



Measurement of the ratios $\mathcal{B}(B_s^0 \rightarrow K^- \mu^+ \nu)/\mathcal{B}(B_s^0 \rightarrow D_s^- \mu^+ \nu)$ and $|V_{ub}|/|V_{cb}|$

Adlene Hicheur, Basem Khanji, Franz Muheim, Iwan Smith, Marta Calvi, Michel de Cian, Mika Vesterinen, Paul Seyfert, Svende Braun

Abstract

The first observation of a suppressed B_s^0 semileptonic decay, $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$, is reported. Using 2 fb^{-1} of data recorded in 2012, the measurement of the branching fraction ratio $\mathcal{B}(B_s^0 \rightarrow K^- \mu^+ \nu)/\mathcal{B}(B_s^0 \rightarrow D_s^- \mu^+ \nu) = (4.89 \pm 0.21(\text{stat})^{+0.20}_{-0.21}(\text{syst}) \pm \pm 0.14(D_s)) \times 10^{-3}$ and the ratio of the Cabibbo-Kobayashi-Maskawa (CKM) matrix elements at low and high momentum transfer, $|V_{ub}|/|V_{cb}|(\text{low}) = 0.0607 \pm 0.0015(\text{stat}) \pm 0.0013(\text{syst}) \pm 0.0008(D_s) \pm 0.0030(\text{FF})$ and $|V_{ub}|/|V_{cb}|(\text{high}) = 0.0946 \pm 0.0030(\text{stat})^{+0.0024}_{-0.0025}(\text{syst}) \pm 0.0013(D_s) \pm 0.0068(\text{FF})$, are performed. Those are obtained with the help of the normalization channel $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$, where D_s^- is reconstructed in the final state $K^+ K^- \pi^-$, and theoretical calculations on the Form Factors of the $B_s^0 \rightarrow K$ and $B_s^0 \rightarrow D_s$ transitions. The absolute value $\mathcal{B}(B_s^0 \rightarrow K \mu \nu) = (1.06 \pm 0.05(\text{stat})) \pm 0.04(\text{syst}) \pm 0.06(\text{expinp}) \pm 0.04(\text{FF}) \times 10^{-4}$ is also derived.

Versions

- Version v0: first draft delivered to the SL WG with core information on the analysis.
- Version v1: after first round of comments from SLWG. Additional information on fits and selection added. Plots added or changed (9,21,24,40,41,57). A bug was discovered in the proton component of the Kaon misID matrix calculation. Accounting for GhostProb and isMuon (trigger) effects in PID studies. A more stable paradigm (unfolding) is now used to estimate the misID, which modified the misID component yields and thus (slightly) the signal fit results.
- Version v1.1: After follow up on the SLWG reviewers comments. More explanation on the Bayes unfolding (main update).
- Version v1.2: After SLWG approval presentation and first RC comments. Added the details on the behaviour of the q^2 solutions for which the discriminant is zero (or set to zero), Fig.12. Added fit uncertainties without fluctuating the templates to estimate the share of the finite template size (Table 25). Added checks on K^* mode Form Factor alternative model (10.8.1, Figs.61,62), hypothetical $K\pi^0$ NR wave (10.8.1, Fig.63), fit systematic studies with toys (10.8.1) and appendix on the MisID matrices elements (D, and corresponding subsection 9.2). Pulls of the MC/data comparisons of the BDT shapes using $B^+ \rightarrow J/\psi K^+$, before any kinematical correction (Fig.24). Illustration of combinatorial shape variation (Fig.64). For the normalization fit, the numbers after the “alignment on the same triggering” strategy have been updated in the results + specification of the use of the McLean *et al.* prediction as a baseline, section 8.1.
- Version v1.3: After “unblinding” the results. Total efficiencies derived, see Table 34 for the baseline scenario. The numbers for the alternative models are in appendix I. Tables 43 and 44 on systematics edited and/or created. Section 11 filled with all the results along with their statistical, systematic and model errors.

Contents

1	Introduction	1
2	Phenomenology	2
2.1	Form Factor Calculations	3
2.1.1	Form Factor Calculations for $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$	3
2.1.2	Form Factor Calculations for $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$	4
2.1.3	Form factor Choice	6
3	Method and kinematics	7
3.1	Corrected Mass	7
3.2	Neutrino Reconstruction and q^2	10
3.3	Linear Regression to Reconstruct q^2	11
4	Dataset and Monte Carlo samples	12
4.1	Data set	12
4.2	Simulated samples	13
4.2.1	Composition of filtered inclusive samples	15
5	Selections	18
5.1	Trigger Selection	18
5.2	Stripping lines	18
5.2.1	Stripping lines for $B_s^0 \rightarrow K^+ \mu^- \nu$	18
5.2.2	Stripping lines for $B_s^0 \rightarrow D_s^+ \mu^- \nu$	19
5.3	Offline $B_s^0 \rightarrow K^+ \mu^- \nu$ selection	19
5.3.1	Vetos on K^{*+} excited states	19
5.3.2	Removal of combinatorial background	21
5.3.3	Specific background from misidentification	22
5.3.4	Charged isolation variable:	23
5.3.5	Summary of the offline selection	24
6	MVA background removal	27
7	Control and Normalization Channels corrections	32
8	Normalisation Fit	36
8.1	Normalisation Fit Model	36
8.2	Background Subtraction	37
8.3	Fit Results	37
9	Signal fit	41
9.1	Combinatorial Modeling	41
9.2	Mis-identification background	43
9.3	Excited K^* states	45
9.4	$B \rightarrow c\bar{c}(\rightarrow \mu\mu)KX$ (B2CC) modeling and constraint	46
9.5	Fitting results & projections	48
9.6	Fit goodness and toys	49

10 Efficiencies and systematics	51
10.1 Generator Efficiency	51
10.2 Particle Identification	51
10.3 Trigger efficiencies	53
10.4 Tracking Correction	54
10.5 $B^+ \rightarrow J/\psi K^+$ corrections	56
10.6 q^2 Migration	60
10.7 Final Corrected Relative Efficiency	62
10.8 Fit systematics	63
10.8.1 Signal fit systematics	63
10.8.2 Normalization fit systematics	70
10.9 Final Systematic Uncertainties	73
11 Extraction of the branching fraction and V_{ub}	76
11.1 Branching fractions	76
11.2 $ V_{ub} $	76
11.3 Discussion	77
A Form Factor Comparisons	79
A.1 Publications Used	80
A.2 z -expansion Fit Parameters	80
A.3 Comparison Plots	81
B Input distributions of the variables used in the BDTs	90
B.1 Charged BDT variables: signal vs background comparison	90
B.2 Same Sign BDT variables: signal vs background comparison	90
C Validation of Combinatoric Modelling	92
D Study of the elements of the MisID matrices	95
D.1 Elements as a function of the kinematics of the track	95
D.2 Elements as a function of m_{corr}	95
E Study of the mis-identification background for $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$	101
F Study of the same sign sample	102
G TisTos efficiency checks	104
H Projections of the Signal Fit	118
I Efficiencies and observables with alternative models	121
I.1 Efficiencies	121
I.2 Observables	121
References	125

1 Introduction

In order to test the unitarity of the CKM matrix and precisely measure the amount of CP violation in the quark sector, the parameters of the CKM matrix must be constrained. The CKM parameters can be constrained by performing measurements of observables sensitive to the magnitudes of the CKM matrix elements. Since $|V_{ub}|$ is the least well known of the CKM matrix elements it is the dominant limiting factor when drawing CKM unitary triangles. An improved uncertainty on $|V_{ub}|$ will improve the global precision of fits to the CKM unitary triangles and test the unitarity of the CKM matrix. Non unitarity of the CKM matrix would be indicative of new physics beyond the standard model.

The $|V_{ub}|$ and $|V_{cb}|$ matrix elements can be determined from inclusive and exclusive semileptonic decays of a B hadron. When performing an exclusive measurement all visible¹ decay products of the B are reconstructed, and an inclusive decay, $B \rightarrow P\ell^-\bar{\nu}X$, contains additional unreconstructed decay particles. Inclusive determinations of $|V_{cb}|$ combine measurements of the semileptonic $b \rightarrow c\mu^-\bar{\nu}_\mu X$ decay rate with the leptonic energy, the hadronic invariant mass spectra and a theoretical input. Inclusive measurements of $|V_{cb}|$ were performed by the B factories operating at the $\Upsilon(4S)$ production energy and by LEP using B mesons produced from the decays of the Z boson. The B factories had the benefit of higher statistics and produced more precise determinations than LEP while the boosted B mesons from the Z allowed measurements to be made in a larger phase space.

An inclusive measurement of $|V_{ub}|$ is complicated due to the enormous backgrounds originating from $B \rightarrow X_c\ell^-\bar{\nu}$ decays. A kinematic approach is usually taken and inclusive measurements are performed in the region where charm backgrounds are kinematically forbidden although statistics can be increased by extending the phase space into the $B \rightarrow X_c e^-\bar{\nu}$ region. CLEO, Belle and BaBar have quoted partial rates of $B \rightarrow X_c\ell^-\bar{\nu}$ for $|\vec{p}_e| \geq 2.0 \text{ GeV}$ and $|\vec{p}_e| \geq 1.9 \text{ GeV}$ which is well below the charm kinematic endpoint.

Exclusive determinations of $|V_{cb}|$ are based on semileptonic $B \rightarrow D^{(*)}\ell^-\bar{\nu}$ decays in the limit $m_{b,c} \gg \Lambda_{\text{QCD}}$. Exclusive measurements of $|V_{ub}|$ are made by combining the exclusive decay rate of B hadrons combined with form factor predictions. The decays $B^0 \rightarrow \pi^-\mu^+\nu_\mu$, $B_s^0 \rightarrow K^-\mu^+\nu_\mu$ and $\Lambda_b^0 \rightarrow p\mu^-\bar{\nu}_\mu$ which contain a ground state hadron in the final state are “golden modes” for lattice QCD predictions and have the lowest theoretical uncertainties.

Form factors provided by lattice QCD are most accurate in the kinematic region with high momentum transfer.

The averaged $|V_{cb}|$ measurements are [1]

$$|V_{cb}|_{Incl} = (42.2 \pm 0.8) \times 10^{-3}, \quad |V_{cb}|_{Excl} = (41.9 \pm 2.0) \times 10^{-3}, \quad (1)$$

and the averaged $|V_{ub}|$ measurements are

$$|V_{ub}|_{Incl} = (4.49 \pm 0.28) \times 10^{-3}, \quad |V_{ub}|_{Excl} = (3.70 \pm 0.16) \times 10^{-3}. \quad (2)$$

The difference between inclusive and exclusive measurements of $|V_{cb}|$ and $|V_{ub}|$ of approximately 3σ has been a long-standing puzzle in particle physics.

Exclusive determinations of $|V_{ub}|$ is possible with the decays $\Lambda_b^0 \rightarrow p\mu^-\bar{\nu}_\mu$ and $B_s^0 \rightarrow K^-\mu^+\nu_\mu$. The goal of this analysis is a first observation of the decay $B_s^0 \rightarrow K^-\mu^+\nu_\mu$

¹A visible particle is reconstructible by the detector. The neutrino is considered invisible.

40 with a measurement of the ratio of branching fractions $\frac{\mathcal{B}_{B_s^0 \rightarrow K^- \mu^+ \nu_\mu}}{\mathcal{B}_{B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu}}$ and a ratio of the
 41 CKM matrix elements $|V_{ub}|/|V_{cb}|$. This measurement uses data collected from pp collision
 42 events collected by the LHCb experiment in year 2012. The measured ratio of branching
 43 fractions is combined with theoretical inputs from Lattice QCD and Light-Cone Sum
 44 Rules allowing $|V_{ub}|/|V_{cb}|$ to be determined. This ratio provides an important constraint
 45 when performing global fits testing the unitarity of the CKM matrix.
 46 Section 2 discusses the phenomenology involved in the measurement, while the kinematic
 47 (fit and discrimination) variables are detailed in section 3. The data and simulation sam-
 48 ples are shown in section 4. The selection of the signal $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ and normalization
 49 $B_s^0 \rightarrow D_s(\rightarrow KK\pi)\mu\nu$ channels is described in section 5. The fits to the $D_s(KK\pi)\mu$ and
 50 $K\mu$ data samples are exhibited in sections 8 and 9. The derivation of the efficiencies
 51 and the systematics is shown in section 10, and the extraction of the main observable is
 52 explained in section 11.

53 2 Phenomenology

54 In order to extract the electroweak parameters $|V_{ub}|$ and $|V_{cb}|$ from the physically observable
 55 decay rates hadronic form factors are required. This section will present an overview of
 56 the current form factor calculations for $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ and $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$.

57 The amplitude of the semileptonic decay $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ can be written as a term
 58 proportional to the product of a leptonic current L^μ and a hadronic current H_μ [2]. When
 59 $q^2 \ll m_{W^\pm}^2$ the matrix element of the decay $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ may be written

$$\begin{aligned}\mathcal{M} &= -i \frac{G_F}{\sqrt{2}} V_{ub} L^\mu H_\mu \\ &= -i \langle K^-(p') \mu^+(k') \nu_\mu(k) | \mathcal{H}_{\text{eff}} | B_s^0 \rangle,\end{aligned}\quad (3)$$

60 where the leptonic current is

$$L^\mu = \bar{u} \gamma^\mu (1 - \gamma_5) \nu_\mu, \quad (4)$$

61 and the hadronic current is

$$H_\mu = \langle K^-(p') | \bar{u} \gamma^\mu b | B_s^0(p) \rangle - \langle K^-(p') | \bar{u} \gamma^\mu \gamma_5 b | B_s^0(p) \rangle, \quad (5)$$

62 which leads to an effective Hamiltonian of

$$\mathcal{H}_{\text{eff}} = \frac{G_F}{\sqrt{2}} V_{ub}^L [\bar{u} \gamma^\mu b - \bar{u} \gamma^\mu \gamma_5 b] \mu^+ \gamma_\mu (1 - \gamma_5) \nu_\mu, \quad (6)$$

63 where $\gamma_5 = i \gamma_0 \gamma_1 \gamma_2 \gamma_3$ which separates ψ into left and right handed currents, $\psi_L = \frac{1-\gamma_5}{2} \psi$
 64 and $\psi_R = \frac{1+\gamma_5}{2} \psi$. Since $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ is a pseudoscalar meson transition, $B_s^0(J^P =$
 65 $0^-) \rightarrow K^-(J^P = 0^-)$, the axial-vector component of H_μ is zero due to constraints on the
 66 spin of the outgoing u quark. The vector component of H^μ may be written as:

$$\langle K^-(p') | \bar{u} \gamma^\mu b | B_s^0(p) \rangle = f_+(q^2) \left(p^\mu + p'^\mu - \frac{m_{B_s^0}^2 - m_{K^-}^2}{q^2} q^\mu \right) + f_0(q^2) \frac{m_{B_s^0}^2 - m_{K^-}^2}{q^2} q^\mu, \quad (7)$$

67 where $q^\mu = p_{B_s^0}^\mu - p_{K^-}^\mu$ is the momentum transfer. The determination of the vector and
 68 scalar form factors, f_+ , f_0 are given in what follows.

69 2.1 Form Factor Calculations

70 The current non perturbative methods for the calculations of form factors for
 71 $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ include lattice QCD [3, 4] and light-cone sum rules [5]. The two cal-
 72 culation methods provide predictions which are complimentary in phase space, with
 73 calculations from lattice QCD restricted to high values of q^2 and calculations from light-
 74 cone sum rules restricted to low values of q^2 . Requirements on unitarity and analyticity
 75 can be used to extrapolate results out of the restricted q^2 region. Consequently two mea-
 76 surements of $|V_{ub}|/|V_{cb}|$ will be performed. One measurement restricted to high momentum
 77 transfer ($q^2 > 7$) using an average of the three LQCD predictions with an additional
 78 uncertainty originating from the extrapolation to lower values of q^2 . The other method
 79 restricted to low momentum transfer using form factor calculations from light-cone sum
 80 rules. The decay $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ is normalised to the decay $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$ for which form
 81 factor calculations from LQCD are available [6], [7], [8]. The form factor calculations
 82 for $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$ need not be restricted to high momentum transfer and the full phase
 83 space in q^2 is used.

84 2.1.1 Form Factor Calculations for $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$

85 The differential decay rate for $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ in the B_s^0 rest frame is

$$\frac{d\Gamma(B_s^0 \rightarrow K^- \mu^+ \nu_\mu)}{dq^2} = \frac{G_F^2 |V_{ub}|^2}{24\pi^3} \frac{(q^2 - m_\mu^2)^2 \sqrt{E_K^2 - m_K^2}}{q^4 m_{B_s^0}^2} \times \\ \left[\left(1 + \frac{m_\mu^2}{2q^2} \right) m_{B_s^0}^2 (E_K^2 - m_K^2) |f_+(q^2)|^2 + \frac{3m_\mu^2}{8q^2} (m_{B_s^0}^2 - m_K^2)^2 |f_0(q^2)|^2 \right], \quad (8)$$

86 which in the limit, $m_\mu^2 \ll q^2$, becomes

$$\frac{d\Gamma(B_s^0 \rightarrow K^- \mu^+ \nu_\mu)}{dq^2} = \frac{G_F^2 |V_{ub}|^2}{24\pi^3} (E_K^2 - m_K^2)^{3/2} |f_+(q^2)|^2, \quad (9)$$

87 where G_F is the Fermi coupling constant, E_K is the kaon energy in the B_s^0 rest frame, q
 88 is the momentum transfer or the invariant mass of the muon and neutrino, m_{μ,K,B_s^0} , are
 89 the masses of the muon, kaon and B_s^0 respectively. $|f_+|$ and $|f_0|$ are the vector and scalar
 90 form factors which parametrise the hadronic contributions to the electroweak decay and
 91 are calculated nonperturbatively using either lattice QCD or Light-Cone Sum Rules.

92 The form factors are parametrised using a $K = 2$ or $K = 3$ BCL parametrisation [9]
 93 to the variable z

$$z = (q^2, t_0) = \frac{\sqrt{1 - q^2/t_+} - \sqrt{1 - t_0/t_+}}{\sqrt{1 - q^2/t_+} + \sqrt{1 - t_0/t_+}}. \quad (10)$$

94 The $K = 3$ BCL parameterisation for the vector and scalar form factors are

$$f_+(q^2) = \frac{1}{1 - q^2/m_{B^*}^2} \sum_{k=1}^{K-1} b_+^{(k)} \left[z^k - (-1)^{k-K} \frac{k}{K} z^k \right], \\ f_0(q^2) = \frac{1}{1 - q^2/m_{B^*}^2} \sum_{k=1}^{K-1} b_0^{(k)} z^k, \quad (11)$$

where a pole is included at the theoretically predicted $m_{B^*} = 5.63$ GeV. The $K = 2$ BCL parameterisation for the vector and scalar form factors are

$$f_{+,0}(q^2) = \frac{f_{+,0}(0)}{1 - q^2/m_{B^*}^2} \left\{ 1 + b_{+,0}^{(1)} \left[z(q^2) - z(0) + \frac{1}{2} (z(q^2)^2 - z(0)^2) \right] \right\}. \quad (12)$$

At large recoil (low q^2) the vector and scalar form factors may be described by a single independent form factor

$$f_0(q^2) = \frac{m_{B_s^0}^2 - q^2}{m_{B_s^0}^2} f_+(q^2). \quad (13)$$

The coefficients, $b_{+,0}^{(k)}, f_{+,0}(0)$, for all models discussed in this section are given in Appendix A.

Three sets of form factor calculations for $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ are used in the determination of $|V_{ub}|/|V_{cb}|$. Lattice QCD predictions provide a precise determination of the form factors at low recoil transfer (high q^2) and are provided by Bouchard et al. [10] Flynn et al. [11] and A.X.Khadra et al. [12]. Calculations from QCD light-cone sum rules are most precise at large recoil (low q^2) and are provided by Khodjamirian and Rusov (K&R) [13].

The predicted form factors are plotted in Figure 1 and the predicted decay rates are plotted in Figure 2. The results of the form factor calculations are given at the end of this section.

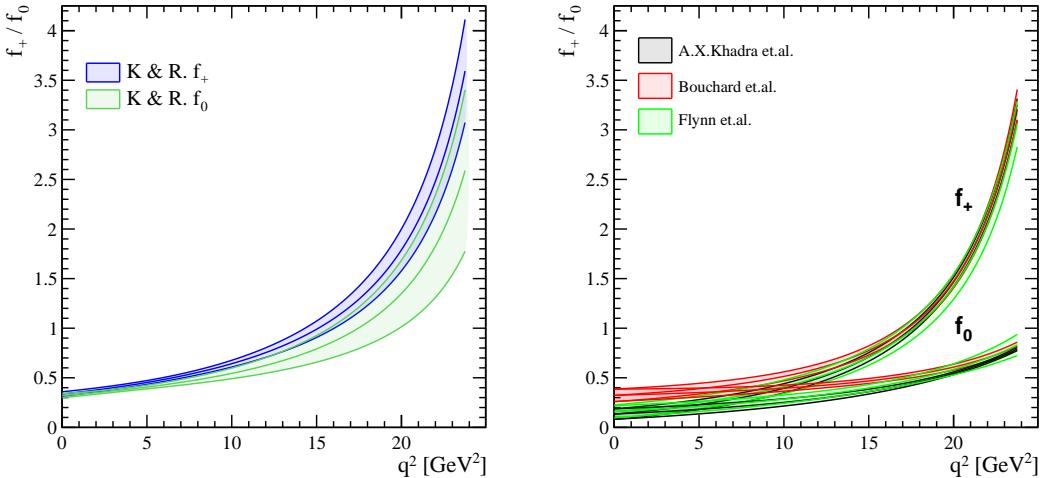


Figure 1: The form factor predictions for $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ calculated using QCD sum rules (left) and lattice QCD (right) from references [10–13].

2.1.2 Form Factor Calculations for $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$

The $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$ differential decay rate is given by

$$\frac{d\Gamma(B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu)}{d\omega} = \frac{G_F^2 |V_{cb}|^2}{48\pi^3} m_{D_s^+}^3 (m_{B_s^0} + m_{D_s^+})^2 (\omega^2 - 1)^{3/2} |\mathcal{G}(\omega)|^2, \quad (14)$$

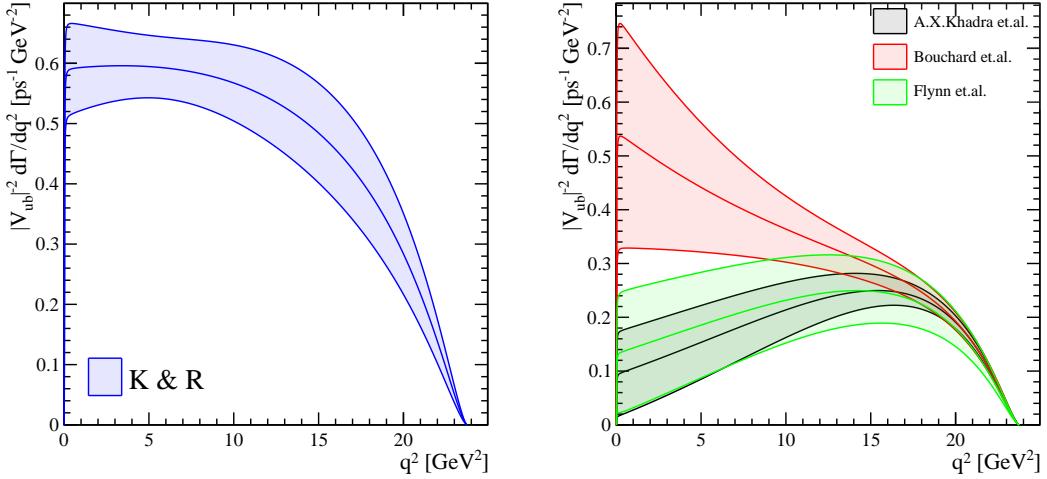


Figure 2: The predicted differential decay rates for $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ calculated using QCD sum rules (left) and lattice QCD (right) from references [10–13].

112 where $\mathcal{G}(\omega)$ is conventionally introduced as

$$\mathcal{G}(\omega) = \frac{2\sqrt{r}}{1+r} f_+(\omega), \quad (15)$$

113 with

$$\omega(q^2) = 1 + \frac{q^2_{\max} - q^2}{2m_{B_s^0} m_{D_s^+}} \quad \text{and} \quad r = \frac{m_{D_s^+}}{m_{B_s^0}}. \quad (16)$$

114 The form factor f_+ is parametrised using a modification of the BCL parametrisation [9]
115 with $J=3$

$$f_+(q^2) = \frac{1}{P_+} \sum_{j=0}^{J-1} a_j^{(+)} \left[z^j - (-1)^{j-J} \frac{j}{J} z^j \right], \quad (17)$$

116 where P_+ is called the Blashcke factor. $P_+(q^2)$ and z are given by

$$P_+(q^2) = \left(1 - \frac{q^2}{m_+^2} \right) \quad \text{and} \quad z(q^2) = \frac{\sqrt{t_+ - q^2} - \sqrt{t_+ - t_0}}{\sqrt{t_+ - q^2} + \sqrt{t_+ - t_0}}, \quad (18)$$

117 where $m_+ = m_{B_c^*} = 6.3309 \text{ GeV}$, $t_+ = (m_{B_s^0} + m_{D_s^+})^2$ and $t_0 = (m_{B_s^0} - m_{D_s^+})^2$. The
118 coefficients, $a_{+,0}^{(k)}$, for the models discussed in this section are given in Appendix A.

119
120 Three sets of lattice QCD form factor calculations for $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$ are used in the
121 determination of $|V_{ub}|/|V_{cb}|$ from Bailey et al. [6], Monahan et al. [7] and McLean et al [8].
122 The calculated form factors and differential decay rates are plotted in Figure 3.

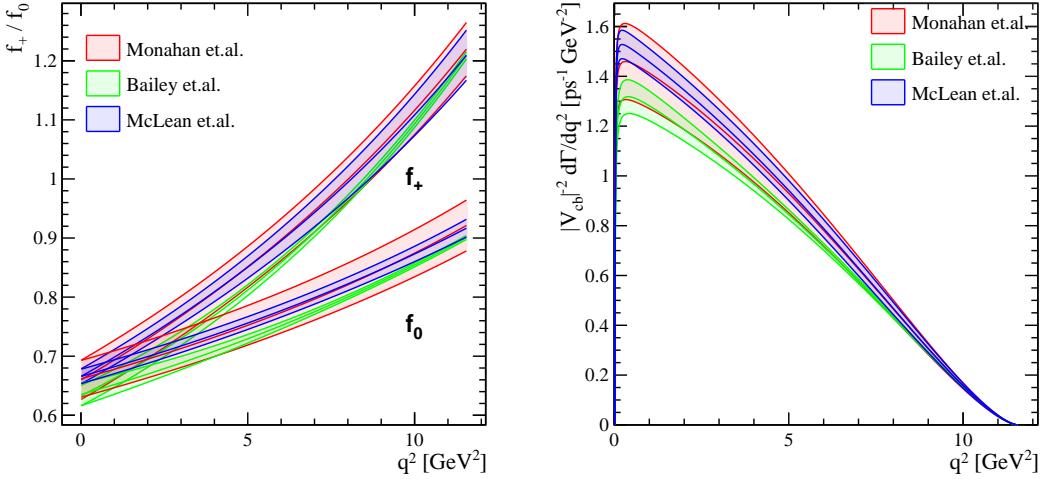


Figure 3: The form factor predictions (left) and differential decay rates (right) for $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$ calculated using lattice QCD from references [6–8].

	$\Gamma V_{ub} ^{-2} [ps^{-1}]$	$\Gamma V_{ub} ^{-2} [ps^{-1}] \Big _{q^2 > 7 \text{ GeV}^2}$	$\Gamma V_{ub} ^{-2} [ps^{-1}] \Big _{q^2 < 7 \text{ GeV}^2}$	$\mathcal{B}(B_s^0 \rightarrow K^- \mu^+ \nu_\mu)^1$
Flynn et al. [11]	4.54 ± 1.29	3.37 ± 0.74	1.18 ± 0.63	0.000093 ± 0.000027
A.X.Khadra et al. [12]	4.26 ± 0.92	3.32 ± 0.46	0.94 ± 0.48	0.000087 ± 0.000020
Bouchard et al. [10]	7.75 ± 1.56	4.47 ± 0.58	3.29 ± 1.00	0.000159 ± 0.000032
K & R [13]	11.07 ± 1.14	6.94 ± 1.04	4.14 ± 0.38	0.000229 ± 0.000024

Table 1: The predicted decay widths and branching fractions of $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ are presented for the form factor predictions given in References [10–13] for the full q^2 region and the high and low bins.

2.1.3 Form factor Choice

The form factor choice for $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ is to take the LCSR calculation at low q^2 and the latest LQCD one at high q^2 :

$$\Gamma|V_{ub}|^{-2} \Big|_{q^2 < 7 \text{ GeV}^2} = 4.14 \pm 0.38 \text{ ps}^{-1}, \quad \Gamma|V_{ub}|^{-2} \Big|_{q^2 > 7 \text{ GeV}^2} = 3.32 \pm 0.46 \text{ ps}^{-1}, \quad (19)$$

As for $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$, the latest calculation from McLean et al. is taken:

$$\Gamma|V_{cb}|^{-2} = 9.15 \pm 0.37 \text{ ps}^{-1}, \quad (20)$$

¹The exclusive average of $|V_{ub}|$ and $|V_{cb}|$ as determined by the PDG are used in the calculation of branching fractions [14].

	$\Gamma V_{cb} ^{-2} [ps^{-1}]$	$\mathcal{B}(B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu)^1$
Bailey et al. [6]	8.17 ± 0.24	0.0215 ± 0.0006
Monahan et al. [7]	8.98 ± 0.73	0.0238 ± 0.0020
McLean et al. [8]	9.15 ± 0.37	0.0242 ± 0.0011

Table 2: The predicted decay widths and branching fractions of $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$ are presented for the form factor predictions given in References [6, 7].

et al.

The full set of results from the form factor calculations including decay widths in different regions of phase space and predicted branching fractions are given in Tables 1 and 2 for $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ and $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$ respectively.

The Form Factor ratio (19 over 20) in low q^2 region has a relative 10% uncertainty, while it is 14% at high q^2 region, the theoretical uncertainty on $|V_{ub}|/|V_{cb}|$ will be 5% and 7%

3 Method and kinematics

Due to the missing neutrino, a partial reconstruction of the $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ decay chain is performed. The B_s^0 vertex is defined by the K^+ and μ^- tracks. With the knowledge of the B_s^0 flight direction one can use the symmetry of the decay to measure the transverse momentum of the invisible neutrino and built a reconstituted mass of the B_s^0 meson: the corrected mass. Furthermore, one can use the knowledge of the true mass of the B_s^0 meson to reconstruct the full kinematics of the invisible neutrino with a two fold ambiguity. Similar techniques are used for the normalization channel $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$ which selection criteria are aligned maximally with the signal selection to minimize the systematic uncertainties. The decay $B^+ \rightarrow J/\psi(\mu\mu)K^+$ is used as a control and calibration channel.

3.1 Corrected Mass

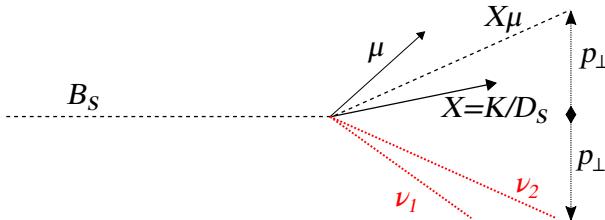


Figure 4: Visualisation of conservation of momentum with respect to the B_s^0 flight direction. Two neutrino solutions are compatible with the reconstructed decay.

As visualised in Figure 4, the event is rotated such that the B_s^0 meson flies in the z direction, and from the symmetry of the decay the transverse momentum of the neutrino must be equal and opposite to the transverse momentum of the visible system. The corrected mass is defined as

$$M_{corr} = \sqrt{M_{X\mu}^2 + p_\perp^2} + p_\perp, \quad (21)$$

with uncertainty

$$\sigma_{M_{corr}} = \left(\frac{p_\perp}{\sqrt{M_{X\mu}^2 + p_\perp^2}} + 1 \right) \sigma_{p_\perp} \quad (22)$$

where $M_{X\mu}$ is the visible invariant mass and p_\perp is the visible momentum transverse to the B_s^0 flight direction. If the only missing particle is a neutrino the corrected mass distribution will peak at the B_s^0 mass.

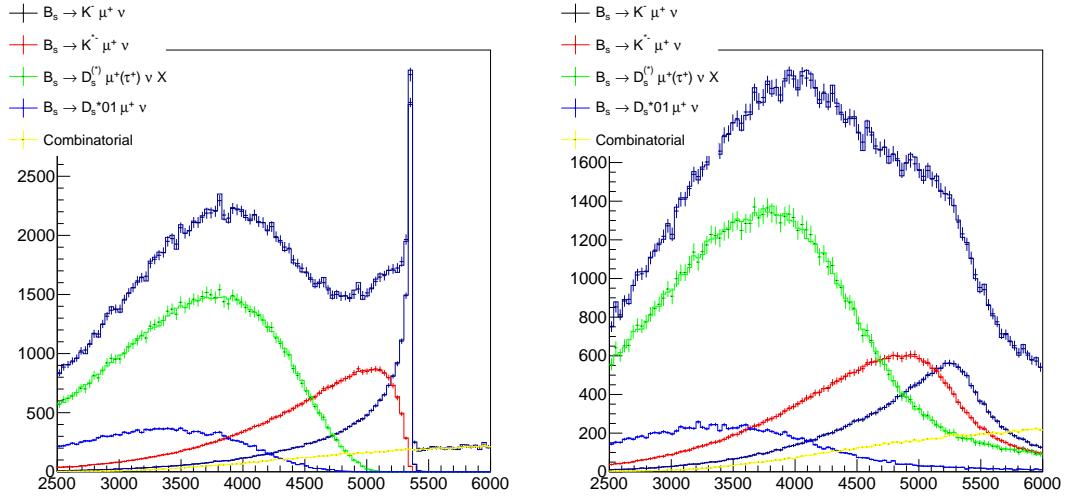


Figure 5: The corrected mass distribution for simulated signal and background events reconstructed as $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ before (left) and after (right) the modelling of vertex resolution.

The dominant source of uncertainty on the corrected mass comes from the uncertainty on the secondary (B) vertex position resulting in a large uncertainty on the momentum transverse to the B flight direction. The corrected mass distributions for simulated events reconstructed as $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ are plotted in Figure 5 before and after the simulation of resolution effects, the very sharp signal peak becomes significantly broader and harder to resolve with the addition of resolution effects.

Unsurprisingly the resolution on the plotted corrected mass is significantly improved if one rejects events with a large corrected mass uncertainty. The distributions of signal Monte Carlo decays and same sign data candidates are plotted in Figure 6. The distribution consisting of same sign candidates is shifted to the left when rejecting events with a large uncertainty on the corrected mass and the peak corresponding to the signal decay becomes significantly sharper. The additional resolution and separating power obtained by rejecting events with a high corrected mass uncertainty result in reduced systematics when performing a fit to the corrected mass.

For this analysis candidates with a corrected mass uncertainty greater than $100 \text{ MeV}/c^2$ are rejected. This selection has an efficiency of $\approx 45\%$ for both signal and partially reconstructed background decays while backgrounds from random combinations are significantly reduced. Although this selection doesn't increase signal purity the separation between signal and background decays is significantly improved in the corrected mass distribution resulting in a fit with significantly reduced systematics. The efficiency of this selection is verified using a kaon and muon combination from the decay $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+$ and this method is quantified later in Section 7. The distribution of the corrected mass uncertainty is plotted in Figure 7 for signal Monte Carlo and $K^- \mu^+$ combination from $B^+ \rightarrow J/\psi K^+$ using Monte Carlo and data. This selection is very effective at rejecting backgrounds from combinatoric combinations of a $K^- \mu^+$ pair. Combinatorics originate from two sources, the combination of prompt tracks originating from the primary vertex or from $b\bar{b}$ production with one b decaying semileptonically producing a muon and the other decaying hadronically producing a kaon. Combinations formed from the former production mechanism will have a very short decay length and

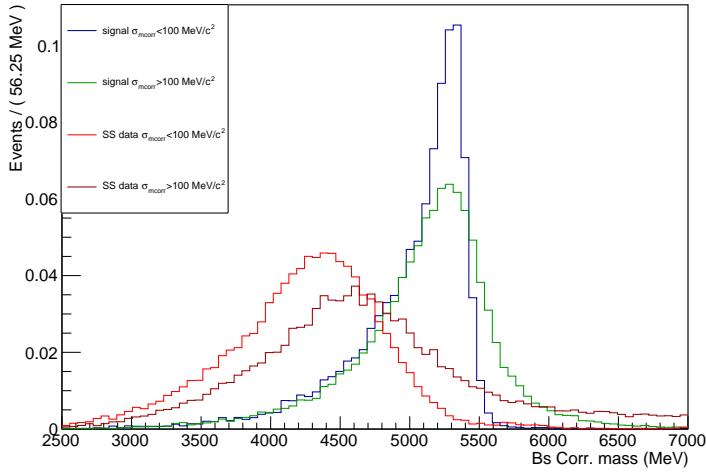


Figure 6: The corrected mass for Monte Carlo signal decays and same sign candidates from data. The effect of rejecting events with a high or low corrected mass uncertainty is demonstrated.

183 large corrected mass uncertainty while those produced from the latter method will form a
 184 secondary vertex with a large uncertainty resulting in a large corrected mass uncertainty.

185 When considering the decay $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$ no selection is made on the corrected mass
 186 uncertainty of the $K^- \mu^+$ pair or the $D_s^- \mu^+$ pair, since all or most of the combinatoric is
 187 removed by the D_s^- mass peak fit and particle ID requirement on the muon. Furthermore,
 188 contrary to the $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ case ($b \rightarrow u$), there is no need to separate further with a
 189 shape of a $b \rightarrow c$ background. The main background in this case, $B_s^0 \rightarrow D_s^{*-} \mu^+ \nu_\mu$, has a
 190 similar $\sigma_{m_{\text{corr}}}$ distribution as shown in Fig.8.

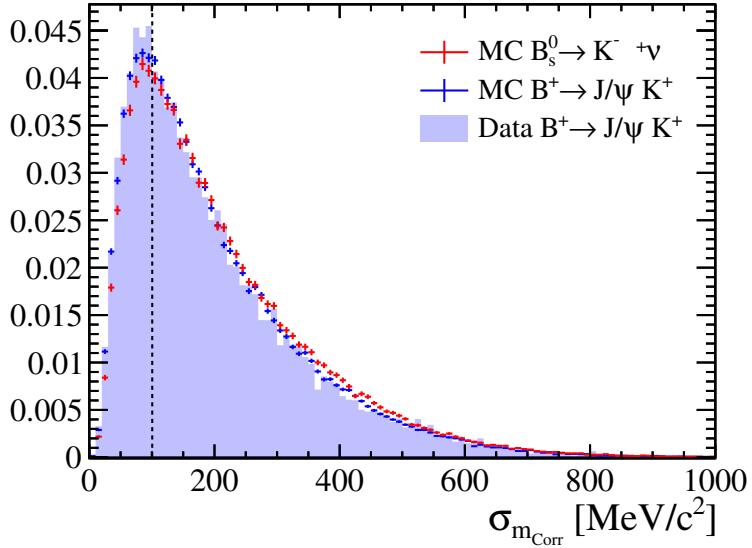


Figure 7: The corrected mass uncertainty for signal decays and $B^+ \rightarrow J/\psi K^+$ decays reconstructed as $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$.

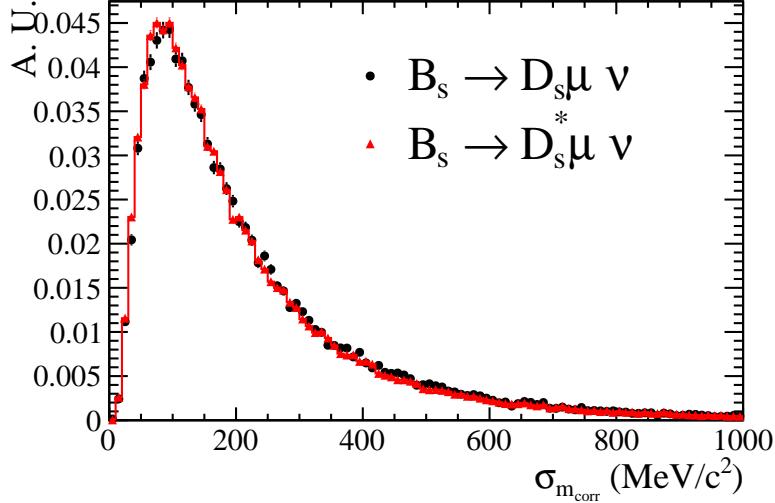


Figure 8: $\sigma_{m_{\text{corr}}}$ distributions of $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$ and $B_s^0 \rightarrow D_s^{-*} \mu^+ \nu_\mu$.

3.2 Neutrino Reconstruction and q^2

The $|V_{ub}|$ measurement will be performed by measuring the signal yield of $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ candidates in two regions of phase space separated by $q^2 \leq 7 \text{ GeV}^2$. A calculation of q^2 first requires the neutrino momentum be reconstructed. As discussed previously, the transverse component of the neutrino momentum p_\perp is trivial to calculate. The longitudinal component, p_\parallel , is less trivial and may be determined up to a two-fold ambiguity with the quadratic equation

$$p_\parallel = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}, \quad (23)$$

where a , b and c are defined as

$$\begin{aligned} a &= |2p_{\parallel,X\mu} m_{X\mu}|^2, \\ b &= 4p_{\parallel,X\mu}(2p_\perp p_{\parallel,X\mu} - m_{miss}^2), \\ c &= 4p_\perp^2(p_{\parallel,X\mu}^2 + m_{B_s^0}^2) - |m_{miss}|^2, \\ m_{miss}^2 &= m_{B_s^0}^2 - m_{X\mu}^2. \end{aligned} \quad (24)$$

The kinematics of the B_s^0 and q^2 of the event may now be calculated with a two fold ambiguity [15, 16]. When performing a physics analysis it is desirable to resolve this ambiguity without the introduction of a bias in q^2 . A choice must be made on which of the two solutions of q^2 will be used when performing an analysis.

The simplest approach is to randomly select one of the two solutions² which while unbiased has a poor resolution in q^2 . A significantly improved method uses a linear regression model to predict the B_s^0 momentum and the ambiguity is resolved by selecting the solution most consistent with the regression value. The full details of the regression method are given in Ref [17].

Due to the detector resolution effects approximately 20% of the candidates have an unphysical solution (i.e. $b^2 < 4ac$) for P_\parallel . Those candidates happen to have $m_{corr}(B_s^0) >$

²To ensure determinism the “random” component is calculated from a hash of the event.

210 $m(B_s^0)$. Those events are dealt with by setting $b^2 = 4ac$, leading to a unique solution.
211 Those events are evenly spread across q^2 and their q^2 resolution is slightly degraded
212 with respect to the non-zero discriminant solutions, with a slight bias that is uniform
213 throughout the q^2 range, as shown in the next section.

214 3.3 Linear Regression to Reconstruct q^2

215 Linear regression analysis is a statistical technique for predicting the value of a target or
216 response variable based on relationships with predictor or regressor variables [18–20]. For
217 this analysis the momentum of the B_s^0 is inferred from the flight information of the B_s^0
218 with a resolution of 60% which is sufficient to select the correct solution of the quadratic
219 equation 70% of the time [17], on average (q^2 dependency).

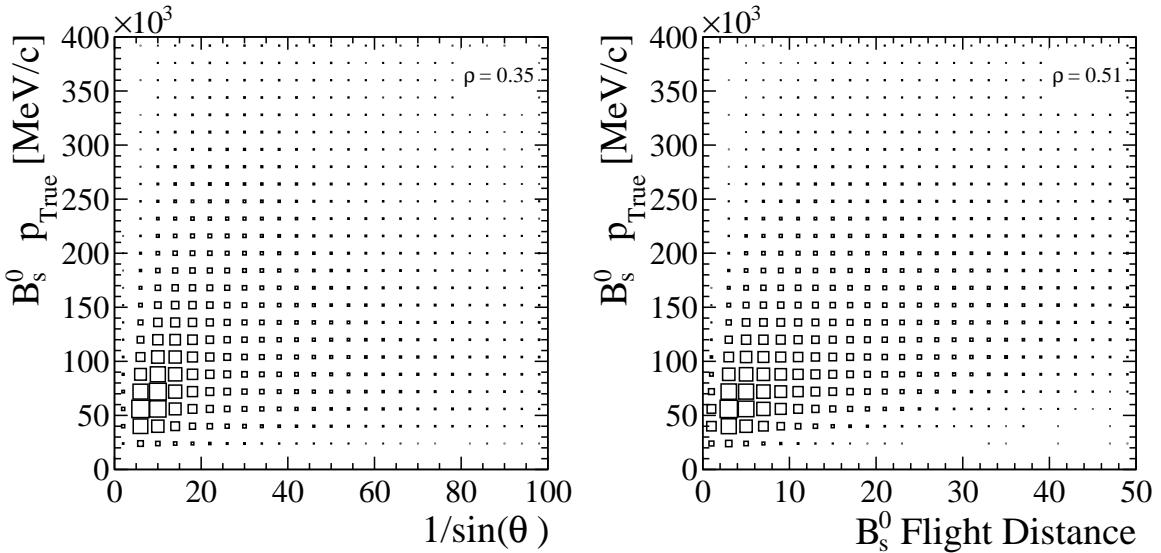


Figure 9: Input variables used to predict the B_s^0 momentum with a linear regression model.

220 The B_s^0 momentum may be weakly estimated from its polar angle, θ_{flight} ,

$$P = \frac{p_T}{\sin \theta_{\text{flight}}}, \quad (25)$$

221 and flight length, $|\vec{F}|$, and decay time, t (M is the B_s^0 mass),

$$P = \frac{M|\vec{F}|}{t}. \quad (26)$$

222 The two flight variables discussed above are plotted in Figure 9 and considered in a least
223 squares linear regression model [21]

$$P = \beta_0 + \beta_1 / \sin \theta_{\text{flight}} + \beta_2 |\vec{F}|, \quad (27)$$

224 where β_n are parameters to be found. In this analysis, those parameters are found by
225 applying the regression to the $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ simulation. The predicted value of the
226 B_s^0 momentum is compared to the two solutions derived from the quadratic equation in

227 Section 3.2 and the solution most consistent with the regression value, q^2_{Best} , is selected.
 228 The use of regression in the selection of a solution to the quadratic equation significantly
 229 improves the resolution on the reconstructed q^2 as plotted in Figure 10. The “correct”
 230 solution, set up for comparison purposes, is defined as being the one closest to the true q^2 .
 231 The resolution on the reconstructed q^2 for different methods of selecting a solution is given
 232 in Table 3. Using the output of the linear regression model to select a solution improves
 233 the resolution on the reconstructed q^2 by 38% when compared to a random selection.
 234 Figure 11 shows the resolution for the different q^2 bins: there is no real difference and the
 235 only seen effect is the inter-bin migration producing either asymmetric left or right tails. The
 236 cases (20% of the total) with solutions with a zero discriminant, as discussed in previous
 237 section, exhibit a slightly degraded resolution and a small uniform (throughout q^2) bias
 238 of $-0.28 \text{ GeV}^2/\text{c}^4$ as compared to the other cases, Fig.12. An adequate bin migration
 239 procedure, described in section 10.6, corrects for any impact of resolution and/or bias,
 240 particularly near the bin boundary, $7 \text{ GeV}^2/\text{c}^4$.

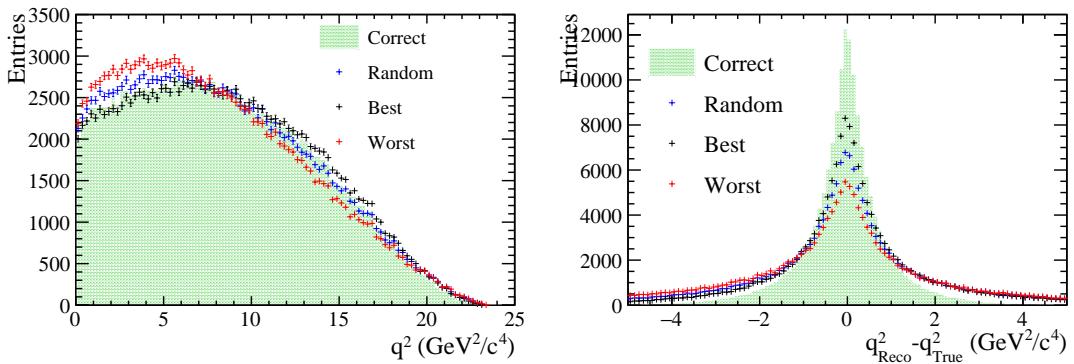


Figure 10: q^2 distributions with different methods of selecting the B_s^0 momentum solution (left) and the resolution on q^2 is shown using different selections (right).

Solution	RMS
Correct	$1.07 \text{ GeV}^2/\text{c}^4$
Regression	$2.21 \text{ GeV}^2/\text{c}^4$
Random	$3.06 \text{ GeV}^2/\text{c}^4$
Incorrect	$4.23 \text{ GeV}^2/\text{c}^4$

Table 3: Resolution on reconstructed q^2 after selecting one of the two solutions. Resolutions are given for the correct solution, solution obtained from regression, randomly selecting a solution and the incorrect solution.

241 4 Dataset and Monte Carlo samples

242 4.1 Data set

243 This analysis is performed using Run I data collected at LHCb in 2012, the sample
 244 corresponds to a luminosity of 2 fb^{-1} at 8 TeV . The full selection chain is described in section

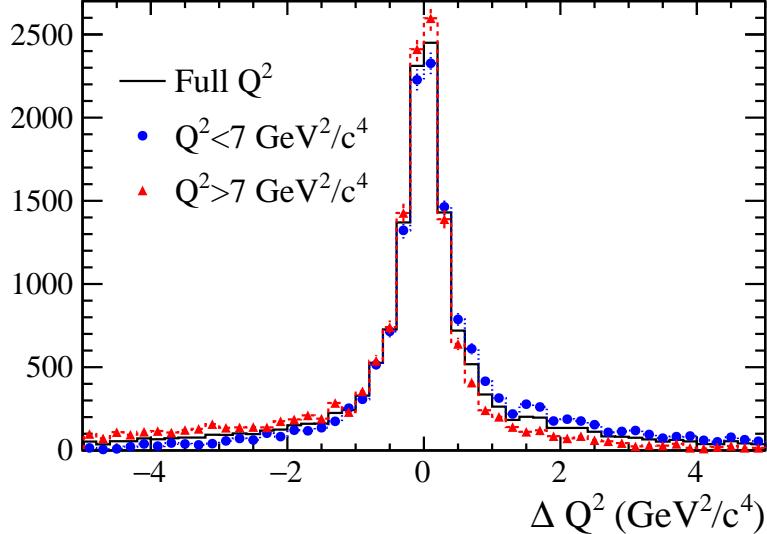


Figure 11: The resolution on q^2 for the full range, high and low q^2 bins.

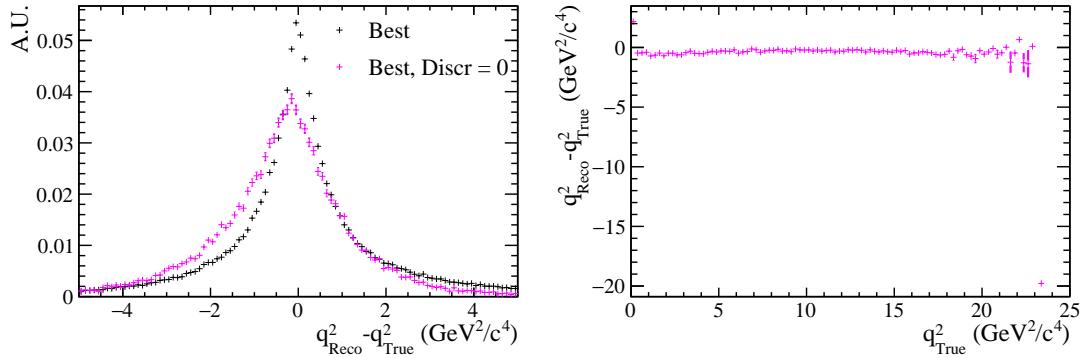


Figure 12: (Left) q^2 resolution distributions for all the cases with the “best” solution and the cases with a zero discriminant. (Right) Evolution of the small bias for the cases with a zero discriminant as a function of true q^2 .

245 5. Stripping lines are set up for both the signal and the normalization channel. To study
 246 misidentification backgrounds, prescaled lines are also considered with no K or μ PID
 247 requirements. Lines with same sign combination ($K^+\mu^+$) are also introduced to study
 248 specific combinatorial background sources which are present both in the pposite and same
 249 sign samples.

250 4.2 Simulated samples

251 The simulated samples include the $B_s^0 \rightarrow K^+\mu^-\nu$ and $B_s^0 \rightarrow D_s^+\mu^-\nu$ signals, exclusive
 252 backgrounds, and inclusive filtered samples, as listed in Tables 4 and 5.

253
 254 The signal $B_s^0 \rightarrow K^+\mu^-\nu$ sample and $B_s^0 \rightarrow K^{*+}\mu\nu$ background samples are generated
 255 using the ISGW2 model [22]³. Finally the inclusive samples have been designed as follows

³Re weighting procedure is used a posterior to account for different Form Factors models available in

Decay	Number of events (2011)	Number of events (2012)
$B_s^0 \rightarrow K^+ \mu^- \nu$ event type 13512010	1.5M	6M
$H_b \rightarrow H_c(\rightarrow K^+ \mu^- X) X'$ event type 10010032	-	250k (filtered)
$H_b \rightarrow K^+ \mu^+ X$ event type 10010035	-	536k (filtered)
$H_b \rightarrow H_c(\rightarrow K^+ X) \mu^- X'$ event type 10010037	-	1.492 M (filtered)
$B_s^0 \rightarrow K^{*+}(\rightarrow K^+ \pi^0) \mu \nu$ event type 13512400	-	4M
$B_s^0 \rightarrow K_2^{*+}(1430)(\rightarrow K^+ \pi^0) \mu \nu$ event type 13512410	2M	4M
$B_s^0 \rightarrow K_0^{*+}(1430)(\rightarrow K^+ \pi^0) \mu \nu$ event type 13512420	2M	4M
$B^0 \rightarrow J/\psi(\rightarrow \mu\mu) K^{*0}(K^+ \pi^-)$ event type 11144001	8M	10M
$B^+ \rightarrow J/\psi(\rightarrow \mu\mu) K^+$ event type 12143001	20M	20M
$B^+ \rightarrow J/\psi(\rightarrow \mu\mu) K^{*+}(K^+ \pi^0)$ event type 12143401	22M	20M
$B^+ \rightarrow c\bar{c}(\rightarrow \mu\mu) K X$ event type 12445022	-	1146302 Down + 1151704 Up (filtered)
$B^0 \rightarrow \pi \mu \nu$ event type 11512011	2M	4M
$B^0 \rightarrow \rho^+(\pi^+ \pi^0) \mu^- \nu$ event type 11512400	2M	4M
$B^+ \rightarrow \rho \mu^+ \nu$ event type 12513001	-	5M
$\Lambda_b^0 \rightarrow p \mu \nu$ event type 15512013	-	5M (FF from LQCD)
$\Lambda_b^0 \rightarrow p \mu \nu$ event type 15512014	-	10M (FF from LCSR)

Table 4: Number of generated events for the different samples related to the $B_s^0 \rightarrow K^+ \mu^- \nu$ fit.

