

Search for the $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu_\mu$ decay at LHCb

Slavomira Stefkova¹, Patrick Owen², Ulrik Egede¹.

¹*Imperial College London, London, United Kingdom*

²*Universitat Zurich, Zurich, Switzerland*

Abstract

This analysis note describes a search for the fully-leptonic annihilation-type $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu_\mu$ decay using data collected with the LHCb detector, corresponding to integrated luminosity $\approx 4.7 \text{ fb}^{-1}$. This measurement is performed in a specific range of dimuon mass, where there is an upper limit for the minimum dimuon mass combination to be less $980 \text{ MeV}/c^2$. This decay has never been observed before and is sensitive to the least precisely determined CKM matrix element, $|V_{ub}|$. The analysis had blinding procedure in place where signal region was not seen until full strategy was set. Upon unblinding no significant signal was observed and hence upper limit on the branching fraction was set yielding $\mathcal{B}(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu) < 1.4 \times 10^{-8}$ at 95% confidence level.

Versions - Comments

- **v0r0 :**
 - Circulation to SLWG
- **v0r1 :**
 - Added subsubsection 5.3.1 showing efficiency variation across the phase space for partially reconstructed background
 - Added subsubsection 5.3.2 showing considerations about partially reconstructed $D^0 \rightarrow \eta X$ background
 - Added formula for extraction of systematic uncertainty for trigger Equation 46
 - Added systematic for HLT2 one TCK generalisation, see Table 59
 - Added Figure 47 showing the selection efficiency for signal as a function of $minq^2$
 - Added correlation matrices for final simultaneous fits to blinded data Figures 49 and 50
- **v0r2 :**
 - Added subsection 3.3 to show full data sample logic
 - Implemented comments from Eduardo, first batch
- **v0r3 :**
 - Responded to all of the comments from 1st iteration of full comments
- **v0r4:**
 - 2011 and 2012 MC equivalence of efficiencies discussion in subsection 8.4
 - Efficiency corrected yield of Run 1 vs 2016 shown in subsection 8.3
 - List of systematics and new sensitivities that were used in the fit discussed in section 13
 - Corrected mass split on signal simulation discussed in subsubsection 9.1.1
 - Systematics associated with partially reconstructed backgrounds subsection 10.5
 - New sensitivities reflecting all systematics section 13
- **v0r5:**
 - Implemented RC comments on v4r0. Last version before unblinding.
- **v0r6:**
 - Unblinded data
 - Set limit
 - Added sections: Results(section 14) and Conclusion(section 15)
 - Added result into the abstract
- **v0r7:**
 - Re-explained the procedure for the fitting in subsection 9.3

Contents

1	Introduction	1
2	Analysis strategy	3
3	Data and Simulation Samples	5
3.1	Data Samples	5
3.1.1	Signal and Normalisation Data Samples	5
3.1.2	Misidentified (misID) background samples	6
3.1.3	Combinatorial Sample	6
3.1.4	Control Samples for misID weights	6
3.2	Simulation Samples	6
3.2.1	Phasespace Signal MC	7
3.2.2	More Realistic Signal MC	7
3.2.3	Signal MC - Model from Nikolai Nikitin	8
3.2.4	Partially Reconstructed MC Sample	8
3.2.5	Inclusive $b\bar{b}$ MC dimuon sample	9
3.3	Analysis flow	9
4	Selection	11
4.1	Choice of q^2 region	11
4.2	Stripping	12
4.3	Trigger Selection	13
4.4	Pre-selection	14
4.4.1	<code>nShared</code> variable	17
4.5	Multivariate Selection	19
4.5.1	Isolation	20
4.5.2	Combinatorial BDT	21
4.5.3	MisID BDT	27
4.6	Fitting Region Selection	31
4.7	<code>Probnn</code> Optimization	31
4.8	Fractional Corrected Mass Window Split	32
4.9	Multiple Candidates	33
5	Background	34
5.1	Inclusive simulation studies	34
5.2	Combinatorial Background	35
5.3	Partially Reconstructed $B^+ \rightarrow D\mu^+\nu$ type backgrounds	36
5.3.1	Efficiency Across Phase Space	39
5.3.2	Partially Reconstructed Backgrounds, where $D^0 \rightarrow \eta/\eta'X$, where $\eta/\eta' \rightarrow \mu\mu\gamma$	41
5.4	MisID type background	41
5.5	Rare $B^+ \rightarrow \pi^+/K^+\mu^-\mu^+$ Background	45
5.6	$B \rightarrow \eta(')V$ Backgrounds	45
5.7	Summary	46

6 Specific control sample for $K/\pi \rightarrow \mu$ misID rates	46
6.1 Selection for $B^0 \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)K^*$	47
6.2 Fitting Strategy for $B^0 \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)K^*$ decay	47
6.3 Results of $B^0 \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)K^*$ control sample for $K/\pi \rightarrow \mu$ misID rates	50
7 Efficiency	54
7.1 Efficiency Ratio	54
7.2 Detector Acceptance Efficiencies (GEN)	55
7.3 Reconstruction (REC) Efficiency	56
7.4 Trigger Efficiency (TRG)	57
7.5 Offline Selection (OFF)	61
7.6 Combinatorial BDT and Misid BDT Efficiencies	61
7.7 Fitting Range Efficiency (FR)	65
7.8 Fractional Corrected Mass Split Efficiency	66
7.9 PID Efficiencies (PID)	66
7.10 Summary of Individual MC and/or data efficiencies necessary for <i>Single Event Sensitivity</i>	67
8 Normalisation channel	71
8.1 $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+$ Fit	71
8.2 Single Event Sensitivity	74
8.3 Efficiency Corrected Yields	75
8.4 Efficiency equivalence between Stripping 21 and 21r1	76
9 Corrected Mass Fit	77
9.1 Signal	77
9.1.1 FCME split consideration	79
9.2 Backgrounds	80
9.2.1 MisID background	80
9.2.2 Partially Reconstructed Background	83
9.2.3 Combinatorial Background	85
9.3 Blinded Data Fit	85
10 Systematic uncertainties	88
10.1 Trigger	88
10.2 Signal Decay Model	89
10.3 Fit Bias	91
10.4 Kinematic Reweighting	95
10.5 Partially Reconstructed Background	97
11 Crosschecks	98
11.1 Validation of q^2 cut	98
11.2 AntiMisidBDT study (Control sample study)	99
11.3 Ghosts	102
11.4 Comparison of final shapes of components for final fits	102

12 PID Tuning	104
12.1 Comparison of misID distributions after final selection for Run1 and 2016	104
12.2 Probnnmu V2 vs V3 for Run 1	106
13 Expected Exclusion Limits - Sensitivity studies	107
14 Results	109
14.1 Unblinded Mass Fits	109
14.2 Limit Setting	112
14.3 Additional Results: Non-simultaneous fit limits	113
15 Conclusion	113
A BDT input variables Run 1 and 2016 - Combinatorial BDT	114
A.1 All	114
A.2 Signal MC	117
A.3 Combinatorial data	120
B BDT input variables Run 1 and 2016 - Misid BDT	122
B.1 All	123
B.2 Signal MC	126
B.3 MisID data	129
C BDT Characteristics	131
C.1 Combinatorial BDT	131
C.2 MisID BDT	137
D Partially Reconstructed Background - break down of efficiencies	142
E MC Truth Matching	142
E.1 Signal MC truth matching	142
E.2 Normalisation MC truth matching	143
E.3 PartReco MC truth matching	143
F TCK-dependant Efficiencies 2016	143
F.1 BDT efficiencies	144
F.2 Fit Range Efficiencies	145
F.3 FCME split	146
G Kinematic Reweighting	146
References	149

1 Introduction

In this analysis note, the summary of the search for $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ decay at LHCb using Run 1 and 2016 data is presented. This decay is sensitive to modulus of the matrix element V_{ub} , $|V_{ub}|$.

At present, the largest relative uncertainty of CKM magnitudes is the value of $|V_{ub}|$. The current value of $|V_{ub}|$ is measured through inclusive [1] (from PDG, as they added the last uncertainty) and exclusive [2] semileptonic b decays and yields $|V_{ub}| = (3.94 \pm 0.36) \times 10^{-3}$ [1]. It should be noted that these measurements are in tension:

$$|V_{ub}| = (4.49 \pm 0.15^{+0.16}_{-0.17} \pm 0.17) \times 10^{-3} \text{ (inclusive),} \quad (1)$$

$$|V_{ub}| = (3.70 \pm 0.10 \pm 0.12) \times 10^{-3} \text{ (exclusive).} \quad (2)$$

Even though semileptonic decays provide the best approach currently of how to access $|V_{ub}|$, theoretical uncertainties that are related to the measurements still remain high. An alternative approach of measuring $|V_{ub}|$, with lower theoretical uncertainties, is therefore of particular interest. This could be achieved by constraining $|V_{ub}|$ through fully leptonic decay modes. Different approaches of measuring $|V_{ub}|$ can be seen in Figure 1.

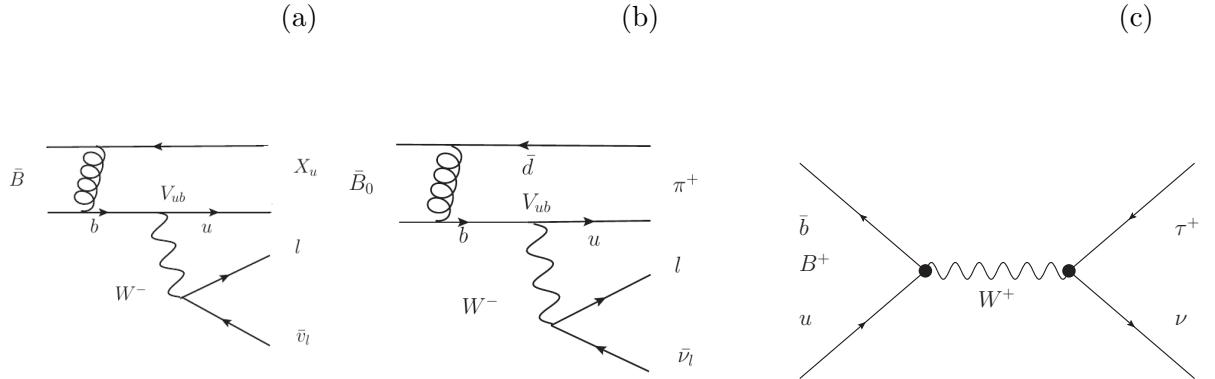


Figure 1: Feynman diagrams of (a) inclusive semileptonic, (b) exclusive semileptonic as well as (c) leptonic decays conducted by flavour experiments to measure $|V_{ub}|$.

One such search, namely the measurement of the branching fraction $\mathcal{B}(B^+ \rightarrow \tau^+ \nu_\tau)$, had already been conducted by Belle and BaBar [3] [4]. In the SM the branching fraction for $B^+ \rightarrow \tau^+ \nu_\tau$ is calculated as

$$\mathcal{B}(B^+ \rightarrow \tau^+ \nu_\tau) = \frac{G_F^2 m_B m_{\tau^+}^2}{8\pi} \left[1 - \frac{m_{\tau^+}^2}{m_B^2} \right]^2 f_B^2 |V_{ub}|^2 \tau_{B^+}, \quad (3)$$

where G_F is the Fermi constant, m_B and m_{τ^+} are the B^+ meson and τ^+ lepton masses, respectively, and τ_{B^+} is the B^+ lifetime.

The experimental difficulty connected with the measurement of the purely leptonic decays of the heavy pseudoscalars is due to helicity suppression where the decay rate is suppressed by a factor of $(m_l/m_h)^2$ where m_l is the mass of the lepton and m_h is the mass of the hadron. Decays with a τ in a final state in LHCb are more complicated to reconstruct as they are short-lived and involve decays with further neutrino(s). The muon identification performance of the LHCb detector makes final states with muons such as $B^+ \rightarrow \mu^+ \nu_\mu$ much easier to reconstruct. But

25 helicity suppression results in a very low expected branching fraction for such a decay, ≈ 225
 26 smaller compared to the tauonic mode. The helicity suppression can be lifted by considering the
 27 decay with an additional photon radiated from the B^+ meson, at the cost of the electromagnetic
 28 suppression with coupling constant α_{em} . Consequently, the branching fraction for radiative decays
 29 can be comparable or even larger than the corresponding fraction for purely leptonic decays. It
 30 has been shown that $R_B^\mu = \frac{\Gamma(B \rightarrow \mu\nu\gamma)}{\Gamma(B \rightarrow \mu\nu)} \approx (1 - 20)$ making $\mathcal{B}(B \rightarrow \mu\nu\gamma) \approx (10^{-7} - 10^{-6})$ [5].

31 Recent measurement of the radiative $B^+ \rightarrow l^+ \nu_l \gamma$, where l^+ is either e^+ or μ^+ was performed
 32 by Belle using hadronic tagging on their full data sample [6]. The differential branching fraction
 33 is given by the following formula:

$$\frac{d\Gamma}{dE_\gamma} = \frac{\alpha_{em} G_F^2 |V_{ub}|^2}{48\pi^2} m_B^4 (1 - x_\gamma) x_\gamma^3 [F_A^2 + F_V^2] \quad (4)$$

34 where $x_\gamma = 2E_\gamma/m_B$, F_A is axial form factor and F_V is vector form factor defined as

$$F_V(E_\gamma) = \frac{Q_u m_B f_B}{2E_\gamma \lambda_B(\mu)} R(E_\gamma, \mu) + [\xi(E_\gamma) + \frac{Q_u m_B f_B}{(2E_\gamma)^2} + \frac{Q_b m_B f_B}{2E_\gamma m_b}], \quad (5)$$

$$F_A(E_\gamma) = \frac{Q_u m_B f_B}{2E_\gamma \lambda_B(\mu)} R(E_\gamma, \mu) + [\xi(E_\gamma) - \frac{Q_u m_B f_B}{(2E_\gamma)^2} - \frac{Q_b m_B f_B}{2E_\gamma m_b} + \frac{Q_l f_B}{E_\gamma}]. \quad (6)$$

35 Here Q_l, Q_u, Q_b are the charges of the lepton, up quark, and bottom quark, respectively and
 36 $R(E_\gamma, \mu)$ is a radiative correction calculated at the energy scale μ . The search yielded $\mathcal{B}(B^+ \rightarrow$
 37 $\mu^+ \nu_\mu \gamma) < 3.4 \times 10^{-6}$ [6].

38 In LHCb, the most optimal approach due to the detector capabilities is to measure this
 39 kind of decay by decaying the photon into pair of muons, see Figure 2(a). This though
 40 introduces suppression by α_{em} . If the naive expectation of only taking into account photon
 41 conversion into two muons is adopted, then the expected branching fraction for this analysis
 42 is $\mathcal{B}(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu_\mu) \approx 10^{-8}$ (will be referred to as pessimistic branching fraction (BF)).
 43 However this estimate is of course not correct because there are other contributions to the total
 44 branching fraction as it was shown in the first theoretical prediction for $\mathcal{B}(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu_\mu)$.

45 The first theoretical prediction published for the $\mathcal{B}(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu_\mu)$ decay branching
 46 fraction [7] [8] yields approximately 1.3×10^{-7} where the majority of the contribution actually
 47 comes from the ρ and ω resonances, which are included as a part of the signal (will be referred
 48 to as optimistic BF). An example of such contribution can be seen in Figure 2(b), where
 49 $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ proceeds via the hadronic contributions to the photon pole.

50 Moreover, once the theoretical formulation for this branching fraction is predicted, extraction
 51 of the value of $|V_{ub}|$ using $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ decay mode will be possible. First results of analysis
 52 towards the measurement of $\mathcal{B}(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu_\mu)$ are presented.

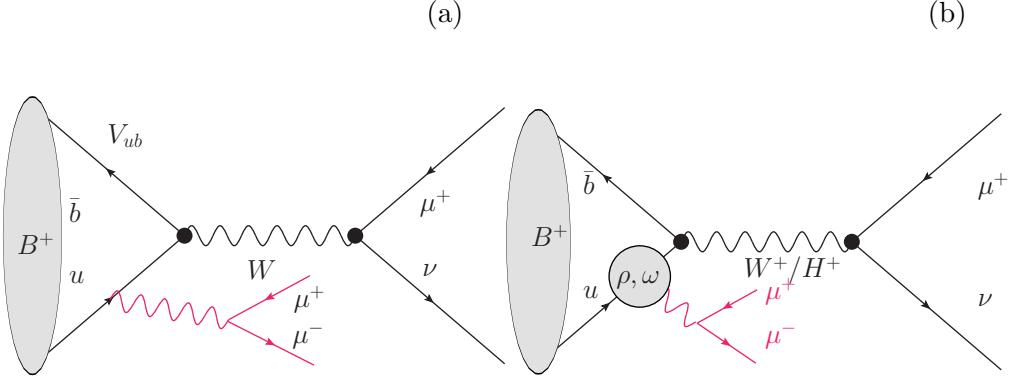


Figure 2: (a) Annihilation diagram where the initial state quark radiates off a virtual photon which decays into a pair of muons with small invariant mass (q^2) and the W^+ turns into a muon and muon neutrino.
 (b) Most of the contribution to the rate comes from hadronic contribution to photon.

53 2 Analysis strategy

54 The analysis of the $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu_\mu$ decay is divided into several different parts; signal
 55 selection, optimisation, normalisation, fitting and limit setting. Throughout this document,
 56 charge conjugates of the decays are assumed unless stated otherwise. Results presented are
 57 based on the analysis of the full 3 fb^{-1} Run 1 dataset as well $\approx 1.7 \text{ fb}^{-1}$ Run 2 data (not
 58 using 2015 dataset due to very low sensitivity (high muon trigger thresholds)). Additionally the
 59 search will be conducted in a particular $minq = \sqrt{\min(q^2(\mu_1^+, \mu^-), q^2(\mu^-, \mu_2^+))}$ region, described
 60 in subsection 4.1.

61 To perform the search for $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu_\mu$, a specific preselection was applied to form
 62 potential signal candidates. To reconstruct the mass of the B^+ with missing information about
 63 the neutrino, a corrected mass variable $M_{\text{B corr}} = \sqrt{M_{3\mu}^2 + |\vec{p}_\perp^2| + |\vec{p}_\perp|}$, where $M_{3\mu}^2$ is the invariant
 64 visible mass squared and \vec{p}_\perp^2 is the missing momentum squared transverse to the direction of
 65 flight of B^+ , is introduced. A simulation sample that mimics the decay of the $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu_\mu$
 66 passing through preselection was used to develop further discriminating selection. To get the
 67 selection efficiency for different types of backgrounds, different proxy samples are used. For more
 68 details about samples used see section 3.

69 Combinatorial background, which arises as random combinations of tracks passing the
 70 preselection, is taken from the upper corrected $\mu^+ \mu^- \mu^+$ mass side band, $M_{\text{B corr}} > 5500 \text{ MeV}/c^2$
 71 , where very few signal candidates are expected. Another significant background arises from
 72 misidentification of particle type. This includes ‘cascade decays’ coming from $b \rightarrow c \rightarrow s$ decay
 73 chains, where the kaon is misidentified as a muon. Distributions for this type of background
 74 are obtained from data, where Particle Identification (PID) preselection on the last muon is
 75 not applied. Finally, for the most dangerous partially reconstructed backgrounds, such as
 76 $B^+ \rightarrow D^0(K^+ \pi^- \mu^+ \mu^-) \mu^+ \nu_\mu$, an inclusive simulation sample was used to study this background.
 77 The selection is described in more detail in section 4 and background studies are summarized
 78 in section 5.

79 After selection, the $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu_\mu$ decay is normalised to the $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+$.
 80 The fit to $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+$ is described in section 8. The relative efficiency between the

81 signal and normalisation channels are computed using simulation, as described in subsection 7.1.
82 The ratio of these efficiencies, together with results of the normalisation channel fit are used
83 to parametrise the expected signal yield, which can be found in section 7.

84 Finally, in subsection 9.3, the fitting strategy that provides the best sensitivity was developed
85 and the expected exclusion limit on the branching fraction is calculated and can be found
86 in section 13. Systematic studies are summarized in section 10.

87 In summary, the aim of the analysis is to fit corrected mass of B^+ in bins of fractional
88 corrected mass error. This is the fitting strategy that provides the best sensitivity (described in
89 detail in subsection 4.8) but the motivation for this particular strategy is that the signal data is
90 split into two bins of resolution, increasing signal separation from background.

91 Throughout the analysis (until v5r0) there was a blinding procedure put in the place, where
92 signal region defined between $4500 \text{ MeV}/c^2 < M_{B_{\text{corr}}} < 5500 \text{ MeV}/c^2$ was blinded, as discussed
93 in subsection 4.6. The anticipated result was either to observe the decay, with Wilk's theorem or
94 set a limit with CLs method.

95 Upon unblinding, no significant signal was observed and stringent limit $\mathcal{B}(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu) <$
96 1.4×10^{-8} at 95% confidence level was set using CLs. All information about the result can be
97 found in section 14. This result is hence does not confirm the prediction for the decay branching
98 fraction [7] [8], where $\mathcal{B}(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu) \approx 1.3 \times 10^{-7}$.

99 3 Data and Simulation Samples

100 The different samples that are used in this analysis are defined in Table 1. The search for
 101 $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ decay is normalised to $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+$ decay, hence data samples
 102 for Signal and Normalisation are necessary. In order to obtain background from misidentified
 103 background (misID), different data samples listed in Table 1, are used. Combinatorial background
 104 is taken from upper mass sideband above $5.5 \text{ GeV}/c^2$. In order to actually obtain the weights for
 105 the $\pi/K \rightarrow \mu$ misID probability $B^0 \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)(K^* \rightarrow \pi^- K^+)$ is necessary.

Channel	Sample Name	More information
$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$	Signal	subsubsection 3.1.1
$B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+$	Normalisation	subsubsection 3.1.1
$B^+ \rightarrow \mu^+ \mu^- X^+ \nu$	MisID SS	subsubsection 3.1.2
$B^+ \rightarrow \mu^+ X^- \mu^+ \nu$	MisID OS	subsubsection 3.1.2
$B^+ \rightarrow \mu^+ \mu^- X^+ \nu$ small	MisID small	subsubsection 3.1.2
$B^+ \rightarrow \mu^+ X^- \mu^+ \nu$ small	MisID small	subsubsection 3.1.2
$B^+ \rightarrow X^+ X^- \mu^+ \nu$ small	MisID small	subsubsection 3.1.2
$B^+ \rightarrow X^+ \mu^- X^+ \nu$ small	MisID small	subsubsection 3.1.2
$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ above $5.5 \text{ GeV}/c^2$	COMBI	subsubsection 3.1.3
$B^0 \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)(K^* \rightarrow \pi^- K^+)$	misID probability CC	subsubsection 3.1.4

Table 1: List of data samples used for this analysis in addition to control samples from **PIDCalib**.

106 Simulation events are used to study the background contaminations, train the multivariate
 107 classifiers and most importantly to determine the efficiencies for efficiency ratios between the
 108 different decays considered.

109 3.1 Data Samples

110 3.1.1 Signal and Normalisation Data Samples

111 The signal data samples for $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ were stripped using the **B23MuNuTriMuLine** (SL
 112 stream) and for normalisation channel $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+$ corresponding stripping line was
 113 **BetaSBu2JpsiKDetachedLine** (Dimuon stream) for both Run 1 and 2016 data. All properties
 114 can be seen in Table 2.

Year	\sqrt{s} TeV	Integrated luminosity(SL) [pb $^{-1}$]	Integrated luminosity(Dimuon) [pb $^{-1}$]	Reco Version	Stripping Version
2011	7	981	976	14	21r1
2012	8	1990	1987	14	21
2016	13	1794	1793	15a	26

Table 2: Integrated luminosities of data samples used in this analysis collected each year at the LHCb experiment. Values are taken from the output of the luminosity tool stored in the ntuples used in the analysis. The uncertainty on the quoted numbers is about 1.5% for Run 1 and 4% for Run 2.

115 **3.1.2 Misidentified (misID) background samples**

116 To obtain the shape and level of the misID backgrounds wherever necessary for this analysis,
117 stripping line `B23MuNu_TriFakeMuLine` was added, which is the same line as `B23MuNuTriMuLine`,
118 but with last muon candidate having no PID applied (taken from `StdNoPIDMuons` container -
119 selects all the tracks in the given data sample and assigns them the muon mass hypothesis with
120 no PID imposed). A prescale of 0.01 is used due to the size of the sample. This misID sample
121 consists of two components: $B^+ \rightarrow \mu^+\mu^-X^+\nu$, $B^+ \rightarrow \mu^+X^-\mu^+\nu$. The sample with misidentified
122 muon having the same charge as the mother $B(B^+ \rightarrow \mu^+\mu^-X^+\nu)$ is referred to as same sign
123 misID sample (SS misID) or opposite charge misID sample ($B^+ \rightarrow \mu^+X^-\mu^+\nu$) (OS misID).

124 Additionally, small data samples ($B^+ \rightarrow \mu^+\mu^-X^+\nu$ small, $B^+ \rightarrow \mu^+X^-\mu^+\nu$ small), with
125 5.8 pb⁻¹ luminosity of Run 1 data with no prescale ¹ was used and yielded same results taking
126 into account the relative normalisation with respect to the other sample. Moreover this small
127 data sample was also used in assessing the misID rate from double misID processes, where either
128 two same sign ($B^+ \rightarrow X^+\mu^-X^+\nu$ small) or opposite sign muons ($B^+ \rightarrow X^+X^-\mu^+\nu$ small) were
129 misidentified, and this background was found negligible compared to single misID.

130 **3.1.3 Combinatorial Sample**

131 A data sample with corrected mass above 5500 MeV/c². This sample will be denoted as COMBI
132 data sample.

133 **3.1.4 Control Samples for misID weights**

134 In this analysis, misID PID weights cannot be obtained from standard control samples. This is
135 because correlations of three muons in a system are not taken into account given that most of
136 these samples have no other muon in the decay. What is done instead, $B^0 \rightarrow (J/\psi \rightarrow \mu^+\mu^-)(K^* \rightarrow$
137 $\pi^-K^+)$ decays are used for this purpose since there are two real muons along with kaon or pion.
138 Hence misID probability of kaon are pion can be determined. For more details, see section 6.

139 **3.2 Simulation Samples**

140 Full (*generator-level + detector-level*) simulation samples used in this analysis are listed in Table 3.
141 These simulations are used for signal shapes for mass fits and for most of the efficiency estimations.
142 There is no signal simulation generated for signal in 2011, so in efficiency estimations 2012 signal is
143 representative of 2011 and 2012 data. This assumption is validated with normalisation simulation,
144 where 2012 selection efficiency is very similar to 2011 selection efficiency.

145 *Generator-level* simulations used in the analysis are listed Table 4. Some of them are used for
146 evaluating *generator – level* efficiencies and some of them are used to cross-check q^2 selection.

147 For signal simulation, three different decay model were applied and are summarized in Section
148 3.2.1, 3.2.2, 3.2.3. As default, event type 12513070 is used.

¹Included runs: 130390, 130391, 130392, 130393, 130394, 130395, 130396

Channel	Year	Pythia	ProcessingPass
$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$	2012	Pythia6	/Sim08h/Digi13/Trig0x409f0045/Reco14c/Stripping20NoPrescalingFlagged
$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$	2012	Pythia8	/Sim08h/Digi13/Trig0x409f0045/Reco14c/Stripping20NoPrescalingFlagged
$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$	2016	Pythia8	/Sim09b/Trig0x6138160F/Reco16/Turbo03/Stripping26NoPrescalingFlagged
$B^+ \rightarrow J/\psi K^+$	2011	Pythia6	/Sim08c/Digi13/Trig0x40760037/Reco14a/Stripping20r1NoPrescalingFlagged
$B^+ \rightarrow J/\psi K^+$	2011	Pythia8	/Sim08c/Digi13/Trig0x40760037/Reco14a/Stripping20r1NoPrescalingFlagged
$B^+ \rightarrow J/\psi K^+$	2012	Pythia6	/Sim08h/Digi13/Trig0x409f0045/Reco14c/Stripping20NoPrescalingFlagged
$B^+ \rightarrow J/\psi K^+$	2012	Pythia8	/Sim08h/Digi13/Trig0x409f0045/Reco14c/Stripping20NoPrescalingFlagged
$B^+ \rightarrow J/\psi K^+$	2016	Pythia8	/Sim09b/Trig0x6138160F/Reco16/Turbo03/Stripping26NoPrescalingFlagged
$B^+ \rightarrow J/\psi \pi^+$	2012	Pythia6	/Sim08a/Digi13/Trig0x409f0045/Reco14a/Stripping20NoPrescalingFlagged
$B^+ \rightarrow J/\psi \pi^+$	2012	Pythia8	/Sim08a/Digi13/Trig0x409f0045/Reco14a/Stripping20NoPrescalingFlagged
$B^0 \rightarrow J/\psi K^*$	2011	Pythia8	/Sim08f/Digi13/Trig0x40760037/Reco14a/Stripping20r1NoPrescalingFlagged
$B^0 \rightarrow J/\psi K^*$	2012	Pythia8	/Sim08f/Digi13/Trig0x409f0045/Reco14a/Stripping20NoPrescalingFlagged
$B^0 \rightarrow J/\psi K^*$	2016	Pythia8	/Sim09b/Trig0x6138160F/Reco16/Turbo03/Stripping26NoPrescalingFlagged
PartReco	2012	Pythia8	/Sim08g/Digi13/Trig0x409f0045/Reco14a/Stripping20NoPrescalingFlagged

Table 3: List of full MC samples used in this analysis with processing conditions. All candidates are also required to have their daughters in the LHCb acceptance, i.e within [10,400]mrad along the beam pipe. For both $B^+ \rightarrow J/\psi K^+$ and $B^0 \rightarrow J/\psi K^*$ samples, $J/\psi \rightarrow \mu^+ \mu^-$. PartReco stands for partially reconstructed sample as discussed in prpr.

Channel	Year	Pythia	EVTGEN	Size	Stage
Simulation used for evaluating <i>generator-level</i> efficiencies					
$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$	2012	Pythia 6	PHSP	25000	<i>generator-level</i>
$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$	2012	Pythia 6	DEFAULT	25000	<i>generator-level</i>
$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$	2012	Pythia 8	DEFAULT	25000	<i>generator-level</i>
Simulation used for ratification of <i>minq</i> selection					
$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$	2012	Pythia 6	NIKI	25000	<i>generator-level</i>

Table 4: Summary of all *generator-level* signal simulation samples used in this analysis with different decay models. In all cases the daughters of B^\pm are required to be within LHCb acceptance. All of this samples are mixture under magnetic polarity up and magnetic polarity down conditions.

149 3.2.1 Phasespace Signal MC

150 Event type: 12513020.

151 Decay model: Full phase space model : PHSP.

152 Requested : 1M events.

153 3.2.2 More Realistic Signal MC

154 Event type: 12513070.

155 Decay model: B^+ decays into $K^*(\mu^+, \nu_\mu) \mu^- \mu^+$ using BTOSLLBALL decay model and $K^*(\mu^+, \nu_\mu)$

then decays into $\mu^+\nu_\mu$ using phase space decay model (PHSP). BTOSLLBALL decay model [9], is traditionally used for $B \rightarrow (K, K^*)l^+l^-$ decay, with the form factor calculations using the light cone QCD sum rule approach.

Requested: 1M events.

A more representative MC sample was produced with more appropriate decay model. The decay proceeds through a virtual W decaying to $\mu^+\nu$ and a virtual photon decaying to a muon pair. This has similar structure to $B^+ \rightarrow (K^{*+})\mu^+\mu^-$ decay, where the K^{*+} can take the role of the virtual W decay. The aim is hence to produce a distribution that is flat for $K^*(\mu^+, \nu_\mu)$ (in Figure 3(a)) and has photon pole for low $q(\mu^+, \mu^-)$ (in Figure 3 (b)) to resemble behaviour of other semileptonic decays. In order to do this a new particle, $K^*(\mu^+, \nu_\mu)$, was introduced to EvtGen with specific properties. Output with the flattest distribution for K^* while having a photon pole for q can be seen to be for a particle $K^*(\mu^+, \nu_\mu)$ with mass set to $0.1 \text{ GeV}/c^2$ and width corresponding to $\tau = 1.3 \times 10^{-17}$ nanoseconds. This procedure was also applied for the charge conjugate case.

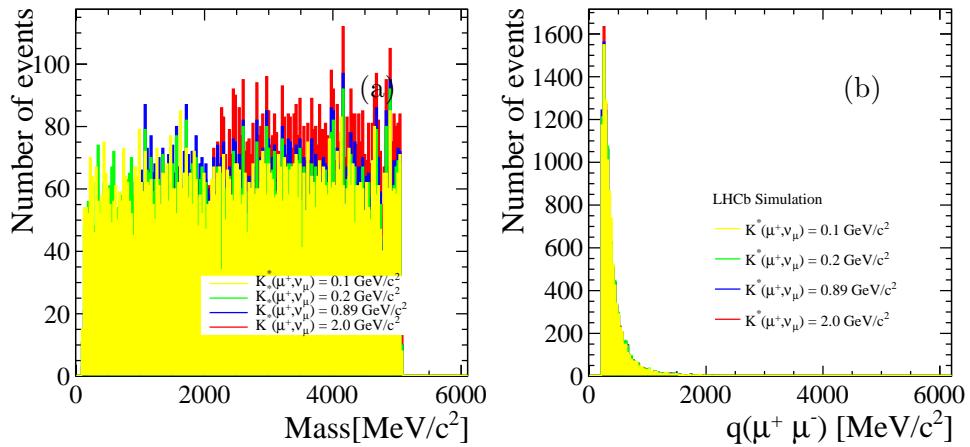


Figure 3: Distributions for signal MC in using Pythia 6 conditions. (a) $K^*(\mu^+, \nu_\mu)$ (b) $q(\mu^+, \mu^-)$ distributions under different K^* mass hypotheses. The most flat distribution in $K^*(\mu^+, \nu_\mu)$ is plotted in yellow.

3.2.3 Signal MC - Model from Nikolai Nikitin

Signal MC model that was produced exclusively for this analysis and committed to EVTGEN, as seen in this [JIRA task link](#), that uses decay model by Nikolai Nikitin and Anna Danilina [7] which relies on meson vector dominance model that models hadronic contributions such as ρ and ω to virtual photon. As seen in 4 in this case only *generator-level* simulation was produced due to timing of the model release. However, this generator-level simulation is useful in order to validate q^2 selection, discussed in 11.1.

3.2.4 Partially Reconstructed MC Sample

An inclusive MC cocktail sample simulating $B^+ \rightarrow (D^0 \rightarrow K_S^0 \pi^+ \pi^- \pi^0)\mu^+\nu_\mu$ decays is used as a proxy for contamination of $B^+ \rightarrow (D^0 \rightarrow K^+ \pi^- \mu^+ \mu^-)\mu^+\nu$ decays. Two charged pions from the

180 D^0 decay are assumed to be signal muons. This inclusive MC contains higher excited resonance
181 of D such as D^{*0} , D_2^{*0} , etc. (see Table 13 for more details). This sample can be found under
182 `EventType:12875540`. This sample will be denoted as PartReco MC throughout the analysis.

183 **3.2.5 Inclusive $b\bar{b}$ MC dimuon sample**

184 In order to gauge the severity of different backgrounds, a sample of simulated inclusive $b\bar{b}$
185 dimuon events is analysed (`EventType:10012004`). This sample has following generator selection:
186 the $p > 3$ GeV/c for the two muons forming the dimuon candidate and no sign requirement.
187 Generator level cuts used with this sample are `LHCb acceptance`, `DiLeptonInAcceptance`.
188 `DiLeptonInAcceptance` requires that in the event there are two leptons - muons, within 10-400
189 mrad of fiducial region, having dimuon mass constraint between 0-100 GeV/c². `LHCb acceptance`
190 make all other tracks to be within 0-400 mrad.

191 **3.3 Analysis flow**

192 The flowchart of analysis can be seen in Figure 4. This analysis requires lot of simulation (yellow
193 boxes) as well as data samples (soft red boxes) in order to complete it.

194 Other samples used in this analysis, but not shown in Figure 4, are rare $B^+ \rightarrow \pi^+ \mu^- \mu^+$
195 simulation which cross-checks the shape and amount of backgrounds events coming from decay-in
196 flight events. In practice this is already incorporated in misID samples, but cross-check has been
197 performed.

198 And finally samples for normalisation channel $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+$ are required as well.
199 For these, $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+$ data samples are obtained and in order to model the signal
200 component $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+$ is used, and $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)\pi^+$ simulation models
201 $\pi \rightarrow \mu$ misID component to the invariant mass fit.

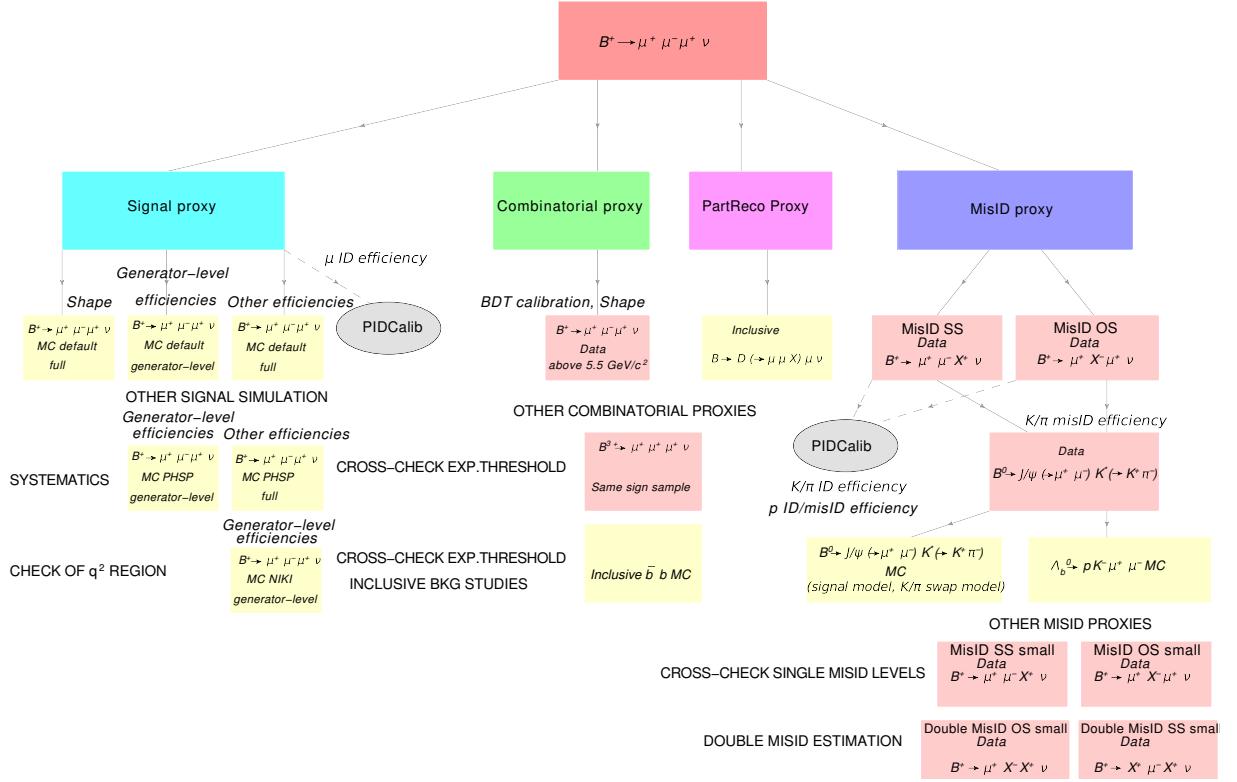


Figure 4: Flow chart of analysis. In yellow all the simulation samples related to signal $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ decay are listed. All the data samples directly analysed are in soft red boxes. (PIDCalib data samples are in grey).

202 **4 Selection**

203 The selection for this analysis was performed in the few steps:

204 • stripping and trigger requirements

205 • pre-selection including PID requirements

206 • two consecutive multivariate classifiers to remove combinatorial and misID background

207 • final PID optimisation

208 Selection for different particles and their tracks for the data signal sample $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu_\mu$
209 are summarized in Table 5.

210 **4.1 Choice of q^2 region**

211 In $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$, two pairs of oppositely-charged muon combinations can be formed, where
212 the invariant mass squared q^2 is either $q^2(\mu_1, \mu_2)$ or $q^2(\mu_2, \mu_3)$ where $\mu_1 = \mu^+, \mu_2 = \mu^-, \mu_3 = \mu^+$.
213 From the two invariant mass squared pairs the combination with lower value is, $minq^2 =$
214 $min[q^2(\mu_1, \mu_2), q^2(\mu_2, \mu_3)]$ and with higher $maxq^2 = max[q^2(\mu_1, \mu_2), q^2(\mu_2, \mu_3)]$. This measure-
215 ment is made in region where $\sqrt{minq^2} = minq < 980$ MeV/ c^2 because of two main reasons:
216 most of the contributions to the amplitude of the decay is below this value and combinatorial
217 background is greatly reduced if $minq < 980$ MeV/ c^2 , see Figure 5. The effect on all the major
218 backgrounds, such as combinatorial background (COMBI DATA) misID background (MISID
219 DATA) and partially reconstructed background (PARTRECO MC) can be compared to the effect
220 on signal simulation (SIGNAL MC), as seen in Figure 6.

221 Cutting harder on the $minq$ variable is not advisable as according to the prediction for
222 $\mathcal{B}(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu)$ [7] most of the contribution would come from ρ and ω resonances as can be
223 seen also in subsection 11.1.

224 In order to remove backgrounds that proceed via resonant J/ψ and $\Psi(2S)$ contributions,
225 vetoes in invariant mass are placed in the corresponding regions, see Table 7 for more details.

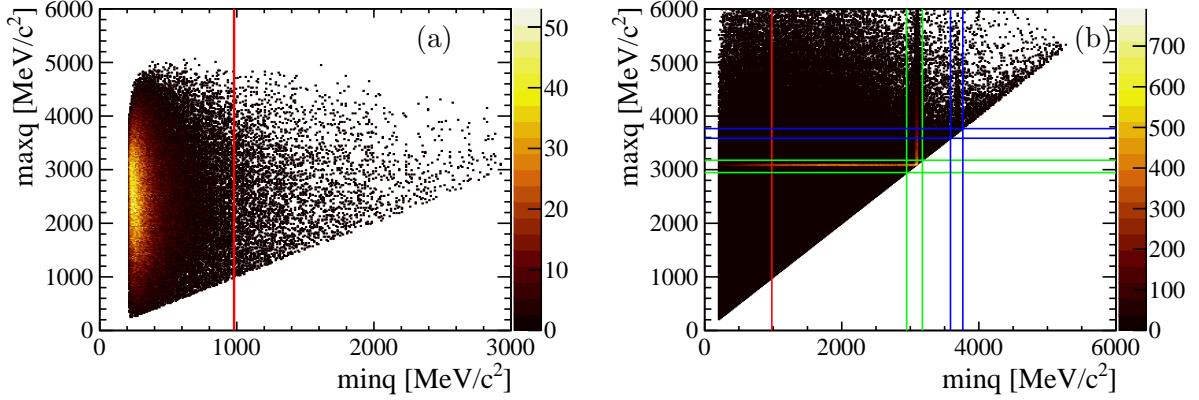


Figure 5: (a) Signal simulation sample distribution in $minq$ and $maxq$ variables. Values below 980 MeV/ c^2 (red line) are accepted. (b) Combinatorial data sample after *stripping* selection with no other cuts shows clearly the J/ψ (green) and $\Psi(2S)$ (blue) resonances which are vetoed and the measurement region (red).

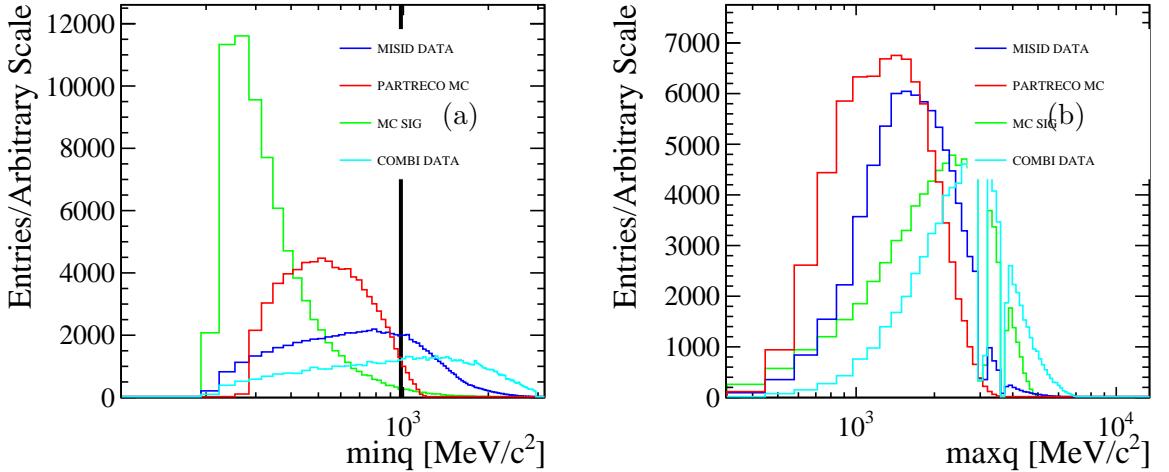


Figure 6: (a) $minq$ and (b) $maxq$ distributions for different samples in Stripping 21 for MC and Stripping 21 and 21r1 for data.

4.2 Stripping

Set of initial requirements for signal $B^+ \rightarrow \mu^+\mu^-\mu^+\nu$ summarized in Table 5, also known as *stripping* selection was developed in order to improve signal to background ratio.

Firstly, all three muon tracks are required to have a significant IP with respect to the primary vertex. Minimum Impact Parameter χ^2 , χ^2_{minIP} , gives the minimum significance of a particle's trajectory to the primary vertex. Hence by requiring $\chi^2_{minIP} > 9$ for muons is consistent with the hypothesis that the muon is 3σ away from the primary vertex and hence can be well differentiated. In addition, the change in the χ^2 if primary vertex (PV) and secondary vertex (SV) vertices are fitted separately as opposed to common vertex fit, χ^2_{FD} , suppresses prompt backgrounds.

235 Each muon track is required to have good track χ^2 per number of degrees of freedom of the
 236 fit (`ndof`), χ^2_{tr}/dof , as well as low $P_{ghost} < 0.35$. This removes spurious tracks as well as tracks
 237 with low quality.

238 Each muon candidate is also identified with initial basic PID variables. Firstly muons are
 239 chosen due to their signature in the muons stations with the binary `isMuon` decision. Secondly,
 240 muons candidates are PID using $\Delta LL(\mu - \pi)$ and $\Delta LL(\mu - K)$. ΔLLx are PID variables,
 241 which quantify the strength of the species hypothesis against pion hypothesis using especially
 242 information from RICH, tracking and calorimetry. This reduces the background from misidentified
 243 muons.

244 In order to only select events which are compatible with the three muons originating from
 245 the same point in the space, $\chi^2/dof < 4$, the χ^2 of the trimuon vertex per degree of freedom fit
 246 required to be small. This decreases the contamination from *cascade decays* where the particle
 247 with the c quark content from $b \rightarrow c \rightarrow s$, such as D , would have non-negligible lifetime leading
 248 to higher χ^2/dof .

249 Requiring that B^+ direction points in the same direction as the line from PV to SV, $(\cos(\theta_B)$
 250 - which measures the angle between these two vectors), is close to unity translates into a well
 251 reconstructed event, which minimizes combinatorial background, where random track makes this
 252 pointing worse. Putting bounds on mass window, whether it is *visible* or *corrected* mass, also
 253 suppresses combinatorial events. This is because of on average higher momentum combinatorial
 254 muon.

255 It is important to note that the stripping selection is the same for both Run 1 and 2016 data.

Muon Cuts	Mother/CombCut
$\chi^2_{minIP} > 9$	$2500 \text{ MeV}/c^2 < M_{B_{\text{corr}}} < 10000 \text{ MeV}/c^2$
$\Delta LL(\mu - \pi) > 0$	$\cos(\theta_B) > 0.999$
$\Delta LL(\mu - K) > 0$	$p_T > 2000 \text{ MeV}/c$
$P_T > 0$	$\chi^2_{FD} > 50$
$\chi^2_{tr}/dof < 3$	$\chi^2/dof < 4$
$P_{ghost} < 0.35$	$0 \text{ MeV}/c^2 < M_B < 7500 \text{ MeV}/c^2$

Table 5: Cuts applied to muons and the B^+ candidates. Stripping selection for the signal decay $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu_\mu$ is the same for both Run 1 and 2016 data. Mother/CombCut refers to vertex variables, or variables where combination of all four-vector of the decay that provides B candidates are considered.

256 4.3 Trigger Selection

257 In order to obtain triggered data, $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu_\mu$ candidates are required to pass certain set
 258 of trigger selection. No special optimisation procedure is put in place.

259 Just for illustration, in the Table 6 are fractions of triggered-on-signal (TOS) $B^+ \rightarrow (J/\psi \rightarrow$
 260 $\mu^+ \mu^-)K^+$ data with respect to $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+$ candidates. Prior to trigger require-
 261 ments, candidates have been obtained from Stripping 21r1 (2011), Stripping 21 (2012), Stripping
 262 26 (2016) via its `B2XMuMu_Line` (stripping for $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+$) and misID strip-
 263 ping, detailed in Table 16 (known `B23MuNu_TriFakeMuLine`). In `B23MuNu_TriFakeMuLine`, kaon
 264 candidate is required not to be muon (`isMuon==0`) but it is in the muon acceptance (`inMuon==1`).

265 For MC efficiencies of trigger selection, see subsection 7.4.

Trigger selection line	2011	2012	2016
Bplus_L0MuonDecision_TOS	0.915	0.895	0.74
Bplus_Hlt1TrackMuonDecision_TOS	0.874	0.929	0.931
Or of HLT2 lines below	0.986	0.987	0.996
Bplus_Hlt2TopoMu2BodyBBDTDecision_TOS	0.859	0.892	0.94*
Bplus_Hlt2TopoMu3BodyBBDTDecision_TOS	0.677	0.76	0.886*
Bplus_Hlt2DiMuonDetachedDecision_TOS	0.809	0.769	0.988
Bplus_Hlt2DiMuonDetachedHeavyDecision_TOS	0.94	0.929	0.99

Table 6: Trigger Selection for the $B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+$ normalisation data channel after B2XMuMu_Line stripping and B23MuNu_TriFakeMuLine stripping applied. The reason behind relatively high Bplus_Hlt2DiMuonDetachedHeavyDecision_TOS trigger rate is its application of the previous stripping selection (it already requires good mass vertex and mass window to be in J/ψ region). *Name different in 2016.

266 4.4 Pre-selection

267 A few additional selections are placed upon the signal after the stripping, namely muon identification cuts - such as `nShared` and `Probnn`-based variables. For `Probnn`-based variables In Run
268 1 tuning `MC12TuneV2` is used (motivated in subsection 12.2) and in 2016 tuning `MC15TuneV1` is
269 employed. This selection differs between Run 1 and 2016 as there was a redefinition of some
270 of these variables. Namely, the `nShared` variable, number of tracks with shared hits in the
271 muon stations, had a bug in the Run 1 sample which was removed in Run 2, but unfortunately
272 new bugs were introduced in Run 2, yielding worse signal purity. More about this topic is
273 described in subsubsection 4.4.1. In 2016, `isMuonTight` variable, which is the same as `isMuon`
274 but using crossed hits with information from x, y pads simultaneously, is exploited. Full further
275 pre-selection is given in Table 7.

Idea	Object	Selection Run 1	Selection 2016
J/ ψ veto	Dimuon	$!(2946.0 \text{ MeV}/c^2 < M(\mu^+ \mu^-) < 3176.0 \text{ MeV}/c^2)$	$!(2946.0 \text{ MeV}/c^2 < M(\mu^+ \mu^-) < 3176.0 \text{ MeV}/c^2)$
$\Psi(2S)$ veto	Dimuon	$!(3586.0 \text{ MeV}/c^2 < M(\mu^+ \mu^-) < 3766.0 \text{ MeV}/c^2)$	$!(3586.0 \text{ MeV}/c^2 < M(\mu^+ \mu^-) < 3766.0 \text{ MeV}/c^2)$
Clean	Muon	-	<code>IsMuonTight==1.0</code>
Clone and ghost	Muon	<code>Nshared==0</code>	<code>Nshared<2</code>
Bkg Removal	event	Combinatorial BDT selection	Combinatorial BDT selection
Bkg Removal	event	Misid BDT selection	Misid BDT selection
Fit Region	B	$4000 \text{ MeV}/c^2 < M_{B_{\text{corr}}} < 7000 \text{ MeV}/c^2$	$4000 \text{ MeV}/c^2 < M_{B_{\text{corr}}} < 7000 \text{ MeV}/c^2$
Optimize FOM	Muon	<code>Probnnmu>0.35</code>	<code>Probnnmu>0.35</code>

Table 7: Offline selection performed after stripping. Differences can be seen between Run 1 and 2016 datasets. Here $M(\mu^+ \mu^-)$ refers to any combination of oppositely charged muons. In practise, as the $ming$ is below $980 \text{ MeV}/c^2$, it is a requirement on $maxq$.

277 This different Selection 2016 from Table 7 was optimised to regain the best signal purity
278 possible for 2016. Different strategies for optimisation are discussed in Table 8. To see the drop
279 in signal purity(FOM), **Selection Run 1** is compared to **Selection 2016 - Same as Run 1**
280 where the efficiency of applying `nShared==0` drops from around 80% to around 50% for signal,

281 whereas the background drop is not as big. More specifically, pre-selection in 2016 was optimized
282 to match the closest $FOM = \frac{\epsilon_s}{\sqrt{(B)+3/2}}$ [10] to 2013, where ϵ_s is the signal efficiency (denoted as
283 SIG MC in table) and B is the number of background events, estimated from upper mass side
284 band (denoted as COMBI DATA in the table) and misID sample (denoted as MISID DATA (SS)
285 in the table). In the Selection Run 1 table, the FOM is calculated. As it can be seen from all 2016
286 strategies the **Alternative Selection 2016 - 1** was chosen due to the best (closest to 2012)
287 FOM.

Selection Run 1					
Sample	SIG MC 2012	COMBI DATA 2012	MISID (SS) DATA 2012	NORM MC 2012	NORM DATA 2012
nShared==0	0.805	0.570	0.828	0.944	0.945
$\min q(\mu^+, \mu^-) < 980 \text{ MeV}/c^2$	0.958	0.291	0.576	1	1
total	0.772	0.166	0.477	0.944	0.945
FOM = 0.00134					
Selection 2016 - Same as Run 1					
Sample	SIG MC 2016	COMBI DATA 2016	MISID (SS) DATA 2016	NORM MC 2016	NORM DATA 2016
nShared==0	0.530	0.399	0.628	0.705	0.711
$\min q(\mu^+, \mu^-) < 980 \text{ MeV}/c^2$	0.961	0.321	0.606	1	1
total	0.509	0.128	0.381	0.705	0.711
FOM = 0.00101					
Alternative Selection 2016 - 1					
Sample	SIG MC 2016	COMBI DATA 2016	MISID (SS) DATA 2016	NORM MC 2016	NORM DATA 2016
IsMuonTight==1.0	0.960	0.859	0.934	0.976	0.964
nShared<2	0.890	0.766	0.880	0.897	0.910
$\min q(\mu^+, \mu^-) < 980 \text{ MeV}/c^2$	0.960	0.308	0.602	1	1
total	0.745	0.203	0.495	0.876	0.877
FOM = 0.00129					
Alternative Selection 2016 - 2					
Sample	SIG MC 2016	COMBI DATA 2016	MISID (SS) DATA 2016	NORM MC 2016	NORM DATA 2016
IsMuonTight==1.0	0.960	0.860	0.934	0.976	0.964
$+:-\text{nShared}<2 +:\text{nShared}==0$	0.694	0.617	0.880	0.897	0.910
$\min q(\mu^+, \mu^-) < 980 \text{ MeV}/c^2$	0.960	0.315	0.602	1	1
total	0.639	0.167	0.495	0.876	0.877
FOM = 0.00111					
Alternative Selection 2016 - 3					
Sample	SIG MC 2016	COMBI DATA 2016	MISID (SS) DATA 2016	NORM MC 2016	NORM DATA 2016
IsMuonTight==1.0	0.960	0.860	0.934	0.976	0.964
$+:+\text{nShared}<2 -:\text{nShared}==0$	0.695	0.655	0.741	0.791	0.800
$\min q(\mu^+, \mu^-) < 980 \text{ MeV}/c^2$	0.960	0.306	0.597	1	1
total	0.641	0.172	0.413	0.772	0.771
FOM = 0.00121					

Table 8: This table lists pre-selection efficiencies for Run 1 and different scenarios in 2016. Pre-selection in 2016 was optimized to match the closest $FOM = \frac{\epsilon_s}{\sqrt{(B)+3/2}}$ [10] to 2012, where ϵ_s is the signal efficiency (SIG MC) and B is the number of background events, estimated from upper mass side band (COMBI DATA) and misID sample (MISID DATA (SS)). **Alternative Selection 2016 - 1** was chosen. Big signal efficiency penalty for choosing nShared==0 for 2016 can be also seen in normalisation channel (NORM MC, NORM DATA), where there are two muons that are affected. In **Alternative Selection 2016 - 2** and **Alternative Selection 2016 - 3** + sign, means applied on μ^+ , - sign mean applied on μ^- . For outcome of optimisation, see the main text.

288 **4.4.1 nShared variable**

289 The **nShared** variable for muons gives number of tracks with shared hits in the muon stations.
290 To identify muons for the signal sample, we require **nShared** for all 3 muons to be 0, i.e not
291 to share hits in muon stations with any other particle. This greatly reduces clones as well as
292 combinatorial background, as seen in Table 8, where for Selection Run 1 the signal efficiency
293 of this requirement is 80.5% and combinatorial efficiency is 57%. The same is true for the
294 Alternative Selection 2016 - 1, which is used, where signal efficiency is 89% and combinatorial
295 efficiency is 77%. For misID backgrounds, the effect can be seen in Figure 28, where the misID
296 probabilities of $K/\pi \rightarrow \mu$ are calculated. By comparing red and black points (without and with
297 **nShared** cuts), the misID probability is reduced by factor of at least 2 in most bins of momentum.
298 In this section, discussion on the differences between Run 1 and 2016 are discussed. In analysing
299 Run 1 and 2016 data, there were features that changed between them.

300 The first feature that is different comes from the definition of **nShared**. Muon identification
301 works on extrapolation of tracks into the muon system. If there is a hit saved in one of the muon
302 stations, and this hit belongs to several tracks' field of interest(FoI), then it is decided which
303 track owns the hit - owner track. To assign the hit to the track, the extrapolation of the tracks
304 to the muon system is performed and the track with the smallest distance to the hit will own
305 the hit. Other tracks with FoI covering this hit, but not being owner of the hit, increase by one
306 unit. In Stripping 21 and 21r1, it was discovered that the distance between extrapolated track
307 and hit was wrongly calculated, and hence this was corrected before Stripping 23. Additionally,
308 information from M1 station was used to calculate distances, even though it is not usually used
309 for muon id algorithm. For analysts, this feature was present across all reconstruction software
310 and hence it is consistent within stripping version.

311 In Stripping 23, the MuonID algorithm, which gives information about PID variables related
312 to muon candidates, was rewritten to adapt for parallelization that needs to be done in order to
313 meet criteria for upgrade. There were two mistakes introduced prior to 2015 datataking.

314 Firstly, defining array with 4 elements [0, 3] to store information about x and y coordinates
315 of the hits, but filling arrays with 5 elements (M2-M5 muon stations)(Iterating from [1, 4]). It
316 has been proved to have no impact on physics.

317 Further in the process, however, this information is used to calculate the sum and average of
318 distances per station between the hits and extrapolations, this again iterates over [0, 3] arrays,
319 meaning that no information is used from M5 muon station. This obviously has effect, but again
320 it is consistent across the reconstruction version.

321 Summary of these features can be found in [this presentation](#)². The interplay between all these
322 features for normalisation channel can be seen in Figure 7. In the plots, **nShared** distribution for
323 one of the muons from $B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+$ is plotted for different stripping versions. Run 1
324 (Stripping 21 and Stripping 21r1) behaves in the same way, however, as it can be seen there is a
325 shift of distribution in Stripping 26 compared to 21/21r1, which indicates that in 2016 the muon
326 tracks are less isolated, as bigger proportion of the tracks have **nShared** value greater than 0.

327 One can also asses the impact on the PID variables, using **PIDCalib** package, [twiki reference](#)
328 [here](#), which houses high-yield, relatively background-free data samples of decays, where the
329 species of the particle is determined purely from kinematics of the decay. Performance of
330 **nShared** variable (its probability of correct identification for muon $\mu \rightarrow \mu$ ID and as compared

²Presentation about **nShared** by R. Gomez, can be accessed at
https://indico.cern.ch/event/612764/contributions/2567244/attachments/1449649/2234804/20170426_nShared.pdf

331 to misidentification probability $\pi \rightarrow \mu$ misID) can be seen in Figure 8. Left plot compares
 332 2015 versus 2012 data (this plot is just for bookkeeping, 2015 data are not used anywhere),
 333 whereas right plot compares 2016 and 2012 performance. In black, the performance between
 334 2012 and 2016 for the lowest kinematical bin of p and η is compared. In the red, the second
 335 lowest kinematical bin is compared. The most important result is in the right plot, where two
 336 red points should be compared. It can be noted that for the same *misID* efficiency, the ID rate
 337 is much higher in 2012 than in 2016.

338 In order to keep the signal efficiency high in 2016 data, as shown in Table 8, softer condition,
 339 `nShared < 2`, is applied for 2016 as compared to Run 1, where `nShared == 0`. The reason
 340 for imposing the softer condition is that, in this analysis, already small statistics for signal is
 341 expected, hence there would not be much signal left if harsher condition was imposed. This is
 342 also partly due to the fact that this condition is applied on all the muons (multiply the effect \times
 343 3), which is unaffordable.

