


Figure B.3: Combinatorial BDT efficiency as a function of $\cos\theta_\ell$, $\cos\theta_K$, and ϕ for 2016 $B^0 \to K^{*0} J/\psi$ simulation and 2016 $B^0 \to K^{*0} J/\psi$ data, where the uncertainties are statistical.

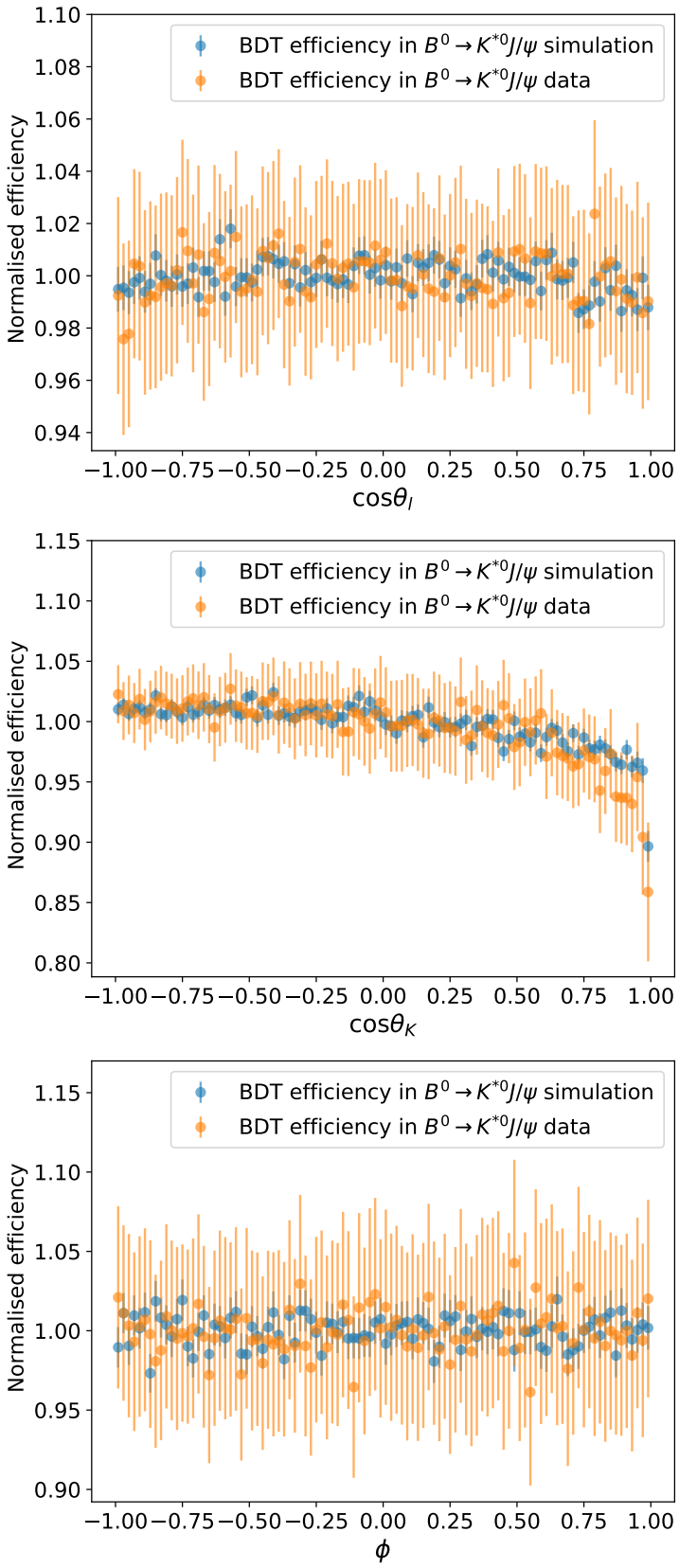


Figure B.4: Combinatorial BDT efficiency as a function of $\cos\theta_\ell$, $\cos\theta_K$, and ϕ for 2016 $B^0 \to K^{*0} J/\psi$ simulation and 2016 $B^0 \to K^{*0} J/\psi$ data, in the region $876 < m_{K\pi} < 916$ MeV/c^2, where the uncertainties are statistical. This better aligns the data with the simulated P-wave state.

