ATLAS Note

ANA-BPHY-2018-09-INT1

13th April 2018

Draft version 0.1

Study of the rare decays of B0s and B0 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 $B_{s,d} \rightarrow \mu\mu$ 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

1 Note Changes

- **v**0.1
    - 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 17.

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 14, while the relative efficiencies and acceptances of $B^0_{(s)} \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$ will be extracted in 15. The $B^0_{(s)} \to \mu^+\mu^-$ yield will be derived from invariant mass distribution fits in section 16, implementing the selections optimized in sections 10, 12 and 9.

Given these ingredients, the branching ratio to the left of the reference formula above will be derived in section 17, where relevant systematic effects will be propagated as well. Dedicated sections will discuss the tuning of Montecarlo samples 5, the data-driven 13 extrapolation of MC models to the signal region 11.

99 # 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 the analysis optimisation ("blinding").[2]

The derived AOD (DAOD) samples used in this analysis are produced by the BPHY8 derivation format:[3]

- 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.

---

[2] This sentence will need to be adjusted once we have the reprocessed DAODs.

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

Channel	Type	Events	Generator	Simulation
$B_s^0 \to \mu^+\mu^-$	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.).

Further details on the datasets are given in Table 2 (for AODs) and Table 3 (for DAODs)[4].

[4] Please note that the DAOD datasets will need adjustment once we will have produced the final round of MC DAODs with the additional variables.

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
BdPimunu	30,932	mc16_13TeV.300433.Pythia8BEvtGen_A14_CTEQ6L1_Bd_pi3p5mu3p5nu.deriv.DAOD_BPHY8.e4720_e5984_a875_r9364_r9315_p3371
LbPmunu	15,768	mc16_13TeV.300434.Pythia8BEvtGen_A14_CTEQ6L1_Lambda0b_p3p5mu3p5nu.deriv.DAOD_BPHY8.e4911_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:

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

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

162

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

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

5.2 Data Driven Weights

169 The Data Driven Weights (DDW) have been evaluated in a similar way as it has been done in [6] and
170 [4].[5]

171 The DDW for the MC signal samples are determined with an iterative method, by comparing sideband-
172 subtracted $B^\pm \to J/\psi K^\pm$ odd-numbered events with MC events, after the latter are corrected with the
173 W_{QLC} weights.[6]

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

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

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

3: calculaton of DDW under-way

---

[5] 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 ....

2: still work in progress

[6] 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,
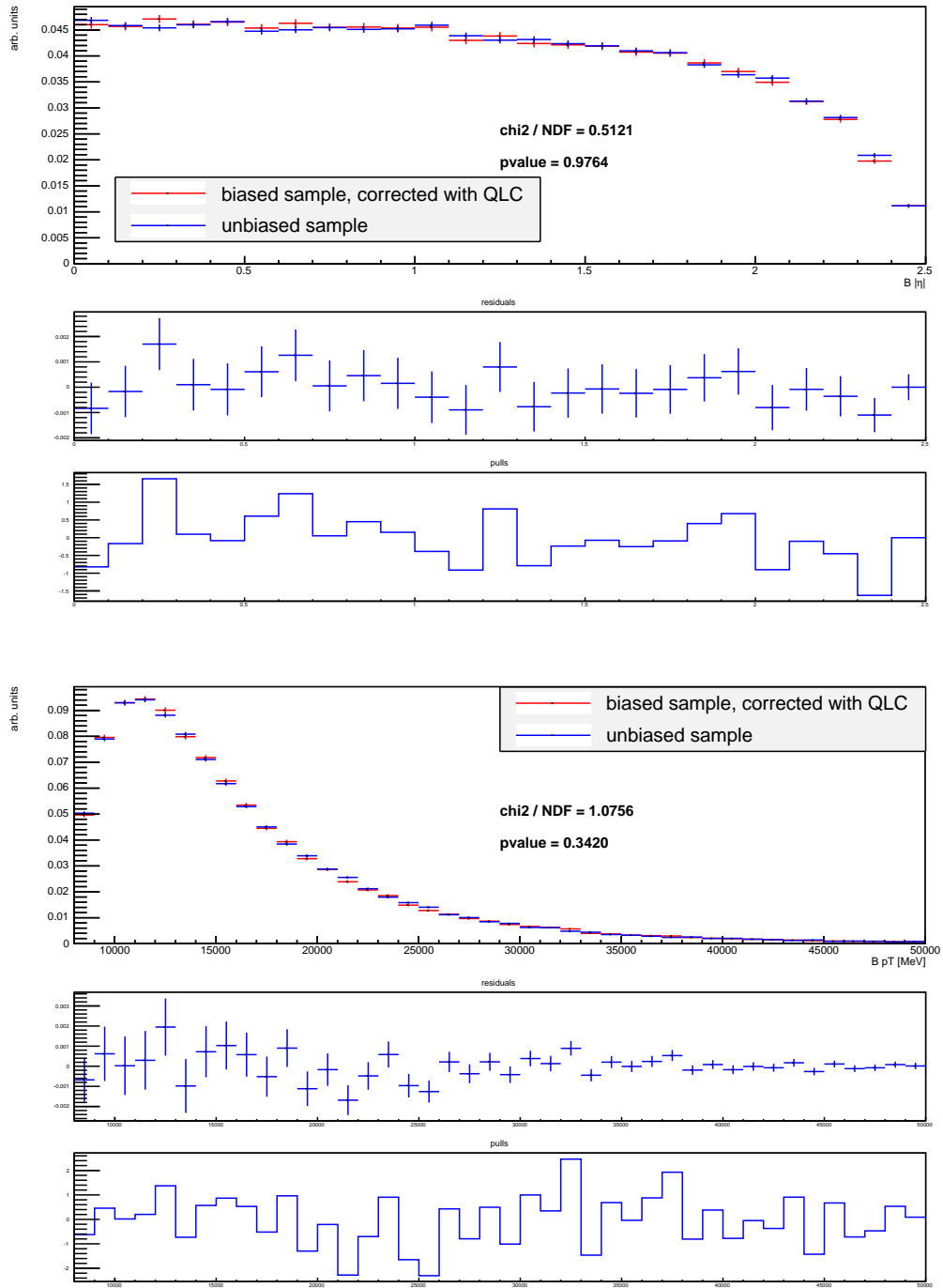and the uncertainty in the signal yield is dominated statistical errors.

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 Preselection

7 Candidate preselection

7.1 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 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).

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
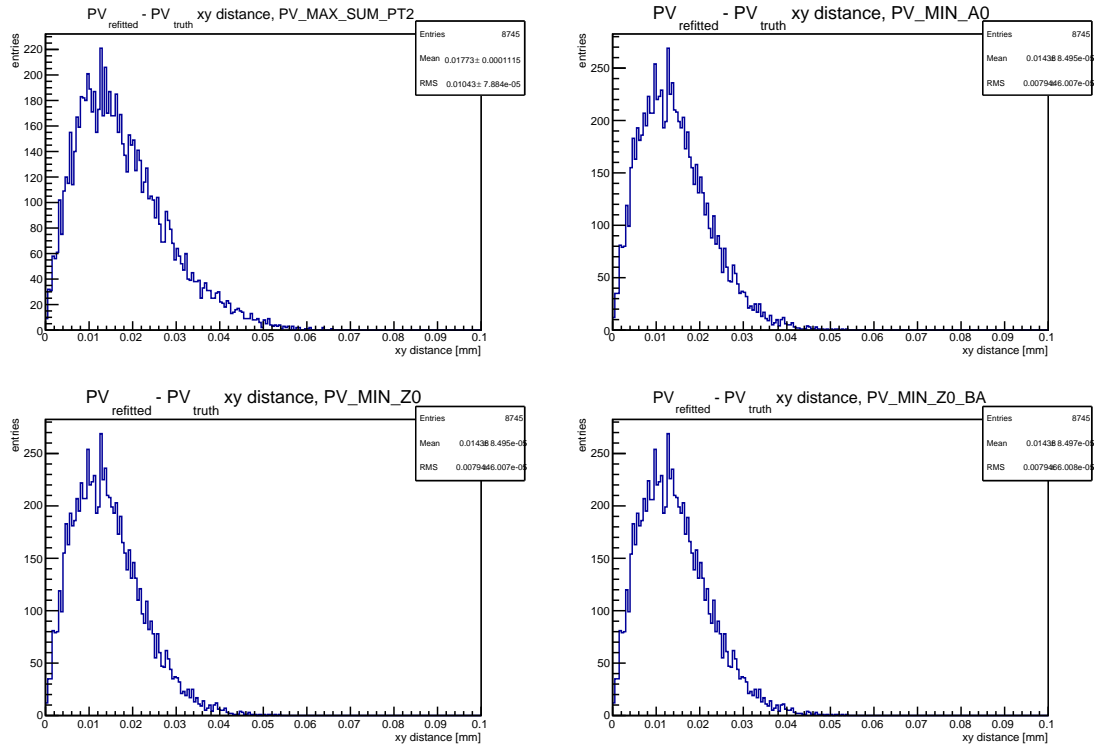

Figure 2: xy distance between selected PV and truth PV for the four approaches.