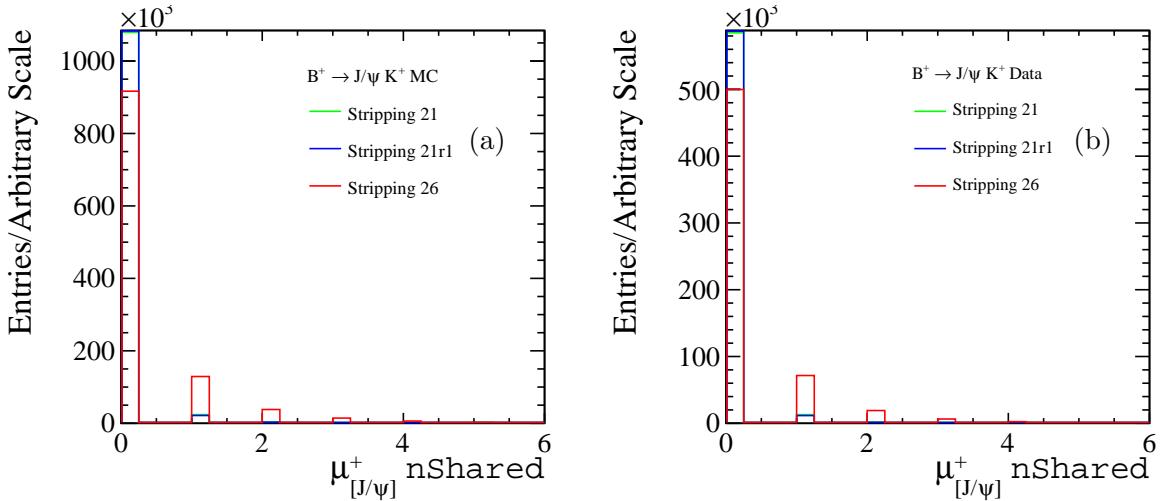


Figure 7: (a) The `nShared` variable in MC, (b) data for normalisation channel, in different stripping versions corresponding to 2012 (Stripping 21), 2011 (Stripping 21r1), 2016 (Stripping 26) data-taking. There is shift of distribution in Stripping 26 compared to the two others, which shows that in 2016 the muon tracks are less isolated.

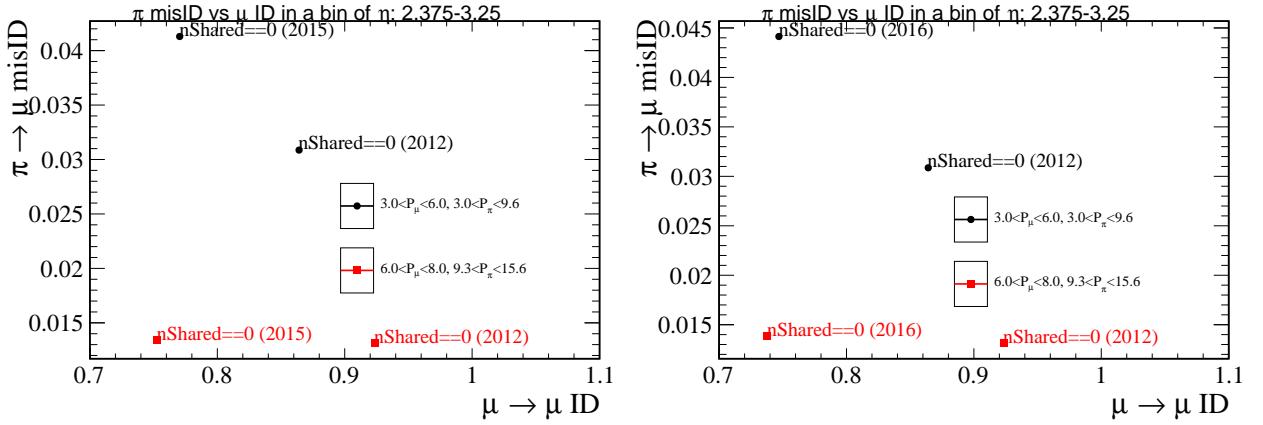


Figure 8: These plots use **PIDCalib** datasets from 2012 (Stripping 21), 2015 (Stripping 23) and 2016 (Stripping 26), binned using default 2-dimensional binning scheme in P and η . Here values chosen are in central bin of η and first and second bin in P . This shows that for same ID efficiency, the misID rate is significantly higher both in 2015 and 2016 efficiency.

344 4.5 Multivariate Selection

345 In this analysis, the background study of inclusive $b\bar{b}$ simulation sample, see subsubsection 3.2.5
 346 for definition and subsection 5.1 for background study, shows that there will be two dominating
 347 backgrounds, combinatorial background and misID background. In order to reduce these
 348 backgrounds, two consecutive multivariate classifiers are used. The first multivariate classifier is
 349 developed to remove efficiently combinatorial background and a second multivariate classifier will
 350 help to control the contamination from misID decays. In order to be able to use all the samples
 351 (all the signal MC and background samples) while reducing bias due to the use of samples
 352 (bias that comes usually from only training on half of dataset), each multivariate classifier uses
 353 *k-folding* technique.

- 354 • Both background and signal sample are randomly split into k similar size sets.
- 355 • Then, the multivariate classifier is trained on the $k - 1$ sets.
- 356 • This is then tested and applied on the remaining set.
- 357 • This process is repeated k -times for all possible samples.
- 358 • Finally, the multivariate classifier is applied on the whole sample.
- 359 • For both combinatorial and misID classifiers, $k=10$.

360 Both of the classifiers use TMVA's AdaBoost algorithm and same set of variables:

- 361 • $B^+ p$
- 362 • $B^+ p_T$
- 363 • p_T of all three muons

- B^+ lifetime
- B^+ direction angle $\cos(\theta_B)$ - the cosine of the angle between the momentum vector of the B^+ meson and the direction of the flight of the B^+ meson from its associated primary vertex to its secondary vertex
- μ minimum impact parameter χ^2 - the minimum significance of a particleless trajectory to the primary vertex for all three muons
- Isolation variable, see subsubsection 4.5.1
- B^+ flight distance χ^2
- B^+ vertex χ^2

The distributions for these variables for both Combinatorial BDT (Run 1 and 2016) can be seen in Appendix A, and for Misid BDT (Run 1 and 2016) in Appendix B.

The reason for two separate BDTs in sequence is that purely combinatorial and purely misID background are different in nature, i.e the variables that are ranked the highest in discriminating power, differ. Using common BDT could lead to loss of discrimination power and also maybe to obscure shaping of the mass spectra as it would be convolution of different shapes. Moreover, for each background the number of background events needs to be estimated and hence keeping the BDTs separate makes it more simple to produce the estimate.

4.5.1 Isolation

It is possible to use multivariate analysis (MVA) techniques to see how ‘isolated’ the three μ vertex(signal) and two μ and fake μ vertex (misidentified background) is. In an event, for both signal and misidentified background there is a possibility that neutral particles, charged tracks or even a combination of both cases had not been reconstructed in the given vertex.

The isolation quality of the vertex is determined using a MVA BDT method, the AdaBoost algorithm, which is employed to classify the isolation given a set of different track and vertex variables. The ‘isolation variables’ obtained for this analysis are based on training output and weights obtained in $\Lambda_b^0 \rightarrow p\mu^-\bar{\nu}$ analysis³ where different variables, such as track p_T , the angle between two 3-momenta of the track and 3-momenta of the signal/background vertex Δ , the track χ^2 , the ghost probability P_{ghost} of the track, IP χ^2 of the track with respect to the signal/background vertex, were used. It is expected that the topology of the $\Lambda_b^0 \rightarrow p\mu^-\bar{\nu}$ and backgrounds are indeed very similar to $B^+ \rightarrow \mu^+\mu^-\mu^+\nu_\mu$ with respect to the isolation hence weights obtained from the previous $\Lambda_b^0 \rightarrow p\mu^-\bar{\nu}$ analysis are used. The only difference is addition of a charged track. In a nutshell, the algorithm looks for other than ”signal tracks (3 muon tracks)” in signal and background samples in the event and compares the vertex information for signal and background in conjunction with the track properties of these other tracks in the event. In summary, the algorithm looks for track density and compatibility with the ”signal tracks”⁴ Adding one more charged track hence should not make difference of the behaviour of

³the analysis note can be accessed at <https://cds.cern.ch/record/1694339/files/LHCb-ANA-2014-048.pdf>. This analysis resulted in a paper [11].

⁴For $\Lambda_b^0 \rightarrow p\mu^-\bar{\nu}$ analysis the isolation variable looks at the vertex of $p\mu^-$ for both signal and background. For signal the density of tracks around the $p\mu^-$ tracks is smaller then for the background where there are other tracks X , i.e the background sample is $\Lambda_b^0 \rightarrow Xp\mu^-\bar{\nu}$. The same idea applies to $B^+ \rightarrow \mu^+\mu^-\mu^+\nu$. For signal the density of tracks around the $\mu^+\mu^-\mu^=$ vertex is also smaller compared to the background case.

400 the algorithm. But of course, each MVA is optimised given the training so what can happen
 401 here is that this isolation variable performs suboptimally. This kind of problem, however, is also
 402 faced in other parts of LHCb software, Probnn algorithm, trigger algorithm and so on.

403 As a part of the crosscheck, isolation distribution for data and simulation of the normalisation
 404 channel $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+$, another 3-track channel is done. As expected, the signal MC
 405 is slightly better isolated but as $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+$ is quite clean channel the distributions
 406 are very close as seen in Figure 10.

407 The BDT response peaks between -1 and 0 for isolated tracks and between 0 and 1 for
 408 non-isolated tracks. The output of the BDT can be seen in Figure 9. The isolation variable is
 409 the most discriminating variable in the combinatorial BDT and ranks high for misID BDT as it
 410 will be shown in following two sections.

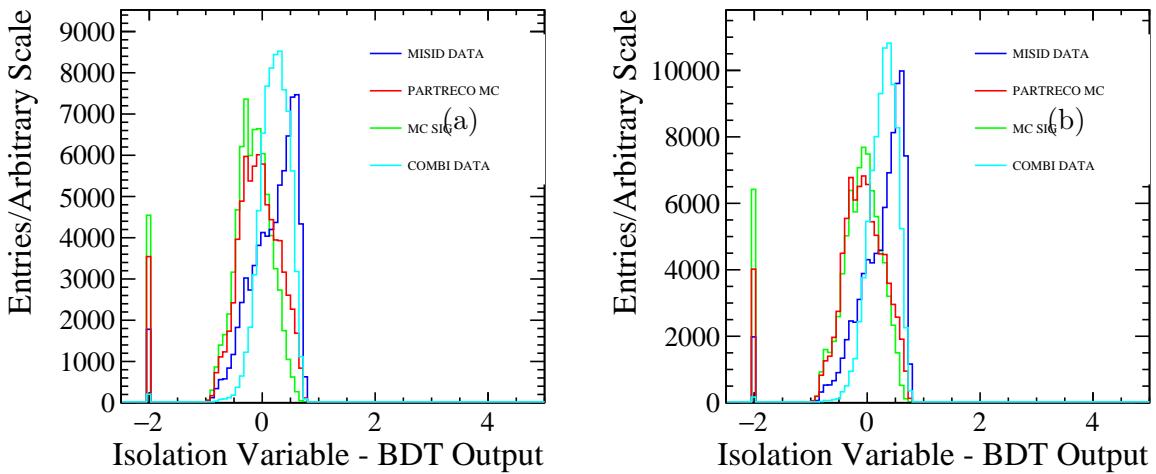


Figure 9: Isolation for both (a) Run 1 and (b) 2016 selections after all previous pre-selection. If isolation fails to find track, by default it gives value -2. The reason why partially reconstructed background is not discriminated with respect to signal is that simulation with neutral kaon and pion is used.

411 4.5.2 Combinatorial BDT

412 To obtain combinatorial BDT discriminant, a simulated sample is used for signal and upper
 413 mass sideband data ($B_{corr m}^+ > 5500 \text{ MeV}/c^2$) for background. These samples have passed through
 414 all previous pre-selection and offline selection. The Punzi figure of merit [10], $\text{FOM} = \frac{\epsilon(\text{signal})}{\sqrt{B + \sigma/2}}$
 415 (for $\sigma = 3$ as default), in blinded region, $4500 \text{ MeV}/c^2 < B_{corr m}^+ < 5500 \text{ MeV}/c^2$, is used to find
 416 optimal working point. To estimate the number of background candidates in blinded region, final
 417 fit strategy described in subsection 9.3 is used to fit the data⁵, yielding around 10000 in Run 1,
 418 and 9000 combinatorial candidates in 2016. The yields are extracted from fits to data as shown

⁵Full fitting framework is necessary for estimate of background component size for BDTS. This is because exponential constant, which describes the shape/slope of the combinatorial background is left freely floating in the data fit. This is shown in the final fitting strategy in subsection 9.3. If there is some error on this slope the optimal point could change. This was investigated and optimal BDT point is not extremely sensitive to the exact number of background events, which is ofcourse true up to a factor.

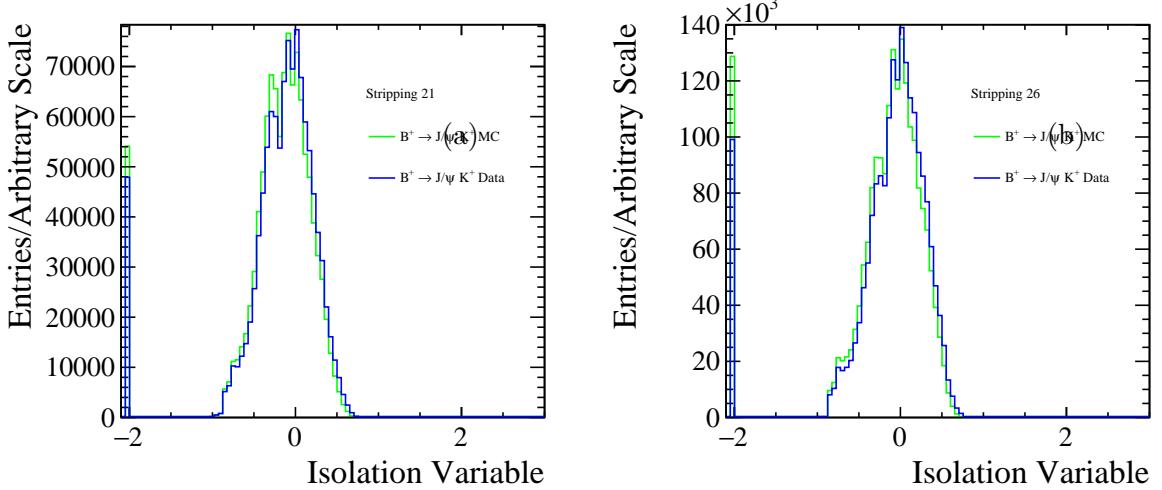


Figure 10: Isolation distribution for both (a) Run 1 and (b) 2016 normalisation channel $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+$ for signal MC and data. As $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+$ is clean channel it has very similar distribution to the signal simulation, which is just slightly shifted to the left.

419 in Figure 11. Data fitted passed through stripping, trigger and pre-selection up to Combinatorial
420 BDT (see Table 7).

421 In order to accommodate different offline selections between Run 1 and 2016, separate BDTs
422 are trained for Run 1 and 2016. Combined training of all of the datasets was also performed
423 but it does not lead to any improvement. Results of the comparison between separate and
424 combined trainings can be seen in Figure 12. Different intrinsic properties of BDT the algorithm
425 (such as number of trees, maximal depth) and also other variables (2-particle vertexes- vertexes
426 between combinations of muons) have been tested to see whether improvement in discrimination
427 is possible but the configuration giving following results proves to be most optimal and give the
428 biggest discrimination power.

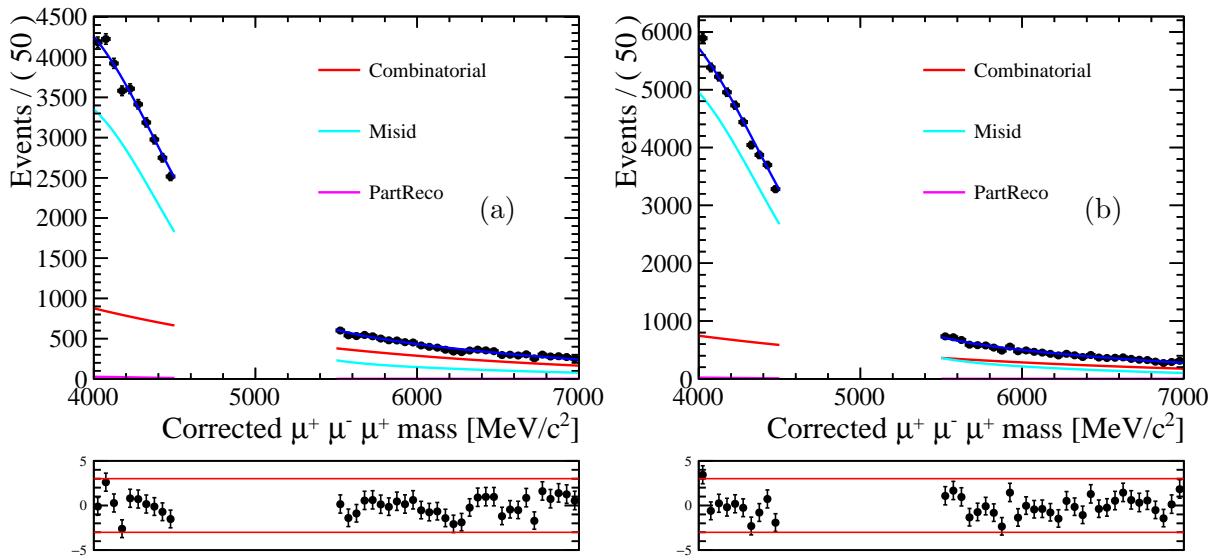


Figure 11: (a) Run 1 (b) 2016 data fit with final fitting strategy. Then number of combinatorial candidates is extracted using extrapolation of the fit into the blinded region. In these fits, data after all the selection up to combinatorial BDT (see Table 7) was applied, hence everything apart from misID BDT cut, fitting range cut and `Probnnmu` cut.

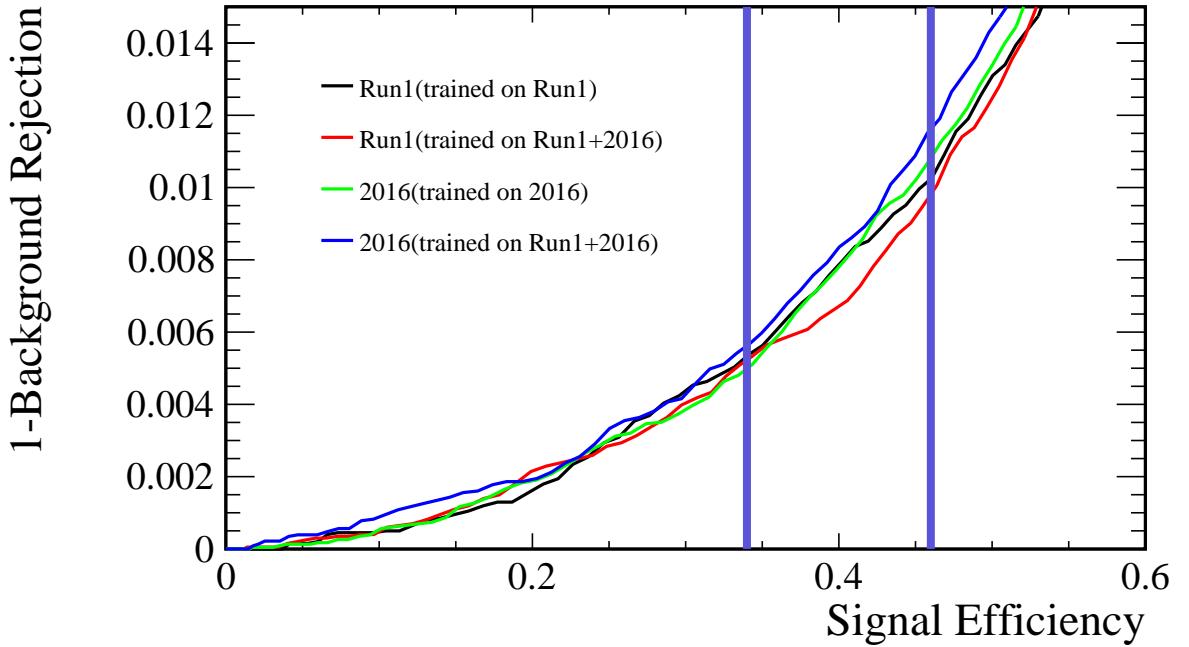


Figure 12: Comparison of separate and combined training samples and performance on different datasets. Two vertical violet lines represent optimal cut points at signal efficiency, for Run 1 (0.47) and for 2016 (0.34) where the working point of the two BDTs are chosen. Separate training provide greater rejection power in 2016. In Run 1 training on both datasets provide comparable performance for given optimal signal efficiency. Taking into the account the fact that offline selection slightly differs for 2016, it is advantageous to keep training separately.

430 The optimal working point of the combinatorial BDT for Run 1 is described in Table 9 in the
 431 Optimal cut row, together with ranking of most discriminative variables for 0th fold, Table 13.
 432 The ranking is very similar in other folds as well, with isolation variable being always the most
 433 discriminating one. The characteristics of individual BDT are seen in Figure 14 for Run 1.
 434 For 2016, the results are very similar, for more information see Appendix C. In conclusion
 435 the optimum is found at BDT output value 0.47 for Run 1 and 0.54 for 2016 Combinatorial
 436 BDT.

Property	Value
Signal sample size	76523
Background sample size	20076
Target On	1
Target signal size	89
Target background size	10134
Max Punzi FOM	0.0411402
Optimal cut	0.473333
Min cut	-1
Max cut	1
Index of cut	222
Signal Efficiency	0.460659
Background Rejection	0.989739

Table 9: Properties of Run 1 BDT.

Combinatorial BDT 2012

Rank : Variable	: Variable Importance
1 : Bplus_pmu_ISOLATION_BDT1_weights	: 1.725e-01
2 : Bplus_PT	: 9.764e-02
3 : Bplus_ENDVERTEX_CHI2	: 8.068e-02
4 : mu2_MINIPCHI2	: 7.856e-02
5 : mu3_PT	: 7.610e-02
6 : Bplus_P	: 7.367e-02
7 : Bplus_FD_CHI2	: 6.969e-02
8 : mu3_MINIPCHI2	: 6.699e-02
9 : mu1_MINIPCHI2	: 6.486e-02
10 : mu1_PT	: 5.546e-02
11 : Bplus_TAU	: 5.238e-02
12 : mu2_PT	: 5.233e-02
13 : mu2_P	: 2.456e-02
14 : mu1_P	: 2.197e-02
15 : mu3_P	: 1.261e-02
16 : Bplus_DIRA_OWNPV	: 0.000e+00

Figure 13: Ranking of the variables in 2012 data, 0-th fold.

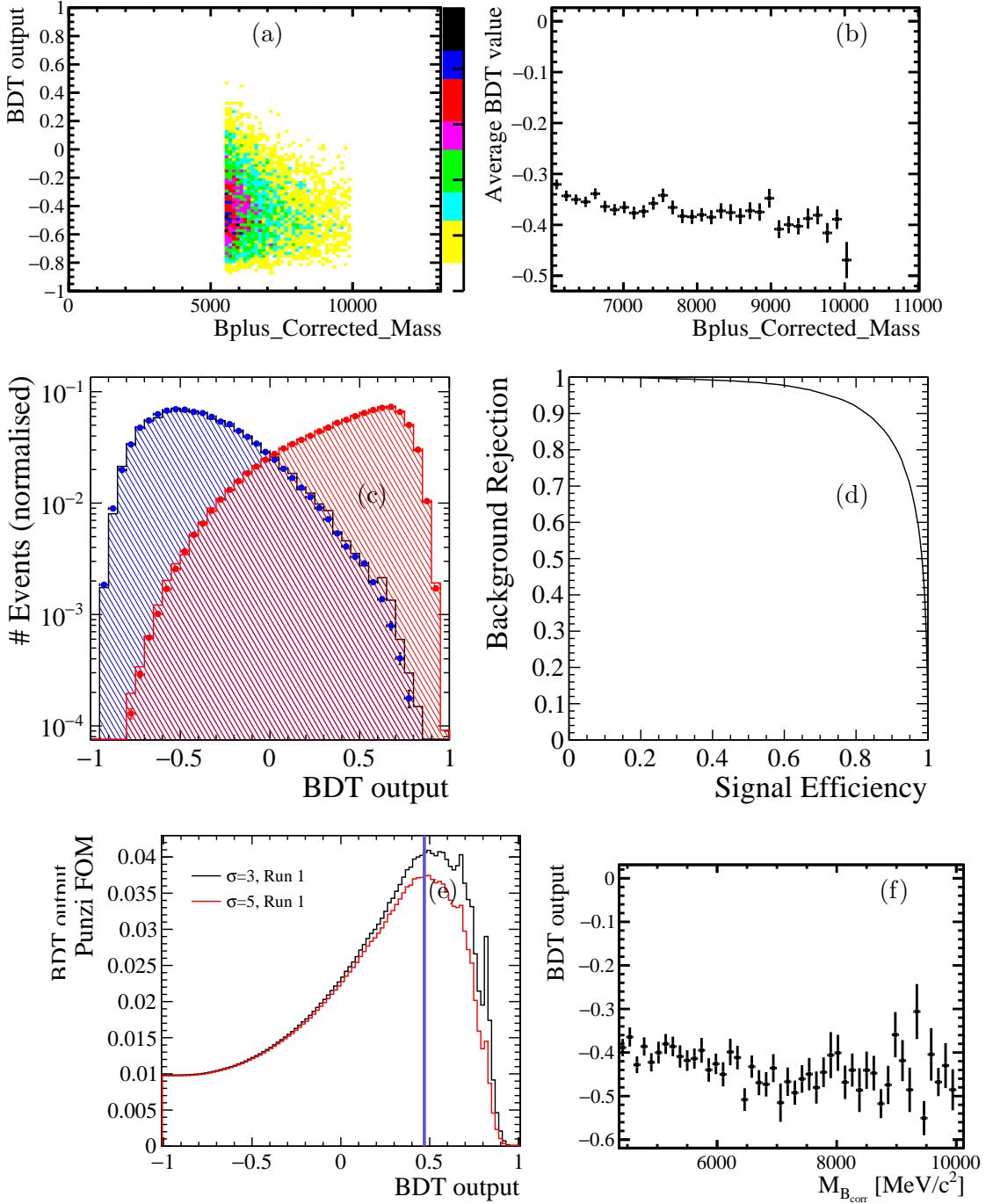


Figure 14: **Properties of the Combinatorial Run 1 BDT** (a) Correlation between B_{corr}^+ and BDT (combinatorial sample) for Run 1 Combinatorial BDT as well as (b) evidence that BDT is not correlated to the B_{corr}^+ . (c) The check for overtraining was passed yielding following (d) ROC-curve. Finally, (e) optimal working point for combinatorial BDT for Run 1 data was obtained, where $\sigma = 3, 5$ of FOM were tested both showing the same optimal point at 0.47. The spike here is due to no background events left and hence shall be ignored. (f) Evidence that BDT is not correlated to the B_{corr}^+ for same-sign combinatorial proxy. This is done to cross-check there are no peaking structures.

437 **4.5.3 MisID BDT**

438 To further reduce misID background, another set of BDTs was used. The signal training samples
 439 are the default MC samples, described in subsubsection 3.2.2, after all previous selections
 440 **including** the combinatorial BDTs. The background training samples are the misID samples
 441 also passing through all previous selections **including** combinatorial BDTs. Before application
 442 of Combinatorial BDTs, misID samples consist of mixture of misID and combinatorial misID
 443 events. The effect of combinatorial BDTs hence should reduce combinatorial misID component
 444 and this can be seen in Figures 15, 16 for Run 1 and 2016 datasets respectively.

445 As before, Run 1 and 2016 are trained and used separately. To obtain number of background
 446 events, the default fitting strategy for misID was used. The misID samples are obtained by
 447 associating the misID candidates, that passed all the previous selection with probability that
 448 takes into account crossfeed between different species. More about this procedure can be seen
 449 in subsection 5.4. The χ^2 binned fits to misID samples with Crystal Ball functions can be seen
 450 in Figure 17. For more information about fitting details, see section 9.2.1.

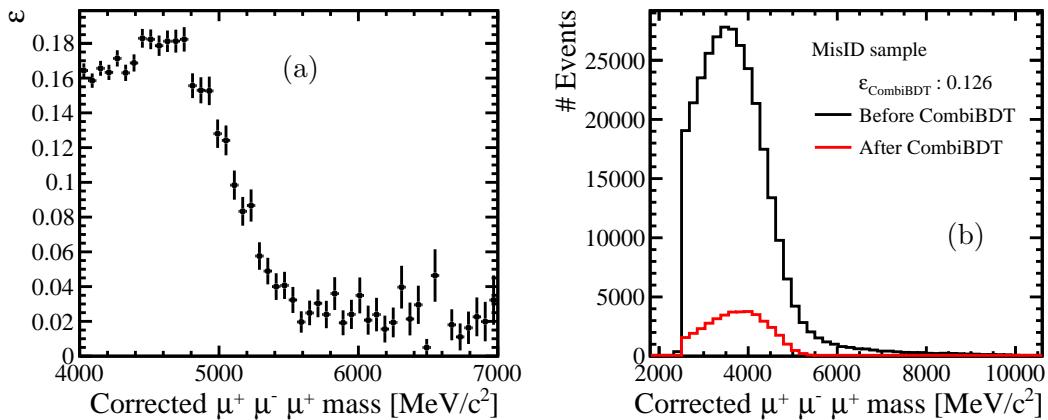


Figure 15: The effect of combinatorial BDTs on the misID samples. Here is (a) the efficiency of applying Run 1 combinatorial BDT at the optimal working point on Run 1 MisID sample. It can be seen that (b) combinatorial component of the MisID sample has been significantly reduced.

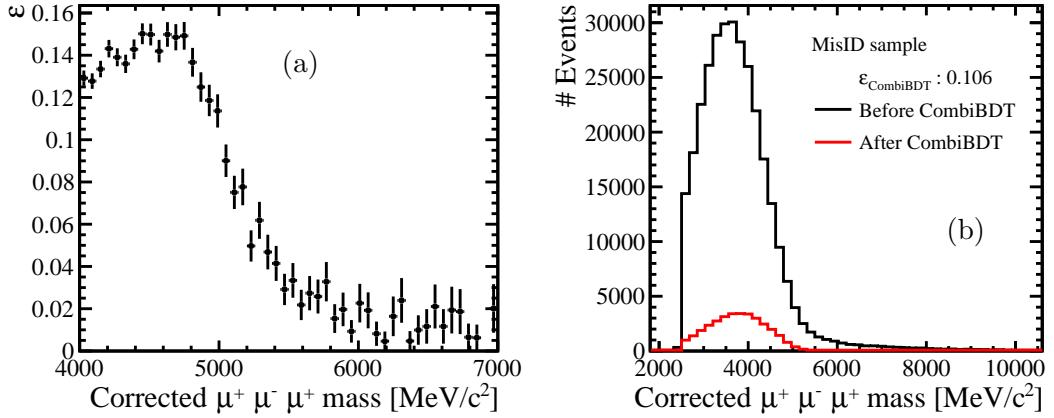


Figure 16: Similarly to Figure 15 (a) the efficiency of applying 2016 Combinatorial BDT at optimal working point on 2016 misID sample. It can be seen that (b) combinatorial component of the MisID sample has also been significantly reduced.

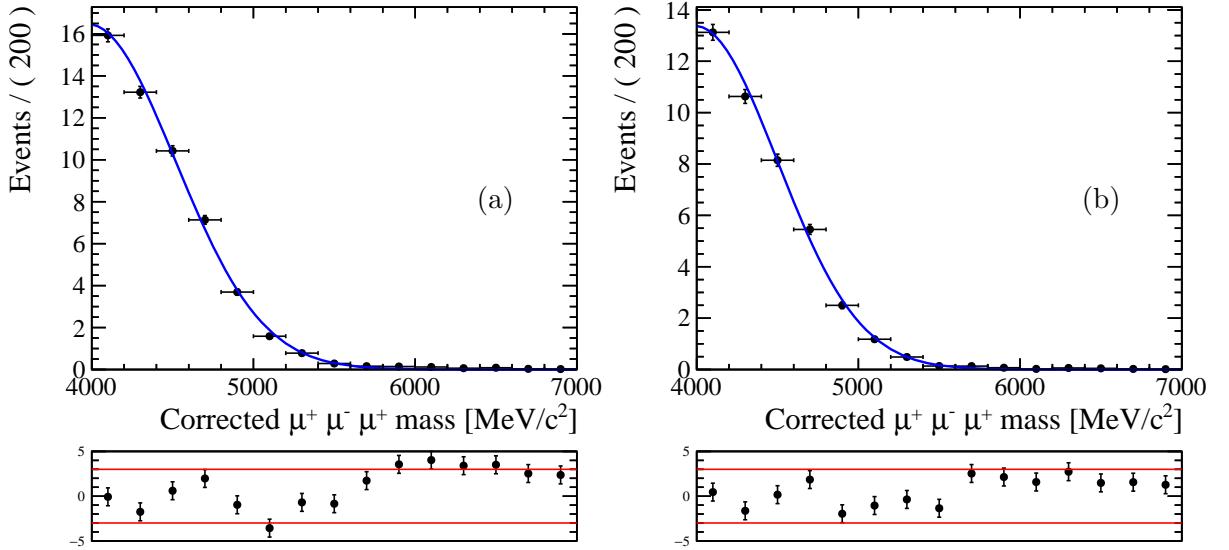


Figure 17: (a) Run 1 (b) 2016 data fit with misID fitting strategy (more information can be found in subsubsection 9.2.1).

451 The optimal working point of the misID BDT for Run 1 is described in Table 10 together with
452 ranking of most discriminative variables for 0th fold, Table 18. The characteristics of individual
453 BDT are seen in Figure 19 for Run 1.

454 For 2016, the results again are very similar. For parallel information about 2016, see Appendix
455 C. In conclusion the optimum is found at BDT output value 0.21 for Run 1 and 0.27 for 2016
456 Misid BDT.

Property	Value
Signal sample size	36160
Background sample size	39140
Target On	1
Target signal size	40
Target background size	2477
Max Punzi FOM	0.0307636
Optimal cut	0.21
Min cut	-1
Max cut	1
Index of cut	197
Signal Efficiency	0.301632
Background Rejection	0.968702

Table 10: MisID BDT summary Run 1

Misid BDT 2012

Rank : Variable	: Variable Importance
1 : mu3_PT	: 1.467e-01
2 : mu1_PT	: 8.875e-02
3 : mu3_P	: 8.326e-02
4 : mu1_MINIPCHI2	: 7.392e-02
5 : mu3_MINIPCHI2	: 7.059e-02
6 : mu1_P	: 6.966e-02
7 : Bplus_pmu_ISOLATION_BDT1_weights	: 6.949e-02
8 : mu2_PT	: 6.936e-02
9 : mu2_P	: 5.710e-02
10 : mu2_MINIPCHI2	: 5.648e-02
11 : Bplus_ENDVERTEX_CHI2	: 4.602e-02
12 : Bplus_P	: 4.572e-02
13 : Bplus_PT	: 4.465e-02
14 : Bplus_FD_CHI2	: 4.155e-02
15 : Bplus_TAU	: 3.676e-02
16 : Bplus_DIRA_OWNPV	: 0.000e+00

Figure 18: Ranking of the variables in 2012 data, 0-th fold.

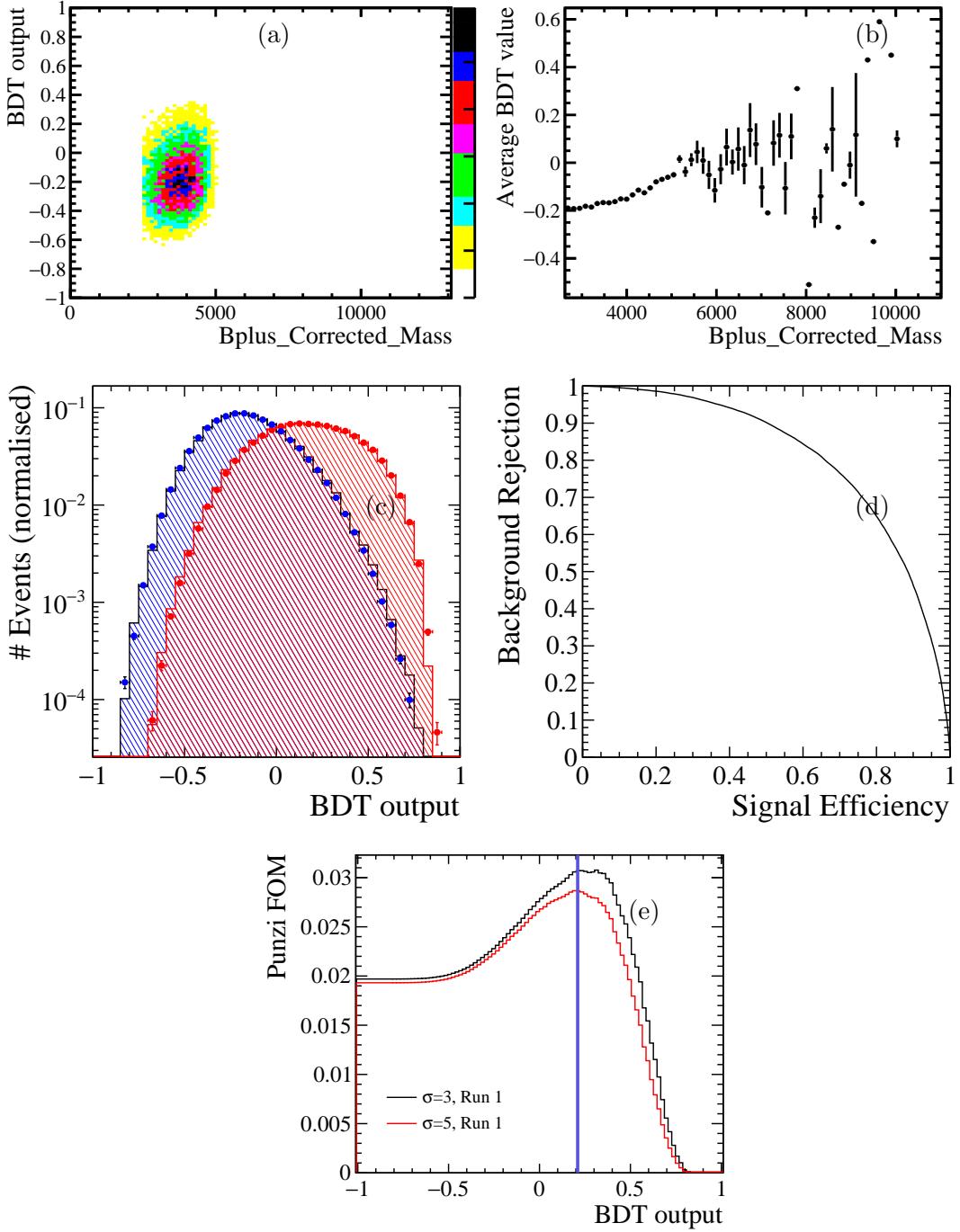


Figure 19: **Properties of Misid Run 1 BDT** (a) Correlation between B_{corr}^+ and BDT (misID sample) for the Run 1 MisID BDT as well as (b) evidence that the BDT is not correlated to the B_{corr}^+ . (c) the check for overtraining was passed yielding following (d) ROC-curve. As can be seen compared to Figure 66 the trade-off is worse however there is less events to discriminate against. Finally, (e) the optimal working point for misID BDT for Run 1 data was obtained. It can be noticed that there is plateau and value of 0.21 was used.

457 **4.6 Fitting Region Selection**

458 The distribution of the signal MC shows that majority of the signal will be in more narrow
459 window around corrected B^+ mass. In defining corrected mass region in which the fit will be
460 done there are two considered aspects. The exponential description of combinatorial background
461 is not correct below 4000 MeV/c² as will be discussed in bigger detail in subsection 5.2, meaning
462 that the fit should only start where exponential behaviour is correct. In the upper mass sideband,
463 there is little signal but also very little background expected and hence there is no point in
464 including events above 7000 MeV/c² as it would not bring any information to the fit, rather it
465 would give these events higher weight when the fit is done. Hence the fitting region selection will
466 be 4000 MeV/c² < $M_{B_{\text{corr}}}$ < 7000 MeV/c².

467 Moreover, there is blinding procedure in place for this search. The signal region 4500 MeV/c² <
468 $M_{B_{\text{corr}}}$ < 5500 MeV/c² is blinded. Hence in all blinded fits only data in regions 4000 MeV/c² <
469 $M_{B_{\text{corr}}}$ < 4500 MeV/c² and 5500 MeV/c² < $M_{B_{\text{corr}}}$ < 7000 MeV/c² are used.

470 **4.7 Probnn Optimization**

471 The current stripping has relatively loose DLL cuts - PID cuts based on log-likelihood, see Table 5
472 ($\Delta LL(\mu - \pi) = DLL_{\mu\pi}$, $\Delta LL(\mu - K) = DLL_{\mu K}$ are examples) - hence it is possible to
473 improve the performance/selection by using cuts on additional PID variables. In the optimisation,
474 different hypotheses were tested, such as cuts on `Probnnmu`, `Probnnpi`, and `ProbnnK` variables
475 and their combinations. Optimisation was performed in a following way:

- 476 • For MisID sample:

- 477 – `PIDCalib` was run with different `Probnn` cuts to get misID efficiency tables with the
478 new hypothesis.
479 – Using the misID algorithm as will be seen in subsection 5.4, new probability of misID
480 was calculated (weight) with the given `Probnn` cuts.
481 – Hence, the weight for misID was altered and total misID was recalculated using this
482 weight.

- 483 • For MC signal sample:

- 484 – To estimate the change in signal MC, the ID efficiency obtained from `PIDCalib` -ID
485 branch - was altered with the new PID hypothesis and added to MC.
486 – Then using these ID efficiencies, the number of signal events with given cuts were
487 recalculated.

- 488 • For combinatorial sample:

- 489 – The samples were cut requiring only of the muons to have relevant `Probnn` cuts.

490 Partially reconstructed background component is small so no weighting of additional PID
491 is performed.

492 The modified samples in this PID optimisation are highlighted in red boxes in Figure 20. In
493 order to select the optimal cut final fit was performed separately to Run 1 and 2016 data. In
494 2016, misID bins above 5800 MeV/c² are not populated. Hence final 6 bins of corrected mass

495 (5800 MeV/c²-7000 MeV/c²) are merged into 1⁶. For each PID hypothesis, FOM is calculated. In
496 both cases, in Run 1 and 2016 Probnnmu > 0.35 yielded highest FOM.

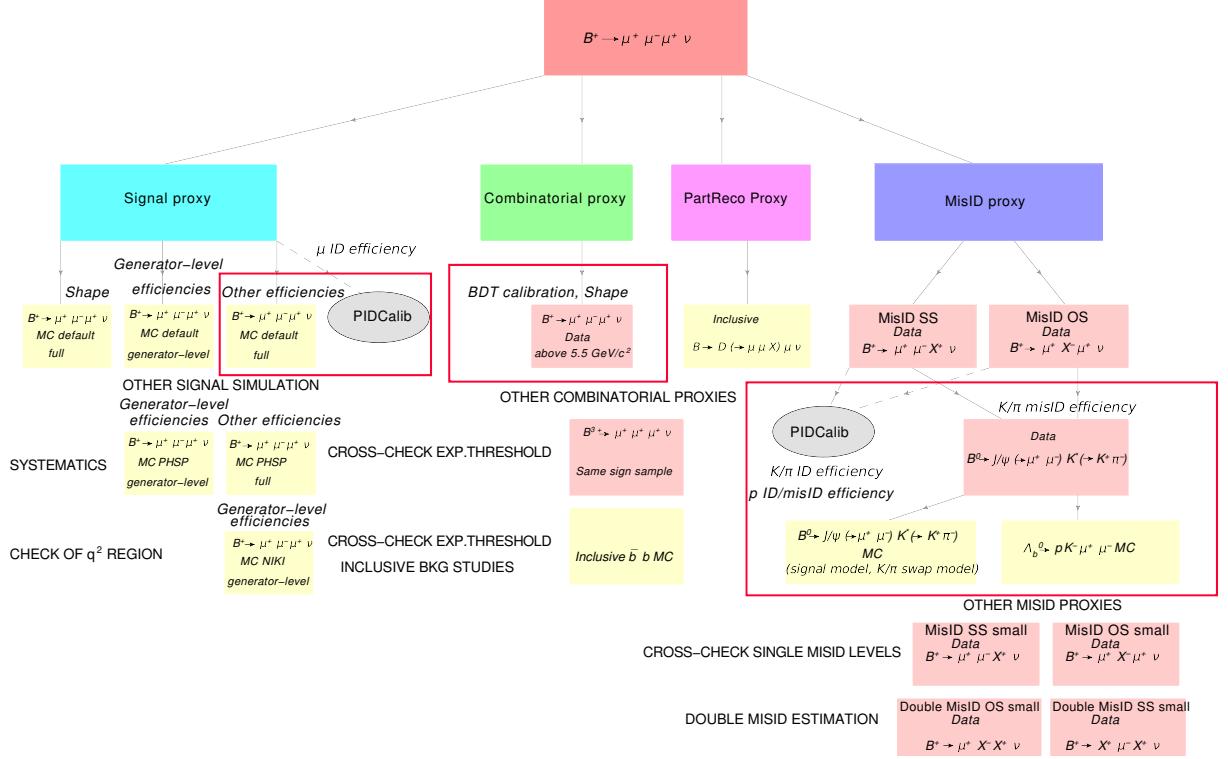


Figure 20: In this flowchart, the samples that are used and modified for PID optimisation are highlighted in red boxes.

497 4.8 Fractional Corrected Mass Window Split

498 In order to increase sensitivity, but not to decrease statistics as this selection leads to low-statistics
499 regime, it was decided that the fitting procedure for the final fit will be in two bins of fractional
500 corrected mass error (FCME), defined as

$$\sigma_{\{\text{lowFCME}, \text{highFCME}\}} = \frac{\delta^{\{\text{lowFCME}, \text{highFCME}\}}}{M_{\text{Bcorr}}^{\{\text{lowFCME}, \text{highFCME}\}}}, \quad (7)$$

501 where δ is the corrected mass error. δ arise as propagated error, whereby the error includes the
502 error on visible 4-momentum, primary and secondary vertexes (from missing p_T), where the
503 last two dominate⁷. As corrected mass error does have a dependence on resolution (see Figure

⁶The last six bins behave in a very similar way, hence this approach does not change the pdf parameters significantly. In all misID fits binned fit is performed reducing the last six bins to one bin only, but when plotting is done, the fit result from reduced binned fit is overlaid with non-reduced binned dataset. Hence, the way misID behaves in the last six bins is always visible in all the misid plot and can be compared within the last six bins.

⁷the analysis note describing δ calculation in more detail on page 8 can be accessed at <https://cds.cern.ch/record/1694339/files/LHCb-ANA-2014-048.pdf>

504 21), this split is expected to increase sensitivity. The split into two bins is done in order for
 505 $\approx 50\%$ of the signal to be in σ_{lowFCME} and $\approx 50\%$ of the signal to be in high σ_{highFCME} . This
 506 corresponds to $\delta/M_{B_{\text{corr}}} < 0.0225$ for σ_{lowFCME} and $\delta/M_{B_{\text{corr}}} > 0.0225$ for σ_{highFCME} . Exact
 507 efficiencies of this split for both signal and normalisation MC are given in Table 42. In order
 508 to look at consistency of this two bin strategy also one bin of fractional corrected mass error
 509 strategy is performed (fit across full region / no split), σ_{NOFCME} .

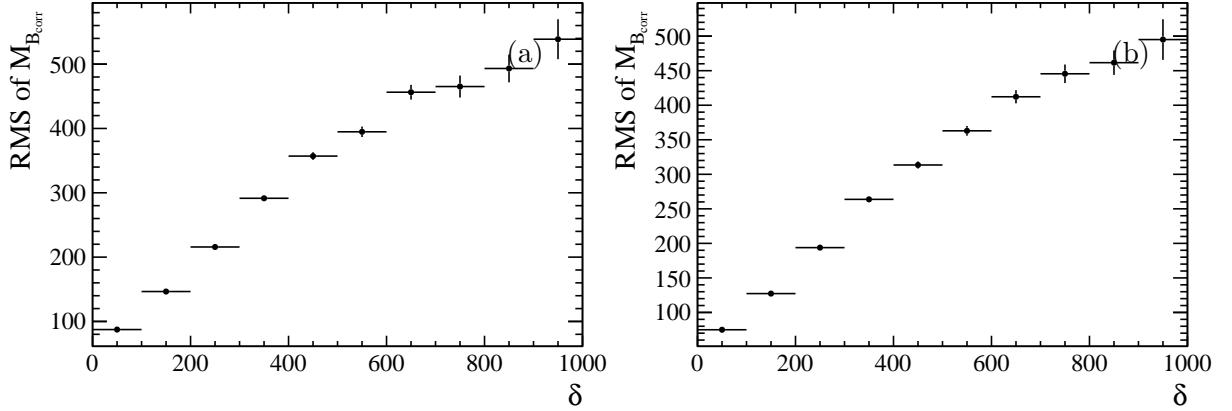


Figure 21: (a) The average resolution in signal MC Stripping 21 in bins of corrected mass error δ . (b) The average resolution in signal MC Stripping 26 in bins of corrected mass error δ .

510 4.9 Multiple Candidates

511 After all the selection performed, in blinded datasets, $M_{B_{\text{corr}}} < 4500 \text{ MeV}/c^2$ and
 512 $M_{B_{\text{corr}}} > 5500 \text{ MeV}/c^2$ there is very few candidates left: 683 in Run 1 and 515 in 2016. No
 513 multiple candidates are seen.

514 **5 Background**

515 **5.1 Inclusive simulation studies**

516 The decay $B^+ \rightarrow \mu^+\mu^-\mu^+\nu_\mu$ is a fully leptonic decay and therefore has a good potential for
 517 eliminating many types of the backgrounds. In order to gauge the severity of different backgrounds,
 518 the inclusive $b\bar{b}$ simulation sample is analysed (detailed in subsubsection 3.2.5).

519 This simulation sample is passed through the stripping selection for the mother B^+ which
 520 is in Table 5. The first two muons are chosen from `StdLooseMuons` and are required to have
 521 $\chi^2_{minIP} > 9$ and a third muon is without any cuts and selected from `StdNoPIDsMuons`. After
 522 truth matching there are 6862 events left. In this sample different types of backgrounds are
 523 investigated, as seen in Table 11. It shows that the most abundant type of background with 4324
 524 events is partially reconstructed background. Combinatorial background (`BKGAT==110`), with
 525 only 655 events is small compared to the partial reconstructed physics background (`BKGAT==40`).
 526 Combinatorial background is expected to be present and have exponential shape. A fit to sample
 with `BKGAT==110` can be seen in Figure 22. Once the third muon is required to pass muon ID

Type	B^+ Background Category	Number of events
1. MisID event	30	33
2. ParReco event	40	4324
3. Low-mass background event	50	68
4. Ghost event	60	499
5. PV event but not due to pile-up event	70	1061
6. Pile-up/Different PV event	100	221
7. Combinatorial bb event	110	655

Table 11: Breakdown of different categories for the 6862 candidates. Most abundant background when two muons are positively identified is partially reconstructed background. This can be modelled with MC and hence it is easier to account for than the combinatorial background.

527 requirements, only 68 candidates are left. In this case there are only two types of background
 528 left. Here combinatorial background is higher but there is substantial contribution from low-
 529 mass-background (`BKGAT==50`), see Table 12. An example of possible low-mass-background in
 530 this case would be $B \rightarrow (D \rightarrow K\mu\nu)\pi^0\mu\nu$. This is treated as misidentified background (misID)
 531 background.
 532

B^+ Background Category	Number of events
1. 50	25
2. 110	43

Table 12: Results when all three muons are truth matched.

533 This study shows that misID background needs to be modelled in order to understand the
 534 composition and kinematic distribution of the background. MisID background contains a group
 535 of decays, cascade decays of the type $B \rightarrow ((D \rightarrow K\mu\nu)\mu\nu)$. Here, kaon is misidentified as a

536 muon. This type of background is difficult to reject as a good vertex in many case will be formed
 537 between the three muon candidates. In the following section, modelling and amount of each type
 538 of backgrounds is described.

539 It should be noted that this study was limited by statistics and also by the fact that two
 540 muons are required to have $p > 3$ GeV/c. p selection for signal muons is performed by trigger
 541 and therefore is different making exact proportion of each background contamination hard to
 542 calculate but in order to identify relevant backgrounds this study is indicative.

543 5.2 Combinatorial Background

544 The combinatorial component needs to be studied in order to estimate its yield and shape in the
 545 final fit. It occurs when two muons from one b-hadron are combined with an additional muon
 546 not originating from that b-hadron. The usual methods of obtaining this component include
 547 extrapolation of invariant upper mass sideband to the signal region. In this analysis this means
 548 an extrapolation of $M_{B_{\text{corr}}} > 5500$ MeV/c² into the signal region. Since this analysis yielded into
 549 low-statistics regime the extrapolation from upper mass sideband introduces a big uncertainty on
 550 the exponential constant and hence cannot be used. However, what can be done is assume shape
 551 of this combinatorial component and use this shape in the data fit with floating parameter. This
 552 method for estimation of the combinatorial component will be re-explained in subsubsection 9.2.3.
 553 In this rest of this Section, exponential parametrisation of this background is motivated. As it
 554 will be shown, this assumption only holds in certain region.

555 Final fitting region was chosen in such a way as to make sure that combinatorial background
 556 is exponential in the entire fitting region. Apart from the nominal upper mass sideband, two
 557 other samples that are proxies - but with low statistics - are studied to determine in which mass
 558 regions the combinatorial background can be considered exponential. These two samples are:
 559 same sign data sample μ^+, μ^+, μ^+ for Run 1 after all selection apart from the MVA selection
 560 (to have sufficient statistics) and inclusive $b\bar{b}$ simulation sample with BKGCAT==110 (discussed
 561 in subsection 5.1). This exponential component stops rising at low $M_{B_{\text{corr}}}$ in all samples. As
 562 seen in Figure 22(a)(c), the drop appears at $M_{B_{\text{corr}}} = 4000$ MeV/c². Hence the choice of fitting
 563 region 4000 MeV/c² $< M_{B_{\text{corr}}} < 7000$ MeV/c².

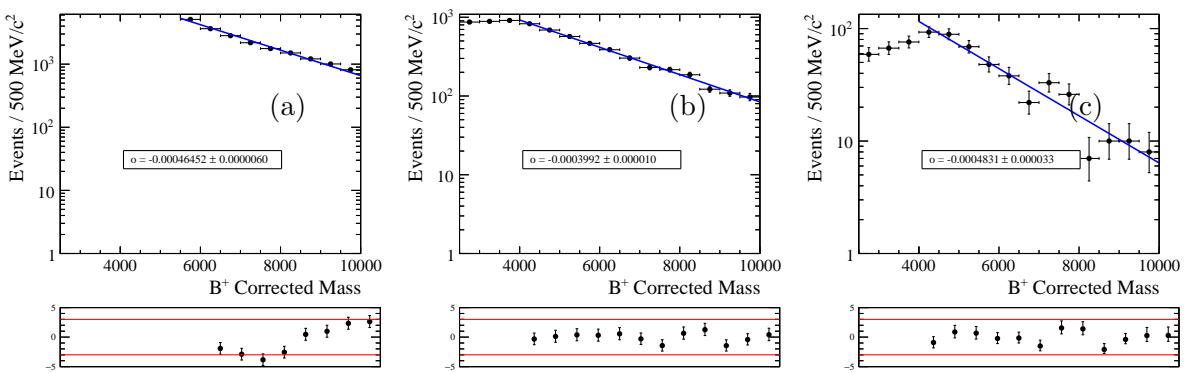


Figure 22: (a) Fit to upper mass side band just before application of combinatorial BDT. (b) Fit to μ^+, μ^+, μ^+ same sign sample. (c) Fit to BKGCAT==110 sample with exponential function from 4000 MeV/c², where the exponential rise stops.

564 **5.3 Partially Reconstructed $B^+ \rightarrow D\mu^+\nu$ type backgrounds**

565 In this section, comparison of partially reconstructed background to signal is used to demon-
566 strate that this type of background peaks at much lower corrected mass than signal. In the
567 final signal data fit, in order to obtain the expected yield of this partially reconstructed back-
568 ground, normalisation to $B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+$ is performed. More details can be found
569 in subsubsection 9.2.2.

570 In order to estimate the contamination of partially reconstructed backgrounds, one of the most
571 dangerous example is studied: $B^+ \rightarrow (D^0 \rightarrow K^-\pi^+\mu^+\mu^-)\mu\nu$. $\mathcal{B}(B^+ \rightarrow (D^0 \rightarrow K^-\pi^+\mu^+\mu^-)\mu^+\nu)$
572 is obtained by amalgamating $\mathcal{B}(D^0 \rightarrow K^-\pi^+\mu^+\mu^-) = (4.17 \pm 0.12 \pm 0.40) \times 10^{-6}$ [12] and $\mathcal{B}(B^+ \rightarrow$
573 $Dl^+\nu X) = (9.8 \pm 0.7) \times 10^{-2}$ [1] yielding $\mathcal{B}(B^+ \rightarrow (D^0 \rightarrow K^-\pi^+\mu^+\mu^-)\mu^+\nu) \approx (4.10 \pm 0.50) \times 10^{-7}$.

574 Even though there is no simulation sample exactly specifying this decay, there is an official
575 simulation that can be used as a proxy: an inclusive MC cocktail sample (containing higher
576 excited resonance of D^{*0}, D_2^{*0} , etc. see Table 13) of $B^+ \rightarrow (D^0 \rightarrow K_S^0\pi^+\pi^-\pi^0)\mu^+\nu_\mu$, where two
577 charged pions from the D^0 decay are reconstructed as signal. This sample can be found under
578 `EventType:12875540` in official production. As muons have similar mass to pions, this sample
579 will be a good approximation for contamination from $B^+ \rightarrow (D^0 \rightarrow K^-\pi^+\mu^+\mu^-)\mu\nu$. This is
580 valid approximation to make as here the impact of partially reconstructed background is being
581 studied. More specifically, what impact do missing particles have on the shape of this background.
582 Now if pion track is used instead of the muon track, the mass of the B from this decay a bit higher
583 assuming momentum spectrum is nearly the same, but nothing more. What could be a problematic
584 is, however, if the selection efficiency is not constant as a function of $M(\pi\pi)/M(\pi\mu)/M(\mu\pi)$ mass
585 then there is a danger of shaping of the background possible underestimating the contribution
586 from ω and ρ region. As it will be shown in the subsubsection 5.3.1, this is not the case.

587 For this reason all muon cuts from selection are also applied to pions, apart from the PID
588 variables which are removed in both cases. The invariant masses for possible combinations are
589 seen in Figure 23 and show expected shapes. So altogether the difference for the selection used
590 in this case as compared to signal selection is the container from which the pion candidates are
591 obtained from, as seen in Table 14. Relative efficiency ratios including all the efficiencies up to
592 BDT stage are obtained, where for signal $\epsilon_{sig}^{tot} = (1.3 \pm 0.09)\%$ and for partially reconstructed
593 background $\epsilon_{bkg}^{tot} = (0.440 \pm 0.002)\%$. After BDTs, $\epsilon_{sig}^{tot} = (1.3 \pm 0.09)\% \times (20 \pm 0.14)\%$ and for
594 partially reconstructed background $\epsilon_{bkg}^{tot} = (0.440 \pm 0.002)\% \times (15 \pm 0.15)\%$. More information
595 about exact efficiency breakdown can be obtained in Table 72. Assumed the branching fractions
596 are $\mathcal{B}(B^+ \rightarrow \mu^+\mu^-\mu^+\nu) = 1 \times 10^{-8}$ (as shown in section 1(pessimistic BF)) and $\mathcal{B}(B^+ \rightarrow (D^0 \rightarrow$
597 $K^+\pi^-\mu^+\mu^-)\mu^+\nu) = (4.10 \pm 0.50) \times 10^{-7}$ (shown just above).

Fraction	Mode	EVTGEN decay model
0.022400	$D^0 \mu^- \nu_\mu$	PHOTOS ISGW2;
0.056800	$D^{*0} \mu^- \nu_\mu$	PHOTOS ISGW2;
0.002072	$D_0^{*0} \mu^- \nu_\mu$	PHOTOS ISGW2;
0.001936	$D'_1 0 \mu^- \nu_\mu$	PHOTOS ISGW2;
0.005244	$D_1 0 \mu^- \nu_\mu$	PHOTOS ISGW2;
0.002792	$D_2 * 0 \mu^- \nu_\mu$	PHOTOS ISGW2;
0.000198	$D^0 \pi^0 \mu^- \nu_\mu$	PHOTOS GOITY_ROBERTS;
0.000969	$D^0 \pi^+ \pi^- \mu^- \nu_\mu$	PHOTOS PHSP;
0.000238	$D^0 \pi^0 \pi^0 \mu^- \nu_\mu$	PHOTOS PHSP;
0.000462	$D^{*0} \pi^0 \mu^- \nu_\mu$	PHOTOS GOITY_ROBERTS;
0.002451	$D^{*0} \pi^+ \pi^- \mu^- \nu_\mu$	PHOTOS PHSP;
0.000602	$D^{*0} \pi^0 \pi^0 \mu^- \nu_\mu$	PHOTOS PHSP;
0.000626	$D^{*+} \pi^- \mu^- \nu_\mu$	PHOTOS GOITY_ROBERTS;
0.000844	$D^{*+} \pi^0 \pi^- \mu^- \nu_\mu$	PHOTOS PHSP;
0.001337	$D^0 \tau^- \nu_\tau$	ISGW2;
0.003646	$D^{*0} \tau^- \nu_\tau$	ISGW2;
0.000246	$D_{-10} \tau^- \nu_\tau$	ISGW2;
0.000099	$D_0^{*0} \tau^- \nu_\tau$	ISGW2;
0.000168	$D'_{10} \tau^- \nu_\tau$	ISGW2;
0.000186	$D_{2*0} \tau^- \nu_\tau$	ISGW2;

Table 13: MC cocktail to look at shape and background for $B^+ \rightarrow (D^0 \rightarrow K^- \pi^+ \mu^+ \mu^-) \mu \nu$ contains contributions from all of these states. As shown in subsubsection 3.2.4, this sample can be found under `EventType:12875540`. PHOTOS, PHSP,

Muon/Pion Cuts	Mother/CombCut
$\chi^2_{minIP} > 9$ $\chi^2_{tr}/dof < 3$ $P_{ghost} < 0.35$	$2500 \text{ MeV}/c^2 < M_{B_{\text{corr}}} < 10000 \text{ MeV}/c^2$ $\cos(\theta_B) > 0.999$ $p_T > 2000 \text{ MeV}/c$ $\chi^2_{FD} > 50$ $\chi^2/dof < 4$ $0 \text{ MeV}/c^2 < M_B < 7500 \text{ MeV}/c^2$
StdLooseMuons/StdLoosePions	

Table 14: Stripping selection for $B^+ \rightarrow \pi^+ \pi^- \mu^+$ from $B^+ \rightarrow (D^0 \rightarrow ((K_s^0 \rightarrow \pi^+ \pi^-) \pi^- \pi^+ \pi^0) \mu^+ \nu$.