256 ($H_{(c,b)}$ denotes charm or bottom hadrons):

- 257 • Filtered $H_b \rightarrow H_c(\rightarrow K^+ \mu^- X) X'$ events.
 258 • Filtered $H_b \rightarrow H_c(\rightarrow K^+ X) \mu^- X'$ events.
 259 • Filtered same sign $H_b \rightarrow K^+ \mu^+ X$ events.

the literature

Decay	Number of events (2011)	Number of events (2012)
$B_s^0 \rightarrow D_s^+ \mu \nu_\mu X$ cocktail event type 13774000	3M	6M
$B^+ \rightarrow D^0 \mu \nu$ event type 12873002	10M	15M
$B^+ \rightarrow D_s^{*+} D^*$ event type 12875601	-	5M
$B_s^0 \rightarrow D_s^{*+} D_s^{*+}$ event type 13873201	-	5M
$B_s^0 \rightarrow D^0 D_s^+ K$ event type 13796000	-	5M
$B_s^0 \rightarrow DD$ event type 13996202	1M	-

Table 5: Number of generated events for the different samples related to the $B_s^0 \rightarrow D_s^+ \mu^- \nu$ fit.

To improve the production efficiency of these samples and increase their yield(after stripping and offline selection), generator level cuts are applied as shown in Table 6. The stripping cuts (see section 5.2) are then applied at the reconstruction level, with the exception of the PID selection cuts ⁴

Particle	p cut	p_T cut
Kaon	> 7 GeV	> 0.3 GeV
Muon	> 5 GeV	> 1 GeV

Table 6: Generator level cuts for the simulated inclusive samples.

For the rest of exclusive backgrounds in Table 4 no special requests were made.

4.2.1 Composition of filtered inclusive samples

The main background for the signal decay comes from the V_{cb} modes $b \rightarrow c$ represented mainly by the decay types $H_b \rightarrow H_c (\rightarrow K^+ \mu^- X) X'$ and $H_b \rightarrow H_c (\rightarrow K^+ X) \mu^- X'$ (Table 4). Tables 7 and 8 show the individual contributions in these two samples after applying stripping cuts (including PID), numbers shown in those Tables are the ones corresponding to the output of the opposite-sign stripping line.

The third and last filtered sample corresponds to the event type: 10010035($H_b \rightarrow K^+ \mu^+ X$ decays). This sample was generated to understand the composition of our Same-Sign data sample. We report in Table 9 the composition of this simulation, numbers shown in the Table are the ones corresponding to the output of the same-sign stripping line.

For the exclusive samples, the Monte Carlo truth is used to match the kaon, the muon, their mother and grand-mother (requiring for a charm and a beauty hadron). For the

⁴To give the possibility to study mis-identified background sources.

Decay	Nb. of events	fraction [%]
All	61683	100
True $K^+\mu^-$	59948	97.19 ± 0.07
Beauty Semileptonic	1772	2.87 ± 0.07
Charm Semileptonic	59225	96.02 ± 0.08
Beauty to Charmonium	108	0.17 ± 0.02

Table 7: Composition of the MC sample corresponding to the event type: 10010032 ($H_b \rightarrow H_c(\rightarrow K^+\mu^-X)X'$ decays).

Decay	Nb. of events	fraction [%]
All	534159	100
True $K^+\mu^-$	523222	97.95 ± 0.02
Beauty to Charmonium	253	0.05 ± 0.003
Charm Semileptonic	101	0.51 ± 0.01
Beauty Semileptonic	520057	97.35 ± 0.02
$B^0 \rightarrow D^0\mu^-\nu_\mu X$	53524	10.02 ± 0.04
$B^0 \rightarrow D^+\mu^-\nu_\mu X$	52907	9.92 ± 0.04
$B^+ \rightarrow D^0\mu^-\nu_\mu X$	105562	19.76 ± 0.05
$B^+ \rightarrow D^+\mu^-\nu_\mu X$	18679	3.50 ± 0.03
$B_s^0 \rightarrow D^0\mu^-\nu_\mu X$	6269	1.17 ± 0.01
$B_s^0 \rightarrow D^+\mu^-\nu_\mu X$	14362	2.69 ± 0.02
$B_s^0 \rightarrow D_s^+\mu^-\nu_\mu X$	126457	23.67 ± 0.06
b Baryon Semileptonic	25202	4.72 ± 0.03

Table 8: Composition of the MC sample corresponding to the event type: 10010037 ($H_b \rightarrow H_c(\rightarrow K^+X)\mu^-X'$ decays).

Decay	Nb. of events	fraction [%]
All	249320	100
True $K^+\mu^-$	249320	98.47 ± 0.02
Beauty to Charmonium	6292	2.49 ± 0.03
Charm Semileptonic	5651	2.23 ± 0.03
Beauty Semileptonic	233784	92.33 ± 0.05
B^+ Semileptonic	121848	24.93 ± 0.07
B^0 Semileptonic	136079	42.28 ± 0.09
B_s^0 Semileptonic	125452	31.42 ± 0.08
b Baryon Semileptonic	25202	0.634 ± 0.01

Table 9: Decomposition of the MC sample corresponding to eventtype: 10010035 ($H_b \rightarrow K^+\mu^+X$ decays). Table is prepared using 2012 and 2011 combined

²⁷⁸ inclusive samples, true kaon and true muon are required but as shown in Table 9, the cut
²⁷⁹ Beauty semileptonic brings little to no change, specially after the full selection (next two

280 sections) is considered.

281 5 Selections

282 5.1 Trigger Selection

283 The Level 2 trigger requirements `Hlt2TopoMu2BodyDecisionTOS` and
 284 `Hlt2SingleMuonDecisionTOS` are requested to select candidates for both signal
 285 and normalisation decays and require as an input a muon having passed the L0 and HLT1
 286 single muon triggers, `Hlt1SingleMuonHighPT`, with a minimum transverse momentum
 287 of 1.76 GeV [23]. It should be mentioned here that `Hlt2SingleMuonDecisionTOS` has
 288 a prescale factor of 0.5 in data while the simulation forces the prescale factors to be
 289 one. To overcome this mismatch we apply a weight of 0.5 for events in the simulation
 290 which are exclusively triggered by this line to match the data. The `TopoMu2BodyBBDT`
 291 trigger is designed to select partially reconstructed decays of B hadrons containing a high
 292 likelihood muon [24, 25]. The trigger algorithm requires a displaced secondary vertex
 293 and a candidate is built from the muon and the additional particle. A bonsai boosted
 294 decision tree (BBDT) is employed to efficiently select signal events using discretised
 295 kinematics of the candidate [26]. To prevent the BDT from making a series of overly
 296 complicated selections the input variables to the trigger are discretised and the BDT may
 297 only introduce selections at the specific intervals listed in Table 10. The discretisation
 298 earns the BBDT its bonsai name. During optimisation of the BBDT the number of
 299 intervals was gradually reduced until a decrease in performance was observed.

Variable	Selection	BBDT Intervals
$\sum p_T [\text{GeV}/c]$	> 3	3.5, 4, 4.5, 5, 6, 7, 8, 9, 10, 15, 20
$p_T^{\min} [\text{GeV}/c]$	> 0.5	0.6, 0.7, 0.8, 0.9, 1, 1.25, 1.5, 1.75, 2, 2.5, 3, 4, 5, 10
$m [\text{GeV}/c^2]$	< 7	2.5, 4.75
$m_{\text{corr}} [\text{GeV}/c^2]$		2, 3, 4, 5, 6, 7, 8, 9, 10, 15
DOCA [mm]	< 0.2	0.05, 0.1, 0.15
IP χ^2		20
FD $\chi^2/100$	> 1	2, 3, 4, 5, 6, 7, 8, 9, 10, 25, 50, 100

Table 10: The variables and intervals used in the BBDT for trigger `TopoMu2BodyBBDT` selecting 2 body decays. Table taken from [26].

300 5.2 Stripping lines

301 5.2.1 Stripping lines for $B_s^0 \rightarrow K^+ \mu^- \nu$

302 The analysis is based on the stripping version "Stripping 21r0p1" (for 2012
 303 data). The stripping line for the signal is `StrippingB2XuMuNuBs2KLine` while
 304 we make use of a same sign line `StrippingB2XuMuNuBs2KSSLine` and MisPID
 305 lines(`StrippingB2XuMuNuBs2K_FakeKLine`, `StrippingB2XuMuNuBs2K_FakeMuLine`) for
 306 background studies. The requirements for the $B_s^0 \rightarrow K^+ \mu^- \nu$ signal stripping line
 307 `StrippingB2XuMuNuBs2KLine` ($K^+ \mu^-$) are shown in Table 11. The kaons and muons are
 308 initially taken from `StdLooseKaons` and `StdLooseMuons` particle containers. On top of
 309 that, the muons are required to pass the `LOMuonDecisionTOS` trigger.

Variables	Stripping cuts
Kaon	
Track P _T	> 0.5 GeV/c
Track P	> 10 GeV/c
Track IP χ^2	> 16
Track $\chi^2/\text{n.d.f.}$	< 6
Track GhostProb	< 0.5
$\Delta\log\mathcal{L}_{K\pi}$	> 5
$\Delta\log\mathcal{L}_{Kp}$	> 5
$\Delta\log\mathcal{L}_{K\mu}$	> 5
muon	
Track P _T	> 1.5 GeV/c
Track P	> 6 GeV/c
Track IP χ^2	> 12
Track $\chi^2/\text{n.d.f.}$	< 4
Track GhostProb	< 0.35
$\Delta\log\mathcal{L}_{\mu\pi}$	> 3
$\Delta\log\mathcal{L}_{\mu p}$	> 0
$\Delta\log\mathcal{L}_{\mu K}$	> 0
$K\mu$ (partial B_s^0)	
M_{CORR} window	[2.5, 7] GeV/c ²
Secondary Vertex χ^2	< 4
pointing angle	> 0.994
Flight Distance χ^2	> 120
HLT2	Hlt2SingleMuonDecisionTOS or Hlt2TopoMu2BodyDecisionTOS

Table 11: Stripping selection criteria for $B_s^0 \rightarrow K^+ \mu^- \nu$

5.2.2 Stripping lines for $B_s^0 \rightarrow D_s^+ \mu^- \nu$

The requirements for the $B_s^0 \rightarrow D_s^+ \mu^- \nu$ stripping line B2DMuNuX_Ds ($D_s^+ \mu^-$) are shown in Table 12. A similar line with misidentified muon, B2DMuNuX_Ds_FakeMu (prescaled by 0.02), has been designed for background studies. Additional requirements are exhibited in Table 13.

5.3 Offline $B_s^0 \rightarrow K^+ \mu^- \nu$ selection

Various types of background must be further reduced by offline cuts. The main backgrounds come from V_{cb} modes, i.e. decays of beauty to charmonium. The backgrounds from resonant $B_s^0 \rightarrow K^{*+} (\rightarrow K^+ \pi^0) \mu^- \nu$ modes should also be dealt with.

5.3.1 Vетос on K^{*+} excited states

π^0 candidates⁵ lying in the window $|M(\gamma\gamma) - 135| < 30 \text{ MeV}/c^2$ are searched around K^+ to form $K^+ \pi^0$ candidates. To sort out the multiple candidate problem, the π^0 with maximal

⁵from StdLoosePi02gg and StdLooseMergedPi0 containers

Variables	Stripping cuts
Event	Long track multiplicity < 250
muon	
Track P_T	> 1 GeV/c
Track P	> 6 GeV/c
Track GhostProb	< 0.35
Track $\chi^2/\text{n.d.f.}$	< 3.0
Track IP χ^2	> 12
$\Delta \log \mathcal{L}_{\mu\pi}$	> 3
Kaon	
Track P_T	> 250 MeV/c
Track P	> 2000 MeV/c
Track GhostProb	< 0.35
Track $\chi^2/\text{n.d.f.}$	< 3.0
Track IP χ^2	> 4
$\Delta \log \mathcal{L}_{K\pi}$	> -2.0
Pion	
Track P_T	> 250 MeV/c
Track P	> 2000 MeV/c
Track GhostProb	< 0.35
Track $\chi^2/\text{n.d.f.}$	< 3.0
Track IP χ^2	> 4
$\Delta \log \mathcal{L}_{K\pi}$	< 20.0
D_s^+ candidate	
$ m_{\text{Cand.}} - m_{D_s^+} $	< 80 MeV/c^2
DOCA χ^2	< 20
Vertex χ^2/N_{dof}	< 6.0
FD_{χ^2}	> 25
pointing angle	> 0.99
$D_s^+ \mu^-$ (partial B_s^0)	
m_{Cand}	[2.2, 8] GeV/c^2
Vertex χ^2/N_{dof}	< 9.0
pointing angle	> 0.999
Vertex(D_s^+) $_Z$ - Vertex(B_s^0) $_Z$	> -0.05

Table 12: Stripping selections applied to $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$ candidates using the `B2DMuNuX_Ds` line. Selections are aligned with those for $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ as closely as possible.

sum of confidence levels of the photons, $CL(\gamma_1) + CL(\gamma_2)$, is chosen. Figure 13 left shows the $M(\gamma\gamma)$ distribution for candidates with best $CL(\gamma_1) + CL(\gamma_2)$. Figure 13 right shows the $M(K^+\pi^0)$ distribution after all the requirements on the π^0 are applied. The candidates falling in the windows $|M(K^+\pi^0) - 892| < 60 \text{ MeV}/c^2$ and $|M(K^+\pi^0) - 1430| < 90 \text{ MeV}/c^2$ are vetoed. This selection only rejects $\approx 20\%$ of the background from higher excited resonances due to the low reconstruction efficiency of soft pions [27]. The remainder has its m_{corr} shape modeled for the final fit. The efficiency of this requirement on signal is $(95.7 \pm 0.1)\%$.

Additional Selections	
K^-	$\ln \mathcal{L}_K - \ln \mathcal{L}_\pi > 5.0$
K^-	$\ln \mathcal{L}_K - \ln \mathcal{L}_p > 5.0$
K^-	$\ln \mathcal{L}_K - \ln \mathcal{L}_\mu > 5.0$
K^-	$p > 10000 \text{ MeV}$
μ^+	$\ln \mathcal{L}_\mu - \ln \mathcal{L}_\pi > 3.0$
μ^+	$\ln \mathcal{L}_\mu - \ln \mathcal{L}_p > 0$
μ^+	$\ln \mathcal{L}_\mu - \ln \mathcal{L}_K > 5.0$
Vetoes	
$D^{*-} \rightarrow D^0 \pi^+$ Veto	$m_{KK\pi} - m_{KK} > 138 \text{ MeV}$
$B_s^0 \rightarrow D_s^- \pi^+$ Veto	$ m_{Ds(\mu \rightarrow \pi)} - m_{B_s^0} > 70 \text{ MeV}$
BDTs	
Isolation	$\min(\text{IsoMinBDT_K}, \text{IsoMinBDT_Mu}) > -0.8$

Table 13: Additional selections and vetoes applied to $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$ candidates.

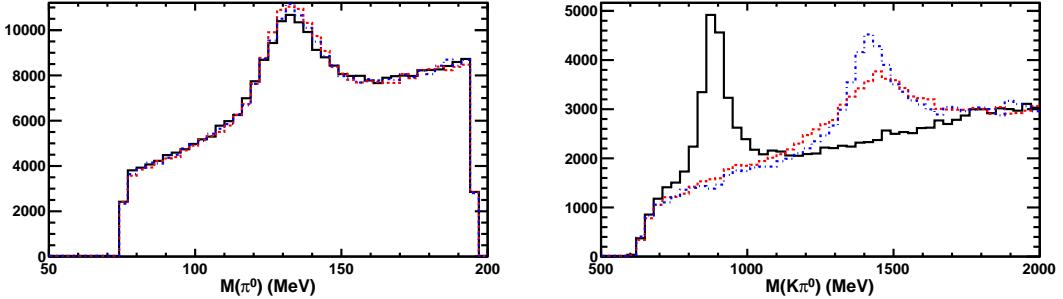


Figure 13: (Left) $M(\gamma\gamma)$ and (right) $M(K^+\pi^0)$ distributions for (black solid) $B_s^0 \rightarrow K^{*+}(892)(\rightarrow K^+\pi^0)\mu^-\nu$, (dashed red) $B_s^0 \rightarrow K_0^{*+}(1430)(\rightarrow K^+\pi^0)\mu^-\nu$ and (dot-dashed blue) $B_s^0 \rightarrow K_2^{*+}(1430)(\rightarrow K^+\pi^0)\mu^-\nu$ simulated samples. The $M(K^+\pi^0)$ distribution is shown with the requirement $|M(\gamma\gamma) - 135| < 30 \text{ MeV}/c^2$ for the π^0 candidates. Both distributions have $\max(CL(\gamma_1) + CL(\gamma_2))$ for the photon pair forming the π^0 .

330

331 5.3.2 Removal of combinatorial background

332 The pure combinatorial background, where the kaon and the muon originate from different
 333 decay chains, are believed to give candidates of low $p_T(B_s^0)$ as it can be seen in Fig.15
 334 It is also believed that the corresponding kaon-muon pairs have transverse momenta
 335 in opposite ‘‘quadrants’’, i.e., $P_X(K^+) \times P_X(\mu^-) < 0$ AND $P_Y(K^+) \times P_Y(\mu^-) < 0$, as
 336 illustrated in Fig.14 and confirmed in Fig.15, while pairs coming from the same (physical)
 337 decay come from the same quadrant, as observed in Fig.16. The efficiency of the removal
 338 of the low $p_T(B_s^0)$ component ranges from 92% for the signal to 97% for the inclusive V_{cb}
 339 $H_b \rightarrow H_c (\rightarrow K^+ X) \mu^- X'$ background.

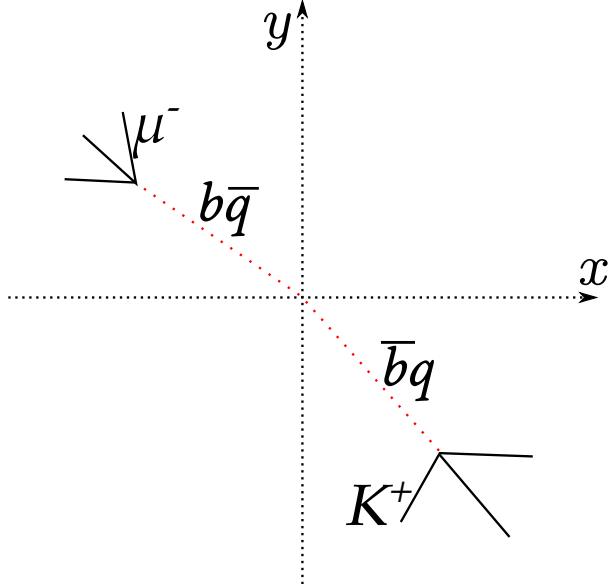


Figure 14: The topology of a combinatoric candidate looking down the z axis, beam line axis, with kaon and muon originating from the decay of different B mesons. The two B mesons are produced *back to back* in the transverse plane. The kaon and muon are visualised in opposite quadrants of the xy plane.

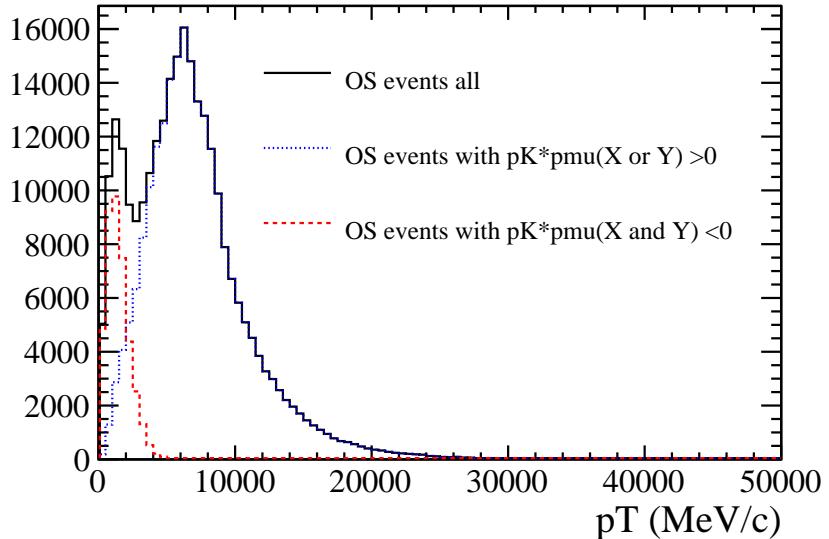


Figure 15: p_T distribution of the $B_s^0 \rightarrow K^+ \mu^-$ candidates for the (black solid) full data sample, (dashed red) events with $PX(K^+) \times P_X(\mu^-) > 0$ AND $P_Y(K^+) \times P_Y(\mu^-) > 0$ and (dotted blue) events with $PX(K^+) \times P_X(\mu^-) < 0$ OR $P_Y(K^+) \times P_Y(\mu^-) < 0$

³⁴⁰ 5.3.3 Specific background from misidentification

³⁴¹ Residual misidentified μ^+ as K^+ are present in the sample after applying PID requirements.
³⁴² Their corresponding m_{corr} contribution is treated in Section 9. However a specific category
³⁴³ of such misIDed events, $J/\psi \rightarrow \mu\mu$ decays where one muon is misidentified as a kaon, can
³⁴⁴ be vetoed easily. Figure 17 shows the $K\mu$ invariant mass (or the B_s^0 visible mass) where

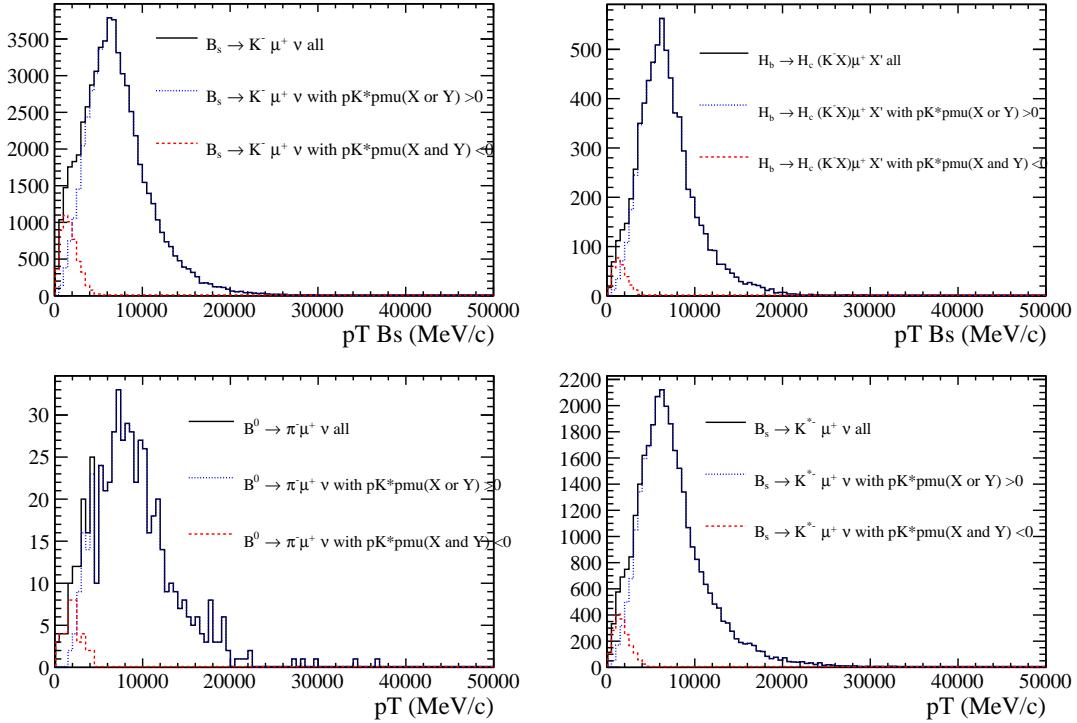


Figure 16: p_T distribution of the $B_s^0 \rightarrow K^+\mu^-$ candidates for different simulated physics samples.

345 the mass of the kaon is forced to be the muon mass and the isMuon criterion required
 346 to be true, along with the m_{corr} distribution for events falling into the J/ψ mass window
 (3072 $< m(\mu K \rightarrow \mu) < 3130$ MeV/ c^2). The events falling in that window are vetoed.

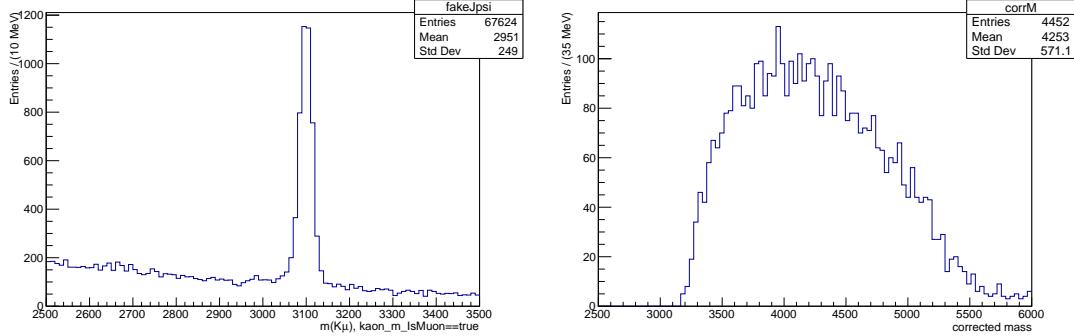


Figure 17: Left: $m(K\mu)$ invariant mass distribution where the kaon is required to pass the isMuon criterion and its mass hypothesis is forced to the muon mass. Right: m_{corr} distribution corresponding to the events in the mass window $3072 < m(\mu K \rightarrow \mu) < 3130$ MeV/ c^2

347

348 5.3.4 Charged isolation variable:

349 The use of such variable was proved useful in various analysis at LHCb [28–30]. The
 350 same training samples ($B^0 \rightarrow D^*\mu\nu$ and $B^+ \rightarrow D^*\mu\pi^+\nu$) are used here. As explained
 351 in [31] and [32]: The aim of the charged-track isolation is to verify that a signal track

352 is coming from the signal B vertex and not any other vertex. An event is regarded
 353 as a background if a signal track makes a "good" vertex with any non-signal track.
 354 In practical terms, underlying tracks are those particles found in StdAllNoPIDPions
 355 and StdAllNoPIDPionsUp containers after removing the signal tracks. This variable is
 356 important as most of our background is coming from Beauty decays with multiple charged
 357 tracks in the final states. The philosophy and training samples is the same as in [29]. Two
 358 main changes were done for this study:

- 359 • The definition of underlying tracks was extended to include particles in StdAllNoPID-
360 PionsUp on top of those in StdAllNoPIDPions container.
- 361 • The list of input variables for the classifier (BDTG) was shortened, Indeed, some of
362 those variables where relying on the existence of D^* , D^0 particles.

363 The following set of variables is considered for the training:

- 364 • minimum of impact parameter χ^2 of the track with respect to any PV
365 (track_minIPchi2)
- 366 • Pseudo-rapidity (η), azimuthal angle (ϕ) and Transverse momentum (p_T) of the
367 track
- 368 • Match- χ^2 per nDoF of the track (track_MatchChi2)
- 369 • Distance between the (signal track, track) vertex and the PV (track_pvdis_mu)
- 370 • Distance between the (signal track, track) vertex and the SV (tracksvidis_mu)
- 371 • Distance of closest approach between the signal track and the underlying track
- 372 • Angle between signal track(strack) and the underlying track(track) (track_angle_mu)
- 373 • $FC = \frac{(p_{strack}+p_{track}) \times \alpha^{strack+track,PV}}{(p_{strack}+p_{track}) \times \alpha^{strack+track,PV} + p_{T_{strack}} + p_{T_{track}}}$ where $\alpha^{strack+track,PV}$ is the angle
374 between the between the sum of momenta ($p_{strack} + p_{track}$) and the straight line
375 from the primary vertex to the vertex reconstructed using the signal track and the
376 track. (track_fc_mu)

377 The distributions of these variables are shown in Figure 18, while the output of the
378 classifier training is shown in Figure 19 along with its ROC curve.

379 The next step is to convert the BDT output from being attached to the underly-
380 ing track to being attached to the signal track. For this, the minimum BDT of the
381 pair (signal, underlying) tracks is chosen: this defines the variable IsoMinBDT, asso-
382 ciated to the signal track. Figure 20 shows the distribution of the minimum value of
383 ($\text{IsoMinBDT}(K^+)$, $\text{IsoMinBDT}(\mu^-)$) for the signal, different sources of backgrounds and SS
384 data. It clearly shows the separation power of such variable for exclusive backgrounds like
385 $B^+ \rightarrow J/\psi K^+$ or $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$ as the IsoMinBDT distribution for these backgrounds
386 exhibits a sharp peak at low values. Those correspond to events where the isolation
387 algorithm finds a track pair which is compatible with B vertex.

388 5.3.5 Summary of the offline selection

389 The offline requirements described in the previous subsections are summarized in Table 14.
 390 The effect of the $\sigma_{m_{\text{corr}}}$ cut on the m_{corr} distribution is illustrated in Fig.21.

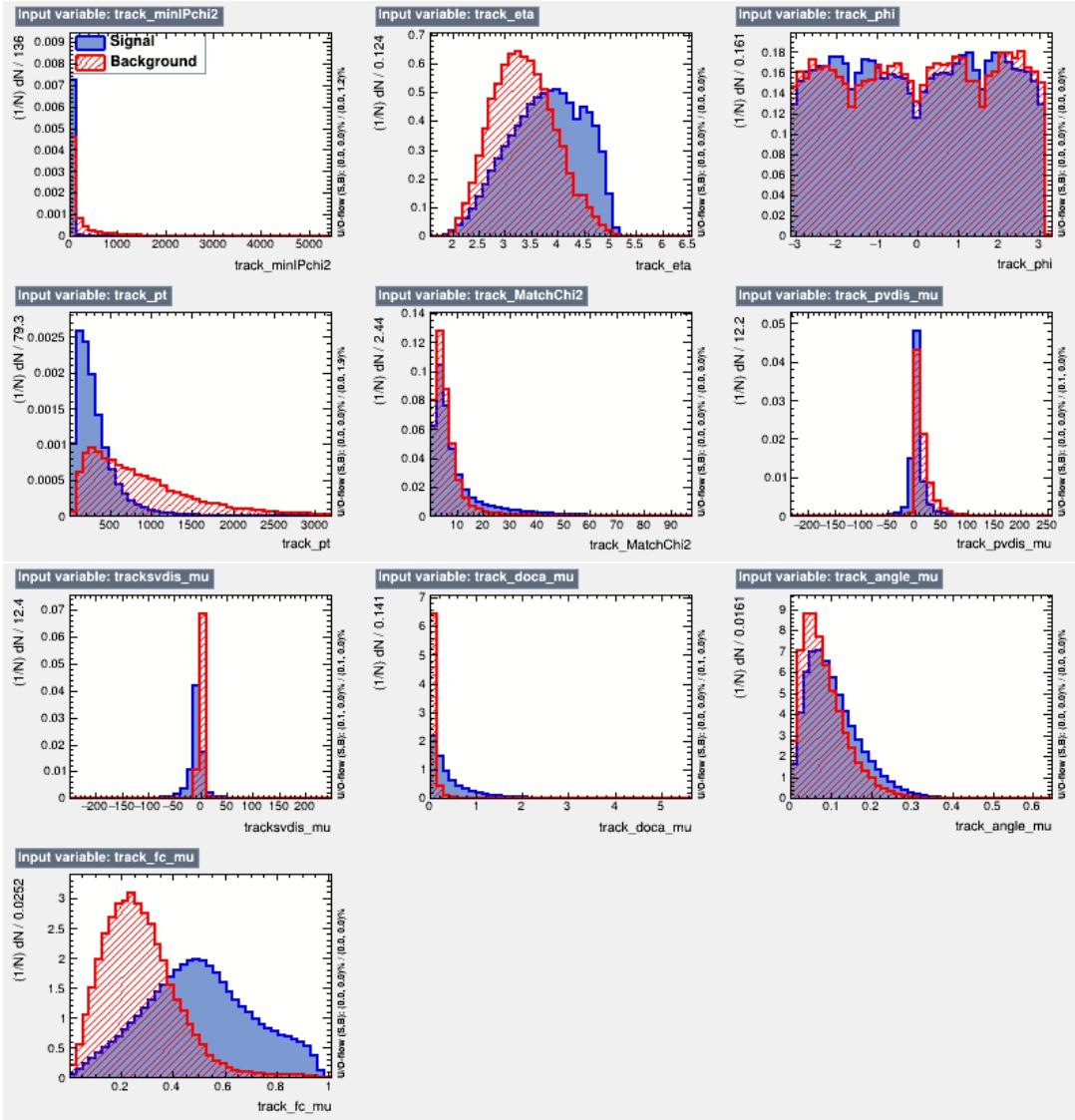


Figure 18: Input variables for the BDT training of the isolation variable. Plain blue is for signal (in blue) and hatched red is for background.

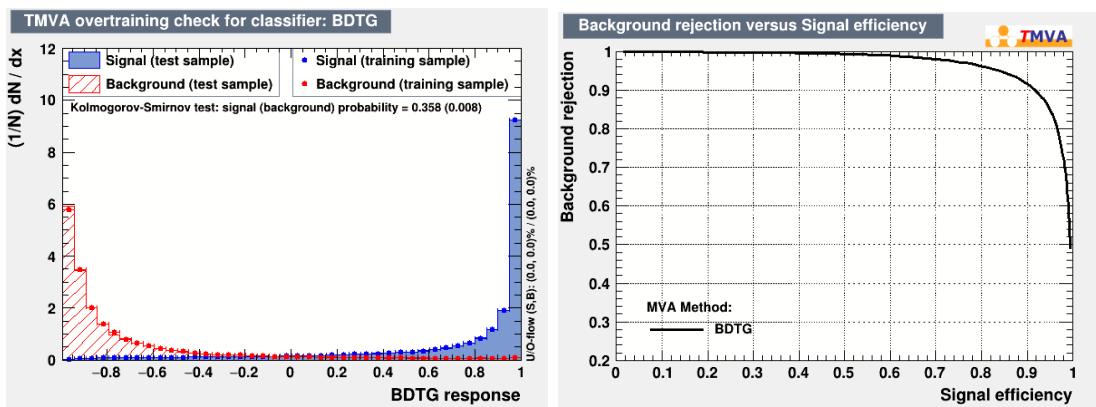


Figure 19: Right: BDTG Output distribution for signal (in blue) and background in (red). Left: ROC curve corresponding to the BDTG output described.

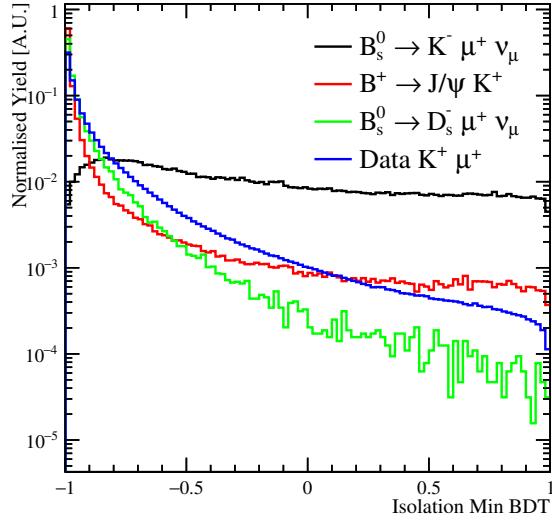


Figure 20: distribution of $\min(\text{IsoMinBDT}(K^+), \text{IsoMinBDT}(\mu^-))$ for $B_s^0 \rightarrow K^+ \mu^-$ (black), $B^+ \rightarrow J/\psi K^+$ (red) $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$ (green) and SS data (blue).

Variables	offline cuts
$ M(\pi^0)_{nIso} - 135 $	< 30 MeV
$\& M(K^+, \pi^0_{nIso}) - 892 $	> 60 MeV
$\& M(K^+, \pi^0_{nIso}) - 1430 $	> 90 MeV
K^+ IsMuon	= 0
$\& M(K^+_{m_{K^+}=m_{\mu^-}}, \mu^-) - 3096 $	> 40 MeV
$PX(K^+) \times P_X(\mu^-)$	> 0
$OR P_Y(K^+) \times P_Y(\mu^-)$	> 0
$\min(\text{IsoMinBDT}(K^+), \text{IsoMinBDT}(\mu^-))$	> -0.9
Event	
Nb. candidates = 1	
$\sigma_{m_{corr}}$	< 100 MeV/c ²

Table 14: Offline selection criteria for $B_s^0 \rightarrow K^+ \mu^- \nu$

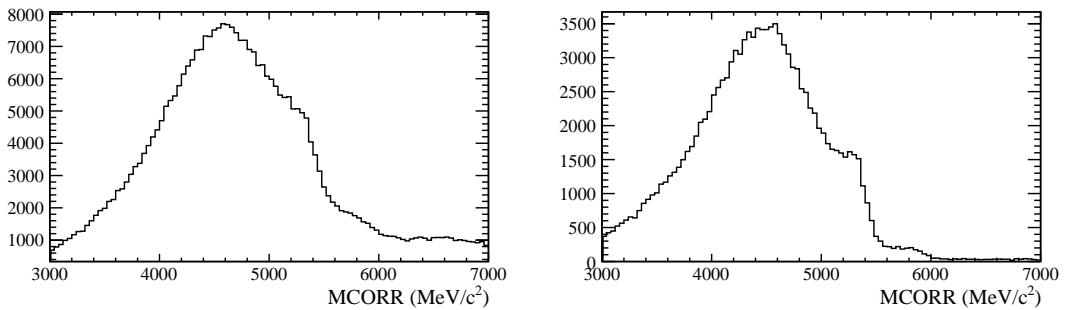


Figure 21: m_{corr} distribution (left) before the $\sigma_{m_{corr}}$ cut and (right) after.

391 6 MVA background removal

392 Two BDTs are trained to separate signal from background for this analysis. Both BDTs
393 use signal $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ decays as the signal sample during training. The background
394 sample used when training the first BDT is a cocktail mix of background Monte Carlo
395 decays reconstructed as $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ and the second BDT uses same sign (SS) data
396 reconstructed as $B_s^0 \rightarrow K^+ \mu^+ \nu_\mu$. The strategy is to apply a loose selection to the output
397 of the isolation BDT, `min(kaon_m_IsoMinBDT, muon_p_IsoMinBDT) > -0.9`, train the
398 first BDT to reduce backgrounds from partially reconstructed decays containing additional
399 charged tracks, apply a selection on the output of this BDT and train the second BDT to
400 provide additional discriminating power and reduce additional backgrounds seen in data
401 such as combinatorics and *feed-down* decays of higher excited particle states.

402 The first BDT, refereed to as the *charged BDT* is trained to separated signal Monte
403 Carlo from a cocktail of Monte Carlo backgrounds, the events used in the training are
404 detailed in Table 17. The charged BDT uses the variables listed in Table 15 to separate
405 signal from background, the corresponding distributions are detailed in appendix B.1.
406 All Monte Carlo samples used in the training have their kinematics corrected using the
407 reweighting procedure detailed in Section 7. During the training 850 trees with a maximum
408 depth of 3 levels and a minimum node size of 2.5% are trained using the AdaBoost boosting
409 method [33]. The effects of over training are removed through the use of 2 factor k -folding
410 with the data divide by magnet polarity for training and testing.

411 The second BDT is referred to as the same sign (SS) BDT since it is trained with
412 $K^- \mu^-$ data candidates as the background sample. The variables used in training the SS
413 BDT are listed in Table 16, the corresponding distributions are detailed in appendix B.2.
414 The SS BDT follows the same training procedure as the charged BDT, and a selection is
415 placed on the output of the charged BDT before training, thus minimising correlations in
416 training between the two BDTs.

417 The response of both BDTs with the selection are plotted in Figures 22 and 23 with
418 the corresponding Receiver operating characteristic (ROC) curves plotting the signal
419 efficiency against the background efficiency. A selection of `BDT_Charged > 0.05` is placed
420 on the charged BDT and `BDT_SS > 0` on the same sign BDT.

421 The BDTs are validated by comparing the BDT response in *sPlot* unfolded data with
422 the BDT response in Monte Carlo for $B^+ \rightarrow J/\psi K^+$ decays using a fully reconstructed
423 $K^+ \mu^- \mu^+$ triad or the $K^+ \mu^-$ pair. The validation using $B^+ \rightarrow J/\psi K^+$ is plotted in
424 Figure 24 and a slight discrepancy is seen in the BDT response between Monte Carlo
425 simulation and data for the three-prong reconstruction while the two prong reconstruction
426 is correctly behaved. A correction factor and systematic uncertainty are applied to the
427 calculated BDT efficiency using the $B^+ \rightarrow J/\psi K^+$ decay, the details of which are in
428 Section 7. The corrected three-prong BDT shapes can be seen in the corresponding
429 subsection (10.5) of the efficiency section, Fig.55.

Variable	Separation
Minimum isolation BDT response	1.90×10^{-1}
Invariant Mass, $K^- + T_{\text{Least Iso.}}$	1.00×10^{-1}
Isolation BDT summed over all underlying tracks	9.39×10^{-2}
B_s^0 Transverse Isolation	3.54×10^{-2}
Transverse isolation ⁶ between K^- and cone	3.00×10^{-2}
Kaon transverse momentum	2.84×10^{-2}
Transverse isolation between K^- and charged cone	2.58×10^{-2}
$p_T(B_s^0) - 1.5 \times p_T(\mu^+)$	2.17×10^{-2}
$p_T(B_s^0)$	1.94×10^{-2}
$\Delta\eta$ between K^- and cone	1.89×10^{-2}
Momentum asymmetry between μ^+ and cone	1.85×10^{-2}
K^- Isolation BDT response - μ^+ Isolation BDT response	1.63×10^{-2}
B_s^0 Decay vertex fit χ^2	1.11×10^{-2}
Invariant Mass, $\mu^+ + T_{\text{Least Iso.}}$	7.67×10^{-3}
B_s^0 helicity angle	2.03×10^{-3}

Table 15: The input variables used as inputs for the BDT trained to reject charged backgrounds are listed with their separating power. Several variables use information obtained from a cone draw around candidate tracks with $\Delta R = 0.5$. The least isolated track is referred to as $T_{\text{Least Iso.}}$.

Variable	Separation
$p_T(K^-)$	2.76×10^{-2}
DIRA $_{B_s^0}$	1.98×10^{-2}
Momentum asymmetry between K^- and neutral cone	1.36×10^{-2}
Transverse isolation between K^- and neutral cone	1.30×10^{-2}
Invariant mass K^- and additional π^0	1.02×10^{-2}
$p_T(B_s^0) - 1.5 \times p_T(\mu^+)$	9.31×10^{-3}
B_s^0 Flight distance significance	7.56×10^{-3}
B_s^0 transverse momentum	6.20×10^{-3}
B_s^0 helicity angle	5.96×10^{-3}
B_s^0 Decay vertex fit χ^2	3.68×10^{-4}

Table 16: The input variables used as inputs for the BDT trained to reject backgrounds found within the same sign sample. Several variables use information obtained from a cone draw around candidate tracks with $\Delta R = 0.5$.

Sample	Entries
$B_s^0 \rightarrow K^- \mu^+ \nu_\mu$	165k
Background Samples	
$b \rightarrow (c \rightarrow K\mu X)X$	47k
$b \rightarrow K^\pm \mu^\pm X$	190k
$b \rightarrow K\mu X$	550k
$B^0 \rightarrow J/\psi K^{*0}$	226k
$B^0 \rightarrow (D^* \rightarrow K\pi\pi\pi X)\mu\nu$	50k
$B^0 \rightarrow (D^* \rightarrow K\pi X)\mu\nu$	1210k
$B^0 \rightarrow (D^* \rightarrow K\pi)\mu\nu$	34k
$B^0 \rightarrow (D \rightarrow K\pi\pi)\mu\nu$	75k
$B^+ \rightarrow J/\psi K^+$	445k
$B^+ \rightarrow J/\psi K^{*+}$	135k
$B^+ \rightarrow D^0 \mu^+ \nu_\mu X$	230k
$B_s^0 \rightarrow J/\psi \phi$	2500k
$B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu X$	96k

Table 17: The simulated events used during the training of the BDT are summarised.

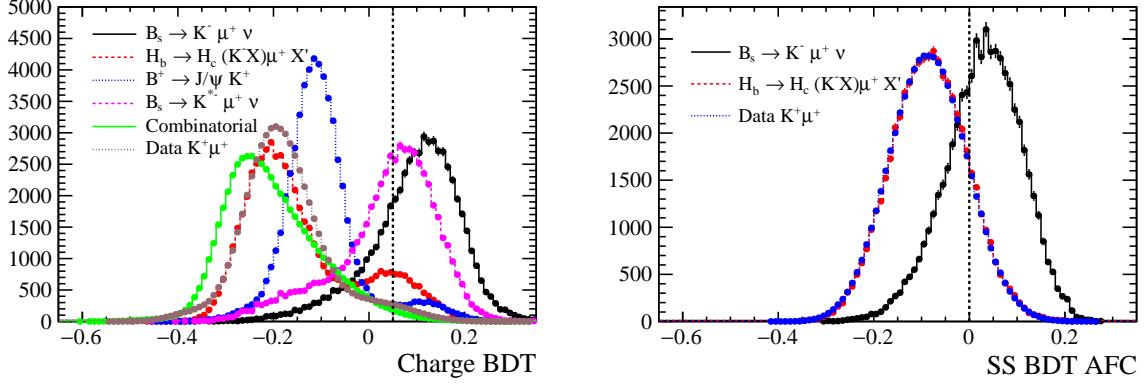


Figure 22: Response for the BDT trained to reject backgrounds (left) containing additional charged tracks and (right) within the same sign sample.

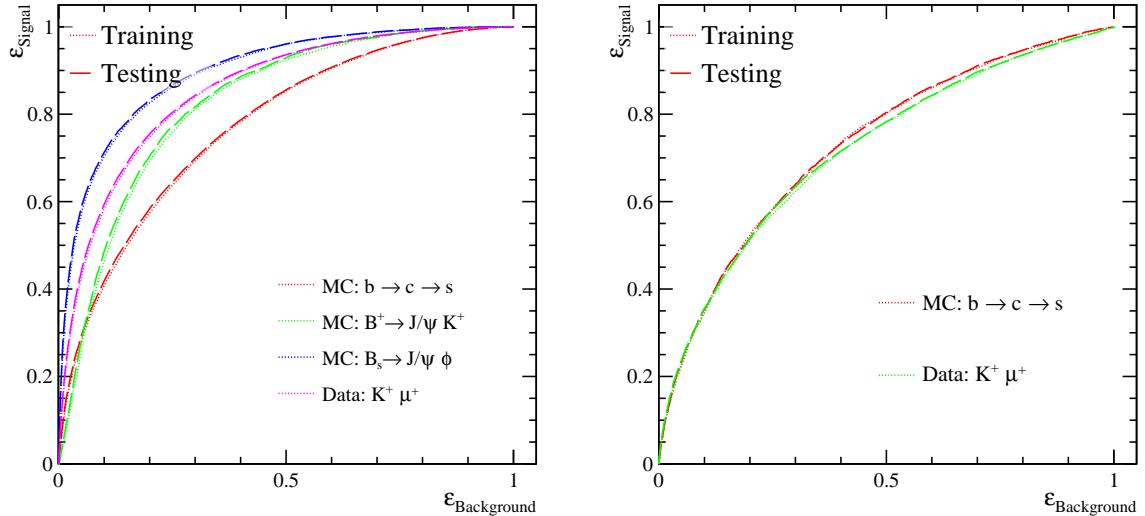


Figure 23: ROC curves for the BDT trained to reject backgrounds (left) containing additional charged tracks and (right) within the same sign sample.

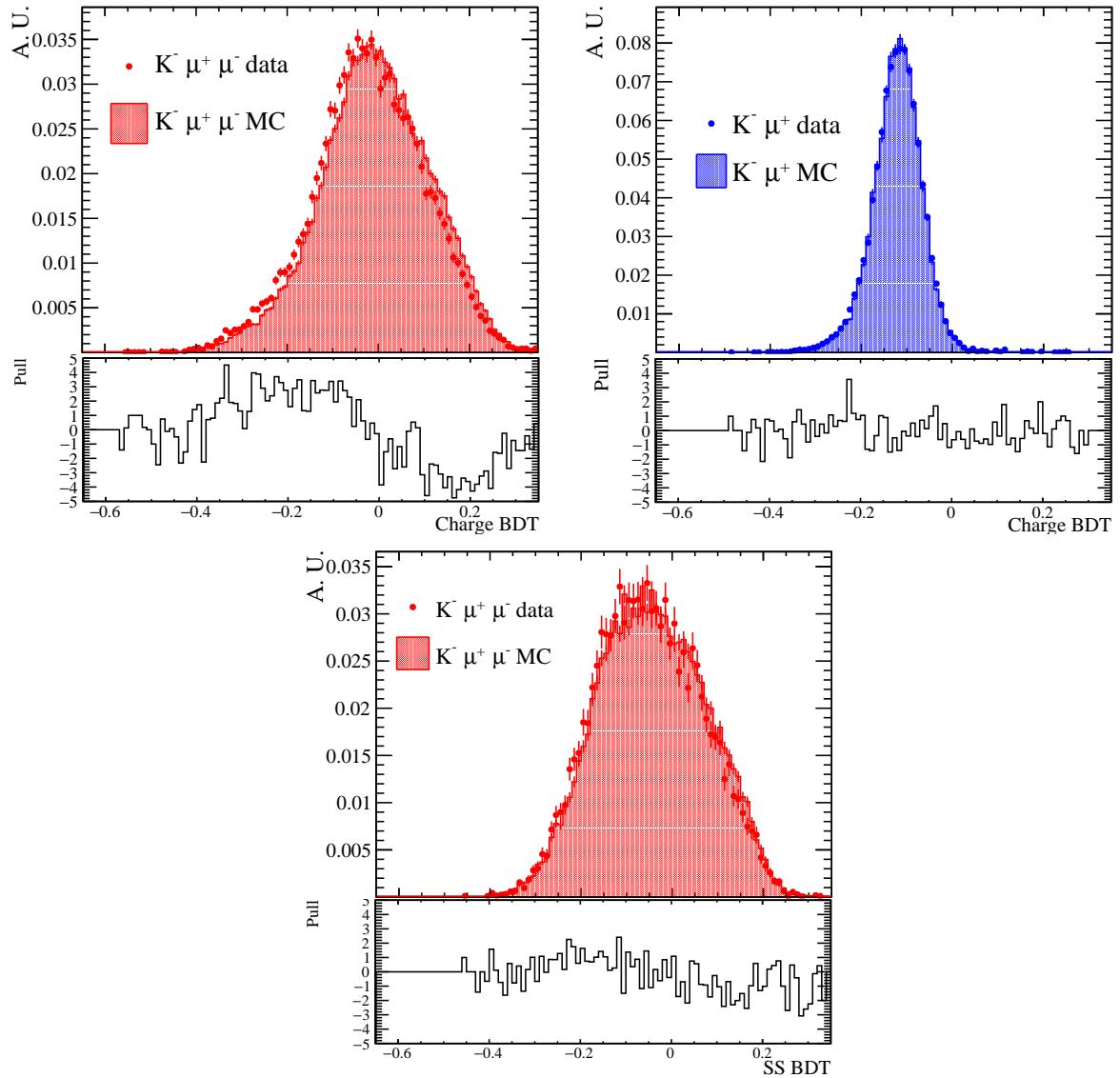


Figure 24: Data/MC BDT comparison using $B^+ \rightarrow J/\psi K^+$ decays. The blue histogram shows the decays reconstructed as $B \rightarrow K\mu$ while the red ones shows fully reconstructed $B^+ \rightarrow J/\psi (\mu\mu)K^+$. The fully reconstructed or three-prong shapes are later kinematically corrected to give appropriate matching, Fig.55.

430 7 Control and Normalization Channels corrections

431 The simulation of Monte Carlo data is not perfect and the distributions of several variables
 432 show disagreements between Monte Carlo and data. It is essential that differences between
 433 Monte Carlo and data are corrected to ensure that the distributions of variables used
 434 for fitting are correct, and that the determination of efficiencies are accurate. A simple
 435 approach to reweighting would be to plot the histogram of a variable for Monte Carlo
 436 and data, and assign each Monte Carlo event a weight corresponding to the ratio of data
 437 and Monte Carlo yields at that point. The simple method fails for multidimensional
 438 reweighting as a multidimensional histogram with granular bins will face problems due to
 439 low bin statistics and a histogram with coarse bins will produce biases due to variations
 440 within the bins. For this analysis a boosted decision tree (BDT) is trained to separate pure
 441 data and Monte Carlo. If the simulation is perfectly modelled, the BDT will return an
 442 output variable with no separating power⁷. If, however the simulation does not perfectly
 443 model the data the BDT will return an output variable with a significant separating power.
 444 The driving assumption behind this method for correcting the simulation is that the
 445 output variable of the BDT will completely encapsulate all Monte Carlo/data differences
 446 in a single discriminating variable. By performing the simple one dimensional correction
 447 on the BDT output it is possible to correct all the variables used in training the BDT
 448 simultaneously [34].

449 In order to perform an effective comparison between simulated Monte Carlo and
 450 Data a pure, signal only data sample is needed. The decay $B^+ \rightarrow J/\psi K^+$ is used
 451 to correct $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$, data and simulation are reconstructed as $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ and
 452 the candidate B_s^0 is combined with a muon found using the isolation BDT detailed in
 453 Section 5.3.4 allowing the B^+ candidate to be constructed. An *sPlot* [35] background
 454 subtraction is performed on both the data and simulated samples using the reconstructed
 455 B^+ mass peak, as illustrated in Fig.25 left (for data).

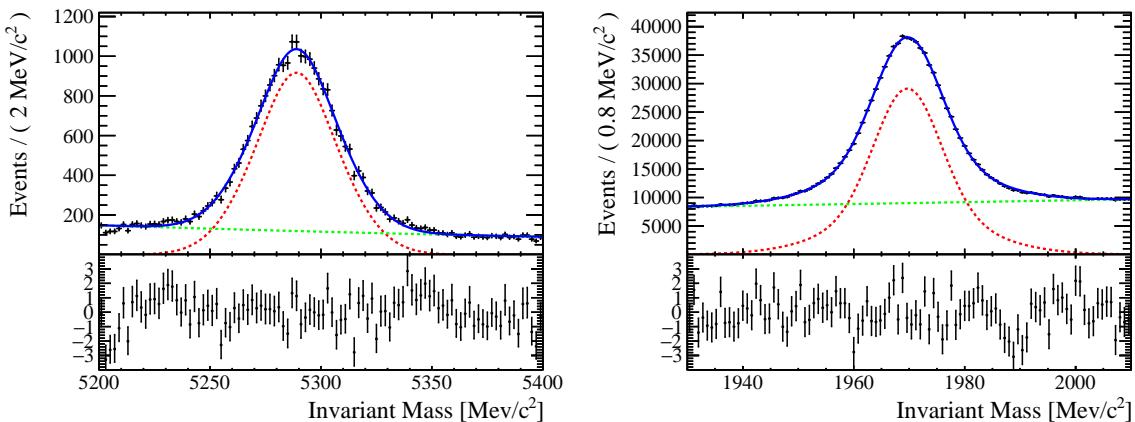


Figure 25: Fits used to generate the sWeights for (left) $B^+ \rightarrow J/\psi K^+$ (invariant mass $K\mu\mu$) and (right) $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu X$ (invariant mass $K^+ K^- \pi$).

456 When correcting the $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$ simulation, a simulated cocktail of decays
 457 $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu X$ is compared to data containing a well reconstructed D_s^+ in associa-
 458 tion with a muon. Backgrounds are reduced by applying a selection on the output of

⁷Over training effects and statistical fluctuations will return some *false* separating power.

