

Search for the $B^0 \rightarrow \mu^+ \mu^-$ decay and measurement of the $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction and effective lifetime

F. Archilli¹, M.-O. Bettler², X. Cid Vidal³, H. V. Cliff⁴, F. Dettori², H. M. Evans⁴, V. Gibson⁴, G. Mancinelli⁵, D. Martínez Santos³, A. Mordà⁵, M. Mulder¹, M. Palutan⁶, A. Pellegrino¹, M. Rama⁷, M. Santimaria⁶, A. Sarti^{6,8}, M. Schlupp⁹, B. Sciascia⁶, S. Tolk⁴, R. Vazquez Gomez⁶, J. B. Zonneveld¹.

¹ Nikhef, National Institute for Subatomic Physics, Amsterdam, Netherlands.

² European Organisation for Nuclear Research (CERN), Geneva, Switzerland.

³ Universidade de Santiago de Compostela, Santiago, Spain.

⁴ Cavendish Laboratory, University of Cambridge, Cambridge, United Kingdom

⁵ CPPM, Aix-Marseille Université, CNRS/IN2P3, Marseille, France.

⁶ Laboratori Nazionali di Frascati, INFN, Frascati, Italy.

⁷ Sezione di Pisa dell'INFN, Pisa, Italy.

⁸ Dipartimento di Scienze di Base e Applicate per l'Ingegneria, "La Sapienza" Università di Roma, Italy.

⁹ Fakultät Physik, TU Dortmund, Dortmund, Germany

Abstract

A search for the rare decays $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ is performed at the LHCb experiment using data collected in pp collisions corresponding to a total integrated luminosity of 4.4 fb^{-1} . An excess of $B_s^0 \rightarrow \mu^+ \mu^-$ events is observed with a significance of 7.8 standard deviations, representing the first observation of this decay in a single experiment. The branching fraction $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (2.8 \pm 0.6) \times 10^{-9}$ is measured, with the uncertainty including both the statistical and systematic components. The first measurement of the $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime, $\tau(B_s^0 \rightarrow \mu^+ \mu^-) = 2.04 \pm 0.44 \pm 0.05 \text{ ps}$, is reported. No excess of $B^0 \rightarrow \mu^+ \mu^-$ events is found and a 95 % confidence level upper limit $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) < 6.1 \times 10^{-10}$ is determined. All results are in agreement with the Standard Model expectations.

1 Contents

2	0	Preface	4
3	1	Introduction	7
4	2	Analysis strategy	8
5	2.1	BF measurement	8
6	2.2	Effective lifetime measurement	9
7	3	Data and Monte Carlo samples	11
8	3.1	Data sample	11
9	3.2	Monte Carlo samples	11
10	3.3	Trigger	13
11	4	Selection	14
12	4.1	Stripping and pre-selection	14
13	4.2	The BDTS discriminant	14
14	4.3	Particle Identification	16
15	4.3.1	Muon Identification	16
16	4.3.2	Pion, kaon and proton misID	18
17	4.3.3	PID selection strategy for 2016 data	19
18	4.3.4	About the use of MC12TuneV2 ProbNN version for 2015 data	23
19	5	Peaking backgrounds	27
20	5.1	$B \rightarrow h^+ h^-$	27
21	5.1.1	$B \rightarrow h^+ h^-$ estimate as a function of BDT	28
22	5.1.2	Systematics on $B \rightarrow h^+ h^-$ estimate	29
23	5.1.3	$B \rightarrow h^+ h^-$ mass pdf	33
24	5.2	Peaking background other than $B \rightarrow h^+ h^-$	34
25	5.2.1	$B^0 \rightarrow \pi^- \mu^+ \nu_\mu$	35
26	5.2.2	$B_s^0 \rightarrow K^- \mu^+ \nu_\mu$	38
27	5.2.3	$B^+ \rightarrow \pi^+ \mu^+ \mu^-$ and $B^0 \rightarrow \pi^0 \mu^+ \mu^-$	41
28	5.2.4	$\Lambda_b^0 \rightarrow p \mu^- \nu$	43
29	5.2.5	$B_c^+ \rightarrow J/\psi \mu^+ \nu$	45
30	5.3	Background summary tables	49
31	6	Signal classification	56
32	6.1	Track isolations	56
33	6.1.1	Cut-based long track isolation used in the 2013 analysis	56
34	6.1.2	Development of new long track isolation	57
35	6.1.3	Development of VELO-only track isolation	60
36	6.1.4	Performance on data	60
37	6.2	Investigation of other isolation variables	61

38	6.3	Global BDT definition	63
39	6.3.1	Procedure to identify the best discriminating variables	63
40	6.3.2	BDT characterization and comparison with BDT 2013	65
41	6.3.3	Performance in Run I and Run II data	65
42	6.4	BDT calibration	70
43	6.4.1	$\Delta LL_{K-\pi}$ cut efficiency determination	71
44	6.4.2	BDT calibration with $B^0 \rightarrow K^+\pi^-$	73
45	6.4.3	Systematics	80
46	6.5	Invariant mass	90
47	6.5.1	Mean	90
48	6.5.2	Invariant mass resolution	93
49	6.5.3	Crystal Ball parameters α and n	96
50	6.5.4	Combination of Run I and Run II parameters	99
51	7	Normalisation	101
52	7.1	Normalisation channels yields	102
53	7.1.1	$B^+ \rightarrow J/\psi K^+$ channel	102
54	7.1.2	$B^0 \rightarrow K^+\pi^-$ channel	104
55	7.2	Geometrical detector acceptance	106
56	7.3	Ratio of reconstruction and selection efficiencies	106
57	7.3.1	Corrections to the simulated selection efficiency	107
58	7.4	Ratio of trigger efficiencies	109
59	7.5	Normalisation cross-checks	111
60	7.6	Ratio of hadronisation fractions	113
61	7.7	Normalization factors	114
62	8	Time dependent effects	123
63	8.1	Effect of effective lifetime on the distribution of the BDT response	123
64	9	BF results	127
65	9.1	Toy MC studies	127
66	9.2	Unblinded results	138
67	9.3	Probability to observe N events in the 2015 signal region	138
68	9.4	Kolmogorov-Smirnov test of the BDT distribution in the 2015 and 2016 datasets	138
69	9.5	Compatibility of fit results from different data samples	139
70	9.6	Fit to Run 1 and Run 2 datasets	140
72	10	Effective lifetime measurement	143
73	10.1	Data sample and selection	143
74	10.1.1	PID requirements	144
75	10.1.2	BDT requirement optimisation	145
76	10.2	Decay time acceptance	149
77	10.3	Mass and decay time fit	155

78	10.3.1 Correlation of mass and decay time	156
79	10.3.2 Mass PDFs	156
80	10.3.3 Decay time fit	157
81	10.3.4 The trouble with τ (at low statistics)	157
82	10.3.5 Optimising the fit	161
83	10.4 Systematic uncertainties	165
84	10.4.1 Fit accuracy	165
85	10.4.2 Contamination from exclusive backgrounds	168
86	10.4.3 Effect of the production asymmetry	170
87	10.4.4 The effect of decay time acceptance on the mix of light and heavy mass eigenstates	172
88	10.4.5 Mass model	173
89	10.4.6 Decay time acceptance	173
90	10.4.7 Combinatorial background decay time model	177
91	10.4.8 Incorrect assignment of primary vertices cross check	179
92	10.4.9 Total systematic uncertainty	182
93	10.5 Results	183
94	11 Conclusions	185
95	A Description of the toy model used for the lifetime measurement	186
96	B Distributions of the long track isolation input variables	191
97	C Definition of BDT 2013 input variables	193
98	D Distribution of BDT input variables in MC training sample, more recent MC and data	194
99	E Comparison of BDT training strategies	195
100	F Comparison of BDT input variables in Run I, 2015 and 2016 data	198
101	G Trigger BDT pdf's for signal and double misID	199
102	H $\pi - \mu$ and $K - \mu$ fits for 2015 and 2016 data	200
103	I $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$ mass pdf's with FF weights	200
104	J J/ψ veto	201
105	References	210

¹⁰⁸ **0 Preface**

¹⁰⁹ **Version 2 (14 September 2016)**

¹¹⁰ **Update after first round of comments from WG review of lifetime analysis**

¹¹¹

¹¹² Sections related to the lifetime analysis: 2, 3, 4, 6.1-6.3, 10.

¹¹³

¹¹⁴ Changes related to the BF analysis (not yet under review):

¹¹⁵ • Updated section on muon Id selection (sec. 4.3.1)

¹¹⁶ • New section on peaking backgrounds (sec. 5)

¹¹⁷ • Updated section on BDT definition (some more info relevant to BF analysis, sec. 6.3)

¹¹⁸ • New section on BDT calibration (sec. 6.4)

¹¹⁹ • Section on BF normalization (sec. 7)

¹²⁰ **Version 3 (23 September 2016)**

¹²¹ **BF analysis version for WG circulation and update of lifetime analysis after
122 WG approval**

¹²³

¹²⁴ Sections related to the lifetime analysis: 2, 3, 4, 6.1-6.3, 10.

¹²⁵

¹²⁶ Changes related to the BF analysis compared to Version 2:

¹²⁷ • Updated section on muon Id selection (sec. 4.3.1)

¹²⁸ • Updated section on peaking backgrounds (sec. 5)

¹²⁹ • Minor update of BDT definition section (sec. 6.3)

¹³⁰ • Minor update of BDT calibration section (sec. 6.4)

¹³¹ • Added section on invariant mass calibration (sec. 6.5)

¹³² • Updated section on BF normalization (sec. 7)

¹³³ • Added section on lifetime-dependent corrections (sec. 8)

¹³⁴ **Version 4 (7 October 2016)**

¹³⁵ **BF analysis version after WG approval**

¹³⁶

¹³⁷ Changes with respect to Version 3:

¹³⁸ • Updated sections 4-7

- Part related to Run1 mostly complete (see 7.6-7.7 for work in progress on normalisation)
- Part related to Run 2 2015 mostly complete, some more work in progress in sec. 7
- Initial analysis of 2016 data in sec. 4.3.1 and 6.1-6.3

Version 5 (24 October 2016)

Updated version after comments from the RC

Main changes with respect to Version 4:

- Updated sections 4,6-8,10 taking into account comments from RC.
- Invariant mass calibration finalized including 2016 data
- $BF(B \rightarrow K\pi)/BF(B \rightarrow J/\psi K)$ and f_s/f_d stability checks performed including 2016 data (preliminary)

Version 6 (28 October 2016)

Addressed RC comments on Peaking bkg section, updated normalisation section

Main changes with respect to Version 5

- Updated sec. 5 on exclusive backgrounds. Addressed remarks from RC.
- Updated PID systematics of $B \rightarrow K\pi$ yields
- Updated normalisation sec. 7, including estimate of $BF(B \rightarrow K\pi)/BF(B \rightarrow J/\psi K)$ for 2016 (although not final yet), evaluation of α/β parameters and of expected number of signal events.

Version 7 (7 November 2016)

Main changes with respect to Version 6

- Updated BDT calibration including 2016
- Updated normalisation section with all inputs for 2016, new $BF(B_s \rightarrow J/\psi\phi)/BF(B \rightarrow J/\psi K)$ check and updated $BF(B \rightarrow K\pi)/BF(B \rightarrow J/\psi K)$ check after a bug fix on the way the PID eff correction to $B \rightarrow K\pi$ yields was computed
- Updated lifetime section including $B \rightarrow K\pi$ lifetime measurement in whole sample and comparison between τ^{-1} and τ extraction

Version 8 (9 November 2016)

Main changes with respect to Version 7

- 171 ● Updated background section (sec. 5) following the reply to RC comments sent on 3
172 Nov 2016

173 **Version 9 (30 November 2016)**

174 Main changes with respect to Version 8

- 175 ● Updated the $B^0 \rightarrow \pi^0 \mu^+ \mu^-$ vs $B^+ \rightarrow \pi^+ \mu^+ \mu^-$ efficiencies, now agreeing (sec. 5.2.3)
176 ● New procedure to determine the PID efficiency in the $B^0 \rightarrow K^+ \pi^-$ selection used
177 for normalisation and to determine the BDT pdf. Updated BDT pdfs, updated
178 $B^0 \rightarrow K^+ \pi^-$ normalisation and $BF(B_s \rightarrow J/\psi \phi)/BF(B \rightarrow J/\psi K)$ check (sec.6.4)
179 ● Added additional information on J/ψ veto performance in sec. 5.2.5 and appendix J.
180 ● Updated trigger selection of lifetime analysis (now TIS or TOS at all trigger levels)
181 following the discussion with RC on Nov 10. Added systematic check of mis-assigned
182 PV (sec. 10).

183 **Version 10 (2 December 2016)**

184 Version to ask permission to unblind.

185 **Version 11 (29 January 2017)**

186 Version with final results before the 1st paper circulation

- 187 ● Unblinded branching fractions and effective lifetime results
188 ● Checks done after the branching fraction unblinding

1 Introduction

The $B_s^0 \rightarrow \mu^+ \mu^-$ decay is a flavour-changing neutral current (FCNC) process that is highly suppressed in the Standard Model (SM). The decay proceeds via loop-diagrams and so can be significantly affected by the presence of non-SM particles, particularly additional Higgs bosons, entering at tree or loop level. The branching fraction of this decay mode has attracted considerable theoretical and experimental interest, culminating in a recent observation of the decay mode by the LHCb and CMS experiments [1]. At the current level of precision the measured branching fraction is in agreement with the SM predictions [2, 3], putting strong constraints on new physics models.

Though this result is disappointing for those hoping to find evidence of physics beyond the SM, all is not lost. More precise measurements of the $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction using the 13 TeV data currently being collected by LHCb, CMS and ATLAS may yet produce deviations from the SM. Measurements of the ratio of the branching fraction of the even rarer $B^0 \rightarrow \mu^+ \mu^-$ decay to the $B_s^0 \rightarrow \mu^+ \mu^-$ decay may also reveal new physics. Furthermore, even if these measurements stubbornly agree with the SM, the effective lifetime of the $B_s^0 \rightarrow \mu^+ \mu^-$ decay could yet provide a lifeline, potentially revealing new physics effects that are not apparent in the branching fraction measurements [3, 4]. This note details measurements of the $B^0 \rightarrow \mu^+ \mu^-$ branching fraction and the $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime using data taken by the LHCb experiment during 2011, 2012 and 2015 at centre of mass energies of 7, 8 and 13 TeV respectively and the possibility of using data collected during 2016 at 13 TeV as well.

The effective lifetime is simply the mean decay time of an unbiased sample of $B_s^0 \rightarrow \mu^+ \mu^-$ decays;

$$\tau_{\mu^+ \mu^-} \equiv \frac{\int_0^\infty t \langle \Gamma(B_s^0 \rightarrow \mu^+ \mu^-) \rangle dt}{\int_0^\infty \langle \Gamma(B_s^0 \rightarrow \mu^+ \mu^-) \rangle dt} \quad (1)$$

$$= \frac{\tau_{B_s}}{1 - y_s^2} \left[\frac{1 + 2\mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-} y_s + y_s^2}{1 + \mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-} y_s} \right] \quad (2)$$

where t is the proper decay time of the B_s^0 meson and y_s and $\mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-}$ are defined as

$$y_s \equiv \frac{\Delta\Gamma_s}{2\Gamma_s} \quad (3)$$

$$\mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-} \equiv \frac{R_H^{\mu^+ \mu^-} - R_L^{\mu^+ \mu^-}}{R_H^{\mu^+ \mu^-} + R_L^{\mu^+ \mu^-}} \quad (4)$$

and $R_H^{\mu^+ \mu^-}$ and $R_L^{\mu^+ \mu^-}$ set the contributions of the heavy and light mass eigenstates of the B_s system to the untagged $B_s^0 \rightarrow \mu^+ \mu^-$ decay rate. The $\mu^+ \mu^-$ final state is CP odd and so in the SM $\mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-} = +1$ and the effective lifetime is equal to the lifetime of the heavy B_s mass eigenstate. As discussed in [?] $\mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-}$ may be moved away from 1 by new physics effects, particularly those related to (pseudo-)scalar $l^+ l^-$ densities of four-fermion operators, revealing new physics even in the event that the branching fraction agrees with the SM. This is illustrated by Figure 1.

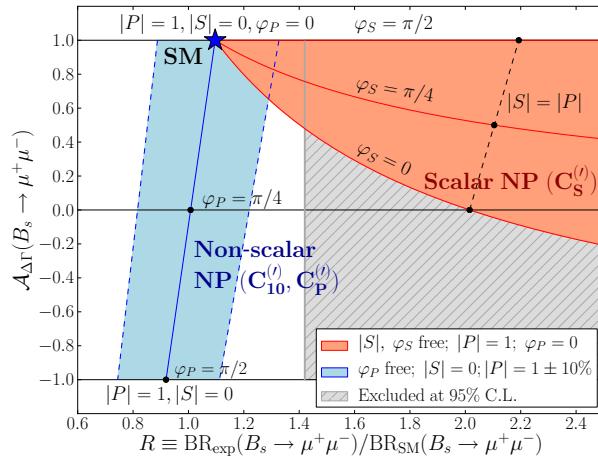


Figure 1: Illustration of allowed regions in the $R - \mathcal{A}_{\Delta\Gamma}(B_s \rightarrow \mu^+\mu^-)$ plane for scenarios with scalar or non-scalar NP contributions [?].

2 Analysis strategy

2.1 BF measurement

The $B_s^0 \rightarrow \mu^+\mu^-$ decay has been recently observed with a significance of 6.2σ through the combined analysis of LHCb and CMS data of Run I [1]. The measured BR, $2.8_{-0.6}^{+0.7} \times 10^{-9}$, is in agreement with the SM prediction, $(3.65 \pm 0.23) \times 10^{-9}$ [5]. With the same data, an evidence of the $B^0 \rightarrow \mu^+\mu^-$ decay with 3.2σ significance was observed (1.8σ and 2.6σ from LHCb and CMS data, respectively). The measured BR is in this case $3.9_{-1.4}^{+1.6} \times 10^{-10}$, which is 2.2σ above the SM prediction, $(1.06 \pm 0.09) \times 10^{-10}$ [5].

We remind here that a departure from the SM prediction on the B_d/B_s ratio would falsify the Minimal Flavour Violation (MFV) hypothesis, consisting of identifying the flavour symmetry and symmetry breaking structure of the SM and enforcing it also for physics beyond the SM (BSM). The latter hypothesis is indeed an excellent phenomenological explanation for the lack of BSM physics signals in the flavor physics data so far. For the above reasons, the clarification of the experimental picture on the $B^0 \rightarrow \mu^+\mu^-$ decay rate is the challenge for this analysis during LHC Run II.

The main backgrounds in this analysis are the dimuon combinatorial events (combinatorics of two muons coming from different B -hadron decay chains), which affect both the B^0 and B_s^0 signal regions, and the $B_{(s)}^0 \rightarrow h^+h^{(\prime)-}$ peaking background ($h, h' = K, \pi$), which affect mainly the B^0 .

For the combinatorial background, the rejection power essentially rests on the distance of flight and impact parameter resolutions, and on the muon isolation criteria, which are able to disentangle the muon tracks from b or c semileptonic decays. All the relevant variables are combined into a BDT, which is used, together with the dimuon invariant mass, to rate events according to their signal likeness. For the muon isolation variable,

244 a substantial improvement has been achieved in this work with respect to the previous
245 published analysis [6], where the isolation was defined using rectangular cuts. This
246 improvement reflects into a much better combinatorial background rejection in the signal
247 sensitive region at high output values for the BDT. This will be discussed in Sec. 6.1.

248 For the $B_{(s)}^0 \rightarrow h^+ h^{(\prime)-}$ peaking background, the rejection power is strongly dependent
249 on the PID selection. While in the previous version of the analysis [6] strong emphasis
250 was put on keeping a high efficiency on B_s^0 , which translated into a loose PID selection,
251 in this work we reoptimized the PID cuts in order to maximize the sensitivity on B^0 . In
252 this way, a large reduction of $B_{(s)}^0 \rightarrow h^+ h^-$ background has been achieved, accepting a
253 mild signal reduction. This will be discussed in Sec. 5.1. For what concern instead the
254 $B_{(s)}^0 \rightarrow h^+ h^{(\prime)-}$ normalization, the strategy is unchanged: it is determined by measuring
255 the yields of the various $B_{(s)}^0 \rightarrow h^+ h^{(\prime)-}$ modes directly on data, and combining them with
256 the $\pi \rightarrow \mu$ and $K \rightarrow \mu$ misID probabilities, as measured from data control samples.

257 Besides the $B_{(s)}^0 \rightarrow h^+ h^{(\prime)-}$, there are other exclusive backgrounds which are relevant
258 to the analysis: $\Lambda_b^0 \rightarrow p\mu^-\nu$ with proton misidentified as a muon, which pollutes the
259 signal region, and other semileptonic decays polluting the lower mass sideband only:
260 $B^0 \rightarrow \pi^-\mu^+\nu_\mu$, $B_s^0 \rightarrow K^-\mu^+\nu_\mu$ and $B^{0(+)} \rightarrow \pi^{0(+)}\mu^+\mu^-$. Eventhough they only pollute
261 the lower mass sideband, they can introduce a bias in the determination of the combinatorial
262 background from the mass fit, so they have to be carefully estimated. To this purpose,
263 these decays are normalised relative to the $B^+ \rightarrow J/\psi K^+$, according to their measured
264 (or predicted) BFs, and using the MC efficiencies for all selection steps but for the PID
265 cuts, which are evaluated instead from data control samples. This analysis contains many
266 improvements in the treatment of these backgrounds, the most important being: the use
267 of proton PID to reject $\Lambda_b^0 \rightarrow p\mu^-\nu$, the use of several new LHCb BF measurements to
268 constrain the yields of the events, and the introduction of a simultaneous fit to data control
269 samples to further constrain the yields of $B^0 \rightarrow \pi^-\mu^+\nu_\mu$ and $B_s^0 \rightarrow K^-\mu^+\nu_\mu$, reducing the
270 dependence on the form factors used in the simulation. Details will be given in Sec. 5.2.

271discussion follows on the normalization

272 2.2 Effective lifetime measurement

273 The strategy adopted for the analysis is informed to a large extent by the expected precision
274 of the measurement. In the Standard Model $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} = +1$, meaning that $\tau_{\mu^+\mu^-}$ is equal to
275 the lifetime of the heavy B_s mass eigenstate, which is evaluated by the Heavy Flavour
276 Averaging Group (HFAG) as $\tau_H = (1.610 \pm 0.012)$ ps [7].

277 New physics may move $\tau_{\mu^+\mu^-}$ towards the lifetime of the light B_s mass eigenstate,
278 which is similarly evaluated as $\tau_L = (1.422 \pm 0.008)$ ps [7]. The difference between these
279 two extreme cases is $\Delta\tau = 0.188$ ps, so to distinguish $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} = +1$ from $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} = -1$ at
280 five standard deviations would require a precision smaller than 0.038 ps on the effective
281 lifetime.

282 In contrast, sensitivity studies have indicated a likely precision for a measurement of
283 $\tau_{\mu^+\mu^-}$ with the 3 fb^{-1} Run I data sample of around 0.4 ps. Based on the LHCb branching

284 fraction analysis, the number of candidates in the Run I data is around 15. The key
285 challenge of this measurement is therefore how to design a reliable lifetime fit that produces
286 accurate results with such a small number of candidates. Although 2015 data will be
287 included in the measurement as well, only a small increase is expected in the number of
288 candidates.

289 The approach is therefore driven by simplicity. First a requirement is made on a
290 multivariate classifier to remove combinatorial background. An extended maximum
291 likelihood fit to the di-muon invariant mass spectrum is performed and sWeights are
292 calculated using the *sPlot* method [8]. A maximum likelihood fit is then performed on the
293 sWeighted decay time distribution to determine $\tau_{\mu^+\mu^-}$.

294 More complex approaches were investigated, including two-dimensional fits to invariant
295 mass and decay time, as well as simultaneous fits in bins of the multivariate classifier. It
296 was found that these approaches did not produce stable results due to the small number
297 of candidates.

298 **3 Data and Monte Carlo samples**

299 **3.1 Data sample**

300 The measurements described in this note are made using pp collision data collected
301 by the LHCb experiment in 2011, 2012 and 2015 at $\sqrt{s} = 7\text{ TeV}$, 8 TeV and 13 TeV
302 respectively. Table 1 lists the integrated luminosities collected each year and the versions
303 of the reconstruction software used.

Year	\sqrt{s} (TeV)	Integrated luminosity Dimuon stream (pb $^{-1}$)	Integrated luminosity Leptonic (pb $^{-1}$)	Reco version	Stripping version
2011	7	962.7	978.62	Reco14	21r1
2012	8	1973.2	1984.64	Reco14	21
2015	13	290.2	not used	Reco15a	24
2016	13	~ 1100	not used	Reco15a	26

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

304 **3.2 Monte Carlo samples**

305 A number of different samples of simulated events are used in the analysis of $B_s^0 \rightarrow \mu^+ \mu^-$
306 decays. Most samples have been regenerated since the previous 3 fb^{-1} analysis using
307 more recent versions of PYTHIA and the LHCb simulation software. The most significant
308 exception to this is the large 100 million event sample of $b\bar{b} \rightarrow X\mu^+\mu^-$ events that were
309 generated under 2010 conditions. It has not been possible to generate a similar sample
310 with up-to-date software and conditions due to the prohibitive cost in computer resources.

311 Details of all the samples used in the analysis are listed in Table 2.

Decay	Year	Sim	Reco	TCK	Stripping	Events generated
$B_s^0 \rightarrow \mu^+ \mu^-$	2011	08a	14a	0x40760037	20r1	0.5 M
	2012	06b	14	0x40990042	20	2 M
	2015	09a	15a	0x411400a2	24	2 M
$B_s^0 \rightarrow K^+ K^-$	2012	08g	14a	0x409f0045	20	7 M
$B_s^0 \rightarrow \pi^+ K^-$	2012	08g	14a	0x409f0045	20	7 M
$B^0 \rightarrow K^+ \pi^-$	2011	08b	14a	0x40760037	20r1	0.8 M
	2012	08g	14a	0x409f0045	20	8.5 M
	2015	09a	15a	0x411400a2	24	4 M
$B^0 \rightarrow \pi^+ \pi^-$	2012	08g	14a	0x409f0045	20	7 M
$B^0 \rightarrow \pi^- \mu^+ \nu_\mu$ $M_{\mu^+ \mu^-} > 4.5 \text{ GeV}$	2012	08e	14a	0x409f0045	20	6 M
$B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ $M_{\mu^+ \mu^-} > 4.5 \text{ GeV}$	2012	08e	14a	0x409f0045	20	6 M
$A_b^0 \rightarrow p \mu^- \bar{\nu}_\mu$ $M_{\mu^+ \mu^-} > 4.5 \text{ GeV}$	2012	08e	14a	0x409f0045	20	2 M
$B^0 \rightarrow \pi^0 \mu^+ \mu^-$ $\mu^+ \mu^-$ in acceptance	2012	08e	14a	0x409f0045	20	2 M
$B^+ \rightarrow \pi^+ \mu^+ \mu^-$ $\mu^+ \mu^-$ in acceptance	2012	08e	14a	0x409f0045	20	2 M
$b\bar{b} \rightarrow X \mu^+ \mu^-$ $p > 3 \text{ GeV}, M_{\mu^+ \mu^-} > 4.7 \text{ GeV}$	2010	01	08	0x002e002a	12	100.0 M
$b\bar{b} \rightarrow X \mu^+ \mu^-$ $p > 3 \text{ GeV}, 4.7 < M_{\mu^+ \mu^-} < 6.0 \text{ GeV}$ $\text{doca} < 0.4\text{mm}, 1 < \text{PtProd} < 16 \text{ GeV}^2$	2012	06a	13a	0x4097003d	17filt	2.6 M
$b\bar{b} \rightarrow X \mu^+ \mu^-$ $p > 3 \text{ GeV}, 4.7 < M_{\mu^+ \mu^-} < 6.0 \text{ GeV}$ $\text{doca} < 0.4\text{mm}, 1 < \text{PtProd} < 16 \text{ GeV}^2$	2012	06b	14	0x4097003d	20filt	4.0 M
$b\bar{b} \rightarrow X \mu^+ \mu^-$ $p > 3 \text{ GeV}, 4.7 < M_{\mu^+ \mu^-} < 6.0 \text{ GeV}$ $\text{doca} < 0.4\text{mm}, \text{PtProd} > 16 \text{ GeV}^2$	2012	06b	14	0x4097003d	20filt	3.3 M

Table 2: Monte Carlo samples used in the analysis of $B_s^0 \rightarrow \mu^+ \mu^-$ decays. Requirements imposed at the generator stage are listed beneath the decay descriptor. All candidates are also required to have their daughters in the LHCb angular acceptance.

312 3.3 Trigger

313 No specific trigger requirements are imposed on the $B_s^0 \rightarrow \mu^+ \mu^-$ data in order to maximise
 314 the available statistics. The most efficient lines for candidates passing the $B_s^0 \rightarrow \mu^+ \mu^-$ and
 315 $B \rightarrow h^+ h^-$ stripping lines are listed below in Tables 3 and 4.

Trigger line	2011	2012	2015	2016
L0Muon	90%	87%	58%	69.3%
L0DiMuon	74%	70%	62%	66.1%
Hlt1TrackAllL0	61%	62%	-	-
Hlt1TrackMuon	76%	84%	90%	93.5%
Hlt1DiMuonLowMass	76%	69%	76%	79.5%
Hlt1DiMuonHighMass	80%	76%	73%	79.3%
Hlt1TrackMVA	-	-	97.3 %	94.2 %
Hlt2DiMuonB	94%	95%	100%	100%

Table 3: The most efficient trigger lines for $B_s^0 \rightarrow \mu^+ \mu^-$ candidates in 2011 , 2012, 2015 and 2016 data. Values are fractions of triggered-on-signal (TOS) $B_s^0 \rightarrow \mu^+ \mu^-$ events with respect to $B_s^0 \rightarrow \mu^+ \mu^-$ events passing the stripping requirements (Stripping 21r1, Stripping 21 and Stripping 24 respectively). Note that not all trigger lines are present in each year, in which case the fraction is marked with a dash.

Trigger line	2011	2012	2015	2016
L0Hadron	39%	37%	39%	36.4 %
L0Global TIS	65%	67%	69%	64.2%
Hlt1TrackAllL0	78%	76%	-	-
Hlt1TrackMVA	-	-	96.1%	89.2 %
Hlt2Topo2BodyBBDT	34%	35%	-	-
Hlt2B2HH	34%	29%	60%	77.5 %

Table 4: The most efficient trigger lines for $B \rightarrow h^+ h^-$ candidates in 2011, 2012, 2015 and 2016 data. Values are fractions of triggered-on-signal (unless otherwise indicated) $B \rightarrow h^+ h^-$ events with respect to events passing the $B \rightarrow h^+ h^-$ stripping requirements (Stripping 21r1, Stripping 21 and Stripping 24 respectively). Note that not all trigger lines are present in each year, in which case the fraction is marked with a dash.

316 **4 Selection**

317 **4.1 Stripping and pre-selection**

318 The data samples used in the analysis were produced using Stripping 21r1, 21 and 24 for
319 2011, 2012 and 2015 respectively. The requirements imposed in all three versions of the
320 stripping are almost identical and all versions differ from the previously used version of
321 the stripping (20(r1)) in that softer distance of flight χ^2 and muon IP χ^2 requirements are
322 imposed, see [9] for full study. Also for 2015 data taking the track χ^2/ndf and ghost
323 probability requirements are softer to take advantage of changes in the reconstruction
324 for Run II. Separate lines are used to select $B_s^0 \rightarrow \mu^+\mu^-$, $B_{(s)}^0 \rightarrow h^+h^-$ and $B^+ \rightarrow J/\psi K^+$
325 where the requirements are kept as similar as possible, for $B_{(s)}^0 \rightarrow h^+h^-$ to ensure selection
326 is similar to $B_s^0 \rightarrow \mu^+\mu^-$, the hadrons are required to be within the muon acceptance.
327 Additional requirements are then imposed on top of the requirements in the stripping lines
328 to further reduce the size of the samples. The full stripping requirements are detailed in
329 Table 5.

330 **4.2 The BDTS discriminant**

331 A cut on the response of a multivariate classifier ('BDTS') is used to further reduce the
332 size of the background sample before the final BDT classifier is trained. This discriminant
333 is highly efficient on signal. The variables entering the BDTS are:

- 334 - the impact parameter (IP(B)) and impact parameter χ^2 (IP χ^2 (B)) of the B candidate;
- 335 - the χ^2 of the secondary vertex (VCHI2);
- 336 - the angle between the direction of the momentum of the B candidate and the
337 direction defined by the secondary and the primary vertices (DIRA);
- 338 - the minimum distance between the two daughter tracks (DOCA);
- 339 - the minimum impact parameter each muon with respect to any primary vertex
340 (minIP(μ)).

341 In order to minimise the systematic uncertainty in the normalization factors, the same
342 BDTS cut is applied to the three normalization channels; $B^+ \rightarrow J/\psi K^+$, $B_s^0 \rightarrow J/\psi \phi$ and
343 $B^0 \rightarrow K^+\pi^-$. For the $B^+ \rightarrow J/\psi K^+$ mode, the χ^2 of the secondary vertex is substituted
344 by the χ^2 of the J/ψ vertex, the flight distance is computed between the J/ψ vertex and
345 the primary vertex and the DOCA is computed between the two muons from the J/ψ
346 decay. In this way, the distributions of all the variables but minIP and DOCA, are very
347 similar for $B_s^0 \rightarrow \mu^+\mu^-$, $B \rightarrow h^+h^-$ and $B^+ \rightarrow J/\psi K^+$, resulting in a similar efficiency for
348 signal and normalization channels. The ratio of the efficiencies for a range of BDTS cut
349 values of $B_s^0 \rightarrow \mu^+\mu^-$ and $B^+ \rightarrow J/\psi K^+$ unsmeared, smeared and oversmeared simulated
350 samples agree within 0.4%.

Cut	applied on	value	applied on	value
		$B_s^0 \rightarrow \mu^+ \mu^-$ and $B_{(s)}^0 \rightarrow h^+ h^-$		$B^+ \rightarrow J/\psi K^+$
track χ^2/ndf ghost prob DOCA IP χ^2 p_T p ISMUON	μ / h 	<3 (<4) < 0.3 (< 0.4) <0.3 mm >25 > 0.25 and < 40 GeV/c <500 GeV/c true	μ / h 	< 3
vertex χ^2 VDS ΔM	$B_{(s)}^0$	<9 > 15 $ M(hh, \mu\mu) - m_B < 60 \text{ MeV}/c^2$	J/ψ	<9 > 15 $ M(\mu\mu) - m_{J/\psi} < 60 \text{ MeV}/c^2$
IP χ^2 t BDTS ΔM $p_T (B_s^0)$	$B_{(s)}^0$	< 25 $< 9 \cdot \tau(B_s^0)$ > 0.05 $> 0.5 \text{ GeV}/c$	B^+	< 25 $< 9 \cdot \tau(B_s^0)$ > 0.05 $ M(J/\psi K) - m_B < 100 \text{ MeV}/c^2$

Table 5: Selection for $B_{(s)}^0 \rightarrow \mu^+ \mu^-$, $B_{(s)}^0 \rightarrow h^+ h^-$ and $B^+ \rightarrow J/\psi K^+$ channels; DOCA is the distance of closest approach between the two tracks, VDS is the secondary vertex flight distance significance, and DLL is the combined PID likelihood to discriminate different particle hypotheses. Cut values in parenthesis for track χ^2/ndf and ghost probability show softer cuts used for 2015 data taking. The BDTS discriminant is described in Section 4.2.

351 The BDTS cut has not been optimized since Ref. [10] where it was chosen by looking
 352 at the combined performance of the BDTS and the BDT; a BDTS cut $BDTS > 0.05$ was
 353 found to be optimal. This cut choice has an efficiency of 92% on smeared $B_s^0 \rightarrow \mu^+ \mu^-$
 354 MC .

355 **4.3 Particle Identification**

356 The muon identification procedure (muonID) is a key ingredient of the analysis. It consists
 357 of two steps: the requirement for the muon candidate to satisfy the ISMUON algorithm [11],
 358 and a more restrictive PID selection based on a combination of RICH, CALO and MUON
 359 information. In the present version of the analysis, we switched from *combDLL*, used in
 360 Ref. [6], to the NNpid variables [12]. The guide principle of the optimization done in [13]
 361 has been the strong rejection of backgrounds while preserving high enough signal efficiency.
 362 The best performances for the BF analysis have been obtained with the requirement

$$\text{PID}_\mu \equiv \text{ProbNN}_\mu \times (1 - \text{ProbNN}_p) \times (1 - \text{ProbNN}_K) > 0.4, \quad (5)$$

363 referred to as $\text{PID}_{\mu,4}$ in the following. In this combination of variables, the ProbNN_K is
 364 very effective to reject $B_{(s)}^0 \rightarrow h^+ h^{(\prime)-}$, while ProbNN_p has been introduced specifically to
 365 reject $\Lambda_b^0 \rightarrow p \mu^- \nu$. For the effective lifetime analysis, the above condition has been relaxed
 366 into $\text{PID}_\mu > 0.2$ (see Sec. 10.1.1), the reason being that in this case only the B_s^0 signal
 367 region is relevant, which is less polluted by the $B_{(s)}^0 \rightarrow h^+ h^{(\prime)-}$ events.

368 PID performance for all species has been evaluated from data control samples. The
 369 whole data set has been used for Run 1 and 2015/Run 2, while only the first 600 pb^{-1} of
 370 2016 have been produced by the PID group. The possible systematic effect of using the
 371 first part of 2016 data set also for the second 500 pb^{-1} will be evaluated using suitable
 372 subsamples of data.