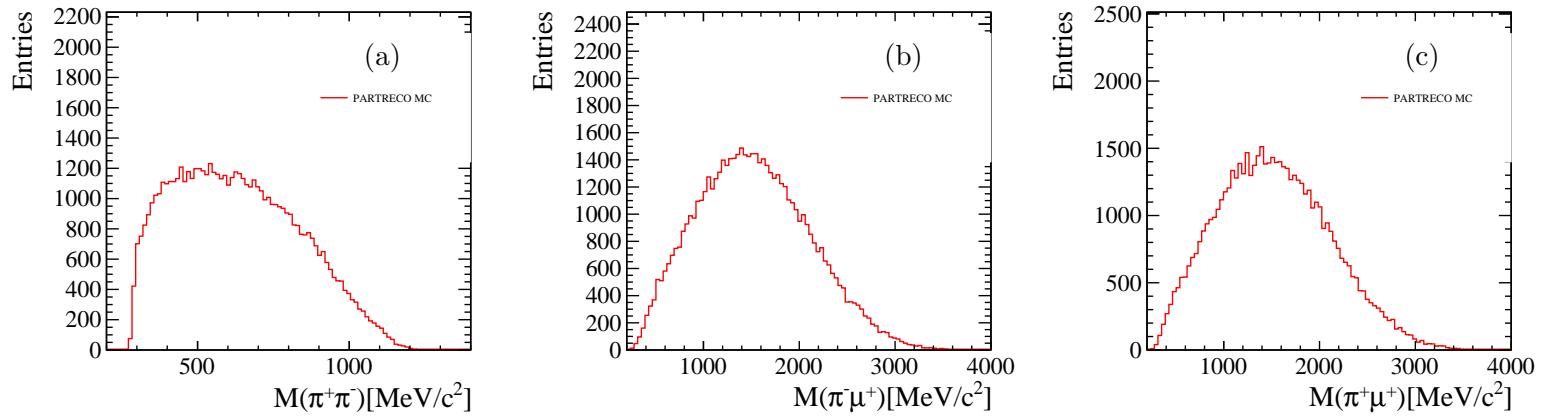


Figure 23: Invariant mass for (a) $\pi^+\pi^-$ (b) $\pi^-\mu^+$ and (c) $\pi^+\mu^+$ distributions.

This results in the following estimates.

Before BDTs, the $\frac{\mathcal{B}(B^+\rightarrow\mu^+\mu^-\mu^+\nu)\times\epsilon_{sig}^{tot}}{\mathcal{B}(B^+\rightarrow(D^0\rightarrow K^+\pi^-\mu^+\mu^-)\mu^+\nu)\times\epsilon_{bkg}^{tot}} = \frac{1.41\times10^{-10}}{1.83\times10^{-9}} = 0.077$.

After BDTs, the $\frac{\mathcal{B}(B^+\rightarrow\mu^+\mu^-\mu^+\nu)\times\epsilon_{sig}^{tot}}{\mathcal{B}(B^+\rightarrow(D^0\rightarrow K^+\pi^-\mu^+\mu^-)\mu^+\nu)\times\epsilon_{bkg}^{tot}} = \frac{2.91\times10^{-11}}{2.85\times10^{-10}} = 0.102$.

The relative expected shape and amount contamination with respect to the signal channel can be seen in Figure 24. As it can be seen, the partially reconstructed background peaks at much lower corrected mass.

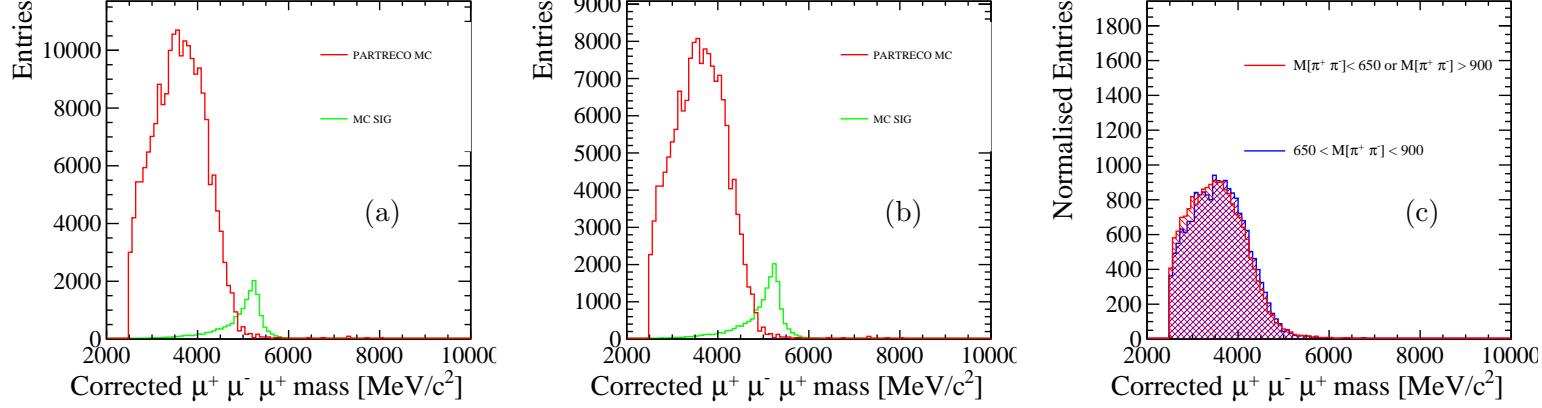


Figure 24: (a) Before BDT cut (scaled to expected ratio) and (b) after both bdt cuts distributions (scaled to expected ratio) and (c) check for correlation in corrected mass. Performing multivariate selection improved both separation as well as the relative contamination, no correlation to corrected mass has been observed.

In this section, contamination from $B^+\rightarrow(D^0\rightarrow K^+\pi^-\mu^+\mu^-)\mu^+\nu$ decays (charged case) were considered. However, partially reconstructed background could also come from neutral K and π decays via D (neutral case). Considering that in this section simulation with neutral K and π was used as a proxy for the $B^+\rightarrow(D^0\rightarrow K^+\pi^-\mu^+\mu^-)\mu^+\nu$ decays (charged case) it is expected that the isolation variable underperforms. Isolation would have been more rejective if

609 the simulation had charged K and π track instead of neutral tracks.

610 Therefore, the contribution for charged case is an overestimate of the expected number
611 of $B^+ \rightarrow (D^0 \rightarrow K^+\pi^-\mu^+\mu^-)\mu^+\nu$ decays. Therefore the contribution for the neutral case is
612 added into the consideration by counterbalancing this overestimate. This means that using
613 $\mathcal{B}(B^+ \rightarrow (D^0 \rightarrow K^+\pi^-\mu^+\mu^-)\mu^+\nu) = (4.10 \pm 0.50) \times 10^{-7}$ is a good cumulative estimate for
614 partially reconstructed background coming from D .

615 **5.3.1 Efficiency Across Phase Space**

616 As discussed previously, there are assumptions that are tied with using this particular simulation
617 to estimate the contamination of partially reconstructed background. More specifically if the
618 selection efficiency is not constant as a function of $M(\pi\pi)/M(\pi\mu)/M(\mu\pi)$ mass then there is
619 a danger of shaping of the backgrounds in ρ and ω regions, as the real background comes via
620 $K^+\pi^-\mu^+\mu^-$ decays.

621 The selection efficiency of this partially reconstructed sample proxy consists of following
622 efficiencies: combinatorial BDT, misID BDT, fitting range window selection. It can be seen
623 in Figure 25, that this particular selection is constant as a function of key kinematic variables for
624 dipion which emulates two muons from D (they are the proxies for the two opposite sign muons).
625 The selection efficiency as a function of this dipion mass is flat and can be seen in Figure 25
626 ((a)(b)). This is very important in the low dipion region where the real two muons could come
627 also from γ^* and hence have a photon pole. It can be seen that in this region the efficiency is
628 still constant within errors. So in conclusion, as the efficiency is flat across the important dipion
629 invariant mass spectrum, this selection does not have model dependence and hence it is safe to
630 use for shape estimates for the partially reconstructed backgrounds. In addition, combination of
631 pions with the last muons are shown Figure 25(c)(d)(e)(f).

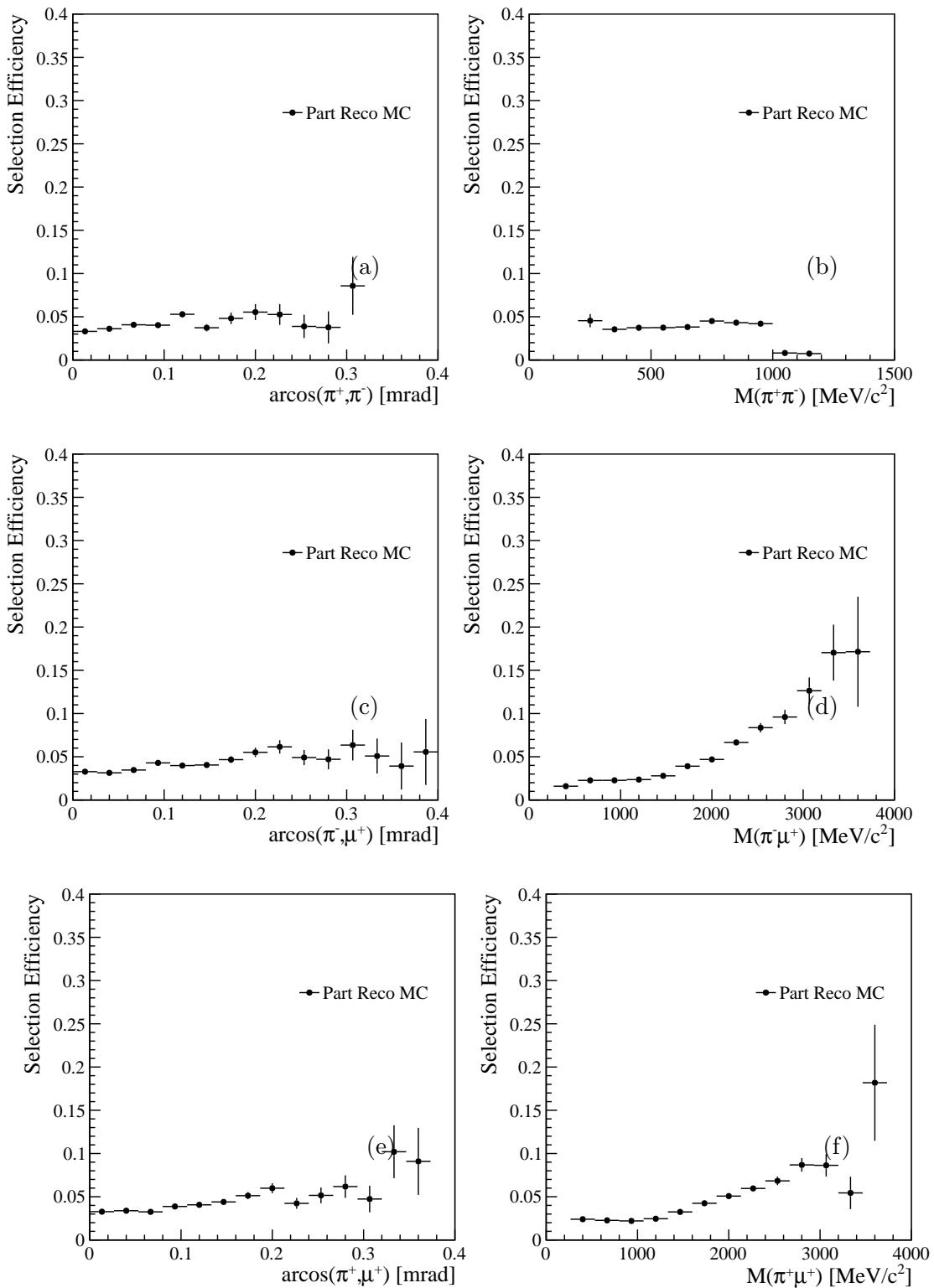


Figure 25: (a) Selection efficiency of partially reconstructed background ($B^+ \rightarrow (D^0 \rightarrow K^+\pi^-\mu^+\mu^-)\mu^+\nu$ decays) as a function of the opening angle between two opposite sign pions (the proxy for two opposite sign muons coming from D) and (b) its invariant mass. (c) Opening angle between oppositely charged pion and muon (here pion is from D and muon from \bar{B}) and its (d) invariant mass and finally (e) between same charged pion and muon (here pion is from D and muon from B). Efficiency is relatively constant across the opening angles as well as invariant mass of dipion invariant mass (at least where most of the events are located).

632 **5.3.2 Partially Reconstructed Backgrounds, where $D^0 \rightarrow \eta/\eta' X$, where $\eta/\eta' \rightarrow \mu\mu\gamma$**

633 In the previous partially reconstructed sample, the background that proceed via η/η' from D
 634 resonance is not considered, as it is not part of the inclusive simulation. The selection efficiency
 635 of such decays is expected to have very similar values as in the partially reconstructed sample
 636 proxy, because the reconstructed particles are the same.

637 In this section, the total estimate for the BF of partially reconstructed backgrounds proceeding
 638 via η/η' from D resonances is computed. Full inclusive rate $\mathcal{B}(D^0 \rightarrow \eta/\eta' X)$ is $\sim 10\%$. However,
 639 the most relevant decay chains are the ones where the mass of the missed particle is small. This
 640 is because if only light particle is missed, the shape of corrected mass of partially reconstructed
 641 background comes closest to the signal region. Such decay chains are considered in Table 15.

642 It can be seen that total cumulative contribution is much smaller than the one considered with
 643 $D^0 \rightarrow K^+\pi^-\mu^+\mu^-$, where $\mathcal{B}(D^0 \rightarrow K\pi^+\mu^+\mu^-) = (4.17 \pm 0.12(\text{stat}) \pm 0.40(\text{syst})) \times 10^{-6}$ [12].

644 Systematic uncertainty due to this type of background is evaluated post-unblinding in
 645 subsection 10.5.

Process	\mathcal{B}	Contribution to $D^0 \rightarrow (\eta/\eta' \rightarrow \mu\mu\gamma)X$
$\mathcal{B}(\eta \rightarrow \mu\mu\gamma)$	$(3.10 \pm 0.40) \times 10^{-4}$	-
$\mathcal{B}(\eta' \rightarrow \mu\mu\gamma)$	$(1.08 \pm 0.27) \times 10^{-4}$	-
$\mathcal{B}(D^0 \rightarrow \eta'\pi^0)$	$(9.10 \pm 1.40) \times 10^{-4}$	$(9.80 \pm 2.90) \times 10^{-8}$
$\mathcal{B}(D^0 \rightarrow \eta'\pi^+\pi^-)$	$(4.50 \pm 1.70) \times 10^{-4}$	$(4.90 \pm 2.20) \times 10^{-8}$
$\mathcal{B}(D^0 \rightarrow 2\eta)$	$(1.70 \pm 0.02) \times 10^{-3}$	$(5.30 \pm 0.70) \times 10^{-7}$
$\mathcal{B}(D^0 \rightarrow 2\eta)$	$(1.70 \pm 0.02) \times 10^{-3}$	$(5.30 \pm 0.70) \times 10^{-7}$
$\mathcal{B}(D^0 \rightarrow \eta\eta')$	$(1.06 \pm 0.27) \times 10^{-3}$	$(3.30 \pm 0.90) \times 10^{-7}$
$\mathcal{B}(D^0 \rightarrow \eta\eta')$	$(1.06 \pm 0.27) \times 10^{-3}$	$(1.10 \pm 0.40) \times 10^{-7}$
$\mathcal{B}(D^0 \rightarrow \eta\phi)$	$(1.40 \pm 0.50) \times 10^{-4}$	$(4.30 \pm 1.60) \times 10^{-8}$
Total	-	$(1.69 \pm 0.15) \times 10^{-6}$

Table 15: Contribution to total $D^0 \rightarrow (\eta/\eta' \rightarrow \mu\mu\gamma)X$ rate made from all the considered decays above.
 In total, this cumulative contribution is around three times smaller than $D^0 \rightarrow K^+\pi^-\mu^+\mu^-$.

646 **5.4 MisID type background**

647 MisID background is one of the most prominent backgrounds that is expected to be present. This
 648 type of background proceeds mostly via cascade decays, where $B^+ \rightarrow (\overline{D^0} \rightarrow [K^+, \pi^+] \mu^-\nu) \mu^+\nu$
 649 and then $[K^+, \pi^+]$ are misidentified as muons. The contributions from decays where two muons
 650 are correctly identified as muons and third track is consistent with proton passing all the selection
 651 is also considered, however it will be shown that this contribution will be very low. In this case
 652 the sign of the misidentified particle agrees with the sign of the mother B , this will be denoted
 653 as same sign misID background (SS misID) background. It is possible to also have the opposite
 654 sign particle misidentified (OS misID), however this contribution is expected to have smaller
 655 rate as the misidentified particle would have to proceed via decays with additional particles.
 656 These two types of backgrounds are studied using data-driven method described below using
 657 dedicated stripping line `StrippingB23MuNu_TriFakeMuLine` with a 1% prescale. Finally, also
 658 double misID, where two of the muons are misidentified was studied using *small data sample*,

659 with luminosity $5.8 \pm 0.3 \text{ pb}^{-1}$. This sample is specific as there was no stripping applied to it.
 660 The *Small data sample* was also used to cross-check the consistency of results for single muon
 661 misID both (SS an OS).

662 The stripping selection for this sample is similar but with `no PID cut` for the misidentified
 663 muon (SS or OS), see Table 16.

Table 16: Stripping Cuts for Misidentified Background

B^+ candidate cut	First two μ candidates	Third μ candidate
$2500 \text{ MeV}/c^2 < M_{B_{\text{corr}}} < 10000 \text{ MeV}/c^2$ $\theta_B > 0.999$ $p_T > 2000 \text{ MeV}/c$ $\chi^2_{FD} > 50$ $\chi^2/\text{dof} < 4$ $0 \text{ MeV}/c^2 < M_B < 7500 \text{ MeV}/c^2$	$\chi^2_{minIP} > 9$ $\chi^2_{tr}/\text{dof} < 3$ $P_{ghost} < 0.35$ $P_T > 0$ $\Delta LL(\mu - \pi) > 0$ $\Delta LL(\mu - K) > 0$ <code>StdLooseMuons</code>	$\chi^2_{minIP} > 9$ $\chi^2_{tr}/\text{dof} < 3$ $P_{ghost} < 0.35$ $P_T > 0$ <code>isMuon==0.0</code> <code>inMuon==1.0</code> <code>StdNoPIDsMuons</code>

664 The same trigger as well as offline selection (apart from PID) selection as for signal is applied.
 665 The shape and amount of misID contamination is used at many different stages of this analysis,
 666 namely:

- 667 • Extract the number of events for combinatorial BDT
 668 • Extract the number of events for misID BDT
 669 • Fix the shape for signal fit

670 To estimate the shape of the background and size of it following procedure is applied:

- 671 • The sample is subdivided into 3 different samples where the third "muon" is classified to
 672 be in proton-, pion- or kaon-like region, see Table 17.

Region	ID cuts
Proton-like	$\Delta LL(p - \pi) > 5, \Delta LL(p - K) > 5$
Kaon-like	$\Delta LL(K - \pi) > 0, \Delta LL(p - K) < 5$
Pion-like	$\Delta LL(K - \pi) < 0, \Delta LL(p - \pi) < 5$

Table 17: Species region definitions. These cuts define regions in $\Delta LL(p), \Delta LL(K), \Delta LL(\pi)$.

- 673 • ID efficiencies are obtained from `PIDCalib` in bins of p, η for all three regions.
 674 • MisID efficiencies are obtained from the calibration sample in section 6 in the bins of p, η .
 675 • Some of the entries in the tuples are outside of default bins of `PIDCalib`/Calibration
 676 samples. For events below $p < 3 \text{ GeV}/c$, these are not considered and cut out (in signal
 677 data this is done by requiring `isMuon`). For events $p > 100 \text{ GeV}/c$, the same id/misID
 678 probability as the last bin is assigned.

- 679 • Some bins of **PIDCalib**/Calibration samples yield probabilities that are unphysical. These
 680 bins have either negative probabilities or probabilities above 1 (can arise due to *Splot*
 681 technique used in **PIDCalib**). For bins with negative probabilities, probability is set to
 682 0.00001. For bins with probability above 1 are set 1. And for bins which were not populated,
 683 the value from neighbouring bin was used. Remark: negative bins only appear for proton
 684 misID. As it will be shown in the misID fit section there is very small contribution from
 685 protons so there is no bias for the pion/kaon samples. The misID probabilities for different
 686 years will be discussed and shown in Figure 28 and as it can be seen there are no negative
 687 bins.

- 688 • It is also possible to have cross-contamination between the three species themselves. In
 689 order to account for this, the following procedure is applied:

- 690 – The data in each region is binned to obtain two dimensional $N(p, \eta)$ distributions,
 691 where p is momentum and η is pseudorapidity. Taking an example for $K\pi$ cross-
 692 contamination, the true kinematical distributions for kaons and pions are given by

$$n(p, \eta)_{\pi/K}^0 = \frac{N(p, \eta)_{\pi/K}}{\epsilon(p, \eta)_{\pi/K}}. \quad (8)$$

693 where $\epsilon(p, \eta)_{\pi/K}$ are efficiencies obtained from **PIDCalib** tables.

- 694 – To correct for the cross-feed between pion and kaon regions, following algorithm is
 695 applied:

$$n(p, \eta)_{\pi}^{i+1} = \frac{N(p, \eta)_{\pi} - M(p, \eta)_{K \rightarrow \pi} n(p, \eta)_K^i}{\epsilon(p, \eta)_{\pi}}. \quad (9)$$

$$n(p, \eta)_K^{i+1} = \frac{N(p, \eta)_K - M(p, \eta)_{\pi \rightarrow K} n(p, \eta)_{\pi}^i}{\epsilon(p, \eta)_K}. \quad (10)$$

696 Here, $n(p, \eta)_{\pi}^i$ $n(p, \eta)_K^i$ together with the misID binned efficiencies $M(p, \eta)_{K \rightarrow \pi}$ and
 697 $M(p, \eta)_{\pi \rightarrow K}$ are estimating the-cross contamination between two regions.

- 699 – The next order distributions $n(p, \eta)_{\pi/K}^{i+1}$ are obtained by correcting the original
 700 distributions with the cross-contamination and then correcting for the ID binned
 701 efficiency.
 702 – At each iteration, the total number of misID particles of the type $\pi \rightarrow \mu$ and $K \rightarrow \mu$
 703 events are given by

$$\sum_{p, \eta} n(p, \eta)_{\pi}^i M(p, \eta)_{\pi \rightarrow \mu} \quad (11)$$

$$\sum_{p, \eta} n(p, \eta)_K^i M(p, \eta)_{K \rightarrow \mu} \quad (12)$$

- 705 – This procedure is repeated until the change in total misID between iterations is less
 706 than 0.1%. Typical number of iterations depend also on the size of the sample. For big
 707 samples the convergence is achieved after two or three iterations. For small samples
 708 this is achieved after six iterations on average.

- 709 – For each event in both kaon-like and pion-like sample, $w_{crossfeed}$ = probability of
 710 being misidentified particle including the cross-contamination correction is added.

$$w_{crossfeed} = \frac{n(p, \eta)_{\pi}^{final} \times M(p, \eta)_{\pi \rightarrow \mu}}{N(p, \eta)_{\pi}^0} \quad (13)$$

$$w_{crossfeed} = \frac{n(p, \eta)_K^{final} \times M(p, \eta)_{K \rightarrow \mu}}{N(p, \eta)_K^0} \quad (14)$$

- 712 • The number of misidentified events and the shape are obtained by reweighting the datasets
 713 of the pion and kaon by crossfeed weight. The difference between unweighted, weighted by
 714 PID efficiencies with no crossfeed, and weighted with crossfeed distributions can be seen
 715 in Figure 26.

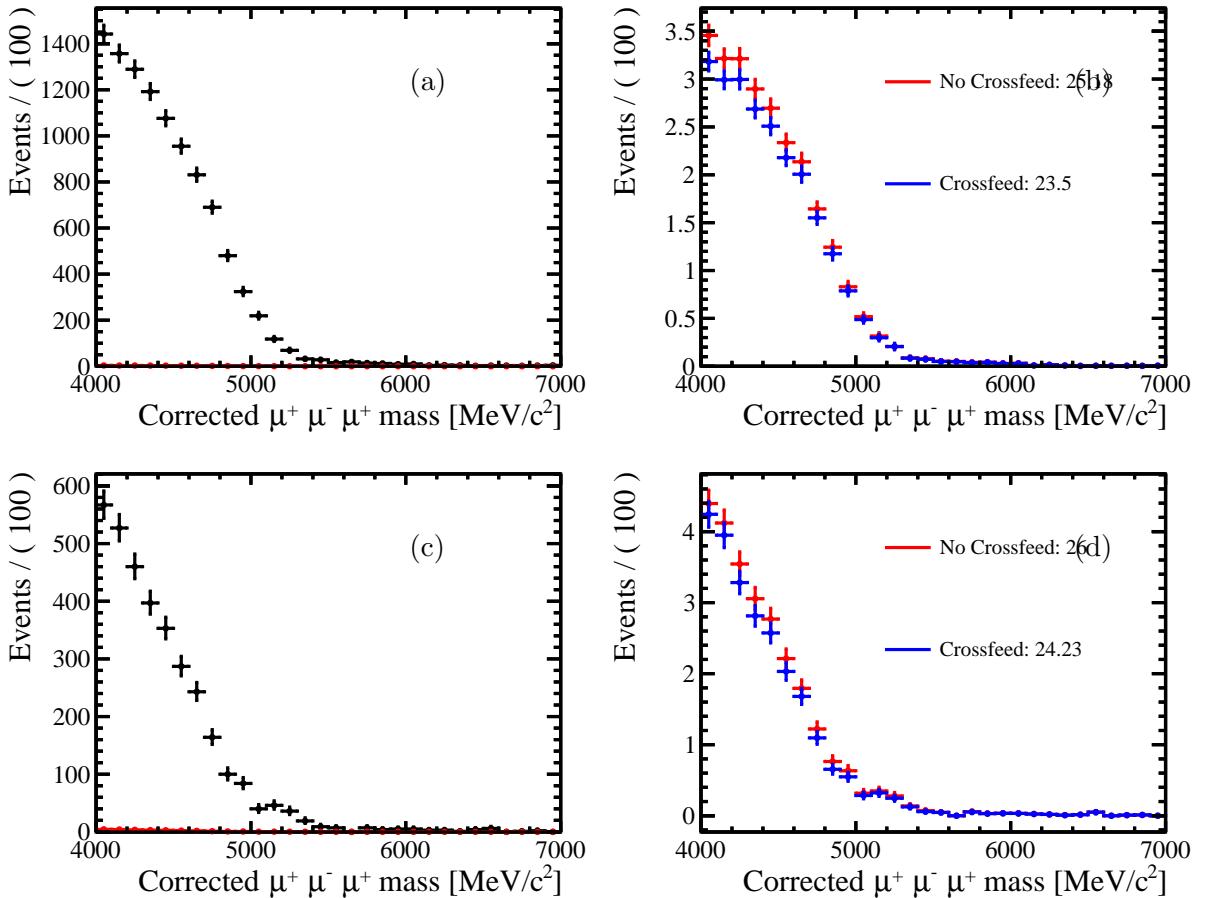


Figure 26: Crossfeed between kaon-like and pion-like regions. Examples of misID procedure applied to obtain yields for Run 1 misID BDT. On the left, unweighted misID distributions for (a) kaon SS (c) pion SS (e) kaon OS (g) pion OS Run 1 data. On the right, weighted misID distributions for Run 1 (b) kaon SS (d) pion SS (f) kaon OS (h) pion OS, both with and without crossfeed. These shapes are obtained after combinatorial BDT was applied, but before MisidBDT was applied. Total yields need to be multiplied by 100.

716 **5.5 Rare $B^+ \rightarrow \pi^+/K^+\mu^-\mu^+$ Background**

717 Despite the fact that background coming from $B^+ \rightarrow \pi^+(J/\psi \rightarrow \mu^-\mu^+)$ is eliminated because
 718 of the $c\bar{c}$ veto, it is necessary to evaluate the impact of the background coming from the rare
 719 $B^+ \rightarrow \pi^+\mu^-\mu^+$ decay as this kind of background can cause potential problem as it peaks just
 720 under the corrected mass of B . For equivalent rare decay but with kaon, $B^+ \rightarrow K^+\mu^-\mu^+$, the
 721 mass will be shifted away from this peak.

722 $B^+ \rightarrow \pi^+(J/\psi \rightarrow \mu^-\mu^+)$ MC in 2012 and 2016 was run through signal stripping and
 723 all the signal selection. Using PDG's values for both branching fractions (i.e multiplying
 724 $\mathcal{B}(B^+ \rightarrow J/\psi K^+) = (1.026 \pm 0.031) \times 10^{-3}$ [13] and $\mathcal{B}(J/\psi \rightarrow \mu^-\mu^+) = (5.961 \pm 0.0033) \times 10^{-2}$ [13]
 725 yielding $\mathcal{B}(B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+) = (6.12 \pm 0.19) \times 10^{-5}$), one can normalise to normalisation
 726 and obtain number of events in a given year,

$$N_{year} = \frac{N(B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+)}{\varepsilon_{gen_{B^+ \rightarrow \pi^+\mu^+\mu^-}} \times \varepsilon_{sel_{B^+ \rightarrow \pi^+\mu^+\mu^-}}} \times \frac{\mathcal{B}(B^+ \rightarrow \pi^+\mu^+\mu^-)}{\mathcal{B}(B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+)}. \quad (15)$$

727 It was shown in Table 18 that this component is negligible compared to the number of signal
 728 events that are expected⁸. Hence, this component does not need to be treated separately.

Properties	Run 1 (2012 MC)	2016 (2012 MC)
$\mathcal{B}(B^+ \rightarrow \pi^-\mu^+\mu^-)$	$(1.79 \pm 0.23) \times 10^{-8}$	$(1.79 \pm 0.23) \times 10^{-8}$
$\mathcal{B}(B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+)$	$(6.12 \pm 0.19) \times 10^{-5}$	$(6.12 \pm 0.19) \times 10^{-5}$
$\varepsilon_{gen_{B^+ \rightarrow \pi^+\mu^+\mu^-}}$	0.183 ± 0.002	Using 2012
$\varepsilon_{sel_{B^+ \rightarrow \pi^+\mu^+\mu^-}}$	$(3.85 \pm 0.62) \times 10^{-5}$	Using 2012
$\varepsilon_{gen_{B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+}}$	0.162 ± 0.0002	Using 2012
$\varepsilon_{sel_{B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+}}$	$(3.58 \pm 0.007) \times 10^{-2}$	Using 2012
$\mathcal{B} \times \varepsilon_{pr}$	$(1.26 \pm 0.26) \times 10^{-14}$	Using 2012
$\mathcal{B} \times \varepsilon_{norm}$	$(3.55 \pm 0.11) \times 10^{-7}$	Using 2012
$\frac{\mathcal{B} \times \varepsilon_{pr}}{\mathcal{B} \times \varepsilon_{norm}}$	$(3.55 \pm 0.74) \times 10^{-7}$	Using 2012
$N(B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+)$	173000 ± 446	94500 ± 314
$N(B^+ \rightarrow \pi^-\mu^+\mu^-)$	0.0616 ± 0.0128	0.0335 ± 0.00697

Table 18: Calculation of number of events that comes from rare $B^+ \rightarrow \pi^-\mu^+\mu^-$ backgrounds, assuming 2012 efficiencies but extrapolating to 2011,2012,2016 samples. $\mathcal{B}(B^+ \rightarrow \pi^-\mu^+\mu^-)$ is obtained from PDG $B^+ \rightarrow \pi^-\mu^+\mu^- = (1.79 \pm 0.23) \times 10^{-8}$. Normalisation channel $B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+$ was used to normalize to using Run 1 and 2016 yields. Ratio of PID efficiencies is 1.

729 **5.6 $B \rightarrow \eta(')V$ Backgrounds**

730 The backgrounds with $\eta(')$ resonances from partially reconstructed decays that proceed via D
 731 were considered in subsubsection 5.3.2. In this section backgrounds from $\eta(')$ along with vector
 732 resonances ω/ρ had come directly from B are estimated. The total branching fractions are seen

⁸For full Run 1 + 2016, around 17 signal events are expected if $\mathcal{B}(B^+ \rightarrow \mu^+\mu^-\mu^+\nu) = 1 \times 10^{-8}$

733 in Table 19 and since they are small this background is discarded and will not be considered
 734 further.

Process	\mathcal{B}
$\mathcal{B}(B^0 \rightarrow \omega\eta')$	$(1.00 \pm 0.50) \times 10^{-6}$
$\mathcal{B}(B^0 \rightarrow \rho\eta')$	$< 5 \times 10^{-7}$
$\mathcal{B}(B^0 \rightarrow \omega\eta)$	$(9.00 \pm 4.00) \times 10^{-7}$
$\mathcal{B}(B^0 \rightarrow \rho\eta)$	$< 5 \times 10^{-7}$
$\mathcal{B}(\eta \rightarrow \mu\mu\gamma)$	$(3.10 \pm 0.40) \times 10^{-4}$
$\mathcal{B}(\eta' \rightarrow \mu\mu\gamma)$	$(1.08 \pm 0.27) \times 10^{-4}$
$\mathcal{B}(\rho \rightarrow \mu\mu)$	$(4.55 \pm 0.28) \times 10^{-5}$
$\mathcal{B}(\omega \rightarrow \mu\mu)$	$(9.00 \pm 3.10) \times 10^{-5}$
$\mathcal{B}(B^0 \rightarrow (\omega \rightarrow \mu\mu)(\eta \rightarrow \mu\mu\gamma))$	$(7.10 \pm 1.00) \times 10^{-15}$
$\mathcal{B}(B^0 \rightarrow (\omega \rightarrow \mu\mu)(\eta' \rightarrow \mu\mu\gamma))$	$(2.50 \pm 0.60) \times 10^{-15}$
$\mathcal{B}(B^0 \rightarrow (\rho \rightarrow \mu\mu)(\eta \rightarrow \mu\mu\gamma))$	$<(2.50 \pm 1.40) \times 10^{-14}$
$\mathcal{B}(B^0 \rightarrow (\rho \rightarrow \mu\mu)(\eta' \rightarrow \mu\mu\gamma))$	$<(1.00 \pm 0.60) \times 10^{-14}$
Total	$< (4.50 \pm 1.60) \times 10^{-14}$

Table 19: Different and total contribution to $B^0 \rightarrow \eta(')\rho(\omega)$.

735 5.7 Summary

736 In summary, in this section different backgrounds that could be of potential threat to signal were
 737 studied. From all these possible backgrounds only combinatorial (more in subsection 5.2), misID
 738 (more in subsection 5.4) and partially reconstructed backgrounds (more in subsection 5.3) have
 739 still contribution and need to be modelled after all of the selection as will be shown in section 9.

740 6 Specific control sample for $K/\pi \rightarrow \mu$ misID rates

741 In order to assign weights to misID shapes, misID probabilities for kaons, pions and protons into
 742 muons need to be known. At LHCb, **PIDCalib** package can usually provide these misID rates
 743 as it contains K/π control samples. These are usually statistically populated background-free
 744 *sWeighted* samples, for which it is possible to extract misID and ID rates as a function of
 745 kinematics given certain PID criteria. In this analysis, some control samples from **PIDCalib** are
 746 used (for proton to muons misID estimates), however for the K/π misID rates use of **PIDCalib**
 747 samples is not satisfactory for reasons below.

748 In **PIDCalib** for K/π misID rates, tracks from $D^{*+}(\rightarrow D^0(\rightarrow K^+\pi^-)\pi^+)$ decays are used. To
 749 obtain the rates the distributions of the PID variables for each track can be studied. Unfortunately,
 750 the topology of the misID background component is two real muons with an additional fake
 751 muon, which is very different to **PIDCalib** sample $D^0 \rightarrow K^+\pi^-$. This influences rest of misID
 752 rates as some of the PID variables are strongly correlated with number of muons in the decay,
 753 due to the fact that the misidentified particle can share hits with other muons in the rest of the
 754 decay. This is reflected mostly in high momenta region, where the three particles are collimated.
 755 An alternative sample, $B^0 \rightarrow J/\psi(\rightarrow \mu^+\mu^-)K^*$, which mimics the rates for fake muon correctly,

756 will be used to obtain misID rates. This is a clean decay that can be fit to obtain *sWeights* to
 757 measure PID efficiencies. Moreover, its the signal topology and kinematics are much closer to
 758 the actual misID background component than that of the **PIDCalib** control samples.

759 **6.1 Selection for $B^0 \rightarrow J/\psi(\rightarrow \mu^+\mu^-)K^*$**

760 As in this chapter, PID performance ($K/\pi \rightarrow \mu$ misID/ID rates) is to be studied, both the
 761 data and simulation needs to be obtained from such stripping selections where no PID is
 762 applied on pion or kaon. This is achieved by looking at samples from **B2Xmumuline** stripping
 763 line, where pions are from **Phys/StdAllNoPIDsPions/Particles** container and kaons from
 764 **Phys/StdAllNoPIDsKaons/Particles** container, meaning that no PID variable was applied on
 765 kaon/pion candidate track. So in total, the candidates have **B2Xmumuline** stripping line selection
 766 applied to them. Further, more stringent $B^+ \rightarrow \mu^+\mu^-\mu^+\nu$ pre-selection, selection and trigger
 767 (on J/ψ level) was applied. Finally to remove most of the backgrounds from this sample cuts
 768 summarized in Table 20 are used.

Idea	Cut
ID K^*	$ M(K\pi) - M_{PDG}(K_0^*) < 100 \text{ MeV}/c^2$
Compatible with PIDCalib	for K, π , $p_T > 250 \text{ MeV}/c$
Compatible with PIDCalib	for μ , $p_T > 800 \text{ MeV}/c$
Muon swap	$ M((h \rightarrow \mu)\mu) - M_{PDG}(J/\psi) > 60 \text{ MeV}/c^2$
$B^+ \rightarrow K^+\mu^+\mu^-$	$\max(M(K^+\mu^+\mu^-)), M((\pi^+ \rightarrow K)\mu^+\mu^-)) < 5100 \text{ MeV}/c^2$
$B_s^0 \rightarrow \phi\mu^+\mu^-$	$M(K(\pi \rightarrow K)) > 1040 \text{ MeV}/c^2$
Id muons	$\text{mu1_Probnnmu} > 0.5$ and $\text{mu2_Probnnmu} > 0.5$
For kaon misID rates:Id pion	$\text{PIDK} < 0$ $\text{PIDp} < 0$ and $\text{IsMuon} == 0$
For pion misID rates:Id kaon	$\text{PIDK} > 0$ and $\text{PIDK-PIDp} > 0$ and $\text{IsMuon} == 0$

Table 20: Offline selection for $B^0 \rightarrow J/\psi(\rightarrow \mu^+\mu^-)K^*$ decay.

769 **6.2 Fitting Strategy for $B^0 \rightarrow J/\psi(\rightarrow \mu^+\mu^-)K^*$ decay**

770 One more procedure is put in place before fitting the spectrum, known as mass constraint. The
 771 mass of J/ψ is constrained to its nominal mass, which yields new estimates for track parameters
 772 of the final state particles, from which a new kinematic refit is done. In order to obtain these
 773 K/π misID rates fit to constrained mass between 5150 - 5450 MeV/c^2 was performed, where all
 774 of the yields are left floating. Components to the fit are:

- 775 • Signal - $B^0 \rightarrow J/\psi K^*$ - fixed shape from MC apart from mean μ and σ , which is fitted
 776 with double sided Ipatia function [14]
- 777 • Combinatorial background - exponential function
- 778 • Background $B_s^0 \rightarrow \pi^+ K^- J/\psi \rightarrow \mu^+\mu^-$ - the shape is same as signal, but with offset of μ
 779 by PDG difference between B_s^0 and B^0

- 780 • Background from $\Lambda_b \rightarrow K^- p \mu^+ \mu^-$ - fix shape from MC, fitted with `RooKeys` pdf (non-
 781 parametric function composed of superposition of Gaussians with equal surface, but with
 782 different widths, which are established by data at a given point)
- 783 • Background coming from $K \leftrightarrow \pi$ swaps - fixed from MC, fitted with double sided Crystal
 784 Ball function

785 In order to obtain misID rates two approaches were considered: either *fit twice method*
 786 or *sPlot* method (the same as in `PIDCalib` package). The *fit twice method* consists of fitting
 787 $B^0 \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)K^*$ before and after PID cut in each kinematic bin separately. MisID
 788 probabilities are then obtained from ratio of signal yields from these two fits. *sWeight* method,
 789 the distribution is fitted once and each event is assigned *sWeights* through *sPlot* technique. The
 790 misID probabilities are then obtained by looking at *sWeights* in each bin. This was done in
 791 order to make sure that *sWeight* procedure is not introducing biases. This could happen if PID
 792 variables (discriminating variables) for the background are correlated with the mass (fitting
 793 variable). It turns out that these two methods yields very similar results, hence, for purposes of
 794 this analysis *sWeight* values will be used. Fits to Run 1 and 2016 data can be seen in Figure 27.

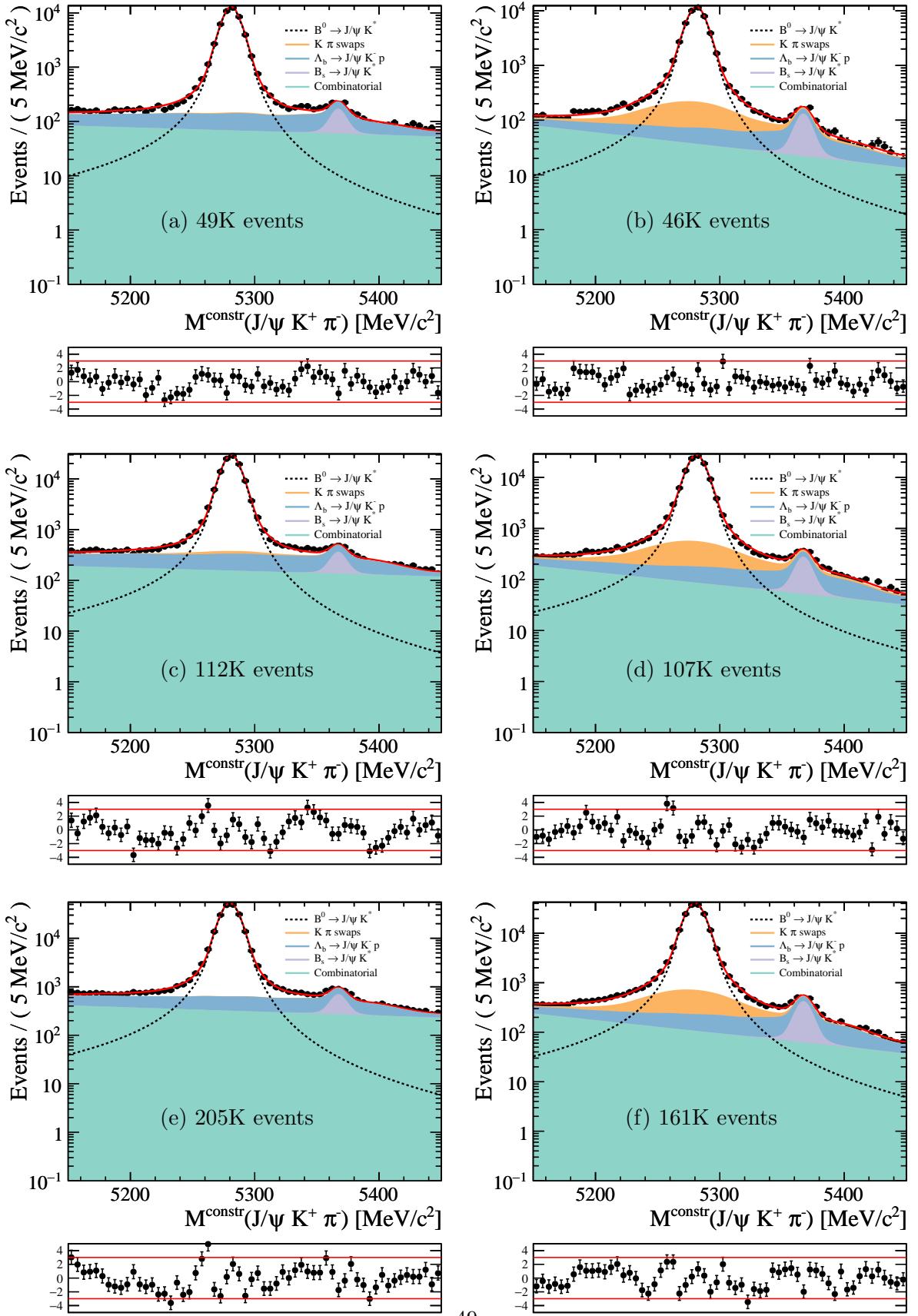


Figure 27: Fit to constrained $J/\psi(\rightarrow \mu^+ \mu^-)K^*(\rightarrow \pi^+ K^-)$ mass with all the components for (a)(b) 2011, (c)(d) 2012, (e)(f) 2016. On the left, fit to data with pion ID (giving kaon misid probabilities), on right data with kaon ID (pion misid rates).

795 **6.3 Results of $B^0 \rightarrow J/\psi(\rightarrow \mu^+\mu^-)K^*$ control sample for $K/\pi \rightarrow \mu$ misID rates**

796 Using the *sWeight* method, misID rates for K/π are obtained. In Figure 28, the rates for different
797 PID cuts are seen. In Figure 29, these rates are compared to rates obtained from **PIDCalib**. In
798 the low momenta region, there is generally good agreement between the two samples. This can be
799 explained as these particles are softer, and they will not be collimated, causing less interference
800 with other two real muons in decay. In high momenta region, particles will tend to be more
801 collimated and hence the influence of other two real muons will lead to bigger disagreement.
802 This is because the other two real muons leave hits in the muon chambers and hence the rate of
803 **IsMuon==1.0** is higher. The disagreement between the two samples in high momentum region is
804 decreased by requiring **nShared==0.0**, as having 2 other muons to share hits will be more likely.
805 The effect of other PID variables can also be seen, but it is harder to interpret as these depend
806 on several variables.

807 In conclusion, the misID rates obtained from $B^0 \rightarrow J/\psi(\rightarrow \mu^+\mu^-)K^*$ control sample will be
808 used for misID probabilities.

809 Finally, it was also checked that these two samples contain the same fraction of kaon
810 tracks within fiducial region of muon acceptance. This was checked by requiring that the
811 **InMuonAcceptance==1.0** for the kaon tracks and the result can be seen in Table 21.

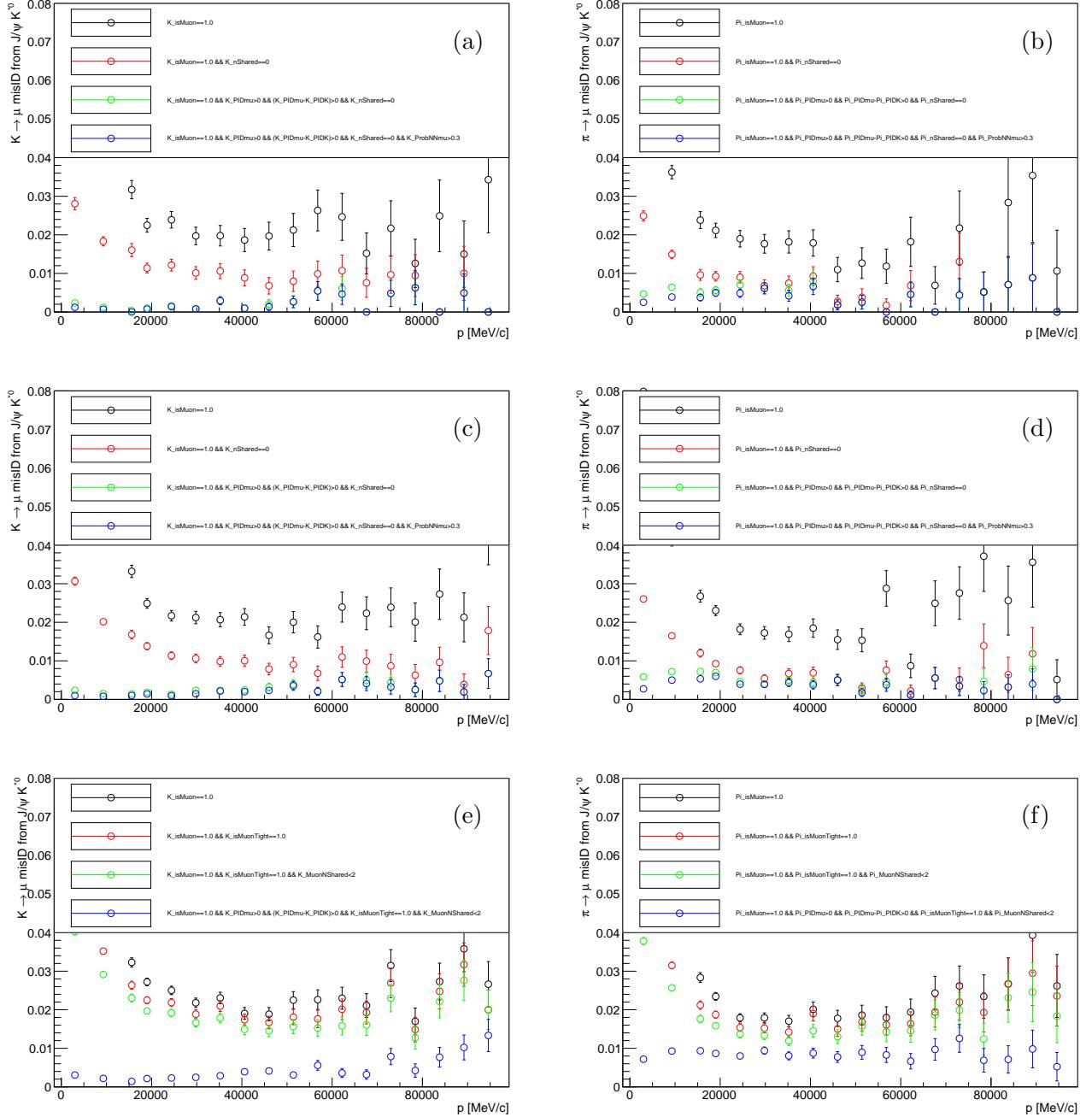


Figure 28: Weights for different PID cuts from $B^0 \rightarrow J/\psi(\rightarrow \mu^+ \mu^-) K^*$ for (a)(b) 2011, (c)(d) 2012, (e)(f) 2016 data in different bins of p . On the left, kaon misID rates, on right pion misID rates.

For (a)(b)(c)(d) In black: `isMuon==1.0` is probed, in red: `isMuon==1.0 && nShared==0`, in green: `isMuon==1.0 && nShared==0 && Δ LL(μ - π) > 0 && ((Δ LL(μ - π) - (Δ LL(K - π)) > 0`, in blue: `isMuon==1.0 && nShared==0 && Δ LL(μ - π) > 0 && ((Δ LL(μ - π) - (Δ LL(K - π)) > 0 && ProbNNμ > 0.3`.

For (e)(f) In black: `isMuon==1.0` is probed, in red: `isMuon==1.0 && isMuonTight==1.0`, in green: `isMuon==1.0 && isMuonTight==1.0 && nShared<2 && Δ LL(μ - π) > 0 && ((Δ LL(μ - π) - (Δ LL(K - π)) > 0`, in blue: `isMuon==1.0 && nShared<2 && isMuonTight==1.0 && Δ LL(μ - π) > 0 && ((Δ LL(μ - π) - (Δ LL(K - π)) > 0`.

These plots show that misID in low p region is highly suppressed with ΔLL (defined in subsection 4.2) and `Probnnmu` cuts leading to misID probabilities that are less than 0.5%.

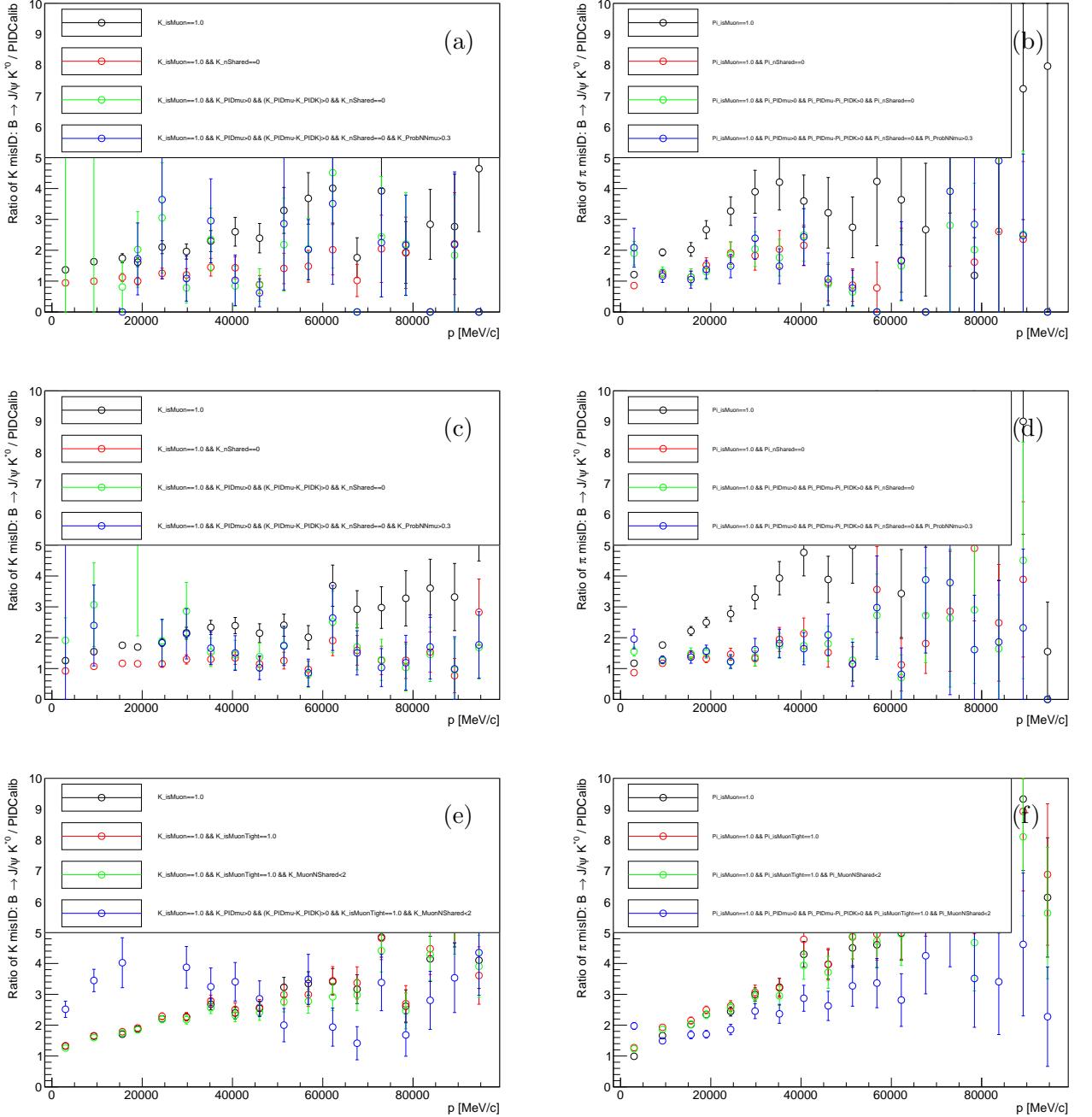


Figure 29: Ratio of weights from $B^0 \rightarrow J/\psi(\rightarrow \mu^+\mu^-)K^*$ as compared to those from PIDCalib for (a)(b) 2011, (c)(d) 2012, (e)(f) 2016 data in different bins of p . It can be seen that at low momenta region the agreement is very good, however in high momenta region, where particles become collimated this is not anymore the case.

For (a)(b)(c)(d) In black: `isMuon==1.0` is probed, in red: `isMuon==1.0 && nShared==0`, in green: `isMuon==1.0 && nShared==0 && Δ LL(μ - π) > 0 && ((Δ LL(μ - π) - (Δ LL(K - π)) > 0`, in blue: `isMuon==1.0 && nShared==0 && Δ LL(μ - π) > 0 && ((Δ LL(μ - π) - (Δ LL(K - π)) > 0 && ProbNNμ > 0.3`.

For (e)(f) In black: `isMuon==1.0` is probed, in red: `isMuon==1.0 && isMuonTight==1.0`, in green: `isMuon==1.0 && isMuonTight==1.0 && nShared<2 && Δ LL(μ - π) > 0 && ((Δ LL(μ - π) - (Δ LL(K - π)) > 0`, in blue: `isMuon==1.0 && nShared<2 && isMuonTight==1.0 && Δ LL(μ - π) > 0 && ((Δ LL(μ - π) - (Δ LL(K - π)) > 0`. Δ LL variables are defined in subsection 4.2.

p low [MeV/c]	p high[MeV/c]	PIDCalib	$B^0 \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)K^*$	Ratio
3000	6000	0.77±0.0016	0.83±0.0047	1.1±0.0065
6000	9300	0.93±0.00030	0.95±0.0019	1.0±0.0020
9300	10000	0.96±0.00037	0.97±0.0031	1.0±0.0033
10000	12600	0.97±0.00014	0.97±0.0017	1.0±0.0017
12600	15600	0.98±0.00011	0.97±0.0017	0.99±0.0018
15600	17500	0.98±0.00013	0.96±0.0024	0.98±0.0025
17500	21500	0.98±8.9e-05	0.96±0.0018	0.98±0.0018
21500	27000	0.98±7.8e-05	0.96±0.0018	0.98±0.0019
27000	32000	0.98±8.8e-05	0.96±0.0024	0.98±0.0025
32000	40000	0.98±8.0e-05	0.96±0.0022	0.98±0.0022
40000	60000	0.97±7.5e-05	0.95±0.0021	1.0±0.0022
60000	70000	0.96±0.00016	0.96±0.0043	1.0±0.0046
70000	100000	0.95±0.00013	0.94±0.0044	0.99±0.0046

Table 21: K_InMuonAcc==1.0 shows the interpolation of K tracks into muon chambers for 2012 MagDown data samples with given binning. It can be seen that both samples agree with each other, showing that the fraction of the kaons tracks within muon acceptance is very similar. The same check was performed for pions. This measurement is in a bin $1.5 < \eta < 5.0$.

813 **7 Efficiency**

814 In this section overview of the efficiencies is discussed. Firstly relative efficiency fraction will
 815 be defined and then individual efficiency calculations are described. As 2011 signal simulation
 816 MC was not generated, it was checked that the efficiencies in normalisation channel reveal no
 817 significant difference between Stripping 21 and 21r1 (more on this can be seen in subsection 8.4).
 818 Hence, it is assumed that Stripping 21r1 MC efficiencies (2011) will be the same efficiencies as
 819 Stripping 21 (2012).

820 **7.1 Efficiency Ratio**

821 To be able to translate observed signal events into branching fraction estimate, the normalisation
 822 channel of $B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+$ is used. Both, for signal and normalisation channel the
 823 absolute efficiencies, luminosity, the b-quark cross-section or fragmentation fractions will cancel.
 824 There are, however, efficiencies that will not cancel and will be necessary for the final limit setting
 825 procedure. These are shown in the calculation od relative efficiencies.

826 The relative efficiency is calculated as follows:

827

$$R_{\{\text{NOFCME}\}}^{(21,26)}(\varepsilon) = \frac{\varepsilon_s}{\varepsilon_n} = \frac{\varepsilon_s^{\text{GEN}}}{\varepsilon_n^{\text{GEN}}} \times \frac{\varepsilon_s^{\text{REC}}}{\varepsilon_n^{\text{REC}}} \times \frac{\varepsilon_s^{\text{TRG}}}{\varepsilon_n^{\text{TRG}}} \times \frac{\varepsilon_s^{\text{OFF}}}{\varepsilon_n^{\text{OFF}}} \times \frac{\varepsilon_s^{\text{CombiBDT}}}{\varepsilon_n^{\text{CombiBDT}}} \times \frac{\varepsilon_s^{\text{MisidBDT}}}{\varepsilon_n^{\text{MisidBDT}}} \times \frac{\varepsilon_s^{\text{FR}}}{\varepsilon_n^{\text{FR}}} \times \frac{\varepsilon_s^{\text{PID}}}{\varepsilon_n^{\text{PID}}}, \quad (16)$$

828

$$R_{\{\text{lowFCME,highFCME}\}}^{(21,26)}(\varepsilon) = \frac{\varepsilon_s}{\varepsilon_n} \times \frac{\varepsilon_s^{\text{FCME}}}{\varepsilon_n^{\text{FCME}}} = \frac{\varepsilon_s^{\text{GEN}}}{\varepsilon_n^{\text{GEN}}} \times \frac{\varepsilon_s^{\text{REC}}}{\varepsilon_n^{\text{REC}}} \times \frac{\varepsilon_s^{\text{TRG}}}{\varepsilon_n^{\text{TRG}}} \times \frac{\varepsilon_s^{\text{OFF}}}{\varepsilon_n^{\text{OFF}}} \times \frac{\varepsilon_s^{\text{CombiBDT}}}{\varepsilon_n^{\text{CombiBDT}}} \times \frac{\varepsilon_s^{\text{MisidBDT}}}{\varepsilon_n^{\text{MisidBDT}}} \times \frac{\varepsilon_s^{\text{FR}}}{\varepsilon_n^{\text{FR}}} \frac{\varepsilon_s^{\text{FCME}}}{\varepsilon_n^{\text{FCME}}} \times \frac{\varepsilon_s^{\text{PID}}}{\varepsilon_n^{\text{PID}}}, \quad (17)$$

829 where NOFCME means one bin of fractional corrected mass error, lowFCME and highFCME
 830 fractional corrected mass error are two bins of fractional corrected mass error, see subsection 4.8
 831 for more details. 21, 26 are the versions of stripping used. As seen in equations 16 and 17,
 832 absolute selection efficiencies ε_s , ε_n have several components.

833 Selection efficiency includes contribution from the detector acceptance efficiency (measured
 834 at the generator level and labelled (GEN)); the reconstruction selection efficiency (REC); the
 835 efficiency of the offline selection (OFF) comprising of trigger, J/ψ and $\Psi(2S)$ veto, MVA
 836 based selection (CombiBDT and MisidBDT); fitregion selection (FR); the efficiency of the PID
 837 requirement (PID). Since this analysis will be performed in two bins of FCME, and with two
 838 stripping versions there will be 4 efficiency ratios.

839 In order to obtain these individual efficiencies different methods were used. In particular for
 840 signal efficiencies, as this measurement was done in a particular $minq$ region, all signal efficiencies
 841 are calculated given the $minq$ selection. As the exact decay model is not known there is an
 842 uncertainty on how much of the signal sattisfies $minq$ requirement even though this efficiency
 843 is not part of efficiency calculations *per se*. Using default decay model this efficiency is around
 844 93%. This high efficiency is compared to Nikitin model shown in subsection 11.1, where also
 845 most of the events are below $q^2 < 1(\text{GeV}/c^2)^2$. The summary of method used to extract signal
 846 efficiency is shown in Table 22 and discussed below.

Component	Method
$\varepsilon_{GEN}, \varepsilon_{REC}$	1
$\varepsilon_{TRG}, \varepsilon_{OFF}, \varepsilon_{BDTs}, \varepsilon_{FR}$	2
ε_{PID}	3

Table 22: Method of obtaining efficiencies. Most of these efficiencies are evaluated using simulation, however, TRG and PID efficiencies are evaluated using data and/or simulation techniques.

