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

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]

[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*

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

Contents

3

1 Note Changes

61 • **v**0.1

62   – Initial version

2 Introduction

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

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

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

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

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

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

$$\mathcal{B}(B^0_{(s)} \to \mu^+\mu^-) = \mathcal{B}(B^\pm \to J/\psi K^\pm \to \mu^+\mu^- K^\pm) \times \frac{f_u}{f_s} \times N_{\mu^+\mu^-} \times \left(N_{J/\psi K^\pm} \frac{(A\varepsilon)^k_{\mu^+\mu^-}}{(A\varepsilon)^k_{J/\psi K^\pm}} \right)^{-1}, \qquad (1)$$

and simplified with respect to the final Run 1 analysis thanks to the fact that we employ one single trigger category encompassing $\sim 80 - 85\%$ of the signal for the 2015+2016 Run 2 dataset.
While the branching ratio $BR(B^\pm \to J/\psi K^\pm \to \mu^+\mu^- K^\pm)$ and the relative B_u/B_d production fraction $\frac{f_u}{f_s}$ 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 16.

of the remaining ingredients to this formula and their uncertainties.

The measured $B^+ \to J/\psi K^+$ yield $N_{J/\psi K^\pm}$ will be derived in section 13, while the relative efficiencies and acceptances of $B^0_{(s)} \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$ will be extracted in 14. The $B^0_{(s)} \to \mu^+\mu^-$ yield will be derived from invariant mass distribution fits in section 15, implementing the selections optimized in sections 9, 11 and 8.

Given these ingredients, the branching ratio to the left of the reference formula above will be derived in section 16, where relevant systematic effects will be propagated as well. Dedicated sections will discuss the tuning of Montecarlo samples 5, the data-driven 12 extrapolation of MC models to the signal region 10.

3 AnalysisMethodology

4 Data and Monte Carlo Samples

4.1 Data Samples

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

For the $B^0_{(s)} \to \mu^+\mu^-$ signal channel a region of 360 MeV width around the B^0_s mass is omitted during the analysis development to avoid any bias ("blinding")[2].

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

- Data 2015, physics_Main stream:
  data15_13TeV.period[P].physics_Main.PhysCont.DAOD_BPHY8.grp15_v01_p3372
  with periods P A, C, D, E, F, G, H, J.

- Data 2106, physics_Main stream:
  data16_13TeV.period[P].physics_Main.PhysCont.DAOD_BPHY8.grp16_v01_p3372
  with periods P A, B, C, D, E, F, G, I, K, L.

- Data 2016, physics_BphysDelayed stream:
  data16_13TeV.period[P].physics_BphysDelayed.PhysCont.DAOD_BPHY8.grp16_v01_p3372
  with periods P D, E, F, G, I, K, L.

4.2 Monte Carlo Samples

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

For each sample, the number of events generated is given together with the details on the generation: Pythia8B plus EvtGen is used for most samples except for the $B \to hh'$ peaking background and the $b\bar{b} \to \mu^+\mu^- X$ continuum background channels where only Pythia8B is used. EvtGen is used for the reference channels including a J/ψ in the final state in order to correctly account for the J/ψ polarisation effects. The $B^0_s \to J/\psi\phi$ control channel, using Pythia8B and Photos is taken from the samples produced for the $B^0_s \to J/\psi\phi$ analysis within the J/ψ analysis subgroup: the sample is generated flat from the angular point of view and maps are used to obtain the correct angular distribution. Most of the samples are processed with Atlfast-II that employs the fast detector simulation for the calorimeter by means of parameterisations of the longitudinal and lateral energy profile, while the muon and tracking parts are fully simulated. Due the less accurate simulation of calorimetry in Atlfast-II, this cannot be used to estimate the muon fake rates, so the $B \to hh'$ peaking background are processed with the full simulation in order to have an accurate description of the hadronic contributions.

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

---

[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.

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.

Channel	Type	Events	Generator	Simulation
$B_s^0 \to \mu^+\mu^-$	signal	1,000,000	Pythia8B + EvtGen	Atlfast-II
$B_{(s)}^0 \to \mu^+\mu^-$	signal	1,000,000	Pythia8B + EvtGen	Atlfast-II
$B^+ \to J/\psi K^+$ with $J/\psi \to \mu^+\mu^-$	reference	1,997,000	Pythia8B + EvtGen	Atlfast-II
$B^- \to J/\psi K^-$ with $J/\psi \to \mu^+\mu^-$	reference	1,999,500	Pythia8B + EvtGen	Atlfast-II
$B^+ \to J/\psi \pi^+$ with $J/\psi \to \mu^+\mu^-$	reference	498,000	Pythia8B + EvtGen	Atlfast-II
$B^- \to J/\psi \pi^-$ with $J/\psi \to \mu^+\mu^-$	reference	500,000	Pythia8B + EvtGen	Atlfast-II
$B_s^0 \to J/\psi \phi$ with $J/\psi \to \mu^+\mu^-, \phi \to K^+K^-$	control	5,000,000	Pythia8B + Photospp	Atlfast-II
$B \to hh'$	peaking bkg.	5,000,000	Pythia8B	full simulation
$B_s^0 \to K^-\mu^+\nu$	part. rec. bkg.	250,000	Pythia8B + EvtGen	Atlfast-II
$B_s^0 \to \pi^-\mu^+\nu$	part. rec. bkg.	500,000	Pythia8B + EvtGen	Atlfast-II
$\Lambda^0 \to p\mu^-\overline{\nu}$	part. rec. bkg.	250,000	Pythia8B + EvtGen	Atlfast-II
$b\overline{b} \to J/\psi X$ with $J/\psi \to \mu^+\mu^-$	cont. bkg.	10,000,000	Pythia8B + EvtGen	Atlfast-II
$b\overline{b} \to \mu^+\mu^- X$	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.).

Channel	#events	Dataset
Bsmumu	1,000,000	mc16_13TeV.300426.Pythia8BEvtGen_A14_CTEQ6L1_Bs_mu3p5mu3p5.merge.AOD.e4889_e5984_a875_r9364_r9315
Bdmumu	1,000,000	mc16_13TeV.300430.Pythia8BEvtGen_A14_CTEQ6L1_Bd_mu3p5mu3p5.merge.AOD.e4889_e5984_a875_r9364_r9315
BpJpsiKp	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.Pythia8BPhotospp_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.300431.Pythia8B_A14_CTEQ6L1_B_hh.merge.AOD.e4889_e5984_s3126_r9364_r9315
BsKmunu	250,000	mc16_13TeV.300432.Pythia8BEvtGen_A14_CTEQ6L1_Bs_K3p5mu3p5nu.merge.AOD.e4720_e5984_a875_r9364_r9315
BdPimunu	500,000	mc16_13TeV.300433.Pythia8BEvtGen_A14_CTEQ6L1_Bd_pi3p5mu3p5nu.merge.AOD.e4720_e5984_a875_r9364_r9315
LbPmunu	250,000	mc16_13TeV.300434.Pythia8BEvtGen_A14_CTEQ6L1_Lambda0b_p3p5mu3p5nu.merge.AOD.e4720_e5984_a875_r9364_r9315
bbmumu	49,999,000	mc16_13TeV.300306.Pythia8B_A14_CTEQ6L1_bb_mu3p5mu3p5_Py8RepDec.merge.AOD.e4911_e5984_a875_r9364_r9315
bbJpsimumu	10,000,000	mc16_13TeV.300203.Pythia8BPhotospp_A14_CTEQ6L1_bb_Jpsimu3p5mu3p5.merge.AOD.e4889_e5984_a875_r9364_r9315
bbmumuX	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_r9364_r9315_p3371
Bdmumu	913,507	mc16_13TeV.300430.Pythia8BEvtGen_A14_CTEQ6L1_Bd_mu3p5mu3p5.deriv.DAOD_BPHY8.e4889_e5984_a875_r9364_r9315_p3371
BpJpsiKp	1,610,177	mc16_13TeV.300404.Pythia8BEvtGen_A14_CTEQ6L1_Bp_Jpsi_mu3p5mu3p5_Kp_BMassFix.deriv.DAOD_BPHY8.e4862_e5984_a875_r9364_r9315_p3371
BmJpsiKm	1,617,566	mc16_13TeV.300405.Pythia8BEvtGen_A14_CTEQ6L1_Bm_Jpsi_mu3p5mu3p5_Km_BMassFix.deriv.DAOD_BPHY8.e4862_e5984_a875_r9364_r9315_p3371
BsJpsiPhi	3,555,942	mc16_13TeV.300438.Pythia8BPhotospp_A14_CTEQ6L1_Bs_Jpsimu3p5mu3p5_phi.deriv.DAOD_BPHY8.e4922_e5984_a875_r9364_r9315_p3371
BpJpsiPip	399,607	mc16_13TeV.300406.Pythia8BEvtGen_A14_CTEQ6L1_Bp_Jpsi_mu3p5mu3p5_Pip_BMassFix.deriv.DAOD_BPHY8.e4862_e5984_a875_r9364_r9315_p3371
BmJpsiPim	405,710	mc16_13TeV.300437.Pythia8BEvtGen_A14_CTEQ6L1_Bm_Jpsi_mu3p5mu3p5_Pim_BMassFix.deriv.DAOD_BPHY8.e4862_e5984_a875_r9364_r9315_p3371
Bhh	4,002,651	mc16_13TeV.300431.Pythia8B_A14_CTEQ6L1_B_hh.deriv.DAOD_BPHY8.e4889_e5984_s3126_r9364_r9315_p3371
BsKmumu	15,786	mc16_13TeV.300432.Pythia8BEvtGen_A14_CTEQ6L1_Bs_K3p5mu3p5mu.deriv.DAOD_BPHY8.e4720_e5984_a875_r9364_r9315_p3371
BdPimumu	30,932	mc16_13TeV.300433.Pythia8BEvtGen_A14_CTEQ6L1_Bd_pi3p5mu3p5mu.deriv.DAOD_BPHY8.e4720_e5984_a875_r9364_r9315_p3371
LbPmumu	15,768	mc16_13TeV.300434.Pythia8BEvtGen_A14_CTEQ6L1_Lambda0b_p3p5mu3p5mu.deriv.DAOD_BPHY8.e4720_e5984_a875_r9364_r9315_p3371
bbmumuX	14,666,736	mc16_13TeV.300306.Pythia8B_A14_CTEQ6L1_bb_mu3p5mu3p5_Py8RepDec.deriv.DAOD_BPHY8.e4911_e5984_a875_r9364_r9315_p3371
bbJpsimumuX	6,807,750	mc16_13TeV.300203.Pythia8BPhotospp_A14_CTEQ6L1_bb_Jpsimu3p5mu3p5.deriv.DAOD_BPHY8.e4889_e5984_r9364_r9315_p3371

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

5 MCTuning

In this section we describe how the Monte Carlo samples have been tuned and cross-checked against data. The tuning procedure is two-fold: a first set of weights ("Quark Level Corrections" or QLC) accounts for the (pT_B, η_B) selection bias introduced to expedite the MC generation of b quarks; a second set of weights ("Data Driven Weights" or DDW) accounts for residual data-MC (pT_B, η_B) discrepancies. All MC samples will be consistently used across these studies after applying both sets of complementary corrections: Quark Level and Data Driven, which are meant to be used in combination in order to obtain meaningful responses from the MC, with a clear way of accounting for residual discrepancies through systematic uncertainties.

5.1 Quark Level Corrections

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

The QLC are evaluated using two different MC samples, generated with looser quark-level cuts with respect to the default MC and no cuts on the final state particles ((*unbiased* MC), and with the same quark-level cuts as the default MCs and without final state particle cuts (*quark biased* MC). Since QLC are meant to correct generator-level biases, the production of the unbiased and quark biased MCs is limited to generation, without simulation of the detector response and reconstruction; table 4 shows the sets of cuts applied at generation level for the different MCs.

	$\hat{pT}min$	anti-b η	anti-b pT	muons η	muons pT	final h η	final h pT
default $B^+ \to J/\psi K^+$	7 GeV	2.6	7 GeV	2.6	3.5 GeV	2.6	900 MeV
unbiased $B^+ \to J/\psi K^+$	5 GeV	4	2.5 GeV				
quark biased $B^+ \to J/\psi K^+$	7 GeV	2.6	7 GeV				
default $B_s \to \mu^+\mu^-$	5 GeV	2.6	5 GeV	2.6	3.5 GeV		
unbiased $B_s \to \mu^+\mu^-$	5 GeV	4	2.5				
quark biased $B_s \to \mu^+\mu^-$	5 GeV	2.6	5 GeV				
default $B-s \to J/\psi\phi$	11 GeV	2.5	9 GeV	2.6	3.5 GeV	2.6	0.8 GeV
unbiased $B-s \to J/\psi\phi$	5 GeV	4	2.5 GeV				
quark biased $B-s \to J/\psi\phi$	11 GeV	2.5	9 GeV				

Table 4: Quark-level and final state particle cuts per MC sample.

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

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

164 Where ν is the number of entries in a (pT_B, η_B) from the unbiased or quark biased samples after applying
165 the final state particle cuts. The term $\frac{\sigma^{Pythia}}{N^{tot}}$ is used to normalise the two MCs to the same integrated
166 luminosity: σ^{Pythia} is the Pythia generation cross-section and N^{tot} is the total number of generated
167 events.
168 The inverse of these weights should be used to weight events individually, thus correcting with event-
169 weights the QL cut biases.
170

171 Figure **??** show the QLC calculated for B^+, $B_s \to \mu\mu$, and $B_s J/\psi\phi$ and their uncertainty. This computation
172 of the QLC has been cross-check applying QLC calculated using odd events, from both the quark biased
173 and unbiased samples, to the even events of the quark biased sample. The resulting distributions have
174 been compared to unbiased distributions obtained using only even events. Figure 1 shows the checks
175 performed on the B^+ sample.

4: B+ QLC ready, BsMuMu and BsJpsiPhi QLC still work in progress

5.2 Data Driven Weights

177 The Data Driven Weights (DDW) have been evaluated in a similar way as it has been done in [6] and
178 [4].[3]

179 The DDW for the MC signal samples are determined with an iterative method, by comparing sideband-
180 subtracted $B^\pm \to J/\psi K^\pm$ odd-numbered events with MC events, after the latter are corrected with the
181 W_{QLC} weights.[4]

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

$$W_{DD}(p_T, |\eta|) = w(p_T) \cdot w(\eta) \tag{2}$$

184 where $w = \nu^{data}/(\nu^{MC} \times W_{QLC})$ and ν is the normalised number of entries in either the data or MC
185 histograms for p_T or η. These weights are then used as per-event weights on the MC events, and the
186 procedure is iterated until these weights stabilise. The stability of the weights can be seen from the
187 convergence of the second iteration and it is shown in Figure .....

6: calculaton of DDW under-way

---

[3] Due to the limited statistics of the B^+ data sample, the DDW corrections are not computed in a two-dimensional grid of the variables p_T and $|y|$, but are extracted with the one-dimensional times one-dimensional procedure described in the text. This approach was already used in the analyses [6] based on the 4.7 fb^{-1} sample of data collected in 2011 and [4] based on the full Run1 sample. The validity of the approach is tested by the stability of the recursive procedure, and fundamentally it works because the production cross-section depends strongly on p_T, but much less on $|y|$, with small correlation between the two variables. The systematic deviation in the computation of the average acceptance times efficiency due to the method chosen for this analysis rather than a true two-dimensional computation of the DDW has been estimated to be at the level of ....