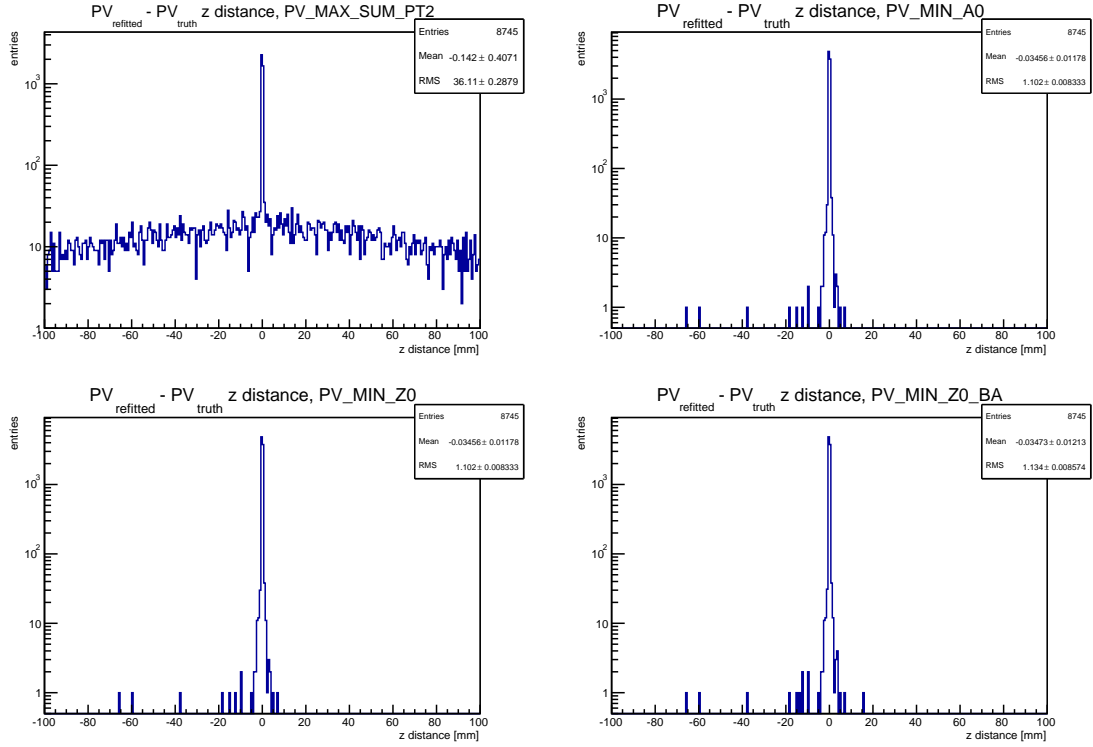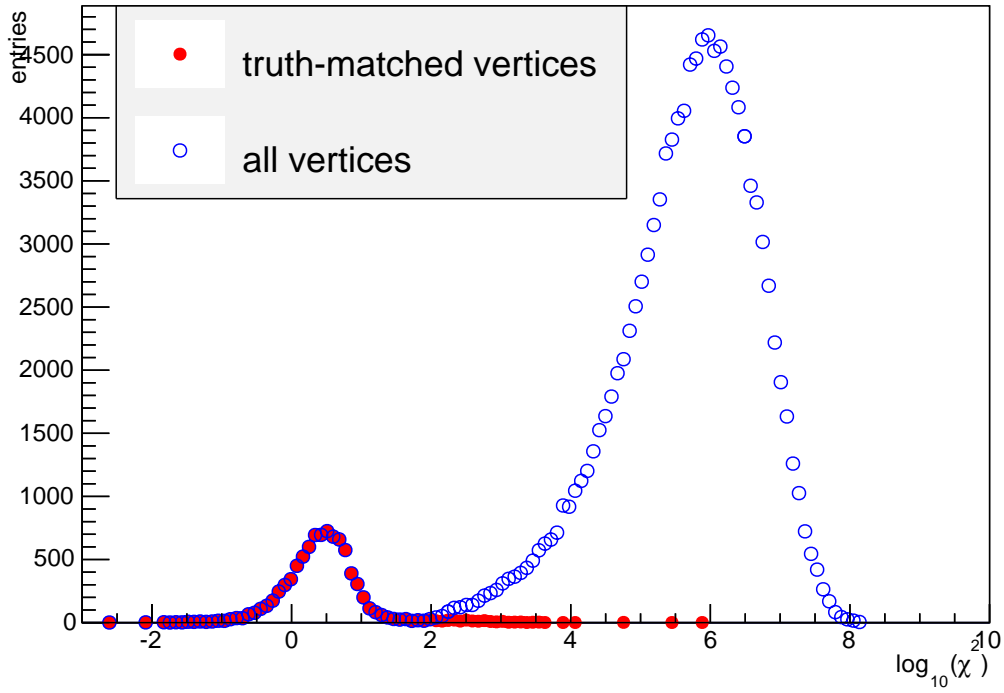Figure 3: z distance between selected PV and truth PV for the four approaches.

