The search for $B_c^+ \to \pi^+\mu^+\mu^-$ and updating the measurement $\mathcal{B}(B_c^+ \to \psi(2S)\pi^+)/\mathcal{B}(B_c^+ \to J/\psi\,\pi^+)$ with the LHCb experiment

by

Vedanshu Mahajan

Thesis

Submitted to the University of Warwick

for the degree of Masters (by Research) in Physics

Master of Science

Department of Physics

December 2023

CERN-THESIS-2023-410
10/04/2024

Contents

Acknowledgments

First, I would like to thank Tim Gershon for believing in me and giving me the fantastic opportunity to work with Warwick LHCb - I wouldn't have met such an amazing group of brilliant physicists who rekindled a brighter love for my subject without him. Next, conversations with Tom Blake provided me with many of the insights into academia and research that form the foundation of my career aspirations for the future too.

The biggest mention has to be to my day-to-day supervisor Fernando Abudinén, who's fierce but steady guidance (and lots of tools from previous analyses to adapt from!) kept me on track and achieving what I describe in this thesis.

Additionally, I need to make known the fact that I could not have done this without the support of George Sotiropoulos and Matthew Jeffries who have been close counsel and my rocks to lean on through this. Calling George on an almost daily basis and hearing what he has been up to in his own amazing ventures kept me humble and looking forward to doing my own job to the best of my abilities. Having less frequency of correspondence with Matt however, did not change his impact to the motivation he gave me, as whenever we would talk, we would have so much to catch up on and his determination in his own research was most definitely contagious.

Finally, a heartfelt thank you to every friend I brought with me and every new friend I have made at Warwick (especially those in my office) for making the time I spent here a time I will not forget.

Declarations

This thesis is the result of my own work, except where direct and explicit reference has been made to the work of others. The work described in Sec. 3.2.1, 3.2.3, 3.2.4, Tables 3.4, and E.1 was completed by Dr Fernando Abudinén and summer student Callum White. Fernando was also involved in managing, requesting and correcting the methods and simulated data described in Sec. 2.3, 3.1, Tables A.1, A.2, and Appendix B however everything mentioned has been included for completeness' sake. This thesis has not been submitted for any other awards, degrees, or qualifications outside of the submission made here.

Vedanshu Mahajan

Abstract

We perform a search for non-resonant $B_c^+ \to \pi^+\mu^+\mu^-$ decays. We reconstruct $B_c^+ \to \pi^+\mu^+\mu^-$ decays and use the invariant mass of the B_c^+ candidates as discriminating observable to search for a signal using extended, unbinned maximum likelihood fit in dimuon mass. The signal yield, normalized to the yield of reconstructed $B_c^+ \to J/\psi\,(\mu^+\mu^-)\,\pi^+$ decays, is used to obtain a result on the relative branching fractions. As no significant signal for non-resonant decays is seen, we provide a limit for the ratio of branching fractions, $\mathcal{B}(B_c^+ \to \pi^+\mu^+\mu^-)/\mathcal{B}(B_c^+ \to J/\psi\,\pi^+)$. We also provide an update on the ratio of branching fractions $\mathcal{B}(B_c^+ \to \psi(2S)\pi^+)/\mathcal{B}(B_c^+ \to J/\psi\,\pi^+)$.

Chapter 1

Overview of Standard Model

1.1 The Standard Model

The Standard Model (SM) is what the particle zoo of the 20^{th} century culminated in. The SM is derived from the mathematical framework of quantum field theory (QFT) as a collection of related gauge theories [1][1]. It is the most tested model of particles in existence so far.

1.1.1 Particles

The SM describes the picture of all particles as fermions (particles with spin half-integer) and bosons (particles with integer spin). Bosons have the property that they can occupy the same state while fermions obey the Pauli exclusion principle (that each particle must have a unique state) [2].

Table 1.1 shows the particles contained in the SM with the most up-to-date measurement of their properties available from the Particle Data Group (PDG) [3].

[1] Gauge theories are QFTs that satisfy Local Gauge Invariance. This is when the theory produces the same observable results when the field is invariant under some local transformations. The SM combines Quantum Electrodynamics, Electroweak Unification, and Quantum Chromodynamics put together, however deep understanding/explanation of these topics is not necessary for the purpose of this thesis, which is to describe the search for rare decays at the LHCb.

Table 1.1: The Standard Model Particles Properties, with values taken from PDG [3] as of 2023.

Particle	Mass (GeV/c^2)	Charge (e)	Spin
Quarks			
Up (u)	$2.16^{+0.49}_{-0.26} \times 10^{-3}$	$+2/3$	1/2
Down (d)	$4.67^{+0.48}_{-0.17} \times 10^{-3}$	$-1/3$	1/2
Charm (c)	1.27 ± 0.02	$+2/3$	1/2
Strange (s)	$93.4^{+8.6}_{-3.4} \times 10^{-3}$	$-1/3$	1/2
Top/Truth (t)	172.69 ± 0.30	$+2/3$	1/2
Bottom/Beauty (b)	$4.18^{+0.3}_{-0.2}$	$-1/3$	1/2
Leptons			
Electron (e)	$0.51099895000 \pm 0.00000000015 \times 10^{-3}$	-1	1/2
Electron Neutrino (ν_e)	$< 0.8 \times 10^{-9}$ (estimated)	0	1/2
Muon (μ)	$105.6583755 \pm 0.0000023 \times 10^{-3}$	-1	1/2
Muon Neutrino (ν_μ)	$< 0.19 \times 10^{-3}$ (estimated)	0	1/2
Tau (τ)	1.77686 ± 0.00012	-1	1/2
Tau Neutrino (ν_τ)	$< 18.2 \times 10^{-3}$ (estimated)	0	1/2
Gauge Bosons			
Photon (γ)	0	0	1
W Boson ($W^{\pm}$)	80.377 ± 0.012	± 1	1
Z Boson (Z^0)	91.1876 ± 0.0021	0	1
Gluon (g)	0	0	1
Higgs Boson			
Higgs (H)	125.25 ± 0.17	0	0

As table 1.2 shows, the description of fermions gives way to a structure of generations with each succession having increased mass of same type of charged particles than previous ones. It is still currently an unanswered question as to why we observe only three generations of particles.

Generation	1st	2nd	3rd
Up-style quarks	u	c	t
Down-style quarks	d	s	b
Leptons	e	μ	τ
Lepton Neutrinos	ν_e	ν_μ	ν_τ

Table 1.2: Particles sorted by generation.

Quarks are found in hadron states as mesons (states with quark and anti-quark and integer/scalar spin) and baryons (states with 3 quarks and half integer spin) which all fall under the term: hadrons[2]. Below, in figure 1.1, we show how particles are classified based on their spin states or properties.

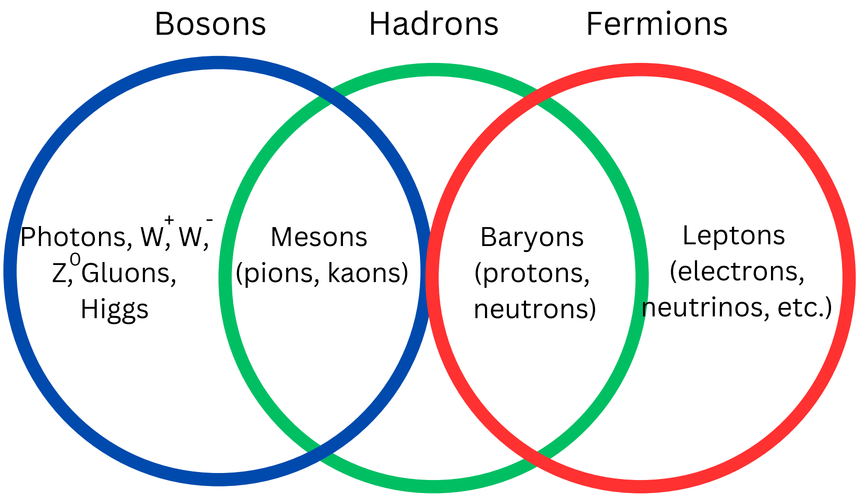


Figure 1.1: Venn diagram showing the classifications of particles.

[2]However, studies have found evidence for tetra-quark and penta-quark states [4]. The current scientific open question on these > 3 quark states is the structure on whether they are like overlapping mesons+baryons or 4/5 quarks in the same (assumed) spherical bound.

1.1.2 Fundamental Forces

To our current knowledge, reality consists of four forces only: gravitational, electromagnetic, weak, and the strong force. Gravitational forces cause attraction between mass/energy in the macro-scale however a quantum mechanical picture of gravitation is still yet to be formed. In the SM, forces are mediated by bosons and sometimes referred to as 'virtual force mediating particles' (or some variation of such).

The electromagnetic force is felt between all particles with an electric charge. It can be attractive or repulsive and is described by Quantum Electrodynamics (QED). QED combines our knowledge of relativistic quantum mechanics with electromagnetism and describes the interactions of all charged particles by the exchange of photons.

The weak force is generally responsible for particle decays [3], and the weak nuclear interaction acts only on left handed quarks and leptons or right handed antiquarks or antileptons (right handed particles have spin pointing in the direction of motion while left handed particles have spin pointing in the opposite direction). This is due to the fact that the W and Z bosons couple differently to the chirality of the fermions due to the V-A (vector minus axial vector) theory [5][4]. Interestingly, the weak force is the only force that is known to violate Charge-Parity (CP) Symmetry (later described in section 1.2.1).

The weak force originates from the QFT of Electroweak Theory, that combines the weak with the electromagnetic force. This stems from the early universe cosmology (the Electroweak Epoch), where it was so hot, the electromagnetic and the weak force were indistinguishable, and upon 'condensation' (where electroweak symmetry breaks) separated into the two forces we see today.

Finally, the strong force from QCD is responsible for holding the nuclei of atoms together and trapping quarks in bound states. Interactions of quarks are mediated by gluons due to an inherent charge-like property of quarks and gluons, known as colour, that can have three values (six including anti-values): red, blue, green.

[3] This can also occur due by EM or strong force.

[4] The vector (V) and axial (A) interaction separately couple particles' spin and momentum in a manner that is independent of chirality. However, the V−A or V+A interaction couples only to left-handed or right-handed chiral particles respectively.

1.2 Particle Decays

From Einstein's famous $E = mc^2$, and relativistic kinematics, we know that the energy from $p+p$ collisions (conventionally also pp or p–p) can be converted into masses of particles which can be heavy, exotic, and previously unobserved. These then decay into more stable states. The B_c^+ ($\bar{b}c$) meson, for example, can decay into many different final states, also known as decay modes. The total decay rate is the sum of the decay rates of each individual decay rate for each mode:

$$\Gamma = \Sigma\Gamma_i. \tag{1.1}$$

Then, the lifetime of a particle is the inverse of its decay rate,

$$\tau = \frac{1}{\Gamma}. \tag{1.2}$$

The probability of a particle going down a specific decay path is known as the branching fraction of that decay path,

$$\mathcal{B} = \frac{\Gamma_i}{\Gamma}. \tag{1.3}$$

For this thesis, we are aiming to measure a relative branching fraction (i.e. a ratio of two branching fractions) and updating another channel's relative branching fraction. Relative branching fractions are useful to measure as they do not require detailed knowledge of how many B_c^+ mesons are produced within the experiment.

An ideal normalisation channel should have similar decay topology to the signal, as it minimises other potential sources of systematic uncertainties. Later, we see that this study satisfies this convention (see Sec. 1.3.2).

1.2.1 Feynman Diagrams as Matrix Elements

A number of properties are conserved in every particle interaction including charge, baryon number, lepton number, mass-energy, and momentum.

Feynman diagrams can be used to not only represent decays diagrammatically but also calculate decay rate coefficients. They are conventionally drawn with time going from left to right with fermions as solid lines and the arrows indicating the direction of time they are travelling in. Antiparticles are drawn as if they are the same as their regular counterpart but travelling backwards in time. Bosons are drawn with a dashed line except for gluons that drawn with loops, and photons and Z bosons by a waved line. Examples are shown in Fig. 1.2.

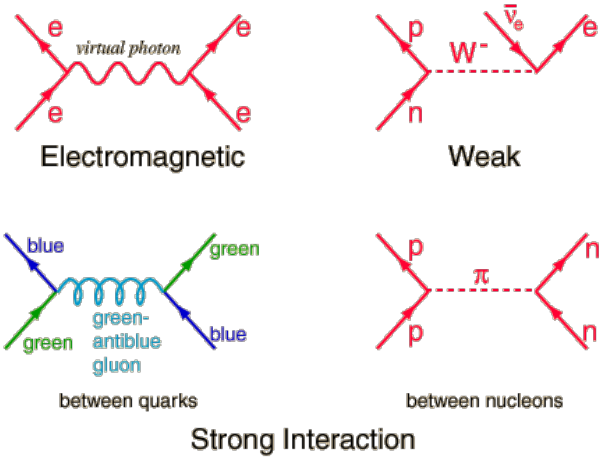


Figure 1.2: Example Feynman diagrams for each fundamental force of the SM. Image sourced from http://hyperphysics.phy-astr.gsu.edu/hbase/Particles/expar.html.

The factors associated to the vertices of each Feynman diagram (such as incoming and outgoing particle lines, propagators, etc.) can also help us calculate expected transition rate. The total decay rate for a given interaction is given by Fermi's golden rule [2] as follows,

$$\Gamma_{fi} = 2\pi|M_1 + M_2 + M_3 + ...|^2\rho(E), \tag{1.4}$$

where Γ_{fi} is the transition rate from initial states to final states of the quantum system, M_i is the amplitude coefficient of a specific Feynman diagram, and $\rho(E)$ is the energy density of states. The amplitude coefficient for each photon propagator is given as $1/q^2$ while each vertex with fermions contribute a factor of Qe, where Q is the electric charge, and e is the charge of an electron. This means that for an EM interaction with two vertices, we can show that the amplitude coefficient is given by

$$M_{EM} = \frac{4\pi\alpha}{q^2}, \tag{1.5}$$

where α is the electromagnetic parameter $e^2/4\pi$ ($\approx 1/137$), and q^2 is the invariant energy of the photon.

Similarly, the contribution from strong pairs of vertices is:

$$M_S = \alpha_S = \frac{g_s^2}{4\pi q^2} \approx 1, \tag{1.6}$$

where g_s is the coupling of the gluon. Finally, each electroweak vertex would introduce a factor of g_W or $g_W V_{CKM}$ to our Feynman diagram if the interaction is leptonic or quark-flavour changing respectively. The Cabibbo-Kobayashi-Maskawa matrix is given as

$$V_{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \approx \begin{pmatrix} 0.975 & 0.220 & 0.01 \\ -0.220 & 0.975 & 0.05 \\ 0.01 & -0.05 & 1 \end{pmatrix} \tag{1.7}$$

where, while the values on the left are complex, the magnitudes are provided on the right.

Weak interactions between quarks of the same family are 'Cabibbo allowed', while quarks differing by one family are considered 'Cabibbo suppressed'. Interactions of quarks differing by two families are considered 'doubly Cabibbo suppressed'. This impacts the expected number of events in some of the backgrounds described in Sec 4.2 and our signal non-resonant decay.

The weak force allows for quark flavour changes and for charge and parity violations.

Charge-Parity Violation

Charge-Parity (CP) violation refers to the breaking of the universal symmetry of charge conjugation and parity transformations. According to CPT symmetry theorem (that is Charge, Parity, and Time), symmetry under the combined C, P, and time reversal (T) operations must be preserved[5].

While CP-violation is not a priority in this thesis, it provides useful context to why rare decays are interesting as explained later in section 1.3. As explained in subsection 1.1.2, in our hot, early universe (quark epoch) we would expect the process of $p + \bar{p} \leftrightarrow \gamma + \gamma$ to occur regularly. This forward/backward decay does not break any conservation rules. However as the universe cooled, the equilibrium shifted to the left by some degree (while photons still outnumber baryons significantly) - yet we see much more p than $\bar{p}$. This is due to CP violation. The mechanism for this phenomenon was proposed by Kobayashi and Maskawa as it provided explanations for the observations of CP violation in the mixing of neutral kaons [6][6] (first indirect CP violation measurement by Christenson, Cronin, Fitch, and Turlay) and their decays (direct measurement of CP violation, which came much later).

The three conditions proposed by Sakharov [7], state that, to provide the 1-billion anti-baryons to 1-billion-and-1 baryons ratio that explains observations in

[5] As long as certain assumptions are met.

[6] The assumed initial and final states have different values of CP, hence suggesting CP violation.

today's universe:

- Baryon number is not conserved.
- C and CP symmetries are violated.
- The (early) universe is no longer in thermal equilibrium.

The LHCb experiment focuses on the second point above.

1.3 Rare Decays

To understand the fundamentals of particle physics, rare decay measurements at the LHCb experiment are crucial as they offer the opportunity to explore new physics and test predictions of the SM.

Rare decays often contain loop processes or flavour changing neutral currents which are largely suppressed by the SM [2]. This means that they are particularly sensitive to effects Beyond the Standard Model (BSM). Precise measurements of the branching ratios can allow us to compare nature with theory, with significant discrepancies indicating potentially new physics.

1.3.1 Beauty Quarks as Probes into BSM

The LHCb experiment focuses on studying particles that contain the b quark. Having high rest mass means that they are more sensitive to new particles and the interactions that can arise at higher energy scales. Rare decays involving b-hadrons provide insight into CP violation, flavour physics, and mixing:

- CP violation is explained through the CKM mechanism however additional sources of CP violation would be a flag for BSM physics.
- Flavour physics aims to understand the physical properties and behaviours of quarks and leptons. Inconsistencies between theoretical predictions from the SM and nature could indicate the presence of new particles or physics.
- Neutral B mesons can mix and oscillate between their matter and antimatter states. Transitions between B and $\bar{B}$ mesons could be sensitive to new physics effects and precise measurements of B mixing parameters (frequency, phase, etc.) can be a location of deviations from the SM and constrains on new physics models.

1.3.2 B_c^+ Mesons

The B_c^+ mesons are the heaviest ground state mesons that form in the SM. Made of the beauty and charm quarks, they weigh at 6274.47 ± 0.32 MeV [3]. Ground state B_c^+ are known to decay in various ways as each quark can individually decay while the other spectates, or the $\bar{b}$ and c can annihilate into a virtual W boson. Having two heavy quarks in the meson has particular effects on the theoretical calculations of B_c^+ decay properties, however for this search for $B_c^+ \to \pi^+\mu^+\mu^-$, there are no published predictions at the time of writing.

Since the discovery of B_c^+ mesons at the CDF experiment [8], several B_c^+-decay modes have been observed [3]. The first observation of the $B_c^+ \to J/\psi\,\pi^+$ decay was performed at the CDF experiment [9]. Since then, this channel has been used to measure the lifetime [10, 11], the mass [12] and the production rate of B_c^+ mesons [13, 14].

The LHCb experiment has observed several new B_c^+-decay modes which include the first observation of the $B_c^+ \to \psi(2S)\pi^+$ decay. The J/ψ and $\psi(2S)$ mesons are spin-1 bound states formed of a $c\bar{c}$ quark anti-quark pair, where the former is the ground state, and the latter the first excited state.

The first observation of the $B_c^+ \to \psi(2S)\pi^+$ decay was made at LHCb using $1\,\mathrm{fb}^{-1}$of Run I data collected at $\sqrt{s} = 7$ TeV [15]. An update of this measurement [16] exploited the full Run I data sample resulting in the currently most precise measurement of the ratio

$$R_{\psi(2S)/J/\psi} \equiv \frac{\mathcal{B}(B_c^+ \to \psi(2S)\pi^+)}{\mathcal{B}(B_c^+ \to J/\psi\,\pi^+)} = 0.268 \pm 0.032\,(\mathrm{stat}) \pm 0.007\,(\mathrm{syst}) \pm 0.006\,(\mathrm{BF})\,,$$

where the third uncertainty is due to limited knowledge of the branching fractions of the $J/\psi \to \mu^+\mu^-$ and $\psi(2S) \to \mu^+\mu^-$ decays.

The non-resonant $B_c^+ \to \pi^+\mu^+\mu^-$ decay has never been observed, nor has a dedicated search for it been made. In the SM, this decay can happen through the annihilation of the $\bar{b}$ and c quarks into a virtual W boson (that decays into a charged pion) and the radiation of a virtual photon or Z boson (that decays into a muon pair). An example diagram for this decay is shown in Fig. 1.3; similar diagrams where the γ/Z is radiated from one of the initial or final state quarks also contribute to the total decay rate.

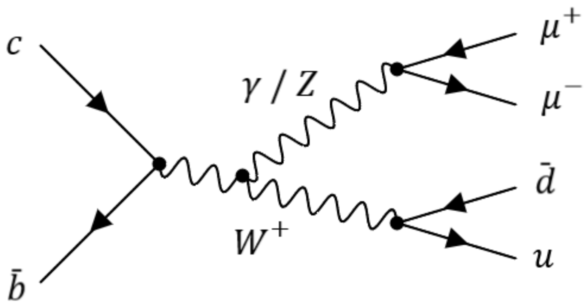


Figure 1.3: Example Feynman diagram for the non-resonant $B_c^+ \to \pi^+\mu^+\mu^-$ decay. The γ/Z particle can couple with any of the incoming or outgoing quarks.

Chapter 2

The LHCb Detector

2.1 CERN and the LHC

The European organization for Nuclear Research (CERN) hosts the largest particle accelerator in the world, known as the Large Hadron Collider (LHC). Using the subterranean 27 km tunnel that previously held the Large Electron Positron (LEP) collider, LHC has been operating since 2010 collecting pp and heavy-ion collision data, with the aim of probing the SM for new particles, testing theory, and searching for BSM physics.

There are four major experiments located at four particle collision points which are known as (alphabetically), ALICE [17], ATLAS [18], CMS [19], and LHCb [20]. The LHC focuses on accelerating particles to very high energies and when such speeds are reached, deflected into one of the several set collision points.

Figure 2.1 shows the CERN complex and not-to-scale locations of the various experiments and facilities.

The ALICE experiment is focused on heavy ion collisions to study unique states of matter while ATLAS, CMS, and LHCb focus on $p-p$ collisions with aims of impact into the current particle physics model, rare decays, charge-parity mixing/violations etc. ATLAS/CMS distinguish themselves from the LHCb experiment as they intend to study high p_T signatures such as those of electroweak particles (including the Higgs boson) or new candidates that the LHC could potentially produce. The LHCb detector is primarily for the study of decay products from hadrons containing b and/or c quarks, which are close to the beamline (relatively low p_T). Over time the center of mass energies for the accelerated particles have increased run by run of the LHC, from 8 TeV in run I (2010-2012), 13 TeV in run II (2015-2018), and now 13.6 TeV for run III.

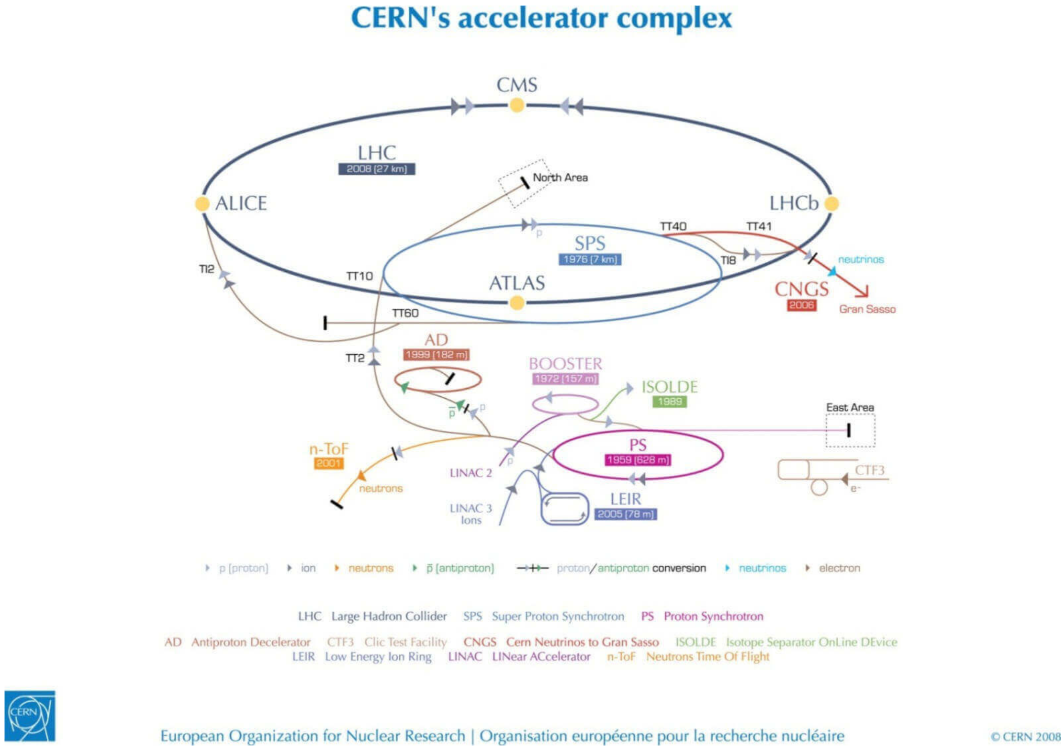


Figure 2.1: Map showing the CERN accelerator complex. Image sourced from the European Organization for Nuclear Research (CERN) 2008.

2.2 The LHCb Detector

The LHCb detector [20, 21] is a single-arm forward spectrometer designed to study particles containing b or c quarks with the intention to understand the asymmetry between matter and antimatter along with testing the SM with measurements of rare decays. The detector is additionally being used to pursue many more objectives such as heavy ion physics or lepton flavour universality violation.

As production of hadrons with b or c quarks is boosted into the forward and backward directions due to their large masses, the forward arm spectrometer can exploit that property covering the pseudo-rapidity ranges of $2 < \eta < 5$, where $\eta = -\ln(\tan(\theta/2))$ and θ is the polar angle with respect to the beam.

The detector is formed of many sub-detectors with various functionality such as particle identification (PID), tracking, or triggering with the schematic shown in figure 2.2 and explained in the (sub)sections below.

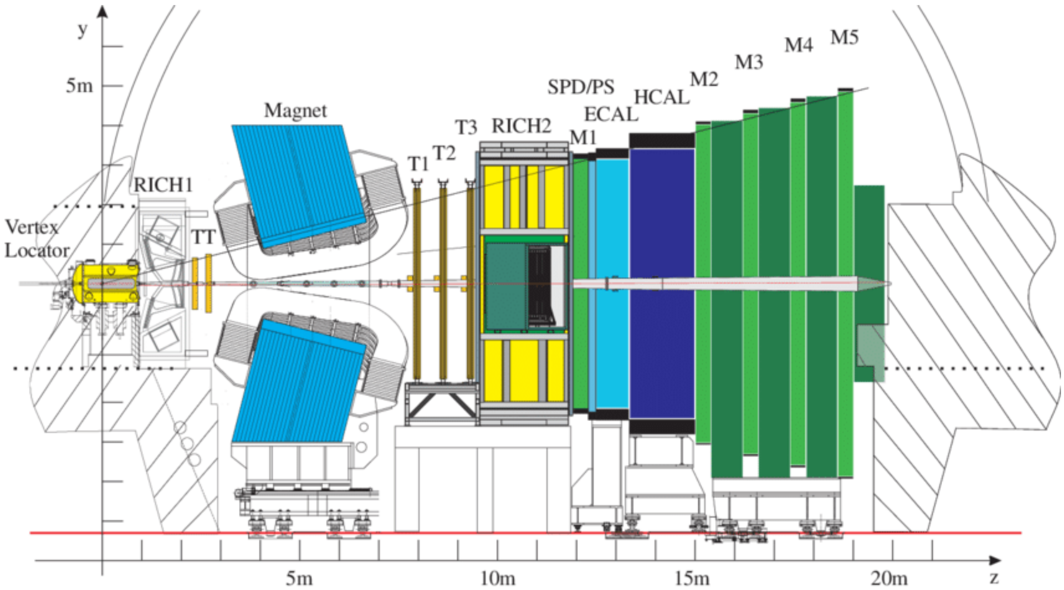


Figure 2.2: Schematic view of the detector [20] and the locations of its components as described in this chapter.

2.2.1 Tracking Detectors

The position of the $p - p$ collision is commonly known as the primary vertex and for later decays inside the tracking detectors, the secondary, tertiary, etc. vertices. Responsible for detecting, identifying, and reconstructing the vertices and tracks, the tracking system is is compromised of several sub-detectors: the Vertex Locator (VELO), the Tracker Turicensis (TT), the magnet, and the T-stations (T1-T3), which are listed from the beam collision point and outwards (downstream).

The first stage of the detector is the VELO, which provides precise measurements of track coordinates close to the $p-p$ collision point. A detailed cross section of the VELO silicon sensors is provided in figure 2.3 below. Made primarily using silicon strip detection, tracks passing through cause ionisation which creating paired energy holes in doped silicon wafers, which travel to electrodes that determine the locations of the track passing through.

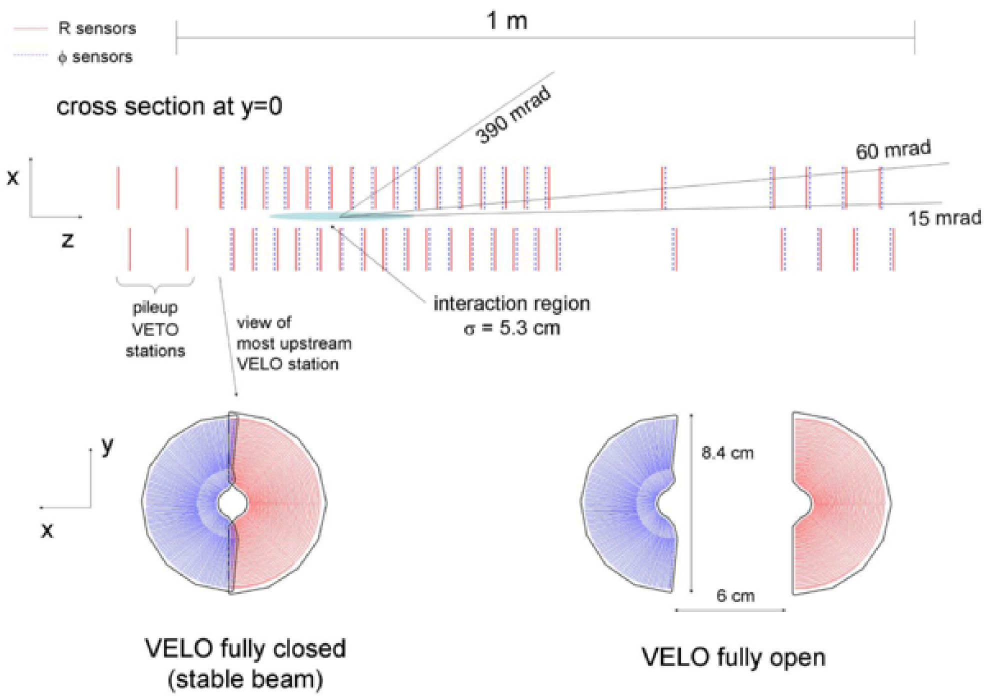


Figure 2.3: Cross section of the VELO sensors at $y = 0$ while the detector has a stable beam running. The front face of the VELO is shown underneath in both closed and open positions. [20]

To allow measurements of r and ϕ, each of the 21 micro-strip tracking modules contain 300 µm thick sensors along the z direction in vacuo (maintained by a thin aluminium wall). This minimizes the amount of material charged particles we wish to study have to travel through while balancing how much we protect our sensors from RF pickup from the beam. The VELO sensors cover a pseudo-rapidity range of $1.6 < \eta < 4.9$. With some simple angle calculations of the paths, we can assert that any tracks in the LHCb detector must pass through at least 3 of the VELO tracking modules. Another feature, as shown by figure 2.3, is that the modules are spaced more tightly near the collision point, which means that particle tracks can be more accurately ascertained for reconstruction straight after production.

Tracks, before passing through the bending magnet, pass through the TT, which is another silicon strip detector. This determines the initial location of the particles before being deflected by the dipole magnet and passing through T1-T3. The schematic view of the magnet is provided in figure 2.4 below. The 'warm' magnet [1], with its vertical field, along with the TT has two purposes: to reconstruct long lived particles that decay outside of the VELO, and to reconstruct particles with low momentum that would be deflected outside of the later tracking stations.

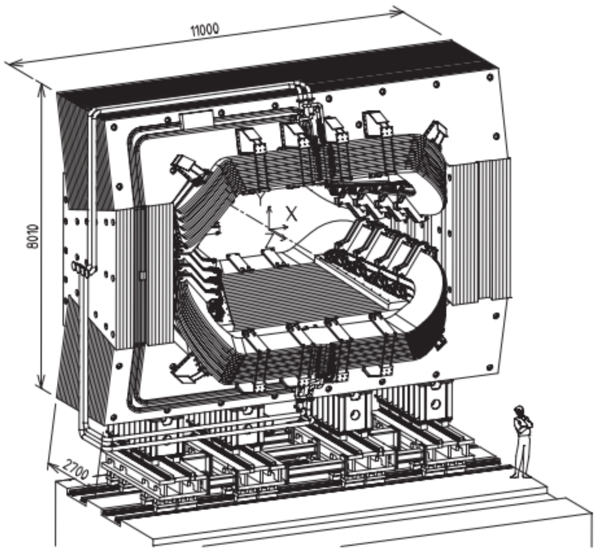


Figure 2.4: Schematic view of the magnet between TT and T1. It provides a 4 Tm magnetic field [20].

Figure 2.5 below shows a sketch from the tracking system and tracking categories according to where hits are registered.

T1-T3 stations have elements of the 'inner' and 'outer' tracking (respectively IT and OT) systems, as they are sometimes called. The TT and the IT systems are

[1] The initial proposal for a super-cooled, super-conducting magnet was denied due to having unacceptably high construction costs and build time [20].

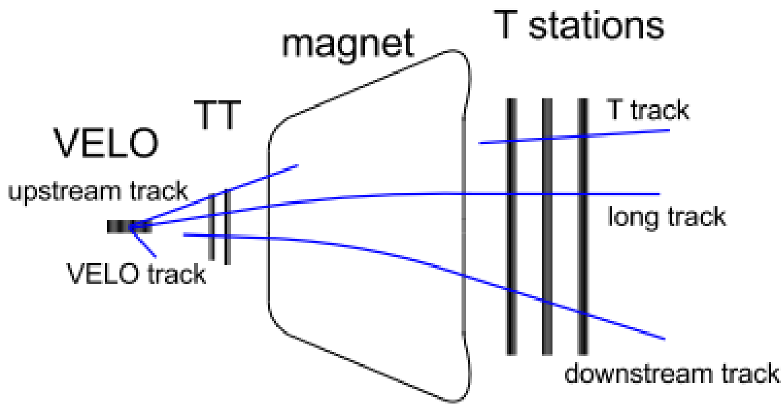


Figure 2.5: Example tracks and nomenclature used for tracks coming through the system [20].

of the same type and therefore are also known as the silicon tracking system (ST) together. The area covered by the IT has approximately 20% all particles in the LHCb acceptance range as it is close to the beam line.

The OT is a drift-time detector which relies on the gas straw-tube design which is a long tube filled with ionizable *Argon* and CO_2 [2] gas with a wire in the middle of the tube. The OT forms the outer region of the T1-3 detectors with a total of 64 drift tubes arranged in four layers. Particles penetrating this detection system create moving charges which constitutes a measurable current indicating that a track has gone through. The maximum occupancy of this system is approximately 10% of the total candidates.

2.2.2 Ring Imaging Cherenkov Detectors

Ring Imaging Cherenkov Detectors (RICH) use the phenomenon of Cherenkov radiation, which occurs when a charged particle travels through a medium faster than the speed of light in that medium. This causes energy radiation with photons emitted at an angle θ which is given by $\cos(\theta) = c/nv_p$, where c is the speed of light in a vacuum, n is the refractive index of the medium, and v_p is the velocity of the particle.

In both RICH sub-detectors, Cherenkov radiation is emitted as a cone, so it is reflected off spherical mirrors to focus the light into a ring shape. These rings can be detected using Hybrid Photon Detectors (HPDs) which use photo-cathodes that, as their name suggests, emit photo-electrons when activated and those photo-electrons

[2] *Argon* and CO_2 as gases were chosen due to their fast drift time ($\approx$ 50ns) and relative inexpense.

are pulled along an electric field to anodes, that can then detect the signal.

The RICH sub-detectors are shielded from the LHCb magnet described earlier in subsection 2.2.1 as the field would distort the photo-electron paths. RICH1 is located before the magnet and is designed to observe low momentum particles in the range of $2 - 60\,\text{GeV}/c$ while RICH2, located after the magnet, is designed to observe higher momentum particles with momentum range $15 - 100\,\text{GeV}/c$.

Particles are identified by the RICH system using a likelihood ratio test and a global pattern-recognition algorithm. The measured pattern of hit pixels (of the photon sensors) across both RICH detectors is compared against known particle hypotheses which are of the electron, muon, pion, kaon, and proton. The hypothesis that maximises the likelihood is assigned as the identity of the particle (overall known as the PID system).

2.2.3 Calorimeters

There are two main calorimeters in the LHCb detector, the Electromagnetic Calorimeter (ECAL), and the Hadron Calorimeter (HCAL) downstream from the tracking system. Just upstream lies the Scintillating Pad Detector (SPD) and the PreShower detector (PS) which, along with the HCAL and ECAL, all use scintillation material to detect particle showers as they pass through and lose energy. The density of detector pads perpendicular increases as they get closer to the beam line as we expect production of particles to be higher in that region as shown in figure 2.6.

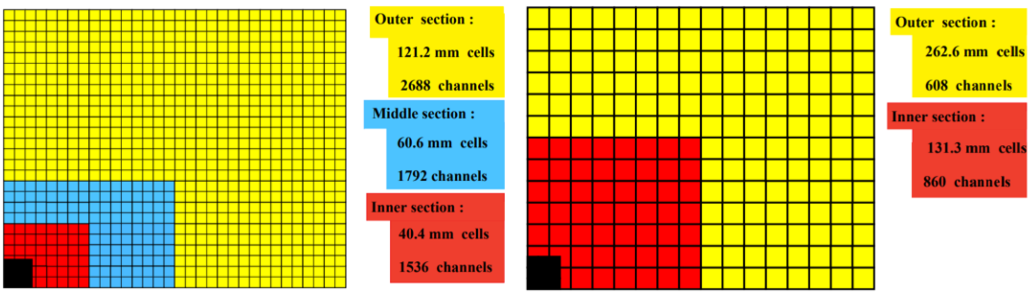


Figure 2.6: Segmentation schematic of SPD, PS, and ECAL on the left, and ECAL on the right of one quarter 'radius' of the detector. The dimensions on the left figure are for only the ECAL [20].

The SPD/PS detector system contain two planes of scintillator pads with a layer of lead between them, where the lead induces showers of neutral particles; the SPD only detects charged tracks while the PS can detect neutral particles too when electromagnetic showers pass through the SPD. Used in particle identification, SPD separates charged and neutral particles while PS aims to determine what they really are.

All four sub-detectors alternate pads of organic scintillators (made of doped polystyrene) as active medium and absorption layers. The processes occurring include (in electromagnetic interactions) Compton scattering, pair production, and bremsstrahlung radiation. Additionally, for hadronic interactions, we can see inelastic processes with various results such as pion production and particle showers.

As figure 2.6 shows, the pads for SPD, PS, and ECAL projectively match. These three sub-detectors work together to specifically discriminate electrons, photons, and pions. Photons do not scintillate but produce showers at the lead barrier which can then be picked up by the PS before heading downstream. Electrons are identified by the fact they scintillate. As the material and thickness of the absorption layer is 15mm, electrons predominantly shower in the PS while pions do not. Neutral particles tend to deposit their energy in the ECAL. Together, the PS and ECAL provide a pion rejection rate of $\approx 99\%$ while an electron acceptance rate of $\approx 95\%$.

The HCAL's main purpose is to contribute to the hardware trigger. The absorber material (this time, iron) and scintillator tiles are parallel to the beam line. To improve LHCb's PID performance, the energy of the electrons and photons are measured in the ECAL while those of hadrons are measured by the ECAL and HCAL together. The calorimetry system also provides important information for the hardware trigger, which is described later in section 2.3.

2.2.4 Muon Detectors

The muon detectors M1-5, which are the most downstream of the subsystems. Each station is separated by a sheet of iron which measures the energy and momentum of the muons that are not picked up by the calorimeters. A side view of the arrangement is shown in figure 2.7. The iron sheets act as energy filters, only allowing energetic muons to pass through each one as lower energy muons get stopped.

In a similar fashion to the OT, the muon detectors are a gas based system using $Argon, CF_4$ and CO_2 with ionisation of these gasses being detected by electrodes. The five sub-stations are split into regions R1-4 and as shown in figure 2.7, the later regions correspond to larger angular distance. Despite this, each region has approximately the same acceptance with appropriate resolution/granularity dependent on particle density. Granularity is greater in the horizontal plane as the

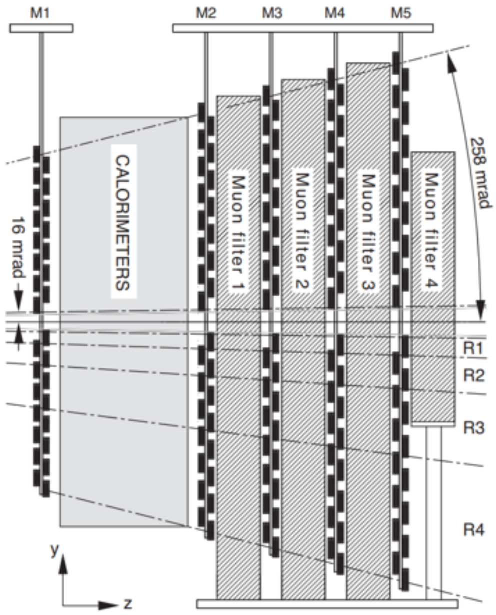


Figure 2.7: Side view of the muon detectors [20].

magnet deflection is in the same plane, thus would provide better information of the momentum of the tracks.

M1 is stationed before the SPD while M2-5 are stationed post-HCAL. The muon system is used in the first level of the trigger system, so must provide fast signals to be interpreted by hardware logic (quicker than 20 ns). The detectors provide information on the muon momentum for the first level of the trigger and muon identification for successive levels.

Compared to electrons, due to their larger mass, muons do not lose energy as quickly and appear as charged particles in the tracking system. Additionally, as muons only interact electromagnetically or weakly, they don't lose energy from interactions with nuclear matter. They are also minimally ionising so placing the detectors far downstream, they are expected to be the only particles interacting with these substations.

2.3 Trigger, Reconstruction, and Particle Identification

Given that we expect around 20-40 million events per second at any given point in the LHC, storing detailed information on each of them is impossible due to limits on how fast data can be transmitted and written, and actual available storage space. Less than 1 in every 100 inelastic events contain b quarks [22] so the trigger looks for evidence of these to allow rejection of most events. In LHC Run I+II, the LHCb experiment used a combination of ultra-fast electronics to form a hardware trigger, and software for the software trigger, to select just the most interesting collision events. This reduced the number of events saved from 30 MHz to 15 kHz.

Run III radically overhauls the solution to this problem by removing the hardware trigger. The only-software based trigger read out is the full LHC bunch-crossing rate of 40 MHz, which allows real-time (online) analysis to be much more precise and flexible.

2.3.1 Hardware Level 0 Trigger

Due to the requirement of the Level 0 (L0) trigger needing to reduce the rate of saved events by a factor of 30, it looks for evidence of high transverse momentum (p_T) events. The muon trigger only accepts events with muons of $p_T > 1.48\,\mathrm{GeV}/c$ in run I or $p_T > 1.76\,\mathrm{GeV}/c$ for run II, while the electronic and hadronic calorimeters have thresholds of $3\,\mathrm{GeV}$ and $3.68\,\mathrm{GeV}$ transverse energy (E_T) respectively.

The hadron calorimeter trigger is also designed to reject high p_T neutral pions as these represent a large source of background, which is done by applying a selection on events with high p_T electrons. This is because b-hadron decays tend to produce particles with significantly higher p_T than other decays.

Overall, the L0 is designed to output events at the rate of 1 MHz, where the muon system accounts for 400 kHz, the ECAL 150 kHz, and the HCAL 450 kHz. Data passing through L0 is stored in front-end electronics temporarily.

2.3.2 Software High Level Trigger

Events that pass the L0 trigger are moved to the Event Filter Farm (EFF). The software based High Level Trigger (HLT) works in two stages on these events, split into HLT1 and HLT2. HLT1 works by partially reconstructing events from their primary vertices and tracks in the VELO, before tracks with large impact parameters (IP) are traced back to the main tracking system, along with events that can be matched to hits in the muon detectors, are consistent with a high p_T event. This reduces the output from 1 MHz to 70 kHz.

Next, the HLT2 forms b hadron candidates from at least two tracks, with at least one having high p_T ($p_T > 300\,\mathrm{MeV}/c$), that form a common vertex displaced from all PVs, as expected in b-hadron decays. This further reduces output to 3-5 kHz in run I and 12.5 kHz in run II[3], and the events that pass HLT2 are written to LHCb storage [23].

2.3.3 Particle Identification at LHCb

The information gathered by detectors at the LHCb experiment can be combined to create a hypothesis of a track or final-state particle's nature. While LHCb has two methods of constructing the hypothesis, this analysis only uses the more recent of the methods, that is to feed sub-detector data into an artificial neural network, that assigns correlations between hits in the RICH detectors, calorimeters, and the muon system. The PID variables used in this thesis are the LHCb standard `ProbNN` variables obtained by this method.

PID performance is calibrated using a data-driven approach [24][4]. Figures 2.8 and figure 2.9 shows identification efficiencies of π, K, and μ final states. These are determined using well known decay processes, such as $D^{*+} \to (D^0 \to K^-\pi^+)\pi^+$ as the $K^-\pi^+$ can be identified using only kinematic information. Similarly, protons are reconstructed using data samples from the $\Lambda_c^+ \to pK^-\pi^+$ and $\Lambda \to p\pi$ decays.

The muon identification efficiency is measured from the $J/\psi \to \mu^+\mu^-$ decay, as it is plentiful at the LHCb and high quality samples can be obtained with relative ease (by selecting events with high primary vertex impact parameter, a large J/ψ flight distance, and a good vertex decay quality). To do this, the *tag and probe* method is used, where one muon is labelled a 'tag' muon while the remaining muon candidate is labelled the 'probe'. The muon identification efficiency is shown in figure 2.8 across the momentum range for run II.

[3] The increase in Run II is largely due to triggering more c-hadron candidates (due to innovations that allowed this to be done in an efficient way without using too much storage)

[4] Corrections, as acknowledged, have all been completed by Fernando Abudinén. Appendix B covers the rest of the work done on this topic.

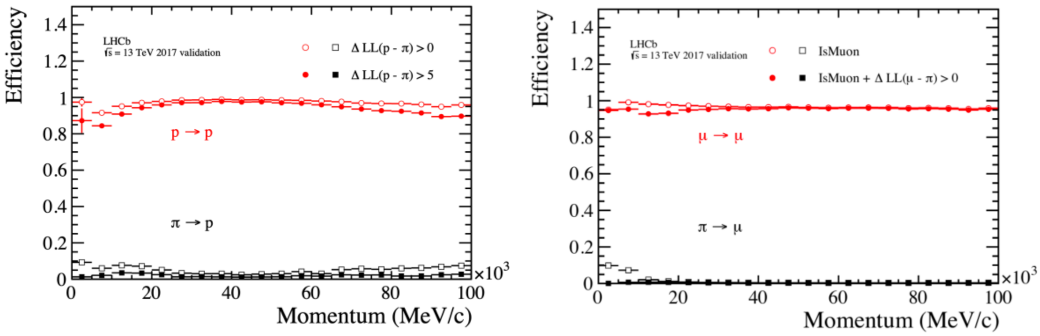


Figure 2.8: Efficiency of pion and muon identification in red, and misidentification in black based on data samples. Two different $\Delta \log \mathcal{L}(K-\pi)$ and `IsMuon`+$\Delta \log \mathcal{L}(\mu-\pi)$ requirements are applied to the data respectively, presented using open and filled markers [20].

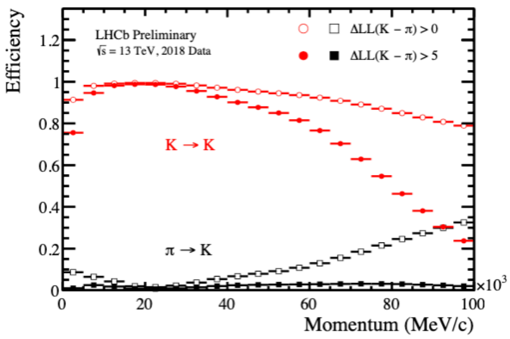


Figure 2.9: Efficiency of kaon identification in red, and misidentification in black based on data samples. Two different $\Delta \log \mathcal{L}(K - \pi)$ requirements are applied to the data, presented using similar open and filled markers [20].

Chapter 3

Data Preparation for Analysis

3.1 Data Set Description and Tools

The data used in this analysis come from the full Run 1 and Run 2 samples. The `B2XMuMu` stripping line involves a series of requirements (cuts) applied just after the trigger level to store linked events for offline analysis. The data we is taken from this `B2XMuMu` stripping line from where we extract events with final state $K^+\mu^+\mu^-$ or $\pi^+\mu^+\mu^-$ that are consistent with having originated from a b-meson decay.

The full simulation samples used are given in Tables A.1 and A.2, and we use approximately six million events for each decay channel specified in total, except for $B_c^+ \to J/\psi\pi^+$, for which we use double the number (twelve million) of events. The properties of the particles used in simulation are specified in Table 3.2 below. Then, as simulated samples are not necessarily perfect, raw simulated samples are re-weighted to improve the agreement with the data, as described in App. B.

Non-resonant decays are generated with a choice of distribution in the dimuon invariant mass squared (q^2). We consider the uncertainty in the true phase space distribution as a source of systematic error later (Sec. 5) as this isn't expected to be an accurate model of the true decays. The search for the non-resonant decay $B_c^+ \to \pi^+\mu^+\mu^-$ is done in the q^2 intervals shown in table 3.1.

Table 3.1: Discrete intervals of q^2 (bins) that are used in this analysis.

	$q^2/(\text{GeV}/c^2)^2$
Very low	$0.1 < q^2 < 1.1$
Low	$1.1 < q^2 < 8$
Central	$11 < q^2 < 12.5$
High	$15 < q^2 < 35$
Combined	Above bins combined

We use DaVinci (version `v45r6`) to process all data. DaVinci is LHCb software that can reconstruct events, identify particles, filter and select data, and reconstruct simulated Monte-Carlo data.

Table 3.2: Particle masses, lifetimes and widths used in simulation compared to Particle Data Group (PDG) [3] values. The PDG value of the reduced Planck's constant is $\hbar = 6.582119569 \cdot 10^{-16}$ eVs.

	Simulation		PDG	
Particle	Mass [MeV]	Lifetime [s]	Mass [MeV]	Lifetime [s]
B_c^+	6273.7	5.07×10^{-13}	6274.47 ± 0.32	$(5.10 \pm 0.09) \times 10^{-13}$
J/ψ	3096.90	7.09×10^{-21}	3096.9 ± 0.01	N/A
$\psi(2S)$	3686.10	2.20×10^{-21}	3686.10 ± 0.06	N/A
B_s^{*0}	5415.40	1.00×10^{-19}	$5415.4^{+1.8}_{-1.5}$	N/A
B^{*0}	5325.20	1.00×10^{-19}	5324.71 ± 0.21	N/A

	Simulation	PDG
Particle	$\hbar$/Lifetime [keV]	Width [keV]
B_c^+	$1.3 \cdot 10^{-6}$	N/A
J/ψ	92.9	92.6 ± 1.7
$\psi(2S)$	299	294 ± 8
B_s^{*0}	6.58	N/A
B^{*0}	6.58	N/A

3.2 Pre-selection and Kinematic Selection

We reconstruct B_c^+ candidates by combining charged-particle candidates and performing kinematic fits according to the topologies of all of the considered decay modes: non-resonant $B_c^+ \to \pi^+\mu^+\mu^-$, and resonant $B_c^+ \to J/\psi\,(\mu^+\mu^-)\,\pi^+$, $B_c^+ \to \psi(2S)(\mu^+\mu^-)\,\pi^+$ decay modes. We also use simulation samples of $B_c^+ \to B^{*0}(\mu^+\mu^-)\pi^+$ and $B_c^+ \to B_s^{*0}(\mu^+\mu^-)\pi^+$ decays to provide enough data in the high q^2 region to train the MultiVariate Analysis (MVA) algorithm.

Most selection criteria are adopted from analyses of $B^+ \to \pi^+\mu^+\mu^-$ decays [25], which involve the same final state particles, hence provide methodologies and tools for us to inspire from in this analysis. However, we do make a dedicated optimization of the PID criteria and the MVA used to suppress combinatorial background for this analysis, as documented in Sec. 3.3.

The selection plan is as follows:

- B_c^+ candidates are obtained from the `B2XMuMu` stripping line and passed through a common loose pre-selection, for all decay modes. We additionally apply loose requirements to select candidates in the kinematic and track-multiplicity ranges supported by the data-simulation discrepancy corrections.
- Then, a common MVA is used to reduce the combinatorial backgrounds. This is followed by a selection on specific trigger lines.
- Finally, we apply PID and invariant mass requirements to separate the signal and normalisation modes.

3.2.1 Pre-selection

We require each B-candidate daughter to be inside the ranges of momentum and pseudo-rapidity $5 < p < 100\,\text{GeV}/c$ and $1.9 < \eta < 4.9$. We further require the event track multiplicity to be below 300 tracks. The calibration histograms cover up to 500 tracks, but the region between 300 and 500 is too sparsely populated to extract kinematic data-simulation corrections (see Sec. B.3). These requirements are also included in Table 3.3. `isMuonLoose` involves requiring muons to be within their dimuon pairs, which we require that the pion does not satisfy this requirement. `InAccMuon` involves the muon to have a trajectory such that it passes through the acceptance of the muon system.

Table 3.3: Preselection and fiducial requirements applied to all data sets.

Candidate	Selection criteria
all tracks	$p_{\mathrm{T}} > 300\,\mathrm{MeV}/c$
pion	`isMuonLoose` $== 0$
	`InAccMuon` $== 1$
all tracks	$5 < p < 100\,\mathrm{GeV}/c$
	$1.9 < \eta < 4.9$
event	$N_{\mathrm{tracks}} < 300$

3.2.2 MVA

With the aim to expel combinatorial background, MVA algorithms are trained for the Run 1 and for Run 2 samples.

The MVA algorithm for Run 1 was trained using 2011 and 2012 samples and the one for Run 2 using 2016, 2017 and 2018 samples in the same relative proportions as in the collision data for each run, where each MVA combines samples from both magnet polarities.

The variables used in the MVA training are given in Table 3.4 and their choice is motivated by a dedicated optimization study (see App. E). Figures 3.1 and 3.2 show normalized distributions of these variables for signal and background.

The signal sample used to train the MVA is taken from simulated non-resonant $B_c^+ \to \pi^+\mu^+\mu^-$ and resonant $B_c^+ \to J/\psi\,(\mu^+\mu^-)\pi^+$, $B_c^+ \to B^{*0}(\mu^+\mu^-)\pi^+$ and $B_c^+ \to B_s^{*0}(\mu^+\mu^-)\pi^+$ events (corrected as described in App. B).

The signal sample contains the same amount of events for each of these four decay modes. We use the $B_c^+ \to B^{*0}(\mu^+\mu^-)\pi^+$ and $B_c^+ \to B_s^{*0}(\mu^+\mu^-)\pi^+$ decay modes because they populate the high dimuon-mass region close to the kinematic boundary. We find that, by including them into the signal training sample, the performance of the MVA does not depend strongly on dimuon invariant mass over the whole phase space (see App. E). If we exclude them, we observe a decrease in the performance of the MVA at dimuon masses above $5000\,\mathrm{MeV}/c^2$.

The background sample is taken from a combination of data in low and high B_c^+ candidate mass sidebands, $5550 < m(\pi^+\mu^+\mu^-) < 5850\,\mathrm{MeV}/c^2$ and $6700 < m(\pi^+\mu^+\mu^-) < 7000\,\mathrm{MeV}/c^2$, with vetoes to remove candidates containing $J/\psi \to \mu^+\mu^-$ or $\psi(2S) \to \mu^+\mu^-$ decays as those samples are used elsewhere in the analysis.

For each run of data taking, the signal and background are sampled to contain about 10^5 candidates each, in order to match the background sample size to the available statistics of the signal MC.

We evaluate the performance of the MVA using statistically independent testing samples drawn from the same data as the training samples and of about the same sizes. All samples are required to have passed the pre-selection. We observe no considerable increase in the performance of the MVA when we apply the trigger selection to the training and testing samples and thus avoid applying them at this stage. The MVA algorithm used is a gradient boosted decision tree (BDT) [26] as implemented in the XGBoost library [27].

Table 3.4: Variables used as inputs in the MVA training.