5: still work in progress

[4] The extraction followed similar lines used in [6] and [4]. The B^+ yeld was extracted in p_T intervals and alternatively, in $|\eta|$ intervals. The event selection followed the same line discussed in the following sections of this note, with the addition of an additional cut $L_{xy} > 0.3$ mm on the transverse separation between primary and secondary vertex, and without applying the multivariate selection against combinatorial background. The signal is described with two Gaussians with equal mean; the continuum background is described by an exponential; the background due to partially reconstructed decays (B to $J/\psi X$, with m($J/\psi h^+ < 5.200$ GeV) is described with an error function. All shape and amplitude parameters are extracted form the fit, and the uncertainty in the signal yield is dominated statistical errors.
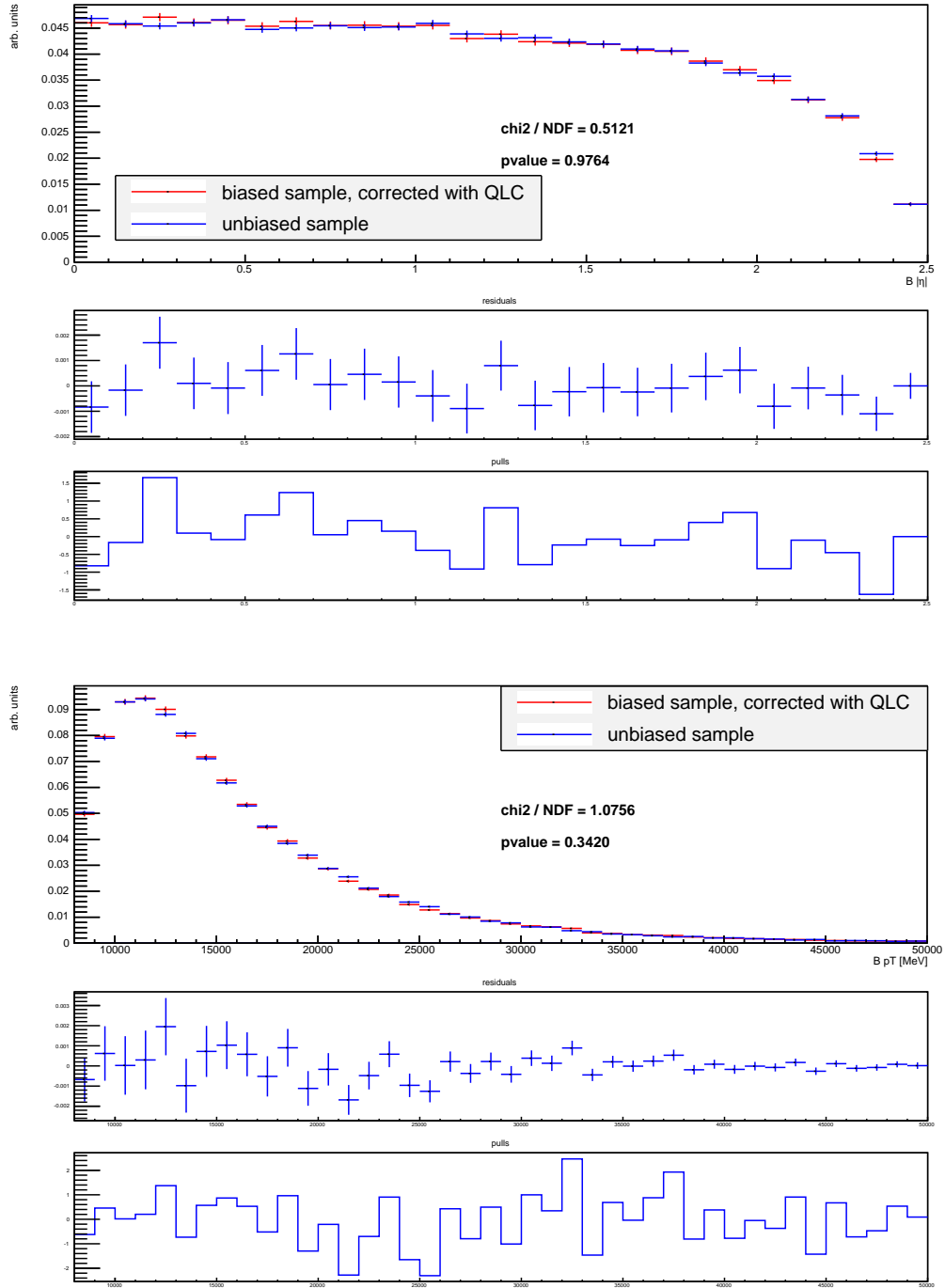
Figure 1: Plots show the comparison of the η_B (first plot) and pT_B (second plot) QLC corrected quark biased distribution using the two samples approach QLC and the unbiased distribution. In order to avoid correlations between the distributions, QLC have been calculated using odd numbered events from the unbiased and quarkBiased samples, and the remaining events in the two samples have been weighted and used for the comparison.

6 Candidate Preselection

6.1 Candidate building and preselection on derivation level

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

- $B^0_{(s)} \to \mu^+\mu^-$ signal channel

  A $B^0_{(s)}$ candidate is build from two oppositely-charged muons. Before vertexing the muon momenta are corrected by the MuonCalibrationAndSmearingTool as recommended by the muon combined performance group, applying the appropriate configurations[5] for data15 and data16. Using the JpsiFinder tool, the $B^0_{(s)}$ candidates are constructed by a fit of the two muon's inner detector tracks to a common vertex requiring $\chi^2_{B^0_{(s)}}/NDF < 15$ and $3500 \text{ MeV} < m_{B^0_{(s)}} < 7000 \text{ MeV}$. In addition, the so called muon-based mass $m^{\text{MUCALC}}_{B^0_{(s)}}$ and the corresponding uncertainty are calculated as the invariant mass from the four-momentum sum using the combined muon information of the two muons comprising the candidate. The PDG-mass for muons is applied in this calculation. On data, the candidates two mass values are <mark>blinded</mark>[6] if both, $m_{B^0_{(s)}}$ and $m^{\text{MUCALC}}_{B^0_{(s)}}$ fall into the blinded region around the nominal B^0_s, i.e. into the interval from 5166 MeV to 5526 MeV.

- $B^\pm \to J/\psi K^\pm$ reference channel with $J/\psi \to \mu^+\mu^-$

  First J/ψ candidates are reconstructed using the JpsiFinder similarily as for the $B^0_{(s)} \to \mu^+\mu^-$ signal channel with an adjusted mass range of $2000 \text{ MeV} < m_{J/\psi} < 7000 \text{ MeV}$. Then the JpsiPlus1TrackFinder tool is used to combine the J/ψ candidates with an additional track[7] which is required to have $p^{Jpsi}_{\text{T}} > 1 \text{ GeV}$ and $|\eta < 2.5|$. A mass constraint of the J/ψ candidate's mass to the J/ψ PDG mass is applied. The $B^\pm$ candidates are accepted if $p^{B^\pm}_{\text{T}} > 1 \text{ GeV}$, $\chi^2_{B^\pm}/NDF < 15$ and $3500 \text{ MeV} < m_{B^\pm} < 7000 \text{ MeV}$. Again, muon-based mass values $m^{MUCALC}_{B^\pm}$ and corresponding uncertainties are calculated from the four-momenta of the two muons and the additional kaon track, assuming PDG values for the masses of the muons and the kaon as appropriate. No mass-value blinding is applied to the reference channel candidates.

- $B^0_s \to J/\psi\phi$ control channel with $J/\psi \to \mu^+\mu^-$

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

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

---

[5] 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.

[6] 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.

[7] 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 $B \to hh'$ 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 $B^+ \to J/\psi \pi^+$ Monte Carlo sample, two instances of the $B^\pm \to J/\psi K^\pm$ 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 $B^\pm \to J/\psi K^\pm$ reference sample on data will contain a contribution of $B^+ \to J/\psi \pi^+$ 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 6.2), 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 14. 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 acount. 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 longtidinal 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 14 for the definition of L_{xy}.

- Proper decay time τ and its uncertainty

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

6.2 Primary vertex determination

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

- PV_MAX_SUM_PT2: predefined in ATLAS, considers the sum of the squared transverse momentum of the tracks associated to each PV, the chosen PV is the one with the highest sum;

- PV_MIN_A0: a backward extrapolation of the B momentum from the decay vertex is considered, the PV is chosen as the one with the shortest 3D distance from the point of closest approach (POCA) of the B extrapolation to each of the reconstructed PVs;

- PV_MIN_Z0: similar to PV_MIN_A0, but uses the distance along z from the POCA of the B extrapolation to each of the reconstructed PVs;

- PV_MIN_Z0_BA: approach developed for full Run1 analysis, the associated PV is chosen as the one with the shortest separation, along z, from the POCA of the B extrapolation to the beam line.

The performances of the four PV association procedure have been tested on signal MC sample.
A first comparison was performed using only the information regarding the coordinates of the selected PVs and the position of the truth PV.
Figures 2 and 3 show respectively the distance on the xy plane and along the z direction between the selected PV and the truth PV for the four approaches. PV_MAX_SUM_PT2 shows clearly large distributions, which tells that the selected PV is often the wrong one. The other three approaches show almost identical narrow distributions.
In order to have a quantitative estimation of the performance of the approaches, for each PV reconstructed in the events, a χ^2 is computed to estimate the compatibility with the MC truth. The PV with the lowest χ^2 in each event is found to correspond to the PV correctly associated to the B candidate and is considered "truth-matched". Figure 4 shows the distribution of the χ^2 of all the PVs (blue points) and the distribution of the χ^2 of the "truth-matched" PVs. Figure 5 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 5 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 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

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

proxy for the pile-up). Figure 6 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).
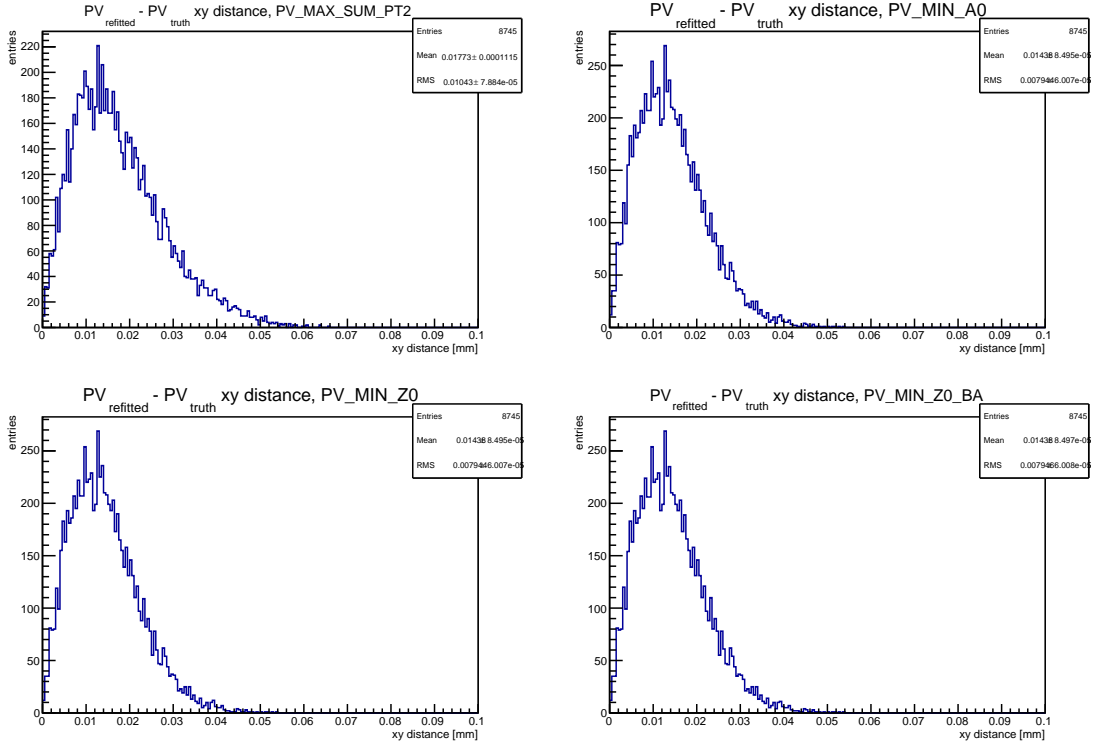
Figure 2: xy distance between selected PV and truth PV for the four approaches.
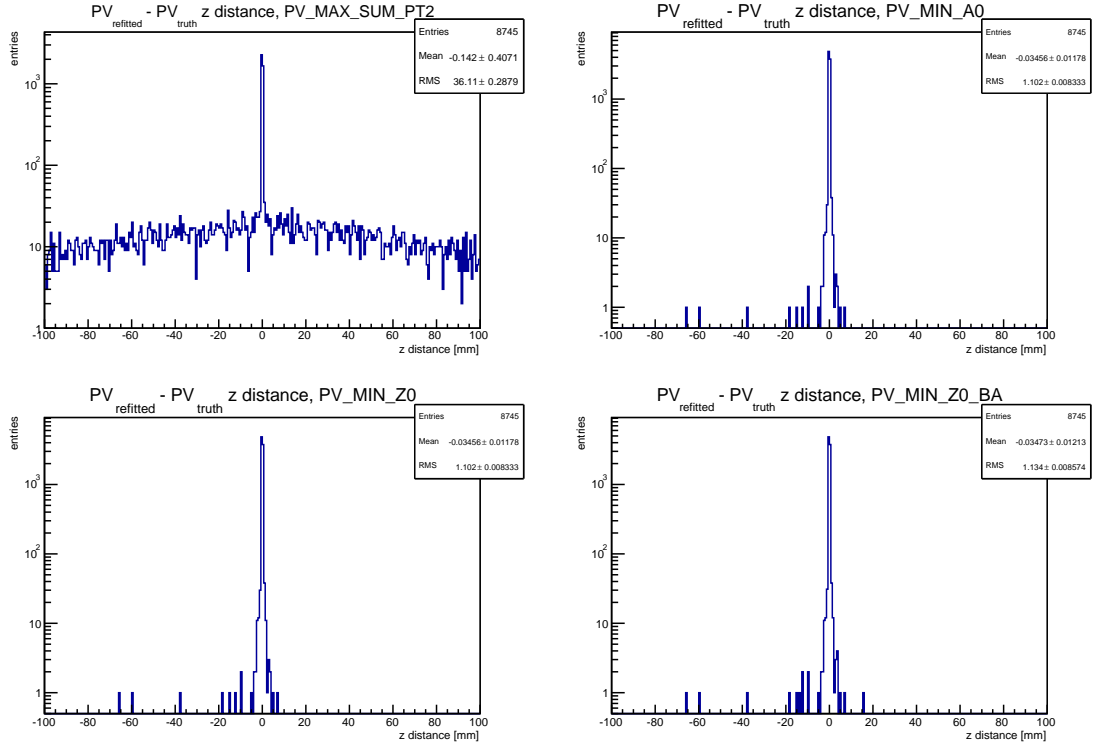
Figure 3: z distance between selected PV and truth PV for the four approaches.

Figure 4: χ^2 of all the PVs (blue points) and the distribution of the χ^2 of the "truth-matched" PVs.

Figure 5: χ^2 distribution of the chosen PVs using the 4 approaches (green distributions) superimposed to the truth-matched distribution (red).

Figure 6: purity as a function of number of reconstructed PVs for the four approaches.