Figure 4: χ^2 of all the PVs (blue points) and the distribution of the χ^2 of the "truth-matched" PVs.
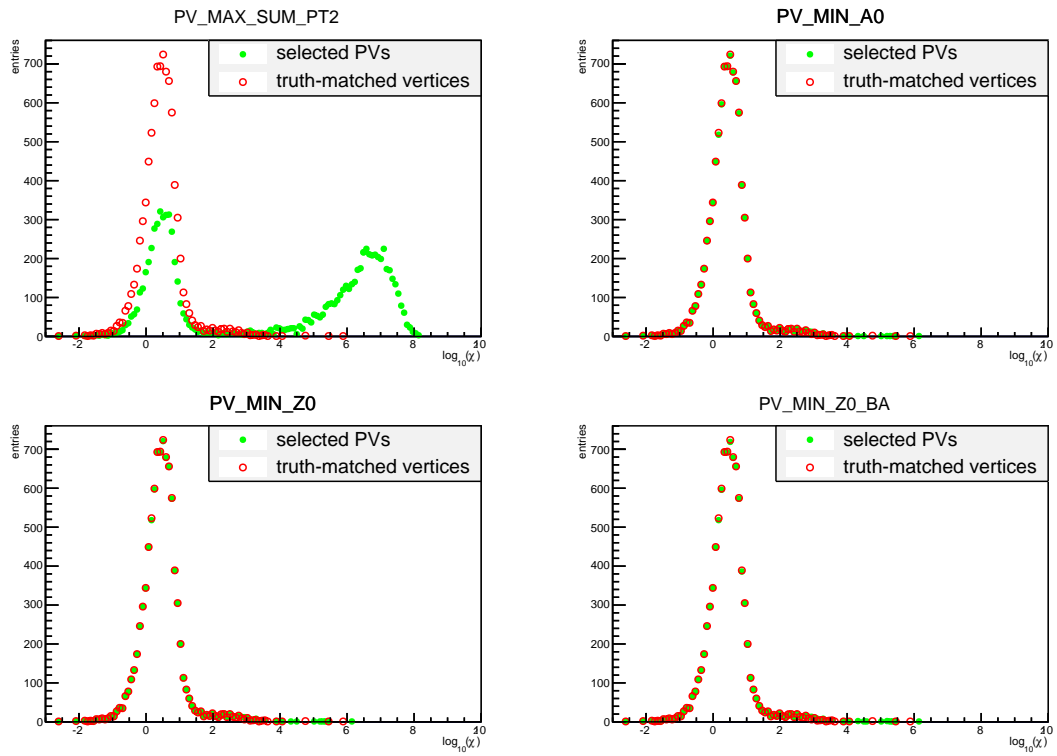
Figure 5: χ^2 distribution of the chosen PVs using the 4 approaches (green distributions) superimposed to the truth-matched distribution (red).
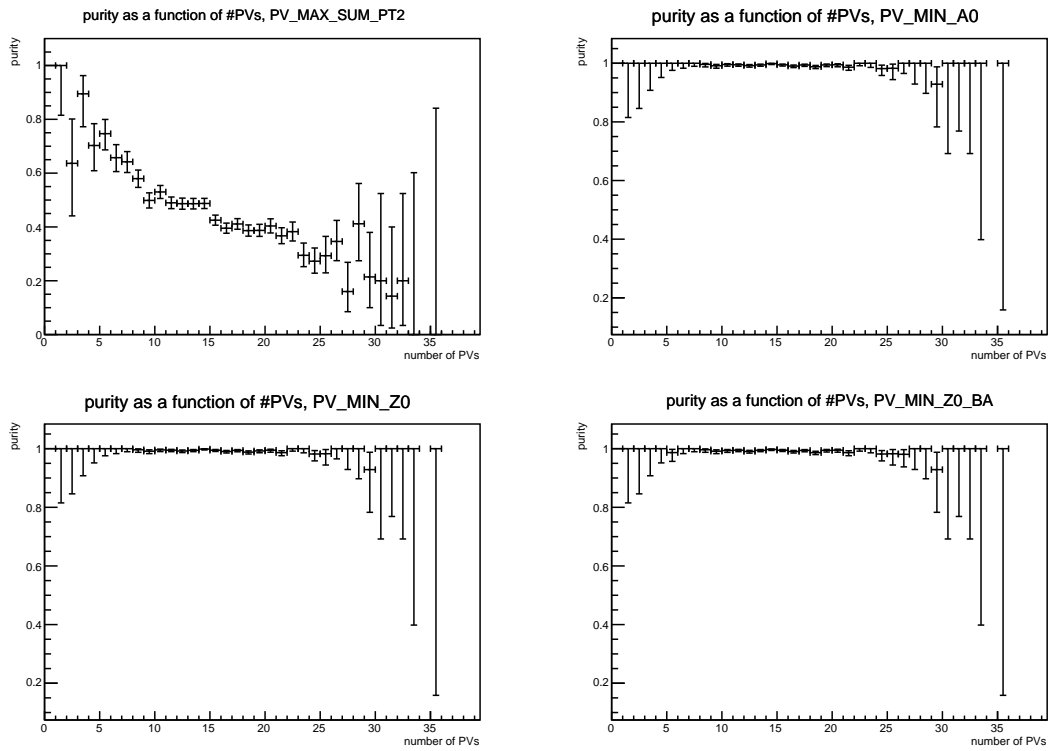
Figure 6: purity as a function of number of reconstructed PVs for the four approaches.

8 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 estimation of the available statistics for this iteration of the analysis compared to Run1 and the description of the different HLT algorithms (Lxy0, dalayed)

4: add

9 CutFlow

10 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 $\sim \times 1.6$ in the number of

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

In order to understand the effect of the tight muons usage, we compared the effect of a Run1-like BDT and the tight muons in the analysis.

Knowing the available statistics before and after the fake-BDT application in previous analysis [4] and the muon efficiency and fake-rate, we computed the Run1 background composition in terms of fake and real muons; applying the properties of tight muons we can estimate the effect tight muons would have had on the Run1 analysis. Tables 8 and 9 show the amount of statistics for the different components of the likelihood used in the fit for the signal yield extraction and the estimation of the statistics using tight muons. The combinatorial background doesn't increase significantly, the SS-SV (ans semileptonic) background increases of a factor $\sim \times 1.1$ and the paking background increases of a factor $\sim \times 2.6$. Signal 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.

Run1 analysis likelihood with the available statistics modified according to the estimation of the available stastics in this analysis performed in section 8 has been run, varying the number of events associated to the different background and signal models according to the usage of tight muons or a Run1-like BDT. Table 10 shows the root mean square of the distribution of the fitted number of B_s and B_d events. No significant effect on analysis sensitivity is visible for B_s, while a $\sim 1\%$ broader RMS is visible on B_d count.

Given the negligible difference due to the usage of tight muons instead of a Run1-like BDT for fake muons 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 8.

11 BackgroundModeling

11.1 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 $\approx$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 $\approx$20 events/100 MeV expected at the peak of the B_s^0 signal (section 16.3).

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 16.3, 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 [**PythiaB**] 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 PYTHIAand 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 and multiplicity of tracks of B-candidates 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 $B_s^0 \rightarrow \mu^+\mu^-$, 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 also to exclusive decays such as $B_c^+ \rightarrow J/\psi \, \mu^+ \nu$, $B \rightarrow K \, \mu^+ \mu^- (\gamma)$, $B_s^0 \rightarrow \mu^+ \mu^- (\gamma)$ (250 signal events are present in the four-corner MC sample).

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

12 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 [8] package of ROOT. We use 15 physical input variables to obtain the signal-to-background discriminator. These 15 input variables are summarised in Table 11 along with a short description of each variable. The BDT ranking of input variable importance is given in Table 12 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 $B_s^0 \to \mu^+ \mu^-$ 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. To associate tracks to pile-up vertices we use information from the standard ATLAS vertex reconstruction algorithms. As a measure of proximity of a track to the B vertex we define the variable $\ln \chi^2 = \ln((d_0/\sigma_{d_0})^2 + (z_0/\sigma_{z_0})^2)$, with the track perigee parameters d_0 and z_0 calculated with respect to the reconstructed B vertex. Figure 7 shows the $\ln \chi^2$ distribution for all reconstructed tracks in $B_s^0 \to \mu^+ \mu^-$ signal (red) and four-corner background (blue) MC events with at least one reconstructed B candidate. The large peak away from zero belongs to the tracks from other vertices, mostly pile-up. The presence of tracks associated with the B vertex (peak around zero) is higher in case of background events, while for signal there should be almost no tracks close to the B vertex. By introducing a "loose" association requirement, we can refine our search getting rid of most of the pile-up tracks without losing the interesting ones. The selection $\ln \chi^2(\text{SV,trk}) < 7$ is used in the definition of a new variable, $DOCA_{xtrk}$, which is the distance of closest approach (DOCA) of the extra track closest to the B vertex (also defined secondary vertex (SV)). Exploiting this SV-to-track association further, we define another discriminating variable N_{trks}^{close} as the number of tracks with "strong" B vertex association. The "strong" association is defined by requiring the tracks to have $\ln \chi^2(\text{SV,trk}) < 1$.

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.

The other 12 variables in Table 11 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
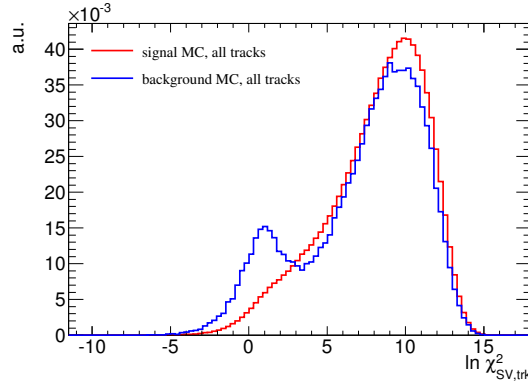- χ^2 between PV and B vertex in z
- p_T^B significance

Figure 7: Distribution of $\ln \chi^2 = \ln((d_0/\sigma_{d_0})^2 + (z_0/\sigma_{z_0})^2)$, with the track perigee parameters d_0 and z_0 calculated relative to SV, for all reconstructed tracks of di-muon events in $B_s^0 \to \mu^+\mu^-$ signal (red) and four-corner background (blue) MC.

