Determination of the $r_{J/\psi}^{K}$ and $r_{J/\psi}^{K^*}$ Ratios for Lepton Universality Measurements at LHCb

Ryan Bernard Calladine

Thesis submitted for the degree of
Doctor of Philosophy

Particle Physics Group,
School of Physics and Astronomy,
University of Birmingham.

May 19, 2021

Abstract

Lepton Universality is an integral part of the Standard Model of Particle Physics. This thesis describes new measurements carried out as part of a programme to test this by comparing ratios of branching fractions of rare B meson decays that include either electrons or muons in their otherwise equivalent final states. The ratios, denoted R_{K^*} and R_K for $B^0 \to K^{*0}\ell^+\ell^-$ and $B^+ \to K^+\ell^+\ell^-$ decays respectively, have been measured previously by LHCb, showing tension with the Standard Model expectation at a level of 2–3σ.

This thesis investigates the simultaneous extraction of both of these ratios and focuses on measurements of $r^K_{J/\psi}$ and $r^{K^*}_{J/\psi}$. These have the same final state particles as the rare decay but are not expected to be sensitive to any physics beyond the Standard Model as they are made in intervals of the squares of the dilepton invariant masses corresponding to the decay of J/ψ mesons. Hence, they are the ideal and mandatory environment in which to study all aspects of the analysis to avoid bias in the primary measurements or R_{K^*} and R_K.

All pp collision data recorded by the LHCb detector, corresponding to an integrated luminosity of approx. $9\,\text{fb}^{-1}$ at energies 7–13 TeV, are used. Measurements of the integrated $r_{J/\psi}$ value generally agree within 3σ, while $r_{J/\psi}$ itself is flat as a function of most kinematic and topological variables considered. Results obtained using corrections derived from B^+ or B^0 modes are consistent. These detailed measurements of $r_{J/\psi}$ demonstrate that it is now acceptable to proceed with the high-profile determination of R_K and R_{K^*}.

In addition, as validation of the LHCb detector simulation is vitally important for reliable physics measurements, a test of the muon multiple scattering in the LHCb detector is presented and its implementation into LHCbPR, a browser-based monitoring system which collates regular test results, is presented.

Acknowledgements

This thesis would not be possible without the help and support of a number of individuals who I will thank here. Firstly I would like to thank Nigel Watson who was always there to help and support me with his supervision, often at short notice, and made me feel very welcome in the Birmingham group. Secondly I owe a huge debt of gratitude to Simone Bifani who taught me a huge amount and without whom almost none of this thesis would have been possible. Thank you for your help, patience and support and putting up with me on a daily basis for around 3 years, it's greatly appreciated. I would also like to thank everyone involved in the RX analysis group, it was a pleasure to work with you all over the years. In particular I would like to thank Renato Quagliani for our many useful discussions as well as my fellow students Fabrice, Alex, Stephan and Da Yu who joined the group around the same time as me.

I would also like to thank everyone in the Birmingham group particularly Russell, Rob, John and Jack who came through the ranks with me and Tim, Georgios, Kristian, Nat, Jon, Naomi who have populated the West 313 office over the years for their company and advice. I should also mention everyone who I met whilst on LTA in Switzerland, of which there are too many to list, that helped me settle in and were great company over the 18 months I spent there, particularly Nicholas, Greg and Jack who I lived with and Charlie who I played many frames of snooker with. It would also be remiss of me to not mention an excursion to Interlaken with Dan, Matt and Esh of which I will have fond memories of for years to come.

Lastly I would like to thank my parents, the rest of my family and all my close friends who have always been very supportive of me, not only through my years as a PhD student but through my whole life.

Author's Contribution

I am one of the main authors of the analysis presented in this thesis actively participating with Simone Bifani or Renato Quagliani in most aspects. I developed and implemented the code for the MVA training and validation as well as the MVA cut optimisation. I developed the code which computes the weights which account for the q^2 smearing. I was involved with the the production and processing of various simulated samples and attached various corrections to these samples. I also helped to develop the code which produced the final flatness plots presented in this thesis.

With regards to my service work for the LHCb experiment I updated the code for the Muon Multiple Scattering test to work with the current simulation framework and wrote a handler to export the results of the test to LHCbPR.

Contents

Chapter 1

Introduction

1.1 The Standard Model of Particle Physics

The Standard Model (SM) of particle physics is a quantum field theory based on three gauge groups, $U(1)_Y \times SU(2)_L \times SU(3)_C$, which describes our current knowledge of particle physics [1–9]. The SM provides a description of three of the four fundamental forces of nature (excluding gravity) and how they interact with the fundamental particles of matter. In the SM, there are two categories of particles, namely 'matter' particles and so-called 'gauge bosons', the quanta that are exchanged between matter particles and are responsible for mediating the three forces of the SM. The matter particles are known as fermions, have half-integer spins and exist in two types: quarks, which are subject to the strong force, and leptons, which are not. The force carriers are known as bosons and have integer spins. The fermions and bosons of the SM and their properties are summarised in Fig. 1.1.

1.1.1 Fermions

Quarks and Hadrons

Quarks [11–13] are the building blocks of particles such as the proton and neutron which form atomic nuclei. There are six types of quark, known as flavours, which can be divided into two categories based upon their electric charge (in units of the magnitude of the charge of the electron, $|e|$). The up-type quarks up, charm and top (u, c, t) have a charge of $+2/3$. The down-type quarks down, strange and bottom (d, s, b) have a charge of $-1/3$. Quarks also carry what is known as colour charge and this can be red, green or blue (r, g, b). The internal (or additive)

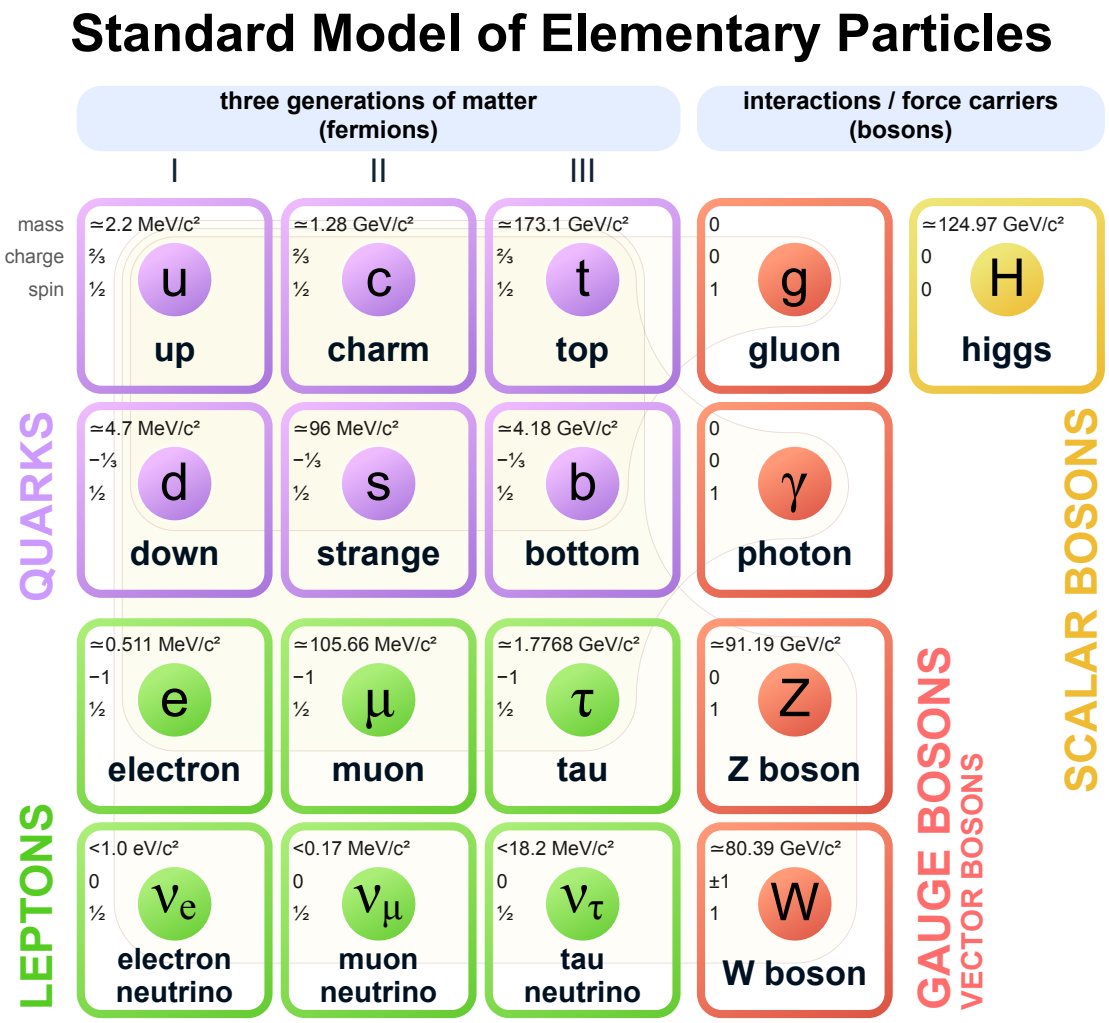


Figure 1.1: Table of the fundamental particles within the SM [10].

quantum numbers of antiparticle quarks (antiquarks) are equal in magnitude but inverted in sign relative to those of matter quarks, meaning an anti-down quark ($\bar{d}$) has an electric charge of $+1/3$ and may have a colour charge of anti-red ($\bar{r}$). Quarks may also be divided by generation with the first generation consisting of u and d, the second consisting of c and s and the third consisting of t and b. Each generation contains an up-type and down-type quark. Masses are lowest in the first generation and highest in the third generation, with the top quark being by far the most massive.

Unbound, or free, quarks have not been observed [14] and the most common arrangements are groups of either a quark-antiquark pair or three quarks or three antiquarks. Particles consisting of a quark-antiquark pair, in general of differing flavours, are

known as mesons. Particles consisting of three quarks (antiquarks), such as the proton (anti-proton), are known as baryons. Each quark (antiquark) is assigned a so-called baryon number [15–17], an additive quantum number based on phenomenological observations, of $1/3$ $(-1/3)$ and its conservation in all interactions ensures the stability of the proton. The family of particles which are composed of quarks are collectively known as hadrons and this includes the recently discovered pentaquarks, which are five quark states, claimed by LHCb [18, 19]. For a particle to exist freely as a hadron it is hypothesised that it must be in a colour neutral state, which can be satisfied by having bound states consisting of quark-antiquark pairs or three quarks (three antiquarks). Since a single quark carries one colour charge it can never be colour neutral and therefore should not exist in isolation, which is consistent with experimental evidence [20].

Leptons

Leptons [1] are divided into two categories: charged leptons, such as the electron, and neutral leptons (neutrinos). The charged leptons have electric charges of -1 and come in three flavours, electron, muon and tau (e, μ, τ). The neutrinos have no electric charge and also come the in same three flavours, namely electron, muon and tau neutrinos (ν_e, ν_μ, ν_τ). Leptons do not carry colour charge. The leptons are also split into three generations each containing a charged lepton and a neutrino, the first generation consists of e and ν_e, the second consists of μ and ν_μ and the third consists of τ and ν_τ. As before the only difference between the generations is the masses of the particles with the tau being the most massive charged lepton.

Leptons carry an additive quantum number known as a lepton flavour number [21–23] and it is postulated that this must be conserved in interactions within the SM. The flavour number must be conserved overall and also separately within each generation. Anti-leptons have opposite charge and opposite lepton numbers. An example of lepton flavour conservation is that of the muon decay $\mu^- \to e^- + \overline{\nu}_e + \nu_\mu$. In the initial state, the muon lepton number, L_μ is $+1$. In the final state, L_μ is still $+1$ and the electron lepton number, $L_e = 1 - 1 = 0$ hence overall lepton number is conserved and this decay is allowed. This means that even though all the neutrinos are massless in the SM they cannot all be the same particle since three flavours are required to ensure lepton flavour number conservation. It is however still unclear whether the neutrino is its own antiparticle [24].

1.1.2 Bosons

There are five bosons in the SM and these can be divided into the vector bosons, which arise from gauge symmetry, and a scalar boson resulting from spontaneously breaking a symmetry [6–8]. The vector bosons consist of gluons [4, 5] and the so-called electroweak bosons [1–3]: the photon, $W^{\pm}$ bosons and Z boson, and are responsible for three of the fundamental forces. The scalar boson is the Higgs boson responsible for giving particles mass. Forces manifest themselves via the exchange of virtual particles, which are quantum fluctuations, that may exist at a different mass to the 'real' particle, or off mass-shell, for a time dictated by the uncertainty principle.

The photon and the electromagnetic force

The photon is a massless, spin-1 boson responsible for the electromagnetic (EM) force which binds electrons and the nucleus together within an atom. The photon couples to any electrically charged particle making neutrinos the only fermions which do not interact via the EM force. Since the photon is massless it travels at the speed of light, c. Electromagnetism is the second strongest force, characterised by an inverse square law on a macroscopic scale with an infinite range.

Electromagnetism arises naturally from the requirement that the theory should be invariant under a local phase transformation $e^{iq\alpha(\mathbf{x},t)}$, known as a $U(1)$ transformation, see e.g. Ref. [25]. The extra term added to the Lagrangian to satisfy this requirement is associated with electromagnetic interactions.

Gluons and the strong nuclear force

The gluons are massless, spin-1 bosons responsible for the strong nuclear force which binds protons and neutrons together in the nucleus. Gluons couple to any particle having colour charge, such as quarks. However, as gluons also have colour charge they self-interact, in contrast to the electromagnetic interaction of the photon. In contrast to quarks, gluons carry two colours, such as $r\bar{g}$ for example. Gluons may exist in eight colour permutations to satisfy the overall symmetry properties of baryons, see e.g. Ref. [26]. The strong force is the strongest fundamental interaction and it has a very short range due to the potential that is itself a consequence of the gluons' self-interaction. The strong potential between a quark antiquark pair can be written as (e.g. Ref. [27])

$$V = -\frac{\alpha}{r} + br, \tag{1.1}$$

where r is the distance between the quarks and a and b are coefficients. In Eq. 1.1 we see at short distances (or high energies) the $1/r$ term dominates. At large distances (low energies) the term linear in r dominates, such that at a distance of around 2 fm it is energetically favourable to produce a pair of lighter quarks than to further increase the separation of a pair of colour-connected fermions. Due to this linear term in r the force effectively only has a very small range. This term is the source of the postulated confinement that prevents the existence of quarks outside of bound states. Gluons are a natural consequence of invariance of the theory under $SU(3)_C$ transformations where the C stands for colour.

The $W^{\pm}$ and Z bosons and the weak nuclear force

The $W^{\pm}$ and Z are massive, spin-1 bosons, responsible for the weak force which causes β decay of the neutron. All fermions are subject to the weak force. The weak force is weaker than both electromagnetism and the strong force at everyday energy scales. Given the $W^{\pm}$ and Z are so massive (around 80 GeV and 91 GeV, respectively, [14]), it follows from the Uncertainty Principle that the effective range of the force is very small therefore it appears to be weak. However, intrinsically the force has a comparable coupling strength to electromagnetism and at energy scales greater than the mass of the $W^{\pm}$ the weak and EM forces contribute roughly equally. It is possible to describe both the weak and EM interactions as manifestations of a unified electroweak (EW) interaction [1–3].

An interesting aspect of the weak interaction is that it violates parity symmetry, namely, the transformation $(x,y,z) \rightarrow (-x,-y,-z)$. As a result of this observation [28], it was determined that the weak interaction only interacts with the left-handed chiral component of particle states, hence the L in $SU(2)_L$. Although parity was found not to be respected in nature, it was expected that the combined operation of parity and charge conjugation would remain intact. However, this was demonstrated not to be the case by experiment [29]. Since CP violation is the only known mechanism of arriving at an imbalance of matter and antimatter in the universe this is a particularly interesting area of study.

Another novel property of the weak interaction is that it is the only force able to change quark flavour. Flavour is generally conserved in every interaction and for leptons this is true even for the weak interaction. However, when the weak interaction is applied to quarks they may change flavour with varying probabilities depending on the transition occurring. Changing flavour within a generation is

preferred but non-zero probabilities exist for flavour changes between the generations. This will be further discussed in Sec. 1.2. This opens the door for the study of rare decays where due to the low rate in the SM it may be possible to detect effects from Beyond the Standard Model (BSM) particles.

The Higgs boson and the Yukawa Interaction

The Higgs boson is a massive, spin-0 boson, responsible for giving mass to the $W^{\pm}$ and Z bosons as well as all the fermions through so-called Yukawa interactions. The discovery of the Higgs boson in 2012 [30,31] completed the SM and the world-average measured mass is $125.10 \pm 0.14\,\text{GeV}$ [14]. In contrast to the other bosons it does not result from a gauge symmetry of the Lagrangian but from a spontaneous breaking of a symmetry, namely the electroweak symmetry. It is postulated that the vacuum state spontaneously breaks this symmetry, giving the $W^{\pm}$ and Z bosons their masses and creating the Higgs boson. The Higgs boson then must also act to give the fermions their masses. This is done by having the fermion and Higgs fields interacting with each other in so-called Yukawa interactions which are gauge invariant. The strength of the interaction is proportional to the mass of the fermion in question, meaning the Higgs field interacts much more strongly with the t quark, $W^{\pm}$ and Z bosons than the leptons or first generation quarks.

1.2 Flavour and the weak interaction

1.2.1 The Cabibbo angle

As noted previously, the weak force is the only known interaction able to change quark flavour. At a time when only the u, d and s quarks were known Cabibbo [32] suggested that the weak coupling constant is universal but shared between quarks with interactions proceeding between a u or d quark and an s quark being suppressed relative to interactions between u and d quarks. The reason for introducing this came from trying to explain the two weak decays $K \to \mu\nu_{\mu}$ and $K \to \pi^0 e\nu_e$, the first corresponding to a $d \to u + W^-$ vertex and the second a $s \to u + W^-$. The first process is around 12 times more likely to occur than the second [14].

In order to keep the universality of the weak coupling but simultaneously explain the discrepancy in these decay rates, Cabibbo introduced the Cabibbo angle, θ_C. He proposed that the mass eigenstates which take part in the weak interactions are

rotations of the flavour eigenstates and suggested the doublet,

$$\begin{pmatrix} u \\ d' \end{pmatrix} = \begin{pmatrix} u \\ d \cdot \cos\theta_C + s \cdot \sin\theta_C \end{pmatrix}. \tag{1.2}$$

This quark mixing of the flavour eigenstates could explain the discrepancies in the branching ratios of $d \to u$ and $s \to u$ transitions, since $\sin\theta_C \sim 0.22$.

1.2.2 The GIM mechanism

The problem with Cabibbo's solution was that it predicted very similar rates for $K \to \mu\nu_\mu$ and $K_L^0 \to \mu\mu$ with the second decay being a $s \to d$ transition, known as a flavour changing neutral current (FCNC) since the transition occurs between two quarks of the same electric charge. Cabibbo's model permitted these types of transitions to occur at tree level, which would violate charge conservation at the vertex and lead to a decay rate much larger than observed. The solution was proposed by Glashow, Iliopoulus and Maiani and known as the GIM mechanism [33]. They proposed a second doublet with an up type partner for the strange quark,

$$\begin{pmatrix} c \\ s' \end{pmatrix} = \begin{pmatrix} c \\ -d\sin\theta_C + s\cos\theta_C \end{pmatrix}, \tag{1.3}$$

denoted by c and called the charm quark, which was at this time undiscovered. This second doublet cancels the FCNC term coming from Cabibbo's doublet and demands that the FCNC must proceed via a 'box' diagram with virtual u and c propagators, explaining the small branching fraction in $K_L^0 \to \mu\mu$ decays. This cancellation can be seen in Fig. 1.2. The charm quark was later discovered [34, 35], vindicating this idea.

1.2.3 The CKM matrix

Secs. 1.2.1 and 1.2.2 describe the hypothetical case of only two generations of quarks; however, it is known that there are three generations. In this case a single mixing angle will not suffice. To describe the mixing a 3×3 unitary matrix called the Cabibbo-Kobayashi-Maskawa (CKM) matrix [37] is used. The unitarity of the matrix ensures that the universality of the weak interaction is respected. A general 3×3 complex matrix has $2N^2$ real parameters; however, the requirement of unitarity, $A(A_T)^* = I$, results in nine constraints of which only five are independent of each other. This means there are four independent real parameters, three angles and one

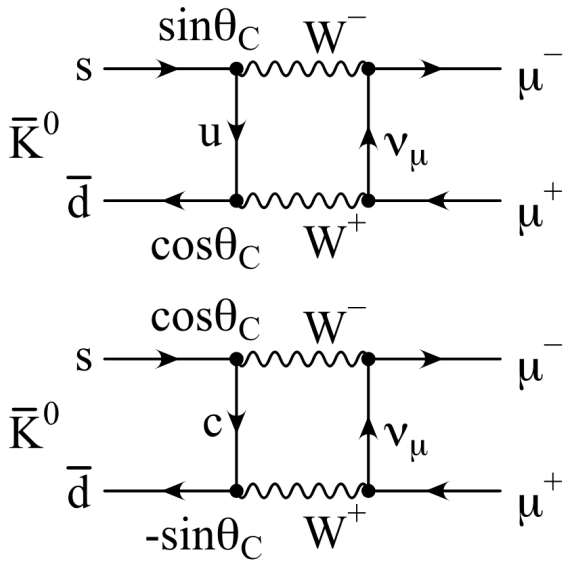


Figure 1.2: Feynman diagrams illustrating how the GIM mechanism accounts for $K_L^0 \to \mu\mu' s$ low branching fraction [36].

complex phase. This complex phase, δ, is responsible for CP violation, which is only present in the weak interaction. In general a complex unitary matrix has $(N-1)^2$ real parameters, interestingly, this means that three generations is the minimum number required for CP violation to exist.

The CKM matrix, V_{CKM}, consists of elements V_{ij} which describe the amplitude of a weak interaction between two quarks i and j. Eq. 1.4 reports the current measured values of each of the CKM elements [14]. Transitions occurring within the same generation are heavily favoured whilst transitions between the first and third generations are heavily suppressed. A useful parametrisation of the CKM matrix is that of Wolfenstein [38], seen in Eq. 1.5. The three real independent parameters are represented by λ, A and ρ whilst the imaginary parameter is represented by η; they

current values, determined from experiments, are stated in Eq. 1.6 [14].

$$V_{CKM} \equiv \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}$$

$$= \begin{pmatrix} 0.97446 \pm 0.00010 & 0.22452 \pm 0.00044 & 0.00365 \pm 0.00012 \\ 0.22438 \pm 0.00044 & 0.97359^{+0.00010}_{-0.00011} & 0.04214 \pm 0.00076 \\ 0.00896^{+0.00024}_{-0.00023} & 0.04133 \pm 0.00074 & 0.999105 \pm 0.000032 \end{pmatrix} \quad (1.4)$$

$$V_{CKM} = \begin{pmatrix} 1 - \lambda^2/2 & \lambda & A\lambda(\rho - i\eta) \\ -\lambda & 1 - \lambda^2/2 & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4) \quad (1.5)$$

$$\lambda = 0.22650 \pm 0.00048, \quad \eta = 0.357 \pm 0.011, \quad \rho = 0.141^{+0.016}_{-0.017}, \quad A = 0.790^{+0.017}_{-0.012}. \quad (1.6)$$

Again considering unitarity, two types of conditions are imposed on the elements of the CKM matrix,

$$\Sigma_i V_{ij} V^*_{ik} = \delta_{jk} \quad \text{and} \quad \Sigma_j V_{ij} V^*_{kj} = \delta_{ik}. \quad (1.7)$$

There are six combinations which equal zero and they can be represented as the so-called unitary triangles on the complex plane. The most commonly used triangle comes from the condition,

$$V_{ud} V^*_{ub} + V_{cd} V^*_{cb} + V_{td} V^*_{tb} = 0 \quad (1.8)$$

and by dividing by $V_{cd} V^*_{cb}$, which is the best known, the triangle in Fig. 1.3 is obtained.

With measured values for the various CKM elements, the angles of the unitarity triangle, (α, β, γ) can be determined. The current state of things is seen in Fig. 1.4. At LHCb in particular a substantive effort is focused on determining the least well known of the unitarity triangle angles, γ, which is currently determined to be $(71.1^{+4.6}_{-5.3})^\circ$ [39].

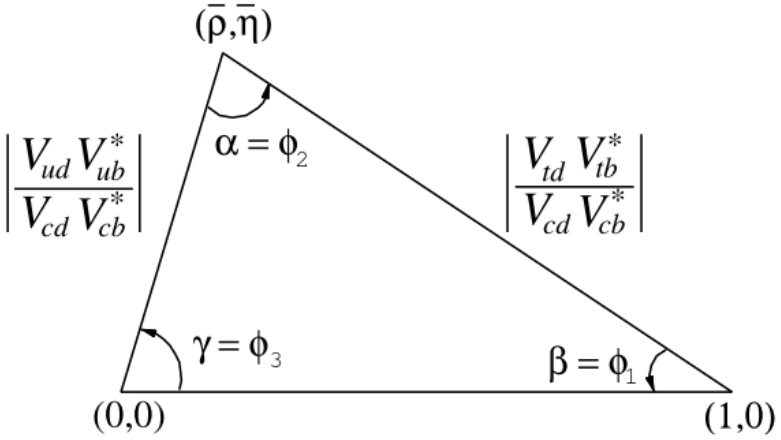


Figure 1.3: Example of the most common unitarity triangle, in the complex plane [14].

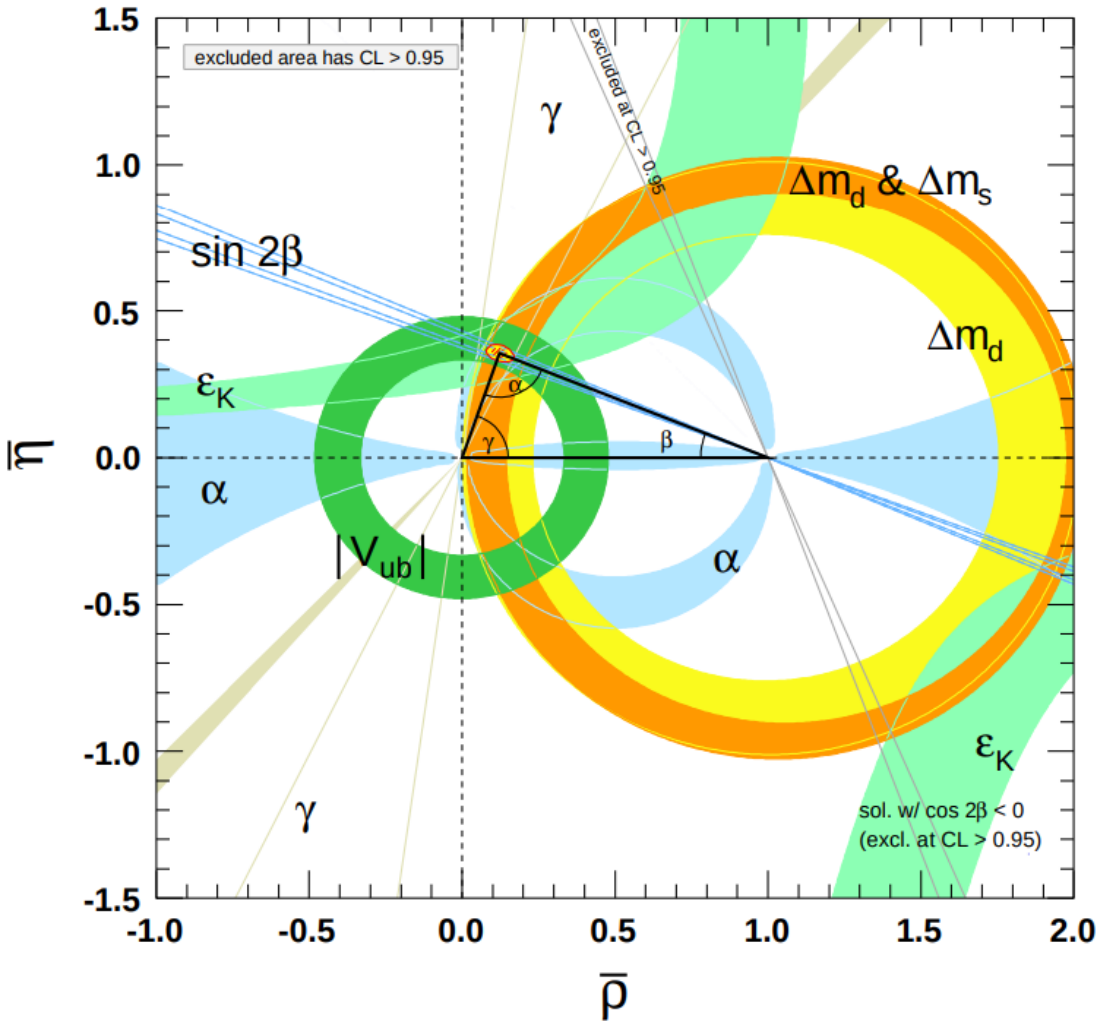


Figure 1.4: Summary of current experimental knowledge of observables related to the CKM matrix. Shaded areas have a 95% confidence limit [14].

1.3 Limitations of the SM

The Standard Model is a very successful theory and well tested from the eV to the low TeV scale. However, there are a number of unexplained questions that strongly suggest that the SM is not a complete theory of nature.

Dark Matter and Dark Energy The SM describes the fundamental fermions and bosons, however, this only makes up $\sim 5\%$ of the known universe [40]. A number of astronomical observations, such as the flat rotation curves of galaxies, motivate the existence of dark matter but the SM can provide no candidate dark matter particles. Many searches have been performed but so far no dark matter candidates have been found, see e.g. Refs. [41–45]. Dark energy makes up an even larger part of the known universe, used to explain the further acceleration of the expansion of the universe. Here as well the SM can provide no explanation.

Why three generations As previously noted, the quarks and leptons come in three generations. This is the minimum number of generations required to cause CP violation but the SM does not explain why there are only three generations.

The Higgs Mass Hierarchy Problem In the SM the Higgs boson mass is not predicted, which already hints at an incomplete theory. The hierarchy problem is concerned with why the measured Higgs mass is not close to some very large energy scale such as the Planck scale. If new physics exists at such a scale, the Higgs boson should be subject to quantum corrections pushing its mass upwards. The fact that its mass is so low suggests there is some fine-tuning which exactly cancels out these corrections. It is this fine-tuning that is one motivation for favouring supersymmetric (SUSY) models [46] as extensions to the SM.

Matter-Antimatter asymmetry When the universe was created, matter and antimatter are assumed to have been produced in equal amounts. Today the universe is dominated by matter particles. Clearly this means there must have been some interactions which violate CP symmetry. It is known that the weak interaction violates CP symmetry but the amount of CP violation observed is insufficient to account for the asymmetry. There is no fundamental reason why the strong interaction cannot also violate CP but this has not been observed and this is known as the strong CP problem.

Neutrinos In the SM neutrinos were expected to have zero mass. However, it is known from the discovery of neutrino oscillations [47–49] that neutrinos do indeed have masses, so already the SM does not correctly describe neutrinos. We also know that neutrino oscillations do not respect lepton number conservation as neutrinos may change flavour spontaneously during flight. This has motivated searches for lepton flavour violating decays at LHCb such as $B^0 \to \mu^{\pm} e^{\mp}$ [50].

Flavour The values of the CKM matrix and the equivalent lepton mixing PMNS matrix [51] are known from experiment only. The SM cannot predict the values of the elements of these matrices. Lepton flavour number conservation is not a fundamental symmetry of the SM and has just been observed empirically to be true. This coupled with the fact that this conservation law appears to be broken for neutrino oscillations motivates tests of lepton universality, which probe whether the electroweak coupling is in fact the same for all generations of leptons, disregarding their mass differences.

1.4 Rare Decays as probe of the Standard Model

There are two main ways in which collider experiments search for evidence of physics beyond the SM: direct and indirect detection.

Direct Detection Direct detection is the production of new particles and requires sufficient energy in the centre-of-mass system of the experiment, whether collider or fixed target, where they can be produced on mass-shell and consequently reconstructed from their decay products, or inferred by an enhancement in the observed cross-section. The drawback of this technique is that it is limited by the maximum collision energy a particular particle accelerator may reach. However, it has been very successful and was responsible for discovering the W, Z and Higgs bosons, as well as all of the quarks.

Indirect Detection Without having a hard limit imposed by the collision energy, indirect detection acts as an alternative through which BSM effects may be discovered. The two main approaches are by making precision measurements and the study of rare decays. Both types allow for probing of much higher energy scales than direct detection can provide, however, if BSM effects are observed there is some ambiguity as to what the cause of these effects may be. If BSM physics is detected in this way it can inform the next generation of direct detection machines.

Precise measurements aim to reduce the uncertainty on the values of measured observables of the SM to achieve an experimental precision similar to or better than that of the theoretical predictions. Discrepancies found between these can be an indication of BSM physics. An example of this type of experiment is the measurement of the anomalous magnetic dipole moment of the electron, predicted by quantum electrodynamics (QED), in this case theory and experiment agree with each other to an unprecedented precision [52]. However, the analogous quantity for the muon has been measured [53] and is in tension with the SM. If confirmed, this may suggest BSM physics is playing some role.

The study of rare decays, that is, decays which in the SM have a relatively low branching fraction, are considered a good laboratory for testing the SM to its limits. Since the SM contribution to these decays is low, any contributions from BSM physics are expected to be observable in deviations from the SM branching fractions or angular distributions of the particles. Rare decays will be a main focus of this work with FCNCs being studied in particular. FCNCs are a promising area to study as in the SM they do not occur at tree level, but instead occur at loop order, and thus have a small branching fraction. As particles within the loop are virtual this means potential BSM particles with masses much higher than direct detection methods may access can contribute, altering predicted properties of the decay.

1.5 Lepton Flavour Universality

Lepton Flavour Universality (LFU) describes the property that all (charged) leptons have the same electro-weak coupling constant in the SM, meaning that the photon and the Z and W bosons have the same coupling strength to all leptons. The only difference between the lepton couplings in the SM comes from their respective Yukawa couplings as a result of the different masses of the leptons. Therefore any deviation from LFU, once differences from the Yukawa couplings are accounted for, indicates the presence of BSM physics. Of particular interest in this work is probing LFU with FCNC decays which involve $b \to s\,\ell^+\ell^-$ quark transitions. These types of decays are rare enough that BSM physics may be disentangled from SM effects but also numerous enough given LHCb's integrated luminosity to have statistically significant samples for study.

1.5.1 Experimental status

Many tests of LFU and searches for its violation have been conducted in various experimental settings over the years. The main focus of this work is related to LFU searches in the B sector, discussed further in Sec. 1.6.1, however a very broad overview will be given here.

Z decays Measurements of the ratios of the partial widths of the decays of Z bosons to pairs of charged leptons have been performed at both LEP [54] and ATLAS [55] with both measurements supporting LFU in Z decays.

W decays Measurements of the ratio of the branching fractions of W bosons decaying to an electron and a $\overline{\nu}_e$ or a μ and a $\overline{\nu}_\mu$ have been performed by LEP [56], CDF [57] LHCb [58] and ATLAS [55] experiments with all measurements supporting LFU in W decays. Measurements involving W boson decays to a τ and a $\overline{\nu}_\tau$ are dominated by LEP [56], the precision of these measurements is about an order of magnitude lower for W decays to the first two generations of leptons and is consistent with the SM at the 2.6σ level.

Psuedoscalar meson decays Measurements of the ratios of the decays of kaons and pions to an electron and anti-electron neutrino or a muon and anti-muon neutrino are dominated by the NA62 [59] and PIENU [60] results respectively. Both results support LFU in kaon and pion decays, however, the result for pion decays is still an order of magnitude away from the SM prediction. Measurements of a similar ratio [39] for the D_s^- meson involving muonic and tauonic decay modes also support LFU.

Purely leptonic decays Using tau decays to purely leptonic final states following the analysis of Ref. [61] ratios of coupling constants of the various leptons in flavour changing charged current interactions can be determined. All permutations of the ratios within the three lepton generations were consistent with unity.

J/ψ decays The ratio of the partial widths of decays from Ref. [14] of the J/ψ meson to pairs of electrons or muons are found to be consistent with LFU. This is exploited by the LHCb measurements of R_K and R_{K^*} ratios, discussed in Sec. 1.6.2.

1.5.2 Lepton Universality in b-quark decays

Since this work will focus on $b \to s\,\ell^+\ell^-$ quark transitions it is useful to separate effects coming from the b quark mass scale, m_b, and below from those scales above this such as the electroweak scale, the top quark mass and the Higgs mass. This can be done using an effective field theory approach [62, 63] which is similar to the Fermi theory of weak interactions [64], allowing a low energy equivalent theory to be constructed. In the Fermi theory the short distance propagation of the W boson is replaced by a four-fermion contact interaction and for $b \to s\,\ell^+\ell^-$ transitions contributions from scales above m_b are dealt with in a similar way. Contributions from lower scales are allowed to propagate and constitute the long distance effects. The result is to have coefficients, known as Wilson coefficients, which encapsulate the short distance effects above a set energy scale, μ, and operators describing propagation of fields over distances larger than $1/\mu$. Here, the renormalisation scale $\mu = m_b$ but in general it may be set to whatever is convenient for a given situation.

For $b \to s\,\ell^+\ell^-$ decays, which are FCNCs, the effective Hamiltonian is of the form [65]

$$\mathcal{H}_{\text{eff}}(b \to s\,\ell^+\ell^-) = -\frac{4G_F}{\sqrt{2}} V_{tb}V_{ts}^* \Sigma_i \mathcal{C}_i \mathcal{O}_i, \tag{1.9}$$

where the $\mathcal{O}_i$ are the various four-fermion operators describing the fields propagating over the long distances and the $\mathcal{C}_i$ are the Wilson coefficients which describe the short distance effects from fields such as the W, Z, H and t. The Wilson coefficients may be calculated peturbatively. In the SM due to lepton universality the values of these coefficients should be identical for all three lepton generations. The CKM factor $V_{tb}V_{ts}^*$ is from the dominant loop contribution, whilst the factor $V_{ub}V_{us}^*$ is much smaller as seen in Eq. 1.4 and is hence neglected.

There are three dominant operators contained in Eq. 1.9 in the SM:

$$\mathcal{O}_7 = \frac{e}{16\pi^2} m_b (\bar{s}\sigma_{\mu\nu} P_R b) F^{\mu\nu} \tag{1.10}$$

$$\mathcal{O}_{9l} = \frac{e^2}{16\pi^2} (\bar{s}\gamma_\mu P_L b)(\bar{l}\gamma^\mu l) \tag{1.11}$$

$$\mathcal{O}_{10l} = \frac{e^2}{16\pi^2} (\bar{s}\gamma_\mu P_L b)(\bar{l}\gamma^\mu\gamma_5 l), \tag{1.12}$$

where $P_{L,R} = (1 \mp \gamma_5)/2$, which projects left or right handed chirality, is included in $\mathcal{O}_{9l}$ and $\mathcal{O}_{10l}$ because the weak interaction acts on left handed particles. The index l denotes the lepton flavour involved. The electromagnetic field tensor is denoted $F^{\mu\nu}$ whilst $\sigma_{\mu\nu} = i/2[\gamma_\mu, \gamma_\nu]$ and e is the coupling constant of QED. Note that Eq. 1.10 does not contain any leptonic terms since this interaction is mediated by the photon which is allowed to propagate as a light field; in contrast, Eqs. 1.11 and 1.12 are full 4 fermion interactions analogous to Fermi theory as they would proceed via a W boson interaction, which is a heavy field for this choice of μ. These interactions are illustrated in Fig. 1.5.

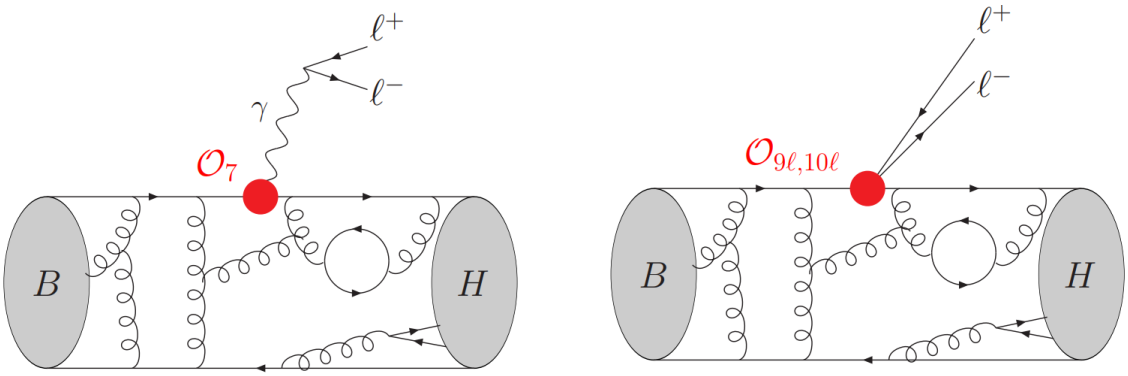


Figure 1.5: Illustrations of $b \to s\,\ell^+\ell^-$ transitions with the effective Hamiltonian method [65]. The red dot denotes a local operator. Transition corresponding to: (left) $\mathcal{O}_7$ with a photon propagator; (right) $\mathcal{O}_{9l,10l}$ which is a four fermion contact interaction.

The advantage of this approach is that BSM effects can be included fairly simply: the values of the Wilson coefficients can be shifted from their SM values and additional operators, which are either absent from the SM or heavily suppressed, can be enhanced. In particular for LFU it could be envisaged that the Wilson coefficients for each of the lepton generations are not identical, which would generate lepton flavour violating effects.

A further complication to this approach comes from soft gluons, which correspond to long distance QCD effects. The short distance QCD effects are resummed in the Wilson coefficients but the long distance effects are responsible for hadronisation from the initial B hadron into the final state hadrons; the decay amplitude is generally written as $\mathcal{C}_i \times \langle F|\mathcal{O}_i|I\rangle$ where I and F are the initial and final state respectively. Expressions of this type can be computed and generally include what are known as form factors, which are scalar quantities that depend on some Lorentz invariant kinematic quantity such as the squared dilepton invariant mass, q^2. However, in the FCNC case of $b \to s\,\ell^+\ell^-$ decays the lepton pair may come from the annihilation

of a $c\bar{c}$ pair (charm loop). An example of how a charm loop may contribute can be seen in Fig. 1.6. In the SM the J/ψ and $\psi(2S)$ mesons have large amplitudes, so that cases involving the charm loops dominate for q^2 close to these meson masses and therefore these regions are generally excluded when studying rare, non-resonant decays to inclusive dilepton final states. These regions are often useful as control channels due to their large amplitudes and having the same final state particles as the signal channel.

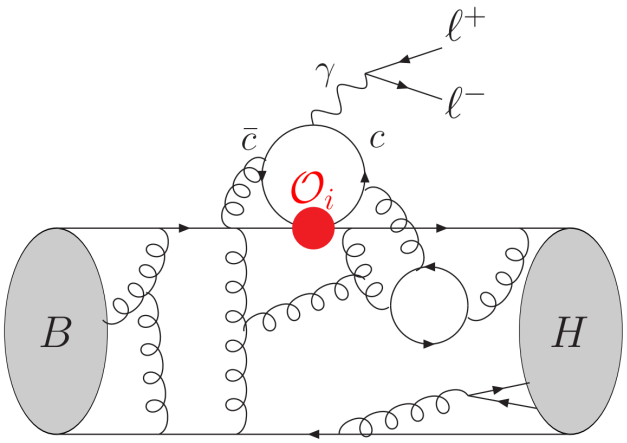


Figure 1.6: Illustration of how a charm loop may contribute to a $b \to s\,\ell^+\ell^-$ transition with the effective Hamiltonian method [65]. The red point denotes a four quark operator in place of a W exchange in the full theory.

In order to reduce dependence on form factors, which must be calculated non-perturbatively and thus are subject to large theoretical uncertainties, it was proposed to measure ratios of the form [66]

$$R_X = \frac{\int_{q^2_{\min}}^{q^2_{\max}} \frac{\mathrm{d}\Gamma(B^+\to X\mu^+\mu^-)}{\mathrm{d}q^2}\mathrm{d}q^2}{\int_{q^2_{\min}}^{q^2_{\max}} \frac{\mathrm{d}\Gamma(B^+\to Xe^+e^-)}{\mathrm{d}q^2}\mathrm{d}q^2}, \tag{1.13}$$

where X represents a hadron that contains an s quark and Γ represents the decay rate which is integrated over a specific range of q^2. In this work the R_K and R_{K^*} ratios will be considered in particular. The advantage of these observables is that the hadronic form factors largely cancel out making this very clean from a theoretical standpoint. In the SM it is expected that these ratios should have values close to unity, after taking into account both the accessible q^2 regions in which measurements are performed and lepton mass effects, and therefore any significant departure from this is expected to be a signal of BSM physics.

1.6 LFU measurements in the B sector

1.6.1 e^+e^- and p-p colliders

The motivation for continued study of LFU comes primarily from measurements obtained with B meson decays, the so-called B sector. Generally outside the B sector experimental results have agreed very well with SM predictions, with the exception of LFU tests involving W boson decays to tau leptons at LEP where a 2.6σ deviation was observed. A review of the current measurements outside of the B sector can be found in Ref. [65].

Within the B sector there are two types of initial states contributing to results: experiments at e^+e^- colliders (Belle, Belle-II and BaBar), and those at the LHC p-p collider, dominated by LHCb. The e^+e^- colliders generally provide a cleaner experimental environment, however, LHCb is able to reach higher collision energies. Also at the LHC from 2018 ATLAS and CMS used a new B trigger in order to provide LFU measurements but their precision is not yet known.

The particle identification performance necessary to separate kaons and pions and the reconstructed B mass resolution from charged hadrons are quite similar between the two types of experiments, however, muon and electron reconstruction presents the most important differences for LFU measurements. The e^+e^- colliders operate at a centre of mass energy around the $\Upsilon(4S)$ meson (10.58 GeV), which decays to B-$\bar{B}$ pairs more than 96% of the time. In contrast, the LHC operates on the TeV scale so the decay products of B mesons produced there are much more energetic. This means that for decays involving electrons LHC experiments experience much more bremsstrahlung than e^+e^- colliders. Both types of experiments employ a bremsstrahlung recovery algorithm to offset this effect. As a result of this difference, at Belle the reconstructed resolution of $J/\psi\,(ee)$ events is only slightly wider than in $J/\psi\,(\mu\mu)$ events. However, reconstructing $B^0 \to K^{*0}\mu^+\mu^-$ and $B^0 \to K^{*0}e^+e^-$ at LHCb shows that the decay involving electrons has a much wider resolution than the decay involving muons and also has a long radiative tail. Bremsstrahlung at LHCb is discussed in detail in Sec. 2.2.4. Furthermore, $B^+ \to K^+\ell^+\ell^-$ decays the e^+e^- colliders, judging by their reported yields, result in a roughly equal number of events split between the muon and electron modes. At LHCb the recorded yield for the decay with muons is more than seven times larger than the same with electrons. These discrepancies motivate using double ratio-based observables in LHCb analyses, such as the analysis presented in this work which will be outlined

Table 1.1: Comparison of LHCb and Belle-II R_X ratio measurements in similar q^2 windows. The first uncertainty is statistical and the second is systematic.

Experiment (year)	Hadron	q^2 (GeV/c^2)	Value	Ref.
LHCb (2014)	K^+	1.0–6.0	$0.745^{+0.090}_{-0.074} \pm 0.036$	[68]
LHCb (2019)	K^+	1.1–6.0	$0.846^{+0.060\ +0.016}_{-0.054\ -0.014}$	[69]
Belle-II (2019)	K^+	1.0–6.0	$1.31^{+0.34}_{-0.31} \pm 0.07$	[70]
LHCb (2017)	K^{*0}	0.045–1.1	$0.66^{+0.11}_{-0.03} \pm 0.05$	[71]
Belle-II (2019)	K^{*0}	0.045–1.1	$0.46^{+0.55}_{-0.27} \pm 0.07$	[72]
LHCb (2017)	K^{*0}	1.1–6.0	$0.69^{+0.11}_{-0.07} \pm 0.05$	[71]
Belle-II (2019)	K^{*0}	1.1–6.0	$1.06^{+0.63}_{-0.38} \pm 0.13$	[72]

in Sec. 3.

1.6.2 $b \to s\,\ell^+\ell^-$ transitions

A survey of the current experimental status of LFU in $b \to s\,\ell^+\ell^-$ transitions will follow, as they are the focus of this work. A recent review of current measurements in other areas of the B sector can be found in Ref. [65]. An area in which both types of experiment have contributed results is in taking ratios of branching fractions of decays differing only by the pair of final state leptons present such as in Eq. 1.13.

Although there are differences between the two collider types mentioned above, there is some common ground between the two experimental procedures. The areas where the J/ψ and $\psi(2S)$ resonances dominate in q^2 are vetoed in both cases as these areas have very large SM contributions. Also, the K^{*0} is only reconstructed in a region where the K^{*0} (892) resonance is dominant. The main area of focus so far has been on R_K and R_{K^*}, corresponding to $B^+ \to K^+\ell^+\ell^-$ and $B^0 \to K^{*0}\ell^+\ell^-$ decays respectively, due to their relatively large branching fractions. However, LHCb has also recently published a measurement for R_{pK} [67], corresponding to $\Lambda_b^0 \to pK\ell^+\ell^-$ decays. Belle and Babar took measurements in different q^2 ranges than LHCb, current Belle-II measurements have been made in q^2 ranges comparable to both types of experiment and the results can be seen in Tabs. 1.1 and 1.2.

1.6.3 Standard model predictions: R_K and R_{K^*}

Accurate theoretical predictions of R_X ratios can be made due to form factor cancellations discussed in Sec. 1.5. However, this is only true if there are no large LFU violating effects. If there are then the cancellation of the form factors may

Table 1.2: Comparison of Belle, BaBar and Belle-II R_X measurements. The first uncertainty is statistical and the second is systematic.

Experiment (year)	Hadron	q^2 (GeV/c^2)	Value	Ref.
Belle (2009)	K	0–kin. endpoint	$1.03 \pm 0.19 \pm 0.06$	[73]
Belle-II (2019)	K^+	0.1–kin. endpoint	$1.04^{+0.16}_{-0.15} \pm 0.06$	[72]
BaBar (2012)	K	0.1–8.12	$0.74^{+0.40}_{-0.31} \pm 0.06$	[74]
Belle-II (2019)	K^+	0.1–4.0	$0.92^{+0.27}_{-0.24} \pm 0.05$	[72]
Belle-II (2019)	K	4.0–8.12	$1.22^{+0.42}_{-0.37} \pm 0.07$	[72]
BaBar (2012)	K	>10.11	$1.43^{+0.65}_{-0.44} \pm 0.12$	[74]
Belle-II (2019)	K^+	>14.18	$1.08^{+0.30}_{-0.27} \pm 0.06$	[72]
Belle (2009)	K^*	0–kin. endpoint	$0.83 \pm 0.17 \pm 0.08$	[73]
Belle-II (2019)	K^{*0}	0.045–kin. endpoint	$1.12 \pm^{+0.27}_{-0.21} \pm 0.09$	[72]
Babar (2012)	K^*	0.10–8.12	$1.06^{+0.48}_{-0.33} \pm 0.08$	[74]
Belle-II (2019)	K^{*0}	0.1–8.0	$0.86^{+0.33}_{-0.24} \pm 0.08$	[72]
Babar (2012)	K^*	>10.11	$1.18^{+0.55}_{-0.37} \pm 0.11$	[74]
Belle-II (2019)	K^{*0}	15–19	$1.12^{+0.61}_{-0.36} \pm 0.10$	[72]

not be as complete as in the SM case. The LFU violating effects may come from BSM physics or lepton mass effects, but the latter would only occur at low-q^2 where m_l^2/q^2 (m_l denotes the mass of the lepton) is not small compared to unity, typically below a q^2 value of $1\,\mathrm{GeV}/c^2$. This effect in particular would affect the first bin in LHCb's R_{K^*} measurement and leads to larger theoretical uncertainties in this region. It has also been established that radiative corrections one might expect in these calculations are small and under control [75]. A selection of SM predictions can be found in Tab. 1.3, most of which are compatible with unity except in the lowest q^2 region for R_{K^*}.

With comparison to the experimental results, the latest R_K and R_{K^*} measurements at LHCb are found to be consistent with the SM at levels of 2.5σ and 2.1–2.5σ, respectively, leaving open the possibility of new physics. The Belle and BaBar measurements are all in agreement with the SM, as are the Belle-II results, however, all of these measurements lack the statistical precision of the LHCb results. These tensions motivate the main body of this work, which is probing lepton universality with the R_K and R_{K^*} ratios simultaneously using the full LHCb Run 1 and Run 2 datasets.

Table 1.3: Standard Model predictions for the LFU ratios R_K and R_{K^*}

q^2 region (GeV/c^2)	Ref. [76] (2018)	Ref. [77] (2017)	Ref. [78] (2017)	Ref. [75] (2016)
R_K $(1.0 < q^2 < 6.0)$	1.00±0.01	$1.0004^{+0.0008}_{-0.0007}$	-	1.000±0.010
R_{K^*} $(0.045 < q^2 < 1.1)$	0.92±0.02	$0.920^{+0.007}_{-0.006}$	0.9259±0.0041	0.906±0.028
R_{K^*} $(1.1 < q^2 < 6.0)$	1.00±0.01	0.996±0.002	0.9965±0.0006	1.000±0.010

Chapter 2

The Large Hadron Collider and the LHCb Detector

2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [79] is a two ring hadron accelerator and collider located at CERN in Geneva, Switzerland. The LHC is located in a 27 km tunnel at a depth varying 45–175 m around its circumference. The collider is capable of colliding proton-proton and also heavy ion beams, with the two ring system allowing the beams to collide at four designated interaction points along its circumference. Around the four interaction points are the four largest experiments associated with the LHC, the two general purpose detectors, ATLAS and CMS, and two detectors looking at more specific areas of particle physics, LHCb and ALICE as can be seen in Fig. 2.1.

The protons used within the LHC originate from bottles of hydrogen gas. The hydrogen is stripped of its electrons using an electric field leaving behind protons which are then accelerated using Linac2 up to 50 MeV. The protons are then injected into the Proton Synchrotron Booster (PSB) and accelerated to 1.4 GeV, from here the protons are further accelerated to 25 GeV in the Proton Synchrotron (PS) and then to 450 GeV in the Super Proton Synchrotron (SPS). At this stage the beam is injected into the two rings of the LHC where the protons are accelerated to their nominal energy. In 2018 operation the collision energy was 13 TeV so each proton beam was accelerated to an energy of 6.5 TeV. With the design specifications of the LHC a maximum of 2808 bunches consisting of 10^{11} protons can be circulated per

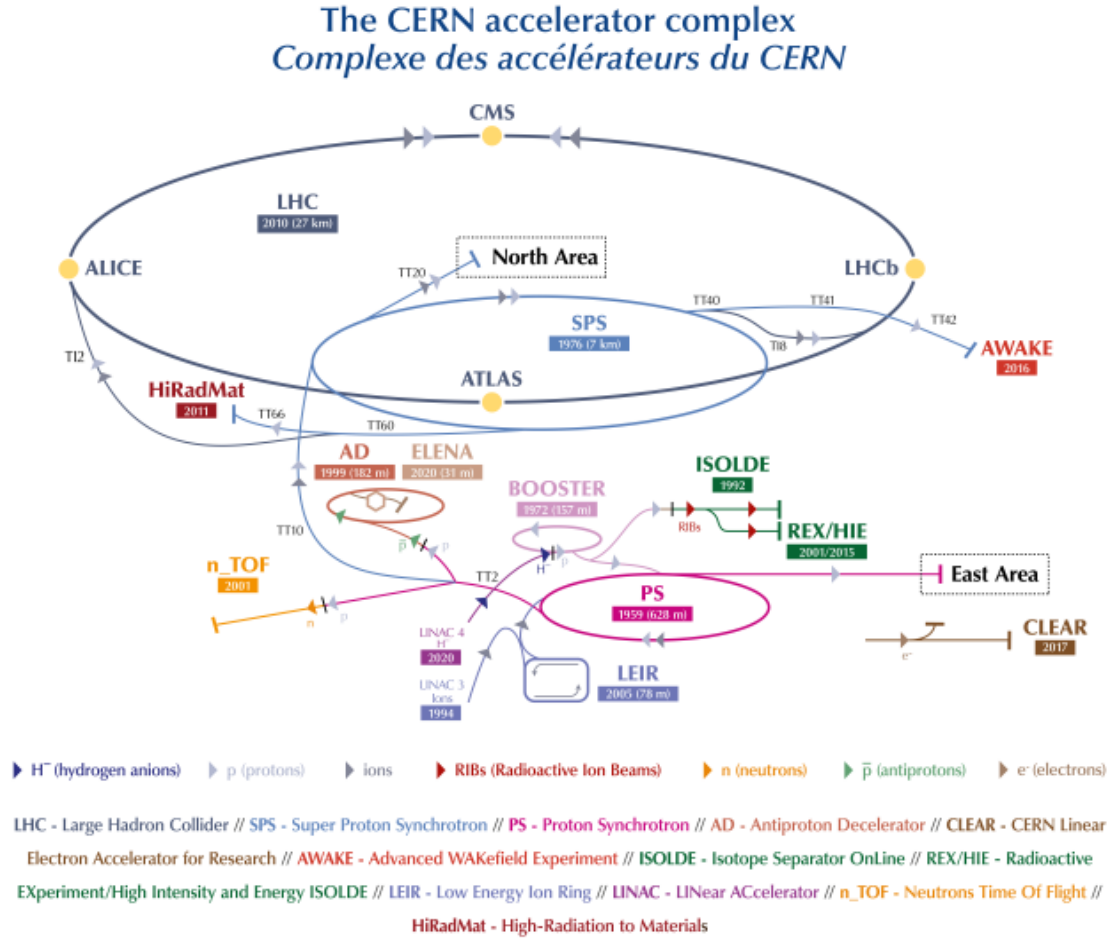


Figure 2.1: The accelerator complex of the LHC and associated experiments [80].

proton beam with a typical bunch crossing every 25 ns, corresponding to a collision rate of 40 MHz at each interaction point. The proton beams are bent around the LHC tunnel with 8.3T superconducting dipole magnets, quadrupole magnets are used to focus the beams and higher order multipole magnets are used for further small corrections.

The design centre of mass energy ($\sqrt{s}$) and instantaneous luminosity of the LHC are 14 TeV and $10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ respectively. In 2011 $\sqrt{s}$ at the LHC was 7 TeV corresponding to 3.5 TeV proton beams, in 2012 this was increased to 8 TeV and between 2015 and 2018 this was further increased to 13 TeV. During 2011–2012 the peak instantaneous luminosity reached $1.7 \times 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ and in 2015–2018 this value was $2.1 \times 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, the latter value being more than twice the design luminosity. Currently the LHC is in its third long shutdown with many detectors undergoing upgrades in anticipation of the third run of the LHC in 2022.

2.2 The LHCb Detector

The LHCb detector [81, 82] was designed specifically to look for CP-violation in predominantly B meson decays and now also contributes in many areas including D meson decays, rare decays and hadron spectroscopy. The detector is a single arm forward spectrometer, it is designed in this manner as at high energies proton-proton interactions produce highly boosted virtual gluons, when the resulting gluon interactions produce $b\bar{b}$ pairs they will remain at small angles due to the large boost. The detector is sensitive to an angular range of 10–300 mrad which corresponds to a pseudorapidity, η, in the range of $2 < \eta < 5$. The pseudorapidity is defined as

$$\eta = -\ln(\tan(\theta/2)). \tag{2.1}$$

In Eq. 2.1 θ is the angle between the momentum of the particle and the beam direction (the z-axis in Fig. 2.2).

Run 1 took place during 2011 and 2012 recording a combined integrated luminosity of $3\,\text{fb}^{-1}$. In 2011 when the centre of mass energy was 7 TeV the cross-section for $b\bar{b}$ production was measured to be 295 μb which results in a production cross-section for b flavoured hadrons in LHCb's η range of $72 \pm 7\,\mu$b [83]. Run 2 took place from 2015 to 2018, recording a combined integrated luminosity of $6\,\text{fb}^{-1}$. In Run 2 when the centre of mass energy was 13 TeV the cross-section for $b\bar{b}$ production was measured to be 600 μb which results in a production cross-section for b flavoured hadrons in LHCb's η range of $154 \pm 14\,\mu$b [83].

The LHCb detector itself consists of several sub-detectors visible in Fig. 2.2 which will be described in more detail in the following sections. The sub-detectors fall into two main categories, tracking and particle identification (PID). The tracking system handles vertex detection and momentum measurements. The minimum distance of a track to the primary vertex (PV), called the impact parameter (IP), can be measured with a resolution of $(15 + 29/p_T)\mu$m. The measurements of the momentum of charged particles have relative uncertainties of 0.5% at low momentum increasing to 2% at 200 GeV. The PID system consists of Ring Imaging Cherenkov (RICH) detectors [84], a calorimeter system and a muon system. The RICH detectors provide PID information for different types of charged particles and are crucial for differentiating between protons, kaons and pions. The calorimeter system provides further PID information for hadrons, electrons and photons. This system is also essential for identifying neutral particles which will deposit energy here but otherwise

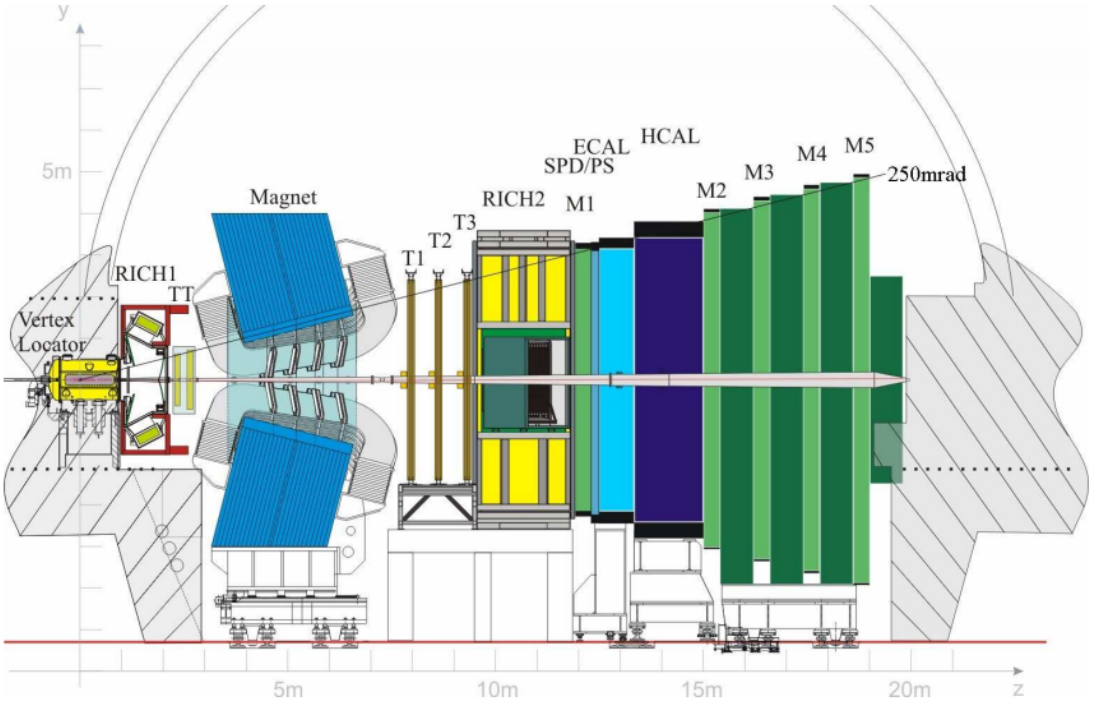


Figure 2.2: A side view of the LHCb detector [81].

cannot be tracked in the detector. The muon system is located towards the end of the detector as muons are the most penetrating particles commonly dealt with.

At nominal operation the LHC provides an event rate of 40 MHz. In order to reduce this to a manageable rate of 1 MHz a two stage trigger is employed. The first stage, implemented in hardware, is primarily to limit the event rate by looking for evidence of interesting activity within the detector, while the second stage performs an online reconstruction and verifies that tight quality and momentum requirements of tracks are met. Once an event has passed the trigger system it is written to long-term storage.

2.2.1 Luminosity levelling

One key difference for LHCb compared to the general purpose detectors is that it runs at a lower instantaneous luminosity, $3.5 \times 10^{32}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ in 2011 and $4 \times 10^{32}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ for all other years of running. This is around two orders of magnitude lower than the LHC is able to provide but twice the design luminosity of LHCb ($2 \times 10^{32}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$). Since this luminosity is much lower than the LHC is capable of delivering at the interaction point, it also means that the luminosity during the course of a fill can be kept constant whereas if you take the maximum luminosity

the LHC provides this will decay over time, as is illustrated in Fig. 2.3. In order to run at this lower luminosity constantly a technique called luminosity levelling is employed. At the beginning of a run the beams are displaced slightly from head on collisions whilst keeping the same crossing angle. As the luminosity begins to drop from the desired operating luminosity the beams can be brought closer together again.

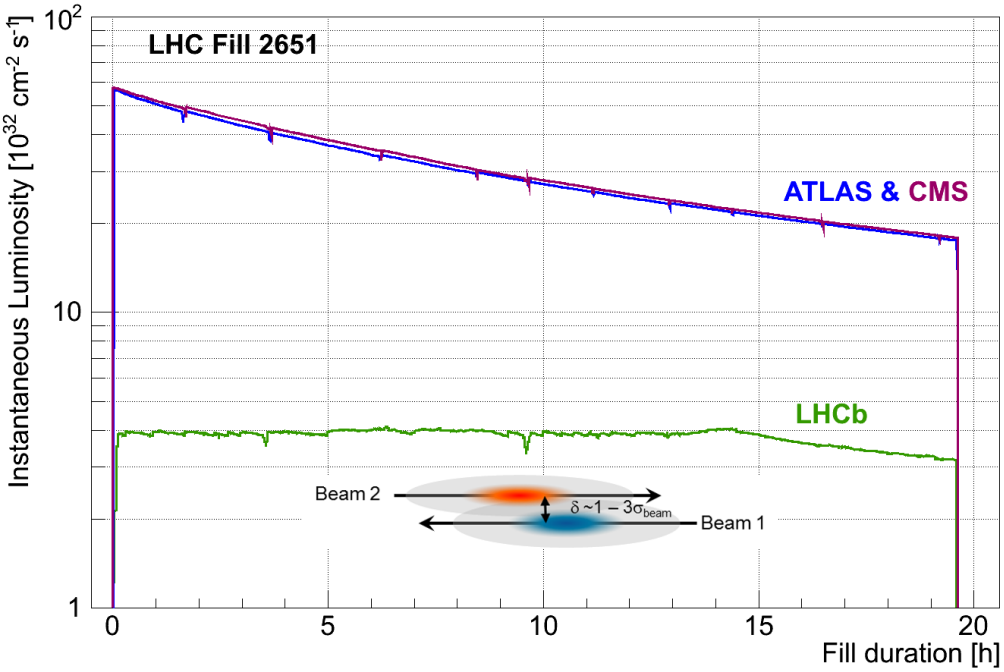


Figure 2.3: A plot showing luminosity levelling in LHCb compared to the luminosity of ATLAS and CMS. It can be seen the luminosity of LHCb is kept stable for roughly 15 hours [82].

Using luminosity levelling minimises issues arising from luminosity decay and allows the same trigger configuration to be kept during a fill of the LHC. It also means that the detector occupancy remains roughly constant over the course of a fill reducing systematics associated with changes in performance due to occupancy. Another advantage of this procedure is that the number of interactions per bunch crossing, or pile-up, can be regulated. By keeping the luminosity relatively low the pile-up remains small, which is important for LHCb because bunch crossings with many primary vertices (PVs) reduce the efficiency for detecting secondary vertices (SVs), which are crucial to detect b hadrons.

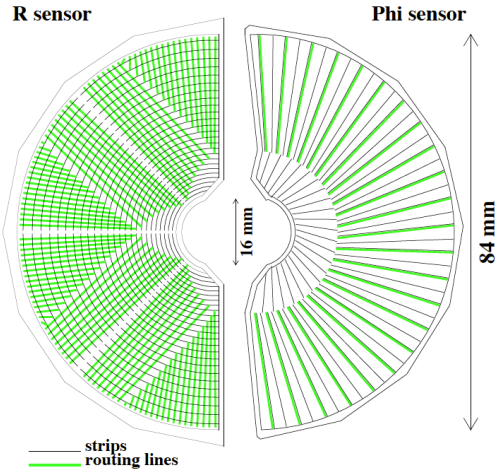
2.2.2 Tracking

The tracking system consists of the Vertex Locator (VELO) [85] and four planar tracking stations, the Tracker Turicensis (TT) [86] upstream of the magnet and tracking stations T1-T3 downstream of the magnet. The VELO and TT use only silicon microstrip detectors, these are also used in T1-T3 but only in the region close to the beampipe known as the Inner Tracker (IT) [87]. The outer regions of T1-T3 use straw-tubes and these are known as the Outer Tracker (OT) [88].

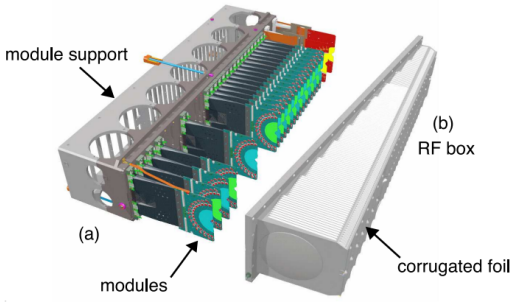
VELO

To be able to reconstruct vertices is of paramount importance for LHCb and the VELO facilitates this with precise measurements of track coordinates using silicon strip detector technology. This allows for production and decay vertices to be located, particle lifetimes to be measured and for IPs of particles used for flavour tagging to be determined. In order to cover the angular acceptance of the downstream detectors the VELO covers a pseudorapidity range of $1.6 < \eta < 4.9$ and can detect particles emerging from primary vertices in the range of $|z| < 10.6\,\text{cm}$.

The VELO is split into two halves and to cover the required azimuthal acceptance these must overlap. The minimum radial distance from the beamline to the sensitive area is 7mm. This distance is smaller than the beam aperture while the proton beams are being ramped to full energy and therefore the sensors must be retractable. A cylindrical geometry (r, ϕ, z) is used to reconstruct the tracks as this is the most efficient in software. The r and ϕ coordinates are determined using dedicated silicon sensors arranged along the beam direction, therefore the z-coordinate is given by the location of the sensor. The R-sensors are concentric semi-circles with their centres located on the beam axis, split into 45 degree sections to keep occupancy to a minimum. The ϕ-sensors read out the orthogonal coordinate to the R-sensors having strips which run radially but subdivided into two regions called the inner region and the outer region, which begins at $r = 17.25$mm, to keep the occupancy at an acceptable level. The sensor layout can be seen in Fig. 2.4a. In each VELO half there are 21 modules which contain both types of detectors. The modules are subsequently mounted in RF-boxes which are enclosed in a secondary vacuum to the machine vacuum by an RF-foil which faces the beam as seen in Fig. 2.4b. The purpose of the RF-boxes and foils is threefold: to provide shielding from RF pick up due to the beams; to guide wakefields in a way which prevents impedance disruptions to the beams, and to protect the LHC vacuum from any outgassing from the VELO modules which may occur.



(a) R and ϕ sensor layouts [82].



(b) Modules mounted in RF-boxes and protected by the RF-foil [81].

Figure 2.4: The VELO sub-detector.

Studies performed during RUN 1 determined that the track finding efficiency of the VELO is generally greater than 98%. A PV resolution of 13 µm in the transverse plane and 71 µm along the beam axis can be expected for vertices with 25 tracks and an IP resolution of less than 35 µm can be achieved for particles with transverse momenta above 1 GeV/c [89].

Silicon Tracker and Outer Tracker

The Silicon Tracker (ST) consists of the TT and the IT, both of which use silicon microstrip technology. The TT is 150 cm wide and 130 cm high, has an active area of roughly 8.4 m^2 and covers the full angular acceptance of the experiment. As it is located upstream of the magnet it can reconstruct tracks corresponding to low momentum particles which are lost in the downstream acceptance. The IT is a 120 cm wide and 40 cm high cross shaped region in the centre of each of the T1-T3 tracking stations and has a total active area of 4 m^2. It has a much higher granularity than the OT as a greater particle flux is present closer to the beamline. Each of the ST stations have four layers of detectors in a (x-u-v-x) arrangement meaning they have vertical strips in the first and last layers and the middle two layers (u and v) are rotated by $\pm 5°$ respectively. This feature allows for the y-coordinate to be determined. Using RUN 1 data it was found that the TT and IT both have hit efficiencies of well over 99% and in 2011 the hit resolutions were 52.6 µm and 50.3 µm for the TT and IT respectively [82].

The OT is a gaseous straw tube detector [88] which tracks charged particles over a large acceptance area and provides measurements of their momenta. It is arranged

as an array of gas-tight straw tube modules, these modules also being placed in a (x-u-v-x) arrangement. The area covered is around 30 m^2 with 12 double layers of straw tubes. The gas used is a 70:30 mixture of argon and carbon dioxide as this yields a sufficiently fast drift time of 50 ns and a good drift-coordinate resolution of 200 µm in the x-direction. Studies using RUN 1 data estimated that the hit efficiency in the central half of the straw tube was 99.2% and hit resolution was 200 µm with no signs of ageing due to radiation detected [90]. In RUN 2 a real time calibration of the time alignment of the detector and the alignment of the layers composing detector modules was introduced and studies with 2015 and 2016 data determined this increased the resolution on the drift time and the position measurements of the OT by 20% [91].

LHCb's Magnet

LHCb's Magnet [92] is essential in providing momentum measurements for charged particles and has a bending power of 4 Tm (over a 10m track length). The magnet covers the full acceptance of LHCb and consists of two coils placed symmetrically and an iron yoke. A nominal current of 5.85kA is used. In order to achieve the design momentum resolution for charged particles the magnetic field was required to be known with a relative precision of a few times 10^{-4}. Using an array of hall probes the precision of the magnetic field measurement is able to be determined to a relative precision of 4×10^{-4} making the desired momentum resolution achievable. To control systematic effects coming from the detector the polarity of the magnet is often changed for data taking runs with the aim of collecting a roughly equal amount of data from both polarities (referred to as MagUp and MagDown).

Track Types in LHCb

Tracks in LHCb are classified into types depending on the sub-detectors in which hits are recorded. The classifications of tracks, summarised in Fig. 2.5, are:

- **Long Tracks** - Hits detected in the VELO and T stations and may also have hits detected in the TT.
- **Downstream Tracks** - Hits detected in the TT and T stations but not in the VELO.
- **VELO Tracks** - Hits detected only in the VELO.
- **Upstream Tracks** - Hits detected in only the VELO and the TT.

- **T Tracks** - Hits detected in one or more of the T stations only.

For physics analyses long and downstream tracks are used, long tracks are preferred if possible as they contain more information but for long lived particles downstream tracks may be the best tracks available. The other types of tracks are mainly used for sub-detector studies and calibrations.

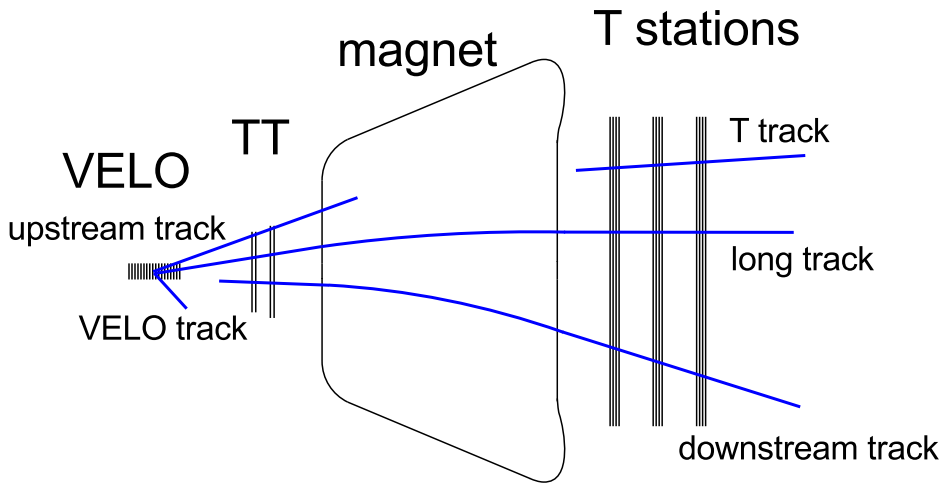


Figure 2.5: Diagram showing the different type of track classifications in LHCb, with relevant detectors labelled [93].

2.2.3 RICH

In LHCb distinguishing between pions and kaons is essential as they are often key decay products of many physics analyses. The RICH system [94] provides this distinction and consists of two detectors providing PID over a momentum range of 2–100 GeV/c and covering the whole angular acceptance of the experiment. As low momentum particles are likely to be deflected beyond the downstream acceptance, RICH 1 is placed upstream of the magnet and is used to identify particles with momenta of 2 to 40 GeV/c. RICH 1 covers an angular acceptance of 25–300 mrad. RICH 2 is downstream of the magnet and covers the higher momentum tracks, less affected by the magnet, with an angular acceptance of 12–120 mrad.

The RICH detectors make use of the Cherenkov effect [95] to detect charged particles. The detectors are filled with gas (C_4F_{10} in RICH 1 and CF_4 in RICH 2) and when charged particles travelling faster than the speed of light in the gas pass through the detector, Cherenkov photons are produced. The angle, θ, at which Cherenkov photons are produced depends on the refractive index, n, of the medium and the

velocity, v, of the particle expressed as

$$\cos(\theta) = \frac{1}{\beta n}, \tag{2.2}$$

where β is v/c. Since the refractive index is approximately constant the emission angle of the photons contains information about the velocity of the particle traversing the detector. Due to the mass differences between the various particles each starts to emit Cherenkov light at different momentum thresholds leading to the bands which can be seen in Fig. 2.6. In RUN 1 as well as C_4F_{10} gas, there was also aerogel in RICH 1. The aerogel was more performant in the low momentum region complimenting the gas that was present. However in RUN 2 the aerogel was removed as in a higher multiplicity environment its ability to provide PID for kaons below the C_4F_{10} momentum threshold was compromised by the sheer number of photons in RICH 1. This aerogel was also blocking a portion of Cherenkov photons from between the entrance window and the aerogel location as shown in Fig. 2.7. The removal of the aerogel also reduced the time taken for RICH reconstruction as it reduced the number of photon candidates associated with tracks by more than a factor of two.

Once Cherenkov photons are produced they are focused into ring images using a combination of flat and spherical mirrors which reflect the image out of the spectrometer acceptance. This reduces the material budget in the active portion of the detector. Hybrid Photon Detectors (HPDs) are used to detect Cherenkov photons in the wavelength range 200–600 nm. The HPD is a vacuum photon detector in which a Cherenkov photon produces a photoelectron which is accelerated onto a silicon detector. A total of 484 HPDs are used in total by both detectors. As protection the HPDs are surrounded by iron shields and placed in MuMetal cylinders which allow them to operate in magnetic fields of up to 50mT.

2.2.4 Calorimetry

The calorimeter system [96] is an important input to the hardware trigger providing it with transverse energy measurements of hadrons, electrons and photons. It also provides PID information for electrons, photons and hadrons (including neutrals) as well as their energies and positions. The system consists of four detectors, the first an incoming particle will see is the Scintillating Pad Detector (SPD) which is used to detect charged particles. This is followed by a thin sheet of lead designed to convert photons to electrons. After this is the Pre-Shower Detector (PS) and the

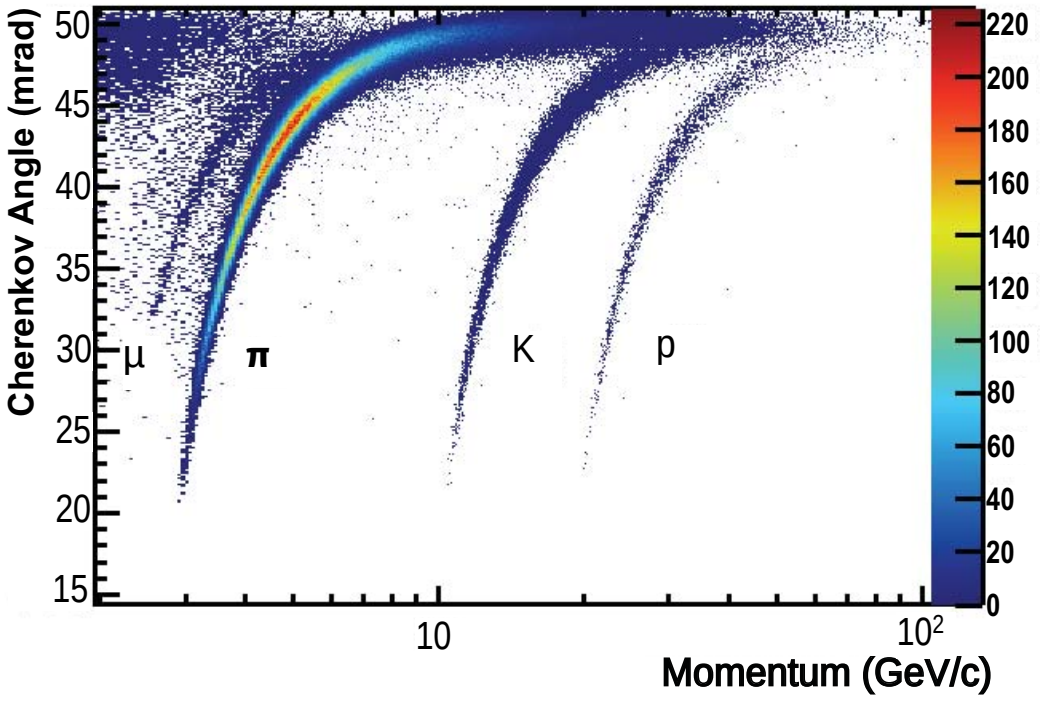


Figure 2.6: The reconstructed Cherenkov angle as a function of track momentum for isolated tracks in C_4F_{10} gas. Bands for muons, pions, kaons and protons can be seen [82].

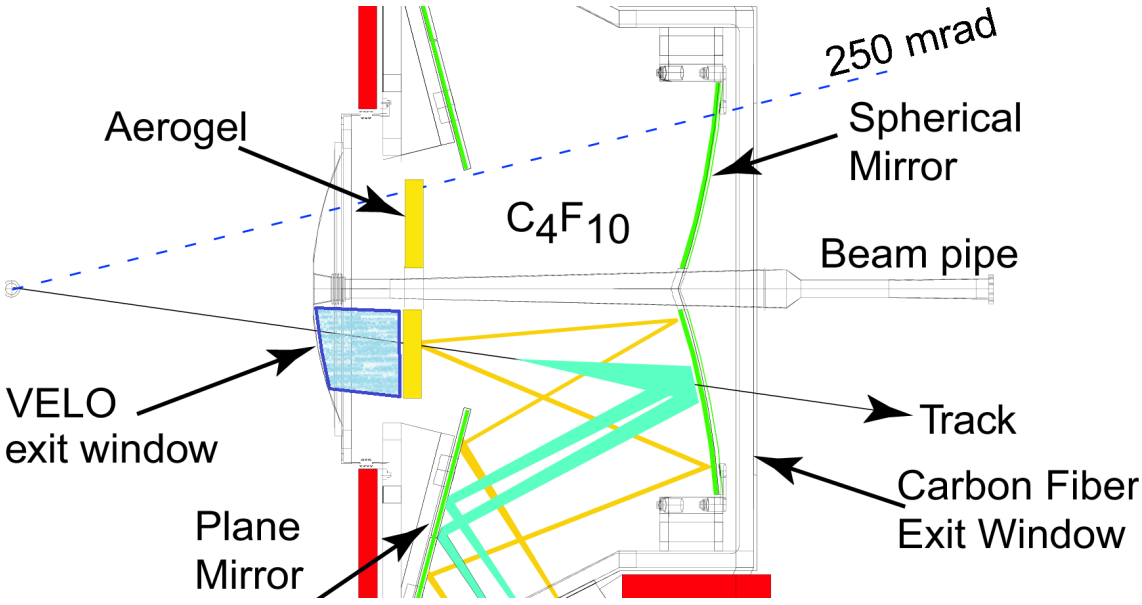


Figure 2.7: A schematic of the side view of RICH 1, when the aerogel radiator was in place Cherenkov photons produced in the blue shaded area filled with C_4F_{10} would be blocked [84].

Electromagnetic Calorimeter (ECAL) which are required to provide some information on the longitudinal development of electromagnetic showers, helping to reject the large background of charged pions. Lastly the particle will pass through the Hadronic Calorimeter (HCAL). All of the detectors measure the energy by transmitting scintillation light resulting from particle interactions to photomultiplier tubes via wavelength shifting fibres.

The SPD/PS detect charged particles and consist of two scintillator pads which are identical except that the SPD dimensions are all 0.45% smaller than the PS to maintain the projective geometry of LHCb. In between the two is a 15mm lead converter which corresponds to 2.5 radiation lengths ($2.5X_0$). Overall the detector is 7.6 m wide and 6.2 m high. Both planes of the SPD/PS are segmented into inner, middle and outer regions in order to match the segmentation of the ECAL. No significant degradation in performance of the PS or SPD was observed after Run 1.

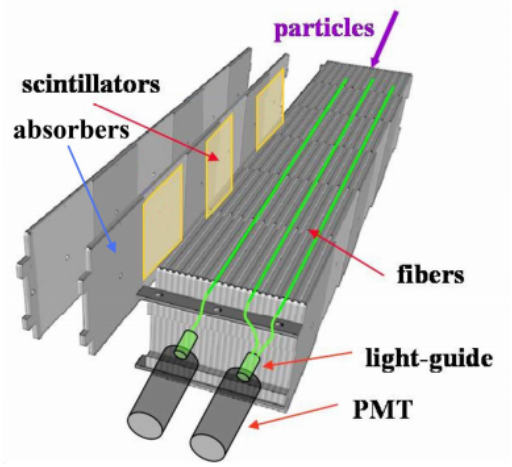
The ECAL detects and measures the energy of primarily electrons, photons and π^0's, although small amounts of energy are also deposited here by charged and neutral hadrons. The detector is located 12.5 m from the interaction point and uses shashlik calorimeter technology [97] which alternates between scintillator material and lead converters. The outer dimensions are such that they projectively match the upper limit of the angular acceptance but the lower limit is at 25 mrad to protect against radiation damage. As the hit density will be higher close to the beam axis the ECAL is split into three sections: the inner, middle and outer, with higher granularity in the inner section and reducing as distance from the beam axis increases. ECAL modules for these regions can be seen in Fig. 2.8a. The SPD and PS also match this projective segmentation. The ECAL has an area of 7.8 m $\times$ 6.3 m and a total thickness of 42 cm, consisting of 66 alternating layers of lead and scintillator material, resulting in a radiation length of 25 X_0. The ECAL was designed with an energy resolution of

$$\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}} \oplus 1\% \tag{2.3}$$

with E in GeV.

The HCAL detects and measures the energy of both charged and neutral hadrons. The detector is positioned 13.33 m from the interaction point and is made from alternating layers of iron and scintillator materials but in contrast to the ECAL they run parallel to the beam axis. This layout can be seen in Fig. 2.8b. It is 1.65 m thick with an area of 8.4 m $\times$ 6.8 m. Due to the limited requirements of the HCAL

(a) ECAL modules from inner, middle and outer regions [81].



(b) Exploded view of two scintillator-absorber layers and their placement within a HCAL module [81].

Figure 2.8: The LHCb calorimeters.

the interaction length (λ_I) was chosen to be $5.6\lambda_I$ with the ECAL adding a further $1.2\lambda_I$. The energy resolution determined from test beam studies was

$$\frac{\sigma_E}{E} = \frac{(69 \pm 5)\%}{\sqrt{E}} \oplus (9 \pm 2)\% \tag{2.4}$$

with E in GeV.

Bremsstrahlung recovery for electrons

Bremsstrahlung is electromagnetic radiation produced by deceleration of a charged particle typically when interacting with the electric field of an atomic nucleus. The probability to emit bremsstrahlung is proportional to the inverse of the squared mass of the particle making electrons the most susceptible to this radiation within LHCb. Emitted bremsstrahlung is treated in two categories, emission before the magnet and emission after the magnet, seen as E_1 and E_2 in Fig. 2.9. In the latter case photons will follow the trajectory of the electron and finish in the same ECAL cluster, however in the former case bremsstrahlung photons will appear in a different position in the ECAL to the electron. The reconstructed mass resolution suffers without these photons, hence, a dedicated algorithm to add these clusters was designed. The algorithm checks the trajectory of the electron before being deflected by the magnet and checks (using a χ^2 figure of merit) whether any neutral clusters

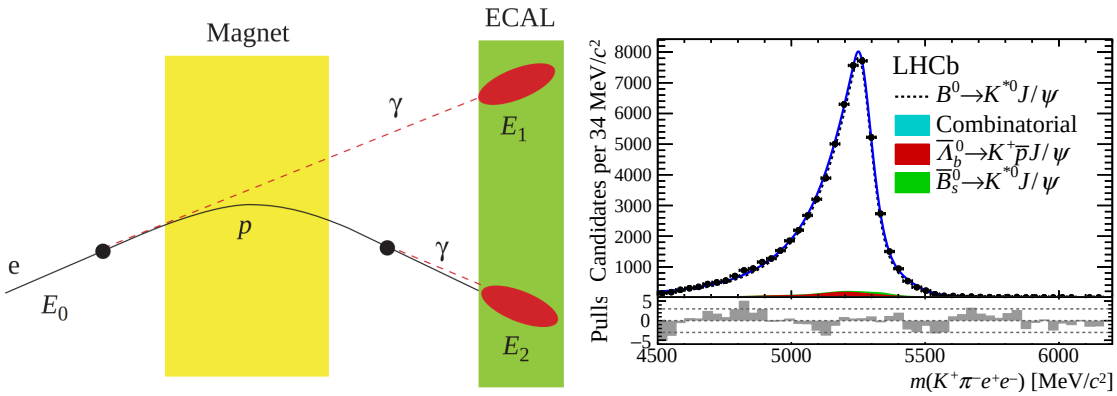


Figure 2.9: (left) Diagram of two categories of bremsstrahlung for electrons in LHCb [81]. (right) Reconstructed invariant mass distribution of $B^0 \to K^{*0} J/\psi (\to e^+e^-)$ taken from Ref. [71].

in the ECAL match this trajectory. If a matching cluster is found it is added to the cluster where the electron was detected and the deposited energy and momentum are recalculated. This procedure is not perfect and when too few photons are added a long tail in the lower sideband of the reconstructed mass is observed, similarly when too many photons are added a small tail, as this occurs less often, in the upper sideband of the reconstructed mass is observed. An example of these tails can be seen in Fig. 2.9. More details on the recovery procedure can be found in Ref. [86].

2.2.5 Muon system

Identification of muons in LHCb is extremely important as they are present in many key decay modes. The muon system is designed for this purpose and they are the final detectors a particle may traverse in LHCb. Typically these particles are high energy muons which are by far the most penetrating observable particle to be identified. The muon system consists of five muon stations (M1-M5) of rectangular shape corresponding to a total active area of 435 m^2, as seen in Fig. 2.10. The first station M1 is located before the calorimeters. Stations M2 to M5 are located downstream of the HCAL and between these stations are 80 cm thick iron absorbers designed to only allow penetrating muons through. Multi-wire proportional chambers are equipped in more than 99% of the total area of the system except for the part of M1 which is closest to the beamline which is equipped with triple-GEM detectors which have better ageing properties. Each station is divided into four regions (R1-R4) with R1 being closest to the beam axis, as seen in Fig. 2.10. The dimensions and segmentation of the regions from R1 to R4 follow the ratio 1:2:4:8.

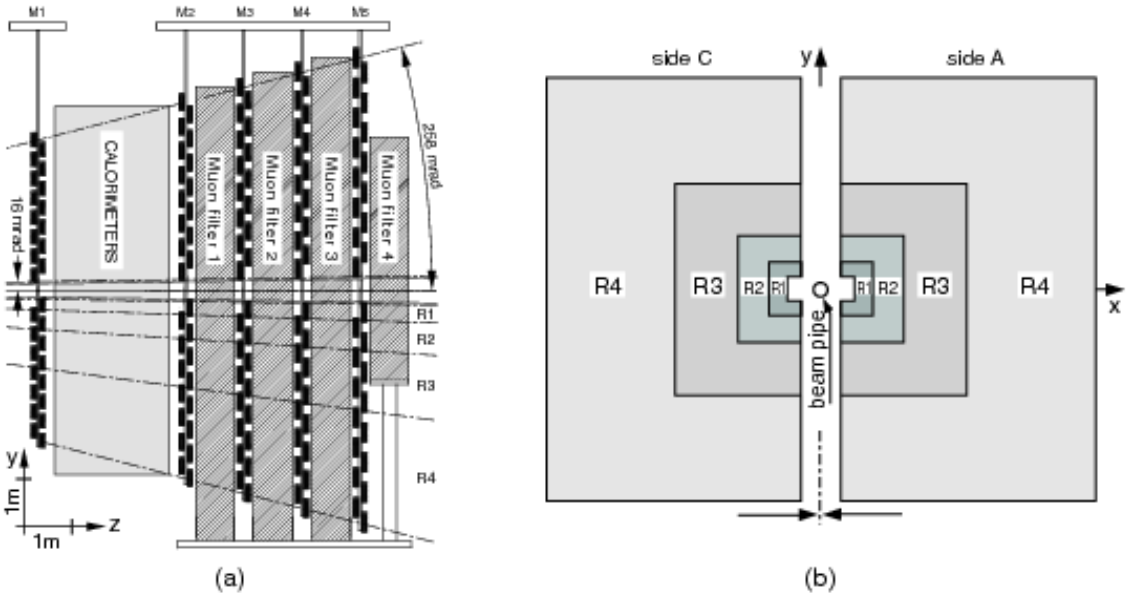


Figure 2.10: (a) Side view of the muon system [81]. (b) A schematic view of how the regions R1-R4 are positioned in each muon station [81].

The muon system provides an input to the hardware based trigger based on the p_{T} of muon candidates. In order to achieve a trigger efficiency of at least 95% the efficiency from each station must be $> 99\%$ within a time of $< 25\,\mathrm{ns}$. Using the full 2010 dataset the efficiency in each of the regions of each station was measured to be $> 99\%$ within this time constraint [98].

2.2.6 Trigger system

The LHCb trigger system [99] is designed to reduce the event rate from 40MHz down to a more manageable level. The online selection requires the use of a trigger which is split into two levels, the Level-0 (L0) trigger and the High Level Trigger (HLT), which is itself split into two levels, HLT1 and HLT2. The L0 trigger is purely implemented in hardware and is based on information from the calorimeters and muon system. Its primary role is to reduce the event rate from 40 MHz to 1 MHz retaining events likely to be of physics interest. Events passing the L0 trigger are then passed to HLT1 and HLT2 which are purely software based with HLT1 performing a partial event reconstruction and HLT2 performing a full event reconstruction. After both L0 and HLT trigger stages the event rate is reduced to 2–5 kHz. The trigger system is where most of the changes between Run 1 and Run 2 occurred for LHCb and Fig.2.11 summarises the differences in trigger strategy which will be discussed below.

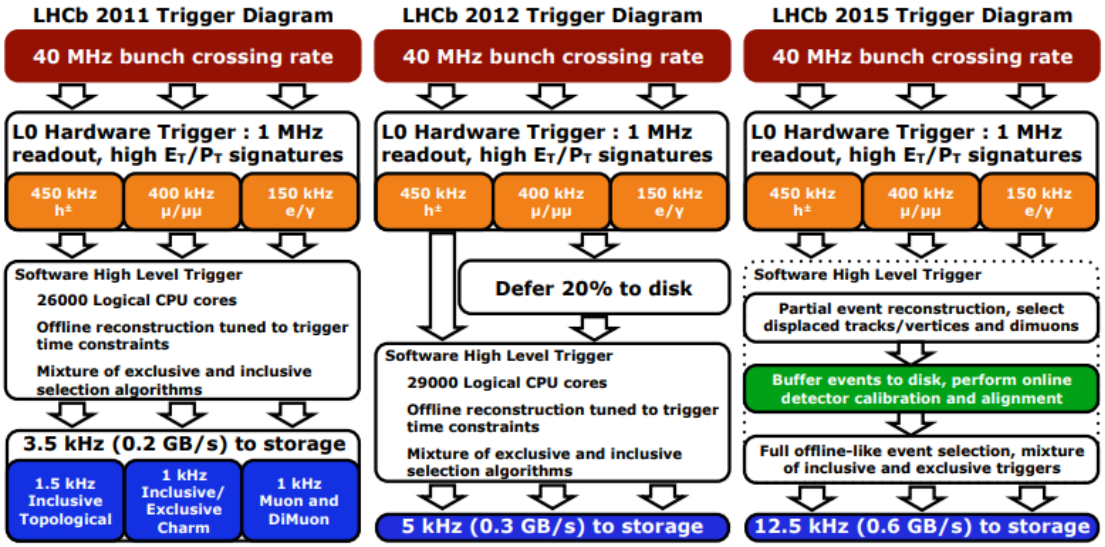


Figure 2.11: A diagram showing the differences in overall trigger strategy between 2011, 2012 and 2015 (which is representative of RUN 2 operation) [100].

Level-0 Trigger

The L0 trigger system has three main components: `L0Muon` is derived from information from the muon system, `L0Calo` from the elements of the calorimeter system and `L0PileUp` is used solely for luminosity calculations.

`L0Muon` is based on the p_{T} of muon candidates and requires a muon candidate track in all five of the muon stations. The p_{T} can be measured with a resolution of around 25% from the first two muon stations. The two muon candidates with the highest p_{T} are selected and p_{T} thresholds on two separate categories are set. One threshold is set on the single muon candidate with highest p_{T} and is known as the `L0Muon` trigger, another threshold is set on the product of the p_{T} of the two highest p_{T} muon candidates and is known as `L0Dimuon` trigger. In RUN 2 an additional trigger category for very high p_{T} muons was also introduced.

`L0Calo` is based on the E_{T} of candidates deposited in the calorimeter system. There are three types, `L0Hadron`, `L0Electron` and `L0Photon`. `L0Hadron` sets a lower limit on the HCAL cluster with the largest E_{T}; this also includes the small amount of energy deposited by the hadron in the ECAL. `L0Photon` sets a threshold on the ECAL cluster with the highest E_{T} that has associated hits in the PS detector but no hits in the SPD cells in front of these PS cells. `L0Electron` is similar to `L0Photon` but also requires a hit to be present in the SPD detector.

In all cases if the threshold set for any of the trigger categories is exceeded by any

Table 2.1: Minimum L0 trigger thresholds for Run 1 and Run 2 operation [101, 102].

L0 Trigger	E_T or p_T (GeV; GeV^2 for DiMuon)					SPD hits	
	2011	2012	2015	2016	2017	Run 1	Run 2
Hadron	3.50	3.70	3.60	3.70	3.46	600	450
Photon	2.50	3.00	2.70	2.78	2.47	600	450
Electron	2.50	3.00	2.70	2.40	2.11	600	450
Muon	1.48	1.76	2.80	1.80	1.35	600	450
DiMuon	1.30	1.60	1.69	2.25	1.69	900	900

candidate in the event this event is saved and sent on to the HLT step. A further requirement on the SPD hits is also made and this is to veto events which would take a prohibitively long time to compute in the HLT. This requirement does not result in significant loss of signal however, as the ability of the L0 to discriminate between signal and background is greatly reduced with increasing event multiplicity. The `L0DiMuon` and high p_T `L0Muon` triggers have a much lower rate of accepted events hence have reduced SPD hit requirements. L0 Thresholds for 2011–2017 data taking years can be seen in Tab. 2.1.

High Level Trigger

Events passing the L0 trigger are processed by the event filter farm, comprising 1700 nodes of which 800 were added in Run 2 for improved computing performance. The software used in the HLT is the same as used in offline reconstruction. HLT1 performs a partial event reconstruction under tight time constraints, these events are then subject to a full reconstruction by HLT2 and if all available nodes are busy are sent to a disk buffer which in Run 2 could hold an event for up to two weeks.

As HLT1 is subject to a short time constraint it is limited to a few operations. The trajectories of long track particles can be reconstructed if they have a p_T of greater than 500 MeV/c, the PV can be reconstructed and muons can be identified. Any tracks of vertices that meet the p_T requirements are also required to be of good quality. No particle identification algorithms can be executed at this step except for muon identification, which leaves a very clear pattern in the detector. In Run 2 HLT1 also performed a real time calibration and alignment for various detectors including the VELO, which requires realignment when the VELO is extended and retracted during data taking. In Run 1 this information had to be computed offline and applied after the fact.

Any events passing HLT1 are then processed by HLT2 which performs a full event reconstruction. This is split into reconstruction of tracks of charged particles,

reconstruction of neutral particles and particle identification. In RUN 2 the reconstruction algorithms were optimised with respect to RUN 1 such that the same or better performance was achieved in a shorter period of time. Improved performance is also due to calibration and alignment information coming from HLT1 resulting in full offline quality data being available directly from HLT2. After the HLT2 a preselection known as the stripping, which will be described later in this work, is performed online so it is ready for analysis immediately and this is known as the TURBO stream. Full details of the reconstruction in RUN 1 and the improvements made in RUN 2 for the HLT can be found in Refs. [101, 102] respectively.

Measuring trigger efficiencies - TISTOS Method

It is important to be able to calculate the efficiency of the various trigger steps but as data which are not selected by the triggers is not saved a direct efficiency determination is not possible. One way to measure this efficiency is to use the TISTOS method [103]. When events pass through the trigger system they are designated two categories TIS and TOS, with respect to each trigger (L0, HLT1, HLT2). Some events may be TIS and TOS. They are defined as:

- **TOS** - Those events for which a signal candidate in the event passes the trigger, without using information from the rest of the event.
- **TIS** - Those events for which the rest of the event passes the trigger, without needing the signal candidate to be present. As such these are generally unbiased with regards to the trigger selection but the signal candidate and the rest of the event may be correlated. An example of when this correlation would be important is when the event is triggered on the other B produced in the p-p collisions to the signal candidate B. Quantities such as the momentum distributions would be correlated as they originate from the same decay chain.
- **TIS&TOS** - Those events for which the both the signal candidate and the rest of the event pass the trigger independently of each other

Using the TISTOS method the TIS, ϵ^{TIS}, and TOS, ϵ^{TOS}, efficiencies can be calculated:

$$\epsilon^{TIS} = \frac{N^{TIS\&TOS}}{N^{TOS}} \qquad \text{and} \qquad \epsilon^{TOS} = \frac{N^{TIS\&TOS}}{N^{TIS}}, \tag{2.5}$$

where N^{TOS}, N^{TIS}, and $N^{TIS\&TOS}$ are the number of events classified as TOS, TIS and TIS&TOS respectively. This method will be used to measure the various trigger

efficiencies required to be calculated in the remainder of this work.

2.2.7 Particle Identification in LHCb

Particle identification in LHCb uses information from three areas, the calorimeter system, the two RICH detectors and the muon system, which were described in the previous sections. The calorimeter system provides PID information for electrons, photons and π^0's. The RICH detectors provide PID information for charged hadrons (p, K and π) and also contribute to electron and muon identification in conjunction with the calorimeter and muon systems. The muon system provides PID information for muons. Finally the PID information from all these detectors can be combined into a set of combined likelihoods which offer a more powerful variable than considering each detector system separately.

Particle Identification from the Calorimeter System

The calorimeter system is the only place within LHCb in which neutral particles can be detected. A deposited energy cluster within the calorimeter is considered belonging to a neutral particle if there is no associated charged track. If such a cluster is located within the ECAL it is very likely due to a photon. If a photon converts to an electron in the lead between the SPD and PS a charged track will only appear in the PS indicating a converted photon. Identifying upstream converted photons requires identifying the electron and positron pair produced. To achieve photon identification a photon hypothesis likelihood is built from probability density functions representing signal and background. For unconverted photons this is based on variables that characterise how well a cluster in the ECAL matches with a charged particle track, the ratio of the energy of an ECAL cluster to the total energy deposited in the ECAL, and the energy deposited in the PS in front of a cluster in the ECAL. The difference between the estimators for the background and photon hypothesis is then used to determine if the particle was a photon.

Two cases of photon identification in particular are of interest, those of high energy prompt photons or those originating from a π^0. For the first case, using $B^0 \to K^{*0}\gamma$ decays from 1 fb^{-1} of LHCb data results in a mass resolution of 93 MeV/c^2 [104]. This is strongly dependent on the energy resolution of the ECAL. For π^0's there are two cases, those of 'resolved' π^0's which have low p_{T} and therefore result in well separated photons and those of 'merged' π^0's which have high p_{T} and where due to the limited granularity of the ECAL it is not possible to resolve both photons individually.

Based on $D^0 \to K^-\pi^+\pi^0$ decays the invariant mass resolution is $30\,\mathrm{MeV}/c^2$ for the merged case and $20\,\mathrm{MeV}/c^2$ for the resolved case [105]. To avoid misidentifying photons and merged π^0's a dedicated neural network classifier is trained with relevant simulated samples. A photon identification efficiency of 95% can be obtained whilst rejecting 45% of π^0 candidates which are reconstructed as photons [82].

The most problematic charged particles to identify are electrons so a number of design choices were made with this in mind. The electron L0 trigger, which takes information from the calorimeter system, selects electrons with high transverse energy but must still reject large amounts of charged and neutral pions. Charged pion rejection requires longitudinal separation of the electromagnetic shower detection which is realised by the PS and the main section of the ECAL. Neutral pion rejection is achieved by a charged particle detector, the SPD, being placed in front of the PS. If hits are seen in the PS and ECAL but not the SPD, these hits are unlikely to be due to an electron from the decay of interest. Electrons resulting from photon conversions in the thin layer of lead are also rejected in this manner. Electrons arising from photon conversions upstream of the SPD cannot be rejected with this system. Using information from the PS, ECAL and HCAL, combined likelihoods are built from signal and background distributions from each of the detectors. A particularly discriminating variable for the ECAL is the distribution of E/pc as can be seen in Fig. 2.12 . A combined likelihood for the calorimeter system, $\Delta\log\mathcal{L}^{CALO}(e-h)$, based on the difference between an electron, e, and hadron, h, hypothesis can be constructed as:

$$\Delta\log\mathcal{L}^{CALO}(e-h) = \Delta\log\mathcal{L}^{ECAL}(e-h)+\Delta\log\mathcal{L}^{HCAL}(e-h)+\Delta\log\mathcal{L}^{PS}(e-h). \quad (2.6)$$

In Eq.2.6 $\Delta\log\mathcal{L}^{ECAL}(e-h)$ is based on the E/pc distribution and the how well a charged track matches with an ECAL cluster, $\Delta\log\mathcal{L}^{HCAL}(e-h)$ and $\Delta\log\mathcal{L}^{PS}(e-h)$ are based on energy deposits in each of the detectors. Using 2011 data the average identification efficiency of electrons from $J/\psi \to e^+e^-$ originating from $B^\pm \to J/\psi\, K^\pm$ candidates was determined to be $91.9 \pm 1.3\%$ with a rate of misidentification of $4.54 \pm 0.02\%$ for a requirement of $\Delta\log\mathcal{L}^{CALO}(e-h) > 2$ applied [82].

Particle identification from the RICH system

The primary aim of the two RICH detectors is to distinguish between protons, kaons and pions. The detectors can also contribute to electron and muon identification. The identification procedure uses all the tracks in the event and in both the RICH

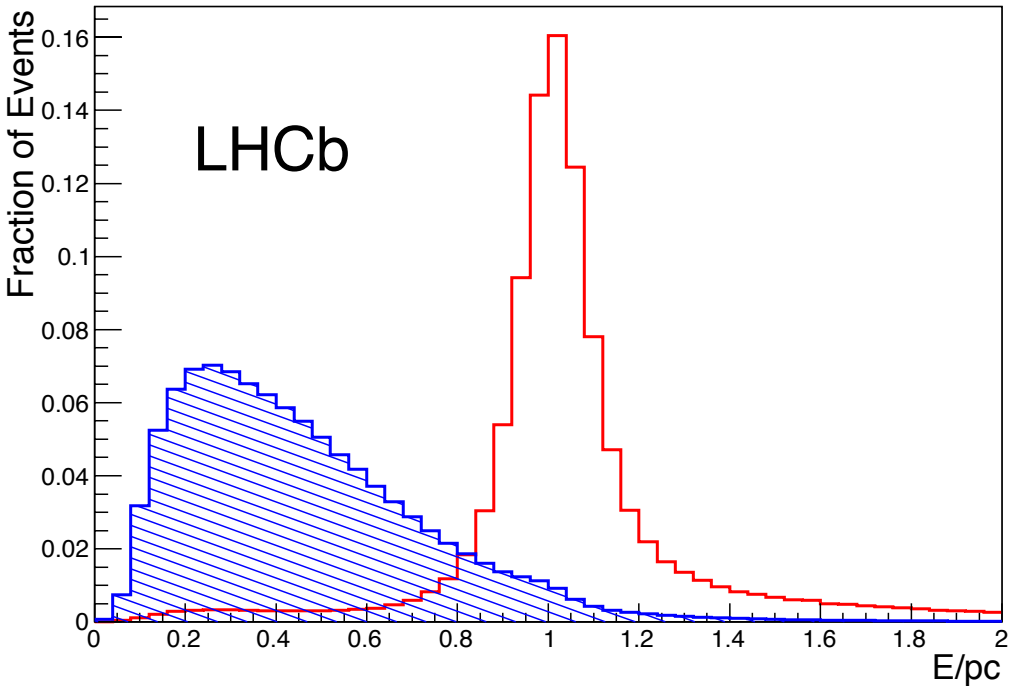


Figure 2.12: E/pc distribution from the ECAL with 2011 data. (red) Electron distribution (blue) hadron distribution [82].

detectors simultaneously, allowing to distinguish between overlapping Cherenkov cones. Since pions are the most numerous particles produced in p-p collisions all candidates begin with a mass hypothesis of a pion. An overall likelihood is determined from the distribution of RICH photon hits, the tracks associated to these and their errors. The mass hypothesis for each track in turn is changed to kaon, proton, electron and muon whilst leaving the mass hypothesis of all other tracks unchanged. Within the set of all tracks the change in mass hypothesis yielding the largest increase in the likelihood is determined and the mass hypothesis for that track is set to that of the preferred particle. Full details of this procedure can be found in Ref. [106].

Using samples of kaon tracks selected with purely kinematic selections from decay modes with large branching fractions the PID performance of the RICH detectors can be evaluated. The plot in Fig. 2.13 shows the efficiency of identifying a true kaon as a kaon and a true kaon as a pion with different requirements on the $\Delta\log\mathcal{L}(K-\pi)$, often shortened to $DLL_K\pi$. The larger this value the more likely the track is a kaon. With a $\Delta\log\mathcal{L}(K-\pi) > 0$ the average (over the momentum range of 2-100 GeV/c) kaon efficiency is around 95% with a pion mis-identification rate of around 10%.

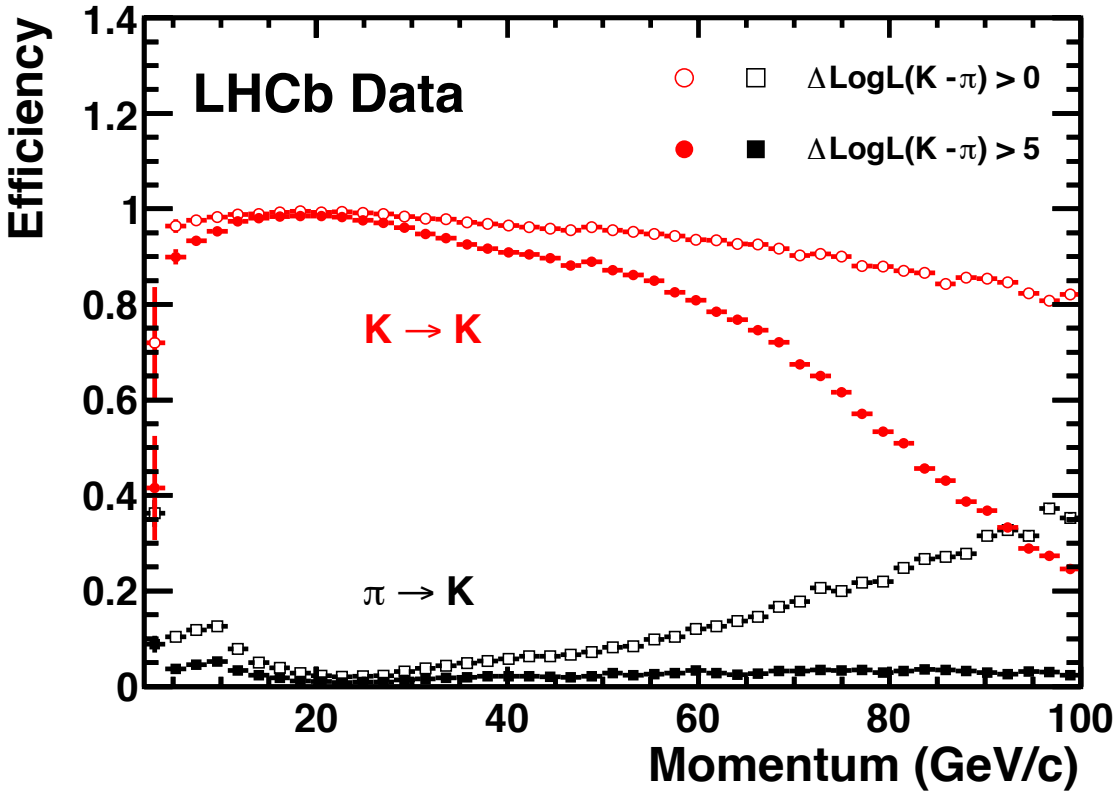


Figure 2.13: The kaon identification efficiency (red) and pion misidentification rate (black) determined from collision data as a function of track momentum. Unfilled markers represent a looser $\Delta\log\mathcal{L}(K-\pi)$ requirement whilst filled markers represent a tighter requirement [82].

Increasing this requirement to $\Delta\log\mathcal{L}(K-\pi) > 5$ the average results yield a reduction to around a 3% misidentification rate and a kaon efficiency of around 90% [82].

Particle identification from the Muon System

To determine whether a particle traversing the muon system is likely to be a muon, the association of a tracks' trajectory with hits in the muon system and the number of stations in which a hit would be expected (given the track momentum) are considered [107]. A genuine muon requires a momentum of at least 3 GeV/c to pass through the calorimeters to the M2 or M3 stations, whilst a momentum of above 6 GeV/c is required to pass through the entire muon system. Likelihoods for each muon candidate are computed using the average squared distance of hits closest to a track's extrapolated trajectory in the muon system, for those stations it is expected to reach.

Identification of a proton as a muon may arise due to random hits in the muon

detectors being aligned with the proton direction or if a genuine muon produced in the event travels in the same direction as a proton in the muon system. The main source of misidentifying pions or kaons as muons is from muon decays in flight to either of these particles. The total misidentification rate is roughly the sum of these two sources of error. In order to gain power to reject these sources of misidentification a likelihood is determined based on the pattern of hits around the extrapolated track direction in the muon system with muon and non-muon hypotheses. The logarithm of the ratio of the muon and non-muon hypothesis, $\Delta\log\mathcal{L}(\mu)$, is then used a variable to distinguish between muons and other particles.

Combining Particle Identification Information

In a similar manner to the case for electrons, a combined likelihood can be determined for each particle using combined information from the RICH, calorimeter and muon systems. These are designated as $\Delta\log\mathcal{L}_{\text{comb}}(X-\pi)$ where the X is the particle of interest.

Another complementary method is to take the information from the various detectors as well as other information and train multivariate classifiers [108] so that correlations between variables are properly handled. These variables are designated ProbNNX, where the X defines the particle of interest. In some cases these variables outperform the combined log likelihood approach with better background rejection and signal efficiencies. Various tunes of these variables have been computed for Run 1 and Run 2 as the variable distributions entering the classifiers change with the run conditions.

Requiring an electron selection efficiency of 90%, the likelihood obtained using the combined detector information yields around an order of magnitude lower misidentification than that obtained using calorimeter system information alone, as can be seen in Fig. 2.14 [82].

2.3 Simulation in LHCb

In order to determine the efficiency of various selections on the decay mode of interest, LHCb makes use of simulated data. Pythia 8 [109] is used to simulate p-p collisions with a specific LHCb configuration [110] applied. EvtGen [111] is used to simulate hadronic decays and final state radiation is dealt with using Photos [112]. The Geant4 toolkit [113] is used to describe generated particle interactions with

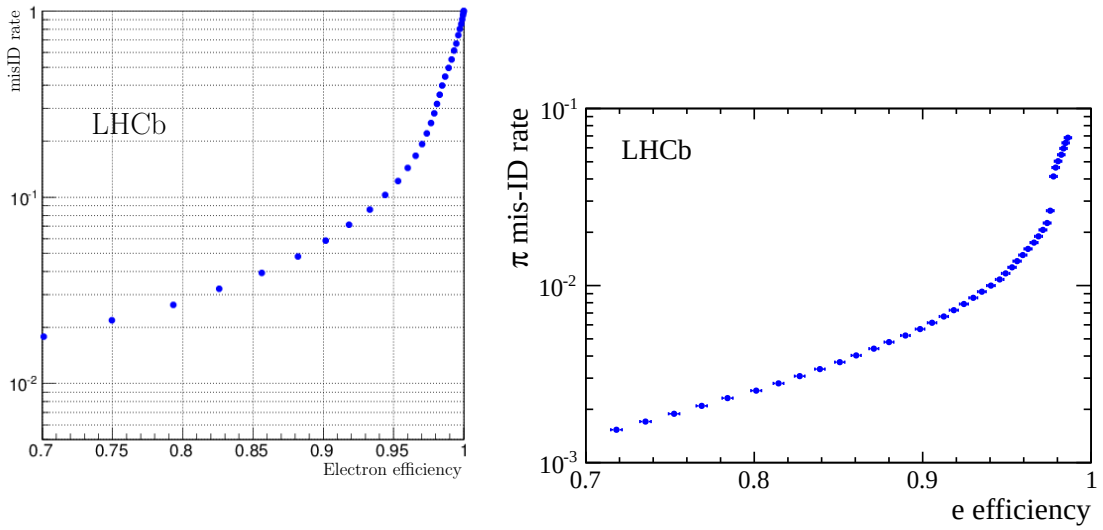


Figure 2.14: Pion misidentification rate versus electron efficiency, (left) using just PID information from the calorimeter system, (right) using PID information from all detectors [82]

the detector and the detector response. A software package known as GAUSS [114] incorporates all of these elements and is used for simulation in LHCb.

In order that simulated data undergo the same processing as collision data the detector response and reconstruction must be emulated well. Therefore, after events have been generated with GAUSS they are then processed by BOOLE [115] which digitises the data thus emulating the detector response to data. Emulation of the trigger system is performed after the digitisation step by MOORE [115]. Reconstruction is performed by BRUNEL [115] using as input either digitised simulated data from BOOLE or directly from the detector. Reconstructed items include calorimeter clusters, charged tracks and information about PID. Finally to select a subset of events which may be relevant to a particular analysis DAVINCI [115] is used in the stripping preselection, which is based on particle types and kinematic quantities. The data flow in LHCb for collision and simulated data is summarised in Fig. 2.15.