BDT variables
π^+ p_T
μ^+ χ^2_IP
μ^- χ^2_IP
B_c^+ χ^2_IP
B_c^+ $\chi^2_\mathrm{vtx}/n_\mathrm{dof}$
B_c^+ χ^2_FD
max(μ^+ p_T, μ^- p_T)
max(DOCA(μ^+,π^+), DOCA(μ^-,π^+), DOCA(μ^+,μ^-))

The performance of the MVA classifiers is shown in Figs. 3.3 and 3.4 for data from/MC corresponding to Run I and Run II, respectively. These show the distributions of the MVA response from the signal and background samples evaluated both on the sample used to train the BDT and on the independent sample. No overtraining is observed. Also shown are the signal efficiency *vs.* background rejection (Receiver Operating Characteristic - ROC) curves for all classifiers, each with a similar performance. The Run I and Run II ROC curves are compared in Fig. 3.5. Slightly better performance is seen in Run II, which may be due to the higher centre-of-mass energy.

The optimisation of the requirement on the MVA output is described in Sec. 3.3.

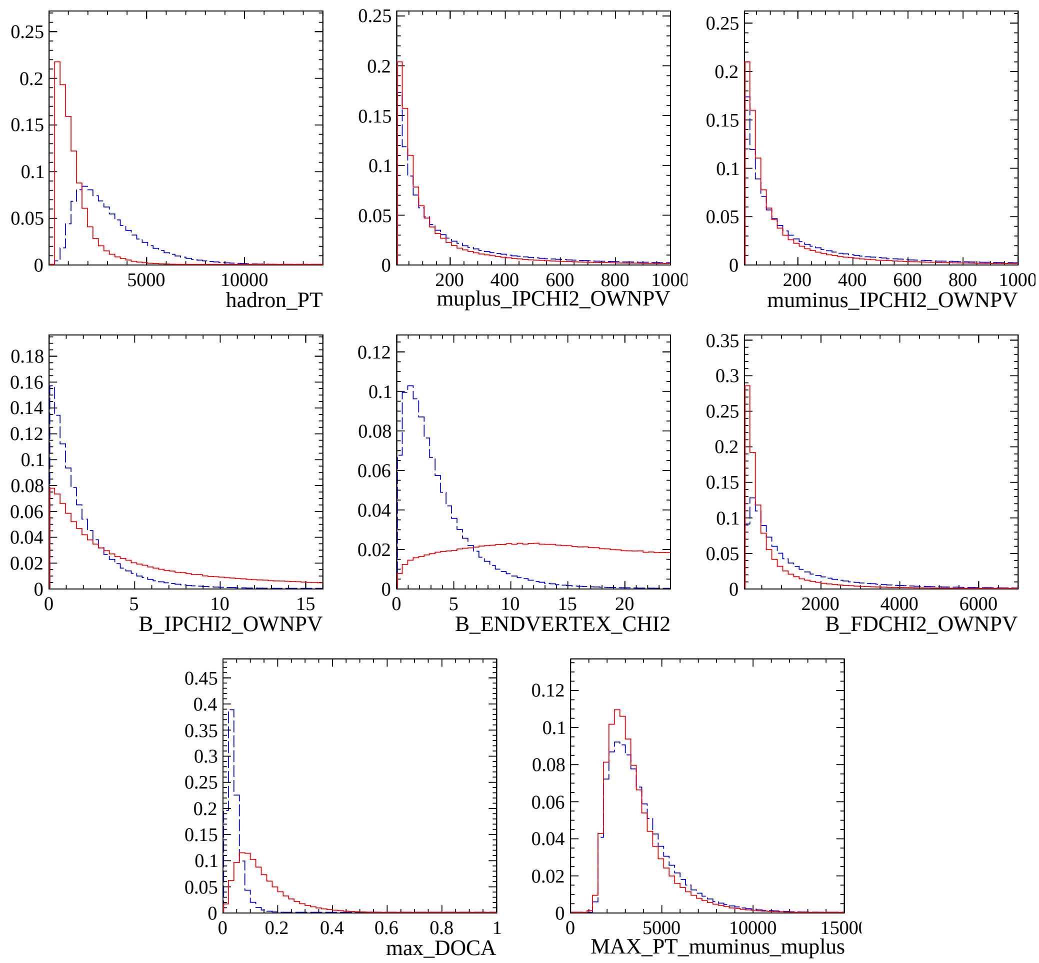


Figure 3.1: Normalized distributions of variables used as inputs in the MVA training. Signal distributions are shown in dashed blue, with background in solid red.

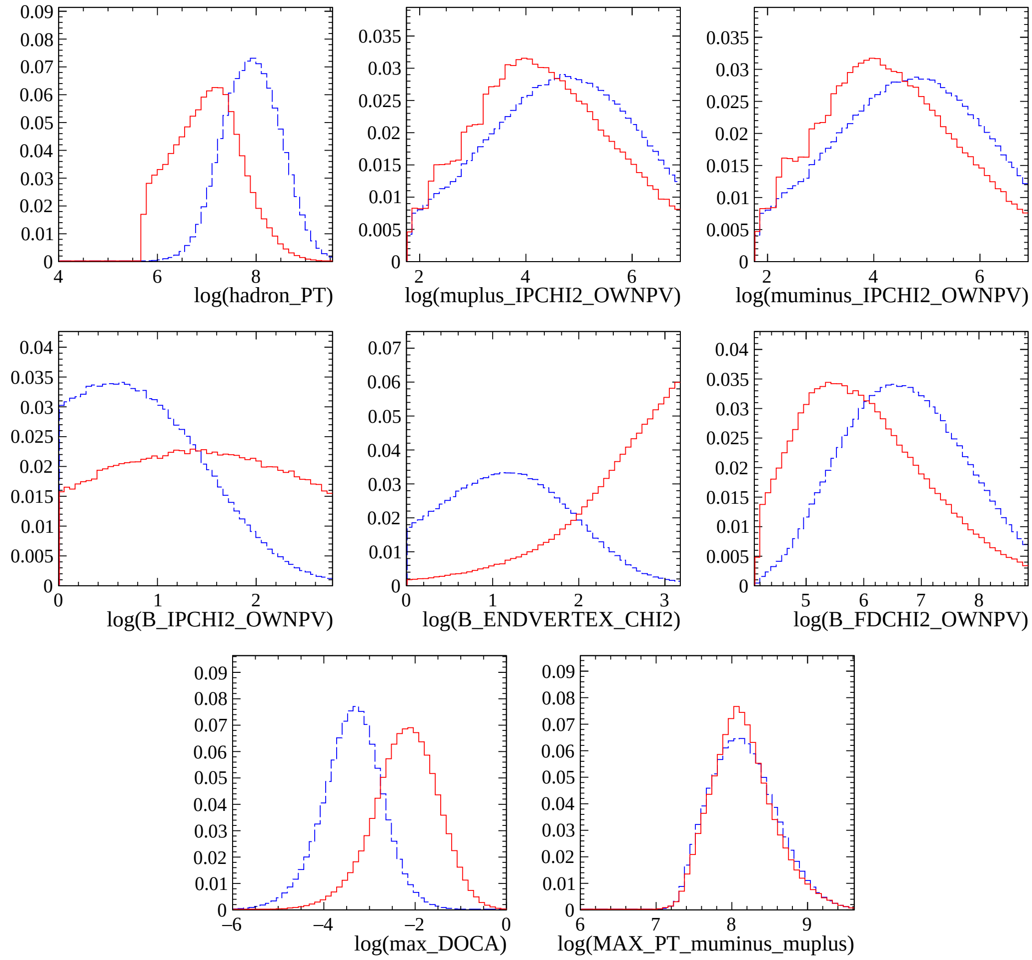


Figure 3.2: Normalized distributions of variables used as inputs in the MVA training, with logarithmic x-axis scale. Signal distributions are shown in dashed blue, with background in solid red.

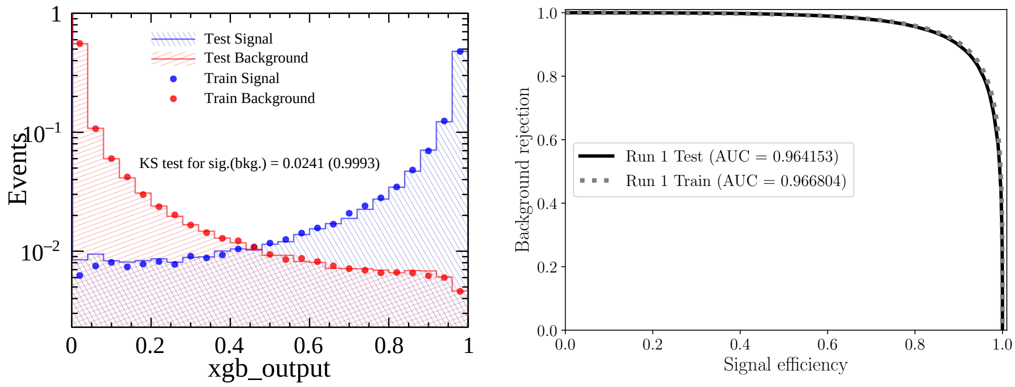


Figure 3.3: Performance plots of the MVA trained for data from/MC corresponding to Run I. Left: MVA response on the training and test categories of the signal and background samples. No overtraining is evident. Right: ROC curve.

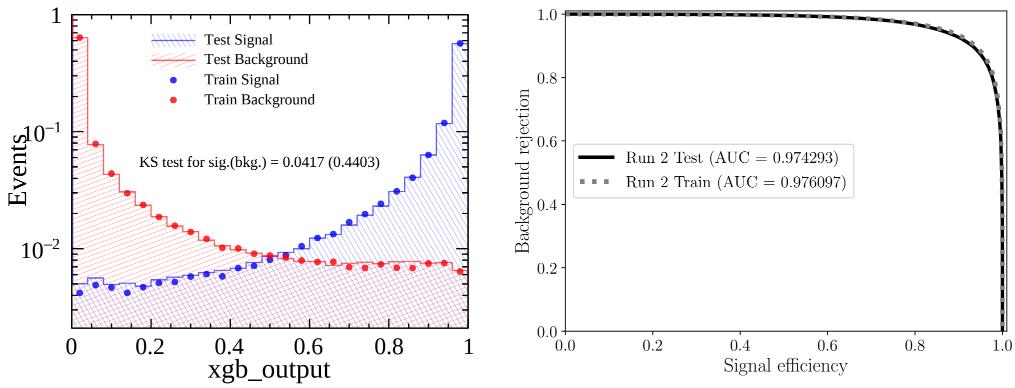


Figure 3.4: Performance plots of the MVA trained for data from/MC corresponding to Run II. Left: MVA response on the training and test categories of the signal and background samples. No overtraining is evident. Right: ROC curve.

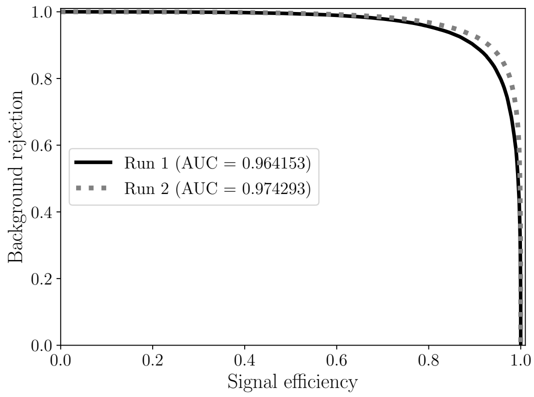


Figure 3.5: ROC curves for Run I and Run II overlaid.

3.2.3 Trigger

All candidates are required to be specifically triggered on one of the relevant trigger lines at each of the `L0`, `HLT1` and `HLT2` stages, as these have high efficiency.

At `L0` stage, we use only the `L0Muon` line. Including the `L0DiMuon` line recovers only less than 1% of the signal events for Run I and about 3% of the signal events for Run II, taking into account the preselection and fiducial requirements (see Sec. 3.2.1). These fractions are the same for resonant and non-resonant signal modes.

The recovered events populate mostly the high occupancy region where the PID performance is worse, and therefore we do not include it. Moreover, to obtain the trigger corrections (see Sec. B.2), we would need to assume that the efficiency is the same for low and high occupancies.

3.2.4 Particle Identification

The particle identification criteria used in this analysis are given in Table 3.5. The PID variables used are the LHCb standard `ProbNN` variables obtained from neutral networks trained to identify particular particle species.

To select a good quality dimuon pair, both muons are required to pass the `IsMuon` requirement at the stripping level. To reduce further potential contamination from kaons and pions, both muons must also satisfy a requirement on `ProbNNmu`.

To select the signal pions, two particle identification cuts are made: one to select pions using `ProbNNpi` and one to reject kaons using `ProbNNk`. They must also not satisfy the `IsMuon` criteria to reject misidentified backgrounds.

More complex selections based on combinations of the `ProbNNpi` and `ProbNNk` variables, such as $\texttt{ProbNNpi} \times (1 - \texttt{ProbNNk})$ and $(\texttt{ProbNNpi})^2 + (1 - \texttt{ProbNNk})^2$ have been tested and found to offer no significant improvements.

Table 3.5: Particle identification and MVA selection criteria

Particle	$B_c^+ \to \pi^+\mu^+\mu^-$	$B_c^+ \to \psi(2S)(\mu^+\mu^-)\pi^+$
π	$\texttt{ProbNNpi} > 0.3$	$\texttt{ProbNNpi} > 0.2$
	$\texttt{ProbNNk} < 0.6$	$\texttt{ProbNNk} < 0.8$
	$\texttt{IsMuon} == 0$	$\texttt{IsMuon} == 0$
μ	$\texttt{ProbNNmu} > 0.2$	$\texttt{ProbNNmu} > 0.2$
B_c^+	$\texttt{BDT} > 0.989$	$\texttt{BDT} > 0.910$

The PID requirement used to select muons is adopted from analyses of $B^+ \to \pi^+\mu^+\mu^-$ decays [25]. The requirements used to select the signal pion are optimised simultaneously with the MVA output selection due to the interaction between the misidentified and combinatorial backgrounds (see Sec. 3.3). Table 3.5 shows also the selection criteria for the MVA output. We perform separate optimisations for the search for non-resonant $B_c^+ \to \pi^+\mu^+\mu^-$ decays and for the measurement of $B_c^+ \to \psi(2S)(\to \mu^+\mu^-)\pi^+$ decays (*i.e.* of $R_{\psi(2S)/J/\psi}$).

3.2.5 Dimuon- and B-mass fit range

For the B_c^+-candidate invariant mass fits, we require all B_c^+ candidates to be in the range $6150 < M(\pi^+\mu^+\mu^-) < 6700\,\mathrm{MeV}/c^2$. The lower bound of this range excludes many potential backgrounds from partially reconstructed B_c^+ decays, while the upper bound gives a broad range that is dominated by combinatorial background and hence controls the contribution from this component in the invariant mass fit.

The B_c^+ candidates are sorted into regions of dimuon mass, some of which are defined in terms of $q^2 = m(\mu^+\mu^-)^2$, depending on the considered decay mode. The ranges used are shown in Table 3.6.

Table 3.6: Ranges of dimuon mass or q^2 for the considered decay modes.

Decay mode		dimuon-mass range
$B_c^+ \to \pi^+\mu^+\mu^-$	(very low)	$0.1 < q^2 < 1.1\,\mathrm{GeV}^2$
	(low)	$1.1 < q^2 < 8.0\,\mathrm{GeV}^2$
	(central)	$11 < q^2 < 12.5\,\mathrm{GeV}^2$
	(high)	$15 < q^2 < 35\,\mathrm{GeV}^2$
$B_c^+ \to J/\psi\,(\mu^+\mu^-)\,\pi^+$		$\lvert m(\mu^+\mu^-) - m_{J/\psi}\rvert < 50\,\mathrm{MeV}$
$B_c^+ \to \psi(2S)(\mu^+\mu^-)\,\pi^+$		$\lvert m(\mu^+\mu^-) - m_{\psi(2S)}\rvert < 50\,\mathrm{MeV}$

A check for events with multiple candidates has been done (see appendix C for more details), and we move forward using all candidates that fulfil the requirements here.

3.3 MVA and PID Optimization

3.3.1 Optimization for Non-resonant $B_c^+ \to \pi^+\mu^+\mu^-$ Final State

The pion PID and B-candidate MVA selection criteria are optimised with the goal of identifying the criteria that maximize the sensitivity to the signal non-resonant decay $B_c^+ \to \pi^+\mu^+\mu^-$.

Peaking backgrounds are expected to be negligible (see Sec. 4.2), so the PID information serves to reduce further combinatorial background. Since there is no significant difference in the MVA and PID performances for Run I and Run II (see

Sec. 3.2.2 and App. D), we perform an optimisation for the two runs together. Having the same selection criteria for the two runs also simplifies the estimation of systematic uncertainties (which will be discussed in Sec. 5).

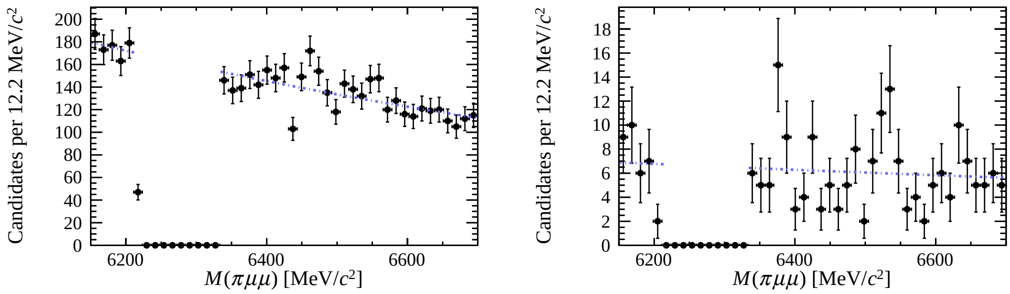
The quantity that is maximized is the Punzi Figure of Merit (FoM) [28]

$$\mathrm{FoM} = \frac{\hat{\varepsilon}}{\frac{5}{2} + \sqrt{N_{\mathrm{bkg}}}}, \tag{3.1}$$

where the efficiency $\hat{\varepsilon}$ includes the effects of the acceptance and selection averaged over all years and polarities (see Sec. 4.1) and N_{bkg} is the expected background yield inside a signal region in B mass corresponding to $6215 < m(\pi^+\mu^+\mu^-) < 6335\,\mathrm{GeV}/c^2$. The signal region has a width corresponding to about ±3 times the expected signal resolution and is centered at the expected B_c^+ mass. For the optimisation we take all B_c^+ candidates in the three considered non-resonant q^2 intervals (see Table 3.6). The value of N_{bkg} at each working point is obtained by interpolating the results of fits to sideband data into the signal region.

The sideband data sample is obtained by applying the full selection including the respective MVA and PID criteria and taking only the B_c^+ candidates outside the signal region. In the fit, the combinatorial PDF is modelled using an exponential function. The fit parameters are the exponential slope and the combinatorial-background yield. Figure 3.6 shows example projections of fits to sideband data.

We perform a scan on a 3D grid with resolution 0.1 in `ProbNNpi` (from > 0.0 to 0.9) and in `ProbNNk` (from < 0.1 to 1.0) and 0.001 in MVA cut (from > 0.983 to 0.995). Table 3.7 shows the results of the scan for the top ten selection criteria ranked according to the FoM values. Figures 3.7 and 3.8 show the variation of the efficiency $\hat{\varepsilon}$ and the FoM with BDT cut value for the two best PID working points. Figure 3.9 shows a 2D visualization of the results on the `ProbNNpi` and `ProbNNk` plane for various MVA cuts. These figures show that the selected optimal point is in the middle of a FoM plateau and is therefore a reasonable choice. For the BDT cut, the value of 0.989 is preferred. At this working point the BDT rejects 99.9% of the background, while keeping 27.1% of the signal.



(BDT>0.983, `ProbNNpi`>0.0, `ProbNNk`<1.0) (BDT>0.989, `ProbNNpi`>0.3, `ProbNNk`<0.6)

Figure 3.6: Distribution of B_c^+-candidate mass in $9\,\text{fb}^{-1}$ of sideband data for (left) an example loose working point and for (right) the chosen working point. The projections of unbinned maximum likelihood fits are overlaid.

Table 3.7: Top ten best selection criteria together with the respective values of expected background yield N_{bkg}, efficiency $\hat{\varepsilon}$ and Punzi FoM.

Selection criteria						FoM
Rank	BDT>	`ProbNNpi`>	`ProbNNk`<	N_{bkg}	$\hat{\varepsilon}$ [%]	$\hat{\varepsilon}/\left(\frac{5}{2}+\sqrt{N_{\text{bkg}}}\right)$ $[10^{-4}]$
1	0.989	0.3	0.6	64.68	0.24	2.29
2	0.989	0.2	0.7	65.55	0.24	2.29
3	0.989	0.4	0.6	63.41	0.24	2.28
4	0.989	0.4	0.7	64.15	0.24	2.28
5	0.989	0.3	0.5	64.68	0.24	2.28
6	0.989	0.3	0.4	64.68	0.24	2.27
7	0.989	0.4	0.5	63.41	0.24	2.27
8	0.989	0.5	0.6	61.86	0.24	2.27
9	0.989	0.4	0.4	62.41	0.24	2.26
10	0.988	0.4	0.6	71.42	0.25	2.26

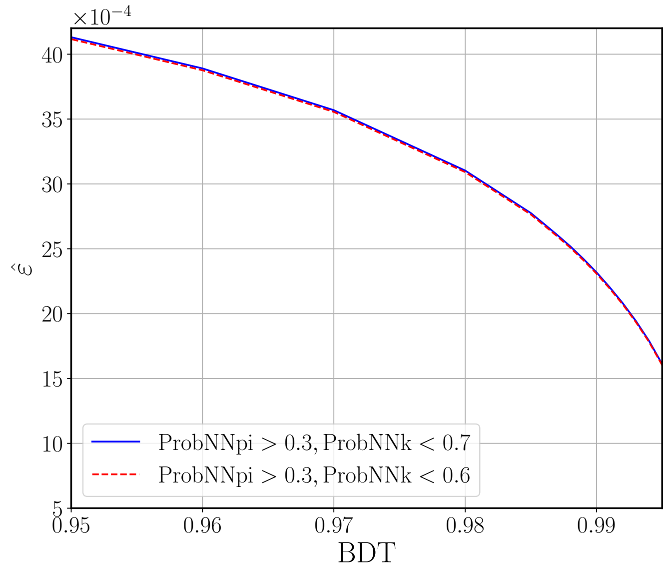


Figure 3.7: Scan of efficiency, $\hat{\varepsilon}$, against BDT, taking our two best PID working points.

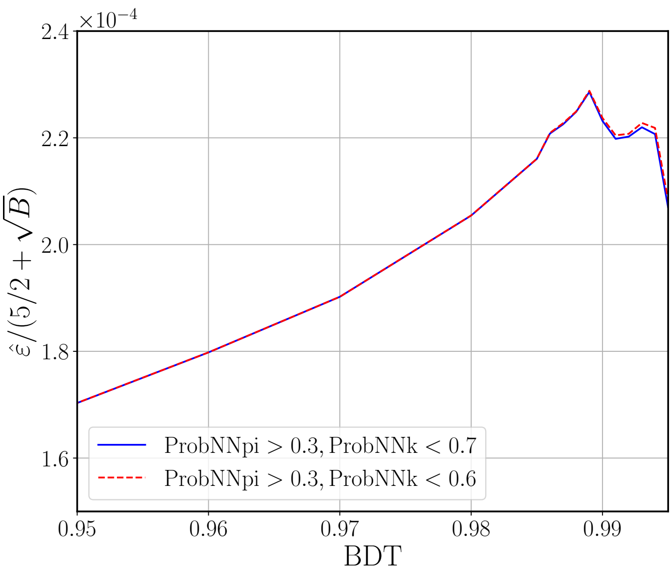
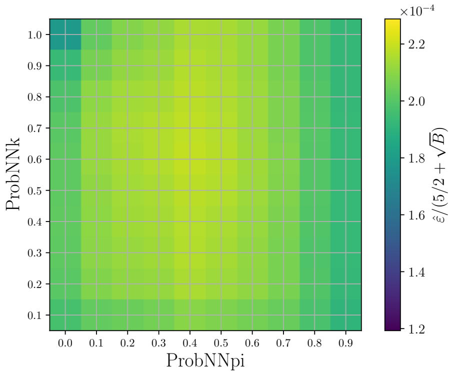
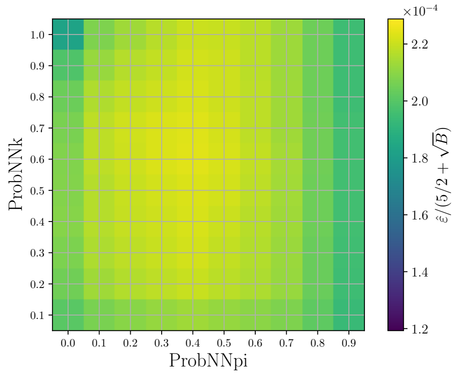


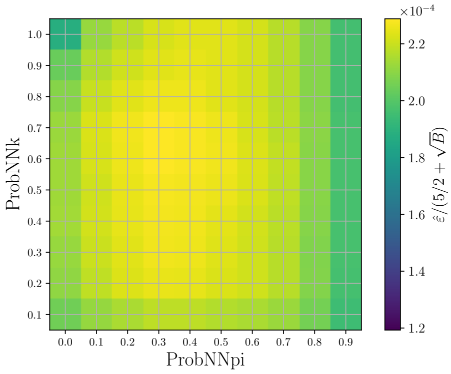
Figure 3.8: Scan of Punzi FoM against BDT, taking our two best PID working points.



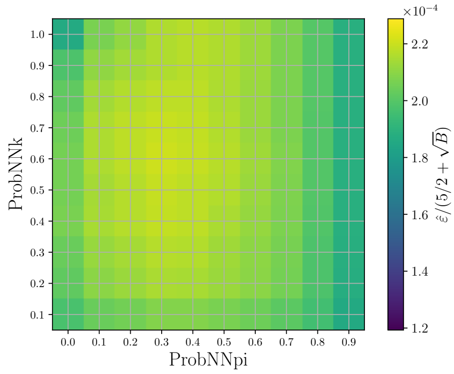
a) BDT>0.985



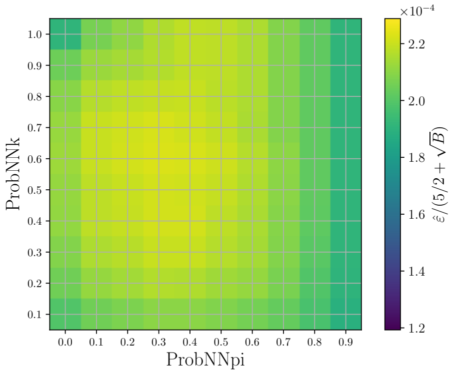
b) BDT>0.987



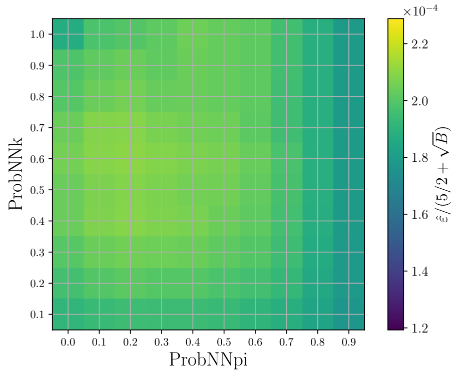
c) BDT>0.989



d) BDT>0.991



c) BDT>0.993



d) BDT>0.995

Figure 3.9: Visualization of the results of the FoM optimization scan for six different MVA selection requirements.

3.3.2 Optimization for $B_c^+ \to \psi(2S)(\mu^+\mu^-)\pi^+$ Final State

We perform an additional optimization to obtain selection criteria that maximise the sensitivity for the $B_c^+ \to \psi(2S)\pi^+$ decay. The pion PID and B-candidate MVA selection criteria are optimised simultaneously. Also for this channel, we expect peaking backgrounds to be negligible (see Sec. 4.2), so the PID information serves to reduce further combinatorial background. In this case, we maximise the FoM

$$\text{FoM} = \frac{S}{\sqrt{S+B}}, \tag{3.2}$$

where B is the background expectation and S is the expected signal yield computed as

$$S = \hat{\varepsilon} \cdot \frac{N^{\text{ref}}_{\psi(2S)\pi^+}}{\hat{\varepsilon}_{\text{ref}}}, \tag{3.3}$$

where $\hat{\varepsilon}$ stands for the average efficiency over all years and polarities (accounting for acceptance, reconstruction and selection as in Sec. 4.1) and $N_{\psi(2S)\pi^+}$ stands for the $\psi(2S)\pi^+$ yield obtained from a fit to data using the B_c^+-candidate mass calculated with $\psi(2S)$ constraint (as described in Sec. 4.3.3). In the equation above, the suffix "ref" refers to a reference working point, which we choose to correspond to a soft MVA requirement BDT>0.5 without hadron PID requirements (`ProbNNpi`>0 and `ProbNNk`<1.0). Table 3.8 shows the result of a fit to data at the reference working point, where the fit includes signal and combinatorial background components (with shapes described similarly as in Sec. 4.3). Figure 3.10 shows the respective fit projection.

We choose this FoM as it optimizes the precision of measuring the branching fraction ratio while the Punzi FoM described in Sec. 3.3.1 is more suitable for unobserved decay modes, as it optimizes for observing a signal hypothesis.

Table 3.8: Results from the reference fit to $B_c^+ \to \psi(2S)(\mu^+\mu^-)\pi^+$ data.

Fit parameter	Fit result
$B_c^+ \to \psi(2S)(\mu^+\mu^-)\pi^+$ yield	309 ± 23
Combinatorial background yield	1423 ± 41
Comb. bkg. exponent $[(\text{MeV}/c^2)^{-1}]$	$(-1.49 \pm 0.18) \times 10^{-3}$

To obtain the background expectation B, we perform for each working point a fit to the distribution of the B_c^+-candidate mass, calculated with a $\psi(2S)$ mass constraint applied to the dimuon pair, using sideband data. The sideband excludes the signal region $6215 < m(\pi^+\mu^+\mu^-) < 6335\,\text{GeV}/c^2$. The value of B at each working point is obtained by interpolating the result of the fit to sideband data into the

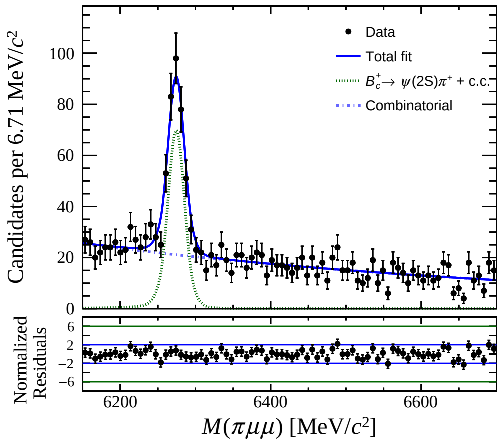
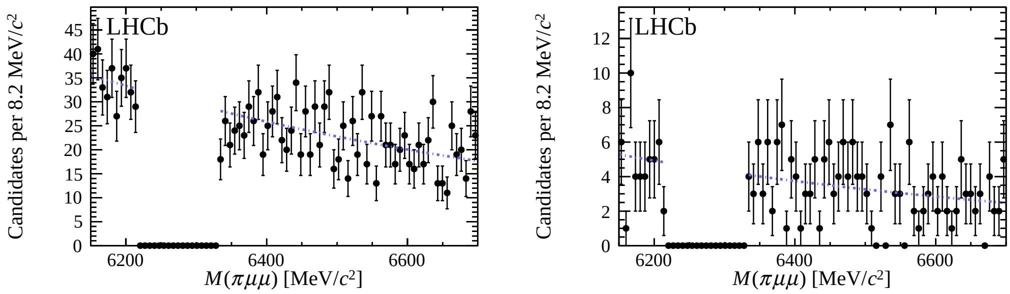


Figure 3.10: Distribution of B_c^+-candidate mass in the $9\,\mathrm{fb}^{-1}$ data sample for the reference working point. The projection of an unbinned maximum-likelihood fit is overlaid.

signal region.

The sideband data sample is obtained by applying the full selection including the respective MVA and PID criteria and taking only the B_c^+ candidates outside the signal region. In the fit, the combinatorial PDF is modelled using an exponential function. The fit parameters are the exponential slope and the combinatorial-background yield. Figure 3.11 shows example projections for the fits to sideband data.

We perform a scan on a 3D grid with resolution 0.1 in `ProbNNpi` (from 0.0 to 0.9) and in `ProbNNk` (from 0.1 to 1.0) and 0.01 in MVA cut (from 0.80 to 0.99). Table 3.9 shows the results of the scan for the top ten selection criteria ranked according to the FoM values. Figures 3.12 and 3.13 show the variation of the efficiency $\hat{\varepsilon}$ and the FoM with BDT cut value for the two best PID working points. Figure 3.14 shows a 2D visualization of the results on the `ProbNNpi` and `ProbNNk` plane for various MVA cuts. These figures show that the selected optimal point is inside a FoM plateau and is therefore a reasonable choice. For the BDT cut, the value of 0.91 is preferred. At this working point the BDT rejects 98.8% of the background, while keeping 63.3% of the signal.



(BDT> 0.5, `ProbNNpi`> 0.0, `ProbNNk`< 1.0) (BDT> 0.91, `ProbNNpi`>0.2, `ProbNNk`< 0.8)

Figure 3.11: Distribution of B_c^+-candidate mass in $9\,\mathrm{fb}^{-1}$ of sideband data for (left) an example loose working point and for (right) the chosen working point. The projections of unbinned maximum likelihood fits are overlaid.

Table 3.9: Top ten best selection criteria together with the respective values of expected background yield N_{bkg}, efficiency $\hat{\varepsilon}$ and figure of merit.

	Selection criteria					FoM
Rank	BDT>	`ProbNNpi`>	`ProbNNk`<	B	S	$S/\sqrt{S+B}$
1	0.91	0.2	0.8	81.09	244.8	13.56
2	0.93	0.2	0.8	61.60	232.5	13.56
3	0.92	0.2	0.8	72.10	239.1	13.55
4	0.91	0.2	1.0	83.62	246.0	13.55
5	0.93	0.2	1.0	64.12	233.7	13.54
6	0.92	0.2	1.0	74.62	240.3	13.54
7	0.93	0.3	0.8	58.68	230.1	13.54
8	0.91	0.3	1.0	79.31	243.1	13.54
9	0.91	0.3	0.8	77.89	242.2	13.54
10	0.93	0.3	1.0	60.09	230.9	13.54

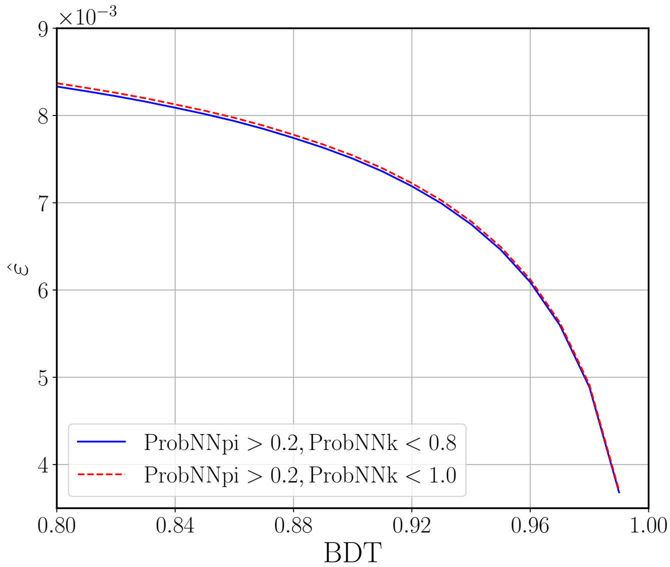


Figure 3.12: Scan of efficiency, $\hat{\varepsilon}$, against BDT, taking the best and another close PID working point.

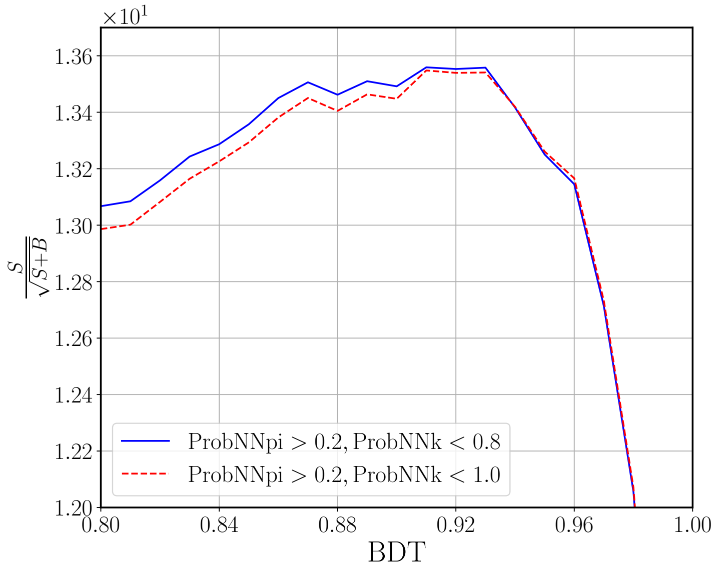
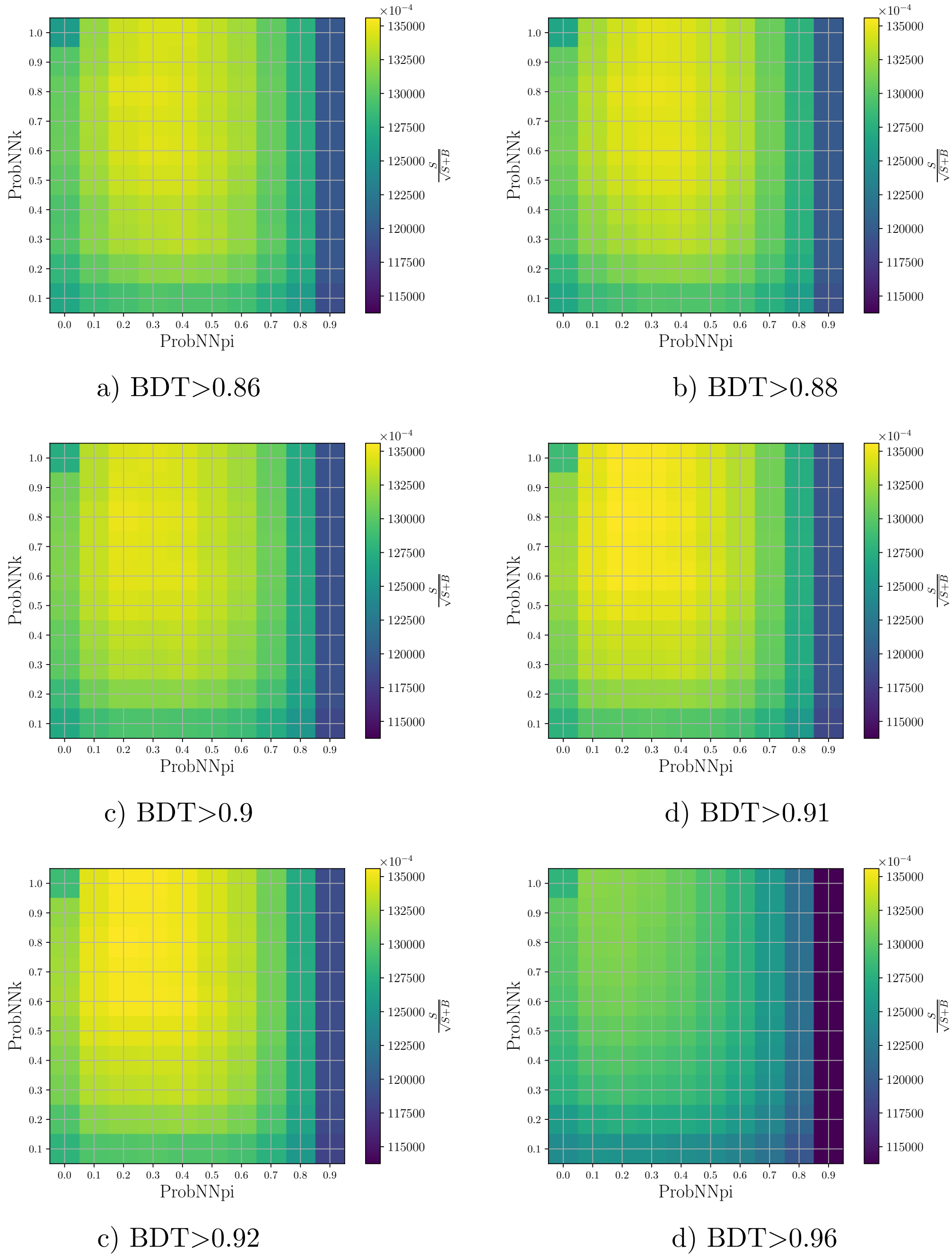


Figure 3.13: Scan of figure of merit against BDT, taking the best and another close PID working point.



a) BDT>0.86

b) BDT>0.88

c) BDT>0.9

d) BDT>0.91

c) BDT>0.92

d) BDT>0.96

Figure 3.14: Visualization of the results of the FoM optimization scan for six different MVA selection requirements.

Chapter 4

Signal Extraction

4.1 Efficiency Determination

We determine the efficiency for each year and polarity as the product

$$\varepsilon = \varepsilon_{\mathrm{acc}} \cdot \varepsilon_{\mathrm{sel}}, \tag{4.1}$$

where $\varepsilon_{\mathrm{acc}}$ is the acceptance efficiency and $\varepsilon_{\mathrm{sel}}$ is the selection efficiency. Acceptance and selection efficiencies are estimated using simulation that has been corrected for data-MC discrepancies.

4.1.1 Detector Acceptance

The efficiency associated with the detector acceptance is estimated using simulated samples without applying acceptance cuts and without detector simulation. We calculate the acceptance efficiency as

$$\varepsilon_{\mathrm{acc}} = \frac{\sum_k^{N_{\mathrm{acc}}} w_{p_{\mathrm{T}}}^k}{\sum_k^{N_{\mathrm{gen}}} w_{p_{\mathrm{T}}}^k}, \tag{4.2}$$

where N_{acc} is the number of signal B_c^+ mesons where all B_c^+ daughters fulfill the acceptance requirement, N_{gen} is the total number of generated signal B-mesons and $w_{p_{\mathrm{T}}}$ are the p_{T} weights introduced in Sec. B.3. The acceptance requirement (`BcDaughtersInLHCb`) corresponds to a polar angle $\theta = \arccos(p_z/p)$ in the lab frame within the range $0.01 < \theta < 0.40\,\mathrm{rad}$.

Tables 4.1 and 4.2 show the acceptance efficiencies for all considered modes. The acceptance efficiency for the non-resonant mode is shown also for each q^2 bin separately, in which case the denominator includes only the particles generated in the respective q^2 bin. The uncertainties on the acceptance efficiencies are only statistical and systematic uncertainties will be quantified in Sec. 5. A small difference is seen between years of data-taking at different centre-of-mass energies, but otherwise these are stable as expected.

Table 4.1: Acceptance efficiency (%) for resonant decay modes for each year and polarity.

Year	Magnet polarity	$B_c^+ \to J/\psi \pi^+$	$B_c^+ \to \psi(2S)\pi^+$
2011	up	12.63 ± 0.08	12.67 ± 0.08
2011	down	12.49 ± 0.08	12.48 ± 0.08
2012	up	13.06 ± 0.08	13.08 ± 0.08
2012	down	13.15 ± 0.08	13.16 ± 0.08
2015	up	14.92 ± 0.08	14.91 ± 0.08
2015	down	14.97 ± 0.08	15.05 ± 0.08
2016	up	14.96 ± 0.08	14.98 ± 0.08
2016	down	15.04 ± 0.08	15.10 ± 0.08
2017	up	14.93 ± 0.08	14.98 ± 0.08
2017	down	14.92 ± 0.08	14.99 ± 0.08
2018	up	14.89 ± 0.08	14.89 ± 0.08
2018	down	14.87 ± 0.08	14.94 ± 0.08

Table 4.2: Acceptance efficiency (%) for each year and polarity for non-resonant $B_c^+ \to \pi^+\mu^+\mu^-$ decays. The efficiencies are given for each q^2 interval and for all intervals combined.

		$B_c^+ \to \pi^+\mu^+\mu^-$ — q^2 interval				
Year	Mag. pol.	All	Very Low	Low	Central	High
2011	up	13.26 ± 0.08	13.86 ± 0.08	13.16 ± 0.14	12.82 ± 0.33	13.32 ± 0.13
2011	down	13.05 ± 0.08	14.07 ± 0.08	13.07 ± 0.14	12.98 ± 0.33	12.90 ± 0.12
2012	up	13.53 ± 0.08	14.52 ± 0.08	13.46 ± 0.14	12.65 ± 0.33	13.58 ± 0.13
2012	down	13.69 ± 0.08	14.69 ± 0.08	13.63 ± 0.13	13.26 ± 0.34	13.66 ± 0.13
2015	up	15.41 ± 0.08	16.34 ± 0.08	15.22 ± 0.14	15.43 ± 0.35	15.45 ± 0.13
2015	down	15.45 ± 0.08	16.37 ± 0.08	15.18 ± 0.14	14.69 ± 0.34	15.67 ± 0.14
2016	up	15.39 ± 0.08	16.29 ± 0.08	15.44 ± 0.15	15.16 ± 0.34	15.26 ± 0.14
2016	down	15.56 ± 0.08	17.08 ± 0.08	15.44 ± 0.15	15.45 ± 0.36	15.49 ± 0.13
2017	up	15.43 ± 0.08	16.67 ± 0.08	15.39 ± 0.15	14.85 ± 0.34	15.39 ± 0.14
2017	down	15.54 ± 0.08	17.05 ± 0.08	15.28 ± 0.14	15.21 ± 0.34	15.62 ± 0.14
2018	up	15.40 ± 0.08	16.38 ± 0.08	15.35 ± 0.15	14.75 ± 0.36	15.41 ± 0.13
2018	down	15.48 ± 0.08	16.36 ± 0.08	15.58 ± 0.14	15.48 ± 0.35	15.22 ± 0.13

4.1.2 Selection Efficiency

The efficiency associated with the selection is estimated from the full simulation samples on which we apply the full selection (see Sec. 3.2) except the PID requirements. We account for the effect of the PID requirements by weighting the candidates with efficiencies, corresponding to the selection applied, where the selection efficiency can be expressed as

$$\varepsilon_{\text{sel}} = \frac{\sum_{k}^{N_{\text{sel}}} w_{\text{PID}}^{k} \cdot w_{\mu\text{ID}}^{k} \cdot w_{\text{trigger}}^{k} \cdot w_{\text{track}}^{k} \cdot w_{\text{kin}}^{k} \cdot}{\sum_{k}^{N_{\text{acc}}} w_{p_{\text{T}}}^{k}}, \tag{4.3}$$

where N_{sel} is the number of selected candidates, N_{acc} is the number of signal B_c^+-mesons generated in the detector acceptance (*i.e.* passing the `BcDaughtersInLHCb` requirement), and w_{PID}, $w_{\mu\text{ID}}$, w_{trigger}, w_{track}, w_{kin}, and $w_{p_{\text{T}}}$ are the PID, muon ID, trigger, track reconstruction, kinematic and p_{T} weights detailed in Sec B.

Tables 4.3 and 4.4 show the selection efficiencies for all considered modes. The efficiency for the non-resonant mode is shown also for each q^2 bin separately, in which case the denominator includes only the particles generated in the respective q^2 bin. The statistical uncertainties on the selection efficiencies are around 0.03% in absolute terms. (Similarly to the section above, systematic uncertainties will be quantified in Sec. 5.)

We observe small differences between years of data-taking which we mainly attribute to the differences in `L0` muon trigger conditions, which have varying muon p_{T} thresholds. For instance, the muon p_{T} threshold is highest for 2015 data.

The products of acceptance and selection efficiencies (using all correction weights) are shown in Tables 4.5 and 4.6. We calculate efficiency averages over all years and polarities, taking into account the relative proportions in the data. The statistical uncertainties on these efficiencies are around 0.01%, again in absolute terms.

To calculate the efficiency averages over years and polarities, we use the proportions of $B^+ \to J/\psi K^+$ decays from Ref. [29].

Table 4.3: Selection efficiency in percent for each year and polarity for resonant modes. The efficiencies are given for the two working points (WP) defined in Table 3.5.

		$B_c^+ \to J/\psi\pi^+$		$B_c^+ \to \psi(2S)\pi^+$	
Year	Magnet polarity	$\pi^+\mu^+\mu^-$ WP	$\psi(2S)\pi^+$ WP	$\pi^+\mu^+\mu^-$ WP	$\psi(2S)\pi^+$ WP
2011	up	1.77	3.98	1.93	4.29
2011	down	1.76	3.97	1.96	4.35
2012	up	1.74	3.92	1.91	4.26
2012	down	1.74	3.97	1.93	4.32
2015	up	1.52	2.88	1.74	3.32
2015	down	1.63	3.17	1.90	3.68
2016	up	2.07	4.36	2.30	4.70
2016	down	2.03	4.20	2.26	4.60
2017	up	2.11	4.28	2.35	4.67
2017	down	2.13	4.29	2.37	4.66
2018	up	2.06	4.21	2.29	4.63
2018	down	2.05	4.21	2.32	4.65

Table 4.4: Selection efficiency in percent for each year and polarity for the non-resonant mode. The efficiencies are given for each q^2 interval and for all combined.

		$B_c^+ \to \pi^+\mu^+\mu^-$ q^2 range				
Year	Magnet polarity	Very Low	Low	Central	High	All
2011	up	1.11	1.36	1.90	1.93	1.66
2011	down	1.31	1.41	1.86	1.93	1.69
2012	up	1.22	1.36	1.91	1.95	1.67
2012	down	1.34	1.32	1.80	1.90	1.64
2015	up	1.12	1.12	1.65	2.00	1.57
2015	down	1.15	1.24	1.97	2.02	1.66
2016	up	1.43	1.60	2.38	2.23	1.96
2016	down	1.46	1.50	2.44	2.26	1.92
2017	up	1.55	1.66	2.34	2.33	2.03
2017	down	1.58	1.69	2.24	2.40	2.06
2018	up	1.43	1.60	2.19	2.34	1.99
2018	down	1.54	1.59	2.22	2.35	2.00

Table 4.5: Product of acceptance and selection efficiency in percent for each year and polarity for resonant modes. The efficiencies are given for the two working points (WP) defined in Table 3.5.

		$B_c^+ \to J/\psi\pi^+$		$B_c^+ \to \psi(2S)\pi^+$	
Year	Magnet polarity	$\pi^+\mu^+\mu^-$ WP	$\psi(2S)\pi^+$ WP	$\pi^+\mu^+\mu^-$ WP	$\psi(2S)\pi^+$ WP
2011	up	0.22	0.50	0.24	0.54
2011	down	0.22	0.50	0.24	0.54
2012	up	0.22	0.51	0.25	0.56
2012	down	0.23	0.52	0.25	0.57
2015	up	0.23	0.43	0.26	0.49
2015	down	0.24	0.47	0.29	0.55
2016	up	0.31	0.65	0.34	0.70
2016	down	0.31	0.63	0.34	0.69
2017	up	0.32	0.64	0.35	0.70
2017	down	0.32	0.64	0.35	0.70
2018	up	0.31	0.63	0.34	0.69
2018	down	0.31	0.63	0.34	0.69
Average		0.28	0.59	0.32	0.65

Table 4.6: Product of acceptance and selection efficiency in percent for each year and polarity for the non-resonant mode. The efficiencies are given for each q^2 interval and for all combined.

		$B_c^+ \to \pi^+\mu^+\mu^-$ q^2 range				
Year	Magnet polarity	Very Low	Low	Central	High	All
2011	up	0.15	0.18	0.24	0.26	0.22
2011	down	0.18	0.19	0.24	0.25	0.22
2012	up	0.18	0.18	0.24	0.27	0.23
2012	down	0.20	0.18	0.24	0.26	0.22
2015	up	0.18	0.17	0.26	0.31	0.24
2015	down	0.19	0.19	0.29	0.32	0.26
2016	up	0.23	0.25	0.36	0.35	0.30
2016	down	0.25	0.23	0.35	0.35	0.30
2017	up	0.26	0.26	0.35	0.36	0.31
2017	down	0.27	0.26	0.34	0.37	0.32
2018	up	0.23	0.25	0.32	0.36	0.31
2018	down	0.26	0.25	0.34	0.36	0.31
Average		0.23	0.23	0.31	0.33	0.28

4.2 Backgrounds

This section is dedicated to the study of background decays that are expected to remain after the selection described in Sec. 3.2. Three main sources are considered; partially reconstructed decays (Sec. 4.2.1), peaking background from misidentified resonant or non-resonant $B_c^+ \to K^+\mu^+\mu^-$ decays (Sec. 4.2.2) and combinatorial background (Sec. 4.2.3).

4.2.1 Partially reconstructed background

Partially reconstructed decays, where one or more final state particles are missed in the reconstruction, populate the lower region of the B_c^+-candidate invariant mass distribution. For all modes, we require candidates to have a B_c^+-candidate mass $m(\pi^+\mu^+\mu^-) > 6150\,\mathrm{MeV}/c^2$. Since the threshold is above $m_{B_c^+} - m_\pi$, this requirement removes most potential partially reconstructed backgrounds, with only a small residual tail remaining in the fit range. Although this tail could be partially absorbed by the shape of the combinatorial background, we study the possibility of explicitly modelling the partially reconstructed backgrounds using simulation.

One source of partially reconstructed background is considered for each mode, with the final state π^+ replaced by $\rho^+ \to \pi^+\pi^0$ and the π^0 missing. These are, explicitly, for the signal mode

- $B_c^+ \to \rho^+(\pi^+\pi^0)\,\mu^+\mu^-$,

for the $B_c^+ \to \psi(2S)(\mu^+\mu^-)\,\pi^+$ mode

- $B_c^+ \to \psi(2S)(\mu^+\mu^-)\,\rho^+(\pi^+\pi^0)$,

and for the normalisation mode

- $B_c^+ \to J/\psi\,(\mu^+\mu^-)\,\rho^+(\pi^+\pi^0)$.

In all cases the polarisation of the ρ meson can affect the momentum of the missing pion and hence the B_c^+-candidate mass shape of the partially reconstructed background; as this is unknown it is a potential source of systematic uncertainty. Although there are other possible backgrounds with a single missing particle, only in these cases is the momentum of the missing particle small enough that a non-negligible component may remain within the allowed B_c^+-candidate mass range. This is true for modes with a missing photon or with missing neutrinos, such as $B_c^+ \to J/\psi\,\mu^+\nu$.

The B_c^+-candidate invariant mass shapes of these backgrounds, after the event selection criteria are applied, are studied with simulation. Figure 4.1 shows the B_c^+-candidate mass distribution for the $B_c^+ \to J/\psi\,\rho^+$ background to the $B_c^+ \to J/\psi\,\pi^+$ sample (note that the range shown goes beyond the applied B_c^+-candidate mass requirement). We include this decay mode as a fit component in Sec. 4.3.2. We also include the $B_c^+ \to \psi(2S)\rho^+$ component in the $B_c^+ \to \psi(2S)\pi^+$ fit, but we do not include the $B_c^+ \to \rho\mu^+\mu^-$ component in the $B_c^+ \to \pi^+\mu^+\mu^-$ search as we do not observe enough events in the B-mass range to create an accurate model (see App. F).

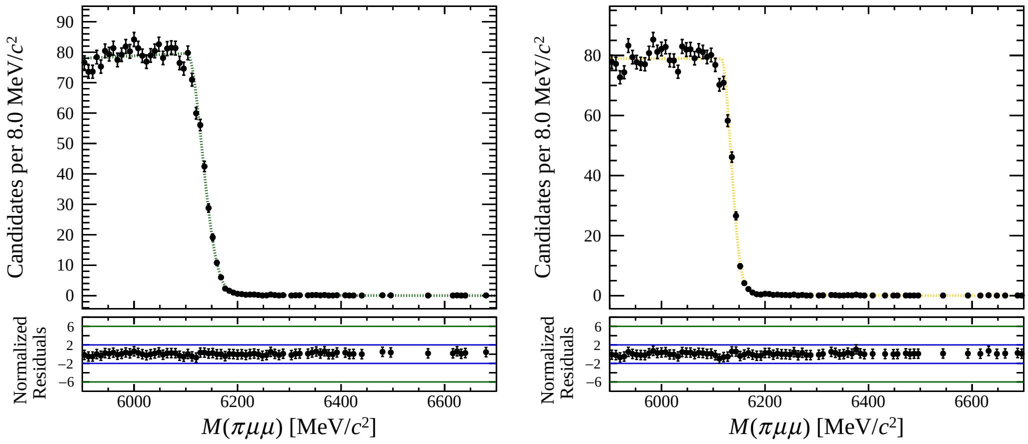


Figure 4.1: Distributions of B_c^+-candidate mass for $B_c^+ \to J/\psi\,\rho^+$ candidates reconstructed in simulation. These are shown (left, green) without and (right, gold) with the dimuon mass being constrained to $m_{J/\psi}$.