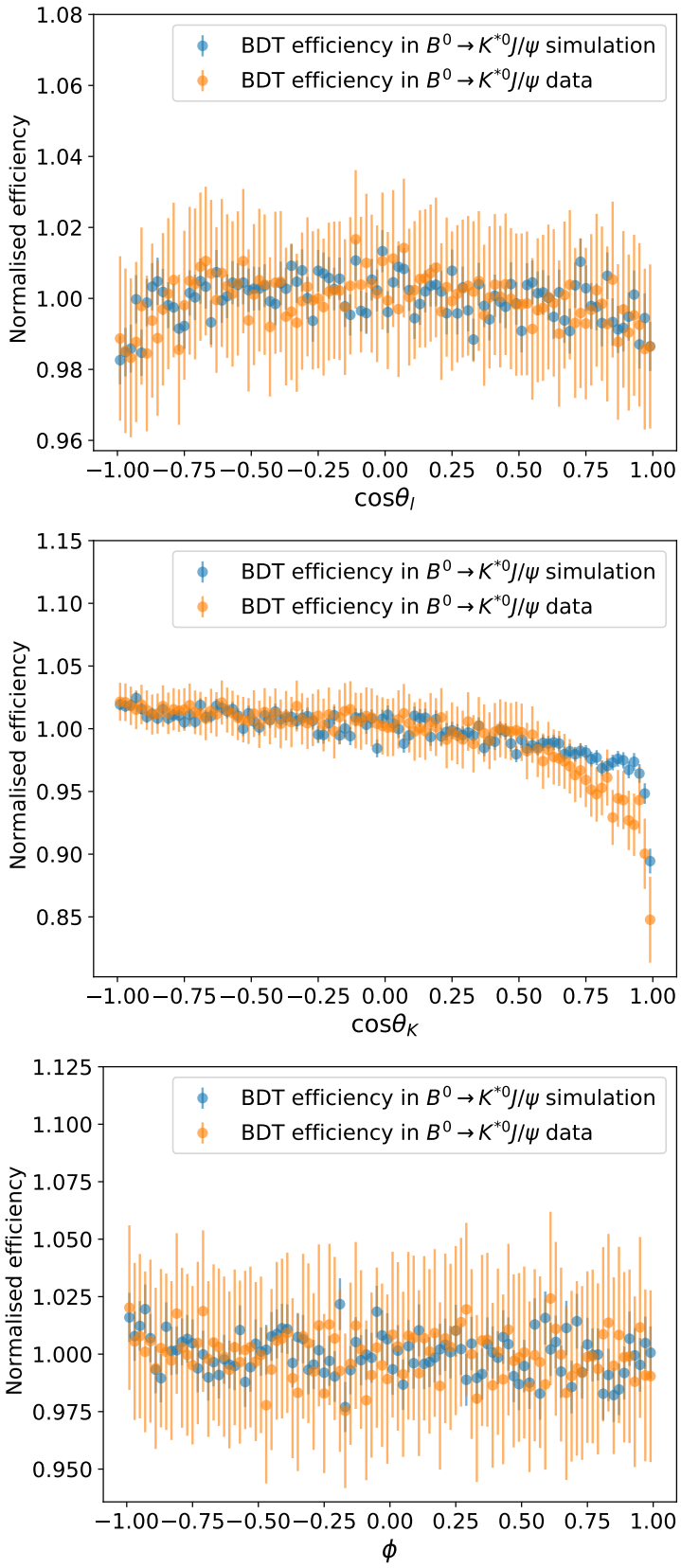


Figure B.5: Combinatorial BDT efficiency as a function of $\cos\theta_\ell$, $\cos\theta_K$, and ϕ for 2018 $B^0 \to K^{*0} J/\psi$ simulation and 2018 $B^0 \to K^{*0} J/\psi$ data, where the uncertainties are statistical.