- pointing angle $|\alpha_{3D}|$
- B vertex fit χ_B^2/NDF
- coplanarity of $B_s^0 \to \mu^+\mu^-$' decays
- coplanarity of $B_s^0 \to \mu^+\mu^-$' 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 $b\bar{b} \to \mu^+\mu^- X$ background MC (described in Sec. 11) 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 from the right mass sideband.[7] The left plot in Figure 8 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 2011 analysis: the new training results to be significantly more efficient.

The final choice of the configuration parameters of the *continuum*-BDT, as used by TMVA, is shown in Table 13. The optimal values for the configuration parameters *MinNodeSize* and *AdaBoostBeta* were found with the help of a grid scan using background rejection at 36% signal efficiency on the ROC-curve as a figure of merit of classifier performance. The *MaxDepth* parameter has also been studied in the similar way. For this parameter, it has been found that the performance of classifier improves with increasing of *MaxDepth*, whereas the discrepancy between its performance on the training and testing sample, accessed

---

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

Variable	Description
$\lvert\alpha_{2D}\rvert$	Absolute value of the angle in the transverse plane between $\Delta\vec{x}$ and $\vec{p}^B$
Isolation $I_{0.7}$	Ratio of $\lvert\vec{p}_T^B\rvert$ to the sum of $\lvert\vec{p}_T^B\rvert$ 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
$\chi^2_{\mu,xPV}$ $\log[\chi^2_{\mathrm{PV,SV}}]_{xy}$	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
$\lvert d_0\rvert^{max}$ 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
$\mathrm{DOCA}_{\mu\mu}$	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
$\lvert d_0\rvert^{min}$ 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 $(\Delta\vec{x} \cdot \vec{p}^B)/\lvert\vec{p}_T^B\rvert$
P_L^{min}	Minimum momentum of the two muon candidates along the B direction

Table 11: Description of the 15 discriminating variables used in the discrimination between signal and continuum background.

by the Kolmogorov-Smirnov test, becomes larger, leading to the over-training. In order to not compromise the generality of the classifier, the value *MaxDepth=3* has been chosen. The value of *NTrees* has been chosen large enough to allow the training to converge. The choice of other parameters does not have any significant impact on the training result.

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

Imp. Rank	Variable	Importance	Sep. Rank	Separation
1	$\lvert\alpha_{2D}\rvert$	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	$\lvert\Delta\phi_{\mu\mu}\rvert$	0.06478	12	0.01947
8	$\lvert IP_{B}^{3D}\rvert$	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 12: 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.

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 13: Configuration parameters of the *continuum*-BDT.

407 of the BDT proving that the chosen configuration is very robust.

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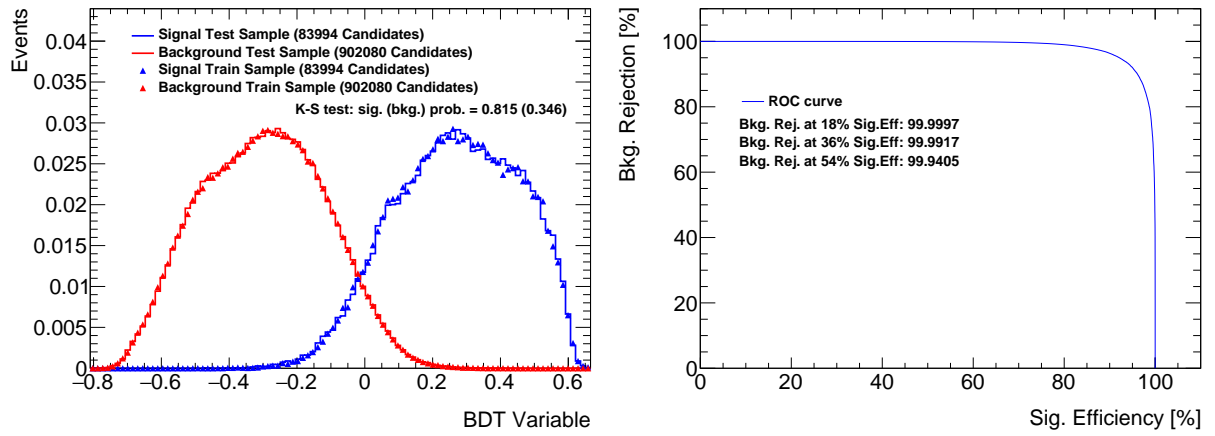410 described in this section is the one used in the analysis.

Figure 8: *Left*: cross-checks for over-training for the continuum BDT variable. *Right*: comparison of ROC curves trained on the combinatorial background MC and the signal MC. The ROCs are evaluated on the high-mass data sideband.

13 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].

13.1 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}, η ^{B} and B Isolation variables simultaneously with a gradient boosted re-weighting technique. Figure 9 shows the corrected p_T, η, and B Isolation variables as a cross-check.

Figure 10 shows the distributions of few discriminating variables. The rest can be checked in Appendix **??**.
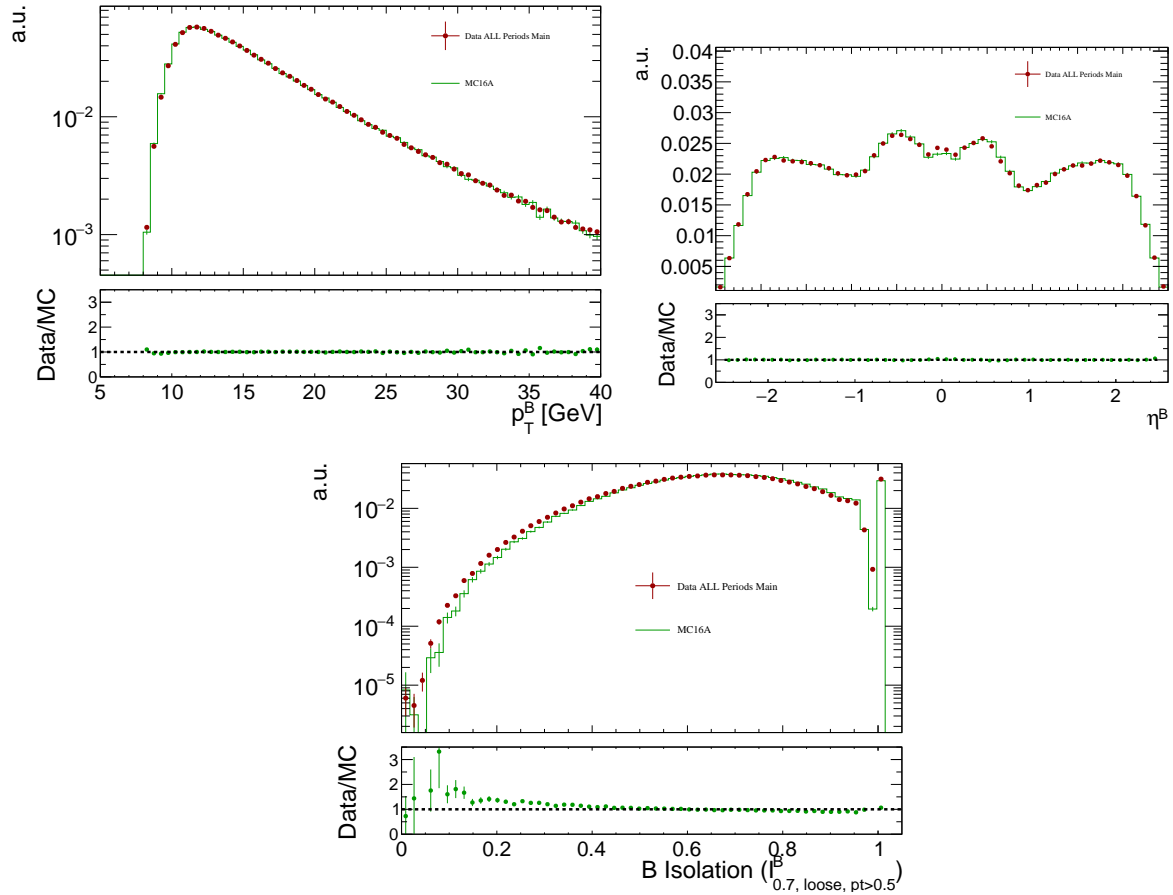