4.2.2 Misidentified peaking background

Misidentified (misid) backgrounds occur when the particle identification algorithms incorrectly identify one of the final state particles. Misidentified decays from non-resonant $B_c^+ \to K^+\mu^+\mu^-$ and resonant $B_c^+ \to \psi(2S)(\mu^+\mu^-)K^+$ decays (see [16]) are expected to be negligible due to two factors. First, the branching fractions for these decays are Cabibbo suppressed (see Sec. 1.2.1) with respect to those for the signal decays with the kaon replaced by a pion. Second, the misid probability to identify a true kaon as a pion is typically low. Multiplying these factors together gives a suppression of well below a percent, so in modes where the signal yield is $\mathcal{O}(100)$ or less the background is negligible.

Due to the high statistics in the normalisation channel, however, misidentified $B_c^+ \to J/\psi\,(\mu^+\mu^-)K^+$ decays should be considered in order to obtain a good fit. We thus include this component in the normalisation-mode fit (see Sec. 4.3.2). The contribution appears as a bump on the low-mass side of the B_c^+-mass signal peak, since the misid process reduces the B_c^+-candidate invariant mass.

We calculate the expected yield of misidentified $B_c^+ \to J/\psi K^+$ decays for each working point as

$$N_{J/\psi K^+} = R_{J/\psi K^+/J/\psi\pi^+} \cdot \hat{\varepsilon}_{J/\psi K^+} \cdot \frac{N_{J/\psi\pi^+}}{\hat{\varepsilon}_{J/\psi\pi^+}}, \tag{4.4}$$

where N denotes the yield and $\hat{\varepsilon}$ the average efficiency (or misidentification rate) for the mode indicated in the subscript, and $R_{J/\psi K^+/J/\psi\pi^+}$ is the ratio of branching fractions

$$R_{J/\psi K^+/J/\psi\pi^+} = \frac{\mathcal{B}\,(B_c^+ \to J/\psi\, K^+)}{\mathcal{B}\,(B_c^+ \to J/\psi\,\pi^+)} = 0.079 \pm 0.007\,(\text{stat}) \pm 0.003\,(\text{syst})\ [30].$$

The misidentification rate for $B_c^+ \to J/\psi K^+$ is given in Table 4.7. This information is used to obtain the predictions for $N_{J/\psi K^+}$ given in Table 4.8, where the value of $N_{J/\psi\pi^+}$ in Eq. (4.4) is taken from a fit with the $B_c^+ \to J/\psi\,(\mu^+\mu^-)K^+$ component omitted. The $B_c^+ \to J/\psi K^+$ decay is included as a fit component in Sec. 4.3.2.

Tables 4.7 and 4.8 show also the misidentification rate and expected yields for $B_c^+ \to \psi(2S)K^+$, respectively. The expected yields are obtained using Eq. (4.4), assuming that $R_{\psi(2S)K^+/\psi(2S)\pi^+}$ is equal to $R_{J/\psi K^+/J/\psi\pi^+}$, and using $\hat{\varepsilon}_{\psi(2S)K^+}$ from Tab. 4.7, $\hat{\varepsilon}_{\psi(2S)\pi^+}$ from Tab. 4.5 and the value of $N_{\psi(2S)\pi^+}$ from a fit with the $B_c^+ \to \psi(2S)(\mu^+\mu^-)K^+$ component omitted.

We include all these components in their respective channels.

4.2.3 Combinatorial Background

Combinatorial background is comprised of random combinations of tracks wrongly combined together to form a B_c^+ candidate. This is suppressed by the BDT selection described in Sec. 3.2.2, but the remnants must be fitted. This background is present over the full B_c^+-candidate mass range, and drops off slowly, following a nearly exponential distribution, towards higher masses.

Table 4.7: Misidentification rates (%) taking into account detector acceptance and selection including particle identification, given for each year and polarity. Values are given for both of the two working points (WP) from Table 3.5.

		$B_c^+ \to J/\psi K^+$		$B_c^+ \to \psi(2S)K^+$	
Year	Magnet polarity	$\pi^+\mu^+\mu^-$ WP	$\psi(2S)\pi^+$ WP	$\pi^+\mu^+\mu^-$ WP	$\psi(2S)\pi^+$ WP
2011	up	0.04	0.11	0.04	0.11
2011	down	0.04	0.11	0.03	0.10
2012	up	0.04	0.11	0.04	0.11
2012	down	0.04	0.12	0.04	0.11
2015	up	0.01	0.04	0.02	0.05
2015	down	0.02	0.05	0.02	0.05
2016	up	0.02	0.07	0.02	0.07
2016	down	0.02	0.07	0.02	0.07
2017	up	0.03	0.08	0.03	0.07
2017	down	0.03	0.08	0.03	0.07
2018	up	0.03	0.08	0.02	0.07
2018	down	0.03	0.08	0.03	0.08
Average ($\hat{\varepsilon}$)		0.03	0.08	0.03	0.08

Table 4.8: Expected yields for misidentified decays at each working point. The uncertainties take into account the uncertainty on $R_{J/\psi K^+/J/\psi\pi^+}$ and the statistical uncertainties on the yields of $J/\psi\pi^+$ and $\psi(2S)\pi^+$ candidates.

	$B_c^+ \to J/\psi K^+$		$B_c^+ \to \psi(2S)K^+$	
	$\pi^+\mu^+\mu^-$ WP	$\psi(2S)\pi^+$ WP	$\pi^+\mu^+\mu^-$ WP	$\psi(2S)\pi^+$ WP
Expected yield	28.7 ± 2.8	74.0 ± 7.2	1.1 ± 0.1	2.5 ± 0.3

4.3 Fitting

We determine the yields of non-resonant $B_c^+ \to \pi^+\mu^+\mu^-$ decays, and resonant $B_c^+ \to J/\psi\,\pi^+$ and $B_c^+ \to \psi(2S)\pi^+$ decays from extended maximum-likelihood fits to the unbinned distribution of the B_c^+-candidate mass $m(\pi^+\mu^+\mu^-)$.

For the resonant modes $B_c^+ \to J/\psi\,\pi^+$ and $B_c^+ \to \psi(2S)\pi$, we obtain nominal results using the B_c^+-candidate mass $m(\pi^+\mu^+\mu^-)$ calculated with J/ψ and $\psi(2S)$ constraints (such that the $\mu^+\mu^-$ mass is constrained within the appropriate mass regions), respectively, due to the improvement in resolution. For the resonant modes, we also perform fits using the unconstrained $m(\pi^+\mu^+\mu^-)$ mass as a crosscheck, both, for our tools and our results on the update channel.

For the non-resonant mode of course no such constraint is possible. The decay mode $B_c^+ \to J/\psi\,\pi^+$ serves as normalisation mode and also to obtain correction factors to account for data-simulation discrepancies associated with the signal peak position and width.

4.3.1 Fits for $B_c^+ \to \pi^+\mu^+\mu^-$ Non-resonant Mode

We perform five fits for the $B_c^+ \to \pi^+\mu^+\mu^-$ mode, one fit for each of the four q^2 bins (see Tab. 3.6), and one fit combining all four q^2 intervals.

We consider two components in each fit: signal decays and combinatorial background as we do not expect any other contributions in our fit range (see Sec. 4.2). We perform a study including an additional $B_c^+ \to \rho^+\mu^+\mu^-$ background component in App. F. The results of this study confirm that the possible contribution from this component is negligible and we therefore do not consider it further in this section.

Fits to Simulation

The B_c^+-mass Probability Density Functions (PDFs) for each component are modelled using empirical analytical functions. The PDF parameters for the signal are determined from fits to the distributions of the B_c^+ candidates reconstructed and selected in the full simulation samples.

We weight the full simulation samples to mirror the proportions for the various years and different polarities in the collision data.

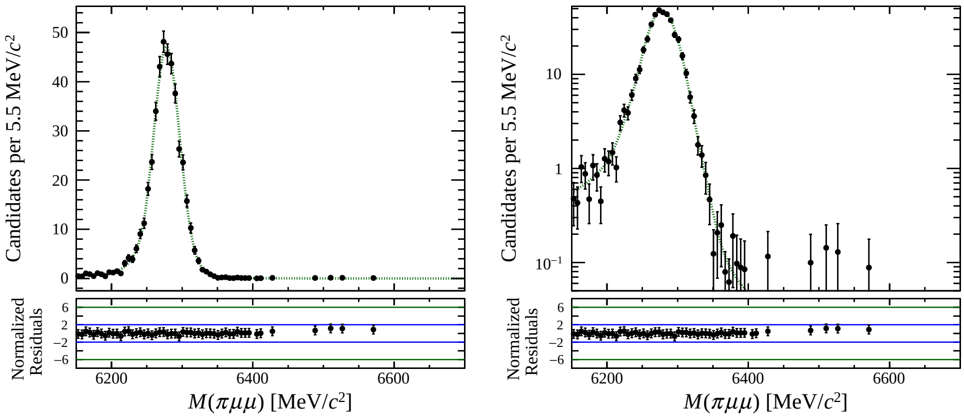
We model the B_c^+-mass distribution for the signal using the sum of a Gaussian with a double-sided Crystal Ball function. Figures 4.2 and 4.3 show the B_c^+ mass distributions in full simulation for the signal with PDF fit projections overlaid. Table 4.9 shows the PDF parameters obtained from the fits to simulation.

The B_c^+-mass PDFs for the signal includes a global shift of peak position and a global width scaling factor to account for possible data-simulation differences. The values for these parameters are fixed to 0 and 1 respectively in the fits to simulation (which will later be free but constrained in our fits to data from values from the normalisation mode (see Sec. 4.3.2)).

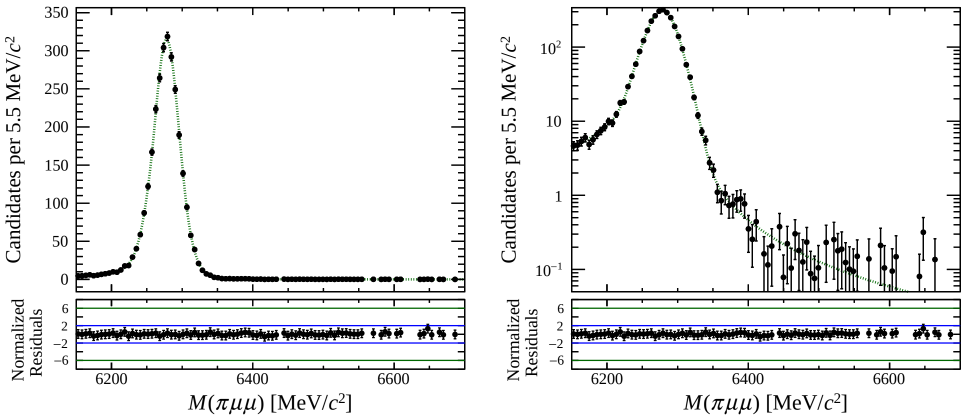
We model the combinatorial background B_c^+-candidate mass distribution using an exponential function. In the fit to data, the B_c^+-mass PDF exponent varies freely. In total, the signal-mode fit includes three free parameters (the yields for each component and the exponent of the combinatorial background) and two parameters with Gaussian constraints (the shift and scale factors).

Table 4.9: Signal PDF parameters obtained by fitting non-resonant $B_c^+ \to \pi^+\mu^+\mu^-$ simulation, in separate q^2 bins and combined. The parameters μ and σ are the mean and width of the Gaussian, while the double-sided Crystal Ball function has mean and width μ and $R\sigma$, respectively, and tail parameters $\alpha_{1,2}$, $n_{1,2}$; the parameter f_1 gives the fraction of the PDF in the double-sided Crystal Ball. Units of GeV^2/c^4 for q^2 are implied.

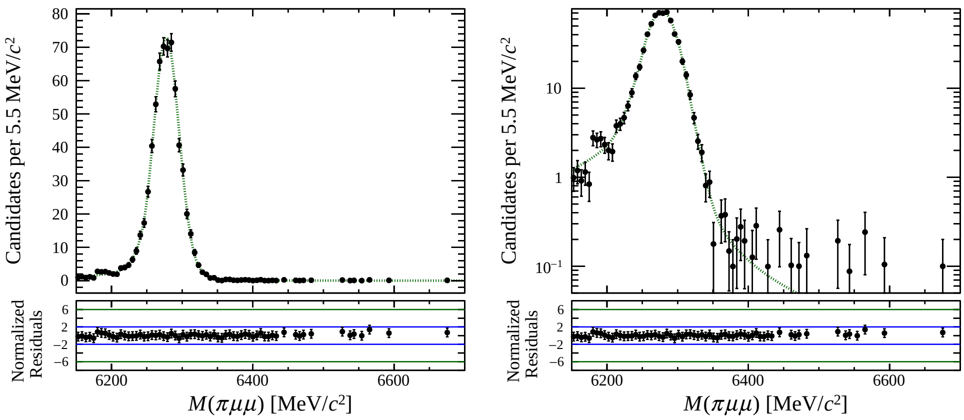
Fit parameters	$0.1 < q^2 < 1.1$	$1.1 < q^2 < 8$	$11 < q^2 < 12.5$	$15 < q^2 < 35$	All
α_1	-1.76 ± 0.88	-1.14 ± 0.68	-1.62 ± 0.57	-0.88 ± 0.70	-1.11 ± 0.51
α_2	2.51 ± 1.32	1.88 ± 0.62	2.34 ± 0.72	1.58 ± 0.55	1.81 ± 0.41
f_1	0.67 ± 0.77	0.37 ± 0.18	0.62 ± 0.38	0.35 ± 0.19	0.39 ± 0.15
$\mu\,[\mathrm{MeV}/c^2]$	6277.9 ± 1.3	6278.0 ± 0.7	6277.6 ± 1.0	6278.2 ± 0.6	6278.0 ± 0.5
n_1	1.14 ± 1.46	1.37 ± 0.83	1.00 ± 0.86	1.81 ± 1.21	1.44 ± 0.63
n_2	1.65 ± 2.05	1.99 ± 0.76	1.54 ± 1.06	3.22 ± 0.98	2.43 ± 0.58
$\sigma\,[\mathrm{MeV}/c^2]$	24.34 ± 8.81	20.45 ± 1.75	21.79 ± 5.24	19.28 ± 1.30	19.96 ± 1.25
R	0.71 ± 0.26	0.74 ± 0.13	0.78 ± 0.22	0.83 ± 0.14	0.80 ± 0.08



a) $0.1 < q^2 < 1.1\,\mathrm{GeV}^2/c^4$

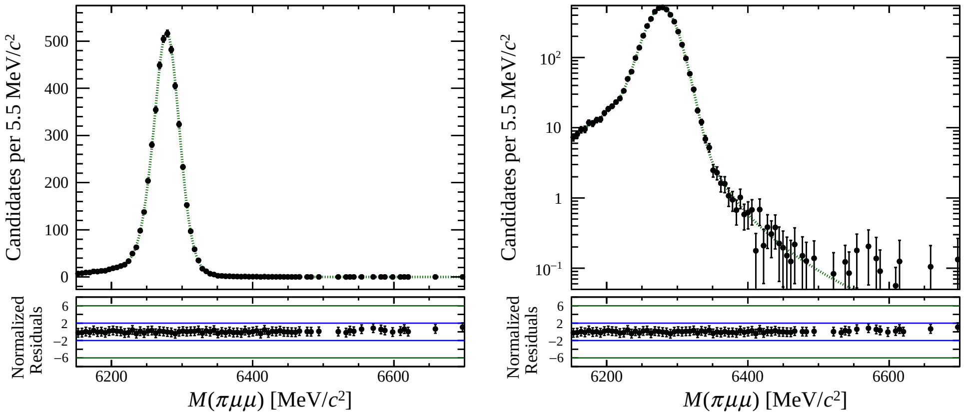


b) $1.1 < q^2 < 8\,\mathrm{GeV}^2/c^4$

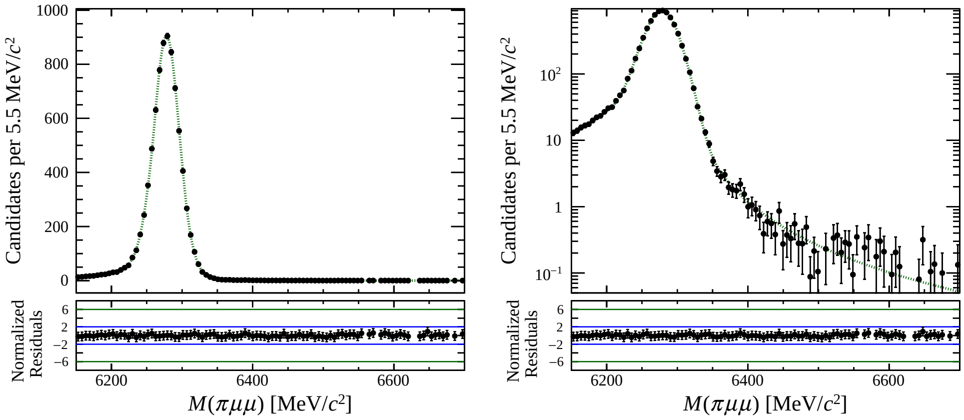


c) $11 < q^2 < 12.5\,\mathrm{GeV}^2/c^4$

Figure 4.2: Distributions of B_c^+-candidate mass for non-resonant $B_c^+ \to \pi^+\mu^+\mu^-$ simulation with fit projections overlaid. The distributions are shown for the four q^2 bins separately, in (left) linear and (right) log scales.



$15 < q^2 < 35\,\mathrm{GeV}^2/c^4$



All q^2 bins together

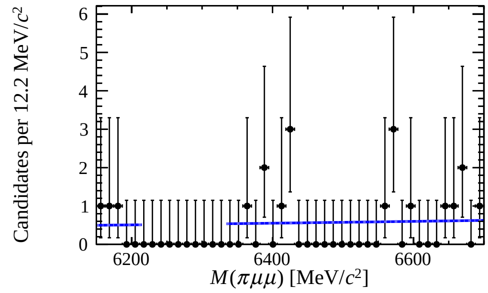
Figure 4.3: Distributions of B_c^+-candidate mass for non-resonant $B_c^+ \to \pi^+\mu^+\mu^-$ simulation with fit projections overlaid. The distribution is shown for the very low bin and all four q^2 bins together in (left) linear and (right) log scales.

Fit to Sideband Data

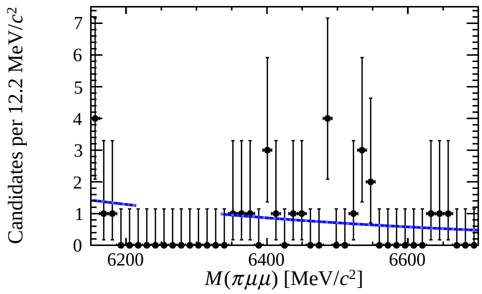
We estimate the expectations for the background by performing a maximum-likelihood fit of the combinatorial background PDF to the sideband data. Such fits are used in the selection optimization (Sec. 3.3), to obtain estimates for yields that can be used in pseudo-experiments prior to unblinding the signal region, and to set initial values in the final fits. To obtain the sideband data we apply the full selection to data and take only the candidates that are outside the signal region $6215 < m(\pi^+\mu^+\mu^-) < 6335\,\mathrm{MeV}/c^2$, which corresponds to an interval with a width of about ±3 times the expected B_c^+ signal resolution and centered at the B_c^+ mass.

Figure 4.4 shows the sideband data with the fit projection overlaid where good agreement between the fit model and the sideband data is achieved. Table 4.10 shows the results for the PDF exponent and the number of candidates in the fit sample.

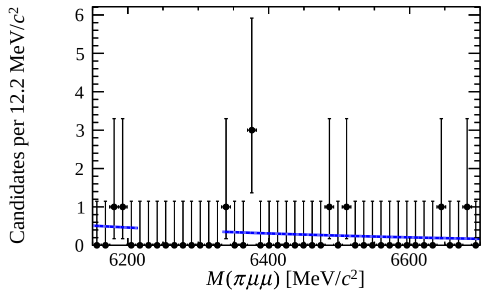
To estimate the background expectation, we interpolate the background PDF to obtain the fraction of background in the signal region. The integral of the normalized background PDF in the signal region is about 20%, *i.e.* the yield in Table 4.10 corresponds to about 80% of the total expected combinatorial background.



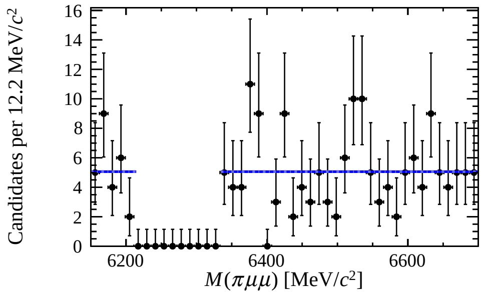
$0.1 < q^2 < 1.1\,\mathrm{GeV}^2/c^4$



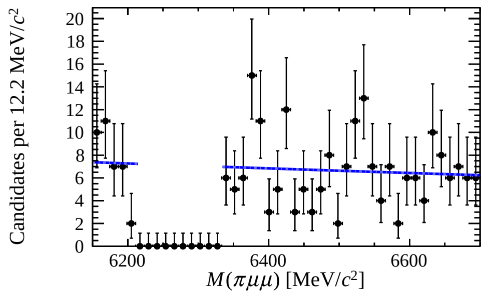
$1.1 < q^2 < 8\,\mathrm{GeV}^2/c^4$



$11 < q^2 < 12.5\,\mathrm{GeV}^2/c^4$



$15 < q^2 < 35\,\mathrm{GeV}^2/c^4$



All q^2 bins

Figure 4.4: Distributions of B_c^+-candidate mass in sideband data for the three q^2 bins and for all q^2 bins combined. Projection of maximum-likelihood fits are overlaid.

Table 4.10: Results for the combinatorial background exponent obtained from a fit to sideband data and number of candidates in the sideband data.

q^2-bin $[\text{GeV}^2/c^4]$	Comb. bkg. exponent $[\text{MeV}^{-1}]$	Candidates
$0.1 - 1.1$	$(-0.4 \pm 1.5) \cdot 10^{-3}$	20
$1.1 - 8$	$(-2.0 \pm 1.2) \cdot 10^{-3}$	28
$11 - 12.5$	$(-2.0 \pm 1.9) \cdot 10^{-3}$	10
$15 - 35$	$(-0.0 \pm 0.5) \cdot 10^{-3}$	178
All bins	$(-0.3 \pm 0.4) \cdot 10^{-3}$	236

Fit Estimator Properties

To assess the estimator properties of our fits, we perform a series of experiments using pseudo-data samples. The pseudo-data samples are produced with the expected combinatorial background, and without or with signal decays. The pseudo-experiments are performed for each of the three q^2 bins, and combining all three bins.

We perform pseudo-experiments for various expectations of signal decays, $N_{\text{sig}}^{\text{true}} = 0, 5, 10, 20, 30, 40$ and 50. The expected numbers of candidates for the combinatorial background is obtained by interpolating the results of the fit to sideband data (Table 4.10). We use Poisson distributions to model the number of events for each component in each pseudo-data sample.

The expectation for the combinatorial PDF slope is also obtained from the results of the fit to sideband data (Table 4.10). The expectations for the global shift of peak position and the width scaling factor are obtained from the fit to data in Sec. 4.3.2 for the $\pi^+\mu^+\mu^-$ working point.

For each pseudo-data sample, we perform two fits of the full model including signal and the two backgrounds, one fit where the signal yield varies, and one fit where it is fixed to zero. The fit with signal yield fixed to zero is used to estimate the significance of the fit results based on the Wilks' theorem [31].

We perform 1000 pseudo-experiments and calculate residuals and pull distributions for the fit parameters from the results of the fit with varying signal yield. The fits to the B_c^+ mass distribution converge for all pseudo-experiments. We then fit a Gaussian function to the pull distributions to obtain the mean and the width of the distributions. Figures 4.6–4.8 show the pull distributions for all fit parameters combining the three q^2 bins.

We generally observe unbiased distributions and proper uncertainty estimation for all fit parameters. In App. H, we provide pull plots for all free parameters

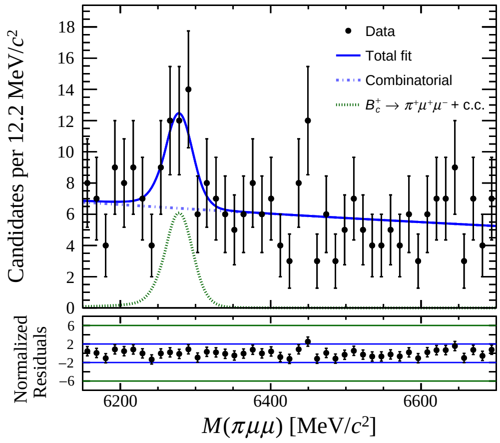


Figure 4.5: Example B_c^+-candidate mass distribution for a simulated pseudo-data sample generated with signal events combining all three q^2 bins. The number of signal events is drawn from a Poisson distribution with expectation equal to $N_{\text{sig}}^{\text{true}} = 30$.

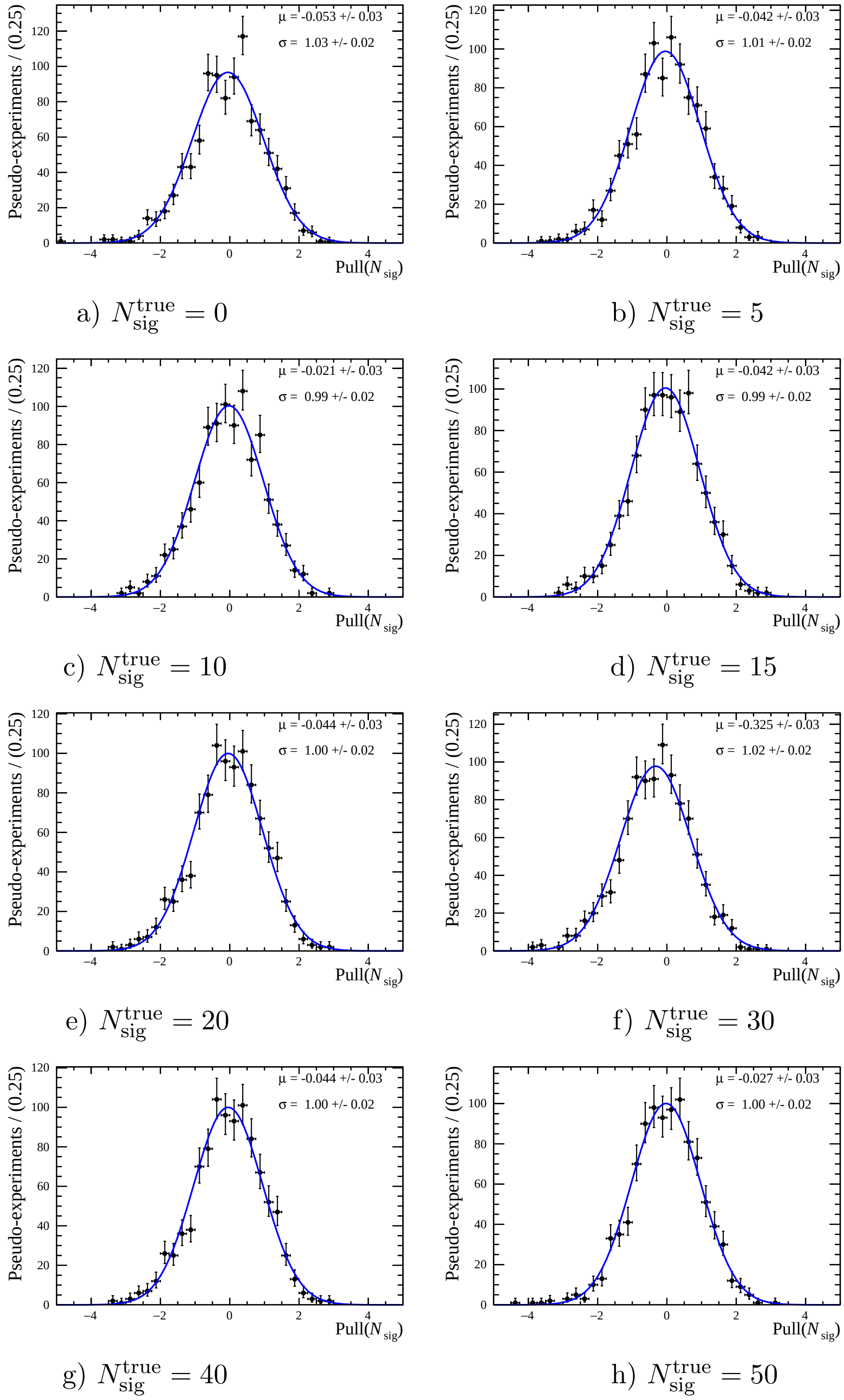
for the q^2 intervals separately. We note that we do not expect symmetric pull distributions for the cases where both signal and background are very small due to the behaviour of extended maximum-likelihood fits in this regime [32, 33].

Upon unblinding, we estimate the significance of the measurement only when the fit converges into a signal yield above zero. For the sake of completeness, we calculate the significance also for pseudo-experiments for which the fit converges into a negative signal yield value.

Figure 4.9 shows the p-value distributions obtained from the pseudo-experiments. They are calculated as `Prob(`$2 \cdot \Delta \log \mathcal{L}$`, 1)`, where `Prob` is a probability assuming a χ^2 distribution for one degree of freedom and $\Delta \log \mathcal{L}$ is the difference between the logarithms of the likelihoods for the fit with signal yield varying and with signal yield fixed to zero. Table 4.11 shows the fraction $f_{3\sigma}$ of pseudo-experiments in which the fit converged into a positive signal yield with a significance of or above 3σ, that is p-value $< 2.7 \cdot 10^{-3}$. The p-value distributions for each q^2 bin are provided in App. H.

Table 4.11: Fraction of pseudo-experiments (%) in which a significance of 3σ or more is achieved.

$N_{\text{sig}}^{\text{true}}$	0	5	10	15	20	30	40	50
$f_{3\sigma}$	0.2	0.6	4.0	10.6	24.2	58.9	87.7	97.6



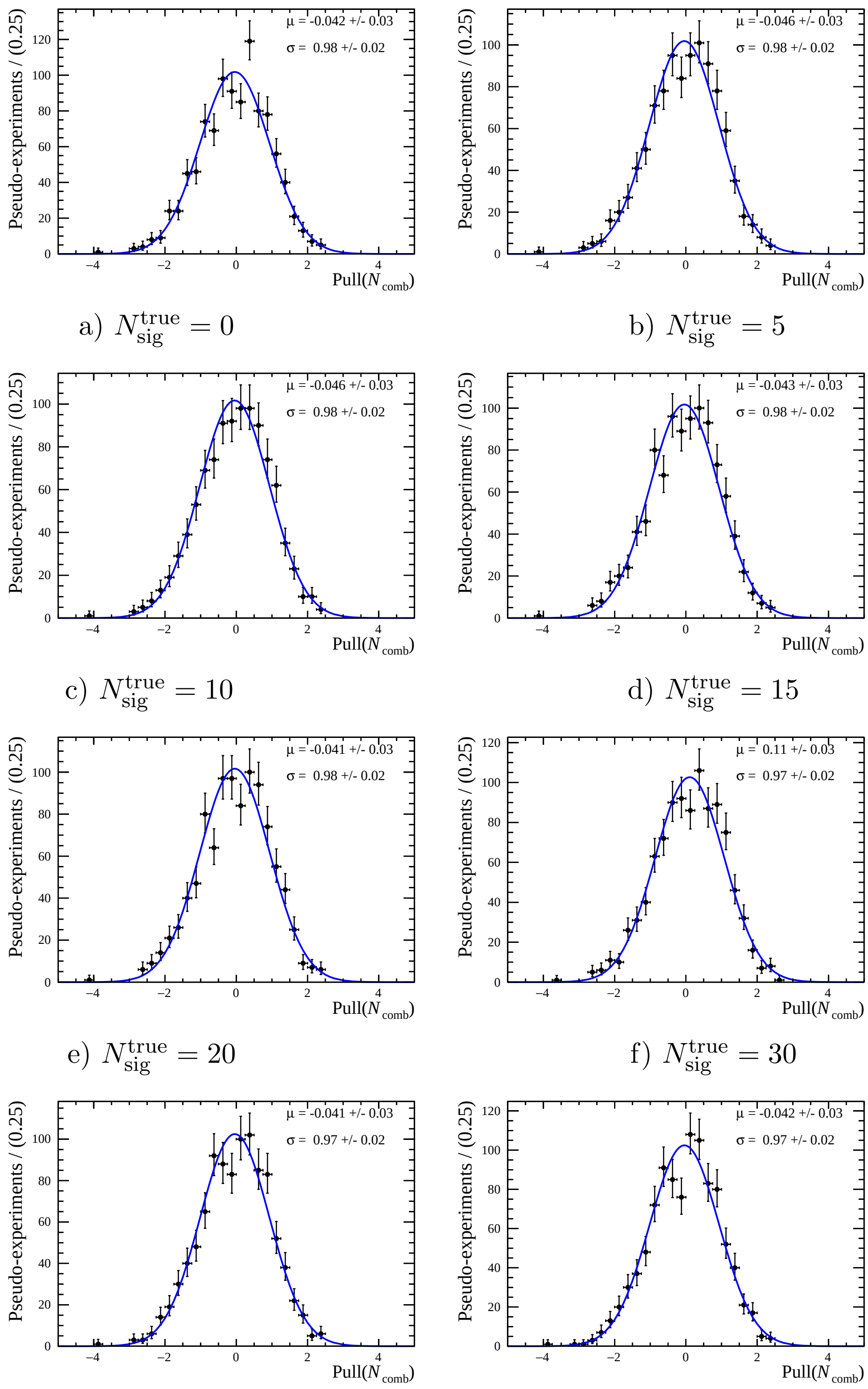
a) $N_{\text{sig}}^{\text{true}} = 0$ b) $N_{\text{sig}}^{\text{true}} = 5$

c) $N_{\text{sig}}^{\text{true}} = 10$ d) $N_{\text{sig}}^{\text{true}} = 15$

e) $N_{\text{sig}}^{\text{true}} = 20$ f) $N_{\text{sig}}^{\text{true}} = 30$

g) $N_{\text{sig}}^{\text{true}} = 40$ h) $N_{\text{sig}}^{\text{true}} = 50$

Figure 4.6: Distributions of pulls for the signal yield obtained from fits to simulated pseudo-data samples. The number of signal events in each pseudo-data sample is drawn from a Poisson distribution with expectation equal to $N_{\text{sig}}^{\text{true}}$.



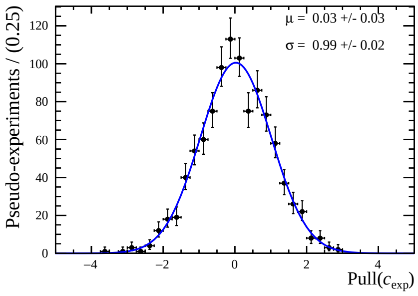
a) $N_{\mathrm{sig}}^{\mathrm{true}} = 0$ b) $N_{\mathrm{sig}}^{\mathrm{true}} = 5$

c) $N_{\mathrm{sig}}^{\mathrm{true}} = 10$ d) $N_{\mathrm{sig}}^{\mathrm{true}} = 15$

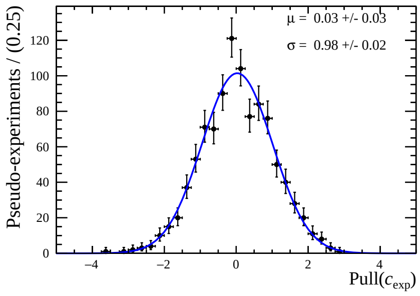
e) $N_{\mathrm{sig}}^{\mathrm{true}} = 20$ f) $N_{\mathrm{sig}}^{\mathrm{true}} = 30$

g) $N_{\mathrm{sig}}^{\mathrm{true}} = 40$ h) $N_{\mathrm{sig}}^{\mathrm{true}} = 50$

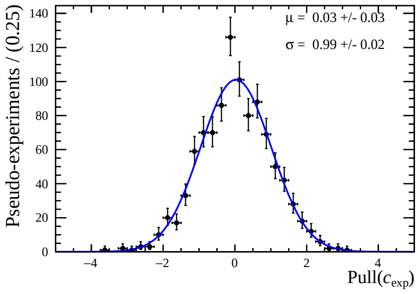
Figure 4.7: Distributions of pulls for the combinatorial background yield obtained from fits to simulated pseudo-data samples. The number of signal events in each pseudo-data sample is drawn from a Poisson distribution with expectation equal to $N_{\mathrm{sig}}^{\mathrm{true}}$.



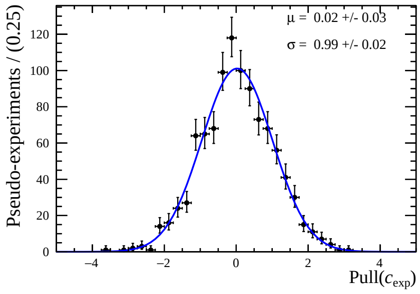
a) $N_{\rm sig}^{\rm true} = 0$



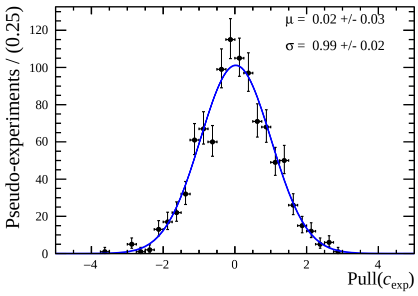
b) $N_{\rm sig}^{\rm true} = 5$



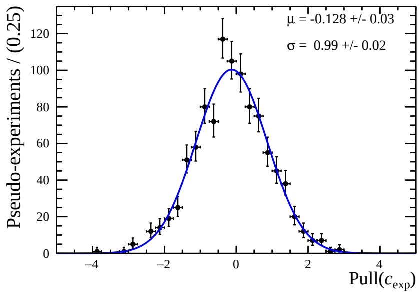
c) $N_{\rm sig}^{\rm true} = 10$



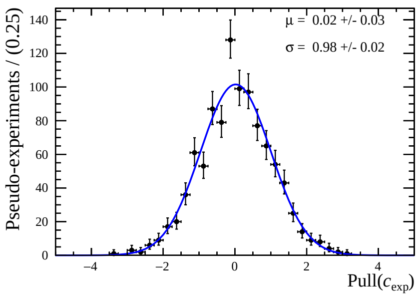
d) $N_{\rm sig}^{\rm true} = 15$



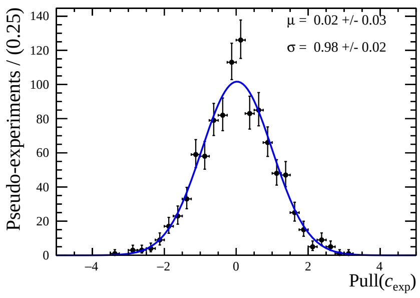
e) $N_{\rm sig}^{\rm true} = 20$



f) $N_{\rm sig}^{\rm true} = 30$



g) $N_{\rm sig}^{\rm true} = 40$



h) $N_{\rm sig}^{\rm true} = 50$

Figure 4.8: Distributions of pulls for the combinatorial PDF exponent obtained from fits to simulated pseudo-data samples. The number of signal events in each pseudo-data sample is drawn from a Poisson distribution with expectation equal to $N_{\rm sig}^{\rm true}$.

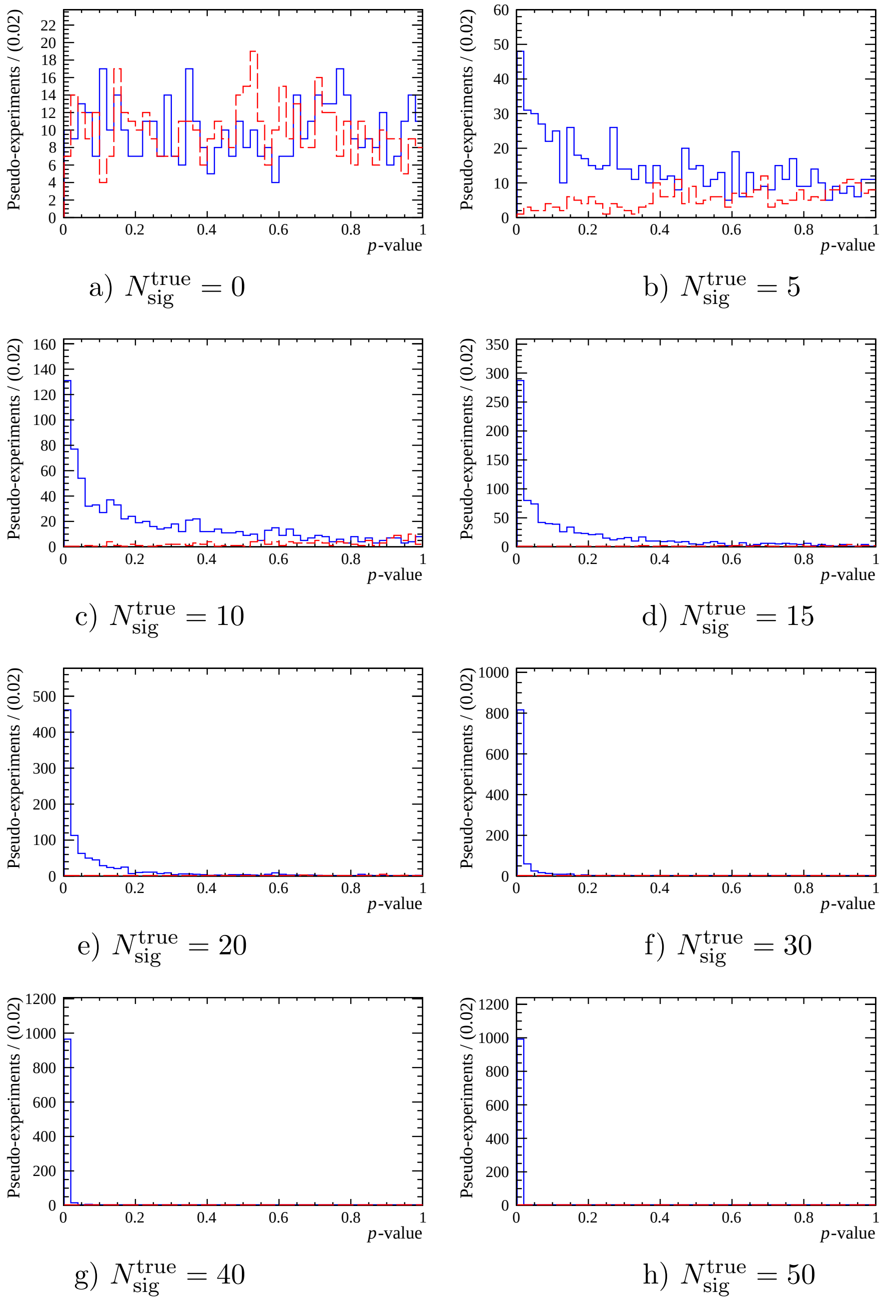


Figure 4.9: Distributions of p-values obtained from the difference in $2 \cdot \log \mathcal{L}$ between fits performed with fixed and with varying signal yield to simulated pseudo-data samples. The number of signal events in each pseudo-data sample is drawn from a Poisson distribution with expectation equal to $N_{\rm sig}^{\rm true}$. The solid blue line shows the distributions for pseudo-experiments where the nominal fit converges to a positive signal yield and the dashed red curve shows the distributions for those where the nominal fit converges to a negative signal yield.

4.3.2 Fits for $B_c^+ \to J/\psi\,\pi^+$ Normalization mode

We consider four components in the fit models for the normalisation mode:

- $B_c^+ \to J/\psi\,(\mu^+\mu^-)\,\pi^+$ decays
- misidentified $B_c^+ \to J/\psi\,(\mu^+\mu^-)\,K^+$ decays
- partially reconstructed $B_c^+ \to J/\psi\,(\mu^+\mu^-)\,\rho^+$ decays
- combinatorial background

The PDFs for the B_c^+-candidate mass fits for each component are modelled using empirical analytical functions similar to the ones used in the signal-mode fit. The misidentified background is modelled using the sum of a Gaussian with a double-sided Crystal Ball function, *i.e.* the same analytic function as the signal, but of course with different parameters; the partially reconstructed background is modelled using the sum of a Gaussian with a single-sided Crystal Ball function.

Fits to Simulation

The PDF parameters for the $B_c^+ \to J/\psi\,(\mu^+\mu^-)\,\pi^+$, $B_c^+ \to J/\psi\,(\mu^+\mu^-)\,K^+$ and $B_c^+ \to J/\psi\,(\mu^+\mu^-)\,\rho^+$ components are determined from fits to the distributions of the B_c^+ candidates reconstructed and selected in the full simulation samples, weighted to mirror the relative proportions for the various years and different polarities in data. We select B_c^+ candidates that are associated with the generated B_c^+ decay.

The fit is performed for the two selection working points defined in Table 3.5. For the $B_c^+ \to \psi(2S)\pi^+$ working point, the fit is performed two times, first using as fit observable the B_c^+-candidate mass calculated with J/ψ-mass constraint applied to the dimuon system, then, as a cross check, using the B_c^+-candidate mass calculated without constraint.

Figures 4.10–4.12 show the B_c^+ mass distributions in full simulation for the $B_c^+ \to J/\psi\,(\mu^+\mu^-)\,\pi^+$, the $B_c^+ \to J/\psi\,(\mu^+\mu^-)\,K^+$ and the $B_c^+ \to J/\psi\,(\mu^+\mu^-)\,\rho^+$ components with PDF fit projections overlaid. Tables 4.12–4.14 show the PDF parameters obtained from the fits. The PDFs for these three components share a global shift of peak position and a width scaling factor. The values for the shift of peak position and width scaling factor are fixed respectively to 0 and 1 in the fits to simulation and left free in the fits to data.

The possible remaining combinatorial background is modelled using an exponential function, whose exponent is free in the fits to data. In total, the fit model includes seven free parameters: the yields for each component, the global peak shift and the width scaling factor, and the exponent of the combinatorial background.

Table 4.12: Signal PDF parameters (defined as in Table 4.9) obtained by fitting $B_c^+ \to J/\psi\,\pi^+$ simulation.

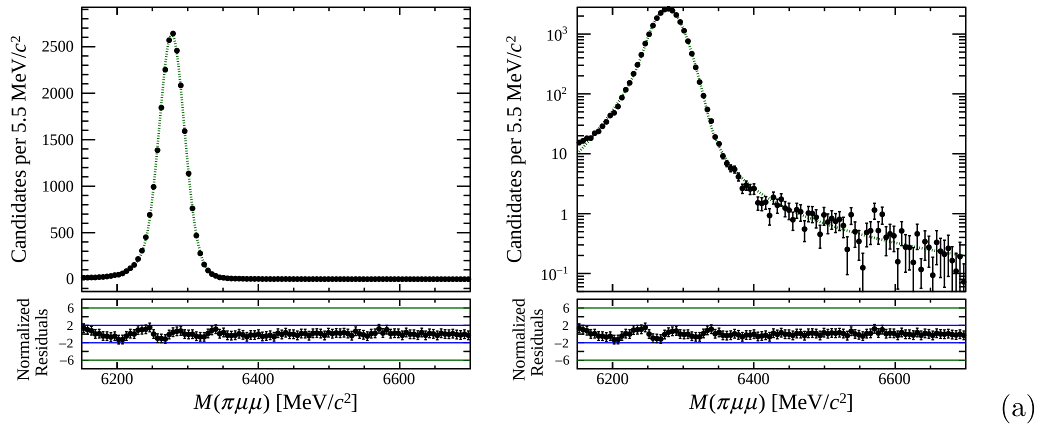
Fit parameters	$\pi^+\mu^+\mu^-$ working point	$\psi(2S)\pi^+$ working point	
	Unconstr. $\mu^+\mu^-$ mass	Unconstr. $\mu^+\mu^-$ mass	Constr. $\mu^+\mu^-$ mass
α_1	-1.26 ± 0.58	-1.56 ± 0.07	-1.81 ± 0.11
α_2	2.54 ± 0.24	2.60 ± 0.06	2.30 ± 0.06
f_1	0.53 ± 0.37	0.81 ± 0.08	0.46 ± 0.09
$\mu\,[\mathrm{MeV}/c^2]$	6277.5 ± 0.2	6277.5 ± 0.1	6276.7 ± 0.1
n_1	5.92 ± 7.09	3.53 ± 0.47	1.35 ± 0.12
n_2	1.67 ± 0.26	1.77 ± 0.17	1.76 ± 0.13
$\sigma\,[\mathrm{MeV}/c^2]$	15.9 ± 2.69	12.9 ± 1.31	9.6 ± 0.42
R	1.41 ± 0.07	1.61 ± 0.13	1.64 ± 0.04

Table 4.13: Misidentified background PDF parameters (defined as in Table 4.9) obtained by fitting $B_c^+ \to J/\psi\,K^+$ simulation. The parameter n_1 is fixed to 10 for this decay mode to be treated as a very large factor in the DCB PDF.

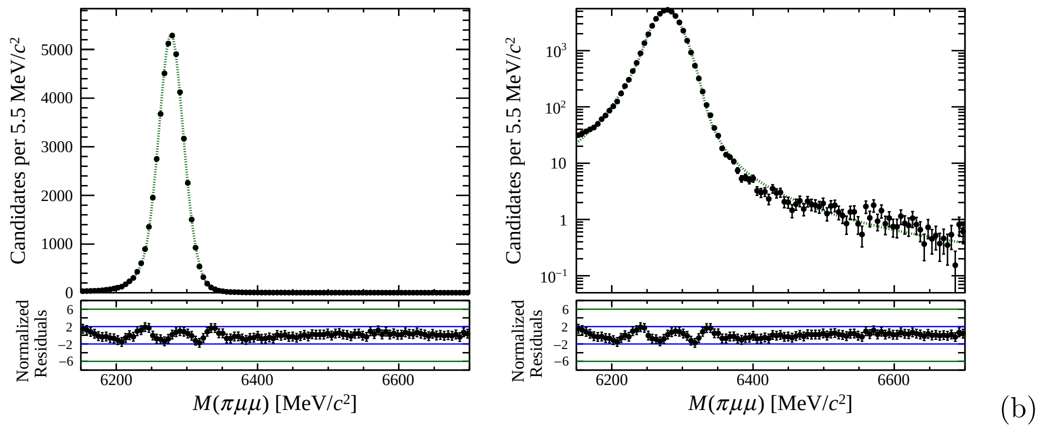
Fit parameters	$\pi^+\mu^+\mu^-$ working point	$\psi(2S)\pi^+$ working point	
	Unconstr. $\mu^+\mu^-$ mass	Unconstr. $\mu^+\mu^-$ mass	Constr. $\mu^+\mu^-$ mass
α_1	-0.56 ± 0.35	-0.58 ± 0.23	-0.77 ± 0.48
α_2	2.17 ± 0.70	2.26 ± 0.41	2.17 ± 0.43
f_1	0.41 ± 0.25	0.47 ± 0.18	0.50 ± 0.21
$\mu\,[\mathrm{MeV}/c^2]$	6242.9 ± 1.4	6242.6 ± 0.9	6241.6 ± 0.7
n_1	10.00 ± 0.00	10.00 ± 0.00	2.94 ± 2.51
n_2	1.87 ± 1.24	1.79 ± 0.70	1.81 ± 0.53
$\sigma\,[\mathrm{MeV}/c^2]$	22.79 ± 1.82	23.01 ± 1.30	16.33 ± 1.42
R	0.80 ± 0.25	0.81 ± 0.14	0.85 ± 0.14

Table 4.14: Partially reconstructed background PDF parameters obtained by fitting $B_c^+ \to J/\psi\,\rho^+$ simulation using a double Gaussian function. The first Gaussian has mean and width corresponding to the parameters μ and σ, while the second Gaussian has mean and the width corresponding to μ and $R\sigma$. The parameter f_1 gives the fraction of the first Gaussian function in the PDF.

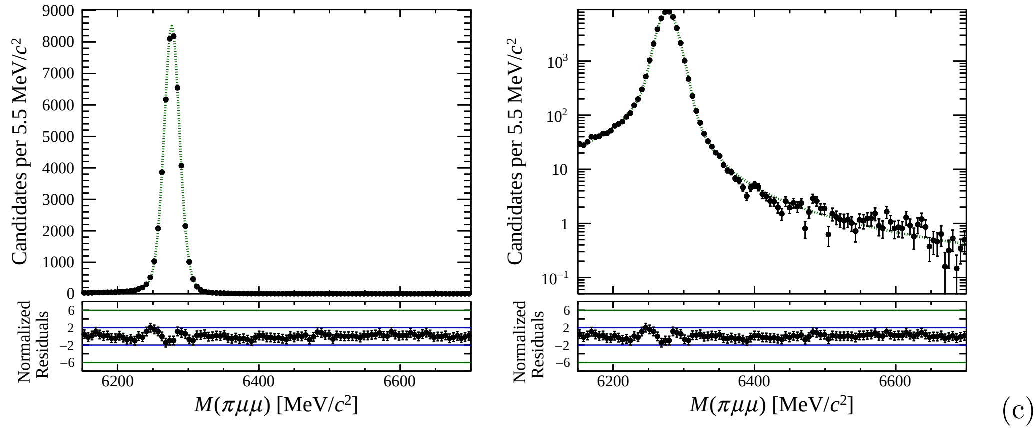
Fit parameters	$\pi^+\mu^+\mu^-$ working point	$\psi(2S)\pi^+$ working point	
	Unconstr. $\mu^+\mu^-$ mass	Unconstr. $\mu^+\mu^-$ mass	Constr. $\mu^+\mu^-$ mass
f_1	0.82 ± 0.11	0.84 ± 0.08	0.65 ± 0.14
$\mu\,[\mathrm{MeV}/c^2]$	5975.2 ± 19.3	5974.6 ± 27.7	5974.5 ± 18.6
$\sigma\,[\mathrm{MeV}/c^2]$	48.36 ± 16.0	48.65 ± 6.42	42.87 ± 9.65
R	4.95 ± 2.26	5.11 ± 1.66	5.66 ± 1.79



$B_c^+ \to \pi^+\mu^+\mu^-$ working point, unconstrained dimuon mass

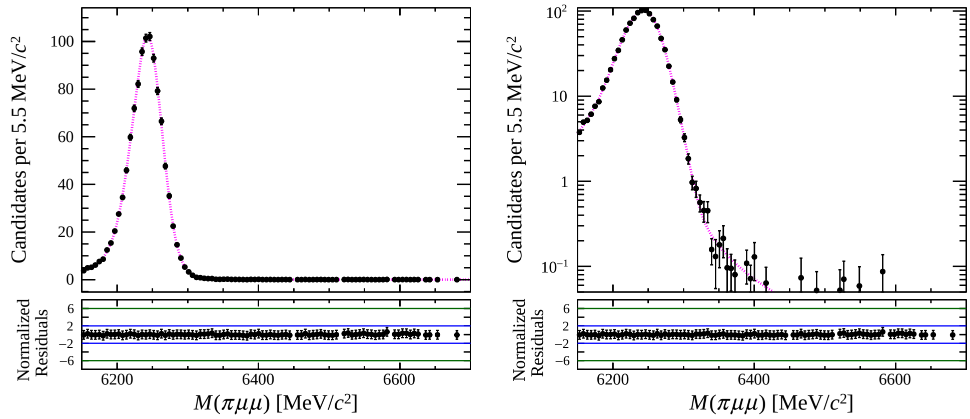


$B_c^+ \to \psi(2S)\pi^+$ working point, unconstrained dimuon mass



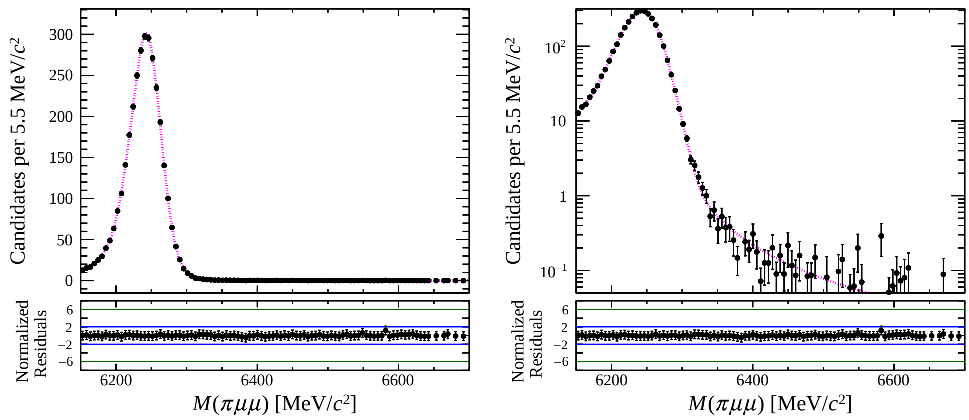
$B_c^+ \to \psi(2S)\pi^+$ working point, constrained dimuon mass

Figure 4.10: Distributions of B_c^+-candidate mass for $B_c^+ \to J/\psi\,\pi^+$ simulation with fit projections overlaid. The distributions are shown in (left) linear and (right) log scales.



