



ATLAS Note

ANA-BPHY-2018-09-INT1

14th June 2018



Draft version 0.1

1

2

Study of the rare decays of B_s^0 and B^0 into muon pairs from data collected during 2015 and 2016 with the ATLAS detector

3

4

5

6

7

A. Cerri^a, I. Ibragimov^b, A. Grummer^c, F. Tresoldi^a, U. de Sanctis^d, A. Campoverde^b, W. Walkowiak^b, P. Buchholz^b, S. Seidel^c, S.Yu. Sivoklokov^e,
XAuthor list incomplete. Please let us know if you're missing!¹

8

9

10

11

12

^a*Department of Physics and Astronomy, University of Sussex*

^b*Department Physik, Universität Siegen*

^c*Department of Physics and Astronomy, University of New Mexico*

^d*INFN Sezione di Roma Tor Vergata and Dipartimento di Fisica Università di Roma Tor Vergata*

^e*D.V. Skobeltsyn Institute of Nuclear Physics, M.V. Lomonosov Moscow State University*

13 This note documents the analysis performed to search for $B_{s,d} \rightarrow \mu\mu$ decays. The study is
14 based on the data collected by ATLAS in 2015 and 2016. This analysis complements the
15 previous result based on Run 1 data, and is mostly based on the same tools and techniques.

© 2018 CERN for the benefit of the ATLAS Collaboration.

16 Reproduction of this article or parts of it is allowed as specified in the CC-BY-4.0 license.

17 Contents

18	1 Note Changes	5
19	2 Introduction	6
20	3 Data and Monte Carlo Samples	8
21	3.1 Data Samples	8
22	3.2 Monte Carlo Samples	8
23	3.3 Ntuples	9
24	4 Candidate Preselection	19
25	4.1 Candidate building and preselection on derivation level	19
26	4.2 Candidate Preselection	21
27	4.3 Primary vertex determination	22
28	4.4 Invariant Mass Ranges and Blinding Window	24
29	4.5 Additional selection	24
30	5 Kinematic Corrections to Signal MC	28
31	5.1 Preliminary assessments	28
32	5.2 Quark Level Corrections	29
33	5.3 Data Driven Weights	30
34	6 Trigger	35
35	7 CutFlow	37
36	8 Studies on muon fake rates	38
37	9 Background Modeling	41
38	9.1 Non-resonant Background	41
39	9.2 Background Classification	42
40	9.2.1 Truth Matching	42
41	9.2.2 Cumulative Decay Probability Distribution	43
42	10 ContinuumBDT	47
43	11 DataMCComparison	52
44	11.1 Reference channel as control sample for signal: $B^+ \rightarrow J/\psi K^+$	52
45	11.2 Alternate reference channel: $B_s^0 \rightarrow J/\psi \phi$	56
46	11.3 Yield stability during run	56
47	12 BPlus Yield	60
48	12.1 Preselection	60
49	12.2 Procedure	61
50	12.3 Systematic Uncertainties in the Extraction of the B^\pm Yields	71
51	12.4 $B^\pm \rightarrow J/\psi K^\pm$ yield results	72
52	13 EfficiencyAndAcceptance	73

53	13.1 Systematic uncertainties on $R_{A\epsilon}$	74
54	14 SignalFit	75
55	14.1 Signal and peaking background	75
56	14.2 Parametrisation of background components	75
57	14.3 Fit to background components from MC and to sideband data	76
58	14.4 Summary of the fit configuration	76
59	14.5 Systematic uncertainties on the fit in the simultaneous fit to B_s and B_d	76
60	15 BranchingRatio	77
61	Appendices	80
62	A QLC weights	80
63	A.1 Comparison with Run1 procedure	80
64	A.2 Checks on QLC	81
65	B DDW weights	85
66	B.1 SBW weights	85
67	B.1.1 B^+ SBW weights	85
68	B.1.2 B_s^0 SBW weights	86
69	B.2 DDW weights extraction	89
70	B.3 Checks on DDW	89
71	C Data-MC comparisons in the sidebands	92
72	D Data-MC comparisons in the reference channel $B^+ \rightarrow J/\psi K^+$	98
73	E Data-MC comparisons in the alternate reference channel $B_s^0 \rightarrow J/\psi \phi$	102
74	F Continuum BDT	102
75	G Fit for the reference channel yield extraction: detailed discussion of the analysis	124
76	G.1 Datasets and event selection	124
77	G.2 Mass Fit Strategy	124
78	G.2.1 Fit Structure	124
79	G.2.2 Fit Initialisation	124
80	G.2.3 Goodness of fit (GoF)	124
81	G.3 Choice of fit models	124
82	G.3.1 Signal model	124
83	G.3.2 Splitting of partially reconstructed decays	124
84	G.3.3 $J/\psi \pi$ and Combinatorial background models	124
85	G.4 Default Fit Parameter counts	124
86	G.5 Evaluation of systematic uncertainties	124
87	G.6 Fit validation studies	124
88	G.7 Other crosschecks	124
89	G.8 $B^\pm \rightarrow J/\psi K^\pm$ reference channel yield	124

90	G.9 $B^\pm \rightarrow J/\psi K^\pm$ Appendix	124
91	H 2D Neyman belt confidence region construction	124
92	H.1 The Neyman approach to confidence region construction	125
93	H.2 2D neyman belt construction	125
94	H.3 Checks on coverage	127
95	H.4 Systematics uncertainties in 2D Neyman belt	128
96	H.4.1 Checks on systematics uncertainties introduction	131

97 1 Note Changes

- 98 • v0.1
99 – Initial version

100 2 Introduction

101 This document is the supporting note for the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis of the 2015+2016 part of the Run 2
 102 dataset. The aim is to obtain a first intermediate Run 2 ATLAS result on the $B_s \rightarrow \mu\mu$ and $B \rightarrow \mu\mu$ final
 103 states.

104 The strategy for this updated analysis is mostly following what adopted for the complete run 1 version,
 105 focusing on the possibility of a measurement of the $B_s \rightarrow \mu\mu$ branching fraction and taking advantage of
 106 the Run 2 statistics available in this first Run 2 iteration. Wherever possible the approach will be simplified
 107 in favor of a leaner analysis, leaving – within reason – the ultimate exploitation of the sample sensitivity
 108 to for the full Run 2 dataset study. The trigger and muon quality selections – for instance – are simplified
 109 with little loss in statistical power and background rejection, but with a much simpler analysis strategy.

110 Theoretical prediction on the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching ratios are $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (3.65 \pm 0.23) \times 10^{-9}$
 111 and $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) = (1.06 \pm 0.09) \times 10^{-10}$ [1]. CMS and LHCb have a combined Run 1 result [2, 3]
 112 showing a $>5\sigma$ effect for the $B_s \rightarrow \mu\mu$ final state giving an average branching ratio of $(2.8^{+0.7}_{-0.6}) \times 10^{-9}$,
 113 and $>3\sigma$ evidence for $B^0 \rightarrow \mu^+ \mu^-$ with a central BR value of $3.6^{+1.6}_{-1.4} \times 10^{-10}$. ATLAS has limited trigger
 114 efficiency and mass resolution, resulting in a degraded sensitivity to these decays.¹ The combined Run
 115 1 ATLAS result shows a sensitivity comparable to expectations, with a measured $B_s^0 \rightarrow \mu^+ \mu^-$ BR of
 116 $0.9^{+1.1}_{-0.8} \times 10^{-9}$ and an upper bound on $B^0 \rightarrow \mu^+ \mu^-$ of 4.2×10^{-10} at 95% CL. In order to exploit at
 117 best ATLAS data, we updated the analysis re-optimising the background rejection and the BR extraction
 118 methodology.

119 The main idea guiding this analysis is to increase the sensitivity to the signal by using a mass fit on the widest
 120 possible set of events, and improve the $B^0 \rightarrow \mu^+ \mu^- / B_s^0 \rightarrow \mu^+ \mu^-$ correlation with a better exploitation
 121 of the signal subsample with the smallest mass resolution. A loose selection will be applied to retain a
 122 maximum of signal events, and the final fit will distinguish between signal and the various backgrounds,
 123 as well as different resolution components within the signal sample. Finally the peaking background
 124 will be a major contributor under the signal peaks, differing from dimuonic decays only through muon
 125 identification. The improved Run 2 muon reconstruction allows this iteration of the analysis to rely on
 126 standard MCP categories for the separation of $B \rightarrow hh'$ from $B_{(s)}^0 \rightarrow \mu^+ \mu^-$.

127 In this note, we refer mainly to the internal documentation on the previous analysis on the full Run 1
 128 dataset [4, 5], on the full 2011 data set [6] and the studies carried out for the first ATLAS analysis in this
 129 mode [7].

130 The reference formula for the branching ratio measurement is similar to the one previously used in [7]:

$$\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-) = \mathcal{B}(B^\pm \rightarrow J/\psi K^\pm \rightarrow \mu^+ \mu^- K^\pm) \times \frac{f_u}{f_s} \times N_{\mu^+ \mu^-} \times \left(N_{J/\psi K^\pm} \frac{(A\epsilon)_{\mu^+ \mu^-}^k}{(A\epsilon)_{J/\psi K^\pm}^k} \right)^{-1}, \quad (1)$$

131 and simplified with respect to the final Run 1 analysis thanks to the fact that we employ one single trigger
 132 category encompassing $\sim 80 - 85\%$ of the signal for the 2015+2016 Run 2 dataset.
 133 While the branching ratio $BR(B^\pm \rightarrow J/\psi K^\pm \rightarrow \mu^+ \mu^- K^\pm)$ and the relative B_u/B_d production fraction $\frac{f_u}{f_s}$
 134 are derived from other experimental results, a good fraction of this document is devoted to the derivation

¹ We performed tests showing the detailed breakdown of the mass resolution and signal statistics contributions to the analysis, reported later in section 15.

135 of the remaining ingredients to this formula and their uncertainties.

136 The measured $B^+ \rightarrow J/\psi K^+$ yield $N_{J/\psi K^\pm}$ will be derived in section 12, while the relative efficiencies
137 and acceptances of $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ and $B^+ \rightarrow J/\psi K^+$ will be extracted in 13. The $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ yield will
138 be derived from invariant mass distribution fits in section 14, implementing the selections optimized in
139 sections 8, 10 and 7.

140 Given these ingredients, the branching ratio to the left of the reference formula above will be derived in
141 section 15, where relevant systematic effects will be propagated as well. Dedicated sections will discuss
142 the tuning of Montecarlo samples 5, the data-driven 11 extrapolation of MC models to the signal region
143 9.

¹⁴⁴ 3 Data and Monte Carlo Samples

¹⁴⁵ 3.1 Data Samples

¹⁴⁶ This analysis uses the first part of the ATLAS Run 2 dataset consisting of $\sqrt{s} = 13$ TeV collision data
¹⁴⁷ taken with stable LHC beams in the years 2015 and 2016. The ATLAS muon and tracking detectors,
¹⁴⁸ which are essential for the reconstruction of the $B_{(s)}^0$ and B^\pm mesons, are required to be fully operational.
¹⁴⁹ These requirements yield an integrated luminosity of $\approx 39 \text{ fb}^{-1}$.

¹⁵⁰ For the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ signal channel a region of 360 MeV width around the B_s^0 mass is omitted during the
¹⁵¹ analysis development to avoid any bias (“blinding”)².

¹⁵² The derived AOD (DAOD) samples used in this analysis are produced by the BPHY8 derivation format:

- ¹⁵³ • Data 2015, physics_Main stream:
¹⁵⁴ data15_13TeV.period[P].physics_Main.PhysCont.DAOD_BPHY8.grp15_v01_p3372
¹⁵⁵ with periods P A, C, D, E, F, G, H, J.
- ¹⁵⁶ • Data 2106, physics_Main stream:
¹⁵⁷ data16_13TeV.period[P].physics_Main.PhysCont.DAOD_BPHY8.grp16_v01_p3372
¹⁵⁸ with periods P A, B, C, D, E, F, G, I, K, L.
- ¹⁵⁹ • Data 2016, physics_BphysDelayed stream:
¹⁶⁰ data16_13TeV.period[P].physics_BphysDelayed.PhysCont.DAOD_BPHY8.grp16_v01_p3372
¹⁶¹ with periods P D, E, F, G, I, K, L.

¹⁶² 3.2 Monte Carlo Samples

¹⁶³ Simulated Monte Carlo data samples are required for most of the analysis steps. Dedicated MC data
¹⁶⁴ samples were produced, see Table 1.

¹⁶⁵ For each sample, the number of events generated is given together with the details on the generation:
¹⁶⁶ PYTHIA8B plus EvtGen is used for most samples except for the $B \rightarrow hh'$ peaking background and the
¹⁶⁷ $b\bar{b} \rightarrow \mu^+ \mu^- X$ continuum background channels where only PYTHIA8B is used. EvtGen is used for the
¹⁶⁸ reference channels including a J/ψ in the final state in order to correctly account for the J/ψ polarisation
¹⁶⁹ effects. The $B_s^0 \rightarrow J/\psi \phi$ control channel, using PyTHIA8B and Photos is taken from the samples produced
¹⁷⁰ for the $B_s^0 \rightarrow J/\psi \phi$ analysis within the J/ψ analysis subgroup: the sample is generated flat from the
¹⁷¹ angular point of view and maps are used to obtain the correct angular distribution. Most of the samples
¹⁷² are processed with Atlfast-II that employs the fast detector simulation for the calorimeter by means of
¹⁷³ parameterisations of the longitudinal and lateral energy profile, while the muon and tracking parts are fully
¹⁷⁴ simulated. Due the less accurate simulation of calorimetry in Atlfast-II, this cannot be used to estimate
¹⁷⁵ the muon fake rates, so the $B \rightarrow hh'$ peaking background are processed with the full simulation in order
¹⁷⁶ to have an accurate description of the hadronic contributions.

¹⁷⁷ Further details on the datasets are given in Table 2 (for AODs) and Table 3 (for DAODs).

1: This sentence will need to be adjusted once we have the re-processed DAODs.

2: Please note that the p-tags of these data samples will change again as we reprocess our data DAODs to include the mass values from the blinded region in a blinded/encrypted way.

3: Please note that the DAOD datasets will need adjustment once we will have produced the final round of MC DAOD.

² Future footnote text: Only once the unblinding decision will be given we will decrypt the blinded mass values using our unblinding key. This procedure avoids the time needed to re-run the derivation on data once we are allowed to unblind while it still keeps us from accidentally looking at the mass values in the blinded region.

Channel	Type	Events	Generator	Simulation
$B_s^0 \rightarrow \mu^+ \mu^-$	signal	1,000,000	PYTHIA8B + EvtGen	Atlfast-II
$B_{(s)}^0 \rightarrow \mu^+ \mu^-$	signal	1,000,000	PYTHIA8B + EvtGen	Atlfast-II
$B^+ \rightarrow J/\psi K^+$ with $J/\psi \rightarrow \mu^+ \mu^-$	reference	1,997,000	PYTHIA8B + EvtGen	Atlfast-II
$B^- \rightarrow J/\psi K^-$ with $J/\psi \rightarrow \mu^+ \mu^-$	reference	1,999,500	PYTHIA8B + EvtGen	Atlfast-II
$B^+ \rightarrow J/\psi \pi^+$ with $J/\psi \rightarrow \mu^+ \mu^-$	reference	498,000	PYTHIA8B + EvtGen	Atlfast-II
$B^- \rightarrow J/\psi \pi^-$ with $J/\psi \rightarrow \mu^+ \mu^-$	reference	500,000	PYTHIA8B + EvtGen	Atlfast-II
$B_s^0 \rightarrow J/\psi \phi$ with $J/\psi \rightarrow \mu^+ \mu^-$, $\phi \rightarrow K^+ K^-$	control	5,000,000	PYTHIA8B + Photospp	Atlfast-II
$B \rightarrow hh'$	peaking bkg.	5,000,000	PYTHIA8B	full simulation
$B_s^0 \rightarrow K^- \mu^+ \nu$	part. rec. bkg.	250,000	PYTHIA8B + EvtGen	Atlfast-II
$B_s^0 \rightarrow \pi^- \mu^+ \nu$	part. rec. bkg.	500,000	PYTHIA8B + EvtGen	Atlfast-II
$\Lambda^0 \rightarrow p \mu^- \bar{\nu}$	part. rec. bkg.	250,000	PYTHIA8B + EvtGen	Atlfast-II
$b\bar{b} \rightarrow J/\psi X$ with $J/\psi \rightarrow \mu^+ \mu^-$	cont. bkg.	10,000,000	PYTHIA8B + EvtGen	Atlfast-II
$b\bar{b} \rightarrow \mu^+ \mu^- X$	cont. bkg.	200,000,000	PYTHIA8B	Atlfast-II

Table 1: Monte Carlo data samples for signal, reference, control and background channels. The background channels are sub-divided into peaking background (peaking bkg.), partially reconstructed background (part. rec. bkg.) and continuum background (cont. bkg.).

178 3.3 Ntuples

179 These DxAODs are processed using an EventLoop based package; the resulting flat ntuples are placed
 180 both in the Grid and in the storage facilities of the HEP group of the University of Siegen. Table 4 lists
 181 the MC ntuples corresponding to the data used as well as the number of events and candidates.

182 For most studies, these ntuples end up been too large. Therefore some cuts are applied to skim them:

$$183 |\eta(K^{+/-})| < 2.5$$

$$184 p_T(K^{+/-}) > 800\text{MeV}$$

185 The kaons are required to be identified as loose tracks.

$$186 \chi^2/Ndof(J/\psi) < 13$$

$$187 M(J/\psi) \in [2900, 3300]\text{MeV}$$

$$188 M[\phi] \in [959, 1079]\text{MeV}$$

$$189 M[B] \in [4700, 6000]\text{MeV}$$

$$190 p_T(B) > 6000\text{MeV}$$

$$191 |\eta(B)| < 2.5$$

$$192 \chi^2/Ndof(B) < 10$$

193 Here B means either $B_{(s)}^0$ or B^+ and the mass cut is applied to the muon based mass.

$$194 p_T(\mu) > 4000\text{MeV}$$

$$195 |\eta(\mu)| < 2.5$$

196 The muons are expected to be loose and pass the MCP requirements. Additionally, the events are required
 197 to pass the GRL and to be collected only when the pixel, SCT, TRT, LAr, Tile and Muon subdetectors had
 198 no problems. The list of MC skimmed samples is shown in 5, together with the skimming efficiencies.

199 The data was treated in exactly the same way as the MC. Table 6 shows the ntuples produced from the
 200 DxAODs for data collected in 2015. After skimming the statistics vary as shown in table 7.

201 In 2016, period D4, the delayed stream was turned on and from this period we have two copies of the data,
202 collected in different streams. Table 8 shows the data for the delayed stream and for different periods,
203 after skimming we get what is shown in table 9.

204

205 On the other hand the delayed stream, before the skimming, is showing in 10, 11 shows what we get after
206 skimming.

Channel	#events	Dataset
Bsmumu	1,000,000	mc16_13TeV.300426.Pythia8BEvtGen_A14_CTEQ6L1_Bs_mu3p5mu3p5.merge.AOD.e4889_e5984_a875_r9364_r9315
Bdmunu	1,000,000	mc16_13TeV.300430.Pythia8BEvtGen_A14_CTEQ6L1_Bd_mu3p5mu3p5.merge.AOD.e4889_e5984_a875_r9364_r9315
BplpsiKp	1,997,000	mc16_13TeV.300404.Pythia8BEvtGen_A14_CTEQ6L1_Bp_Jpsi_mu3p5mu3p5_Kp_BMassFix.merge.AOD.e4862_e5984_a875_r9364_r9315
BmJpsiKm	1,999,500	mc16_13TeV.300405.Pythia8BEvtGen_A14_CTEQ6L1_Bm_Jpsi_mu3p5mu3p5_Km_BMassFix.merge.AOD.e4862_e5984_a875_r9364_r9315
BsJpsiPhi	5,000,000	mc16_13TeV.300438.Pythia8BPhotosp_A14_CTEQ6L1_Bs_Jpsimu3p5mu3p5_phi.merge.AOD.e4922_e5984_a875_r9364_r9315
BpJpsiPip	498,000	mc16_13TeV.300406.Pythia8BEvtGen_A14_CTEQ6L1_Bp_Jpsi_mu3p5mu3p5_Pip_BMassFix.merge.AOD.e4862_e5984_a875_r9364_r9315
BmJpsiPim	500,000	mc16_13TeV.300437.Pythia8BEvtGen_A14_CTEQ6L1_Bm_Jpsi_mu3p5mu3p5_Pim_BMassFix.merge.AOD.e4862_e5984_a875_r9364_r9315
Bhh	500,0000	mc16_13TeV.300432.Pythia8BEvtGen_A14_CTEQ6L1_B_hh.merge.AOD.e4889_e5984_s3126_r364_r9315
BsKmmunu	250,000	mc16_13TeV.300433.Pythia8BEvtGen_A14_CTEQ6L1_Bs_K3p5mu3p5nu.merge.AOD.e4720_e5984_a875_r9364_r9315
BdPmmunu	500,000	mc16_13TeV.300433.Pythia8BEvtGen_A14_CTEQ6L1_Bd_pi3p5mu3p5nu.merge.AOD.e4720_e5984_a875_r9364_r9315
LbPmmunu	250,000	mc16_13TeV.300434.Pythia8BEvtGen_A14_CTEQ6L1_Lambda0b_p3p5mu3p5nu.merge.AOD.e4720_e5984_a875_r9364_r9315
bbmmunu	49,999,000	mc16_13TeV.300306.Pythia8B_A14_CTEQ6L1_bb_mu3p5mu3p5_Py8RepDec.merge.AOD.e4911_e5984_a875_r9364_r9315
bbJpsimunu	10,000,000	mc16_13TeV.300203.Pythia8BPhotosp_A14_CTEQ6L1_bb_Jpsimu3p5mu3p5.merge.AOD.e4889_e5984_a875_r9364_r9315
bbmmunuX	200,000,000	mc16_13TeV.300307.Pythia8B_A14_CTEQ6L1_bb_mu3p5mu3p5_Py8RepDec_4to6p5GeV.merge.AOD.e6179_e5984_a875_r9364_r9315

Table 2: Monte Carlo data samples used (AOD).

Channel	#events	Dataset
Bsmumu	918,488	mc16_13TeV.300426.Pythia8BEvtGen_A14_CTEQ6L1_Bs_mu3p5mu3p5.deriv.DAOD_BPHY8_e4889_e5984_a875_i9364_i9315_p3371
Bdmunu	913,507	mc16_13TeV.300430.Pythia8BEvtGen_A14_CTEQ6L1_Bd_mu3p5mu3p5.deriv.DAOD_BPHY8_e4889_e5984_a875_i9364_i9315_p3371
Bp1psiKp	1,610,177	mc16_13TeV.300404.Pythia8BEvtGen_A14_CTEQ6L1_Bp_jpsi_mu3p5mu3p5_Kp_BMassFix.deriv.DAOD_BPHY8_e4862_e5984_a875_i9364_i9315_p3371
BmpsiKm	1,617,566	mc16_13TeV.300405.Pythia8BEvtGen_A14_CTEQ6L1_Bm_jpsi_mu3p5mu3p5_Km_BMassFix.deriv.DAOD_BPHY8_e4862_e5984_a875_i9364_i9315_p3371
Bs1psiPhi	3,555,942	mc16_13TeV.300408.Pythia8BPhotospp_A14_CTEQ6L1_Bs_jpsi_mu3p5mu3p5_phi.deriv.DAOD_BPHY8_e4922_e5984_a875_i9364_i9315_p3371
Bp1psiPip	399,607	mc16_13TeV.300406.Pythia8BEvtGen_A14_CTEQ6L1_Bp_jpsi_mu3p5mu3p5_Pip_BMassFix.deriv.DAOD_BPHY8_e4862_e5984_a875_i9364_i9315_p3371
BmpsiPim	405,710	mc16_13TeV.300437.Pythia8BEvtGen_A14_CTEQ6L1_Bm_jpsi_mu3p5mu3p5_Pim_BMassFix.deriv.DAOD_BPHY8_e4862_e5984_a875_i9364_i9315_p3371
Bhh	4,002,651	mc16_13TeV.300431.Pythia8B_A14_CTEQ6L1_B_hh.deriv.DAOD_BPHY8_e4889_e5984_s3126_r9364_i9315_p3371
BsKmuu	15,786	mc16_13TeV.300432.Pythia8BEvtGen_A14_CTEQ6L1_Bs_K3p5mu3p5mu.deriv.DAOD_BPHY8_e4720_e5984_a875_i9364_i9315_p3371
BdPmuu	30,932	mc16_13TeV.300433.Pythia8BEvtGen_A14_CTEQ6L1_Bd_p3p5mu3p5mu.deriv.DAOD_BPHY8_e4720_e5984_a875_i9364_i9315_p3371
LbFmuu	15,768	mc16_13TeV.300434.Pythia8BEvtGen_A14_CTEQ6L1_Lambda0b_p3p5mu3p5mu.deriv.DAOD_BPHY8_e4720_e5984_a875_i9364_i9315_p3371
bbnunuX	14,666,736	mc16_13TeV.300306.Pythia8B_A14_CTEQ6L1_bb_mu3p5mu3p5_Py8RepDec.deriv.DAOD_BPHY8_e4911_e5984_a875_i9364_i9315_p3371
bbJpsimuX	6,807,750	mc16_13TeV.300203.Pythia8BPhotospp_A14_CTEQ6L1_bb_jpsi_mu3p5mu3p5.deriv.DAOD_BPHY8_e4889_e5984_a875_i9364_i9315_p3371

Table 3: Derived Monte Carlo data samples used (DAOD).

Process	DSID	Events	Candidates
BJpsiK	300203	5809625	17024595
Bmumu	300203	5809625	180305
BsJpsiPhi	300203	5809625	9561422
BJpsiK	300306	12884089	1650341
Bmumu	300306	12884089	12859940
BsJpsiPhi	300306	12884089	1409159
BJpsiK	300307	298737322	23934328
Bmumu	300307	298737322	302068067
BsJpsiPhi	300307	298737322	21794420
BJpsiK	300404	1469407	3892138
BJpsiK	300405	1475675	3899669
BJpsiK	300406	364370	962642
BJpsiPi	300406	364370	943066
Bmumu	300426	835007	845514
Bmumu	300430	830497	841378
Bhh	300431	4002651	4546553
Bhmunu	300432	11116	11259
Bhmunu	300433	21801	22132
Bhmunu	300434	11037	11186
BJpsiK	300437	369854	979292
BJpsiPi	300437	369854	959678
BsJpsiPhi	300438	2933804	8432713

Table 4: The table shows the statistics for each MC sample after processing the corresponding DxAOD.

Process	DSID	Events	Candidates	ϵ_{evt}	ϵ_{cand}
BJpsiK	300203	1416618	2194907	0.244	0.230
Bmumu	300203	6815	6823	0.001	0.001
BsJpsiPhi	300203	209968	243234	0.036	0.025
BJpsiK	300306	40852	68857	0.003	0.049
Bmumu	300306	2514304	2514717	0.195	1.785
BsJpsiPhi	300306	6795	8303	0.001	0.006
BJpsiK	300307	538086	933586	0.002	0.043
Bmumu	300307	71988765	72000498	0.241	3.304
BsJpsiPhi	300307	96572	121657	0.000	0.006
BJpsiK	300404	885111	1236759	0.602	0.318
BJpsiK	300405	886421	1238101	0.601	0.317
BJpsiK	300406	216982	303968	0.595	0.322
BJpsiPi	300406	220599	304779	0.605	0.323
Bmumu	300426	531149	531246	0.636	0.628
Bmumu	300430	524753	524870	0.632	0.624
BJpsiK	300437	220615	309061	0.596	0.322
BJpsiPi	300437	224180	309469	0.606	0.322
BsJpsiPhi	300438	1773151	1903220	0.604	0.226

Table 5: The table shows the statistics for each MC sample after skimming as well as the corresponding efficiencies.

Process	Period	Events	Candidates
Bmumu	data15D	2051499	938220
BJpsiK	data15D	2051499	5227354
BsJpsiPhi	data15D	2051499	5609694
Bmumu	data15E	8933537	3200040
BJpsiK	data15E	8933537	24306111
BsJpsiPhi	data15E	8933537	23328266
Bmumu	data15F	5976830	2092234
BJpsiK	data15F	5976830	16297978
BsJpsiPhi	data15F	5976830	15265921
Bmumu	data15G	13088718	4469795
BJpsiK	data15G	13088718	35691929
BsJpsiPhi	data15G	13088718	32781065
Bmumu	data15H	8605283	2891503
BJpsiK	data15H	8605283	24098490
BsJpsiPhi	data15H	8605283	21953503
Bmumu	data15J	23303091	7500796
BJpsiK	data15J	23303091	63679808
BsJpsiPhi	data15J	23303091	58153686

Table 6: The table shows the statistics for each data sample after processing the corresponding DxAOD.

Process	Period	Events	Candidates	ϵ_{evt}	ϵ_{cand}
BsJpsiPhi	data15D	32242	43424	0.016	0.008
Bmumu	data15D	30794	30814	0.015	0.005
BJpsiK	data15D	242668	450908	0.118	0.080
BsJpsiPhi	data15E	292214	392172	0.033	0.017
Bmumu	data15E	295506	295629	0.033	0.013
BJpsiK	data15E	2272880	4229780	0.254	0.181
Bmumu	data15F	197056	197147	0.033	0.013
BsJpsiPhi	data15F	195349	261670	0.033	0.017
BJpsiK	data15F	1519996	2827015	0.254	0.185
Bmumu	data15G	466846	467059	0.036	0.014
BsJpsiPhi	data15G	462125	618805	0.035	0.019
BJpsiK	data15G	3601782	6683521	0.275	0.204
BJpsiK	data15H	1219020	2251621	0.142	0.103
Bmumu	data15H	155498	155573	0.018	0.007
BsJpsiPhi	data15H	157055	210229	0.018	0.010
BsJpsiPhi	data15J	851357	1143935	0.037	0.020
Bmumu	data15J	807398	807810	0.035	0.014
BJpsiK	data15J	6398341	11725613	0.275	0.202

Table 7: The table shows the statistics for each data sample after skimming as well as the corresponding efficiencies.

Process	Period	Events	Candidates
BsJpsiPhi	data16A	8370006	21009705
BJpsiK	data16A	8370006	22775873
Bmumu	data16A	8370006	2747636
BJpsiK	data16B	14119467	36001490
BsJpsiPhi	data16B	14119467	33868823
Bmumu	data16B	14119467	5041451
BJpsiK	data16C	22271287	56988999
Bmumu	data16C	22271287	7857371
BsJpsiPhi	data16C	22271287	52738628
BJpsiK	data16D	23419964	53311428
Bmumu	data16D	23419964	10841102
BsJpsiPhi	data16D	23419964	56002153
BJpsiK	data16E	6754013	14158864
Bmumu	data16E	6754013	3544544
BsJpsiPhi	data16E	6754013	15928262
Bmumu	data16F	13657285	7121722
BJpsiK	data16F	13657285	29266016
BsJpsiPhi	data16F	13657285	33105811
Bmumu	data16G	17607835	8279994
BJpsiK	data16G	17607835	40806289
BsJpsiPhi	data16G	17607835	43598490
Bmumu	data16I	43516442	15912079
BJpsiK	data16I	43516442	117290037
BsJpsiPhi	data16I	43516442	107292923
Bmumu	data16K	20164603	7619686
BJpsiK	data16K	20164603	54124782
BsJpsiPhi	data16K	20164603	50176431
Bmumu	data16L	43199842	18315574
BJpsiK	data16L	43199842	109996718
BsJpsiPhi	data16L	43199842	108405448

Table 8: The table shows the statistics for each data sample after processing the corresponding Dx AOD.

Process	Period	Events	Candidates	ϵ_{evt}	ϵ_{cand}
Bmumu	data16A	248559	248691	0.030	0.091
BsJpsiPhi	data16A	302999	410399	0.036	0.149
BJpsiK	data16A	2144228	3865576	0.256	1.407
BsJpsiPhi	data16B	604215	846929	0.043	0.168
Bmumu	data16B	652888	653210	0.046	0.130
BJpsiK	data16B	3553973	6261145	0.252	1.242
Bmumu	data16C	1031250	1031757	0.046	0.020
BJpsiK	data16C	5787028	10226250	0.260	0.194
BsJpsiPhi	data16C	948201	1322220	0.043	0.025
BsJpsiPhi	data16D	927638	1352621	0.040	0.024
BJpsiK	data16D	4709076	8265283	0.201	0.148
Bmumu	data16D	1622116	1622912	0.069	0.029
Bmumu	data16E	580855	581128	0.086	0.036
BJpsiK	data16E	1147406	2007278	0.170	0.126
BsJpsiPhi	data16E	247591	368970	0.037	0.023
BJpsiK	data16F	2336963	4083902	0.171	0.123
BsJpsiPhi	data16F	524275	786354	0.038	0.024
Bmumu	data16F	1093008	1093545	0.080	0.033
Bmumu	data16G	1174664	1175251	0.067	0.027
BJpsiK	data16G	3480530	6222078	0.198	0.143
BsJpsiPhi	data16G	699884	1029623	0.040	0.024
Bmumu	data16I	2187799	2188938	0.050	0.020
BJpsiK	data16I	11697980	21887603	0.269	0.204
BsJpsiPhi	data16I	2066692	2922492	0.047	0.027
BsJpsiPhi	data16K	876925	1226538	0.043	0.024
BJpsiK	data16K	5262863	9876865	0.261	0.197
Bmumu	data16K	972602	973089	0.048	0.019
Bmumu	data16L	2414501	2415995	0.056	0.022
BJpsiK	data16L	10399239	19340053	0.241	0.178
BsJpsiPhi	data16L	1816353	2584612	0.042	0.024

Table 9: The table shows the statistics for each data sample after skimming as well as the corresponding efficiencies.

Process	Period	Events	Candidates
Bmumu	data16D	23919490	4594980
BsJpsiPhi	data16D	23919490	58899128
BJpsiK	data16D	23919490	72781108
BJpsiK	data16E	11830122	35866106
BsJpsiPhi	data16E	11830122	29062384
Bmumu	data16E	11830122	2343103
Bmumu	data16F	21596238	4455413
BJpsiK	data16F	21596238	65557864
BsJpsiPhi	data16F	21596238	54617526
BJpsiK	data16G	24486001	74754106
Bmumu	data16G	24486001	5045472
BsJpsiPhi	data16G	24486001	62548191
BsJpsiPhi	data16I	35909554	93394650
Bmumu	data16I	35909554	7471927
BJpsiK	data16I	35909554	110040675
Bmumu	data16K	15632434	3327032
BsJpsiPhi	data16K	15632434	40446754
BJpsiK	data16K	15632434	47896820
Bmumu	data16L	37920208	8257564
BJpsiK	data16L	37920208	117150856
BsJpsiPhi	data16L	37920208	101644186

Table 10: The table shows the statistics for each data sample after processing the corresponding DxAOD.

Process	Period	Events	Candidates	ϵ_{evt}	ϵ_{cand}
Bmumu	data16D	303263	303532	0.013	0.004
BJpsiK	data16D	8602722	15439040	0.360	0.212
BsJpsiPhi	data16D	1296704	1756392	0.054	0.024
Bmumu	data16E	159352	159549	0.013	0.068
BsJpsiPhi	data16E	618808	834377	0.052	0.356
BJpsiK	data16E	4216850	7585223	0.356	3.237
Bmumu	data16F	291083	291462	0.013	0.005
BsJpsiPhi	data16F	1171630	1593864	0.054	0.029
BJpsiK	data16F	7583636	13600482	0.351	0.249
BsJpsiPhi	data16G	1294595	1763975	0.053	0.028
Bmumu	data16G	319273	319707	0.013	0.005
BJpsiK	data16G	8323786	14948099	0.340	0.239
Bmumu	data16I	471665	472359	0.013	0.004
BJpsiK	data16I	12418032	22363822	0.346	0.203
BsJpsiPhi	data16I	2089667	2878929	0.058	0.026
BJpsiK	data16K	5234098	9466493	0.335	0.198
Bmumu	data16K	197661	197937	0.013	0.004
BsJpsiPhi	data16K	865415	1184687	0.055	0.025
Bmumu	data16L	470676	471516	0.012	0.005
BJpsiK	data16L	12896768	23406743	0.340	0.230
BsJpsiPhi	data16L	2335331	3235032	0.062	0.032

Table 11: The table shows the statistics for each data sample after skimming as well as the corresponding efficiencies.

207 4 Candidate Preselection

208 4.1 Candidate building and preselection on derivation level

209 For the Run 2 $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis the $B_{(s)}^0$ and J/ψ candidates are build simultaneously by the derivation
 210 format BPHY8 (for data and continuum background) in the three relevant decay modes:

- 211 • $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ signal channel

212 A $B_{(s)}^0$ candidate is build from two oppositely-charged muons. Before vertexing the muon momenta
 213 are corrected by the MuonCalibrationAndSmearingTool as recommended by the muon combined
 214 performance group, applying the appropriate configurations³ for data15 and data16. Using the
 215 JpsiFinder tool, the $B_{(s)}^0$ candidates are constructed by a fit of the two muon's inner detector
 216 tracks to a common vertex requiring $\chi^2_{B_{(s)}^0} / NDF < 15$ and $3500 \text{ MeV} < m_{B_{(s)}^0} < 7000 \text{ MeV}$. In
 217 addition, the so called muon-based mass $m_{B_{(s)}^0}^{\text{MUCALC}}$ and the corresponding uncertainty are calculated
 218 as the invariant mass from the four-momentum sum using the combined muon information of the
 219 two muons comprising the candidate. The PDG-mass for muons is applied in this calculation. On
 220 data, the candidates two mass values are blinded⁴ if both, $m_{B_{(s)}^0}$ and $m_{B_{(s)}^0}^{\text{MUCALC}}$ fall into the blinded
 221 region around the nominal B_s^0 , i.e. into the interval from 5166 MeV to 5526 MeV.

- 222 • $B^\pm \rightarrow J/\psi K^\pm$ reference channel with $J/\psi \rightarrow \mu^+ \mu^-$

223 First J/ψ candidates are reconstructed using the JpsiFinder similarly as for the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$
 224 signal channel with an adjusted mass range of $2000 \text{ MeV} < m_{J/\psi} < 7000 \text{ MeV}$. Then the
 225 JpsiPlus1TrackFinder tool is used to combine the J/ψ candidates with an additional track⁵
 226 which is required to have $p_T^{J\psi} > 1 \text{ GeV}$ and $|\eta| < 2.5$. A mass constraint of the J/ψ candid-
 227 ate's mass to the J/ψ PDG mass is applied. The B^\pm candidates are accepted if $p_T^{B^\pm} > 1 \text{ GeV}$,
 228 $\chi^2_{B^\pm} / NDF < 15$ and $3500 \text{ MeV} < m_{B^\pm} < 7000 \text{ MeV}$. Again, muon-based mass values $m_{B^\pm}^{\text{MUCALC}}$
 229 and corresponding uncertainties are calculated from the four-momenta of the two muons and the ad-
 230 ditional kaon track, assuming PDG values for the masses of the muons and the kaon as appropriate.
 231 No mass-value blinding is applied to the reference channel candidates.

- 232 • $B_s^0 \rightarrow J/\psi \phi$ control channel with $J/\psi \rightarrow \mu^+ \mu^-$

233 To reconstruct $B_s^0 \rightarrow J/\psi \phi$ candidates, the J/ψ candidates reconstructed as described for the
 234 $B^\pm \rightarrow J/\psi K^\pm$ channel above are combined with two oppositely-charged inner detector tracks
 235 fulfilling the same requirements as the kaon track in the $B^\pm \rightarrow J/\psi K^\pm$ reconstruction by the
 236 JpsiPlus2TracksFinder tool. Again, a mass constraint of the J/ψ candidate's mass to the J/ψ
 237 PDG mass is applied. Muon-based mass values $m_{B_s^0}^{\text{MUCALC}}$ and corresponding uncertainties are

4: The
footnote
text already
explains
the new
blinding
method
with the
encrypted
mass
values.

³ For data15: McstYear = "Data15", McstRelease = "Recs2017_08_02", McstStatComb = False, McstSagittaCorr = True, McstSagittaRelease = "sagittaBiasDataAll_25_07_17", McstDoSagittaMCDistortion = False;
 for data16: McstYear = "Data16", McstRelease = "Recs2017_08_02", McstStatComb = False, McstSagittaCorr = True, McstSagittaRelease = "sagittaBiasDataAll_25_07_17", McstDoSagittaMCDistortion = False;
 for MC the data16 settings are applied.

⁴ The mass values are encrypted using a simple RSA algorithm with asymmetric keys, multiplied by -1, so that they appear on the negative side of the mass spectrum, and stored in the same float variable.

⁵ We apply the BLS group's standard inner detector track preselection for vertexing as defined in configureVertexing.py, among them pTMin = 400.0 MeV, and the minimal track hit requirements nHitBLayer = 0, nHitPix = 1, nHitBLayerPlusPix = 1, nHitSct = 2, nHitSi = 3, NHitTrt = 0 as well as TrtMaxEtaAcceptance = 1.9.

calculated from the four-momenta of the two muons and the two additional kaon tracks, assuming PDG values for the masses of the muons and the kaons as appropriate. No mass-value blinding is applied to the control channel candidates.

For MC signal channels only the relevant decay modes are run. For the $B \rightarrow hh'$ MC the requirement that the inner detector tracks must be identified as muons is dropped. Instead two oppositely-charged tracks with $p_T > 3.5$ GeV and $|\eta| < 2.5$ are required. For the $B^+ \rightarrow J/\psi\pi^+$ Monte Carlo sample, two instances of the $B^\pm \rightarrow J/\psi K^\pm$ mode are run in parallel. The first one is assuming the PDG pion mass for the additional track, while the second one assumes the (wrongly-assigned) PDG kaon mass. The second version is needed as – due to the lack of pion-kaon separation in ATLAS – our $B^\pm \rightarrow J/\psi K^\pm$ reference sample on data will contain a contribution of $B^+ \rightarrow J/\psi\pi^+$ events located to the right of the B^\pm peak due to the wrongly-assigned mass values for the pion track.

At derivation level each B and J/ψ candidate is decorated with the additional information about e.g. the separation w.r.t. the associated primary vertex (see Section 4.3), the isolation of the candidate and others:

- Isolation of the decay candidate $I_{0.7}$ and number of tracks around the candidate $N_{0.7}^{trks}$
Both variables are defined in Table 22. Only tracks of the CP-type “loose” being part of the primary vertex associated to the secondary vertex and tracks not associated to any primary vertex are taken into account. The tracks which are not part of the secondary vertex candidate are required to have $p_T > 0.5$ GeV and to stem from within a cone of $\Delta R < 0.7$ around the B candidates momentum vector. The tracks considered also need to be close enough to the primary vertex, i.e. $\log \chi^2_{DCA} < 5$ where

$$\log \chi^2_{DCA} = \log \left(\left(\frac{d_0}{\sigma_{d_0}} \right)^2 + \left(\frac{z_0}{\sigma_{z_0}} \right)^2 \right)$$

with the transverse and longitudinal impact parameters d_0 and z_0 respectively. $N_{0.7}^{trks}$ records the number of non-decay-candidate tracks within the cone.

- Isolation of the muons $I_{0.7}^{\mu_i}$ and number of tracks around the muons $N_{0.7,\mu_i}^{trks}$
Similar to the isolation variables for the decay candidate, $I_{0.7}^{\mu_i}$ are calculated as the isolation values for the two muons within a cone of $\Delta R < 0.7$ around each muon’s momentum vector w.r.t. inner detector tracks. The number of tracks within the cone is recorded by $N_{0.7,\mu_i}^{trks}$.

- Closest track to the vertex candidate and number of close tracks N_{trks}^{close}
Link to the track closest to the decay vertex candidates from the set of tracks fulfilling the same requirements as for the determination of the isolation variable except for the missing cone size requirement and an adjusted $\log \chi^2_{DCA} < 7$ cut. In addition the number of all qualifying closest track candidates with $\log \chi^2_{DCA,SV} < 1$ is recorded as $N_{0.7,\mu_i}^{trks}$ where $\log \chi^2_{DCA,SV}$ is calculated w.r.t. the secondary vertex’ position.

- `minLogChi2ToAnyPV`
The minimum distance in χ^2 of the candidate to any primary vertex, using the refitted primary vertices associated to reconstructed secondary vertices where existing.
- Transverse, longitudinal and 3-dimensional impact parameters w.r.t. the associated primary vertex and corresponding uncertainties
- Transverse decay distance L_{xy} w.r.t. the associated primary vertex and its uncertainty
See Table 22 for the definition of L_{xy} .

- 278 • Proper decay time τ and its uncertainty

279 Only events with at least one decay candidate in one of the three decay modes are stored in the derivation
 280 output (DAOD) file. In order to save disk space on the computing grid, from the primary vertex collection
 281 only primary vertices associated to decay vertices by the primary vertex association method (see Section ??
 282 are saved. The muon collection and track collections are thinned to only contain muons or tracks which
 283 are either part of a reconstructed decay vertex, part of an associated primary vertex or identified as a track
 284 closest to a decay vertex.

285 **4.2 Candidate Preselection**

286 For $B_s^0 \rightarrow \mu^+ \mu^-$ candidates we require two muons with the candidate preselection proceeding as follows:

- 287 • first we apply the Muon Combined Performance (MCP) quality criteria:
 - 288 – 'Tight' working point
- 289 • muons kinematics cuts:
 - 290 – combined muons
 - 291 – $p_T > 4, 6 \text{ GeV}$
 - 292 – $|\eta| < 2.5$
- 293 • B_s^0 kinematics:
 - 294 – $p_T^B > 8.0 \text{ GeV}$
 - 295 – $|\eta^B| < 2.5$.
- 296 • B_s^0 decay vertex:
 - 297 – $\chi^2/NDF < 6$

298 $B \rightarrow J/\psi K^\pm$ candidates are created considering two muons forming a J/ψ and then adding a charged
 299 track for the kaon. The candidates are selected requiring all the above for muons in their final state, the
 300 B kinematics and the B candidate vertex. In addition we include specific cuts on the J/ψ vertex and the
 301 kaon in the final state:

- 302 • quality cuts on additional kaon track:
 - 303 – 'Loose' track quality selection
- 304 • kinematic cuts on additional kaon track:
 - 305 – $p_T > 1.0 \text{ GeV}$
 - 306 – $|\eta| < 2.5$
- 307 • J/ψ mass:
 - 308 – $2915 < m(J/\psi) < 3275 \text{ MeV}$
- 309 • J/ψ decay vertex:
 - 310 – $\chi^2/NDF < 10$

311 For $B_s^0 \rightarrow J/\psi\phi$ candidates we consider two muons forming a J/ψ and two charged tracks forming the ϕ .
 312 The candidates are selected requiring all the above cuts for muons in their final state, the B kinematics,
 313 the B candidate vertex, the J/ψ vertex and the kaon cuts on both kaons in the final state. In addition we
 314 have:

- 315 • ϕ mass:
 316 – $1005 < m(\phi) < 1035$ MeV
- 317 • ϕ decay vertex:
 318 – $\chi^2/NDF < 10$

319 The above selection contains minor differences with respect to the one used in the previous analysis:
 320 namely the cut on any second muon p_T in the pair forming B candidate has been raised to 6 GeV in
 321 accordance with the raised trigger selection criteria. Also we request tighter quality selection criteria
 322 for muons to achieve fake rejection comparable to the previous round of analysis without a need for a
 323 dedicated BDT, as detailed in Sec. 8.

324 4.3 Primary vertex determination

325 In full Run1 analysis a new approach to determine the primary vertex (PV) associated to the B candidate
 326 was developed, in order to exploit it also in Run2 its performance has to be tested again. The Run1
 327 approach performance has been compared to other three possible methods, in order to check for possible
 328 improvements. The 4 approaches considered are:

- 329 • PV_MAX_SUM_PT2: predefined in ATLAS, considers the sum of the squared transverse mo-
 330 mentum of the tracks associated to each PV, the chosen PV is the one with the highest sum;
- 331 • PV_MIN_A0: a backward extrapolation of the B momentum from the decay vertex is considered,
 332 the PV is chosen as the one with the shortest 3D distance from the point of closest approach (POCA)
 333 of the B extrapolation to each of the reconstructed PVs;
- 334 • PV_MIN_Z0: similar to PV_MIN_A0, but uses the distance along z from the POCA of the B
 335 extrapolation to each of the reconstructed PVs;
- 336 • PV_MIN_Z0_BA: approach developed for full Run1 analysis, the associated PV is chosen as the
 337 one with the shortest separation, along z , from the POCA of the B extrapolation to the beam line.

338 The performances of the four PV association procedure have been tested on signal MC sample.
 339 A first comparison was performed using only the information regarding the coordinates of the selected
 340 PVs and the position of the truth PV.

341 Figures 1 and 2 show respectively the distance on the xy plane and along the z direction between the selected
 342 PV and the truth PV for the four approaches. PV_MAX_SUM_PT2 shows clearly large distributions,
 343 which tells that the selected PV is often the wrong one. The other three approaches show almost identical
 344 narrow distributions.

345 In order to have a quantitative estimation of the performance of the approaches, for each PV reconstructed in
 346 the events, a χ^2 is computed to estimate the compatibility with the MC truth. The PV with the lowest χ^2 in
 347 each event is found to correspond to the PV correctly associated to the B candidate and is considered “truth-
 348 matched”. Figure 3 shows the distribution of the χ^2 of all the PVs (blue points) and the distribution of the
 349 χ^2 of the “truth-matched” PVs. Figure 4 shows the distribution of the chosen PVs using the 4 approaches

(green distributions) superimposed to the truth-matched distribution (red). While the distribution of chosen PVs for PV_MAX_SUM_PT2 shows two clear peaks, that can be identified as correct associations (left peak) and wrong associations (right peak), the other three distributions show one peak that follows the behaviour of the “truth-matched” vertices. Table 12 shows the purity for each approach, defined as the ratio between the number of correct associations and the total number of candidates. As expected, purity

approach	purity
PV_MAX_SUM_PT2	0.451 ± 0.0053
PV_MIN_A0	0.9938 ± 0.0008
PV_MIN_Z0	0.9937 ± 0.0008
PV_MIN_Z0_BA	0.9931 ± 0.0009

Table 12: purity of the four considered approaches to perform PV-SV association

for PV_MAX_SUM_PT2 is low, while for the other approaches the purities are compatible within the error. Due to the increasing pile-up environment in Run2, we checked the stability of the four approaches as a function of the number of reconstructed primary vertices in each event (which is a good proxy for the pile-up). Figure 5 shows the result of this study. PV_MAX_SUM_PT2 shows a remarkable dependence on the number of PVs, while the other approaches are stable. Given the lack of significant improvements with the new approaches and aiming for minimal changes on the analysis approach, we decide to stick to the same algorithm used in the last round of the analysis (PV_MIN_Z0_BA).