373 **4.3.1 Muon Identification**

374 The efficiency for muons to pass the above selection is evaluated using the PIDCalib
 375 package [14], which is based on data control samples of $J/\psi \rightarrow \mu^+ \mu^-$ from b decays. To
 376 ensure that no bias arises from their selection the muons from this sample are selected as
 377 probes by a *tag-and-probe* method. Trigger unbias is also guaranteed, by requiring the
 378 muon *probe* candidates to satisfy the condition

$$\text{L0Global_TIS} \&& \text{Hlt1Phys_TIS} \&& \text{Hlt2Phys_TIS}. \quad (6)$$

379 The efficiencies are shown in Fig. 2 for the calibration samples collected in 2011, 2012, and
 380 2015.

381 The kinematic distribution of the probe muon in the control samples is different from
 382 the one belonging to $B_{d,s}^0 \rightarrow \mu^+ \mu^-$. To accommodate this difference, the single muon
 383 efficiency is evaluated from PIDCalib in 11 bins in momentum¹, p , and 4 bins in transverse
 384 momentum², p_T , and then convoluted with the proper signal MC spectrum. For each
 385 MC signal $B_{(s)} \rightarrow \mu^+ \mu^-$ event the efficiency from PIDCalib is applied to both μ^+ and μ^-
 386 according to their p, p_T , so that the output of such convolution gives the overall signal
 387 efficiency under the muonID selection. The efficiency is computed given the geometrical

¹ $p = (0,5), (5,10), (10,15), (15,20), (20,25), (25,30), (30,35), (35,40), (40,50), (50,60)$, and $>60 \text{ GeV}$
² $p_T = (0.8,1.7), (1.7,3), (3,5)$, and $(5,10) \text{ GeV}$

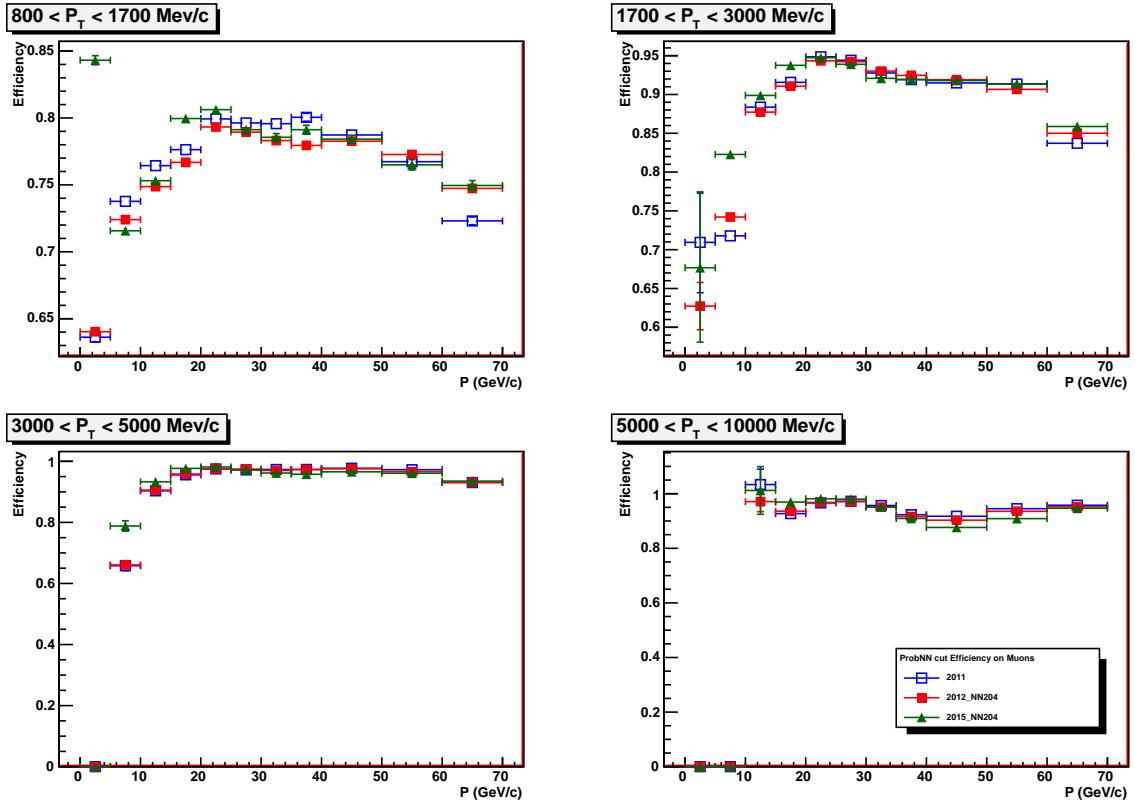


Figure 2: MuonID vs momentum for the four p_T bins of the probe track, evaluated from *PIDCalib* for 2011, 2012, and 2015 calibration data. (MagUp and MagDown performance tables have been averaged.)

acceptance of the candidate tracks in the muon detector. The results for Run I and Run II are listed in Tab. 6, for the whole BDT range and for each BDT bin separately. A very good stability of the efficiencies is observed between Run I and Run II data; in all cases, a $\sim 10\%$ slope is observed as a function of BDT. Notice that the PID selection applied to 2016 data is different from the one applied to Run I and 2015 data. See later, Sect. 4.3.3, for further details.

The muonID efficiencies extracted from data will be used in Sec. 7 to evaluate data/MC correction factors when computing the ratio of efficiencies between the signal and normalization channels. A 2% systematic uncertainty is assigned to the above ratio of efficiencies. This has been obtained in the past rounds of the analysis varying the variables used to parametrise the muonID efficiency, the binning, the sample used for the muonID evaluation, the trigger unbias requests (TIS-probe against TOS-tag). This is still a valid, and even “conservative”, evaluation of the systematic error coming from the muonID. It is “conservative” in the sense it includes also the trigger unbias part which should be negligible in Run 2 given the changing in the selection strategy of PID samples (from Stripping in

Table 6: ISMUON \times PID $_{\mu,4}$ efficiency on $B_{d,s}^0 \rightarrow \mu^+\mu^-$ for Run I and 2015 and ISMUON \times PID $_{\mu,8}$ for 2016 data, as evaluated convoluting the single muon efficiencies from PIDCalib with the MC signal spectra: results are given in the whole BDT range, and for each BDT bin separately. (Notice that the PID selection applied to 2016 data is different from the one applied to Run I and 2015 data. See Sect. 4.3.3.)

BDT range	<i>RunI</i>	2015	2016
0-1.0	0.8231 ± 0.0003	0.8347 ± 0.0004	0.8788 ± 0.0002
0-0.25	0.7971 ± 0.0003	0.8088 ± 0.0004	0.8537 ± 0.0002
0.25-0.4	0.8091 ± 0.0003	0.8205 ± 0.0004	0.8642 ± 0.0002
0.4-0.5	0.8198 ± 0.0003	0.8310 ± 0.0004	0.8750 ± 0.0002
0.5-0.6	0.8270 ± 0.0003	0.8367 ± 0.0004	0.8803 ± 0.0002
0.6-0.7	0.8353 ± 0.0003	0.8438 ± 0.0004	0.8871 ± 0.0002
0.7-0.8	0.8467 ± 0.0003	0.8524 ± 0.0005	0.8962 ± 0.0002
0.8-0.9	0.8599 ± 0.0004	0.8631 ± 0.0005	0.9071 ± 0.0003
0.9-1.0	0.8764 ± 0.0004	0.8725 ± 0.0005	0.9173 ± 0.0003

403 Run 1 to HLT2 in Run2).

404 4.3.2 Pion, kaon and proton misID

405 The evaluation of $\pi, K, p \rightarrow \mu$ probabilities under the PID selection of eq. 5, misID in
 406 the following, is crucial to the determination of the $B_{(s)}^0 \rightarrow h^+h^{(\prime)-}$ and the other peaking
 407 backgrounds. The misID is evaluated from data using the PID calibration lines: $D^0 \rightarrow K\pi$
 408 from the $D^* \rightarrow D^0\pi$ decays are used for pions and kaons (`NoPIDDstarWithD02RSKPi`
 409 stripping line). $\Lambda \rightarrow p\pi$ decays (`Lam0LLLine1V0ForPID` stripping line) as well as protons
 410 from Λ_c decays, have been used to evaluate the proton misID. (For details on the selection
 411 of the PID calibration samples in Run II see Sect. 4.3.3.)

412 For pions and kaons, the extraction of the misID from data relies on the *PIDCalib*
 413 package [14], and it is performed using the same $p - p_T$ binning as introduced for muons.
 414 Further splitting of the phase space using N_{track} bins has been also investigated, and it
 415 is included in the systematic error. Different from muons, the trigger unbias condition
 416 of eq. 6 is implemented in this case without including a requirement on the Hlt2, which
 417 has been proven not to bias the kaon and pion PID probabilities [13]. The pion and kaon
 418 misID in the binned phase space are shown in fig. 3 and fig. 4, respectively.

419 For protons, the trigger unbias condition has to be implemented with inclusion of Hlt2
 420 requirement [13], and this reduces the available statistics. In this condition, for Run I
 421 data more stable results on the proton misID are obtained from a direct fit of the Λ mass
 422 (with and w/o PID selection) instead of a use of the sWeights from *PIDCalib*. For Run II
 423 data, the new PID selection implemented in HLT [15] provides enough statistics also for
 424 protons and the sWeighted samples from *PIDCalib* have been used. The proton misID in
 425 the binned phase space is shown in fig. 5, for Run I and 2015 data.

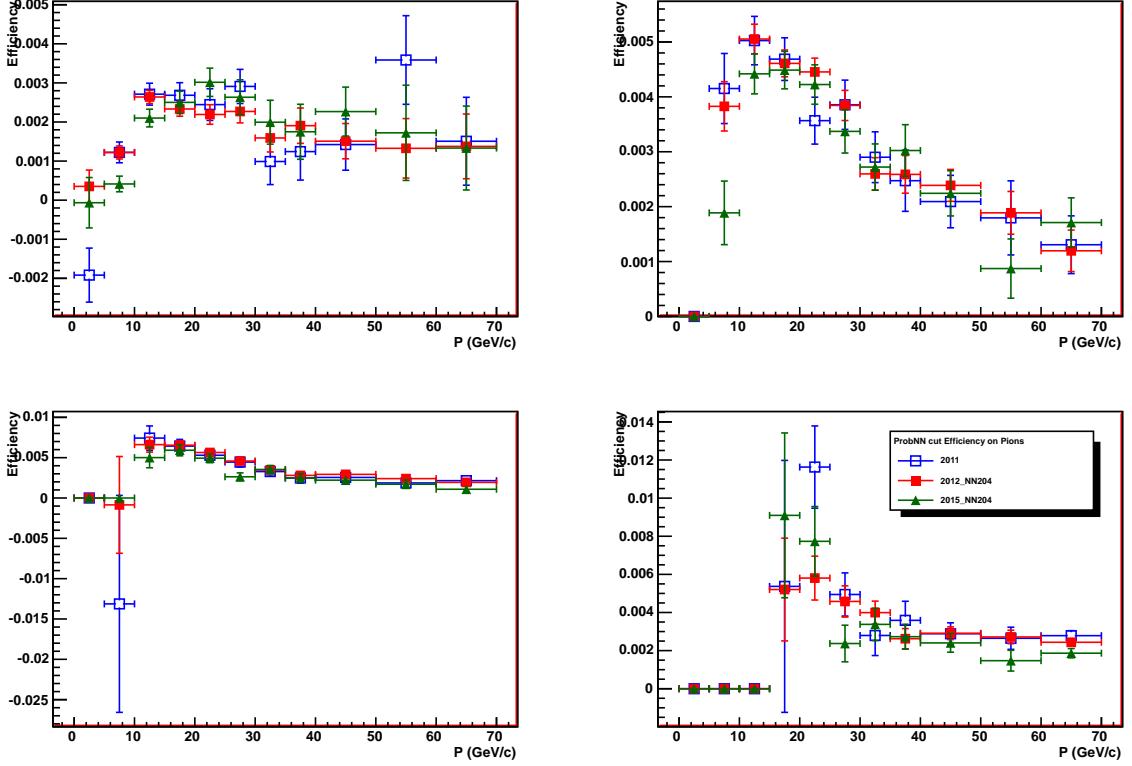


Figure 3: Pion misID vs momentum for the four p_T bins of the probe track, evaluated from *PIDCalib* for 2011, 2012, and 2015 calibration data. (MagUp and MagDown performance tables have been averaged.) (Panels refer to p_T bins (0.8,1.7), (1.7,3), (5,10), and (3,5) GeV clockwise from top left.)

426 Systematic studies have been carried on in order to assess the effect of the multiplicity in
 427 the evaluation of the misID. In particular, for a subsample of Run I data, the pion and kaon
 428 misID have been recomputed by further splitting the momentum and transverse momentum
 429 phase space into four bins of ntrack ($0, 150, 250, inf$), and the misID probabilities thus
 430 obtained have been used to recompute the background estimates. A comparison with the
 431 values obtained with the standard 2D misID evaluation is used to asses the systematic
 432 error. For protons coming from the Λ calibration line, the results have been also computed
 433 as a function of the proton production point: a negligible effect has been found in this
 434 case [6].

435 4.3.3 PID selection strategy for 2016 data

436 In Run I, PID samples were selected during the Stripping, while during LS1 the selection
 437 has been moved in the HLT [15]. The HLT2 lines used are PIDDO2KPiTag line for K and

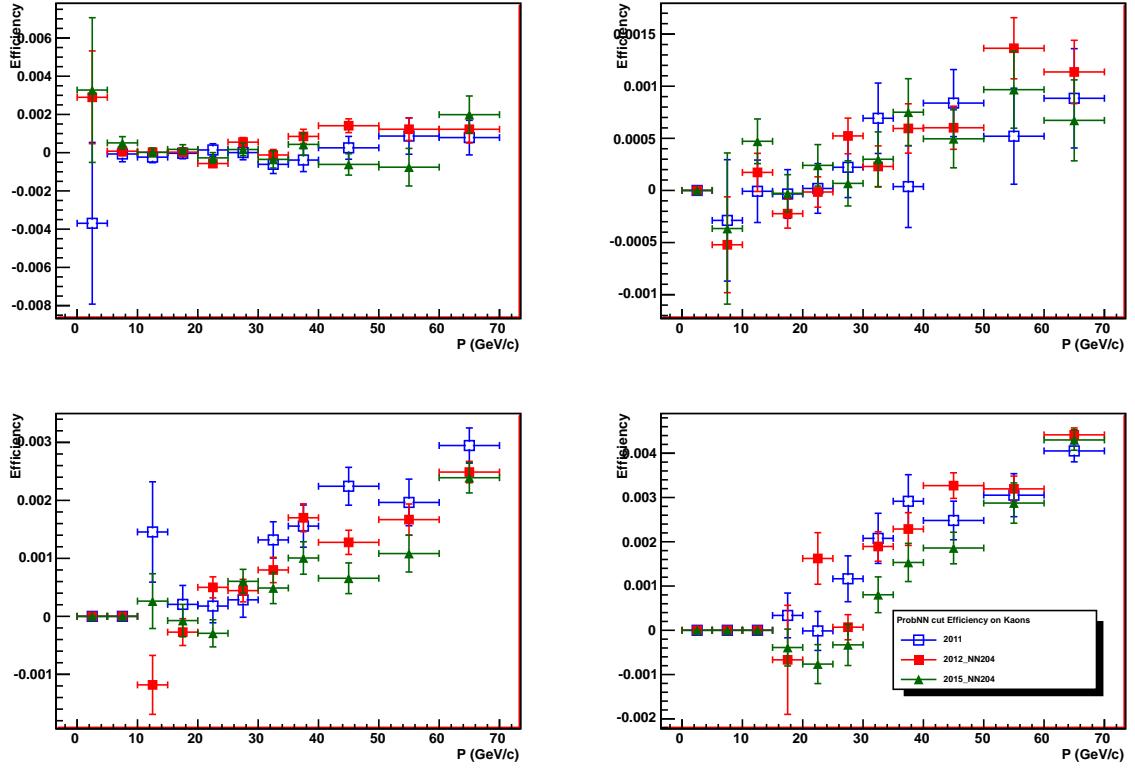


Figure 4: Kaon misID vs momentum for the four p_T bins of the probe track, evaluated from *PIDCalib* for 2012, 2015, and 2016 calibration data. (MagUp and MagDown performance tables have been averaged.) (Panels refer to p_T bins (0.8,1.7), (1.7,3), (5,10), and (3,5) GeV clockwise from top left.)

438 π , PIDLambda2PPiLL, PIDLambda2PPiLLhighPT, and PIDLambda2PPiLLveryhighPT lines
 439 for p , and PIDDetJPsiMuMuNegTagged and PIDDetJPsiMuMuPosTagged for μ . Besides
 440 several technical advantages, this change includes the larger accessible statistics, the better
 441 kinematical coverage, and the chance to better take under control the decorrelation of the
 442 PID sample with respect to the trigger (the so called unbias condition).

443 Another important change during LS1, has been the removal of the Aerogel radiator in
 444 RICH1. This forced to change the list of the input variables used by the *ANN* algorithm,
 445 which produces the *ProbNN* variables. Beyond this needed change, the *ANN* algorithm
 446 has been optimized exploiting the study done with Run 1 and MC 2015 data. The new
 447 algorithm (avaialble as MC15TuneV1 for Run II data, and as MC12TuneV4 for Run I
 448 data) performs better than the old one. This implies that the cut defined in Eq. 5 is
 449 sub-optimal once the new algorithm is used, as can be seen from the magenta open circles
 450 in Figs. 6, 7, and 8 showing the mis-identification of pions, kaons and protons, respectively.
 451 For convenience, the different tunings used for this analysis are listed here.

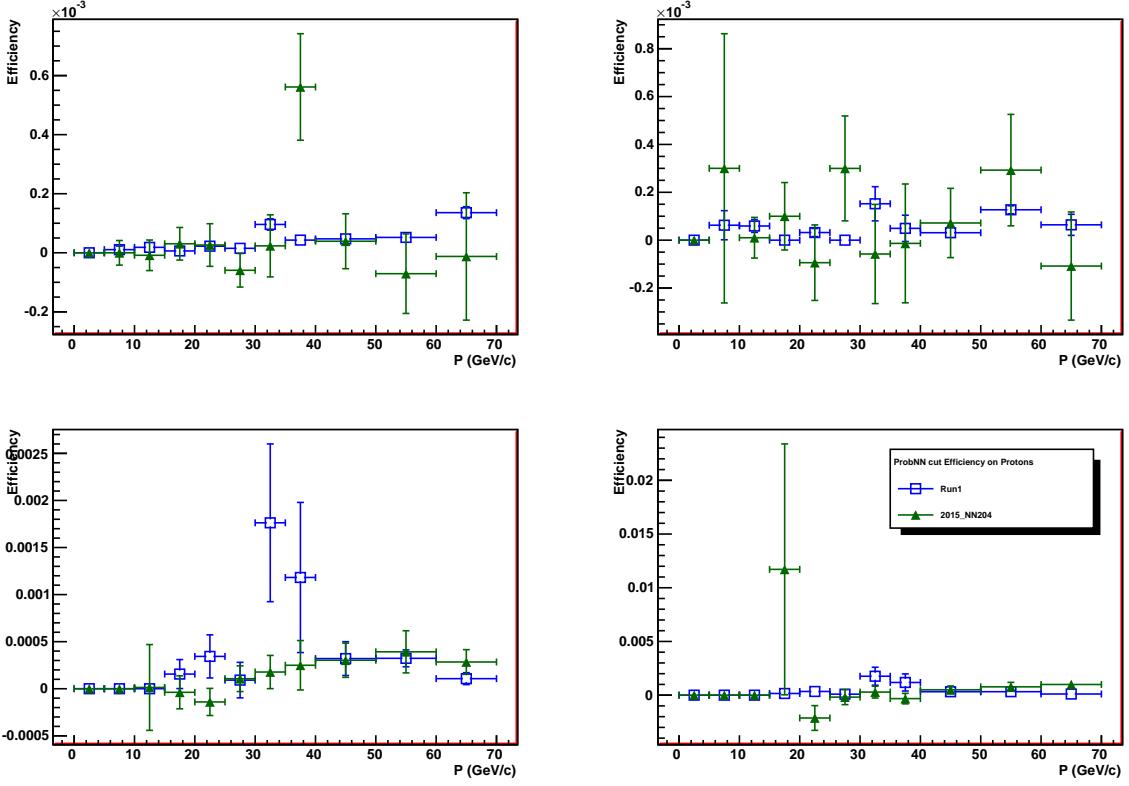


Figure 5: Proton misID vs momentum for the four p_T bins of the probe track, evaluated from *PIDCalib* for Run I and 2015 calibration data. (Run I curves are from a direct fit of the Λ mass (with and w/o PID selection).) (Panels refer to p_T bins (0.8,1.7), (1.7,3), (5,10), and (3,5) GeV clockwise from top left.)

- 452 • For Run I (2011 and 2012) data, the MC12TuneV2 (ProbNNv2) tune has been
453 used. (Notice that as far as the electron ID or misID is not relevant, the other tune
454 available for Run I data, MC12TuneV3, is completely equivalent.)
- 455 • For 2015 data, the MC12TuneV2 tune has been used. This is due to the choice of
456 producing the nTuples using the same Da Vinci version, v38rX, used during the
457 Stripping. (MC 2015 wasn't available at the time of 2015 Stripping campain.)
- 458 • For 2016 data, the MC15TuneV1 tune has been used. This tune, produced using the
459 MC 2105, is available since Da Vinci version v40. Due to the change of the input
460 used by the ANN algorithms, the older MC12TuneV2 and MC12TuneV3 tunings are
461 not well defined once calculated in Da Vinci versions later than v40.

462 Given the choice of tunings described above, the cut defined in Eq. 5 can be still used for
463 2015 data, while a new working point has been studied for 2016 data for which only the

464 latest MC15TuneV1 is available. The requirement

$$\text{PID}_\mu \equiv \text{ProbNN}_\mu \times (1 - \text{ProbNN}_p) \times (1 - \text{ProbNN}_K) > 0.8 \quad (7)$$

465 applied on the MC15TuneV1 version of the ProbNN variables provides in Run II performance
 466 similar to the one provided by Eq. 5 in Run I data with MC12TuneV2 version. In particular
 467 the cut in Eq. 7 has been chosen to have same or lower $\pi, K, p \rightarrow \mu$ probabilities. This is
 468 evident from Figs. 6, 7, and 8 for the pion, kaon and proton misID, respectively, where the
 469 different performance curves are superimposed: $\text{PID}_\mu(\text{MC12TuneV2}) > 0.4$ applied on 2012
 470 samples (in red closed squares), $\text{PID}_\mu(\text{MC15TuneV1}) > 0.4$ applied on 2016 samples (in
 471 magenta open circles), and $\text{PID}_\mu(\text{MC15TuneV1}) > 0.8$ applied on 2016 samples (in purple
 472 closed circles). As a result, with the same cut as defined in Eq. 7, the muon efficiency for
 473 2016 calibration data is larger by $\sim 10\%$ for p_T below 3 GeV/c (see Fig. 9).

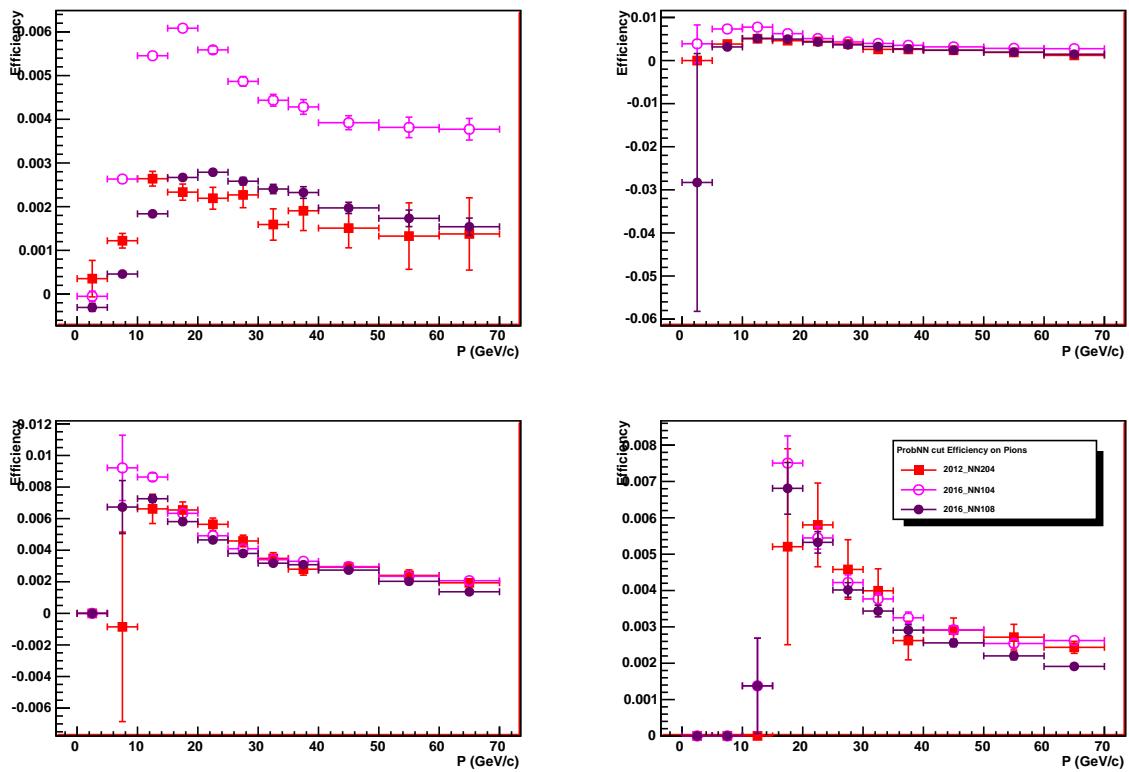


Figure 6: Pion misID vs momentum for the four p_T bins of the probe track, evaluated from $PIDCalib$ (see text for details). (Panels refer to p_T bins (0.8,1.7), (1.7,3), (5,10), and (3,5) GeV clockwise from top left.)

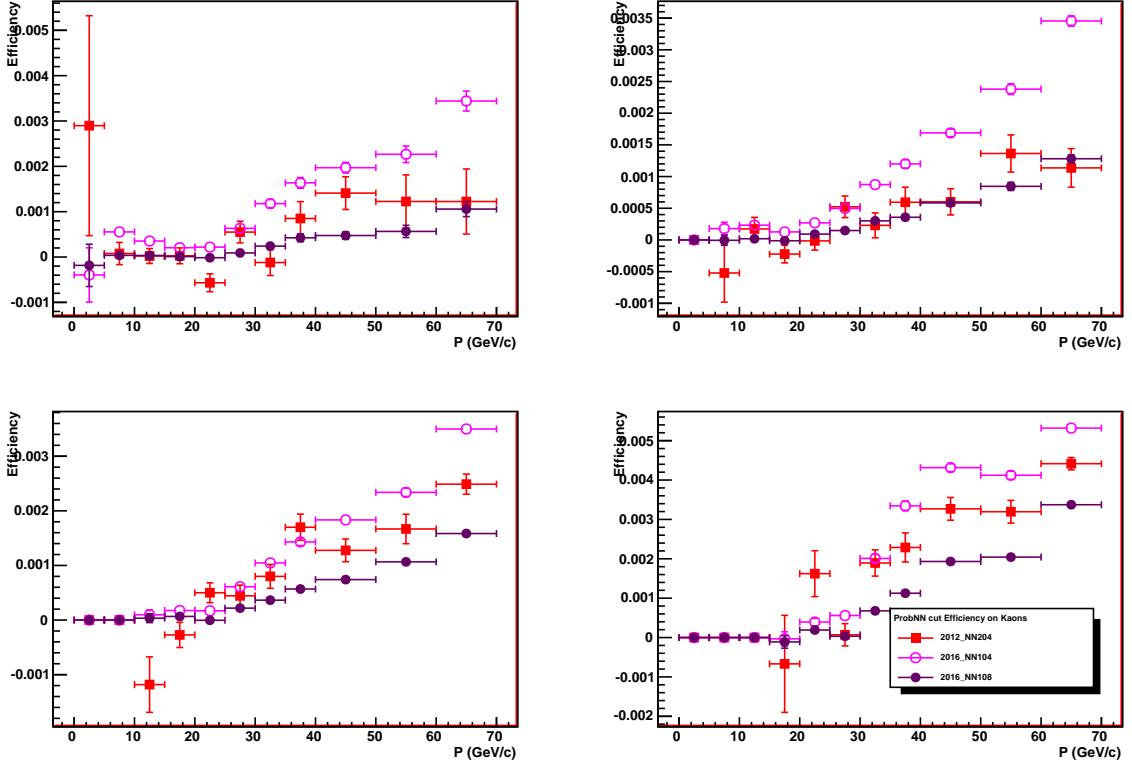


Figure 7: Kaon misID vs momentum for the four p_T bins of the probe track, evaluated from *PIDCalib* (see text for details). (Panels refer to p_T bins (0.8,1.7), (1.7,3), (5,10), and (3,5) GeV clockwise from top left.)

474 4.3.4 About the use of MC12TuneV2 ProbNN version for 2015 data

475 The ProbNN tunings available for 2015 depends on the Da Vinci version. Before v40 only
 476 the old tuning (the one with the aeogel info) MC12TuneV2 is available. From v40 on the
 477 new/more correct one, MC15TuneV1, becomes the default. The main difference between
 478 MC12TuneV2 and MC15TuneV1 is that the information from the aerogel is not available
 479 in Run 2 (the aerogel has been removed during LS1) and that the list input variables has
 480 been modified between Run 1 and Run 2 to allow to have the same version of the ProbNN
 481 variables in HLT2 (with some timing constraint) and in Brunel. These differences can
 482 produce “weird” behaviors like the one visible in a few p_T bins in the lower p_T bin of Fig. 2
 483 (top left panel). A “normal” behavior is recovered when the right tuning is used, see
 484 Fig. 10.

485 2015 data we are using were stripped using DaVinci v38r1p1. For consistency we are
 486 producing the nTuples using the same DV version and this implies that we can only access
 487 the MC12TuneV2 ProbNN version. Consistently we evaluate the PID performance on the

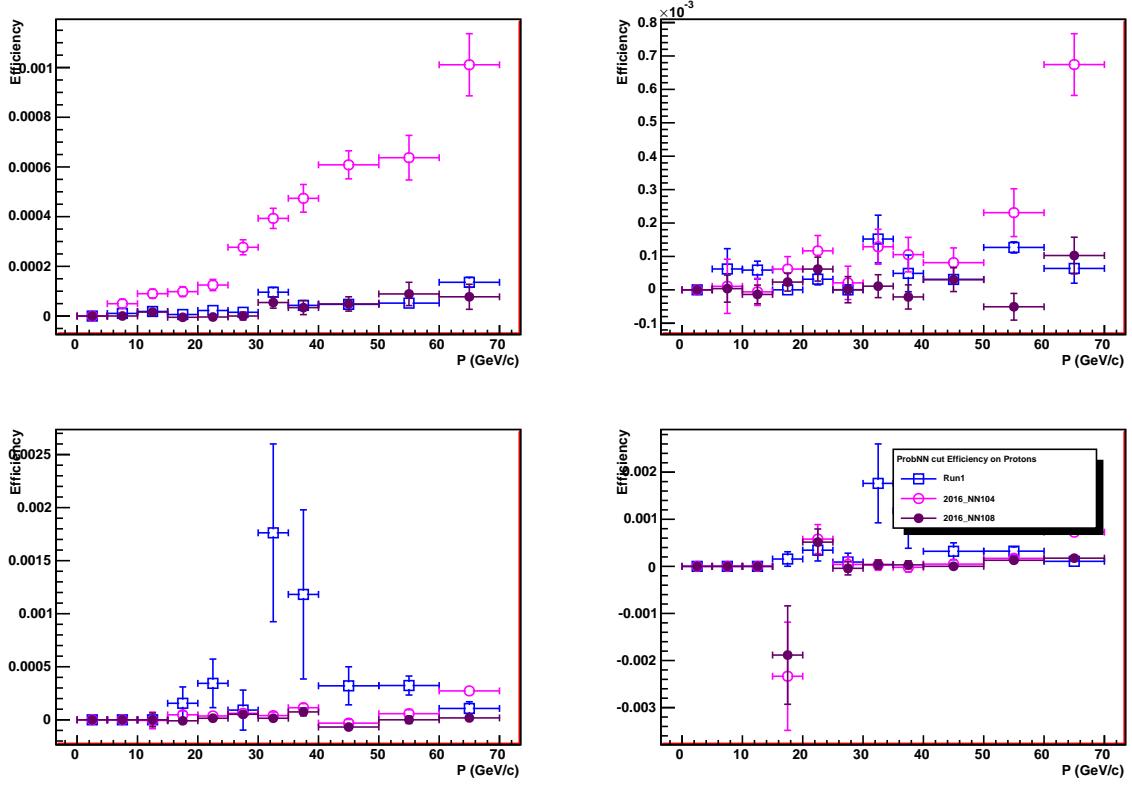


Figure 8: Proton misID vs momentum for the four p_T bins of the probe track, evaluated from *PIDCalib* for 2016 and from a direct fit of the Λ mass for Run I (see text for details). (Panels refer to p_T bins (0.8,1.7), (1.7,3), (5,10), and (3,5) GeV clockwise from top left.)

488 old PIDCalib samples ³.

³see here https://twiki.cern.ch/twiki/bin/view/LHCb/PIDCalibPackage#PIDCalib_sample_for_2015_and_201 for details

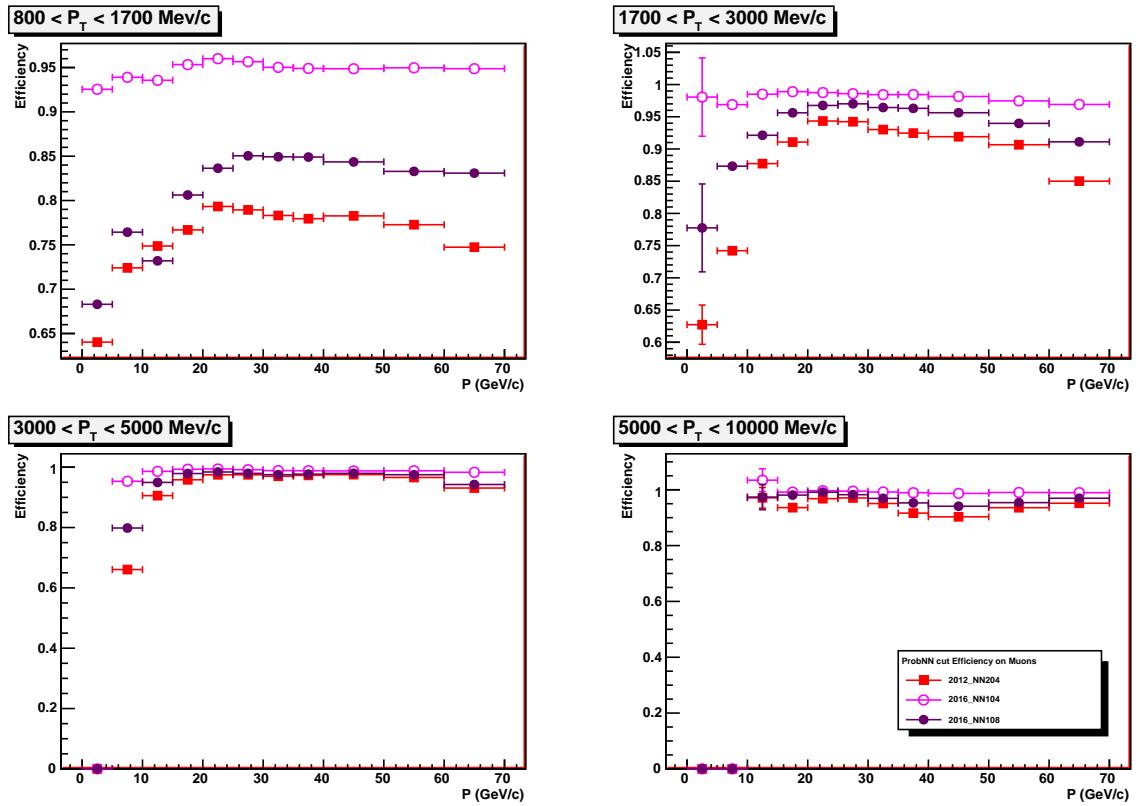


Figure 9: Muon efficiency vs momentum for the four p_T bins of the probe track, evaluated from *PIDCalib* (see text for details).

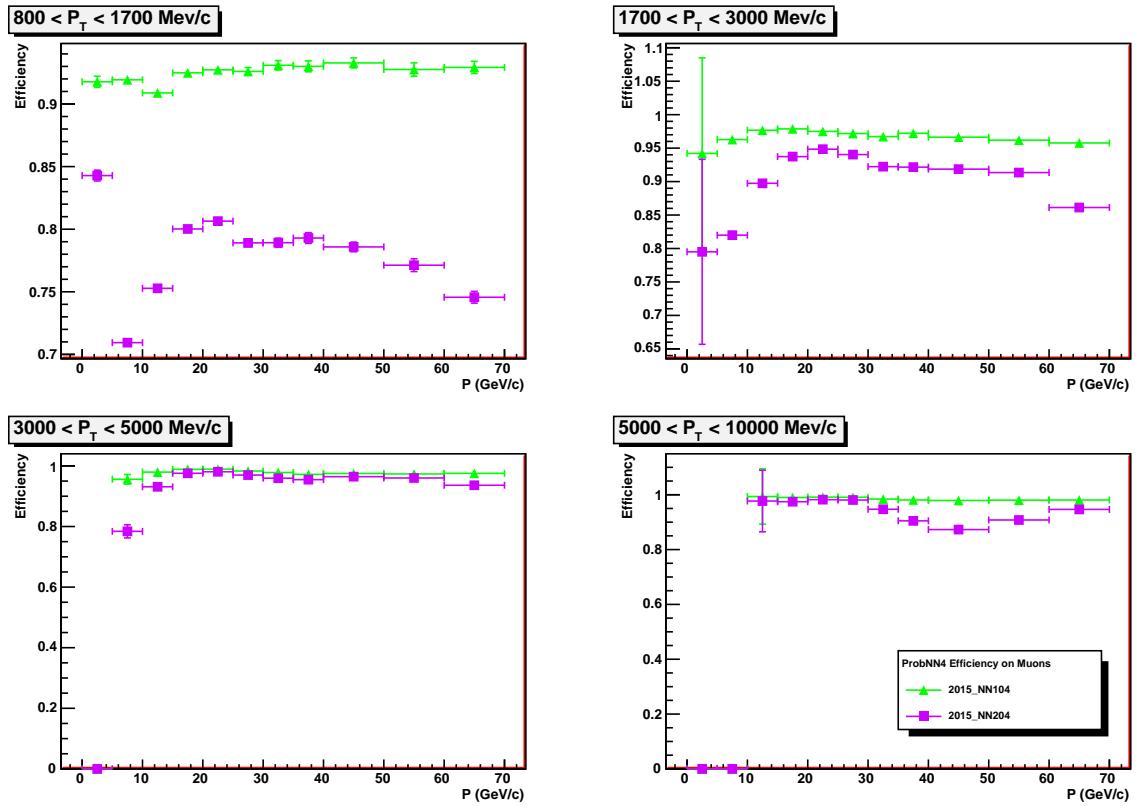


Figure 10: MuonID efficiency for the $PID_\mu > 0.4$ applied to 2015 calibration data using MC12TuneV2 (purple squares) or MC15TuneV1 (green triangles).