2.3.1 Tests of the simulation - Muon Multiple Scattering

As new versions of software, such as GAUSS or GEANT4, are released it is important to compare the performance of these new versions to older ones to avoid introducing errors and to quantify any improvements that may have been made. In this work tests of muon multiple scattering using simulated data from different software versions

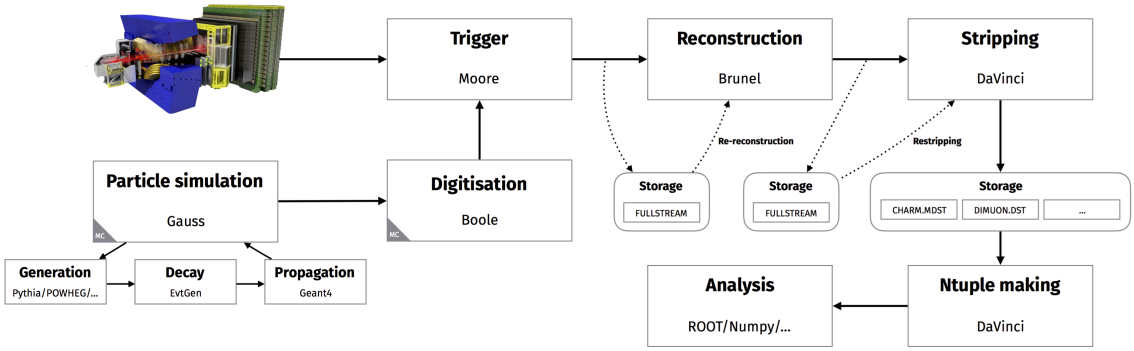


Figure 2.15: Representation of the data flow within LHCb for simulated and real collision data.

will be discussed as well as how this test has been implemented into LHCbPR [116], a web client developed to present the results of this and similar tests.

Muon Multiple Scattering

Charged particles such as muons experience small angle scatters from nuclei in the material of detectors in LHCb, primarily the calorimeters and muon filters as they have the greatest amount of material. The vast majority of these scatters are due to Coulomb scattering and the resulting distribution is well described by Molière's theory [117]. For small deflection angles, θ_0, the distribution is approximately Gaussian. For scattering angles exceeding a few θ_0, the distribution resembles Rutherford scattering, which has larger tails than a Gaussian distribution. Therefore, in most cases it is safe to use a Gaussian to approximate the central 98% of the angular distribution resulting in a width of [118, 119]

$$\theta_0 = \frac{13.6 \text{ MeV}}{\beta cp} z \sqrt{\frac{x}{X_0}} \left[1 + 0.038 \ \ln \frac{x z^2}{X_0 \beta^2} \right]. \tag{2.7}$$

In Eq. 2.7, p is the momentum, βc the velocity and z the charge (units of $|e|$) of the incident particle. The thickness in radiation lengths of the medium in which scattering takes place is represented by x/X_0. This form of θ_0 is determined from a fit to the Molière distribution [117] for singly charged particles with $\beta = 1$ and is found to be accurate to 11% or better in the range of $10^{-3} < x/X_0 < 100$, well within the limits of what is encountered in LHCb.

Whilst Eq. 2.7 describes scattering from a single material, the calorimeters in LHCb are alternating layers of scintillator and other materials. In this case it is better to use Eq. 2.7 with the combined x and X_0 over the whole scatterer rather than sum

in quadrature θ_0 values for each alternating layer of scatterer. Since this result was derived from a fit the end result would be systematically too small. This approach has been further improved by Lynch and Dahl [119] and is now able to achieve results that are considered accurate to better than 2%.

The muon multiple scattering test

Muon multiple scattering in the calorimeters and muon filters of LHCb are the main areas where muon trajectories will be affected by multiple scattering. If the amount of scattering is high this can affect the momentum resolution. The aim of the test is to monitor the effect of this scattering on muon trajectories and the variation of this with changing software revisions.

The test is designed to compute the differences in angle and displacement of muon trajectories between the five muon stations, resulting in four regions in which this is monitored. Between M1 and M2 scattering is observed in the calorimeter system whilst the remaining three regions probe scattering in muon filters one, two and three. A particle gun is used to simulate 5000 muons uniformly in the range 0–2000 GeV and a pseudorapidity range of 1.8–5.2. For each of the four regions, six types of plots are produced. These are

- δx vs $\delta\theta_x$ (or $\delta\theta_y$)
- p vs δx (or δ_y)
- p vs $\delta\theta_x$ (or $\delta\theta_y$),

where p is the muon momentum, δ_x is the difference in displacement in the x direction between any two consecutive muon stations and $\delta\theta_x$ is the difference in angle in the x-z plane between any two consecutive muon stations. For ease of comparison, profile histograms of these scatter plots are also produced.

As well as comparing the results of the test with different software versions, it is also possible to compare the variation when different physics lists provided by GEANT4 are used. If an alternative list is used then physics processes may be simulated by different models in specific energy ranges. The default physics list in LHCb is EmNoCuts but EmStd, EmLHCb, EmLHCbNoCuts and EmOpt1, 2 and 3 can also be used. The EMStd physics list uses the G4UrbanMscModel for multiple scattering of e^+ and e^- below 100 MeV, the G4WentzelVIModel for multiple scattering of combined scatterers with G4eCoulombScatteringModel for large angle

scattering for muons, pions, kaons, protons, and anti-protons at all energies and the G4UrbanMscModel for all other charged particles. By default the physics lists include production thresholds for EM processes but these can be disabled in the lists which contain NoCuts. The EmLHCb physics lists have been tailored to LHCb's specific configuration and are not publicly available and uses a different algorithm to model bremsstrahlung. The EmOpt physics lists are standard presets that come with GEANT4 and are publicly available, EmOpt1 is tailored to a CMS configuration, EmOpt2 is tailored to a LHCb configuration, although not as finely tuned as the EMLHCb list and EmOpt3 is tailored for situations that require higher electron, hadron and ion tracking in the absence of a magnetic field. More information about the physics lists can be found in Ref. [120].

The results displayed in Figs. 2.16 and 2.17 were created to compare the results of the test for GEANT4 v10.3.3 denoted by red and GEANT4 v10.4.1 denoted by black. As can be seen the results of test were in good agreement between the versions confirming that the muon multiple scattering did not significantly change with this upgrade. As expected high momentum particles suffer much less deflection (in angle and displacement) than lower momentum particles.

To allow users to view the results of this and other tests, and compare and contrast results over versions of software, a web client known as LHCbPR has been developed. In addition to allowing users to download the files produced by a given test users are able to select jobs interactively and view the created data online. If only a single job is selected the plots that are set to be produced by that job can be viewed. If more than one job is selected plots can be viewed side by side, superimposed and ratio plots can be produced, amongst other features. An example of the web interface can be seen in Fig. 2.18. LHCbPR is now being used to validate GAUSS, GEANT4, the HLT and BRUNEL with plans to expand further into LHCb.

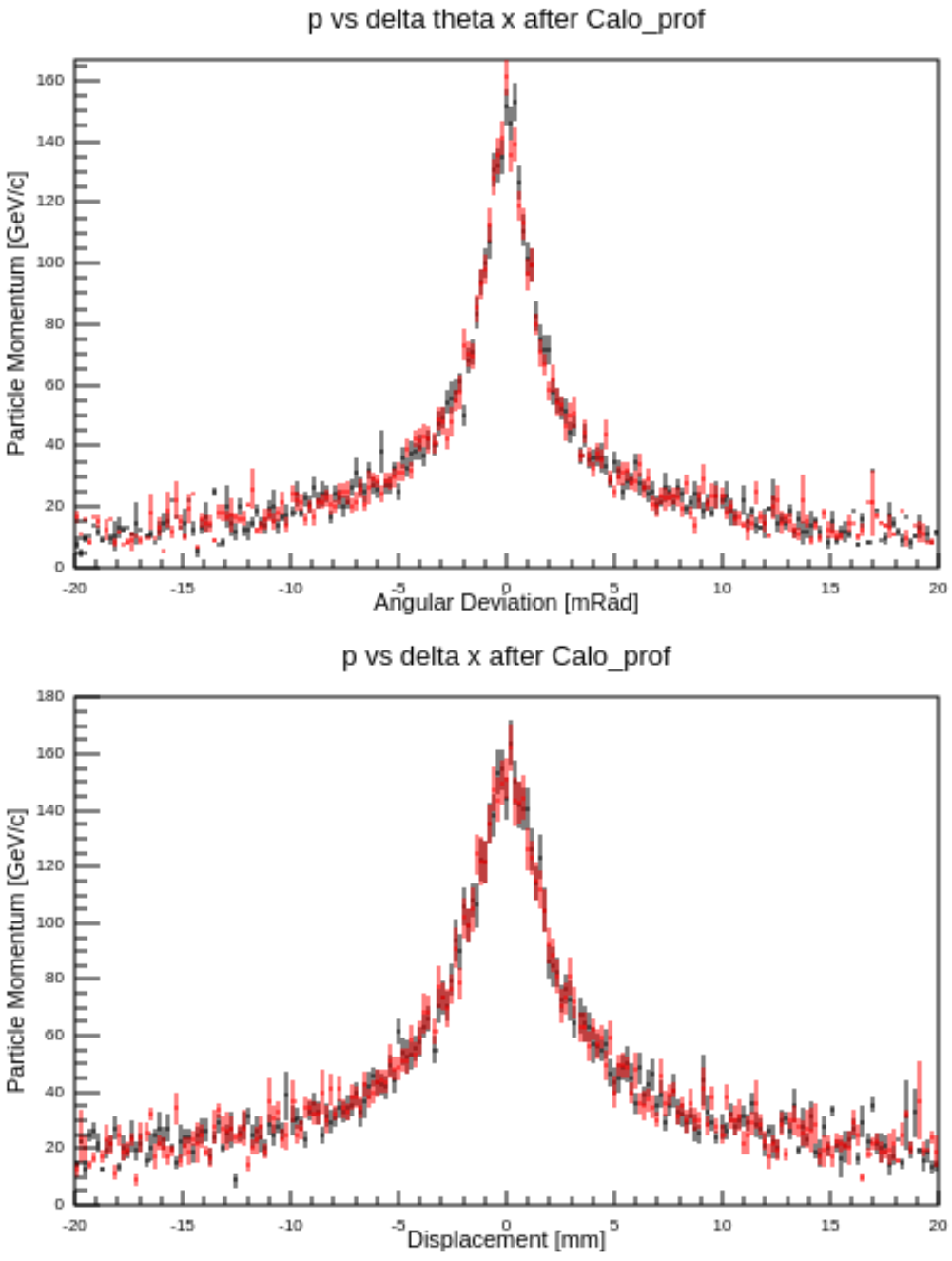


Figure 2.16: Comparison of muon multiple scattering in GEANT4 v10.3.3 (red) and GEANT4 v10.4.1 (black) probing the calorimeter region. (top) Profile histogram of momentum versus $\delta\theta_x$, (bottom) Profile histogram of momentum versus δ_x.

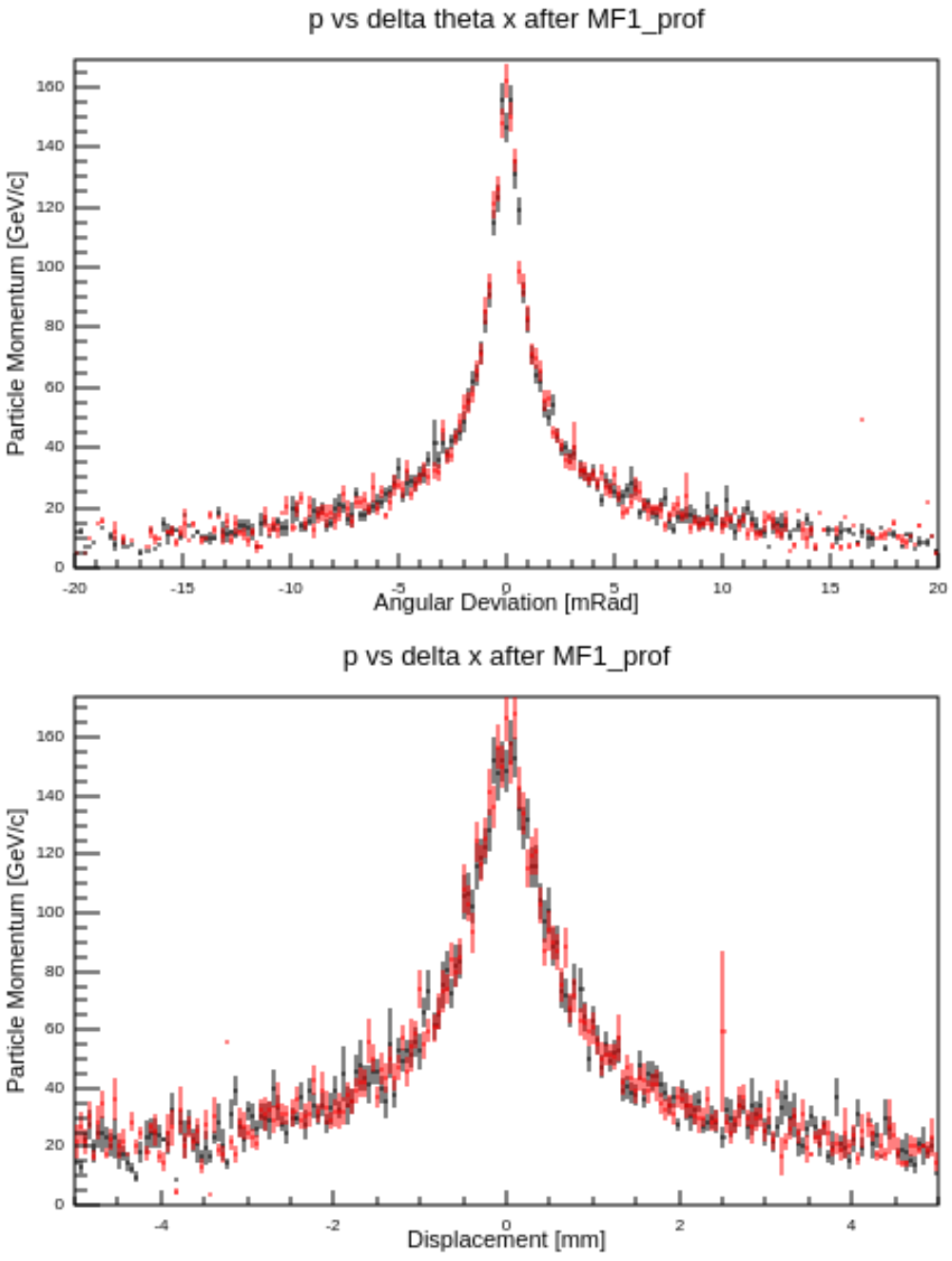


Figure 2.17: Comparison of muon multiple scattering in GEANT4 v10.3.3 (red) and GEANT4 v10.4.1 (black) probing the first muon filter. (top) Profile histogram of momentum versus $\delta\theta_x$, (bottom) Profile histogram of momentum versus δ_x.

Figure 2.18: An example of the LHCbPR web client showcasing comparing two plots side by side.

Chapter 3

Analysis Strategy

This analysis will make use of the full LHCb dataset from 2011 to 2018 corresponding to roughly $9\,\text{fb}^{-1}$. The ultimate aim of the analysis is to use Eq. 1.13 and simultaneously measure the R_K and R_{K^*} ratios, using $B^+ \to K^+\ell^+\ell^-$ and $B^0 \to K^{*0}\ell^+\ell^-$ decays from the aforementioned dataset. These decays both involve $b \to s\,\ell^+\ell^-$ transitions with the leading order Feynman diagrams displayed in Fig. 3.1. The motivation for making these measurements is outlined in Sec. 1.5.1.

The main focus of the work presented in this thesis is the determination of similar ratios known as $r^K_{J/\psi}$ and $r^{K^*}_{J/\psi}$, where the lepton pair is produced via a J/ψ resonance, that correspond to $B^+ \to K^+ J/\psi (\to \ell^+\ell^-)$ and $B^0 \to K^{*0} J/\psi (\to \ell^+\ell^-)$ decays, respectively. The tree level Feynman diagrams for these processes are displayed in Fig. 3.2. This measurement must be performed before making the final determinations of R_K and R_{K^*} which are discussed in more detail in Sec. 3.2. The vast majority of this work is dedicated to the measurement of $r^K_{J/\psi}$ and $r^{K^*}_{J/\psi}$ so that confident measurements of R_K and R_{K^*} may be made. Where possible, further studies have also been made for $B^+ \to K^+\ell^+\ell^-$ and $B^0 \to K^{*0}\ell^+\ell^-$ decays, as well as similar modes where the lepton pair proceeds via a $\psi(2S)$ meson.

3.1 Definition of q^2 regions

Ultimately the measurement of R_K and R_{K^*} will be performed in two separate regions ('bins'), of q^2. The reasons for doing this are:

- Regions in q^2 where $c\bar{c}$ resonances dominate can be excluded since the high rate

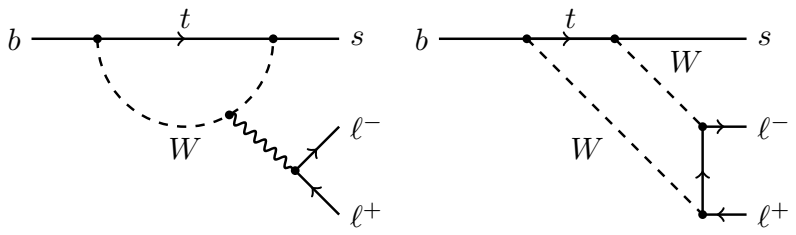


Figure 3.1: 'Penguin' (left) and 'Box' (right) Feynman diagrams for $b \to s\,\ell^+\ell^-$ decays in the SM.

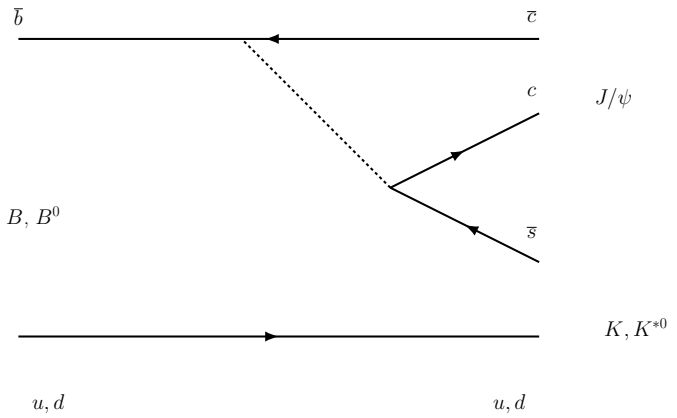


Figure 3.2: Tree-level Feynman diagrams for the control mode processes which proceed via a J/ψ resonance.

of the resonant decays will mask any non-SM behaviour of the non-resonant decays in these regions.

- Regions where different Wilson coefficients dominate can be considered largely independently from each other, as illustrated in Fig. 3.3 using $B^0 \to K^{*0}\mu^+\mu^-$ as an example. If non-SM behaviour is observed this will give some indication of which coefficients may be responsible.

- Regions where the ratios can have some dependence on q^2, depending on the BSM scenario, can be considered separately from regions where there will be effectively no dependence on q^2. The R_{K^*} ratio in particular can have some dependence on q^2, especially in the low-q^2 region, depending on the BSM scenario considered as seen in Fig. 3.4. In contrast the R_K ratio is fairly flat independent of the BSM scenario.

As the measurement of R_K and R_{K^*} is to be simultaneous the binning in q^2 is identical for both cases. The measurement is thus carried out in two bins of q^2 inspired by the Run 1 LHCb R_{K^*} analysis [71]. They are defined as:

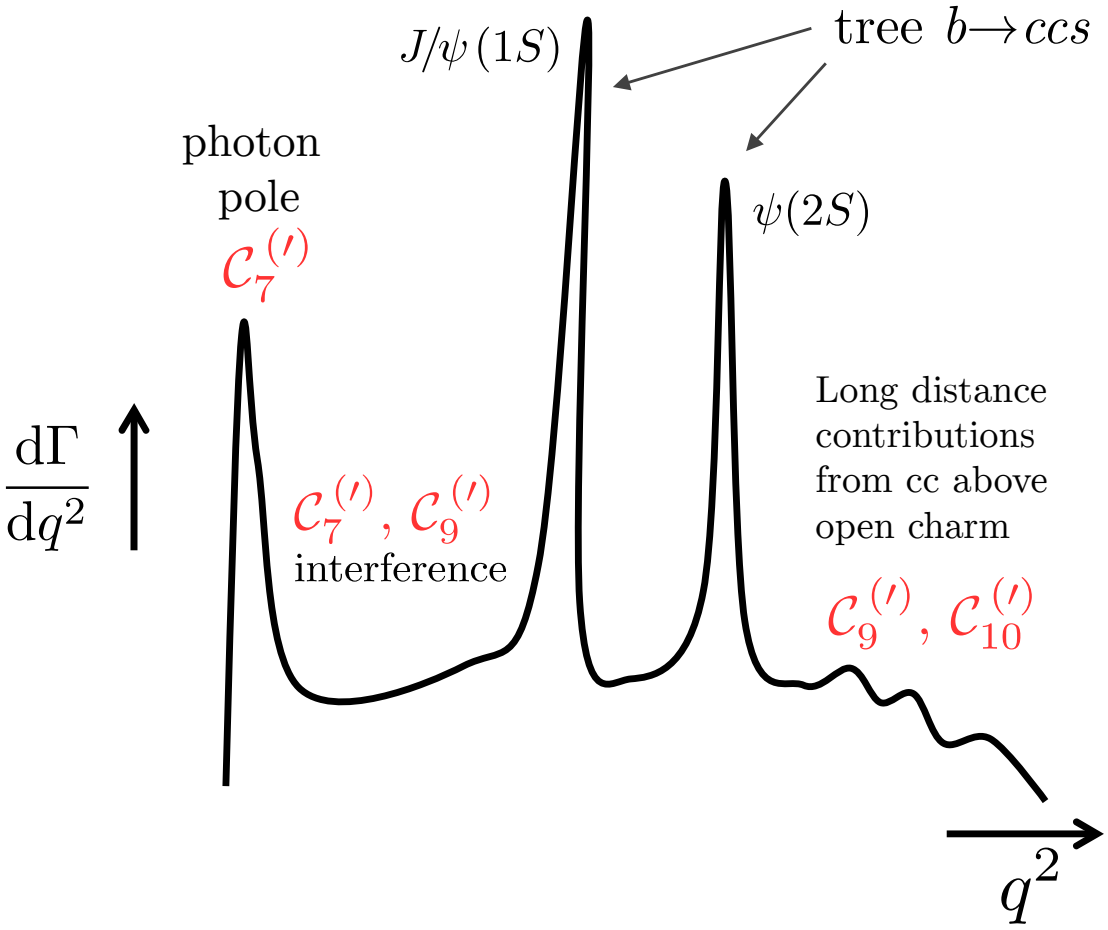


Figure 3.3: A sketch outlining where different Wilson coefficients contribute in the q^2 spectrum of the $B^0 \to K^{*0}\mu^+\mu^-$ decay [121].

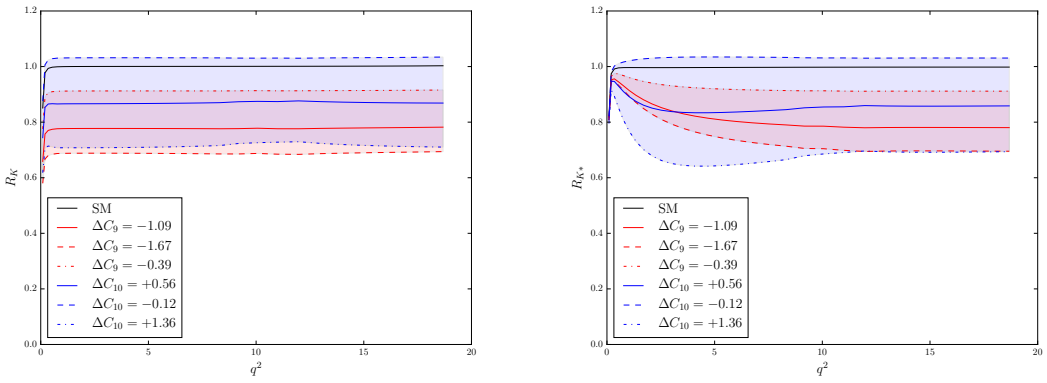


Figure 3.4: How (left) R_K and (right) R_{K^*} vary as a function of q^2 under various BSM scenarios which modify values of the Wilson coefficients. Contributions from $c\bar{c}$ have been subtracted in both cases.

- **low-q^2** - 0.1–1.1 GeV/c^2.

- **central-q^2** - 1.1–6.0 GeV/c^2.

The rationale behind these choices is as follows:

- **low-q^2** - The lower bin edge is chosen to correspond to where the SM prediction of R_{K^*} is unity. The upper edge is chosen such that the $\phi \to \ell\ell$ resonances are included in this bin alone.

- **central-q^2** - The upper bin edge is chosen to be as high as possible without significant contamination from the J/ψ resonance.

As well as the low and central-q^2 bins, designed for the R_K and R_{K^*} measurements, there are also selections made to isolate regions where the J/ψ and $\psi(2S)$ resonances dominate, where ratios involving the resonant modes may be determined. These selections differ if the leptons are muons or electrons, due to the bremsstrahlung experienced by the electrons. The selections are:

- **Muons** - $\pm 100\,\text{MeV}$ around the J/ψ ($\psi(2S)$) mass

- **Electrons** - For J/ψ resonances a region 6 GeV $< q^2 <$ 11 GeV is selected and for $\psi(2S)$ resonances a region 11 GeV $< q^2 <$ 15 GeV is selected.

3.2 Single and Double ratios

The definitions of R_K and R_{K^*}, in both the low and central-q^2 bins, can be expressed in experimentally accessible quantities such as the measured yields, $\mathcal{N}$, and efficiencies, ε, from both the electron and muon decay modes as seen in Eqs. 3.1 and 3.2. The measured efficiency for a particular decay mode acts to correct its measured yield making the efficiency determination a crucial part of the analysis procedure. The SR subscript denotes that these are so-called 'Single Ratios' since the ratios involve decays where the only difference is the final lepton pair.

$$R_K^{SR} = \frac{\mathcal{N}_{B^+\to K^+\mu^+\mu^-}}{\mathcal{N}_{B^+\to K^+e^+e^-}} \cdot \frac{\varepsilon_{B^+\to K^+e^+e^-}}{\varepsilon_{B^+\to K^+\mu^+\mu^-}}\,. \tag{3.1}$$

$$R_{K^*}^{SR} = \frac{\mathcal{N}_{B^0\to K^{*0}\mu^+\mu^-}}{\mathcal{N}_{B^0\to K^{*0}e^+e^-}} \cdot \frac{\varepsilon_{B^0\to K^{*0}e^+e^-}}{\varepsilon_{B^0\to K^{*0}\mu^+\mu^-}}\,. \tag{3.2}$$

These types of ratios can also be defined when the lepton pair originates from a J/ψ

resonance, these are known as $r^{K}_{J/\psi}$ and $r^{K^*}_{J/\psi}$ and are defined in Eqs. 3.3 and 3.4.

$$r^{K}_{J/\psi} = \frac{\mathcal{N}_{B^+\to K^+ J/\psi(\to\mu^+\mu^-)}}{\mathcal{N}_{B^+\to K^+ J/\psi(\to e^+e^-)}} \cdot \frac{\varepsilon_{B^+\to K^+ J/\psi(\to e^+e^-)}}{\varepsilon_{B^+\to K^+ J/\psi(\to\mu^+\mu^-)}}\,. \tag{3.3}$$

$$r^{K^*}_{J/\psi} = \frac{\mathcal{N}_{B^0\to K^{*0} J/\psi(\to\mu^+\mu^-)}}{\mathcal{N}_{B^0\to K^{*0} J/\psi(\to e^+e^-)}} \cdot \frac{\varepsilon_{B^0\to K^{*0} J/\psi(\to e^+e^-)}}{\varepsilon_{B^0\to K^{*0} J/\psi(\to\mu^+\mu^-)}}\,. \tag{3.4}$$

In order to perform a more experimentally robust measurement it is advantageous to use double ratios (DR), which are the ratios of the single ratios of the non-resonant and J/ψ resonant modes defined for both K^+ and K^{*0} modes in Eqs. 3.5 and 3.6. The ratio of the resonant mode, $r_{J/\psi}$, is expected to be unity even if BSM effects are present as discussed in 1.5.2. Hence the double ratio acts to cancel biases in the measured efficiencies.

$$\begin{aligned} R^{DR}_{K} &\equiv \frac{R^{SR}_{K}}{r^{K}_{J/\psi}} \qquad (3.5)\\ &= \frac{\mathcal{N}_{B^+\to K^+\mu^+\mu^-}}{\mathcal{N}_{B^+\to K^+ J/\psi(\to\mu^+\mu^-)}} \cdot \frac{\mathcal{N}_{B^+\to K^+ J/\psi(\to e^+e^-)}}{\mathcal{N}_{B^+\to K^+ e^+e^-}} \cdot \frac{\varepsilon_{B^+\to K^+ J/\psi(\to\mu^+\mu^-)}}{\varepsilon_{B^+\to K^+\mu^+\mu^-}} \cdot \frac{\varepsilon_{B^+\to K^+ e^+e^-}}{\varepsilon_{B^+\to K^+ J/\psi(\to e^+e^-)}} \\ R^{DR}_{K^*} &\equiv \frac{R^{SR}_{K^*}}{r^{K^*}_{J/\psi}} \qquad (3.6)\\ &= \frac{\mathcal{N}_{B^0\to K^{*0}\mu^+\mu^-}}{\mathcal{N}_{B^0\to K^{*0} J/\psi(\to\mu^+\mu^-)}} \cdot \frac{\mathcal{N}_{B^0\to K^{*0} J/\psi(\to e^+e^-)}}{\mathcal{N}_{B^0\to K^{*0} e^+e^-}} \cdot \frac{\varepsilon_{B^0\to K^{*0} J/\psi(\to\mu^+\mu^-)}}{\varepsilon_{B^0\to K^{*0}\mu^+\mu^-}} \cdot \frac{\varepsilon_{B^0\to K^{*0} e^+e^-}}{\varepsilon_{B^0\to K^{*0} J/\psi(\to e^+e^-)}} \end{aligned}$$

Since $r_{J/\psi}$ is used in the double ratio it is of vital importance that its value is unity and that it is flat in various decays that are used as proxies of the rare decay kinematics and topology (influences geometrical acceptance). This acts as a stringent test of the efficiency correction procedure. Therefore, before measuring R_K and R_{K^*} it is essential that $r^{K}_{J/\psi}$ and $r^{K^*}_{J/\psi}$ defined in Eqs. 3.3 and 3.4 meet these criteria and this is the main thrust of the work presented here.

3.3 General Stratergy

As can be seen in the expressions in Sec. 3.2 the analysis has two main parts: evaluating the efficiencies of the various decay modes and measuring their yields. The yield measurement depends on the efficiencies because the accuracy of Probability Density Functions (PDFs) used in the maximum likelihood fits depends on the

detector efficiencies that are measured. The measurement of the efficiencies themselves depends on the selection criteria introduced to isolate potential signal candidates in the dataset available. The strategy for each of these steps will now be discussed in turn.

3.3.1 Selection Strategy

The selection process in LHCb generally factorises into three stages, which in order of application are: a trigger selection encompassing the L0 and HLT; a so-called stripping selection and a final 'offline' selection. The general principles are outlined below, with more details presented in Sec. 4.

The Stripping (Sec. 4.2) and HLT (Sec. 4.5) selections are common to all signal and resonant decay modes. The stripping selection acts as an efficient preselection for certain types of decay modes. Stripping lines are thus set up to isolate certain families of decays which can then be further refined by analysts in the offline selection. HLT selections require that candidates are TOS with respect to a given HLT line, as discussed in Sec. 2.2.6. In practice only a subset of the HLT lines are of interest to any given analysis, for example certain lines may be configured to fire on 3 body events including a muon.

The HLT and stripping selections depend on properties of the signal candidates that are very similar or identical to the offline selection, so determining the efficiencies of these selections is not much more involved than computing offline efficiencies. However, determining the efficiency of the L0 trigger in hardware is more challenging. The L0 trigger selects events based on the momentum, transverse energy and event occupancy. However, the L0 resolution of these features is significantly worse than the offline resolution. Simulation does not always accurately model this resolution and the resolution can also vary between trigger categories for muons, electrons and hadrons. Hence the choice of L0 categories (Sec. 4.4) is driven by the need to align the muon and electron states as much as possible, whilst discarding the least signal.

The offline selection (Sec. 4.6), consists of PID criteria, various exclusive background vetoes and multivariate classifiers (MVAs). PID criteria (Sec. 4.7) are used to reduce backgrounds from other B decays which have similar topologies to the signal modes of interest, for example where a pion may have been misidentified as a kaon. Exclusive background vetoes (Secs. 4.9 and 4.10) are used to exclude decays that peak under or near to the signal peak and are not removed by the other

selection requirements. There are two types of MVA (Sec. 4.11), one designed to reduce combinatorial background and another to reduce partially reconstructed backgrounds in electron modes that cannot easily be vetoed. PID variables are not used in the classifiers because they are poorly modelled in simulation and a per event efficiency correction will be used to account for this, rather than a resampling procedure which would correct the underlying distributions. The main inputs of the MVA are variables based on kinematics and track quality, as well as isolation variables for MVAs designed to reject partially reconstructed background. The optimal working points of the MVAs are chosen by maximising the signal significance (Sec. 4.12); in the case of electron modes both MVAs are optimised simultaneously.

3.3.2 Efficiency Strategy

Simulated samples are always used to evaluate efficiencies in this analysis. Therefore the simulated samples should emulate collision data as well as possible to obtain a reliable efficiency estimate. There are a number of approaches available to correct simulated samples where they do not agree sufficiently well with collision data. For PID corrections (Sec. 6.4) efficiency corrections can be determined from calibration samples collected in each of the data taking periods. For L0 (Sec. 6.5) and HLT (Sec. 6.6) corrections the TISTOS method can be employed using J/ψ resonant simulation and collision data selected for the J/ψ mode in question. Kinematic, Multiplicity and Reconstruction corrections (Sec. 5.3) are based on a multivariate reweighting approach, trained on J/ψ resonant simulation and collision data selected for the J/ψ mode in question, to account for residual differences between simulation and collision data. The correction chain and strategy is discussed in more detail in Sec. 5. The final efficiencies are always determined with all corrections applied. The efficiency determination is discussed in more detail in Sec. 6.

As well as determining the efficiency for signal candidates, efficiencies are also determined for misidentified backgrounds which enter the final fits. This is done to constrain the yields (Sec. 7.8) where they may be too small to be allowed to vary freely, and to determine the shape of these components. The latter point is particularly important: although the selection aims to not change the shape of the signal component in the fit, this same selection can distort the background shapes, hence the need for reliable modelling of their shapes.

As the measurement will be simultaneous, corrections to the simulation can be taken from both types of resonant mode allowing cross-checks to be made between

the two. Corrections to the R_K ratio can be taken from $B^0 \to K^{*0} J/\psi (\to \ell^+\ell^-)$ modes and corrections to the R_{K^*} ratio can be taken from $B^+ \to K^+ J/\psi (\to \ell^+\ell^-)$ modes to avoid any correlations. This work will aim to show the compatibility of taking corrections from either mode for the $r_{J/\psi}$ ratios so that this procedure can be followed with confidence.

3.3.3 Yield Strategy

To determine the yields of the various signal and resonant modes one dimensional maximum likelihood fits to the B hadron mass are performed (Sec. 7). The fits are performed simultaneously to all modes, this approach allows the handling of backgrounds from non resonant $B^0 \to K^{*0}\ell^+\ell^-$ and $B^+ \to K^+\ell^+\ell^-$ decays which enter as backgrounds to each other. The fits are also performed simultaneously over trigger categories, q^2 bins and different data taking periods. The fits can be configured to either extract the signal yields or use measured efficiencies and extract R_K and R_{K^*}. Once values are determined their compatibilities with the SM or BSM theories can be evaluated using likelihood scans.

3.4 Data samples

The analysis relies on pp collision data from three main data taking periods, 2011 and 2012 (referred to as Run 1), 2015 and 2016 (referred to as Run 2p1) and 2017 and 2018 (referred to as Run 2p2). Each data taking period collected a different amount of data, sometimes at different collision energies. The software versions used for reconstruction and stripping of the data also changed throughout the years and this information is summarised in Tab. 3.1.

Table 3.1: Summary of the data samples used, along with relevant software versions used.

Year	$\mathcal{L}$ [fb^{-1}]	$\sqrt{s}$ [TeV]	Reconstruction version	Stripping Version
2011	1.1	7	14	21r1p1
2012	2.1	8	14	21r0p1
2015	0.3	13	15a	24r1p1
2016	1.7	13	16	28r1p1
2017	1.7	13	17	29r2
2018	2.2	13	18	34

During the data taking period different Trigger Configuration Keys (TCKs) were used to vary the thresholds of the L0 and HLT triggers. This means that even within the same year of data taking, the recording conditions were not always the same. The TCKs with which the largest amount of data was taken are `0x760037` (2011), `0x990042` (2012), `0x10600a2` (2015), `0x1138160f` (2016), `0x11611709` (2017) and `0x117a18a2` (2018). A full table of the TCKs used and the corresponding amount of recorded data can be found in Appendix A.1.

3.5 Simulated samples

Simulated samples are required to measure efficiencies, to obtain shapes for PDFs used in fits and as inputs to train MVAs. The creation of simulated data samples was discussed in Sec. 2.3. Simulated samples for each year are produced with the same version of reconstruction and stripping software as for collision data detailed in Tab. 3.1. It should be noted that as simulation poorly models PID-related variables any PID requirements imposed in the stripping are removed and the efficiency of these selections is determined in a data driven way. The most important simulated samples are those produced for the rare modes of interest, $B^0 \to K^{*0}\ell^+\ell^-$ and $B^+ \to K^+\ell^+\ell^-$, and the analogous decays where the lepton pair is produced from an intermediate J/ψ or $\psi(2S)$ resonance. The use of other simulated samples will be noted where relevant.

When producing simulated data a single TCK is used for each year and these are `0x40760037` (2011), `0x409f0045` (2012), `0x411400a2` (2015), `0x6138160F` (2016), `0x62661709` (2017) and `0x617d18a4` (2018).

3.6 Tuples

The data structures used to hold the data used for this analysis are referred to as tuples and based on ROOT's TTrees [122]. For each event various properties are measured and the values of these properties are then recorded within the tuple and referred to as branches. For example the branch named `B0_PT` contains the values of B meson p_{T} for each event. There are two main types of tuples which are important for this analysis, referred to as `MCDecayTuple` and `DecayTuple` described as follows:

- `MCDecayTuple`: The `MCDecayTuple` applies to simulated data only and includes the generator or truth level information of a simulated event. At this level

the event has not passed through the detector so contains no reconstruction information. This tuple is important for various efficiency steps and also for correcting differences between collision data and simulation at the generator level.

- `DecayTuple`: The `DecayTuple` applies to both collision and simulated data and contains the reconstructed information of an event and are the tuples used for the majority of the analysis. The `DecayTuple` of simulated data will also contain the truth level information from the corresponding `MCDecayTuple` if that event was successfully reconstructed, which is important for the estimation of the reconstruction efficiency.

Chapter 4

Selection

4.1 Geometric Acceptance

When collisions take place in the LHC particles are produced in the full 4π solid angle, however, LHCb only covers the forward region so only those particles which fall within LHCb's acceptance can be used for analysis. Simulated samples are used to evaluate the efficiency of this requirement, ε_{geo}, detailed in Sec. 6.2.

4.2 Stripping

In order to save disk space and to act as an efficient preselection, a stripping step is applied to all data collected online and the same selection is applied to all simulated samples. However, for reasons discussed in Sec. 3.5, PID criteria from the stripping are not applied to simulated data. The stripping requirements come from the `Bu2LLKmmLine` and `Bu2LLKeeLine2` lines for muons and electrons respectively and are listed in Tab 4.1. The stripping criteria can be summarised as:

- In order to form good quality vertices for the B meson and di-lepton resonances the χ^2/ndf (Number of Degrees of Freedom), which represents the χ^2 of their respective vertex fits, is required to be small.

- To avoid the selection of random tracks coming from a vertex other tracks are required to be significantly displaced from it. This is achieved by requiring the vertex χ^2 separation to be large, particularly so for the B meson since it has a relatively long lifetime compared to other particles resulting in a long flight

distance. Another variable which helps here is the χ^2_{IP} (primary): requiring this be to large ensures the impact parameter between a particle and its mother (denoted by primary) is sufficiently large.

- Since produced B mesons are highly boosted it is expected that the angle between the direction from the primary vertex to the B decay vertex and the momentum of the B is small. The variable known as DIRA is defined as the cosine of this angle, hence the requirement of accepting candidates which have DIRA values of very close to unity. It is also required that χ^2_{IP} (primary) for the B meson itself is small.

- A large mass window of $\pm 1500\,\text{MeV}/c^2$ around the B mass is required so that the sidebands can be studied, in particular the upper mass sideband can be used to train an MVA to reject combinatorial background and the bremsstrahlung tail for electrons stretches into the lower mass sideband. As this stripping line is also used by multiple analyses this wide mass window provides them with a large tolerance.

- Requirements on the multiplicity of events are reliant on the `nSPDHits` variable which records the number of hits in the SPD detector. High multiplicity events are removed as the detector performance suffers and the HLT algorithms begin to take too much time to run, this also aligns the stripping to the `nSPDHits` requirements in the L0 trigger.

- Two boolean variables are required to be true for muons at the stripping stage. `HasMuon` is true if a track has hits in the muon system, whilst `IsMuon` is true if the hits in the muon system are consistent with the PID hypothesis of the candidate being a muon.

- Loose PID requirements are applied to collision data samples only, largely to reduce the size of the samples. The only variables used at this stage are of the type $DLL_X\pi$, which were discussed in Sec. 2.2.7. The offline selection in Sec. 4.6 will further optimise PID requirements.

- For the R_K case the χ^2_{IP} (primary) criteria for the kaon is looser than for the kaon from the K^{*0} in the R_{K^*} case. This is because the stripping deals differently with kaons coming from a K^{*0} and kaons coming from a B meson. For the sake of consistency in the offline selection for R_K this criteria is tightened from $\chi^2_{\text{IP}} > 4$ to $\chi^2_{\text{IP}} > 9$.

Table 4.1: Summary of the `Bu2LLKmmLine` and `Bu2LLKeeLine2` stripping requirements, for R_K and R_{K^*}. PID requirements are only applied to collision data and are noted.

Type	**Requirement**
Global	`nSPDHits` $< 600(450)$ Run 1 (Run 2)
B	$\lvert m - m_B^{\mathrm{PDG}} \rvert < 1500\,\mathrm{MeV}/c^2$
	`DIRA` > 0.9995
	χ^2_{IP} (primary) < 25
	end vertex χ^2/ndf < 9
	primary vertex χ^2 separation > 100
K^{*0} (R_{K^*})	$\lvert m - m_{K^{*0}}^{PDG} \rvert < 300\,\mathrm{MeV}/c^2$
	$p_{\mathrm{T}} > 500\,\mathrm{MeV}/c$
	origin vertex χ^2/ndf < 25
K	$\mathtt{DLL}_{K\pi} > -5$ (only data)
	$\chi^2_{\mathrm{IP}}(primary) > 9(4)$ R_{K^*} (R_K)
	$p_{\mathrm{T}} > 400\,\mathrm{MeV}/c$ (only for R_K)
π (R_{K^*})	χ^2_{IP} (primary) > 9
$\ell\ell$	$m < 5500\,\mathrm{MeV}/c^2$
	end vertex χ^2/ndf < 9
	origin vertex χ^2 separation > 16
μ	`isMuon` (only data), `hasMuon`
	$p_{\mathrm{T}} > 300\,\mathrm{MeV}/c$
	$\chi^2_{\mathrm{IP}}(primary) > 9$
e	$\mathtt{DLL}_{e\pi} > 0$ (only data)
	$p_{\mathrm{T}} > 300\,\mathrm{MeV}/c$
	χ^2_{IP} (primary) > 9

4.3 Truth Matching for Simulated Samples

Reconstructed candidates from simulated samples are subject to a 'truth matching' procedure that varies depending on the use case. Truth matching is performed using the `TupleToolMCBackgroundInfo` (BKGCAT) tool which is an indicator of how well a particle decay has been reconstructed.

The BKGCAT tool provides a number of categories which are based on the following criteria:

- **A** - All final state particles used to construct the candidate are matched to decay products of the true MC particle. Note this does not necessarily have to be the signal decay.

- **B** - All final state MC particles originating from the true MC particle in **A** are matched to particles used to form the candidate.

- **C** - All final state particles which form the candidate have been correctly identified, so are assigned their true masses.

- **D** - The true MC particle in **A** is the signal decay of interest and may only differ by the presence or absence of intermediate resonances.

- **E** - Identical to **D** except all intermediate resonances, if there are any, are reconstructed.

- **F** - The true MC particle in **A** has a mass which does not exceed the mass of the mother of the signal decay by $100\,\mathrm{MeV}/c^2$

- **G** - At least one final state particle is a ghost[1].

- **K** - At least two final state daughters are matched to the same MC particle.

- **L** - At least one final state particle is matched to an MC particle which is the MC mother of an MC particle matched to another final state particle.

The BKGCAT tool assigns a value to the reconstructed B meson based on these criteria as follows:

- **0** (Signal) - A && B && C && D && E && !G && !K && !F

- **10**(Quasi-Signal) - A && B && C && D && !E && !G && !K && !F

- **50**(Low mass background) - A && C && F && !B && !G && !K && !F

- **60**(Ghosts) - G

Combinations of these categories are then used for different parts of the analysis:

- **Corrections to simulation (Sec. 5** - Performed with BKGCAT values of 0

[1]Defined as a reconstructed particle for which there is no corresponding particle at generator level.

&& 10 && 50

- **MVA training (Sec. 4.11 and Mass Fits(Sec. 7** - Performed with BKGCAT values of 0 && 10 && 50 && 60
- **Efficiency determination (Sec. 6** - Performed on signal modes with BKGCAT values of 0 && 10 && 50 && 60. Additionally it is required that $n-1$ of the final state particles are correctly identified using the `TRUE_ID` variable, which returns a unique number for each particle. It is also required that there is at most one fake track, designated by `TRUE_ID` $= 0$, amongst the final state particles.

When more than one candidate per event is present the candidate with the lower BKGCAT is chosen. If more than one candidate is present and they have the same BKGCAT one candidate is chosen at random to ensure that no bias is introduced.

Reconstructed candidates that pass the selection for background modes, often used to create background PDFs for the mass fits, are also truth matched checking particle `TRUE_ID` values are correct and that particle mother and daughter relationships are valid.

4.4 L0 Trigger categories

The analysis is performed in two trigger categories in each q^2 bin, henceforth referred to as L0I and L0L! (or sometimes L0I inclusive and L0L exclusive), where the second L (for lepton) in L0L can be E for electrons or M for muons. They correspond to the following L0 trigger lines:

- **L0I**: `L0Global_TIS` (B),
- **L0M!**: (`L0MuonTOS` (μ_1) `||` `L0MuonTOS` (μ_2)) `&& !L0Global_TIS` (B),
- **L0E!**: (`L0ElectronTOS` (e_1) `||` `L0ElectronTOS` (e_2)) `&& !L0Global_TIS` (B),

where having an exclamation mark at the end of trigger category is equivalent to requiring a logical NOT on the other trigger category. This allows for a secondary set of trigger categories to be defined, L0I! and L0L (or sometimes L0I exclusive and L0L inclusive), where the logical NOT requirements are reversed, which can be useful for cross-checks especially for the L0L category which is also used to derive some corrections.

This choice of trigger categories amounts to choosing L0I as the primary trigger category of the analysis and L0L! as the secondary trigger category. This choice is motivated by the fact that L0I trigger efficiency is more similar between electrons and muons than the L0L! trigger efficiency, leading to a main trigger category where the differences from the L0 trigger are minimised.

Due to the fact that only one TCK is used in simulation but multiple TCKs are used during each data taking period, in some years this can cause a misalignment between simulation and collision data for a significant fraction of the data. How this is handled for the various correction steps is outlined in Sec. 5.1. The thresholds relating to various trigger lines used in these categories were discussed in Sec. 2.2.6.

4.5 HLT Lines

Various HLT lines can be used to isolate events containing interesting candidates, and these lines may be of general use or specific to certain decay topologies or PID hypotheses. The line names and configurations are slightly different between data taking periods, Tab. 4.2 lists those used in this analysis. The lines are combined in a logical OR configuration for each data taking period with common lines.

For HLT1 in Run 1 the `HLT1TrackAllL0` line acts to select B decays by triggering if a single high p_{T} track of good quality which is also displaced from the interaction point is present in the event. In Run 2, the only change is that `HLT1TrackAllL0` becomes `HLT1TrackMVA`; these act very similarly except the latter employs an MVA instead of a cut-based approach.

For HLT2 all the used trigger lines are designed to find a n-body decay topology [123, 124]. However, with the intent of being very inclusive, they are not designed to find all the daughters of a given decay in all cases, so do select some partially reconstructed background. They are the `HLT2TopoLXBodyBBDT` lines where the `L` may be blank or stand for `Mu`, for muon, or `E`, for electron whilst the `X` can be either `2` or `3` for 2- or 3-body decays. The `BBDT` is an acronym for Bonsai Boosted Decision Tree [125] and is trained using a variety of kinematic and geometric variables, of particular use is the Distance Of Closest Approach (DOCA) which is used to decide whether the candidate is 2, 3 or 4 body in nature. The lines containing a `Mu` or `E` demand that one of the daughter particles is a muon or electron candidate, since this reduces the overall number of candidates it is possible to cut less tightly on the BBDT to gain efficiency. In Run 2 the choice of MVA classifier used in the

topological lines was re-optimised [126] and this time a MatrixNet classifier was chosen. From 2016 onwards the addition of `MuMu` or `EE` lines was introduced which required at least two of the daughters to be identified as muon or electron candidates.

Table 4.2: Summary of the HLT lines. Those lines which straddle both columns are not specific to muon or electron modes and can therefore be used in both.

	Muon modes	**Electron modes**
	Run 1	
HLT1	`Hlt1TrackAllL0`	
HLT2	`Hlt2Topo[2,3]BodyBBDT`	
	`Hlt2TopoMu[2,3]BodyBBDT`	`Hlt2TopoE[2,3]BodyBBDT`
	2015	
HLT1	`Hlt1TrackMVA`	
HLT2	`Hlt2Topo[2,3]Body`	
	`Hlt2TopoMu[2,3]Body`	
	2016,2017,2018	
HLT1	`Hlt1TrackMVA`	
HLT2	`Hlt2Topo[2,3]Body`	
	`Hlt2TopoMu[2,3]Body`	`Hlt2TopoE[2,3]Body`
	`Hlt2TopoMuMu[2,3]Body`	`Hlt2TopoEE[2,3]Body`

Similarly to the L0 trigger selection sometimes simulation and collision data are not aligned due to different TCK thresholds. For the HLT selection only HLT1 lines in 2016 and 2012 are affected. Since in these cases the simulated TCK always contains the loosest threshold values it is possible to align simulation and collision data by emulating the tighter TCKs on the relevant fractions of simulation.

4.5.1 2016 HLT 1 alignment

In 2016 changes of TCKs through the data taking year resulted in a misalignment between simulation and collision data in the `HLT1TrackMVA` line. The threshold for triggering this line is defined in Eq. 4.1, the parameter being varied with changing TCKs is denoted b. The logical OR of this criterion must be taken with the χ^2_{IP} and p_{T} variables from each of the final state particles of interest, for example for B^+ modes there are three expressions, one for the kaon and two for each lepton, and each of these three expressions must be subject to a logical OR. In the simulated

data $b = 1.1$ throughout, however, in data b takes three values: 1.1, 1.6 and 2.3. The fractions of data taken with each magnet polarity under each threshold is noted in Tab. 4.3. Using these fractions it is possible for simulation to emulate the tighter thresholds seen in collision data, resulting in an aligned TCK selection.

$$\begin{aligned} &(p_{\mathrm{T}}(\,\mathrm{GeV}) > 25 \quad \& \quad \log\chi^2_{IP} > 7.4) \\ &\qquad\qquad \text{OR} \\ &(\log\chi^2_{IP} > \frac{1}{(p_{\mathrm{T}}(\,\mathrm{GeV}) - 1)^2} + \frac{b}{25} \cdot (25 - p_{\mathrm{T}}(GeV)) + \log(7.4)) \end{aligned} \tag{4.1}$$

Table 4.3: Fractions of candidates with HLT threshold cuts corresponding to the parameter b on 2016 data.

Polarity	$b = 1.1$ [%]	$b = 1.6$ [%]	$b = 2.3$ [%]
MagDown	89.8	0.0	10.2
MagUp	30.0	15.5	54.5

4.5.2 2012 HLT 1 alignment

In 2012 changes in the TCK over the data taking period resulted in a misalignment between simulation and collision data for the `Hlt1TrackAllL0` line. The thresholds for each TCK in collision data and simulation can be seen in Tab. 4.4. Since the TCK in simulation corresponds to the loosest thresholds it is possible emulate the percentage of collision data that were taken with tighter TCKs in simulation resulting in an aligned TCK selection.

Table 4.4: `TCK` thresholds of the `HLT1AllTrackL0` line for the 2012 data taking period, (left column) magnet polarity down (right column) magnet polarity up. The gray row highlights the thresholds in simulation, and in blue the `TCK` values in collision data which have different thresholds to simulation.

HLT TCK value	% data	p [GeV]	p_{T} [GeV]	IP [mm]	IPχ^2	$\chi^2_{ndof}(track)$
MC	100.0	> 3.0	> 1.6	> 0.1	>16.0	<2.0
0x990042	19.2	> 3.0	> 1.6	> 0.1	>16.0	<2.0
0x97003d	15.19	> 10	> 1.7	> 0.1	>16.0	<1.5
0xa30044	14.67	> 3.0	> 1.6	> 0.1	>16.0	<2.0
0x990044	11.21	> 3.0	> 1.6	> 0.1	>16.0	<2.0
0xac0046	10.43	> 3.0	> 1.6	> 0.1	>16.0	<2.0
0x94003d	9.7	> 10	> 1.7	> 0.1	>16.0	<1.5
0xa90046	6.37	> 3.0	> 1.6	> 0.1	>16.0	<2.0
0x8c0040	6.31	> 10	> 1.7	> 0.1	>16.0	<2.0
0xab0046	5.24	> 3.0	> 1.6	> 0.1	>16.0	<2.0
0xa30046	1.12	> 3.0	> 1.6	> 0.1	>16.0	<2.0
0x860040	0.45	> 1.0	> 1.7	> 0.1	>16.0	<2.25
0x95003d	0.1	> 10	> 1.7	> 0.1	>16.0	<2.0
0x8e0040	0.05	> 10	> 1.7	> 0.1	>16.0	<2.0
0x7f0040	0.02	> 1.0	> 1.7	> 0.1	>16.0	<2.25
0xad0046	0.01	> 3.0	> 1.6	> 0.1	>16.0	<2.0
0x7e003a	0.01	> 1.0	> 1.3	> 0.1	>16.0	<2.25

HLT TCK value	% data	p [GeV]	p_{T} [GeV]	IP [mm]	IPχ^2	$\chi^2_{ndof}(track)$
MC	100.0	>3	>1.6	>0.1	>16.0	<2.0
0x990042	35.75	>3	>1.6	>0.1	>16.0	<2.0
0xa30044	17.52	>3	>1.6	>0.1	>16.0	<2.0
0x94003d	16.03	>10	>1.7	>0.1	>16.0	<1.5
0x97003d	11.98	>10	>1.7	>0.1	>16.0	<1.5
0xa10045	7.19	>3	>1.6	>0.1	>16.0	<2.0
0x990044	3.03	>3	>1.6	>0.1	>16.0	<2.0
0xa10044	2.67	>3	>1.6	>0.1	>16.0	<2.0
0x9f0045	2.66	>3	>1.6	>0.1	>16.0	<2.0
0xa30046	1.89	>3	>1.6	>0.1	>16.0	<2.0
0xac0046	0.88	>3	>1.6	>0.1	>16.0	<2.0
0x990043	0.26	>3	>1.6	>0.1	>16.0	<2.0
0xa20044	0.17	>3	>1.6	>0.1	>16.0	<2.0
0xad0046	0.01	>3	>1.6	>0.1	>16.0	<2.0
0x9a0042	0.01	>3	>1.6	>0.1	>16.0	<2.0
0x95003d	0.01	>10	>1.7	>0.1	>16.0	<2.0

4.6 Offline Selection

These criteria are heavily inspired by the Run 1 R_{K^*} analysis [71], listed in Tab. 4.5 and fall into four broad categories. The selections are the same for both B^+ and B^0 modes except that for B^+ modes the criteria regarding pions are not used. All criteria are applied to both collision data and simulated samples except for `ProbNN` variables which are not applied to simulated samples as they are poorly modelled. An efficiency correction is applied by the `PIDCalib` package described in Sec. 6.4 to account for this. PID and clone requirements are discussed further in Secs. 4.7 and 4.8 respectively. The remaining requirements are motivated as follows:

- To improve the quality of selected tracks and vertices the requirement on χ^2/ndf is tightened to select the best vertices and only tracks which have a low probability to be ghosts, parametrised by the `GhostProb` variable, are selected. Further, electron and muon tracks are required to be in the geometrical acceptance of the ECAL or muon system respectively.

- To ensure the calorimeter acceptance is modelled correctly, electrons must be in the ECAL acceptance by requiring $\texttt{region}^{L0CaloTool}_{\text{ECAL}} \geq 0$. There is also a region of the calorimeter where cells are not read out. To align the simulation to this, the region corresponding to $|\texttt{xProjection}^{L0CaloTool}_{\text{ECAL}}| < 363.6\,\text{mm}$ && $|\texttt{yProjection}^{\text{L0CaloTool}}_{\text{ECAL}}| < 282.6\,\text{mm}$, that is the x and y projection of the calorimeter, is vetoed. All the `L0CaloTool` variables are provided by LHCb's Calorimeter Objects Group.

- To ensure the K^{*0} (892) is selected rather than other kaon resonances, a mass window of $100\,\text{MeV}/c^2$ is taken around the known value of its mass [14].

- To ensure particles will have the relevant PID information it is required that the following booleans are true: `HasRich` for all tracks, `HasCalo` for electron tracks and `HasMuon` for muon tracks. `HasRich` requires that a particle has hits in the RICH system, `HasCalo` that the electron has hits in the calorimeter system and `HasMuon` that the muon has hits in the muon system.

4.7 Particle Identification

A loose PID selection is applied in the stripping for collision data (for simulated data this is not the case as efficiencies must be evaluated with `PIDCalib`) but by imposing tighter cuts a higher level of background suppression can be achieved. The main

Table 4.5: Summary of the selection requirements.

Type		**Requirement**
Quality	all tracks	$\chi^2/\text{ndf} < 3$
		$\texttt{GhostProb} < 0.4$
	e	$\texttt{InAccEcal==1}$
	μ	$\texttt{InAccMuon==1}$
	K/π	$\texttt{InAccMuon==1}$
	e	$\texttt{region}^{L0CaloTool}_{\text{ECAL}} \geq 0$
	e	$!(\lvert\texttt{xProjection}^{L0CaloTool}_{\text{ECAL}}\rvert < 363.6\,\text{mm}$
		$\&\&\ \lvert\texttt{yProjection}^{L0CaloTool}_{\text{ECAL}}\rvert < 282.6\,\text{mm})$
ID	K^{*0}	$\lvert m(K\pi) - m^{PDG}_{K^{*0}}\rvert < 100\,\text{MeV}/c^2$
PID	all	$\texttt{hasRich==1}$
	μ	$\texttt{hasMuon==1}$
	e	$\texttt{hasCalo==1}$
	K, π	$p_\text{T} > 250\,\text{MeV}/c$ ($> 500\,\text{MeV}/c$ for R_K), $p > 2000\,\text{MeV}/c$
	μ	$p_\text{T} > 800\,\text{MeV}/c$, $p > 3000\,\text{MeV}/c$
	e	$p_\text{T} > 500\,\text{MeV}/c$, $p > 3000\,\text{MeV}/c$
	K	$\texttt{ProbNNk} \cdot (1 - \texttt{ProbNNp}) > 0.05$
	π	$\texttt{ProbNNpi} \cdot (1 - \texttt{ProbNNk}) \cdot (1 - \texttt{ProbNNp}) > 0.1$
	μ	$\texttt{ProbNNmu} > 0.2$
	e	$\texttt{ProbNNe} > 0.2$
Clones	all tracks	$\theta(\ell_{1,2}, h) > 0.5$ mrad $\&\&$ $\theta(\ell_1, \ell_2) > 0.5$ mrad

additional PID variables which are used with respect to the stripping are ProbNN variables which are based on neural networks discussed in Sec. 2.2.7. For RUN 1 two different tunings are used, v2 and v3, the main difference between the two being that in v3 ghost tracks were not included in the background samples used to train the neural networks. In practice v2 is used for hadrons whilst v3 is used for leptons. In RUN 2 there is only one tuning which is always used. To achieve the maximum PID performance probabilities to identify and misidentify particles are combined, these criteria are inspired heavily by the RUN 1 R_{K^*} analysis [71]. For example the kaon criterion of $\texttt{ProbNNk} \cdot (1 - \texttt{ProbNNp})$ will be very small if both ProbNNk and $1 - \texttt{ProbNNp}$ are both small, that is if the particle is very unlikely to be a kaon and also very likely to be a proton. The following list of requirements is a combination of the tightest cuts required on each particle species from the stripping and offline selection found in Tabs 4.1 and 4.5. Criteria for the pion are only relevant to the B^0 modes.

$\boldsymbol{K}$: $\texttt{ProbNNk} \cdot (1 - \texttt{ProbNNp}) > 0.05 \quad \&\& \quad \mathrm{DLL}_{K\pi} > 0$

$\boldsymbol{\pi}$: $\texttt{ProbNNpi} \cdot (1 - \texttt{ProbNNk}) \cdot (1 - \texttt{ProbNNp}) > 0.1$

$\boldsymbol{\mu}$: $\texttt{ProbNNmu} > 0.2$

$\boldsymbol{e}$: $\texttt{ProbNNe} > 0.2 \quad \&\& \quad \mathrm{DLL}_{e\pi} > 2$

In order to obtain reasonable efficiency estimates for simulated data, where modelling of ProbNN variables is poor the PIDCalib package is used. This package relies on calibration collision data samples selected to contain large amounts of specific decay modes in order to obtain efficiencies for certain particle track types. These samples often have selection criteria imposed upon them to obtain clean samples and good quality tracks. Therefore, to obtain a reliable efficiency estimate the simulated and collision data samples must be aligned to these criteria, which are known as prior cuts. The only exception to this are the boolean variables UnbiasHLT1, MuonUnbiased and ElectronUnbiased. The first two variables are relevant to muons in RUN 1 and RUN 2 respectively, whilst the last variable is relevant to electrons. These variables are only applied to the PIDCalib calibration samples and when they are true require that hadron candidates did not trigger the event. This allows for the trigger efficiency and the misidentification efficiency of leptons to hadrons to be decoupled. The prior cuts required for each particle species is shown in Tab. 4.6, where again for B^+ modes the pion criteria are not relevant.

Table 4.6: `PIDCalib` prior cuts.

Particle	Prior cut
All	`nSPDHits` $< 600(450)$ for Run 1 (Run 2) `hasRICH` $== 1$ $\chi^2_{\text{track}}/ndof < 3$, `TRACK_GhostProb` < 0.4, $\chi^2_{IP} > 9$
K	$p_{\text{T}} > 250(400)\,\text{MeV}/c$ for B^0 (B^+), $p > 2000\,\text{MeV}/c$, `InAccMuon` $== 1$
π	$p_{\text{T}} > 250\,\text{MeV}/c$, $p > 2000\,\text{MeV}/c$, `InAccMuon` $== 1$
μ	`InMuonAcc` $== 1$, `IsMuon` $== 1$ `HLTUnbias` $== 1$ (`MuonUnbiased` $== 1$) Run 1 (Run 2) (Only on PIDCalib Calibration samples), $p_{\text{T}} > 800\,\text{MeV}/c$, $p > 3000\,\text{MeV}/c$
e	`hasCalo` $== 1$, `ElectronUnbiased` $== 1$ (Only on PIDCalib Calibration samples), $p_{\text{T}} > 500\,\text{MeV}/c$, $p > 3000\,\text{MeV}/c$

4.8 Clones

Clone tracks are the result of multiple tracks being assigned to the same particle as the output of tracking algorithms. In this case the track which is the longest, that is the one with the most hits over the most subdetectors, is taken to be used in subsequent studies as it has the highest information content. The remaining tracks are defined as clones. Within LHCb clone tracks are defined as those which share 70% of the hits that make up the tracks; for tracks which span multiple subdetectors this criterion is applied to each in turn and the tracks are called clones if the criterion is true in each case. An algorithm in LHCb known as the Clone Killer [127] is implemented in tracking procedures and acts to remove most clone tracks before data is passed on to analysts.

However, in some cases the Clone Killer does not catch all the clones in particular for electrons emitting bremsstrahlung. Bremsstrahlung can act to change the direction of the track and this can cause one set of hits in the VELO to be associated to multiple sets of hits in the T-stations. The momenta and PID hypotheses can differ between two clone tracks so this is an important source of background to consider. Since clone tracks have the same origin the opening angle between two clone tracks will be close to, if not exactly, zero. Therefore, to test for the presence of clone tracks the opening angle between tracks for all final state particles is calculated, in order to remove the clones a criterion that the angle between two final state particles should be $> 0.500\,\text{mrad}$ is applied.

4.9 Exclusive Backgrounds for $B^0 \to K^{*0}\ell^+\ell^-$

Backgrounds to $B^0 \to K^{*0}\ell^+\ell^-$ decays can originate from a number of sources: misidentification of hadrons or leptons, partially reconstructed decays which mimic the signal and over-reconstructed decays contributing to the combinatorial background. In the following, vetoes to deal with these backgrounds, where this is possible, will be discussed. Some backgrounds can not be vetoed without significant signal loss and are instead modelled in the mass fits. Each background is studied using collision data and simulated samples of the signal and the background modes considered. In the following only 2016 electron modes will be shown, which are largely representative of the general effect of the various vetoes that will be discussed.

4.9.1 Strategy

After the selection in Tab. 4.5 is made it is imperative that what remains is a genuine sample of only $B^0 \to K^{*0}\ell^+\ell^-$ decays and not polluted with other decays which may mimic the signal well enough to pass the selection. It is therefore crucial that any backgrounds which may contribute are studied and suppressed as much as possible whilst retaining the maximum amount of signal possible. Usually these backgrounds differ by some kinematic variable on which a criterion can be imposed to significantly reduce them. If this is not the case, the background must be modelled in the final fit.

The remaining background in collision data after the selection is made, N_B, can be estimated by Eq. 4.2. This is computed using the efficiency of the background in question, ϵ_B, evaluated on simulated data after the selection. This is then scaled by a normalisation factor α_{norm} based on the muonic control mode and the ratio of the production fraction of the b-hadron of the background to that of the signal. In Eq. 4.2 $\mathcal{B}_x$ denotes the branching fraction of the background or control mode, f_x denotes the production fraction of the background or signal hadron, $N_{control}$ denotes the yield of the $B^0 \to K^{*0}J/\psi(\to \mu^+\mu^-)$ (or $B^+ \to K^+J/\psi(\to \mu^+\mu^-)$ for B^+ modes) determined from the ${}_s\mathcal{P}lot$ method, see Appendix B, and $\epsilon_{control}$ is the efficiency of the selection on corrected $B^0 \to K^{*0}J/\psi(\to \mu^+\mu^-)$ simulated samples.

$$N_B = \epsilon_B \cdot \alpha_{norm} = \epsilon_B \cdot \frac{\mathcal{B}_B}{\mathcal{B}_{control}} \cdot \frac{N_{control}}{\epsilon_{control}} \cdot \frac{f_B}{f_d} \tag{4.2}$$

For Eq. 4.2 to be valid requires that ϵ_B is properly defined, in some cases due to

low statistics in the simulated sample this efficiency may be zero. In order to obtain a meaningful result in these cases an upper limit, at a certain confidence level, can be set using Bayes' Theorem by calculating a PDF for the background efficiency. This technique is only used if the number of events remaining is < 20, details of this method can be found in Appendix C.1.

When dealing with a particular background as well as considering a dedicated veto the suppression that full selection is producing is also examined. This is split into categories to disentangle the differing parts of the selection. In Figs **??** to **??** keys are used in the x-axis to refer to these different parts of the selection and they are explained in Tab. 4.7.

Table 4.7: Differing parts of the full selection as referred to in the figures in Secs. 4.9 and 4.10.

Part of Selection	**Selection**
L0	Cuts defined in Sec. 4.4
HLT1	Cuts defined in Tab. 4.2 in the HLT1 rows
HLT2	Cuts defined in Tab. 4.2 in the HLT2 rows
m_Kst (Sec.4.9 only)	Cut defined in Tab. 4.5 in the ID row
Quality	Cuts defined in Tab. 4.5 in the Quality row
PID_strip	Cuts defined in Tab. 4.5 in the PID row, not containing `ProbNN`
ProbNNk	`ProbNNk (1-ProbNNp) > 0.05`
ProbNNPi (Sec.4.9 only)	`ProbNNpi(1 - ProbNNk)(1-ProbNNp)> 0.1`
ProbNNl	`ProbNNmu > 0.2` for muons or `ProbNNe > 0.2` for electrons
q2	Cuts defined in Sec. 3.1
m_b	Cuts defined in Sec. 7.6
MVA	Cuts defined in Tab. 4.22

4.9.2 Misidentification

As $B^0 \to K^{*0}\ell^+\ell^-$ decays are reconstructed from charged particles misidentification of pions, kaons, protons, muons and electrons must be considered. The misidentification (misID) rate can reach up to a few percent, depending on the particle type, and if the branching ratio of the misidentified decay is large enough this can cause significant pollution. Typically by changing the mass hypothesis on a particular misidentified particle to its true identity an intermediate resonance can be reconstructed allowing a veto in a certain mass window to be applied. In the interest of keeping the signal efficiency high often the whole resonance is not vetoed but it is possible to gain background rejection by applying stricter PID requirements on the particles which have been misidentified. Tab. 4.8 summarises the sources of misidentification for $B^0 \to K^{*0}\ell^+\ell^-$ decays.

Table 4.8: Relevant mis-identifications of final state particles. Only the dominant decay modes are considered here, i.e. large branching ratio and/or similarity to the signal decay. The branching ratios are taken from the PDG [14]. Where the whole q^2 region is affected the first number corresponds to taking B_{SIG} from the signal mode, while the number in brackets takes B_{SIG} from the control mode.

MisID	**Decay mode**	$\sim \mathcal{B}_{BKG}/\mathcal{B}_{SIG}$	**Important for q^2 region**
$K \to \pi$	$B_s^0 \to (\phi \to K^+K^-)(J/\psi \to \ell^+\ell^-)$	0.6	J/ψ
	$B_s^0 \to (\phi \to K^+K^-)\ell^+\ell^-$	0.6	signal
$p \to \pi$	$\Lambda_b^0 \to pK(J/\psi \to \ell^+\ell^-)$	0.4	J/ψ
	$\Lambda_b^0 \to pK\ell^+\ell^-$	B_{BKG} unknown	signal
$K \leftrightarrow \pi$ swap	$B^0 \to (K^{*0} \to K^+\pi^-)(J/\psi \to \ell^+\ell^-)$	N/A	J/ψ
	$B^0 \to (K^{*0} \to K^+\pi^-)\ell^+\ell^-$	N/A	signal
$\pi \to \ell$	$B^0 \to (D^{*-} \to (\overline{D}^0 \to K^+\pi^-)\pi^-)\ell^+\nu_\ell$	180 (2.4)	all
	$B^0 \to (\overline{D}^0 \to K^+\pi^-)\pi^-\ell^+\nu_\ell$	250 (3.2)	all
	$B^0 \to (D^- \to (K^{*0} \to K^+\pi^-)\pi^-)\ell^+\nu_\ell$	360 (5.0)	all
$\pi \to \ell$ double	$B^0 \to (K^{*0} \to K^+\pi^-)(X \to \pi^+\pi^-)$	B_{BKG} unknown	all
	$B^0 \to (D^- \to (K^{*0} \to K^+\pi^-)\pi^-)\pi$	4.0 (0.05)	all
$K \to \ell$	$B_s^0 \to (D_s^- \to (K^{*0} \to K^+\pi^-)K^-)\ell^+\nu_\ell$	B_{BKG} unknown	all
$h \leftrightarrow \ell$ swap	$B^0 \to (K^{*0} \to K^+\pi^-)(J/\psi \to \ell^+\ell^-)$	N/A	all
	$B^0 \to (K^{*0} \to K^+\pi^-)(\psi(2S) \to \ell^+\ell^-)$	N/A	all

$K \to \pi$ misID

The source of $K \to \pi$ misID comes from $B_s^0 \to \phi(\to K^+K^-)\ell^+\ell^-$ or $B_s^0 \to \phi(\to KK)J/\psi(\to \ell\ell)$ decays, where one of the kaons is misidentified as a pion. The mass shift involved in the $K \to \pi$ misID is larger than the $m(B_s^0) - m(B^0)$ mass shift, but due to the bremsstrahlung experienced by electrons this background will peak underneath the signal. The K^{*0} mass window applied in Tab. 4.5 already removes a large amount these types of decays as can be seen in Fig. 4.1. To remove residual background events a requirement on $m(KK_{\to\pi})$, the invariant mass of the kaon system when one kaon has been misidentified as a pion, can be applied. A clear peak at the ϕ mass can seen in this distribution also shown in Fig. 4.1. In order to remove this background whilst keeping the signal efficiency as high as possible a pion PID requirement is also applied in the region to be vetoed so the selection is formulated as $!(m(KK_{\to\pi}) < 1040\,\text{MeV}/c^2$ && `Pi_ProbNNpi` $< 0.8)$.

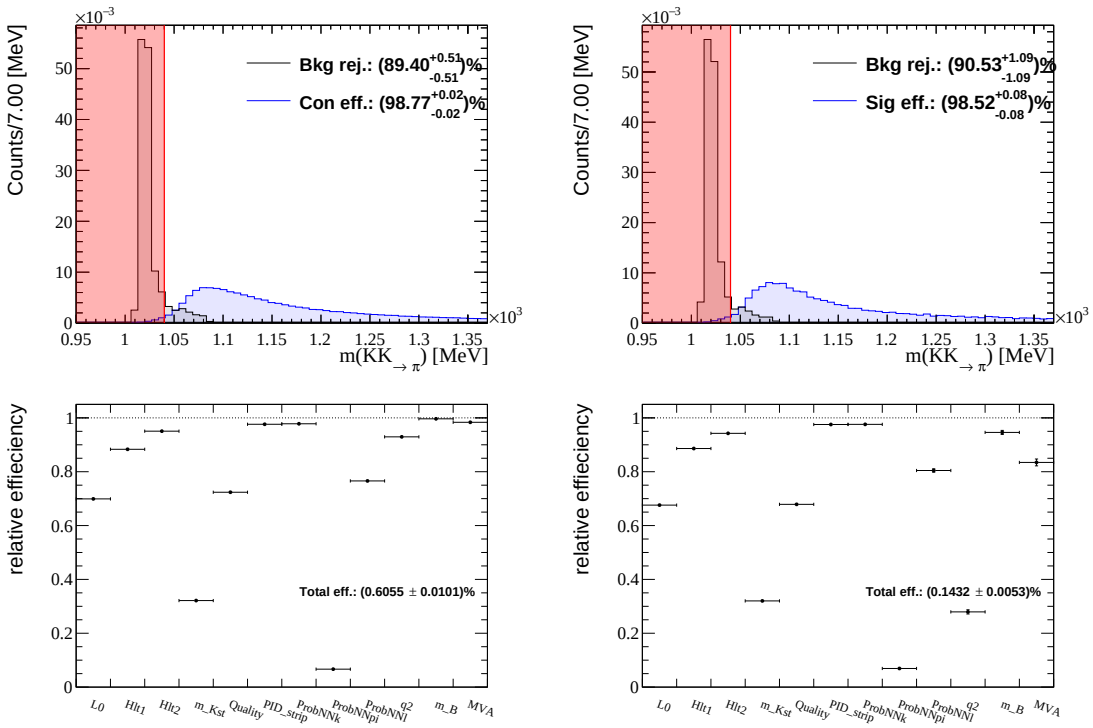


Figure 4.1: Top: Veto against $\phi \to KK$ decays in 2016 data. The black/blue distribution shows the background/signal MC sample. The red area is rejected by the veto. Left: $B_s^0 \to \phi(J/\psi \to ee)$ resonant mode plotted in $J/\psi\, q^2$ region. Right: $B_s^0 \to \phi ee$ rare mode plotted in central-q^2 region. Background events peak in $m(KK_{\to\pi})$ at the ϕ mass. Bottom: (left) Offline selection efficiency in 2016 for $B_s^0 \to \phi(J/\psi \to ee)$ (right) Offline selection efficiency for $B_s^0 \to \phi ee$ in central-q^2 . The cuts defined by the x-axis labels are defined in Tab. 4.7 and these cuts are applied separately.

$p \to \pi$ misID

Final states of the form $pX\ell^+\ell^-$ (where $X = K, \pi$) can mimic the signal decay if the proton is misidentified as a kaon or a pion. Due to baryon number conservation present in the SM this final state must originate from a baryonic decay. In the region around the B^0 mass the most likely candidate is the Λ_b^0. Since $\Lambda_b^0 \to p\pi\ell^+\ell^-$ contains a $b \to d$ transition it is suppressed with respect to $\Lambda_b^0 \to pK\ell^+\ell^-$, hence, $p \to \pi$ misID is considered. The resonant decay $\Lambda_b^0 \to pKJ/\psi(\to \ell^+\ell^-)$ must also be considered for the control mode. When assigning the pion mass hypothesis to the proton, the background has a broad distribution under the signal and has no hadronic resonances so an efficient veto is not possible. The selection requirements on the $m(K\pi)$ mass window and PID criteria suppress this background somewhat as can be seen in Fig. 4.2. Even with this suppression residual events still survive under the signal peak and are modelled in the final mass fits in Sec. 7 for J/ψ modes; for non-resonant modes, the relative contribution is much smaller and therefore this background is not expected to be significant.

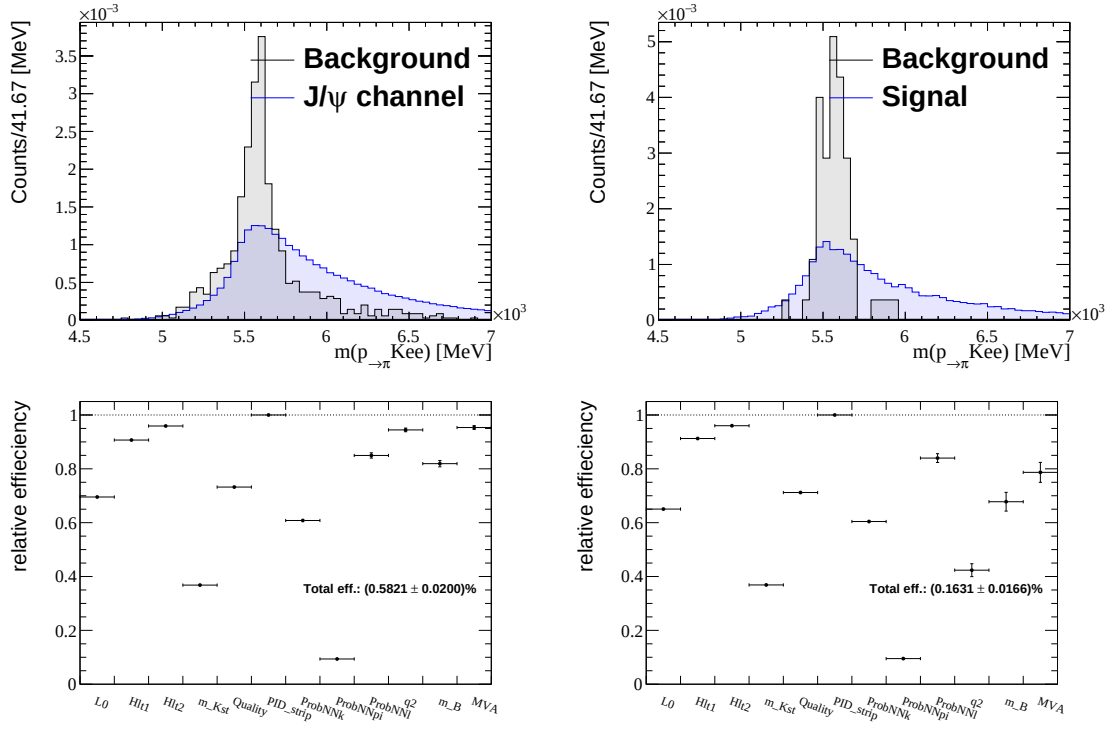


Figure 4.2: Top: (Black) Background and (Blue) signal shapes in $m(p_{\to\pi}Kee)$ for (left) $\Lambda_b^0 \to pKJ/\psi(\to \ell^+\ell^-)$ (right) $\Lambda_b^0 \to pK\ell^+\ell^-$. Right: Offline selection efficiency for (left) $\Lambda_b^0 \to pKJ/\psi(\to \ell^+\ell^-)$ in $J/\psi\, q^2$ region (right) $\Lambda_b^0 \to pK\ell^+\ell^-$ in central-q^2 region. The cuts defined by the x-axis labels are defined in Tab. 4.7 and these cuts are applied separately.

$K \leftrightarrow \pi$ swaps

When a kaon and pion originating from a K^{*0} are both misidentified as each other ($K \leftrightarrow \pi$ swaps), the resulting invariant mass no longer peaks at the K^{*0} mass making this a source of background. This type of background is heavily suppressed by PID requirements and after the whole selection the remaining pollution is around 0.1%. When looking at the distribution of $m(K_{\to\pi}\pi_{\to K})$ a broad distribution that completely overlaps the K^{*0} mass region results from this background. Introducing a veto around the K^{*0} mass was considered, as can be seen in Fig. 4.3, however, it was deemed to reject too much of the signal and therefore is not applied in the final selection. Mismodelling in the fit due to this remaining source of misidentification is assigned as a systematic uncertainty.

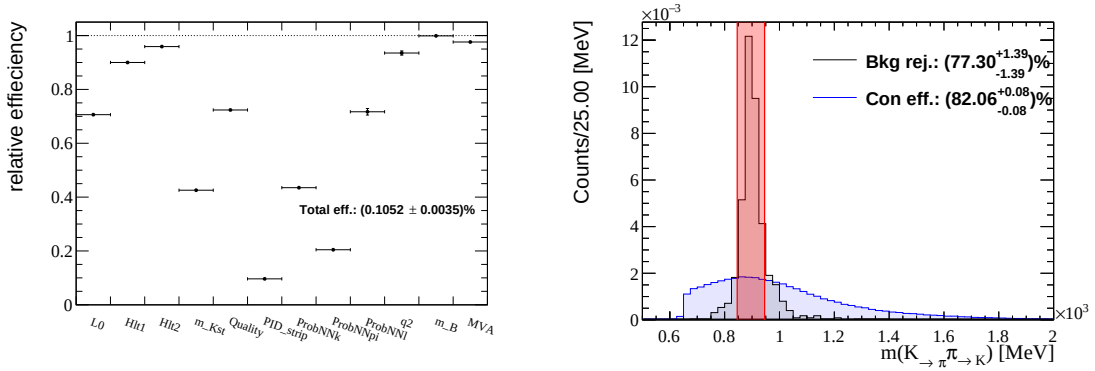


Figure 4.3: Left: Offline selection efficiency for $K \leftrightarrow \pi$ swap background in $J/\psi\, q^2$ region. The cuts defined by the x-axis labels are defined in Tab. 4.7 and these cuts are applied separately. Right: Shape in $m(K_{\to\pi}\pi_{\to K})$, the black/blue distribution shows the background/signal MC sample. The shown veto is not included in the final selection.

In addition, sources of this background from the more numerous $B^0 \to K^{*0} J/\psi\,(\to \ell^+\ell^-)$ decays can leak into the central-q^2 bin as a consequence of resolution effects. Applying a veto in this case, as seen in Fig. 4.4, would also lead to an unacceptable loss of signal. Since the amount of leakage is small and only contributes to the lower mass region a veto is not applied in this case either.

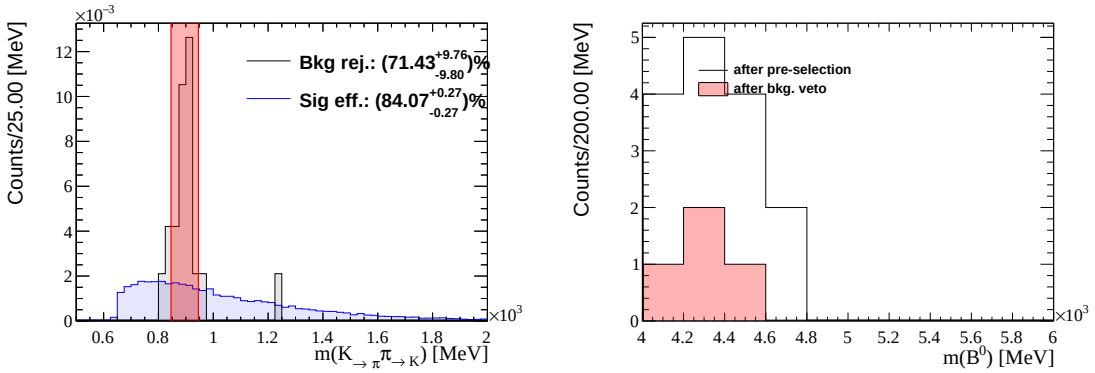


Figure 4.4: Left: Shape of $m(K_{\to\pi}\pi_{\to K})$ in central-q^2 , the black/blue distribution shows the background/signal MC sample. The shown veto is not included in the final selection. Right: Distribution of $K \leftrightarrow \pi$ swap in $m(K\pi ee)$.

$\pi \to \ell$ misID

A number of decays contribute to $\pi \to \ell$ misID, they are semi leptonic and proceed via D meson resonances. The main decay modes contributing are $B^0 \to (D^{*-} \to (\overline{D}^0 \to K^+\pi^-)\pi^-)\ell^+\nu_\ell$, $B^0 \to (\overline{D}^0 \to K^+\pi^-)\pi^-\ell^+\nu_\ell$ and $B^0 \to (D^- \to (K^{*0} \to K^+\pi^-)\pi^-)\ell^+\nu_\ell$. Neutrinos cannot be detected in LHCb so if a pion is misidentified as a lepton of the same type produced in any of these decays they will mimic the signal decay. These decays have much higher branching ratios than the signal decays and due to the undetected neutrinos tend to populate the lower mass sideband. Without a dedicated veto $\overline{D}^0 \to K^+\pi^-$ modes are already very strongly suppressed by the K^{*0} mass cut. To suppress these backgrounds a strict PID requirement (`L_ProbNNL` > 0.8, where `L` is `E` or `Mu`) is applied within a $30\,\mathrm{MeV}/c^2$ window around the D^0 mass in $m(K^+\pi^-{}_{\to\ell})$ or the D^- mass in $m(K^+\pi^-\pi^-{}_{\to\ell})$. The D meson mass peaks in 2012 electron and muon data can be seen in Fig. 4.5, before and after the veto is applied. It can be seen that the veto significantly reduces the peaks and the signal efficiency, evaluated on $B^0 \to K^{*0}e^+e^-$ simulated samples, is $> 98\%$. Purely hadronic decays such as $B^0 \to (K^{*0} \to K^+\pi^-)(\rho \to \pi^+\pi^-)$ and $B^0 \to (D^- \to (K^{*0} \to K^+\pi^-)\pi^-)\pi$ can also mimic the signal if two pions are misidentified both as muons or electrons. In this case the PID requirements strongly suppress them and they have negligible contribution, for the latter decay the $m(K^+\pi^-\pi^-{}_{\to\ell})$ veto adds further suppression.

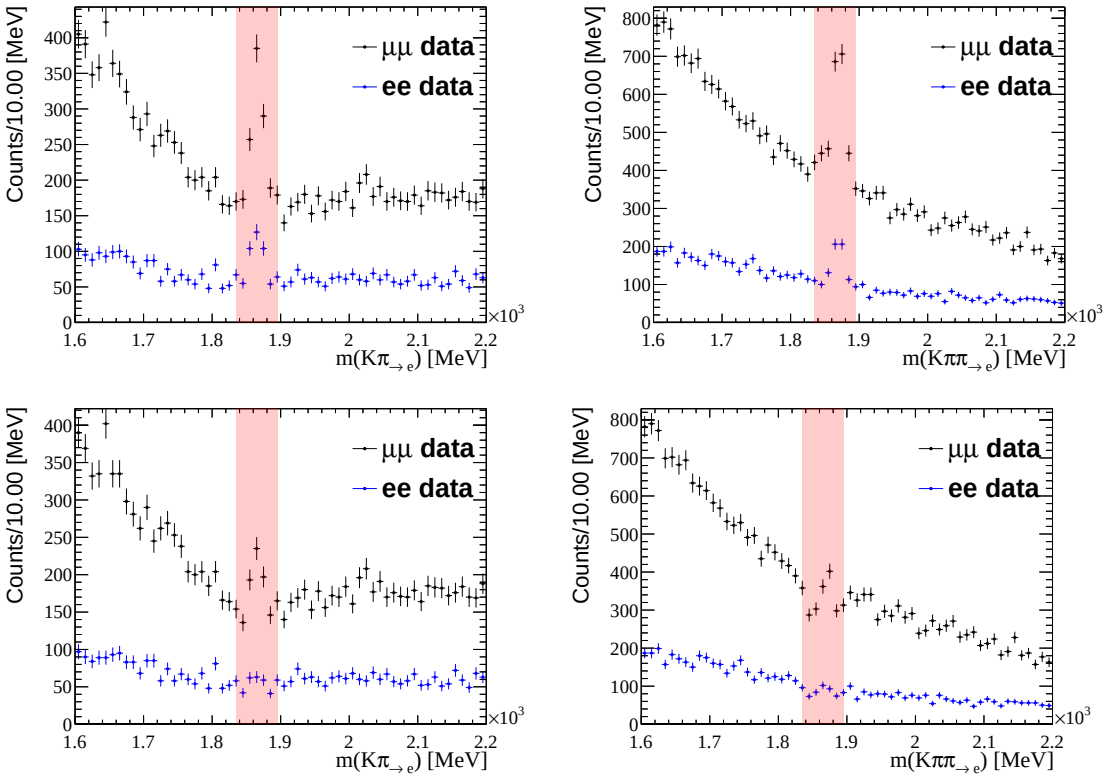


Figure 4.5: 2012 data in central-q^2 for muon/electron in black/blue. The left (right) plots show the $m(K^+\ell^-_{\to\pi})$ ($m(K^+\pi^-\ell^-_{\to\pi})$) mass system. Top: Distributions shown after selection in Tab. 4.5. Bottom: Additional PID requirement around D mass applied.

$h \leftrightarrow \ell$ swaps

It is possible that a hadron and a lepton can be misidentified as each other ($h \leftrightarrow \ell$ swap). If these particles come from a resonant decay mode, $m(\ell\ell)$ will no longer display a peak at the resonance mass. To catch all the cases $K \leftrightarrow \ell$ and $\pi \leftrightarrow \ell$ misidentifications must be considered. The selection suppresses this type of background somewhat as seen in Fig. 4.6. For muons when looking at the distributions of $m(\mu_{\rightarrow K}\mu)$ and $m(\mu\mu_{\rightarrow \pi})$ a sharp peak at the J/ψ or $\psi(2S)$ mass can be observed. A veto of $\pm 60\,\mathrm{MeV}/c^2$ around the resonance mass can then be applied as seen in Fig. 4.7. For electrons the same approach will not work as the resonance mass is not as well resolved due to bremsstrahlung effects. In this case the background is seen to peak in the J/ψ or $\psi(2S)$ constrained fit in the $m(K_{\rightarrow e}\pi e_{\rightarrow K})$ and $m(K\pi_{\rightarrow e}e_{\rightarrow \pi})$ distributions with a much better resolution. A veto of $\pm 60\,\mathrm{MeV}/c^2$ around the B^0 mass can then be applied. Comparisons of the vetoes for electrons in both types of mass system can be seen in Fig. 4.8.

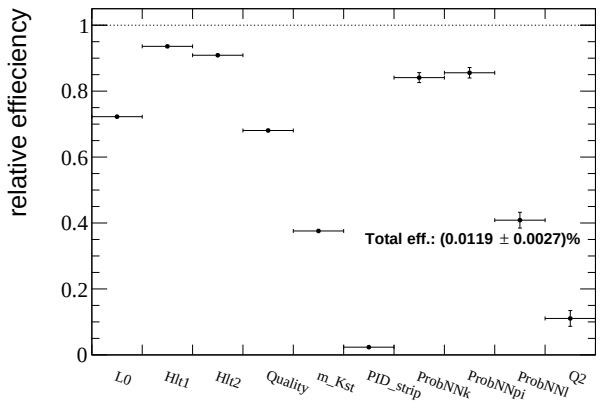


Figure 4.6: Offline selection efficiency for $h \leftrightarrow e$ swap background in central-q^2 region. The cuts defined by the x-axis labels are defined in Tab. 4.7 and these cuts are applied separately.

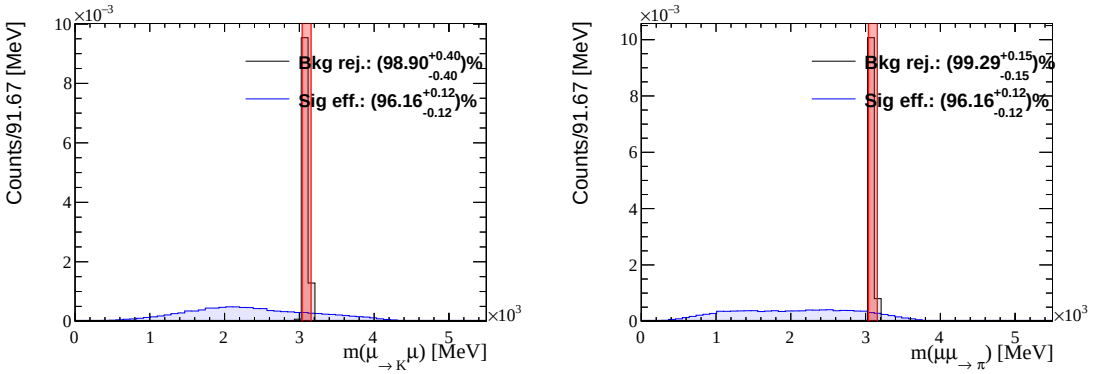


Figure 4.7: Veto against $K \leftrightarrow \mu^+$ (left) and $\pi \leftrightarrow \mu^-$ (right), the black/blue distribution shows the background/signal MC sample.

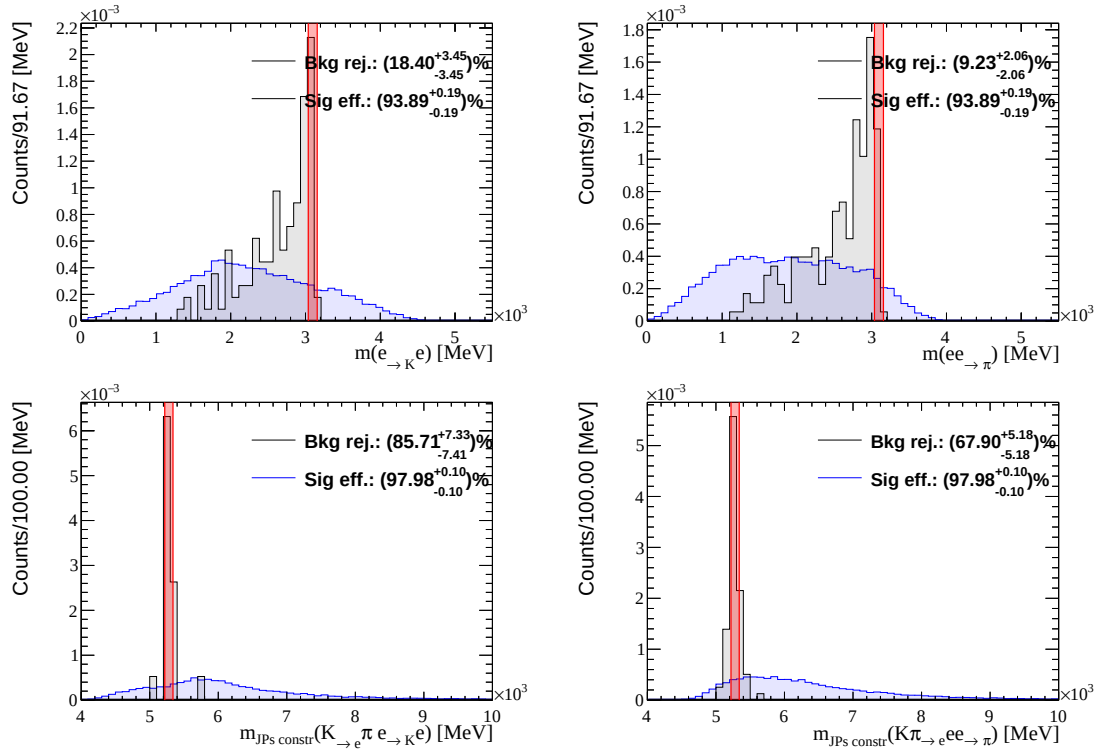


Figure 4.8: Veto against $K \leftrightarrow e^+$ (left) and $\pi \leftrightarrow e^-$ (right), the black/blue distribution shows the background/signal MC sample. Top: Distributions of $m(e_{\to K}e)$ and $m(ee_{\to \pi})$ showing the poor resolution due to bremsstrahlung. Bottom: Distribution of J/ψ mass constraint fits showing a much better resolution.

4.9.3 Partially Reconstructed Backgrounds

Partially reconstructed decays contributing as backgrounds to $K\pi\ell\ell$ final states result from ≥ 5 body final states in which one or more particles are not reconstructed by the LHCb detector and the remaining final state particles mimic the signal decay. As these backgrounds have as mother particles B and D mesons they will appear in the lower mass sideband of $m(K\pi\ell\ell)$. Due to bremsstrahlung this is a particular issue for electrons so it is important to suppress these types of background as much as possible. The most relevant partially reconstructed backgrounds are listed in Tab. 4.9.

Table 4.9: Possible partially-reconstructed decay modes which can contribute. Only the relevant decays are considered here, *i.e.* those having a large branching ratio and/or similarity to the signal decay. The branching ratios are taken from the PDG [14]. Where the whole q^2 region is affected the first number corresponds to taking B_{SIG} from the signal mode, while the number in brackets takes B_{SIG} from the control mode.

Decay mode	$\frac{\mathcal{B}_{BKG}}{\mathcal{B}_{Signal}}$	**Important for q^2 region**
$B^0 \to (D^- \to (K^{*0} \to K^+\pi^-)\ell^-\bar{\nu}_\ell)\ell^+\nu_\ell$	1200 (16)	all
$B^+ \to K^+\pi^+\pi^-(J/\psi \to \ell^+\ell^-)$	1.0	J/ψ
$B^+ \to K^+\pi^+\pi^-\ell^+\ell^-$	0.7	signal

$B^0 \to (D^- \to (K^{*0} \to K^+\pi^-)\ell^-\bar{\nu}_\ell)\ell^+\nu_\ell$

The decay $B^0 \to (D^- \to (K^{*0} \to K^+\pi^-)\ell^-\bar{\nu}_\ell)\ell^+\nu_\ell$ mimics the signal decay as LHCb cannot detect neutrinos. This decay has a large branching ratio and efficiently passes PID and $m(K\pi)$ mass window requirements in the selection so would be a significant contribution to the lower mass sideband. However, the $m(K\pi\ell^-)$ system resulting from this decay must have a mass lower than the D^- mass so an efficient veto of $m(K\pi\ell^-) < 1780\,\text{MeV}/c^2$ can be applied as seen in Fig. 4.9.

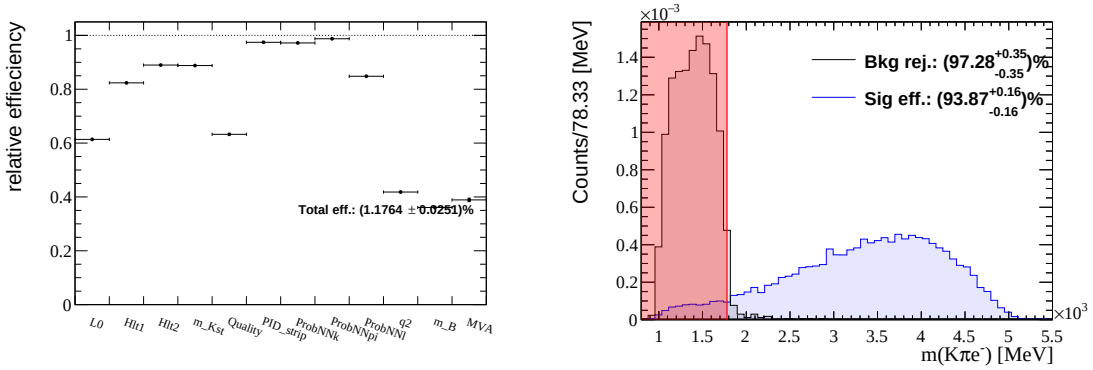


Figure 4.9: Left: Offline selection efficiency for $B^0 \to (D^- \to (K^{*0} \to K^+\pi^-)e^-\bar{\nu}_\ell)e^+\nu_\ell$ in central-q^2 region. Right: Shape in $m(K\pi e^-)$, the black/blue distribution shows the background/signal MC sample. The cuts defined by the x-axis labels are defined in Tab. 4.7 and these cuts are applied separately.

$B^+ \to K\pi\pi J/\psi(\to \ell^+\ell^-)$ and $B^+ \to K\pi\pi\ell^+\ell^-$

The decays $B^+ \to K\pi\pi J/\psi(\to \ell^+\ell^-)$ and $B^+ \to K\pi\pi\ell^+\ell^-$ can mimic the signal decay if one pion is not reconstructed. The $K\pi\pi$ system may be produced directly or via a resonance such as $K_1(1270)$. For the J/ψ control mode a veto on the J/ψ constrained mass system of $m(K\pi\ell\ell)$ can be applied, as seen in Fig. 4.10. However, as this sculpts the combinatorial background shape, no veto is applied and this background is modelled in the fit to control modes. In the signal region a dedicated MVA classifier is trained to reject this type of background and is discussed in Sec. 4.11. However, as the MVA is not as performant as the one for reducing combinatorial background, the residual partially reconstructed background must be accounted for in the final fits. This MVA is also applied to $B^+ \to K\pi\pi J/\psi(\to \ell^+\ell^-)$ decays with a loose selection, as a dedicated MVA is not trained for this mode.

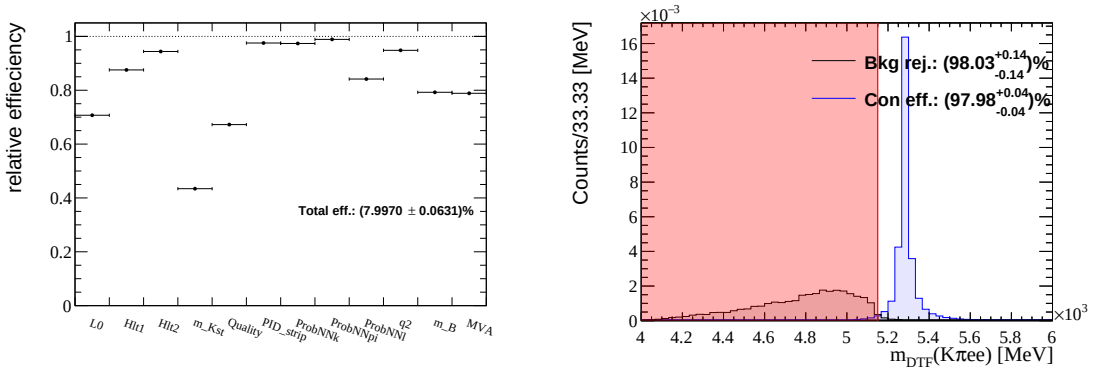


Figure 4.10: Left: Offline selection efficiency for $B^+ \to K^+\pi^+\pi^-(J/\psi \to \ell^+\ell^-)$ in $J/\psi\, q^2$ region. The cuts defined by the x-axis labels are defined in Tab. 4.7 and these cuts are applied separately. Right: Shape in $m_{JPsconstr}(K\pi ee)$, the black/blue distribution shows the background/signal MC sample. The displayed veto is not used.

4.9.4 Over-Reconstructed Background

An over-reconstructed decay is a ≤ 3 body decay which has a low momentum particle added to its decay products in order to mimic the signal decay. The main source of over-reconstructed background is from $B^+ \to K^+\ell^+\ell^-$ decays which have a low momentum pion from the rest of the event reconstructed along with it. This background can be considered combinatorial in nature and is heavily suppressed by an MVA classifier trained to reject combinatorial background, discussed in Sec. 4.11. A veto of $> 5100\,\text{MeV}/c^2$ is also applied to whichever of $m(K\ell\ell)$ and $m(K_{\to\pi}\ell\ell)$ is greater. This veto corresponds to $\sim m(B^0) - m(\pi)$ where very little signal would be expected as seen in Fig. 4.11.

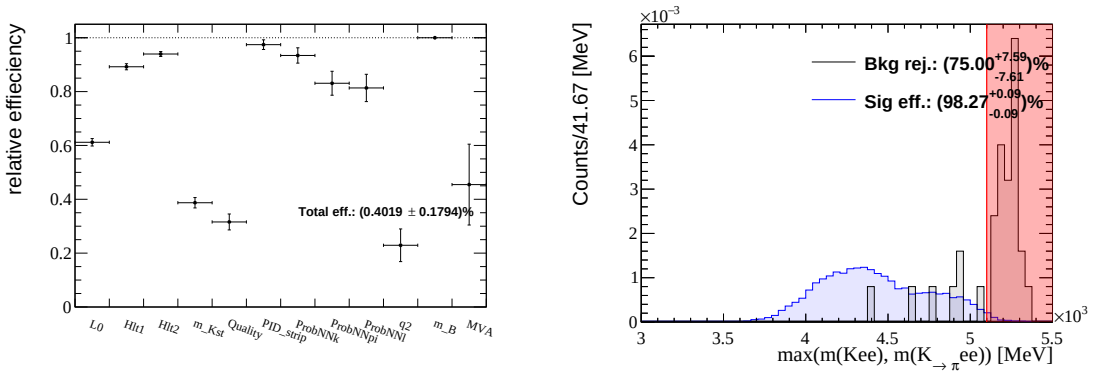


Figure 4.11: Left: Offline selection efficiency for $B^+ \to Kee$ in central-q^2 region. The cuts defined by the x-axis labels are defined in Tab. 4.7 and these cuts are applied separately. Right: Shape in $\max(m(Ke^-e^+), m(K_{\to\pi}e^-e^+))$, the black/blue distribution shows the background/signal MC sample.

4.9.5 Other

One other type of background which must be considered is that of $B_s^0 \to K^{*0}\ell^+\ell^-$ decays. These decays only differ by the mother particle and no veto has been found to suppress them in an acceptable manner. Therefore this background must be considered in the mass fits of Sec. 7 for J/ψ modes; for non-resonant modes with the relative contribution presented, this background is not expected to be important.

4.9.6 Summary of backgrounds considered for fits

As has been discussed in the previous sections a number of backgrounds either cannot be fully vetoed or are not able to be vetoed at all and therefore must be considered in the final fits. Tab. 4.10 shows the relevant decay modes and which regions of q^2 they must be considered for.

Table 4.10: Sources of background which must be taken into account in fits, along with the q^2 region of importance.

Decay mode	**Important for fit in q^2 region**
$B_s^0 \to K^{*0} J/\psi (\to \ell^+\ell^-)$	J/ψ
$\Lambda_b^0 \to pK J/\psi (\to \ell^+\ell^-)$	J/ψ
$B^+ \to K\pi\pi J/\psi (\to \ell^+\ell^-)$	J/ψ
$B_s^0 \to K^{*0}\psi(2S)(\to \ell^+\ell^-)$	$\psi(2S)$
$\Lambda_b^0 \to pK\psi(2S)(\to \ell^+\ell^-)$	$\psi(2S)$
$B^+ \to K\pi\pi\psi(2S)(\to \ell^+\ell^-)$	$\psi(2S)$
$B^+ \to K\pi\pi\ell^+\ell^-$	signal

4.9.7 Summary of exclusive background selection for R_{K^*}

The various vetoes discussed in the previous section are presented in Tab. 4.11 with the relevant q^2 region and decay modes where applicable. The total signal efficiency for $B^0 \to K^{*0}e^+e^-$ in both low and central-q^2 regions is shown in Fig. 4.12. The expected number of background events in the various q^2 regions for all the backgrounds considered here can be found in Appendix C.2

Table 4.11: Summary of the exclusive background selection requirements.

Background	**Requirement**	**Applied for**
$B_s^0 \to \phi\ell^+\ell^-$	$!(m(K(\pi \to K)) < 1040\,\text{MeV}$ && `Pi_ProbNNpi` $< 0.8)$	all
$B^0 \to (D^0 \to K\pi)\pi^-\ell^+\nu$	$!(\lvert m(K^+\ell^-_{\to\pi}) - m^{PDG}(D^0)\rvert < 30\,\text{MeV}$ && `L_ProbNNl` $< 0.8)$	all
$B^0 \to (D^- \to K\pi\pi)\ell^+\nu$	$!(\lvert m(K^+\pi^+\ell^-_{\to\pi}) - m^{PDG}(D^-)\rvert < 30\,\text{MeV}$ && `L_ProbNNl` $< 0.8)$	all
$h \leftrightarrow \ell$ swap	$!(\lvert m(\mu_{\to h}\mu) - m^{PDG}_{J/\psi\,,(\psi(2S))}\rvert < 60\,\text{MeV}$ && `M_ProbNNmu` $< 0.8)$	all $\mu\mu$
	$!(\lvert m_{J/\psi\,,(\psi(2S))constr.}(h_{\to e}he_{\to h}e) - m^{PDG}_{B^0}\rvert < 60\,\text{MeV}$ && `E_ProbNNe` $< 0.8)$	all ee
$B^0 \to (D^- \to K^{*0}\ell\nu)\ell\nu$	$m(K\pi\ell^-) < 1780\,\text{MeV}$	low + central-q^2
$B^+ \to K^+\ell^+\ell^-$	$max(m(K\ell\ell), m(K_{\to\pi}\ell\ell)) < 5100\,\text{MeV}$	all

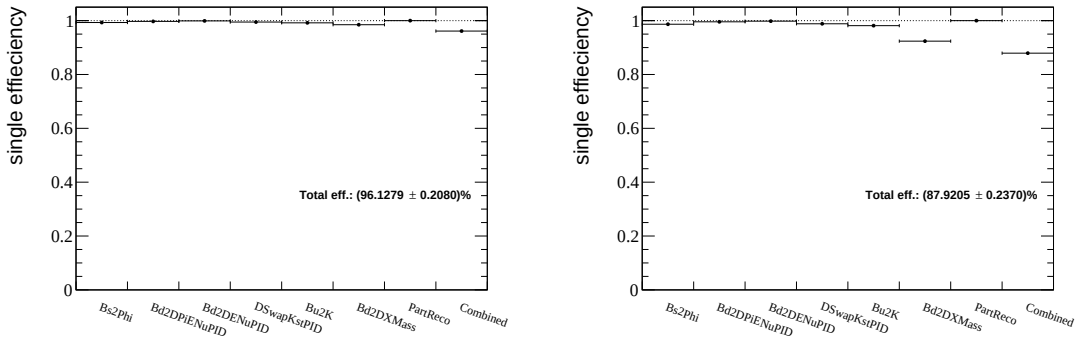


Figure 4.12: Signal efficiencies for $K^{*0}ee$ of all applied background vetoes for low (left) and central (right) q^2.

4.10 Exclusive Backgrounds for $B^+ \to K^+\ell^+\ell^-$

Background studies undertaken for $B^+ \to K^+\ell^+\ell^-$ modes follow the same strategy as for $B^0 \to K^{*0}\ell^+\ell^-$ described in Sec. 4.9.1. For these type of decays the main sources of background are from misidentification and partial reconstruction of decays.

4.10.1 Misidentification

For $B^+ \to K^+\ell^+\ell^-$ decays fewer types of misidentification can occur than for $B^0 \to K^{*0}\ell^+\ell^-$ decays since it is a 3-body final state containing only one hadron. Table 4.12 shows possible sources of misidentification, the relevant branching ratios and the q^2 regions in which they are particularly important.

Table 4.12: Relevant mis-identification of final state particles. Only the dominant decay modes are considered here, *i.e.* those with a large branching ratio and/or similarity to the signal decay. The branching ratios are taken from PDG [14]. Where the whole q^2 region is affected the first number corresponds to taking B_{SIG} from the signal mode, while the number in brackets takes B_{SIG} from the control mode.

MisID	**Decay mode**	$\sim \mathcal{B}_{BKG}/\mathcal{B}_{SIG}$	**Important for q^2 region**
$\pi \to K$	$B^+ \to \pi(J/\psi \to \ell^+\ell^-)$	0.04	J/ψ
	$B^+ \to \pi\ell^+\ell^-$	0.04	signal
$\pi \to \ell$	$B^+ \to (\overline{D}^0 \to K^+\pi^-)\ell^+\nu$	2060 (16)	all
	$B^+ \to (\overline{D}^0 \to K^+\ell^-\overline{\nu})\pi^+$	350 (3)	all
$\pi \to \ell$ double	$B^+ \to (\overline{D}^0 \to K^+\pi^-)\pi^+$	410 (3)	all
$K \leftrightarrow \ell$ swap	$B^+ \to K^+(J/\psi \to \ell^+\ell^-)$	N/A	all
	$B^+ \to K^+(\psi(2S) \to \ell^+\ell^-)$	N/A	all

$\pi \to K$ misID

If a pion is misidentified as a kaon, $B^+ \to \pi^+ J/\psi (\to \ell^+\ell^-)$ or $B^+ \to \pi^+\ell^+\ell^-$ decays can mimic the signal decay. Compared to $B^+ \to K^+\ell^+\ell^-$ decays, which are $b \to s\,\ell^+\ell^-$ transitions, $B^+ \to \pi^+\ell^+\ell^-$ decays are $b \to d$ decays which are suppressed and the PID requirements in the selection also act to reduce this as can be seen in Fig. 4.13. The remaining contributions from these backgrounds are modelled in the final mass fits, discussed in Sec. 7. However, they are only considered for J/ψ modes because with the current statistics in the non-resonant modes they are not expected to be resolvable.

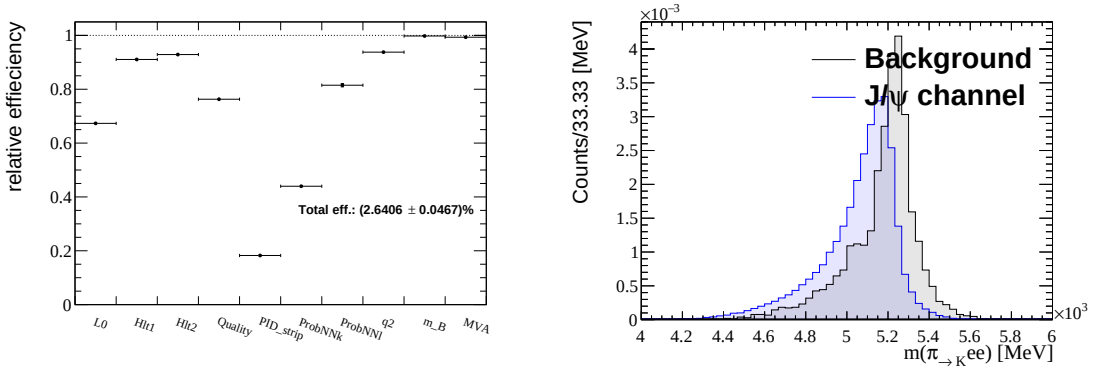


Figure 4.13: Left: Offline selection efficiency for $B^+ \to \pi(J/\psi \to \ell^+\ell^-)$ in J/ψ q^2 region. The cuts defined by the x-axis labels are defined in Tab. 4.7 and these cuts are applied separately. Right: Shape in $m(\pi_{\to K}\ell\ell)$, the black/blue distribution shows the background/signal MC sample (no dedicated veto applied).

$\pi \to \ell$ misID

Sources of $\pi \to \ell$ misID come from B decays proceeding via an intermediate D resonance, such as $B^+ \to (\overline{D}^0 \to K^+\pi^-)\ell^+\nu$ and $B^+ \to (\overline{D}^0 \to K^+\ell^-\nu)\pi^+$. These decays can mimic the signal decay if a pion is misidentified as a lepton of the same flavour as the other lepton in the final state. To suppress $B^+ \to (\overline{D}^0 \to K^+\pi^-)\ell^+\nu$ decays the $m(K^+\pi_{\to\ell^-})$ mass system can be reconstructed. A peak at the D mass in this spectrum is a clear signal for this misidentified background. To veto this background a PID requirement of L_ProbNNl > 0.8 is applied using a mass window of $m(D^0) \pm 40\,\text{MeV}/c^2$ within the $m(K^+\pi_{\to\ell^-})$ mass spectrum, as can be seen in Fig. 4.14. A veto on $m(K\ell^-)$ is used to suppress $B^+ \to (\overline{D}^0 \to K^+\ell^-\nu)\pi^+$ as can be seen in Fig. 4.16.

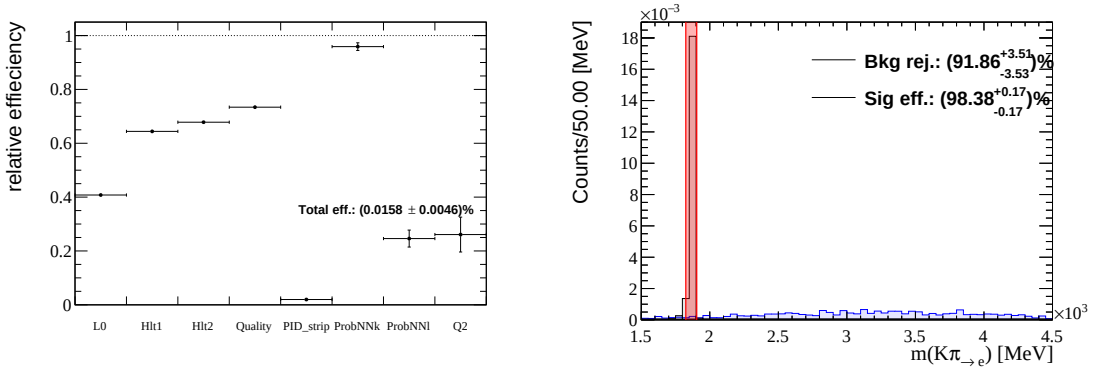


Figure 4.14: Left: Offline selection efficiency for $B^+ \to (\overline{D}^0 \to K^+\pi^-)\ell^+\nu$ in central-q^2 region. The cuts defined by the x-axis labels are defined in Tab. 4.7 and these cuts are applied separately. Right: Shape in $m(K^+\ell^-_{\to\pi})$, the black/blue distribution shows the background/signal MC sample.