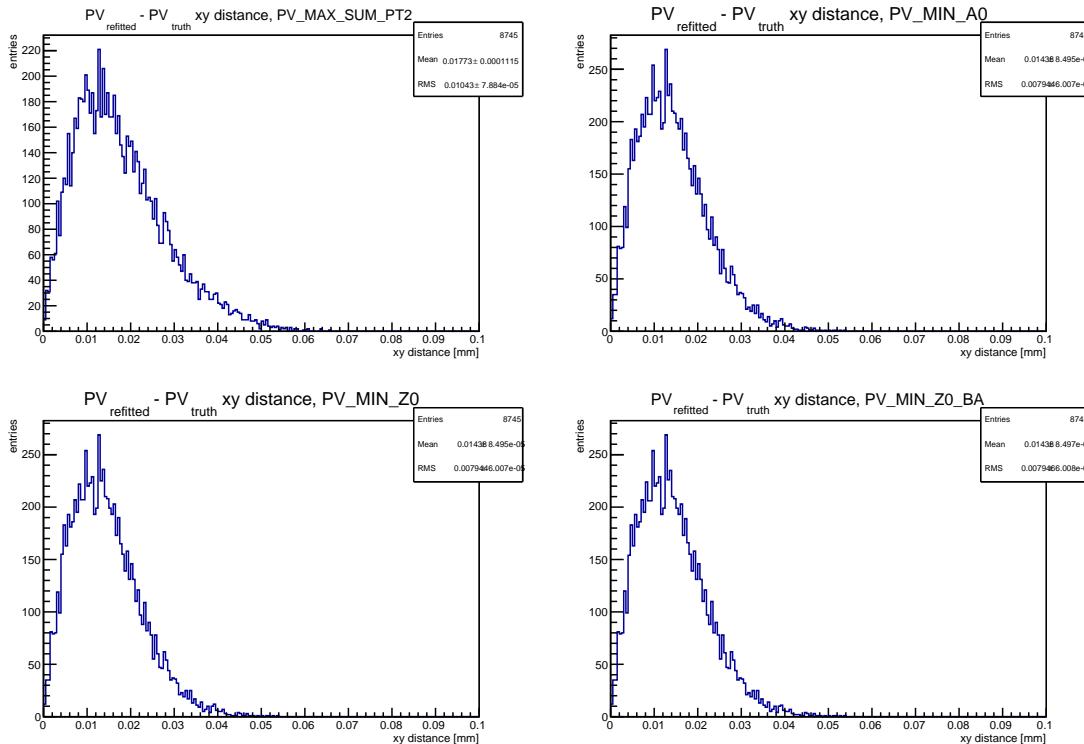


Figure 1: xy distance between selected PV and truth PV for the four approaches.

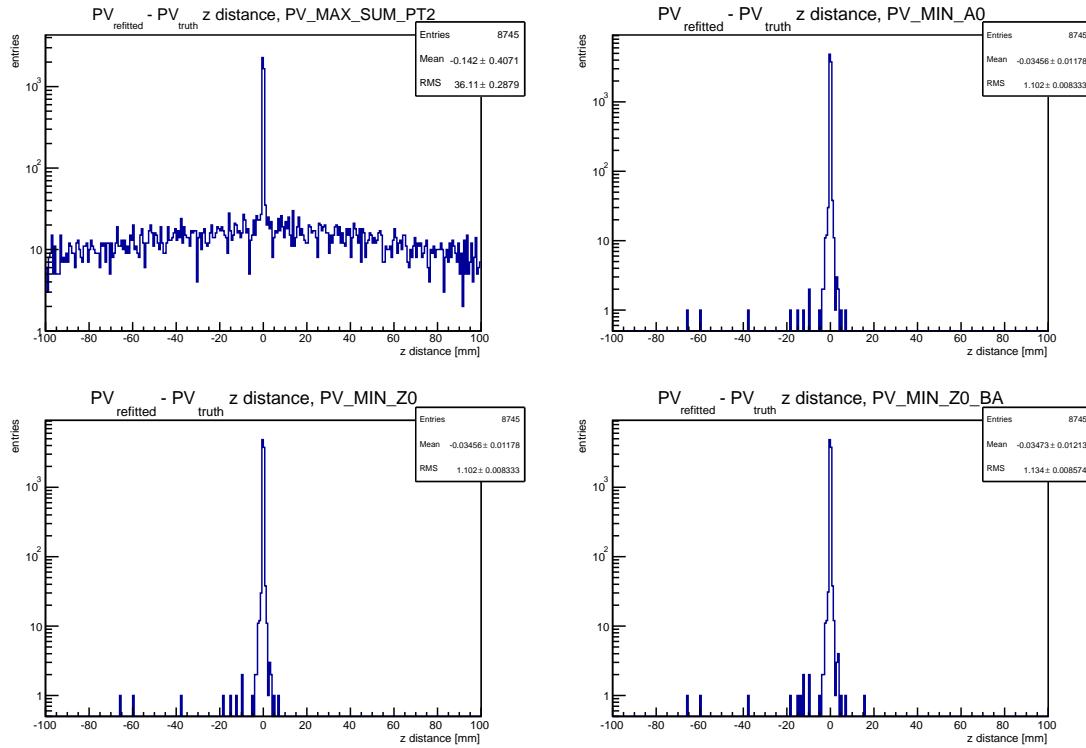


Figure 2: z distance between selected PV and truth PV for the four approaches.

362 4.4 Invariant Mass Ranges and Blinding Window

363 For the analysis we define a mass window wider than the signal one in order to be able to model the
 364 background contributions both for the signal and for the reference channel. The very loose ntuple-level
 365 window is between 3500 – 7000 MeV for all the samples.

366 The analysis windows are:

- 367 • for $B_s^0 \rightarrow \mu^+ \mu^-$ reconstruction: 4766 – 5966 MeV
 - 368 – signal region: 5166 – 5526 MeV. This is the blinded region.
- 369 • For $B^+ \rightarrow J/\psi K^+$ reconstruction: 4930 – 5630 MeV for the B^+ mass
 - 370 – for the J/ψ mass: 2915 – 3275 MeV

371 4.5 Additional selection

372 To be considered as B candidate, we require any events to pass the baseline selection cuts listed in Sec. 4.2.
 373 Then, there are additional cuts applied to reduce the background without reducing the signal. Combined,
 374 these cuts reduce the background by a factor of 2.5, without cutting into the signal (about 95% efficient).
 375 These are detailed below:

- 376 • $\Delta R < 1.5$
- 377 • $|\alpha_{2D}| < 1.0$

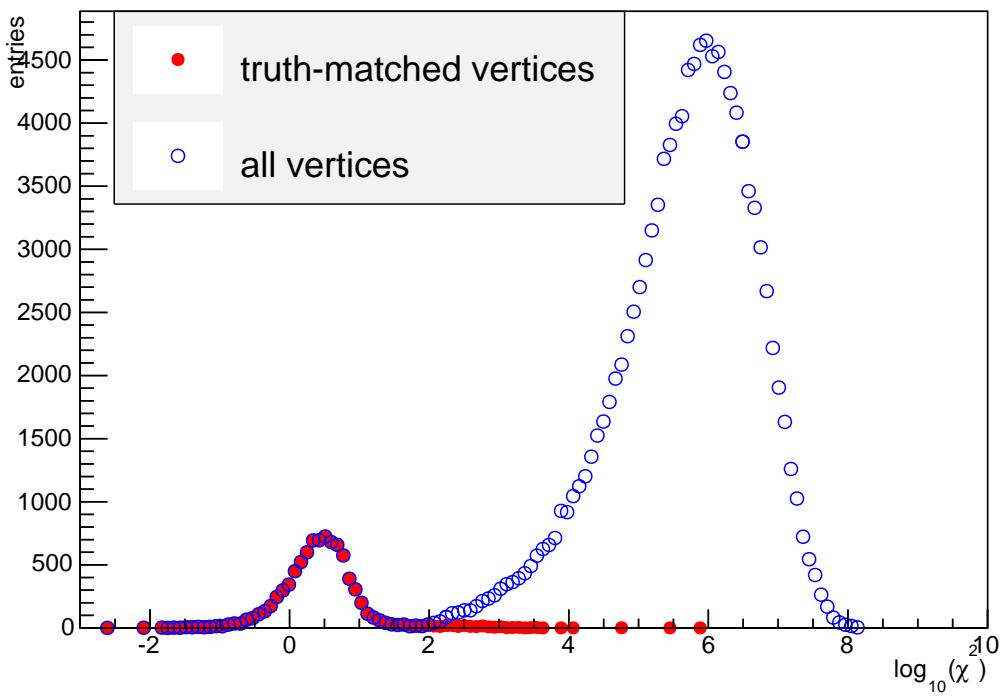


Figure 3: χ^2 of all the PVs (blue points) and the distribution of the χ^2 of the “truth-matched” PVs.

378

- $L_{xy} > 0$

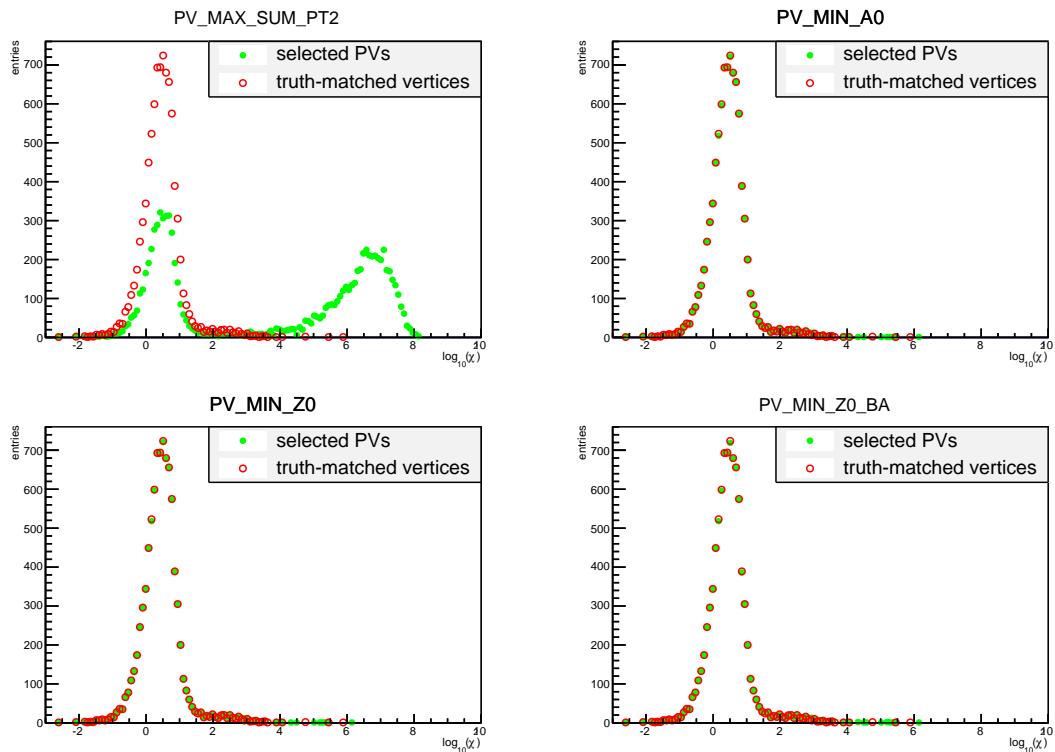


Figure 4: χ^2 distribution of the chosen PVs using the 4 approaches (green distributions) superimposed to the truth-matched distribution (red).

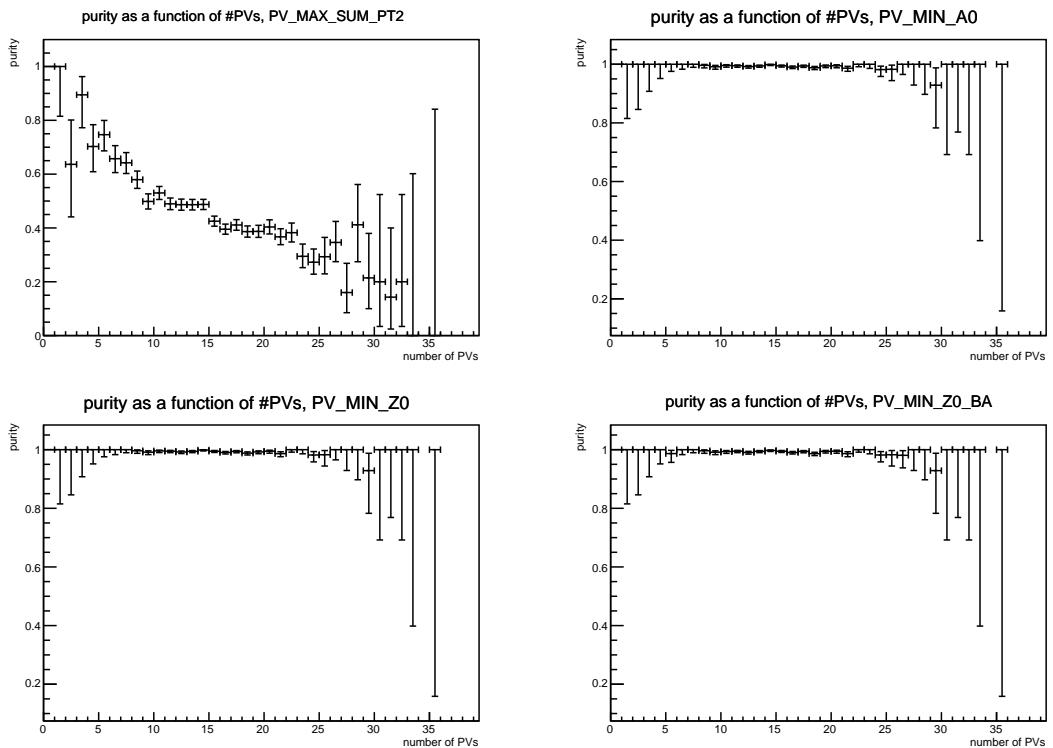


Figure 5: purity as a function of number of reconstructed PVs for the four approaches.

379 5 Kinematic Corrections to Signal MC

380 In this section we describe how the Monte Carlo samples have been tuned and cross-checked against data.
 381 The tuning procedure is two-fold: a first set of weights ("Quark Level Corrections" or QLC) accounts for
 382 the (pT_B , η_B) selection bias introduced to expedite the MC generation of b quarks; a second set of weights
 383 ("Data Driven Weights" or DDW) accounts for residual data-MC (pT_B , η_B) discrepancies, mostly arising
 384 from differences in the b quark phase space and hadronisation models from reality.
 385 All MC samples will be consistently used across the analysis after applying both QLC and DDW, which
 386 are meant to be used in combination in order to obtain the most faithful responses from the MC, as well
 387 as a direct way of accounting for residual data-MC discrepancies through the propagation of QLC/DDW
 388 uncertainties.
 389 In section 5.1 some preliminary assessments are reported, defining the dataset used, the framework used
 390 in the weights calculation and clarifying the procedure we are going to follow. In sections 5.2 and 5.3
 391 we describe the calculation of the two sets of weights, respectively QLC and DDW. More details on the
 392 procedures followed and on the checks performed are reported in appendices A and B.

393 5.1 Preliminary assessments

394 The most important use of MC in this analysis will be for the determination of relative efficiencies
 395 for the signal and control ($B^+ \rightarrow J/\psi K^+$) channels. Since the MC corrections are determined using
 396 $B^+ \rightarrow J/\psi K^+$ events, this would introduce a correlation between the calculation of the relative efficien-
 397 cies and the $B^+ \rightarrow J/\psi K^+$ signal yield measured on data. In order to avoid possible correlations between
 398 data and MC after the reweighting procedure, only odd (even) numbered MC and data events are used in
 399 the DDW calculation, while the remaining data and MC events are used for the B^+ yield fit and efficiency
 400 and acceptance calculation, respectively.
 401 The DDW consistency is checked by applying the weights calculated using odd (even) numbered events
 402 to the even (odd) numbered MC events and comparing the resulting distributions with the even (odd)
 403 numbered data events (see Appendix B).
 404 The QLC corrections are based ad-hoc generator-level only MC samples, generated either with very loose
 405 or tighter generator-level cuts. The tighter set of cuts reproduces what used to generate the corresponding
 406 full-simulation samples used in the analysis to determine relative efficiencies. The looser set of cuts is
 407 chosen as not to introduce a significant bias in the B meson yield within the kinematical acceptance of the
 408 analysis selections. From the ratio of these two samples one can therefore measure the bias introduced
 409 by generator-level selections and extract the corresponding corrections. Since the QLC are calculated
 410 using these dedicated MC, there is no need for special arrangements in the repartition of simulated events
 411 between the QLC derivation and the analysis. The consistency of the QLC weights is checked in a similar
 412 way as the DDW, by calculating the corrections using only half of the samples and using the other halfs
 413 to check the effects of the weighting procedure.
 414 The QLC and DDW calculation involves several ingredients affected by uncertainty (data, MCs, side-
 415 band weights). The weights have been extracted taking care of the error propagation, and including the
 416 correlation between the quantities involved. The final weights are finally available for the rest of the
 417 analysis and propagated with their full covariance matrix. For instance, section 13 the correlation between
 418 these weights is and taken into account in the calculation of signal efficiency and acceptance.
 419

420

421 Despite the difference in principle, We expect the B^\pm DDW to be usable also to correct the B_s^0 MCs,
 422 as was done in the previous rounds of the analysis [4]. This hypothesis is verified by calculating the DDW
 423 for the $B_s^0 \rightarrow J/\psi\phi$ process and comparing them with the B^\pm DDW. The weights will be shown in sec-
 424 tion ?? to be compatible within stastistical uncertainty: the fragmentation functions used for the B^\pm and B_s^0
 425 have differences between MC and data which are similar for the $B^\pm \rightarrow J/\psi K^\pm$ and $B_s^0 \rightarrow \mu^+ \mu^-$ signals,
 426 and the DDW corrections are probably dominated by data-MC differences in the b quark kinematics before
 427 hadronisation.

428 The DDW consistency between $B_s^0 \rightarrow J/\psi\phi$ and $B_s^0 \rightarrow \mu^+ \mu^-$ can be considered also as a consistency
 429 check on the QLC: since the $B^\pm \rightarrow J/\psi K^\pm$ and $B_s^0 \rightarrow J/\psi\phi$ MCs have been generated with different
 430 kinematic biases, any mistake in the QLC on either sample would reflect in an inconsistency between the
 431 two sets of DDW.

432 5.2 Quark Level Corrections

433 In order to enhance the MC production efficiency the default MCs are generated with relatively tight
 434 parameters of the quark-level process and with a selection applied to the particles in the final state. Quark
 435 Level Corrections (QLC) are intended to correct and evaluate the systematics uncertainty due to the
 436 quark-level cuts, while effects due to the cuts on the final state particles are included in the efficiency and
 437 acceptance terms of equation 1.

438 The QLC are evaluated using two different MC samples, generated with looser quark-level cuts with
 439 respect to the default MC and no cuts on the final state particles (*unbiased* MC), and with the same quark-
 440 level cuts as the default MCs and without final state particle cuts (*quark biased* MC). The quark-level cuts
 441 chosen for the *unbiased* MC are the same as the ones used to generate this sample for the previous version
 442 of the analysis. In order to check that the unbiased quark-level cuts do not introduce any bias, a sample
 443 with even looser cuts has been generated and the kinematic distributions of the B meson from the two
 444 samples have been compared after applying the analysis fiducial volume cuts on the B meson properties
 445 ($p_T > 8$ GeV, $|\eta| < 2.5$) and the final state particle cuts. This test showed that the Run1 unbiased quark-level
 446 cuts do not introduce any residual bias.

447 Since QLC are meant to correct generator-level biases, the production of the unbiased and quark biased
 448 MCs is limited to generation, without simulation of the detector response and reconstruction; table 13
 449 shows the sets of cuts applied at generation level for the different MCs.

The computation of the QLC is performed using the unbiased and the quark biased samples according to

	$\hat{p}_T min$	anti-b η	anti-b p_T	muons η	muons p_T	final h η	final h p_T
default $B^+ \rightarrow J/\psi K^+$	7 GeV	2.6	7 GeV	2.6	3.5 GeV	2.6	0.9 GeV
unbiased $B^+ \rightarrow J/\psi K^+$	5 GeV	4	2.5 GeV				
quark biased $B^+ \rightarrow J/\psi K^+$	7 GeV	2.6	7 GeV				
default $B_s \rightarrow \mu^+ \mu^-$	5 GeV	2.6	5 GeV	2.6	3.5 GeV		
unbiased $B_s \rightarrow \mu^+ \mu^-$	5 GeV	4	2.5 GeV				
quark biased $B_s \rightarrow \mu^+ \mu^-$	5 GeV	2.6	5 GeV				
default $B_s \rightarrow J/\psi\phi$	11 GeV	2.5	9 GeV	2.6	3.5 GeV	2.6	0.8 GeV
unbiased $B_s \rightarrow J/\psi\phi$	5 GeV	4	2.5 GeV				
quark biased $B_s \rightarrow J/\psi\phi$	11 GeV	2.5	9 GeV				

Table 13: Quark-level and final state particle cuts per MC sample. \hat{p}_T .

450 the following formula:

$$W_{QL} = v_{quarkBiased}^{FScuts} \cdot \left(\frac{\sigma_{quarkBiased}^{Pythia}}{N_{quarkBiased}^{tot}} \right) / \left[v_{unbiased}^{FScuts} \cdot \left(\frac{\sigma_{unbiased}^{Pythia}}{N_{unbiased}^{tot}} \right) \right] \quad (2)$$

451 Where v is the number of entries in a (pT_B, η_B) from the unbiased or quark biased samples after applying
 452 the final state particle cuts. σ^{Pythia} and N^{tot} are, respectively, the Pythia-calculated generation cross-
 453 section and the number of generated events for the given samples. The $\frac{\sigma^{Pythia}}{N^{tot}}$ terms in equation 2
 454 account for the relative 'integrated luminosity' of the two MC samples considered. Effectively the QLC
 455 are correcting not only the $pT_B - \eta_B$ shape distribution of the events, but also their normalization in order
 456 to account for the specific generator-level cuts chosen.

457 The inverse of these weights should be used to weight events individually, thus correcting with event-
 458 weights the QL cut biases.

459 The computation of the QLC correction is limited to the B fiducial volume of the analysis ($pT_B > 8$
 460 GeV and $|\eta_B| < 2.5$); the binning scheme chosen is not regular, but the bin size is chosen to maintain a
 461 comparable statistical uncertainty in every bin. Figure 6 shows the QLC and their uncertainties for the
 462 three processes $B^+ \rightarrow J/\psi K^+$, $B_s \rightarrow J/\psi \phi$ and $B_s \rightarrow \mu\mu$.

463 5.3 Data Driven Weights

464 Due to the limited statistics of the B^+ data sample, the DDW corrections are not computed in a two-
 465 dimensional grid of the variables p_T and $|\eta|$, but are extracted with the one-dimensional times one-
 466 dimensional iterative method, by comparing sideband-subtracted $B^\pm \rightarrow J/\psi K^\pm$ odd(even)-numbered
 467 events with odd(even)-numbered MC events, after the latter are corrected with the QLC weights. This
 468 approach was already used in the analyses [6] based on the 4.7 fb^{-1} sample of data collected in 2011
 469 and [4] based on the full Run1 sample.

470 The validity of the approach is tested by the stability of the recursive procedure, and fundamentally it
 471 works because the production cross-section depends strongly on p_T , but much less on $|\eta|$, with small
 472 correlation between the two variables.

473 The event selection followed is the same as the selection explained in the following sections, with the
 474 addition of an additional cut $L_{xy} > 0.3 \text{ mm}$ on the transverse separation between primary and secondary
 475 vertex to remove most of the combinatorial background, and without applying the multivariate selection
 476 against combinatorial background.

477 The extraction of the SideBand Weights (SBW) for the sideband subtraction follows similar lines used in
 478 [6] and [4], with the difference that the B^+ yield is extracted in combined p_T and $|\eta|$ bins, after a small
 479 dependence of the SBW on $|\eta|$ was found. Two $|\eta|$ bins are chosen, in order to separate the two regions
 480 where the $|\eta|$ SBW behave differently. For a more detailed discussion see B.1.

481 We derive two sets of DDW, in p_T and $|\eta|$ of the B meson, determined by the ratio of the normalised p_T
 482 and $|\eta|$ spectra in data and MC. The final weights that are going to be used are therefore

$$W_{DD}(p_T, |\eta|) = W_{p_T} \cdot W_{\eta} \quad (3)$$

483 Figures 41(a) and 41(b) show the result weights for B^+ . A more detailed discussion on the DDW derivation
 484 can be found in B.2.

487

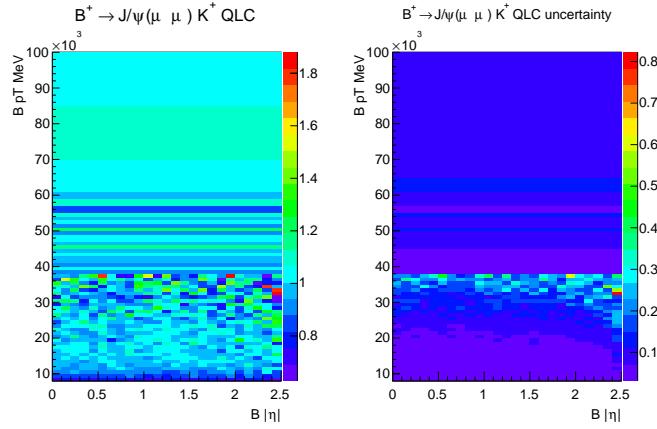
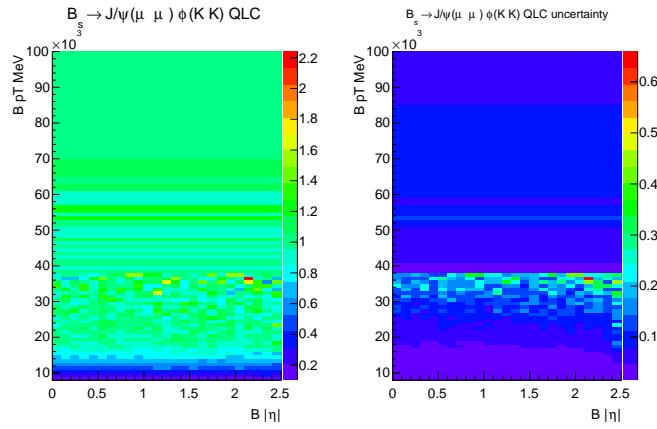
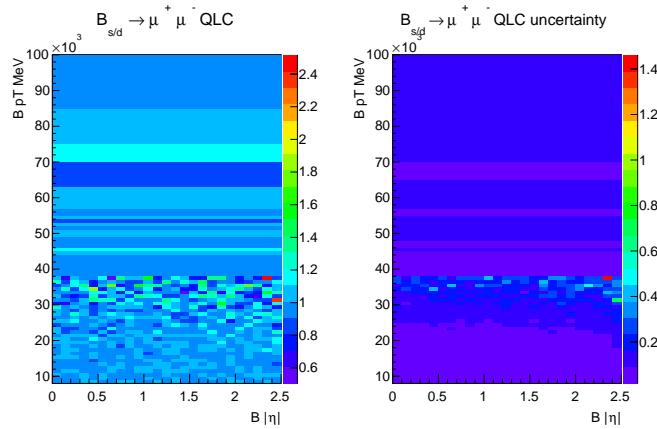
Events from B_s^0 decays into $J/\psi \phi \rightarrow \mu^+ \mu^- K^+ K^-$ can be studied and compared to B^+ decays to $J/\psi K^+$. In particular, DDW can be extracted in the $J/\psi \phi$ channel and compared the standard ones from $J/\psi K^+$. This is shown in Figure 9. The two sets of weights agrees well within the the statistical error, confirming that the differences in the MC generation for B^+ and B_s^0 are small and providing a consistency check on our procedures. Furthermore, the comparison supports the assumption of using DDW derived from B^+ also for the B_s^0 channel.

A further check on the DDW can be performed by comparing the B^+ DDW with the weights calculated and used in the Run 1 analysis. Figures 9(a) and 9(b) show the comparison of the p_T and η DDW calculated with Run2 data and MC (red), calculated using Run1 data and MC (blue) and calculated using Run1 data and MC after the application of the trigger weights (green).

There is a good agreement between the p_T DDW for Run2 and Run1 DDW without applying the trigger weights, while the Run1 and Run2 η DDW show a different behaviour.

The error associated to the DDW takes into account the uncertainty on all the quantities involved in the calculation and their correlation, it is therefore possible to calculate the covariance matrix of the DDW. Figure 10 shows the covariance matrix of the DDW weights, where the diagonal has been set to 0 to make the plot more readable. The correlation of some of the DDW results to be quite high, it is therefore taken into account in the systematic uncertainty evaluation. Studies have also been performed to asses the correlation of the QLC and DDW weights. This was proven to be low and therefore can be omitted in systematic uncertainty calculation.

6: Trigger
weights are
not applied
yet
to the Run2
investiga-
tion
DDW

(a) $B^+ \rightarrow J/\psi K^+$ QLC (left) and the corresponding uncertainty (right).(b) $B_s \rightarrow J/\psi\phi$ QLC (left) and the corresponding uncertainty (right).(c) $B_s \rightarrow \mu\mu$ QLC (left) and the corresponding uncertainty (right).Figure 6: The three sets of plots show the QLC (left plot) and the relative uncertainty (right plot), calculated for the three processes, respectively, $B^+ \rightarrow J/\psi K^+$ figure 6(a), $B_s \rightarrow J/\psi\phi$ figure 6(b) and $B_s \rightarrow \mu\mu$ figure 6(c).

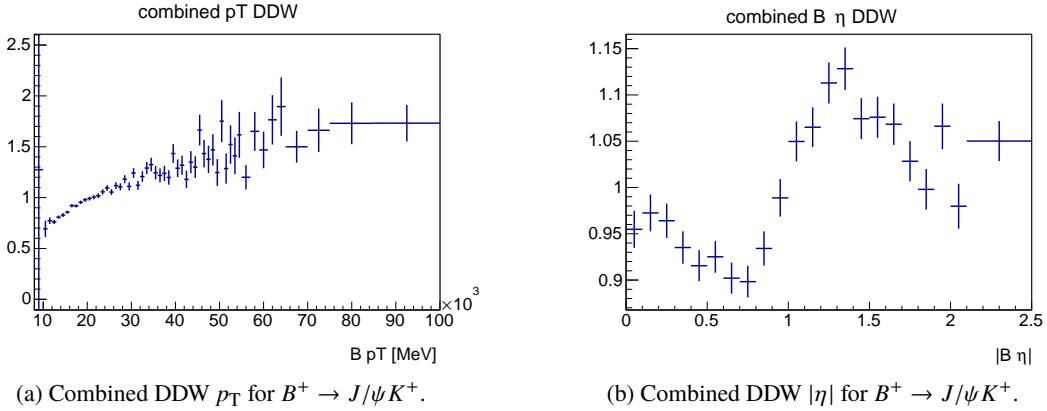


Figure 7: Result of the calculation of the DDW for the $B^+ \rightarrow J/\psi K^+$ process. Figures 7(a), 7(b) show the combination of the first two iterations of the DDW calculation.

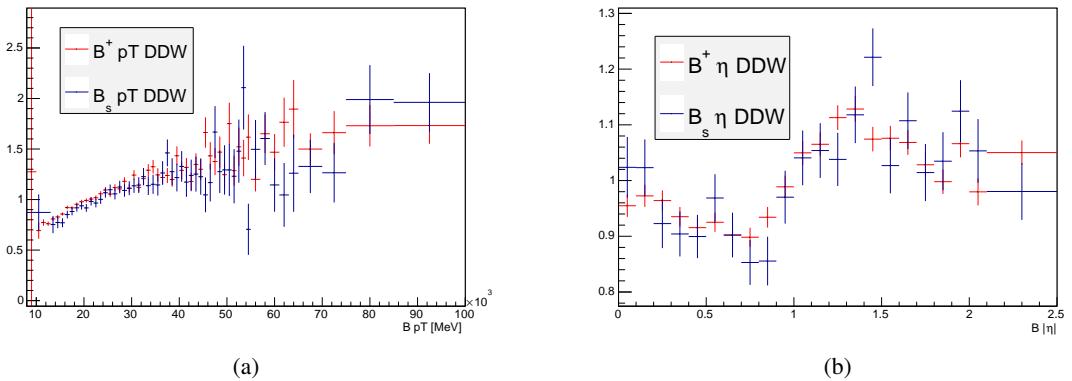


Figure 8: Comparison of p_T DDW, figure 8(a), and η DDW, figure 8(b), for $B^+ \rightarrow J/\psi K^+$ (red points) and $B_s^0 \rightarrow J/\psi \phi$ (blue points).

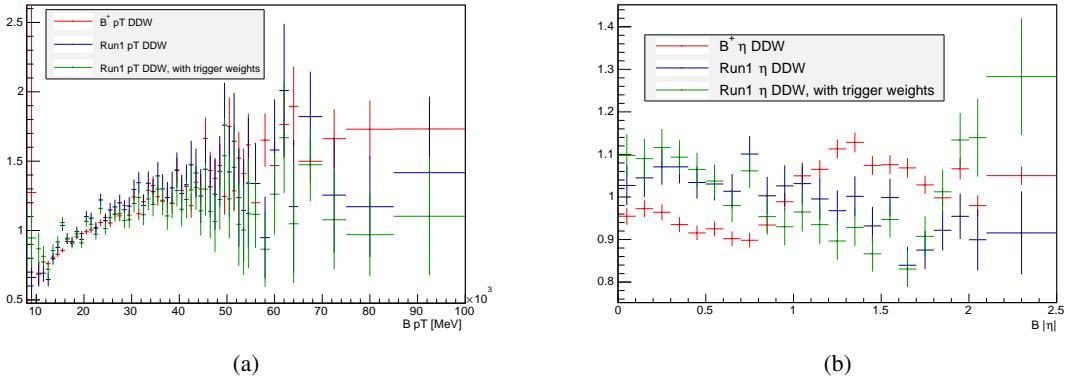


Figure 9: Comparison of p_T DDW, figure 8(a), and η DDW, figure 8(b), calculated using Run2 data and MC (red points), using Run1 data and MC (blue points) and using Run1 data and MC applying also the trigger weights (green). There is a good agreement between the Run and Run2 p_T DDW when trigger weights are not applied, while the η DDW show a different behaviour. This is currently under investigation.

DDW correlation matrix

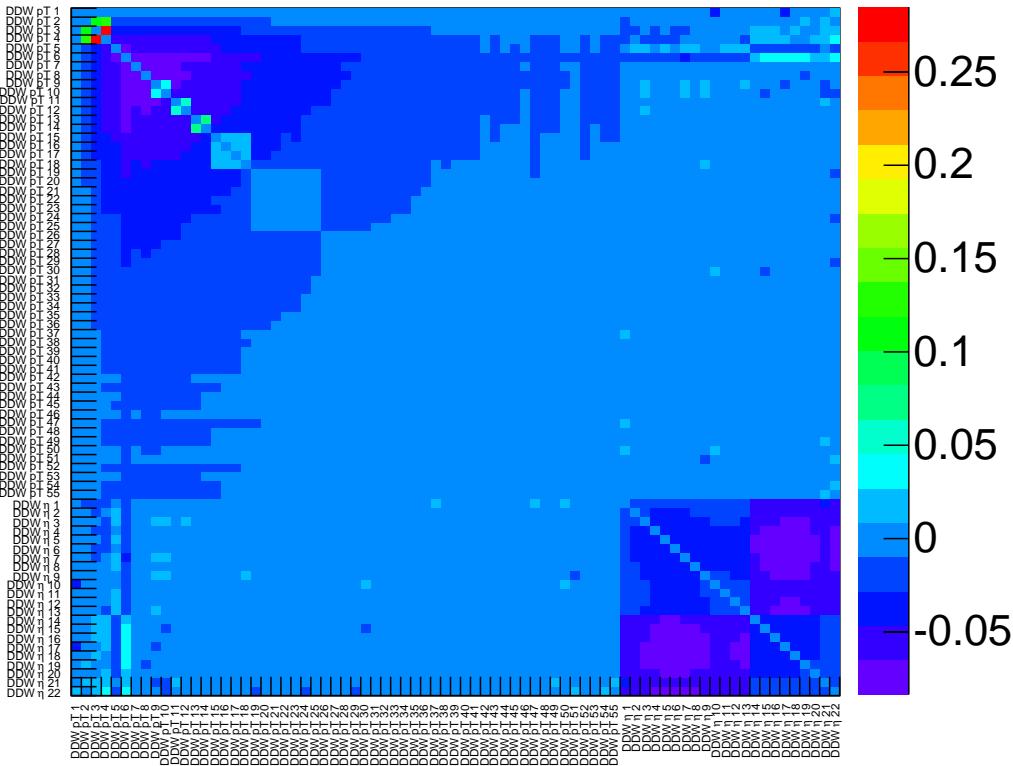


Figure 10: Covariance matrix of $B^+ \rightarrow J/\psi K^+$ DDW. The diagonal has been set to 0 to improve plot readability.

507 6 Trigger

508 In 2015-16 the main low p_T di-muon triggers were:

- 509 • 2mu4 prescaled (2015) or not active (2016)
- 510 • mu6_2mu4 prescaled during 2016, $\sim 82\%$ efficient w.r.t. 2mu4
- 511 • 2mu6 unprescaled, $\sim 35\%$ efficient w.r.t. 2mu4

The collected luminosity of each trigger per year are shown in table 14.

	2015	2016
2mu4	3.17 fb^{-1}	0 fb^{-1}
mu6mu4	3.93 fb^{-1}	37.13 fb^{-1}
2mu6	3.93 fb^{-1}	26.03 fb^{-1}

Table 14: Collected luminosity in 2015 and 2016.

512
513 We can compare the available collected luminosity provided by the different triggers by calculating their
514 effective collected luminosity, the total luminosity we would need to have the same amount of statistics
515 using 2mu4 as main trigger.

- 516 • consider mu6mu4 only would get an effective collected luminosity of 24.6 fb^{-1}
- 517 • consider the combination of mu6mu4 and 2mu6 would get an effective collected luminosity of 28.5 fb^{-1}
- 519 • 2mu4 addition would increase the 2mu6+mu6mu4 statistics of a factor $\sim 1\%$

520 The prescaled trigger mu6mu4 provides the bulk of the statistics, which is $\sim 85\%$ of total. The addition of
521 2mu6 would allow to use basically the full dataset but would also increase the complexity of the analysis,
522 given the tight schedule, this round of the analysis will use mu6mu4 only.

523 Diffrent version of the mu6mu4 triggers are available at the HLT stage in the 2015-16 data taking period,
524 while the L1 seed reamins the same: L1_MU6_2MU4. The available HLT triggers for the $B^0 \rightarrow \mu\mu$
526 process are:

- 527 • HLT_mu6_mu4_bBmumu: basic dimuon trigger for muons pT thresolds of 6 and 4 GeV;
- 528 • HLT_mu6_mu4_bBmumu_Lxy0: similar to the previous trigger, in order ot reduce the rate a Lxy
529 > 0 cut was added;

530 The first item was active in 2015 and heavily prescaled in 2016, in favour of the second item.
531 The triggers used to collect the data for the $B^+ \rightarrow J/\psi K^+$ and $B_s \rightarrow J/\psi \phi$ processes are based on the
532 same L1 item as the ' $B \rightarrow \mu\mu$ trigger, L1_MU6_2MU4, while the HLT items change:

- 533 • HLT_mu6_mu4_bJpsimumu: basic dimuon trigger for muons pT thresolds of 6 and 4 GeV;
- 534 • HLT_mu6_mu4_bJpsimumu_Lxy0: similar to the previous trigger, in order ot reduce the rate a
535 Lxy > 0 cut was added;

- 536 • HLT_mu6_mu4_bJpsimumu_Lxy0_delayed: same as the previous trigger, but for the delayed
537 stream.

538 The first item is used in 2015, while it was heavily prescaled in 2016 in favour of the second item. The
539 third item started being used when the delayed stream was activated.

540 In order compare the available statistics for this analysis with respect to the previous version, three
541 main ingredients are needed.

- 543 1. **B production cross section with respect to Run-1**, due to the Run-2 increased center of mass
544 energy a value ~ 1.7 times higher is expected, according to studies performed using FONLL [8].
- 545 2. **The efficiency of the dimuon triggers available in Run-2 with respect to the Run-1 triggers**,
546 already shown in this section.
- 547 3. **The collected luminosity for the selected trigger in Run-2**, an integrated luminosity of $\approx 40\text{fb}^{-1}$
548 has been collected during the 2015 and 2016 data-taking period.

549 The same signal over background ratio as the analysis performed on the Run-1 dataset is assumed for this
550 study.

551 With the three ingredients listed above and the previous assumptions we estimate the expected statistics
552 available for the 2015-16 dataset by scaling the Run-1 analysis statistics. This yields an estimated 2-fold
553 increase with respect to Run-1 statistics.

554 **7 CutFlow**

555 8 Studies on muon fake rates

556 One of the most problematic backgrounds in this study is represented by the charmless two-body B decays
 557 referred to as $B \rightarrow hh'$, h being a charged K or π . This background is topologically identical to, and peaks
 558 under, the signal. The only handle we can exploit is the muon identification capability of the ATLAS
 559 detector. For these decays to feed into our events, the charged K or π has to be misidentified as a muon.
 560 In the previous analysis [4] a dedicated analysis was carried out in order to reduce the fake rate thus
 561 reducing the contamination from these background events. The misidentification fraction was found to be
 562 respectively 0.00076 and 0.00101 for negative and positive K , and 0.00044 and 0.00042 for negative and
 563 positive π . The average misidentification fraction was 0.00067 ± 0.00001 , for a muon efficiency of 95%.
 564 Some of the variables used in the BDT of the previous analysis are now used in the muon quality definition,
 565 therefore we decide to check the misidentification fraction for the different muon qualities.

566

567 The study of fake muons has been performed on two MC samples of signal ($B_s^0 \rightarrow \mu^+ \mu^-$) and charmless
 568 two-body decays ($B \rightarrow hh'$). These samples have been produced with full GEANT simulation in order
 569 to accurately describe the hadrons after they leave the Inner Detector. The preselection described in
 570 Section ?? is applied to the events entering this study (except for the muon quality requirement). A di-
 571 muon trigger request (mu6mu4) is applied to the $B_s^0 \rightarrow \mu^+ \mu^-$ sample, while events containing preselected
 572 hadrons from $B \rightarrow hh'$ that are misidentified as muons are required to satisfy a single muon trigger (mu4).
 573 Once an event has passed the preselection, the two final state particles are considered separately.

574

Table 15 show the misidentification fraction and the muon efficiency for the different muon qualities
 available and for the previous analysis. Tight muons have a comparable misidentification fraction and

	run1 presel + trigger match	run1 fake-BDT	loose muons	medium muons	tight muons
total	0.00181	0.00067	0.00221	0.00221	0.00109
total +			0.00229	0.00228	0.00114
total -			0.00214	0.00213	0.00105
total π		0.0004	0.0018	0.0018	0.00101
total K		0.0009	0.00264	0.00263	0.00118
π^+	0.00121	0.00042	0.00177	0.00177	0.00101
π^-	0.00116	0.00044	0.00183	0.00182	0.00101
K^+	0.00263	0.00101	0.00281	0.00281	0.00127
K^-	0.00207	0.00076	0.00246	0.00245	0.00109
μ eff		0.95	0.997	0.996	0.935
μ^+ eff			0.997	0.996	0.935
μ^- eff			0.997	0.997	0.935

Table 15: misidentification fraction and muon efficiency for the three available muon qualities (loose, medium and tight) and, for comparison, from the previous analysis [4], for different emdmixtures of positive and negative π and K . Misidentification fraction is calculated as the ratio of the number of hadrons after all selections and the quality requirements divided by the total number of hadrons that pass all selections. Muons efficiency is defined as the number of muons that pass all selections and the quality requirements divided by the total number of muons that pass all selections.

575

576 muon efficiency to the Run1 BDT; their usage would imply an increase of about $\sim \times 1.6$ in the number of

577 fakes than using a Run1-like BDT, for about the same muon efficiency.
 578 In order to understand the effect of the tight muons usage, we compared the effect of a Run1-like BDT
 579 and the tight muons in the analysis.
 580 Knowing the available statistics before and after the fake-BDT application in previous analysis [4] and
 581 the muon efficiency and fake-rate, we computed the Run1 background composition in terms of fake and
 582 real muons; applying the properties of tight muons we can estimate the effect tight muons would have
 583 had on the Run1 analysis. Tables 16 and 17 show the amount of statistics for the different components
 584 of the likelihood used in the fit for the signal yield extraction and the estimation of the statistics using
 585 tight muons. The combinatorial background doesn't increase significantly, the SS-SV (and semileptonic)
 586 background increases of a factor $\sim \times 1.1$ and the peaking background increases of a factor $\sim \times 2.6$. Signal
 587 reduction due to lower muon efficiency is almost negligible.

In order to asses the impact of tight muons usage on the analysis performance a set of toy-MC based on the

	comb bkg run1	comb bkg tight mu	SS-SV bkg run 1	SS-SV bkg tight mu
bin 1	1455.3	1460.7	205.5	229.0
bin 2	110.5	110.9	105.6	117.7
bin 3	11.6	11.6	51.2	57.1

Table 16: breakdown of the background contributions in the Run1 likelihood after the BDT application. Columns labelled tight mu show the estimated statisticcs we would have gotten using tight muons instead of the BDT.

	run 1	run 1 with tight mu
nBs (SM expected)	41	39.7
nBd (SM expected)	5	4.8
peaking bkg	1	2.6

Table 17: Breakdown of the expected signal and peaking background contributions in the Run1 likelaihood after the BDT application. Columns labelled tight mu show the estimated statisticcs we would have gotten using tight muons instead of the BDT.

588 Run1 analysis likelihood with the available statistics modified according to the estimation of the available
 589 stastics in this analysis performed in section 6 has been run, varying the number of events associated to
 590 the different background and signal models according to the usage of tight muons or a Run1-like BDT.
 591 Table 18 shows the root mean square of the distribution of the fitted number of B_s and B_d events. No
 592 significant effect on analysis sensitivity is visible for B_s , while a $\sim 1\%$ broader RMS is visible on B_d
 593 count.
 594 Given the negligible difference due to the usage of tight muons instead of a Run1-like BDT for fake muons
 595 reduction, we decide to use tight muons instead of developing and ad-hoc BDT for this analysis.

	Run1 fake-BDT	Run1 tight muons	Run2 fake-BDT	Run2 tight muons
RMS nBs	15.20 ± 0.01	15.16 ± 0.01	21.260 ± 0.015	21.260 ± 0.015
RMS nBd	13.88 ± 0.01	13.96 ± 0.01	18.850 ± 0.015	19.130 ± 0.015

Table 18: RMS of distributions of number of fitted B_s and B_d events in toy-MC study. The columns labelled Run1 refer to the statistics available in the previous analysis, modified for the usage of tight muons in case of the second column. Columns labelled Run2 refer to the estimation of the statistics available for this analysis performed in section 6.

597 9 Background Modeling

598 9.1 Non-resonant Background

599 The measurement of the rare decay of B mesons into muon pairs requires the rejection of a very large
 600 combinatorial background. After the preliminary and additional selection, the amount of combinatorial
 601 background is at the level of ≈ 140 k events per 100 MeV interval in the muon pair invariant mass (this
 602 is shown in the context of the normalisation study for the signal fit in Figure ??). This figure should be
 603 compared to ≈ 20 events/100 MeV expected at the peak of the B_s^0 signal (section 14.3).

604 To have a statistically meaningful sample of this combinatorial background, a very large number of events
 605 needs to be generated (one of the largest MC production performed so far by ATLAS for a single analysis),
 606 thus specific procedures have been used in order to reduce the event generation time, as discussed below.
 607 The MC sample was generated inclusively to provide a realistic composition of muons from different
 608 sources, including non combinatorial sources of muon pairs. On the other hand, because of uncertainties
 609 in cross sections and branching ratios, as well as for those related to the generation procedures, no attempt
 610 is made to use the MC to draw quantitative conclusions on the amount in which each type of background is
 611 present in real data. For that purpose, events collected in the sidebands are used. As shown in section 14.3,
 612 after the background reduction developed on MC is applied, the remaining background in real data is
 613 sufficiently low and separated in its different components to be used for an effective interpolation in the
 614 signal region.

615 In order to optimise the production speed, *repeated hadronisation* and *repeated decay* procedures were
 616 used for the inclusive MC, in order to reduce the typical generation time from a several minutes per event
 617 to only about 40 seconds.

618 Repeated hadronisation, performed in PYTHIA, was applied with 10 repetitions. This is technically
 619 implemented by means of PythiaB package [9] and had already been widely tested and used for various
 620 MC production in B physics. This procedure scarcely affects the characteristics of the generated samples.

621 The next step, decays of the unstable particles, has been done differently with respect to Run 1, since
 622 the procedure employed back then effectively enhanced the acceptance at lower p_T , which required a
 623 correction. Currently, each hadronised event is cloned fixed number of times (200) by PYTHIA and each
 624 cloned event decays independently. In the next step a generator level filtering (e.g. cuts on muons
 625 transverse momenta) is applied, and all cloned events which pass it are stored in the resulting sample. This
 626 approach allows events with identical b hadron kinematics if more than one of those 200 clones produce
 627 a muon pair, which passes the filter. However, this approach does not introduce the kinematic bias we
 628 had in Run 1. With PYTHIA(unlike EvtGen used in Run 1) we have less up-to-date decay tables and no
 629 angular decay models. However the last effect we found to be negligible for the decays we require - with
 630 two muons in the final state. The residual discrepancies in kinematic of B-candidates and multiplicity of
 631 tracks in the vicinity of reconstructed di-muon vertices are corrected on data.

632 If only standard cuts are applied to the muons and loose cuts are applied to the muon pair, the MC
 633 sample is dominated by opposite side combinatorial background (about 20 M events in the mass interval
 634 4.8–5.9 GeV, with 50 k events of different nature). When a multivariate selection aiming at the rejection of
 635 combinatorial background is applied (as discussed in Sec 10), the relevance of the additional contribution
 636 becomes more evident (e.g., for a selection aiming at 60 % (20 %) efficiency for the signal $B_s^0 \rightarrow \mu^+ \mu^-$,
 637 opposite side events account for about 95 % (40 %) of the muon pairs in the mass interval 4.8–5.9 GeV).
 638 The additional events are mainly due to same-side combinatorial background from B meson decays, and

8: numbers
9b4b.d.

10:
numbers
t.b.u.

11:
numbers
t.b.u.

also to exclusive decays such as $B_c^+ \rightarrow J/\psi \mu^+ \nu$, $B \rightarrow K \mu^+ \mu^- (\gamma)$, $B_s^0 \rightarrow \mu^+ \mu^- (\gamma)$ (XXX signal events are present in the background MC sample).

Because of the uncertainties discussed above, information extracted from the relative normalisation of the different types of backgrounds and of the signal events present in the inclusive sample are not used in any part of the analysis. On the other hand, the samples of data obtained from MC have shown remarkable consistency with real data collected in the side-bands of the signal region, in their dependence on the di-muon invariant mass and on the BDT classifier. It is also remarkable that the simulation shows that exclusive semileptonic decays $B^0 \rightarrow \pi \mu \nu$, $B_s^0 \rightarrow K \mu \nu$, with the hadron being misidentified as muon, provide a background of negligible size compared to same-side cascade events, or B_c background.

9.2 Background Classification

The three processes of interest are $B_s^0 \rightarrow \mu^+ \mu^-$, $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi \phi$. In order to estimate the different background contributions to each of them, the MC in table 19 is used.

Process	DSID	Events	Generator
$b\bar{b} \rightarrow \mu^+ \mu^- X$	300307	600M	PYTHIA + PHOTOS
$b\bar{b} \rightarrow J/\psi X$	300203	10M	PYTHIA

Table 19: Inclusive background samples used to find background contributions.

The same reconstruction algorithms that would be applied to the three exclusive MC samples are applied to the inclusive samples. From this, containers with candidates misidentified as B_s^0 and B^+ are created. The reconstructed objects, both muons and tracks can be matched to the corresponding truth objects. In this way, one can create a string representing the decay process responsible for the reconstructed B candidate. Figure 11(a) shows this, in this case, given that the upper most truth B candidate does not exist (there are two, each associated to a b quark), this reconstructed B_s^0 would be combinatorial background.

On the other hand, figure 11(b) shows a reconstructed B^+ candidate originating from a partially reconstructed decay. In this case the decay string would be $B0[K^*0[pi:-K+]Jpsi[mu+:mu-]]$.

659

9.2.1 Truth Matching

Figures 11 show, as a dashed black line, the truth and reconstructed objects matching. In order to match them one requires $\Delta R < 0.05$. Preliminary studies have shown that the optimal value for this cone is 0.01; but a larger radius should prevent us from loosing any truth candidate. In order to prevent truth objects that are not associated with the reconstructed object, from entering the cone, we require:

$$\left| 1 - \frac{p_T^{truth}}{p_T^{reco}} \right| < 0.15 \quad (4)$$

Also, reconstructed muons are required to be matched to muons with the same charge. The kaons are required to be matched with truth light hadrons, either mesons or baryons.