7 Trigger

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

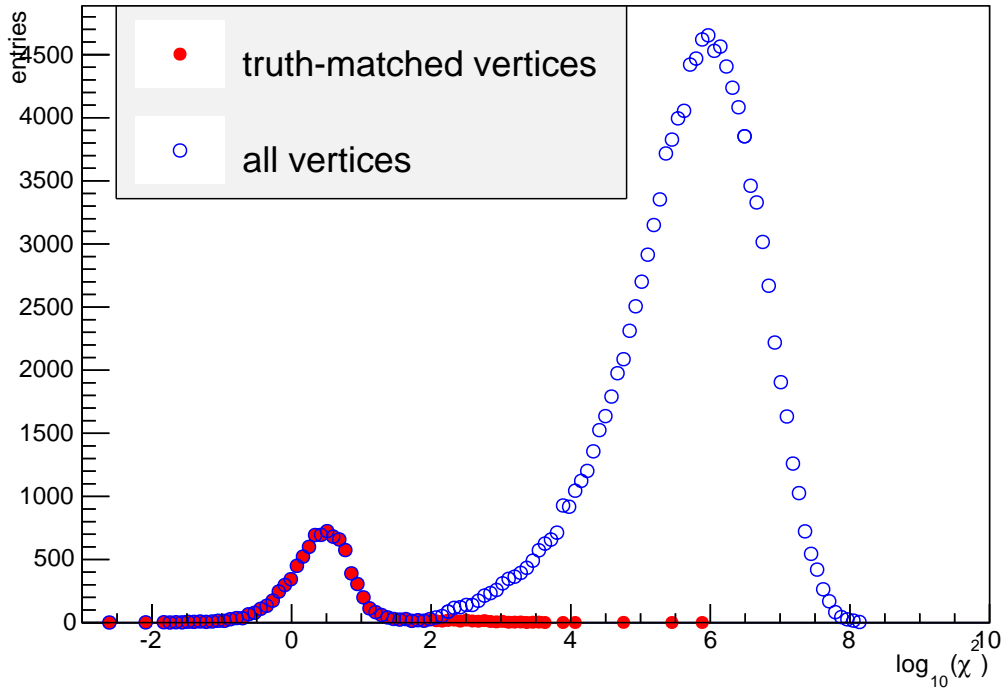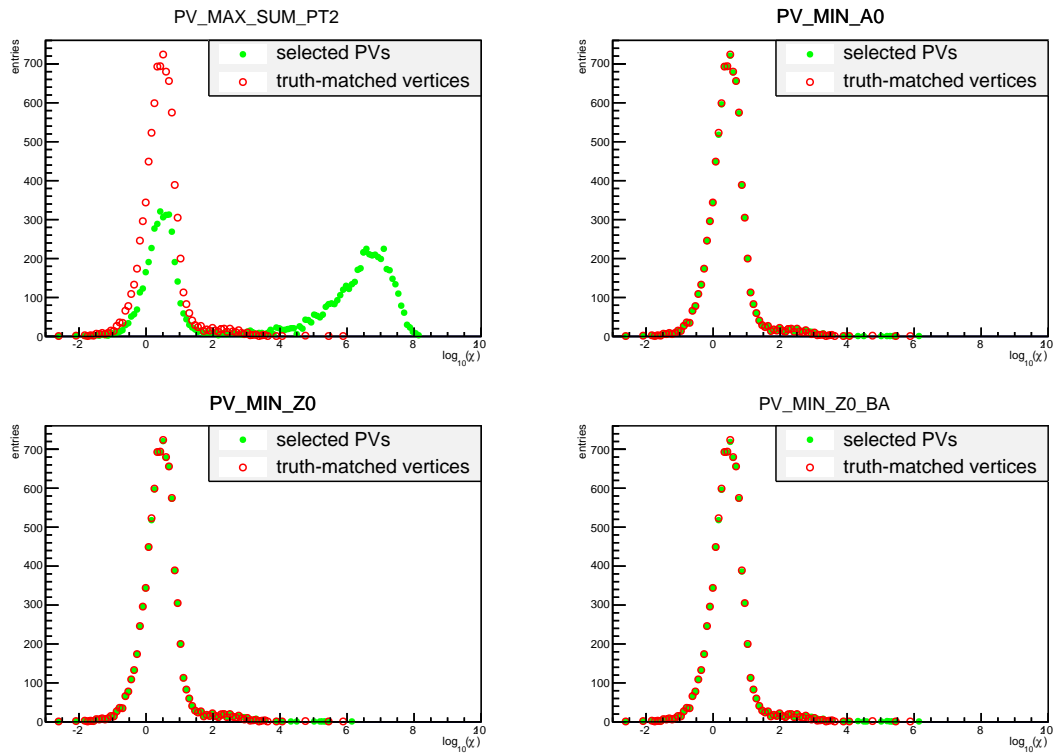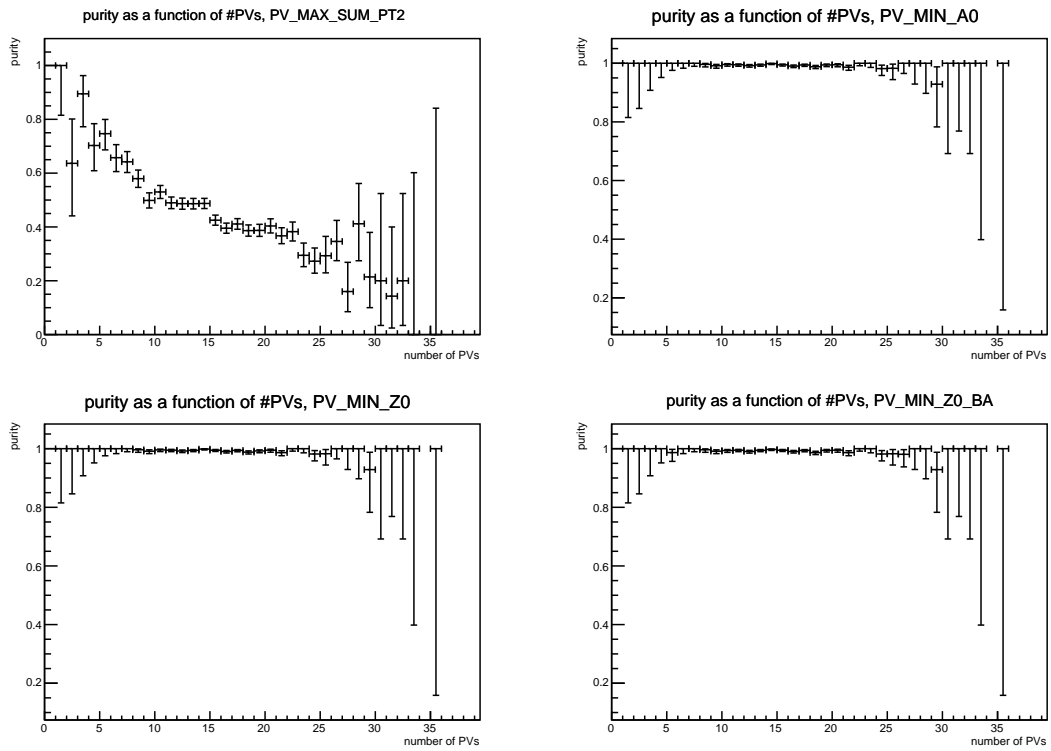- *2mu4* prescaled (2015) or not active (2016)

- *mu6_2mu4* prescaled during 2016, $\sim 82\%$ efficient w.r.t. 2mu4

- *2mu6* unprescaled, $\sim 35\%$ efficient w.r.t. 2mu4

The collected luminosity of each trigger per year are shown in table 6.

	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 6: Collected luminosity in 2015 and 2016.

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

- consider mu6mu4 only would get an effective collected luminosity of 24.6 fb^{-1}

- consider the combination of mu6mu4 and 2mu6 would get an effective collected luminosity of 28.5 fb^{-1}

- 2mu4 addition would increase the 2mu6+mu6mu4 statistics of a factor $\sim 1\%$

The prescaled trigger mu6mu4 provides the bulk of the statistics, which is $\sim 85\%$ of total. The addition of 2mu6 would allow to use basically the full dataset but would also increase the complexity of the analysis, given the tight schedule, this round of the analysis will use mu6mu4 only. still have to add the description of the different HLT algorithms (Lxy0, dalayed)

8: add

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

1. **B production cross section with respect to Run-1**, due to the Run-2 increased center of mass energy a value ~ 1.7 times higher is expected, according to studies performed using FONLL .....

9: to be added in file bib

2. **The efficiency of the dimuon triggers available in Run-2 with respect to the Run-1 triggers**, already shown in this section.

3. **The collected luminosity for the selected trigger in Run-2**, an integrated luminosity of $\approx 40\text{fb}^{-1}$ has been collected during the 2015 and 2016 data-taking period.

330 The same signal over background ratio as the analysis performed on the Run-1 dataset is assumed for this
331 study.
332 With the three ingredients listed above and the previous assumptions we estimate the expected statistics
333 available for the 2015-16 dataset by scaling the Run-1 analysis statistics. This yields an estimated 2-fold
334 increase with respect to Run-1 statistics.

335 # 8 CutFlow

9 Studies on muon fake rates

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

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

Table 7 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 7: 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 edmixtures 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.

muon efficiency to the Run1 BDT; their usage would imply an increase of about ∼ ×1.6 in the number of

358 fakes than using a Run1-like BDT, for about the same muon efficiency.

359 In order to understand the effect of the tight muons usage, we compared the effect of a Run1-like BDT

360 and the tight muons in the analysis.

361 Knowing the available statistics before and after the fake-BDT application in previous analysis [4] and

362 the muon efficiency and fake-rate, we computed the Run1 background composition in terms of fake and

363 real muons; applying the properties of tight muons we can estimate the effect tight muons would have

364 had on the Run1 analysis. Tables 8 and 9 show the amount of statistics for the different components

365 of the likelihood used in the fit for the signal yield extraction and the estimation of the statistics using

366 tight muons. The combinatorial background doesn't increase significantly, the SS-SV (ans semileptonic)

367 background increases of a factor $\sim \times 1.1$ and the paking background increases of a factor $\sim \times 2.6$. Signal

368 reduction due to lower muon efficiency is almost negligibe.

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 8: breakdown of the background contributions in the Run1 likelihood after the BDT application. Columns labelled tight mu show the estimated statistcs 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 9: Breakdown of the expected signal and peaking background contributions in the Run1 likelaihood after the BDT application. Columns labelled tight mu show the estimated statistcs we would have gotten using tight muons instead of the BDT.

369

370 Run1 analysis likelihood with the available statistics modified according to the estimation of the available

371 stastics in this analysis performed in section 7 has been run, varying the number of events associated to

372 the different background and signal models according to the usage of tight muons or a Run1-like BDT.

373 Table 10 shows the root mean square of the distribution of the fitted number of B_s and B_d events. No

374 significant effect on analysis sensitivity is visible for B_s, while a $\sim 1\%$ broader RMS is visible on B_d

375 count.

376 Given the negligible difference due to the usage of tight muons instead of a Run1-like BDT for fake muons

377 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 10: RMS of distributions of number of fitted B_s and B_d events in toy-MC study. The columns labelled Run1 refer to the statistcs 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 7.

10 Background Modeling

10.1 Non-resonant Background

The measurement of the rare decay of B mesons into muon pairs requires the rejection of a very large combinatorial background. After the preliminary and additional selection, the amount of combinatorial background is at the level of ≈140 k events per 100 MeV interval in the muon pair invariant mass (this is shown in the context of the normalisation study for the signal fit in Figure **??**). This figure should be compared to ≈20 events/100 MeV expected at the peak of the B_s^0 signal (section 15.3).

To have a statistically meaningful sample of this combinatorial background, a very large number of events needs to be generated (one of the largest MC production performed so far by ATLAS for a single analysis), thus specific procedures have been used in order to reduce the event generation time, as discussed below. The MC sample was generated inclusively to provide a realistic composition of muons from different sources, including non combinatorial sources of muon pairs. On the other hand, because of uncertainties in cross sections and branching ratios, as well as for those related to the generation procedures, no attempt is made to use the MC to draw quantitative conclusions on the amount in which each type of background is present in real data. For that purpose, events collected in the sidebands are used. As shown in section 15.3, after the background reduction developed on MC is applied, the remaining background in real data is sufficiently low and separated in its different components to be used for an effective interpolation in the signal region.

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

Repeated hadronisation, performed in Pythia, was applied with 10 repetitions. This is technically implemented by means of PythiaB package [8] and had already been widely tested and used for various MC production in B physics. This procedure scarcely affects the characteristics of the generated samples.

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

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

10:
numbers
t.b.u.

11:
numbers
t.b.u.

12:
numbers
t.b.u.

13:
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 four-corner MC sample).

14: numbers t.b.u.

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.

10.2 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 11 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 11: 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 7(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 7(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-]]*.

10.2.1 Truth Matching

Figures 7 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:

$$\left| 1 - \frac{p_T^{truth}}{p_T^{reco}} \right| < 0.15 \tag{3}$$

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.
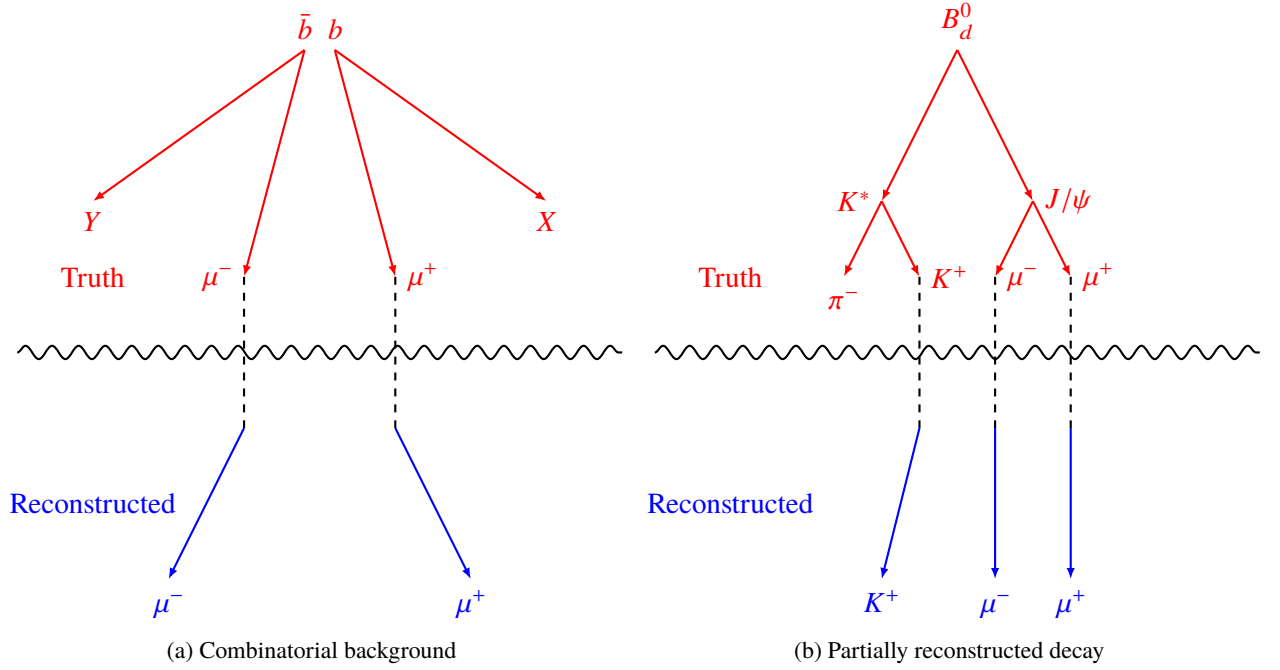
(a) Combinatorial background   (b) Partially reconstructed decay

Figure 7: Both diagrams represent a reconstructed B_s^0 or B^+, which in reality are combinatorial background or a partially reconstructed decay.

10.2.2 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 8 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 $|\eta| < 2.5$, $p_T > 4 GeV$.

- **Kaons:** Must pass loose requirements and $|\eta| < 2.5$, $p_T > 1 GeV$.

- **Candidates:** $|\eta| < 2.5$, $p_T > 8 GeV$, $\chi^2 < 6$.

Another thing that we might want to know is the number of decays present in the samples. Figure 9 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 $b\bar{b} \to \mu^+\mu^- + X$ sample contains very few $B_s^0 \to \mu^+\mu^-$ candidates and mostly $B^+ \to J/\psi + K^+$ and $B_s^0 \to J/\psi + \phi$ 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 7(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 $B_s^0 \to J/\psi + \phi$ candidates than to the $B^+ \to J/\psi + K^+$ candidates.

Table 12 shows the main sources of background for $B^+ \to J/\psi + K^+$. Table **??** shows the same for $B_s^0 \to J/\psi + \phi$. 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
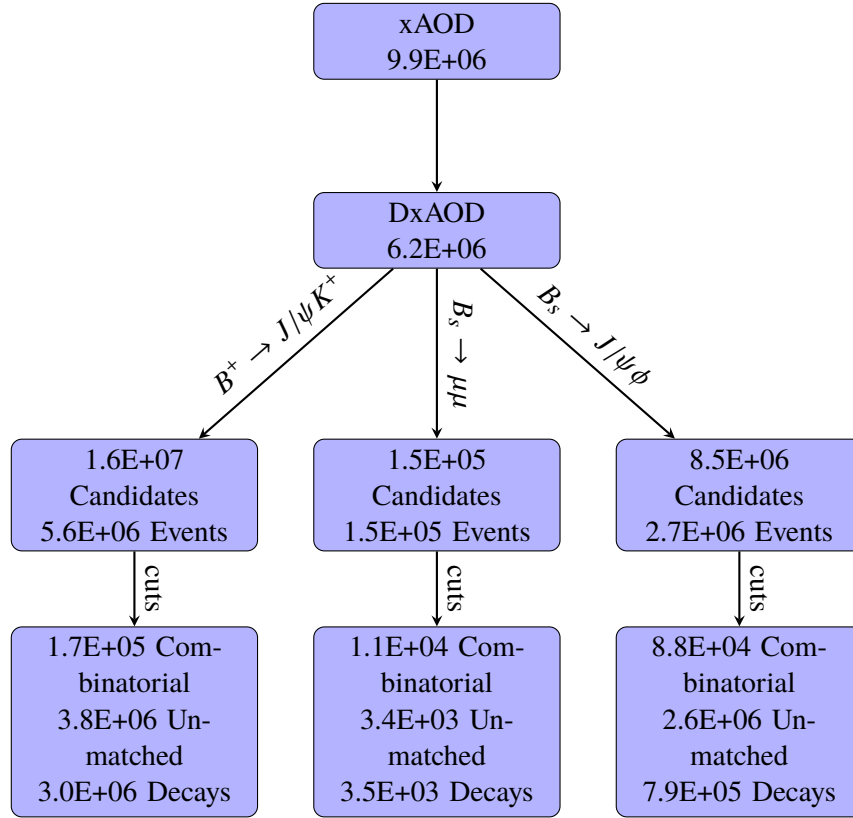
Figure 8: Cutflow corresponding to $b\bar{b} \rightarrow J/\psi + X$ sample.

removed 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 10.