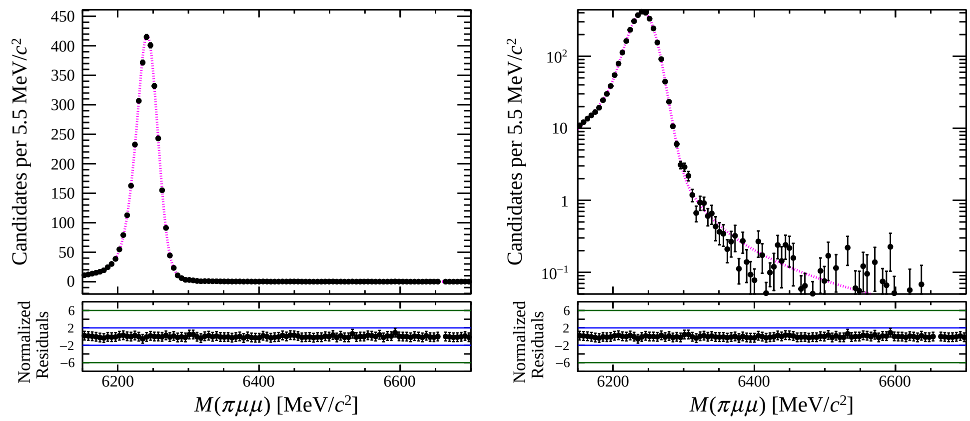
(a)

$B_c^+ \to \pi^+\mu^+\mu^-$ working point, unconstrained dimuon mass



(b)

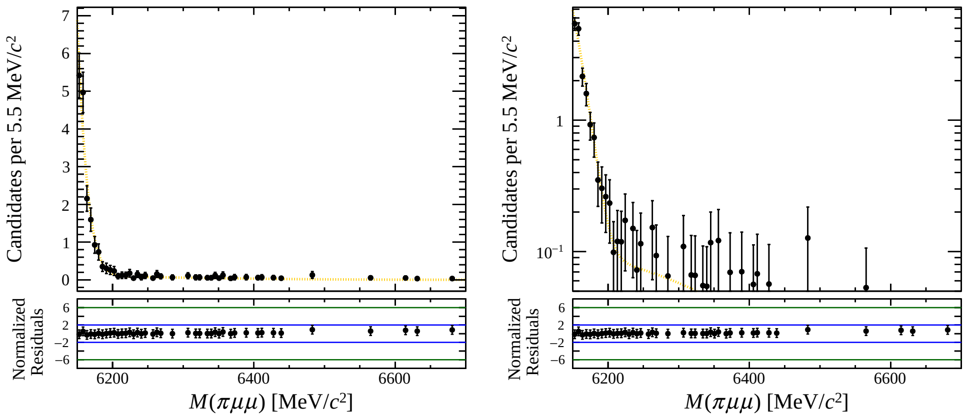
$B_c^+ \to \psi(2S)\pi^+$ working point, unconstrained dimuon mass



(c)

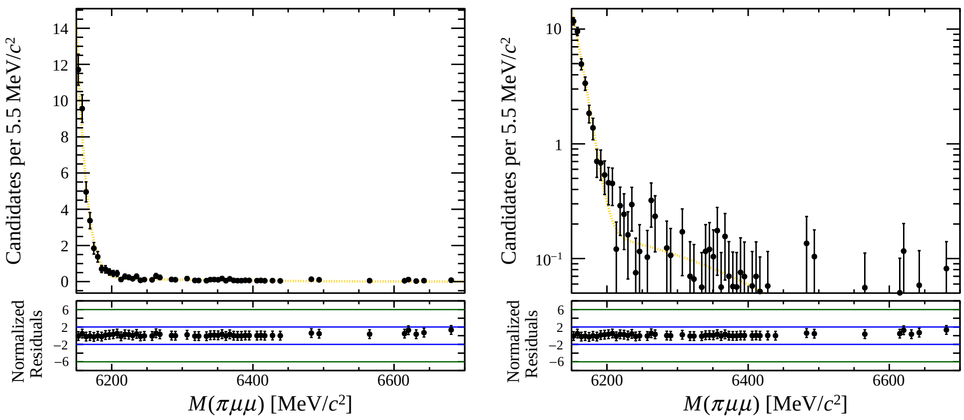
$B_c^+ \to \psi(2S)\pi^+$ working point, constrained dimuon mass

Figure 4.11: Distributions of B_c^+-candidate mass for $B_c^+ \to J/\psi\, K^+$ simulation with fit projections overlaid. The distributions are shown in (left) linear and (right) log scales.



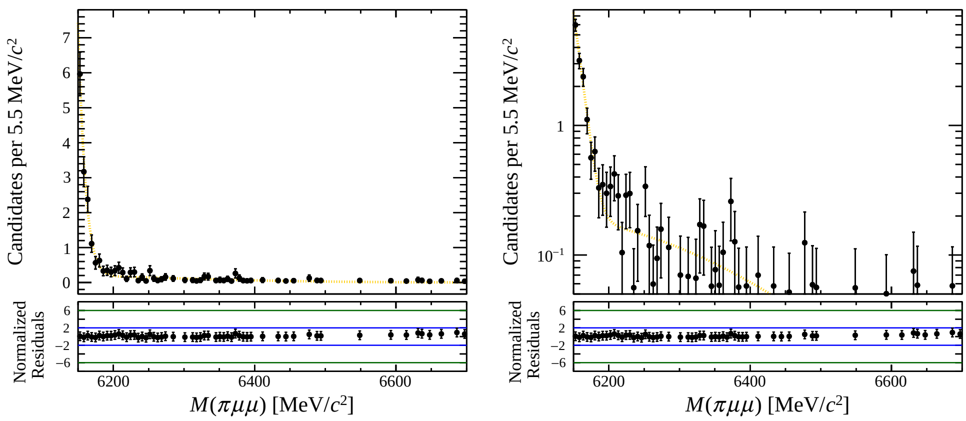
(a)

$B_c^+ \to \pi^+\mu^+\mu^-$ working point, unconstrained dimuon mass



(b)

$B_c^+ \to \psi(2S)\pi^+$ working point, unconstrained dimuon mass



(c)

$B_c^+ \to \psi(2S)\pi^+$ working point, constrained dimuon mass

Figure 4.12: Distributions of B_c^+-candidate mass for $B_c^+ \to J/\psi\,\rho^+$ simulation with fit projections overlaid. The distributions are shown in (left) linear and (right) log scales.

Fits to Data

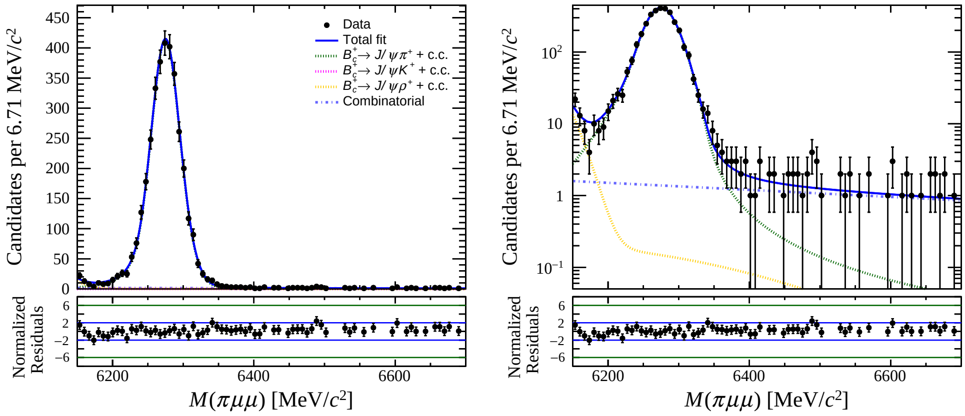
We fit the model for the normalisation mode described above to the $B_c^+ \to J/\psi\,\pi^+$ candidates in the full data sample. Table 4.15 shows the fit results. Figure 4.13 shows the full data sample with fit projections overlaid.

We note that the yield of the misidentified background from $B_c^+ \to J/\psi\,K^+$ decays is found to be slightly negative (consistent with zero at $< 1.5\sigma$) in the fit at the $\pi^+\mu^+\mu^-$ working point. The expectation for this quantity, as given in Table 4.8 is 26.7 ± 3.5, and therefore the yield is also $< 2\sigma$ below the expectation. Since the yield at the $\psi(2S)$ working point is consistent with the expectation of 78 ± 10, and the overall agreement of the data with the fit result is excellent, this does not appear to indicate any missing component in the fit. As such, we attribute it to a fluctuation.

Another small region of disagreement between the data and the results of the fit is at the lowest values of B_c^+-candidate mass in the fit at the $\pi^+\mu^+\mu^-$ working point. This is the region populated by the $B_c^+ \to J/\psi\,\rho^+$ partially reconstructed background, and as such may be due to mis-modelling of this background. As noted in Sec. 4.2.1, the polarisation of the final state in this decay is unknown, and can affect the shape of the component in the fit. This is accounted for as a source of systematic uncertainty, and therefore not considered further here.

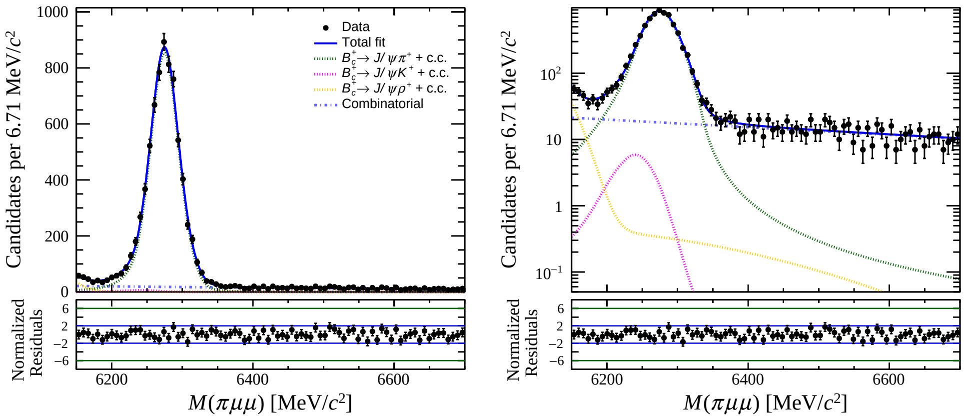
Table 4.15: Results of the fit to the B_c^+-candidate mass distribution of $B_c^+ \to J/\psi\,\pi^+$ candidates in data. The combinatorial background exponent ("Comb. bkg. exp.") is quoted in units of $(\,\mathrm{MeV}/c^2)^{-1}$.

Fit parameters	$\pi^+\mu^+\mu^-$ working point	$\psi(2S)\pi^+$ working point	
	Unconstr. $\mu^+\mu^-$ mass	Unconstr. $\mu^+\mu^-$ mass	Constr. $\mu^+\mu^-$ mass
$J/\psi\,\pi^+$ yield	3508 ± 82	6926 ± 131	6887 ± 93
$J/\psi\,K^+$ yield	-81 ± 58	58 ± 110	90 ± 43
$J/\psi\,\rho^+$ yield	41 ± 11	99 ± 22	56 ± 22
Scale	1.17 ± 0.03	1.13 ± 0.02	1.14 ± 0.02
Shift $[\,\mathrm{MeV}/c^2]$	-3.17 ± 0.59	-2.66 ± 0.47	-1.93 ± 0.20
Comb. bkg. exp.	-0.0014 ± 0.0014	-0.0014 ± 0.0003	-0.0014 ± 0.0003
Comb. bkg. yield	101 ± 25	1242 ± 69	1254 ± 60



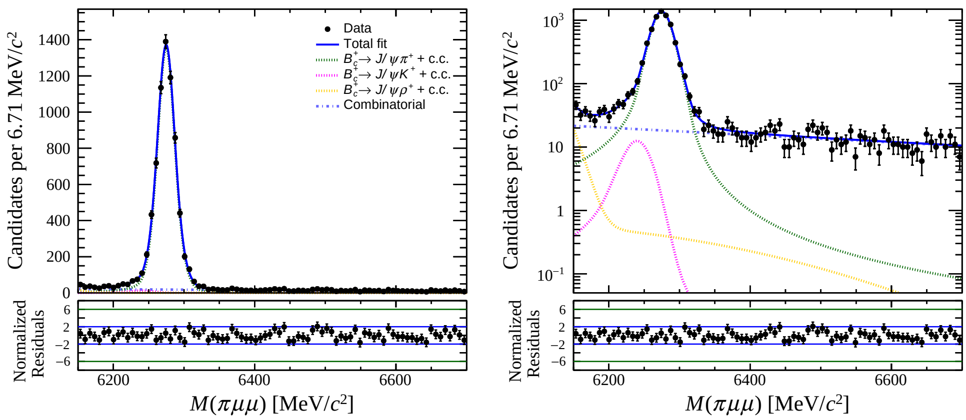
(a)

$\pi^+\mu^+\mu^-$ working point, unconstrained dimuon mass



(b)

$\psi(2S)\pi^+$ working point, unconstrained dimuon mass



(c)

$\psi(2S)\pi^+$ working point, constrained dimuon mass

Figure 4.13: Distributions of B_c^+-candidate mass for $B_c^+ \to J/\psi\,\pi$ candidates reconstructed in collision data with fit projections overlaid. The distributions are shown in (left) linear and (right) log scales.

Fit Estimator Properties

We perform pseudo-experiments to assess the fit estimator properties. The pseudo-data samples are generated with the expected four components: $B_c^+ \to J/\psi\,\pi^+$ decays, misidentified $B_c^+ \to J/\psi\,K^+$ decays, partially reconstructed $B_c^+ \to J/\psi\,\rho^+$ decays and combinatorial background.

The pseudo-data samples are generated from the PDF shapes obtained from simulation in Sec. 4.3.2 for the B_c^+-decay components. The expectations for the shape of the combinatorial background, the number of candidates for each component, the global shift of peak position, and the width scaling factor are obtained from the fit to data in Sec. 4.3.2 for the $\psi(2S)\pi^+$ working point. We use Poisson distributions to model the number of events for each component in each pseudo-data sample.

For each pseudo-data sample, we perform a fit of the full model as in Sec. 4.3.2. We perform 1000 pseudo-experiments and calculate the pull distributions for each fit parameter. The fit converges for all pseudo-experiments. We fit a Gaussian function to the pull distributions to obtain the mean and the width of the distributions.

Figure 4.14 shows the pull distribution for each fit parameter. We generally observe unbiased distributions and proper uncertainty estimation for all fit parameters.

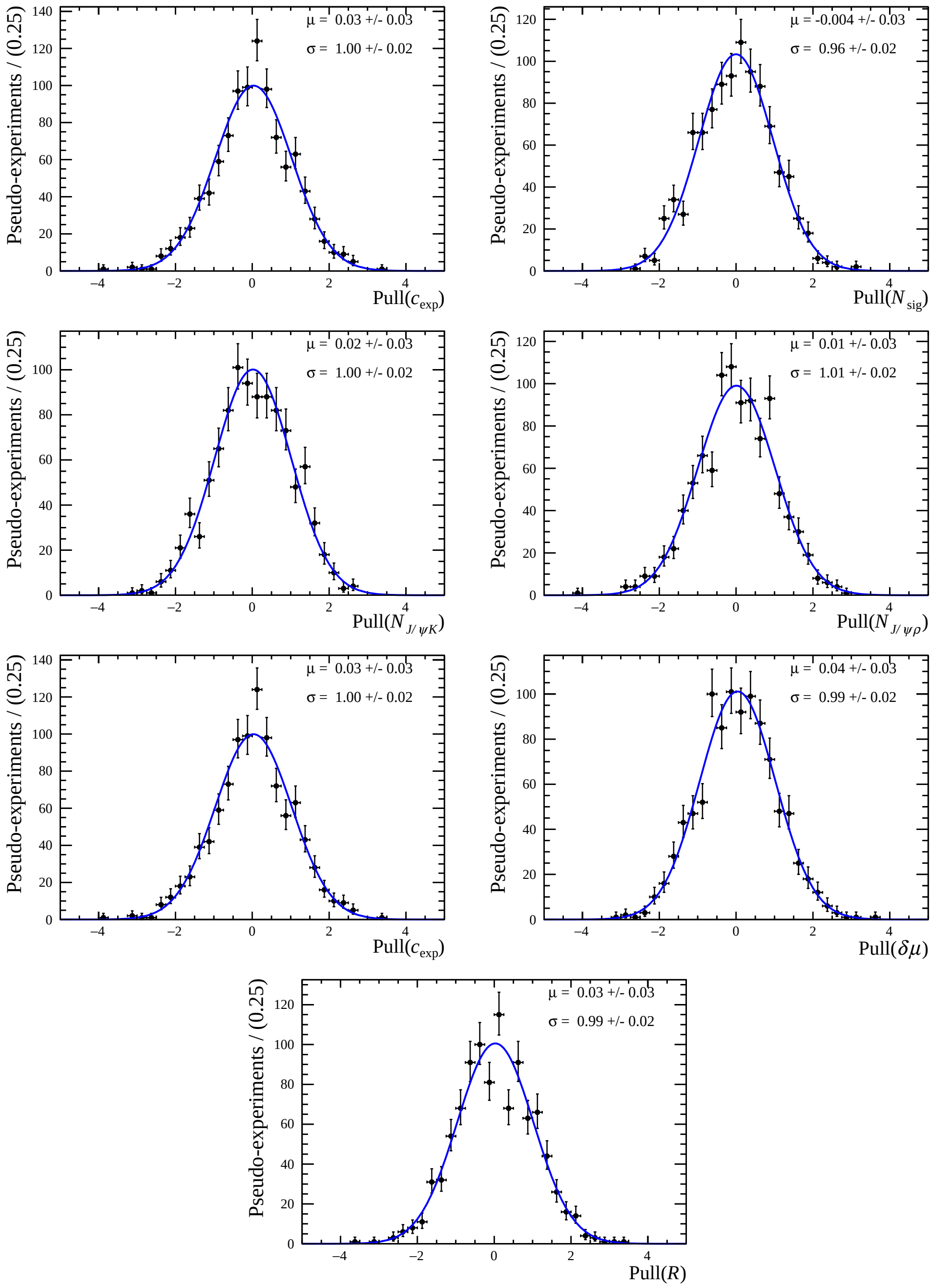


Figure 4.14: Pull distributions obtained from fits to pseudo-data samples with $B_c^+ \to J/\psi \pi^+$ candidates. The pull distributions are shown for all fit parameters: N_i represents the yield of the component in the respective subscript, $c_{\rm exp}$ the exponent of the combinatorial background PDF, $\delta\mu$ the global shift of peak position, and R the width scaling factor.

4.3.3 Fits for $B_c^+ \to \psi(2S)\pi^+$ Resonant mode

The fit for the $B_c^+ \to \psi(2S)\pi^+$ mode is performed in a similar way as the normalisation-mode fit including four components in the fit model:

- $B_c^+ \to \psi(2S)(\mu^+\mu^-)\,\pi^+$ decays,
- misidentified $B_c^+ \to \psi(2S)(\mu^+\mu^-)K^+$ decays,
- partially reconstructed $B_c^+ \to J/\psi\,(\mu^+\mu^-)\rho^+$ decays,
- combinatorial background.

The PDFs for the B_c^+-candidate mass fits for each component are modelled using the same empirical analytical functions used for the normalisation mode fit, but with different parameter values. The parameters for the three B_c^+-decay components are determined from fits to simulation, with the samples weighted to mirror the relative proportions for the various years and different polarities in data, and including only B_c^+ candidates that are associated with the generated B_c^+ decays.

The fit is performed for the two selection working points defined in Table 3.5. For the $B_c^+ \to \psi(2S)\pi$ working point, the fit is performed twice, using the B_c^+-candidate mass calculated both without and with the dimuon mass constrained to $m_{\psi(2S)}$.

Fits to simulation

Figures 4.15–4.17 show the B_c^+ mass distributions in full simulation for the $B_c^+ \to \psi(2S)(\mu^+\mu^-)\,\pi^+$, the $B_c^+ \to \psi(2S)(\mu^+\mu^-)\,K^+$ and the $B_c^+ \to \psi(2S)(\mu^+\mu^-)\,\rho^+$ components with PDF fit projections overlaid. Tables 4.16–4.18 show the PDF parameters obtained from the fits. The PDFs for these three components share a global shift of peak position and a width scaling factor. The values for the shifts of peak position and width scaling factors are fixed respectively to 0 and 1 in the fits to simulation. The values for the fit to data will be constrained using the results of the fits for the normalisation mode (see Sec. 4.3.2).

The possible remaining combinatorial background is modelled using an exponential function. The B_c^+-candidate mass PDF exponent is free in the fits to data. In total, the fit model includes five free fit parameters, the yields for each component and the exponent of the combinatorial background, and two parameters with Gaussian constraints (the shift and scale factors).

Table 4.16: Signal PDF parameters (defined as in Table 4.9) determined by fitting the $B_c^+ \to \psi(2S)\pi^+$ simulation.

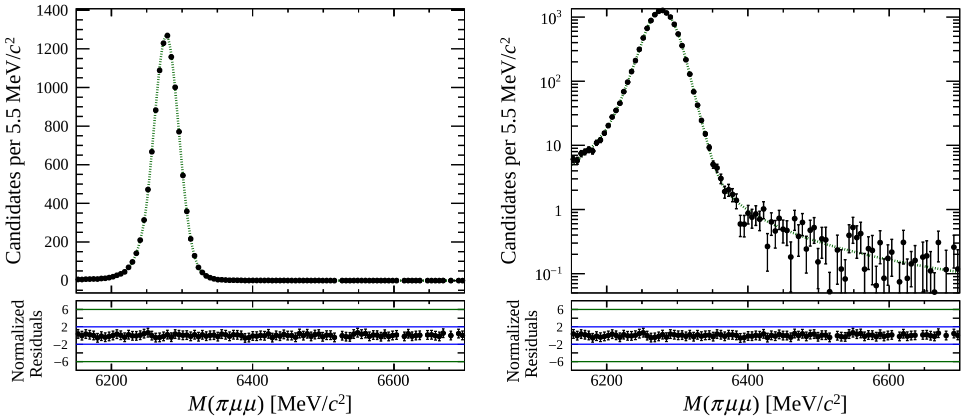
Fit parameters	$\pi^+\mu^+\mu^-$ working point	$\psi(2S)\pi^+$ working point	
	Unconstr. $\mu^+\mu^-$ mass	Unconstr. $\mu^+\mu^-$ mass	Constr. $\mu^+\mu^-$ mass
α_1	-1.58 ± 0.17	-1.43 ± 0.16	-1.73 ± 0.24
α_2	2.64 ± 0.25	2.54 ± 0.21	1.90 ± 0.24
f_1	0.66 ± 0.11	0.55 ± 0.09	0.51 ± 0.12
$\mu\,[\text{MeV}/c^2]$	6277.8 ± 0.3	6277.8 ± 0.2	6276.7 ± 0.1
n_1	2.52 ± 0.61	2.59 ± 0.48	1.35 ± 0.20
n_2	1.50 ± 0.38	1.52 ± 0.29	2.13 ± 0.23
$\sigma\,[\text{MeV}/c^2]$	22.8 ± 1.66	21.7 ± 0.99	11.4 ± 0.73
R	0.73 ± 0.05	0.73 ± 0.03	0.75 ± 0.03

Table 4.17: Misidentified background PDF parameters (defined as in Table 4.9) determined by fitting the $B_c^+ \to \psi(2S)K^+$ simulation. When fitting to the constrained B_c^+-candidate mass for the $\psi(2S)\pi^+$ working point, the parameter n_1 is fixed to 10 for this decay mode to be treated as a very large factor in the DCB PDF.

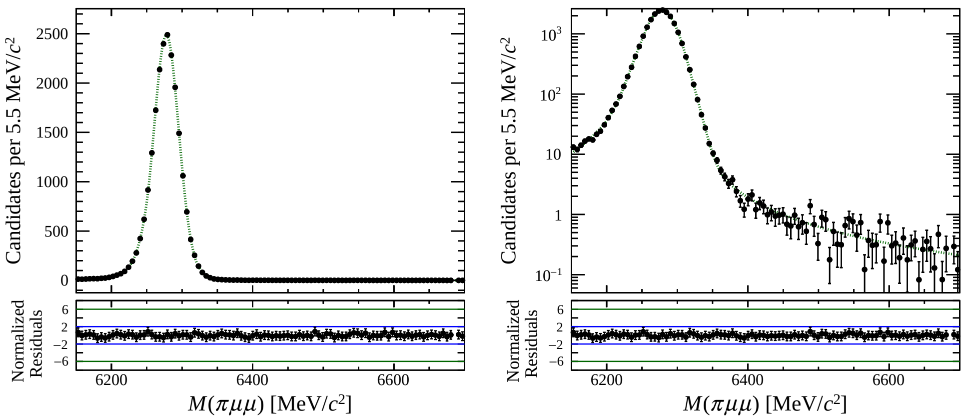
Fit parameters	$\pi^+\mu^+\mu^-$ working point	$\psi(2S)\pi^+$ working point	
	Unconstr. $\mu^+\mu^-$ mass	Unconstr. $\mu^+\mu^-$ mass	Constr. $\mu^+\mu^-$ mass
α_1	-0.56 ± 0.64	-0.51 ± 0.21	-0.68 ± 0.32
α_2	2.37 ± 0.78	2.41 ± 0.45	2.24 ± 0.32
f_1	0.46 ± 0.29	0.47 ± 0.17	0.54 ± 0.16
$\mu\,[\text{MeV}/c^2]$	6240.5 ± 1.9	6239.9 ± 0.9	6239.5 ± 0.6
n_1	7.33 ± 45.9	10 ± 0	2.51 ± 1.34
n_2	1.65 ± 1.24	1.60 ± 0.74	1.71 ± 0.44
$\sigma\,[\text{MeV}/c^2]$	22.4 ± 2.86	22.7 ± 1.22	14.4 ± 1.00
R	0.79 ± 0.23	0.80 ± 0.15	0.80 ± 0.12

Table 4.18: Partially reconstructed background PDF parameters (defined as in Table 4.14) determined by fitting the $B_c^+ \to \psi(2S)\rho^+$ simulation.

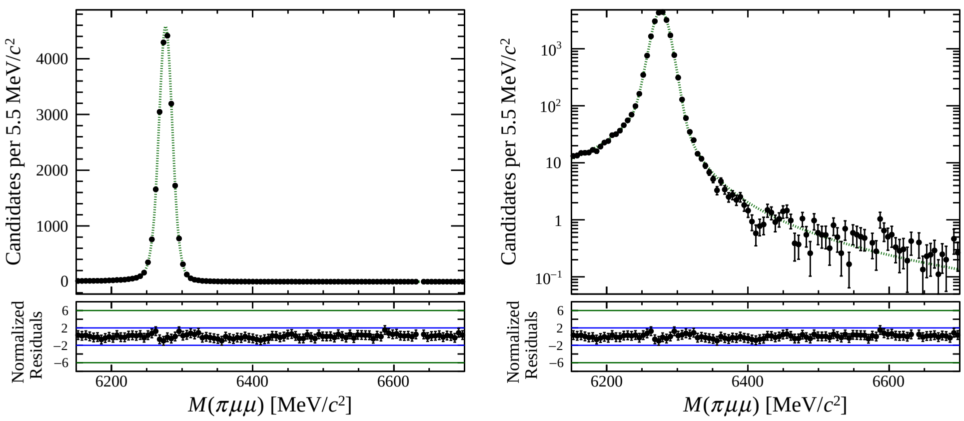
Fit parameters	$\pi^+\mu^+\mu^-$ working point	$\psi(2S)\pi^+$ working point	
	Unconstr. $\mu^+\mu^-$ mass	Unconstr. $\mu^+\mu^-$ mass	Constr. $\mu^+\mu^-$ mass
f_1	0.13 ± 0.10	0.12 ± 0.06	3.10 ± 0.14
$\mu\,[\text{MeV}/c^2]$	6090.4 ± 26.3	6088.7 ± 18.2	6028.2 ± 53.6
$\sigma\,[\text{MeV}/c^2]$	31.3 ± 5.14	30.5 ± 3.5	29.8 ± 5.73
R	6.33 ± 3.35	6.17 ± 2.31	6.53 ± 2.11



(a) $\pi^+\mu^+\mu^-$ working point, unconstrained dimuon mass

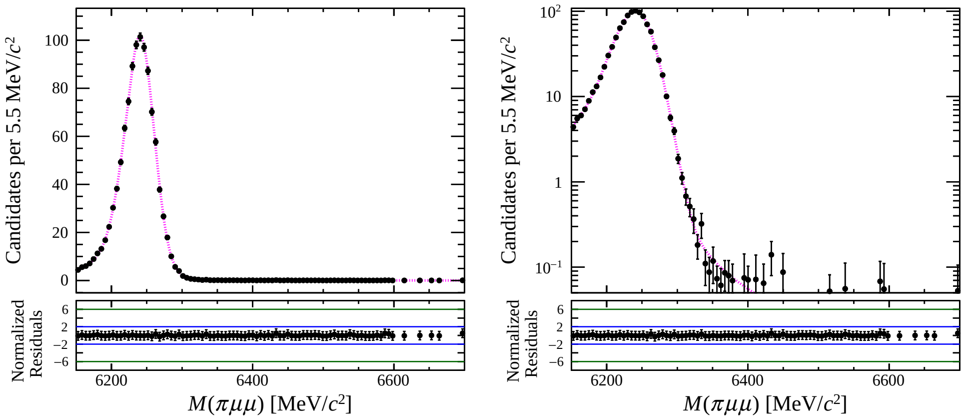


(b) $\psi(2S)\pi^+$ working point, unconstrained dimuon mass

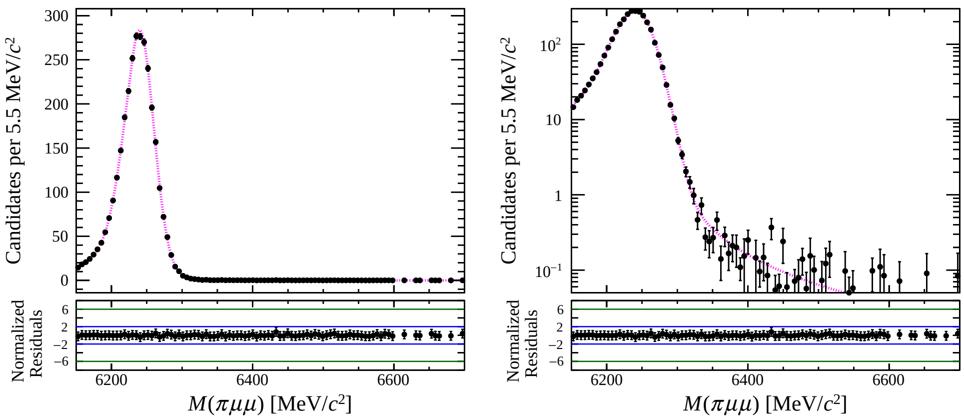


(c) $\psi(2S)\pi^+$ working point, constrained dimuon mass

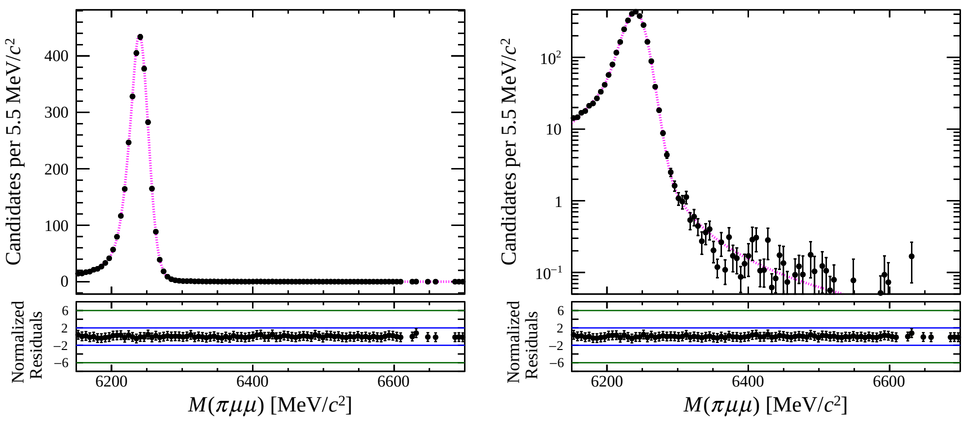
Figure 4.15: Distributions of B_c^+ mass for $B_c^+ \to \psi(2S)\pi$ candidates reconstructed in simulation with fit projections overlaid. The distributions are shown in (left) linear and (right) log scales.



(a) $\pi^+\mu^+\mu^-$ working point, unconstrained dimuon mass

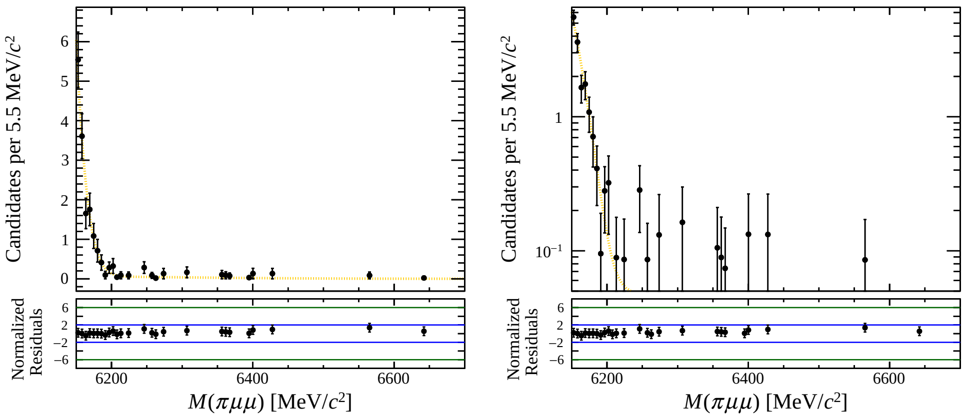


(b) $\psi(2S)\pi^+$ working point, unconstrained dimuon mass

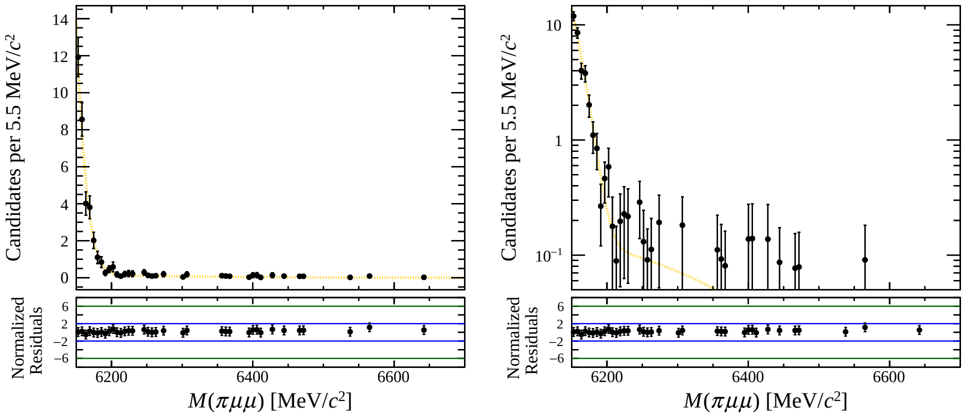


(c) $\psi(2S)\pi^+$ working point, constrained dimuon mass

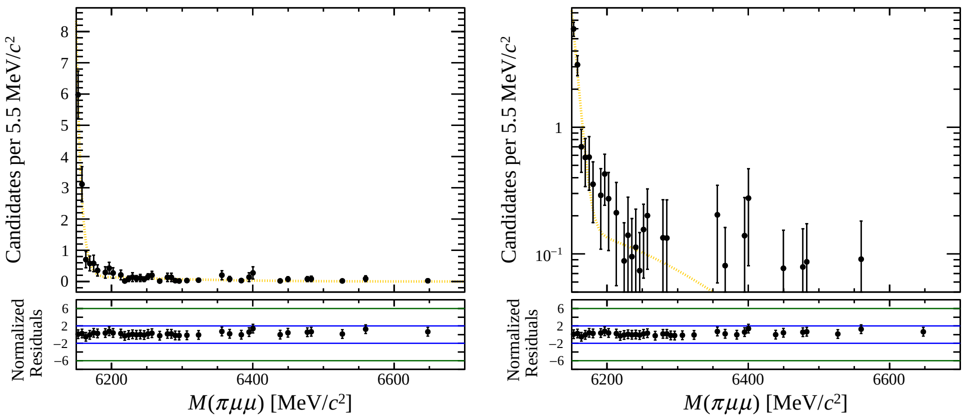
Figure 4.16: Distributions of B_c^+ mass for $B_c^+ \to \psi(2S)K^+$ candidates reconstructed in simulation with fit projections overlaid. The distributions are shown in (left) linear and (right) log scales.



(a) $\pi^+\mu^+\mu^-$ working point, unconstrained dimuon mass



(b) $\psi(2S)\pi^+$ working point, unconstrained dimuon mass



(c) $\psi(2S)\pi^+$ working point, constrained dimuon mass

Figure 4.17: Distributions of B_c^+ mass for $B_c^+ \to \psi(2S)\rho^+$ candidates reconstructed in simulation with fit projections overlaid. The distributions are shown in (left) linear and (right) log scales.

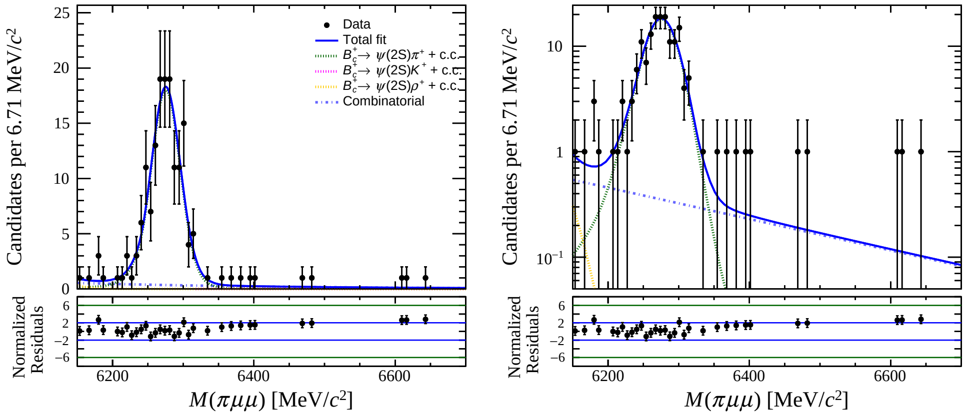
Fits to Data

We fit the model for the $B_c^+ \to \psi(2S)\pi$ mode to the candidates selected in the full data sample. Table 4.19 shows the fit results. Figure 4.18 shows the full data sample with fit projections overlaid.

The agreement between the data and the fit model is excellent. We observe that the yields for the $B_c^+ \to \psi(2S)(\mu^+\mu^-)\,K^+$ and the $B_c^+ \to \psi(2S)(\mu^+\mu^-)\,\rho^+$ background components are consistent with zero in all cases, indicating that the contributions from these background sources are negligible, as expected.

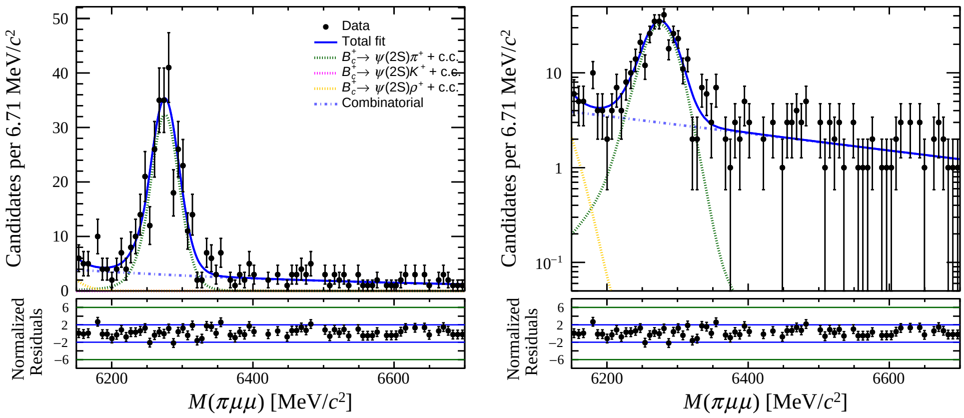
Table 4.19: Results of the fit to the B_c^+-candidate mass distribution of $B_c^+ \to \psi(2S)\pi^+$ candidates in data. The combinatorial background exponent ("Comb. bkg. exp.") is quoted in units of $(\,\mathrm{MeV}/c^2)^{-1}$.

Fit parameters	$\pi^+\mu^+\mu^-$ working point	$\psi(2S)\pi^+$ working point	
	Unconstr. $\mu^+\mu^-$ mass	Unconstr. $\mu^+\mu^-$ mass	Constr. $\mu^+\mu^-$ mass
$\psi(2S)\pi^+$ yield	149 ± 14	263 ± 20	256 ± 18
Bkg. yield	21 ± 10	189 ± 24	197 ± 19
$\psi(2S)K^+$ yield	-5 ± 9	0 ± 16	13 ± 10
$\psi(2S)\rho^+$ yield	0 ± 3	5 ± 7	-4 ± 5
Comb. bkg. exp.	-0.004 ± 0.002	-0.002 ± 0.0007	-0.002 ± 0.001



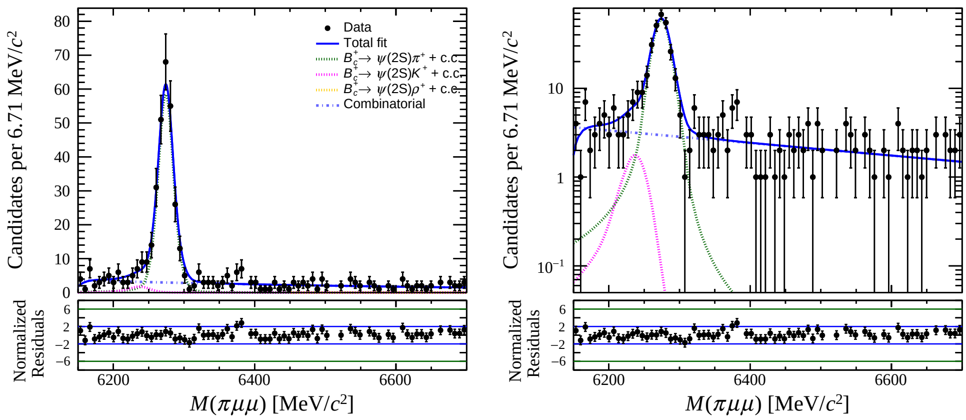
(a)

$\pi^+\mu^+\mu^-$ working point, unconstrained dimuon mass



(b)

$\psi(2S)\pi^+$ working point, unconstrained dimuon mass



(c)

$\psi(2S)\pi^+$ working point, constrained dimuon mass

Figure 4.18: Distributions of B_c^+-candidate mass for $B_c^+ \to \psi(2S)\pi$ candidates reconstructed in collision data with fit projections overlaid. The distributions are shown in (left) linear and (right) log scales.

Fit Estimator Properties

We perform pseudo-experiments to assess the fit estimator properties, in a similar way as in Sec. 4.3.2. The pseudo-data samples are generated with the expected four components: $B_c^+ \to \psi(2S)\pi^+$ decays, misidentified $B_c^+ \to \psi(2S)K^+$ decays, partially reconstructed $B_c^+ \to \psi(2S)\rho^+$ decays and combinatorial background.

The pseudo-data samples are generated from the PDF shapes obtained from simulation in Sec. 4.3.3 for the B_c^+-decay components. The expectations for the shape of the combinatorial background and the number of candidates for each component are obtained from the fit to data in Sec. 4.3.3 for the $\psi(2S)\pi^+$ working point. The expectations for the global shift of peak position and width scaling factor are obtained from the $B_c^+ \to J/\psi\pi^+$ fit in Sec. 4.3.2. We use Poisson distributions to model the number of events for each component in each pseudo-data sample.

For each pseudo-data sample, we perform a fit of the full model as in Sec. 4.3.3. We perform 1000 pseudo-experiments and calculate the pull distributions for each fit parameter. The fit converges for all pseudo-experiments. We fit a Gaussian function to the pull distributions to obtain the mean and the width of the distributions.

Figure 4.19 shows the pull distribution for each fit parameter. We generally observe unbiased distributions and proper uncertainty estimation for all fit parameters.

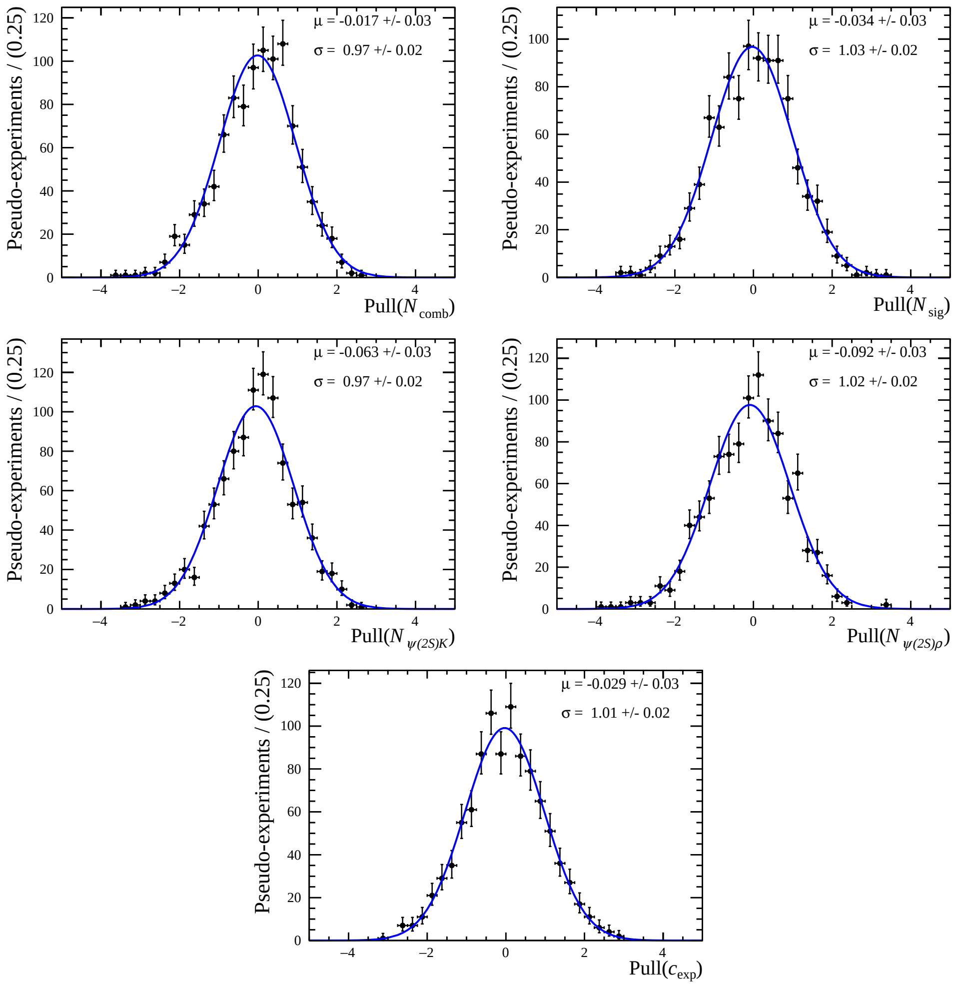


Figure 4.19: Pull distributions obtained from fits to pseudo-data samples with $B_c^+ \to \psi(2S)\pi^+$ candidates. The pull distributions are shown for all fit parameters: N_i represents the yield of the component in the respective subscript and c_{exp} the exponent of the combinatorial background PDF.

Key takeaways from this chapter

The non-resonant mode is blinded in this analysis, as it is a previously unobserved decay channel; we do not present fits to data yet. However, as the resonant modes have been previously observed thus we perform the analysis unblinded, the end of this section provides us with all the information needed for the central value and statistical uncertainty for the update of the ratio of branching fractions $\mathcal{B}(B_c^+ \to \psi(2S)\pi^+)/\mathcal{B}(B_c^+ \to J/\psi\,\pi^+)$.

Chapter 5

Systematic Uncertainties

Systematic uncertainties on the ratios of branching fractions arise from potential imperfections in the fit configuration, and in the efficiency determination. As for the efficiencies, we consider systematic effects associated with

- Monte Carlo simulation statistics
- data-simulation efficiency corrections,
- truth-matching criteria,
- variation of acceptance and selection efficiency with q^2,
- uncertainty of B_c^+ lifetime.

Related to the fit, we consider systematic effects due to

- PDF modelling,
- polarisation in the partially reconstructed background.

The systematic effect associated with the B_c^+-mass peak shift and width scaling factors is taken into account by introducing Gaussian constraints into the signal-mode branching-fraction fits. A final source of systematic uncertainty is due to external inputs in the calculation of the ratio of branching fractions summarized in Table 5.1. Below we describe how the uncertainties are estimated.

5.1 Uncertainties from Efficiency

The systematic uncertainties associated with the efficiencies are obtained with a method common to all $B_{(c)}^+ \to \pi^+\mu^+\mu^-$ analyses. We estimate the systematic

Table 5.1: External branching fractions obtained from the PDG [3].

Parameter	Value
$\mathcal{B}(J/\psi \to \mu^+\mu^-)$ $[10^{-3}]$	59.61 ± 0.33
$\mathcal{B}(\psi(2S) \to \mu^+\mu^-)$ $[10^{-3}]$	8.0 ± 0.6
$\mathcal{B}(J/\psi \to e^+e^-)$ $[10^{-3}]$	59.71 ± 0.32
$\mathcal{B}(\psi(2S) \to e^+e^-)$ $[10^{-3}]$	7.93 ± 0.17

uncertainty on the efficiency ratios between signal and normalisation modes for each source and in total which is done by adding the individual uncertainties in quadrature. Tables 5.2 and 5.3 summarizes the systematic uncertainties for each source.

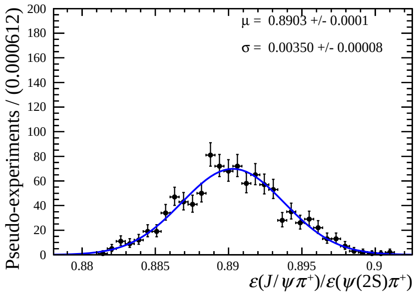
Table 5.2: Systematic uncertainties on the efficiency ratio $r_{J/\psi\pi^+/\pi^+\mu^+\mu^-} = \hat{\varepsilon}_{J/\psi\pi^+}/\hat{\varepsilon}_{\pi^+\mu^+\mu^-}$ between the non-resonant $\pi^+\mu^+\mu^-$ decay and the resonant $J/\psi\pi^+$ decay mode. The uncertainties are given for each q^2 interval and for all combined.

	$B_c^+ \to \pi^+\mu^+\mu^-$ q^2 range				
Source	Very low	Low	Central	High	Combined
Simulation sample size	0.0231	0.0092	0.0142	0.0053	0.0047
Truth-matching efficiency	0.0000	0.0000	0.0000	0.0000	0.0000
Multivariate kinematic correction	0.0347	0.0219	0.0010	0.0242	0.0091
p_{T}-correction binning	0.0004	0.0001	0.0000	0.0002	0.0001
p_{T}-correction statistical uncertainty	0.0008	0.0004	0.0009	0.0002	0.0001
Trigger correction binning	0.0043	0.0021	0.0003	0.0003	0.0006
Trigger correction statistical uncertainty	0.0029	0.0016	0.0006	0.0011	0.0010
PID correction statistical uncertainty	0.0001	0.0000	0.0000	0.0002	0.0001
Muon ID correction statistical uncertainty	0.0018	0.0010	0.0005	0.0018	0.0011
Tracking efficiency uncertainty	0.0018	0.0011	0.0002	0.0005	0.0007
B_c^+ lifetime	0.0006	0.0013	0.0009	0.0012	0.0011
HLT1 TCK variation	0.0030	0.0014	0.0014	0.0015	0.0016
Phase space spread	0.0701	0.0989	0.0302	0.2326	0.2710
Polarisation assumption	0.2312	0.1440	0.0271	0.0336	0.0032
Total	0.2442	0.1761	0.0406	0.2363	0.2712

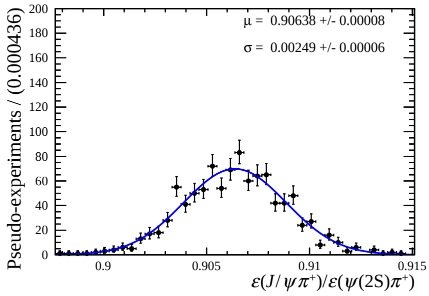
Table 5.3: Systematic uncertainties on the efficiency ratio $r_{J/\psi\pi^+/\psi(2S)\pi^+} = \hat{\varepsilon}_{J/\psi\pi^+}/\hat{\varepsilon}_{\psi(2S)\pi^+}$ between the resonant modes. The efficiencies are given for the two working points (WP) defined in Table 3.5.

Source	$\pi^+\mu^+\mu^-$ WP	$\psi(2S)\pi^+$ WP
Simulation sample size	0.0035	0.0025
Truth-matching efficiency	0.0000	0.0000
Multivariate kinematic correction	0.0112	0.0103
p_{T}-correction binning	0.0000	0.0000
p_{T}-correction statistical uncertainty	0.0001	0.0001
Trigger correction binning	0.0001	0.0001
Trigger correction statistical uncertainty	0.0008	0.0010
PID correction statistical uncertainty	0.0000	0.0000
Muon ID correction statistical uncertainty	0.0006	0.0006
Tracking efficiency uncertainty	0.0001	0.0001
B_c^+ lifetime	0.0011	0.0005
HLT1 TCK variation	0.0018	0.0057
Total	0.0114	0.0118

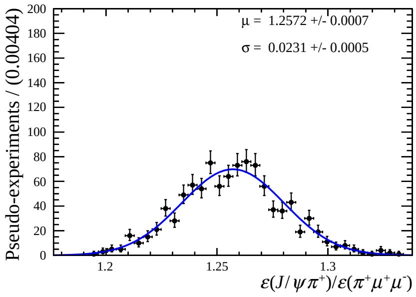
The uncertainty due to the finite size of the simulated sample is estimated by generating 1000 pseudo-data samples from the full simulation samples by bootstrapping [34]. For this, we follow the Poisson-bootstrap technique [35]: each simulation event in the parent samples gets assigned a weight drawn from a Poisson distribution with the expectation equal to unity. For each pseudo-data sample, we estimate the efficiencies for signal and normalisation modes as described in 4.1, and calculate the efficiency ratio. The width of the efficiency ratio distribution is taken as uncertainty. Figure 5.1 shows the efficiency ratio distributions, with the projection of a fit of a Gaussian function overlaid, for the two selection working points (see 3.5), in case of the ratio between the resonant modes, and for the four q^2 bins, in case of the ratio between the non-resonant mode and the $B_c^+ \to J/\psi\pi^+$ mode. The efficiency ratio distributions for all other variations described hereafter, are shown in App. G.



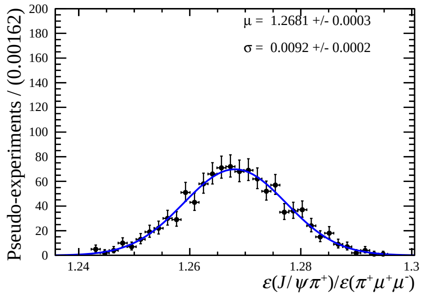
a) $\pi^+\mu^+\mu^-\pi^+$ WP



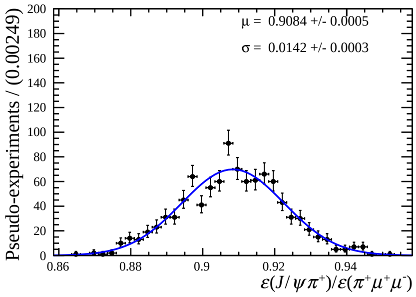
b) $\psi(2S)\pi^+$ WP



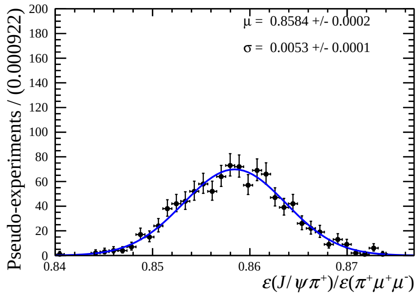
c) $\pi^+\mu^+\mu^-$ Very Low



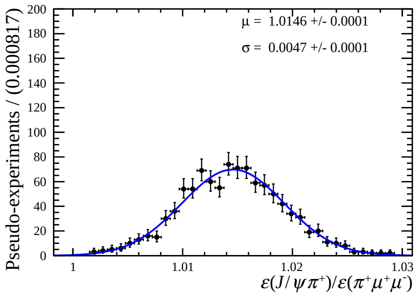
d) $\pi^+\mu^+\mu^-$ Low



e) $\pi^+\mu^+\mu^-$ Central



f) $\pi^+\mu^+\mu^-$ High



g) $\pi^+\mu^+\mu^-$ Bins Combined

Figure 5.1: Distributions of efficiency ratio obtained from pseudo-data samples with default configurations to study the effect of MC size.

The truth-matching efficiency uncertainty is estimated by including the background category `BKGCAT`$=60$ as signal, which corresponds to candidates for which one or more particles is not correctly truth matched and is classified as a `ghost`. However, the fraction of `BKGCAT`$=60$ candidates for which more than one particle is classified as a ghost is negligible. The fraction of candidates associated with `BKGCAT`$=60$ is about 2% for the signal and normalisation modes. We generate 1000 pseudo-data samples and estimate the mean efficiency ratio from them as described above. The difference between the mean ratio estimated with and without `BKGCAT`$=60$ is taken as uncertainty.