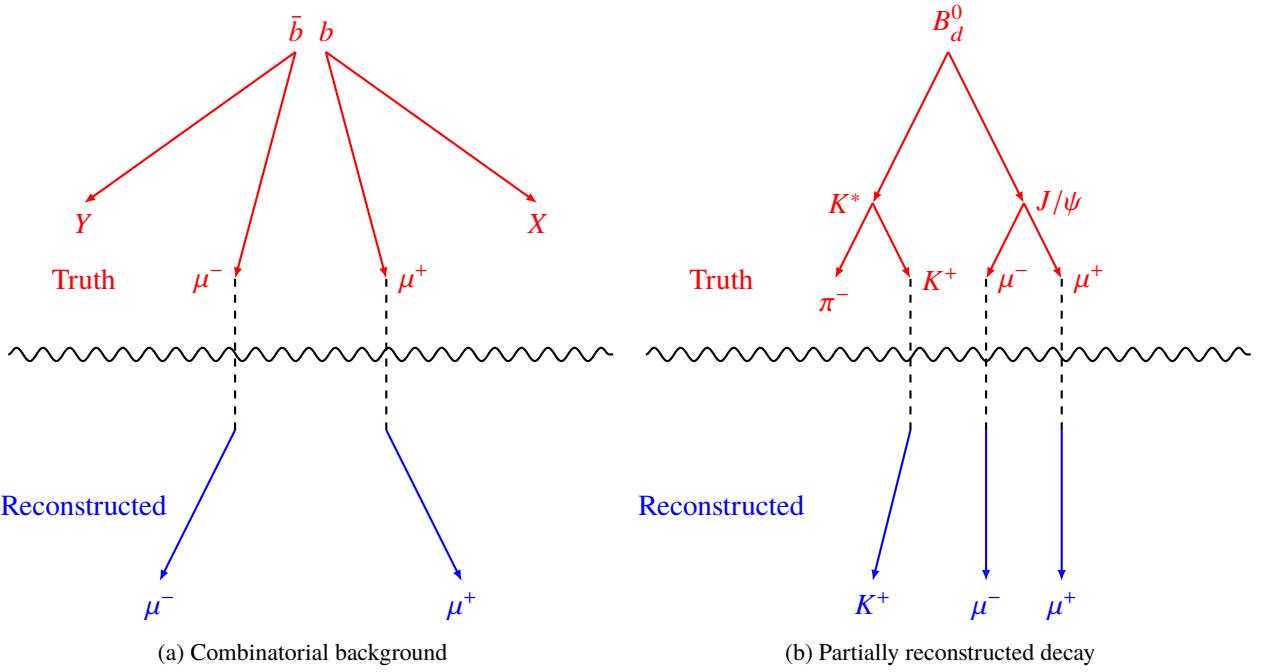


Figure 11: Both diagrams represent a reconstructed B_s^0 or B^+ , which in reality are combinatorial background or a partially reconstructed decay.

9.2.2 Cumulative Decay Probability Distribution

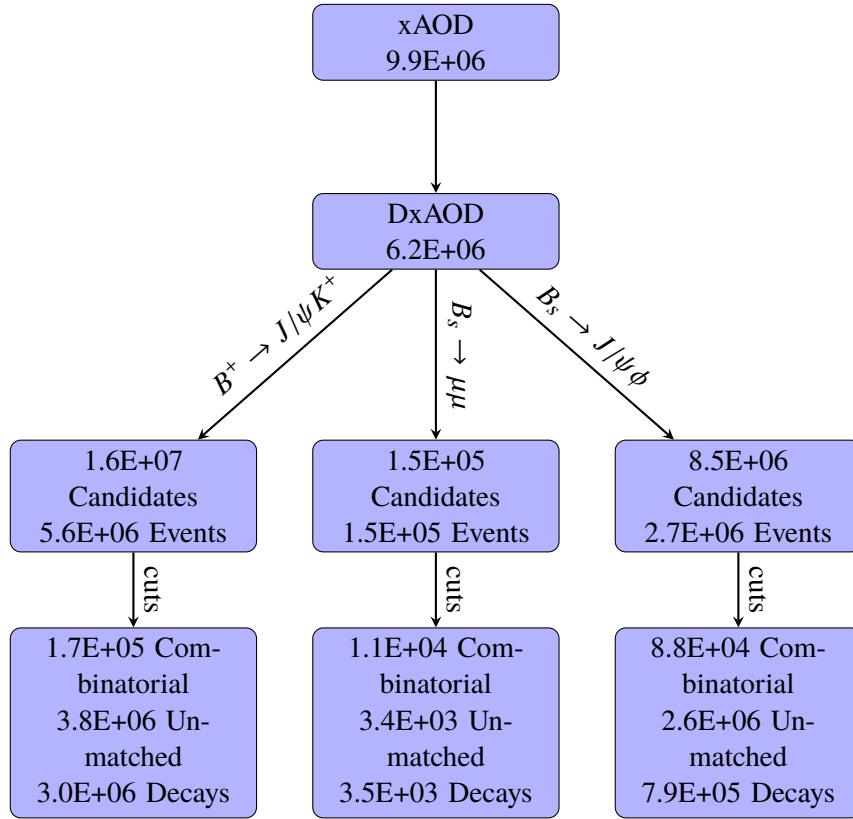
A natural question now would be, how many different events in the inclusive background survive the candidate selection and get reconstructed? Figure 12 shows the number of events and candidates at each stage of the analysis, from the xAOD level until the ntuples.

The cuts used in the last level are:

- **Muons:** Combined muons that pass MCP requirements and $|\eta| < 2.5, p_T > 4\text{GeV}$.
- **Kaons:** Must pass loose requirements and $|\eta| < 2.5, p_T > 1\text{GeV}$.
- **Candidates:** $|\eta| < 2.5, p_T > 8\text{GeV}, \chi^2 < 6$.

Another thing that we might want to know is the number of decays present in the samples. Figure 13 shows the cumulative distribution of decays vs the decay index, where the most common decays are put first. From this we can conclude that, as expected, the $b\bar{b} \rightarrow \mu^+\mu^-X$ sample contains very few $B_s^0 \rightarrow \mu^+\mu^-$ candidates and mostly $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi \phi$ candidates. Two mass windows were used, the wide window contains much more decays than the narrow one. This is because in the low end of the mass spectrum, partially reconstructed decays, like the one in 11(b), accumulate. The narrow mass window removes all of them. Another thing that the plot tells us is that there are many more background processes associated to the $B_s^0 \rightarrow J/\psi \phi$ candidates than to the $B^+ \rightarrow J/\psi K^+$ candidates.

Table 20 shows the main sources of background for $B^+ \rightarrow J/\psi K^+$. Table 21 shows the same for $B_s^0 \rightarrow J/\psi \phi$. In these tables *unmatched* refers to decays in which the matching could not be made. This happens because one of the hadron tracks, identified as the K^+ , corresponds to a particle that was removed

Figure 12: Cutflow corresponding to $b\bar{b} \rightarrow J/\psi X$ sample.

from the samples during the derivation. In order to slim the truth particles container, particles that do not come from a B decay are removed. In practice this means that unmatched candidates can be classified as combinatorial background. This can be seen in figure 14.

Process	Candidates
unmatched	864837
B+[K+:Jpsi[mu+:mu-]]	192462
B0[K*0[pi-:K+]Jpsi[mu+:mu-]]	117439
combinatorial	41869
B+[K*+[pi0[gamma:gamma]K+]Jpsi[mu+:mu-]]	30063
B+[K+:chi_1c[gamma:Jpsi[mu+:mu-]]]	26241
B+[K+:psi'[pi-:pi+Jpsi[mu+:mu-]]]	9477
B0[K*0[pi-:pi-:gamma:K+:K+]Jpsi[mu+:mu-]]	8545
B0[pi-:K+:Jpsi[mu+:mu-]]	7886
B0[K*_20[pi-:K+]Jpsi[mu+:mu-]]	7435
B+[pi+:Jpsi[mu+:mu-]]	5956
B+[K*+[pi+:K0[K_L0]]Jpsi[mu+:mu-]]	5198
B+[K*+[pi+:K0[K_S0]]Jpsi[mu+:mu-]]	5087
B+[K+:psi'[pi0[gamma:gamma]pi0[gamma:gamma]Jpsi[mu+:mu-]]]	4811
B+[rho+[pi0[gamma:gamma]pi+]Jpsi[mu+:mu-]]	3022

Table 20: Processes most commonly misidentified as $B^+ \rightarrow J/\psi K^+$.

Process	Candidates
unmatched	759877
B0[K*0[pi-:K+]Jpsi[mu+:mu-]]	147001
combinatorial	49979
B0[Jpsi[mu+:mu-]K_10[rho-[pi-:pi0[gamma:gamma]]K+]]	43372
B+[Jpsi[mu+:mu-]K_1+[rho0[pi-:pi+]K+]]	33521
B0[pi-:K+:Jpsi[mu+:mu-]]	17924
B0[K*0[pi-:K+]psi'[pi-:pi+Jpsi[mu+:mu-]]]	16922
B+[Jpsi[mu+:mu-]K_1+[pi+:K*0[pi-:K+]]]	16325
B0[pi-:K+:psi'[pi-:pi+Jpsi[mu+:mu-]]]	14931
B0[K*_20[pi-:K+]Jpsi[mu+:mu-]]	14856
B+[K+:psi'[pi-:pi+Jpsi[mu+:mu-]]]	13086
B+[Jpsi[mu+:mu-]K_1+[omega[pi-:pi0[gamma:gamma]pi+]K+]]	12826
B0[pi-:K+:chi_1c[gamma:Jpsi[mu+:mu-]]]	12027
B_s0[phi[K-:K+]Jpsi[mu+:mu-]]	11752
B0[K*0[pi-:gamma:K+:K+]Jpsi[mu+:mu-]]	11022

Table 21: Processes most commonly misidentified as $B_s^0 \rightarrow J/\psi \phi$.

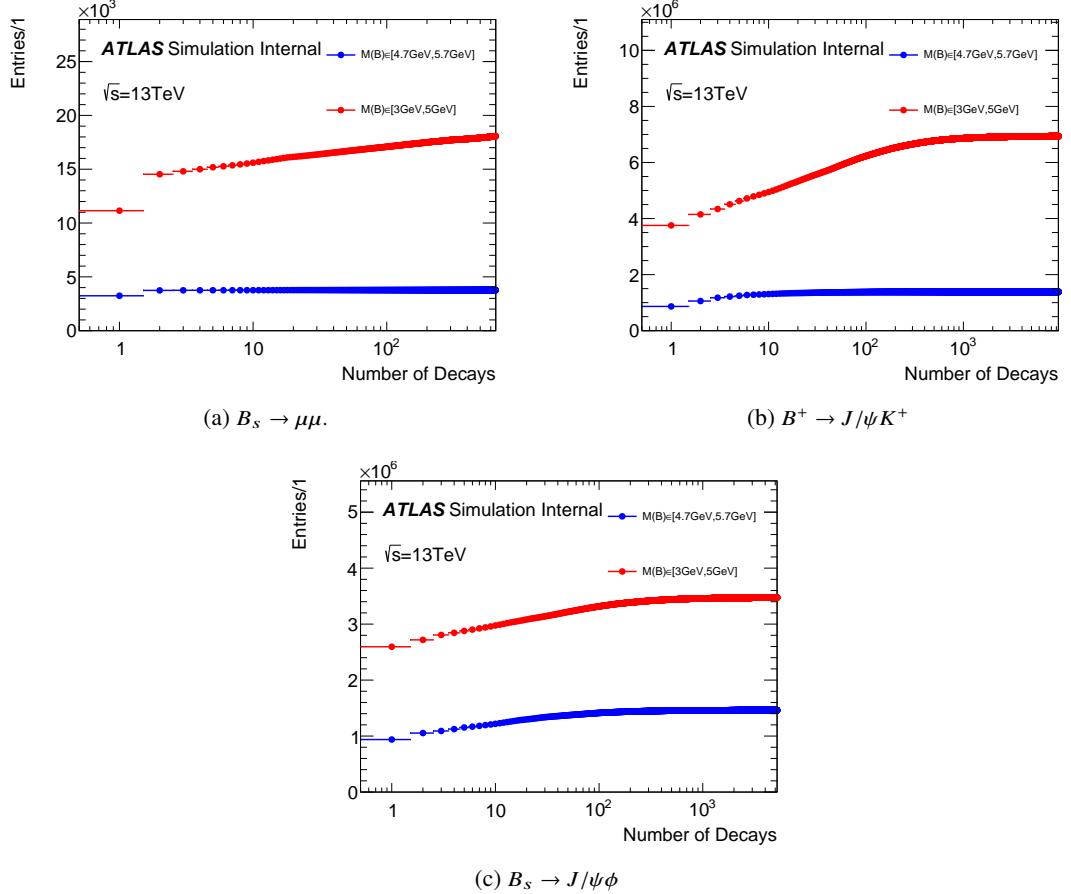


Figure 13: Both figures above show the cumulative number of events per decay. The first bin has all the events for decay number 1. The next bin, those plus decay 2, etc. The decays are ordered in decreasing order of events.

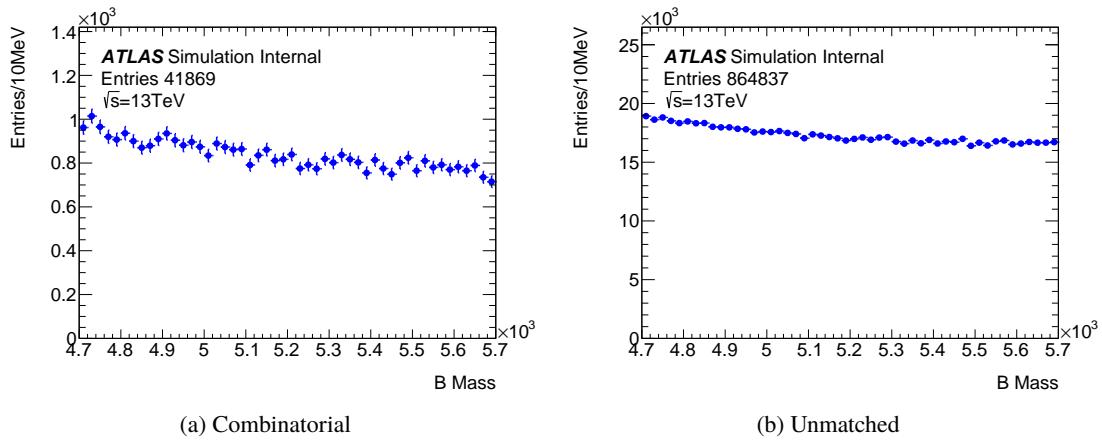


Figure 14: The figures show the mass distributions of B candidates classified as *combinatorial* and *unmatched*. Both distributions are identical in shape.

689 10 ContinuumBDT

690 In order to discriminate against the continuum background we employ a Multi-Variate Analysis (MVA)
 691 strategy, based on a boosted decision tree (BDT) algorithm, as implemented in the TMVA [10] package
 692 of ROOT. We use 15 physical input variables to obtain the signal-to-background discriminator. These
 693 15 input variables are summarised in Table 22 along with a short description of each variable. The
 694 BDT is trained using the B to $\mu\mu$ signal MC and the sideband collision data to obtain the optimal BDT
 695 performance.

696 The following summarises the input variables which were considered for use in the BDT. After discussing
 697 the individual variables, details of training and validation of the classifier are provided.

698 The variables can be subdivided into two groups, one related to isolation properties of the B candidates or
 699 the final state particles, and another describing topological and kinematic properties of the $B_s^0 \rightarrow \mu^+ \mu^-$
 700 decay.

701 To check the isolation of the signal decay we look at non-signal tracks in the vicinity of reconstructed
 702 B vertices, excluding the tracks associated to the pile-up vertices. The procedure of computation of the
 703 variable, outlined in section 4.1 requires information from all tracks in an event and therefore performed
 704 on the derivation level.

705 Since individual muons in the signal decay are fairly well separated, isolation variables of individual
 706 muon candidates have also been considered. However, they were rejected in the final training, since the
 707 B isolation variable $I_{0.7}$, highly correlated to the single-muon isolation variables, contains nearly all the
 708 separation power of this type of variables by itself.

709 The other 12 variables in Table 22 are related to B -decay topology and kinematics. They were chosen
 710 from the larger pool of variables as the ones ensuring the best BDT performance. The variables that have
 711 been considered but dismissed are listed below for future reference: they do not contribute significantly
 712 to the signal/background separation, either by nature or due to the presence of other strongly correlated
 713 but more powerful variables.

- 714 • L_{xy} significance
- 715 • flight length significance in 2D and 3D
- 716 • d_0^{Max}, d_0^{Min}
- 717 • the significance of the closest track DOCA
- 718 • χ^2 between PV and B vertex in z
- 719 • p_T^B significance
- 720 • pointing angle $|\alpha_{3D}|$
- 721 • B vertex fit χ_B^2/NDF
- 722 • coplanarity of $B_s^0 \rightarrow \mu^+ \mu^-$ decays
- 723 • coplanarity of $B_s^0 \rightarrow \mu^+ \mu^-$ decays with Z-axis
- 724 • minimum signed d_0 significance of the signal candidate

725 Lastly in order to arrive at the BDT to use for this analysis, a number of configurations have been studied,
 726 and the best (with respect to background rejection and algorithm behavior) has been selected for use in
 727 the final result.

728 Appendix F summarises the correlation matrices calculated for signal and background events for the 15
 729 discriminating variables used in this analysis.

14:
t.b.c.

Variable	Description
$ \alpha_{2D} $	Absolute value of the angle in the transverse plane between $\Delta\vec{x}$ and \vec{p}_T^B
Isolation $I_{0.7}$	Ratio of $ \vec{p}_T^B $ to the sum of $ \vec{p}_T^B $ and the transverse momenta of all tracks with $p_T > 0.5$ GeV within a cone $\Delta R < 0.7$ from the B direction, excluding B decay products
ΔR	Angle $\sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$ between $\Delta\vec{x}$ and \vec{p}_T^B
p_T^B	B transverse momentum
$\chi_{\mu,xPV}^2$	Minimum χ^2 between the muon candidates and any PV
$\log[\chi_{PV,SV}^2]_{xy}$	separation between production (PV) and decay (SV) vertices, $\Delta\vec{x}^T \cdot (\sigma_{\Delta\vec{x}}^2)^{-1} \cdot \Delta\vec{x}$, in the transverse plane (x, y)
IP_B^{3D}	3-dimensional impact parameter (POCA) of the B candidate
$ d_0 ^{max}$ sig.	Significance of the maximum absolute value of impact parameters of the two muon candidates in the transverse plane of the B decay products relative to the primary vertex
$\Delta\phi(\mu\mu)$	Difference in ϕ between the two muon candidates
$DOCA_{xtrk}$	DOCA of the track closest (“xtrk”) to the B vertex. The tracks associated to pile-up vertices are excluded
$DOCA_{\mu\mu}$	DOCA of the two ID tracks forming the B vertex
N_{trks}^{close}	Number of (“close”) tracks with $\ln(\chi^2) < 1$ where χ^2 is a test of association of a track to the reconstructed B vertex. The tracks associated to pile-up vertices are excluded
$ d_0 ^{min}$ sig.	Significance of the minimum absolute value of impact parameters of the two muon candidates in the transverse plane of the B decay products relative to the primary vertex
L_{xy}	Scalar product in the transverse plane of $(\Delta\vec{x} \cdot \vec{p}_T^B)/ \vec{p}_T^B $
P_L^{min}	Minimum momentum of the two muon candidates along the B direction

Table 22: Description of the 15 discriminating variables used in the discrimination between signal and continuum background.

- 730 The sideband data and the signal MC events, with all selection cuts applied, have been used to train, test,
 731 and evaluate the BDT. It is important to use statistically independent samples for each of these three steps.
 732 In order to utilize the full datasets a rotation strategy has been implemented. Both the sideband data and
 733 the signal MC were split into three independent sets with approximately equal number of candidates and
 734 used to create three BDTs. The process is summarized in table 23.
- 735 After the three BDTs are trained and tested, they are evaluated on a third independent dataset which

	Trained	Tested	Evaluated
<i>BDT 1</i>	Sample 1	Sample 2	Sample 3
<i>BDT 2</i>	Sample 2	Sample 3	Sample 1
<i>BDT 3</i>	Sample 3	Sample 1	Sample 2

Table 23: Training, testing, and evaluation of the 3 BDTs. Each data sample contains 1/3 of the sideband data and 1/3 of the signal MC. For the evaluation sample, 1/3 of the signal MC and only the high-mass sideband of the collision data is used.

contains 1/3 of the signal MC and only the high-mass sideband of the collision data. The evaluation samples are investigated separately and then together, always enforcing the requirement that the BDT is applied to statistically independent data from its training and testing samples.

The importance of each variable in training BDT 1 is given in Table 24 and compared to the separation power of the signal and background. The other two BDTs have similar importance values.

Imp. Rank	Variable	Importance	Sep. Rank	Separation
1	ΔR	0.11780	1	0.59710
2	$ \alpha_{2D} $	0.11570	2	0.57890
3	B Isolation ($I_{0.7}^B$)	0.09715	7	0.36160
4	$\log(\chi_{\mu,xPV}^2)$	0.08314	3	0.44190
5	$\log(\chi_{xy}^2)$	0.07723	9	0.30730
6	$ \Delta\phi_{\mu\mu} $	0.06692	15	0.02661
7	p_T^B	0.06338	13	0.04200
8	N_{trk}^{clos}	0.05866	4	0.42660
9	$DOCA_{\mu\mu}$	0.05599	12	0.08282
10	$d_{max,sig}^0$	0.05087	10	0.17080
11	$ IP_B^{3D} $	0.05032	11	0.15100
12	P_L^{min}	0.04700	14	0.03469
13	L_{xy}	0.04241	6	0.38910
14	$DOCA_{xtrk,loose,pt>0.5}$	0.03732	8	0.31300
15	$d_{min,sig}^0$	0.03611	5	0.39140

Table 24: The BDT ranking of input variable importance and compared with the signal and background separation power of each variable as calculated by TMVA before training.

The left plot in Figure 15 shows the test for over-training of BDT 1, where the BDT outputs for training and testing samples are shown, confirming that the BDT is not over-trained as the training and testing samples are in good agreement with each other. Similar results are seen for the other two BDTs. The right plot contains the training result via the Receiver Operating Characteristic (ROC) curve and the comparison with the BDT used in the 2012 analysis. The ROC curve of the BDT used in the 2012 analysis applied to the 2015 and 2016 high-mass sideband data is also compared to in this figure. We see a large increase in BDT performance compared to the BDT used in 2012 applied to the same data.

The final choice of the configuration parameters of the *continuum*-BDT, as used by TMVA, is shown in Table 25.

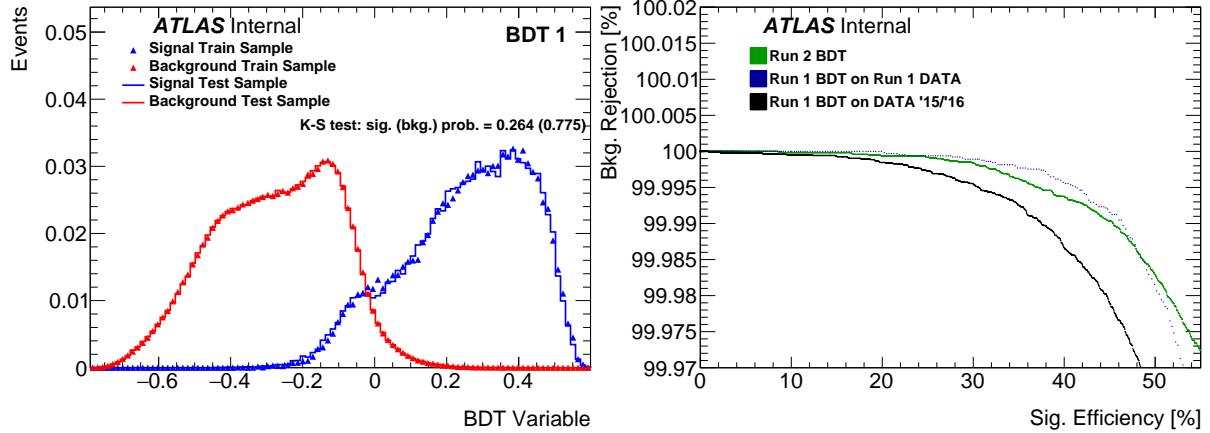


Figure 15: *Left:* Cross-check for over-training for the continuum BDT variable for BDT 1. *Right:* Comparison of the Run 2 ROC curve trained on the sideband collision data and the signal MC with the ROC curve from the Run 1 BDT and from the Run 1 BDT applied to the 2015 and 2016 data. The ROCs are evaluated on the high-mass data sideband.

Parameter	Value
<i>NTrees</i>	500
<i>MinNodeSize</i>	0.2[%]
<i>MaxDepth</i>	4
<i>BoostType</i>	AdaBoost
<i>AdaBoostBeta</i>	0.5
<i>UseBaggedBoost</i>	True
<i>BaggedSampleFraction</i>	0.6
<i>SeparationType</i>	GiniIndex
<i>nCuts</i>	100
<i>NormMode</i>	EqualNumEvents

Table 25: Configuration parameters of the *continuum*-BDT.

The optimal values for the configuration parameters *MinNodeSize* and *AdaBoostBeta* were found with the help of a grid scan using background rejection at 36% signal efficiency on the ROC-curve as a figure of merit of classifier performance. The *MaxDepth* parameter has also been studied in the similar way. For this parameter, it has been found that the performance of classifier improves with increasing of *MaxDepth*, whereas the discrepancy between its performance on the training and testing sample, accessed by the Kolmogorov-Smirnov test, becomes larger, leading to the over-training. In order to not compromise the generality of the classifier, the value *MaxDepth*=4 has been chosen. The value of *NTrees* has been chosen large enough to allow the training to converge. The choice of other parameters does not have any significant impact on the training result.

15:
t.b.c.16:
t.b.c.

759 We have tested the chosen BDT configuration against different training samples including various sample
760 re-weighting, previous selection cuts and increase of statistics. We found no real impact on the performance
761 of the BDT proving that the chosen configuration is very robust.

762 In addition we tested a BDT variable built without the use of the isolation variable as input: the study
763 is described in Appendix ???. As expected, the separation power is reduced thus the BDT configuration
764 described in this section is the one used in the analysis.

765 11 DataMCComparison

766 The shapes of distributions of the discriminating variables used to separate out the combinatorial back-
 767 ground are compared in data and MC samples. This analysis is detailed in [6] [App. A and C].

768 11.1 Reference channel as control sample for signal: $B^+ \rightarrow J/\psi K^+$

769 The MC sample for the reference channel $B^+ \rightarrow J/\psi K^+$ is used to compare the signal MC shapes with
 770 the collision data. The MC sample is re-weighted with the QLC and the DDW described in Secs. 5.2
 771 and 5.3. Also data-driven scaling factors correcting muon reconstruction efficiencies [MCP reference] and
 772 the pile-up weights [PRW reference] are applied. As shown in the plot 50 there is a residual discrepancy
 773 in the number of reconstructed primary vertices after applying all analysis selections, however, as shown
 774 in the appendix plots, it is not responsible for the residual data MC discrepancies observed in the variables
 775 and shown below. For the data sample, the shape of the background distribution for each discriminating
 776 variable is estimated using the events falling into the left ($5080 \text{ MeV} < m_{J/\psi K^\pm} < 5180 \text{ MeV}$) and the
 777 right ($5380 \text{ MeV} < m_{J/\psi K^\pm} < 5480 \text{ MeV}$) sidebands.

778 The left sideband contains also a fraction of mis-reconstructed decays that behave signal-like from the
 779 point of view of the discriminating variables. We re-weight the left sideband considering the combinatorial
 780 contribution only: the net effect is that some of the signal will be also subtracted as mis-reconstructed
 781 B events.⁶ It is proven that the mis-reconstructed events have the same shape as the signal in the
 782 discriminating variables considered, so no distortion of the signal shapes is created by this subtraction.

783 The shape from the rescaled sideband events is subtracted statistically from the signal region. The number
 784 of background events feeding into the signal region is obtained by a binned maximum likelihood fit to the
 785 mass distribution. Figure 16 shows these binned fits to the mass distributions in four bins of p_T and η of
 786 B -mesons. These bins effectively separate the data sample by mass resolution and signal-to-background
 787 ratios, which improves subtraction of the background; the binning is taken from Figure 38(a). The fit
 788 model consists of a double Gaussian for the signal, an error function for the mis-reconstructed events
 789 and an exponential for the continuum background. The results of these binned fit analysis are perfectly
 790 compatible with the unbinned fit described in Section 12 and used for the yield extraction.

791 The distributions of p_T and η of the B is rechecked in figures 17, the distributions of few discriminating
 792 variables are shown in figures 18, and the distribution of the continuum BDT variable is shown in figure 19.
 793 The distributions of all the other variables are shown in the appendix D.

794 Typically, except for deviations in individual bins, the overall shapes of distributions agree well between
 795 data and MC. The agreement is very much alike to what we obtained in the Run 1 analysis. The observed
 796 differences in shape are accounted for as systematics with the procedure described in Section 13. In
 797 addition, the discrepancy seen in the data-MC comparison of the continuum BDT in the $B^+ \rightarrow J/\psi K^+$
 798 events is assessed and exploited in the evaluation of the efficiency in the $B_s^0 \rightarrow \mu^+ \mu^-$ signal in Section ??.

799 A statistical test has been run over the background-subtracted data and MC shapes. Due to vast amount of
 800 available data a χ^2 statistical technique, which takes into account bin errors and therefore is less sensitive
 801 to the bins with low statistics, has been chosen. The tests results are shown in Table 26. Some of the

17:
 ¶8.f.
 ¶9.f plots to
 be added to
 ap-
 pendix

20:
 t.b.c.

21: t.b.u.
 after error
 propaga-
 tion,
 trigger
 scale
 factors,
 etc.

⁶ From the binned fit we extract the number of combinatorial background events in the left and in the right sidebands (let's call them C and D) and also in the signal region (let's call it A). Then we normalise the sideband distributions to A/(C+D), effectively subtracting the right amount of combinatorial background feeding into the signal region.

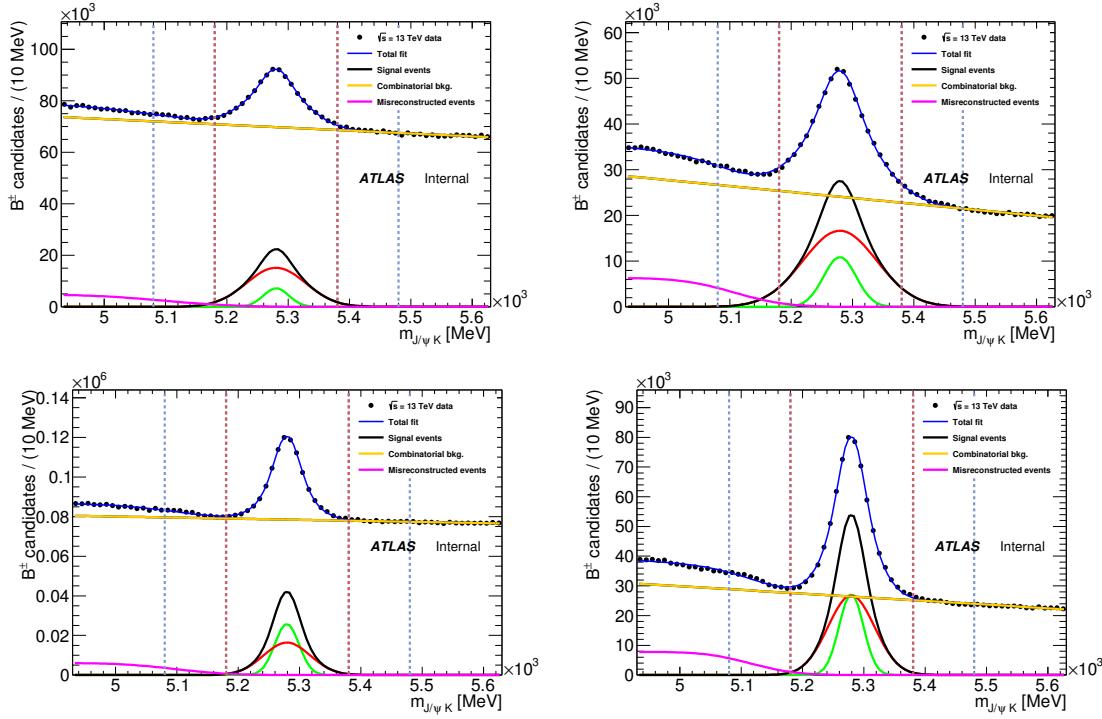


Figure 16: Fits to the invariant mass distributions of $B^+ \rightarrow J/\psi K^+$ events in bins of p_T and η of B-mesons. For the plots on the left (right) $p_T^B < 19\text{GeV}$ ($p_T^B \geq 19\text{GeV}$), for the plots on the top (bottom) $\eta^B \geq 1.3$ ($\eta^B < 1.3$), respectively. All other analysis cuts are applied.

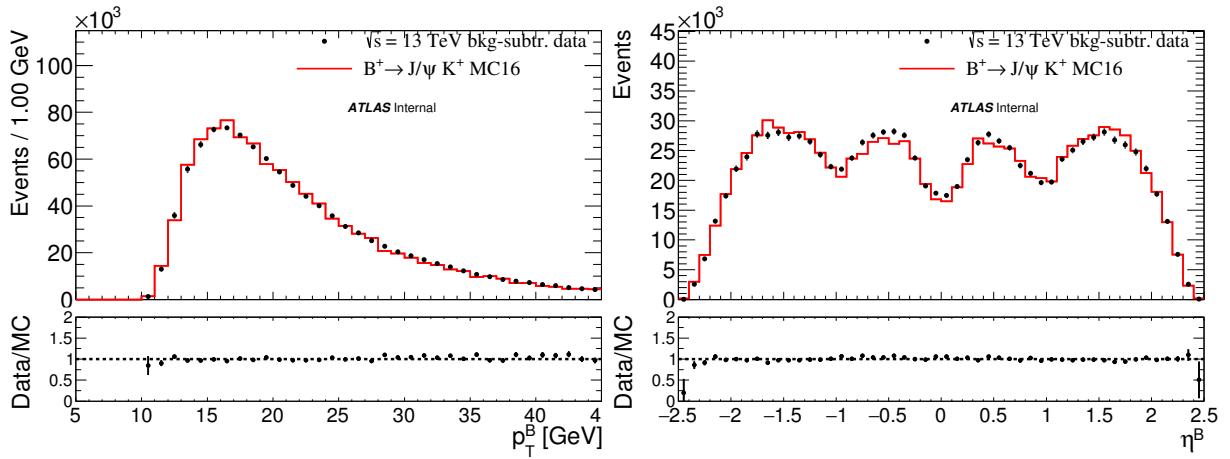


Figure 17: Cross-check on the p_T^B and η^B distributions of the $J/\psi K$ candidates in data and signal MC. The black dots correspond to the sideband subtracted data, while the red histograms correspond to reweighted MC normalised to the number of data events.

802 errors, in particular errors on the B-kinematic weights, are not propagated at the moment, which might
 803 explain some of the discrepancies we observe. Also MC is missing few important corrections (trigger
 804 efficiency scale factors, etc.)

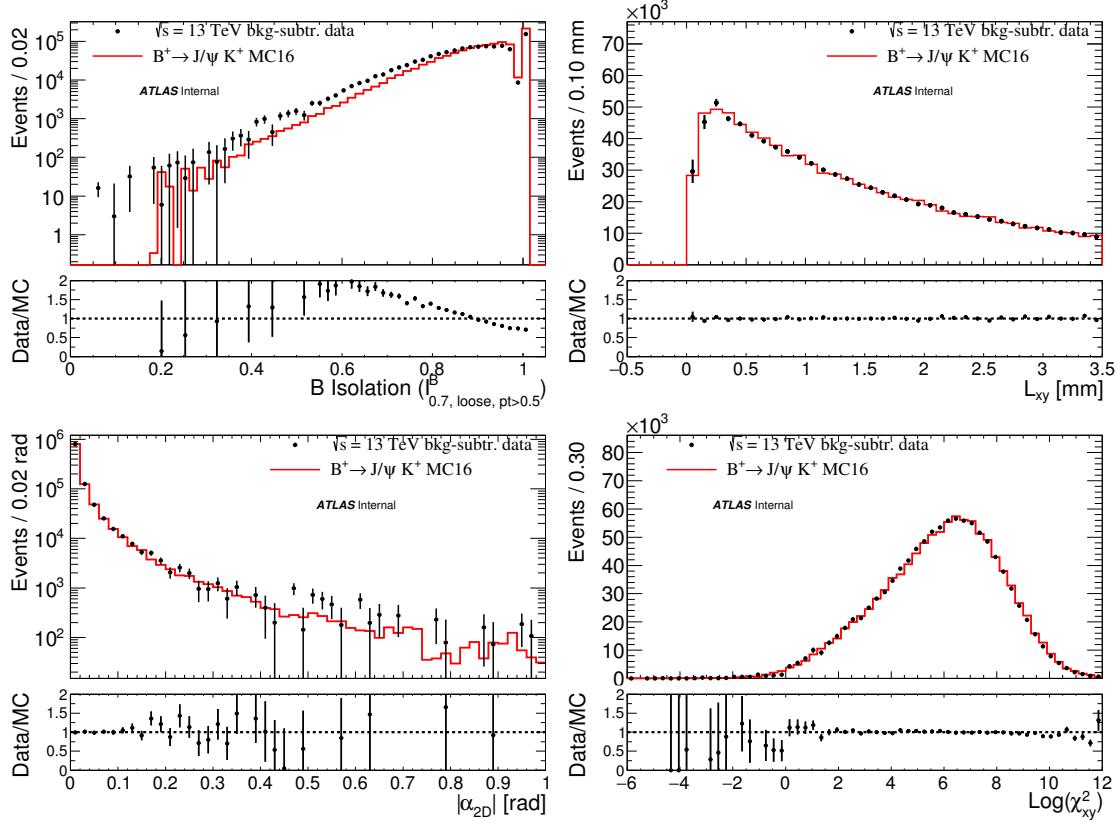


Figure 18: From left to right, from top to bottom: Data and signal MC distributions in $B^+ \rightarrow J/\psi K^+$ events for the B isolation variable, the transverse decay length (L_{xy}), $|\alpha_{2D}|$ and χ^2_{xy} that represents the separation between production (PV) and decay (SV) vertices. The black dots correspond to the sideband subtracted data, while the red points correspond to reweighted MC normalised to the number of data events.

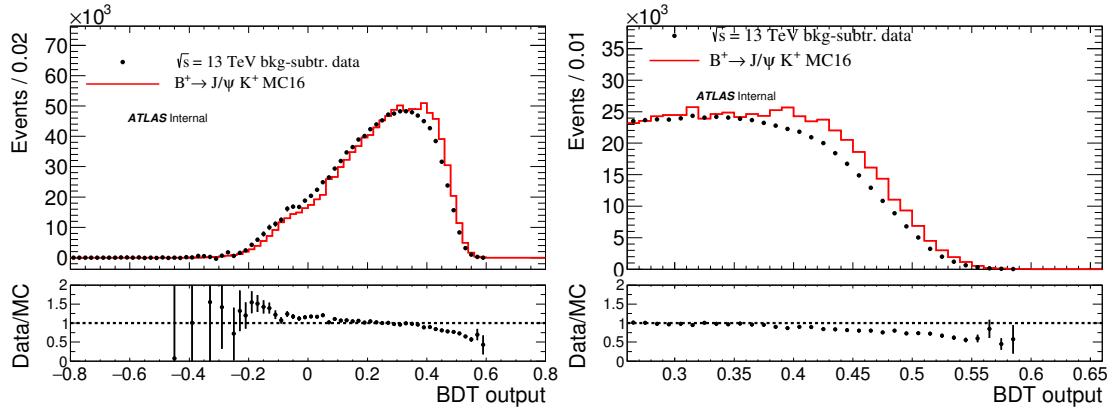


Figure 19: Data and signal MC distribution in $B^+ \rightarrow J/\psi K^+$ events for the continuum BDT variable. The black dots correspond to the sideband subtracted data, while the red points correspond to reweighted MC normalised to the number of data events. The left plot shows the same distribution zoomed in the region of interest of the analysis.

Table 26: χ^2 test results by discriminating variable of the MC versus background-subtracted data distributions shown in the figures 18 and the appendix D. Separation variables are ordered by their importance in the BDT ranking; the topmost variable has the highest importance.

Variable	χ^2/ndf	p-value
ΔR	23.7/29	0.743304
$ \alpha_{2D} $	75.8/49	0.00832271
$BIsolation(I_{0.7,\text{loose},pt>0.5}^B)$	3178.5/57	0
$\text{Log}(\chi_{\mu,xPV}^2)$	153.0/79	1.18809e-06
$\text{Log}(\chi_{xy}^2)$	93.2/59	0.00301653
$ \Delta\phi_{\mu\mu} $	31.7/34	0.581102
p_T^B	74.5/35	0.000112134
$ IP_B^{3D} $	212.4/59	3.48804e-19
$N_{\text{trk,loose},pt>0.5}^{\text{clos}}$	151.3/59	4.76158e-10
$DOCA_{\mu\mu}$	287.4/79	1.90241e-25
L_{xy}	39.8/34	0.227679
$ d_0 ^{\text{max sig.}}$	60.0/49	0.134521
P_L^{min}	19.0/21	0.586505
$ d_0 ^{\text{min sig.}}$	66.6/39	0.00384554
$DOCA_{x\text{trk,loose},pt>0.5}$	188.6/39	1.64653e-21

805 11.2 Alternate reference channel: $B_s^0 \rightarrow J/\psi\phi$

806 The same procedure can be applied on data reconstructed as $B_s^0 \rightarrow J/\psi\phi$. In this case, a few extra
 807 selection cuts are necessary to address the extra kaon in the final state and to select the ϕ mass window:
 808 both kaon p_T are requested > 1 GeV and for the ϕ mass we require $|m_{hh} - m_\phi| < 15$ MeV. The additional
 809 cuts defined in Section 4.5 are also applied. The invariant mass fits in this case are shown in figure 20
 810 for the three trigger categories. The fits are performed binned and a Gaussian PDF is used for the signal
 811 shape while a third order Chebyshev is used for the combinatorial background PDF.

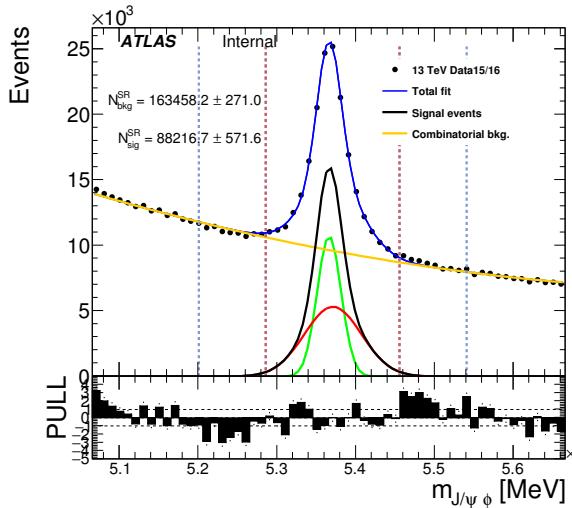


Figure 20: Fit to the invariant mass distribution of $B_s^0 \rightarrow J/\psi\phi$ events.

812 The data/MC comparisons of p_T and η distributions for the B_s^0 are shown in Figure 21, while the
 813 distributions of few discriminating variables are shown in Figure 22. Finally the distribution of the
 814 continuum BDT variable is shown in Figure 23. The signal distributions in data are obtained by subtracting
 815 the sideband shapes in each variable.

816 Additional studies of the $B_s^0 \rightarrow J/\psi\phi$ are presented in the appendix E.

22:
t.b.d.

817 11.3 Yield stability during run

818 The stability of the gain in the B_s^0 sidebands in the 2015 and 2016 data taking periods is shown in Figure 24.
 819 The gain is calculated as the yield of events per period divided by the luminosity in that period. The gain
 820 appears to be stable during the run, within statistical uncertainties.

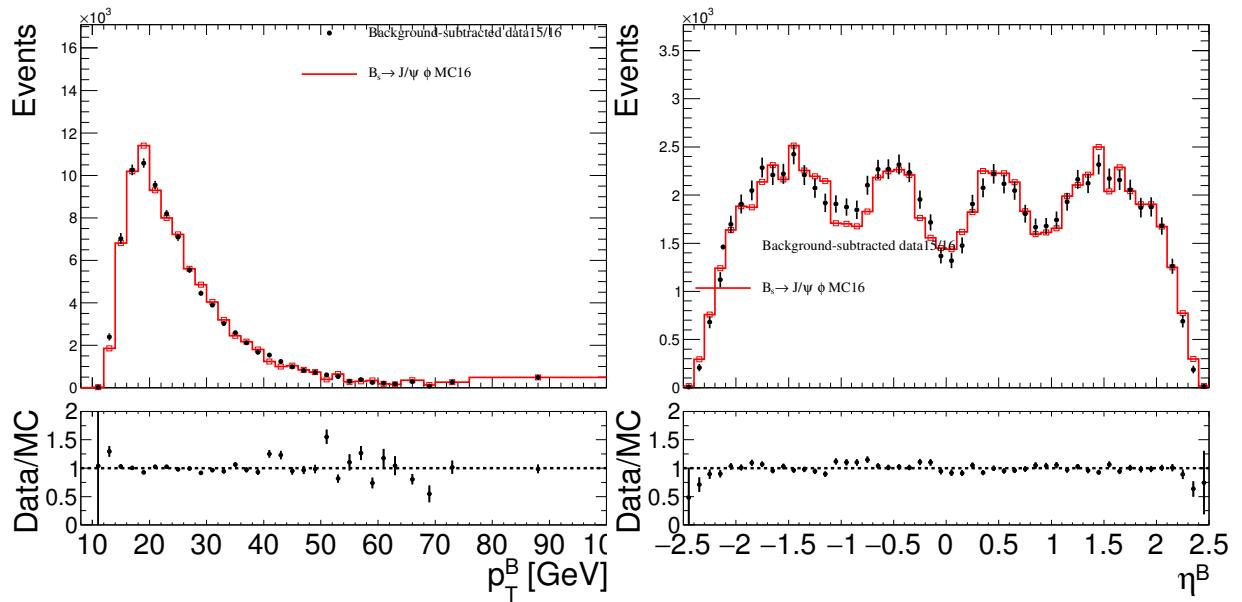


Figure 21: Cross-check on the p_T^B and η^B distributions of the $J/\psi\phi$ candidates in data and signal MC. The black dots correspond to the sideband subtracted data, while the red histograms correspond to reweighted MC normalised to the number of data events.

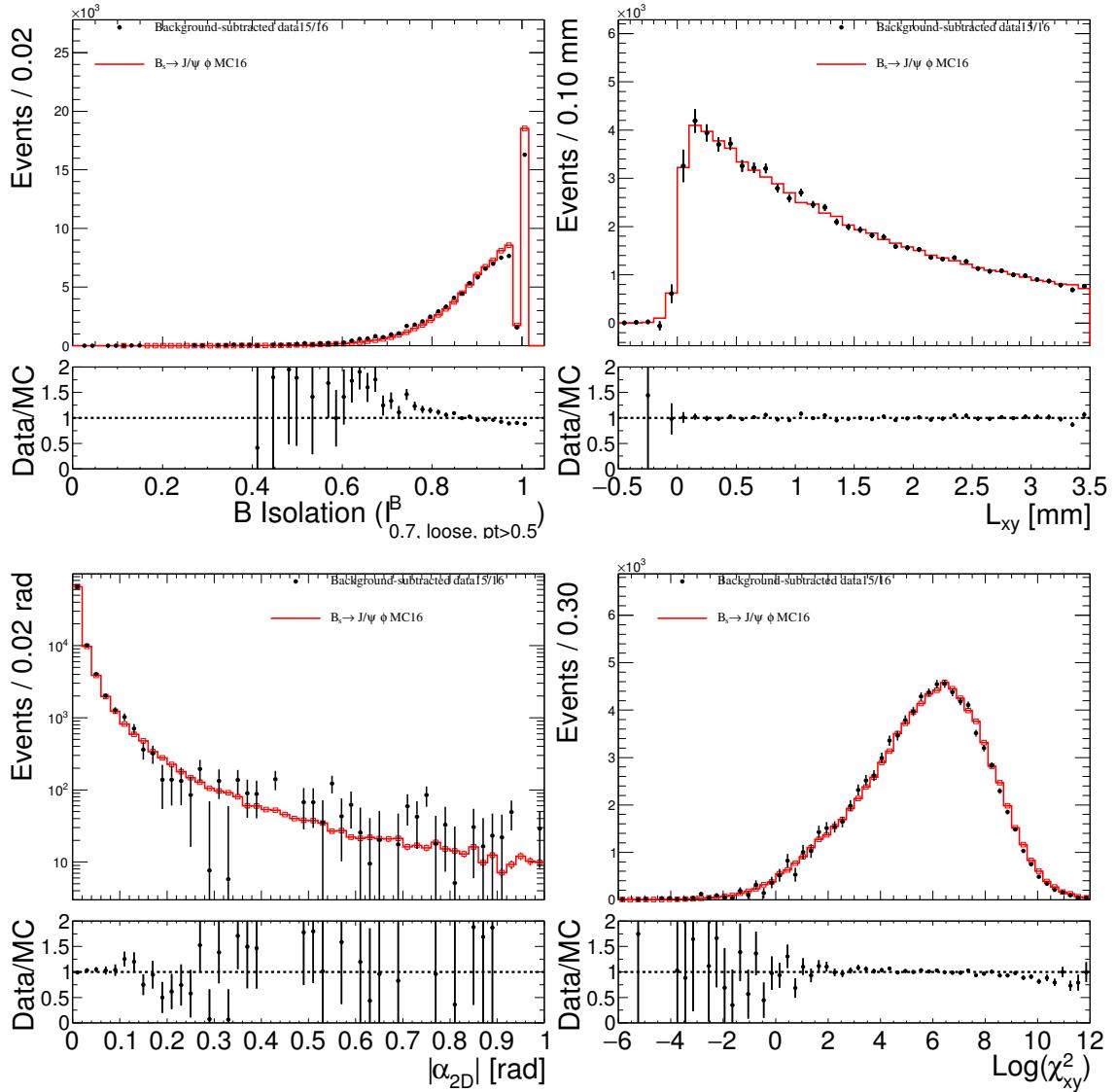


Figure 22: From left to right, from top to bottom: Data and signal MC distributions of $B_s^0 \rightarrow J/\psi \phi$ events for the B isolation variable, the transverse decay length (L_{xy}), $|\alpha_{2D}|$ and χ_{xy}^2 that represents the separation between production (PV) and decay (SV) vertices. The black dots correspond to the sideband subtracted data, while the red points correspond to reweighted MC normalised to the number of data events.