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 12: Processes most commonly misidentified as $B^+ \to J/\psi + K^+$.

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-:pi-:gamma:K+:K+]Jpsi[mu+:mu-]]	11022

Table 13: Processes most commonly misidentified as $B_s^0 \to J/\psi + \phi$.

(a) $B_s \to \mu\mu$.

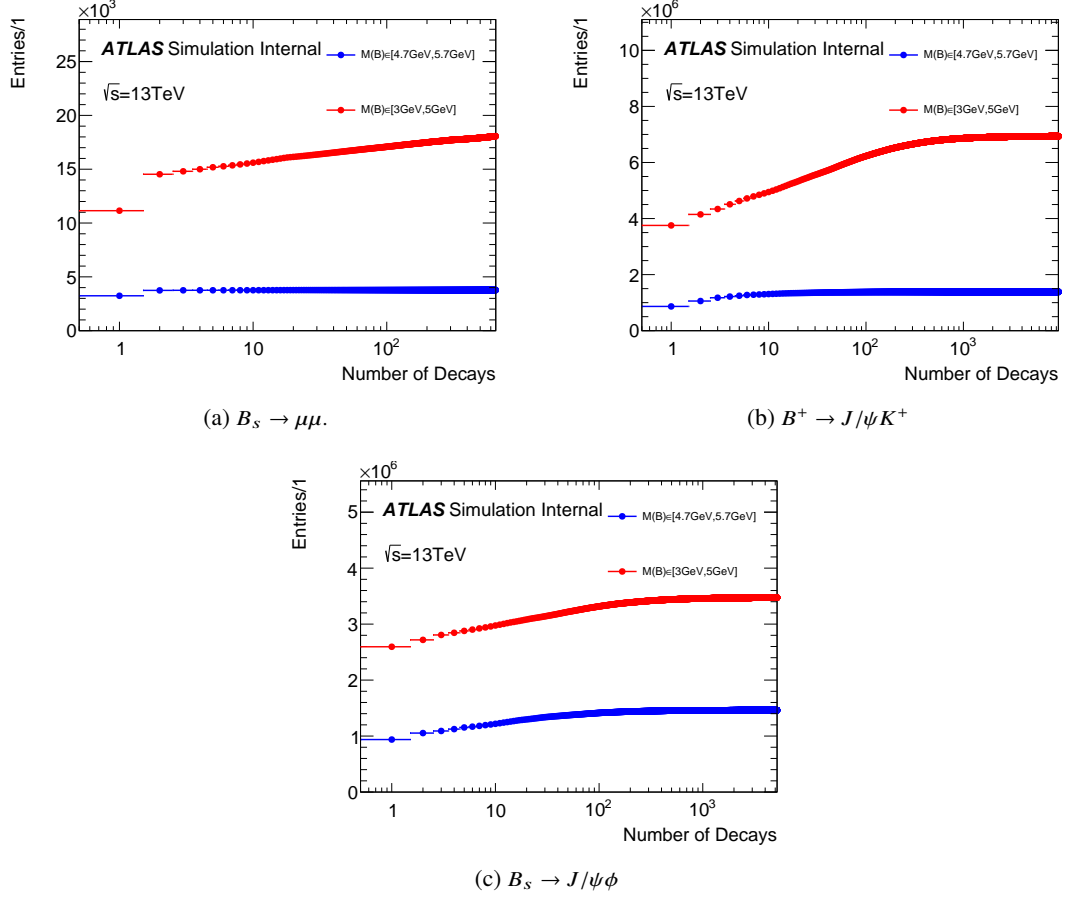(b) $B^+ \to J/\psi K^+$

(c) $B_s \to J/\psi\phi$

Figure 9: 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.
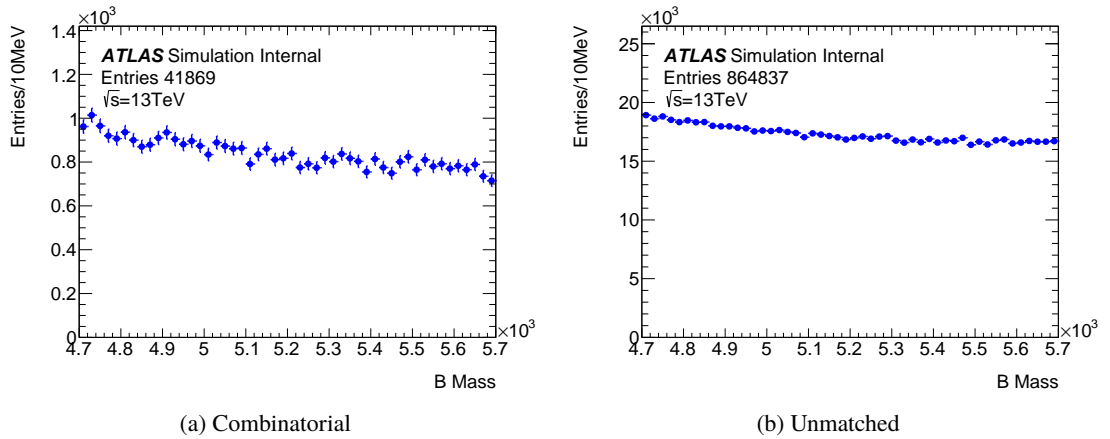

(a) Combinatorial

(b) Unmatched

Figure 10: The figures show the mass distributions of *B* candidates classified as *combinatorial* and *unmached*. Both distributions are identical in shape.

11 ContinuumBDT

In order to discriminate against the continuum background we employ a Multi-Variate Analysis (MVA) strategy, based on a boosted decision tree (BDT) algorithm, as implemented in the TMVA [9] package of ROOT. We use 15 physical input variables to obtain the signal-to-background discriminator. These 15 input variables are summarised in Table 14 along with a short description of each variable. The BDT ranking of input variable importance is given in Table 15 and compared to the separation power of the signal and background.

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

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

To check the isolation of the signal decay we look at non-signal tracks in the vicinity of reconstructed B vertices, excluding the tracks associated to the pile-up vertices. The procedure of computation of the variable, outlined in section 6.1 requires information from all tracks in an event and therefore performed on the derivation level.

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

15: t.b.c.

The other 12 variables in Table 14 are related to B-decay topology and kinematics. They were chosen from the larger pool of variables as the ones ensuring the best BDT performance. The variables that have been considered but dismissed are listed below for future reference: they do not contribute significantly to the signal/background separation, either by nature or due to the presence of other strongly correlated but more powerful variables.

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

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

Appendix **??** summarises the correlation matrices calculated for signal and background events for the 15 discriminating variables used in this analysis. The BDT training is done using the $b\bar{b} \rightarrow \mu^+\mu^- X$

Variable	Description
$|\alpha_{2D}|$	Absolute value of the angle in the transverse plane between $\Delta\vec{x}$ and $\vec{p}^{\,B}$
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 $\vec{p}^{\,B}$
p_T^B	B transverse momentum
$\chi^2_{\mu,xPV}$ $\log[\chi^2_{\text{PV,SV}}]_{xy}$	Minimum χ^2 between the muon candidates and any PV separation between production (PV) and decay (SV) vertices, $\Delta\vec{x}^T \cdot \left(\sigma^2_{\Delta\vec{x}}\right)^{-1} \cdot \Delta\vec{x}$, 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
$\text{DOCA}_{\mu\mu}$	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 $(\Delta\vec{x} \cdot \vec{p}^{\,B})/|\vec{p}_T^{\,B}|$
P_L^{min}	Minimum momentum of the two muon candidates along the B direction

Table 14: Description of the 15 discriminating variables used in the discrimination between signal and continuum background.

background MC (described in Sec. 10) and signal MC events. All selection cuts are applied to both the combinatorial events and the signal MC events.

The training is done using the TMVA tool, splitting the two input samples into equal halves. The first half is used for training the BDT and the second for validation. The 1/4th of the signal MC was kept off the training and validation samples and it has been used for evaluation, together with the data candidates

Imp. Rank	Variable	Importance	Sep. Rank	Separation
1	$|\alpha_{2D}|$	0.13780	1	0.57930
2	ΔR	0.10620	2	0.56750
3	B Isolation ($I_{0.7}^{B}$)	0.10270	3	0.43390
4	$\log(\chi_{xy}^2)$	0.08076	9	0.18830
5	p_T^B	0.07600	14	0.00967
6	$d_{min,sig}^0$	0.06921	4	0.32230
7	$|\Delta\phi_{\mu\mu}|$	0.06478	12	0.01947
8	$|IP_B^{3D}|$	0.06255	8	0.26380
9	$DOCA_{\mu\mu}$	0.05848	10	0.10090
10	L_{xy}	0.05694	6	0.29220
11	N_{trk}^{clos}	0.05230	5	0.30720
12	$d_{max,sig}^0$	0.05211	11	0.07930
13	$DOCA_{xtrk,loose,pt>0.5}$	0.04400	7	0.26980
14	P_L^{min}	0.03603	13	0.01274

Table 15: 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.

from the right mass sideband.[8] The left plot in Figure 11 shows the test for over-training, where the BDT outputs for training and validation samples are shown, confirming that the BDT is not over-trained as the training and validation samples are in good agreement with each other. The right plot contains the training result via the Receiver Operating Characteristic (ROC) curve and the comparison with the BDT used in 2012 analysis: the new training results to be significantly more efficient.

16: t.b.d.

Parameter	Value
NTrees	500
MinNodeSize	1.0[%]
MaxDepth	3
BoostType	AdaBoost
AdaBoostBeta	0.5
UseBaggedBoost	True
BaggedSampleFraction	0.6
SeparationType	GiniIndex
nCuts	100
NormMode	EqualNumEvents

Table 16: Configuration parameters of the *continuum*-BDT.

The final choice of the configuration parameters of the *continuum*-BDT, as used by TMVA, is shown in Table 16. 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

---

[8] To clarify, the signal MC sample events are used: 37.5% for training, 37.5% for validation, and 25% for evaluation.

524 figure of merit of classifier performance. The *MaxDepth* parameter has also been studied in the similar
525 way. For this parameter, it has been found that the performance of classifier improves with increasing of
526 *MaxDepth*, whereas the discrepancy between its performance on the training and testing sample, accessed
527 by the Kolmogorov-Smirnov test, becomes larger, leading to the over-training. In order to not compromise
528 the generality of the classifier, the value *MaxDepth=3* has been chosen. The value of *NTrees* has been
529 chosen large enough to allow the training to converge. The choice of other parameters does not have any
530 significant impact on the training result.

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

534 In addition we tested a BDT variable built without the use of the isolation variable as input: the study
535 is described in Appendix **??**. As expected, the separation power is reduced thus the BDT configuration
536 described in this section is the one used in the analysis.
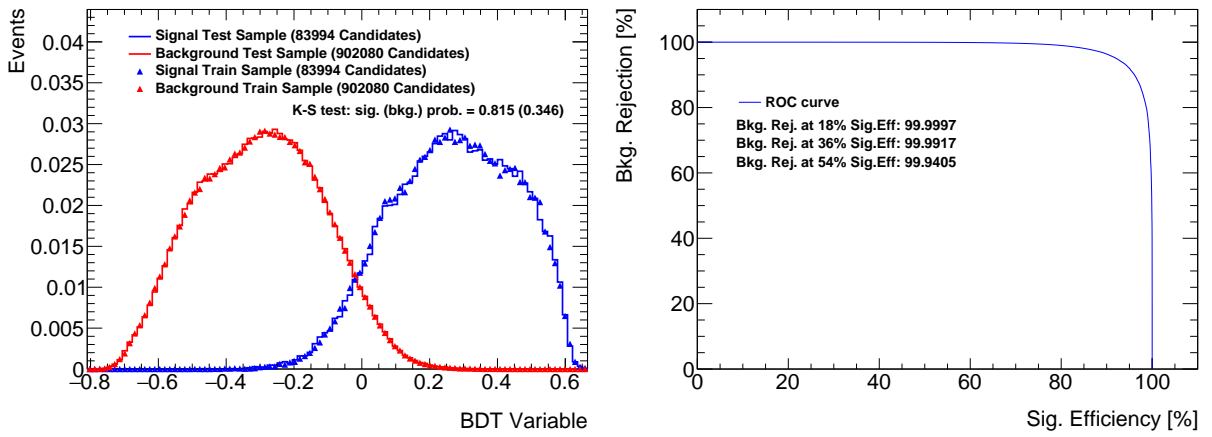
17:
t.b.c.

Figure 11: *Left*: cross-checks for over-training for the continuum BDT variable. *Right*: **Only ROC curve at the moment!** comparison of ROC curves trained on the combinatorial background MC and the signal MC. The ROCs are evaluated on the high-mass data sideband.

12 DataMCComparison

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

12.1 Continuum events

To check the shapes of the discriminating variables for the combinatorial background we compare the signal MC sample with the data sidebands. The signal sample is re-weighted on PV and separately on the $p_T{}^B$, $\eta{}^B$ and B Isolation variables simultaneously with a gradient boosted re-weighting technique [10]. Figure 12 shows the corrected p_T, η, and B Isolation variables as a cross-check.

Figure 13 shows the distributions of few discriminating variables. The rest can be checked in Appendix ??.
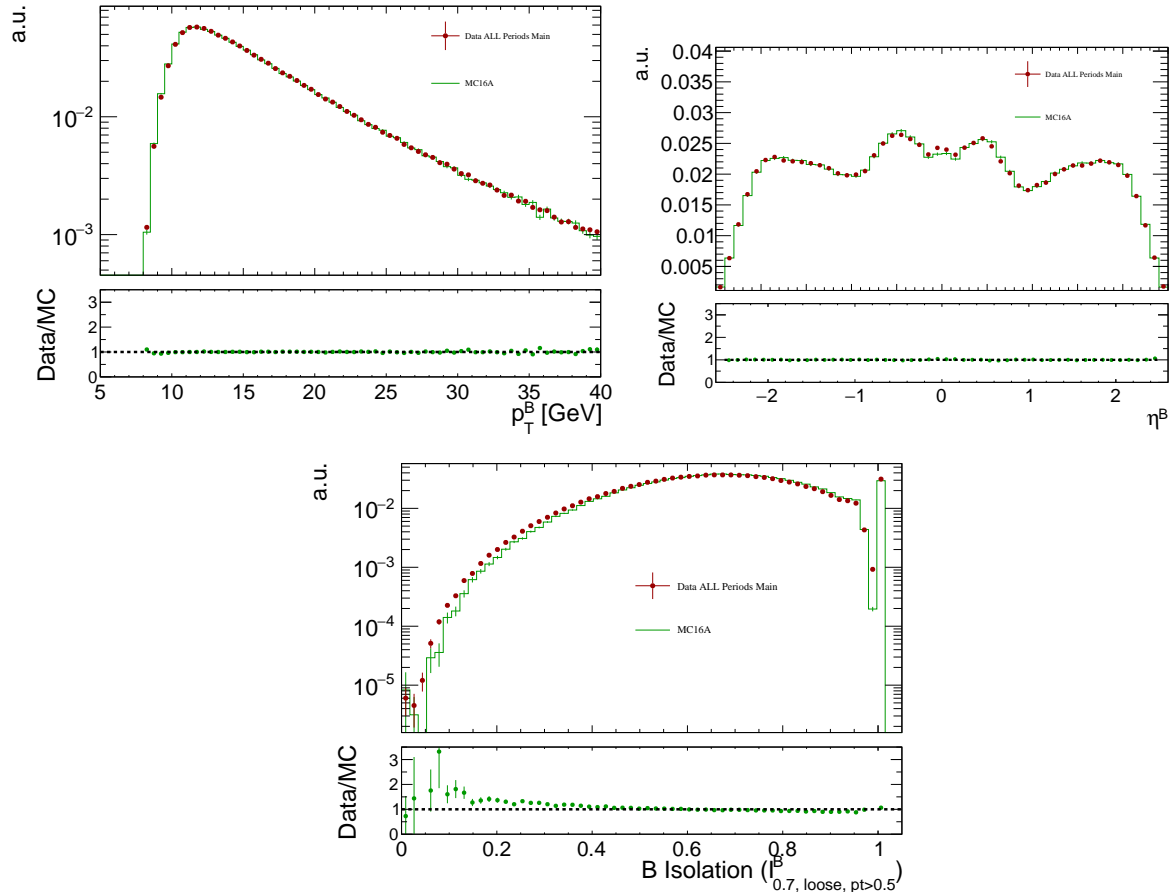
18: t.b.d.