489 5 Peaking backgrounds

490 In this section the evaluation of the peaking backgrounds is described. As discussed in
 491 sec. 2, there are several different decays contributing to this category:

- 492 • $B_{(s)}^0 \rightarrow h^+h'^{-}$ ($h, h' = K, \pi$), with both kaon and pion misidentified as muons,
 493 which pollute mainly the B^0 signal region;
- 494 • $B^0 \rightarrow \pi^-\mu^+\nu_\mu$ and $B_s^0 \rightarrow K^-\mu^+\nu_\mu$, with the pion or kaon misidentified as a muon,
 495 which pollute the left mass sideband only;
- 496 • $\Lambda_b^0 \rightarrow p\mu^-\nu$, with the proton misidentified as a muon, which pollutes both the left
 497 mass sideband and B^0 - B_s^0 mass regions;
- 498 • $B^{0(+)} \rightarrow \pi^{0(+)}\mu^+\mu^-$ and $B_c^+ \rightarrow J/\psi\mu^+\nu$, both with a couple of true muons selected
 499 in the final state, which pollute the left sideband only and the full mass region,
 500 respectively.

501 Since the signal fit is made in BDT bins, all of the above background sources have to be
 502 estimated in each BDT bin (*i.e.* the BDT *pdf* has to be determined for each background
 503 source).

504 The section outline is the following: in sec. 5.1 the $B_{(s)}^0 \rightarrow h^+h'^{-}$ contamination is
 505 estimated, and in sec. 5.2 all of the other channels are evaluated.

506 5.1 $B \rightarrow h^+h^-$

507 The double misID probability is estimated by convoluting the kaon and pion misID
 508 curves given above with the momentum and p_T spectra of the two hadrons from MC
 509 $B_{(s)}^0 \rightarrow h^+h'^{-}$ decays, selected as described in Section 4. Such a convolution is performed
 510 with a toy technique that takes into account the misID uncertainties as evaluated from
 511 calibration data.

512 The working point for the PID selection, the $\text{PID}_{\mu,4}$ cut of eq. 5, reduces the $B_{(s)}^0 \rightarrow$
 513 $h^+h'^{-}$ yield by almost a factor of 5 (on the whole BDT range) wrt previous *combDLL*
 514 selection [6], with a signal loss of about 11 %; the increase of sensitivity is of $\sim 10\%$ on
 515 $B^0 \rightarrow \mu^+\mu^-$ signal, as evaluated from toy MC studies [13]. Different cuts have been
 516 explored, $\text{PID}_\mu > 0.1, 0.2, 0.3, 0.4, 0.5$, and the configuration giving the maximum increase
 517 of sensitivity on B^0 signal has been chosen.

518 The double misID probability for Run I and Run II data is given in Table 7, separately
 519 for the four exclusive $B^0 \rightarrow K^+\pi^-$, $B^0 \rightarrow \pi^+\pi^-$, $B_s^0 \rightarrow \pi^+K^-$ and $B_s^0 \rightarrow K^+K^-$ decays
 520 modes, after the $\text{PID}_{\mu,4}$ selection. The average double misID for the inclusive $B_{(s)}^0 \rightarrow h^+h^-$
 521 decays, $\epsilon_{hh \rightarrow \mu\mu}$, is also shown in last column. This has been obtained by weighting the
 522 values obtained for exclusive decays according to their relative production rates⁴.

⁴ The following branching fractions are assumed: $\mathcal{B}(B^0 \rightarrow K^+\pi^-) = (19.6 \pm 0.5)10^{-6}$, $\mathcal{B}(B^0 \rightarrow \pi^+\pi^-) = (5.12 \pm 0.19)10^{-6}$ $\mathcal{B}(B_s^0 \rightarrow \pi^+K^-) = (5.6 \pm 0.6)10^{-6}$, $\mathcal{B}(B_s^0 \rightarrow K^+K^-) = (25.2 \pm 1.7)10^{-6}$ [16], and $f_s/f_d = 0.259 \pm 0.016$ [17].

Table 7: Double misID probability integrated over the full BDT range, in units of 10^{-6} , for Run I, 2015 and 2016 data; $\text{PID}_{\mu,4}$ selection is used for Run I and 2015, $\text{PID}_{\mu,8}$ for 2016; the quoted errors include MC statistics and PID efficiency statistical uncertainties for the single modes, and the BRs and f_s/f_d for the average.

	$B^0 \rightarrow \pi\pi$	$B_s^0 \rightarrow KK$	$B^0 \rightarrow K\pi$	$B_s^0 \rightarrow K\pi$	Total($\epsilon_{hh \rightarrow \mu\mu}$)
Run I	10.8 ± 0.4	1.9 ± 0.1	3.9 ± 0.2	4.0 ± 0.2	4.6 ± 0.2
2015	7.1 ± 0.5	1.9 ± 0.2	3.2 ± 0.3	3.3 ± 0.3	3.6 ± 0.2
2016	8.6 ± 0.1	1.22 ± 0.03	2.79 ± 0.04	2.83 ± 0.05	3.39 ± 0.07

523 From the above values of $\epsilon_{hh \rightarrow \mu\mu}$, the number of $B_{(s)}^0 \rightarrow h^+h^-$ double misidentified
 524 events is evaluated as

$$N_{B_{(s)}^0 \rightarrow h^+h^- \rightarrow \mu^+\mu^-} = \epsilon_{B_{(s)}^0 \rightarrow \mu^+\mu^-}^{\text{TRIG|SEL}} \frac{N_{hh}^{\text{TIS}}}{\epsilon^{\text{TIS}} \epsilon^{\text{HLT2,MC}}} \epsilon_{hh \rightarrow \mu\mu} \quad (8)$$

525 where N_{hh}^{TIS} is the number of $B_{(s)}^0 \rightarrow h^+h^-$ TIS events (evaluated from $B^0 \rightarrow K^+\pi^-$ TIS
 526 events and correcting for the expected fraction of this mode), ϵ^{TIS} is the L0 and HLT1
 527 TIS efficiency for $B_{(s)}^0 \rightarrow h^+h^-$ events, $\epsilon^{\text{HLT2,MC}}$ is the HLT2 efficiency for $B_{(s)}^0 \rightarrow h^+h^-$
 528 events and $\epsilon_{B_{(s)}^0 \rightarrow \mu^+\mu^-}^{\text{TRIG|SEL}}$ is the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ trigger efficiency. All of the above inputs are
 529 evaluated in Sec. 7, and summarized in Table 8, together with the background events
 530 estimated for each dataset separately.

Table 8: Numbers entering into the computation of the $B_{(s)}^0 \rightarrow h^+h^- \rightarrow \mu\mu$ peaking background, for Run I, 2015 and 2016 data.

	2011	2012	2015	2016
$N_{B^0 \rightarrow K^+\pi^-}$ TIS	6833 ± 662	18012 ± 1221	8552 ± 848	28411 ± 2548
$\epsilon_{B_{(s)}^0 \rightarrow \mu^+\mu^-}^{\text{TRIG SEL}}$	$(94.2 \pm 2.0)\%$	$(94.1 \pm 2.0) \%$	$(96.1 \pm 3.2) \%$	$(97.3 \pm 1.3) \%$
ϵ^{TIS}	$(4.97 \pm 0.05) \%$	$(5.68 \pm 0.04) \%$	$(9.19 \pm 0.20) \%$	$(7.94 \pm 0.09) \%$
$\epsilon^{\text{HLT2,MC}}$	$(72.31 \pm 0.38)\%$	$(75.49 \pm 0.10) \%$	$(88.85 \pm 0.11) \%$	$(91.20 \pm 0.07) \%$
$\epsilon_{hh \rightarrow \mu\mu}$	$(4.6 \pm 0.2) \times 10^{-6}$	$(4.6 \pm 0.2) \times 10^{-6}$	$(3.6 \pm 0.2) \times 10^{-6}$	$(3.39 \pm 0.07) \times 10^{-6}$
$N_{hh \rightarrow \mu\mu}$	1.47 ± 0.16	2.92 ± 0.32	0.59 ± 0.06	2.12 ± 0.10

531 5.1.1 $B \rightarrow h^+h^-$ estimate as a function of BDT

532 In order to take into account the dependence of double misID from the BDT bins, we
 533 evaluated it as a function of BDT bin, as listed in Table 9. The observed effect is large,
 534 and it is given by the correlation between PID and BDT via the momentum of the selected
 535 candidates.

Table 9: Double misID probability in units of 10^{-6} as a function of the BDT bin, for Run I, 2015 and 2016 data; the quoted errors include MC statistics, PID efficiency statistical uncertainties, and the errors from the BRs and f_s/f_d .

BDT range	Run I	2015	2016
[0-0.25]	3.30 ± 0.15	2.52 ± 0.16	2.59 ± 0.08
[0.25-0.4]	3.77 ± 0.15	2.97 ± 0.17	2.92 ± 0.08
[0.4-0.5]	4.26 ± 0.15	3.39 ± 0.18	3.26 ± 0.08
[0.5-0.6]	4.65 ± 0.16	3.67 ± 0.19	3.49 ± 0.08
[0.6-0.7]	5.06 ± 0.16	4.00 ± 0.20	3.67 ± 0.07
[0.7-0.8]	5.57 ± 0.17	4.31 ± 0.22	3.99 ± 0.07
[0.8-0.9]	6.12 ± 0.18	4.66 ± 0.25	4.23 ± 0.07
[0.9-1.0]	6.69 ± 0.19	4.94 ± 0.28	4.44 ± 0.07

536 The number of $B_{(s)}^0 \rightarrow h^+h^- \rightarrow \mu\mu$ events in each BDT bin is then computed using
537 the same Eq. 8, where the double misID in bins of BDT is used, as given above. The BDT
538 dependence of Eq. 8 is indeed more complex, and appears in two factors more:

- 539 • The ratio of trigger efficiencies $\epsilon_{B_{(s)}^0 \rightarrow \mu^+\mu^-}^{\text{TRIG|SEL}} / (\epsilon^{\text{TIS}} \cdot \epsilon^{\text{HLT2,MC}})_{B^0 \rightarrow K^+\pi^-}$: while the BDT-
540 integrated values (given in Table 8 and computed in Sec. 7) are partially extracted
541 from data control samples, the BDT dependence for all these quantities is taken
542 from $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow K^+\pi^-$ MC samples.
- 543 • The actual value of the BDT *pdf* for $B_{(s)}^0 \rightarrow h^+h^-$ double misID events, before PID
544 selection: this is assumed the same as for signal and is given, with inclusion of the
545 above trigger correction, in Tables 46, 47 and 48 for Run I, 2015 and 2016 data,
546 respectively.

547 Finally, in Table 10 our estimates of $B_{(s)}^0 \rightarrow h^+h^-$ double misID background per BDT
548 bin is given, for Run I and 2015 data. Here and in the following we sum up the background
549 estimates for 2011 and 2012 datasets, since the signal fit is treating Run I data as a whole.
550 The same estimates on 2016 will follow once the data will be available to compute the
551 N_{hh} TIS normalization.

552 5.1.2 Systematics on $B \rightarrow h^+h^-$ estimate

553 A second independent estimate of the $B_{(s)}^0 \rightarrow h^+h^-$ peaking background is obtained by
554 fitting the mass spectrum on data in bins of BDT under a $\pi - \mu$ or $K - \mu$ selection. This
555 estimate will be used to assess the systematic error on the $B_{(s)}^0 \rightarrow h^+h^-$ double misID.
556 The selection is applied on the $B_{(s)}^0 \rightarrow h^+h^-$ stream and requiring the following additional

Table 10: $N_{B_{(s)}^0 \rightarrow h^+h^- \rightarrow \mu^+\mu^-}$ as extracted from $B^0 \rightarrow K^+\pi^-$ TIS events as a function of the BDT bin, for Run I, 2015 and 2016 data.

BDT range	Run I	2015	2016
[0-0.25]	0.640 ± 0.115	0.112 ± 0.019	0.441 ± 0.068
[0.25-0.4]	0.556 ± 0.067	0.074 ± 0.012	0.300 ± 0.041
[0.4-0.5]	0.398 ± 0.051	0.060 ± 0.009	0.202 ± 0.034
[0.5-0.6]	0.505 ± 0.044	0.055 ± 0.009	0.222 ± 0.026
[0.6-0.7]	0.482 ± 0.048	0.064 ± 0.009	0.221 ± 0.025
[0.7-0.8]	0.551 ± 0.050	0.066 ± 0.009	0.216 ± 0.027
[0.8-0.9]	0.585 ± 0.055	0.075 ± 0.010	0.242 ± 0.027
[0.9-1.0]	0.676 ± 0.057	0.084 ± 0.012	0.274 ± 0.030

557 PID cuts to the two candidate tracks to select either a $\pi - \mu$ or a $K - \mu$ couple.

$$\mu : \quad \text{ProbNN}_\mu > 0.5, \quad (9)$$

$$\pi : \quad \text{ProbNN}_\pi \times (1 - \text{ProbNN}_K) \times (1 - \text{ProbNN}_p) > 0.5, \quad (10)$$

$$K : \quad \text{ProbNN}_K \times (1 - \text{ProbNN}_\pi) \times (1 - \text{ProbNN}_p) > 0.5. \quad (11)$$

558 The mass fits on Run I data for the $\pi - \mu$ and $K - \mu$ selections are shown in Figs. 11 and
559 12, respectively (see App. H for 2015 and 2016 data): the contribution from $B_{(s)}^0 \rightarrow h^+h^-$
560 with a single misID appears as a very clear peak in these spectra, with semileptonic
561 $B^0 \rightarrow \pi^-\mu^+\nu_\mu$ and $B_s^0 \rightarrow K^-\mu^+\nu_\mu$ contributing as a shoulder on the left side. Data in the
562 first BDT bin, [0 – 0.25], are not considered by this method, as they're fully dominated by
563 combinatorial background.

564 The $B_{(s)}^0 \rightarrow h^+h^-$ component in the fit is parameterized with a double CB function,
565 with mean and resolution free to vary in the fit, and tail parameters determined on MC
566 from a fit to a cocktail of the four $B_{(s)}^0 \rightarrow h^+h^-$ modes, each one weighted according to the
567 corresponding BF and PID efficiency;. The semileptonic $B^0 \rightarrow \pi^-\mu^+\nu_\mu$ and $B_s^0 \rightarrow K^-\mu^+\nu_\mu$
568 mass shapes are fitted using Argus functions [18], with parameters fixed from MC and,
569 finally, an exponential fuction is used for combinatorial background. The slope of the
570 combinatorial background is a common parameter among the different BDT bins, which
571 are fitted simultaneously. This strategy is used to dump the strong correlation among the
572 semileptonic component and the combinatorial events, which would bias the fit, especially
573 in the high BDT bins. To justify this approach the right sidebands have been fitted
574 independently with an exponential functions, giving slopes in agreement between the
575 different BDT bins ⁵.

576 The $B_{(s)}^0 \rightarrow h^+h^-$ yields extracted from the above fits, characterized by a single muon
577 misidentification, are converted into an estimate of double misidentified $B_{(s)}^0 \rightarrow h^+h^-$ by

⁵bin2 = -0.00200 ± 0.00036 ; bin3 = -0.00126 ± 0.00074 ; bin4 = -0.00119 ± 0.00123 ; bin5 = -0.00344 ± 0.00196 ; bin6 = -0.00603 ± 0.00231 ; bin7 = -0.01000 ± 0.00806 ; bin8 = -0.01000 ± 0.00640 .

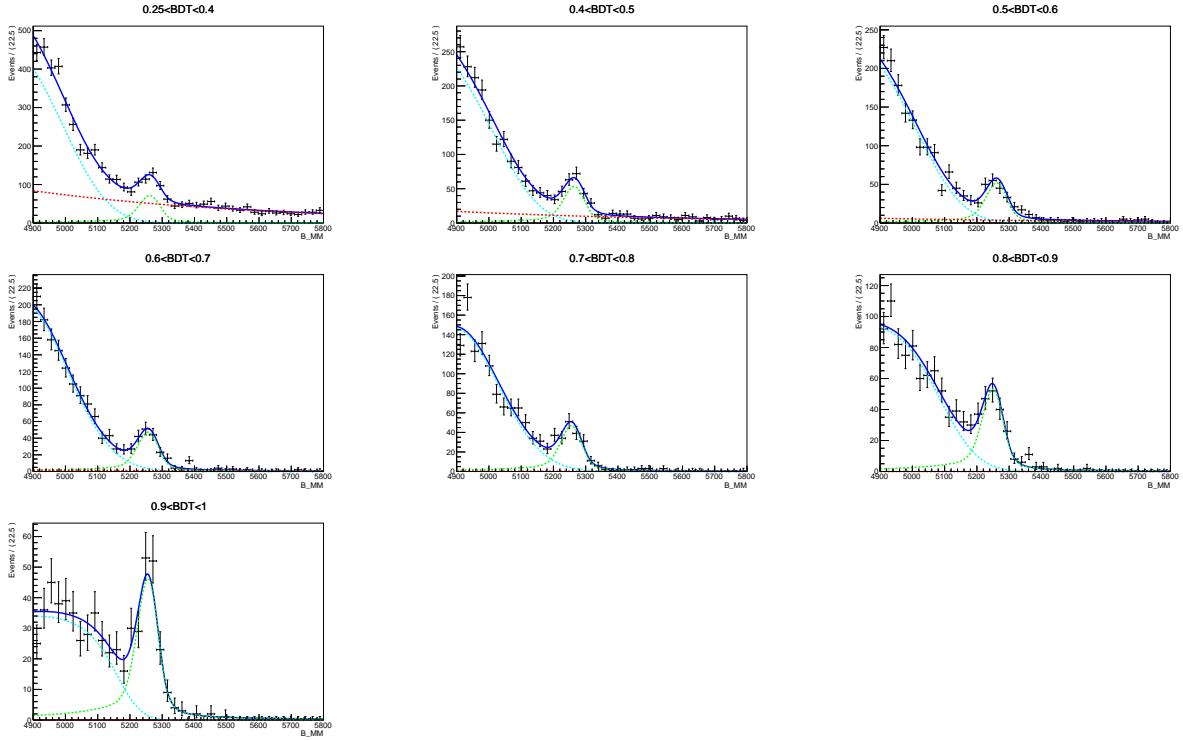


Figure 11: Fit to Run I data under $\pi - \mu$, for BDT bins > 0.25 ; the 3 components of the fit are: $B_{(s)}^0 \rightarrow h^+h^-$ (green), semileptonic (cyan), and combinatorial (red).

correcting for the ratio of $\frac{hh' \rightarrow \mu\mu}{hh' \rightarrow h\mu}$ PID and trigger selections: $PIDcalib$ binned efficiencies are convoluted with the relevant MC spectra for PID, while MC after full selection is used for trigger. The results are shown for $\pi - \mu$ fit in Table 11, for Run I and Run II data. Similar results for $K - \mu$ fit are listed in Table 12. The agreement between the two estimates is generally good, half of the bins being in agreement within the quoted fit errors. For the rest of the bins, and especially for 2015 sample, we introduce a fit systematic error, defined as the error which is needed to recover a 1σ difference between the two independent estimates. As central values, we assume instead the ones obtained from the $K - \mu$ fit, since the $B_{(s)}^0 \rightarrow h^+h^-$ over semileptonic ratio is much more favorable in this case.

The above estimate of the double misidentified $B_{(s)}^0 \rightarrow h^+h^-$ events from $K - \mu$ fit, Table 12, is used to assign a systematic error to the value obtained from TIS events, Table 10, which will be used as benchmark value in the signal fit. The same criterium as used for the semileptonic fit is adopted to compute the systematic error; the systematic from the semileptonic fit itself is also added in quadrature. The above procedure does not apply to the first BDT bin, for which the semileptonic fit is not available; however, the determination of the exclusive background is mostly irrelevant for this bin, since it is fully dominated by combinatorial background, at 99% level. The final values for $B_{(s)}^0 \rightarrow h^+h^-$ double misID with statistical and systematic errors are listed in Table 13 for Run I, 2015

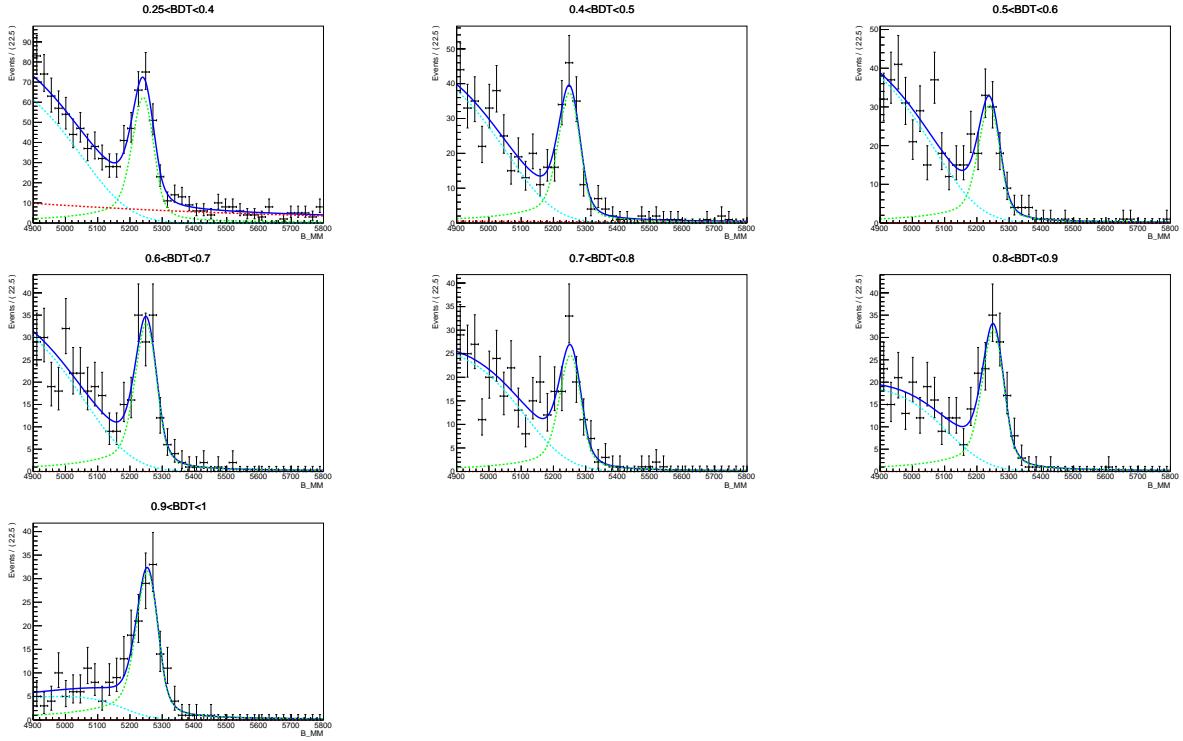


Figure 12: Fit to Run I data under $K - \mu$, for BDT bins > 0.25 ; the 3 components of the fit are: $B_{(s)}^0 \rightarrow h^+h^-$ (green), semileptonic (cyan), and combinatorial (red).

Table 11: $N_{B_{(s)}^0 \rightarrow h^+h^- \rightarrow \mu^+\mu^-}$ as extracted from $\pi - \mu$ fit as a function of the BDT bin, for Run I and Run II data; errors are statistical only.

BDT range	Run I	2015	2016
[0.25-0.4]	0.802 ± 0.106	0.201 ± 0.033	0.437 ± 0.046
[0.4-0.5]	0.637 ± 0.068	0.093 ± 0.019	0.333 ± 0.033
[0.5-0.6]	0.671 ± 0.068	0.102 ± 0.020	0.309 ± 0.029
[0.6-0.7]	0.644 ± 0.064	0.114 ± 0.021	0.272 ± 0.028
[0.7-0.8]	0.683 ± 0.070	0.084 ± 0.019	0.273 ± 0.041
[0.8-0.9]	0.819 ± 0.072	0.072 ± 0.017	0.387 ± 0.031
[0.9-1.0]	0.731 ± 0.064	0.085 ± 0.016	0.329 ± 0.028

and 2016 data. The procedure for calculating the systematic error gives zero for a couple of bins, since we do not observe any discrepancy in the comparison between TIS and semileptonic fit, nor in the semileptonic fit itself. This effect can be of course due to fluctuations in the uncertainties of the various estimates, and since it is not reasonable to have large discontinuities in the systematic error as a function of BDT, we assign in this case the largest systematic error among the two continuous bins.

Table 12: $N_{B_{(s)}^0 \rightarrow h^+h^- \rightarrow \mu^+\mu^-}$ as extracted from $K - \mu$ fit as a function of the BDT bin, for Run I and Run II data; errors are statistical only.

BDT range	Run I	2015	2016
[0.25-0.4]	0.949 ± 0.095	0.132 ± 0.028	0.51 ± 0.044
[0.4-0.5]	0.623 ± 0.064	0.101 ± 0.022	0.328 ± 0.033
[0.5-0.6]	0.534 ± 0.056	0.126 ± 0.026	0.357 ± 0.034
[0.6-0.7]	0.625 ± 0.061	0.064 ± 0.019	0.275 ± 0.030
[0.7-0.8]	0.516 ± 0.066	0.061 ± 0.02	0.319 ± 0.034
[0.8-0.9]	0.696 ± 0.070	0.114 ± 0.025	0.305 ± 0.033
[0.9-1.0]	0.713 ± 0.071	0.125 ± 0.025	0.272 ± 0.030

603 The fractional statistical errors amount to $\sim 6\%$, $\sim 11\%$ and $\sim 5\%$ for Run I, 2015
604 and 2016, respectively. The fractional systematic errors are in the range $15 - 25\%$ in the
605 last four BDT bins of Run 1, and in the range $15 - 35\%$ in the last four bins of 2016 data;
606 the situation is worst for 2015, but this sample is subdominant, corresponding to $\sim 10\%$
 of the total integrated luminosity.

Table 13: $N_{B_{(s)}^0 \rightarrow h^+h^- \rightarrow \mu^+\mu^-}$ used in the signal fit as a function of the BDT bin, for Run I, 2015 and 2016 data; first error is statistical, second is systematic.

BDT range	Run I	2015	2016
[0-0.25]	$0.640 \pm 0.115 \pm 0$	$0.112 \pm 0.019 \pm 0$	$0.441 \pm 0.068 \pm 0$
[0.25-0.4]	$0.556 \pm 0.067 \pm 0.377$	$0.074 \pm 0.012 \pm 0.073$	$0.300 \pm 0.041 \pm 0.204$
[0.4-0.5]	$0.398 \pm 0.051 \pm 0.210$	$0.060 \pm 0.009 \pm 0.033$	$0.202 \pm 0.034 \pm 0.117$
[0.5-0.6]	$0.505 \pm 0.044 \pm 0.105$	$0.055 \pm 0.009 \pm 0.065$	$0.222 \pm 0.026 \pm 0.129$
[0.6-0.7]	$0.482 \pm 0.048 \pm 0.120$	$0.064 \pm 0.009 \pm 0.041$	$0.221 \pm 0.025 \pm 0.037$
[0.7-0.8]	$0.551 \pm 0.050 \pm 0.137$	$0.066 \pm 0.009 \pm 0.042$	$0.216 \pm 0.027 \pm 0.093$
[0.8-0.9]	$0.585 \pm 0.055 \pm 0.097$	$0.075 \pm 0.010 \pm 0.041$	$0.242 \pm 0.027 \pm 0.083$
[0.9-1.0]	$0.676 \pm 0.057 \pm 0.112$	$0.084 \pm 0.012 \pm 0.040$	$0.274 \pm 0.030 \pm 0.040$

607

608 5.1.3 $B \rightarrow h^+h^-$ mass pdf

609 The $B_{(s)}^0 \rightarrow h^+h^-$ mass pdf is obtained from MC events without IsMuon and trigger
610 requests, and applying a momentum smearing procedure, tuned on MC for kaons and
611 pions separately, which accounts for the hadron decay in flight. This procedure overcomes
612 the obvious lack of MC statistics which would occur when requiring both hadrons to
613 satisfy the IsMuon algorithm. On top of the momentum smearing, event-by-event PID
614 weights are also applied. The mass spectra are calculated separately for the $B^0 \rightarrow K^+\pi^-$,
615 $B^0 \rightarrow K^+K^-$, $B_s^0 \rightarrow K^+K^-$ and $B_s^0 \rightarrow \pi^+K^-$ decay modes, and the results are then
616 combined according to their BRs and average PID efficiencies. The results are shown in

617 Fig. 13. In the same plot the mass spectrum is also superimposed which has been used in
 618 the past analysis round [6], from a similar momentum smearing procedure. The observed
 619 differences come mainly from the PID weighting procedure, the expected bias on the B^0
 620 BR being however less than 5%.

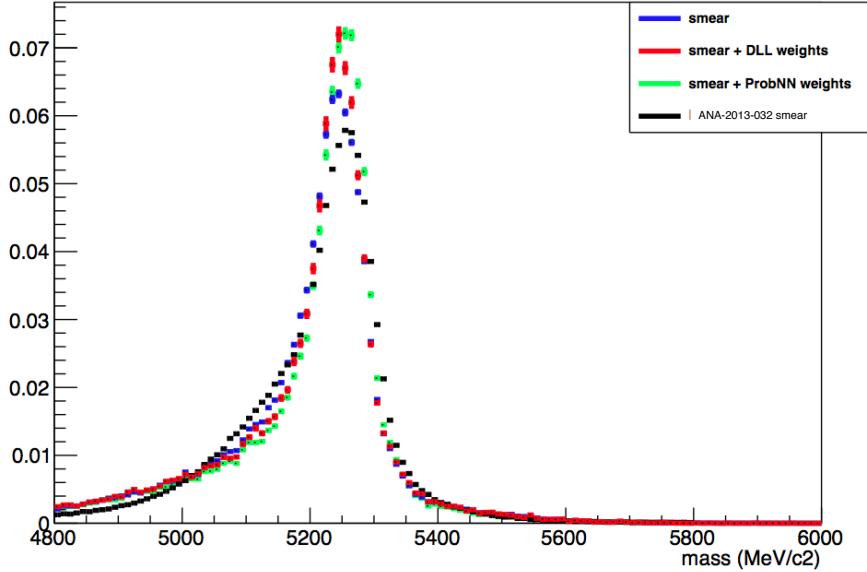


Figure 13: $B_{(s)}^0 \rightarrow h^+h^-$ inclusive mass spectrum, with double hadron misID simulated via a momentum smearing (blue), with additional event-by-event DLL (red) and ProbNN (green) weights, which is the one used in this analysis; the spectrum used in Ref. [6] is also shown (black).

621 5.2 Peaking background other than $B \rightarrow h^+h^-$

622 In this section the estimate of the exclusive decays with one or two real muons in the
 623 final state is given: $B^0 \rightarrow \pi^-\mu^+\nu_\mu$, $B_s^0 \rightarrow K^-\mu^+\nu_\mu$, $B^+ \rightarrow \pi^+\mu^+\mu^-$ and $B^0 \rightarrow \pi^0\mu^+\mu^-$,
 624 polluting the lower mass sideband only; $A_b^0 \rightarrow p\mu^-\nu$ and $B_c^+ \rightarrow J/\psi\mu^+\nu$, spanning the
 625 whole mass region. In all cases, the number of expected candidates has been estimated by
 626 normalising to the $B^+ \rightarrow J/\psi K^+$ channel:

$$N^{Exp} = N_{B^+ \rightarrow J/\psi K^+} \frac{f_x}{f_u} \frac{\mathcal{B}_x}{\mathcal{B}_{B^+ \rightarrow J/\psi K^+}} \frac{\epsilon_x^{tot}}{\epsilon_{B^+ \rightarrow J/\psi K^+}^{tot}} \\ = \beta_x \cdot \epsilon_x^{tot} \cdot \mathcal{B}_x \quad (12)$$

627 where x indicates a specific background channel. In practice, the β factor is calculated as
 628 $\beta_u = N_{B^+ \rightarrow J/\psi K^+} / (\mathcal{B}_{B^+ \rightarrow J/\psi K^+} \cdot \epsilon_{B^+ \rightarrow J/\psi K^+}^{tot})$ for B^0 and B^+ , and as $\beta_s = \beta_u \cdot f_s / f_d$ for B_s^0 ;
 629 for B_c^+ and Λ_b^0 decays β_u will be used as well, and the different hadronization factors will
 630 be absorbed either in the branching fraction or in the efficiency definition, respectively.

631 The total efficiency (ϵ_x^{tot}) for the given channel includes generation efficiency, reconstruc-
 632 tion and selection, including PID, and trigger. All of the different steps are estimated from
 633 MC simulation, but for PID, which is computed with the same method as for $B_{(s)}^0 \rightarrow h^+ h^-$.
 634 The trigger is evaluated from MC after the full selection is required, including PID. For Run
 635 I, the same MC efficiency is used for 2011 and 2012, while PID efficiencies are measured
 636 from the corresponding calibration samples, and convoluted with MC spectra. MC15
 637 samples and PID corrections from the 2015 calibration sample are used for 2015 data.
 638 Finally, MC15 samples and 2016 PID calibration samples are used for 2016 data. As for
 639 $B_{(s)}^0 \rightarrow h^+ h^-$, the background expectations have been computed in bins of BDT: in this
 640 case, the above efficiencies, but generation, are evaluated in bins of BDT. For channels
 641 implying misID, $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$, $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ and $\Lambda_b^0 \rightarrow p \mu^- \nu$, the PID requirement
 642 implies a strong reduction in statistics: for this reason the trigger efficiency cannot be
 643 computed from MC separately for each BDT bin, but BDT < 0.5 and BDT > 0.5 regions
 644 are considered.

645 The mass shapes to be used in the signal fit for each background source are extracted
 646 by fitting MC samples. An Argus function [18] has been used for all channels, with
 647 parameters determined for each BDT bin separately.

648 5.2.1 $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$

649 The semileptonic decay $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$, with a branching fraction of $(1.44 \pm 0.05) \cdot 10^{-4}$ [19],
 650 can contribute to the backgrounds of $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ when the pion is misidentified as
 651 muon. The invariant mass of the dimuon candidate is shifted to the left due to the
 652 missing neutrino and for the same reason the BDT is shifted to low values: however these
 653 two requirements are correlated as more “pointing” decays have also less missing mass,
 654 therefore the contribution and the invariant mass distribution of this decay have to be
 655 carefully estimated in bins of BDT.

656 MC samples of about 6 million events (for both MC12 and MC15) equally split in the
 657 two magnetic polarities were used to study this channel. The sample was produced with a
 658 cut at generator level requiring the $\pi\mu$ invariant mass to satisfy $m(\pi\mu) > 4500$ MeV/ c^2 ,
 659 corresponding to approximately 190 M events produced with standard DecProdCut, i.e.
 660 only acceptance. The MC sample was produced with ISGW2 form factor model, which is
 661 not in good agreement with recent data. This is shown in Fig. 14, where the ISGW2 q^2
 662 spectrum obtained at generator level is compared with a q^2 spectrum generated according
 663 to recent data and lattice results, fitted with a BCL z parameterization [20]⁶; in the
 664 same figure, also the original fit from Ref [20] is shown. From the above comparison we
 665 obtain per-event-weights to rescale our MC events. For the events surviving our invariant
 666 mass cut $M_{\mu\mu} > 4900$ MeV/ c^2 these weights are: 0.747 ± 0.019 ($0 < q^2 < 2.5$ GeV 2), and

⁶We thank Marcello Rotondo for providing us with the curve corresponding to the BCL fit of Ref [20]

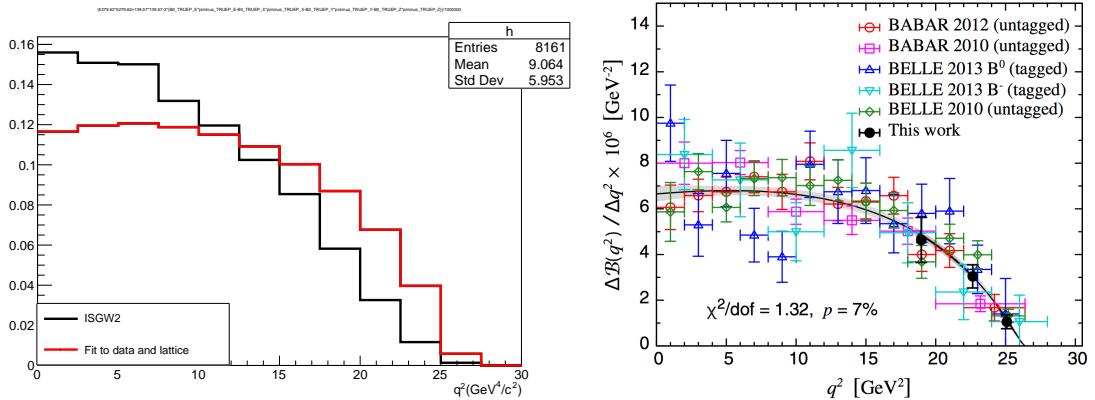


Figure 14: Left: $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$ q^2 spectra for ISGW2 (black) and for a fit to data and recent lattice results (red), histograms are normalized to the same area; Right: fit on data and lattice results used in the left plot, taken from Ref. [20].