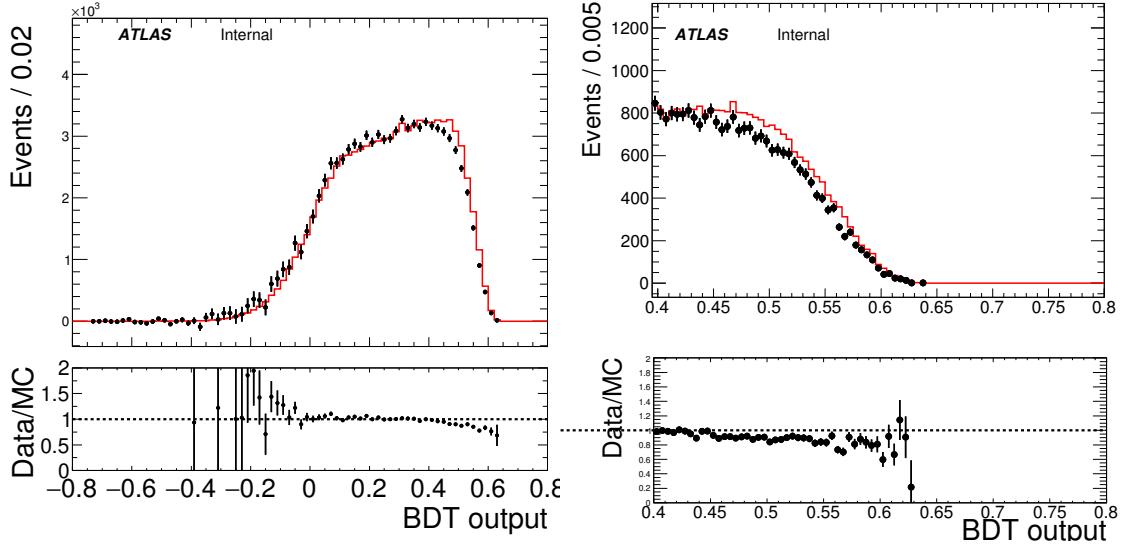


Figure 23: Data and signal MC distribution in $B_s^0 \rightarrow J/\psi\phi$ events for the continuum BDT variable. The black dots correspond to the sideband subtracted data, while the red points correspond to reweighted MC normalised to the number of data events. The left plot shows the same distribution zoomed in the region of interest of the analysis.

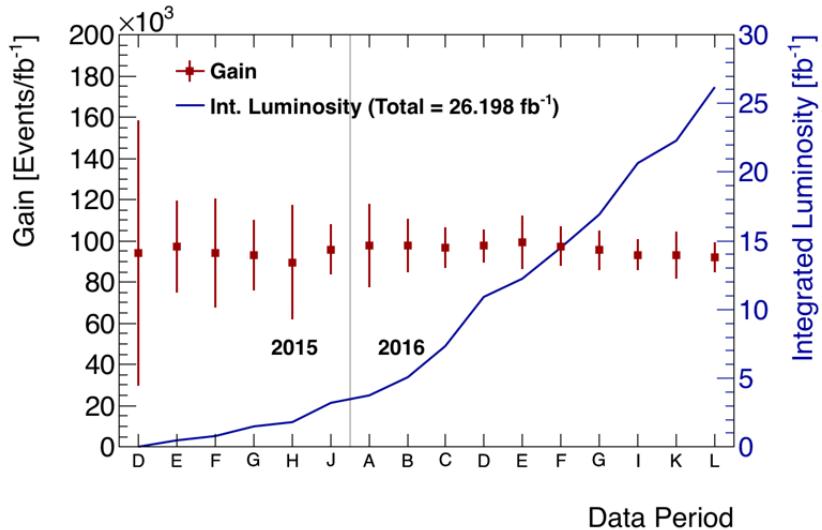


Figure 24: Stability of B_s^0 gain for 2015 and 2016 data taking periods. Also shown is the integrated luminosity for the analysis when the HLT_mu4_mu6_bBmumu and HLT_mu4_mu6_bBmumu_Lxy0 triggers are applied to the 2015 and 2016 data respectively.

821 12 BPlus Yield

822 12.1 Preselection

823 Before carrying out the fit, a reasonably large $B^+ \rightarrow J/\psi K^+$ signal has to be isolated in the data. In order
824 to do this we apply the following selection:

- 825 • **Trigger:** We require the events, collected up to run number 284484, to pass HLT_mu6_mu4_bJpsimumu;
826 events collected above this run and before 302956, and belonging to the main stream, should pass
827 HLT_mu6_mu4_bJpsimumu_Lxy0.
828 Events above or in this run are required to pass HLT_mu6_mu4_bJpsimumu_Lxy0_delayed.
- 829 • **Requirements on B:** $M(B^+) \in [4930, 5630]\text{MeV}$ and $\chi^2/N_{\text{d.o.f.}} < 6$ and $p_T(B^+) > 8\text{GeV}$.
- 830 • **Requirements on J/ψ :** $M(J/\psi) \in [2915, 3275]\text{MeV}$ and $\chi^2/N_{\text{d.o.f.}} < 10$.
- 831 • **Kinematics on B:** $DR < 1.5$ and $|a_{2D}| < 1$ and $L_{xy} > 0$.
- 832 • **BDT:** $BDT > 0.24$.

833 This selection is applied to 2015 and 2016 data with the background as the dimuon combinatorial
834 contribution to the $B_s^0 \rightarrow \mu^+ \mu^-$ signal (see figure 11(a)). Figure 25 shows the normalized distributions of
835 the BDT score for both signal MC, data and the background.

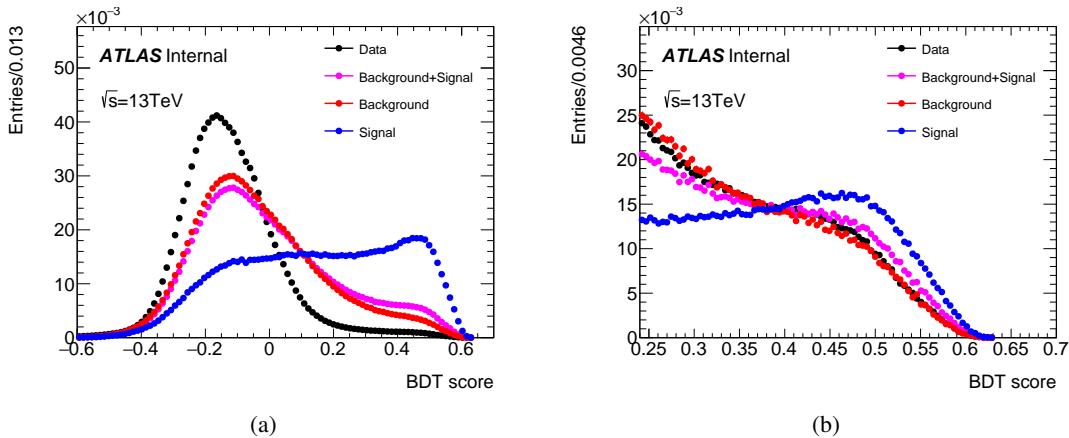


Figure 25: BDT score distribution for signal MC, background and data.

836 The background was extracted from the $b\bar{b} \rightarrow J/\psi X$ sample, after the signal component was filtered out.
837 The cutflow for the three samples can be seen in figure 26.

838 In these figures *No signal* implies removing the $B^+ \rightarrow J/\psi K^+$ component from the respective sample, this
839 obviously only makes sense for the $b\bar{b} \rightarrow J/\psi X$ sample, otherwise it has no effect. The corresponding
840 efficiencies are in tables 27, 28 and 29. After applying this selection the background composition is as
841 shown in table 30.

842 Additionally figure 27 shows the distribution of the number of decays versus the index of the decay, when
843 the most common decays are put at the beginning. The plot also shows the corresponding cumulative
844 distribution.

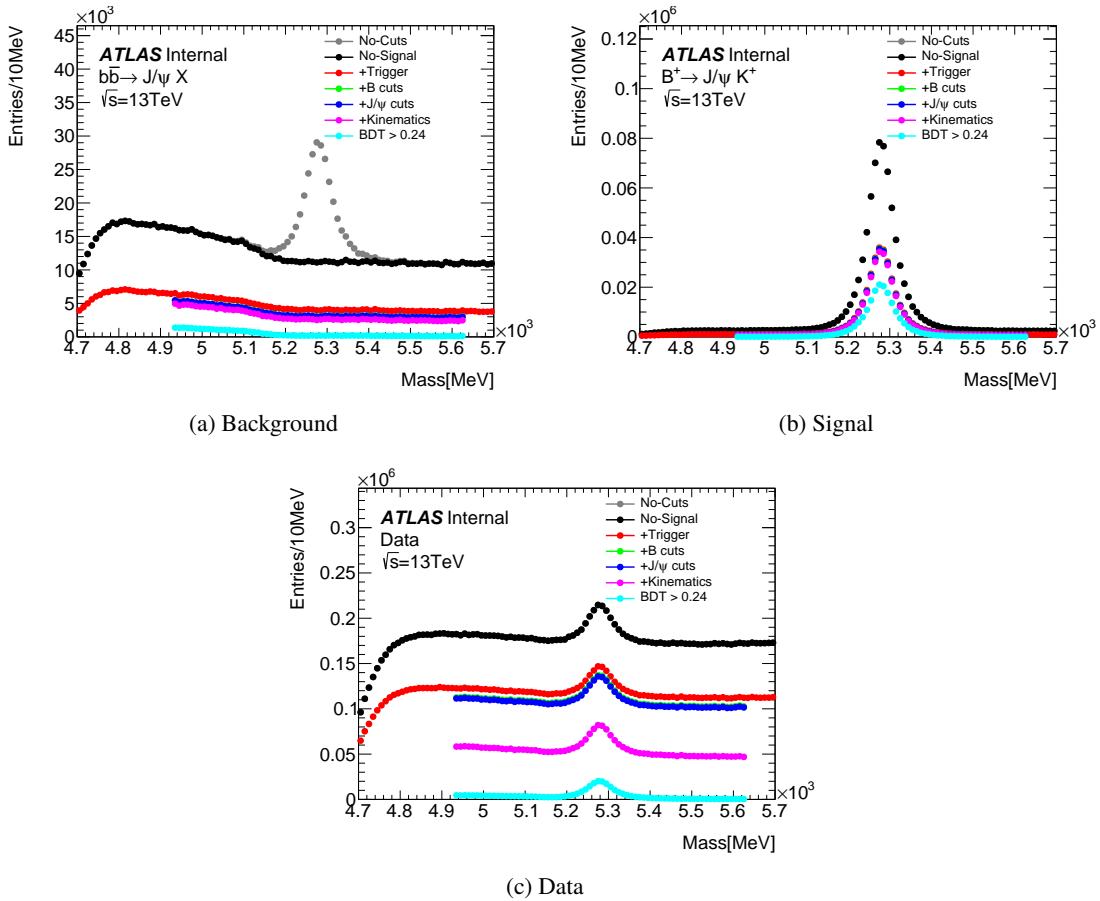


Figure 26: Cutflow associated to the simulated $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^+ \rightarrow J/\psi K^+$ samples, and for the data.

Cut	Efficiency	Number of Events
No Signal	0.91	1695814
Trigger	0.37	629375
Cuts on B	0.15	251257
Cuts on J/ψ	0.15	249617
Cuts on kinematic quantities	0.13	215851
BDT	0.02	31101

Table 27: Cutflow efficiency for $b\bar{b} \rightarrow J/\psi X$ sample.

845 There are 81 candidates classified as *missing particle* for which the matching is not possible given that
 846 there are no particles in the event record, that satisfy the requirements.

847 12.2 Procedure

848 To extract the B^+ yield for the reference channel, we use a unbinned maximum likelihood fit to the mass
 849 distribution. The fit is performed simultaneously on data and the MC samples from which the models are

Cut	Efficiency	Number of Events
No Signal	1.00	990211
Trigger	0.43	429273
Cuts on B	0.36	353872
Cuts on J/ψ	0.35	351176
Cuts on kinematic quantities	0.34	335628
BDT	0.19	188707

Table 28: Cutflow efficiency for $B^+ \rightarrow J/\psi K^+$ sample.

Cut	Efficiency	Number of Events
No Signal	1.00	23470163
Trigger	0.66	15553565
Cuts on B	0.33	7691019
Cuts on J/ψ	0.32	7602035
Cuts on kinematic quantities	0.16	3860800
BDT	0.01	292550

Table 29: Cutflow efficiency for data.

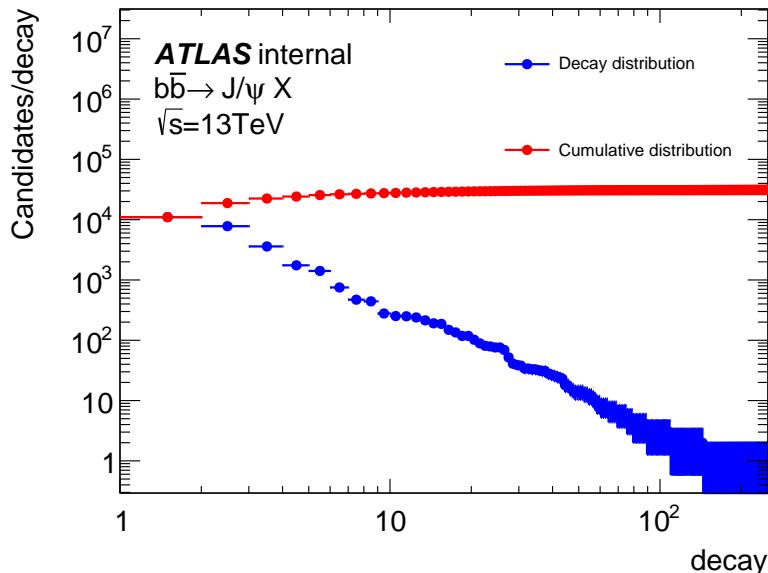


Figure 27: Decay distribution and the corresponding cumulative distribution.

taken. With respect to the 2011 analysis, the event-by-event error is no longer used as it provides very limited or no separation.

Figure 28 shows the B^+ invariant mass for the three categories of partially reconstructed decays. Figure 29 shows the same distribution in data; the B^\pm signal is quite evident, but with visible contributions from at least three background categories. On the left of the B^\pm peak we find the partially reconstructed B decays (PRD, e.g. $B^{+0} \rightarrow K^{*+/0} J/\psi$, $B^+ \rightarrow K^+ \chi_{c1,2}$) where one or more of the final state particles are missed in the reconstruction. On the right side, it is expected a contribution from the reflection of the

Frequency	Decay
11015	$B0[K^*0[\pi:-K+]J/\psi[\mu+:\mu-]]$
7797	unmatched
3598	$B+[K^*+[pi0[\gamma:\gamma]K+]J/\psi[\mu+:\mu-]]$
1750	$B+[K+:chi_1c[\gamma:\gamma]J/\psi[\mu+:\mu-]]$
1411	$B+[pi+:J/\psi[\mu+:\mu-]]$
751	$B0[K^*0[\pi:-\gamma:K+]J/\psi[\mu+:\mu-]]$
471	$B0[pi:-K+:J/\psi[\mu+:\mu-]]$
442	$B+[rho+[\pi0[\gamma:\gamma]\pi+]J/\psi[\mu+:\mu-]]$
278	$B0[K^*_{-20}[\pi:-K+]J/\psi[\mu+:\mu-]]$
252	$B+[K^*+[pi+:K0[K_S0]]J/\psi[\mu+:\mu-]]$
250	$B+[K^*+[pi+:K0[K_L0]]J/\psi[\mu+:\mu-]]$
239	combinatorial
214	$B+[pi0[\gamma:\gamma]K+:J/\psi[\mu+:\mu-]]$
191	$B0[rho0[pi:-pi+]J/\psi[\mu+:\mu-]]$
187	$B+[\gamma:pi+:J/\psi[\mu+:\mu-]]$
149	$B_s0[pi:-pi+:J/\psi[\mu+:\mu-]]$
135	$B_s0[\eta:[\gamma:\rho0[pi:-pi+]]J/\psi[\mu+:\mu-]]$
118	$B+[\gamma:K^*+[pi0[\gamma:\gamma]K+]J/\psi[\mu+:\mu-]]$
118	$B+[K+:chi_0c[\gamma:\gamma]J/\psi[\mu+:\mu-]]$
102	$B+[K^*_{-2}[\pi0[\gamma:\gamma]K+]J/\psi[\mu+:\mu-]]$
89	$B^*-[B-[K:-J/\psi[\mu+:\mu-]]\gamma]$
81	missing particles
79	$B+[\gamma:K+:chi_1c[\gamma:\gamma]J/\psi[\mu+:\mu-]]$
76	$B+[K^*_{-2}+[pi+:K0[K_L0]]J/\psi[\mu+:\mu-]]$
76	$B+[K^*_{-2}+[pi+:K0[K_S0]]J/\psi[\mu+:\mu-]]$
69	$B+[pi+:K0[K_L0]J/\psi[\mu+:\mu-]]$
52	$B+[pi+:K0[K_S0]J/\psi[\mu+:\mu-]]$
41	$B+[K^*+[pi0[e+:e-:\gamma]K+]J/\psi[\mu+:\mu-]]$
39	$Xi_b+[Xi_bar:J/\psi[\mu+:\mu-]]$

Table 30: Background composition of $b\bar{b} \rightarrow J/\psi X$ sample selection is applied.

857 Cabibbo-suppressed $B^\pm \rightarrow J/\psi \pi^\pm$ decay with the assignment of the kaon mass to the final state pion.

858 Finally there is the combinatorial background, which MC studies suggest to be composed, after our
 859 selection cuts, mostly of $b\bar{b} \rightarrow J/\psi X$ that spans on the whole mass range, and consists of random
 860 combination of J/ψ (produced promptly in pp collisions or in feed-down from B-decays) with a track.

861 For the extraction of the B^+ yield, we define the following event categories:

- 862 • $N_{J/\psi K^\pm}$: number of $B^\pm \rightarrow J/\psi K^\pm$ events (signal events for this fit);
- 863 • $N_{J/\psi \pi^\pm}$: number of $B^+ \rightarrow J/\psi \pi^+$ events;
- 864 • N_{pr} : number of partially reconstructed events (PRD);
- 865 • N_{comb} : number of combinatorial events.

866 The fit is based on an unbinned extended maximum likelihood fit to data with simultaneous inclusion of three
 867 MC samples to guide the modeling of several fit components. The MC samples are introduced to model

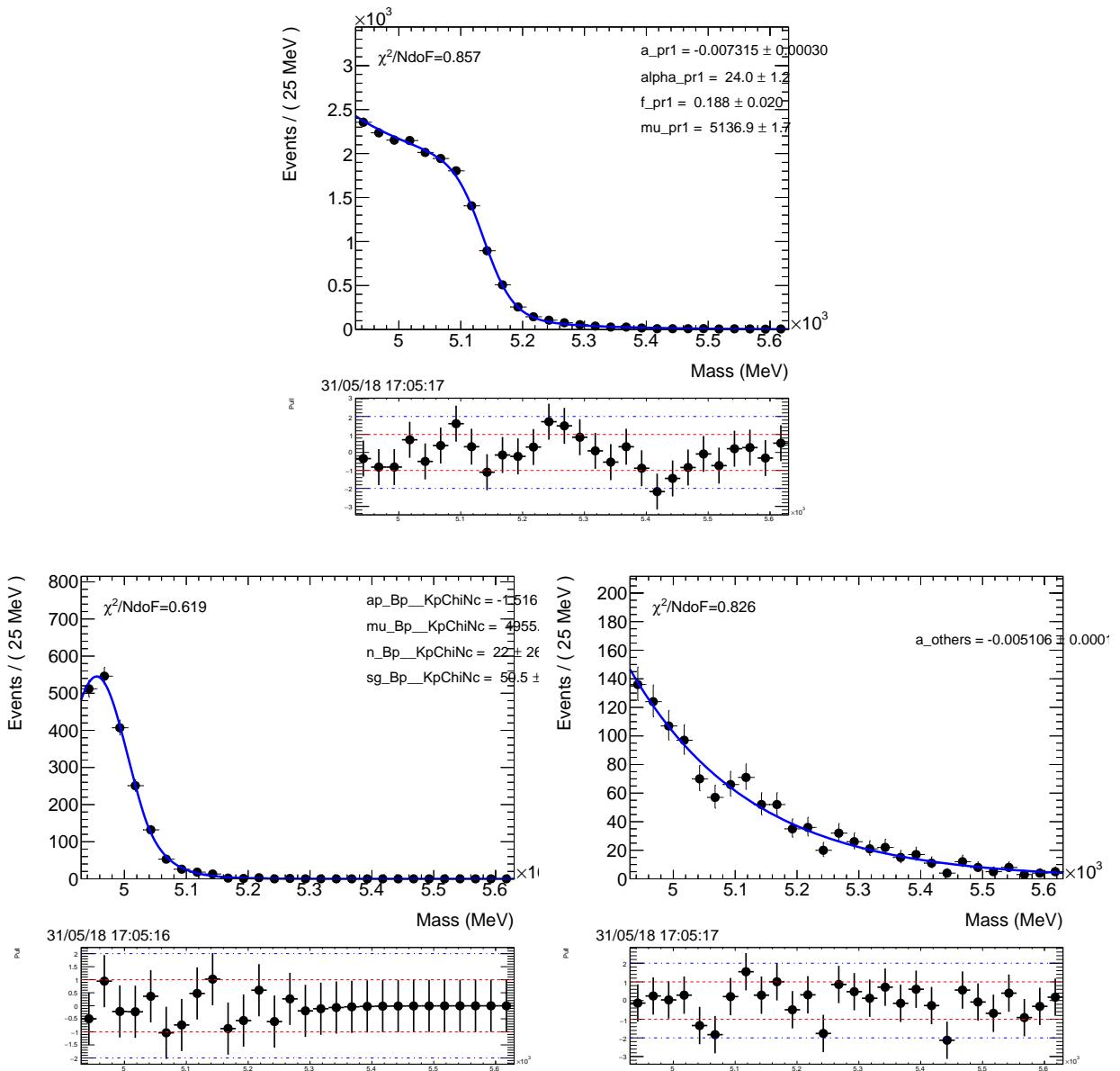


Figure 28: *Left:* $J/\psi K^\pm$ invariant mass distribution for all B^\pm candidates in 2012 data. *Right:* partially reconstructed B decays contributing to the background as described by Monte Carlo.

868 accurately the shapes of the $B^\pm \rightarrow J/\psi K^\pm$ events as well as the most critical background components: the
 869 PRD and the $J/\psi \pi^\pm$ decays. By fitting the above mentioned MC samples simultaneously, we constrain
 870 the fit parameters of the corresponding fit components to their MC values. This results in an *MC assisted*
 871 determination of the background shapes, while automatically accounting for the statistical uncertainties
 872 of the MC. Two free parameters, one for the mass scale and the other for the mass resolution are extracted
 873 from data to accommodate for the possible data-MC difference in the shapes for $B^\pm \rightarrow J/\psi K^\pm$, PRD and
 874 $J/\psi \pi^\pm$ decays.

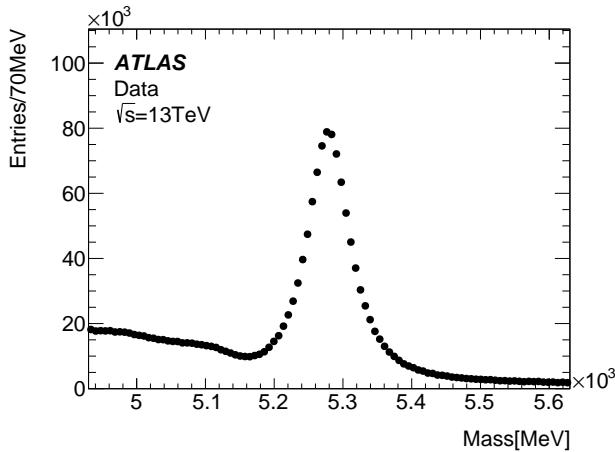


Figure 29: B^+ invariant mass distribution in data.

875 The probability density functions (PDF's) used in this fit are described below. The parameters of
 876 each function are tied among the data and MC samples, so that effectively the parameter values that are
 877 determined by the fit on the MC components are propagated to functions used to fit the data components.

878 The three MC samples included in the simultaneous fit are:

- $B^\pm \rightarrow J/\psi K^\pm$ MC events: $B^+ \rightarrow J/\psi K^+$ ($J/\psi \rightarrow \mu^+ \mu^-$) exclusively generated. These events include also radiative decays where the B^\pm radiates a γ . The two cases are modeled separately in the fit but they are both included in the signal definition.

- $B^\pm \rightarrow J/\psi \pi^\pm$ MC: used to describe this reflection when the K mass is mis-assigned to the final state pion.

- $bb \rightarrow J/\psi X$ MC sample: PRD events are characterized by this sample. Where we can distinguish the true origin of each reconstructed B^\pm candidate. In order to model the PRD contribution to the mass fit, we identify three classes of decays contributing with slightly different mass shapes. The three classes are as follows:

- PRD1: these events produce a step shape at 5150 MeV: these are mainly $B^0 \rightarrow J/\psi K^{*0}$, $B^+ \rightarrow J/\psi K^{*+}$, $B^0 \rightarrow J/\psi \rho^0$, $B^+ \rightarrow J/\psi \rho^+$, $B^0 \rightarrow J/\psi K^+ \pi^-$, and $B^+ \rightarrow J/\psi K^+ \pi^0$.

- PRD2: these events produce a bump at 5050 MeV. These events are mainly $B^+ \rightarrow \chi_{c1} K^+$.

- PRD3: this component includes everything else from the $J/\psi X$ sample and has a smoother shape in the mass distribution.

893 The PDF's used are the following:

- $B^\pm \rightarrow J/\psi K^\pm$ signal: Johnson S_U [**Johnson**], this PDF contains the core of the signal events and it is parameterized as follows:

$$\text{Johnson } S_U = \frac{1}{\sqrt{1+t^2}} \left[t + \sqrt{1+t^2} \right]^{-\gamma - \frac{1}{2}\delta \ln(t+\sqrt{1+t^2})}, \text{ where } t = \frac{m - \xi}{\lambda} \quad (5)$$

896 Parameters λ , ξ , δ and γ determine the shape of the Johnson S_U distribution. The parameters ξ
 897 and λ control the position and width of the distribution, respectively. The sign of γ determines the
 898 position of the tail (left or right to the peak).

899 A data-MC discrepancy is allowed in the resolution and scale parameters; the Johnson S_U distribution
 900 is convoluted with a Gaussian resolution function with μ mean and s_σ width, where the
 901 former parameter describes the scale difference between data and MC and the latter, describes the
 902 resolution differences.

903 The model of the shape is taken from $B^\pm \rightarrow J/\psi K^\pm$ MC events and its parameters are left floating
 904 in the fit to be determined from the $B^\pm \rightarrow J/\psi K^\pm$ MC sample in the simultaneous fit. The
 905 normalization and the 2 scaling factors allow the distribution to adjust to the data peak. Details on
 906 the specific implementation of these parameterizations are found in Appendix G.3.1.

- 907 • radiative contribution to $B^\pm \rightarrow J/\psi K^\pm$ decays: when the B radiates a γ , the mass shape results
 908 skewed on the left so we need to consider this component separately. A single Johnson S_U is used.
 909 The shape parameters are determined from the radiative decays selected in the $J/\psi K^\pm$ MC sample.
 910 The normalization and the 2 scaling factors allow the distribution to adjust to the data peak. Details on
 911 the specific implementation of these parameterizations are found in Appendix G.3.1.

912 The relative abundances of the two signal components (non-radiative and radiative) are parametrized
 913 using relative fractions so that the total number of $B^\pm \rightarrow J/\psi K^\pm$ events is extracted from the fit
 914 together with the three fractions.

- 915 • $J/\psi \pi^\pm$ final states: similarly to the $B^\pm \rightarrow J/\psi K^\pm$ component, a Johnson S_U . Also in this case the
 916 shape parameters are determined from the $J/\psi \pi^\pm$ MC sample in the simultaneous fit.
 917 The 2 overall scaling factors on the mass scale and resolution are also applied to these parameter-
 918 izations: in the fit they are driven by the effect on the $B^\pm \rightarrow J/\psi K^\pm$ component.

919 • PRD component:

- 920 – PRD1 with 5150-step shape. We parameterize these events with a Fermi-Dirac (FD) plus an
 921 Exponential The FD distribution is parametrize as follows:

$$M_{FD}(m|\mu_{FD}, \alpha_{FD}) = \frac{1}{1 + e^{\frac{m - \mu_{FD}}{\alpha_{FD}}}}$$

922 where α_{FD} accounts for the slope, while μ_{FD} accounts for the mass scale at which the step-like
 923 effect occurs. The shape parameters are taken from the PRD1 events in the $(bb \rightarrow \mu^- \mu^+ X)_{pr}$
 924 MC sample in the simultaneous fit. In the parameterization used for data the FD mass scale
 925 parameter μ_{FD} is redefined as $\mu_{FD}^{MC} + s_\mu$ where s_μ is the scale factor already defined above
 926 and it is common to the $B^\pm \rightarrow J/\psi K^\pm$ and $J/\psi \pi^\pm$ events. The PDF is then convoluted with
 927 the Gaussian resolution function with the s_σ width already defined above and common to the
 928 $B^\pm \rightarrow J/\psi K^\pm$ and $J/\psi \pi^\pm$ events.

- 929 – PRD2 with 5000-bump shape. We parameterize these events with a FD plus an Exponential
 930 Again the 2 overall scaling factors on the mass scale and resolution are applied to these
 931 parameterizations. The shape parameters are taken from the PRD2 events in the $(bb \rightarrow \mu^- \mu^+ X)_{pr}$
 932 MC sample in the simultaneous fit.
- 933 – PRD3 with a falling shape. We parameterize these events with an Exponential plus a constant
 934 term. Again the 2 overall scaling factors on the mass scale and resolution are applied to
 935 these parameterizations. The shape parameters are taken from the PRD3 events in the $(bb \rightarrow \mu^- \mu^+ X)_{pr}$
 936 MC sample in the simultaneous fit.

937 Details on the specific implementation of these parameterizations are found in Appendix G.3.2.
 938 The relative abundances of the PRD three components are taken from the MC predictions, while the
 939 PDG values of these are used as a systematic check. The total number of PRD events is extracted
 940 from the data fit.

- 941 • combinatorial background: exponential function

$$M_c(m|a) = e^{am}$$

942 with the shape parameter that is left floating in the fit to be extracted from data.

943 Figure 30 shows the results of the fits: the projection on the B^+ mass are shown for data. Appendix G
944 contains the complete fit results including the projections on the mass for the MC samples. As an example,
945 we show in Figure 31 all the simultaneous fit results. Table 31 shows the parameters from the fit, again as
946 an example. In Appendix G similar tables are shown.

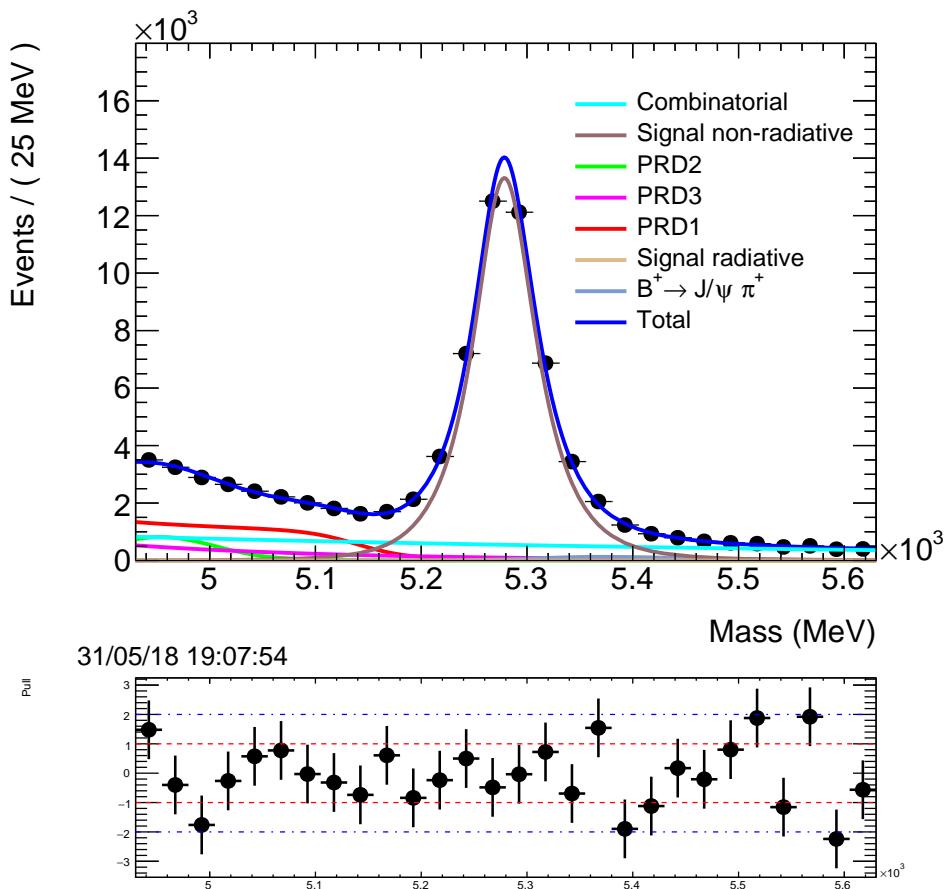


Figure 30: Fit to 2015 and 2016 data. The various background and signal components are shown.

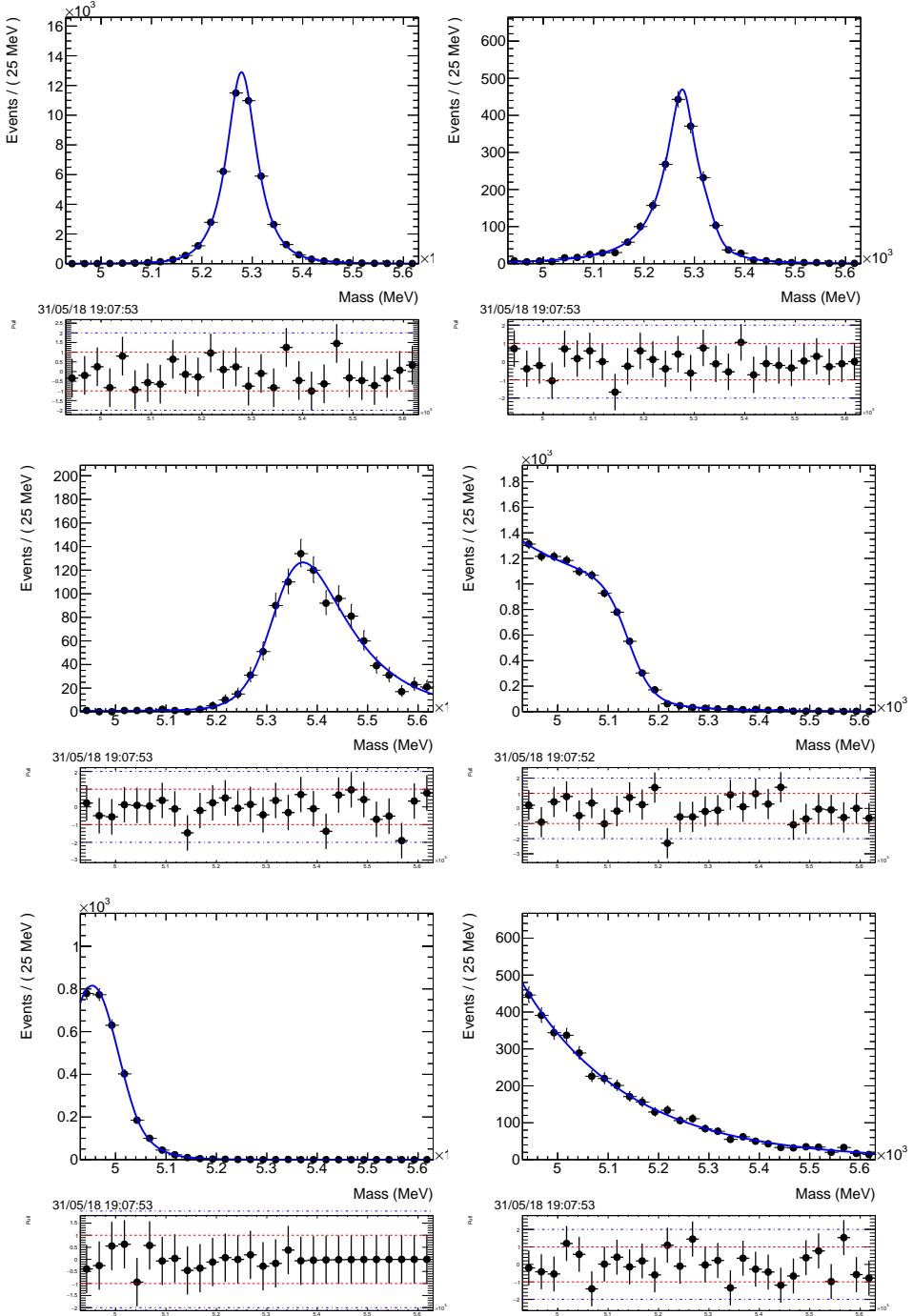


Figure 31: Fit projections on the MC samples fitted simultaneously with the data. From left to right, from top to bottom: non-radiative $B^+ \rightarrow J/\psi K^+$ signal, radiative $B^+ \rightarrow J/\psi K^+$ signal, peaking background $B^+ \rightarrow J/\psi \pi^+$, PRD1, PRD2 and PRD3.

N_{Signal}	46856 ± 287
N_{JPSIPI}	1326 ± 214
N_{PRD}	14959 ± 791
N_{Comb}	15107 ± 1050
s_μ	$0.2 \pm 0.2 \text{ MeV}$
s_σ	$9.1 \pm 0.6 \text{ MeV}$

Signal	
Johnson $S_U + \text{Gaussian}$	
ξ	5276.2 ± 0.3
λ	38.0 ± 0.5
δ	1.795 ± 0.032
γ	-0.198 ± 0.018
μ	5240.4 ± 9.4
σ	33.8 ± 4.2
Radiative signal	
Johnson $S_U + \text{Gaussian}$	
ξ	5282.3 ± 1.5
λ	17.7 ± 2.0
δ	0.348 ± 0.055
γ	0.489 ± 0.041
μ	5278.1 ± 3.2
σ	40.0 ± 1.9
Signal pdf fractions	
Signal f_A	0.9751 ± 0.0019
Signal f_B	0.0122 ± 0.0019
Signal f_C	0.0345 ± 0.0127
$J/\psi\pi$	
Johnson $S_U + \text{Gaussian}$	
ξ	5317.8 ± 2.8
λ	80.4 ± 3.7
δ	1.754 ± 0.190
γ	-1.620 ± 0.124
μ	5491.5 ± 309.9
σ	305.7 ± 116.5
Gaussian frac	0.035 ± 0.017
PDR1 Fermi-Dirac + Exponential	
μ_{FD}	5140.1 ± 1.0
α_{FD}	20.5 ± 0.8
FD frac	0.893 ± 0.011
a	-0.0036 ± 0.0004
PDR2 Fermi-Dirac + Exponential	
μ_{FD}	5013.9 ± 2.3
α_{FD}	17.0 ± 1.8
FD frac	0.944 ± 0.037
a	-0.0099 ± 0.0028
PDR3 Exponential + constant	
a	-0.0079 ± 0.0004
Expo frac	0.72 ± 0.02
PRD fractions	
PRD f_a	0.896 ± 0.002
PRD f_b	0.107 ± 0.002
Combinatorial Exponential	
a	-0.00221 ± 0.00016
Yields of control samples	
$N_{\text{NON-RadiativeSignal}}^{\text{ctl}}$	137455 ± 371
$N_{\text{JPSIPI}}^{\text{ctl}}$	48332 ± 220
$N_{\text{PRD}_{\text{tot}}}^{\text{ctl}}$	23741 ± 154

Table 31: **Example, Run1!** Results for the parameters of the fit to the N3 category. The definitions of all the parameters are found in the text and tables in Appendix G.

947 12.3 Systematic Uncertainties in the Extraction of the B^\pm Yields

948 Some of the systematic effects are taken care of automatically in the fit: the MC effect of limited statistics,
 949 for example, is included in the statistical fit error from the simultaneous fit. In addition the data-MC
 950 discrepancy in the mass scale and resolution is included in the fit using the aforementioned scaling
 951 factors.

952 Systematic effects on the reference channel yield are coming mainly from the parametrization of the mass
 953 distributions and the data-MC discrepancies. The approximation of mass shapes in the simultaneous fit by
 954 PDFs is a source of systematic uncertainty on the fit result which we evaluate by repeating the fit varying
 955 the fit models. The difference between the modified and default fit results is then taken as the systematic
 956 uncertainty. The systematic uncertainties can be found in Table 33 as an example.

957 The list of systematic checks is the following:

- 958 • data-MC discrepancies in B kinematics: discrepancies on the B meson kinematics between data and
 959 MC introduce a source of uncertainty. GLC- and DDW-weighted $J/\psi K^+$ and $J/\psi \pi^+$ MC samples
 960 account for biases introduced by generator level preselection cuts and discrepancies in 2D (p_T, η)
 961 spectrum of the B^\pm with respect to data. These reweighted samples are used as alternative MC
 962 samples in the simultaneous fit.
- 963 • Alternative model for the PRD3 component: in the PRD3 mass distribution we observe a structure
 964 not modeled by the Exponential PDF used. We account for this structure by adding a Gaussian to
 965 the default model.
- 966 • Combinatorial background model: the combinatorial background smoothly crosses our fit window
 967 and it is modeled with an exponential function in the default fit. This shape is not MC driven. We
 968 evaluate the systematic uncertainty coming from the unknown combinatorial background model by
 969 changing the model to a linear function.
- 970 • $B^\pm \rightarrow J/\psi K^\pm$ signal peak charge asymmetry: the MC sample used in the default fit is from
 971 $B^+ \rightarrow J/\psi K^+$ events only. In order to account for shape variations relative to $B^- \rightarrow J/\psi K^-$, we
 972 repeat the fit using the latter.
- 973 • Alternative models for PRD1 and PRD2 components: systematics from the choice of PRD1 and
 974 PRD2 models are assessed by replacing the principal (default) FD function with complementary
 975 error function ($\text{Erfc}(m) = \int_m^\infty e^{-t^2} dt$) in each of the two models.
- 976 • data-MC PRD composition discrepancies: the PRD MC sample has a relative decay modes abund-
 977 ances different than what would be expected according to current PDG measurements. Different
 978 abundances imply different mass shapes: we account for this data-MC shape discrepancy by intro-
 979 ducing per-event weights derived from the Pythia and PDG differences. The weights correct for the
 980 MC shapes as well as for the relative abundances of the three PRD components. Two additional free
 981 parameters take care of the relative PRD fractions: these parameters have unit central values and
 982 are allowed to vary (via an external Gaussian constraint) within the PDG uncertainties on the BRs.

983 The relative difference on the B^+ yields are given in table 32.

984 The main systematic uncertainties on the B^+ yield come from the MC reweighting, the PRD reweighting
 985 and the combinatorial background parametrization. The overall systematic uncertainties result to be
 986 between 0.8 – 0.9%, depending on the data categories, as it is summarized in Table 33.

Systematic uncertainties in trigger category N3				
	Systematic	Signal	$J/\psi\pi^\pm$	ratio
nr.1	MC reweighting (GLC&DDW)	0.24%	2.21%	2.46%
nr.2	PRD re-weighting	0.38%	4.44%	4.08%
nr.3	PRD3 alternate model	0.02%	0.10%	0.12%
nr.4	Combinatorial alternate model	0.39%	28.26%	27.98%
nr.5	Signal peak charge asymmetry	0.02%	10.93%	10.91%
nr.6	PRD1&2 alternate models	0.03%	1.11%	1.14%
Total		0.60%	30.72%	30.43%

Table 32: **Example, Run1!** Relative changes with respect to the default fit obtained with each systematic check described in the text. The total effect is given both in relative effect and in the absolute number of events (or value of π/K ratio). Systematic uncertainties in the N3 category shown.

	$N_{J/\psi K^\pm}$ total syst.	$N_{J/\psi\pi^\pm}$ total syst.	ratio total syst.
N1	1.69%	34.96%	34.75%
N2	2.00%	47.77%	48.67%
N3	0.60%	30.72%	30.43%
2011	1.10%	38.11%	37.63%

Table 33: **Example, Run1!** Total systematic uncertainties on the number of events for both $J/\psi K^\pm$ and $J/\psi\pi^\pm$ channels and on the π/K ratio.

987 12.4 $B^\pm \rightarrow J\psi K^\pm$ yield results

988 We measure the $B^\pm \rightarrow J\psi K^\pm$ reference channel yield with full systematic uncertainty evaluation in the
 989 four measurement categories as shown in Table 34.

Measured reference channel yield			
Trigger Category	Yield	stat.uncert.	syst.uncert.
N1	2512	± 91	± 42
N2	5173	± 84	± 104
N3	46856	± 287	± 279
2011	95867	± 423	± 1050

Table 34: **Example, Run1!** Result of the reference channel yield measurement in the three trigger categories.

990 13 EfficiencyAndAcceptance

991 Acceptances and efficiencies for both reference and signal channels enter in Eq. 1 through their ratio $R_{A\epsilon}$.
 992 $R_{A\epsilon}$ is evaluated using $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^+ \rightarrow J/\psi K^+$ signal MC samples, after applying to both MC
 993 samples the GLC and data driven corrections (see Section ??).

994 For the sake of clarity, we decided to split each component of $R_{A\epsilon}$ into two separate terms defined as the
 995 product of a pure acceptance term A and a pure efficiency term ϵ , as stated in Eq. 1. $R_{A\epsilon}$ is then defined
 996 as:

$$R_{A\epsilon} = \frac{A_{B^+ \rightarrow J/\psi K^+} \times \epsilon_{B^+ \rightarrow J/\psi K^+}}{A_{B_s^0 \rightarrow \mu^+ \mu^-} \times \epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}} \quad (6)$$

997 More precisely, the definition of the A and ϵ terms can be summarised as follows:

- 998 • A takes into account the loss in acceptance, with respect to the chosen phase-space fiducial volume,
 999 due to the MC generator cuts applied on the final state particles (the two muons for both channels
 1000 and the kaon for the reference channel only) in order to produce the MC samples used in the analysis.
 1001 The fiducial volume is defined as $p_T^B > 8.0$ GeV and $|\eta_B| < 2.5$, while the final state particle cuts
 1002 are: $p_T^\mu > 4.0$ GeV and $|\eta_\mu| < 2.5$ for muons and $p_T^K > 1.0$ GeV and $|\eta_K| < 2.5$ for the kaon. The A
 1003 term is defined as the ratio between the number of events passing the final state particle cuts and the
 1004 number of events in the fiducial volume. This term has been evaluated using a specific "un-biased"
 1005 MC sample where only the phase-space fiducial volume cuts on the B meson were applied and
 1006 it is different between the reference and the signal channels. Data driven corrections to the p_T^B ,
 1007 $|\eta_B|$ distribution are applied to the MC sample, both in the numerator and the denominator, in
 1008 order to reproduce the spectrum observed in data. The kinematic cuts are applied to the various
 1009 particles only at the truth level. The values of the A term are reported in Table 35. **They refer**
 1010 **to Run1 analysis and are here just to give an idea of the order of magnitude of values and**
 1011 **uncertainties.**

- 1012 • ϵ takes into account all the reconstruction effects (selection cuts, trigger efficiency, reconstruction
 1013 efficiency) affecting the two channels. It is defined as the ratio between the number of events passing
 1014 the final selection cuts listed in Section 4 including the cut on the continuum BDT once fixed(listed
 1015 in Section 10) and the number of events passing the final state particle kinematic cuts applied
 1016 to the appropriate truth level quantities. Both terms are evaluated using the simulated MC samples
 1017 available. Since this term involves reconstructed quantities, QLC (see Section 5.2) and data driven
 1018 corrections (see Section ??) have been applied to both the numerator (to reconstructed quantities)
 1019 and the denominator (to truth quantities) in order to properly take into account the effects of a
 1020 modified p_T and $|\eta|$ spectra of the B -meson. Trigger efficiency corrections, that will be taken from
 1021 data and will appear as Scale Factors to be applied to MC, will be added as well once available (see
 1022 Section 6). They will be applied to the numerator only.

1023 The values for the $A \times \epsilon$ and $R_{A\epsilon}$ terms are summarised in Table 36, together with the statistical and
 1024 systematic uncertainties described in the next section. The statistical uncertainties come from the finite
 1025 statistic available for the simulated samples used in the analysis. **They refer to Run1 analysis and are**
 1026 **here just to give an idea of the order of magnitude of values and uncertainties.** The sources of
 1027 systematic uncertainties affecting the values reported in Table 36 are described in Section 13.1.

$A_{B^+ \rightarrow J/\psi K^+}^{2012}$	0.0865 ± 0.0002
$A_{B_s^0 \rightarrow \mu^+ \mu^-}^{2012}$	0.2902 ± 0.0005
$A_{B^+ \rightarrow J/\psi K^+}^{2011}$	0.0819 ± 0.0003
$A_{B_s^0 \rightarrow \mu^+ \mu^-}^{2011}$	0.3035 ± 0.0005

Table 35: A -terms for B^+ and B_s^0 channels for 2012 and 2011 MC samples. The statistical uncertainty due to the finite size of the simulated samples is also reported.

channel	ϵ	$A \times \epsilon$	$R_{A\epsilon}$
B^+	0.0928	$0.0080 \pm 0.28\% \pm 14.2\%$	$0.180 \pm 0.56\% \text{ (stat)} \pm 5.2\% \text{ (syst)}$
B_s^0	0.1522	$0.0441 \pm 0.49\% \pm 10.3\%$	

Table 36: ϵ , $A \times \epsilon$ and $R_{A\epsilon}$ values for B^+ and B_s^0 channels for 2012 (split into the three trigger categories) and 2011 samples. Where present, the first uncertainty is statistical and second one is systematic.

1028 13.1 Systematic uncertainties on $R_{A\epsilon}$

1029 The systematic uncertainty affecting $R_{A\epsilon}$ comes mainly from two sources: the uncertainty related to the
 1030 corrections (GLC, data driven and trigger efficiencies) and the residual discrepancy between data and MC
 1031 samples after the application of these corrections. The total systematic uncertainty quoted in Table 36 is
 1032 the sum in quadrature of these two components.

1033 The first source of systematic uncertainty considered originates from the uncertainty of the GLC, DDW
 1034 and trigger efficiency corrections. For evaluation of the effect a toy study has been performed by varying
 1035 the corrections within their statistical uncertainties and recomputing each term of the $R_{A\epsilon}$ ratio after each
 1036 toy. The RMS of the values obtained in each toy has been quoted as the systematic uncertainty for the
 1037 various quantities.

1038 Another source of systematic uncertainty arises from the discrepancies between data and MC. The sys-
 1039 tematic uncertainty on $R_{A\epsilon}$ was assessed by observing the variation in the efficiency of the final selection
 1040 when re-weighting both the MC samples to the observed data for each of the 15 variables used in the
 1041 continuum BDT. Data are extracted from B^+ events after the subtraction of the background as shown in
 1042 section 11. $R_{A\epsilon}$ is recomputed after the reweighting of each variable one at a time. The discrepancy
 1043 between the values obtained with this procedure and the central values has been considered as the system-
 1044 atic uncertainty due to the specific variable mis-modelling.

1045

1046 The total systematic uncertainty on $A \times \epsilon$ and $R_{A\epsilon}$ due to the reweighting procedure is the sum in quadrature
 1047 of the single systematic uncertainties and it is quoted in Table 36.

1048 14 SignalFit

1049 To extract the signal yield, an unbinned maximum-likelihood fit is performed on the selected events,
 1050 mostly based on the fit performed in the previous analysis, based on the full Run1 dataset. The fit is
 1051 performed on the invariant mass distribution, classifying the events according to different intervals in the
 1052 continuum-BDT output. This is similar to the strategy used by CMS and LHCb.

1053 The events have been classified according to the three bins in continuum BDT_{ss} output. The bins are
 1054 chosen to correspond to a signal efficiency equal to %, and they result ordered according to increasing
 1055 signal-to-noise ratio.

1056 The sensitivity of our study is discussed in this chapter. The model for describing signal and background
 1057 is based on MC and on data collected in the sidebands of the search region.

- 1058 • The models used for the signal and the backgrounds are discussed in Sections 14.1 and 14.2.
- 1059 • The results of the fit to the data in the sidebands and the interpolation in the signal region are
 1060 presented in Section 14.3.
- 1061 • A summary of the baseline fit configuration is given in Section 14.4.
- 1062 • Systematic uncertainties on the fit are discussed in Section 14.5.

1063 14.1 Signal and peaking background

1064 The mass shape of the $B_s^0 \rightarrow \mu^+ \mu^-$, as well as the one of $B^0 \rightarrow \mu^+ \mu^-$, is described by a superposition
 1065 of two Gaussian distributions, both centred at the world average value of the mass. The parameters of this
 1066 distribution will be extracted from MC.