Figure 9: Cross-check on the p_T ^{B} and η^B distributions of the $J/\psi K$ candidates in data and signal MC. The red dots correspond to the sideband data during all data taking periods in Run 2 and the green histogram corresponds to the MC $b\bar{b} \to \mu\mu X$ sample. Both histograms are normalised to one.
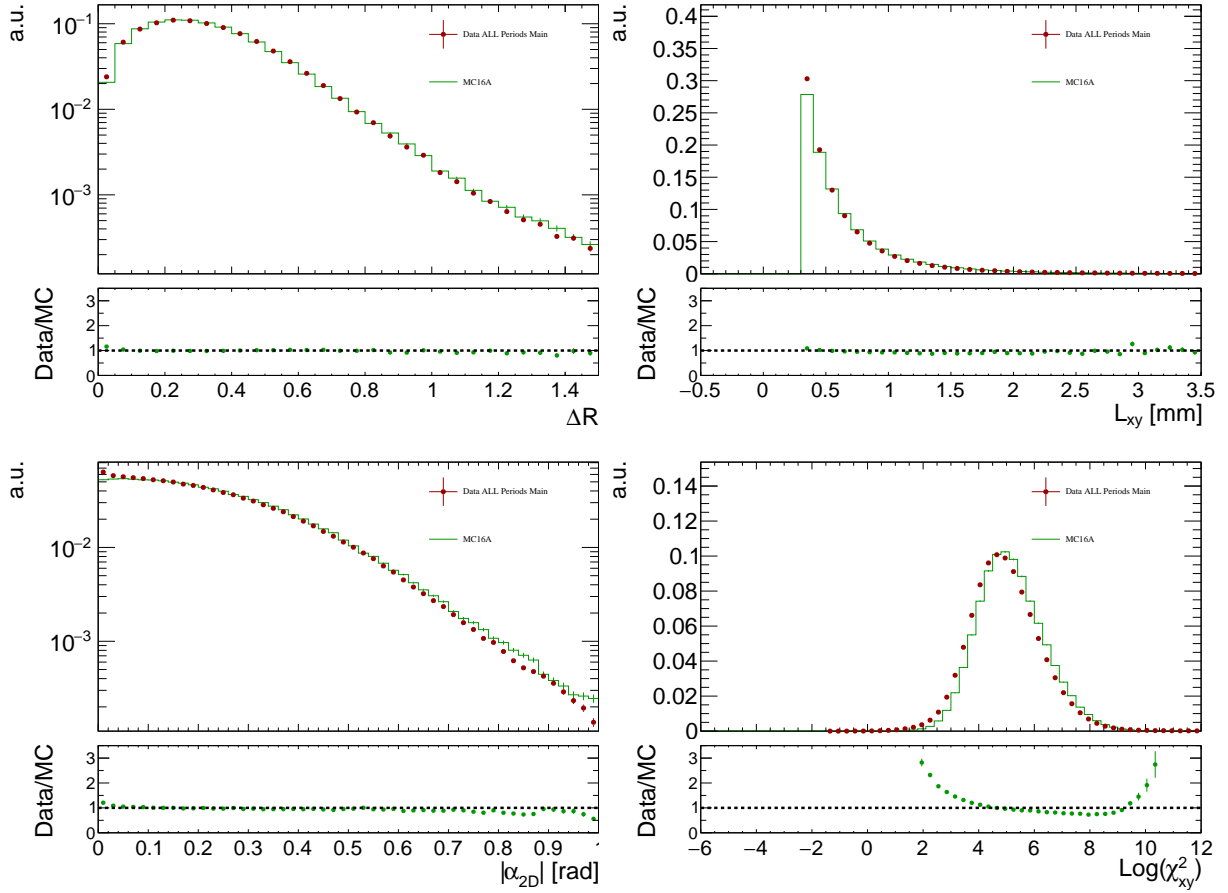
Figure 10: From left to right, from top to bottom: Data and signal background MC distributions of the ΔR, the transverse decay length (L_{xy}), $|\alpha_{2D}|$ and χ^2_{xy} that represents the separation between production (PV) and decay (SV) vertices. The green histogram corresponds to the sideband data, while the red points correspond to the re-weighted signal MC.

13.2 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.

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. [8] It is proven that the mis-reconstructed events have the same shape as the signal in the

---

[8] 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.

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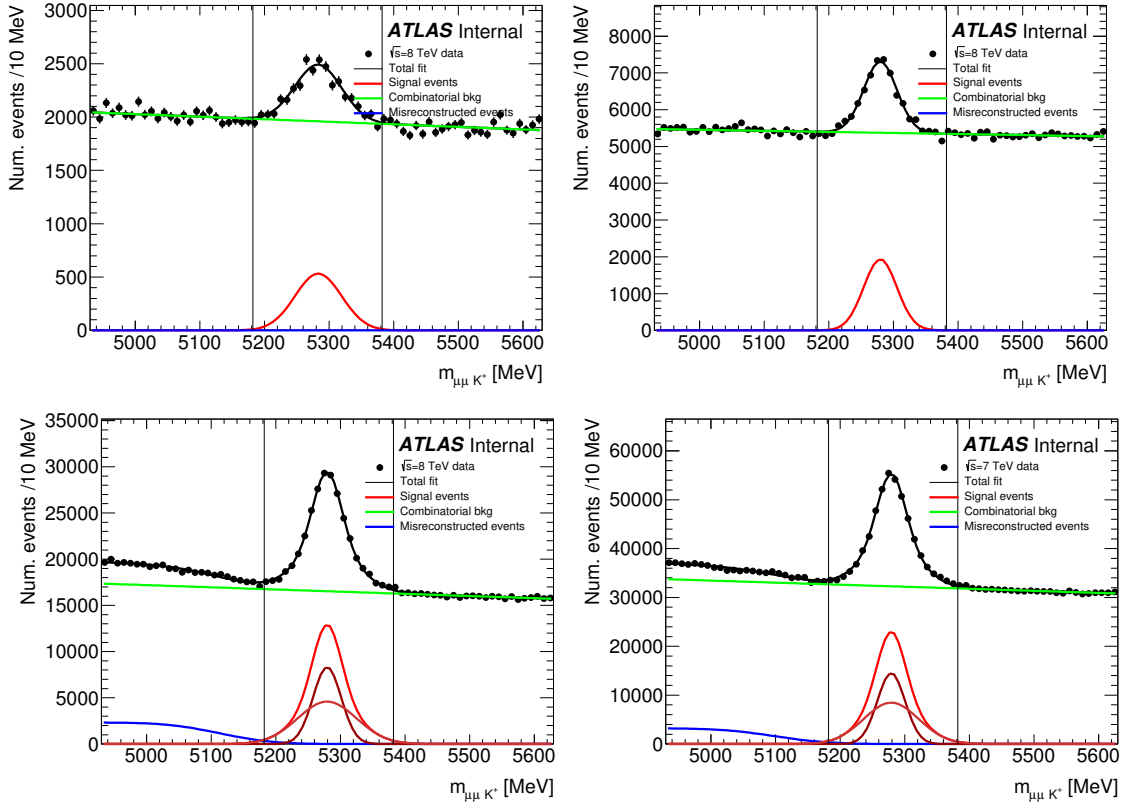