459 the isolation BDT and an *sPlot* background subtraction is performed on the D_s^+ mass
 460 peak as shown in Fig.25 right (for data). For both the signal and normalisation modes a
 461 BDT containing 200 trees with a maximum depth of 3 and a minimum leaf size of 6% for
 462 $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$ and 4% for $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ is trained to separate simulated Monte Carlo
 463 and data. For the training and evaluation a $k = 2$ k -folding is used. The BDT is trained
 464 to separate, and hence correct the following variables in the simulation:

465

$B_s^0 \rightarrow K^- \mu^+ \nu_\mu$	$B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$
Track Multiplicity	Track Multiplicity
ηB_s^0	ηB_s^0
$p_T B_s^0$	$p_T B_s^0$
$p_T K^-$	$p_T D_s^-$
$p_T \mu^+$	$p_T \mu^+$

466 Figure 26 shows the BDT responses for Data/MC comparison and the corresponding
 467 correction weights.

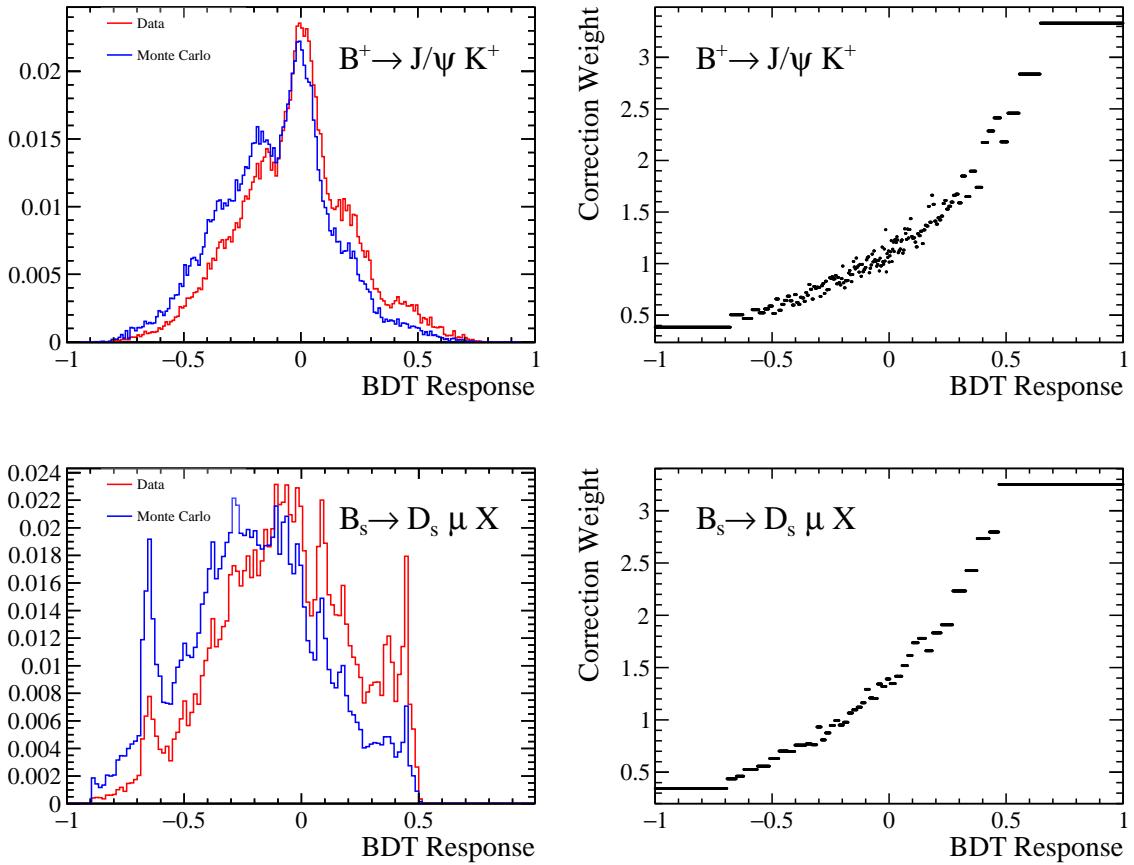


Figure 26: The BDT response used to separate Monte Carlo and data (left) and the weights used to correct the simulation (right) for (top) $B^+ \rightarrow J/\psi K^+$ and (bottom) $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu X$. The $J/\psi K^+$ is reconstructed using only the $K^+ \mu^-$ pair.

468 Figures 27 and 28 show the effect of the BDT reweighting on the $B^+ \rightarrow J/\psi K^+$ and
 469 $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu X$ simulations.

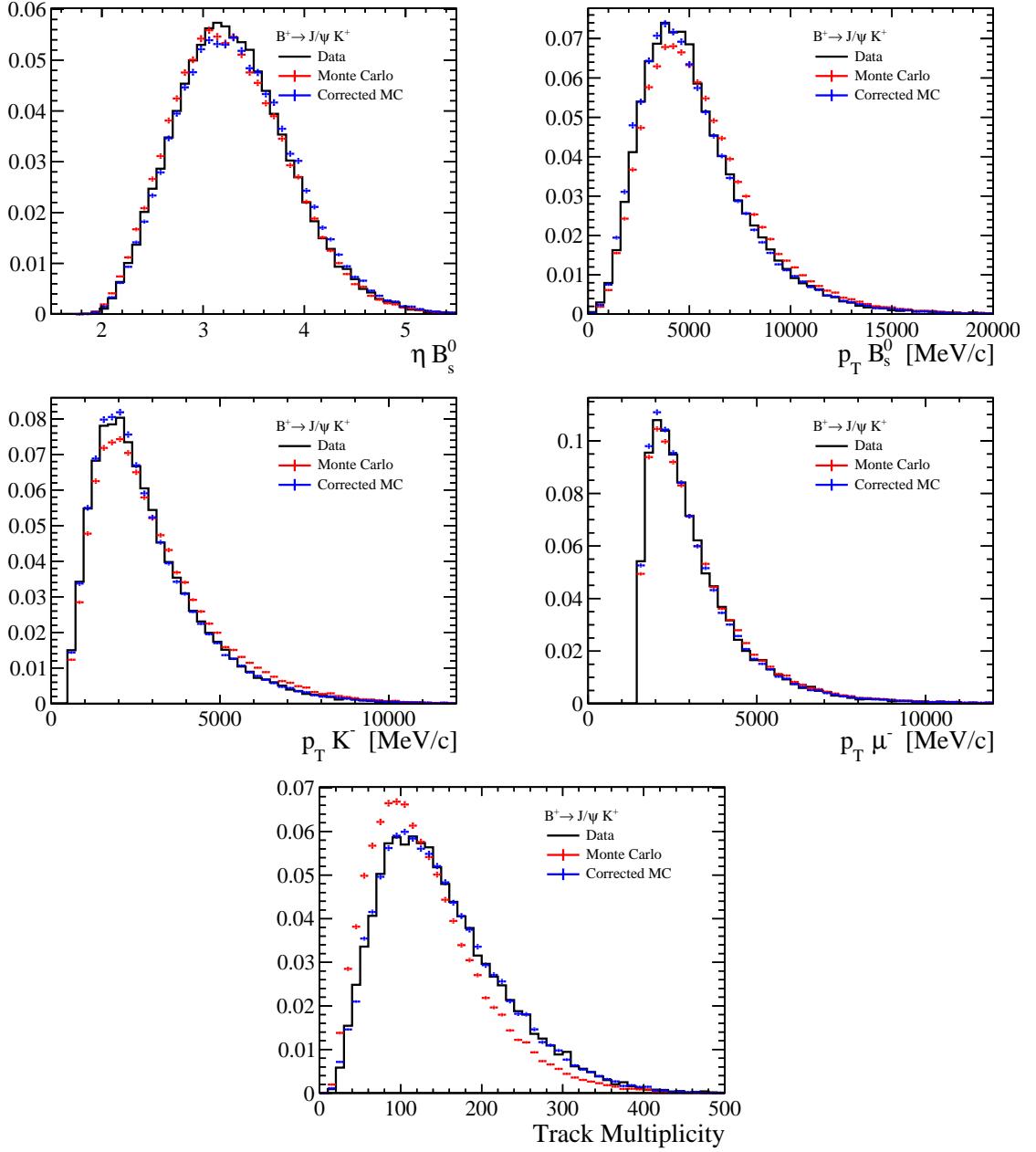


Figure 27: Kinematic distributions of $B^+ \rightarrow J/\psi K^+$ reconstructed using the $K^+\mu^-$ pair are plotted for data and simulation before and after correction.

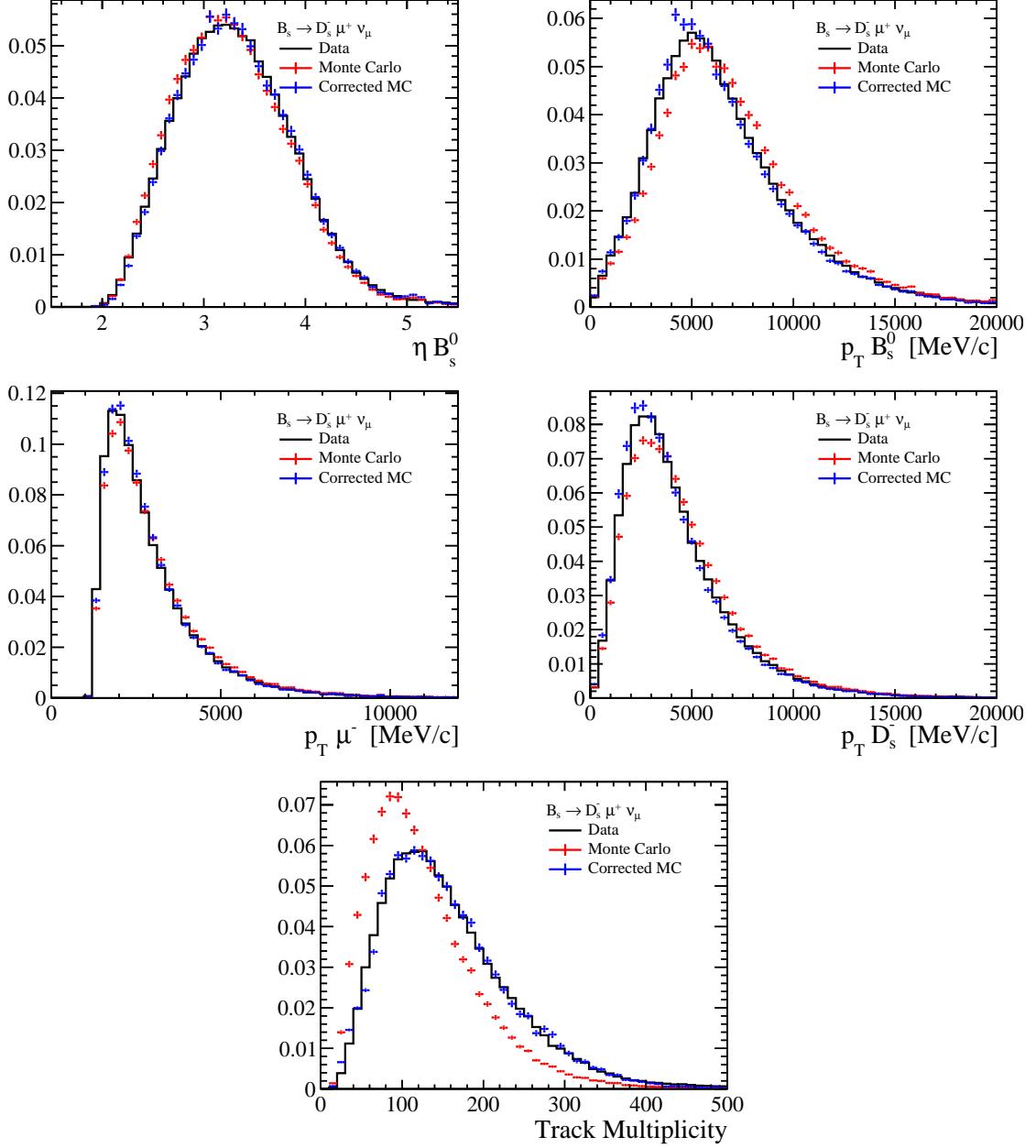


Figure 28: Kinematic distributions of $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu X$ before and after correction.

470 8 Normalisation Fit

471 8.1 Normalisation Fit Model

472 A maximum likelihood, Beeston-Barlow binned template fit is performed on the corrected
 473 $D_s^- \mu^+$ mass distribution to extract the $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$ signal yield. A bias free background
 474 subtraction is performed using the $K^- K^+ \pi^-$ invariant mass distribution to remove the
 475 $K^- K^+ \pi^-$ combinatoric contribution. The yield in each bin of the corrected $D_s^- \mu^+$
 476 histogram is the result of a fit to the $K^- K^+ \mu^-$ invariant mass to determine the D_s^+
 477 yield. The fit to the corrected $D_s^- \mu^+$ mass distribution is used to separate the signal
 478 $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$ signal component from the background contributions. The signal shape
 479 is determined from the last available FF prediction, from McLean *et al.* as presented
 480 in section 2.1.2. The backgrounds predominantly originate from semileptonic B_s^0 decays
 481 containing higher excited D_s^- resonances. Backgrounds consisting of partially reconstructed
 482 $B \rightarrow D_s^+ DX$ candidates and tauonic decays are considered, as are candidates containing
 483 misidentified muons. The background from misidentified μ ⁸ is studied using the stripping
 484 line `B2DMuNuX_Dp_FakeMuon`, the method is identical to that explained in 9.2, the
 485 contribution from such background is found to be approximately of the order of 3 per
 486 mil of the signal yield, its yield is constrained in the fit. Random combinations of real
 487 muons with real D_s^+ mesons are neglected: when investigating the $K^- K^+ \pi^-$ invariant
 488 mass distribution using the same sign, $D_s^- \mu^-$ sample no D_s^- peak is seen. The same sign
 489 sample may be assumed to be purely combinatoric as very few decays contain a same
 490 sign D_s^- and muon. The fit components and the sources of templates used in the fit are
 491 summarised in Table 18.

492 The templates used in the fit contain 40 bins in corrected mass ranging from 3000 MeV
 493 to 6500 MeV with an equal bin width. Backgrounds having similar shapes are combined
 494 into a common template; $B_s^0(B^+B^0) \rightarrow D_s^+ \tau^+ \nu_\mu(D_s^+ DX)$ are combined into a single
 495 template, the higher excitations of the D_s^+ above the D_s^{*+} are combined into along with
 496 $B_s^0 \rightarrow D_s^+ DX$ a single category. The combinations are plotted in Figure 29. The fit to
 497 data is performed with templates reweighted for kinematic (BDT), PID and tracking
 498 corrections. Besides, all component yields are left free.

Component	Source
$B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$	Monte Carlo
$B_s^0 \rightarrow D_s^{*-} \mu^+ \nu_\mu$, with $D_s^{*-} \rightarrow D_s^- \gamma$	Monte Carlo
$B_s^0 \rightarrow D_s^{*-} \mu^+ \nu_\mu$, with Higher D_s^+ Excitations	Monte Carlo
$B_s^0 \rightarrow D_s^- \tau^+ \nu_\mu X$, with $\tau^+ \rightarrow \mu^+ \nu_\mu \bar{\nu}_\tau$	Monte Carlo
$B_q^0 \rightarrow D_s^- D_q^{(*)} X$, with $D_q \rightarrow \mu^+ \nu_\mu X$	Monte Carlo
Misidentified Muons	Fake Muon Data

Table 18: Fit components for the normalisation fit and sources of data used when drawing templates.

⁸misidentified D_s^- is excluded as we subtract out none- D_s^- background

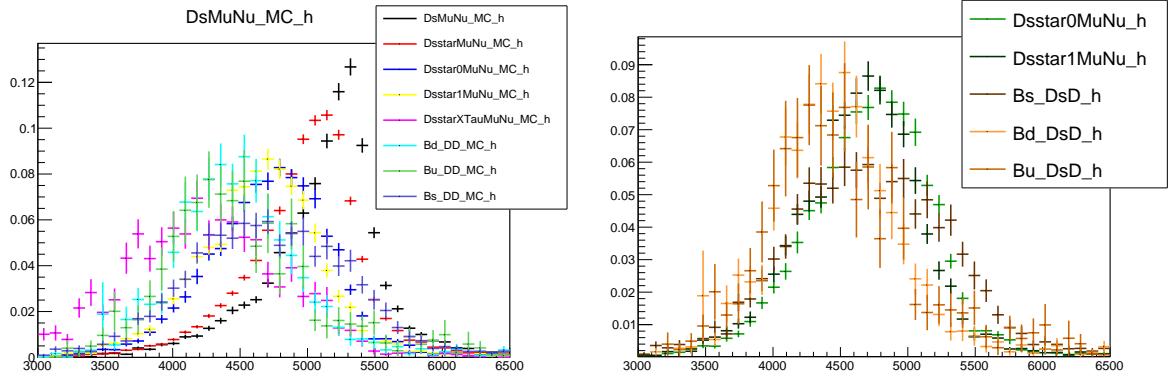


Figure 29: (Left) Individual templates from MC in the $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$ fit. (Right) Similar templates from $B_s^0 \rightarrow D_s^{*-} \mu^+ \nu_\mu$, with higher D_s^+ excitations and $B_q^0 \rightarrow D_s^- D_q^{(*)} X$, with $D_q \rightarrow \mu^+ \nu_\mu X$.

499 8.2 Background Subtraction

500 Correlations between the $K^+ K^- \pi^+$ invariant mass and the $D_s^- \mu^+$ corrected mass mean
 501 that the *sPlot* method for subtracting backgrounds cannot be used. Instead a *divide and*
 502 *fit* method is used whereby the data is divided into n smaller subsets, each corresponding
 503 to a specific bin in the $D_s^- \mu^+$ corrected mass spectrum. A binned maximum likelihood
 504 fit is performed to the $K^+ K^- \pi^-$ invariant mass distribution for each dataset. A double-
 505 Gaussian with free parameters models the D_s^- shape and an exponential models the
 506 combinatoric background shape. The $D_s^- \mu^+$ yield in the corrected mass histogram for
 507 each bin is set as the signal yield from the fit. The fits from the *divide and fit* method are
 508 plotted in Figure 30 while the resulting corrected mass distribution is shown in Figure 31
 509 (n.b: no background subtraction is required for the Monte Carlo samples).

510 8.3 Fit Results

Sample	Yield / 10^3
$B_s^0 \rightarrow D_s^+ \mu^+ \nu_\mu$	201.5 ± 5.2
$B_s^0 \rightarrow D_s^{*-} \mu^+ \nu_\mu$	349.0 ± 7.7
$B_s^0 \rightarrow D_s^{**-} \mu^+ \nu_\mu (D_s^+ DX)$	32.2 ± 5.7
$B_s^0 (B^+ B^0) \rightarrow D_s^+ \tau^+ \nu_\mu (D_s^+ DX)$	45.2 ± 2.5

Table 19: Fit results for all components of the fit used to obtain the $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$ yield.

511 The results of the maximum likelihood fit to the $D_s^+ \mu^-$ corrected mass are plotted in
 512 Figure 31, fitting to all events passing the signal selection. The signal and background
 513 yields obtained from the fit are provided in Table 19, the correlation matrix between these
 514 yields is shown in Table 20. The results of the $D_s^- \mu^+$ fit are validated by performing 1000
 515 fits to pseudo-data. The data template in each pseudo-data fit is replaced with a *toy*
 516 *template* generated by randomly selecting points from the fit templates. Consequently the
 517 yields of each fit component are known precisely. The fit templates used in the fits to
 518 pseudo-data are *statistically compatible copies*, i.e. the contents of each bin is replaced by

519 a random number sampled from a Gaussian distribution centred on the bin contents with
 520 width equal to the bin uncertainty. The $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$ yield in the pseudo-data is fixed
 521 to 191.0×10^3 while the yields of all backgrounds are chosen by randomly sampling a
 522 Gaussian distribution centred on the yield determined from the fit and a width set to the
 523 component's uncertainty. The mean and width of the Gaussians are taken from Table 19.
 524 The distribution of the $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$ yield for all 1000 fits to the pseudo-data is plotted
 525 in Figure 32 alongside the pull distribution. The pull is defined as $(N_{\text{Fit}} - N_{\text{In}})/\sigma_{\text{Fit}}$.
 526 where N_{Fit} and σ_{Fit} are the yield and uncertainty obtained from the fit to pseudo-data
 527 and N_{In} is the true number of $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$ events in the pseudo-data. The pulls should
 528 be centred at 0 and follow a Gaussian distribution with a width of 1. As seen in Figure 32
 529 the pull distribution of the toy fits is well fit by a Gaussian which is well centered around
 530 zero but has a width < 1 : this is due to the impact of the inclusion of the statistical
 531 uncertainties of the various templates involved in the fit. Indeed, when turning off the
 532 statistical uncertainty in the fit, a pull width value of 1 is recovered.

	Param.	1	2	3	4	5
$B_s^0 \rightarrow D_s^+ \mu^+ \nu_\mu$	1	1	-0.858	0.461	-0.263	-0.002
$B_s^0 \rightarrow D_s^{*-} \mu^+ \nu_\mu$	2		1	-0.769	0.471	-0.004
$B_s^0 \rightarrow D_s^{**-} \mu^+ \nu_\mu (D_s^+ DX)$	3			1	-0.819	-0.001
$B_s^0 (B^+ B^0) D_s^+ \tau^+ \nu_\mu (D_s^+ DX)$	4				1	0.001
MisPID	5					1

Table 20: Fit correlation coefficients obtained for all free yields in the fit.

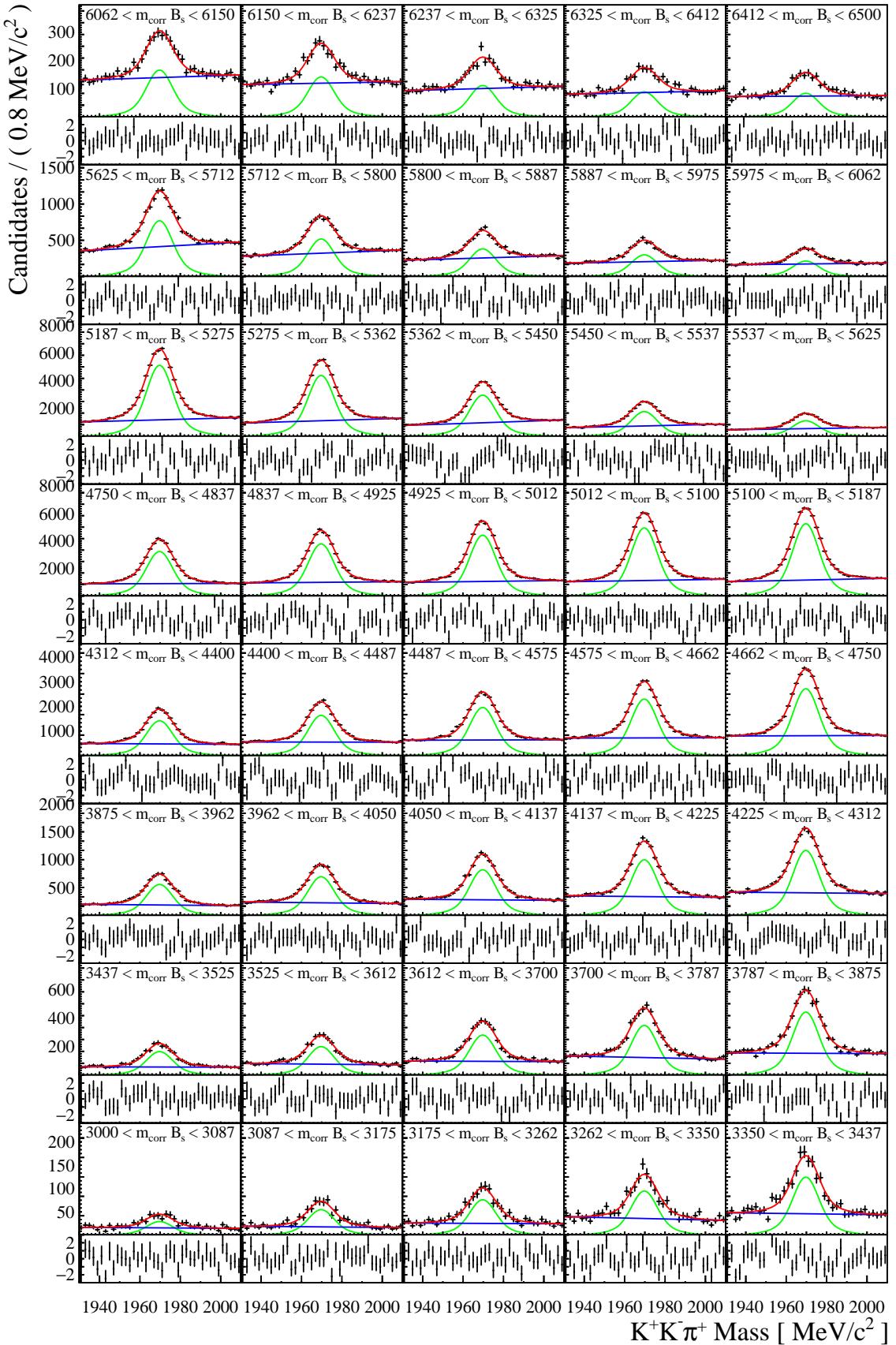


Figure 30: Fits performed as part of a combinatoric background subtraction on the $K^+K^-\pi^+$ invariant mass with pulls underneath.

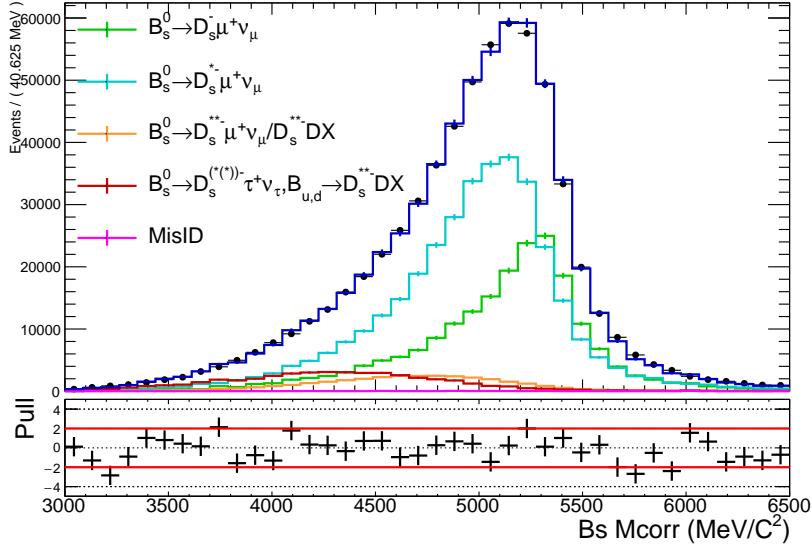


Figure 31: A fit to the $D_s^- \mu^+$ corrected mass for candidates in data passing the selections.

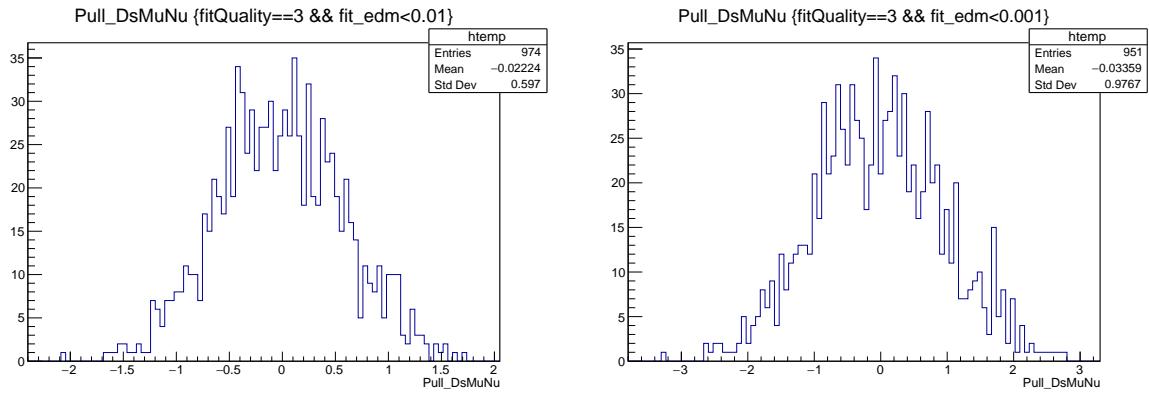


Figure 32: Distributions of pulls from fits to 1000 pseudo datasets. The input $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$ yield is fixed for each toy fit and the variation in results is due to statistical uncertainties in the fit, for (left) the case where the uncertainties in the templates is included or (right) excluded.

533 9 Signal fit

534 A fit to the m_{corr} distribution of the $K^-\mu^+$ data is performed in the range $2500 \text{ MeV}/c^2 <$
 535 $m_{\text{corr}} < 5700 \text{ MeV}/c^2$, simultaneously in two bins of q^2 , with the bin boundary placed at
 536 $q^2 = 7 \text{ GeV}^2/c^4$ in order to have approximately the same $B_s^0 \rightarrow K^-\mu^+\nu_\mu$ yield in each bin.
 537 For each fit component, the yields in the two bins, $N_{q^2 < 7}$ and $N_{q^2 > 7}$, are treated in a specific
 538 way explained in the next subsections. As in the normalization fit 8, the fit is implemented
 539 using the HistFactory package in ROOT [36]. The package is based on the Beeston-Barlow
 540 likelihood method which accounts correctly for the statistical uncertainties of various
 541 MC templates [37]. The dominant backgrounds in the fits to extract the $B_s^0 \rightarrow K^-\mu^+\nu_\mu$
 542 yields include $b \rightarrow c$ decays, decays from the excited K^* resonances, decays to charmonium,
 543 combinatoric combination of a kaon and muon, and candidates containing misidentified
 544 kaons. The most concerning background are the partially reconstructed decays involving
 545 a charmonium, $B \rightarrow c\bar{c}(\mu^+\mu^-)K^+X$, dominated by $B^+ \rightarrow J/\psi K^+$, the latter having a
 546 m_{corr} distribution close to signal. The yields and/or shapes of many backgrounds may be
 547 constrained using data driven methods. The different ways of yield constraining will be
 548 described in the coming subsections. The shapes inferred by data-driven methods, for the
 549 combinatorial and misidentified backgrounds, are explained in subsections 9.1 and 9.2.

550 The components and data sources used to generate fit templates are summarized in
 551 Table 21. Events originating from Simulated Monte Carlo decays are weighted to correct
 552 for differences in kinematics and mismodelling between the simulation and data.

	Component	Source
Signal	$B_s^0 \rightarrow K^-\mu^+\nu_\mu$	Monte Carlo
	$B_s^0 \rightarrow K^*-\mu^+\nu_\mu$	Monte Carlo
	$B \rightarrow c\bar{c}(\mu^+\mu^-)K^+X$	Monte Carlo
	Combinatorics	<i>Mixing</i> of K^- and μ^+ from different events (data driven)
	Misidentified particles	Signal candidates in data failing PID selections (data driven)
	$b \rightarrow c \rightarrow s$	Monte Carlo

Table 21: Sources of data used to generate the corrected mass histograms for each $B_s^0 \rightarrow K^-\mu^+\nu_\mu$ fit component.

553 9.1 Combinatoric Modeling

554 Though the selection described in subsection 5.3.2 removes a big part of the combinatorial
 555 background, the remainder still needs to be modeled. Due to the nature of a partially
 556 reconstructed decay there is no calibration sample from which a representative sample
 557 of combinatoric combinations of kaons and muons may be obtained, although there are
 558 *corners* in phase space where a pure sample of combinatorics may be obtained. By
 559 requiring the $m_{K^-\mu^+} > m_{B_s^0}$, a pure combinatoric sample may be produced however the
 560 reconstructed corrected mass will be larger than the invariant mass making the sample
 561 irrelevant when considering the full phase space sample. For this analysis combinatorics
 562 are modeled using event mixing of an independent sample (2011 data) selected with the
 563 same requirements, a candidate kaon track from one event is combined with a candidate
 564 muon from another event and a new B_s^0 is reconstructed from this combination. The

effectiveness of this method is validated by comparing the kinematics of the mixed events with candidates in data for $m_{K^-\mu^+} > m_{B_s^0}$ where lies a pure combinatoric sample.

To construct a new combinatoric candidate from existing candidates each candidate is split into quarters containing the saved information on the B_s^0 , K^- , μ^+ and information on the underling event. A new combinatoric candidate is created by *stitching* different sub events together into a new event from which a combinatoric B_s^0 candidate is reconstructed by combining the muon and kaon from different events recalculating its kinematics. The flight distance of the new B_s^0 is chosen by randomly sampling the flight distance of reconstructed B_s^0 candidates with $m_{K^-\mu^+} > m_{B_s^0}$ and the uncertainty on the corrected $K^-\mu^+$ mass is calculated using a toy study; the secondary vertex location is randomly varied within its quoted uncertainty with a new value of the corrected mass calculated each time. The uncertainty on the corrected mass is the rms of the corrected mass toys.

When reconstructing combinatoric candidates a cyclic approach is used with each event providing information that will be used in the generation of five candidates. The cyclic approach is visualized in Figure 33. After generation of the candidates they are passed through a loose selection mimicking the preselection detailed in subsections 5.2, 5.1.

Event #	n-1	n	n+1	n+2	n+3	n+4	n+5	n+6	n+8	n+9	n+10	n+12	n+15
Combi. 1	B_s^0	μ^+	K^-	Ev.									
Combi. 2	B_s^0	μ^+		K^-			Ev.						
Combi. 3	B_s^0		μ^+		K^-					Ev.			
Combi. 4	B_s^0			μ^+		K^-					Ev.		
Combi. 5	B_s^0				μ^+				K^-			Ev.	

Figure 33: Each event is used to generate five combinatoric combinations. For the first combination consecutive events are mixed, for the second combination alternate events are used.

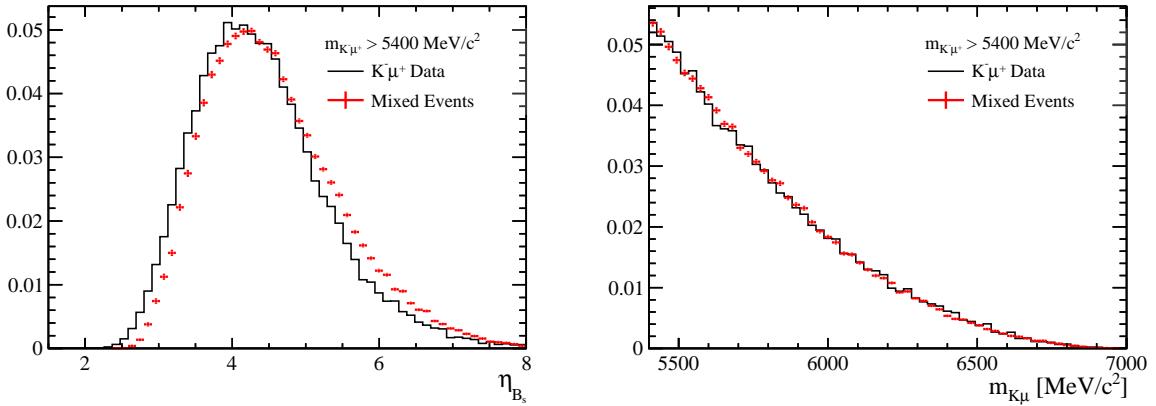


Figure 34: The distributions of $K^-\mu^+$ candidates in data (black) are plotted alongside simulated combinatoric candidates (red). All distributions are drawn requiring that the reconstructed $K^-\mu^+$ mass is greater than $m_{B_s^0}$.

This method of modeling combinatorics doesn't accurately reproduce the kinematics of the B_s^0 meson and a two dimensional reweighting is used to correct the momentum and transverse momentum of the B_s^0 candidate. The plotted distributions of the $K^-\mu^+$

585 invariant mass and corrected mass are invariant under a correction of the kinematics of
 586 the B_s^0 . The details of the reweighting procedure and validation plots may be found in
 587 Appendix C.

588 The invariant mass and corrected mass of the $\mu^+ K^-$ pair obtained from the event
 589 mixing procedure is plotted in Figure 35.

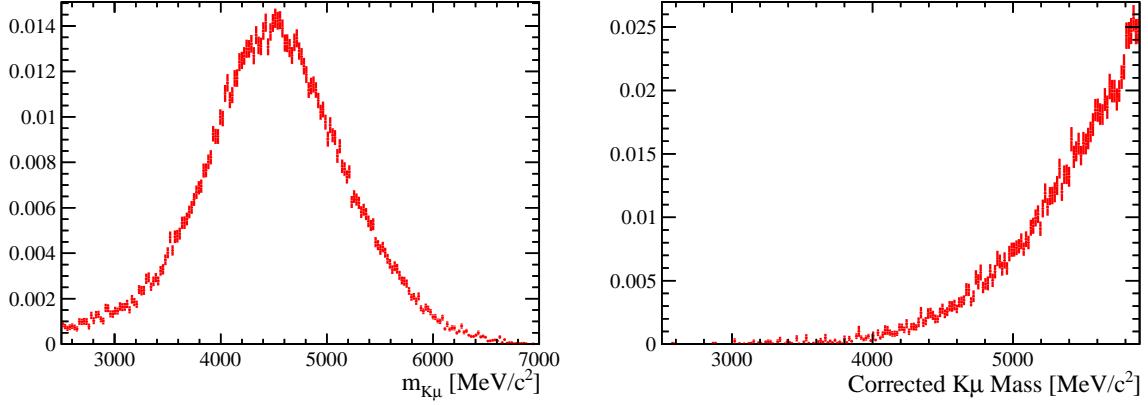


Figure 35: The invariant mass and corrected mass of a $K^- \mu^+$ pair is plotted for candidates modeling combinatoric backgrounds.

590 9.2 Mis-identification background

591 In this section we measure the contribution of the background involving particles that are
 592 mis-identified as kaon or muon (and to a much lesser extent, both) and provide its shape
 593 using a data-driven technique. Our approach to account for such background is similar to
 594 the one adopted in the $\mathcal{R}(J/\psi)$ analysis [38], based on Bayes unfolding [39]. An alternative
 595 method, based on SVD decomposition [40], is also used to evaluate the systematics due
 596 to unfolding. The expected contribution from this background is estimated to be $\sim 5\%$
 597 of our sample ⁹, this is less than the one featured in $\Lambda_b^0 \rightarrow p \mu^-$ [41], due to a better K^+
 598 identification compared to the proton identification in LHCb.

599 For our analysis the mis-identification background is composed of two components:
 600 the first is coming from particles misidentified as K^+ (dominant contribution), the sec-
 601 ond is coming from particles misidentified as μ^- (much smaller contribution). The
 602 case where a double misidentification of K^+ and μ^- in the same event is considered to
 603 be negligible. The fake K and μ stripping lines `StrippingB2XuMuNuBs2K_FakeKLine`,
 604 `StrippingB2XuMuNuBs2K_FakeMuLine`¹⁰ are used to create enriched regions of the follow-
 605 ing particles types: π , μ , p (for the K^+ misid component) and π , K , p (for the μ^- misid
 606 component), Table 22 shows the PID cuts used to define such regions.

607 The estimated yield of misidentified events due to $\zeta = K, \mu$ in the signal region from a
 608 given particle type i (e.g π) would be:

$$609 \quad \mathcal{Y}_i(\zeta) = N_i \times \frac{P(\hat{\zeta}|i)}{P(\hat{i}|i)} \quad (28)$$

⁹this is based on naive values of the misID rates on K^+ and μ^- at LHCb.

¹⁰where the analysis PID requirements are not applied on the K^+ and μ^- respectively

Region	Cut applied (FakeKLine)
Signal	$K_PIDK > 5 \& (K_PIDK - K_PIDp) > 5 \& (K_PIDK - K_PIDmu) > 5$
π	$K_PIDK < 0 \& (K_PIDK - K_PIDp) < 0 \& (K_PIDK - K_PIDmu) < 0$
p	$K_PIDK < 5 \& (K_PIDK - K_PIDp) > 0 \& (K_PIDK - K_PIDmu) < 0$
μ	$K_PIDK < 0 \& (K_PIDK - K_PIDp) < 0 \& (K_PIDK - K_PIDmu) > 0$
Region	Cut applied (FakeMuLine)
Signal	$Mu_PIDmu > 3 \& (Mu_PIDmu - Mu_PIDp) > 0 \& (Mu_PIDmu - Mu_PIDK) > 0$
π	$Mu_PIDmu < 3 \& (Mu_PIDmu - Mu_PIDp) > 0 \& (Mu_PIDmu - Mu_PIDK) > 0$
p	$Mu_PIDmu < 3 \& (Mu_PIDmu - Mu_PIDp) > 0 \& (Mu_PIDmu - Mu_PIDK) < 0$
K	$Mu_PIDmu < 3 \& (Mu_PIDmu - Mu_PIDp) < 0 \& (Mu_PIDmu - Mu_PIDK) > 0$

Table 22: PID requirements applied to obtain enriched regions with particle-type for the K^+ Fakelines(Top) and μ^- Fakelines(Bottom). On top of these, a common $\text{ProbNNghost} < 0.35$ requirement is applied to all the regions, to reproduce what is done in the stripping.

610 Where $N_{\hat{i}}$: number of candidates passing all offline cuts in the particle type i enriched
611 region, $P(\hat{\zeta}|i)$: Mis-identification rate of the particle type i in the $\zeta = K, \mu$ signal region,
612 $P(i|i)$: efficiency to select the particle type i in the enriched region of type \hat{i} .

613 The fact that "enriched" regions themselves are polluted with misidentified particles
614 must be accounted for. This "cross-feed" between various types of misIDed particles(π
615 $\rightarrow p$, $\mu \rightarrow \pi$, etc ...) adds complications to Equation 28. $N_{\hat{i}}$ is calculated with the formula:

$$N_{\hat{i}} = \sum_j N_j \times P(\hat{i}|j) \quad (29)$$

Particle/Region	K	π	μ	p
$\epsilon K [\%]$	57.5	1.1	0.3	1.4
$\epsilon \pi [\%]$	2.1	7.8	0.6	0.4
$\epsilon \mu [\%]$	0.5	1.3	7.8	0.2
$\epsilon p [\%]$	0.4	2.9	0.3	8.2
$\epsilon K [\%]$	95.1	17.5	3.1	87.3
$\epsilon \pi [\%]$	94.5	96.1	14.9	93.9
$\epsilon \mu [\%]$	0.1	0.6	78.0	0.05
$\epsilon p [\%]$	68.0	16.7	5.1	94.7

Table 23: PIDCalib ϵ for each particle-type in every enriched region calculated for the K^+ particle (Top) and for μ^- particle (Bottom)

616 To get N_i , the actual yields of misID of the types i , we must use unfolding techniques
617 on the equations 29 where $N_{\hat{i}}$ and $P(\hat{i}|j)$ are known from our data set and from PIDCalib.
618 Table 23 shows the efficiency of selecting any particle-type in any enriched region. The
619 numbers are obtained after similar cuts to the ones in the stripping are applied to
620 the PIDCalib samples, in particular `isMuon`, `GhostProb` and `PIDK>-5` (to mimic the
621 `StdLooseKaons` cuts) requirements are done. Appendix D shows that if the matrix
622 elements sampled by PIDCalib do vary with the track momentum and η , their profile vs.

623 m_{corr} is flat. The Bayes unfolding is an “inversion” which relies on Bayes theorem. We
 624 consider $N_i(\hat{k}) = N_{\hat{k}} P(i|\hat{k})$, where $P(i|\hat{k}) = \frac{P(\hat{k}|i)P_{prior}(i)}{\sum_{\ell} P(\hat{k}|\ell)P_{prior}(\ell)}$. Then an iterative process is
 625 performed starting from a uniform $P_{prior}^0(\ell) = 1/4$ (4 being the number of track categories)
 626 and where $N_i(\hat{k})$ is computed successively, replacing $P_{prior}^0(\ell)$ with $P_{prior}(\ell) = N_{\ell}/N$
 627 ($N = \sum_k N_k$), until it stabilizes. This solving gives the N_i in their enriched regions. By
 628 applying equation 28 we obtain the yield of each misID-ed type i . Summing up those
 629 yields, we infer the total yield for K^+ , $N_K = \sum_i \mathcal{Y}_i(K)$, (2455.5 candidates) and μ^- ,
 630 $N_{\mu} = \sum_i \mathcal{Y}_i(\mu)$, (625.1 candidates) of the mis-identification components. Those are shown
 631 in Figure 36: adding both contributions we get the final yield of the mis-identification
 632 component in our signal region and its shape. The yields for the low and high q^2 regions
 633 are found to be: 1326 ± 36 and 1757 ± 42 . The central values are fixed in the fit and
 634 their uncertainty constitutes a source of systematic uncertainty. The mis-ID background
 635 is thus largely dominated by misIDed tracks to kaon, and its shape is very similar to
 636 the background $b \rightarrow c \rightarrow s$ and therefore, the effect on the signal yield in the fit is very
 637 limited.

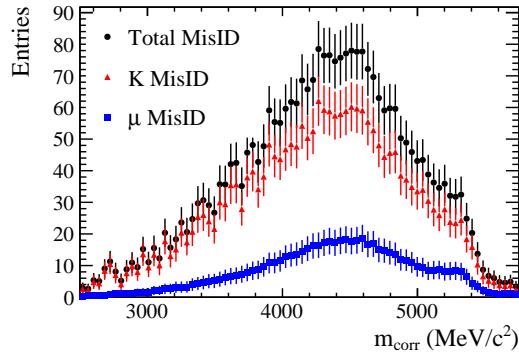


Figure 36: Distributions of corrected mass variable for K^+ misidentified component(Red), μ^- misidentified component(in Blue) and the total(in Black)

638 Figure 37 shows the comparison for misID derivation between the Bayes and SVD
 639 method, with the different ranks for the construction of the regularization parameter, as
 640 described in ref. [40].

641 9.3 Excited K^* states

642 Three excited kaon resonances, $K^{*+}(892)$, $K_2^{*+}(1430)$ and $K_0^{*+}(1430)$, have their templates
 643 combined into a single template with contributions weighted according to their spin and
 644 secondary branching fractions, expressed here as relative fraction with respect to the
 645 $K\mu\nu$ signal: $2.5 \times 0.33(K^{*+}(892) \rightarrow K^+\pi^0)$ for $K^{*+}(892)$, $0.93 (K_0 \rightarrow K\pi) \times 0.33$
 646 ($K_0^{*+} \rightarrow K^+\pi^0$) for $K_0^{*+}(1430)$ and $0.499 (K_2^* \rightarrow K\pi) \times 0.33 (K_2^{*+} \rightarrow K^+\pi^0)$ for
 647 $K_2^{*+}(1430)$, which makes these three modes in the proportions $K^*(892) : K_0^{*+}(1430) : K_2^{*+}(1430) = 1 : 0.37 : 0.20$. The factor 2.5 for $K^{*+}(892)$ comes from the average ratio
 648 $K^*(892)\mu\mu/K\mu\mu$ (for B^+ and B^0) [1]. The merging of templates is motivated by a lack
 649 of knowledge on the relative branching fractions of the different decays and in part by
 650 the similarity of the template shapes. As the template shapes of K^{*-} decays are almost
 651 identical it is impossible to distinguish between the different excited resonances, therefore

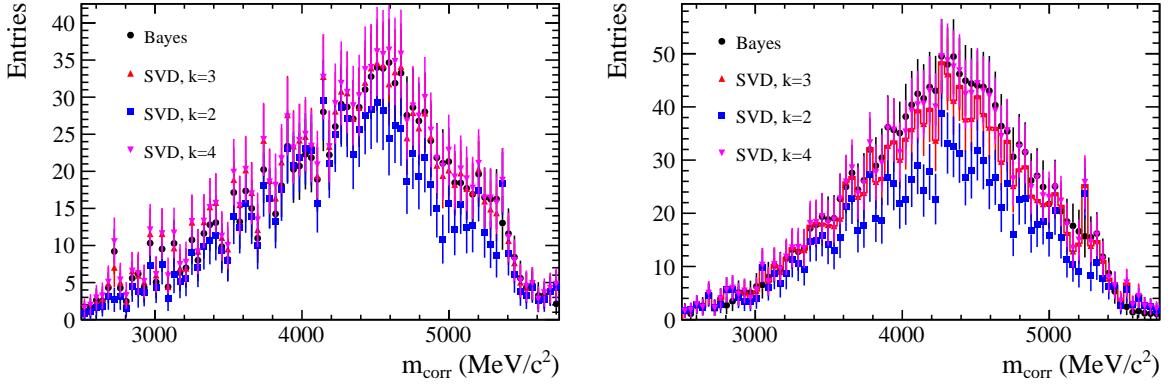


Figure 37: Distribution of the corrected mass shapes for (left) low and (right) high q^2 , for the total misID background inferred by Bayes and SVD methods. The recipe explained in ref. [40] suggests the rank $k = 3$ for the regularization parameter $\tau = s_{kk}^2$.

653 the $B_s^0 \rightarrow K^{*-} \mu^+ \nu_\mu$ decay is taken to mean the combination of all excited K^{*-} decays.
 654 The corrected $K^- \mu^+$ mass for each template and the combination is plotted in Figure 38.

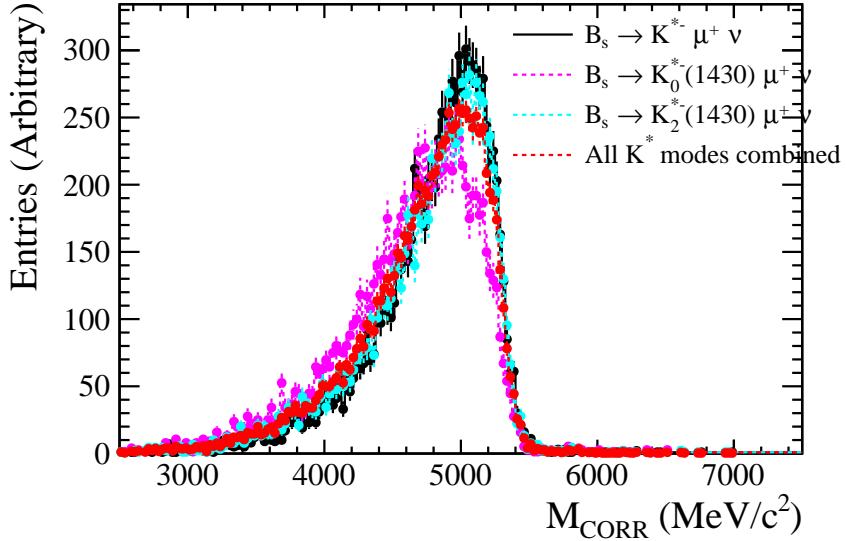


Figure 38: The corrected mass shapes for $B_s^0 \rightarrow K^{*-} \mu^+ \nu_\mu$ modes.

655 9.4 $B \rightarrow c\bar{c}(\rightarrow \mu\mu)KX$ (B2CC) modeling and constraint

656 The B2CC background includes all the decays $B \rightarrow J/\psi/\psi(2S)(\rightarrow \mu^+\mu^-)K^+X$ (where X
 657 denotes one or several light hadrons). It is suppressed with the combination of kinematic
 658 and isolation cuts as described in subsection 5.3.4. However, due to their relatively
 659 high branching fraction, a substantial amount of these events is expected to survive the
 660 selection. The strategy to estimate this background is described.

661
 662 The $B^+ \rightarrow J/\psi K^+$ background is very effectively removed through the use of charged
 663 isolation and BDTs. These selections, however, are only effective when the additional

664 muon is reconstructible, which is often the case. But events where the muon is produced
 665 outside the acceptance of the detector or inside the acceptance but not reconstructed
 666 (tracking inefficiency) present a major concern, as they appear similar to a signal decay
 667 with an outgoing neutrino.

668 Using an approach similar to that detailed in Section 3.2 it is possible to reconstruct
 669 a B^+ mass peak. The momentum of the invisible muon perpendicular to the B^+ flight
 670 direction, p_\perp , must be equal and opposite the momentum of the visible particles. The
 671 momentum of the invisible muon parallel to the B^+ flight direction, or longitudinal
 672 momentum, p_\parallel , may be found from a knowledge of the J/ψ mass. When calculating the
 673 J/ψ mass from the $K^+\mu^+\mu^-$ four vectors, the only unknown component is the longitudinal
 674 momentum of the invisible muon. Solving the four vector equation yields p_\parallel with a two
 675 fold ambiguity,

$$p_\perp(\mu^-) = -p_\perp(K^-\mu^+), \\ p_\parallel(\mu^-) = \frac{\pm\sqrt{A^2 + BC^2 - B} - AC}{C^2 - 1}, \quad (30)$$

676 where,

$$A = \frac{m_{J/\psi}^2 - m_\mu^2 + p_\perp^2(K^-) + E^2(\mu^+) + p_\parallel^2(\mu^+) - p_\perp^2(\mu^+K^-)}{2E(\mu^+)}, \\ B = m_\mu^2 + p_\perp^2(\mu^+K^-), \\ C = \frac{p_\parallel(\mu^+)}{E(\mu^+)}. \quad (31)$$

677 This method of reconstructing the B^+ peak will be referred to as *the neutrino method*. Due
 678 to an imperfect vertex resolution approximately 15% of events have unphysical solutions
 679 for p_\parallel , i.e. when $A^2 + BC^2 - B < 0$, we set $A^2 + BC^2 - B = 0$ when this happens
 680 (similarly to what is done in section 3).

681 The $B^+ \rightarrow J/\psi K^+$ yield is obtained by performing a binned maximum likelihood fit to
 682 a histogram containing the positive solution of the B^+ invariant mass (which corresponds
 683 to the correct one in 70% of the cases). The signal peak is modeled using a double
 684 Gaussian and the background shape is modeled using Chebychev function of the second
 685 order.

686 The fit is performed in a two stage process. An initial fit is performed to the Monte
 687 Carlo distribution to determine the signal shape and estimate the efficiency of the method,
 688 where the latter is defined as: signal yield from the fit divided by the number of true
 689 $B^+ \rightarrow J/\psi K^+$ in the MC sample. A second fit is performed on the data distribution
 690 to determine the signal and background yields. The plots of these fits in data and
 691 simulation are shown in Figure 39. To determine the expected $B^+ \rightarrow J/\psi K^+$ yield with
 692 a $K\mu$ reconstruction, the fit results must account for the efficiency of reconstructing the
 693 additional muon. The full expected $B \rightarrow c\bar{c}(\rightarrow \mu\mu)KX$ yield in the data fit is estimated
 694 by deriving the fraction of $B^+ \rightarrow J/\psi K^+$ in the $B \rightarrow c\bar{c}(\rightarrow \mu\mu)KX$ background source,
 695 this fraction is estimated using a simulation sample where the underlying physics of
 696 $B \rightarrow c\bar{c}(\rightarrow \mu\mu)KX$ is generated. Table 24 reports the individual yields and efficiencies
 697 related to the steps discussed in this section in bins of q^2 .

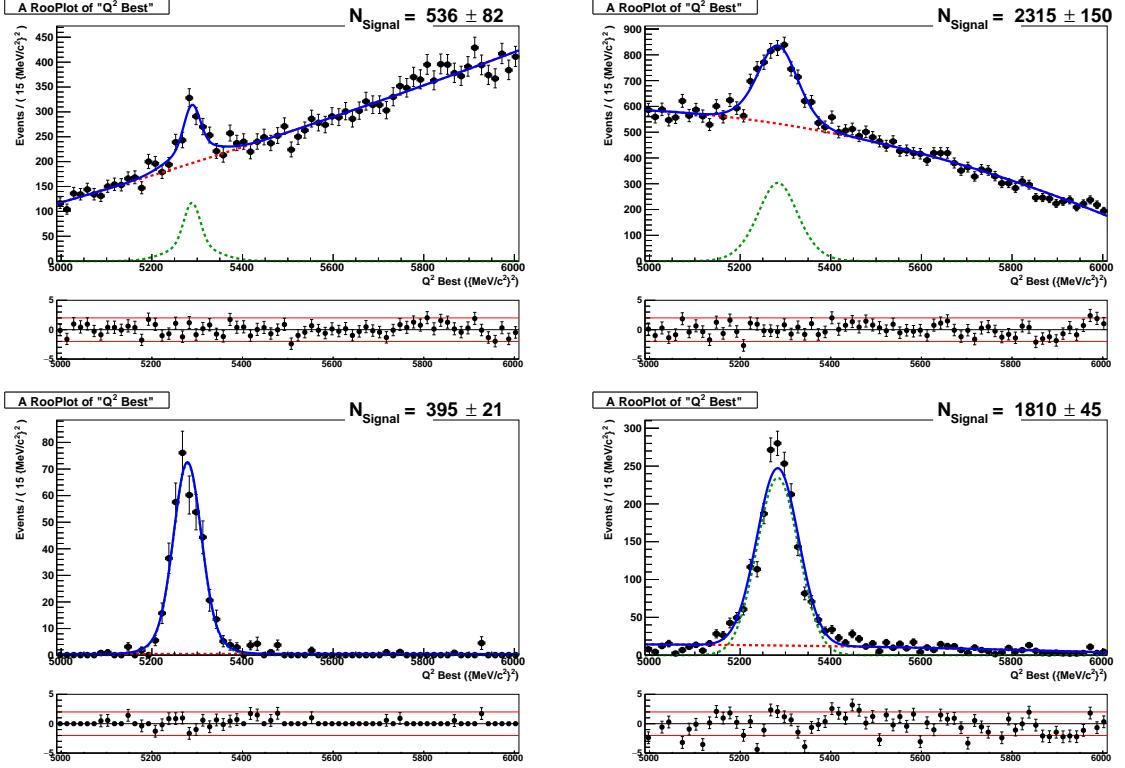


Figure 39: The $K^- \mu^+ \mu^-$ invariant reconstructed from a $K^- \mu^+$ pair using a knowledge of the B flight direction. Fit results are plotted for (top) data and (bottom) MC, for $q^2 < 7 \text{ GeV}^2/c^4$ (left) and $q^2 > 7 \text{ GeV}^2/c^4$ (right).