1067 The **peaking background** is composed of $B \rightarrow hh'$, mainly $B_s \rightarrow K^+ K^-$ and $B_d \rightarrow K^\pm \pi^\mp$, in which both
 1068 hadrons are misidentified as muons. Due to the mass distortion related to the $K \rightarrow \mu$ mass assignment,
 1069 and the smaller one for $\pi \rightarrow \mu$, the mass distribution of these events is substantially superimposed with
 1070 the B_d signal.

1071 14.2 Parametrisation of background components

1072 The **combinatorial** (opposite-side) background, following the Run1 analysis, will be described with a
 1073 Chebychev first order polynomial like $f(x) = 1 + \alpha T_1(x) = 1 + \alpha x/1200\text{MeV}$.

1074 The **same-side and same-vertex (SS+SV)** background includes double semileptonic cascade events (e.g.,
 1075 $B \rightarrow D\mu X \rightarrow \mu\mu X'$), which we call SS, where the muons do not originate from the same vertex, and
 1076 events where the muons come from the same vertex (e.g., $B \rightarrow K\mu\mu$), which we call SV. In both cases,
 1077 in Run1 the mass distribution of the two muons was peaked far below the signal region, and the analysis
 1078 was sensitive to a tail of the distribution determined by kinematic limits and detector resolution effects.
 1079 In Run1 this background was fitted with an exponential PDF $f(x) = \exp(\alpha x)$ used for MC events of this
 1080 class, we expect to be able to use the same model.

1081

1082 The **semileptonic** background is due to few-body semileptonic B decays feeding into our final selections
 1083 though a misidentification $h \rightarrow \mu$, in the limit of low energy neutrinos. In particular $B_d \rightarrow \pi\mu\nu$ and
 1084 $B_s \rightarrow K\mu\nu$ can contribute, together with $\Lambda_b \rightarrow p\mu\nu$. The mass distribution for the last process extends
 1085 closer to the signal region, but it is highly suppressed because of a very low probability of misidentifying

23: still
work in
progress

1086 the proton as muon in ATLAS. The contribution from this background is expected to be significantly
 1087 smaller than the SS+SV and the combinatorial background contributions in all bins of the BDT output.
 1088 It is expected to be described with sufficient accuracy by the first-order polynomial and the exponential
 1089 PDF used for the main background components, without adding an extra PDF.

1090

1091 **14.3 Fit to background components from MC and to sideband data**

1092 In each bin in the continuum BDT, the background will be fitted in the sideband of the data sample, and
 1093 interpolated in the search region. This is done in order to optimise the analysis and evaluate its sensitivity
 1094 before proceeding to the unblinding of the signal region and performing a simultaneous fit to signal and
 1095 background.

1096

1097 **14.4 Summary of the fit configuration**

1098 The baseline signal fit to the number of events is expected to include the following PDFs:

- 1099 1. signal PDF: the mass dependence is described by the sum of 2 Gaussians centred at the B_s (or
 1100 B_d) mass. The widths of the Gaussians and their relative fraction, assumed to be identical in all
 1101 continuum-BDT bins, will be taken from MC and fixed in the fit.
- 1102 2. Continuum background PDF: the mass dependence is first order polynomial. The normalisation and
 1103 the slope will be extracted independently in each bin of the continuum-BDT. In full Run1 analysis
 1104 Gaussian constraints were placed on the uniformity of the slope, so that the slope in bin-2 (bin-3)
 1105 was equal to the one in bin-1 within $\pm 40\%$ ($\pm 80\%$).
- 1106 3. Low-mass background PDF: exponential dependence on the mass. The normalisation will be
 1107 extracted independently in each bin of the continuum-BDT, while the shape will be assumed to be
 1108 uniform.
- 1109 4. Peaking background: the mass dependence is described with a Gaussian describing the total
 1110 background.

1111 **14.5 Systematic uncertainties on the fit in the simultaneous fit to B_s and B_d**

1112 The evaluation of the systematic uncertainties due to the fitting procedure will be evaluated as in the
 1113 full Run1 analysis, by applying variations to the baseline model and testing the result behaviour with
 1114 toy-experiments. The corresponding variations in the result of the fit in the baseline configuration will be
 1115 taken as systematic uncertainties.

1116 15 BranchingRatio

1117 The $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction is obtained by means of the formula 1. We list below the various
 1118 inputs appearing in this formula starting from the ones we need to consider from external sources and then
 1119 detailing the ingredients we collected through the various steps of this analysis.

- 1120 • The external inputs needed are the branching fraction for the reference channel that is obtained
 1121 from the PDG [PDG2014] as the product of $\mathcal{B}(B^\pm \rightarrow J/\psi K^\pm) = (1.027 \pm 0.031) \times 10^{-3}$ and
 1122 $\mathcal{B}(J/\psi \rightarrow \mu^+ \mu^-) = (5.961 \pm 0.033)\%$. The relative hadronisation probability f_u/f_s is taken from
 1123 the best experimental result [lhcbfsu14] available: $f_s/f_d = 0.259 \pm 0.015$ using $f_d/f_u = 1$.
 1124 The product of these external inputs gives:

$$\mathcal{F}_{\text{ext}} = \mathcal{B}(B^\pm \rightarrow J/\psi K^\pm \rightarrow \mu^+ \mu^- K^\pm) \times \frac{f_u}{f_s} = (2.36 \pm 0.15) \times 10^{-4}$$

1125 which corresponds to a relative uncertainty of 6.6%

- 1126 • Number of signal events, $N_{\mu^+ \mu^-}$: at this blinded stage we consider the expected number of events
 1127 as ... derived from the SM branching ratio and the error from the default fit as The relative
 1128 statistical uncertainty is about
 1129 • Efficiency weighted number of events for the reference channel (D_{norm}).

1130 Putting together the three terms, the branching ratio is obtained as:

$$\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-) = \frac{\mathcal{F}_{\text{ext}} \times N_{\mu^+ \mu^-}}{D_{\text{norm}}} = (\dots \pm \dots) \times 10^{-9}$$

1131 where the relative uncertainty is about

1132 A two-dimensional Neyman construction [11] based on likelihood ratio ranking is used to estimate the
 1133 68.3%, 95.5% and 99.7% confidence level regions for the combined measurement of $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ and
 1134 $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$. Pseudo-MC experiments are used in the Neyman construction procedure and to verify
 1135 the coverage. See appendix H for a more detailed explanation.

25:
possibly to
be updated
26:
possibly to
be updated

1136 References

- 1137 [1] C. Bobeth, M. Gorbahn, T. Hermann, M. Misiak, E. Stamou et al.,
 1138 $B_{s,d} \rightarrow l^+l^-$ in the Standard Model with Reduced Theoretical Uncertainty,
 1139 *Phys.Rev.Lett.* **112** (2014) 101801, arXiv: [1311.0903 \[hep-ph\]](https://arxiv.org/abs/1311.0903).
- 1140 [2] R. Aaij et al., Measurement of the $B_s^0 \rightarrow \mu^+\mu^-$ branching fraction and effective lifetime and search
 1141 for $B^0 \rightarrow \mu^+\mu^-$ decays, *Phys. Rev. Lett.* **118** (2017) 191801, arXiv: [1703.05747 \[hep-ex\]](https://arxiv.org/abs/1703.05747).
- 1142 [3] V. Khachatryan et al.,
 1143 Observation of the rare $B_s^0 \rightarrow \mu^+\mu^-$ decay from the combined analysis of CMS and LHCb data,
 1144 *Nature* **522** (2015) 68, arXiv: [1411.4413 \[hep-ex\]](https://arxiv.org/abs/1411.4413).
- 1145 [4] C. Alpigiani et al., $B \rightarrow \mu\mu$ analysis on the complete RUN1 dataset,
 1146 tech. rep. ATL-COM-PHYS-2014-1179, CERN, 2014,
 1147 URL: <https://cds.cern.ch/record/1756291>.
- 1148 [5] M. Aaboud et al., Study of the rare decays of B_s^0 and B^0 into muon pairs from data collected
 1149 during the LHC Run 1 with the ATLAS detector, *Eur. Phys. J. C* **76** (2016) 513,
 1150 arXiv: [1604.04263 \[hep-ex\]](https://arxiv.org/abs/1604.04263).
- 1151 [6] Limit on $B_s \rightarrow \mu\mu$ branching ratio based on 4.9 fb^{-1} of integrated luminosity,
 1152 tech. rep. ATL-COM-PHYS-2013-053, CERN, 2013.
- 1153 [7] G. Aad et al., Search for the decay $B_s^0 \rightarrow \mu\mu$ with the ATLAS detector,
 1154 *Phys.Lett. B* **713** (2012) 387, arXiv: [1204.0735 \[hep-ex\]](https://arxiv.org/abs/1204.0735).
- 1155 [8] M. Cacciari, M. Greco and P. Nason, The $P(T)$ spectrum in heavy flavor hadroproduction,
 1156 *JHEP* **05** (1998) 007, arXiv: [hep-ph/9803400 \[hep-ph\]](https://arxiv.org/abs/hep-ph/9803400).
- 1157 [9] M. Smizanska, PythiaB: an interface to Pythia 6 dedicated to simulation of beauty events,
 1158 ATL-COM-PHYS-2003-038, 2003.
- 1159 [10] A. Hoecker et al., TMVA: Toolkit for Multivariate Data Analysis, PoS *ACAT* (2007) 040,
 1160 arXiv: [physics/0703039](https://arxiv.org/abs/physics/0703039).
- 1161 [11] J. Neyman,
 1162 *Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability*,
 1163 *Phil. Trans. Roy. Soc. Lond. A* **236** (1937) 333.
- 1164 [12] A. Andrew, Another efficient algorithm for convex hulls in two dimensions,
 1165 *Information Processing Letters* **9** (1979) 216, ISSN: 0020-0190,
 1166 URL: <http://www.sciencedirect.com/science/article/pii/0020019079900723>.

1167 The supporting notes for the analysis should also contain a list of contributors. This information should
1168 usually be included in `mydocument-metadata.tex`. The list should be printed either here or before the
1169 Table of Contents.

1170 **List of contributions**

A. Cerri	Analysis coordination, Sussex activities coordination, BR extraction, statistical tools, muon identification, MC validation and corrections
I. Ibragimov	Analysis coordination, data-MC comparison, signal BDT
A. Grummer	signal BDT
F. Tresoldi	BR extraction, statistical tools, muon identification, MC corrections, MC Validation
<small>1171</small> U. de Sanctis	Efficiency extraction
A. Campoverde	Ntuple production, B^+ reference yield, data-MC comparisons
W. Walkowiak	Siegen activities coordination, derivation (DAOD) production Ntuple production, statistical tools
P. Buchholz	Siegen activities coordination
S. Seidel	UNM activities coordination
<small>1172</small> S.Yu. Sivoklokov	$J/\Psi\phi$ control sample studies

1173 Appendices

1174 A QLC weights

1175 In this appendix we compare the approach used to calculate the quark level corrections (QLC) to the ap-
 1176 proach used in Run1 [A.1](#) and we show the checks performed on the QLC to verify their consistency [A.2](#)

1177 A.1 Comparison with Run1 procedure

In the previous version of the analysis [4] QLC have been evaluated using only one sample generated with unbiased quark-level cuts and no final state particle cuts, according to the following formula:

$$W_{QL} = \frac{\nu^{unbiased+QLsel+FScuts}(pT_B, \eta_B)}{\nu^{unbiased+FScuts}(pT_B, \eta_B)}$$

1178 Where ν is the number of entries in a (pT_B, η_B) bin and the quark-level selection (QLsel) and final state
 1179 particle cuts (FScuts), implicit in the default MC sample, are explicitly applied to the unbiased sample.
 1180 The inverse of such efficiencies was used to weight events individually, thus correcting with event-weights
 1181 the QL cut biases.

1182 We applied this procedure using our unbiased sample (see [5.2](#)) and as a cross-check, QLC calculated
 1183 using odd events from the $B^+ \rightarrow J/\psi K^+$ unbiased MC have been applied to the quark biased sample
 1184 and the result has been compared to the distribution of the even unbiased events [32](#). The B meson η
 1185 distribution shows good agreement, while the p_T distribution shows an inconsistency at low p_T . The same
 1186 effect is visible swapping odd and even unbiased events.

1187 The source of this discrepancy was found to be the \hat{p}_T ⁷ cut introduced at generator level, figure [33](#)
 1188 shows the \hat{p}_T distribution of the unbiased sample and the \hat{p}_T distribution from a new sample, semi biased,
 1189 generated with the same cuts as the unbiased sample but with tighter \hat{p}_T (7. GeV instead of 5.), the
 1190 distribution of the unbiased sample has been cut at 7 GeV in order to compare the two distributions, that
 1191 are clearly not compatible.

1192 This is due to a regularisation in the Pythia generation for low \hat{p}_T , basically the parton-parton cross
 1193 section becomes too high when \hat{p}_T decreases, causing a violation of unitarity. This regularisation in
 1194 Pythia smoothly takes care of this divergence.

1195 The effect of this feature on the QLC is that corrections calculated using a single unbiased sample are not
 1196 usable, because they would introduce an additional bias due to this \hat{p}_T discrepancy.
 1197 In order to properly calculate the QLC we have to use two samples, as described in [5.2](#), but first we
 1198 have to make sure that the unbiased sample we have is not affected by the \hat{p}_T bias. Figure [34](#) shows the
 1199 comparison of the \hat{p}_T distribution from the unbiased sample and a new sample with the same quark-level
 1200 cuts as the unbiased but with looser \hat{p}_T (3 GeV instead of 5), the distribution of the new sample has been
 1201 cut at 5 GeV in order to compare the two distributions. We are interested in not having the bias in the
 1202 parameter space used in the analysis, therefore the quark-level cuts, final state cuts and B fiducial volume
 1203 cuts ($pT_B > 8$ GeV and $|\eta_B| < 2.5$) are applied. The two distributions look compatible, therefore we can
 1204 use the unbiased sample for the QLC calculation.

1205

⁷ \hat{p}_T is the transverse momentum of the particles outgoing from the hard scattering, in the reference frame of the incoming particles.

1206 A.2 Checks on QLC

1207 The computation of the QLC exploiting two samples has been cross-check applying QLC calculated using
1208 odd events, from both the quark biased and unbiased samples, to the even events of the quark biased
1209 sample. The resulting distributions have been compared to unbiased distributions obtained using only
1210 even events. Figure 35 shows the checks performed on the B^+ sample.
1211 a similar result can be obtained switching even events with odd events or performing these checks for the
1212 other processes, $B_s \rightarrow \mu^+ \mu^-$ and $B_s \rightarrow J/\psi \phi$.

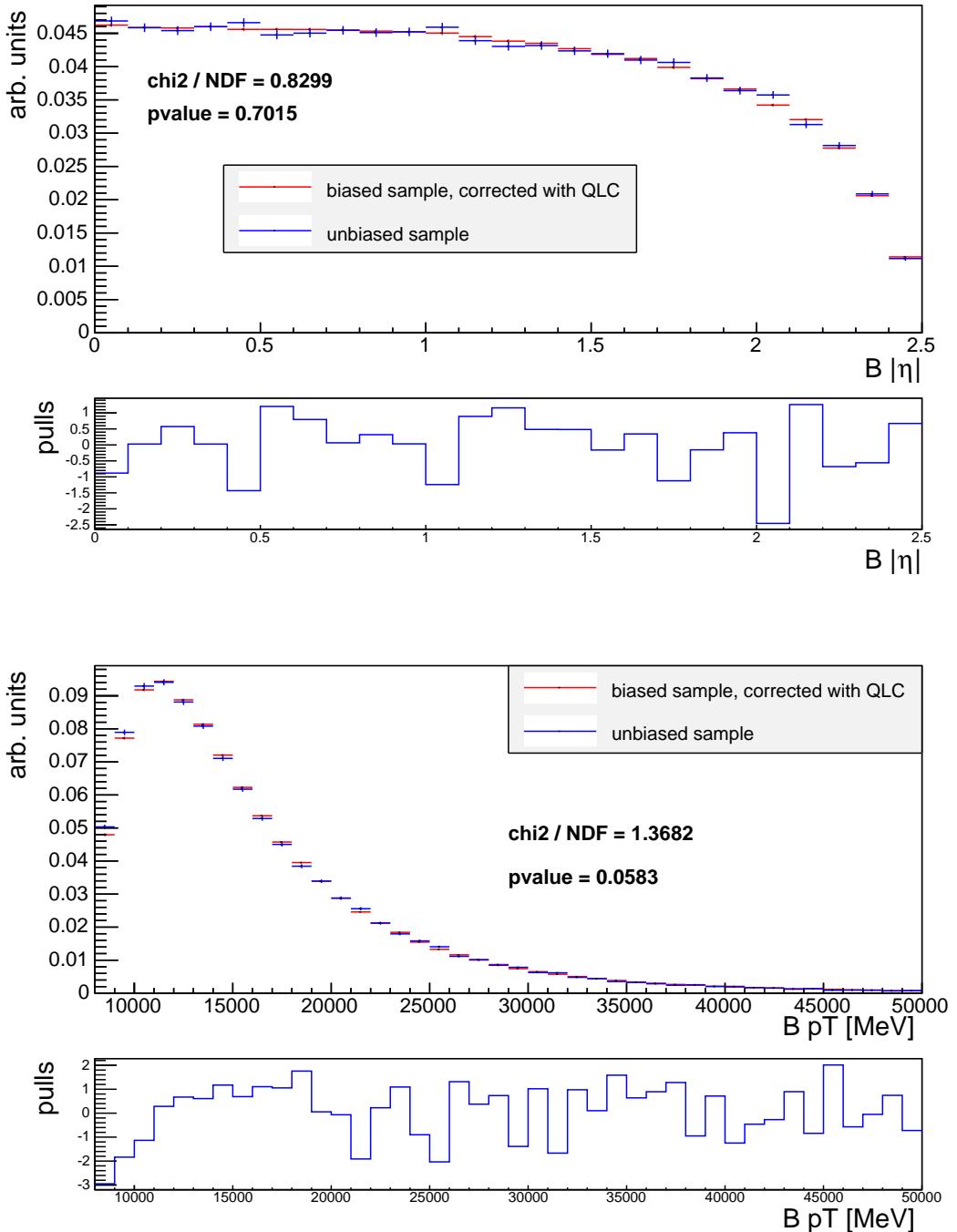


Figure 32: Plots show the comparison of the η_B (first plot) and pT_B (second plot) QLC corrected quark biased distribution and the unbiased distribution. In order to avoid correlations between the distributions, QLC have been calculated using odd numbered events from the unbiased sample and the QLC corrected quark biased distributions are compared with even numbered unbiased events.

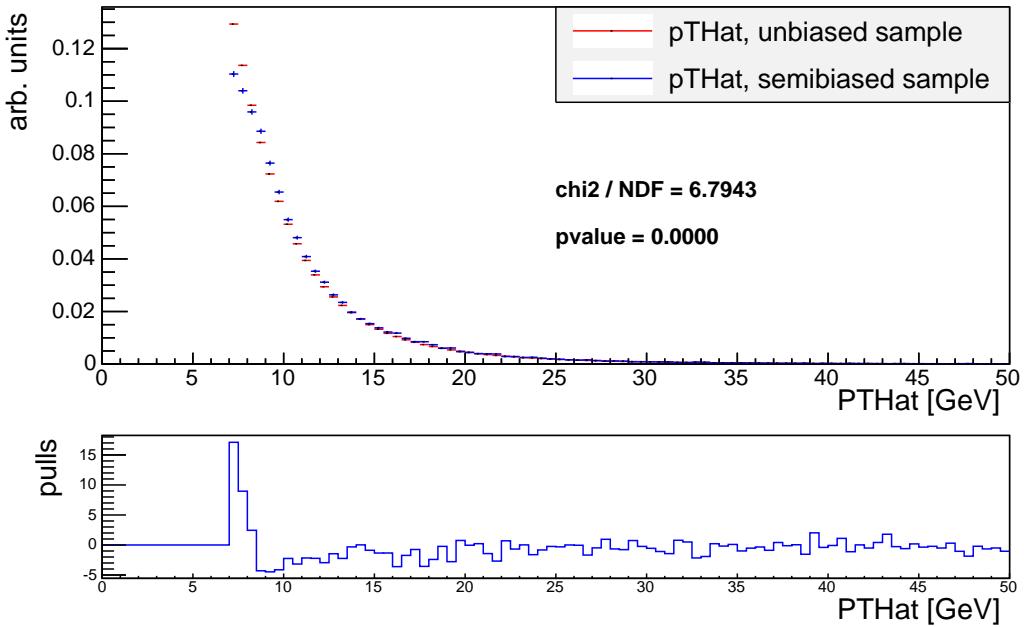


Figure 33: Plot show the comparison of the \hat{p}_T distribution from the unbiased sample and the \hat{p}_T distribution from a new sample, semi biased, with the same quark-level cuts as the unbiased but with tighter \hat{p}_T (7 GeV insted of 5). The unbiased distribution has been cut at 7 GeV in order to compare the two distributions.

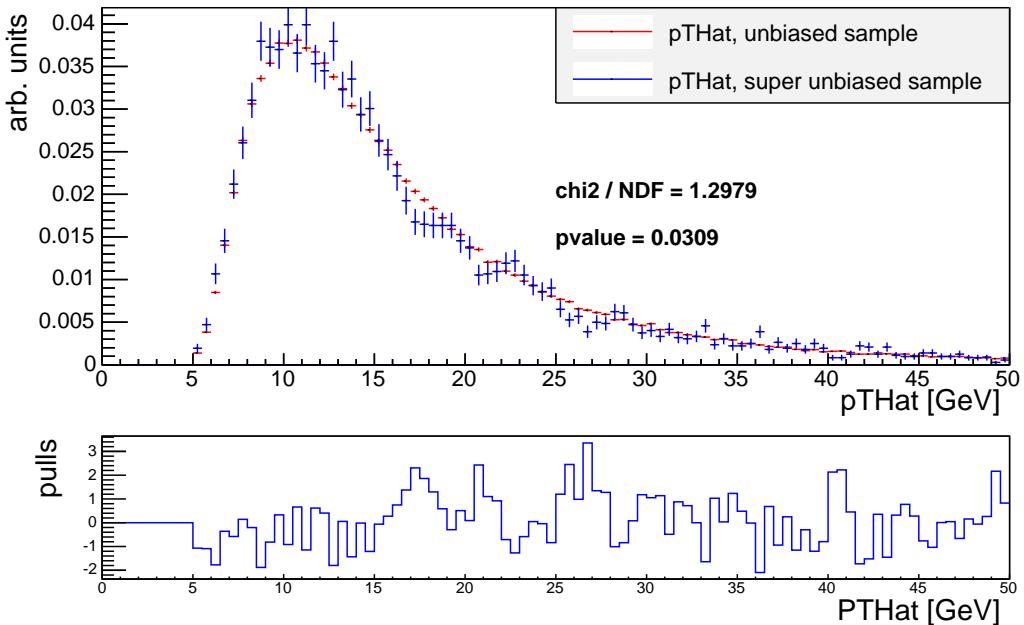


Figure 34: Plot show the comparison of the \hat{p}_T distribution from the unbiased sample and the \hat{p}_T distribution from a new sample, super unbiased, with the same quark-level cuts as the unbiased but with looser \hat{p}_T (3 GeV insted of 5). The super unbiased distribution has been cut at 5 GeV in order to compare the two distributions. Quark-level cuts, final state cuts and B fiducial volume cuts ($pT_B > 8$ GeV and $|\eta_B| < 2.5$) are applied to both distributions.

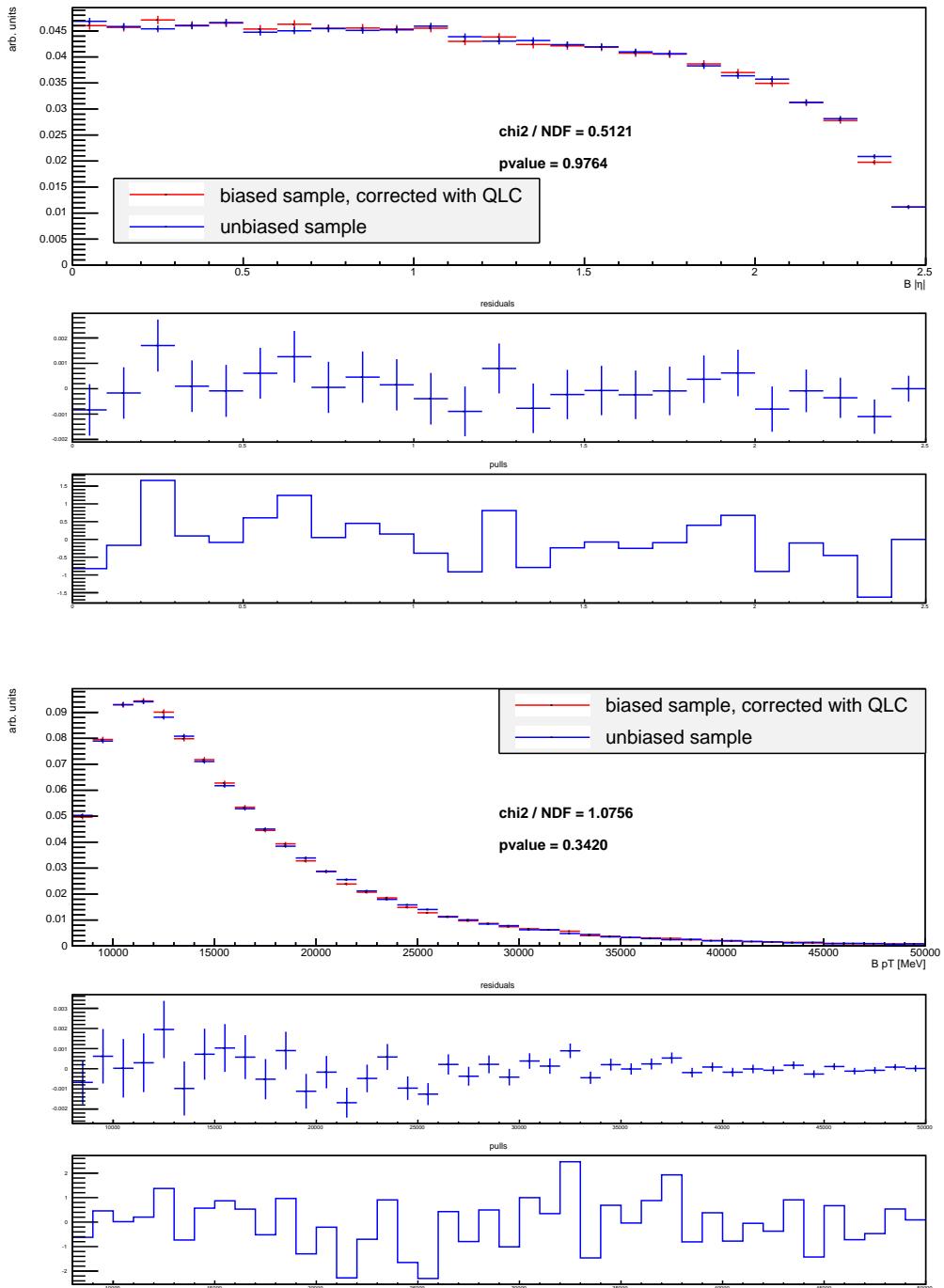


Figure 35: Plots show the comparison of the η_B (first plot) and pT_B (second plot) QLC corrected quark biased distribution using the two samples approach QLC and the unbiased distribution. In order to avoid correlations between the distributions, QLC have been calculated using odd numbered events from the unbiased and quarkBiased samples, and the remaining events in the two samples have been weighted and used for the comparison.

1213 B DDW weights

1214 This appendix provides additional information on the DDW calculation. The SideBand Weights (SBW)
 1215 extraction is discussed in B.1 for both the $B^\pm \rightarrow J/\psi K^\pm$, B.1.1, and the $B_s \rightarrow J/\psi \phi$, B.1.2, processes.
 1216 In B.2 a detailed discussion on the DDW calcualtion approach is reported and in B.3 a set of checks on
 1217 the DDW are reported.

1218 B.1 SBW weights

1219 The number of signal events in data for the DDW calculation is obtained through the sideband subtraction
 1220 procedure, following the approach used in the Run1 analysis [4].
 1221 The number of signal and background events is esimated with a binned extended maximum likelihood fit
 1222 on the mass distribution of the B meson in different pT_B and η_B bins to calculate the SideBand Weights
 1223 (SBW).
 1224 The SBW are per event weights defined as 1 if the mass of the B candidate falls into a defined region
 1225 around the mass peak (signal region), while if it falls into the regions outside the mass peak (left and right
 1226 sidebands), the SBW are defined as:

$$W_{SBW} = -1 \times \frac{\int_{\text{signal region}} PDF_{bkg}}{\int_{\text{left sideband}} PDF_{bkg} + \int_{\text{right sideband}} PDF_{bkg}} \quad (7)$$

1227 where PDF_{bkg} is the probability distribution function associated to the background.
 1228 The SBW are calculated for both the $B^\pm \rightarrow J/\psi K^\pm$ and $B_s^0 \rightarrow J/\psi \phi$ processes, the fit on the data is
 1229 modified accordingly to the features of the dataset and the definition of the signal and sideband regions is
 1230 modified according to the position and the width of the mass peak.

1231 B.1.1 B^+ SBW weights

1232 For the $B^\pm \rightarrow J/\psi K^\pm$ process, the signal and sideband regions are defined as:

- 1233 • signal region: [5182.96, 5382.96] MeV
- 1234 • left sideband: [5082.96, 5182.96] MeV
- 1235 • right sideband: [5382.96, 5482.96] MeV.

1236 The fit on the data is performed using the following configuration: the signal is described with two
 1237 Gaussians with equal mean; the continuum background is described by an exponential; the background
 1238 due to partially reconstructed decays (B to $J/\psi X$, with $m(J/\psi h^+ < 5.200$ GeV) is described with an
 1239 error function. All shape and amplitude parameters are extracted from the fit, and the uncertainty in the
 1240 signal yield is dominated statistical errors.

1241 Following the procedure applied in the Run1 analysis, the SBW have been calculated separately in pT_B
 1242 and η_B bins, but a dependence of the SBW on η_B was found, so the B^+ SBW have been extracted in
 1243 combined $pT_B - \eta_B$ bins.

1244 Figures 38 show the SBW extracted separately in pT_B and η_B bins. Three linear fits are performed on the
 1245 η_B SBW. The fit performed on the full X axis (red line) range has a low compatibility with the data (p-value
 1246 $\sim 0.00\%$), so two other fits have been performed in the two η regions that show a different behaviour. The

compatibility of these two fits is much higher than the previous (p-value $\sim 9.6\%$ for $\eta < 1.3$, green line, and p-value $\sim 24.9\%$ for $\eta > 1.3$, blue line) so the SBW extraction is performed in two eta bins combined with pT bins.

Figure 37 shows some of the fits performed to extract the B^+ yield and evaluate the SBW.

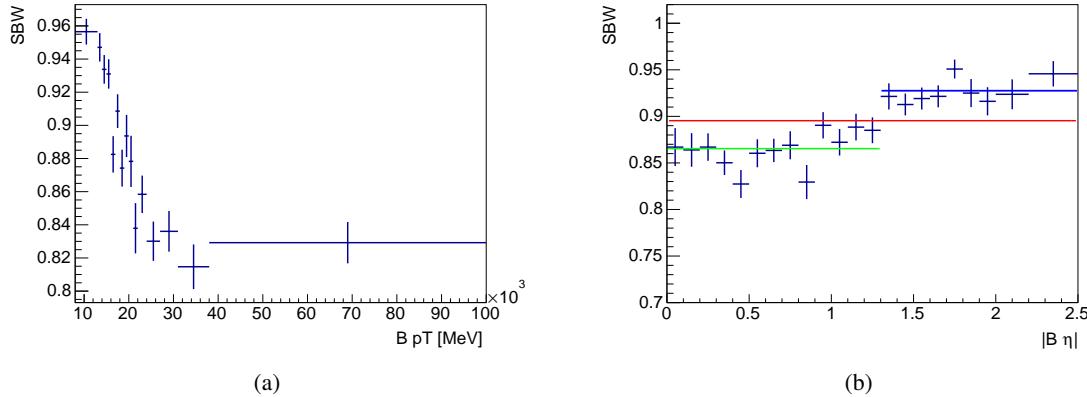


Figure 36: SBW calculated for the $B^\pm \rightarrow J/\psi K^\pm$ process. The SBW have been calculated separately in pT_B 36(a) and η_B 36(b) bins. The dependence of the SBW on η_B has been checked with a linear fit on the SBW, and it does not show a good compatibility with the histogram (red line, p-value $\sim 0.00\%$). Two other fits have been performed, dividing the η range in two, and they show a much higher compatibility, p-value $\sim 9.6\%$ for $\eta < 1.3$, green line, and p-value $\sim 24.9\%$ for $\eta > 1.3$, blue line. The SBW extraction is therefore performed in combined pT- η bin. In particular two eta bins are chosen, following the behaviour of 36(b).

The result of the SBW extraciton is shown in figure 38.

B.1.2 B_s^0 SBW weights

For the $B_s \rightarrow J/\psi \phi$ process, the signal and sideband regions are defined as:

- signal region: [5286, 5456] MeV
- left sideband: [5201, 5286] MeV
- right sideband: [5456, 5541] MeV.

The fit on the data is performed using the following configuration: the signal is described with two Gaussians with equal mean and the continuum background is described with a third order Chebyshev polynomial. All shape and amplitude parameters are extracted from the fit, and the uncertainty in the signal yield is dominated statistical errors.

Also in this case the SBW have been firstly calculated separately in pT_B and η_B bins. Figure 39 shows the resulting SBW, in this case the linear fit performed on the η_B SBW shows compatibility with the data (p-value $\sim 60\%$). For the $B_s \rightarrow J/\psi \phi$ process the sideband subtraction procedure is therefore independent on the specific η bin chosen, so only the pT SBW will be used in the DDW calculation. Figure 40 shows some of the fits performed to extract the $B_s \rightarrow J/\psi \phi$ yield and evaluate the SBW.

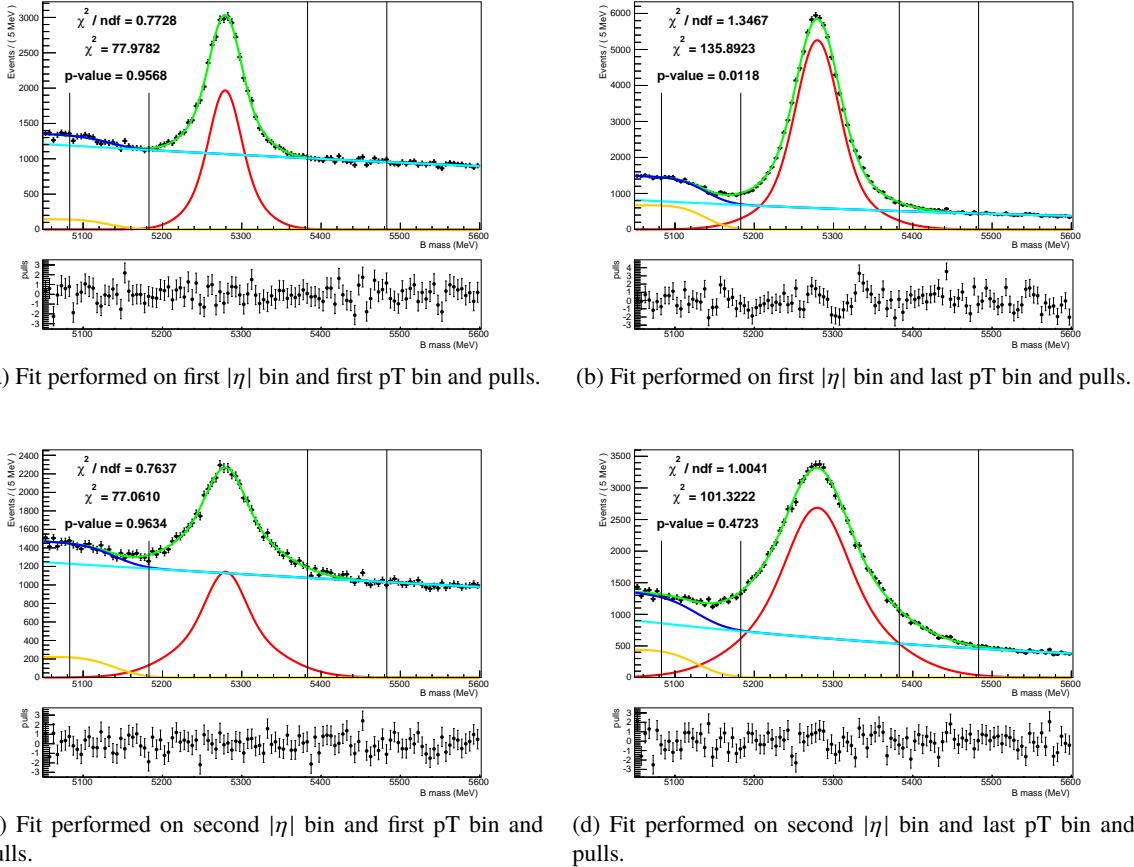


Figure 37: The plots show some of the fits performed on the B^+ sample. Each plot shows the data points (black dots) and the PDFs used to perform the fit: the red line is the signal PDF, the light blue line is the PDF for the combinatorial background, the yellow line is the PDF for the partially reconstructed decays, the blue line is the total background PDF and the green line is the total PDF. Each plot shows also the compatibility of the data with the fitted curve and the pulls. The vertical lines show the sideband and signal regions.

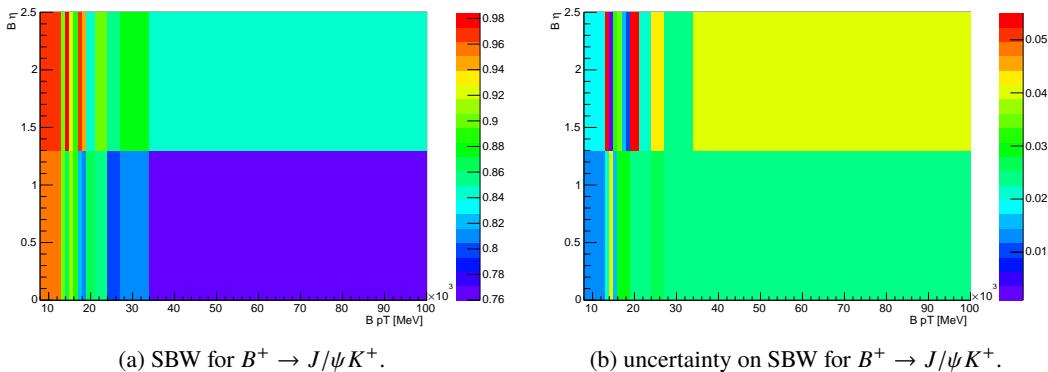


Figure 38: Plots show the sideband weights calculated for the $B^+ \rightarrow J/\psi K^+$ 38(a) and their uncertainty 38(b).

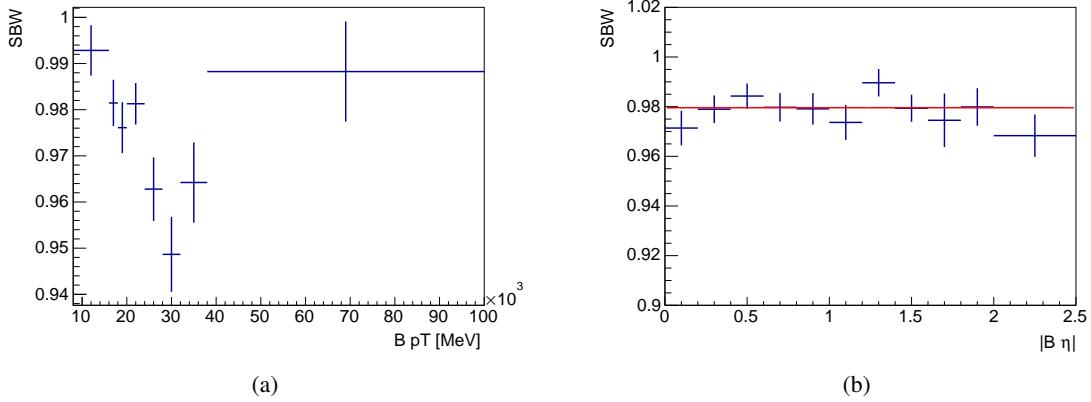


Figure 39: SBW calculated for the $B_s \rightarrow J/\psi\phi$ process. The SBW have been calculated separately in pT_B 39(a) and η_B 39(b) bins. The dependence of the SBW on η_B has checked with a linear fit on the SBW, that shows a good compatibility with the histogram (p-value $\sim 60\%$) proving that the SBW for this process don't depend on η .

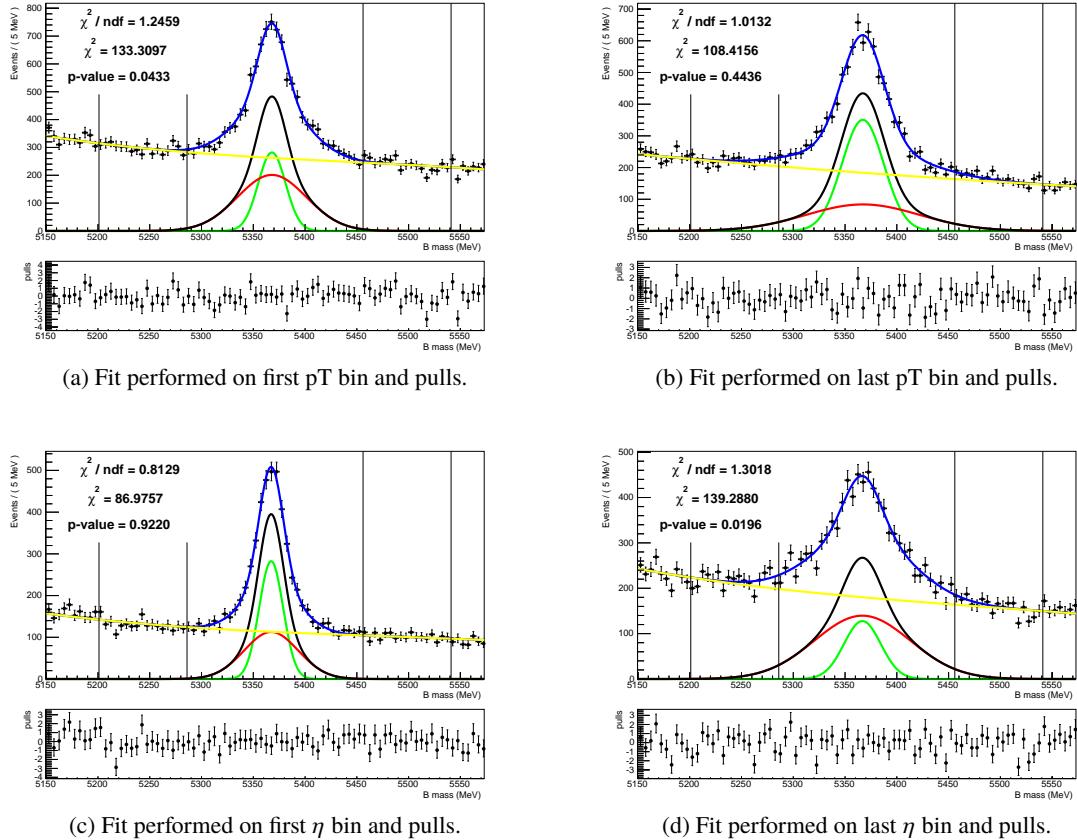


Figure 40: The plots show some of the fits performed on the $B_s \rightarrow J/\psi\phi$ sample. Each plot shows the data points (black dots) and the PDFs used to perform the fit: the green and red lines are the two gaussians for the signal PDF, while the black line is the total signal PDF, the yellow line is the PDF for the combinatorial background and the blue line is the total PDF. Each plot shows also the compatibility of the data with the fitted curve and the pulls. The vertical lines show the sideband and signal regions.

1267 B.2 DDW weights extraction

1268 The DDW calcualtion follows the same line as in [6] and [4]. Given the low statistics available, the weights
 1269 are extracted with a one-dimensional times one-dimensional iterative method, by comparing sideband-
 1270 subtracted $B^\pm \rightarrow J/\psi K^\pm$ odd(even)-numbered data events with odd(even)-numbered MC events, after the
 1271 latter are corrected with the QLC weights. The sideband subtraction procedure is also performed on the
 1272 MC, in order to avoid possible biases due to the leakage of signal events in the sidebands.

1273 The DDW calculation proceeds according to the following formulas:

$$W_{p_T}^1 = \frac{\sum_\eta D_{\eta,p_T}}{\sum_{\eta,p_T} D_{\eta,p_T}} \frac{\sum_{\eta,p_T} MC_{\eta,p_T}}{\sum_\eta MC_{\eta,p_T}} \quad (8)$$

$$W_{p_T}^n = \frac{\sum_\eta D_{\eta,p_T}}{\sum_{\eta,p_T} D_{\eta,p_T}} \frac{\sum_{\eta,p_T} (MC_{\eta,p_T} \prod_{m=1}^{m=n-1} W_{p_T}^m W_\eta^m)}{\sum_\eta (MC_{\eta,p_T} \prod_{m=1}^{m=n-1} W_{p_T}^m W_\eta^m)} \quad (9)$$

1274 Where equation 8 is the formula for the first iteration and equation 9 is the formula for the n_{th} iteration.
 1275 MC_{η,p_T} refers to a particular η, p_T bin of the MC ditribution after the application of the SBW and the QLC
 1276 weights, while D_{η,p_T} refers to a η, p_T bin of the data distribution after the sideband subtraction.

1277 Equivalent formulas hold for calculating the η weights.

1278 Due to the nature of the iterative procedure, combined weights are $W_{p_T} = \prod_{m=1}^{m=3} W_{p_T}^m$ for p_T and
 1279 $W_\eta = \prod_{m=1}^{m=3} W_\eta^m$ for η , in the hypotesys of a convergence after the third iteration.

1280 The final weights that are going to be used are therefore

$$W_{DDW}(p_T, |\eta|) = W_{p_T} \cdot W_\eta \quad (10)$$

1282 Figures 41(c) and 41(d) show the first 4 iterations for the p_T and η weights for B^+ , the procedure converges
 1283 at the second iteration, as the third iteration is compatible with 1. Figures 41(a) and 41(b) show the
 1284 combination of the first and second iteration.

1285 B.3 Checks on DDW

1286 As was performed with the QLC weights, the DDW calculation has been cross-checked applying DDW
 1287 calculated using odd MC and data events to the even MC events and comparing the result with the even
 1288 numbered events data.

1289 Both the odd and even MC distributions have been firstly weighted using the QLC weigths, in order to
 1290 apply the complete procedure; the sideband subtraction procedure is also applied to both data and MC.
 1291 Figure 42 shows the result of this cross-check applied to the $B^+ \rightarrow J/\psi K^+$ DDW and the weighted MC
 1292 distribution is compatible with the data distribution.

1293 A similar result is otained switching odd and even numbered events.

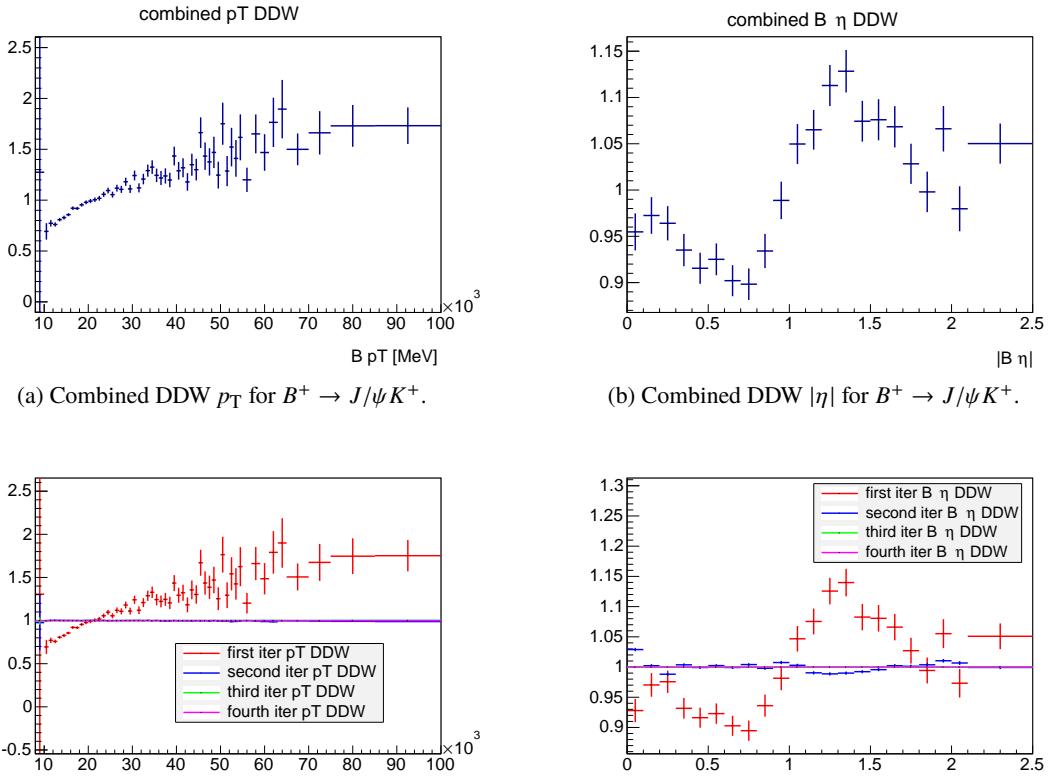


Figure 41: Result of the calculation of the DDW for the $B^+ \rightarrow J/\psi K^+$ process. Figures 7(a), 7(b) show the combination of the first two iterations of the DDW calculation, while figures 41(c), 41(d) show the first four iterations of the DDW calculation. The third and fourth iterations are compatible with one therefore the weights calculation is interrupted after the second iteration.