Figure 11: From left to right, top to bottom: fits to the invariant mass distributions of $B^+ \to J/\psi K^+$ events in the three trigger categories N1, N2 and N3, and in 2011 data, respectively.
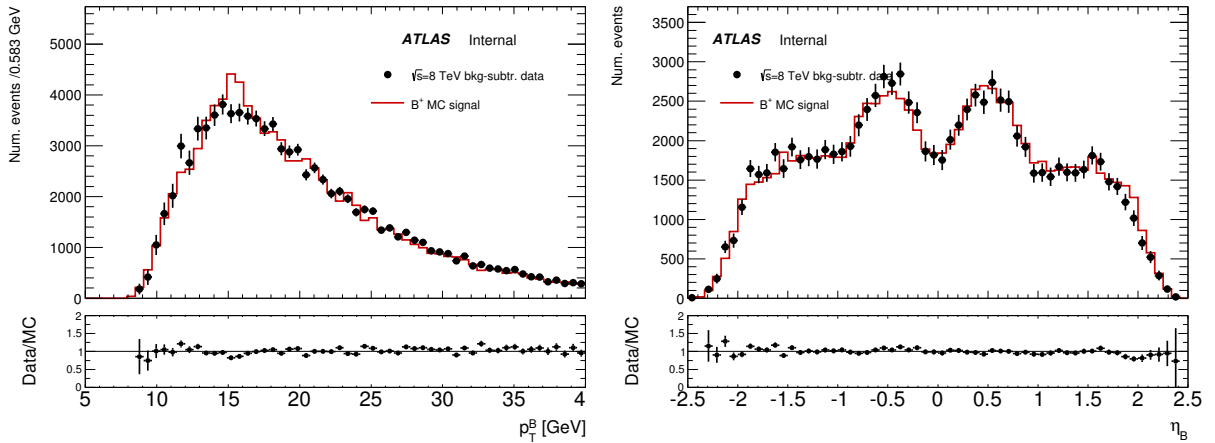

Figure 12: Cross-check on the $p_T^{\ B}$ and η^B distributions of the $J/\psi K$ candidates in data and signal MC. The black dots correspond to the sideband subtracted data, while the red histograms correspond to reweighted MC normalised to the number of data events.
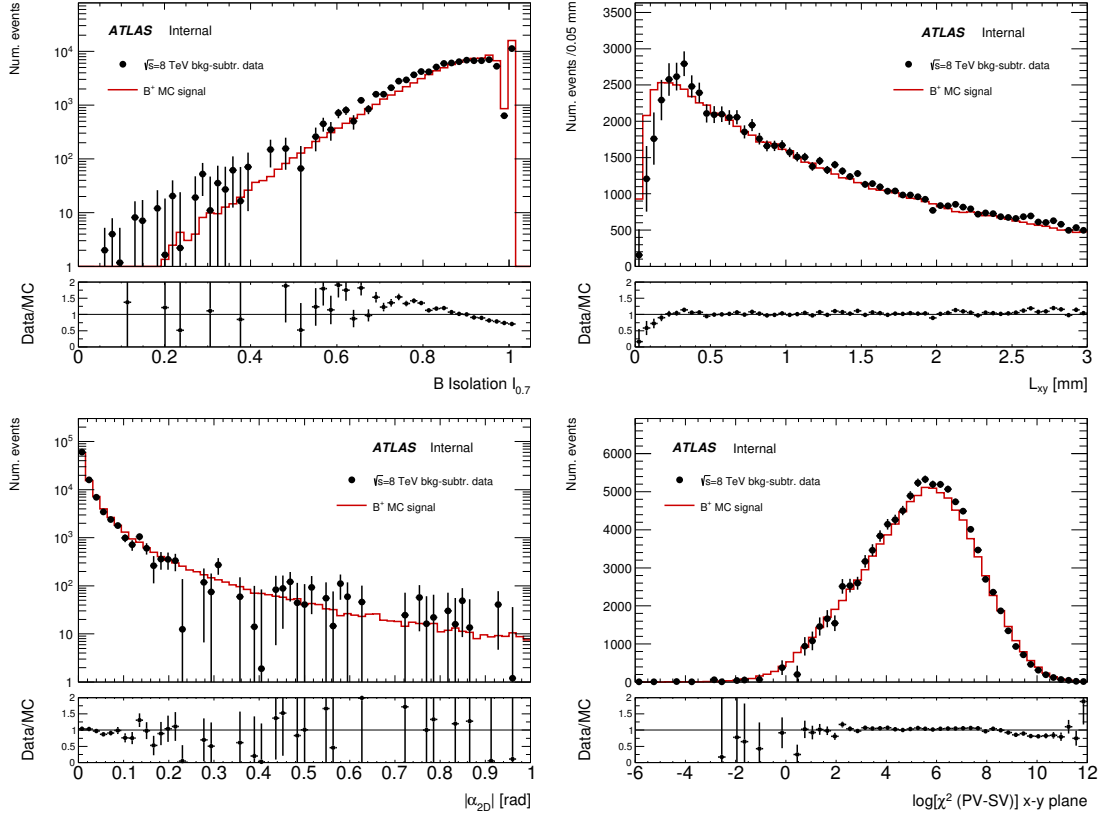
Figure 13: *From left to right, from top to bottom*: Data and signal MC distributions in $B^+ \rightarrow J/\psi K^+$ events for the B isolation variable, the transverse decay length (L_{xy}), $|\alpha_{2D}|$ and χ^2_{xy} that represents the separation between production (PV) and decay (SV) vertices. The black dots correspond to the sideband subtracted data, while the red points correspond to reweighted MC normalised to the number of data events.
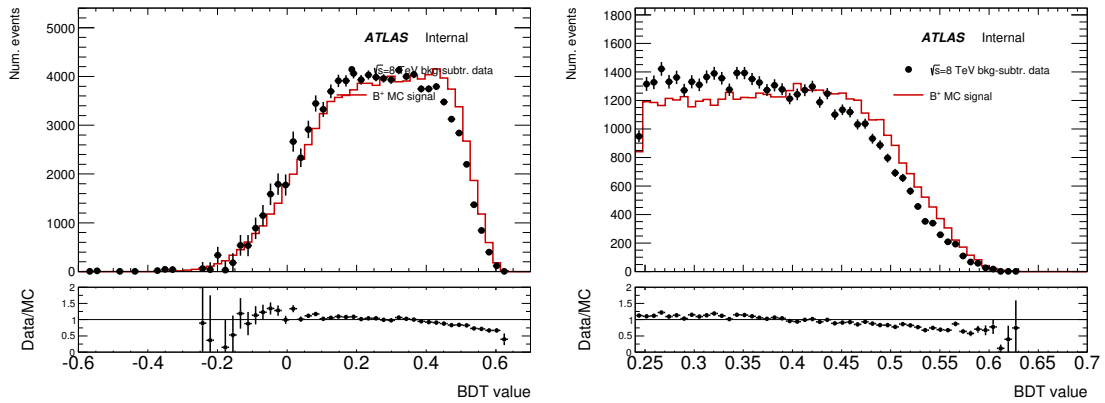

Figure 14: Data and signal MC distribution in $B^+ \rightarrow J/\psi K^+$ events for the continuum BDT variable. The black dots correspond to the sideband subtracted data, while the red points correspond to reweighted MC normalised to the number of data events. The left plot shows the same distribution zoomed in the region of interest of the analysis.

The shape from the rescaled sideband events is subtracted statistically from the signal region. The number of background events feeding into the signal region is obtained by a binned maximum likelihood fit to the mass distribution. Figure 11 shows these binned fits to the mass distributions in the three trigger categories. The fit model consists of a single (double) Gaussian for the signal in N1 and N2 (N3), an error function for the mis-reconstructed events and an exponential for the continuum background. The results of these binned fit analysis are perfectly compatible with the unbinned fit described in Section ?? and used for the yield extraction.

The distributions of p_T and η of the B is rechecked in figures 12, the distributions of few discriminating variables are shown in figures 13, and the distribution of the continuum BDT variable is shown in figure 14. The distributions of all the other variables are shown in the appendix ??.

Typically, except for deviations in individual bins, the overall shapes of distributions agree well between data and MC. The observed differences in shape are accounted for as systematics with the procedure described in Section ??. In addition, the discrepancy seen in the data-MC comparison of the continuum BDT in the $B^+ \to J/\psi K^+$ events is assessed and exploited in the evaluation of the efficiency in the $B_s^0 \to \mu^+ \mu^-$ signal in Section ??.