667 0.791 ± 0.021 ($2.5 < q^2 < 5$ GeV 2), where the error is due to the MC statistics. Of course,
 668 such a procedure (“MC-driven” estimate in the following) needs to be validated, and for
 669 this reason we use the independent estimate of $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$ events from the data fitted
 670 under the $\pi - \mu$ selection (see Fig. 11), and correcting for the different PID and trigger
 671 efficiencies between the $\pi - \mu$ and the $\mu - \mu$ selections (‘“data-driven” estimate in the
 672 following).

673 The full breakdown of the numbers of the MC-driven estimate is listed in Table 14,
 674 in the full BDT range. The numbers of expected events in the invariant mass range
 675 [4900-6000] MeV/ c^2 and in the full BDT range for the Run I and Run II samples are:

$$\begin{aligned} N(B^0 \rightarrow \pi^- \mu^+ \nu_\mu)_{\text{RunI}} &= 44.0 \pm 2.3, \\ N(B^0 \rightarrow \pi^- \mu^+ \nu_\mu)_{2015} &= 6.2 \pm 0.4, \\ N(B^0 \rightarrow \pi^- \mu^+ \nu_\mu)_{2016} &= 25.2 \pm 1.5. \end{aligned}$$

676 The quoted errors include the uncertainty on the branching fraction, on the $B^+ \rightarrow J/\psi K^+$
 677 normalization (stat+syst) and on the total efficiency from Monte Carlo and data control
 678 samples (stat); this applies to all of the background estimates given in the rest of this
 679 section.

680 The expected events in bins of BDT from the same MC-driven estimate are listed in
 681 Table 15, while the data-driven estimates are listed in Table 16, for BDT bins above 0.25.
 682 For the extraction of the signal, we’ll assume as a reference value the MC-driven estimates,
 683 and we’ll use the data-driven estimates to assess the bin-by-bin systematic error, with
 684 the same criterium as used for $B^0_{(s)} \rightarrow h^+ h^-$ (including the treatment of bins with zero
 685 calculated systematic error). The resulting systematic errors are also listed in Table 15,
 686 and are below 15% for the last five BDT bins, for all datasets.

687 The invariant mass distributions in bins of BDT have been fitted on MC with an Argus

Table 14: Numbers entering into the computation of the $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$ peaking background, and estimated background yields for Run I, 2015 and 2016 data.

	Run I	2015	2016
β_u	$(7.18 \pm 0.24) \cdot 10^{11}$	$(1.27 \pm 0.06) \cdot 10^{11}$	$(4.62 \pm 0.21) \cdot 10^{11}$
Branching fraction	$(1.44 \pm 0.05) \cdot 10^{-4}$	$(1.44 \pm 0.05) \cdot 10^{-4}$	$(1.44 \pm 0.05) \cdot 10^{-4}$
ϵ_{gen}	$(6.745 \pm 0.0036) \cdot 10^{-3}$	$(6.995 \pm 0.0029) \cdot 10^{-3}$	$(6.995 \pm 0.0029) \cdot 10^{-3}$
ϵ_{sele}	$(6.80 \pm 0.12) \cdot 10^{-5}$	$(5.16 \pm 0.17) \cdot 10^{-5}$	$(5.79 \pm 0.04) \cdot 10^{-5}$
ϵ_{trig}	0.926 ± 0.011	0.935 ± 0.011	0.935 ± 0.011
$N(B^0 \rightarrow \pi^- \mu^+ \nu_\mu)$	44.0 ± 2.3	6.2 ± 0.4	25.2 ± 1.5

Table 15: $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$ expected yields per BDT bin, for Run I, 2015 and 2016 data; first error is statistical, second is systematic, evaluated by comparison with the data-driven estimates.

BDT range	Run I	2015	2016
[0-0.25]	$13.57 \pm 0.73 \pm 0$	$1.90 \pm 0.13 \pm 0$	$7.81 \pm 0.47 \pm 0$
[0.25-0.4]	$7.00 \pm 0.38 \pm 3.37$	$1.02 \pm 0.07 \pm 0.84$	$4.18 \pm 0.25 \pm 3.29$
[0.4-0.5]	$4.56 \pm 0.25 \pm 1.96$	$0.64 \pm 0.05 \pm 0.37$	$2.64 \pm 0.16 \pm 2.41$
[0.5-0.6]	$4.78 \pm 0.26 \pm 0.87$	$0.66 \pm 0.05 \pm 0.14$	$2.68 \pm 0.16 \pm 0.92$
[0.6-0.7]	$4.72 \pm 0.25 \pm 0.81$	$0.65 \pm 0.05 \pm 0.14$	$2.64 \pm 0.16 \pm 0.99$
[0.7-0.8]	$4.41 \pm 0.24 \pm 0.28$	$0.59 \pm 0.04 \pm 0.03$	$2.41 \pm 0.15 \pm 0.42$
[0.8-0.9]	$3.46 \pm 0.19 \pm 0.27$	$0.47 \pm 0.04 \pm 0.02$	$1.92 \pm 0.12 \pm 0.20$
[0.9-1.0]	$1.54 \pm 0.09 \pm 0.12$	$0.23 \pm 0.02 \pm 0.02$	$0.91 \pm 0.06 \pm 0.12$

Table 16: $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$ expected yields per BDT bin as extracted from $\pi - \mu$ fit, for Run I, 2015 and 2016 data.

BDT range	Run I	2015	2016
[0.25-0.4]	10.4 ± 0.6	1.87 ± 0.12	7.59 ± 0.34
[0.4-0.5]	6.56 ± 0.27	1.03 ± 0.08	5.12 ± 0.19
[0.5-0.6]	5.71 ± 0.22	0.81 ± 0.07	3.68 ± 0.14
[0.6-0.7]	5.60 ± 0.21	0.69 ± 0.06	3.71 ± 0.14
[0.7-0.8]	4.83 ± 0.20	0.60 ± 0.06	2.93 ± 0.14
[0.8-0.9]	3.66 ± 0.17	0.54 ± 0.05	2.21 ± 0.10
[0.9-1.0]	1.72 ± 0.11	0.18 ± 0.03	0.93 ± 0.07

function: the fit results are shown in Fig. 15, for RUN I MC sample. Possible biases in the mass pdf from the FF reweighting have been investigated: no effect has been found in any of the BDT bin (see App. I). This is expected, given the reduced q^2 range accessed by

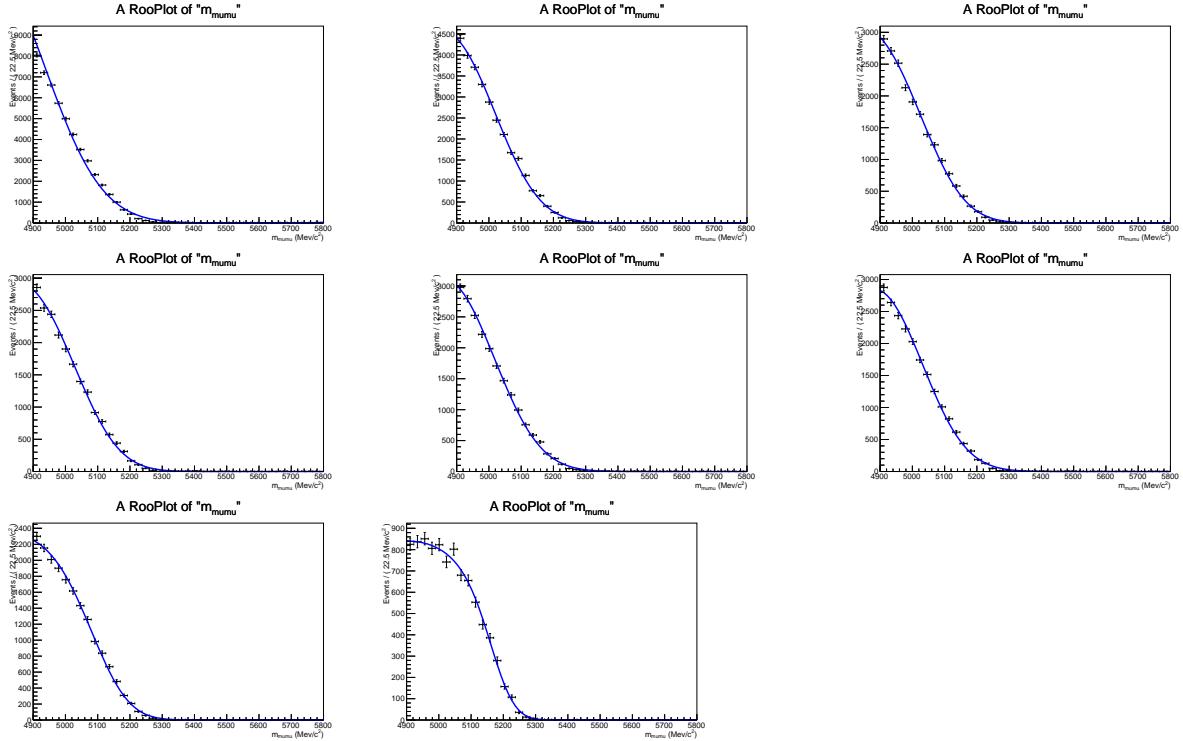


Figure 15: Invariant mass distributions and Argus fits to the $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$ channel.

the selected events, $q^2 < 5 \text{ GeV}^2$.

5.2.2 $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$

Similarly to $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$, also the semileptonic decay $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ can represent a significant peaking background for the analysis, if the kaon is misidentified as muon. The branching fraction of this decay, which has been never measured so far, is expected to be of the same order of $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$. Using an average of recent determinations of the $B_s \rightarrow K \ell \nu$ form factors from lattice QCD [21] [20] and $V_{ub} = 4.09 \pm 0.39$ from PDG [16], we obtain a branching fraction of $(1.42 \pm 0.35) \cdot 10^{-4}$, which is used in our estimates. Nevertheless, we expect a smaller contribution from $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ due to the larger mass shift and lower fragmentation fraction of the B_s^0 .

MC samples of about 6 million events (for both MC12 and MC15) were used, produced requiring the $K\mu$ invariant mass to satisfy $m(K\mu) > 4500 \text{ MeV}/c^2$. Similarly to the $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$ case, a form-factor reweighting is applied to the q^2 ISGW2 spectrum produced by our MC simulation. Different from the previous case, however, we rely on lattice results only. This is shown in Fig. 16, where the q^2 spectrum from ISGW2 is compared with the lattice prediction based on a recent FLAG average [22], which makes use of the same lattice data from which we extract the branching fraction estimate, Refs. [21] [20]; in the same figure, the original spectrum from Ref. [20] is also shown. From

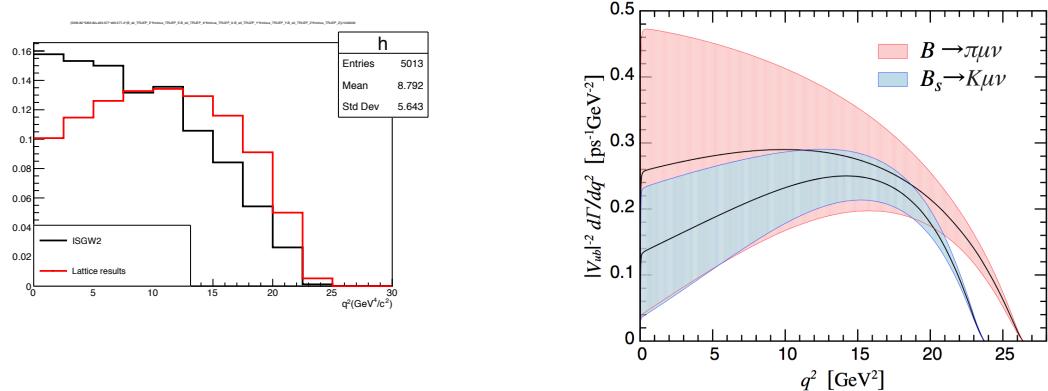


Figure 16: Left: $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ q^2 spectra for ISGW2 (black) and for recent lattice results (red), histograms are normalized the same area; Right: $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ and $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$ q^2 spectra from lattice results, as from Ref. [20].

the above comparison we obtain per-event-weights to rescale our MC events. For the events surviving our invariant mass cut $M_{\mu\mu} > 4900 \text{ MeV}/c^2$ these weights are: 0.637 ± 0.020 ($0 < q^2 < 2.5 \text{ GeV}^2$), and 0.750 ± 0.025 ($2.5 < q^2 < 5 \text{ GeV}^2$), where the error is due to the MC statistics. Even if there's a large uncertainty in the FF theoretical prediction, we do not include it in the reweighting procedure, since it is partly correlated to the BF uncertainty, and also because it will largely be covered by the systematic uncertainties from the data-driven estimate. The data-driven estimates are extracted from data fitted under the $K - \mu$ selection (see Fig. 12), and treated as for the $\pi - \mu$ case.

The full breakdown of the numbers of the MC-driven estimate is listed in Table 17. The numbers of expected events in the full BDT range for the Run I and Run II samples are:

$$\begin{aligned} N(B_s^0 \rightarrow K^- \mu^+ \nu_\mu)_{RunI} &= 5.04 \pm 1.31, \\ N(B_s^0 \rightarrow K^- \mu^+ \nu_\mu)_{2015} &= 0.92 \pm 0.24, \\ N(B_s^0 \rightarrow K^- \mu^+ \nu_\mu)_{2016} &= 2.81 \pm 0.74. \end{aligned}$$

The expected events in bins of BDT are listed in Table 18, while the data-driven estimates are listed in Table 19, for BDT bins above 0.25. For the extraction of the signal, we'll assume as a reference value the MC-driven estimates, and we'll use the data-driven estimates to assess the bin-by-bin systematic error, with the same criterium as used for $B_{(s)}^0 \rightarrow h^+ h^-$ and $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$ (including the treatment of bins with zero calculated systematic error, which are the majority for 2015 data). The resulting systematic errors are also listed in Table 15, and are larger than $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$ case. The overall impact is however expected to be small since this background source amounts to $\sim 20 - 25\%$ of $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$ in the most sensitive bins.

Table 17: Numbers entering into the computation of the $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ peaking background, and estimated background yields for Run I, 2015 and 2016 data.

	Run I	2015	2016
β_s	$(1.86 \pm 0.13) \cdot 10^{11}$	$(3.36e + 10 \pm 0.26) \cdot 10^{10}$	$(1.20e + 10 \pm 0.08) \cdot 10^{11}$
Branching fraction	$(1.42 \pm 0.35) \cdot 10^{-4}$	$(1.42 \pm 0.35) \cdot 10^{-4}$	$(1.42 \pm 0.35) \cdot 10^{-4}$
ϵ_{gen}	$(9.141 \pm 0.0051) \cdot 10^{-3}$	$(9.465 \pm 0.0044) \cdot 10^{-3}$	$(9.465 \pm 0.0044) \cdot 10^{-3}$
ϵ_{sele}	$(2.34 \pm 0.07) \cdot 10^{-5}$	$(2.18 \pm 0.11) \cdot 10^{-5}$	$(1.82 \pm 0.02) \cdot 10^{-5}$
ϵ_{trig}	0.898 ± 0.020	0.961 ± 0.012	0.961 ± 0.012
$N(B_s^0 \rightarrow K^- \mu^+ \nu_\mu)$	5.04 ± 1.31	0.92 ± 0.24	2.81 ± 0.74

Table 18: $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ expected yields per BDT bin, for Run I, 2015 and 2016 data.

BDT range	Run I	2015	2016
[0-0.25]	$0.93 \pm 0.24 \pm 0$	$0.16 \pm 0.04 \pm 0$	$0.50 \pm 0.13 \pm 0$
[0.25-0.4]	$0.62 \pm 0.16 \pm 0.40$	$0.12 \pm 0.03 \pm 0.09$	$0.36 \pm 0.09 \pm 0.54$
[0.4-0.5]	$0.47 \pm 0.12 \pm 0.31$	$0.10 \pm 0.03 \pm 0.08$	$0.29 \pm 0.08 \pm 0.41$
[0.5-0.6]	$0.63 \pm 0.16 \pm 0.36$	$0.11 \pm 0.03 \pm 0.10$	$0.34 \pm 0.09 \pm 0.23$
[0.6-0.7]	$0.71 \pm 0.18 \pm 0.17$	$0.13 \pm 0.03 \pm 0.09$	$0.40 \pm 0.10 \pm 0.20$
[0.7-0.8]	$0.76 \pm 0.20 \pm 0.29$	$0.14 \pm 0.04 \pm 0.09$	$0.41 \pm 0.11 \pm 0.28$
[0.8-0.9]	$0.65 \pm 0.17 \pm 0.19$	$0.12 \pm 0.03 \pm 0.08$	$0.35 \pm 0.09 \pm 0.24$
[0.9-1.0]	$0.27 \pm 0.07 \pm 0.08$	$0.05 \pm 0.01 \pm 0.03$	$0.16 \pm 0.04 \pm 0.13$

Table 19: $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ expected yields per BDT bin as extracted from $K - \mu$ fit, for Run I, 2015 and 2016 data.

BDT range	Run I	2015	2016
[0.25-0.4]	1.07 ± 0.13	0.23 ± 0.06	0.92 ± 0.07
[0.4-0.5]	0.82 ± 0.07	0.19 ± 0.03	0.72 ± 0.05
[0.5-0.6]	1.03 ± 0.08	0.10 ± 0.03	0.60 ± 0.05
[0.6-0.7]	0.97 ± 0.08	0.17 ± 0.04	0.64 ± 0.05
[0.7-0.8]	1.13 ± 0.10	0.24 ± 0.04	0.72 ± 0.06
[0.8-0.9]	0.91 ± 0.09	0.10 ± 0.03	0.62 ± 0.06
[0.9-1.0]	0.32 ± 0.06	0.07 ± 0.03	0.31 ± 0.04

The invariant mass distributions in bins of BDT have been fitted on MC with an Argus function: the fit results are shown in Fig. 17, for RUN I MC sample. As the mass shapes are very similar to the $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$ ones, while the yield being a factor of four smaller, this background source will be summed up to $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$ in the signal fit, as in the previous version of the analysis [6].

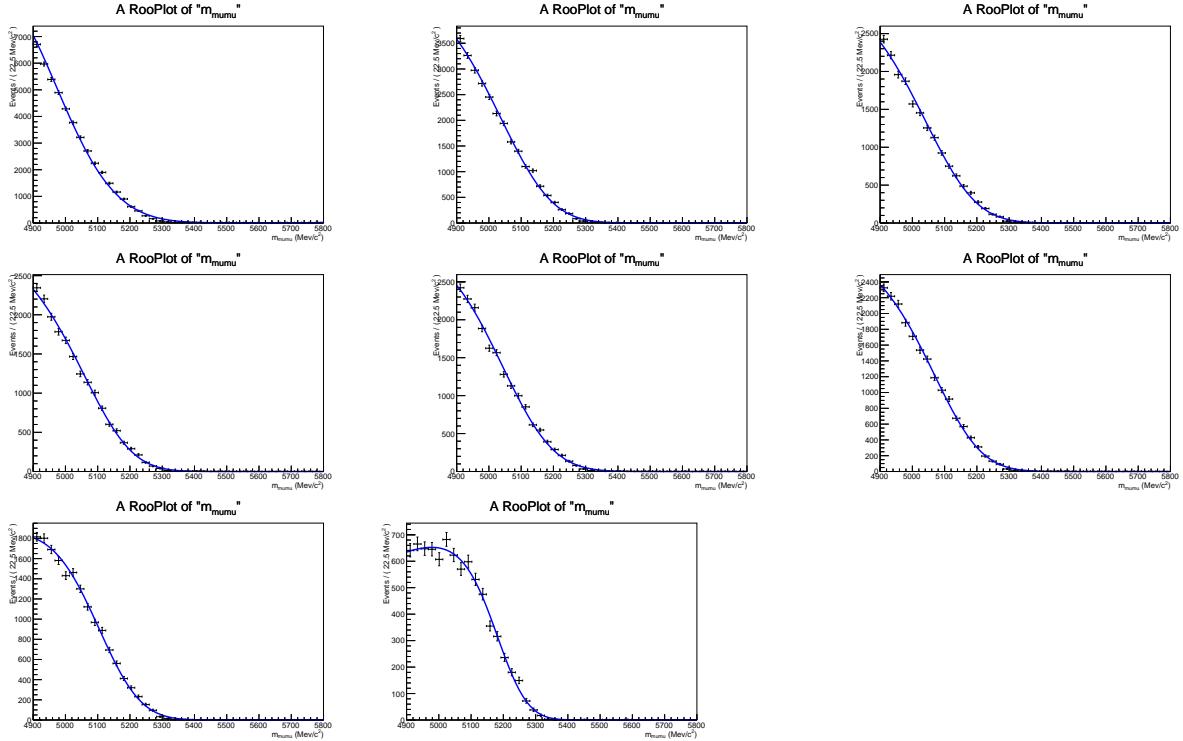


Figure 17: Invariant mass distributions and Argus fits to the $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ channel.

734 5.2.3 $B^+ \rightarrow \pi^+ \mu^+ \mu^-$ and $B^0 \rightarrow \pi^0 \mu^+ \mu^-$

735 The $B^{0(+)} \rightarrow \pi^{0(+)} \mu^+ \mu^-$ decays can fake the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ thanks to two final state muons
 736 forming a good vertex. While the dimuon invariant mass will not reach the signal region,
 737 it could affect the left sideband.

738 The $B^+ \rightarrow \pi^+ \mu^+ \mu^-$ has been measured by the LHCb collaboration with a branching
 739 fraction [23]:

$$\mathcal{B}(B^+ \rightarrow \pi^+ \mu^+ \mu^-) = (1.83 \pm 0.25) \times 10^{-8}.$$

740 The $B^0 \rightarrow \pi^0 \mu^+ \mu^-$ has not been observed so far, but we can profit of a theoretical
 741 estimate [24] of its rate relative to $B^+ \rightarrow \pi^+ \mu^+ \mu^-$:

$$\frac{\mathcal{B}(B^0 \rightarrow \pi^0 \mu^+ \mu^-)}{\mathcal{B}(B^+ \rightarrow \pi^+ \mu^+ \mu^-)} = 0.47^{+0.22}_{-0.18} \quad , \quad (13)$$

742 from which we obtain

$$\mathcal{B}(B^0 \rightarrow \pi^0 \mu^+ \mu^-) = (0.86 \pm 0.36) \times 10^{-8}.$$

743 As far as their contribution as $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ background is concerned, in terms of total
 744 selection efficiency and mass spectra, these two modes are very similar. For this reason,
 745 up to the previous analysis round [6], we simply scaled the $B^+ \rightarrow \pi^+ \mu^+ \mu^-$ normalization

according to the $B^0 \rightarrow \pi^0\mu^+\mu^-$ theoretical expectation. The present isolation algorithm, however, has much better performances than before, and for this reason $B^+ \rightarrow \pi^+\mu^+\mu^-$ events are more efficiently rejected by the BDT than $B^0 \rightarrow \pi^0\mu^+\mu^-$ ones. The above mechanism would generate a 25% underestimation of the overall $B^{0(+)} \rightarrow \pi^{0(+)}\mu^+\mu^-$, $\sim 1.5\sigma$ of the estimated error, if treated with the same procedure as before. This would not be a serious problem, since the expected yield is at the level of $\sim 20\%$ of $B^0 \rightarrow \pi^-\mu^+\nu_\mu$, nevertheless we decided to estimate $B^+ \rightarrow \pi^+\mu^+\mu^-$ and $B^0 \rightarrow \pi^0\mu^+\mu^-$ separately, the main reason being that a bias is avoided at almost no cost. Once the normalizations have been computed, the two background sources will be treated a single component in the signal fit, the dimuon mass spectra being identical.

MC samples of about 2 million events each have been used for the two decays, with both muons required in the detector acceptance; the same statistics has been generated for both MC12 and MC15. The full breakdown of the numbers of the MC-driven estimate is listed in Table 20. The numbers of expected $B^+ \rightarrow \pi^+\mu^+\mu^-$ events in the full BDT range for the Run I and Run II samples are:

$$\begin{aligned}
 N(B^+ \rightarrow \pi^+\mu^+\mu^-)_{RunI} &= 11.8 \pm 1.7, \\
 N(B^+ \rightarrow \pi^+\mu^+\mu^-)_{2015} &= 1.9 \pm 0.3, \\
 N(B^+ \rightarrow \pi^+\mu^+\mu^-)_{2016} &= 7.3 \pm 1.1.
 \end{aligned}$$

Table 20: Numbers entering into the computation of the $B^{0(+)} \rightarrow \pi^{0(+)}\mu^+\mu^-$ peaking background, and estimated background yields for Run I and 2015 data.

	Run I	2015	2016
β_u	$(7.18 \pm 0.24) \cdot 10^{11}$	$(1.27 \pm 0.06) \cdot 10^{11}$	$(4.62 \pm 0.21) \cdot 10^{11}$
$BF(B^+ \rightarrow \pi^+\mu^+\mu^-)$	$(1.83 \pm 0.25) \times 10^{-8}$	$(1.83 \pm 0.25) \times 10^{-8}$	$(1.83 \pm 0.25) \times 10^{-8}$
ϵ_{gen}	0.2486 ± 0.0011	0.2503 ± 0.0010	0.2503 ± 0.0010
ϵ_{sele}	$(3.75 \pm 0.004) \cdot 10^{-3}$	$(3.51 \pm 0.004) \cdot 10^{-3}$	$(3.70 \pm 0.004) \cdot 10^{-3}$
ϵ_{trig}	0.958 ± 0.002	0.933 ± 0.003	0.933 ± 0.003
$N(B^+ \rightarrow \pi^+\mu^+\mu^-)$	11.8 ± 1.7	1.9 ± 0.3	7.3 ± 1.1
$BF(B^0 \rightarrow \pi^0\mu^+\mu^-)$	$(0.86 \pm 0.36) \times 10^{-8}$	$(0.86 \pm 0.36) \times 10^{-8}$	$(0.86 \pm 0.36) \times 10^{-8}$
ϵ_{gen}	0.251 ± 0.003	0.251 ± 0.003	0.251 ± 0.003
ϵ_{sele}	$(3.80 \pm 0.004) \cdot 10^{-3}$	$(3.54 \pm 0.004) \cdot 10^{-3}$	$(3.73 \pm 0.004) \cdot 10^{-3}$
ϵ_{trig}	0.954 ± 0.002	0.940 ± 0.003	0.940 ± 0.003
$N(B^0 \rightarrow \pi^0\mu^+\mu^-)$	5.6 ± 2.3	0.9 ± 0.4	3.5 ± 1.5

The expected events in bins of BDT are listed in Table 21 for $B^+ \rightarrow \pi^+\mu^+\mu^-$ and in Table 22 for $B^0 \rightarrow \pi^0\mu^+\mu^-$, respectively.

The invariant mass distributions in bins of BDT have been fitted on MC with an Argus function: the fit results are shown in Fig. 18, for RUN I MC sample. The same mass pdf is used for both $B^+ \rightarrow \pi^+\mu^+\mu^-$ and $B^0 \rightarrow \pi^0\mu^+\mu^-$.

Table 21: $B^+ \rightarrow \pi^+ \mu^+ \mu^-$ expected yields per BDT bin, for Run I, 2015 and 2016 data.

BDT range	Run I	2015	2016
[0-0.25]	5.06 ± 0.72	0.82 ± 0.12	3.14 ± 0.46
[0.25-0.4]	2.33 ± 0.33	0.38 ± 0.06	1.47 ± 0.21
[0.4-0.5]	1.24 ± 0.18	0.19 ± 0.03	0.74 ± 0.11
[0.5-0.6]	1.06 ± 0.15	0.17 ± 0.03	0.64 ± 0.10
[0.6-0.7]	0.84 ± 0.12	0.13 ± 0.02	0.50 ± 0.07
[0.7-0.8]	0.61 ± 0.09	0.11 ± 0.02	0.40 ± 0.06
[0.8-0.9]	0.44 ± 0.07	0.076 ± 0.012	0.29 ± 0.05
[0.9-1.0]	0.19 ± 0.03	0.034 ± 0.006	0.13 ± 0.02

Table 22: $B^0 \rightarrow \pi^0 \mu^+ \mu^-$ expected yields per BDT bin, for Run I, 2015 and 2016 data.

BDT range	Run I	2015	2016
[0-0.25]	1.66 ± 0.70	0.26 ± 0.11	0.98 ± 0.41
[0.25-0.4]	0.86 ± 0.36	0.15 ± 0.06	0.56 ± 0.24
[0.4-0.5]	0.60 ± 0.25	0.09 ± 0.04	0.36 ± 0.15
[0.5-0.6]	0.60 ± 0.25	0.10 ± 0.04	0.38 ± 0.16
[0.6-0.7]	0.61 ± 0.26	0.10 ± 0.04	0.38 ± 0.16
[0.7-0.8]	0.57 ± 0.24	0.10 ± 0.04	0.37 ± 0.16
[0.8-0.9]	0.47 ± 0.20	0.080 ± 0.034	0.31 ± 0.13
[0.9-1.0]	0.19 ± 0.08	0.040 ± 0.017	0.15 ± 0.07

5.2.4 $\Lambda_b^0 \rightarrow p \mu^- \nu$

The $\Lambda_b^0 \rightarrow p \mu^- \nu$ can fake the signal if the proton is mis-identified as a muon. This can happen due to noise or *punch-through* muons from the proton shower in the calorimeters. The branching fraction has been recently measured for the first time by the LHCb collaboration in Ref. [25]:

$$\mathcal{B}(\Lambda_b^0 \rightarrow p \mu^- \nu) = (4.1 \pm 1.0) \cdot 10^{-4}, \quad (14)$$

where the uncertainty is dominated by theoretical error on the extrapolation to the full phase space.

As far as the hadronization fraction is concerned we use LHCb measurement [26] of:

$$r_\Lambda \equiv \frac{f_\Lambda}{f_u + f_d} = (0.404 \pm 0.017 \pm 0.027 \pm 0.105) \times [1 - (0.031 \pm 0.004 \pm 0.003) \times p_T(\text{GeV})], \quad (15)$$

and assuming $f_d = f_u$ we would have $\beta_\Lambda = \beta_u \cdot 2r_\Lambda$. However, given the p_T dependence of the r_Λ measurement, we include this correction in the efficiency computation, by assigning a per-event-weight given by $2 * r_\Lambda(p_T)$, multiplied by the already mentioned misID probability,

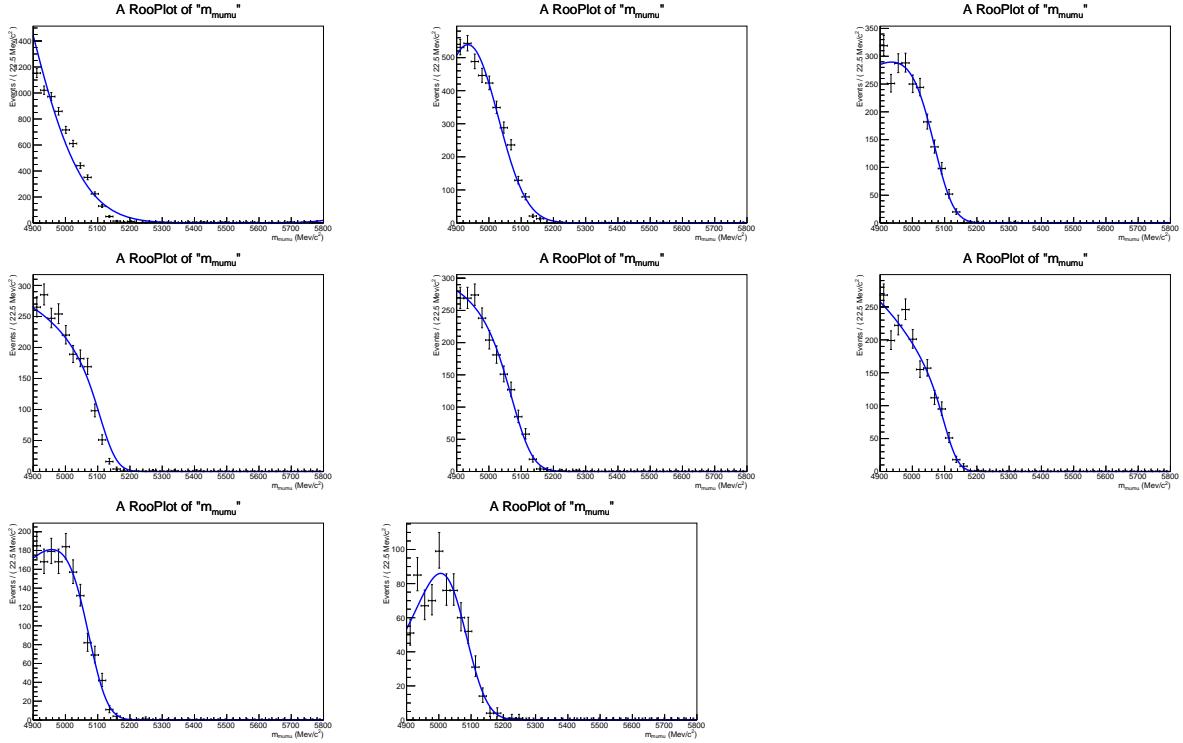


Figure 18: Invariant mass distributions and Argus fits to the $B^+ \rightarrow \pi^+ \mu^+ \mu^-$ channel for each BDT bin.

and use simply β_u to normalise. In the above equation, the transverse momentum of the combination formed by the muon and the proton is used, which is appropriate, since in the LHCb measurement the transverse momentum of the $\Lambda_c^+ - \mu$ in $\Lambda_b^0 \rightarrow \Lambda_c^+ \mu \nu X$ decays is used.

The $\Lambda_b^0 \rightarrow p \mu^- \nu$ has been studied with a sample of 2M events generated using LQCD form factors, for both MC12 and MC15. The events have been generated requiring $m(p\mu) > 4500$ MeV/ c^2 at generator level. The full breakdown of the numbers of the MC-driven estimate is listed in Table 23. The number of expected events in the full BDT range for the Run I sample is:

$$\begin{aligned} N((\Lambda_b^0 \rightarrow p \mu^- \nu)_{RunI}) &= 4.30 \pm 2.11, \\ N((\Lambda_b^0 \rightarrow p \mu^- \nu)_{2015}) &= 2.09 \pm 0.75, \\ N((\Lambda_b^0 \rightarrow p \mu^- \nu)_{2016}) &= 1.23 \pm 0.48. \end{aligned}$$

The large uncertainty is dominated by the systematic effects in the measurement of the proton misID probability which, for this extreme selection cut, amount to $\sim 40\%$.

The expected events in bins of BDT are listed in Table 24. The invariant mass distributions in bins of BDT have been fitted on MC with an Argus function: the fit results are shown in Fig. 19, for RUN I MC sample.

Table 23: Numbers entering into the computation of the $\Lambda_b^0 \rightarrow p\mu^-\nu$ peaking background, and estimated background yields for Run I, 2015 and 2016 data.

	Run I	2015	2016
β_u	$(7.18 \pm 0.24) \cdot 10^{11}$	$(1.27 \pm 0.06) \cdot 10^{11}$	$(4.62 \pm 0.21) \cdot 10^{11}$
Branching fraction	$(4.1 \pm 1.0) \cdot 10^{-4}$	$(4.1 \pm 1.0) \cdot 10^{-4}$	$(4.1 \pm 1.0) \cdot 10^{-4}$
ϵ_{gen}	$(1.34 \pm 0.01e - 04) \cdot 10^{-2}$	$1.32 \pm 0.01 \cdot 10^{-2}$	$1.32 \pm 0.01 \cdot 10^{-2}$
ϵ_{sele}	$(1.4 \pm 0.6) \cdot 10^{-6}$	$(1.3 \pm 0.5) \cdot 10^{-6}$	$(0.65 \pm 0.15) \cdot 10^{-6}$
ϵ_{trig}	0.801 ± 0.002	0.759 ± 0.002	0.759 ± 0.002
$N(\Lambda_b^0 \rightarrow p\mu^-\nu)$	4.30 ± 2.11	2.09 ± 0.75	1.23 ± 0.48

Table 24: $\Lambda_b^0 \rightarrow p\mu^-\nu$ expected yields per BDT bin, for Run I, 2015 and 2016 data.

BDT range	Run I	2015	2016
[0-0.25]	1.107 ± 0.551	0.374 ± 0.139	0.246 ± 0.100
[0.25-0.4]	0.616 ± 0.310	0.266 ± 0.099	0.154 ± 0.062
[0.4-0.5]	0.433 ± 0.219	0.225 ± 0.083	0.114 ± 0.047
[0.5-0.6]	0.556 ± 0.270	0.279 ± 0.102	0.176 ± 0.069
[0.6-0.7]	0.558 ± 0.271	0.315 ± 0.115	0.191 ± 0.075
[0.7-0.8]	0.496 ± 0.239	0.290 ± 0.105	0.172 ± 0.068
[0.8-0.9]	0.361 ± 0.176	0.219 ± 0.080	0.121 ± 0.049
[0.9-1.0]	0.124 ± 0.062	0.086 ± 0.032	0.046 ± 0.019

5.2.5 $B_c^+ \rightarrow J/\psi\mu^+\nu$

Partially reconstructed decays of the B_c^+ meson could be a dangerous background to the $B_s^0 \rightarrow \mu^+\mu^-$ signal due to the larger mass. However the hadronisation fraction of a b quark to a B_c^+ is about two orders of magnitude lower than B^+ mesons, though with great uncertainties.

The $B_c^+ \rightarrow J/\psi\mu^+\nu$ decay, with $J/\psi \rightarrow \mu^+\mu^-$, could pass the signal selection in case a good vertex is reconstructed between the muon from the semileptonic decay and the oppositely charged one from the J/ψ . Given the above topology, the muon isolation is expected to be rather effective in rejecting this background; however, a simple “ J/ψ veto” is used to reject further these events. The cut consists of vetoing events in which one of the two candidate muons, coupled to any other oppositely-charged muon in the event (selected with $ProbNN_\mu > 0.3$, which has an efficiency of about 97% compared to isMuon=1), falls in a window $(m_{\mu\mu} - M_{J/\psi}) < 30 \text{ MeV}/c^2$. This veto is expected to reject almost 64% of $B_c^+ \rightarrow J/\psi\mu^+\nu$ events (on the whole BDT range), with a negligible signal loss, 0.2%. All of the other exclusive backgrounds are unaffected as well, with the appropriate efficiencies being included in all the estimates given throughout the section. It should be noted that the majority of the vetoed $B_c^+ \rightarrow J/\psi\mu^+\nu$ events fall in the low BDT region because of the long track isolation variable discussed in sec. 6.1. This is shown in fig. 117 of appendix J.