$K \leftrightarrow \ell$ swap

This situation is analogous to that of $h \leftrightarrow \ell$ swaps in the R_{K^*} case. These types of backgrounds, largely from the resonant modes $B^+ \to K^+ J/\psi (\to \ell^+\ell^-)$, are problematic as the resulting decays in the resonant mode no longer peak at the B mass or the J/ψ mass in the control mode. The vetoes for these decays are directly analogous to the vetoes for $K \leftrightarrow \ell$ swaps in the R_{K^*} case and can be seen in Fig. 4.15.

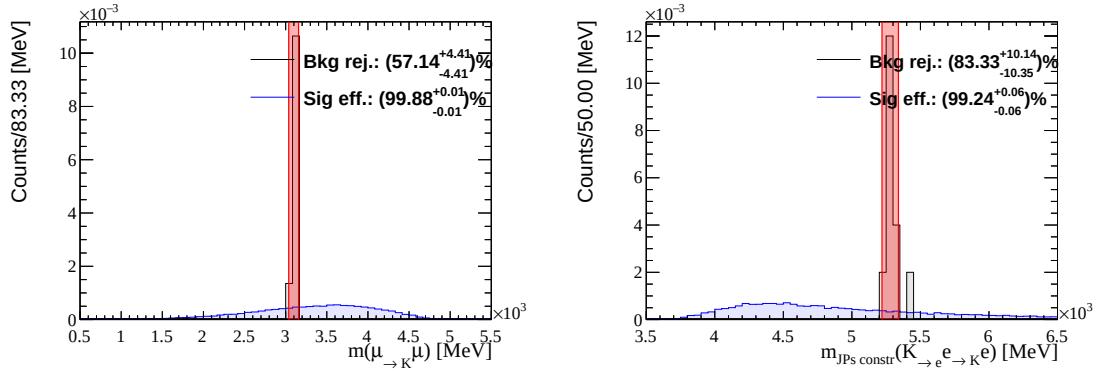


Figure 4.15: Veto against $K \leftrightarrow \ell$ swaps for $\mu\mu$ (left) and ee (right), the black/blue distribution shows the background/signal MC sample.

4.10.2 Partially Reconstructed

The relevant partially reconstructed background processes are summarised in Tab. 4.13.

Table 4.13: Possible partially-reconstructed decay modes which can contribute. Only the relevant decays are considered here, *i.e.* those having a large branching ratio and/or similarity to the signal decay. The branching ratios are taken from PDG [14]. Where the whole q^2 region is affected the first number corresponds to taking B_{SIG} from the signal mode, while the number in brackets takes B_{SIG} from the control mode.

Decay mode	$\sim \mathcal{B}_{BKG}/\mathcal{B}_{Signal}$	**Important for q^2 region**
$B^+ \to (\overline{D}^0 \to K^+\ell^-\overline{\nu})\ell^+\nu$	1777 (13)	all
$B^0 \to K^{*0}(J/\psi \to \ell^+\ell^-)$	0.8	J/ψ
$B^0 \to K^{*0}\ell^+\ell^-$	1.5	signal

$B^+ \to (\overline{D}^0 \to K^+\ell^-\overline{\nu})\ell^+\nu$

The $B^+ \to (\overline{D}^0 \to K^+\ell^-\overline{\nu})\ell^+\nu$ decay mimics the signal decay in LHCb as neutrinos are not detected. In a similar fashion to the R_{K^*} style veto for this sort of background, a selection of $m(K\ell^-)$ <1885 MeV/c^2 is applied as can be seen in Fig. 4.16. It can also be seen that this helps to suppress $B^+ \to (\overline{D}^0 \to K^+\ell^-\nu)\pi^+$ decays.

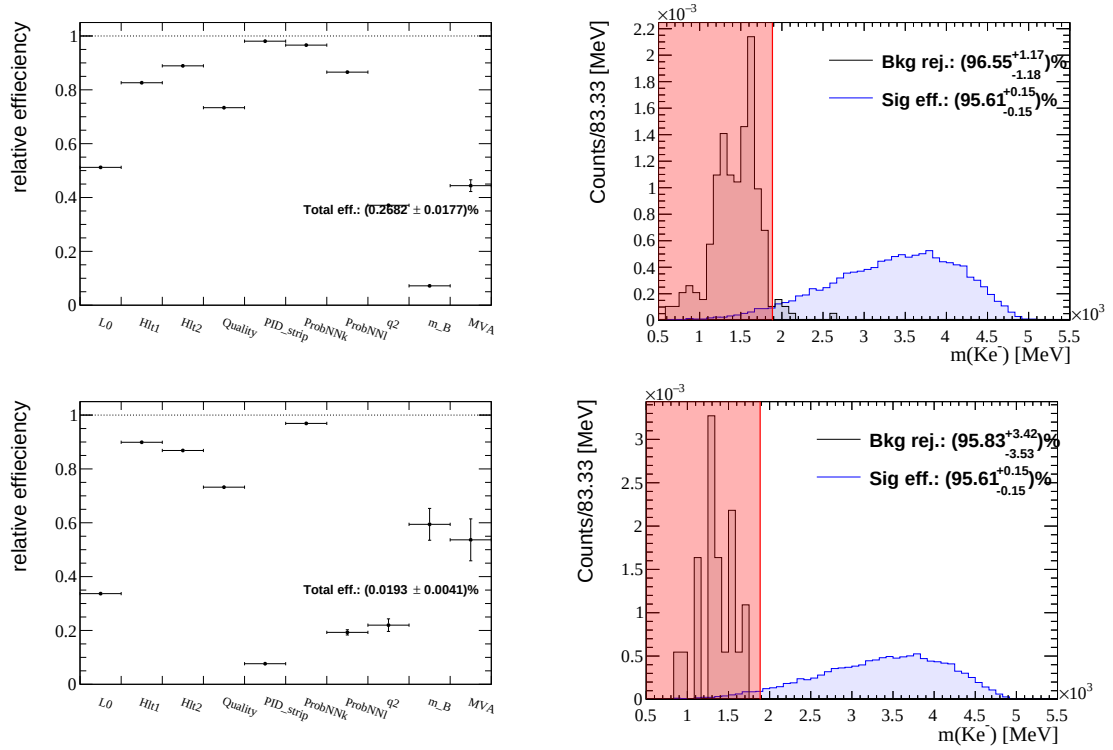


Figure 4.16: Top: Left: Offline selection efficiency for $B^+ \to (\overline{D}^0 \to K^+e^-\overline{\nu})e^+\nu$ in the central-q^2 region. The cuts defined by the x-axis labels are defined in Tab. 4.7 and these cuts are applied separately.Right: Shape in $m(Ke^-)$, the black/blue distribution shows the background/signal MC sample. Bottom: same for $B^+ \to (\overline{D}^0 \to K^+e^-\nu)\pi^+$

$B^0 \to K^{*0} J/\psi (\to \ell^+\ell^-)$

The decay $B^0 \to K^{*0} J/\psi (\to \ell^+\ell^-)$ can mimic the signal decay if a pion from the K^{*0} is not reconstructed. These decays can be vetoed in the J/ψ control mode in an analogous way to R_{K^*} as seen in Fig. 4.17. However, as this also sculpts the combinatorial shape, no veto is applied and the source is modelled in the control mode fits.

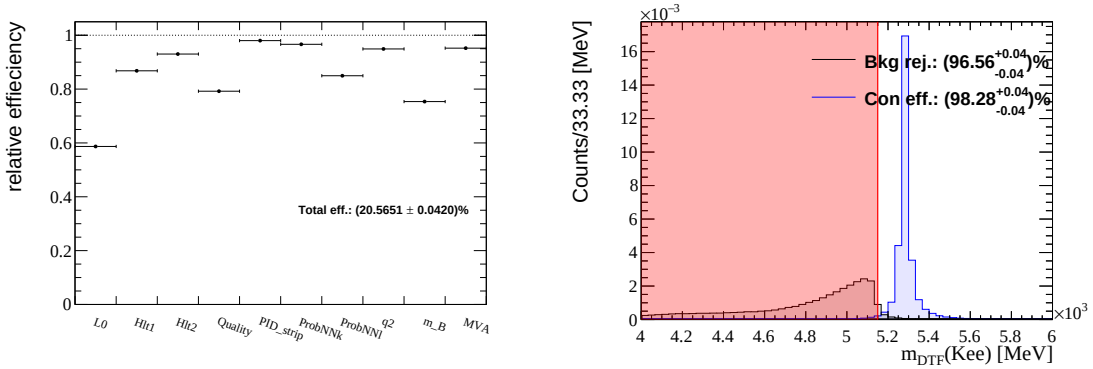


Figure 4.17: Left: Offline selection efficiency for $B^0 \to K^{*0}(J/\psi \to \ell^+\ell^-)$ in $J/\psi\, q^2$ region. The cuts defined by the x-axis labels are defined in Tab. 4.7 and these cuts are applied separately. Right: Shape in $m_{JPsconstr}(K\pi ee)$, the black/blue distribution shows the background/signal MC sample.

$B^0 \to K^{*0} e^+ e^-$ and $B^0 \to K\pi e^+ e^-$

The decays $B^0 \to K^{*0} e^+ e^-$ and $B^0 \to K\pi e^+ e^-$ mimic the signal decay if a pion, from the K^{*0} resonance for the former decay, is not reconstructed. A dedicated MVA is designed to suppress partially reconstructed $B^0 \to K^{*0} e^+ e^-$ decays and is discussed in Sec. 4.11. Since this MVA cannot suppress all of this background the remaining component is modelled in the final mass fits discussed in Sec. 7. For the resonant case it is also necessary to consider its isospin partner $B^+ \to K^{*+} e^+ e^-$ where the K^{*+} decays to a charged kaon and a neutral pion. Since none of these backgrounds can be fully vetoed they are modelled in the final fits.

4.10.3 Backgrounds unable to veto

For R_K there are fewer backgrounds which remain after the vetoes previously discussed that must be considered in the final fits than the R_{K^*} case, Tab 4.14 lists these backgrounds and the q^2 regions in which they are important.

Table 4.14: Backgrounds that cannot be vetoed that must be modelled in the mass fits, together with the q^2 regions they are important for

Decay mode	**Important for q^2 region**
$B^+ \to \pi^+ J/\psi (\to \ell^+\ell^-)$	J/ψ
$B^0 \to K^{*0} J/\psi (\to \ell^+\ell^-)$	J/ψ
$B^0 \to K^{*0}\psi(2S)(\to \ell^+\ell^-)$	$\psi(2S)$
$B^0 \to K^{*0}\ell^+\ell^-$	signal

4.10.4 Summary of exclusive background selection for R_K

A summary of all the vetoes applied for R_K and the q^2 region in which they are applied is shown in Tab. 4.15. The combined efficiency of these vetoes on the low and central-q^2 regions for $B^+ \to K^+ e^+ e^-$ decays is shown in Fig. 4.18.

Table 4.15: Summary of the exclusive background selection requirements.

Background	**Requirement**	**Applied for**
$B^+ \to (\overline{D}^0 \to K^+\pi^-)\ell^+\nu$	$!(\lvert m(K^+\ell^-_{\to\pi}) - m^{PDG}(D^0)\rvert < 40\,\text{MeV}$ && `L_ProbNNl` $< 0.8)$	all
	$!(\lvert m(\mu_{\to K}\mu) - m^{PDG}_{J/\psi,(\psi(2S))}\rvert < 60\,\text{MeV}$ && (`M_ProbNNmu` $< 0.8)$	all $\mu\mu$
$K \leftrightarrow \ell$ swap	$!(\lvert m_{J/\psi,(\psi(2S))constr.}(K_{\to e} e_{\to K} e) - m^{PDG}_{B^+}\rvert < 60\,\text{MeV}$ && `E_ProbNNe` $< 0.8)$	all ee
$B^+ \to (\overline{D}^0 \to K^+\ell^-\overline{\nu})\ell^+\nu$	$m(K\ell^-) < 1885\,\text{MeV}$	low + central-q^2

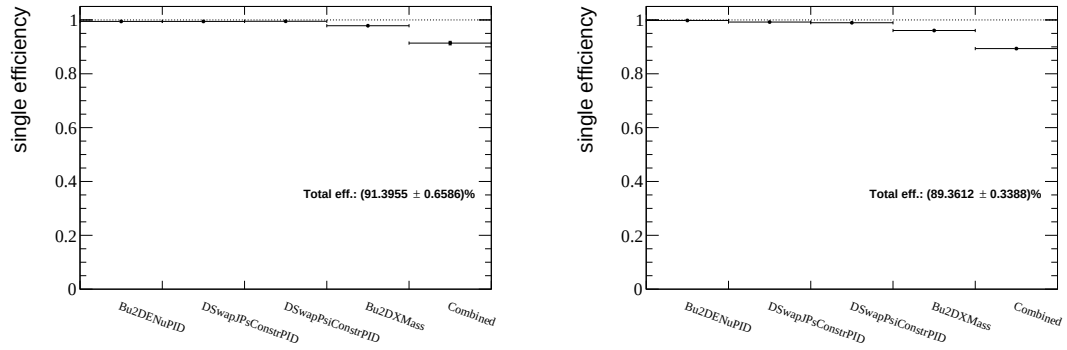


Figure 4.18: Signal efficiencies for Kee of all applied background vetos for low (left) and central (right) q^2.

4.11 Multivariate classifiers

CatBoost classifiers developed by Yandex [128] are used in conjunction with the Reproducible Experiment Platform [129] to reduce combinatorial background (MVA_{Comb}) for $\ell^+\ell^-$ modes and to reduce partially reconstructed backgrounds ($\text{MVA}_{\text{PReco}}$) for ee modes only. Separate classifiers are always trained for B^0 and B^+ decays, for $\mu\mu$ and ee modes, and in the three separate run periods. This leads to 12 separate trainings for classifiers designed to reduce combinatorial background and six separate trainings for classifiers designed to reduce partially reconstructed backgrounds.

Samples used to train the classifiers have the full selection applied and consist of proxies for the signal mode in question and the type of background desired to be suppressed. The signal proxy is a simulated sample of the signal mode of interest. The background proxy for combinatorial background comes from the upper sideband in collision data. The background proxy for partially reconstructed backgrounds is taken from simulated samples of the decay modes known to contribute.

In order to have a reasonable sample size for the combinatorial background to train with, especially in Run 1, trigger categories are combined as well as low- and central-q^2 regions for MVA_{Comb}. For consistency the same procedure is followed for $\text{MVA}_{\text{PReco}}$. Training is split by run periods with a separate training performed for Run 1, Run 2p1 and Run 2p2 datasets.

When creating samples for training, the number of signal and background events are chosen so that equal amounts are used. Therefore, the limit on the statistics used in the training is restricted by which of the signal or background datasets has fewer entries. This is only a real issue for the combinatorial training where there are relatively few events in the upper sideband in collision data after the full selection compared to the large statistics of the signal simulation produced. Where the full statistics of the sample are not used due to this approach, the events are chosen at random to avoid taking events from only one year of the run period or only one magnet polarity.

To avoid biases, a *k-fold* approach [130], where 10 classifiers are iteratively trained, is implemented. A classifier uses 90% of the events from the training sample to train with and is applied to the remaining 10% that was not used in the training to evaluate its performance. The ten folds are chosen using the modulo of the `eventNumber`, a unique number given to each event, with 10. For the first fold, all

events for which this operation yields 0 are chosen to be in the 10% of data to be used to evaluate performance, whilst events for which the operation yields numbers from 1-9 are used for training. This process is then repeated another nine times, each time leaving out a different tenth of the data from the training sample.

Once trainings are performed for each of the run periods for muon and electron modes, the MVA responses that are produced are then applied to all samples. Classifiers to suppress combinatoric backgrounds, MVA_{Comb}, are applied to all modes whilst those suppressing partially reconstructed backgrounds, $\text{MVA}_{\text{PReco}}$, are only applied to ee modes. Classifiers trained with Run 1 data are applied to 2011 and 2012 samples, those trained with Run 2p1 data are applied to 2015 and 2016 samples, and those trained with Run 2p2 data are applied to 2017 and 2018 samples. For events that were used in training the ten MVAs, the MVA which did not have that event in its training sample is applied.

4.11.1 MVA_{Comb}

Signal and background samples

Samples representative of the signal and the background one wishes to reduce are required as inputs to train the classifier. The proxy for signal is fully reconstructed, selected and truth matched (Sec. 4.3) $B^0 \to K^{*0}\ell^+\ell^-$ (for R_{K^*}) or $B^+ \to K^+\ell^+\ell^-$ (for R_K) simulated events.

As a proxy for combinatorial background, events in the upper-mass sideband of collision data are taken. The lower sideband is not suitable for training as it contains a significant fraction of partially-reconstructed background, as well as signal for electron modes. The upper sidebands are defined as $m(B) > 5400\,\text{MeV}/c^2$ and $m(B) > 5600\,\text{MeV}/c^2$ for muon and electron modes, respectively, with the upper limit being set by the stripping requirements (Tab. 4.1). The full selection is applied to the collision data in the mass ranges defined. An exception to this is for trainings with a signal proxy of $B^0 \to K^{*0}e^+e^-$ decays, where, to increase the available statistics in the upper sideband, the requirement on the K^{*0} invariant mass is relaxed to a mass window of $\pm 200\,\text{MeV}/c^2$ around the true K^{*0} mass instead of $\pm 100\,\text{MeV}/c^2$. A summary of the available signal and upper sideband statistics can be seen in Tab. 4.16, showing that in every case the upper sideband statistics are the limiting factor following the strategy of training with an equal number of signal and background events.

Table 4.16: Statistics of Signal and Background samples used to train MVA_{COMB}. The number in brackets denotes the full size of the sample compared to what was used. The upper sideband in data is defined as $> 5400\,\text{MeV}$ for muon modes and as $> 5600\,\text{MeV}$ for electron modes.

Training	**Signal events (MC)**	**Background events (Data)**
$B^+ \to K^+\mu^+\mu^-$ RUN 1	23217 (102991)	23217 (limiting)
$B^+ \to K^+\mu^+\mu^-$ RUN 2P1	40417 (65970)	40417 (limiting)
$B^+ \to K^+\mu^+\mu^-$ RUN 2P2	68222 (122538)	68222 (limiting)
$B^+ \to K^+e^+e^-$ RUN 1	4715 (18498)	4715 (limiting)
$B^+ \to K^+e^+e^-$ RUN 2P1	10489 (26291)	10489 (limiting)
$B^+ \to K^+e^+e^-$ RUN 2P2	16286 (63019)	16286 (limiting)
$B^0 \to K^{*0}\mu^+\mu^-$ RUN 1	3884 (43132)	3884 (limiting)
$B^0 \to K^{*0}\mu^+\mu^-$ RUN 2P1	5331 (66247)	5331 (limiting)
$B^0 \to K^{*0}\mu^+\mu^-$ RUN 2P2	9382 (118276)	9382 (limiting)
$B^0 \to K^{*0}e^+e^-$ RUN 1	1051 (42800)	1051 (limiting)
$B^0 \to K^{*0}e^+e^-$ RUN 2P1	2095 (62512)	2095 (limiting)
$B^0 \to K^{*0}e^+e^-$ RUN 2P2	3869 (66997)	3869 (limiting)

Input Variables

The classifier for reducing combinatorial background has as its input 23 variables for trainings involving B^0 modes and 16 variables for trainings involving B^+ modes. For both B^+ and B^0 trainings the same variables are used for ee and $\mu\mu$ modes and the variables remain unchanged for trainings performed in different run periods.

The input variables are selected from a larger set of variables which all showed differences in the distribution of that variable in the respective signal and background samples. To reduce this set of variables a backward elimination is performed. This involves using all variables in training an MVA from one run period for both electron and muon modes for one fold. Using feature importances provided by CATBOOST the least important variables are removed and trainings are repeated until a significant change is observed in the area under the ROC (Receiver Operating Characteristic) curve (around 1%) for either the electron or muon mode. The variables which result from this elimination procedure describe the kinematics of the decay and quality of the tracks and vertices. The variables are listed in Tab. 4.17 and Tab. 4.18 for B^0 and B^+ trainings, respectively. The comparison of these variables between signal and background samples with the full preselection applied for muon and electron modes can be found in Appendix D.1.

Table 4.17: Summary of the input variables to MVA_{COMB} for B^0 trainings. The variable p_{T} corresponds to the transverse momentum, $\chi^2_{IP\_OWNPV}$ to the quality of the impact parameter fit with respect to the particles' own primary vertex, $\chi^2_{FD\_OWNPV}$ to the quality of the flight distance fit with respect to the particles' own primary vertex, χ^2_{DTF}/ndf to the quality of the vertex fit divided by the number of degrees of freedom and DIRA to the DIRA angle.

Particle	**Variables**
B^0	$p_{\text{T}},\ \chi^2_{IP\_OWNPV},\ \chi^2_{FD\_OWNPV},\ \chi^2_{\text{vtx}}/\text{ndf},\ \chi^2_{DTF}/\text{ndf},\ \text{DIRA}$
K^{*0}	$p_{\text{T}},\ \chi^2_{IP\_OWNPV},\ \chi^2_{FD\_OWNPV},\ \chi^2_{\text{vtx}}/\text{ndf},\ \text{DIRA}$
$\ell^+\ell^-$	$p_{\text{T}},\ \chi^2_{IP\_OWNPV},\ \chi^2_{FD\_OWNPV},\ \chi^2_{\text{vtx}}/\text{ndf},\ \text{DIRA}$
h	$min, max(p_{T,K},\ p_{T,\pi}),\ min, max\ (\chi^2_{IP\_OWNPV,K},\ \chi^2_{IP\_OWNPV,\pi})$
ℓ	$min, max(p_{T,\ell^+},\ p_{T,\ell^-}),\ min(\chi^2_{IP\_OWNPV,\ell^+},\ \chi^2_{IP\_OWNPV,\ell^-})$

Table 4.18: Summary of the input variables to MVA_{COMB} for B^+ trainings. The variable p_{T} corresponds to the transverse momentum, $\chi^2_{IP\_OWNPV}$ to the quality of the impact parameter fit with respect to the particles' own primary vertex, $\chi^2_{FD\_OWNPV}$ to the quality of the flight distance fit with respect to the particles' own primary vertex, χ^2_{DTF}/ndf to the quality of the vertex fit divided by the number of degrees of freedom and DIRA to the DIRA angle.

Particle	**Variables**
B^+	$p_{\text{T}},\ \chi^2_{IP\_OWNPV},\ \chi^2_{FD\_OWNPV},\ \chi^2_{\text{vtx}}/\text{ndf},\ \chi^2_{DTF}/\text{ndf},\ \text{DIRA}$
K	$p_{\text{T}},\ \chi^2_{IP\_OWNPV}$
$\ell^+\ell^-$	$p_{\text{T}},\ \chi^2_{IP\_OWNPV},\ \chi^2_{FD\_OWNPV},\ \chi^2_{\text{vtx}}/\text{ndf},\ \text{DIRA}$
ℓ	$min(p_{T,\ell^+},\ p_{T,\ell^-}),\ min, max(\chi^2_{IP\_OWNPV,\ell^+},\ \chi^2_{IP\_OWNPV,\ell^-})$

Training and Checks

After the training has been completed a number of checks can be made to determine the effectiveness of the trained classifier. For $\mu\mu$ and ee modes in RUN 1 Figs. 4.19 and 4.21, respectively, show ROC curves resulting from each of the ten folds used for training. These are representative of all run periods used for training. For cases with the highest statistics the ROC curves from each fold agree very well with each other and all cases show consistent behaviour between folds. In all cases the average area under the ROC curves is above 0.98 indicating very good performance (high signal classification and high background rejection).

The CATBOOST classifier can return the importance of the variables passed to the classifier for training, identifying the most discriminating variables. For B^0 trainings,

the most discriminating variable is the χ^2 of the kinematic fit to the B^0 performed using `DecayTreeFitter` followed by the p_{T} of the B^0 for both $\mu\mu$ and ee modes. For B^+ trainings, the most discriminating variable is the χ^2 of the kinematic fit to the B^+ performed using `DecayTreeFitter` followed by the χ^2_{IP} of the K for the $\mu\mu$ mode and the p_{T} of the B^+ for the ee mode.

Overtraining checks are performed to test how well the classifier responds to unseen data, examples from one fold for $\mu\mu$ and ee modes in Run 1 can be seen in Figs. 4.20 and 4.22, respectively. In general, these are representative of the other folds in Run 1 and folds from other periods. The best agreement is between the training and testing samples for the higher statistic cases, however the other samples do not show signs of overtraining. The reduced performance of the B^0 ee mode here is due to very low statistics in Run 1, with higher statistics in Run 2p1 or Run 2p2 the agreement is much better.

Further plots detailing the relative importance of the input variables from one representative fold and the correlation of the MVA output with the B meson mass and q^2 in Run 2p1 can be found in Appendix D.2. No significant correlation can be seen in either the B mass or q^2 and this is seen in the other run periods as well.

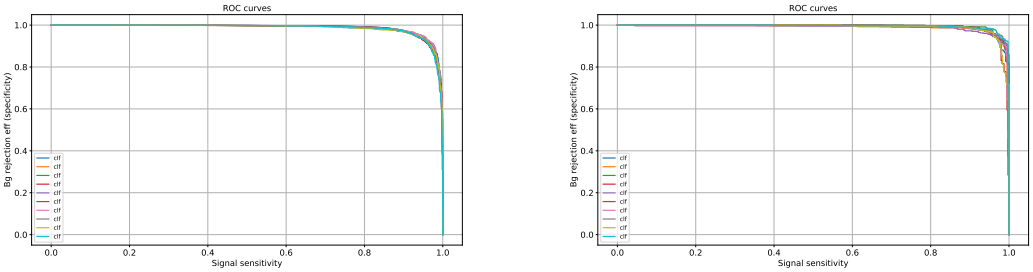


Figure 4.19: ROC curves for MVA_{COMB} for $\mu\mu$ modes in RUN 1 for (left) B^+ and (right) B^0 trainings for all 10 folds, represented by the 10 different coloured lines in each plot.

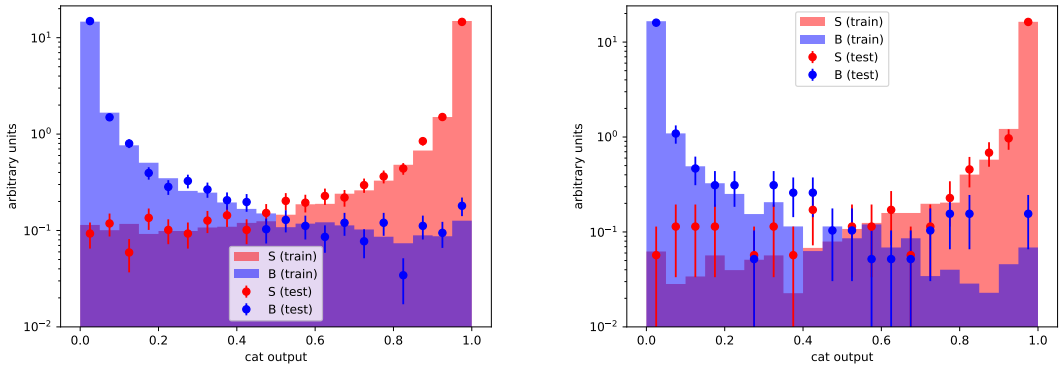


Figure 4.20: Overtraining plots for MVA_{COMB} for one fold in RUN 1 for (left) B^+ and (right) B^0 trainings for $\mu\mu$ modes. The red and blue histograms represent signal and background data used for training respectively, corresponding to $9/10^{th}$s of the total training data. The red and blue points represent signal and background data used for testing respectively, corresponding to the tenth of the training sample left out of this particular training.

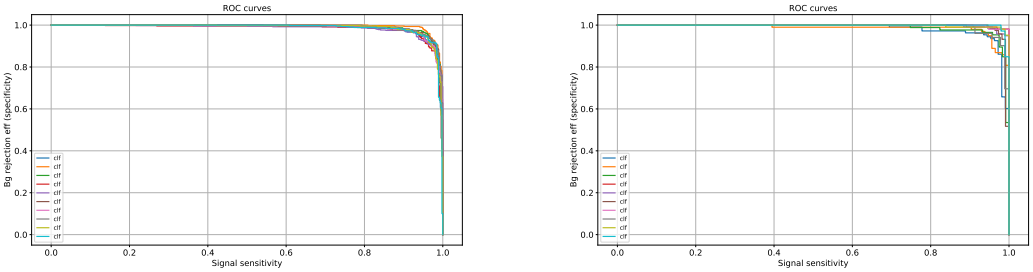


Figure 4.21: ROC curves for MVA_{COMB} for ee modes in RUN 1 for (left) B^+ and (right) B^0 trainings for all 10 folds, represented by the 10 different coloured lines in each plot.

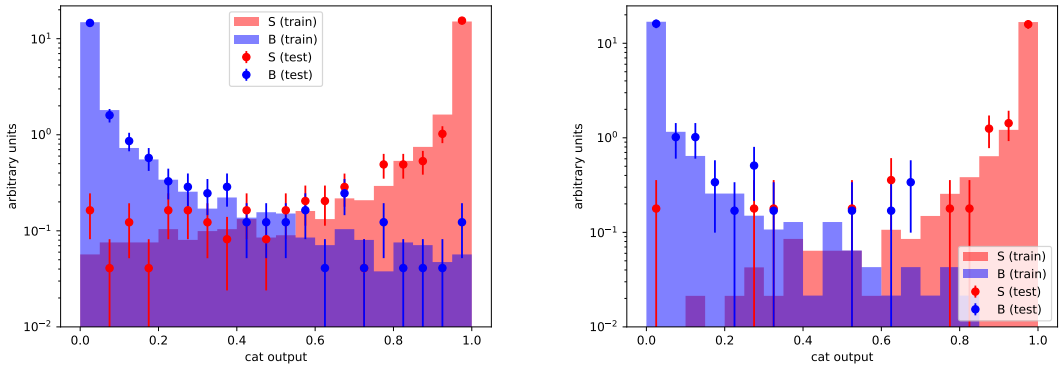


Figure 4.22: Overtraining plots for MVA_{COMB} for one fold in RUN 1 for (left) B^+ and (right) B^0 trainings for ee modes. The red and blue histograms represent signal and background data used for training respectively, corresponding to $9/10^{th}$s of the total training data. The red and blue points represent signal and background data used for testing respectively, corresponding to the tenth of the training sample left out of this particular training.

4.11.2 $\text{MVA}_{\text{PRECO}}$

Signal and background samples

Samples representative of the signal and the background one wishes to reduce are required as inputs to train the classifier. The proxy for signal is fully reconstructed, selected and truth matched (Sec.4.3) $B^0 \to K^{*0}\ell^+\ell^-$ (for R_{K^*}) or $B^+ \to K^+\ell^+\ell^-$ (for R_K) simulated events.

As a proxy for partially-reconstructed background for ee modes in the low- and central-q^2 regions, simulated samples of the relevant decays are used. These samples are fully selected with the additional requirement on `TupleToolMCBackgroundInfo` to be (20 || 30 || 40 || 50) to ensure partially reconstructed events are selected for the training. For trainings with a signal proxy of $B^+ \to K^+e^+e^-$ decays, a sample of $B^0 \to K^{*0}e^+e^-$ MC is used as a proxy for background. If a π from the K^{*0} decay is missed when reconstructing this decay it will mimic the signal. For trainings with a signal proxy of $B^0 \to K^{*0}e^+e^-$ decays, a sample of $B^+ \to K\pi\pi e^+e^-$ MC is used as a proxy for background. If a π is missed when reconstructing this decay it will mimic the signal. A summary of the available signal and background statistics can be seen in Tab. 4.19.

Table 4.19: Statistics of Signal and Background samples used to train $\text{MVA}_{\text{PRECO}}$. The number in brackets denotes the full size of the sample compared to what was used.

Training	**Signal events (MC)**	**Background events (MC)**
$B^+ \to K^+e^+e^-$ RUN 1	18498 (limiting)	18498 (35568)
$B^+ \to K^+e^+e^-$ RUN 2P1	26291 (limiting)	26291 (96348)
$B^+ \to K^+e^+e^-$ RUN 2P2	61975 (63019)	61975 (limiting)
$B^0 \to K^{*0}e^+e^-$ RUN 1	26824 (42800)	26824 (limiting)
$B^0 \to K^{*0}e^+e^-$ RUN 2P1	62512 (limiting)	62512 (73913)
$B^0 \to K^{*0}e^+e^-$ RUN 2P2	27027 (66997)	27027 (limiting)

Input Variables

The classifier for reducing partially reconstructed background has as its input two different sets of 14 variables for B^0 and B^+ trainings. The variables used for B^0 or B^+ trainings do not change for trainings performed in different run periods.

Variables are selected based on the procedure outlined in Sec. 4.11.1 and describe the kinematics of the decay, quality of the tracks and vertices, and isolation. Cone isolation variables are computed in a 0.5 mrad cone around the particle of interest.

MULT type variables denote the multiplicity within the cone excluding the particle of interest. SPT type variables denote the scalar sum of p_{T} of tracks within the cone not related to the particle of interest. IT type variables denote how much of the p_{T} of the cone is of the particle of interest. The variables are listed in Tab. 4.20 and Tab. 4.21 for B^0 and B^+ trainings, respectively. There is good agreement between data and MC for these variables and plots can be found in Appendix E. The comparison of these variables from the relevant signal and background samples with the full preselection applied can be found in Appendix D.1.

Table 4.20: Summary of the input variables to $\mathrm{MVA}_{\mathrm{PReco}}$ for B^0 trainings. The variable *pt* corresponds to the transverse momentum, $\chi^2_{IP\_OWNPV}$ to the quality of the impact parameter fit with respect to the particles' own primary vertex, $\chi^2_{FD\_OWNPV}$ to the quality of the flight distance fit with respect to the particles' own primary vertex, χ^2_{DTF}/ndf to the quality of the vertex fit divided by the number of degrees of freedom and DIRA to the DIRA angle.

Particle	**Variables**
B^0	$\chi^2_{IP\_OWNPV}$, χ^2_{DTF}/ndf, χ^2_{vtx}/ndf, DIRA, vertex isolation one-track-χ^2, vertex isolation one-track-mass
K^{*0}	DIRA, χ^2_{vtx}/ndf, $\chi^2_{IP\_OWNPV}$
$\ell^+\ell^-$	DIRA, $\chi^2_{IP\_OWNPV}$, $\chi^2_{FD\_OWNPV}$
h	$min(p_{T,K},\ p_{T,\pi})$
l	charged-cone isolation min(MULT l^+, MULT l^-)

Table 4.21: Summary of the input variables to $\mathrm{MVA}_{\mathrm{PReco}}$ for B^+ trainings. The variable *pt* corresponds to the transverse momentum, $\chi^2_{IP\_OWNPV}$ to the quality of the impact parameter fit with respect to the particles' own primary vertex, $\chi^2_{FD\_OWNPV}$ to the quality of the flight distance fit with respect to the particles' own primary vertex, χ^2_{DTF}/ndf to the quality of the vertex fit divided by the number of degrees of freedom and DIRA to the DIRA angle.

Particle	**Variables**
B^+	p_{T}, $\chi^2_{IP\_OWNPV}$, DIRA, vertex isolation one-track-χ^2, vertex isolation one-track-mass
$\ell^+\ell^-$	DIRA, $\chi^2_{IP\_OWNPV}$, $\chi^2_{FD\_OWNPV}$
K	p_{T}
l	$min(\chi^2_{IP\_OWNPV,\ell^+},\ \chi^2_{IP\_OWNPV,\ell^-})$, charged-cone isolation min, max(APT l^+, APT l^-), charged-cone isolation max(MULT l^+, MULT l^-), charged-cone isolation min(IT l^+, IT l^-)

Training and Checks

ROC curves for RUN 1 for B^0 and B^+ trainings can be seen in Fig. 4.23. The performance is reduced compared to the combinatorial case as this background is very similar to the signal in a number of distributions. These are representative of all the run periods.

For both B^0 and B^+ the most discriminating variable is the B vertex isolation one-track-mass followed by the B vertex isolation one-track-χ^2.

Overtraining plots for B^0 and B^+ trainings from one fold from RUN 1 can be seen in Fig. 4.24. There is good agreement and overtraining is not present. These are representative of all folds and run periods.

The relative importance of in the input variables from one representative fold and the correlation of the MVA output with the B meson mass and q^2 can be found in Appendix D.2.

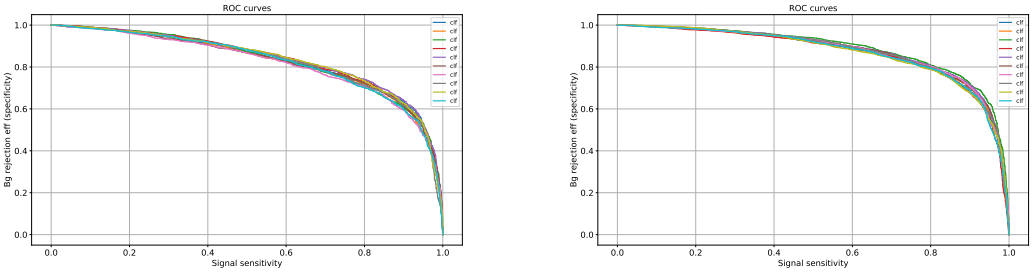


Figure 4.23: ROC curves for $\mathrm{MVA}_{\mathrm{PReco}}$ for ee modes in Run 1 for (left) B^+ and (right) B^0 trainings for all 10 folds, represented by the 10 different coloured lines in each plot.

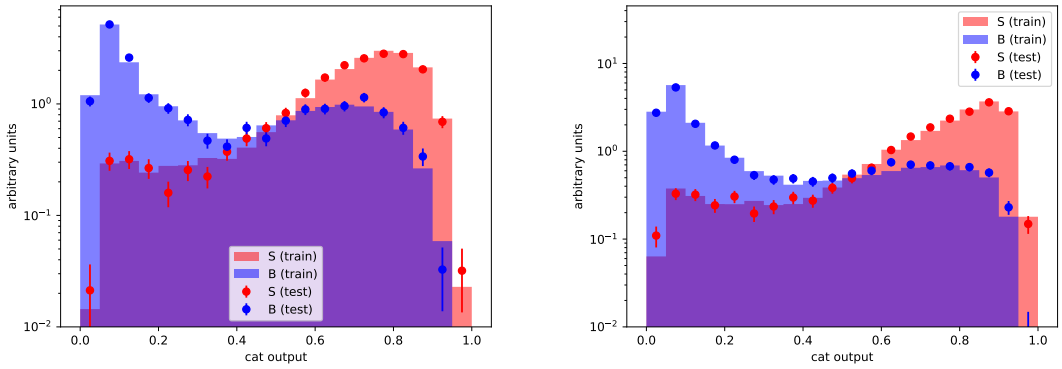


Figure 4.24: Overtraining plots for $\mathrm{MVA}_{\mathrm{PReco}}$ for one fold in Run 1 for (left) B^+ and (right) B^0 trainings for ee modes. The red and blue histograms represent signal and background data used for training respectively, corresponding to $9/10^{th}$s of the total training data. The red and blue points represent signal and background data used for testing respectively, corresponding to the tenth of the training sample left out of this particular training.

4.12 MVA working point optimisation

4.12.1 $\mathrm{MVA}_{\mathrm{Comb}}$ and $\mathrm{MVA}_{\mathrm{PReco}}$

To determine the optimum value of the MVA responses to cut upon to retain the maximum amount of signal whilst removing the most background, the significance is chosen as the figure-of-merit. The significance is defined as $N_{\mathrm{S}}/\sqrt{N_{\mathrm{S}} + N_{\mathrm{B}}}$ where N_{S} and N_{B} are the expected number of signal and background events, respectively.

Fully selected MC can be used to extract N_{S}, whilst fits to the reconstructed B meson mass in data, where possible to just the upper and lower sidebands but also full fits where this is necessary, are used to extract N_{B} in a defined signal region. This region is chosen to contain a large fraction of the genuine signal to determine the amount of background lying directly underneath it.

The optimisation is performed, with trigger categories combined, separately for B^0 and B^+, for ee and $\mu\mu$ modes, for the three run periods and in four regions of q^2 corresponding to low and central (rare mode) regions and J/ψ and $\psi(2S)$ (control modes) regions. This results in 48 separate optimisations. Depending on the $\ell^+\ell^-$ mode and the q^2 region, the following can differ: the MC sample to be used to determine N_{S}; whether or not the data fit is to the reconstructed B meson mass or to the B mass determined from `DecayTreeFitter`; or whether the optimisation will be performed in 1 dimension ($\mathrm{MVA}_{\mathrm{Comb}}$) or 2 dimensions ($\mathrm{MVA}_{\mathrm{Comb}}$ and $\mathrm{MVA}_{\mathrm{PReco}}$). In general all $\mu\mu$ and control mode optimisations are done in 1D.

The optimisation procedure consists of iterating through cut values on the MVA response, which can take values from 0 to 1, and determining N_{S} and N_{B} for each iteration and from that the significance. When all iterations have been completed a maximum is then located in the significance and the cut value at that maximum is concluded to be the optimum cut value. As N_{S} and N_{B} are essential ingredients to this procedure how they are determined will be elaborated below.

The number of signal events, N_{S}, is determined using inputs from both simulation and data. Firstly the number of events, N_{MC}, for the fully selected and truth matched MC sample of interest in a defined signal window are counted. For $\mu\mu$ modes the signal window is defined as $m(B) \pm 50\,\mathrm{MeV}/c^2$. The signal window for ee modes has a wider range of 5150–5350 MeV/c^2 in order to account for bremsstrahlung. Before any iterations of the optimisation can proceed however, N_{MC} must be scaled by a factor that accounts for the actual amount of signal, N_{expected}, we would expect

in this signal window. To determine this factor yields of $B \to K^{(*)} J/\psi (\to \ell\ell)$ modes, $N_{B\to K^{(*)} J/\psi (\to \ell\ell)}$, are taken from ${}_s\mathcal{P}lot$ted data (Appendix B) and summed over both trigger categories and the years making up the given run period. The $B \to K^{(*)} J/\psi (\to \ell\ell)$ yield is then scaled by the ratio of the branching fraction of the signal mode of interest (defined by the q^2 region), $\mathcal{B}_S$, and the particular J/ψ resonant mode, $\mathcal{B}_{B\to K^{(*)} J/\psi (\to \ell\ell)}$. A further scaling is applied by the ratio of the efficiencies of the fully selected signal mode of interest, ϵ_S, and the fully selected $B \to K^{(*)} J/\psi (\to \ell\ell)$ mode, $\epsilon_{B\to K^{(*)} J/\psi (\to \ell\ell)}$. If a $B \to K^{(*)} J/\psi (\to \ell\ell)$ optimisation is performed both of these ratios will be unity. So for R_{K^*} in a particular run period

$$N_{\text{expected}} = N_{B^0\to K^{*0} J/\psi (\to \ell^+\ell^-)} \times \frac{\mathcal{B}_S}{\mathcal{B}_{B^0\to K^{*0} J/\psi (\to \ell^+\ell^-)}} \times \frac{\epsilon_S}{\epsilon_{B^0\to K^{*0} J/\psi (\to \ell^+\ell^-)}} \,.$$

This results in a scale factor $\alpha_S = N_{\text{expected}}/N_{\text{MC}}$ which is applied to the number of simulated events given the MVA cut of that iteration of the optimisation, $N_{\text{MC|MVA Cut}}$, to determine N_{S} such that

$$N_{\text{S}} = \alpha_S \times N_{\text{MC}|MVA\ Cut} \,.$$

The number of background events, N_{B}, is determined from fits to the B meson mass in data. Where possible fits are performed to the upper and lower sidebands and then extrapolated into the signal regions defined previously to determine N_{B}. This is the general procedure for all $\mu\mu$ modes and control modes where an estimate of the combinatorial background in the signal region is required and using an exponential to model this is sufficient. However, for ee modes in low- and central-q^2 regions an estimate for the combined combinatorial and partially-reconstructed background in the signal region is required. This necessitates using a component to model the partially-reconstructed background. To obtain stable fits in these cases requires performing a full fit to the data with three components: a double sided crystal ball for the signal; an exponential for the combinatorial background; a `RooKeysPdf` of the partially reconstructed background of interest. In the central-q^2 region it is also necessary to add a further component, a `RooKeysPdf` to model the $B \to K^{(*)} J/\psi (\to \ell\ell)$ leakage with its yield constrained. In all cases the background samples used to produce `RooKeysPdf`'s are fully selected with the MVA cut of the current iteration applied.

To summarise all the cases:

- Low- and central-q^2 $\mu\mu$ modes: An exponential is used to model combinatorial background and a fit is performed to the lower and upper sidebands of the reconstructed B meson mass. A 1D optimisation in $\mathrm{MVA}_{\mathrm{ComB}}$ is performed. An example fit and optimisation scan can be seen in Fig. 4.25.

- $J/\psi\,(\mu\mu)$ modes: An exponential is used to model combinatorial background and additionally a `RooKeysPdf` of $B^0 \to XJ/\psi(\to \mu^+\mu^-)$ (R_{K^*}) or $B^+ \to XJ/\psi(\to \mu^+\mu^-)$ (R_K) is used for better fit stability in the lower sideband. Fits are performed to the lower and upper sidebands of the `DecayTreeFitter` B meson mass. A 1D optimisation in $\mathrm{MVA}_{\mathrm{ComB}}$ is performed. An example fit and optimisation scan be seen in Fig. 4.26.

- $\psi(2S)(\mu\mu)$ modes: An exponential is used to model combinatorial background and a fit is performed to the lower and upper sidebands of the `DecayTreeFitter` B meson mass. A 1D optimisation in $\mathrm{MVA}_{\mathrm{ComB}}$ is performed.

- low-q^2 ee modes: A full fit to the reconstructed B meson mass is performed consisting of three components. The signal is modelled by a double sided crystal ball, a fit to simulated signal in the low-q^2 region is used to determine the final parameters used for the fit to data. The combinatorial background is modelled by an exponential and partially-reconstructed background is modelled by a `RooKeysPdf` of partially-reconstructed simulated $B^+ \to K\pi\pi e^+e^-$ for R_{K^*} or $B^0 \to K^{*0}e^+e^-$ for R_K. A 2D optimisation in $\mathrm{MVA}_{\mathrm{ComB}}$ and $\mathrm{MVA}_{\mathrm{PReco}}$ is performed. As a cross-check two 1D optimisations are also performed with the same fit configurations, first a 1D scan in $\mathrm{MVA}_{\mathrm{ComB}}$ is performed and the optimum cut is extracted. Secondly fixing the optimum $\mathrm{MVA}_{\mathrm{ComB}}$ cut a scan in $\mathrm{MVA}_{\mathrm{PReco}}$ is performed. The results of these "1D+1D" scans can then be compared to the full 2D scans. An additional cross-check performed is to undertake the 2D scan with all run periods combined.

- central-q^2 ee modes: A full fit to the reconstructed B meson mass is performed consisting of four components. The first three components are identical to the low-q^2 case above, the fourth component is a `RooKeysPdf` of J/ψ leakage into the central-q^2 region obtained from applying the selection of the central-q^2 region to the simulated J/ψ data. This leakage component has its yield constrained by determining a scale factor between the ${}_s\mathcal{P}lot$ted J/ψ yield and the number of simulated J/ψ events and then applying this factor to the number of simulated J/ψ events determined to pass the selection of the

central-q^2 region. A 2D optimisation in MVA_{COMB} and in $\text{MVA}_{\text{PRECO}}$ is performed. The 1D cross-checks as described in the low-q^2 case above are also performed here as well as running with all run periods combined. An example fit and optimisation scan be seen in Fig. 4.27.

- $J/\psi\,(ee)$ modes - MVA_{COMB}: An exponential is used to model combinatorial background and additionally a `RooKeysPdf` of $B^0 \to XJ/\psi(\to e^+e^-)$ (R_{K^*}) or $B^+ \to XJ/\psi(\to e^+e^-)$ (R_K) is used for better fit stability in the lower sideband. For the `DecayTreeFitter` B meson mass, fits are performed to the lower and upper sidebands, and a 1D optimisation in MVA_{COMB} is performed.

- $J/\psi\,(ee)$ modes - $\text{MVA}_{\text{PRECO}}$: The $\text{MVA}_{\text{PRECO}}$ has some power to reduce partially reconstructed background in the J/ψ case also, though not as much as for the rare modes. This requires a full fit to the reconstructed B meson mass with three components. The signal component is a double-sided crystal ball, with its final parameters decided by a fit to the simulated J/ψ data. An exponential is used to model the combinatorial component whilst a `RooKeysPdf` of $B^0 \to XJ/\psi(\to e^+e^-)$, $B^+ \to XJ/\psi(\to e^+e^-)$ and $B_s^0 \to XJ/\psi(\to e^+e^-)$ combined is used to model partially-reconstructed background. This optimisation has as a fixed cut the optimal MVA_{COMB} cut from the previous bullet and a 1D optimisation in $\text{MVA}_{\text{PRECO}}$ is performed. This case together with the previous case defines a "1D+1D" optimisation. An example fit and optimisation scan be seen in Fig. 4.28.

- $\psi(2S)(ee)$ modes: Identical to the $\psi(2S)(\mu\mu)$ modes a 1D optimisation in MVA_{COMB} is performed.

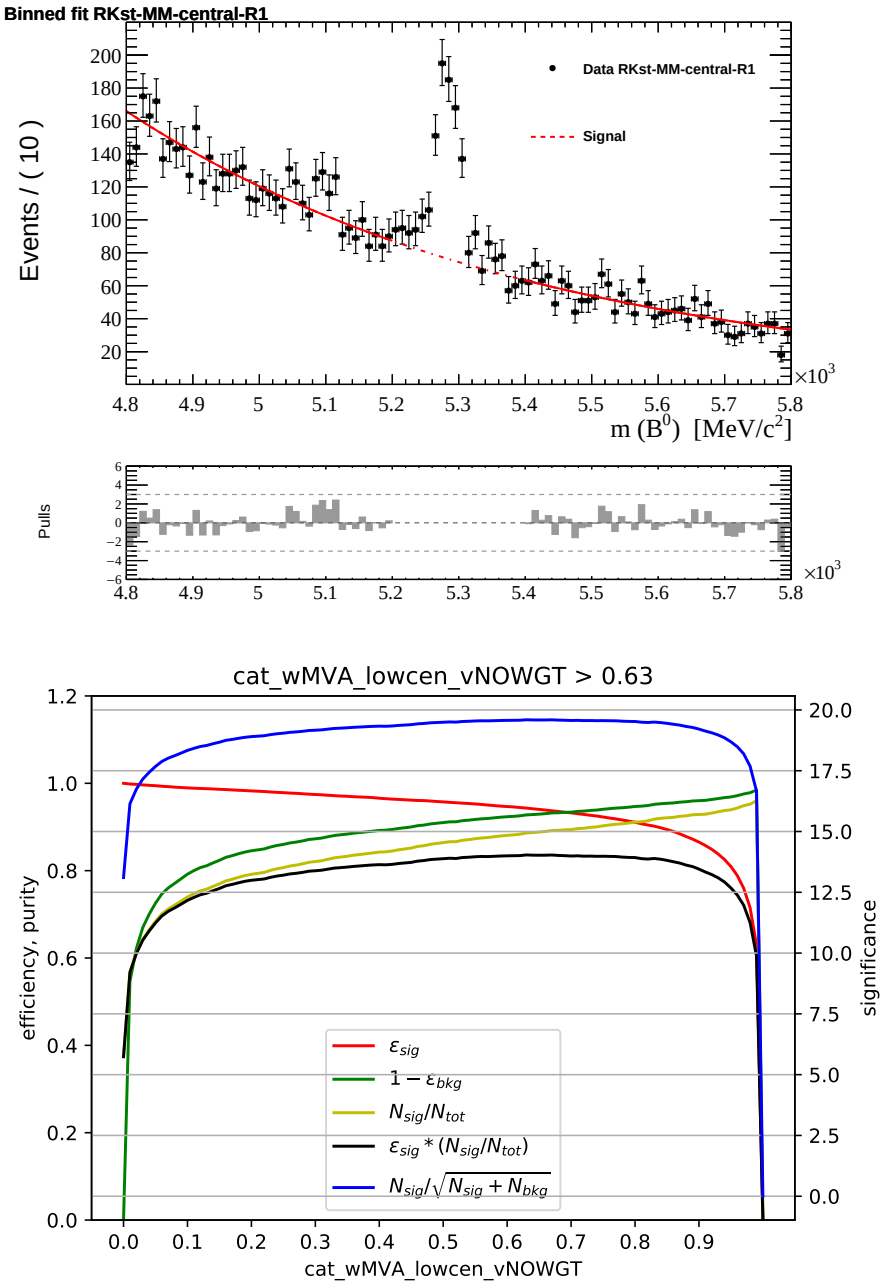


Figure 4.25: (top) Example fit before any MVA_{COMB} cut is applied for central-q^2 optimisation in RUN 1 for $B^0 \to K^{*0}\mu^+\mu^-$.The red line represents an exponential, with the solid lines denoting the fit region and the dotted line denoting the region the fit is extrapolated into. (bottom) Plot showing how the significance (blue), signal efficiency (red) and (1 − background efficiency) (green) evolve over the space of cuts. The MVA_{COMB} response is on the x-axis with the efficiency and significance on the left and right y-axis scales, respectively.

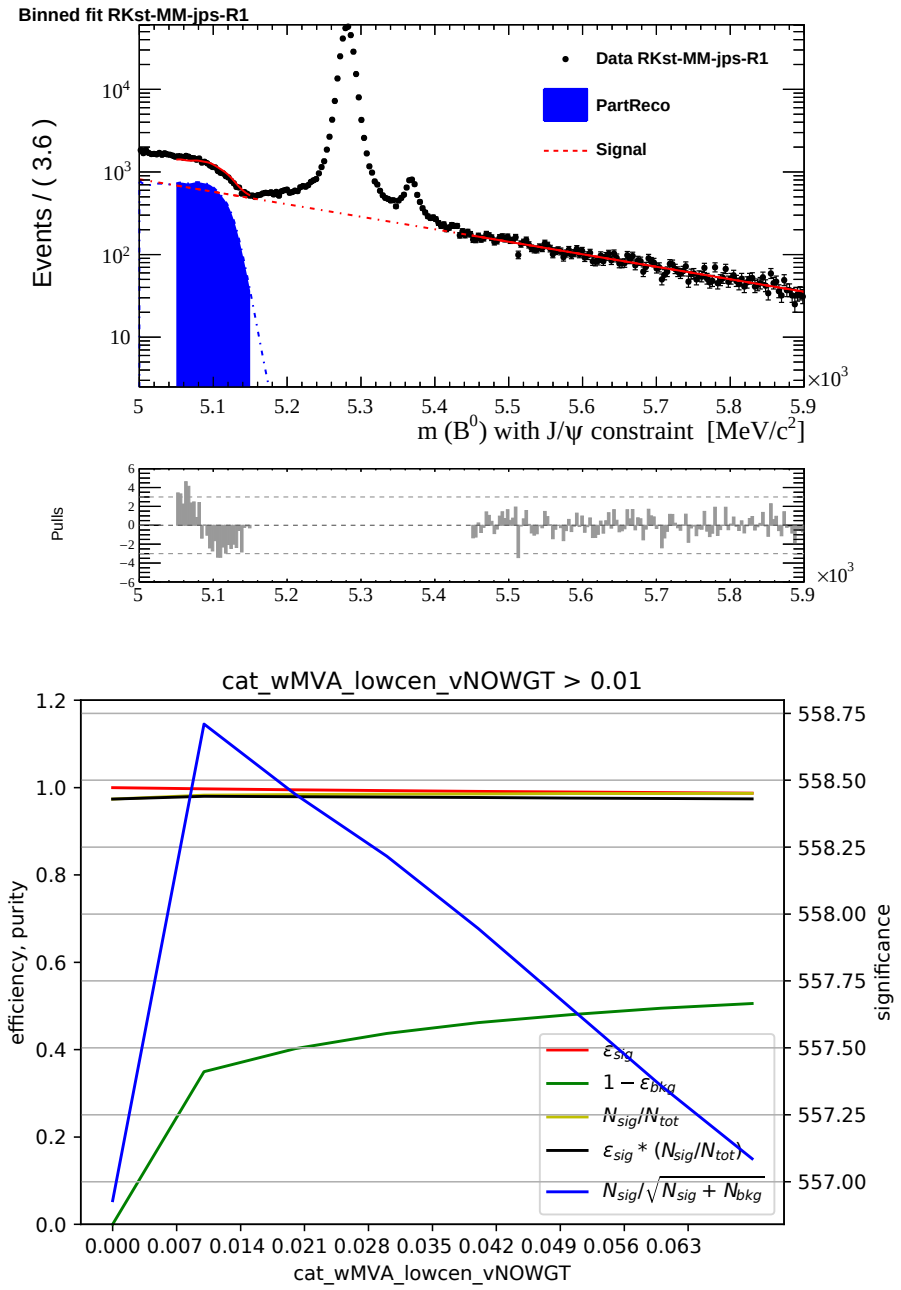


Figure 4.26: (top) Example fit before any MVA_{COMB} cut is applied for J/ψ optimisation in RUN 1 for $B^0 \to K^{*0} J/\psi\,(\to \mu^+\mu^-)$. The solid red line denotes the full model with all components combined, the red dotted line denotes an exponential and the blue shaded area is a `RooKeysPdf` of $B^0 \to X J/\psi\,(\to \mu^+\mu^-)$ modelling partially reconstructed background with the blue dotted line being this component outside of the fit region. (bottom) Plot showing how the significance (blue), signal efficiency (red) and (1 − background efficiency) (green) evolve over the space of cuts. The MVA_{COMB} response is on the x-axis with the efficiency and significance on the left and right y axis scales, respectively.

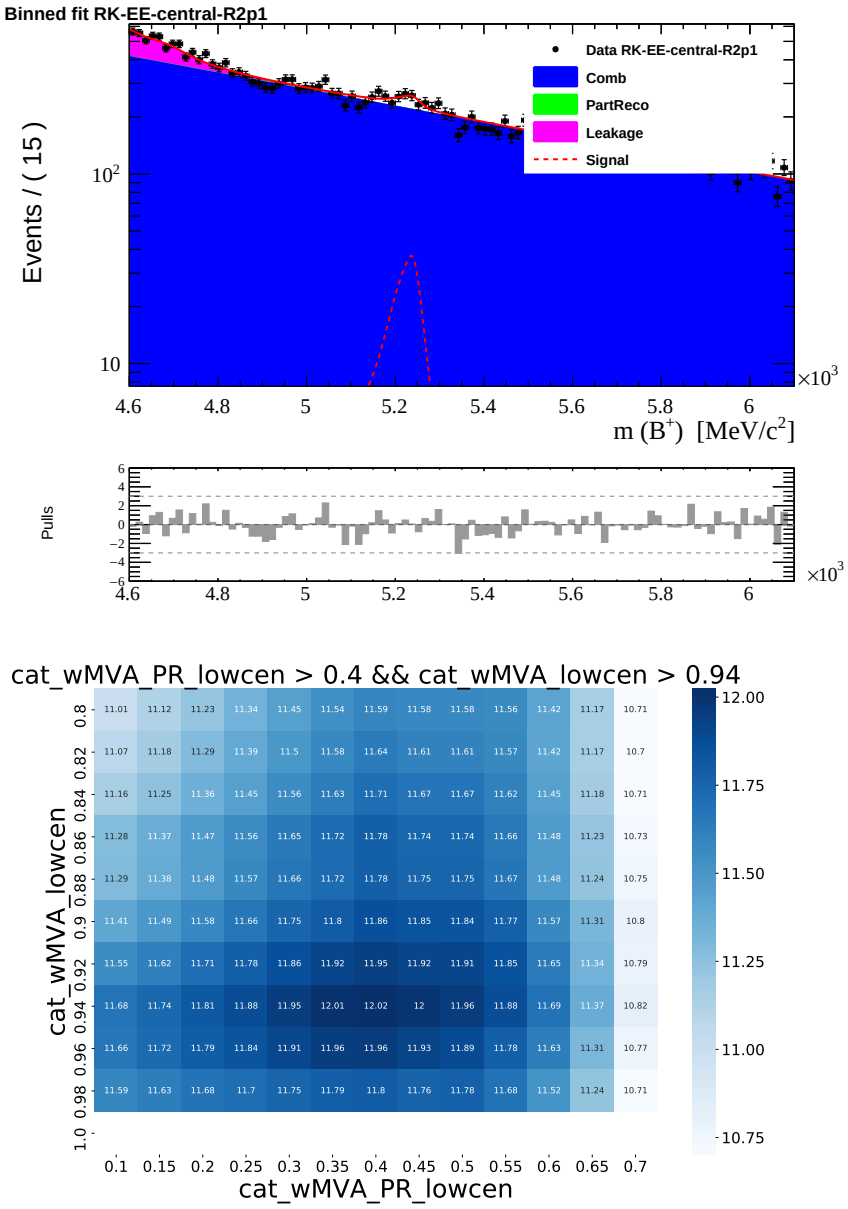


Figure 4.27: (top) Example fit before any MVA_{Comb} or $\text{MVA}_{\text{PReco}}$ cut is applied for central-q^2 optimisation in Run 2p1 for $B^+ \to K^+e^+e^-$. The solid red line represents the whole model with all components combined. The red dotted line represents the signal modelled by a double sided crystal ball function. The blue shaded area represents combinatorial background modelled by an exponential. The green shaded area represents partially reconstructed background modelled by a `RooKeysPdf` of partially reconstructed $B^0 \to K^{*0}e^+e^-$ MC. The pink shaded area represents J/ψ leakage modelled by a `RooKeysPdf` of $B^+ \to K^+J/\psi(\to e^+e^-)$ MC under the central-q^2 selection. (bottom) Plot showing the significance over the space of cuts scanned with $\text{MVA}_{\text{PReco}}$ and MVA_{Comb} on the x- and y-axes, respectively. A darker blue denotes a greater significance.

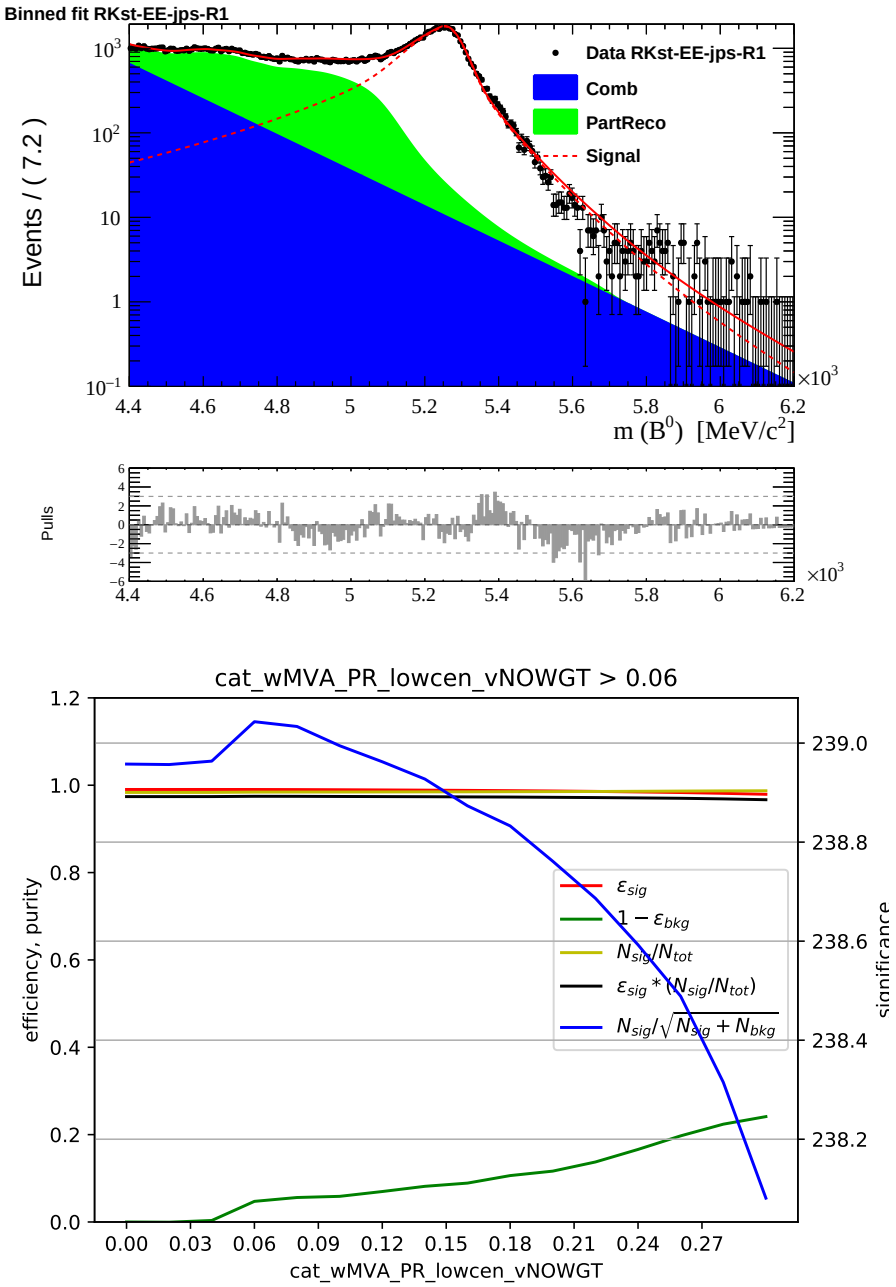


Figure 4.28: (top) Example fit with no $\text{MVA}_{\text{PRECO}}$ cut applied for J/ψ optimisation in RUN 1 for $B^0 \to K^{*0} J/\psi (\to e^+e^-)$ with the optimal MVA_{COMB} cut already applied. The solid red line represents the full model with all components combined. The dotted red line represents the signal modelled with a double sided crystal ball function. The blue shaded area represents the combinatorial background modelled by an exponential. The green shaded area represents the partially-reconstructed background modelled by a `RooKeysPdf` of $B^0 \to X J/\psi (\to e^+e^-)$, $B^+ \to X J/\psi (\to e^+e^-)$ and $B_s^0 \to X J/\psi (\to e^+e^-)$ MC combined. (bottom) Plot showing how the significance (blue), signal efficiency (red) and (1 − background efficiency) (green) evolve over the space of cuts. The $\text{MVA}_{\text{PRECO}}$ response is on the x-axis with the efficiency and significance on the left and right y-axis scales, respectively.

The optimal cuts resulting from the optimisation can seen in Tab. 4.22. Some of the cuts presented in the table have been unified to a common value, the reasons are as follows:

- Unified cuts for $\mu\mu$ control modes: For $J/\psi(\mu\mu)$ and $\psi(2S)(\mu\mu)$ modes the optimal cut values are always close to 0 and very compatible between years. As can be seen from the optimisation plots the signal efficiency is fairly flat and the significance does not change a great deal numerically. However the loss in combinatorial background does increase to plateau around a cut value of 0.05 which rejects around 50% of the combinatorial background according to the fits. Due to this all the cuts for $\mu\mu$ control modes are moved to > 0.05 in MVA_{Comb}.

- Unified cuts for ee control modes: For the MVA_{Comb} cut the same approach as the previous point is taken. For the $\text{MVA}_{\text{PReco}}$ a conservative cut of > 0.05 is applied as the majority of optimisations for both B^0 and B^+ find optimum cut values around 0.05.

- Unified cuts for ee rare modes: As a cross-check of the 2D optimisation in separate run periods another optimisation was performed combining all run periods together. The result of this optimisation was compatible with the results of the optimisations per run period.

- $B^0 \to K^{*0}e^+e^-$ low-q^2 : a conservative cut value is chosen to retain enough background for stable fits to be performed.

4.12.2 HOP

The HOP variables [131] are known to be able to further reduce partially reconstructed backgrounds. They exploit the expectation that B meson decay products should have no component of momentum orthogonal to the flight direction of the B meson. For a $B^0 \to K^{*0}\ell^+\ell^-$ decay if this is true then α_{HOP} defined in Eq. 4.3 should be unity, noting that p_{T} in this equation is the transverse momentum relative to the flight direction of the B meson and not the beam axis. This is very similar to how J/ψ mass constraints are used except α would be $\frac{m(J/\psi)}{m(ee)}$.

$$\alpha_{HOP} = \frac{p_{\text{T}}(B)}{p_{\text{T}}(\ell\ell)} \tag{4.3}$$

Table 4.22: Summary of the MVA_{Comb} and $\text{MVA}_{\text{PReco}}$ cut values resulting from the optimisation procedure.

Optimisation	**Run 1**	**Run 2p1**	**Run 2p2**
$B^+ \to K^+e^+e^-$ low-q^2	$\text{MVA}_{\text{Comb}} > 0.90$ && $\text{MVA}_{\text{PReco}} > 0.40$		
$B^+ \to K^+e^+e^-$ central-q^2	$\text{MVA}_{\text{Comb}} > 0.90$ && $\text{MVA}_{\text{PReco}} > 0.40$		
$B^+ \to K^+J/\psi(\to e^+e^-)$	$\text{MVA}_{\text{Comb}} > 0.10$ && $\text{MVA}_{\text{PReco}} > 0.05$		
$B^+ \to K^+\psi(2S)(\to e^+e^-)$	$\text{MVA}_{\text{Comb}} > 0.47$	$\text{MVA}_{\text{Comb}} > 0.36$	$\text{MVA}_{\text{Comb}} > 0.59$
$B^+ \to K^+\mu^+\mu^-$ low-q^2	$\text{MVA}_{\text{Comb}} > 0.70$	$\text{MVA}_{\text{Comb}} > 0.85$	$\text{MVA}_{\text{Comb}} > 0.85$
$B^+ \to K^+\mu^+\mu^-$ central-q^2	$\text{MVA}_{\text{Comb}} > 0.70$	$\text{MVA}_{\text{Comb}} > 0.80$	$\text{MVA}_{\text{Comb}} > 0.80$
$B^+ \to K^+J/\psi(\to \mu^+\mu^-)$		$\text{MVA}_{\text{Comb}} > 0.05$	
$B^+ \to K^+\psi(2S)(\to \mu^+\mu^-)$		$\text{MVA}_{\text{Comb}} > 0.05$	
$B^0 \to K^{*0}e^+e^-$ low-q^2	$\text{MVA}_{\text{Comb}} > 0.50$ && $\text{MVA}_{\text{PReco}} > 0.50$		
$B^0 \to K^{*0}e^+e^-$ central-q^2	$\text{MVA}_{\text{Comb}} > 0.90$ && $\text{MVA}_{\text{PReco}} > 0.40$		
$B^0 \to K^{*0}J/\psi(\to e^+e^-)$	$\text{MVA}_{\text{Comb}} > 0.20$ && $\text{MVA}_{\text{PReco}} > 0.05$		
$B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$	$\text{MVA}_{\text{Comb}} > 0.30$	$\text{MVA}_{\text{Comb}} > 0.55$	$\text{MVA}_{\text{Comb}} > 0.67$
$B^0 \to K^{*0}\mu^+\mu^-$ low-q^2	$\text{MVA}_{\text{Comb}} > 0.29$	$\text{MVA}_{\text{Comb}} > 0.54$	$\text{MVA}_{\text{Comb}} > 0.55$
$B^0 \to K^{*0}\mu^+\mu^-$ central-q^2	$\text{MVA}_{\text{Comb}} > 0.63$	$\text{MVA}_{\text{Comb}} > 0.77$	$\text{MVA}_{\text{Comb}} > 0.64$
$B^0 \to K^{*0}J/\psi(\to \mu^+\mu^-)$		$\text{MVA}_{\text{Comb}} > 0.05$	
$B^0 \to K^{*0}\psi(2S)(\to \mu^+\mu^-)$		$\text{MVA}_{\text{Comb}} > 0.05$	

When α_{HOP} deviates from unity, this means that energy is missing from the final state and, if this is greater than unity, this is most likely to be due to bremmstrahlung from electrons. Using this the measured momentum of the di-electron, $p(ee)_{meas}$, can be corrected, $p(ee)_{corr}$, as seen in Eq. 4.4. Since bremmstrahlung photons are emitted close to the direction of the electron this correction can be used to correct the longitudinal component of its momentum.

$$p(ee)_{corr} = \alpha_{HOP} \cdot p(ee)_{meas} \tag{4.4}$$

Using this corrected momentum the invariant mass can be recomputed resulting in the HOP mass variable. The resolution of this variable depends on the χ^2_{FD} since it is dependent on the vertex reconstruction and the resolution of the flight distance of the B meson. The resolution also degrades at larger q^2 values as the angle between the electron pair reduces making this harder to resolve, for this reason HOP is best suited to be used in the low- and central-q^2 regions.

When the hadronic system is only partially reconstructed α_{HOP} becomes less than unity. This results in much lower HOP mass values, thus, by cutting on the HOP mass partially reconstructed background can be further reduced after the $\text{MVA}_{\text{PReco}}$ is applied. This also provides some rejection power for combinatorial background,

however, this is already greatly suppressed by $\mathrm{MVA}_{\mathrm{COMB}}$.

In order to test whether the HOP mass has some rejection power a further optimisation is performed in this variable. The method is to take exactly the same optimisation procedure as before but to fix the $\mathrm{MVA}_{\mathrm{COMB}}$ and $\mathrm{MVA}_{\mathrm{PRECO}}$ cuts of the previous optimisation step. The optimisations are carried out in each of the run periods separately and an extra optimisation is performed on all the run periods combined. An example can be seen in Fig 4.29. A conservative cut value is chosen to be applied to the low-q^2 and central-q^2 cases, respectively, as can seen in Tab. 4.23.

Table 4.23: Summary of HOP mass cut values for ee low and central-q^2 regions.

Optimisation	**HOP mass cut** [MeV/c^2]
ee low-q^2	> 4800
ee central-q^2	> 4700

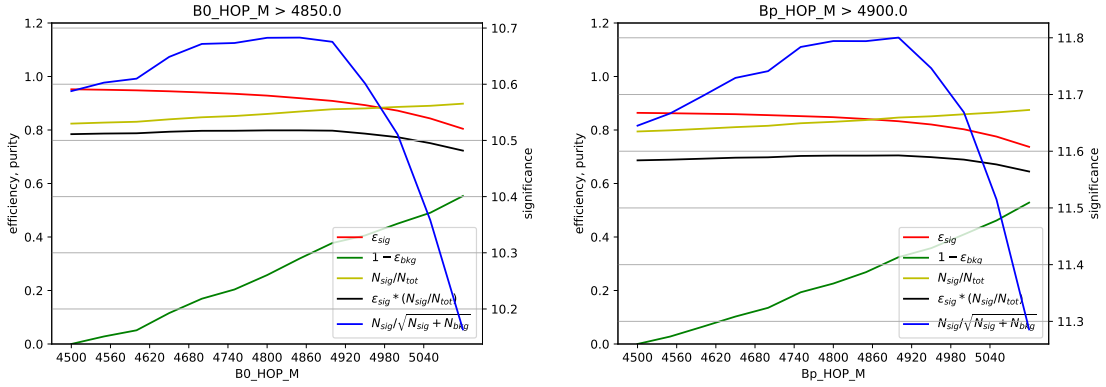


Figure 4.29: (left) Plot for low-q^2 optimisation with all runs combined for $B^0 \to K^{*0}e^+e^-$, showing how the significance (blue), signal efficiency (red) and (1 − background efficiency) (green) evolve over the space of cuts. The HOP mass is on the x-axis with the efficiency and significance on the left and right y-axes, respectively. (right) The same plot but for $B^+ \to K^+e^+e^-$ low-q^2 optimisation with all runs combined.

Chapter 5

Corrections

The determination of efficiencies, detailed in Sec. 6, is based on simulation so it is vital that the simulation used reproduces distributions from the collision data correctly. Therefore, where this is not the case corrections need to be made so that the two are compatible. There are five main correction steps, which can be grouped into two categories: efficiency corrections and reweighting of distributions. The steps are summarised as follows:

- **PID** - The `ProbNN` type variables are used in the selection but are not well modelled in simulation. Therefore a per-event efficiency correction, w_{PID}, is applied making use of `PIDCalib` calibration samples. This procedure is described in detail in Sec. 6.4.

- **Tracking** - An efficiency correction for reconstructing electrons as long tracks within LHCb is applied, w_{TRK}, these corrections are computed according to Ref. [132] using a Tag and Probe approach. Currently no tracking corrections are available for muon modes. This procedure is described in Sec. 5.2.

- **Trigger** - Variables related to the L0 and HLT trigger decisions are not well modelled in simulation. Thus per-event weights, w_{L0} and w_{HLT}, are determined using a Tag and Probe approach and utilising the TISTOS method on J/ψ resonant mode samples. This procedure is described in detail in Secs. 6.5 and 6.6.

- **Kinematics and multiplicity** - The simulation is known to not provide a good description of some kinematic and multiplicity related variables at

generator level. A BDT reweighter is used to produce corrections, $w_{\text{Mult\&Kin}}$, for these variables based on J/ψ resonant mode samples, a description of which can be found in Sec. 5.3

- **Reconstruction** - The simulation is known to badly model variables related to reconstruction of vertices and flight distances. Dedicated BDT reweighters are used to produce corrections, w_{Reco}, to correct for this in each trigger category for $\mu\mu$ and ee modes, respectively, based on J/ψ resonant mode samples, a description of the procedure can be found in Sec. 5.3.

5.1 Strategy

The analysis makes use of two correction chains known as the prior and nominal chain. The prior chain is used to determine a kinematic and multiplicity correction in the most unbiased way possible. The nominal chain then uses the kinematic and multiplicity corrections determined in the prior chain instead of calculating them again. The various steps of the two chains are shown in Fig. 5.1 and discussed in more detail as follows:

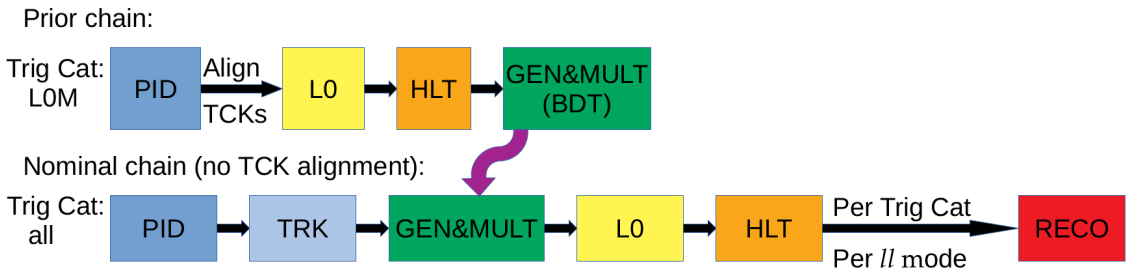


Figure 5.1: A diagrammatic representation of the the prior and nominal chains.

- **Prior Chain** - The aim of the prior chain is to access kinematic and multiplicity corrections in the most unbiased way possible in order to correct for generator level misalignments introduced by PYTHIA. These corrections use only the L0M (L0M inclusive) trigger category, which corresponds to a large sample size is relatively clean. The chain then proceeds as follows: PID corrections are computed; the L0-TCKs thresholds are aligned so that simulation and data are in perfect agreement; the L0 corrections are computed followed by the HLT corrections and finally the kinematic and multiplicity corrections are determined. With this procedure the kinematic and multiplicity corrections can be safely used for both muon and electron modes.

- **Nominal Chain** - The aim of the nominal chain is to correct simulation so that it matches collision data. These corrections are performed on both trigger categories, L0I and L0L! and no TCK alignment is performed. The chain then proceeds as follows, PID corrections are computed, then tracking corrections for the electron modes are added, after this the kinematic and multiplicity corrections determined in the prior chain are applied, next the L0 and HLT corrections are computed and finally reconstruction corrections are computed per $\ell^+\ell^-$ mode per trigger category. The key aspect of this chain is to ensure the trigger corrections are taken from samples with corrected kinematics and multiplicity.

Once the corrections have been determined they are always used in the calculation of efficiencies, but not when determining mass shapes. Corrections are determined in a sequential manner, that is, the n^{th} correction is always applied after the $(n-1)^{th}$ correction. To reduce the dependence on binning schemes corrections are interpolated when applied to simulated samples, except for tracking corrections which were produced externally. Comparisons between corrected simulation and collision data can be found in Appendix E., for the Run 2p1 L0L! category, the level of agreement here is representative of the other trigger category and run periods. The rest of this section aims to detail the tracking, kinematic, multiplicity and reconstruction corrections, after a discussion of the TCK alignment in the prior chain.

5.1.1 TCK alignment for the prior chain

As discussed previously, during the various data taking periods the TCK often changed, but in simulation only one TCK is used. This can lead to misalignments between collision data and simulation. Depending on the TCK thresholds there are two situations that can occur:

- **Case 1** - In this case the simulation TCK encodes the loosest thresholds with some fractions of collision data being recorded with tighter TCKs. In this case it is easy to resample the relevant fractions of simulation with the tighter thresholds to align it with collision data. This was the case for the HLT1 alignment discussed in Sec. 4.5.

- **Case 2** - In this case the simulation TCK encodes tighter thresholds than some fractions of collision data being recorded with looser TCKs. In this case to

align simulation with collision data the data taken with looser TCK thresholds must be removed and in some data-taking years this can be a large portion of the available statistics. This alignment is only performed in the prior chain on L0M to access an unbiased as possible kinematic and multiplicity correction and not in the nominal chain due to the large loss of statistics this would incur.

For the TCK alignment performed in the prior chain the situation arrived at is Case 2. Since L0M is the trigger of interest, the threshold of interest is that of the muon p_{T}. It is found that in 2011 and 2012 a very small amount of collision data ($< 1\%$) is taken with different trigger thresholds whilst in 2018 everything is perfectly aligned. In 2015 the situation is slightly worse whilst 2016 has only 70% of its data in one magnet polarity aligned and 2017 has around 40% of its data aligned. There are tables presenting each of the TCKs and their thresholds in Appendix A.2, showing which TCKs are aligned and what fraction of data was recorded with them. In order to align simulation and collision data a cut is made on the TCK value of the simulated TCK. In some cases, such as 2017, fractions of the data were recorded with a tighter TCK than simulation which would make it a Case 1 situation, however, since this is not the case for all the TCKs in any of the years a blanket cut on the simulation TCK is made.

5.1.2 L0E trigger alignment

As can be seen in the tables in Appendix A.2, the $E_T^{L0ADC}(e)$, which is the transverse energy of the electron in ADC counts, is misaligned between simulation and collision data in a number of years. In order to unify the kinematics of the electron trigger decision whenever an L0E type trigger category is used, which is only in the nominal chain, an extra condition is added when requiring the trigger decision as seen in Eq. 5.1.

$$\texttt{E\{1,2\}\_L0Electron\_TOS} \rightarrow \texttt{E\{1,2\}\_L0Electron\_TOS} \,\&\, E_T^{L0}(e\{1,2\}) > \texttt{TCK\_Threshold} \tag{5.1}$$

Determining the `TCK_Threshold` requires converting from ADC counts to values of transverse momentum. In 2011 and 2012 this conversion is 24 MeV per ADC count and for the remaining years it is 20 MeV per ADC count. It should be noted that the transverse energy at the L0 trigger is not accessible so the projected transverse energy at the calorimeter is used in the tuples using the `L0CaloTool`. For years such as 2018 where everything is perfectly aligned a simple conversion may be made. However, in other years a compromise must be made which aligns collision data and simulation so there are not regions were there is no simulation but there is collision

data and vice versa. By avoiding these situations large w_{L0} can be avoided. The E_{T} values which are selected are presented in Tab 5.1.

Year	$E_T^{L0}[MeV]$
2011	>2500.
2012	>3000.
2015	>3000.
2016	>2700.
2017	>2700.
2018	>2400.

Table 5.1: L0E threshold alignment selections applied offline to both data and simulation to re-define the E{1, 2}_L0Electron_TOS trigger decision.

5.2 Tracking Corrections (w_{TRK})

Tracking efficiency corrections for electron final states are taken from the work detailed in Ref. [132] and are based on a tag and probe method. These tracking corrections, w_{TRK}, encode the efficiency of reconstructing electron tracks as long tracks within LHCb given that the track was reconstructed in the VELO. This efficiency is then calculated as function of p_{T}, η and ϕ so that in each bin the following formula is used:

$$w_{\mathrm{TRK}} = \frac{\epsilon(long|velo)^{data}}{\epsilon(long|velo)^{MC}}. \tag{5.2}$$

The maps produced using this procedure can be seen in Appendix F. The final tracking weight is the product of the tracking weights for each electron such that:

$$w_{\mathrm{TRK}} = w_{\mathrm{TRK}}(e_1) \times w_{\mathrm{TRK}}(e_2) \tag{5.3}$$

There are currently no muon tracking weights available, although LHCb is actively working to produce them and they will be included in this analysis at a later date. Tracking weights for the hadrons are not used as they are expected to cancel in the ratios.

5.3 Kinematic, Multiplicity and Reconstruction Corrections ($w_{\text{Mult\&Kin}}$ & w_{Reco})

Kinematic and multiplicity corrections are computed using the prior chain, whilst reconstruction corrections are computed as the final step of the nominal chain after all other corrections have been applied. In order to correct for the kinematics the B meson p_{T} and η are used as proxies and corrected simultaneously. In simulated data, the p_{T} distributions are generally harder than in collision data whilst the η distributions are shifted to lower values.

Multiplicity corrections are required due to a number of effects which are not well modelled in simulation, such as low momentum particles, particles originating from photon conversions or other secondary interactions and backsplash from particle showers with various parts of the detector. Since these factors affect different sub-detectors in differing degrees there are a number of choices of occupancy proxies that can be made. One commonly used proxy is `nSPDHits`, however, correcting for this degrades the level of agreement for other occupancy proxies with collision data. Also the agreement in `nSPDHits` is affected by mismodelling of the material budget in the M1 station in simulation, so, this variable is not used to correct for the multiplicity. In this analysis multiplicity corrections are based on the number of unique tracks in an event (`nTracks`) as this is also used to parametrise the PID corrections. This choice of variable leads to much better agreement between collision data and simulation in most occupancy variables. However, for the reasons outlined above, `nSPDHits` is still not in good agreement with collision data, this is important as there is a `nSPDHits` requirement on the L0 trigger thresholds. However, the corrected distributions of the `nSPDHits` variable for both muons and electrons show similar behaviour so a large amount of this mismodelling should cancel in the ratios. For reconstruction corrections the χ^2 of the B vertex fit, χ^2_{vtx}, and impact parameter, χ^2_{IP}, are used as proxies and corrected simultaneously. This correction is designed to correct for differences between data and simulation when candidates are reconstructed.

To compute the corrections, in both cases, a GBReweighter, which is a type of BDT, from the hep-ml library [133] is chosen and its parameters are optimised. By computing the weights in this manner, correlations between the input variables are taken into account, in a similar manner to the way in which the MVA of Sec. 4.11 exploits correlations between input variables to determine whether an

event is signal-like or background-like.

The kinematics and multiplicity corrections are not expected to vary between muon and electron modes, hence the use of the L0M trigger category to compute the corrections that will be applied to both. However, there are differences in reconstructing muon and electron modes meaning that corrections computed using one mode cannot be ported to the other safely and this is also the case within trigger categories. Due to this the process is split into two steps:

- **Kinematics and multiplicity correction** - Reweighting of the B meson kinematics, using the B meson p_{T} and η, and multiplicity, using `ntracks`. This is trained using data from J/ψ resonant muon modes in the L0M trigger category based upon the prior chain.

- **Reconstruction correction** - Reweighting of the reconstructed quantities, the B meson χ^2_{vtx} and χ^2_{IP}. This is trained using data from J/ψ resonant modes for both lepton species and for both trigger categories, L0I and L0L!.

Kinematics and multiplicity The sample used to derive kinematics and multiplicity corrections is the J/ψ resonant muon mode using the L0M trigger category, which provides the highest statistics and is relatively clean. Taking corrections from the L0I trigger category yields compatible results, however, the `L0Muon` inclusive category is used as it yields larger statistics. The full selection, as described in Sec. 4, minus the MVA requirements, is applied so that a clear peak can be seen at the B meson mass. There will still be a small amount of residual background, in order to reduce this the ${}_s\mathcal{P}lot$ method [134] is used. This technique can be used to subtract statistically the background from the collision data by performing fits to the invariant mass of the final state particles involved, this provides an even cleaner sample from which to determine corrections. The fits used in the ${}_s\mathcal{P}lot$ procedure can be found in Appendix B. As described in Sec. 4.11 a *k-fold* technique with four folds is used in order to prevent biases. In this case 3/4 of the training sample is used to determine the correction and this is then applied to the 1/4 which is left out. If the correction is applied to samples not used in the training the average of the folds is taken. During the training process the reconstructed p_{T} and η are used to produce the corrections. However when applying the weights the true p_{T} and η, which are the generator values of these variables before they are propagated and reconstructed through the detector, are used. This is

because the aim of this step is to correct for generator level differences. This is especially important for electrons which have a poor kinematic resolution due to bremsstrahlung. For signal candidates which are classified as ghosts, which do not exist at generator level, the corrections must be applied to the reconstructed information. This correction is applied to both `MCDecayTuple` and `DecayTuple`. In the RUN 2 trainings there is a cutoff applied for the `nTracks` variable of < 330 as the data distribution shows a long tail towards high values whilst the simulation has almost no events in this area. Using this cutoff has a negligible effect on efficiencies but prevents very large weights in the region > 330.

Reconstruction To correct for reconstruction effects the same general procedure is followed but with some differences. In this case a BDT is trained for $\mu\mu$ and ee modes with one per trigger category, L0I and L0L!, so that four separate BDTs are trained. These BDTs are then only applied back to the trigger category that was used to train them. Using the ${}_s\mathcal{P}lot$ method to select the collision data sample used in the training here is not feasible due to a bias introduced by correlations between the ${}_s\mathcal{W}eights$ computed and the training variables, hence, a $60\,\text{MeV}/c^2$ window around the J/ψ mass is applied to select a clean data sample. To account for possible correlations between the training variables the reconstructed B meson p_T and η are also added as training variables. The corrections are then applied back to the reconstructed variables so they can only be applied to `DecayTuple`.

Validation There are three main checks performed to validate the effectiveness of the corrections. Kolmogorov-Smirnov (K-S) tests can be performed before and after the correction. This test can quantify how alike two distributions are and a smaller value is indicative of more similar distributions. Profiles of the weights in the training variables can be compared per fold, an example seen in Fig. 5.2 for kinematic weights in 2012 shows that the folds have consistent weight profiles, this is seen in general over all years and correction steps. Data and simulation comparison plots in a number of variables can be produced showing the agreement between tightly selected data (as in the reconstruction step) and corrected simulated data, some examples are shown in Figs. 5.3 and 5.4 which show the improvement of the simulated data with the corrections applied. Further evidence of the improvement gained by these corrections can be seen in the various flatness plots presented in Sec. 9.2.

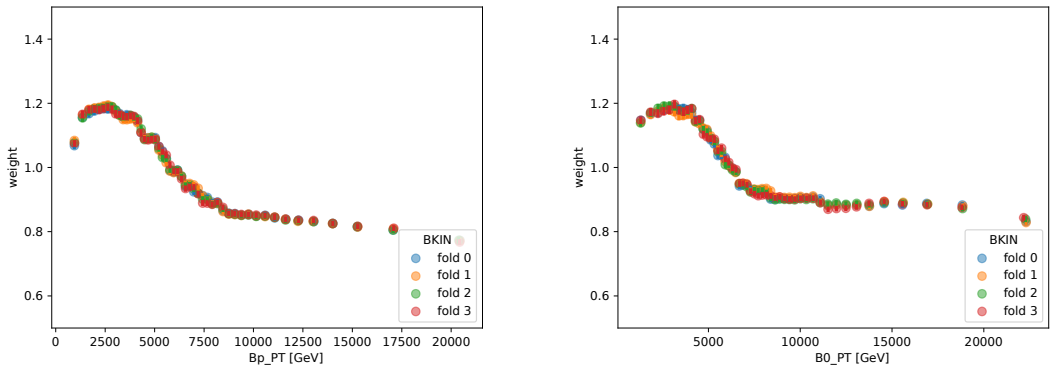


Figure 5.2: Weight profiles for all trained folds for kinematic weights, the different coloured points represent the different folds. Left: 2012 R_K, right: 2012 R_{K^*}

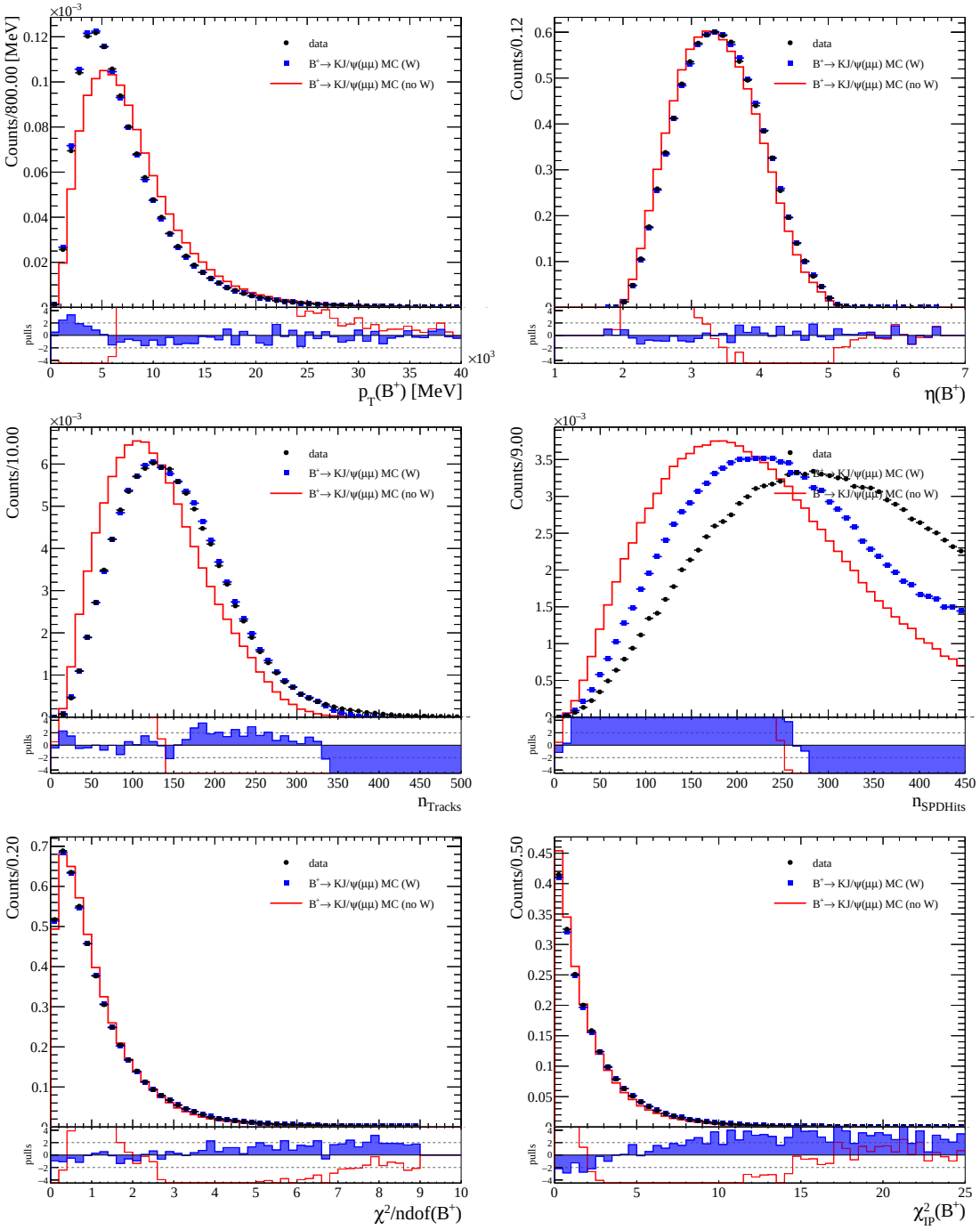


Figure 5.3: Data/simulation comparison plots for 2016 `L0Muon` R_K samples. The variables shown are the input variables used in the reweighter (and `nSPDHits`).

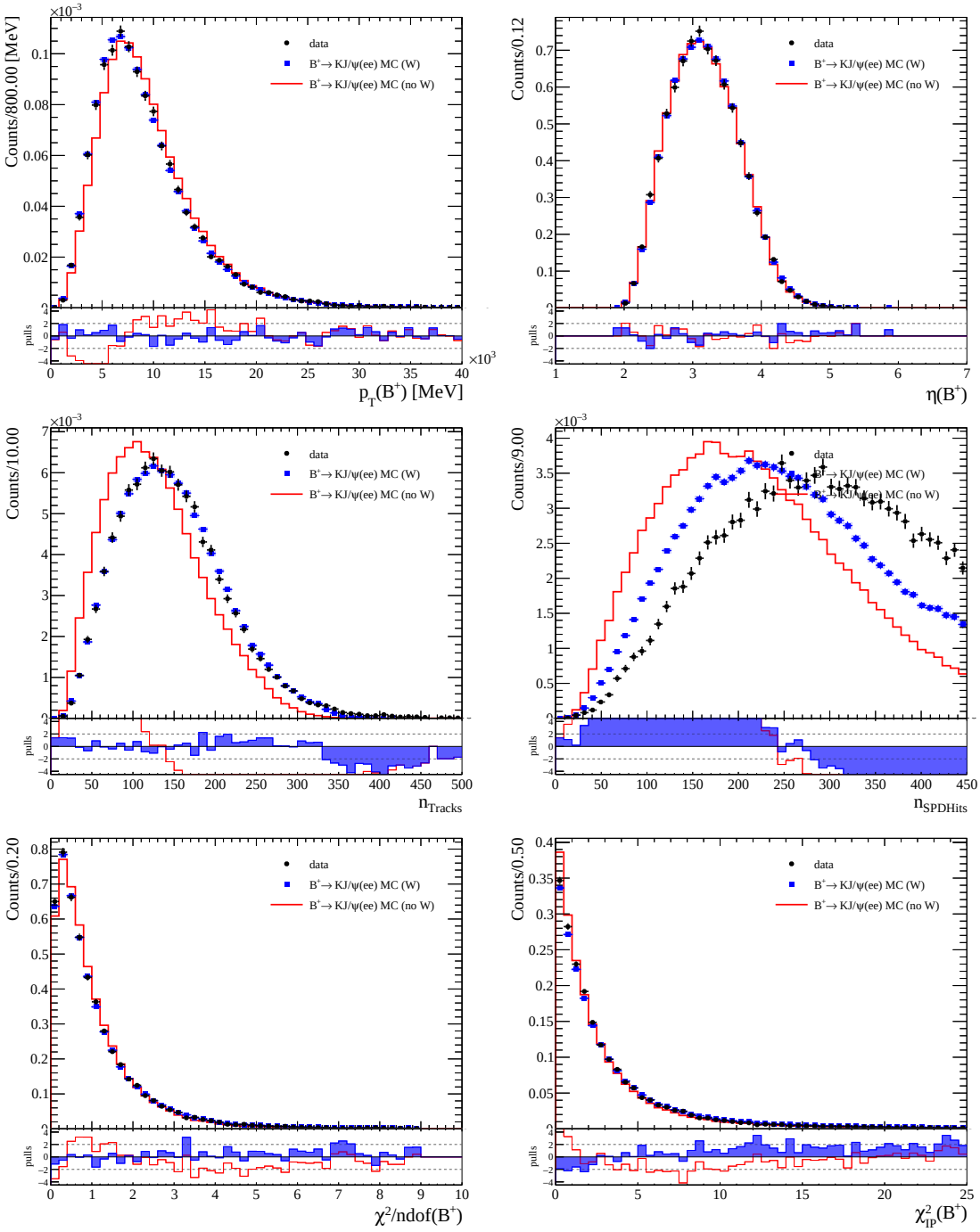


Figure 5.4: Data/simulation comparison plots for 2016 `L0Electron` R_K samples. The variables shown are the input variables used in the reweighter (and `nSPDHits`).

Chapter 6

Efficiencies

To determine the total efficiency for each signal decay mode of interest Eq. 6.1 is used.

$$\varepsilon_{\text{tot}} = \varepsilon_{\text{geo}} \times \varepsilon_{\text{flt,reco,strip|geo}} \times \varepsilon_{\text{MVA\&sel\&trg\&PID\&reco,strip|flt}} \tag{6.1}$$

The terms in Eq. 6.1 correspond to the following:

$\boldsymbol{\varepsilon_{\text{geo}}}$ - Is the efficiency relating to having all the final state particles within LHCb's acceptance (Sec. 6.2),

$\boldsymbol{\varepsilon_{\text{flt,reco,strip|geo}}}$ - Is the efficiency of events passing the filtering step given that the final state particles are within LHCb's acceptance (Sec. 6.3),

$\boldsymbol{\varepsilon_{\text{MVA\&sel\&trg\&PID\&reco,strip|flt}}}$ - Is the efficiency of events passing the full selection, with all corrections applied, given that events have passed the filtering step (Sec. 6.8).

The term $\varepsilon_{\text{MVA\&sel\&trg\&PID\&reco,strip|flt}}$ encompasses many different terms relating to the selection and corrections to simulation which can be broken down as follows:

PID - Efficiency corrections for simulated samples for the PID cuts chosen are applied as PID weights, w_{PID}, discussed further in Sec. 6.4,

Tracking - Tracking corrections, w_{TRK}, for samples involving electrons are added as discussed in Sec. 5.2,

Kinematics and Multiplicity - Corrections due to the mismodelling of multiplicity and kinematic variables for the B meson, $w_{\text{Mult\&Kin}}$, originating from the

generator level, are applied as discussed in Sec. 5.3,

L0 Trigger - Efficiency corrections for simulated samples due to the mismodelling of the L0 trigger decision in simulation are applied as L0 weights, w_{L0}, discussed further in Sec. 6.5,

HLT Trigger - Efficiency corrections for simulated samples due to the mismodelling of the HLT trigger decision in simulation are applied as HLT weights, w_{HLT}, discussed further in Sec. 6.6,

Reconstruction - Corrections due to mismodelling for reconstructed variables for the B meson, w_{Reco}, are applied as discussed in Sec. 5.3

Selection - The efficiency of the selection given events passing the filtering step is determined, this includes:

- The selection criteria outlined in Tabs. 4.1 and 4.5, corresponding to the stripping criteria and various selection criteria, are applied with the full correction chain applied. The vetoes for exclusive backgrounds outlined in Tabs. 4.14 and 4.10 are also applied with the full correction chain applied,
- When PID corrections are applied the PID selection is replaced by per event PID weights, w_{PID},
- When L0 trigger corrections are applied the L0 trigger category selection, outlined in Sec. 4.4, is modified by w_{L0} to obtain a better estimate of the L0 trigger efficiency,
- When HLT trigger corrections are applied the HLT line selection, outlined in Sec. 4.5, is modified by w_{HLT} to obtain a better estimate of the HLT trigger efficiency,

MVA - The efficiency of the MVA selection, detailed in Tab 4.22, is applied.

It should be noted that efficiencies for the non-resonant modes must remain blinded at this stage but those for the J/ψ and $\psi(2S)$ resonant modes are not blinded. The remainder of this section will discuss the geometric efficiency (Sec. 6.2), filtering efficiency (Sec. 6.3), PID efficiency corrections (Sec. 6.4), trigger efficiency corrections (Secs. 6.5 and 6.6) and discuss in further detail the final efficiency calculation

(Secs. 6.8). A discussion of the various sources of correlations is also included in Sec. 6.1. The effect of smearing, due to an incompatible resolution between simulation and data, is also discussed in Sec. 6.7, although this has not yet been factored into the efficiency calculation.

6.1 Sources of Correlations

A number of sources of correlations due to the selection and correction procedure arise and should be in consideration when determining the final efficiencies or extracting results. Some of the main sources of correlation are as follows:

Porting corrections between muon and electron final states - A correlation is introduced where corrections are determined using the muon mode and applied to both the muon and the electron mode. This is the case for the w_{L0} correction in the L0I trigger category and for the kinematic and multiplicity correction which is evaluated on L0M.

Determining efficiencies on samples used to determine corrections - For a correction which involves collision data and simulation ratios a correlation is present if the corrections are derived on a simulated sample to which those same corrections will then be applied back to. To minimise this effect the correction can be taken from the other decay mode, so that B^0 modes take corrections from B^+ modes and vice versa. This relies on the results of the corrections being compatible between B^0 and B^+ modes. This type of correlation is also present for mass fits on collision data for the J/ψ resonant modes, since the same data are used to determine the corrections.

Overlap of events in samples used to determine corrections - Where corrections make use of the L0L (L0L inclusive) category this includes events which are classified as L0I. If corrections are also taken from the L0I category a correlation is present due to events which are both L0L and L0I and are present in both samples used to derive the corrections. This can also be the case for PID where the calibration data samples are collected from the same decay mode as studied in this work, the selection is not exactly the same but there is still some overlap (and the overlap is not always possible to determine due to lack of information in some of these calibration samples) which causes a correlation. These correlations are expected to be purely statistical.

In order to fully account for correlations, for each efficiency term an associated efficiency and correlation matrix would be computed which could be injected into the final fit. At this stage a bootstrapping procedure is followed to evaluate the statistical correlation. The bootstrapping accounts for the limited size of the samples used to calculate the corrections and overlaps between samples used to derive corrections which induce a correlation. To perform the bootstrapping in collision data and simulation 100 Poission weights are generated for each reconstructed candidate. The Poisson distribution used to generate these weights has a mean of 1 and the generator seed is set by a combination of the `RunNumber` and `EventNumber` which yields a unique seed that can be used to repeat the process. In this way, 100 different corrections can be obtained associated with 100 different collision data bootstrapped samples. How these bootstrapped corrections are handled in the final efficiency calculation is discussed in Sec. 6.8.

6.2 Geometric Efficiency (ε_{geo})

As was discussed in Sec. 4.1, there is an efficiency associated with whether the final state particles of a decay end up in LHCb's acceptance. Since this requirement is imposed whilst generating the simulation, the geometric efficiency, ε_{geo}, is produced centrally and can be looked up by analysts on a central webpage. The values obtained from the webpage are presented in Tab. 6.1, they are given per magnet polarity and per year.

6.3 Filtering Efficiency ($\varepsilon_{\text{flt,reco,strip|geo}}$)

In order to save disk space some simulation productions are subject to a filtering step where a version of the stripping is run before the simulated samples are given to analysts. Events which do not pass the filtering stage are not retained so an efficiency, $\varepsilon_{\text{flt,reco,strip|geo}}$, is provided to analysts centrally. Not all the samples used in this analysis are filtered and in the future those samples which are filtered will be reproduced unfiltered so that all steps after the geometric selection can be run by analysts and not centrally managed. The reason for this is that for filtered samples a selection on `nSPDHits` is made to the `MCDecayTuple`, this makes it very difficult to assess the efficiency of the `nSPDHits` cut. In Tab. 6.2 the filtering efficiencies can be seen, if the value is 100% the sample was not filtered.

Table 6.1: Geometric efficiency (ε_{geo}) for $B^0 \to K^{*0} J/\psi (\to \ell^+\ell^-)$, $B^+ \to K^+ J/\psi (\to \ell^+\ell^-)$, $B^0 \to K^{*0}\ell^+\ell^-$ and $B^+ \to K^+\ell^+\ell^-$ samples. MD and MU refer two magnet polarities.

Period	$B^0 \to K^{*0} J/\psi (\to \mu^+\mu^-)$	$B^0 \to K^{*0} J/\psi (\to e^+e^-)$	$B^+ \to K^+ J/\psi (\to \mu^+\mu^-)$	$B^+ \to K^+ J/\psi (\to e^+e^-)$
2011-MD	(15.39±0.03) %	(15.30±0.03) %	(15.90±0.03) %	(15.81±0.04) %
2011-MU	(15.39±0.04) %	(15.33±0.04) %	(15.95±0.03) %	(15.89±0.06) %
2012-MD	(15.75±0.03) %	(15.60±0.03) %	(16.28±0.02) %	(16.26±0.06) %
2012-MU	(15.75±0.03) %	(15.66±0.03) %	(16.28±0.02) %	(16.18±0.06) %
2015-MD	(16.70±0.04) %	(16.66±0.04) %	(17.29±0.05) %	(17.62±0.63) %
2015-MU	(16.79±0.04) %	(16.61±0.04) %	(17.38±0.05) %	(16.80±0.50) %
2016-MD	(16.77±0.04) %	(16.68±0.04) %	(17.35±0.06) %	(17.20±0.06) %
2016-MU	(16.70±0.04) %	(16.68±0.04) %	(17.24±0.06) %	(17.39±0.06) %
2017-MD	(16.71±0.06) %	(16.56±0.06) %	(17.39±0.06) %	(17.27±0.06) %
2017-MU	(16.70±0.06) %	(16.62±0.06) %	(17.40±0.06) %	(17.38±0.06) %
2018-MD	(16.68±0.06) %	(16.56±0.06) %	(17.31±0.06) %	(17.36±0.06) %
2018-MU	(16.63±0.06) %	(16.74±0.06) %	(17.40±0.06) %	(17.25±0.06) %
Period	$B^0 \to K^{*0}\mu^+\mu^-$	$B^0 \to K^{*0}e^+e^-$	$B^+ \to K^+\mu^+\mu^-$	$B^+ \to K^+e^+e^-$
2011-MD	(15.70±0.03) %	(15.96±0.03) %	(16.16±0.04) %	(16.10±0.04) %
2011-MU	(15.71±0.04) %	(16.01±0.04) %	(16.27±0.04) %	(16.03±0.04) %
2012-MD	(16.03±0.03) %	(16.33±0.03) %	(16.88±0.06) %	(16.90±0.04) %
2012-MU	(15.99±0.03) %	(16.37±0.03) %	(16.98±0.06) %	(16.86±0.04) %
2015-MD	(17.09±0.04) %	(17.36±0.04) %	(17.65±0.06) %	(17.55±0.06) %
2015-MU	(17.07±0.04) %	(17.35±0.04) %	(17.64±0.06) %	(17.61±0.06) %
2016-MD	(17.00±0.04) %	(17.33±0.04) %	(17.66±0.06) %	(17.60±0.07) %
2016-MU	(17.05±0.04) %	(17.30±0.06) %	(17.72±0.06) %	(17.58±0.06) %
2017-MD	(17.09±0.06) %	(17.42±0.06) %	(17.64±0.06) %	(17.50±0.06) %
2017-MU	(17.16±0.06) %	(17.23±0.06) %	(17.70±0.06) %	(17.54±0.06) %
2018-MD	(17.04±0.06) %	(17.36±0.06) %	(17.64±0.06) %	(17.52±0.06) %
2018-MU	(17.00±0.06) %	(17.34±0.06) %	(17.61±0.06) %	(17.50±0.06) %

6.4 PID efficiency

The PID responses in simulated samples are not reliably modelled, with `ProbNN` type variables particularly affected. In order to compute reliable efficiencies for the PID cuts detailed in Sec. 4.7 calibration samples collected during run periods are used. Using these samples efficiency maps for each of the relevant particle species can be computed. For muons and hadrons these maps are computed with a KDE based approach using the aforementioned calibration samples. For electrons a method known as fit and count is employed making use of the calibration samples and some necessary simulated samples. For the efficiencies used to be valid the selection of the calibration samples and the samples used in the analysis are aligned, referred to as prior cuts in Sec. 4.7. The calculated efficiencies act as weights, w_{PID}, to correct for the actual efficiency that should have been measured if the simulation accurately reproduced the PID responses.

Table 6.2: Filtering efficiency ($\varepsilon_{\text{flt,reco,strip|geo}}$) for the $B^0 \to K^{*0}J/\psi(\to \ell^+\ell^-)$, $B^+ \to K^+J/\psi(\to \ell^+\ell^-)$, $B^0 \to K^{*0}\psi(2S)(\to \ell^+\ell^-)$, $B^+ \to K^+\psi(2S)(\to \ell^+\ell^-)$, $B^0 \to K^{*0}\ell^+\ell^-$ and $B^+ \to K^+\ell^+\ell^-$ samples

Period	$B^0 \to K^{*0}J/\psi(\to \mu^+\mu^-)$	$B^0 \to K^{*0}J/\psi(\to e^+e^-)$	$B^+ \to K^+J/\psi(\to \mu^+\mu^-)$	$B^+ \to K^+J/\psi(\to e^+e^-)$
2011-MD	(56.28±0.05) %	(44.73±0.03) %	100 %	100 %
2011-MU	(56.23±0.05) %	(44.65±0.03) %	100 %	100 %
2012-MD	(56.26±0.03) %	(45.85±0.02) %	100 %	100 %
2012-MU	(56.26±0.03) %	(45.85±0.02) %	100 %	100 %
2015-MD	(56.23±0.06) %	(44.29±0.03) %	100 %	100 %
2015-MU	(55.65±0.06) %	(43.76±0.03) %	100 %	100 %
2016-MD	(57.48±0.03) %	(45.22±0.01) %	100 %	100 %
2016-MU	(57.15±0.03) %	(44.78±0.01) %	100 %	100 %
2017-MD	100 %	100 %	100 %	100 %
2017-MU	100 %	100 %	100 %	100 %
2018-MD	100 %	100 %	100 %	100 %
2018-MU	100 %	100 %	100 %	100 %

Period	$B^0 \to K^{*0}\mu^+\mu^-$	$B^0 \to K^{*0}e^+e^-$	$B^+ \to K^+\mu^+\mu^-$	$B^+ \to K^+e^+e^-$
2011-MD	(56.01±0.06) %	(44.35±0.03) %	100 %	100 %
2011-MU	(55.99±0.06) %	(44.32±0.03) %	100 %	100 %
2012-MD	(55.98±0.04) %	(45.45±0.02) %	100 %	100 %
2012-MU	(56.01±0.05) %	(45.47±0.02) %	100 %	100 %
2015-MD	(55.93±0.08) %	(43.77±0.03) %	100 %	100 %
2015-MU	(55.34±0.08) %	(43.12±0.03) %	100 %	100 %
2016-MD	(57.12±0.04) %	(44.75±0.02) %	100 %	100 %
2016-MU	(56.75±0.03) %	(44.28±0.02) %	100 %	100 %
2017-MD	100 %	100 %	100 %	100 %
2017-MU	100 %	100 %	100 %	100 %
2018-MD	100 %	100 %	100 %	100 %
2018-MU	100 %	100 %	100 %	100 %

6.4.1 Calibration samples

Correcting the PID efficiency relies on centrally produced calibration collision data samples which were recorded during data taking periods. Low multiplicity decay modes with large branching fractions and final states composed of charged particles are preferred. In order to use the calibration samples to calculate the PID efficiency no PID requirements can be used in their selection. Simply avoiding PID requirements on the final selection of these samples is not enough, as trigger algorithms can also use PID information to select events. To avoid the trigger selections biasing the samples either the trigger algorithms do not use PID requirements or PID requirements are applied to a particle that is not used to measure the PID performance. A strategy used to employ the latter idea is known as tag and probe. Taking the example of muons for which the calibration sample usually involves a $J/\psi \to \mu\mu$ decay; tagged tracks are well-identified muons of a particular charge and probe tracks are tracks with the opposite charge selected without PID requirements. The tracks can then be combined into J/ψ candidates if the invariant mass is consistent with the J/ψ mass with a good quality vertex fit. The efficiencies must then be computed using only probe tracks. Table 6.3 shows the relevant calibration samples used for each

type of particle track. In RUN 2, for muons, there was the choice between taking prompt $J/\psi \to \mu^+\mu^-$ decays or those from B decays (in RUN 1 the prompt option is not available) but no large differences were found between the samples so the choice of the prompt sample was taken due to its larger size. In order to remove excess background in the calibration samples the ${}_s\mathcal{P}lot$ method is used centrally in all cases except for that of electrons. The electron calibration samples are much smaller than the others and with fewer statistics the ${}_s\mathcal{P}lot$ method is no longer reliable. This necessitates the need for the fit and count method described in Sec. 6.4.3.

Table 6.3: PID calibration samples

Particle	Description
K/π	$D^0 \to K^-\pi^+$ from $D^{*+} \to D^0\pi^+$
μ	$J/\psi \to \mu^+\mu^-$ from B decays (prompt J/ψ in RUN 2)
e	$J/\psi \to e^+e^-$ from B decays

6.4.2 Hadron and muon PID efficiencies

Efficiency maps are produced in dimensions of particle momentum, pseudorapidity and `nTracks`. The sample is split into isopopulated bins in `nTracks` and within each bin two KDEs are used. One KDE for the momentum versus pseudorapidity distribution for all candidates in the calibration sample is produced and another for the same distribution but only on candidates which pass the full PID selection are produced. The KDEs are determined using the `RooNDKeysPdfs` method, using two dimensional Gaussian distributions as kernels and then transformed into fine binned histograms, the ratio of which gives the final PID efficiency map. The final weight maps are checked with the usual LHCb approach of using PIDCalib, the two approaches are found to agree. The KDE approach is preferred as the PIDCalib procedure involves changing the binning by hand until a reasonable binning is attained, this is time consuming and can easily lead to biases. Efficiency maps from the KDE approach in 2016 for B^+ modes can be found in Appendix G.1, they are fairly representative of other years and B^0 modes. There is not always a large dependence on the multiplicity, as can be seen at the top of Fig. 6.1 for muon ID maps. However, it can be noted that for $\pi \to \mu$ misID maps, seen at the bottom of Fig. 6.1, and $K \to \mu$ misID maps there is some variation in the region centred around an η of 4 and a momentum of 100 MeV.

Using the efficiency maps per particle for a given year it is possible to extract the total PID efficiency, w_{PID}. The procedure to do this is as follows:

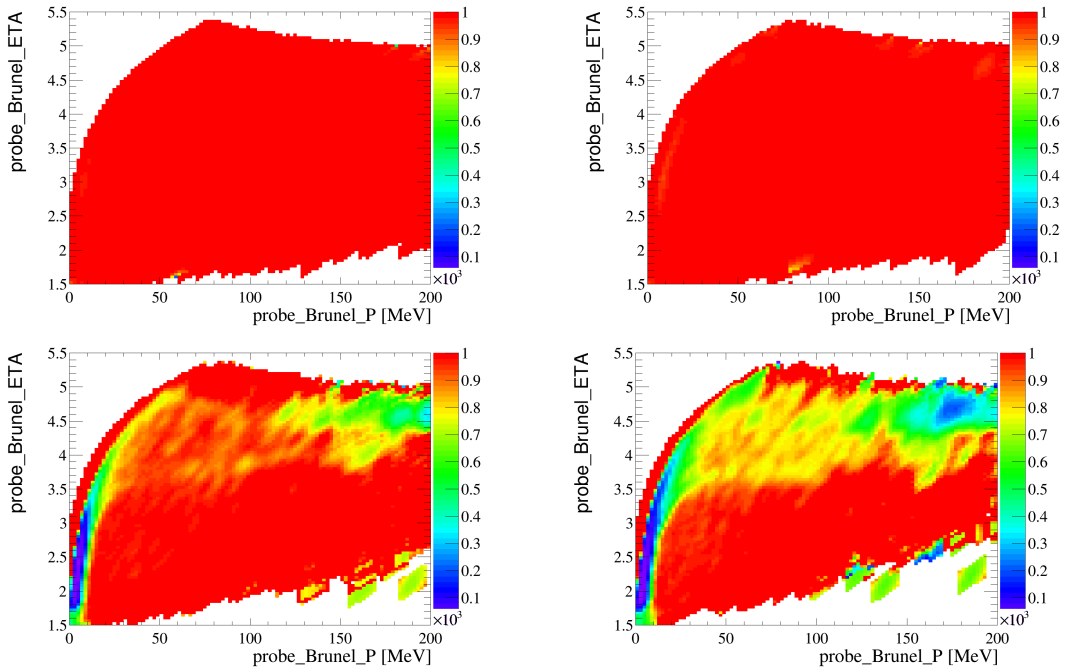


Figure 6.1: Example weight maps produced by the KDE approach. (top) μ ID maps, (bottom) $\pi \to \mu$ misID maps, (left) first bin in `nTracks`, (right) last bin in `nTracks`

- For each event, for each track that is relevant, the corresponding efficiency map is used to look up the efficiency value. First the `nTracks` value for that event is used to determine which bin of `nTracks` to look in. After this the momentum and η values are used to find the per event efficiency value. To reduce the dependence on the binning scheme for momentum and η a linear interpolation between bins is used when looking up the efficiency.