$(\text{Gev}^2/\text{c}^4)$	$B^+ \rightarrow J/\psi K^+$ (neutrino method)	Method Eff. (MC)	Expected $B^+ \rightarrow J/\psi K^+$	$J/\psi K^+/\text{B2CC}$ fraction(MC)	Expected B2CC
No q^2	2827.3 ± 168.2	0.63 ± 0.02	4490.7 ± 295.7	0.76 ± 0.01	5908 ± 295.7
$q^2 < 7$	535.9 ± 82.3	0.87 ± 0.06	616.5 ± 104.1	0.86 ± 0.02	716.9 ± 104.1
$q^2 > 7$	2315.0 ± 149.8	0.60 ± 0.02	3848.9 ± 275.8	0.75 ± 0.01	5131.8 ± 275.8

Table 24: Fitted (neutrino) and corrected yields of the $B^+ \rightarrow J/\psi K^+$ contribution. The gaussian constraint on the expected B2CC yield in the $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ fit is fed by the statistical uncertainty of the fits and the systematic uncertainty from Monte Carlo discrepancies.

698 9.5 Fitting results & projections

699 As mentioned in previous sections the fit contains six different components. All the yields
700 are left free in the fit except for two components: the MisID & B2CC contributions,
701 which are estimated using the methods shown in the previous subsections. The templates
702 are obtained from simulation except for the MisID and combinatorial which shapes are
703 obtained from uncorrelated data sets. The fit results are shown in Table 25, Figure 40
704 shows the projection of the fit onto the corrected mass variable in high and low bins of
705 q^2 . Those results, and the corresponding uncertainties, include the fluctuations of the fit
706 templates according to the Beaston Barlow method. To estimate the contribution of the
707 finite size of the templates in the uncertainties, the fit is performed without fluctuating the
708 templates. Projections for other variables are documented in appendix H. The correlation

matrix of the fit is shown in Table 26.

Component	low q^2	high q^2
Signal $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$	$6921.9 \pm 285.0(208.0)$	$6399.4 \pm 370.0(256.0)$
$B_s^0 \rightarrow K^{*-} \mu^+ \nu_\mu$	$3216.4 \pm 464.0(312.0)$	$4381.8 \pm 575.0(393.0)$
$B \rightarrow c\bar{c}(\mu^+ \mu^-)K^+ X$	$744.9 \pm 103.0(101.0)$	$4910.8 \pm 268.0(265.0)$
Combinatorics	$2546.6 \pm 141.0(94.7)$	$1089.5 \pm 112.0(76.3)$
$b \rightarrow c \rightarrow s$	$30102 \pm 544(268)$	$68335 \pm 726(320)$

Table 25: Signal fit output yields for the different components. The numbers between parentheses are the uncertainties obtained when the fit templates are not fluctuated.

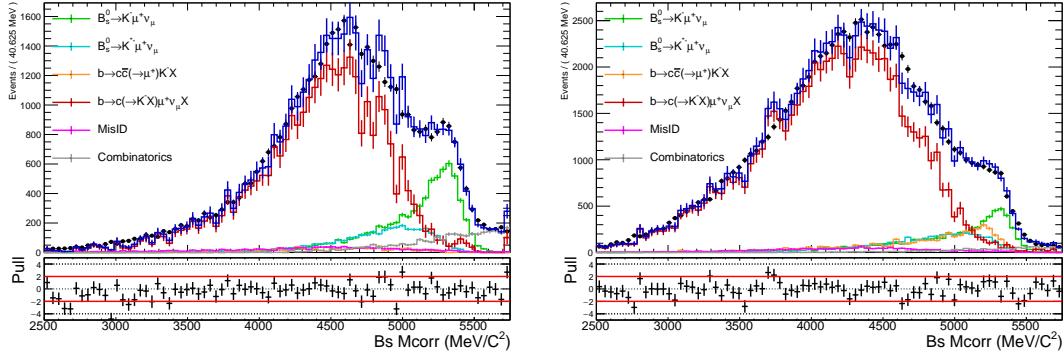


Figure 40: Fit projection onto corrected mass variable for low (Left) and high (Right) q^2 bins. The total pdf is shown in blue along its individual components: Signal (Green), B2CC (Orange), K^* (Cyan), MisID(Magenta), Combinatorial (Grey) and Vcb (red) backgrounds.

High q^2	Param.	1	2	3	4	5
$B \rightarrow c\bar{c}(\mu^+ \mu^-)K^+ X$	1	1	0.030	-0.396	-0.110	0.020
Combinatorics	2		1	0.070	-0.350	-0.028
$B_s^0 \rightarrow K^{*-} \mu^+ \nu_\mu$	3			1	-0.636	-0.436
Signal $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$	4				1	0.199
$b \rightarrow c \rightarrow s$	5					1
Low q^2	Param.	1	2	3	4	5
$B \rightarrow c\bar{c}(\mu^+ \mu^-)K^+ X$	1	1	-0.041	-0.084	0.026	-0.113
Combinatorics	2		1	0.087	-0.414	-0.047
$B_s^0 \rightarrow K^{*-} \mu^+ \nu_\mu$	3			1	-0.617	-0.595
Signal $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$	4				1	0.217
$b \rightarrow c \rightarrow s$	5					1

Table 26: Fit correlation coefficients for $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ fit obtained for all free yields in the fit.

9.6 Fit goodness and toys

To test the validity of the fit, we run 1250 pseudoexperiments with two configurations: the first is identical to the fitter's configuration where the Barlow-Beeston method is applied, the second is identical to the fitter's configuration except the Barlow-Beeston method is deactivated. The pull distributions for these two sets of toys are shown in Figure 41. No bias is observed. The uncertainty from the fit is over estimated for the nominal toys,

716 due to the fact that the Barlow-Beeston method is activated hence the uncertainty due
 717 to limited MC statistics is propagated to the final fit uncertainties: the effect of the MC
 718 statistics is estimated to be 30% of the total uncertainty.

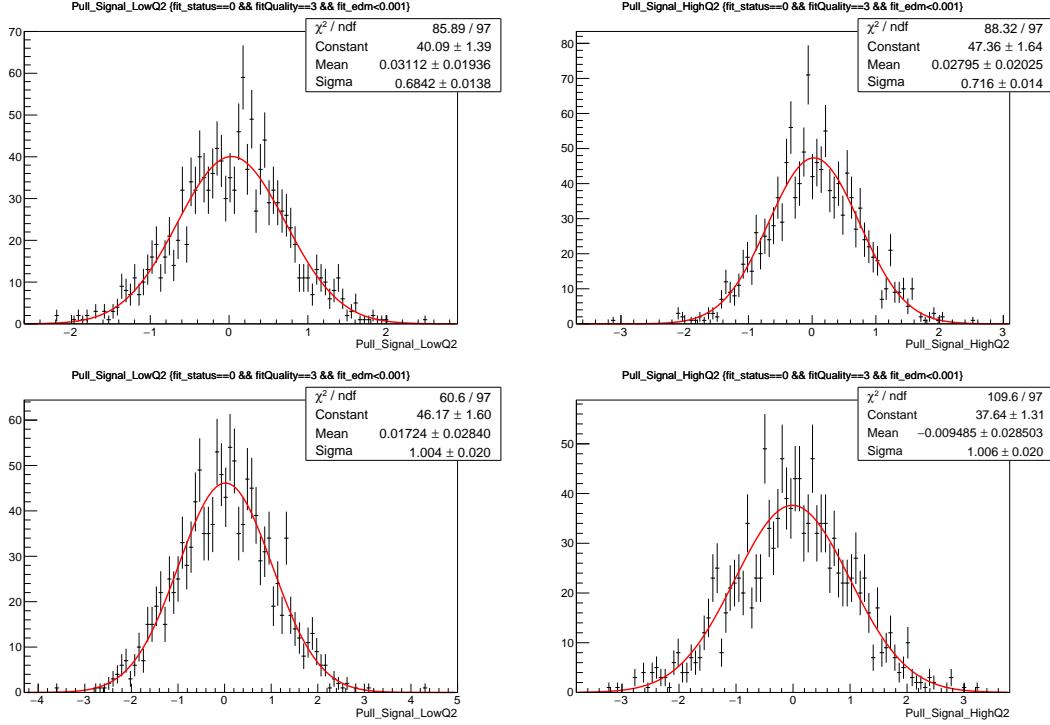


Figure 41: Pull distributions for Signal yields for (left) low q^2 and (right) high q^2 for (top) nominal fit with Barlow-Beeston activated, and (bottom) with Barlow-Beeston method deactivated.

719 An alternative check was done to verify that the yield uncertainties are correctly
 720 determined by the BB fit. The fit templates were varied from pseudoexperiment to
 721 pseudoexperiment along with the toy data. The obtained distribution of the signal yield
 722 pulls can be seen in Fig.42. It can be observed that the central values and widths correctly
 723 account for the fit results in data.

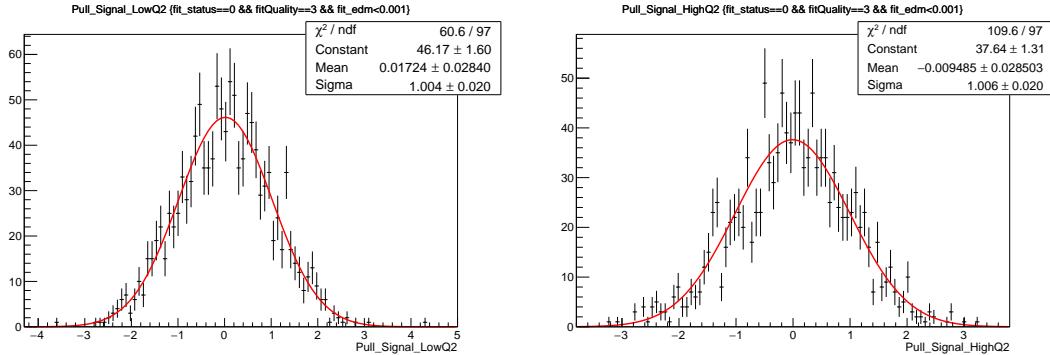


Figure 42: Pull distributions for the signal yields for (left) low q^2 and (right) high q^2 obtained from pseudoexperiments where fit templates were varied along the toy data.

724 10 Efficiencies and systematics

725 In this section, the corrections for the efficiencies and the systematics are presented. For
 726 $B_s^0 \rightarrow K^+ \mu^- \nu$, the chosen default model is based on the LCSR prediction (Rusov *et al.*)
 727 for the low q^2 bin and the most recent LQCD prediction (A.X.Khadra *et al.*), for the higher
 728 q^2 bin.

729 10.1 Generator Efficiency

	$\varepsilon_{\text{Gen.}} [\%]$	$\sigma_{\text{stat.}} [\%]$	$\sigma_{\text{FF.}} [\%]$	$\sigma_{\text{comb.}} [\%]$
$B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$	17.87	0.08		
$B_s^0 \rightarrow K^- \mu^+ \nu_\mu$	20.58	0.08	0.23	0.24
$B_s^0 \rightarrow K^- \mu^+ \nu_\mu \quad q^2 < 7 \text{ GeV}^2/c^4$	19.67	0.13	0.03	0.12
$B_s^0 \rightarrow K^- \mu^+ \nu_\mu \quad q^2 > 7 \text{ GeV}^2/c^4$	20.02	0.09	0.16	0.19

Table 27: Generator efficiencies for $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ and $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$ in different q^2 regions. The uncertainties originate from Monte Carlo statistics, form factor parametrisation and the quadrature sum. For $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$ there is minimal variation in Form Factors between parametrisations and the form factor uncertainty is ignored.

730 A pre-selection is applied to Monte Carlo events before the simulation of particle
 731 interactions with the detector. These selections are called *generator cuts* as they are
 732 applied immediately after the generation of the decay. A selection is made on the polar
 733 angle, θ_{flight} ,

$$0.01 < \theta_{\text{flight}} < 0.4, \quad (32)$$

734 on some or all of the final state particles. When simulating $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ events,
 735 generator cuts are applied to all charged final state particles, and when simulating
 736 $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu X$ events the cuts are applied to the muon and daughters of the D_s^+ . For all
 737 other backgrounds the generator cuts are applied on all charged final state tracks. As this
 738 analysis measures $|V_{ub}|$ with respect to a given q^2 selection the generator efficiency must
 739 be determined for that region rather than the whole sample. To measure the generator
 740 efficiencies for $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ and $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$ small Monte Carlo samples of 250,000
 741 events are generated without a simulation of the detector and without the generator
 742 cuts. The Generator efficiencies are plotted against q^2 in Figure 43 with the simulated q^2
 743 distributions overlaid before and after the selection. The generator efficiencies are quoted
 744 in Table 27.

745 10.2 Particle Identification

746 The PID efficiencies are calculated using the PIDCalib package [42]. The PID selections
 747 applied to data and simulation are listed in Table 28. To minimise systematic effects,
 748 tight PID selections are only applied to the opposite sign kaon and muon while for
 749 $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$, very soft selections are applied to the opposite sign $\pi^- K^+$ pair. Conse-
 750 quently the efficiency of PID selections will be similar for both the signal and normalisation
 751 decays and systematic effects are reduced when calculating corrections to the ratio of
 752 efficiencies.

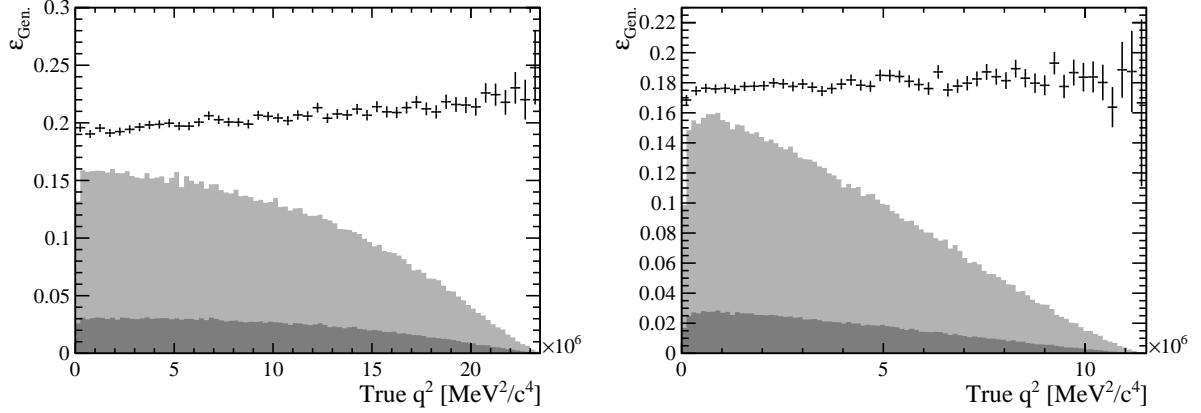


Figure 43: The Generator efficiencies plotted against the true q^2 for $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ (left) and $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$ (right). The q^2 distributions are plotted in grey before and after the selections are applied.

μ^+	$\ln \mathcal{L}_\mu / \mathcal{L}_\pi > 3$	and	$\ln \mathcal{L}_\mu / \mathcal{L}_p > 0$	and	$\ln \mathcal{L}_\mu / \mathcal{L}_K > 0$
K^-	$\ln \mathcal{L}_K / \mathcal{L}_\pi > 5$	and	$\ln \mathcal{L}_K / \mathcal{L}_p > 5$	and	$\ln \mathcal{L}_K / \mathcal{L}_\mu > 5$
K^+	$\ln \mathcal{L}_K / \mathcal{L}_\pi > -2$				
π^-	$\ln \mathcal{L}_K / \mathcal{L}_\pi < 20$				

Table 28: The PID likelihood selections applied to all particles. Selections are aligned between $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ and $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$ minimising systematics when taking the ratio of efficiencies.

753 Lookup tables binned in momentum, pseudorapidity and track multiplicity in the
 754 underlying events is generated with each entry containing the PID efficiency for that
 755 region of data. Figures 44 and 45 show a 2D projection of these lookup tables.

756 To minimise systematic effects from the intra bin variations in efficiency and the
 757 *sWeight* background subtraction a MC/Data driven correction is used instead of the pure
 758 data driven correction. The data driven correction returns a *true* efficiency value for
 759 a given PID cut, the MC/Data driven correction returns the ratio of PID efficiencies
 760 obtained from data and Monte Carlo. This ratio is used to correct the PID efficiency in
 761 the simulation.

762 When determining efficiencies from Monte Carlo each track from each event is weighted
 763 by the correction factor obtained from the lookup table. The corrected Monte Carlo yield
 764 is taken as the sum of the correction weights. Systematic uncertainties are quantified by
 765 performing 1000 pseudo-experiments, each time varying the contents of the lookup tables
 766 within the obtained errors. The PID corrections for each q^2 bin used in the fits is given in
 767 Table 29. Figure 46 shows the variation of the PID efficiencies with M_{corr} .

	$B_s^0 \rightarrow K^- \mu^+ \nu_\mu$	$B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$	Ratio
No Sel.	0.850 ± 0.005	0.798 ± 0.006	1.067 ± 0.007
$q^2_{K^- \mu^+} < 7 \text{ GeV}^2/\text{c}^4$	0.848 ± 0.005	0.798 ± 0.006	1.062 ± 0.007
$q^2_{K^- \mu^+} > 7 \text{ GeV}^2/\text{c}^4$	0.854 ± 0.005	0.798 ± 0.006	1.074 ± 0.007

Table 29: PID correction factors averaged over all tracks and all events applied to Monte Carlo.

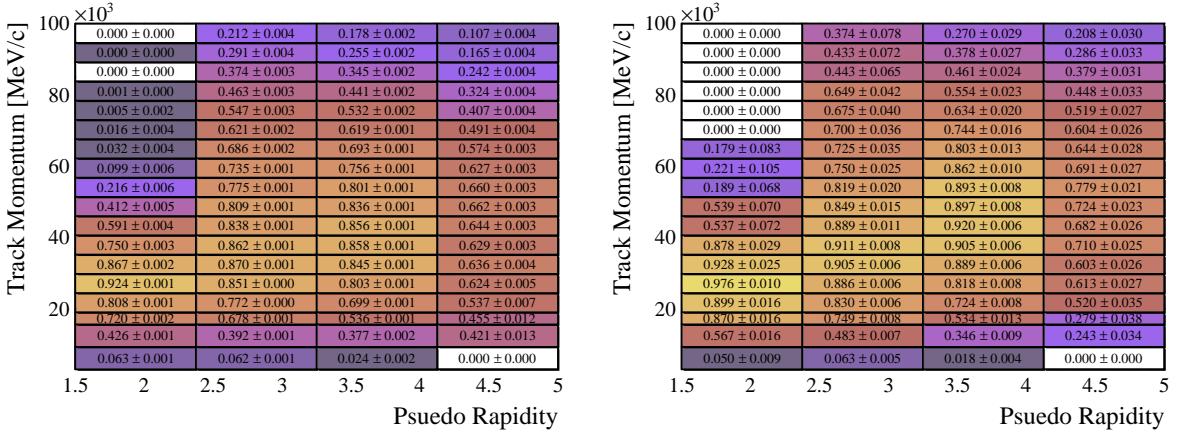


Figure 44: A two dimensional projection of the PID efficiency lookups for kaons determined from data (left) and Monte Carlo (right) $D^{*+} \rightarrow D^0\pi^+$ decays.

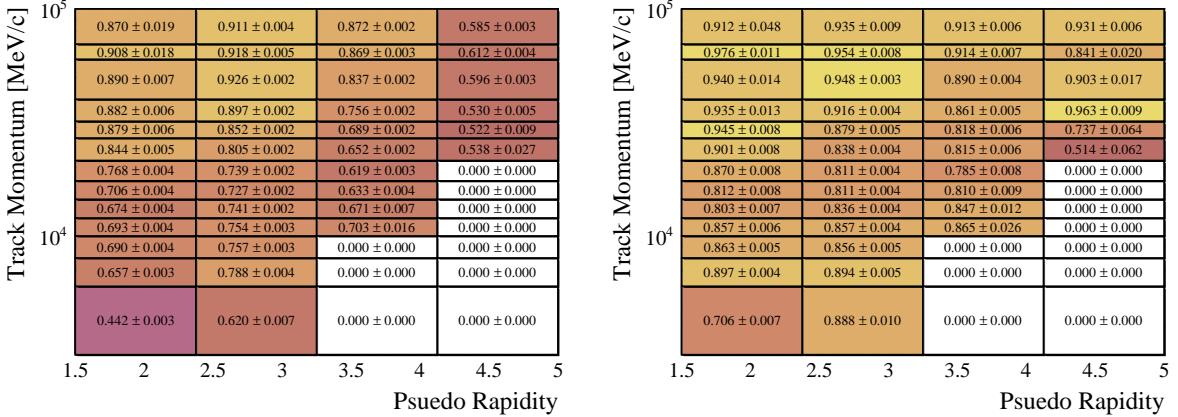


Figure 45: A two dimensional projection of the PID efficiency lookups for muons determined from data (left) and Monte Carlo (right) $J/\psi \rightarrow \mu^+\mu^-$ decays.

Additional systematic sources are due to the choice of the binning scheme in the variables p , η and nTracks in the PIDCalib lookup Tables. They are estimated by varying the binning scheme: as systematic uncertainty of 0.4% is found.

10.3 Trigger efficiencies

To account for a possible mismodeling of our trigger efficiency in the simulation we follow similar approach to [43]. The method is based on estimating the TISTOS efficiency [44] in data and simulation using our control channel ($B^+ \rightarrow J/\psi K^+$ decay). To map the variations of the TISTOS efficiency for the different trigger lines, the correction (data/MC) is evaluated in bins of p_T and IP of the muon. Figures 48 show the distribution of TISTOS efficiency in data and MC along with their ratio.

This correction is then used for the simulated signal and normalization channels to obtain the ratio of the correction factors between signal and normalization (in bins of μ , p_T and IP), Figure 49 shows the distribution of signal and normalization events.

The correction map (Figure 48, Bottom) allows to derive the correction factor for signal and normalization decays. To estimate the systematic uncertainty due to this correction,

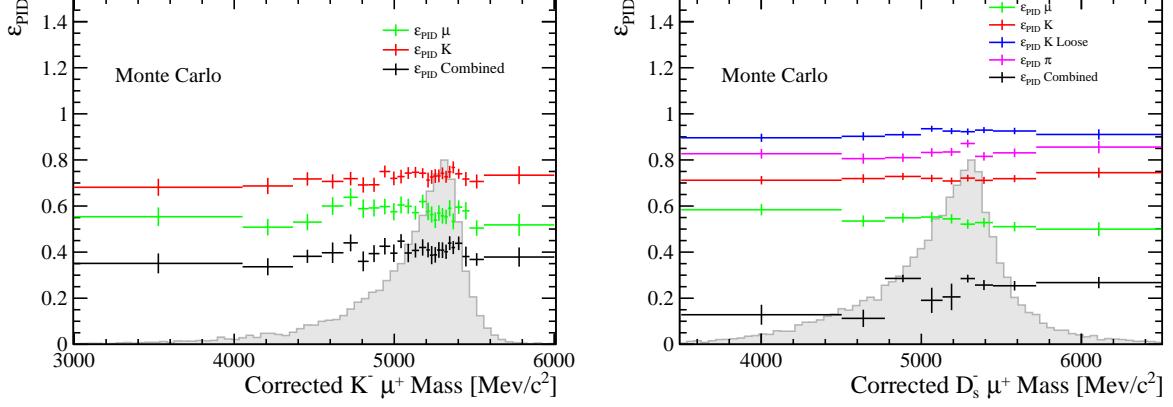


Figure 46: The efficiencies of PID selections are plotted against the B_s^0 corrected mass for $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ (left) and $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$ (right). The signal corrected mass distribution is shaded in light grey.

783 the correction map is varied within the bin uncertainties using 1000 pseudoexperiments.
 784 The resulting distribution of the ratio Signal/Normalization is shown in Figure 50, right:
 785 the mean of the distribution is quoted to be the correction to the trigger mismodeling
 786 in MC and the RMS is the systematic uncertainty on that number. This excercise is
 787 repaeated for the two q^2 regions, Table 30 reports the results for the full sample and the
 788 q^2 regions.

789 To assess the systematic uncertainty on the correction factor we redid the excercise
 790 binning the sample in p_T and IPCHI2 of the muon, the variation w.r.t to the values quoted
 791 in the Table 30 are(for full, High and low q^2 regions): 0.3, 0.2, 0.6%. Using $B^+ \rightarrow J/\psi K^+$
 792 decays, we also performed cross checks for TISTOS and TOS efficiencies for individual
 793 trigger lines used in this analysis. We also study the variation of the TOS fraction for the
 794 trigger decisions in bins of η and p_T , the trends are compatible between data and MC, as
 795 documented in Appendix G.

	$B_s^0 \rightarrow K^- \mu^+ \nu_\mu$	$B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$	Ratio
No Sel.	0.698 ± 0.007	0.693 ± 0.006	1.008 ± 0.003
$q^2_{K^- \mu^+} < 7 \text{ GeV}^2/c^4$	0.703 ± 0.008	0.693 ± 0.006	1.013 ± 0.006
$q^2_{K^- \mu^+} > 7 \text{ GeV}^2/c^4$	0.694 ± 0.007	0.693 ± 0.006	1.002 ± 0.002

Table 30: Trigger correction factors averaged over all events applied to Monte Carlo in bins of q^2 .

796 10.4 Tracking Correction

797 It is of vital importance that the efficiency of reconstructing tracks is well understood when
 798 performing a cross section or branching fraction measurement. The track reconstruction
 799 efficiency is over 95% and is determined from Monte Carlo. A data driven correction is
 800 applied to the simulation using clean $J/\psi \rightarrow \mu^+ \mu^-$ decays. The tracking reconstruction
 801 efficiency is measured using a tag and probe method, the tag muon is fully reconstructed
 802 as well identified muon and the probe track is partially reconstructed without information

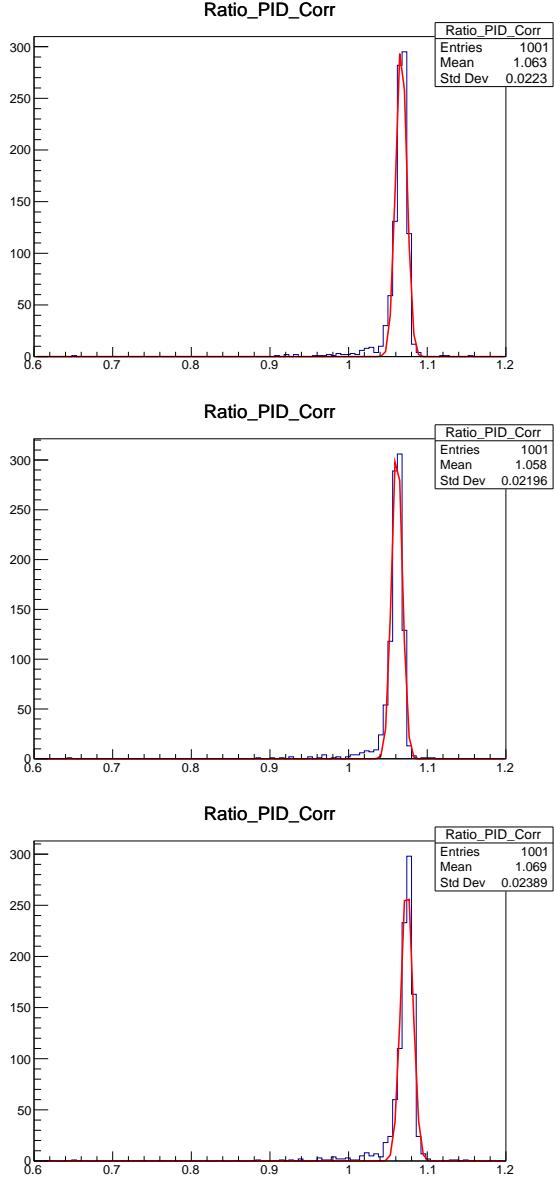


Figure 47: Distribution of PID correction ratio varying the individual efficiency corrections from the lookup tables within their uncertainties for Full q^2 region (Top Left), Low q^2 region (Top Right) and for High q^2 region (Bottom).

803 from at least one subdetector which is being probed. The tracking efficiency is determined by
 804 counting the amount of fully reconstructed tracks corresponding to the partially reconstructed
 805 probe track. Performing the tag and probe analysis on both simulation and data yields a
 806 discrepancy of approximately 2% [45].

807 A lookup table of tracking efficiency ratios and their ratios is provided by the LHCb
 808 collaboration. The two dimensional table binned in momentum and pseudorapidity is
 809 visualised in Figure 51. The tracking efficiency corrections are applied as a weight on
 810 each track as determined from the lookup table and efficiencies are corrected by taking
 811 the product of the weights for each track. As $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ is a two body decay and
 812 $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$ is a four body decay the corrections partially cancel when taking the ratio

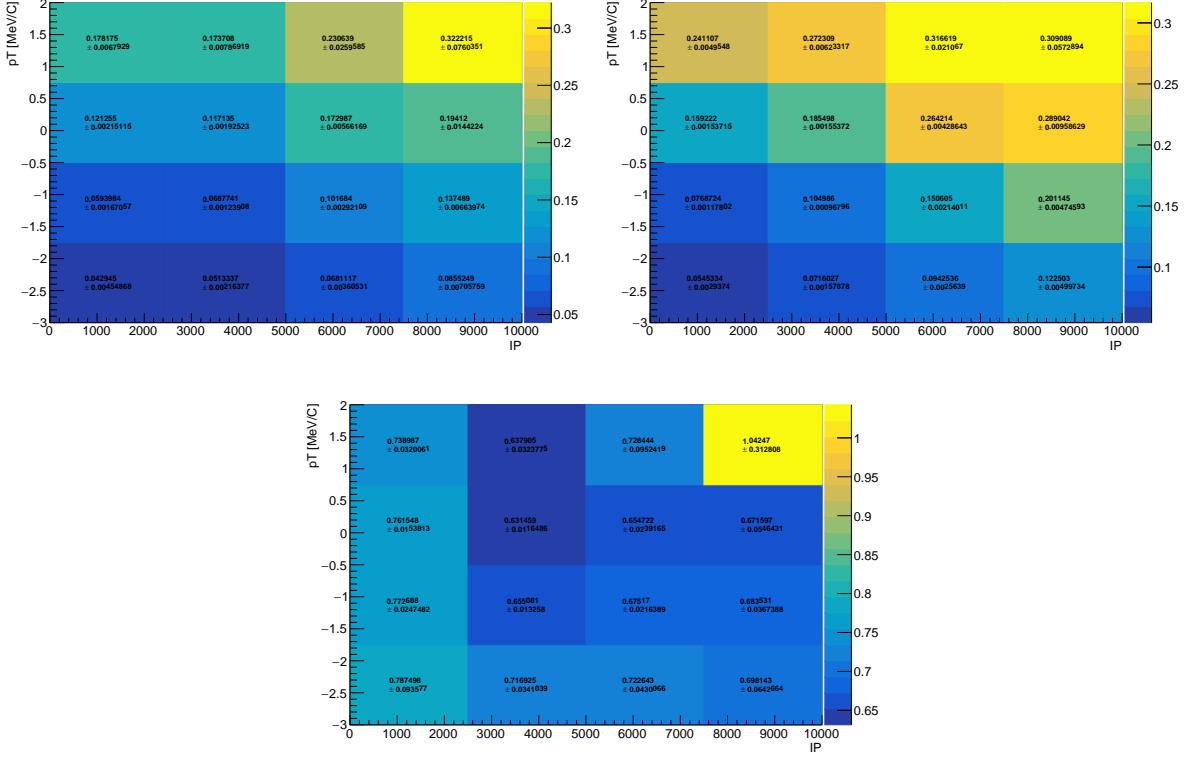


Figure 48: Top: TISTOS efficiency for $B^+ \rightarrow J/\psi K^+$ decay in bins of μ p_T and IP for the trigger requirement in data (left) and simulation (right). Bottom: the ratio $TISTOS_{data}/TISTOS_{MC}$ for $B^+ \rightarrow J/\psi K^+$ decays.

$K^- \mu^+ q^2$ Sel.	$ B_s^0 \rightarrow K^- \mu^+ \nu_\mu $	$ B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu $	Ratio
No Sel.	0.999 ± 0.001	0.975 ± 0.004	1.024 ± 0.004
$q^2_{K^- \mu^+} < 7 \text{ GeV}^2/c^4$	0.999 ± 0.001	0.975 ± 0.004	1.025 ± 0.004
$q^2_{K^- \mu^+} > 7 \text{ GeV}^2/c^4$	0.999 ± 0.001	0.975 ± 0.004	1.025 ± 0.004

Table 31: Tracking efficiency corrections applied to Monte Carlo events.

813 of the efficiencies. As instructed in [46], tracking corrections are applied to our simulation
814 samples where re-weighting BDT(explained in 7) is applied as weight.

815 The uncertainties on the overall correction factor are determined by performing 1000
816 pseudo-experiments, each time the efficiencies in the lookup table are varied within their
817 uncertainties 52. The tracking corrections to the efficiency calculations are summarised
818 in Table 31 for each of the q^2 bins used in the fits. The systematic uncertainties from
819 hadronic interactions with detector material are partially canceled between the two modes
820 in the ratio, the remaining ones would be the kaon's (1.1%) and pion's (1.4%) they are
821 estimated to be 0.02

822 10.5 $B^+ \rightarrow J/\psi K^+$ corrections

823 The decays $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ and $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$ are partially reconstructed and have broad
824 distributions making it difficult or impossible to isolate a pure signal sample in data. In

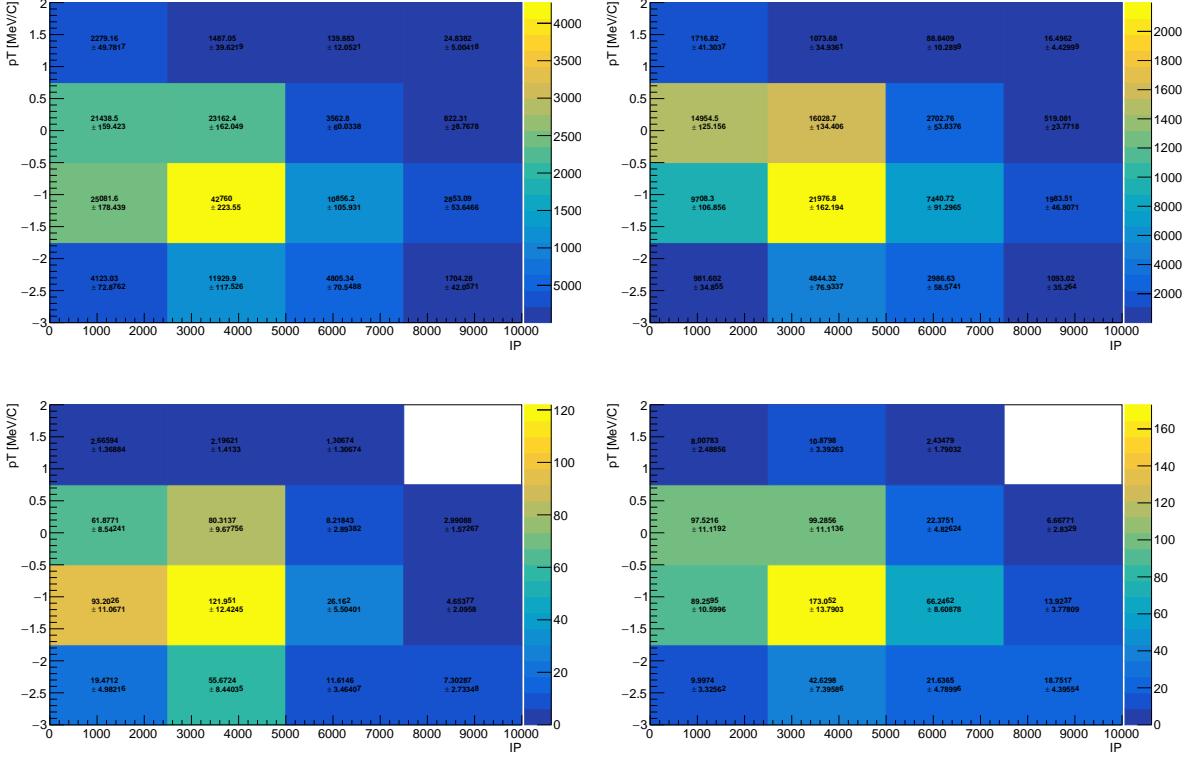


Figure 49: Top: Signal(Left) and Normalization (Right) distribution of simulated events in bins of p_T and IP in the full q^2 region. Bottom: Signal distribution of simulated events in bins of p_T and IP for the low q^2 region(Left) and High q^2 region(Right).

order to validate the efficiencies of a selection and ensure that biases between data and simulation are corrected, the decay $B^+ \rightarrow J/\psi K^+$ is used as a proxy for the signal decay. When reconstructed using only one muon the B^+ decay is kinematically very similar to the signal B_s^0 decay, allowing the efficiencies of selections on kinematic variables to be validated. When fully reconstructed the efficiencies of selecting B^+ decays is similar to the signal B_s^0 decay as there are no additional tracks which can be associated with the secondary vertex.

	$K^- \mu^+$	$D_s^- \mu^+$
$\sigma m_{\text{Corr.}}$	1.036 ± 0.005	
Isolation BDT	0.993 ± 0.001	0.988 ± 0.002
Charged Track BDT	0.988 ± 0.006	
Same Sign BDT	0.958 ± 0.011	

Table 32: Correction factors applied to Monte Carlo determined from simulated and real decays of $B^+ \rightarrow J/\psi K^+$.

Efficiency corrections are calculated for the corrected mass uncertainty cut and the BDT response variables. The efficiency of a selection is calculated for $B^+ \rightarrow J/\psi K^+$ by performing a fit to the invariant mass distribution of the $\mu^- \mu^+ K^+$ triad before and after a selection. The correction factor is the ratio of the efficiency in data and Monte Carlo and

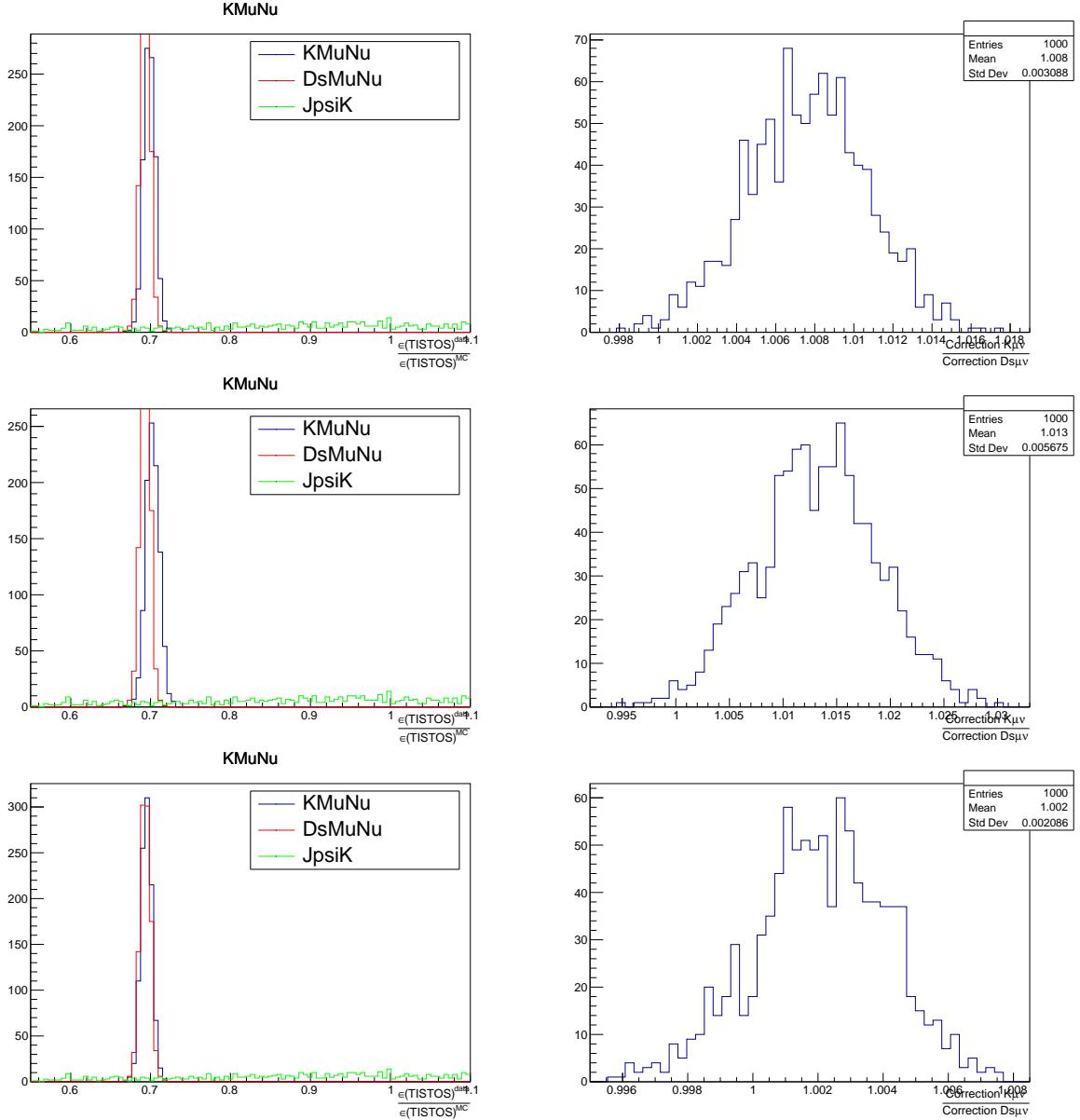


Figure 50: Left: Distribution of correction factors(data/MC) for Signal (red), normalization (Green) and control decay(black). Right: distribution of the ratio of correction factors (Signal/Normalization) for 1000 toy experiments. Top plots represent the full q^2 region, center plots are for the low q^2 region while the bottom plots are for the high q^2 region.

836 the efficiency for $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ obtained from Monte Carlo is scaled by the correction
 837 factor. A uncertainty obtained from the correction factor is applied as a systematic
 838 correction. The corrections are listed in Table 32 and plots displaying *sPlot* background
 839 subtracted $B^+ \rightarrow J/\psi K^+$ data alongside Monte Carlo are given in Figs 53 - 55. The q^2 of
 840 $B^+ \rightarrow J/\psi K^+$ peaks at $m_{J/\psi}^2$ resulting in very few events being reconstructed in the low
 841 q^2 bin. The corrections applied in the low q^2 are set equal to those in the high q^2 bin.

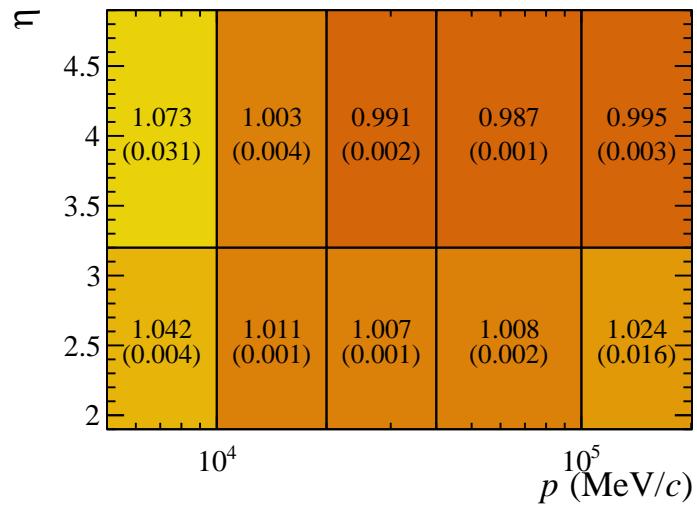


Figure 51: The look-up table used to correct the tracking efficiency of charged tracks, binned in momentum and pseudorapidity.

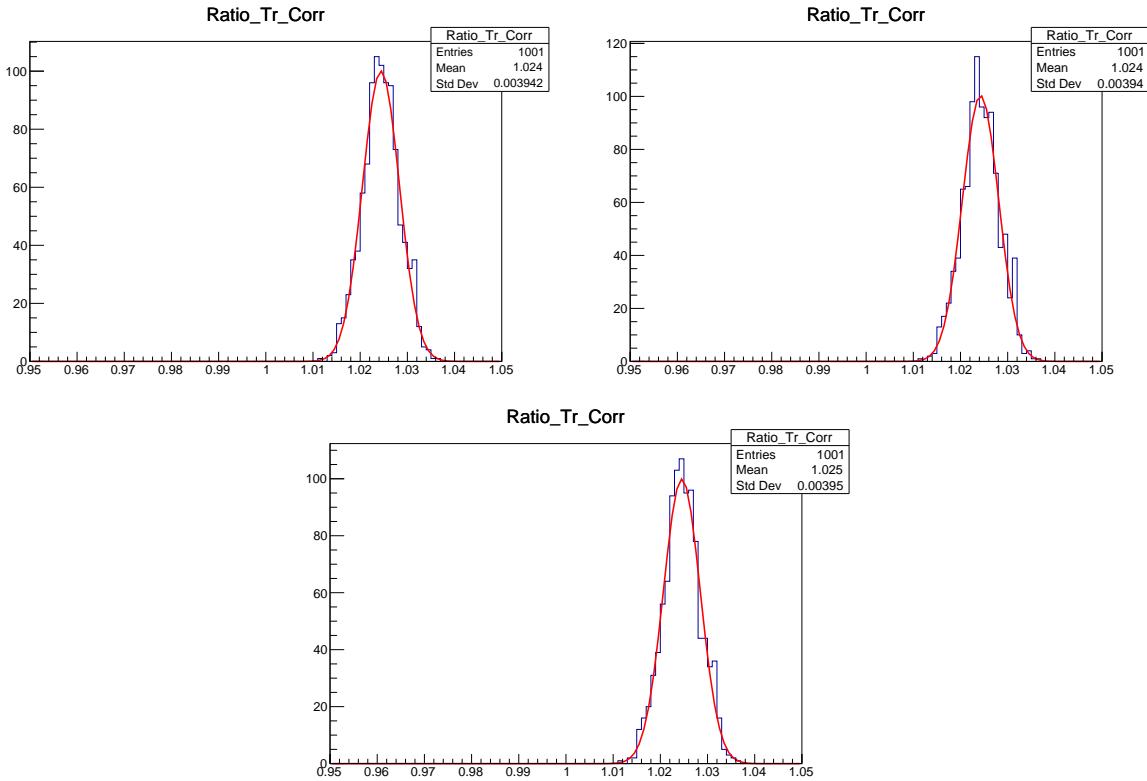


Figure 52: Distribution of tracking correction Ratio varying the individual efficiency corrections from the lookup table within their uncertainties for Full q^2 region(Top Left), Low q^2 region(Top Right) and for High q^2 region(Bottom).

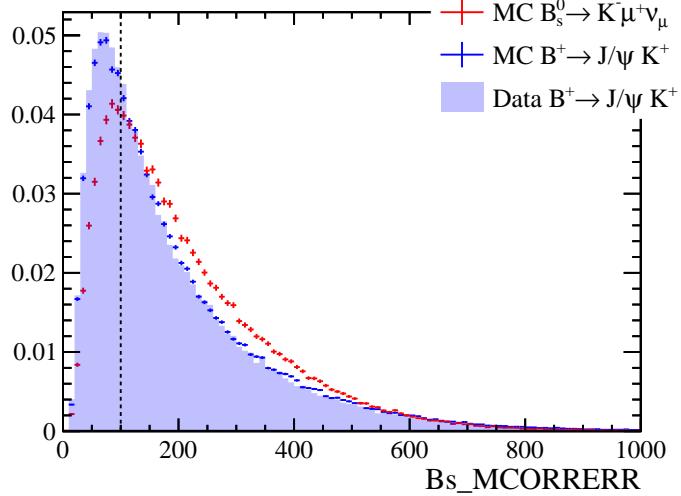


Figure 53: The corrected mass uncertainty for $B^+ \rightarrow J/\psi K^+$ (blue) and $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ (red). A *sPlot* background subtraction is performed on the data. The correction factor is taken as the ratio of $B^+ \rightarrow J/\psi K^+$ data and Monte Carlo decays passing the selection. As opposed to Figure 7, the $B^+ \rightarrow J/\psi K^+$ decay is fully reconstructed to allow for quantitative calculations of the cut efficiency.

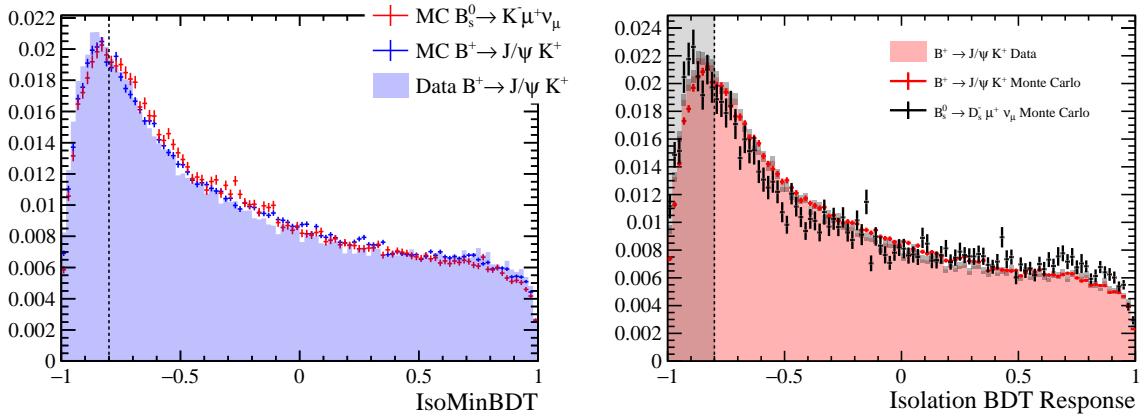


Figure 54: The response of the isolation BDT for $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ (left) and $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$ (right) is plotted against the $B^+ \rightarrow J/\psi K^+$ calibration samples.

842 10.6 q^2 Migration

843 A neutrino solution is selected using the linear regression method detailed in Section 3.2.
 844 The resolution on the reconstructed q^2 will result in some migration of events across the
 845 selection boundary with some events rejected that should have been selected and vice
 846 versa. The distribution of the true q^2 is plotted against the reconstructed q^2 in Figure 56,
 847 the region containing events migrating either in or out of the high q^2 region are illustrated.
 848 Inward migration is defined by the events with a true q^2 outside the region of interest
 849 but are reconstructed inside due to the resolution. Outward migration is defined by the
 850 events which are truly in the region of interest but are reconstructed out. A correction
 851 factor is calculated from simulated Monte Carlo events by taking the ratio of events truly

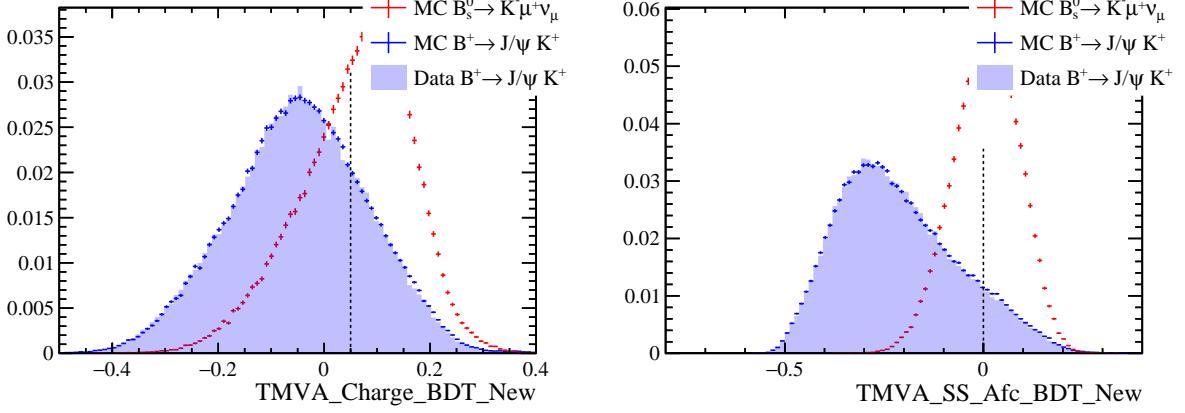


Figure 55: The response of the BDTs rejecting charged (left) and same sign (right) backgrounds for $B_s^0 \rightarrow K^-\mu^+\nu_\mu$ are plotted against the $B^+ \rightarrow J/\psi K^+$ calibration samples.

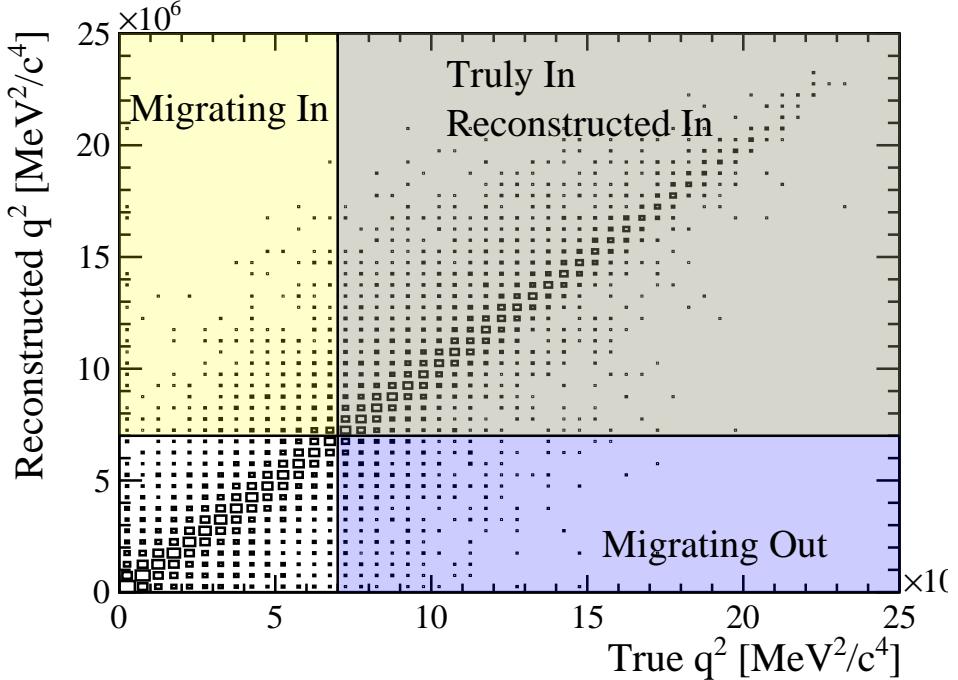


Figure 56: The reconstructed q^2 solution selected using the regression model is plotted against the true q^2 for simulated $B_s^0 \rightarrow K^-\mu^+\nu_\mu$ events. The regions of inward and outward migration are shaded when requiring the q^2 to be above 7 GeV $^2/c^4$.

in the high q^2 with events reconstructed in the q^2 region. The Monte Carlo is reweighted to be consistent with form factor predictions from Lattice QCD and light cone sum rules, and the percentages of events migrating in and out are listed in Table 33.