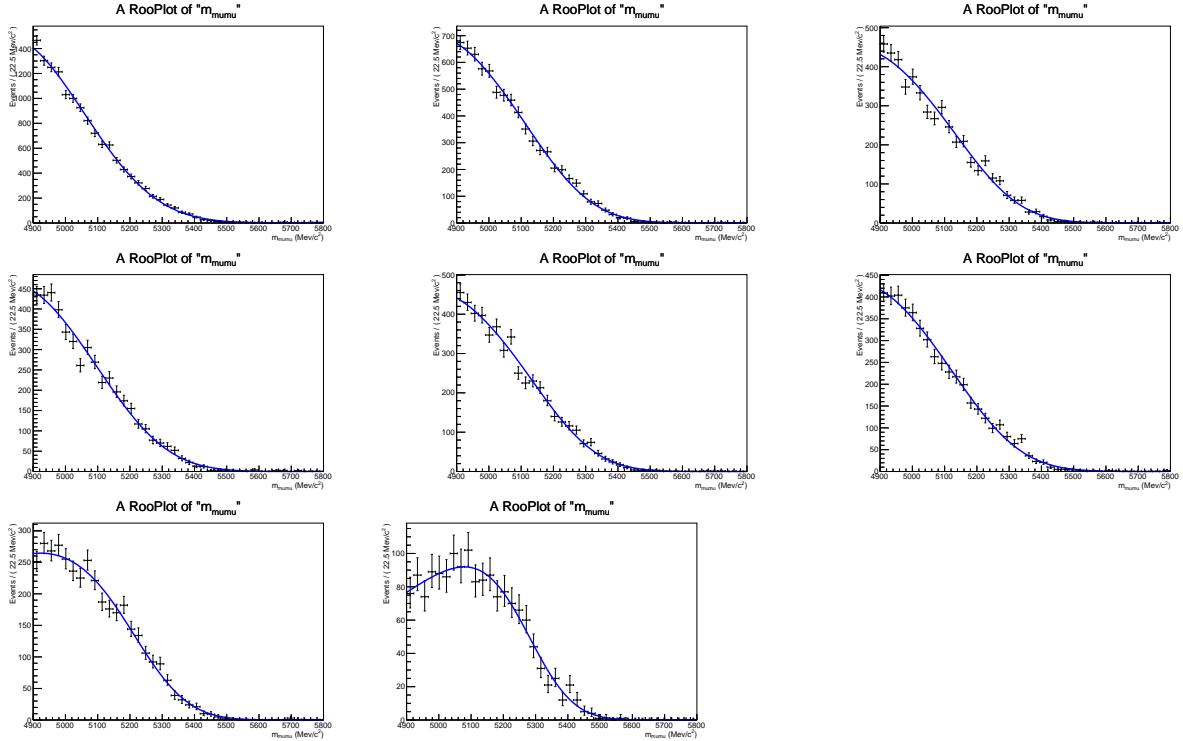


Figure 19: Invariant mass distributions and Argus fits to the $\Lambda_b^0 \rightarrow p\mu^-\nu$ channel for each BDT bin.

809 The absolute branching fraction of this decay is not known. But as we normalise
 810 directly to $B^+ \rightarrow J/\psi K^+$, we can use a couple of recent measurements from LHCb: the
 811 ratio [27]

$$\mathcal{R}_1 = \frac{\sigma(B_c^+) \mathcal{B}(B_c^+ \rightarrow J/\psi \pi)}{\sigma(B^+) \mathcal{B}(B^+ \rightarrow J/\psi K^+)} = (0.68 \pm 0.12) \times 10^{-2},$$

812 for $p_T > 4 \text{ GeV}/c$ and $2.5 < \eta < 4.5$, and the ratio [28]

$$\mathcal{R}_2 = \frac{\mathcal{B}(B_c^+ \rightarrow J/\psi \pi)}{\mathcal{B}(B_c^+ \rightarrow J/\psi \mu\nu)} = 0.0469 \pm 0.0054.$$

813 From the above, one can define an effective $B_c^+ \rightarrow J/\psi \mu^+\nu$ branching fraction as

$$\mathcal{B} = \mathcal{R}_1 / \mathcal{R}_2 \cdot \alpha \cdot \mathcal{B}(B^+ \rightarrow J/\psi K^+) \cdot \mathcal{B}(J/\psi \rightarrow \mu^+\mu^-) = (9.5 \pm 2.0) \times 10^{-6}, \quad (16)$$

814 where $\alpha = 1.049 \pm 0.028$ is a correction factor accounting for the ratio of acceptances
 815 between $B^+ \rightarrow J/\psi K^+$ and $B_c^+ \rightarrow J/\psi \mu^+\nu$ under the kinematic selection of Ref [27], as
 816 evaluated from ad hoc MC generations without acceptance cuts.

817 The MC generation has been done with BcDaughter-in-LHCbAcceptance, with a
 818 dimuon minimum mass of $4.5 \text{ GeV}/c^2$, and using a phase space model for Bc decay, which
 819 is not ideal for acceptance calculation. Indeed using an ad hoc MC production with FF

parameterization according to Kiselev model, a $\sim 20\%$ bias has been found and corrected in the generation efficiency. A part from this, the rest proceeds as in the previous cases, with relevant numbers listed in Table 25. The numbers of expected events in teh full BDT range for the Run I and Run II samples are:

$$\begin{aligned} N(B_c^+ \rightarrow J/\psi \mu^+ \nu)_{RunI} &= 117.0 \pm 25.1, \\ N(B_c^+ \rightarrow J/\psi \mu^+ \nu)_{2016} &= 13.3 \pm 2.9, \\ N(B_c^+ \rightarrow J/\psi \mu^+ \nu)_{2016} &= 50.8 \pm 11.1. \end{aligned}$$

Despite the large total number of events, 80% of the yield is concentrated in the first BDT bin, so that in the most sensitive region this background source is still suppressed with respect to $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$.

Table 25: Numbers entering into the computation of the $B_c^+ \rightarrow J/\psi \mu^+ \nu$ peaking background, and estimated background yields for Run I, 2015 and 2016 data.

	Run I	2015	2016
β_u	$(7.18 \pm 0.24) \cdot 10^{11}$	$(1.27 \pm 0.06) \cdot 10^{11}$	$(4.62 \pm 0.21) \cdot 10^{11}$
Eff. Branching fraction	$(9.5 \pm 2.0) \cdot 10^{-6}$	$(9.5 \pm 2.0) \cdot 10^{-6}$	$(9.5 \pm 2.0) \cdot 10^{-6}$
ϵ_{gen}	$(4.35 \pm 0.006) \cdot 10^{-3}$	$(4.35 \pm 0.006) \cdot 10^{-3}$	$(4.35 \pm 0.006) \cdot 10^{-3}$
ϵ_{sele}	$(4.09 \pm 0.006) \cdot 10^{-3}$	$(2.64 \pm 0.006) \cdot 10^{-3}$	$(2.78 \pm 0.007) \cdot 10^{-3}$
ϵ_{trig}	0.964 ± 0.002	0.956 ± 0.003	0.956 ± 0.003
$N(B_c^+ \rightarrow J/\psi \mu^+ \nu)$	117.0 ± 25.1	13.3 ± 2.9	52.0 ± 11.1

The expected events in bins of BDT are listed in Table 26. No MC events have been selected in the most sensitive BDT bin; most probably, however, this bin will be merged in the final fit with the previous one, as no combinatorial background on the sidebands is observed.

Table 26: $B_c^+ \rightarrow J/\psi \mu^+ \nu$ expected yields per BDT bin, for Run I and 2015 data.

BDT range	Run I	2015	2016
[0-0.25]	92.32 ± 19.9	10.10 ± 2.22	38.74 ± 8.48
[0.25-0.4]	13.17 ± 2.88	1.56 ± 0.36	5.97 ± 1.36
[0.4-0.5]	4.68 ± 1.06	0.71 ± 0.17	2.70 ± 0.65
[0.5-0.6]	3.45 ± 0.80	0.44 ± 0.11	1.69 ± 0.44
[0.6-0.7]	2.06 ± 0.50	0.22 ± 0.07	0.85 ± 0.25
[0.7-0.8]	1.02 ± 0.27	0.19 ± 0.06	0.75 ± 0.23
[0.8-0.9]	0.30 ± 0.11	0.026 ± 0.016	0.10 ± 0.06
[0.9-1.0]	0	0	0

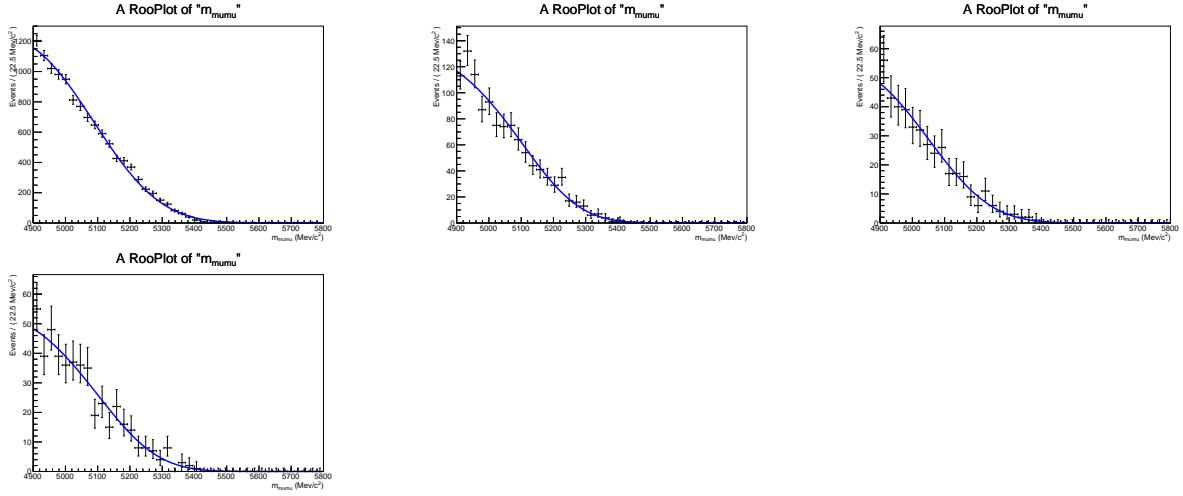


Figure 20: Invariant mass distributions and Argus fits to the $B_c^+ \rightarrow J/\psi \mu^+ \nu$ channel for each BDT bin; bins above 0.6 have been merged because of lack of MC statistics.

831 The invariant mass distributions in bins of BDT have been fitted on MC with an Argus
 832 function: the fit results are shown in Fig. 20, for RUN I MC sample. BDT bins above 0.6
 833 have been merged together because of lack of MC statistics.

834 **5.3 Background summary tables**

835 The background estimates in bins of BDT and in the mass range is $(4900 - 6000)$ MeV/ c^2
836 for $B_{(s)}^0 \rightarrow h^+h^-$ (Table 13), $B^0 \rightarrow \pi^-\mu^+\nu_\mu$ (Table 15), $B_s^0 \rightarrow K^-\mu^+\nu_\mu$ (Table 18),
837 $B^+ \rightarrow \pi^+\mu^+\mu^-$ (Table 21), $B^0 \rightarrow \pi^0\mu^+\mu^-$ (Table 22), $B_c^+ \rightarrow J/\psi\mu^+\nu$ (Table 26), and
838 $\Lambda_b^0 \rightarrow p\mu^-\nu$ (Table 24), are summarized in the following Tables 5.3, 5.3, 5.3 for Run I,
839 2015 and 2016 data, respectively. The exclusive background estimates are also shown in
840 blinded region, $(5200 - 5450)$ MeV/ c^2 , in Tables 5.3, 5.3, 5.3 for Run I, 2015 and 2016
841 data, respectively; in this mass interval the contribution of both $B^+ \rightarrow \pi^+\mu^+\mu^-$ and
842 $B^0 \rightarrow \pi^0\mu^+\mu^-$ is negligible, and it is not listed.

Table 27: Exclusive background yields per BDT bin in the mass range (4900 – 6000) MeV/ c^2 , with their total estimated uncertainties, for Run I data.

BDT range	$B_{(s)}^0 \rightarrow h^+h^-$	$B^0 \rightarrow \pi^-\mu^+\nu_\mu$	$B_s^0 \rightarrow K^-\mu^+\nu_\mu$	$B^+ \rightarrow \pi^+\mu^+\mu^-$	$B^0 \rightarrow \pi^0\mu^+\mu^-$	$B_c^+ \rightarrow J/\psi\mu^+\nu$	$A_b^0 \rightarrow p\mu^-\nu$
[0-0.25]	0.64 ± 0.12	13.57 ± 0.74	0.93 ± 0.24	5.06 ± 0.72	1.66 ± 0.70	92.32 ± 19.9	1.11 ± 0.55
[0.25-0.4]	0.56 ± 0.38	7.00 ± 3.39	0.62 ± 0.43	2.33 ± 0.33	0.86 ± 0.36	13.17 ± 2.88	0.62 ± 0.31
[0.4-0.5]	0.40 ± 0.22	4.56 ± 1.98	0.47 ± 0.34	1.24 ± 0.18	0.60 ± 0.25	4.68 ± 1.06	0.43 ± 0.22
[0.5-0.6]	0.51 ± 0.11	4.78 ± 0.91	0.63 ± 0.40	1.06 ± 0.15	0.60 ± 0.25	3.45 ± 0.80	0.56 ± 0.27
[0.6-0.7]	0.48 ± 0.13	4.72 ± 0.85	0.71 ± 0.25	0.84 ± 0.12	0.61 ± 0.26	2.06 ± 0.50	0.56 ± 0.27
[0.7-0.8]	0.55 ± 0.15	4.41 ± 0.37	0.76 ± 0.35	0.61 ± 0.09	0.57 ± 0.24	1.02 ± 0.27	0.50 ± 0.24
[0.8-0.9]	0.59 ± 0.11	3.46 ± 0.33	0.65 ± 0.25	0.44 ± 0.07	0.47 ± 0.20	0.30 ± 0.11	0.36 ± 0.18
[0.9-1.0]	0.68 ± 0.13	1.54 ± 0.15	0.27 ± 0.11	0.19 ± 0.03	0.19 ± 0.08	0	0.12 ± 0.06

Table 28: Exclusive background yields per BDT bin in the mass range (4900 – 6000) MeV/ c^2 , with their total estimated uncertainties, for 2015 data.

BDT range	$B_{(s)}^0 \rightarrow h^+h^-$	$B^0 \rightarrow \pi^-\mu^+\nu_\mu$	$B_s^0 \rightarrow K^-\mu^+\nu_\mu$	$B^+ \rightarrow \pi^+\mu^+\nu_\mu$	$B^0 \rightarrow \pi^0\mu^+\mu^-$	$B_c^+ \rightarrow J/\psi\mu^+\nu$	$A_b^0 \rightarrow p\mu^-\nu$
[0-0.25]	0.1112 ± 0.019	1.90 ± 0.13	0.16 ± 0.04	0.82 ± 0.12	0.26 ± 0.11	10.10 ± 2.22	0.158 ± 0.079
[0.25-0.4]	0.074 ± 0.074	1.02 ± 0.84	0.12 ± 0.10	0.38 ± 0.06	0.15 ± 0.06	1.56 ± 0.36	0.092 ± 0.046
[0.4-0.5]	0.060 ± 0.035	0.64 ± 0.38	0.10 ± 0.09	0.19 ± 0.03	0.09 ± 0.04	0.71 ± 0.17	0.066 ± 0.033
[0.5-0.6]	0.055 ± 0.066	0.66 ± 0.14	0.11 ± 0.10	0.17 ± 0.03	0.10 ± 0.04	0.44 ± 0.11	0.089 ± 0.043
[0.6-0.7]	0.064 ± 0.042	0.65 ± 0.14	0.13 ± 0.09	0.13 ± 0.02	0.10 ± 0.04	0.22 ± 0.07	0.090 ± 0.044
[0.7-0.8]	0.066 ± 0.043	0.59 ± 0.05	0.14 ± 0.10	0.11 ± 0.02	0.10 ± 0.04	0.19 ± 0.06	0.080 ± 0.039
[0.8-0.9]	0.075 ± 0.042	0.47 ± 0.04	0.12 ± 0.08	0.076 ± 0.012	0.080 ± 0.034	0.026 ± 0.016	0.059 ± 0.029
[0.9-1.0]	0.084 ± 0.042	0.23 ± 0.03	0.05 ± 0.04	0.034 ± 0.006	0.040 ± 0.017	0	0.021 ± 0.011

Table 29: Exclusive background yields per BDT bin in the mass range (4900 – 6000) MeV/ c^2 , with their total estimated uncertainties, for 2016 data.

BDT range	$B_{(s)}^0 \rightarrow h^+h^-$	$B^0 \rightarrow \pi^-\mu^+\nu_\mu$	$B_s^0 \rightarrow K^-\mu^+\nu_\mu$	$B^+ \rightarrow \pi^+\mu^+\mu^-$	$B^0 \rightarrow \pi^0\mu^+\mu^-$	$B_c^+ \rightarrow J/\psi\mu^+\nu$	$A_b^0 \rightarrow p\mu^-\nu$
[0-0.25]	0.44 ± 0.07	7.81 ± 0.47	0.50 ± 0.13	3.14 ± 0.46	0.98 ± 0.41	38.74 ± 8.48	0.25±0.10
[0.25-0.4]	0.30 ± 0.21	4.18 ± 3.40	0.36 ± 0.55	1.47 ± 0.21	0.56 ± 0.24	5.97 ± 1.36	0.15±0.06
[0.4-0.5]	0.20 ± 0.12	2.64 ± 2.47	0.29 ± 0.42	0.74 ± 0.11	0.36 ± 0.15	2.70 ± 0.65	0.11±0.05
[0.5-0.6]	0.22 ± 0.13	2.68 ± 0.99	0.34 ± 0.25	0.64 ± 0.10	0.38 ± 0.16	1.69 ± 0.44	0.18±0.07
[0.6-0.7]	0.22 ± 0.04	2.64 ± 1.06	0.40 ± 0.23	0.50 ± 0.07	0.38 ± 0.16	0.85 ± 0.25	0.19±0.08
[0.7-0.8]	0.22 ± 0.10	2.41 ± 0.50	0.41 ± 0.31	0.40 ± 0.06	0.37 ± 0.16	0.75 ± 0.23	0.17±0.07
[0.8-0.9]	0.24 ± 0.09	1.92 ± 0.28	0.35 ± 0.26	0.29 ± 0.05	0.31 ± 0.13	0.10 ± 0.06	0.12±0.05
[0.9-1.0]	0.27 ± 0.05	0.91 ± 0.13	0.16 ± 0.14	0.13 ± 0.02	0.15 ± 0.07	0	0.05±0.05

Table 30: Exclusive background yields per BDT bin in the mass range (5200 – 5450) MeV/ c^2 , with their statistical uncertainties, for Run I data.

BDT range	$B_{(s)}^0 \rightarrow h^+ h^-$	$B^0 \rightarrow \pi^- \mu^+ \nu_\mu$	$B_s^0 \rightarrow K^- \mu^+ \nu_\mu$	$B_c^+ \rightarrow J/\psi \mu^+ \nu$	$A_b^0 \rightarrow p \mu^- \nu$
[0-0.25]	0.43 ± 0.08	0.201 ± 0.013	0.030 ± 0.008	11.1 ± 2.43	0.142 ± 0.071
[0.25-0.4]	0.37 ± 0.26	0.091 ± 0.007	0.020 ± 0.005	1.37 ± 0.35	0.090 ± 0.045
[0.4-0.5]	0.27 ± 0.15	0.068 ± 0.006	0.016 ± 0.004	0.36 ± 0.12	0.062 ± 0.031
[0.5-0.6]	0.34 ± 0.08	0.076 ± 0.006	0.023 ± 0.006	0.41 ± 0.13	0.082 ± 0.040
[0.6-0.7]	0.33 ± 0.09	0.078 ± 0.007	0.024 ± 0.006	0.10 ± 0.05	0.085 ± 0.041
[0.7-0.8]	0.37 ± 0.10	0.077 ± 0.007	0.029 ± 0.008	0.13 ± 0.06	0.079 ± 0.038
[0.8-0.9]	0.39 ± 0.08	0.062 ± 0.006	0.033 ± 0.009	0	0.079 ± 0.038
[0.9-1.0]	0.46 ± 0.08	0.051 ± 0.005	0.022 ± 0.006	0	0.035 ± 0.018

Table 31: Exclusive background yields per BDT bin in the mass range (5200 – 5450) MeV/ c^2 , with their statistical uncertainties, for 2015 data.

BDT range	$B_{(s)}^0 \rightarrow h^+ h^-$	$B^0 \rightarrow \pi^- \mu^+ \nu_\mu$	$B_s^0 \rightarrow K^- \mu^+ \nu_\mu$	$B_c^+ \rightarrow J/\psi \mu^+ \nu$	$A_b^0 \rightarrow p \mu^- \nu$
[0-0.25]	0.076 ± 0.013	0.027 ± 0.002	0.005 ± 0.001	1.25 ± 0.29	0.046 ± 0.017
[0.25-0.4]	0.050 ± 0.050	0.014 ± 0.001	0.004 ± 0.001	0.17 ± 0.05	0.035 ± 0.013
[0.4-0.5]	0.040 ± 0.023	0.008 ± 0.001	0.003 ± 0.001	0.065 ± 0.027	0.031 ± 0.011
[0.5-0.6]	0.037 ± 0.045	0.010 ± 0.001	0.004 ± 0.001	0.051 ± 0.024	0.037 ± 0.014
[0.6-0.7]	0.043 ± 0.028	0.010 ± 0.001	0.004 ± 0.001	0.034 ± 0.019	0.043 ± 0.016
[0.7-0.8]	0.045 ± 0.029	0.010 ± 0.001	0.005 ± 0.001	0.025 ± 0.016	0.045 ± 0.017
[0.8-0.9]	0.051 ± 0.028	0.009 ± 0.001	0.005 ± 0.001	0	0.045 ± 0.017
[0.9-1.0]	0.057 ± 0.028	0.008 ± 0.001	0.004 ± 0.001	0	0.025 ± 0.009

Table 32: Exclusive background yields per BDT bin in the mass range (5200 – 5450) MeV/ c^2 , with their statistical uncertainties, for 2016 data.

BDT range	$B_{(s)}^0 \rightarrow h^+ h^-$	$B^0 \rightarrow \pi^- \mu^+ \nu_\mu$	$B_s^0 \rightarrow K^- \mu^+ \nu_\mu$	$B_c^+ \rightarrow J/\psi \mu^+ \nu$	$A_b^0 \rightarrow p \mu^- \nu$
[0-0.25]	0.30 ± 0.05	0.112 ± 0.008	0.016 ± 0.004	4.8 ± 1.1	0.030 ± 0.012
[0.25-0.4]	0.20 ± 0.14	0.057 ± 0.005	0.011 ± 0.003	0.64 ± 0.20	0.020 ± 0.008
[0.4-0.5]	0.14 ± 0.08	0.033 ± 0.003	0.010 ± 0.003	0.25 ± 0.10	0.015 ± 0.006
[0.5-0.6]	0.15 ± 0.09	0.042 ± 0.004	0.011 ± 0.003	0.20 ± 0.09	0.023 ± 0.009
[0.6-0.7]	0.15 ± 0.03	0.039 ± 0.004	0.013 ± 0.004	0.13 ± 0.07	0.026 ± 0.010
[0.7-0.8]	0.15 ± 0.07	0.040 ± 0.004	0.015 ± 0.004	0.10 ± 0.06	0.027 ± 0.011
[0.8-0.9]	0.16 ± 0.06	0.036 ± 0.003	0.015 ± 0.004	0	0.025 ± 0.010
[0.9-1.0]	0.18 ± 0.03	0.030 ± 0.003	0.013 ± 0.003	0	0.013 ± 0.006

843 **6 Signal classification**

844 **6.1 Track isolations**

845 One of the most powerful tool to distinguish the signal from the combinatorial background
846 exploits the muon *isolation*, that is the property of the muons from $B \rightarrow \mu\mu$ to be relatively
847 isolated from the other tracks of the event compared to the background decay where one or
848 more daughter particles can be produced close to the muon(s). In this section we present
849 the implementation and the performance of new muon isolations developed for $B \rightarrow \mu\mu$. In
850 Sec. 6.1.1 we briefly describe the “cut-based” isolation used in the measurement published
851 in 2013. In Sec. 6.1.2, we describe the new isolation based on long tracks and in Sec. 6.1.3
852 the new isolation based on Velo tracks. Next, in Sec. 6.1.4 we discuss the performance
853 on data.

854 **6.1.1 Cut-based long track isolation used in the 2013 analysis**

855 The muon isolation used in the previous measurement is described in detail in [29]. The
856 isolation of a muon candidate from $B \rightarrow \mu\mu$ is defined as the number of long tracks *close*
857 to the muon, where *close* is specified by a set of rectangular cuts on five variables grouped
858 below, and defined for future reference as set A:

859
860 **set A:**

- 861 • trk_ips: the minimum of the square root of the impact parameter χ^2 of the track
862 with respect to any primary vertex (PV) of the event
- 863 • trk_pvdis: the signed distance between the (track,muon) vertex and the PV
- 864 • trk_svdis: the signed distance between the (track,muon) vertex and the $B \rightarrow \mu\mu$
865 vertex.
- 866 • trk_doca: the DOCA between the track and the muon
- 867 • trk_angle: the angle between the track and the muon
- 868 • $f_c = \frac{|\vec{P}_\mu + \vec{P}_{trk}| \sin(\alpha^{\mu+trk,PV})}{|\vec{P}_\mu + \vec{P}_{trk}| \sin(\alpha^{\mu+trk,PV}) + P_{T_\mu} + P_{T_{trk}}}$, where $\alpha^{\mu+trk,PV}$ is the angle between the sum of
869 the muon and track momenta ($\vec{P}_\mu, \vec{P}_{trk}$) and the direction defined by the the PV and
870 the (track,muon) vertex [29]. $f_c \rightarrow 0$ if the track+muon system originates from the
871 PV.

872 The muon isolation of the $B \rightarrow \mu\mu$ candidate is defined as the sum of the isolations of the
873 two muons. From the physics point of view, the muon isolation defined above is sensitive
874 to the number of reconstructed tracks originating from the same decay tree as the muon
875 candidate. These tracks are labelled as *non-isolating* as opposed to the *isolating* tracks
876 which do not share a *b* quark (or *c* quark) ancestor with the muon [29]. In the following

877 sections we will keep using this notation. The non-isolating tracks are more likely to be
 878 close to the muon candidate compared to the other tracks of the event. If the $B \rightarrow \mu\mu$
 879 candidate is signal, there are no other tracks sharing the same decay tree of the two muons
 880 and the isolation is close to zero. In combinatorial background, where most of the muon
 881 candidates originate from semileptonic b decays, the tracks belonging to the same decay
 882 tree of the muon are likely to be closer to it and consequently on average the isolation
 883 variable has larger values. This is shown in Fig. 21.

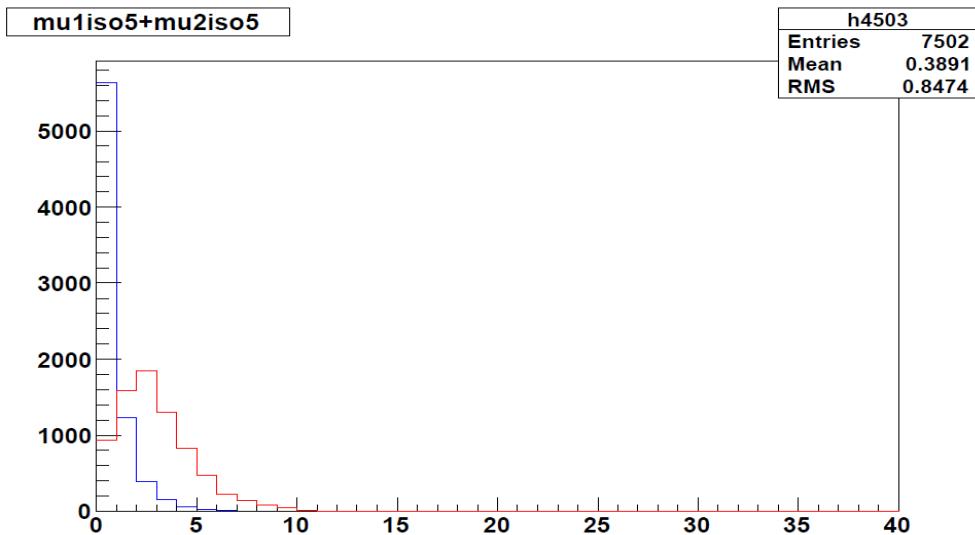


Figure 21: Example of cut-based muon isolation distribution for MC $B_s \rightarrow \mu^+\mu^-$ signal (blue) and simulated $b\bar{b} \rightarrow \mu^+\mu^-X$ background events (red) [29].

884 6.1.2 Development of new long track isolation

885 As an alternative approach, the non-isolating and isolating tracks can be separated using
 886 a multivariate classifier taking as input the variables used to define the cut-based isolation
 887 discussed in Sec. 6.1.1, set A, or additional sets of variables as discussed in the following.
 888 In this section we describe the implementation of a BDT-based muon isolation using
 889 long tracks as input. More details can be found in [13]. In Sec. 6.1.3 we describe the
 890 implementation of a second BDT-based isolation using VELO tracks. Both are used in
 891 the global BDT to classify the events, as discussed in Sec. 6.3.

892 A Boosted Decision Tree classifier⁷ with no variable transformation, is used to maximize
 893 the separation between the categories composed of non-isolating tracks in $b\bar{b} \rightarrow \mu^+\mu^-X$
 894 MC events and isolating tracks in $B_s \rightarrow \mu\mu$ MC events.

⁷Using the *Adaboost* boosting algorithm and developed using the TMVA toolkit.

895 **Discriminating variables** Four sets of variables were considered as input. The first
896 one is “set A” which is defined in Sec. 6.1.1 and includes the variables that were used to
897 compute the cut-based muon isolation of the past analysis. In addition, sets “B”, “C” and
898 “D” defined below have been considered.

899

900 **set B:**

- 901 • trk_gho: track ghost probability
902 • trk_chi: track χ^2

903 **set C:**

- 904 • trk_dphi: difference between the azimuthal angles of track and muon (absolute
905 value)
906 • trk_deta: difference between the pseudorapidities of track and muon (absolute value)
907 • trk_pt: transverse momentum of the track

908 **set D:**

- 909 • trk_ch: relative charge of the track with respect to the muon candidate charge
910 • trk_nnmu, trk_nnpi, trk_nnk: ProbNNmu, ProbNNpi and ProbNNk of the track,
911 respectively

912 The distributions of the variables in set A, B and C are reported in Figs. 104-106 in
913 appendix B. Each plot compares the distribution for non-isolating and isolating tracks
914 in $b\bar{b} \rightarrow \mu^+\mu^-X$ MC, and isolating tracks in $B \rightarrow \mu\mu$ MC. The BDT is trained taking as
915 input the non-isolating tracks of $b\bar{b} \rightarrow \mu^+\mu^-X$ sim06 MC events and isolating tracks of
916 $B_s \rightarrow \mu\mu$ sim06 MC events and is defined so that it takes larger values for tracks more
917 likely to belong to the former category, as shown in Fig. 22.

918 **BDT isolation of the event** The BDT returns an isolation score value for a pair of
919 tracks: the track whose isolation it is computed for and the muon it is computed against.
920 For each track, $iso1(track)$ and $iso2(track)$ correspond to the isolation relative to the μ^+ and
921 μ^- , respectively. Several criteria have been compared to classify the event as signal-like or
922 background-like through a single number computed from the set of iso1 and iso2 values.
923 These include:

- 924 • iso_max2: maximum value of $iso1(track_i)+iso2(track_j)$, i and j are any pair of
925 tracks in the event (not required to be the same track)
926 • iso_ave2: mean of the two largest iso1 plus mean of the two largest iso2
927 • iso_ave3: mean of the three largest iso1 plus mean of the three largest iso2

- 928 • iso_ave4: mean of the four largest iso1 plus mean of the four largest iso2
 929 • iso_max: iso1(track.i)+iso2(track.i) for any track in the event, but the track is
 930 constrained to be the same in iso1 and iso2
 931 • iso_frac_x: fraction of tracks with BDT> x ($x = 0, 0.1, 0.2\dots$).

932 The iso_max2 and iso_ave2 criteria give the best performance in terms of separation
 933 between $B_s \rightarrow \mu\mu$ and combinatorial background (for more details please see [30]). The
 934 iso_max2 criterion has been chosen and from now on we will refer to it when the BDT
 935 long or VELO isolations are discussed.

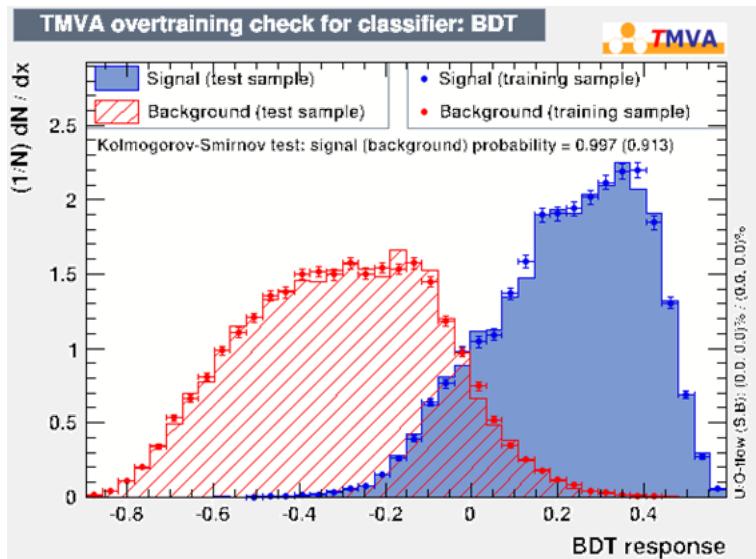


Figure 22: Example of muon isolation BDT training: superposition of training and test samples for “signal” (non-isolating long tracks from $b\bar{b} \rightarrow \mu^+\mu^-X$ MC, blue histograms) and “background” (isolating long tracks from $B_s \rightarrow \mu\mu$ MC, red histograms).

936 **Choice of the input variables** The performance of the track isolation was evaluated
 937 by comparing, by eye, the ROC⁸ for the sets of input variables A, A+B, A+B+C and
 938 A+B+C+D, both in MC and in data. In MC, the ROC was computed using $b\bar{b} \rightarrow \mu^+\mu^-X$
 939 and $B_s \rightarrow \mu\mu$ (sim06) events, while in data the background sample was selected in the
 940 right mass sideband defined as $m(\mu\mu) \in [5447, 6000]$ MeV of stripping 20+20r1. The
 941 standard selection was applied except for the PID requirement, where the requirement
 942 used in the previous analysis was applied. In MC, the performance improves as more
 943 discriminating variables are added, as expected. Hence the best set of input variables is

⁸Receiver Operating Characteristic: background rejection vs signal efficiency

944 A+B+C+D. On the contrary, the performance on data worsens instead of improving when
945 set B is added to A. Furthermore, it does not improve when set D is included. The reason
946 of this behaviour can be explained by noticing that the data/MC agreement is particularly
947 poor for the variables in sets B and D.⁹ Based on these results, A+C has been chosen as
948 set of input variables for the long track isolation. The interested reader can find more
949 details in [13].

950 6.1.3 Development of Velo-only track isolation

951 A BDT-based isolation taking as input the list of “Velo tracks”¹⁰ was developed with
952 the same procedure described in Sec. 6.1.2 for the long track isolation, i.e. with the
953 same definition of non-isolating/isolating tracks and the same criterion to define the event
954 isolation from the BDT output associated to each muon. The set A of input variables¹¹
955 has been adopted.

956 A muon isolation taking as input the upstream tracks has also been developed but the
957 performance was found to be significantly worse than that achieved with the Velo tracks
958 so that eventually it has not been used.

959 6.1.4 Performance on data

960 Figure 23 shows the long track and Velo track isolation distributions for $B_s \rightarrow \mu\mu$ MC
961 events and for candidates selected in the right mass sidebands of Run1 and Run II data.
962 The selection described in Sec. 4 is applied except for the PID requirement on the muon
963 candidates, which is looser (isMuon requirement). Figure 24 shows the ROC of the long
964 track isolation separately in Run I, Run II 2015 and Run II 2016 data, as well as the
965 comparison with the track isolation used in the past analysis. The latter appears as a
966 small set of points as a consequence of the “discrete” nature of the variable (see fig. 21).
967 Figure 25 shows the ROC of the Velo track isolation in Run I and Run II data.

968 The background rejection power of the new long track isolation is significantly higher
969 compared to that of the old isolation. At $\epsilon_{sig} \sim 0.8$ it rejects approximately 40% more
970 background events. On top of this, the continuous nature of the BDT-based isolation, as
971 opposed to the “discrete” nature of the old one, gives an additional important advantage
972 when used in the global BDT, which is discussed in Secs. 6.3 and 6.3.3.

⁹An attempt was made to train the BDT isolation after replacing the distributions of set B variables in the training samples with distributions extracted from data control samples, but no significant improvement was found.

¹⁰Track segments composed of Velo hits and not used to form the long tracks.

¹¹Since the P_T of Velo tracks is not measured, it is set to 400 MeV in the computation of f_c . This value is not far from the mean of the long track P_T distribution, which is about 600 MeV.