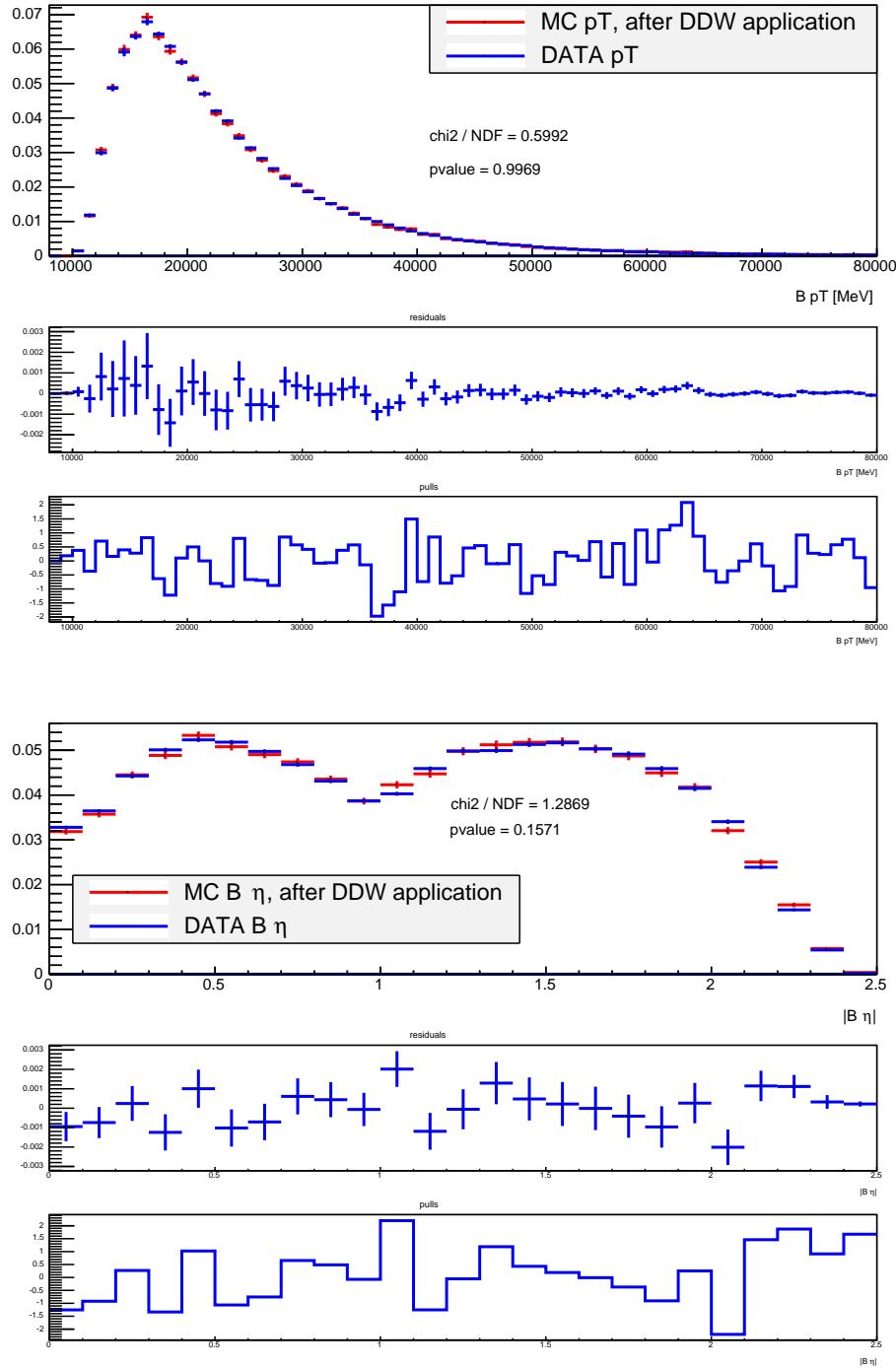
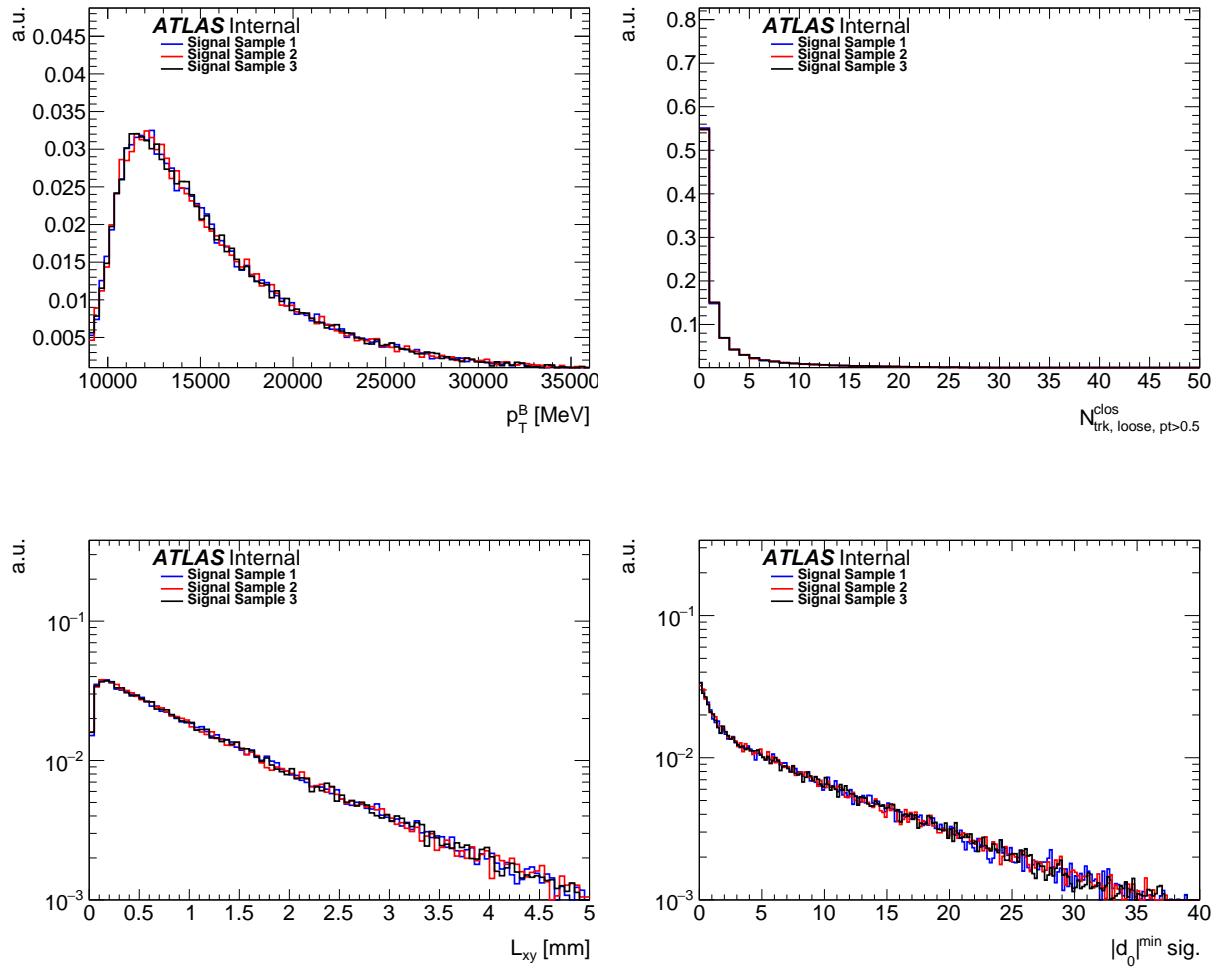
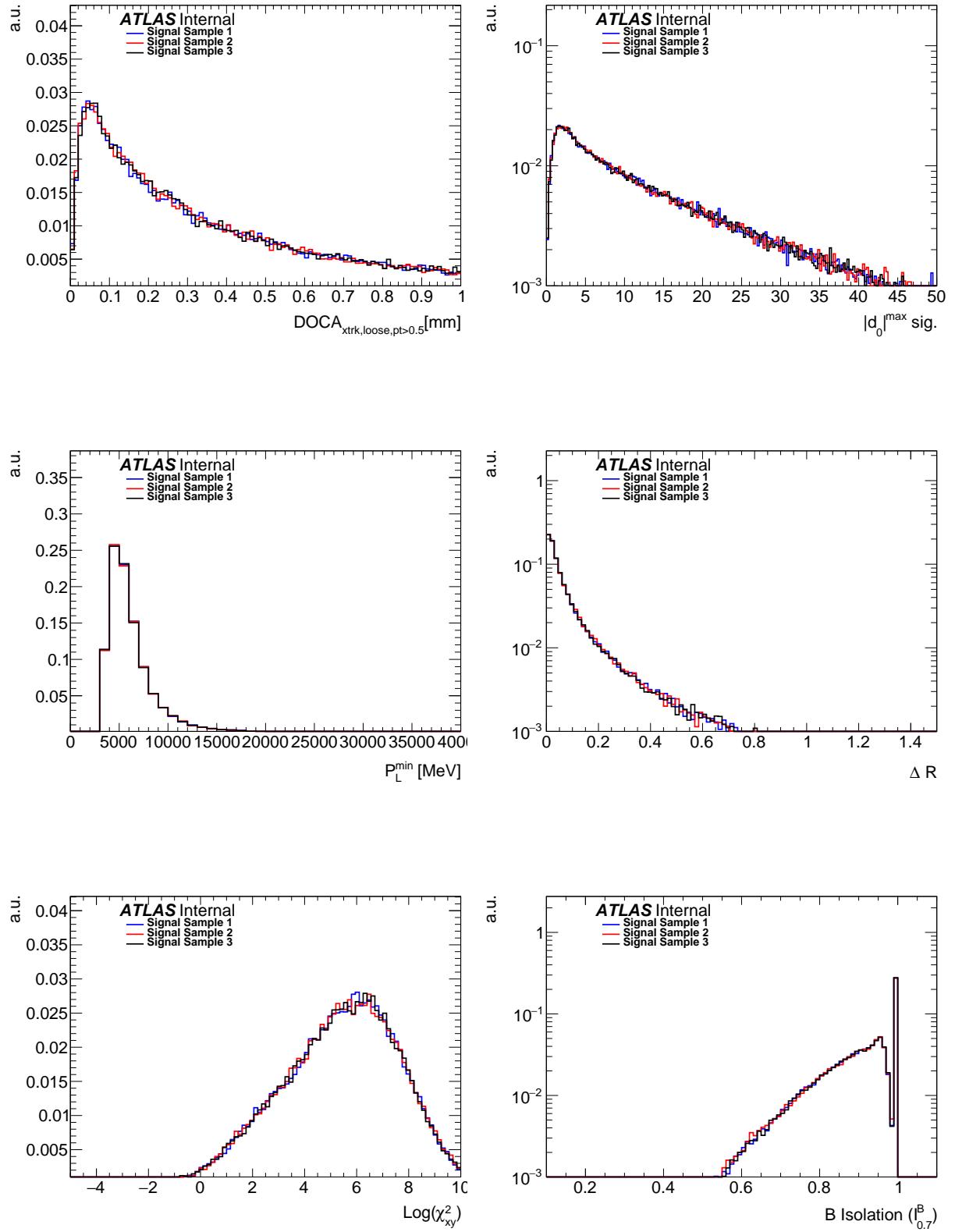


Figure 42: Plots show the comparison of $\eta(B^+)$ (first plot) and $p_T(B^+)$ (second plot) of the even numbered data events and even numbered MC events corrected using QLC and DDW, with the DDW calculated using odd MC and data events. The sideband subtraction procedure has been applied to both MC and data.

1294 C Data-MC comparisons in the sidebands

1295 The first set of figures in this section compare the 3 BDT signal samples and the second set compares the
 1296 3 BDT background samples.





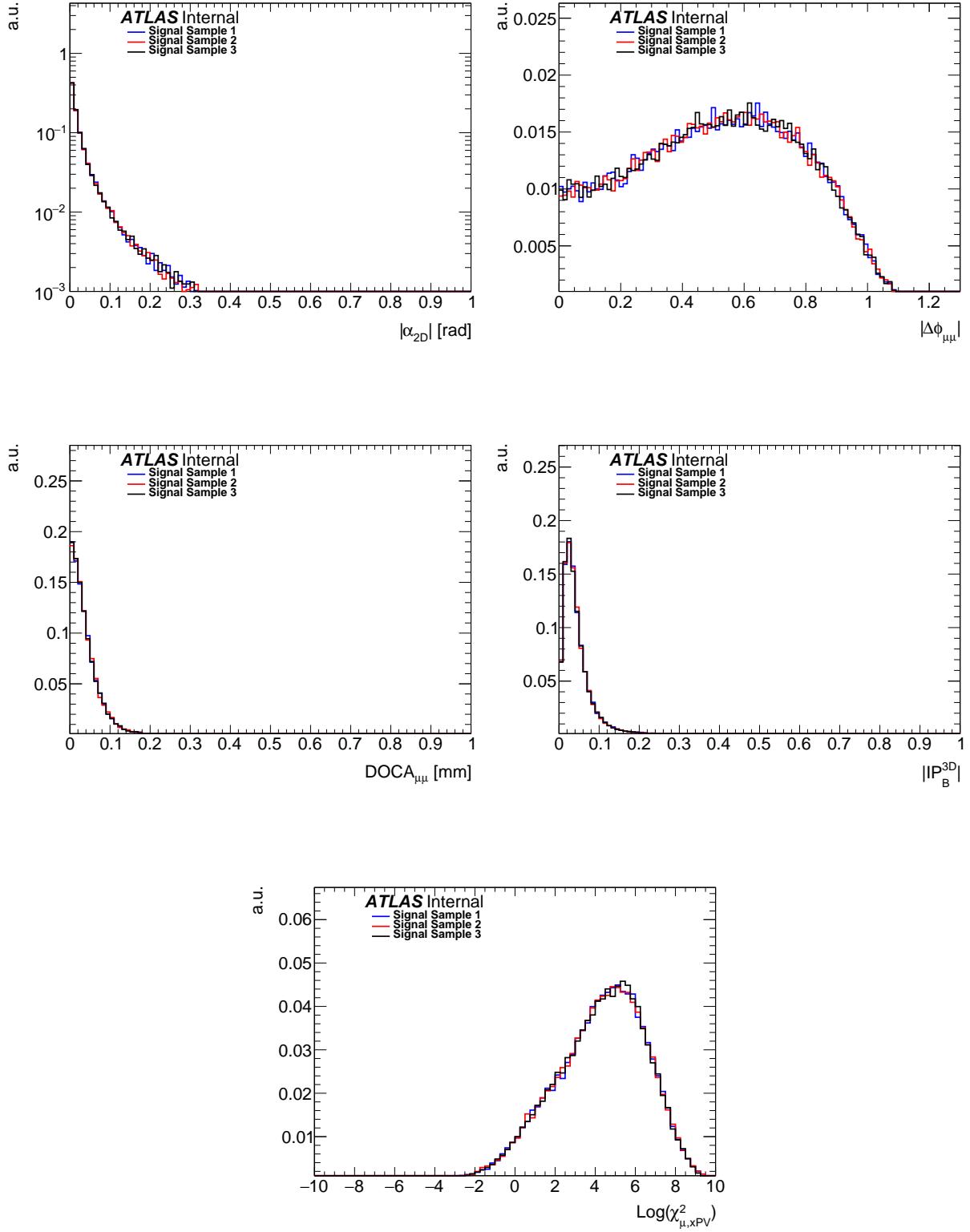
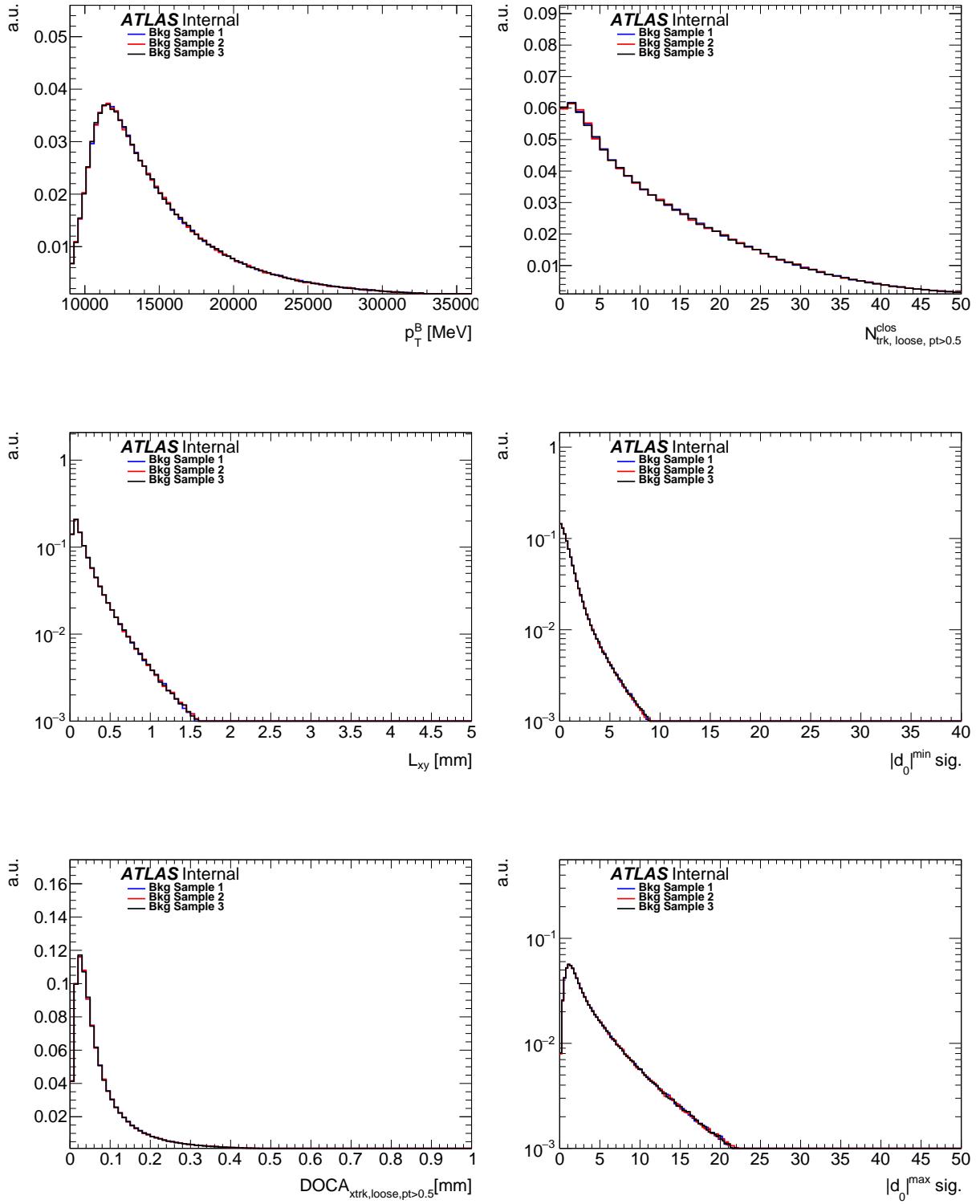
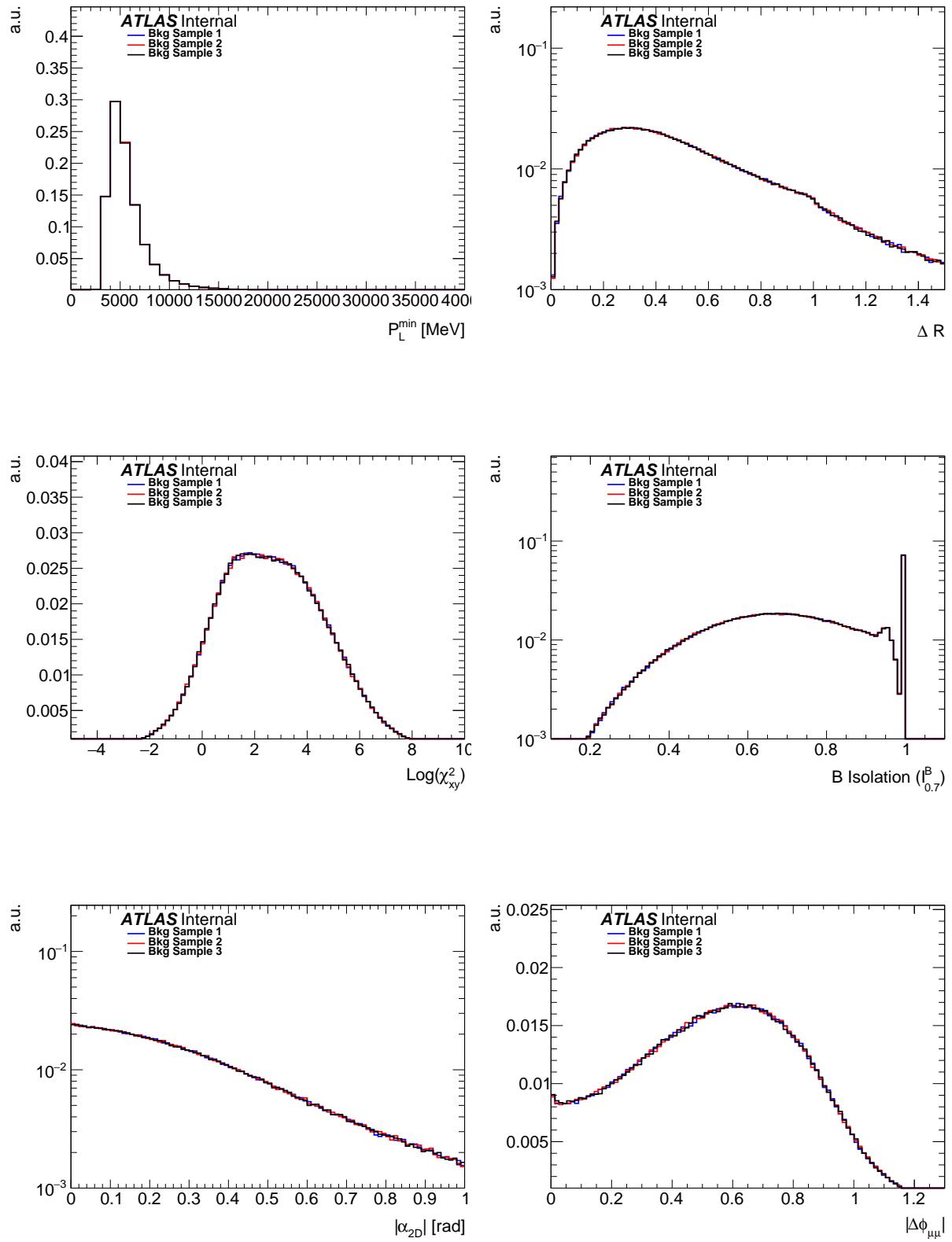


Figure 43: These 15 figures compare the 3 BDT signal samples.

1297 The following 15 figures compare the 3 background samples.





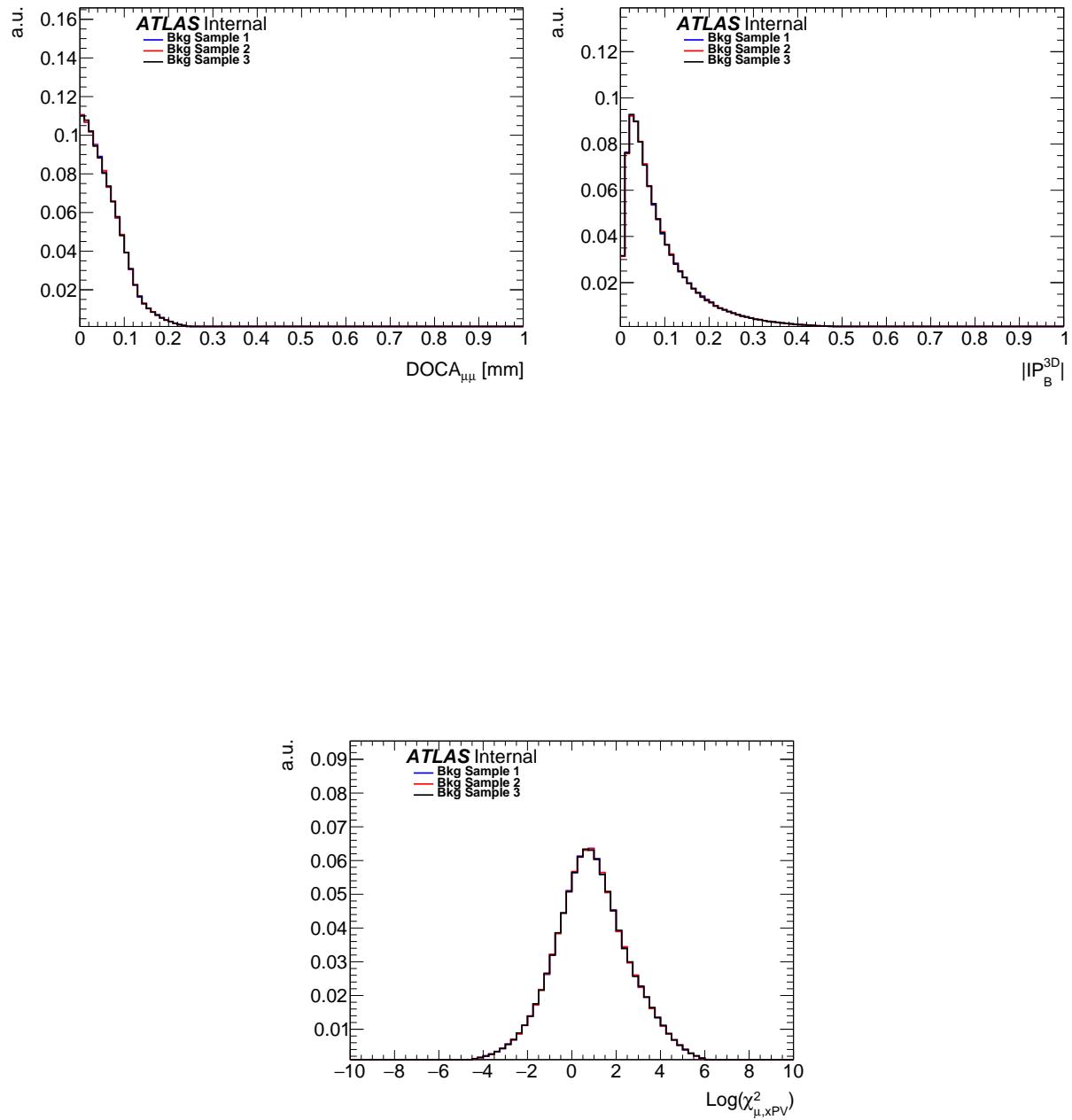


Figure 44: These 15 figures compare the 3 BDT background samples.

1298 D Data-MC comparisons in the reference channel $B^+ \rightarrow J/\psi K^+$

- 1299 This appendix integrates Section 11.1, where the shapes of distributions of the discriminating variables
 1300 used to separate out the combinatorial background are compared in data and MC samples for the reference
 1301 channel.
- 1302 In addition on the comparisons in figure 18, the distributions of all the other variables are included here.
 1303 Since the B^+ vertex is 3-prong, certain variables (e.g. the secondary vertex resolution) are not directly
 1304 equivalent to their counterparts in $B_x \rightarrow \mu\mu$. In these cases we repeat the vertex fit in data and MC also
 1305 for the J/ψ in $B^\pm \rightarrow J/\psi K^\pm$, using only the two muons in the vertex.

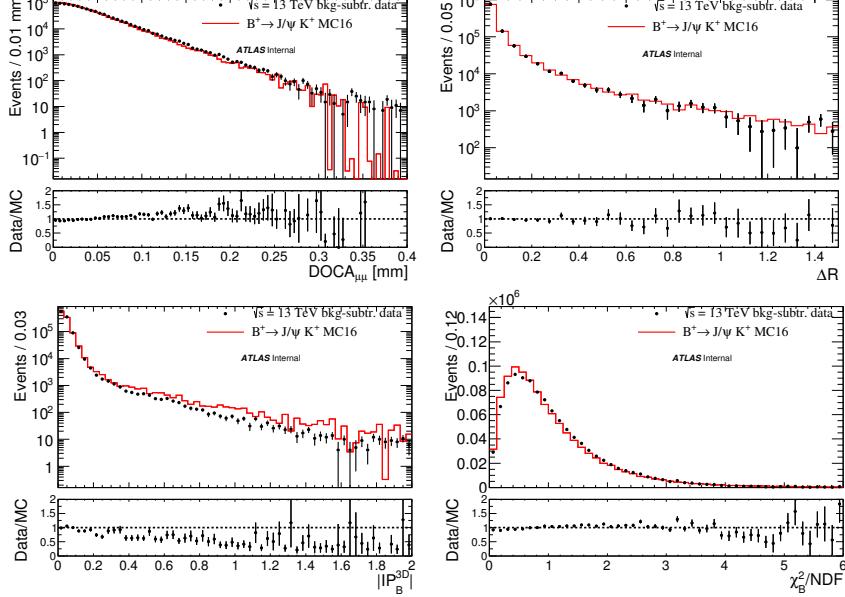


Figure 45: The black dots correspond to the sideband subtracted data, while the red histograms correspond to reweighted $B^+ \rightarrow J/\psi K^+$ MC events normalised to the number of data events.

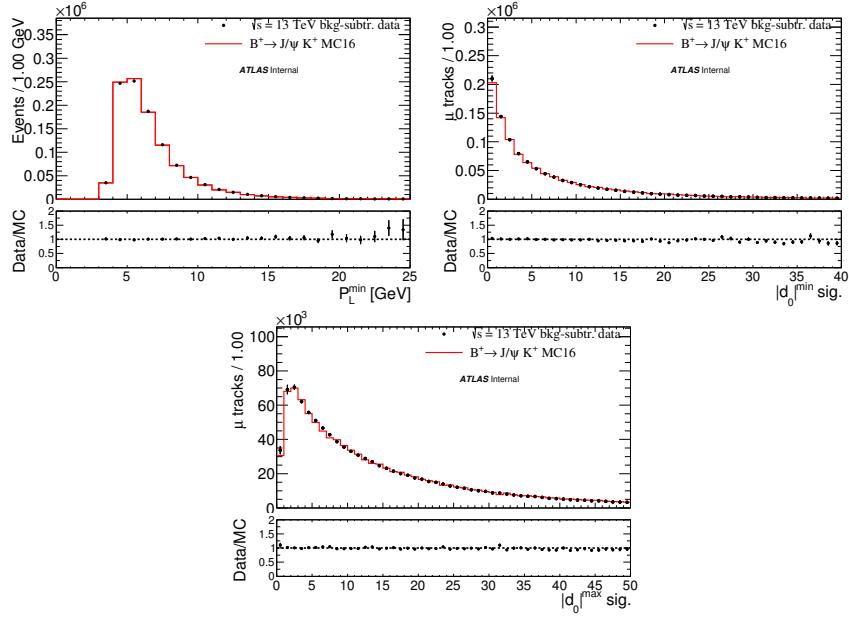


Figure 46: The black dots correspond to the sideband subtracted data, while the red histograms correspond to reweighted $B^+ \rightarrow J/\psi K^+$ MC events normalised to the number of data events.

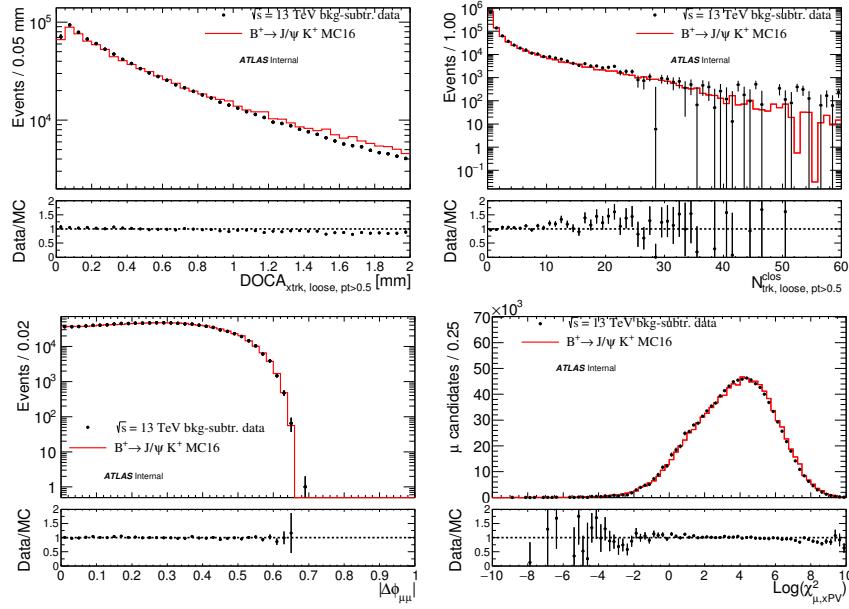


Figure 47: The black dots correspond to the sideband subtracted data, while the red histograms correspond to reweighted $B^+ \rightarrow J/\psi K^+$ MC events normalised to the number of data events.

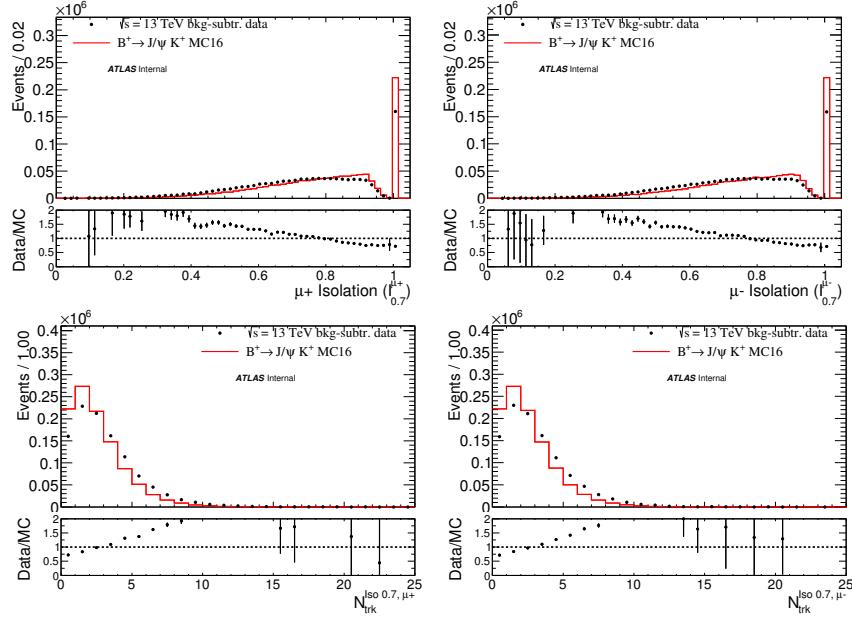


Figure 48: Distributions of isolation and number of tracks in the isolation cone for the μ^+ (plots on the left) and μ^- (plots on the right) forming the B candidates. The black dots correspond to the sideband subtracted data, while the red histograms correspond to reweighted $B^+ \rightarrow J/\psi K^+$ MC events normalised to the number of data events.

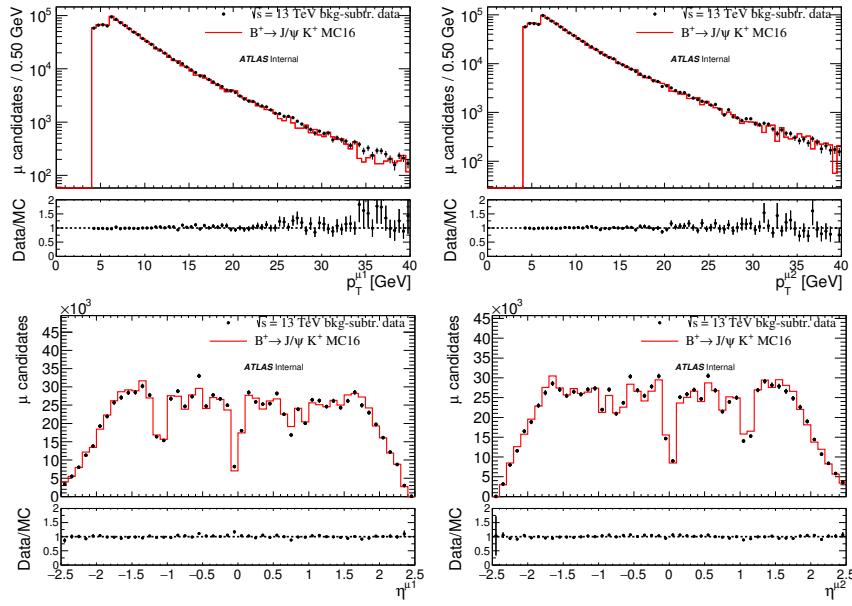


Figure 49: p_T and η distributions for the μ^+ (plots on the left) and μ^- (plots on the right) forming the B candidates. The black dots correspond to the sideband subtracted data, while the red histograms correspond to reweighted $B^+ \rightarrow J/\psi K^+$ MC events normalised to the number of data events.

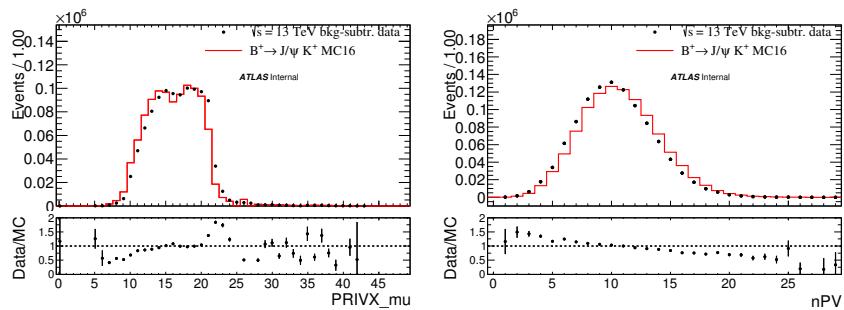


Figure 50: Distributions of the number of interactions per bunch crossing (left) and the number of reconstructed primary vertices (right). The black dots correspond to the sideband subtracted data, while the red histograms correspond to reweighted $B^+ \rightarrow J/\psi K^+$ MC events normalised to the number of data events.

1306 E Data-MC comparisons in the alternate reference channel $B_s^0 \rightarrow J/\psi\phi$

1307 F Continuum BDT

1308 This appendix show the BDT input variable comparisons of the signal and background. Also shown are
 1309 the variable efficiencies for the signal and background samples with 18%, 36%, and 54% signal efficiency
 1310 cut on the BDT variable.

1311 In figure 51 are the correlation matrices for the signal (B to $\mu\mu$) and background (high-mass sideband)
 1312 samples.

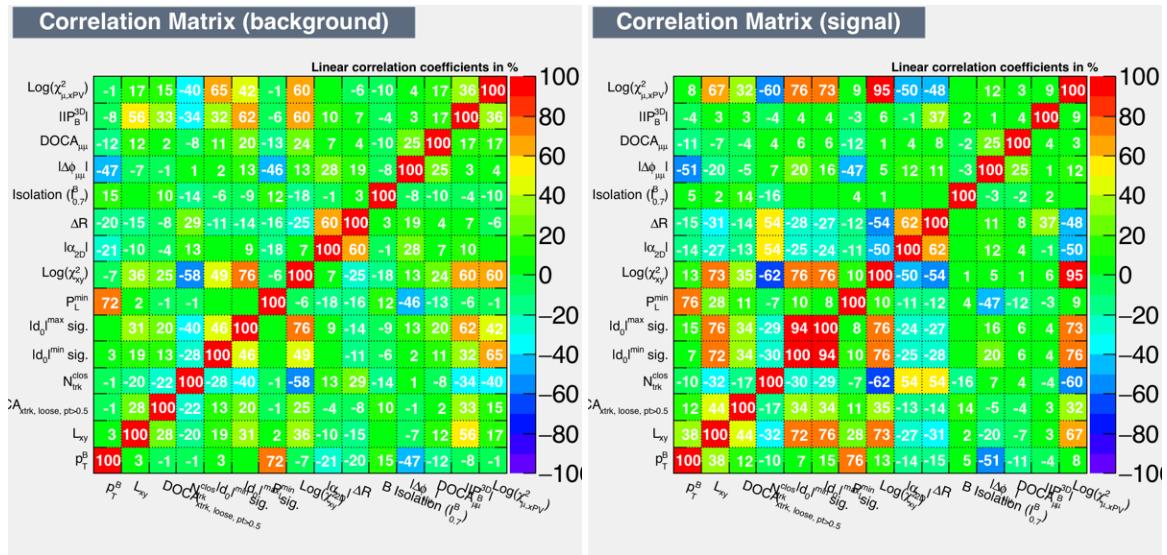
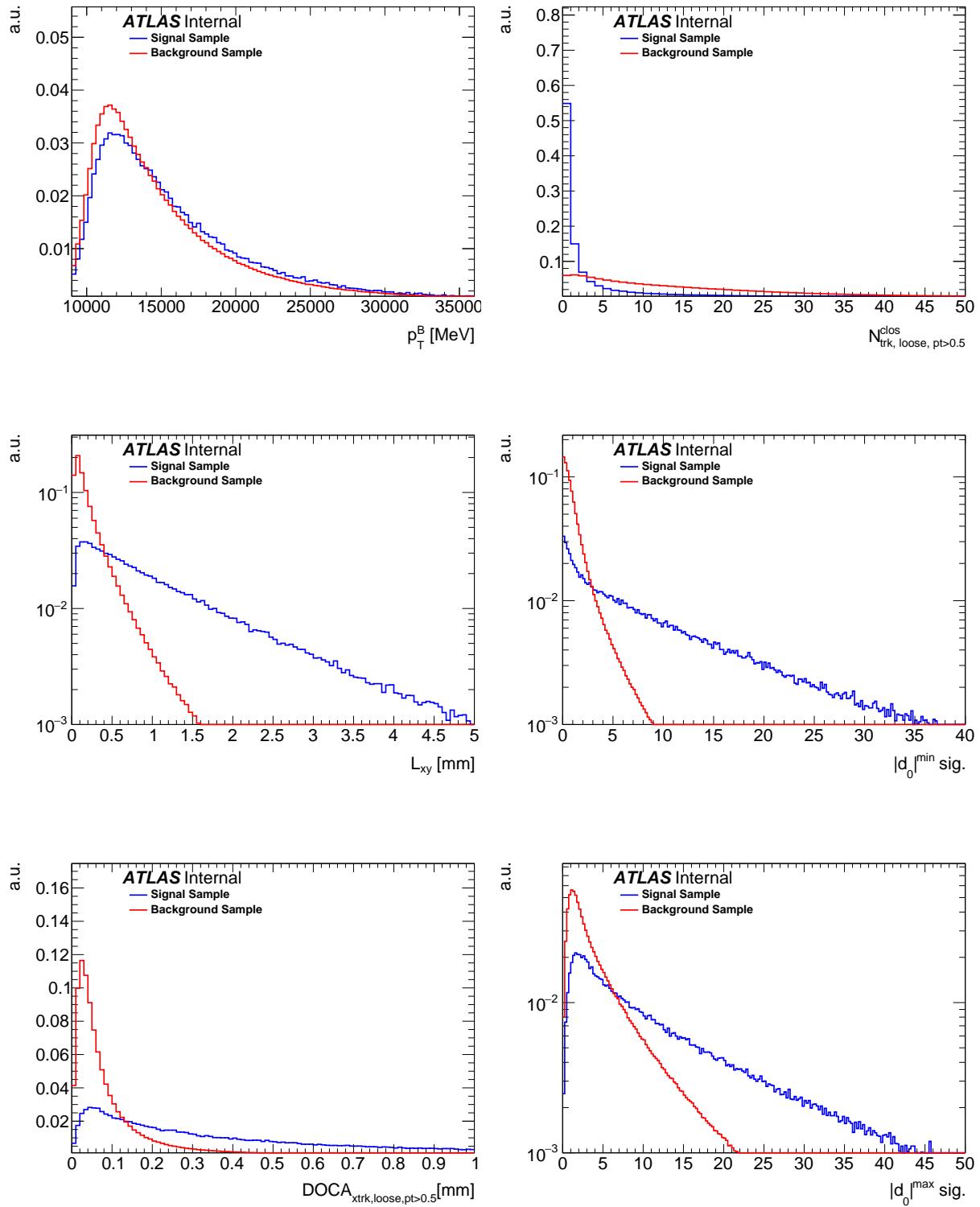
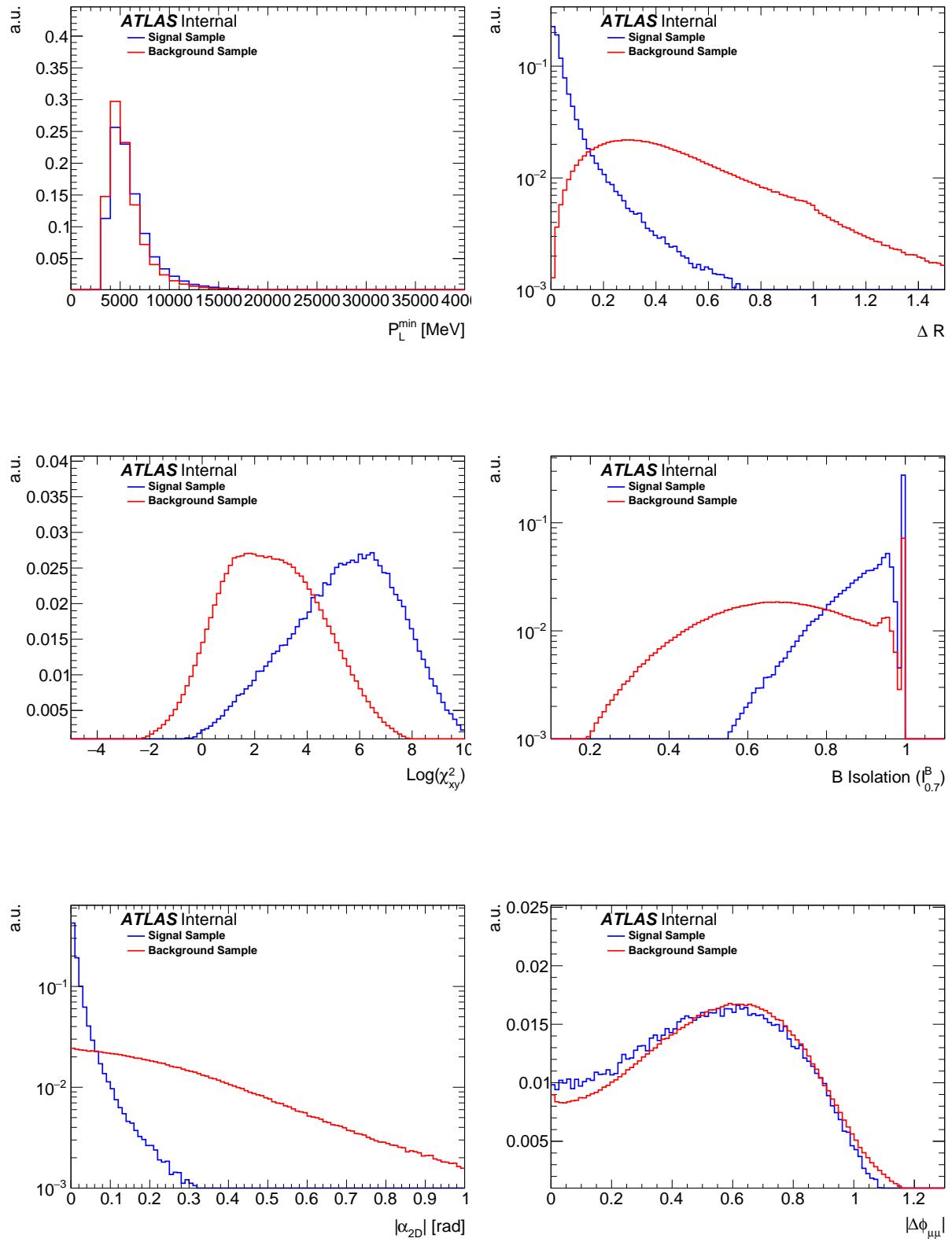


Figure 51: On the left, is the correlation matrix for the input variable to the BDT from the background sample. On the right, is the correlation matrix for the signal sample.

1313 In the following 15 figures are the variable comparisons of the signal and background.





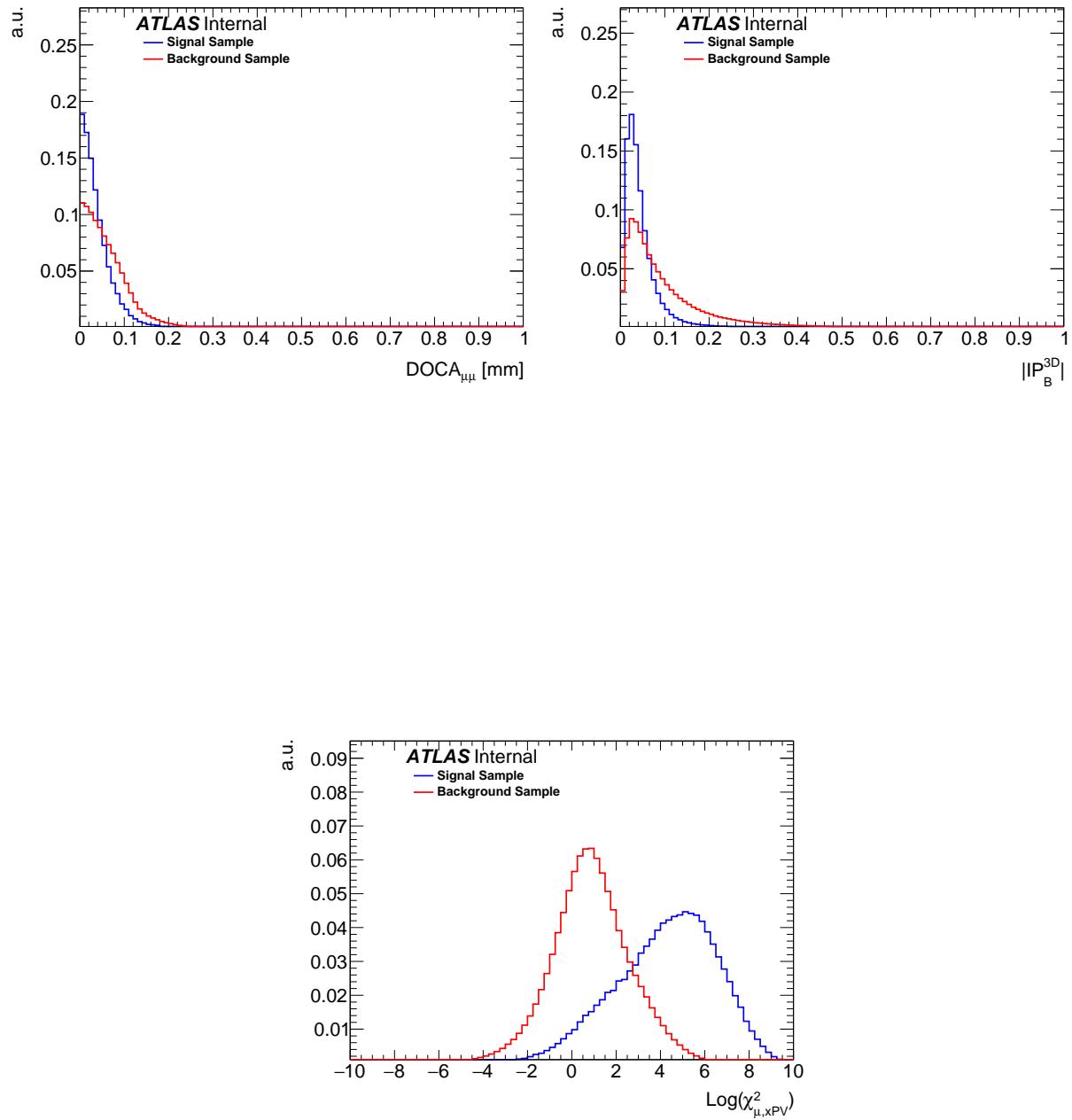
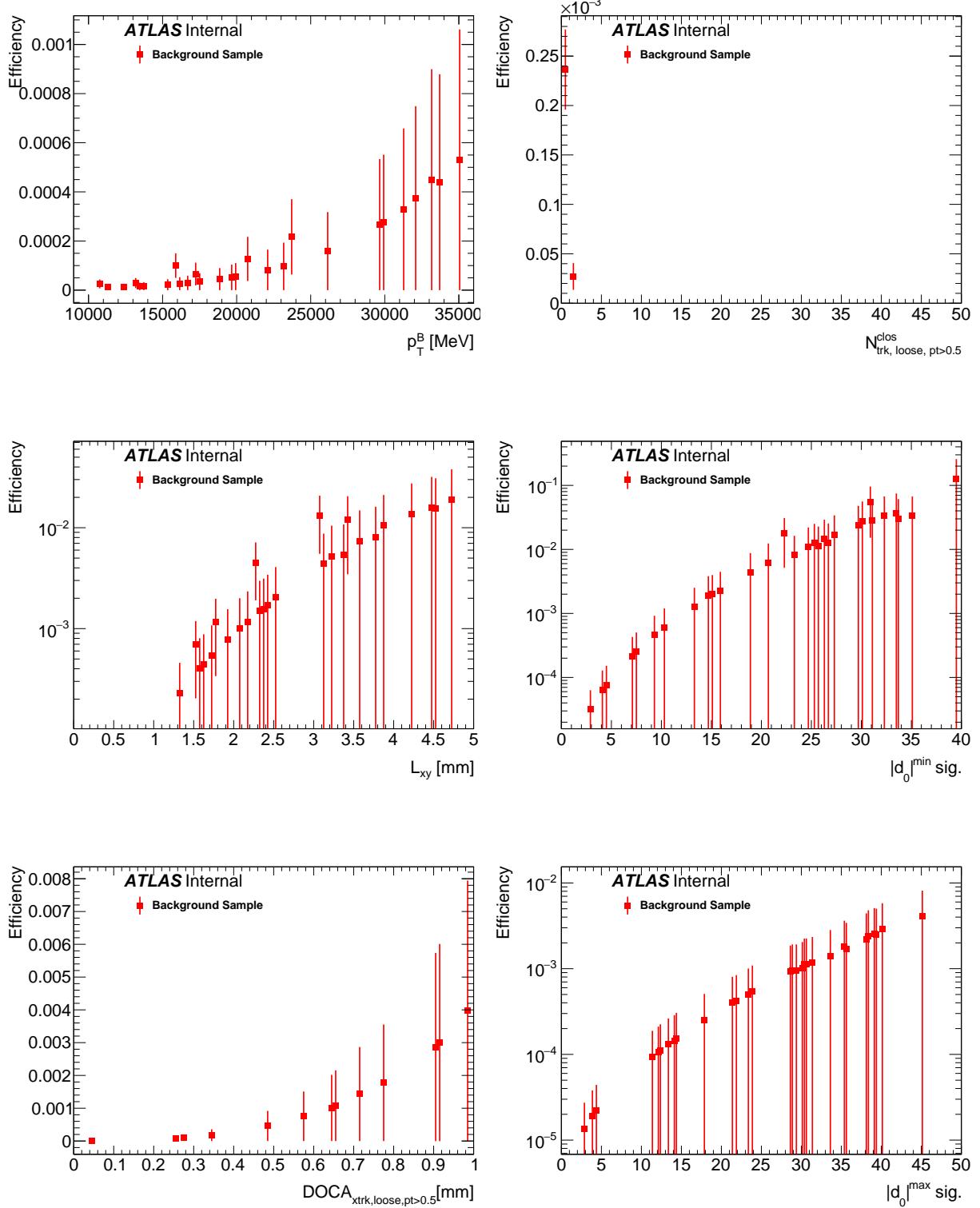
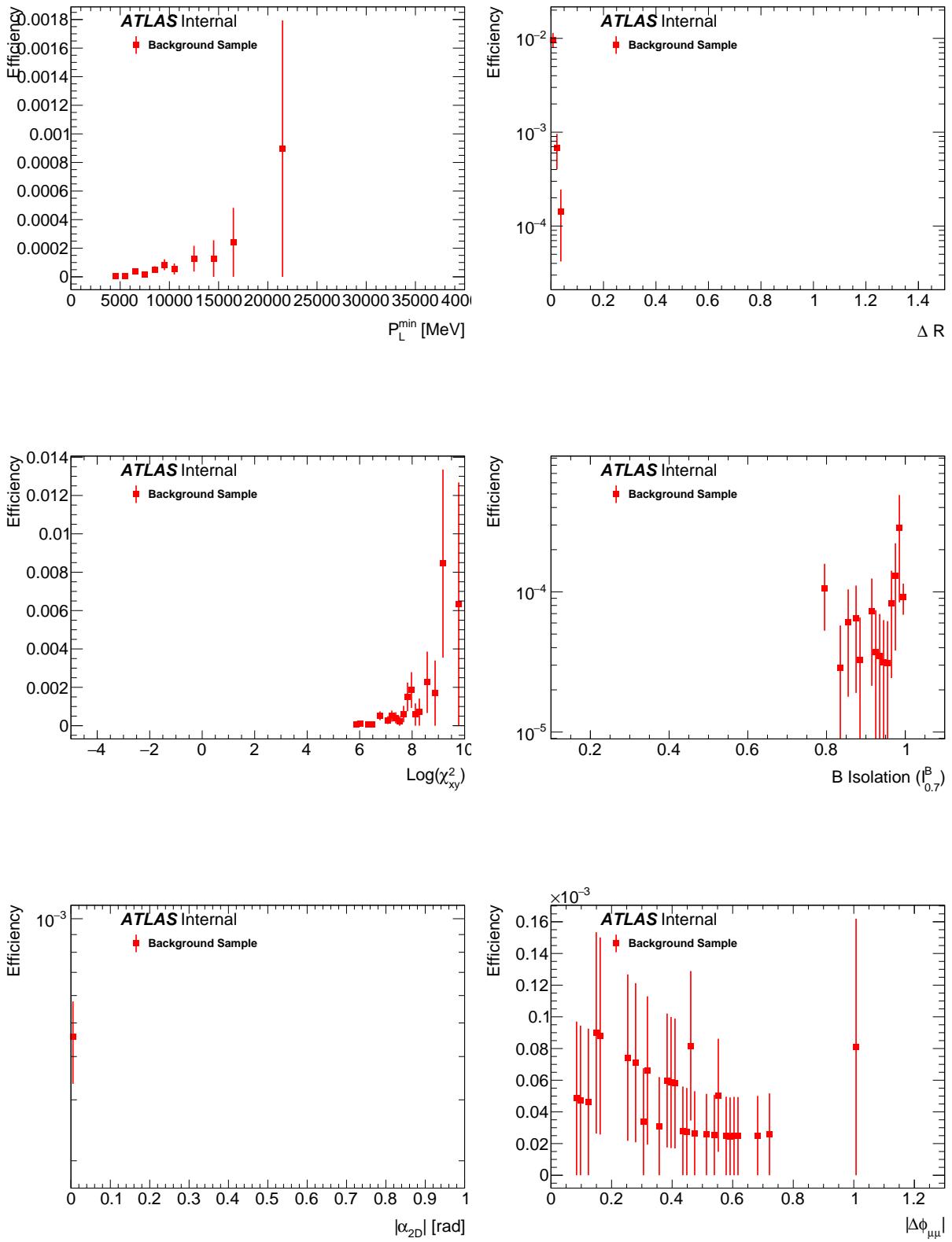


Figure 52: These 15 figures show the BDT input variable comparisons of the signal and background.