847 The three methods for signal efficiency determination are listed:

- 848 • Method 1 - The first two efficiencies, $\varepsilon_{GEN}, \varepsilon_{REC}$, for signal are obtained using privately
849 generated simulation from Table 3 using

$$\varepsilon_{GEN,minq} \times \varepsilon_{REC,minq} = \frac{N_{in_acc,minq}}{N_{generated,minq}} \times \frac{N_{REC,minq}}{N_{in_acc,minq}}, \quad (18)$$

$$N_{in_acc,minq} = N_{in_acc} \times \varepsilon_{minq}. \quad (19)$$

850 In Equation 50, ε_{minq} is obtained by dividing number of generated events in *generator-level*
851 simulation (mentioned in Table 3) with *minq* condition imposed, $N_{generated,minq}$, to total
852 number of generated events, $N_{generated}$. N_{in_acc} is the number of events in *generator-level+detector*
853 simulation before reconstruction, $N_{REC,minq}$ is the number of events after
854 reconstruction with *minq* condition.

- 855 • Method 2 - Divide number of events that passed the selection by total number of events
856 prior to this particular selection step.
- 857 • Method 3 - Data-driven approach using **PIDCalib** package of determining PID efficiency is
858 used. Using **PIDCalib** efficiencies in this case is valid way to obtain efficiencies because
859 here ID efficiencies are needed (these are obtained from $J/\psi \rightarrow \mu^+\mu^-$ in **PIDCalib**) rather
860 than misID efficiencies.

861 Values for different efficiencies and more information about individual efficiencies are discussed
862 below.

863 7.2 Detector Acceptance Efficiencies (GEN)

864 As LHCb detector is forward-arm-spectrometer, it accepts event only in certain angular region.
865 For charged particles detector acceptance efficiency describes the fraction of decays contained
866 in the polar angle region of [10, 400] mrad. For neutral particles, the corresponding angular
867 region is [0, 400] mrad. These efficiencies were obtained by generating statistics tables , looking
868 at generator level cut efficiency. The summary of these efficiencies are shown in Table 23.

Channel	Year	Conditions	ε_{up}	ε_{down}	ε_{gen}
$B^+ \rightarrow J/\psi K^+$	2011	Sim08, Pyth6	0.15511 ± 0.00029	0.15489 ± 0.00029	0.15921 ± 0.00015
$B^+ \rightarrow J/\psi K^+$	2011	Sim08, Pyth8	0.16379 ± 0.00030	0.16366 ± 0.00030	
$B^+ \rightarrow J/\psi K^+$	2012	Sim08, Pyth6	0.15723 ± 0.00047	0.15776 ± 0.00047	0.16216 ± 0.00024
$B^+ \rightarrow J/\psi K^+$	2012	Sim08, Pyth8	0.16734 ± 0.00049	0.16712 ± 0.00049	
$B^+ \rightarrow J/\psi K^+$	2016	Sim09b, Pyth8	0.17380 ± 0.00063	0.17402 ± 0.00063	0.1739 ± 0.0004
$B^+ \rightarrow J/\psi \pi^+$	2012	Sim08, Pyth6	0.1519 ± 0.000400	N/A	
$B^+ \rightarrow J/\psi \pi^+$	2012	Sim08, Pyth8	0.1622 ± 0.000424	0.1611 ± 0.000421	0.15816 ± 0.00024
$B^0 \rightarrow J/\psi K^*$	2012	Sim08, Pyth8	0.16141 ± 0.00043	0.16050 ± 0.00043	0.16095 ± 0.00030
PartReco subsubsection 3.2.4	2012	Sim08, Pyth8	0.16058 ± 0.00051	0.16058 ± 0.00050	0.1606 ± 0.0004

Table 23: Geometrical detector acceptance-generator level efficiencies. For 2011 and 2012 simulation samples, the overall efficiency will be the average for MagDown, MagUp and Pythia 6 and Pythia 8 conditions. This is because between Pythia 6 and 8 and MagDown and MagUp each number of simulation events accounts to 1/4 of the simulation statistics.

869 Privately generated simulation is used to determine detector acceptance and reconstruction
 870 efficiency with Method 1 in subsection 7.1. The reason for usage of privately generated generator
 871 level samples is that official full samples yield generator level efficiencies in full $minq$ region.
 872 However, in this case what is necessary is the efficiency of generator level cuts **given** the $minq$
 873 selection as this measurement is made in $minq < 980 \text{ MeV}/c^2$. Summary of efficiencies from the
 874 privately generated tables given the $minq$ selection is summarized in Table 24.

Channel	Year	Conditions	ε_{gen}	Average
$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$	2012	Sim08, Pyth6	0.1828 ± 0.0015	0.1856 ± 0.0011
$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$	2012	Sim08, Pyth8	0.1884 ± 0.0015	
$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$	2016	Sim09b, Pyth8	0.1959 ± 0.0016	0.1959 ± 0.0016

Table 24: Geometrical detector acceptance-generator level efficiency, estimated as the fraction of decays contained in the polar angle region of [10, 400] mrad. These numbers were privately generated with 25000 events and $minq$ selection is applied beforehand.

875 7.3 Reconstruction (REC) Efficiency

876 The reconstruction efficiency is calculated on simulated events which have passed the detector
 877 acceptance. For signal, this efficiency consists of stripping. For normalisation it consists from
 878 stripping **and on the top** signal stripping is applied. This is done so that selections in
 879 normalisation and signal channel are kept as similar as possible and the fact that the signal
 880 selection has tighter cuts. This is reflected in definition of reconstruction efficiency seen in
 881 Equation 20. The reconstruction efficiency here reflects stripping selection **without the PID**
 882 **cuts**, in order to avoid applying PID in the simulation (as MC is known to be badly modelled in
 883 simulation). The stripping with no PID requirements is listed in Table 25. For reconstruction
 884 efficiency, candidates' tracks are also required to be **truth-matched offline**. For more details

885 about **truth-matching** see Appendix E.

$$\frac{\varepsilon_s^{REC}}{\varepsilon_n^{REC}} = \frac{\varepsilon_s^{stripping}}{\varepsilon_n^{stripping}} \times \frac{1}{\varepsilon_{signalstripping}} \quad (20)$$

886 Summary of REC efficiencies for normalisation channel in different years of data-taking can
887 be seen in Table 26.

Muon Cuts	Mother Cut
$\chi^2_{minIP} > 9$	$2500 \text{ MeV}/c^2 < M_{B_{corr}} < 10000 \text{ MeV}/c^2$
$\chi^2_{tr}/dof < 3$	$\cos(\theta_B) > 0.999$
$P_{ghost} < 0.35$	$p_T > 2000 \text{ MeV}/c$
	$\chi^2_{FD} > 50$
	$\chi^2/dof < 4$
	$0 \text{ MeV}/c^2 < M_B < 7500 \text{ MeV}/c^2$

Table 25: Stripping selection for the MC signal decay $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu_\mu$ with no PID requirements.

Channel	Year	Conditions	$N_genInAcc$	N_RecSel	ε_{REC}
$B^+ \rightarrow J/\psi K^+$	2011	Sim08, Pyth6+8	4074487.0	782018.0	0.19193 ± 0.00020
$B^+ \rightarrow J/\psi K^+$	2012	Sim08, Pyth6+8	8043820.0	1427090.0	0.17741 ± 0.00013
$B^+ \rightarrow J/\psi K^+$	2016	Sim09b, Pyth8	13231591.0	2650397.0	0.20031 ± 0.00011

Table 26: Reconstruction and preselection efficiencies.

888 As previously the summary of privately generated tables given the $minq$ selection summarized
889 in Table 27.

Channel	Year	Conditions	ε_{REC}
$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$	2012	Sim08, Pyth6+8	0.10841 ± 0.00030
$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$	2016	Sim09b, Pyth8	0.12417 ± 0.00032

Table 27: Reconstruction and preselection efficiencies where these numbers were privately generated with 25000 events and $minq$ cut is applied beforehand. The 2011 efficiencies are assumed to be the same as 2012.

890 7.4 Trigger Efficiency (TRG)

891 The trigger efficiency is calculated on the top of (GEN) and (REC) efficiency. In order to extract
892 the trigger efficiency, MC for both signal and normalisation was used. This approach is valid if
893 default MC setting is representative of the datataking for the year. This is true for Stripping
894 21 and 21r1 data. However in 2016 trigger thresholds have been changing often resulting in
895 16 different configuration keys with very different thresholds, see Table 28. The default MC in

TCK dec	TCK hex	% lumi	lumi [pb ⁻¹]	TrackMuon				SPD_{mult}	L0Muon $p_T^{max} \mu$ $\times 0.5 [\text{GeV}]$
				<i>GhostProb</i>	p %	p_T [GeV]	MIP χ^2		
2016 MD				0.859656 fb ⁻¹					
287905280	0x11291600	0.769	12.74	—	6.0	0.91	10	450	14
287905283	0x11291603	2.11	35.01	—	6.0	0.91	10	450	23
287905284	0x11291604	1.50	24.78	—	6.0	0.91	10	450	27
287905285	0x11291605	4.73	78.42	—	6.0	0.91	10	450	31
288822793	0x11371609	4.35	72.14	0.2	6.0	1.1	35	450	27
288822798	0x1137160e	1.37	22.756	0.2	6.0	1.1	35	450	27
288888329	0x11381609	0.414	6.86	0.2	6.0	1.1	35	450	31
288888334	0x1138160e	1.912	31.70	0.2	6.0	1.1	35	450	31
288888335	0x1138160f	34.7	575.25	0.2	6.0	1.1	35	450	37
2016 MU				0.798156 fb ⁻¹					
288495113	0x11321609	6.45	107.00	—	6.0	0.91	10	450	27
288626185	0x11341609	7.12	118.06	—	6.0	0.91	10	450	27
288691721	0x11351609	1.42	23.46	0.2	6.0	1.1	35	450	27
288757257	0x11361609	25.0	414.62	0.2	6.0	1.1	35	450	27
288888337	0x11381611	2.66	44.13	0.2	6.0	1.1	35	450	31
288888338	0x11381612	5.41	89.75	0.2	6.0	1.1	35	450	33
288888339	0x11381613	0.0685	1.136	0.2	6.0	1.1	35	450	27
MC Default									
1362630159	0x5138160f	—	—	0.2	6.0	1.1	35	450	37

Table 28: Summary of 16 different trigger configurations listing properties of candidates necessary to pass L0 and HLT1 selection. In the final row, the default configuration for 2016 can be seen. The reason for common multiplying factor for L0 is the conversion from A/D count to GeV which is 0.5 or to MeV 50.

896 2016 only represents around 35% of the data. This can be seen by comparing the default MC
 897 conditions for L0 and HLT1, which are the same requirements as TCK 288888335. The rest is
 898 taken with very different thresholds. For this reason, the trigger efficiencies for 2016 data have
 899 been obtained by emulation of the trigger on MC for L0 and HLT1 level. For the HLT2 level, there
 900 was no change of thresholds and hence these will be taken from MC regardless.

901 As it can be seen in the list of trigger lines in Table 29, `L0MuonDecision` needs to be modelled.
 902 This trigger line selects event candidates only if candidate has certain muon p_T and $nSPD$ hits.
 903 This information is accessible via the `TupleToolL0Data` tool, which code can be accessed [here](#)⁹.
 904 This tool takes MC sample and adds all necessary branches for L0 trigger to make decisions,
 905 such as p_T of three highest p_T muons in the event, dimuon p_T , hadron transverse energy E_T ,
 906 electron E_T , SPD multiplicity and so on.

⁹Source code is accessed at

<https://gitlab.cern.ch/lhcb/Analysis/blob/9f7a8759f10ee4967e375c4452bb0a6a9fc41d17/Phys/DecayTreeTuple/src/TupleToolL0D>

Trigger selection line
L0: Bplus_L0MuonDecision_TOS and
HLT1: Bplus_Hlt1TrackMuonDecision_TOS and
HLT2: or of
Bplus_Hlt2TopoMu2BodyBBDTDecision_TOS
Bplus_Hlt2TopoMu3BodyBBDTDecision_TOS
Bplus_Hlt2DiMuonDetachedDecision_TOS
Bplus_Hlt2DiMuonDetachedHeavyDecision_TOS

Table 29: Trigger Selection that needs to be emulated by MC.

907 HLT1 selection was also emulated offline as HLT1 requires an L0 requirement which would
 908 be the default case and hence is L0-dependent. The efficiency of HLT1TrackMuonDecision
 909 is then determined on the top of L0MuonDecision as only events which have passed either
 910 L0MuonDecision or L0DimuonDecision would be considered. HLT1TrackMuonDecision trigger
 911 lines requires events with only certain muon p_T, p , *ghost probability* and χ^2_{minIP} (all defined
 912 in subsection 4.2). There are other variables that are included in HLT1 filter such as whether
 913 the track is Velo track or how many hits have been missed in Velo, however, these have
 914 not changed across the Stripping 26 and are not likely to be different between signal and
 915 normalisation channel and hence only efficiency for the relevant cuts are included in emulation of
 916 HLT1TrackMuonDecision. It should be noted that there is difference between *offline* and HLT1
 917 container for PVs which stores the information about χ^2_{minIP} fast Kalman fitter rather than full
 918 is used. Therefore the emulation will not be exact, however, the effect is very small as it seems
 919 to be only resolution that will slightly shift the distribution, see Figure 30.

920 This trigger emulation to extract efficiencies was tested with the default trigger configuration
 921 to validate the emulation and the correct efficiencies have been recovered. In more detail, using
 922 default MC for signal for 2016, before the L0 emulation there were 785 132 candidates and after
 923 emulation: 117402. If instead just cut-based approach (cutting on L0MuonDecision_dec) was
 924 used 117402 events survived the selection. Using default MC for normalisation for 2016, before
 925 the L0 emulation there were 4 114 418 candidates and after emulation: 3 255 259. If instead just
 926 cut-based approach was used 3 255 259 events survived the selection.

927 TCK dependent efficiency breakdown for signal and normalisation channel can be seen
 928 in Table 30. In order to obtain the average efficiency for Stripping 26, these efficiencies are
 929 weighted by the % of lumi for which this luminosity was ran on. These numbers have been
 930 obtained by looking at API version of the rundatabase where one can obtain luminosity per
 931 TCK.

932 Stripping 21 and stripping 21r1 trigger efficiency are determined directly by looking at default
 933 TCK and are summarized in Table 31. It should be noted that in the rest of the efficiencies
 934 Stripping 21 will be representative of 21+21r1 dataset as it was noticed that for normalisation
 935 channel these efficiencies are equivalent. In this section following ratio will be calculated,

$$\frac{\varepsilon_s^{TRG}}{\varepsilon_n^{TRG}} = \frac{\varepsilon_s^{L0}}{\varepsilon_n^{L0}} \times \frac{\varepsilon_s^{HLT1}}{\varepsilon_n^{HLT1}} \times \frac{\varepsilon_s^{HLT2}}{\varepsilon_n^{HLT2}} \quad (21)$$

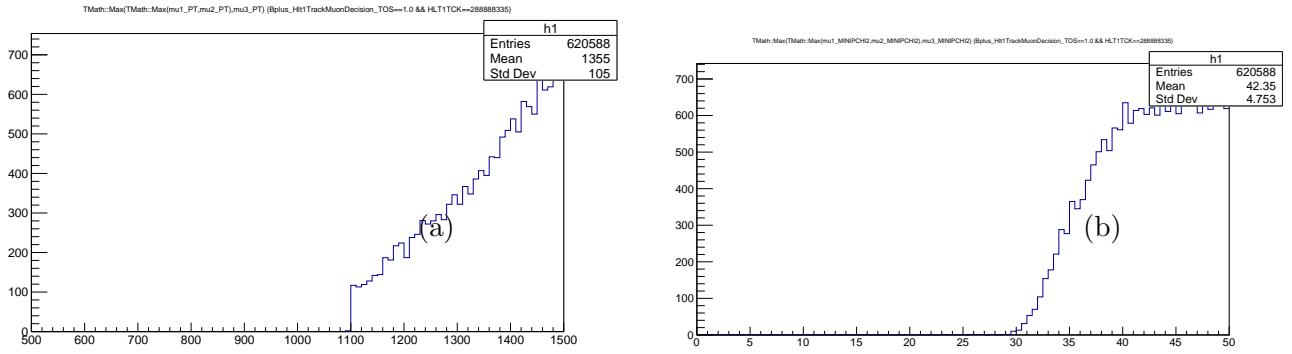


Figure 30: On the left p_T of $\max(\mu^+, \mu^-, \mu^+)$ at 0x1138160f trigger configuration after HLT1_TOS was applied. There is a sharp cut at 1.1 GeV as expected, requiring HLT1TrackMuonDecision_TOS (see Table 28 for expected thresholds). On the right χ^2_{minIP} does not mirror the expected cut of $\chi^2_{minIP} > 35$. This can be explained by the fact that there is difference between *offline* and HLT1 container for PVs. HLT1 container stores the information about χ^2_{minIP} with fast Kalman fitter, whereas *offline* uses full Kalman fitter.

TCK	$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$			$B^+ \rightarrow J/\psi K^+$		
	ε_{L0}	ε_{HLT1}	ε_{HLT2}	ε_{L0}	ε_{HLT1}	ε_{HLT2}
287905280	0.921	0.999	0.831	0.891	0.997	0.943
287905283	0.905	0.999	0.845	0.878	0.998	0.953
287905284	0.894	0.999	0.855	0.867	0.998	0.962
287905285	0.88	0.999	0.868	0.854	0.998	0.973
288495113	0.894	0.999	0.855	0.867	0.998	0.962
288626185	0.894	0.999	0.855	0.867	0.998	0.962
288691721	0.894	0.957	0.873	0.867	0.94	0.965
288757257	0.894	0.957	0.873	0.867	0.94	0.965
288822793	0.894	0.957	0.873	0.867	0.94	0.965
288822798	0.88	0.957	0.886	0.854	0.941	0.976
288888329	0.894	0.957	0.873	0.867	0.94	0.965
288888334	0.88	0.957	0.886	0.854	0.941	0.976
288888335	0.848	0.958	0.911	0.821	0.941	0.999
288888337	0.88	0.957	0.886	0.854	0.941	0.976
288888338	0.871	0.957	0.895	0.844	0.941	0.984
288888339	0.89	0.957	0.877	0.864	0.94	0.968
Weighted efficiency	0.876	0.967	0.884	0.849	0.953	0.978

Table 30: Efficiencies of 2016 trigger emulation on MC. Depending on TCK, the efficiencies vary up 10% for L0 level for signal MC and up to 5% for normalisation TCK. This is important as *single event sensitivity* is sensitive to the ratio of these two efficiencies. This configuration is describing correctly only 35% data with high p_T threshold.

Efficiency	$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ (2012)	$B^+ \rightarrow J/\psi K^+$ (2012)	$B^+ \rightarrow J/\psi K^+$ (2011)
ε_{L0}	0.900	0.873	0.907
ε_{HLT1}	0.934	0.908	0.879
ε_{HLT2}	0.883	0.981	0.973
ε_{TRG}	0.742	0.775	0.778

Table 31: 2012 default TCK efficiencies for $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ and $B^+ \rightarrow J/\psi K^+$ and 2011 default TCK $B^+ \rightarrow J/\psi K^+$ efficiencies. It can be noticed that the $B^+ \rightarrow J/\psi K^+$ values of cumulative efficiency ε_{TRG} for 2011 and 2012 are nearly identical.

936 7.5 Offline Selection (OFF)

937 In this section offline efficiencies will be discussed. These include J/ψ and $\Psi(2S)$ veto signal
 938 efficiency that were mentioned in pre-selection section in Table 7, where $2946.0 \text{ MeV}/c^2 < |M(\mu^+ \mu^-)| < 3176.0 \text{ MeV}/c^2$ for J/ψ veto and $3586 \text{ MeV}/c^2 < |M(\mu^+ \mu^-)| < 3766 \text{ MeV}/c^2$ for $\Psi(2S)$
 939 veto. For normalisation channel this is non applicable, and hence will not be included in the
 940 efficiency ratio. As trigger efficiencies for Stripping 26 will be TCK-dependant all the Stripping
 941 26 veto efficiencies are TCK-weighted averages. The summary for Stripping 21 is in Table 32 and
 942 for Stripping 26 in Table 33. Dependence upon decay model for the veto efficiencies is discussed
 943 in subsection 10.2.

$$\frac{\varepsilon_s^{OFF}}{\varepsilon_n^{OFF}} = \frac{\varepsilon_s^{cc}}{1} \quad (22)$$

Efficiency	$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$
ε^{OFF}	0.882

Table 32: Efficiency of J/ψ and $\Psi(2S)$ veto selections in Stripping 21.

Weighted Efficiency	$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$	$B^+ \rightarrow J/\psi K^+$
ε^{OFF}	0.883	N/A

Table 33: Efficiency of J/ψ and $\Psi(2S)$ veto selections in Stripping 26. This was naturally only applied to the signal channel as the normalisation channel is decay including intermediate J/ψ resonance.

945 7.6 Combinatorial BDT and Misid BDT Efficiencies

946 Efficiencies of MVA selection are also evaluated on simulation samples. These efficiencies are
 947 obtained using samples that passed (GEN),(REC) and (OFF) cuts. Specific MVA for combi-
 948 natorial background suppression (see subsubsection 4.5.2) and misID background suppression
 949 (subsubsection 4.5.3) are applied to the MC samples. As the optimisation led to different
 950 BDTs between Stripping versions, these are then applied parametrically to relevant Stripping

samples. So the Combinatorial BDT trained on Run 1 sample is applied to Run 1 samples and the Combinatorial BDT for 2016 is applied to 2016 samples. The results are listed in Tables 34 for Stripping 21 and 35 for Stripping 26.

In both Misid and Combinatorial BDT selection, normalisation $B^+ \rightarrow J/\psi K^+$ channel retains more signal, or in other words, the BDT selection efficiency for the normalisation is higher. This can only happens if normalisation channel important variables are more signal-like than signal itself as compared to background considered. To put it differently, higher BDT normalisation efficiency can be obtained if normalisation distribution for that variable is even further separated from the background distribution than the signal distribution. This is true for the kaon/muon p_T and p_T kinematics difference as seen in Figure 31(c)(d) and Figure 32(d), where the kaon track is generally harder than the muon track. Moreover as kaon reconstruction efficiency is worse than muon reconstruction efficiency because about 11% of the kaons cannot be reconstructed due to hadronic interactions that occur before the last T station, this implies that the p_T of B is on average harder for normalisation channel as seen in Figure 31(b) and Figure 32(b). As these two quantities are high in Combinatorial BDT importance ranking, this makes normalisation MC more efficient. In other words, the Combinatorial BDT assigns to the event BDT score with highest emphasis/weight on these two distributions. As the normalisation channel is even further separated to the background, it is more likely to classify the events as signal, i.e being more efficient. This can be seen in Figure 31(a). In Misid BDT selection, the IP χ^2 of the oppositely charged muon to B will come from D and hence will be also making it more signal-like than signal itself, see Figure 32(c). Again this will make normalisation retain more of its candidates, shown in Figure 32(a).

In the following Combinatorial BDT plots in Figure 31, μ^1 and μ^3 are the muons with the charge that agrees with mother B , so if the mother is B^+ then μ^1 and μ^3 are the positively charged muons (μ^+), and μ^2 is negatively charged muon (μ^-).

In the following Misid BDT plots in Figure 32, for signal samples μ^1 and μ^3 are the muons with the charge that agrees with mother B , so if the mother is B^+ then μ^1 and μ^3 are the positively charged muons (μ^+), and μ^2 is negatively charged muon (μ^-). For misID SS data samples, μ^1 is true muon with the charge that agrees with the charge of mother B , μ^2 is true muon with the charge that does not agree with the charge mother B and finally μ^3 is the misidentified particle whose charge agrees with the charge of mother B .

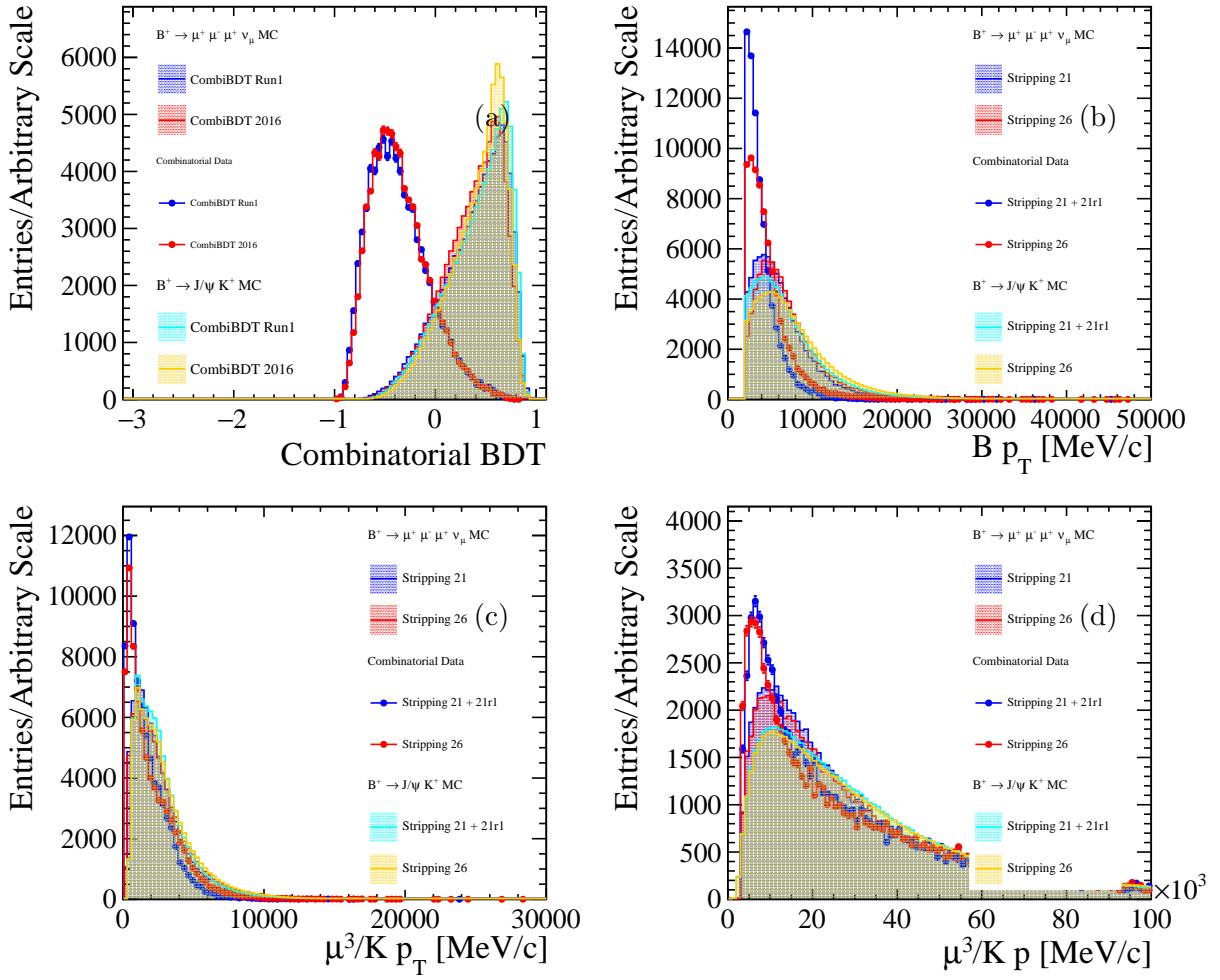


Figure 31: (a) Combinatorial BDT response for signal MC and upper mass sideband as well as for normalisation channel MC for Stripping 21 and Stripping 26. The most discriminative variables are (b) p_T of B , (c) muon/kaon p_T and (d) muon/kaon p

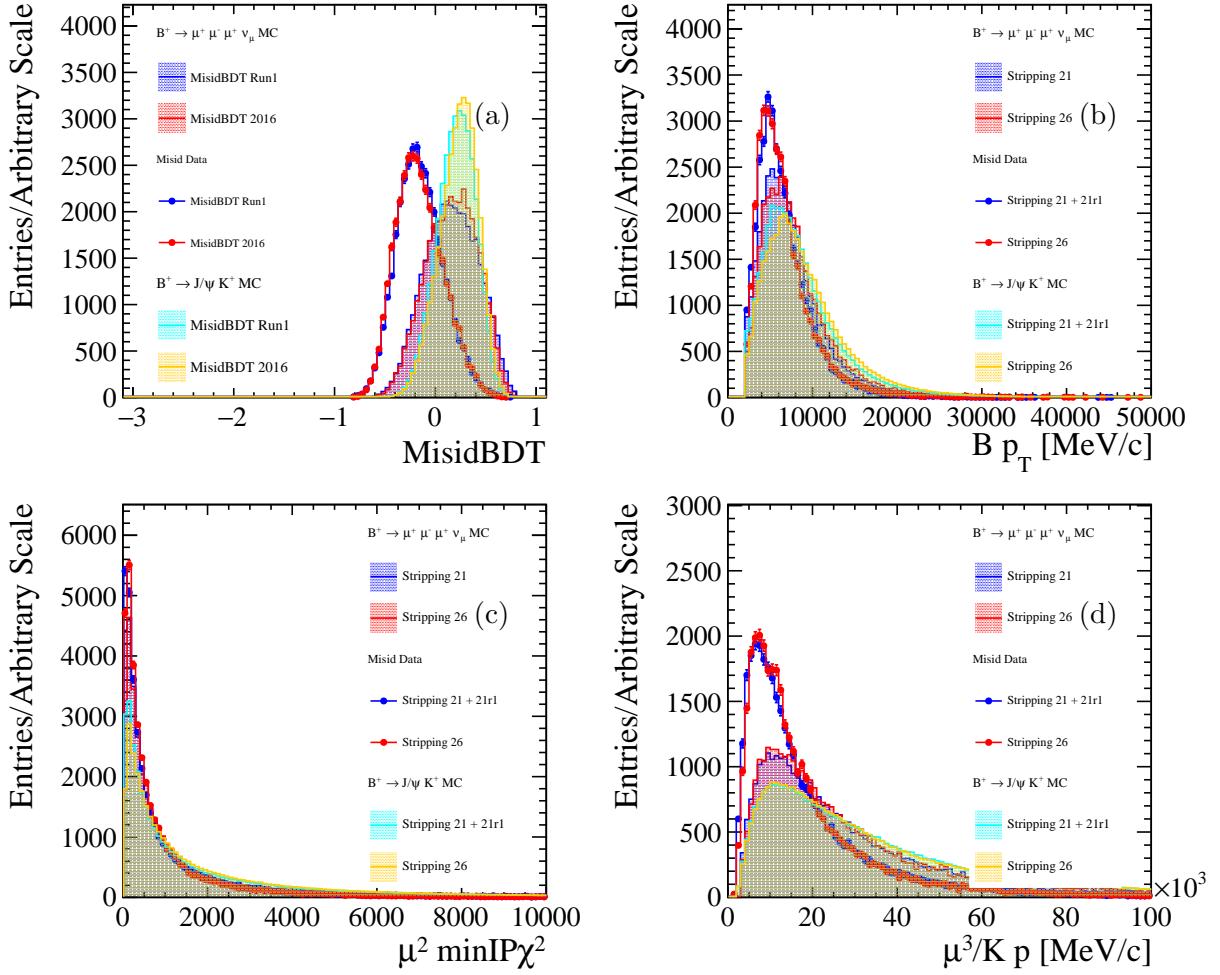


Figure 32: (a) Misid BDT response for signal MC and upper mass sideband as well as for normalisation channel MC for Stripping 21 and Stripping 26. The most discriminative variables are (b) p_T of B, (c) muon IP χ^2 and (d) muon/kaon p . Misid BDT responses are plotted with combinatorial BDT already applied.

Efficiency	$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$	$B^+ \rightarrow J/\psi K^+$
$\varepsilon_{\text{CombiBDT}}$	0.473	0.509
$\varepsilon_{\text{MisidBDT}}$	0.436	0.511

Table 34: 2012 Combinatorial and Misid BDT selection efficiency.

Weighted Efficiency	$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$	$B^+ \rightarrow J/\psi K^+$
$\varepsilon_{CombiBDT}$	0.343	0.397
$\varepsilon_{MisidBDT}$	0.368	0.446

Table 35: 2016 Combinatorial BDT selection efficiency. These MVA selection were trained on events with 288888335 TCK. Here are the luminosity weighted average efficiencies. TCK-dependent efficiencies are listed in Table 76 in subsection F.1.

982 7.7 Fitting Range Efficiency (FR)

983 As discussed in Sections 4.6, 5.2, fitting region was chosen firstly, in order to avoid modelling
984 combinatorial background drop in low corrected mass region (exclusion below 4.0 GeV/c²) and
985 secondly in order to not include region where there are very few/no events (exclusion above 7.0
986 GeV/c²) in **corrected mass**. As seen in Tables 36 and 37 normalisation channel does not
987 loose any candidates. This is expected as the **visible mass** is more constrained in normalisation
988 channel then in signal channel (see Figure 33).

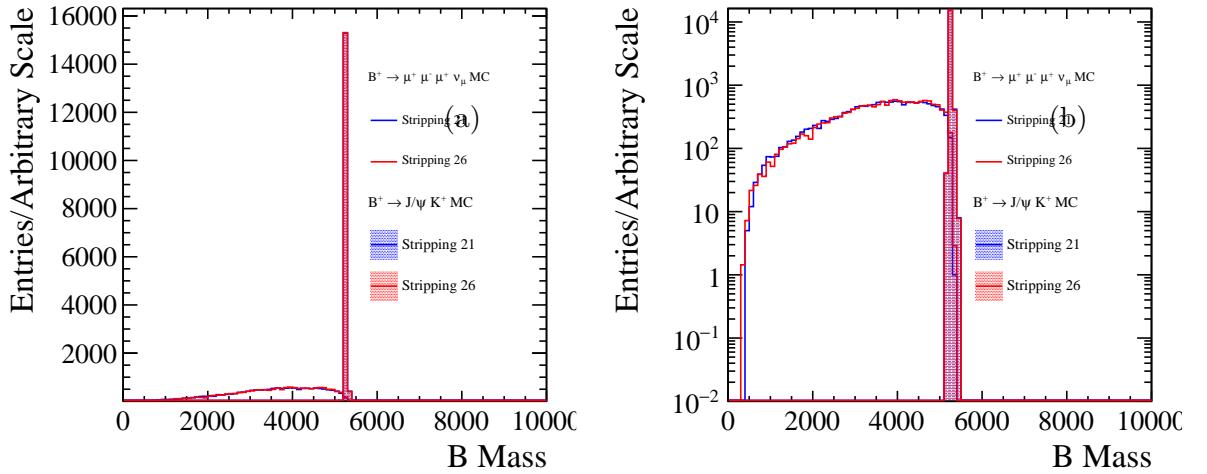


Figure 33: (a) Visible mass of normalisation and signal MC. It can be seen that normalisation previous selection has a sharp cut around B mass leading to much higher fitting region efficiency.(b) The corresponding log plot of (a).

Efficiency	$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$	$B^+ \rightarrow J/\psi K^+$
ε_{FR}	0.923	0.996

Table 36: 2012 Fitting Range efficiency.

Weighted Efficiency	$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$	$B^+ \rightarrow J/\psi K^+$
ε_{FR}	0.938	0.999

Table 37: Weighted 2016 Fitting Range efficiencies. TCK-dependent efficiencies are listed in Table 77 in subsection F.2.

989 7.8 Fractional Corrected Mass Split Efficiency

990 As discussed in Section 4.8, in order to increase sensitivity, there will be two bins of fractional
 991 corrected mass. The split was tuned to contain $\approx 50\%$ of the signal in $\sigma_{lowFCME}$ and $\approx 50\%$ of
 992 the signal in high $\sigma_{highFCME}$ in Run 1. Numerically this corresponds to $\delta/M_{B_{corr}} < 0.0225$ for
 993 $\sigma_{lowFCME}$ and $\delta/M_{B_{corr}} > 0.0225$ for $\sigma_{highFCME}$. This can be seen in Table 38 for 2012 simulation
 994 and in ?? for 2016 simulation. As normalisation channel has much higher visible mass, it will
 995 also have higher corrected mass, leading to bigger proportion of the normalisation channel in the
 996 $\sigma_{lowFCME}$ bin.

Efficiency	$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$	$B^+ \rightarrow J/\psi K^+$
$\varepsilon_{lowFCME}$	0.484	0.593
$\varepsilon_{highFCME}$	0.516	0.407

Table 38: 2012 High and Low Fractional Corrected Mass Error Split.

Weighted Efficiency	$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$	$B^+ \rightarrow J/\psi K^+$
$\varepsilon_{lowFCME}$	0.548	0.673
$\varepsilon_{highFCME}$	0.452	0.327

Table 39: Weighted fractional corrected mass efficiencies in Stripping 26 for the split. TCK-dependent efficiencies are listed in Table 78 in subsection F.3.

997 7.9 PID Efficiencies (PID)

998 As PID variables are not correctly modelled in MC, data-driven approach of extracting PID
 999 efficiency can be taken. To not introduce any biases in previous steps (especially in MVA
 1000 selection) PID efficiencies are evaluated at the end of the chain using `PIDCalib` package where
 1001 data samples can be used to extract this efficiency. Efficiency for signal and background are
 1002 summarized in Table 40 and Table 41 respectively.

species	2012 PID MC	2016 PID MC
muon	$\Delta LL(\mu - \pi) > 0$	$\Delta LL(\mu - \pi) > 0$
muon	$\Delta LL(\mu - K) > 0$	$\Delta LL(\mu - K) > 0$
muon	-	<code>IsMuonTight==1.0</code>
muon	<code>Nshared==0</code>	<code>Nshared<2</code>
muon	<code>Probnnmu> 0.35</code>	<code>Probnnmu> 0.35</code>
	0.631 ± 0.005	0.623 ± 0.006

Table 40: Signal MC efficiency using `PIDCalib` efficiencies.

species	2012 PID MC	2016 PID MC
muon	$\Delta LL(\mu - \pi) > 0$	$\Delta LL(\mu - \pi) > 0$
muon	$\Delta LL(\mu - K) > 0$	$\Delta LL(\mu - K) > 0$
muon	-	<code>IsMuonTight==1.0</code>
muon	<code>Nshared==0</code>	<code>Nshared<2</code>
muon	<code>Probnnmu> 0.35</code>	<code>Probnnmu> 0.35</code>
kaon	$\Delta LL(K - \pi) > 0$	$\Delta LL(K - \pi) > 0$
kaon	$\Delta LL(p - K) < 5$	$\Delta LL(p - K) < 5$
	0.6853 ± 0.0011	0.6563 ± 0.0010

Table 41: Normalisation MC efficiency using `PIDCalib` efficiencies.

1003 7.10 Summary of Individual MC and/or data efficiencies necessary for *Single* 1004 *Event Sensitivity*

1005 Different selection that was optimised for Run 1 and 2016 yield different efficiencies. The
1006 summary of these as a function of Stripping are summarized in the Figure 34 (no FCME splitting)
1007 and Table 42 (FCME splitting). Here offline selection (OFF) comprises of trigger efficiency, J/ψ
1008 and $\Psi(2S)$ veto efficiency. For 2016, it is the TCK weighted average efficiency.

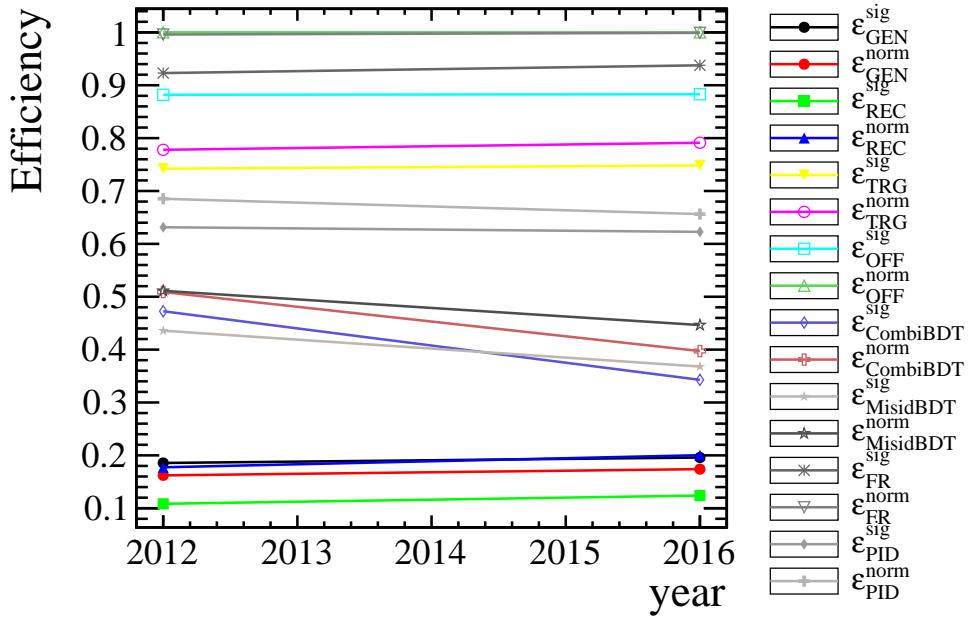


Figure 34: Summary of Individual MC and/or data efficiencies necessary for *Single Event Sensitivity* for signal (*sig*) and normalisation channel (*norm*) with no FCME split. Efficiency values for 2016 are TCK-weighted averaged efficiencies.

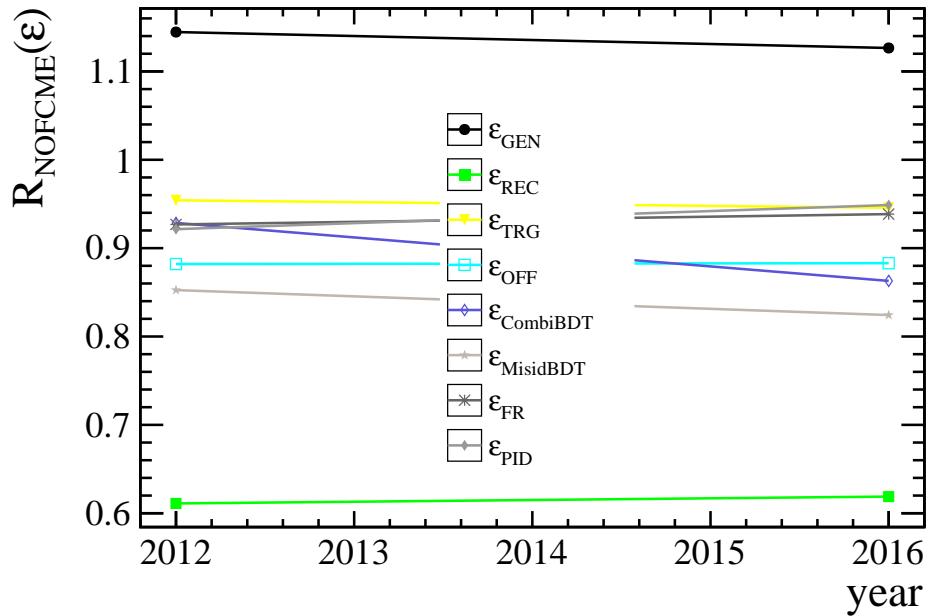


Figure 35: Summary of ratio of efficiencies between 2012 MC and 2016 MC with no FCME split. Efficiency values for 2016 are TCK-weighted averaged efficiencies.

Efficiency	Low FCME		High FCME	
	2012	2016	2012	2016
ε_{GEN}^{sig}	0.18560 ± 0.00110	0.19590 ± 0.00072	0.18560 ± 0.00110	0.19590 ± 0.00072
ε_{GEN}^{norm}	0.16216 ± 0.00024	0.17390 ± 0.00018	0.16216 ± 0.00024	0.17390 ± 0.00018
ε_{REC}^{sig}	0.10841 ± 0.00030	0.12417 ± 0.00014	0.10841 ± 0.00030	0.12417 ± 0.00014
ε_{REC}^{norm}	0.17741 ± 0.00013	0.20031 ± 0.00005	0.17741 ± 0.00013	0.20031 ± 0.00005
ε_{TRG}^{sig}	0.74224 ± 0.00128	0.74829 ± 0.00054	0.74224 ± 0.00128	0.74829 ± 0.00054
ε_{TRG}^{norm}	0.77790 ± 0.00035	0.79125 ± 0.00011	0.77790 ± 0.00035	0.79125 ± 0.00011
ε_{OFF}^{sig}	0.88199 ± 0.00110	0.88299 ± 0.00046	0.88199 ± 0.00110	0.88299 ± 0.00046
ε_{OFF}^{norm}	1.00000 ± 0.00000	1.00000 ± 0.00000	1.00000 ± 0.00000	1.00000 ± 0.00000
$\varepsilon_{CombiBDT}^{sig}$	0.47254 ± 0.00180	0.34281 ± 0.00072	0.47254 ± 0.00180	0.34281 ± 0.00072
$\varepsilon_{CombiBDT}^{norm}$	0.50892 ± 0.00047	0.39727 ± 0.00015	0.50892 ± 0.00047	0.39727 ± 0.00015
$\varepsilon_{MisidBDT}^{sig}$	0.43581 ± 0.00261	0.36797 ± 0.00125	0.43581 ± 0.00261	0.36797 ± 0.00125
$\varepsilon_{MisidBDT}^{norm}$	0.51121 ± 0.00067	0.44642 ± 0.00025	0.51121 ± 0.00067	0.44642 ± 0.00025
ε_{FR}^{sig}	0.92303 ± 0.00212	0.93769 ± 0.00103	0.92303 ± 0.00212	0.93769 ± 0.00103
ε_{FR}^{norm}	0.99592 ± 0.00012	0.99912 ± 0.00002	0.99592 ± 0.00012	0.99912 ± 0.00002
ε_{FCME}^{sig}	0.48350 ± 0.00414	0.54812 ± 0.00220	0.51650 ± 0.00414	0.45188 ± 0.00220
$\varepsilon_{FCME}^{norm}$	0.59321 ± 0.00092	0.67311 ± 0.00035	0.40679 ± 0.00092	0.32689 ± 0.00035
ε_{PID}^{sig}	0.61463 ± 0.00500	0.62455 ± 0.00269	0.64720 ± 0.00500	0.62041 ± 0.00269
ε_{PID}^{norm}	0.67695 ± 0.00110	0.65421 ± 0.00045	0.69740 ± 0.00110	0.66057 ± 0.00045
$\varepsilon_{total}^{sig}$	0.000744 ± 0.000012	0.000651 ± 0.000005	0.000837 ± 0.000013	0.000533 ± 0.000005
$\varepsilon_{total}^{norm}$	0.002329 ± 0.000008	0.002149 ± 0.000003	0.001645 ± 0.000006	0.001054 ± 0.000002

Table 42: Summary of Individual MC and/or data efficiencies necessary for *Single Event Sensitivity* for signal (s) and normalisation channel (n) with FCME split. Efficiency values for 2016 are TCK-weighted averaged efficiencies.

Hence resulting values for relative no fractional corrected mass split efficiency ratios defined in Equation 16 (ratios of total signal efficiency to normalisation efficiency) are

$$\begin{aligned}
R_{\text{NOFCME}}^{21}(\varepsilon) &= \frac{(1.58 \pm 0.02) \times 10^{-3}}{(3.97 \pm 0.01) \times 10^{-3}} = (3.98 \pm 0.05) \times 10^{-1}, \\
R_{\text{NOFCME}}^{26}(\varepsilon) &= \frac{(1.18 \pm 0.01) \times 10^{-3}}{(3.20 \pm 0.00) \times 10^{-3}} = (3.69 \pm 0.03) \times 10^{-1}.
\end{aligned} \tag{23}$$

and including fractional corrected mass split efficiency ratios defined in Equation 17

$$\begin{aligned}
R_{\text{lowFCME}}^{21}(\varepsilon) &= \frac{(7.44 \pm 0.12) \times 10^{-4}}{(2.33 \pm 0.01) \times 10^{-3}} = (3.20 \pm 0.05) \times 10^{-1}, \\
R_{\text{highFCME}}^{21}(\varepsilon) &= \frac{(8.37 \pm 0.13) \times 10^{-4}}{(1.65 \pm 0.01) \times 10^{-3}} = (5.09 \pm 0.08) \times 10^{-1}, \\
R_{\text{lowFCME}}^{26}(\varepsilon) &= \frac{(6.51 \pm 0.05) \times 10^{-4}}{(2.15 \pm 0.00) \times 10^{-3}} = (3.03 \pm 0.02) \times 10^{-1}, \\
R_{\text{highFCME}}^{26}(\varepsilon) &= \frac{(5.33 \pm 0.05) \times 10^{-4}}{(1.05 \pm 0.00) \times 10^{-3}} = (5.06 \pm 0.04) \times 10^{-1}.
\end{aligned} \tag{24}$$

1012 It can be seen that 2012 and 2016 efficiencies slightly differ, hence the sensitivity will be
1013 slightly different. The individual efficiency ratios can be seen Figure 35. It can be seen that
1014 difference can be attributed to nature of different BDT.

1015 These relative efficiency ratios represent relative efficiencies in bins of fractional corrected
1016 mass and will be used later for *Single Event Sensitivity*.

1017 **8 Normalisation channel**

1018 The normalisation channel used in this analysis is $B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+$ as it is clean,
1019 well understood, and well populated channel that is similar to signal. This means that many
1020 systematic uncertainties will cancel. Normalising the signal decay to this decay also means that
1021 absolute efficiencies, luminosity, the b-quark cross-section or fragmentation fractions will be
1022 cancelled. With the same number of tracks it will also give a small uncertainty in the tracking
1023 efficiency. One of the differences in selection compared to signal is that the kaon candidates are
1024 required to have PID criteria consistent with being a kaon. ϵ_{PID} is the efficiency of PID, that is
1025 taken from control data samples and then parametrically applied to simulation samples.

1026 **8.1 $B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+$ Fit**

1027 After all selection, PID weighted samples for Run 1 and 2016 are fitted in the following way. The
1028 shape for the signal is best described by double-sided Ipatia function, [14]:

$$I(m, \mu_{IP}, \sigma_{IP}, \lambda, \zeta, \beta, a_1, n_1, a_2, n_2) \quad (25)$$

1029 where all the parameters apart from μ_{IP}, σ_{IP} are fixed from MC. These MC have passed all of
1030 the selection for MC (as described in section 7). Since the PID requirement on kaon is very
1031 loose, there will background contribution from $B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)\pi^+$. This contribution is
1032 modelled by double-sided Crystal Ball function to $B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)\pi^+$ MC under kaon
1033 mass hypothesis. Again, all the parameters apart from μ_{CB}, σ_{CB} will be fixed from the fit of
1034 this sample. In Figure 36, fits to signal MC and misID MC from Stripping 21 are showed. For
1035 signal, Run 1 (Stripping 21) and 2016 (Stripping 26 - TCK 288888335) MC samples are used, for
1036 misID Stripping 21 is used for all samples. Combinatorial background is modelled by exponential
1037 function. The $B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+$ signal yield is extracted by performing an unbinned
1038 extended maximum likelihood fit to the invariant $\mu^+\mu^-K^+$ distribution in $5150 < M_{B^+} < 5450$.
1039 Fits to the $B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+$ for Run 1 and 2016 are shown in Figure 37.

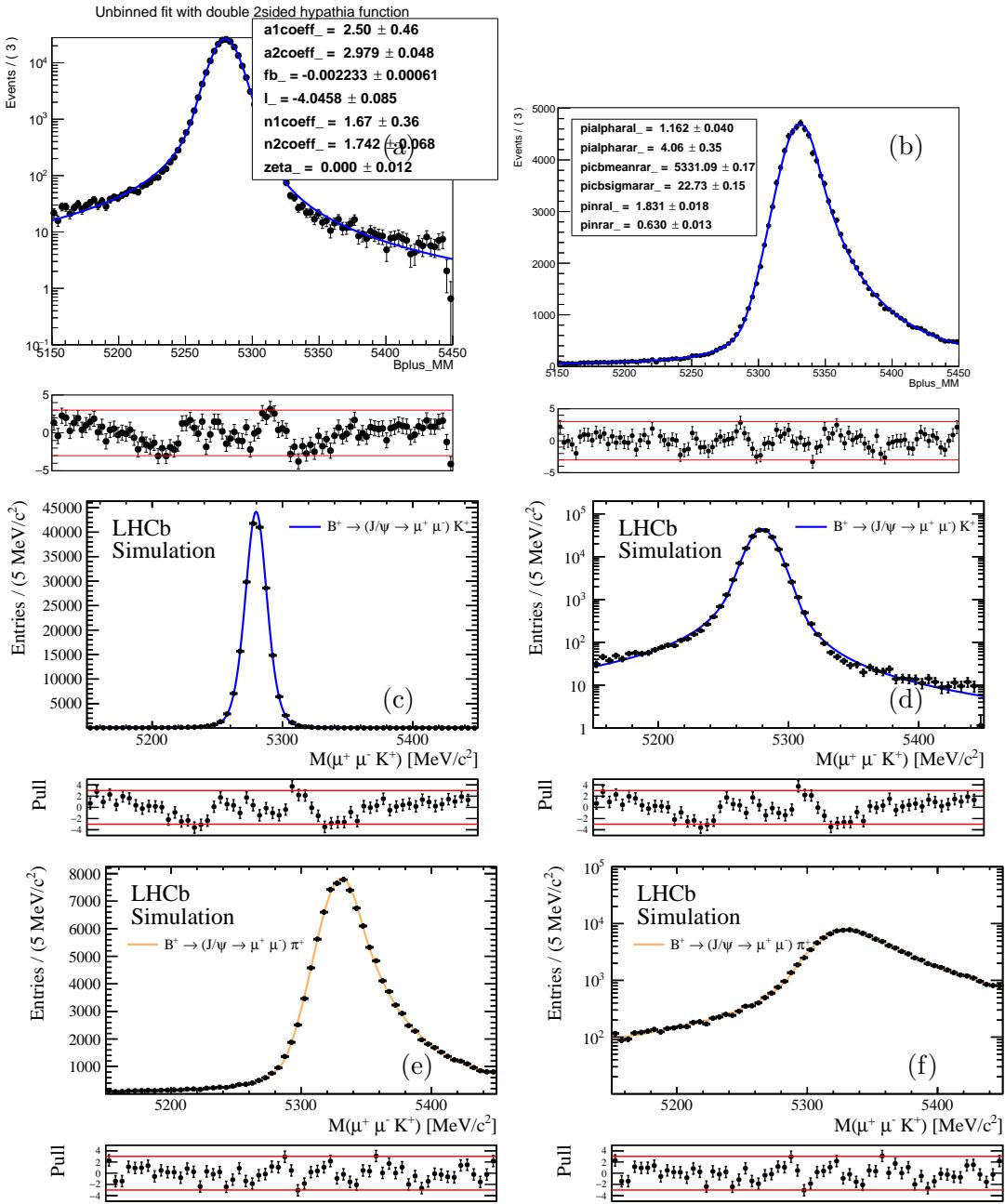


Figure 36: Fit to Stripping 21 (a) $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+$ MC and (b) $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) \pi^+$ MC under kaon mass hypothesis with parameters. Fit to 2012 (c) $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+$ simulation and (e) $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) \pi^+$ simulation under kaon mass hypothesis with no parameters listed. On right (d)(f), the same plots but on logarithmic scale instead.

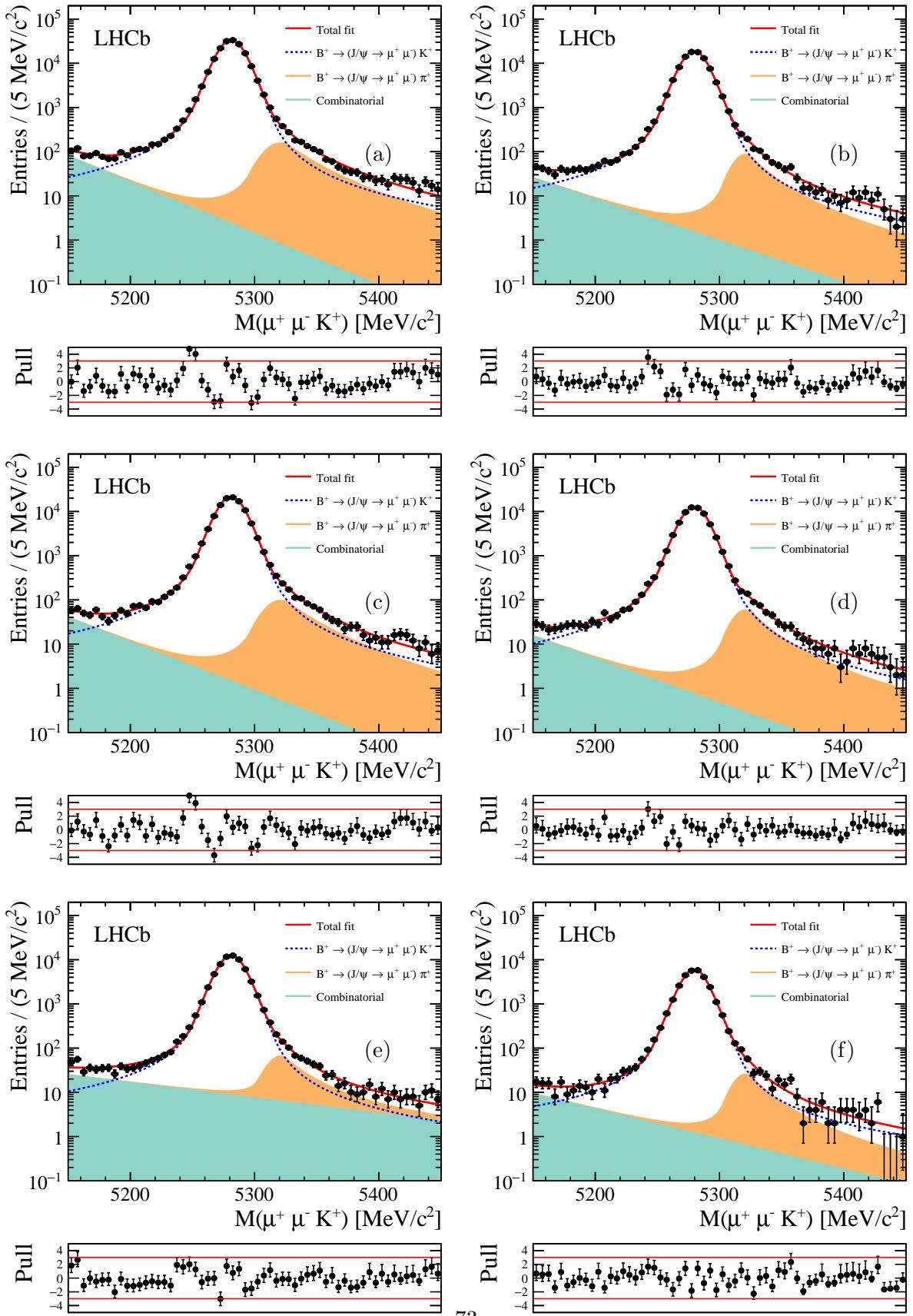


Figure 37: Fit results in logarithmic scale to (a) Run 1 (b) Run 2 $\mu^+ \mu^- K^+$ mass spectrum with no fractional corrected mass split, (c)(d) low FCME bin, (e)(f) high FCME bin.

1040 Yields from fit to $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+$ are obtained and summarized in Table 43.

Sample	Stripping	Split	Yields
$N_{B^+ \rightarrow J/\psi K^+}$	21+21r1	NOFCME	173422 ± 446.444
$N_{B^+ \rightarrow J/\psi K^+}$	26	NOFCME	94490.7 ± 313.926
$N_{B^+ \rightarrow J/\psi K^+}$	21+21r1	lowFCME	109224 ± 337.016
$N_{B^+ \rightarrow J/\psi K^+}$	26	lowFCME	64723.2 ± 259.027
$N_{B^+ \rightarrow J/\psi K^+}$	21+21r1	highFCME	64077.8 ± 257.3
$N_{B^+ \rightarrow J/\psi K^+}$	26	highFCME	29760.1 ± 176.067

Table 43: Signal Yield $B^+ \rightarrow J/\psi K^+$

1041 8.2 Single Event Sensitivity

1042 The branching fraction of $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$, normalized to $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+$, can be
1043 written

$$\mathcal{B}(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu) = \alpha \times N_{B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu} \quad (26)$$

$$1044 = \underbrace{\frac{\mathcal{B}(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+) \times \epsilon_{B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+}}{N_{(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+)} \times \epsilon_{B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu}}}_{\alpha} \times N_{B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu}, \quad (27)$$

1045 where $N_{(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+)}$ and $N_{B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu}$ are the observed yields of the normalisation and
1046 the signal channel. To obtain (*single event sensitivity*) α , which is necessary for setting limit
1047 procedure, yields from $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+$ are combined with efficiency ratios obtained
1048 in subsection 7.10. Together with $\mathcal{B}(B^+ \rightarrow J/\psi K^+) = (1.026 \pm 0.031) \times 10^{-3}$ [13] and $\mathcal{B}(J/\psi \rightarrow$
1049 $\mu^- \mu^+) = (5.961 \pm 0.0033) \times 10^{-2}$ [13] yielding $\mathcal{B}(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+) = 6.12 \pm 0.19) \times 10^{-5}$,

$$\alpha_{NOFCME}^{Run1} = \frac{\mathcal{B}(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+)}{N_{NOFCME}^{Run1}(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+) \times R_{NOFCME}^{21}(\epsilon)} = (8.86 \pm 0.297) \times 10^{-10} \quad (28)$$

$$\alpha_{NOFCME}^{26} = \frac{\mathcal{B}(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+)}{N_{NOFCME}^{26}(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+) \times R_{NOFCME}^{26}(\epsilon)} = (17.5 \times 0.557) \times 10^{-10} \quad (29)$$

$$\alpha_{lowFCME}^{Run1} = \frac{\mathcal{B}(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+)}{N_{lowFCME}^{Run1}(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+) \times R_{lowFCME}^{21}(\epsilon)} = (17.5 \pm 0.609) \times 10^{-10} \quad (30)$$

$$\alpha_{highFCME}^{Run1} = \frac{\mathcal{B}(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+)}{N_{highFCME}^{Run1}(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+) \times R_{highFCME}^{21}(\epsilon)} = (18.8 \pm 0.650) \times 10^{-10} \quad (31)$$

$$\alpha_{lowFCME}^{2016} = \frac{\mathcal{B}(B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+)}{N_{lowFCME}^{Run1}(B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+) \times R_{lowFCME}^{21}(\epsilon)} = (31.2 \pm 1.00) \times 10^{-10} \quad (32)$$

$$\alpha_{highFCME}^{2016} = \frac{\mathcal{B}(B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+)}{N_{highFCME}^{Run1}(B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+) \times R_{highFCME}^{21}(\epsilon)} = (40.6 \pm 1.32) \times 10^{-10} \quad (33)$$

1051 8.3 Efficiency Corrected Yields

1052 Using integrated luminosity collected for Run 1 and 2016, it is possible to obtain efficiency
 1053 corrected yield per pb^{-1} as seen in Table 44. It can be seen that overall selection efficiency in
 1054 2016 drops by 1/4 as compared to Run 1.