The difference of migration correction factors due to different Form Factor models is evaluated in the Table 33, results lead to a correction that is compatible with 1 for all form factor models. The relative uncertainty is 2%.

Model	Corr. Mig. $q^2 < 7 \text{ GeV}^2/\text{c}^4$	Corr. Mig. $q^2 > 7 \text{ GeV}^2/\text{c}^4$
A.X.Khadra et al.	1.02 ± 0.02	0.96 ± 0.02
Bouchard	1.00 ± 0.02	0.99 ± 0.02
Rusov	1.01 ± 0.02	0.99 ± 0.02
Flynn	1.02 ± 0.02	0.97 ± 0.02

Table 33: Migration of q^2 regions due to resolution on the reconstructed q^2 using the choice closest to the regression value. Results obtained from simulated $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ events after a full selection is applied. The events have been reweighted to be consistent with predictions from Lattice QCD and LCSR.

10.7 Final Corrected Relative Efficiency

The uncertainties on the corrections are taken as systematic uncertainties when calculating the final ratio of branching fractions. The corrected efficiency is plotted against the true q^2 in Figure 57 for $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ (left) and $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$ (right). The unfortunately large bias on the efficiency of the signal mode combined with a lack of knowledge on the shape of true q^2 distribution results in the assignment of a systematic uncertainty on the final corrected efficiency. The systematic uncertainty on the final corrected efficiency originating from an uncertainty on the knowledge of the q^2 distribution is determined by calculating the corrected $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ efficiency under each of the four form factor models and taking the standard deviation.

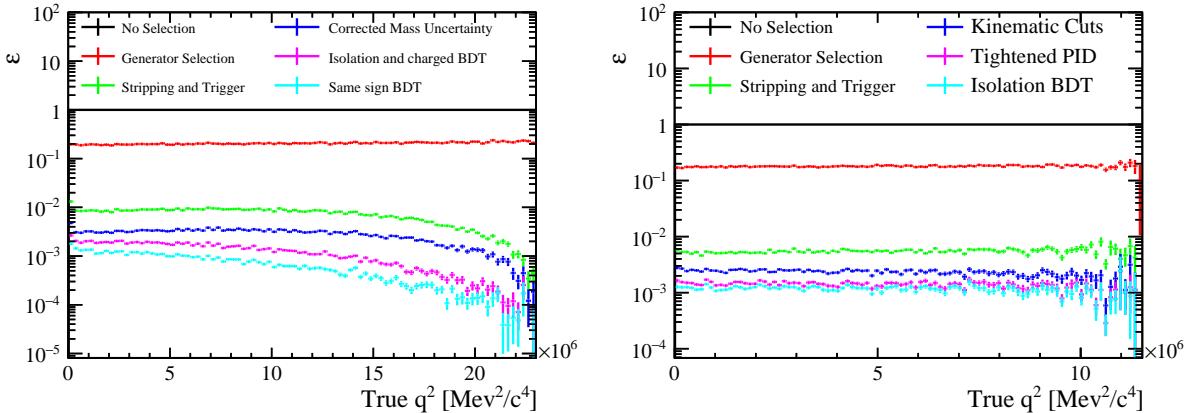


Figure 57: The corrected Efficiencies for successive selections on $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ (left) and $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$ (right) candidates are plotted against the true q^2 .

The final corrected efficiencies for $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$, for the baseline model, are listed in Table 34. The generator and reconstruction+selection efficiencies for $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$ are 0.1787 ± 0.0008 ($17.87 \pm 0.08\%$) and $0.649 \pm 0.005\%$. The numbers for the other FF models are shown in appendix I.1. The reconstruction+selection efficiency are corrected for tracking, PID, data/MC discrepancy factors. The efficiency ratio includes, on top of those, the effects of the trigger, $\sigma_{mc_{corr.}}$, Isolation, BDTs and q^2 migration.

	$\varepsilon_{\text{gen}} (\%)$	$\varepsilon_{\text{recsel}} (\%)$	$\varepsilon_{B_s^0 \rightarrow K^- \mu^+ \nu_\mu} / \varepsilon_{B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu}$
No q^2 sel.	20.582 ± 0.075	0.413 ± 0.003	0.728 ± 0.009
$q^2 < 7 \text{GeV}^2/c^4$	19.669 ± 0.130	0.654 ± 0.008	1.113 ± 0.018
$q^2 > 7 \text{GeV}^2/c^4$	21.021 ± 0.092	0.305 ± 0.004	0.527 ± 0.009

Table 34: Final generator efficiency, reconstruction/selection efficiency and global ratio $\varepsilon_{B_s^0 \rightarrow K^- \mu^+ \nu_\mu} / \varepsilon_{B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu}$. The global ratio is obtained after applying the corrections due to the effects of the trigger, $\sigma_{m_{\text{corr}}}$, Isolation, BDTs and q^2 migration.

Yield range	mean	sigma
$q^2 > 7 \text{GeV}^2/c^4$	6426.7 ± 4.4	257.5 ± 3.1
$q^2 < 7 \text{GeV}^2/c^4$	6946.6 ± 3.5	202.9 ± 2.5

Table 35: Mean and width of the distribution of signal yields in the default setup.

874 10.8 Fit systematics

875 The evaluation of the uncertainties on the shape and yields of the components in the
876 signal fit addresses the following points:

- 877 • The uncertainty in the K^* mixture for the resonant $K^* \mu \nu$ modes.
- 878 • misID component(s): misID rates and efficiencies, derivation of the template shapes
879 with unfolding, uncertainty on the estimated yields of the misID components.
- 880 • B2CC component: alternative way of extracting the yield from the neutrino method.
- 881 • Combinatorial component: alternative template obtained from the control sample
882 defined by the reverse of the anti-quadrant cut explained in Sec. 5.3.2.
- 883 • Variation of the Form Factor Models.

884 For the normalization channel fit, the main noticeable uncertainty lies in the choice of
885 gathering the components that have very close shapes.

886

887 10.8.1 Signal fit systematics

888 To assess the uncertainties, we generate a set of 2500 pseudo datasets using the default
889 set of templates normalized according to the fractions and yields of the components and
890 sub-components found from the fit to data. Then, for the studies, alternative templates
891 are used to fit the pseudo datasets and compare the mean values of the distributions of
892 the signal yields (in both low and high q^2 regions) to calculate the uncertainty for each
893 source of systematics. As a null test, the distribution of the values of the yields when
894 the default set of templates is used for the fit is shown in Figure 58, Table 35 shows the
895 mean values of the yields for the signal, starting points for the assessment of the systematics.

896

897 Mis-PID components:

898

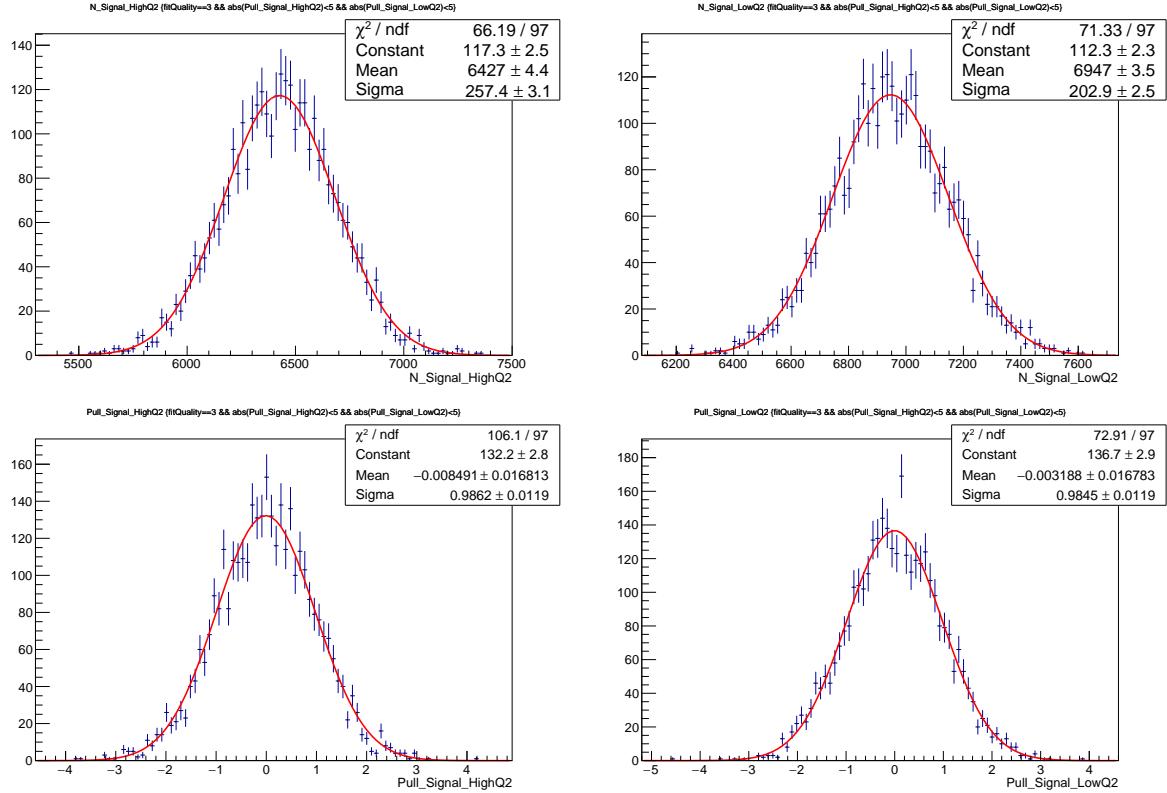


Figure 58: Distributions of (top) signal yields and (bottom) pulls for (left) high and (right) low q^2 regions based on ~ 2500 toys generated using the default setup.

899 The precision on the ID efficiencies and misID rates are affected by two factors: the
900 size of the PIDCalib control samples which induces a very negligible uncertainty and the
901 systematic variation due to the choice of the binning scheme (*under assessment*).

902 The extraction of the misID shapes involves the Bayes unfolding explained in section 9.2:
903 to estimate the unfolding systematics, the shapes obtained with the Bayes method are
904 replaced by the ones obtained with the SVD in the signal fit, using the toy setup (Figure 59)
905 we find that the signal yields are not modified as shown in the Table 36

906 The yields of the misID components are fixed in the base signal fit. To assess the
907 impact of the statistical uncertainty of the yields on the signal yields, we vary the yields
908 of the misID components by $\pm 1\sigma$ for each q^2 independently. As in the previous case, we
909 find similar negligible effects on both signal yields.

911 Excited K^* states composition:

913 The m_{corr} shapes for $B_s^0 \rightarrow K^*(892)\mu\nu$ and $B_s^0 \rightarrow K_2^*(1430)\mu\nu$ are similar while the
914 shape of $B_s^0 \rightarrow K_0^*(1430)\mu\nu$ differs from them. Therefore, only the relative weight of the
915 latter with respect to the formers will play a role in the variation of the overall shape of
916 $B_s^0 \rightarrow K^*\mu\nu$ resonances. Alternatively to what is done in subsection 9.3, we use a mixture
917 of resonances with equal weights, accounting for the secondary branchings $K^* \rightarrow K\pi^0$,
918 thus with the proportions $K^*(892) : K_0^*(1430) : K_2^*(1430) = 1 : 0.93 : 0.499$. With
919 this new $B_s^0 \rightarrow K^*\mu\nu$ shape, the central values of the $B_s^0 \rightarrow K\mu\nu$ signal yields obtained

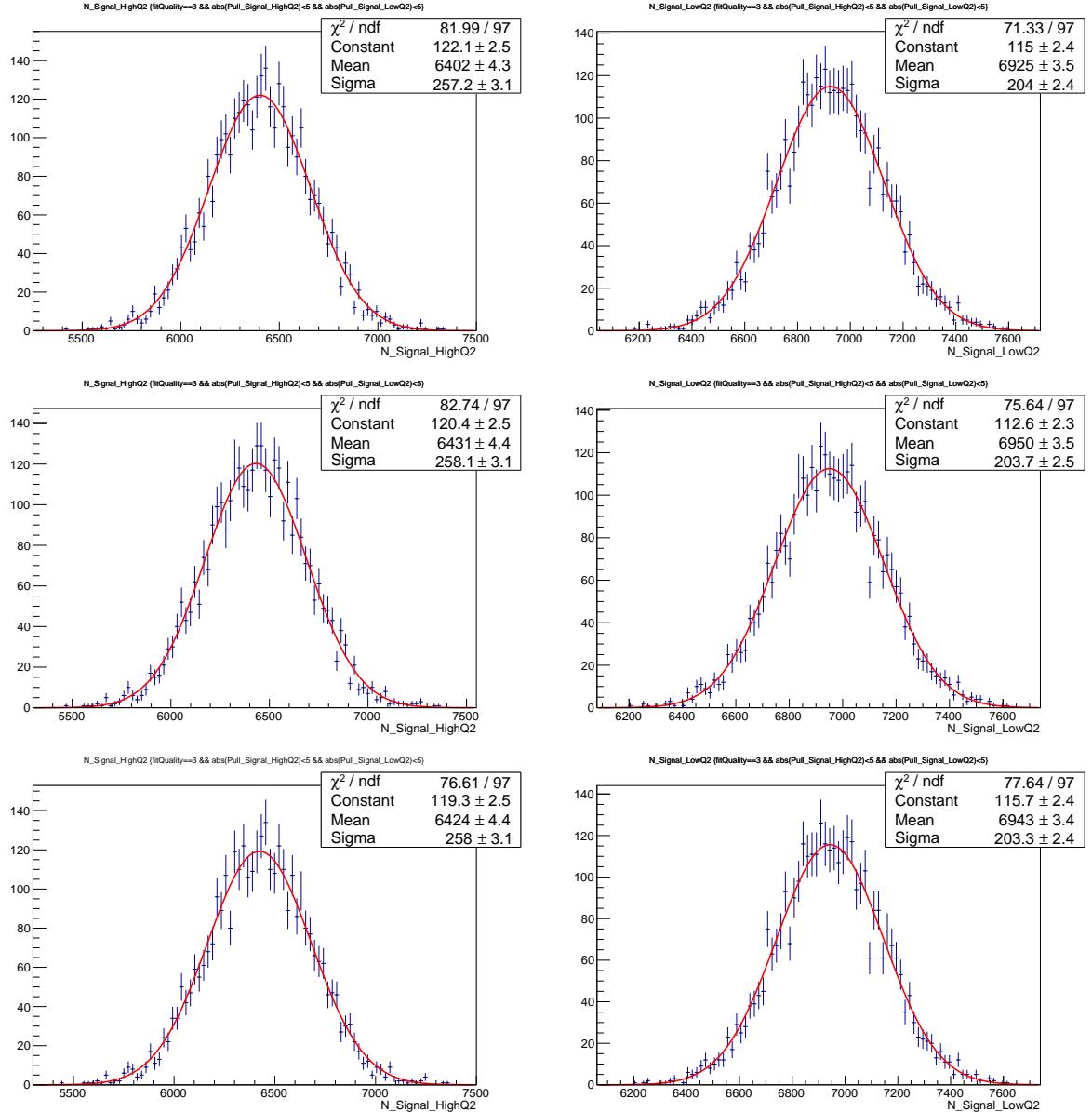


Figure 59: Distributions of signal yields for (top) the fit with the MisPID SVD alternative shape, (middle) decreasing and (bottom) increasing the MisPID yields by one standard deviation in the default set up. Left and right are for high and low q^2 regions, respectively.

in the low and high q^2 bins are 6591.1 and 6651.2, respectively. This corresponds to relative increases of 1% and 1.8% for the integrated yield. Another possibility which has been tested is to use the average of the relative weights between K^* , derived from the branching fraction for the decays $B^+ \rightarrow K^{*+}(892)/K_0^{*+}(1430)/K_2^{*+}(1430)\phi$ [1] giving the proportions $K^*(892) : K_0^*(1430) : K_2^*(1430) = 1 : 0.65 : 0.419$: in this case, the central values of the $B_s^0 \rightarrow K\mu\nu$ yields are found to be 6570.3 (+0.5%) and 6672.1 (-1.4%). Toy distributions are shown in Figure 60, and the results of the systematic studies for this source are summarized in Table 37.

928

929 Excited K^* states Form Factor:

Systematic type	mean	sigma
SVD $q^2 > 7 \text{ GeV}^2/c^4$	6401.6 ± 4.4	257.2 ± 3.1
	6924.8 ± 3.5	204.0 ± 2.4
Base misID $+1\sigma$ $q^2 > 7 \text{ GeV}^2/c^4$	6423.6 ± 4.4	258.0 ± 3.1
	6943.2 ± 3.4	203.3 ± 2.4
Base misID -1σ $q^2 > 7 \text{ GeV}^2/c^4$	6431.0 ± 4.4	258.1 ± 3.1
	6950.4 ± 3.5	203.7 ± 2.5

Table 36: Mean and width of the distribution of the signal yields for the different misID systematic variations.

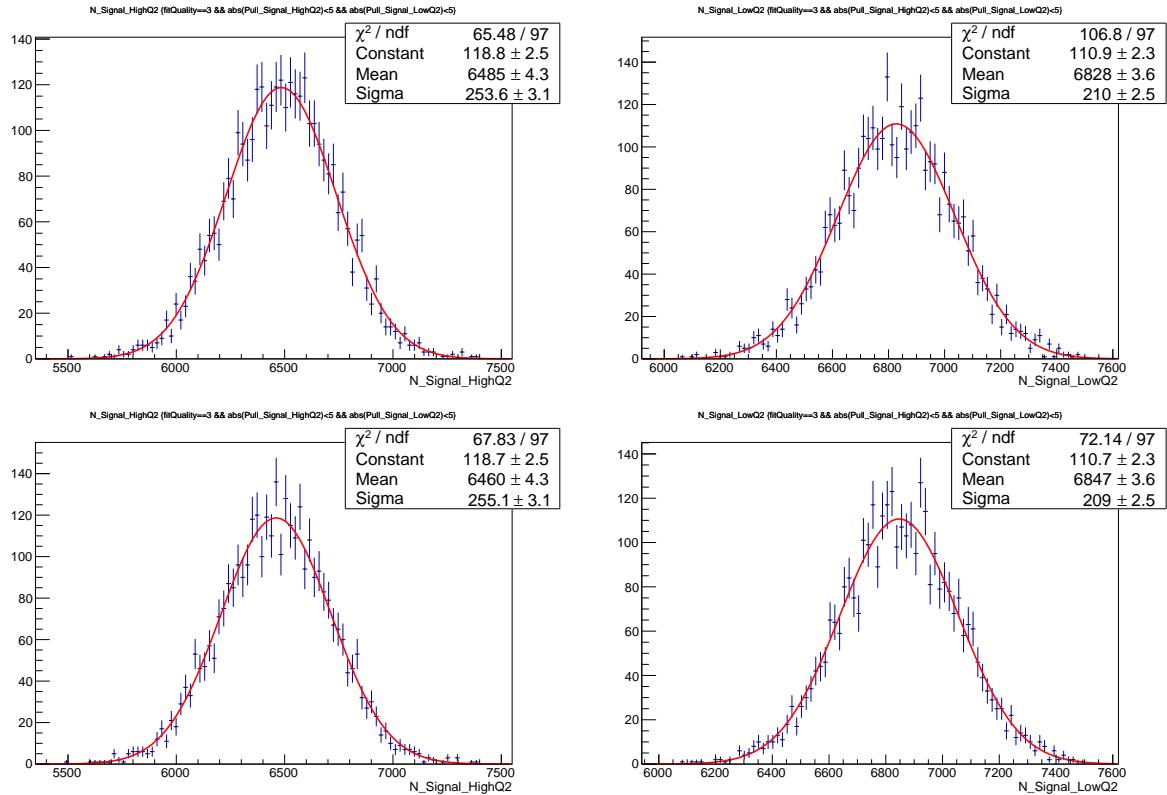


Figure 60: Distributions of signal yields for the two tests of K^* mixture: (Top) using equal weights for K^* sub-components and (Bottom) using weights inspired from $B^+ \rightarrow K^{*+}(892)/K_0^{*+}(1430)/K_2^{*+}(1430)\phi$ decays. Left and right represents the regions $q^2 > 7 \text{ GeV}^2/c^4$ and $q^2 < 7 \text{ GeV}^2/c^4$, respectively.

930

931 Similarly to the signal $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$, the excited states have been produced
 932 with the default model **ISGW2**. The influence of more up-to-date models on the m_{corr}
 933 distribution has been studied. The **flavio** package [47] provides a form factor model for
 934 $B_s^0 \rightarrow K^*(892)\mu\nu$ based on a prediction from 2016 [48], as illustrated in Fig.61. Based on
 935 that, the q^2 distribution with the $K\mu\nu$ reconstruction hypothesis was derived and the
 936 **ISGW2** distribution reweighted. Figure 62 shows the comparison between the different
 937 m_{corr} distributions. We conclude that the impact of the Form Factor for the excited
 938 states, within the $K\mu\nu$ reconstruction can be neglected.

Yield range	mean	sigma
$q^2 > 7 \text{ GeV}^2/c^4$	6484.7 ± 4.3	253.6 ± 3.1
$q^2 < 7 \text{ GeV}^2/c^4$	6827.6 ± 3.6	210.0 ± 2.5
$q^2 > 7 \text{ GeV}^2/c^4$	6460.4 ± 4.3	255.1 ± 3.1
$q^2 < 7 \text{ GeV}^2/c^4$	6874.1 ± 3.6	209.0 ± 2.5

Table 37: Mean of the distributions of signal yields for the two tests of K^* mixture: (Top) using equal weights for K^* sub-components and (Bottom) using weights inspired from $B^+ \rightarrow K^{*+}(892)/K_0^{*+}(1430)/K_2^{*+}(1430)\phi$ decays.

939

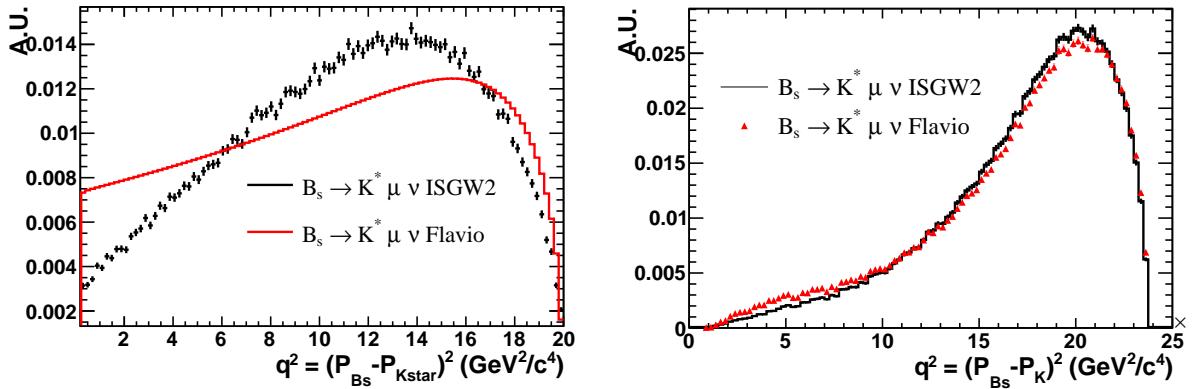


Figure 61: The generator-level q^2 distributions for $B_s^0 \rightarrow K^*(892)\mu\nu$ sample for $K^*\mu\nu$ (left) and $K\mu\nu$ (right) hypotheses, for the different Form Factor models.

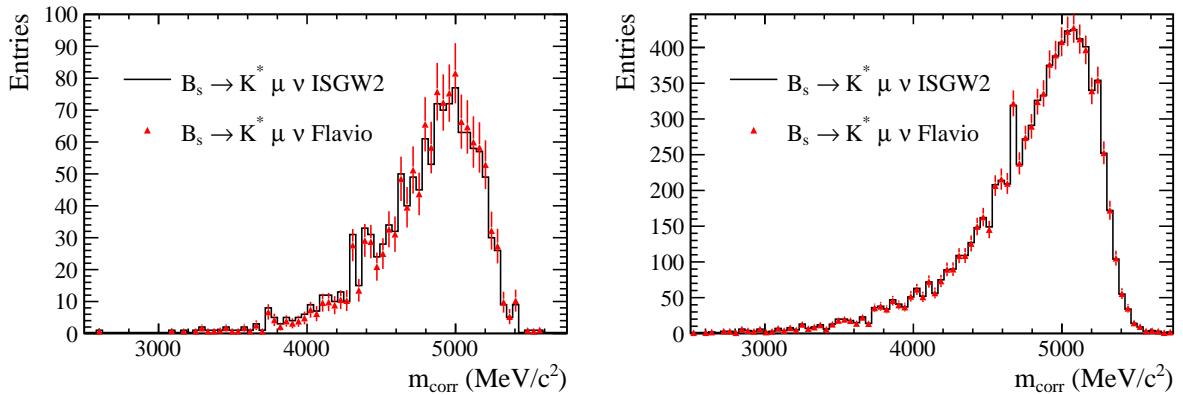


Figure 62: The m_{corr} distributions for the $B_s^0 \rightarrow K^*(892)\mu\nu$ sample for $q^2 < 7 \text{ GeV}^2/c^4$ (left) and $q^2 > 7 \text{ GeV}^2/c^4$ (right), for the different Form Factor models.

940

Hypothetical $K\pi^0$ non-resonant wave:

941

942

943

944

$B_s^0 \rightarrow K\pi^0\mu\nu$ events were generated using a phase space model with a cut $m(K\pi^0) < 2 \text{ GeV}/c^2$ to infer the shape of such a component. Figure 63 shows that its shape would match the shape of $B_s^0 \rightarrow K_0^{*+}\mu\nu$ and is therefore absorbed by this

component, and the corresponding systematics studied in previous paragraphs.

946

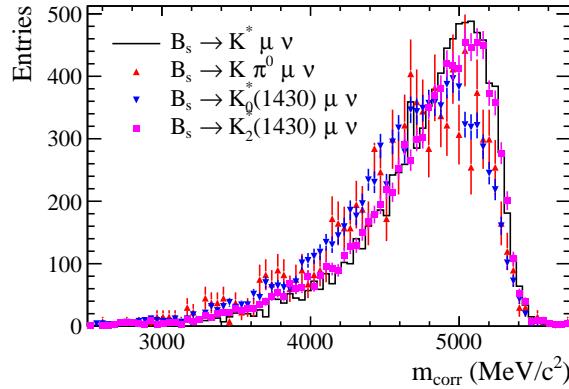


Figure 63: The m_{corr} distributions for the $B_s^0 \rightarrow K^* \mu \nu$ and $B_s^0 \rightarrow K \pi^0 \mu \nu$ samples.

947 **Event mixing template:**

948

949 A reversed cut on the quadrant selection explained in section 5.3.2 creates a
 950 region which is enriched in combinatorial events. Though the region has some small
 951 contamination of physics, notably V_{cb} events, we consider its m_{corr} distribution as a good
 952 alternative for the (conservative) variation of the shape of the combinatorial background,
 953 Fig.64. The distribution of the signal yields from the toys are shown in Fig. 65 and Ta-
 954 ble 38 shows the (conservative) amplitude of the expected systematics due to this variation.
 955

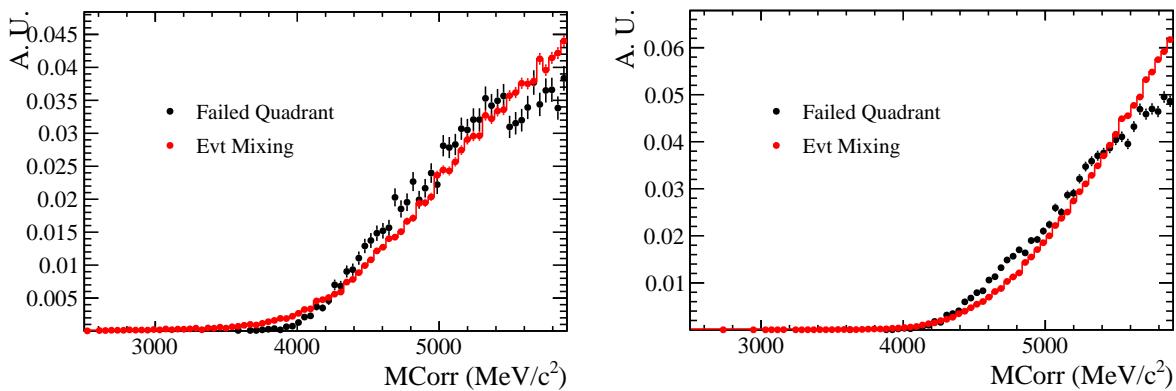


Figure 64: m_{corr} shape comparison between the event mixing and the anti-quadrant cut for (left) high and (right) low q^2 regions.

956 **B2CC background:**

957

958 An alternative method to estimate the yields of B2CC is considered. We established
 959 in section 9.4 that $B^+ \rightarrow J/\psi K^+$ invariant mass can be fully reconstructed using the
 960 isolation algorithm, producing a peak to fit in data: Figure 66 shows these peaks in
 961 the low and high q^2 regions, after the stripping stage. However, the offline isolation

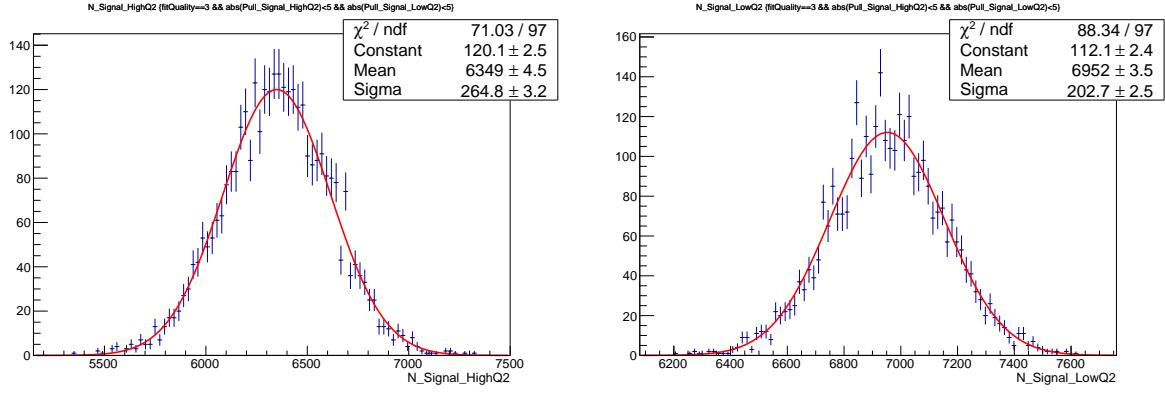


Figure 65: Distributions of signal yields for (left) high and (right) low q^2 regions based on ~ 2500 toys using the alternative combinatorial template.

Yield range	mean	sigma
$q^2 > 7 \text{ GeV}^2/c^4$	6349.3 ± 4.5	264.8 ± 3.2
$q^2 < 7 \text{ GeV}^2/c^4$	6952.1 ± 3.5	202.7 ± 2.5

Table 38: Distribution of the mean and width of the signal yields when using the alternative combinatorial template.

requirement suppresses such peaks. To infer the yield of the B2CC component after applying the full selection, we compute the efficiency of the isolation requirement in the B2CC simulation in each q^2 region. The data yields shown in the fit of Figure 66 are then multiplied by this efficiency. This procedure gives the following numbers:

Low q^2

$B^+ \rightarrow J/\psi K^+$ Yield (data) before the isolation: 10750.4 ± 195.4

Efficiency of the isolation = 0.298 ± 0.007

Efficiency of the offline cuts = 0.019 ± 0.001

$B^+ \rightarrow J/\psi K^+$ Yield (data) after isolation = 685.4 ± 41.4

High q^2

$B^+ \rightarrow J/\psi K^+$ Yield (data) before the isolation: 175449.8 ± 3134.4

Efficiency of the isolation = 0.169 ± 0.002

Efficiency of the offline cuts = 0.0057 ± 0.0002

$B^+ \rightarrow J/\psi K^+$ Yield (data) after isolation = 5917.5 ± 226.5

Figure 67 and Table 39 summarize the results of the toys performed with this new normalization of the B2CC template.

978

979 Form Factor model:

Yield range	mean	sigma
$q^2 > 7 \text{ GeV}^2/c^4$	6261.2 ± 4.3	255.8 ± 3.1
$q^2 < 7 \text{ GeV}^2/c^4$	6943.2 ± 3.5	203.3 ± 2.5

Table 39: Mean and width of the distributions of the signal yields obtain in the toy studies using alternative B2CC yields from the isolation method.

980

The m_{corr} distribution for the different Form Factors is shown in Fig 68.

The results of the toys performed with the different LQCD and LCSR models are shown in Fig.69 and summarized in Table 40.

984

985 Summary of systematics of signal fit:

986

987 Table 41 summarizes the individual systematic yield variation along with the relative
 988 uncertainty contribution of each systematic source. To derive the final relative uncertainties,
 989 the sums in quadrature of the negative and positive variations are performed separately.

990 10.8.2 Normalization fit systematics

991 A similar toy exercise to what is done for the signal in 10.8.1 is performed for the
 992 normalization channel. The null test gives an average and a σ for the DsMuNu signal of
 993 197524.2 ± 32.0 and 1903.3 ± 22.7 , respectively.

994 MisID component:

995

996 All components in the normalization fit are left free except for the MisPID, which
 997 yield is very small. Using a $\pm 1\sigma$ variation of the MisPID yield around its central value,

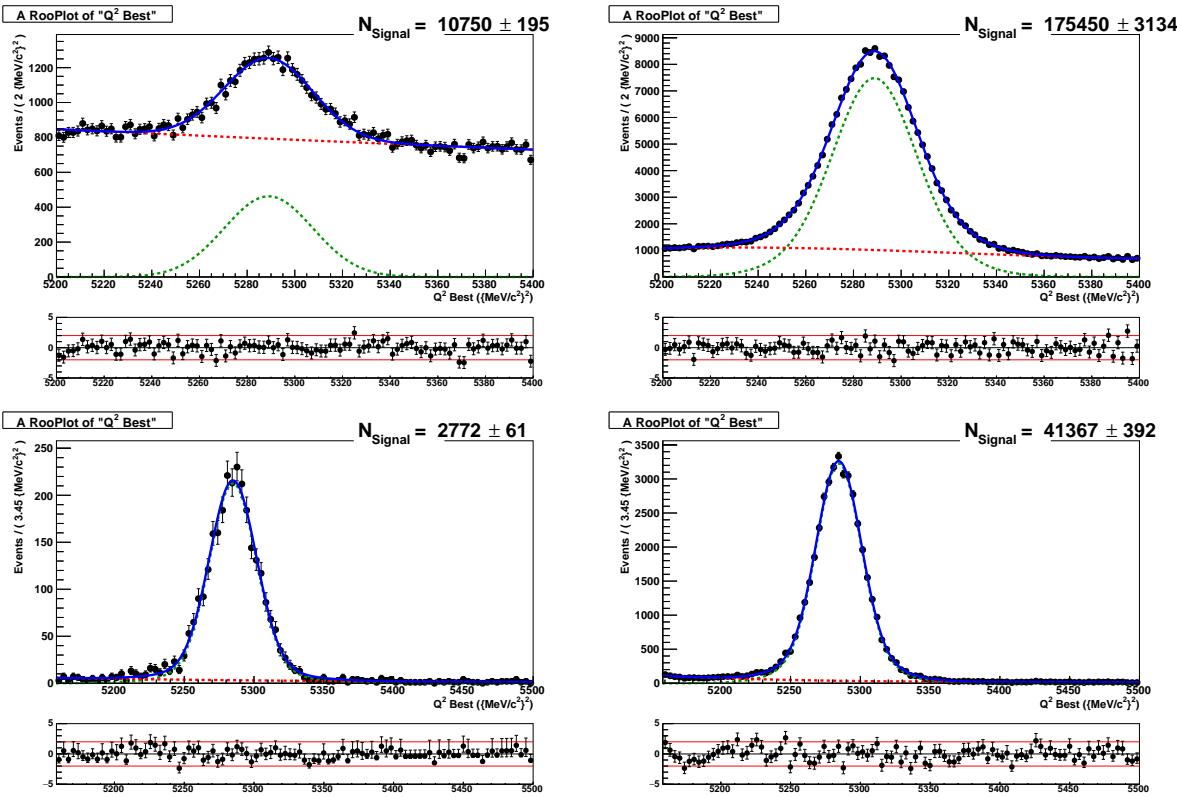


Figure 66: Fit to $B^+ \rightarrow J/\psi K^+$ peak in data using the isolation algorithm for low q^2 region (Top Left) and high q^2 region (Top right). The same fits are applied to B2CC simulation for low q^2 region (Bottom Left) and High q^2 region (Bottom right)

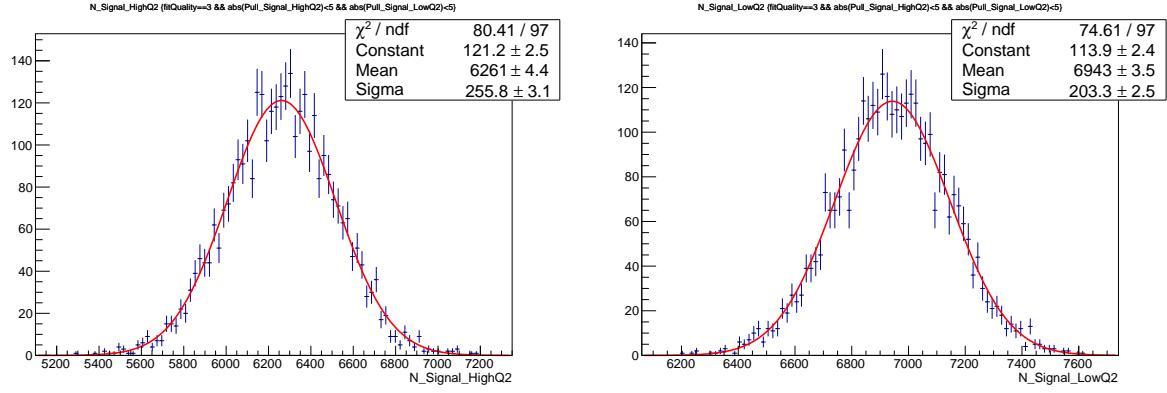


Figure 67: Distributions of signal yields for (left) high and (right) low q^2 regions using the alternative constrain on the B2CC yields

FF model	Yield range	mean	sigma
A.X.Khadra <i>et al.</i>	$q^2 > 7 \text{ GeV}^2/c^4$	6428.0 ± 4.4	257.8 ± 3.1
	$q^2 < 7 \text{ GeV}^2/c^4$	7512.9 ± 3.7	219.3 ± 2.6
Bouchard <i>et al.</i>	$q^2 > 7 \text{ GeV}^2/c^4$	6566.7 ± 4.5	264.0 ± 3.2
	$q^2 < 7 \text{ GeV}^2/c^4$	6738.1 ± 3.3	197.7 ± 2.4
Witzel <i>et al.</i>	$q^2 > 7 \text{ GeV}^2/c^4$	6449.9 ± 4.5	257.6 ± 3.2
	$q^2 < 7 \text{ GeV}^2/c^4$	7327.1 ± 3.7	213.7 ± 2.6
Rusov <i>et al.</i>	$q^2 > 7 \text{ GeV}^2/c^4$	6518.4 ± 4.5	261.1 ± 3.2
	$q^2 < 7 \text{ GeV}^2/c^4$	6948.1 ± 3.5	203.1 ± 2.5
ISGW2	$q^2 > 7 \text{ GeV}^2/c^4$	6554.3 ± 4.4	263.2 ± 3.1
	$q^2 < 7 \text{ GeV}^2/c^4$	6873.6 ± 3.4	201.4 ± 2.4

Table 40: Mean and width of the distributions of the signal yields obtained from toys using different Form Factor models. The ISGW2 model is mentioned for documentation and does not enter in the assessment of the systematics.

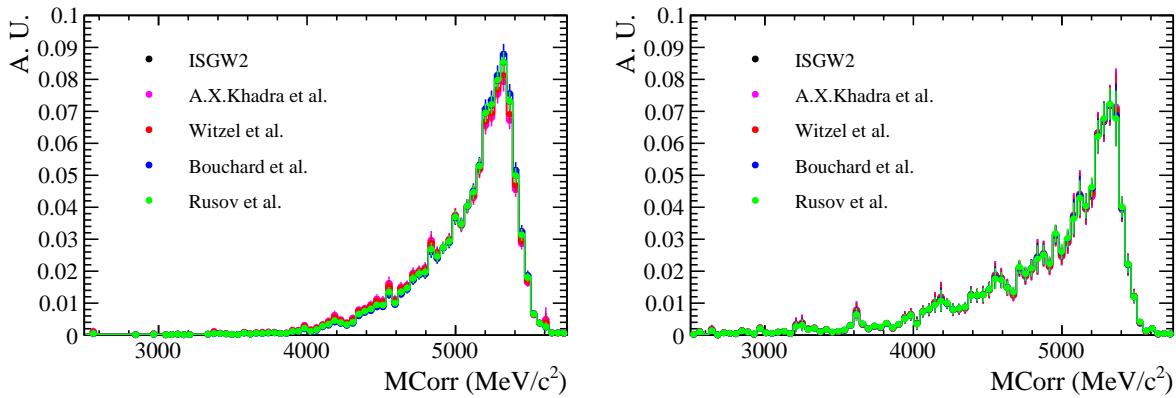


Figure 68: The m_{corr} distributions for $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ sample for $q^2 < 7 \text{ GeV}^2/c^4$ (left) and $q^2 > 7 \text{ GeV}^2/c^4$ (right), for the different Form Factor models.

Source	Yield range	Variation	Relative var. %
MisPID(SVD)	$q^2 > 7 \text{ GeV}^2/c^4$	-25.1	-0.4
	$q^2 < 7 \text{ GeV}^2/c^4$	+2.9	+0.04
MisPID(Y= $\pm 1\sigma$)	$q^2 > 7 \text{ GeV}^2/c^4$	-3.2(+4.3)	-0.05(+0.07)
	$q^2 < 7 \text{ GeV}^2/c^4$	+21.3(+28.5)	+0.3(+0.4)
K*(Equal W)	$q^2 > 7 \text{ GeV}^2/c^4$	+57.9	+0.9
	$q^2 < 7 \text{ GeV}^2/c^4$	-94.3	-1.3
K*(from phiK*)	$q^2 > 7 \text{ GeV}^2/c^4$	+33.7	+0.5
	$q^2 < 7 \text{ GeV}^2/c^4$	-74.8	-1.1
B2CC(Iso)	$q^2 > 7 \text{ GeV}^2/c^4$	-165.5	-2.6
	$q^2 < 7 \text{ GeV}^2/c^4$	+21.3	+0.3
Comb(Quad cut)	$q^2 > 7 \text{ GeV}^2/c^4$	-77.4	-1.2
	$q^2 < 7 \text{ GeV}^2/c^4$	+30.2	+0.4
FF(A.X.K.)	$q^2 > 7 \text{ GeV}^2/c^4$	+1.3	neg.
	$q^2 < 7 \text{ GeV}^2/c^4$	NA	NA
FF(Bouchard)	$q^2 > 7 \text{ GeV}^2/c^4$	+139.9	+2.2
	$q^2 < 7 \text{ GeV}^2/c^4$	NA	NA
FF(Rusov)	$q^2 > 7 \text{ GeV}^2/c^4$	NA	NA
	$q^2 < 7 \text{ GeV}^2/c^4$	+26.2	+0.4
FF(Witzel)	$q^2 > 7 \text{ GeV}^2/c^4$	+23.2	0.4
	$q^2 < 7 \text{ GeV}^2/c^4$	NA	NA
FF(ISGW2)	$q^2 > 7 \text{ GeV}^2/c^4$	+127.6	NA
	$q^2 < 7 \text{ GeV}^2/c^4$	-48.3	NA
Total	$q^2 > 7 \text{ GeV}^2/c^4$	—	+2.5/-2.9
	$q^2 < 7 \text{ GeV}^2/c^4$	—	+0.7/-1.7

Table 41: Absolute and relative variations of the signal yields due to the different sources of systematics. The ISGW2 model is mentioned for documentation and does not enter in the assessment of the systematics.

998 we get a negligible variation on the DsMuNu signal yield, $\sim \pm 100$ events.

999

1000 Background template grouping

1001

1002 The "grouping" of templates of physics backgrounds according to the similarity of
 1003 their m_{corr} shape is examined. A grouping different from the one performed in the base
 1004 DsMuNu fit is performed as follows: All $B \rightarrow D_s D$ decays are grouped in one template,
 1005 and $B \rightarrow D^{(*)}\tau\nu X$ is fitted as a single component. This approach is motivated by the
 1006 unique shape of the $B \rightarrow D^{(*)}\tau\nu X$ and similarities between $B \rightarrow D s D$ decays. The
 1007 pseudoexperiments performed with this configuration lead to an average and a σ for the
 1008 DsMuNu signal of 190757 ± 45 and 2474 ± 32 .

1009

1010 The systematic variation of the DsMuNu yields due to the alternative template
 1011 grouping is $-6.8k$ events (-3.4% of the signal yield).

1012

1013 FF influence

FF model	$B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$ mean ($\times 10^3$)	$B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$ sigma ($\times 10^3$)
HQET2	198.87 ± 0.03	1.92 ± 0.02
Monahan <i>et al.</i>	198.25 ± 0.03	1.91 ± 0.02
Bailey <i>et al.</i>	198.87 ± 0.03	1.92 ± 0.02

Table 42: Averages and means obtained in the pseudoexperiments for the $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$ yields with different Form Factor Models.

Uncertainty [%]	Relative uncertainty [%]			
	$B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$	$B_s^0 \rightarrow K^- \mu^+ \nu_\mu$		
		No q^2 sel.	$q^2 < 7$	$q^2 > 7$
Tracking	0.4 (+had)	0.1 (+had)	0.1 (+had)	0.1 (+had)
Trigger	1.0	1.0	1.0	1.3
Particle Identification	0.8	0.6	0.6	0.6
m_{corr} error		0.5	0.5	0.5
Isolation	0.2	0.1	0.1	0.1
Charged BDT		0.6	0.6	0.6
Same Sign BDT		1.1	1.1	1.1
q^2 migration			2	2
ε gen& reco	0.9	0.8	1.4	1.4
Fit template	1.7	+1.6/-2.3	+0.7/-1.7	+2.5/-2.9

Table 43: Systematic uncertainties on the evaluated yields at production N/ϵ for $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$ and $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$. “had” stands for the uncertainties due to hadronic interactions. When taking the ratio of branching fractions some of the systematic uncertainties will partially cancel, and when calculating the ratio of $|V_{ub}|/|V_{cb}|$ the uncertainties will be approximately halved.

1014
1015 The fit to the $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$ signal has been repeated with the different FF reweighting
1016 shown in Table 42, with the corresponding averages and widths of the signal yields.
1017 The maximum variation of the yield is $+1.4k$ events ($+0.7\%$ of the signal yield).
1018
1019 The total systematic uncertainty in the $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$ fit is the sum in quadrature of
1020 half the maximum variations of the signal yield due to the different sources, i.e. $\pm 1.7\%$.

1021 10.9 Final Systematic Uncertainties

1022 Table 43 summarizes the systematic uncertainties involved in the analysis. Table 44
1023 summarizes the systematic uncertainties for the measured ratio of branching fractions.
1024 For this latter table, the tracking efficiency accounts for the hadronic interactions since
1025 the corresponding uncertainty does not cancel fully due to the extra kaon and pion coming
1026 from the D_s decay as explained in subsection 10.4. The systematic uncertainty on $\frac{|V_{ub}|}{|V_{cb}|}$
1027 is easily inferred from the uncertainty on the branching fraction ratio as shown in the
1028 formulae of next chapter (11).

Uncertainty [%]	$\frac{\mathcal{B}(B_s^0 \rightarrow K\mu\nu)}{\mathcal{B}(B_s^0 \rightarrow D_s\mu\nu)}$		
	No q^2 sel.	$q^2 < 7$	$q^2 > 7$
Tracking	2.0	2.0	2.0
Trigger	1.4	1.2	1.6
Particle Identification	1	1	1
m_{corr} error	0.5	0.5	0.5
Isolation	0.2	0.2	0.2
Charged BDT	0.6	0.6	0.6
Same Sign BDT	1.1	1.1	1.1
q^2 migration		2	2
ε gen& reco	1.2	1.6	1.6
Fit template	(.) _{-2.9} ^{+2.3}	(.) _{-2.4} ^{+1.8}	(.) _{-3.4} ^{+3.0}
Total exp	(.) _{-4.3} ^{+4.0}	(.) _{-4.5} ^{+4.3}	(.) _{-5.3} ^{+5.0}
$\mathcal{B}(D_s^- \rightarrow K^- K^+ \pi^-)$	2.8	2.8	2.8

Table 44: Systematic uncertainties on the branching fraction ratio

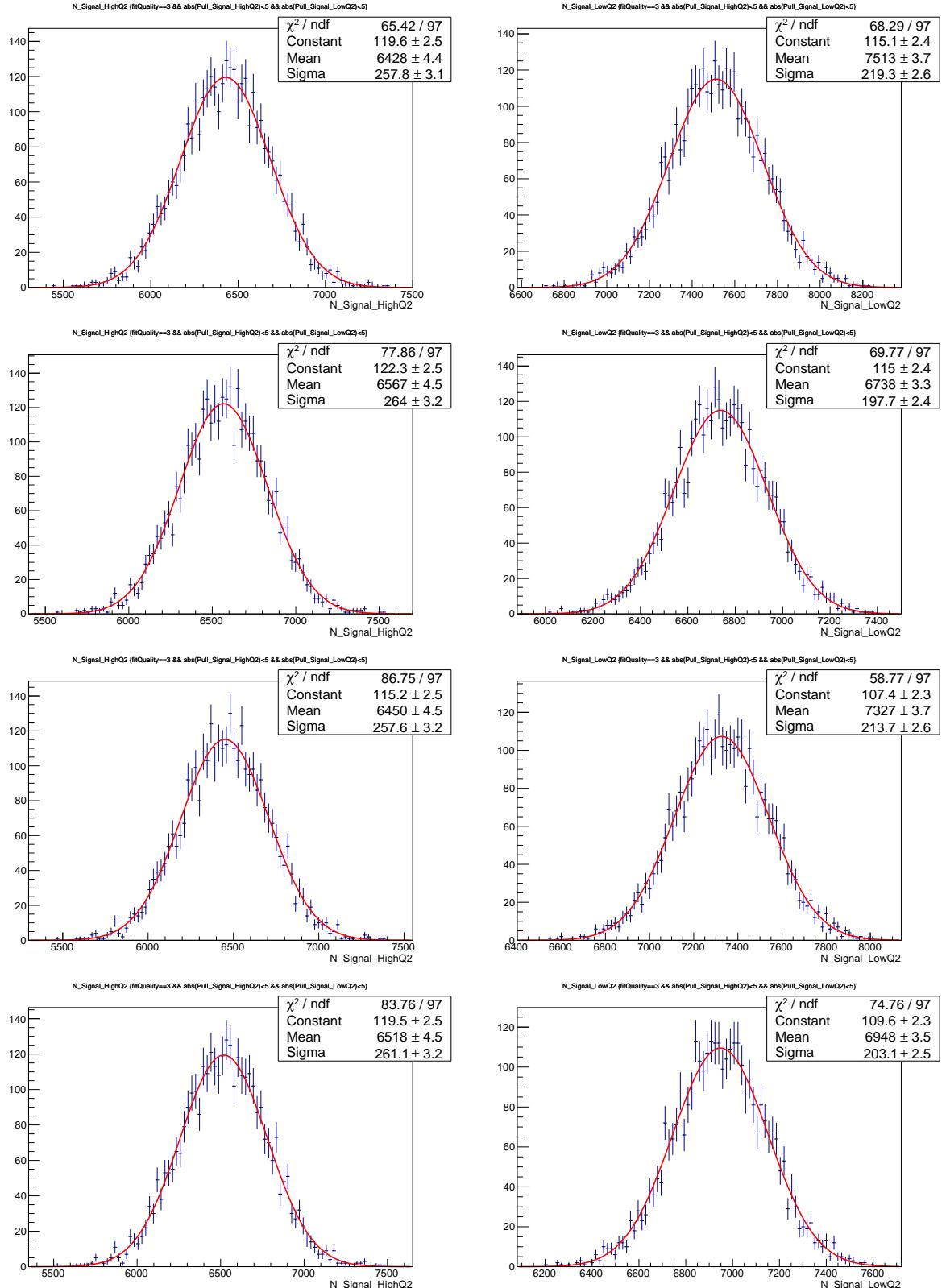


Figure 69: Distributions of the signal yields for (left) high and (right) low q^2 region for the different Form Factor models, from top to bottom: (LQCD) A.X.Khadra *et al.*, Bouchard *et al.*, Witzel *et al.* and (LCSR) Rusov *et al.*.

1029 11 Extraction of the branching fraction and $|V_{ub}|$

1030 11.1 Branching fractions

1031 The ratio of branching fractions is determined as:

$$1032 R_{BF} = \frac{\mathcal{B}(B_s^0 \rightarrow K\mu\nu, \text{ bin})}{\mathcal{B}(B_s^0 \rightarrow D_s\mu\nu)} = \frac{N_K}{N_{D_s}} \times \frac{\epsilon_{D_s}}{\epsilon_K(\text{bin})} \times \mathcal{B}(D_s^+ \rightarrow K^+K^-\pi^+) \quad (33)$$

1032 Where N_i are the fitted yields and ϵ_i the overall efficiencies for the signal and normalization
 1033 channels; The ratio is computed for $\text{bin} = q^2 < 7 \text{ GeV}^2/c^4$ or low, $q^2 > 7 \text{ GeV}^2/c^4$ or high,
 1034 and integrated (all). Using the fit yields shown in Secs.9.5,8.3, the efficiency ratio of
 1035 Sec.10.7 and $\mathcal{B}(D_s^+ \rightarrow K^+K^-\pi^+) = (5.39 \pm 0.15)\%$ [49], we obtain:

$$\begin{aligned} 1036 R_{BF}(\text{low}) &= (1.66 \pm 0.08(\text{stat}) \pm 0.07(\text{syst}) \pm 0.05(D_s)) \times 10^{-3}, \\ 1037 R_{BF}(\text{high}) &= (3.25 \pm 0.21(\text{stat})^{+0.16}_{-0.17}(\text{syst}) \pm 0.09(D_s)) \times 10^{-3}, \\ 1038 R_{BF} &= (4.89 \pm 0.21(\text{stat})^{+0.20}_{-0.21}(\text{syst}) \pm 0.14(D_s)) \times 10^{-3}, \end{aligned}$$

1039 where (D_s) stands for the error on the D_s decay branching fraction.