The corrections applied to the simulation are described in Sec. B. They comprise corrections for PID efficiencies, muon ID efficiencies, track reconstruction efficiency, trigger efficiency, kinematic and generator-level p_{T} distributions. We perform two kinds of variations of them. First, a systematic uncertainty is estimated due to some of the choices made when determining the efficiency model: the choice of the binning of the trigger and p_{T} correction histograms and the choice of the inputs of the BDT correcting MVA inputs. We calculate new efficiency models by making variations of these choices. For the trigger and p_{T} correction histograms, we vary the binning by increasing the number of bins by two in each axis. For the BDT correcting MVA inputs, we train an alternative BDT that receives as input the χ^2_{IP} of the two muons together with the default inputs (p_{T}, n_{tracks} and B_c^+ $\chi^2_{\mathrm{vtx}}/n_{\mathrm{dof}}$). For each new efficiency model, we generate 1000 pseudo-data samples and estimate the mean efficiency ratio from them as described above. The difference between the mean ratio estimated with the new alternative and with the nominal efficiency model is taken as uncertainty. Second, a systematic uncertainty is estimated due to the statistical uncertainty on the correction histograms. For this, each histogram is varied within its uncertainty 100 times by varying the content of each bin following a Gaussian distribution. The Gaussian mean and width correspond to the central value and width of the respective bin in the nominal histogram. The varied histograms are used to estimate a new efficiency model. We calculate the efficiency ratio for each of the 100 efficiency models and take the width of the efficiency ratio distribution as systematic uncertainty.

The uncertainty associated with the B_c^+ lifetime is determined by weighting all B_c^+ candidates such that the simulated decay-time distribution follows an exponential decay with a lifetime corresponding to the PDG lifetime. The PDG lifetime is varied by $\pm 1\sigma$ and the difference between the mean efficiency ratio for the two cases is taken as systematic uncertainty.

We assign an additional systematic uncertainty associated with the misalignment in HLT1 trigger requirements between data and simulation for 2012 and 2016. We apply the tightest set of HLT1 requirements for 2012 and 2016 to all B_c^+ candidates and assign the difference between the nominal efficiency ratio and the mean efficiency ratio calculated with these requirements as systematic uncertainty [1].

As for the simulation model assumption, the non-resonant signal simulation samples are generated with a flat phase-space distribution (PHSP). However, the true q^2 distribution of $B_c^+ \to \pi^+\mu^+\mu^-$ decays is unknown, and there is no clear theoretical guidance for its expectation. We therefore assign a systematic uncertainty associated with the variation of efficiency with q^2 (phase space spread), which is shown in Fig. 5.2. We calculate the efficiency as a function of q^2 in $0.5\,\mathrm{GeV}^2$ intervals and use the efficiency values to estimate a mean and a standard deviation for each search q^2 bin and for all bins combined. The standard deviation is used as systematic uncertainty associated with the phase-space spread. The approach of determining the differential branching fraction within q^2 bins helps to reduce this uncertainty.

A similar uncertainty arises due to the different possible polarisations of the dimuon system, which is illustrated by the different coloured distributions in Fig. 5.2. The two extreme possibilities are: a fully longitudinally polarised dimuon system that is distributed as $\frac{1}{2}(1-\cos^2\theta_l)$, where θ_l is the helicity angle of the μ^+; and an unpolarised dimuon system that is distributed uniformly in $\cos\theta$. We estimate the mean efficiency for each q^2 bin and for all bins combined in the two extreme cases and assign the difference between the two cases as systematic uncertainty associated with the polarisation.

The dependence of the efficiency on q^2 arises from several different sources: from intrinsic momentum requirements of $p \gtrsim 3\,\mathrm{GeV}/c$ to reconstruct the two muons; from p_T requirements on one or both muons in the hardware trigger; and from selection requirements. The increase in efficiency from low to moderate q^2 is driven by the trigger requirements. The sudden decrease in efficiency at high q^2 is in part due to the impact parameter requirement on the pion in the stripping line. At the q^2 end-point, the pion will be co-linear with the direction of the B_c^+ in the laboratory frame and the resulting efficiency is small. Due to the short B_c^+ lifetime the effect is more marked here than in *e.g.* $B^+ \to \pi^+\mu^+\mu^-$ decays. The muon momentum and trigger requirements also sculpt the angular distribution. At low q^2, decays are preferentially selected with θ_l away from 0 and π. At larger q^2, the opposite is true. This results in a lower efficiency for the unpolarised case than pure longitudinal

[1]This work involving HLT1 was completed by Fernando towards the end of my time working on this analysis.

polarisation in Fig. 5.2 at low q^2.

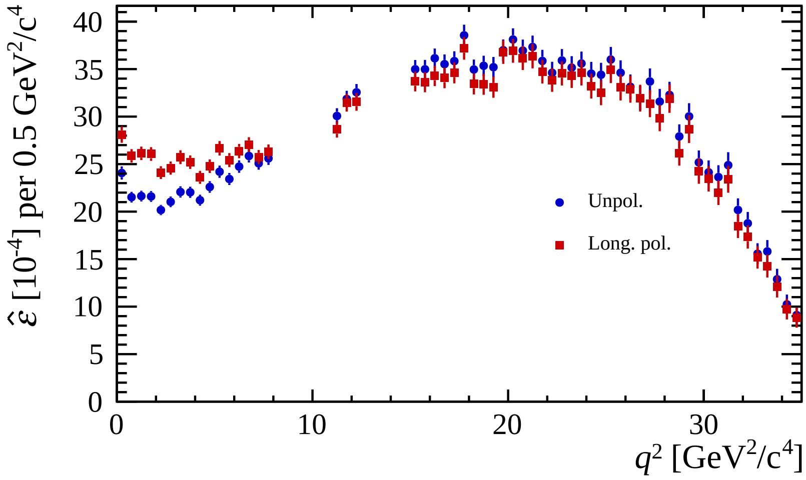
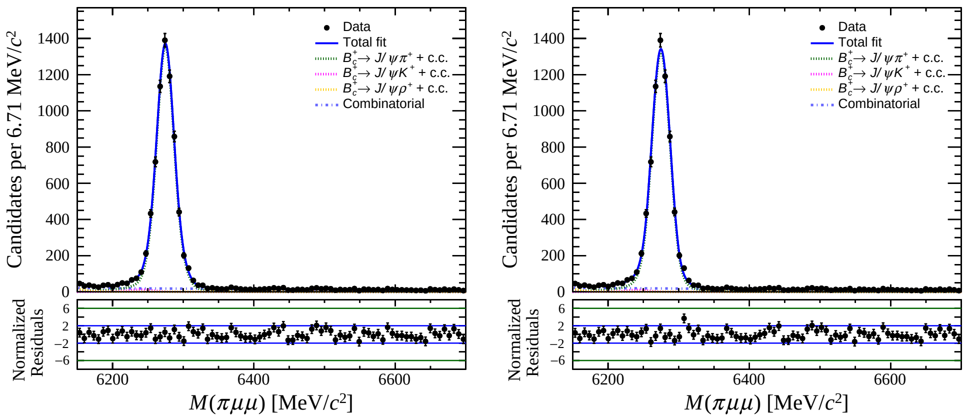


Figure 5.2: Efficiency as a function of q^2 for $B_c^+ \to \pi^+\mu^+\mu^-$ decays, evaluated with both longitudinally polarised and unpolarised models for the dimuon system.

5.2 Uncertainties from Modelling

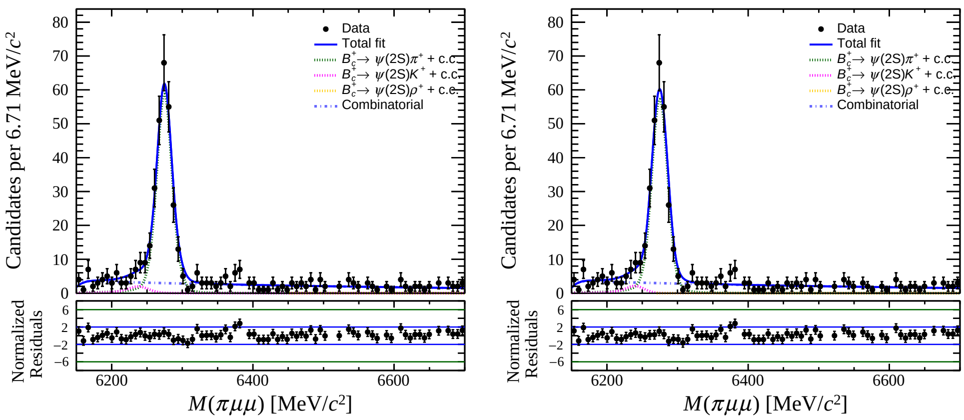
We perform fits using alternate models using the full data sample. Table 5.4 shows an overview of the used analytical functions for the default and the alternate models for the resonant $B_c^+ \to J/\psi\pi^+$ and $B_c^+ \to \psi(2S)\pi^+$ modes. The fits to simulation and data are performed in a similar way as described in Secs. 4.3.2 and 4.3.3. We use consistently the same model shapes for both modes, but with different internal parameters. The alternate models include a global shift of peak position and a global width scaling factor, such that the alternate fits have the same degrees of freedom as the default ones. The global shift of peak position and width scaling factor in the alternate fit for the $B_c^+ \to \psi(2S)\pi^+$ mode is constrained to the values obtained in the fit for the $B_c^+ \to J/\psi\pi^+$ mode in the same way as in the default fit (see Sec. 4.3.3).

Figure 5.3 shows the distributions of B_c^+ candidate mass for the two resonant modes with fit projections overlaid for default and alternate fit models. Tables 5.5 and 5.6 show the fit results for the $B_c^+ \to J/\psi\pi^+$ and $B_c^+ \to \psi(2S)\pi^+$ modes for default and alternate models and for $\psi(2S)\pi^+$ and $\pi^+\mu^+\mu^-$ working points. The resulting values of the ratio $F_{\psi(2S)/J/\psi}$ (Eq. (6.1)) and the resulting systematic uncertainty associated with the choice of the model PDFs are given in Table 5.7.



(a)

$B_c^+ \to J/\psi\,\pi^+$ candidates



(b)

$B_c^+ \to \psi(2S)\pi^+$ candidates

Figure 5.3: Distributions of B_c^+-candidate mass for candidates reconstructed at the $\psi(2S)\pi^+$ WP in collision data with fit projections overlaid. The distributions are shown with the (left) default and (right) alternate models as described above.

Table 5.4: Summary of empirical analytical functions used as PDFs in the fits for the resonant $B_c^+ \to J/\psi\,\pi^+$ and $B_c^+ \to \psi(2S)\pi^+$ modes.

Component	Default	Alternate
$B_c^+ \to J/\psi\,(\to \mu^+\mu^-)\,\pi^+$	Gaussian plus double-sided Crystal Ball	Two single-sided Crystal Balls
$B_c^+ \to J/\psi\,(\to \mu^+\mu^-)\,K^+$	Gaussian plus double-sided Crystal Ball	Two single-sided Crystal Balls
$B_c^+ \to J/\psi\,(\to \mu^+\mu^-)\,\rho^+$	Double Gaussian	Single-sided Crystal Ball
Combinatorial background	Exponential	Exponential

Table 5.5: Results of the fits to the B_c^+-candidate mass distribution of $B_c^+ \to J/\psi\,\pi^+$ candidates in data for default and alternate models and for $\psi(2S)\pi^+$ and $\pi^+\mu^+\mu^-$ working points.

	$\psi(2S)\pi^+$ WP		$\pi^+\mu^+\mu^-$ WP	
Fit component	Default	Alternate	Default	Alternate
Comb. exp. $[(10^2\,\mathrm{MeV}/c^2)^{-1}]$	-0.14 ± 0.03	-0.14 ± 0.03	-0.14 ± 0.14	-0.14 ± 0.14
Comb. yield	1254 ± 60	1277 ± 57	101 ± 25	101 ± 24
$J/\psi\,K^+$ yield	90 ± 43	70 ± 44	-81 ± 58	-122 ± 85
$J/\psi\,\pi^+$ yield	6887 ± 92	6896 ± 93	3508 ± 82	3552 ± 104
$J/\psi\,\rho^+$ yield	56 ± 22	45 ± 18	41 ± 11	37 ± 9
Global shift $[\,\mathrm{MeV}/c^2]$	-1.93 ± 0.20	-1.97 ± 0.20	-3.17 ± 0.59	-3.47 ± 0.77
Width scale	1.14 ± 0.02	1.14 ± 0.02	1.17 ± 0.03	1.18 ± 0.03

Table 5.6: Results of the fits to the B_c^+-candidate mass distribution of $B_c^+ \to \psi(2S)\pi^+$ candidates in data for default and alternate models and for $\psi(2S)\pi^+$ and $\pi^+\mu^+\mu^-$ working points.

	$\psi(2S)\pi^+$ WP		$\pi^+\mu^+\mu^-$ WP	
Fit component	Default	Alternate	Default	Alternate
Comb. exp. $[(10^2\,\mathrm{MeV}/c^2)^{-1}]$	-0.16 ± 0.06	-0.16 ± 0.06	-0.36 ± 0.27	-0.33 ± 0.27
Comb. yield	197 ± 19	196 ± 18	21 ± 10	20 ± 9
$\psi(2S)K^+$ yield	13 ± 10	12 ± 9	-5 ± 9	-4 ± 9
$\psi(2S)\pi^+$ yield	256 ± 18	256 ± 18	149 ± 14	150 ± 14
$\psi(2S)\rho^+$ yield	-4 ± 5	-3 ± 4	1 ± 3	1 ± 3

Table 5.7: Values of $F_{\psi(2S)/J/\psi}$ obtained with the default and with the alternative fit model as well as the difference between the two.

$F_{\psi(2S)/J/\psi}$	$\psi(2S)\pi^+$ WP	$\pi^+\mu^+\mu^-$ WP
Default	0.03369	0.03782
Alternate	0.03365	0.03760
$\delta F_{\psi(2S)/J/\psi}$	±0.00004	±0.00022

5.2.1 Polarisation of Partially Reconstructed Backgrounds

In the fits for the resonant $B_c^+ \to J/\psi\,\pi^+$ and $B_c^+ \to \psi(2S)\pi^+$ modes, the model parameters for the partially reconstructed backgrounds are obtained respectively from $B_c^+ \to J/\psi\,\rho^+$ and $B_c^+ \to \psi(2S)\rho^+$ simulation in which the final sate is unpolarised, that is in which the absolute values of the three helicity amplitudes are equal and the relative phases between the amplitudes are set to zero.

We study the effect of a possible polarization of the partially reconstructed $B_c^+ \to J/\psi\,\rho^+$ and $B_c^+ \to \psi(2S)\rho^+$ decays on the modeled $m(\pi^+\mu^+\mu^-)$ mass distributions. We consider the extreme cases of a fully longitudinally polarised dipion system, which follows a $\frac{3}{2}\cos^2\theta_\pi$ distribution, where θ_π is the helicity angle of the π^+ in the ρ frame, and a fully transversely polarised di-pion system, which follows a $\frac{3}{4}(1-\cos^2\theta_\pi)$ distribution. The resulting $m(\pi^+\mu^+\mu^-)$ mass distributions for longitudinally and transversely polarised decays as well as for unpolarised $B_c^+ \to J/\psi\,\pi^+$ and $B_c^+ \to \psi(2S)\pi^+$ decays are shown in Fig. 5.4

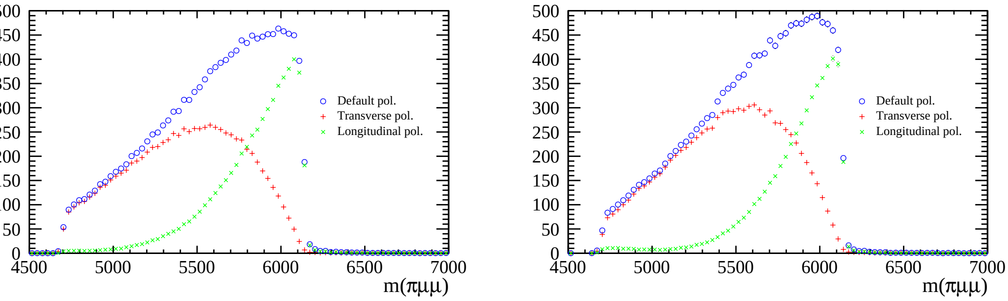


Figure 5.4: Distributions of B_c^+-candidate mass for B_c^+ candidates reconstructed in (left) $B_c^+ \to J/\psi\,\rho^+$ and (right) $B_c^+ \to \psi(2S)\rho^+$ simulation with different polarisation models.

We perform fits for the resonant $B_c^+ \to J/\psi\,\pi^+$ and $B_c^+ \to \psi(2S)\pi^+$ modes as described in Secs. 4.3.2 and 4.3.3, but modeling the shapes for the partially reconstructed backgrounds assuming a fully longitudinal or transverse polarisation consistently for the $B_c^+ \to J/\psi\,\rho^+$ and $B_c^+ \to \psi(2S)\rho^+$ decays. This study is performed only for the $\psi(2S)\pi$ working point as the non-resonant mode is not affected by partially reconstructed backgrounds (see App. F).

Figure 5.5 shows the B_c^+ mass distributions in full simulation for the $B_c^+ \to J/\psi\,\rho^+$ and the $B_c^+ \to \psi(2S)\rho^+$ components with PDF fit projections overlaid. Tables 5.8 and 5.9 show the fit results for the $B_c^+ \to J/\psi\,\pi^+$ and $B_c^+ \to \psi(2S)\pi^+$ modes for longitudinally and transversely polarised $B_c^+ \to J/\psi\,\rho^+$ and $B_c^+ \to \psi(2S)\rho^+$ decays. The resulting values of the ratio $F_{\psi(2S)/J/\psi}$ (Eq. (6.1)) and the resulting sys-

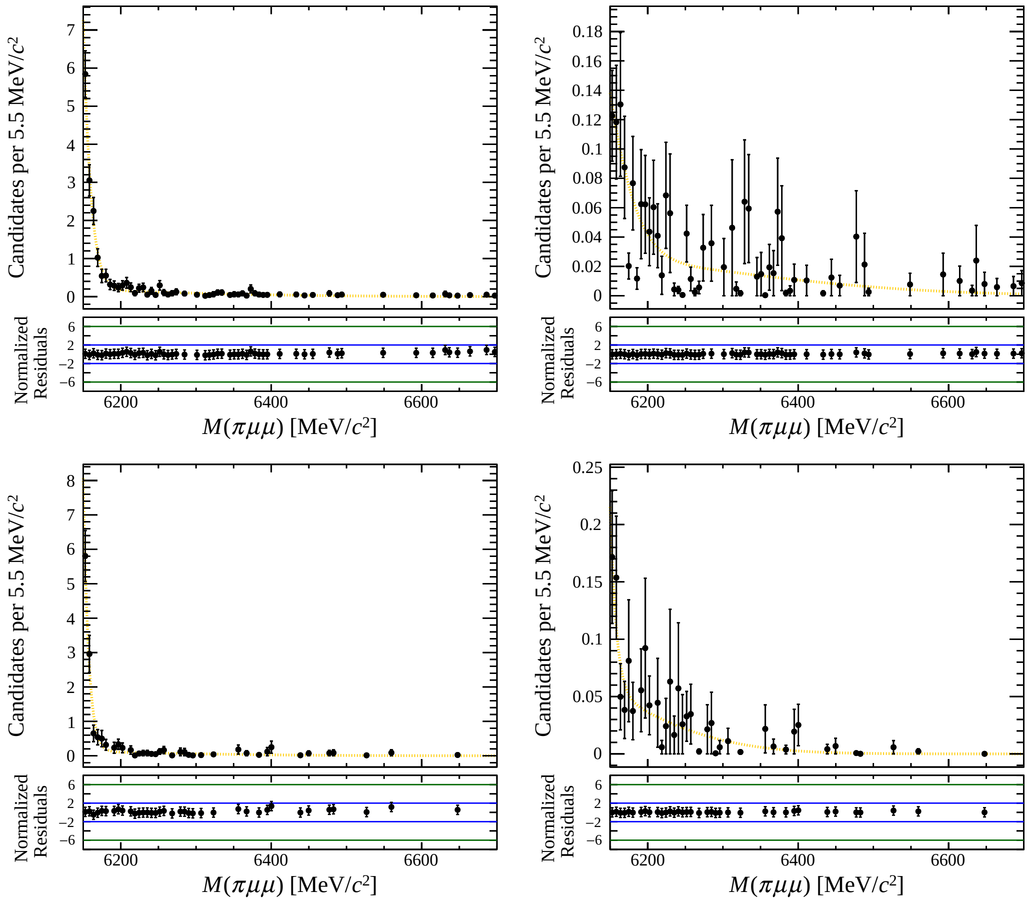


Figure 5.5: Distributions of B_c^+-candidate mass for $B_c^+ \to J/\psi\,\rho^+$ and $B_c^+ \to \psi(2S)\rho^+$ candidates reconstructed in simulation with (left) longitudinal and (right) transverse polarisation.

tematic uncertainty associated with the polarisation of the partially reconstructed backgrounds are given in Table 5.10.

Table 5.8: Results of the fits to the B_c^+-candidate mass distribution of $B_c^+ \to J/\psi\,\pi^+$ candidates in data for longitudinally and transversely polarised $B_c^+ \to J/\psi\,\rho^+$ decays.

Fit component	Longitudinal	Transverse
Comb. bkg. exp. $[(10^2\,\mathrm{MeV}/c^2)^{-1}]$	-0.13 ± 0.03	-0.10 ± 0.04
Comb. bkg. yield	1258 ± 59	1146 ± 99
$J/\psi\,K^+$ yield	90 ± 44	78 ± 43
$J/\psi\,\pi^+$ yield	6888 ± 92	6896 ± 92
$J/\psi\,\rho^+$ yield	52 ± 20	167 ± 75
Global shift $[\,\mathrm{MeV}/c^2]$	-1.93 ± 0.20	-1.94 ± 0.20
Width scale	1.14 ± 0.02	1.14 ± 0.02

Table 5.9: Results of the fits to the B_c^+-candidate mass distribution of $B_c^+ \to \psi(2S)\pi^+$ candidates in data for longitudinally and transversely polarised $B_c^+ \to \psi(2S)\rho^+$ decays.

Fit component	Longitudinal	Transverse
Comb. bkg. exp. $[(10^2\,\mathrm{MeV}/c^2)^{-1}]$	-0.16 ± 0.06	-0.18 ± 0.08
Comb. bkg. yield	197 ± 19	202 ± 26
$\psi(2S)K^+$ yield	13 ± 10	13 ± 10
$\psi(2S)\pi^+$ yield	256 ± 18	256 ± 18
$\psi(2S)\rho^+$ yield	-4 ± 4	-10 ± 16
Global shift $[\,\mathrm{MeV}/c^2]$	-1.98 ± 0.20	-1.98 ± 0.02
Width scale	1.14 ± 0.02	1.14 ± 0.02

Table 5.10: Values of $F_{\psi(2S)/J/\psi}$ obtained for longitudinally and transversely polarised $B_c^+ \to J/\psi\,\rho^+$ and $B_c^+ \to \psi(2S)\rho^+$ decays as well as the difference between the two.

$F_{\psi(2S)/J/\psi}$	$\psi(2S)\pi^+$ WP
Longitudinal	0.03369
Transverse	0.03365
$\delta F_{\psi(2S)/J/\psi}$	± 0.00004

Chapter 6

Results and Summary

6.1 Update of $\mathcal{B}(B_c^+ \to \psi(2S)\pi^+)/\mathcal{B}(B_c^+ \to J/\psi\,\pi^+)$

Since the study for the resonant modes is performed in an unblind way, we obtain results for the fraction

$$F_{\psi(2S)/J/\psi} = \frac{N_{\psi(2S)\pi^+}}{N_{J/\psi\,\pi^+}} \cdot \frac{\hat{\varepsilon}_{J/\psi\,\pi^+}}{\hat{\varepsilon}_{\psi(2S)\pi^+}}, \tag{6.1}$$

using the results in Tables 4.5, 4.15 and 4.19. We also obtain results for the ratio

$$R_{\psi(2S)/J/\psi} \equiv \frac{\mathcal{B}(B_c^+ \to \psi(2S)\pi^+)}{\mathcal{B}(B_c^+ \to J/\psi\,\pi^+)} = F_{\psi(2S)/J/\psi} \cdot \frac{\mathcal{B}(J/\psi \to e^+e^-)}{\mathcal{B}(\psi(2S) \to e^+e^-)}\,, \tag{6.2}$$

using the PDG values in Tab. 5.1. As in the previous analysis [16], we assume electroweak universality and use the ratio between the electron modes which is more precisely measured than the ratio between the muon modes. The results are shown in Table 6.1. The $\pi^+\mu^+\mu^-$ working point is a cross check that has served it's purpose hence we will only quote the more important $\psi(2S)\pi^+$ WP full result. The results for the $\psi(2S)\pi$ working point are obtained using the fit results with constrained B_c^+-candidate mass. The results for the two selection working points agree within the statistical precision. Both values are consistent with the value of $F_{\psi(2S)/J/\psi}$ and $R_{\psi(2S)/J/\psi}$ measured in the previous analysis [16], which are $0.0354 \pm 0.0042\,(\text{stat}) \pm 0.0010\,(\text{syst})$, and $0.268 \pm 0.032\,(\text{stat}) \pm 0.007\,(\text{syst}) \pm 0.006\,(\text{BF})$, respectively.

Table 6.1: Results for fraction $F_{\psi(2S)/J/\psi}$ and the ratio $R_{\psi(2S)/J/\psi}$.

	$\psi(2S)\pi^+$ working point
$F_{\psi(2S)/J/\psi}$	$0.0337 \pm 0.0024\,(\text{stat}) \pm 0.0004\,(\text{syst})$
$R_{\psi(2S)/J/\psi}$	$0.2537 \pm 0.0182\,(\text{stat}) \pm 0.0033\,(\text{syst}) \pm 0.0053\,(\text{BF})$

	$\mu^+\mu^-\pi^+$ working point
$F_{\psi(2S)/J/\psi}$	$0.0378 \pm 0.0037\,(\text{stat}) \pm 0.0005\,(\text{syst})$
$R_{\psi(2S)/J/\psi}$	$0.2847 \pm 0.0276\,(\text{stat}) \pm 0.0037\,(\text{syst}) \pm 0.0059\,(\text{BF})$

6.2 Search for $B_c^+ \to \pi^+\mu^+\mu^-$

Using the results for the normalisation mode in Table 4.15 and the average efficiency combining all q^2 intervals in Table 4.6, we estimate the sensitivity for the ratio of branching fractions for the non-resonant signal decay mode and the normalisation mode to be

$$\frac{\mathcal{B}(B_c^+ \to \pi^+\mu^+\mu^-)}{\mathcal{B}(B_c^+ \to J/\psi\,(\mu^+\mu^-)\pi^+)} = \frac{1}{N_{J/\psi\,\pi^+}} \cdot \frac{\hat{\varepsilon}_{J/\psi\,\pi^+}}{\hat{\varepsilon}_{\pi^+\mu^+\mu^-}} \cdot \mathcal{B}(J/\psi \to \mu^+\mu^-) \approx 2 \cdot 10^{-5},$$

where we have assumed one single signal decay and used the branching fraction $\mathcal{B}\,(J/\psi \to \mu^+\mu^-)$ in Tab. 5.1.

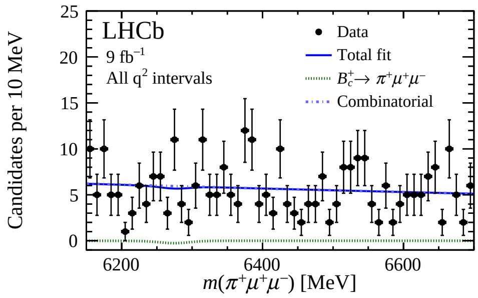
After completing the studies on backgrounds, fit estimator properties (see Sec. 4.3.1) and systematic uncertainties (see Sec. 5), we perform an unblind fit to the data for the non-resonant mode as described in Sec. 4.3.1. Figure 6.1 shows the B_c^+-candidate mass distributions for each q^2 interval and for all intervals combined, with fit projections overlaid. As a cross check, we perform fits in which we fit a background-only model, that is the combinatorial background PDF, to the data (see Fig 6.2 and Tab. 6.2).

No signal over the background-only hypothesis is observed. We parametrise the signal yield in the fits to data (default signal plus background fits) in terms of $R_{\pi^+\mu^+\mu^-/J/\psi\,\pi^+}$ using

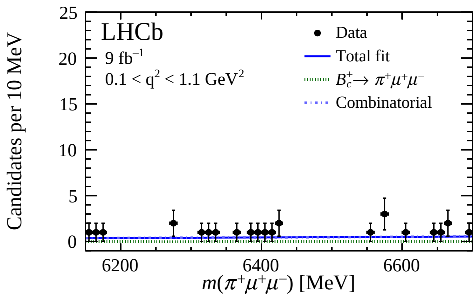
$$R_{\pi^+\mu^+\mu^-/J/\psi\,\pi^+} \equiv \frac{\mathcal{B}(B_c^+ \to \pi^+\mu^+\mu^-)}{\mathcal{B}(B_c^+ \to J/\psi\,(\mu^+\mu^-)\pi^+)} = \frac{N_{\pi^+\mu^+\mu^-}}{N_{J/\psi\,\pi^+}} \cdot \frac{\varepsilon_{J/\psi\,\pi^+}}{\varepsilon_{\pi^+\mu^+\mu^-}} \cdot \mathcal{B}(J/\psi \to \mu^+\mu^-).$$

Table 6.3 shows the fit results. An upper limit on the branching fraction is obtained following the Feldmann–Cousins prescription [36]: pseudoexperiments are generated for various values of $R_{\pi^+\mu^+\mu^-/J/\psi\,\pi^+}$ and the resulting distribution of measured

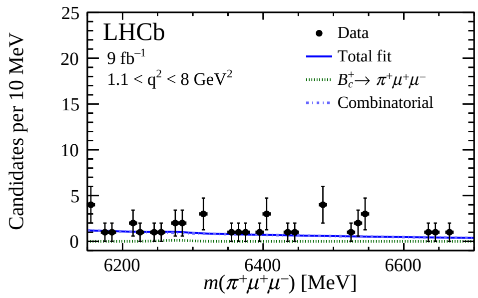
$R_{\pi^+\mu^+\mu^-/J/\psi\pi^+}$ is used to form confidence belts. Figure 6.3 shows confidence belts at 90% and 95% CL. Table 6.4 shows the results obtained for the upper limits.



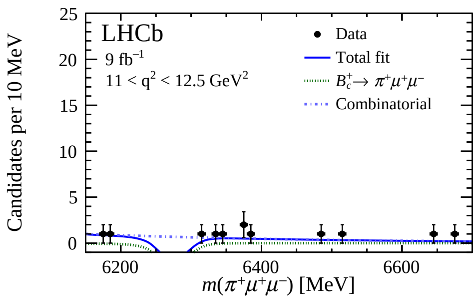
All q^2 bins

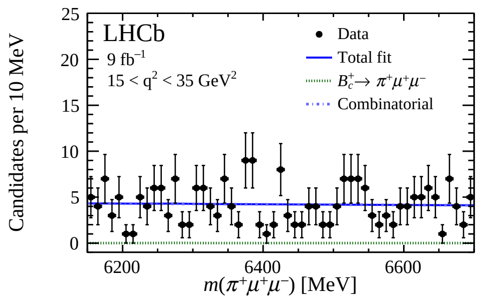


$0.1 < q^2 < 1.1\,\mathrm{GeV}^2/c^4$



$1.1 < q^2 < 8\,\mathrm{GeV}^2/c^4$

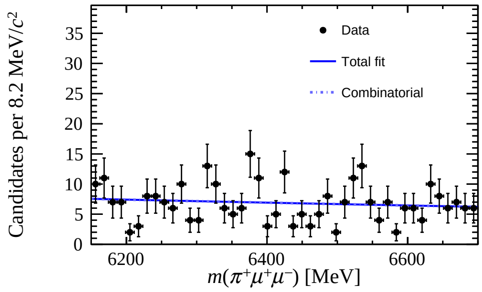




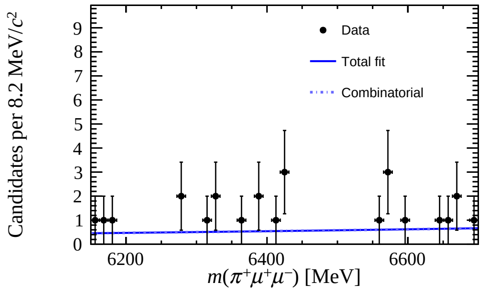
$11 < q^2 < 12.5\,\mathrm{GeV}^2/c^4$

$15 < q^2 < 35\,\mathrm{GeV}^2/c^4$

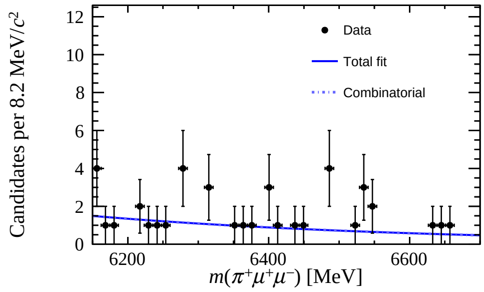
Figure 6.1: Distributions of B_c^+-candidate mass in data for the four q^2 bins and for all q^2 bins combined. Projection of maximum-likelihood fits of the signal plus background model are overlaid.



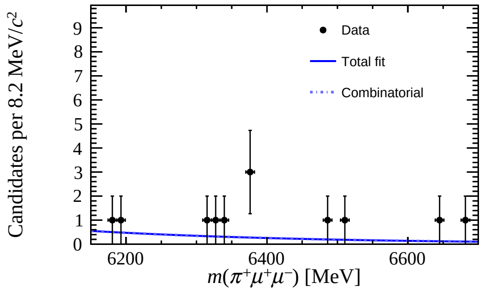
All q^2 bins



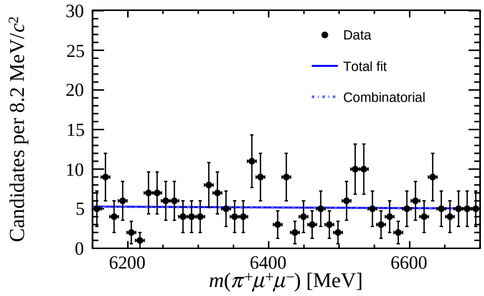
$0.1 < q^2 < 1.1\,\mathrm{GeV}^2/c^4$



$1.1 < q^2 < 8\,\mathrm{GeV}^2/c^4$



$11 < q^2 < 12.5\,\mathrm{GeV}^2/c^4$



$15 < q^2 < 35\,\mathrm{GeV}^2/c^4$

Figure 6.2: Distributions of B_c^+-candidate mass in data for the four q^2 bins and for all q^2 bins combined. Projection of maximum-likelihood fits of a background-only model are overlaid.

Table 6.2: Results of the fit of the background-only model to the B_c^+-candidate mass distribution of $B_c^+ \to \pi^+\mu^+\mu^-$ candidates in data. The combinatorial background exponent ("Comb. bkg. exp.") is quoted in units of $(\mathrm{MeV}/c^2)^{-1}$. The number of candidates in the sample is also provided.

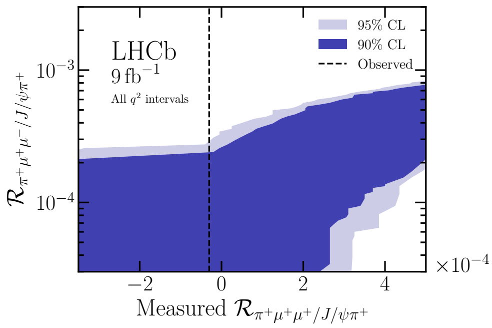
Fit parameters	$0.1 < q^2 < 1.1$	$1.1 < q^2 < 8$	$11 < q^2 < 12.5$	$15 < q^2 < 35$	All
Comb. bkg. exp. $[10^{-3}]$	0.71 ± 1.26	-2.16 ± 1.03	-1.01 ± 1.83	-0.09 ± 0.41	-0.32 ± 0.36
Candidates	25	40	12	232	309

Table 6.3: Results of the fit to the B_c^+-candidate mass distribution of $B_c^+ \to \pi^+\mu^+\mu^-$ candidates in data. The combinatorial background exponent ("Comb. bkg. exp.") is quoted in units of $(\text{MeV}/c^2)^{-1}$.

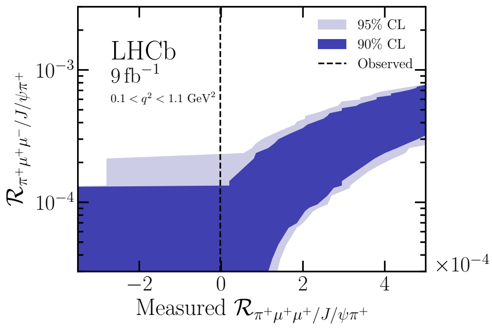
Fit parameters	$0.1 < q^2 < 1.1$	$1.1 < q^2 < 8$	$11 < q^2 < 12.5$	$15 < q^2 < 35$	All
$R_{\pi^+\mu^+\mu^-/J/\psi\pi^+}$	-0.20 ± 4.04	1.46 ± 7.10	-28.4 ± 12.7	0.22 ± 11.0	-3.02 ± 14.4
Comb. bkg. exp. $[10^{-3}]$	0.69 ± 1.35	-2.09 ± 1.09	-3.29 ± 1.30	-0.01 ± 0.46	-0.35 ± 0.39
Comb. bkg. yield	25 ± 5	39 ± 7	30 ± 9	232 ± 17	311 ± 20
Shift $[\text{MeV}/c^2]$	-3.17 ± 0.45	-3.17 ± 0.45	-3.47 ± 0.47	-3.17 ± 0.45	-3.17 ± 0.45
Singl. evt. sens. $[10^{-5}]$	2.14 ± 0.57	2.16 ± 0.58	1.55 ± 0.42	1.46 ± 0.39	1.71 ± 0.46
Scale	1.17 ± 0.03	1.17 ± 0.03	1.14 ± 0.03	1.17 ± 0.03	1.17 ± 0.03
Signal yield	-0 ± 2	1 ± 3	-18 ± 8	0 ± 8	-2 ± 8

Table 6.4: Upper limit (UL) on $R_{\pi^+\mu^+\mu^-/J/\psi\pi^+}$ obtained at 90% and 95% CL. from the results of the fits with nuisance parameters constrained.

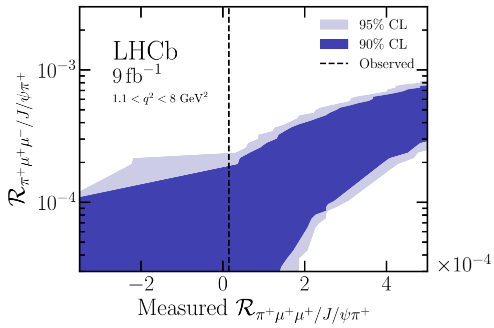
Fit parameters	$0.1 < q^2 < 1.1$	$1.1 < q^2 < 8$	$11 < q^2 < 12.5$	$15 < q^2 < 35$	All
UL at 90% CL	1.28×10^{-4}	1.73×10^{-4}	0.62×10^{-4}	1.90×10^{-4}	2.06×10^{-4}
UL at 95% CL	2.14×10^{-4}	2.16×10^{-4}	0.73×10^{-4}	2.34×10^{-4}	2.74×10^{-4}



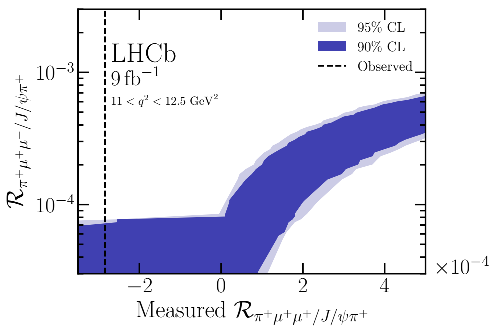
All q^2 bins



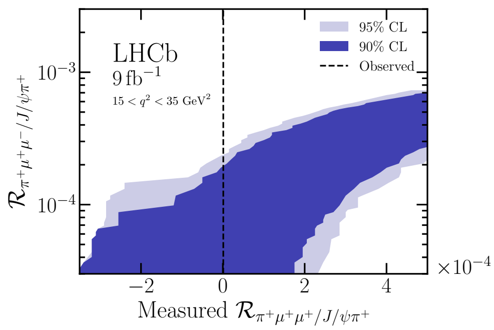
$0.1 < q^2 < 1.1\,\text{GeV}^2/c^4$



$1.1 < q^2 < 8\,\text{GeV}^2/c^4$



$11 < q^2 < 12.5\,\text{GeV}^2/c^4$



$15 < q^2 < 35\,\text{GeV}^2/c^4$

Figure 6.3: Confidence bands at the 90 and 95% interval using Feldman-Cousins hypothesis testing.

6.3 Summary

No evidence for an excess of signal events over background is observed and an upper limit is set on the branching fraction ratio

$$\frac{\mathcal{B}(B_c^+ \to \pi^+\mu^+\mu^-)}{\mathcal{B}(B_c^+ \to J/\psi\,\pi^+)} < 2.1 \times 10^{-4} \tag{6.3}$$

at the 90% confidence level. Additionally, we can finally report an updated measurement of the ratio of the $\mathcal{B}(B_c^+ \to \psi(2S)\pi^+)/\mathcal{B}(B_c^+ \to J/\psi\,\pi^+)$ branching fractions. The ratio

$$\frac{\mathcal{B}(B_c^+ \to \psi(2S)\pi^+)}{\mathcal{B}(B_c^+ \to J/\psi\,\pi^+)} = 0.254 \pm 0.018 \pm 0.003 \pm 0.005 \tag{6.4}$$

is reported, where the first uncertainty is statistical, the second systematic, and the third is due to the uncertainties on the branching fractions of the leptonic J/ψ and $\psi(2S)$ decays. This measurement is the most precise to date and is consistent with previous LHCb results.

Appendix A

Simulated Data-set Specifics

Tables A.1-A.2 show the full Monte-Carlo simulation data specification on the mode, year, and number of events used for this study.

Table A.1: Simulated samples used in the analysis (1/2).

Mode	Year	Events
$B_c^+ \to \pi^+\mu^+\mu^-$	2011	1004025
$B_c^+ \to \pi^+\mu^+\mu^-$	2012	1028704
$B_c^+ \to \pi^+\mu^+\mu^-$	2015	1008671
$B_c^+ \to \pi^+\mu^+\mu^-$	2016	1024898
$B_c^+ \to \pi^+\mu^+\mu^-$	2017	1006686
$B_c^+ \to \pi^+\mu^+\mu^-$	2018	994797
$B_c^+ \to J/\psi\,(\mu^+\mu^-)\,\pi^+$	2011	2098499
$B_c^+ \to J/\psi\,(\mu^+\mu^-)\,\pi^+$	2012	2008771
$B_c^+ \to J/\psi\,(\mu^+\mu^-)\,\pi^+$	2015	2020299
$B_c^+ \to J/\psi\,(\mu^+\mu^-)\,\pi^+$	2016	2006687
$B_c^+ \to J/\psi\,(\mu^+\mu^-)\,\pi^+$	2017	2034929
$B_c^+ \to J/\psi\,(\mu^+\mu^-)\,\pi^+$	2018	2079883
$B_c^+ \to \psi(2S)(\mu^+\mu^-)\,\pi^+$	2011	504004
$B_c^+ \to \psi(2S)(\mu^+\mu^-)\,\pi^+$	2012	1019292
$B_c^+ \to \psi(2S)(\mu^+\mu^-)\,\pi^+$	2015	519134
$B_c^+ \to \psi(2S)(\mu^+\mu^-)\,\pi^+$	2016	1012380
$B_c^+ \to \psi(2S)(\mu^+\mu^-)\,\pi^+$	2017	1118750
$B_c^+ \to \psi(2S)(\mu^+\mu^-)\,\pi^+$	2018	1011110
$B_c^+ \to B_s^{*0}(\mu^+\mu^-)\,\pi^+$	2011	1000522
$B_c^+ \to B_s^{*0}(\mu^+\mu^-)\,\pi^+$	2012	1004874
$B_c^+ \to B_s^{*0}(\mu^+\mu^-)\,\pi^+$	2015	988270
$B_c^+ \to B_s^{*0}(\mu^+\mu^-)\,\pi^+$	2016	1009989
$B_c^+ \to B_s^{*0}(\mu^+\mu^-)\,\pi^+$	2017	1005911
$B_c^+ \to B_s^{*0}(\mu^+\mu^-)\,\pi^+$	2018	1008931
$B_c^+ \to B^{*0}(\mu^+\mu^-)\,\pi^+$	2011	1002597
$B_c^+ \to B^{*0}(\mu^+\mu^-)\,\pi^+$	2012	992994
$B_c^+ \to B^{*0}(\mu^+\mu^-)\,\pi^+$	2015	973157
$B_c^+ \to B^{*0}(\mu^+\mu^-)\,\pi^+$	2016	980121
$B_c^+ \to B^{*0}(\mu^+\mu^-)\,\pi^+$	2017	1005878
$B_c^+ \to B^{*0}(\mu^+\mu^-)\,\pi^+$	2018	1013680

Table A.2: Simulated samples used in the analysis (2/2).

Mode	Year	Events
$B_c^+ \to J/\psi\,(\mu^+\mu^-)\,K^+$	2011	992696
$B_c^+ \to J/\psi\,(\mu^+\mu^-)\,K^+$	2012	979923
$B_c^+ \to J/\psi\,(\mu^+\mu^-)\,K^+$	2015	1010903
$B_c^+ \to J/\psi\,(\mu^+\mu^-)\,K^+$	2016	1061419
$B_c^+ \to J/\psi\,(\mu^+\mu^-)\,K^+$	2017	991500
$B_c^+ \to J/\psi\,(\mu^+\mu^-)\,K^+$	2018	1004524
$B_c^+ \to \psi(2S)(\mu^+\mu^-)\,K^+$	2011	946601
$B_c^+ \to \psi(2S)(\mu^+\mu^-)\,K^+$	2012	1010514
$B_c^+ \to \psi(2S)(\mu^+\mu^-)\,K^+$	2015	1001427
$B_c^+ \to \psi(2S)(\mu^+\mu^-)\,K^+$	2016	1003804
$B_c^+ \to \psi(2S)(\mu^+\mu^-)\,K^+$	2017	1008553
$B_c^+ \to \psi(2S)(\mu^+\mu^-)\,K^+$	2018	1009803
$B_c^+ \to J/\psi\,(\mu^+\mu^-)\,\rho^+(\pi^+\pi^0)$	2011	1033320
$B_c^+ \to J/\psi\,(\mu^+\mu^-)\,\rho^+(\pi^+\pi^0)$	2012	2006796
$B_c^+ \to J/\psi\,(\mu^+\mu^-)\,\rho^+(\pi^+\pi^0)$	2015	111092
$B_c^+ \to J/\psi\,(\mu^+\mu^-)\,\rho^+(\pi^+\pi^0)$	2016	1964636
$B_c^+ \to J/\psi\,(\mu^+\mu^-)\,\rho^+(\pi^+\pi^0)$	2017	1965071
$B_c^+ \to J/\psi\,(\mu^+\mu^-)\,\rho^+(\pi^+\pi^0)$	2018	2015590
$B_c^+ \to \psi(2S)(\mu^+\mu^-)\,\rho^+(\pi^+\pi^0)$	2011	894423
$B_c^+ \to \psi(2S)(\mu^+\mu^-)\,\rho^+(\pi^+\pi^0)$	2012	923206
$B_c^+ \to \psi(2S)(\mu^+\mu^-)\,\rho^+(\pi^+\pi^0)$	2015	934466
$B_c^+ \to \psi(2S)(\mu^+\mu^-)\,\rho^+(\pi^+\pi^0)$	2016	864851
$B_c^+ \to \psi(2S)(\mu^+\mu^-)\,\rho^+(\pi^+\pi^0)$	2017	990743
$B_c^+ \to \psi(2S)(\mu^+\mu^-)\,\rho^+(\pi^+\pi^0)$	2018	842384
$B_c^+ \to \rho^+(\pi^+\pi^0)\,\mu^+\mu^-$	2011	996869
$B_c^+ \to \rho^+(\pi^+\pi^0)\,\mu^+\mu^-$	2012	997334
$B_c^+ \to \rho^+(\pi^+\pi^0)\,\mu^+\mu^-$	2015	1004301
$B_c^+ \to \rho^+(\pi^+\pi^0)\,\mu^+\mu^-$	2016	1010375
$B_c^+ \to \rho^+(\pi^+\pi^0)\,\mu^+\mu^-$	2017	1009482
$B_c^+ \to \rho^+(\pi^+\pi^0)\,\mu^+\mu^-$	2018	991782

Appendix B

Simulation Corrections

The simulated samples have several known deficiencies, in particular the event occupancy and $B_{(c)}^+$ kinematics are known to be mismodelled. To ensure the correct calculation of efficiencies these deficiencies are corrected in a data-driven way. Each simulated event is given a weight such that the distributions of the problematic variables match those of a clean data sample. The total weight for each event is a product of five individual weights, each of them correcting for one particular discrepancy.

We account for discrepancies associated with the PID selection, with the track reconstruction, with the muon ID selection in the stripping line, with the trigger selection and with the simulation of B_c^+ kinematics. Corrections related with the discrepancies due to PID, track reconstruction, muon ID and trigger selection are derived in a way common to all $B_{(c)}^+ \to \pi^+\mu^+\mu^-$ decays [25]. These corrections are derived for each year of data-taking and magnet polarity to study systematic effects separately for different center-of-mass energies and particle charges. Corrections related with the B_c^+ kinematics are derived specifically for this analysis, separately for each run of data taking and combining both magnet polarities.

We find another small discrepancy between data and simulation associated with the B_c^+ lifetime. At generator level, the B_c^+ lifetime in simulation is 0.507 ps, while the PDG lifetime is (0.510 ± 0.009) ps [3] (see Table 3.2). Since the value used in MC is well within the 1σ range of the current best knowledge, we do not apply any correction, but rather consider the effect of the uncertainty on the PDG value as as source of systematic uncertainty (see Sec. 5).

In this section we show example histograms illustrating the data-simulation corrections. The comprehensive collection of histograms illustrating the data-simulation corrections can be found in Ref. [25].

B.1 PID Selection

We weight the candidates reconstructed in simulation using PID efficiency histograms obtained from data with the PIDCalib package [37]. The histograms containing the efficiencies are obtained separately for each year, magnet polarity and track charge. For the B-candidate, the weight corresponds to the product of the three weights corresponding to the three daughters.

The histograms for the muon candidates are obtained by requiring that the muon candidates in the PIDCalib samples survive the muon PID requirements applied by the `B2XMuMu` stripping line.

We account separately for the effect of the stripping line PID selection using a different set of weights (see next section).

Figure B.1 shows example 2D projections of the PIDCalib histograms for muons and pions.

Table B.1: Selections applied to PIDCalib samples to obtain efficiencies. The preselection is applied to numerator and denominator. The selection only to the numerator. For muons in the PIDCalib samples we additionally require `Unbias_HLT1` $== 1$ for run 1 and `MuonUnbiased` $== 1$ for run 2. The values of x and y are adapted based on the chosen working point (see Tab. 3.5).

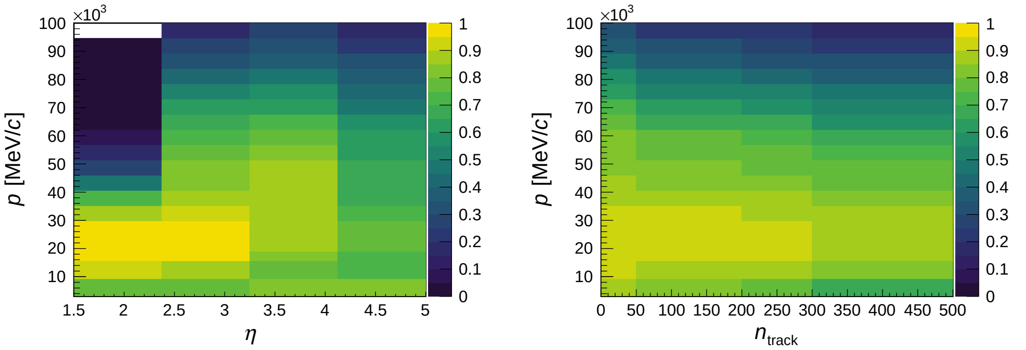
Candidate	Preselection	Selection
μ	`DLLmu` > -3.0 `&&` `IsMuon` $== 1$	`ProbNNmu` > 0.2
π		`ProbNNpi` $> x$ `&&` `ProbNNK` $< y$ `&&` `IsMuon` $== 0$ `&&` `IsMuonLoose` $== 0$ `&&` `InMuonAcc` $== 1$

B.1.1 Muon ID Requirements

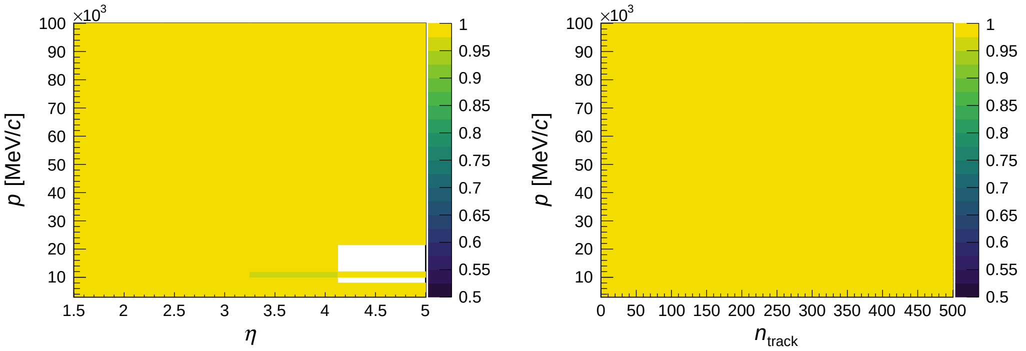
The `B2XMuMu` stripping line applies the PID requirement `PIDmu > -3` and `IsMuon` on muon candidates. To correct for possible data-simulation discrepancies, we extract muon ID weights in bins of p, η and n_{tracks} as the ratio of the efficiency between data and simulation. The weights containing the efficiencies are obtained separately for each year, magnet polarity and track charge.

The efficiency in data is estimated using the PIDCalib package [37].

The efficiency in simulation is obtained from the $B^- \to J/\psi\,(\to \mu^+\mu^-)\,K^-$ simulation samples. For this, we reconstruct J/ψ candidates such that they fulfill the requirements in the `B2XMuMu` stripping line except for the `PIDmu > -3` and `IsMuon` requirements on the muon candidates. Figure B.2 compares the kinematic



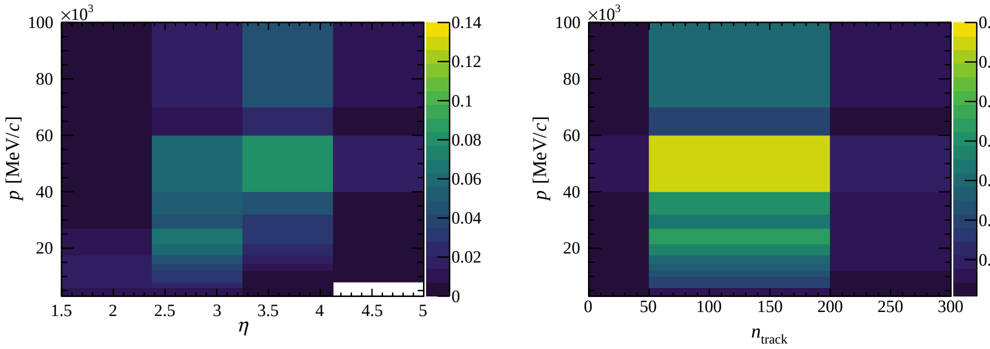
(a) π^+ PID efficiency, for non-resonant signal selection requirements



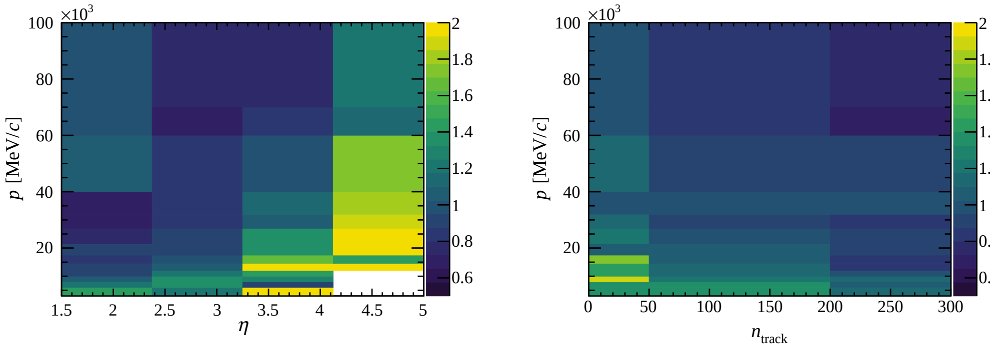
(b) μ^+ PID efficiency

Figure B.1: 2D projections of PIDCalib efficiency histograms for data collected in 2016 with magnet polarity up.

distribution for muons in $B^- \to J/\psi\,(\to \mu^+\mu^-)\,K^-$ and decay$B_c^+\pi^+\mu^+\mu^-$ simulation, showing that they are very similar. Figure B.3 shows example 2D projections of the muon ID correction histograms. The histogram is empty for regions that are not populated in the simulation.