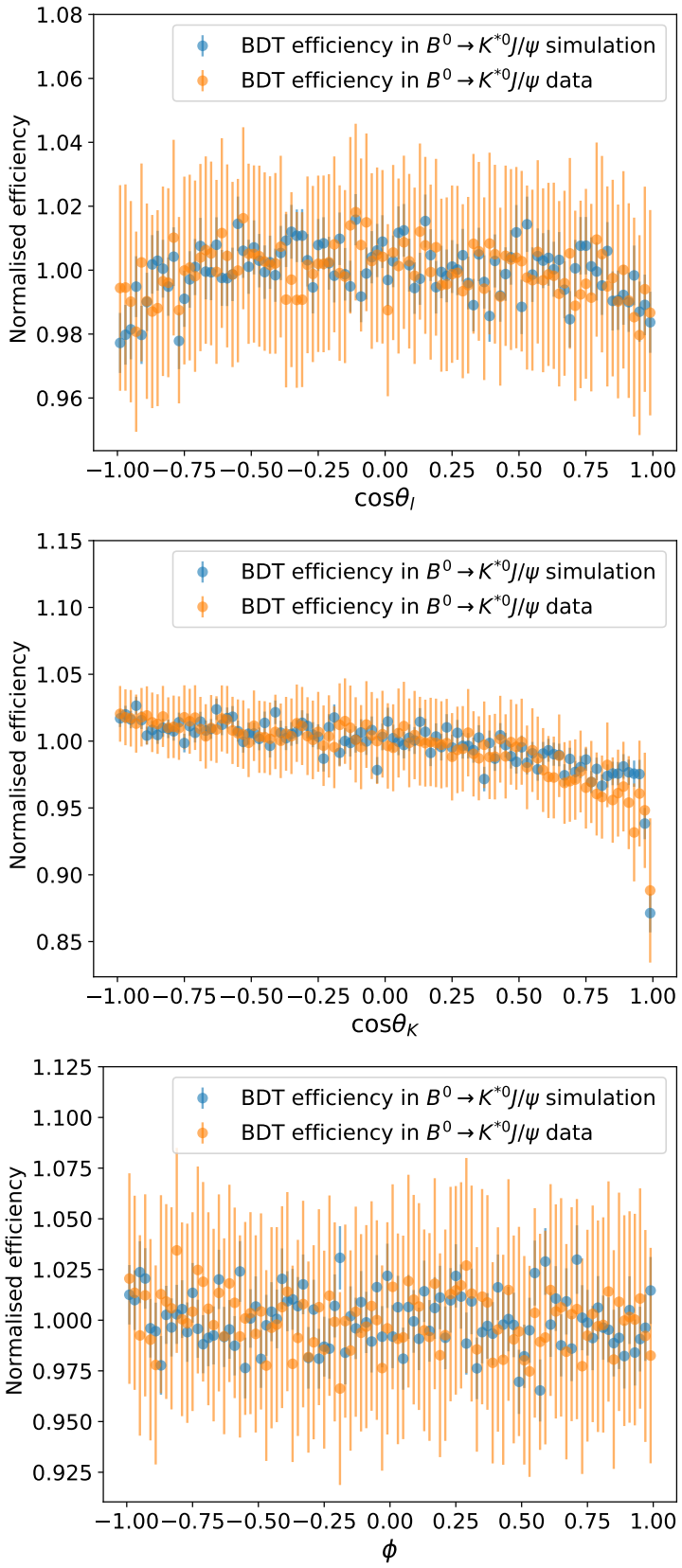


Figure B.6: Combinatorial BDT efficiency as a function of $\cos\theta_\ell$, $\cos\theta_K$, and ϕ for 2018 $B^0 \to K^{*0} J/\psi$ simulation and 2018 $B^0 \to K^{*0} J/\psi$ data, in the region $876 < m_{K\pi} < 916$ MeV/c^2, where the uncertainties are statistical. This better aligns the data with the simulated P-wave state.