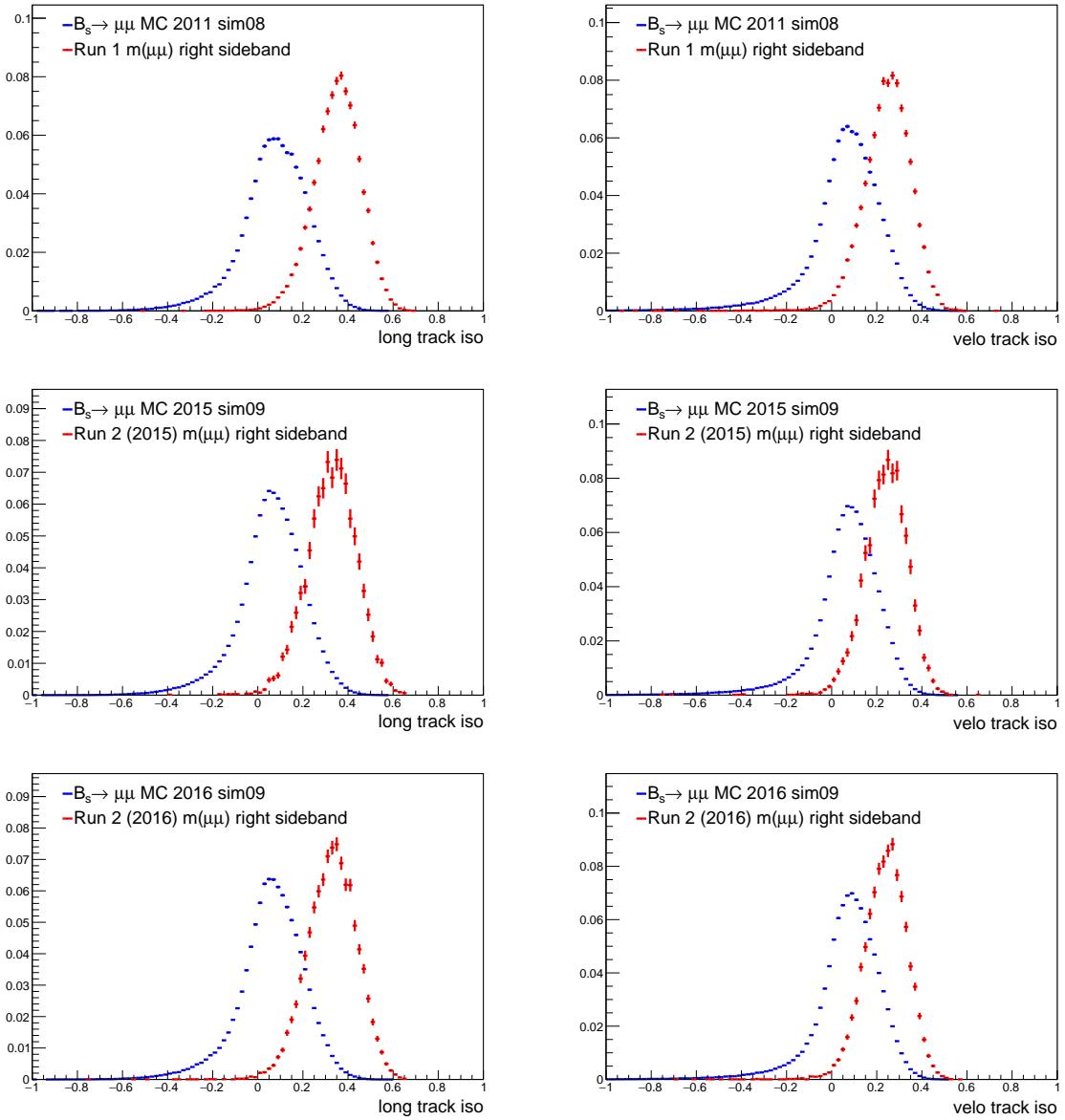


Figure 23: Top row: long (left) and VELO (right) track isolation distribution of $B_s \rightarrow \mu\mu$ MC events and of candidates in the right mass data sideband of Run I. Middle row: same for Run II 2015. Bottom row: same for Run II 2016.

6.2 Investigation of other isolation variables

Two additional isolation tools have been evaluated. One exploits the *ZVTop* algorithm based on the vertex information between the tracks of the event [31], the other exploits the reconstruction of jets [32]. Although both isolations have shown a good discriminating power against the combinatorial background, their use in the global BDT has not shown

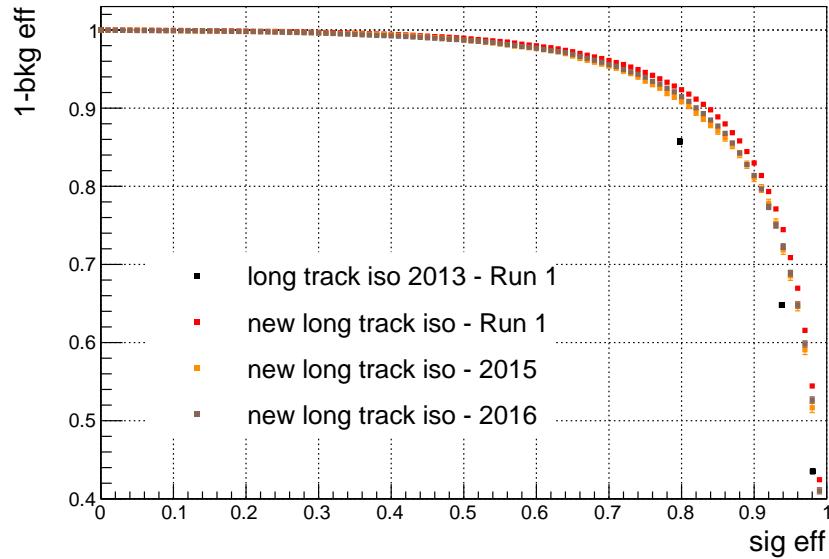


Figure 24: ROC of the new long track isolation in Run1 data (red), Run II 2015 data (orange), Run II 2016 data (brown) and ROC of the old track isolation used in the previous analysis (black).

978 a statistically significant improvement of the overall performance. This is due to the
 979 correlation with other variables entering the BDT, in particular – but not only – the track
 980 isolation discussed in the previous sections. Therefore, in the end they have not been
 981 included in the global BDT.

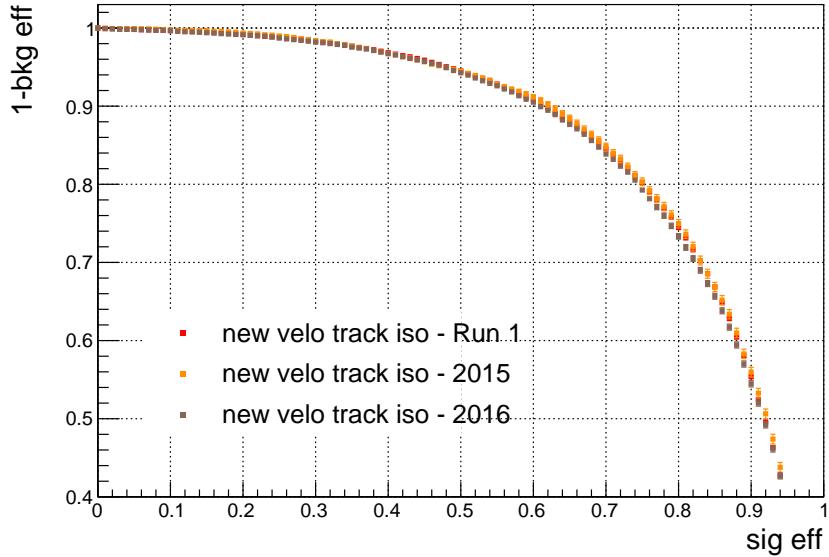


Figure 25: ROC of the new VELO track isolation in Run1 data (red), Run II 2015 data (orange) and Run II 2016 data (brown).

6.3 Global BDT definition

We use a multivariate classifier, a BDT, to separate the $B_{(s)}^0 \rightarrow \mu\mu$ signal from the combinatorial background, which is mostly composed of muons originating from different b quark decays. In this section we discuss the procedure used to optimize the background rejection power of the BDT. The resulting performance in data is discussed in Sec. 6.3.3.

The BDT was trained using $b\bar{b} \rightarrow \mu\mu X$ and $B_s \rightarrow \mu\mu$ simulated events (sim06) with the standard selection applied. The BDT variable is subsequently *flattened* to range between 0 and 1 such that the signal is uniformly distributed while the background peaks at zero. This is a simple transformation which does not change the signal/background separation but is convenient to easily split the BDT range in subregions of known signal efficiency. In the following we will usually refer to the flattened BDT unless otherwise specified. In the 2013 analysis the range was split in 8 subregion: [0,0.25], [0.25,0.4], [0.4,0.5], [0.5,0.6], [0.6,0.7], [0.7,0.8], [0.8,0.9], [0.9,1.0] and a combined fit to the events in each subregion was performed. In the optimization procedure discussed below the same binning is used.

6.3.1 Procedure to identify the best discriminating variables

We start with the long track isolation, which is the most discriminating variable.

1 Put the selected variable in the list of BDT input variables

1000 2 Add (temporarily) one of the other variables to the input list
 1001 3 Train the BDT and measure the background yield for $b\bar{b} \rightarrow \mu\mu X$ MC in the most
 1002 sensitive bin. If the yield is zero, gets the yield in the next lower bin (and so on till
 1003 a yield > 0 is found)
 1004 4 Remove the variable from the input list and repeat from 2 till all the variables have
 1005 been tested
 1006 5 Select the variable which gave the smallest background yield in 2-4 and repeat from 1
 1007 6 The process stops when no additional variables are found that, added to the BDT
 1008 input list, can further decrease the yield in the most sensitive bin.
 1009 This procedure allows to easily test and compare a very large number of discriminating
 1010 variables. However, care is needed when interpreting the results since purely down
 1011 statistical fluctuations of the yields need to be pruned away in the process of identifying
 1012 the best variables. In other words, the above procedure is a powerful investigation tool
 1013 but the final choice of variables is made by hand.
 1014 A total of seven discriminating variables are selected:
 1015 • Long track isolation (see Sec. 6.1)
 1016 • mu_DeltaR: $\sqrt{\Delta\phi^2 + \Delta\eta^2}$, where $\Delta\phi$ and $\Delta\eta$ are the azimuthal angle and pseudora-
 1017 pidity differences between the two muons.
 1018 • mu_MINIPS: smallest value among the muon impact parameter significance of the
 1019 two muons with respect to the primary vertex associated to the $B \rightarrow \mu\mu$ candidate.
 1020 • B_ACOSDIRA_OWNPV: angle between the B direction and the vector joining the
 1021 primary and secondary vertex positions.
 1022 • B_ENDVERTEX_CHI2: vertex χ^2 of the B candidate
 1023 • B_IPS_OWNPV: impact parameter significance of the B candidate with respect to
 1024 the primary vertex.
 1025 • VELO track isolation (see Sec. 6.1)
 1026 After the choice of the input variables the BDT parameters have been scanned to
 1027 search for a set giving a significantly better performance, but no improvement was found
 1028 compared to the set used during the variable choice procedure. The BDT parameter values
 1029 are: NTrees=1000, MinNodeSize=1%, MaxDepth=3, nCuts=30, AdaBoostBeta=0.75
 1030 (BoostType=AdaBoost).

1031 **6.3.2 BDT characterization and comparison with BDT 2013**

1032 The BDT used in the previous analysis takes as input 12 variables, which are listed
 1033 in Tab. 33. Their definition is reported in appendix C. No significant improvement in
 1034 performance has been observed when one or more of the BDT 2013 input variables are
 1035 added to the 7 input vars of the new BDT.

Var category	new BDT	BDT 2013
Isolation	BDT long track iso BDT VELO track iso	Cut-based long track iso B_yet_another_CDF_iso
$\mu^+ \mu^-$ angle	mu_DeltaR	mu_DeltaEta mu_AbsPhi
Correlated to PV-SV distance	mu_MINIPS B_ACOSDIRA_OWNPV B_IPS_OWNPV	mu_MINIPS B_TAU_ps B_IP_OWNPV
Others	B_ENDVERTEX_CHI2	B_doca B_PT B_cosnk B_otherB_ang B_otherB_boo_ang

Table 33: Input variables of the new BDT and the BDT used in 2013.

1036 The correlation between BDT and $m(\mu\mu)$ in simulated $b\bar{b} \rightarrow \mu\mu X$ events and in the
 1037 right mass data sidebands ranges between -3% and -4% (Fig. 26), close to the -2-3%
 1038 correlation found with the 2013 selection and BDT. The top-left plot of Fig. 26 shows that
 1039 the BDT average as a function of $m(\mu\mu)$ does not have any peaking structure in the signal
 1040 mass region. The correlation among the input variables of the new BDT in $B_s \rightarrow \mu\mu$ MC
 1041 signal events and in $b\bar{b} \rightarrow \mu\mu X$ events is reported in Fig. 27.

1042 Figure 28 compares the BDT input variable distributions in $B_s \rightarrow \mu\mu$ MC events (2011,
 1043 2015 and 2016 conditions) and in the right mass data sidebands (Run I, Run II 2015 and
 1044 Run II 2016). Figure 29 shows the resulting BDT distribution before the flattening is
 1045 applied. Figures 107 and 108 in appendix D show the distributions of the BDT and its
 1046 input variables in the MC sample used to train the BDT ($B_s \rightarrow \mu\mu$ MC12 sim06 and
 1047 $b\bar{b} \rightarrow \mu\mu X$ MC sim06) as well as in the right mass data sideband (Run1) and in a more
 1048 recent version of $B_s \rightarrow \mu\mu$ MC (sim08). The possibility of training the BDT using events
 1049 in the mass data sidebands as opposed to $b\bar{b} \rightarrow \mu\mu X$ MC events was investigated, but no
 1050 clear evidence of improvement was found. This is dicussed in appendix E.

1051 **6.3.3 Performance in Run I and Run II data**

1052 Figure 30 shows the ROC of the new BDT in Run1 data and compares it with the ROC of
 1053 the BDT used in the 2013 analysis. The new BDT rejects approximately 50% additional
 1054 background compared to the old one. This improvement is observed in all BDT bins, as

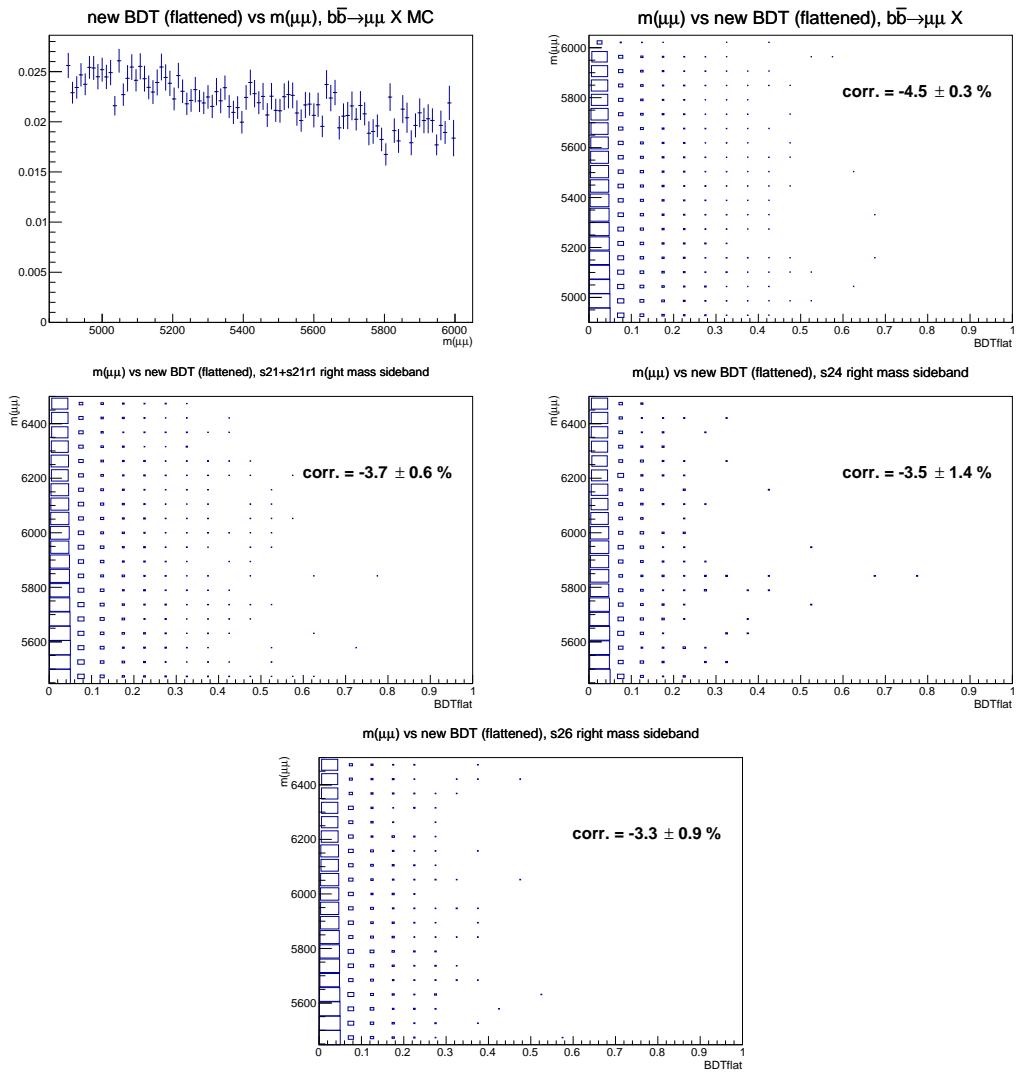


Figure 26: Top row: average value of the BDT output as a function of $m(\mu\mu)$ (left) and $m(\mu\mu)$ as a function of BDT (right) in $b\bar{b} \rightarrow \mu\mu X$ MC events. Middle row: $m(\mu\mu)$ as a function of BDT in Run1 (left) and Run II 2015 (right) right mass data sideband. Bottom row: $m(\mu\mu)$ as a function of BDT in Run II 2016 right mass data sideband. The BDT variable is flattened to range in [0,1] (see sec. 6.3).

discussed below. Figure 31 compares the new BDT ROCs in Run 1 and Run 2 data. The signal/background discriminating power in the selected 2016 data looks slightly higher than in Run 1 data.

It is worth to mention that the procedure to develop and optimize the new BDT has been entirely done on simulated events, therefore no possible biases are introduced in data.

Tables 34-36 show the combinatorial background yields in the right mass data sidebands

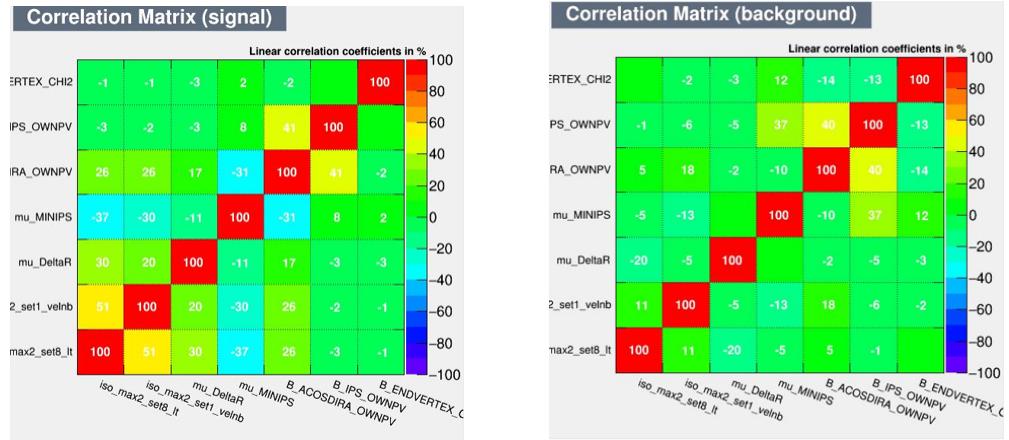


Figure 27: Correlation between the input variables of the new BDT in $B_s \rightarrow \mu\mu$ MC events (left) and simulated $b\bar{b} \rightarrow \mu\mu X$ events (right). $iso_max2_set8_lt$ is the long track isolation, $iso_max2_set1_velnb$ is the VELO track isolation.

as a function of the BDT bin. The BDT flattening is performed using $B_s \rightarrow \mu\mu$ 2011 sim08 MC for stripping 21 data, and $B_s \rightarrow \mu\mu$ 2015 sim09 MC for stripping 24 (2015) and stripping 26 (2016) data. Each table shows the yields for different analysis configurations: “old BDT + DLL” corresponds to the 2013 analysis, where a DLL-based requirement was applied to muons. “new BDT + DLL” corresponds to using the old analysis selection with the new BDT. Finally, “new BDT + ProbNN” corresponds to the current BF analysis where the new BDT and a tighter ProbNN-based cut is used (see sec. 4.3).

Let us consider the Run I data (tab. 34). Based on bins which are sufficiently populated, between 0.25 and 0.7, the new BDT removes about 50% of background compared to the one used in 2013, as also seen in Fig. 30. Passing from the DLL cut to the ProbNN cut removes approximately another 20%. Overall, the combinatorial background in the new BF analysis is about 40% the one in the previous analysis. If we compare the $BDT > 0.5$ region we pass from 64 to 16 events, to be compared with the “expected” $64 \times 0.4 = 26$ events. This can be a down fluctuation or a possible hint that the background suppression in the sensitive region might be even better than 60%. There is no way to draw a conclusion.

Comparing the yields in 2015 data (tab. 35) with those in Run I (tab. 34) one can see that the ratio of the total yields is about 0.15. After normalizing for the corresponding integrated luminosity one gets a ratio of about 1.5, which is not far from the 1.6 value which has been measured using the $B^+ \rightarrow J/\psi K^+$ and $B_s \rightarrow J/\psi \phi$ control samples. The yields observed in 2015 data in each BDT bin are consistent with those observed in Run 1 multiplied by 0.15. Similar considerations apply to the 2016 data (tab 36) compared to Run 1. In this case the ratio of the total yields is 0.37, which becomes about 1.6 after normalizing for the integrated luminosities. The yields observed in 2016 data for $BDT > 0.25$ are consistently below those observed in Run 1 multiplied by 0.37.

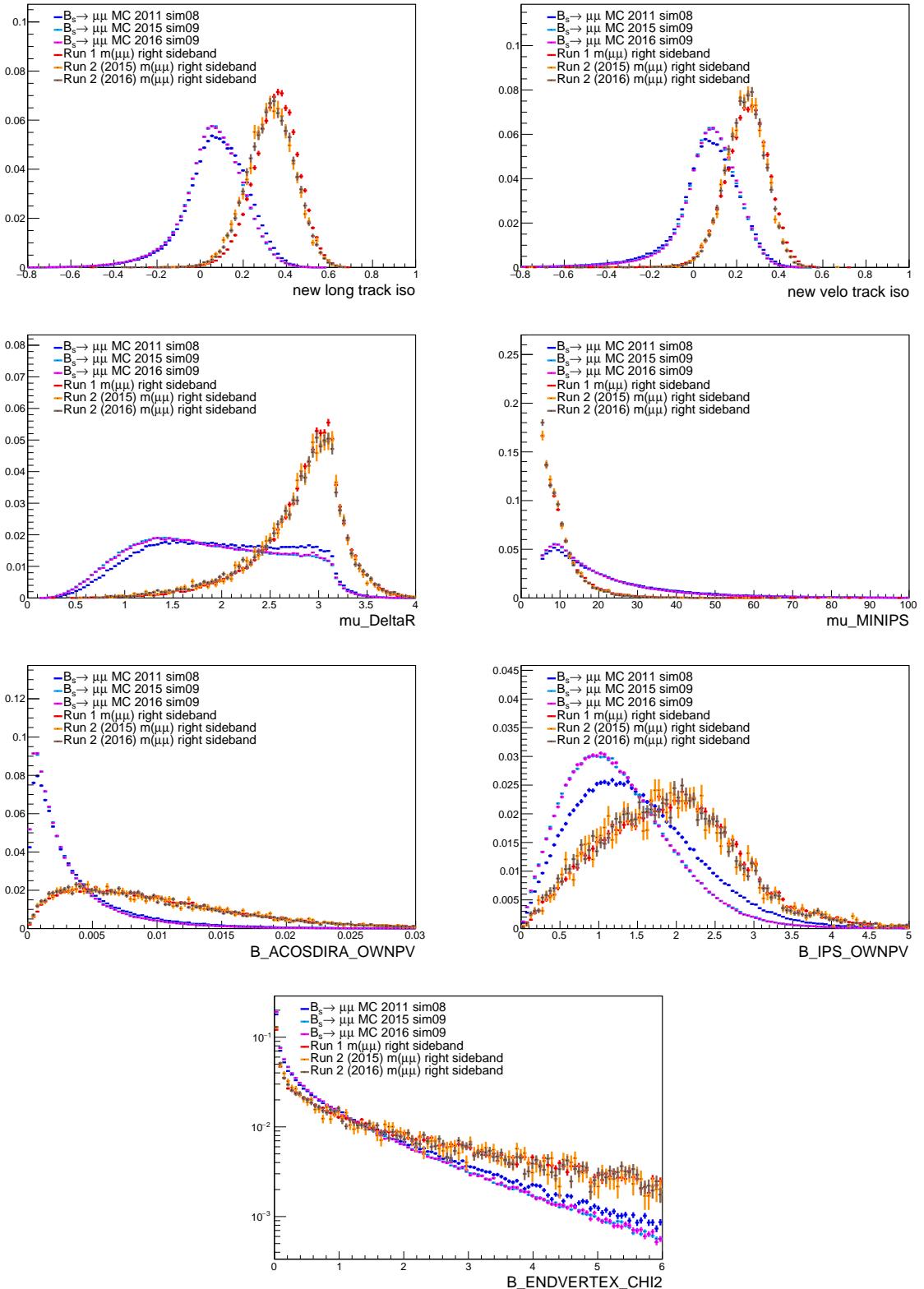


Figure 28: Distribution of the BDT input variables in $B_s \rightarrow \mu\mu$ MC events (2011, 2015 and 2016 conditions) and in the right mass data sideband (Run I, Run II 2015 and Run II 2016).

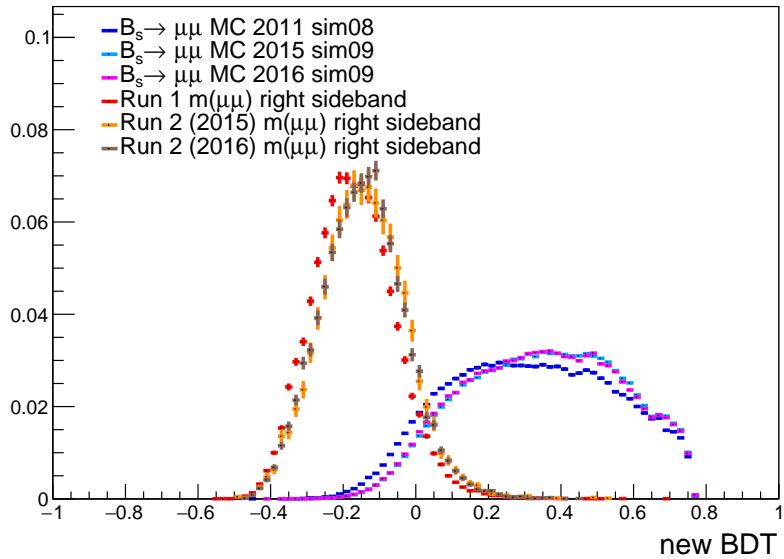


Figure 29: BDT distribution in $B_s \rightarrow \mu\mu$ MC events (2011, 2015 and 2016 conditions) and in the right mass data sidebands (Run I, Run II 2015 and Run II 2016).

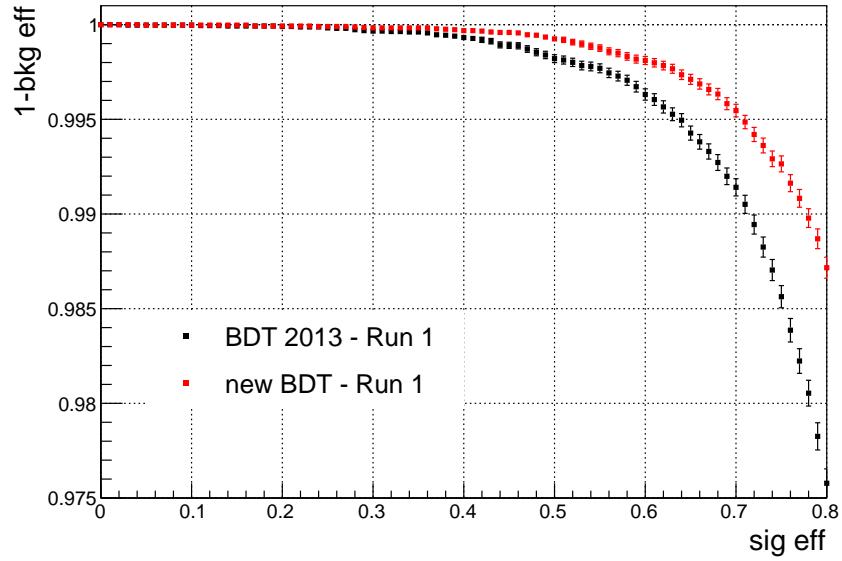


Figure 30: ROC of the new BDT in Run1 data and comparison with the ROC of the BDT used in the past analysis.

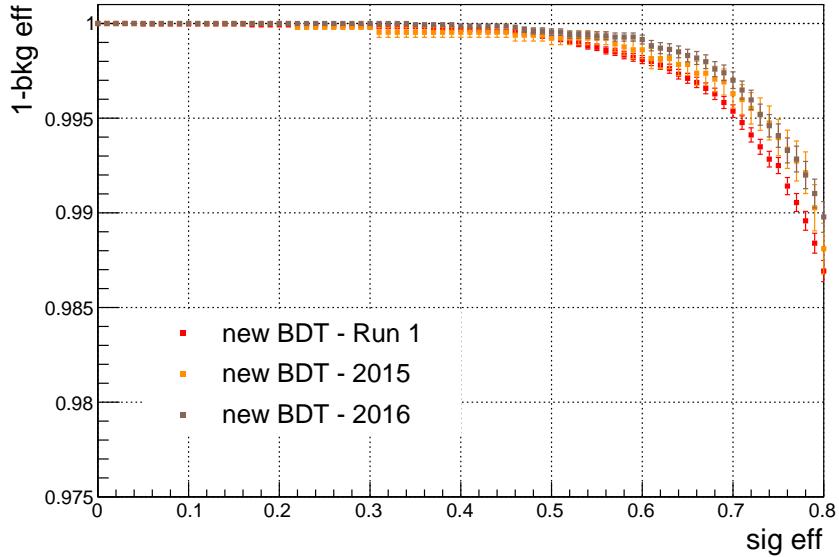


Figure 31: ROC of the new BDT in Run I, Run II 2015 and Run II 2016 data.

Selection	0.0-0.25	0.25-0.4	0.4-0.5	0.5-0.6	0.6-0.7	0.7-0.8	0.8-0.9	0.9-1.0
old BDT + DLL	37442	403	76	41	11	9	3	0
new BDT + DLL	37701	213	46	16	4	3	2	0
new BDT + ProbNN	30631	172	34	13	3	2	0	0

Table 34: Number of events as a function of the BDT bin in stripping 21+21r1 data (Run I) in the mass region $m(\mu\mu) \in [5447, 6500]$, for three selections: BDT and PID selection used in 2013 (“old BDT + DLL”), new BDT and PID selection used in 2013 (“new BDT + ProbNN”), new BDT and PID selection used in this BF analysis (“new BDT + ProbNN”).

1085 6.4 BDT calibration

1086 While the BDT classifier is trained using Monte Carlo simulated events ($B_s^0 \rightarrow \mu^+\mu^-$ for
 1087 signal and $b\bar{b} \rightarrow \mu^+\mu^- X$ for background), its PDF for signal and background is evaluated
 1088 using the data. Since BDT is designed to only use kinematic information of a two-body
 1089 neutral B meson decay, to calibrate our classifier we rely on the $B_{(s)}^0 \rightarrow h^+h^{(\prime)-}$ that act as
 1090 a proxy for the $B_s^0 \rightarrow \mu^+\mu^-$ decay. However, given the branching fractions of the neutral
 1091 B mesons into two hadrons, only the most frequent $B^0 \rightarrow K^+\pi^-$ process is considered
 1092 in the following. In order to distinguish this channel from the others and therefore to
 1093 identify the correct mass hypothesis of the final state particles a cut on the $\Delta LL_{K-\pi}$
 1094 variable of the final state particles is applied. The effect of misidentified components in
 1095 the $B^0 \rightarrow K^+\pi^-$ yield per BDT bin is evaluated varying this cut. Moreover, the cut on
 1096 the $\Delta LL_{K-\pi}$ variable can induce a bias as its separation power is highly correlated with

Selection	0.0-0.25	0.25-0.4	0.4-0.5	0.5-0.6	0.6-0.7	0.7-0.8	0.8-0.9	0.9-1.0
old BDT + DLL	5821	73	9	3	3	2	0	0
new BDT + DLL	5877	26	4	2	1	1	0	0
new BDT + ProbNN	4720	22	3	2	1	1	0	0

Table 35: Number of events as a function of the BDT bin in stripping 24 data (Run II, 2015) in the mass region $m(\mu\mu) \in [5447, 6500]$, for three selections: BDT and PID selection used in 2013 (“old BDT + DLL”), new BDT and PID selection used in 2013 (“new BDT + ProbNN”), new BDT and PID selection used in this BF analysis (“new BDT + ProbNN”).

Selection	0.0-0.25	0.25-0.4	0.4-0.5	0.5-0.6	0.6-0.7	0.7-0.8	0.8-0.9	0.9-1.0
old BDT + DLL	13927	108	18	13	1	1	1	0
new BDT + DLL	14000	63	5	1	0	0	0	0
new BDT + ProbNN	11392	48	3	1	0	0	0	0

Table 36: Number of events as a function of the BDT bin in stripping 26 data (Run II 2016, $\sim 0.7 \text{ fb}^{-1}$) in the mass region $m(\mu\mu) \in [5447, 6500]$, for three selections: BDT and PID selection used in 2013 (“old BDT + DLL”), new BDT and PID selection used in 2013 (“new BDT + ProbNN”), new BDT and PID selection used in this BF analysis (“new BDT + ProbNN”).

the kinematics of the final state particle and the multiplicity of the event. To reduce this effect the efficiency of the $\Delta LL_{K-\pi}$ cut is evaluated as a function of particle momenta, the pseudo rapidities and the number of best tracks in the event.

6.4.1 $\Delta LL_{K-\pi}$ cut efficiency determination

The efficiency is estimated using the PIDCalibTool provided by PID group, separately per Stripping, polarity and year of data taking. We use $D^{*\pm} \rightarrow (D^0 \rightarrow K^\pm \pi^\mp)\pi^\pm$ events to calculate the efficiency as a function of the momentum of the final state hadron, p , its pseudo rapidity, η , and number of best tracks, N_{tracks} . The binning scheme used throughout this analysis to determine the efficiency of the $\Delta LL_{K-\pi}$ cuts is:

- p : 2 bins for $0 \text{ GeV}/c < p < 10 \text{ GeV}/c$; 45 bins for $10 \text{ GeV}/c < p < 100 \text{ GeV}/c$; 20 bins for $100 \text{ GeV}/c < p < 150 \text{ GeV}/c$; 4 bins for $150 \text{ GeV}/c < p < 500 \text{ GeV}/c$ (71 bins);
- η [1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 5.5, 6.0] (10 bins);
- N_{tracks} [0, 100, 200, 300, 400, 600] (5 bins).

For systematic checks, the binning from the previous analysis was used to determine the efficiency as well:

- p [5, 9.3, 15.6, 17.675, 20, 23, 26, 29.65, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 125, 150, 200] GeV/c (25 bins);

- 1115 • η [1.5, 2.5, 3.0, 3.5, 4.0, 5.0] (5 bins);
 1116 • N_{tracks} [0, 50, 100, 200, 400, 800] (5 bins).

1117 The strange boundary values for the binning in p are reflecting the radiator thresholds in
 1118 the two RICH detectors.

1119 In order to determine the PID efficiency for $B^0 \rightarrow K^+\pi^-$, the three dimensional distri-
 1120 bution in $p, \eta, N_{\text{tracks}}$ is determined from sWeighted $B_{(s)}^0 \rightarrow h^+h^-$ kinematic distributions.
 1121 The $\Lambda_b^0 \rightarrow ph$ modes are not included in the $B_{(s)}^0 \rightarrow h^+h^-$ signal, and are separated in
 1122 mass from the $B_{(s)}^0 \rightarrow h^+h^-$ modes. In the previous version of the note simulated events
 1123 were used for the phase-space, but since these are not correctly reproducing the true data
 1124 distribution, an alternative strategy is used and described in this section. The kinematic
 1125 distributions are now obtained by fitting the inclusive $B_{(s)}^0 \rightarrow h^+h^-$ sample per BDT bin
 1126 in the $\pi\pi$ mass hypothesis without any PID selection. This procedure was already used
 1127 in the $B^0 \rightarrow K^+K^-$ search [33] and slightly modified in order to take into account the
 1128 different data selection. Since the sWeighted kinematic distributions are used to evaluate
 1129 the PID efficiency for the $B^0 \rightarrow K^+\pi^-$ sample, the same trigger requirements are applied
 1130 (L0Global_TIS & HLT1Phys_TIS & HLT2B2HH_TOS). Template models evaluated from
 1131 simulated events are used for neutral B meson and Λ_b baryon two body decays and smeared
 1132 with a gaussian resolution model to take into account the discrepancy between real data
 1133 and simulated events. The relative yield of each $B_{(s)}^0 \rightarrow h^+h^-$ decay is fixed according
 1134 to their branching fractions and production fractions. The same is done for $\Lambda_b^0 \rightarrow ph$
 1135 backgrounds. For the combinatorial background, an exponential shape is used, while for
 1136 the $B \rightarrow 3 - \text{body}$ decays an Argus shape convoluted with a gaussian pdf is implemented.