(a) Kinematic distributions for μ^+ in $B^+ \to J/\psi\,K^+$ MC



(c) Kinematic distributions for μ^+ in the $B^+ \to J/\psi\,K^+$ MC divided by the kinematic distributions for μ^+ in $B_c^+ \to \pi^+\mu^+\mu^-$ MC

Figure B.2: Normalized 2D distributions of (left) p vs. η and (right) p vs. n_{tracks} obtained from simulation.

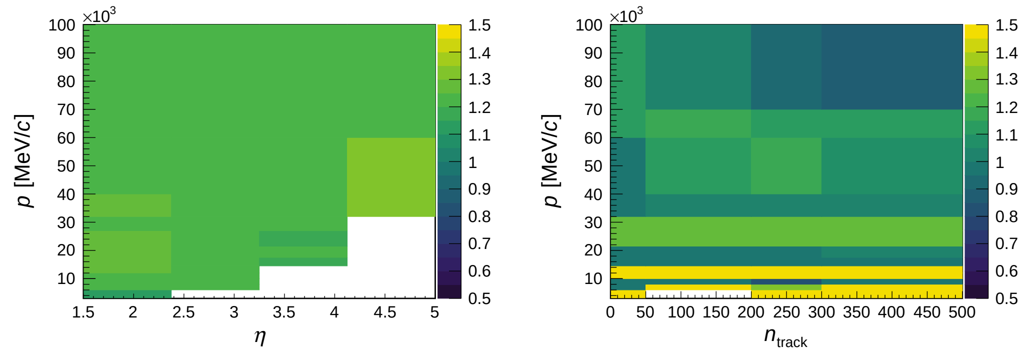


Figure B.3: 2D projections of the muon ID correction histogram for data collected in 2016 with magnet polarity up.

B.1.2 Track Reconstruction

We correct for data-simulation discrepancies associated with the efficiency of the charged-particle track reconstruction by giving to each track candidate a weight obtained from the available TrackCalib histograms [38]. The histograms containing the weights are available separately for each year as functions of p and η. Example histograms are shown in Figure B.4. For the B-candidate, the weight corresponds to the product of the three weights corresponding to the three daughters.

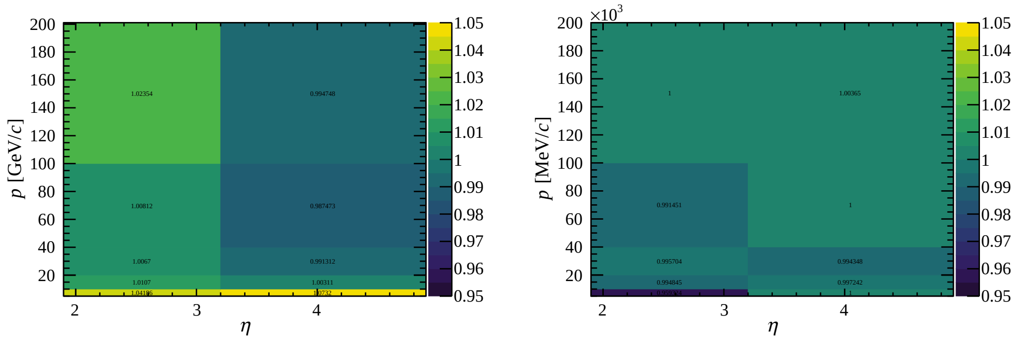


Figure B.4: TrackCalib histograms for (left) 2012 and (right) 2016 data.

B.2 Trigger Selection

The trigger cuts are also split into two types, Trigger On Signal (TOS) and Trigger Independent of Signal (TIS). TOS class is given if a signal candidate is sufficient to activate the trigger itself while TIS is classified if something else in the event activates the trigger.

For this analysis, we only use the `L0Muon` line of the L0 trigger.

The `L0Muon` trigger efficiency is determined in data and simulation with the `TIS TOS` method using samples of $B^- \to J/\psi\,(\to \mu^+\mu^-)\,K^-$ decays selected by the `B2XMuMu` stripping line. The B candidates are required to be `TOS` on `Hlt1TrackMVA` (that is a μ or K candidate must be responsible for the trigger decision) and `Hlt2Topo2Body`. The J/ψ candidate is required to be `L0Global TIS` at `L0`. In addition, the K candidate is required to have `ProbNNk` > 0.1 to reject background from misidentified $B^- \to J/\psi\,(\to \mu^+\mu^-)\,\pi^-$ decays. The `HLT` 1 and 2 lines are chosen to avoid biasing the `L0Muon` trigger efficiency estimate. The efficiency is then determined in bins of the transverse momenta (p_{T}) of the μ^+ and the μ^-.

In data, the efficiency is determined by fitting separately candidates that satisfy `TIS` and `TOS` and candidates that are `TIS` and not `TOS`. The resulting efficiency and uncertainty are given by

$$\varepsilon = \frac{N_{\rm TOS}}{N_{\rm TOS} + N_{\rm NOT}}, \quad \text{and} \quad \sigma_\varepsilon = \varepsilon\left(\left(\sigma_{\rm TOS}\cdot\frac{N_{\rm NOT}}{N_{\rm TOS}}\right)^2 + \sigma^2_{\rm NOT}\right)^{\frac{1}{2}},$$

where $N_{\rm TOS}$ is the yield of candidates that are `TIS` and `TOS`, and $N_{\rm NOT}$ is the yield of candidates that are `TIS` and not `TOS`. In the fits (see example in Fig. B.5), the $B^- \to J/\psi(\to \mu^+\mu^-)K^-$ signal shape is described by a Gaussian function with power-law tails. The tail parameters are fixed from fits to simulation. The peak position and width of the shape are allowed to vary between $p_{\rm T}$ bins and are fixed from the combined `TIS` sample.

Figure B.6 shows example correction histograms corresponding to ratio between the efficiency in data and simulation.

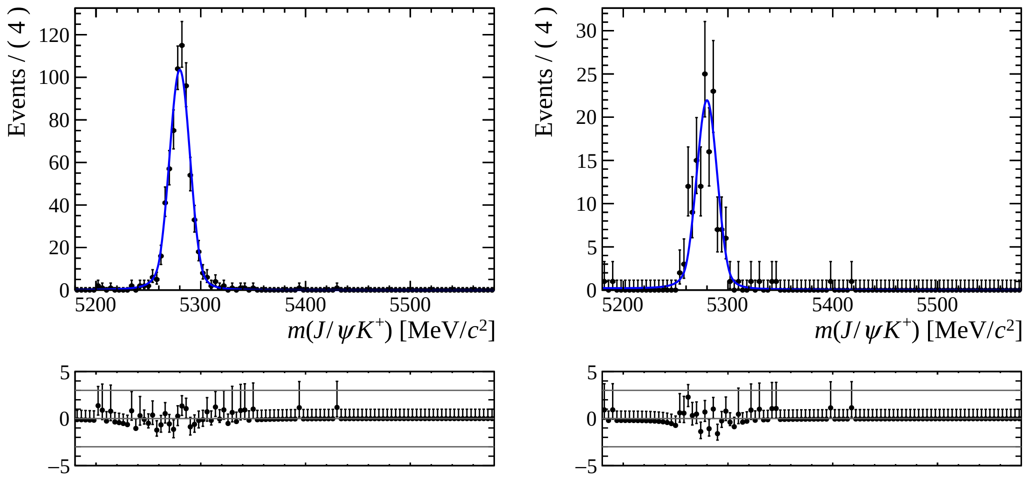


Figure B.5: Example of fits to determine the yield of candidates that are (left) `TIS` and `TOS` and (right) `TIS` and not `TOS` for the interval $1905 < p_{\rm T} < 2170\,{\rm MeV}/c$ for μ^+ and $870 < p_{\rm T} < 1150$ for μ^-.

B.3 Corrections for Data-simulation Disagreements in MVA Inputs

To account for data-simulation discrepancies in the occupancy and in the B_c^+ kinematics, we correct for discrepancies in the distributions of $n_{\rm tracks}$, $p_{\rm T}$ and $\chi^2_{\rm vtx}/n_{\rm dof}$. To maintain correlations between these variables, an MVA is used to derive the weight for each event. The MVA is trained to separate simulated samples and real data, and a weight (This method also known as *sWeighting*) taken as the

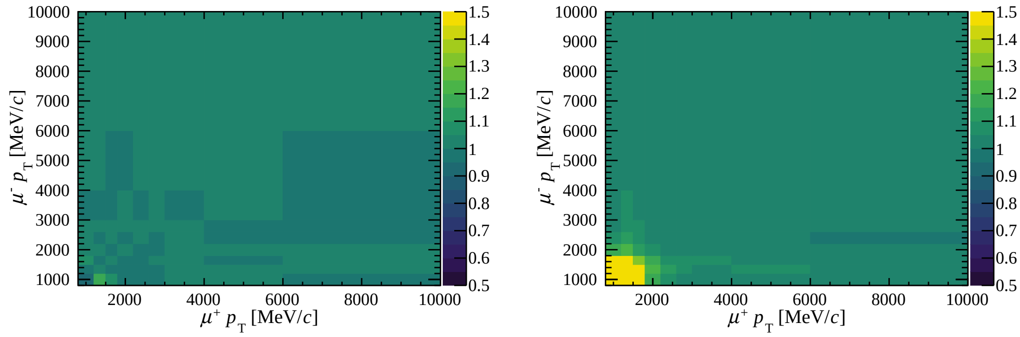


Figure B.6: Ratio between `L0Muon` efficiency in data and simulation obtained with the `TIS TOS` method for (left) 2012 and (right) 2016 data with magnet polarity up.

ratio of probabilities of an event to be data-like and simulation-like, based only on these variables.

The data sample for calibration is $B_c^+ \to J/\psi\,\pi^+$. This sample is very loosely selected by applying the selection requirements shown in Table B.2 together with the trigger selection. The background is statistically subtracted using the *sPlot* technique [39]. Figure B.7 shows example fits to the loosely selected $B_c^+ \to J/\psi\,\pi^+$ decays from which the *sWeights* are derived. A simulated sample of $B_c^+ \to J/\psi\,(\mu^+\mu^-)\pi^+$ corrected for muon-ID, trigger and tracking efficiency (see previous subsections) is used to derive the weights. Separate sets of weights are derived for each run of data taking, but with both magnet polarities combined. The obtained weights are used to reweight all simulated modes, not only $B_c^+ \to J/\psi\,(\mu^+\mu^-)\pi^+$.

Using the *sWeighted* data and the full simulation corrected for muon-ID, trigger and tracking efficiency, we train a gradient boosted decision tree algorithm as implemented in the hep_ml reweighting package [40]. Figure B.8 shows the distribution of weights obtained from the MVA on simulated $B_c^+ \to J/\psi\,\pi^+$ events.

Figures B.9–B.12 show distributions of various observables for *sWeighted data*, raw simulation (with muon-ID, trigger and tracking efficiency corrections) and reweighted simulation (with all corrections). Figures B.9 and B.11 include the variables that are explicitly part of the reweighting procedure; other variables are also shown to examine for residual disagreements. In a similar way, Figure B.13 shows the output distribution of the BDT used to suppress combinatorial background for Run I and for Run II of data taking. We observe generally good data-simulation agreement after applying all corrections to simulation, except for a residual distribution in the χ^2_{IP} of the muons and in the BDT output to suppress combinatorial background. Therefore, we have trained an alternative reweighting algorithm using χ^2_{IP} of the two muons together with the default inputs. Using the alternative algorithm removes almost fully the residual discrepancies. The differences between the nominal and the alternative algorithm will be considered as systematic uncertainties (see Sec. 5).

So far we have discussed kinematic corrections for full simulation samples and we also produce dedicated p_{T} correction histograms for generator-level samples, which are needed for the estimation of efficiencies (see Sec. 4.1). For this, we obtain normalized p_{T} histograms for the *sWeighted* data and for the full simulation corrected for muon-ID, trigger and tracking efficiency. We divide the histogram for data by that for simulation and use the resulting ratio histogram to obtain p_{T} weights for generator-level B_c^+ particles. Figure B.14 shows the p_{T} correction histograms for Run I and Run II. We choose the binning such that the bins for the p_{T} histograms in data are nearly equally populated. The histograms are obtained independently for each run of data taking combining both magnet polarities. As a cross check we compare the p_{T}, η and azimuthal angle ϕ distributions for *sWeighted* data, raw simulation (with muon-ID, trigger and tracking efficiency corrections) and reweighted simulation (with muon-ID, trigger, tracking efficiency and p_{T} corrections) as shown in Fig. B.15.

Table B.2: Selection requirements used to isolate a clean sample of $B_c^+ \to J/\psi\,(\mu^+\mu^-)\pi^+$ decays to correct the simulation.

Particle	Parameter	Selection
All	p_{T}	$> 300\,\mathrm{MeV}/c$
π^+	`isMuonLoose`	`False`
π^+	`InAccMuon`	`True`
J/ψ	$\lvert m(\mu\mu) - m_{J/\psi}^{\mathrm{PDG}}\rvert$	$< 100\,\mathrm{MeV}/c^2$

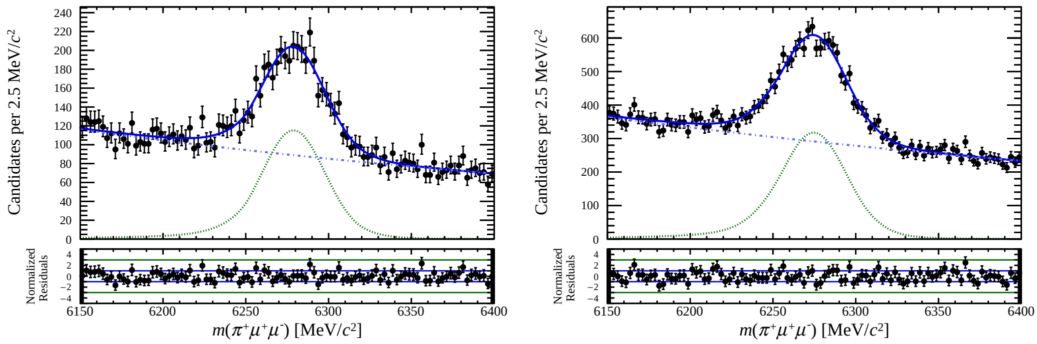


Figure B.7: Loosely selected $B_c^+ \to J/\psi\,(\mu^+\mu^-)\pi^+$ events for (left) Run I and (right) Run II, with both polarities combined, with results of the maximum-likelihood fit used to obtain *sWeights* superimposed. The distributions are shown for the collision data (dots), the full fit model (solid blue line), the signal PDF (dotted green line) and the combinatorial background PDF (dash-dotted blue line).

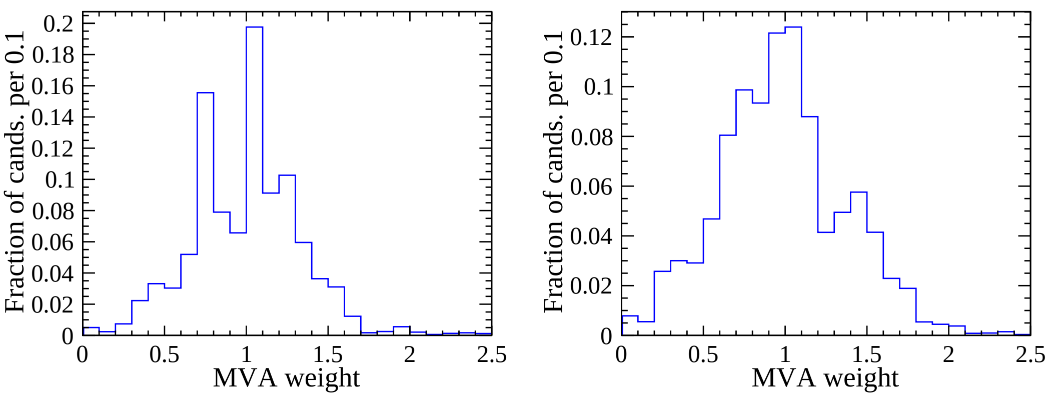


Figure B.8: Distribution of kinematic correction weights on simulated $B_c^+ \to J/\psi\,(\mu^+\mu^-)\pi^+$ events for (left) Run I and (right) Run II.

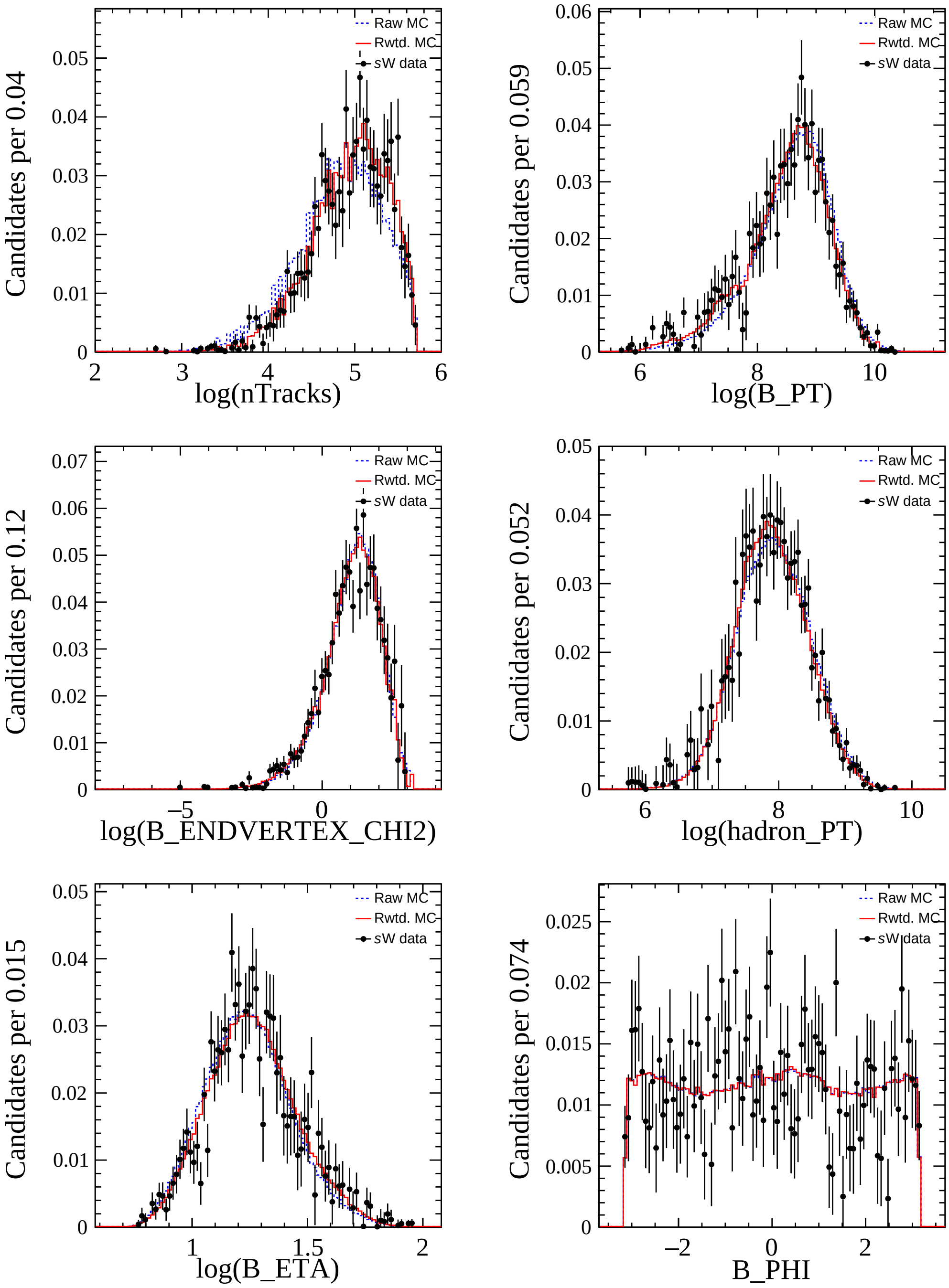


Figure B.9: Distributions of observables for $B_c^+ \to J/\psi\,(\mu^+\mu^-)\pi^+$ candidates reconstructed in *sWeighted* data (black dots), partially corrected simulation (dashed blue line), fully corrected simulation (solid red line) for Run I.

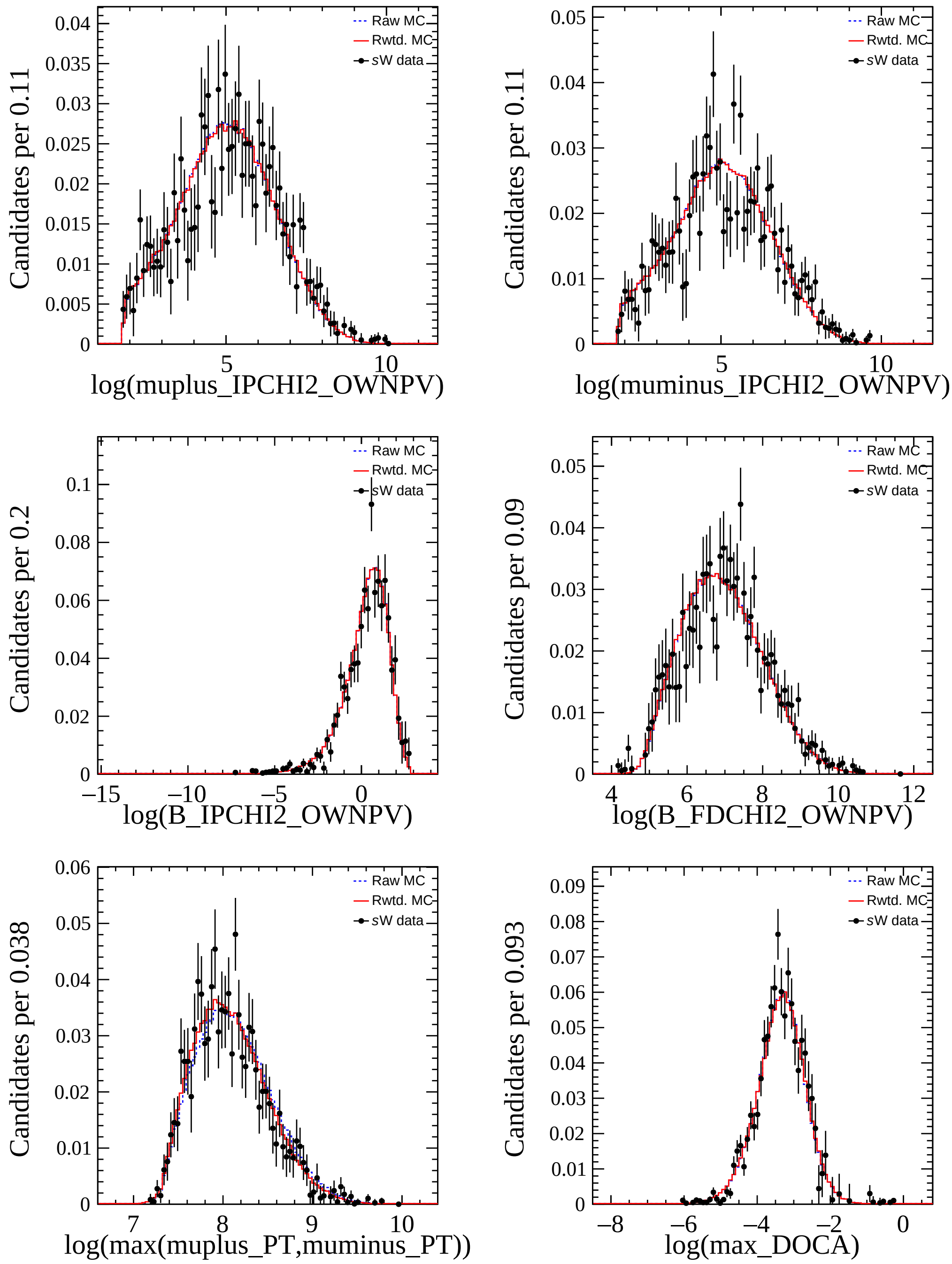


Figure B.10: Distributions of observables for $B_c^+ \to J/\psi(\mu^+\mu^-)\pi^+$ candidates reconstructed in *sWeighted* data (black dots), partially corrected simulation (dashed blue line), fully corrected simulation (solid red line) for Run I.

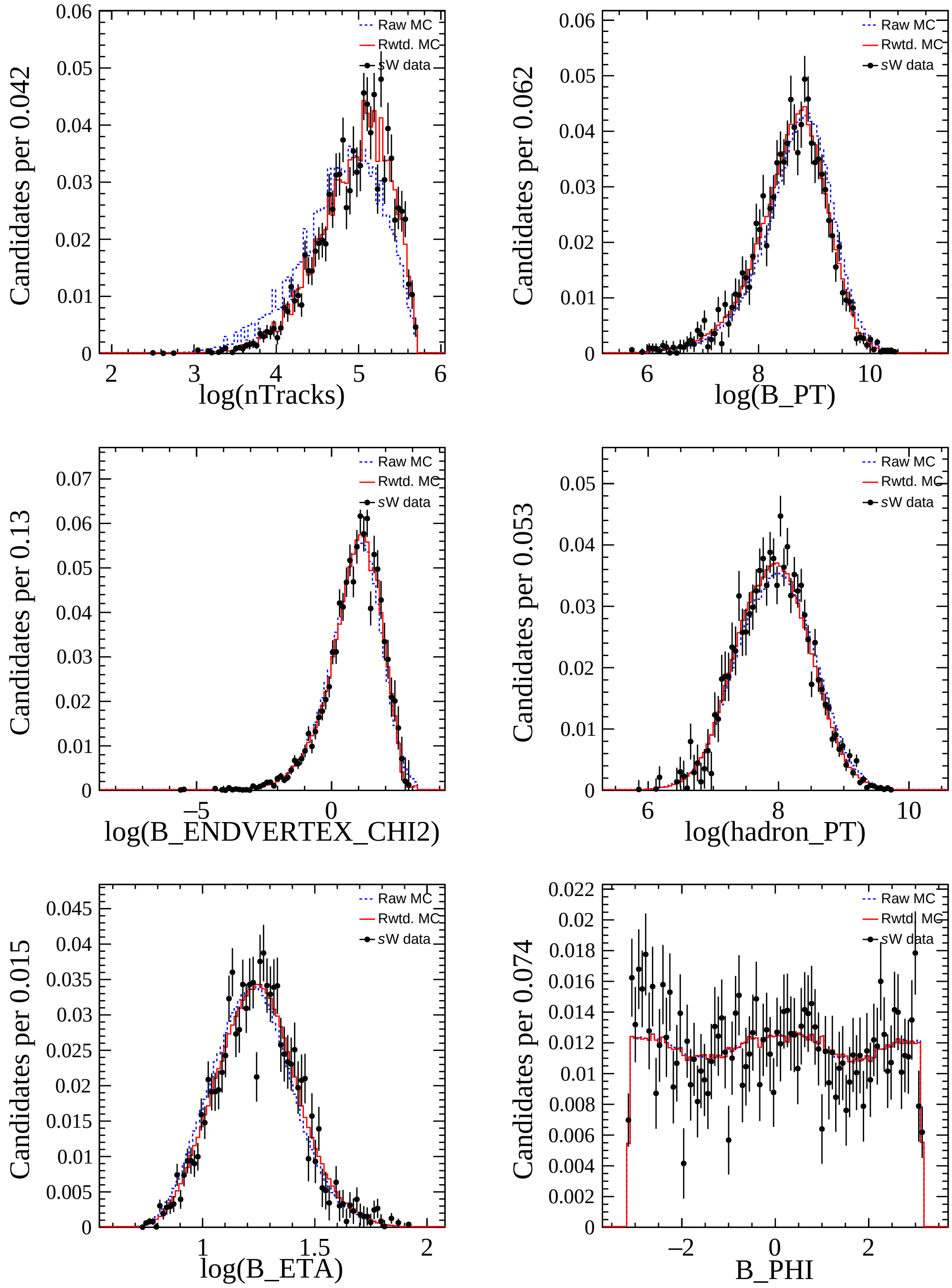


Figure B.11: Distributions of observables for $B_c^+ \to J/\psi(\mu^+\mu^-)\pi^+$ candidates reconstructed in *sWeighted* data (black dots), partially corrected simulation (dashed blue line), fully corrected simulation (solid red line) for Run II.

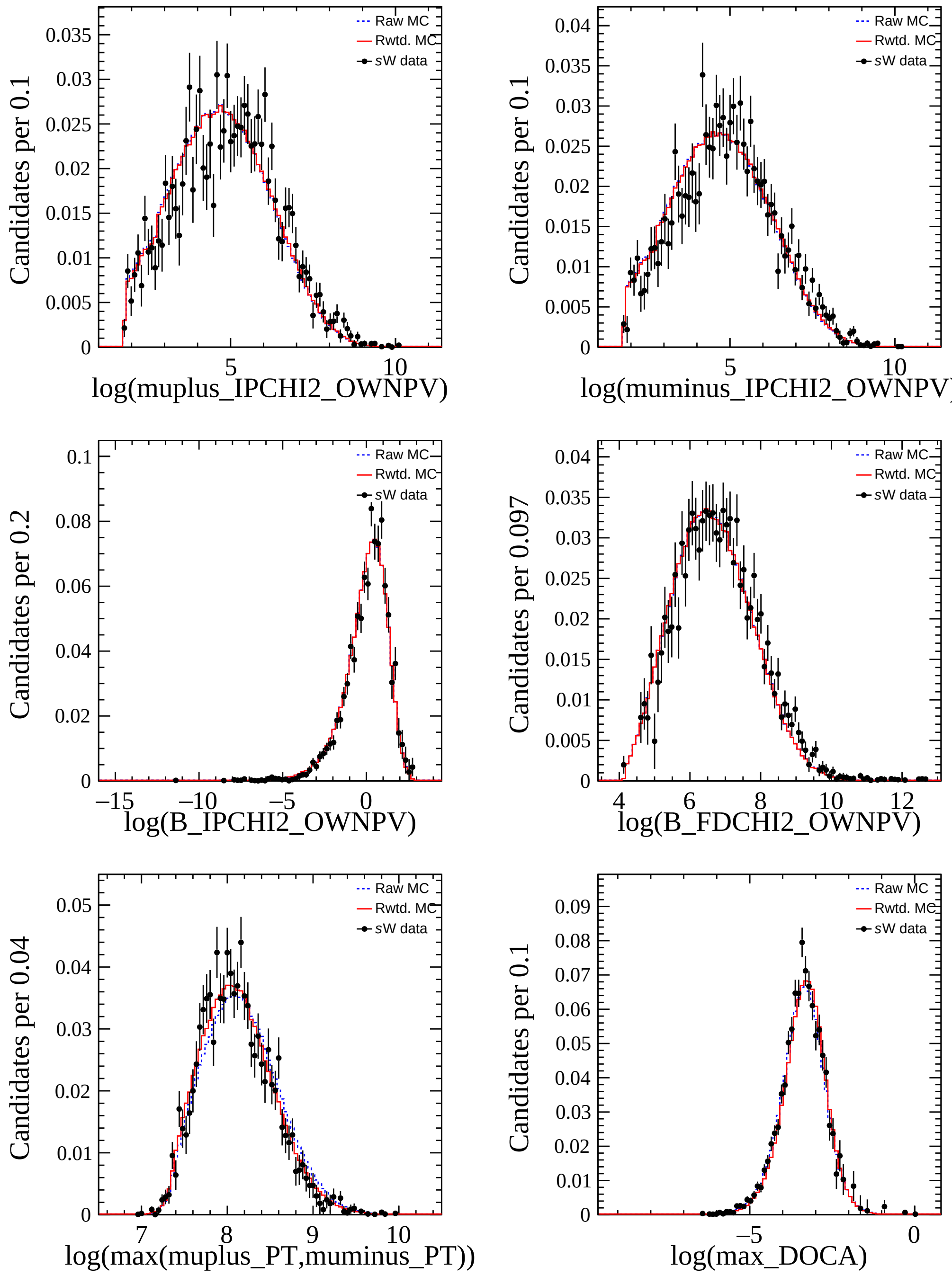


Figure B.12: Distributions of observables for $B_c^+ \to J/\psi(\mu^+\mu^-)\pi^+$ candidates reconstructed in *sWeighted* data (black dots), partially corrected simulation (dashed blue line), fully corrected simulation (solid red line) for Run II.

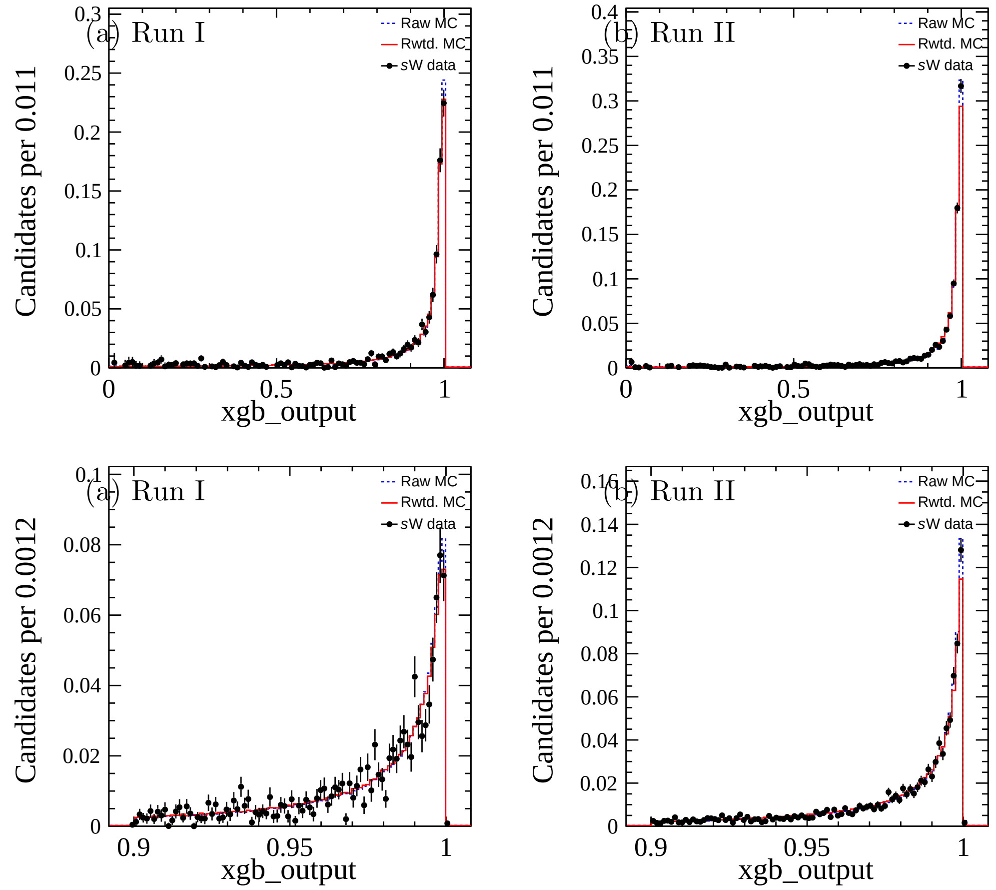


Figure B.13: Distributions of BDT output for $B_c^+ \to J/\psi(\mu^+\mu^-)\pi^+$ candidates reconstructed in *sWeighted* data (black dots), partially corrected simulation (dashed blue line), fully corrected simulation (solid red line) for (left) Run I and (right) Run II.

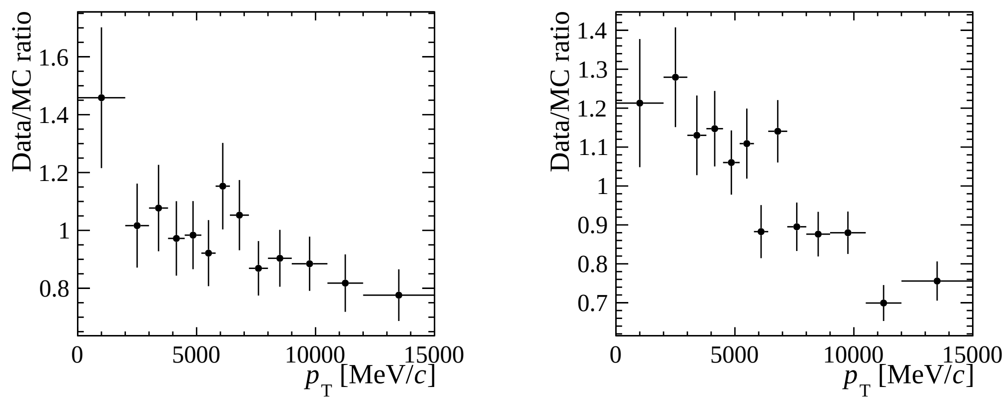


Figure B.14: data-simulation corrections to the B_c^+ p_{T} distribution for (left) Run I and (right) Run II.

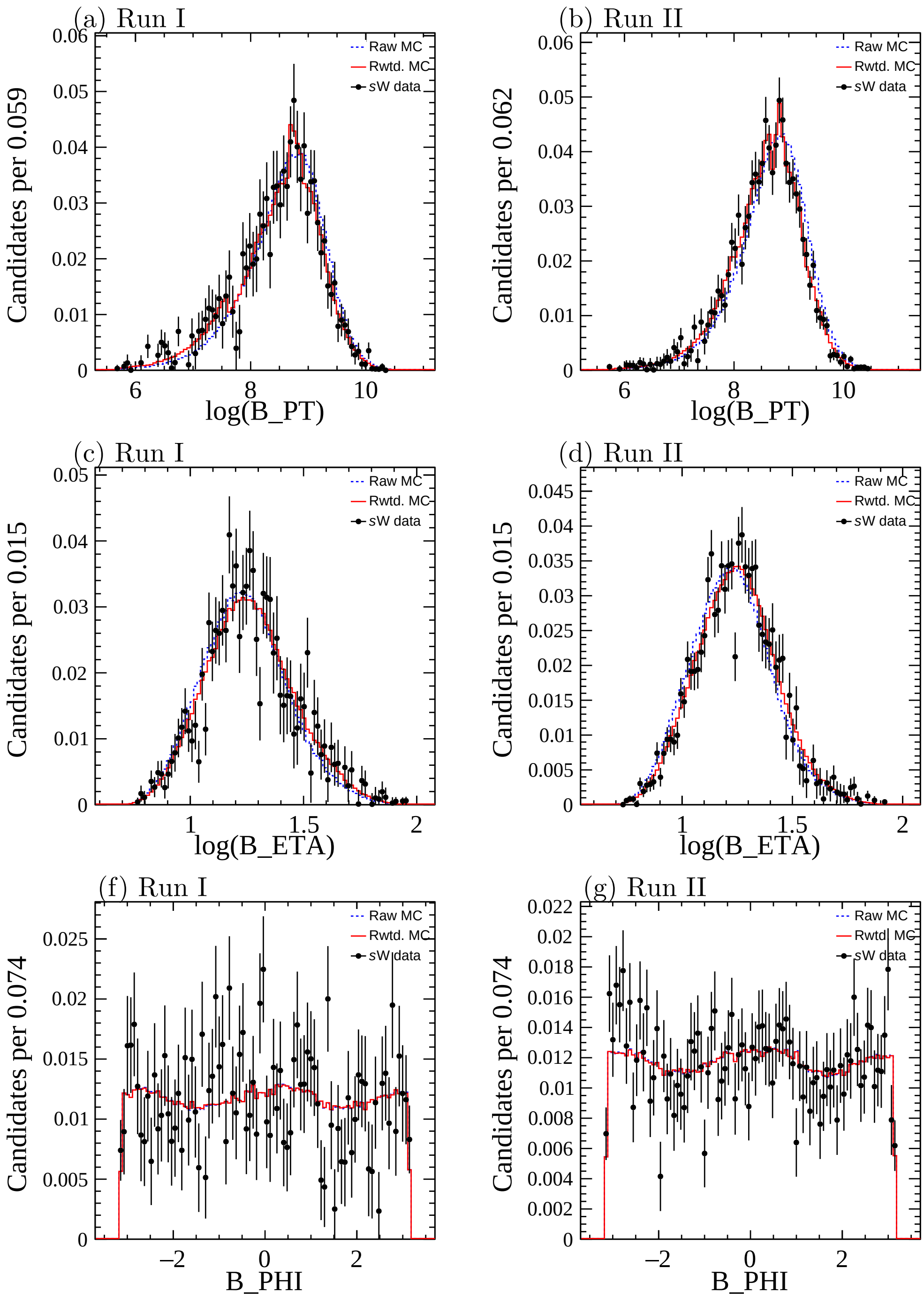


Figure B.15: Distributions of p_{T}, η and ϕ for $B_c^+ \to J/\psi\,(\mu^+\mu^-)\pi^+$ candidates reconstructed in *sWeighted* data (black dots), partially corrected simulation (dashed blue line), fully corrected simulation (solid red line) for (left) Run I and (right) Run II of data taking.

Appendix C

Multiple Candidates

After applying the full selection, the fraction of selected events with multiple candidates for the non-resonant $B_c^+ \to \pi^+\mu^+\mu^-$ decay mode is around 0.05% in simulation. No events with multiple candidates are found in sideband data.

For the $B_c^+ \to \psi(2S)\pi^+$ and $B_c^+ \to J/\psi\,\pi^+$ decay modes, no events with multiple candidates are found in simulation, while the rates in data are around 0.05% or below at both working points. The check on simulation has been done including and excluding the ghost category. No significant difference is found between the two cases. Additionally, we have checked for the presence of cloned tracks. None are present in the final data and simulation samples. Since these rates are negligible, we use all candidates that fulfill the full selection criteria throughout the analysis.

Appendix D

PID Efficiencies

PID efficiencies are directly obtained from PIDCalib histograms for each year and polarity. We check the PID efficiency on our signal samples for the respective working points (see Table 3.5) for Run I and Run II of data taking separately. As Figure D.1 and Table D.1 show, the PID performance for Run I and Run II is comparable.

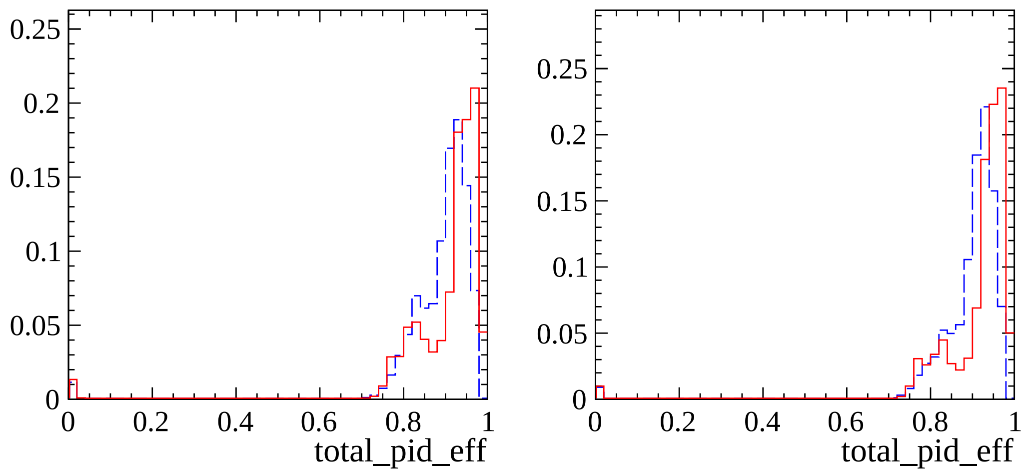


Figure D.1: Normalized histograms of PID efficiency for each signal sample ($\pi^+\mu^+\mu^-$ left, $\psi(2S)\pi^+$ right) at their respective working points for Run I (dashed blue) and Run II (solid red).

Table D.1: Average PID efficiencies for each run for each signal sample.

	$\pi^+\mu^+\mu^-$	$\psi(2S)\pi^+$
Run 1	0.883	0.892
Run 2	0.898	0.910

Appendix E

MVA Optimisation Studies

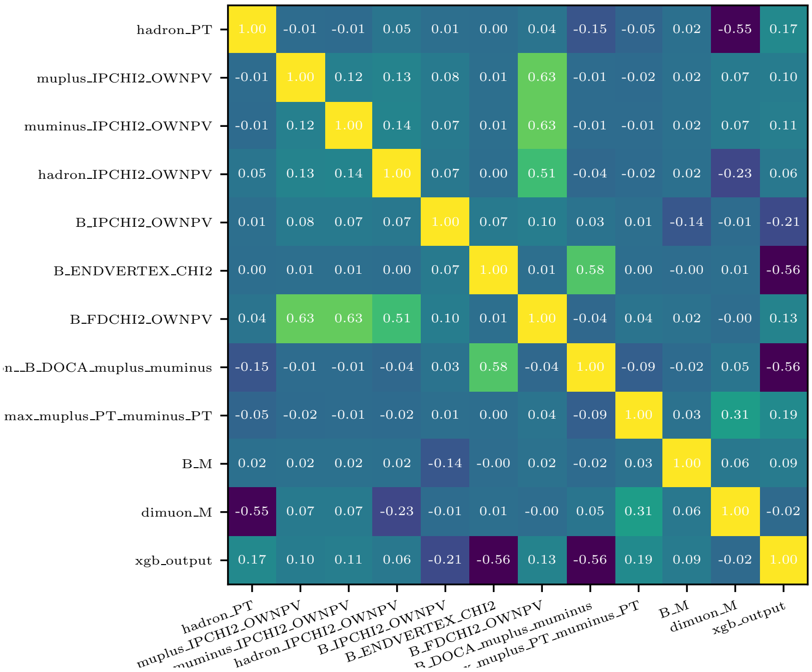
E.1 MVA Input Choice

We performed a study in which we trained a BDT to suppress combinatorial background using all variables listed in Tab. E.1. For this study we used only 2016 data and simulation with all selections as described in Sec. 3.2.2. The signal and background were sampled to contain each about $5 \cdot 10^4$ events, for a training and for a testing sample. We iteratively retrained the BDT by removing one of the inputs at each iteration.

We ranked inputs based on the difference in the area under the ROC curve (AUC) on the testing samples. The fact that this difference can become negative after removing inputs with little separation power is attributed to statistical fluctuations. The top eight variables on the list are selected as inputs for the final BDT configuration since no significant improvement is observed in the BDT performance by adding more variables.

Dependencies between inputs and between MVA output and fit observables

We check for possible dependencies between the inputs of used for the BDT and the observables used for fits. Figure E.1 shows the correlation matrix between BDT inputs and dimuon and B_c^+ candidate mass. Figures E.2 and E.3 show the BDT efficiencies for signal and background for different ranges of dimuon and B_c^+-candidate mass. We observe no significant dependency for the signal over the whole observable ranges. For the background, we observe a small tendency for the B_c^+-candidate mass, but this is expected as the combinatorial background receives various contributions at different ranges.



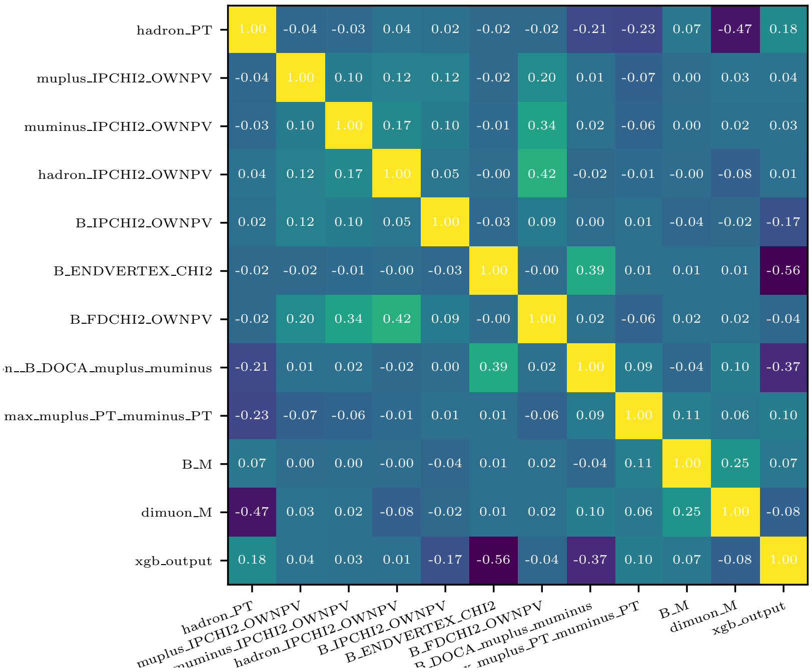
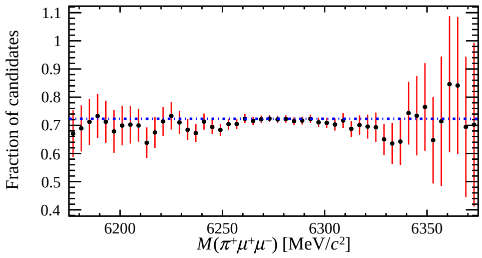
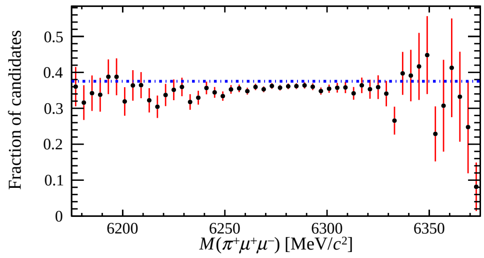


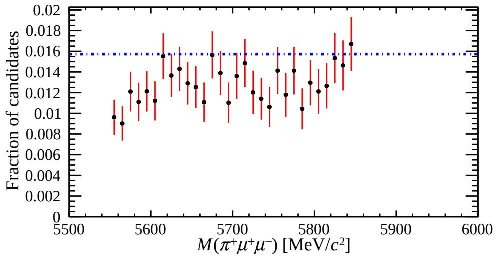
Figure E.1: Correlation plots for (top) signal and (bottom) background



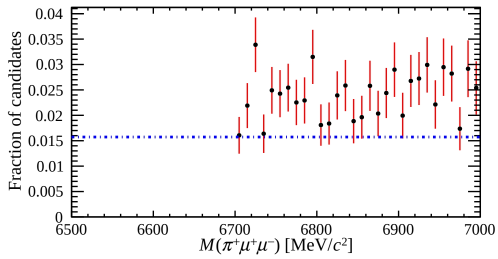
a) Signal (BDT > 0.91)



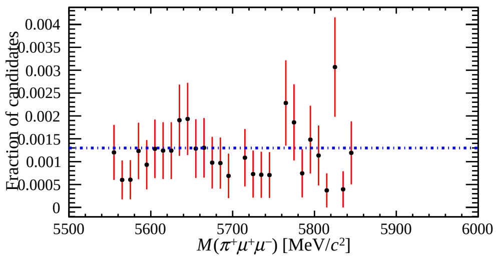
b) Signal (BDT > 0.989)



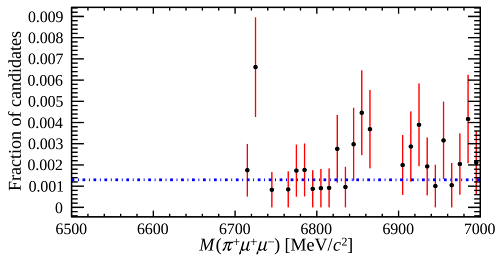
c) Sideband Low (BDT > 0.91)



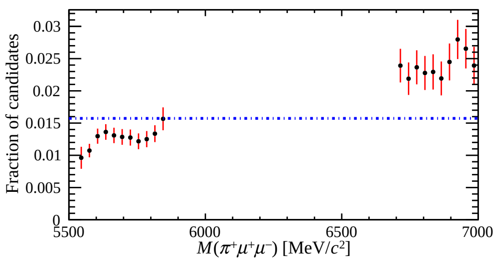
d) Sideband High (BDT > 0.91)



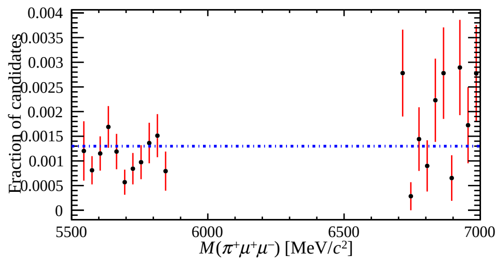
e) Sideband low (BDT > 0.989)



f) Sideband high (BDT > 0.989)



g) Sideband Full (BDT > 0.91)

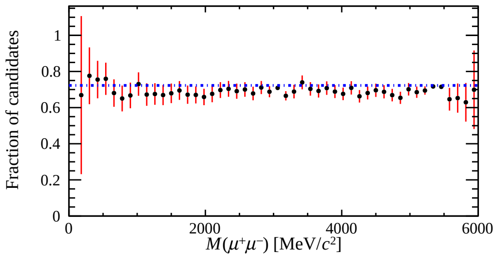


h) Sideband Full (BDT > 0.989)

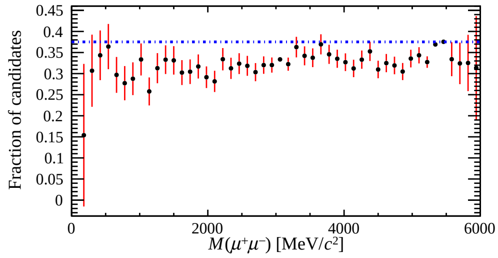
Figure E.2: BDT efficiency for various ranges of reconstructed B_c^+ mass for signal simulation and sideband data

Table E.1: BDT performance evaluated by removing one input variable at the time. The value of the area under the ROC curve (AUC) is shown for the training and for the testing samples. The difference ΔAUC corresponds to the difference between AUC for the testing sample using all variables in the list and using all variables except the one in the respective row. The score corresponds to ΔAUC devided by its maximal value. All values are given in percentage.

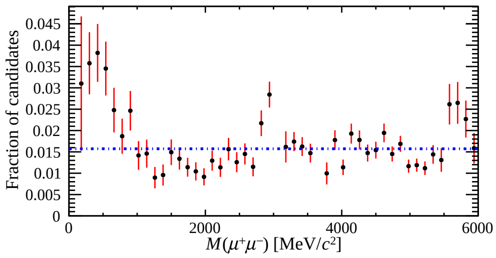
Removed variable	AUC training	AUC testing	ΔAUC	Score
B_ENDVERTEX_CHI2	99.2922	98.3233	0.3872	100.00
B_IPCHI2_OWNPV	99.5312	98.5367	0.1738	44.89
muminus_IPCHI2_OWNPV	99.5015	98.5711	0.1394	36.00
muplus_IPCHI2_OWNPV	99.5606	98.5753	0.1352	34.92
B_FDCHI2_OWNPV	99.5727	98.5917	0.1188	30.68
hadron_PT	99.5699	98.6571	0.0534	13.80
max(muplus_PT,muminus_PT)	99.5872	98.6946	0.0159	4.11
max_DOCA	99.6040	98.7032	0.0073	1.88
B_PT	99.6042	98.7048	0.0057	1.48
hadron_CosTheta	99.6117	98.7064	0.0041	1.05
B_VTXISOBDTHARDFIRSTVALUE	99.5998	98.7155	−0.0050	−1.29
B_CONEP	99.5707	98.7168	−0.0063	−1.63
B_DOCACHI2_muminus_hadron	99.6304	98.7175	−0.0070	−1.81
B_DOCACHI2_muplus_hadron	99.5820	98.7194	−0.0089	−2.30
B_VTXISONUMVTX	99.6223	98.7213	−0.0108	−2.78
B_CONEDELTAPHI	99.6053	98.7233	−0.0128	−3.31
B_CONEMULT	99.6201	98.7240	−0.0135	−3.48
B_VTXISODCHI2TWOTRACK	99.6794	98.7308	−0.0203	−5.24
B_VTXISOBDTHARDSECONDVALUE	99.6116	98.7336	−0.0231	−5.97
muplus_CosTheta	99.6472	98.7337	−0.0232	−6.00
B_CONEPASYM	99.6266	98.7342	−0.0237	−6.12
B_CONEPT	99.6431	98.7344	−0.0239	−6.18
B_DOCACHI2_muplus_muminus	99.6377	98.7346	−0.0241	−6.22
B_CONEPTASYM	99.6260	98.7360	−0.0255	−6.59
B_VTXISODCHI2ONETRACK	99.6012	98.7390	−0.0285	−7.36
B_MU_SLL_ISO_1	99.6431	98.7402	−0.0297	−7.68
muminus_CosTheta	99.6466	98.7410	−0.0305	−7.86
B_VTXISOBDTHARDTHIRDVALUE	99.6284	98.7450	−0.0345	−8.92
B_MU_SLL_ISO_2	99.6421	98.7507	−0.0402	−10.39
B_DIRA_OWNPV	99.6461	98.7509	−0.0404	−10.43
B_CONEDELTAETA	99.6551	98.7522	−0.0417	−10.78



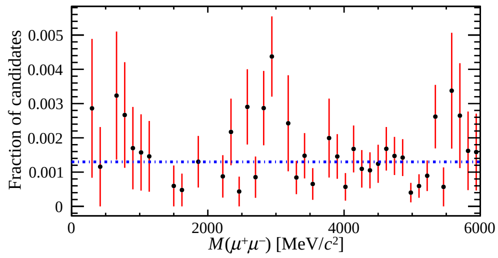
a) Signal (BDT > 0.91)



b) Signal dimuon (BDT > 0.989)



c) Background (BDT > 0.91)



d) Background dimuon (BDT > 0.989)

Figure E.3: BDT efficiency for various ranges of reconstructed dimuon mass for signal simulation and sideband data

Appendix F

Studies with $B_c^+ \to \rho^+\mu^+\mu^-$ Background

We perform a fit to sideband data as in Sec. 4.3.1 including an additional $B_c^+ \to \rho^+\mu^+\mu^-$ component to the fit model. The B_c^+ mass distribution for this component is modelled using an exponential function whose slope is determined from truth-matched simulation taking into account the effect of the PID selection using PIDCalib histograms.