- If unphysical efficiency values are encountered, which can occur due to effects stemming from the ${}_s\mathcal{P}lot$ method or division by zero errors in bins with a very small number of events, the nearest physical value (0 or 1) is assigned.

- Once the efficiency for each relevant track has been extracted, they can be multiplied together to produce w_{PID} for that event. The process is then repeated for all events.

Weight maps are produced for particles which are correctly identified and those which are misidentified. The correctly identified (ID maps) encode the efficiency for candidates to pass the PID selection cuts defined in Sec. 4.7. The misidentification maps (misID maps) encode the efficiency for candidates to pass the PID selection for another particle. For example a kaon ID map is produced using the kaon calibration sample with the kaon PID selection applied. The kaon to muon misID map is

produced from the candidates in the kaon calibration sample which pass the muon PID selection. If the reconstructed particle ID matches the `TRUE_ID` of the particle then the identification weight map is used. If the reconstructed particle ID does not match the `TRUE_ID` of the particle then the corresponding misID map is used.

6.4.3 Electron PID efficiency

As mentioned in Sec. 6.4.1 calibration samples for electrons suffer from low statistics, which affects the ${}_s\mathcal{P}lot$ method used centrally to reduce residual background. A further complication results from the fact the mass shapes differ depending on the number of bremsstrahlung photons added, this would result in multiple ${}_s\mathcal{P}lot$ determinations on even smaller subsets of data. To avoid these issues a method known as fit and count is employed. In this approach the electron yields are fitted before and after the PID selection is applied in bins of momentum, pseudorapidity and `nTracks` in order to obtain a similar efficiency map as for hadrons and muons. The method still relies on a calibration data sample, in this case $J/\psi \to ee$ from B decays, but without the ${}_s\mathcal{W}eights$ applied. The PID prior cuts yield good identification of the kaon (from the B decay) and tag electron.

Two types of maps are produced for the electron PID, one with electrons with 0 bremsstrahlung photons added and one with electrons with > 1 bremsstrahlung photon added. The binning schemes are determined by isobinning in `nTracks` and within each bin computing separate 1D binning schemes in psuedorapidity and momentum. These 1D binning schemes are determined by first isobinning in the given dimension before merging adjacent bins which have a less than 5σ difference in efficiency, based on the statistical error. This is performed iteratively until no more merging is required. The two 1D binning schemes produced are then combined to make a 2D binning scheme within each `nTracks` bin. The binning schemes are determined before the split based on bremsstrahlung photons. When optimising the binning schemes yields are not used but a tight cut of $\pm 40\,\text{MeV}$ around the B meson mass is used instead.

A simultaneous fit in each bremsstrahlung category to candidates which either pass the PID requirements or fail to pass the PID requirements in a single bin is performed and the ratio of the electron yields determines the efficiency in that bin. The fits consist of a signal shape, combinatorial background and partially reconstructed background. This signal shape is modelled by a bifurcated crystal ball function, the mean and sigma are allowed to float in the fit, whilst the other parameters are fixed

from a fit to simulated $B^+ \to K^+ J/\psi (\to e^+e^-)$ (the decay the calibration sample is based on). The combinatorial background is modelled by an exponential function with its yield and slope allowed to float. The partially reconstructed background is described by an Argus function, its parameters are fixed to be the same as those from a fit to $B^0 \to K^{*0} J/\psi (\to e^+e^-)$ reconstructed as $B^+ \to K^+ J/\psi (\to e^+e^-)$. The fractions of the yields of the signal and partially reconstructed background are constrained to be equal in both the failed and passed sample fits. Figs. 6.2 and 6.3 show example simultaneous fits for the pass and fail categories for the no bremsstrahlung photons added and >1 bremsstrahlung photons added categories. The larger background for the no bremsstrahlung added case comes from charged pions misidentified as electrons, when bremsstrahlung photons are required this background is much reduced. The full set of electron efficiency maps can be found in Appendix G.2. These maps can then be used in the method described in Sec. 6.4.2 to determine w_{PID}.

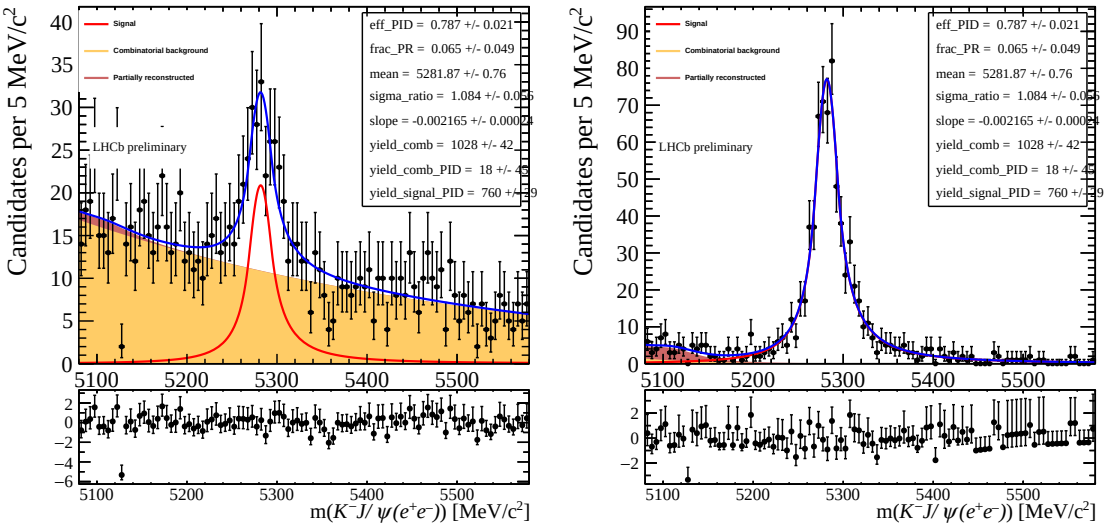


Figure 6.2: Example plots of simultaneous fits to calculate the electron fit and count PID efficiencies for 2018 with 0 bremsstrahlung photons added. With (left)/(right) dataset that failed/passed the PID requirement.

6.5 L0 trigger efficiency

The analysis uses two main L0 trigger categories, L0I and L0L, the choice of which was discussed in Sec. 4.4. Since any collision data that are not triggered are lost permanently simulated samples must be used to estimate this efficiency. The simulated efficiencies must describe the real efficiencies well and since the L0 trigger decisions are not well modelled in simulation this requires that correction factors

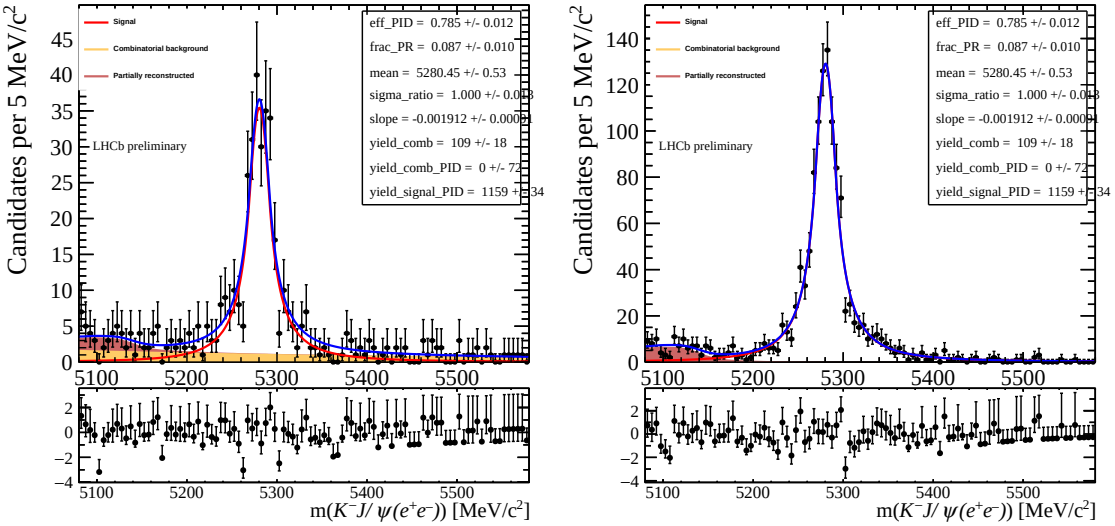


Figure 6.3: Example plots of simultaneous fits to calculate the electron fit and count PID efficiencies for 2018 with ≥ 1 bremsstrahlung photons added. With (left)/(right) dataset that failed/passed the PID requirement.

are applied to simulated samples.

To determine correction factors for the trigger efficiencies ratios of the efficiencies of collision data and simulation are used. The efficiencies are calculated using the TISTOS method, described in Sec. 2.2.6, using $B^+ \to K^+\ell^+\ell^-$ and $B^0 \to K^{*0}\ell^+\ell^-$ modes which are in agreement with each other. Therefore, the following sections will show only plots from $B^+ \to K^+\ell^+\ell^-$ modes.

6.5.1 Strategy

The L0 trigger weights, w^x_{L0}, where x denotes a trigger category, are obtained from the ratio of trigger efficiencies in data and simulation, ϵ^y_x, where x again denotes a trigger category and y denotes data or simulation (denoted MC here). The full selection, minus the MVA requirements, is applied to both simulated and collision data samples. To further suppress background a cut of $\pm 45\,\text{MeV}/c^2$ is applied around the J/ψ constrained B mass. Trigger efficiencies are only derived from the inclusive categories L0I or L0L; the TISTOS method will not hold in the L0L! case as it only contains TOS events so a tag and probe cannot be easily defined. Equation 6.2 shows how the efficiencies are determined on simulation and data, N_{TISTOS} denotes the number of candidates which are both TIS and TOS, N_{TIS} denotes the number of candidates triggered by the TIS category only and N_{TOS} denotes the number of candidates triggered by the TOS category only. In the following sections tag

and probe terminology will be adopted, as was introduced in Sec. 6.4.1, the trigger category in the denominator will be referred to as the tag whilst the trigger category only present in the numerator will be referred to as the probe. For example in the expression for ϵ_{L0L} the tag category is TIS and the probe category is TOS.

$$\epsilon_{\text{L0L}}^{\text{data,MC}} = \frac{N_{\text{TIS\&TOS}}}{N_{\text{TIS}}}^{\text{data,MC}}, \quad \epsilon_{\text{L0I}}^{\text{data,MC}} = \frac{N_{\text{TIS\&TOS}}}{N_{\text{TOS}}}^{\text{data,MC}} \tag{6.2}$$

By taking the ratios of these efficiencies the trigger weights in the L0L and L0I categories can be determined as seen in Eq. 6.3.

$$w_{L0}^{\text{L0I}} = \frac{\epsilon_{\text{L0I}}^{\text{data}}}{\epsilon_{\text{L0I}}^{\text{MC}}}, \quad w_{L0}^{\text{L0L}} = \frac{\epsilon_{\text{L0L}}^{\text{data}}}{\epsilon_{\text{L0L}}^{\text{MC}}} \tag{6.3}$$

Using a combination of these weights and efficiencies the weights for the L0L! category can be determined as seen in Eq. 6.4.

$$w_{L0}^{\text{L0L!}} = w_{L0}^{\text{L0L}} \cdot \frac{1-\epsilon_{\text{L0I}}^{\text{data}}}{1-\epsilon_{\text{L0I}}^{\text{MC}}} = \frac{\epsilon_{\text{L0L}}^{\text{data}}}{\epsilon_{\text{L0L}}^{\text{MC}}} \cdot \frac{1-\epsilon_{\text{L0I}}^{\text{data}}}{1-\epsilon_{\text{L0I}}^{\text{MC}}}. \tag{6.4}$$

The rest of the section will proceed as follows: first a short description of the L0 trigger corrections for the prior chain is presented in Sec. 6.5.2. The following sections will deal with the determination of the efficiencies for `L0Electron` (Sec. 6.5.4), `L0Muon` (Sec. 6.5.3) and L0I (Sec. 6.5.5) trigger categories. The choice of the tag in each case is described. The corrections are evaluated in each year separately as the trigger lines were not always configured in the same way per year. Finally, application of the weights is discussed in Sec. 6.5.6.

6.5.2 Prior chain

A set of L0 corrections is computed for the prior chain, described in Sec. 5.1, using the L0M trigger category. In this case only PID corrections are applied and the L0-TCKs are aligned so the muon p_{T} thresholds are identical in collision data and simulation. The resulting weight maps can be seen in Fig. 6.4 and comparing these to the maps obtained when the TCK thresholds are not aligned in Fig. 6.6, especially in 2016 and 2017, it can be seen that maps with aligned TCKs are much flatter.

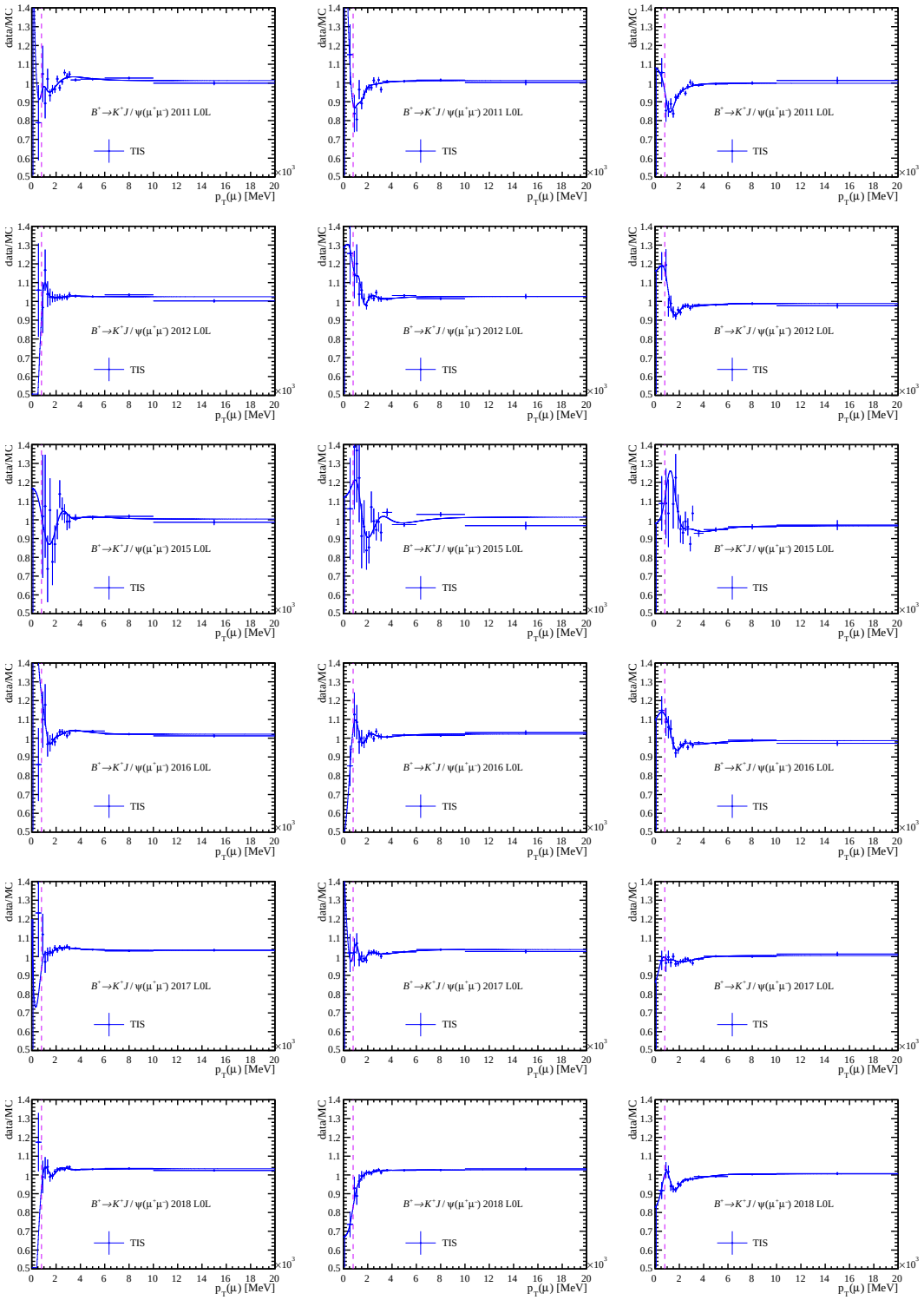


Figure 6.4: Prior chain data over simulation efficiency ratio of the `L0Muon` line as a function of the $p_{\mathrm{T}}(\mu)$ in the three $\eta(\mu)$ regions from left to right for $B^+ \to K^+ J/\psi\,(\to \mu^+\mu^-)$ events for all years from top to bottom. The three $\eta(\mu)$ regions from left to right are $0 < \eta < 2.75$, $2.75 < \eta < 3.25$ and $3.25 < \eta < 6$. The red vertical dashed line denotes the trigger threshold. The efficiencies are measured with the **TIS** tag.

6.5.3 L0Muon

Muons must have a p_{T} above a certain threshold for the L0Muon trigger line to be fired. In this case the efficiency is parametrised as a function of p_{T} in three regions of η. The TISTOS method is used with three different tag samples described in the following:

TIS tag: Events triggered at L0 level by a non-signal hadron or electron (B_L0Hadron_TIS or B_L0Electron_TIS lines).

Hadron tag: Events triggered at L0 level by a signal hadron (B_L0Hadron_TOS line).

Lepton tag: Events triggered at L0 level by the other muon in the $J/\psi \to \mu\mu$ decay (M(other)_L0Muon_TOS line).

A characteristic feature of the efficiency as a function of a threshold variable, muon p_{T} in this case, is that it forms a 'turn-on' curve. There is a plateau at or close to zero below the threshold p_{T} value, after this there is a strong 'turn-on' where the efficiency rapidly increases with p_{T} until a saturation efficiency is reached. The binning in p_{T} in each region in η reflects this, in particular a fine binning is employed in the turn on region to accurately model the sharp rise of the efficiency curve. In order to reduce the dependency of the corrections on the binning scheme the efficiency curve is fitted using a combination of a Gompertz function and an Error function as displayed in Eq. 6.5,

$$f_1 \cdot \left(1 + \mathrm{erf}\left(\frac{x-t}{\sqrt{2}\cdot\sigma_1}\right)\right) + f_2 \cdot s^{\exp\left(\frac{-(x-t)}{\sqrt{2}\cdot\sigma_2}\right)} + a. \tag{6.5}$$

In the equation $f_{1,2}$ denote fractions of the error function and Gompertz function, respectively, which are to be combined, t denotes the trigger threshold, $\sigma_{1,2}$ denotes the resolutions of the error function and Gompertz function respectively, s denotes the skewness of the Gompertz function and a represents a constant offset. The error function is well suited to model the turn-on curve since there are large plateau regions at zero efficiency for $\mathtt{real}_{E_{\mathrm{T}}}^{\mathtt{L0CaloTool}}$ values below the trigger threshold and at the saturation efficiency at large values of $\mathtt{real}_{E_{\mathrm{T}}}^{\mathtt{L0CaloTool}}$. It also possesses a symmetry around the the value of t and the steep increase is dominated by the resolution σ_1. Unity is added to the error function to ensure the efficiencies do not drop below zero values. The Gompertz function is required as the turn-on curves are not exactly symmetric, the entry to the right plateau is generally more gradual

than that of the left plateau, this asymmetry is dealt with by the skewness term s. The offset a is added as the low p_{T} plateau does not always end up at zero and may be slightly higher.

An example of the fis to `L0Muon` efficiencies for data and simulation in 2012 is shown in Fig. 6.5 with each of three tags shown. The trigger threshold of a given year is marked by a dotted line. In Fig. 6.6 the ratio of data over simulation efficiencies for all years can be seen. The agreement between the various choices of tag samples is very good across all years. Due to its high statistics the **TIS** tag is chosen as the default.

6.5.4 `L0Electron`

An electron must deposit an amount of transverse energy in the ECAL, labelled $\texttt{real}_{E_{\mathrm{T}}}^{\texttt{L0CaloTool}}$ in the analysis tuples, above a certain trigger threshold for the `L0Electron` trigger to be fired. The $\texttt{real}_{E_{\mathrm{T}}}^{\texttt{L0CaloTool}}$ variable is measured with a low resolution by the ECAL, which is expected to vary between the three ECAL regions since the granularity varies between them. Hence the efficiency correction is parametrised as a function of $\texttt{real}_{E_{\mathrm{T}}}^{\texttt{L0CaloTool}}$ and the ECAL region. The corrections are derived using two types of tag as follows:

- **TIS** tag: Events triggered at L0 level by a non-signal hadron or muon (`B_L0Hadron_TIS` or `B_L0Muon_TIS` lines).
- **Hadron** tag: Events triggered at L0 level by a signal hadron (`B_L0Hadron_TOS`)

The efficiency 'turn-on' curves are fitted with Eq. 6.5 as in the muon case and an example for the 2012 case be seen in Fig. 6.7. Ratios of data over simulation efficiencies in each calorimeter region can be seen in Fig. 6.8. In these figures the trigger threshold is denoted by a dashed vertical line, the reason that events can be seen below this threshold is due the ECAL resolution effects that are part of the $\texttt{real}_{E_{\mathrm{T}}}^{\texttt{L0CaloTool}}$ variable, plus the fact that in some years the TCK thresholds are not aligned. At large values above the threshold data and simulation agree well with each other, however, near and below the threshold significant deviations can be observed. The reasons for the deviation stem from poorly simulated ECAL response and also because when the data were collected a range of trigger thresholds were sometimes used. The agreement between the two choices of tag samples is good in all cases. As the **TIS** tag has higher statistics it is used as the default.

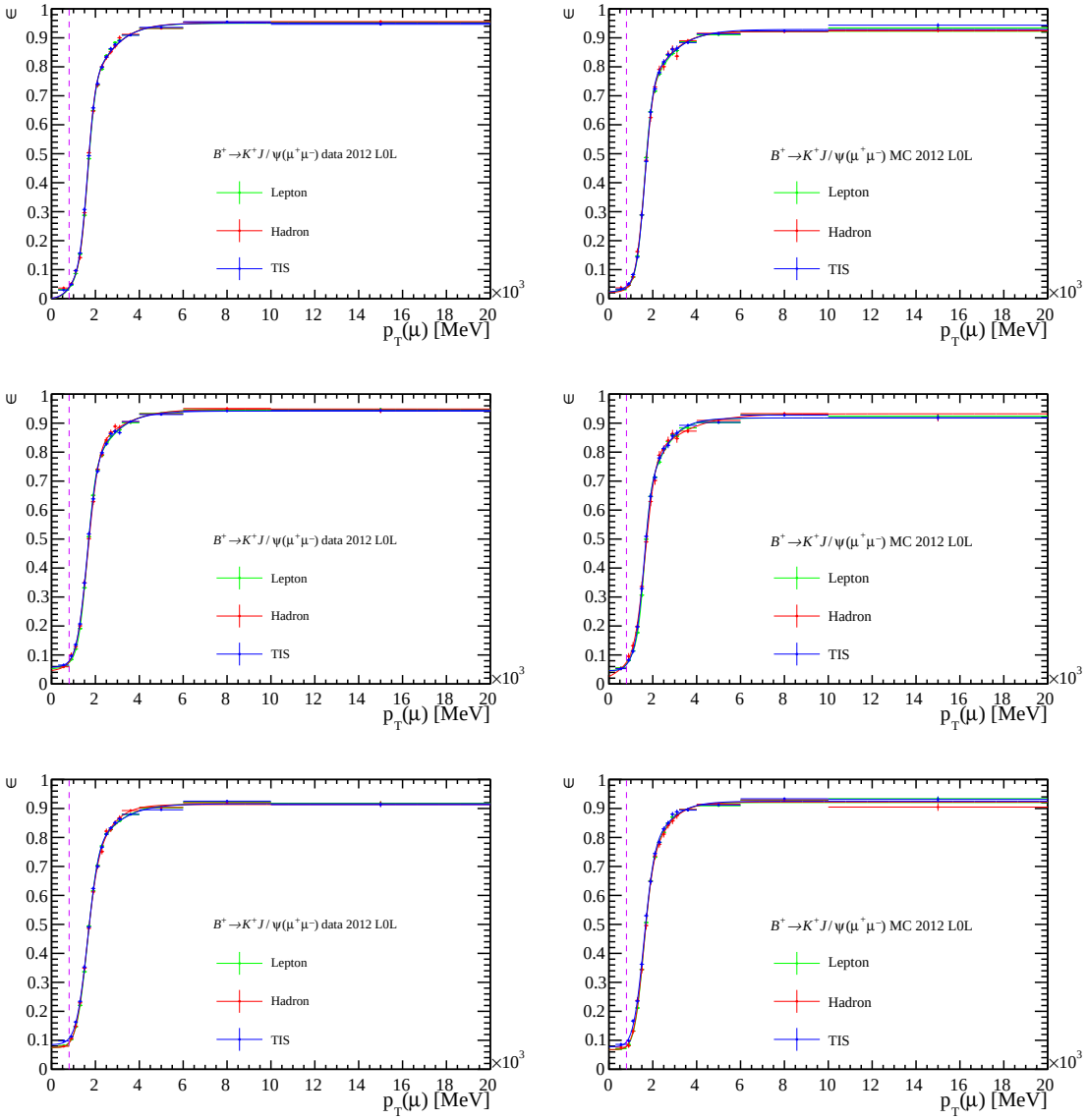


Figure 6.5: Efficiency of the `L0Muon` line as a function of the transverse momentum of the muon in three bins of $\eta(\mu)$ from top to bottom. The left (right) plots show 2012 $B^+ \to K^+ J/\psi(\to \mu^+\mu^-)$ data (simulation) events. The three $\eta(\mu)$ regions from top to bottom are $0 < \eta < 2.75$, $2.75 < \eta < 3.25$ and $3.25 < \eta < 6$.The red vertical dashed line denotes the trigger threshold. The efficiencies are measured with the **TIS** tag in blue, **Hadron** tag in red and **Lepton** tag in green.

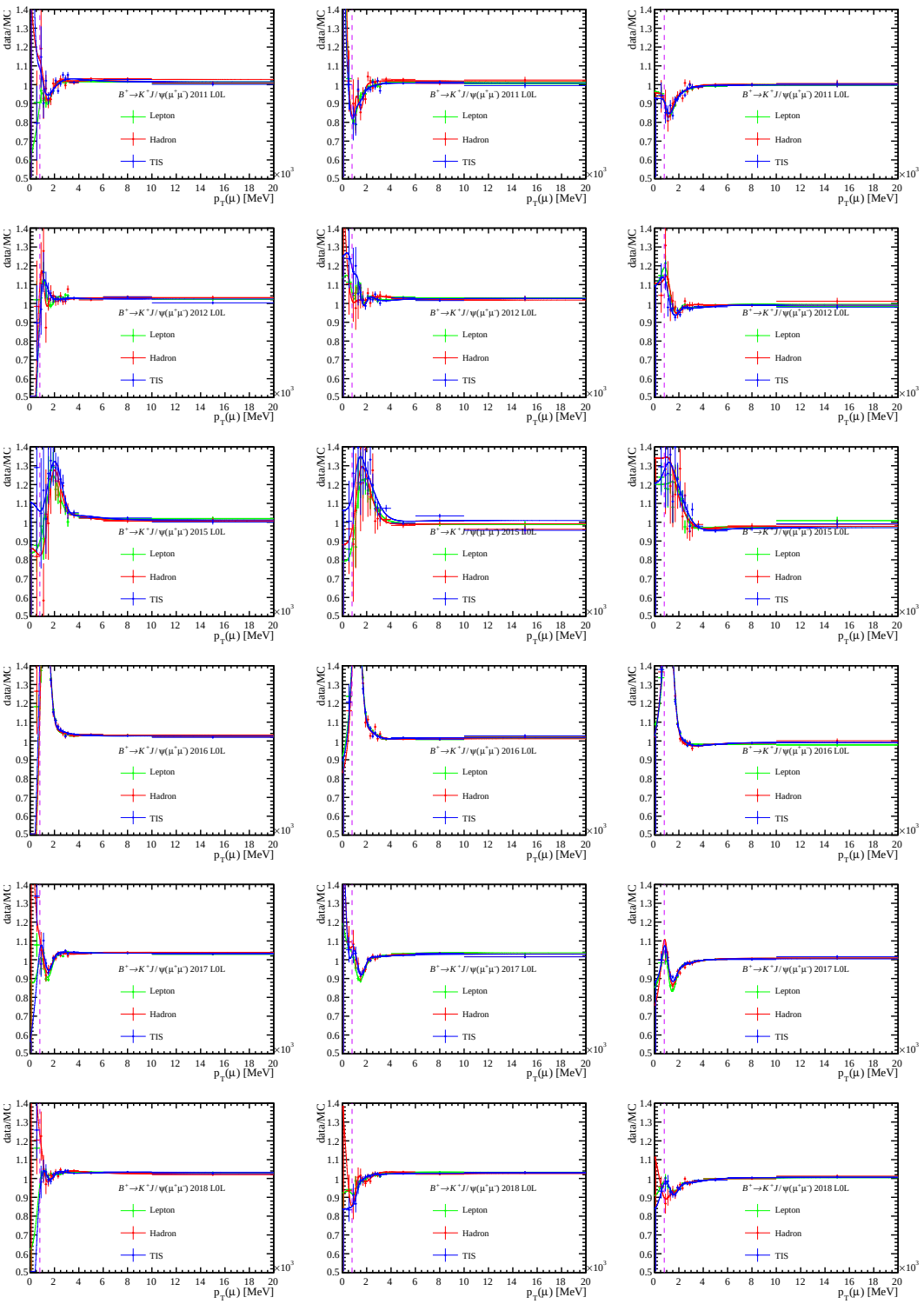


Figure 6.6: Data over simulation efficiency ratio of the `L0Muon` line as a function of the transverse momentum of the muon in three bins of $\eta(\mu)$ from left to right for $B^+ \to K^+ J/\psi (\to \mu^+\mu^-)$ events for all years from top to bottom. The three $\eta(\mu)$ regions from left to right are $0 < \eta < 2.75$, $2.75 < \eta < 3.25$ and $3.25 < \eta < 6$. The red vertical dashed line denotes the trigger threshold. The efficiencies are measured with the **TIS** tag in blue, **Hadron** tag in red and **Lepton** tag in green.

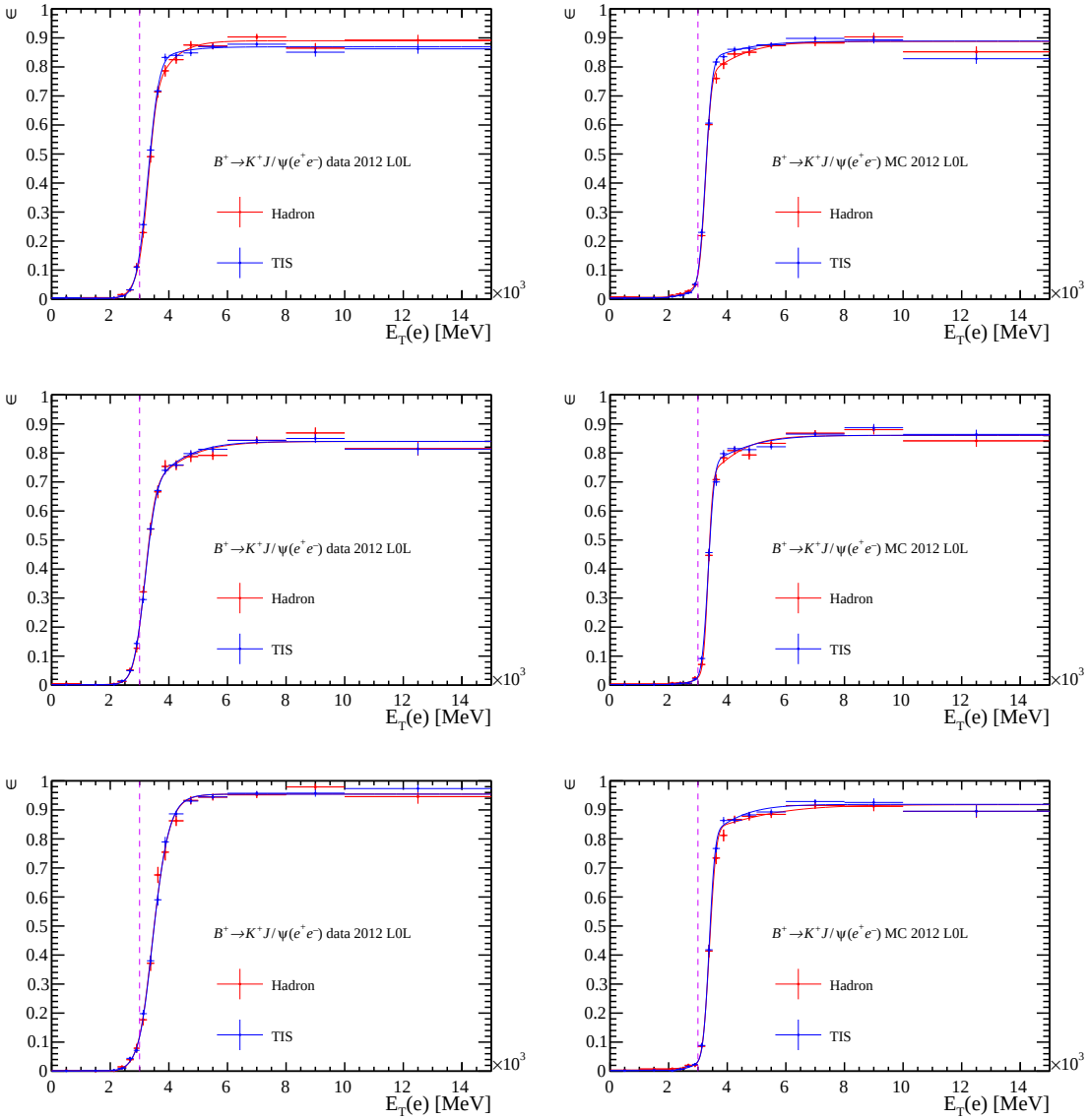


Figure 6.7: Efficiency of the `L0Electron` line as a function of the transverse energy of the electron in the three ECAL regions from top to bottom. The left (right) plots show 2012 $B^+ \to K^+ J/\psi (\to e^+e^-)$ data (simulation) events.The three $\eta(\mu)$ regions from top to bottom are $0 < \eta < 2.75$, $2.75 < \eta < 3.25$ and $3.25 < \eta < 6$. The red vertical dashed line denotes the trigger threshold. The efficiencies are measured with the **TIS** (**Hadron**) tag in blue (red).

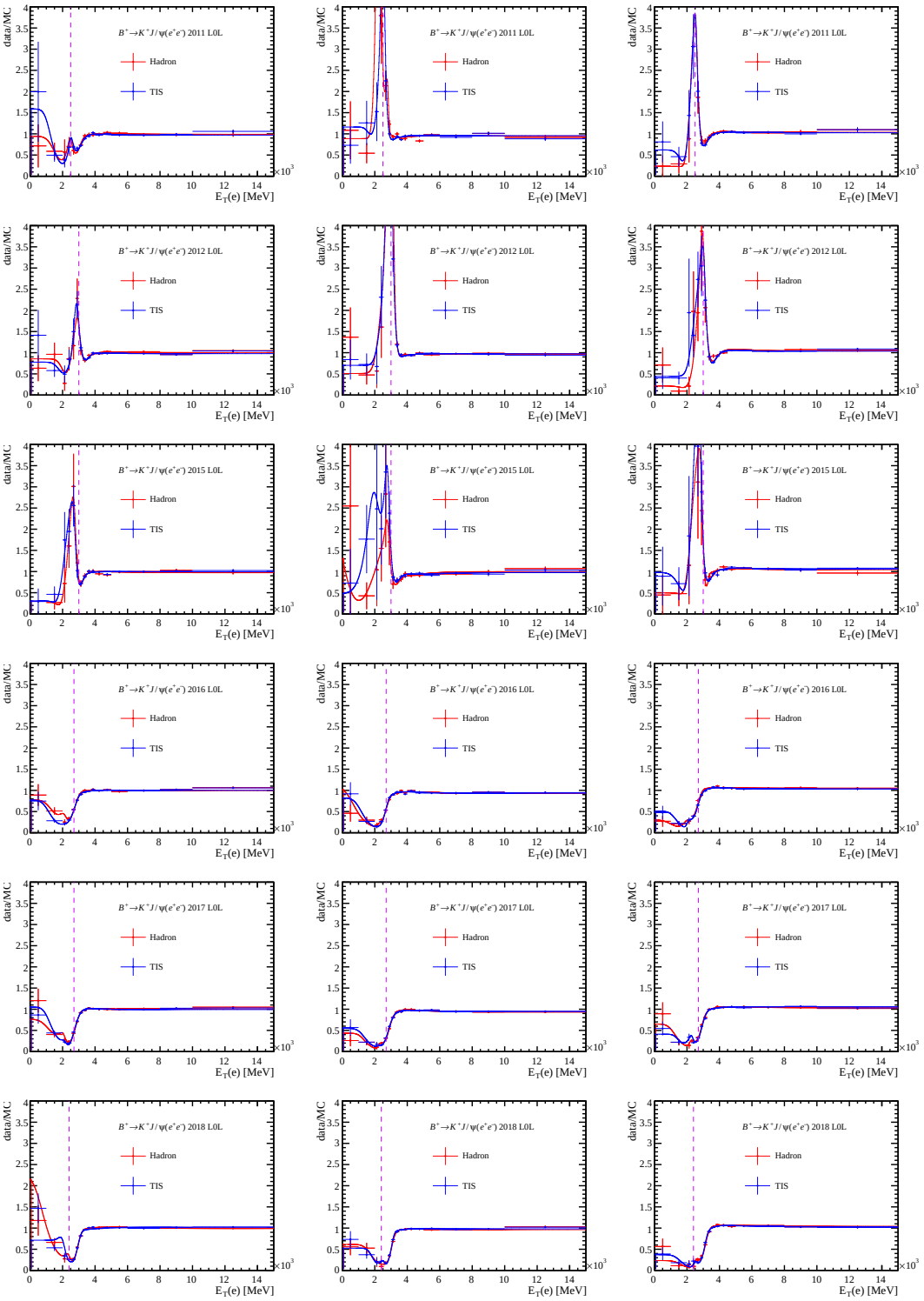


Figure 6.8: Data over simulation efficiency ratio of the `L0Electron` line as a function of the transverse momentum of the muon in three bins of $\eta(\mu)$ from left to right for $B^+ \to K^+ J/\psi\,(\to e^+e^-)$ events for all years from top to bottom. The three $\eta(\mu)$ regions from left to right are $0 < \eta < 2.75$, $2.75 < \eta < 3.25$ and $3.25 < \eta < 6$. The red vertical dashed line denotes the trigger threshold. The efficiencies are measured with the **TIS** (**Hadron**) tag in blue (red).

6.5.5 L0I

The L0I trigger category uses the `L0Global_TIS` trigger line. The trigger is mostly fired by the B hadron which results from the hadronisation of the opposite b quark. Therefore its efficiency is chosen to be parametrised in the p_{T} of the signal B meson, which is expected to be correlated to the p_{T} of the other B meson. The p_{T} is parametrised in six regions of `nTracks` because the non-signal trigger is more likely to fire on events with higher multiplicity. In 2015 only four regions of `nTracks` are used due to the low statistics in that year. The L0I efficiency is expected to be independent of the signal final state (it is by definition triggered on the rest of the event) so the same correction can be extracted from both muon and electron final states. It was verified, as shown in Fig. 6.9, that corrections from L0I using just electron or muon final states are compatible. As the muon sample has larger statistics and is generally a cleaner sample it is used to extract all the corrections in the following and these are then used for the electron mode as well. The TISTOS method is performed using three types of tag:

- **Lepton tag** : Events triggering the L0 trigger on one of the signal leptons
- **Hadron tag** : Events triggering the L0 trigger on a signal hadron
- **Combined tag** : Events triggering the L0 trigger by a signal lepton or signal hadron.

The comparison plots for the various tags, with example fits can be seen in Fig. 6.10 for 2012, 2016 and 2018. Due to the small samples size of the **Hadron** tag, these are not differential in bins of `nTracks`. All the tags appear to be in good agreement, the **Combined** tag is used as it is the sample with the highest statistics. In these figures a different function is used to fit the efficiency turn-on curves, a single Gompertz function is used as shown in Eq. 6.6,

$$f \cdot s^{\frac{-(x-t)}{\sqrt{2}\cdot\sigma}} + a. \tag{6.6}$$

In the equation f is the amplitude of the function, σ is the resolution, t the trigger threshold, s the skewness and a a constant offset. In order to stabilise the fits the skewness parameter, s is fixed to 0.5. Fig. 6.11 shows the ratio of data and simulation efficiencies for 2012, 2016 and 2018. The ratios performed differential in `nTracks` can be seen in Appendix H. The differential ratio plots all have a rather flat area at high p_{T} around unity, with an increasing trend at low p_{T} feeding into

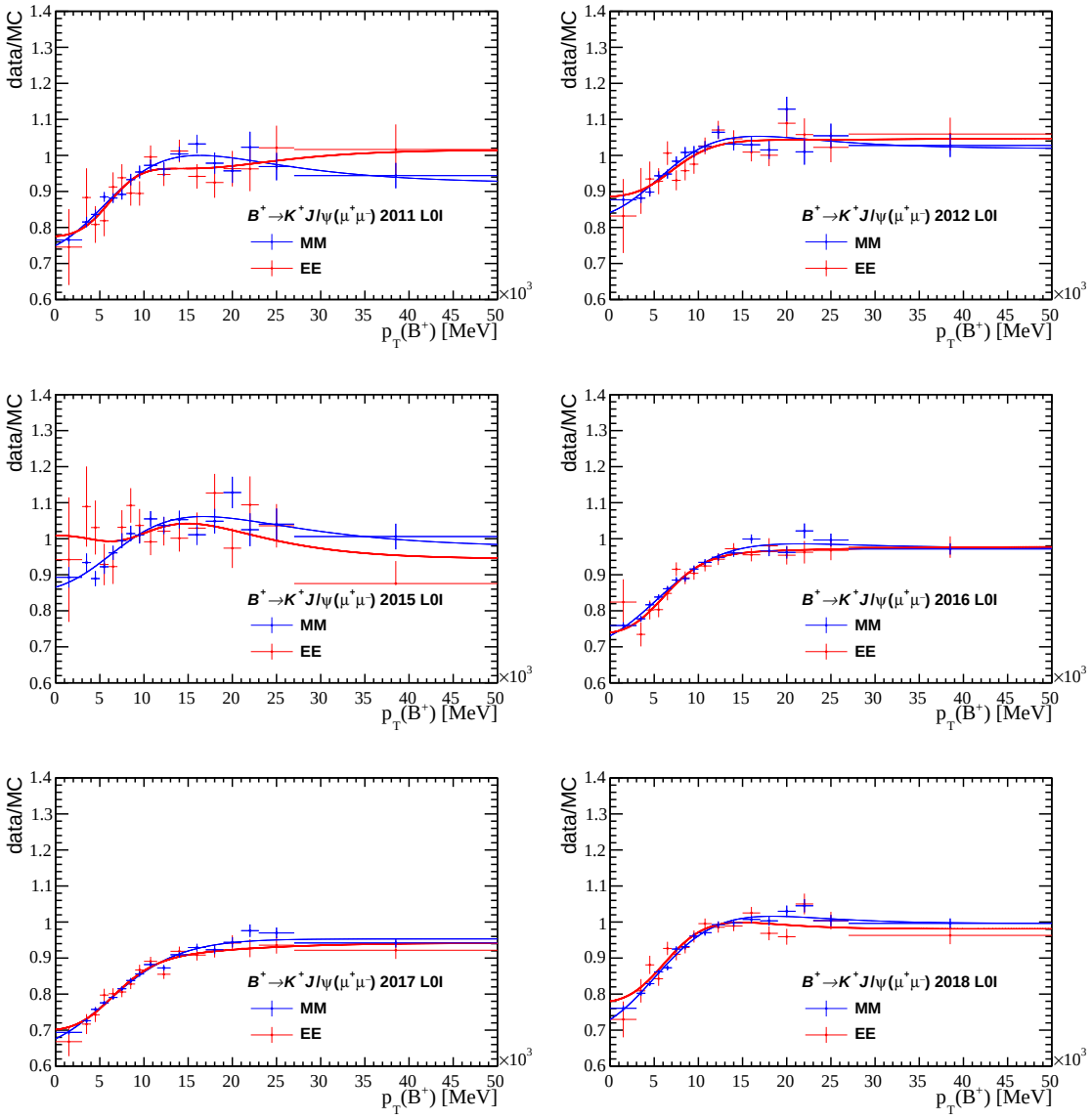


Figure 6.9: Data over simulation efficiency ratio of the `L0Global_TIS` line as a function of the transverse momentum of the B meson for $B^+ \to K^+ J/\psi (\to \mu^+\mu^-)$ ($B^+ \to K^+ J/\psi (\to e^+e^-)$) events in blue (red) for all years. The efficiencies are measured with the **Combined** tag.

the plateau. As the multiplicity increases this increasing trend begins to start at a higher value of the ratio resulting in flatter ratio plots at high multiplicity.

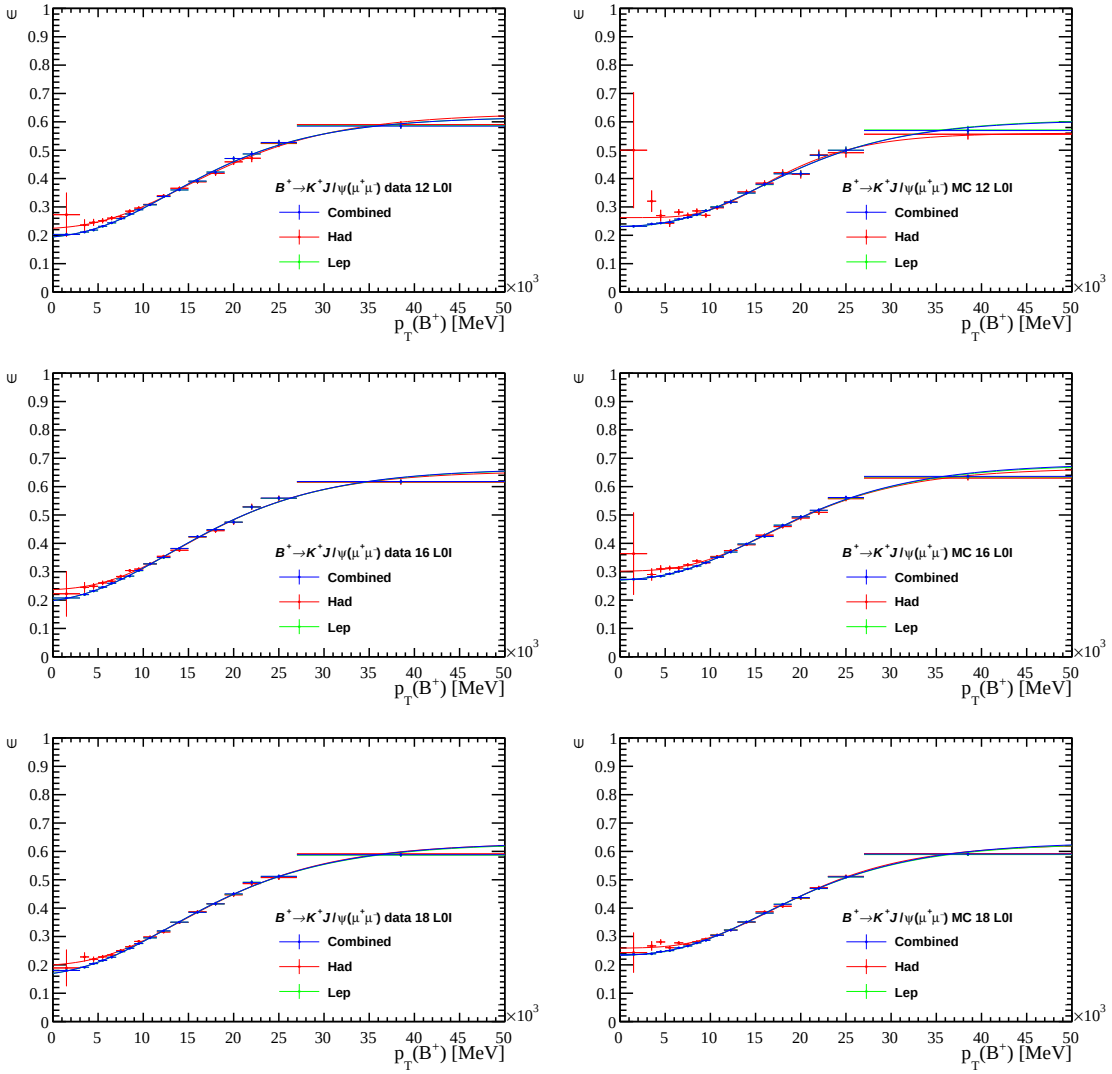


Figure 6.10: Efficiency of the `L0Global_TIS` line as a function of the transverse momentum of the B meson. The left (right) plots show $B^+ \to K^+ J/\psi (\to \mu^+\mu^-)$ data (simulation) events for 2012, 2016 and 2018 from top to bottom. The efficiencies are measured with the **Combined** tag in blue, **Hadron** tag in red and **Lepton** tag in green.

6.5.6 Application of corrections

With the corrections determined, some care must be taken when applying the corrections to obtain an accurate L0 efficiency particularly in the L0L and L0L! categories.

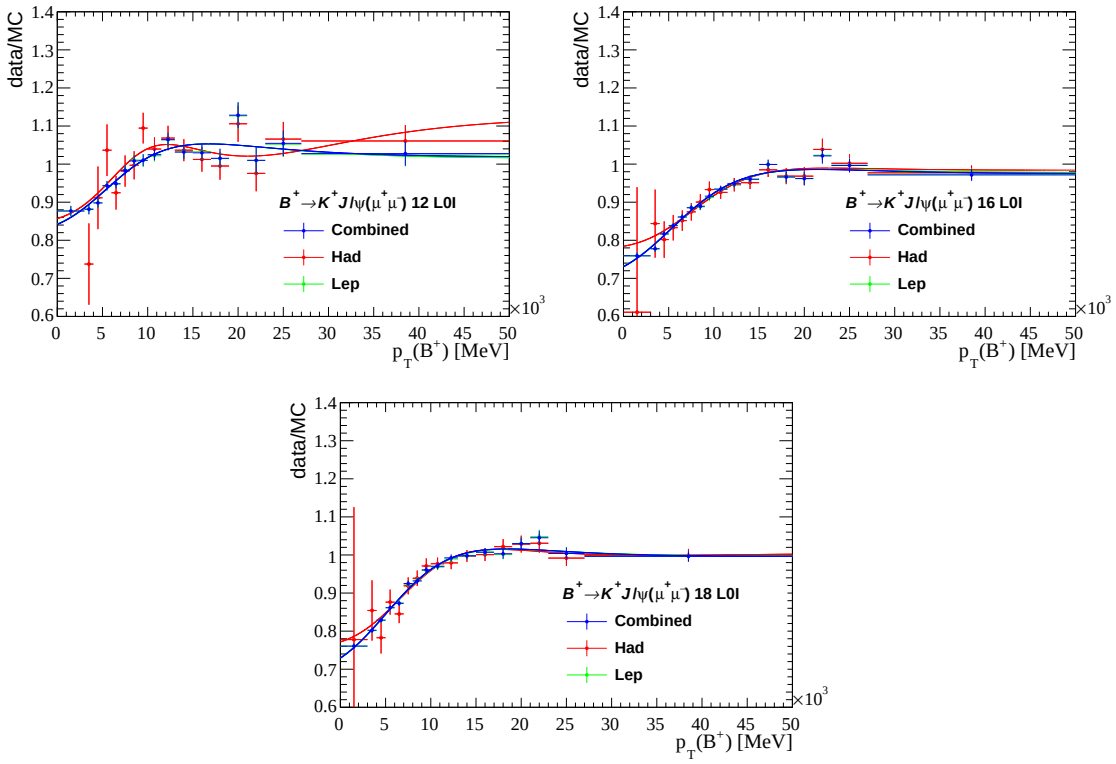


Figure 6.11: Data over simulation efficiency ratio of the `L0Global_TIS` line as a function of the transverse momentum of the B meson for $B^+ \to K^+ J/\psi (\to \mu^+\mu^-)$ events for (top-bottom) 2012, 2016, and 2018. The efficiencies are measured with the **Combined** tag in blue, **Hadron** tag in red and **Lepton** tag in green.

The L0L trigger category is defined by the logical `OR` condition for one of the two leptons to be TOS with respect to this trigger. Since it is possible for both leptons to be TOS with respect to the trigger decision the probability to trigger them is not mutually exclusive so $P(A||B) = P(A) \cdot P(B) - P(A) \cdot P(B)$ holds. Using this the efficiency terms in Eq. 6.3 can be written as:

$$\begin{aligned}\epsilon_{\mathrm{L0L}}^{\mathrm{data,MC}} &= \epsilon_{\mathrm{L0L}}^{\mathrm{data,MC}}(\ell^+) \cdot \ell^+{}_{\mathrm{TOS}} + \epsilon_{\mathrm{L0L}}^{\mathrm{data,MC}}(\ell^-) \cdot \ell^-{}_{\mathrm{TOS}} \\ &\quad - \epsilon_{\mathrm{L0L}}^{\mathrm{data,MC}}(\ell^+) \cdot \ell^+{}_{\mathrm{TOS}} \cdot \epsilon_{\mathrm{L0L}}^{\mathrm{data,MC}}(\ell^-) \cdot \ell^-{}_{\mathrm{TOS}} \\ &= \left(1 - \left(1 - \epsilon_{\mathrm{L0L}}^{\mathrm{data,MC}}(\ell^+) \cdot \ell^+{}_{\mathrm{TOS}}\right) \cdot \left(1 - \epsilon_{\mathrm{L0L}}^{\mathrm{data,MC}}(\ell^-) \cdot \ell^-{}_{\mathrm{TOS}}\right)\right) , \end{aligned} \tag{6.7}$$

where the $\ell^{+,-}_{TOS}$ denotes the trigger decisions for the two leptons which are either `E{1,2}_L0Electron_TOS` or `Mu{1,2}_L0Electron_TOS`. Therefore the expression for the L0L trigger weights can be rewritten:

$$\omega^{\mathrm{L0L}} = \frac{\epsilon_{\mathrm{L0L}}^{\mathrm{data}}}{\epsilon_{\mathrm{L0L}}^{\mathrm{MC}}} = \frac{\left(1 - \left(1 - \epsilon_{\mathrm{L0L}}^{\mathrm{data}}(\ell^+) \cdot \ell^+{}_{\mathrm{TOS}}\right) \cdot \left(1 - \epsilon_{\mathrm{L0L}}^{\mathrm{data}}(\ell^-) \cdot \ell^-{}_{\mathrm{TOS}}\right)\right)}{(1 - (1 - \epsilon_{\mathrm{L0L}}^{\mathrm{MC}}(\ell^+) \cdot \ell^+{}_{\mathrm{TOS}}) \cdot (1 - \epsilon_{\mathrm{L0L}}^{\mathrm{MC}}(\ell^-) \cdot \ell^-{}_{\mathrm{TOS}}))} . \tag{6.8}$$

The weight expression for the L0I can be directly computed as in Eq. 6.3 as:

$$\omega^{\mathrm{L0I}} = \frac{\epsilon_{\mathrm{L0I}}^{\mathrm{data}}}{\epsilon_{\mathrm{L0I}}^{\mathrm{MC}}} . \tag{6.9}$$

Combining Eqs. 6.8 and 6.9 the weight expression for the L0L! category can be built:

$$\omega^{\mathrm{L0L!}} = \frac{\left(1 - \left(1 - \epsilon_{\mathrm{L0L}}^{\mathrm{data}}(\ell^+) \cdot \ell^+{}_{\mathrm{TOS}}\right) \cdot \left(1 - \epsilon_{\mathrm{L0L}}^{\mathrm{data}}(\ell^-) \cdot \ell^-{}_{\mathrm{TOS}}\right)\right)}{(1 - (1 - \epsilon_{\mathrm{L0L}}^{\mathrm{MC}}(\ell^+) \cdot \ell^+{}_{\mathrm{TOS}}) \cdot (1 - \epsilon_{\mathrm{L0L}}^{\mathrm{MC}}(\ell^-) \cdot \ell^-{}_{\mathrm{TOS}}))} \cdot \frac{1 - \epsilon_{\mathrm{L0I}}^{\mathrm{data}}}{1 - \epsilon_{\mathrm{L0I}}^{\mathrm{MC}}} . \tag{6.10}$$

Using the equations in this section the various efficiencies which have been calculated can be used to create L0 correction weights.

6.6 HLT efficiency

As outlined in Tab. 4.2 candidates are required to pass the selection configured in certain HLT lines. Similarly to the case for L0 corrections simulated samples are required to evaluate the trigger efficiency since in collision data only those candidates which are TOS with respect to a given line are retained. The HLT decisions are more complex than for the L0 case however, since some reconstruction occurs in both HLT1 and HLT2 steps, the various lines have multiple selections encoded within

Table 6.4: Tag categories definition used for the HLT correction for electron and muon mode in data (HLTPHYS) and in simulated events (HLTOR). The HLT1OR and HLT2OR categories are defined as the logical OR of their listed single lines.

Label	Run 1	15	16 & 17 &18
HLTPHYS	HLT1_PHYS_TIS & HLT2_PHYS_TIS		
HLT1OR	HLT1TrackAllL0_TIS HLT1Track{Muon,DiMuon}_TIS	HLT1Track{MVA,Muon}_TIS HLT1TwoTrack_TIS	HLT1Track{MVA,Muon,MuonMVA}_TIS HLT1TwoTrack_TIS
HLT2OR	HLT2Topo{2,3,4}BodyBBDT_TIS HLT2TopoMu{2,3,4}BodyBBDT_TIS HLT2TopoE{2,3,4}BodyBBDT_TIS HLT2DiMuonDetached_TIS	HLT2Topo{2,3,4}Body_TIS HLT2TopoMu2,3,4Body_TIS Hlt2DiMuonDetachedHeavy_TIS	HLT2Topo{2,3,4}Body_TIS HLT2TopoMu{2,3,4}Body_TIS HLT2TopoMuMu{2,3,4}Body_TIS Hlt2DiMuonDetachedHeavy_TIS HLT2TopoE{2,3,4}Body_TIS HLT2TopoEE{2,3,4}Body_TIS
HLTOR	HLT1OR & HLT2OR		

them and there is some interplay between HLT1 and HLT2. Despite this the same TISTOS technique can also be used to evaluate efficiency corrections here. As for the L0 case a tag and probe approach is again used with the probe being TIS&TOS and the tag being TIS. The final HLT correction is computed in Eq. 6.11, ϵ_x denotes the efficiency, where x can be data or MC for collision data or simulation respectively and N_y denotes the number of events where y can be TIS&TOS or TIS.

$$w_{\text{HLT}} = \frac{\epsilon_{data}}{\epsilon_{MC}} = \frac{(\frac{N_{TIS\&TOS}}{N_{TIS}})_{data}}{(\frac{N_{TIS\&TOS}}{N_{TIS}})_{MC}} \tag{6.11}$$

The TOS decision for the HLT trigger is defined as the logical AND of the HLT1 and HLT2 trigger line decisions in Tab. 4.2, whilst the TIS decision is taken as the logical OR of several lines which are fired independently of the signal. A list of the lines used to define the TIS decision can be seen in Tab. 6.4.

Note that the definition of the tag category used for collision data and simulation is different. In collision data the logical AND of the HLT1PHYS and HLT2PHYS lines are selected, whilst for simulation the logical AND of HLT1OR and HLT2OR is taken. The HLTOR type lines are the logical OR of the various lines listed in Tab. 6.4. This choice is made since HLTPHYS can be very different between collision data and simulation. The main reasons being that:

- The HLTPHYS lines are subject to prescaling of 0 in data and 1 in simulation. A prescale acts so that if there are more events than the trigger can handle only a fraction of them are retained. So in this case the data is not prescaled at all whilst the simulation is.

- The `HLTPHYS` line criteria change with TCKs so that in certain years there will be a misalignment between data and simulation.

6.6.1 Strategy

The HLT correction makes use of $B^0 \to K^{*0} J/\psi (\to \ell^+\ell^-)$ and $B^+ \to K^+ J/\psi (\to \ell^+\ell^-)$ decays separated into L0I and L0L! L0 trigger categories. The full selection is applied to these samples, minus the MVA selection, with correction steps applied which differ if the corrections are for the prior chain or the nominal chain. The ϵ_{MC} is extracted using the ratio between probe and tag events on simulation. The ϵ_{data} is extracted in a similar way but fits are performed to the tag and probe categories to determine the yields of each. For B^+ modes for both muons and electrons the signal is modelled with a double sided crystal ball (DSCB) function, the $K \to \pi$ misID is also modelled by a DSCB, a partially reconstructed background component is added using a `RooKeysPdf` from the inclusive resonant samples ($B^0 \to X J/\psi (\to \ell^+\ell^-)$, $B^+ \to X J/\psi (\to \ell^+\ell^-)$ and $B_s^0 \to X J/\psi (\to \ell^+\ell^-)$) and the combinatorial background is modelled with an exponential function. All the parameters in these models are floating where this is allowed. For B^0 modes the situation is similar, for muon modes the fit model is almost identical to the B^+ case except that instead of a misID peak a DSCB is used to model the B_s^0 background. For electron modes the fit model is the same as for the muon modes except that there is no partially reconstructed background component. In both these cases where possible parameters are allowed to float.

Two proxy variables have been used to obtain the HLT corrections, the transverse momentum of the B meson and `nTracks`. The binning schemes are optimised separately for electrons and muons ensuring there are enough events in the TIS&TOS category to allow fits to be performed. When computing efficiencies using transverse momentum a trend can be seen in collision data which is reproducible in the simulation. However, when the same procedure is performed using `nTracks` a trend in 2011 and 2012 is present in data which is not reproduced in simulation and this is also reproduced using `nSPDHits` as a cross-check. This suggests that the HLT selection is dependant on the multiplicity in collision data in Run 1 and therefore the HLT corrections are parametrised in bins of `nTracks`. Corrections are then interpolated between bins of `nTracks` to reduce the dependence on the binning scheme.

The rest of this section will discuss the HLT corrections from the prior chain (Sec. 6.6.2) and after that the HLT corrections for the nominal chain for L0M

(Sec. 6.6.3), L0E (Sec. 6.6.4) and L0I (Sec. 6.6.5) trigger categories.

6.6.2 Prior Chain

As discussed in Sec. 5.1 a HLT correction is applied to the prior chain. In this case the data and simulation samples are from the L0M trigger category and are subject to the full selection, minus MVA requirements, with PID and L0 corrections applied to the simulated sample. The TCKs are also aligned in the prior chain and due to this the 2012 and 2016 HLT alignments, discussed in Sec. 4.5, are dropped since the alignment of the TCKs already accounts for this. The HLT efficiencies in collision data and simulation and the resulting correction weight maps are displayed in Fig. 6.12.

6.6.3 L0M

For the nominal chain collision data and simulation are fully selected, minus the MVA selection, and simulated samples are subject to PID, tracking, L0 and kinematic and multiplicity corrections. For the L0M category (and L0E also) it was verified that ϵ_{data} is the same for L0M and L0M! and the decision was made to take corrections from L0M and use them for L0M!. The inclusive category has the benefit of larger statistics which helps with overall fit stability when extracting ϵ_{data}. The HLT efficiencies in collision data and simulation and the resulting correction weight maps are displayed in Fig. 6.13.

6.6.4 L0E

For the L0E trigger category the situation is the same as for the L0M category, the L0E category is used to derive the correction and this is applied to the L0E! category. The HLT efficiencies in collision data and simulation and the resulting correction weight maps are displayed in Fig. 6.14.

6.6.5 L0I

For the L0I trigger category the selection and corrections applied is the same for for the L0M and L0E cases. For the L0I case for the L0 trigger corrections were taken from the muon mode in the L0I category and applied to both muons and electrons. However, for the HLT corrections the electron and muon efficiencies in the L0I category do not always agree, therefore, separate corrections are taken from muon and electron modes. The HLT efficiencies in collision data and simulation and

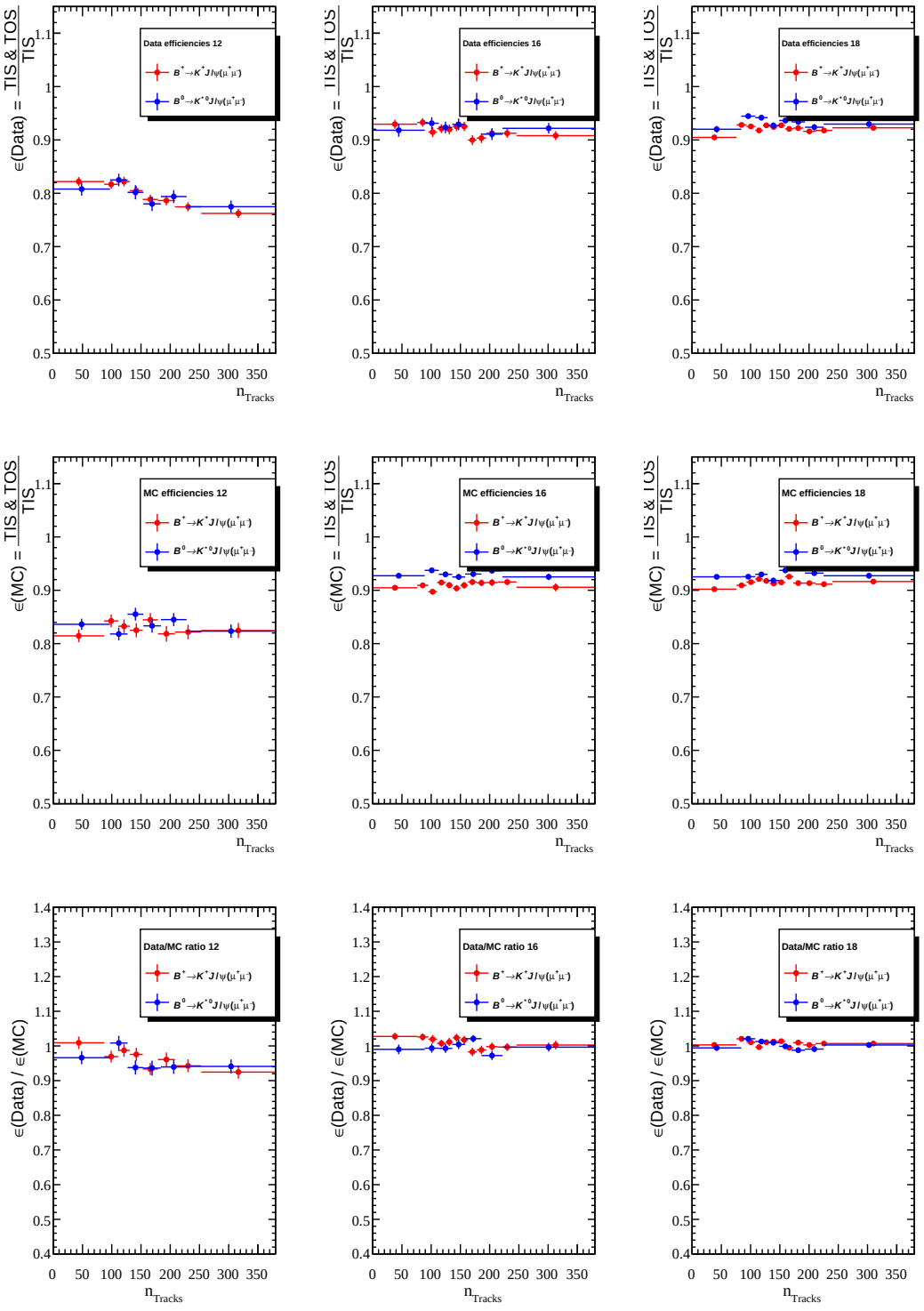


Figure 6.12: (top) HLT efficiencies measured in data, (centre) HLT efficiencies measured in simulation, (bottom) ratio of efficiencies between data and simulation as a function of the event occupancy n_{Tracks}. From left to right 2012, 2016 and 2018 results are displayed. In red(blue) the efficiencies and ratios measured using the $B^+ \to K^+ J/\psi (\to \mu^+\mu^-)$ ($B^0 \to K^{*0} J/\psi (\to \mu^+\mu^-)$) modes selected in the `L0Muon` inclusive category. Data are selected in the TCKs where the `L0MuonDecision` thresholds are aligned to simulation.

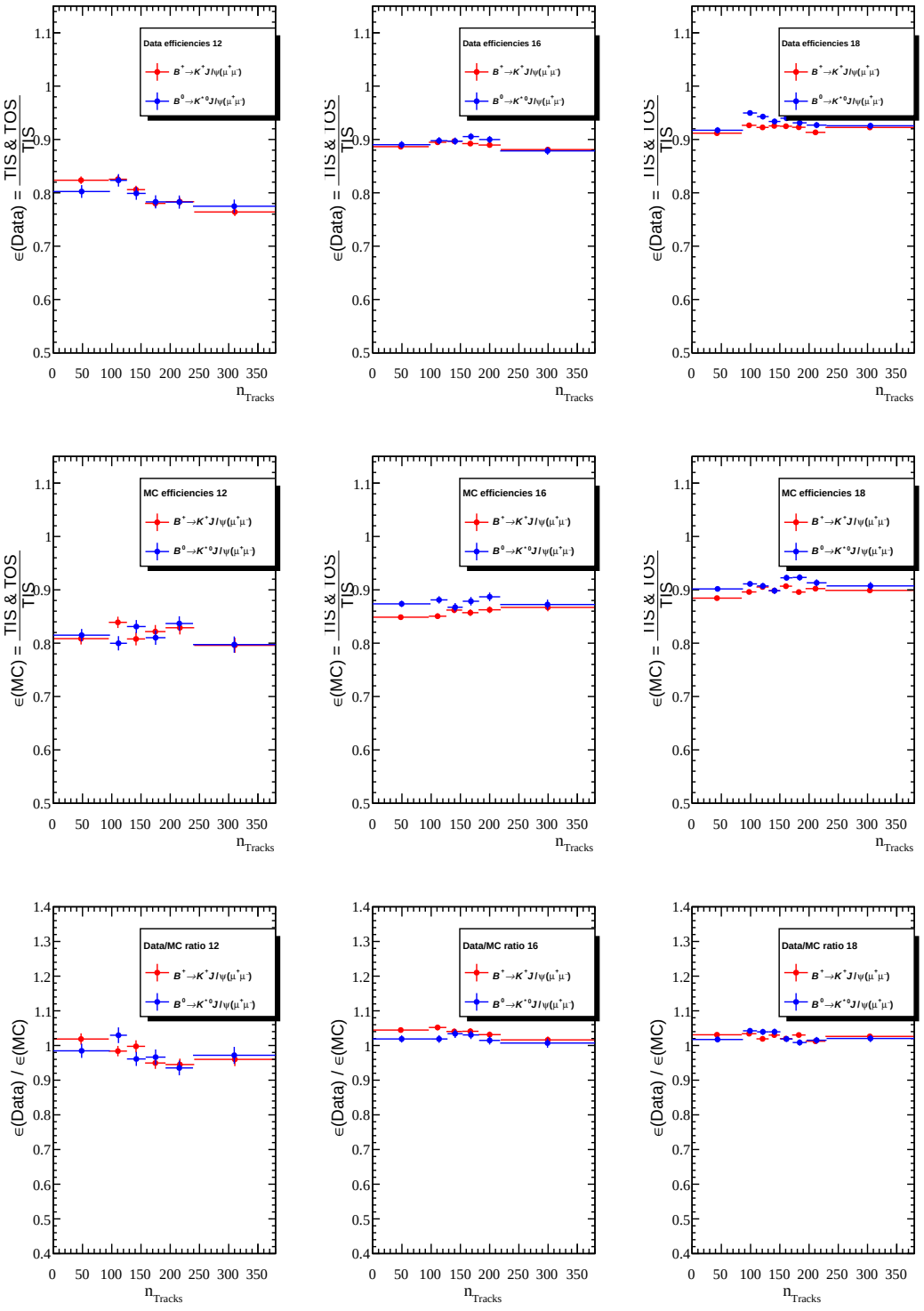


Figure 6.13: (top) HLT efficiencies measured in data, (center) HLT efficiencies measured in simulation, (bottom) ratio of efficiencies between data and simulation as a function of the event occupancy `nTracks`. From top to bottom 2012, 2016 and 2018 results are shown. In red(blue) the efficiencies and ratios measured using the $B^+ \to K^+ J/\psi (\to \mu^+\mu^-)$ ($B^0 \to K^{*0} J/\psi (\to \mu^+\mu^-)$) modes selected in the `L0Muon` inclusive category.

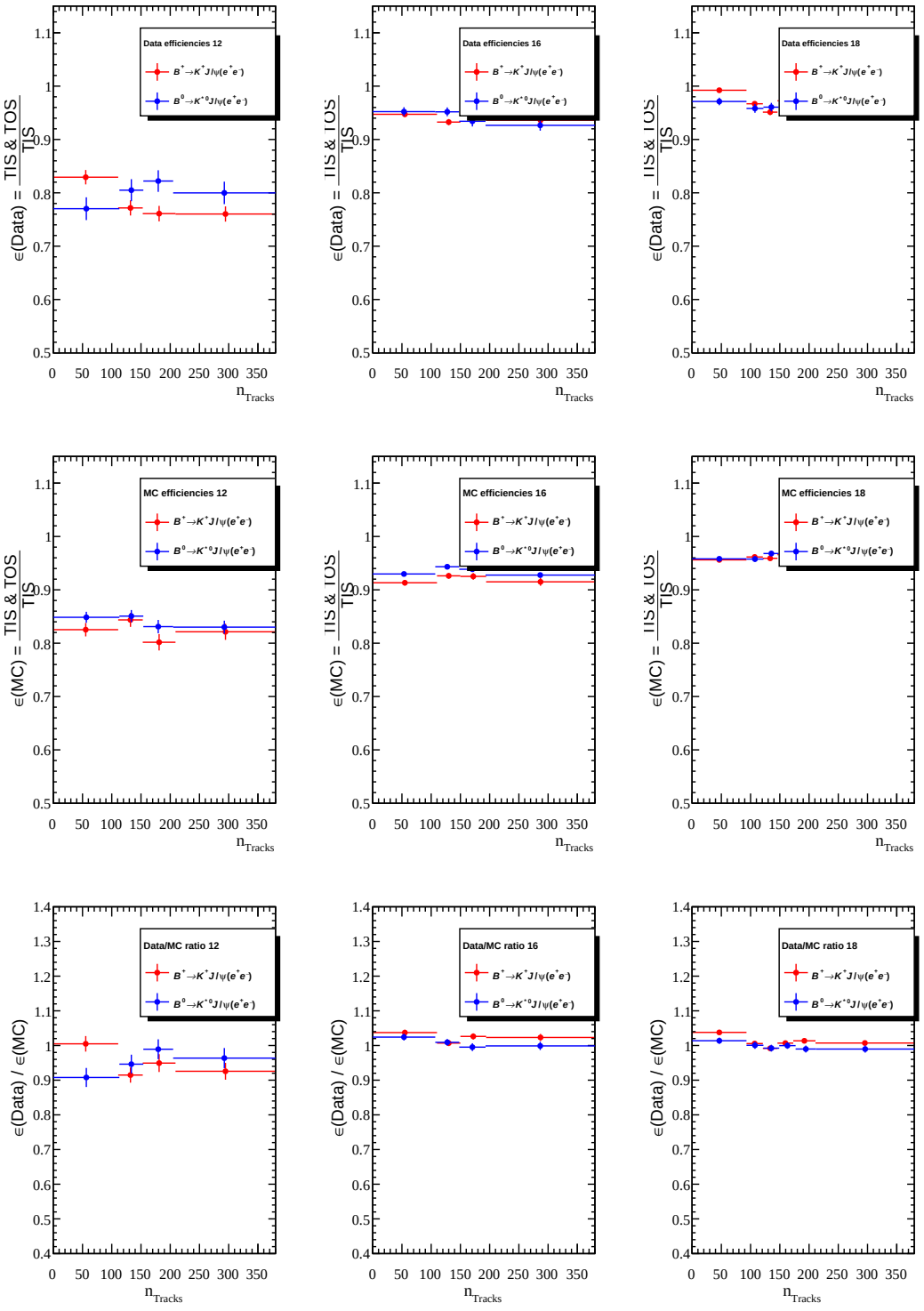


Figure 6.14: (top) HLT efficiencies measured in data, (centre) HLT efficiencies measured in simulation, (bottom) ratio of efficiencies between data and simulation as a function of the event occupancy `nTracks`. From top to bottom 2012, 2016 and 2018 results are shown. In red(blue) the efficiencies and ratios measured using the $B^+ \to K^+ J/\psi(\to e^+e^-)$ ($B^0 \to K^{*0} J/\psi(\to e^+e^-)$) modes selected in the `L0Electron` inclusive category.

the resulting correction weight maps are displayed in Fig. 6.15 for muon modes and Fig. 6.16 for electron modes.

6.7 q^2 Smearing

Simulated samples are used when computing the efficiency of cuts selecting part of the q^2 spectrum, such as the low or central-q^2 regions, however, the resolution of $m(B)$ and $m(ee)$ is generally better in simulated samples when compared to collision data. Due to this mismatch the efficiency of events to lie in a certain q^2 window may be overestimated if evaluated on simulated samples. In order to evaluate the size of this effect a smearing can be applied to simulated samples based on inputs from fits to the $m(ee)$ spectrum in both simulated samples and collision data. Additionally studying this effect can improve constraints in fits where leakage is expected to play a role.

The differences in resolution stem from two main sources. The first source is that the simulated material budget of the detector and the actual material budget of the detector are not exactly aligned. Due to this the effect of bremsstrahlung is different in the two cases which affects the momentum resolution. The second source is due to imperfect simulation of occupancy and calorimeter responses which affect the efficacy of the bremsstrahlung recovery procedure which can lead to changes in the B mass shape.

Strategy

In order to perform the smearing the following formula will be used:

$$m^{smeared} = m^{true} + s_\sigma \cdot (m^{reco} - m^{true}) + \Delta\mu + (1 - s_\sigma) \cdot (\mu^{MC} - m(J/\psi)_{PDG}) \quad (6.12)$$

where m^{true} is the true (generated) dilepton mass, m^{reco} is the reconstructed mass in simulation, s_σ is the sigma scale, $\Delta\mu$ is the mass shift, μ^{MC} is value of the mass determined from a fit to simulated samples and $m(J/\psi)_{PDG}$ is the PDG value of the J/ψ mass.

In order to extract s_σ, $\Delta\mu$ and μ^{MC} fits to the $m(ee)$ spectrum in simulated samples and collision data must be performed. Fits are performed for B^+ and B^0 modes in the 3 bremsstrahlung categories in each of the two primary trigger categories for each year separately. The full selection is applied except for the q^2 and MVA selection criteria and only vetoes for peaking and semileptonic backgrounds are retained. An

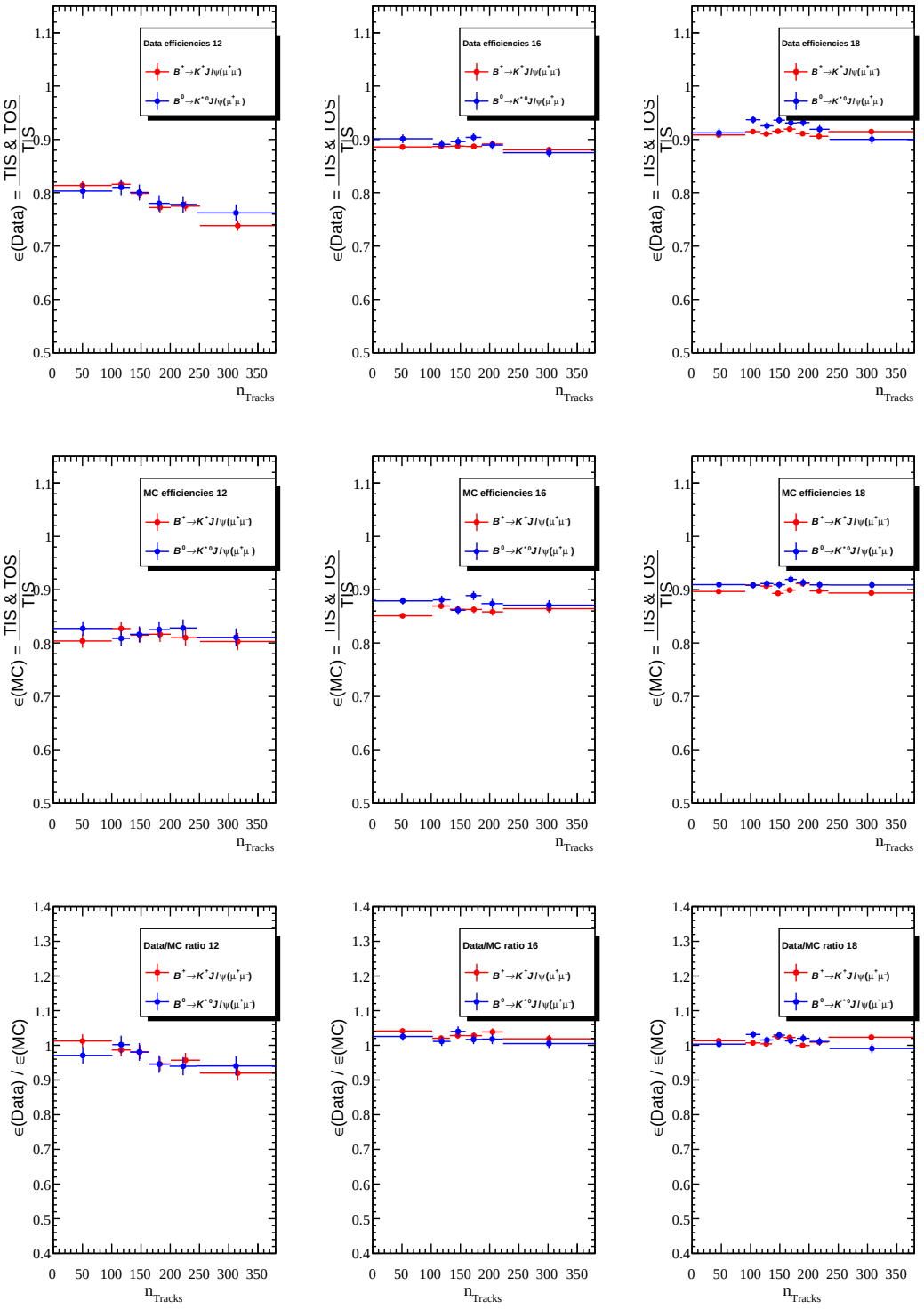


Figure 6.15: (top) HLT efficiencies measured in data, (centre) HLT efficiencies measured in simulation, (bottom) ratio of efficiencies between data and simulation as a function of the event occupancy `nTracks`. From top to bottom 2012, 2016 and 2018 results are shown. In red(blue) the efficiencies and ratios measured using the $B^+ \to K^+ J/\psi (\to \mu^+\mu^-)$ ($B^0 \to K^{*0} J/\psi (\to \mu^+\mu^-)$) modes selected in the L0I category for the μ mode.

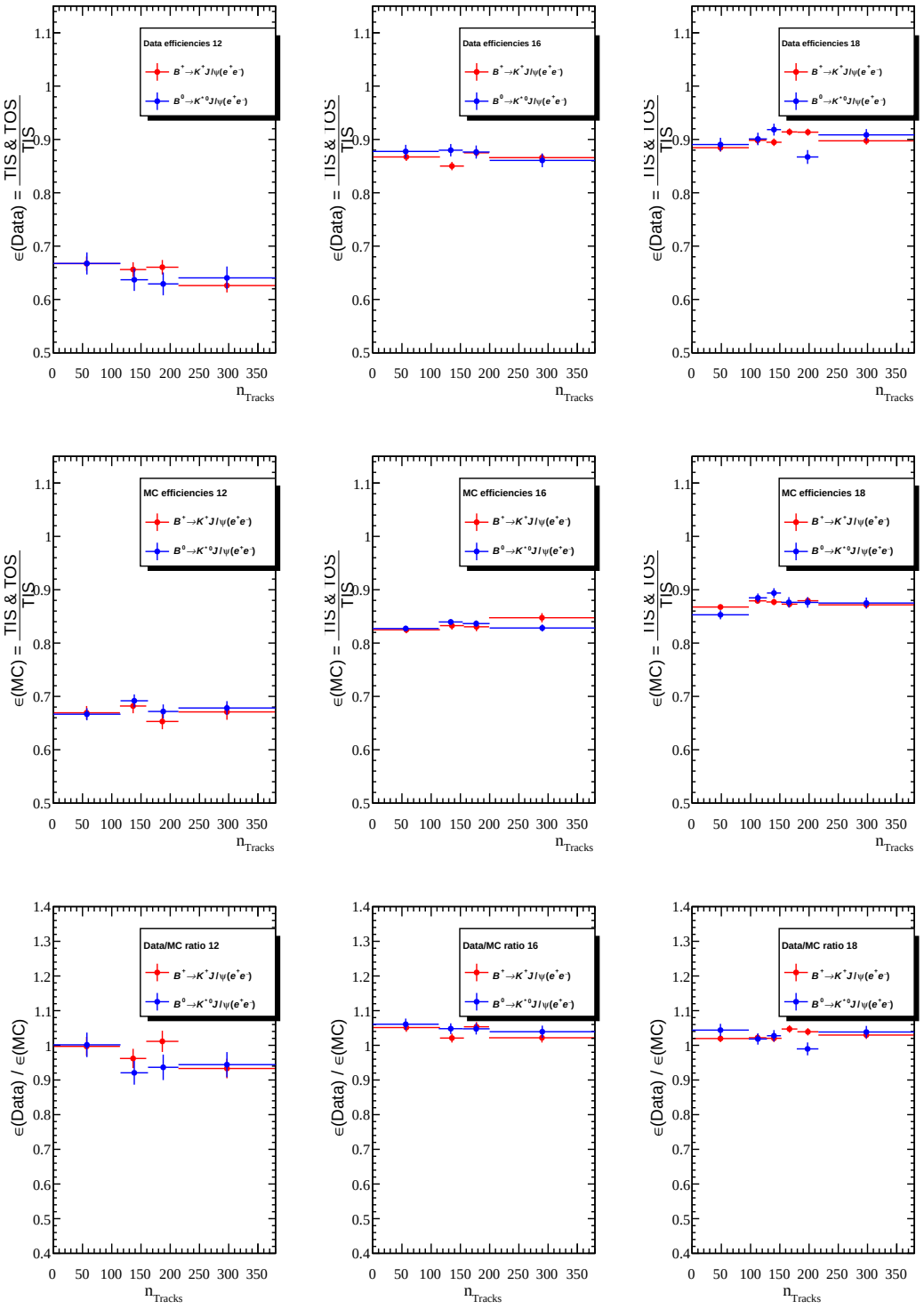


Figure 6.16: (top) HLT efficiencies measured in data, (centre) HLT efficiencies measured in simulation, (bottom) ratio of efficiencies between data and simulation as a function of the event occupancy `nTracks`. From top to bottom 2012, 2016 and 2018 results are shown. In red(blue) the efficiencies and ratios measured using the $B^+ \to K^+ J/\psi(\to e^+e^-)$ ($B^0 \to K^{*0} J/\psi(\to e^+e^-)$) modes selected in the L0I category for the e mode.

additional cut of $5.20 < m(B) < 5.68$ is applied to both B^+ and B^0 modes to suppress partially reconstructed backgrounds.

A double sided crystal ball is used to model the $m(ee)$ spectrum in both simulated samples and collision data and an exponential function is used to model remaining combinatorial background. Fits to the simulated samples share the same shape parameters. Fits to the simulated samples are performed with BKGCAT of 0 or 10 or 50 or 60 with the additional requirements that $n-1$ of the final state particles have the correct TRUEID and at most 1 final state particle has a TRUEID = 0. For the fits to data for the 0 bremsstrahlung photons category all the parameters from the corresponding fit to simulation are fixed except for the mean and sigma which are allowed to float. For the fits to data for 1 bremsstrahlung photon and greater than 2 bremsstrahlung photons all the parameters from the corresponding MC fits are fixed except for the mean, sigma and high mass tail parameter which are allowed to float. The fits to both simulated samples and collision data can be found in Appendix I. Using the fits the values of s_σ, $\Delta\mu$ and μ^{MC} can be extracted and these values are then averaged over both trigger categories in a particular year. The s_σ and $\Delta\mu$ values are shown in Tab. 6.5, whilst the μ^{MC} values are shown in Tab. 6.6. Generally the values obtained for B^+ or B^0 modes agree within their errors, with 2015 and 2016 values of s_σ and $\Delta\mu$ lying furthest away from unity and zero respectively. The values of μ^{MC} are very similar in the 0 γ and 1 γ cases with a slight increase in the > 2 γ case, as can be seen in Appendix I the right tail is most prominent in the > 2 γ case which likely accounts for this.

Table 6.5: Values of the sigma scale, s_σ, and mass shift, $\Delta\mu$, from fits to data and simulation of the relevant $m(ee)$ spectrum. The values have been averaged over both trigger categories for each year.

	(0 γ) s_σ	(0 γ) $\Delta\mu$	(1 γ) s_σ	(1 γ) $\Delta\mu$	(> 2 γ) s_σ	(> 2 γ) $\Delta\mu$
R_K-11	1.15 ± 0.03	-4.90 ± 0.89	1.16 ± 0.02	-1.80 ± 0.92	1.15 ± 0.03	2.54 ± 1.45
R_{K^*}-11	1.14 ± 0.04	-2.75 ± 1.44	1.21 ± 0.04	-1.07 ± 1.54	1.14 ± 0.05	2.65 ± 2.36
R_K-12	1.13 ± 0.02	-2.18 ± 0.59	1.15 ± 0.01	0.06 ± 0.64	1.14 ± 0.02	5.78 ± 1.05
R_{K^*}-12	1.11 ± 0.03	-2.04 ± 0.93	1.16 ± 0.02	-0.39 ± 1.08	1.15 ± 0.04	7.24 ± 1.81
R_K-15	1.26 ± 0.03	-7.99 ± 1.15	1.18 ± 0.03	-9.51 ± 1.24	1.17 ± 0.03	-10.92 ± 1.60
R_{K^*}-15	1.20 ± 0.05	-6.60 ± 1.95	1.18 ± 0.04	-9.39 ± 2.01	1.14 ± 0.07	-13.62 ± 3.33
R_K-16	1.20 ± 0.01	-6.45 ± 0.45	1.19 ± 0.01	-11.54 ± 0.48	1.18 ± 0.02	-13.70 ± 0.79
R_{K^*}-16	1.17 ± 0.02	-5.97 ± 0.75	1.21 ± 0.02	-12.68 ± 0.84	1.19 ± 0.03	-12.14 ± 1.36
R_K-17	1.16 ± 0.01	-6.37 ± 0.44	1.16 ± 0.01	-8.66 ± 0.46	1.13 ± 0.01	-9.24 ± 0.74
R_{K^*}-17	1.15 ± 0.02	-5.36 ± 0.74	1.13 ± 0.02	-7.13 ± 0.78	1.12 ± 0.03	-7.64 ± 1.27
R_K-18	1.16 ± 0.01	-5.83 ± 0.41	1.16 ± 0.01	-7.22 ± 0.42	1.12 ± 0.01	-6.16 ± 0.69
R_{K^*}-18	1.13 ± 0.02	-4.97 ± 0.65	1.17 ± 0.02	-5.90 ± 0.73	1.14 ± 0.02	-5.94 ± 1.16

Table 6.6: Values of the mass value from fits to simulation μ^{MC} from fits to data and simulation of the relevant $m(ee)$ spectrum. The values have been averaged over both trigger categories for each year.

	(0 γ) μ^{MC}	(1 γ)μ^{MC}	(>2 γ) μ^{MC}
R_K-11	3073.95 ± 1.58	3066.65 ± 1.46	3078.95 ± 1.54
R_{K^*}-11	3071.82 ± 1.49	3069.71 ± 1.36	3078.62 ± 1.25
R_K-12	3073.45 ± 1.52	3070.01 ± 1.22	3084.06 ± 1.12
R_{K^*}-12	3074.32 ± 1.19	3069.33 ± 1.11	3080.89 ± 1.11
R_K-15	3074.29 ± 1.43	3069.83 ± 1.34	3083.30 ± 1.37
R_{K^*}-15	3074.88 ± 1.40	3068.84 ± 1.33	3081.10 ± 1.55
R_K-16	3075.59 ± 0.85	3070.01 ± 0.81	3078.84 ± 0.81
R_{K^*}-16	3073.38 ± 0.69	3069.26 ± 0.59	3079.71 ± 0.61
R_K-17	3073.66 ± 0.60	3069.14 ± 0.55	3080.10 ± 0.56
R_{K^*}-17	3073.37 ± 1.04	3069.28 ± 0.91	3079.48 ± 0.91
R_K-18	3073.67 ± 0.64	3070.66 ± 0.58	3080.77 ± 0.58
R_{K^*}-18	3075.67 ± 1.09	3070.32 ± 0.95	3079.15 ± 0.89

To test the effect of the smearing the data is refit using smeared MC with the smearing based on the values presented in Tabs. 6.5 and 6.6. Smearing based on the values from both B^+ and B^0 modes are used to cross-check that smearing corrections from either mode are compatible. If the smearing has been successful the s_σ should be compatible with unity and the $\Delta\mu$ should be compatible with zero. Tables of these values are shown in Tabs. 6.7 and 6.8 for B^+ modes fitted using simulated samples employing a smearing from B^+ and B^0 modes respectively, similar tables are obtained for smearings applied to the B^0 modes. In all cases s_σ is in very good agreement with unity for all bremsstrahlung categories for smearings derived from both B^+ and B^0 modes. The best agreement for $\Delta\mu$ is generally found in the 0 γ bremsstrahlung category, when the smearing is taken from the same mode as is being fit to the agreement is better than when taken from a different mode, however, using corrections from either B^+ or B^0 modes results in a reasonable agreement. The agreement in the 1 γ and > 2 γ categories is worse when compared to the 0 γ category but is much improved with respect to the values in Tab. 6.5. Considering all the above the smearing from both B^+ and B^0 modes can be said to considerably improve the modelling of the resolution in simulation.

Table 6.7: Values of the sigma scale ,s_σ and mass shift, $\Delta\mu$, from fits to B^+ data based on smearing derived from B^+ samples

	(0γ) s_σ	(1γ) s_σ	(2γ) s_σ	(0γ) $\Delta\mu$	(1γ) $\Delta\mu$	(2γ) $\Delta\mu$
R_K-11-L0I	1.00 ± 0.03	1.00 ± 0.02	1.00 ± 0.03	0.51 ± 1.10	-1.35 ± 1.06	-2.70 ± 1.73
R_K-11-L0L	1.01 ± 0.02	1.00 ± 0.02	1.00 ± 0.02	-0.31 ± 0.76	0.78 ± 0.82	2.84 ± 1.28
R_K-12-L0I	0.99 ± 0.02	0.99 ± 0.01	1.00 ± 0.02	-0.23 ± 0.63	-0.75 ± 0.69	-3.50 ± 1.07
R_K-12-L0L	1.01 ± 0.01	1.01 ± 0.01	1.01 ± 0.02	0.13 ± 0.56	0.69 ± 0.58	4.73 ± 1.00
R_K-15-L0I	1.00 ± 0.03	1.02 ± 0.02	1.03 ± 0.01	-0.67 ± 1.33	-1.50 ± 1.39	-4.47 ± 1.09
R_K-15-L0L	1.00 ± 0.02	1.00 ± 0.02	0.98 ± 0.01	0.82 ± 1.06	0.70 ± 1.12	6.85 ± 0.69
R_K-16-L0I	0.98 ± 0.01	1.00 ± 0.01	1.01 ± 0.01	-0.08 ± 0.54	-2.21 ± 0.55	-3.53 ± 0.86
R_K-16-L0L	1.01 ± 0.01	1.00 ± 0.01	1.02 ± 0.01	0.09 ± 0.39	1.73 ± 0.42	5.51 ± 0.75
R_K-17-L0I	0.99 ± 0.01	1.00 ± 0.01	1.01 ± 0.01	0.87 ± 0.51	-2.84 ± 0.52	-3.60 ± 0.79
R_K-17-L0L	1.01 ± 0.01	1.00 ± 0.01	1.02 ± 0.02	-0.60 ± 0.40	2.33 ± 0.41	9.13 ± 0.90
R_K-18-L0I	1.02 ± 0.01	0.99 ± 0.01	1.01 ± 0.01	-0.67 ± 0.52	-1.87 ± 0.48	-2.81 ± 0.75
R_K-18-L0L	0.99 ± 0.01	1.01 ± 0.01	1.00 ± 0.01	0.45 ± 0.35	1.38 ± 0.37	3.73 ± 0.63

Table 6.8: Values of the sigma scale ,s_σ and mass shift, $\Delta\mu$, from fits to B^+ data based on smearing derived from B^0 samples

	(0γ) s_σ	(1γ) s_σ	(2γ) s_σ	(0γ) $\Delta\mu$	(1γ) $\Delta\mu$	(2γ) $\Delta\mu$
R_K-11-L0I	1.01 ± 0.03	0.96 ± 0.02	1.01 ± 0.03	-1.97 ± 1.10	-1.53 ± 1.07	-2.87 ± 1.72
R_K-11-L0L	1.01 ± 0.02	0.96 ± 0.01	1.00 ± 0.02	-2.74 ± 0.76	0.68 ± 0.82	2.67 ± 1.28
R_K-12-L0I	1.01 ± 0.02	0.99 ± 0.01	0.99 ± 0.02	-0.23 ± 0.63	-0.40 ± 0.69	-5.37 ± 1.07
R_K-12-L0L	1.03 ± 0.01	1.01 ± 0.01	1.00 ± 0.02	0.05 ± 0.56	1.01 ± 0.58	2.79 ± 1.00
R_K-15-L0I	1.05 ± 0.04	1.02 ± 0.02	1.06 ± 0.03	-1.82 ± 1.31	-1.83 ± 1.39	-2.40 ± 2.04
R_K-15-L0L	1.05 ± 0.02	1.00 ± 0.02	1.01 ± 0.02	-0.51 ± 1.05	0.40 ± 1.12	9.12 ± 1.27
R_K-16-L0I	1.01 ± 0.01	0.99 ± 0.01	1.00 ± 0.01	-0.94 ± 0.54	-1.24 ± 0.55	-4.88 ± 0.86
R_K-16-L0L	1.04 ± 0.01	0.98 ± 0.01	1.01 ± 0.01	-0.75 ± 0.39	2.67 ± 0.42	4.08 ± 0.75
R_K-17-L0I	1.00 ± 0.01	1.03 ± 0.01	1.02 ± 0.01	-0.16 ± 0.51	-4.38 ± 0.52	-5.37 ± 0.79
R_K-17-L0L	1.03 ± 0.01	1.03 ± 0.01	1.02 ± 0.01	-1.64 ± 0.40	0.85 ± 0.41	3.61 ± 0.70
R_K-18-L0I	1.04 ± 0.01	0.99 ± 0.01	0.99 ± 0.01	-1.27 ± 0.51	-3.24 ± 0.48	-3.13 ± 0.75
R_K-18-L0L	1.01 ± 0.01	1.01 ± 0.01	0.99 ± 0.01	-0.12 ± 0.35	0.00 ± 0.37	3.28 ± 0.63

6.8 Total Efficiency

6.8.1 $\varepsilon_{\text{MVA\&sel\&trg\&PID\&reco,strip|flt}}$

The efficiency for $\varepsilon_{\text{MVA\&sel\&trg\&PID\&reco,strip|flt}}$ in Eq. 6.1 is computed using fully corrected simulated samples in accordance with Eq. 6.13. This equation should be used for both B^+ and B^0 modes, muon and electron modes, different trigger categories, different q^2 regions and different years or run periods. It should be noted that if the simulated sample has been filtered then the `MCDecayTuple` is also subject to this filtering.

$$\varepsilon_{\text{MVA\&sel\&trg\&PID\&reco,strip|flt}} = \frac{A}{B} \times \frac{C}{D},$$

$$A = \sum_{w}^{\text{DecayTuple}} (w_{\text{PID}} \times w_{\text{TRK}} \times w_{\text{L0}} \times w_{\text{HLT}} \times w_{\text{Mult\&Kin}} \times w_{\text{Reco}}) \quad |\text{Selection(A)}$$

$$B = \sum_{w}^{\text{MCDecayTuple}} (w_{\text{Mult\&Kin}}) \quad |\text{Selection(B)}$$

$$C = \sum_{w}^{\text{DecayTuple}} (w_{\text{PID}} \times w_{\text{TRK}} \times w_{\text{L0}} \times w_{\text{HLT}} \times w_{\text{Mult\&Kin}}) \quad |\text{Selection(NORM)}$$

$$D = \sum_{w}^{\text{DecayTuple}} (w_{\text{PID}} \times w_{\text{TRK}} \times w_{\text{L0}} \times w_{\text{HLT}} \times w_{\text{Mult\&Kin}} \times w_{\text{Reco}}) \quad |\text{Selection(NORM).} \tag{6.13}$$

The full selection applied to reconstructed (i.e. `DecayTuple`) simulated samples which pass the stripping is split into categories in Tab. 6.9. This can be used to explain what the various selections referenced in Eq. 6.13 correspond to.

Term A - In the term A, the selection $Selection(A)$ corresponds to all the selections presented in the table, applied to the `DecayTuple`, with some caveats depending on which correction weights are enabled which will be described below and also which q^2 region is considered. This term can be considered the fully selected and corrected term hence its presence as a numerator. This term is also important for the normalisation of $w_{\text{Mult\&Kin}}$ weights described more below.

Term B - In the term B, the selection $Selection(B)$ corresponds to `cutSPD`. Ideally this term which acts as the denominator in the usual efficiency calculation would have no selection applied and just have the correction to the generator level multiplicity and kinematics applied. However, as was previously discussed in Sec. 6.3, when simulation is centrally produced it often comes with a selection on `nSPDHits` already applied. To align things in the efficiency calculation for the term B this selection must be applied, effects coming from this should cancel in the ratios. This term is also important for the normalisation of $w_{\text{Mult\&Kin}}$ weights described more below.

Terms C and D - In the terms C and D, the selection $Selection(NORM)$ is defined as $Selection(A)$ with some selections removed. The selections which are removed are `cutBKG`, `cutQ2`, `cutMVA`, `cutEXTRA` and `cutKSTMASS`. This is

Selection label	Content
Occupancy(`cutSPD`)	$n_{SPD} < 600(450)$ for RUN 1 (RUN 2)
Truth-matching(`cutMCT`)	Truth matching selection for efficiencies (See Sec. 4.3).
Preselection(`cutPS`)	Quality cut from Tab. 4.5 Clone cut from Tab. 4.5 `PriorPID` cuts from Tab. 4.6
m(K^{*0}) (`cutKSTMASS`)	$\lvert m(K\pi) - m(K^{*0})_{pdg}\rvert < 100\,\text{MeV}/c^2$ for R_{K^*}
L0 (`cutL0`)	L0-selected trigger category (see Sec. 4.4)
HLT (`cutHLT`)	HLT-selection (see Tab. 4.2)
Trigger TCK alignments (`cutALIGN`)	HLT alignment in 2012 and 2016 (see Sec. 4.5) Alignment for L0E category based on $\texttt{real}_{E_\text{T}}^{\texttt{L0CaloTool}}$ (see Sec. 5.1.2)
PID (`cutPID`)	PID-selection from Tab. 4.5
q^2 (`cutQ2`)	$m^2(\ell\ell)$ depending on q^2 region selected (see Sec. 3.1)
BKG (`cutBKG`)	Background cuts (see Sec. 4.9 for R_{K^*} and Sec. 4.10 for R_K)
MVA (`cutMVA`)	MVA selection from Tab. 4.22
HOP (`cutHOP`)	HOP mass cut applied only for rare modes (see Tab. 4.23)
FitRange (`cutEXTRA`)	Fit range cut for `DTF` and `noDTF` fits (see Sec. 7.6)

Table 6.9: Selections applied for efficiency computation in R_K and R_{K^*} reconstructed ntuples passing the Stripping requirements.

important for the normalisation of the w_Reco weights discussed more below.

Using Tab. 6.9 each correction can be linked to aspects of the selection as follows:

- $\mathbf{w}_\text{PID}$ - PID corrections encode a per-event, per-particle efficiency correction factor based on the selections in `cutPID`. Whenever the PID weights are enabled the selections in `cutPID` are removed and replaced by the PID weights. The selections in `PriorPID` are retained as they align the analysis samples to the `PIDCalib` calibration samples so ensure the PID weights are being used under the same conditions they are determined in.

- $\mathbf{w}_\text{TRK}$ - The tracking weights are unrelated to any selection used in this analysis and can be used at any step in the correction chain.

- $\mathbf{w}_\text{L0}$ - These weights are only used if simulated events are selected for a L0 trigger category (`cutL0`) or when L0-TCKs are aligned (`cutALIGN`)

- $\mathbf{w}_\text{HLT}$ - Very similar to the L0 weights, only used when HLT lines (`cutHLT`) are selected or TCK alignment (`cutALIGN`) for HLT is performed.

- $\mathbf{w}_\text{Mult\&Kin}$ - This correction employs a BDT to correct kinematic and multiplicity related variables. In contrast to the other weights already mentioned the weights are not derived from an efficiency ratio, so a normalisation must be applied to correctly set the efficiency scale. This is achieved by the A/B term in Eq. 6.13.

- $\mathbf{w}_{\text{Reco}}$ - The same line of reasoning holds for the BDT to correct reconstructed related variables except in this case there is no corresponding `MCDecayTuple` to normalise the BDT to. Since this BDT is not normalised in the A/B term in Eq. 6.13 the C/D term serves this purpose. Ideally an `MCDecayTuple` would be used for the normalisation of this step but since events are not reconstructed in the `MCDecayTuple` a compromise must be reached. The next logical place to start would be just after the stripping has been performed, however, since the reconstruction corrections are specific to each trigger category for muon and electron modes the L0 trigger categories must also be selected which in turn require the PID corrections, `cutPS` and `cutHLT` selections to be applied. Following on from this the choice is made to start with the selection of L0 trigger categories and this motivates the removal of criteria from $Selection(A)$ that are not in $Selection(NORM)$. Also selections which are not identical in the different trigger categories and q^2 regions are removed.

6.8.2 Total efficiency (ε_{tot})

The total efficiencies are calculated per year and per polarity. To combine these into a composite efficiency for a given run period the following equation is used which weights the efficiencies by the relevant luminosities:

$$\epsilon_{tot} = \frac{\sum_{i=year,j=polarity} \mathcal{L}^i_j \times \epsilon^{i,j}_{tot}}{\sum_{i=year,j=polarity} \mathcal{L}^i_j} \tag{6.14}$$

where, $\mathcal{L}^i_j$ denotes the luminosity in given year and polarity.

6.8.3 Bootstrapped $\varepsilon_{\text{MVA\&sel\&trg\&PID\&reco,strip|flt}}$ efficiency

When bootstrapping is used Eq. 6.13 is modified in the following way:

$$\begin{aligned}
\varepsilon_{\text{MVA\&sel\&trg\&PID\&reco,strip|flt}}[i] &= \frac{A[i]}{B[i]} \times \frac{C[i]}{D[i]}, \\
A[i] &= \sum_{w}^{\text{DecayTuple}} (w_{\text{BS}}[i] \times w_{\text{PID}}[i] \times w_{\text{TRK}}[i] \times w_{\text{L0}}[i] \times w_{\text{HLT}}[i] \times w_{\text{Mult\&Kin}}[i] \times w_{\text{Reco}}[i]) && \text{|Selection(A)} \\
B[i] &= \sum_{w}^{\text{MCDecayTuple}} (w_{\text{BS}}[i] \times w_{\text{Mult\&Kin}}[i]) && \text{|Selection(B)} \\
C[i] &= \sum_{w}^{\text{DecayTuple}} (w_{\text{BS}}[i] \times w_{\text{PID}}[i] \times w_{\text{TRK}}[i] \times w_{\text{L0}}[i] \times w_{\text{HLT}}[i] \times w_{\text{Mult\&Kin}}) && \text{|Selection(NORM)} \\
D[i] &= \sum_{w}^{\text{DecayTuple}} (w_{\text{BS}}[i] \times w_{\text{PID}}[i] \times w_{\text{TRK}}[i] \times w_{\text{L0}}[i] \times w_{\text{HLT}}[i] \times w_{\text{Mult\&Kin}}[i] \times w_{\text{Reco}}[i]) && \text{|Selection(NORM)}
\end{aligned} \tag{6.15}$$

where the index $i = 1, ...100$ refers to one of the 100 generated Poission-smeared weights discussed previously in Sec. 6.1. Using the formulation in Eq. 6.15 this leads to 100 values of the efficiency. The mean and standard deviation of these values are quoted as the bootstrapped efficiency and its error. As discussed previously in Sec. 6.1, this approach accounts for some of the correlations introduced by the correction procedure as well as indicating the level of stability of the correction procedure. At present the only corrections for which bootstrapping has been employed for is w_{L0} and w_{HLT} as they have the smallest sample sizes to derive corrections from. Bootstrapping for the other correction steps will be added to the analysis in the future.

Chapter 7

Mass Fits

In the previous section the determination of the efficiency was discussed. To compute single or double ratios the signal yields must also be determined from mass fits. The fits in this section are all one dimensional maximum extended likelihood fits to the reconstructed 4-body (B^0) or 3-body (B^+) invariant masses, produced using RooFit. Fits to the resonant modes are performed binned whilst fits to non-resonant modes are performed unbinned.

When fitting there are multiple options:

- Fits can be performed to B^+ and B^0 modes,
- Fits can be performed to muon and electron modes,
- Fits can be performed in the different L0 trigger categories,
- Fits can be performed to different years or data taking periods,
- Fits can be performed in the different q^2 regions,
- Fits can be performed simultaneously in all the above categories or individually,
- Fits can be performed to extract the signal yields or to extract the double or single ratios directly.

Depending on which single or double ratio is being determined, there are three situations which can be arrived at. The following describes the default configuration for each:

$r_{J/\psi}$ In this case both $B^+ \to K^+ J/\psi (\to \ell^+\ell^-)$ and $B^0 \to K^{*0} J/\psi (\to \ell^+\ell^-)$ modes are fitted separately. Within each group of decay modes the run periods (Run 1, Run 2p1 and Run 2p2) and trigger categories (L0I and L0L!) are fitted separately. However the muon and electron modes in each subdivision are fitted simultaneously. In this case the leptonic system's invariant mass is constrained to the known value of the J/ψ mass using `DecayTreeFitter`. This improves the resolution considerably for electron modes and is discussed further in Sec. 7.2.

R_X In this case $B^+ \to K^+\ell^+\ell^-$, $B^0 \to K^{*0}\ell^+\ell^-$, $B^+ \to K^+ J/\psi (\to \ell^+\ell^-)$ and $B^0 \to K^{*0} J/\psi (\to \ell^+\ell^-)$ will be fitted simultaneously in both muon and electron modes, trigger categories and run periods. In this case the resonant mode leptonic system is not constrained to the known value of the J/ψ mass as some parameters are shared between the resonant and non-resonant mode in the simultaneous fits.

$r_{\psi(2S)}$ In this case $B^+ \to K^+\psi(2S)(\to \ell^+\ell^-)$, $B^0 \to K^{*0}\psi(2S)(\to \ell^+\ell^-)$, $B^+ \to K^+ J/\psi (\to \ell^+\ell^-)$ and $B^0 \to K^{*0} J/\psi (\to \ell^+\ell^-)$ are all fitted seperately. Similar to the case for $r_{J/\psi}$ within each group of decay mdoes the run periods and trigger categories are fitted seperately. However, the muon and electron modes in each subdivision are fitted simultaneously. In this case the leptonic systems are constrained to the known value of the J/ψ or $\psi(2S)$ mass as appropriate.

The rest of this section will discuss: Common fit functions (Sec. 7.1), a discussion of the effect of constraining the lepton system to the PDG J/ψ mass (Sec. 7.2), the PDFs used to model signal shapes (Sec. 7.3), the PDFs used to model background shapes (Secs. 7.4 and 7.5), a discussion of the fit ranges (Sec. 7.6), how differences between data and simulation are handled in the fits (Sec. 7.7), how constraints are determined for certain backgrounds (Sec. 7.8) and results from pseudoexperiments validating the fitting procedure (Sec. 7.10).

7.1 Common Fit functions

The analytical functions which are commonly used to fit the signal shapes of interest are defined in the following sections.

7.1.1 Crystal Ball

Crystal Ball functions [135] are commonly used to model processes which include some sort of energy loss, such as bremsstrahlung for ee modes. The crystal ball consists of a Gaussian core with a power law tail taking effect after a certain threshold, α, away from the central value. The function is defined as follows:

$$f(x;\alpha,n,\mu,\sigma) = N \cdot \begin{cases} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, & \text{for } \frac{(x-\mu)}{\sigma} > -\alpha \\ A \cdot (B - \frac{(x-\mu)}{\sigma})^{-n} & \text{for } \frac{(x-\mu)}{\sigma} < -\alpha \end{cases} \tag{7.1}$$

where the constants A and B are defined:

$$A = \left(\frac{n}{|\alpha|}\right)^n \cdot e^{-\frac{|\alpha|^2}{2}} \tag{7.2}$$

$$B = \frac{n}{|\alpha|} - |\alpha| \tag{7.3}$$

As can be seen in Eq. 7.1 changing the sign of α causes the tail to appear on opposite sides of the function. Since the bremsstrahlung recovery algorithm also causes a smaller tail to higher mass values where too many bremsstrahlung photons were recovered it is useful to use double-sided crystal call functions. These functions consist of two Crystal Ball functions where the central values are shared between the two and different α values are defined to capture the behaviour of both the low and high mass tails.

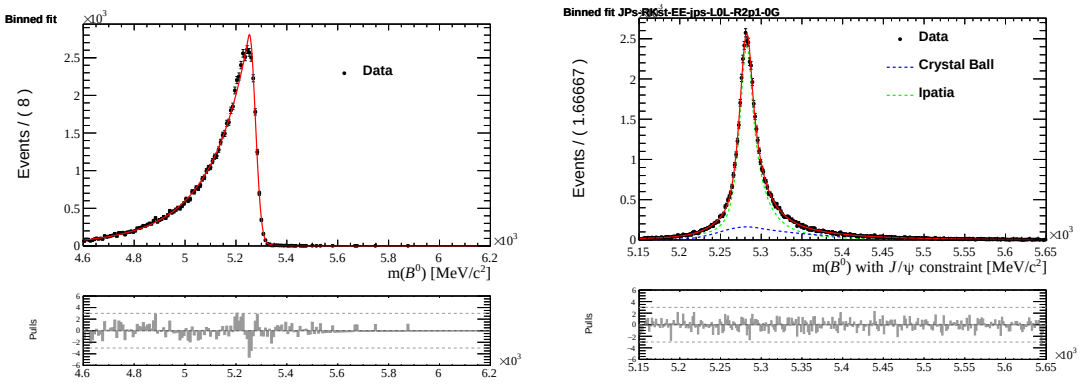
7.1.2 Ipatia

The Ipatia function is similar to the crystal ball except that it has a hyperbolic rather than a Gaussian core, whilst retaining the same crystal ball style tail descriptions. The function is described in detail in Ref. [136].

7.2 Leptonic system constraints

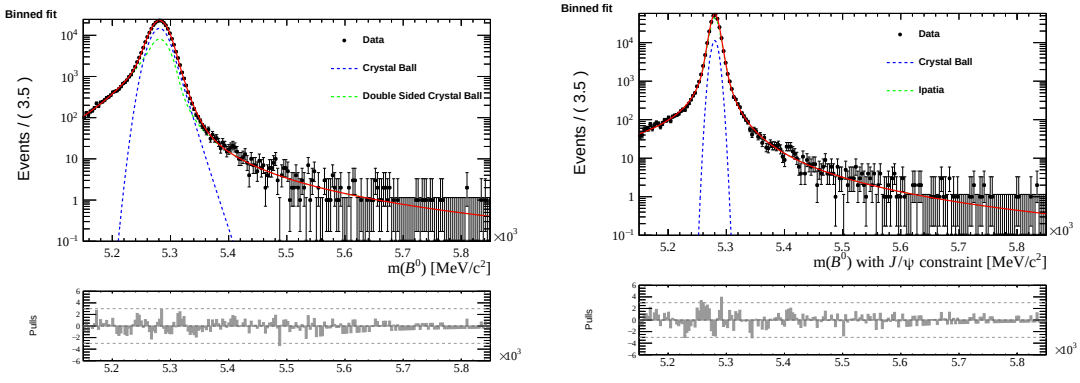
A number of the mass fits performed for the various single and double ratios make use of constraining the invariant mass of the leptonic system to the corresponding PDG mass of the resonance they originate from. This is achieved by using the `DecayTreeFitter` tool in LHCb and denoted $m^{J/\psi}_{DTF}(m^{\psi(2S)}_{DTF})$ for the J/ψ ($\psi(2S)$) resonant modes. Since electrons are affected by bremsstrahlung the affect of this constraint is to correct for the momentum and energy loss due to this effect, improving

the resolution significantly. Muons are not affected by bremsstrahlung so the effect on muon modes is much more marginal although the resolution can be seen to improve. The comparisons between electron modes fitted with and without a J/ψ constraint can be seen in Fig. 7.1 and the same situation for muons can be seen in Fig. 7.2. Fitting with the mass constraint is the preferred option where only resonant modes are available, however, for the determination of R_X where non-resonant modes are also used this can no longer be used in the fits since the resonant and non-resonant modes share certain fit parameters.



(a) $m(K\pi ee)$ (b) $m(K\pi ee)^{J/\psi}_{DTF}$

Figure 7.1: Fits to simulated $B^0 \to K^{*0} J/\psi (\to e^+e^-)$ in the 0γ category, without (left) and with (right) J/ψ DTF mass constraint. With DTF J/ψ mass constraint, the bremsstrahlung radiated lower mass tail is removed and mass resolution is significantly improved. Note that the x-axis of both plots are different. The fits shown above uses $15 + 16$ simulation passing the electron L0L triggers.



(a) $m(K\pi\mu\mu)$ (b) $m(K\pi\mu\mu)^{J/\psi}_{DTF}$

Figure 7.2: Fits to simulated $B^0 \to K^{*0} J/\psi (\to \mu^+\mu^-)$, without (left) and with (right) J/ψ DTF mass constraint. Unlike the electron mode in Fig. 7.1, improvements in the mass resolution are not as significant because the muon modes do not suffer from bremsstrahlung radiation. The fits shown above uses Run 2p1 simulation passing the electron L0L triggers.