1314 The following figures show the variable efficiencies for the background sample with an 18% signal
 1315 efficiency cut on the BDT variable.





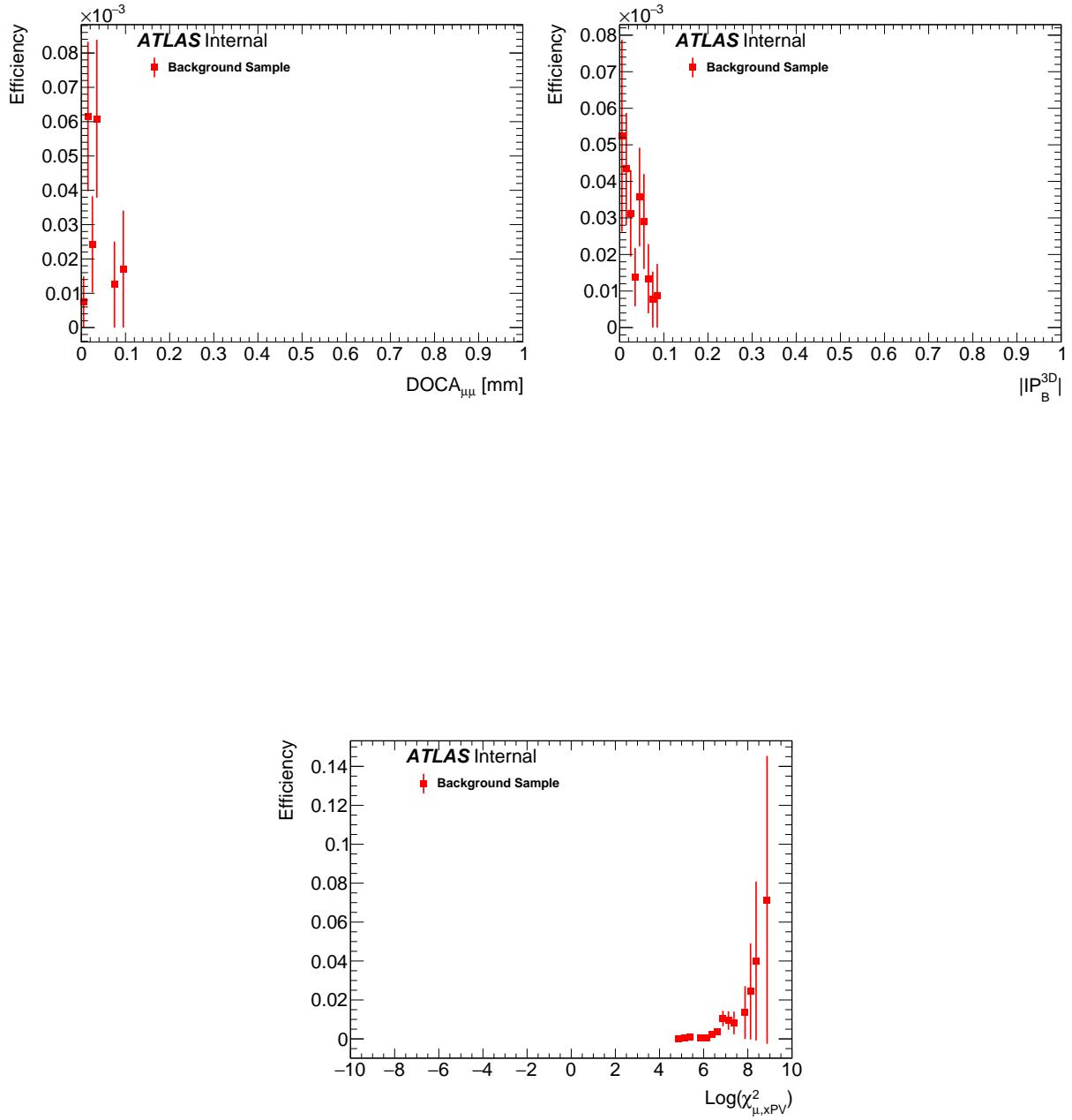
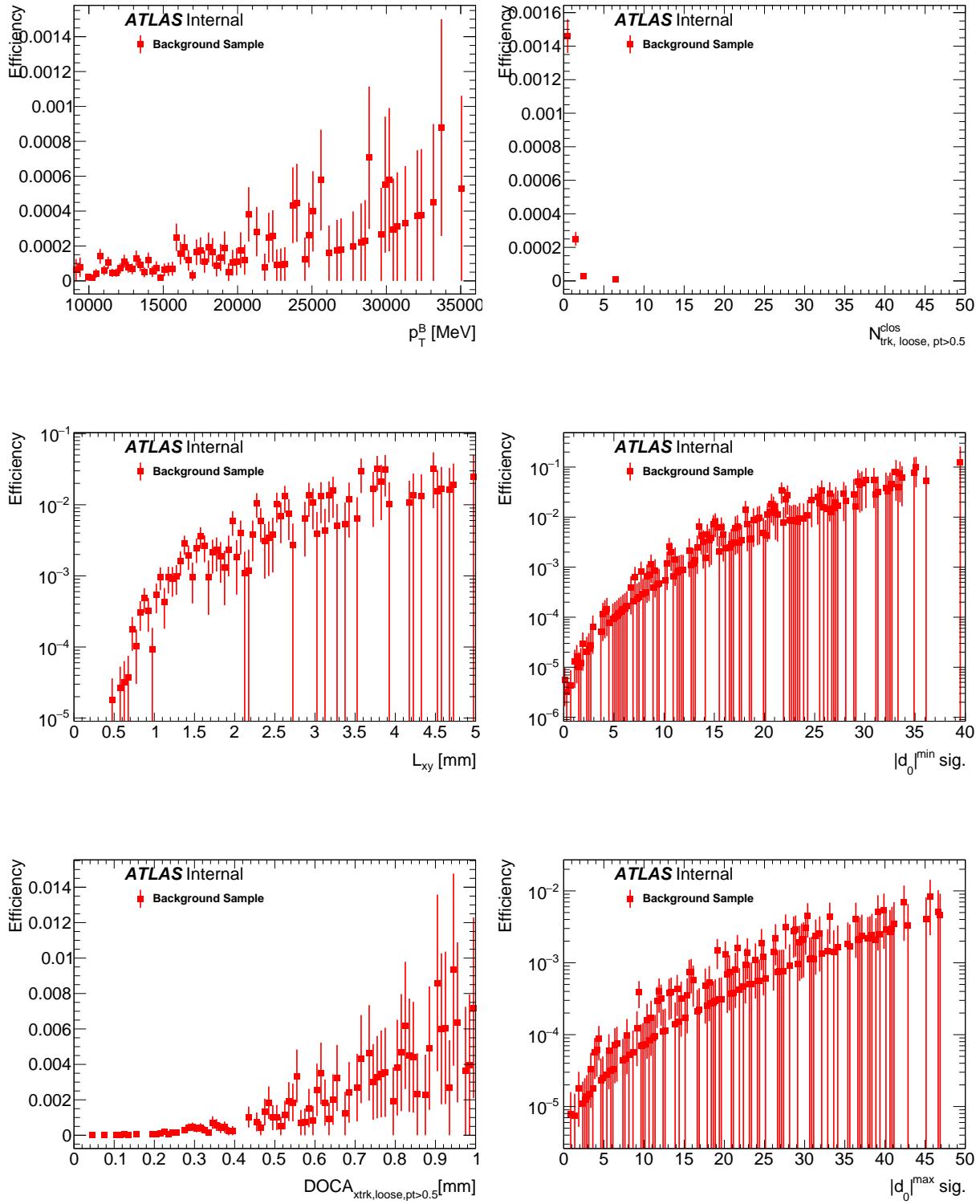
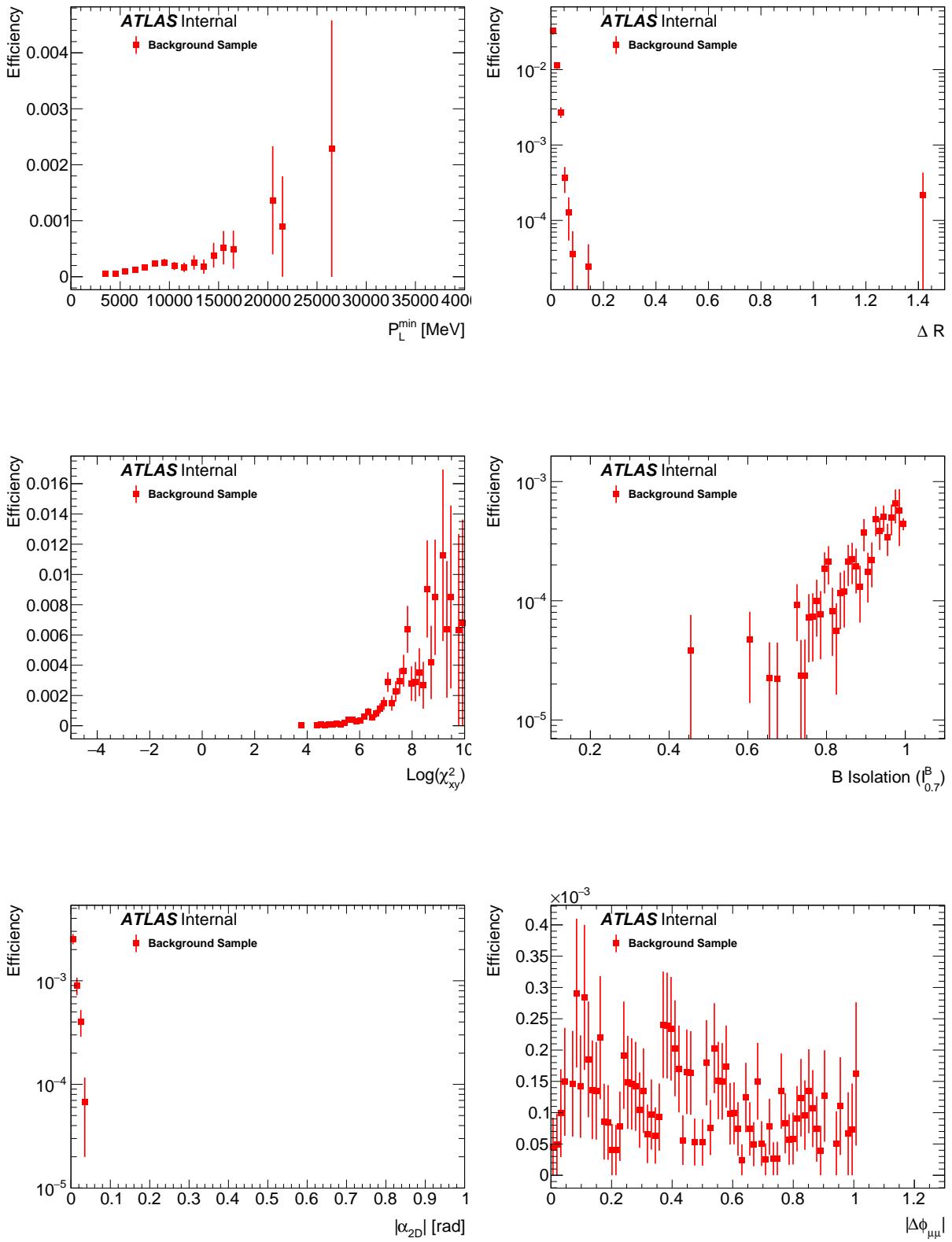


Figure 53: These 15 figures show the the variable efficiencies for the background sample with an 18 % signal efficiency cut on the BDT variable.

1316 The following figures show the variable efficiencies for the background sample with an 36% signal
 1317 efficiency cut on the BDT variable.





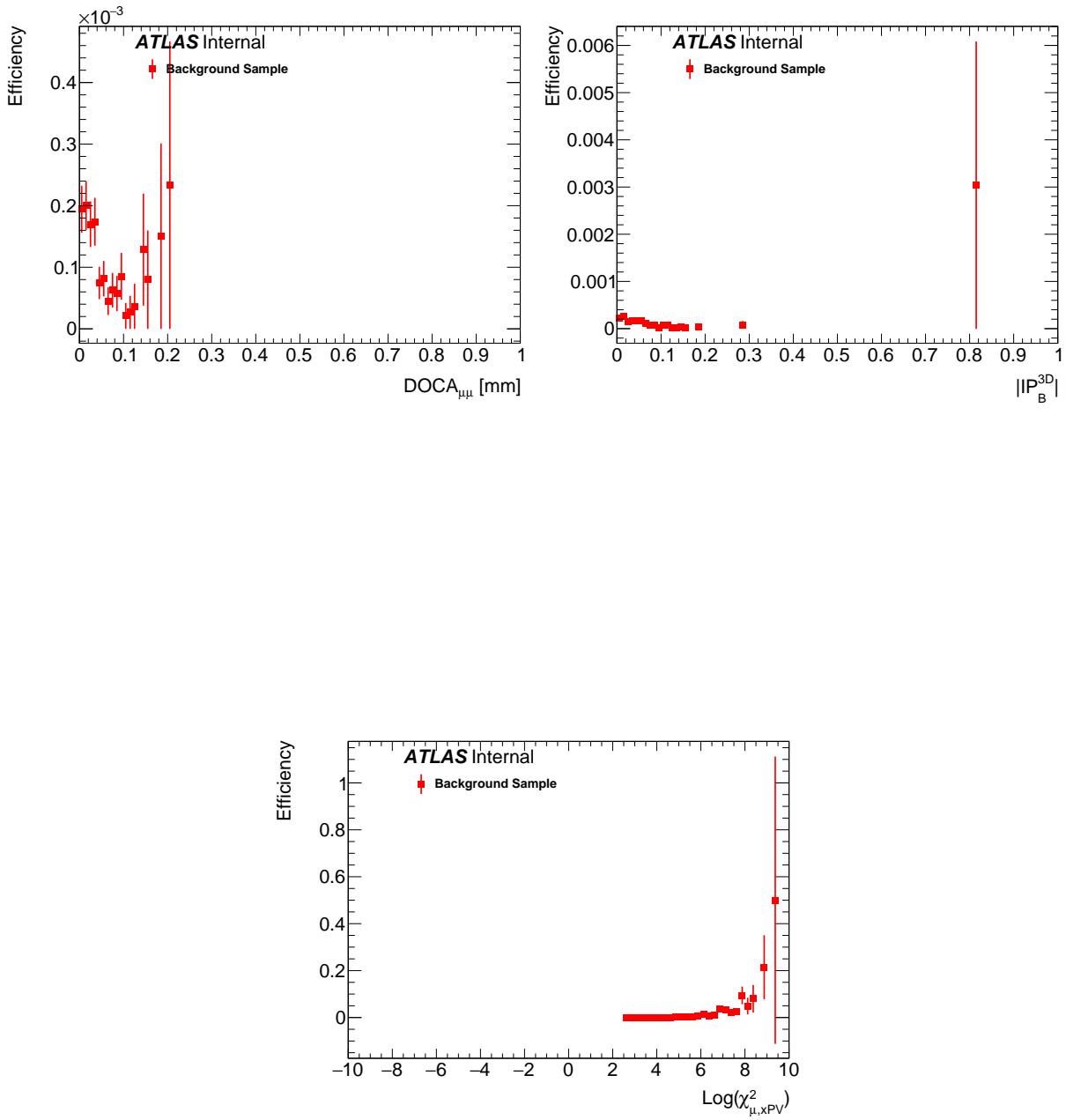
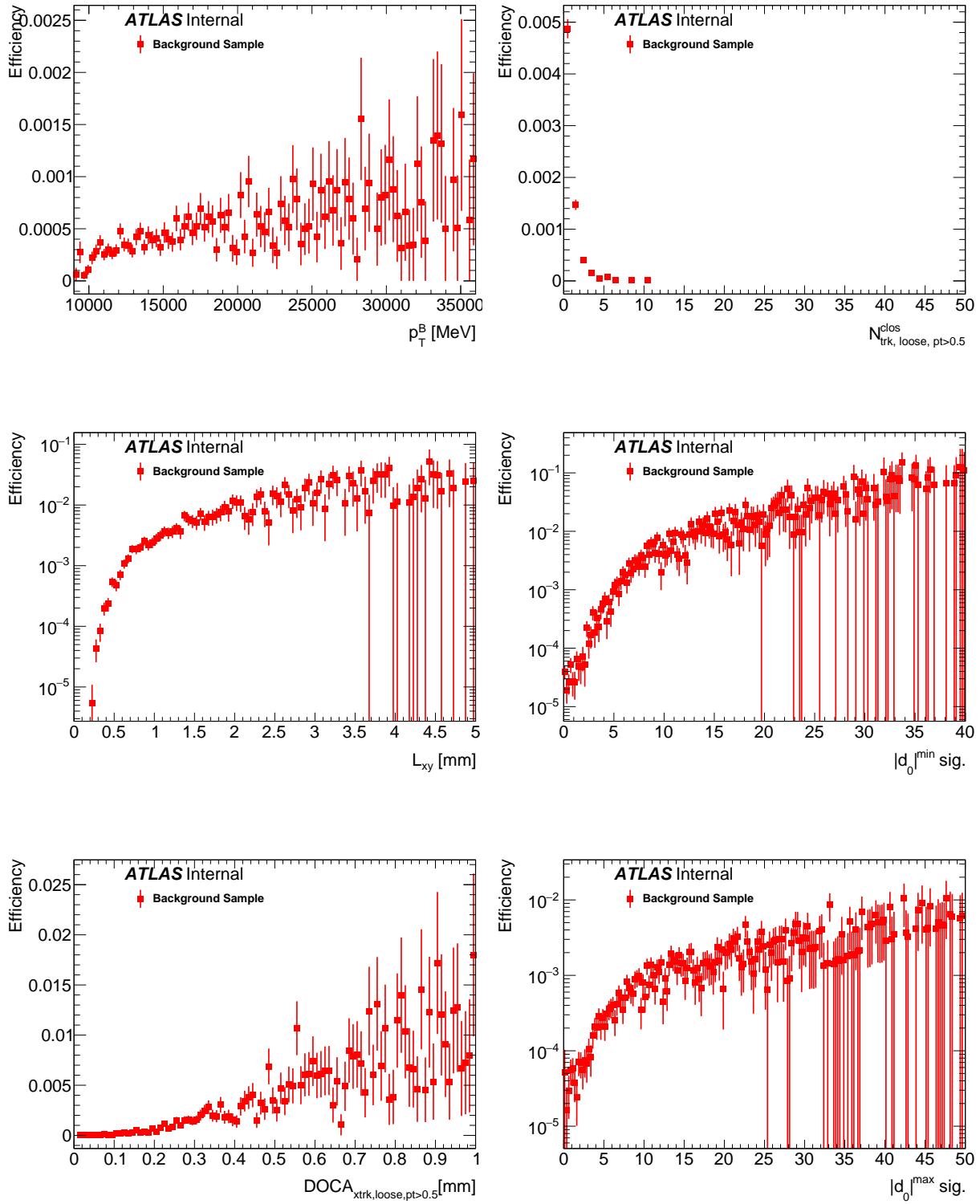
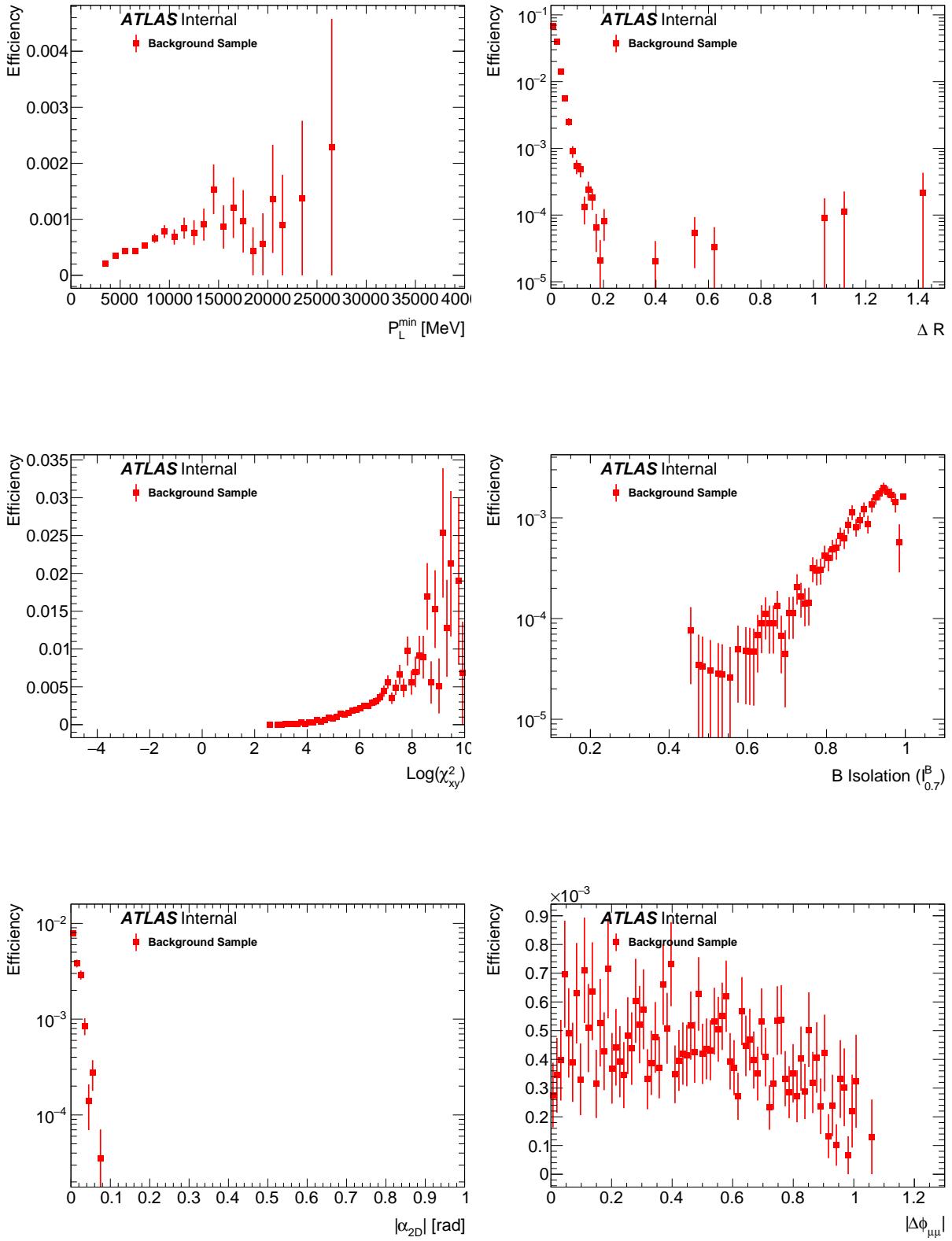


Figure 54: These 15 figures show the the variable efficiencies for the background sample with an 36 % signal efficiency cut on the BDT variable.

1318 The following figures show the variable efficiencies for the background sample with an 54% signal
 1319 efficiency cut on the BDT variable.





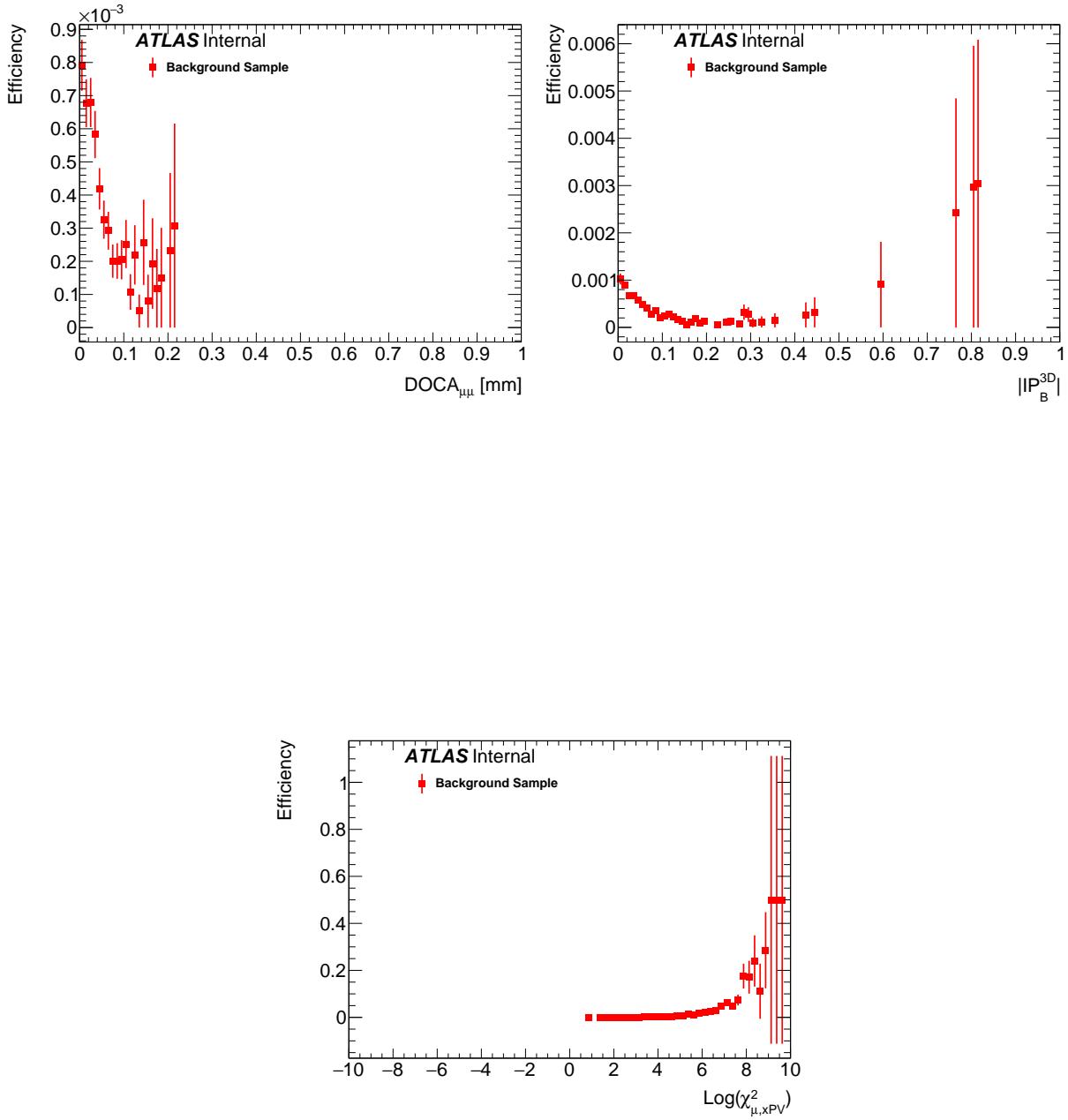
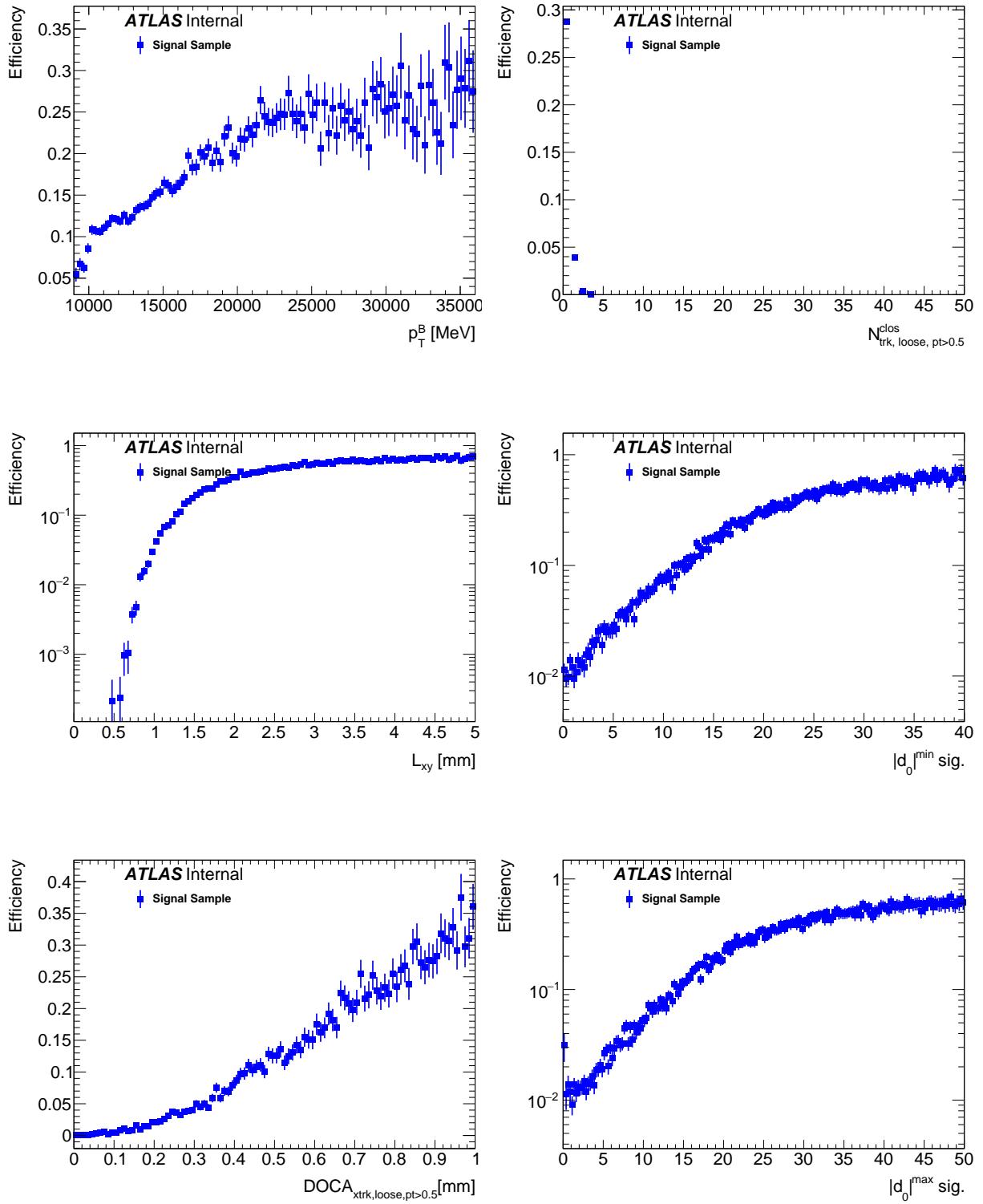
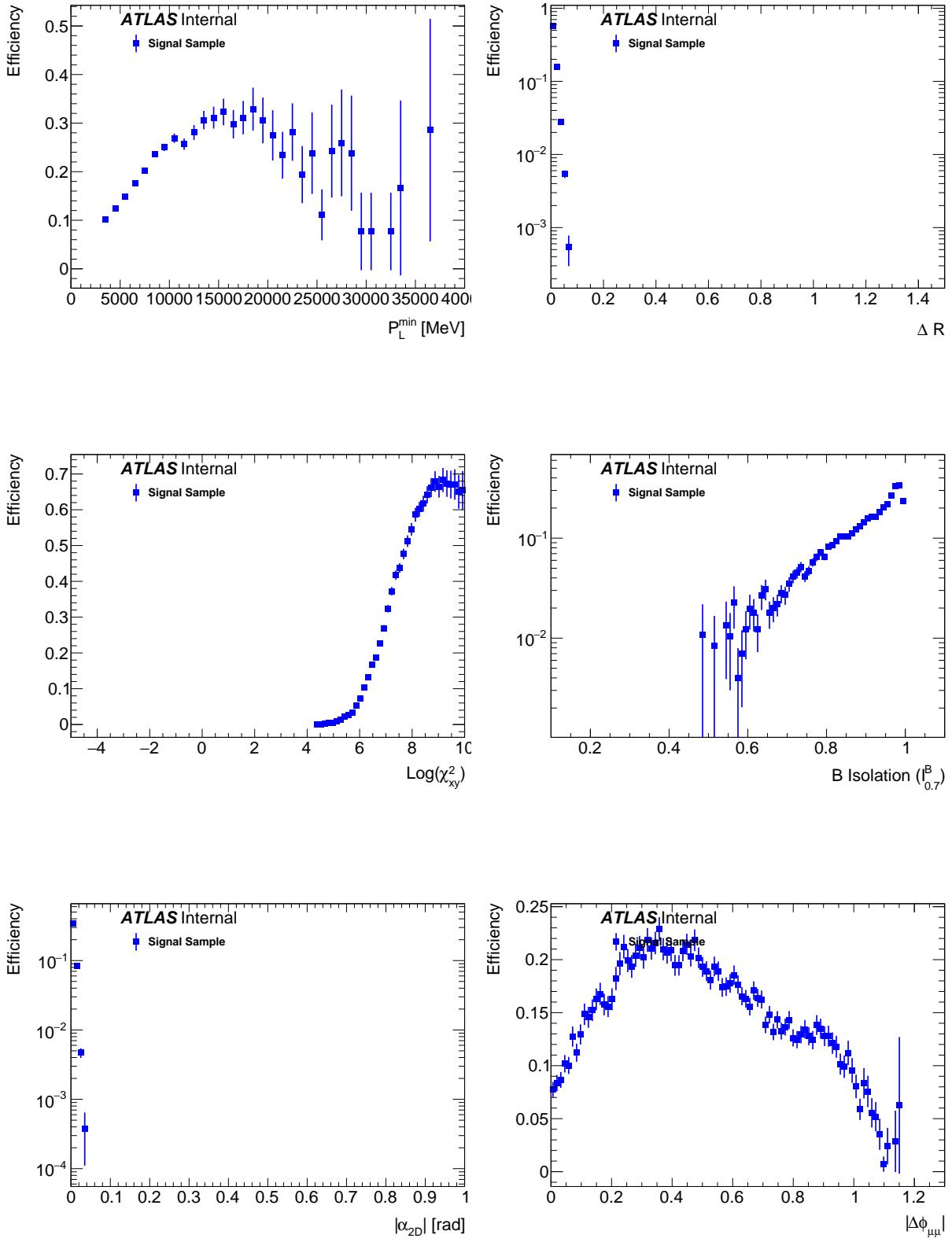


Figure 55: These 15 figures show the the variable efficiencies for the background sample with an 54 % signal efficiency cut on the BDT variable.

1320 The following figures show the variable efficiencies for the signal sample with an 18% signal efficiency
 1321 cut on the BDT variable.





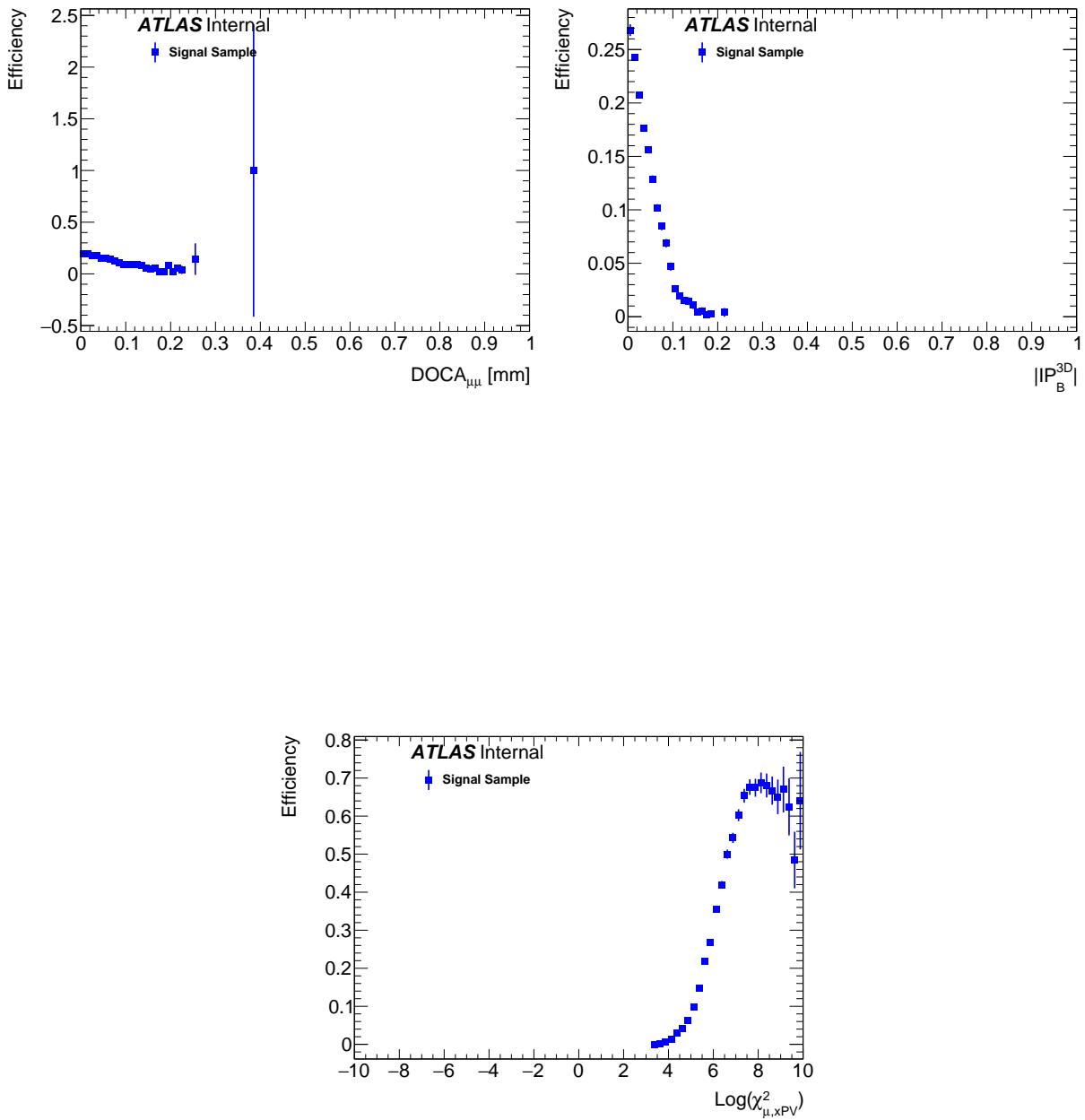
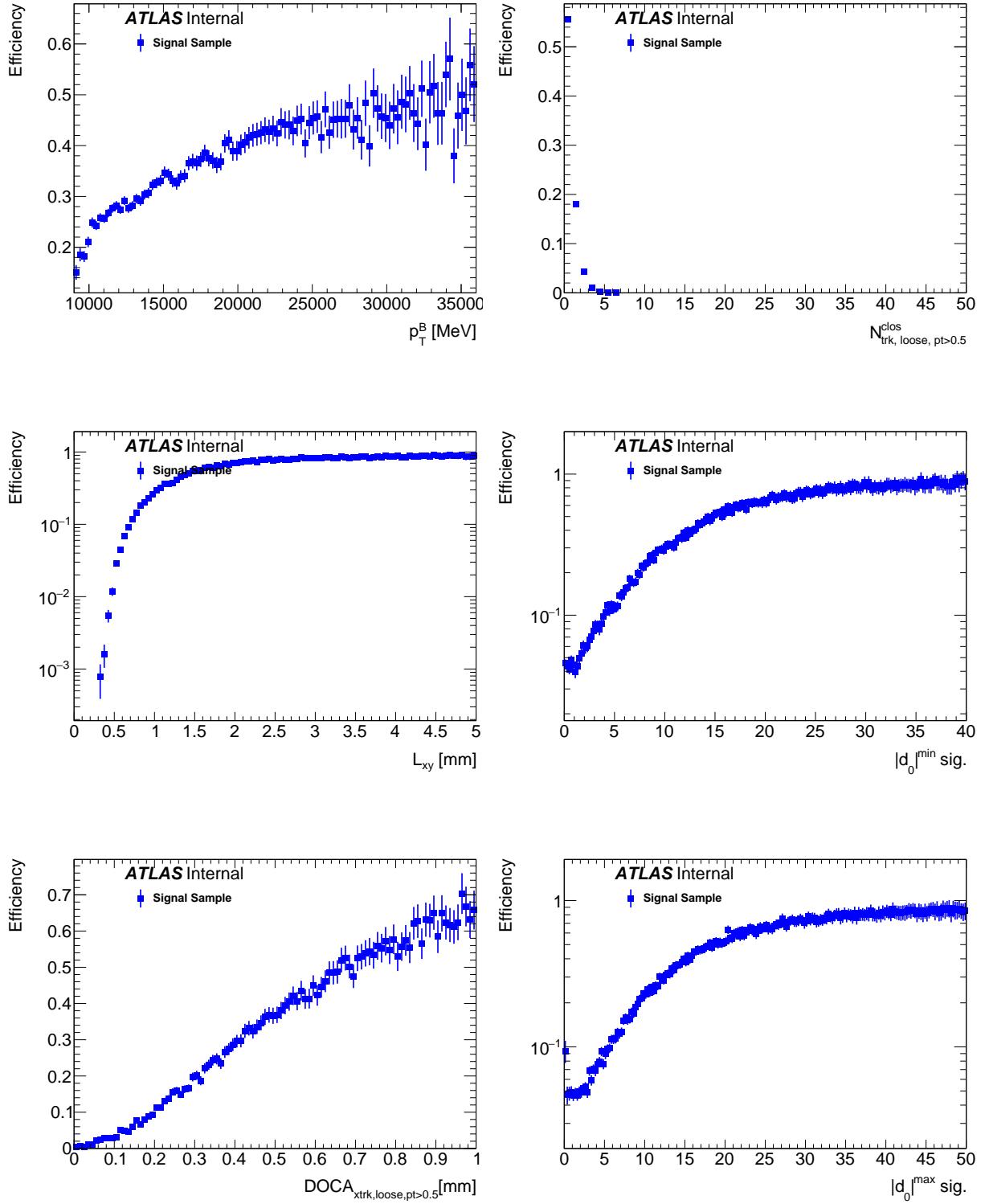
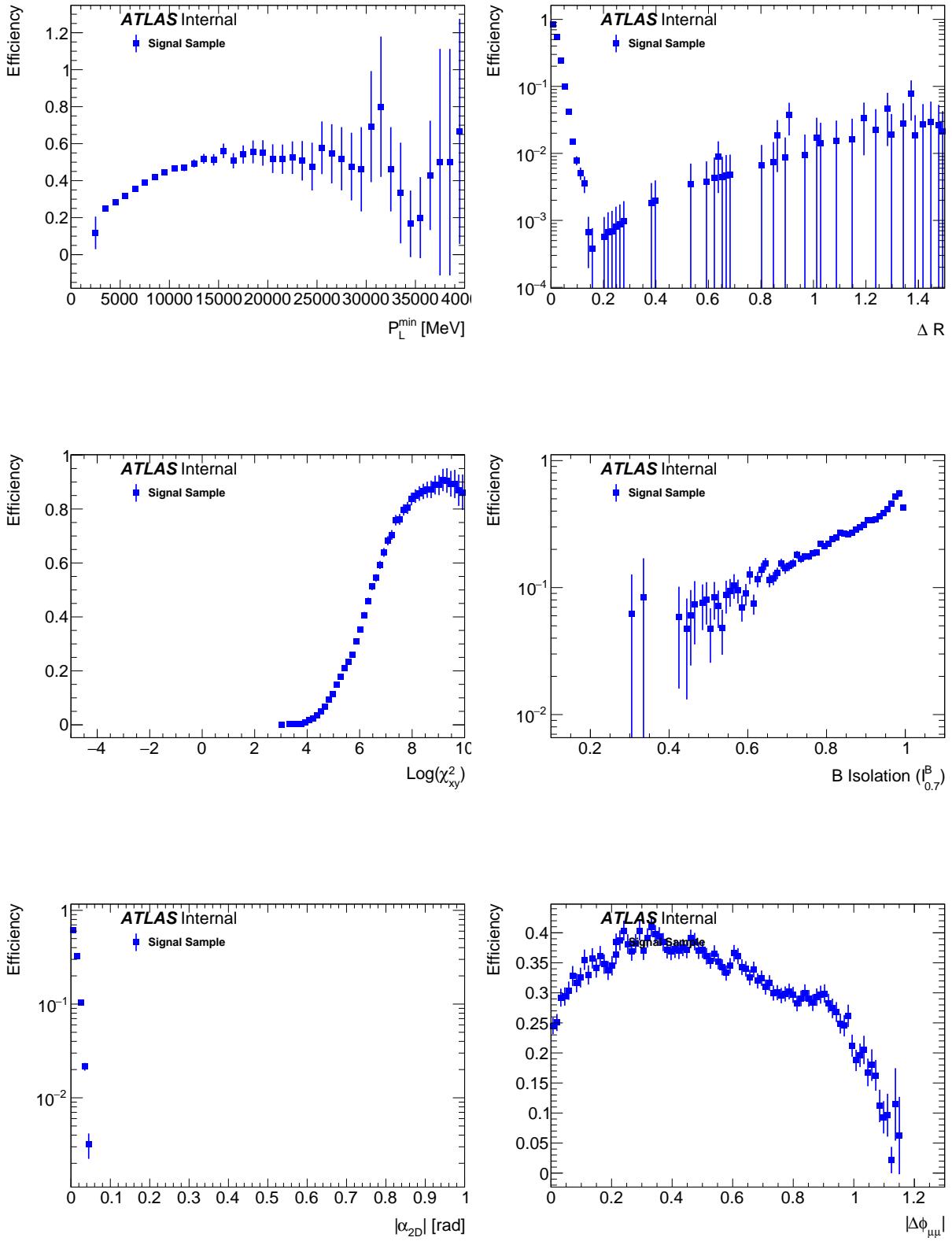


Figure 56: These 15 figures show the the variable efficiencies for the signal sample with an 18 % signal efficiency cut on the BDT variable.

1322 The following figures show the variable efficiencies for the signal sample with an 36% signal efficiency
 1323 cut on the BDT variable.