1055 From increase in collision energy (13 TeV in 2016 as compared to 7.8 TeV in Run 1) it is
 1056 expected that efficiency corrected yield per pb^{-1} mirrors the increase in collision energy. In this
 1057 case, as seen in Table 44 the efficiency corrected yield per pb^{-1} in 2016 is low.

Summary	Run 1	2016
$B^+ \rightarrow J/\psi K^+$ events	173422	94490.7
ε_s	0.001581	0.001182
ε_n	0.003974	0.003203
R_{NOFCME}	0.397836	0.369029
Luminosity \mathcal{L}	2968 pb^{-1}	1612 pb^{-1}
$B^+ \rightarrow J/\psi K^+$ events per pb^{-1}	58.4308	58.6171
Efficiency corrected yield per pb^{-1}	14703.3	18300.7

Table 44: Efficiency corrected yields for Run 1 and 2016.

1058 The reason for lower efficiency corrected yield per pb^{-1} than expected in 2016 is due to the
 1059 big uncertainty on nSPD hits in simulation. In Run 1 only signal events with nSPD below 600
 1060 were expected and in 2016 this was below 450.

1061 As simulation does not reproduce nSPD distribution well there is a big uncertainty on this
 1062 variable. This directly affects efficiency corrected yield per pb^{-1} . To estimate this uncertainty, the
 1063 efficiency of requirement of $nSPD < 450$ in Run 1 $B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+$ data at the end of all the
 1064 selection is studied. Efficiency of this cut on data is computed by fitting $B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+$
 1065 spectrum before and after this cut, which is highlighted in Table 46. This is to be compared
 1066 with simulation efficiency of $nSPD < 450$ cut, see highlighted row in Table 45.

Efficiency	$B^+ \rightarrow J/\psi K^+$ 2011	$B^+ \rightarrow J/\psi K^+$ 2012	$B^+ \rightarrow J/\psi K^+$ 2016
ε_{nSPD}	0.9811 ± 0.0003	0.9597 ± 0.0004	1.0000 ± 0.0000
Ratios	-	$\varepsilon_{nSPD}^{2012}/\varepsilon_{nSPD}^{2011} = 0.9782 \pm 0.0005$	$\varepsilon_{nSPD}^{2016}/\varepsilon_{nSPD}^{2012} = 1.0420 \pm 0.0004$

Table 45: Efficiencies of nSPD selection for $B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+$ MC for 2011 2012 and 2016, where nSPD<450 is imposed on Run 1 MC and for 2016 only MC efficiencies with TCK==288888335 are listed.

Properties	2011	2012	2016
N total	$(5.36 \pm 0.01) \times 10^5$	$(1.21 \pm 0.00) \times 10^6$	$(5.86 \pm 0.01) \times 10^5$
N FR	$(5.51 \pm 0.02) \times 10^4$	$(1.29 \pm 0.00) \times 10^5$	$(3.25 \pm 0.04) \times 10^4$
N nSPD	$(4.71 \pm 0.02) \times 10^4$	$(1.06 \pm 0.00) \times 10^5$	$(3.25 \pm 0.04) \times 10^4$
ε_{nSPD}	0.8544 ± 0.0055	0.8229 ± 0.0035	1.0000 ± 0.0000
Ratios	-	$\varepsilon_{nSPD}^{2012}/\varepsilon_{nSPD}^{2011} = 0.9632 \pm 0.0074$	$\varepsilon_{nSPD}^{2016}/\varepsilon_{nSPD}^{2012} = 1.2152 \pm 0.0202$

Table 46: $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+$ yields and efficiency of nSPD selection in 2011, 2012, 2016 samples. The yields are obtained by fitting the normalisation channel. For 2016 only data with TCK==288888335 are analysed. N total is the number of $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+$ candidates fitted right from the stripping, N FR is the number of $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+$ candidates fitted after ε_{FR} is imposed in the selection chain.

1067 As efficiencies are taken from 2012 simulation, the efficiency on data is compared to 2012
 1068 data. In 2012, MC efficiency of nSPD requirement is 96% efficient and in data it is 82% efficient.
 1069 In conclusion the efficiency corrected yield per pb^{-1} in 2016 should have an uncertainty of 14.5%
 1070 due to the nSPD mismodelling in MC, which makes expected efficiency corrected yield in 2016
 1071 compatible with that of Run 1.

1072 8.4 Efficiency equivalence between Stripping 21 and 21r1

1073 In the following efficiency calculations, it is assumed that the efficiencies in Stripping 21 are the
 1074 same as in Stripping 21r1 for both signal and normalisation. This assumption is tested with
 1075 normalisation $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+$ data which was passed through the selection and fitted
 1076 before and after each selection step using fit model described in subsection 8.1. The results of
 1077 this study are summarized in Table 47. As it can be noticed ratio of the efficiencies at each step
 1078 is very close to 1 and the total efficiency of the selection listed in Table 47 also is the same.

Properties	2011	2012	2012/2011
N total	$(5.36 \pm 0.01) \times 10^5$	$(1.21 \pm 0.00) \times 10^6$	2.2576 ± 0.0073
N misidstripping	$(3.17 \pm 0.01) \times 10^5$	$(7.22 \pm 0.01) \times 10^5$	2.2757 ± 0.0051
N PIDinmuonacc	$(2.97 \pm 0.01) \times 10^5$	$(6.76 \pm 0.01) \times 10^5$	2.2756 ± 0.0053
N TRG	$(2.38 \pm 0.01) \times 10^5$	$(5.62 \pm 0.01) \times 10^5$	2.3642 ± 0.0061
N PIDnshared	$(2.17 \pm 0.01) \times 10^5$	$(5.11 \pm 0.01) \times 10^5$	2.3568 ± 0.0063
N CombiBDT	$(1.05 \pm 0.00) \times 10^5$	$(2.51 \pm 0.01) \times 10^5$	2.3811 ± 0.0089
N MisidBDT	$(5.53 \pm 0.02) \times 10^4$	$(1.29 \pm 0.00) \times 10^5$	2.3351 ± 0.0121
N FR	$(5.51 \pm 0.02) \times 10^4$	$(1.29 \pm 0.00) \times 10^5$	2.3346 ± 0.0121
$\varepsilon_{misidstripping}$	0.5928 ± 0.0019	0.5975 ± 0.0014	1.0080 ± 0.0040
$\varepsilon_{PIDinmuonacc}$	0.9353 ± 0.0025	0.9353 ± 0.0017	1.0000 ± 0.0032
ε_{TRG}	0.8012 ± 0.0023	0.8323 ± 0.0016	1.0389 ± 0.0036
$\varepsilon_{PIDnshared}$	0.9106 ± 0.0028	0.9077 ± 0.0019	0.9969 ± 0.0037
$\varepsilon_{CombiBDT}$	0.4864 ± 0.0019	0.4914 ± 0.0012	1.0103 ± 0.0046
$\varepsilon_{MisidBDT}$	0.5245 ± 0.0028	0.5144 ± 0.0018	0.9807 ± 0.0063
ε_{FR}	0.9976 ± 0.0061	0.9974 ± 0.0040	0.9998 ± 0.0073
ε_{total}	0.1285 ± 0.0013	0.1279 ± 0.0009	
Ratios	-	$\varepsilon_{tot}^{2012}/\varepsilon_{tot}^{2011} = 0.9953 \pm 0.0124$	

Table 47: $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+$ yields and efficiencies of selection in 2011 and 2012 samples after using fitting framework for normalisation fit. N total is the number of $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+$ candidates fitted right from the stripping. ε_{total} is total efficiency excluding trigger efficiency from this calculation (as this should not be taken in this way). It can be seen that the total efficiency ratio between $\varepsilon_{tot}^{2012}/\varepsilon_{tot}^{2011}$ is very close to 1.

9 Corrected Mass Fit

A simultaneous unbinned maximum likelihood fit to the $\mu^+ \mu^- \mu^+$ corrected mass after the full selection in two bins of fractional corrected mass error (FCME), $\sigma_{low,high} = \frac{\sigma_{err}}{M_{B_{corr}}}$, will be performed. As the corrected mass error have a dependence on resolution as mentioned in subsection 4.8, this split will increase sensitivity. Shapes and yields used for different components of necessary for the signal corrected mass fit, will be described in the following sections. Components are constrained either with data-driven methods or methods that use simulation samples.

Blinding Procedure: For now, the signal region defined $4500 \text{ MeV}/c^2 < M_{B_{corr}} < 5500 \text{ MeV}/c^2$ is blinded and hence in data plots in this region is omitted. In blinded unbinned likelihood fit, the shapes and yields are extrapolated to the blinded region in order to asses sensitivity, summarized in subsection 9.3. Once given permission to unblind, unbinned likelihood fit to full data will be performed with signal yield floating and mean μ_{sig} and width σ_{sig} constrained from MC.

9.1 Signal

Signal shape for Run 1 and 2016 is described by pdf in corrected mass taken from $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu_\mu$ 2012 MC and 2016 MC cocktail. After all respective selection, this cocktail is created by assigning event-by-event weights. Signal event-by-event weights, w^i , to create cocktail come from differences between 2012 MC and 2016 MC. To obtain signal weights, following conditions

1097 must be satisfied,

$$n^{2012} = \mathcal{L}^{2012} \times \sigma_{pp \rightarrow b\bar{b}}^{2012} \quad (34)$$

$$n^{2016} = \mathcal{L}^{2016} \times \sigma_{pp \rightarrow b\bar{b}}^{2016} \quad (35)$$

$$w^{2012} \times N^{2012} + w^{2016} \times N^{2016} = N^{2012} + N^{2016} = N^{tot} \quad (36)$$

$$\frac{w^{2012} \times N^{2012}}{w^{2016} \times N^{2016}} = \frac{n^{2012}}{n^{2016}} \quad (37)$$

1098 yielding

$$w^{2012} = \frac{N^{2012} + N^{2016}}{N^{2012} \times (1.0 + \frac{n^{2016}}{n^{2012}})} = 0.931 \quad (38)$$

$$w^{2016} = \frac{N^{2016} + N^{2012}}{N^{2016} \times (1.0 + \frac{n^{2012}}{n^{2016}})} = 1.073, \quad (39)$$

1099 where N^{2012}, N^{2016} number of events at *generator level* and can be obtained by dividing
1100 number of reconstructed events by ϵ_{rec} , $\mathcal{L}^{2012}, \mathcal{L}^{2016}$ are integrated luminosities, and $\sigma_{pp \rightarrow b\bar{b}}^{2012}, \sigma_{pp \rightarrow b\bar{b}}^{2016}$
1101 are cross-sections in a given year. Values for these variables are summarized in Table 48, noting
1102 that for cross-section numbers only the ratio is important hence using assumption that 2016
1103 cross-section is double that of 2012. Other differences between 2012 and 2016 that need to be
1104 accounted is PID efficiency ϵ_{PID} , which depends on the kinematics of the final state particles.

Summary	MC 2012	MC 2016
ϵ_{PID}	taken from PID	taken from PID
ϵ_{gen}	0.1856 ± 0.0011	0.1959 ± 0.0016
N_{rec}	1114130.0	1107715.0
\mathcal{L}	$2968 pb^{-1}$	$1612 pb^{-1}$
$\sigma_{pp \rightarrow b\bar{b}}$	1	2

Table 48: Signal MC weights used to create cocktail of mixed 2012+2016 MC.

1105 The weight of an event can be calculated

$$w_{total}^{ie\{2012,2016\}} = w^{ie\{2012,2016\}} * w_{\epsilon_{PID}}^{ie\{2012,2016\}} \quad (40)$$

1106 These samples have passed all the selection and the PID efficiency is extracted from PIDCalib
1107 data samples, as this is for the ID efficiency efficiency rather than misID. The shape in the
1108 corrected mass is modelled by double-sided Crystal Ball function.

1109 The signal yield in the final data sample can be estimated using the *single event sensitivity* (α
1110 shown in). For a given assumption of branching fraction \mathcal{B} ,

$$N_{sig} = \frac{\mathcal{B}}{\alpha}. \quad (41)$$

1111 Only the yield will be floating in the final fit to the data. The fit to the weighted MC signal
 1112 cocktail in different bins of FCME can be seen Figure 38.

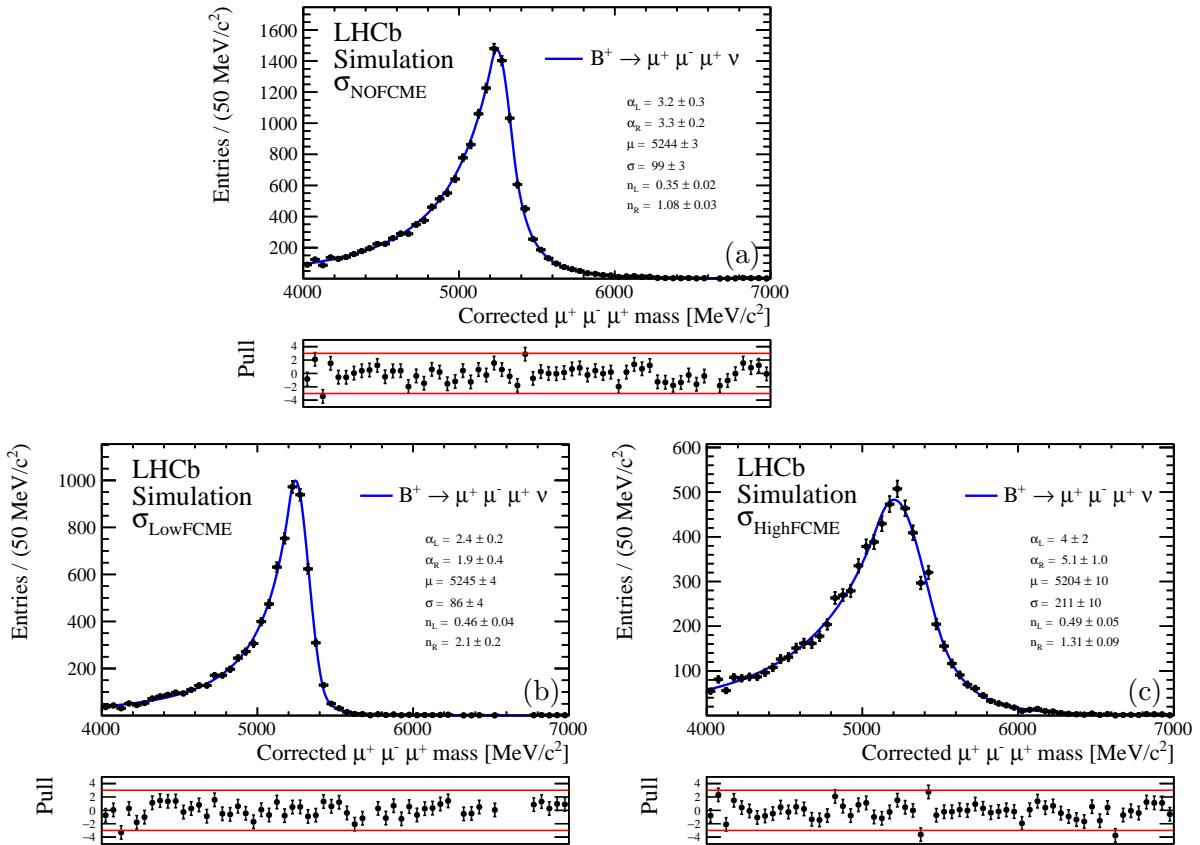


Figure 38: Fit to weighted combined signal *cocktail* for (a) NO FCME (b) Low FCME and (c) High FCME split.

1113 9.1.1 FCME split consideration

1114 In this section, the effect of FCME split for signal into $\sigma_{LowFCME}$ and $\sigma_{HighFCME}$ bin is
 1115 discussed.

1116 In subsection 7.8, the efficiency of this split was computed using the default simulation. If
 1117 this is not representative of the true signal in the data (the efficiency is wrong), then it would
 1118 manifest itself with poorer quality simultaneous fit in the signal region as the signal events would
 1119 be differently redistributed in the two bins. Simultaneous fit does take this into account and
 1120 hence no systematic uncertainty is assigned.

1121 To check whether no significant deviation is seen, comparison to PHSP simulation is used for
 1122 the purpose of this study. As shown in Table 49 there is not any marginal difference. Given that
 1123 the expected signal is rather small, this should not manifest itself very much in the quality of the
 1124 data fit.

	$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ (Nominal)	$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ (Alternative)
$\varepsilon_{LowFCME}$	0.483501	0.522695
$\varepsilon_{HighFCME}$	0.516499	0.477305

Table 49: Comparison between FCME split efficiencies using two different signal simulations.

1125 9.2 Backgrounds

1126 9.2.1 MisID background

1127 The level and the shape of misID background is extracted by fitting the misID data samples
 1128 obtained using the method described in subsection 5.4. A binned χ^2 fit is used to extract the
 1129 shape and yields parameters. The reason why a binned χ^2 fit is used is that these misID samples
 1130 are low-statistics weighted samples and the shape and yield needs be propagated to the final
 1131 data fit while preserving the fit parameter correlations. Since there is prescale factor of 1% at
 1132 stripping level, to obtain the correct yield, the final number needs to be multiplied by this factor.

1133 As the misID weights obtained from kinematically binned $B^+ \rightarrow J/\psi K^*$ sample have
 1134 uncertainties associated with them, see Figure 28. These are accounted for in the fit by Gaussian
 1135 variation of these weights in a given kinematic bin of p, η for each particle species and then folded
 1136 in to the misID calculation. Each variation results in different template for misID shape. This
 1137 misID template is subsequently binned in 15 bins of corrected mass. From each corrected mass
 1138 bin, mean μ and error σ from Gaussianly distributed number of misID events is obtained. The
 1139 total uncertainty due to the weight, σ_{tot} , for given bin of corrected mass is hence obtained using
 1140 $\sqrt{\sigma^2 + \sqrt{\sum w_i^2}}$, where $\sigma_{par} = \sqrt{\sum w_i^2}$ is the associated error per bin and σ is the standard
 1141 deviation obtained from variation of misID weights. Finally, The binned χ^2 is fitted to this shape.
 1142 The number of misID events for different species-regions after all selections are seen in Table 50.

1143 As it can be seen, crossfeedweight is only considered for SS (same sign) kaon and pion misID.
 1144 This arises as a consequence of two characteristics of the misID crossfeedweight procedure : the
 1145 convergence criteria makes unbalanced (one sample very high in misID events and other sample
 1146 very small in number of misID events) samples hard to satisfy (case for most of OS samples),
 1147 and secondly, proton-like region samples are very sparse, see Figure 39.

1148 The binned χ^2 fit to different bins of FCME with full errors split due to different source are
 1149 shown in Figure 41. The binned χ^2 fit to different bins of FCME with full errors only is shown
 1150 in Figure 41. For this fit, Crystal Ball function was used. Results of the fits in different bins
 1151 of FCME can be seen in Table 51, Table 52, Table 53. Note that if the values in these tables
 1152 differ to Figure 40, it is due to different rounding. These fit results are propagated into final
 1153 data fit preserving correlations between parameters, which are hence set as multidimensional
 1154 Gaussian constraints in the final fit. This means that all uncertainties due to misID will be
 1155 directly accounted for in the data fit.

Sample	Region	PID	weight	Cummulative Number of Misid events / 100
Run1 SS	Kaon-region	Run 1 PID	crossfeedweight	1.98
Run1 SS	Pion-region		crossfeedweight	3.01
Run1 SS	Proton-region		no-crossfeedweight	3.07
Run1 OS	Kaon-region		no-crossfeedweight	3.10
Run1 OS	Pion-region		no-crossfeedweight	3.52
Run1 OS	Proton-region		no-crossfeedweight	3.53
2016 SS	Kaon-region		crossfeedweight	4.89
2016 SS	Pion-region		crossfeedweight	5.65
2016 SS	Proton-region		no-crossfeedweight	5.65
2016 OS	Kaon-region		no-crossfeedweight	5.73
2016 OS	Pion-region		no-crossfeedweight	6.18
2016 OS	Proton-region		no-crossfeedweight	6.19

Table 50: The final misID template is constructed by summing the contribution from Run 1 and 2016 kaon, pion and proton-like regions for both SS and OS misID contributions. The last column adds cumulatively the contributions with respect to the previous row. In order to obtain correct yield these numbers need to be corrected for by the prescale of 0.01.

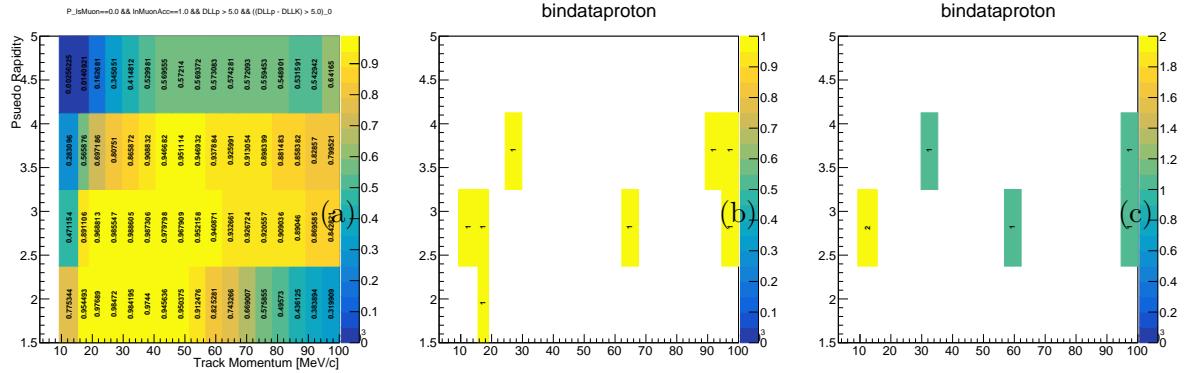


Figure 39: (a) PID efficiency map in p, η obtained from `PIDCalib` using Stripping 26 for protons to correctly identify them with following ID cut: `IsMuon==0&InMuonAcc==1&DLLp > 5.0&DLLp - DLLK > 5.0`. Regions below momentum of 9.3 GeV/ c^2 have tiny ID probability as in this case the $DLLp - DLLK > 5.0$ in order to identify protons. In any case, looking at relevant binned proton data in 2016 (b) *MagUp*, (c) *MagDown* samples, there is very low event yield, and as shown in Table 50, protons misID contribute significantly less than other species.

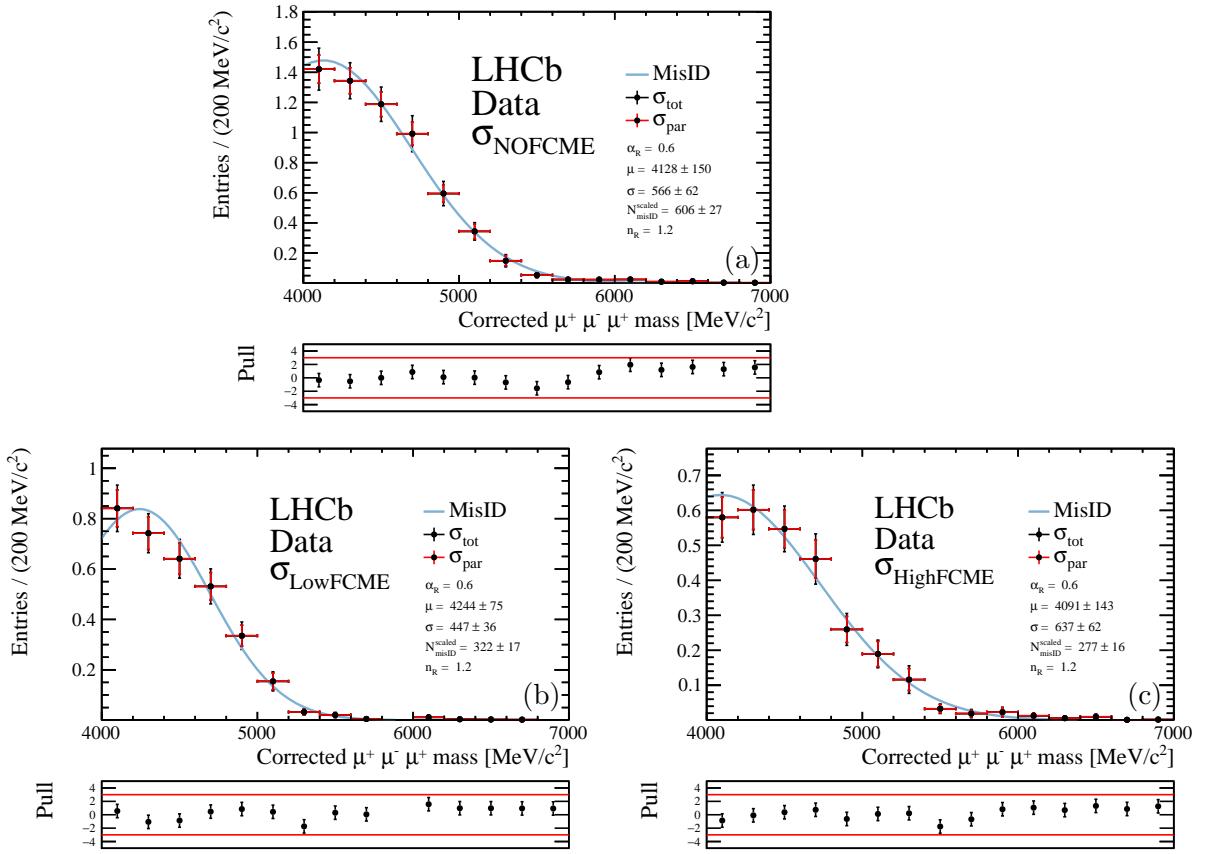


Figure 40: Binned χ^2 to misid template with no FCME split (b) Low FCME (c) High FCME. In high FCME bin, the distribution of misid is pollutes signal window more than in the low FCME bin. Both full weight error σ_{tot} and partial weight error σ_{par} can be seen.

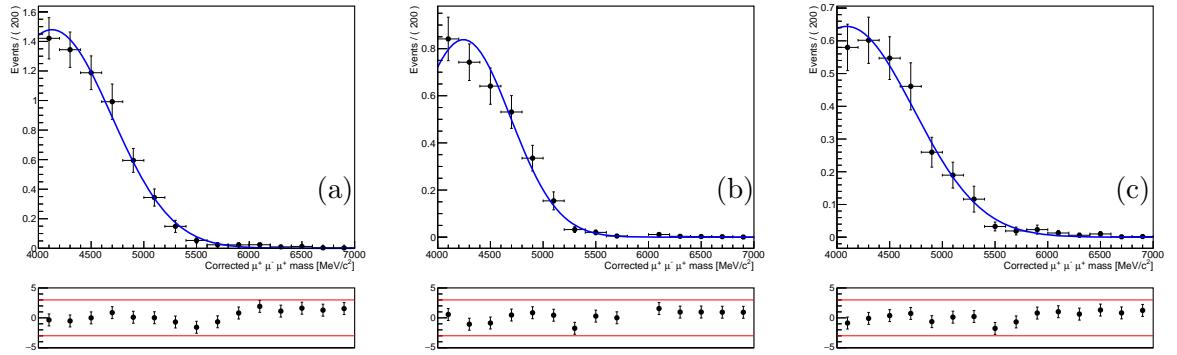


Figure 41: (a) Binned χ^2 to misID template with no FCME split (b) Low FCME (c) High FCME. In high FCME bin, the distribution of misID is pollutes signal window more than in the low FCME bin.

Variable	Value
mean	4130 ± 150
σ	566 ± 62
number	606 ± 26.8
n (fixed)	1.24
α (fixed)	0.6

Variable	Value
mean	4240 ± 74.7
σ	447 ± 35.9
number	322 ± 16.9
n (fixed)	1.24
α (fixed)	0.6

Variable	Value
mean	4090 ± 143
σ	637 ± 61.9
number	277 ± 15.5
n (fixed)	1.24
α (fixed)	0.6

Table 51: Binned χ^2 fit to the NOFCME sample.
 Table 52: Binned χ^2 fit to the LowFCME sample.
 Table 53: Binned χ^2 fit to the HighFCME sample.

1156

1157 9.2.2 Partially Reconstructed Background

1158 Simulation sample for partially reconstructed background that originates from B^+ proceeding
 1159 through D was described in subsection 5.3. This sample passed all of the selection and total
 1160 selection efficiency $\varepsilon_{totalsel}$ is listed in Table 54. The Table 72 in Appendix D details the efficiencies
 1161 in individual steps of selection. To obtain number of partially reconstructed background events
 1162 following equation is used:

$$N^{year} = \frac{N(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+)}{\varepsilon_{gen_{norm}} \times \varepsilon_{sel_{norm}}} \times \frac{\mathcal{B}(B^+ \rightarrow (D^0 \rightarrow K^+ \pi^- \mu^+ \mu^-) \mu^+ \nu)}{\mathcal{B}(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+)} \quad (42)$$

1163 The shape is fitted with sum of two Crystal Ball functions with free μ and σ as seen
 1164 in Figure 42. In Figure 43 the same fits are shown with both constant and floating parameters.
 1165 This estimation shows that partially reconstructed background that have been mediated by D
 1166 are the least contaminating background, but need to be accounted for in the final fit.

1167 In Figure 44, fit to wider range of corrected mass is shown. As it can be seen sum of two
 1168 Crystal Ball functions described the behaviour of partially reconstructed backgrounds well and
 1169 hence serves as motivation for the fitting model for this background.

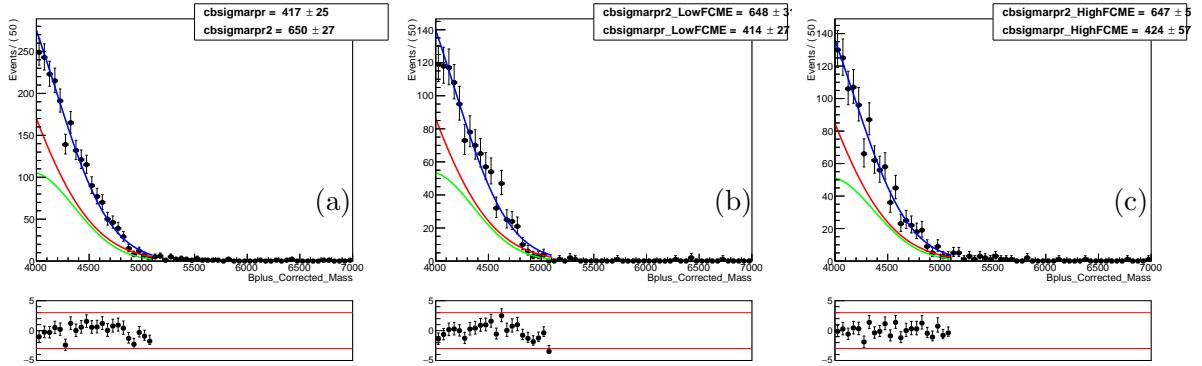


Figure 42: Fit to 2012 MC sample for partially reconstructed background for (a) NO FCME (b) Low FCME and (c) High FCME split.

Properties	Run 1 (2012 MC)	2016 (2012 MC)
$\mathcal{B}(B^+ \rightarrow (D^0 \rightarrow K^+ \pi^- \mu^+ \mu^-) \mu^+ \nu)$ see subsection 5.3	$(4.10 \pm 0.50) \times 10^{-7}$	$(4.10 \pm 0.50) \times 10^{-7}$
$\mathcal{B}(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+)$ see subsection 8.2	$(6.12 \pm 0.19) \times 10^{-5}$	$(6.12 \pm 0.19) \times 10^{-5}$
$\varepsilon_{gen,pr}$	$(1.61 \pm 0.00) \times 10^{-1}$	Using 2012
$\varepsilon_{sel,pr}$	$(1.16 \pm 0.02) \times 10^{-3}$	Using 2012
$\varepsilon_{gen,norm}$	$(1.62 \pm 0.00) \times 10^{-1}$	Using 2012
$\varepsilon_{sel,norm}$	$(3.58 \pm 0.01) \times 10^{-2}$	Using 2012
lowFCME		
$\mathcal{B} \times \varepsilon_{pr}$	$(3.80 \pm 0.50) \times 10^{-11}$	Using 2012
$\mathcal{B} \times \varepsilon_{norm}$	$(2.10 \pm 0.06) \times 10^{-7}$	Using 2012
$\frac{\mathcal{B} \times \varepsilon_{pr}}{\mathcal{B} \times \varepsilon_{norm}}$ low	$(1.81 \pm 0.24) \times 10^{-4}$	Using 2012
$N(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+)$	109000 ± 337	64700 ± 259
$N(B^+ \rightarrow (D^0 \rightarrow K^+ \pi^- \mu^+ \mu^-) \mu^+ \nu)$ low	19.8 ± 2.58	11.7 ± 1.53
highFCME		
$\mathcal{B} \times \varepsilon_{pr}$	$(3.80 \pm 0.50) \times 10^{-11}$	Using 2012
$\mathcal{B} \times \varepsilon_{norm}$	$(1.44 \pm 0.04) \times 10^{-7}$	Using 2012
$\frac{\mathcal{B} \times \varepsilon_{pr}}{\mathcal{B} \times \varepsilon_{norm}}$ high	$(2.60 \pm 0.40) \times 10^{-4}$	Using 2012
$N(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+)$	64100 ± 257	29800 ± 176
$N(B^+ \rightarrow (D^0 \rightarrow K^+ \pi^- \mu^+ \mu^-) \mu^+ \nu)$	17 ± 2.21	7.89 ± 1.03
noFCME		
$\mathcal{B} \times \varepsilon_{pr}$ no	$(7.63 \pm 0.95) \times 10^{-11}$	Using 2012
$\mathcal{B} \times \varepsilon_{norm}$ no	$(3.55 \pm 0.11) \times 10^{-7}$	Using 2012
$\frac{\mathcal{B} \times \varepsilon_{pr}}{\mathcal{B} \times \varepsilon_{norm}}$ no	$(2.15 \pm 0.28) \times 10^{-4}$	Using 2012
$N(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+)$ no	173000 ± 446	94500 ± 314
$N(B^+ \rightarrow (D^0 \rightarrow K^+ \pi^- \mu^+ \mu^-) \mu^+ \nu)$ no	37.3 ± 4.8	20.3 ± 2.61

Table 54: Calculation of number of events that comes from partially reconstructed backgrounds, assuming 2012 efficiencies but extrapolating to 2011,2012,2016 samples. $\mathcal{B}(B^+ \rightarrow (D^0 \rightarrow K^+ \pi^- \mu^+ \mu^-) \mu^+ \nu)$ is obtained by multiplying $\mathcal{B}(D^0 \rightarrow K^+ \pi^- \mu^+ \mu^-)$ and $\mathcal{B}(B^+ \rightarrow Dl^+\nu X) = (4.10 \pm 0.50) \times 10^{-7}$. Normalisation channel $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+$ was used to normalize to using Run 1 and 2016 yields. This calculation assumes ε_{PID} between normalisation and partially reconstructed sample ratio to be 1 which is a rough estimate.

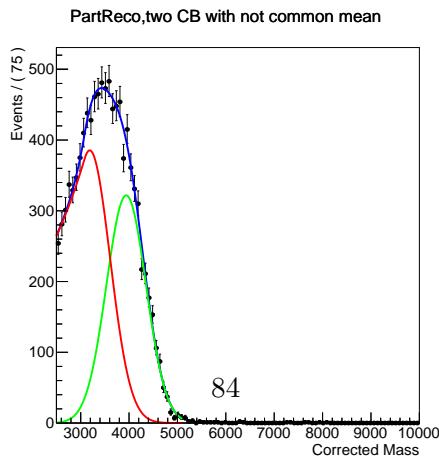


Figure 44: Fit to 2012 MC sample for partially reconstructed background before fit range was imposed on MC.

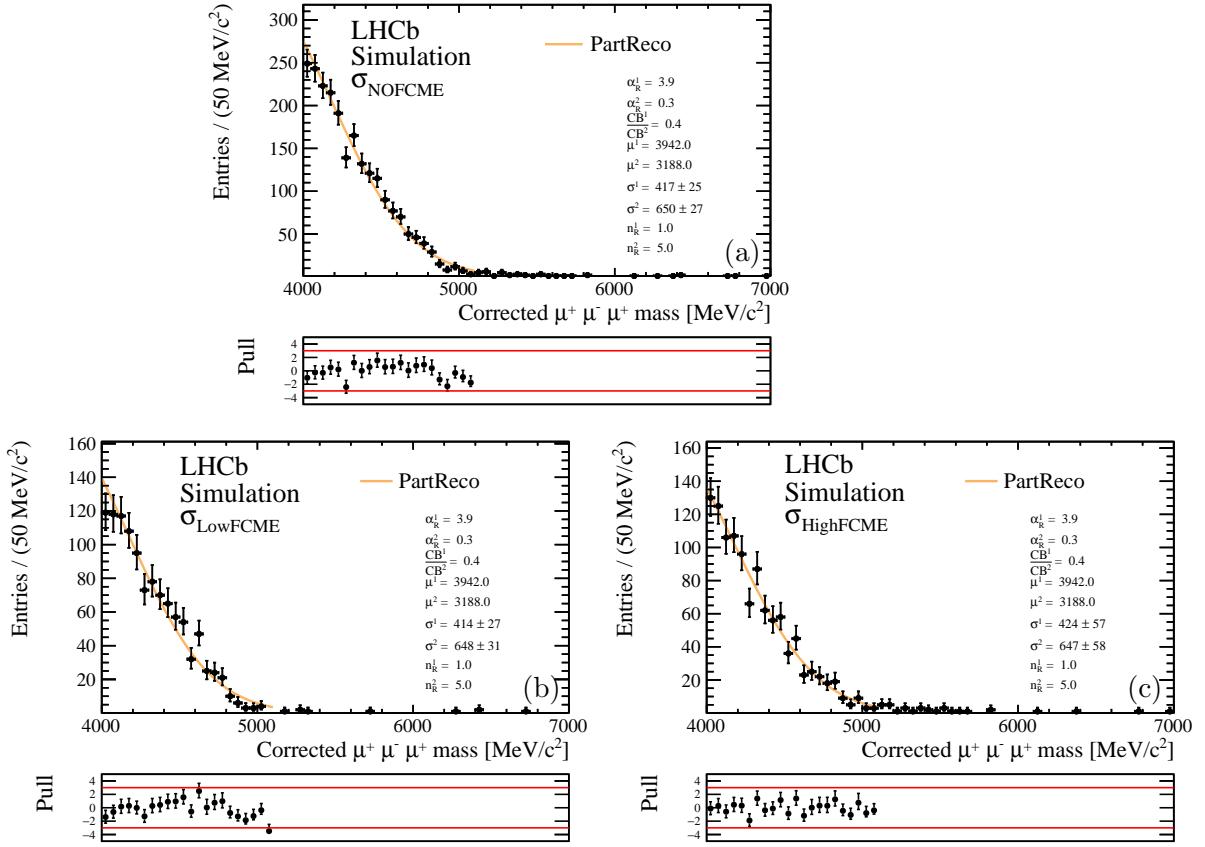


Figure 43: Fit to weighted combined partially reconstructed background simulation proxy for (a) NO FCME (b) Low FCME and (c) High FCME split.

1170 9.2.3 Combinatorial Background

1171 Data fit so far includes components for signal component, partially reconstructed background
 1172 component, and misID background component. The only component left is combinatorial
 1173 background. To model the combinatorial component exponential will be fitted to all unblinded
 1174 region. For now, it is fitted to blinded data.

1175 9.3 Blinded Data Fit

1176 A simultaneous unbinned extended maximum likelihood fit to the blinded data of corrected mass
 1177 after the full selection in two bins of fractional corrected mass error (FCME), $\sigma_{low,high} = \frac{\sigma_{err}}{M_{B_{corr}}}$, is
 1178 performed here. Let x be the corrected mass, and $y^i(z^k)$ all the parameters for non-simultaneous
 1179 (simultaneous) fit. The total PDF for the non-simultaneous fit can be written

$$f^{NS}(x, y^i) = N^{sig} \times f^{sig} + N^{MisID} f^{MisID} + N^{PR} f^{PR} + N^{Combi} \times f^{Combi}, \quad (43)$$

1180 The summary of all the parameters for the non-simultaneous data fit as well as its modelling is
 1181 shown in Table 55.

Fit Parameter	Status	Constraint	Obtained in
Yields			
N^{sig}	$\mathcal{B}(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu)$	Free	-
	$\mathcal{B}(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+)$	Constrained	gaussian
	$R_{FCME}^{21}(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu)$	Constrained	gaussian
	$R_{FCME}^{26}(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu)$	Constrained	gaussian
	$N(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+)^{Run1}_{FCME}$	Constrained	gaussian
	$N(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+)^{2016}_{FCME}$	Constrained	gaussian
N^{PR}	$\mathcal{B}(B^+ \rightarrow (D^0 \rightarrow K^+ \pi^- \mu^+ \mu^-) \mu^+ \nu)$	Constrained	gaussian
	$\mathcal{B}(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+)$	Constrained	gaussian
	$R_{FCME}^{21}(B^+ \rightarrow (D^0 \rightarrow K^+ \pi^- \mu^+ \mu^-) \mu^+ \nu)$	Constrained	gaussian
	$N(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+)^{Run1}_{FCME}$	Constrained	gaussian
	$N(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+)^{2016}_{FCME}$	Constrained	gaussian
	N^{Misid}	Constrained	mvg_gaussian
MisID Shape Parameters (Crystal Ball function)			
μ_{misid}	Constrained	mvg_gaussian	Figure 41
σ_{misid}	Constrained	mvg_gaussian	Figure 41
Others	Fixed	fixed to values seen in Figure 41	
PartReco Shape Parameters (sum of two Crystal Ball functions)			
All	Fixed	fixed to values seen in Figure 42	
Signal Shape Parameters (Double-sided Crystal Ball function)			
All	Fixed	fixed to values seen in Figure 38	
Combinatorial Shape Parameters (exponential function)			
β	Free	-	This fit

Table 55: Summary of all parameters going into the non-simultaneous fit. For all constrained variables the range is set to be within $\pm 5\sigma$. For simultaneous fit, all the shape parameters double. For the yield parameters, all of them double apart from the branching fractions.

1182 For the simultaneous fit the total PDF can be written as

$$f^S(x, y^k) = \sum_{j \in \sigma} (N_j^{sig} \times f_j^{sig} + N_j^{MisID} f_j^{MisID} + N_j^{PartReco} f_j^{PartReco} + N_j^{Combi} \times f_j^{Combi}), \quad (44)$$

1183 where σ is the bin of fractional corrected mass. The shared parameters in the two bins of
1184 fractional corrected mass error for the simultaneous fit are the branching fractions $\mathcal{B}(B^+ \rightarrow$
1185 $\mu^+ \mu^- \mu^+ \nu)$, $\mathcal{B}(B^+ \rightarrow (D^0 \rightarrow K^+ \pi^- \mu^+ \mu^-) \mu^+ \nu)$, $\mathcal{B}(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+)$.

1186 The fit to blinded data after all selection and with modelling of its components is shown
1187 in Figure 45 for all categories of FCME. The same fits but using stacked options is shown
1188 in Figure 46. As the signal region is blinded, the fit model for now is reduced by the signal
1189 shape, which leaves background-only model. Number of expected background events, N_b , is then
1190 obtained by integrating the total background PDF (misid PDF + partreco PDF + combinatorial
1191 PDF) in the signal region.

1192 In Table 55, it should be noted that both $N_{partreco}$ and N_{sig} are calculated using normalisation
1193 to $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+$, which include following parameters for the final data fit: the
1194 number of $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+$ events, efficiency ratios, \mathcal{B} for normalisation, partially
1195 reconstructed channel and signal. The last fit parameter of course is the parameter of interest.

A RooPlot of "Bplus_Corrected_Mass"

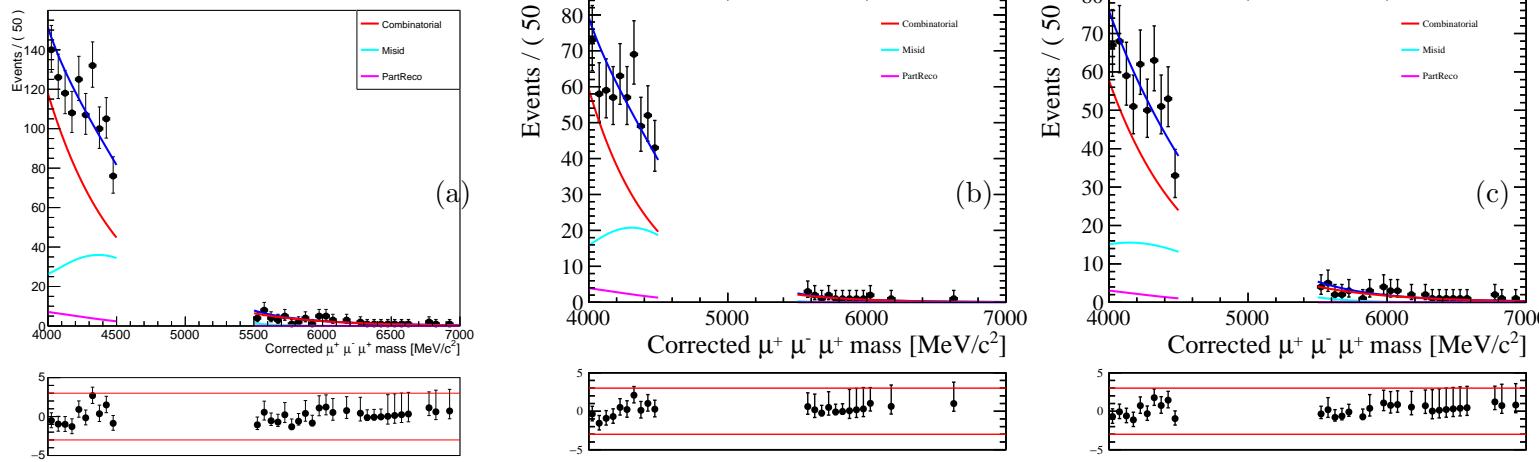


Figure 45: (a) Unbinned maximum likelihood fit to the blinded data in one bin of FCME. Simultaneous unbinned maximum likelihood fit to blinded data after full selection chain in two bins of FCME, with (b) fit to σ_{low} bin, (c) σ_{high} bin.

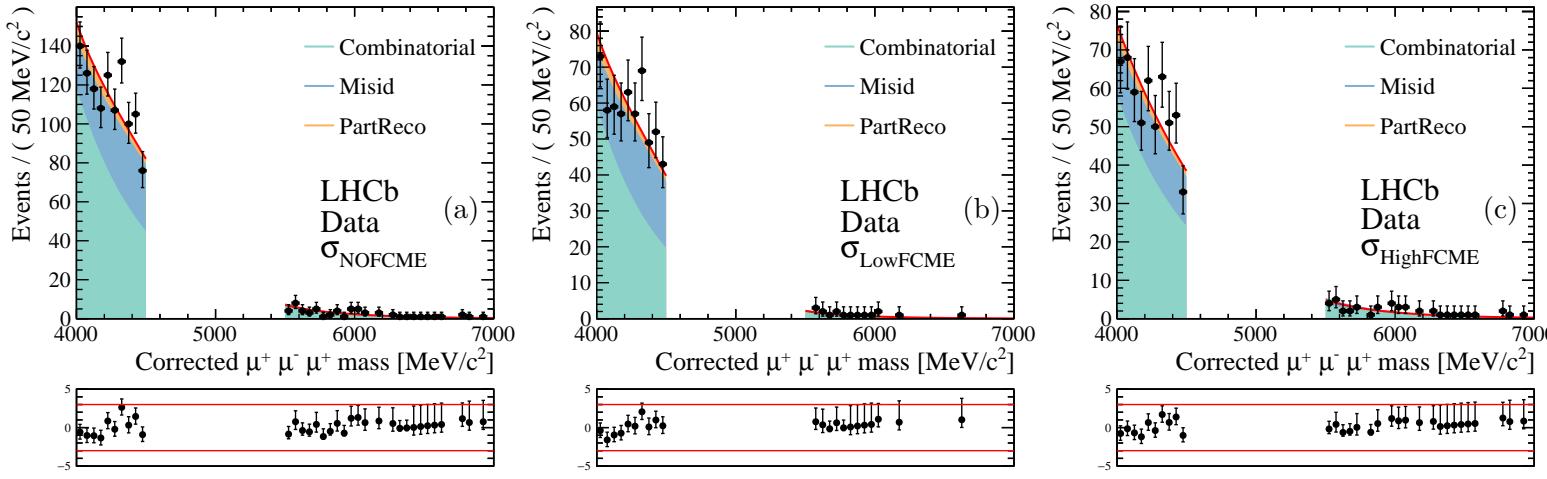


Figure 46: Stacked version of Figure 45. (a) Unbinned maximum likelihood fit to the blinded data in one bin of FCME. Simultaneous unbinned maximum likelihood fit to blinded data after full selection chain in two bins of FCME, with (b) fit to σ_{low} bin, (c) σ_{high} bin.

1198 10 Systematic uncertainties

1199 In this Section, systematics studies are performed in order to account for possible shortcomings
 1200 of the methods used. The summary of systematic uncertainties is summarized in Table 56. This
 1201 measurement will be statistically dominated.

Systematic source	Run 1/%	2016/%	Error Overall
$\mathcal{B}(J/\psi \rightarrow \mu^+ \mu^-)$	0.6	0.6	0.6
$\mathcal{B}(B^+ \rightarrow J/\psi K^+)$	3.0	3.0	3.0
Signal Decay Model	4.6	4.6	4.6
Trigger Data/Sim	5.1	1.3	3.2
Trigger HLT2	-	1.5	1.5
Kaon interaction probability	2.0	2.0	2.0
Kinematic Reweighting	1.0	2.0	1.5
Simulation Statistics	1.3	0.7	0.8
Fit bias	1.0	1.0	1.0
Total	8.0	6.6	7.1
Statistical source	Run 1/%	2016/%	Error Overall
Error on $B^+ \rightarrow J/\psi K^+$ yield	0.3	0.3	0.2

Table 56: Summary of systematic uncertainties.

1202 10.1 Trigger

1203 Systematic uncertainty due to trigger is calculated using **TISTOS method**, whereby the events
 1204 that are TIS (triggered independent of signal) and TOS (triggered on signal) are assumed to be
 1205 independent. Hence, the efficiency for selection of TOS candidates hence can be calculated by

$$\varepsilon_{TOS} = \frac{N_{TIS\&TOS}}{N_{TIS}}. \quad (45)$$

1206 This method allows estimation of the accuracy of using MC efficiencies to estimate trigger
 1207 efficiencies in the first place. In this measurement the trigger efficiencies are used to calculate
 1208 the ratio of signal and normalisation efficiencies. Comparing MC and data for normalisation
 1209 channel $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+$ hence provides measurement for this uncertainty and assumes
 1210 that it is of the same order for the signal channel.

1211 To calculate this uncertainty $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+$ truth-matched simulation is compared
 1212 to $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+$ *sWeighted* data for both 2012 and 2016 data. As discussed in
 1213 subsection 7.4 frequent change of TCK results in very inconsistent projection for simulation and
 1214 whole 2016 dataset comparisons. To decouple this effect from study of the efficiency, for this
 1215 systematic study, only the data that was taken with TCK 288888335 (MC) representing 35%
 1216 data is used together with default TCK. In Table 57 efficiency summary is listed. The efficiencies
 1217 listed are applied on the top of each other, so the efficiency is cumulative.

1218 Hence the systematic uncertainty in a given year is obtained by

$$\delta_{sim/data} = \left(\varepsilon_{HLT2}(Sweighted\ Data) - \varepsilon_{HLT2}(Simulation) \right) / \varepsilon_{HLT2}(Simulation), \quad (46)$$

1219 where ε_{HLT2} is the cumulative efficiency of L0, HLT1 and HLT2. This gives an upper limit on the
 1220 uncertainty due to trigger on the ratio between the signal and normalisation.

Level	N_TIS&TOS	N_TIS	ε_{TOS}
Simulation			
L0	109994	124150	0.886 ± 0.001
HLT1	20259	24777	0.818 ± 0.003
HLT2	4046	5367	0.754 ± 0.006
<i>Sweighted</i> Data			
L0	66109	74377	0.889 ± 0.001
HLT1	10297	12695	0.811 ± 0.004
HLT2	1709	2389	0.715 ± 0.009

Table 57: 2012 efficiencies with total systematic uncertainty 5.1%

1221 The total systematic uncertainty for 2012 due to use of simulation efficiency is 5.1%.

Level	N_TIS&TOS	N_TIS	ε_{TOS}
Simulation			
L0	182650	210021	0.870 ± 0.001
HLT1	38289	47326	0.809 ± 0.002
HLT2	13035	16564	0.787 ± 0.003
<i>Sweighted</i> Data			
L0	21267	24756	0.859 ± 0.002
HLT1	3638	4575	0.795 ± 0.006
HLT2	1094	1408	0.777 ± 0.011

Table 58: 2016 efficiencies with total systematic uncertainty 1.3%

1222 In order to evaluate systematic on using HLT2 selection that is only valid for one TCK in
 1223 2016, the same TISTOS method was run on all the data rather than just the data with TCK
 1224 288888335. It was then compared to efficiency results from simulation with only TCK 288888335.
 1225 Total systematic uncertainty can be estimated using Equation 45 **but** requiring beforehand
 1226 L0_TIS and HLT1_TIS. This leads to systematic uncertainty of 1.5% as seen in Table 59.

1227 10.2 Signal Decay Model

1228 Choice of the signal model used can have an effect on the efficiencies. In order to evaluate the
 1229 systematic uncertainty associated with this measurement two following models will be compared.
 1230 The nominal model relies B^+ decaying into $K(\mu^+, \nu_\mu)\mu^-\mu^+$ using BTOSLLBALL decay model
 1231 and $K(\mu^+, \nu_\mu)$ then decays into $\mu^+\nu_\mu$ using phase space decay model (PHSP), described in
 1232 subsubsection 3.2.2.
 1233 The alternative model described in subsubsection 3.2.1 relies fully on phase space model (PHSP).

Level	N_TIS&TOS	N_TIS	ε_{TOS}
Simulation			
HLT2	20073	22224	0.9032 ± 0.0020
Sweighted Data			
HLT2	6136	6899	0.889 ± 0.004

Table 59: 2016 efficiencies with total systematic uncertainty 1.5% for HLT2 only.

1234 Since this measurement is valid only for $minq < 980 \text{ MeV}/c^2$, the comparison between efficiencies
 1235 will be done with this condition met. Since generator level efficiencies may be correlated with this
 1236 $minq$ choice, generator level tuples with 25000 events with no generator level cuts, $N_{\text{generated}}$,
 1237 were produced for both nominal and alternative model with **Pythia6**. In this case, this is valid
 1238 approach to use as long as it is done for both signal simulations. To calculate overall efficiency
 1239 for nominal and alternative model, see Equations 47 and 48.

$$\varepsilon_{\text{nom}} = \varepsilon_{\text{nom}}^{\text{GEN}, minq} \times \varepsilon_{\text{nom}}^{\text{REC}, minq} \times \varepsilon_{\text{nom}}^{\text{OFF}, minq} \times \varepsilon_{\text{nom}}^{\text{CombiBDT}, minq} \times \varepsilon_{\text{nom}}^{\text{MisidBDT}, minq} \times \varepsilon_{\text{nom}}^{\text{fitrange}, minq} \times \varepsilon_{\text{nom}}^{\text{PID}, minq} \quad (47)$$

$$\varepsilon_{\text{alt}} = \varepsilon_{\text{alt}}^{\text{GEN}, minq} \times \varepsilon_{\text{alt}}^{\text{REC}, minq} \times \varepsilon_{\text{alt}}^{\text{OFF}, minq} \times \varepsilon_{\text{alt}}^{\text{CombiBDT}, minq} \times \varepsilon_{\text{alt}}^{\text{MisidBDT}, minq} \times \varepsilon_{\text{alt}}^{\text{fitrange}, minq} \times \varepsilon_{\text{alt}}^{\text{PID}, minq} \quad (48)$$

1240 The first two efficiencies are obtained in the following way:

$$\varepsilon^{\text{GEN}, minq} \times \varepsilon^{\text{REC}, minq} = \frac{N^{\text{in_acc}, minq}}{N_{\text{generated}, minq}} \times \frac{N^{\text{REC}, minq}}{N^{\text{in_acc}, minq}} \quad (49)$$

$$N^{\text{in_acc}, minq} = N^{\text{in_acc}} \times \varepsilon_{\text{minq}} \quad (50)$$

1241 In Equation 50, $\varepsilon_{\text{minq}}$ is obtained by $\frac{N_{\text{generated}, minq}}{N_{\text{generated}}}$, $N^{\text{in_acc}}$ is the number of events in
 1242 bookkeeping, $N^{\text{REC}, minq}$ is the number of reconstructed events with $minq$ condition. Breakdown
 1243 of the study is presented in Table 60 yielding to systematic uncertainty 4.5%.

	$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ (Nominal)	$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ (Alternative)
$N^{generated}$	25000	25000
$N^{generated,minq}$	24194	7463
$N^{in_{acc},minq}$	8844	2582
$N^{in_{acc}}$	1.11413e+06	1.24046e+06
$N^{REC,minq}$	116892	43874
$\varepsilon^{GEN,minq}$	0.183 ± 0.00155	0.173 ± 0.00275
$\varepsilon^{REC,minq}$	0.108 ± 0.000299	0.118 ± 0.000531
ε_{TRG}	0.742	0.773
ε_{OFF}	0.882	0.894
$\varepsilon_{CombiBDT}$	0.473	0.479
$\varepsilon_{MisidBDT}$	0.436	0.416
ε_{FR}	0.923	0.92
$\varepsilon^{other,minq}$	0.124 ± 0.000965	0.127 ± 0.00159
N^{final}	14546	5552
$\epsilon^{PID,minq}$	0.631 ± 0.005	0.628 ± 0.008
ε_{total}	$\epsilon_{nom} = (1.56 \pm 0.02) \times 10^{-3}$	$\epsilon_{alt} = (1.64 \pm 0.04) \times 10^{-3}$

Table 60: Comparison of efficiencies of nominal and alternative signal model samples using 2012 datasets with $minq < 980$ MeV/ c^2 leads to systematic uncertainty -4.6%. The detector acceptance efficiency was computed on the top of $minq$ selection and is obtained generating privately simulation with Pythia6 conditions.

1244 In addition, the selection efficiency, $\varepsilon^{other,minq} = \varepsilon^{OFF,minq} \times \varepsilon^{CombiBDT,minq} \times$
 1245 $\varepsilon^{MisidBDT,minq} \times \varepsilon^{FR,minq}$, as a function of $minq^2$ be seen in Figure 47 for both samples. It is
 1246 flat for both samples.

1247 10.3 Fit Bias

1248 In this section systematic uncertainty due to signal bias is evaluated. The pull is calculated
 1249 using pseudo-experiments where the data is generated for signal branching fraction $\mathcal{B} = 1.0e^{-8}$
 1250 corresponding to 16.48 signal events for Run 1 + 2016 data. These MC datasets are then fitted
 1251 with floating \mathcal{B} and corresponding number of fitted events is obtained. The pull is defined as
 1252 the difference in number of fitted events with the number events that pseudo-experiment was
 1253 created with, divided by the error on the number of signal given by the fit:

$$\frac{N_{sig}^{orig} - N_{sig}^{fit}}{\sigma^{fit}}. \quad (51)$$

1254 For this study 10000 pseudo-experiments were created and the results for both extended
 1255 non-simultaneous fit (fit to NOFCME) and extended simultaneous fit (fit to lowFCME and
 1256 highFCME) can be seen in Figure 48.

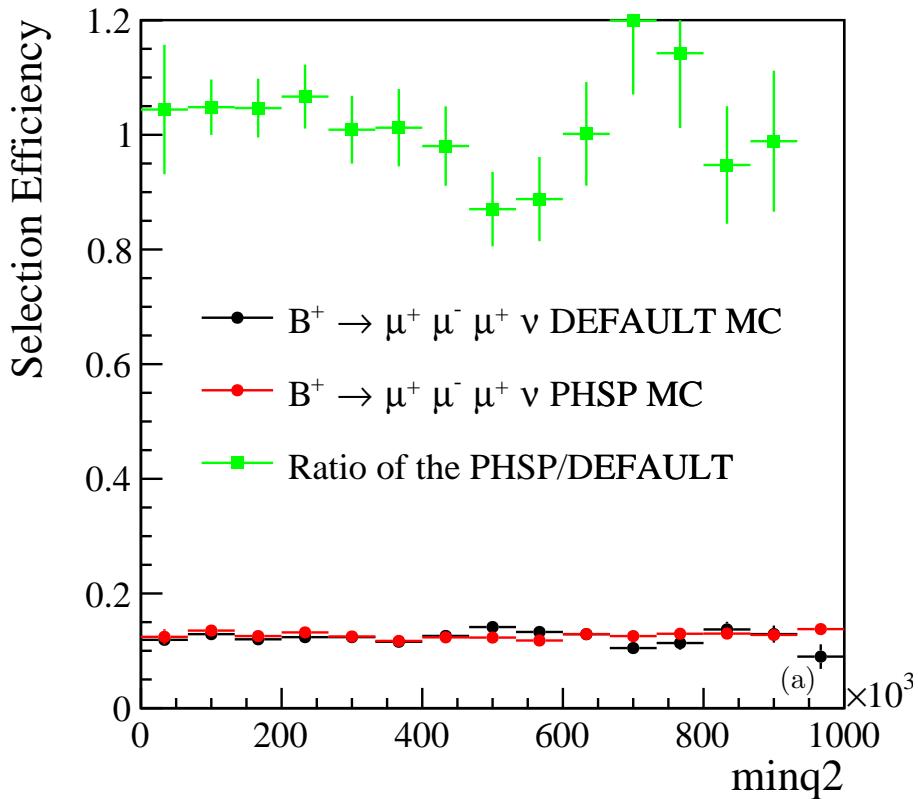


Figure 47: Selection efficiency as a function of minq^2 for the two models default and phase space simulation

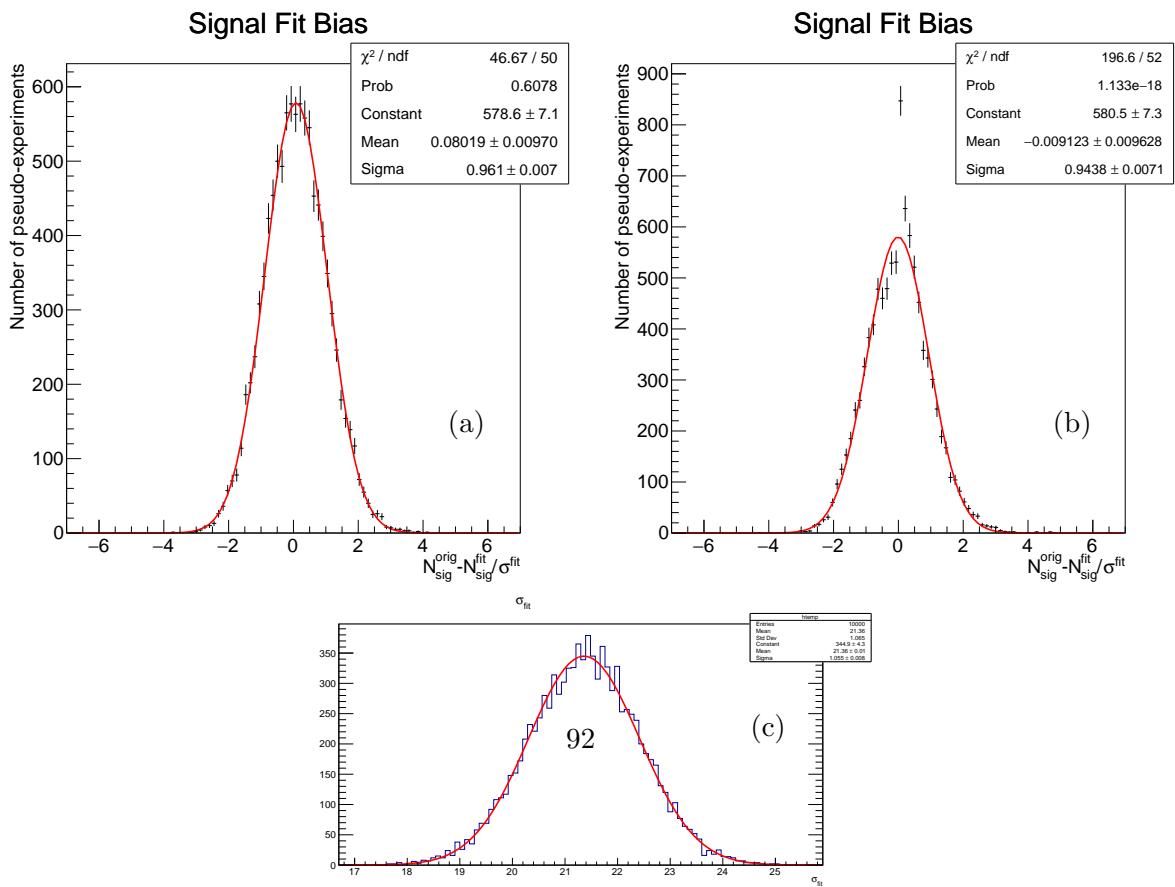


Figure 48: (a) Non-simultaneous fit, (b) simultaneous fit pulls. (c) Error in non-simultaneous fit.