1137 Given the high contamination of combinatorial background in the first BDT bin, it is
 1138 split into two bins to improve the purity for the sWeights:

- 1139 • **Bin 1a:** BDT $\in [0.0, 0.1)$ with $\sim 96\%$ of the total combinatorial background;
 1140 • **Bin 1b:** BDT $\in [0.1, 0.25)$ with $\sim 3\%$ of the total combinatorial background;

1141 The fit performed on the $B_{(s)}^0 \rightarrow h^+h^-$ mass distribution in Bin 1a is unstable and does
 1142 not give reliable sWeights. For just this bin the kinematic distributions are taken from
 1143 simulated events. Following this procedure, the impact of any large discrepancy between
 1144 data and MC will be diluted by a factor 1/10.

1145 The fit result for the inclusive $B_{(s)}^0 \rightarrow h^+h^-$ samples of 2011, 2012, 2015 and 2016 are
 1146 reported in Figs. 32, 33, 34, 35. The $B_{(s)}^0 \rightarrow h^+h^-$ yields are reported in Tabs. 38, 39, 40, 41.
 1147 These are compared with the yields from the exclusive $B^0 \rightarrow K^+\pi^-$ fit, scaled by the
 1148 relative fraction. The relative fractions of $H_b \rightarrow h^+h^-$ decays are shown in Table 37. The
 1149 difference is assigned as a systematic for the BDT calibration per BDT bin and for the
 1150 normalisation on the total yield.

1151 The PID efficiencies are then obtained per BDT bin and $\Delta LL_{K-\pi}$ cut convolving the
 1152 phase-space obtained as described above with PIDcalib efficiency histograms, and are used
 1153 to correct the obtained yields before calculating the fraction of events per BDT bin.

Table 37: Relative fraction of different $H_b \rightarrow h^+h^-$ modes. For the inclusive $B_{(s)}^0 \rightarrow h^+h^-$ fit, all fractions are fixed. The $B^0 \rightarrow K^+\pi^-/B_{(s)}^0 \rightarrow h^+h^-$ fraction is used to compare the inclusive with the exclusive fit result and assign a systematic.

Decay mode	$B_{(s)}^0 \rightarrow h^+h^-$ fraction	$H_b \rightarrow h^+h^-$ fraction
$H_b \rightarrow h^+h^-$	-	1.000 ± 0.000
$B_{(s)}^0 \rightarrow h^+h^-$	1.000 ± 0.000	0.935 ± 0.012
$B^0 \rightarrow \pi^+\pi^-$	0.156 ± 0.008	0.146 ± 0.008
$B^0 \rightarrow K^+\pi^-$	0.599 ± 0.009	0.560 ± 0.010
$B^0 \rightarrow K^+K^-$	0.0024 ± 0.0005	0.0022 ± 0.0004
$B_s^0 \rightarrow K^+K^-$	0.0055 ± 0.0007	0.0052 ± 0.0006
$B_s^0 \rightarrow K^+\pi^-$	0.044 ± 0.004	0.041 ± 0.004
$B_s^0 \rightarrow \pi^+\pi^-$	0.192 ± 0.008	0.180 ± 0.008
$\Lambda_b^0 \rightarrow p\pi^-$	-	0.025 ± 0.007
$\Lambda_b^0 \rightarrow pK^-$	-	0.040 ± 0.010

Table 38: Yields per BDT bin from 2011 $B_{(s)}^0 \rightarrow h^+h^-$ fit

BDT bin	$B_{(s)}^0 \rightarrow h^+h^-$ fit yield
[0.10, 0.25]	1718 ± 80
[0.25, 0.40]	1737 ± 68
[0.40, 0.50]	1144 ± 56
[0.50, 0.60]	1231 ± 47
[0.60, 0.70]	1261 ± 45
[0.70, 0.80]	1268 ± 53
[0.80, 0.90]	1531 ± 45
[0.90, 1.00]	1658 ± 41

6.4.2 BDT calibration with $B^0 \rightarrow K^+\pi^-$

The sample used to calibrate the BDT is the $B^0 \rightarrow K^+\pi^-$ sample defined in 4.1 with the following additional cuts applied:

- both hadrons in the Muon acceptance;
- L0Global TIS;
- Hlt1Physics TIS;
- Hlt2B2HHDecision TOS;

Currently, the BDT calibration is performed separately for Run I (2011+2012), 2015 and 2016 data. The total sample is divided into the eight BDT bins, with bounds given by [0,0.25,0.4,0.5,0.6,0.7,0.8,0.9,1.0]. As already mentioned in the previous subsection,

Table 39: Yields per BDT bin from 2012 $B_{(s)}^0 \rightarrow h^+h^-$ fit

BDT bin	$B_{(s)}^0 \rightarrow h^+h^-$ fit yield
[0.10, 0.25]	4824 ± 186
[0.25, 0.40]	4607 ± 189
[0.40, 0.50]	2997 ± 94
[0.50, 0.60]	3346 ± 81
[0.60, 0.70]	3222 ± 96
[0.70, 0.80]	3296 ± 83
[0.80, 0.90]	3874 ± 68
[0.90, 1.00]	4059 ± 69

Table 40: Yields per BDT bin from 2015 $B_{(s)}^0 \rightarrow h^+h^-$ fit

BDT bin	$B_{(s)}^0 \rightarrow h^+h^-$ fit yield
[0.10, 0.25]	1921 ± 85
[0.25, 0.40]	2186 ± 81
[0.40, 0.50]	1383 ± 69
[0.50, 0.60]	1414 ± 64
[0.60, 0.70]	1497 ± 44
[0.70, 0.80]	1325 ± 47
[0.80, 0.90]	1604 ± 49
[0.90, 1.00]	1788 ± 44

the first BDT bin is treated as a special case and splitted into two sub bins. The yield for bin 1a is obtained by fitting the full $B^0 \rightarrow K^+\pi^-$ mass range and subtracting the yield obtained in all the other bins including bin 1b. The goal of the BDT calibration is to determine the fraction of $B^0 \rightarrow K^+\pi^-$ events for each BDT bin through a fit of the invariant mass of the two hadrons $m(K\pi)$. The correct mass hypothesis for each hadron is chosen by cutting on the $\Delta LL_{K-\pi}$ value. If $\Delta LL_{K-\pi} > \kappa$, with κ positively defined, the hadron is identified as kaon otherwise if $\Delta LL_{K-\pi} < -\kappa$ the hadron is identified as a pion. All the events with $|\Delta LL_{K-\pi}| < \kappa$ are rejected. In order to cut away the partially reconstructed background, the mass window is chosen to run from 5200 to 5850 MeV/ c^2 . The upper bound is chosen taking into account the invariant mass cut in the stripping selection.

The stability of the result under different $\Delta LL_{K-\pi}$ cuts is investigated from $\kappa = 5$ to $\kappa = 10$ in steps of 0.5. To determine the number of events per bin, a maximum likelihood fit is performed.

For the full BDT range and the first BDT bin, this fit is binned due to the large number of events. For all other BDT bins excluding bin 1a, a maximum unbinned likelihood fit is performed to obtained the maximal amount of information from the data.

Table 41: Yields per BDT bin from 2016 $B_{(s)}^0 \rightarrow h^+h^-$ fit

BDT bin	$B_{(s)}^0 \rightarrow h^+h^-$ fit yield
[0.10, 0.25]	7384 ± 230
[0.25, 0.40]	7590 ± 155
[0.40, 0.50]	4906 ± 118
[0.50, 0.60]	4914 ± 97
[0.60, 0.70]	4547 ± 80
[0.70, 0.80]	4678 ± 89
[0.80, 0.90]	5227 ± 76
[0.90, 1.00]	5744 ± 76

The resulting invariant mass distributions are fitted with a Double Sided Crystal Ball function for the B^0 and B_s^0 signal peaks, with the tail parameters constrained from $B^0 \rightarrow K^+\pi^-$ MC. As the B_s^0 yield is small compared to fluctuations of the combinatorial background, the B_s^0 mean is given by the B^0 mean plus the mass difference between the B^0 and B_s^0 mass in the PDG, while the $B_s^0 \sigma$ is given by the B^0 sigma times the factor determined in the mass calibration (see Sect. 6.5). For the combinatorial background, an exponential function is used, while the description of the $\Lambda_b^0 \rightarrow ph$ background is the same used in the previous analysis [6]. The $B^0 \rightarrow \pi^+\pi^-$ and $B_s^0 \rightarrow K^+K^-$ components are found negligible for $\kappa = 5$, therefore excluded from the fit model. In Figs. 36, 37, 38, and 39 the fit results for $\kappa = 5$ are shown for 2011, 2012, 2015, and 2016 data respectively.

From the fit, a number of B^0 events per bin is obtained. This number of events is then corrected for the cut on $\Delta LL_{K-\pi}$ to obtain a PID-independent quantity. The number of B^0 events along with the PID efficiency are evaluated as described before and the corrected yield are reported in Tabs. 42, 43, 44, 45.

Table 42: Yields per BDT bin from 2011 $B^0 \rightarrow K^+\pi^-$ fit

BDT bin	$B^0 \rightarrow K^+\pi^-$ fit yield	PID efficiency	corrected yield
[0.00, 0.10]	240 ± 89	0.6161 ± 0.0385	390 ± 144
[0.10, 0.25]	504 ± 30	0.5860 ± 0.0014	860 ± 52
[0.25, 0.40]	531 ± 27	0.5645 ± 0.0001	940 ± 49
[0.40, 0.50]	351 ± 26	0.5490 ± 0.0001	639 ± 47
[0.50, 0.60]	355 ± 19	0.5075 ± 0.0001	700 ± 38
[0.60, 0.70]	309 ± 18	0.4486 ± 0.0002	688 ± 39
[0.70, 0.80]	320 ± 18	0.4253 ± 0.0001	753 ± 42
[0.80, 0.90]	326 ± 18	0.3846 ± 0.0001	848 ± 47
[0.90, 1.00]	371 ± 19	0.3655 ± 0.0001	1015 ± 53

The fraction of events for the i-th BDT bin and PID cut κ is then determined with:

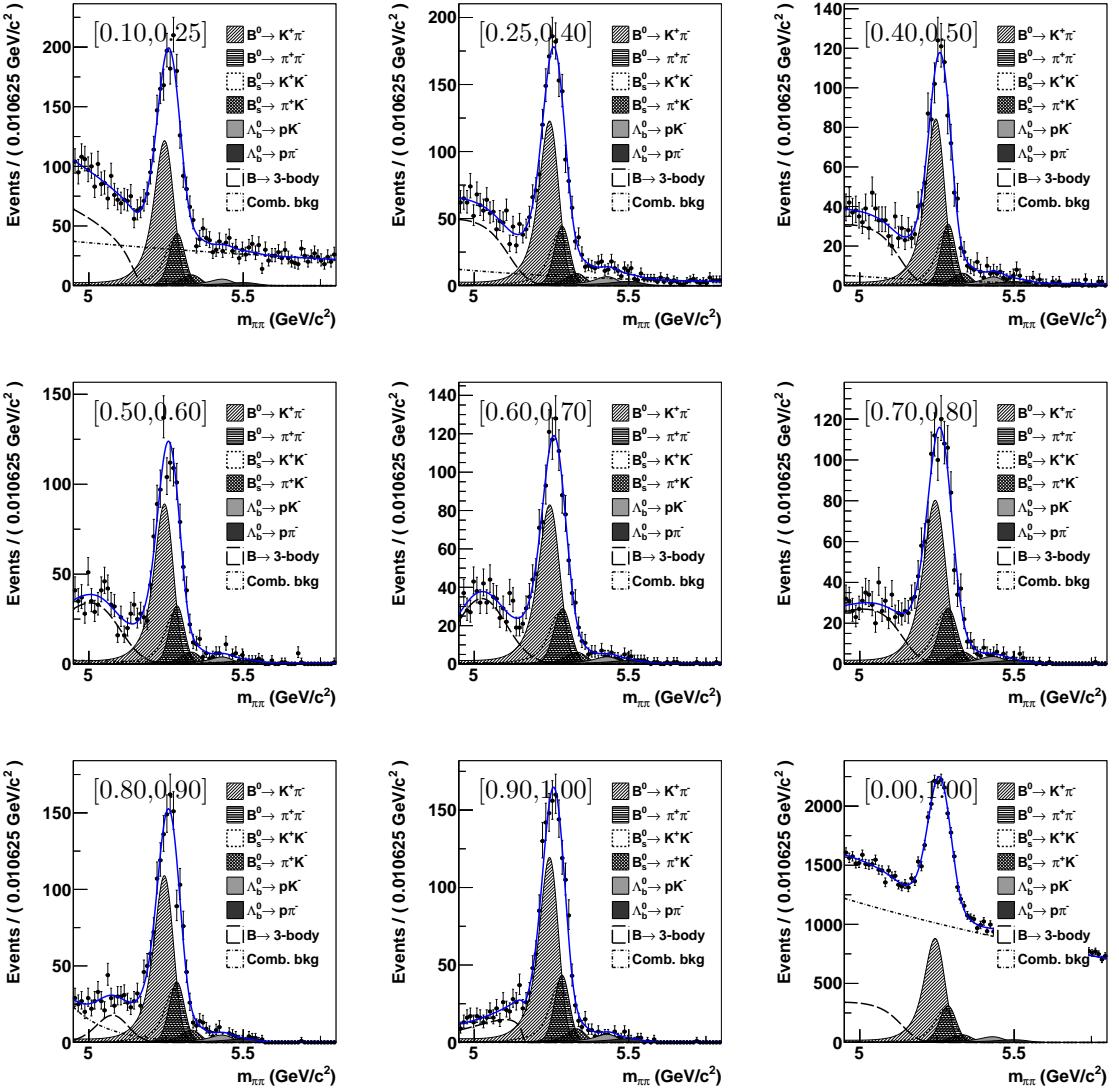


Figure 32: Invariant mass distributions of $B_{(s)}^0 \rightarrow h^+ h^-$ from 2011 data combined without any PID cut. The fit result is superimposed.

$$r_i(\kappa) = N_{B^0,i}(\kappa) / \sum_j (N_{B^0,j}(\kappa)) \quad (17)$$

where $N_{B_d,i}(\kappa)$ is the number of events after dividing by the PID efficiency for i-th BDT bin and PID cut κ . In Fig. 40 the fraction of events per BDT bin as a function of PID cut is shown for Run I (2011+2012), 2015 and 2016 data. The fraction of events is found stable over the range of $\Delta LL_{K-\pi}$ cuts considered.

To take into account all of the information from the fits at different PID cuts, a weighted average \bar{r}_i is determined over the whole PID cut range, taking into account the

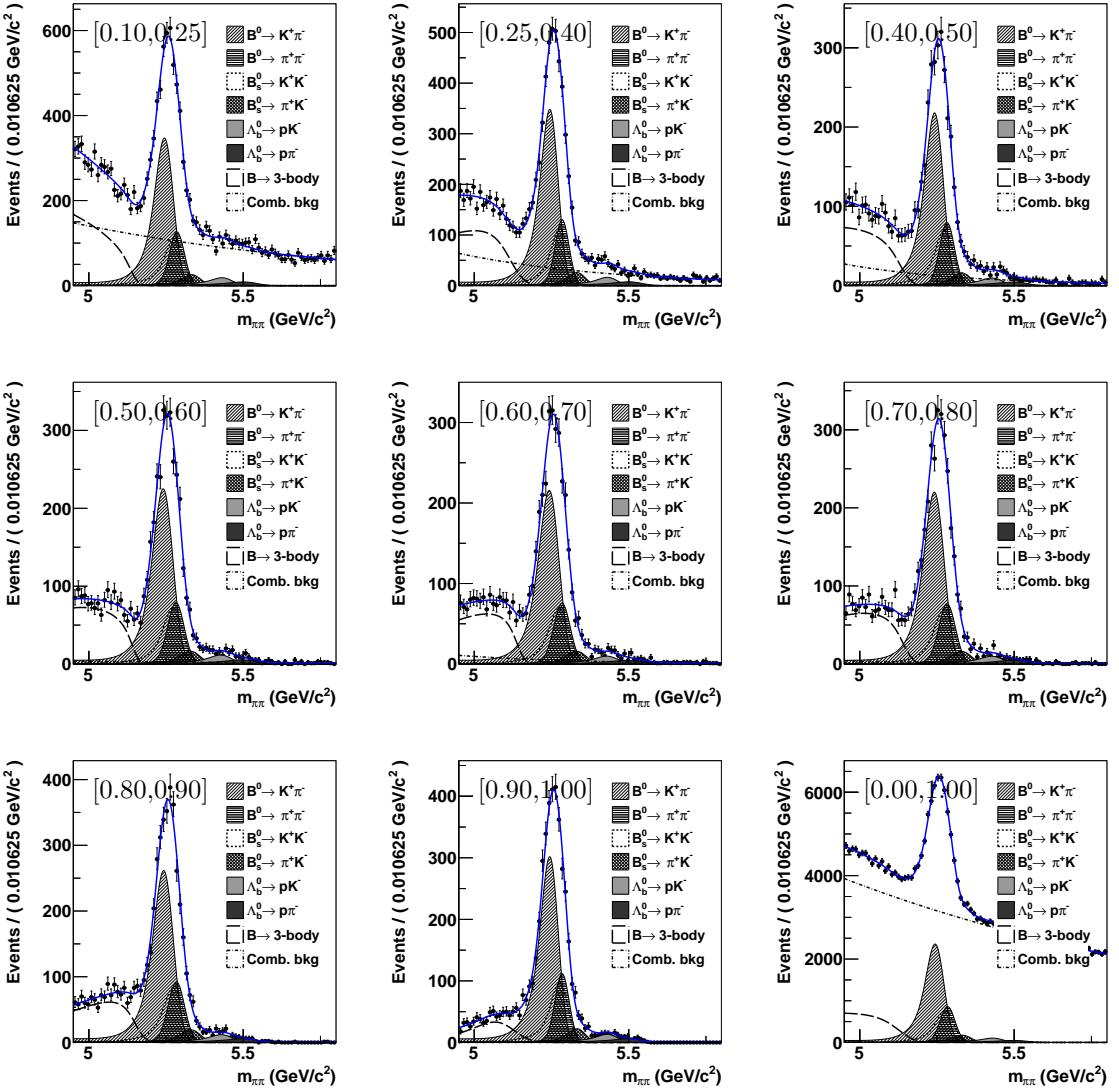


Figure 33: Invariant mass distributions of $B_{(s)}^0 \rightarrow h^+ h^-$ from 2012 data combined without any PID cut. The fit result is superimposed.

1202 correlations between r_i for different values of κ :

$$\bar{r}_i = \sigma_{\bar{r}_i}^2 \sum_a c_{i;a,b}^{-1} r_{i,a}$$

$$\sigma_{\bar{r}_i}^2 = \frac{1}{\sum_{a,b} c_{i;a,b}^{-1}}$$

with $c_{i;a,b}^{-1}$: element of the inverted covariance matrix

a, b : indices of the different data points for the different $\Delta LL_{K-\pi}$ cuts

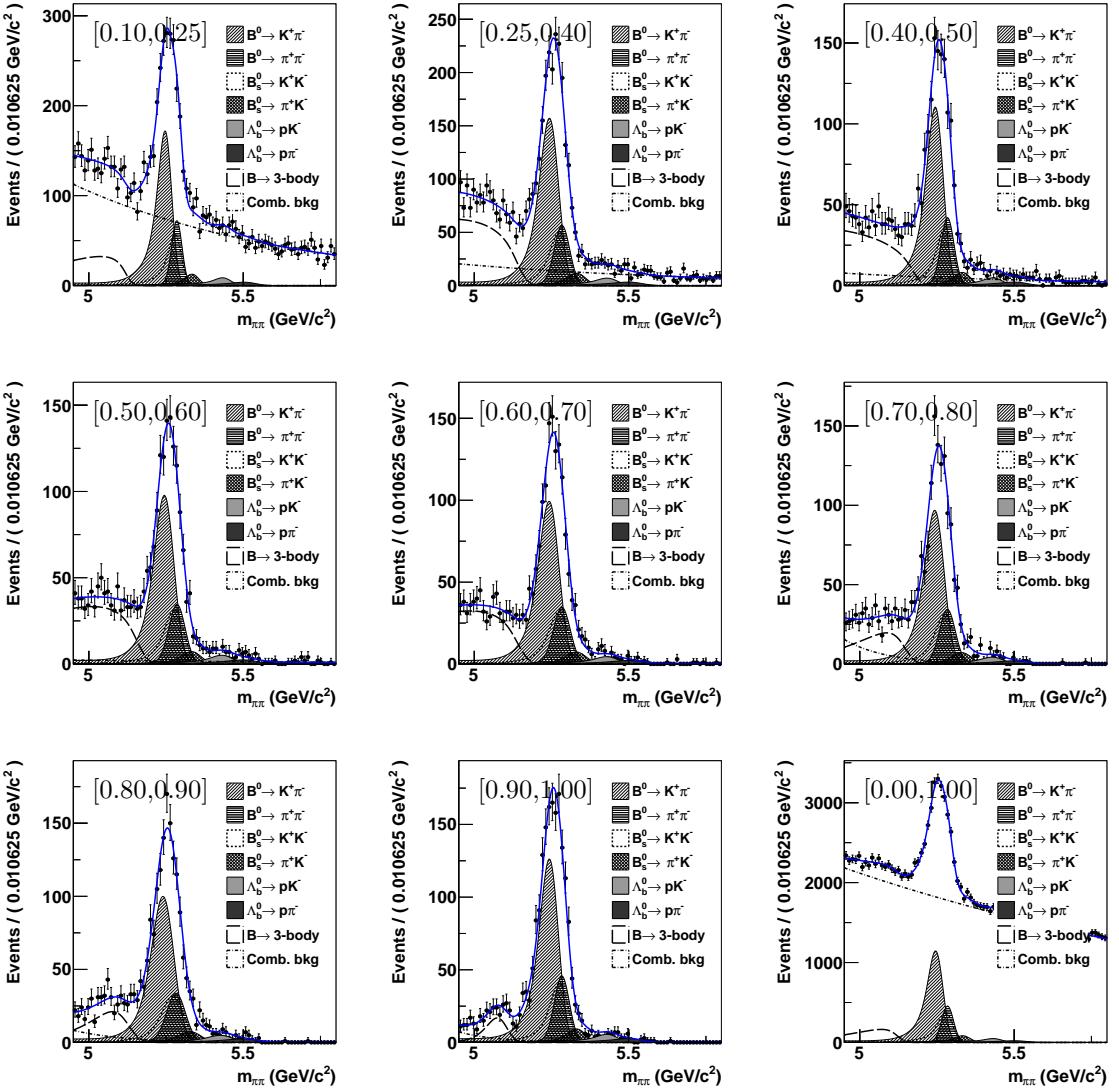


Figure 34: Invariant mass distributions of $B_{(s)}^0 \rightarrow h^+ h^-$ from 2015 data combined without any PID cut. The fit result is superimposed.

¹²⁰³ $\sigma_{\bar{r}_i}$ is also the estimated statistical error on \bar{r}_i .

¹²⁰⁴ The covariance matrix for the i -th BDT bin is defined as

$$c_{i;a,b} = \sigma_{r_{i,a}} \cdot \sigma_{r_{i,b}} \cdot \rho_{i,a,b}$$

with $\sigma_{r_{i,a}}$: statistical error on $r_{i,a}$ propagated from ((17))

$\rho_{i,a,b}$ is the correlation estimated using the number of events in the i -th BDT surviving $\Delta LL_{K-\pi}$ cuts of κ ($n_{i;a}$) and those surviving cuts of $\kappa' > \kappa$ ($n_{i;b}$):

$$\rho_{i,a,b} = \sqrt{\frac{n_{i;b}}{n_{i;a}}}$$

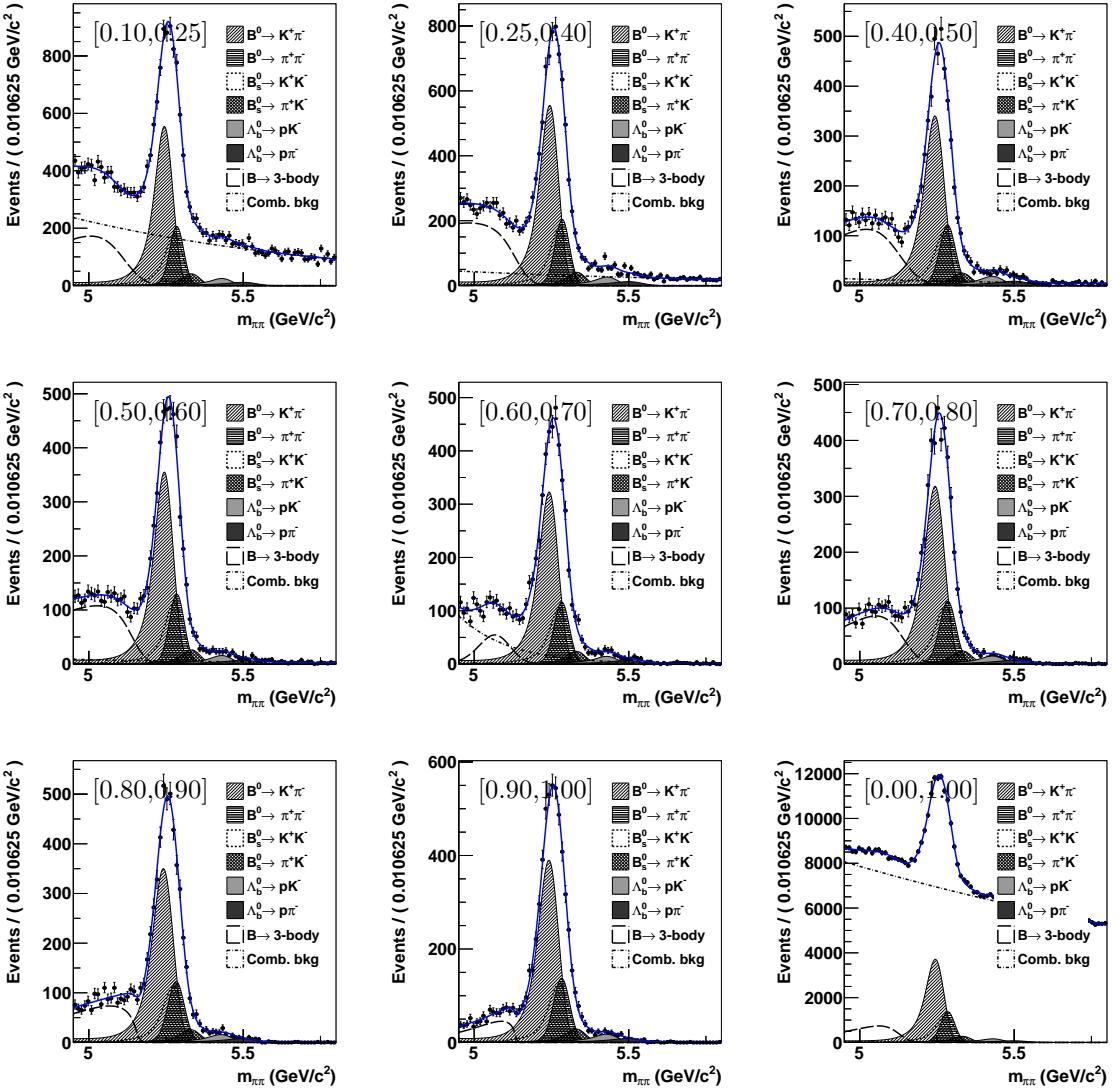


Figure 35: Invariant mass distributions of $B_{(s)}^0 \rightarrow h^+ h^-$ from 2016 data combined without any PID cut. The fit result is superimposed.

The formula for the correlation is derived from comparing the binomial uncertainty on $\varepsilon = n_{i;b}/n_{i;a}$ which is $\sigma_\varepsilon^2 = \varepsilon(1 - \varepsilon)/n_{i;a}$ with the result of gaussian error propagation:

$$\sigma_\varepsilon^2 = \left(\frac{\partial \varepsilon}{\partial n_{i;b}} \sigma_{n_{i;b}} \right)^2 + \left(\frac{\partial \varepsilon}{\partial n_{i;a}} \sigma_{n_{i;a}} \right)^2 + 2\rho \left(\frac{\partial \varepsilon}{\partial n_{i;b}} \sigma_{n_{i;b}} \right) \left(\frac{\partial \varepsilon}{\partial n_{i;a}} \sigma_{n_{i;a}} \right)$$

1205 using $\sigma_{n_{i;b}} = \sqrt{n_{i;b}}$ and $\sigma_{n_{i;a}} = \sqrt{n_{i;a}}$.

1206

1207 In Fig. 41 the BDT signal PDFs are shown for Run I (2011+2012), 2015 and 2016
1208 data separately.

Table 43: Yields per BDT bin from 2012 $B^0 \rightarrow K^+ \pi^-$ fit

BDT bin	$B^0 \rightarrow K^+ \pi^-$ fit yield	PID efficiency	corrected yield
[0.00, 0.10]	467 \pm 143	0.6134 \pm 0.0410	761 \pm 233
[0.10, 0.25]	1559 \pm 55	0.6238 \pm 0.0192	2498 \pm 117
[0.25, 0.40]	1499 \pm 47	0.5939 \pm 0.0055	2523 \pm 83
[0.40, 0.50]	886 \pm 35	0.5489 \pm 0.0001	1613 \pm 64
[0.50, 0.60]	946 \pm 35	0.4674 \pm 0.0000	2024 \pm 75
[0.60, 0.70]	870 \pm 30	0.4699 \pm 0.0001	1852 \pm 63
[0.70, 0.80]	853 \pm 30	0.4101 \pm 0.0000	2079 \pm 73
[0.80, 0.90]	841 \pm 29	0.3753 \pm 0.0001	2240 \pm 78
[0.90, 1.00]	855 \pm 30	0.3530 \pm 0.0006	2422 \pm 84

Table 44: Yields per BDT bin from 2015 $B^0 \rightarrow K^+ \pi^-$ fit

BDT bin	$B^0 \rightarrow K^+ \pi^-$ fit yield	PID efficiency	corrected yield
[0.00, 0.10]	627 \pm 137	0.6494 \pm 0.0794	966 \pm 242
[0.10, 0.25]	755 \pm 37	0.7038 \pm 0.0002	1072 \pm 53
[0.25, 0.40]	731 \pm 34	0.6258 \pm 0.0001	1169 \pm 54
[0.40, 0.50]	443 \pm 25	0.5078 \pm 0.0002	872 \pm 50
[0.50, 0.60]	376 \pm 24	0.4894 \pm 0.0002	769 \pm 49
[0.60, 0.70]	349 \pm 19	0.4101 \pm 0.0002	852 \pm 45
[0.70, 0.80]	357 \pm 19	0.4222 \pm 0.0001	845 \pm 45
[0.80, 0.90]	367 \pm 19	0.3807 \pm 0.0001	965 \pm 50
[0.90, 1.00]	425 \pm 21	0.4074 \pm 0.0001	1042 \pm 51

1209 The PDFs shown in these plots have been corrected for the trigger requirements placed
 1210 on $B^0 \rightarrow \mu^+ \mu^-$ relative to $B^0 \rightarrow K^+ \pi^-$.

1211 The relative trigger efficiencies per bin are evaluated from a nTrack reweighted MC, as
 1212 they are for the exclusive background estimation.

1213 The BDT distribution for $B_s^0 \rightarrow \mu^+ \mu^-$ is corrected later for the lifetime acceptance, as
 1214 discussed in Section 8.1. Both the $B^0 \rightarrow \mu^+ \mu^-$ and $B_s^0 \rightarrow \mu^+ \mu^-$ BDT distributions are
 1215 corrected later for the relative PID efficiency per BDT bin.

1216 The statistical uncertainty is included as a red band and the systematic uncertainty as
 1217 a grey band. The results are summarised in Tabs. 46, 47, 48. The difference between the
 1218 MC and data distribution, in terms of χ^2 , is 9.02, 3.62 and 11.41 for Run I, 2015, and
 1219 2016 respectively. As there are 7 degrees of freedom, no significant difference between MC
 1220 and data is found.

1221 6.4.3 Systematics

1222 For the BDT calibration, these systematics were considered:

1223 1. **Consistency with $B_{(s)}^0 \rightarrow h^+ h^-$ fits:** The dominant systematic comes from the

Table 45: Yields per BDT bin from 2016 $B^0 \rightarrow K^+\pi^-$ fit

BDT bin	$B^0 \rightarrow K^+\pi^-$ fit yield	PID efficiency	corrected yield
[0.00, 0.10]	2023 ± 232	0.6654 ± 0.0501	3040 ± 416
[0.10, 0.25]	2614 ± 70	0.6730 ± 0.0000	3885 ± 104
[0.25, 0.40]	2443 ± 61	0.5907 ± 0.0000	4136 ± 103
[0.40, 0.50]	1355 ± 46	0.5281 ± 0.0000	2566 ± 87
[0.50, 0.60]	1381 ± 47	0.4940 ± 0.0000	2795 ± 82
[0.60, 0.70]	1263 ± 36	0.4458 ± 0.0001	2833 ± 81
[0.70, 0.80]	1065 ± 39	0.4084 ± 0.0001	2609 ± 95
[0.80, 0.90]	1203 ± 35	0.3984 ± 0.0000	3020 ± 89
[0.90, 1.00]	1388 ± 38	0.3935 ± 0.0001	3527 ± 96

Table 46: Summary of signal BDT PDF (value with statistical and systematic uncertainty) for RunI data (2011+2012).

BDT bin	value	stat.	syst.
[0.00, 0.25]	0.2025	0.0177	0.0280
[0.25, 0.40]	0.1540	0.0049	0.0140
[0.40, 0.50]	0.0975	0.0037	0.0096
[0.50, 0.60]	0.1133	0.0039	0.0043
[0.60, 0.70]	0.0994	0.0033	0.0063
[0.70, 0.80]	0.1033	0.0035	0.0049
[0.80, 0.90]	0.0998	0.0033	0.0065
[0.90, 1.00]	0.1055	0.0034	0.0039

comparison of the $B^0 \rightarrow K^+\pi^-$ yields evaluated from the exclusive $B^0 \rightarrow K^+\pi^-$ and inclusive $B_{(s)}^0 \rightarrow h^+h^-$ fits in bins 1b to 8. The $B^0 \rightarrow K^+\pi^-$ yields from the inclusive fits were obtained by scaling the total $B_{(s)}^0 \rightarrow h^+h^-$ yields with the expected $B^0 \rightarrow K^+\pi^-$ fraction: $f_{B^0 \rightarrow K^+\pi^-} = 0.598 \pm 0.012$.

2. **PID consistency:** a linear χ^2 fit to the data points in the PID cut range for each BDT bin is performed. if r_i at $\kappa = 5$ is found to be significantly different from the baseline value \bar{r}_i , the difference is taken as a systematic uncertainty. Significant is defined as $|r_{\text{inter}} - \bar{r}| > \sqrt{\sigma_{\text{inter}}^2 + \sigma_r^2}$ where σ_{inter} is the uncertainty extracted from the fit on the extrapolated value at $\kappa = 5$. It is found that all BDT bins are consistent over their PID range; therefore, this systematic is negligible.
3. **PID efficiency determination:** To check the PID efficiency, including the re-weighting procedure, used to correct the yield per BDT bin, an alternative way to determine the PID efficiency was also used, with the binning mentioned at the start of this section. The difference with the efficiency determined using the nominal binning is taken as a systematic per BDT bin.

Table 47: Summary of signal BDT PDF (value with statistical and systematic uncertainty) for 2015.

BDT bin	value	stat.	syst.
[0.00, 0.00]	0.2655	0.0276	0.0162
[0.25, 0.40]	0.1482	0.0079	0.0157
[0.40, 0.50]	0.1047	0.0065	0.0056
[0.50, 0.60]	0.0896	0.0061	0.0087
[0.60, 0.70]	0.0951	0.0056	0.0053
[0.70, 0.80]	0.0916	0.0055	0.0029
[0.80, 0.90]	0.0963	0.0056	0.0029
[0.90, 1.00]	0.1011	0.0056	0.0039

Table 48: Summary of signal BDT PDF (value with statistical and systematic uncertainty) for 2016.

BDT bin	value	stat.	syst.
[0.00, 0.25]	0.2674	0.0146	0.0287
[0.25, 0.40]	0.1612	0.0046	0.0150
[0.40, 0.50]	0.0973	0.0035	0.0133
[0.50, 0.60]	0.1002	0.0033	0.0063
[0.60, 0.70]	0.0947	0.0030	0.0051
[0.70, 0.80]	0.0849	0.0032	0.0067
[0.80, 0.90]	0.0897	0.0029	0.0048
[0.90, 1.00]	0.0971	0.0030	0.0045

- 1239 4. **Fraction of events outside the mass window:** The amount of B^0 events outside
 1240 of the mass window, due to the cutoff at $5200 \text{ MeV}/c^2$, is about 0.8%. This number
 1241 is stable within $\sim 0.3\%$. The BDT calibration is performed with a correction factor
 1242 which is the reciprocal of the number of events in the mass window, and the result per
 1243 bin is compared with the nominal result. The difference is assigned as a systematic.
- 1244 5. **Fit model:** The analysis is repeated with a different fit model. Instead of the
 1245 nominal fit model, the signal shapes are described by Double Gaussians. As in
 1246 the nominal fit model, both B_s^0 resolution parameters are scaled relative to the B^0
 1247 resolution parameters, where the scale factor is obtained from quarkonia decays.
 1248 The difference between the BDT fractions obtained with the nominal fit model and
 1249 those obtained with this model is taken as a systematic uncertainty. The mass
 1250 distributions for $\kappa = 5$ are shown in Figures ?? and ?? for Run I (2011+2012) and
 1251 2015 data respectively.
- 1252 6. **Trigger corrections:** To correct for the trigger requirements placed on $B^0 \rightarrow \mu^+ \mu^-$
 1253 relative to $B^0 \rightarrow K^+ \pi^-$, the relative efficiencies per bin are evaluated from MC, as
 1254 they are for the exclusive background estimation. The uncertainty on this correction

1255 is assigned as a systematic. An additional systematics is added from the comparison
 1256 of the (HLT2B2HH_TOS—HLT2B2HH_DEC) efficiency, $\varepsilon_{\text{HLT2B2HH_TOS}|\text{HLT2B2HH_DEC}}$
 1257 in data and MC evaluated by fitting the $B_{(s)}^0 \rightarrow h^+h^-$ mass distribution in the full
 1258 BDT range. In Tab. 49 the $\varepsilon_{\text{HLT2B2HH_TOS}|\text{HLT2B2HH_DEC}}$ is reported for data and MC
 1259 in 2011, 2012, 2015 and 2016.