Figure 12: Cross-check on the $p_T{}^B$ and η^B distributions of the $J/\psi K$ candidates in data and signal MC. The red dots correspond to the sideband data during all data taking periods in Run 2 and the green histogram corresponds to the MC $b\bar{b} \to \mu\mu X$ sample. Both histograms are normalised to one.
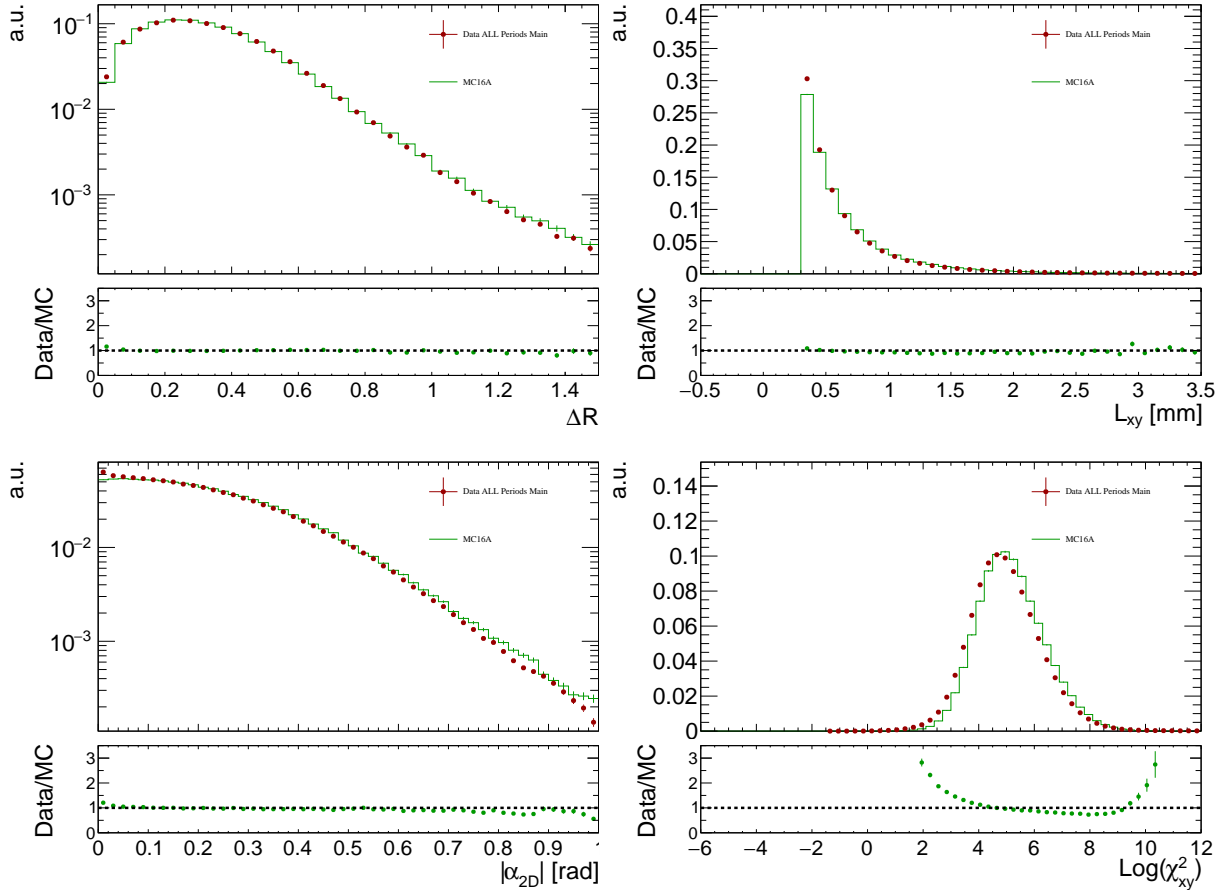
Figure 13: From left to right, from top to bottom: Data and signal background MC distributions of the ΔR, the transverse decay length (L_{xy}), $|\alpha_{2D}|$ and χ^2_{xy} that represents the separation between production (PV) and decay (SV) vertices. The green histogram corresponds to the sideband data, while the red points correspond to the re-weighted signal MC.

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

The MC sample for the reference channel $B^+ \rightarrow J/\psi K^+$ is used to compare the signal MC shapes with the collision data. The MC sample is re-weighted with the GLC and the DDW described in Secs. **??** and 5.2. For the data sample, the shape of the background distribution for each discriminating variable is estimated using the events falling into the left (5080MeV $< m_{J/\psi K^\pm} <$ 5180MeV) and the right (5380MeV $< m_{J/\psi K^\pm} <$ 5480MeV) sidebands.

19: t.b.d. - currently brute-force approach

The left sideband contains also a fraction of mis-reconstructed decays that behave signal-like from the point of view of the discriminating variables. We re-weight the left sideband considering the combinatorial contribution only: the net effect is that some of the signal will be also subtracted as mis-reconstructed B events. [9] It is proven that the mis-reconstructed events have the same shape as the signal in the

---

[9] 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.

557 discriminating variables considered, so no distortion of the signal shapes is created by this subtraction.
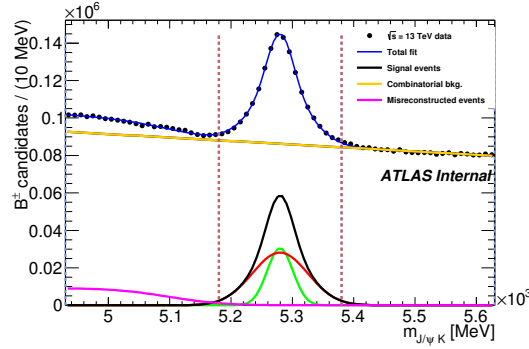

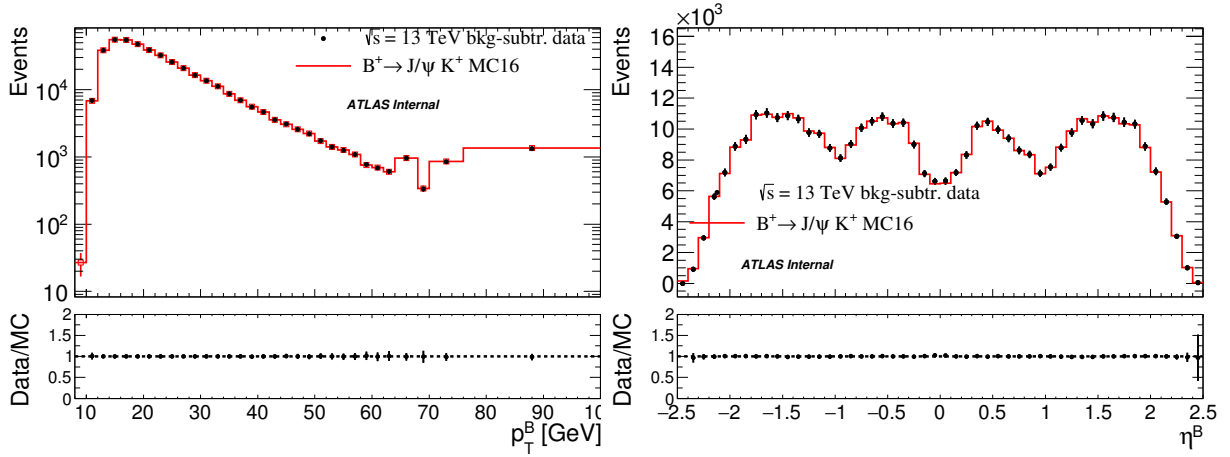Figure 14: Fit to the invariant mass distributions of $B^+ \to J/\psi K^+$ events.

Figure 15: 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.

558 The shape from the rescaled sideband events is subtracted statistically from the signal region. The number
559 of background events feeding into the signal region is obtained by a binned maximum likelihood fit to the
560 mass distribution. Figure 14 shows these binned fits to the mass distributions in the three trigger categories.
561 The fit model consists of a double Gaussian for the signal, an error function for the mis-reconstructed
562 events and an exponential for the continuum background. The results of these binned fit analysis are
563 perfectly compatible with the unbinned fit described in Section ?? and used for the yield extraction.

564 The distributions of p_T and η of the B is rechecked in figures 15, the distributions of few discriminating
565 variables are shown in figures 16, and the distribution of the continuum BDT variable is shown in figure 17.
566 The distributions of all the other variables are shown in the appendix ??.

567 Typically, except for deviations in individual bins, the overall shapes of distributions agree well between
568 data and MC. The observed differences in shape are accounted for as systematics with the procedure
569 described in Section ??. In addition, the discrepancy seen in the data-MC comparison of the continuum

20:
t.b.d.
21:
t.b.d.

BDT in the $B^+ \rightarrow J/\psi K^+$ events is assessed and exploited in the evaluation of the efficiency in the $B_s^0 \rightarrow \mu^+\mu^-$ signal in Section **??**.

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

The same procedure can be applied on data reconstructed as $B_s^0 \rightarrow J/\psi\phi$. In this case, a few extra selection cuts are necessary to address the extra kaon in the final state and to select the ϕ mass window: both kaon p_T are requested > 1 GeV and for the ϕ mass we require $|m_{hh} - m_\phi| < 15$ MeV. The additional cuts defined in Section **??** are also applied. The invariant mass fits in this case are shown in figure 18 for the three trigger categories. The fits are performed binned and a Gaussian PDF is used for the signal shape while a third order Chebychev is used for the combinatorial background PDF.

The data/MC comparisons of p_T and η distributions for the B_s^0 are shown in Figure 19, while the distributions of few discriminating variables are shown in Figure 20.. Finally the distribution of the continuum BDT variable is shown in Figure 21. The signal distributions in data are obtained by subtracting the sideband shapes in each variable.
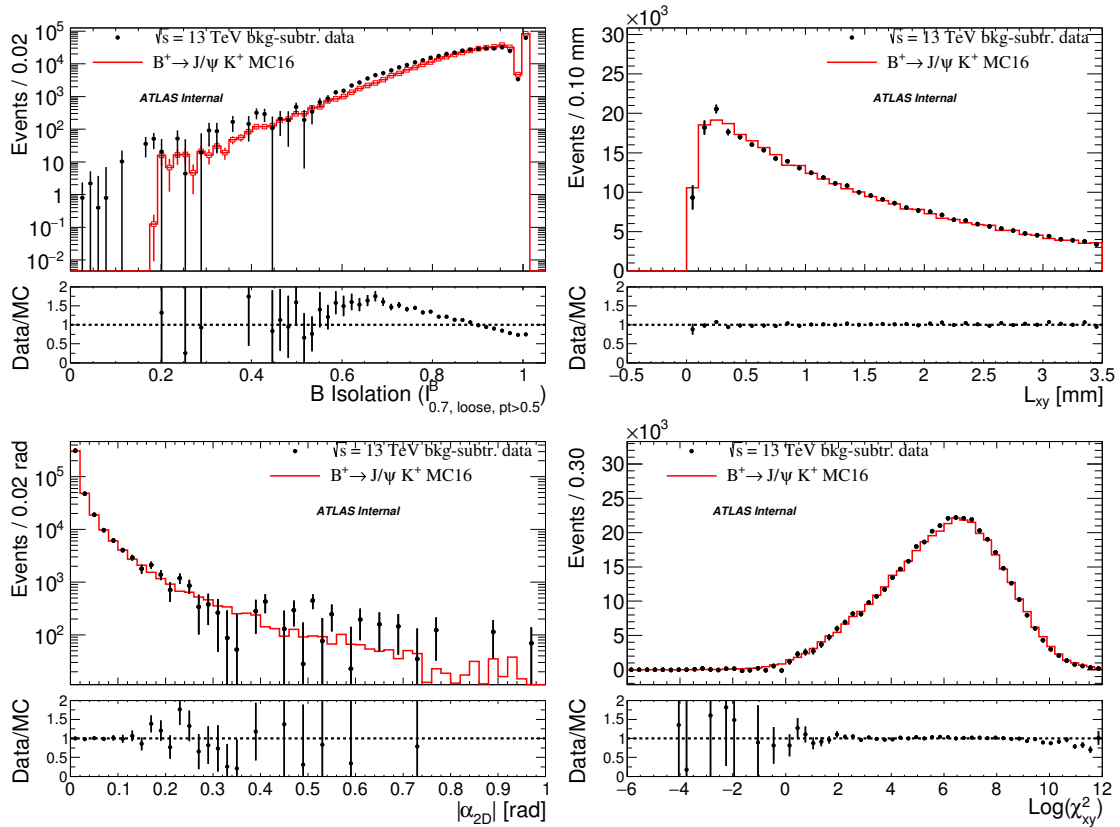
22:
t.b.d.

Figure 16: *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_{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.

583 Additional studies of the $B_s^0 \rightarrow J/\psi\phi$ are presented in Section **??**.

12.4 Yield stability during run

585 The stability of the gain in the B_s^0 sidebands in the 2015 and 2016 data taking periods is shown in Figure 22.
586 The gain is calculated as the yield of events per period divided by the luminosity in that period. The gain
587 appears to be stable during the run, within statistical uncertainties.

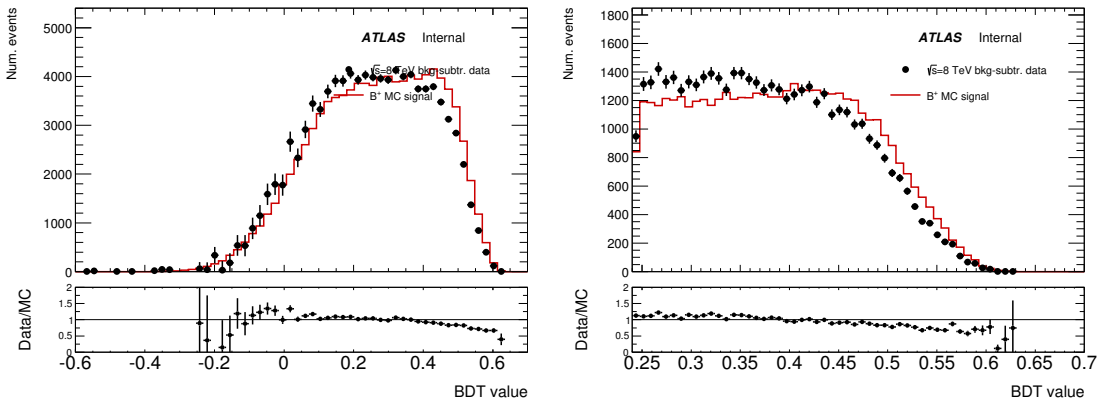Figure 17: **Run 1 result, t.b.u.!**Data and signal MC distribution in $B^+ \rightarrow J/\psi K^+$ events for the continuum BDT variable. The black dots correspond to the sideband subtracted data, while the red points correspond to reweighted MC normalised to the number of data events. The left plot shows the same distribution zoomed in the region of interest of the analysis.
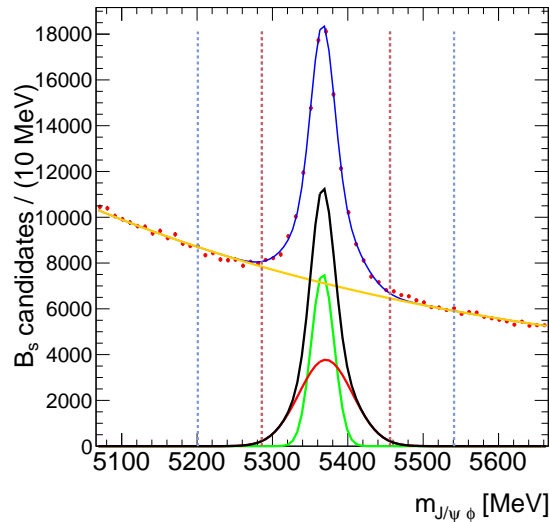

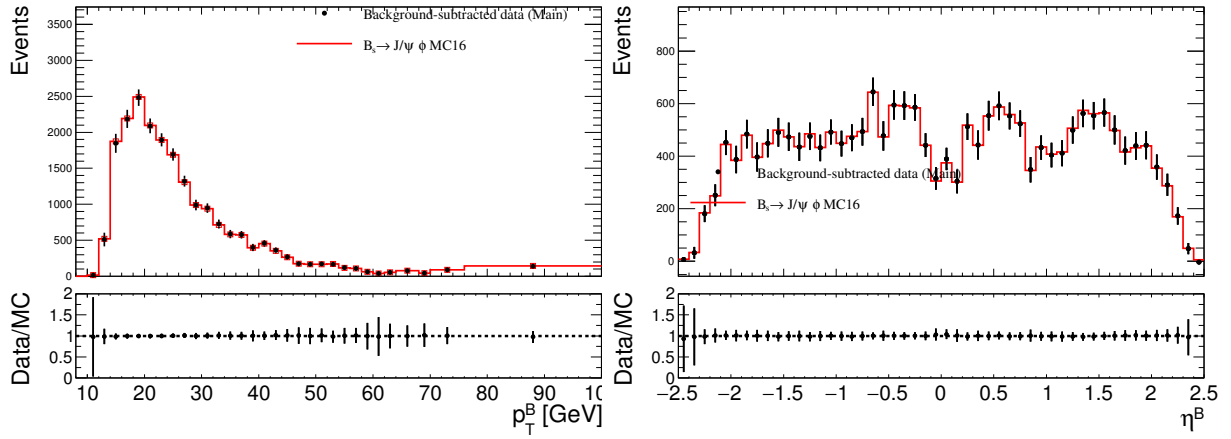Figure 18: Fit to the invariant mass distribution of $B_s^0 \rightarrow J/\psi\phi$ events.