Table F.1 shows the slope of the exponential function for the three q^2 bins and combining the three bins together. Figures F.1 and F.2 show the B_c^+ mass distributions in simulation with fit projections overlaid.

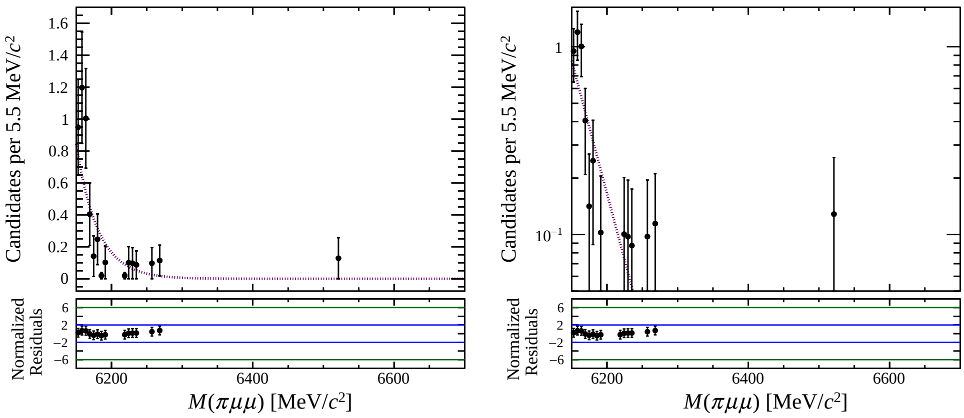
Table F.2 shows the results from fits to sideband data. Figure F.2 shows the B_c^+ mass distributions in sideband data with fit projections overlaid. The results confirm that the possible contributions from $B_c^+ \to \rho^+\mu^+\mu^-$ decays are negligible.

Table F.1: Background PDF parameters (defined as in Table 4.9) obtained by fitting $B_c^+ \to \rho^+\mu^+\mu^-$ simulation.

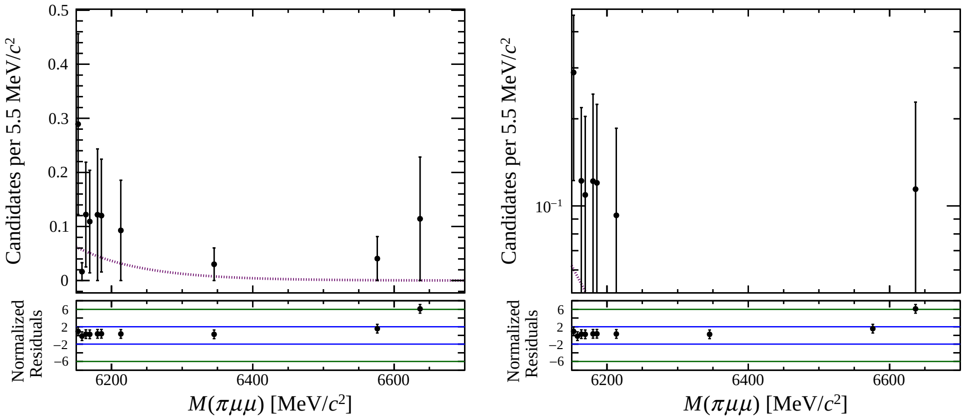
Fit parameters	$0.1 < q^2 < 1.1$	$1.1 < q^2 < 8$	$11 < q^2 < 12.5$	$15 < q^2 < 35$	All
Part. bkg. exp. [MeV^{-1}]	-0.02 $\pm$ 0.02	-0.03 ± 0.02	-0.01 ± 0.01	-0.06 ± 0.03	-0.03 ± 0.01

Table F.2: Results for the $B_c^+ \to \rho^+\mu^+\mu^-$ background yield and combinatorial background exponent and yield obtained from a fit to sideband data.

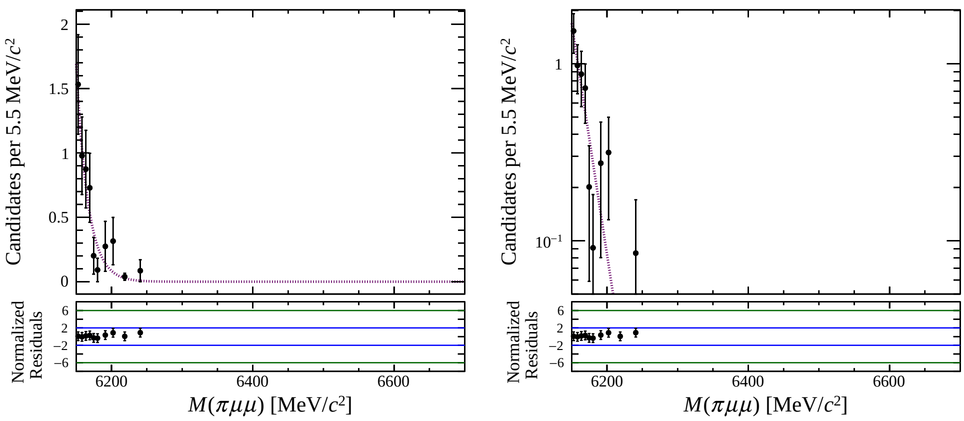
Fit parameters	$0.1 < q^2 < 1.1$	$1.1 < q^2 < 8$	$11 < q^2 < 12.5$	$15 < q^2 < 35$	All
Comb. bkg. exp. [10^{-3} MeV^{-1}]	1.9 ± 2.2	-0.9 ± 1.6	-4.8 ± 4.9	-0.2 ± 0.6	0.03 ± 0.6
Comb. bkg. yield	21 ± 6	33 ± 8	26 ± 31	222 ± 20	267 ± 23
$\rho^+\mu^+\mu^-$ yield	2 ± 3	3 ± 3	-11 ± 29	3 ± 5	8 ± 8



a) $1.1 < q^2 < 8\,\mathrm{GeV}^2/c^4$

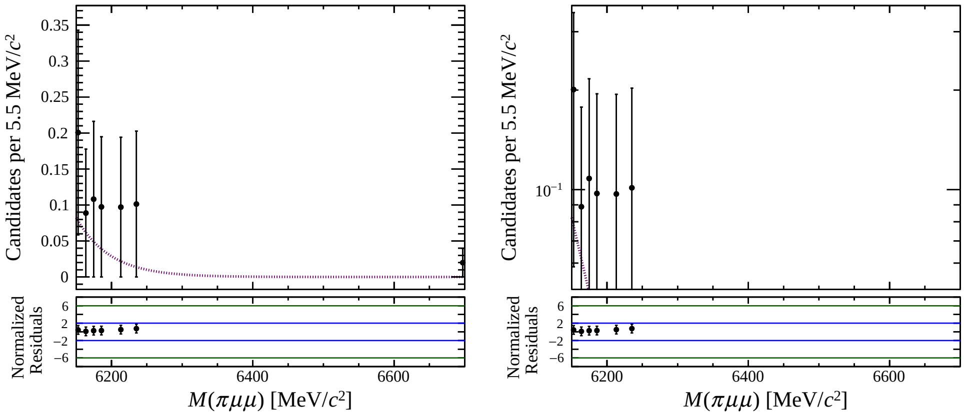


b) $11 < q^2 < 12.5\,\mathrm{GeV}^2/c^4$

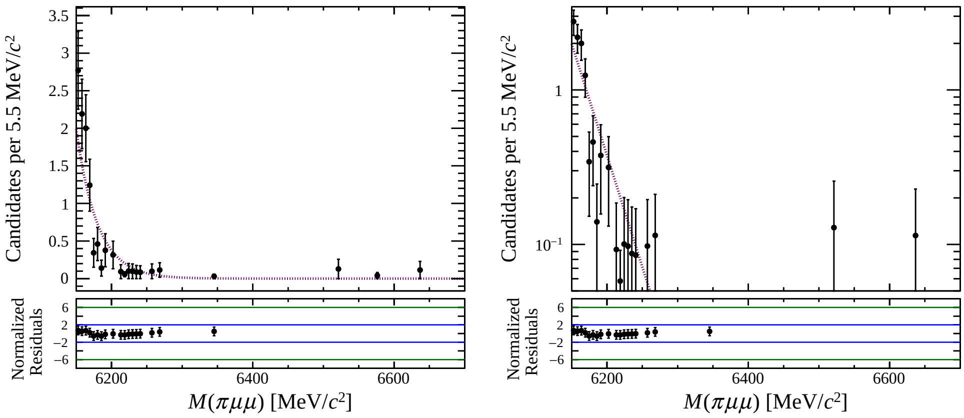


c) $15 < q^2 < 35\,\mathrm{GeV}^2/c^4$

Figure F.1: Distributions of B_c^+-candidate mass for non-resonant $B_c^+ \to \rho^+\mu^+\mu^-$ simulation with fit projections overlaid. The distributions are shown for the three q^2 bins separately, in (left) linear and (right) log scales.

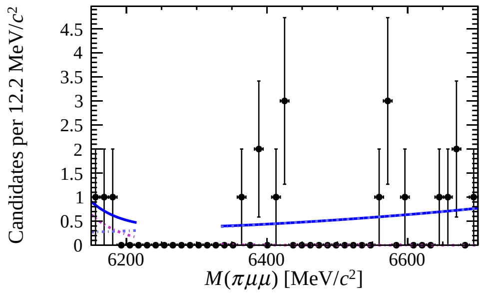


a) $0.1 < q^2 < 1.1\,\mathrm{GeV}^2/c^4$

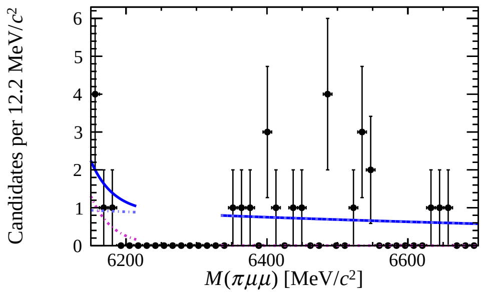


All q^2 bins together

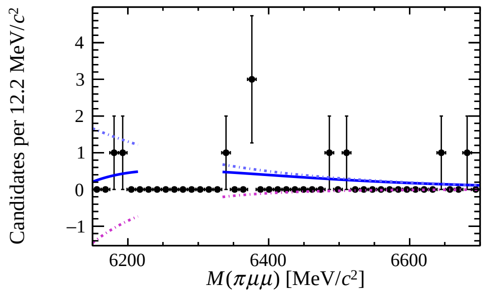
Figure F.2: Distributions of B_c^+-candidate mass for non-resonant $B_c^+ \to \rho^+\mu^+\mu^-$ simulation with fit projections overlaid. The distribution is shown for all three together and the very low q^2 bins in (left) linear and (right) log scales.



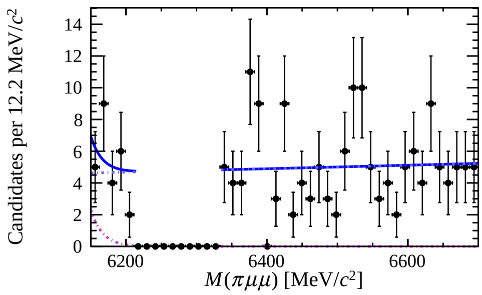
$0.1 < q^2 < 1.1\,\mathrm{GeV}^2/c^4$



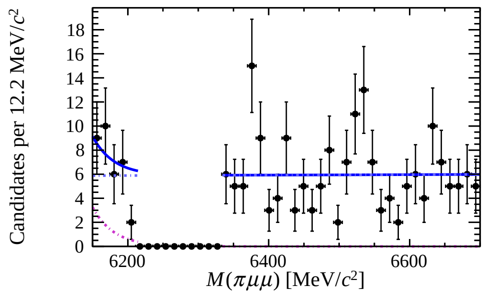
$1.1 < q^2 < 8\,\mathrm{GeV}^2/c^4$



$11 < q^2 < 12.5\,\mathrm{GeV}^2/c^4$



$15 < q^2 < 35\,\mathrm{GeV}^2/c^4$



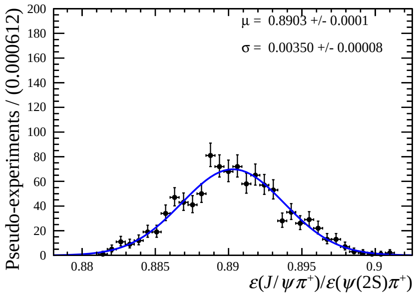
All q^2 bins

Figure F.3: Distributions of B_c^+-candidate mass in sideband data for the three q^2 bins and for all q^2 bins combined. Projection of maximum-likelihood fits are overlaid.

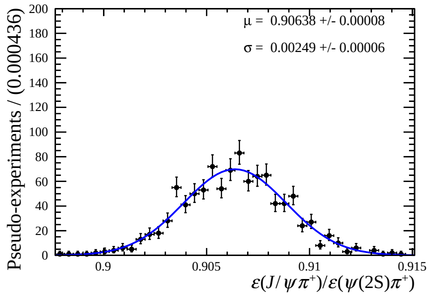
Appendix G

Systematic Uncertainty on Efficiency Ratio

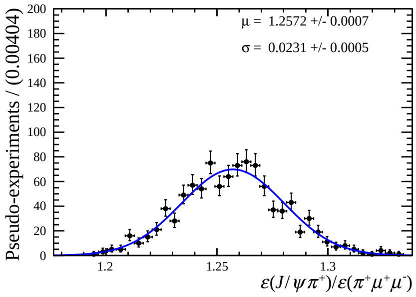
In this appendix we show the distributions from which we estimate the systematic uncertainties on the efficiency ratio as described in Sec. 5. Figures G.1–G.11 show the efficiency ratio distributions, with the projection of a fit of a Gaussian function overlaid, for the two selection working points (see 3.5), in case of the ratio between the resonant modes, and for the four q^2 bins, in case of the ratio between the non-resonant mode and the $B_c^+ \to J/\psi\,\pi^+$ mode. The distributions are shown for various systematic variations. Figures G.12–G.13 show the acceptance, selection and total efficiency (see Sec 4.1) as a function of q^2 for both unpolarised and longitudinally polarised models for the dimuon system.



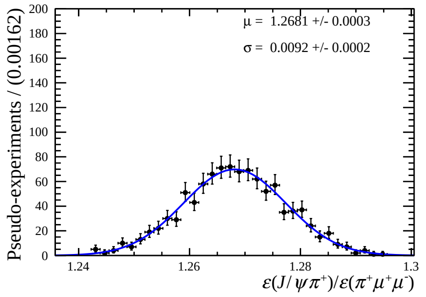
a) $\pi^+\mu^+\mu^-$ WP



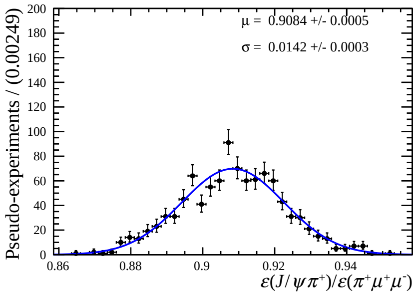
b) $\psi(2S)\pi^+$ WP



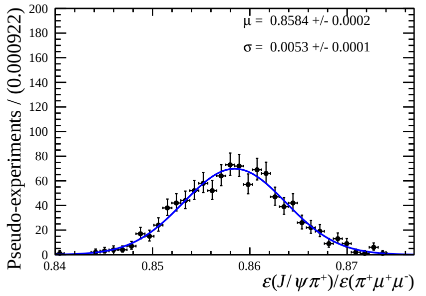
c) $\pi^+\mu^+\mu^-$ Very Low



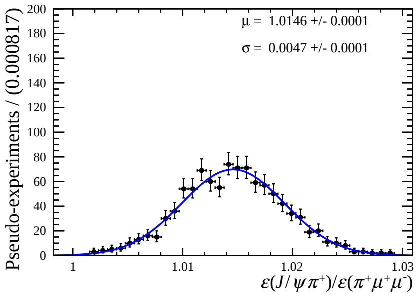
d) $\pi^+\mu^+\mu^-$ Low



e) $\pi^+\mu^+\mu^-$ Central

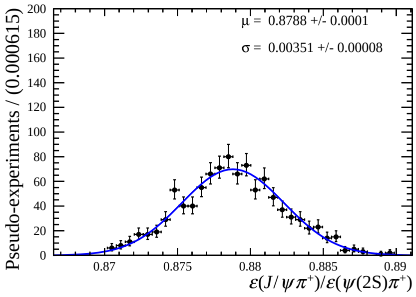


f) $\pi^+\mu^+\mu^-$ High

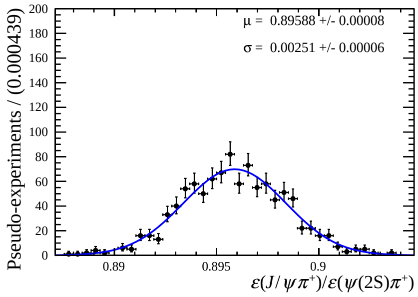


g) $\pi^+\mu^+\mu^-$ Bins Combined

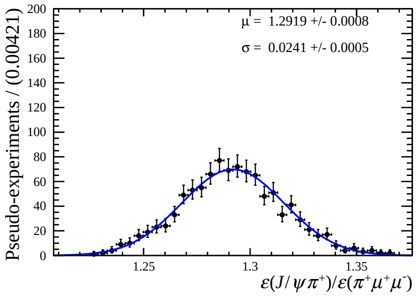
Figure G.1: Distributions of efficiency ratio obtained from pseudo-data samples including `BKGCAT` 60.



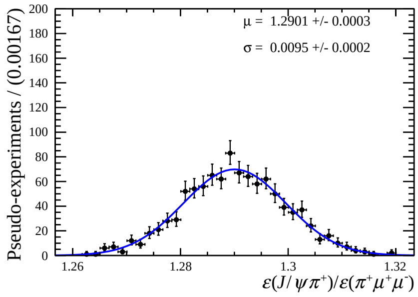
a) $\psi(2S)\pi^+$ WP



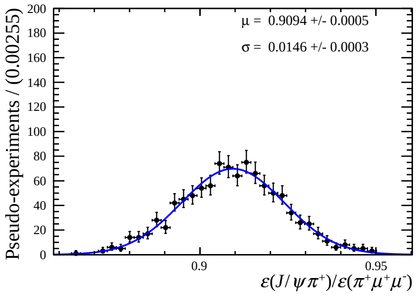
b) $\psi(2S)\pi^+$ WP



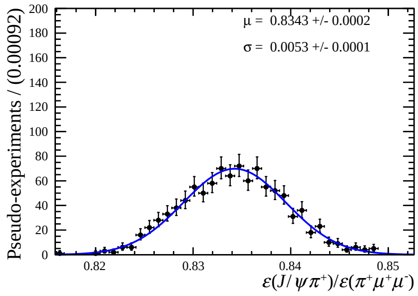
c) $\pi^+\mu^+\mu^-$ Very Low



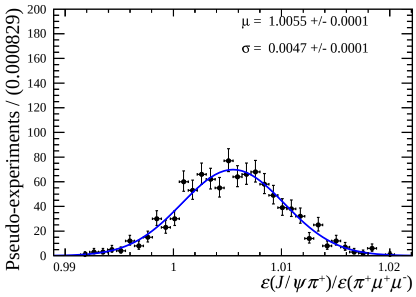
d) $\pi^+\mu^+\mu^-$ Low



e) $\pi^+\mu^+\mu^-$ Central

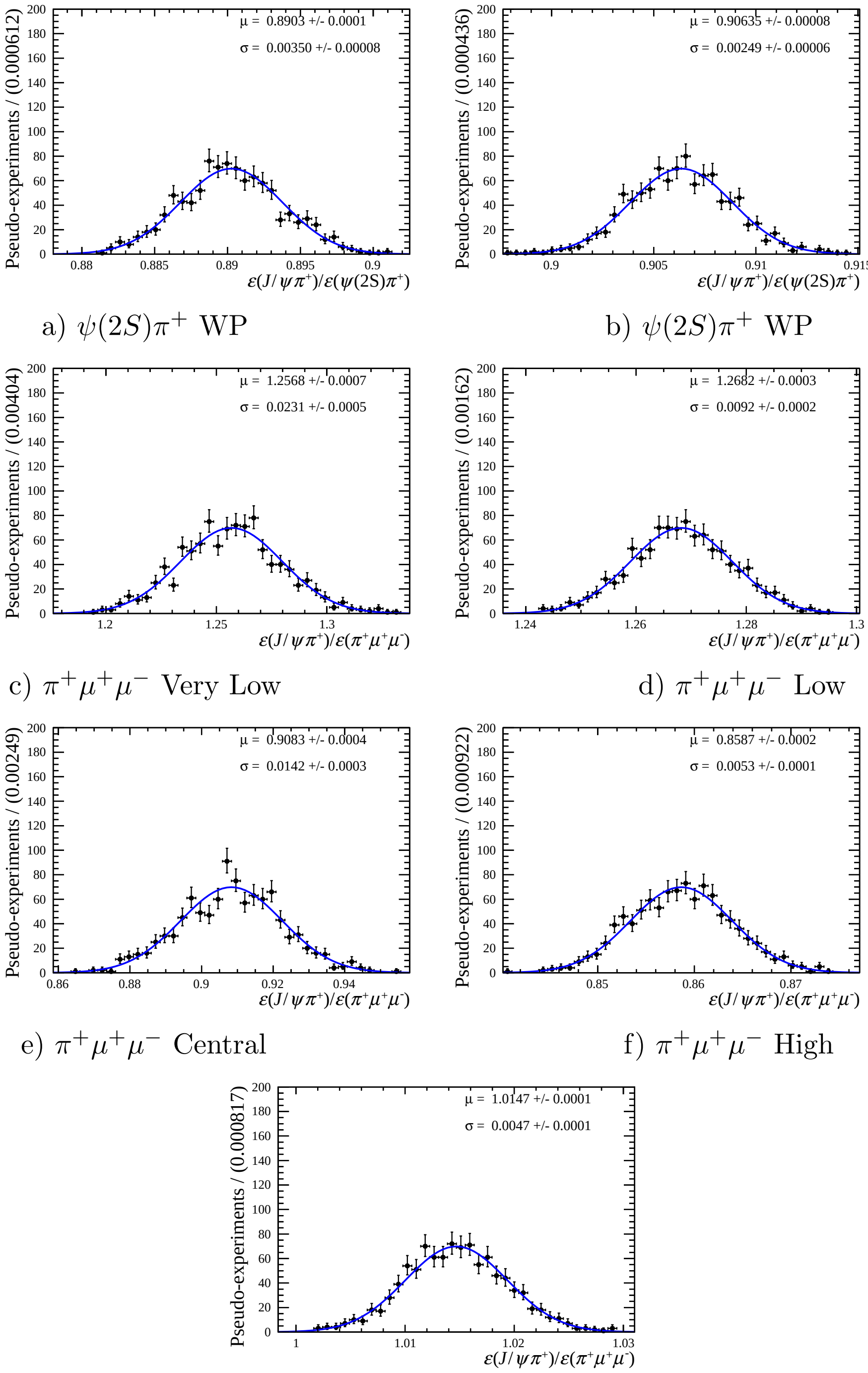


f) $\pi^+\mu^+\mu^-$ High



g) $\pi^+\mu^+\mu^-$ Bins Combined

Figure G.2: Distributions of efficiency ratio obtained from pseudo-data samples using alternative kinematic weights.



a) $\psi(2S)\pi^+$ WP

b) $\psi(2S)\pi^+$ WP

c) $\pi^+\mu^+\mu^-$ Very Low

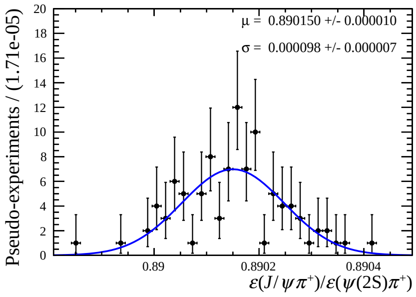
d) $\pi^+\mu^+\mu^-$ Low

e) $\pi^+\mu^+\mu^-$ Central

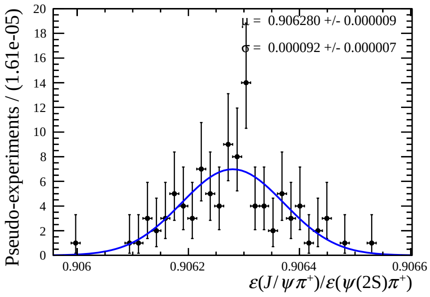
f) $\pi^+\mu^+\mu^-$ High

g) $\pi^+\mu^+\mu^-$ Bins Combined

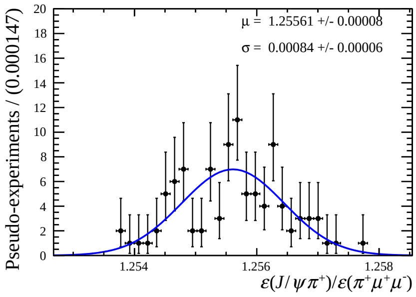
Figure G.3: Distributions of efficiency ratio obtained from pseudo-data samples using p_{T} histograms with alternative binning.



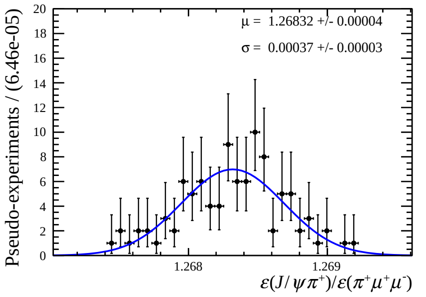
a) $\psi(2S)\pi^+$ WP



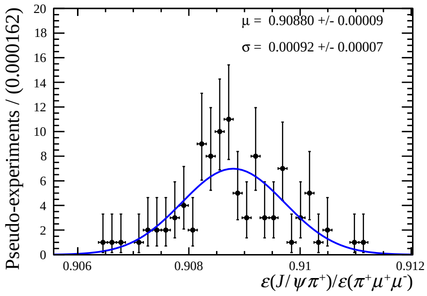
b) $\psi(2S)\pi^+$ WP



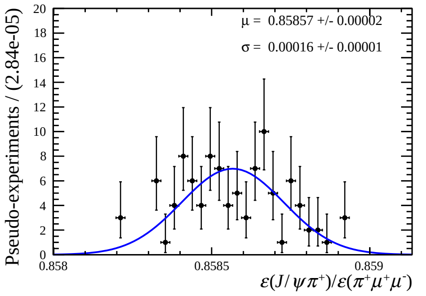
c) $\pi^+\mu^+\mu^-$ Very Low



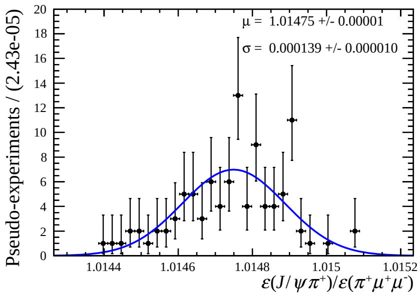
d) $\pi^+\mu^+\mu^-$ Low



e) $\pi^+\mu^+\mu^-$ Central

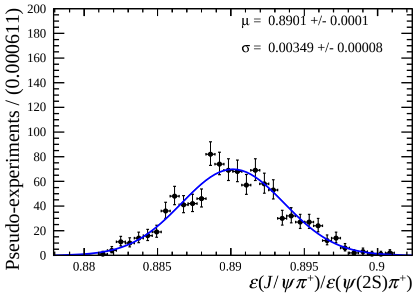


f) $\pi^+\mu^+\mu^-$ High

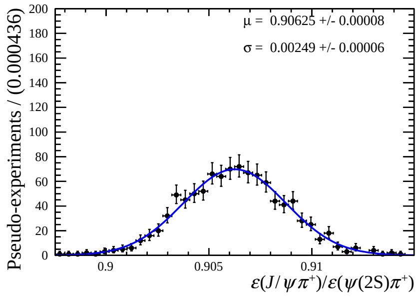


g) $\pi^+\mu^+\mu^-$ Bins Combined

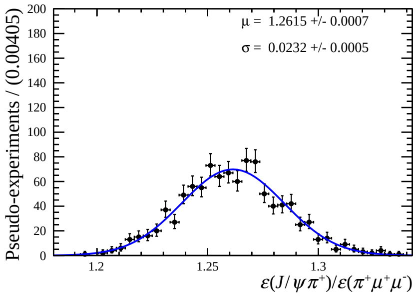
Figure G.4: Distributions of efficiency ratio obtained from pseudo-data samples varying the content of the p_{T} histograms within their statistical uncertainties.



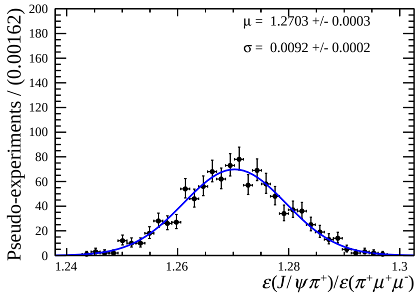
a) $\psi(2S)\pi^+$ WP



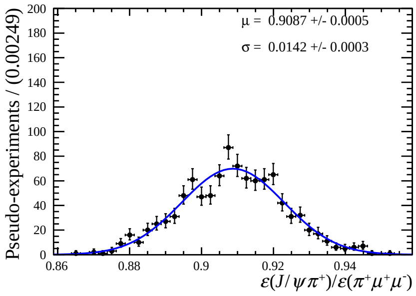
b) $\psi(2S)\pi^+$ WP



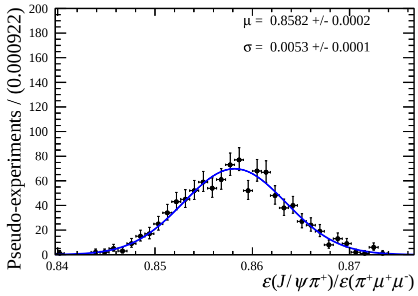
c) $\pi^+\mu^+\mu^-$ Very Low



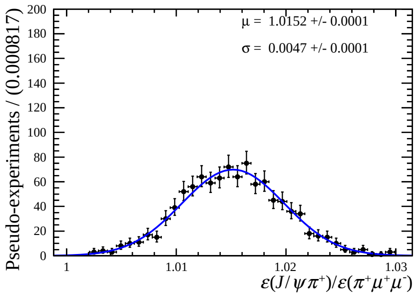
d) $\pi^+\mu^+\mu^-$ Low



e) $\pi^+\mu^+\mu^-$ Central

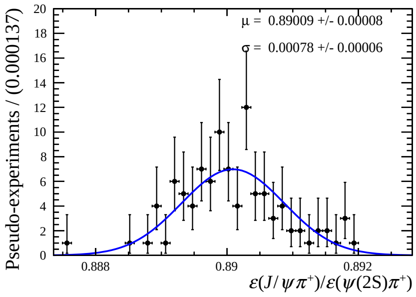


f) $\pi^+\mu^+\mu^-$ High

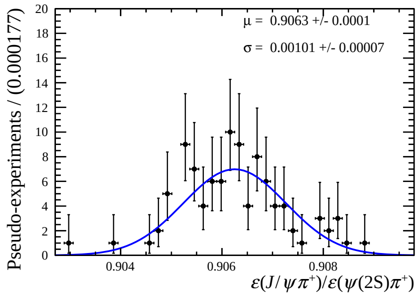


g) $\pi^+\mu^+\mu^-$ Bins Combined

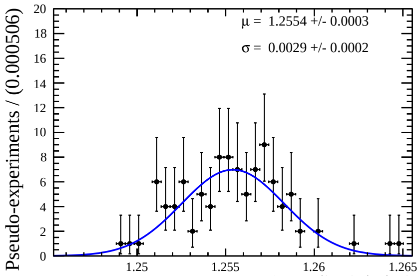
Figure G.5: Distributions of efficiency ratio obtained from pseudo-data samples using trigger correction histograms with alternative binning.



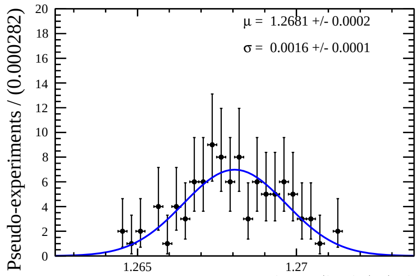
a) $\psi(2S)\pi^+$ WP



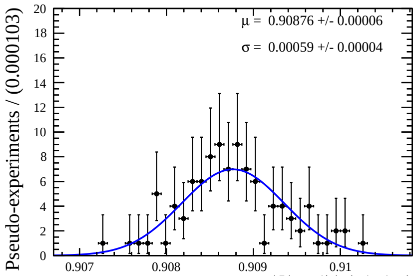
b) $\psi(2S)\pi^+$ WP



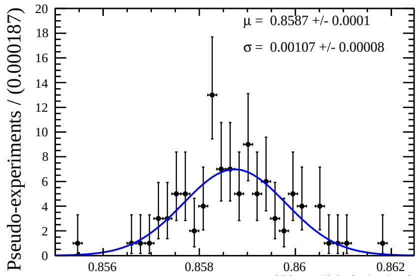
c) $\pi^+\mu^+\mu^-$ Very Low



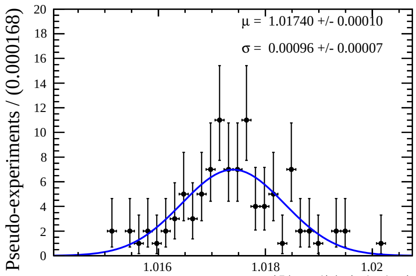
d) $\pi^+\mu^+\mu^-$ Low



e) $\pi^+\mu^+\mu^-$ Central

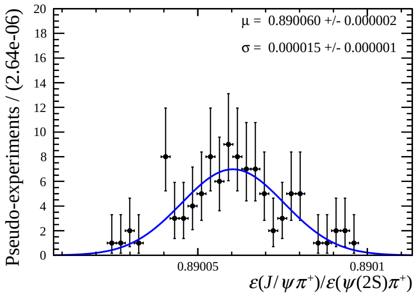


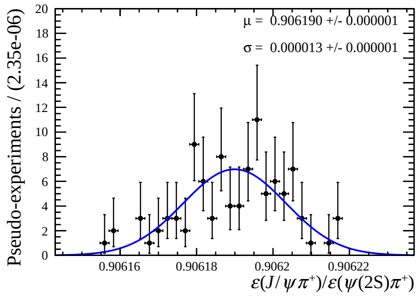
f) $\pi^+\mu^+\mu^-$ High



g) $\pi^+\mu^+\mu^-$ Bins Combined

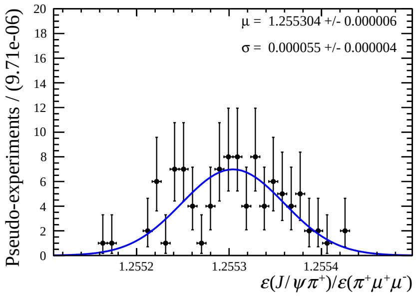
Figure G.6: Distributions of efficiency ratio obtained from pseudo-data samples varying the content of the trigger correction histograms within their statistical uncertainties.

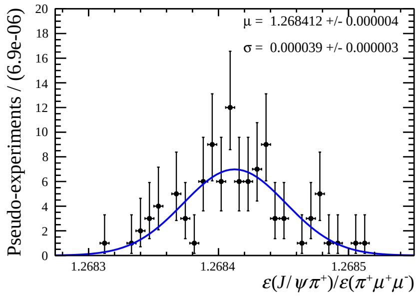




a) $\psi(2S)\pi^+$ WP

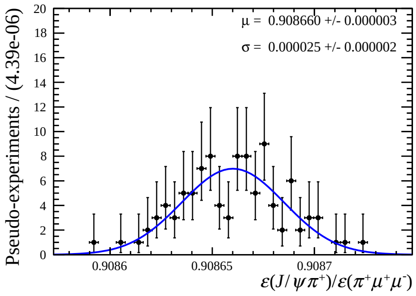
b) $\psi(2S)\pi^+$ WP

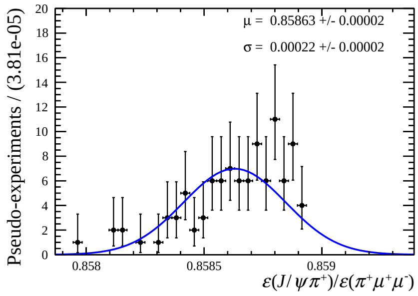




c) $\pi^+\mu^+\mu^-$ Very Low

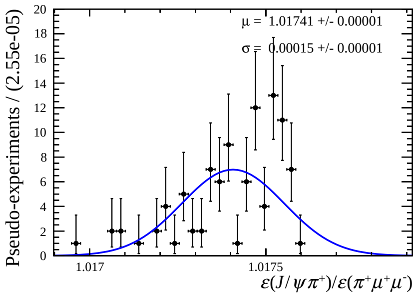
d) $\pi^+\mu^+\mu^-$ Low





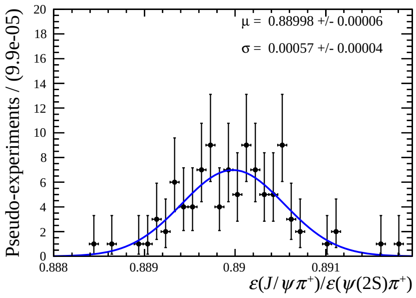
e) $\pi^+\mu^+\mu^-$ Central

f) $\pi^+\mu^+\mu^-$ High

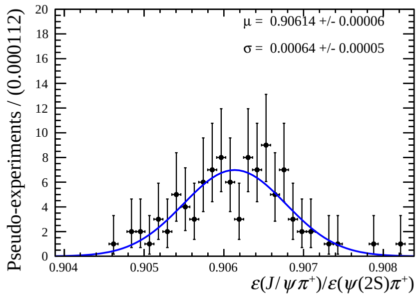


g) $\pi^+\mu^+\mu^-$ Bins Combined

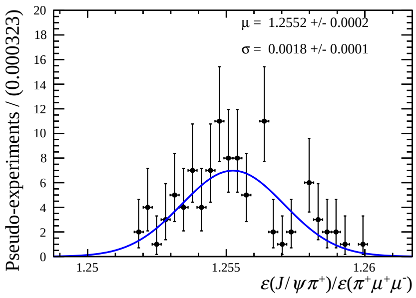
Figure G.7: Distributions of efficiency ratio obtained from pseudo-data samples varying the content of the PID efficiency histograms within their statistical uncertainties.



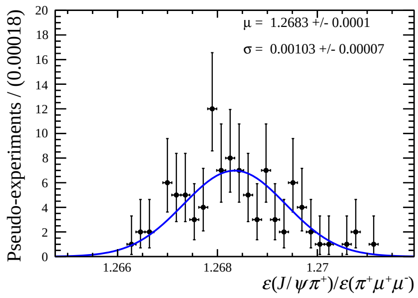
a) $\psi(2S)\pi^+$ WP



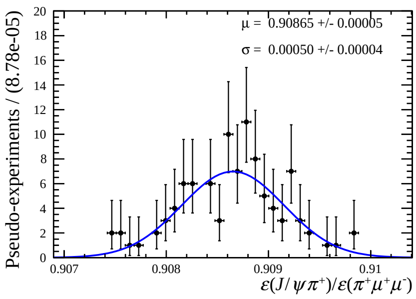
b) $\psi(2S)\pi^+$ WP



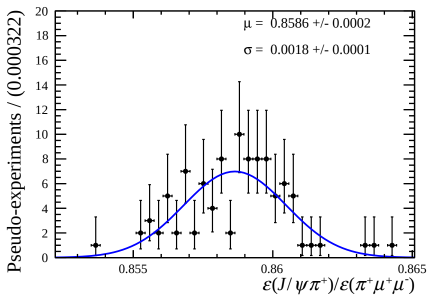
c) $\pi^+\mu^+\mu^-$ Very Low



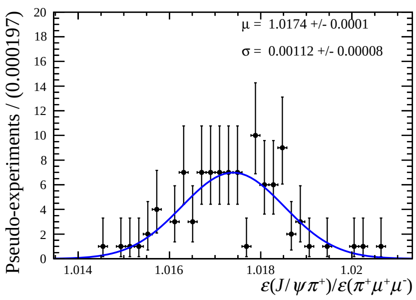
d) $\pi^+\mu^+\mu^-$ Low



e) $\pi^+\mu^+\mu^-$ Central

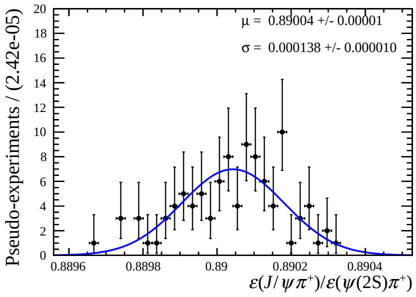


f) $\pi^+\mu^+\mu^-$ High

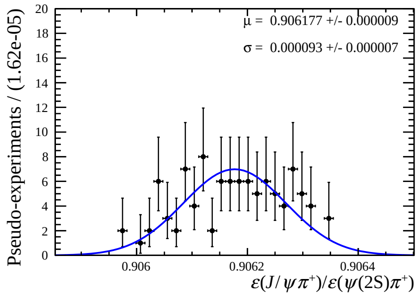


g) $\pi^+\mu^+\mu^-$ Bins Combined

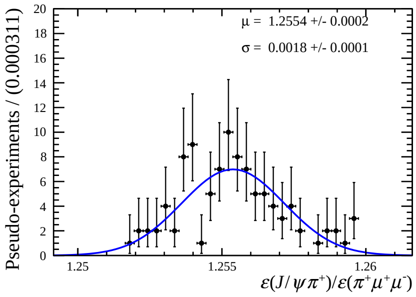
Figure G.8: Distributions of efficiency ratio obtained from pseudo-data samples varying the content of the muon ID correction histograms within their statistical uncertainties.



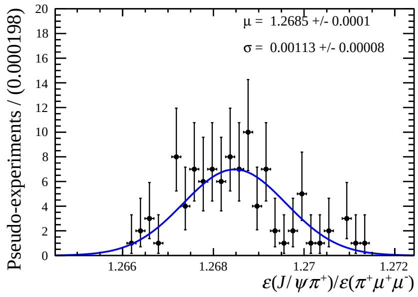
a) $\psi(2S)\pi^+$ WP



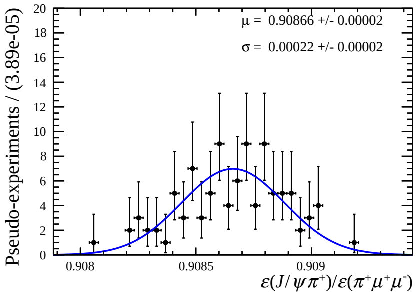
b) $\psi(2S)\pi^+$ WP



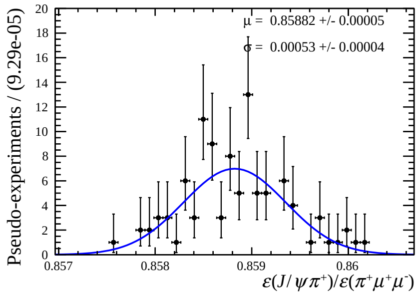
c) $\pi^+\mu^+\mu^-$ Very Low



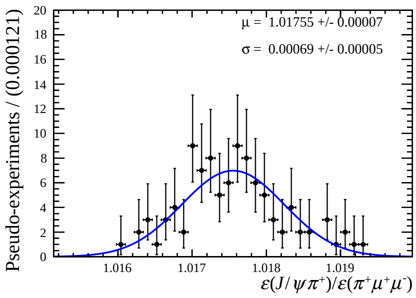
d) $\pi^+\mu^+\mu^-$ Low



e) $\pi^+\mu^+\mu^-$ Central

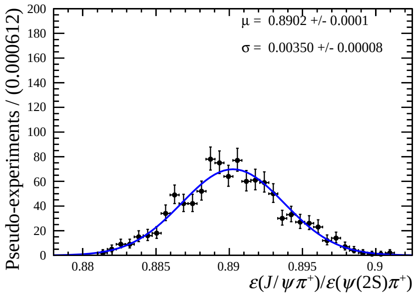


f) $\pi^+\mu^+\mu^-$ High

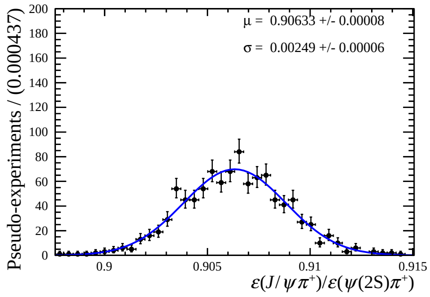


g) $\pi^+\mu^+\mu^-$ Bins Combined

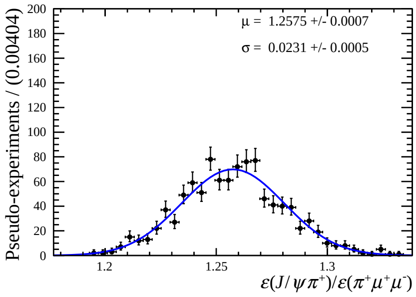
Figure G.9: Distributions of efficiency ratio obtained from pseudo-data samples varying the content of the tracking efficiency correction histograms within their statistical uncertainties.



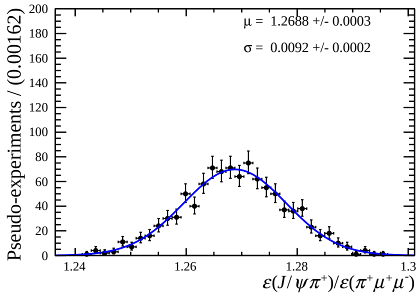
a) $\psi(2S)\pi^+$ WP



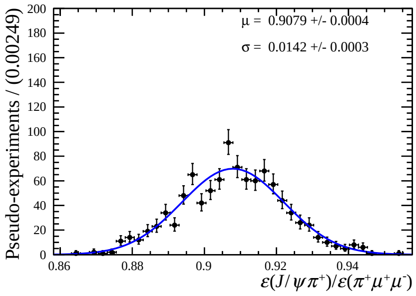
b) $\psi(2S)\pi^+$ WP



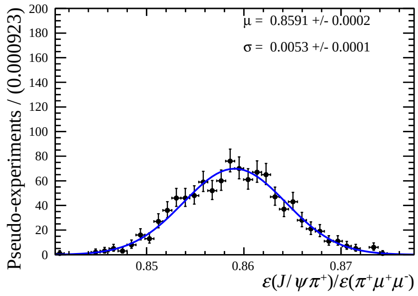
c) $\pi^+\mu^+\mu^-$ Very Low



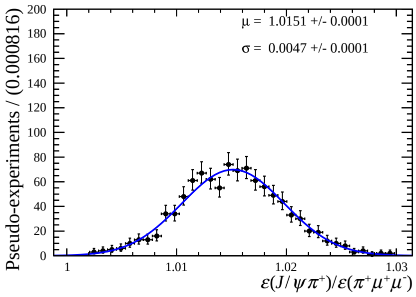
d) $\pi^+\mu^+\mu^-$ Low



e) $\pi^+\mu^+\mu^-$ Central

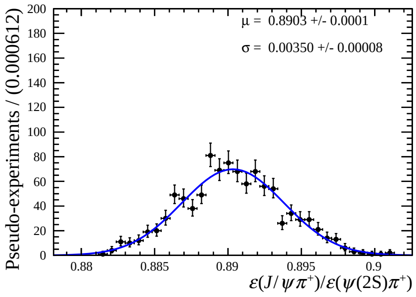


f) $\pi^+\mu^+\mu^-$ High

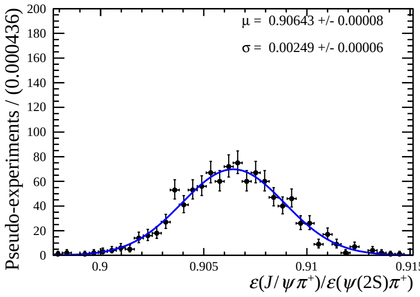


g) $\pi^+\mu^+\mu^-$ Bins Combined

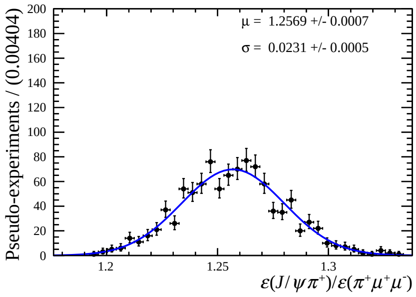
Figure G.10: Distributions of efficiency ratio obtained from pseudo-data samples lowering the B_c^+-lifetime by 1σ.



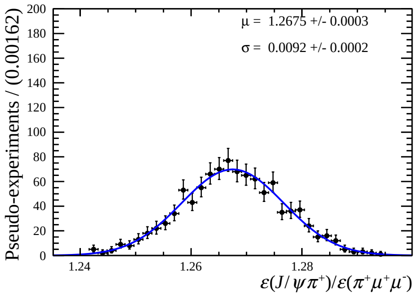
a) $\psi(2S)\pi^+$ WP



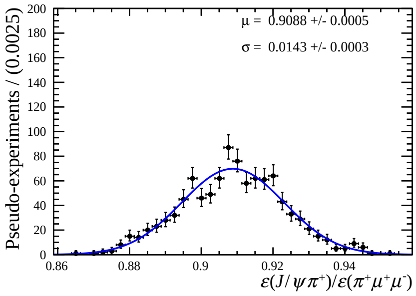
b) $\psi(2S)\pi^+$ WP



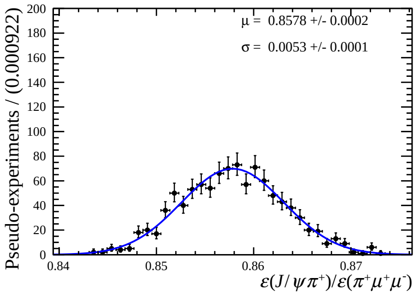
c) $\pi^+\mu^+\mu^-$ Very Low



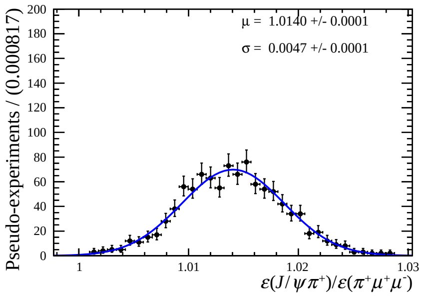
d) $\pi^+\mu^+\mu^-$ Low



e) $\pi^+\mu^+\mu^-$ Central

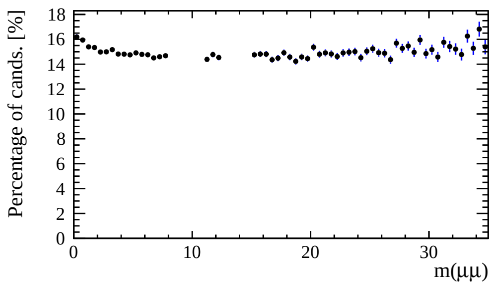


f) $\pi^+\mu^+\mu^-$ High

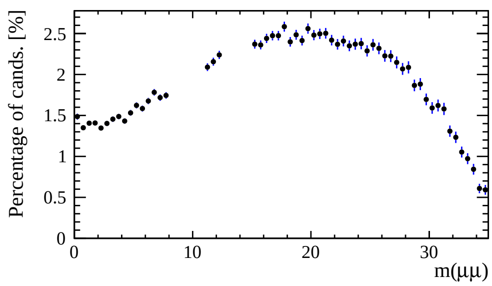


g) $\pi^+\mu^+\mu^-$ Bins Combined

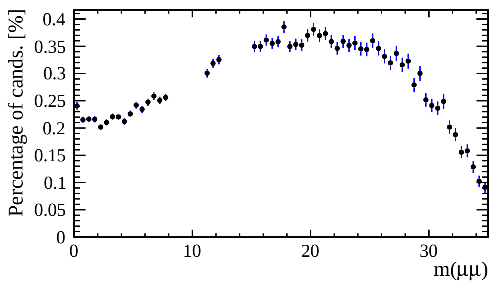
Figure G.11: Distributions of efficiency ratio obtained from pseudo-data samples increasing the B_c^+-lifetime by 1σ.



a) Acceptance efficiency

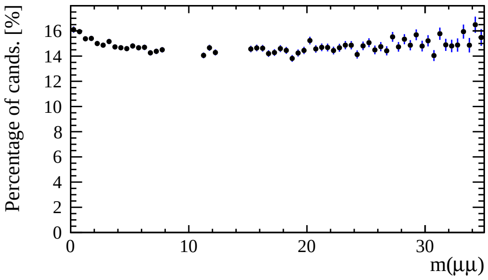


b) Selection efficiency

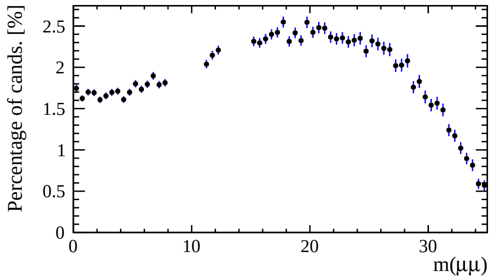


c) Product of acceptance and selection efficiencies

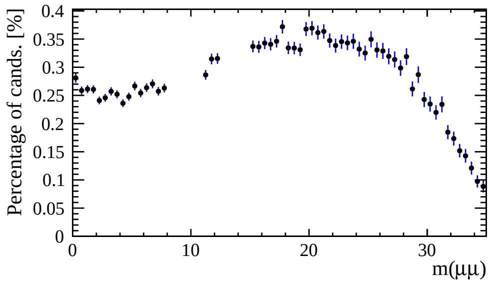
Figure G.12: Efficiency as a function of q^2 for $B_c^+ \to \pi^+\mu^+\mu^-$ decays, evaluated with the unpolarised model for the dimuon system.



a) Acceptance efficiency



b) Selection efficiency



c) Product of acceptance and selection efficiencies

Figure G.13: Efficiency as a function of q^2 for $B_c^+ \to \pi^+\mu^+\mu^-$ decays, evaluated with the longitudinally polarised model for the dimuon system.

Appendix H

Fit Estimator Properties for Separate q^2 Bins

We perform a study of fit estimator properties for the non-resonant $B_c^+ \to \pi^+\mu^+\mu^-$ fit as in Sec. 4.3.1 for each of the three q^2 bins separately (see Sec 3.2.5). We perform pseudo-experiments for various expectations of signal decays, $N_{\rm sig}^{\rm true} = 0, 5, 10, 20, 30, 40$ and 50. The simulated pseudo-data samples are generated without (background only) and with signal decays. The number of signal events in each pseudo-data sample is drawn from a Poisson distribution with expectation equal to $N_{\rm sig}^{\rm true}$. Figures H.1–H.9 show pull distributions for all fit parameters.