1257 As it can be seen, for non-simultaneous fit there is overcoverage of 4 % but since the statistical
1258 error is of the order of 100% this is insignificant. The bias shows 8% preference for lower signal
1259 yields, and hence this will be added as a systematic uncertainty. This systematic will be redone
1260 post-unblinding.

Fit	\mathcal{B}	Statistical Error	Overcoverage	Bias
Not Simultaneous	$1.0e^{-8}$	$\approx 100\%$	4%	8%
Simultaneous	$1.0e^{-8}$	$\approx 100\%$	6%	1.0%

Table 61: Signal Bias estimate from 10 000 pseudo-experiments for both simultaneous and not simultaneous fit.

Floating Parameter	InitialValue	FinalValue +/-	Error	Gbl.Corr.
BNorm	6.160e-05	6.1134e-05 +/- 4.184e-06	<none>	
BBRpt	2.5785e-07	4.1079e-07 +/- 4.938e-08	<none>	
cbeamne4 HighFCM	4.2325e+03	4.1653e+03 +/- 4.141e+02	<none>	
cbeamne4 LowFCM	4.2095e+03	4.3073e+03 +/- 6.911e+01	<none>	
cbigsigma4 HighFCM	8.2871e+02	5.9915e+02 +/- 6.244e+01	<none>	
cbigsigma4 LowFCM	5.9195e+02	4.1599e+02 +/- 3.466e+01	<none>	
efficiency	2.1000e-02	2.1000e-02 +/- 1.0000e-02	<none>	
eff rate ratio pplus	2.196e-02	2.7086e-02 +/- 7.946e-04	<none>	
jpiski_2016_high	2.9746e-04	2.7064e-04 +/- 1.732e-02	<none>	
jpiski_2016_low	6.4703e+04	6.4723e+04 +/- 2.546e+02	<none>	
jpiski_run1_high	6.4048e+04	6.4078e+04 +/- 2.532e+02	<none>	
jpiski_run1_low	1.0919e+05	1.0922e+05 +/- 3.312e+02	<none>	
isid_scattered_HighFCM	3.5580e+08	3.0888e+08 +/- 1.154e+01	<none>	
isid_scattered_LowFCM	3.5580e+08	3.0888e+08 +/- 1.154e+01	<none>	
newexpcoson HighFCM	3.8776e-10	1.7462e-10 +/- 3.376e-05	<none>	
newexpcoson LowFCM	6.2216e-08	2.0483e-08 +/- 1.370e-04	<none>	
nexp HighFCM	1.0000e-01	6.3890e-02 +/- 4.106e+01	<none>	
nexp LowFCM	1.0000e-01	5.2245e-02 +/- 4.006e+01	<none>	

Figure 49: Correlation matrix for fit with background p.d.f only

Floating Parameter	Initial Value	Final Value +/-	Error	G0Corr.
beta	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
B0R0	1.1070e-07	2.5785e-07 +/- 1.25e-08	2.59e-08	0.000000e+00
clsmpr0_lowHIGH	1.5873e-03	4.2352e-03 +/- 1.46e-03	1.46e-01	0.000000e+00
clsmpr0_highHIGH	1.5873e-03	4.2352e-03 +/- 1.46e-03	1.46e-01	0.000000e+00
clsmpr1_lowHIGH	6.0130e-02	8.2871e-02 +/- 1.76e-01	1.76e-01	0.000000e+00
clsmpr1_highHIGH	6.0130e-02	8.0192e-02 +/- 1.34e-01	1.34e-01	0.000000e+00
eff ratio 2016 low	3.3696e-08	3.3818e-08 +/- 1.26e-08	1.26e-02	0.000000e+00
eff ratio 2016 high	3.3696e-08	3.2652e-08 +/- 1.43e-08	1.43e-02	0.000000e+00
eff ratio 2017 low	2.9078e-08	2.6395e-08 +/- 1.41e-08	1.41e-02	0.000000e+00
eff ratio 2017 high	2.9078e-08	2.0557e-08 +/- 1.96e-08	1.96e-02	0.000000e+00
jsipk 2018 low	2.9768e-08	1.9661e-08 +/- 1.19e-08	1.19e-02	0.000000e+00
jsipk 2018 high	2.9768e-08	2.0746e-08 +/- 1.64e-08	1.64e-02	0.000000e+00
jsipk 2019 low	4.7723e-04	4.6703e-04 +/- 4.22e-04	4.22e-02	0.000000e+00
jsipk 2019 high	4.7723e-04	4.6703e-04 +/- 4.22e-04	4.22e-02	0.000000e+00
jsipk run low	1.0022e-05	1.0916e-05 +/- 5.14e-06	5.14e-01	0.000000e+00
jsipk run high	2.0072e-05	5.3998e-06 +/- 1.07e-05	1.07e-01	0.000000e+00
newexp00 HIGHONE	1.7458e-03	3.8776e-10 +/- 2.38e-03	2.38e-02	0.000000e+00
newexp01 HIGHONE	1.7458e-03	3.8776e-10 +/- 2.38e-03	2.38e-02	0.000000e+00
newexp02 HIGHONE	3.8861e-02	1.0009e-01 +/- 2.51e-01	2.51e-01	0.000000e+00
new exp HIGH ONE	5.2216e-02	1.0009e-01 +/- 2.51e-01	2.51e-01	0.000000e+00

Figure 50: Correlation matrix for fit with signal and background p.d.f.

Correlation matrix from fit to blinded data with both (signal and background pdf, background only pdf) can be seen in Figures 49 and 50

1263 **10.4 Kinematic Reweighting**

1264 Simulation is known to have several shortcomings that lead to data-simulation disagreement.
1265 This will have inevitably impact on the efficiencies for both signal and normalisation channel
1266 leading to systematic uncertainty on the efficiency ratios. To asses the impact of correcting
1267 simulation following strategy is implemented.

1268 Both signal and normalisation simulation is reweighted to better reproduce the variables used in
1269 the selection. For 2016 simulations, `default` trigger configuration was used. Corrected variables
1270 are momentum, p , transverse momentum p_T and vertex χ^2 of B. No reweighting of occupancy
1271 is performed as the PID had not used occupancy information. Firstly p and p_T are reweighted
1272 simultaneously as there is correlation that needs to be taken into the account. Then, vertex
1273 χ^2 of B is reweighted. Normalisation data $B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+$ is used as target data
1274 distribution.

1275 The weights are calculated using normalised histograms of target data $B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+$,
1276 and normalized unweighted simulation $B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+$ in given bin,

$$w_{(bin)} = \frac{w_{B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+ data,bin}}{w_{B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+ simulation,bin}} \quad (52)$$

1277 These are then applied to correct both $B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+$, $B^+ \rightarrow \mu^+\mu^-\mu^+\nu$ simulation.
1278 The weighted and unweighted distributions can be seen in Figure 51. As it can be seen, agreement
1279 between simulation and target data is very good after reweighting.

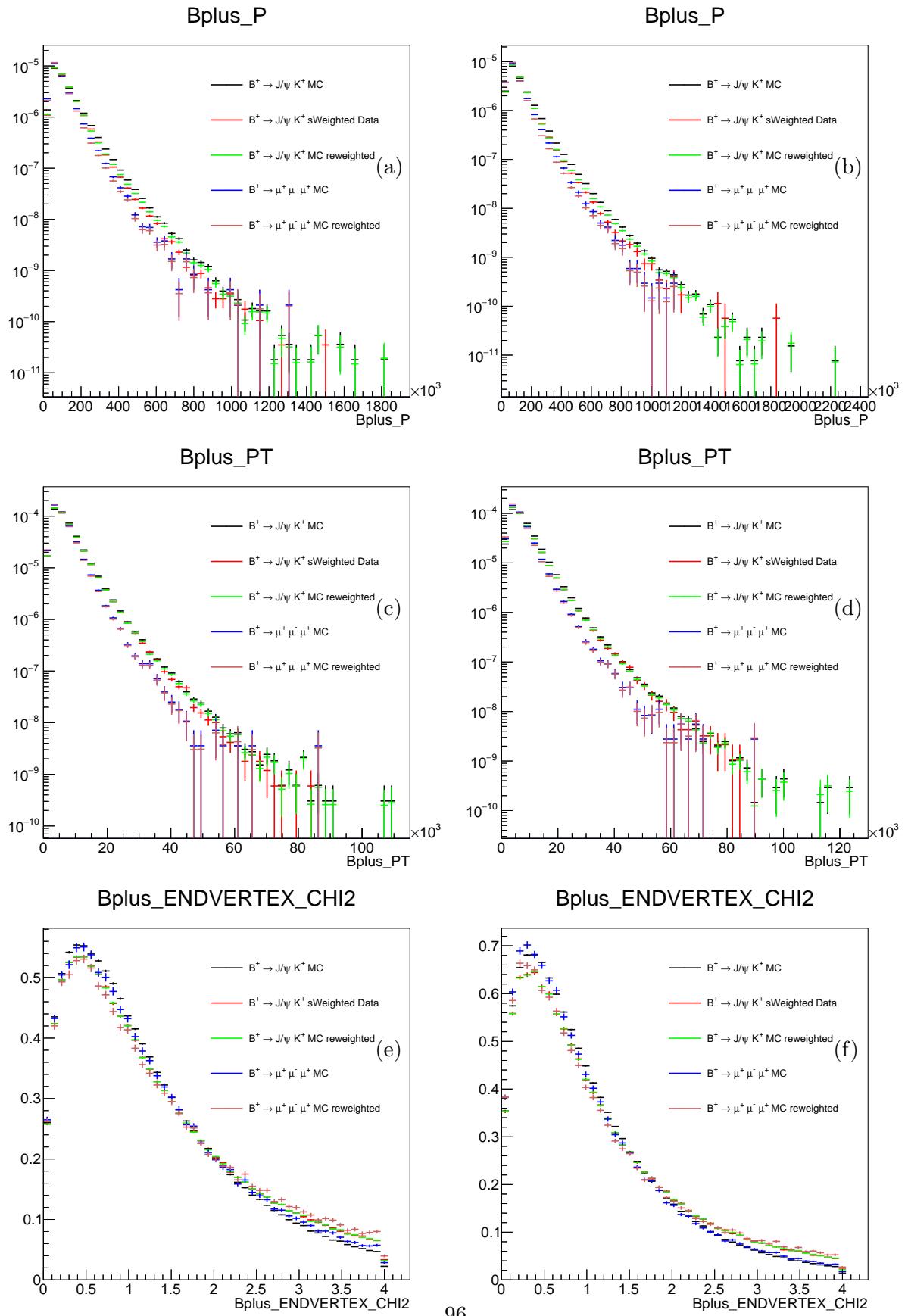


Figure 51: (a)(c)(e) Comparison of reweighted quantities for 2012 simulation to target data (red). (b)(d)(e) 2016.

1280 To asses systematics due to MC kinematic corrections, the weights are generated after
 1281 reconstruction for both normalisation and signal channel and then selection is performed. The
 1282 systematic uncertainties for efficiency ratios in 2012 and 2016 are summarized in Table 62. More
 1283 detailed breakdown of efficiencies is summarized in Appendix G.

Simulation year	Efficiency Ratio	Value	Systematic uncertainty
2012	$R_{NOFCME}^{21}(\varepsilon)$ nominal	0.380 ± 0.004	-1.0%
2012	$R_{NOFCME}^{21}(\varepsilon)$ reweighted	0.384 ± 0.005	
2016	$R_{NOFCME}^{26}(\varepsilon)$ nominal	0.361 ± 0.005	-2.0%
2016	$R_{NOFCME}^{26}(\varepsilon)$ reweighted	0.368 ± 0.005	

Table 62: Efficiency ratios for nominal and kinematically corrected samples and associated systematics for both 2012 and 2016 simulations.

1284 10.5 Partially Reconstructed Background

1285 In subsubsection 5.3.2, partially reconstructed decays that proceed via $\eta(')$ were considered. In
 1286 this section, there are two assumptions being made.

1287 First assumption is that the shape of this type of background is the same as the shape of
 1288 the partially reconstructed backgrounds that do not proceed via D decays. Using **RapidSim**,
 1289 this assumption was tested by comparing the corrected mass shape for the nominal peaking
 1290 background $B^+ \rightarrow (D^0 \rightarrow K^+\pi^-\mu^+\mu^-)\mu^+\nu$ decay and example of an additional one with the
 1291 highest \mathcal{B} listed in Table 15 where $D \rightarrow \eta(\eta \rightarrow \mu\mu)$. The shift of 200 MeV/ c^2 in the mean of the
 1292 corrected mass distribution can be seen in Figure 52.

1293 Another assumption is that the yield of partially reconstructed decays that proceed via $\eta(')$
 1294 backgrounds is not very significant given the same selection efficiency reduction. This assumption
 1295 may not hold and hence doubling the number of events for partially reconstructed background
 1296 can account for challenging of this assumption. This is rather conservative estimate.

1297 The systematic uncertainty will be calculated post-unblinding by allowing the mean of this
 1298 background to float by 200 MeV/ c^2 and doubling the number of partially reconstructed events
 1299 assumed for this type of background.

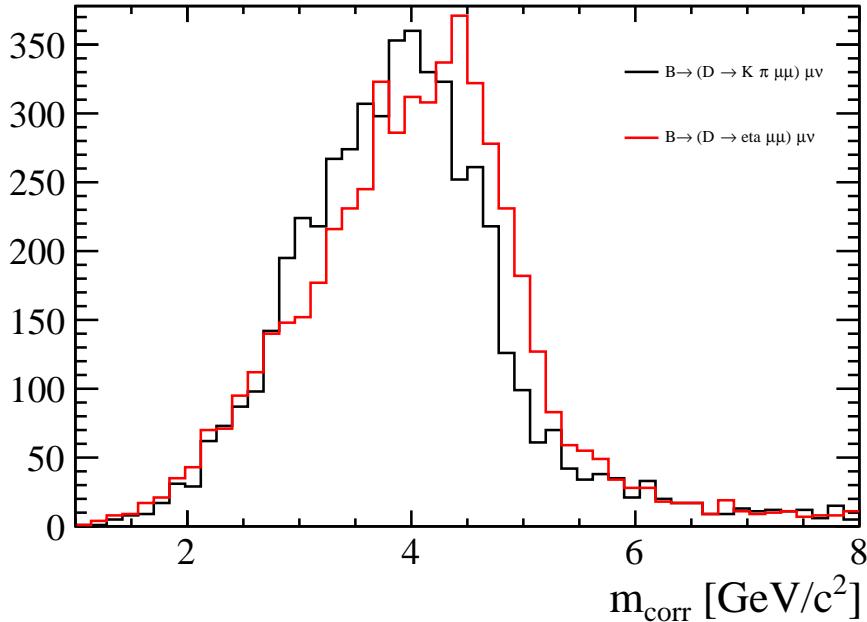


Figure 52: Corrected mass shapes for the two types of the partially reconstructed backgrounds: nominal peaking background $B^+ \rightarrow (D^0 \rightarrow K^+\pi^-\mu^+\mu^-)\mu^+\nu$ decay and $B \rightarrow (D \rightarrow \eta(\eta \rightarrow \mu\mu)\mu\nu)$.

1300 11 Crosschecks

1301 11.1 Validation of q^2 cut

1302 In this analysis the search is performed in $\text{min}q < 980 \text{ MeV}/c^2$ region. To validate this q^2 window,
 1303 the distribution of q^2 is shown using alternative simulation sample, see subsubsection 3.2.3,
 1304 that was produced exclusively for this analysis. In Figure 53 q^2 is plotted and most of the
 1305 contributions are below this threshold. This is important to know as this means that $\text{min}q$
 1306 selection should not be harder, as it would risk to remove large fraction of signal.

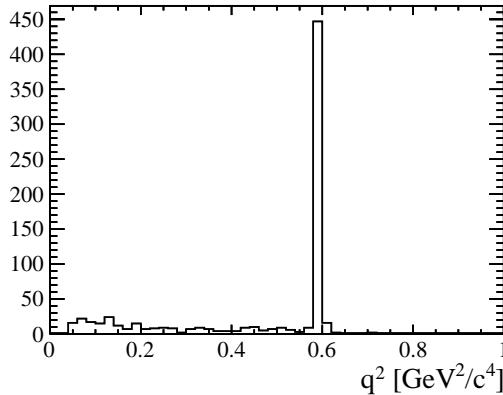


Figure 53: $q^2(\mu^+, \mu^-)$ distribution using model from Nikolai Nikitin. The contribution from photon pole, ρ and ω can be seen.

1307 11.2 AntiMisidBDT study (Control sample study)

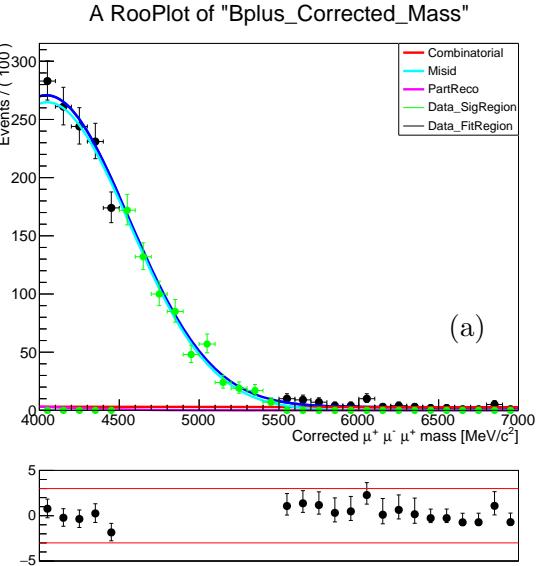
1308 The goal of this study is to validate the fitting framework, namely to check that the misID
 1309 estimation method and its parametrisation works very well. This can be achieved by forming
 1310 a control sample. In selection at the misID BDT stage, see subsubsection 4.5.3, there are two
 1311 samples that are created, the signal sample (which has the blinded region and is used for the
 1312 final signal fit) and antiMisidBDT sample formed of the candidates that did NOT pass relevant
 1313 misID BDT cuts. The second sample should be rich in the candidates that are likely to be of
 1314 misID nature, as they are categorized so by the misID BDT. Combinatorial BDT was already
 1315 applied on both of these samples, so the candidates in the antiMisidBDT sample they are not
 1316 likely to be of combinatorial nature.

1317 The default blinded fitting procedure (background pdf only, see subsection 9.3) is applied on
 1318 Run 1 blinded antiMisidBDT sample. This is the same method as doing blinded fit to the blinded
 1319 signal data, but in a control region. This is then subsequently overlaid with unblinded data of the
 1320 control region. In this case, no misID variation systematic error is taken into account. Updated
 1321 yield of partially reconstructed background is summarized in Table 63. Both simultaneous and
 1322 non-simultaneous fits were performed and are shown in Figure 54.

1323 In conclusion, the blinded fit to the blinded control antiMisidBDT sample overlaid with the
 1324 data points after from the blinded region shows very good agreement with expectations. This
 1325 shows that most of this sample is of misID nature as expected.

Properties	Run 1 (2012 MC)	2016 (2012 MC)
$\mathcal{B}(B^+ \rightarrow (D^0 \rightarrow K^+ \pi^- \mu^+ \mu^-) \mu^+ \nu)$	$(4.10 \pm 0.50) \times 10^{-7}$	$(4.10 \pm 0.50) \times 10^{-7}$
$\mathcal{B}(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+)$	$(6.12 \pm 0.19) \times 10^{-5}$	$(6.12 \pm 0.19) \times 10^{-5}$
ε_{gen}	$(1.61 \pm 0.00) \times 10^{-1}$	Using 2012
ε_{sel}	$(4.48 \pm 0.15) \times 10^{-4}$	Using 2012
$\varepsilon_{gennorm}$	$(1.62 \pm 0.00) \times 10^{-1}$	Using 2012
$\varepsilon_{selnorm}$	$(3.42 \pm 0.01) \times 10^{-2}$	Using 2012
lowFCME		
$\mathcal{B} \times \varepsilon_{pr}$	$(1.62 \pm 0.21) \times 10^{-11}$	Using 2012
$\mathcal{B} \times \varepsilon_{norm}$	$(1.72 \pm 0.05) \times 10^{-7}$	Using 2012
$\frac{\mathcal{B} \times \varepsilon_{pr}}{\mathcal{B} \times \varepsilon_{norm}}$ low	$(9.40 \pm 1.30) \times 10^{-5}$	Using 2012
$N(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+)$	88800 ± 302	69500 ± 269
$N(B^+ \rightarrow (D^0 \rightarrow K^+ \pi^- \mu^+ \mu^-) \mu^+ \nu$ low	8.34 ± 1.12	6.52 ± 0.879
highFCME		
$\mathcal{B} \times \varepsilon_{pr}$	$(1.32 \pm 0.18) \times 10^{-11}$	Using 2012
$\mathcal{B} \times \varepsilon_{norm}$	$(1.67 \pm 0.05) \times 10^{-7}$	Using 2012
$\frac{\mathcal{B} \times \varepsilon_{pr}}{\mathcal{B} \times \varepsilon_{norm}}$	$(7.90 \pm 1.10) \times 10^{-5}$	Using 2012
$N(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+)$	72200 ± 273	53600 ± 248
$N(B^+ \rightarrow (D^0 \rightarrow K^+ \pi^- \mu^+ \mu^-) \mu^+ \nu$	5.72 ± 0.781	4.25 ± 0.58
noFCME		
$\mathcal{B} \times \varepsilon_{pr}$	$(2.90 \pm 0.40) \times 10^{-11}$	Using 2012
$\mathcal{B} \times \varepsilon_{norm}$	$(3.39 \pm 0.10) \times 10^{-7}$	Using 2012
$\frac{\mathcal{B} \times \varepsilon_{pr}}{\mathcal{B} \times \varepsilon_{norm}}$ no	$(8.60 \pm 1.20) \times 10^{-5}$	Using 2012
$N(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+)$	161000 ± 407	123000 ± 356
$N(B^+ \rightarrow (D^0 \rightarrow K^+ \pi^- \mu^+ \mu^-) \mu^+ \nu$	13.9 ± 1.83	10.7 ± 1.4

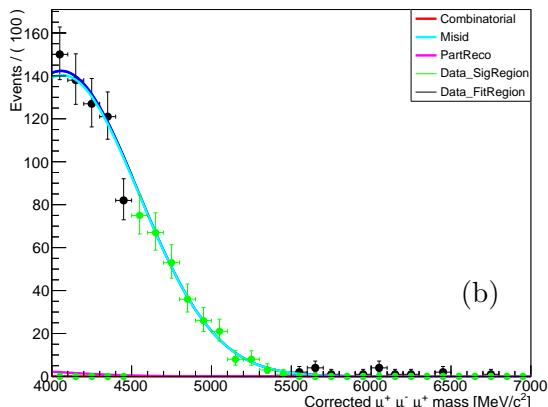
Table 63: Calculation of number of events that comes from partially reconstructed backgrounds, assuming 2012 efficiencies but extrapolating to 2011,2012,2016 samples. $\mathcal{B}(B^+ \rightarrow K^+ \pi^- \mu^+ \mu^-)$ is obtained by multiplying $\mathcal{B}(D^0 \rightarrow K^+ \pi^- \mu^+ \mu^-)$ and $\mathcal{B}(B^+ \rightarrow D l^+ \nu \text{ anything}) = (4.10 \pm 0.50) \times 10^{-7}$. Normalisation channel $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+$ was used to normalize to using Run 1 and 2016 yields. Assumes ε_{PID} between normalisation and partially reconstructed sample ratio 1.



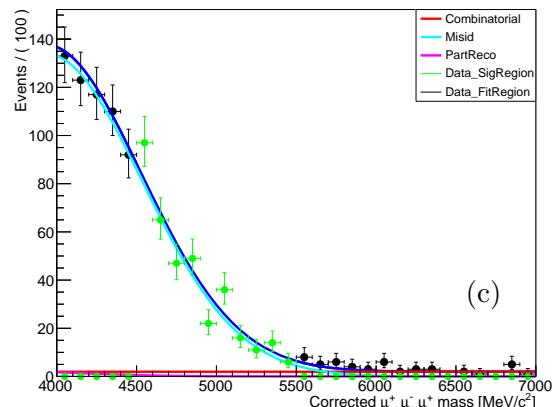
(a)

A RooPlot of "Bplus_Corrected_Mass"

A RooPlot of "Bplus_Corrected_Mass"



(b)



(c)

Figure 54: Fit with background pdf to Run 1 AntiMisidBDT sample for (a) NO FCME (b) Low FCME and (c) High FCME split. The misID is expectedly the dominant contribution in these fits. The fit is performed to blinded dataset and then overlaid with the signal region data. In all cases good agreement with data is achieved.

1326 **11.3 Ghosts**

1327 In this analysis the prediction and shape for misID is crucial. Crossfeed between different species
 1328 is taken into account. One potential group of these, that needs to be taken care of, are ghosts
 1329 and their crossfeed to other species should be evaluated. To see whether the contamination from
 1330 ghosts need to be modelled, it is possible to compare PID variable, `Probnnghost`, in two data
 1331 samples after two BDTs have been applied. MisID and control channel data samples are analysed
 1332 where the misIDed particle distribution for `Probnnghost` is compared to the distribution of
 1333 kaon from control channel. Because normalisation channel $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)K^+$ after all
 1334 selection is very clean, it is not expected to have significant amount of ghosts for any of its
 1335 tracks. This should mean that the distribution for `Probnnghost` should be peaking close to 0 for
 1336 example for kaon track. If misIDed particle from misID control samples has similar distribution
 1337 for `Probnnghost` than no special inclusion of ghost tracks in the misID iteration is necessary.
 1338 In Figure 55(a) difference as a function of `Probnnghost` shows good agreement between the two
 1339 samples.

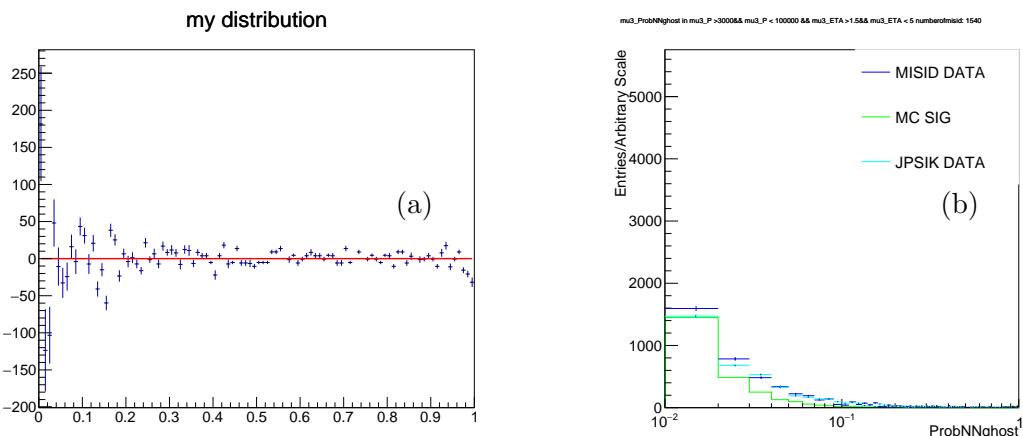


Figure 55: (a) Difference in `Probnnghost` between misIDed particle (from misID samples) and kaon (from normalization sample) as a function of `Probnnghost`. This was done in here cumulatively across momentum p from 3000 100000 MeV/ c and with η 1.5 - 5. If good agreement is achieved especially for high values of `Probnnghost`, then it is expected that the ghosts are taken care of as for normalization channel $B^+ \rightarrow J/\psi K^+$ ghosts are not expected. (b) `Probnnghost` distribution for different samples. It can be seen that not many events have high value.

1340 **11.4 Comparison of final shapes of components for final fits**

1341 To check the shape differences between σ_{lowFCME} and σ_{highFCME} , here is the summary of shapes
 1342 entering final blinded fit for Run 1 and 2016. It can be seen that splitting in these two bins
 1343 improves shape separation. In Figure 56, the signal shape with better resolution as well as misID
 1344 shape with better resolution (in σ_{lowFCME}) are more separated from each other to σ_{highFCME} .

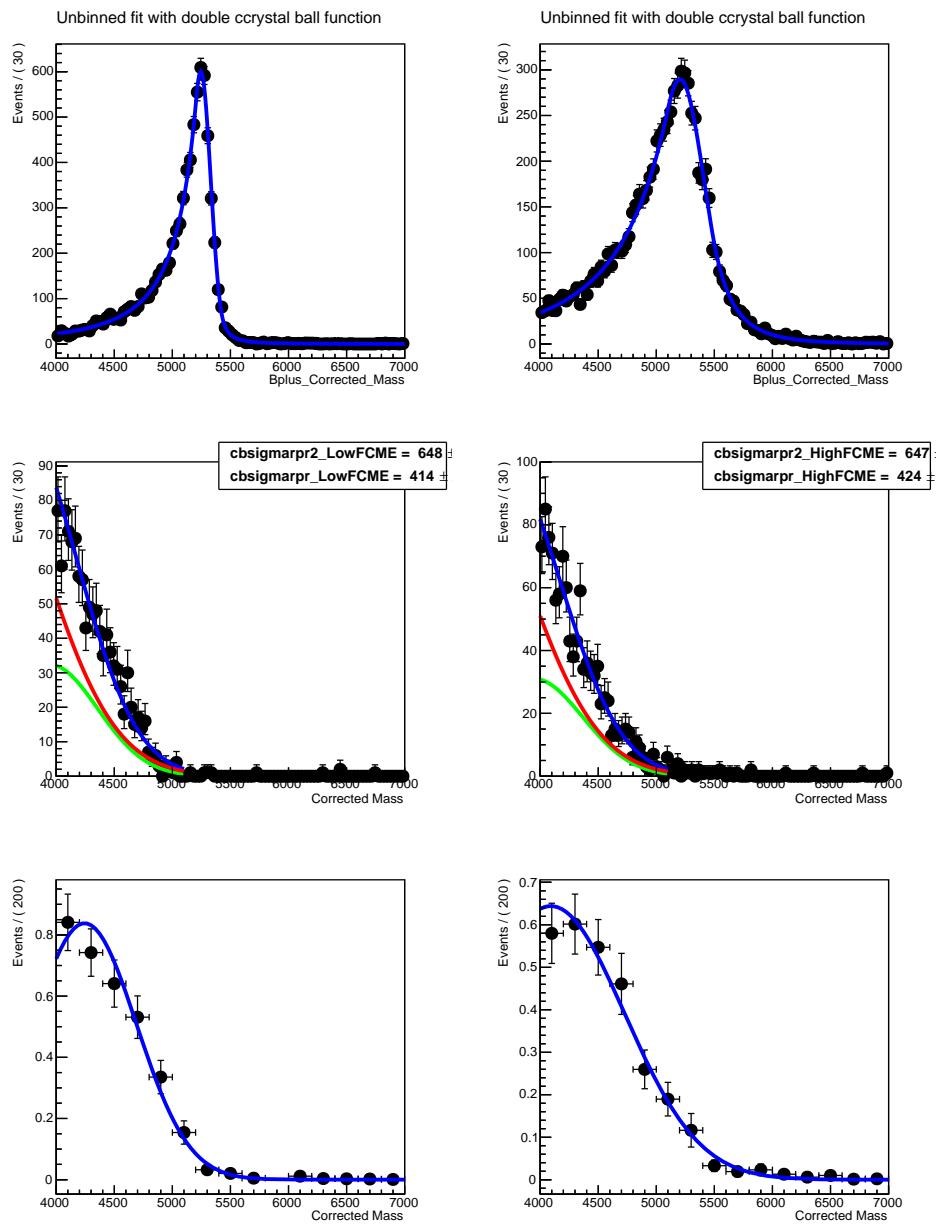


Figure 56: Left: σ_{lowFCME} shapes, Right: σ_{highFCME} shapes. First row: Signal Shape, second row: partially reconstructed background, third row: misID background.

1345 **12 PID Tuning**

1346 **12.1 Comparison of misID distributions after final selection for Run1 and**
1347 **2016**

1348 To check kinematic coverage of final misID distribution, misID distribution is plotted in bins
1349 of momentum p , and pseudorapidity η for different species and for different conditions. These
1350 distributions are then weighted relevantly (see subsection 5.4) to obtain misID shapes. It can
1351 be seen in Figure 57 that most of the misID comes in low momenta region, with very little
1352 differences between Run 1 and 2016, which makes adding Run 1 and 2016 easier.

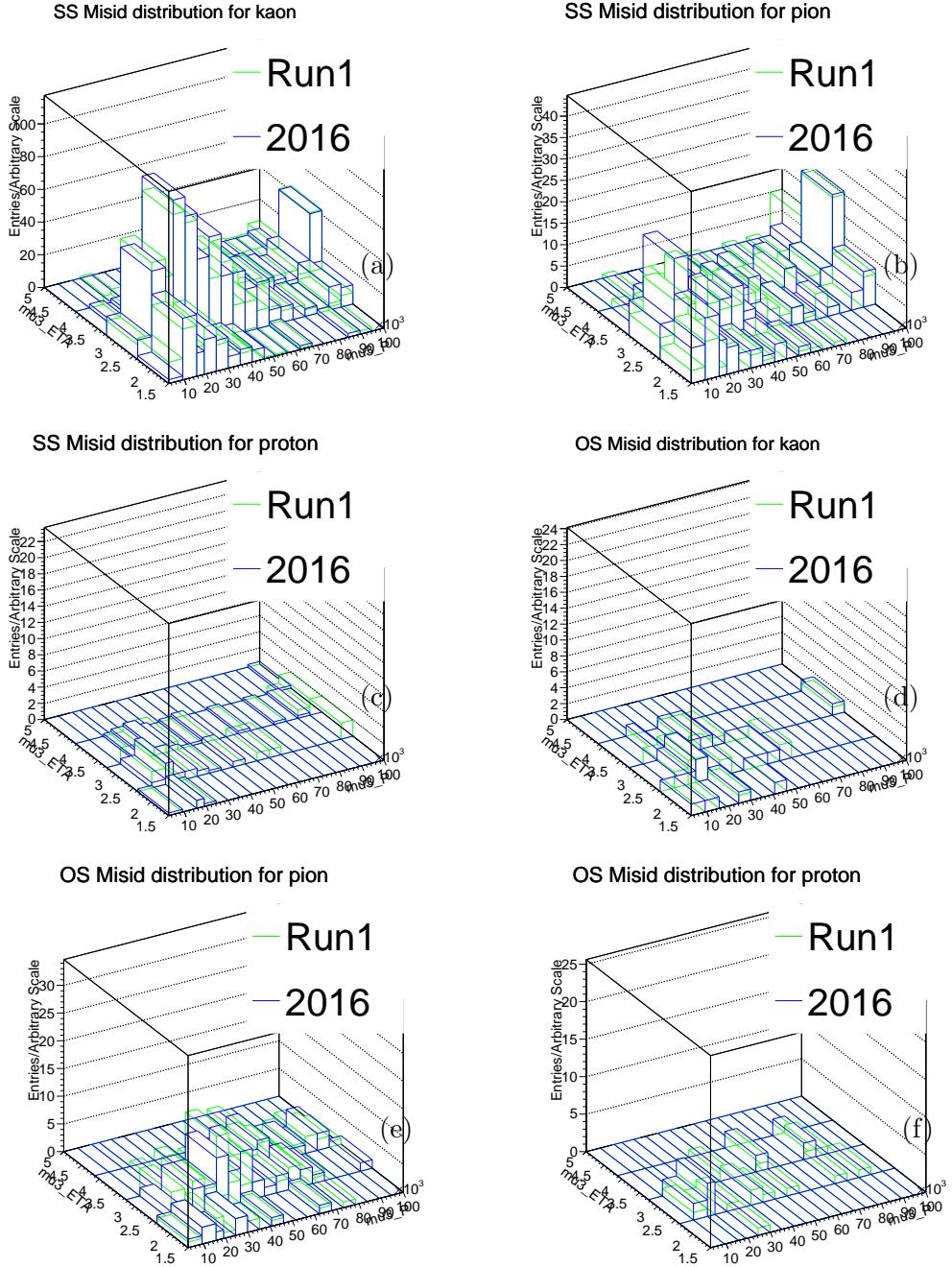


Figure 57: 2D Histogram-normalized non stacked distributions for (a) kaon SS misID and (b) pion SS misID (c) proton SS misID. 2D Histogram-normalized non stacked distributions for (d) kaon OS misID and (e) pion OS misID (f) proton OS misID. In PIDCalib/Calibration sample this binning in p was used, but there was just one bin of η . It can be seen that most of the misID events are in low momentum region.

1353 12.2 Probnnmu V2 vs V3 for Run 1

1354 As it can be seen in 12.1 the misID is in low momenta region hence the tuning for Run 1 is
 1355 chosen as to minimize the misID probability. PIDCalib samples were used for this study and
 1356 results can be seen with pion misID probabilities in first bins of p and in particular bin of η and
 1357 $ntracks$. Probnnmu V2 tuning is chosen since it yields lower misID probability for a given id
 1358 efficiency, see Figure 58.

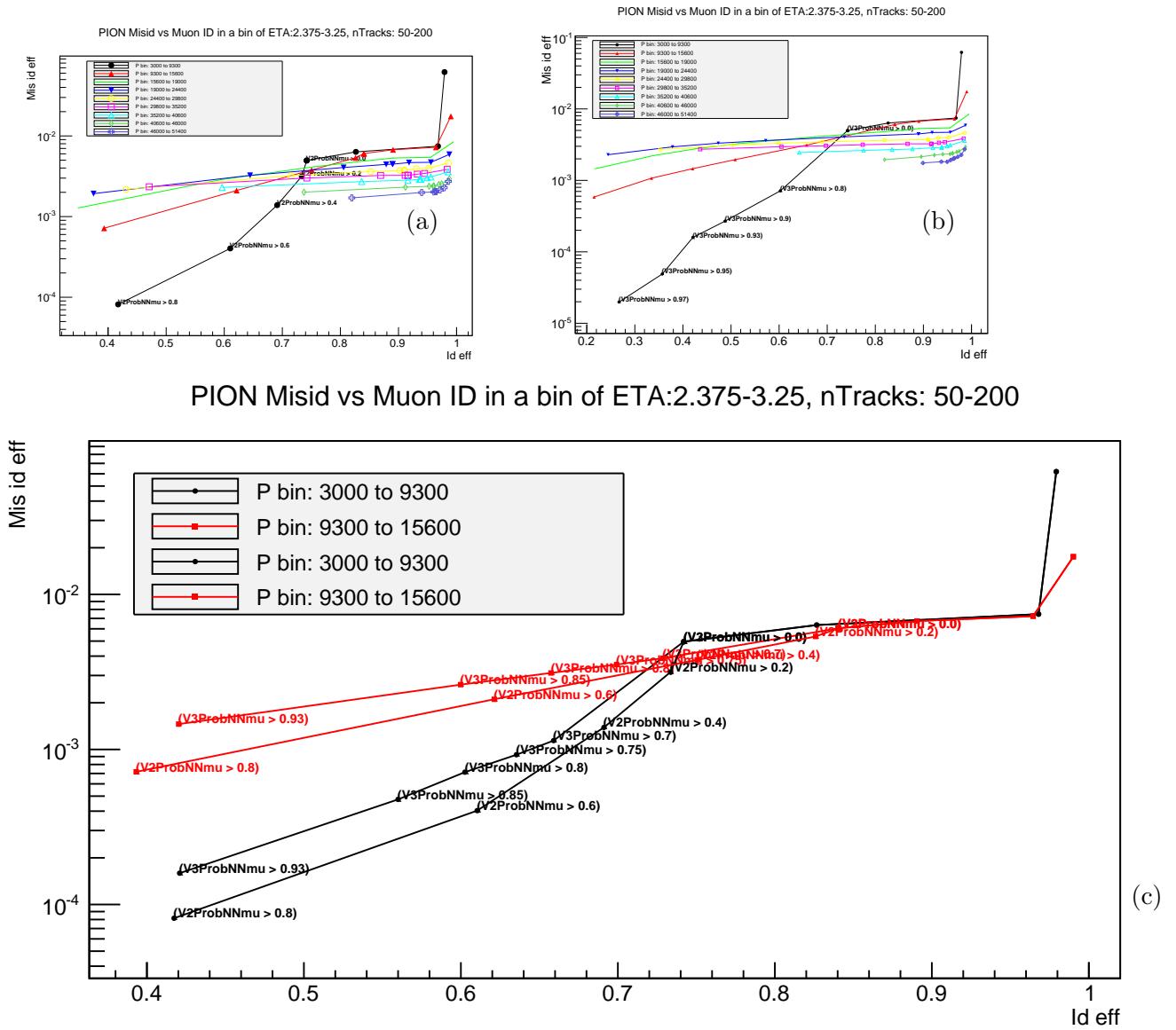


Figure 58: (a) Pion misID probability as a function of id efficiency for Probnnmu V2 (b) and pion misID probability as a function of id efficiency for Probnnmu V3 in first 9 bins. (c) Comparing directly in one plot Probnnmu V2 and Probnnmu V3 for first two bins. As for the same id efficiency the misID Probnnmu V2 is lower, this tuning is chosen.

1359 13 Expected Exclusion Limits - Sensitivity studies

1360 The CLs method [15] is used to produce upper limits. Since the prediction for the branching
1361 fraction is unknown, different branching fraction hypotheses are tested. The expected number of
1362 signal events N_s is calculated using the normalization constant for each branching fraction. The
1363 exclusion limits are shown in Figure 59.

1364 In practice, for a given branching fraction:

- 1365 • Choice of test statistics: profile likelihood ratio: $-2 \ln \frac{\mathcal{L}(s+b)}{\mathcal{L}(b)}$, where $\mathcal{L}(s+b)$ is the likelihood
1366 function under signal and background hypothesis and $\mathcal{L}(b)$ is likelihood under background
1367 only hypothesis. Hence test statistics in this case represents ratio of likelihoods.
- 1368 • Generate 1000 toy datasets no signal injected - bkg_data (pdf for test statistics under
1369 background hypothesis only)
- 1370 • Generate 1000 toy datasets with a given signal branching fraction injected - sig_data (pdf
1371 for test statistics under signal+background hypothesis)
- 1372 • Fit background hypothesis and signal+background hypothesis pdfs to both bkg_data and
1373 sig_data
- 1374 • Extract bkg_ΔLL distribution for test statistics and sig_ΔLL for test statistics under
1375 signal+background hypothesis
- 1376 • Find the median of background bkg_ΔLL
- 1377 • Get CL_b by integrating bkg_ΔLL from the tail to the median. CL_b is the probability to
1378 obtain a result less compatible with the signal than the observed one in the background-only
1379 hypothesis.
- 1380 • Get CL_{s+b} by integrating sig_ΔLL from the tail to the median. CL_{s+b} is the probability to
1381 obtain a result which is less compatible with the signal than the observed result, assuming
1382 the signal hypothesis.
- 1383 • $CL_s = CL_{s+b}/CL_b$

1384 In these studies limits were produced for fits with partial systematics and with full systematics
1385 for both simultaneous and non-simultaneous fit.

1386 Apart from all statistical errors which are propagated into the fit as gaussian constraints,
1387 partial systematics fit include constraints from the \mathcal{B} of the $B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+$ and
1388 $B^+ \rightarrow (D^0 \rightarrow K^+\pi^-\mu^+\mu^-)\mu^+\nu$ and systematics due PID weight for misID data.

1389 Full systematics additionally include: signal model systematics, systematics for trigger due
1390 to difference between simulation and data, trigger systematic due to incorrect emulation of
1391 HLT2 and kaon interaction probability and finally kinematic reweighting systematics. All of
1392 these systematics affect efficiency ratios $R_{NOFCME,LowFCME,HighFCME}^{21,26}$ and hence are added in
1393 quadrature and used as gaussian constraint on the relevant ratios. In the simultaneous fit, all of
1394 these systematics are assumed to be 100% correlated.

1395 As shown in Figure 59(f), the simultaneous fit to both Run 1 and 2016 data with full
1396 systematics allows to set limit up of $\mathcal{B} < 2.75 \times 10^{-8}$.

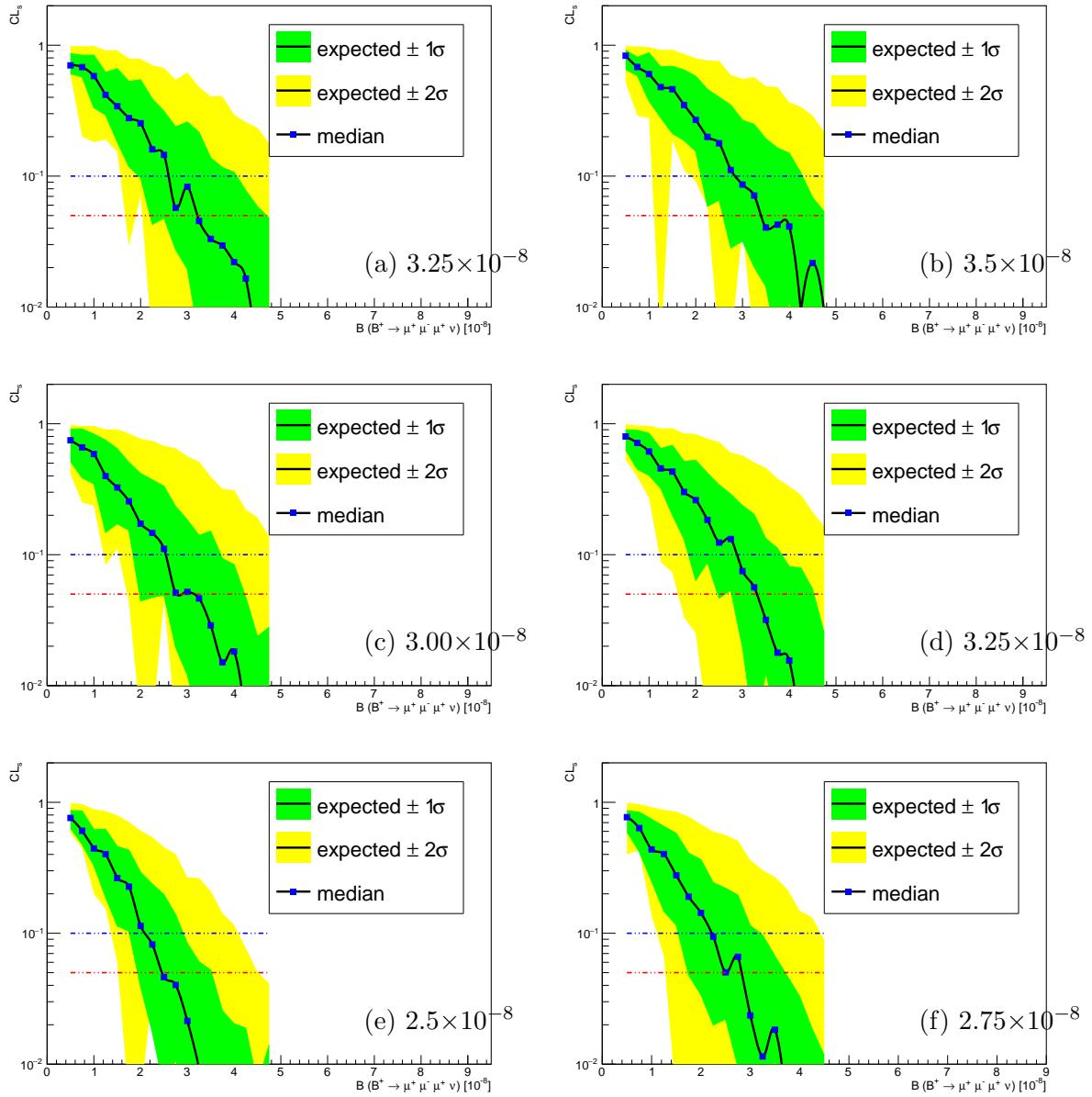


Figure 59: Expected 90% (blue horizontal line) 95% (red horizontal line) CL exclusion limits for full Run 1 data with simultaneous fit (a) with partial systematics and all statistical error constraints and (b) full systematics. Expected 90% (blue horizontal line) 95% (red horizontal line) CL exclusion limits for full Run 1 and 2016 data with non-simultaneous fit (c) with partial systematics and all statistical error constraints and (d) full systematics. And finally expected 90% (blue horizontal line) 95% (red horizontal line) CL exclusion limits for full Run 1 and 2016 data with simultaneous fit (e) with partial systematics and all statistical error constraints and (f) full systematics As expected full systematics sensitivity always yields slightly worse limits.

1397 **14 Results**

1398 **14.1 Unblinded Mass Fits**

1399 After unblinding the signal region, no significant signal was observed.

1400 The simultaneous fit was run to fit the observed dataset. The fit with background-only
1401 hypothesis describes the data well and can be seen in Figure 60.

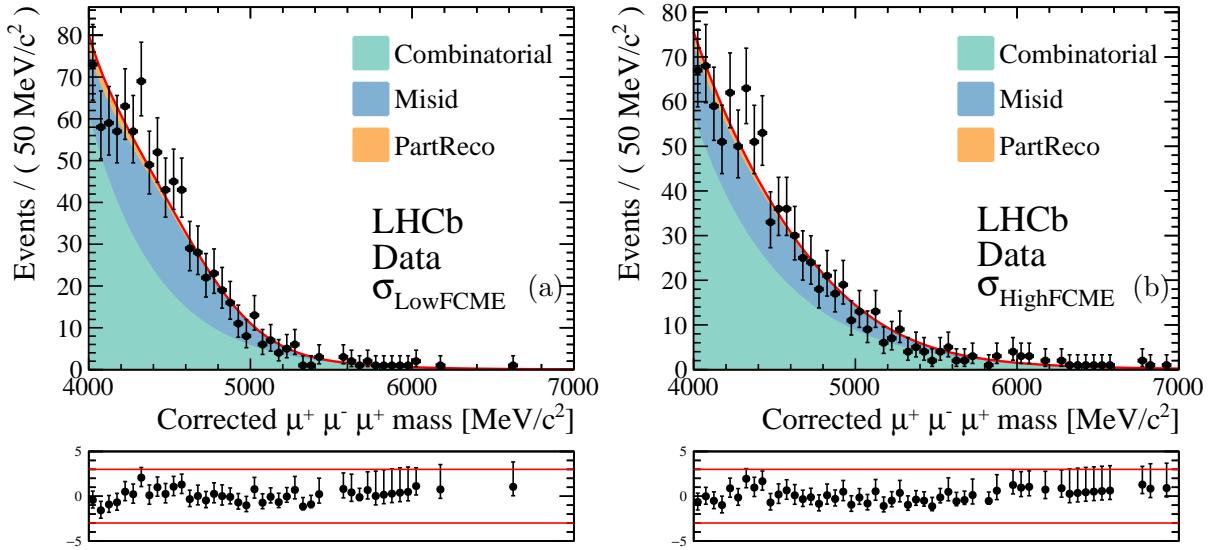


Figure 60: Simultaneous unbinned maximum likelihood fit to unblinded data after full selection chain in two bins of FCME and with $\mathcal{B} = 0$, with (b) fit to σ_{low} bin, (c) σ_{high} bin.

1402 In order to perform fit with signal and background hypothesis good range of the \mathcal{B} needs to
1403 be established. This is done by maximizing the log likelihood values for different \mathcal{B} (also known
1404 as profiling the \mathcal{B}). The results of the profiling can be seen in Figure 61. This shows that the
1405 minimized $-\log(L)$ has a minimum at $\mathcal{B} \sim -2.0 \times 10^{-8}$, hence negative values of \mathcal{B} are preferred
1406 by the fit (negative fluctuation).

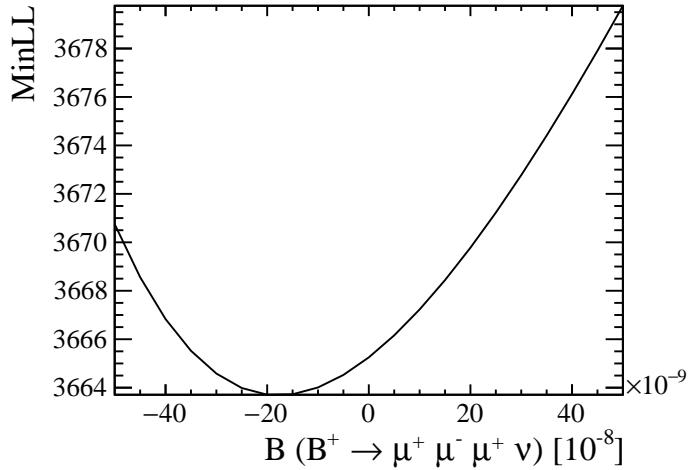


Figure 61: Minimized $-\log(L)$ value at different \mathcal{B} .

With the \mathcal{B} floating the data mass fits results are shown in Figure 62 converging at value of $\mathcal{B} = -1.78 \times 10^{-8}$. The full list of all floating parameters in the fit with its values are shown in Table 64. This fit results then can be translated into yields where the number of signal events, N_{sig} , are obtained by Equation 41 and the number of partially reconstructed events are calculated as in Equation 42. The resulting yields in all the fitting range and specifically blinded region are shown in Table 65.

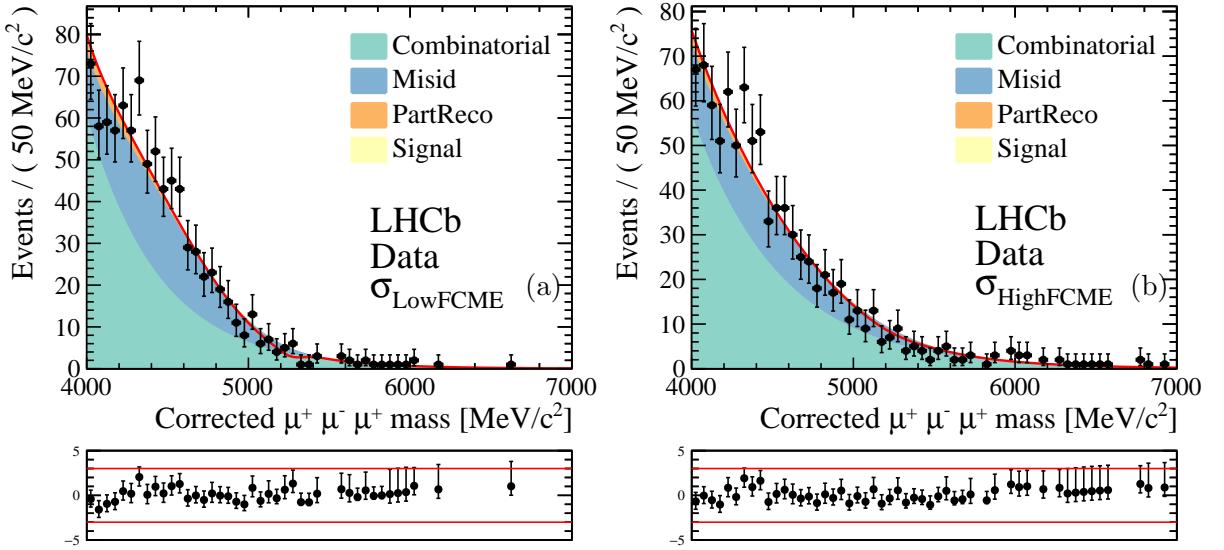


Figure 62: Simultaneous unbinned maximum likelihood fit to unblinded data after full selection chain in two bins of FCME, with (b) fit to σ_{low} bin, (c) σ_{high} bin.

Fit Parameter	Value
$\mathcal{B}(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu)$	$-(1.78 \pm 0.936) \times 10^{-8}$
$\mathcal{B}(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+)$	$(6.11 \pm 0.185) \times 10^{-5}$
$\mathcal{B}(B^+ \rightarrow (D^0 \rightarrow K^+ \pi^- \mu^+ \mu^-) \mu^+ \nu)$	$(4.12 \pm 0.5) \times 10^{-7}$
$\mu_{misID_{high}}$	4100 ± 123
$\mu_{misID_{low}}$	4330 ± 56
$\sigma_{misID_{high}}$	619 ± 56
$\sigma_{misID_{low}}$	401 ± 29.9
$R_{high}^{26}(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu)$	1.97 ± 0.0788
$R_{low}^{26}(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu)$	3.31 ± 0.1
$R_{high}^{21}(B^+ \rightarrow (D^0 \rightarrow K^+ \pi^- \mu^+ \mu^-) \mu^+ \nu)$	0.0397 ± 0.00116
$R_{low}^{21}(B^+ \rightarrow (D^0 \rightarrow K^+ \pi^- \mu^+ \mu^-) \mu^+ \nu)$	0.0271 ± 0.000794
$R_{high}^{21}(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu)$	1.96 ± 0.131
$R_{low}^{21}(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu)$	3.15 ± 0.233
$N(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+)_{high}^{2016}$	29800 ± 173
$N(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+)_{low}^{2016}$	64700 ± 254
$N(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+)_{high}^{Run1}$	64100 ± 253
$N(B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+)_{low}^{Run1}$	109000 ± 331
$N_{misID_{high}}^{scaled}$	278 ± 15.2
$N_{misID_{low}}^{scaled}$	331 ± 16.3
β_{high}	$-0.00183 \pm 9.19 \times 10^{-5}$
β_{low}	-0.00226 ± 0.000129
$N_{combi_{high}}$	620 ± 34.3
$N_{combi_{low}}$	531 ± 35.4

Table 64: Fit results for all floating variables in the unblinded data fit. The $\mathcal{B}(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu) = -1.78 \times 10^{-8}$. Variables R_K^S are the efficiency ratios obtained by normalising the decays to $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+$ decays where S stands for stripping and K for the FCME. $N_{misID_K}^{scaled}$ is the number of misID events, N_{combi_K} is the number of combinatorial events, β_K is the exponential constant, μ_{misID_K} and σ_{misID_K} are the mean and the σ of the CB function.

Component	All ($4 - 7 \text{ GeV}/c^2$)	Signal region ($4.5 - 5.5 \text{ GeV}/c^2$)
$N_{misID_{low}}$	331	139
$N_{sig_{low}}$	-15.8	-14.4
$N_{combi_{low}}$	531	154
$N_{partreco_{low}}$	31.8	6.39
$N_{misID_{high}}$	279	122
$N_{sig_{high}}$	-14.0	-11.0
$N_{combi_{high}}$	620	209
$N_{partreco_{high}}$	25.1	5.17

Table 65: Resulting yields for different components from the corrected mass fit with $\mathcal{B}(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu) = -1.78 \times 10^{-8}$.

14.2 Limit Setting

As no significant signal is observed, limit for the $\mathcal{B}(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu)$ is set. This is achieved by using the CLs method described in section 13 and the CLs p values are shown in Figure 63 together with both expected and observed curves. This gives following limits summarised in Table 66, setting the limit $\mathcal{B}(B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu) < 1.1(1.4) \times 10^{-8}$ at 90%(95%) CL. As it can be seen, the observed limit is better than the expected limit.

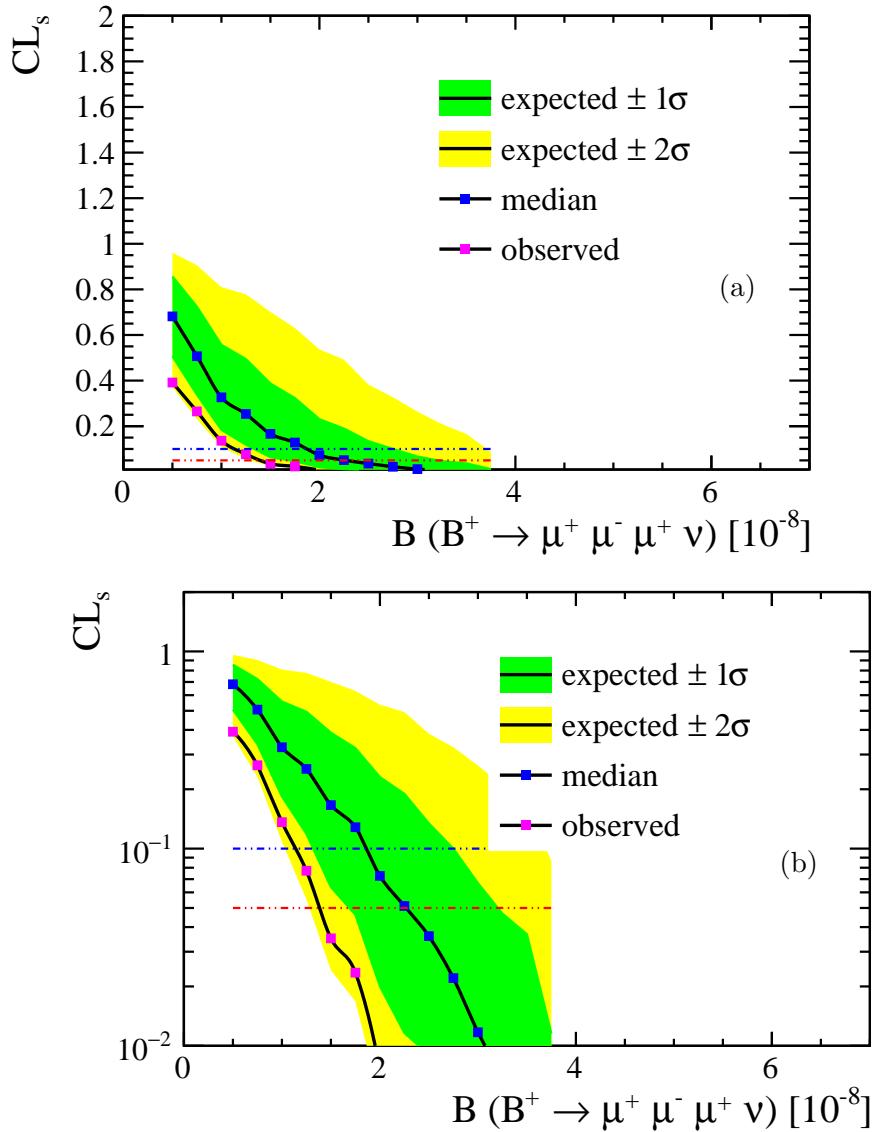


Figure 63: Expected and observed 90% (blue horizontal line) 95% (red horizontal line) CL exclusion limits for full Run 1 and 2016 simultaneous data fit accounting for all systematics with (a) normal and (b) logarithmic y-axis.