13.3 Alternate reference channel: $B_s^0 \to J/\psi\phi$

The same procedure can be applied on data reconstructed as $B_s^0 \to J/\psi\phi$. 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 15 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.
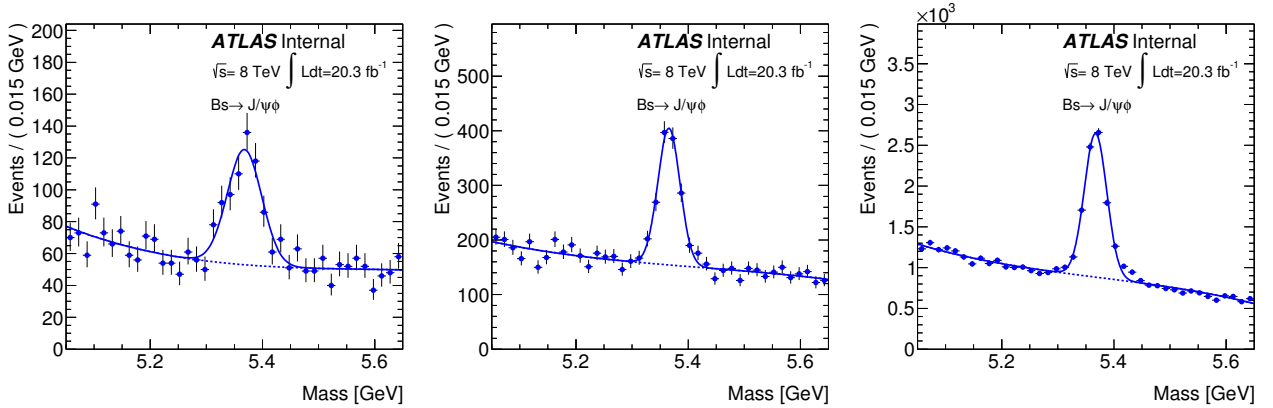

Figure 15: Fit to the invariant mass distributions of $B_s^0 \to J/\psi\phi$ events in each of the trigger categories N1 (*left*), N2 (*center*) and N3 (*right*).

The data/MC comparisons of p_T and η distributions for the B_s^0 are shown in Figure 16, while the distributions of few discriminating variables are shown in Figure 17.. Finally the distribution of the continuum BDT variable is shown in Figure 18. The signal distributions in data are obtained by subtracting the sideband shapes in each variable.

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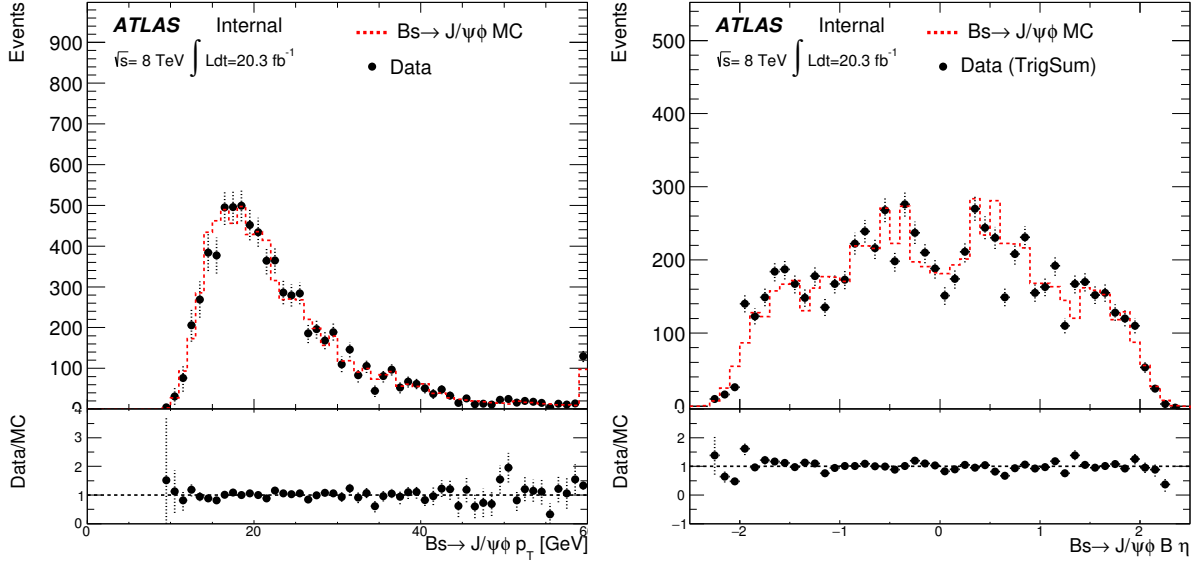

Figure 16: Cross-check on the $p_T{}^B$ and η^B distributions of the $J/\psi\phi$ candidates in data and signal MC. The black dots correspond to the sideband subtracted data, while the red histograms correspond to reweighted MC normalised to the number of data events.

13.4 Yield stability during run

460 The stability of the gain in the B_s^0 sidebands in the 2015 and 2016 data taking periods is shown in Figure 19.
461 The gain is calculated as the yield of events per period divided by the luminosity in that period. The gain
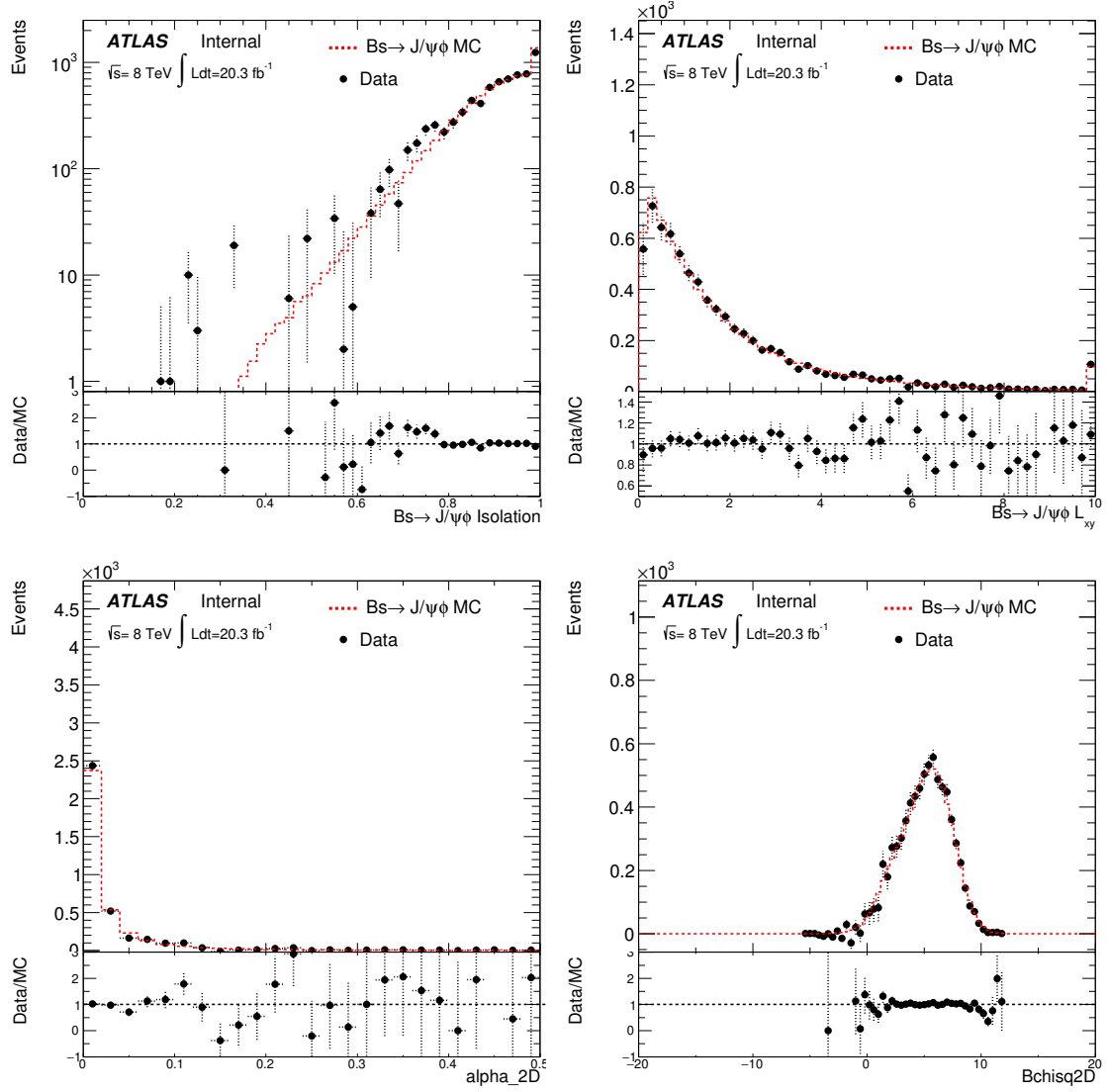462 appears to be stable during the run, within statistical uncertainties.