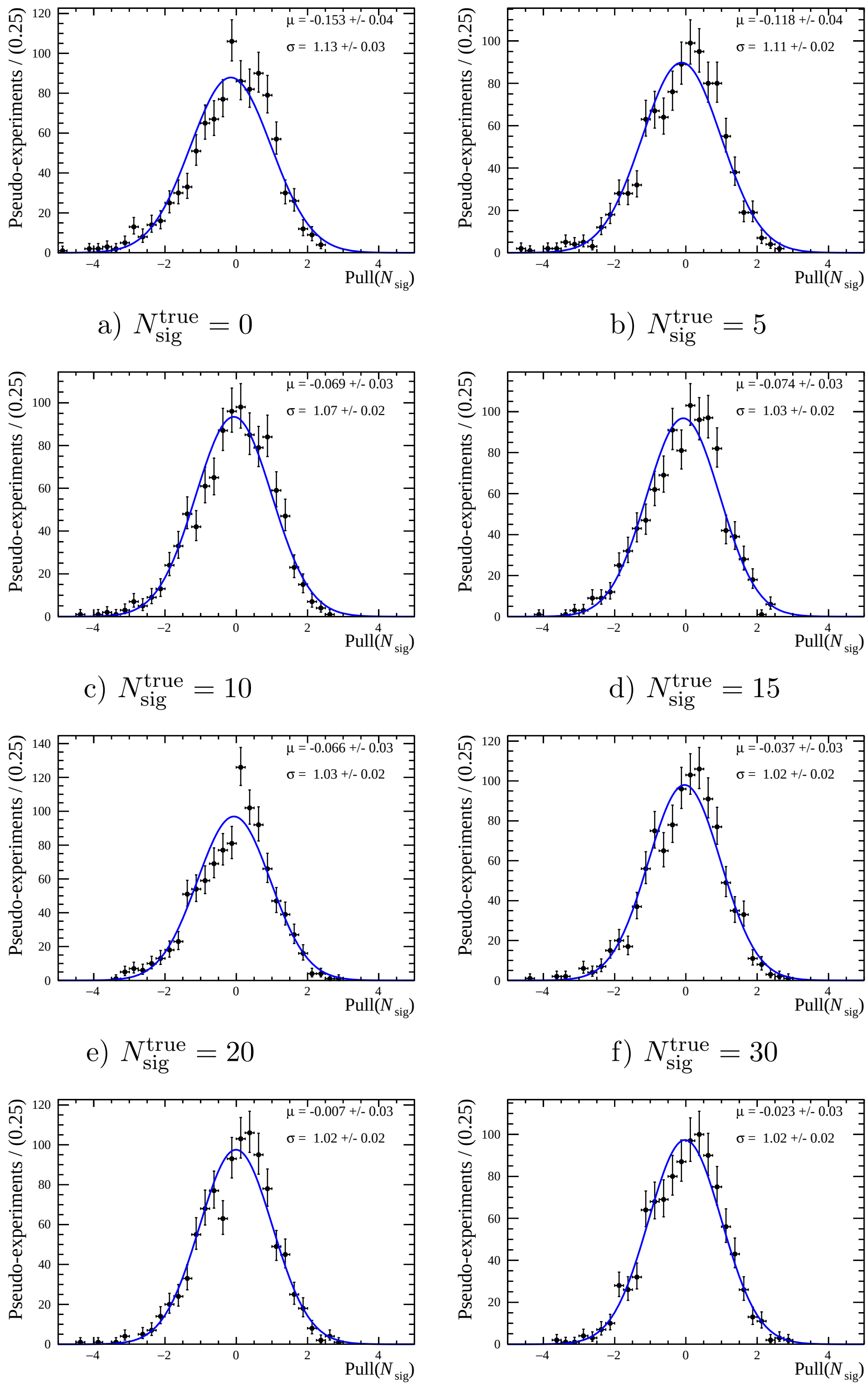
We generally observe unbiased distributions and proper uncertainty estimation for all fit parameters except for the signal yield except for the low and the central q^2 bins when $N_{\rm sig}^{\rm true} = 0$ and the total number of events is small. We attribute this asymmetric tail in the pull distributions to the fact that there are very few events in the signal region and thus the extended maximum-likelihood fit tends to converge towards negative values [32]. This asymmetry is alleviated when we generate pseudo-data samples with non-zero numbers of signal events, and moreover does not preclude us from obtaining appropriate confidence level intervals on the signal branching fraction.

Upon unblinding, we will estimate the significance of the measurement only when the signal yield is larger than zero. For the sake of completeness, we calculate the significance also for pseudo-experiments for which the fit converges into a negative signal yield value. Figures H.10–H.12 show the p-value distributions obtained from the pseudo-experiments. For the low and the central q^2 bins, we observe that the fit can converge to a negative signal yield with a significance above 3σ when $N_{\rm sig}^{\rm true} = 0$. This is attributed to the same effect that that causes the asymmetric tail in the pull distributions in these cases, that is that the extended part of the likelihood causes the fit to converge towards negative values when the total number of events is small.

Table H.1 shows the fractions of pseudo-experiments for which the fit converged to a positive yield with a significance of or above 3σ.

Table H.1: Fraction of pseudo-experiments (%) in which the fit converges to a positive yield with a significance of or above 3σ.

$N_{\text{sig}}^{\text{true}}$	0	5	10	15	20	30	40	50
$f_{3\sigma}$ (low)	1.6	6.0	33.2	65.0	86.7	99.2	100.0	100.0
$f_{3\sigma}$ (central)	0.4	11.6	48.9	79.9	94.9	99.9	100.0	100.0
$f_{3\sigma}$ (high)	0.1	1.2	4.6	13.2	29.4	69.3	91.8	98.4
$f_{3\sigma}$ (all)	0.2	0.6	4.0	10.6	24.2	58.9	87.7	97.6



a) $N_{\text{sig}}^{\text{true}} = 0$

b) $N_{\text{sig}}^{\text{true}} = 5$

c) $N_{\text{sig}}^{\text{true}} = 10$

d) $N_{\text{sig}}^{\text{true}} = 15$

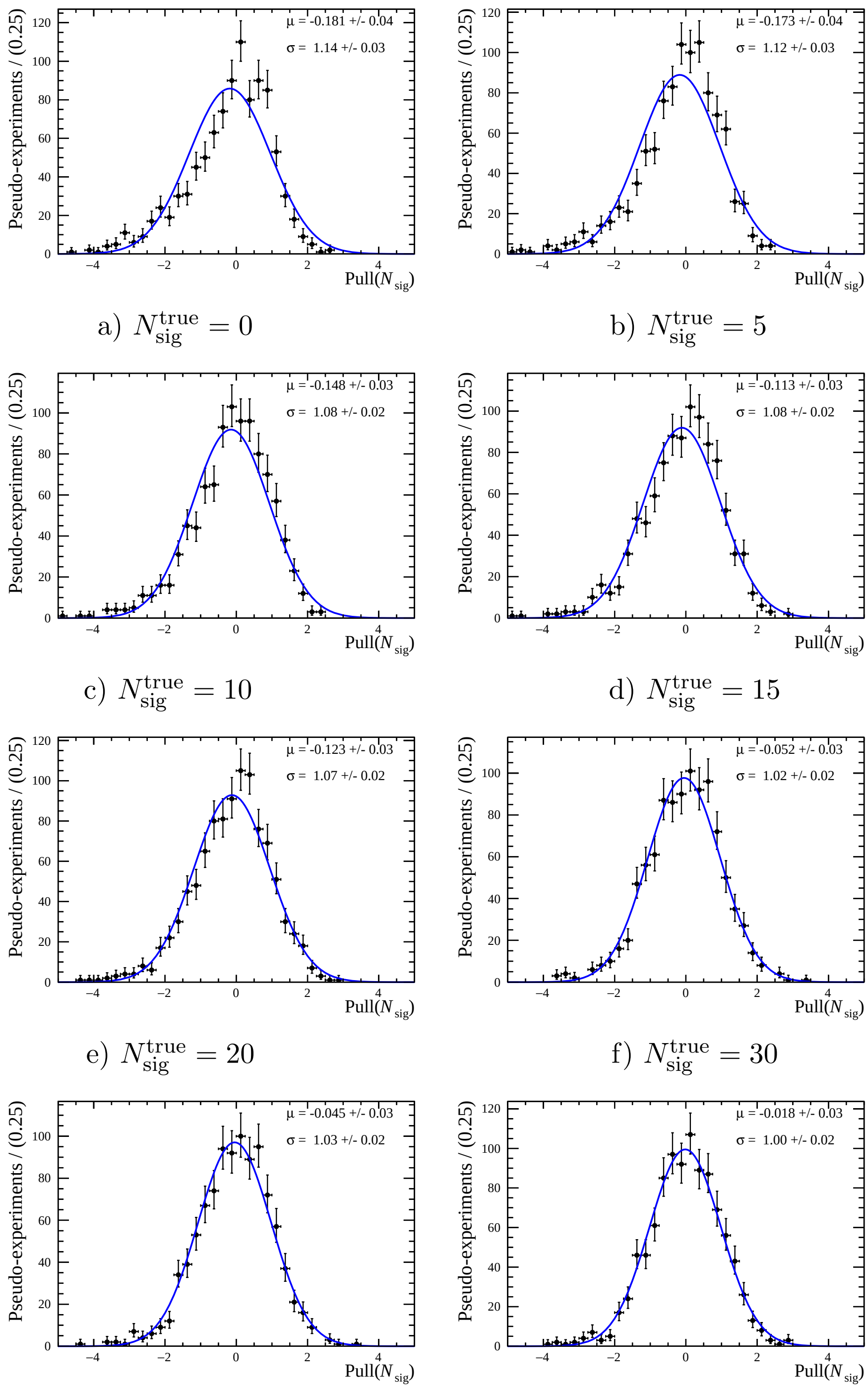
e) $N_{\text{sig}}^{\text{true}} = 20$

f) $N_{\text{sig}}^{\text{true}} = 30$

g) $N_{\text{sig}}^{\text{true}} = 40$

h) $N_{\text{sig}}^{\text{true}} = 50$

Figure H.1: Pull distributions for the signal yield obtained from pseudo-experiments for the low q^2 bin.



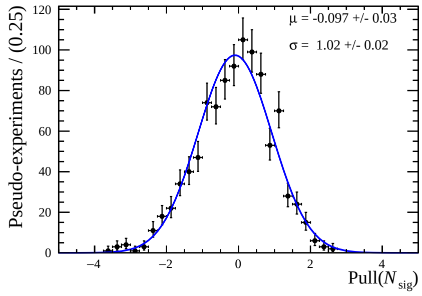
a) $N_{\text{sig}}^{\text{true}} = 0$ b) $N_{\text{sig}}^{\text{true}} = 5$

c) $N_{\text{sig}}^{\text{true}} = 10$ d) $N_{\text{sig}}^{\text{true}} = 15$

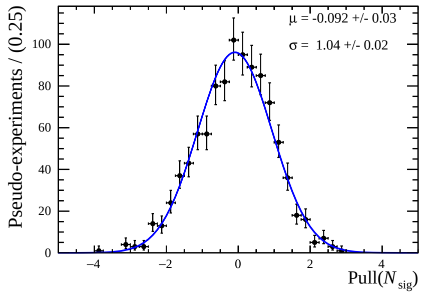
e) $N_{\text{sig}}^{\text{true}} = 20$ f) $N_{\text{sig}}^{\text{true}} = 30$

g) $N_{\text{sig}}^{\text{true}} = 40$ h) $N_{\text{sig}}^{\text{true}} = 50$

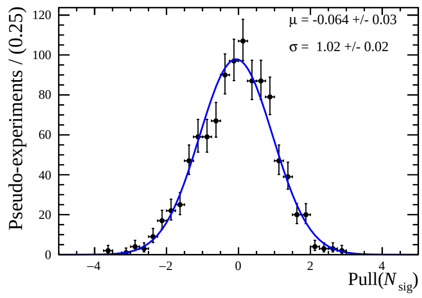
Figure H.2: Pull distributions for the signal yield obtained from pseudo-experiments for the central q^2 bin.



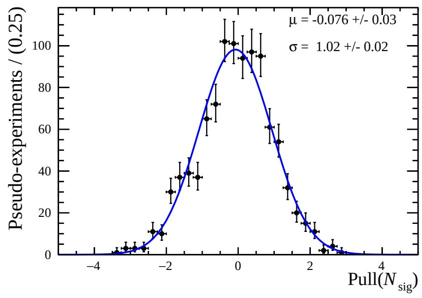
a) $N_{\rm sig}^{\rm true} = 0$



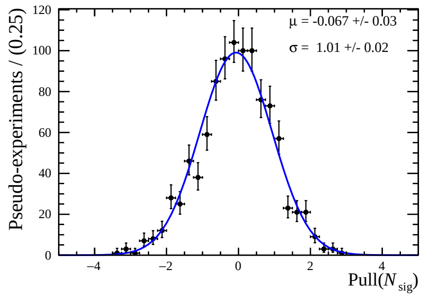
b) $N_{\rm sig}^{\rm true} = 5$



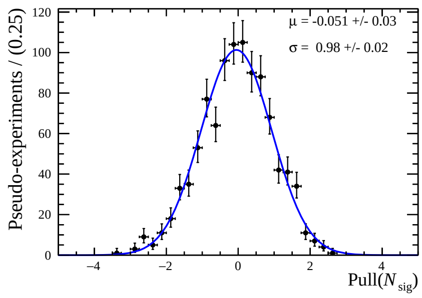
c) $N_{\rm sig}^{\rm true} = 10$



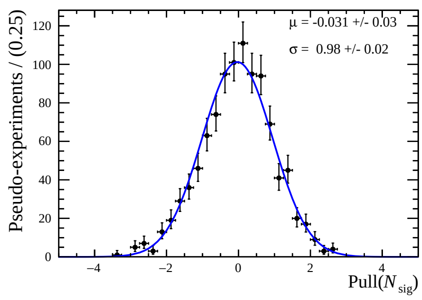
d) $N_{\rm sig}^{\rm true} = 15$



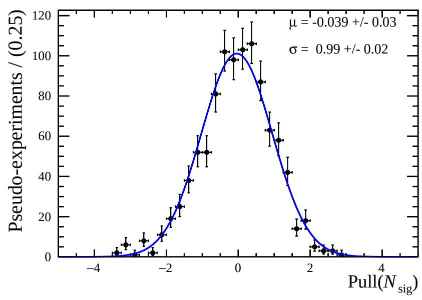
e) $N_{\rm sig}^{\rm true} = 20$



f) $N_{\rm sig}^{\rm true} = 30$



g) $N_{\rm sig}^{\rm true} = 40$



h) $N_{\rm sig}^{\rm true} = 50$

Figure H.3: Pull distributions for the signal yield obtained from pseudo-experiments for the high q^2 bin.

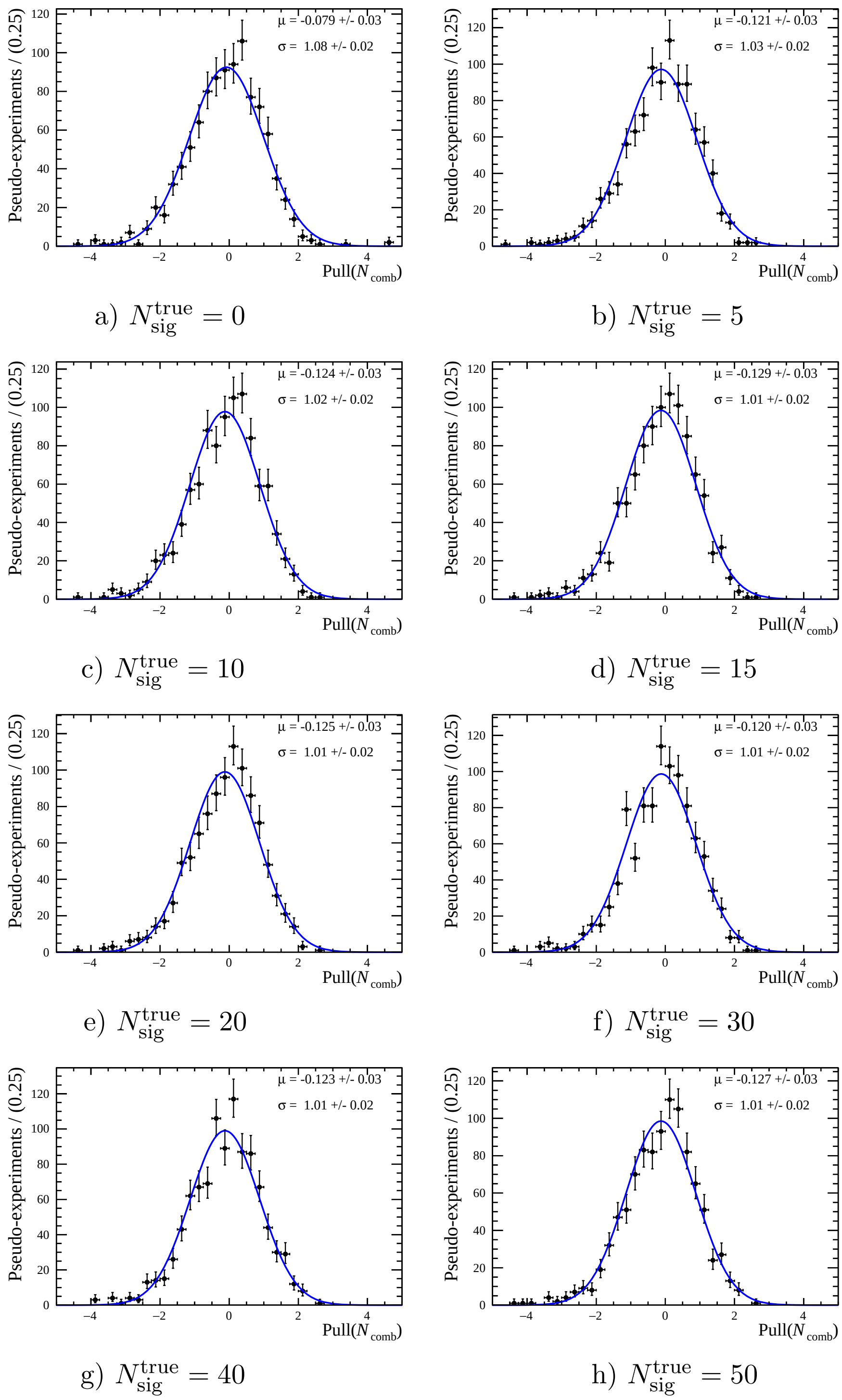
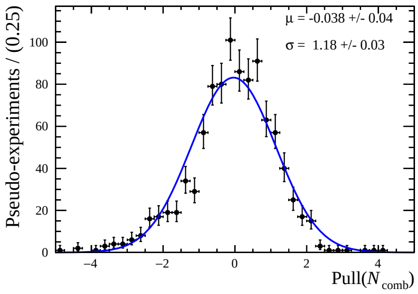
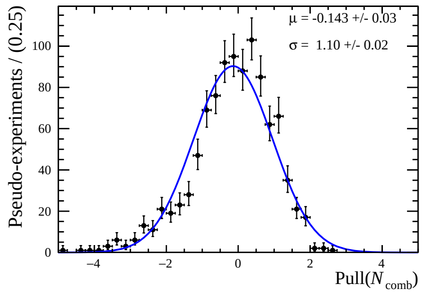


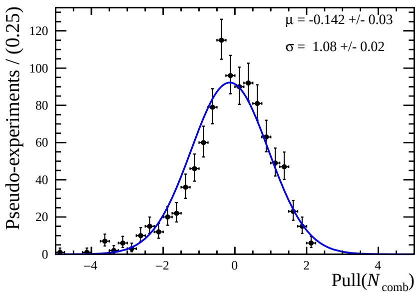
Figure H.4: Pull distributions for combinatorial background yield obtained from pseudo-experiments for the low q^2 bin.



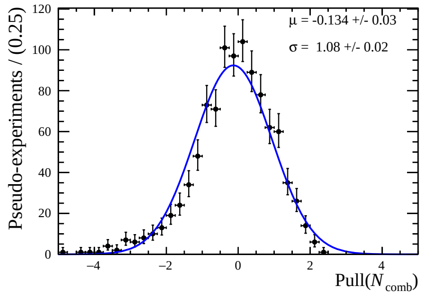
a) $N_{\text{sig}}^{\text{true}} = 0$



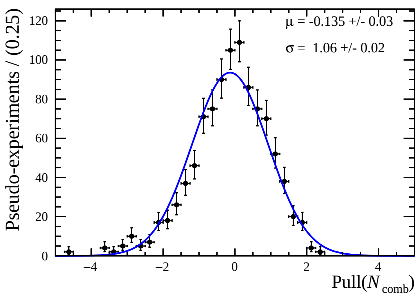
b) $N_{\text{sig}}^{\text{true}} = 5$



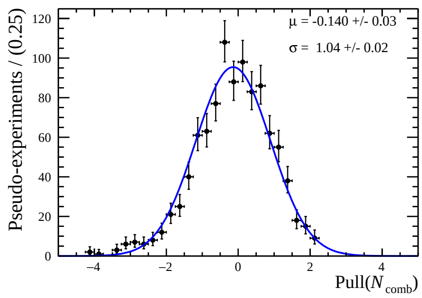
c) $N_{\text{sig}}^{\text{true}} = 10$



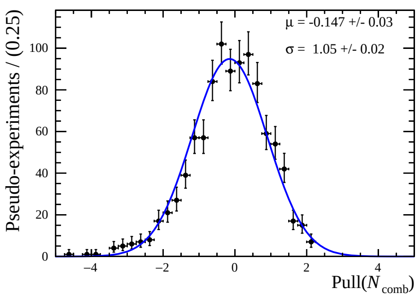
d) $N_{\text{sig}}^{\text{true}} = 15$



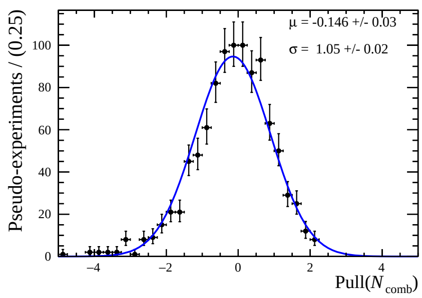
e) $N_{\text{sig}}^{\text{true}} = 20$



f) $N_{\text{sig}}^{\text{true}} = 30$

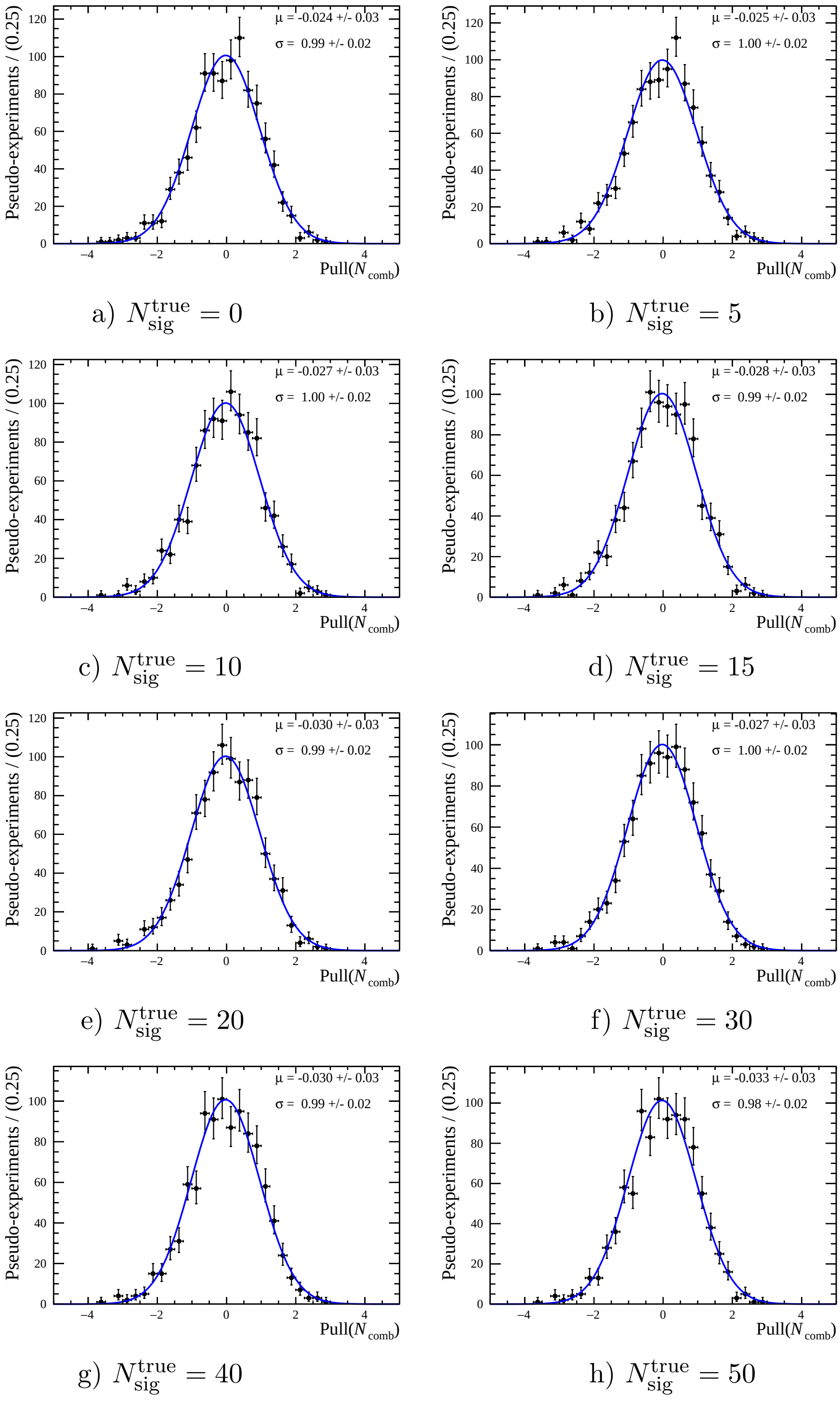


g) $N_{\text{sig}}^{\text{true}} = 40$



h) $N_{\text{sig}}^{\text{true}} = 50$

Figure H.5: Pull distributions for combinatorial background yield obtained from pseudo-experiments for the central q^2 bin.



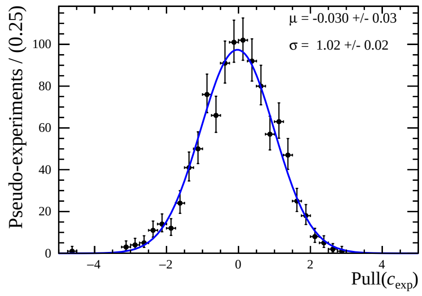
a) $N_{\mathrm{sig}}^{\mathrm{true}} = 0$ b) $N_{\mathrm{sig}}^{\mathrm{true}} = 5$

c) $N_{\mathrm{sig}}^{\mathrm{true}} = 10$ d) $N_{\mathrm{sig}}^{\mathrm{true}} = 15$

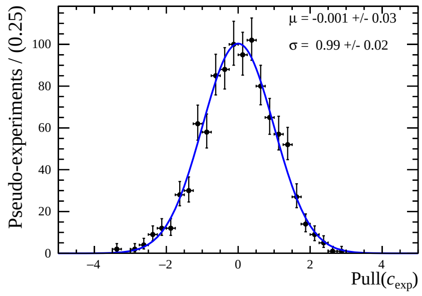
e) $N_{\mathrm{sig}}^{\mathrm{true}} = 20$ f) $N_{\mathrm{sig}}^{\mathrm{true}} = 30$

g) $N_{\mathrm{sig}}^{\mathrm{true}} = 40$ h) $N_{\mathrm{sig}}^{\mathrm{true}} = 50$

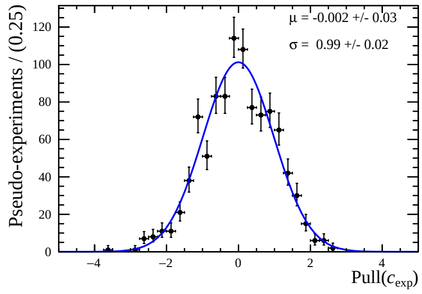
Figure H.6: Pull distributions for combinatorial background yield obtained from pseudo-experiments for the high q^2 bin.



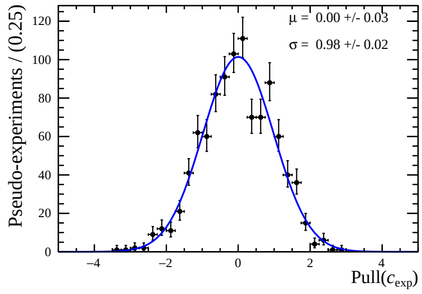
a) $N_{\mathrm{sig}}^{\mathrm{true}} = 0$



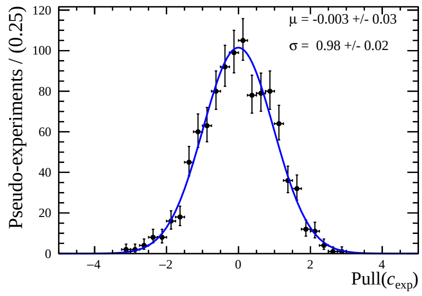
b) $N_{\mathrm{sig}}^{\mathrm{true}} = 5$



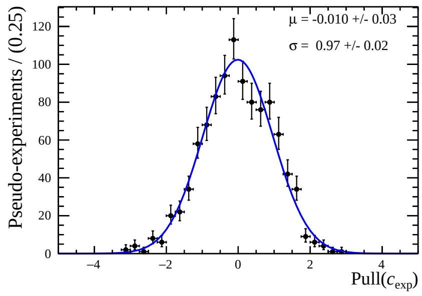
c) $N_{\mathrm{sig}}^{\mathrm{true}} = 10$



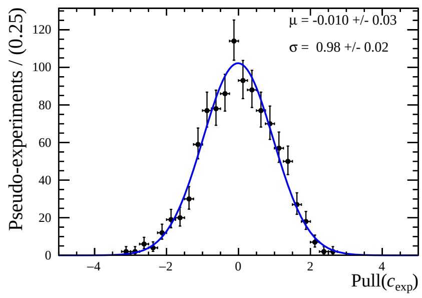
d) $N_{\mathrm{sig}}^{\mathrm{true}} = 15$



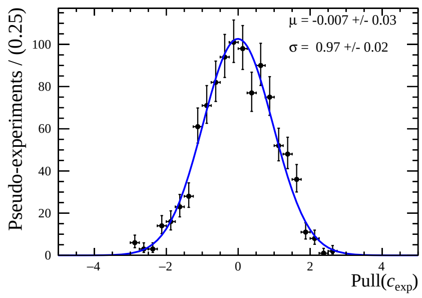
e) $N_{\mathrm{sig}}^{\mathrm{true}} = 20$



f) $N_{\mathrm{sig}}^{\mathrm{true}} = 30$

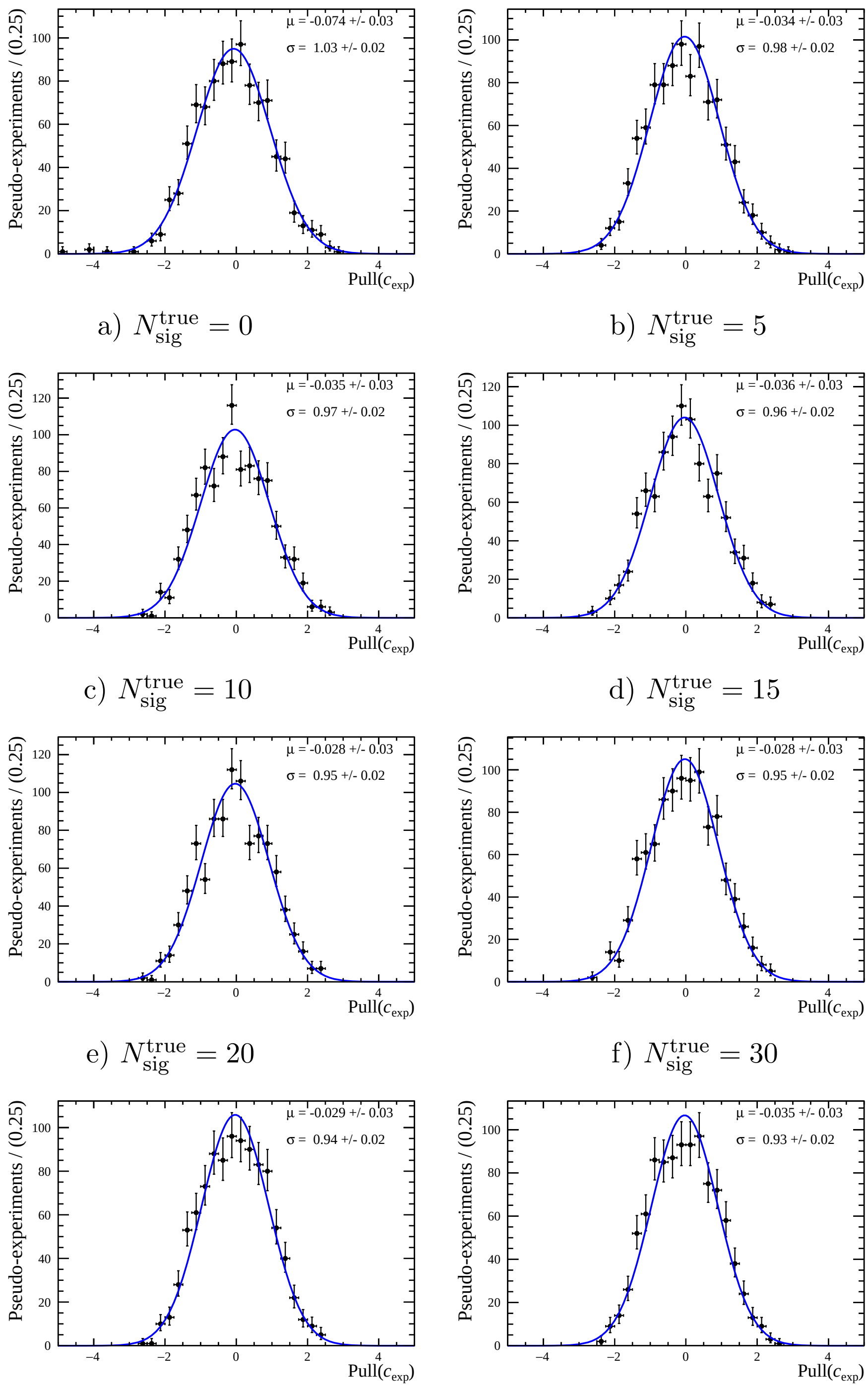


g) $N_{\mathrm{sig}}^{\mathrm{true}} = 40$



h) $N_{\mathrm{sig}}^{\mathrm{true}} = 50$

Figure H.7: Pull distributions for the combinatorial PDF exponent obtained from pesudoexperiments for the low q^2 bin.



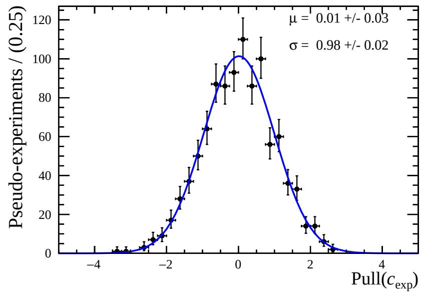
a) $N_{\rm sig}^{\rm true} = 0$ b) $N_{\rm sig}^{\rm true} = 5$

c) $N_{\rm sig}^{\rm true} = 10$ d) $N_{\rm sig}^{\rm true} = 15$

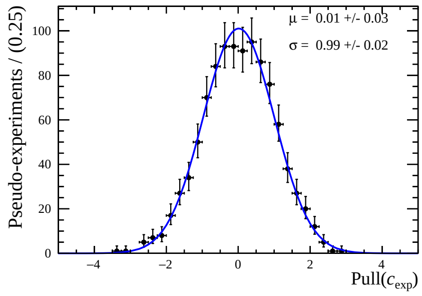
e) $N_{\rm sig}^{\rm true} = 20$ f) $N_{\rm sig}^{\rm true} = 30$

g) $N_{\rm sig}^{\rm true} = 40$ h) $N_{\rm sig}^{\rm true} = 50$

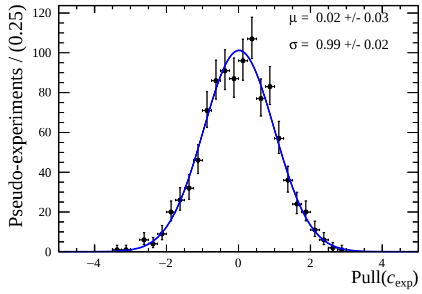
Figure H.8: Pull distributions for the combinatorial PDF exponent obtained from pesudoexperiments for the central q^2 bin.



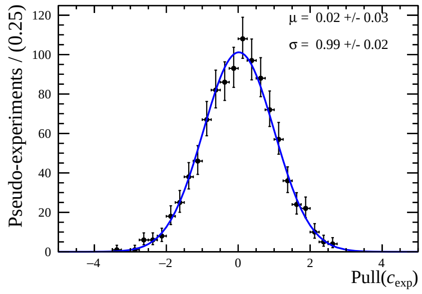
a) $N_{\mathrm{sig}}^{\mathrm{true}} = 0$



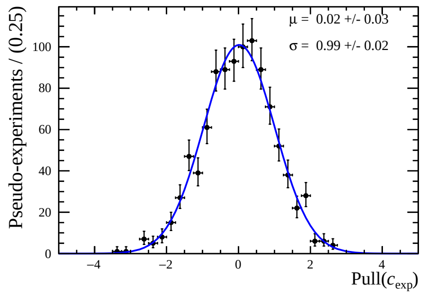
b) $N_{\mathrm{sig}}^{\mathrm{true}} = 5$



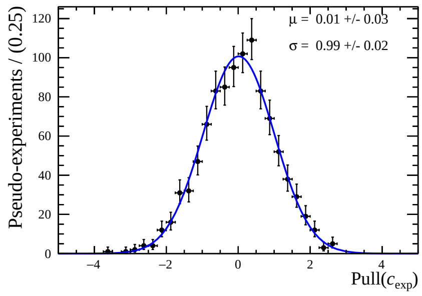
c) $N_{\mathrm{sig}}^{\mathrm{true}} = 10$



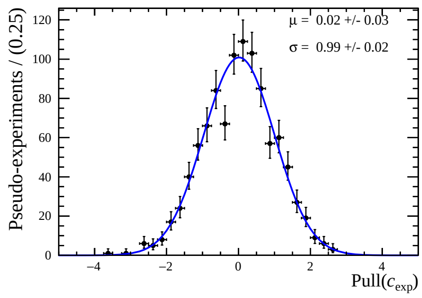
d) $N_{\mathrm{sig}}^{\mathrm{true}} = 15$



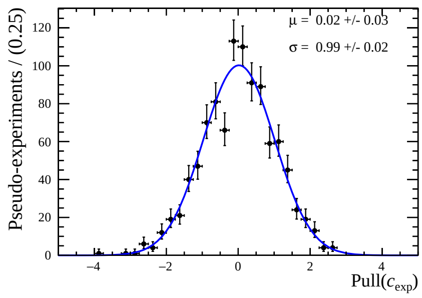
e) $N_{\mathrm{sig}}^{\mathrm{true}} = 20$



f) $N_{\mathrm{sig}}^{\mathrm{true}} = 30$



g) $N_{\mathrm{sig}}^{\mathrm{true}} = 40$



h) $N_{\mathrm{sig}}^{\mathrm{true}} = 50$

Figure H.9: Pull distributions for the combinatorial PDF exponent obtained from pesudoexperiments for the high q^2 bin.

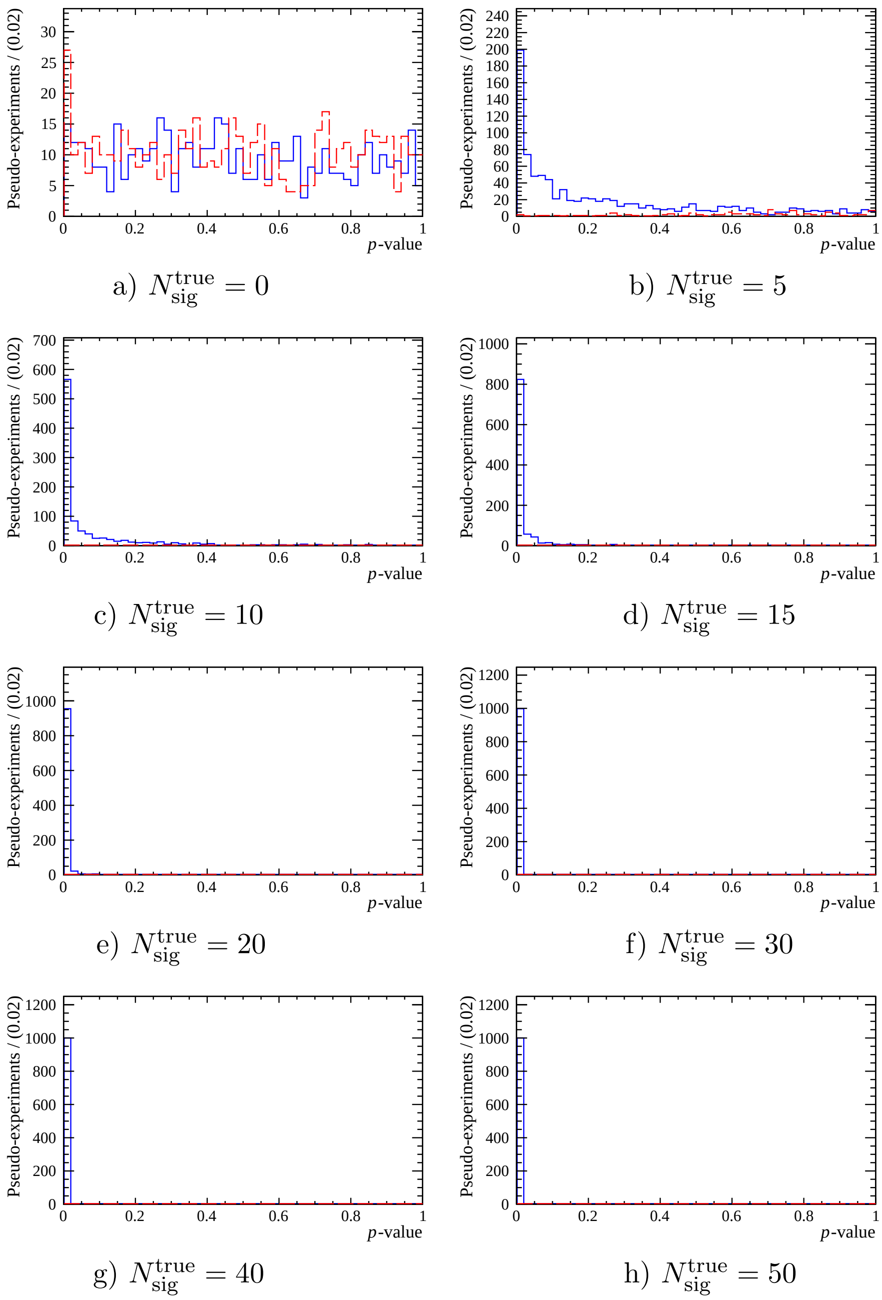


Figure H.10: Distributions of p-value for $B_c^+ \to \pi^+\mu^+\mu^-$ obtained from pseudo-experiments for the low q^2 bin. The solid blue line shows the distributions for pseudo-experiments where the nominal fit converges to a positive signal yield and the dashed red curve shows the distributions for those where the nominal fit converges to a negative signal yield.

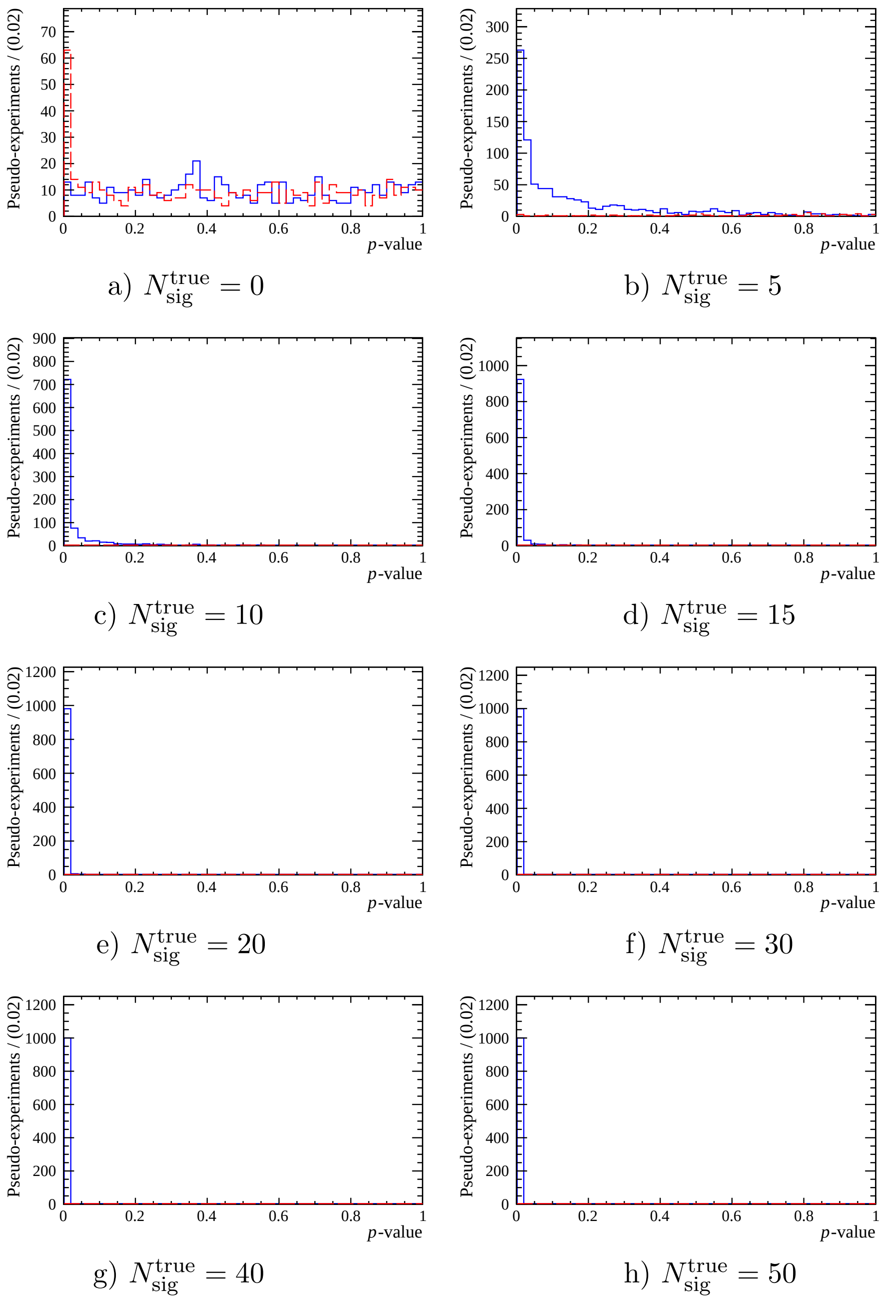
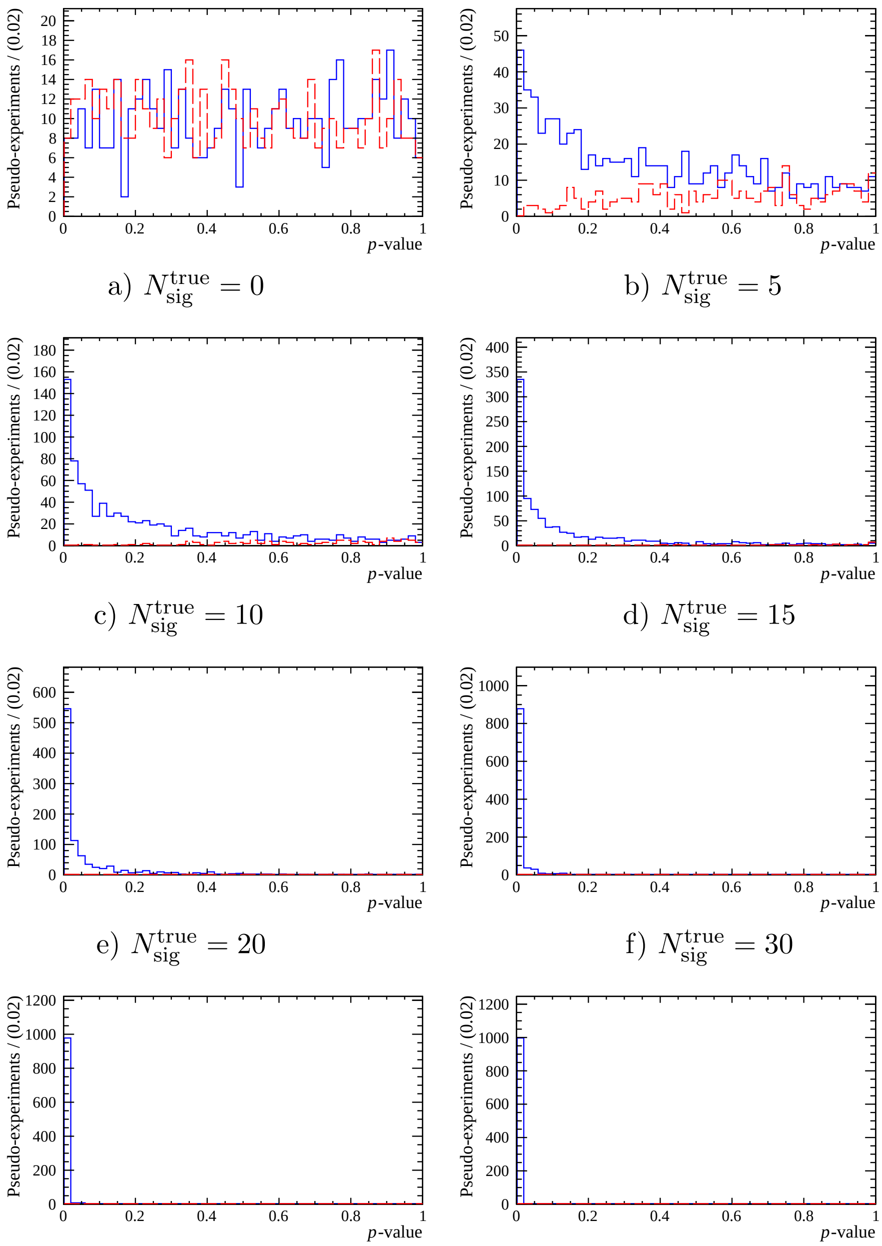


Figure H.11: Distributions of p-value for $B_c^+ \to \pi^+\mu^+\mu^-$ obtained from pseudo-experiments for the central q^2 bin. The solid blue line shows the distributions for pseudo-experiments where the nominal fit converges to a positive signal yield and the dashed red curve shows the distributions for those where the nominal fit converges to a negative signal yield.



a) $N_{\rm sig}^{\rm true} = 0$ b) $N_{\rm sig}^{\rm true} = 5$

c) $N_{\rm sig}^{\rm true} = 10$ d) $N_{\rm sig}^{\rm true} = 15$

e) $N_{\rm sig}^{\rm true} = 20$ f) $N_{\rm sig}^{\rm true} = 30$

g) $N_{\rm sig}^{\rm true} = 40$ h) $N_{\rm sig}^{\rm true} = 50$

Figure H.12: Distributions of p-value for $B_c^+ \to \pi^+\mu^+\mu^-$ obtained from pseudo-experiments for the high q^2 bin. The solid blue line shows the distributions for pseudo-experiments where the nominal fit converges to a positive signal yield and the dashed red curve shows the distributions for those where the nominal fit converges to a negative signal yield.

Bibliography

[1] I. J. R. Aitchison and A. J. G. Hey, *Gauge Theories in Particle Physics: A Practical Introduction; 2nd ed.*, Graduate student series in physics, Hilger, Bristol, 1989.

[2] M. Thomson, *Modern Particle Physics*, Cambridge University Press, 2013.

[3] Particle Data Group, R. L. Workman *et al.*, *Review of particle physics*, Prog. Theor. Exp. Phys. **2022** (2022) 083C01.

[4] Y.-R. Liu *et al.*, *Pentaquark and tetraquark states*, Progress in Particle and Nuclear Physics **107** (2019) 237.

[5] S. Weinberg, *V-a was the key*, Journal of Physics: Conference Series **196** (2009) 012002.

[6] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay, *Evidence for the 2π decay of the k_2^0 meson*, Phys. Rev. Lett. **13** (1964) 138.

[7] A. D. Sakharov, *Violation of cp invariance, c asymmetry, and baryon asymmetry of the universe*, Soviet Physics Uspekhi **34** (1991) 392.

[8] CDF, F. Abe *et al.*, *Observation of the B_c meson in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV*, Phys. Rev. Lett. **81** (1998) 2432, `arXiv:hep-ex/9805034`.

[9] CDF, T. Aaltonen *et al.*, *Observation of the Decay $B_c^+ \to J/\psi\pi^\pm$ and Measurement of the B_c^+ Mass*, Phys. Rev. Lett. **100** (2008) 182002, `arXiv:0712.1506`.

[10] CMS, A. M. Sirunyan *et al.*, *Measurement of b hadron lifetimes in pp collisions at $\sqrt{s} = 8$ TeV*, Eur. Phys. J. C **78** (2018) 457, `arXiv:1710.08949`, [Erratum: Eur.Phys.J.C 78, 561 (2018)].

[11] LHCb collaboration, R. Aaij *et al.*, *Measurement of the lifetime of the B_c^+ meson using the $B_c^+ \to J/\psi\pi^+$ decay mode*, Phys. Lett. **B742** (2015) 29, `arXiv:1411.6899`.

[12] LHCb collaboration, R. Aaij *et al.*, *Precision measurement of the B_c^+ meson mass*, JHEP **07** (2020) 123, arXiv:2004.08163.

[13] ATLAS, M. Aaboud *et al.*, *Measurement of the relative $B_c^{\pm}/B^{\pm}$ production cross section with the ATLAS detector at $\sqrt{s} = 8$ TeV*, Phys. Rev. D **104** (2021) 012010, arXiv:1912.02672.

[14] LHCb collaboration, R. Aaij *et al.*, *Measurement of B_c^+ production in proton-proton collisions at $\sqrt{s} = 8$ TeV*, Phys. Rev. Lett. **114** (2015) 132001, arXiv:1411.2943.

[15] LHCb collaboration, R. Aaij *et al.*, *Observation of the decay $B_c^+ \to \psi(2S)\pi^+$*, Phys. Rev. **D87** (2013) 071103(R), arXiv:1303.1737.

[16] LHCb collaboration, R. Aaij *et al.*, *Measurement of the branching fraction ratio $\mathcal{B}(B_c^+ \to \psi(2S)\pi^+)/\mathcal{B}(B_c^+ \to J/\psi\pi^+)$*, Phys. Rev. **D92** (2015) 057007, arXiv:1507.03516.

[17] ALICE Collaboration, K. A. et al. *The alice experiment at the cern lhc*, JINST **3** (2008).

[18] ATLAS Collaboration, G. A. et al. *The atlas experiment at the cern large hadron collider*, JINST **3** (2008).

[19] CMS Collaboration, S. C. et al. *The cms experiment at the cern lhc*, JINST **3** (2008).

[20] LHCb Collaboration, A. A. et al. *The lhcb detector at the lhc*, JINST **3** (2008).

[21] LHCb Collaboration, *Lhcb technical design report: Reoptimized detector design and performance*, CERN-LHCC- **9** (2003) 2003.

[22] M. P. Altareilli and F. Teubert, *B physics at lhcb*, International Journal of Modern Physics A **23** (2008) 5117, arXiv:https://doi.org/10.1142/S0217751X08042791.

[23] R. Aaij *et al.*, *Performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC*, JINST **14** (2019) P04013, arXiv:1812.10790.

[24] *PIDCalib Packages*, https://twiki.cern.ch/twiki/bin/view/LHCb/PIDCalibPackage.

[25] *BuToPiMuMu analyses LHCb internal Twiki page*, https://twiki.cern.ch/twiki/bin/viewauth/LHCbPhysics/BuToPiMuMuNonresonant.

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

[27] T. Chen and C. Guestrin, *XGBoost: A scalable tree boosting system*, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, (New York, NY, USA), 785–794, ACM, 2016.

[28] G. Punzi, *Sensitivity of searches for new signals and its optimization*, eConf **C030908** (2003) MODT002, arXiv:physics/0308063.

[29] F. Abudinén *et al.*, *Search for $D^{*0} \to \mu^+\mu^-$ decays in reconstructed $B^- \to \pi^-\mu^+\mu^-$ decays*, LHCb-ANA-2022-022, 2022.

[30] LHCb collaboration, R. Aaij *et al.*, *Measurement of the ratio of branching fractions $\mathcal{B}(B_c^+ \to J/\psi K^+)/\mathcal{B}(B_c^+ \to J/\psi \pi^+)$*, JHEP **09** (2016) 153, arXiv:1607.06823.

[31] S. S. Wilks, *The large-sample distribution of the likelihood ratio for testing composite hypotheses*, Ann. Math. Stat. **9** (1938) 60.

[32] F. Porter, *Low statistics likelihood fits are tricky*, 2001. http://www.hep.caltech.edu/~fcp/statistics/maxlikfitnote.ps.

[33] I. Narsky and F. Porter, *Statistical analysis techniques in particle physics: Fits, density estimation and supervised learning*, Wiley, 2013. Likelihood fits with low statistics are discussed in Sec. 2.3.2.

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

[35] J. A. Hanley and B. MacGibbon, *Creating non-parametric bootstrap samples using poisson frequencies*, Computer Methods and Programs in Biomedicine **83** (2006) 57.

[36] G. J. Feldman and R. D. Cousins, *A unified approach to the classical statistical analysis of small signals*, Phys. Rev. **D57** (1998) 3873, arXiv:physics/9711021.

[37] R. Aaij *et al.*, *Selection and processing of calibration samples to measure the particle identification performance of the LHCb experiment in Run 2*, Eur. Phys. J. Tech. Instr. **6** (2018) 1, arXiv:1803.00824.

[38] LHCb collaboration, R. Aaij *et al.*, *Measurement of the track reconstruction efficiency at LHCb*, JINST **10** (2015) P02007, arXiv:1408.1251.

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

[40] A. Rogozhnikov, *Reweighting with Boosted Decision Trees*, J. Phys. Conf. Ser. **762** (2016) , arXiv:1608.05806, https://github.com/arogozhnikov/hep_ml.