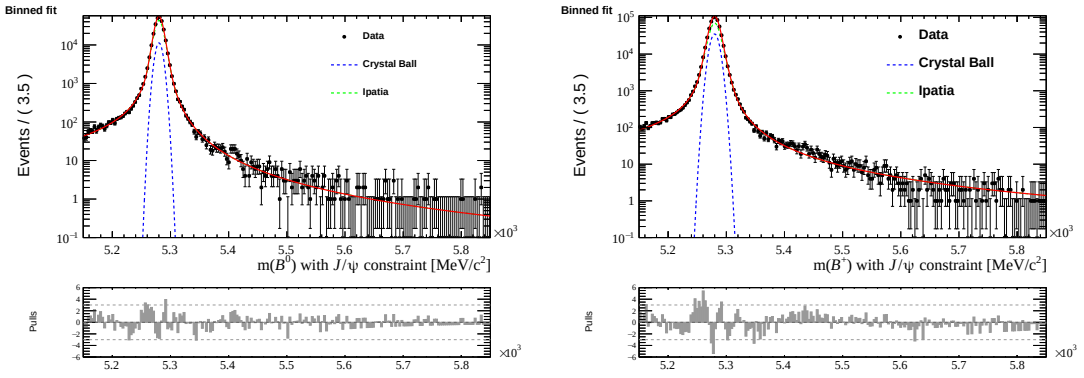
7.3 Signal PDFs

This section introduces the PDFs used to describe the signal components in the resonant and non-resonant fits. To determine the PDF parameters to be used in the collision data fits, first the signal model in question is fitted to simulation then those parameters are used on the collision data fits. As such, in this section fits to simulation will be displayed, also only fits to $m^{J/\psi}_{DTF}$ will be shown for the resonant mode since this work focuses on the measurement of $r_{J/\psi}$.

7.3.1 Muon modes

The signal components in the fits to the reconstructed B meson masses in $B^0 \to K^{*0} J/\psi (\to \mu^+\mu^-)$ and $B^+ \to K^+ J/\psi (\to \mu^+\mu^-)$ modes are modelled by a linear combination of an Ipatia and a Crystal ball function. For $B^0 \to K^{*0}\mu^+\mu^-$ and $B^+ \to K^+\mu^+\mu^-$ modes a single Ipatia function is used to model the signal component.

Examples of fits to simulated resonant and non-resonant modes in the central-q^2 region, respectively, can be seen in Figs. 7.3 and 7.4. The fits in the low-q^2 region are very similar to the central-q^2 case. Only RUN 2P1 samples in the L0L trigger category are shown as they are fairly representative of the other years and trigger categories.

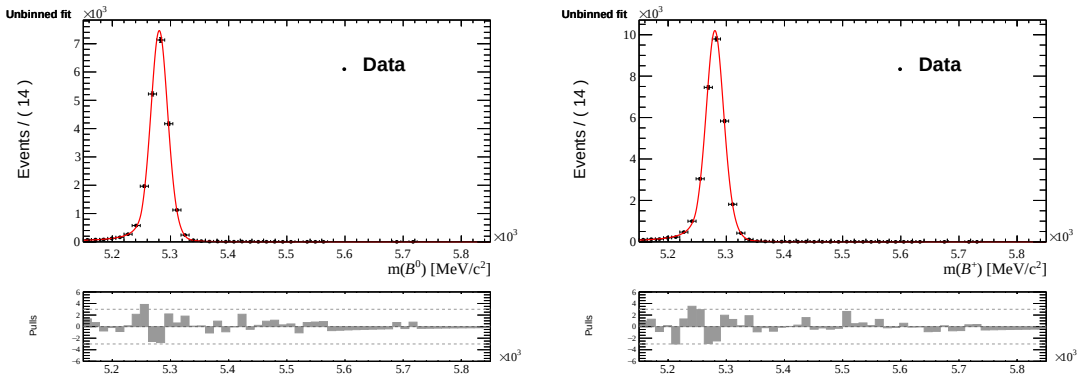


(a) $B^0 \to K^{*0} J/\psi (\to \mu^+\mu^-)$ (b) $B^+ \to K^+ J/\psi (\to \mu^+\mu^-)$

Figure 7.3: Fits to simulated muon control modes $B^0 \to K^{*0} J/\psi (\to \mu^+\mu^-)$ ($B^+ \to K^+ J/\psi (\to \mu^+\mu^-)$) are shown on the left(right). The fits shown above use RUN 2P1 simulation passing the muon L0L triggers.

7.3.2 Electron modes

The reconstructed invariant mass distribution of electron modes depends on how many radiated bremsstrahlung photons are recovered. For this reason three different



(a) $B^0 \to K^{*0}\mu^+\mu^-$, central-q^2 . (b) $B^+ \to K^+\mu^+\mu^-$, central-q^2 .

Figure 7.4: Fits to simulated muon rare modes $B^0 \to K^{*0}\mu^+\mu^-$ ($B^+ \to K^+\mu^+\mu^-$) are shown on the left (right) in the central-q^2 bin. The fits shown above uses Run 2p1 simulation passing the muon L0L triggers.

PDFs are used to model the signal component in each of three bremsstrahlung categories for the non-resonant modes and also resonant modes which are not mass constrained. The 0γ category is modelled with a double-sided Crystal Ball function, while the 1γ and 2γ categories are modelled using a sum of a double-sided Crystal Ball and a Gaussian with independent parameters. To create a total PDF they are added linearly using the relative fractions determined from fits to fully corrected simulated samples. These fractions can be varied within their uncertainties to assign a systematic uncertainty. For J/ψ modes which are subject to mass constraints in each bremsstrahlung category the signal shape is modelled with a combination of an Ipatia and a Crystal Ball function.

Examples of fits to the simulated resonant and the non-resonant mode in the central-q^2 region can be seen in Figs. 7.5 and 7.6. As before the low-q^2 case is very similar to the central-q^2 case. Only the Run 2p1 plots in the L0L category are shown here as the distributions are fairly similar over other years and trigger category.

7.4 Background PDFs: B^0 modes

Background shapes are modelled with either analytical functions or KDEs using the `RooKeysPDF` class in RooFit to estimate the PDF shape from simulated samples.

This section presents plots from Run 2p1 in the L0L trigger category as is the case for Signal PDFs.

7.4.1 $B^0 \to K^{*0} J/\psi (\to \mu^+\mu^-)$

The following backgrounds are considered:

- Combinatorial: Modelled with an exponential function with the slope allowed to float.
- $\Lambda_b^0 \to pKJ/\psi (\to \mu^+\mu^-)$: Modelled with a `RooKeysPdf` using simulated data of these decays corrected for the pK Dalitz plot [137]. The constructed PDF can be seen in Fig. 7.7.
- $B_s^0 \to K^{*0} J/\psi (\to \mu^+\mu^-)$: Modelled using the signal component PDF, described inSec.7.3.1, with the mean of the distribution offset by $m_{B_s^0}^{PDG} - m_{B^0}^{PDG}$.
- Partially Reconstructed: Modelled using a `RooKeysPdf` with inclusive samples of B^0, B^+ and B_s^0 decays. This considers decays of the sort $B \to XJ/\psi (\to \mu^+\mu^-)$ where the X may be any allowed hadron and muons are always required to have originated from a J/ψ resonance. How these samples are combined to generate a PDF is described in detail in Appendix J. The constituents of the combined PDF can be seen in Fig. 7.8.

7.4.2 $B^0 \to K^{*0} J/\psi (\to e^+e^-)$

The following backgrounds are considered:

- Combinatorial: Modelled with an exponential function with the slope allowed to float.
- $\Lambda_b^0 \to pKJ/\psi (\to e^+e^-)$: Modelled with a `RooKeysPdf` using simulated data of these decays corrected for the pK Dalitz plot [137]. The constructed PDF can be seen in Fig. 7.9.
- $B_s^0 \to K^{*0} J/\psi (\to ee)$: Modelled using the signal component PDF, described in 7.3.2, with the mean of the distribution offset by $m_{B_s^0}^{PDG} - m_{B^0}^{PDG}$.
- Partially Reconstructed (J/ψ only): Modelled using a `RooKeysPdf` with inclusive samples of B^0, B^+ and B_s^0 decays. This considers decays of the sort $B \to XJ/\psi (\to \mu^+\mu^-)$ where the X may be any allowed hadron and muons are always required to have originated from a J/ψ resonance. How these samples are combined to generate a PDF is described in detail in Appendix J and the constituents of the combined PDF can be seen in Fig. 7.10.

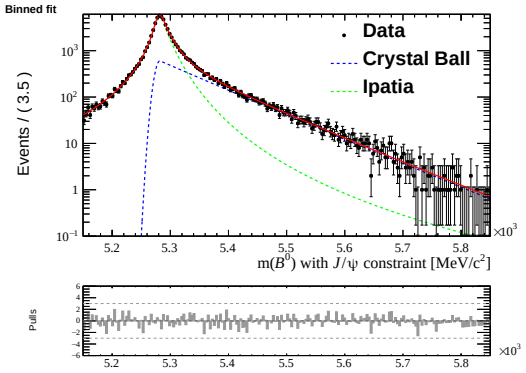
7.4.3 $B^0 \to K^{*0}\mu^+\mu^-$

The only background considered is combinatorial which is modelled by an exponential function with its slope allowed to float.

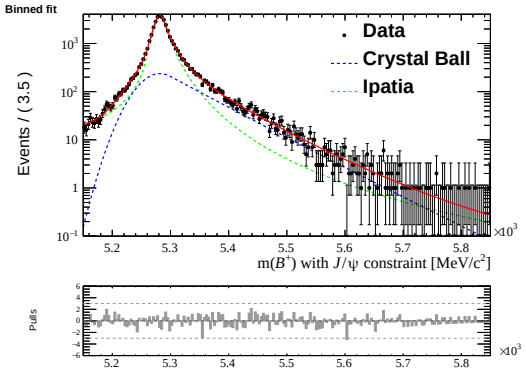
7.4.4 $B^0 \to K^{*0}e^+e^-$

The following backgrounds are considered:

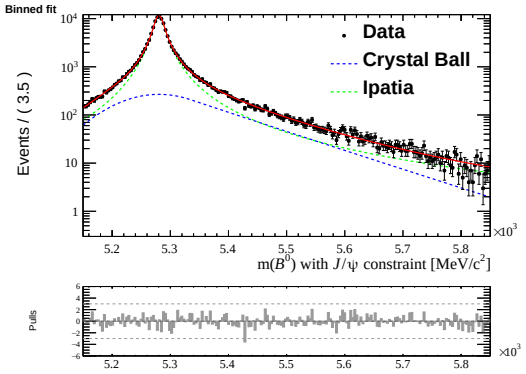
- Combinatorial: Modelled with an exponential function with its slope allowed to float.
- Partially Reconstructed $B^+ \to K\pi\pi e^+e^-$: Modelled with a `RooKeysPdf` using simulated data containing decays of higher K^{*0} resonances, such as $K_1(1270)$ and $K_2^*(1430)$, using BKGCAT categories of 20 or 30 or 40 or 50. The constructed PDF can be seen in Fig. 7.11.
- Leakage from $B^0 \to K^{*0}J/\psi(\to e^+e^-)$ (central-q^2 only): Modelled with a `RooKeysPdf` using simulated $B^0 \to K^{*0}J/\psi(\to e^+e^-)$ data which passes the central-q^2 selection. The constructed PDF can be seen in Fig. 7.12.



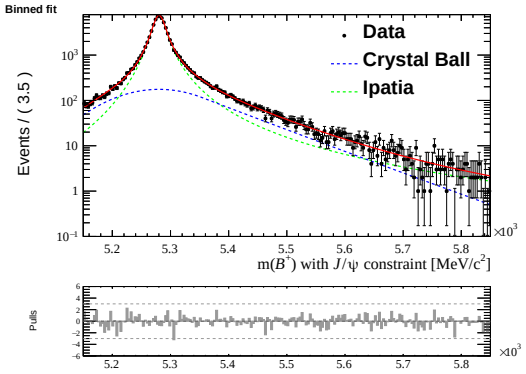
(a) $\mathbf{0\gamma}$ category, $\boldsymbol{B^0 \to K^{*0} J/\psi (\to e^+e^-)}$.



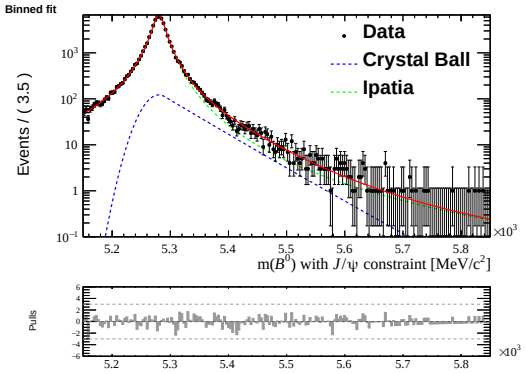
(b) $\mathbf{0\gamma}$ category, $\boldsymbol{B^+ \to K^{+} J/\psi (\to e^+e^-)}$.



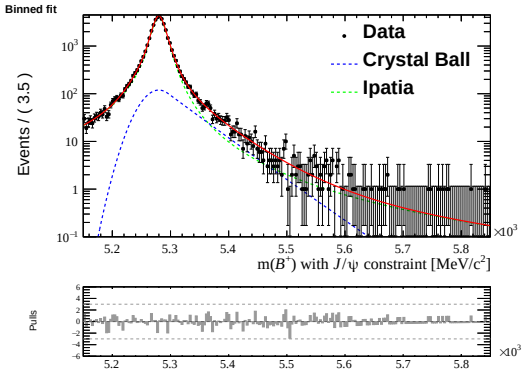
(c) $\mathbf{1\gamma}$ category, $\boldsymbol{B^0 \to K^{*0} J/\psi (\to e^+e^-)}$.



(d) $\mathbf{1\gamma}$ category, $\boldsymbol{B^+ \to K^{+} J/\psi (\to e^+e^-)}$.

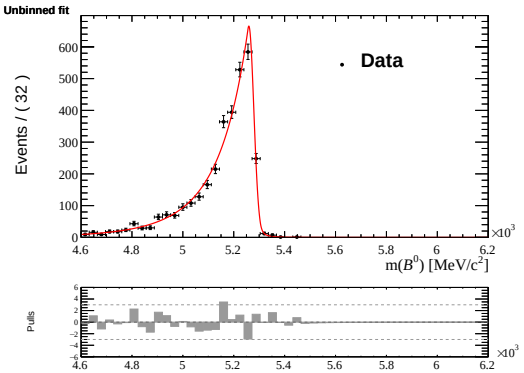


(e) $\mathbf{2\gamma}$ category, $\boldsymbol{B^0 \to K^{*0} J/\psi (\to e^+e^-)}$.

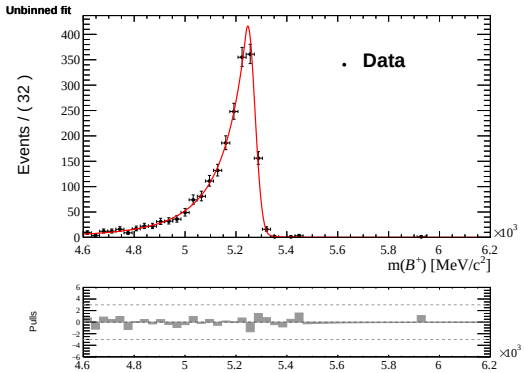


(f) $\mathbf{2\gamma}$ category, $\boldsymbol{B^+ \to K^{+} J/\psi (\to e^+e^-)}$.

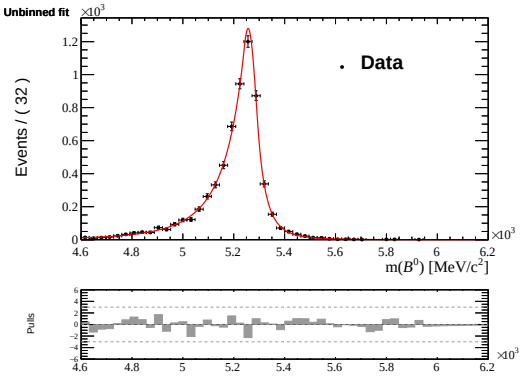
Figure 7.5: Fits to simulated electron J/ψ modes $B^0 \to K^{*0} J/\psi (\to e^+e^-)$ ($B^+ \to K^+ J/\psi (\to e^+e^-)$) are shown on the left(right). A DTF J/ψ mass constraint is applied on the dielectron system. The $\mathbf{0\gamma}$ (top), $\mathbf{1\gamma}$ (middle) and $\mathbf{2\gamma}$ (bottom) bremsstrahlung categories are modelled separately. The fits shown above uses Run 2p1 simulation passing the electron L0L triggers.



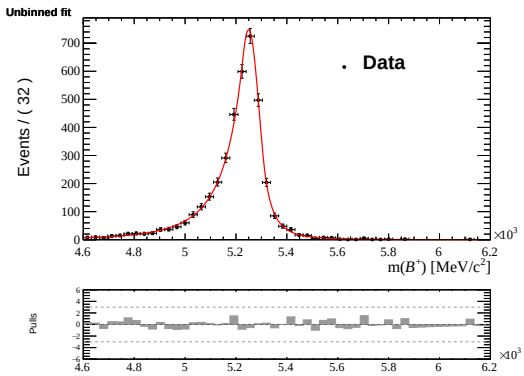
(a) $\mathbf{0\gamma}$ category, central-$\boldsymbol{q^2}$ $\boldsymbol{B^0 \to K^{*0} e^+ e^-}$.



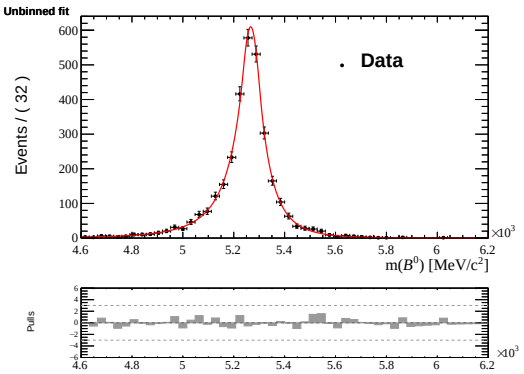
(b) $\mathbf{0\gamma}$ category, central-$\boldsymbol{q^2}$ $\boldsymbol{B^+ \to K^+ e^+ e^-}$.



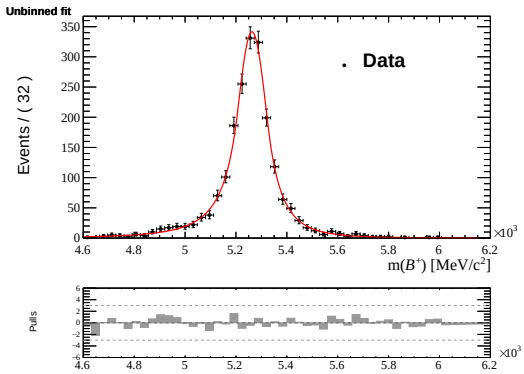
(c) $\mathbf{1\gamma}$ category, central-$\boldsymbol{q^2}$ $\boldsymbol{B^0 \to K^{*0} e^+ e^-}$.



(d) $\mathbf{1\gamma}$ category, central-$\boldsymbol{q^2}$ $\boldsymbol{B^+ \to K^+ e^+ e^-}$.

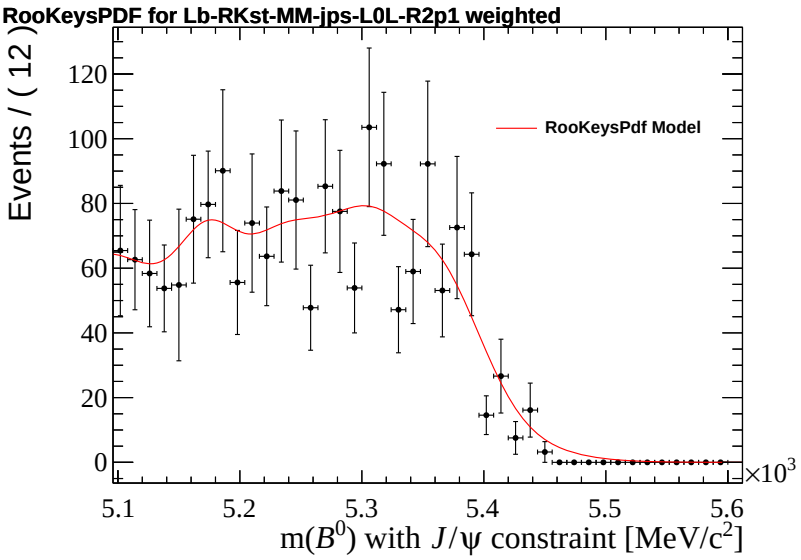


(e) $\mathbf{2\gamma}$ category, central-$\boldsymbol{q^2}$ $\boldsymbol{B^0 \to K^{*0} e^+ e^-}$.



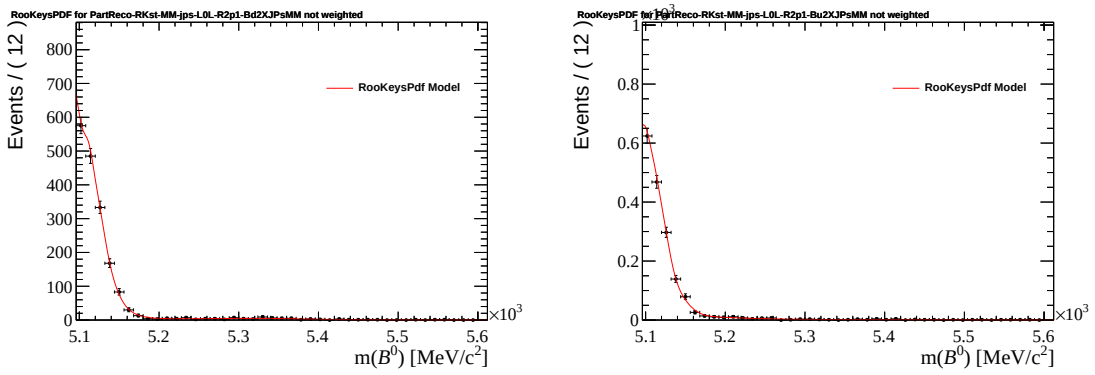
(f) $\mathbf{2\gamma}$ category, central-$\boldsymbol{q^2}$ $\boldsymbol{B^+ \to K^+ e^+ e^-}$.

Figure 7.6: Simulated electron signal modes in central-q^2 bin. The fits shown above use RUN 2P1 simulation passing the electron L0L triggers. Simulated $B^0 \to K^{*0} e^+ e^-$ ($B^0 \to K^{*0} e^+ e^-$) are shown on the left(right), separated into $\mathbf{0\gamma}$ (top), $\mathbf{0\gamma}$ (middle) and $\mathbf{2\gamma}$ (bottom) Bremsstrahlung categories.

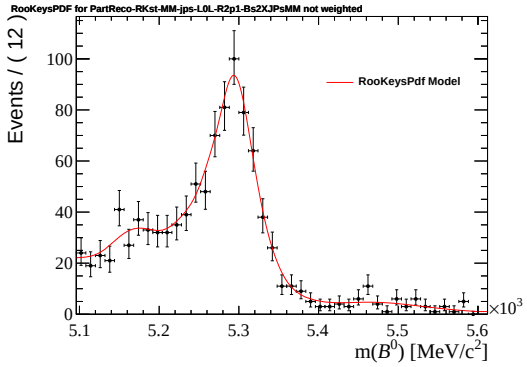


(a) $\Lambda_b^0 \to pKJ/\psi(\to \mu^+\mu^-)$

Figure 7.7: `RooKeysPdf` KDE modelled with simulated mis-identified Λ_b^0 backgrounds in the $\mu\mu$ J/ψ resonant mode, L0L Run 2p1.

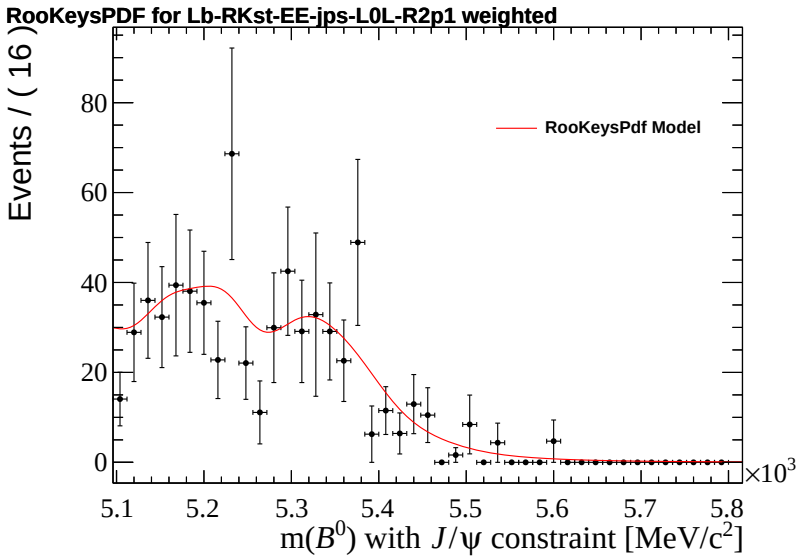


(a) $B^0 \to XJ/\psi(\to \mu^+\mu^-)$ (b) $B^+ \to XJ/\psi(\to \mu^+\mu^-)$



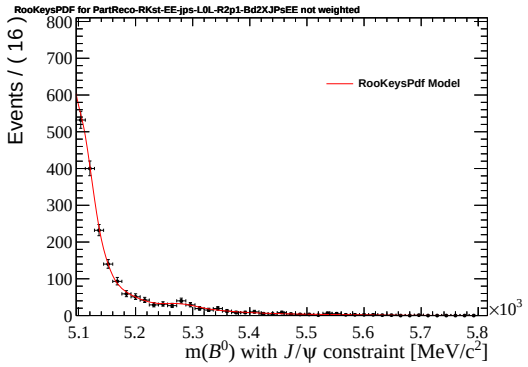
(c) $B_s^0 \to XJ/\psi(\to \mu^+\mu^-)$

Figure 7.8: Partially reconstructed $B^0 \to K^{*0}J/\psi(\to \mu^+\mu^-)$ background modelled using `RooKeysPdf` KDE on cocktail Run 2p1 simulation passing L0L trigger. $B^0 \to XJ/\psi(\to \mu^+\mu^-)$ (top left), $B^+ \to XJ/\psi(\to \mu^+\mu^-)$ (top right) and $B_s^0 \to XJ/\psi(\to \mu^+\mu^-)$ (bottom)

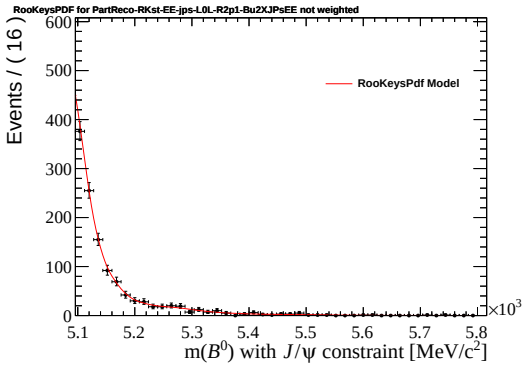


(a) $\Lambda_b^0 \to pKJ/\psi(\to e^+e^-)$

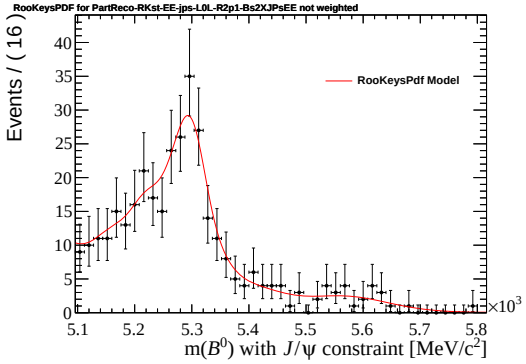
Figure 7.9: `RooKeysPdf` KDE modelled with simulated mis-identified Λ_b^0 backgrounds in the ee J/ψ resonant mode, L0L Run 2p1.



(a) $B^0 \to XJ/\psi(\to e^+e^-)$

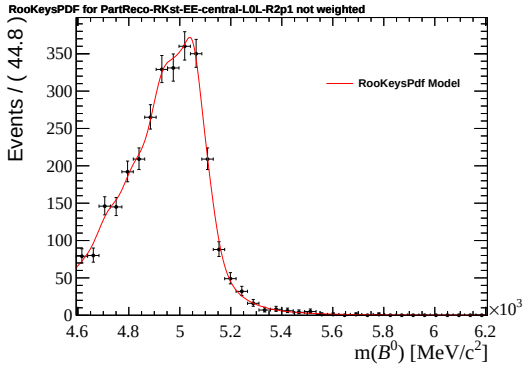


(b) $B^+ \to XJ/\psi(\to e^+e^-)$

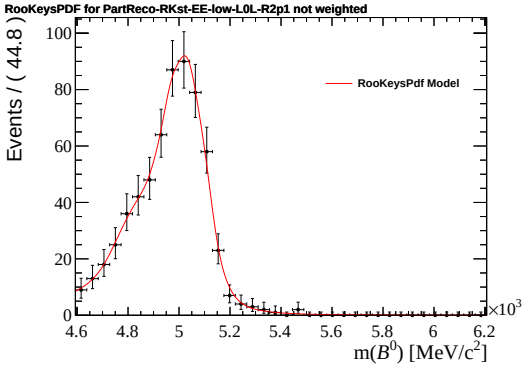


(c) $B^0_s \to XJ/\psi(\to e^+e^-)$

Figure 7.10: Partially reconstructed J/ψ mass constrained $B^0 \to K^{*0}J/\psi(\to e^+e^-)$ background modelled using `RooKeysPdf` KDE on cocktail Run 2p1 Monte-Carlo passing L0L trigger. $B^0 \to XJ/\psi(\to e^+e^-)$ (top left), $B^+ \to XJ/\psi(\to e^+e^-)$ (top right) and $B^0_s \to XJ/\psi(\to e^+e^-)$ (bottom)



(a) low-q^2



(b) central-q^2

Figure 7.11: Partially reconstructed $B^+ \to K\pi\pi e^+e^-$ background, reconstructed as $B^0 \to K^{*0}e^+e^-$ in low-q^2 (left) and central-q^2 (right). Modelled using `RooKeysPdf` KDE on Run 2p1 Monte-Carlo which passes the L0L trigger.

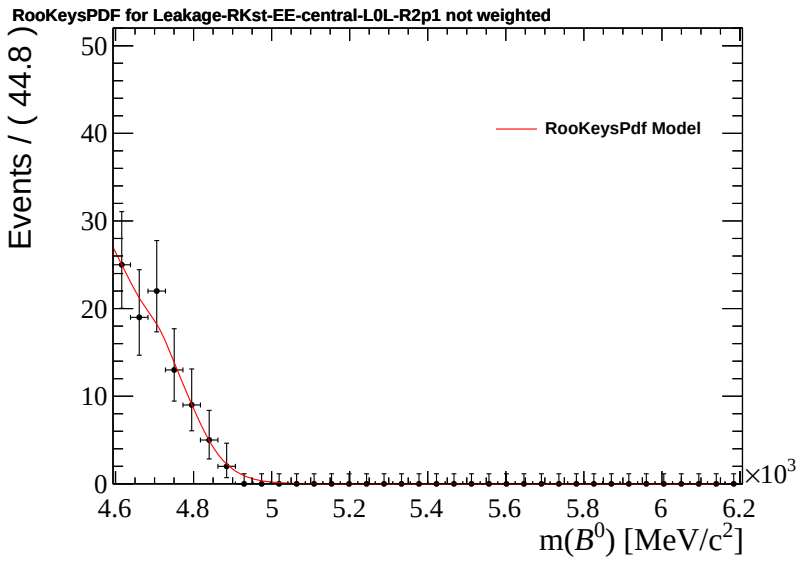


Figure 7.12: $B^0 \to K^{*0} J/\psi (\to e^+e^-)$ q^2 bin leakage into the central-q^2 bin of $B^0 \to K^{*0}e^+e^-$ due to Bremsstrahlung losses. Modelled using `RooKeysPdf` KDE on Run 2p1 Monte Carlo passing the L0L trigger.

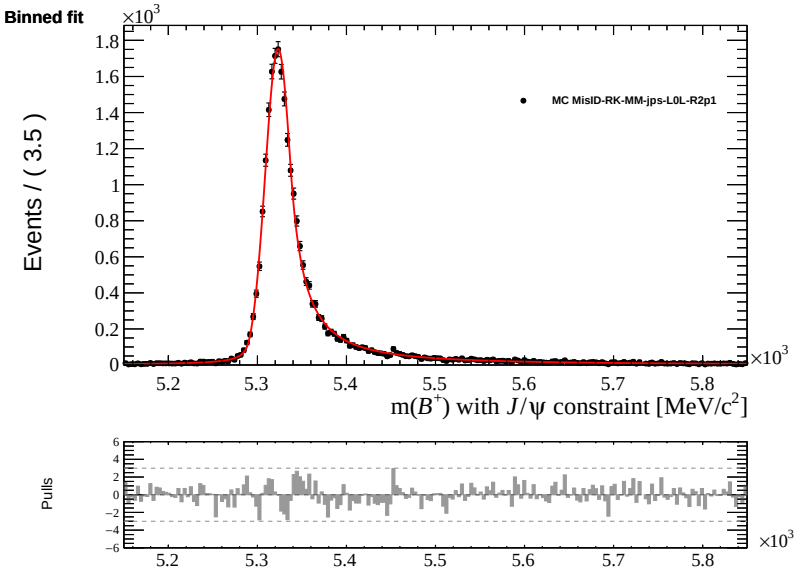
7.5 Background PDFs: B^+ modes

As in the previous section backgrounds are either modelled with an analytical function or a `RooKeysPdf`. The plots in this section show Run 2p1 results in the L0L trigger category.

7.5.1 $B^+ \to K^+ J/\psi (\to \mu^+\mu^-)$

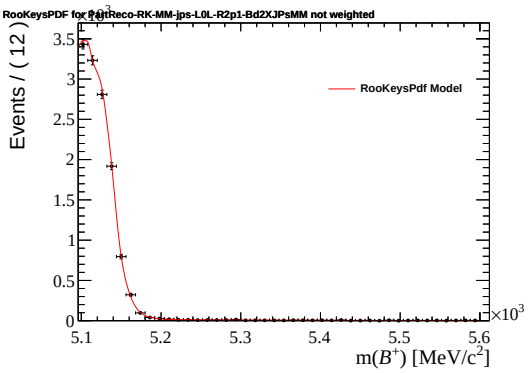
The following backgrounds are considered:

- Combinatorial: Modelled with an exponential function with the slope allowed to float;
- $B^+ \to \pi^+ J/\psi (\to \mu^+\mu^-)$: Modelled with a double sided Crystal Ball fitted to non-corrected simulation. The constructed PDF is shown in Fig. 7.13.
- Partially Reconstructed: Modelled with a `RooKeysPdf` using inclusive simulated samples from B^0, B^+ and B^0 decays of the type $B \to X J/\psi (\to \mu^+\mu^-)$. How these samples are combined to generate a PDF is described in detail in Appendix J. The constituents of the combined PDF are shown in Fig. 7.14.

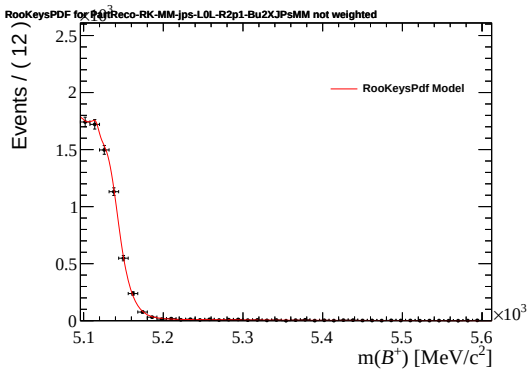


(a) $B^+ \to \pi^+ J/\psi (\to \mu^+\mu^-)$

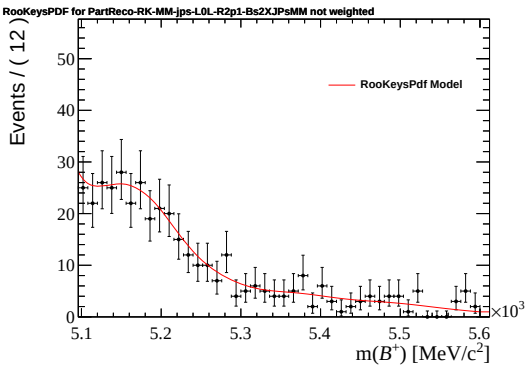
Figure 7.13: Fits to simulated $\pi \to K$ mis-identified $B^+ \to \pi^+ J/\psi (\to \mu^+\mu^-)$ backgrounds reconstructed as $B^+ \to K^+ J/\psi (\to \mu^+\mu^-)$. Modelled using Run 2p1 Monte Carlo passing the L0L trigger.



(a) $B^0 \to XJ/\psi(\to \mu^+\mu^-)$



(b) $B^+ \to XJ/\psi(\to \mu^+\mu^-)$



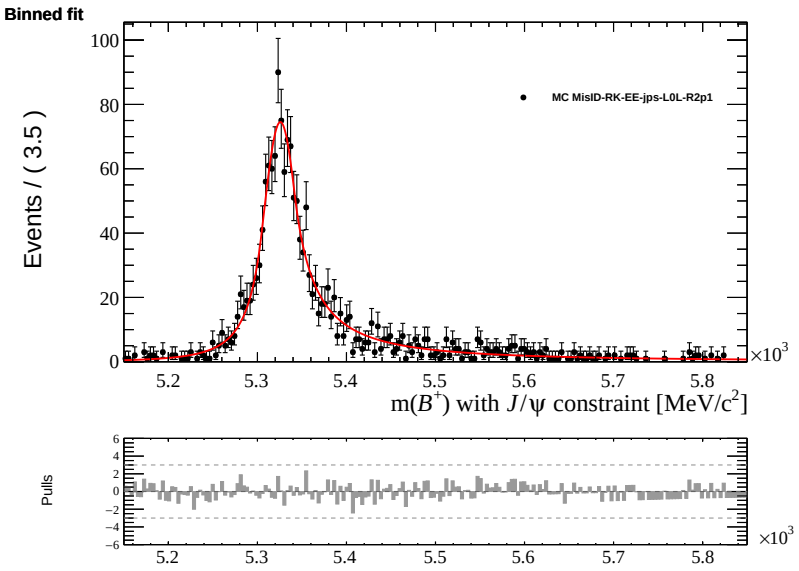
(c) $B_s^0 \to XJ/\psi(\to \mu^+\mu^-)$

Figure 7.14: Partially reconstructed $B^+ \to K^+J/\psi(\to \mu^+\mu^-)$ background modelled using `RooKeysPdf` KDE on cocktail Run 2p1 Monte-Carlo passing L0L trigger. $B^0 \to XJ/\psi(\to \mu^+\mu^-)$ (top left), $B^+ \to XJ/\psi(\to \mu^+\mu^-)$ (top right) and $B_s^0 \to XJ/\psi(\to \mu^+\mu^-)$ (bottom).

7.5.2 $B^+ \to K^+J/\psi(\to e^+e^-)$

The following backgrounds are considered:

- Combinatorial: Modelled with an exponential function with the slope allowed to float
- $B^+ \to \pi^+J/\psi(\to e^+e^-)$: Modelled with a double sided Crystal Ball fitted to non-corrected simulation. The constructed PDF is shown in Fig. 7.15.
- Partially Reconstructed : modelled with a `RooKeysPdf` using inclusive simulated samples from B^0, B^+ and B_s^0 decays of the type $B \to XJ/\psi(\to ee)$, where the electrons are required to originate from the J/ψ resonance. How these samples are combined to generate a PDF is described in detail in Appendix J. The constituents of the combined PDF can be seen in Fig. 7.16.



(a) $B^+ \to \pi^+ J/\psi(\to e^+e^-)$

Figure 7.15: Fits to simulated $\pi \to K$ mis-identified $B^+ \to \pi^+ J/\psi(\to e^+e^-)$ backgrounds reconstructed as $B^+ \to K^+ J/\psi(\to e^+e^-)$. Modelled using Run 2p1 Monte Carlo passing the L0L trigger.

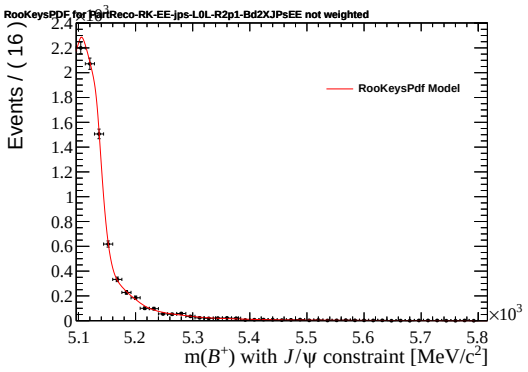
7.5.3 $B^+ \to K^+\mu^+\mu^-$

The only background considered is combinatorial background which is modelled by an exponential function with its slope allowed to float.

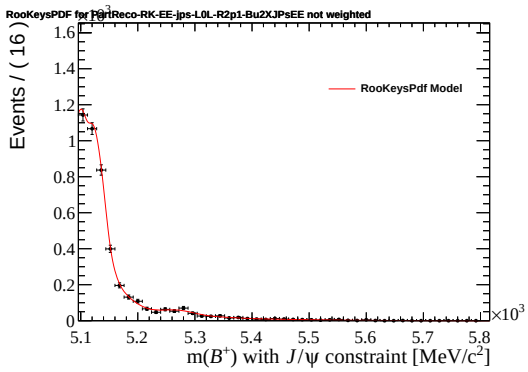
7.5.4 $B^+ \to K^+e^+e^-$

The following backgrounds are considered:

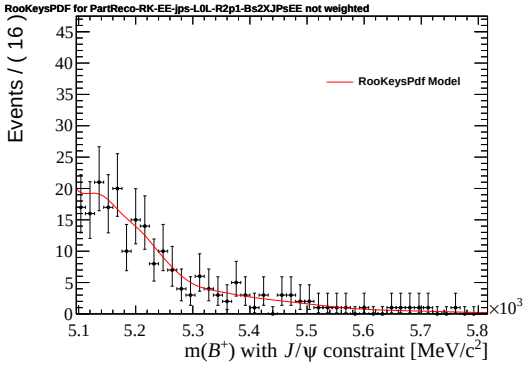
- Combinatorial: modelled with an exponential function with its slope allowed to float.
- Partially Reconstructed $B^0 \to K^{*0}e^+e^-$ or $B^0 \to K^+\pi^-e^+e^-$: modelled with a `RooKeysPdf` using simulated $B^0 \to K^{*0}e^+e^-$ reconstructed as $B^+ \to K^+e^+e^-$ and also requiring the BKGCAT to be 20 or 30 or 40 or 50. The resulting PDF can be seen in Fig. 7.17. The non-resonant decay mode must also be considered in a similar way using $B^0 \to K^+\pi^-e^+e^-$ phase space simulation and is also shown in Fig. 7.17
- Leakage from $B^+ \to K^+ J/\psi(\to e^+e^-)$ (central-q^2 only): modelled with a `RooKeysPdf` using simulated $B^+ \to K^+ J/\psi(\to e^+e^-)$ which passes the central-q^2 selection for $B^+ \to K^+e^+e^-$ applied. The resulting PDF can be



(a) $B^0 \to XJ/\psi(\to e^+e^-)$



(b) $B^+ \to XJ/\psi(\to e^+e^-)$



(c) $B^0_s \to XJ/\psi(\to e^+e^-)$

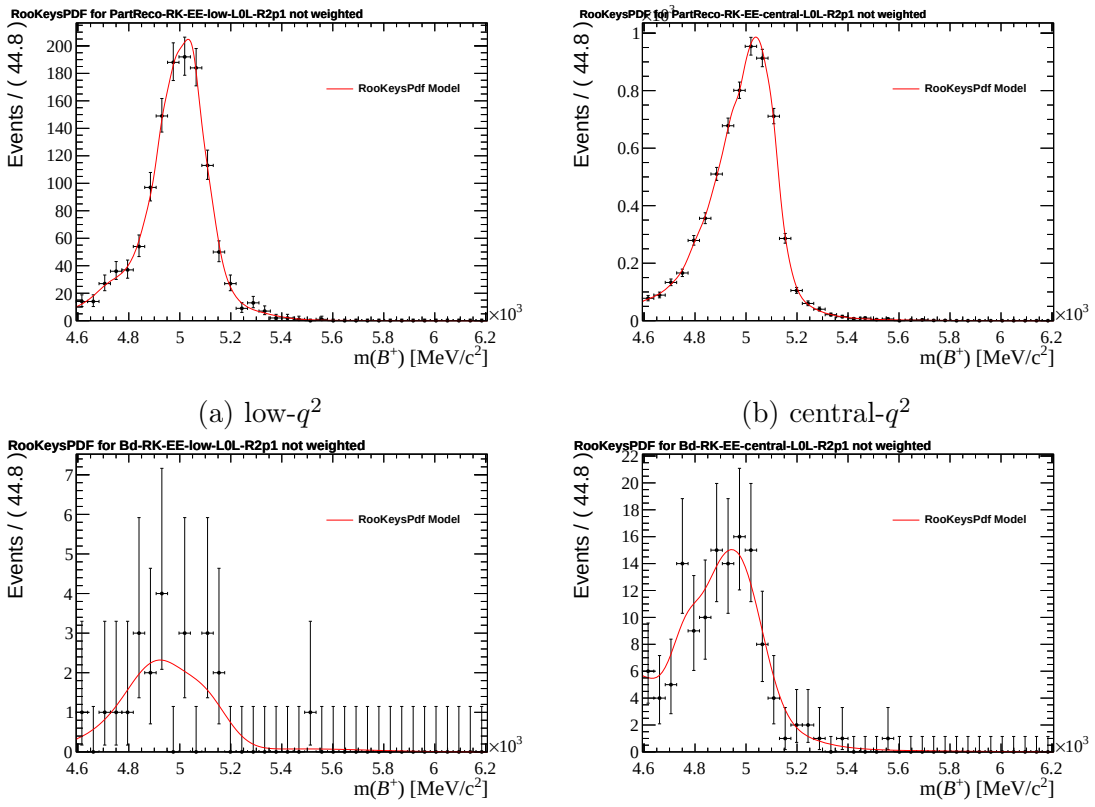
Figure 7.16: Partially reconstructed J/ψ mass constrained $B^+ \to K^+J/\psi(\to e^+e^-)$ background modelled using `RooKeysPdf` KDE on cocktail Run 2p1 Monte-Carlo passing L0L trigger. $B^0 \to XJ/\psi(\to e^+e^-)$ (top left), $B^+ \to XJ/\psi(\to e^+e^-)$ (top right) and $B^0_s \to XJ/\psi(\to e^+e^-)$ (bottom).

seen in Fig. 7.18

7.6 Fit Ranges

The mass windows used for fitting are the same for each q^2 bin per leptonic final state. Electrons have larger mass windows in which the fit is performed due to the effect of bremsstrahlung. The various fit ranges of interest are as follows:

Muonic J/ψ modes : A range of 5100 MeV/c^2 to 5800 MeV/c^2 is used in fits both with and without a J/ψ mass constraint. Partially reconstructed background dominates in the region 5100-5150 MeV/c^2 and begins to leak into the signal region from 5150 MeV/c^2. The lower limit of 5100 MeV/c^2 is sufficient to constrain this background. The upper limit of 5800 MeV/c^2 is chosen to constrain the tail of the combinatorial background.



(a) low-q^2 (b) central-q^2

(c) low-q^2 (d) central-q^2

Figure 7.17: Partially reconstructed $B^0 \to K^{*0}e^+e^-$ (top) and $B^0 \to K^+\pi^-e^+e^-$ (bottom) background, reconstructed as $B^+ \to K^+e^+e^-$ in the low-q^2 (left) and central-q^2 (right). Modelled using `RooKeysPdf` KDE on Run 2p1 simulation which passes the L0L trigger.

Muonic non-resonant modes : A range of 5150 MeV/c^2 to 5850 MeV/c^2 is used. Within this range only the signal modes and combinatorial backgrounds meaningfully contribute. The lower edge is chosen to exclude partially reconstructed background and the upper edge is chosen to sufficiently constrain the tail of the combinatorial background.

Electronic J/ψ modes : There are two ranges, one for the J/ψ mass constrained fit used to determine $r_{J/\psi}$ which is 5100 MeV/c^2–5800 MeV/c^2 and one for the unconstrained fit used in the determination of R_X which is 4400 MeV/c^2 to 6200 MeV/c^2. With the mass constrained fit the situation is much like that of the muonic J/ψ mode, with the lower edge being sufficient to constrain the partially reconstructed background and the upper limit sufficient to constrain the tail of the combinatorial background. In the unconstrained case bremsstrahlung effects cause the tail of the electron signal to extend to 4600 MeV/c^2. Partially

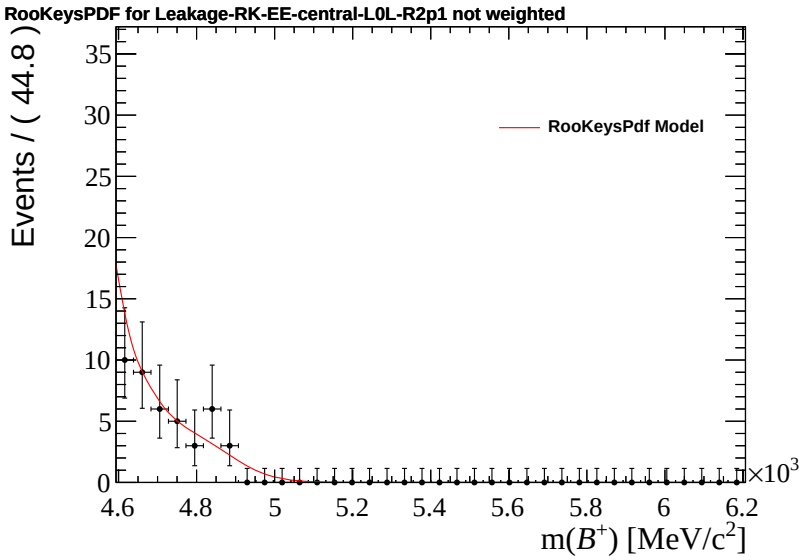


Figure 7.18: $B^+ \to K^+ J/\psi (\to e^+e^-)$ q^2 bin leakage into the central-q^2 bin of $B^+ \to K^+e^+e^-$ due to Bremstrahlung losses. Modelled using `RooKeysPdf` KDE on Run 2p1 Monte Carlo passing the L0L trigger.

reconstructed backgrounds dominate in this region and peak around 4600 MeV/c^2 so the fit range is extended to 4400 MeV/c^2 to constrain this adequately. The upper range is chosen to constrain the tail of the combinatorial background.

Electron non-resonant modes : A range of 4600 MeV/c^2 to 6200 MeV/c^2 is used. As is the case for unconstrained electronic J/ψ modes bremsstrahlung effects cause the signal tail to extend to 4600 MeV/c^2, however, unlike the control mode case the partially reconstructed backgrounds are suppressed by a combination of the partially reconstructed MVA and HOP requirements, meaning 4600 MeV/c^2 is sufficiently low enough to constrain the backgrounds and signal. The upper range is chosen to constrain the tail of the combinatorial background.

7.7 Data/Simulation differences in analytical PDFs

When an analytical function is used to model a signal or background shape, there are many parameters which can be allowed to float but simply allowing every parameter to do this can lead to unstable fits. To avoid this, first the shapes of the relevant fit components are modelled using fits to uncorrected simulation and the parameters of the fits are fixed. These can then be used in the fits to data, however, some differences between the data and simulation are expected. To account for a shift in the central value, $\mu \to \mu + m_{shift}$, a mass shift parameter is introduced. To account

for a scaling of the width ,$\sigma \to \sigma \cdot s_{scale}$, a width scaling parameter is introduced. Both the mass shift m_{shift} and width scaling s_{scale} parameters are left to float in the fit. An attempt to remedy this difference was discussed in Sec. 6.7.

When fitting in the R_X configuration $B^0 \to K^{*0}\ell^+\ell^-$ and $B^+ \to K^+\ell^+\ell^-$ decays, which suffer from low statistics, it is found that m_{shift} and s_{scale} are difficult to determine precisely. Due to this low and central-q^2 signal modes are fitted simultaneously with the relevant J/ψ mode, which is not subject to mass constraints, sharing m_{shift} and s_{scale} between them. This improves the quality and stability of the non-resonant mode fits. Note the m_{shift} and s_{scale} are only shared between modes with the same leptonic final state.

In addition the $B^+ \to \pi^+ J/\psi (\to \ell^+\ell^-)$ misID backgrounds lie beneath the more abundant $B^+ \to K^+ J/\psi (\to \ell^+\ell^-)$ modes so the m_{shift} and s_{scale} are shared between both to further stabilise the fitter.

7.8 Normalisation Constraints

When background yields are fitted directly in data and allowed to float they can have large uncertainties, these are nuisance parameters which can cause fit instabilities and enlarge other uncertainties. Where possible these backgrounds can be constrained, usually as a function of the relevant signal mode yield and ratios of efficiencies. The following sections describe how background yield constraints are calculated for various signal modes of interest.

7.8.1 $B^0 \to K^{*0} J/\psi (\to \ell^+\ell^-)$

The following constraints are considered:

- $\Lambda_b^0 \to pK J/\psi (\to \ell^+\ell^-)$: The normalisation for these backgrounds can be parametrised as a function of the control mode yield, $\mathcal{N}_{B^0 \to K^{*0} J/\psi (\to \ell^+\ell^-)}$ the ratio of the hadronisation fractions of b-quarks, f_d, and Λ_b^0 baryons, f_{Λ_b}, the ratios of the branching fractions of the background, $\mathcal{B}(\Lambda_b^0 \to pK J/\psi (\to \ell^+\ell^-))$, to the control mode, $\mathcal{B}(B^0 \to K^{*0} J/\psi (\to \ell^+\ell^-))$ and the ratio of their selection efficiencies, $\frac{\varepsilon_{background}}{\varepsilon_{control}}$ as seen in Eq. 7.4. The ratio $\frac{f_{\Lambda_b}}{f_d}$ is extracted from Ref. [138] assuming isospin symmetry. The ratios and efficiencies are subject

to Gaussian constraints.

$$\mathcal{N}^{J/\psi}_{\Lambda_b^0} = \frac{f_{\Lambda_b}}{f_d} \cdot \frac{\mathcal{B}(\Lambda_b^0 \to pKJ/\psi(\to \ell^+\ell^-))}{\mathcal{B}(B^0 \to K^{*0}J/\psi(\to \ell^+\ell^-))} \cdot \frac{\varepsilon_{background}}{\varepsilon_{control}} \cdot \mathcal{N}_{B^0 \to K^{*0}J/\psi(\to \ell^+\ell^-)} \quad (7.4)$$

- $B_s^0 \to K^{*0}J/\psi(\to \ell^+\ell^-)$: The normalisation for these backgrounds can be parametrised by the control mode yield, the ratio of the hadronisation fractions of s quarks and b quarks and the ratio of branching fraction of the background to the control mode as seen in Eq. 7.5. The ratios are subject to Gaussian constraints.

$$\mathcal{N}^{J/\psi}_{B_s^0} = \frac{f_s}{f_d} \cdot \frac{\mathcal{B}(B_s^0 \to K^{*0}J/\psi(\to \ell^+\ell^-))}{\mathcal{B}(B^0 \to K^{*0}J/\psi(\to \ell^+\ell^-))} \cdot \mathcal{N}_{B^0 \to K^{*0}J/\psi(\to \ell^+\ell^-)} \quad (7.5)$$

7.8.2 $B^0 \to K^{*0}e^+e^-$

- Leakage from $B^0 \to K^{*0}J/\psi(\to e^+e^-)$ (central-q^2 only): The normalisation for this background can be parametrised by the control mode yield and the ratio of the selection efficiencies for this background in the central-q^2 and J/ψ q^2 bins as shown in Eq. 7.6. The efficiencies are subject to Gaussian constraints.

$$\mathcal{N}_{Leakage} = \frac{\varepsilon^{central-q^2}_{B^0 \to K^{*0}J/\psi(\to e^+e^-)}}{\varepsilon^{J/\psi\, q^2}_{B^0 \to K^{*0}J/\psi(\to e^+e^-)}} \cdot \mathcal{N}_{B^0 \to K^{*0}J/\psi(\to e^+e^-)} \quad (7.6)$$

7.8.3 $B^+ \to K^+J/\psi(\to \ell^+\ell^-)$

- $B^+ \to \pi^+J/\psi(\to \ell^+\ell^-)$: The normalisation for this background can be parametrised by the control mode yield, the ratio of the branching fractions and the selection efficiency ratio as can be seen in Eq. 7.7. The ratios and efficiencies are subject to Gaussian constraints.

$$\mathcal{N}_{B^+ \to \pi^+J/\psi(\to \ell^+\ell^-)} = \frac{\mathcal{B}(B^+ \to \pi^+J/\psi(\to \ell^+\ell^-))}{\mathcal{B}(B^+ \to K^+J/\psi(\to \ell^+\ell^-))} \cdot \frac{\varepsilon_{background}}{\varepsilon_{control}} \cdot \mathcal{N}_{B^+ \to K^+J/\psi(\to \ell^+\ell^-)} \quad (7.7)$$

7.8.4 $B^+ \to K^+e^+e^-$

- Leakage from $B^+ \to K^+J/\psi(\to e^+e^-)$ (central-q^2 only): This background can be parametrised by the control mode yield and the selection efficiencies of

$B^+ \to K^+ J/\psi\,(\to e^+e^-)$ in the central-q^2 and J/ψ q^2 bins as shown in Eq. 7.8. The efficiencies are subject to Gaussian constraints.

$$\mathcal{N}_{Leakage} = \frac{\varepsilon^{central-q^2}_{B^+\to K^+ J/\psi\,(\to e^+e^-)}}{\varepsilon^{J/\psi\,q^2}_{B^+\to K^+ J/\psi\,(\to e^+e^-)}} \cdot \mathcal{N}_{B^+\to K^+ J/\psi\,(\to e^+e^-)} \tag{7.8}$$

- Partially reconstructed $B^0 \to K^{*0} e^+e^-$: The determination of the normalisation for this background and the computation of factors relevant to this calculation are discussed in detail in Appendix K.

- Partially reconstructed $B^0 \to K^+\pi^- e^+e^-$: The determination of the normalisation for this background is also discussed in Appendix K.

7.9 Fits to collision data

Fits are performed to extract the yields in Run 1 L0I, Run 1 L0L!, Run 2p1 L0I, Run 2p1 L0L!, Run 2p2 L0I and Run 2p2 L0L!. These six exclusive samples are fitted independently. Fits to the non-resonant modes in Run 2p1, which are also fitted simultaneously with J/ψ resonant modes without mass constraints can be seen in Figs. 7.19- 7.20. Fits to the mass-constrained J/ψ resonant modes can be seen in Figs. 7.21-7.22. Yields from the simultaneous fits to the J/ψ and non-resonant modes can be seen in Tab. 7.1, the yields from the mass-constrained fits for the J/ψ modes are consistent with these values.

7.10 Fitter validation with pseudoexperiments

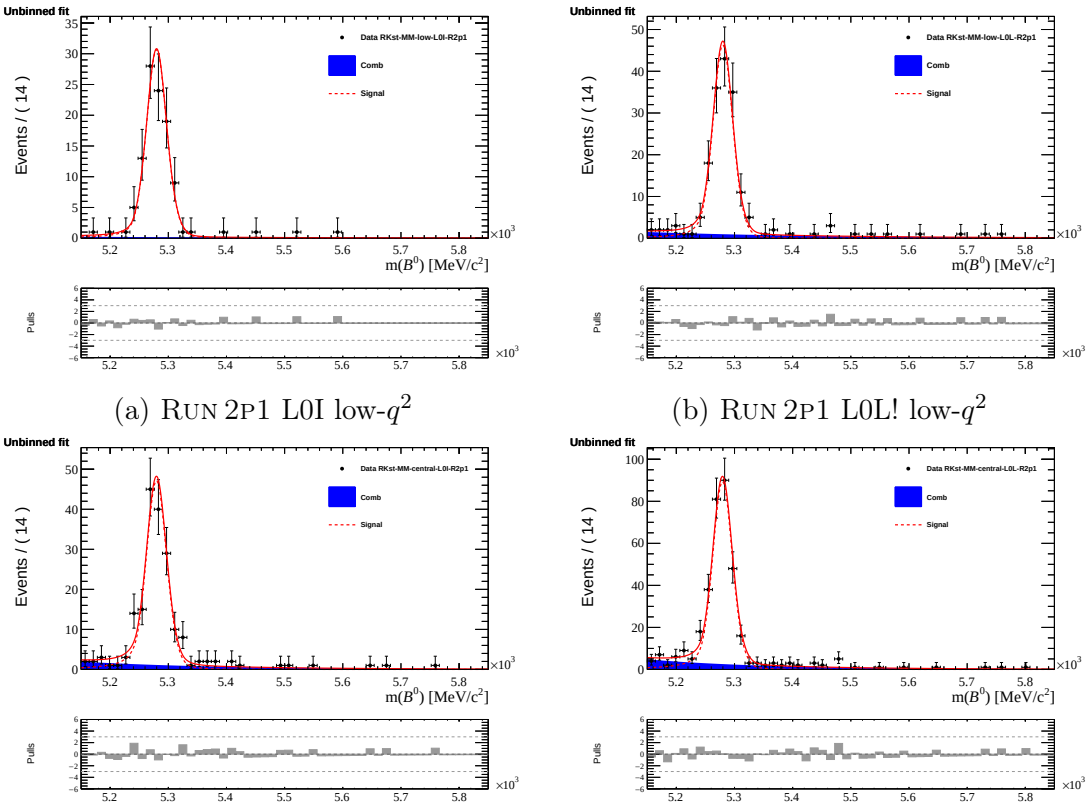
A pseudoexperiment generator was implemented with RooFit which uses the converged fit model PDF as its seed. The outputs can be stored as tuples with each signal and background component able to be stored separately. This allows the user to mix and match background components, or to vary the yields of the individual components. The data generated are then fed back into the same fitting configuration as was used to produce it.

The pseudoexperiments are used to validate the fit stability and to check for bias. These pseudoexperiments are run with the fits in the previous section as seeds and 1000 such pseudoexperiments are generated for each of the six cases. A particular pseudoexperiment is classed as failed if `MIGRAD` or `HESSE` which perform minimisation

final state	q^2bin	lepton	year	L0I	L0L!
$K^{*0}l^+l^-$	J/ψ	ee	Run 1	20850 ± 246	22077 ± 245
			15 + 16	37118 ± 330	39767 ± 331
		$\mu\mu$	Run 1	89620 ± 318	226623 ± 504
			15 + 16	93128 ± 322	197880 ± 467
	low	ee	Run 1	28 ± 7	33 ± 7
			15 + 16	44 ± 8	36 ± 7
		$\mu\mu$	Run 1	111 ± 11	211 ± 16
			15 + 16	119 ± 11	180 ± 14
	central	ee	Run 1	55 ± 10	49 ± 9
			15 + 16	58 ± 11	67 ± 10
		$\mu\mu$	Run 1	157 ± 14	337 ± 20
			15 + 16	181 ± 15	330 ± 20
$K^+l^+l^-$	J/ψ	ee	Run 1	75545 ± 511	76048 ± 492
			15 + 16	94105 ± 548	104075 ± 536
		$\mu\mu$	Run 1	246277 ± 527	660151 ± 860
			15 + 16	242810 ± 520	546319 ± 777
	low	ee	Run 1	41 ± 8	48 ± 8
			15 + 16	54 ± 10	38 ± 8
		$\mu\mu$	Run 1	106 ± 12	272 ± 18
			15 + 16	98 ± 11	227 ± 17
	central	ee	Run 1	161 ± 18	138 ± 15
			15 + 16	172 ± 18	172 ± 16
		$\mu\mu$	Run 1	386 ± 23	991 ± 35
			15 + 16	394 ± 23	840 ± 32

Table 7.1: Signal yields measured in simultaneous fit. The yields of the J/ψ modes are measured without applying the J/ψ DTF mass constraint and suffer from Bremsstrahlung radiation.

in the fitter fail. Pull distributions of fits which do not fail for each of the floating parameters are fitted with a Gaussian function. If the fit is unbiased the pull distribution should have a mean of zero and a standard deviation of unity. A table detailing the failure rate of Run 1 and Run 2p1 pseudoexperiments can be seen in Tab. 7.2, which shows that the fit configurations are quite stable particularly in Run 2p1. The pulls on the signal yields for Run 1 and Run 2p1 can be seen in Figs. 7.23–7.26, no large biases are observed.



(a) Run 2p1 L0I low-q^2

(b) Run 2p1 L0L! low-q^2

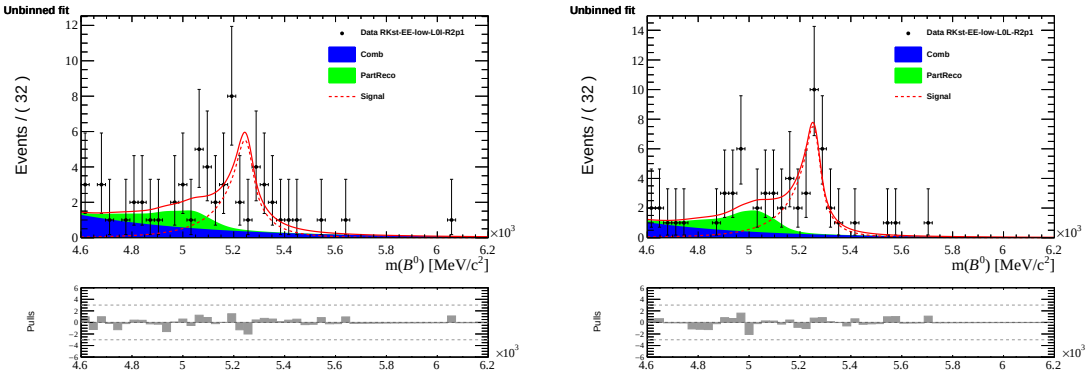
(c) Run 2p1 L0I central-q^2

(d) Run 2p1 L0L! central-q^2

Figure 7.19: (top) Fits to low-q^2 bin data reconstructed as $B^0 \to K^{*0}\mu^+\mu^-$ for Run 2p1 separated into L0I (left) and L0L! (right) trigger categories. (bottom) Fits to central-q^2 bin data reconstructed as $B^0 \to K^{*0}\mu^+\mu^-$ for Run 2p1 separated into L0I (left) and L0L! (right) trigger categories. The model in both cases consists of signal $B^0 \to K^{*0}\mu^+\mu^-$ decays (dashed red line) and a combinatorial background (blue).

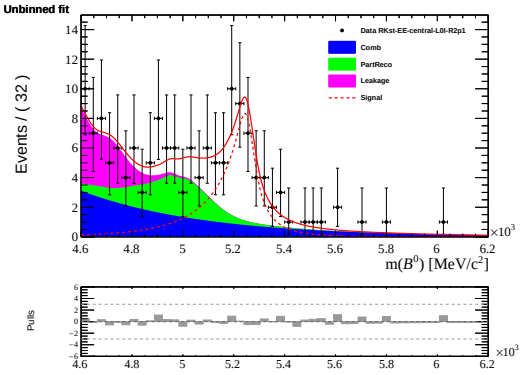
Configuration	Splits	Failure Rate
Yields	Run 1 L0I	5.8%
	Run 1 L0L!	5.4%
	Run 2p1 L0I	1.5%
	Run 2p1 L0L!	3.1%

Table 7.2: Failure rates of 1000 pseudoexperiments generated from fits to data and fitted with the same models.

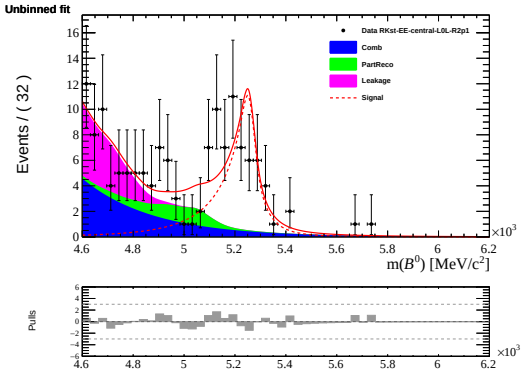


(a) Run 2p1 L0I low-q^2

(b) Run 2p1 L0L! low-q^2

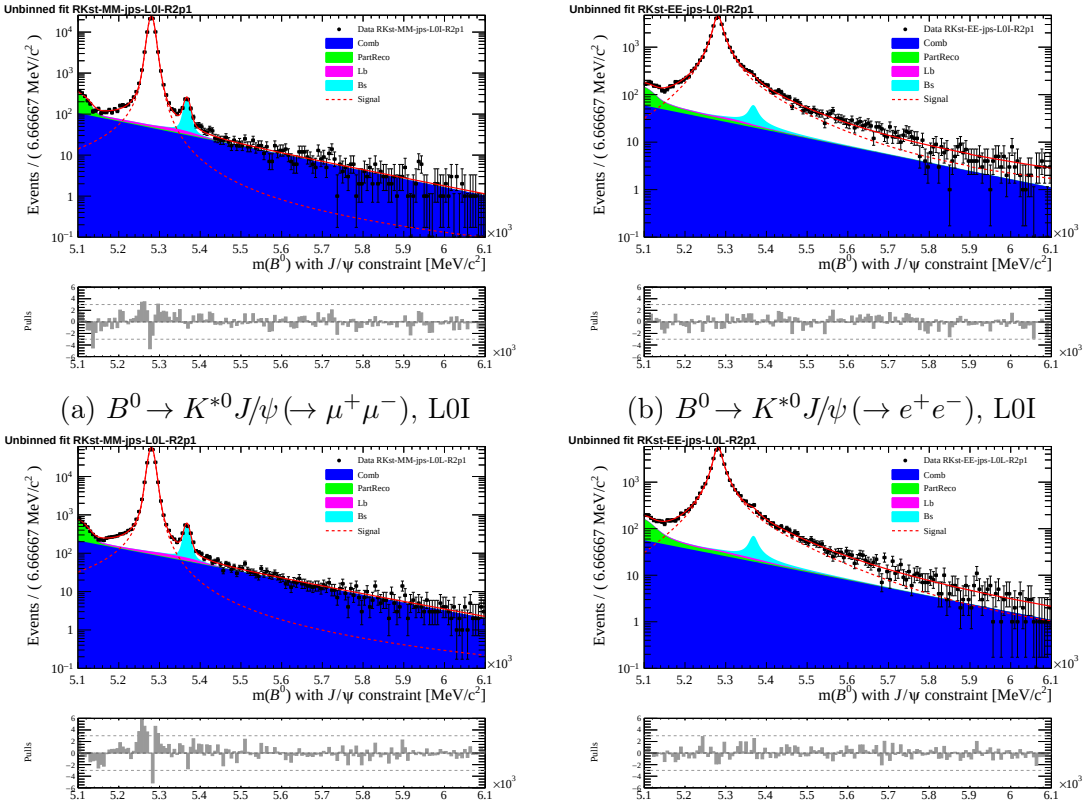


(c) Run 2p1 L0I central-q^2



(d) Run 2p1 L0L! central-q^2

Figure 7.20: (top) Fits to low-q^2 bin data reconstructed as $B^0 \to K^{*0}e^+e^-$ for Run 2p1 separated into L0I (left) and L0L! (right) trigger categories. (bottom) Fits to central-q^2 bin data reconstructed as $B^0 \to K^{*0}e^+e^-$ for Run 2p1 separated into L0I (left) and L0L! (right) trigger categories. The model in both cases consists of signal $B^0 \to K^{*0}e^+e^-$ decays (dashed red line) and a combinatorial background (blue).



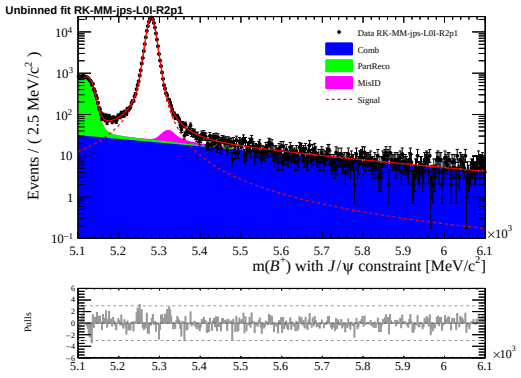
(a) $B^0 \to K^{*0} J/\psi\,(\to \mu^+\mu^-)$, L0I

(b) $B^0 \to K^{*0} J/\psi\,(\to e^+e^-)$, L0I

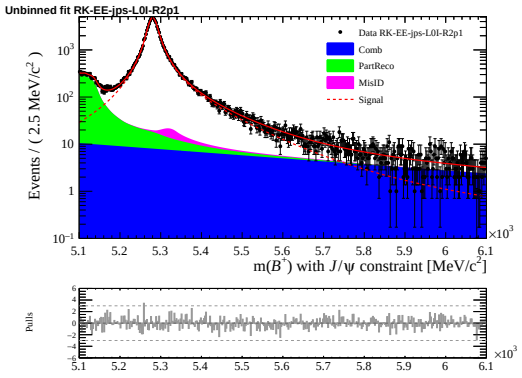
(c) $B^0 \to K^{*0} J/\psi\,(\to \mu^+\mu^-)$, L0L!

(d) $B^0 \to K^{*0} J/\psi\,(\to e^+e^-)$, L0L!

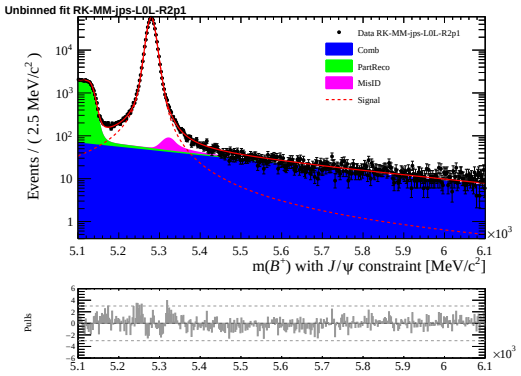
Figure 7.21: Run 2p1 fits to data reconstructed as $B^0 \to K^{*0} J/\psi\,(\to \mu^+\mu^-)$ (left) and $B^0 \to K^{*0} J/\psi\,(\to e^+e^-)$ (right) in the L0I (top) and L0L! (bottom). The DTF constrains the dilepton invariant mass to the J/ψ mass. Electron and muon modes are modelled with the same components: signal $B^0 \to K^{*0} J/\psi\,(\to \ell^+\ell^-)$ (dashed red line), $B_s^0 \to K^{*0} J/\psi\,(\to \ell^+\ell^-)$ decays (light blue) that decays into the same final state, partially-reconstructed $B \to X J/\psi\,(\to \ell^+\ell^-)$ background (green) modelled with inclusive simulation samples, mis-identified $\Lambda_b^0 \to pK J/\psi\,(\to \ell^+\ell^-)$ decays (purple), and a combinatorial background (blue).



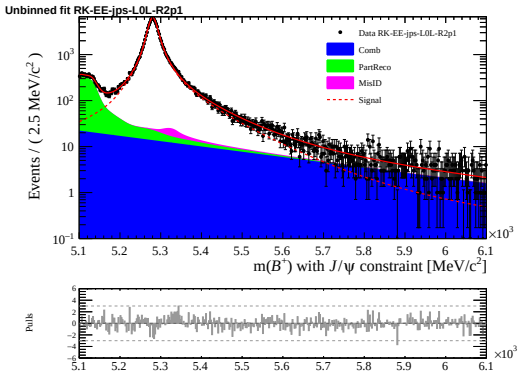
(a) $B^+ \to K^+ J/\psi\,(\to \mu^+\mu^-)$, L0I



(b) $B^+ \to K^+ J/\psi\,(\to e^+e^-)$, L0I

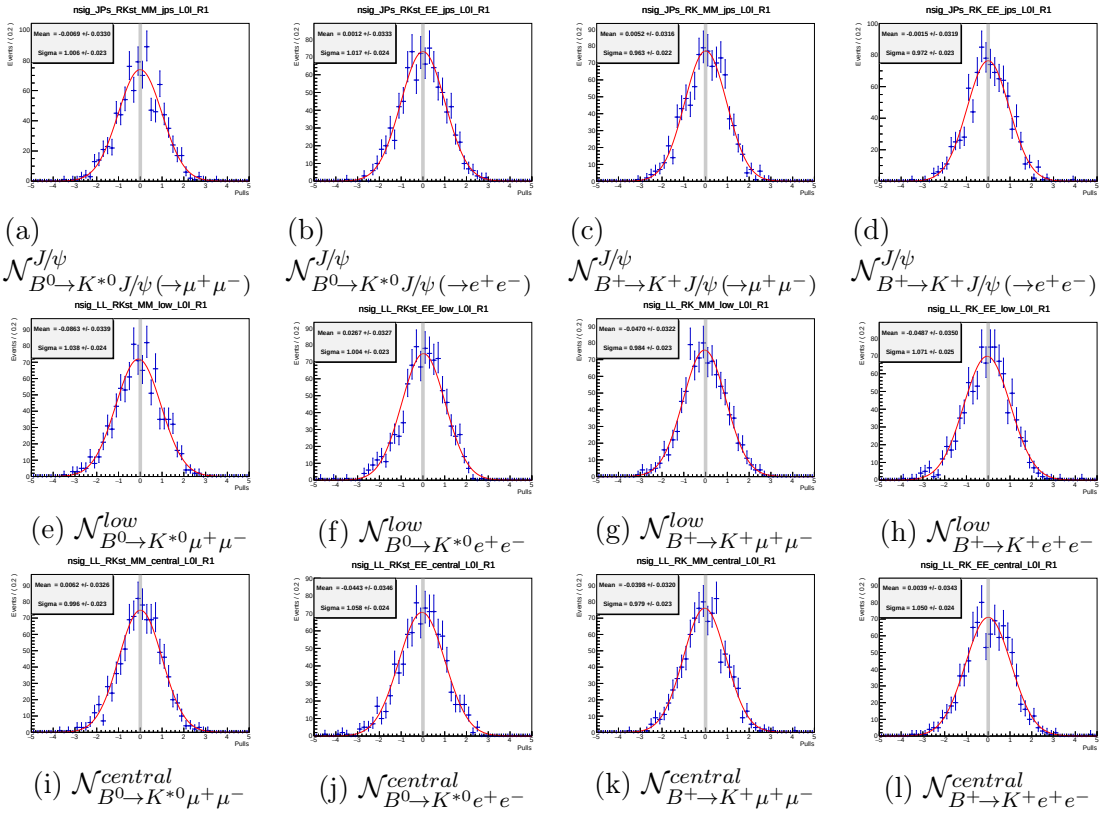


(c) $B^+ \to K^+ J/\psi\,(\to \mu^+\mu^-)$, L0L!



(d) $B^+ \to K^+ J/\psi\,(\to e^+e^-)$, L0L!

Figure 7.22: Run 2p1 fits to data reconstructed as $B^+ \to K^+ J/\psi\,(\to \mu^+\mu^-)$ (left) and $B^+ \to K^+ J/\psi\,(\to e^+e^-)$ (right) in the L0I (top) and L0L! (bottom). The DTF constrains the dilepton invariant mass to the J/ψ mass. Electron and muon modes are modelled with the same components: signal $B^+ \to K^+ J/\psi\,(\to \ell^+\ell^-)$ (dashed red line), partially-reconstructed $B \to X J/\psi\,(\to \ell^+\ell^-)$ background (green) modelled with inclusive simulation samples, mis-identified $B^+ \to \pi^+ J/\psi\,(\to \ell^+\ell^-)$ decays (purple), and a combinatorial background (blue).

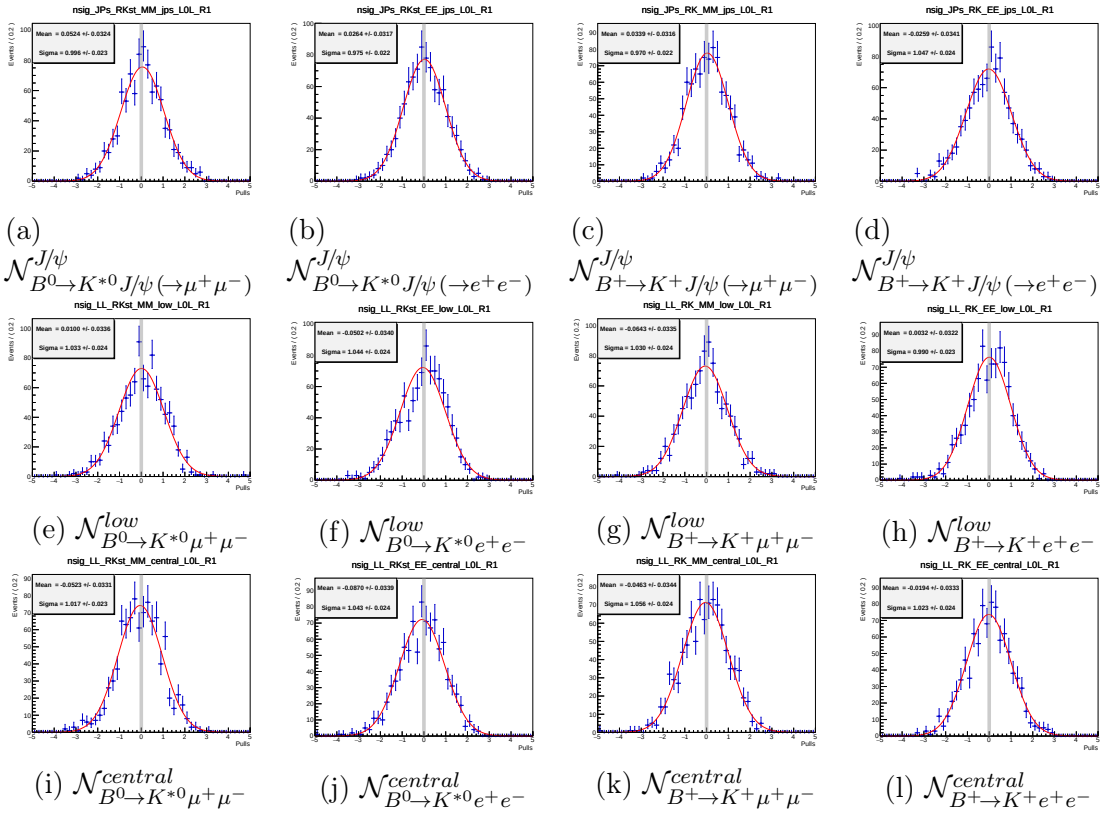


(a) $\mathcal{N}^{J/\psi}_{B^0\to K^{*0}J/\psi(\to\mu^+\mu^-)}$ (b) $\mathcal{N}^{J/\psi}_{B^0\to K^{*0}J/\psi(\to e^+e^-)}$ (c) $\mathcal{N}^{J/\psi}_{B^+\to K^+J/\psi(\to\mu^+\mu^-)}$ (d) $\mathcal{N}^{J/\psi}_{B^+\to K^+J/\psi(\to e^+e^-)}$

(e) $\mathcal{N}^{low}_{B^0\to K^{*0}\mu^+\mu^-}$ (f) $\mathcal{N}^{low}_{B^0\to K^{*0}e^+e^-}$ (g) $\mathcal{N}^{low}_{B^+\to K^+\mu^+\mu^-}$ (h) $\mathcal{N}^{low}_{B^+\to K^+e^+e^-}$

(i) $\mathcal{N}^{central}_{B^0\to K^{*0}\mu^+\mu^-}$ (j) $\mathcal{N}^{central}_{B^0\to K^{*0}e^+e^-}$ (k) $\mathcal{N}^{central}_{B^+\to K^+\mu^+\mu^-}$ (l) $\mathcal{N}^{central}_{B^+\to K^+e^+e^-}$

Figure 7.23: Pull distribution of signal yields from 1000 pseudoexperiments in the RUN 1 L0I configuration. The signal yields are fitted to $K^{*0}\mu\mu$ (leftmost), $K^{*0}ee$ (2nd from left), $K\mu\mu$ (2nd from right) and Kee (rightmost) final states in $J/\psi - q^2$ (top), low-q^2 (middle) and central-q^2 (bottom).

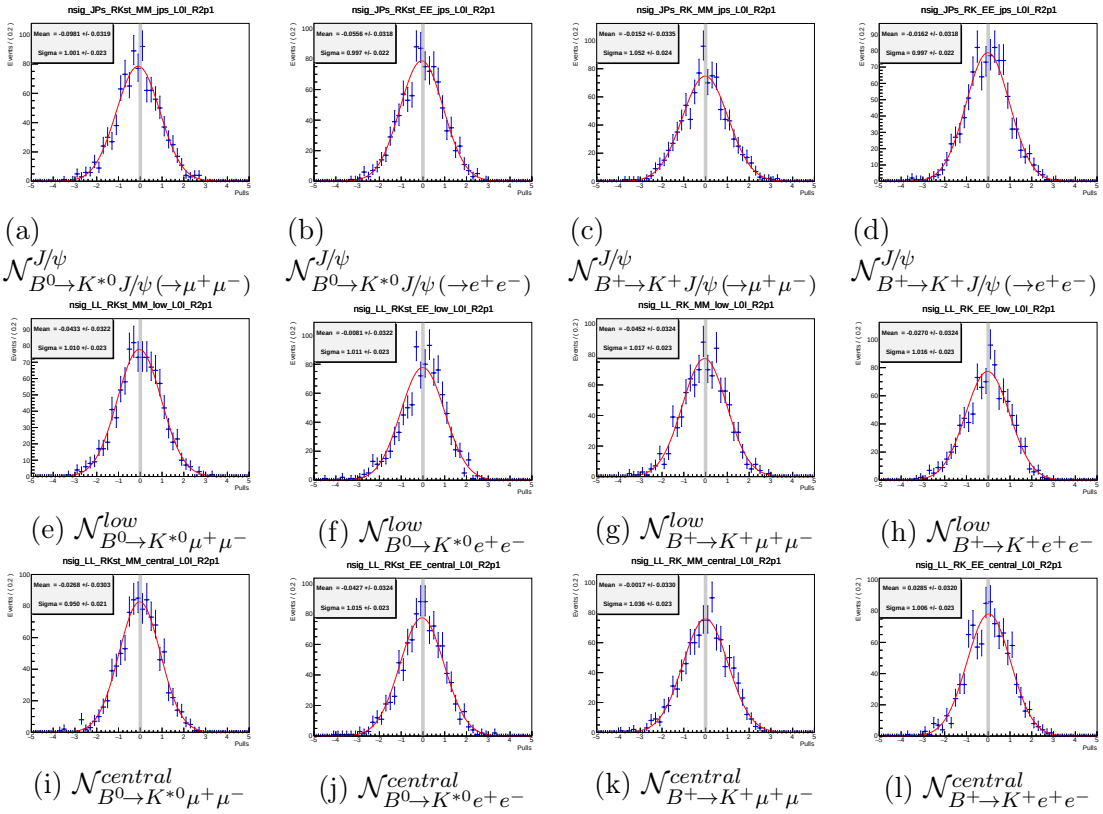


(a) $\mathcal{N}^{J/\psi}_{B^0\to K^{*0}J/\psi(\to\mu^+\mu^-)}$ (b) $\mathcal{N}^{J/\psi}_{B^0\to K^{*0}J/\psi(\to e^+e^-)}$ (c) $\mathcal{N}^{J/\psi}_{B^+\to K^+J/\psi(\to\mu^+\mu^-)}$ (d) $\mathcal{N}^{J/\psi}_{B^+\to K^+J/\psi(\to e^+e^-)}$

(e) $\mathcal{N}^{low}_{B^0\to K^{*0}\mu^+\mu^-}$ (f) $\mathcal{N}^{low}_{B^0\to K^{*0}e^+e^-}$ (g) $\mathcal{N}^{low}_{B^+\to K^+\mu^+\mu^-}$ (h) $\mathcal{N}^{low}_{B^+\to K^+e^+e^-}$

(i) $\mathcal{N}^{central}_{B^0\to K^{*0}\mu^+\mu^-}$ (j) $\mathcal{N}^{central}_{B^0\to K^{*0}e^+e^-}$ (k) $\mathcal{N}^{central}_{B^+\to K^+\mu^+\mu^-}$ (l) $\mathcal{N}^{central}_{B^+\to K^+e^+e^-}$

Figure 7.24: Pull distribution of signal yields from 1000 pseudoexperiments in the Run 1 L0L! configuration. The signal yields are fitted to $K^{*0}\mu\mu$ (leftmost), $K^{*0}ee$ (2nd from left), $K\mu\mu$ (2nd from right) and Kee (rightmost) final states in $J/\psi - q^2$ (top), low-q^2 (middle) and central-q^2 (bottom).

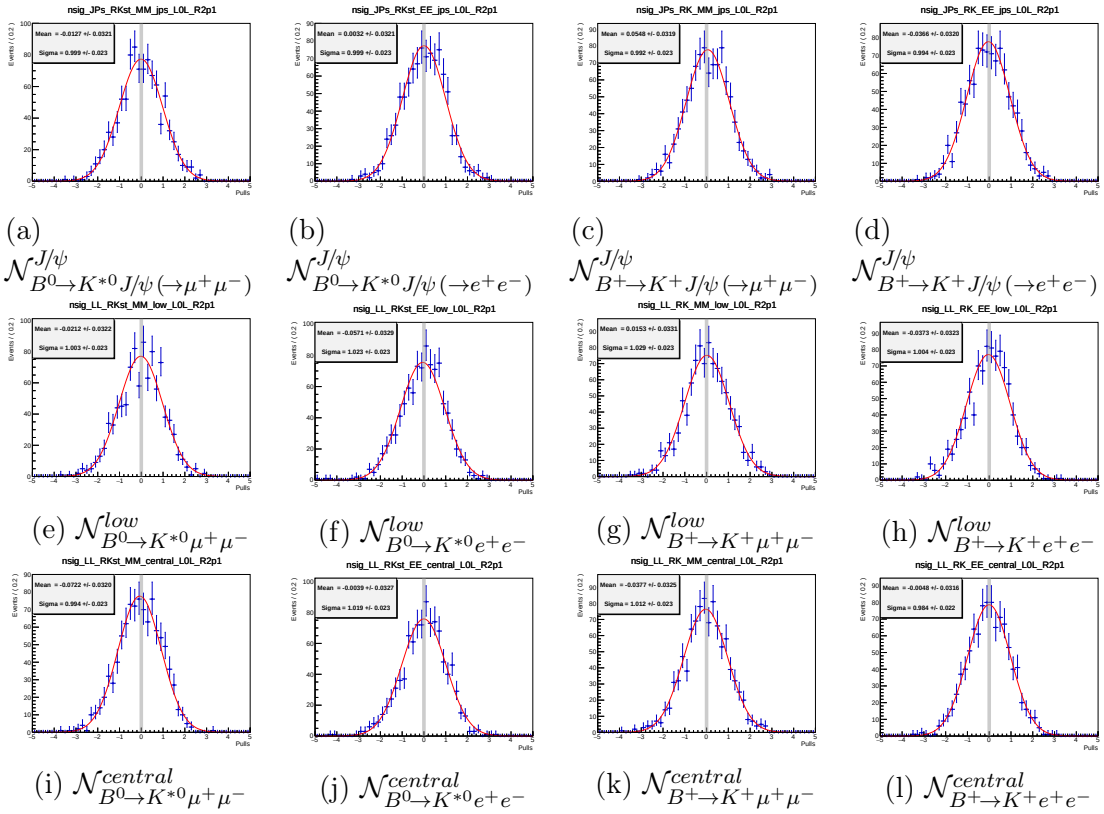


(a) $\mathcal{N}^{J/\psi}_{B^0\to K^{*0}J/\psi(\to\mu^+\mu^-)}$ (b) $\mathcal{N}^{J/\psi}_{B^0\to K^{*0}J/\psi(\to e^+e^-)}$ (c) $\mathcal{N}^{J/\psi}_{B^+\to K^+J/\psi(\to\mu^+\mu^-)}$ (d) $\mathcal{N}^{J/\psi}_{B^+\to K^+J/\psi(\to e^+e^-)}$

(e) $\mathcal{N}^{low}_{B^0\to K^{*0}\mu^+\mu^-}$ (f) $\mathcal{N}^{low}_{B^0\to K^{*0}e^+e^-}$ (g) $\mathcal{N}^{low}_{B^+\to K^+\mu^+\mu^-}$ (h) $\mathcal{N}^{low}_{B^+\to K^+e^+e^-}$

(i) $\mathcal{N}^{central}_{B^0\to K^{*0}\mu^+\mu^-}$ (j) $\mathcal{N}^{central}_{B^0\to K^{*0}e^+e^-}$ (k) $\mathcal{N}^{central}_{B^+\to K^+\mu^+\mu^-}$ (l) $\mathcal{N}^{central}_{B^+\to K^+e^+e^-}$

Figure 7.25: Pull distribution of signal yields from 1000 pseudoexperiments in the Run 2p1 L0I configuration. The signal yields are fitted to $K^{*0}\mu\mu$ (leftmost), $K^{*0}ee$ (2nd from left), $K\mu\mu$ (2nd from right) and Kee (rightmost) final states in $J/\psi - q^2$ (top), low-q^2 (middle) and central-q^2 (bottom).



(a) $\mathcal{N}^{J/\psi}_{B^0\to K^{*0}J/\psi(\to\mu^+\mu^-)}$ (b) $\mathcal{N}^{J/\psi}_{B^0\to K^{*0}J/\psi(\to e^+e^-)}$ (c) $\mathcal{N}^{J/\psi}_{B^+\to K^+J/\psi(\to\mu^+\mu^-)}$ (d) $\mathcal{N}^{J/\psi}_{B^+\to K^+J/\psi(\to e^+e^-)}$

(e) $\mathcal{N}^{low}_{B^0\to K^{*0}\mu^+\mu^-}$ (f) $\mathcal{N}^{low}_{B^0\to K^{*0}e^+e^-}$ (g) $\mathcal{N}^{low}_{B^+\to K^+\mu^+\mu^-}$ (h) $\mathcal{N}^{low}_{B^+\to K^+e^+e^-}$

(i) $\mathcal{N}^{central}_{B^0\to K^{*0}\mu^+\mu^-}$ (j) $\mathcal{N}^{central}_{B^0\to K^{*0}e^+e^-}$ (k) $\mathcal{N}^{central}_{B^+\to K^+\mu^+\mu^-}$ (l) $\mathcal{N}^{central}_{B^+\to K^+e^+e^-}$

Figure 7.26: Pull distribution of signal yields from 1000 pseudoexperiments in the Run 2p1 L0L! configuration. The signal yields are fitted to $K^{*0}\mu\mu$ (leftmost), $K^{*0}ee$ (2nd from left), $K\mu\mu$ (2nd from right) and Kee (rightmost) final states in $J/\psi - q^2$ (top), low-q^2 (middle) and central-q^2 (bottom).

Chapter 8

Systematics

The systematics associated with this analysis can be split into three main areas:

- Ensembles of pseudoexperiments, such as those discussed in Sec. 7.10, can be used to determine systematics associated with the fit configuration. Variations in the results of single fits can be used for systematics, which are negligibly small.

- The methods to obtain the various simulation corrections can be varied. This can include using alternative binning schemes or changing the parameters of the BDTs used in the analysis.

- The residual non-flatness of $r^{K}_{J/\psi}$ and $r^{K^*}_{J/\psi}$ can be used as a closure test to compute a systematic uncertainty that covers aspects of the analysis not covered by the other systematic tests.

It should be noted that the analysis has only recently progressed to the stage of evaluating systematics and as such the only systematics which have been fully implemented are those related to the bootstrapping of the L0 and HLT corrections. The systematics related to these corrections are expected to be dominant since they are extracted from the smallest samples. The rest of the systematics outlined here will be evaluated in the coming months.

8.1 Uncertainty due to simulation corrections

8.1.1 Uncertainty due to samples used to extract corrections

A systematic is assigned for the known limitations of extracting corrections to simulation from J/ψ resonant samples. Correlations should also be taken into account, although this is not yet done, since the J/ψ resonant samples are used to extract yields but also used to compute corrections to simulation. The uncertainty is obtained used the bootstrapping method, described in Sec. 6.1. So far the bootstrapping has been applied to the L0 and HLT trigger correction steps only but should be used for all of the other correction steps, except the tracking which is external to this analysis. Correction maps are produced using the TISTOS method for each of the 100 Poisson weighted samples. The ratio of the i^{th} data and i^{th} simulation fits to the efficiency turn-on curves gives the i^{th} L0 weight map. Example L0 efficiency curve fits and weight maps can be seen in Fig. 8.1. For the HLT correction the same procedure is followed, fits are performed to the i^{th} collision data sample to extract the efficiency and a simple ratio is performed for the i^{th} simulation efficiency. Efficiencies of each of the 100 samples for collision data and simulation can be seen in Figs. 8.2 and 8.3 respectively for the L0L! trigger, the L0I trigger shows similar results.

It should also be noted that in a number of cases it has been shown that a negligible systematic is needed when using different samples to extract the corrections such as:

L0 corrections - It was shown in Sec. 6.5 that no systematic needed to be applied when taking the corrections for the L0I category from a muon selected sample or an electron selected sample since the corrections derived from the two were in agreement, although a systematic related to correlations does need to be evaluated here. It was also shown for each category all sensible choices of tag were compatible with each other so a systematic related to the choice of tag does not need to be added.

HLT corrections - It was noted in Sec. 6.6 that no systematic needed to be applied if the corrections were taken from L0L! or L0L which had compatible corrections, although a systematic related to correlations due to overlaps needs to be evaluated here. Due to reasons outlined in Sec. 6.6 it was not possible to compare different tags as for the L0 case.

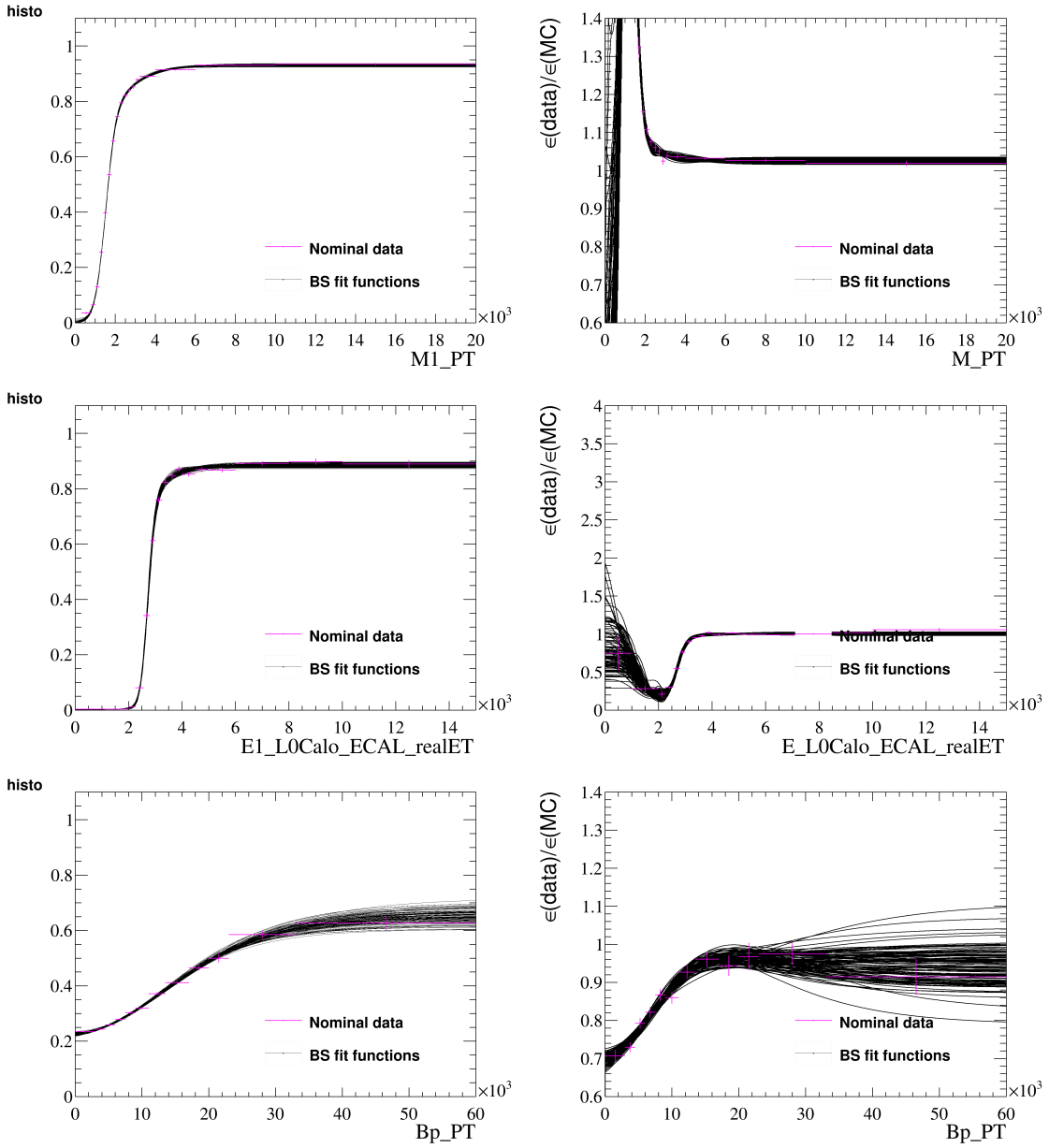


Figure 8.1: Data efficiency (left) and data over simulation efficiency ratio (right) of the `L0Muon`, `L0Electron` and `L0Global_TIS` lines from top to bottom (example plots from 2016 R_K). The fits to the 100 Poisson weight sets are shown in black, the nominal data points are shown in magenta.

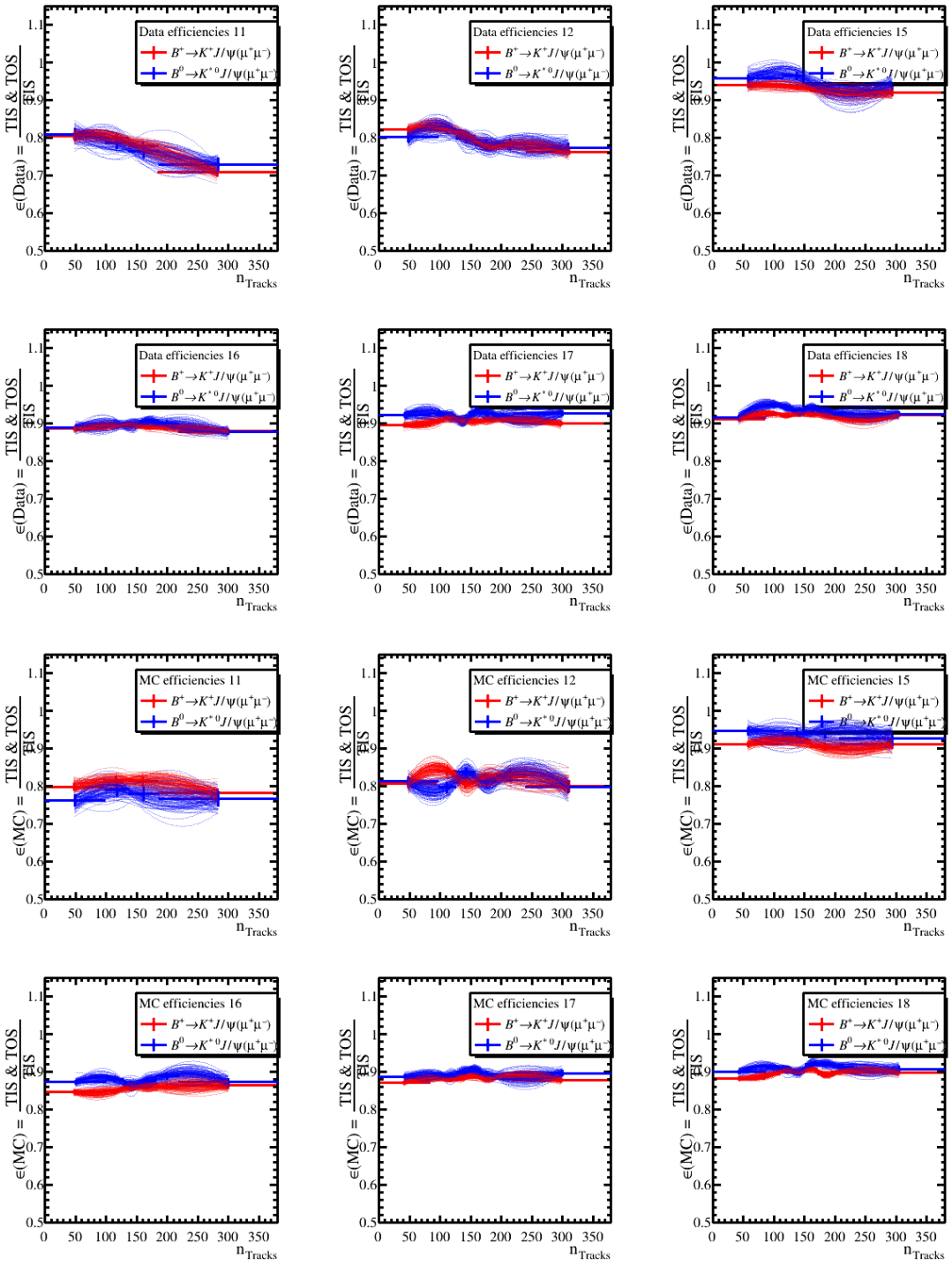


Figure 8.2: Data HLT efficiencies (top half) and simulation HLT efficiencies (bottom half) on L0L categories in all years using the 100 Poisson weight sets for muon modes. Lines across bin centres are drawn for the Poisson smeared efficiencies in data. Nominal HLT efficiencies in data are shown with solid lines and markers.

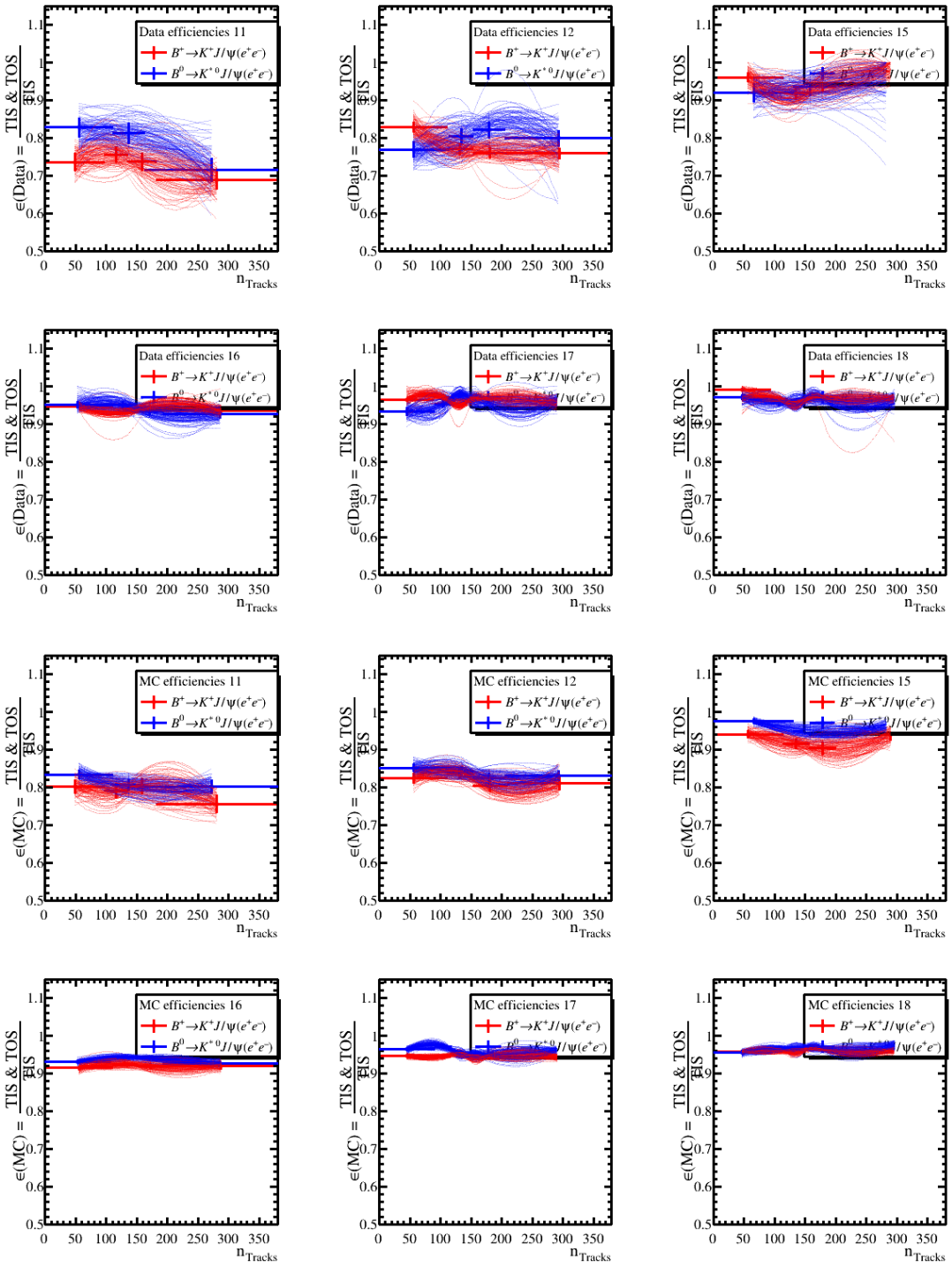


Figure 8.3: Data (top half) and simulation (bottom half) HLT efficiencies on L0L categories in all years using the 100 Poisson weight sets for electron modes. Lines across bin centres are drawn for the Poisson smeared efficiencies in simulation. Nominal HLT efficiencies in simulation points are shown with solid lines and markers.

BDT corrections - It was noted in Sec. 5.3 that the prior chain correction was also computed using L0I instead of L0M and the two gave compatible results so no systematic needs to be added for the choice of sample used to extract the kinematic and multiplicity corrections, however a systematic related to correlations does need to be evaluated.

8.1.2 Uncertainty due to PID corrections

There are a number of systematics that can be evaluated related to the PID corrections concerning the binning scheme used to extract the corrections, the ${}_s\mathcal{P}lot$ procedure, size of the calibration samples and the independence of single track efficiencies.

When computing the PID corrections central calibration samples are used, however, these samples are limited in size and will also have some overlap with $B^+ \to K^+ J/\psi (\to \ell^+\ell^-)$ samples to which the corrections are then applied. To account for the uncertainty due to this a bootstrapping procedure much like that described in the above section must be implemented. The per candidate efficiencies taken from the 100 produced weight maps should then be Gaussianly distributed around their mean values.

An important assumption of the PID correction procedure is that the efficiency of each of the single track types is independent. That is to say:

$$\epsilon = \frac{N(\text{All cuts passed})}{N(\text{total})} = \frac{(N(\text{Pass cut 1}) \cdot N(\text{Pass cut 2}) \cdot ...N(\text{Pass cut n})}{N(\text{total})^n} \quad (8.1)$$

any difference in the the efficiency between the two formulations can be assigned as a systematic.

The calibration samples are subject to a ${}_s\mathcal{P}lot$ procedure in order to remove any residual backgrounds in the samples, the recommended systematic to assign to using this method is 0.2% per track. Also when computing the corrections to capture the efficiency trends a binning optimisation is performed, as described in Sec. 6.4. In order to see how the binning schemes affect the PID corrections, different sensible binning schemes can be used to determine a systematic uncertainty for the binning scheme used. How the ${}_s\mathcal{P}lot$ and binning systematics are handled can be factorised per track type:

K/π **tracks** The PID corrections can be computed with different binning schemes using the standard PIDCalib procedure, at least two different binning schemes

are planned to be used. When applying the corrections derived from the different binning schemes the systematic will be taken as the largest difference on the value of $r_{J/\psi}$ with respect to the KDE approach. Since the hadrons are always present on both sides of the various ratios in the analysis the systematic related to the ${}_s\mathcal{P}lot$ procedure is expected to cancel here.

μ **tracks** The same approach is used as for the hadrons to determine the uncertainty due to the binning scheme. A 0.2% systematic is added per muon track due to the ${}_s\mathcal{P}lot$ approach since this will not cancel for leptonic cases.

e **tracks** The same approach as for the other track types is used to determine the uncertainty due to the binning scheme, here a KDE was not used so the things are compared to the default binning scheme. In order to evaluate the systematic due to the ${}_s\mathcal{P}lot$ procedure the difference in $r_{J/\psi}$ when using fit and count or the ${}_s\mathcal{W}eights$ is taken as the uncertainty.

8.1.3 Uncertainty due to BDT reweighting

A systematic will need to be assigned to the BDT reweighting of data, in addition to bootstrapping, there are a number of checks that will need to be performed:

- The hyperparameters of the BDT can be varied,
- Different proxies to correct for the multiplicity can be used

In order to evaluate these systematics the simulated efficiencies and background shapes would be extracted with the new weight sets associated to the checks above and evaluated as for the efficiency systematic.

8.2 Uncertainty due to fitting procedure

8.2.1 Uncertainty due to fit bias

Pseudoexperiments are generated and fitted, described in Sec. 7.10, to verify the fit stability and to evaluate bias. A bias is considered significant if there is a 3σ deviation in the pull distribution mean from 0 or standard deviation from unity. Any bias in the mean will be corrected for and the bias added to the systematic uncertainty for that parameter. Any bias in the standard deviation can be corrected for in the covariance matrix of the data fit, but no additional systematic will be

added.

8.2.2 Uncertainty due to alternate fit setups

In order to determine the uncertainty associated to performing fits with an alternate fit setup pseudoexperiments are fitted twice, once with the nominal fit model and once with the alternate fit model. Doing so two parameters of interest can be extracted, the signal yield of the nominal fit, y^{nom}, and the signal yield due to the alternate setup, y^{alt}. The pseudoexperiment-by-pseudoexperiment difference ($y^{alt} - y^{nom}$), also called the toy-by-toy difference in shorthand, is fitted with a Gaussian. An example of this can be seen in Fig. 8.4 for varying a background constraint on a $B^+ \to K^+e^+e^-$ fit in Run 2p1 L0L in the central-q^2 region. The mean and standard deviation are summed in quadrature and quoted as the systematic uncertainty. The following uncertainties can be evaluated in this manner:

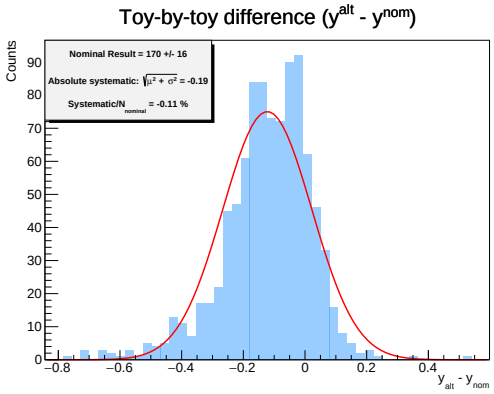


Figure 8.4: Distribution of $N^{alt}_{B^+\to K^+e^+e^-} - N^{nom}_{B^+\to K^+e^+e^-}$ in central-q^2 L0L Run 2p1 where an alternative background constraint is studied as a systematic effect. The distribution can be fitted with a Gaussian model (red line). The mean and width are added in quadrature and quoted as the systematic uncertainty on $N_{B^+\to K^+e^+e^-}$.

Uncertainty due to fixed shape parameters - Instead of fixing the shape parameters from fits to simulation the parameters would be allowed to vary within their uncertainties. For each shape parameter 1000 pseudoexperiments would be produced and the toy-by-toy difference with the nominal setup would be determined. The parameter which yields the largest difference using the toy-by-toy method would be quoted as the systematic uncertainty.

Uncertainty due to `RooKeysPdf` pdf - When using a `RooKeysPdf` to determine the shape of a fit component the smoothing parameter, which determines how many Gaussians are used to compute the shape, can be varied. For

various values of the smoothing parameter 1000 pseudoexperiments would be produced and the largest difference observed using the toy-by-toy method with the nominal fit setup would be quoted as the systematic uncertainty.

Uncertainty on partially reconstructed background normalisation - The normalisation factor computed for the B^0, B^+ and B_s^0 inclusive background in Appendix J is assigned a 10% uncertainty. The factor is varied within its uncertainty as an alternative model and 1000 pseudoexperiments would be used to compare this to the nominal setup using the toy-by-toy difference and a systemic uncertainty assigned.

Uncertainity on the $K\pi\pi$ background - The resonant structure of $b \to K\pi\pi\ell^+\ell^-$ is unknown so the simulated $B^+ \to K\pi\pi e^+e^-$ samples used to extract the shape of this background are reweighted to match the $K\pi\pi$ spectrum in LHCb's latest analysis of this mode detailed in Ref. []. The uncertainty is assigned by generating 1000 pseudoexperiments which were modelled using the reweighted simulation and then fitting back to it with the unweighted simulation and performing a toy-by-toy comparison. A systematic can then be assigned in the low and central-q^2 bins separately.

8.3 Non-flatness of $r_{J/\psi}$

The resonant mode ratios $r^K_{J/\psi}$ and $r^{K^*}_{J/\psi}$ are considered flat in a given variable if the fit results in each bin are consistent with a set of independent measurements centred on unity and there is no significant trend observed in the fit results for one variable. The results of the flatness checks for $r_{J/\psi}$ are shown in Sec. 9.2. If the non-flatness is statistically significant the simulated samples will be reweighted in the relevant variables to obtain flatness and the resulting change in R_X will be taken as the systematic.

The residual non-flatness can be quantified with a flatness parameter, d_f, defined as

$$d_f = \frac{\sum_i^8 \epsilon^i_{rare,\mu} \cdot \mathcal{Y}^i_\mu}{\sum_i^8 \epsilon^i_{rare,\mu}} \cdot \frac{\sum_i^8 \epsilon^i_{J/\psi,\mu}}{\sum_i^8 N^i_\mu} \Bigg/ \frac{\sum_i^8 \epsilon^i_{rare,e} \cdot \mathcal{Y}^i_e}{\sum_i^8 \epsilon^i_{rare,e}} \cdot \frac{\sum_i^8 \epsilon^i_{J/\psi,e}}{\sum_i^8 N^i_e} - 1, \tag{8.2}$$

where N^i_ℓ denotes the control mode yield measured in bin i and $\mathcal{Y}^i_e$ is the efficiency corrected control mode yield,

$$\mathcal{Y}^i_\ell = \frac{N^i_\ell}{\epsilon^i_{J/\psi,\ell}}. \tag{8.3}$$

Essentially, the d_f parameter calculates a pseudo double ratio, but the rare or $\psi(2S)$ mode yields are replaced with $\epsilon^i_{rare,\ell} \cdot \mathcal{Y}^i_\ell$ terms instead. Because $\mathcal{Y}^i_\ell$ is lepton flavour universal, d_f should be compatible with zero if the efficiencies are properly calibrated. In practice, d_f are required to be smaller than the total systematic uncertainty of R_K and R_{K^*}. Otherwise, it hints at additional systematic effects that do not cancel in the efficiency double ratio. The parameter d_f is calculated applying the formula in 8.2 in the low and central bins for both R_K and R_{K^*} decay modes using the full correction chain extracted from the B^0 and B^+, examples, which are representative of the general picture, are shown in Figs. 8.5 and 8.6 corresponding to R_K RUN 2P1 in the L0I and L0L trigger categories respectively. It can be seen that for the vast majority of variables, and in particular those used to derive corrections, the overall contribution is at the sub-percent level.