Figure 17: *From left to right, from top to bottom*: Data and signal MC distributions of $B_s^0 \rightarrow J/\psi\phi$ events for the B isolation variable, the transverse decay length (L_{xy}), $|\alpha_{2D}|$ and χ^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 18: 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 t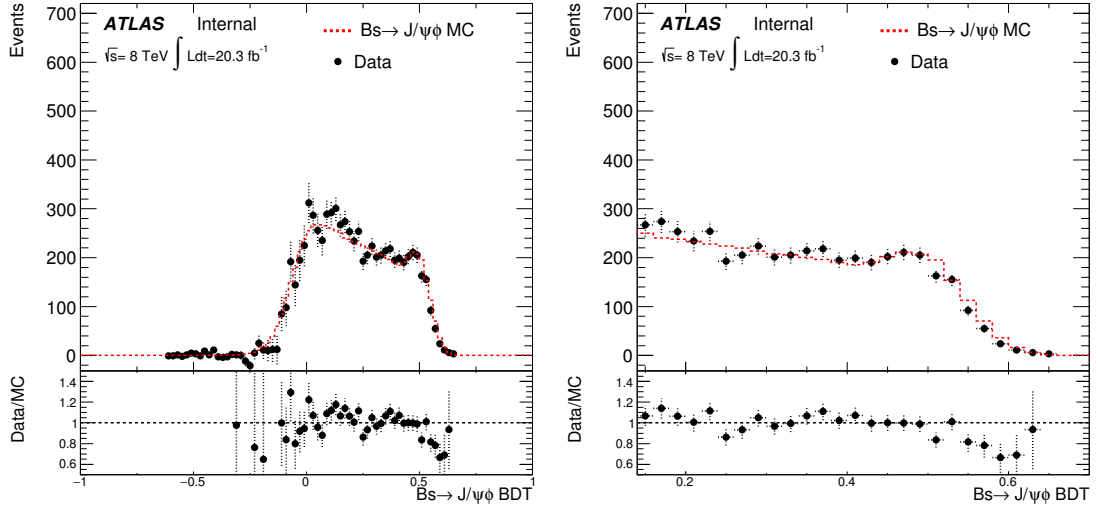o reweighted MC normalised to the number of data events. The left plot shows the same distribution zoomed in the region of interest of the analysis.
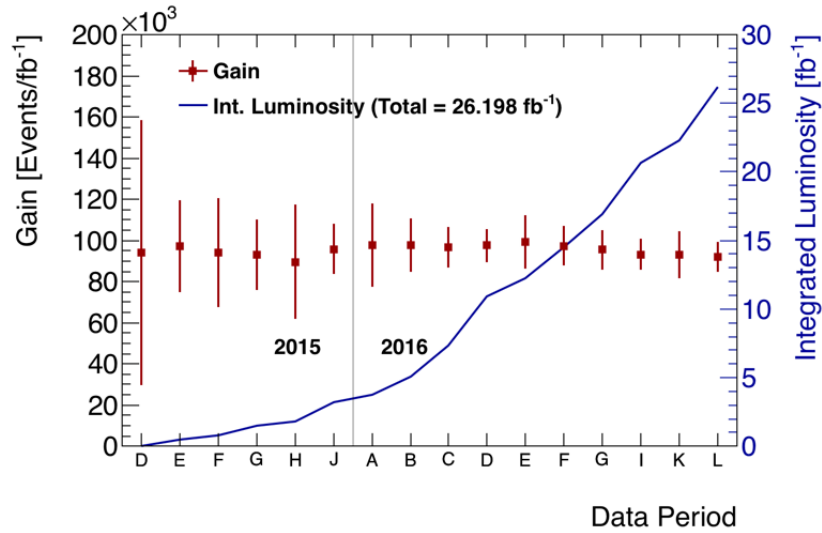

Figure 19: Stability of B_s^0 gain for 2015 and 2016 data taking periods. Also shown is the integrated luminosity for the analysis when the HLT_mu4_mu6_bBmumu and HLT_mu4_mu6_bBmumu_Lxy0 triggers are applied to the 2015 and 2016 data respectively.

463 # 14 BPlusYield

15 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 $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^+ \rightarrow J/\psi K^+$ 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:

$$R_{A\epsilon} = \frac{A_{B^+ \rightarrow J/\psi K^+} \times \epsilon_{B^+ \rightarrow J/\psi K^+}}{A_{B_s^0 \rightarrow \mu^+ \mu^-} \times \epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}} \qquad (3)$$

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 14.**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 12) 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 8). They will be applied to the numerator only.

The values for the $A \times \epsilon$ and $R_{A\epsilon}$ terms are summarised in Table 15, 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 15 are described in Section 15.1.

$A^{2012}_{B^+ \to J/\psi K^+}$	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 14: 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	$0.0080 \pm 0.28\% \pm 14.2\%$	$0.180 \pm 0.56\%$ (stat) $\pm 5.2\%$ (syst)
B^0_s	0.1522	$0.0441 \pm 0.49\% \pm 10.3\%$	

Table 15: 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.

15.1 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 15 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 13. $R_{A\epsilon}$ 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 15.

16 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 16.1 and 16.2.
- The results of the fit to the data in the sidebands and the interpolation in the signal region are presented in Section 16.3.
- A summary of the baseline fit configuration is given in Section 16.4.
- Systematic uncertainties on the fit are discussed in Section 16.5.

16.1 Signal and peaking background

The mass shape of the $B_s^0 \rightarrow \mu^+\mu^-$, as well as the one of $B^0 \rightarrow \mu^+\mu^-$, is described by a superposition 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 $B \rightarrow hh'$, mainly $B_s \rightarrow K^+K^-$ and $B_d \rightarrow K^\pm\pi^\mp$, in which both hadrons are misidentified as muons. Due to the mass distortion related to the $K \rightarrow \mu$ mass assignment, and the smaller one for $\pi \rightarrow \mu$, the mass distribution of these events is substantially superimposed with the B_d signal.

16.2 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., $B \rightarrow D\mu X \rightarrow \mu\mu X'$), 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., $B \rightarrow K\mu\mu$), 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 $h \rightarrow \mu$, in the limit of low energy neutrinos. In particular $B_d \rightarrow \pi\mu\nu$ and $B_s \rightarrow K\mu\nu$ can contribute, together with $\Lambda_b \rightarrow p\mu\nu$. 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.

16.3 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.

16.4 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.

16.5 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.

17 BranchingRatio

References

[1] C. Bobeth, M. Gorbahn, T. Hermann, M. Misiak, E. Stamou et al.,
$B_{s,d} \rightarrow l^+l^-$ 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 $B_s^0 \rightarrow \mu^+\mu^-$ branching fraction and effective lifetime and search for $B^0 \rightarrow \mu^+\mu^-$ decays*, Phys. Rev. Lett. **118** (2017) 191801, arXiv: 1703.05747 [hep-ex].

[3] V. Khachatryan et al.,
Observation of the rare $B_s^0 \rightarrow \mu^+\mu^-$ 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 $\rightarrow \mu\mu$ 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 $B_s \rightarrow \mu\mu$ 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 $B_s^0 \rightarrow \mu\mu$ with the ATLAS detector*, Phys.Lett. **B713** (2012) 387,
arXiv: 1204.0735 [hep-ex].

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

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, 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.