1040 The absolute branching fraction is derived using:

$$R_{BF} = \frac{\mathcal{B}(B_s^0 \rightarrow K\mu\nu, \text{ bin})}{\tau_{B_s} \times |V_{cb}|^2 \times FF_{D_s^+}} \quad (34)$$

1041 which gives:

$$\mathcal{B}(B_s^0 \rightarrow K\mu\nu, \text{ bin}) = R_{BF} \times \tau_{B_s} \times |V_{cb}|^2 \times FF_{D_s^+} \quad (35)$$

1042 Using the exclusive value of $|V_{cb}| = (39.5 \times 0.9) \times 10^{-3}$ [49], the lifetime $\tau_{B_s} =$
 1043 $1.515 \pm 0.004 \text{ ps}$ [49] and $FF_{D_s} = 9.15 \pm 0.37 \text{ ps}^{-1}$ [8] (also section 2.1.3, Table 2):

$$\begin{aligned} 1044 \mathcal{B}(B_s^0 \rightarrow K\mu\nu, q^2 < 7 \text{ GeV}^2/c^4) &= \\ 1045 (0.36 \pm 0.02(\text{stat}) \pm 0.02(\text{syst}) \pm 0.02(\text{expinp}) \pm 0.01(\text{FF})) \times 10^{-4}, \\ 1046 \mathcal{B}(B_s^0 \rightarrow K\mu\nu, q^2 > 7 \text{ GeV}^2/c^4) &= \\ 1047 (0.70 \pm 0.04(\text{stat})^{+0.03}_{-0.04}(\text{syst}) \pm 0.04(\text{expinp}) \pm 0.03(\text{FF})) \times 10^{-4}, \\ 1048 \mathcal{B}(B_s^0 \rightarrow K\mu\nu) &= (1.06 \pm 0.05(\text{stat}) \pm 0.04(\text{syst}) \pm 0.06(\text{expinp}) \pm 0.04(\text{FF})) \times 10^{-4}, \end{aligned}$$

1049 where (expinp) is inclusive of the uncertainty on all the experimental inputs such as the
 1050 D_s decay branching fraction, the lifetime and $|V_{cb}|^2$.

1051 11.2 $|V_{ub}|$

1052 The ratio of branching fractions can be written $R_{BF} = \frac{|V_{ub}|^2}{|V_{cb}|^2} \times \frac{FF_K}{FF_{D_s^+}}$. From this one infers:

$$\frac{|V_{ub}|}{|V_{cb}|} = \sqrt{R_{BF} \times \frac{FF_{D_s^+}}{FF_K}} \quad (36)$$

1053 Our model choice implies $FF_K(q^2 < 7 \text{ GeV}^2/c^4) = 4.14 \pm 0.38 \text{ ps}^{-1}$ [13], $FF_K(q^2 >$
 1054 $7 \text{ GeV}^2/c^4) = 3.32 \pm 0.46 \text{ ps}^{-1}$ [12] (also section 2.1.3, Table 1). The results are:

1055 $\frac{|V_{ub}|}{|V_{cb}|}(q^2 < 7 \text{ GeV}^2/c^4) = 0.0607 \pm 0.0015(\text{stat}) \pm 0.0013(\text{syst}) \pm 0.0008(D_s) \pm 0.0030(FF)$

1056 $\frac{|V_{ub}|}{|V_{cb}|}(q^2 > 7 \text{ GeV}^2/c^4) = 0.0946 \pm 0.0030(\text{stat})^{+0.0024}_{-0.0025}(\text{syst}) \pm 0.0013(D_s) \pm 0.0068(FF)$

1057 Combining the uncertainties, one obtains:

1058 $\frac{|V_{ub}|}{|V_{cb}|}(q^2 < 7 \text{ GeV}^2/c^4) = 0.0607 \pm 0.0037$

1059 $\frac{|V_{ub}|}{|V_{cb}|}(q^2 > 7 \text{ GeV}^2/c^4) = 0.0946 \pm 0.0079$

1060 11.3 Discussion

1061 The branching fraction measurement $\mathcal{B}(B_s^0 \rightarrow K\mu\nu) = (1.06 \pm 0.05(\text{stat})) \pm 0.04(\text{syst}) \pm$
 1062 $0.06(\text{expinp}) \pm 0.04(\text{FF}) \times 10^{-4}$ or $\mathcal{B}(B_s^0 \rightarrow K\mu\nu) = (1.06 \pm 0.10) \times 10^{-4}$ if all the errors
 1063 are combined, is comparable to the average $\mathcal{B}(B^0 \rightarrow \pi^-\ell^+\nu) = (1.50 \pm 0.06) \times 10^{-4}$ [49].

1064 Figure 70 shows the $(|V_{cb}|, |V_{ub}|)$ plane with the PDG averages of $|V_{cb}|$ and $|V_{ub}|$
 1065 overlayed, as well as the $\frac{|V_{ub}|}{|V_{cb}|}$ measurement from the LHCb study $\Lambda_b \rightarrow p\mu\nu$ and from our
 1066 analysis where the form factor uncertainty has been included. The discrepancy between
 1067 low and high q^2 bands is completely driven by the theoretical Form Factor calculations.

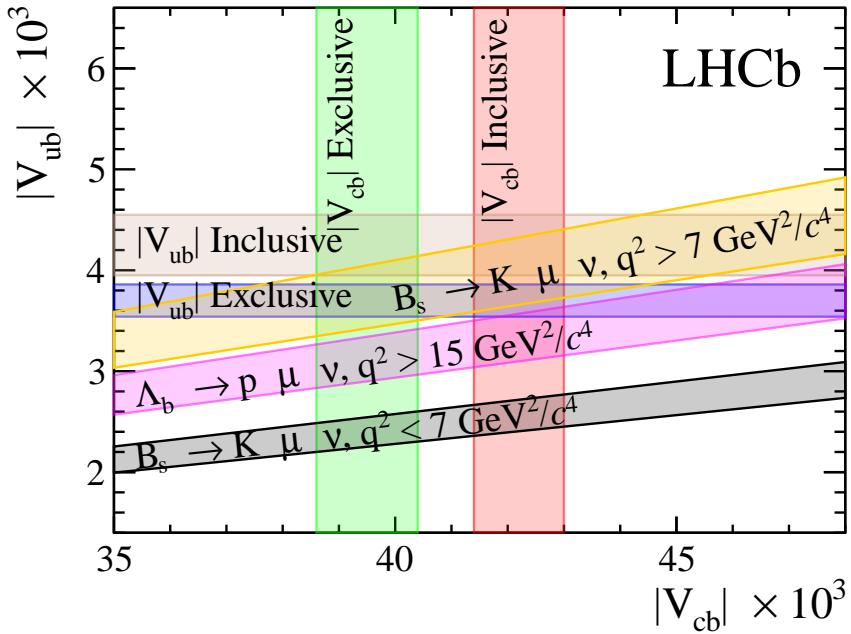


Figure 70: PDG averages (based on exclusive $B \rightarrow \pi\ell\nu$, $B \rightarrow D^*\ell\nu$ and inclusive $B \rightarrow X_{u,c}\ell\nu$ results) and measurements in the $(|V_{cb}|, |V_{ub}|)$ plane.

1068 Appendix

A Form Factor Comparisons

This chapter summarises the results of $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ and $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$ form factor calculations and compares plots presented in the published papers with those generated by the analysis software using results taken from the papers. This is to ensure that results taken from theory have been reproduced accurately and that there are no errors from copying tables of numbers.

This chapter contains

- Names and references of publication used
- Fitted parameters of the z -expansions
- Selected reproductions of plots verifying analysis software

Title	Authors	arXiv
$B_s \rightarrow Kl\nu$ form factors from lattice QCD	C.M. Bouchard, G.Peter Lepage, Christopher Monahan, Heechang Na, Junko Shigemitsu	arXiv:1406.2279v2 [10]
$B \rightarrow \pi l\nu$ and $B_s \rightarrow Kl\nu$ form factors and $ V_{ub} $ from 2 + 1-flavor lattice QCD with domain-wall light quarks and relativistic heavy quarks	J.M. Flynn T. Izubuchi T. Kawanai C. Lehner A. Soni R.S. Van de Water O. Witzel	arXiv:1501.05373v3 [11]
$B_s \rightarrow Kl\nu_l$ and $B_{(s)} \rightarrow \pi(K)l^+l^-$ decays at large recoil and CKM matrix elements	Alexander Khodjamirian Aleksey V. Rusov	arXiv:1703.04765v2 [13]

Table 45: Details of the papers providing form factor results for $B_s^0 \rightarrow K^-\mu^+\nu_\mu$

Title	Authors	arXiv	
$B_s \rightarrow D_s l\nu$ form factors and the fragmentation fraction ratio f_s/f_d .	Christopher J. Monahan Heechang Na Chris M. Bouchard G. Peter Lepage Junko Shigemitsu	arXiv:1703.09728v1 [7]	
$B_s \rightarrow D_s/B \rightarrow D$ semileptonic form-factor ratios and their application to $\text{BR}(B_s^0 \rightarrow \mu^+\mu^-)$	Jon A. Bailey C. Bernard C. DeTar A.X. El-Khadra E.D. Freeland Steven Gottlieb Jongjeong Kim J. Laiho P.B. Mackenzie E. Neil Si-Wei Qiu R. Sugar R.S. Van de Water	A. Bazavov C.M. Bouchard Daping Du J. Foley E. Gamiz U.M. Heller A.S. Kronfeld L. Levkova Y. Meurice M.B. Oktay J.N. Simone D. Toussaint Ran Zhou	arXiv:1202.6346v2 [6]

Table 46: Details of the papers providing form factor results for $B_s^0 \rightarrow D_s^-\mu^+\nu_\mu$

1079 A.1 Publications Used

1080 $B_s^0 \rightarrow K^-\mu^+\nu_\mu$

1081 $B_s^0 \rightarrow D_s^-\mu^+\nu_\mu$

1082 A.2 z -expansion Fit Parameters

1083 $B_s^0 \rightarrow K^-\mu^+\nu_\mu$

1084 **Bouchard et.al.**

1085 **Flynn et.al.**

	$b_1^{(0)}$	$b_2^{(0)}$	$b_3^{(0)}$	$b_1^{(+)}$	$b_2^{(+)}$	$b_3^{(+)}$
Value	0.31500	0.9450	2.3910	0.368000	-0.7500	2.7200
Error	0.12900	1.3050	4.6710	0.021400	0.1930	1.4580
$b_1^{(0)}$	0.01676	0.1462	0.4453	0.001165	0.0214	0.1434
$b_2^{(0)}$	0.14620	1.7020	5.8520	0.009481	0.2255	1.5390
$b_3^{(0)}$	0.44530	5.8520	21.810	0.029630	0.7472	5.3250
$b_1^{(+)}$	0.00117	0.0095	0.0296	0.000458	0.0012	-0.0013
$b_2^{(+)}$	0.02140	0.2255	0.7472	0.001157	0.0372	0.1858
$b_3^{(+)}$	0.14340	1.5390	5.3250	-0.001309	0.1858	2.1240

Table 47: Extrapolated coefficients of a HPChPT z expansion for the $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ form factors with the associated covariance matrix. Results taken from [10].

	$b_{(+)}^0$	$b_{(+)}^1$	$b_{(+)}^2$	$b_{(0)}^0$	$b_{(0)}^1$	$b_{(0)}^2$
Value	0.338	-1.161	-0.458	0.210	-0.169	-1.235
Error	0.024	0.192	1.009	0.024	0.202	0.880
$b_{(+)}^0$	1.000	0.255	0.146	0.873	0.603	0.423
$b_{(+)}^1$	0.255	1.000	0.823	0.311	0.954	0.770
$b_{(+)}^2$	0.146	0.823	1.000	0.346	1.060	0.901
$b_{(0)}^0$	0.873	0.311	0.346	1.000	0.556	0.479
$b_{(0)}^1$	0.603	0.954	1.060	0.556	1.000	0.965
$b_{(0)}^2$	0.423	0.770	0.901	0.479	0.965	1.000

Table 48: Central values, errors, and correlation matrix for the BCL z -parametrisations of f_+ and f_0 for $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$. Results taken from [11].

1086 **Khodjamirian and Rusov**

1087 $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$

1088 **Monahan et.al.**

1089 **Bailey et.al.**

1090 A.3 Comparison Plots

1091 $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$

	$f_{BP}(0)$	$b_{1(BP)}$	Correlation
f_+	0.336(23)	-2.53(1.17)	0.79
f_0	0.320(19)	-1.08(1.53)	0.74

Table 49: Central values, errors, and correlations for the BCL z -parametrisations of f_+ and f_0 for $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$. Results taken from [13].

	$a_0^{(0)}$	$a_1^{(0)}$	$a_2^{(0)}$	$a_0^{(+)}$	$a_1^{(+)}$	$a_2^{(+)}$
Value	0.663	-0.10	1.3	0.868	-3.35	0.6
Error	0.031	0.30	2.8	0.032	0.41	4.7
$a_0^{(0)}$	0.0009534	-0.00303547	-0.00542391	0.000594503	0.00158251	0.0160091
$a_1^{(0)}$	0.00303547	0.0903097	-0.101760	0.000446248	0.0236283	0.0456659
$a_2^{(0)}$	0.00542391	-0.101760	8.02283	0.00848079	0.104246	0.760797
$a_0^{(+)}$	0.000594503	0.000446248	0.00848079	0.00100761	-0.00423358	-0.0264511
$a_1^{(+)}$	0.00158251	0.0236283	0.104246	-0.00423358	0.165251	-0.617234
$a_2^{(+)}$	0.0160091	0.0456659	0.760797	-0.0264511	-0.617234	22.49292

Table 50: Central values, errors, and covariance matrix for the z -parametrisations of f_+ and f_0 for $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$. Results taken from [7].

	$a_0^{(+)}$	$a_1^{(+)}$	$a_2^{(+)}$	$a_0^{(0)}$	$a_1^{(0)}$	$a_2^{(0)}$
Value	0.01191	-0.111	0.47	0.01081	-0.0662	0.18
Error	0.00006	0.002	0.05	0.00004	0.0002	0.06
$a_0^{(+)}$	1.0	-0.055	-0.002	0.593	0.254	0.014
$a_1^{(+)}$	-0.055	1.0	-0.318	-0.067	0.867	-0.180
$a_2^{(+)}$	-0.002	-0.318	1.0	-0.038	-0.307	0.974
$a_0^{(-)}$	0.593	-0.067	-0.038	1.000	-0.050	-0.054
$a_1^{(-)}$	0.254	0.867	-0.307	-0.050	1.000	-0.233
$a_2^{(-)}$	0.014	-0.180	0.974	-0.054	-0.233	1.000

Table 51: Central values, errors, and correlation matrix for the three term z -parametrisations of f_+ and f_0 for $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$. Results taken from [6].

1092 **Bouchard et.al.**

1093 **Flynn et.al.**

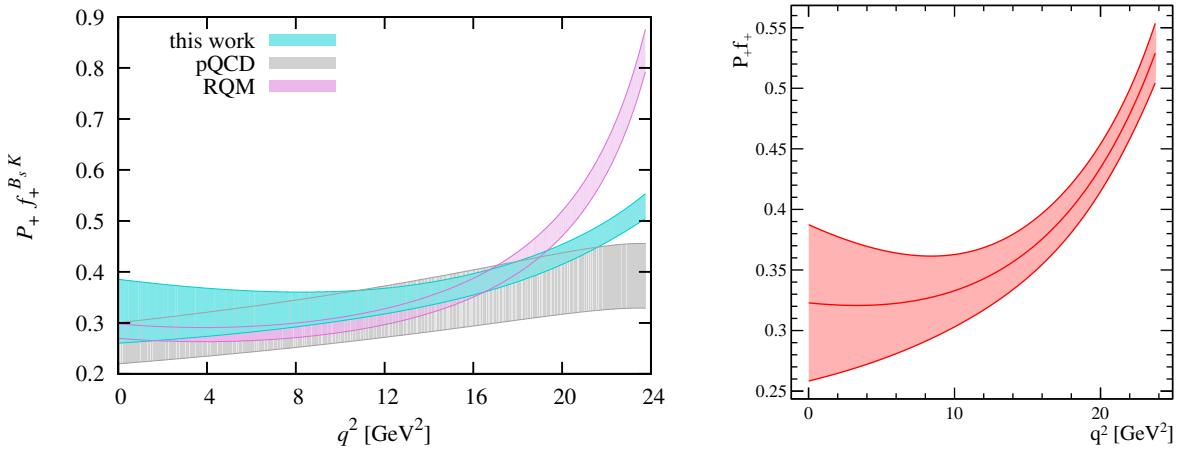


Figure 71: Form factors plotted against z . Image, left, taken from [10] and right, generated using fit parameters taken from [10]. The blue shaded section (left) should be compared to the red section (right).

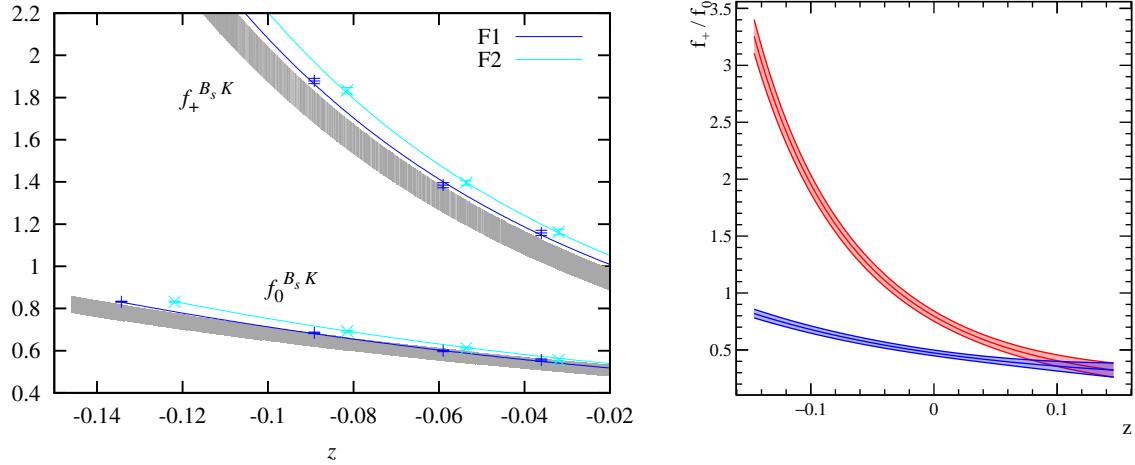


Figure 72: Form factors plotted against z . Image, left, taken from [10] and right, generated using fit parameters taken from [10].

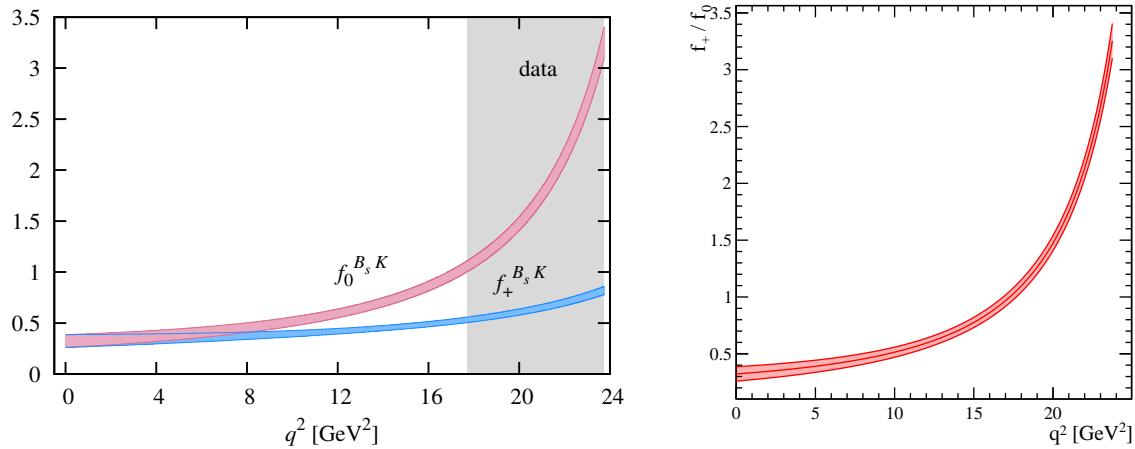


Figure 73: Form factors plotted against q^2 . Image, left, taken from [10] and right, generated using fit parameters taken from [10].

₁₀₉₄ **Khodjamirian and Rusov**

₁₀₉₅ $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$

₁₀₉₆ **Monahan et.al.**

₁₀₉₇ **Bailey et.al.**

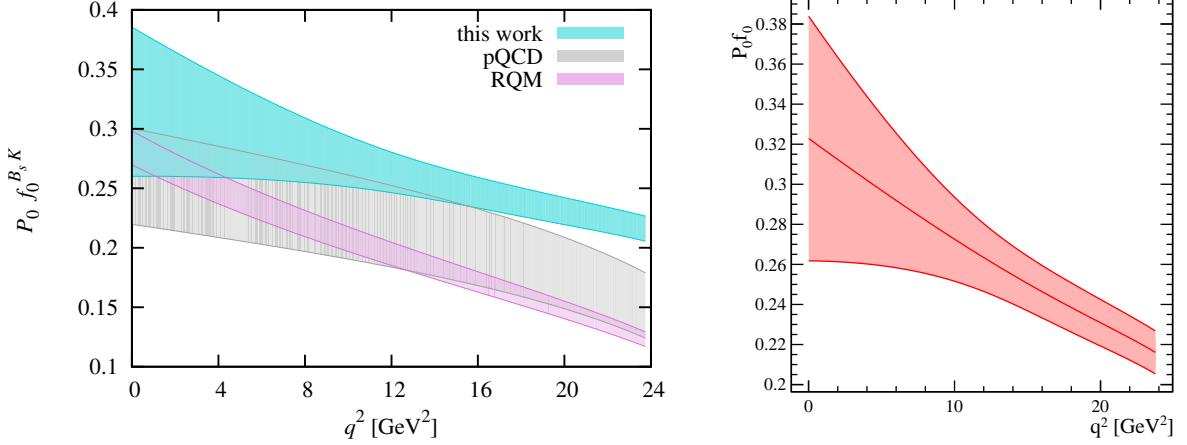


Figure 74: Form factors plotted against q^2 . Image, left, taken from [10] and right, generated using fit parameters taken from [10]. The blue shaded section (left) should be compared to the red section (right).

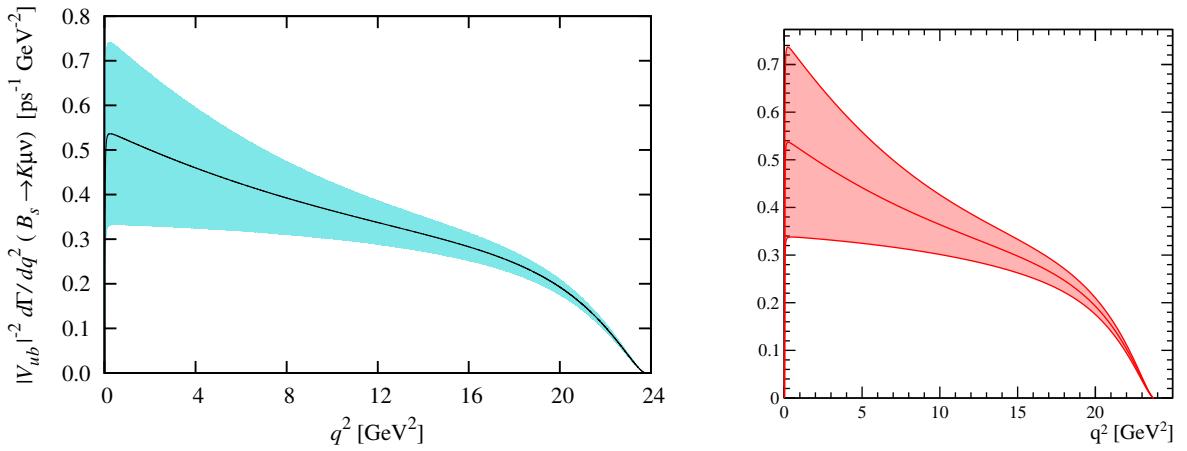


Figure 75: The differential $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ decay rate plotted against q^2 . Image, left, taken from [10] and right, generated using fit parameters taken from [10].

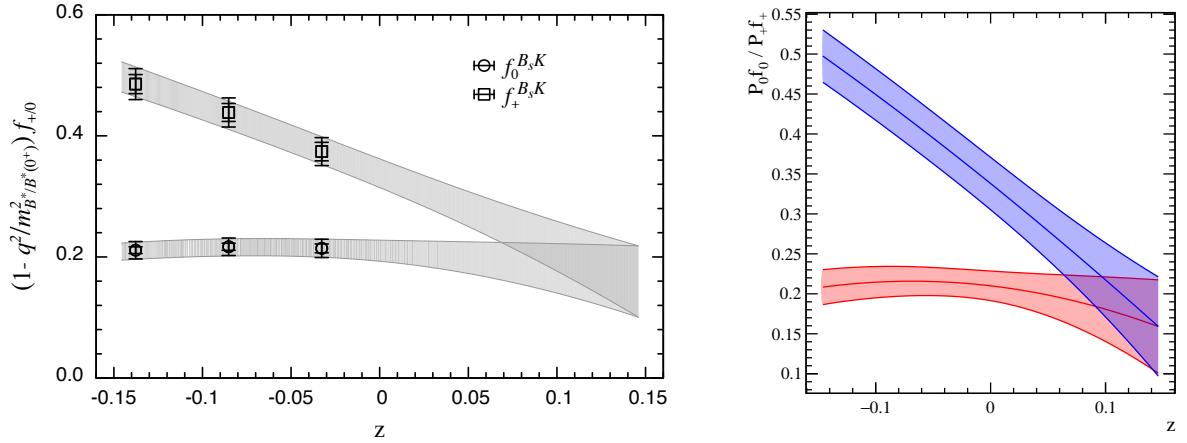


Figure 76: Form factors plotted against z . Image, left, taken from [11] and right, generated using fit parameters taken from [11].

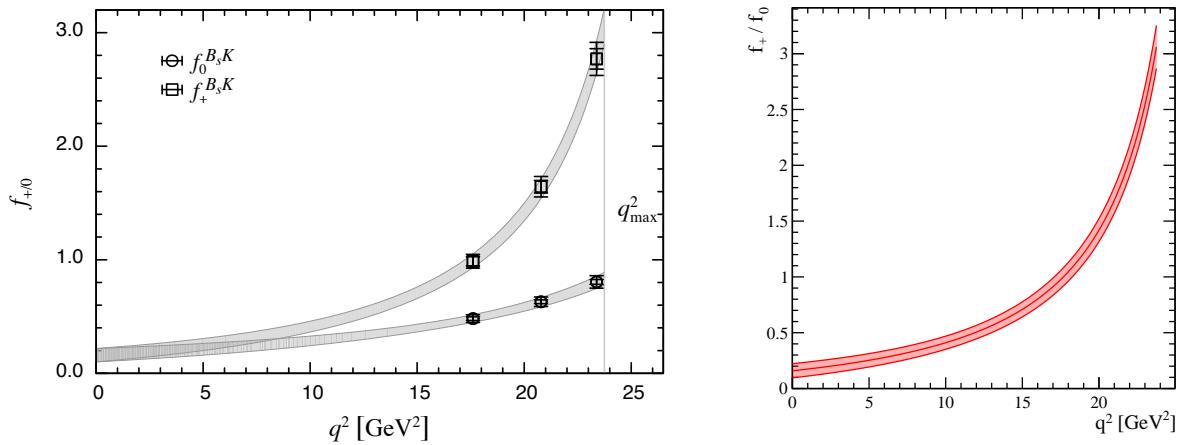


Figure 77: Form factors plotted against q^2 . Image, left, taken from [11] and right, generated using fit parameters taken from [11].

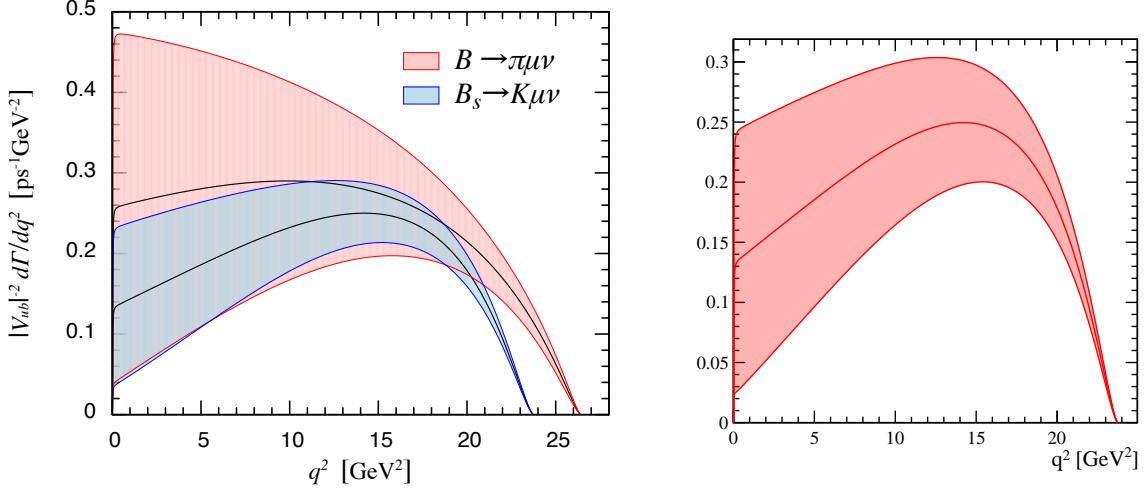


Figure 78: The differential $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ decay rate plotted against q^2 . Image, left, taken from [11] and right, generated using fit parameters taken from [11].

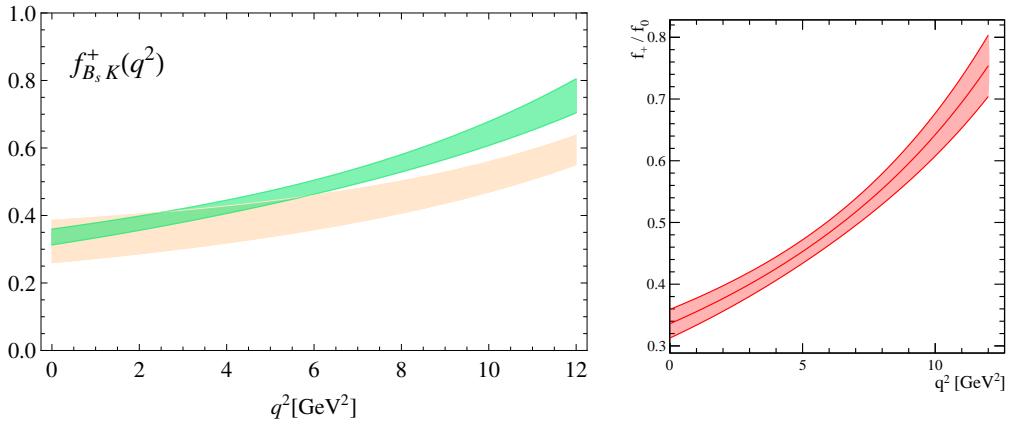


Figure 79: Form factors plotted against q^2 . Image, left, taken from [13] and right, generated using fit parameters taken from [13]. The green shaded region (left) should be compared to the red shaded region (right).

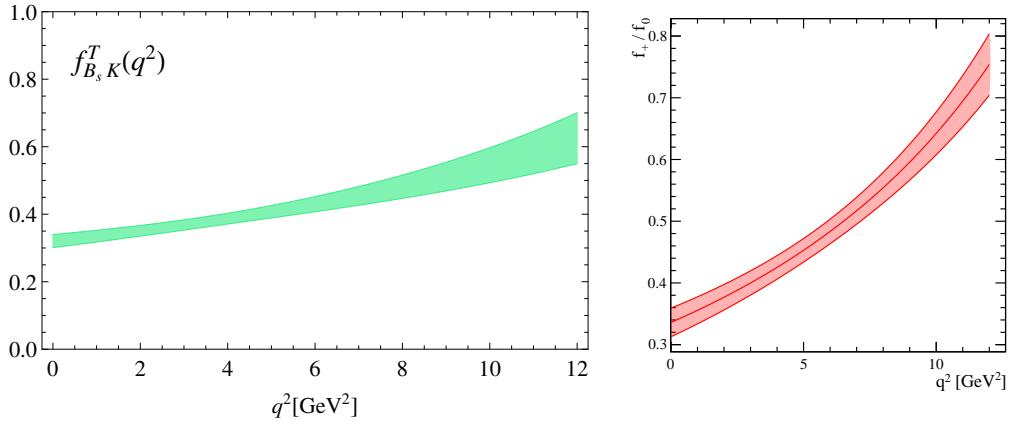


Figure 80: Form factors plotted against q^2 . Image, left, taken from [13] and right, generated using fit parameters taken from [13]. The green shaded region (left) should be compared to the blue shaded region (right).

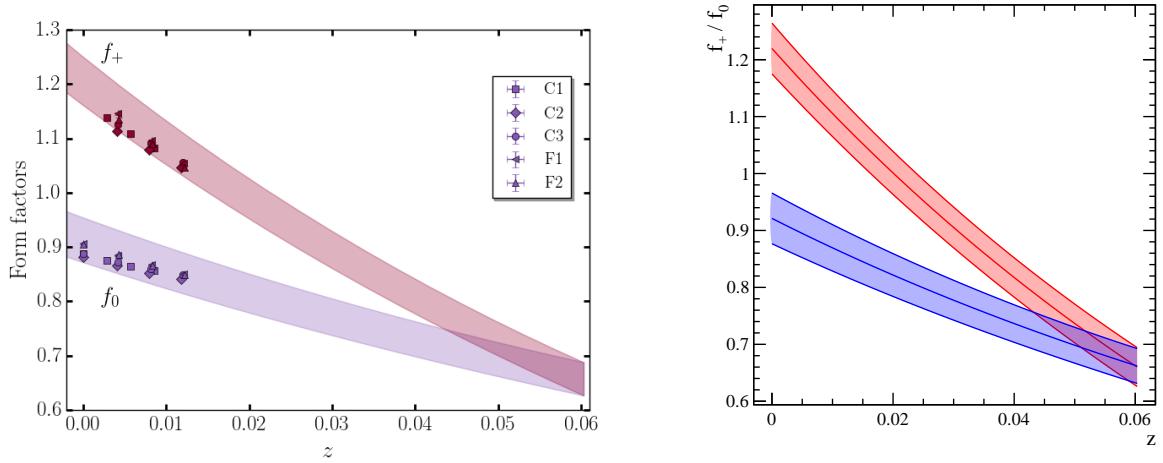


Figure 81: Form factors plotted against z . Image, left, taken from [7] and right, generated using fit parameters taken from [7].

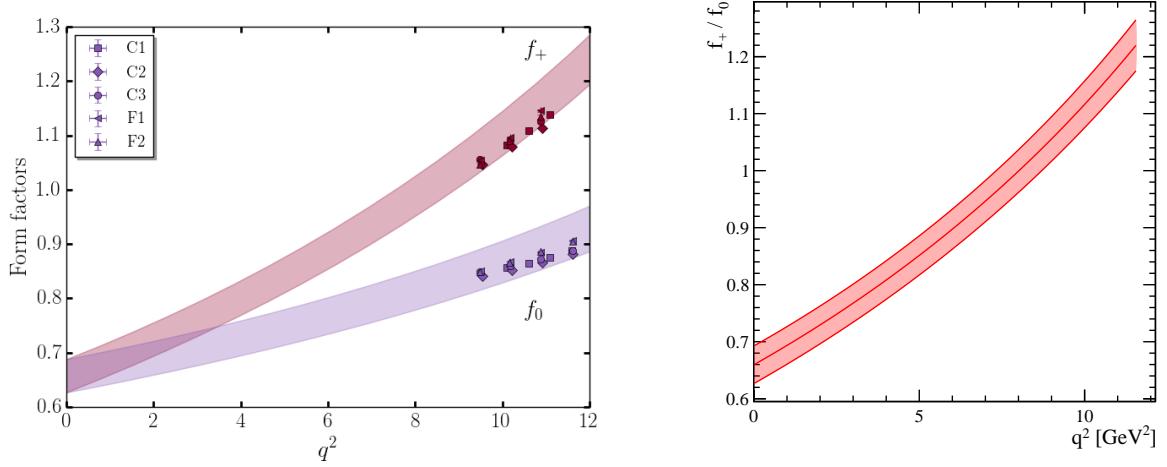


Figure 82: Form factors plotted against q^2 . Image, left, taken from [7] and right, generated using fit parameters taken from [7].

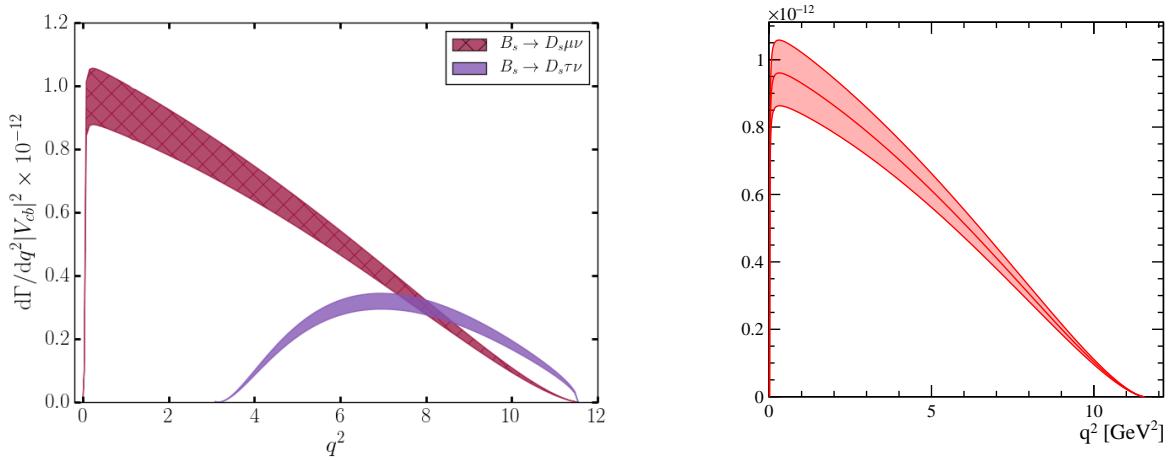


Figure 83: The differential $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$ decay rate plotted against q^2 . Image, left, taken from [7] and right, generated using fit parameters taken from [7].

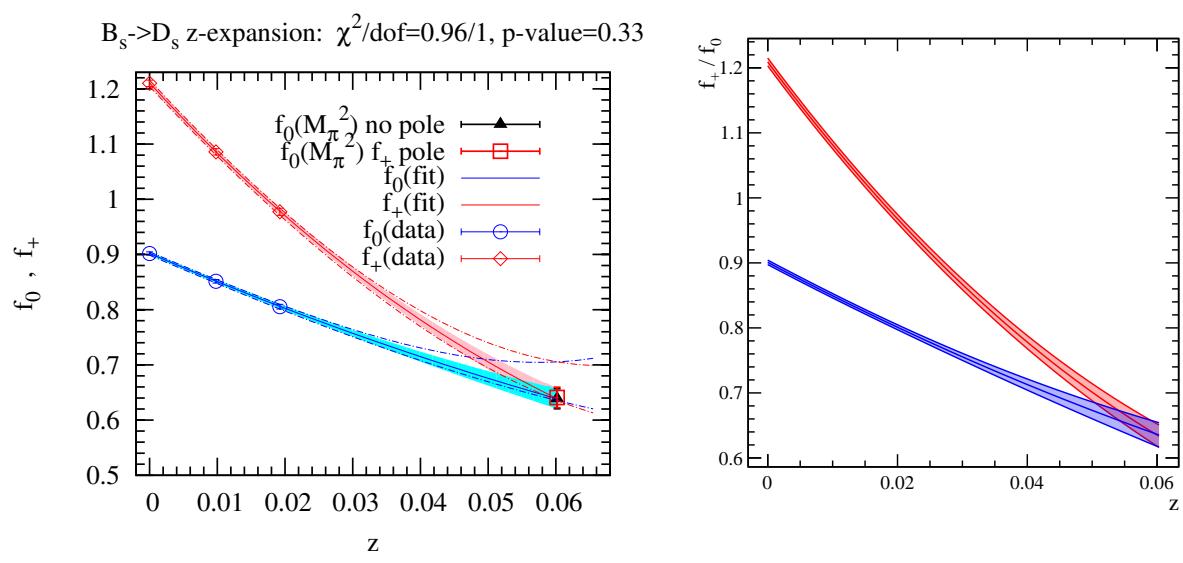


Figure 84: Form factors plotted against z . Image, left, taken from [6] and right, generated using fit parameters taken from [6].

1098 **B Input distributions of the variables used in the**
 1099 **BDTs**

1100 **B.1 Charged BDT variables: signal vs background comparison**

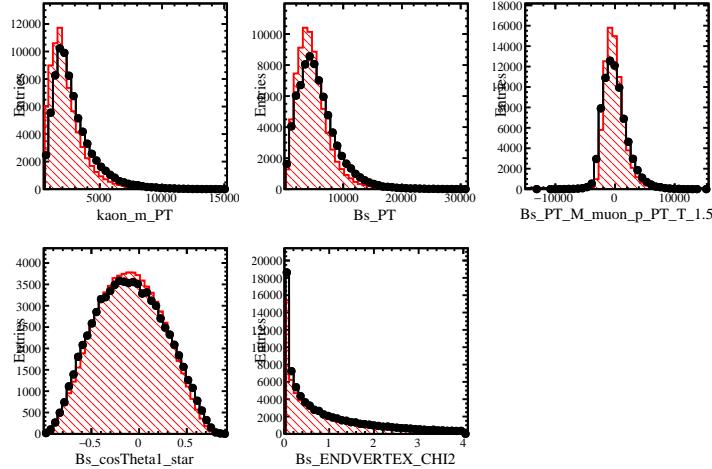


Figure 85: First set of variables used to train the Charged BDT. Signal is represented by black points and background in hatched red histogram.

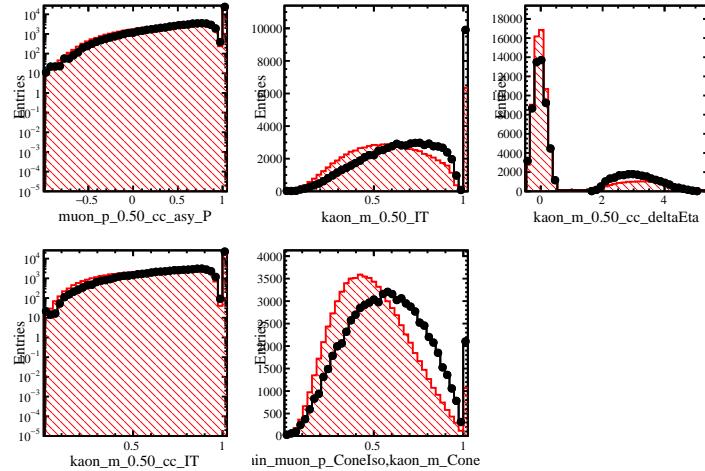


Figure 86: Second set of variables used to train the Charged BDT. Signal is represented by black points and background in hatched red histogram.

1101 **B.2 Same Sign BDT variables: signal vs background compari-**
 1102 **son**

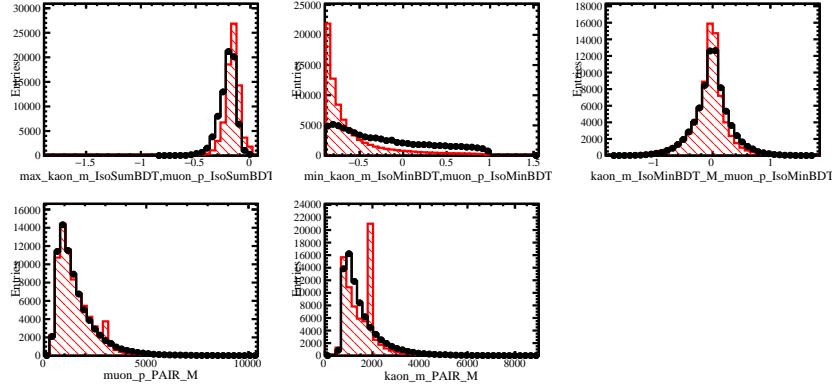


Figure 87: Third set of variables used to train the Charged BDT. Signal is represented by black points and background in hatched red histogram.

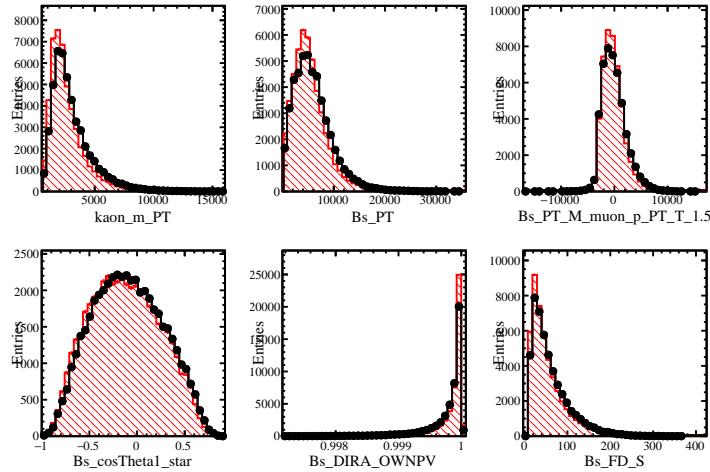


Figure 88: First set of variables used to train the SS BDT. Signal is represented by black points and background in hatched red histogram.

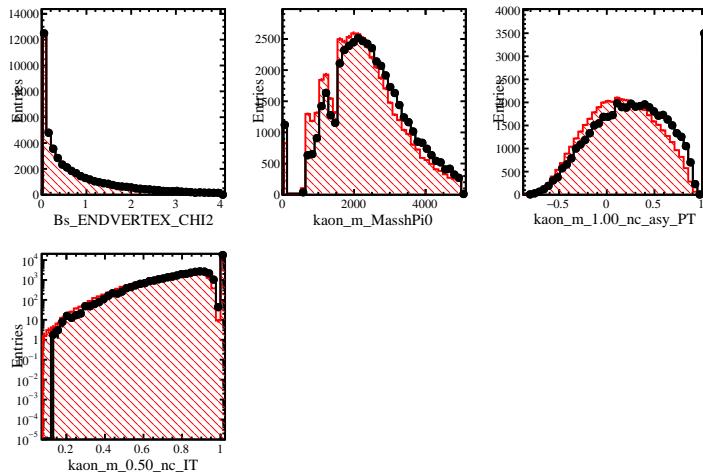


Figure 89: Second set of variables used to train the SS BDT. Signal is represented by black points and background in hatched red histogram.

₁₁₀₃ C Validation of Combinatoric Modelling

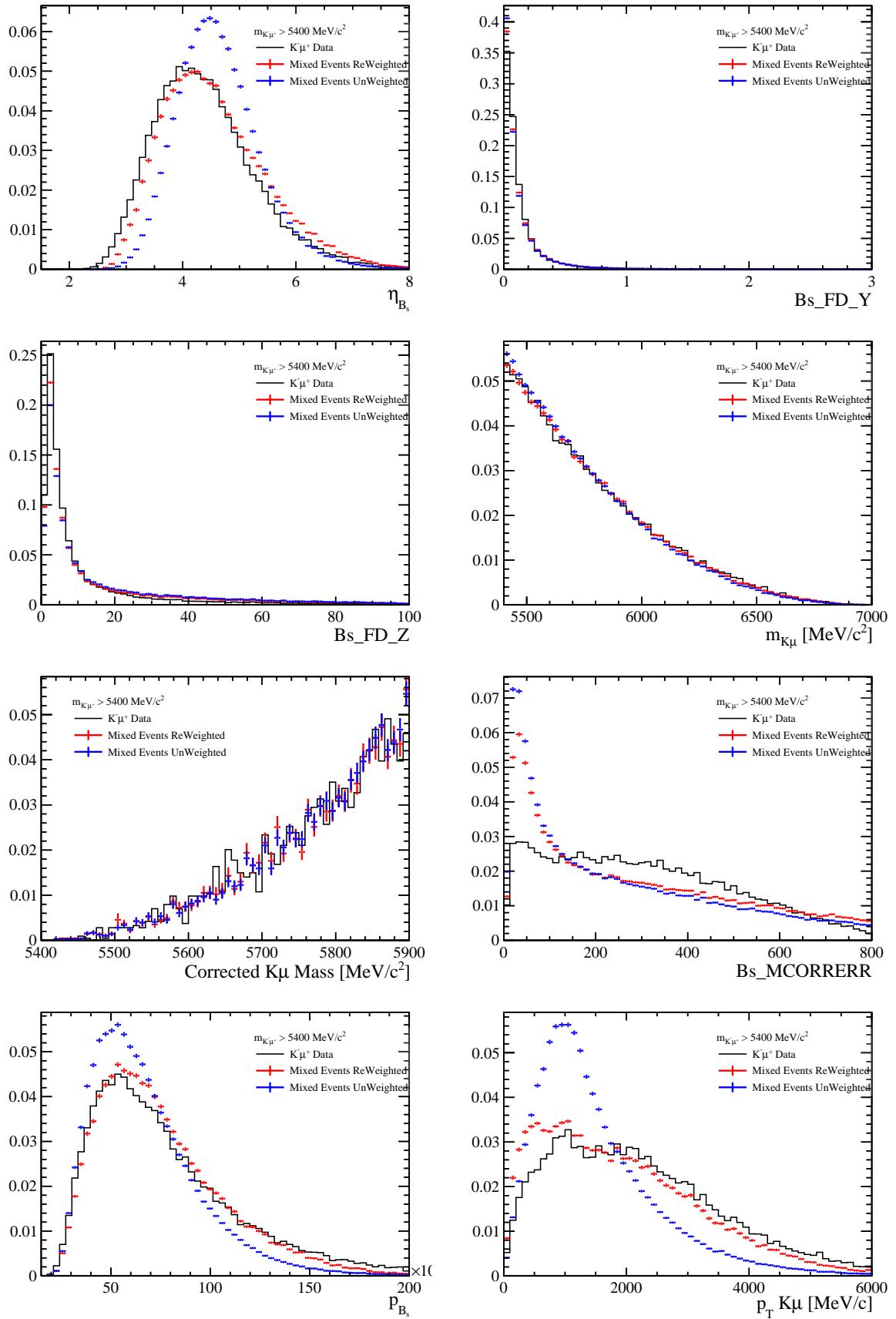


Figure 90: $K^- \mu^+$ candidates in data are plotted with simulated combinatorics before and after a kinematic correction

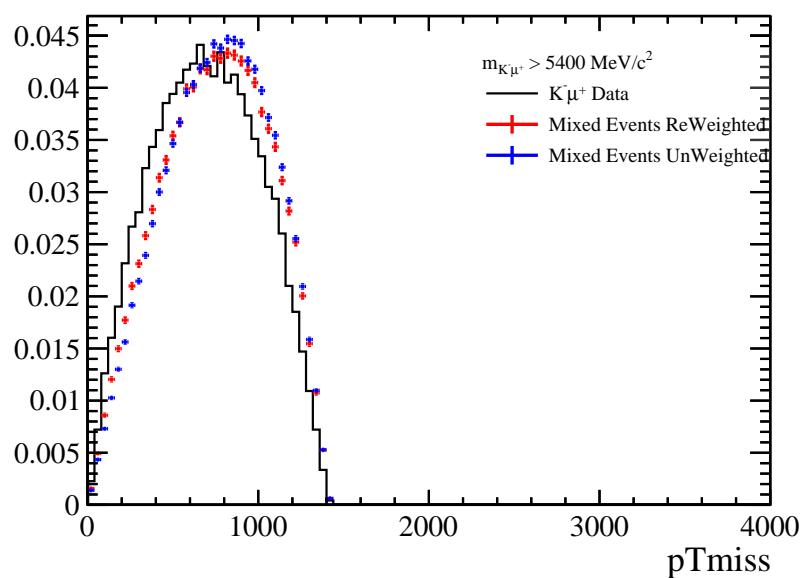


Figure 91: p_\perp for $K^- \mu^+$ candidates in data, plotted with simulated combinatorics before and after a kinematic correction

1104 **D Study of the elements of the MisID matrices**

1105 **D.1 Elements as a function of the kinematics of the track**

1106 Tracks of the fakeK $B_s^0 \rightarrow fakeK\mu$ and fakeMu $B_s^0 \rightarrow KfakeMu$ lines are used to
 1107 study the elements of the misID/PID matrices. Figures 92, 93, 94, 95 show the profile
 1108 of the elements $P(\hat{i}|j)$ ($j \rightarrow i$) as a function of momentum and pseudorapidity. The
 1109 discontinuities correspond to the definition of the binning for the PIDCalib look up tables.

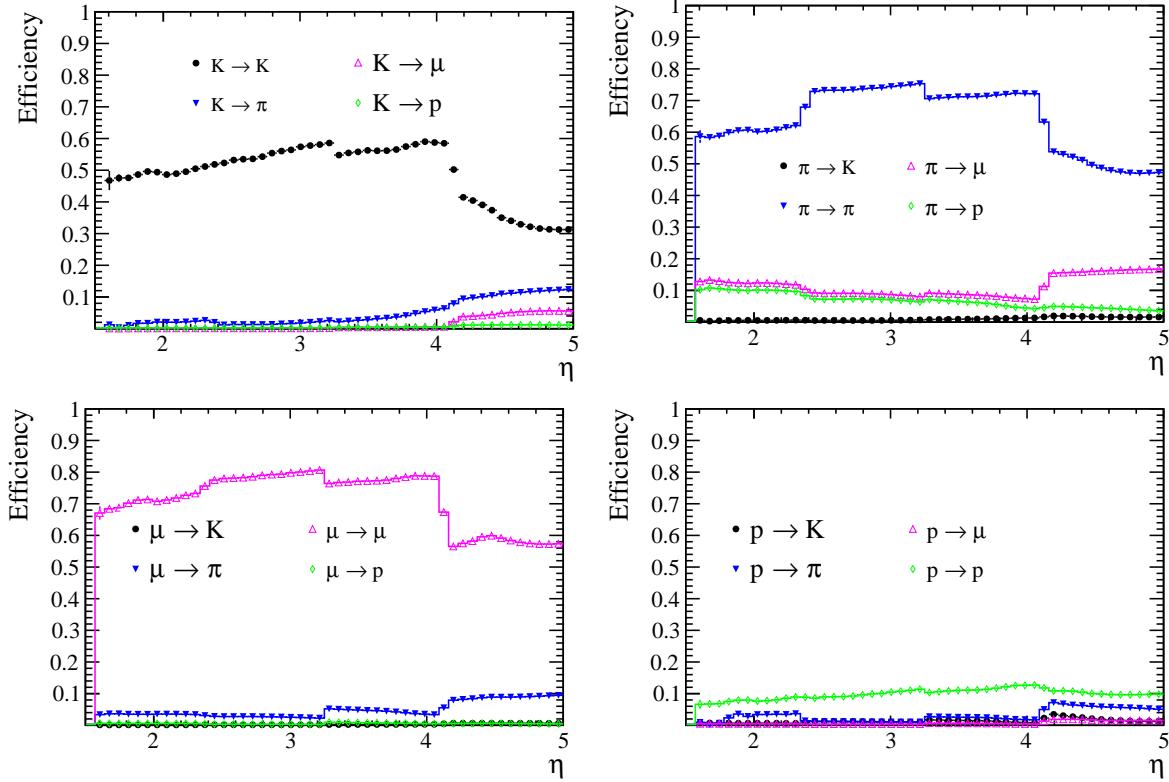


Figure 92: Variations of the elements $P(\hat{i}|j)$ ($j \rightarrow i$) as a function of the pseudorapidity, for the enriched regions defined for the fakeK sample.