Expected/Observed	CI	Value
Expected	90% CL	1.9×10^{-8}
Observed	90% CL	1.1×10^{-8}
Expected	95% CL	2.3×10^{-8}
Observed	95% CL	1.4×10^{-8}

Table 66: Resulting exclusion limits with simultaneous fit.

1419 14.3 Additional Results: Non-simultaneous fit limits

1420 The same procedure was repeated but using non-simultaneous fit. The limits obtained using
 1421 non-simultaneous fit are summarized in Table 67. As it can be seen, expected limits are worse
 1422 compared to the simultaneous fit, but the observed limits are compatible with those seen in
 1423 simultaneous fit.

Expected/Observed	CI	Value
Expected	90% CL	2.3×10^{-8}
Observed	90% CL	0.9×10^{-8}
Expected	95% CL	2.7×10^{-8}
Observed	95% CL	1.4×10^{-8}

Table 67: Resulting exclusion limits with non-simultaneous fit.

1424 15 Conclusion

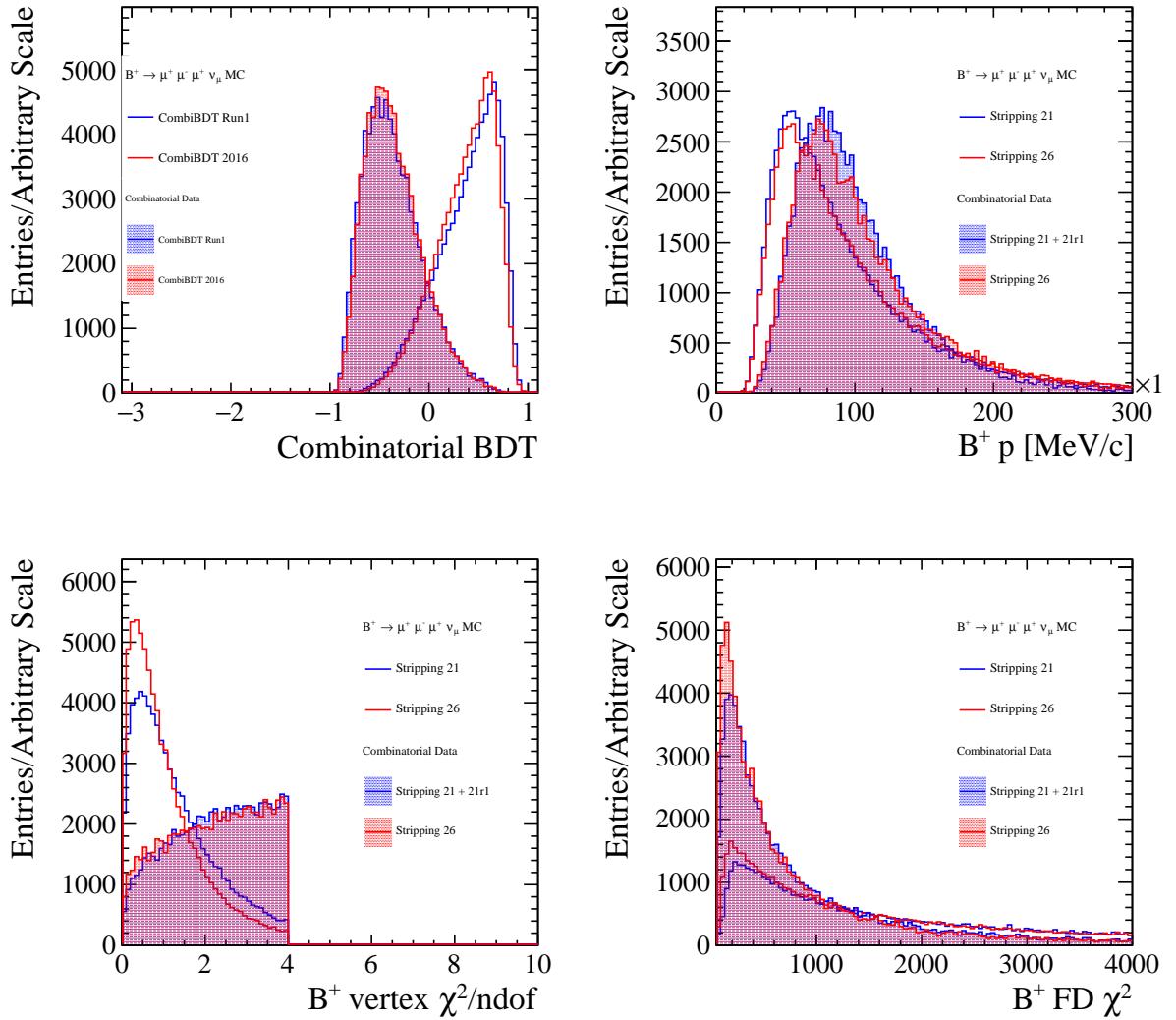
1425 In this analysis first attempt to measure the branching fraction of rare fully-leptonic $B^+ \rightarrow$
 1426 $\mu^+\mu^-\mu^+\nu$ decay was described, using $\approx 4.7\text{fb}^{-1}$ of LHCb data collected in 2011,2012 and 2016.
 1427 This measurement was performed in a particular region of dimuon mass $\min q < 980\text{ MeV}/c^2$.
 1428 Upon unblinding the signal region, no significant signal was seen. Hence, the corrected mass
 1429 spectrum was fitted with a simultaneous fit exploiting the resolution characteristics of background
 1430 and signal distributions and the limit was set using CLs method. The resulting upper limit on
 1431 branching fraction is $\mathcal{B}(B^+ \rightarrow \mu^+\mu^-\mu^+\nu) < 1.1(1.4) \times 10^{-8}$ at 90%(95%) confidence level. This
 1432 does not confirm the only prediction for this decay where the $\mathcal{B}(B^+ \rightarrow \mu^+\mu^-\mu^+\nu) \approx 1.3 \times 10^{-7}$
 1433 [7] [8].

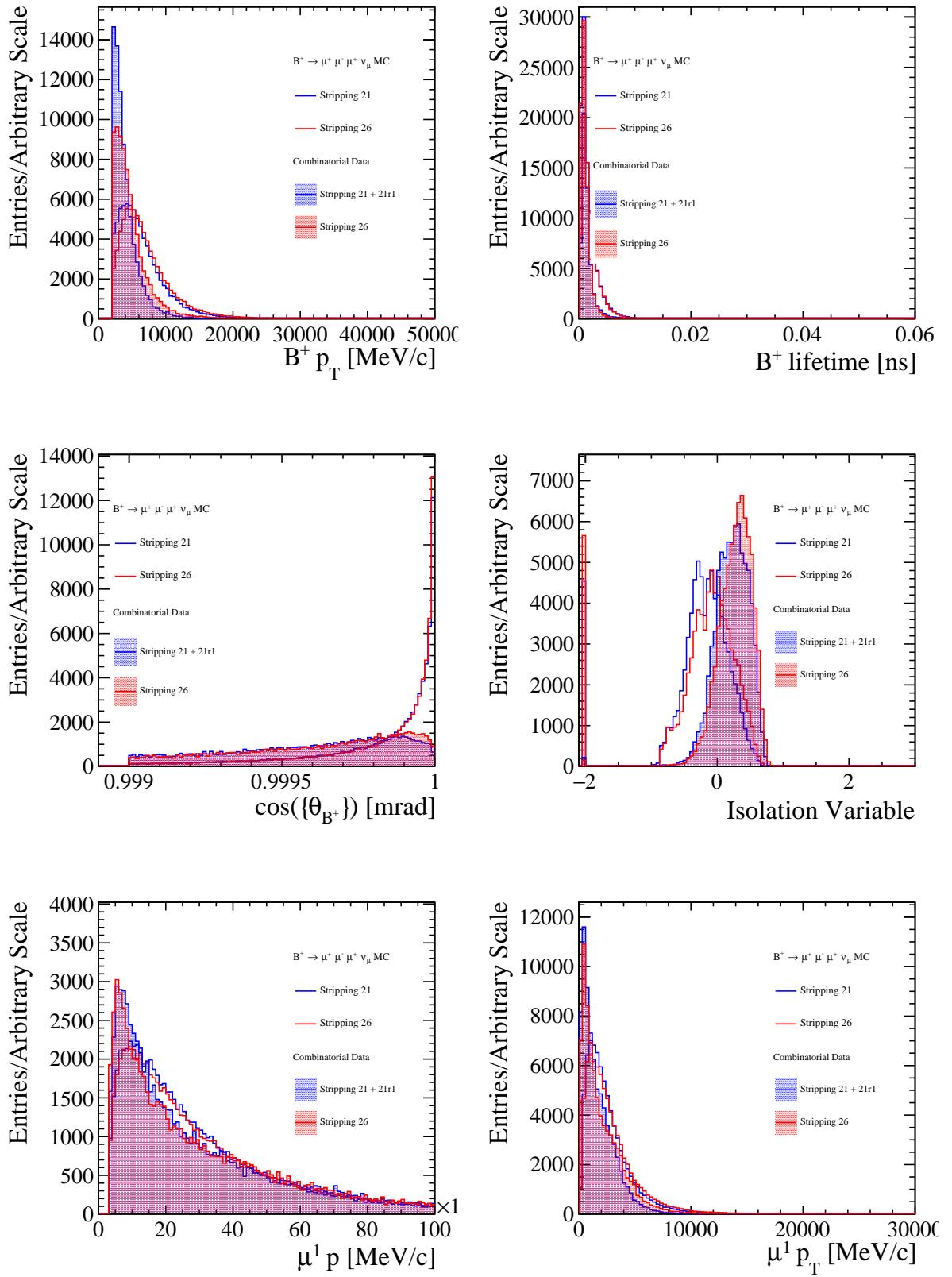
1434 **Appendices**

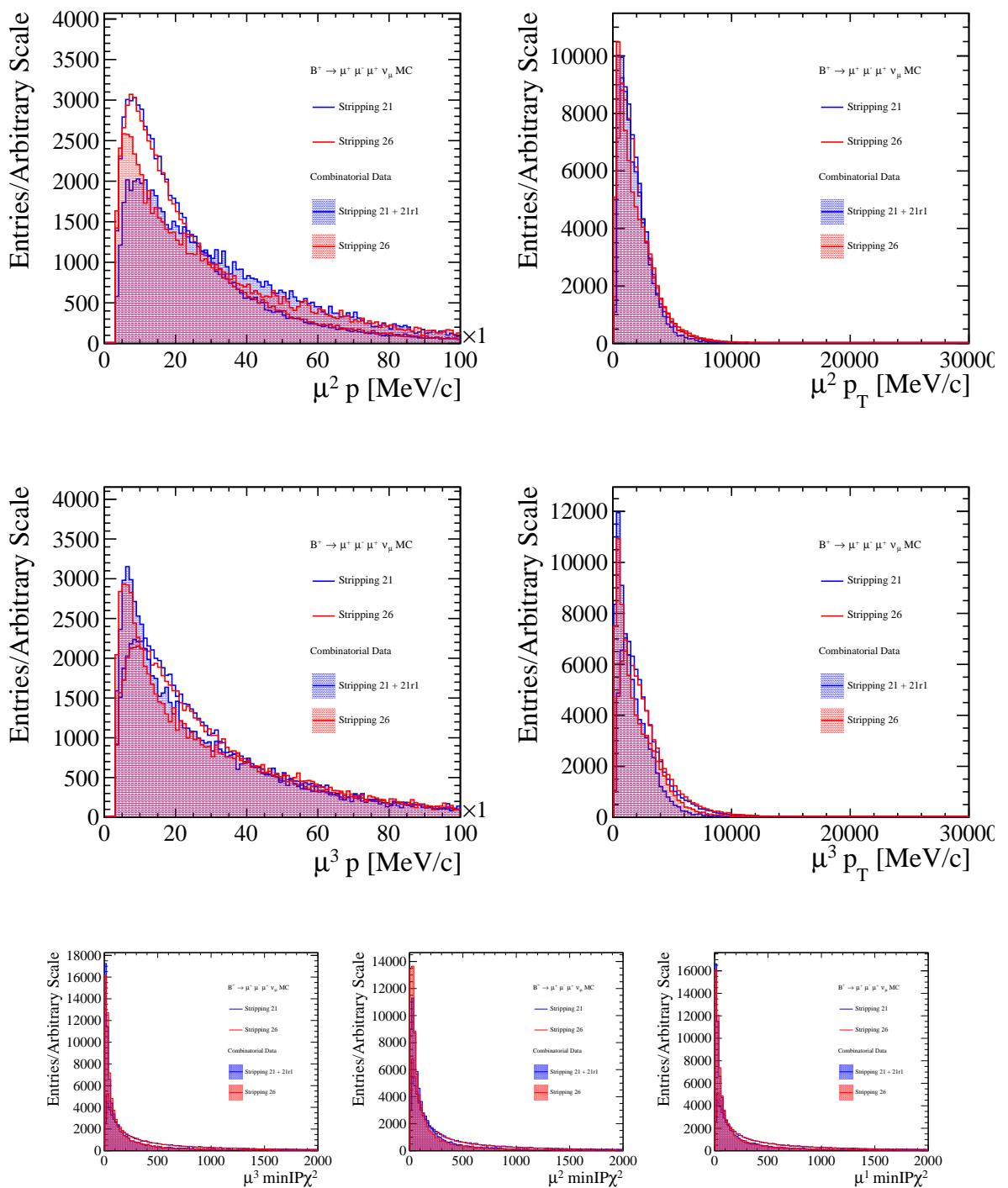
1435 **A BDT input variables Run 1 and 2016 - Combinatorial BDT**

1436 In this section, plots of BDT input variables and Combinatorial BDT output for Run 1 and 2016
 1437 are presented. In Appendix A.1, BDT input variables for both Combinatorial BDTs for Run 1
 1438 and 2016 are plotted for signal MC and combinatorial data sample. In these plots μ^1 and μ^3 are
 1439 the muons with the charge that agrees with mother B , so if the mother is B^+ then μ^1 and μ^3 are
 1440 the positively charged muons (μ^+), and μ^2 is negatively charged muon (μ^-). These plots here are
 1441 for illustration of the distributions that go into Combinatorial BDTs. Comparison of solely the
 1442 distributions between Run 1 and 2016 for signal MC can be seen in Appendix A.2. Furthermore,
 1443 comparison of the distributions between Run 1 and 2016 combinatorial data samples can be seen
 1444 in Appendix A.3.

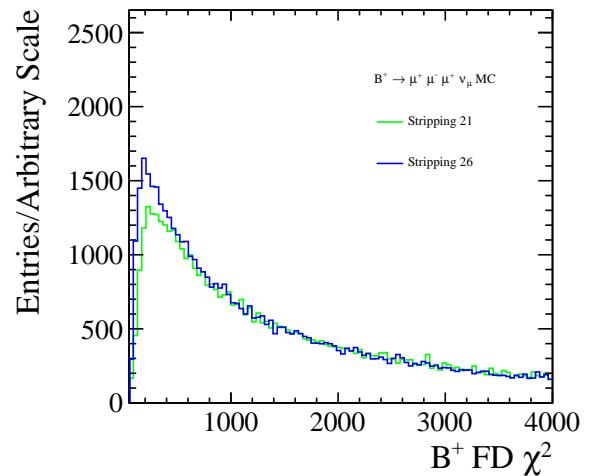
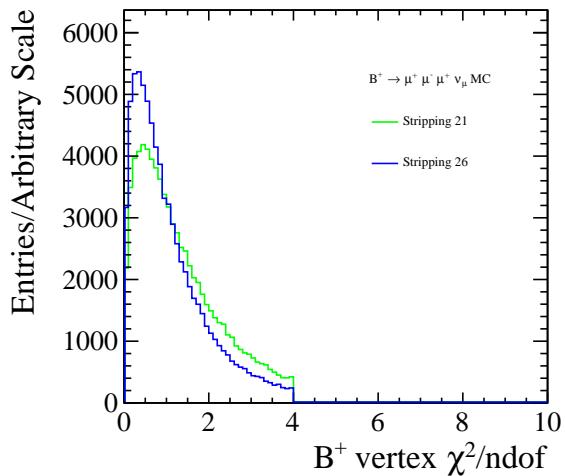
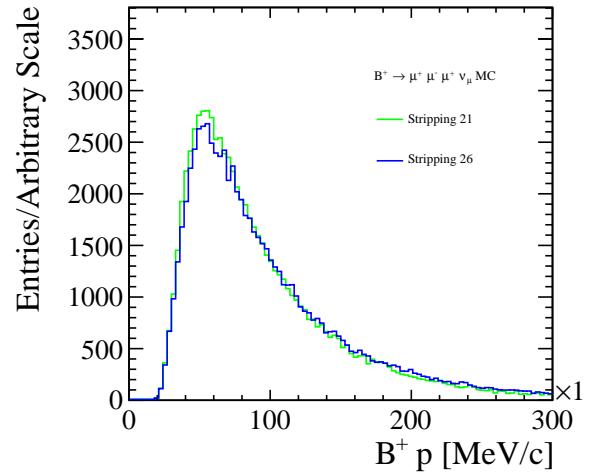
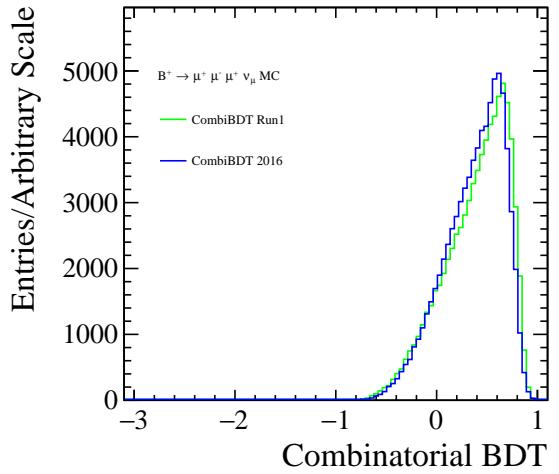
1445 **A.1 All**

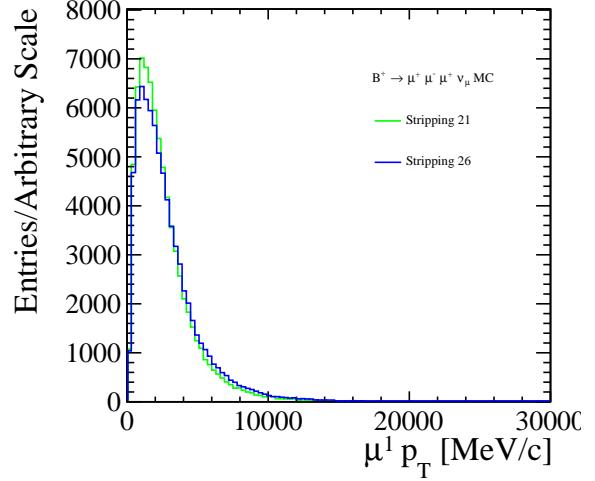
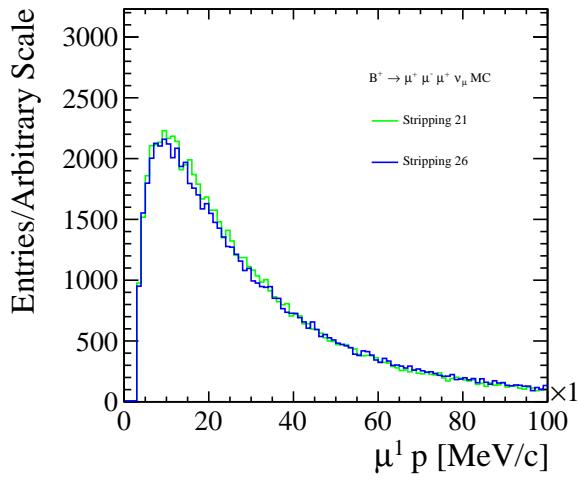
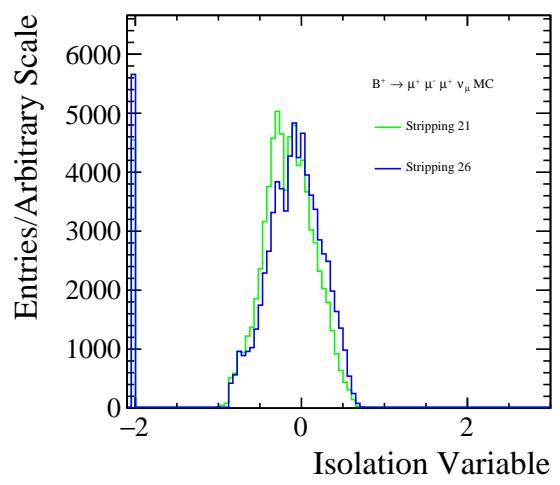
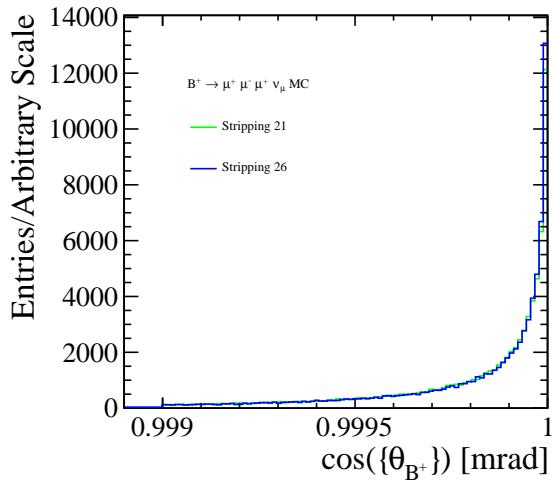
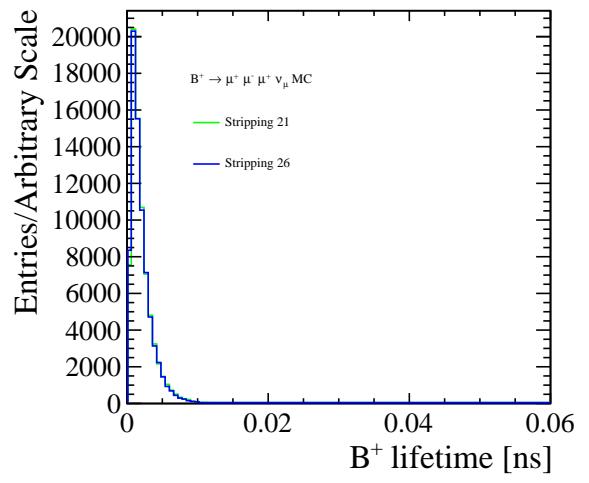
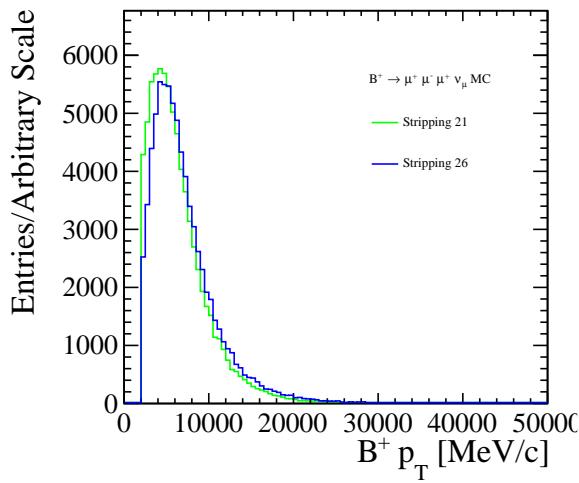


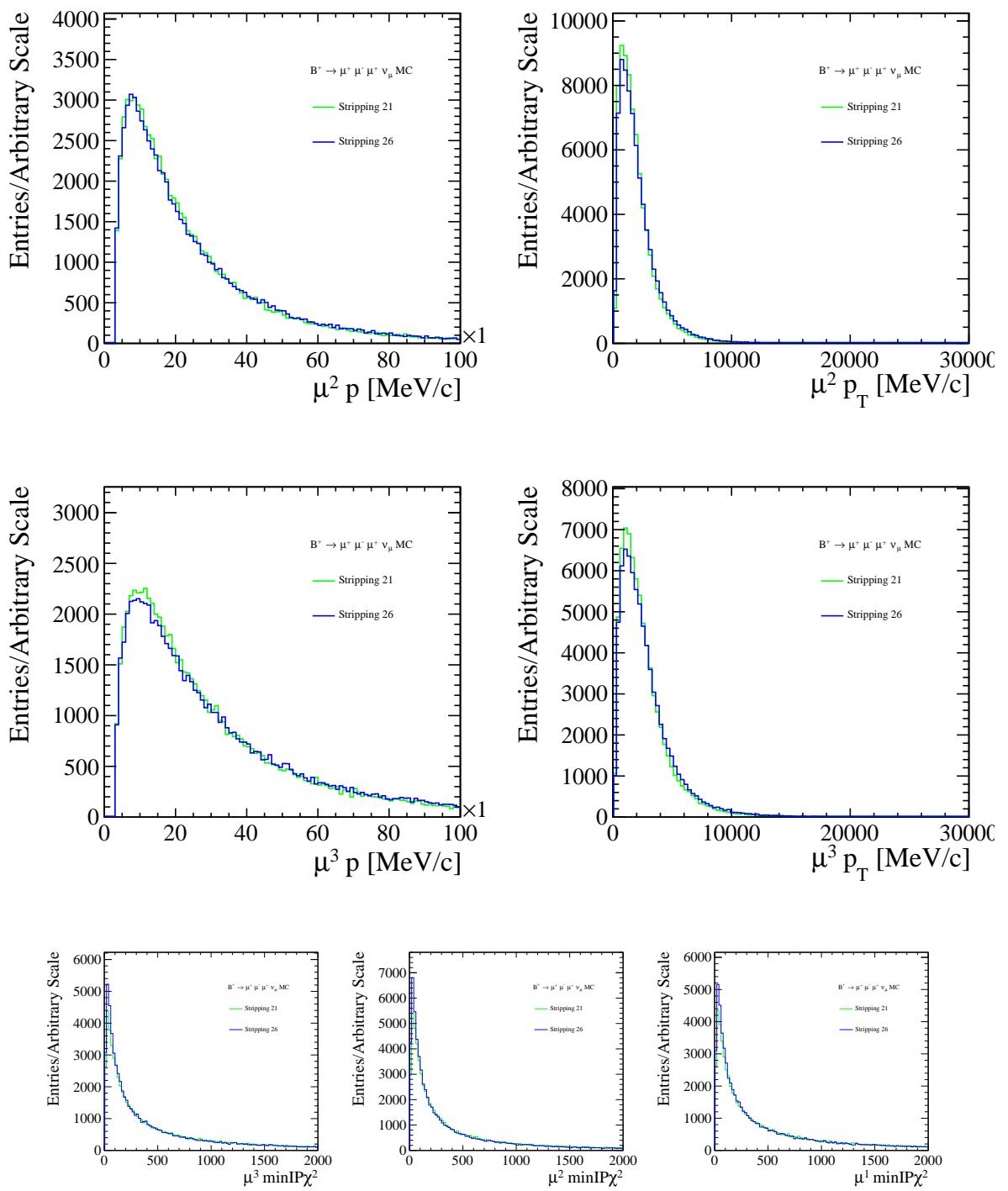




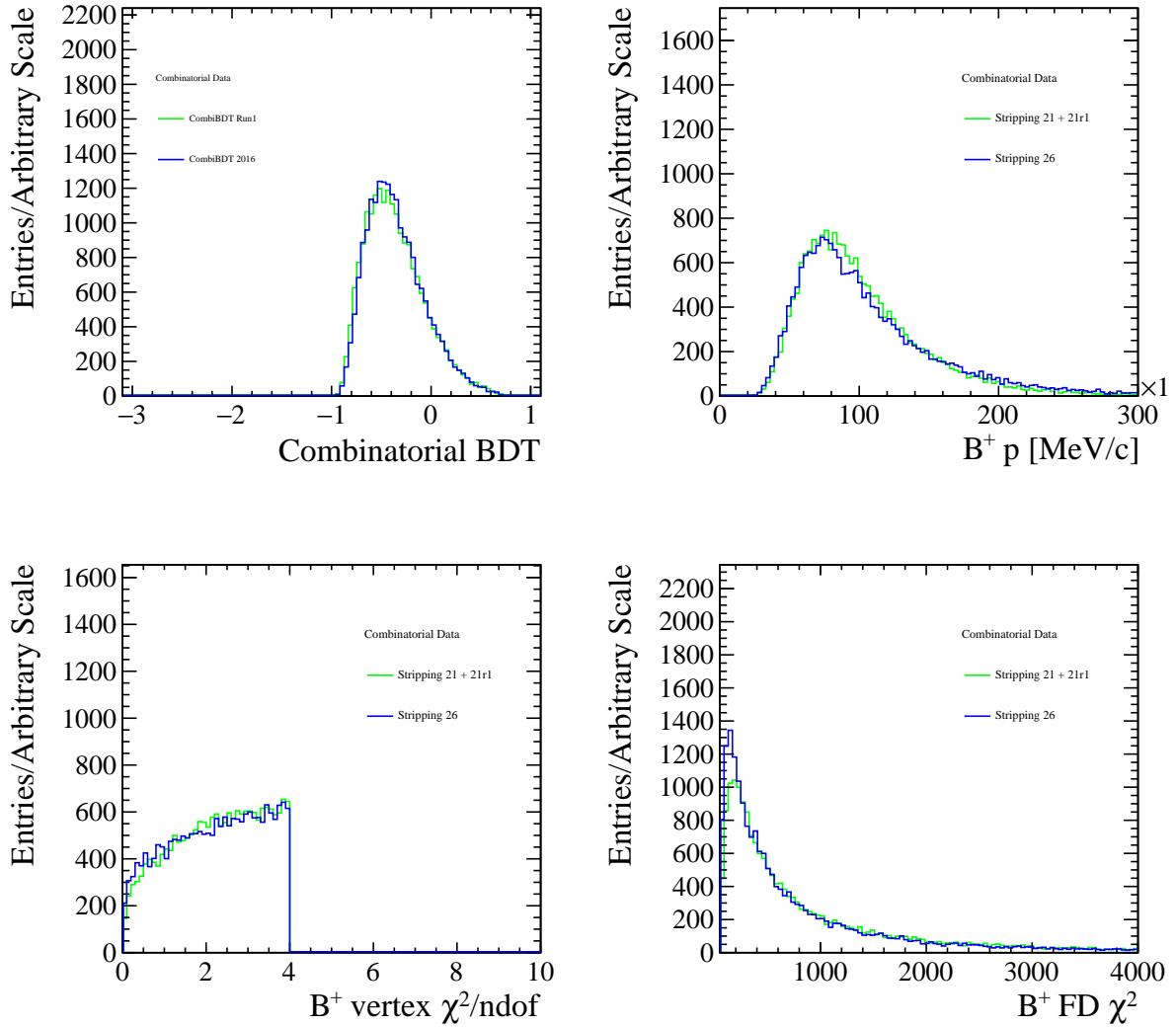
1446 A.2 Signal MC

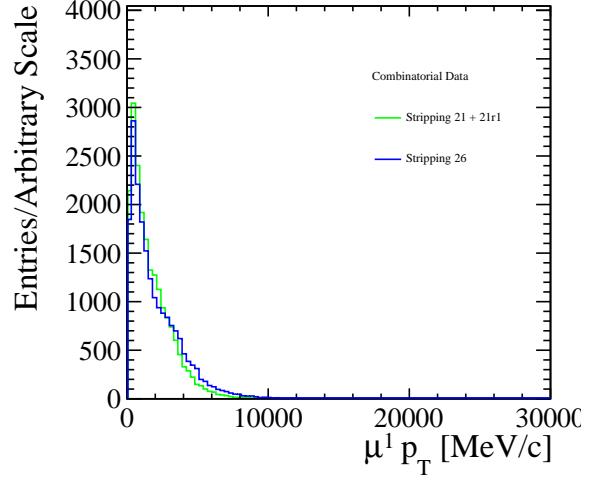
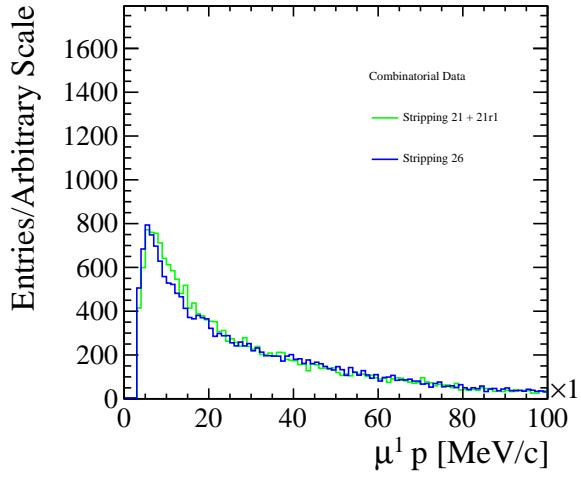
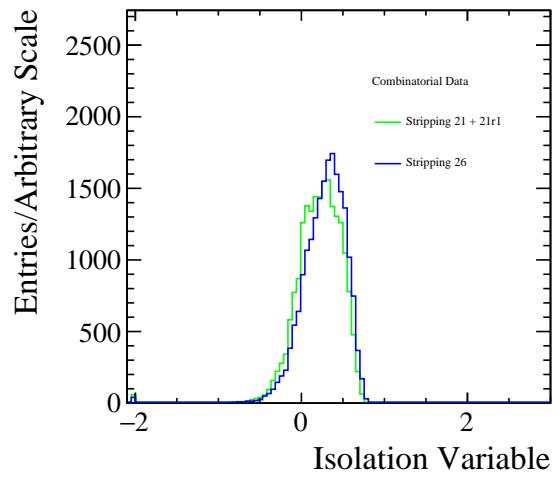
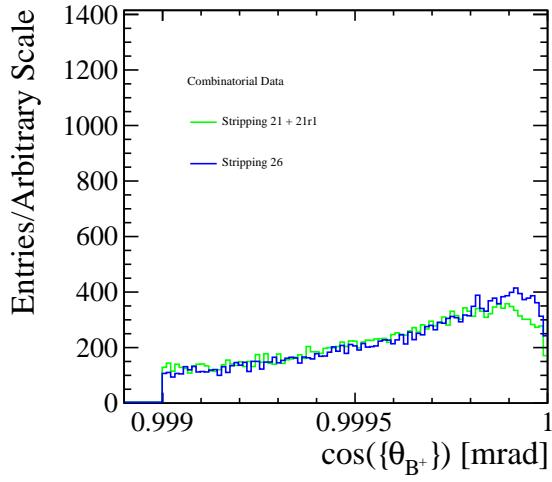
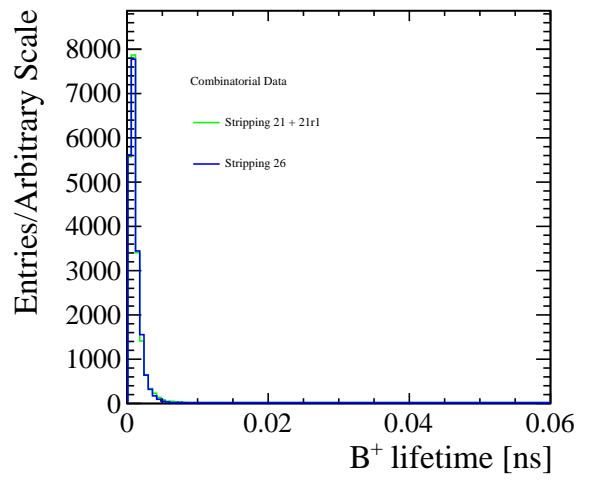
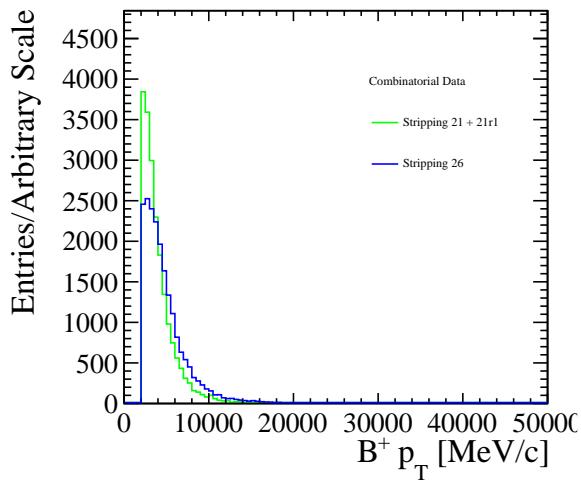


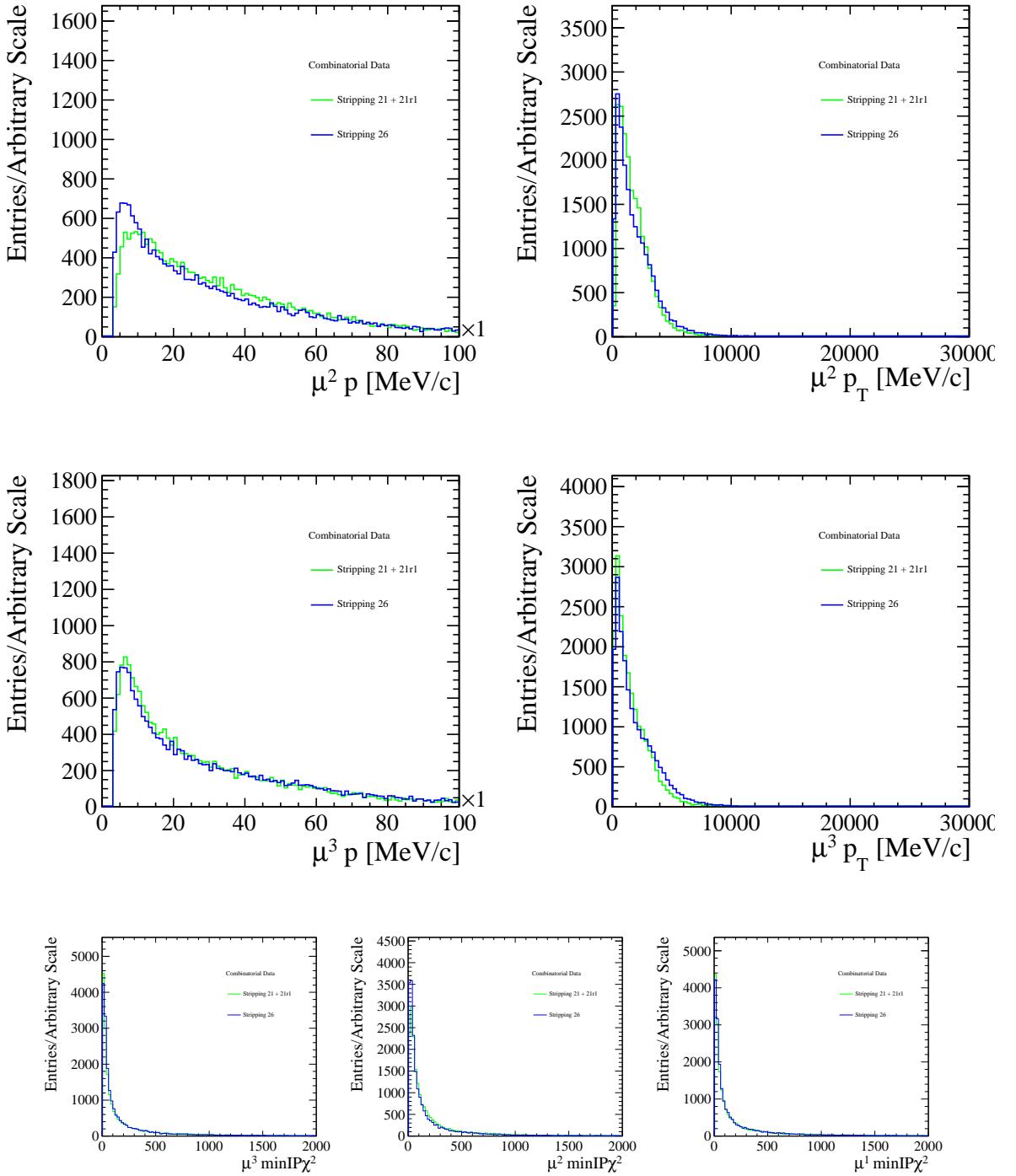




¹⁴⁴⁷ A.3 Combinatorial data





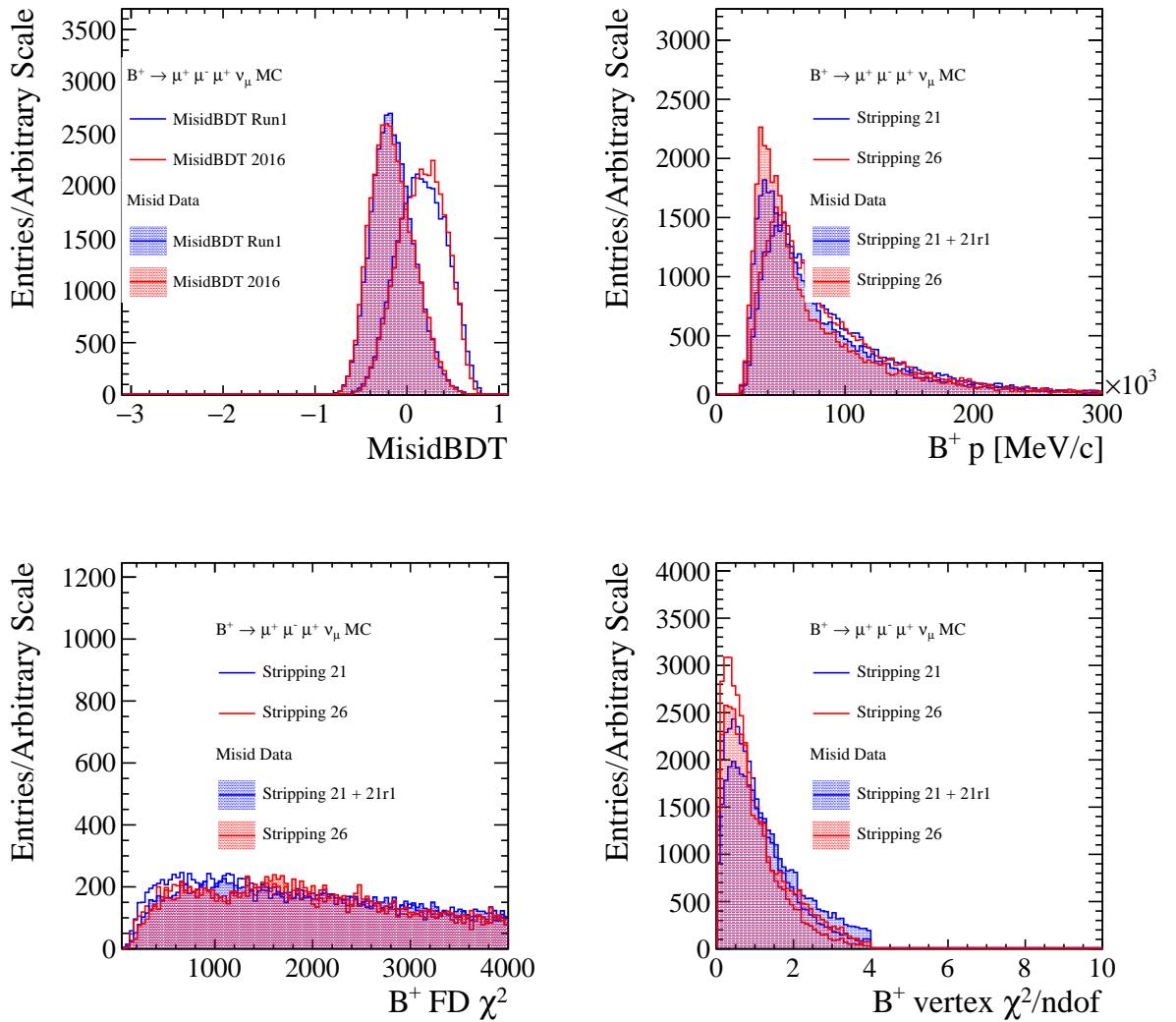


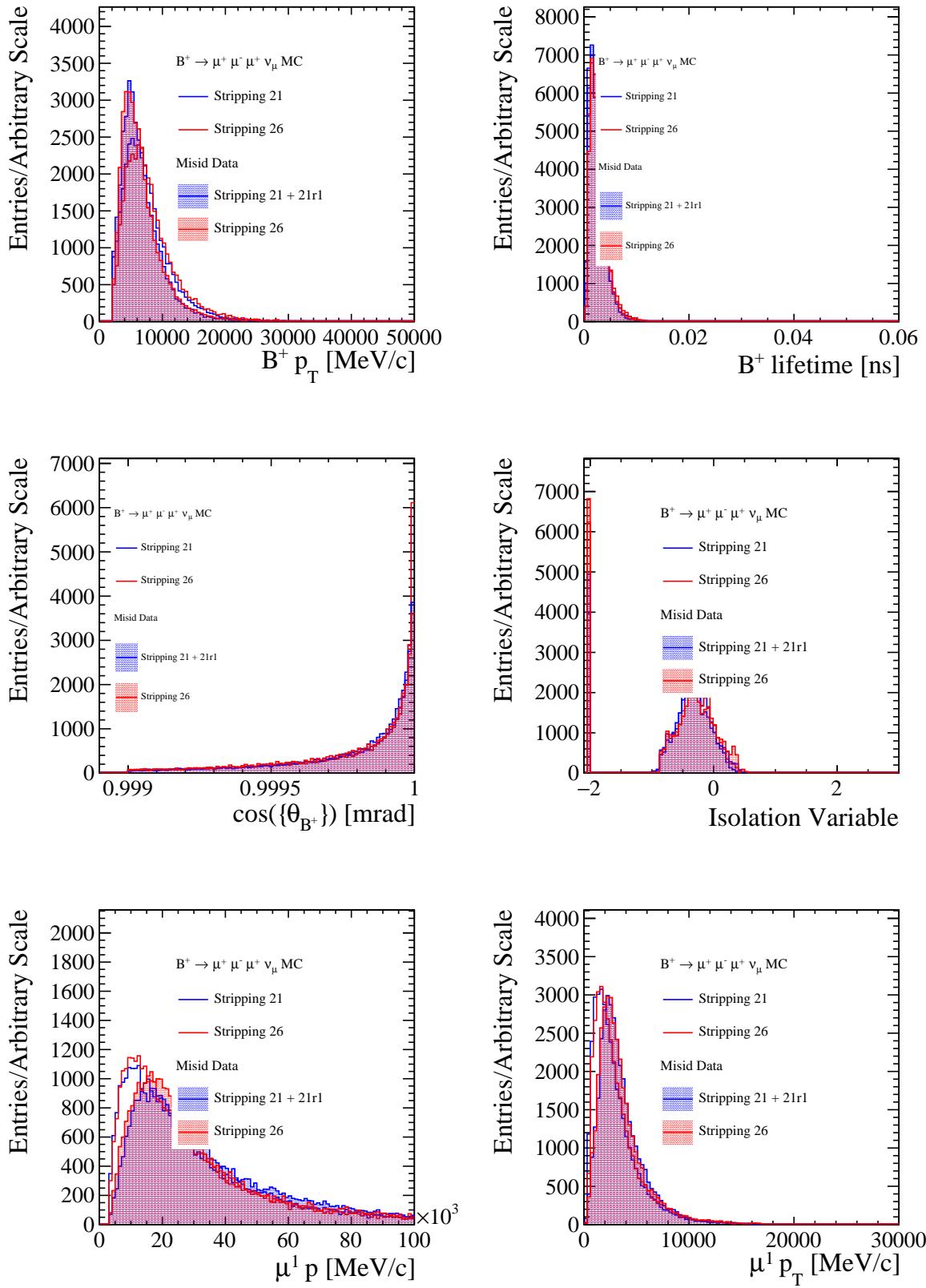
1448 B BDT input variables Run 1 and 2016 - Misid BDT

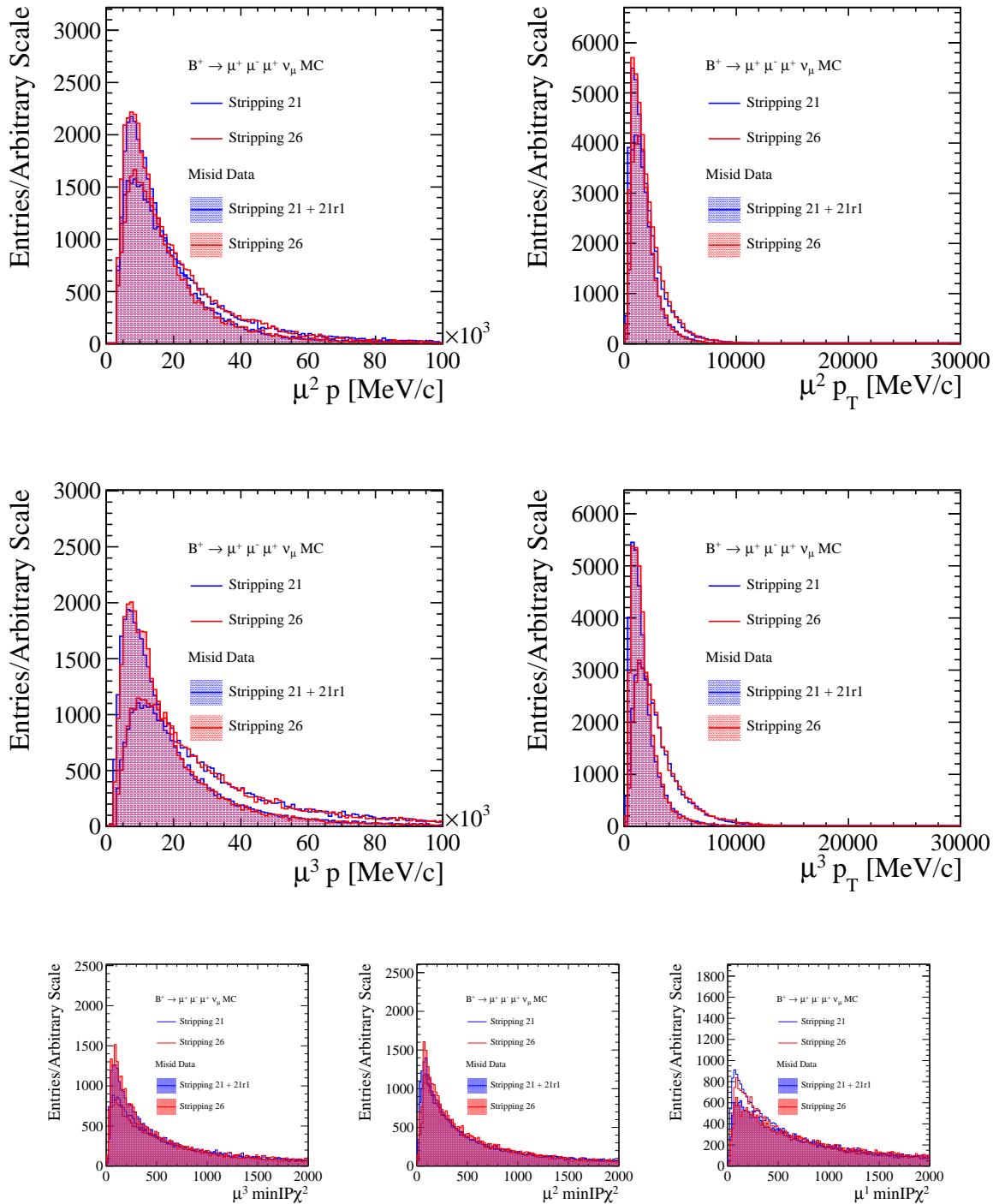
1449 In this section, plots of Misid BDT input variables and Misid BDT output for Run 1 and 2016
 1450 are presented. In Appendix B.1, BDT input variables for both Misid BDTs for Run 1 and 2016
 1451 are plotted for signal MC and misID SS data samples. In these plots, for signal samples μ^1 and
 1452 μ^3 are the muons with the charge that agrees with mother B , so if the mother is B^+ then μ^1 and

1453 μ^3 are the positively charged muons (μ^+), and μ^2 is negatively charged muon (μ^-). For misID
 1454 SS data samples, μ^1 is true muon with the charge that agrees with the charge of mother B , μ^2 is
 1455 true muon with the charge that does not agree with the charge mother B and finally μ^3 is the
 1456 misidentified particle whose charge agrees with the charge of mother B . These plots here are for
 1457 illustration of the distributions that go into Misid BDTs. Comparison of solely the distributions
 1458 between Run 1 and 2016 for signal MC can be seen Appendix B.2. Furthermore, comparison of
 1459 the distributions between Run 1 and 2016 misID data samples can be seen Appendix B.3.