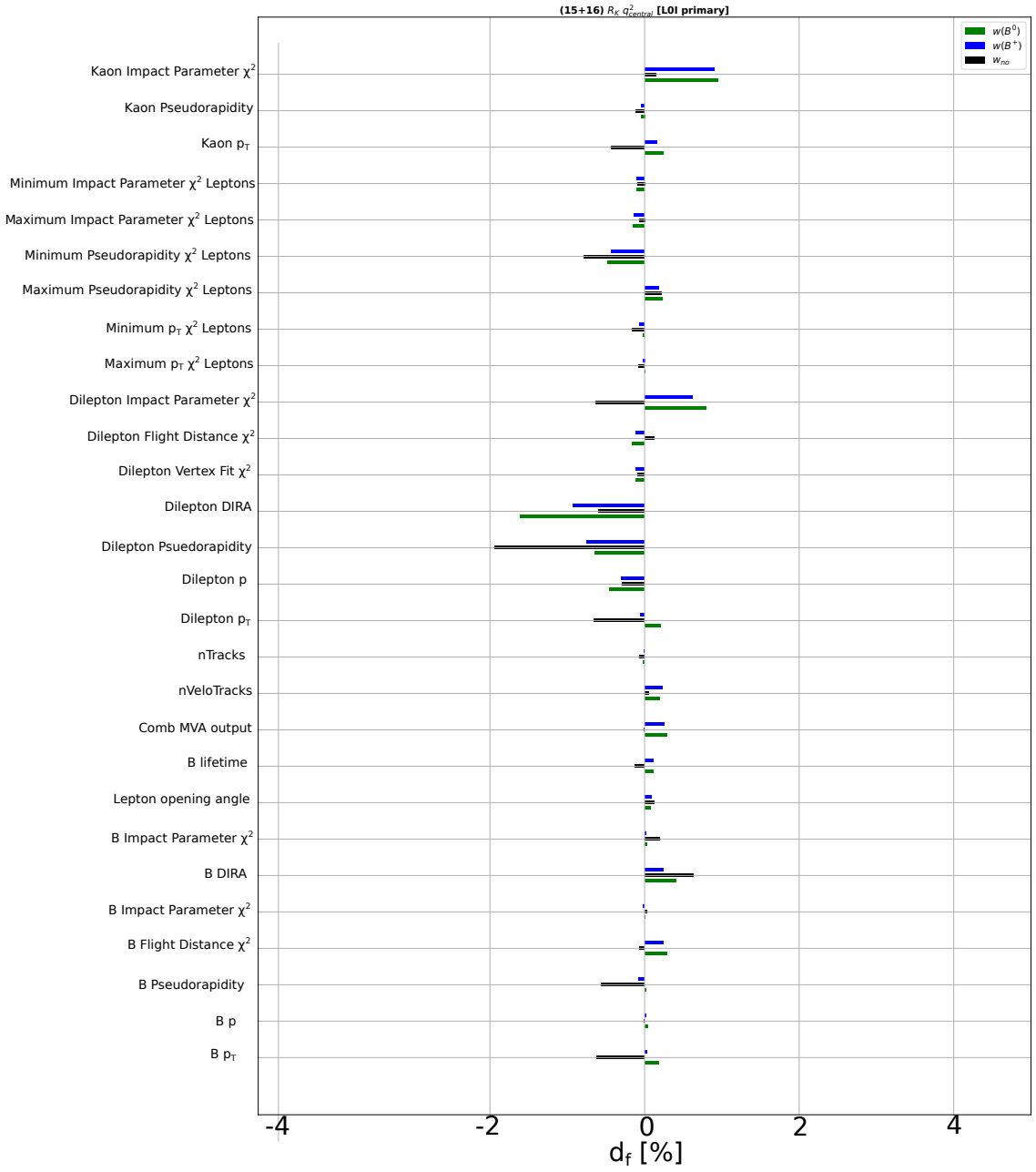


Figure 8.5: The d_f flatness systematic parameter calculated for a number kinematic and geometric variables for R_K mode in the central q^2 bin in RUN 2P1 in the L0I trigger category.

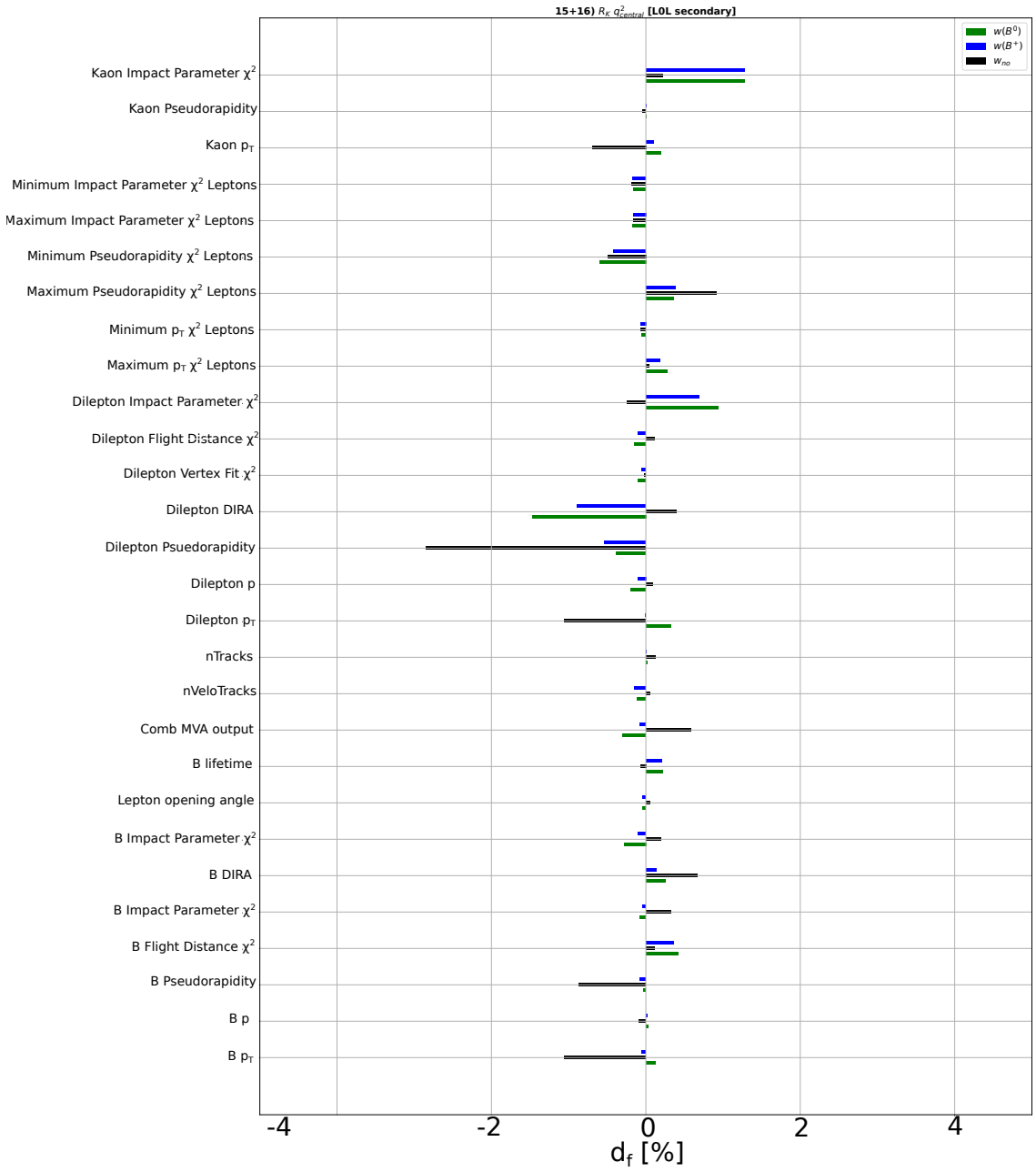


Figure 8.6: The d_f flatness systematic parameter calculated for a number kinematic and geometric variables for R_K mode in the central q^2 bin in Run 2p1 in the L0L trigger category.

Chapter 9

Results

9.1 Integrated $r_{J/\psi}$

The single ratio $r_{J/\psi}$ is defined in Eqs. 3.3 and 3.4 for R_K and R_{K^*} respectively and is expected to be unity. Mass fits as described in Sec. 7 are performed independently for B^+ and $B^0 J/\psi$ resonant modes in Run 1, Run 2p1 and Run 2p2 in the L0I and L0L! trigger categories. The efficiencies are determined using corrected simulation as described in Sec. 6, corrections derived from B^+ and B^0 modes are compared. As a baseline approach the L0I muon category is divided by the L0I electron category and the L0M! category is divided by the L0E! category. The integrated $r_{J/\psi}$ results across the whole J/ψ q^2 region are shown in Tabs. 9.1– 9.2, whilst the evolution of the value of $r_{J/\psi}$ with each subsequent correction step can be seen in Figs. 9.1– 9.6. The quoted uncertainties are the statistical uncertainties from the mass fits and the systematic uncertainties from the bootstrapping of the L0 and HLT correction maps.

Table 9.1: $r^K_{J/\psi}$ calculated with and without the full correction chain for Run 1, Run 2p1 and Run 2p2 with corrections derived from both B^+ and B^0 modes. The first error is the statisical uncertainty associated with the fits and the second error is related to the systematics.

Type	w_{no}	$w_{\text{PID,TRK,BKIN,MULT,RECO,L0,HLT}}$
R1 L0I $w(B^0)$	$1.161 \pm 0.006 \pm 0.008$	$1.061 \pm 0.005 \pm 0.040$
R1 L0I $w(B^+)$	$1.161 \pm 0.006 \pm 0.008$	$1.080 \pm 0.005 \pm 0.020$
R1 L0L! $w(B^0)$	$1.323 \pm 0.006 \pm 0.008$	$1.067 \pm 0.004 \pm 0.020$
R1 L0L! $w(B^+)$	$1.323 \pm 0.006 \pm 0.008$	$1.056 \pm 0.005 \pm 0.019$
R2p1 L0I $w(B^0)$	$1.143 \pm 0.005 \pm 0.009$	$1.032 \pm 0.005 \pm 0.019$
R2p1 L0I $w(B^+)$	$1.143 \pm 0.005 \pm 0.009$	$1.024 \pm 0.005 \pm 0.009$
R2p1 L0L! $w(B^0)$	$1.411 \pm 0.005 \pm 0.009$	$1.038 \pm 0.004 \pm 0.009$
R2p1 L0L! $w(B^+)$	$1.411 \pm 0.005 \pm 0.009$	$1.039 \pm 0.004 \pm 0.009$
R2p2 L0I $w(B^0)$	$1.128 \pm 0.004 \pm 0.002$	$1.020 \pm 0.003 \pm 0.010$
R2p2 L0I $w(B^+)$	$1.128 \pm 0.004 \pm 0.002$	$1.033 \pm 0.003 \pm 0.010$
R2p2 L0L! $w(B^0)$	$1.418 \pm 0.004 \pm 0.002$	$1.005 \pm 0.003 \pm 0.010$
R2p2 L0L! $w(B^+)$	$1.418 \pm 0.004 \pm 0.002$	$1.034 \pm 0.003 \pm 0.010$

Table 9.2: $r^{K^*}_{J/\psi}$ calculated with and without the full correction chain for Run 1, Run 2p1 and Run 2p2 with corrections derived from both B^+ and B^0 modes.The first error is the statisical uncertainty associated with the fits and the second error is related to the systematics.

Type	w_{no}	$w_{\text{PID,TRK,BKIN,MULT,RECO,L0,HLT}}$
R1 L0I $w(B^0)$	$1.135 \pm 0.011 \pm 0.005$	$1.041 \pm 0.010 \pm 0.039$
R1 L0I $w(B^+)$	$1.135 \pm 0.011 \pm 0.005$	$1.050 \pm 0.010 \pm 0.017$
R1 L0L! $w(B^0)$	$1.303 \pm 0.011 \pm 0.005$	$1.067 \pm 0.009 \pm 0.018$
R1 L0L! $w(B^+)$	$1.303 \pm 0.011 \pm 0.005$	$1.056 \pm 0.009 \pm 0.018$
R2p1 L0I $w(B^0)$	$1.143 \pm 0.010 \pm 0.005$	$1.022 \pm 0.008 \pm 0.018$
R2p1 L0I $w(B^+)$	$1.143 \pm 0.010 \pm 0.005$	$1.014 \pm 0.009 \pm 0.004$
R2p1 L0L! $w(B^0)$	$1.383 \pm 0.010 \pm 0.005$	$1.038 \pm 0.007 \pm 0.007$
R2p1 L0L! $w(B^+)$	$1.383 \pm 0.010 \pm 0.005$	$1.039 \pm 0.007 \pm 0.007$
R2p2 L0I $w(B^0)$	$1.133 \pm 0.007 \pm 0.007$	$1.030 \pm 0.006 \pm 0.008$
R2p2 L0I $w(B^+)$	$1.133 \pm 0.007 \pm 0.007$	$1.043 \pm 0.006 \pm 0.008$
R2p2 L0L! $w(B^0)$	$1.414 \pm 0.007 \pm 0.007$	$1.015 \pm 0.005 \pm 0.009$
R2p2 L0L! $w(B^+)$	$1.414 \pm 0.007 \pm 0.007$	$1.044 \pm 0.005 \pm 0.009$

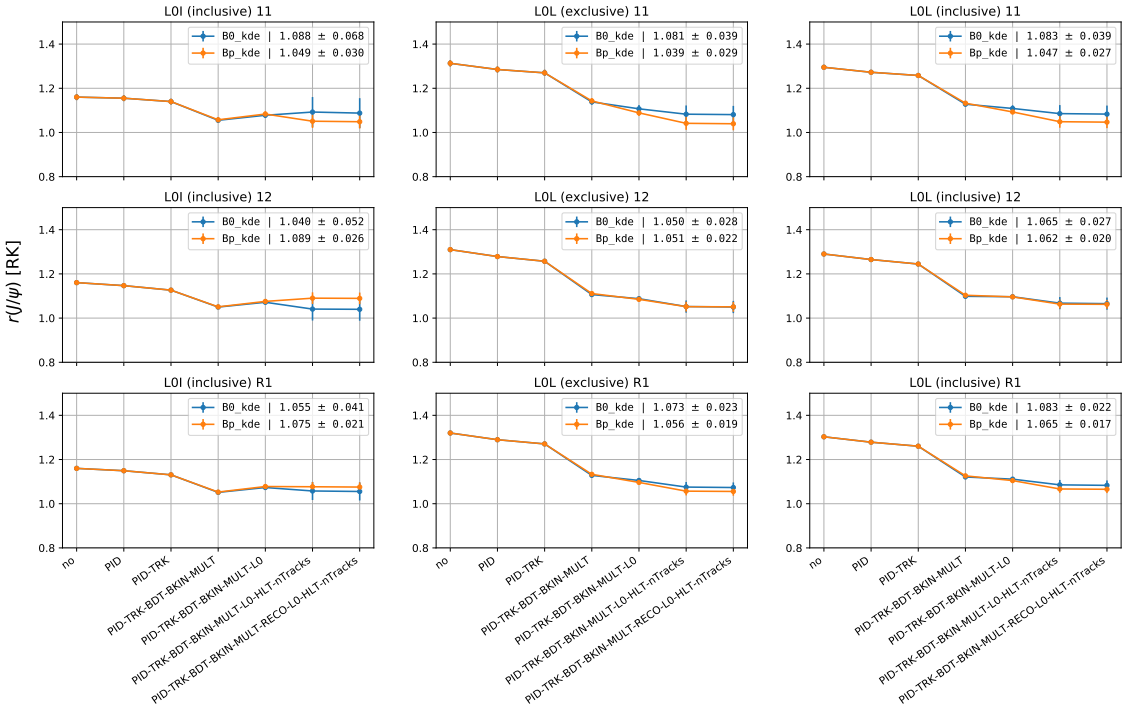


Figure 9.1: $r^K_{J/\psi}$ calculated for all efficiency correction steps for Run 1. The errors are combined statistical uncertainties from the fits and systematic uncertainties from bootstrapping.

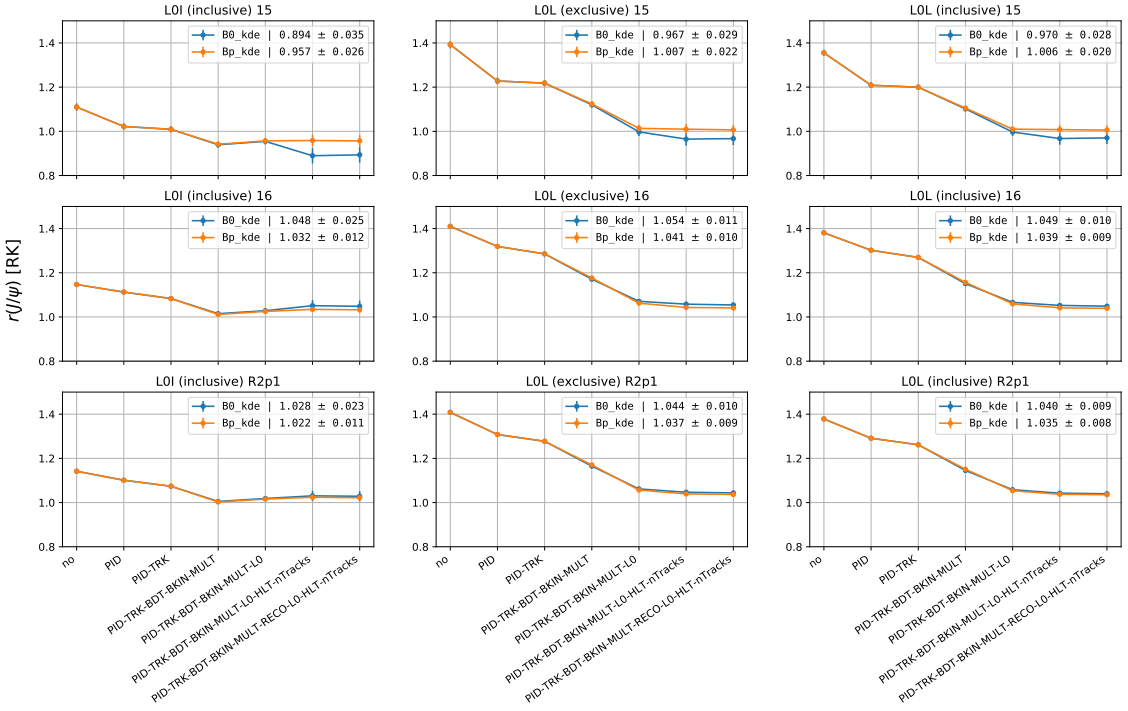


Figure 9.2: $r^K_{J/\psi}$ calculated for all efficiency correction steps for Run 2p1. The errors are combined statistical uncertainties from the fits and systematic uncertainties from bootstrapping.

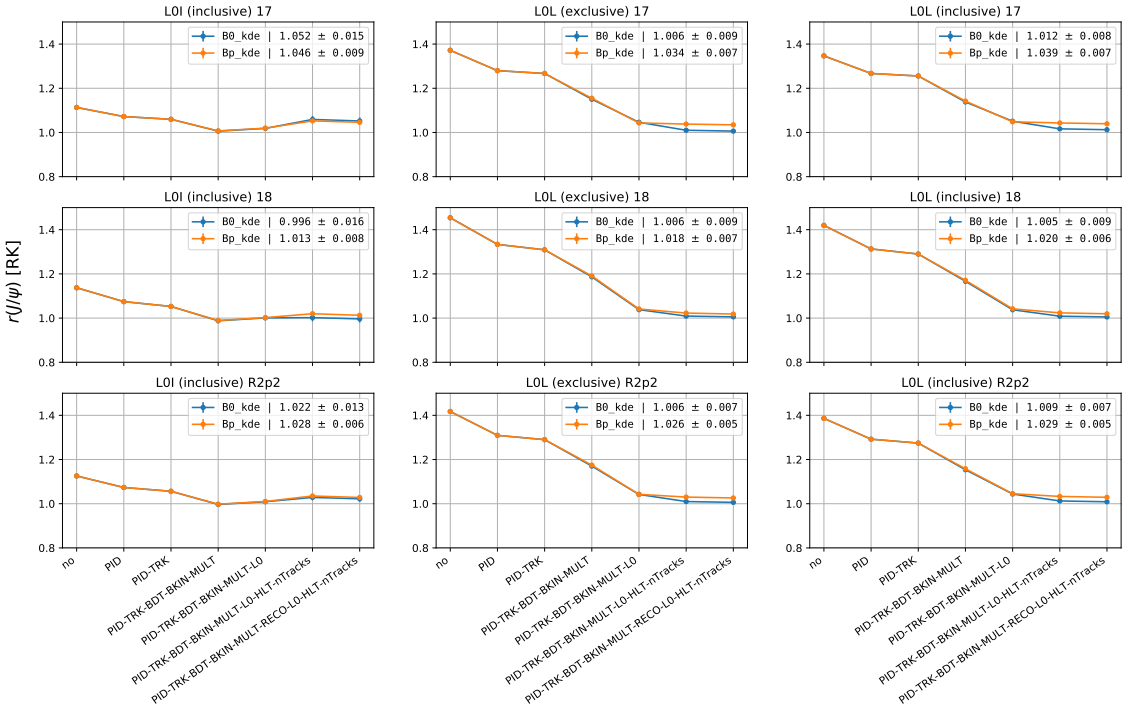


Figure 9.3: $r^K_{J/\psi}$ calculated for all efficiency correction steps for Run 2p2 R_K. The errors are combined statistical uncertainties from the fits and systematic uncertainties from bootstrapping.

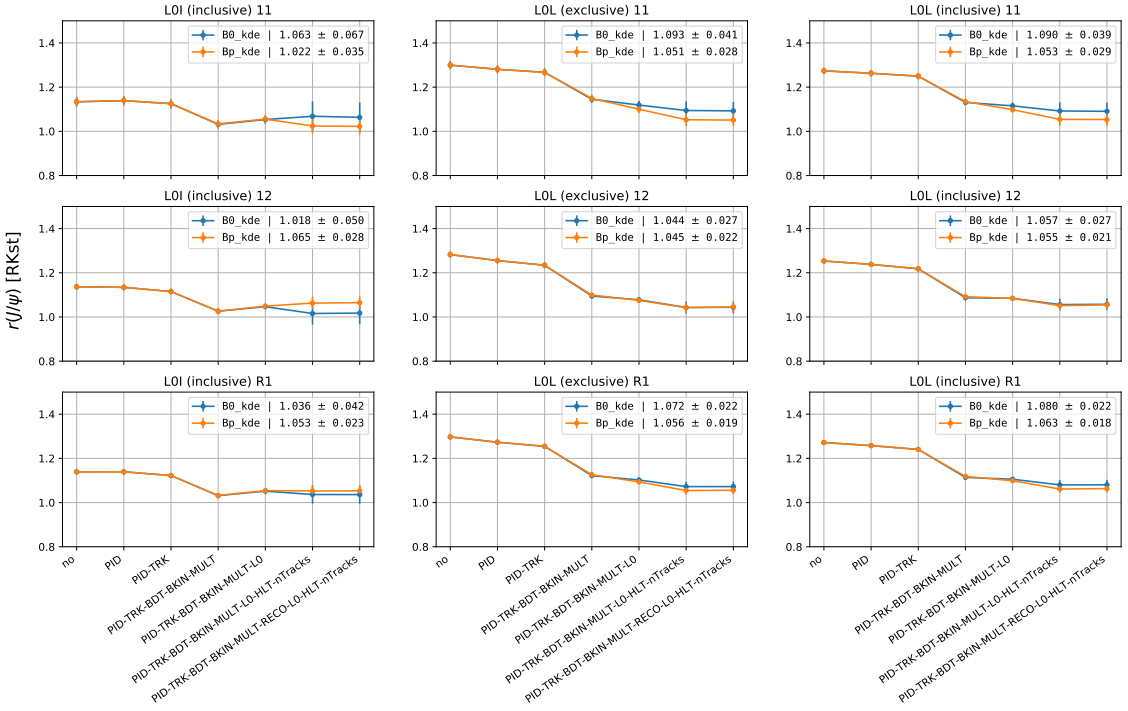


Figure 9.4: $r^{K^*}_{J/\psi}$ calculated for all efficiency correction steps for Run 1. The errors are combined statistical uncertainties from the fits and systematic uncertainties from bootstrapping.

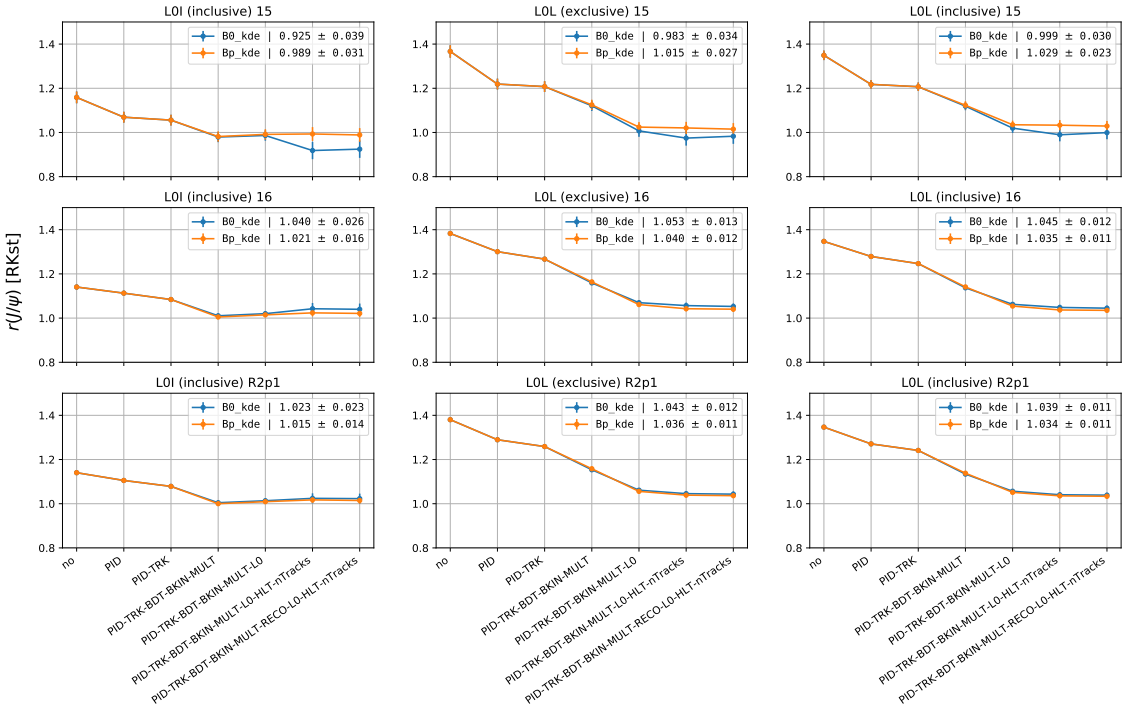


Figure 9.5: $r^{K^*}_{J/\psi}$ calculated for all efficiency correction steps for Run 2p1. The errors are combined statistical uncertainties from the fits and systematic uncertainties from bootstrapping.

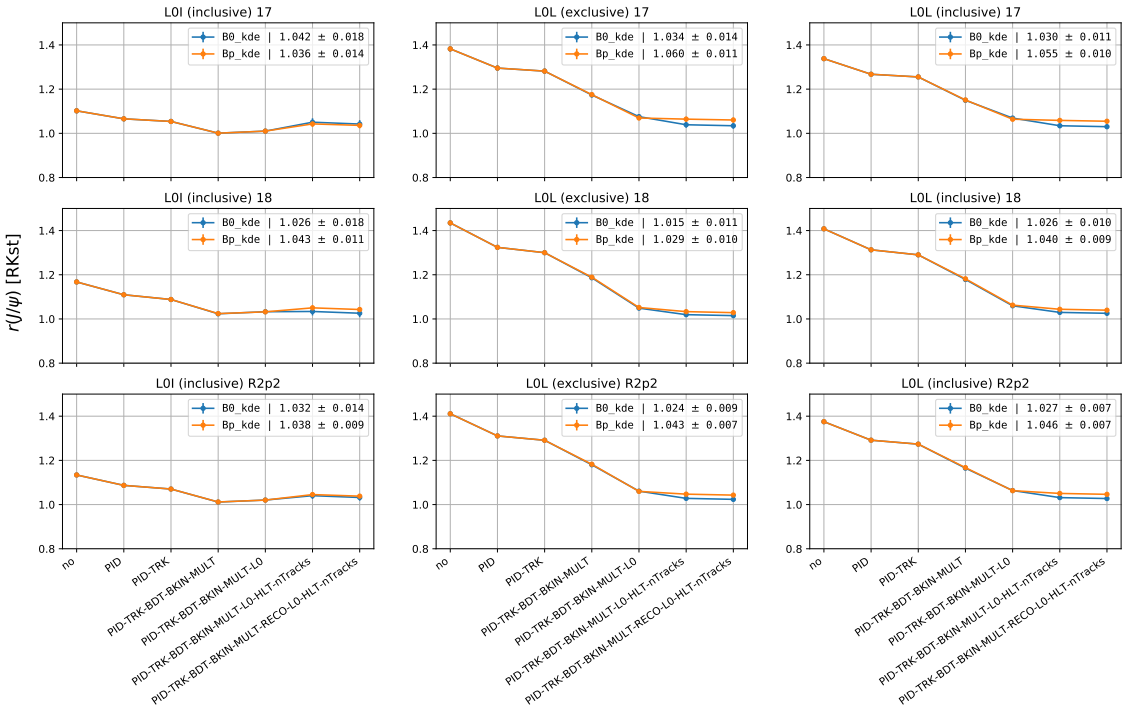


Figure 9.6: $r^{K^*}_{J/\psi}$ calculated for all efficiency correction steps for Run 2p2. The errors are combined statistical uncertainties from the fits and systematic uncertainties from bootstrapping.

9.2 Flatness of $r_{J/\psi}$

As a further cross-check $r_{J/\psi}$ can be plotted as a function of various kinematic variables, many of which are used as proxy variables in the various correction steps or important variables in the MVA training. In order to perform this check $r_{J/\psi}$ must be computed in bins of each kinematic variable, due to the available statistics eight bins are used for both B^+ and B^0 modes. When determining the binning a region corresponding to $\pm 60\,\text{MeV}$ around the B mass is used. To determine the binning scheme in each variable an isobinning approach is used, with one caveat: as there are generally fewer entries in the electron mode tuples than those for the muon mode tuples, the electron mode tuples are weighted so that the sum of the weights is equal to the number of entries in the corresponding muon mode tuple and the isobinning is then performed on the muon and weighted electron tuples. This method of isobinning ensures that there are no cases in which a certain bin may contain events from the muon mode but no events from the electron mode and vice versa. Note that the lower and upper bin values are applied as additional cuts when determining the flatness for a particular variable.

With the binning schemes determined, simultaneous fits to both the muon and electron modes in each bin can be performed. Unbinned fits are performed to the `DecayTreeFitter` B mass with the nominal fit setups and constraints. In the few cases where for electron modes where there are no events in one of the bremsstrahlung categories fits are performed to the merged bremsstrahlung categories instead. The efficiency in each bin, ϵ_{bin}, is calculated as:

$$\epsilon_{bin} = \frac{N_{bin}}{N_{total}} \times \epsilon_{Selection} \tag{9.1}$$

where N_{bin} and N_{total} are the number of events in a given bin and the total number of events in the entire binning range respectively and $\epsilon_{Selection}$ is the efficiency of the fully corrected and selected decay mode. Once the fits are performed it is checked the the sum of the yields in all the bins is consistent with the yield obtained from the integrated yields determined in Sec. 9.1. With this check performed the efficiency corrected yields can be extracted and $r_{J/\psi}$ can be computed in each bin.

If $r_{J/\psi}$ appears to be flat in a given variable, it can be due to one of three scenarios. In the first, most ideal, scenario, the efficiency corrected yields for both the muon and electron modes are flat and at the same scale. In the second scenario both the efficiency corrected yields for the muon and electron modes are flat but there is a

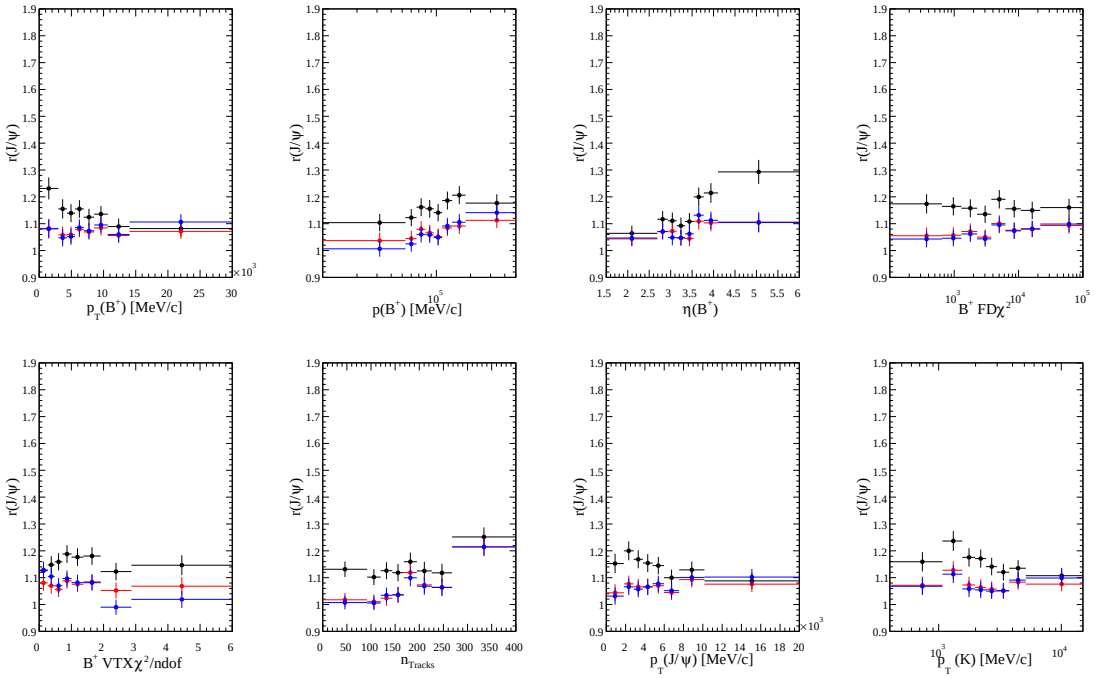


Figure 9.7: $r^K_{J/\psi}$ plotted as a function of a variety of kinematic and other important variables in the L0I trigger category in RUN 1, the variables running from left ro right on the top row are: The transverse momentum of the B^+, the momentum of the B^+, the pseudorapidity of the B^+ and the flight distance χ^2 of the B^+. From left to right on the second row: the χ^2 of the vertex fit to the B^+ divided by the number of degrees of freedom of that fit, the number of tracks in the event, the transverse momentum of the dilepton pair and the transverse momentum of the kaon. The black points represent data with no corrections applied, the red points represent data with corrections from the B^+ mode applied and the blue points represent data with corrections from the B^0 mode applied.

constant offset between the two. In the last scenario both the muon and electron modes have trends in their efficiency corrected yields which cancel out in the ratio. If any non-flat behaviour is also present in the rare mode this will cancel in the double ratio, any residual non-flatness after the double ratio will be assigned to a flatness systematic. Plots showing the flatness in a selected set of variables for $r^K_{J/\psi}$ can be seen in Figs. 9.7 - 9.12 whilst plots showing the flatness of $r^{K^*}_{J/\psi}$ can be seen in Figs. 9.13- 9.18.

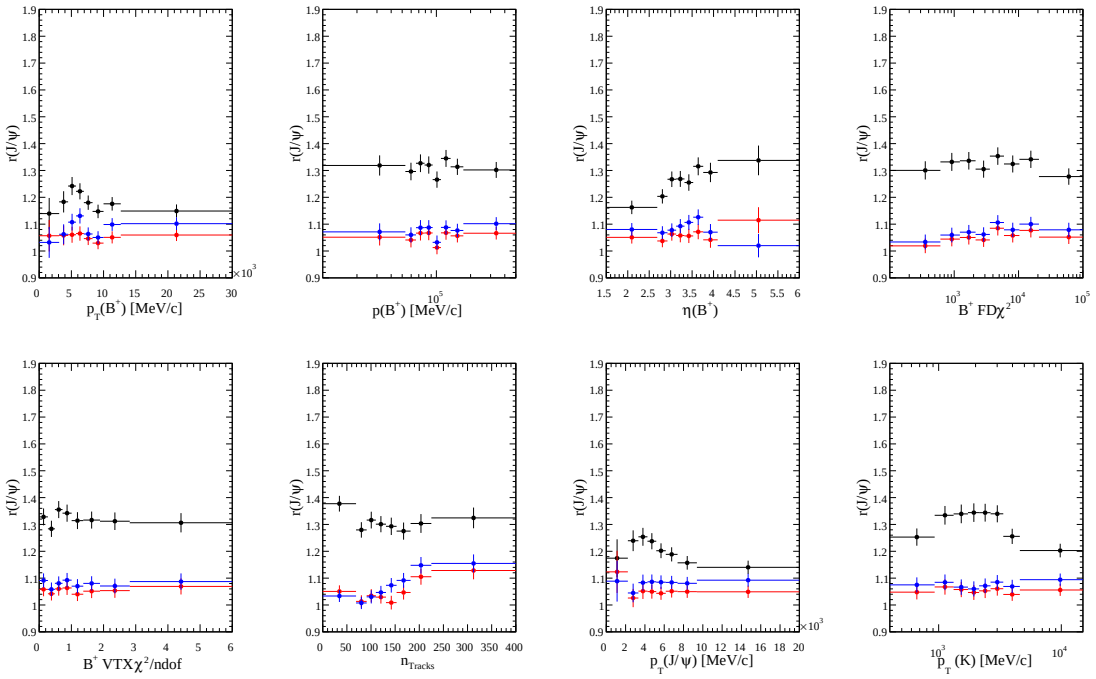


Figure 9.8: $r^K_{J/\psi}$ plotted as a function of a variety of kinematic and other important variables in the L0L! trigger category in Run 1, the variables running from left ro right on the top row are: The transverse momentum of the B^+, the momentum of the B^+, the pseudorapidity of the B^+ and the flight distance χ^2 of the B^+. From left to right on the second row: the χ^2 of the vertex fit to the B^+ divided by the number of degrees of freedom of that fit, the number of tracks in the event, the transverse momentum of the dilepton pair and the transverse momentum of the kaon. The black points represent data with no corrections applied, the red points represent data with corrections from the B^+ mode applied and the blue points represent data with corrections from the B^0 mode applied.

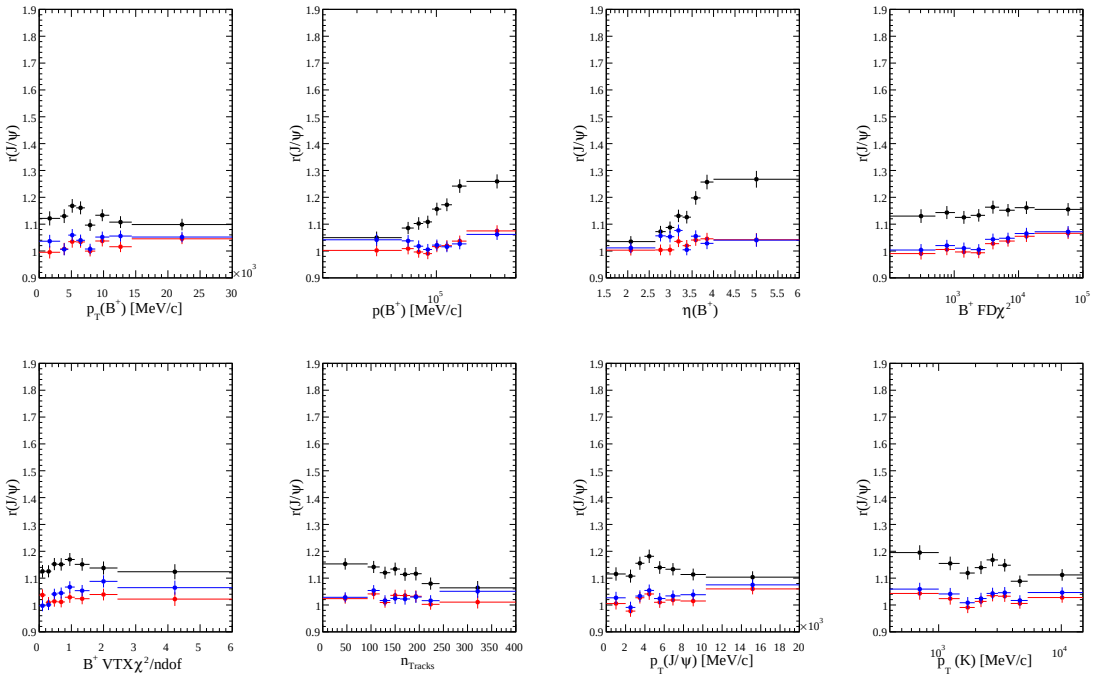


Figure 9.9: $r^K_{J/\psi}$ plotted as a function of a variety of kinematic and other important variables in the L0I trigger category in Run 2p1, the variables running from left ro right on the top row are: The transverse momentum of the B^+, the momentum of the B^+, the pseudorapidity of the B^+ and the flight distance χ^2 of the B^+. From left to right on the second row: the χ^2 of the vertex fit to the B^+ divided by the number of degrees of freedom of that fit, the number of tracks in the event, the transverse momentum of the dilepton pair and the transverse momentum of the kaon. The black points represent data with no corrections applied, the red points represent data with corrections from the B^+ mode applied and the blue points represent data with corrections from the B^0 mode applied.

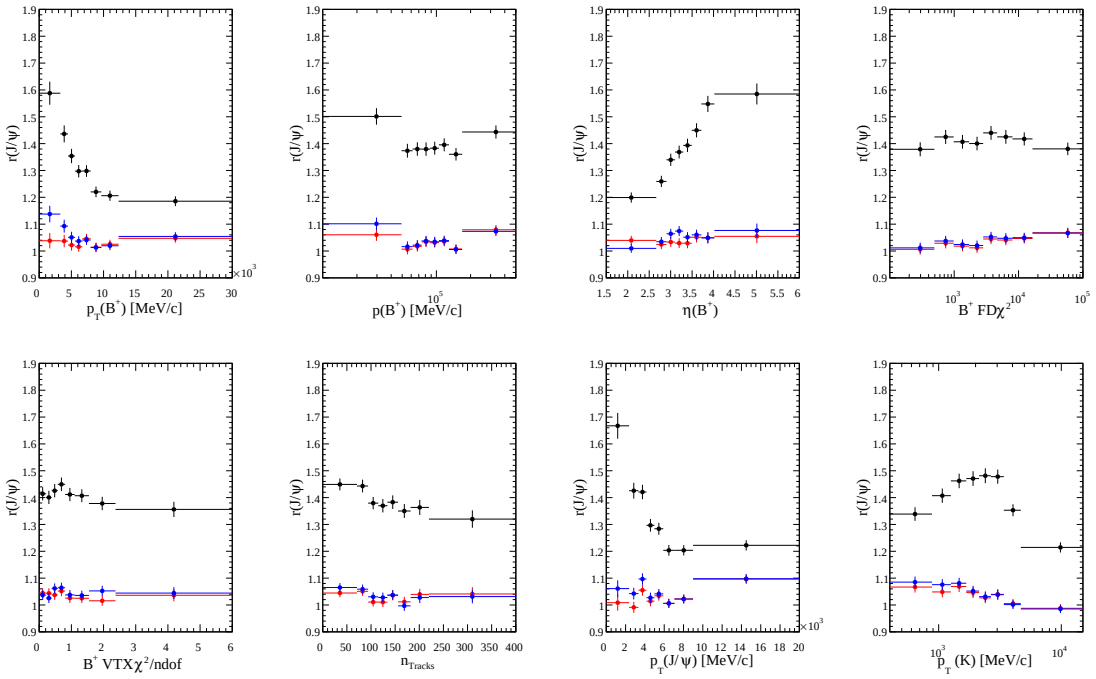


Figure 9.10: $r^K_{J/\psi}$ plotted as a function of a variety of kinematic and other important variables in the L0L! trigger category in Run 2p1, the variables running from left ro right on the top row are: The transverse momentum of the B^+, the momentum of the B^+, the pseudorapidity of the B^+ and the flight distance χ^2 of the B^+. From left to right on the second row: the χ^2 of the vertex fit to the B^+ divided by the number of degrees of freedom of that fit, the number of tracks in the event, the transverse momentum of the dilepton pair and the transverse momentum of the kaon. The black points represent data with no corrections applied, the red points represent data with corrections from the B^+ mode applied and the blue points represent data with corrections from the B^0 mode applied.

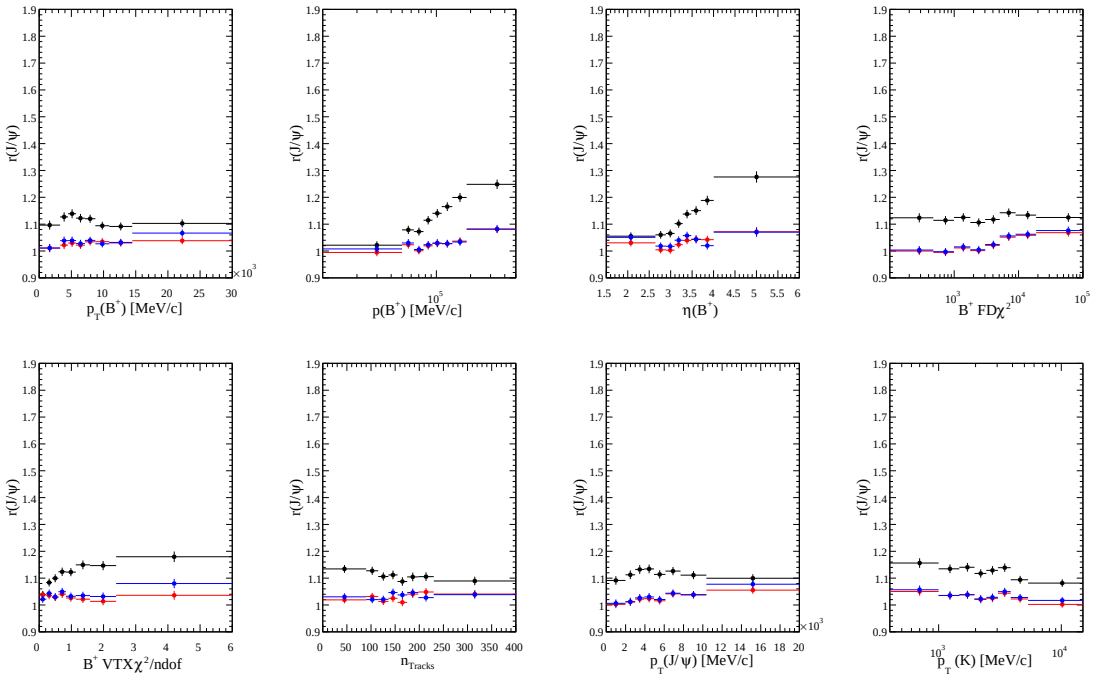


Figure 9.11: $r^K_{J/\psi}$ plotted as a function of a variety of kinematic and other important variables in the L0I trigger category in Run 2p2, the variables running from left ro right on the top row are: The transverse momentum of the B^+, the momentum of the B^+, the pseudorapidity of the B^+ and the flight distance χ^2 of the B^+. From left to right on the second row: the χ^2 of the vertex fit to the B^+ divided by the number of degrees of freedom of that fit, the number of tracks in the event, the transverse momentum of the dilepton pair and the transverse momentum of the kaon. The black points represent data with no corrections applied, the red points represent data with corrections from the B^+ mode applied and the blue points represent data with corrections from the B^0 mode applied.

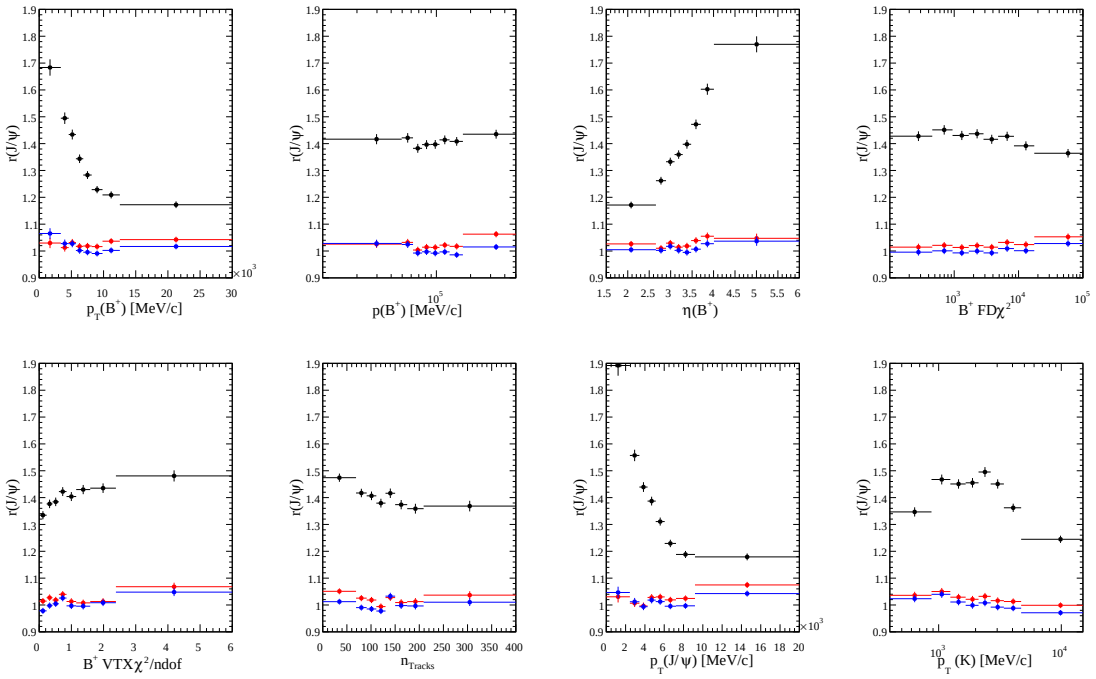


Figure 9.12: $r^K_{J/\psi}$ plotted as a function of a variety of kinematic and other important variables in the L0L! trigger category in Run 2p2, the variables running from left ro right on the top row are: The transverse momentum of the B^+, the momentum of the B^+, the pseudorapidity of the B^+ and the flight distance χ^2 of the B^+. From left to right on the second row: the χ^2 of the vertex fit to the B^+ divided by the number of degrees of freedom of that fit, the number of tracks in the event, the transverse momentum of the dilepton pair and the transverse momentum of the kaon. The black points represent data with no corrections applied, the red points represent data with corrections from the B^+ mode applied and the blue points represent data with corrections from the B^0 mode applied.

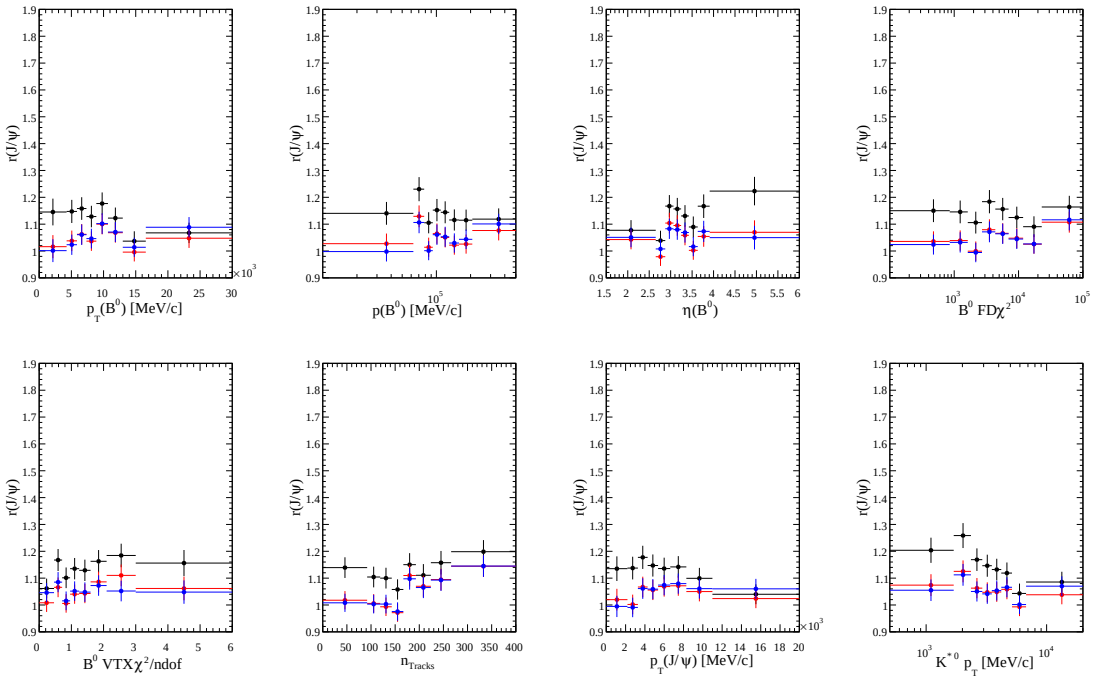


Figure 9.13: $r^{K^*}_{J/\psi}$ plotted as a function of a variety of kinematic and other important variables in the L0I trigger category in RUN 1, the variables running from left ro right on the top row are: The transverse momentum of the B^0, the momentum of the B^0, the pseudorapidity of the B^0 and the flight distance χ^2 of the B^0. From left to right on the second row: the χ^2 of the vertex fit to the B^0 divided by the number of degrees of freedom of that fit, the number of tracks in the event, the transverse momentum of the dilepton pair and the transverse momentum of the K^{*0}. The black points represent data with no corrections applied, the red points represent data with corrections from the B^+ mode applied and the blue points represent data with corrections from the B^0 mode applied.

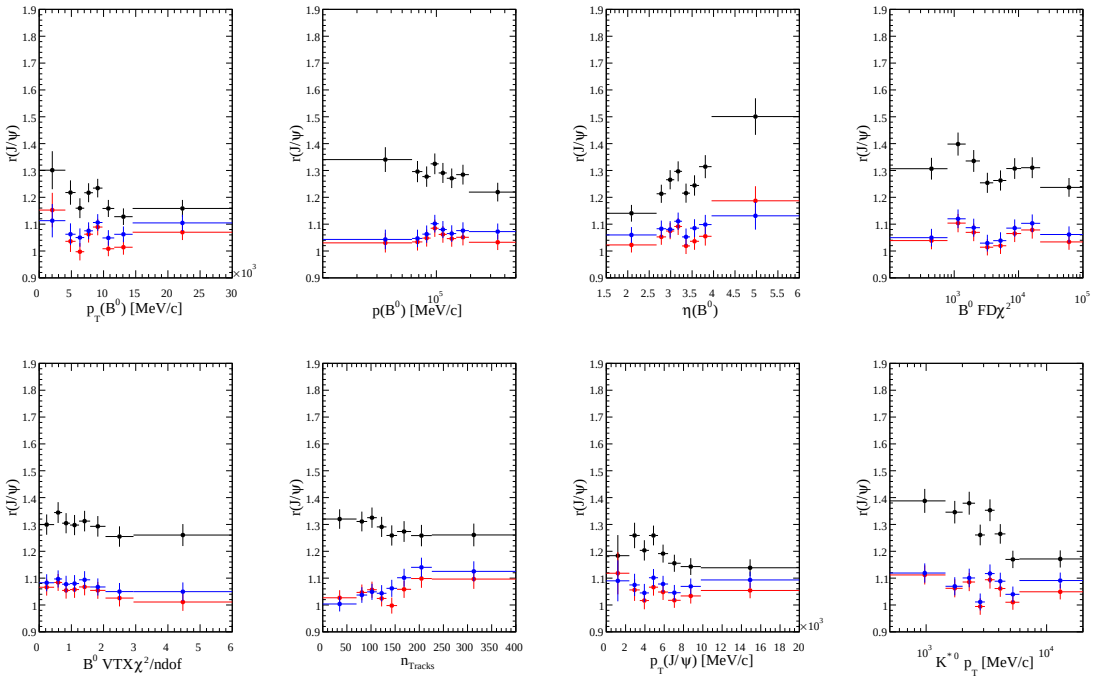


Figure 9.14: $r^{K^*}_{J/\psi}$ plotted as a function of a variety of kinematic and other important variables in the L0L! trigger category in RUN 1, the variables running from left ro right on the top row are: The transverse momentum of the B^0, the momentum of the B^0, the pseudorapidity of the B^0 and the flight distance χ^2 of the B^0. From left to right on the second row: the χ^2 of the vertex fit to the B^0 divided by the number of degrees of freedom of that fit, the number of tracks in the event, the transverse momentum of the dilepton pair and the transverse momentum of the K^{*0}. The black points represent data with no corrections applied, the red points represent data with corrections from the B^+ mode applied and the blue points represent data with corrections from the B^0 mode applied.

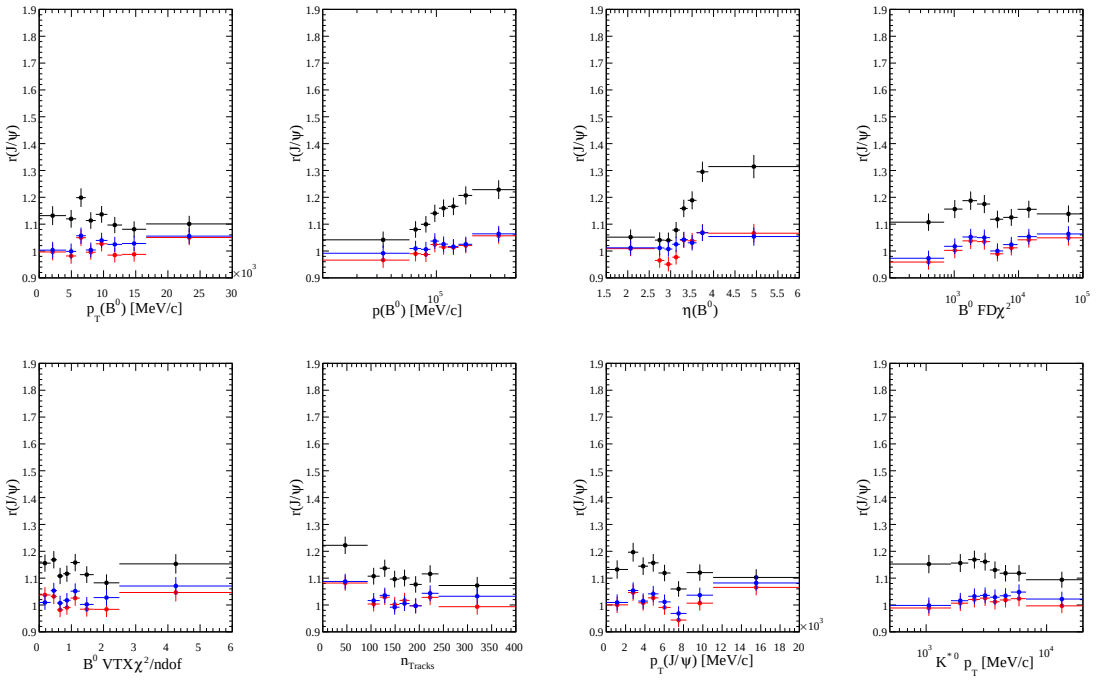


Figure 9.15: $r^{K^*}_{J/\psi}$ plotted as a function of a variety of kinematic and other important variables in the L0I trigger category in Run 2p1, the variables running from left ro right on the top row are: The transverse momentum of the B^0, the momentum of the B^0, the pseudorapidity of the B^0 and the flight distance χ^2 of the B^0. From left to right on the second row: the χ^2 of the vertex fit to the B^0 divided by the number of degrees of freedom of that fit, the number of tracks in the event, the transverse momentum of the dilepton pair and the transverse momentum of the K^{*0}. The black points represent data with no corrections applied, the red points represent data with corrections from the B^+ mode applied and the blue points represent data with corrections from the B^0 mode applied.

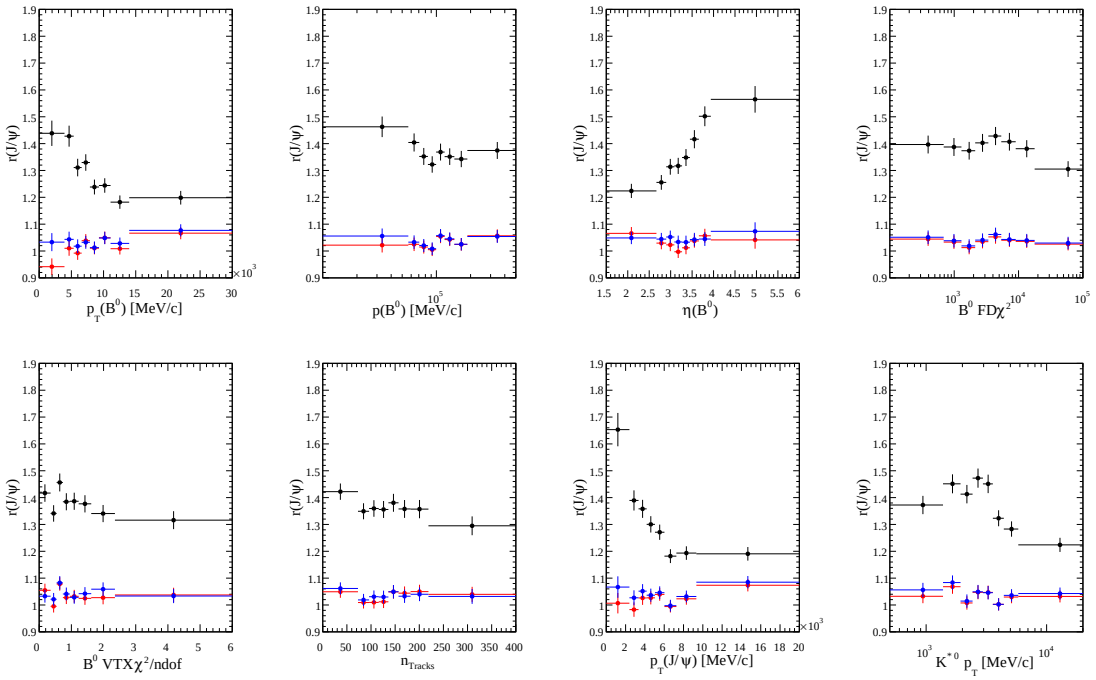


Figure 9.16: $r^{K^*}_{J/\psi}$ plotted as a function of a variety of kinematic and other important variables in the L0L! trigger category in Run 2p1, the variables running from left ro right on the top row are: The transverse momentum of the B^0, the momentum of the B^0, the pseudorapidity of the B^0 and the flight distance χ^2 of the B^0. From left to right on the second row: the χ^2 of the vertex fit to the B^0 divided by the number of degrees of freedom of that fit, the number of tracks in the event, the transverse momentum of the dilepton pair and the transverse momentum of the K^{*0}. The black points represent data with no corrections applied, the red points represent data with corrections from the B^+ mode applied and the blue points represent data with corrections from the B^0 mode applied.

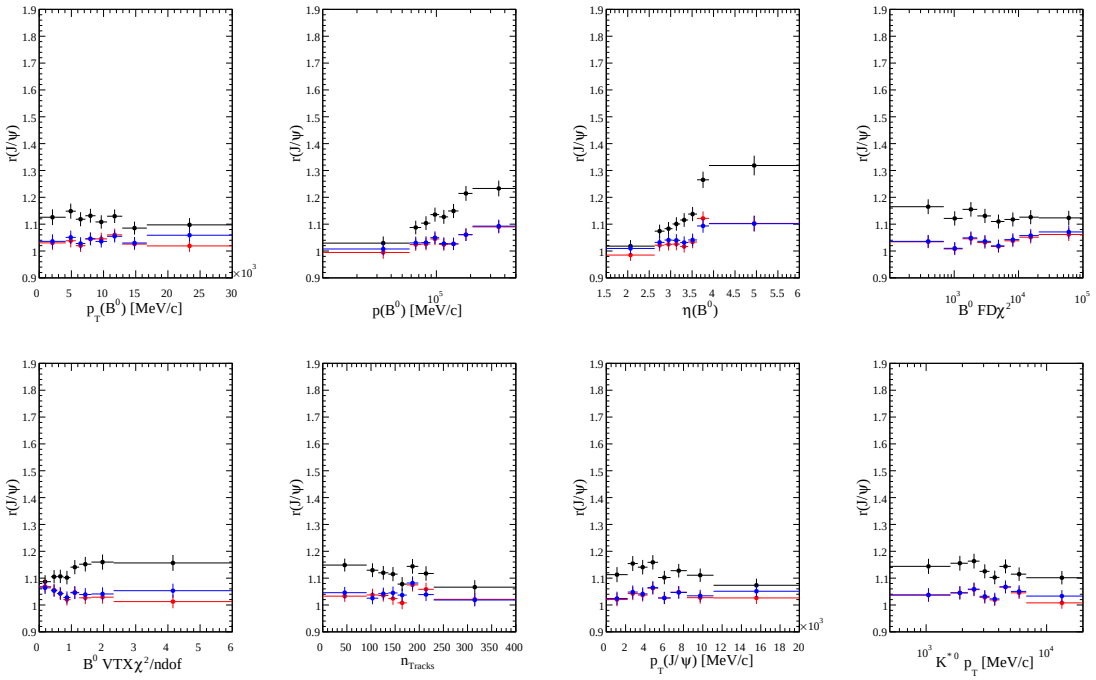


Figure 9.17: $r^{K^*}_{J/\psi}$ plotted as a function of a variety of kinematic and other important variables in the L0I trigger category in Run 2p2, the variables running from left ro right on the top row are: The transverse momentum of the B^0, the momentum of the B^0, the pseudorapidity of the B^0 and the flight distance χ^2 of the B^0. From left to right on the second row: the χ^2 of the vertex fit to the B^0 divided by the number of degrees of freedom of that fit, the number of tracks in the event, the transverse momentum of the dilepton pair and the transverse momentum of the K^{*0}. The black points represent data with no corrections applied, the red points represent data with corrections from the B^+ mode applied and the blue points represent data with corrections from the B^0 mode applied.

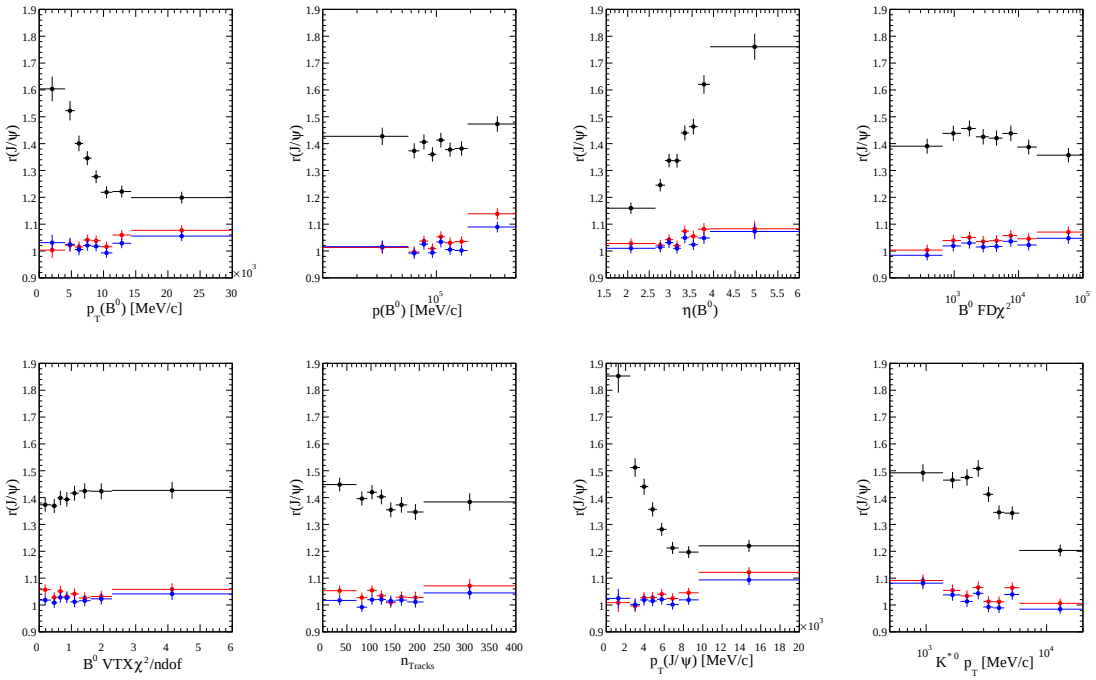


Figure 9.18: $r^{K^*}_{J/\psi}$ plotted as a function of a variety of kinematic and other important variables in the L0L! trigger category in Run 2p2, the variables running from left ro right on the top row are: The transverse momentum of the B^0, the momentum of the B^0, the pseudorapidity of the B^0 and the flight distance χ^2 of the B^0. From left to right on the second row: the χ^2 of the vertex fit to the B^0 divided by the number of degrees of freedom of that fit, the number of tracks in the event, the transverse momentum of the dilepton pair and the transverse momentum of the K^{*0}. The black points represent data with no corrections applied, the red points represent data with corrections from the B^+ mode applied and the blue points represent data with corrections from the B^0 mode applied.

Chapter 10

Conclusion

In this thesis the groundwork for performing the R_K and R_{K^*} measurements is presented, in particular the mandatory measurements of $r^K_{J/\psi}$ and $r^{K^*}_{J/\psi}$. In making these measurements the full LHCb dataset from 2011 to 2018, corresponding to an integrated luminosity of roughly $9\,\text{fb}^{-1}$, was used with collision energies of 7 TeV in 2011, 8 TeV in 2012 and 13 TeV in 2015–2018.

The integrated measurements of $r^K_{J/\psi}$ and $r^{K^*}_{J/\psi}$ are in fairly good agreement with unity in Run 2p1 and Run 2p2 with a small offset present in Run 1. As some systematic uncertainties are still to be evaluated, the level of consistency is likely to improve further. It is reassuring that when computing the flatness systematic parameter d_f that any non-residual flatness contributes at the sub-percent level especially in the variables used to derive correction weights.

The measurement of the flatness in $r^K_{J/\psi}$ and $r^{K^*}_{J/\psi}$ is quite promising, especially in Run 2p1 and Run 2p2 where the the $r_{J/\psi}$ distributions are rather flat in all of the variables used as proxies for the various correction steps. In Run 1 the flatness also appears to be quite promising, although there is a clear trend that can be seen in the `nTracks` variable which has yet to be understood.

In both the measurements of the flatness of $r_{J/\psi}$, and of its integrated value when corrections from either B^+ or B^0 modes have been used, the results are in very good agreement. This is very important as it allows corrections to simulation to be derived from B^+ modes when studying B^0 modes and vice versa with confidence. Doing so eliminates an important source of correlations.

Another important check to perform is the measurement of $r_{\psi(2S)}$, which is the double ratio of the $\psi(2S)$ and J/ψ resonant modes. This is also expected to be unity. Whilst not included in this thesis, the preliminary results look very promising. By looking at the flatness of $r_{\psi(2S)}$ in the same variables as for $r_{J/\psi}$ a flatness systematic can be assigned to those variables which still have a significant non-flatness.

Once the $r_{J/\psi}$ and $r_{\psi(2S)}$ checks are satisfactorily completed and further systematics evaluated, in particular bootstrapping of all the correction steps and fully accounting for correlations, work can proceed on extracting R_K and R_{K^*}. The MVA selection detailed in this thesis is optimised for the non-resonant modes entering into these ratios and the fits to these modes are already fairly advanced as also demonstrated in this work. It is expected that the final result will benefit from processing B^+ and B^0 modes together, to avoid taking corrections from the same modes as being studied and to allow suitable constraints to be applied to the partially reconstructed $B^0 \to K^{*0}\ell^+\ell^-$ background entering the $B^+ \to K^+\ell^+\ell^-$ fits.

In summary, the demonstration of the $r_{J/\psi}$ results in this thesis puts the subsequent simultaneous measurements of R_K and R_{K^*} on a very firm footing. The determination of R_K and R_{K^*} can now proceed with confidence in the efficiency determination and fitting procedure.

Appendices

Appendix A

TCKs

A.1 TCK fractions in collision data

In Tab. A.1 each of the TCKs used for a given data taking year is displayed along with the percentage of that data recorded under that TCK.

Table A.1: Data taking TCKs by year and showing the percentage of data taken under each TCK.

Data Taking Year	TCKs (% of data taken with TCK)
2011	0x790038 (30.6%) 0x760037 (29.4%) 0x730035 (19.3%) 0x6d0032 (10.1%) 0x5a0032 (5.54%) 0x790037 (3.45%) 0x740036 (0.43%) 0x360032 (0.26%) 0x4a0033 (0.20%) 0x5d0033 (0.16%) 0x5b0032 (0.15%) 0x480032 (0.15%) 0x700034 (0.12%) 0x710035 (0.09%)
2012	0x990042 (27.5%) 0xa30044 (16.2%) 0x97003d (13.7%) 0x94003d (12.8%) 0x990044 (7.1%) 0xac0046 (5.6%) 0xa10045 (3.6%) 0xa90046 (3.1%) 0x8c0040 (3.1%) 0xab0046 (2.6%) 0xa30046 (1.5%) 0xa10044 (1.3%) 0x9f0045 (1.3%) 0x860040 (0.21%) 0x990043 (0.13%) 0xa20044 (0.08%) 0x95003d (0.05%) 0x8e0040 (0.02%) 0x7f0040 (0.007%) 0xad0046 (0.003%) 0x7e003a (0.003%) 0x9a0042 (0.0009%)
2015	0x10600a2 (29.6%) 0x11400a8 (25.5%) 0x10600a3 (21.4%) 0x10800a2 (20.8%) 0x10700a1 (2.3%) 0x10600a7 (0.43%) 0x10600a6 (0.02%)
2016	0x1138160f (34.6%) 0x11361609 (25.2%) 0x11341609 (7.6%) 0x11321609 (6.8%) 0x11381612 (5.2%) 0x11291605 (5.1%) 0x11371609 (3.8%) 0x11381611 (2.7%) 0x11291603 (2.2%) 0x1138160e (2.0%) 0x11291604 (1.6%) 0x11351609 (1.4%) 0x1137160e (1.4%) 0x11381609 (0.45%)
2017	0x11611709 (36.4%) 0x11561707 (17.9%) 0x11611708 (8.7%) 0x11501705 (7.2%) 0x11611707 (7.1%) 0x11601707 (6.7%) 0x11601708 (6.4%) 0x11541707 (5.2%) 0x11501704 (1.5%) 0x11501706 (1.5%) 0x11501703 (1.3%) 0x114e1702 (0.17%) 0x114e1703 (0.10%) 0x115417a7 (0.03%)
2018	0x117a18a2 (43.5%) 0x11751801 (14.4%) 0x117a18a4 (13.4%) 0x11741801 (13.2%) 0x11771801 (10.2%) 0x11731801 (5.1%) 0x11711801 (0.22%) 0x11671801 (0.03%) 0x117718a1 (0.02%)

A.2 TCK thresholds

In the tables below the TCKs for the `prior` chain data selection with the muon mode are highlighted in grey.

2011 In 2011, as visible in Tab. A.2 the simulated `L0-TCK` thresholds are aligned to more than 98% of data on the muon and electron mode. Only a tiny fraction of data (1%) is taken with a softer/harder requirement on p_T^{L0ADC}.

Table A.2: `L0-TCK` thresholds for 2011 data taking period. Highlighted in gray the specific `TCKs` selected for the `prior` chain data selection for the muon mode.

2011-(MagUp)

L0-TCK (hex)	% data	$p_T^{L0ADC}(\mu)$	$E_T^{L0ADC}(e)$	$E_T^{L0ADC}(h)$
MC	100%	>37	>125	>175
0x730035	31.08	>37	>125	>175
0x790038	26.84	>37	>125	>175
0x760037	23.99	>37	>125	>175
0x790037	8.68	>37	>125	>175
0x5a0032	7.82	>37	>125	>175
0x740036	0.59	>40	>150	>200
0x5d0033	0.55	>20	>125	>120
0x5b0032	0.45	>37	>125	>175

2011-(MagDown)

L0-TCK (hex)	% data	$p_T^{L0ADC}(\mu)$	$E_T^{L0ADC}(e)$	$E_T^{L0ADC}(h)$
MC	100%	>37	>125	>175
0x790038	33.58	>37	>125	>175
0x760037	31.14	>37	>125	>175
0x6d0032	17.81	>37	>125	>175
0x730035	10.97	>37	>125	>175
0x5a0032	4.75	>37	>125	>175
0x360032	0.55	>37	>125	>175
0x4a0033	0.48	>20	>125	>120
0x480032	0.32	>37	>125	>175
0x700034	0.22	>37	>125	>175
0x710035	0.16	>37	>125	>175
0x5b0032	0.01	>37	>125	>175

2012 For 2012, the whole simulation is aligned in terms of the `L0-TCK` threshold for the `L0MuonDecision`. Only a tiny fraction of data (1%) is taken with a softer/harder requirement on p_T^{L0ADC}. An important misalignment is present for the electron thresholds considering all data. This is one of the main source of the large values obtained on the `L0Electron` correction maps. The thresholds are listed in Tab. A.3.

Table A.3: `L0-TCK` thresholds for 2012 data taking period. Highlighted in gray the specific `TCKs` selected for the `prior` chain data selection for the muon mode.

2012 (MagUp)

L0-TCK (hex)	% data	$p_T^{L0ADC}(\mu)$	$E_T^{L0ADC}(e)$	$E_T^{L0ADC}(h)$
MC	100%	>44	>148	>187
0x990042	37.53	>44	>136	>181
0xa30044	16.77	>44	>148	>184
0x94003d	15.45	>44	>136	>181
0x97003d	11.34	>44	>136	>181
0xa10045	7.36	>44	>148	>187
0x990044	3.03	>44	>148	>184
0xa10044	2.74	>44	>148	>184
0x9f0045	2.73	>44	>148	>187
0xa30046	1.83	>44	>143	>187
0xac0046	0.79	>44	>143	>187
0x990043	0.27	>44	>136	>181
0xa20044	0.15	>44	>148	>184
0x95003d	0.01	>44	>136	>181
0xad0046	<0.01	>44	>143	>187
0x9a0042	<0.01	>44	>136	>181

2012 (MagDown)

L0-TCK (hex)	% data	$p_T^{L0ADC}(\mu)$	$E_T^{L0ADC}(e)$	$E_T^{L0ADC}(h)$
MC	100%	>44	>148	>187
0x990042	19.81	>44	>136	>181
0xa30044	14.22	>44	>148	>184
0x97003d	13.71	>44	>136	>181
0x990044	11.1	>44	>148	>184
0xac0046	9.58	>44	>143	>187
0x8c0040	9.56	>37	>125	>175
0x94003d	9.29	>44	>136	>181
0xa90046	5.97	>44	>143	>187
0xab0046	4.85	>44	>143	>187
0xa30046	1.04	>44	>143	>187
0x860040	0.67	>37	>125	>175
0x95003d	0.09	>44	>136	>181
0x8e0040	0.07	>37	>125	>175
0x7f0040	0.02	>37	>125	>175
0x7e003a	0.01	>37	>125	>165
0xad0046	0.0	>44	>143	>187

2015 For 2015, simulation is aligned in terms of `L0-TCK` threshold for the `L0MuonDecision` fully on the Magnet Down polarity and only a 50% of data in Magnet Up is aligned. The thresholds are listed in Tab. A.4.

Table A.4: `L0-TCK` thresholds for 2015 data taking period. Highlighted in gray the specific `TCKs` selected for the `prior` chain data selection for the muon mode.

2015 (MagUp)				
TCK (hex)	% data	$p_T^{L0ADC}(\mu)$	$E_T^{L0ADC}(e)$	$E_T^{L0ADC}(h)$
MC	100%	>56	>112	>150
0x11400a8	59.08	>56	>112	>167
0x10800a2	40.92	>56	>112	>150

2015 (MagDown)				
TCK (hex)	% data	$p_T^{L0ADC}(\mu)$	$E_T^{L0ADC}(e)$	$E_T^{L0ADC}(h)$
MC	100%	>56	>112	>150
0x10600a2	47.09	>56	>112	>150
0x10600a3	39.93	>48	>95	>129
0x10800a2	6.8	>56	>112	>150
0x10700a1	4.79	>38	>75	>105
0x10600a7	1.33	>30	>58	>95
0x10600a6	0.06	>36	>70	>104

2016 For 2016, simulation is aligned in terms of `L0-TCK` threshold for the `L0MuonDecision` only for Magnet Down in 70% of the data. The thresholds are listed in Tab. A.5.

Table A.5: `L0-TCK` thresholds for 2016 data taking period. Highlighted in gray the specific `TCKs` selected for the `prior` chain data selection for the muon mode.

2016 (MagUp)				
TCK (hex)	% data	$p_T^{L0ADC}(\mu)$	$E_T^{L0ADC}(e)$	$E_T^{L0ADC}(h)$
MC	100%	>36	>100	>156
0x11361609	50.06	>26	>98	>154
0x11341609	15.61	>26	>98	>154
0x11321609	14.94	>26	>98	>154
0x11381612	11.05	>32	>109	>162
0x11381611	5.59	>30	>109	>162
0x11351609	2.75	>26	>98	>154

2016 (MagDown)				
TCK (hex)	% data	$p_T^{L0ADC}(\mu)$	$E_T^{L0ADC}(e)$	$E_T^{L0ADC}(h)$
MC	100%	>36	>100	>156
0x1138160f	69.13	>36	>100	>156
0x11291605	9.09	>30	>108	>154
0x11371609	6.93	>26	>98	>154
0x11291603	4.71	>22	>88	>134
0x1138160e	3.67	>30	>108	>154
0x11291604	3.18	>26	>94	>148
0x1137160e	2.41	>30	>108	>154
0x11381609	0.87	>26	>98	>154

2017 For 2017, simulation is aligned in terms of `L0-TCK` threshold for the `L0MuonDecision` only on $\sim$ 40% of data in both polarities. The thresholds are listed in Tab. A.6.

Table A.6: `L0-TCK` thresholds for 2017 data taking period. Highlighted in gray the specific `TCKs` selected for the `prior` chain data selection for the muon mode.

2017 (MagUp)				
TCK (hex)	% data	$p_T^{L0ADC}(\mu)$	$E_T^{L0ADC}(e)$	$E_T^{L0ADC}(h)$
MC	100	>28	>88	>144
0x11611709	38.83	>28	>88	>144
0x11501705	14.47	>30	>108	>154
0x11601708	13.53	>22	>88	>134
0x11601707	13.16	>34	>96	>155
0x11561707	9.51	>34	>96	>155
0x11501704	3.61	>26	>94	>148
0x11501703	3.35	>22	>88	>134
0x11501706	2.82	>38	>112	>162
0x114e1702	0.46	>14	>78	>124
0x114e1703	0.25	>22	>88	>134

2017 (MagDown)				
TCK (hex)	% data	$p_T^{L0ADC}(\mu)$	$E_T^{L0ADC}(e)$	$E_T^{L0ADC}(h)$
MC	100	>28	>88	>144
0x11611709	34.27	>28	>88	>144
0x11561707	24.13	>34	>96	>155
0x11611708	17.7	>22	>88	>134
0x11611707	13.77	>34	>96	>155
0x11541707	10.06	>34	>96	>155
0x115417a7	0.06	>34	>96	>155

2018 For 2018, simulation is fully aligned to data in both polarities. The thresholds are listed in Tab. A.7.

Table A.7: `L0-TCK` thresholds for 2018 data taking period. Highlighted in gray the specific `TCKs` selected for the `prior` chain data selection for the muon mode.

2018 (MagUp)				
TCK (hex)	% data	$p_T^{L0ADC}(\mu)$	$E_T^{L0ADC}(e)$	$E_T^{L0ADC}(h)$
MC	100%	>35	>99	>158
0x117a18a2	28.06	>35	>99	>158
0x117a18a4	27.31	>35	>99	>158
0x11771801	17.18	>35	>99	>158
0x11741801	15.81	>35	>99	>158
0x11731801	10.72	>35	>99	>158
0x11711801	0.47	>35	>99	>158
0x11671801	0.4	>35	>99	>158
0x117718a1	0.05	>35	>99	>158

2018 (MagDown)				
TCK (hex)	% data	$p_T^{L0ADC}(\mu)$	$E_T^{L0ADC}(e)$	$E_T^{L0ADC}(h)$
MC	100%	>35	>99	>158
0x117a18a2	53.57	>35	>99	>158
0x11751801	27.46	>35	>99	>158
0x117a18a4	8.75	>35	>99	>158
0x11741801	8.71	>35	>99	>158
0x11771801	1.52	>35	>99	>158

Appendix B

sPlot

The various fit models in this section are described in Sec. 7

The ${}_s\mathcal{W}eight$ procedure is only applied to J/ψ resonant modes, a first set of fits is performed in order to determine parameters such as the mass shift and sigma scale which encode differences between collision data and simulation. A second fit is then performed with all the parameters fixed except for the yields which allows for the extract of the ${}_s\mathcal{W}eights$.

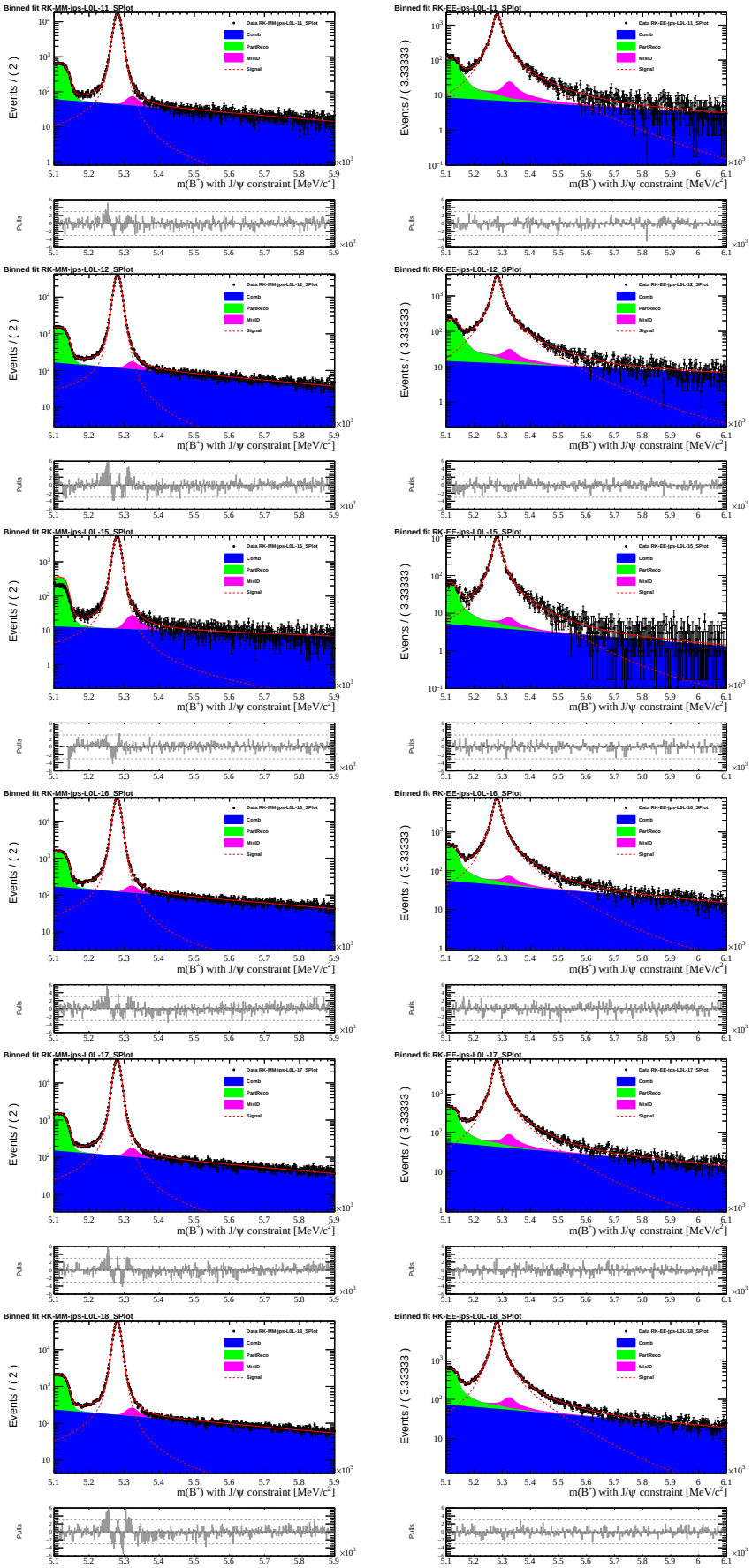


Figure 1: ${}_s\mathcal{P}lot$ fits to $m(K\ell\ell)_{J/\psi}$ for $B^+ \to K^+ J/\psi (\to \mu^+\mu^-)$ (left) and $B^+ \to K^+ J/\psi (\to e^+e^-)$ (right) for L0L!. 2011-2018 (top-bottom)

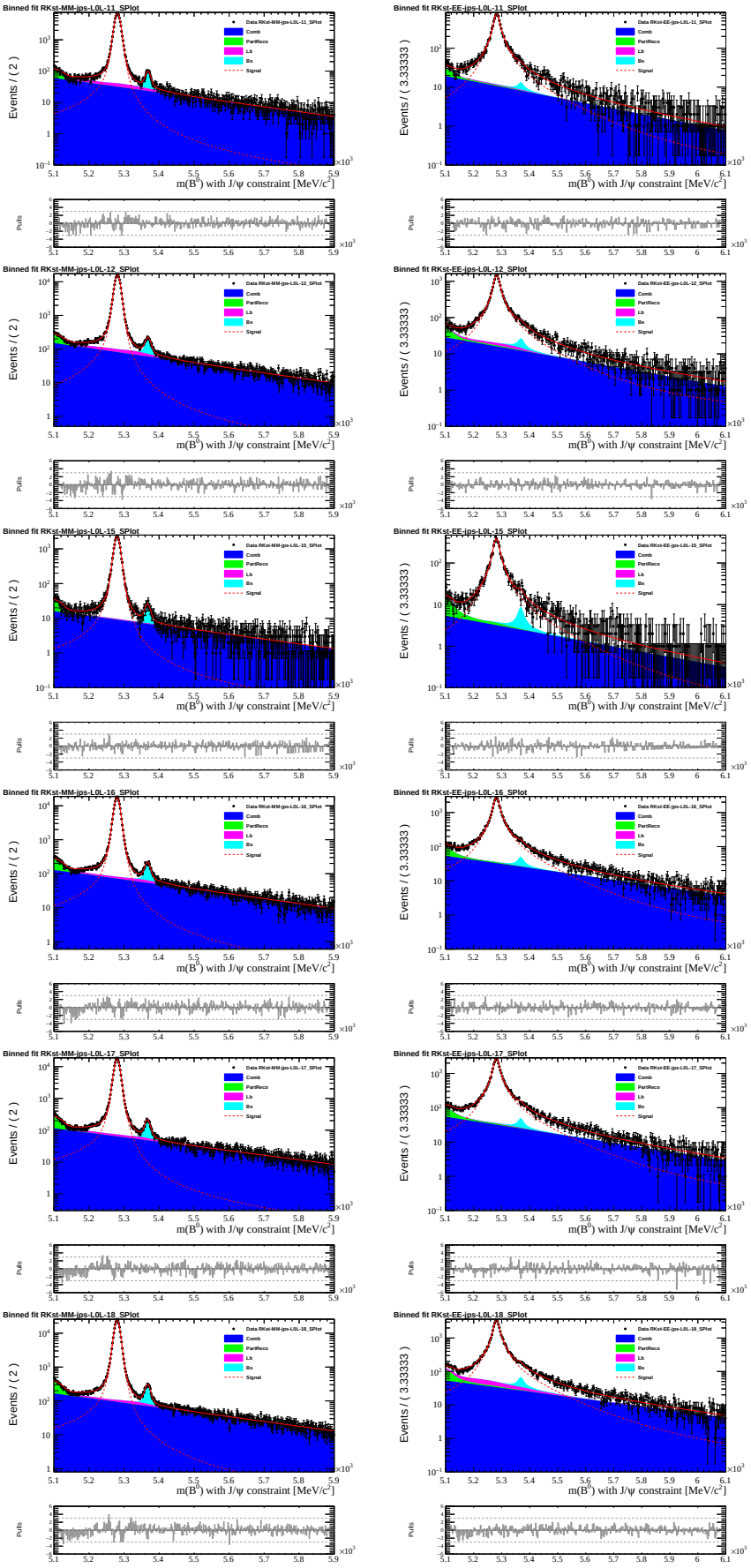


Figure 2: ${}_s\mathcal{P}lot$ fits to $m(K\ell\ell)_{J/\psi}$ for $B^0 \to K^{*0} J/\psi (\to \mu^+\mu^-)$ (left) and $B^0 \to K^{*0} J/\psi (\to e^+e^-)$ (right) for L0L!. 2011-2018 (top-bottom)

Appendix C

Exclusive Backgrounds

C.1 Expected Background Calculation

To calculate the residual amount of a physical background after the selection Eq. 4.2 is used. However, in the limiting case where all generated events n of a given background MC sample are rejected by the selection ($k \to 0$), the efficiency estimate based on binomial statistics is flawed:

$$\bar{\epsilon} = \frac{k}{n} \to 0 \quad \text{and} \quad Var(\bar{\epsilon}) = \frac{k(n-k)}{n^3} \to 0\,. \tag{C.1}$$

Nevertheless, it is possible to determine an upper limit on the selection efficiency and thus on the expected remaining yield of this background. This is done by following an approach described in Ref. [139] in which binomial statistics are combined with the Bayesian theorem to calculate the probability density function

$$\mathcal{P}(\epsilon; k, n) = \frac{(n+1)!}{k!(n-k)!}\epsilon^k(1-\epsilon)^{n-k} \tag{C.2}$$

of the selection efficiency ϵ for a given n and k. This PDF yields the following equations for efficiency and variance:

$$\bar{\epsilon} = \frac{k+1}{n+2} \quad \text{and} \quad Var(\bar{\epsilon}) = \frac{(k+1)(k+2)}{(n+2)(n+3)} - \frac{(k+1)^2}{(n+2^2)}\,, \tag{C.3}$$

which yield both non-zero results in case of $k \to 0$. At large n and $k \gg 1$ the

resulting efficiency approaches the binomial result

$$\text{mode}(\epsilon) = \frac{k}{n}\,. \tag{C.4}$$

In order to estimate the expected background yield, the selection efficiency is sampled from $\mathcal{P}(\epsilon; k, n)$ and α_{norm} (from Eq. 4.2) is sampled from a Gaussian distribution in an independent way, as no correlations between the two distributions are expected. To precisely determine the expected background yield at a chosen confidence level, a large set of sampling iterations is required. Since it is computationally expensive to sample from $\mathcal{P}(\epsilon; k, n)$ directly, a variable transformation $\epsilon \to \lambda := \epsilon n$ is performed yielding a Poisson distribution in the limit of $n \to \infty$

$$\mathcal{P}(\lambda; k, n) = \frac{\lambda^k}{k!} e^{-\lambda}\,, \tag{C.5}$$

which is a valid approximation considering the large number of generated MC events compared to λ since the true selection efficiencies are always small.

For the Gaussian distribution the width is set to the total error of α_{norm}, which is a composition of:

- The uncertainties on the branching ratios taken from PDG
- The uncertainty on the control channel yield taken from the fit
- The uncertainty on the control channel efficiency determined on simulation
- The uncertainties on the production fractions.

Figure C.1 shows exemplary sampled distributions for α_{norm}, ϵ and the expected number of background events $\mathrm{N_{bkg}}$ for hadronic double swaps from $B^0 \to K^{*0} J/\psi\,(\to e^+e^-)$ in 2016.

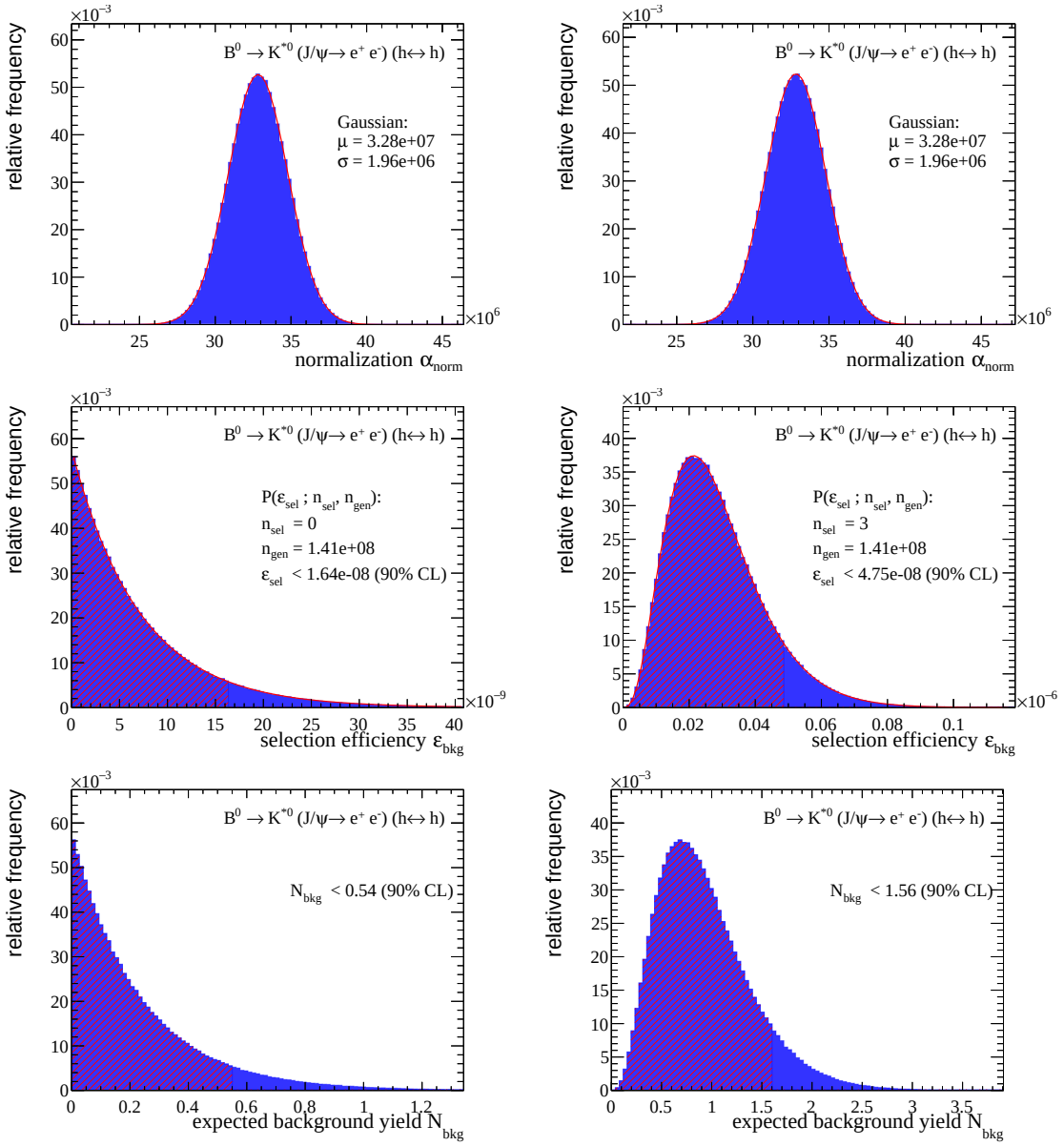


Figure C.1: Sampled distributions for α_{norm}, ϵ and $\mathrm{N_{bkg}}$ for hadronic double swaps from $B^0 \to K^{*0} J/\psi (\to e^+e^-)$ in 2016. Left: low-q^2 region (0 selected events). Right: central-q^2 region (3 selected events). The red hatched area corresponds to the one-sided 90% confidence interval from which upper confidence bound at 90% CL is extracted.

C.2 Expected background yields

R_K EE

Table C.1: Expected background efficiencies and yields in control region for 2012 EE R_K. If nPassed is < 20 a limit at 90% CL is given, else the expected yield is calculated.

MC 2012	q^2 bin	nPassed	efficiency	bkg yield/limit
$B^+ \to K^+(J/\psi \to e^+e^-)(h \leftrightarrow e)$	J/ψ	0	$< 9.29e-08$	< 3.01
$B^+ \to K^+(\psi(2)S \to e^+e^-)(h \leftrightarrow e)$	J/ψ	0	$< 3.84e-07$	< 1.01
$B^+ \to (D^0 \to K^+e^-\bar{\nu}_e)\pi^+$	J/ψ	0	$< 1.69e-09$	< 0.15
$B^+ \to (D^0 \to K^+\pi^-)e^+\nu_e$	J/ψ	0	$< 7.41e-10$	< 0.36
$B^0 \to K^{*0}e^+e^-$	J/ψ	15516	0.000237	87.4 ± 16.4
$B^0 \to K^{*0}(J/\psi \to e^+e^-)$	J/ψ	102356	0.00134	36300 ± 1940

Table C.2: Expected background efficiencies and yields in signal region for 2012 EE R_K. If nPassed is < 20 a limit at 90% CL is given, else the expected yield is calculated.

MC 2012	q^2 bin	nPassed	efficiency	bkg yield/limit
$B^+ \to K^+(J/\psi \to e^+e^-)(h \leftrightarrow e)$	low	0	$< 9.29e-08$	< 3.01
$B^+ \to K^+(J/\psi \to e^+e^-)(h \leftrightarrow e)$	central	0	$< 9.29e-08$	< 3.01
$B^+ \to K^+(\psi(2)S \to e^+e^-)(h \leftrightarrow e)$	low	0	$< 3.84e-07$	< 1.01
$B^+ \to K^+(\psi(2)S \to e^+e^-)(h \leftrightarrow e)$	central	0	$< 3.84e-07$	< 1.01
$B^+ \to (D^0 \to K^+e^-\bar{\nu}_e)\pi^+$	low	0	$< 1.69e-09$	< 0.15
$B^+ \to (\bar{D}^0 \to K^+e^-\bar{\nu}_e)\pi^+$	central	0	$< 1.69e-09$	< 0.15
$B^+ \to (D^0 \to K^+\pi^-)e^+\nu_e$	low	0	$< 7.41e-10$	< 0.36
$B^+ \to (\bar{D}^0 \to K^+\pi^-)e^+\nu_e$	central	0	$< 7.41e-10$	< 0.36
$B^0 \to K^{*0}e^+e^-$	low	968	1.48e-05	5.46 ± 1.04
$B^0 \to K^{*0}e^+e^-$	central	4016	6.13e-05	22.6 ± 4.26
$B^0 \to K^{*0}(J/\psi \to e^+e^-)$	low	0	$< 3.01e-08$	< 0.82
$B^0 \to K^{*0}(J/\psi \to e^+e^-)$	central	7	$< 1.54e-07$	< 4.19

Table C.3: Expected background efficiencies and yields in control region for 2018 EE R_K. If nPassed is < 20 a limit at 90% CL is given, else the expected yield is calculated.

MC 2018	q^2 bin	nPassed	efficiency	bkg yield/limit
$B^+ \to \pi^+(J/\psi \to e^+e^-)$	J/ψ	4325	0.000354	564 ± 28.1
$B^+ \to K^+(J/\psi \to e^+e^-)(h \leftrightarrow e)$	J/ψ	0	$< 3.98e-08$	< 1.65
$B^+ \to K^+(\psi(2)S \to e^+e^-)(h \leftrightarrow e)$	J/ψ	0	$< 6.71e-08$	< 0.23
$B^+ \to (D^0 \to K^+e^-\bar{\nu}_e)\pi^+$	J/ψ	0	$< 1.67e-09$	< 0.19
$B^+ \to (D^0 \to K^+\pi^-)e^+\nu_e$	J/ψ	0	$< 7.42e-10$	< 0.47
$B^+ \to (D^0 \to K^+e^-\bar{\nu}_e)e^+\nu_e$	J/ψ	20	6.27e-10	0.38 ± 0.085
$B^0 \to K^{*0}e^+e^-$	J/ψ	18877	0.000408	193 ± 36.3
$B^0 \to K^{*0}(J/\psi \to e^+e^-)$	J/ψ	141266	0.00234	81500 ± 4360
$B^0 \to K^{*0}(\psi(2)S \to e^+e^-)$	J/ψ	4409	0.000146	313 ± 25.3

Table C.4: Expected background efficiencies and yields in signal region for 2018 EE R_K. If nPassed is < 20 a limit at 90% CL is given, else the expected yield is calculated.

MC 2018	q^2 bin	nPassed	efficiency	bkg yield/limit
$B^+ \to \pi^+(J/\psi \to e^+e^-)$	low	0	$< 1.88e-07$	< 0.3
$B^+ \to \pi^+(J/\psi \to e^+e^-)$	central	19	$< 2.12e-06$	< 3.4
$B^+ \to K^+(J/\psi \to e^+e^-)(h \leftrightarrow e)$	low	0	$< 3.98e-08$	< 1.65
$B^+ \to K^+(J/\psi \to e^+e^-)(h \leftrightarrow e)$	central	0	$< 3.98e-08$	< 1.65
$B^+ \to K^+(\psi(2)S \to e^+e^-)(h \leftrightarrow e)$	low	0	$< 6.71e-08$	< 0.23
$B^+ \to K^+(\psi(2)S \to e^+e^-)(h \leftrightarrow e)$	central	0	$< 6.71e-08$	< 0.23
$B^+ \to (D^0 \to K^+e^-\bar{\nu}_e)\pi^+$	low	0	$< 1.67e-09$	< 0.19
$B^+ \to (\bar{D}^0 \to K^+e^-\bar{\nu}_e)\pi^+$	central	0	$< 1.67e-09$	< 0.19
$B^+ \to (D^0 \to K^+\pi^-)e^+\nu_e$	low	0	$< 7.42e-10$	< 0.47
$B^+ \to (\bar{D}^0 \to K^+\pi^-)e^+\nu_e$	central	0	$< 7.42e-10$	< 0.47
$B^+ \to (D^0 \to K^+e^-\bar{\nu}_e)e^+\nu_e$	low	2	$< 1.67e-10$	< 0.096
$B^+ \to (\bar{D}^0 \to K^+e^-\bar{\nu}_e)e^+\nu_e$	central	4	$< 2.5e-10$	< 0.14
$B^0 \to K^{*0}e^+e^-$	low	1359	2.94e-05	13.9 ± 2.64
$B^0 \to K^{*0}e^+e^-$	central	5214	0.000113	53.3 ± 10
$B^0 \to K^{*0}(J/\psi \to e^+e^-)$	low	0	$< 3.82e-08$	< 1.33
$B^0 \to K^{*0}(J/\psi \to e^+e^-)$	central	5	$< 1.53e-07$	< 5.34
$B^0 \to K^{*0}(\psi(2)S \to e^+e^-)$	low	0	$< 7.62e-08$	< 0.16
$B^0 \to K^{*0}(\psi(2)S \to e^+e^-)$	central	0	$< 7.62e-08$	< 0.16

R_K MM

Table C.5: Expected background efficiencies and yields in control region for 2012 MM R_K. If nPassed is < 20 a limit at 90% CL is given, else the expected yield is calculated.

MC 2012	q^2 bin	nPassed	efficiency	bkg yield/limit
$B^+ \to \pi^+(J/\psi \to \mu^+\mu^-)$	J/ψ	24820	0.00193	2400 ± 114
$B^+ \to K^+(J/\psi \to \mu^+\mu^-)(h \leftrightarrow \mu)$	J/ψ	0	$< 1.78e-07$	< 5.75
$B^+ \to K^+(\psi(2)S \to \mu^+\mu^-)(h \leftrightarrow \mu)$	J/ψ	0	$< 1.88e-07$	< 0.5
$B^+ \to (D^0 \to K^+\mu^-\bar{\nu}_\mu)\pi^+$	J/ψ	0	$< 3.32e-06$	< 284
$B^0 \to K^{*0}\mu^+\mu^-$	J/ψ	826	5.25e-05	17.7 ± 1.29
$B^0 \to K^{*0}(J/\psi \to \mu^+\mu^-)$	J/ψ	25128	0.000895	24200 ± 1300
$B^0 \to K^{*0}(\psi(2)S \to \mu^+\mu^-)$	J/ψ	0	$< 2.64e-07$	< 0.45

Table C.6: Expected background efficiencies and yields in signal region for 2012 MM R_K. If nPassed is < 20 a limit at 90% CL is given, else the expected yield is calculated.

MC 2012	q^2 bin	nPassed	efficiency	bkg yield/limit
$B^+ \to \pi^+(J/\psi \to \mu^+\mu^-)$	low	0	$< 1.79e-07$	< 0.22
$B^+ \to \pi^+(J/\psi \to \mu^+\mu^-)$	central	0	$< 1.79e-07$	< 0.22
$B^+ \to K^+(J/\psi \to \mu^+\mu^-)(h \leftrightarrow \mu)$	low	0	$< 1.78e-07$	< 5.75
$B^+ \to K^+(J/\psi \to \mu^+\mu^-)(h \leftrightarrow \mu)$	central	0	$< 1.78e-07$	< 5.75
$B^+ \to K^+(\psi(2)S \to \mu^+\mu^-)(h \leftrightarrow \mu)$	low	0	$< 1.88e-07$	< 0.5
$B^+ \to K^+(\psi(2)S \to \mu^+\mu^-)(h \leftrightarrow \mu)$	central	0	$< 1.88e-07$	< 0.5
$B^+ \to (D^0 \to K^+\mu^-\bar{\nu}_\mu)\pi^+$	low	0	$< 3.32e-06$	< 284
$B^+ \to (D^0 \to K^+\mu^-\bar{\nu}_\mu)\pi^+$	central	0	$< 3.32e-06$	< 284
$B^0 \to K^{*0}\mu^+\mu^-$	low	15	$< 1.35e-06$	< 0.46
$B^0 \to K^{*0}\mu^+\mu^-$	central	121	7.69e-06	2.61 ± 0.29
$B^0 \to K^{*0}(J/\psi \to \mu^+\mu^-)$	low	0	$< 8.19e-08$	< 2.22
$B^0 \to K^{*0}(J/\psi \to \mu^+\mu^-)$	central	0	$< 8.19e-08$	< 2.22
$B^0 \to K^{*0}(\psi(2)S \to \mu^+\mu^-)$	low	0	$< 2.64e-07$	< 0.45
$B^0 \to K^{*0}(\psi(2)S \to \mu^+\mu^-)$	central	0	$< 2.64e-07$	< 0.45

Table C.7: Expected background efficiencies and yields in control region for 2018 MM R_K. If nPassed is < 20 a limit at 90% CL is given, else the expected yield is calculated.

MC 2018	q^2 bin	nPassed	efficiency	bkg yield/limit
$B^+ \to \pi^+(J/\psi \to \mu^+\mu^-)$	J/ψ	18402	0.00152	2420 ± 116
$B^+ \to K^+(J/\psi \to \mu^+\mu^-)(h \leftrightarrow \mu)$	J/ψ	0	$< 3.99e-08$	< 1.65
$B^+ \to K^+(\psi(2)S \to \mu^+\mu^-)(h \leftrightarrow \mu)$	J/ψ	0	$< 5.07e-08$	< 0.17
$B^+ \to (D^0 \to K^+\mu^-\bar{\nu}_\mu)\pi^+$	J/ψ	0	$< 2.1e-07$	< 23.1
$B^+ \to (D^0 \to K^+\pi^-)\mu^+\nu_\mu$	J/ψ	0	$< 2.11e-07$	< 133
$B^0 \to K^{*0}\mu^+\mu^-$	J/ψ	1523	6.46e-05	27.9 ± 1.92
$B^0 \to K^{*0}(J/\psi \to \mu^+\mu^-)$	J/ψ	65938	0.0011	38200 ± 2040
$B^0 \to K^{*0}(\psi(2)S \to \mu^+\mu^-)$	J/ψ	0	$< 1.56e-07$	< 0.34

Table C.8: Expected background efficiencies and yields in signal region for 2018 MM R_K. If nPassed is < 20 a limit at 90% CL is given, else the expected yield is calculated.

MC 2018	q^2 bin	nPassed	efficiency	bkg yield/limit
$B^+ \to \pi^+(J/\psi \to \mu^+\mu^-)$	low	0	$< 1.9e-07$	< 0.3
$B^+ \to \pi^+(J/\psi \to \mu^+\mu^-)$	central	0	$< 1.9e-07$	< 0.3
$B^+ \to K^+(J/\psi \to \mu^+\mu^-)(h \leftrightarrow \mu)$	low	0	$< 3.99e-08$	< 1.65
$B^+ \to K^+(J/\psi \to \mu^+\mu^-)(h \leftrightarrow \mu)$	central	0	$< 3.99e-08$	< 1.65
$B^+ \to K^+(\psi(2)S \to \mu^+\mu^-)(h \leftrightarrow \mu)$	low	0	$< 5.07e-08$	< 0.17
$B^+ \to K^+(\psi(2)S \to \mu^+\mu^-)(h \leftrightarrow \mu)$	central	0	$< 5.07e-08$	< 0.17
$B^+ \to (D^0 \to K^+\mu^-\bar{\nu}_\mu)\pi^+$	low	0	$< 2.1e-07$	< 23.1
$B^+ \to (\bar{D}^0 \to K^+\mu^-\bar{\nu}_\mu)\pi^+$	central	0	$< 2.1e-07$	< 23.1
$B^+ \to (D^0 \to K^+\pi^-)\mu^+\nu_\mu$	low	0	$< 2.11e-07$	< 133
$B^+ \to (\bar{D}^0 \to K^+\pi^-)\mu^+\nu_\mu$	central	0	$< 2.11e-07$	< 133
$B^0 \to K^{*0}\mu^+\mu^-$	low	44	1.87e-06	0.82 ± 0.13
$B^0 \to K^{*0}\mu^+\mu^-$	central	272	1.15e-05	5 ± 0.44
$B^0 \to K^{*0}(J/\psi \to \mu^+\mu^-)$	low	0	$< 3.83e-08$	< 1.33
$B^0 \to K^{*0}(J/\psi \to \mu^+\mu^-)$	central	0	$< 3.83e-08$	< 1.33
$B^0 \to K^{*0}(\psi(2)S \to \mu^+\mu^-)$	low	0	$< 1.56e-07$	< 0.34
$B^0 \to K^{*0}(\psi(2)S \to \mu^+\mu^-)$	central	0	$< 1.56e-07$	< 0.34

R_{K^*} EE

Table C.9: Expected background efficiencies and yields in control region for 2012 EE R_{K^*}. If nPassed is < 20 a limit at 90% CL is given, else the expected yield is calculated.

MC 2012	q^2 bin	nPassed	efficiency	bkg yield/limit
$B_s^0 \to \phi(J/\psi \to e^+e^-)$	J/ψ	633	5.23e-05	273 ± 33.9
$B_s^0 \to \phi e^+e^-$	J/ψ	382	1.04e-05	0.69 ± 0.13
$\Lambda_b^0 \to pK(\psi(2)S \to e^+e^-)$	J/ψ	221.25	1.11e-05	17.1 ± 4.22
$\Lambda_b^0 \to pK(J/\psi \to e^+e^-)$	J/ψ	1303.75	6.44e-05	480 ± 116
$\Lambda_b^0 \to pKe^+e^-$	J/ψ	333.75	8.58e-06	1.04 ± 0.24
$B^0 \to K^{*0}(J/\psi \to e^+e^-)(h \leftrightarrow h)$	J/ψ	920	1.2e-05	375 ± 25.6
$B^0 \to K^{*0}(J/\psi \to e^+e^-)(h \leftrightarrow e)$	J/ψ	0	$< 3.01e-08$	< 0.94
$B^0 \to K^{*0}(\psi(2)S \to e^+e^-)(h \leftrightarrow h)$	J/ψ	27	1.62e-06	3.22 ± 0.67
$B^0 \to K^{*0}(\psi(2)S \to e^+e^-)(h \leftrightarrow e)$	J/ψ	0	$< 1.38e-07$	< 0.27
$B^0 \to (D^- \to K^{*0}e^-\bar{\nu}_e)e^+\nu_e$	J/ψ	5246	7.67e-07	411 ± 32
$B^+ \to K\pi\pi(J/\psi \to e^+e^-)$	J/ψ	8772.83	0.000339	10100 ± 1690
$B^+ \to K\pi\pi e^+e^-$	J/ψ	5705.54	4.2e-05	11.1 ± 1.16
$B^+ \to K^+(J/\psi \to e^+e^-)$	J/ψ	27	1.09e-06	42 ± 8.25
$B^+ \to K^+e^+e^-$	J/ψ	1	$< 1.29e-06$	< 0.44

Table C.10: Expected background efficiencies and yields in signal region for 2012 EE R_{K^*}. If nPassed is < 20 a limit at 90% CL is given, else the expected yield is calculated.

MC 2012	q^2 bin	nPassed	efficiency	bkg yield/limit
$B_s^0 \to \phi(J/\psi \to e^+e^-)$	low	0	$< 1.9e-07$	< 0.99
$B_s^0 \to \phi(J/\psi \to e^+e^-)$	central	1	$< 3.21e-07$	< 1.69
$B_s^0 \to \phi e^+e^-$	low	134	3.66e-06	0.24 ± 0.047
$B_s^0 \to \phi e^+e^-$	central	292	7.98e-06	0.53 ± 0.097
$\Lambda_b^0 \to pK(\psi(2)S \to e^+e^-)$	low	0	$< 1.16e-07$	< 0.18
$\Lambda_b^0 \to pK(\psi(2)S \to e^+e^-)$	central	0	$< 1.16e-07$	< 0.18
$\Lambda_b^0 \to pK(J/\psi \to e^+e^-)$	low	0	$< 1.14e-07$	< 0.86
$\Lambda_b^0 \to pK(J/\psi \to e^+e^-)$	central	2.5	$< 2.97e-07$	< 2.31
$\Lambda_b^0 \to pKe^+e^-$	low	60	1.54e-06	0.19 ± 0.049
$\Lambda_b^0 \to pKe^+e^-$	central	245	6.3e-06	0.76 ± 0.18
$B^0 \to K^{*0}(J/\psi \to e^+e^-)(h \leftrightarrow h)$	low	0	$< 3.01e-08$	< 0.94
$B^0 \to K^{*0}(J/\psi \to e^+e^-)(h \leftrightarrow h)$	central	0	$< 3.01e-08$	< 0.94
$B^0 \to K^{*0}(J/\psi \to e^+e^-)(h \leftrightarrow e)$	low	0	$< 3.01e-08$	< 0.94
$B^0 \to K^{*0}(J/\psi \to e^+e^-)(h \leftrightarrow e)$	central	0	$< 3.01e-08$	< 0.94
$B^0 \to K^{*0}(\psi(2)S \to e^+e^-)(h \leftrightarrow h)$	low	0	$< 1.38e-07$	< 0.27
$B^0 \to K^{*0}(\psi(2)S \to e^+e^-)(h \leftrightarrow h)$	central	0	$< 1.38e-07$	< 0.27
$B^0 \to K^{*0}(\psi(2)S \to e^+e^-)(h \leftrightarrow e)$	low	0	$< 1.38e-07$	< 0.27
$B^0 \to K^{*0}(\psi(2)S \to e^+e^-)(h \leftrightarrow e)$	central	0	$< 1.38e-07$	< 0.27
$B^0 \to (D^- \to K^{*0}e^-\bar{\nu}_e)e^+\nu_e$	low	14	$< 2.94e-09$	< 1.6
$B^0 \to (D^- \to K^{*0}e^-\bar{\nu}_e)e^+\nu_e$	central	23	3.36e-09	1.88 ± 0.41
$B^+ \to K\pi\pi(J/\psi \to e^+e^-)$	low	0	$< 8.89e-08$	< 2.66
$B^+ \to K\pi\pi(J/\psi \to e^+e^-)$	central	0	$< 8.89e-08$	< 2.66
$B^+ \to K\pi\pi e^+e^-$	low	396.17	2.92e-06	0.77 ± 0.089
$B^+ \to K\pi\pi e^+e^-$	central	1704.05	1.25e-05	3.32 ± 0.35
$B^+ \to K^+(J/\psi \to e^+e^-)$	low	0	$< 9.29e-08$	< 3.45
$B^+ \to K^+(J/\psi \to e^+e^-)$	central	0	$< 9.29e-08$	< 3.45
$B^+ \to K^+e^+e^-$	low	0	$< 7.63e-07$	< 0.26
$B^+ \to K^+e^+e^-$	central	0	$< 7.63e-07$	< 0.26

Table C.11: Expected background efficiencies and yields in control region for 2016 EE R_{K^*}. If nPassed is < 20 a limit at 90% CL is given, else the expected yield is calculated.