Figure 19: 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.
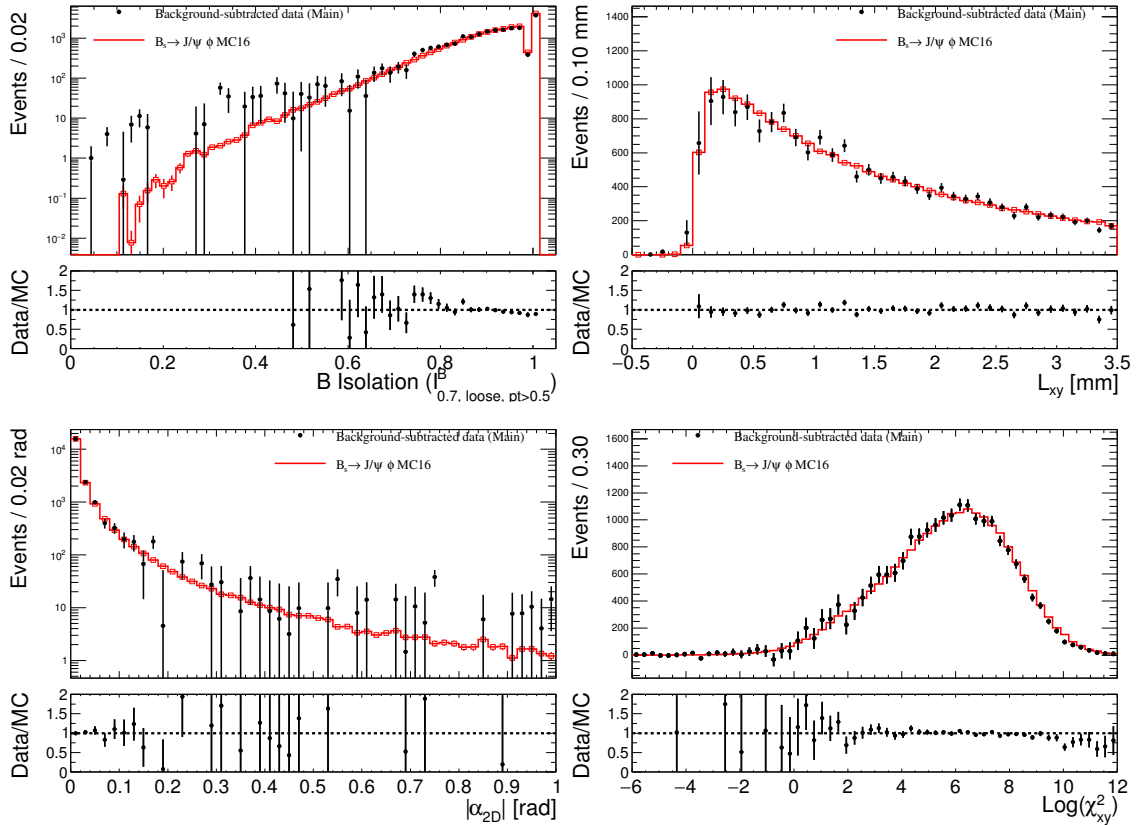
Figure 20: *From left to right, from top to bottom*: Data and signal MC distributions of $B_s^0 \to J/\psi\phi$ events for the B isolation variable, the transverse decay length (L_{xy}), $|\alpha_{2D}|$ and χ^2_{xy} that represents the separation between production (PV) and decay (SV) vertices. The black dots correspond to the sideband subtracted data, while the red points correspond to reweighted MC normalised to the number of data events.

Figure 21: **Run 1 result, t.b.u.!**Data and signal MC distribution in $B_s^0 \to J/\psi\phi$ events for the continuum BDT variable. The black dots correspond to the sideband subtracted data, while the red points correspond to rew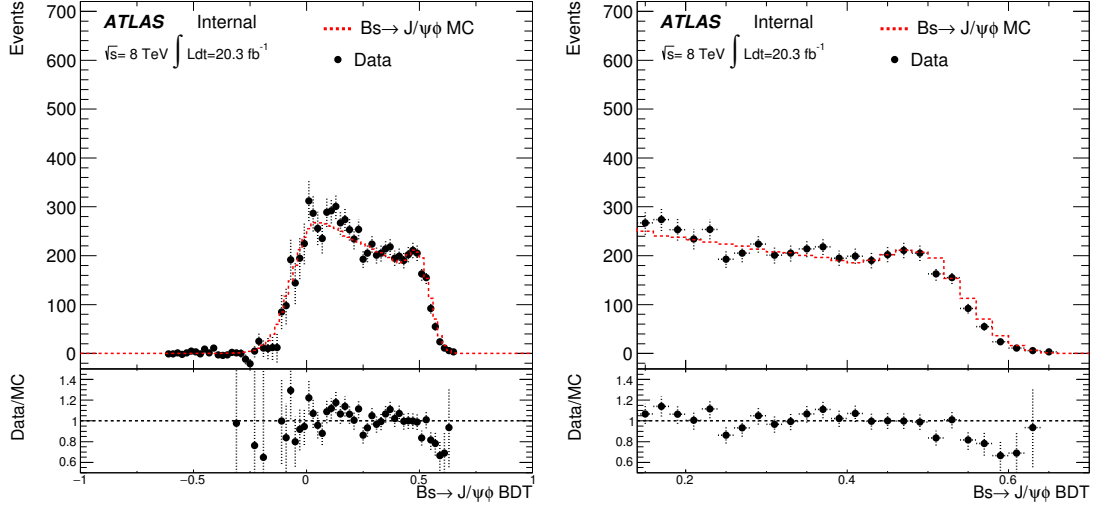eighted MC normalised to the number of data events. The left plot shows the same distribution zoomed in the region of interest of the analysis.
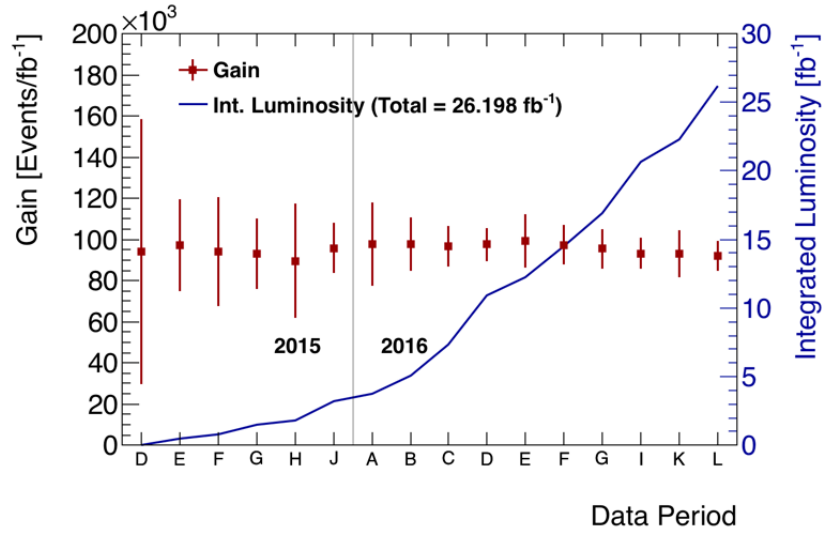

Figure 22: Stability of B_s^0 gain for 2015 and 2016 data taking periods. Also shown is the integrated luminosity for the analysis when the HLT_mu4_mu6_bBmumu and HLT_mu4_mu6_bBmumu_Lxy0 triggers are applied to the 2015 and 2016 data respectively.

13 BPlus Yield

13.1 Preselection

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

- **Trigger:** We require our events to pass `HLT_mu6_mu4_bJpsimumu`.

- **Requirements on B:** $M(B^+) \in [4930, 5630]$MeV and $\chi^2/NdoF < 6$ and $p_T(B^+) > 8$GeV.

- **Requirements on J/ψ:** $M(J/\psi) \in [2915, 3275]$MeV and $\chi^2/NdoF < 10$.

- **Kinematics on B:** $DR < 1.5$ and $|a_{2D}| < 1$ and $L_{xy} > 0$.

- **BDT:** $BDT > 0.24$ We used the same xml file as in Run I, for now.

This selection is applied to 2015 data and the BDT was trained in 2012 data with the background as the combinatorial contribution to the $B_s^0 \rightarrow \mu^+\mu^-$ signal (see figure 7(a)). However this BDT will be updated eventually to use 2015 data. Figure 23 shows the normalized distributions of the BDT score for both signal MC, data and the background.
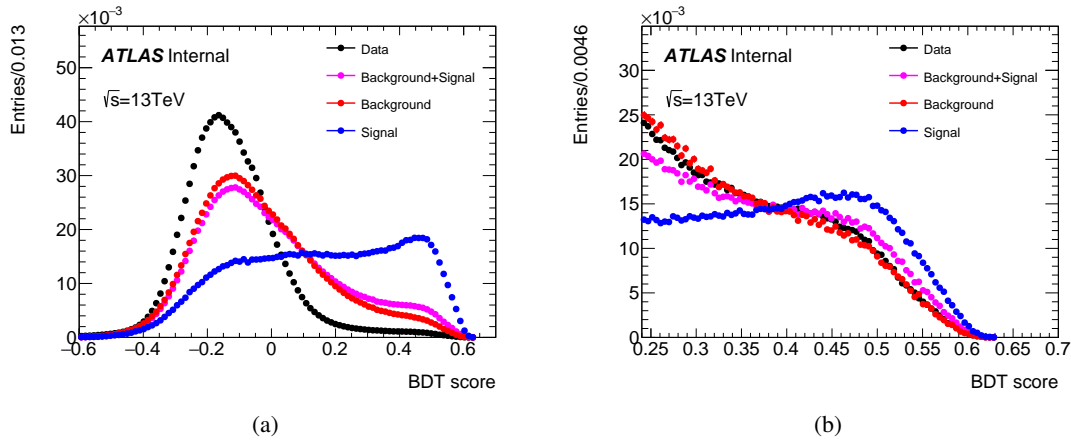
Figure 23: BDT score distribution for signal MC, background and data.

The background was extracted from the $b\bar{b} \rightarrow J/\psi + X$ sample, after the signal component was filtered out. The cutflow for the three samples can be seen in figure 24.

In these figures *No signal* implies removing the $B^+ \rightarrow J/\psi + K^+$ component from the respective sample, this obviously only makes sense for the $b\bar{b} \rightarrow J/\psi + X$ sample, otherwise it has no effect. The corresponding efficiencies are in tables 17, 18 and 19. After applying this selection the background composition is as shown in table 20.

Additionally figure 25 shows the distribution of the number of decays versus the index of the decay, when the most common decays are put at the beginning, the plot also shows the corresponding cumulative distribution.

(a) Background

(b) Signal

(c) Data
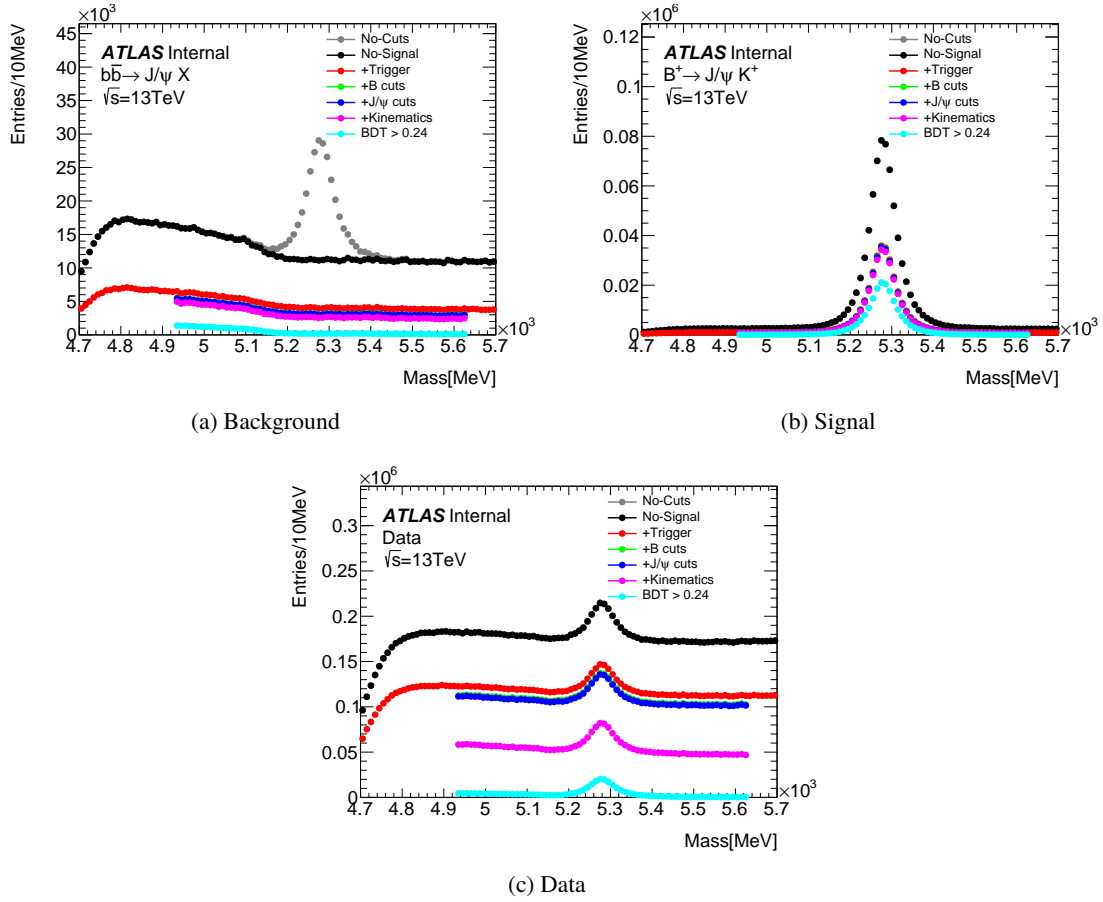
Figure 24: Cutflow associated to the simulated $B_s^0 \to \mu^+\mu^-$ and $B^+ \to J/\psi + K^+$ samples, and for the data.

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 17: Cutflow efficiency for $b\bar{b} \to J/\psi + X$ sample.

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

13.2 Procedure

To extract the B^+ yield for the reference channel, we use a unbinned maximum likelihood fit to the mass 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 18: Cutflow efficiency for $B^+ \to J/\psi + K^+$ 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 19: Cutflow efficiency for data.
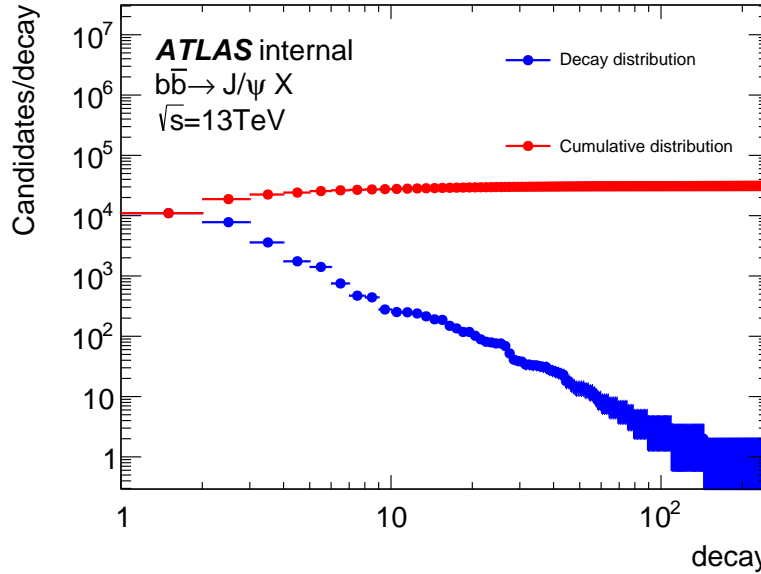
Figure 25: Decay distribution and the corresponding cumulative distribution.