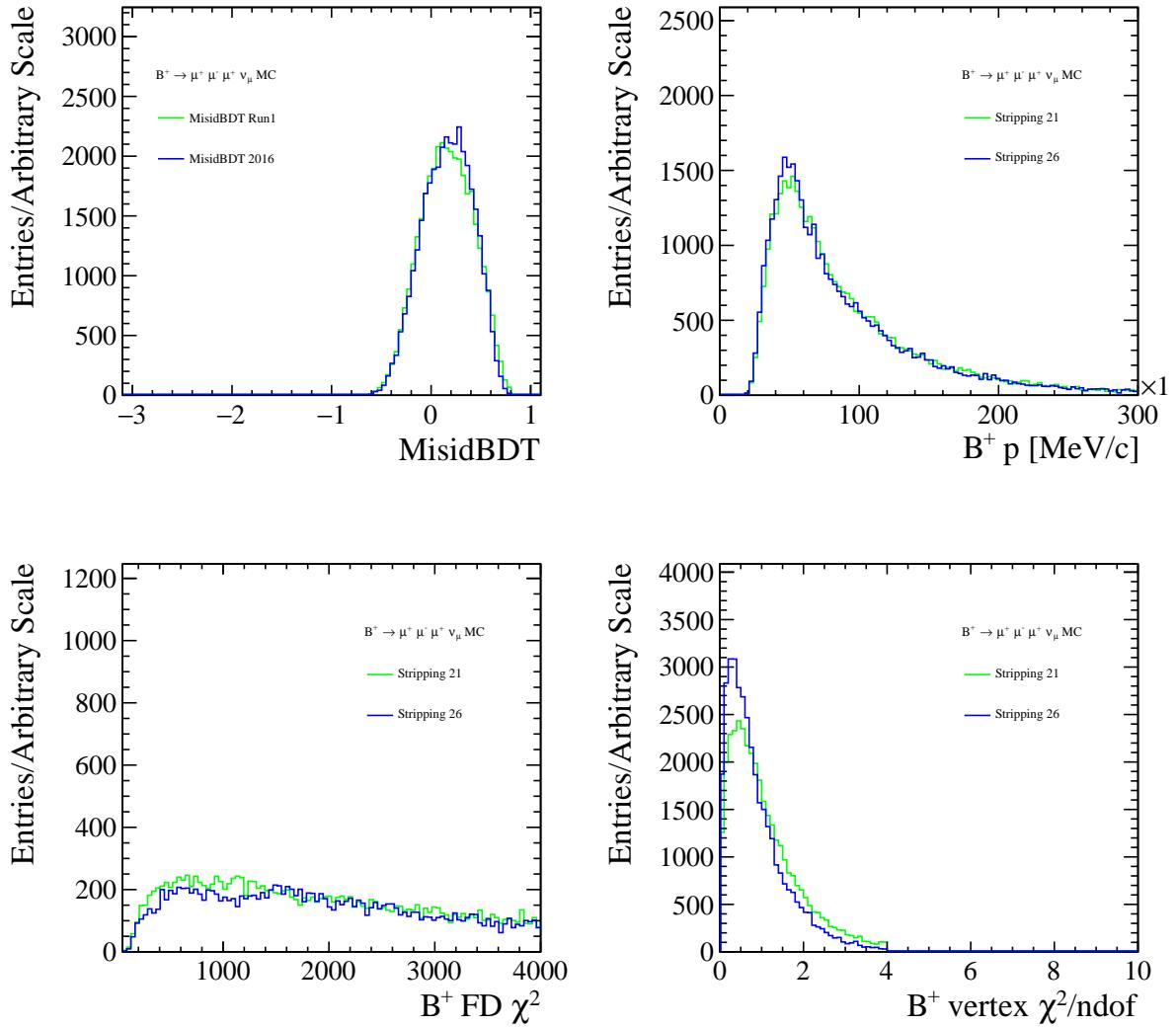
1460 B.1 All

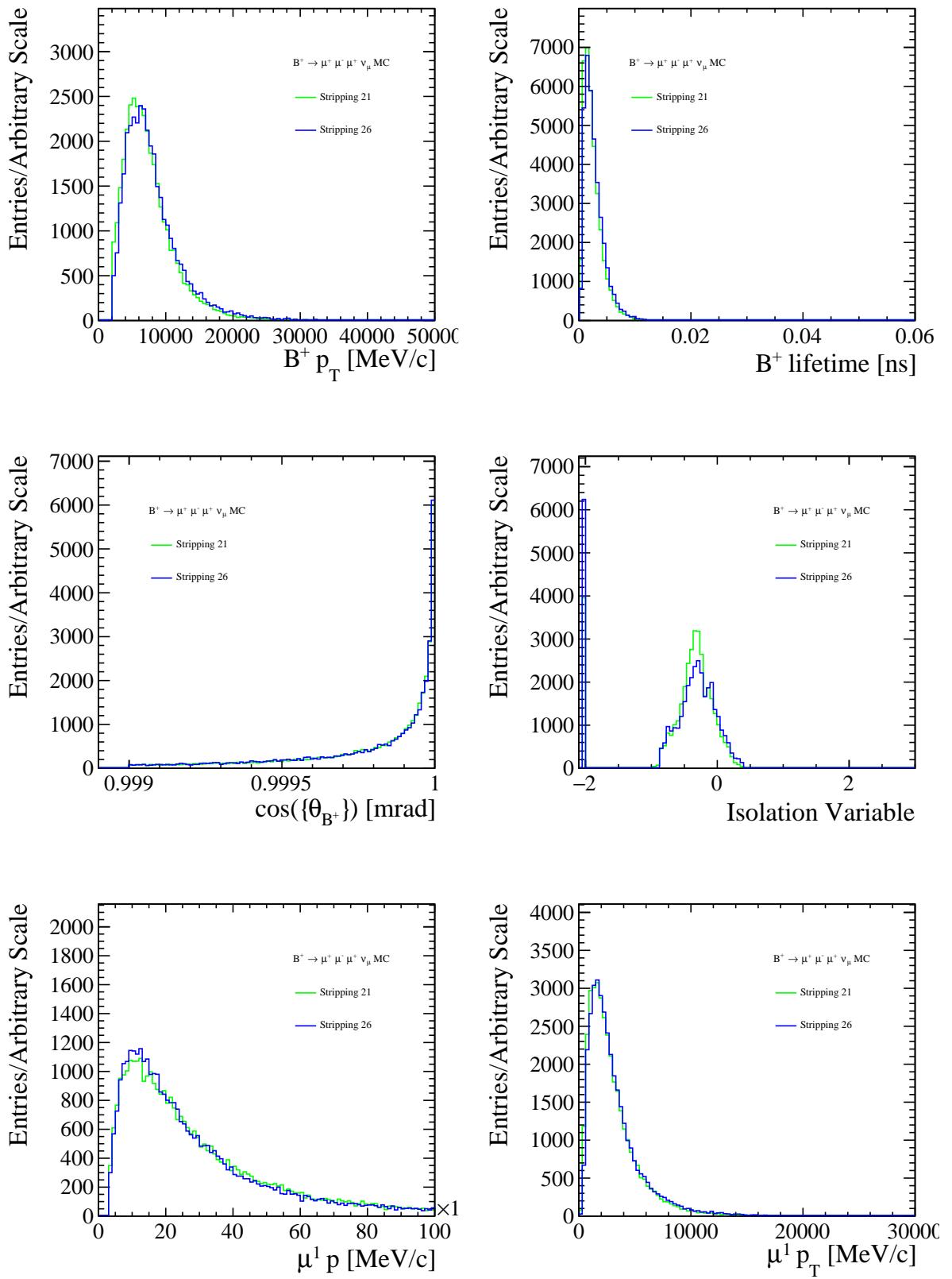


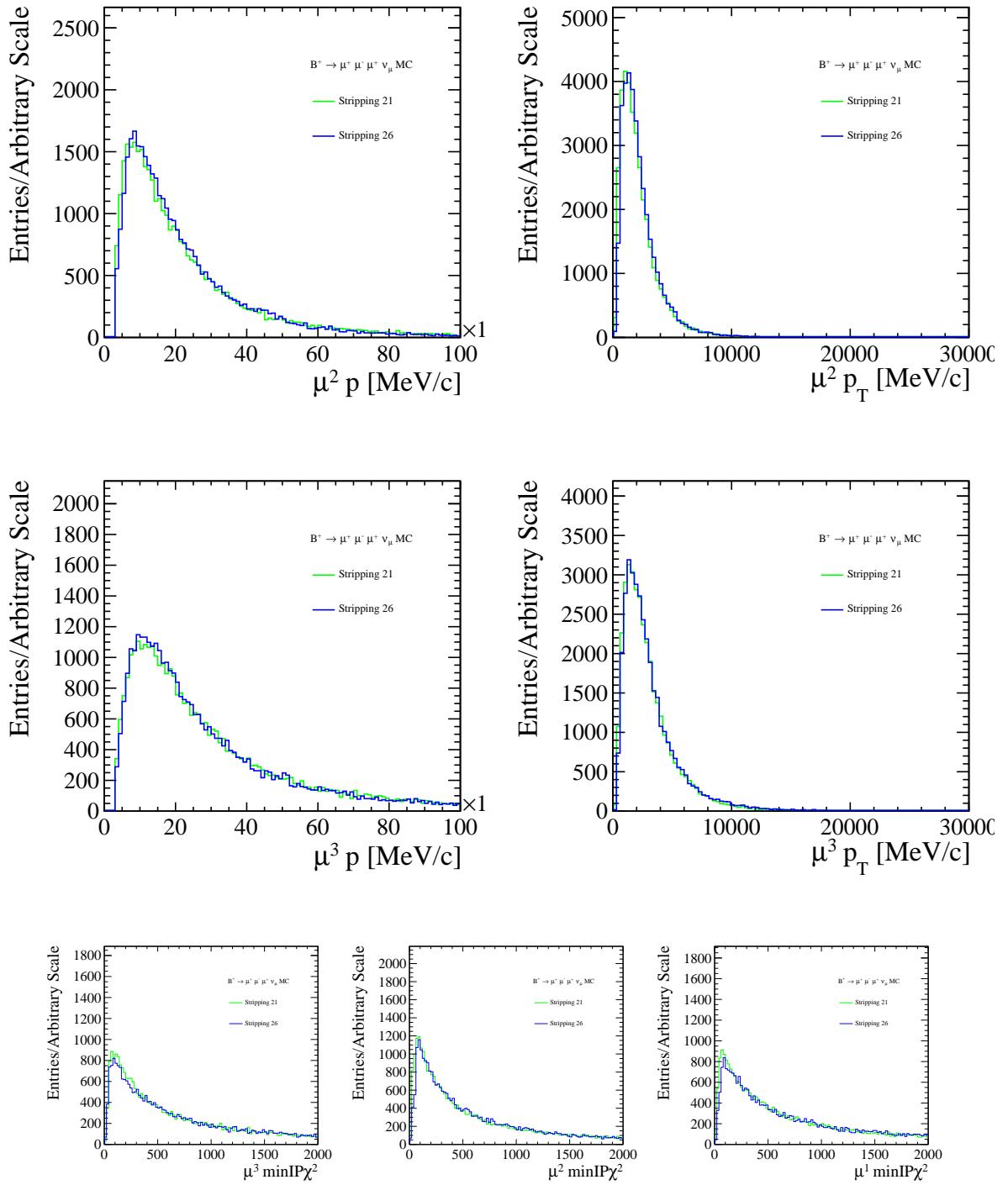




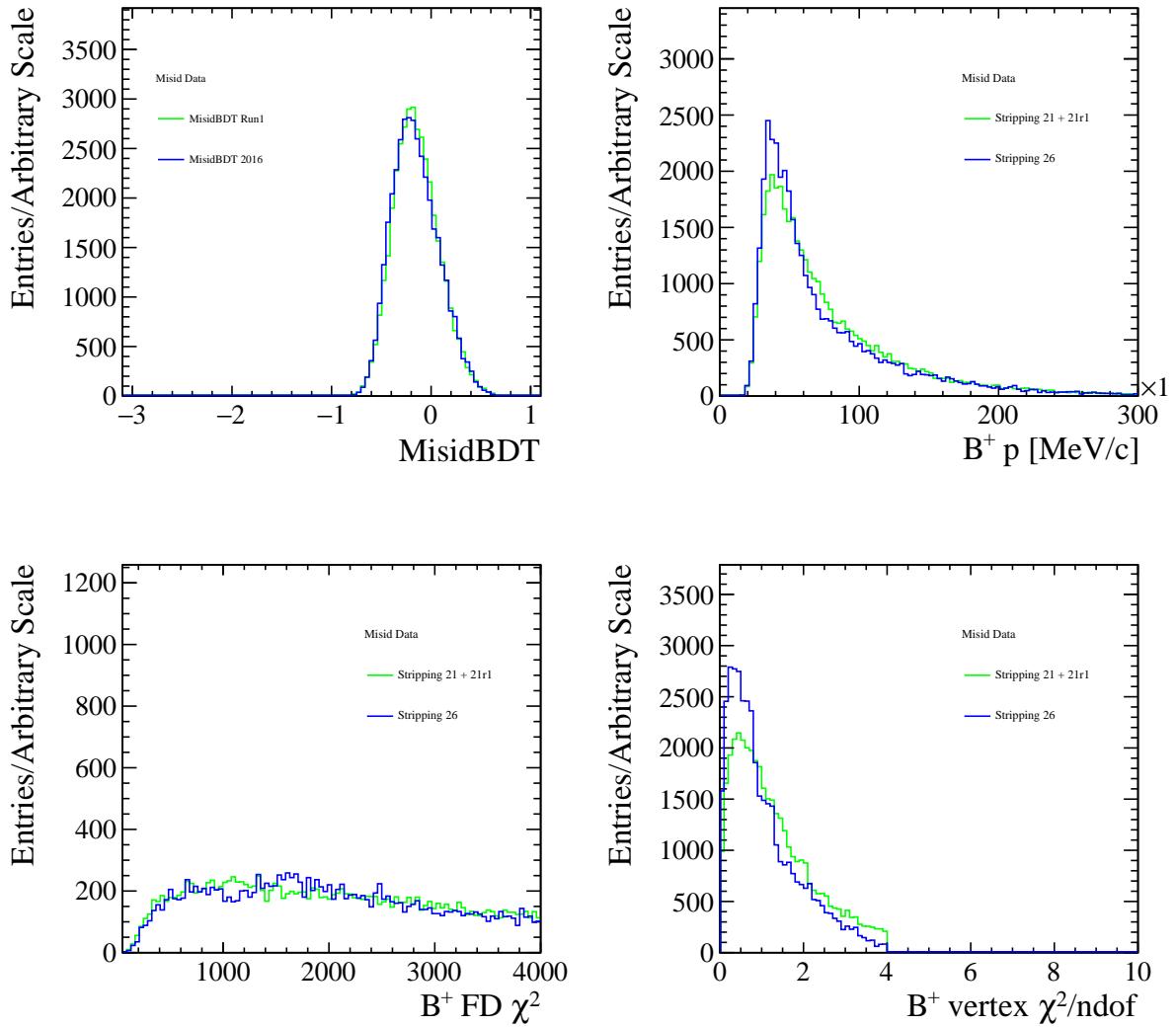
¹⁴⁶¹ **B.2 Signal MC**

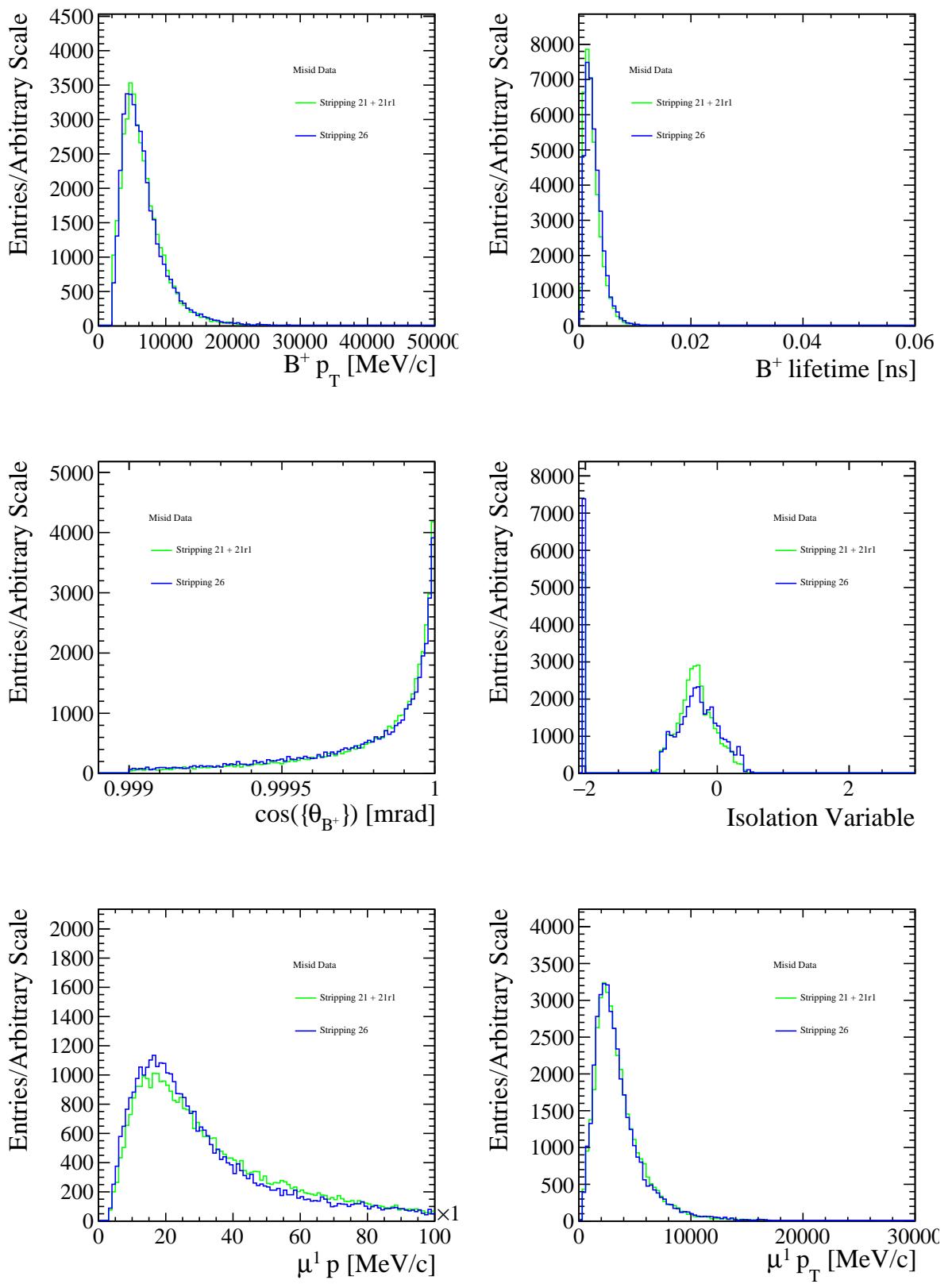


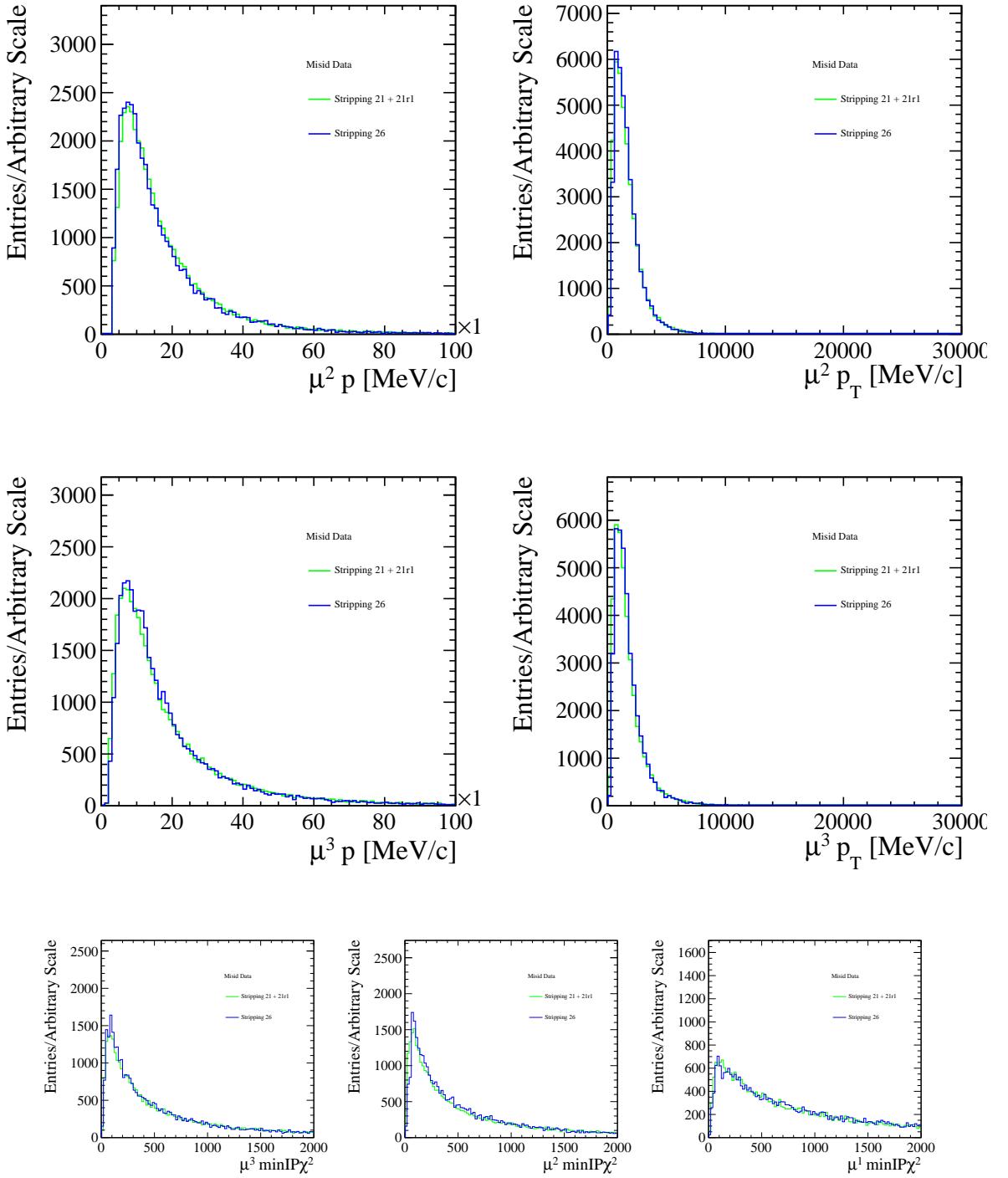




¹⁴⁶² B.3 MisID data







1463 C BDT Characteristics

1464 C.1 Combinatorial BDT

1465 Here are all the properties of Combinatorial BDTs. In the main text, only Run 1 plots/table are
 1466 usually shown. In this Appendix, the information about Run 1 and 2016 Combinatorial BDT is

1467 shown side-by-side to asses if there are any major differences and to show where different working
 1468 points are for the two Combinatorial BDTs. The optimal working point of the combinatorial
 1469 BDT for both Run 1 and 2016 are described in Tables 68,69 together with ranking of most
 1470 discriminative variables for 0th fold, Figures 64, 65. The characteristics of individual BDTs are
 1471 seen in Figure 66 (Run 1), and Figure 67 (2016). The correlations between input variables for
 1472 BDT can be seen in Figure 68.

Property	Value
Signal sample size	76523
Background sample size	20076
Target On	1
Target signal size	89
Target background size	10134
Max Punzi FOM	0.0411402
Optimal cut	0.473
Min cut	-1
Max cut	1
Index of cut	222
Signal Efficiency	0.460659
Background Rejection	0.989739

Property	Value
Signal sample size	86836
Background sample size	23096
Target On	1
Target signal size	89
Target background size	9267
Max Punzi FOM	0.043601
Optimal cut	0.533
Min cut	-1
Max cut	1
Index of cut	231
Signal Efficiency	0.343613
Background Rejection	0.994891

Table 68: Properties of Run 1 BDT.

Table 69: Combinatorial BDT summary 2016

Combinatorial BDT 2012

Combinatorial BDT 2016

Rank : Variable	: Variable Importance	Rank : Variable	: Variable Importance
1 : Bplus_pmu_ISOLATION_BDT1_weights	: 1.725e-01	1 : Bplus_pmu_ISOLATION_BDT1_weights	: 1.475e-01
2 : Bplus_PT	: 9.764e-02	2 : Bplus_ENDVERTEX_CHI2	: 8.697e-02
3 : Bplus_ENDVERTEX_CHI2	: 8.068e-02	3 : mu1_MINIPCHI2	: 8.259e-02
4 : mu2_MINIPCHI2	: 7.856e-02	4 : Bplus_PT	: 7.844e-02
5 : mu3_PT	: 7.610e-02	5 : mu3_PT	: 7.502e-02
6 : Bplus_P	: 7.367e-02	6 : Bplus_FD_CHI2	: 7.494e-02
7 : Bplus_FD_CHI2	: 6.969e-02	7 : Bplus_P	: 7.314e-02
8 : mu3_MINIPCHI2	: 6.699e-02	8 : mu3_MINIPCHI2	: 6.442e-02
9 : mu1_MINIPCHI2	: 6.486e-02	9 : mu1_PT	: 6.318e-02
10 : mu1_PT	: 5.546e-02	10 : mu2_PT	: 6.081e-02
11 : Bplus_TAU	: 5.238e-02	11 : Bplus_TAU	: 5.761e-02
12 : mu2_PT	: 5.233e-02	12 : mu2_MINIPCHI2	: 5.624e-02
13 : mu2_P	: 2.456e-02	13 : mu2_P	: 3.641e-02
14 : mu1_P	: 2.197e-02	14 : mu3_P	: 2.179e-02
15 : mu3_P	: 1.261e-02	15 : mu1_P	: 2.095e-02
16 : Bplus_DIRA_OWNPV	: 0.000e+00	16 : Bplus_DIRA_OWNPV	: 0.000e+00

Figure 64: Ranking of the variables in 2012 data, 0-th fold, for the Combinatorial BDT

Figure 65: Ranking of the variables in 2016 data, 0-th fold, for the Combinatorial BDT.

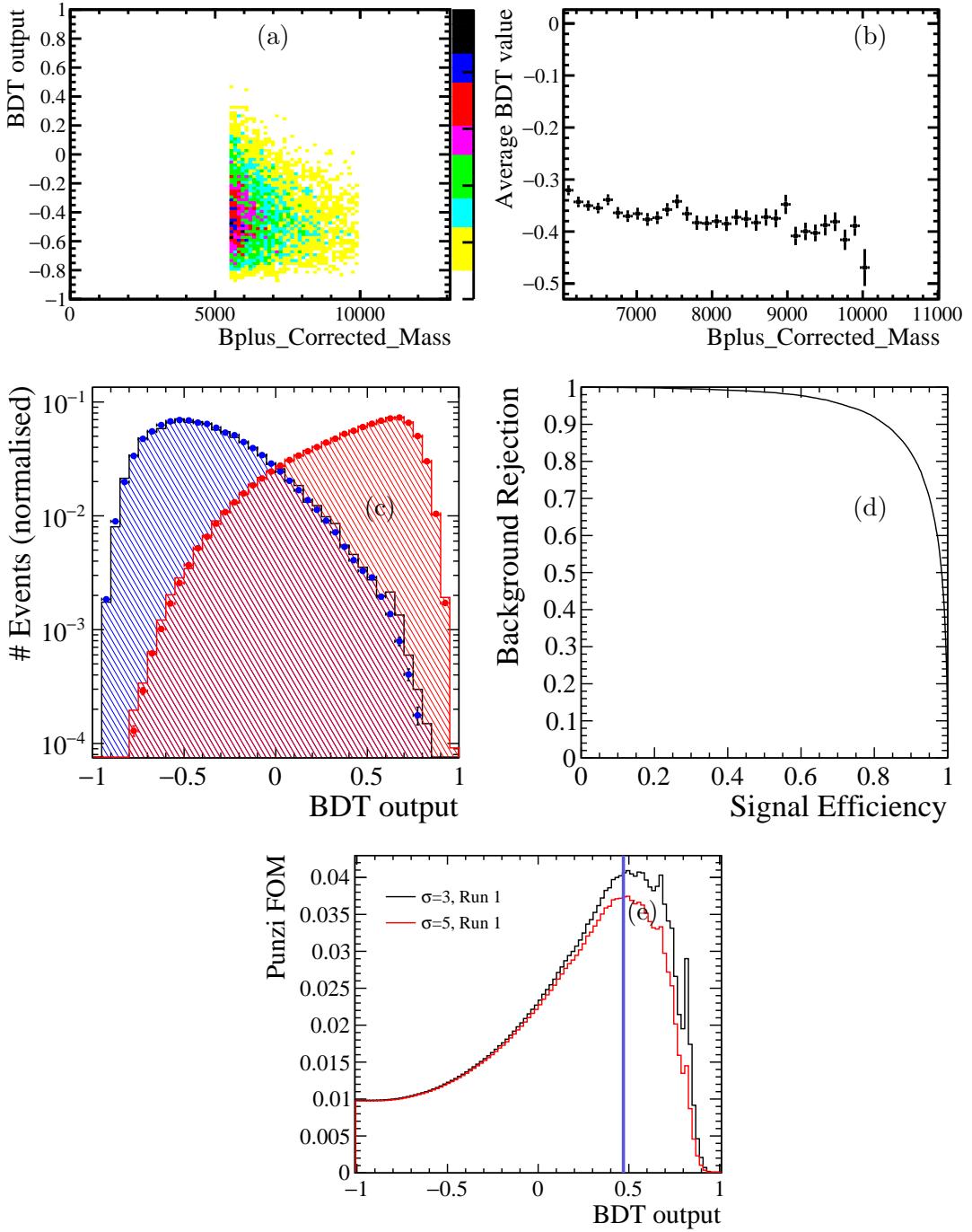


Figure 66: **Properties of Combinatorial Run 1 BDT** (a) Correlation between $B_{corr m}^+$ and BDT (combinatorial sample) for Run 1 Combinatorial BDT as well as (b) evidence that BDT is not correlated to the $B_{corr m}^+$. (c) Check for overtraining was passed yielding following (d) ROC-curve. Finally, (e) optimal working point 0.47 for combinatorial BDT for Run 1 data was obtained. The spike here is due to no background events left and hence shall be ignored.

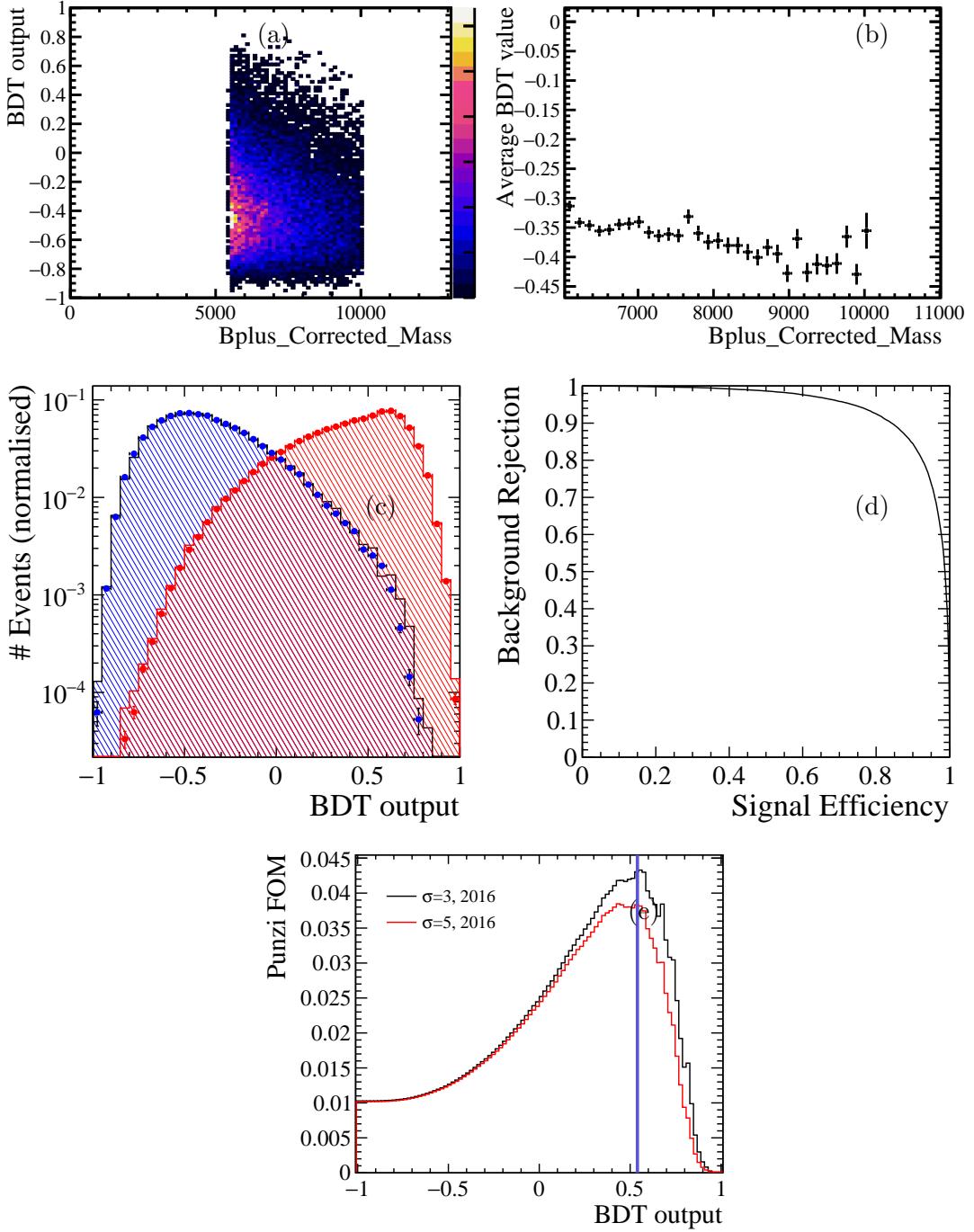


Figure 67: **Properties of Combinatorial 2016 BDT** (a) Correlation between $B_{corr m}^+$ and BDT (combinatorial sample) for 2016 Combinatorial BDT as well as (b) evidence that BDT is not correlated to the $B_{corr m}^+$. (c) Check for overtraining was passed yielding following (d) ROC-curve. Finally, (e) optimal working point 0.54 for combinatorial BDT for Run 1 data was obtained. The spike here is due to no background events left and hence shall be ignored.

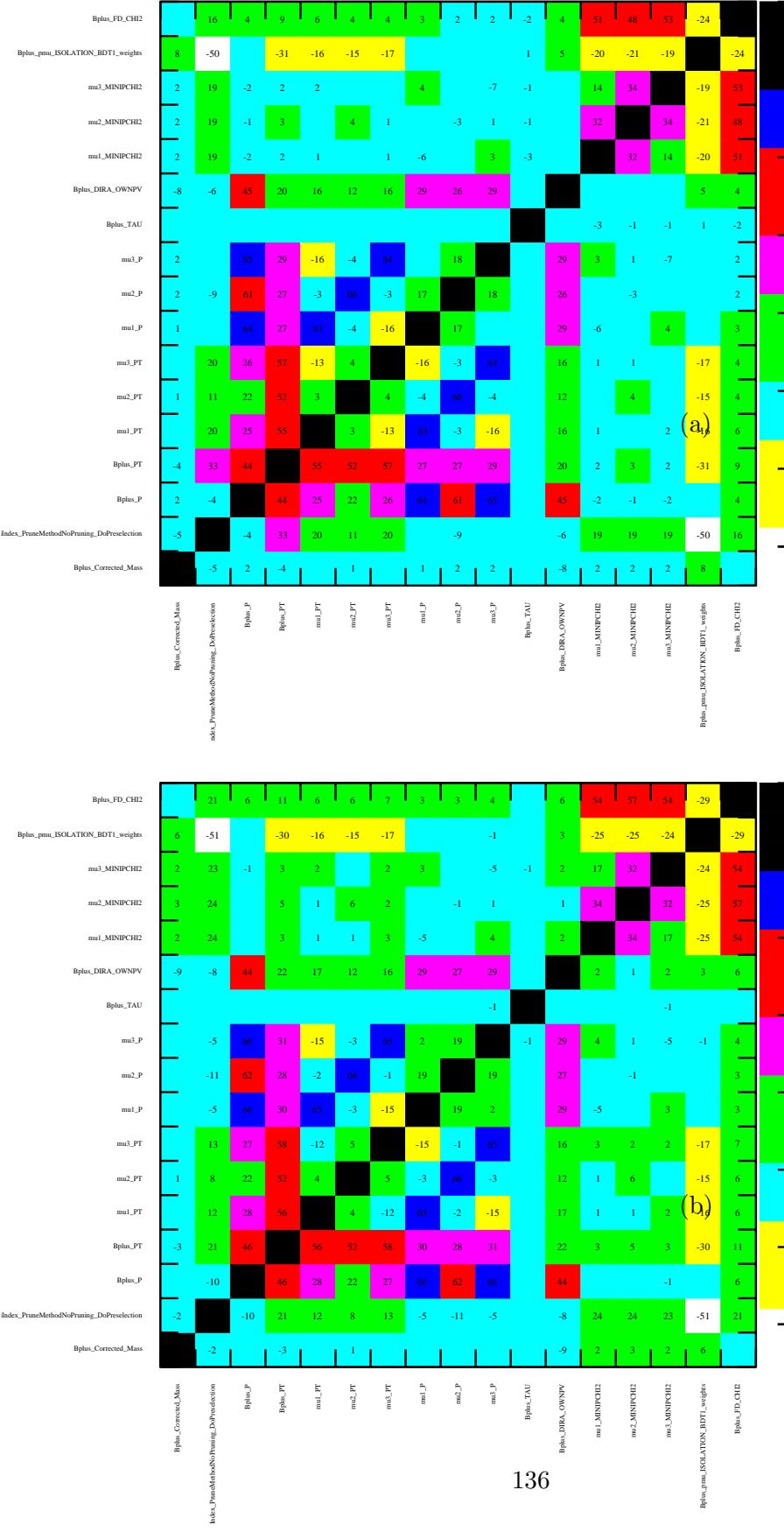


Figure 68: Correlation matrix for all input variables, corrected mass as well resulting BDT variable for both (a) Run 1 Combinatorial BDT (b) 2016 Combinatorial BDT.

1473 **C.2 MisID BDT**

1474 Here are all the properties of Misid BDTs. In this Appendix, the information about Run 1 and
 1475 2016 Misid BDT is shown side-by-side to asses if there are any major differences and to show
 1476 where different working points are for the two Misid BDTs. The optimal working point of the
 1477 misID BDT for both Run 1 and 2016 are described in Tables 70,71 together with ranking of most
 1478 discriminative variables for 0th fold, Figures 69, 70. In this Appendix, the information about
 1479 Run 1 and 2016 Misid BDT is shown side-by-side to asses if there are any major differences
 1480 and to show where different working points are for the two Misid BDTs. The characteristics of
 1481 individual BDTs are seen in Figure 71 (Run 1), and Figure 72 (2016). The correlations between
 1482 input variables for misID BDT for both Run 1 and 2016 can be seen in Figure 73.

Property	Value
Signal sample size	36160
Background sample size	39140
Target On	1
Target signal size	40
Target background size	2477
Max Punzi FOM	0.0307636
Optimal cut	0.21
Min cut	-1
Max cut	1
Index of cut	197
Signal Efficiency	0.301632
Background Rejection	0.968702

Table 70: Misid BDT summary Run 1

Property	Value
Signal sample size	29838
Background sample size	23462
Target On	1
Target signal size	40
Target background size	2263
Max Punzi FOM	0.0331807
Optimal cut	0.27
Min cut	-1
Max cut	1
Index of cut	193
Signal Efficiency	0.343187
Background Rejection	0.961427

Table 71: Misid BDT summary 2016

Misid BDT 2012

Rank : Variable	: Variable Importance
1 : mu3_PT	: 1.467e-01
2 : mu1_PT	: 8.875e-02
3 : mu3_P	: 8.326e-02
4 : mu1_MINIPCHI2	: 7.392e-02
5 : mu3_MINIPCHI2	: 7.059e-02
6 : mu1_P	: 6.966e-02
7 : Bplus_pmu_ISOLATION_BDT1_weights	: 6.949e-02
8 : mu2_PT	: 6.936e-02
9 : mu2_P	: 5.710e-02
10 : mu2_MINIPCHI2	: 5.648e-02
11 : Bplus_ENDVERTEX_CHI2	: 4.602e-02
12 : Bplus_P	: 4.572e-02
13 : Bplus_PT	: 4.465e-02
14 : Bplus_FD_CHI2	: 4.155e-02
15 : Bplus_TAU	: 3.676e-02
16 : Bplus_DIRA_OWNPV	: 0.000e+00

Misid BDT 2016

Rank : Variable	: Variable Importance
1 : mu3_PT	: 1.565e-01
2 : Bplus_pmu_ISOLATION_BDT1_weights	: 8.731e-02
3 : mu3_P	: 8.381e-02
4 : mu1_PT	: 8.160e-02
5 : mu2_PT	: 7.929e-02
6 : mu3_MINIPCHI2	: 7.556e-02
7 : mu2_MINIPCHI2	: 7.310e-02
8 : Bplus_PT	: 6.959e-02
9 : mu1_P	: 6.600e-02
10 : mu1_MINIPCHI2	: 6.178e-02
11 : Bplus_ENDVERTEX_CHI2	: 4.896e-02
12 : Bplus_P	: 4.211e-02
13 : Bplus_FD_CHI2	: 3.299e-02
14 : mu2_P	: 2.292e-02
15 : Bplus_TAU	: 1.029e-02
16 : Bplus_DIRA_OWNPV	: 8.179e-03

Figure 69: Ranking of the variables in 2012 data, 0-th fold, in the Misid BDT.

Figure 70: Ranking of the variables in 2016 data, 0-th fold, in the Misid BDT.

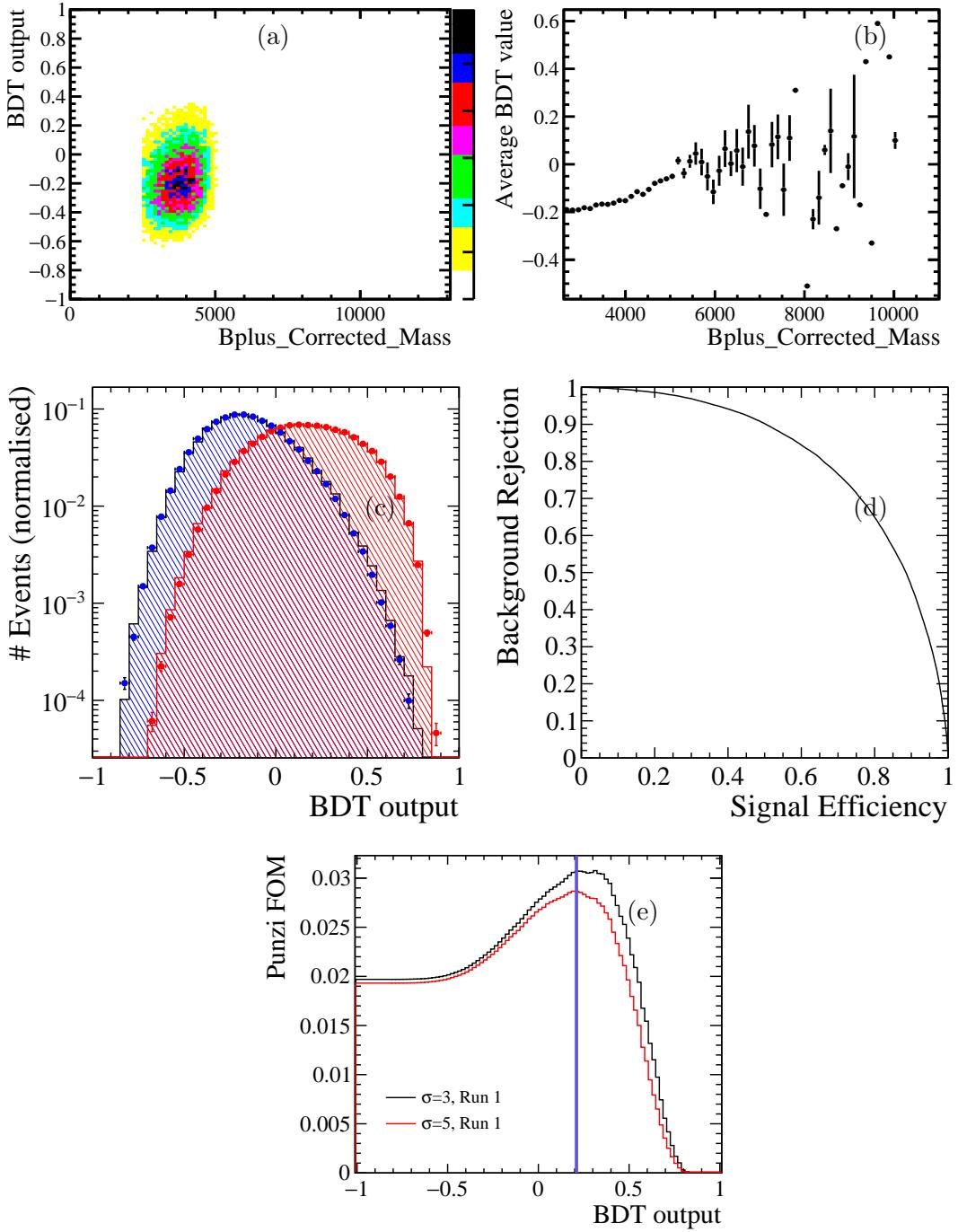


Figure 71: **Properties of Misid Run 1 BDT** (a) Correlation between $B_{corr m}^+$ and BDT (misID sample) for Run 1 Misid BDT as well as (b) evidence that BDT is not correlated to the $B_{corr m}^+$. (c) Check for overtraining was passed yielding following (d) ROC-curve. As can be seen compared to Figure 66 the trade-off is worse however there is less events to discriminate against. Finally, (e) optimal working point for misID BDT for Run 1 0.21 data was obtained.

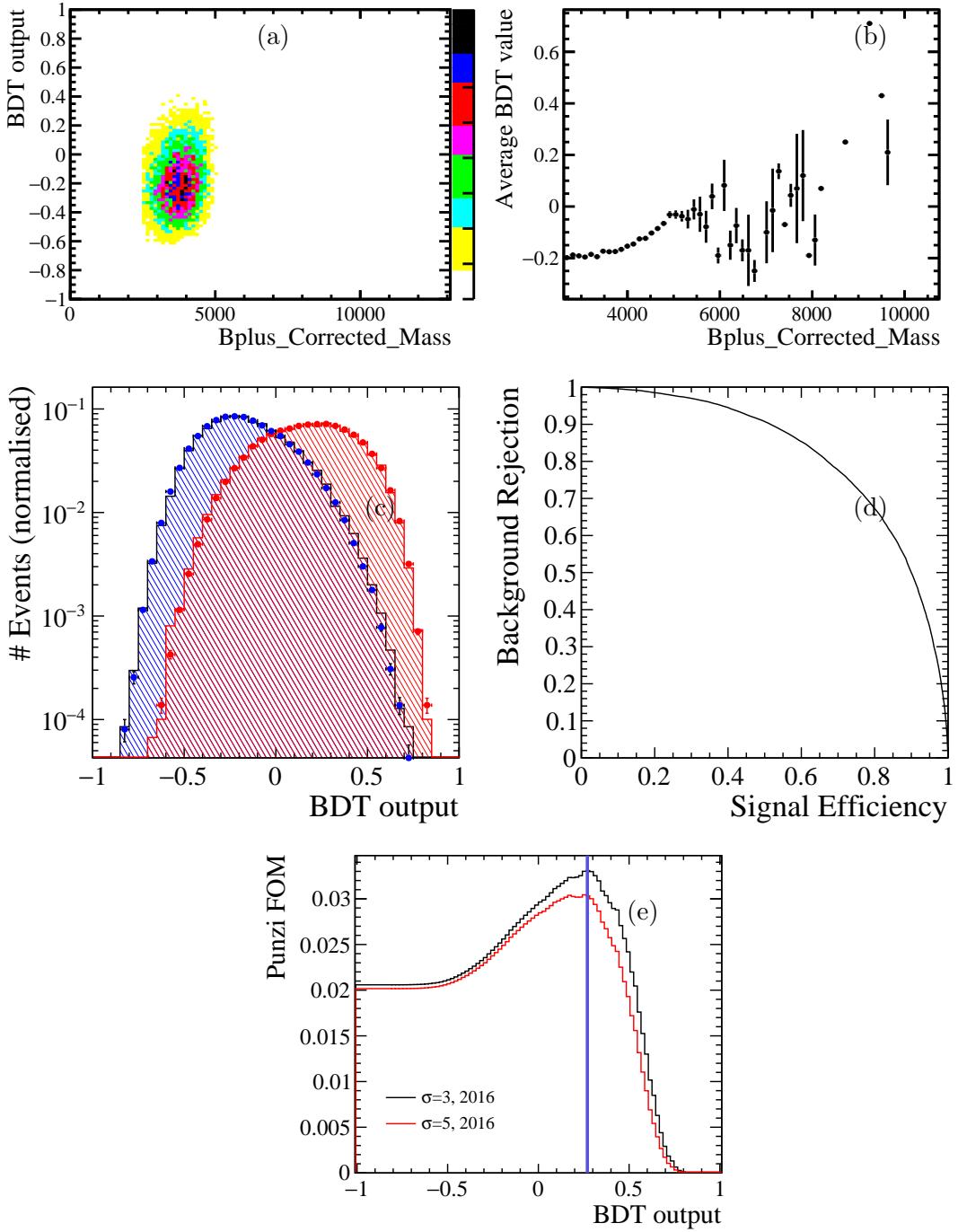
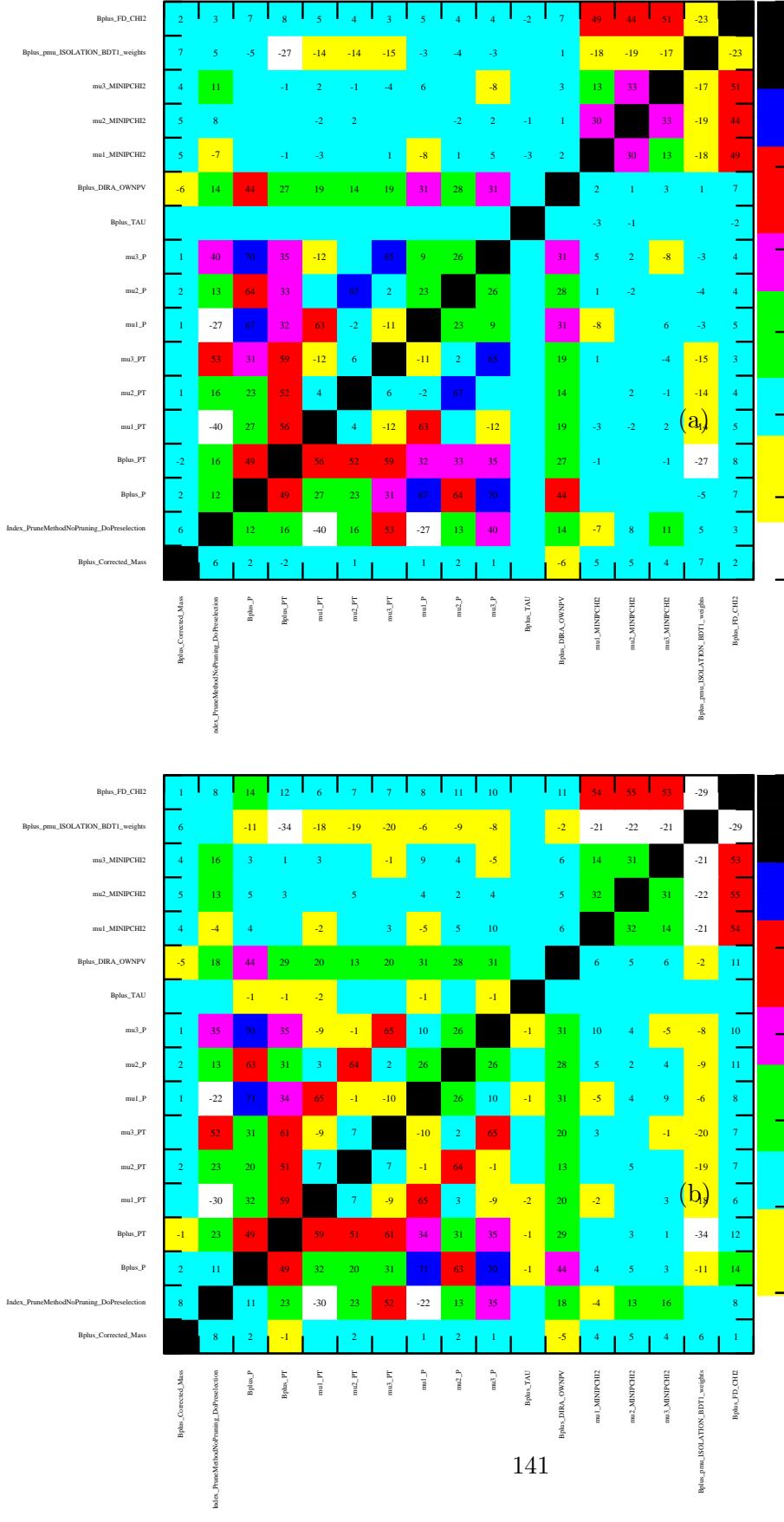


Figure 72: **Properties of Misid 2016 BDT** (a) Correlation between B_{corr}^+ and BDT (misid sample) for 2016 Misid BDT as well as (b) evidence that BDT is not correlated to the B_{corr}^+ . (c) Check for overtraining was passed yielding following (d) ROC-curve. Finally, (e) optimal working point 0.27 for misID BDT for 2016 data was obtained.



1483 **D Partially Reconstructed Background - break down of efficiencies**

1484

1485 The Table 72 in this Appendix shows more detailed efficiency comparison between signal and
 1486 PartReco MC samples up to multivariate selection. There was no trigger applied on the PartReco
 1487 MC sample as it would be mismodelled as most of trigger lines selected look for muon, hence
 1488 taking conservative approach of not requiring trigger will over-estimate the PartReco efficiency
 1489 for this category.

ε	$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$	$B^+ \rightarrow (D^0 \rightarrow K^+ \pi^- \mu^+ \mu^-) \mu^+ \nu$
ε_{gen}	0.185	0.16
Events in Bookkeeping	1.11413e+06	2.05461e+06
ε_{strip}	0.117	0.0749
$\varepsilon_{mctruth}$	0.935	0.389
$\varepsilon_{trigger}$	0.827	1
$\varepsilon_{J/\psi}$	0.882	1
ε_{minq}	0.957	0.955
$\varepsilon_{combibdt}$	0.473	0.245
$\varepsilon_{misidbdt}$	0.436	0.636

Table 72: MC efficiency comparison between signal and PartReco MC samples up to multivariate selection.

1490 **E MC Truth Matching**

1491 **E.1 Signal MC truth matching**

1492 In this Appendix in the Table 73 the offline truth matching for $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu_\mu$ default MC,
 1493 detailed in 3.2.2, is listed. The B^+ background category is required to be less than 60, i.e. to
 1494 be signal-like. Additional requirement that is imposed is that the mother of one of the muon
 1495 candidates with the same charge as B is required to come from W , which in the framework of
 1496 the simulation is denoted with ID: 25 or 35.

Particle	Requirement
mu1	$MC_MOTHER_ID == 25$
OR	
mu3	$MC_MOTHER_ID == 25$
OR	
mu1	$MC_MOTHER_ID == 35$
OR	
mu3	$MC_MOTHER_ID == 35$

Table 73: Offline truth matching for $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu_\mu$ candidates with B $BKG_CAT < 60$.

1497 **E.2 Normalisation MC truth matching**

1498 The normalisation MC is also truth matched offline in a following way, see Table 74: Basically,
 1499 all particle candidates are required to be consistent with their PDG IDs, as do the mother and
 1500 grandmother PDG IDs.

Particle	Requirement
$\text{abs}(\mu^+)$	$MC_TRUE_ID == 13$
$\text{abs}(\mu^-)$	$MC_TRUE_ID == 13$
$\text{abs}(K^+)$	$MC_TRUE_ID == 321$
$\text{abs}(B^+)$	$MC_TRUE_ID == 521$
$\text{abs}(\mu^+)$	$MC_MOTHER_ID == 443$
$\text{abs}(\mu^-)$	$MC_MOTHER_ID == 443$
$\text{abs}(K^+)$	$MC_MOTHER_ID == 521$
$\text{abs}(\mu^+)$	$MC_GD_MOTHER_ID == 521$
$\text{abs}(\mu^-)$	$MC_GD_MOTHER_ID == 521$

Table 74: Truth Match for $B^+ \rightarrow (J/\psi \rightarrow \mu^+ \mu^-) K^+$.

1501 **E.3 PartReco MC truth matching**

1502 For partially reconstructed background $B^+ \rightarrow (D^0 \rightarrow K^-\pi^+\mu^+\mu^-)\mu^+\nu$, inclusive MC sample of
 1503 $B^+ \rightarrow (D^0 \rightarrow ((K_s^0 \rightarrow \pi^+\pi^-)\pi^-\pi^+\pi^0) \mu^+\nu$ is used, as seen in Section 5.3. So one of the pions
 1504 in this decay becomes proxy for signal muon. In this MC sample mu1 = π^+ mu2 = π^- mu3 =
 1505 μ^- . Truth matching of this MC is summarized in Table 75. Again, all particle candidates are
 1506 required to be consistent with their PDG IDs, and they have come from correct mother particle
 1507 B .

Particle	Requirement
mu1	$MC_MOTHER_ID == -421$
mu2	$MC_MOTHER_ID == -421$
mu3	$MC_MOTHER_ID == 521$
OR	
mu1	$MC_MOTHER_ID == 421$
mu2	$MC_MOTHER_ID == 421$
mu3	$MC_MOTHER_ID == -521$

Table 75: Truth Matching for $B^+ \rightarrow (D^0 \rightarrow ((K_s^0 \rightarrow \pi^+\pi^-)\pi^-\pi^+\pi^0) \mu^+\nu$. On the top, two pions are
 matched with their PDG value, as well muon and B^+ candidate.

1508 **F TCK-dependant Efficiencies 2016**

1509 To illustrate the impact of different trigger setting in 2016 on several steps of selection is shown in
 1510 Tables 76, 77,78 for BDT, fit range, FCME split efficiency respectively. In general, the agreement

¹⁵¹¹ of these selection efficiencies show consistency across the different trigger configurations.

¹⁵¹² F.1 BDT efficiencies

TCK	$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$		$B^+ \rightarrow J/\psi K^+$	
	$\epsilon_{combibdt}$	$\epsilon_{misidbdt}$	$\epsilon_{combibdt}$	$\epsilon_{misidbdt}$
287905280	0.337	0.37	0.381	0.447
287905283	0.337	0.371	0.381	0.447
287905284	0.337	0.365	0.381	0.446
287905285	0.338	0.371	0.381	0.446
288495113	0.337	0.365	0.381	0.446
288626185	0.337	0.365	0.381	0.446
288691721	0.345	0.37	0.401	0.447
288757257	0.345	0.37	0.401	0.447
288822793	0.345	0.37	0.401	0.447
288822798	0.344	0.367	0.402	0.447
288888329	0.345	0.37	0.401	0.447
288888334	0.344	0.367	0.402	0.447
288888335	0.344	0.366	0.403	0.446
288888337	0.344	0.367	0.402	0.447
288888338	0.345	0.37	0.402	0.447
288888339	0.345	0.369	0.401	0.447
Weighted efficiency	0.343	0.368	0.397	0.446

Table 76: 2016 Combinatorial BDT selection efficiency. This MVA selection was trained on events with 288888335 TCK. It can be seen that for the trigger configuration 288888335 it is more efficient. This trigger has particularly high p_T threshold.

₁₅₁₃ F.2 Fit Range Efficiencies

TCK	$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ $\epsilon_{fitrange}$	$B^+ \rightarrow J/\psi K^+$ $\epsilon_{fitrange}$
287905280	0.938	0.999
287905283	0.939	0.999
287905284	0.938	0.999
287905285	0.937	0.999
288495113	0.938	0.999
288626185	0.938	0.999
288691721	0.938	0.999
288757257	0.938	0.999
288822793	0.938	0.999
288822798	0.937	0.999
288888329	0.938	0.999
288888334	0.937	0.999
288888335	0.938	0.999
288888337	0.937	0.999
288888338	0.939	0.999
288888339	0.938	0.999
Weighted efficiency	0.938	0.999

Table 77: Fit range TCK-dependant efficiencies.

¹⁵¹⁴ **F.3 FCME split**

TCK	Split	$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$	$B^+ \rightarrow J/\psi K^+$	Split	$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$	$B^+ \rightarrow J/\psi K^+$
287905280	HighFCME	0.452915	0.327096	LowFCME	0.547085	0.672904
287905283	HighFCME	0.453567	0.326762	LowFCME	0.546433	0.673238
287905284	HighFCME	0.451037	0.327404	LowFCME	0.548963	0.672596
287905285	HighFCME	0.452678	0.32732	LowFCME	0.547322	0.67268
288495113	HighFCME	0.451037	0.327404	LowFCME	0.548963	0.672596
288626185	HighFCME	0.451037	0.327404	LowFCME	0.548963	0.672596
288691721	HighFCME	0.452322	0.32656	LowFCME	0.547678	0.67344
288757257	HighFCME	0.452322	0.32656	LowFCME	0.547678	0.67344
288822793	HighFCME	0.452322	0.32656	LowFCME	0.547678	0.67344
288822798	HighFCME	0.457319	0.326268	LowFCME	0.542681	0.673732
288888329	HighFCME	0.452322	0.32656	LowFCME	0.547678	0.67344
288888334	HighFCME	0.457319	0.326268	LowFCME	0.542681	0.673732
288888335	HighFCME	0.450625	0.32704	LowFCME	0.549375	0.67296
288888337	HighFCME	0.457319	0.326268	LowFCME	0.542681	0.673732
288888338	HighFCME	0.452158	0.326756	LowFCME	0.547842	0.673244
288888339	HighFCME	0.453591	0.326175	LowFCME	0.546409	0.673825

Table 78: 2016 High and Low Fractional Corrected Mass Error Split.

¹⁵¹⁵ **G Kinematic Reweighting**

¹⁵¹⁶ In this Appendix, detailed breakdown of efficiencies for both signal and normalisation channel
¹⁵¹⁷ with and without kinematic reweighting for both 2012 and 2016 is listed in Tables 79, 80, 81, 82.
¹⁵¹⁸ Ratios of total efficiencies used to find systematic uncertainty for kinematic reweighting can be
¹⁵¹⁹ seen in Section 10.4.

Efficiency	$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ (Nominal)	$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ (Reweighted JpsiK weights)
$\epsilon_{decprodcut}$	0.18643	0.18643
num bkk	1.11413e+06	1.11413e+06
num reco and mc truth	121779	121779
$\epsilon_{reco+strip+mctruth}$	0.109304	0.109304
num final	14546	14546
ϵ_{TRG}	0.74471	0.742785
ϵ_{OFF}	0.881894	0.880882
$\epsilon_{minqcut}$	0.956789	0.95669
$\epsilon_{CombiBDT}$	0.472538	0.45715
$\epsilon_{MisidBDT}$	0.435813	0.431286
ϵ_{FR}	0.923028	0.924458
$\epsilon_{allofcuts}$	0.119446 ± 0.000929346	0.114094 ± 0.000911044
ϵ_{PID}	0.64815	0.64815
ϵ_{GEN}	$0.18643 + -0.00029$	$0.18643 + -0.00029$
$\epsilon_{reco + strip + mctruth}$	0.109304	0.109304
$\epsilon_{offlinesel}$	$0.1194 + -0.0009$	$0.1141 + -0.0009$
ϵ_{PID}	$0.648 + -0.005$	$0.648 + -0.005$
ϵ_{all}	$0.001577 + -0.000018$	$0.001507 + -0.000017$

Table 79: **2012 Signal Simulation.** Reweighting in B p , p_T , vertex χ^2 has been used.

Efficiency	$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ (Nominal)	$B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ (Reweighted JpsiK weights)
$\epsilon_{decprodcut}$	0.1977	0.1977
num bkk	1.10772e+06	1.10772e+06
num reco and mc truth	138330	138330
$\epsilon_{reco+strip+mctruth}$	0.124879	0.124879
num final	10316	10316
ϵ_{TRG}	0.731447	0.725511
ϵ_{OFF}	0.884168	0.882004
$\epsilon_{minqcut}$	0.958429	0.958325
$\epsilon_{CombiBDT}$	0.345805	0.329075
$\epsilon_{MisidBDT}$	0.371366	0.362917
ϵ_{FR}	0.936881	0.938763
$\epsilon_{alloffcuts}$	$0.0745753 \pm 0.000706334$	0.068752 ± 0.000680326
ϵ_{PID}	0.624	0.624
ϵ_{GEN}	0.1977 ± 0.0004	0.1977 ± 0.0004
$\epsilon_{reco + strip + mctruth}$	0.124879	0.124879
$\epsilon_{offlinesel}$	0.0746 ± 0.0007	0.0688 ± 0.0007
ϵ_{PID}	0.624 ± 0.006	0.624 ± 0.006
ϵ_{all}	0.001149 ± 0.000016	0.001059 ± 0.000015

Table 80: **2016 Signal Simulation.** Reweighting in B p , p_T , vertex χ^2 has been used.

Efficiency	$B^+ \rightarrow J/\psi K^+$ (Nominal)	$B^+ \rightarrow J/\psi K^+$ (Reweighted JpsiK weights)
$\epsilon_{decprodcut}$	0.16216	0.16216
num bkk	8.04382e+06	8.04382e+06
num reco and mc truth	1427090	1427090
$\epsilon_{reco+strip+mctruth}$	0.177414	0.177414
num final	287639	287639
ϵ_{TRG}	0.777899	0.776146
$\epsilon_{CombiBDT}$	0.508916	0.488281
$\epsilon_{MisidBDT}$	0.511213	0.505149
ϵ_{FR}	0.995921	0.995901
$\epsilon_{alloffcuts}$	0.201556 ± 0.000335811	0.190655 ± 0.000328825
ϵ_{PID}	0.716	0.716
ϵ_{GEN}	0.16216 ± 0.00024	0.16216 ± 0.00024
$\epsilon_{reco + strip + mctruth}$	0.177414	0.177414
$\epsilon_{offlinesel}$	0.20156 ± 0.00034	0.19066 ± 0.00033
ϵ_{PID}	0.7152 ± 0.0011	0.7152 ± 0.0011
ϵ_{all}	0.004147 ± 0.000012	0.003923 ± 0.000011

Table 81: **2012 Normalisation Simulation.** Reweighting in B p , p_T , vertex χ^2 has been used.

Efficiency	$B^+ \rightarrow J/\psi K^+$ (Nominal)	$B^+ \rightarrow J/\psi K^+$ (Reweighted JpsiK weights)
$\epsilon_{decprodcut}$	0.1739	0.1739
num bkk	1.32316e+07	1.32316e+07
num reco and mc truth	2650397	2650397
$\epsilon_{reco+strip+mctruth}$	0.200308	0.200308
num final	358161	358161
ϵ_{TRG}	0.744469	0.736892
$\epsilon_{CombiBDT}$	0.406517	0.381997
$\epsilon_{MisidBDT}$	0.446915	0.434232
ϵ_{FR}	0.999118	0.999113
$\epsilon_{alloffcuts}$	0.135135 ± 0.000209992	0.122124 ± 0.000201123
ϵ_{PID}	0.676	0.676
ϵ_{GEN}	0.1739 ± 0.0004	0.1739 ± 0.0004
$\epsilon_{reco + strip + mctruth}$	0.200308	0.200308
$\epsilon_{offlinesel}$	0.13513 ± 0.00021	0.12212 ± 0.00020
ϵ_{PID}	0.6760 ± 0.0010	0.6760 ± 0.0010
ϵ_{all}	0.003182 ± 0.000010	0.002876 ± 0.000009

Table 82: **2016 Normalisation Simulation.** Reweighting in B p , p_T , vertex χ^2 has been used.

References

- 1520 1521 [1] Particle Data Group, C. Patrignani *et al.*, *Review of Particle Physics*, Chin. Phys. **C40** (2016), no. 10 100001.
- 1522 1523 [2] HFLAV, Y. Amhis *et al.*, *Averages of b-hadron, c-hadron, and τ -lepton properties as of summer 2016*, Eur. Phys. J. **C77** (2017), no. 12 895, [arXiv:1612.07233](https://arxiv.org/abs/1612.07233).
- 1524 1525 [3] Belle, I. Adachi *et al.*, *Evidence for $B^- \rightarrow \tau^- \bar{\nu}_\tau$ with a Hadronic Tagging Method Using the Full Data Sample of Belle*, Phys. Rev. Lett. **110** (2013), no. 13 131801, [arXiv:1208.4678](https://arxiv.org/abs/1208.4678).
- 1526 1527 [4] BaBar, J. P. Lees *et al.*, *Evidence of $B^+ \rightarrow \tau^+ \nu$ decays with hadronic B tags*, Phys. Rev. **D88** (2013), no. 3 031102, [arXiv:1207.0698](https://arxiv.org/abs/1207.0698).
- 1528 1529 [5] G. Burdman, J. T. Goldman, and D. Wyler, *Radiative leptonic decays of heavy mesons*, Phys. Rev. **D51** (1995) 111, [arXiv:hep-ph/9405425](https://arxiv.org/abs/hep-ph/9405425).
- 1530 1531 [6] Belle, A. Heller *et al.*, *Search for $B^+ \rightarrow \ell^+ \nu \gamma$ decays with hadronic tagging using the full Belle data sample*, [arXiv:1504.05831](https://arxiv.org/abs/1504.05831).
- 1532 1533 [7] A. Danilina and N. Nikitin, *Rare four-leptonic B-decays with light leptons in the final state in the Standard Model*, EPJ Web Conf. **158** (2017) 03005.
- 1534 1535 [8] A. V. Danilina and N. V. Nikitin, *Four-Leptonic Decays of Charged and Neutral B Mesons within the Standard Model*, Phys. Atom. Nucl. **81** (2018), no. 3 347, [Yad. Fiz. 81, no. 3, 331 (2018)].

- 1538 [9] A. Ali, P. Ball, L. T. Handoko, and G. Hiller, *A Comparative study of the decays $B \rightarrow (K,$*
 1539 *$K^*)\ell^+\ell^-$ in standard model and supersymmetric theories*, Phys. Rev. **D61** (2000) 074024,
 1540 [arXiv:hep-ph/9910221](https://arxiv.org/abs/hep-ph/9910221).
- 1541 [10] G. Punzi, *Sensitivity of searches for new signals and its optimization*, in *Statistical Problems*
 1542 *in Particle Physics, Astrophysics, and Cosmology* (L. Lyons, R. Mount, and R. Reitmeyer,
 1543 eds.), p. 79, 2003. [arXiv:physics/0308063](https://arxiv.org/abs/physics/0308063).
- 1544 [11] LHCb, R. Aaij *et al.*, *Determination of the quark coupling strength $|V_{ub}|$ using baryonic*
 1545 *decays*, [arXiv:1504.01568](https://arxiv.org/abs/1504.01568).
- 1546 [12] LHCb, R. Aaij *et al.*, *First observation of the decay $D^0 \rightarrow K^-\pi^+\mu^+\mu^-$ in the $\rho^0\omega$ region*
 1547 *of the dimuon mass spectrum*, Phys. Lett. **B757** (2016) 558, [arXiv:1510.08367](https://arxiv.org/abs/1510.08367).
- 1548 [13] Particle Data Group, K. A. Olive *et al.*, *Review of particle physics*, Chin. Phys. **C38** (2014)
 1549 090001.
- 1550 [14] D. Martnez Santos and F. Dupertuis, *Mass distributions marginalized over per-event errors*,
 1551 Nucl. Instrum. Meth. **A764** (2014) 150, [arXiv:1312.5000](https://arxiv.org/abs/1312.5000).
- 1552 [15] A. L. Read, *Presentation of search results: The $CL(s)$ technique*, J. Phys. **G28** (2002) 2693,
 1553 [,11(2002)].