MC 2016	q^2 bin	nPassed	efficiency	bkg yield/limit
$B_s^0 \to \phi(J/\psi \to e^+e^-)$	J/ψ	836	2.95e-05	162 ± 19.9
$B_s^0 \to \phi e^+e^-$	J/ψ	192	7.03e-06	0.49 ± 0.093
$\Lambda_b^0 \to pK(\psi(2)S \to e^+e^-)$	J/ψ	27.5	1.89e-06	3.17 ± 0.96
$\Lambda_b^0 \to pK(J/\psi \to e^+e^-)$	J/ψ	652.5	2.18e-05	172 ± 41.9
$\Lambda_b^0 \to pKe^+e^-$	J/ψ	76.25	2.53e-06	0.33 ± 0.082
$B^0 \to K^{*0}(J/\psi \to e^+e^-)(h \leftrightarrow h)$	J/ψ	620	4.4e-06	145 ± 10.4
$B^0 \to K^{*0}(J/\psi \to e^+e^-)(h \leftrightarrow e)$	J/ψ	0	$< 1.64e-08$	< 0.54
$B^0 \to K^{*0}(\psi(2)S \to e^+e^-)(h \leftrightarrow h)$	J/ψ	10	$< 4.87e-07$	< 1
$B^0 \to K^{*0}(\psi(2)S \to e^+e^-)(h \leftrightarrow e)$	J/ψ	0	$< 7.28e-08$	< 0.15
$B^0 \to (D^- \to K^{*0}e^-\bar{\nu}_e)e^+\nu_e$	J/ψ	11590	1.66e-06	939 ± 72.6
$B^0 \to (D^{*-} \to (D^0 \to K^+\pi^-)\pi)e^+\nu_e$	J/ψ	0	$< 2.28e-09$	< 1.9
$B^+ \to K\pi\pi(J/\psi \to e^+e^-)$	J/ψ	23265	0.000566	17800 ± 2960
$B^+ \to K^+(J/\psi \to e^+e^-)$	J/ψ	24	8.3e-07	33.8 ± 7.01
$B^+ \to K^+e^+e^-$	J/ψ	3	$< 1.17e-06$	< 0.43

Table C.12: Expected background efficiencies and yields in signal region for 2016 EE R_{K^*}. If nPassed is < 20 a limit at 90% CL is given, else the expected yield is calculated.

MC 2016	q^2 bin	nPassed	efficiency	bkg yield/limit
$B_s^0 \to \phi(J/\psi \to e^+e^-)$	low	0	$< 8.12e-08$	< 0.45
$B_s^0 \to \phi(J/\psi \to e^+e^-)$	central	1	$< 1.37e-07$	< 0.76
$B_s^0 \to \phi e^+e^-$	low	57	2.09e-06	0.15 ± 0.032
$B_s^0 \to \phi e^+e^-$	central	141	5.16e-06	0.36 ± 0.07
$\Lambda_b^0 \to pK(\psi(2)S \to e^+e^-)$	low	0	$< 1.58e-07$	< 0.26
$\Lambda_b^0 \to pK(\psi(2)S \to e^+e^-)$	central	0	$< 1.58e-07$	< 0.26
$\Lambda_b^0 \to pK(J/\psi \to e^+e^-)$	low	0	$< 7.69e-08$	< 0.61
$\Lambda_b^0 \to pK(J/\psi \to e^+e^-)$	central	0	$< 7.69e-08$	< 0.61
$\Lambda_b^0 \to pKe^+e^-$	low	20	6.64e-07	0.089 ± 0.028
$\Lambda_b^0 \to pKe^+e^-$	central	62.5	2.08e-06	0.27 ± 0.069
$B^0 \to K^{*0}(J/\psi \to e^+e^-)(h \leftrightarrow h)$	low	0	$< 1.64e-08$	< 0.54
$B^0 \to K^{*0}(J/\psi \to e^+e^-)(h \leftrightarrow h)$	central	3	$< 4.75e-08$	< 1.56
$B^0 \to K^{*0}(J/\psi \to e^+e^-)(h \leftrightarrow e)$	low	0	$< 1.64e-08$	< 0.54
$B^0 \to K^{*0}(J/\psi \to e^+e^-)(h \leftrightarrow e)$	central	0	$< 1.64e-08$	< 0.54
$B^0 \to K^{*0}(\psi(2)S \to e^+e^-)(h \leftrightarrow h)$	low	0	$< 7.28e-08$	< 0.15
$B^0 \to K^{*0}(\psi(2)S \to e^+e^-)(h \leftrightarrow h)$	central	0	$< 7.28e-08$	< 0.15
$B^0 \to K^{*0}(\psi(2)S \to e^+e^-)(h \leftrightarrow e)$	low	0	$< 7.28e-08$	< 0.15
$B^0 \to K^{*0}(\psi(2)S \to e^+e^-)(h \leftrightarrow e)$	central	0	$< 7.28e-08$	< 0.15
$B^0 \to (D^- \to K^{*0}e^-\bar{\nu}_e)e^+\nu_e$	low	16	$< 3.22e-09$	< 1.84
$B^0 \to (D^- \to K^{*0}e^-\bar{\nu}_e)e^+\nu_e$	central	52	7.45e-09	4.29 ± 0.68
$B^0 \to (D^{*-} \to (\bar{D^0} \to K^+\pi^-)\pi)e^+\nu_e$	low	0	$< 2.28e-09$	< 1.9
$B^0 \to (D^{*-} \to (\bar{D^0} \to K^+\pi^-)\pi)e^+\nu_e$	central	0	$< 2.28e-09$	< 1.9
$B^+ \to K\pi\pi(J/\psi \to e^+e^-)$	low	0	$< 5.6e-08$	< 1.77
$B^+ \to K\pi\pi(J/\psi \to e^+e^-)$	central	0	$< 5.6e-08$	< 1.77
$B^+ \to K^+(J/\psi \to e^+e^-)$	low	0	$< 7.96e-08$	< 3.12
$B^+ \to K^+(J/\psi \to e^+e^-)$	central	1	$< 1.34e-07$	< 5.28
$B^+ \to K^+e^+e^-$	low	0	$< 4.04e-07$	< 0.14
$B^+ \to K^+e^+e^-$	central	0	$< 4.04e-07$	< 0.14

R_{K^*} MM

Table C.13: Expected background efficiencies and yields in control region for 2012 MM R_{K^*}. If nPassed is < 20 a limit at 90% CL is given, else the expected yield is calculated.

MC 2012	q^2 bin	nPassed	efficiency	bkg yield/limit
$B_s^0 \to \phi(J/\psi \to \mu^+\mu^-)$	J/ψ	2437	0.000203	1060 ± 126
$B_s^0 \to \phi\mu^+\mu^-$	J/ψ	188	1.56e-05	1.04 ± 0.2
$\Lambda_b^0 \to pK(\psi(2)S \to \mu^+\mu^-)$	J/ψ	0	$< 2.79e-07$	< 0.43
$\Lambda_b^0 \to pK(J/\psi \to \mu^+\mu^-)$	J/ψ	1507.5	0.00017	1270 ± 307
$\Lambda_b^0 \to pK\mu^+\mu^-$	J/ψ	90	4.77e-06	0.5 ± 0.11
$B^0 \to K^{*0}(J/\psi \to \mu^+\mu^-)(h \leftrightarrow h)$	J/ψ	1673	5.96e-05	1850 ± 120
$B^0 \to K^{*0}(J/\psi \to \mu^+\mu^-)(h \leftrightarrow \mu)$	J/ψ	0	$< 8.19e-08$	< 2.55
$B^0 \to K^{*0}(\psi(2)S \to \mu^+\mu^-)(h \leftrightarrow h)$	J/ψ	0	$< 2.64e-07$	< 0.51
$B^0 \to K^{*0}(\psi(2)S \to \mu^+\mu^-)(h \leftrightarrow \mu)$	J/ψ	0	$< 2.64e-07$	< 0.51
$B^0 \to (D^- \to K^{*0}\mu^-\bar{\nu}_\mu)\mu^+\nu_\mu$	J/ψ	0	$< 4.47e-08$	< 22.4
$B^+ \to K\pi\pi\mu^+\mu^-$	J/ψ	69.2	4.7e-06	1.26 ± 0.2
$B^+ \to K^+(J/\psi \to \mu^+\mu^-)$	J/ψ	2	$< 4.12e-07$	< 15.3
$B^+ \to K^+\mu^+\mu^-$	J/ψ	0	$< 7.39e-07$	< 0.2

Table C.14: Expected background efficiencies and yields in signal region for 2012 MM R_{K^*}. If nPassed is < 20 a limit at 90% CL is given, else the expected yield is calculated.

MC 2012	q^2 bin	nPassed	efficiency	bkg yield/limit
$B_s^0 \to \phi(J/\psi \to \mu^+\mu^-)$	low	0	$< 1.91e-07$	< 1
$B_s^0 \to \phi(J/\psi \to \mu^+\mu^-)$	central	0	$< 1.91e-07$	< 1
$B_s^0 \to \phi\mu^+\mu^-$	low	224	1.86e-05	1.24 ± 0.23
$B_s^0 \to \phi\mu^+\mu^-$	central	536	4.44e-05	2.95 ± 0.53
$\Lambda_b^0 \to pK(\psi(2)S \to \mu^+\mu^-)$	low	0	$< 2.79e-07$	< 0.43
$\Lambda_b^0 \to pK(\psi(2)S \to \mu^+\mu^-)$	central	0	$< 2.79e-07$	< 0.43
$\Lambda_b^0 \to pK(J/\psi \to \mu^+\mu^-)$	low	0	$< 2.6e-07$	< 1.96
$\Lambda_b^0 \to pK(J/\psi \to \mu^+\mu^-)$	central	0	$< 2.6e-07$	< 1.96
$\Lambda_b^0 \to pK\mu^+\mu^-$	low	86.25	4.57e-06	0.48 ± 0.1
$\Lambda_b^0 \to pK\mu^+\mu^-$	central	326.25	1.73e-05	1.79 ± 0.35
$B^0 \to K^{*0}(J/\psi \to \mu^+\mu^-)(h \leftrightarrow h)$	low	0	$< 8.19e-08$	< 2.55
$B^0 \to K^{*0}(J/\psi \to \mu^+\mu^-)(h \leftrightarrow h)$	central	0	$< 8.19e-08$	< 2.55
$B^0 \to K^{*0}(J/\psi \to \mu^+\mu^-)(h \leftrightarrow \mu)$	low	0	$< 8.19e-08$	< 2.55
$B^0 \to K^{*0}(J/\psi \to \mu^+\mu^-)(h \leftrightarrow \mu)$	central	0	$< 8.19e-08$	< 2.55
$B^0 \to K^{*0}(\psi(2)S \to \mu^+\mu^-)(h \leftrightarrow h)$	low	0	$< 2.64e-07$	< 0.51
$B^0 \to K^{*0}(\psi(2)S \to \mu^+\mu^-)(h \leftrightarrow h)$	central	0	$< 2.64e-07$	< 0.51
$B^0 \to K^{*0}(\psi(2)S \to \mu^+\mu^-)(h \leftrightarrow \mu)$	low	0	$< 2.64e-07$	< 0.51
$B^0 \to K^{*0}(\psi(2)S \to \mu^+\mu^-)(h \leftrightarrow \mu)$	central	0	$< 2.64e-07$	< 0.51
$B^0 \to (D^- \to K^{*0}\mu^-\bar{\nu}_\mu)\mu^+\nu_\mu$	low	0	$< 4.47e-08$	< 22.4
$B^0 \to (D^- \to K^{*0}\mu^-\bar{\nu}_\mu)\mu^+\nu_\mu$	central	0	$< 4.47e-08$	< 22.4
$B^+ \to K\pi\pi\mu^+\mu^-$	low	3.46	$< 4.94e-07$	< 0.13
$B^+ \to K\pi\pi\mu^+\mu^-$	central	8.65	$< 9.35e-07$	< 0.25
$B^+ \to K^+(J/\psi \to \mu^+\mu^-)$	low	0	$< 1.78e-07$	< 6.59
$B^+ \to K^+(J/\psi \to \mu^+\mu^-)$	central	0	$< 1.78e-07$	< 6.59
$B^+ \to K^+\mu^+\mu^-$	low	0	$< 7.39e-07$	< 0.2
$B^+ \to K^+\mu^+\mu^-$	central	0	$< 7.39e-07$	< 0.2

Table C.15: Expected background efficiencies and yields in control region for 2016 MM R_{K^*}. If nPassed is < 20 a limit at 90% CL is given, else the expected yield is calculated.

MC 2016	q^2 bin	nPassed	efficiency	bkg yield/limit
$B_s^0 \to \phi(J/\psi \to \mu^+\mu^-)$	J/ψ	1803	0.000106	584 ± 70
$B_s^0 \to \phi\mu^+\mu^-$	J/ψ	123	7.34e-06	0.52 ± 0.1
$\Lambda_b^0 \to pK(\psi(2)S \to \mu^+\mu^-)$	J/ψ	0	$< 1.89e-07$	< 0.31
$\Lambda_b^0 \to pK(J/\psi \to \mu^+\mu^-)$	J/ψ	388.75	3.17e-05	249 ± 61.4
$\Lambda_b^0 \to pK(J/\psi \to \mu^+\mu^-)$	$\psi(2S)$	0	$< 1.88e-07$	< 1.49
$\Lambda_b^0 \to pK\mu^+\mu^-$	J/ψ	25	5.42e-07	0.061 ± 0.017
$B^0 \to K^{*0}(J/\psi \to \mu^+\mu^-)(h \leftrightarrow h)$	J/ψ	870	2.09e-05	686 ± 47.2
$B^0 \to K^{*0}(J/\psi \to \mu^+\mu^-)(h \leftrightarrow \mu)$	J/ψ	22	5.29e-07	18.1 ± 3.93
$B^0 \to K^{*0}(\psi(2)S \to \mu^+\mu^-)(h \leftrightarrow h)$	J/ψ	0	$< 1.58e-07$	< 0.32
$B^0 \to K^{*0}(\psi(2)S \to \mu^+\mu^-)(h \leftrightarrow \mu)$	J/ψ	0	$< 1.58e-07$	< 0.32
$B^+ \to K\pi\pi\mu^+\mu^-$	J/ψ	57.09	5.1e-06	1.45 ± 0.24
$B^+ \to K^+(J/\psi \to \mu^+\mu^-)$	J/ψ	1	$< 1.34e-07$	< 5.26
$B^+ \to K^+\mu^+\mu^-$	J/ψ	1	$< 6.86e-07$	< 0.2

Table C.16: Expected background efficiencies and yields in signal region for 2016 MM R_{K^*}. If nPassed is < 20 a limit at 90% CL is given, else the expected yield is calculated.

MC 2016	q^2 bin	nPassed	efficiency	bkg yield/limit
$B_s^0 \to \phi(J/\psi \to \mu^+\mu^-)$	low	0	$< 1.36e-07$	< 0.75
$B_s^0 \to \phi(J/\psi \to \mu^+\mu^-)$	central	0	$< 1.36e-07$	< 0.75
$B_s^0 \to \phi\mu^+\mu^-$	low	163	9.73e-06	0.68 ± 0.13
$B_s^0 \to \phi\mu^+\mu^-$	central	393	2.35e-05	1.64 ± 0.3
$\Lambda_b^0 \to pK(\psi(2)S \to \mu^+\mu^-)$	low	0	$< 1.89e-07$	< 0.31
$\Lambda_b^0 \to pK(\psi(2)S \to \mu^+\mu^-)$	central	0	$< 1.89e-07$	< 0.31
$\Lambda_b^0 \to pK(J/\psi \to \mu^+\mu^-)$	low	0	$< 1.88e-07$	< 1.49
$\Lambda_b^0 \to pK(J/\psi \to \mu^+\mu^-)$	central	0	$< 1.88e-07$	< 1.49
$\Lambda_b^0 \to pK\mu^+\mu^-$	low	26.25	5.69e-07	0.064 ± 0.017
$\Lambda_b^0 \to pK\mu^+\mu^-$	central	85	1.84e-06	0.2 ± 0.044
$B^0 \to K^{*0}(J/\psi \to \mu^+\mu^-)(h \leftrightarrow h)$	low	0	$< 5.53e-08$	< 1.81
$B^0 \to K^{*0}(J/\psi \to \mu^+\mu^-)(h \leftrightarrow h)$	central	0	$< 5.53e-08$	< 1.81
$B^0 \to K^{*0}(J/\psi \to \mu^+\mu^-)(h \leftrightarrow \mu)$	low	0	$< 5.53e-08$	< 1.81
$B^0 \to K^{*0}(J/\psi \to \mu^+\mu^-)(h \leftrightarrow \mu)$	central	0	$< 5.53e-08$	< 1.81
$B^0 \to K^{*0}(\psi(2)S \to \mu^+\mu^-)(h \leftrightarrow h)$	low	0	$< 1.58e-07$	< 0.32
$B^0 \to K^{*0}(\psi(2)S \to \mu^+\mu^-)(h \leftrightarrow h)$	central	0	$< 1.58e-07$	< 0.32
$B^0 \to K^{*0}(\psi(2)S \to \mu^+\mu^-)(h \leftrightarrow \mu)$	low	0	$< 1.58e-07$	< 0.32
$B^0 \to K^{*0}(\psi(2)S \to \mu^+\mu^-)(h \leftrightarrow \mu)$	central	0	$< 1.58e-07$	< 0.32
$B^+ \to K\pi\pi\mu^+\mu^-$	low	6.92	$< 1.04e-06$	< 0.3
$B^+ \to K\pi\pi\mu^+\mu^-$	central	5.19	$< 8.47e-07$	< 0.24
$B^+ \to K^+(J/\psi \to \mu^+\mu^-)$	low	0	$< 7.96e-08$	< 3.11
$B^+ \to K^+(J/\psi \to \mu^+\mu^-)$	central	1	$< 1.34e-07$	< 5.26
$B^+ \to K^+\mu^+\mu^-$	low	0	$< 4.06e-07$	< 0.12
$B^+ \to K^+\mu^+\mu^-$	central	0	$< 4.06e-07$	< 0.12

Appendix D

MVA

D.1 Signal/Background comparisons

D.1.1 MVA_{Comb} Signal/Background Comparison

D.1.1.1 B^0 $\mu\mu$ MVA_{Comb}

Run 2p1

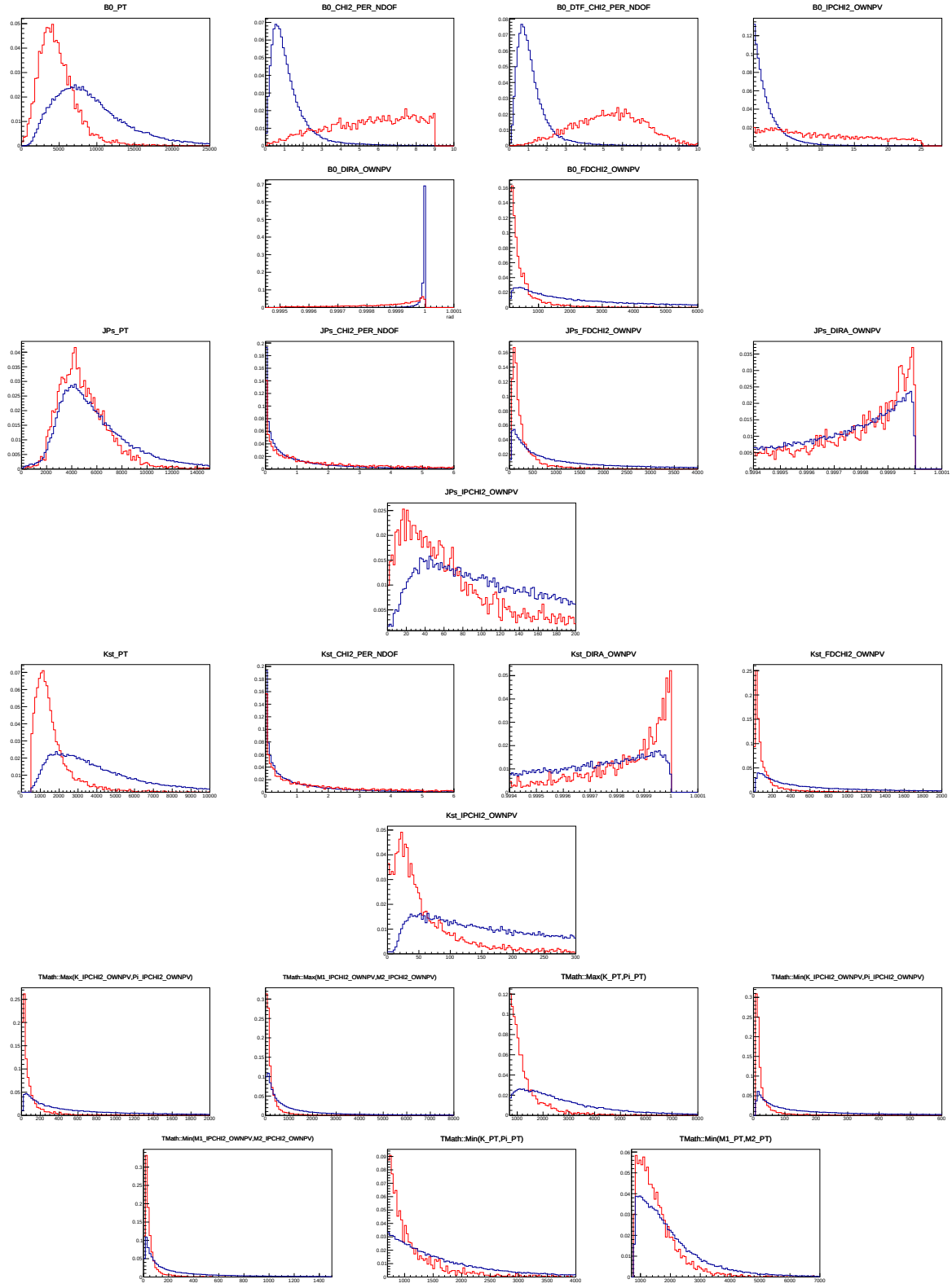


Figure 1: Comparison of variables used in Signal and Background proxies for B^0 $\mu\mu$ MVA_{COMB} trainings. Black respresents truth matched fully selected $B^0 \to K^{*0}\mu^+\mu^-$ MC, red respresents fully selected collision data from the upper sideband, corresponding to $> 5400\,\text{MeV}/c^2$.

D.1.1.2 B^+ $\mu\mu$ MVA$_{\text{Comb}}$

Run 2p1

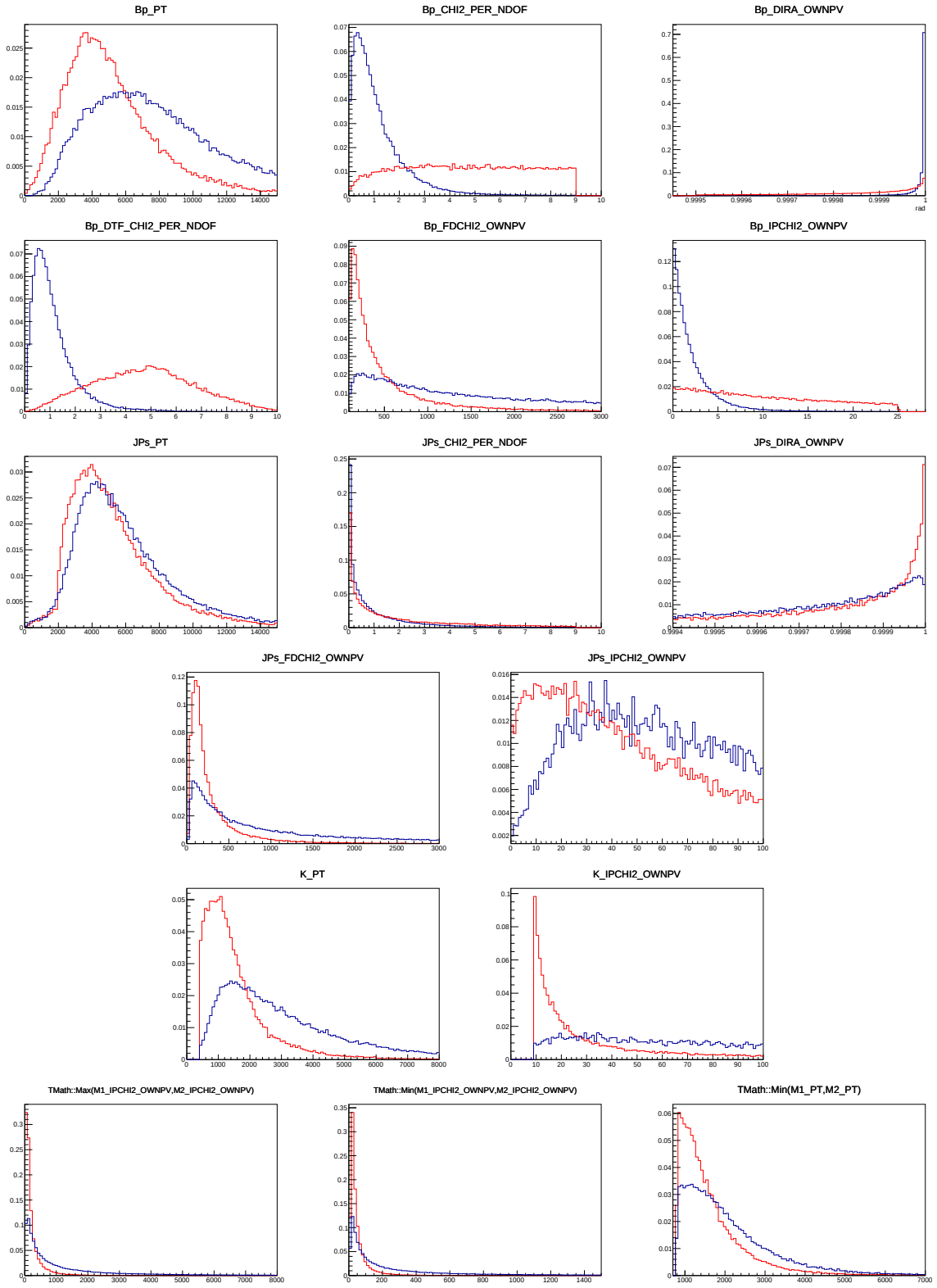


Figure 2: Comparison of variables used in Signal and Background proxies for B^+ $\mu\mu$ MVA$_{\text{COMB}}$ trainings. Black respresents truth matched fully selected $B^+ \to K^+\mu^+\mu^-$ MC, red respresents fully selected collision data from the upper sideband, corresponding to $> 5400\,\text{MeV}/c^2$.

D.1.1.3 B^0 ee MVA$_{\text{Comb}}$

Run 2p1

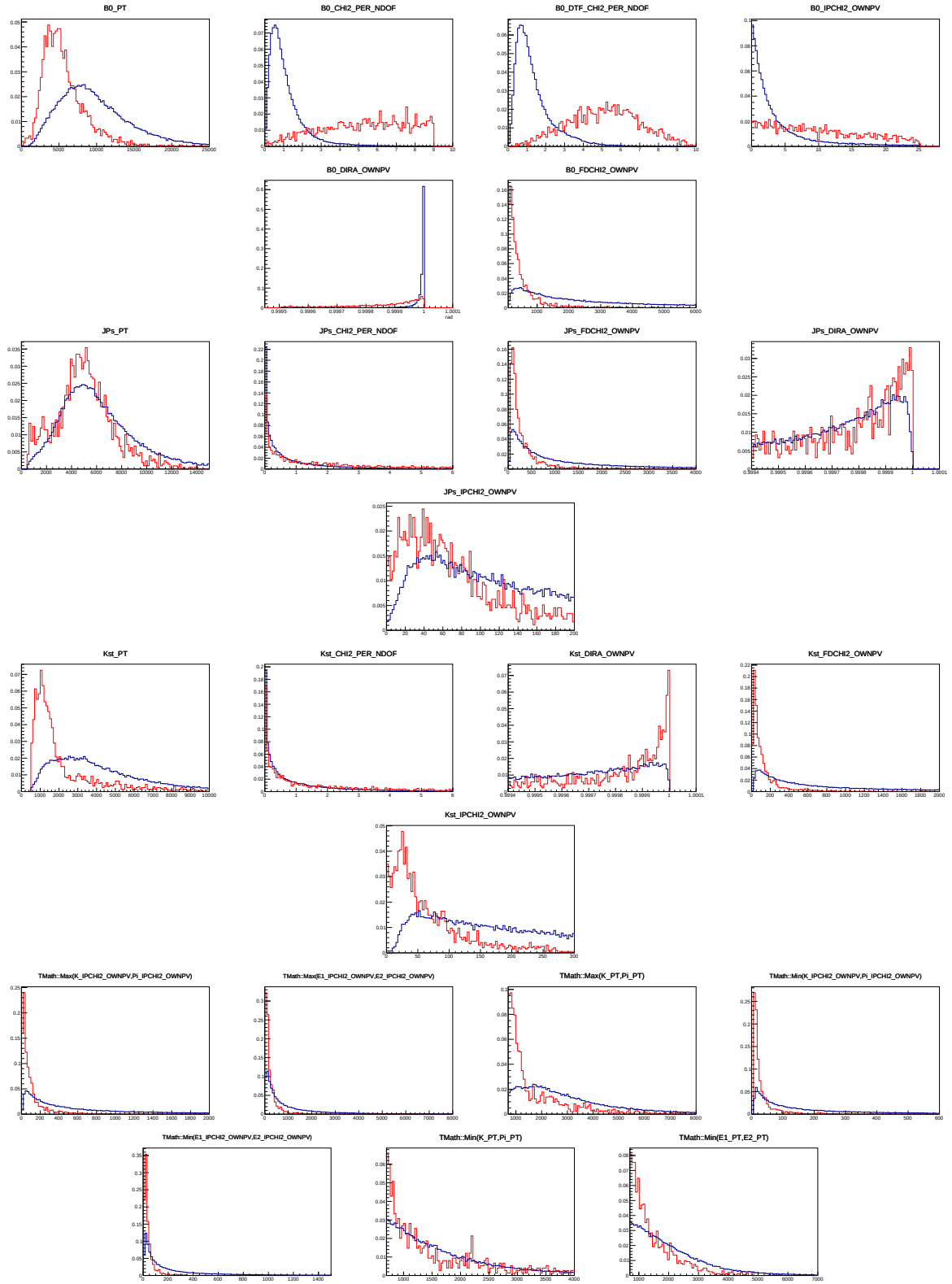


Figure 3: Comparison of variables used in Signal and Background proxies for B^0 ee MVA$_{\text{COMB}}$ trainings. Black respresents truth matched fully selected $B^0 \to K^{*0}e^+e^-$ MC, red respresents fully selected collision data from the upper sideband, corresponding to $> 5600\,\text{MeV}/c^2$.

D.1.1.4 B^+ ee MVA$_{\text{Comb}}$

Run 2p1

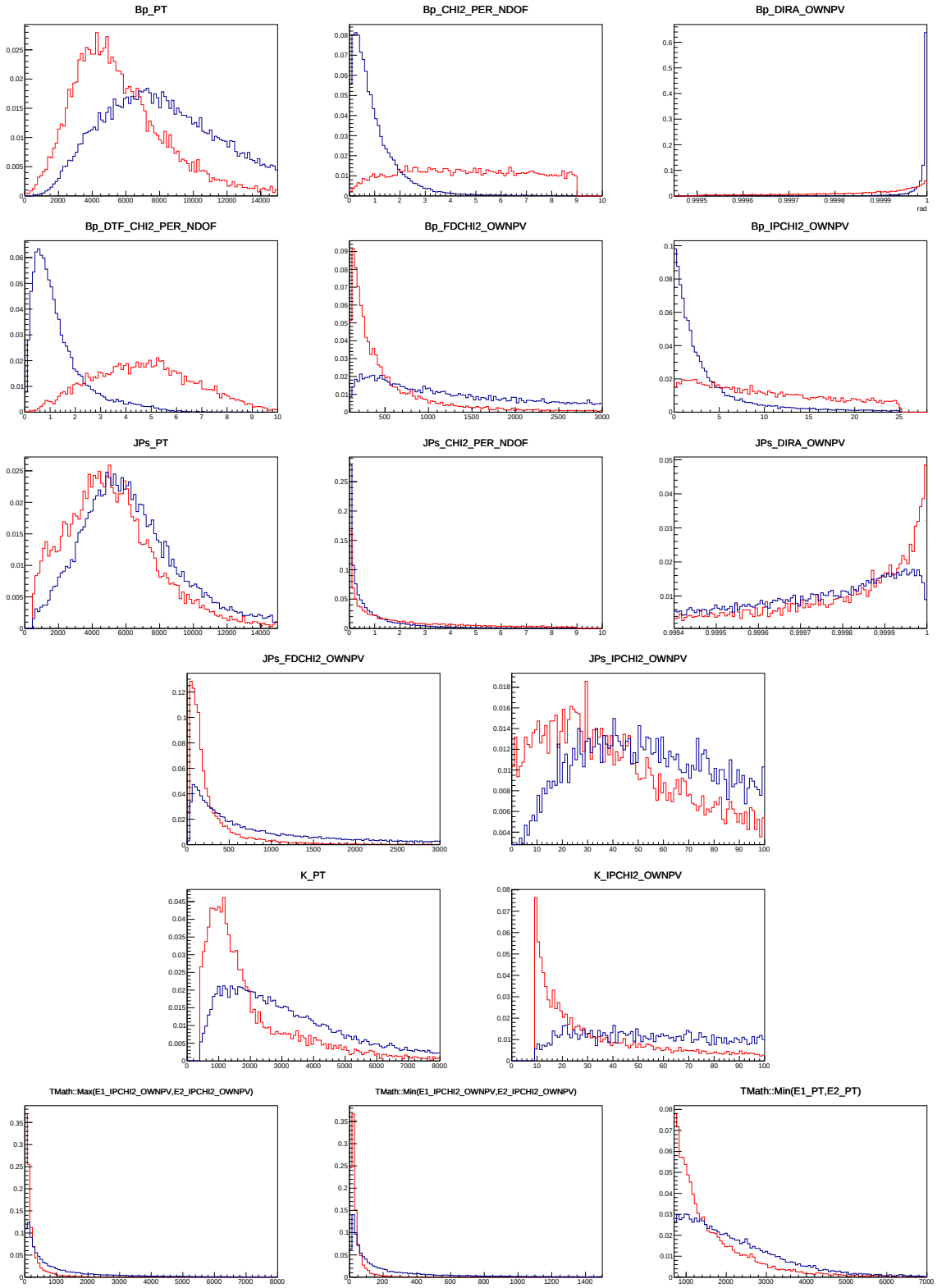


Figure 4: Comparison of variables used in Signal and Background proxies for B^+ ee MVA$_{\text{COMB}}$ trainings. Black respresents truth matched fully selected $B^+ \to K^+e^+e^-$ MC, red respresents fully selected collision data from the upper sideband, corresponding to $> 5600\,\text{MeV}/c^2$.

D.1.2 $\mathrm{MVA_{PReco}}$ Signal/Background Comparison

D.1.2.1 B^0 ee $\mathrm{MVA_{PReco}}$

Run 2p1

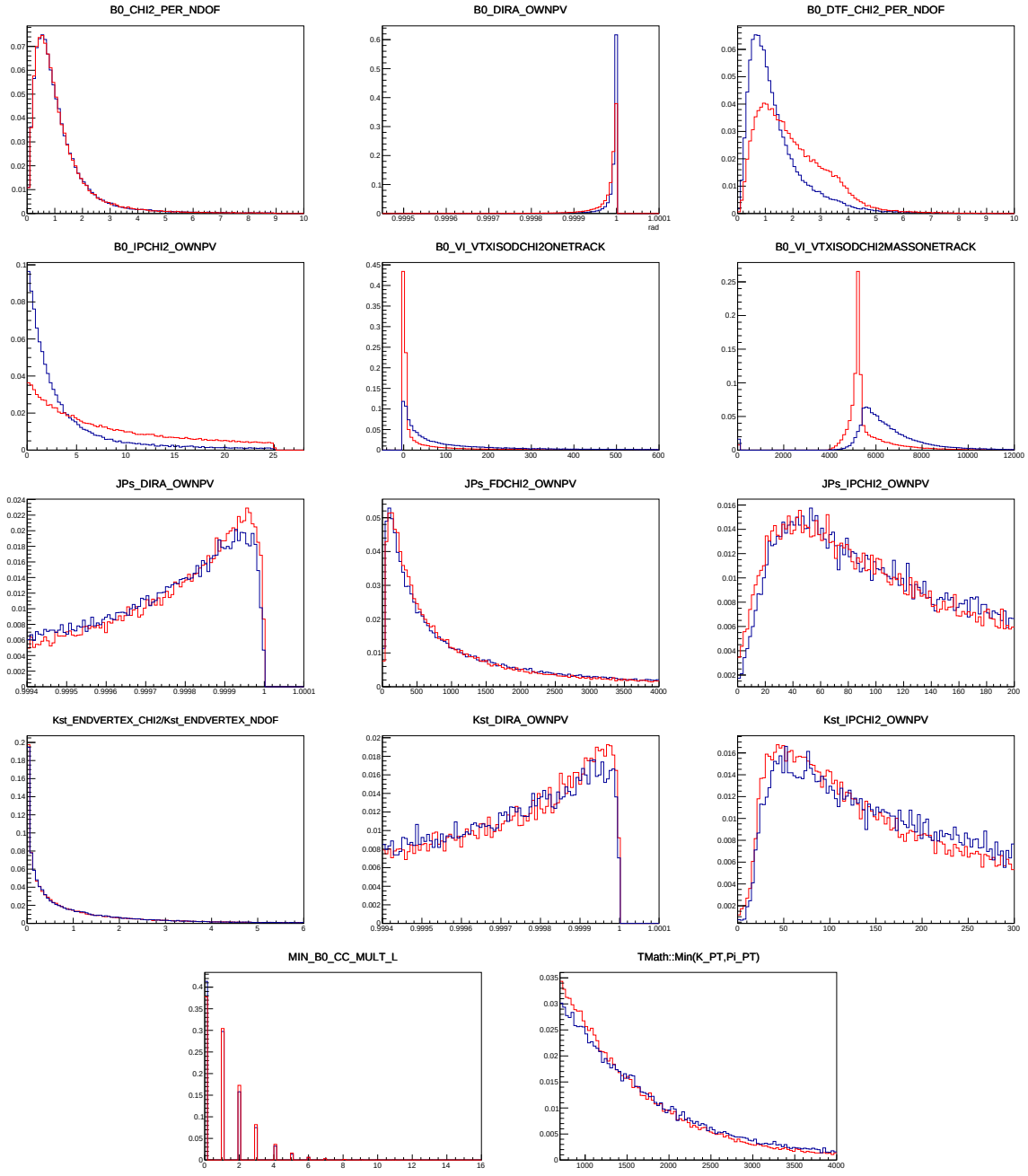


Figure 5: Comparison of variables used in Signal and Background proxies for B^0 ee $\mathrm{MVA_{PRECO}}$ trainings. Black respresents truth matched fully selected $B^0 \to K^{*0}e^+e^-$ MC, red respresents fully selected partially reconstructed $B^+ \to K\pi\pi e^+e^-$ MC.

D.1.2.2 B^+ ee $\mathbf{MVA_{PReco}}$

Run 2p1

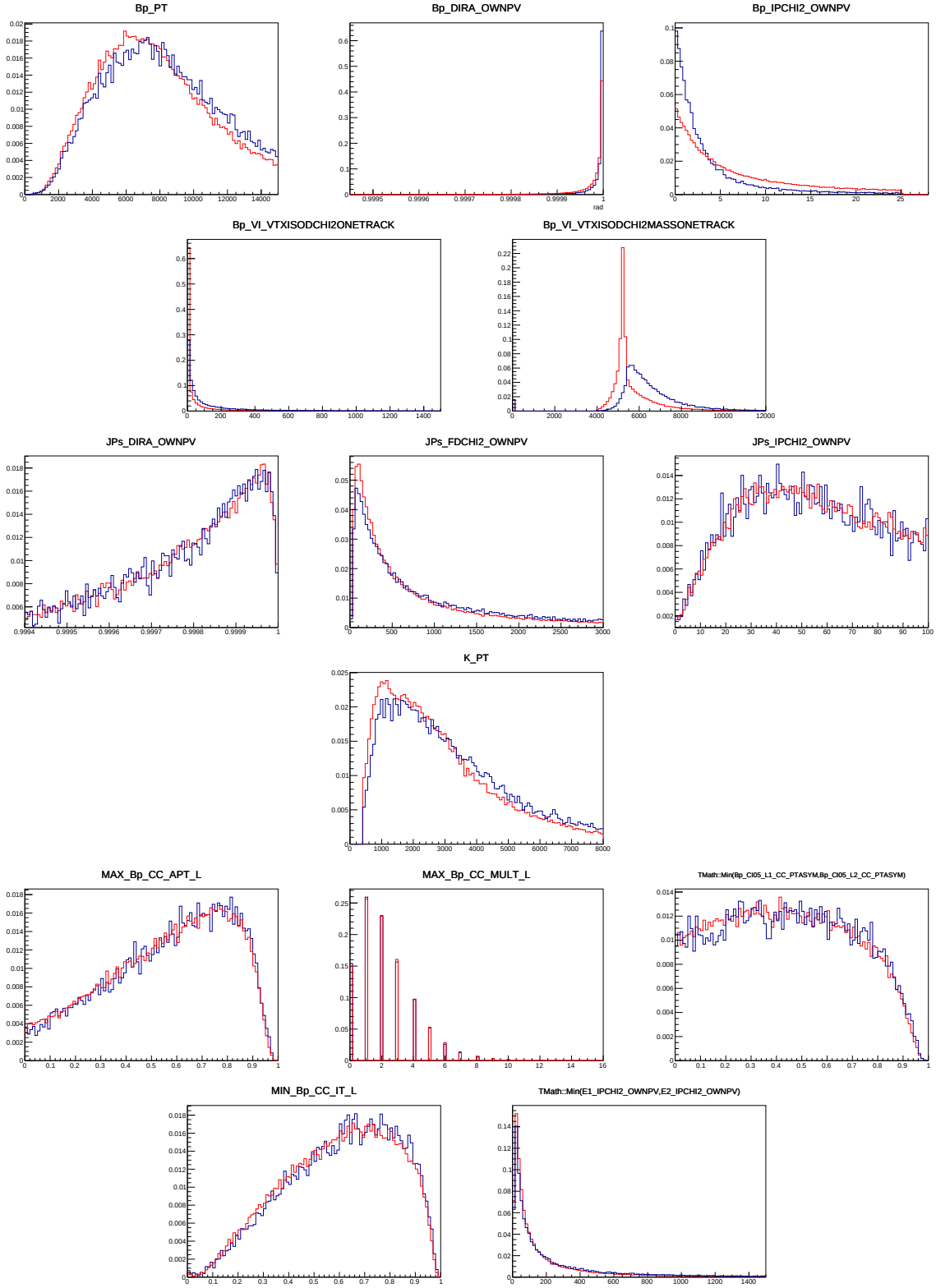


Figure 6: Comparison of variables used in Signal and Background proxies for B^+ ee $\mathrm{MVA_{PRECO}}$ trainings. Black respresents truth matched fully selected $B^+ \to K^+e^+e^-$ MC, red represents fully selected partially reconstructed $B^0 \to K^{*0}e^+e^-$ MC.

D.2 Training

D.2.1 $B^0 \to K^{*0}\mu^+\mu^-$ $\mathbf{MVA_{Comb}}$ Training

D.2.1.1 Feature Importances Run 2p1

Table 1: Feature importances per fold, with fold number increasing from left to right

Variable	Importance
B0 DTF CHI2 PER NDOF	38.273255
Kst FDCHI2 OWNPV	7.588810
B0 PT	6.975776
MAX PT Kst DAUGHTERS	6.785875
MIN IPCHI2 LEPTONS	5.422269
B0 CHI2 PER NDOF	4.331831
B0 DIRA OWNPV	4.081118
MAX IPCHI2 Kst DAUGHTERS	3.657733
Kst PT	3.050851
B0 FDCHI2 OWNPV	3.030965
MIN IPCHI2 Kst DAUGHTERS	2.500212
Kst IPCHI2 OWNPV	2.295404
B0 IPCHI2 OWNPV	2.174148
Kst CHI2 PER NDOF	2.165765
JPs IPCHI2 OWNPV	1.515274
Kst DIRA OWNPV	1.176579
MAX IPCHI2 LEPTONS	0.959059
JPs FDCHI2 OWNPV	0.925766
JPs DIRA OWNPV	0.799611
MIN PT LEPTONS	0.791292
JPs PT	0.703030
MIN PT Kst DAUGHTERS	0.653290
JPs CHI2 PER NDOF	0.142087

Variable	Importance
B0 DTF CHI2 PER NDOF	40.140669
MAX PT Kst DAUGHTERS	9.930499
Kst IPCHI2 OWNPV	7.202289
B0 PT	6.441251
B0 FDCHI2 OWNPV	4.951493
MIN IPCHI2 LEPTONS	4.758021
MIN IPCHI2 Kst DAUGHTERS	4.442199
B0 DIRA OWNPV	3.691737
B0 CHI2 PER NDOF	2.467326
Kst CHI2 PER NDOF	2.444137
Kst PT	2.137006
Kst DIRA OWNPV	1.840635
B0 IPCHI2 OWNPV	1.614482
MAX IPCHI2 LEPTONS	1.505081
MIN PT LEPTONS	1.343517
JPs PT	1.266786
MIN PT Kst DAUGHTERS	1.008668
JPs DIRA OWNPV	0.627231
MAX IPCHI2 Kst DAUGHTERS	0.604255
JPs IPCHI2 OWNPV	0.509739
JPs FDCHI2 OWNPV	0.436663
Kst FDCHI2 OWNPV	0.419353
JPs CHI2 PER NDOF	0.216964

Variable	Importance
B0 DTF CHI2 PER NDOF	40.446992
MAX PT Kst DAUGHTERS	7.277551
Kst IPCHI2 OWNPV	6.118775
B0 FDCHI2 OWNPV	6.111644
B0 PT	5.070128
Kst FDCHI2 OWNPV	4.831109
B0 DIRA OWNPV	3.853090
Kst PT	3.701098
B0 CHI2 PER NDOF	3.634573
MIN IPCHI2 LEPTONS	3.253854
Kst CHI2 PER NDOF	2.294088
JPs IPCHI2 OWNPV	2.164736
MIN IPCHI2 Kst DAUGHTERS	1.988287
B0 IPCHI2 OWNPV	1.895060
Kst DIRA OWNPV	1.881124
MIN PT Kst DAUGHTERS	1.093615
JPs PT	1.077594
MIN PT LEPTONS	0.815667
JPs DIRA OWNPV	0.755870
JPs FDCHI2 OWNPV	0.572479
MAX IPCHI2 LEPTONS	0.489728
MAX IPCHI2 Kst DAUGHTERS	0.415411
JPs CHI2 PER NDOF	0.257529

Variable	Importance
B0 DTF CHI2 PER NDOF	40.468510
B0 PT	8.004842
Kst FDCHI2 OWNPV	7.477213
MAX PT Kst DAUGHTERS	7.384333
JPs IPCHI2 OWNPV	5.041691
Kst PT	4.415360
B0 DIRA OWNPV	4.116160
MIN IPCHI2 LEPTONS	3.901772
B0 FDCHI2 OWNPV	3.639454
MIN IPCHI2 Kst DAUGHTERS	2.675034
Kst CHI2 PER NDOF	2.097346
Kst DIRA OWNPV	1.618026
JPs FDCHI2 OWNPV	1.390058
MIN PT LEPTONS	1.219229
Kst IPCHI2 OWNPV	1.110656
MAX IPCHI2 Kst DAUGHTERS	1.089901
B0 CHI2 PER NDOF	0.945660
MIN PT Kst DAUGHTERS	0.661914
MAX IPCHI2 LEPTONS	0.612040
B0 IPCHI2 OWNPV	0.563784
JPs PT	0.554838
JPs DIRA OWNPV	0.527783
JPs CHI2 PER NDOF	0.484397

Variable	Importance
B0 DTF CHI2 PER NDOF	38.730555
B0 PT	8.321023
Kst IPCHI2 OWNPV	7.054464
Kst FDCHI2 OWNPV	6.693727
Kst PT	5.752437
MAX PT Kst DAUGHTERS	5.251472
JPs IPCHI2 OWNPV	4.240484
MIN IPCHI2 LEPTONS	4.076723
B0 FDCHI2 OWNPV	3.812212
B0 DIRA OWNPV	3.290430
MIN IPCHI2 Kst DAUGHTERS	2.240296
Kst CHI2 PER NDOF	2.020812
B0 CHI2 PER NDOF	1.649279
Kst DIRA OWNPV	1.492367
B0 IPCHI2 OWNPV	1.324146
MIN PT LEPTONS	1.174987
MAX IPCHI2 LEPTONS	0.570516
MIN PT Kst DAUGHTERS	0.554827
JPs PT	0.491283
JPs DIRA OWNPV	0.457686
JPs CHI2 PER NDOF	0.299388
JPs FDCHI2 OWNPV	0.267988
MAX IPCHI2 Kst DAUGHTERS	0.232899

Variable	Importance
B0 DTF CHI2 PER NDOF	38.557295
B0 PT	7.635780
MAX PT Kst DAUGHTERS	6.151846
B0 FDCHI2 OWNPV	5.448605
MIN IPCHI2 LEPTONS	5.089863
B0 DIRA OWNPV	4.897452
B0 CHI2 PER NDOF	4.822588
Kst PT	4.015736
Kst FDCHI2 OWNPV	3.565315
Kst IPCHI2 OWNPV	3.063702
MIN IPCHI2 Kst DAUGHTERS	2.936397
Kst CHI2 PER NDOF	2.096582
MAX IPCHI2 Kst DAUGHTERS	1.609264
B0 IPCHI2 OWNPV	1.552634
MIN PT LEPTONS	1.528947
Kst DIRA OWNPV	1.451938
MIN PT Kst DAUGHTERS	1.124014
JPs DIRA OWNPV	1.053202
JPs IPCHI2 OWNPV	0.985400
MAX IPCHI2 LEPTONS	0.719292
JPs CHI2 PER NDOF	0.651156
JPs FDCHI2 OWNPV	0.528495
JPs PT	0.514498

Variable	Importance
B0 DTF CHI2 PER NDOF	42.261867
B0 PT	9.031102
Kst IPCHI2 OWNPV	7.321748
MAX PT Kst DAUGHTERS	6.192556
Kst FDCHI2 OWNPV	4.020448
MIN IPCHI2 LEPTONS	3.663914
JPs IPCHI2 OWNPV	3.026351
B0 CHI2 PER NDOF	2.926873
B0 DIRA OWNPV	2.890010
MAX IPCHI2 LEPTONS	2.859623
Kst PT	2.374687
MIN IPCHI2 Kst DAUGHTERS	2.258552
Kst CHI2 PER NDOF	2.017955
MIN PT LEPTONS	1.678675
JPs FDCHI2 OWNPV	1.466931
B0 IPCHI2 OWNPV	1.216518
B0 FDCHI2 OWNPV	1.128900
Kst DIRA OWNPV	0.943989
JPs DIRA OWNPV	0.775250
JPs PT	0.648729
JPs CHI2 PER NDOF	0.512414
MAX IPCHI2 Kst DAUGHTERS	0.408183
MIN PT Kst DAUGHTERS	0.374727

Variable	Importance
B0 DTF CHI2 PER NDOF	35.727376
Kst PT	9.299879
MIN IPCHI2 LEPTONS	9.082750
B0 PT	7.047893
B0 DIRA OWNPV	5.358699
MAX IPCHI2 Kst DAUGHTERS	5.264401
B0 CHI2 PER NDOF	4.990019
Kst FDCHI2 OWNPV	2.646999
MIN IPCHI2 Kst DAUGHTERS	2.534970
Kst IPCHI2 OWNPV	2.220591
Kst DIRA OWNPV	2.129826
MAX PT Kst DAUGHTERS	1.963209
B0 IPCHI2 OWNPV	1.642731
JPs DIRA OWNPV	1.583866
Kst CHI2 PER NDOF	1.525073
JPs FDCHI2 OWNPV	1.246329
MIN PT LEPTONS	1.244619
JPs IPCHI2 OWNPV	1.170824
B0 FDCHI2 OWNPV	1.041769
MIN PT Kst DAUGHTERS	0.619476
JPs CHI2 PER NDOF	0.594678
MAX IPCHI2 LEPTONS	0.548305
JPs PT	0.515716

Variable	Importance
B0 DTF CHI2 PER NDOF	42.556933
MAX PT Kst DAUGHTERS	9.791036
B0 PT	7.550382
B0 FDCHI2 OWNPV	7.279631
Kst FDCHI2 OWNPV	4.676870
JPs IPCHI2 OWNPV	3.919193
MIN IPCHI2 LEPTONS	3.836687
B0 DIRA OWNPV	3.096786
Kst IPCHI2 OWNPV	2.776615
MIN IPCHI2 Kst DAUGHTERS	2.361456
Kst PT	1.933522
B0 CHI2 PER NDOF	1.512355
Kst CHI2 PER NDOF	1.503744
Kst DIRA OWNPV	1.439717
B0 IPCHI2 OWNPV	1.067233
MIN PT LEPTONS	0.980242
JPs DIRA OWNPV	0.747372
MAX IPCHI2 Kst DAUGHTERS	0.725318
JPs PT	0.637890
JPs FDCHI2 OWNPV	0.549112
JPs CHI2 PER NDOF	0.496086
MIN PT Kst DAUGHTERS	0.331042
MAX IPCHI2 LEPTONS	0.230777

Variable	Importance
B0 DTF CHI2 PER NDOF	45.498258
B0 FDCHI2 OWNPV	7.692595
B0 PT	7.498333
MAX PT Kst DAUGHTERS	6.516227
Kst IPCHI2 OWNPV	6.172885
MIN IPCHI2 LEPTONS	5.178382
Kst PT	3.327249
B0 DIRA OWNPV	2.456319
B0 CHI2 PER NDOF	2.291291
Kst FDCHI2 OWNPV	1.904714
MIN IPCHI2 Kst DAUGHTERS	1.824370
Kst CHI2 PER NDOF	1.504583
Kst DIRA OWNPV	1.220349
JPs DIRA OWNPV	0.988608
MIN PT LEPTONS	0.985922
MAX IPCHI2 LEPTONS	0.950380
MIN PT Kst DAUGHTERS	0.932138
B0 IPCHI2 OWNPV	0.600180
JPs FDCHI2 OWNPV	0.578238
JPs IPCHI2 OWNPV	0.527626
JPs PT	0.495318
MAX IPCHI2 Kst DAUGHTERS	0.473673
JPs CHI2 PER NDOF	0.382360

D.2.1.2 MVA Correlations with B^0 mass and q^2 Run 2p1

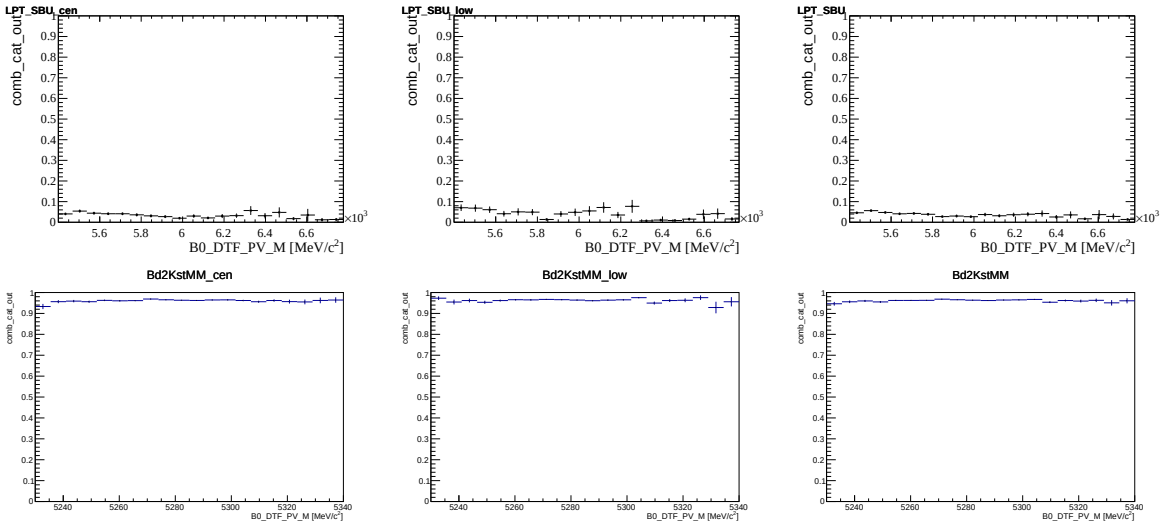


Figure 1: (Top) From left to right, average value of the MVA classifier in collision data sideband as a function of the `DecayTreeFitter` B^0 mass in the central, low and combined low and central regions respectively in Run 2p1. (Bottom) From left to right, average value of the MVA classifier in simulated Bd2KstMM MC as a function of the `DecayTreeFitter` B^0 mass in the central, low and combined low and central regions respectively in Run 2p1.

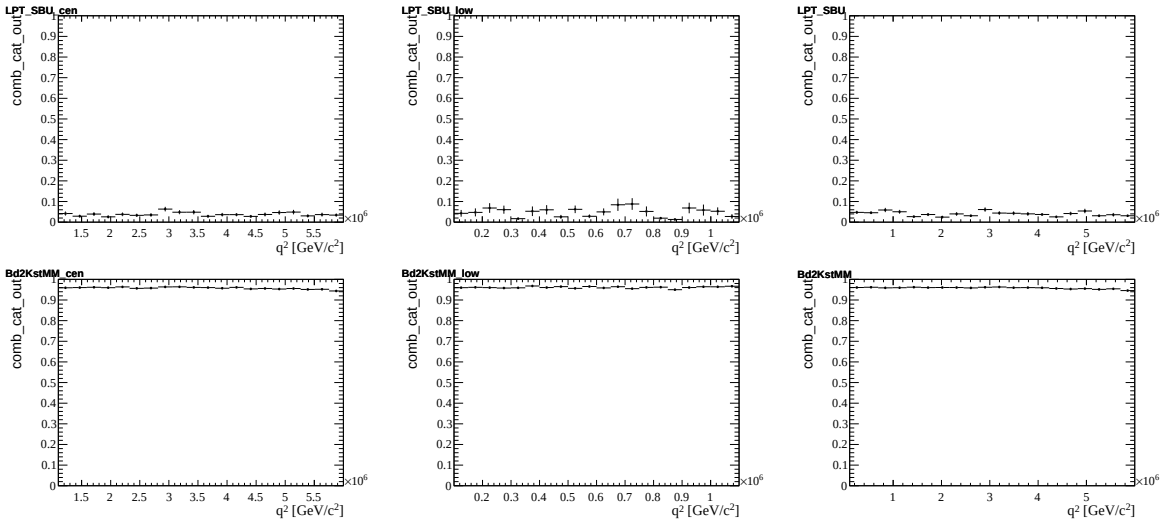


Figure 2: (Top) From left to right, average value of the MVA classifier in collision data sideband as a function of q^2 in the central, low and combined low and central regions respectively in Run 2p1. (Bottom) From left to right, average value of the MVA classifier in simulated Bd2KstMM MC as a function of q^2 in the central, low and combined low and central regions respectively in Run 2p1.

D.2.2 $B^+ \to K^+\mu^+\mu^-$ $\mathbf{MVA_{Comb}}$ Training

D.2.2.1 Feature Importances Run 2p1

Table 2: Feature importances per fold, with fold number increasing from left to right

Variable	Importance
Bp DTF CHI2 PER NDOF	38.847878
K IPCHI2 OWNPV	13.159679
K PT	9.507049
Bp PT	8.687189
Bp FDCHI2 OWNPV	4.741912
MIN IPCHI2 LEPTONS	4.676309
JPs IPCHI2 OWNPV	4.420042
Bp DIRA OWNPV	3.688296
JPs FDCHI2 OWNPV	3.470749
MIN PT LEPTONS	2.577063
JPs PT	2.502602
JPs DIRA OWNPV	1.798383
Bp CHI2 PER NDOF	0.937972
MAX IPCHI2 LEPTONS	0.510132
JPs CHI2 PER NDOF	0.277576
Bp IPCHI2 OWNPV	0.197169

Variable	Importance
Bp DTF CHI2 PER NDOF	33.635930
K IPCHI2 OWNPV	15.641991
Bp PT	9.467542
K PT	9.359377
JPs IPCHI2 OWNPV	6.892695
Bp FDCHI2 OWNPV	4.880137
JPs PT	4.063640
MIN IPCHI2 LEPTONS	3.610005
Bp DIRA OWNPV	3.302705
MIN PT LEPTONS	3.068819
JPs FDCHI2 OWNPV	2.126774
Bp CHI2 PER NDOF	1.591255
JPs DIRA OWNPV	1.250856
MAX IPCHI2 LEPTONS	0.497113
JPs CHI2 PER NDOF	0.344300
Bp IPCHI2 OWNPV	0.266861

Variable	Importance
Bp DTF CHI2 PER NDOF	35.148357
K IPCHI2 OWNPV	16.114270
K PT	9.470941
Bp PT	8.363569
JPs FDCHI2 OWNPV	5.217010
Bp DIRA OWNPV	4.812858
MIN IPCHI2 LEPTONS	4.094601
JPs IPCHI2 OWNPV	3.856057
JPs PT	3.551559
Bp FDCHI2 OWNPV	2.696210
MIN PT LEPTONS	2.601026
JPs DIRA OWNPV	2.309576
Bp CHI2 PER NDOF	1.304795
JPs CHI2 PER NDOF	0.254788
MAX IPCHI2 LEPTONS	0.118179
Bp IPCHI2 OWNPV	0.086205

Variable	Importance
Bp DTF CHI2 PER NDOF	34.601839
K IPCHI2 OWNPV	14.903376
K PT	10.134964
Bp PT	9.080601
JPs IPCHI2 OWNPV	6.226178
Bp FDCHI2 OWNPV	4.180600
MIN IPCHI2 LEPTONS	3.844460
Bp DIRA OWNPV	3.600136
JPs PT	3.548378
JPs FDCHI2 OWNPV	3.519098
MIN PT LEPTONS	2.881949
JPs DIRA OWNPV	1.816912
Bp CHI2 PER NDOF	0.890085
JPs CHI2 PER NDOF	0.327830
Bp IPCHI2 OWNPV	0.230858
MAX IPCHI2 LEPTONS	0.212738

Variable	Importance
Bp DTF CHI2 PER NDOF	34.268958
K IPCHI2 OWNPV	15.118865
Bp PT	9.591098
K PT	9.003764
JPs IPCHI2 OWNPV	5.369853
MIN IPCHI2 LEPTONS	4.671568
Bp FDCHI2 OWNPV	4.326883
Bp DIRA OWNPV	4.175035
JPs FDCHI2 OWNPV	3.521780
JPs PT	3.332418
MIN PT LEPTONS	2.745535
JPs DIRA OWNPV	2.089359
Bp CHI2 PER NDOF	1.202710
JPs CHI2 PER NDOF	0.334459
Bp IPCHI2 OWNPV	0.213832
MAX IPCHI2 LEPTONS	0.033884

Variable	Importance
Bp DTF CHI2 PER NDOF	35.522159
K IPCHI2 OWNPV	13.239337
K PT	9.697237
Bp PT	9.441826
MIN IPCHI2 LEPTONS	5.573474
JPs IPCHI2 OWNPV	4.981142
Bp DIRA OWNPV	4.389426
Bp FDCHI2 OWNPV	4.071497
JPs PT	3.907441
MIN PT LEPTONS	2.698830
MAX IPCHI2 LEPTONS	1.673107
Bp CHI2 PER NDOF	1.638857
JPs DIRA OWNPV	1.226074
JPs FDCHI2 OWNPV	0.951978
Bp IPCHI2 OWNPV	0.680105
JPs CHI2 PER NDOF	0.307509

Variable	Importance
Bp DTF CHI2 PER NDOF	35.787185
K IPCHI2 OWNPV	14.172239
K PT	9.925893
Bp PT	8.394617
Bp FDCHI2 OWNPV	5.208855
MIN IPCHI2 LEPTONS	4.249406
JPs IPCHI2 OWNPV	3.953343
JPs PT	3.945128
Bp DIRA OWNPV	3.661985
JPs FDCHI2 OWNPV	3.165520
MIN PT LEPTONS	2.868277
JPs DIRA OWNPV	1.539833
MAX IPCHI2 LEPTONS	1.537718
Bp CHI2 PER NDOF	0.919420
Bp IPCHI2 OWNPV	0.340747
JPs CHI2 PER NDOF	0.329835

Variable	Importance
Bp DTF CHI2 PER NDOF	37.565053
K IPCHI2 OWNPV	13.307472
K PT	9.358270
Bp PT	8.529899
MIN IPCHI2 LEPTONS	4.462646
Bp FDCHI2 OWNPV	4.255850
JPs IPCHI2 OWNPV	3.870031
JPs PT	3.844494
Bp DIRA OWNPV	3.778157
MIN PT LEPTONS	3.208771
JPs FDCHI2 OWNPV	3.175774
Bp CHI2 PER NDOF	1.570710
JPs DIRA OWNPV	1.337406
MAX IPCHI2 LEPTONS	1.213942
JPs CHI2 PER NDOF	0.361259
Bp IPCHI2 OWNPV	0.160267

Variable	Importance
Bp DTF CHI2 PER NDOF	36.129930
K IPCHI2 OWNPV	14.231876
K PT	10.222276
Bp PT	9.023252
JPs FDCHI2 OWNPV	4.745964
MIN IPCHI2 LEPTONS	4.541480
Bp DIRA OWNPV	4.296162
JPs IPCHI2 OWNPV	3.939593
JPs PT	3.566289
Bp FDCHI2 OWNPV	3.007303
MIN PT LEPTONS	2.640868
JPs DIRA OWNPV	1.444425
Bp CHI2 PER NDOF	1.439022
Bp IPCHI2 OWNPV	0.462215
JPs CHI2 PER NDOF	0.234974
MAX IPCHI2 LEPTONS	0.074372

Variable	Importance
Bp DTF CHI2 PER NDOF	29.517391
K IPCHI2 OWNPV	15.997908
K PT	11.554808
Bp PT	10.212065
MIN IPCHI2 LEPTONS	4.501117
JPs FDCHI2 OWNPV	4.395993
JPs PT	4.360025
JPs IPCHI2 OWNPV	4.026683
Bp DIRA OWNPV	3.983492
Bp FDCHI2 OWNPV	3.246759
MIN PT LEPTONS	2.797788
JPs DIRA OWNPV	1.817667
MAX IPCHI2 LEPTONS	1.594951
Bp CHI2 PER NDOF	1.471509
JPs CHI2 PER NDOF	0.261906
Bp IPCHI2 OWNPV	0.259939

D.2.2.2 MVA Correlations with B^+ mass and q^2 Run 2p1

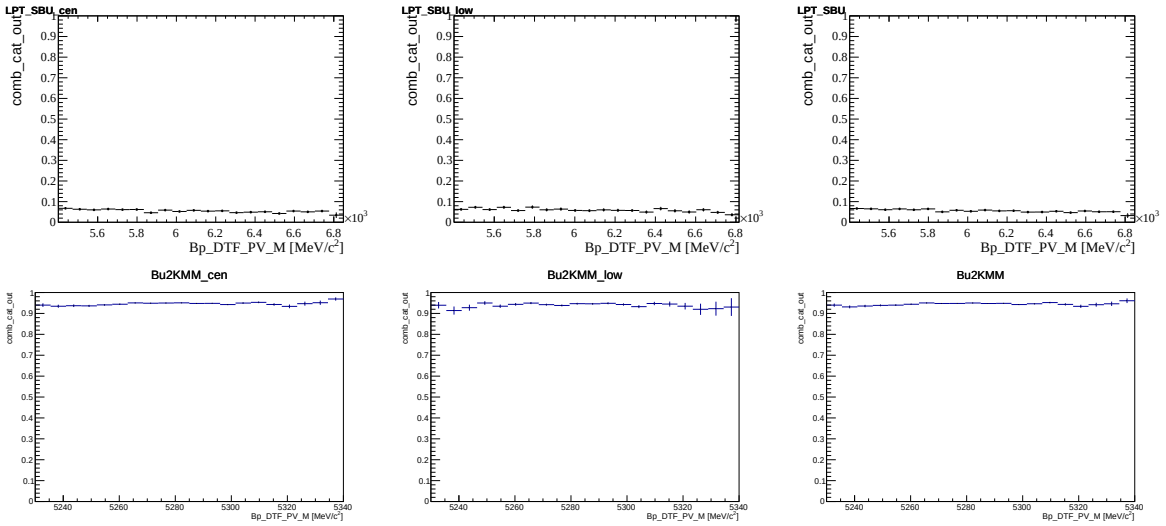


Figure 3: (Top) From left to right, average value of the MVA classifier in collision data sideband as a function of the `DecayTreeFitter` B^+ mass in the central, low and combined low and central regions respectively in Run 2p1. (Bottom) From left to right, average value of the MVA classifier in simulated Bu2KMM MC as a function of the `DecayTreeFitter` B^+ mass in the central, low and combined low and central regions respectively in Run 2p1.

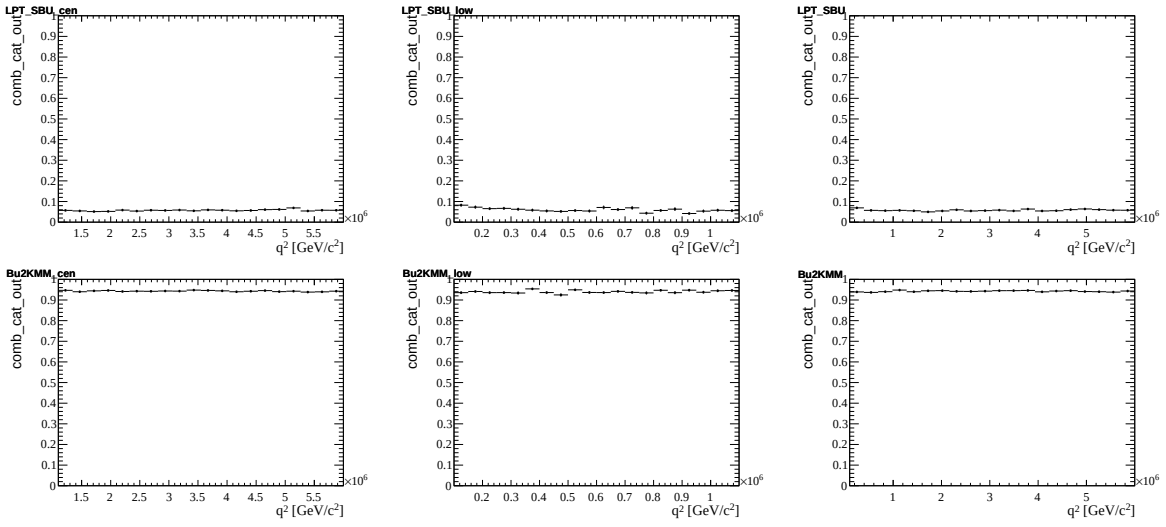


Figure 4: (Top) From left to right, average value of the MVA classifier in collision data sideband as a function of q^2 in the central, low and combined low and central regions respectively in Run 2p1. (Bottom) From left to right, average value of the MVA classifier in simulated Bu2KMM MC as a function of q^2 in the central, low and combined low and central regions respectively in Run 2p1.

D.2.3 $B^0 \to K^{*0} e^+ e^-$ $\mathbf{MVA_{Comb}}$ Training

D.2.3.1 Feature Importances Run 2p1

Table 3: Feature importances per fold, with fold number increasing from left to right

Variable	Importance
B0 DTF CHI2 PER NDOF	23.383922
B0 FDCHI2 OWNPV	11.657189
B0 CHI2 PER NDOF	10.573400
MIN IPCHI2 LEPTONS	8.400452
B0 DIRA OWNPV	7.801556
MAX PT Kst DAUGHTERS	6.842947
B0 PT	4.562260
Kst IPCHI2 OWNPV	3.964662
MIN PT LEPTONS	2.485022
MIN IPCHI2 Kst DAUGHTERS	2.265901
Kst PT	2.211096
JPs DIRA OWNPV	2.174570
B0 IPCHI2 OWNPV	1.880640
MAX IPCHI2 Kst DAUGHTERS	1.535596
JPs PT	1.533356
Kst FDCHI2 OWNPV	1.517955
MIN PT Kst DAUGHTERS	1.421888
Kst DIRA OWNPV	1.368886
Kst CHI2 PER NDOF	1.321724
JPs FDCHI2 OWNPV	1.073796
MAX IPCHI2 LEPTONS	0.969755
JPs CHI2 PER NDOF	0.552644
JPs IPCHI2 OWNPV	0.500784

Variable	Importance
B0 DTF CHI2 PER NDOF	26.218139
B0 CHI2 PER NDOF	12.681316
B0 DIRA OWNPV	7.554046
MIN IPCHI2 LEPTONS	7.203161
MAX PT Kst DAUGHTERS	5.511706
B0 FDCHI2 OWNPV	5.498761
B0 PT	4.879305
Kst IPCHI2 OWNPV	3.744466
Kst PT	3.487637
JPs FDCHI2 OWNPV	3.481740
Kst FDCHI2 OWNPV	2.434365
MAX IPCHI2 Kst DAUGHTERS	2.284679
JPs PT	2.150223
B0 IPCHI2 OWNPV	1.901858
Kst CHI2 PER NDOF	1.766270
Kst DIRA OWNPV	1.619896
MIN PT Kst DAUGHTERS	1.609823
MIN PT LEPTONS	1.520805
MIN IPCHI2 Kst DAUGHTERS	1.386947
MAX IPCHI2 LEPTONS	1.206515
JPs DIRA OWNPV	0.923071
JPs IPCHI2 OWNPV	0.650354
JPs CHI2 PER NDOF	0.284917

Variable	Importance
B0 CHI2 PER NDOF	17.260195
B0 DTF CHI2 PER NDOF	14.984281
B0 FDCHI2 OWNPV	12.386995
MIN IPCHI2 LEPTONS	7.363940
B0 DIRA OWNPV	7.274902
MAX PT Kst DAUGHTERS	6.658653
Kst IPCHI2 OWNPV	5.044158
B0 PT	4.105080
JPs PT	3.551213
Kst FDCHI2 OWNPV	2.993690
MIN PT LEPTONS	2.243741
MAX IPCHI2 Kst DAUGHTERS	2.179135
Kst PT	2.176671
B0 IPCHI2 OWNPV	1.910307
MIN IPCHI2 Kst DAUGHTERS	1.649799
Kst CHI2 PER NDOF	1.524150
Kst DIRA OWNPV	1.482986
MIN PT Kst DAUGHTERS	1.266345
JPs FDCHI2 OWNPV	1.189361
JPs IPCHI2 OWNPV	0.838132
JPs CHI2 PER NDOF	0.744357
JPs DIRA OWNPV	0.664269
MAX IPCHI2 LEPTONS	0.507640

Variable	Importance
B0 DTF CHI2 PER NDOF	22.235356
B0 CHI2 PER NDOF	15.739222
B0 DIRA OWNPV	10.759811
Kst FDCHI2 OWNPV	6.802997
MAX PT Kst DAUGHTERS	6.615959
MIN IPCHI2 LEPTONS	5.666139
B0 FDCHI2 OWNPV	4.827997
B0 PT	3.691093
JPs FDCHI2 OWNPV	3.401561
MIN PT LEPTONS	2.616837
JPs PT	2.223186
B0 IPCHI2 OWNPV	2.028298
Kst IPCHI2 OWNPV	2.013672
Kst PT	1.775746
JPs IPCHI2 OWNPV	1.703264
MAX IPCHI2 Kst DAUGHTERS	1.538701
Kst DIRA OWNPV	1.241517
Kst CHI2 PER NDOF	1.191551
MIN PT Kst DAUGHTERS	1.046228
JPs CHI2 PER NDOF	0.890681
MIN IPCHI2 Kst DAUGHTERS	0.789840
JPs DIRA OWNPV	0.681940
MAX IPCHI2 LEPTONS	0.518403

Variable	Importance
B0 CHI2 PER NDOF	19.319857
B0 DTF CHI2 PER NDOF	17.219889
B0 DIRA OWNPV	11.503529
Kst IPCHI2 OWNPV	6.627865
B0 FDCHI2 OWNPV	6.609148
MIN IPCHI2 LEPTONS	5.957205
Kst PT	5.132735
B0 PT	4.476287
MAX PT Kst DAUGHTERS	3.457656
Kst FDCHI2 OWNPV	3.120402
MAX IPCHI2 Kst DAUGHTERS	3.075516
B0 IPCHI2 OWNPV	2.096180
MIN PT LEPTONS	2.003060
JPs PT	1.831881
MIN PT Kst DAUGHTERS	1.571875
Kst CHI2 PER NDOF	1.492236
Kst DIRA OWNPV	1.050309
JPs FDCHI2 OWNPV	1.018558
JPs DIRA OWNPV	0.729318
MAX IPCHI2 LEPTONS	0.717426
JPs CHI2 PER NDOF	0.642680
MIN IPCHI2 Kst DAUGHTERS	0.176164
JPs IPCHI2 OWNPV	0.170226

Variable	Importance
B0 DTF CHI2 PER NDOF	21.282906
B0 CHI2 PER NDOF	12.077064
MIN IPCHI2 LEPTONS	9.611817
B0 DIRA OWNPV	8.021437
B0 PT	5.765542
MAX PT Kst DAUGHTERS	5.289789
MAX IPCHI2 Kst DAUGHTERS	5.233300
B0 FDCHI2 OWNPV	4.937736
Kst IPCHI2 OWNPV	4.594046
Kst PT	4.092174
JPs PT	3.163775
MIN PT LEPTONS	2.552256
Kst DIRA OWNPV	2.272000
B0 IPCHI2 OWNPV	2.263158
MIN PT Kst DAUGHTERS	1.881584
Kst CHI2 PER NDOF	1.493395
JPs FDCHI2 OWNPV	1.480247
MIN IPCHI2 Kst DAUGHTERS	1.001766
MAX IPCHI2 LEPTONS	0.979264
JPs DIRA OWNPV	0.780633
Kst FDCHI2 OWNPV	0.617990
JPs CHI2 PER NDOF	0.578976
JPs IPCHI2 OWNPV	0.029147

Variable	Importance
B0 DTF CHI2 PER NDOF	24.522575
B0 CHI2 PER NDOF	14.454434
B0 FDCHI2 OWNPV	8.096518
B0 DIRA OWNPV	7.634135
MIN IPCHI2 LEPTONS	5.583471
B0 PT	5.404441
Kst IPCHI2 OWNPV	4.170739
MAX PT Kst DAUGHTERS	3.788540
Kst PT	3.197318
JPs FDCHI2 OWNPV	3.190685
JPs PT	2.742337
MAX IPCHI2 Kst DAUGHTERS	2.604921
MIN PT LEPTONS	2.285838
Kst DIRA OWNPV	2.018839
MIN PT Kst DAUGHTERS	1.856894
B0 IPCHI2 OWNPV	1.822580
Kst FDCHI2 OWNPV	1.741974
Kst CHI2 PER NDOF	1.054074
JPs DIRA OWNPV	1.026894
JPs CHI2 PER NDOF	0.933917
MIN IPCHI2 Kst DAUGHTERS	0.684694
JPs IPCHI2 OWNPV	0.630300
MAX IPCHI2 LEPTONS	0.553881

Variable	Importance
B0 CHI2 PER NDOF	18.634540
B0 DTF CHI2 PER NDOF	15.568392
B0 DIRA OWNPV	12.479786
MIN IPCHI2 LEPTONS	6.450744
MAX PT Kst DAUGHTERS	5.637644
JPs FDCHI2 OWNPV	4.763958
Kst IPCHI2 OWNPV	4.155158
MIN IPCHI2 Kst DAUGHTERS	3.963534
B0 PT	3.251701
MIN PT LEPTONS	3.171969
Kst PT	3.166386
B0 FDCHI2 OWNPV	2.956112
MAX IPCHI2 Kst DAUGHTERS	2.728116
JPs PT	2.515941
MIN PT Kst DAUGHTERS	1.884759
Kst FDCHI2 OWNPV	1.708110
B0 IPCHI2 OWNPV	1.577116
Kst DIRA OWNPV	1.332130
Kst CHI2 PER NDOF	1.264976
MAX IPCHI2 LEPTONS	1.233314
JPs IPCHI2 OWNPV	0.809050
JPs CHI2 PER NDOF	0.475891
JPs DIRA OWNPV	0.270674

Variable	Importance
B0 DTF CHI2 PER NDOF	21.158234
B0 CHI2 PER NDOF	14.633131
B0 FDCHI2 OWNPV	12.631195
B0 DIRA OWNPV	8.285425
MIN IPCHI2 LEPTONS	5.803865
B0 PT	4.835802
Kst FDCHI2 OWNPV	4.520894
Kst PT	4.189076
MAX PT Kst DAUGHTERS	3.559035
Kst IPCHI2 OWNPV	3.375686
MIN PT LEPTONS	2.192994
JPs PT	2.075090
B0 IPCHI2 OWNPV	1.881771
MAX IPCHI2 Kst DAUGHTERS	1.837921
Kst CHI2 PER NDOF	1.390739
Kst DIRA OWNPV	1.341066
JPs FDCHI2 OWNPV	1.223610
MIN PT Kst DAUGHTERS	1.167292
MIN IPCHI2 Kst DAUGHTERS	0.993805
MAX IPCHI2 LEPTONS	0.880843
JPs DIRA OWNPV	0.778680
JPs IPCHI2 OWNPV	0.684297
JPs CHI2 PER NDOF	0.559548

Variable	Importance
B0 DTF CHI2 PER NDOF	22.303124
B0 CHI2 PER NDOF	14.384694
MIN IPCHI2 LEPTONS	8.506604
B0 DIRA OWNPV	8.126715
MAX PT Kst DAUGHTERS	6.614252
B0 PT	4.743232
B0 FDCHI2 OWNPV	4.298967
Kst FDCHI2 OWNPV	3.483458
JPs PT	2.939072
MAX IPCHI2 Kst DAUGHTERS	2.769486
Kst PT	2.612802
MIN PT Kst DAUGHTERS	2.521862
JPs FDCHI2 OWNPV	2.290919
MIN PT LEPTONS	2.231173
Kst IPCHI2 OWNPV	2.145055
B0 IPCHI2 OWNPV	1.849059
Kst CHI2 PER NDOF	1.739310
MIN IPCHI2 Kst DAUGHTERS	1.724509
Kst DIRA OWNPV	1.471048
JPs DIRA OWNPV	1.227973
JPs IPCHI2 OWNPV	0.855782
JPs CHI2 PER NDOF	0.588544
MAX IPCHI2 LEPTONS	0.572359

D.2.3.2 MVA Correlations with B^0 mass and q^2 Run 2p1

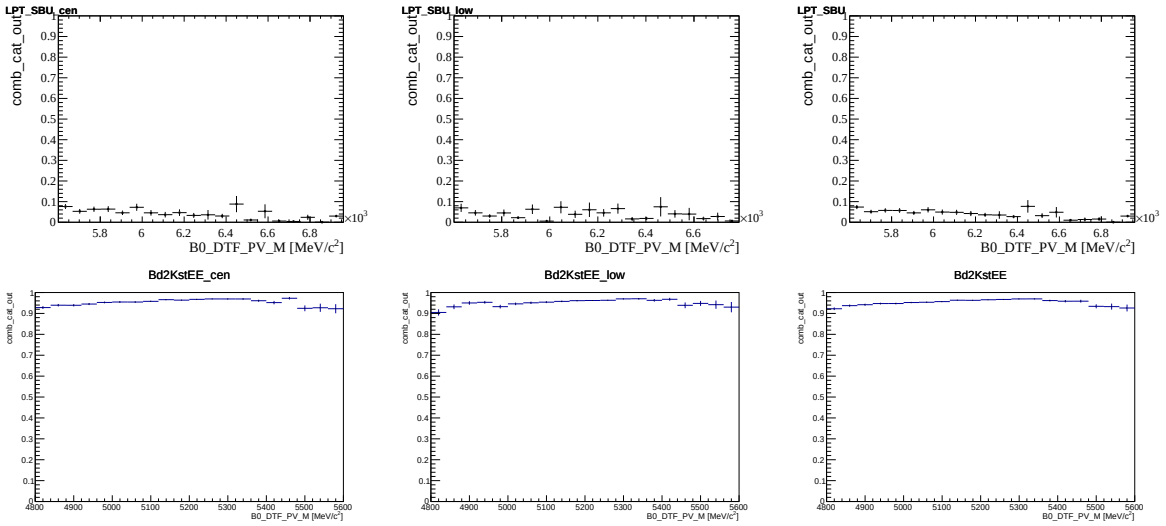


Figure 5: (Top) From left to right, average value of the MVA classifier in collision data sideband as a function of the `DecayTreeFitter` B^0 mass in the central, low and combined low and central regions respectively in Run 2p1. (Bottom) From left to right, average value of the MVA classifier in simulated Bd2KstEE MC as a function of the `DecayTreeFitter` B^0 mass in the central, low and combined low and central regions respectively in Run 2p1.

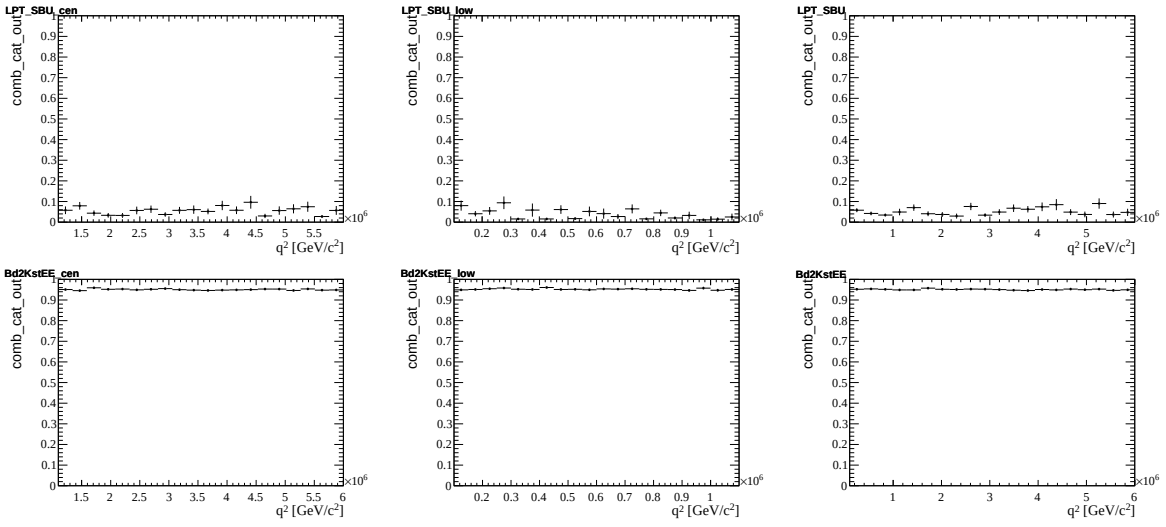


Figure 6: (Top) From left to right, average value of the MVA classifier in collision data sideband as a function of q^2 in the central, low and combined low and central regions respectively in Run 2p1. (Bottom) From left to right, average value of the MVA classifier in simulated Bd2KstEE MC as a function of q^2 in the central, low and combined low and central regions respectively in Run 2p1.

D.2.4 $B^+ \to K^+e^+e^-$ $\mathbf{MVA_{Comb}}$ Training

D.2.4.1 Feature Importances Run 2p1

Table 4: Feature importances per fold, with fold number increasing from left to right

Variable	Importance
Bp DTF CHI2 PER NDOF	20.986322
Bp CHI2 PER NDOF	14.946616
K IPCHI2 OWNPV	11.555545
K PT	9.710681
MIN IPCHI2 LEPTONS	7.504770
Bp PT	7.429543
JPs FDCHI2 OWNPV	6.019837
Bp DIRA OWNPV	5.872263
MIN PT LEPTONS	4.208929
JPs PT	3.108141
Bp FDCHI2 OWNPV	2.353763
JPs IPCHI2 OWNPV	1.991136
MAX IPCHI2 LEPTONS	1.673940
JPs DIRA OWNPV	1.617919
Bp IPCHI2 OWNPV	0.598073
JPs CHI2 PER NDOF	0.422521

Variable	Importance
Bp DTF CHI2 PER NDOF	19.867824
Bp CHI2 PER NDOF	14.772028
K PT	10.704791
K IPCHI2 OWNPV	10.384062
MIN IPCHI2 LEPTONS	9.095050
Bp DIRA OWNPV	7.317797
Bp PT	6.259251
MIN PT LEPTONS	5.435628
JPs FDCHI2 OWNPV	4.460612
Bp FDCHI2 OWNPV	4.144714
JPs PT	3.088133
JPs DIRA OWNPV	1.659410
JPs IPCHI2 OWNPV	1.613387
MAX IPCHI2 LEPTONS	0.675529
JPs CHI2 PER NDOF	0.345317
Bp IPCHI2 OWNPV	0.176467

Variable	Importance
Bp DTF CHI2 PER NDOF	19.200798
Bp CHI2 PER NDOF	15.209729
K IPCHI2 OWNPV	11.133195
K PT	9.673051
MIN IPCHI2 LEPTONS	7.865381
Bp DIRA OWNPV	7.053625
Bp PT	6.951377
JPs PT	4.855062
MIN PT LEPTONS	4.156605
Bp FDCHI2 OWNPV	3.877159
JPs FDCHI2 OWNPV	3.409880
JPs IPCHI2 OWNPV	1.890292
MAX IPCHI2 LEPTONS	1.799865
JPs DIRA OWNPV	1.683524
Bp IPCHI2 OWNPV	0.769117
JPs CHI2 PER NDOF	0.471342

Variable	Importance
Bp DTF CHI2 PER NDOF	18.574625
Bp CHI2 PER NDOF	15.243703
K PT	11.190513
K IPCHI2 OWNPV	10.438998
MIN IPCHI2 LEPTONS	7.971137
Bp PT	6.788005
Bp DIRA OWNPV	6.388835
MIN PT LEPTONS	5.669763
JPs FDCHI2 OWNPV	5.063372
JPs PT	4.190354
Bp FDCHI2 OWNPV	3.106180
JPs IPCHI2 OWNPV	2.393791
JPs DIRA OWNPV	1.630662
MAX IPCHI2 LEPTONS	0.742224
Bp IPCHI2 OWNPV	0.547175
JPs CHI2 PER NDOF	0.060663

Variable	Importance
Bp CHI2 PER NDOF	17.847519
Bp DTF CHI2 PER NDOF	17.728521
K PT	10.383826
K IPCHI2 OWNPV	10.260248
Bp DIRA OWNPV	8.175443
MIN IPCHI2 LEPTONS	7.563445
Bp PT	7.455178
JPs PT	4.164757
MIN PT LEPTONS	3.855445
JPs FDCHI2 OWNPV	3.511913
Bp FDCHI2 OWNPV	3.426429
MAX IPCHI2 LEPTONS	1.957720
JPs DIRA OWNPV	1.805522
JPs IPCHI2 OWNPV	1.020748
Bp IPCHI2 OWNPV	0.492313
JPs CHI2 PER NDOF	0.350970

Variable	Importance
Bp CHI2 PER NDOF	16.942522
Bp DTF CHI2 PER NDOF	13.326320
K IPCHI2 OWNPV	11.482109
K PT	10.649647
MIN IPCHI2 LEPTONS	8.494167
Bp DIRA OWNPV	8.068100
Bp PT	6.570187
MIN PT LEPTONS	5.201596
JPs PT	4.414425
Bp FDCHI2 OWNPV	4.051548
JPs FDCHI2 OWNPV	3.376058
JPs IPCHI2 OWNPV	2.182766
MAX IPCHI2 LEPTONS	1.899644
JPs DIRA OWNPV	1.580227
Bp IPCHI2 OWNPV	1.505103
JPs CHI2 PER NDOF	0.255582

Variable	Importance
Bp DTF CHI2 PER NDOF	18.697397
Bp CHI2 PER NDOF	15.548025
K IPCHI2 OWNPV	10.875217
K PT	10.397279
MIN IPCHI2 LEPTONS	8.395348
Bp DIRA OWNPV	7.089075
Bp PT	6.306125
JPs FDCHI2 OWNPV	5.781352
JPs PT	5.674433
MIN PT LEPTONS	4.234148
JPs DIRA OWNPV	2.054630
JPs IPCHI2 OWNPV	1.875300
Bp FDCHI2 OWNPV	1.633367
MAX IPCHI2 LEPTONS	0.895631
Bp IPCHI2 OWNPV	0.470435
JPs CHI2 PER NDOF	0.072239

Variable	Importance
Bp DTF CHI2 PER NDOF	18.534650
Bp CHI2 PER NDOF	13.600816
K IPCHI2 OWNPV	10.795628
K PT	10.130503
Bp PT	7.540551
MIN IPCHI2 LEPTONS	7.467232
Bp DIRA OWNPV	6.525165
Bp FDCHI2 OWNPV	5.545402
MIN PT LEPTONS	5.171396
JPs PT	4.677906
JPs FDCHI2 OWNPV	4.170451
JPs IPCHI2 OWNPV	2.276838
JPs DIRA OWNPV	2.070900
Bp IPCHI2 OWNPV	0.766376
MAX IPCHI2 LEPTONS	0.565097
JPs CHI2 PER NDOF	0.161087

Variable	Importance
Bp CHI2 PER NDOF	15.837017
Bp DTF CHI2 PER NDOF	14.077389
K PT	11.929834
K IPCHI2 OWNPV	11.564200
MIN IPCHI2 LEPTONS	8.551971
Bp DIRA OWNPV	8.036529
Bp PT	6.914631
JPs PT	5.713394
MIN PT LEPTONS	4.117744
JPs FDCHI2 OWNPV	3.799124
Bp FDCHI2 OWNPV	2.582470
JPs IPCHI2 OWNPV	2.450384
MAX IPCHI2 LEPTONS	1.550947
JPs DIRA OWNPV	1.469809
Bp IPCHI2 OWNPV	0.904152
JPs CHI2 PER NDOF	0.500406

Variable	Importance
Bp DTF CHI2 PER NDOF	20.566687
Bp CHI2 PER NDOF	14.312615
K IPCHI2 OWNPV	11.975107
K PT	10.528271
MIN IPCHI2 LEPTONS	7.526649
Bp PT	6.699532
JPs FDCHI2 OWNPV	5.823341
Bp DIRA OWNPV	5.693169
MIN PT LEPTONS	4.236168
JPs PT	3.303655
JPs IPCHI2 OWNPV	2.877168
Bp FDCHI2 OWNPV	2.416441
MAX IPCHI2 LEPTONS	1.650619
JPs DIRA OWNPV	1.286016
JPs CHI2 PER NDOF	0.605663
Bp IPCHI2 OWNPV	0.498897

D.2.4.2 MVA Correlations with B^+ mass and q^2 Run 2p1

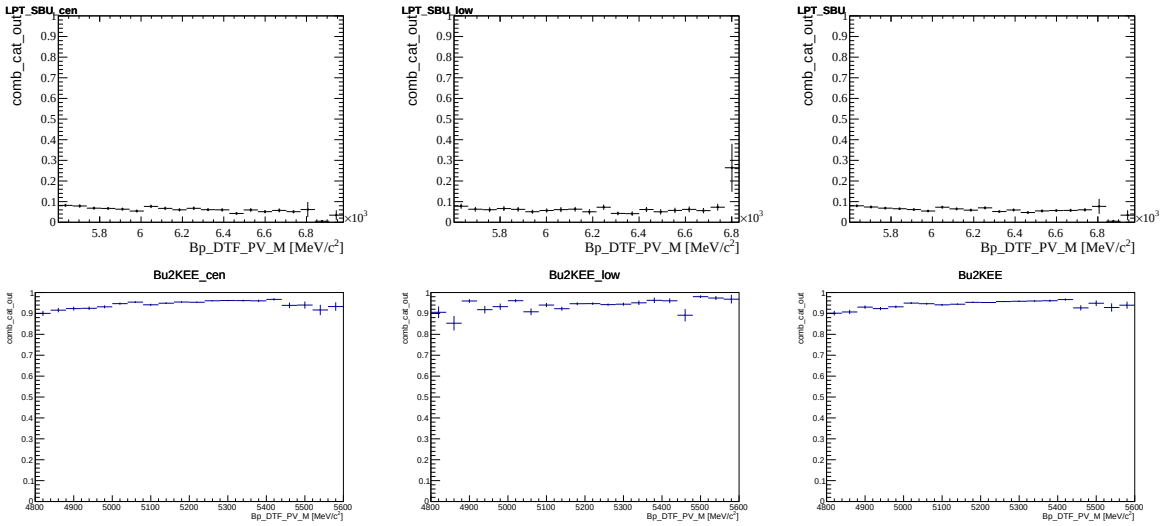


Figure 7: (Top) From left to right, average value of the MVA classifier in collision data sideband as a function of the `DecayTreeFitter` B^+ mass in the central, low and combined low and central regions respectively in Run 2p1. (Bottom) From left to right, average value of the MVA classifier in simulated Bu2KEE MC as a function of the `DecayTreeFitter` B^+ mass in the central, low and combined low and central regions respectively in Run 2p1.

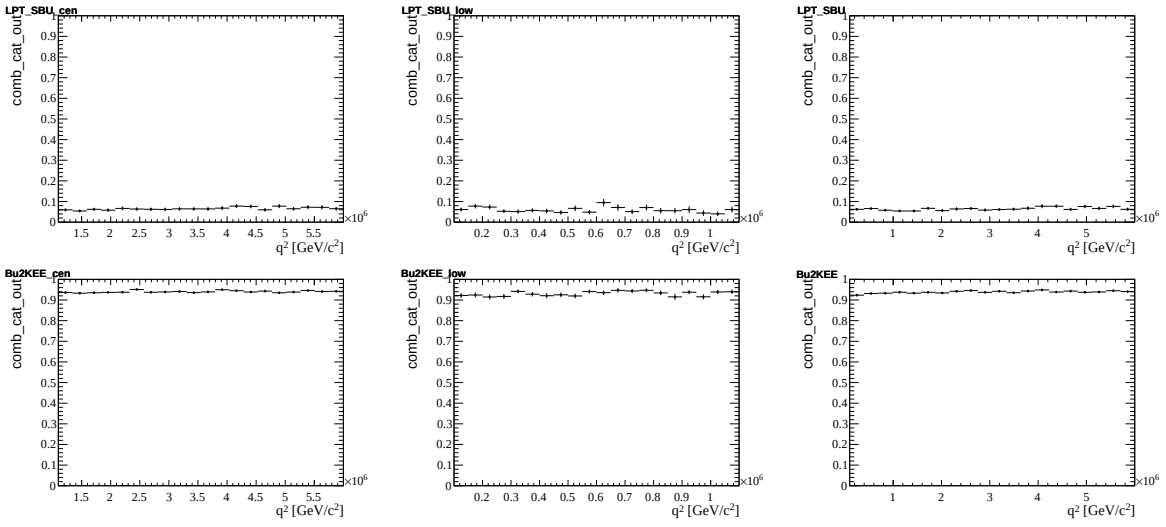


Figure 8: (Top) From left to right, average value of the MVA classifier in collision data sideband as a function of q^2 in the central, low and combined low and central regions respectively in Run 2p1. (Bottom) From left to right, average value of the MVA classifier in simulated Bu2KEE MC as a function of q^2 in the central, low and combined low and central regions respectively in Run 2p1.

D.2.5 $B^0 \to K^{*0} e^+ e^-$ $\mathbf{MVA_{PReco}}$ Training

D.2.5.1 Feature Importances Run 2p1

Table 5: Feature importances per fold, with fold number increasing from left to right

Variable	Importance
B0 VI OTM	54.108907
B0 VI OTC	13.330499
B0 IPCHI2 OWNPV	10.836697
B0 DIRA OWNPV	4.875080
JPs IPCHI2 OWNPV	4.207285
Kst IPCHI2 OWNPV	3.731536
JPs FDCHI2 OWNPV	2.016539
MIN PT Kst DAUGHTERS	1.921139
JPs DIRA OWNPV	1.819822
Kst DIRA OWNPV	1.260707
MIN B0 CC MULT L	1.094286
B0 CHI2 PER NDOF	0.387424
B0 DTF CHI2 PER NDOF	0.315808
Kst CHI2 PER NDOF	0.094272

Variable	Importance
B0 VI OTM	59.171170
B0 VI OTC	11.417193
B0 IPCHI2 OWNPV	10.166204
B0 DIRA OWNPV	4.009043
JPs IPCHI2 OWNPV	3.881266
Kst IPCHI2 OWNPV	3.608483
MIN PT Kst DAUGHTERS	1.731470
JPs FDCHI2 OWNPV	1.638394
JPs DIRA OWNPV	1.352599
Kst DIRA OWNPV	1.240813
MIN B0 CC MULT L	0.990309
B0 CHI2 PER NDOF	0.379217
B0 DTF CHI2 PER NDOF	0.341187
Kst CHI2 PER NDOF	0.072652

Variable	Importance
B0 VI OTM	58.742230
B0 VI OTC	12.154865
B0 IPCHI2 OWNPV	9.402263
B0 DIRA OWNPV	3.881093
Kst IPCHI2 OWNPV	3.787615
JPs IPCHI2 OWNPV	3.740636
MIN PT Kst DAUGHTERS	1.732706
JPs FDCHI2 OWNPV	1.634358
JPs DIRA OWNPV	1.615496
Kst DIRA OWNPV	1.187329
MIN B0 CC MULT L	1.114453
B0 DTF CHI2 PER NDOF	0.578231
B0 CHI2 PER NDOF	0.281092
Kst CHI2 PER NDOF	0.147631

Variable	Importance
B0 VI OTM	60.939633
B0 VI OTC	10.601496
B0 IPCHI2 OWNPV	8.856827
B0 DIRA OWNPV	5.190333
Kst IPCHI2 OWNPV	3.494290
JPs IPCHI2 OWNPV	3.048755
JPs DIRA OWNPV	2.151215
JPs FDCHI2 OWNPV	1.832866
MIN PT Kst DAUGHTERS	1.478257
Kst DIRA OWNPV	1.086176
MIN B0 CC MULT L	0.802912
B0 CHI2 PER NDOF	0.293419
B0 DTF CHI2 PER NDOF	0.132462
Kst CHI2 PER NDOF	0.091359

Variable	Importance
B0 VI OTM	58.326343
B0 IPCHI2 OWNPV	11.751013
B0 VI OTC	11.191360
JPs IPCHI2 OWNPV	4.438736
Kst IPCHI2 OWNPV	3.751293
B0 DIRA OWNPV	2.675272
JPs FDCHI2 OWNPV	1.794156
MIN PT Kst DAUGHTERS	1.705679
JPs DIRA OWNPV	1.482113
Kst DIRA OWNPV	1.262676
MIN B0 CC MULT L	1.012584
B0 CHI2 PER NDOF	0.422913
B0 DTF CHI2 PER NDOF	0.141855
Kst CHI2 PER NDOF	0.044006

Variable	Importance
B0 VI OTM	59.212294
B0 VI OTC	11.214206
B0 IPCHI2 OWNPV	9.404367
B0 DIRA OWNPV	5.107235
Kst IPCHI2 OWNPV	3.773572
JPs IPCHI2 OWNPV	3.580774
JPs DIRA OWNPV	1.937661
MIN PT Kst DAUGHTERS	1.494729
JPs FDCHI2 OWNPV	1.398495
Kst DIRA OWNPV	1.184845
MIN B0 CC MULT L	1.160525
B0 DTF CHI2 PER NDOF	0.271333
B0 CHI2 PER NDOF	0.230380
Kst CHI2 PER NDOF	0.029583

Variable	Importance
B0 VI OTM	58.412878
B0 VI OTC	12.448151
B0 IPCHI2 OWNPV	9.721853
B0 DIRA OWNPV	3.942044
JPs IPCHI2 OWNPV	3.820745
Kst IPCHI2 OWNPV	3.417301
MIN PT Kst DAUGHTERS	1.856521
JPs FDCHI2 OWNPV	1.829513
JPs DIRA OWNPV	1.476761
Kst DIRA OWNPV	1.192915
MIN B0 CC MULT L	0.859698
B0 DTF CHI2 PER NDOF	0.618069
B0 CHI2 PER NDOF	0.308041
Kst CHI2 PER NDOF	0.095510

Variable	Importance
B0 VI OTM	60.970120
B0 VI OTC	10.544439
B0 IPCHI2 OWNPV	9.654856
B0 DIRA OWNPV	4.337851
Kst IPCHI2 OWNPV	3.457084
JPs IPCHI2 OWNPV	3.451936
JPs FDCHI2 OWNPV	1.653595
MIN PT Kst DAUGHTERS	1.621266
JPs DIRA OWNPV	1.576407
Kst DIRA OWNPV	1.203389
MIN B0 CC MULT L	0.876775
B0 CHI2 PER NDOF	0.419108
B0 DTF CHI2 PER NDOF	0.199705
Kst CHI2 PER NDOF	0.033469

Variable	Importance
B0 VI OTM	59.272556
B0 VI OTC	11.142445
B0 IPCHI2 OWNPV	9.872244
B0 DIRA OWNPV	4.229235
Kst IPCHI2 OWNPV	3.767562
JPs IPCHI2 OWNPV	3.386924
JPs DIRA OWNPV	1.982515
JPs FDCHI2 OWNPV	1.882063
MIN PT Kst DAUGHTERS	1.696677
MIN B0 CC MULT L	1.011858
Kst DIRA OWNPV	0.930936
B0 DTF CHI2 PER NDOF	0.382645
B0 CHI2 PER NDOF	0.328518
Kst CHI2 PER NDOF	0.113821

Variable	Importance
B0 VI OTM	59.619803
B0 VI OTC	12.561652
B0 IPCHI2 OWNPV	9.023605
JPs IPCHI2 OWNPV	3.797659
Kst IPCHI2 OWNPV	3.596378
B0 DIRA OWNPV	3.434145
JPs FDCHI2 OWNPV	1.748789
MIN PT Kst DAUGHTERS	1.648075
JPs DIRA OWNPV	1.637721
Kst DIRA OWNPV	1.135246
MIN B0 CC MULT L	0.695475
B0 DTF CHI2 PER NDOF	0.499633
B0 CHI2 PER NDOF	0.403790
Kst CHI2 PER NDOF	0.198029

D.2.5.2 MVA Correlations with B^0 mass and q^2 Run 2p1

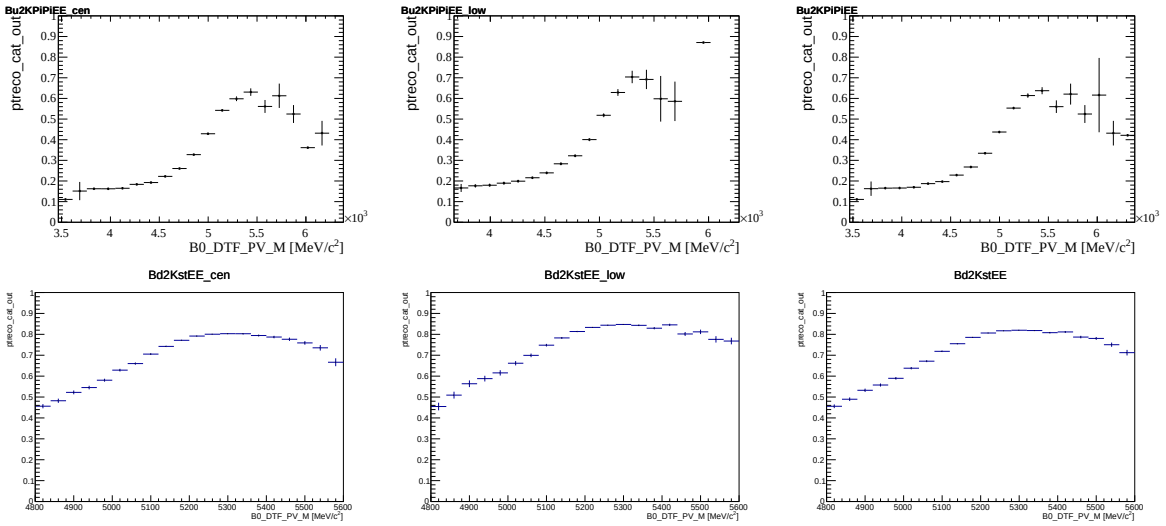


Figure 9: (Top) From left to right, average value of the MVA classifier in partially reconstructed Bu2KPiPiEE MC as a function of the `DecayTreeFitter` B^0 mass in the central, low and combined low and central regions respectively in Run 2p1. (Bottom) From left to right, average value of the MVA classifier in simulated Bd2KstEE MC as a function of the `DecayTreeFitter` B^0 mass in the central, low and combined low and central regions respectively in Run 2p1.

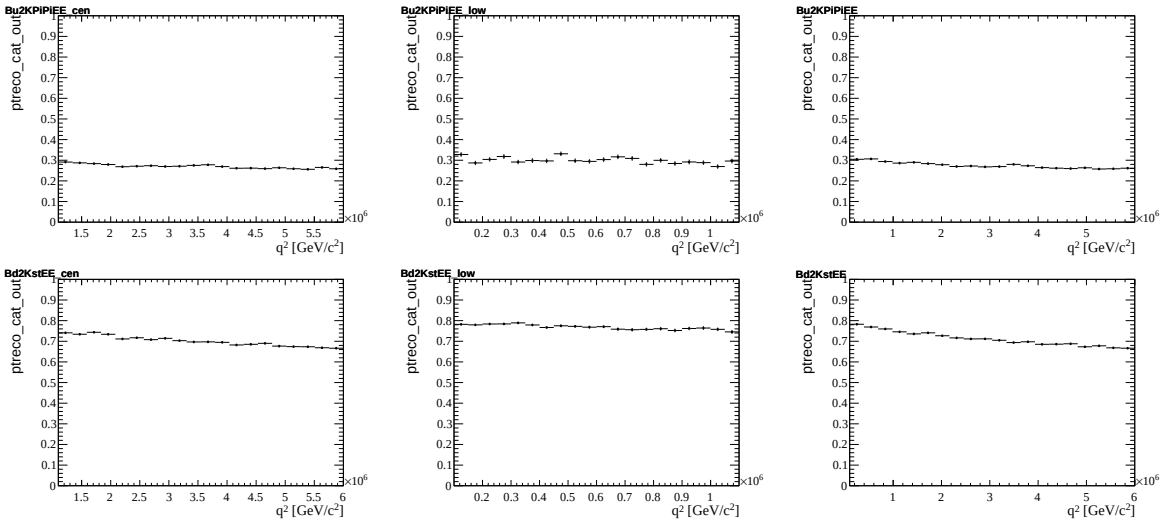


Figure 10: (Top) From left to right, average value of the MVA classifier in partially reconstructed Bu2KPiPiEE MC as a function of q^2 in the central, low and combined low and central regions respectively in Run 2p1. (Bottom) From left to right, average value of the MVA classifier in simulated Bd2KstEE MC as a function of q^2 in the central, low and combined low and central regions respectively in Run 2p1.

D.2.6 $B^+ \to K^+e^+e^-$ $\mathbf{MVA_{PReco}}$ Training

D.2.6.1 Feature Importances Run 2p1

Table 6: Feature importances per fold, with fold number increasing from left to right

Variable	Importance
Bp VI OTM	55.275036
Bp VI OTC	15.586555
Bp IPCHI2 OWNPV	7.207027
Bp DIRA OWNPV	5.404814
K PT	4.665135
JPs FDCHI2 OWNPV	3.067740
JPs IPCHI2 OWNPV	3.014079
JPs DIRA OWNPV	1.220405
MIN IPCHI2 LEPTONS	1.142098
MIN Bp CC IT L	0.967238
Bp PT	0.936022
MIN Bp CC APT L	0.662173
MAX Bp CC APT L	0.577821
MAX Bp CC MULT L	0.273857

Variable	Importance
Bp VI OTM	58.888358
Bp VI OTC	13.666674
Bp IPCHI2 OWNPV	6.110489
Bp DIRA OWNPV	5.443260
K PT	4.504403
JPs IPCHI2 OWNPV	3.589415
JPs FDCHI2 OWNPV	2.293854
Bp PT	1.315806
JPs DIRA OWNPV	1.042140
MAX Bp CC APT L	0.877366
MIN Bp CC IT L	0.824594
MAX Bp CC MULT L	0.669398
MIN IPCHI2 LEPTONS	0.442518
MIN Bp CC APT L	0.331725

Variable	Importance
Bp VI OTM	56.344568
Bp VI OTC	13.913524
Bp DIRA OWNPV	7.024650
Bp IPCHI2 OWNPV	6.057045
K PT	4.562353
JPs IPCHI2 OWNPV	3.954887
JPs FDCHI2 OWNPV	2.490434
MIN Bp CC IT L	1.256515
Bp PT	1.170394
JPs DIRA OWNPV	0.958478
MIN IPCHI2 LEPTONS	0.804571
MAX Bp CC APT L	0.596268
MAX Bp CC MULT L	0.532788
MIN Bp CC APT L	0.333524

Variable	Importance
Bp VI OTM	58.152368
Bp VI OTC	13.272599
Bp DIRA OWNPV	7.436811
Bp IPCHI2 OWNPV	5.589632
K PT	4.467178
JPs IPCHI2 OWNPV	4.064862
JPs FDCHI2 OWNPV	1.912636
JPs DIRA OWNPV	0.898593
MAX Bp CC APT L	0.884691
Bp PT	0.797685
MAX Bp CC MULT L	0.737469
MIN IPCHI2 LEPTONS	0.638830
MIN Bp CC IT L	0.604651
MIN Bp CC APT L	0.541996

Variable	Importance
Bp VI OTM	52.212806
Bp VI OTC	16.026012
Bp IPCHI2 OWNPV	7.592428
Bp DIRA OWNPV	6.290862
K PT	5.050052
JPs IPCHI2 OWNPV	4.043881
JPs FDCHI2 OWNPV	2.303255
JPs DIRA OWNPV	1.351607
Bp PT	1.333139
MAX Bp CC MULT L	1.052699
MIN IPCHI2 LEPTONS	1.037779
MIN Bp CC IT L	0.629292
MIN Bp CC APT L	0.597438
MAX Bp CC APT L	0.478749

Variable	Importance
Bp VI OTM	57.106235
Bp VI OTC	14.112962
Bp IPCHI2 OWNPV	7.149872
Bp DIRA OWNPV	5.978905
K PT	3.413784
JPs IPCHI2 OWNPV	3.150741
JPs FDCHI2 OWNPV	2.194361
Bp PT	1.639099
MIN IPCHI2 LEPTONS	1.298000
JPs DIRA OWNPV	1.153501
MIN Bp CC IT L	0.972226
MAX Bp CC MULT L	0.752324
MAX Bp CC APT L	0.713532
MIN Bp CC APT L	0.364456

Variable	Importance
Bp VI OTM	56.296704
Bp VI OTC	13.567165
Bp DIRA OWNPV	7.185467
Bp IPCHI2 OWNPV	6.148349
K PT	5.232922
JPs IPCHI2 OWNPV	3.023673
JPs FDCHI2 OWNPV	2.386249
JPs DIRA OWNPV	1.214056
Bp PT	1.079663
MIN Bp CC IT L	1.058411
MAX Bp CC APT L	0.775527
MIN IPCHI2 LEPTONS	0.763945
MIN Bp CC APT L	0.665440
MAX Bp CC MULT L	0.602429

Variable	Importance
Bp VI OTM	59.326030
Bp VI OTC	12.670140
Bp DIRA OWNPV	6.957940
Bp IPCHI2 OWNPV	5.334157
K PT	4.685086
JPs IPCHI2 OWNPV	3.225850
JPs FDCHI2 OWNPV	1.876565
Bp PT	1.219020
JPs DIRA OWNPV	1.131445
MIN Bp CC IT L	1.040947
MIN Bp CC APT L	0.905957
MIN IPCHI2 LEPTONS	0.651166
MAX Bp CC APT L	0.496666
MAX Bp CC MULT L	0.479032

Variable	Importance
Bp VI OTM	59.206030
Bp VI OTC	12.093259
Bp DIRA OWNPV	6.993650
Bp IPCHI2 OWNPV	6.193717
K PT	4.386747
JPs IPCHI2 OWNPV	3.358626
JPs FDCHI2 OWNPV	1.698035
MIN IPCHI2 LEPTONS	1.560322
MIN Bp CC IT L	1.261660
Bp PT	1.095810
MAX Bp CC MULT L	0.875149
JPs DIRA OWNPV	0.758164
MAX Bp CC APT L	0.441911
MIN Bp CC APT L	0.076919

Variable	Importance
Bp VI OTM	55.703138
Bp VI OTC	14.139228
Bp IPCHI2 OWNPV	6.781459
Bp DIRA OWNPV	6.769060
K PT	4.276000
JPs IPCHI2 OWNPV	3.323254
JPs FDCHI2 OWNPV	2.951671
MIN IPCHI2 LEPTONS	1.216601
Bp PT	1.146847
JPs DIRA OWNPV	1.064775
MIN Bp CC IT L	0.712155
MAX Bp CC MULT L	0.678799
MIN Bp CC APT L	0.649672
MAX Bp CC APT L	0.587342

D.2.6.2 MVA Correlations with B^+ mass and q^2 Run 2p1

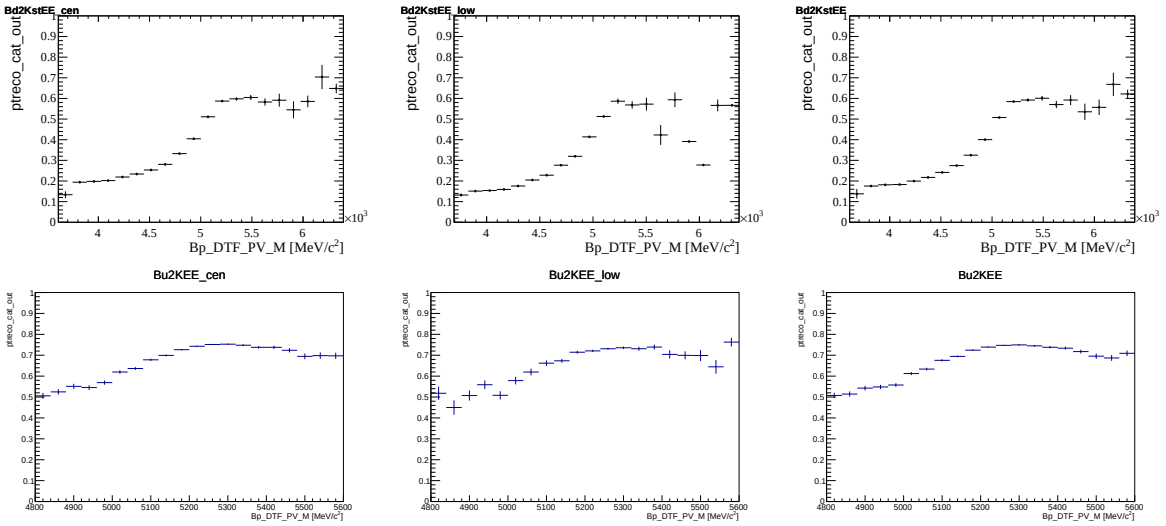


Figure 11: (Top) From left to right, average value of the MVA classifier in partially reconstructed Bd2KstEE MC as a function of the `DecayTreeFitter` B^+ mass in the central, low and combined low and central regions respectively in Run 2p1. (Bottom) From left to right, average value of the MVA classifier in simulated Bu2KEE MC as a function of the `DecayTreeFitter` B^+ mass in the central, low and combined low and central regions respectively in Run 2p1.