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

In the $B^\pm$ invariant mass distribution (shown in Figure 26, left plot) 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. $B^{+/0} \to K^{*+/0} J/\psi$, $B^+ \to K^+ \chi_{c1,2}$) 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 Cabibbo-suppressed $B^\pm \to J/\psi \pi^\pm$ decay with the assignment of the kaon mass to 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: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: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: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 20: Background composition of $b\bar{b} \to J/\psi + X$ sample selection is applied.

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

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

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

Table **??** shows the $B^\pm \to J/\psi K^\pm$ candidates, in the mass region $4930 - 5630$ MeV, passing the complete selection (including the final continuum BDT cut).
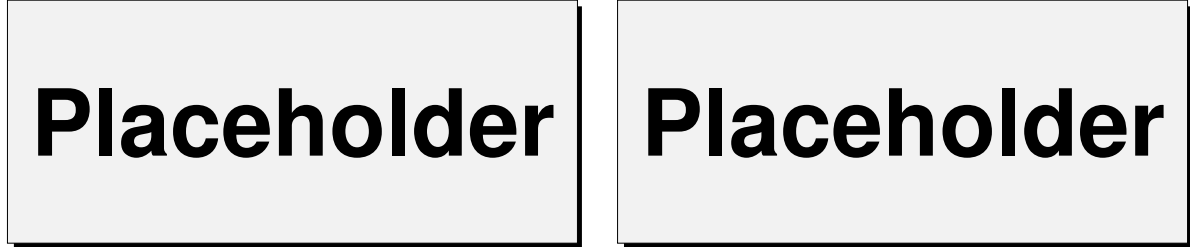
Figure 26: *Left*: $J/\psi K^{\pm}$ invariant mass distribution for all $B^{\pm}$ candidates in 2012 data. *Right*: partially reconstructed B decays contributing to the background as described by Monte Carlo.

The fit is based on a unbinned extended maximum likelihood fit to data with simultaneous inclusion of three MC samples to guide the modeling of several fit components. The MC samples are introduced to model accurately the shapes of the $B^{\pm} \rightarrow J/\psi K^{\pm}$ events as well as the most critical background components: the PRD and the $J/\psi\pi^{\pm}$ decays. By fitting the above mentioned MC samples simultaneously, we constrain the fit parameters of the corresponding fit components to their MC values. This results in an *MC assisted* determination of the background shapes, while automatically accounting for the statistical uncertainties of the MC. Two free parameters, one for the mass scale and the other for the mass resolution are extracted from data to accommodate for the possible data-MC difference in the shapes for $B^{\pm} \rightarrow J/\psi K^{\pm}$, PRD and $J/\psi\pi^{\pm}$ decays.

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

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:

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

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

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

659 The PDF's used are the following:

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

$$\text{Johnson } S_U = \frac{1}{\sqrt{1+t^2}} \left[t + \sqrt{1+t^2} \right]^{-\gamma - \frac{1}{2}\delta \ln\left(t+\sqrt{1+t^2}\right)} \text{ , where } t = \frac{m-\xi}{\lambda} \tag{4}$$

662   where $\xi = s_\mu + \xi^{MC}$ in which s_μ parameterizes the possible mass scale shift between data and MC.
663   The same s_μ scale shift is applied to the Gaussian mean: $s_\mu + \mu^{MC}$ where μ^{MC} is the mean of the
664   Gaussian shape in the $B^\pm \to J/\psi K^\pm$ MC sample. Parameters λ, ξ, δ and γ determine the shape of
665   the distribution Johnson S_U distribution. The parameters ξ and λ control the position and width of
666   the distribution, respectively. The sign of γ determines the position of the tail (left or right to the
667   peak).
668   A data-MC discrepancy is allowed also in the resolution parameters: the Johnson S_U distribution
669   is convoluted with a Gaussian resolution function with zero mean and s_σ width where the latter
670   parameter describes the resolution difference between data and MC. The same effect is taken into
671   account in the Gaussian distribution whose width is thus obtained by summing in quadrature the
672   MC width σ^{MC} and the resolution smearing factor s_σ.
673   The model of the shape is taken from $B^\pm \to J/\psi K^\pm$ MC events and its parameters are left floating
674   in the fit to be determined from the $B^\pm \to J/\psi K^\pm$ MC sample in the simultaneous fit. The
675   normalization and the 2 scaling factors allow the distribution to adjust to the data peak. Details on
676   the specific implementation of these parameterizations are found in Appendix **??**.

677   • radiative contribution to $B^\pm \to J/\psi K^\pm$ decays: when the B radiates a γ, the mass shape results
678   skewed on the left so we need to consider this component separately. A single Johnson S_U is used.
679   The shape parameters are determined from the radiative decays selected in the $J/\psi K^\pm$ MC sample.
680   The normalization and the 2 scaling factors allow the distribution to adjust to the data peak. Details
681   on the specific implementation of these parameterizations are found in Appendix **??**.
682   The relative abundances of the two signal components (non-radiative and radiative) are parametrized
683   using relative fractions so that the total number of $B^\pm \to J/\psi K^\pm$ events is extracted from the fit
684   together with the three fractions.

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

689   • PRD component:

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

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

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

– PRD2 with 5000-bump shape. We parameterize these events with a FD plus an Exponential Again the 2 overall scaling factors on the mass scale and resolution are applied to these parameterizations. The shape parameters are taken from the PRD2 events in the $(bb \to \mu^-\mu^+X)_{\text{pr}}$ MC sample in the simultaneous fit.

– PRD3 with a falling shape. We parameterize these events with an Exponential plus a constant term. Again the 2 overall scaling factors on the mass scale and resolution are applied to these parameterizations. The shape parameters are taken from the PRD3 events in the $(bb \to \mu^-\mu^+X)_{\text{pr}}$ MC sample in the simultaneous fit.

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

• combinatorial background: exponential function

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

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

Figures 27 and 28 show the results of the fits: the projection on the B^+ mass are shown for data. Appendix **??** contains the complete fit results including the projections on the mass for the MC samples. As an example, we show in Figure 29 all the simultaneous fit results. Table **??** shows the parameters from the fit, again as an example. In Appendix **??** similar tables are shown.

Figure 27: Fit projection on The red line represents the $B^\pm \to J/\psi K^\pm$ signal (including both radiative and non-radiative components), while the magenta line represents the $J/\psi\pi$ peaking component. The blue line shows all the three partially reconstructed contributions and the green line represents the combinatorial background. The total of all functions is presented with the black line.

Figure 28: Fit projection on data (left) and for the 2011 data (right). The red line represents the $B^\pm \to J/\psi K^\pm$ signal (including both radiative and non-radiative components), while the magenta line represents the $J/\psi\pi$ peaking component. The blue line shows all the three partially reconstructed contributions and the green line represents the combinatorial background. The total of all functions is presented with the black line.

Placeholder **Placeholder**

Placeholder **Placeholder**

Placeholder **Placeholder**
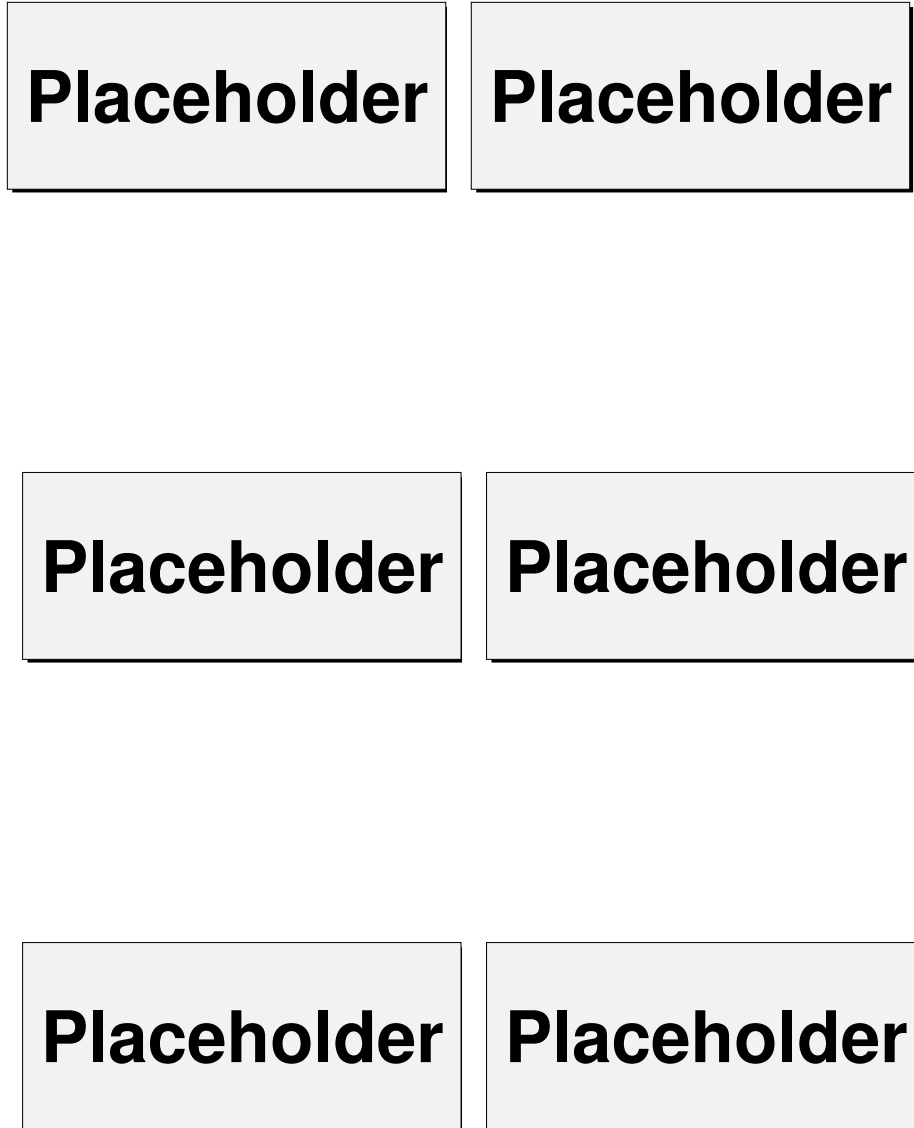
Figure 29: Fit projections on the MC samples simultaneous fitted with the data. From left to right, from top to bottom: non-radiative $B^{\pm} \rightarrow J/\psi K^{\pm}$ signal, radiative $B^{\pm} \rightarrow J/\psi K^{\pm}$ signal, peaking background $J/\psi\pi$, PRD1, PRD2 and PRD3. The red line is used for the $B^{\pm} \rightarrow J/\psi K^{\pm}$ signal (including both radiative and non-radiative components), while the magenta line is for the $J/\psi\pi$ peaking component. The blue lines refer to all the three partially reconstructed contributions. In each plot, the black non-continuous lines show the single functions of the total PDF used to model the given component.

.

13.3 Systematic Uncertainties in the Extraction of the $B^\pm$ Yields

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

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

The list of systematic checks is the following:

- data-MC discrepancies in B kinematics: discrepancies on the B meson kinematics between data and MC introduce a source of uncertainty. GLC- and DDW-weighted $J/\psi K^+$ and $J/\psi \pi^+$ MC samples account for biases introduced by generator level preselection cuts and discrepancies in 2D (p_T, η) spectrum of the $B^\pm$ with respect to data. These reweighted samples are used as alternative MC samples in the simultaneous fit.
- Alternative model for the PRD3 component: in the PRD3 mass distribution we observe a structure not modeled by the Exponential PDF used. We account for this structure by adding a Gaussian to the default model.
- Combinatorial background model: the combinatorial background smoothly crosses our fit window and it is modeled with an exponential function in the default fit. This shape is not MC driven. We evaluate the systematic uncertainty coming from the unknown combinatorial background model by changing the model to a linear function.
- $B^\pm \to J/\psi K^\pm$ signal peak charge asymmetry: the MC sample used in the default fit is from $B^+ \to J/\psi K^+$ events only. In order to account for shape variations relative to $B^- \to J/\psi K^-$, we repeat the fit using the latter.
- Alternative models for PRD1 and PRD2 components: systematics from the choice of PRD1 and PRD2 models are assessed by replacing the principal (default) FD function with complementary error function ($\mathrm{Erfc}(m) = \int_m^\infty e^{-t^2} \, dt$) in each of the two models.
- data-MC PRD composition discrepancies: the PRD MC sample has a relative decay modes abundances different than what would be expected according to current PDG measurements. Different abundances imply different mass shapes: we account for this data-MC shape discrepancy by introducing per-event weights derived from the Pythia and PDG differences. The weights correct for the MC shapes as well as for the relative abundances of the three PRD components. Two additional free parameters take care of the relative PRD fractions: these parameters have unit central values and are allowed to vary (via an external Gaussian constraint) within the PDG uncertainties on the BRs.
- including radiative tails into 2011 fit: When including radiative tails into 2011 fit we need to account for MC11-data11 discrepancies and 2011/2012 kinematic differences. To address the first issue, we use as alternate fitting sample 2011 MC weighted by the appropriate DDW obtained on 2011 data. To account for the MC11/MC12 kinematic differences we take the non-radiative B^+ candidates in the MC12 and the MC11 $B^\pm \to J/\psi K^\pm$ signal, reweight both with the appropriate DDW(&GLC)

and obtain 2011/2012 weights which we can use to reweight our 2012 MC radiative events to 2011 data conditions. These weights are also used in assessing the systematic effects from data-MC discrepancies in B kinematics (first bullet above) for the 2011 fit. In addition, we perform two systematic fits changing the rad/non-rad fractions by 50% and we take the largest variation on the result as systematic uncertainty.

The relative difference on the B^+ yields are given in table **??**.

The main systematic uncertainties on the B^+ yield come from the MC reweighting, the PRD reweighting and the combinatorial background parametrization. The overall systematic uncertainties result to be between $0.8 - 0.9\%$, depending on the data categories, as it is summarized in Table **??**.

13.4 $B^\pm \to J\psi K^\pm$ yield results

We measure the $B^\pm \to J\psi K^\pm$ reference channel yield with full systematic uncertainty evaluation in the four measurement categories as shown in Table **??**.

.

13.5 $J/\psi\pi^\pm/J/\psi K^\pm$ ratio measurement

From the fit described above, we extract both the yields for $B^\pm \to J/\psi K^\pm$ and $B^\pm \to J/\psi\pi^\pm$ and the systematic checks whose results are in Table **??** record the variation on both yields.

$$\mathrm{R}_{\pi/\mathrm{K}} = \frac{N_{J/\psi\pi^\pm} \times I_{ext}^{J/\psi\pi}}{N_{J/\psi K^\pm}} \tag{5}$$

where the $J/\psi\pi^\pm$ yield is corrected by the factor $I_{ext}^{J/\psi\pi}$ explained below. Then we define the ratio:

$$\rho_{\pi/\mathrm{K}} = \frac{\mathcal{BR}(B^\pm \to J/\psi\pi^\pm)}{\mathcal{BR}(B^\pm \to J/\psi K^\pm)} = \frac{N_{J/\psi\pi^\pm} \times I_{ext}^{J/\psi\pi}}{N_{J/\psi K^\pm}} \times \frac{\epsilon_{J/\psi K^\pm}}{\epsilon_{J/\psi\pi^\pm}} = \mathrm{R}_{\pi/\mathrm{K}} \times \left[\frac{\epsilon_{K^+}}{\epsilon_{\pi^+}} \times \frac{1 + \frac{\epsilon_{K^-}}{\epsilon_{K^+}}}{1 + \frac{\epsilon_{\pi^-}}{\epsilon_{\pi^+}}} \right] \tag{6}$$

where N_X is the yield for channel X ($J/\psi\pi^\pm$, $J/\psi K^\pm$), and ϵ_X is the efficiency-times-acceptance product for channel X. In the last equality we have used the asymptotic relation $\epsilon_h^\pm = \frac{\epsilon_h^+ + \epsilon_h^-}{2}$, and $\frac{\epsilon_{K^-}}{\epsilon_{K^+}}$ and $\frac{\epsilon_{\pi^-}}{\epsilon_{\pi^+}}$ are the kaon and pion charge asymmetries, respectively.