1110 **D.2 Elements as a function of m_{corr}**

1111 Profiling (or resampling) the elements $P(\hat{i}|j)$ ($j \rightarrow i$) as a function of m_{corr} leads to the
 1112 curves shown in Figs. 96, 97

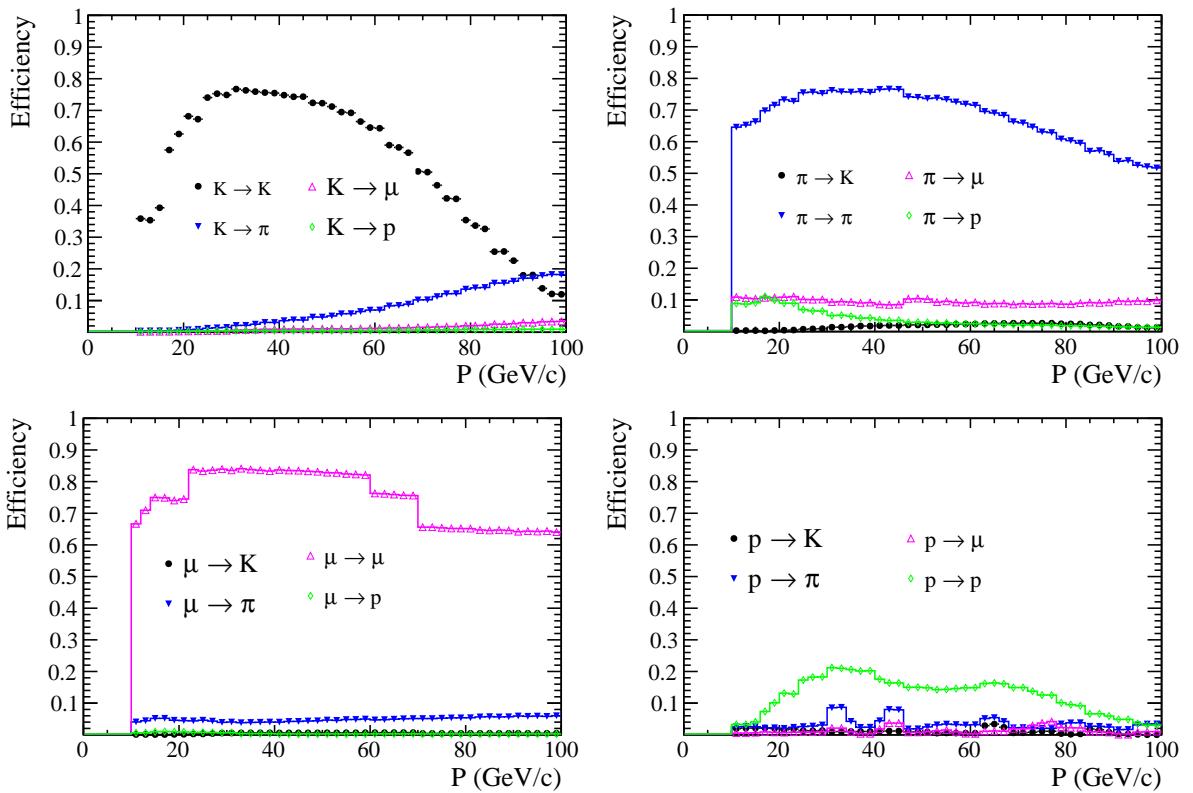


Figure 93: Variations of the elements $P(\hat{i}|j)$ ($j \rightarrow i$) as a function of the momentum, for the enriched regions defined for the fakeK sample.

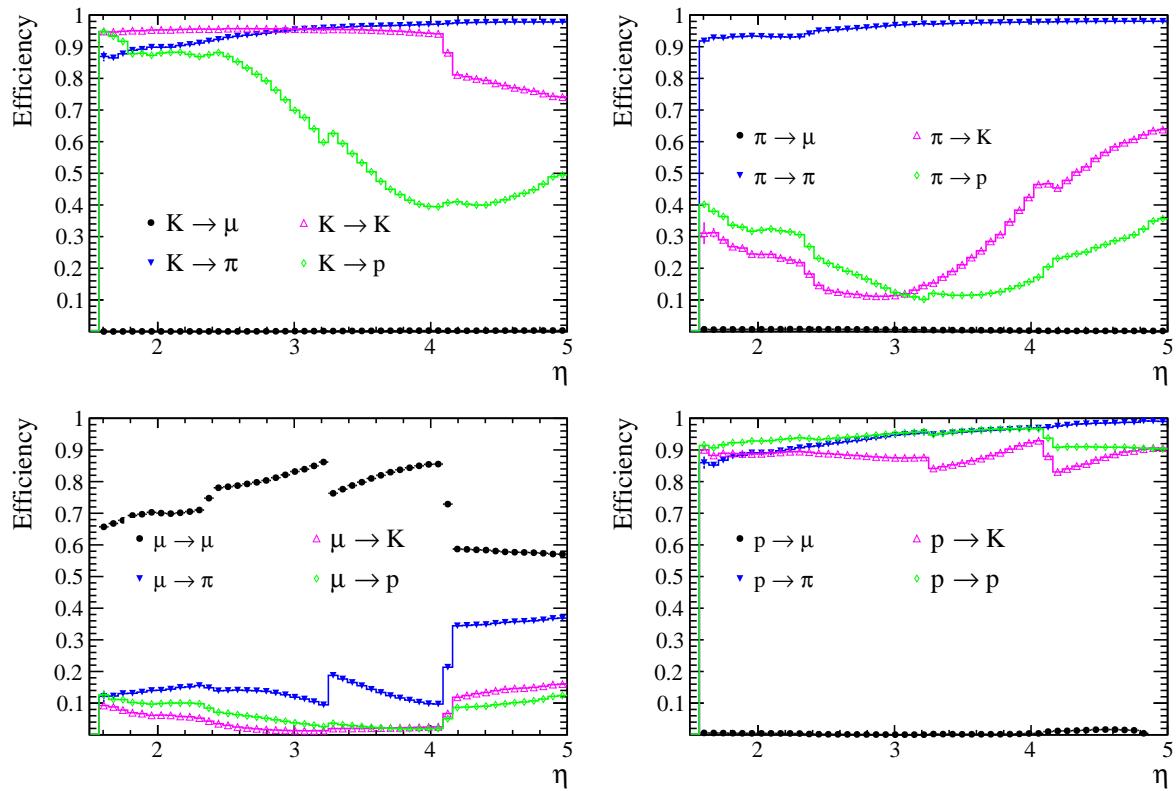


Figure 94: Variations of the elements $P(\hat{i}|j)$ ($j \rightarrow i$) as a function of the pseudorapidity, for the enriched regions defined for the fakeMu sample.

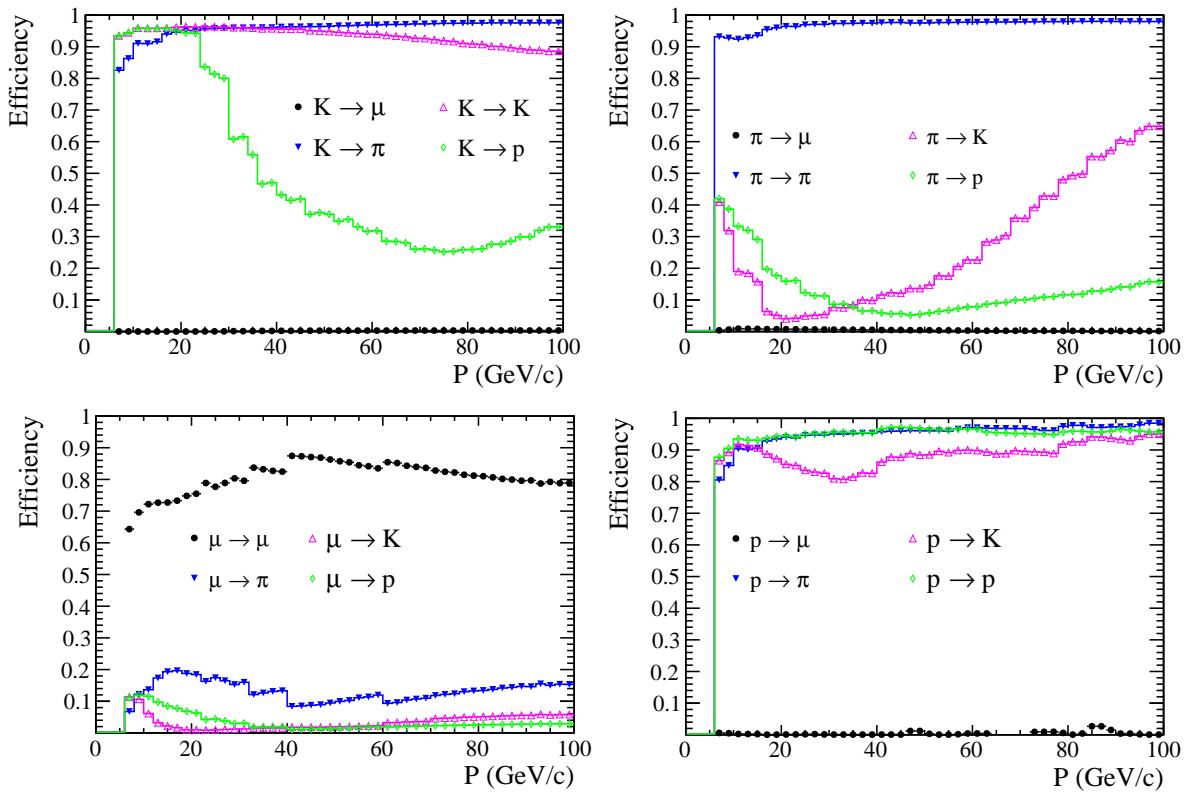


Figure 95: Variations of the elements $P(\hat{i}|j)$ ($j \rightarrow i$) as a function of the momentum, for the enriched regions defined for the fakeMu sample.

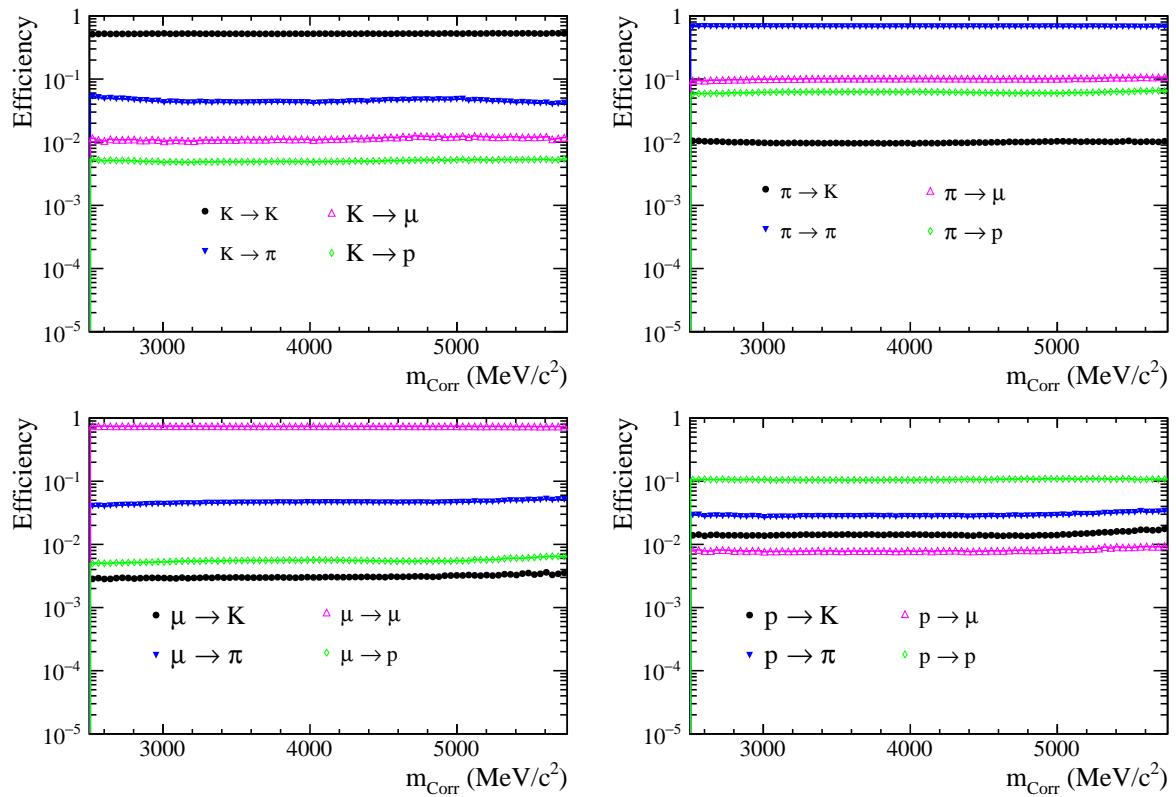


Figure 96: Variations of the elements $P(\hat{i}|j)$ ($j \rightarrow i$) as a function of m_{corr} , for the enriched regions defined for the fakeK sample.

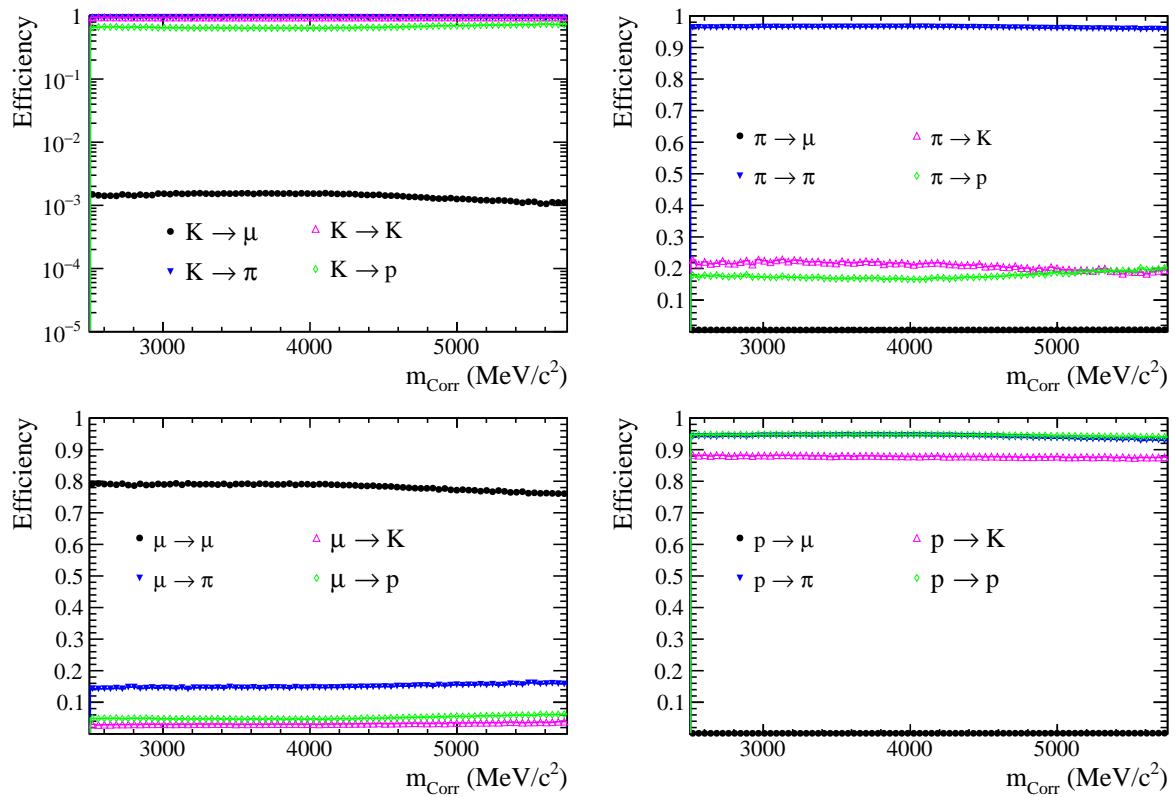


Figure 97: Variations of the elements $P(\hat{i}|j)$ ($j \rightarrow i$) as a function of m_{corr} , for the enriched regions defined for the fakeMu sample.

1113 **E Study of the mis-identification background for**
 1114 $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$

1115 The mis-identification background for $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$ is obtained using the Fake muon
 1116 lines (prescaled to 2%). The approach of subtracting none D_s^- background is not followed
 1117 because of lack of statistics per Bs MCORR bin. We follow the same method used for
 1118 the signal and explained here 9.2, Identical cuts to Table 22 were considered for the μ^+
 1119 misidentification sources(pions, protons and kaons). Figure 98 shows the distribution of
 1120 D_s^- mass for each enriched region

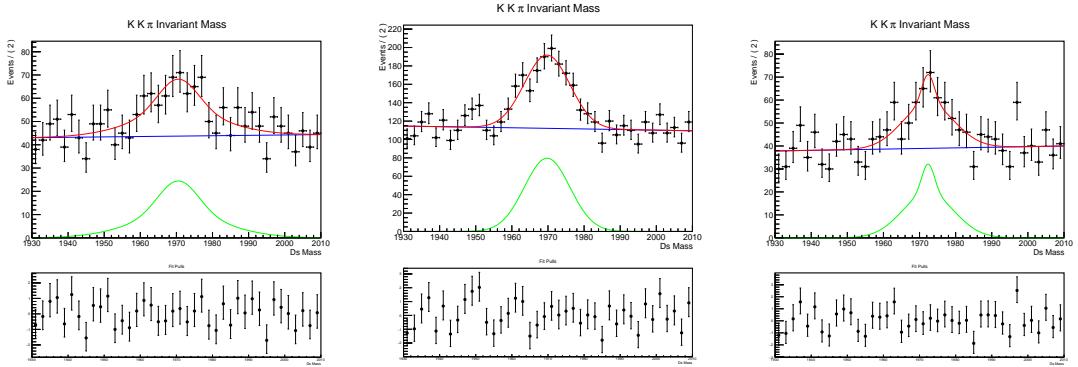


Figure 98: Distributions of D_s^- mass for: enriched kaon region(Left), enriched pion region(centre) and enriched proton region(Right), these are fitted with a sum of a two-Crystal-Balls function(to model the D_s^- peak) and an exponential to model the none- D_s^- background in these regions

1121 sPlot technique is applied to subtract the combinatorial background, the sWeights are
 1122 then used to extract the Bs MCORR distribution for each region. With mis-ID rates and
 1123 efficiencies(from PID calib) as input sweigted Bs MCORR distributions are combined
 1124 to form the shape of the total mis-PID background in the signal region, we estimate
 1125 the contribution of mis-identified background to be ~ 570 events, Figure 99 shows the
 1126 distribution of Bs MCORR for this source of background.

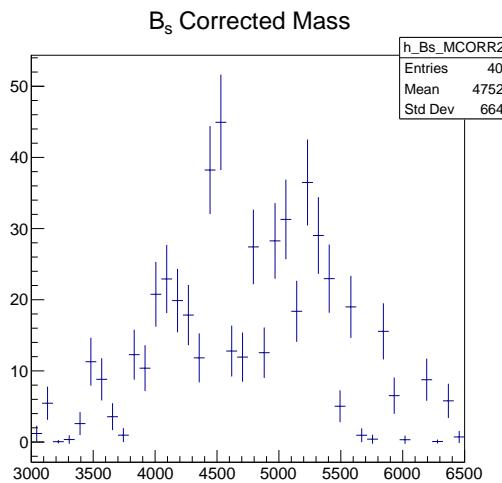


Figure 99: Bs corrected mass shapes for misidentification background for $B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$.

1127 F Study of the same sign sample

1128 Wrong-sign combination of the B_s^0 daughters is traditionally useful to infer the shape
 1129 and yield of the combinatorial background, this approach was used in various analyses at
 1130 LHCb [41], [50] and [51] among many others. The main point there is that the combination
 1131 of wrong-sign daughters should be made of *pure* combinatoric, no physics background
 1132 can distort this assumption. This strategy does not apply to our case; $K^+\mu^+$ sample is
 1133 riddled with Cabibbo-favoured semi-leptonic decays like $B^+ \rightarrow D^0\mu^+\nu_\mu$, $B^0 \rightarrow D^*\mu^+\nu_\mu$ and
 1134 $\Lambda_b^0 \rightarrow \Lambda_c^+\mu^-\bar{\nu}_\mu$, those are processes with high branching fraction and they are also *unique*
 1135 to the wrong-sign sample (they do not usually show up in the signal sample). On the other hand, backgrounds such as $B_s^0 \rightarrow K^-\mu^+\nu_\mu$ and Beauty to charmonium decays
 1136 do show up in both samples, we actually checked that Beauty to charmonium decays
 1137 have the same contribution to both wrong-sign and signal samples, this was achieved by
 1138 calculating the invariant mass between the μ^+ and the non-isolating track chosen by the
 1139 isolation algorithm, Figure 100 shows the distribution of such mass for the wrong sign
 1140 and signal sample, as it is clear the peak around J/ψ belongs dominantly to Beauty to
 1141 charmonium physics (combinatorial J/ψ are present but suppressed due to the tight cut
 1142 on the flight distance χ^2 at the stripping level). Using a fit to that peak we measure the
 1143 yield of the J/ψ peak in both samples, by taking the ratio between the two yields we
 1144 obtain: 0.098 ± 0.001 (compatible with 2σ with stripping prescale factor: 0.1).

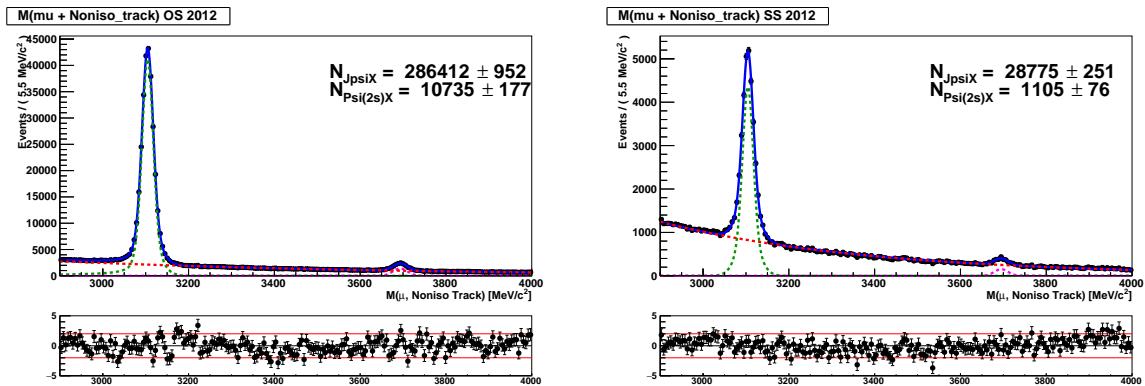


Figure 100: Fit to the invariant mass of the muon + non-isolating track which is chosen by the charged isolation algorithm for signal sample (Right) and for wrong-sign sample (Left)

1146 Another complication is that the misid background on both K^+ and μ^- particles are
 1147 different between signal and wrong-sign samples. Indeed, decays like: $B^0 \rightarrow \pi^-\mu^+\nu_\mu$ and
 1148 $\Lambda_b^0 \rightarrow p\mu^-\bar{\nu}_\mu$ represent an important contribution to the K^+ misidentification background
 1149 in the signal sample but is absent in the wrong-sign sample, the same can be said about
 1150 $B \rightarrow h^+h'^-$ decays contributing to the μ^- misID background component for the signal.
 1151 The complications described here induced the following strategy: subtract physics and
 1152 MisID background sources should be done before using the wrong-sign sample as proxy for
 1153 combinatorial background. Due to the presence of ill-known processes like $B^+ \rightarrow D^0\mu^+\nu_\mu$
 1154 where $D^0 \rightarrow K^+\pi^-\pi^0$ (π^0) and the lack of MC samples made the procedure highly unstable.
 1155 The efficiency is dependent on an accurate description of the final state of such D^0 decays,
 1156 this is lacking in the literature. We decided, finally, to opt for event mixing as a way to
 1157 extract the combinatorial background.

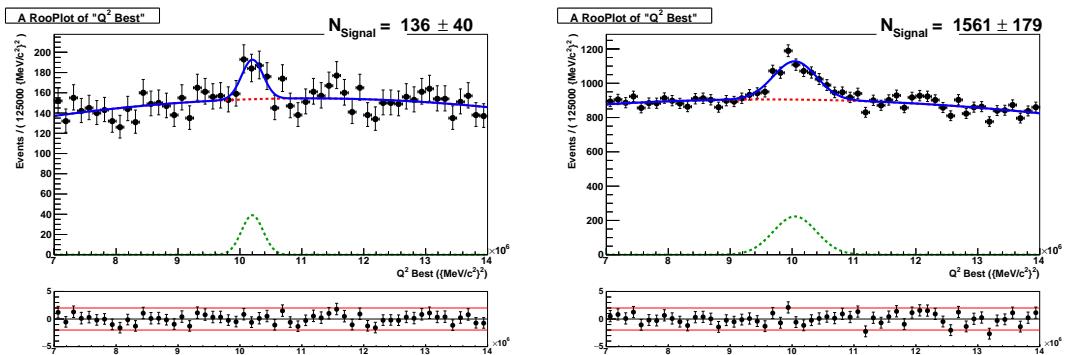


Figure 101: Fit to the Q^2 spectrum of for signal sample (Right) and for wrong-sign sample(Left). The background pdf is described using Chebyshev function of the second order, while the peak is modeled using two gaussians

1158 G TisTos efficiency checks

1159 We estimate the systematics on the trigger correction factor using an alternative variable
 1160 to bin in which is IPChi2 of the muon instead of its IP, Figures ?? shows the correction
 1161 factors form control sample in data/MC along with the Pt, IPCHI2 maps for normalization
 1162 and signal channels in simulation, Table 52 shows the trigger eff. in q^2 bins for signal and
 1163 normalization channels.

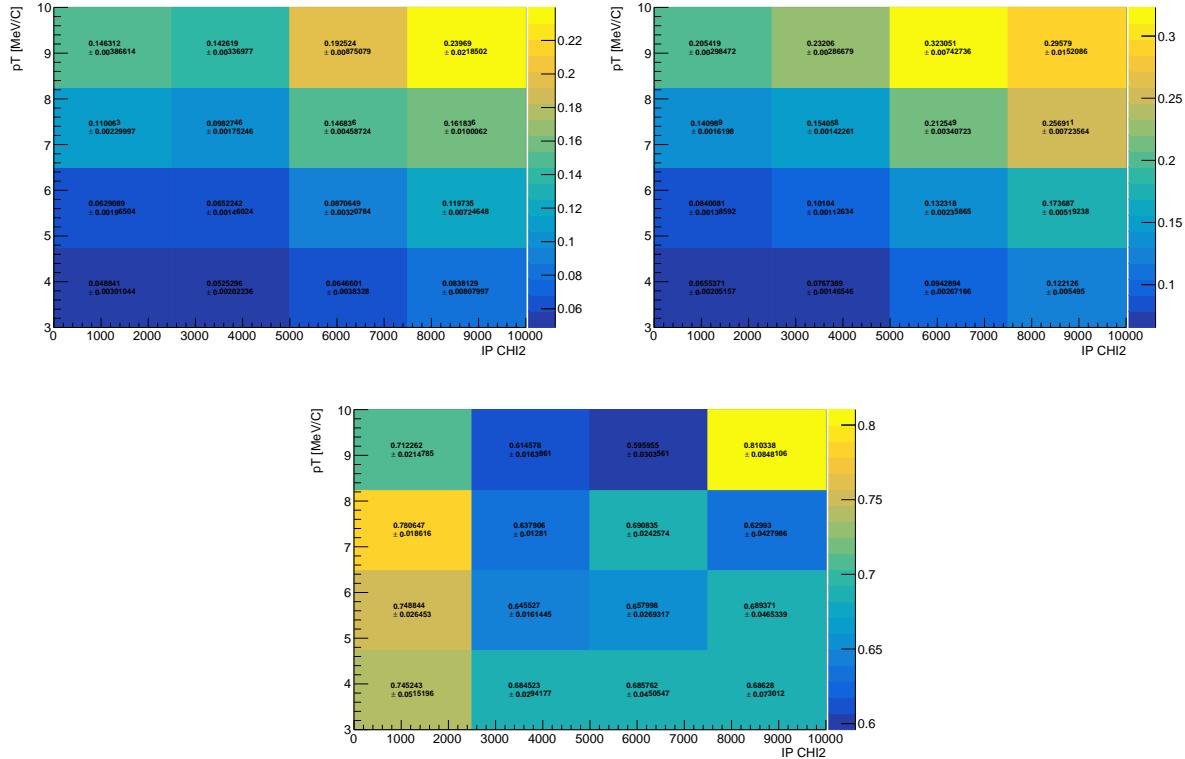


Figure 102: Top: TISTOS efficiency for $B^+ \rightarrow J/\psi K^+$ decay in bins of μ p_T and IPCHI2 for the trigger requirement in data (left) and simulation (right). Bottom: the ratio $TISTOS_{data}/TISTOS_{MC}$ for $B^+ \rightarrow J/\psi K^+$ decays.

	$B_s^0 \rightarrow K^- \mu^+ \nu_\mu$	$B_s^0 \rightarrow D_s^- \mu^+ \nu_\mu$	Ratio
No Sel.	0.688 ± 0.007	0.684 ± 0.006	1.005 ± 0.003
$q^2_{K^- \mu^+} < 7 \text{ GeV}^2/c^4$	0.689 ± 0.008	0.684 ± 0.006	1.007 ± 0.005
$q^2_{K^- \mu^+} > 7 \text{ GeV}^2/c^4$	0.684 ± 0.006	0.684 ± 0.006	1.000 ± 0.003

Table 52: Trigger correction factors averaged over all events applied to Monte Carlo in bins of q^2 .

1164 The check is performed using the $B^+ \rightarrow J/\psi K^+$ decay. The lines L0Muon,
 1165 HLT2SingleMuon and TopoMu2BodyBBDT are considered separately.

1166 The variation of the TISTOS efficiency is further studied in bins of B_s^0 η and p_T
 1167 distributions in Figs. 105,106. Other plots show the efficiency in bins of daughters
 1168 kinematics(p, p_T).

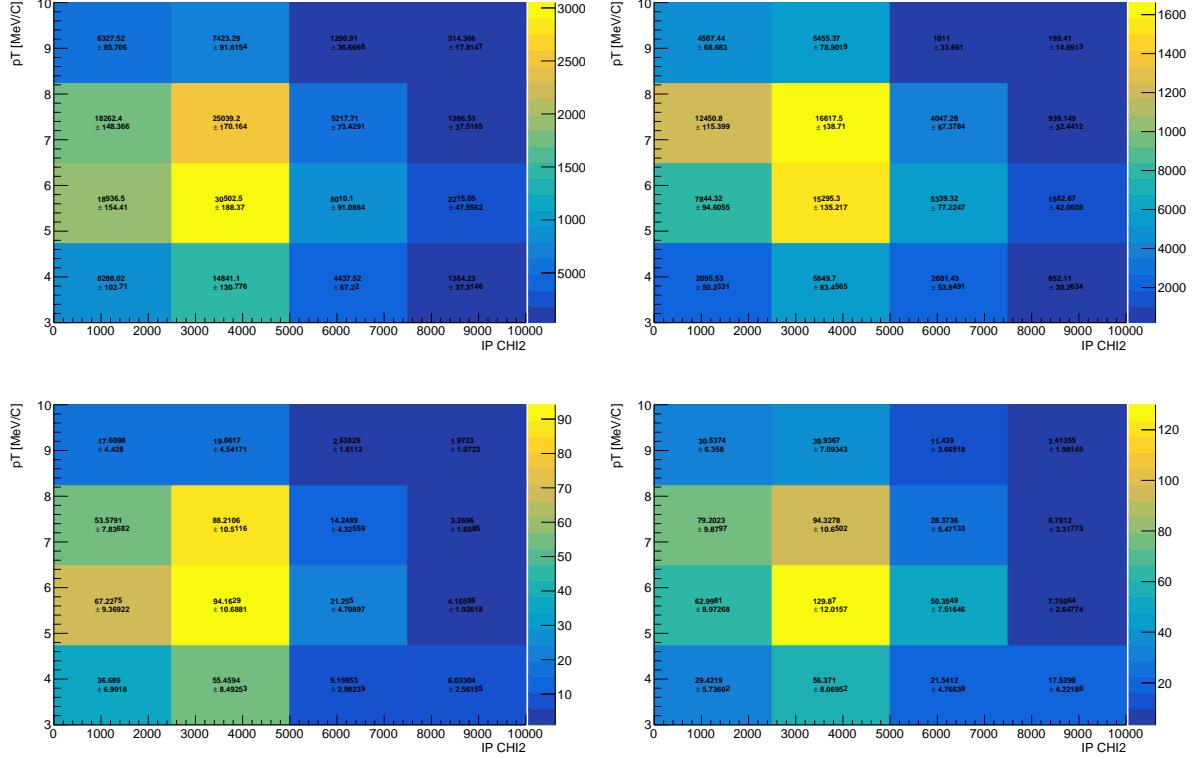


Figure 103: Top: Signal(Left) and Normalization (Right) distribution of simulated events in bins of p_T and IP in the full q^2 region. Bottom: Signal distribution of simulated events in bins of p_T and IP for the low q^2 region(Left) and High q^2 region(Right).

1169 Figures 111 to 115 show the distributions for the TOS fractions for each trigger line in
 1170 bins of $B \eta, p_T$ for $B^+ \rightarrow J/\psi K^+$ control channel

1171 This study insures that no bias is introduced by the TIS decisions and guarantees a
 1172 superior statistical uncertainties.

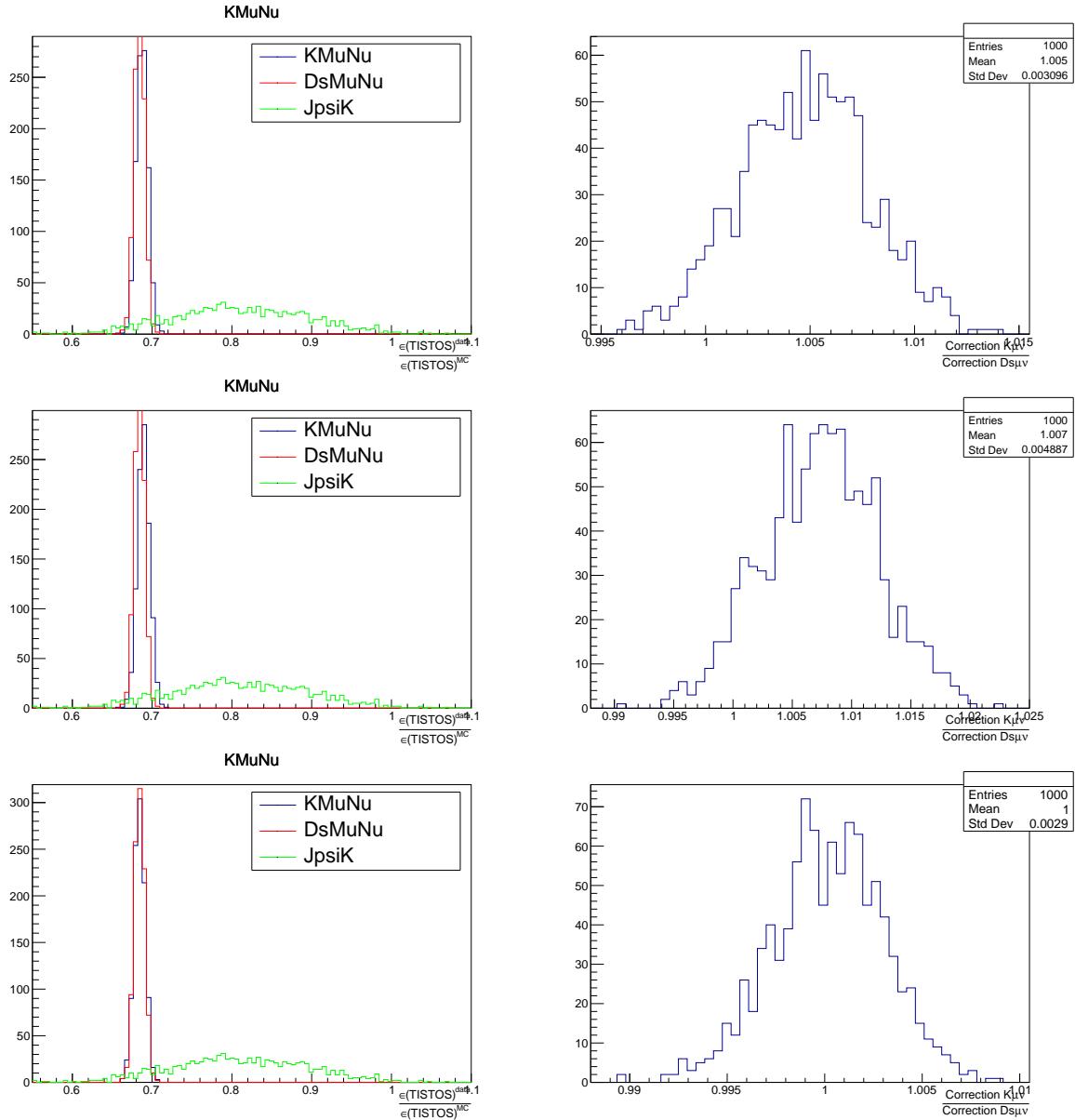


Figure 104: Left: Distribution of correction factors(data/MC) for Signal (red), normalization (Green) and control decay(black). Right: distribution of the ratio of correction factors (Signal/Normalization) for 1000 toy experiments. Top plots represent the full q^2 region, center plots are for the low q^2 region while the bottom plots are for the high q^2 region.

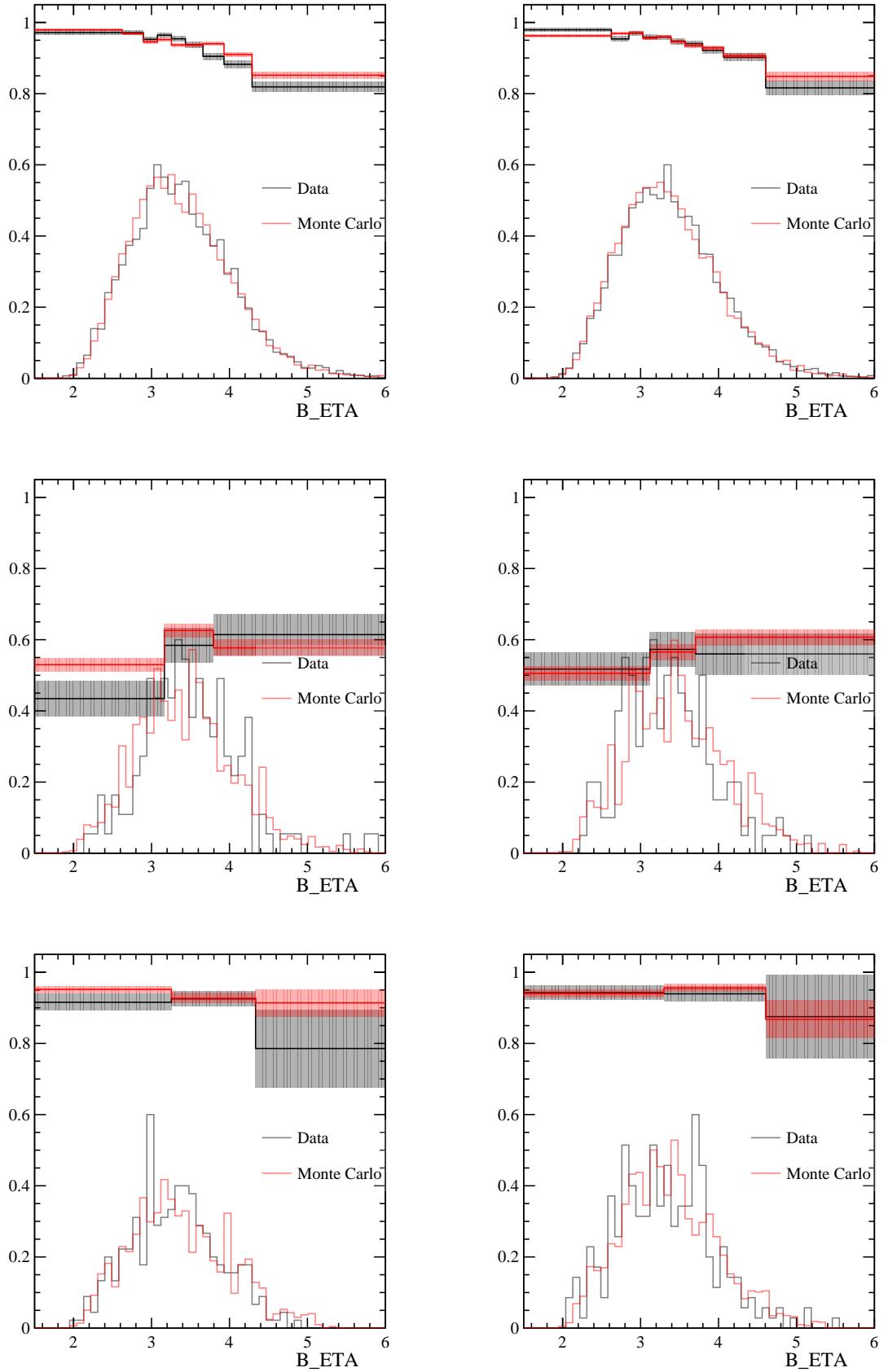


Figure 105: TISTOS efficiency variations in bins of $B_s^0 \eta$ for L0Muon trigger(Top), Hlt2SingleMuon(Centre), TopoMu2BodyBBDT(Bottom), those are split in magnet polarity:
Left: Up, Right: Down polarity

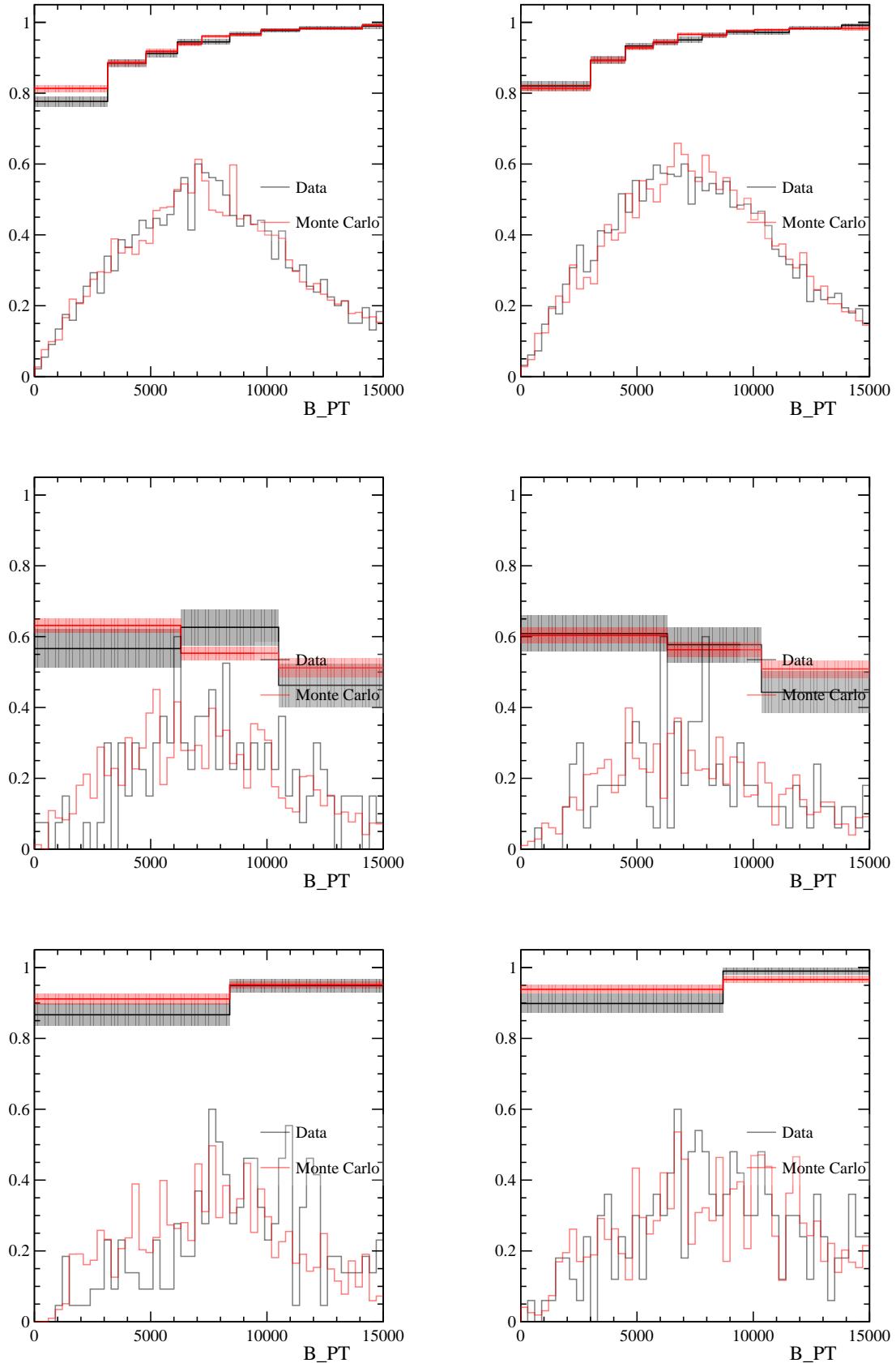


Figure 106: TISTOS efficiency variations in bins of B_s^0 pt for L0Muon trigger(Top), Hlt2SingleMuon(Centre), TopoMu2BodyBBDT(Bottom), those are split in magnet polarity:
Left: Up, Right: Down polarity

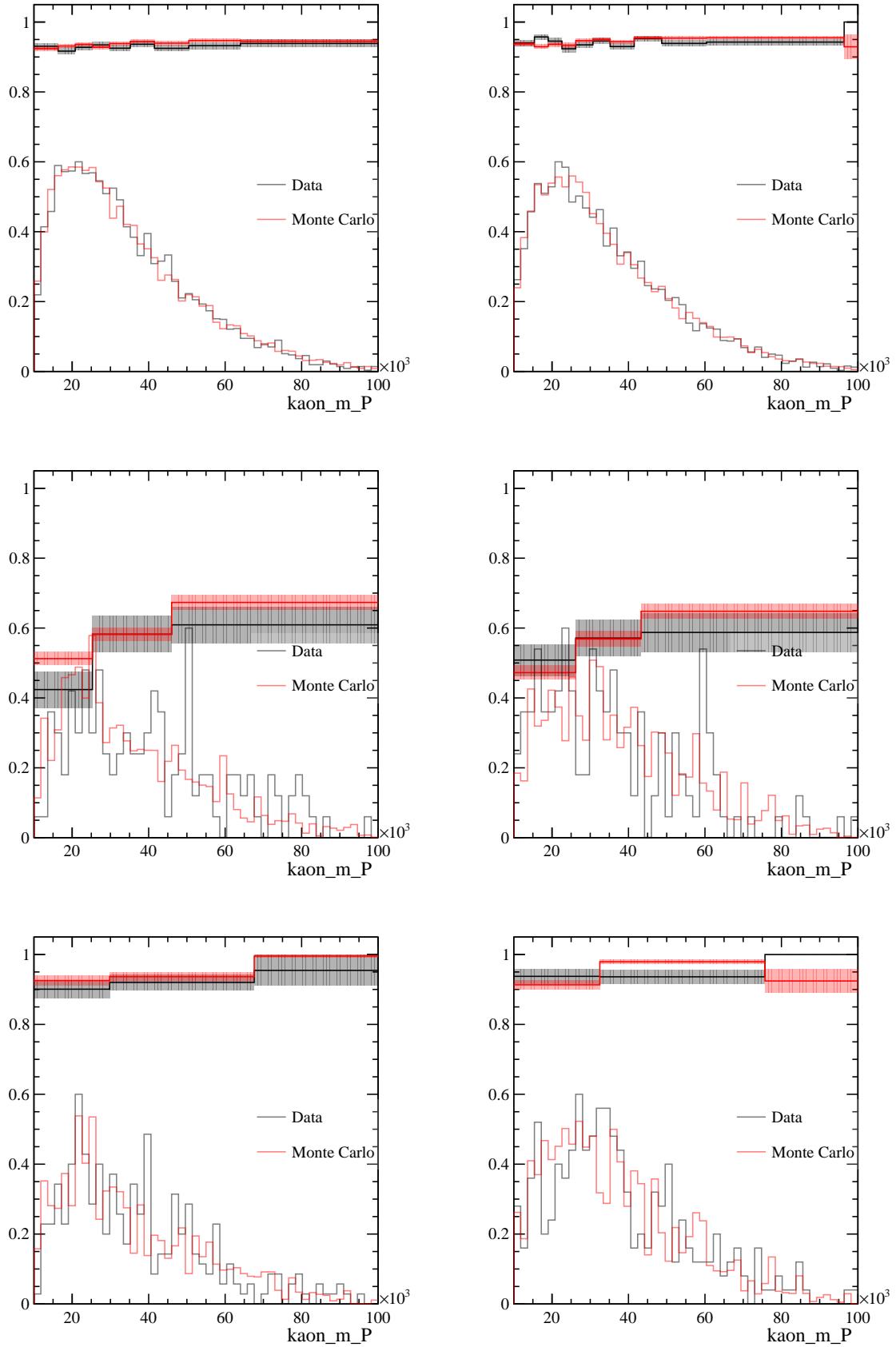


Figure 107: TISTOS efficiency variations in bins of K^- P for L0Muon trigger(Top), Hlt2SingleMuon(Centre), TopoMu2BodyBBDT(Bottom), those are split in magnet polarity:
Left: Up, Right: Down polarity

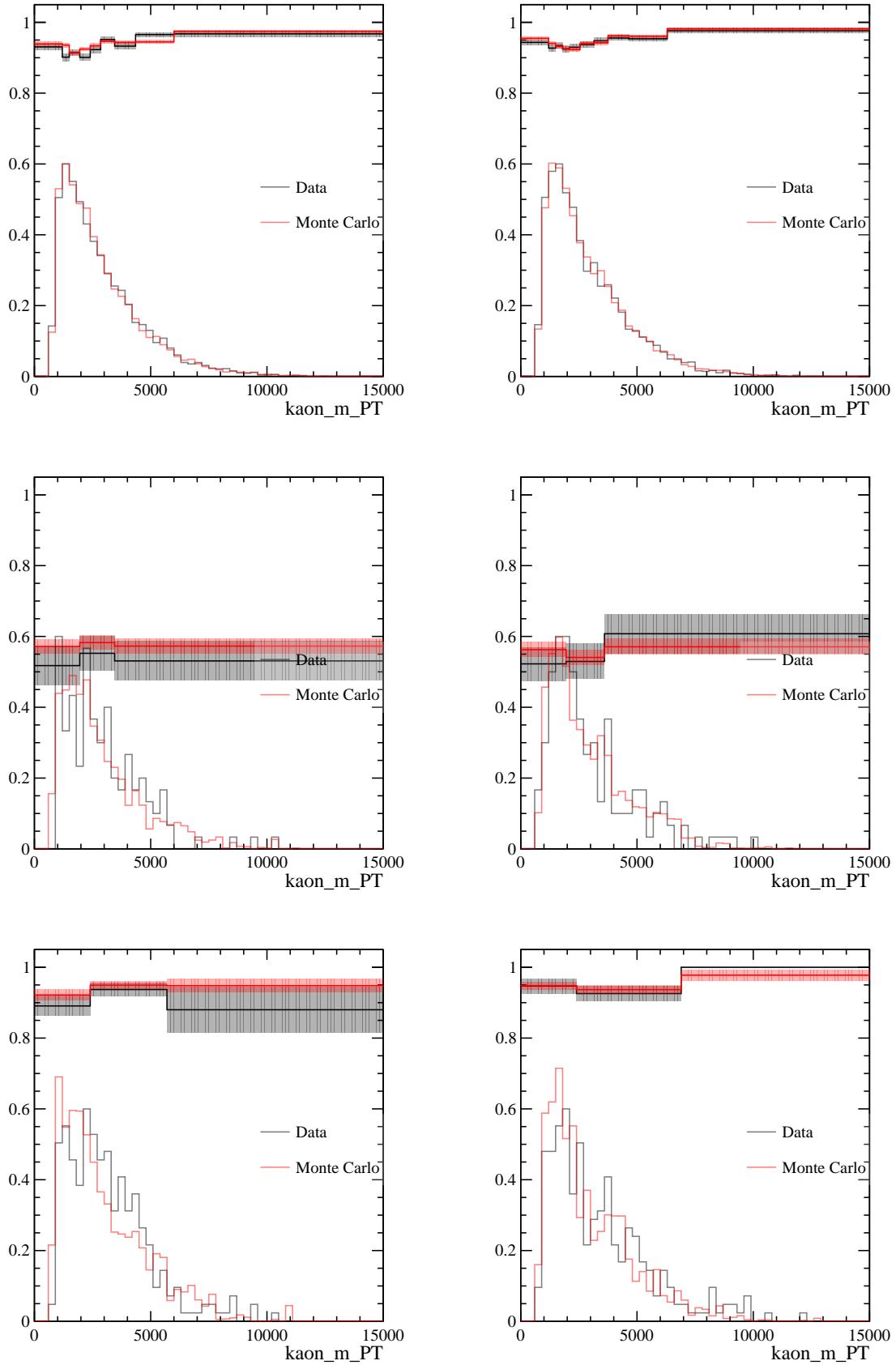


Figure 108: TISTOS efficiency variations in bins of K^- p_T for L0Muon trigger(Top), Hlt2SingleMuon(Centre), TopoMu2BodyBBDT(Bottom), those are split in magnet polarity: Left: Up, Right: Down polarity

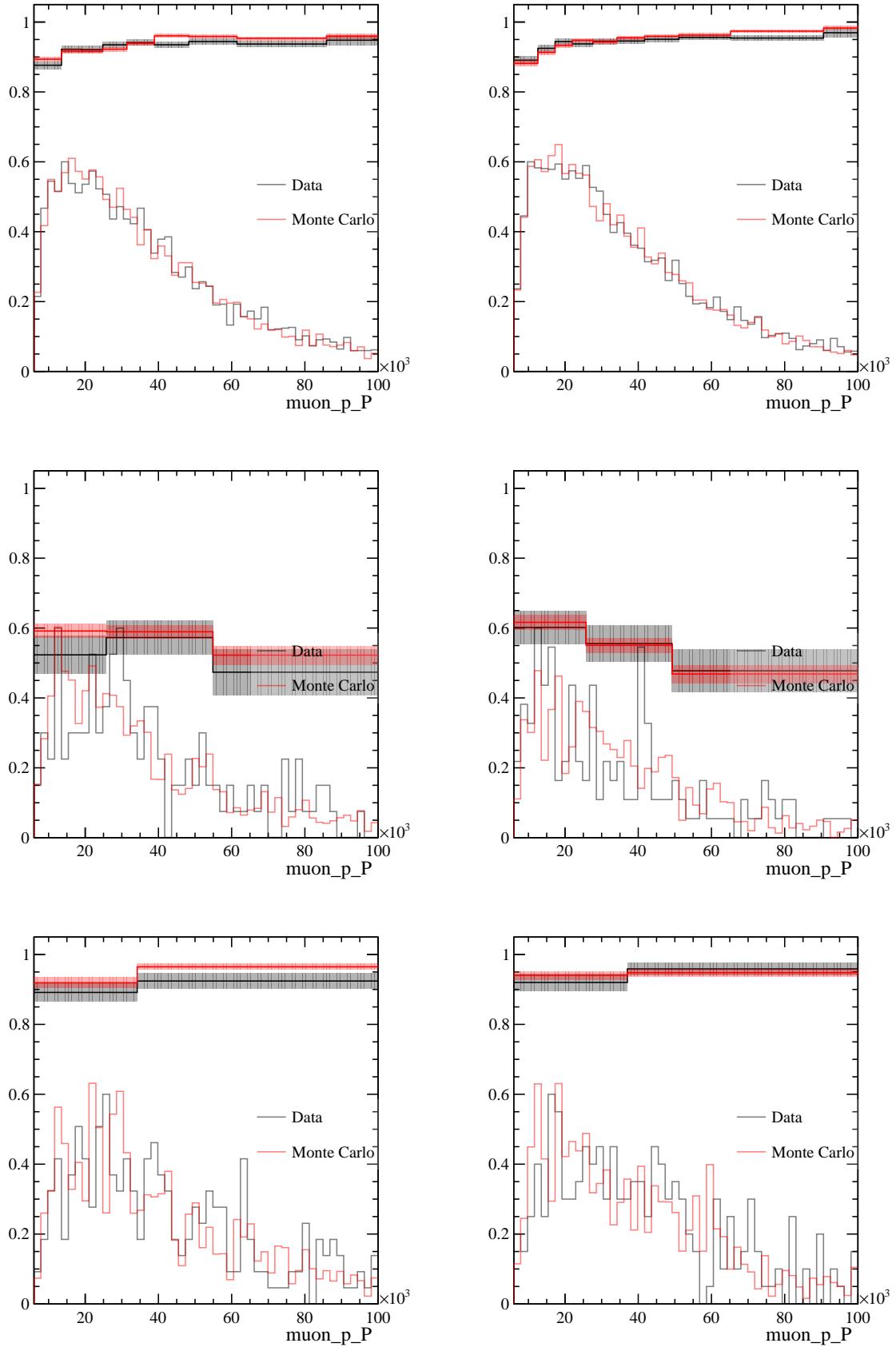


Figure 109: TISTOS efficiency variations in bins of μ^+ P for L0Muon trigger(Top), Hlt2SingleMuon(Centre), TopoMu2BodyBBDT(Bottom), those are split in magnet polarity:
Left: Up, Right: Down polarity