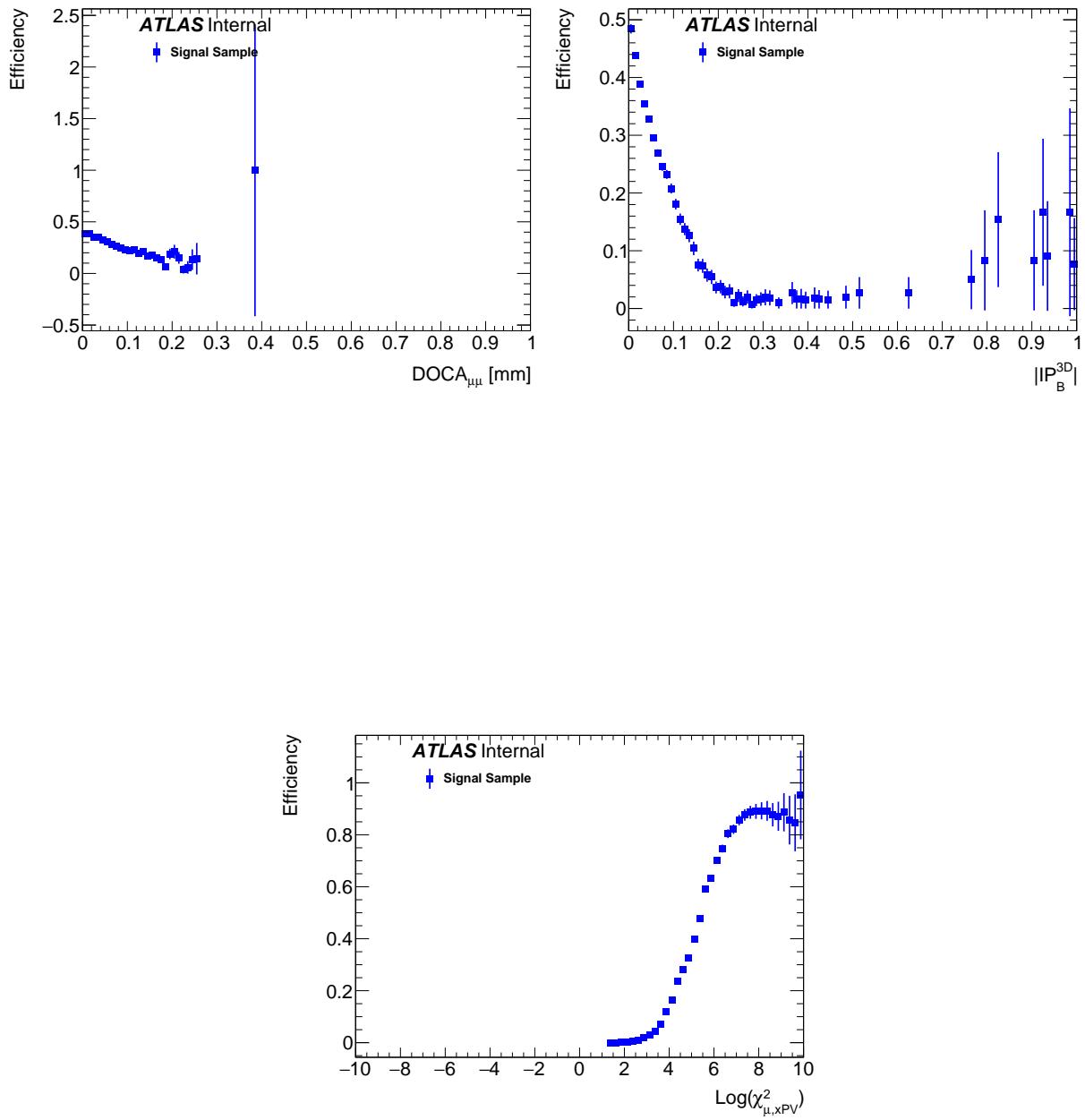
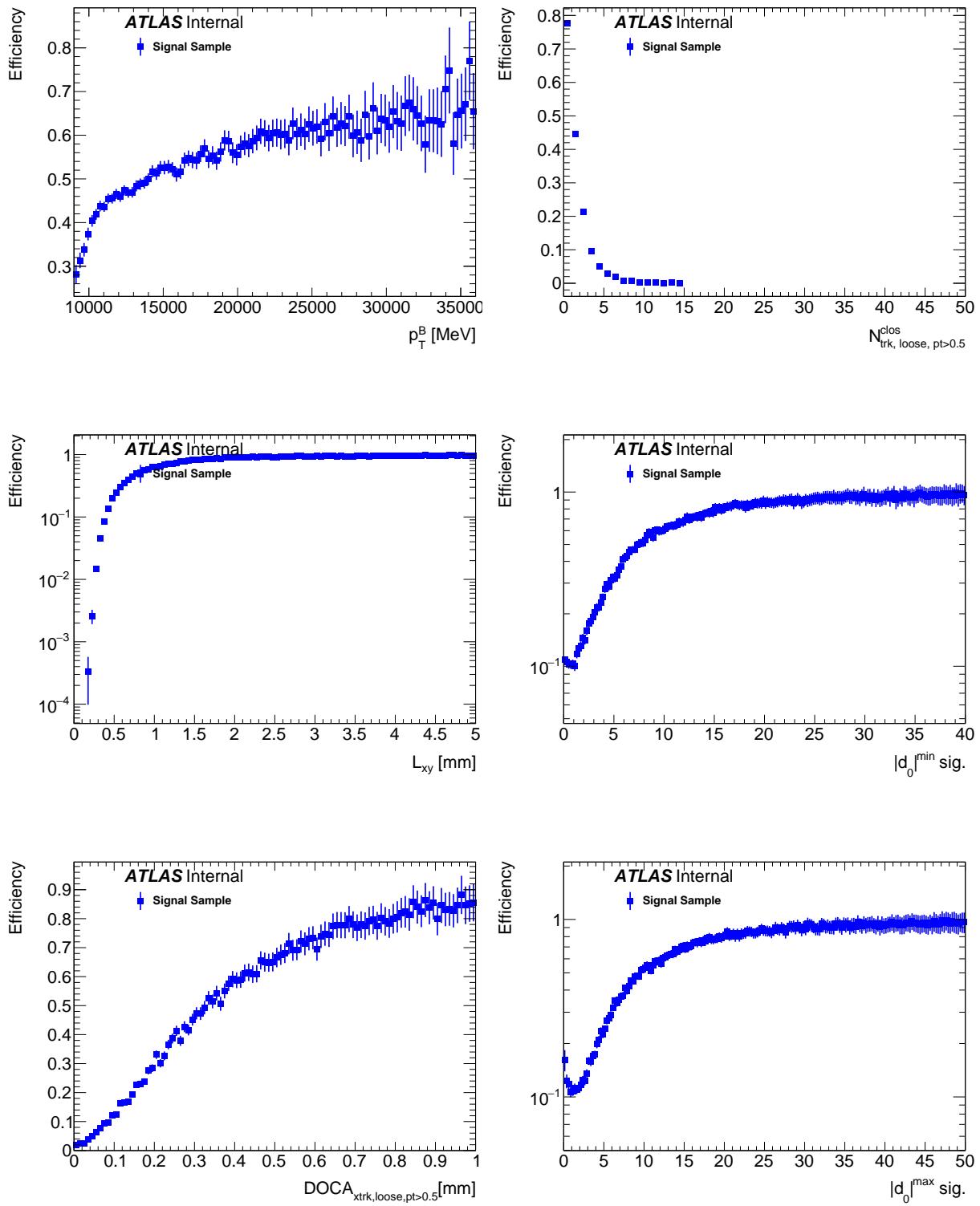
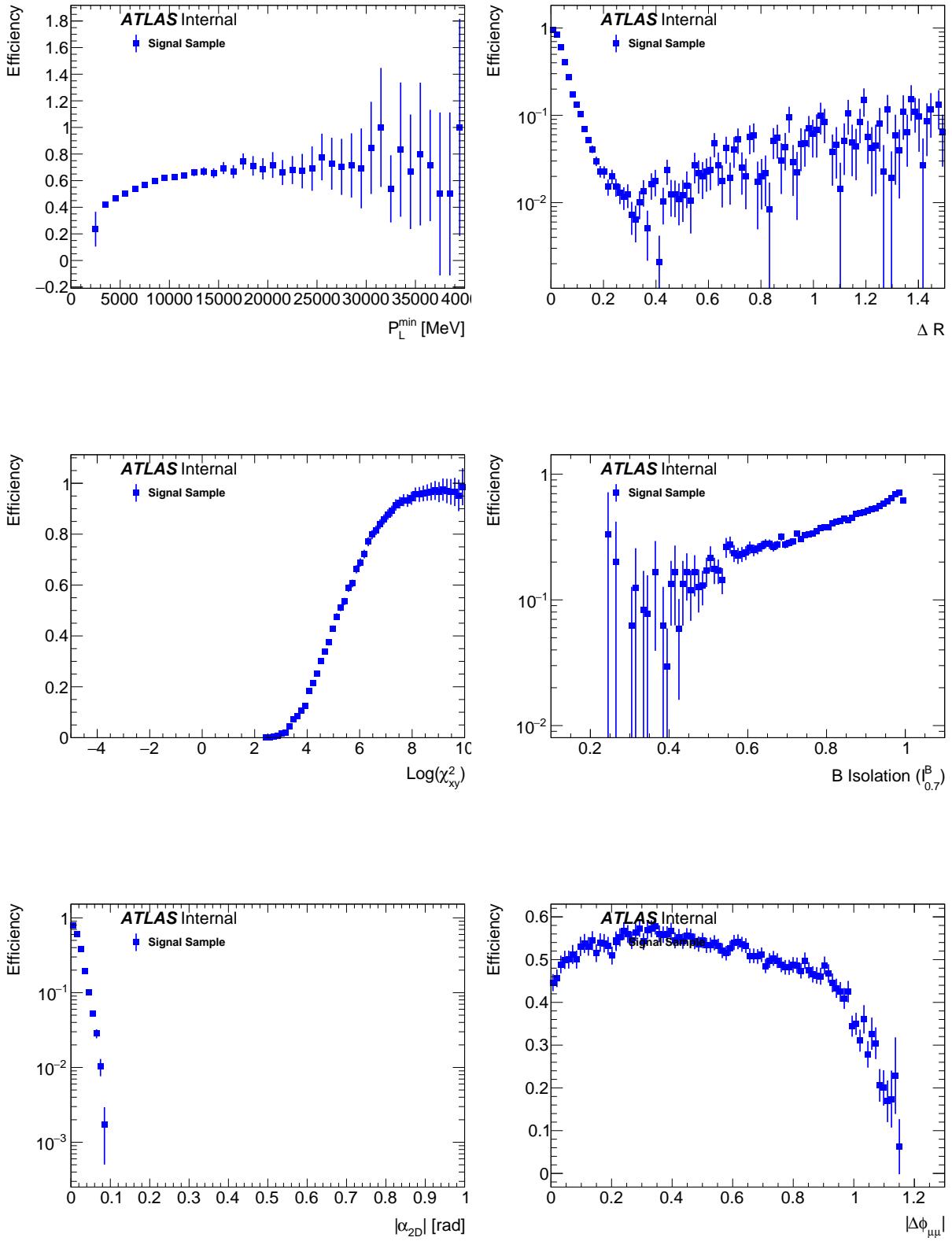


Figure 57: These 15 figures show the the variable efficiencies for the signal sample with an 36 % signal efficiency cut on the BDT variable.

1324 The following figures show the variable efficiencies for the signal sample with an 54% signal efficiency
 1325 cut on the BDT variable.





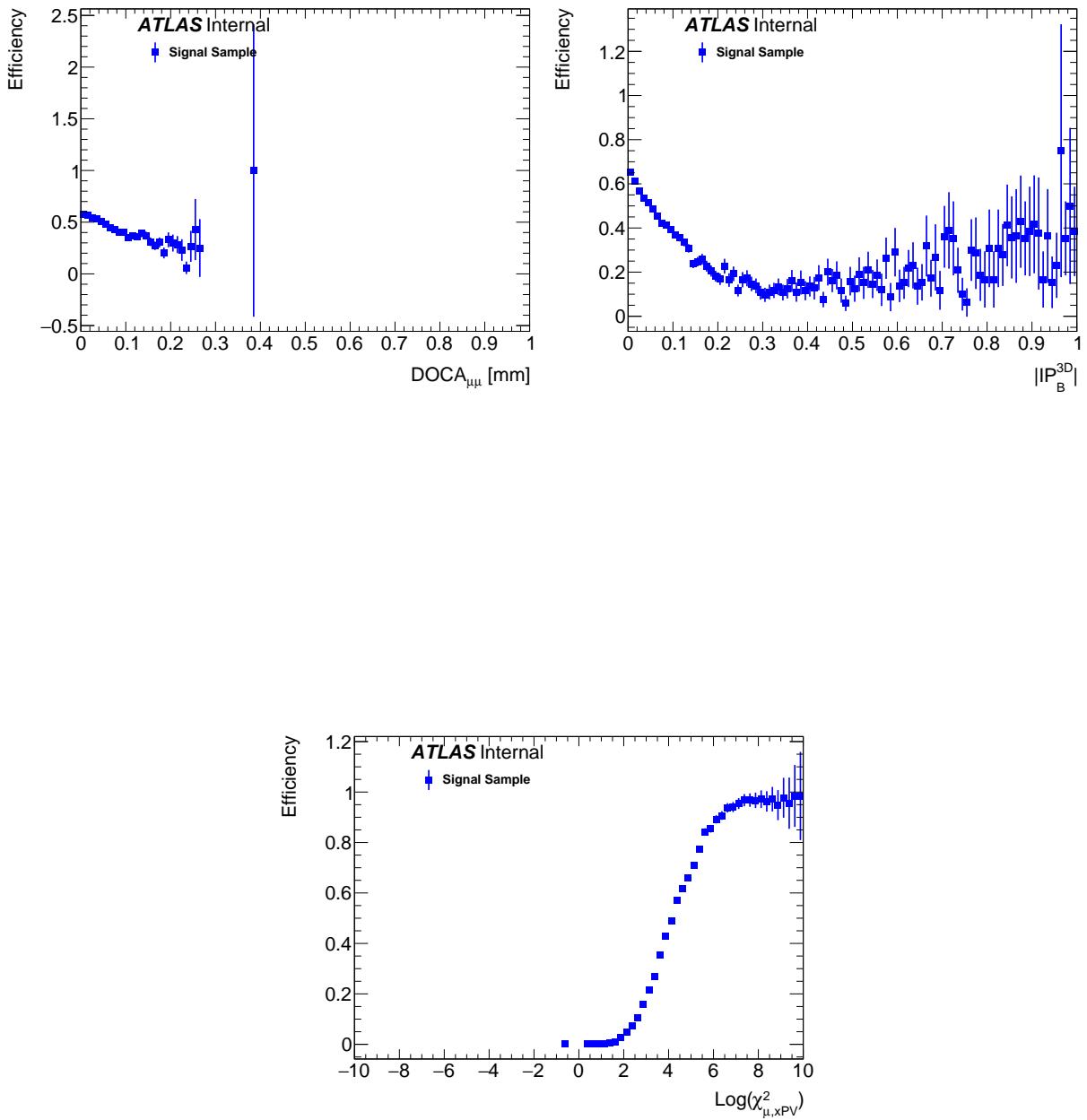


Figure 58: These 15 figures show the the variable efficiencies for the signal sample with an 54 % signal efficiency cut on the BDT variable.

1326 **G Fit for the reference channel yield extraction: detailed discussion of the**
 1327 **analysis**

1328 **G.1 Datasets and event selection**

1329 **G.2 Mass Fit Strategy**

1330 **G.2.1 Fit Structure**

1331 **G.2.2 Fit Initialisation**

1332 **G.2.3 Goodness of fit (GoF)**

1333 **G.3 Choice of fit models**

1334 **G.3.1 Signal model**

1335 **G.3.2 Splitting of partially reconstructed decays**

1336 **G.3.3 $J/\psi\pi$ and Combinatorial background models**

1337 **G.4 Default Fit Parameter counts**

1338 :

1339 **G.5 Evaluation of systematic uncertainties**

1340 **G.6 Fit validation studies**

1341 **G.7 Other crosschecks**

1342 **G.8 $B^\pm \rightarrow J/\psi K^\pm$ reference channel yield**

1343 **G.9 $B^\pm \rightarrow J/\psi K^\pm$ Appendix**

1344 **H 2D Neyman belt confidence region construction**

1345 In this appendix we explain the 2D Neyman belt construction procedure. In [H.1](#) we provide a desription
 1346 of the Neyman approach to confidence regions construction, in [H.2](#) we explain the procedure developed to
 1347 build a 2D Neyman belt, in [H.3](#) we explain some of the checks performed on a Neyman belt based on the
 1348 Run1 likelihood and in [H.4](#) we explain the introduction of the systematics in the contour construction.

1349 H.1 The Neyman approach to confidence region construction

1350 The result of the Run1 analysis [4] ended up in the unphysical region, such a result has a difficult statistical
 1351 interpretation and caused issues in the likelihood contours construction. The Neyman belt approach [11]
 1352 provides an exact frequentist approach to calculate the confidence regions for results near the physical
 1353 boundaries and with low statistics available.

1354 Basically, for a set of possible 'true' values of the parameters of interest (unknown physical constants)
 1355 toy-MC are generated and for each of the true values a region (or interval, for the 1D case) with a given
 1356 coverage is obtained; all the regions are used to build a confidence belt, that lives in the space of the fitted
 1357 value and the true values. Given an experimental result, its confidence region (or interval) is obtained
 1358 cheking if the value is contained in the region (interval) for a given true value of the physical parameter.
 1359 The meaning of a confidence region obtained with the Neyman approach is different from the meaning
 1360 of a region obtained with the likelihood contours approach. The latter doesn't relate to the 'true' value
 1361 of the physical constant(s) we are measuring, but it gives a frequentist estimation of the outcome of the
 1362 experiment, e.g. a 68% CL likelihood contour, tells us that if we repeat the experiment N times, 68% of
 1363 the results will end up inside the likelihood contour.

1364 The confindece region obtained with the Neyman approach provides a connection between the result of
 1365 the experiment (fitted value) and the true value of the unknown physical constant. Basically, a 68%
 1366 confidence region built with the Neyman approach tells us that the true value of the physical constant(s)
 1367 we are measuring lays in the confidence region with a 68% confidence level.

1368 We refer to the regions built for a given true value of the physical constants simply as regions.

1369 H.2 2D neyman belt construction

1370 The construction of the 2D Neyman belt [11] is based on different sets of pseudo-experiments (toy-MC)
 1371 generated using the the likelihood of the analysis, varying the values of $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ and $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$
 1372 according to a grid of values. Basically each toy-MC is generated accoddingly to the likelihood function,
 1373 allowing poissionian fluctuations on the number of events, and then the resulting dataset is fitted using the
 1374 analysis fitting machinery.

1375 For each point of the truth $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ grid a large sample of toy-MC is generated,
 1376 we refer to these samples as toy-MC sets. At early stages of the development of this procedure the toy-MC
 1377 sets were generate using the number of signal events instead of the branching fractions, therefore some
 1378 of the studies performed are based on $N(B_s^0 \rightarrow \mu^+ \mu^-)$ and $N(B^0 \rightarrow \mu^+ \mu^-)$ instead of the branching
 1379 fractions. Since $N(B_s^0 \rightarrow \mu^+ \mu^-)$ and $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ (same for $B_d^0 \rightarrow \mu\mu$) are directly proportional via
 1380 the single event sensitivity (SES), apart for the studies related to the systematics uncertainties that are
 1381 performed using $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ and $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$, the number of events and the branching fraction
 1382 are interchangeable.

1383 Each toy-MC set is ranked according to the profiled likelihood ratio:

$$\frac{\mathcal{L} \left(x | \mathcal{B}(B_s), \mathcal{B}(B_d), \hat{\theta} \right)}{\mathcal{L} \left(x | \widehat{\mathcal{B}(B_s)}, \widehat{\mathcal{B}(B_d)}, \hat{\theta} \right)} \quad (11)$$

1384 where $\mathcal{B}(B_s)$ and $\mathcal{B}(B_d)$ are the branching fraction of the two signal processes and θ represents all the
 1385 nuisance parameters. The denominator of equation 11 is the likelihood at its global minimum, while the
 1386 numerator is the value of the likelihood minimised fixing the two parameters of interest at the truth value

1387 used in the toy-MC set generation.

1388 In order to build a region with a given coverage, for instance the 68.3% confidence level region (1 sigma),
 1389 the 68.3% of the toys in the toy-MC set with the higher profiled likelihood ratio value are taken and a 2D
 1390 region is built in the space of the fitted values of $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) - \mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ ($N(B_s^0 \rightarrow \mu^+ \mu^-) -$
 1391 $N(B^0 \rightarrow \mu^+ \mu^-)$). This is not the actual confidence region, this procedure is run on all the points in the
 1392 truth $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) - \mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ grid.

1393 This approach provides a defined bulk of points, with edges subjected to fluctuations (see figure 59), it is
 1394 also not possible to verify the actual coverage of this region.

1395 In order to solve the latter issue, two sets of toy-MC are generated for each point of the truth $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$
 1396 - $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ grid. The first sample (sample A) is used to build the region, while the second sample
 1397 (sample B) is used to measure the coverage. If the coverage of the region is not compatible with the
 1398 desired coverage, toys from sample A are added or removed from the region, until the coverage measured
 1399 with sample B converges to the desired value.

1400 In order to remove the fluctuations at the edges of the region, the *monotone chain* algorithm [12] (will
 1401 refere to it as convex hull algorithm) is used on the region obtained using sample A to obtain a well defined
 1402 convex hull.

1403 For each point in the truth $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) - \mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ grid, the procedure to obtain a 68.3%
 1404 coverage region is:

- 1405 • take the 68.3% of the toys from sample A with the higher profiled likelihood ratio value;
- 1406 • build a convex hull based on those points, in the fitted $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) - \mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ ($N(B_s^0 \rightarrow$
 1407 $\mu^+ \mu^-) - N(B^0 \rightarrow \mu^+ \mu^-)$) space;
- 1408 • measure the coverage using sample B;
- 1409 • if the coverage is not compatible with the requested coverage (68.3% in this case), add or remove
 1410 sample A points;
- 1411 • build a new convex hull on the resulting region and mesure the coverage;
- 1412 • iterate until the coverage converges to 68.3%.

1413 Figure 59 shows this procedure applied to toy-MC sets generated based the Run1 likelihood using as
 1414 values of the signal branching fractions their SM predictions. The region obtained with one toy-MC set
 1415 (the 2D histogram) is shown and overlaid the convex hull contour obtained with the usage of sample B.
 1416 This approach allows to build a region with well defined contours and a known coverage for each point in
 1417 the truth $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) - \mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ grid.

1418

1419 Once the regions are built for each point in the truth $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) - \mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ grid, the construction
 1420 of the actual confedence region starts. Given an experimental measure, each region is analysed to verify
 1421 if it contains the measurement.

1422 If the experiential measurement is contained in a region, the values of the signal branching fractions used
 1423 in the generation of the toy-MC sets that built this region are taken.

1424 Figure 60 shows the 1,2 and 3 sigma confidence regions obtained generating the toy-MC sets using the
 1425 Run1 likelihood and assuming the SM expectations of the branching fractions as measured value.

1426 Also in this case fluctuations are present at the edges of the regions, so a smoothing procedure is applied;
 1427 the convex hull algorithm is exploited and the set of points that compose the contour obtained with the
 1428 convex hull algorithm are then fitted using the expected shape of the confidence region.

1429 Figure 61 shows the final result for this construction overlaid to the confidence regions affected by the

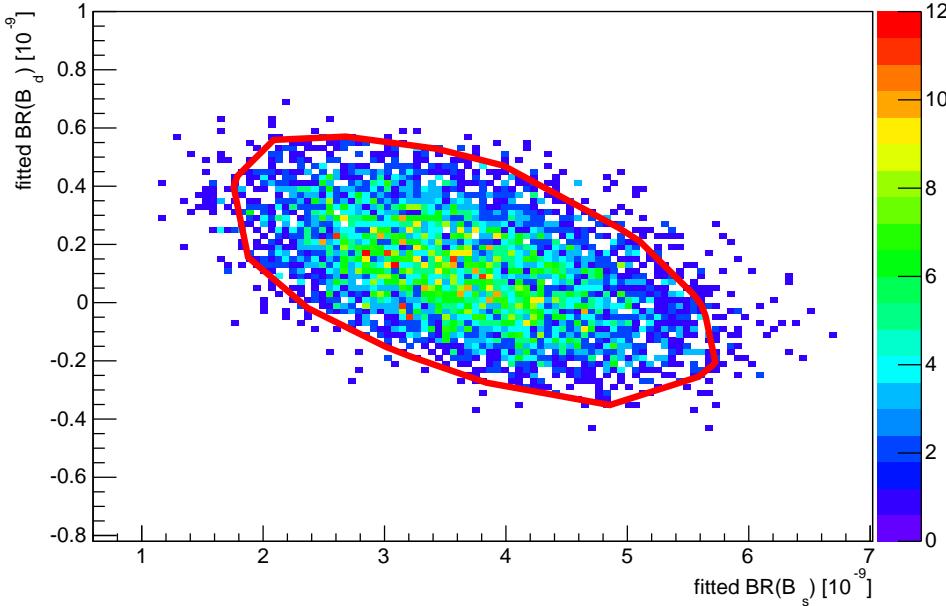


Figure 59: Plot show a region obtained with two set of toy-MCs, generated using as initial signal branching fraction the SM expectation. The 2D histogram show the region obtained using only one toy-MC set, the bulk of the region is well defined, but the edges are subjected to fluctuations. The red contour is obtained using two toy-MC sets and the convex hull algorithm. The region is now well defined and the usage of two toy-MC sets allows a proper measurement of the coverage (as explained in section H.3).

1430 edge fluctuations, again using the run1 likelihood for the toy-MC sets generation and the SM expectations
 1431 of the branching fractions as measured value.

1432

1433 H.3 Checks on coverage

1434 In order to test the validity of this approach, a third set of of toy-MC is generated for each point in the
 1435 truth branching fractions grid. It is used as an unbiased sample, not involved in the region construction
 1436 procedure, to measure the coverage.

1437 Figures 62, 63 and 64. show the coverage of the regions, measured using the three samples, built on
 1438 each point of the truth $N(B_s^0 \rightarrow \mu^+ \mu^-) - N(B^0 \rightarrow \mu^+ \mu^-)$ grid, with the toy-MC sets generated using the
 1439 Run1 likelihood. Regions are built for three coverages, 68.27% (1 sigma), 95.45% (2 sigma), 99.73% (3
 1440 sigma).

1441 The coverage measured using sample A results overestimated, with respect to the coverage measured
 1442 with sample C (unbiased), which is compatible within its uncertainty with the desired coverage. The
 1443 coverage measured with sample B shows a sharp cut at the wanted coverage, this is due to the construction
 1444 procedure, the coverage measured with sample B is not allowed to be lower then the wanted coverage; this
 1445 choice yeilds to a conservative estimation of the confidence region and its coverage.

1446 The uncertainty on the coverage of the regions can be extracted from the 1-dimensional plots in fig-
 1447 ures 62, 63 and 64. The uncertainty on the coverage measured using sample C is compatible with the
 1448 uncertainty of the coverage measured with sample A. This proves that the construction of the regions

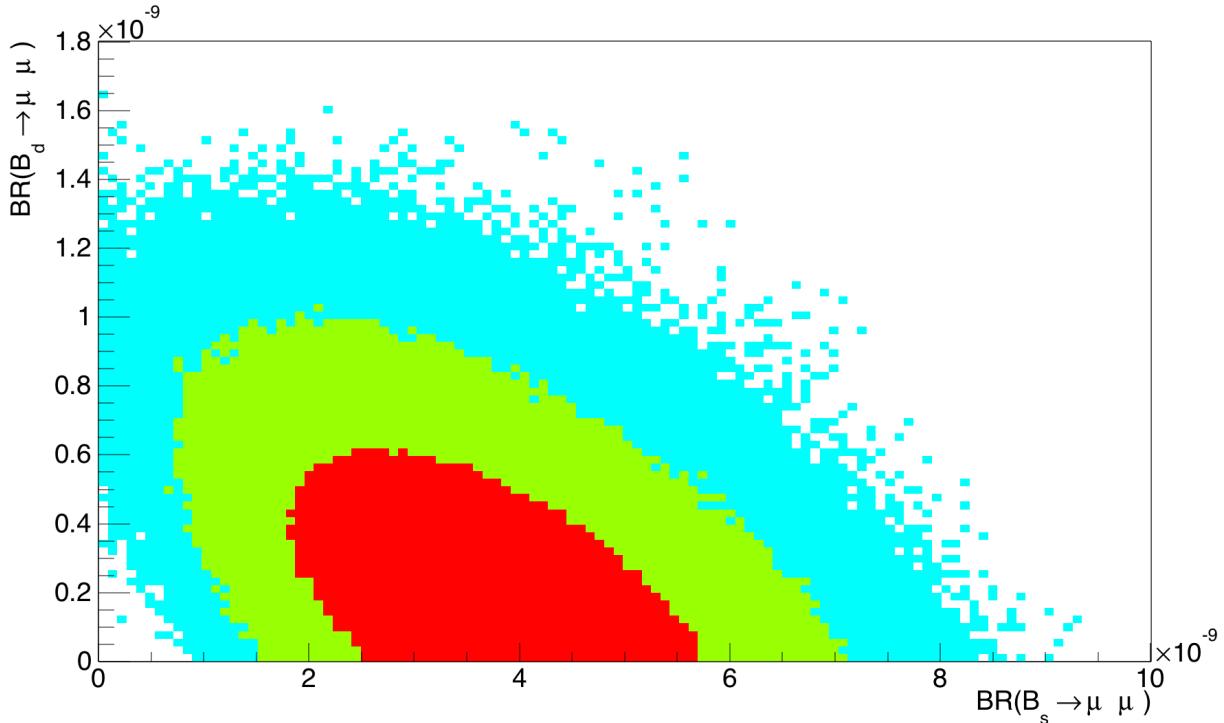


Figure 60: The plot shows the 1, 2 and 3 sigma regions (coverage respectively 68.27%, 95.45% and 99.73%) built using the Run1 likelihood function and using the SM prediction as central value for the construction. Each point is derived applying the two toy-MC sets approach explained in the text. The bulk of the confidence regions is well defined, but the edges are affected by fluctuations.

1449 using two samples provides regions with the desired coverage.
 1450
 1451

1452 H.4 Systematics uncertainties in 2D Neyman belt

1453 Tho sources of systematics uncertainties of the analysis can be gathered in two main categories:

- 1454 • constant systematics, the systematic uncertainty on these quantities doesn't vary with the analysis
 1455 result, e.g. the uncertainty on the acceptance \times efficiency calculation;
- 1456 • systematics correlated with the analysis result, these quantities depend on the result on the signal
 1457 fit, e.g. the systematics on the signal fit.

1458 The first class of systematic uncertainties are gathered into the uncertainty on the single event sensitivity,
 1459 described as a gaussian smearing in the signal fit. The second class of systematic uncertainties affect the
 1460 signal fit results, they are described as a two dimensional gaussian smearing applied to the number of
 1461 fitted signal events.

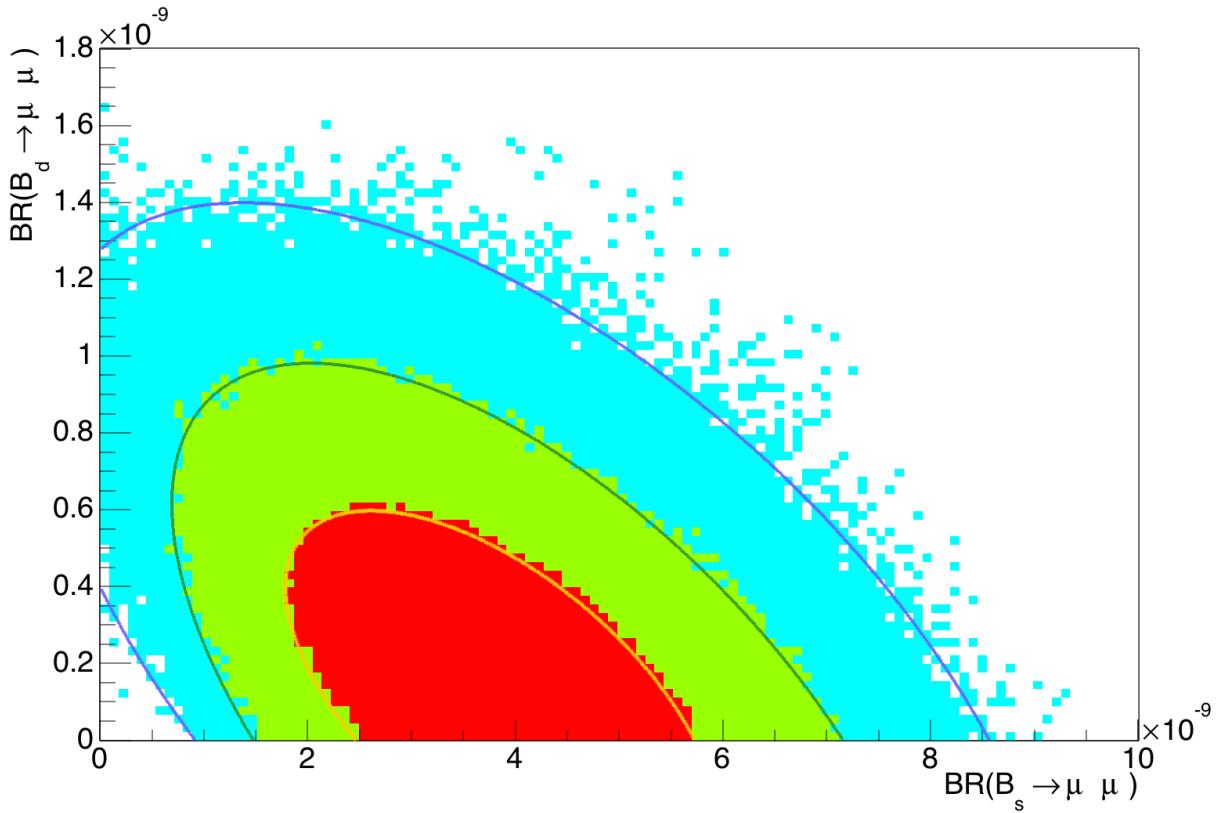


Figure 61: The plot shows the 1, 2 and 3 sigma regions (coverage respectively 68.27%, 95.45% and 99.73%) built using the Run1 likelihood function for the toy-MC sets generation and using the SM prediction as central value for the construction. Each point is derived applying the two toy-MC sets approach explained in the text. The same confidence regions shown in 60 are drawn and the contours obtained applying the somthing procedure explainedin the text are also drawn.

The systematics can not be added to the Neyman belt construction in the toy-MC sets generation, because for certain combinations of the truth values of the signal branching fractions, the number of signal events to be generated might become negative (low values of the signal branching fractions with a underfluctuation of the PDF of the systematic uncertainty). The systematics are therefore introduced in the Neyman belt construction with the following procedure:

- generate a toy-MC without considering the systematic uncertainties influence;
- generate values for the quantities affected by the systematics according to their gaussian constraint (e.g. generate $B_s^0 \rightarrow \mu^+ \mu^-$ single event sensitvity value);
- perform fit using as central value of the quantity the generated value (e.g. set as central value of the $B_s^0 \rightarrow \mu^+ \mu^-$ single event sensitvity its generated smeared value).

In order to properly measure the effect of the systematic uncertainties in the confidence region, four fits are performed on each generated toy-MC:

1. fit on the generated toy-MC, without varying the values of the quantities affected by the systematics (denominator of the likelihood ratio used in the ranking of the toy-MC sets);

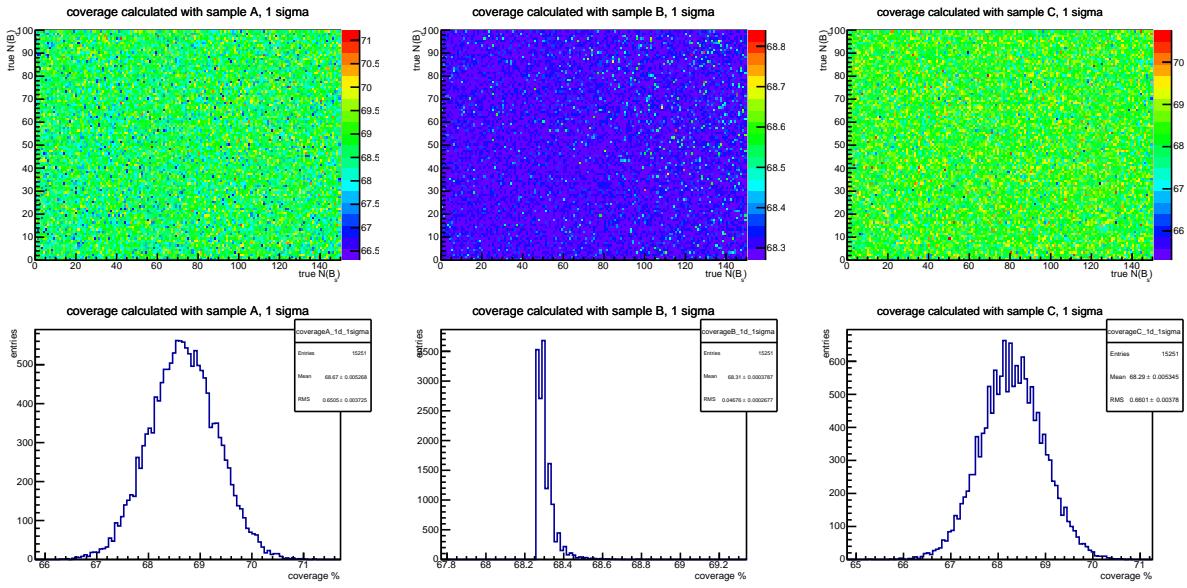


Figure 62: The plots show the coverage of the 1 sigma (68.27% coverage) regions built for all the points in the truth $N(B_s^0 \rightarrow \mu^+ \mu^-) - N(B_s^0 \rightarrow \mu^+ \mu^-)$ grid, all the toy-MC sets are generated using the Run1 analysis likelihood function. The three columns show the coverages measured with the three samples, respectively, starting from the left, coverage measured with sample A, B and C. The upper plots show the coverage of each point in the grid of possible values used in the generation, while the lower plots show the 1D plot of all the coverages. The coverage measured with sample C (not used in the region construction) is compatible with the wanted coverage within its statistical uncertainty, while the coverage measured with sample A overestimates the actual value. The coverage measured with sample B shows a sharp cut, this is due to the region contraction procedure, as explained in the text.

- 1476 2. fit on the generated toy-MC, without varying the values of the quantities affected by the systematics,
 1477 but fixing the signal branching fractions to the value used in the generation (numerator of the
 1478 likelihood ratio used in the ranking of the toy-MC sets);
- 1479 3. fit on the generated toy-MC, applying the systematics uncertainty variation (denominator of the
 1480 likelihood ratio used in the ranking of the toy-MC sets when considering also the systematics
 1481 uncertainties);
- 1482 4. fit on the generated toy-MC, applying the systematics uncertainty variation and fixing the signal
 1483 branching fractions to the value used in the generation (numerator of the likelihood ratio used in the
 1484 ranking of the toy-MC sets when considering also the systematics uncertainties);

1485 These four fits allow to build the confidence regions considering the statistical only uncertainty and
 1486 the statistical + systematics uncertainty. Figure 65 shows the confidence regions obtained considering
 1487 the systematics uncertainty in the regions construction (solid lines) and the statistical only confidence
 1488 regions (dashed lines), using the Run1 analysis likelihood in the toy-MC sets generation and the SM
 1489 theoretical prediction of the signal branching fractions as starting point of the construction. The systematic
 1490 uncertainties introduced are the Run1 analysis systematics.

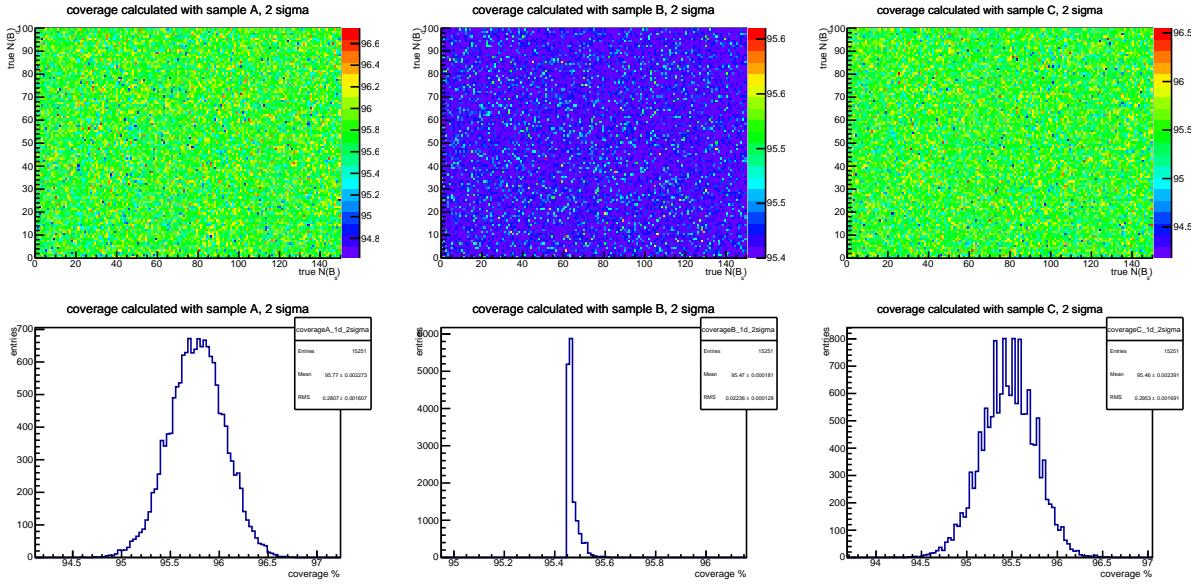


Figure 63: The plots show the coverage of the 2 sigma (95.45% coverage) regions built for all the points in the truth $N(B_s^0 \rightarrow \mu^+ \mu^-) - N(B^0 \rightarrow \mu^+ \mu^-)$ grid, all the toy-MC sets are generated using the Run1 analysis likelihood function. The three columns show the coverages measured with the three samples, respectively, starting from the left, coverage measured with sample A, B and C. The upper plots show the coverage of each point in the grid of possible values used in the generation, while the lower plots show the 1D plot of all the coverages. The coverage measured with sample C (not used in the region construction) is compatible with the wanted coverage within its statistical uncertainty, while the coverage measured with sample A overestimates the actual value. The coverage measured with sample B shows a sharp cut, this is due to the region contraction procedure, as explained in the text.

1491 H.4.1 Checks on systematics uncertainties introduction

1492 In order to check the systematic uncertainties introduction in the Neyman belt construction procedure,
 1493 different sets of toy-MC have been generated using the same truth values of the signal branching fractions
 1494 in the generation (SM prediction) and adding the systematics one at the time; four sets of toy-MC therefore
 1495 are generated:

- 1496 • statistical only uncertainty considered;
- 1497 • statistical uncertainty and systematics on single event sensitivity considered;
- 1498 • statistical uncertainty and systematics on signal fit considered;
- 1499 • statistical uncertainty and all systematics considered.

1500 The distribution of the toy-MC fits are then drawn to evaluate the effect of the systematic uncertainties.
 1501 Each plot in figure 66 shows the resulting distributions of $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$, the measured RMS of the
 1502 distribution and the expected RMS considering the effect of the systematic uncertainties. The measured
 1503 RMS is always consistent with the expectations, showing that the systematics introduction procedure is
 1504 solid. Similar results are obtained for $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$.

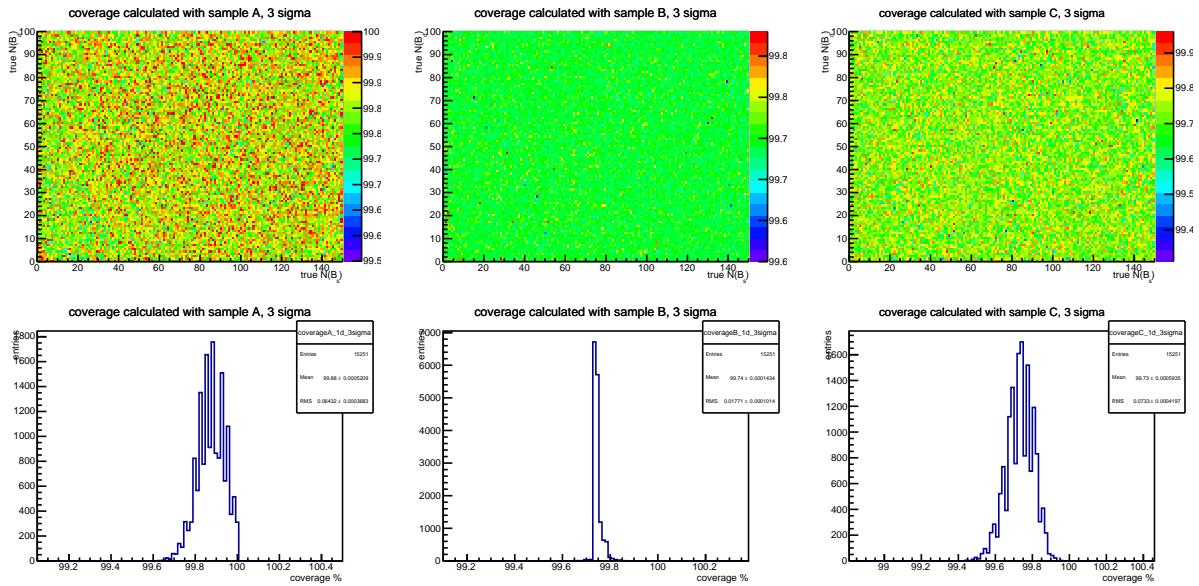


Figure 64: The plots show the coverage of the 3 sigma (99.73% coverage) regions built for all the points in the truth $N(B_s^0 \rightarrow \mu^+ \mu^-)$ - $N(B^0 \rightarrow \mu^+ \mu^-)$ grid, all the toy-MC sets are generated using the Run1 analysis likelihood function. The three columns show the coverages measured with the three samples, respectively, starting from the left, coverage measured with sample A, B and C. The upper plots show the coverage of each point in the grid of possible values used in the generation, while the lower plots show the 1D plot of all the coverages. The coverage measured with sample C (not used in the region construction) is compatible with the wanted coverage within its statistical uncertainty, while the coverage measured with sample A overestimates the actual value. The coverage measured with sample B shows a sharp cut, this is due to the region construction procedure, as explained in the text.

CL contours for SM theoretical prediction

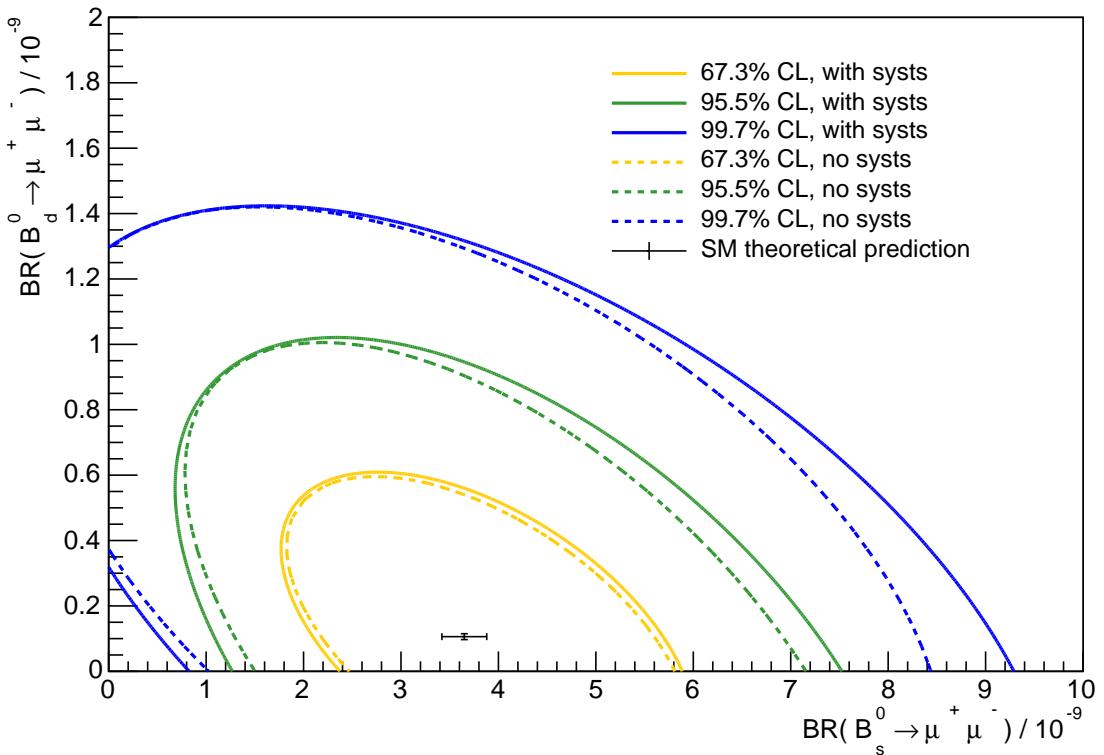


Figure 65: The plot shows the 1, 2 and 3 sigma regions (coverage respectively 68.27%, 95.45% and 99.73%) built using the Run1 likelihood function for the toy-MC sets generation and using the SM prediction as central value for the construction. Two sets of contours are shown, statistical uncertainty only contours (dashed lines) and statistical + systematics uncertainty contours (solid lines). The systematic uncertainties introduced are the Run1 analysis systematics.

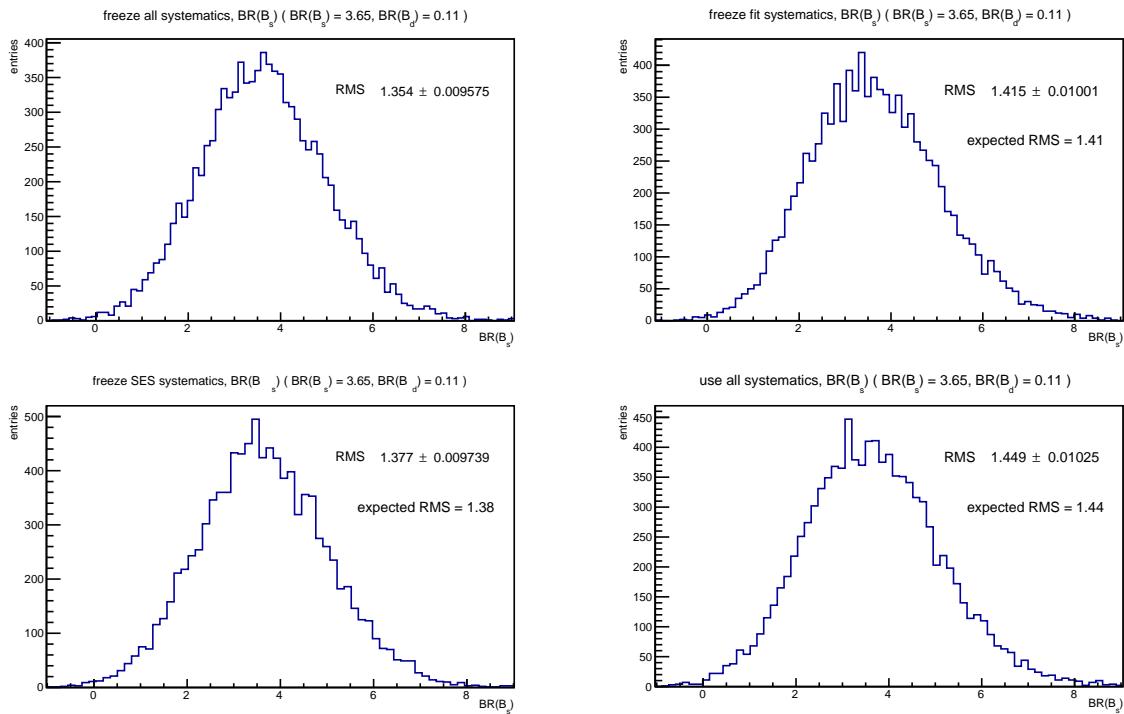


Figure 66: The four plots show the resulting $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ distribution from toy-MC sets generated based the Run1 likelihood function and using the SM theoretical predictions of the signal branching fractions as true values. The top left plot shows the $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ distribution when only the statistical uncertainty is considered, in the top right plot the systematics on the single event sensitivity is added, in the bottom left plot the systematic uncertainty on the signal fit is added to the statistical uncertainty and in the bottom right plot both systematics are considered. The measured and expected RMS of the distributions are shown in each plot. The measured RMS is always in agreement with the expectation.

- 1505 In an ATLAS note, use the appendices to include all the technical details of your work that are relevant
1506 for the ATLAS Collaboration only (e.g. dataset details, software release used). This information should
1507 be printed after the Bibliography.