The fit for the reference channel of Section **??** extracts both yields $J/\psi K^\pm$ and $J/\psi\pi^\pm$ from the fit mass window [4930., 5630.] MeV/c^2. While the $J/\psi K^\pm$ component is accommodated entirely within our mass fit window, the $J/\psi\pi^\pm$ component extends by a small fraction outside the right boundary. A fraction $I_{ext.}^{J/\psi\pi}$ of $J/\psi\pi^\pm$ candidates counted in the extended region [3500., 7000.] MeV/c^2 with respect to the candidates counted in the default fit mass window is introduced as a correction for this small $J/\psi\pi^\pm$ tail. The result of such integration $I_{ext.}^{J/\psi\pi}$ together with the corrected yield are summarized in Table **??**.

783 Regarding the $\frac{\epsilon_{K^\pm}}{\epsilon_{\pi^\pm}}$ ratio, three are the factors contributing: the kaon and pion charge asymmetries ($\frac{\epsilon_{K^-}}{\epsilon_{K^+}}$,

784 $\frac{\epsilon_{\pi^-}}{\epsilon_{\pi^+}}$) and the relative K^+/π^+ efficiency $\frac{\epsilon_{K^+}}{\epsilon_{\pi^+}}$. For the pion charge asymmetry we assume the central value

785 to be $\frac{\epsilon_{\pi^-}}{\epsilon_{\pi^+}} = 1$ and we assign an uncertainty to it as discussed below. We evaluate the remaining two factors

786 using Monte Carlo and estimate systematic uncertainties as described below.

787 The four measurements of $\rho_{\pi/K}$ in the separate data categories are then combined to minimize the

788 measurement uncertainty.

789 Most systematic effects (luminosity, trigger, reconstruction efficiencies) cancel in the measurement of this

790 ratio due to the almost identical topology and kinematics of the two decay channels. Residual systematic

791 uncertainties on the ratio of branching fractions come from uncertainties on the parametrization of the

792 fit PDFs, data-MC discrepancies, the K^-/K^+ and π^-/π^+ charge asymmetries, and the K^+/π^+ relative

793 efficiency. Introducing scaling factors $I_{ext.}^{J/\psi\pi}$ to account for the $(1-4\%)$ of $J/\psi\pi^\pm$ events falling outside

794 of the fit window does not introduce significant source of systematic uncertainties.

795 Table **??** contains the systematic contributions to the final averaged $\overline{\rho_{\pi/K}}$ ratio from the study performed

796 for the reference channel yield extraction described in Sec. 13.3. In addition, a number of ratio-specific

797 systematic checks need to be performed on the kaon and pion charge asymmetries and the relative K^+/π^+

798 efficiency:

799 **K^-/K^+ charge asymmetry** (0.09%) The K^-/K^+ charge asymmetry is measured on $B^\pm \to J/\psi K^\pm$ MC.

800 The uncertainty on the charge asymmetry is driven by the data-MC discrepancies on the kaon

801 kinematics which are connected in the $B^\pm \to J/\psi K^\pm$ decay by the data-MC discrepancies on the B

802 meson kinematics. By reweighting our MC sample using GLC and DDW weights (see Section **??**)

803 we re-evaluate the charge asymmetry and take the resulting variation on $\epsilon_{K^-}/\epsilon_{K^+}$ as systematic

804 uncertainty.

805 **π^-/π^+ charge asymmetry** (1.14%) The π^-/π^+ charge asymmetry is assumed to be $= 1$, compatible

806 within statistical uncertainty with what predicted from MC. We estimate the systematic uncertainty

807 on $\epsilon_{\pi^-}/\epsilon_{\pi^+}$ comparing the central value of the $\epsilon_{K^-}/\epsilon_{K^+}$ (described above) with $\frac{\epsilon_{K^-}/\epsilon_{K^+}}{\epsilon_{\pi^-}/\epsilon_{\pi^+}}$ obtained

808 from the $B \to h^- h^+$ MC12 sample (the highest statistics π/K simulation we have with hadron

809 spectra similar to $J/\psi\pi^\pm/J/\psi K^\pm$). Hadron selection on $B \to h^- h^+$ events are kept as close as

810 possible to those on the B^+ signal selection. We then assume $\frac{\epsilon_{B_d \to K^- \pi^+}}{\epsilon_{B_d \to K^+ \pi^-}} \approx \frac{\epsilon_{K^-}/\epsilon_{K^+}}{\epsilon_{\pi^-}/\epsilon_{\pi^+}}$. The π^-/π^+

811 charge asymmetry is then estimated as :

$$\epsilon_{\pi^-}/\epsilon_{\pi^+} = \frac{\epsilon_{B_d \to K^+ \pi^-}}{\epsilon_{B_d \to K^- \pi^+}} \times \frac{\epsilon_{K^-}}{\epsilon_{K^+}} = 1.011 \pm 0.004 \qquad (7)$$

812 The full difference from 1 is taken as our systematic uncertainty on $\epsilon_{\pi^-}/\epsilon_{\pi^+}$.

813 **K^+/π^+ relative efficiency** (0.60%) $\epsilon_{K^+}/\epsilon_{\pi^+}$ is measured on GLC- and DDW-weighted $B^\pm \to J/\psi K^\pm$

814 MC sample using the same machinery as $\frac{\epsilon_{B^+ \to J/\psi K^+}}{\epsilon_{B_s \to \mu^+ \mu^-}}$ in the main analysis. Discrepancies on this

815 parameter arise predominantly from residual data-MC discrepancies in the B spectrum model.

816 The final efficiency ratios entering in all categories can be found in Table **??**. For comparison and

817 crosscheck, we report also the default fit results for $\frac{N_{J/\psi\pi^-}}{N_{J/\psi\pi^+}}$ and $\frac{N_{J/\psi K^-}}{N_{J/\psi K^+}}$.

Table ?? reports the yield ratio $R_{\pi/K}$ with its statistical uncertainty. After the efficiency correction, the measurement of BR ratio $\rho_{\pi/K}$ with correctly propagated statistical uncertainties can be found in last two columns of the same Table ??. To combine the measurements, we use the squared inverse value of statistical uncertainty on each measurement as a weight and calculate the weighted mean $\overline{\rho_{\pi/K}}$. To evaluate the systematic uncertainty on the result $\overline{\rho_{\pi/K}}$ we re-evaluate the combination for each systematic variation, therefore accounting for correlated effects. The difference with the default value $\overline{\rho_{\pi/K}}$ is taken as the combined systematic uncertainty for each effect. Systematic uncertainties obtained this way are summarized in Table ?? and they are summed in quadrature to obtain the combined uncertainty.

The largest systematic uncertainty on the measured ratio comes from the combinatorial background model parametrization ($\approx 32\%$), followed by the effect of PRD reweighting ($\approx 10\%$), by the $B^+ \rightarrow J/\psi K^+$ signal peak shape charge asymmetry ($\approx 8\%$), by the effect of the radiative tails ($\approx 5\%$) in the signal models and by the effect of the weights (GLC and DDW) applied to signal MC ($\approx 5\%$). All other systematic sources have minor effects ($\approx 3\%$ or less).

The final result on the ratio of branching fractions $\frac{\mathcal{BR}(B^\pm \rightarrow J/\psi \pi^\pm)}{\mathcal{BR}(B^\pm \rightarrow J/\psi K^\pm)}$ is:

$$\overline{\rho_{\pi/K}} = (3.5 \pm 0.3^{stat.} \pm 1.2^{syst.})\% \tag{8}$$

14 EfficiencyAndAcceptance

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

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

$$R_{A\epsilon} = \frac{A_{B^+ \to J/\psi K^+} \times \epsilon_{B^+ \to J/\psi K^+}}{A_{B_s^0 \to \mu^+\mu^-} \times \epsilon_{B_s^0 \to \mu^+\mu^-}} \tag{9}$$

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

- A takes into account the loss in acceptance, with respect to the chosen phase-space fiducial volume, due to the MC generator cuts applied on the final state particles (the two muons for both channels and the kaon for the reference channel only) in order to produce the MC samples used in the analysis. The fiducial volume is defined as $p_T^B > 8.0$ GeV and $|\eta_B| < 2.5$, while the final state particle cuts are: $p_T^\mu > 4.0$ GeV and $\left|\eta_\mu\right| < 2.5$ for muons and $p_T^K > 1.0$ GeV and $|\eta_K| < 2.5$ for the kaon. The A term is defined as the ratio between the number of events passing the final state particle cuts and the number of events in the fiducial volume. This term has been evaluated using a specific "un-biased" MC sample where only the phase-space fiducial volume cuts on the B meson were applied and it is different between the reference and the signal channels. Data driven corrections to the p_T^B, $|\eta_B|$ distribution are applied to the MC sample, both in the numeration and the denominator, in order to reproduce the spectrum observed in data. in The kinematic cuts are applied to the various particles only at the truth level. The values of the A term are reported in Table 21.**They refer to Run1 analysis and are here just to give an idea of the order of magnitude of values and uncertainties**.

- ϵ takes into account all the reconstruction effects (selection cuts, trigger efficiency, reconstruction efficiency) affecting the two channels. It is defined as the ratio between the number of events passing the final selection cuts listed in Section 6 including the cut on the continuum BDT once fixed(listed in Section 11) and the number of events passing the final state particle kinematic cuts applied to the appropriate truth level quantities. Both terms are evaluated using the simulated MC samples available. Since this term involves reconstructed quantities, QLC (see Section 5.1) and data driven corrections (see Section 5.2) have been applied to both the numerator (to reconstructed quantities) and the denominator (to truth quantities) in order to properly take into account the effects of a modified p_T and $|\eta|$ spectra of the B-meson. Trigger efficiency corrections, that will be taken from data and will appear as Scale Factors to be applied to MC, will be added as well once available (see Section 7). They will be applied to the numerator only.

The values for the $A \times \epsilon$ and $R_{A\epsilon}$ terms are summarised in Table 22, together with the statistical and systematic uncertainties described in the next section. The statistical uncertainties come from the finite statistic available for the simulated samples used in the analysis. **They refer to Run1 analysis and are here just to give an idea of the order of magnitude of values and uncertainties**. The sources of systematic uncertainties affecting the values reported in Table 22 are described in Section 14.1.

$A^{2012}_{B^+ \to J/\psi K^+}$	0.0865 ± 0.0002
$A^{2012}_{B^0_s \to \mu^+\mu^-}$	0.2902 ± 0.0005
$A^{2011}_{B^+ \to J/\psi K^+}$	0.0819 ± 0.0003
$A^{2011}_{B^0_s \to \mu^+\mu^-}$	0.3035 ± 0.0005

Table 21: A-terms for B^+ and B^0_s 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	$0.0080\pm0.28\% \pm14.2\%$	$0.180\pm0.56\%$ (stat)$\pm5.2\%$ (syst)
B^0_s	0.1522	$0.0441\pm0.49\% \pm10.3\%$	

Table 22: ϵ, $A \times \epsilon$ and $R_{A\epsilon}$ values for B^+ and B^0_s channels for 2012 (split into the three trigger categories) and 2011 samples. Where present, the first uncertainty is statistical and second one is systematic.

14.1 Systematic uncertainties on $R_{A\epsilon}$

The systematic uncertainty affecting $R_{A\epsilon}$ comes mainly from two sources: the uncertainty related to the corrections (GLC, data driven and trigger efficiencies) and the residual discrepancy between data and MC samples after the application of these corrections. The total systematic uncertainty quoted in Table 22 is the sum in quadrature of these two components.

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

Another source of systematic uncertainty arises from the discrepancies between data and MC. The systematic uncertainty on $R_{A\epsilon}$ was assessed by observing the variation in the efficiency of the final selection when re-weighting both the MC samples to the observed data for each of the 15 variables used in the continuum BDT. Data are extracted from B^+ events after the subtraction of the background as shown in section 12. $R_{A\epsilon}$ is recomputed after the reweighting of each variable one at a time. The discrepancy between the values obtained with this procedure and the central values has been considered as the systematic uncertainty due to the specific variable mis-modelling.

The total systematic uncertainty on $A\times\epsilon$ and $R_{A\epsilon}$ due to the reweighting procedure is the sum in quadrature of the single systematic uncertainties and it is quoted in Table 22.

15 SignalFit

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

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

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

- The models used for the signal and the backgrounds are discussed in Sections 15.1 and 15.2.
- The results of the fit to the data in the sidebands and the interpolation in the signal region are presented in Section 15.3.
- A summary of the baseline fit configuration is given in Section 15.4.
- Systematic uncertainties on the fit are discussed in Section 15.5.

15.1 Signal and peaking background

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

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

15.2 Parametrisation of background components

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

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

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

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

15.3 Fit to background components from MC and to sideband data

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

15.4 Summary of the fit configuration

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

1. signal PDF: the mass dependence is described by the sum of 2 Gaussians centred at the B_s (or B_d) mass. The widths of the Gaussians and their relative fraction, assumed to be identical in all continuum-BDT bins, will be taken from MC and fixed in the fit.

2. Continuum background PDF: the mass dependence is first order polynomial. The normalisation and the slope will be extracted independently in each bin of the continuum-BDT. In full Run1 analysis Gaussian constraints were placed on the uniformity of the slope, so that the slope in bin-2 (bin-3) was equal to the one in bin-1 within ±40% (±80%).

3. Low-mass background PDF: exponential dependence on the mass. The normalisation will be extracted independently in each bin of the continuum-BDT, while the shape will be assumed to be uniform.

4. Peaking background: the mass dependence is described with a Gaussian describing the total background.

15.5 Systematic uncertainties on the fit in the simultaneous fit to B_s and B_d

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

16 BranchingRatio

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

- The external inputs needed are the branching fraction for the reference channel that is obtained 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 $\mathcal{B}(J/\psi \rightarrow \mu^+\mu^-) = (5.961 \pm 0.033)\%$. The relative hadronisation probability f_u/f_s is taken from the best experimental result [**lhcbfsfu14**] available: $f_s/f_d = 0.259 \pm 0.015$ using $f_d/f_u = 1$. The product of these external inputs gives:

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

  which corresponds to a relative uncertainty of 6.6%
- Number of signal events, $N_{\mu^+\mu^-}$: at this blinded stage we consider the expected number of events as ... derived from the SM branching ratio and the error from the default fit as ..... The relative statistical uncertainty is about .....
- Efficiency weighted number of events for the reference channel (D_{norm}).

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

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

where the relative uncertainty is about ......

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

25: possibly to be updated

26: possibly to be updated

References

[1] C. Bobeth, M. Gorbahn, T. Hermann, M. Misiak, E. Stamou et al.,
$B_{s,d} \to l^+ l^-$ in the Standard Model with Reduced Theoretical Uncertainty,
Phys.Rev.Lett. **112** (2014) 101801, arXiv: 1311.0903 [hep-ph].

[2] R. Aaij et al., *Measurement of the $B_s^0 \to \mu^+ \mu^-$ branching fraction and effective lifetime and search for $B^0 \to \mu^+ \mu^-$ decays*, Phys. Rev. Lett. **118** (2017) 191801, arXiv: 1703.05747 [hep-ex].

[3] V. Khachatryan et al.,
Observation of the rare $B_s^0 \to \mu^+ \mu^-$ decay from the combined analysis of CMS and LHCb data,
Nature **522** (2015) 68, arXiv: 1411.4413 [hep-ex].

[4] C. Alpigiani et al., *$B \to \mu\mu$ analysis on the complete RUN1 dataset*,
tech. rep. ATL-COM-PHYS-2014-1179, CERN, 2014,
URL: https://cds.cern.ch/record/1756291.

[5] M. Aaboud et al., *Study of the rare decays of B_s^0 and B^0 into muon pairs from data collected during the LHC Run 1 with the ATLAS detector*, Eur. Phys. J. **C76** (2016) 513,
arXiv: 1604.04263 [hep-ex].

[6] *Limit on $B_s \to \mu\mu$ branching ratio based on 4.9 fb^{-1} of integrated luminosity*,
tech. rep. ATL-COM-PHYS-2013-053, CERN, 2013.

[7] G. Aad et al., *Search for the decay $B_s^0 \to \mu\mu$ with the ATLAS detector*,
Phys.Lett. **B713** (2012) 387, arXiv: 1204.0735 [hep-ex].

[8] M. Smizanska, *PythiaB: an interface to Pythia 6 dedicated to simulation of beauty events*,
ATL-COM-PHYS-2003-038, 2003.

[9] A. Hoecker et al., *TMVA: Toolkit for Multivariate Data Analysis*, PoS **ACAT** (2007) 040,
arXiv: physics/0703039.

[10] A. Rogozhnikov, *Reweighting with Boosted Decision Trees*,
J. Phys. Conf. Ser. **762** (2016) 012036, arXiv: 1608.05806 [physics.data-an].

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

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
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
S.Yu. Sivoklokov	$J/\Psi\phi$ control sample studies

Appendices

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