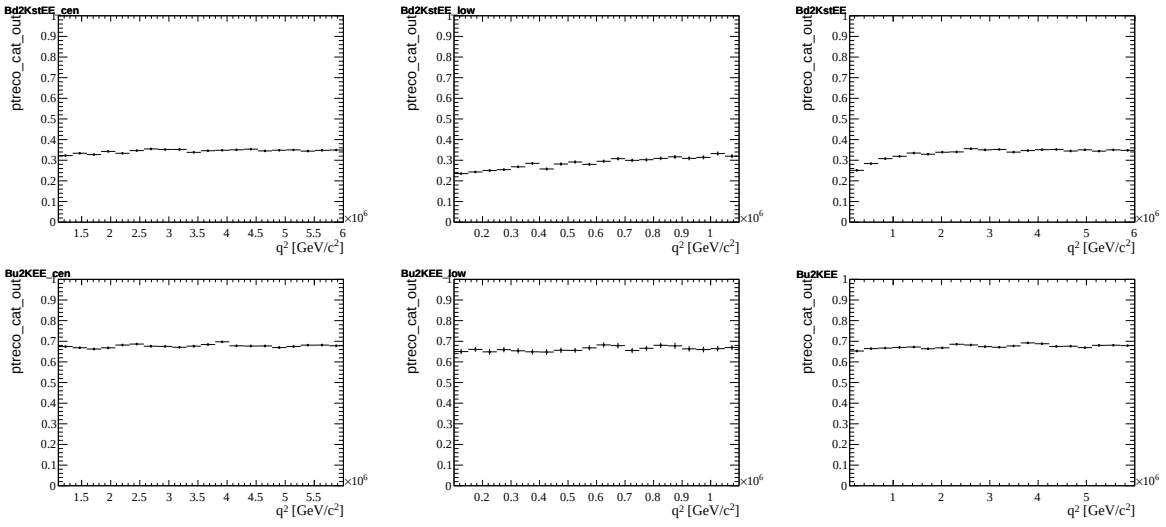


Figure 12: (Top) From left to right, average value of the MVA classifier in partially reconstructed Bd2KstEE MC as a function of q^2 in the central, low and combined low and central regions respectively in Run 2p1. (Bottom) From left to right, average value of the MVA classifier in simulated Bu2KEE MC as a function of q^2 in the central, low and combined low and central regions respectively in Run 2p1.

Appendix E

Data/simulation comparisons

1 Bu2KJPsMM R2p1 Bp kde

1.1 L0Lexclusive

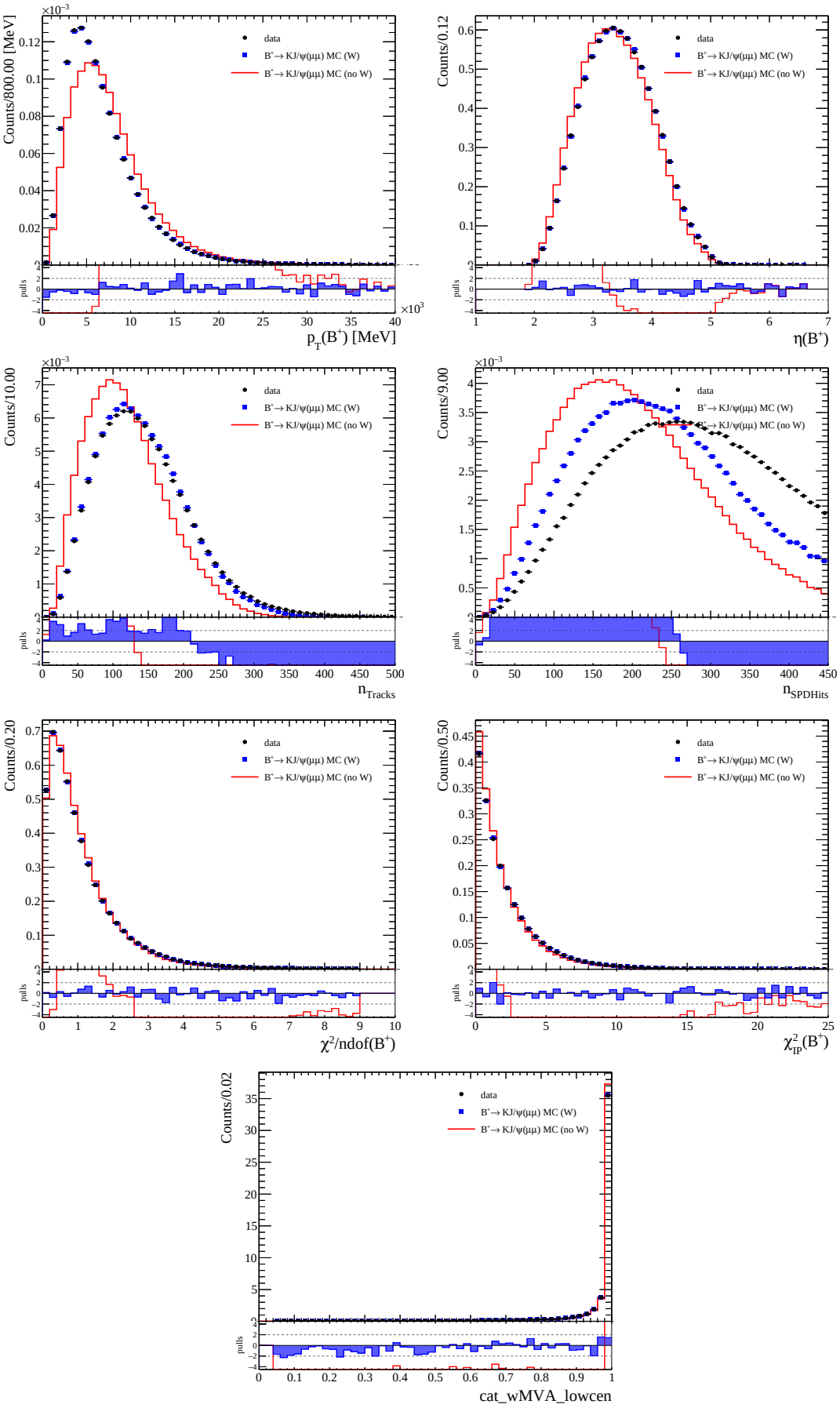


Figure 1

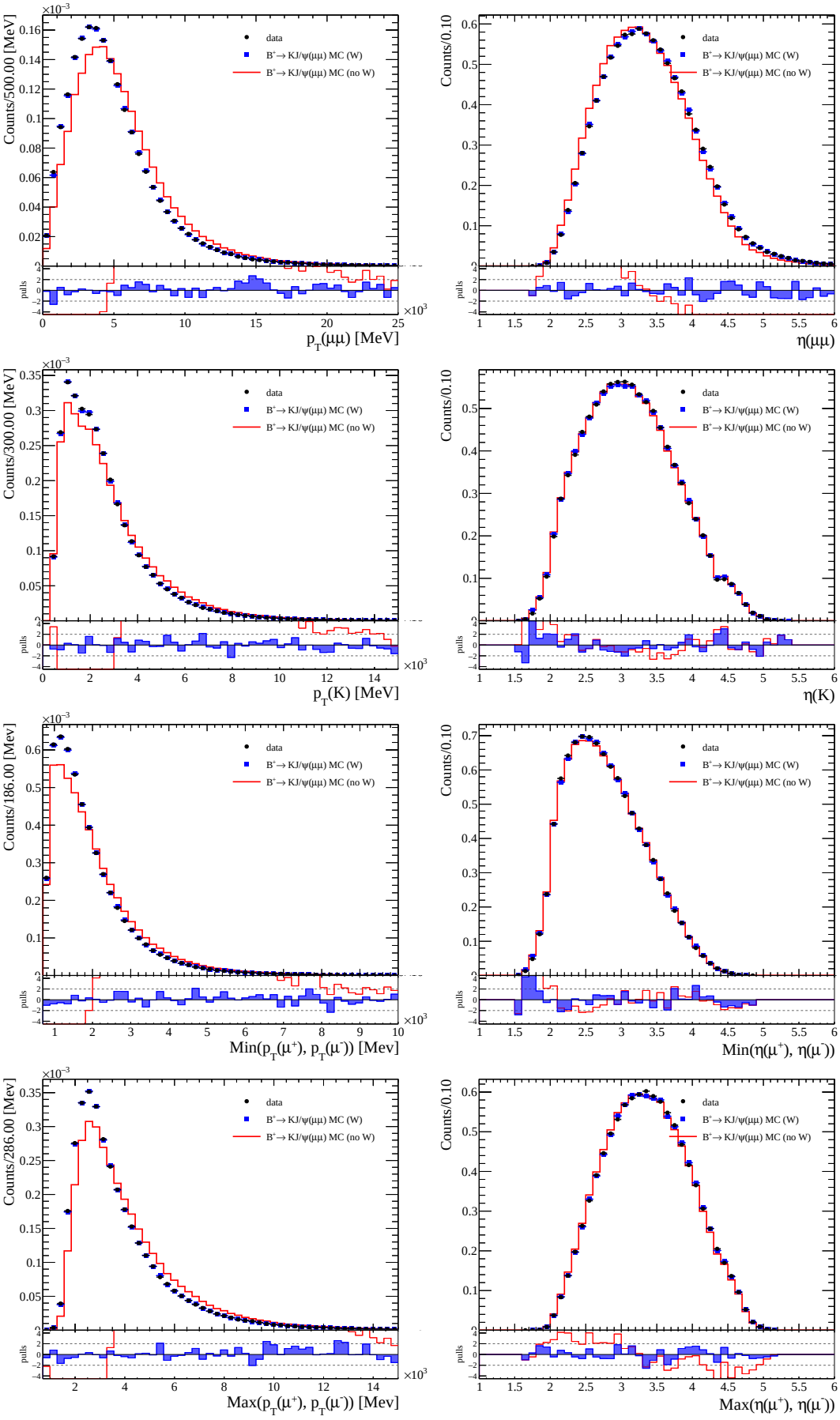

data
B+→ KJ/ψ(μμ) MC (W)
B+→ KJ/ψ(μμ) MC (no W)
Counts/500.00 [MeV]
pT(μμ) [MeV]
Counts/0.10
η(μμ)
Counts/300.00 [MeV]
pT(K) [MeV]
η(K)
Counts/186.00 [Mev]
Min(pT(μ+), pT(μ-)) [Mev]
Min(η(μ+), η(μ-))
Counts/286.00 [Mev]
Max(pT(μ+), pT(μ-)) [Mev]
Max(η(μ+), η(μ-))
pulls

Figure 2

2 Bu2KJPsEE R2p1 Bp kde

2.1 L0Lexclusive

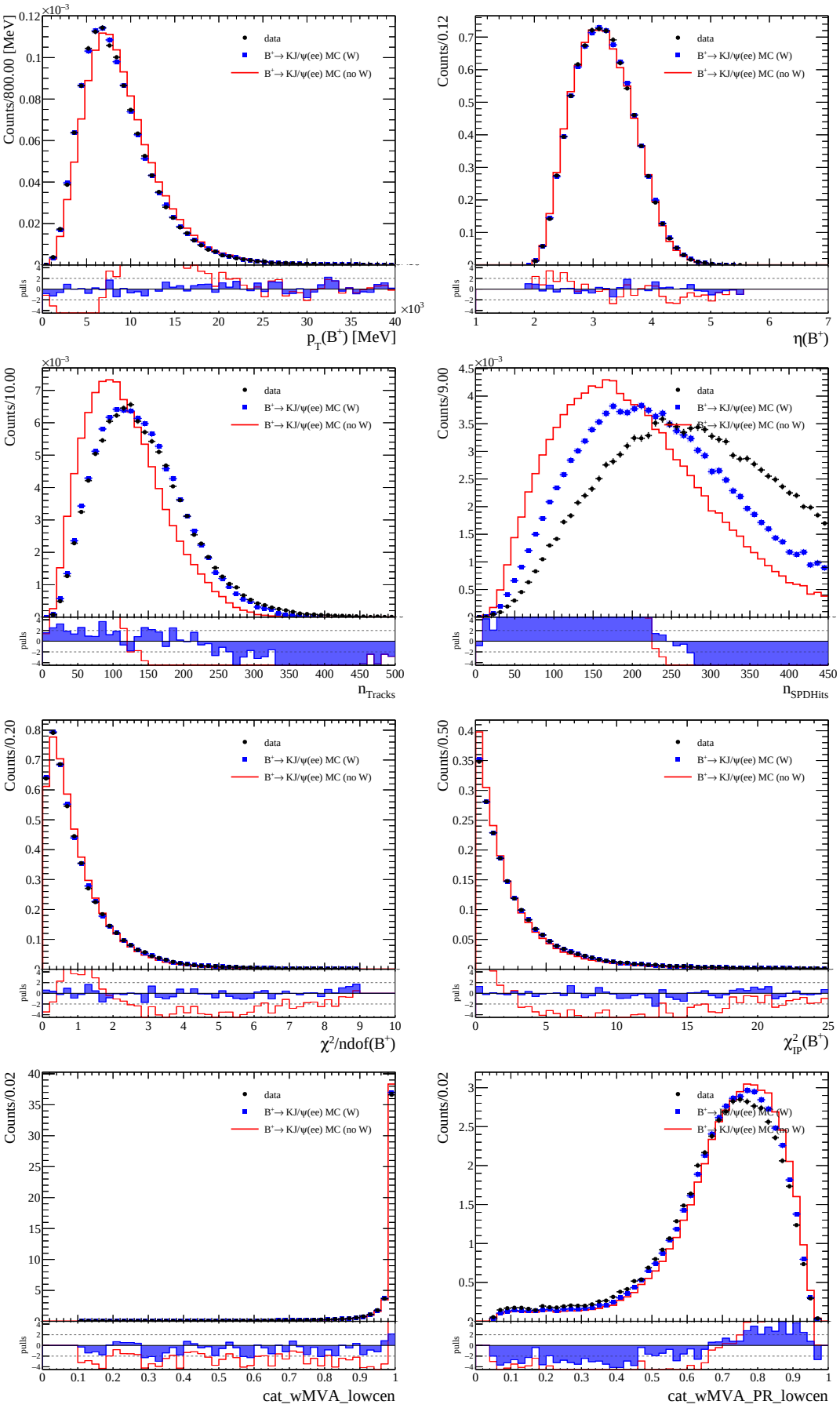


Figure 3

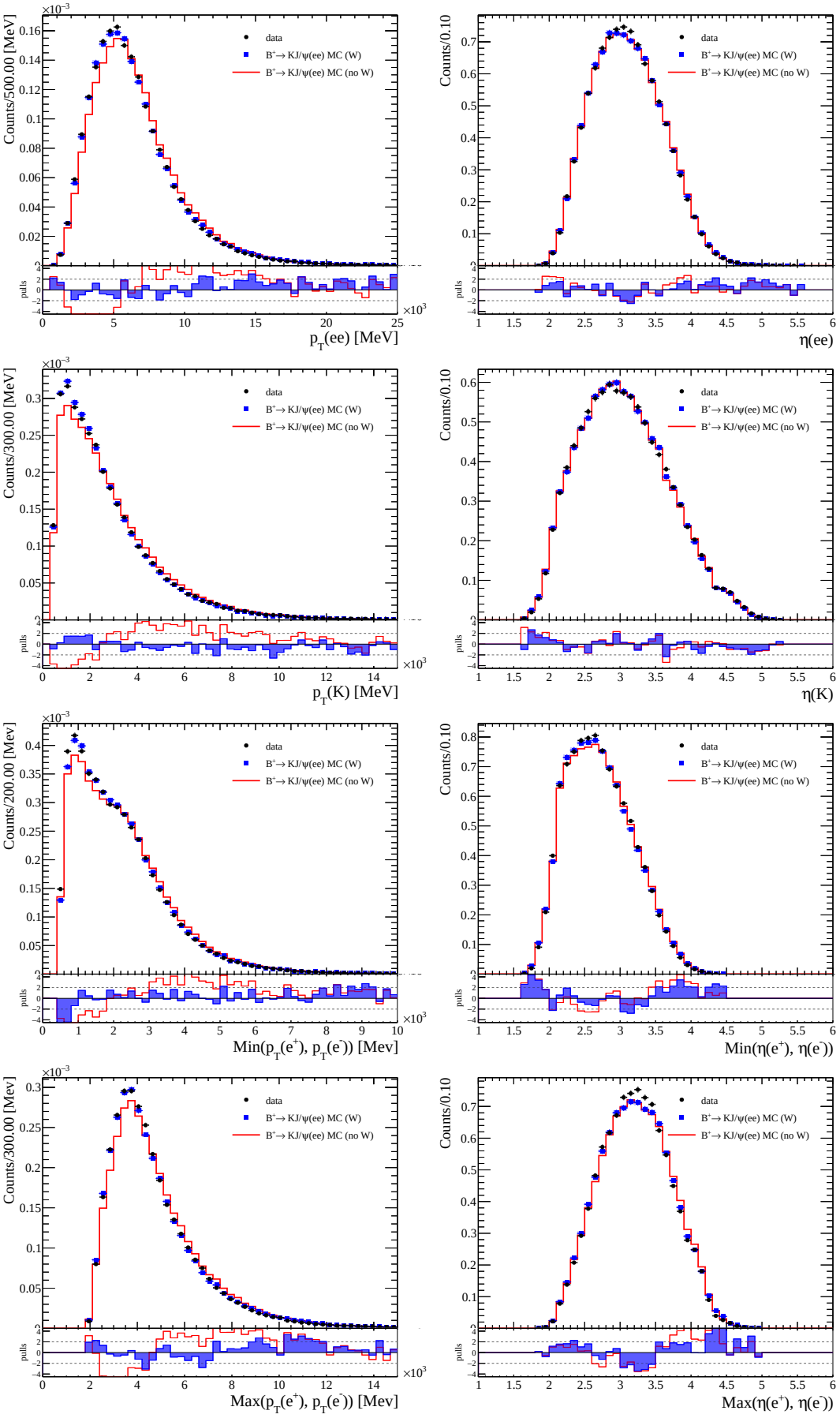

data
B+→ KJ/ψ(ee) MC (W)
B+→ KJ/ψ(ee) MC (no W)
Counts/500.00 [MeV]
pT(ee) [MeV]
Counts/0.10
η(ee)
Counts/300.00 [MeV]
pT(K) [MeV]
η(K)
Counts/200.00 [Mev]
Min(pT(e+), pT(e-)) [Mev]
Min(η(e+), η(e-))
Max(pT(e+), pT(e-)) [Mev]
Max(η(e+), η(e-))
pulls

Figure 4

3 Bd2KstJPsMM R2p1 B0 kde

3.1 L0Lexclusive

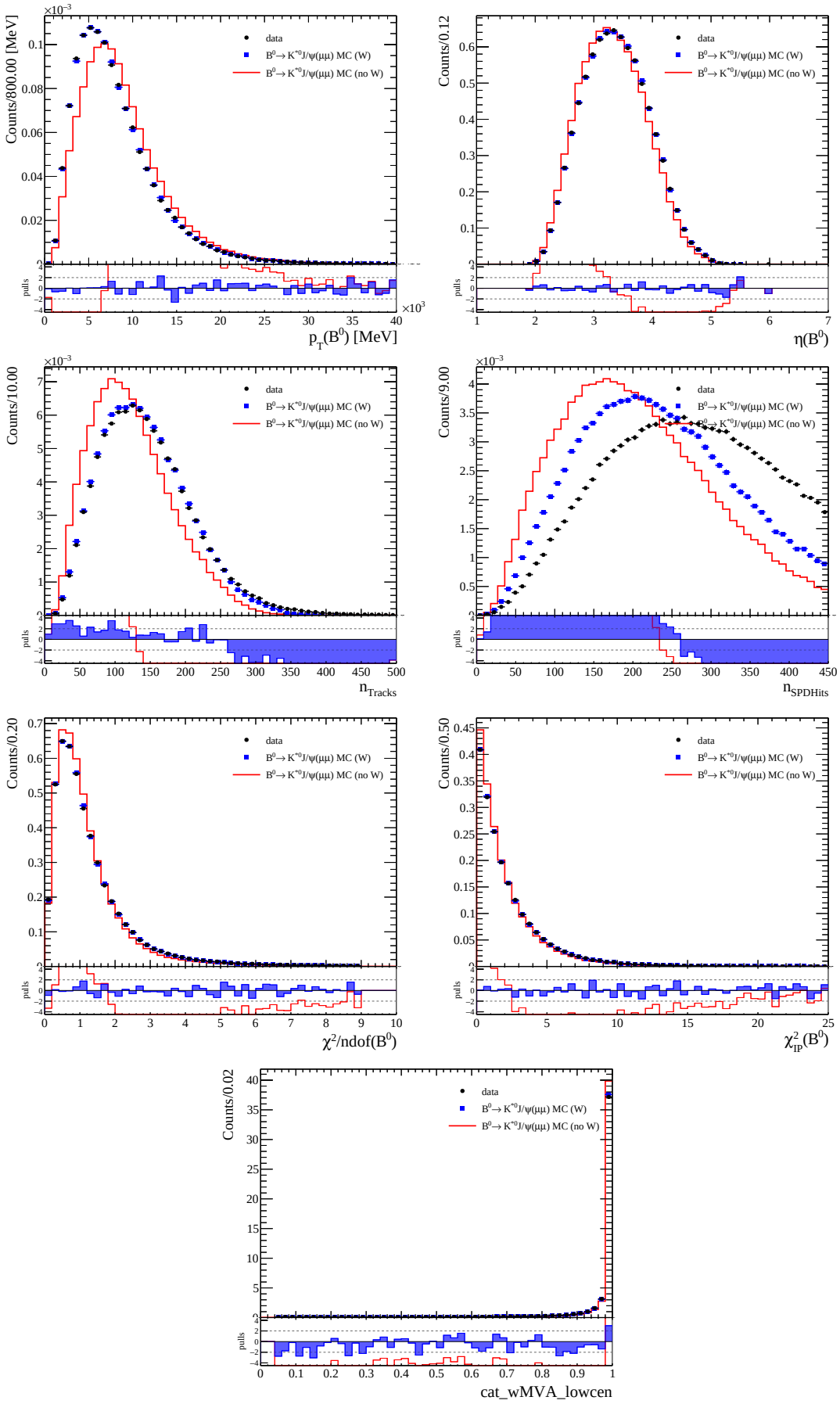


Figure 5

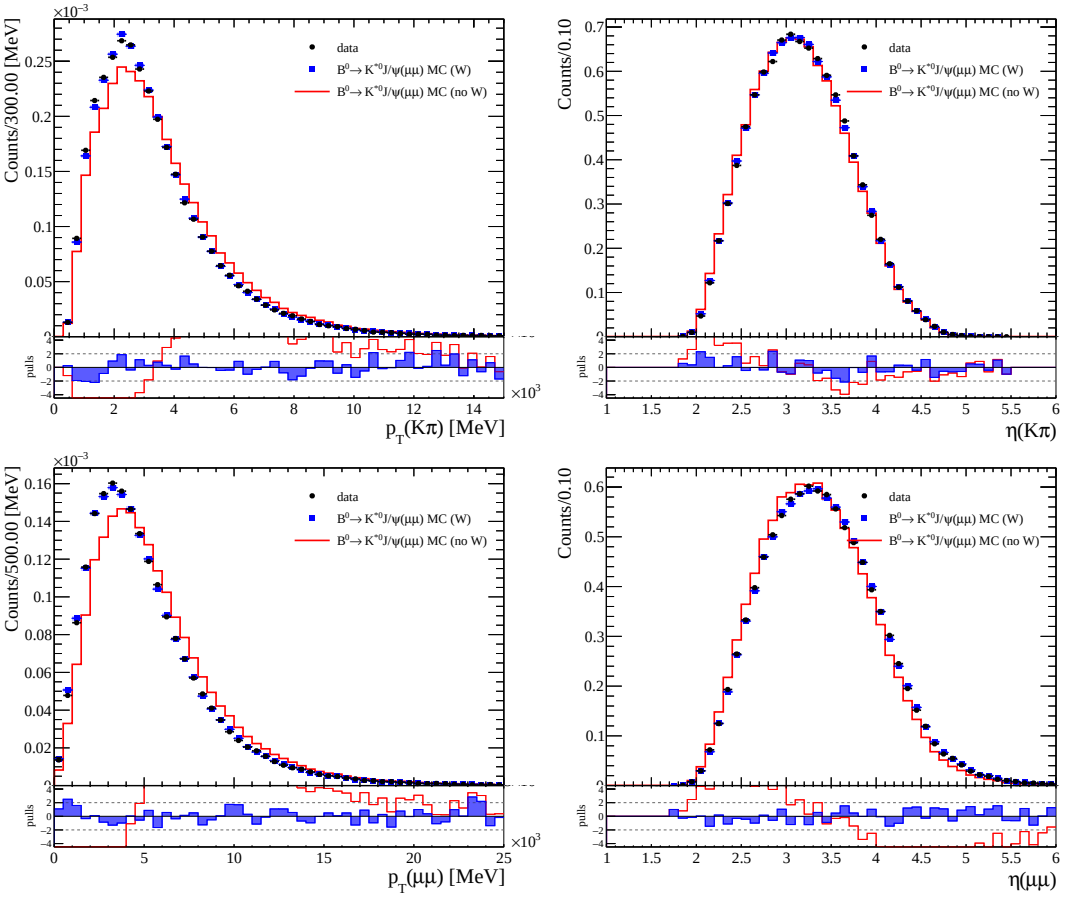

data
$B^0 \to K^{*0} J/\psi(\mu\mu)$ MC (W)
$B^0 \to K^{*0} J/\psi(\mu\mu)$ MC (no W)
Counts/300.00 [MeV]
$p_T(K\pi)$ [MeV]
Counts/0.10
$\eta(K\pi)$
Counts/500.00 [MeV]
$p_T(\mu\mu)$ [MeV]
$\eta(\mu\mu)$
pulls

Figure 6

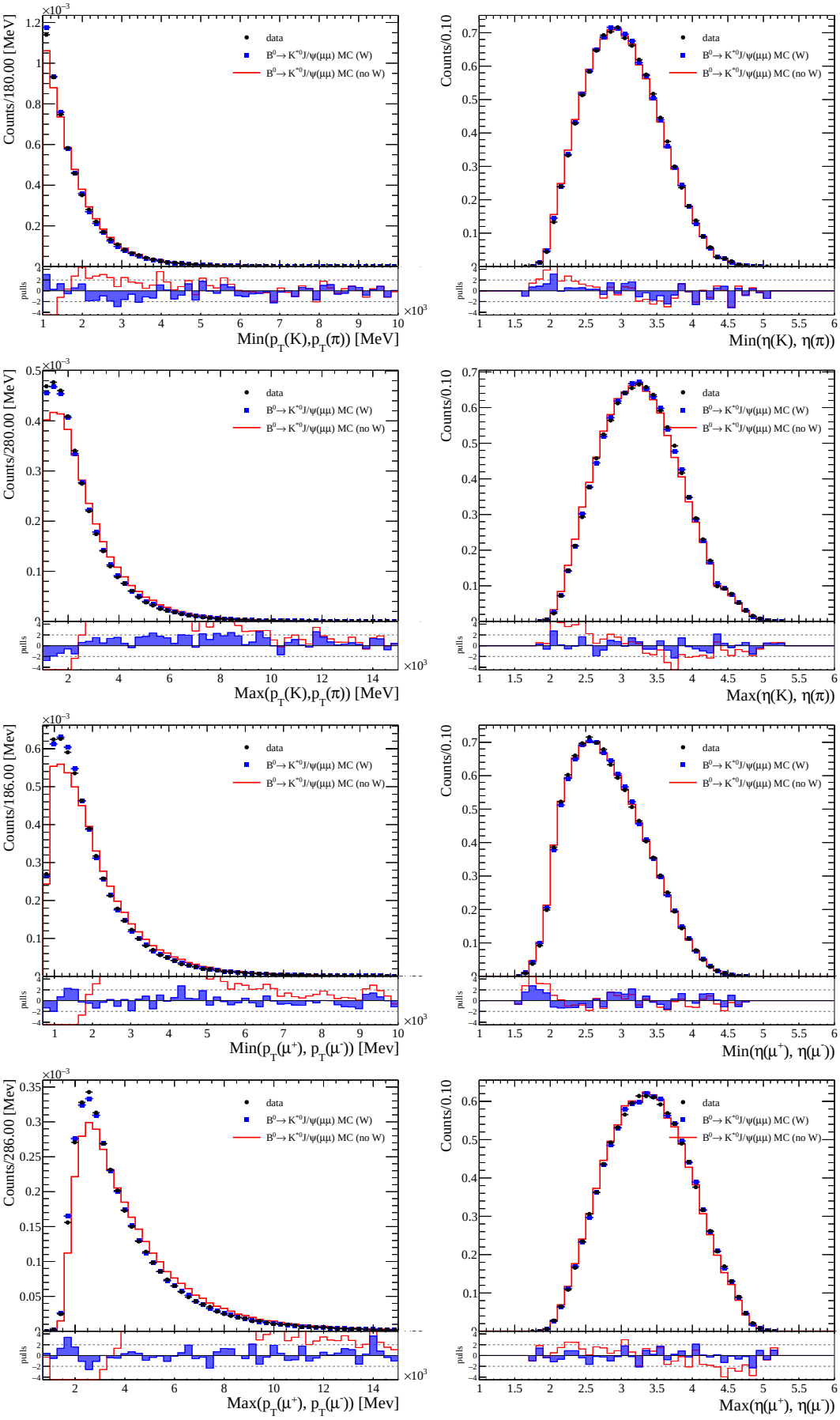

data
B⁰→ K*⁰J/ψ(μμ) MC (W)
B⁰→ K*⁰J/ψ(μμ) MC (no W)
Counts/180.00 [MeV]
Min(p_T(K),p_T(π)) [MeV]
Counts/0.10
Min(η(K), η(π))
Counts/280.00 [MeV]
Max(p_T(K),p_T(π)) [MeV]
Max(η(K), η(π))
Counts/186.00 [Mev]
Min(p_T(μ⁺), p_T(μ⁻)) [Mev]
Min(η(μ⁺), η(μ⁻))
Counts/286.00 [Mev]
Max(p_T(μ⁺), p_T(μ⁻)) [Mev]
Max(η(μ⁺), η(μ⁻))
pulls

Figure 7

4 Bd2KstJPsEE R2p1 B0 kde

4.1 L0Lexclusive

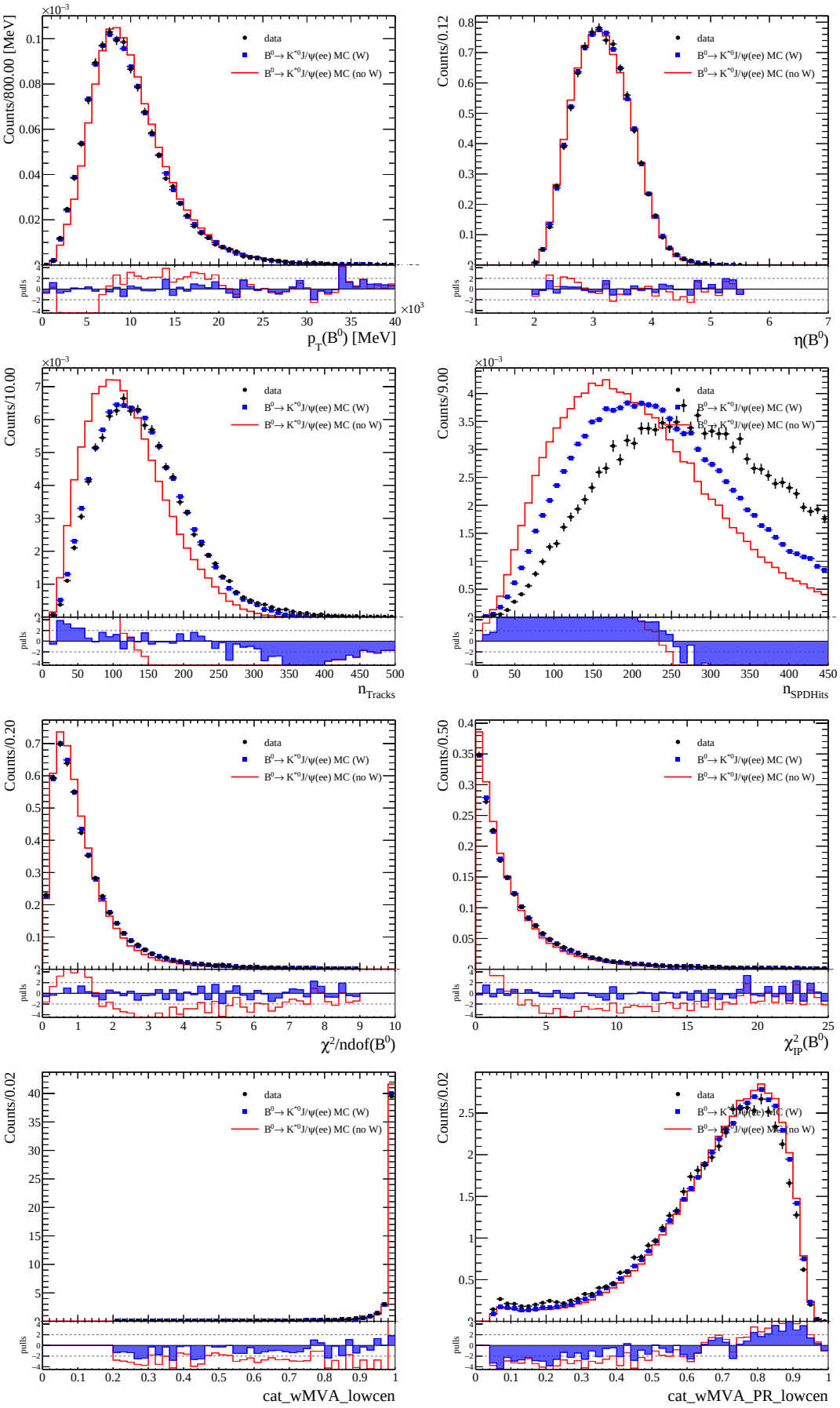


Figure 8

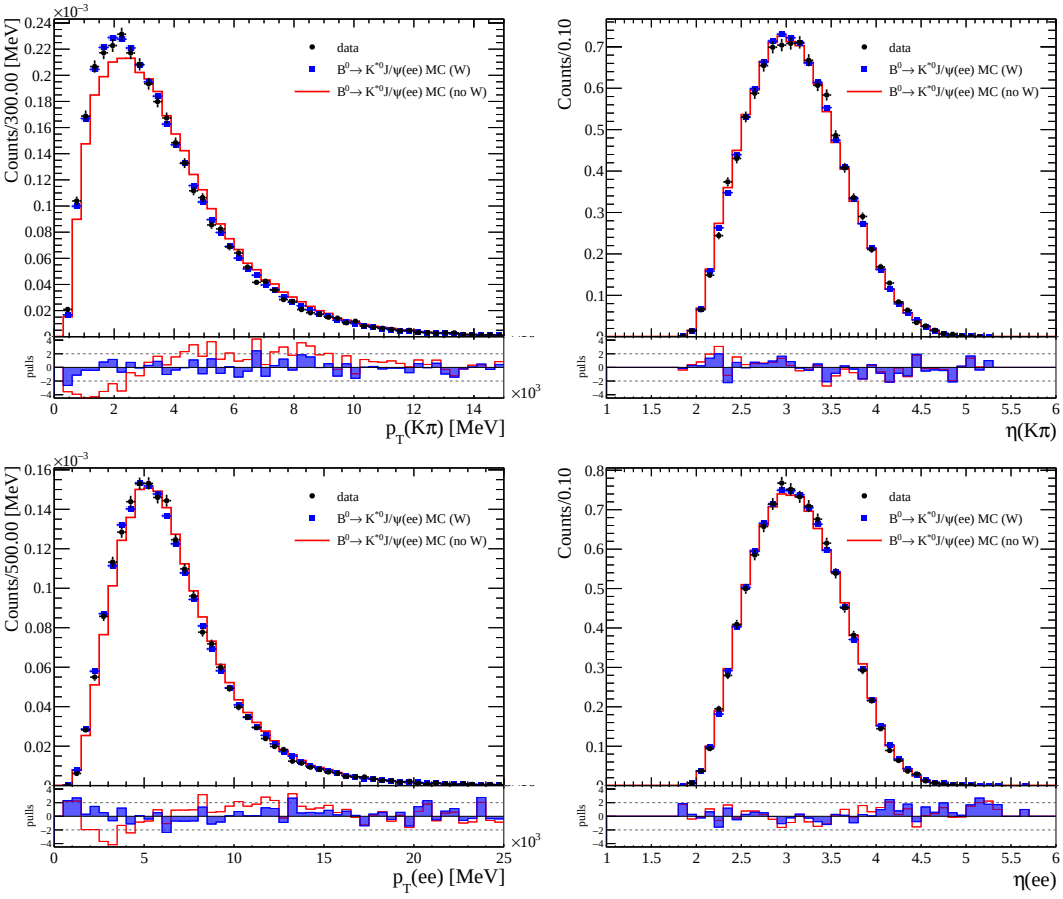

data
$B^0 \to K^{*0} J/\psi(ee)$ MC (W)
$B^0 \to K^{*0} J/\psi(ee)$ MC (no W)
Counts/300.00 [MeV]
$p_T(K\pi)$ [MeV]
Counts/0.10
$\eta(K\pi)$
Counts/500.00 [MeV]
$p_T(ee)$ [MeV]
$\eta(ee)$
pulls

Figure 9

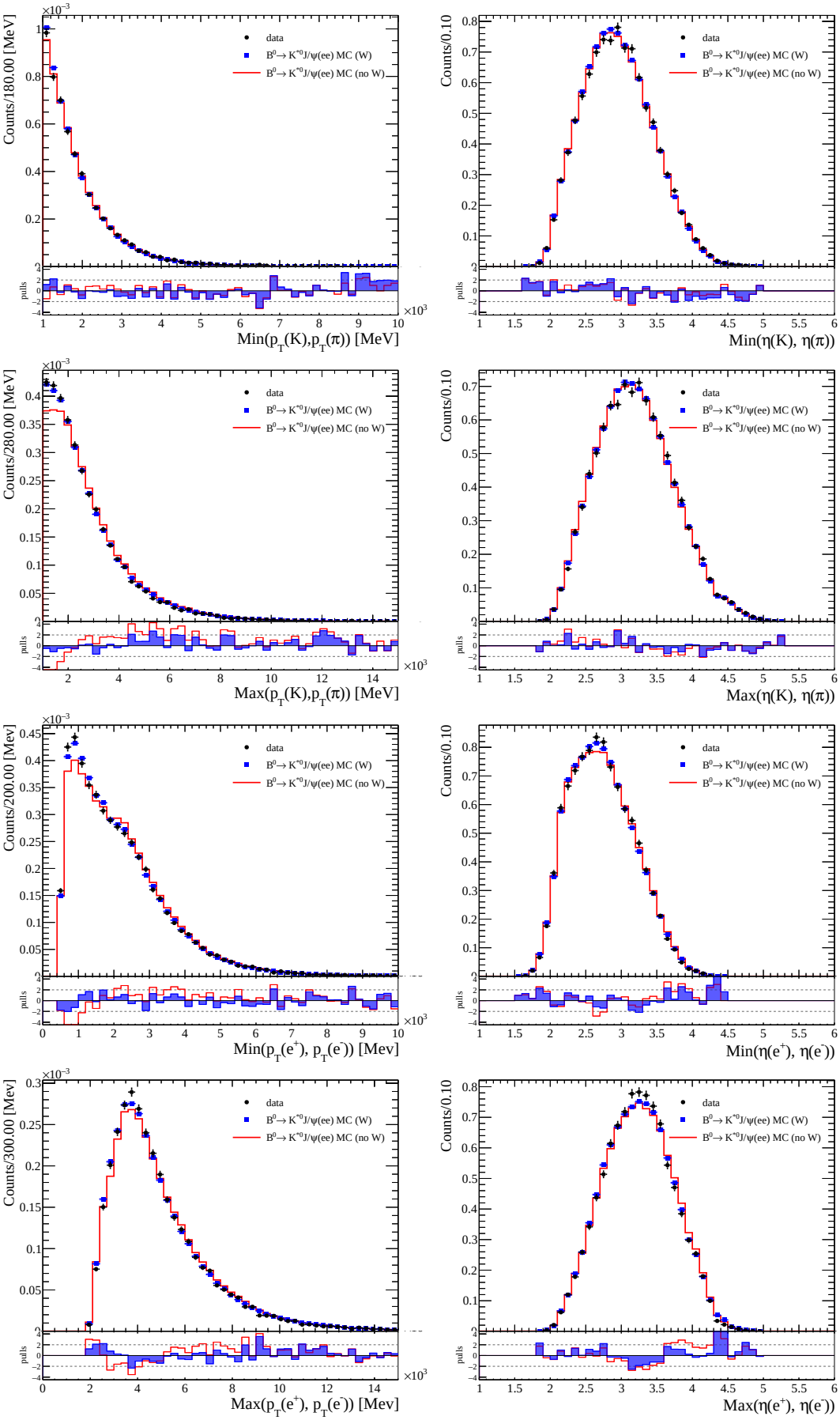

data
B⁰→ K*⁰J/ψ(ee) MC (W)
B⁰→ K*⁰J/ψ(ee) MC (no W)
Counts/180.00 [MeV]
Min(p_T(K),p_T(π)) [MeV]
Counts/0.10
Min(η(K), η(π))
Counts/280.00 [MeV]
Max(p_T(K),p_T(π)) [MeV]
Max(η(K), η(π))
Counts/200.00 [Mev]
Min(p_T(e⁺), p_T(e⁻)) [Mev]
Min(η(e⁺), η(e⁻))
Counts/300.00 [Mev]
Max(p_T(e⁺), p_T(e⁻)) [Mev]
Max(η(e⁺), η(e⁻))
pulls

Figure 10

Appendix F

Tracking

These tracking maps for electrons taken from Ref. [132].

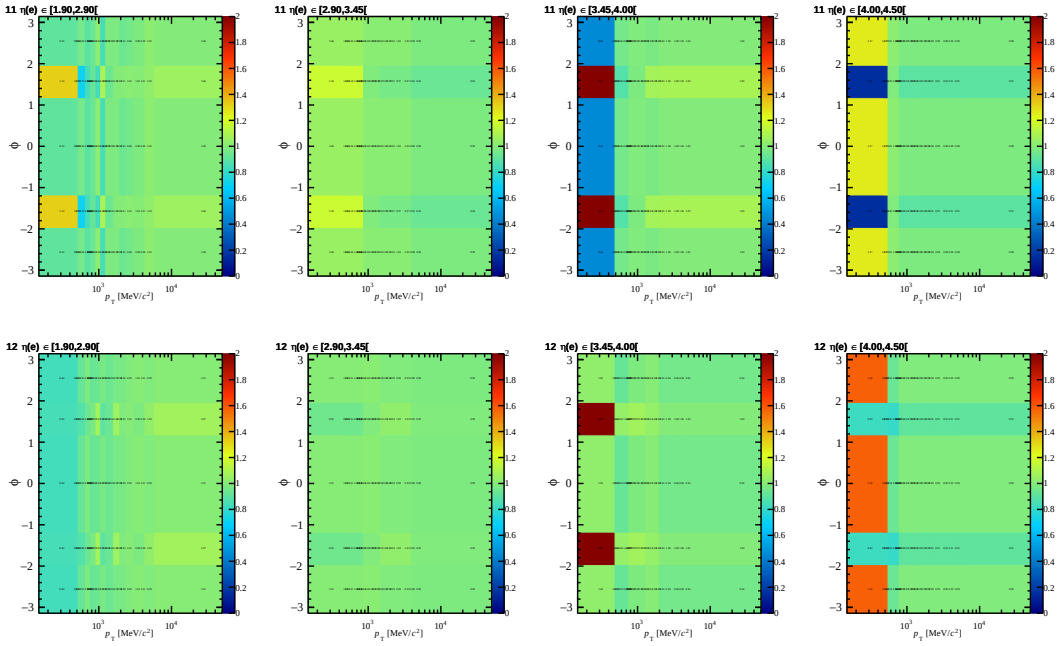


Figure F.1: Electron tracking efficiency corrections for the (top to bottom) years 2011, 2012, in the different bins of electron pseudorapidity

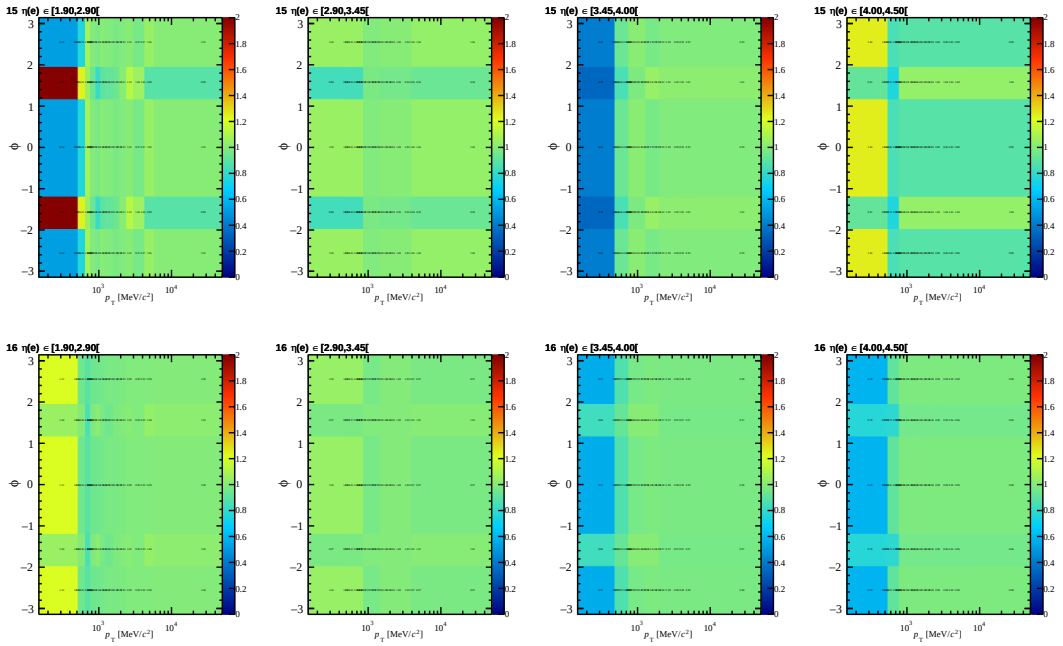


Figure F.2: Electron tracking efficiency corrections for the (top to bottom) years 2015, 2016, in bins of electron pseudorapidity

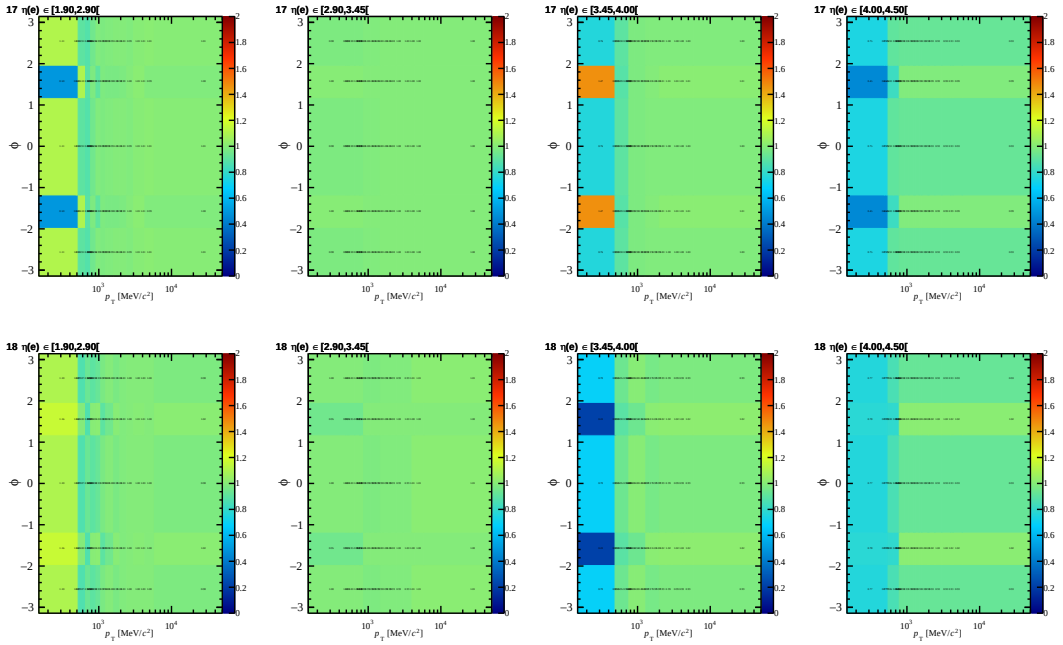


Figure F.3: Electron tracking efficiency corrections for the (top to bottom) years 2017, 2018 in (left to right) bins of electron pseudorapidity

Appendix G

PID

G.1 PIDCalib KDE weight maps

B^+ KDE weight maps

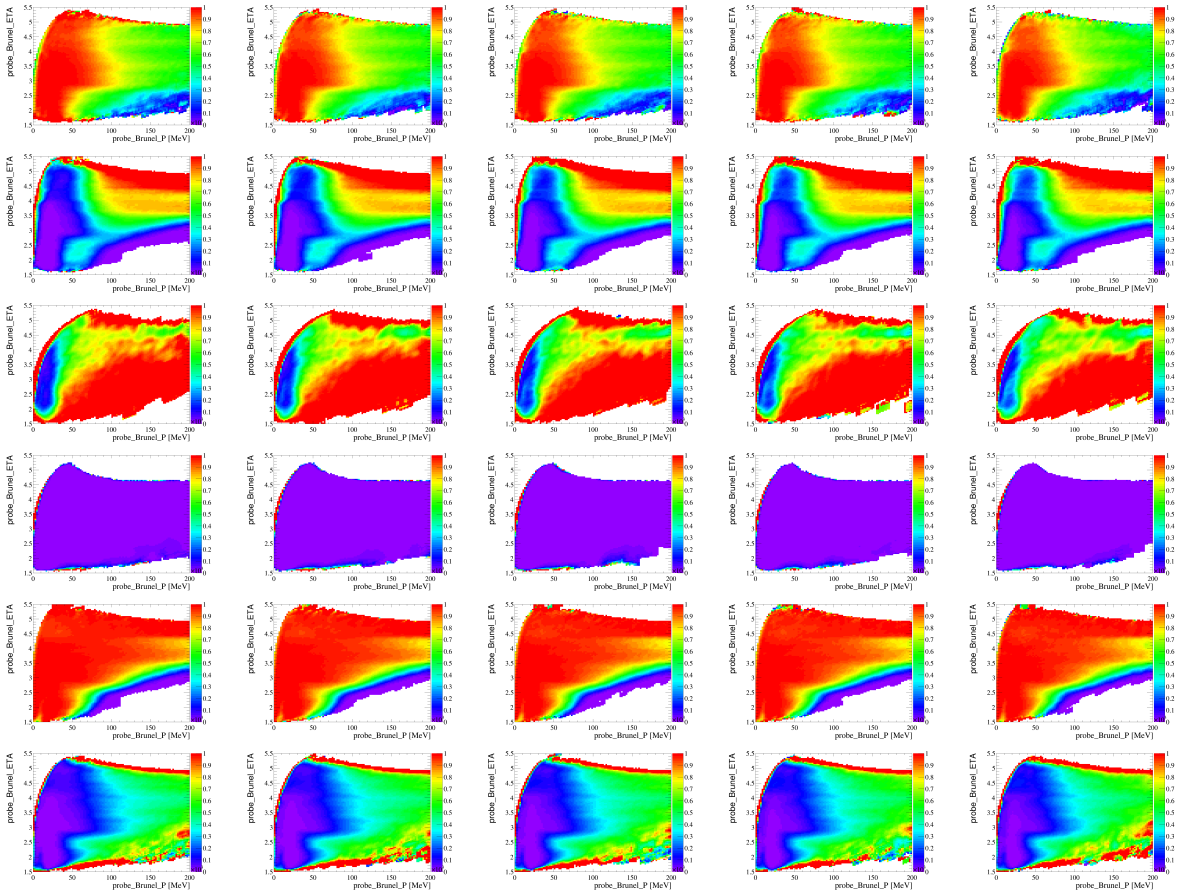


Figure 1: PID weights maps for (top to bottom) kaon ID, kaon → pion misID, kaon → muon misID, kaon → electron misID, pion ID and pion → kaon misID in 2016 for B^+ modes in (left to right) bins of ntracks.

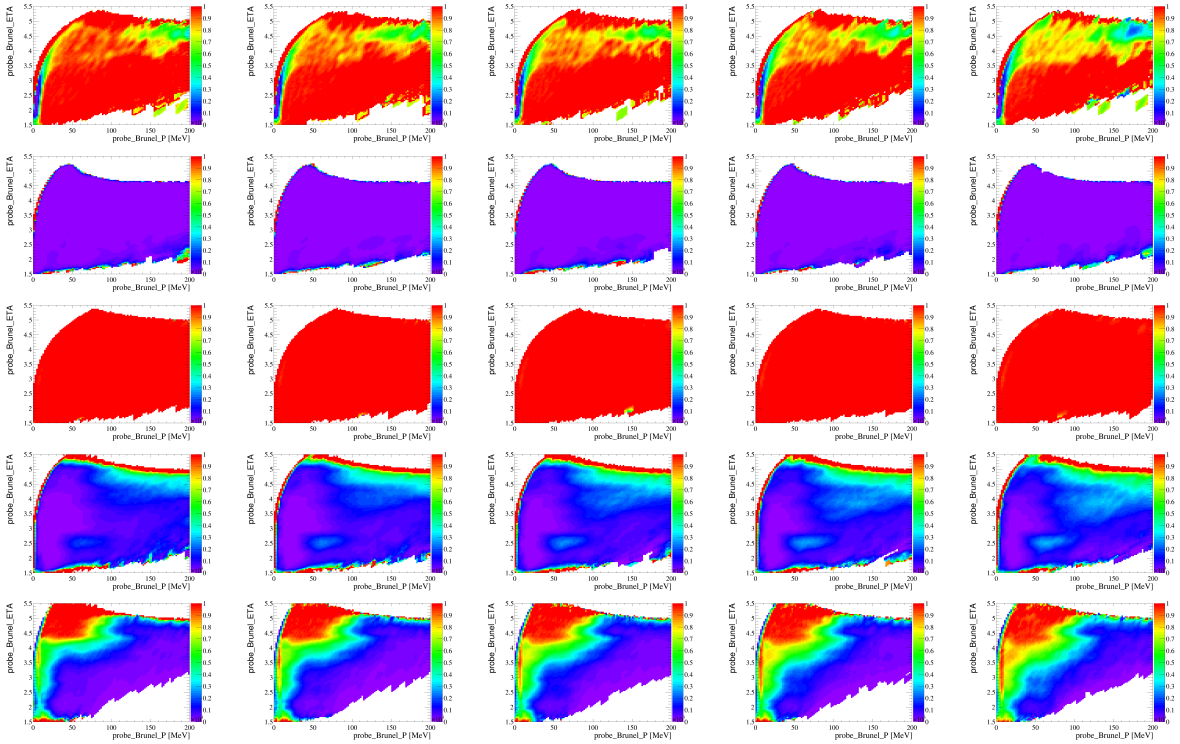


Figure 2: PID weights maps for (top to bottom) pion → muon misID, pion → electron isID, muon ID, muon → kaon misID and muon → pion misID in 2016 for B^+ modes in (left to right) bins of ntracks.

G.2 Electron fit and count

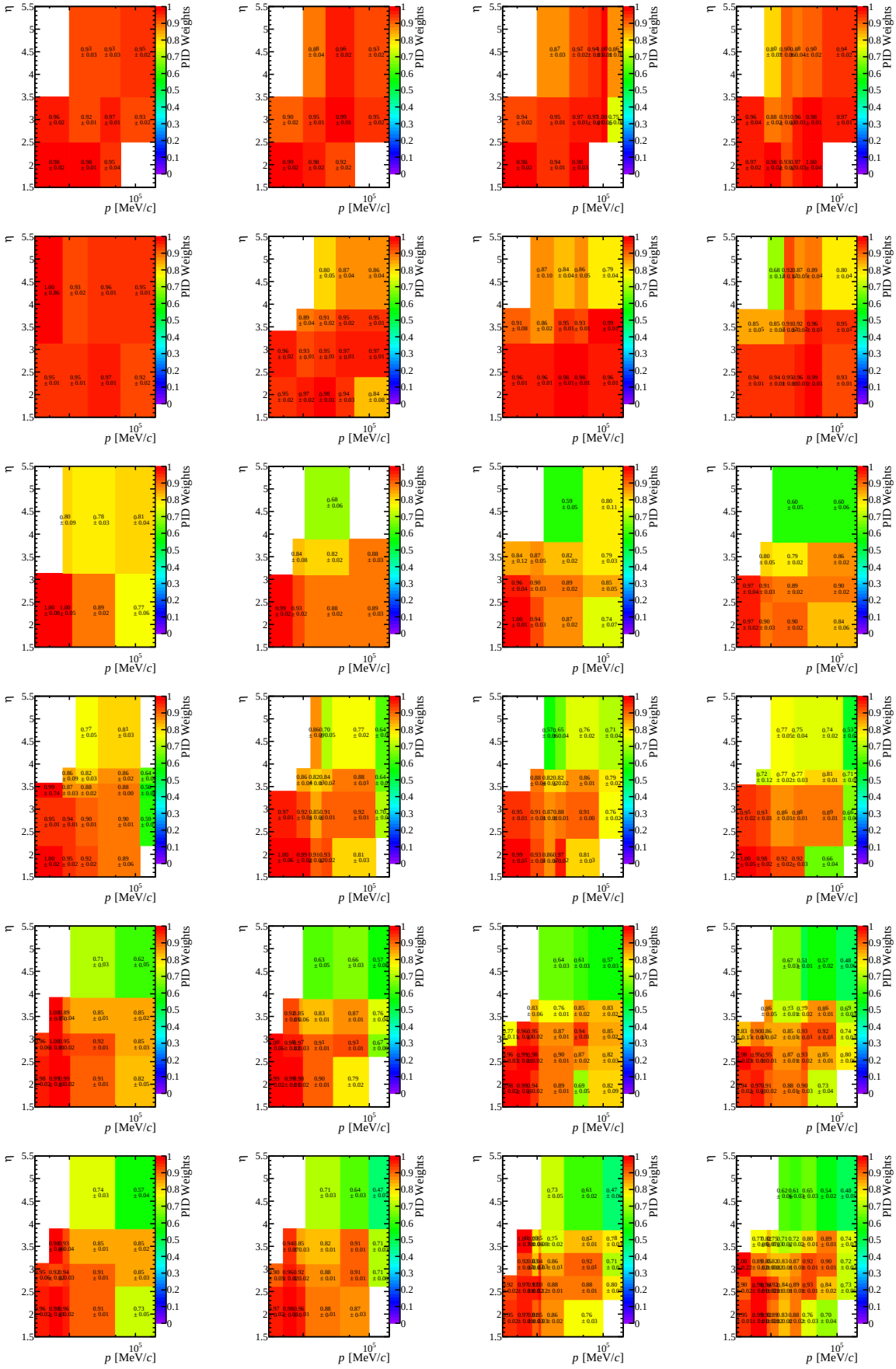


Figure 1: PID ID weight maps for electrons for R_K and R_{K^*} of the years 2011-2018 (top to bottom) for Bremsstrahlung category Brem 0

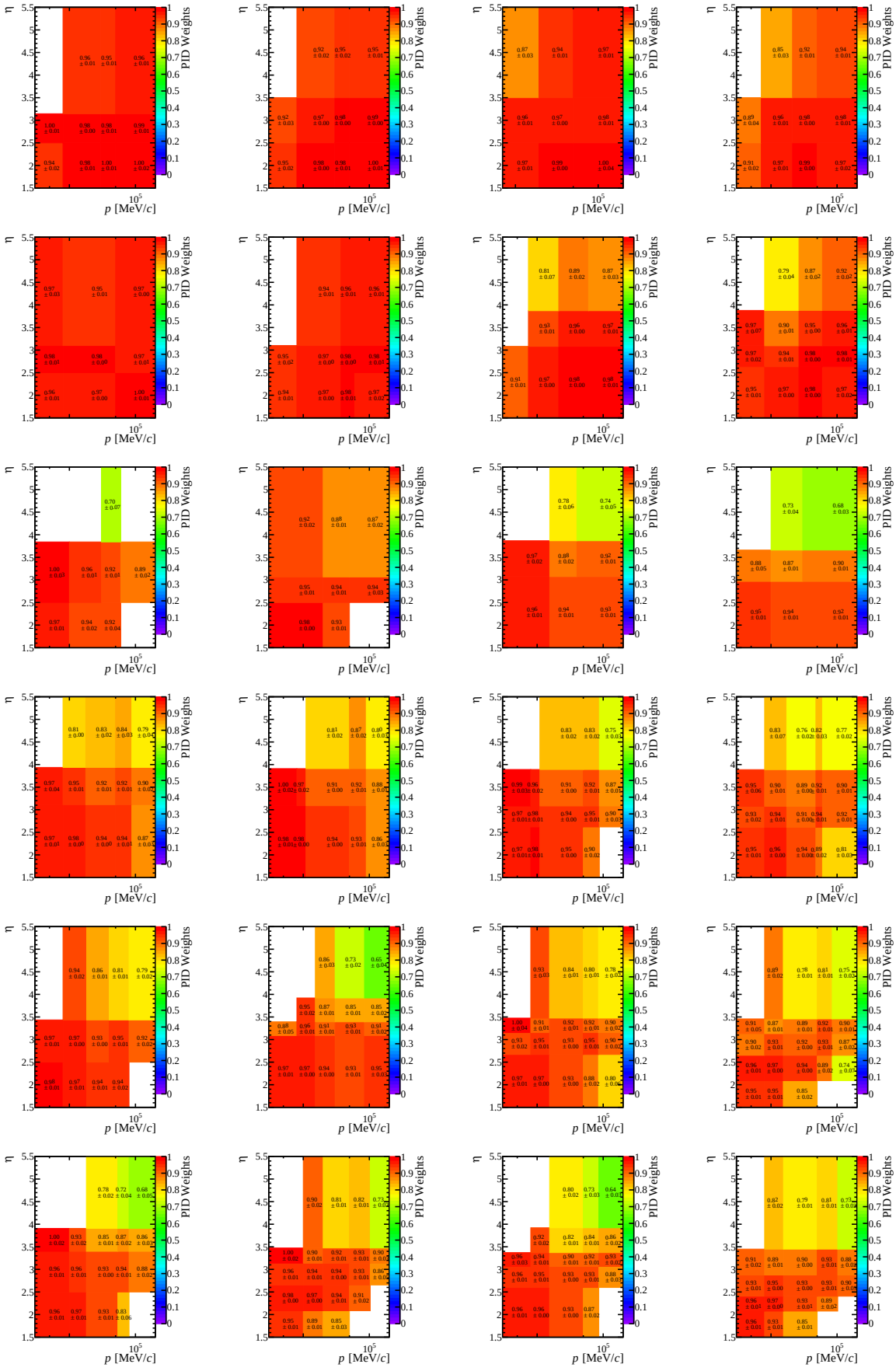


Figure 2: PID ID weight maps for electrons for R_K and R_{K^*} of the years 2011-2018 (top to bottom) for Bremsstrahlung category Brem 1 in bins of n_{Tracks} (left to right)

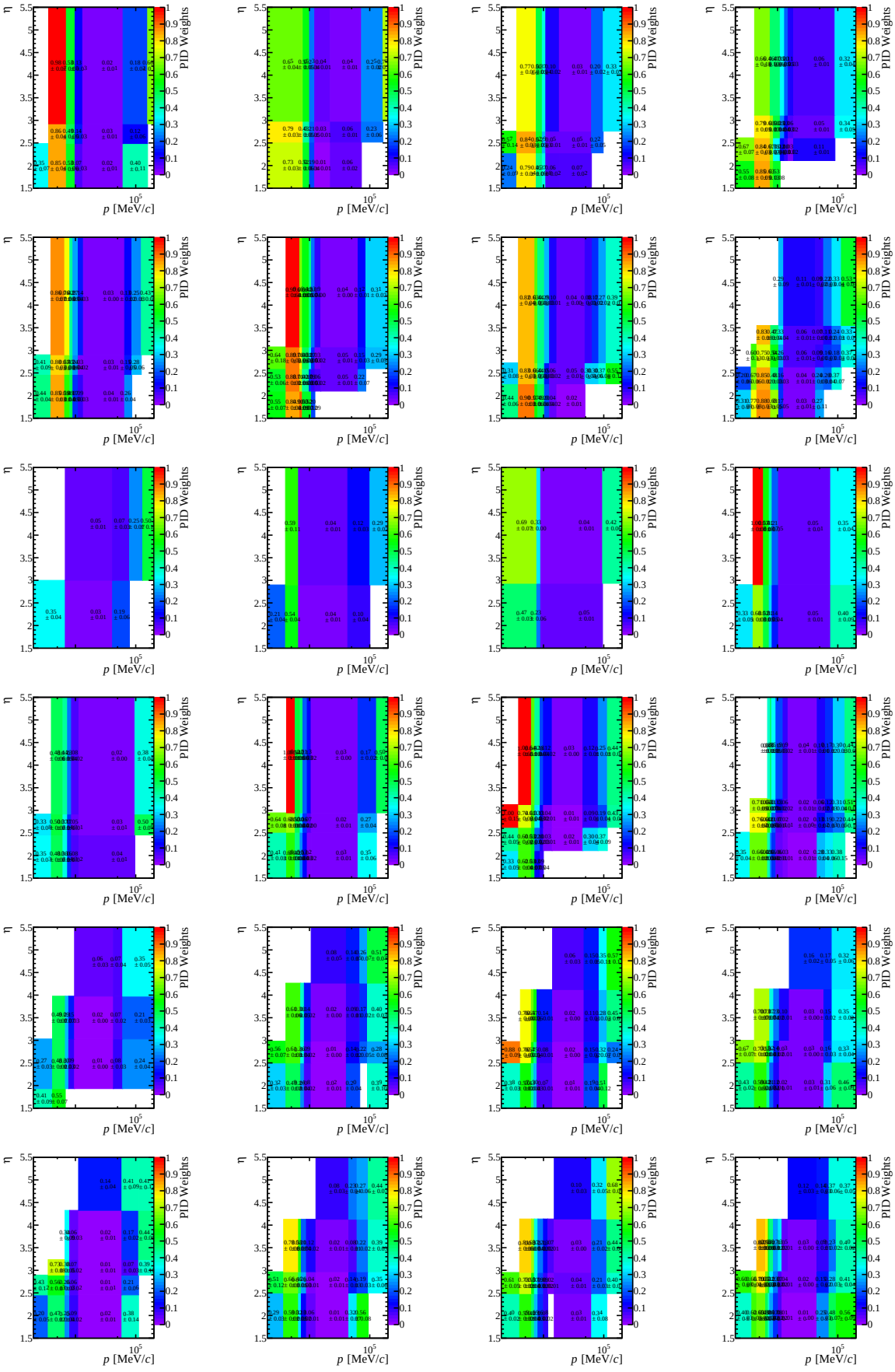


Figure 3: PID misID K weight maps for electrons for R_K of the years 2011-2018 (top to bottom) for Bremsstrahlung category Brem 0 in bins of n_{Tracks} (left to right)

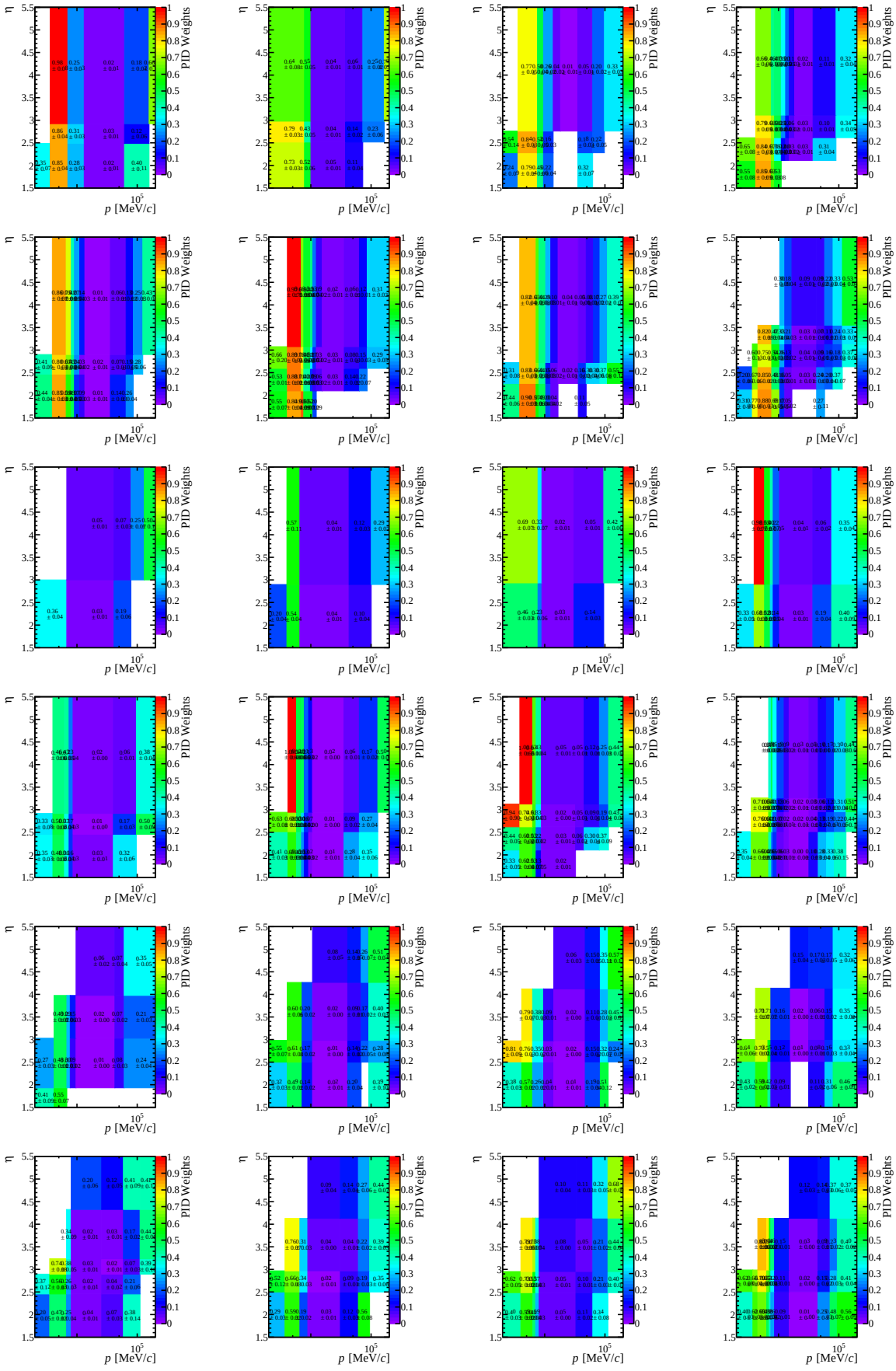


Figure 4: PID misID K weight maps for electrons for R_{K^*} of the years 2011-2018 (top to bottom) for Bremsstrahlung category Brem 0 in bins of n_{Tracks} (left to right)

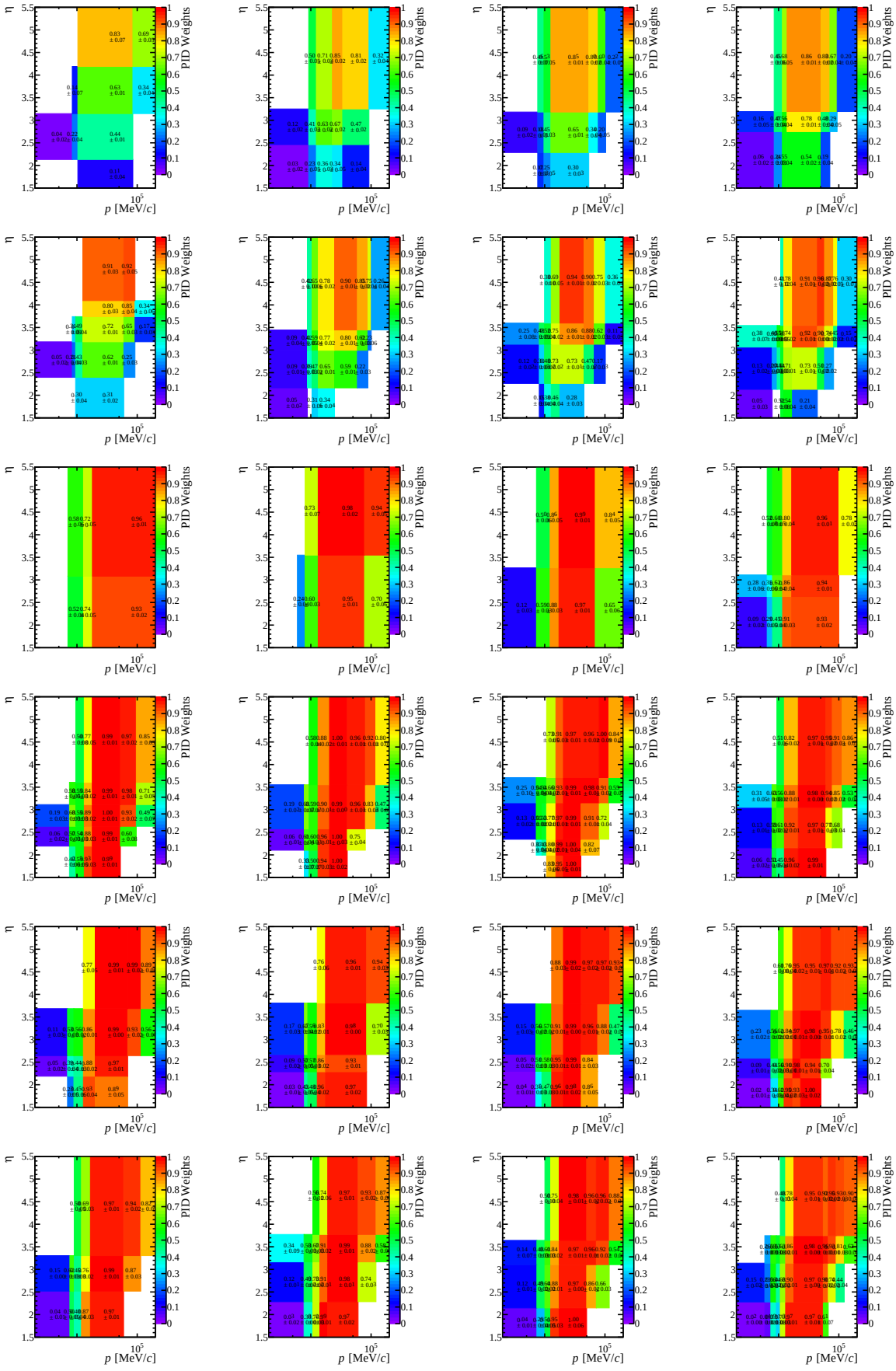


Figure 5: PID misID π weight maps for electrons for R_K and R_{K^*} of the years 2011-2018 (top to bottom) for Bremsstrahlung category Brem 1 in bins of n_{Tracks} (left to right)

Appendix H

L0 trigger

L0I ratio in `nTracks`

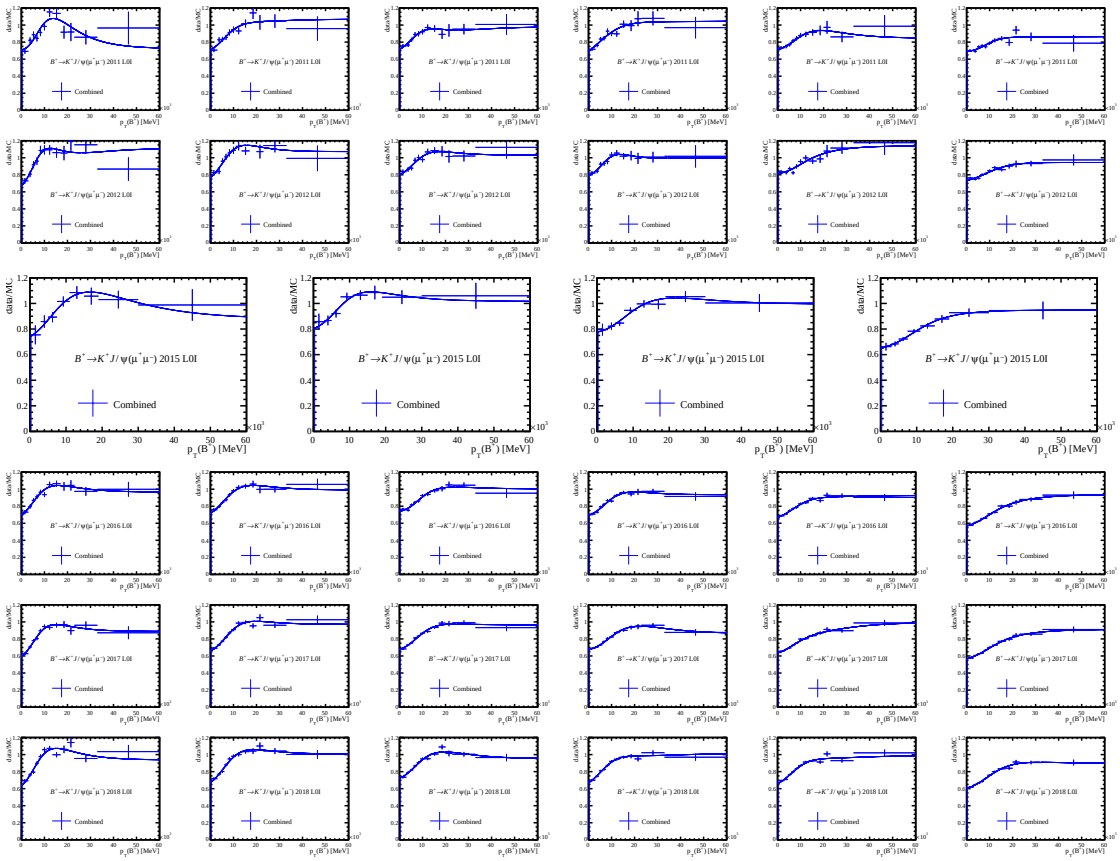


Figure H.1: Data over simulation efficiency ratio of the `L0Global_TIS` line as a function of the transverse momentum of the B meson in 6 regions (4 in 2015) of `nTracks` from left to right for $B^+ \to K^+ J/\psi\,(\to \mu^+\mu^-)$ events for all years from top to bottom. The efficiencies are measured with the **Combined** tag.

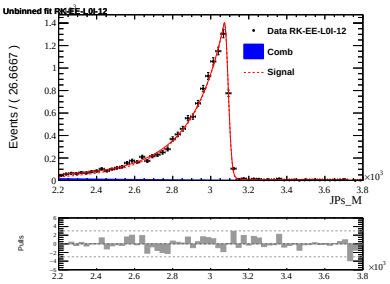
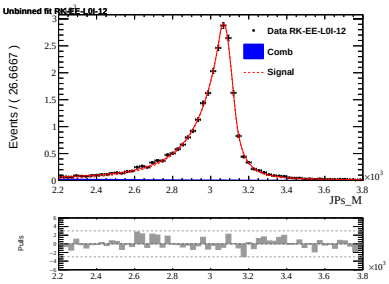
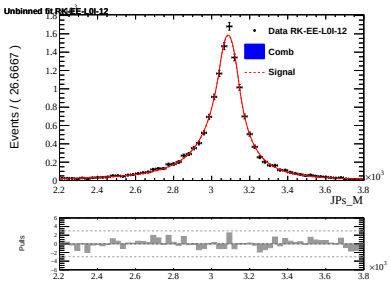
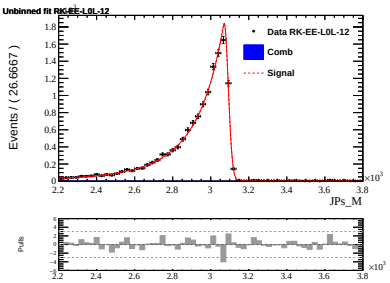
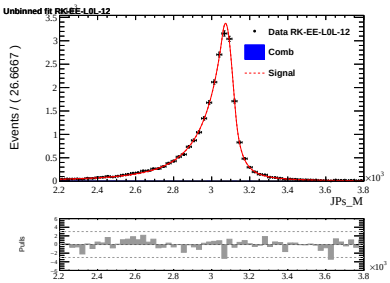
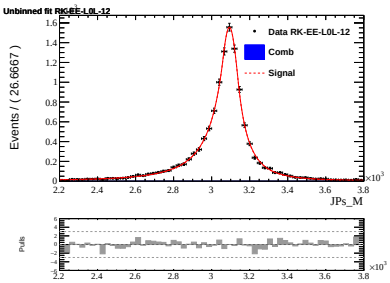
Appendix I

q^2 smearing

$B^+ \to K^+ J/\psi(ee)$

2012

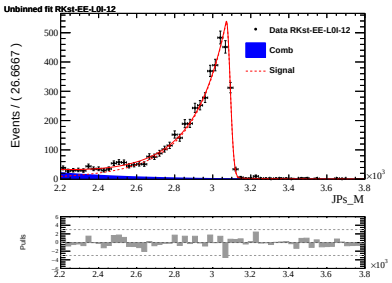
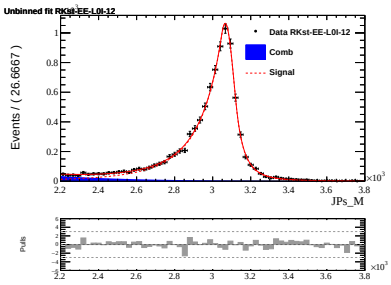
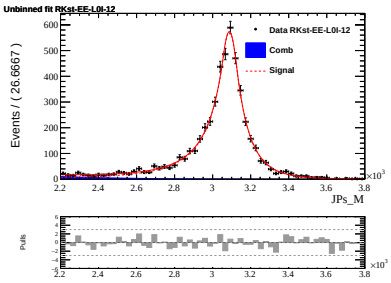
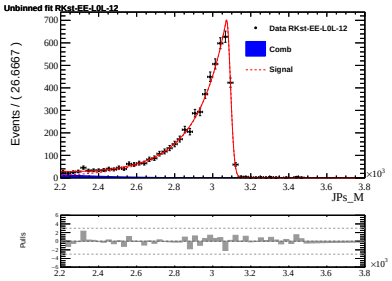
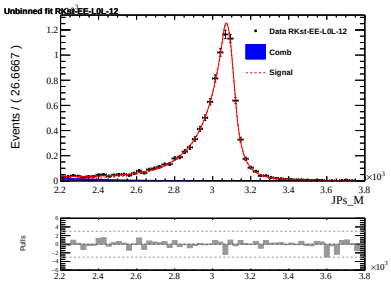
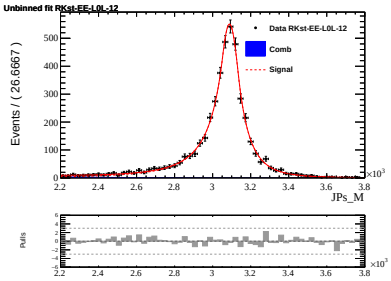
Table 1: A table showing the fits to $m(ee)$ distributions of $B^+ \to K^+ J/\psi(ee)$ in 2012 data. From left to right 0 bremsstrahlung photons added, 1 bremsstrahlung photon added and > 2 bremsstrahlung photons added. From top to bottom are the L0I-inclusive and L0L-exclusive trigger categories.

0γ	1γ	2γ
L0I-exclusive 	L0I-exclusive 	L0I-exclusive 
L0L-exclusive 	L0L-exclusive 	L0L-exclusive 

$B^0 \to K^{*0} J/\psi(ee)$

2012

Table 3: A table showing the fits to $m(ee)$ distributions of $B^0 \to K^{*0} J/\psi(ee)$ in 2012 data. From left to right 0 bremsstrahlung photons added, 1 bremsstrahlung photon added and > 2 bremsstrahlung photons added. From top to bottom are the L0I-inclusive and L0L-exclusive trigger categories.

0γ	1γ	2γ
L0I-exclusive 	L0I-exclusive 	L0I-exclusive 
L0L-exclusive 	L0L-exclusive 	L0L-exclusive 

Appendix J

Combination of inclusive backgrounds

Partially reconstructed background PDFs for J/ψ modes are formed from an admixture of different decay modes, namely simulated inclusive decays of B^0, B^+ and B_s^0 mesons. These simulated samples require the leptons to have been produced by an intermediate J/ψ or $\psi(2S)$ resonance, however, since the constructed PDFs will be used for J/ψ modes it is required that the leptons are produced via a J/ψ resonance only.

The PDFs determined from each of these simulated samples are combined in a linear fashion. Normalisation factors, A_x can be determined using the selection efficiency of the particular inclusive sample corrected by the hadronisation fractions, f_x, and the sum of the branching fractions of the inclusive decay channels, $\sum \mathcal{B}(B_x)$, considered as seen in Eq. J.1.

$$A_x = \frac{N_{pass}(selection)}{N_{simulated}} \times f_x \times \sum \mathcal{B}(B_x) \tag{J.1}$$

The total branching fraction can be determined by listing all of the decay channels that may be included in the particular inclusive sample and summing their individual branching fractions. As this is easily prone to errors the branching fraction of the dominant decay mode within the simulation is divided by the inclusive branching fraction that has been measured elsewhere. For example $B^0 \to K^{*0} J/\psi (\to \mu^+\mu^-)$ has a fraction of 0.1850 in the $B^0 \to X J/\psi (\to \mu^+\mu^-)$ simulated sample and a branching fraction of 1.270×10^3 in Ref. [14]. The branching fraction correction is thus $\frac{1.270\times10^3}{0.1850} = 6.860 \times 10^{-3}$.

Appendix K

S-wave

K.1 Partially reconstructed backgrounds in non-resonant Kee final state

$B^0 \to K^{*0}e^+e^-$ The relative efficiencies of $B^0 \to K^{*0}e^+e^-$ and partially reconstructed $B^0 \to K^{*0}e^+e^-$ which is reconstructed as $B^+ \to K^+e^+e^-$ are computed using simulated samples for each year, trigger and magnet polarity. An extrapolation factor is then required for isospin conjugate background of $B^+ \to K^{*+}e^+e^-$, the calculation of which is split into two parts. First, the relative branching fraction of $B^0 \to K^{*0}e^+e^-$ [140] and $B^+ \to K^{*+}e^+e^-$ [141] is assumed to be 1.0776 according to isospin symmetry. All current measurements support this assumption. This is not exactly unity due to the difference in B^+/B^0 lifetimes. Second, the branching fraction of K^{*0} to $K^+\pi^-$, which LHCb can easily detect, is 2/3 whilst the branching fraction of K^{*+} to $K^+\pi^0$ is 1/3. Using both of the parts the extrapolation factor, $f_{isospin}$, is found to 1.54 from the $B^0 \to K^{*0}e^+e^-$ signal yield and the partially reconstructed yield in the R_K signal region. Note that since the $K^{*+} \to K^+\pi^0$ branching fraction dominates out of the two parts, even if a relative 10% uncertainty is applied to the first part only around a relative uncertainty of 3.5% would be present on the combined extrapolation factor. As a further note when selecting partially reconstructed backgrounds with the $B^+ \to K^+e^+e^-$ background hypothesis there is no 100 MeV mass window applied around the K^{*0} mass to protect against S-wave $B^0 \to K^+\pi^-e^+e^-$ decays. This mass window is present when selecting signal decays. Hence the above calculation does not account for

partially reconstructed $B^0 \to K^+\pi^-e^+e^-$ decays outside the K^{*0} window.

$B^0 \to K^+\pi^-e^+e^-$ Additional contributions from S-wave decays are computed using phase space $B^0 \to K^+\pi^-e^+e^-$ simulation that pass the full $B^+ \to K^+e^+e^-$ selection chain. This allows for an absolute efficiency for these type of events to be calculated. This can then be combined with the relevant b production cross-sections(560 μb at 13 TeV and 295 μb at 7 TeV) in Ref. [142] and the S-wave branching fractions in Ref. [140]. To calculate the expected branching fractions per unit luminosity, year and trigger category. When using the S-wave branching fraction the calculation must proceed in steps. Firstly it can be noted that the S-wave extrapolation in the in a 100 MeV mass window around the K^{*0} mass was already taken into account in the $B^0 \to K^{*0}\ell^+\ell^-$ extrapolation. This is because this mass window is applied in the $B^0 \to K^{*0}\ell^+\ell^-$ collision data fit. Since there is no cut on the K^{*0} helicity it is safe to assume that the resonant and non-resonant and S-wave events in the same K^{*0} mass region is the same. Therefore when computing the S-wave fraction below 1200 MeV we must exclude this window, doing this the S-wave branching fraction is 1.4×10^{-8}. Then the efficiency for events in this region to be selected as partially reconstructed backgrounds to $B^+ \to K^+e^+e^-$ is computed. In this part of the calculation the trigger categories are combined, the yield will be split between them at the end. The calculation gives 1.3 and 2.5 events per unit luminosity respectively at 7 TeV and 13 TeV respectively. This then needs to be scaled to events above 1200 MeV to a value of 2400 MeV. The extrapolation does not need to go higher since due to the lower end of the B^+ mass window being 4600 MeV there is no efficiency to select S-wave events with a $m_{K\pi}$ above 2400 MeV. Using this and S-wave contribution above 1200 MeV can be calculated assuming that the S-wave branching fraction can be scaled linearly to 2400 MeV (broadly justified by Fig. 7.6 of LHCb-ANA-2013-096) and calculating efficiencies from the phase space simulation. This results in 2.7 and 5.2 events per unit luminosity at 7 TeV and 13 TeV respectively for S-wave events above 1200 MeV.

K.2 Normalisation Constraints

The above calculation helps to constrain the amount of background for $B^0 \to K^{*0}e^+e^-$ and $B^0 \to K^+\pi^-e^+e^-$ decays which when partially reconstructed can form a background for $B^+ \to K^+e^+e^-$ decays.

- Partially reconstructed $B^0 \to K^{*0} e^+ e^-$: This background can be parametrised by the yield of $B^0 \to K^{*0} e^+ e^-$ observed in data, an isospin extrapolation factor determined above, and the ratio of the efficiency to select $B^0 \to K^{*0} e^+ e^-$ as a signal and the efficiency to select partially reconstructed $B^0 \to K^{*0} e^+ e^-$ as seen in Eq. K.1.

$$\mathcal{N}^{PartReco}_{B^0 \to K^{*0} e^+ e^-}(q^2) = \frac{\varepsilon^{PartReco}_{B^0 \to K^{*0} e^+ e^-}(q^2)}{\varepsilon^{Signal}_{B^0 \to K^{*0} e^+ e^-}(q^2)} \cdot f_{isospin} \cdot \mathcal{N}^{Signal}_{B^0 \to K^{*0} e^+ e^-}(q^2) \tag{K.1}$$

- Partially reconstructed $B^0 \to K^+ \pi^- e^+ e^-$: The constraint above in Eq. K.1 does not account for $B^0 \to K^+ \pi^- e^+ e^-$ decays outside the K^{*0} mass window. The yield of this background outside this mass window is fixed as discussed above.

K.3 Systematics

Uncertainty on partially reconstructed background in $B^+ \to K^+ e^+ e^-$ mass window - The two partially reconstructed backgrounds in the $B^+ \to K^+ e^+ e^-$ mass window, $B^0 \to K^{*0} e^+ e^-$ and $B^0 \to K^+ \pi^- e^+ e^-$, rely on extrapolating branching fractions to determine their yields as detailed above. In order to evaluate a systematic for this 1000 pseudoexperiments would be generated and an alternative fit model where the extrapolation factors are varied within 1σ of their associated uncertainty would be used to evaluate toy-by-toy differences and assign a systematic. The effect needs to be checked on the signal yield of both $B^+ \to K^+ e^+ e^-$ and $B^0 \to K^{*0} e^+ e^-$, since the signal yield of $B^0 \to K^{*0} e^+ e^-$ is used as a constraint when determining the extrapolation factor. For those backgrounds with $m_{K\pi}$ inside the K^{*0} mass window the extrapolation factor is largely reliant on the relative branching fraction between $B^0 \to K^{*0} e^+ e^-$ and $B^+ \to K^{*+} e^+ e^-$, assigning a conservative uncertainty of 10% on the $B^+ \to K^{*+} e^+ e^-$ branching fraction yields a 3.5% uncertainty on the extrapolation factor, as detailed in above. The systematic uncertainties due to this can be seen in Tab. K.1 and are all at the sub-percent level. For those backgrounds with $m_{K\pi}$ outside the K^{*0} mass window an overall 50% uncertainty was assigned to the yield as detailed above. The resulting systematic uncertainties due to this can be seen in Tab. K.2 and are all at the sub-percent level.

q^2 bin	Data-taking period	L0 Trigger	$B^+ \to K^+e^+e^-$	$B^0 \to K^{*0}e^+e^-$
low	Run 1	L0I	$[-0.14, +0.14]\%$	$[-0.14, +0.13]\%$
		L0L	$[-0.13, +0.14]\%$	$[-0.12, +0.12]\%$
	15 + 16	L0I	$[-0.13, +0.13]\%$	$[-0.18, +0.18]\%$
		L0L	$[-0.12, +0.12]\%$	$[-0.14, +0.14]\%$
central	Run 1	L0I	$[-0.13, +0.13]\%$	$[-0.21, +0.20]\%$
		L0L	$[-0.12, +0.12]\%$	$[-0.25, +0.24]\%$
	15 + 16	L0I	$[-0.05, +0.04]\%$	$[-0.22, +0.15]\%$
		L0L	$[-0.10, +0.10]\%$	$[-0.26, +0.25]\%$

Table K.1: The bias (in percentage) of $B^+ \to K^+e^+e^-$ and $B^0 \to K^{*0}e^+e^-$ signal yields between fits with nominal extrapolation factor and fits with isospin extrapolation factor varied by $\pm 1\sigma$ of their uncertainties. The effect on $B^0 \to K^{*0}e^+e^-$ signal yields is slightly larger than on $B^+ \to K^+e^+e^-$ signal yields in central-q^2 . The overall systematic effect is at the sub-percent level.

q^2 bin	Data-taking period	L0 Trigger	$B^+ \to K^+e^+e^-$	$B^0 \to K^{*0}e^+e^-$
low	Run 1	L0I	$[-0.22, +0.20]\%$	$[-0.27, +0.28]\%$
		L0L	$[-0.58, +0.58]\%$	$[-0.39, 0.40]\%$
	15 + 16	L0I	$[-0.12, +0.12]\%$	$[-0.11, +0.11]\%$
		L0L	$[-0.37, +0.37]\%$	$[-0.31, +0.32]\%$
central	Run 1	L0I	$[-0.07, +0.05]\%$	$[-0.24, +0.16]\%$
		L0L	$[-0.12, +0.08]\%$	$[-0.55, +0.37]\%$
	15 + 16	L0I	$[-x, +x]\%$	$[-x, +x]\%$
		L0L	$[-0.11, +0.11]\%$	$[-0.46, +0.46]\%$

Table K.2: The bias (in percentage) of $B^+ \to K^+e^+e^-$ and $B^0 \to K^{*0}e^+e^-$ signal yields between fits with nominal expected S-wave yields and fits with expected S-wave yields varied by $\pm 1\sigma$ of their uncertainties. The systematic effect on $B^0 \to K^{*0}e^+e^-$ compared to $B^+ \to K^+e^+e^-$ is smaller in low-q^2 but larger in central-q^2 . The overall systematic effect is at the sub-percent level.

Bibliography

[1] Steven Weinberg. A Model of Leptons. *Phys. Rev. Lett.*, 19:1264–1266, 1967.

[2] Abdus Salam. Weak and Electromagnetic Interactions. *Conf. Proc.*, C680519:367–377, 1968.

[3] S. L. Glashow. Partial Symmetries of Weak Interactions. *Nucl. Phys.*, 22:579–588, 1961.

[4] Harald Fritzsch and Murray Gell-Mann. Current algebra: Quarks and what else? *eConf*, C720906V2:135–165, 1972.

[5] H. Fritzsch, Murray Gell-Mann, and H. Leutwyler. Advantages of the Color Octet Gluon Picture. *Phys. Lett. B*, 47:365–368, 1973.

[6] Peter W. Higgs. Broken Symmetries and the Masses of Gauge Bosons. *Phys. Rev. Lett.*, 13:508–509, 1964.

[7] F. Englert and R. Brout. Broken Symmetry and the Mass of Gauge Vector Mesons. *Phys. Rev. Lett.*, 13:321–323, 1964.

[8] G.S. Guralnik, C.R. Hagen, and T.W.B. Kibble. Global Conservation Laws and Massless Particles. *Phys. Rev. Lett.*, 13:585–587, 1964.

[9] Gerard 't Hooft and M.J.G. Veltman. Regularization and Renormalization of Gauge Fields. *Nucl. Phys. B*, 44:189–213, 1972.

[10] MissMJ.

[11] G. Zweig. An SU(3) model for strong interaction symmetry and its breaking. Version 1. 1 1964.

[12] G. Zweig. *An SU(3) model for strong interaction symmetry and its breaking. Version 2*, pages 22–101. 2 1964.

[13] Murray Gell-Mann. The Eightfold Way: A Theory of strong interaction symmetry. 3 1961.

[14] M. Tanabashi et al. Review of particle physics. *Phys. Rev. D*, 98:030001, Aug 2018.

[15] H. Weyl. Electron and Gravitation. 1. (In German). *Z. Phys.*, 56:330–352, 1929.

[16] E.C.G. Stueckelberg. Interaction energy in electrodynamics and in the field theory of nuclear forces. *Helv. Phys. Acta*, 11:225–244, 1938.

[17] Eugene P. Wigner. Invariance in physical theory. *Proceedings of the American Philosophical Society*, 93(7):521–526, 1949.

[18] Roel Aaij et al. Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \to J/\psi K^- p$ Decays. *Phys. Rev. Lett.*, 115:072001, 2015.

[19] Roel Aaij et al. Observation of a narrow pentaquark state, $P_c(4312)^+$, and of two-peak structure of the $P_c(4450)^+$. *Phys. Rev. Lett.*, 122(22):222001, 2019.

[20] M.L. Perl, E.R. Lee, and D. Loomba. A Brief review of the search for isolatable fractional charge elementary particles. *Mod. Phys. Lett. A*, 19:2595–2610, 2004.

[21] G. Marx. Die wechselwirkung der elementar-teilchen und die erhaltungssätze. *Acta Physica*, 3:55–58, 1953.

[22] E.J. Konopinski and H.M. Mahmoud. The Universal Fermi interaction. *Phys. Rev.*, 92:1045–1049, 1953.

[23] Y.B. Zeldovich. O neitrinnom zaryade elementarnykh chastits. *Doklady Akademii Nauk SSSR*, 91(6):1317–1320, 1953.

[24] M. Czakon, M. Zralek, and J. Gluza. Are neutrinos Dirac or Majorana particles? *Acta Phys. Polon.*, B30:3121–3138, 1999.

[25] W. Pauli. Relativistic field theories of elementary particles. *Rev. Mod. Phys.*, 13:203–232, Jul 1941.

[26] David Griffiths. *Introduction to elementary particles*. 2008.

[27] Guido Altarelli. *Collider Physics within the Standard Model: A Primer*, volume 937. Springer, 2017.

[28] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson. Experimental Test of Parity Conservation in Beta Decay. *Phys. Rev.*, 105:1413–1414, 1957.

[29] 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.*, 13:138–140, 1964.

[30] Georges Aad et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett.*, B716:1–29, 2012.

[31] Serguei Chatrchyan et al. Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC. *Phys. Lett.*, B716:30–61, 2012.

[32] Nicola Cabibbo. Unitary Symmetry and Leptonic Decays. *Phys. Rev. Lett.*, 10:531–533, 1963. [,648(1963)].

[33] S. L. Glashow, J. Iliopoulos, and L. Maiani. Weak Interactions with Lepton-Hadron Symmetry. *Phys. Rev.*, D2:1285–1292, 1970.

[34] J. E. Augustin et al. Discovery of a Narrow Resonance in e^+e^- Annihilation. *Phys. Rev. Lett.*, 33:1406–1408, 1974. [Adv. Exp. Phys.5,141(1976)].

[35] J. J. Aubert et al. Experimental Observation of a Heavy Particle J. *Phys. Rev. Lett.*, 33:1404–1406, 1974.

[36] Blibla.

[37] Makoto Kobayashi and Toshihide Maskawa. CP Violation in the Renormalizable Theory of Weak Interaction. *Prog. Theor. Phys.*, 49:652–657, 1973.

[38] Lincoln Wolfenstein. Parametrization of the Kobayashi-Maskawa Matrix. *Phys. Rev. Lett.*, 51:1945, 1983.

[39] Yasmine Sara Amhis et al. Averages of b-hadron, c-hadron, and τ-lepton

properties as of 2018. 9 2019.

[40] Y. Akrami et al. Planck 2018 results. I. Overview and the cosmological legacy of Planck. 7 2018.

[41] Jonathan L. Feng. Dark matter candidates from particle physics and methods of detection. *Annual Review of Astronomy and Astrophysics*, 48(1):495–545, 2010.

[42] Leszek Roszkowski, Enrico Maria Sessolo, and Sebastian Trojanowski. WIMP dark matter candidates and searches—current status and future prospects. *Reports on Progress in Physics*, 81(6):066201, may 2018.

[43] Peter W. Graham, Igor G. Irastorza, Steven K. Lamoreaux, Axel Lindner, and Karl A. van Bibber. Experimental Searches for the Axion and Axion-Like Particles. *Ann. Rev. Nucl. Part. Sci.*, 65:485–514, 2015.

[44] Antonio Boveia and Caterina Doglioni. Dark Matter Searches at Colliders. *Ann. Rev. Nucl. Part. Sci.*, 68:429–459, 2018.

[45] Sebastian Böser, Christian Buck, Carlo Giunti, Julien Lesgourgues, Livia Ludhova, Susanne Mertens, Anne Schukraft, and Michael Wurm. Status of Light Sterile Neutrino Searches. *Prog. Part. Nucl. Phys.*, 111:103736, 2020.

[46] Pierre Fayet and S. Ferrara. Supersymmetry. *Phys. Rept.*, 32:249–334, 1977.

[47] Y. Fukuda et al. Evidence for oscillation of atmospheric neutrinos. *Phys. Rev. Lett.*, 81:1562–1567, 1998.

[48] Q. R. Ahmad et al. Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury Neutrino Observatory. *Phys. Rev. Lett.*, 89:011301, 2002.

[49] K. Eguchi et al. First results from KamLAND: Evidence for reactor anti-neutrino disappearance. *Phys. Rev. Lett.*, 90:021802, 2003.

[50] Roel Aaij et al. Search for the lepton-flavour violating decays $B^0_{(s)} \to e^{\pm}\mu^{\mp}$. *JHEP*, 03:078, 2018.

[51] Ziro Maki, Masami Nakagawa, and Shoichi Sakata. Remarks on the unified model of elementary particles. *Prog. Theor. Phys.*, 28:870–880, 1962.

[,34(1962)].

[52] D. Hanneke, S. Fogwell Hoogerheide, and G. Gabrielse. Cavity Control of a Single-Electron Quantum Cyclotron: Measuring the Electron Magnetic Moment. *Phys. Rev.*, A83:052122, 2011.

[53] G. W. Bennett et al. Final Report of the Muon E821 Anomalous Magnetic Moment Measurement at BNL. *Phys. Rev.*, D73:072003, 2006.

[54] S. Schael et al. Precision electroweak measurements on the Z resonance. *Phys. Rept.*, 427:257–454, 2006.

[55] Morad Aaboud et al. Precision measurement and interpretation of inclusive W^+ , W^- and Z/γ^* production cross sections with the ATLAS detector. *Eur. Phys. J. C*, 77(6):367, 2017.

[56] S. Schael et al. Electroweak Measurements in Electron-Positron Collisions at W-Boson-Pair Energies at LEP. *Phys. Rept.*, 532:119–244, 2013.

[57] A. Abulencia et al. Measurements of inclusive W and Z cross sections in p anti-p collisions at s**(1/2) = 1.96-TeV. *J. Phys. G*, 34:2457–2544, 2007.

[58] Roel Aaij et al. Measurement of forward $W \to e\nu$ production in pp collisions at $\sqrt{s} = 8\,\text{TeV}$. *JHEP*, 10:030, 2016.

[59] C. Lazzeroni et al. Precision Measurement of the Ratio of the Charged Kaon Leptonic Decay Rates. *Phys. Lett. B*, 719:326–336, 2013.

[60] A. Aguilar-Arevalo et al. Improved Measurement of the $\pi \to e\nu$ Branching Ratio. *Phys. Rev. Lett.*, 115(7):071801, 2015.

[61] Antonio Pich. Precision Tau Physics. *Prog. Part. Nucl. Phys.*, 75:41–85, 2014.

[62] Gerhard Buchalla, Andrzej J. Buras, and Markus E. Lautenbacher. Weak decays beyond leading logarithms. *Rev. Mod. Phys.*, 68:1125–1144, 1996.

[63] Andrzej J. Buras. Weak Hamiltonian, CP violation and rare decays. In *Probing the standard model of particle interactions. Proceedings, Summer School in Theoretical Physics, NATO Advanced Study Institute, 68th session, Les Houches, France, July 28-September 5, 1997. Pt. 1, 2*, pages 281–539, 1998.

[64] Enrico Fermi. Tentativo di una teoria dell'emissione dei raggi beta. *Ric. Sci.*, 4:491–495, 1933.

[65] Simone Bifani, Sébastien Descotes-Genon, Antonio Romero Vidal, and Marie-Hélène Schune. Review of Lepton Universality tests in B decays. *J. Phys.*, G46(2):023001, 2019.

[66] Gudrun Hiller and Frank Kruger. More model-independent analysis of $b \to s$ processes. *Phys. Rev.*, D69:074020, 2004.

[67] Roel Aaij et al. Test of lepton universality with $\Lambda_b^0 \to pK^-\ell^+\ell^-$ decays. 2019.

[68] Roel Aaij et al. Test of lepton universality using $B^+ \to K^+\ell^+\ell^-$ decays. *Phys. Rev. Lett.*, 113:151601, 2014.

[69] Roel Aaij et al. Search for lepton-universality violation in $B^+ \to K^+\ell^+\ell^-$ decays. *Phys. Rev. Lett.*, 122(19):191801, 2019.

[70] A. Abdesselam et al. Test of lepton flavor universality in $B \to K\ell^+\ell^-$ decays. 2019.

[71] R. Aaij et al. Test of lepton universality with $B^0 \to K^{*0}\ell^+\ell^-$ decays. *JHEP*, 08:055, 2017.

[72] A. Abdesselam et al. Test of lepton flavor universality in $B \to K^*\ell^+\ell^-$ decays at Belle. 2019.

[73] I. Adachi et al. Measurement of the Differential Branching Fraction and Forward-Backward Asymmetry for B —> K(*)l+l-. 2008.

[74] Bernard Aubert et al. Measurements of branching fractions, rate asymmetries, and angular distributions in the rare decays $B \to K\ell^+\ell^-$ and $B \to K^*\ell^+\ell^-$. *Phys. Rev.*, D73:092001, 2006.

[75] A. Pattori. On the Standard Model prediction for R_K and R_{K^*}. *J. Phys. Conf. Ser.*, 770(1):012033, 2016.

[76] Bernat Capdevila, Andreas Crivellin, Sébastien Descotes-Genon, Joaquim Matias, and Javier Virto. Patterns of New Physics in $b \to s\ell^+\ell^-$ transitions in the light of recent data. *JHEP*, 01:093, 2018.

[77] Li-Sheng Geng, Benjamín Grinstein, Sebastian Jäger, Jorge Martin Camalich, Xiu-Lei Ren, and Rui-Xiang Shi. Towards the discovery of new physics with lepton-universality ratios of $b \to s\ell\ell$ decays. *Phys. Rev.*, D96(9):093006, 2017.

[78] Wolfgang Altmannshofer, Christoph Niehoff, Peter Stangl, and David M. Straub. Status of the $B \to K^*\mu^+\mu^-$ anomaly after Moriond 2017. *Eur. Phys. J.*, C77(6):377, 2017.

[79] Lyndon Evans and Philip Bryant. LHC Machine. *JINST*, 3:S08001, 2008.

[80] Esma Mobs. The CERN accelerator complex - 2019. Complexe des accélérateurs du CERN - 2019. Jul 2019. General Photo.

[81] A. Augusto Alves, Jr. et al. The LHCb Detector at the LHC. *JINST*, 3:S08005, 2008.

[82] Roel Aaij et al. LHCb Detector Performance. *Int. J. Mod. Phys.*, A30(07):1530022, 2015.

[83] Roel Aaij et al. Measurement of the b-quark production cross-section in 7 and 13 TeV pp collisions. *Phys. Rev. Lett.*, 118(5):052002, 2017. [Erratum: Phys. Rev. Lett.119,no.16,169901(2017)].

[84] A. Papanestis and C. D'Ambrosio. Performance of the lhcb rich detectors during the lhc run ii. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 2017.

[85] LHCb VELO TDR: Vertex locator. Technical design report. 2001.

[86] LHCb technical design report: Reoptimized detector design and performance. 2003.

[87] LHCb: Inner tracker technical design report. 2002.

[88] LHCb: Outer tracker technical design report. 2001.

[89] R. Aaij et al. Performance of the LHCb Vertex Locator. *JINST*, 9:P09007, 2014.

[90] R Arink et al. Performance of the LHCb Outer Tracker. *JINST*, 9(01):P01002,

2014.

[91] Ph. d'Argent et al. Improved performance of the LHCb Outer Tracker in LHC Run 2. *JINST*, 12(11):P11016, 2017.

[92] LHCb magnet: Technical design report. 2000.

[93] Roel Aaij et al. Measurement of the track reconstruction efficiency at LHCb. *JINST*, 10(02):P02007, 2015.

[94] LHCb: RICH technical design report. 2000.

[95] P. A. Cherenkov. Visible luminescence of pure liquids under the influence of γ-radiation. *Dokl. Akad. Nauk SSSR*, 2(8):451–454, 1934. [Usp. Fiz. Nauk93,no.2,385(1967)].

[96] LHCb calorimeters: Technical design report. 2000.

[97] H. Fessler, P. Freund, J. Gebauer, K. M. Glas, K. Pretzl, P. Seyboth, J. Seyerlein, and J. C. Thevenin. A SCINTILLATOR - LEAD PHOTON CALORIMETER USING OPTICAL FIBER READOUT SYSTEMS. *Nucl. Instrum. Meth.*, A240:284–288, 1985.

[98] A. A. Alves, Jr. et al. Performance of the LHCb muon system. *JINST*, 8:P02022, 2013.

[99] E. Rodrigues. The lhcb trigger system. *Nuclear Physics B - Proceedings Supplements*, 170:298 – 302, 2007. Proceedings of the 11th International Conference on B-Physics at Hadron Machines.

[100] Barbara Sciascia. Prompt physics analysis from the trigger candidates at LHCb: strategy and new dedicated "TURBO" and PID calibration streams. *PoS*, ICHEP2016:186, 2016.

[101] R Aaij et al. The LHCb Trigger and its Performance in 2011. *JINST*, 8:P04022, 2013.

[102] Roel Aaij et al. Design and performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC. *JINST*, 14(04):P04013, 2019.

[103] S Tolk, J Albrecht, F Dettori, and A Pellegrino. Data driven trigger

efficiency determination at LHCb. Technical Report LHCb-PUB-2014-039. CERN-LHCb-PUB-2014-039, CERN, Geneva, May 2014.

[104] R Aaij et al. Measurement of the ratio of branching fractions $BR(B_0 \to K^{\star 0}\gamma)/BR(B_{s0} \to \phi\gamma)$ and the direct CP asymmetry in $B_0 \to K^{\star 0}\gamma$. *Nucl. Phys.*, B867:1–18, 2013.

[105] Diego Alejandro Roa Romero. Study of charmless $B^0_{d,s} \to h^+h'^-\pi^0$ decays in LHCb, May 2013. Presented 24 May 2013.

[106] R. W. Forty. Ring-Imaging Cherenkov Detectors for LHCb. 1996.

[107] F. Archilli et al. Performance of the Muon Identification at LHCb. *JINST*, 8:P10020, 2013.

[108] Andreas Hocker et al. TMVA - Toolkit for Multivariate Data Analysis. 2007.

[109] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. A Brief Introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178:852–867, 2008.

[110] I Belyaev, T Brambach, N H Brook, N Gauvin, G Corti, K Harrison, P F Harrison, J He, P H Hilten, C R Jones, M Lieng, G Manca, S Miglioranzi, P Robbe, V Vagnoni, M Whitehead, and J Wishahi. Handling of the generation of primary events in Gauss, the LHCb simulation framework. Technical Report LHCb-PROC-2010-056. CERN-LHCb-PROC-2010-056, CERN, Geneva, Nov 2010.

[111] D. J. Lange. The EvtGen particle decay simulation package. *Nucl. Instrum. Meth.*, A462:152–155, 2001.

[112] Piotr Golonka and Zbigniew Was. PHOTOS Monte Carlo: A Precision tool for QED corrections in Z and W decays. *Eur. Phys. J.*, C45:97–107, 2006.

[113] John Allison et al. Geant4 developments and applications. *IEEE Trans. Nucl. Sci.*, 53:270, 2006.

[114] M. Clemencic, G. Corti, S. Easo, C. R. Jones, S. Miglioranzi, M. Pappagallo, and P. Robbe. The LHCb simulation application, Gauss: Design, evolution and experience. *J. Phys. Conf. Ser.*, 331:032023, 2011.

[115] LHCb TDR computing technical design report. 2005.

[116] Alexander Mazurov, Ben Couturier, Dmitry Popov, and Nathanael Farley. Microservices for systematic profiling and monitoring of the refactoring process at the LHCb experiment. *J. Phys. : Conf. Ser.*, 898(7):072037. 8 p, 2017.

[117] H. A. Bethe. Molière's theory of multiple scattering. *Phys. Rev.*, 89:1256–1266, Mar 1953.

[118] Virgil L. Highland. Some Practical Remarks on Multiple Scattering. *Nucl. Instrum. Meth.*, 129:497, 1975.

[119] G R Lynch and O I Dahl. Approximations to multiple Coulomb scattering. *Nucl. Instrum. Methods Phys. Res., B*, 58(LBL-28165-REV):6–10. 14 p, Nov 1990.

[120] Geant4 em physics lists. `http://geant4.cern.ch/collaboration/working_groups/electromagnetic/physlist10.3.shtml`. Accessed: 2010-09-28.

[121] LHCB Rare Decays Working Group.

[122] R. Brun and F. Rademakers. ROOT: An object oriented data analysis framework. *Nucl. Instrum. Meth. A*, 389:81–86, 1997.

[123] M. Williams, V. Gligorov, C. Thomas, H. Dijkstra, J. Nardulli, and P. Spradlin. The HLT2 Topological Lines. 2011.

[124] V. V. Gligorov, Christopher Thomas, and Mike Williams. The HLT inclusive B triggers. 2011.

[125] V. V. Gligorov and Mike Williams. Efficient, reliable and fast high-level triggering using a bonsai boosted decision tree. *JINST*, 8:P02013, 2013.

[126] Tatiana Likhomanenko, Philip Ilten, Egor Khairullin, Alex Rogozhnikov, Andrey Ustyuzhanin, and Mike Williams. LHCb Topological Trigger Reoptimization. *J. Phys. Conf. Ser.*, 664(8):082025, 2015.

[127] E. Rodrigues. Dealing with clones in the tracking. 2006.

[128] Liudmila Prokhorenkova, Gleb Gusev, Aleksandr Vorobev, Anna Veronika Dorogush, and Andrey Gulin. Catboost: Unbiased boosting with categorical features. In *Proceedings of the 32Nd International Conference on Neural Information Processing Systems*, NIPS'18, pages 6639–6649, USA, 2018.

Curran Associates Inc.

[129] Tatiana Likhomanenko, Alex Rogozhnikov, Alexander Baranov, Egor Khairullin, and Andrey Ustyuzhanin. Reproducible Experiment Platform. *J. Phys. Conf. Ser.*, 664(5):052022, 2015.

[130] Avrim Blum, Adam Kalai, and John Langford. Beating the hold-out: Bounds for k-fold and progressive cross-validation. In *Proceedings of the Twelfth Annual Conference on Computational Learning Theory*, COLT '99, pages 203–208, New York, NY, USA, 1999. ACM.

[131] Marie-Helene Schune, Francesco Polci, and Martino Borsato. HOP an additional tool for decays involving electrons. Technical Report LHCb-INT-2015-037. CERN-LHCb-INT-2015-037, CERN, Geneva, Nov 2015.

[132] Roel Aaij et al. Measurement of the electron reconstruction efficiency at LHCb. 9 2019.

[133] A. Rogozhnikov. Reweighting with Boosted Decision Trees. *J. Phys. Conf. Ser.*, 762(1):012036, 2016.

[134] Muriel Pivk and Francois R. Le Diberder. SPlot: A Statistical tool to unfold data distributions. *Nucl. Instrum. Meth. A*, 555:356–369, 2005.

[135] Tomasz Skwarnicki. *A study of the radiative CASCADE transitions between the Upsilon-Prime and Upsilon resonances.* PhD thesis, Cracow, INP, 1986.

[136] Diego Martínez Santos and Frederic Dupertuis. Mass distributions marginalized over per-event errors. *Nucl. Instrum. Meth. A*, 764:150–155, 2014.

[137] Nathan Philip Jurik. *Observation of J/ψ p resonances consistent with pentaquark states in $B^0 \to J/\psi K^+$ decays.* PhD thesis, Syracuse U., 8 2016.

[138] Roel Aaij et al. Measurement of b hadron fractions in 13 TeV pp collisions. *Phys. Rev. D*, 100(3):031102, 2019.

[139] Diego Casadei. Estimating the selection efficiency. *JINST*, 7:P08021, 2012.

[140] Roel Aaij et al. Measurements of the S-wave fraction in $B^0 \to K^+\pi^-\mu^+\mu^-$ decays and the $B^0 \to K^*(892)^0\mu^+\mu^-$ differential branching fraction. *JHEP*,

11:047, 2016. [Erratum: JHEP 04, 142 (2017)].

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

[142] Roel Aaij et al. Measurement of the $B^{\pm}$ production cross-section in pp collisions at $\sqrt{s} = 7$ and 13 TeV. *JHEP*, 12:026, 2017.

[143]