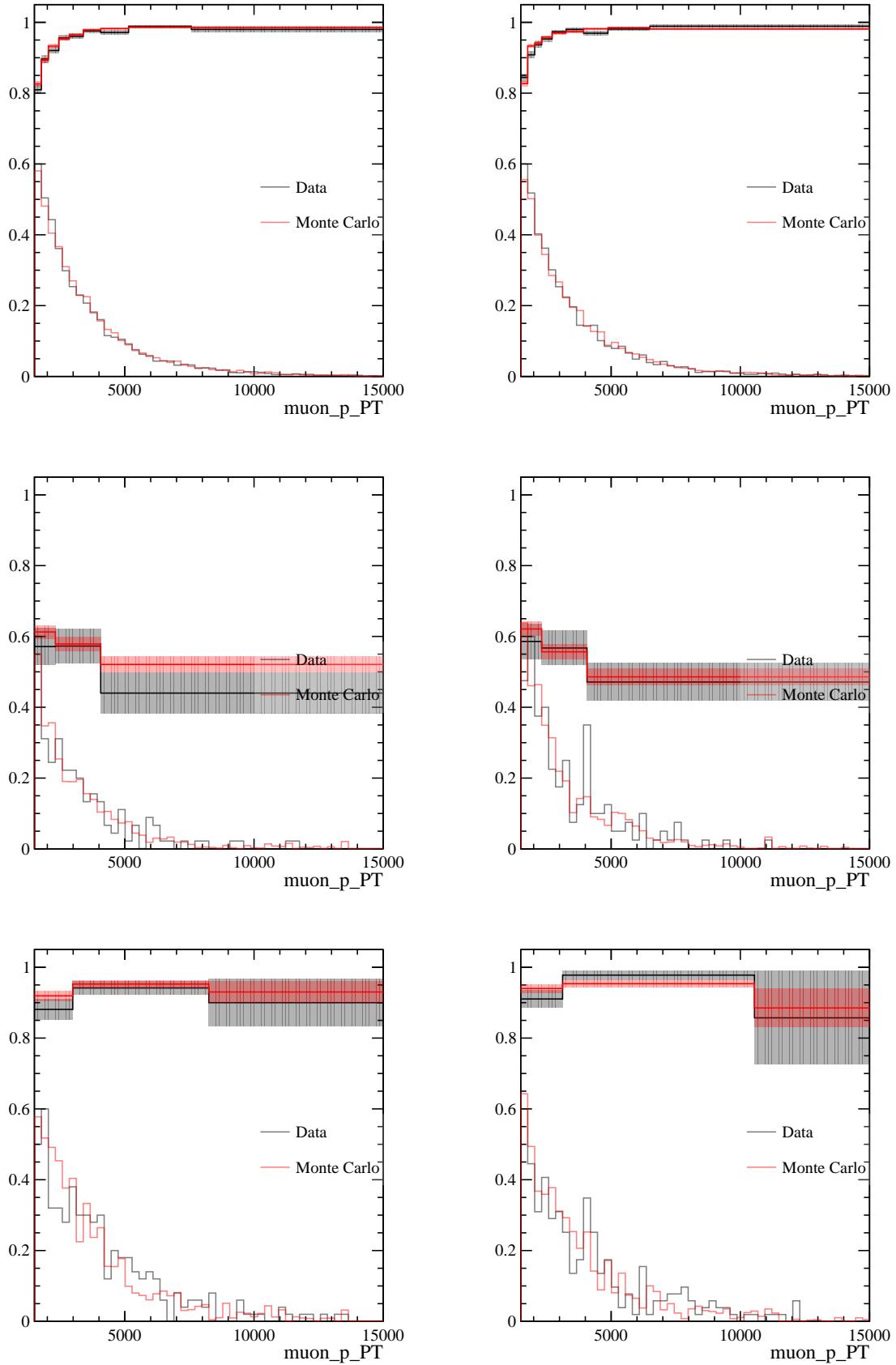


Figure 110: TISTOS efficiency variations in bins of μ^+ p_T for L0Muon trigger(Top), Hlt2SingleMuon(Centre), TopoMu2BodyBBDT(Bottom), those are split in magnet polarity:
Left: Up, Right: Down polarity

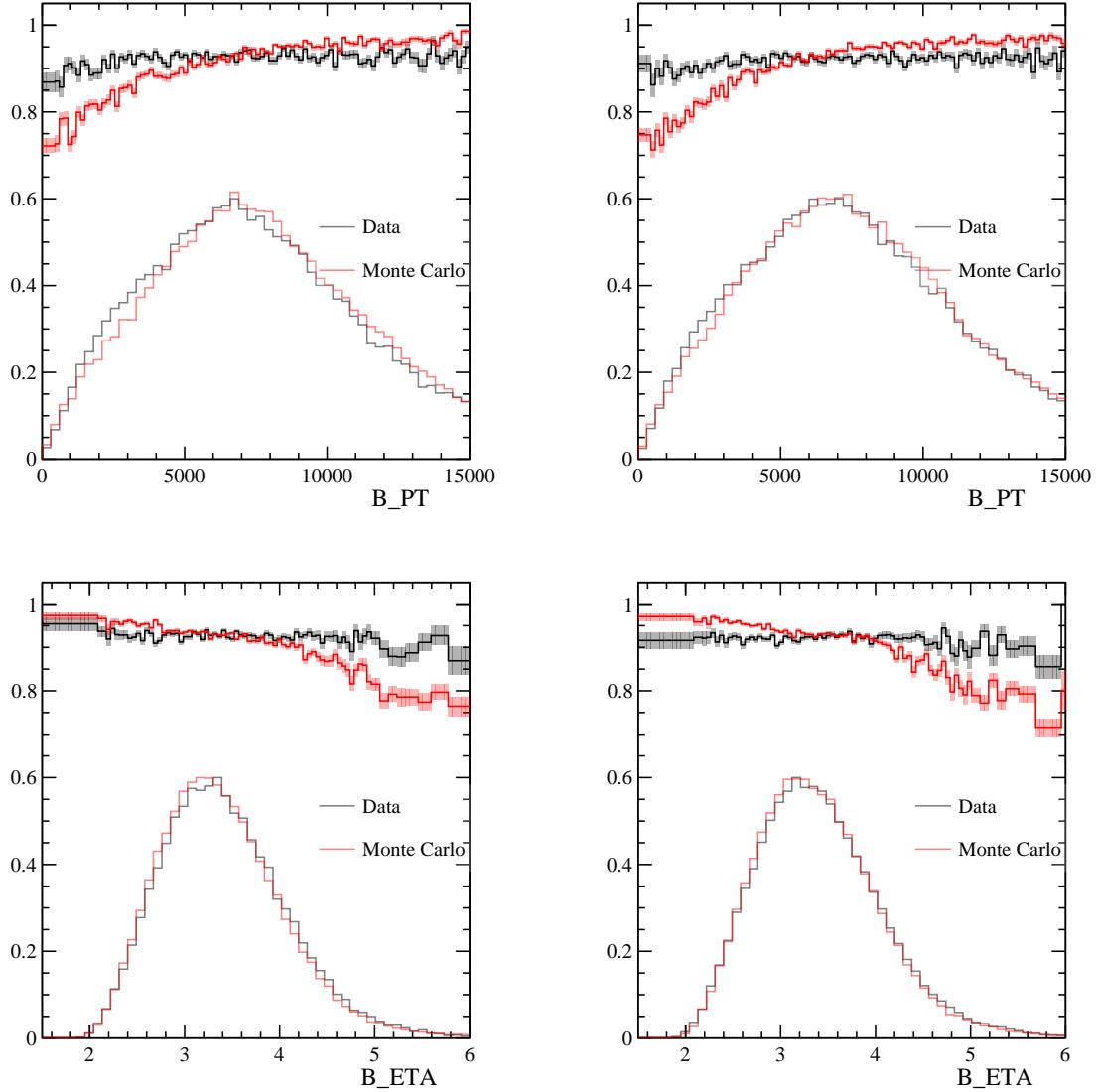


Figure 111: TOS fraction variations in bins of $B \eta, p_T$ for L0Muon trigger split in (Up & Down) polarity

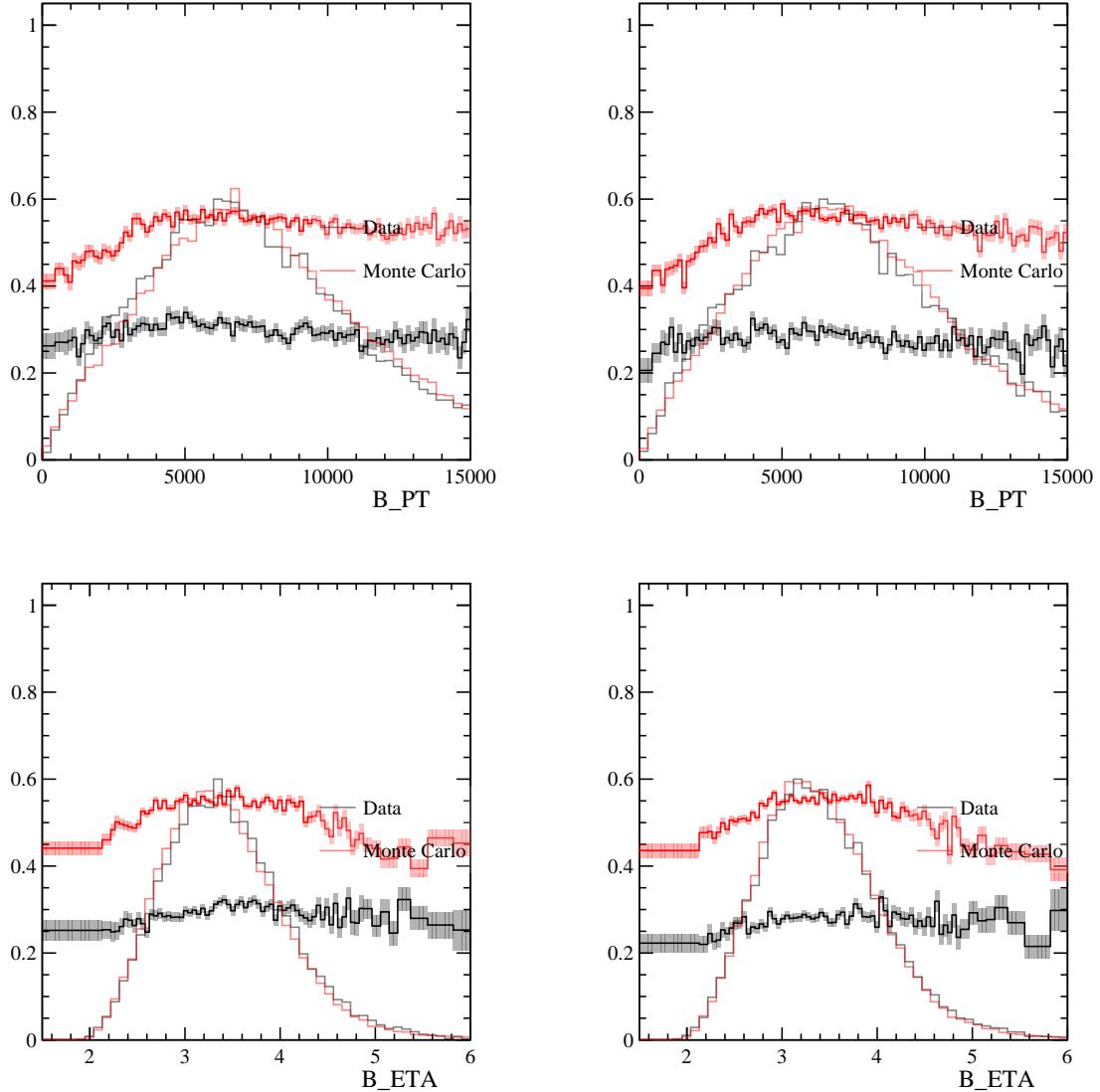


Figure 112: TOS fraction variations in bins of B_{η}, p_T for HLT2SingleMuon trigger split in (Up & Down) polarity

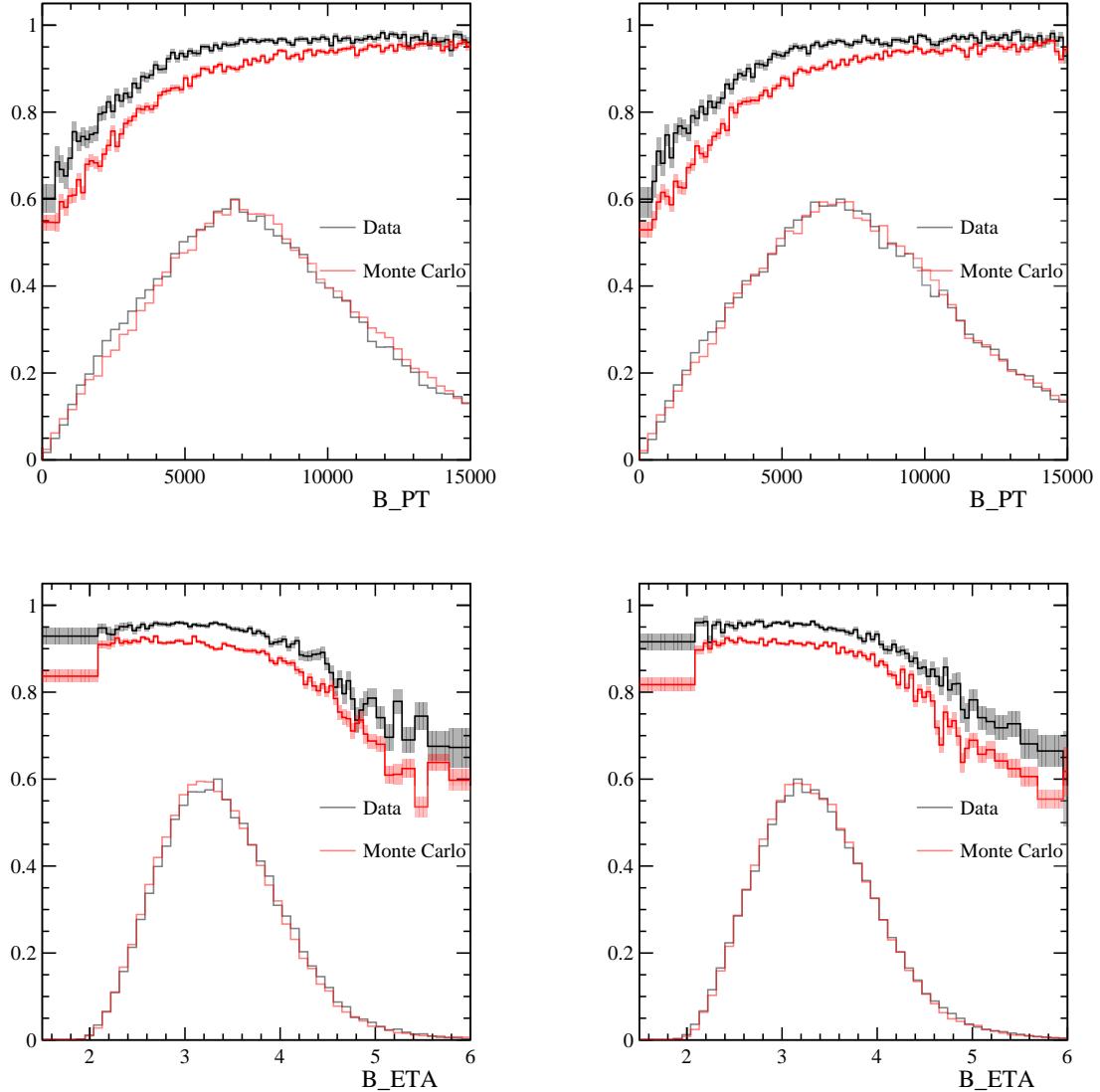


Figure 113: TOS fraction variations in bins of $B \eta, p_T$ for Mu2BodyTopo trigger split in (Up & Down) polarity

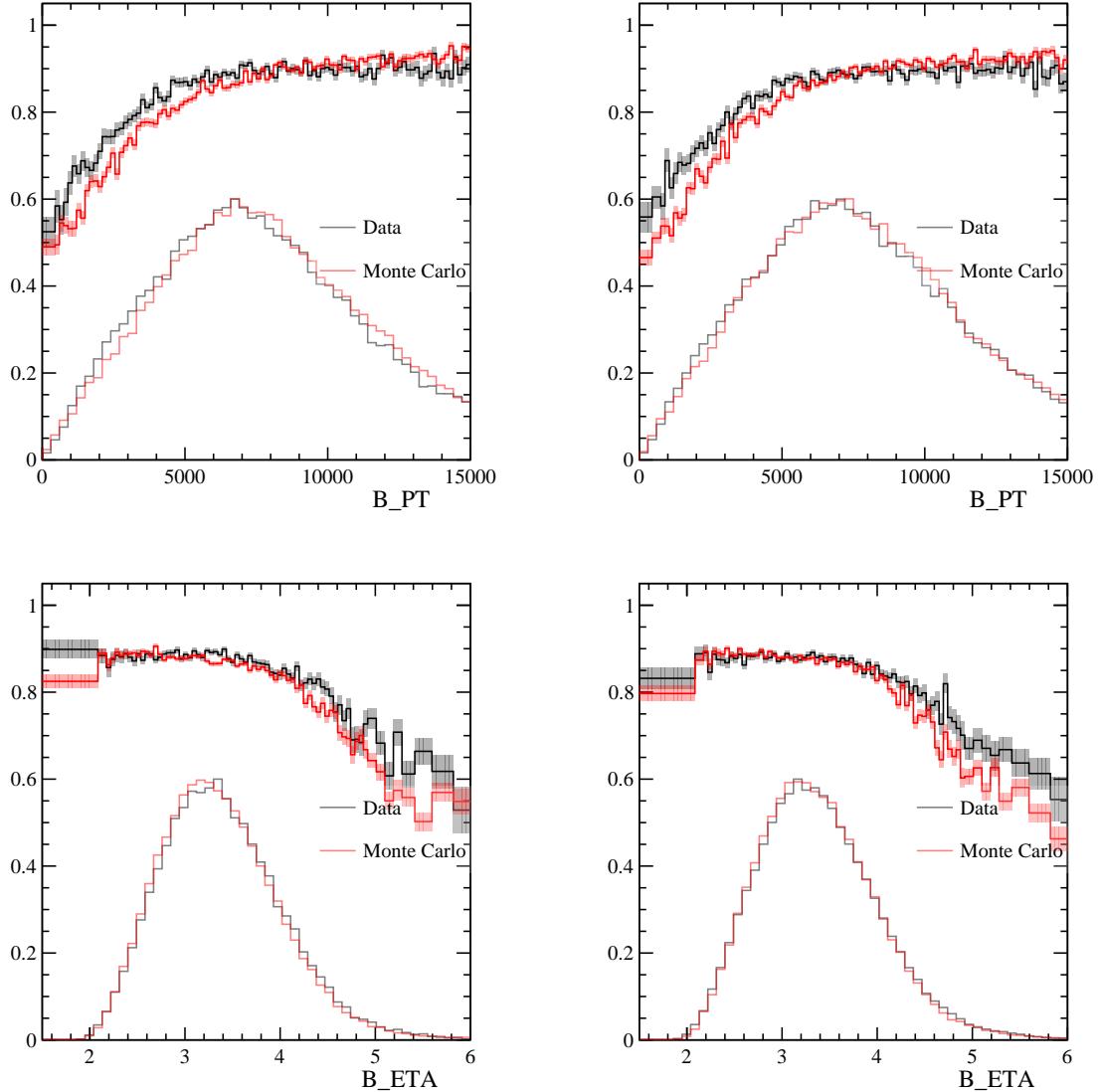


Figure 114: TOS fraction variations in bins of $B \eta, p_T$ for L0Muon & Mu2BodyTopo trigger split in (Up & Down) polarity

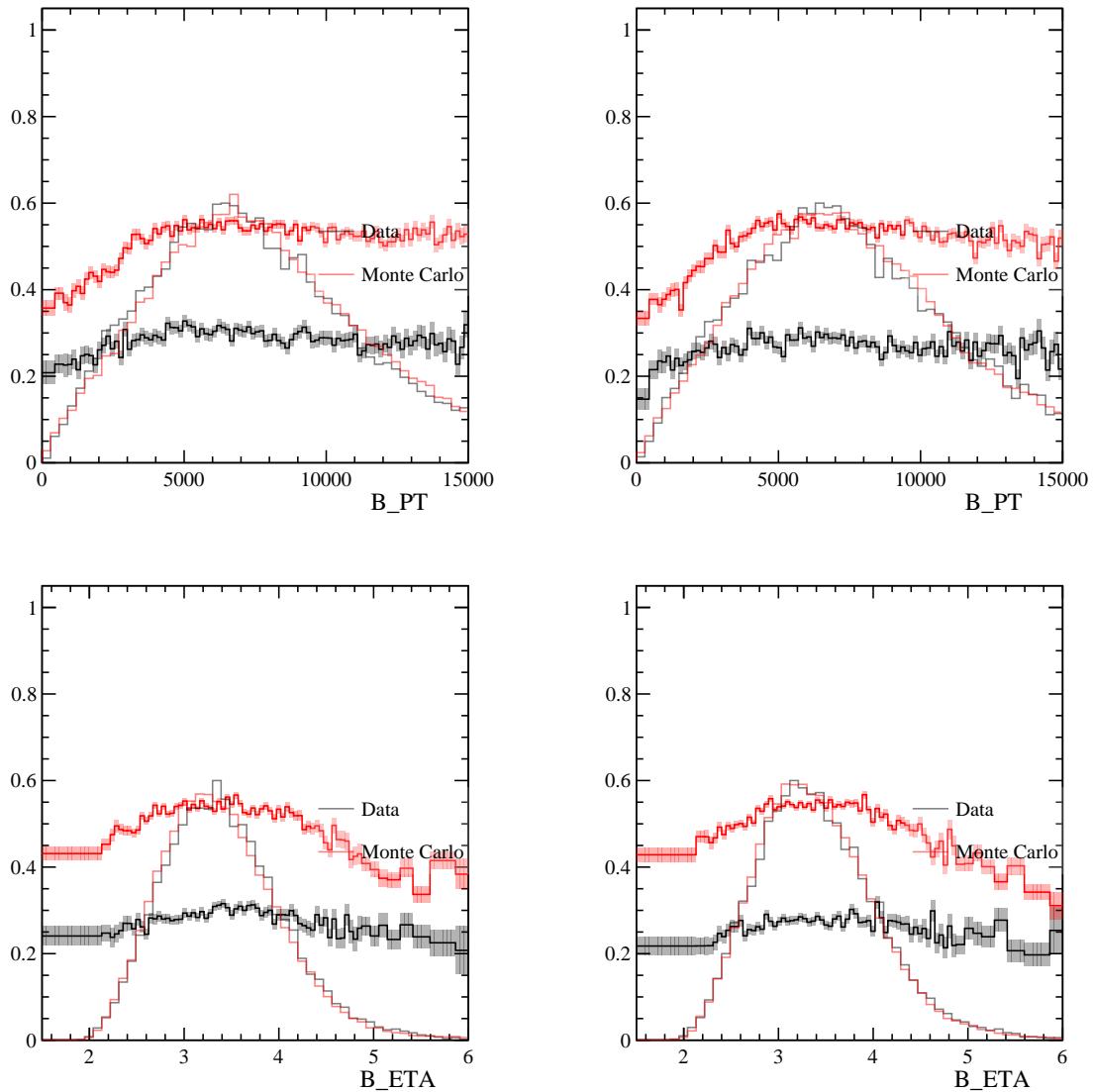


Figure 115: TOS fraction variations in bins of $B \eta, p_T$ for SingleMuon & Mu2BodyTopo trigger split in (Up & Down) polarity

1173 H Projections of the Signal Fit

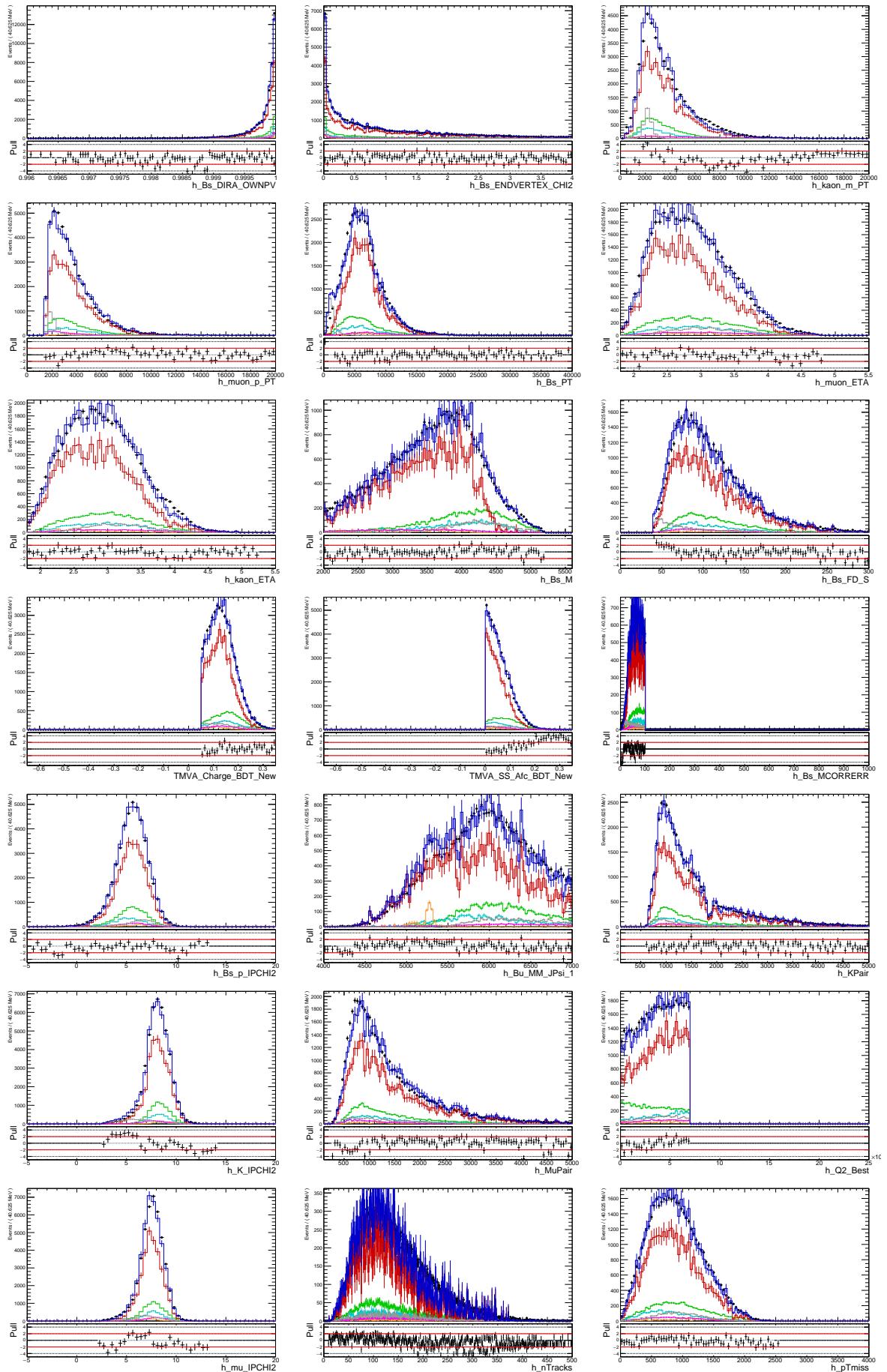


Figure 116: Fit projection onto various variables for the low q^2 bin, the total pdf is shown in blue along with its individual components: Signal (Green), B2CC (Orange), K^* (Cyan), MisPID (Magenta), Combinatorial (Grey) and Vcb (red) backgrounds.

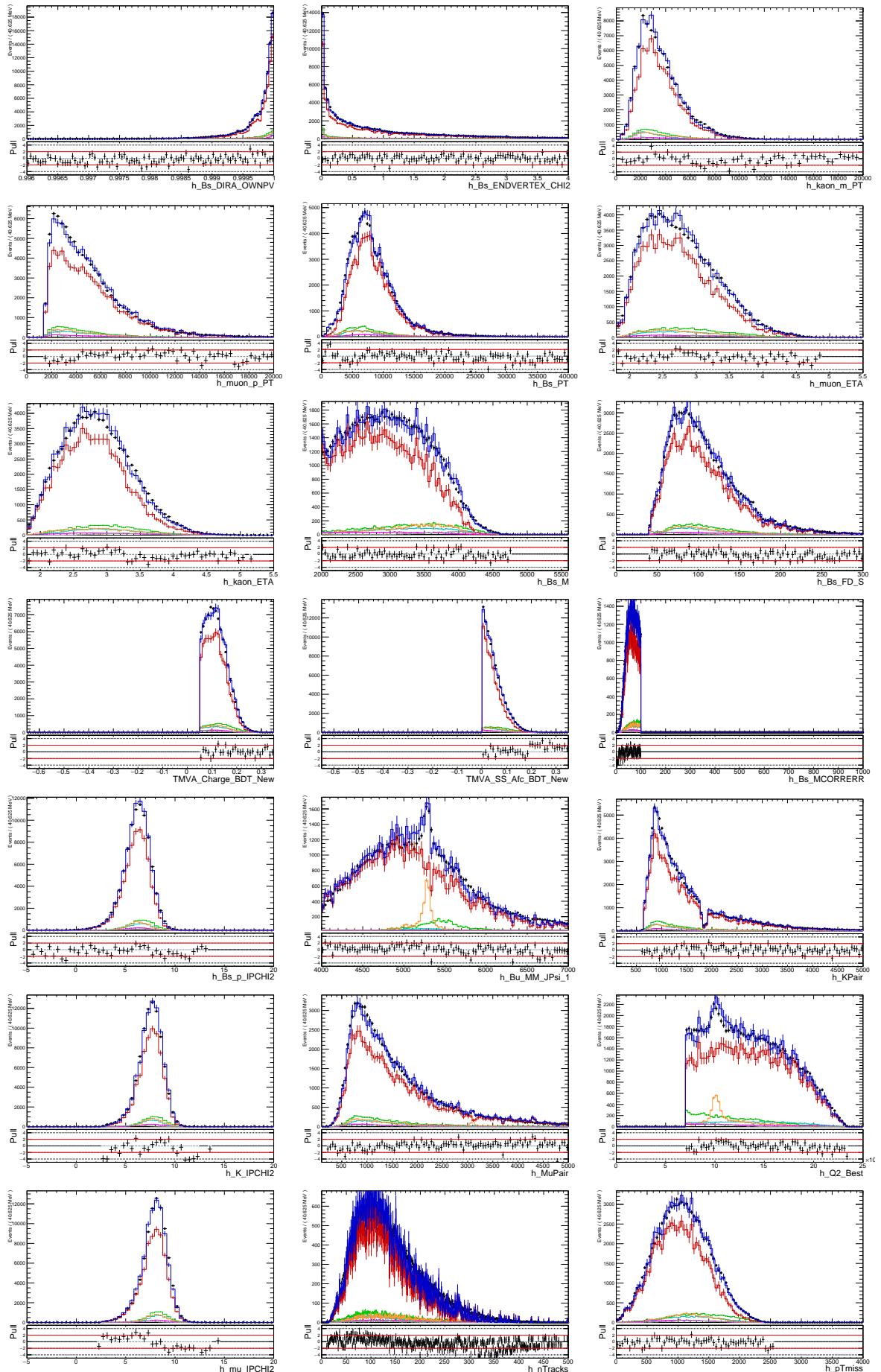


Figure 117: Fit projection onto various variables for the high q^2 bin, the total pdf is shown in blue along its individual components: Signal (Green), B2CC (Orange), K^* (Cyan), MisPID (Magenta), Combinatorial (Grey) and Vcb (red) backgrounds.

¹¹⁷⁴ **I Efficiencies and observables with alternative mod-**
¹¹⁷⁵ **els**

¹¹⁷⁶ **I.1 Efficiencies**

¹¹⁷⁷ As explained in subsection 10.7, though the efficiencies already include tracking, PID,
¹¹⁷⁸ data/MC discrepancy corrections, the efficiency ratio is obtained after accounting for the
¹¹⁷⁹ effects of trigger, $\sigma_{m_{\text{corr.}}}$, Isolation, BDTs and q^2 migration corrections.

A.X.Khadra <i>et al.</i>	$\varepsilon_{\text{gen}} (\%)$	$\varepsilon_{\text{recsel}} (\%)$	$\varepsilon_K/\varepsilon_{D_s}$	Yield
No q^2 sel.	20.735 ± 0.081	0.377 ± 0.003	0.670 ± 0.008	13895 ± 491
$q^2 < 7\text{GeV}^2/c^4$	19.729 ± 0.130	0.673 ± 0.010	1.160 ± 0.021	7497 ± 325
$q^2 > 7\text{GeV}^2/c^4$	21.021 ± 0.092	0.299 ± 0.003	0.517 ± 0.008	6398 ± 368

Table 53: Final generator efficiency, reconstruction efficiency, global ratio $\varepsilon_K/\varepsilon_{D_s}$, and fitted yield.

C.Bouchard <i>et al.</i>	$\varepsilon_{\text{gen}} (\%)$	$\varepsilon_{\text{recsel}} (\%)$	$\varepsilon_K/\varepsilon_{D_s}$	Yield
No q^2 sel.	20.358 ± 0.081	0.470 ± 0.004	0.820 ± 0.011	13249 ± 467
$q^2 < 7\text{GeV}^2/c^4$	19.642 ± 0.122	0.656 ± 0.007	1.104 ± 0.017	6712 ± 285
$q^2 > 7\text{GeV}^2/c^4$	20.886 ± 0.107	0.342 ± 0.004	0.606 ± 0.010	6537 ± 370

Table 54: Final generator efficiency, reconstruction efficiency, global ratio $\varepsilon_K/\varepsilon_{D_s}$, and fitted yield.

Flynn <i>et al.</i>	$\varepsilon_{\text{gen}} (\%)$	$\varepsilon_{\text{recsel}} (\%)$	$\varepsilon_K/\varepsilon_{D_s}$	Yield
No q^2 sel.	20.653 ± 0.081	0.398 ± 0.004	0.704 ± 0.010	13726 ± 482
$q^2 < 7\text{GeV}^2/c^4$	19.711 ± 0.156	0.668 ± 0.009	1.150 ± 0.020	7307 ± 313
$q^2 > 7\text{GeV}^2/c^4$	20.984 ± 0.095	0.309 ± 0.004	0.539 ± 0.009	6419 ± 366

Table 55: Final generator efficiency, reconstruction efficiency, global ratio $\varepsilon_K/\varepsilon_{D_s}$, and fitted yield.

¹¹⁸⁰ **I.2 Observables**

¹¹⁸¹ Tables 57,58 and 59 show the comparisons of the results for the branching fraction ratios
¹¹⁸² while 60, 61 and 62. are showing the same but for $\frac{|V_{ub}|}{|V_{cb}|}$.

¹¹⁸³ Assuming the same systematics as for the baseline, the comparison plot for $\frac{|V_{ub}|}{|V_{cb}|}$ in the
¹¹⁸⁴ full q^2 range is obtained in Fig.118.

Khodjamirian and Rusov	$\varepsilon_{\text{gen}} (\%)$	$\varepsilon_{\text{recsel}} (\%)$	$\varepsilon_K/\varepsilon_{D_s}$	Yield
No q^2 sel.	20.426 ± 0.081	0.454 ± 0.004	0.795 ± 0.011	13412 ± 466
$q^2 < 7\text{GeV}^2/c^4$	19.669 ± 0.130	0.658 ± 0.008	1.120 ± 0.018	6922 ± 294
$q^2 > 7\text{GeV}^2/c^4$	20.878 ± 0.103	0.340 ± 0.004	0.602 ± 0.010	6490 ± 362

Table 56: Final generator efficiency, reconstruction efficiency, global ratio $\varepsilon_K/\varepsilon_{D_s}$, and fitted yield.

Model	$R_{BF}(\text{low})$
Baseline	$0.00166 \pm 0.00008(\text{stat})$
A.X.Khadra <i>et al.</i>	$0.00173 \pm 0.00009(\text{stat})$
C.Bouchard <i>et al.</i>	$0.00163 \pm 0.00008(\text{stat})$
Flynn <i>et al.</i>	$0.00170 \pm 0.00009(\text{stat})$
Khodjamirian and Rusov	$0.00165 \pm 0.00008(\text{stat})$

Table 57: Comparison of branching fraction ratios for low q^2 .

Model	$R_{BF}(\text{high})$
Baseline	$0.00325 \pm 0.00021(\text{stat})$
A.X.Khadra <i>et al.</i>	$0.00331 \pm 0.00021(\text{stat})$
C.Bouchard <i>et al.</i>	$0.00289 \pm 0.00018(\text{stat})$
Flynn <i>et al.</i>	$0.00319 \pm 0.00020(\text{stat})$
Khodjamirian and Rusov	$0.00288 \pm 0.00018(\text{stat})$

Table 58: Comparison of branching fraction ratios for high q^2 .

Model	R_{BF}
Baseline	$0.00489 \pm 0.00021(\text{stat})$
A.X.Khadra <i>et al.</i>	$0.00555 \pm 0.00024(\text{stat})$
C.Bouchard <i>et al.</i>	$0.00433 \pm 0.00019(\text{stat})$
Flynn <i>et al.</i>	$0.00521 \pm 0.00023(\text{stat})$
Khodjamirian and Rusov	$0.00452 \pm 0.00020(\text{stat})$

Table 59: Comparison of branching fraction ratios for the full q^2 range.

Model	$\frac{ V_{ub} }{ V_{cb} }(\text{low})$
Baseline	$0.0607 \pm 0.0015(\text{stat}) \pm 0.0030(\text{FF})$
A.X.Khadra <i>et al.</i>	$0.1297 \pm 0.0033(\text{stat}) \pm 0.0332(\text{FF})$
C.Bouchard <i>et al.</i>	$0.0673 \pm 0.0017(\text{stat}) \pm 0.0103(\text{FF})$
Flynn <i>et al.</i>	$0.1148 \pm 0.0029(\text{stat}) \pm 0.0307(\text{FF})$
Khodjamirian and Rusov	$0.0604 \pm 0.0015(\text{stat}) \pm 0.0030(\text{FF})$

Table 60: Comparison of $\frac{|V_{ub}|}{|V_{cb}|}$ for low q^2 .

Model	$\frac{ V_{ub} }{ V_{cb} } (high)$
Baseline	$0.0946 \pm 0.0030(\text{stat}) \pm 0.0068(\text{FF})$
A.X.Khadra <i>et al.</i>	$0.0955 \pm 0.0030(\text{stat}) \pm 0.0069(\text{FF})$
C.Bouchard <i>et al.</i>	$0.0769 \pm 0.0024(\text{stat}) \pm 0.0052(\text{FF})$
Flynn <i>et al.</i>	$0.0930 \pm 0.0029(\text{stat}) \pm 0.0104(\text{FF})$
Khodjamirian and Rusov	$0.0617 \pm 0.0019(\text{stat}) \pm 0.0048(\text{FF})$

Table 61: Comparison of $\frac{|V_{ub}|}{|V_{cb}|}$ for high q^2 .

Model	$\frac{ V_{ub} }{ V_{cb} }$
Baseline	$0.0775 \pm 0.0017(\text{stat}) \pm 0.0040(\text{FF})$
A.X.Khadra <i>et al.</i>	$0.1092 \pm 0.0024(\text{stat}) \pm 0.0120(\text{FF})$
C.Bouchard <i>et al.</i>	$0.0715 \pm 0.0016(\text{stat}) \pm 0.0073(\text{FF})$
Flynn <i>et al.</i>	$0.1025 \pm 0.0022(\text{stat}) \pm 0.0147(\text{FF})$
Khodjamirian and Rusov	$0.0611 \pm 0.0013(\text{stat}) \pm 0.0034(\text{FF})$

Table 62: Comparison of $\frac{|V_{ub}|}{|V_{cb}|}$ for the full q^2 range.

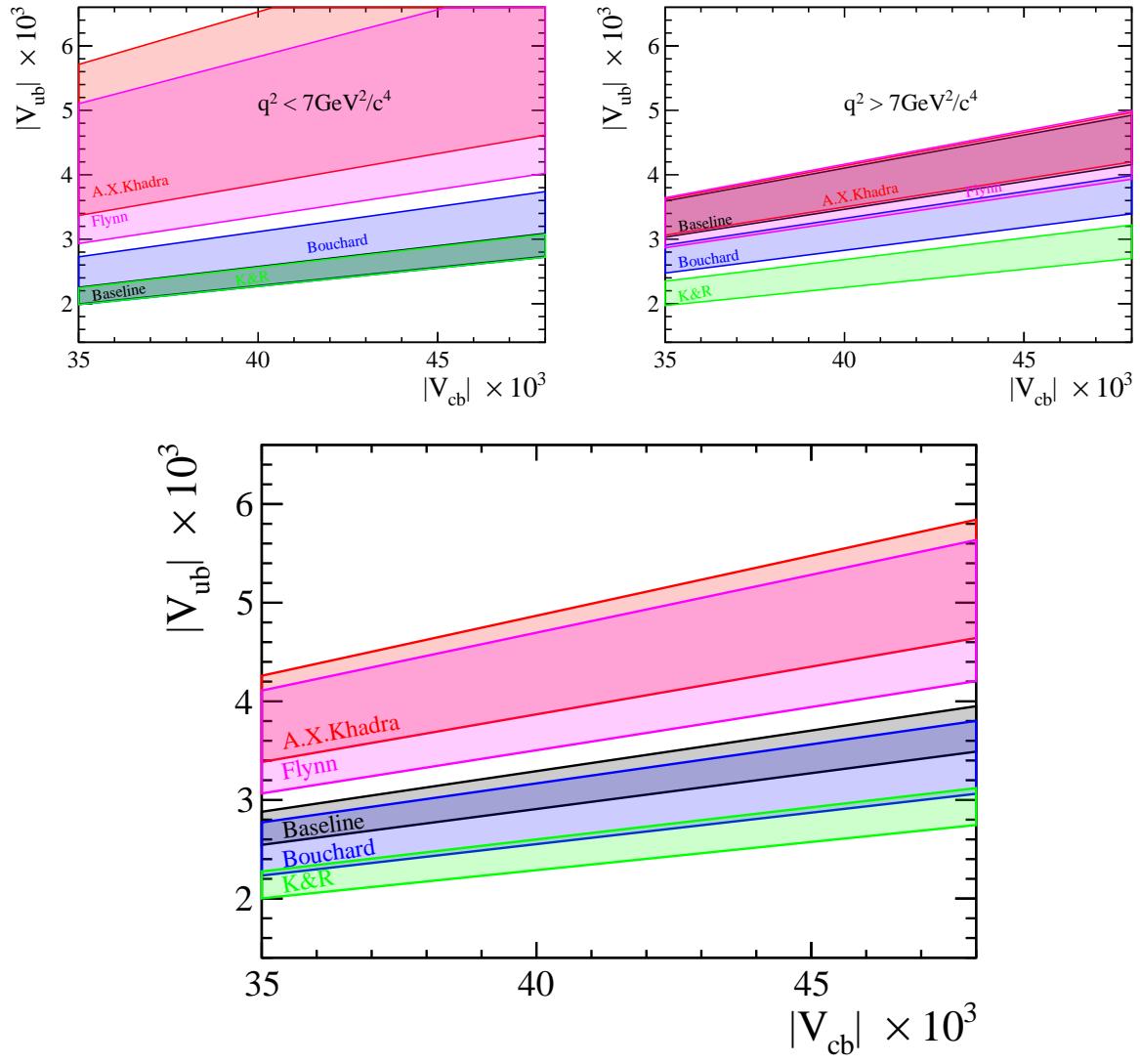


Figure 118: Effect of the Form Factor model choice for the measurement of $\frac{|V_{ub}|}{|V_{cb}|}$ in the $(|V_{cb}|, |V_{ub}|)$ plane.

1185 **References**

- 1186 [1] Particle Data Group, M. Tanabashi *et al.*, *Review of particle physics*, Phys. Rev.
1187 **D98** (2018) 030001.
- 1188 [2] S. Fajfer, J. F. Kamenik, and I. Nišandžić, $b \rightarrow D^* \tau \bar{\nu}_\tau$ sensitivity to new physics,
1189 Phys. Rev. D **85** (2012) 094025.
- 1190 [3] H. J. Rothe, *Lattice gauge theories: An Introduction*, World Sci. Lect. Notes Phys.
1191 **43** (1992) 1, [World Sci. Lect. Notes Phys.82,1(2012)].
- 1192 [4] C. Gatringer and C. B. Lang, *Quantum chromodynamics on the lattice*, Lect. Notes
1193 Phys. **788** (2010) 1.
- 1194 [5] C. A. Dominguez, *Introduction to QCD sum rules*, Mod. Phys. Lett. **A28** (2013)
1195 1360002, [arXiv:1305.7047](#).
- 1196 [6] J. A. Bailey *et al.*, $B_s \rightarrow D_s/B \rightarrow D$ Semileptonic Form-Factor Ratios and Their
1197 Application to $BR(B_s^0 \rightarrow \mu^+ \mu^-)$, Phys. Rev. **D85** (2012) 114502, [arXiv:1202.6346](#),
1198 [Erratum: Phys. Rev.D86,039904(2012)].
- 1199 [7] C. J. Monahan *et al.*, $B_s \rightarrow D_s \ell \nu$ Form Factors and the Fragmentation Fraction
1200 Ratio f_s/f_d , Phys. Rev. **D95** (2017), no. 11 114506, [arXiv:1703.09728](#).
- 1201 [8] E. McLean, C. T. H. Davies, J. Koponen, and A. T. Lytle, $B_s \rightarrow D_s \ell \nu$ Form Factors
1202 for the full q^2 range from Lattice QCD with non-perturbatively normalized currents,
1203 [arXiv:1906.00701](#).
- 1204 [9] C. Bourrely, L. Lellouch, and I. Caprini, Model-independent description of $b \rightarrow \pi \ell \nu$
1205 decays and a determination of $|V_{ub}|$, Phys. Rev. D **79** (2009) 013008.
- 1206 [10] C. M. Bouchard *et al.*, $B_s \rightarrow K \ell \nu$ form factors from lattice QCD, Phys. Rev. **D90**
1207 (2014) 054506, [arXiv:1406.2279](#).
- 1208 [11] J. M. Flynn *et al.*, $B \rightarrow \pi \ell \nu$ and $B_s \rightarrow K \ell \nu$ form factors and $|V_{ub}|$ from 2+1-flavor
1209 lattice QCD with domain-wall light quarks and relativistic heavy quarks, Phys. Rev.
1210 **D91** (2015), no. 7 074510, [arXiv:1501.05373](#).
- 1211 [12] Fermilab Lattice, MILC, A. Bazavov *et al.*, $B_s \rightarrow K \ell \nu$ decay from lattice QCD, Phys.
1212 Rev. **D100** (2019), no. 3 034501, [arXiv:1901.02561](#).
- 1213 [13] A. Khodjamirian and A. V. Rusov, $B_s \rightarrow K \ell \nu_\ell$ and $B_{(s)} \rightarrow \pi(K) \ell^+ \ell^-$ decays at large
1214 recoil and CKM matrix elements, JHEP **08** (2017) 112, [arXiv:1703.04765](#).
- 1215 [14] Particle Data Group, C. Patrignani *et al.*, *Review of Particle Physics*, Chin. Phys.
1216 **C40** (2016), no. 10 100001.
- 1217 [15] S. Dambach, U. Langenegger, and A. Starodumov, *Neutrino reconstruction with
1218 topological information*, Nuclear Instruments and Methods in Physics Research Section
1219 A: Accelerators, Spectrometers, Detectors and Associated Equipment **569** (2006),
1220 no. 3 824 .

- [16] S. Stone and L. Zhang, *Method of Studying Λ_b^0 decays with one missing particle*, Adv. High Energy Phys. **2014** (2014) 931257, [arXiv:1402.4205](#).
- [17] G. Ciezarek, A. Lupato, M. Rotondo, and M. Vesterinen, *Reconstruction of semileptonically decaying beauty hadrons produced in high energy pp collisions*, JHEP **02** (2017) 021, [arXiv:1611.08522](#).
- [18] A. C. Rencher and W. F. Christensen, *Methods of multivariate analysis*, Wiley, 2012.
- [19] D. A. Freedman, *Statistical Models: Theory and Practice*, Cambridge University Press, 2009.
- [20] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to Linear Regression Analysis*, Wiley, 2013.
- [21] F. Pedregosa *et al.*, *Scikit-learn: Machine Learning in Python*, ArXiv e-prints (2012) [arXiv:1201.0490](#).
- [22] D. Scora and N. Isgur, *Semileptonic meson decays in the quark model: an update*, Phys. Rev. D **52** (1995) 2783.
- [23] LHCb HLT project, J. Albrecht, V. V. Gligorov, G. Raven, and S. Tolk, *Performance of the LHCb High Level Trigger in 2012*, J. Phys. Conf. Ser. **513** (2014) 012001, [arXiv:1310.8544](#).
- [24] A. Puig, *The LHCb trigger in 2011 and 2012*, Tech. Rep. LHCb-PUB-2014-046. CERN-LHCb-PUB-2014-046, CERN, Geneva, Nov, 2014.
- [25] M. Williams *et al.*, *The HLT2 Topological Lines*, Tech. Rep. LHCb-PUB-2011-002. CERN-LHCb-PUB-2011-002, CERN, Geneva, Jan, 2011.
- [26] R. Aaij *et al.*, *The LHCb Trigger and its Performance in 2011*, JINST **8** (2013) P04022, [arXiv:1211.3055](#).
- [27] O. Deschamps *et al.*, *Photon and neutral pion reconstruction*, LHCb-2003-091, CERN-LHCb-2003-091.
- [28] J. Albrecht *et al.*, *Search for Lepton Flavor Violation and Baryon Number Violation in τ decays*, LHCb-ANA-2012-095.
- [29] C. Bozzi *et al.*, *Measurement of the mixing frequency Δm_d using semileptonic B^0 decays*, LHCb-ANA-2013-079.
- [30] F. Archilli *et al.*, *Search for the $B^0 \rightarrow \mu^+ \mu^-$ decay and measurement of the $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction and effective lifetime*, LHCb-ANA-2016-038.
- [31] G. Mancinelli and J. Serrano, *Study of Muon Isolation in $B_s^0 \rightarrow \mu^+ \mu^-$ the Channel*, LHCb-INT-2010-011.
- [32] L. Gavardi, M. Calvi, and J. Albrecht, *Search for lepton flavour violation in τ decays at the LHCb experiment*, Nov, 2013. Presented 28 Nov 2013.

- 1256 [33] Y. Freund and R. E. Schapire, *A Decision-Theoretic Generalization of On-Line*
 1257 *Learning and an Application to Boosting*, J. Comput. Syst. Sci. **55** (1997), no. 1 119.
- 1258 [34] A. Rogozhnikov, *Reweighting with Boosted Decision Trees*, J. Phys. Conf. Ser. **762**
 1259 (2016), no. 1 012036, [arXiv:1608.05806](https://arxiv.org/abs/1608.05806).
- 1260 [35] M. Pivk and F. R. Le Diberder, *Splot: A statistical tool to unfold data distributions*,
 1261 Nucl. Instrum. Meth. **A555** (2005) 356, [arXiv:physics/0402083](https://arxiv.org/abs/physics/0402083).
- 1262 [36] G. J. Besjes *et al.*, *HistFitter: a flexible framework for statistical data analysis*, J.
 1263 Phys. Conf. Ser. **664** (2015), no. 7 072004.
- 1264 [37] R. J. Barlow and C. Beeston, *Fitting using finite Monte Carlo samples*, Comput.
 1265 Phys. Commun. **77** (1993) 219.
- 1266 [38] J. Wimberley, B. K. Hamilton, and H. Jawahery, *A measurement of the semitauonic
 1267 branching fraction ratio $\mathcal{R}(J/\psi)$* , LHCb-ANA-2016-059.
- 1268 [39] G. D'Agostini, *A Multidimensional unfolding method based on Bayes' theorem*, Nucl.
 1269 Instrum. Meth. **A362** (1995) 487.
- 1270 [40] A. Hoecker and V. Kartvelishvili, *SVD Approach to Data Unfolding*, Nucl. Instrum.
 1271 Meth. **A372** (1996) 469, [arXiv:hep-ph/9509307](https://arxiv.org/abs/hep-ph/9509307).
- 1272 [41] U. Egede, P. Owen, and W. Sutcliffe, *Determination of the quark coupling strength
 1273 $|V_{ub}|$ using baryonic decays*, LHCb-ANA-2014-048.
- 1274 [42] L. Anderlini *et al.*, *The PIDCalib package*, Tech. Rep. LHCb-PUB-2016-021. CERN-
 1275 LHCb-PUB-2016-021, CERN, Geneva, Jul, 2016.
- 1276 [43] LHCb collaboration, R. Aaij *et al.*, *Determination of the quark coupling strength $|V_{ub}|$
 1277 using baryonic decays*, Nature Physics **11** (2015) 743, [arXiv:1504.01568](https://arxiv.org/abs/1504.01568).
- 1278 [44] S. Tolk, J. Albrecht, F. Dettori, and A. Pellegrino, *Data driven trigger efficiency
 1279 determination at LHCb*, Tech. Rep. LHCb-PUB-2014-039. CERN-LHCb-PUB-2014-
 1280 039, CERN, Geneva, May, 2014.
- 1281 [45] LHCb, R. Aaij *et al.*, *Measurement of the track reconstruction efficiency at LHCb*,
 1282 JINST **10** (2015), no. 02 P02007, [arXiv:1408.1251](https://arxiv.org/abs/1408.1251).
- 1283 [46] M. De Cian *et al.*, *Measurement of the track finding efficiency*, Tech. Rep. LHCb-
 1284 PUB-2011-025. CERN-LHCb-PUB-2011-025, CERN, Geneva, Apr, 2012.
- 1285 [47] D. M. Straub, *flavio package*, arXiv:1810.08132, Oct., 2018.
- 1286 [48] A. Bharucha, D. M. Straub, and R. Zwicky, *$B \rightarrow V\ell^+\ell^-$ decay in the Standard Model from Light-Cone Sum Rules*, J. High Energ. Phys. **08** (2016), no. 98 ,
 1287 [arXiv:1503.05534](https://arxiv.org/abs/1503.05534).
- 1289 [49] Particle Data Group, P. A. Zyla *et al.*, *Review of particle physics*, Prog. Theor. Exp.
 1290 Phys. **2020** (2020) 083C01.

₁₂₉₁ [50] LHCb, R. Aaij *et al.*, *Test of Lepton Flavor Universality by the measurement of*
₁₂₉₂ *the $B^0 \rightarrow D^{*-} \tau^+ \nu_\tau$ branching fraction using three-prong τ decays*, Phys. Rev. **D97**
₁₂₉₃ (2018), no. 7 072013, [arXiv:1711.02505](https://arxiv.org/abs/1711.02505).

₁₂₉₄ [51] LHCb, R. Aaij *et al.*, *Measurement of the shape of the $\Lambda_b^0 \rightarrow \Lambda_c^+ \mu^- \bar{\nu}_\mu$ differential*
₁₂₉₅ *decay rate*, Phys. Rev. **D96** (2017), no. 11 112005, [arXiv:1709.01920](https://arxiv.org/abs/1709.01920).