Table 49: HLT2B2HH_TOS|HLT2B2HH_DEC trigger efficiency for $B_{(s)}^0 \rightarrow h^+h^-$ data and simulated events.

year	data	MC
2011	$81.8 \pm 2.4\%$	$78.2 \pm 0.4\%$
2012	$79.5 \pm 1.5\%$	$80.2 \pm 0.1\%$
2015	$91.5 \pm 1.9\%$	$90.5 \pm 0.1\%$
2016	$90.1 \pm 1.0\%$	$91.20 \pm 0.07\%$

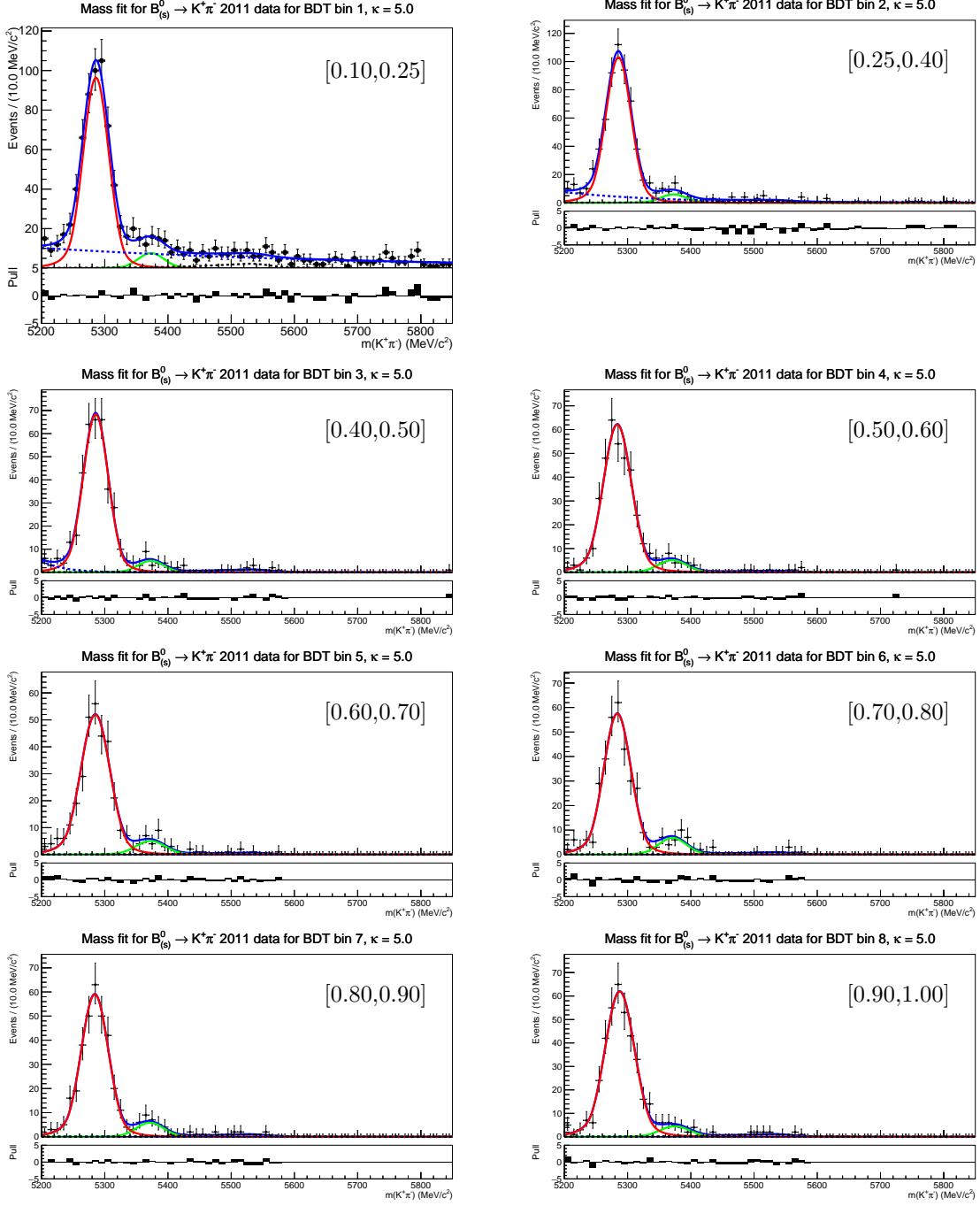


Figure 36: Invariant mass distributions of $B^0 \rightarrow K^+\pi^-$ from 2011 data combined with $\Delta LL_{K-\pi}$ efficiency correction in different BDT bins for $\Delta LL_{K-\pi}$ cut value $\kappa = 5$. The red solid line shows the B^0 and B_s^0 signals, the red dashed one the one from $\Lambda_b^0 \rightarrow ph$ where the proton is misidentified as a kaon or pion. The combinatorial background is shown by the blue dashed line, while the physical background (partially-reconstructed) is shown by the black solid line.

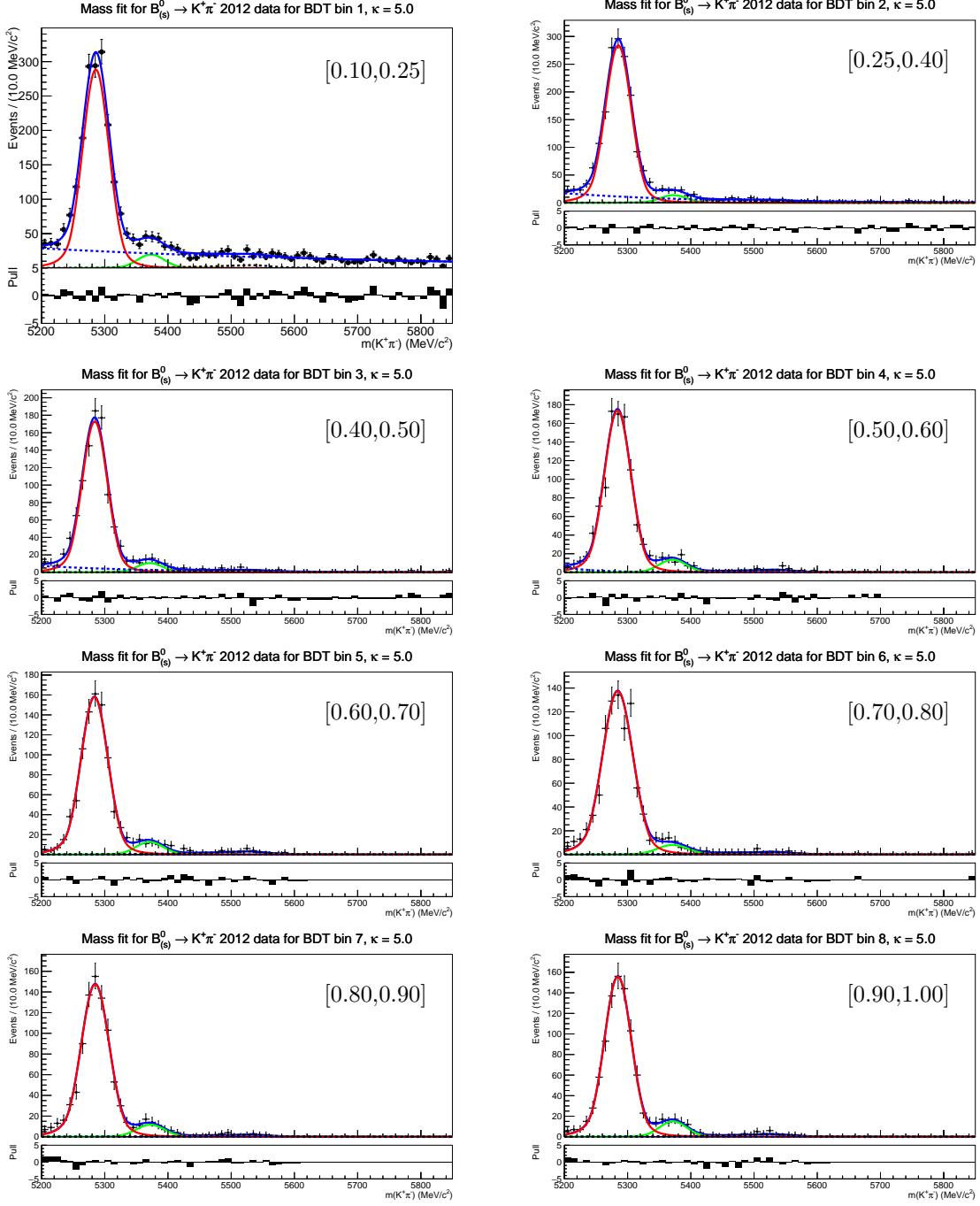


Figure 37: Invariant mass distributions of $B^0 \rightarrow K^+\pi^-$ from 2012 data combined with $\Delta LL_{K-\pi}$ efficiency correction in different BDT bins for $\Delta LL_{K-\pi}$ cut value $\kappa = 5$. The red solid line shows the B^0 and B_s^0 signals, the red dashed one the one from $\Lambda_b^0 \rightarrow ph$ where the proton is misidentified as a kaon or pion. The combinatorial background is shown by the blue dashed line, while the physical background (partially-reconstructed) is shown by the black solid line.

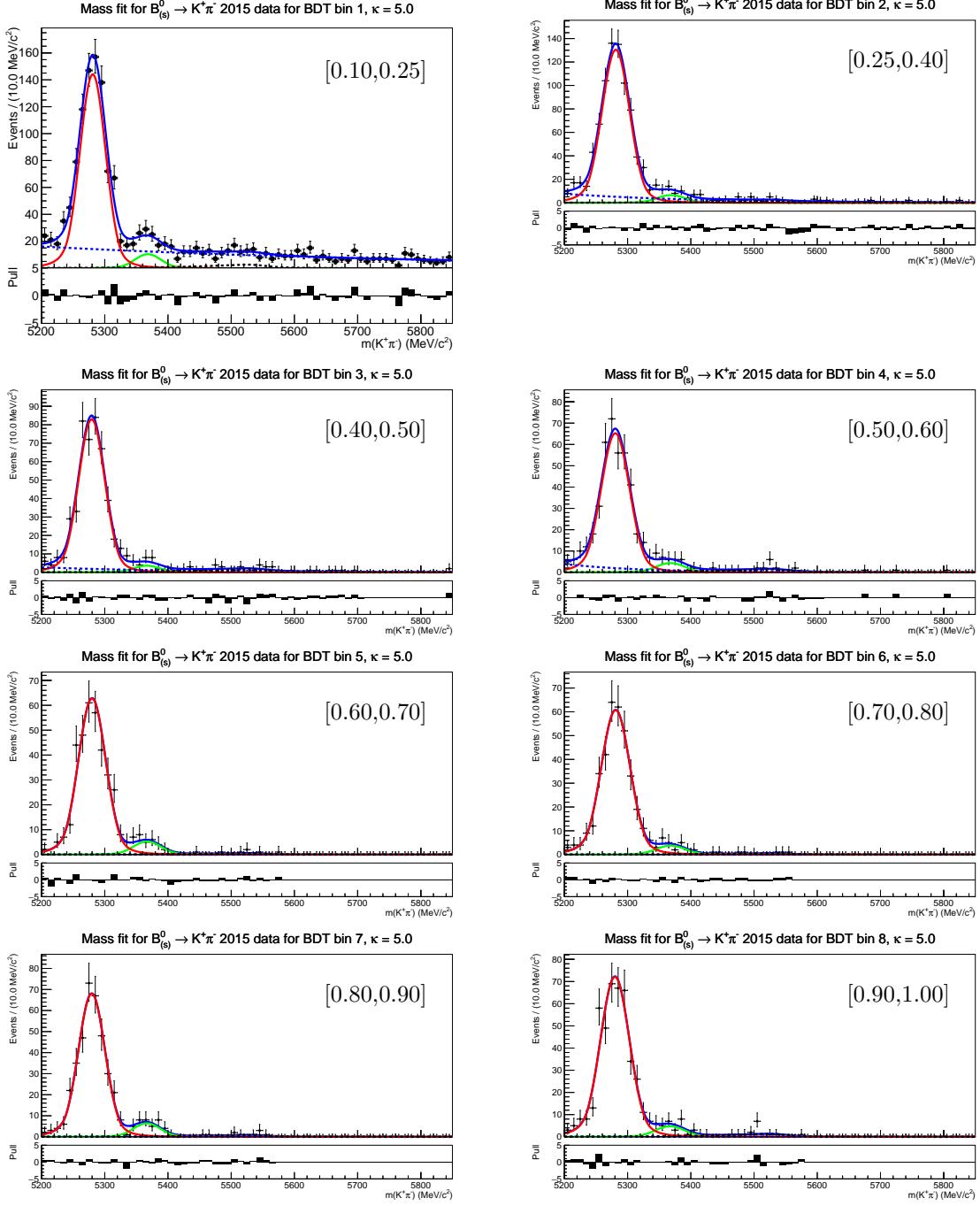


Figure 38: Invariant mass distributions of $B^0 \rightarrow K^+\pi^-$ from 2015 data combined with $\Delta LL_{K-\pi}$ efficiency correction in different BDT bins for $\Delta LL_{K-\pi}$ cut value $\kappa = 5$. The red solid line shows the B^0 and B_s^0 signals, the red dashed one the one from $\Lambda_b^0 \rightarrow ph$ where the proton is misidentified as a kaon or pion. The combinatorial background is shown by the blue dashed line, while the physical background (partially-reconstructed) is shown by the black solid line.

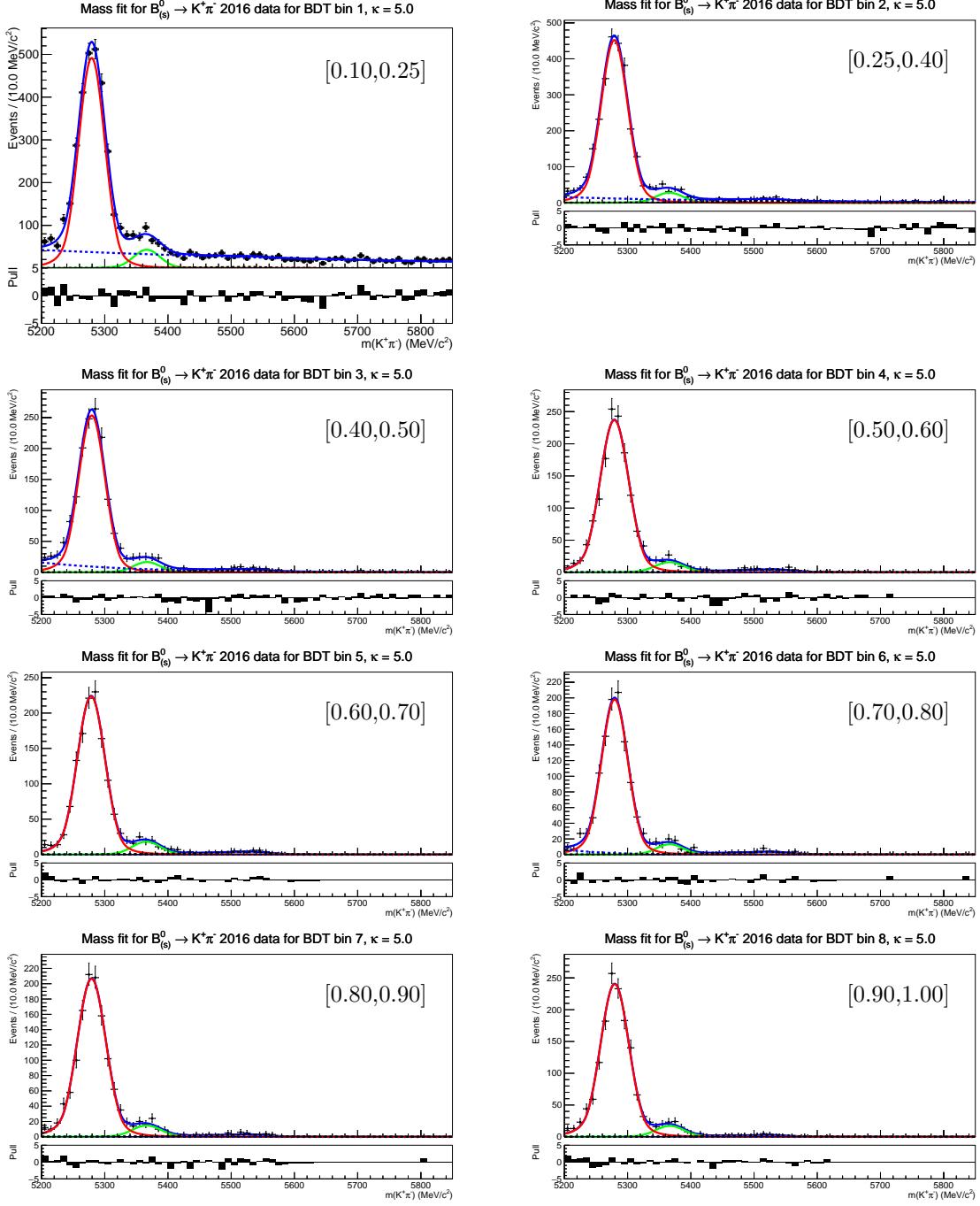


Figure 39: Invariant mass distributions of $B^0 \rightarrow K^+\pi^-$ from 2016 data combined with $\Delta LL_{K-\pi}$ efficiency correction in different BDT bins for $\Delta LL_{K-\pi}$ cut value $\kappa = 5$. The red solid line shows the B^0 and B_s^0 signals, the red dashed one the one from $\Lambda_b^0 \rightarrow ph$ where the proton is misidentified as a kaon or pion. The combinatorial background is shown by the blue dashed line, while the physical background (partially-reconstructed) is shown by the black solid line.

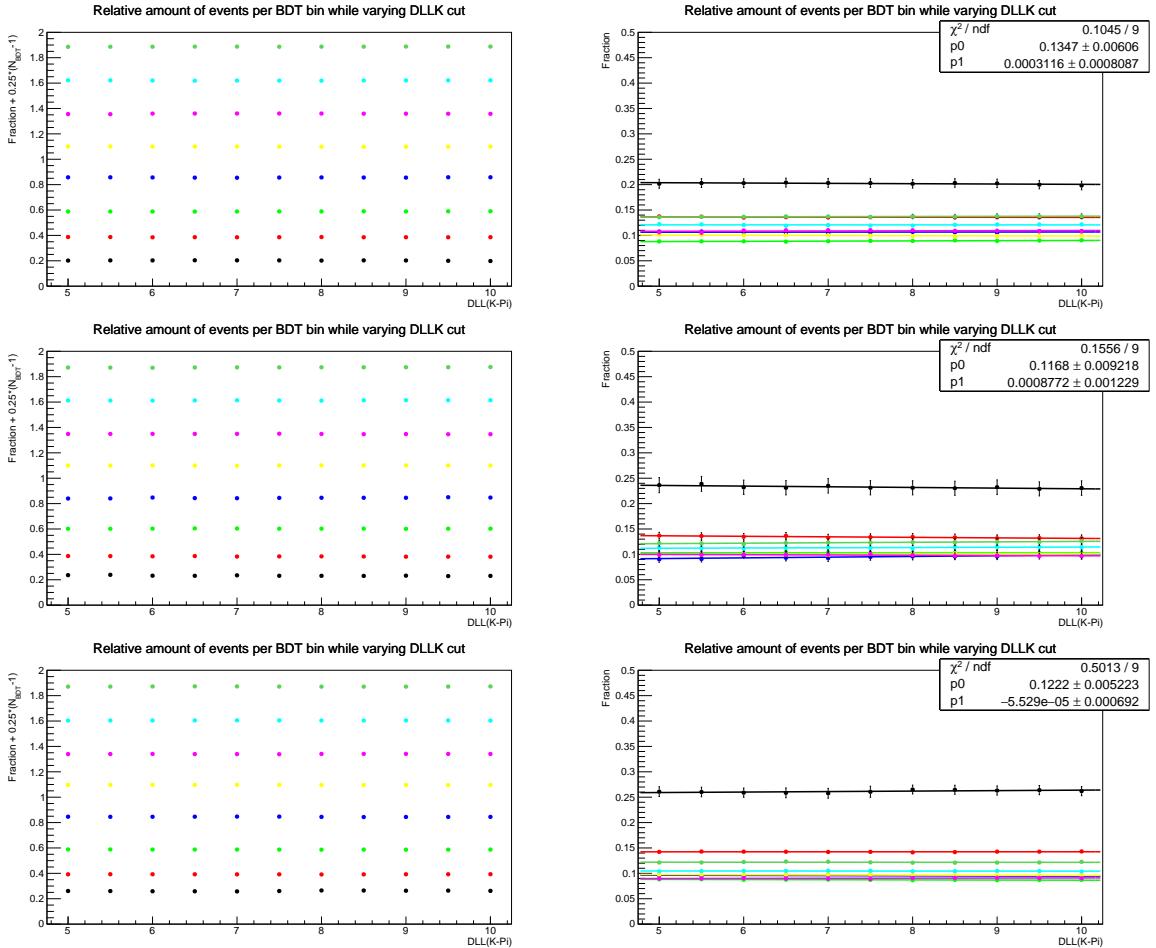


Figure 40: Fraction r_i of signal events from $B^0 \rightarrow K^+\pi^-$ in the i -th BDT bin as a function of the $\Delta LL_{K-\pi}$ cut for Run I data (2011+2012) (top), 2015 (middle), and 2015 (bottom) data samples. Left: For better visualization the values are shifted by $0.25 \cdot (i - 1)$; Right: Data points for r_i are fitted with a linear function to detect systematic drifts in the fraction over the $\Delta LL_{K-\pi}$ range as described in the text.

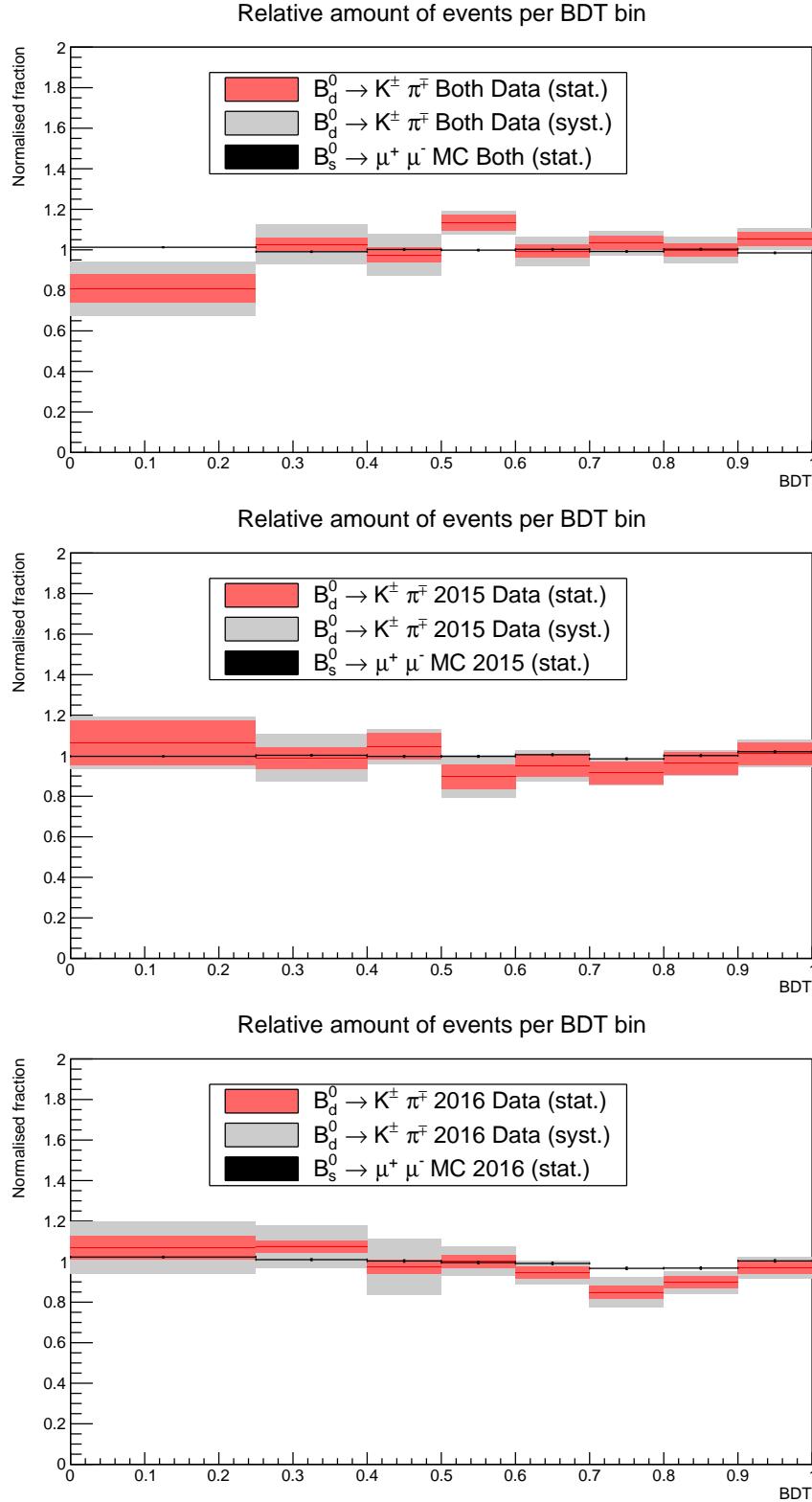


Figure 41: BDT signal pdfs for Run I data (2011 + 2012) (top), 2015 (middle), and 2016 (bottom) data samples. The statistical uncertainty is included as a red band and systematic uncertainty as a gray band. The systematic uncertainties are be discussed in 6.4.3. $B^0 \rightarrow K^+ \pi^-$ distribution (black points) from simulation is superimposed.

1260 **6.5 Invariant mass**

1261 The invariant mass distribution for the signal is described by a Crystal Ball function. The
1262 Crystal Ball function is characterized by the following parameters:

- 1263 • Mean μ
1264 • Mass resolution σ
1265 • Transition point α
1266 • Exponent n

1267 The determination of these parameters is described in the following subsections.

1268 **6.5.1 Mean**

1269 The mean for the invariant mass distribution is estimated from $B^0 \rightarrow K^+ \pi^-$ ($B_s^0 \rightarrow K^+ K^-$)
1270 for $B^0 \rightarrow \mu^+ \mu^-$ ($B_s^0 \rightarrow \mu^+ \mu^-$). In contrast to the normalisation and BDT calibration, no
1271 requirement is placed on the way the events were triggered. The mass hypotheses for these
1272 decays are separated using these cuts on the two hadrons:

- 1273 • a kaon: $\Delta LL_{K-\pi} > \kappa$
1274 • a pion: $\Delta LL_{K-\pi} < -\kappa$

1275 Here, the $K\pi$ and πK mass hypotheses are combined when investigating $B^0 \rightarrow K^+ \pi^-$.
1276 Just as for the normalisation and BDT calibration, κ is varied to investigate the PID
1277 dependence of the mean calibration. Due to the high amount of number of events in the
1278 sample without trigger requirements, κ is varied from 0 to 20. For the nominal fit, $\kappa = 10$
1279 is used. An additional cut is placed to reject $\Lambda_b \rightarrow ph$ backgrounds: $\Delta LL_{h-p} > 0$. This
1280 cut will be investigated as a systematic. The fit model from the normalisation and BDT
1281 calibration is used.

1282 Figures 42,43,44 and 45 for 2011, 2012, 2015, and 2016 data respectively, show the
1283 invariant mass distribution for the two mass hypotheses ($m_{K\pi}$ and m_{KK}) from the selected
1284 events in the $B_{(s)}^0 \rightarrow h^+ h^-$ sample. Tables 50 summarises the results of the mean
1285 calibration.

1286 As the main systematic uncertainty is expected to come from the PID cut that has to
1287 be applied in order to separate the mass hypotheses, this is the only systematic that is
1288 currently investigated.

1289

1290 • **PID dependence:** κ , the cut on $\Delta LL_{K-\pi}$, is varied from 0 to 20. Figures 46, 47,
1291 48 and 49 show the dependence for 2011, 2012, 2015 and 2016 data respectively.
1292 The difference between the mean at $\kappa = 10$ and $\kappa = 20$ is taken as a systematic
1293 uncertainty. Also, the dependence is fitted with a function given by

$$\mu(\kappa) = \mu_0 + \frac{a_0}{1 + r \cdot \exp(\kappa/\kappa_0)}. \quad (18)$$

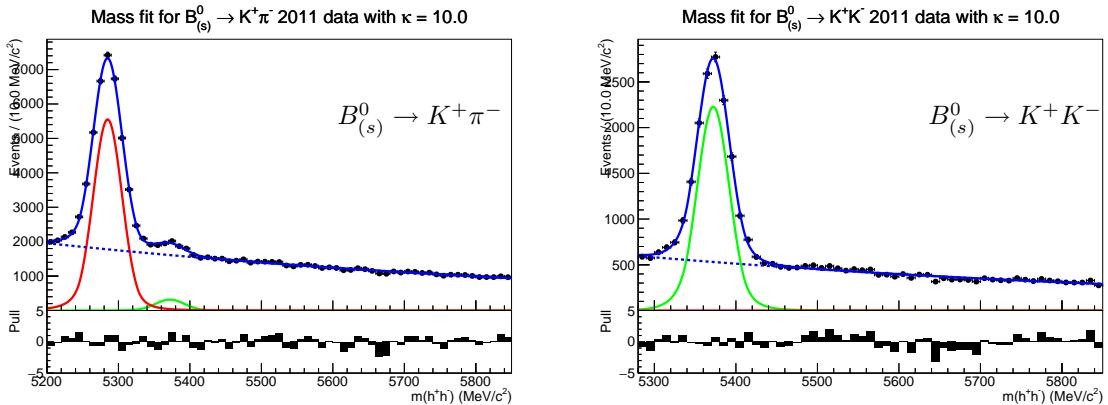


Figure 42: Invariant mass distributions of $B^0 \rightarrow K^+\pi^-$ and $B_{(s)}^0 \rightarrow K^+K^-$ for 2011 data. These distributions are used to determine the mean and resolution of the invariant mass distribution for $B_{(s)}^0 \rightarrow \mu^+\mu^-$. The red line indicates B^0 signal, the green $B_{(s)}^0$ signal, and the blue dashed line combinatorial.

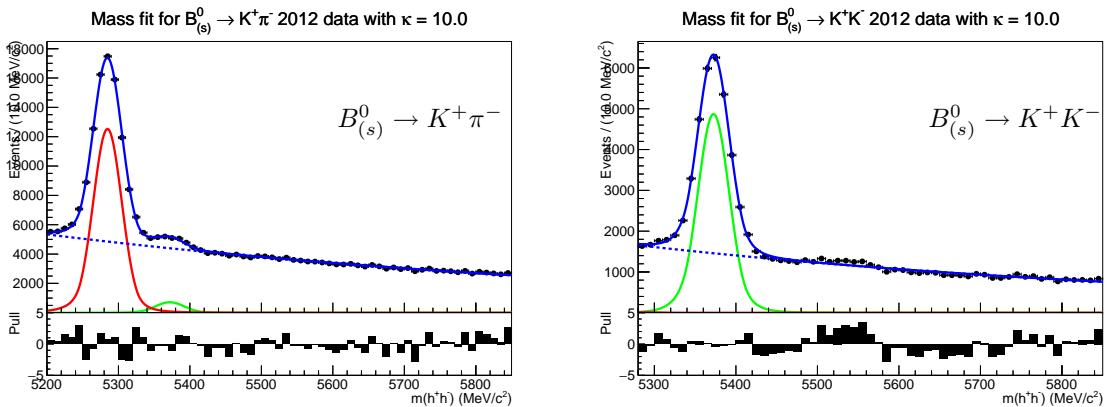


Figure 43: Invariant mass distributions of $B^0 \rightarrow K^+\pi^-$ and $B_{(s)}^0 \rightarrow K^+K^-$ for 2012 data. These distributions are used to determine the mean and resolution of the invariant mass distribution for $B_{(s)}^0 \rightarrow \mu^+\mu^-$. The red line indicates B^0 signal, the green $B_{(s)}^0$ signal, and the blue dashed line combinatorial.

1294 This function models a parameter (mean or resolution) as a constant term and a sec-
 1295 ond term that is caused by misidentified backgrounds and depends exponentially on
 1296 κ . It is only used to model the dependence, not to extract the mean or its uncertainty.
 1297

- 1298 • **Cut on ΔLL_{h-p} :** The cut that is applied to reject $\Lambda_b \rightarrow ph$ background can induce
 1299 a bias on the mean. The fit is repeated at -2 and 2. The variation of the mean over
 1300 this range is taken as a systematic uncertainty.

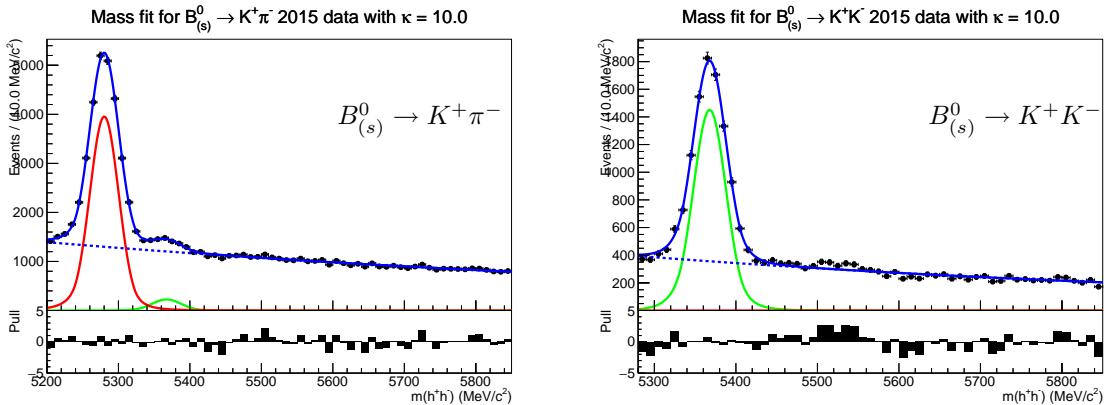


Figure 44: Invariant mass distributions of $B^0 \rightarrow K^+\pi^-$ and $B_s^0 \rightarrow K^+K^-$ for 2015 data. These distributions are used to determine the mean and resolution of the invariant mass distribution for $B_{(s)}^0 \rightarrow \mu^+\mu^-$. The red line indicates B^0 signal, the green B_s^0 signal, and the blue dashed line combinatorial.

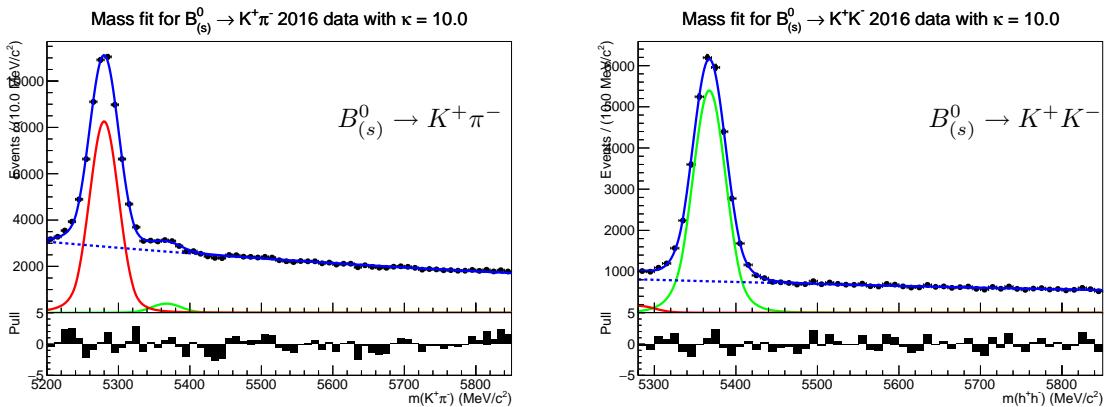


Figure 45: Invariant mass distributions of $B^0 \rightarrow K^+\pi^-$ and $B_s^0 \rightarrow K^+K^-$ for 2016 data. These distributions are used to determine the mean and resolution of the invariant mass distribution for $B_{(s)}^0 \rightarrow \mu^+\mu^-$. The red line indicates B^0 signal, the green B_s^0 signal, and the blue dashed line combinatorial.

1301 The means are as in previous reconstruction versions for 2011 and 2012 data about 0.1 %
 1302 above the nominal value for the B^0 and B_s^0 mass. Interestingly, means from the 2015 data
 1303 are in far better agreement with the nominal mass values, deviating by 0.014 %(0.004 %)
 1304 for B^0 (B_s^0) decays, with a statistical uncertainty of 0.005 %(0.006 %) respectively. Thus,
 1305 the hypothesis that Run II data has a B^0 or B_s^0 mean consistent with the PDG value
 1306 cannot be rejected yet.

Table 50: Summary of mean for the invariant dimuon mass distribution for 2011,2012 and 2015 data.

Dataset	B^0 mean	B_s^0 mean
2011	$(5284.83 \pm 0.27_{\text{stat}} \pm 0.41_{\text{syst}}) \text{ MeV}/c^2$	$(5371.79 \pm 0.29_{\text{stat}} \pm 0.51_{\text{syst}}) \text{ MeV}/c^2$
2012	$(5284.68 \pm 0.18_{\text{stat}} \pm 0.13_{\text{syst}}) \text{ MeV}/c^2$	$(5372.16 \pm 0.19_{\text{stat}} \pm 0.21_{\text{syst}}) \text{ MeV}/c^2$
2015	$(5280.09 \pm 0.28_{\text{stat}} \pm 0.05_{\text{syst}}) \text{ MeV}/c^2$	$(5367.40 \pm 0.33_{\text{stat}} \pm 0.20_{\text{syst}}) \text{ MeV}/c^2$
2016	$(5279.91 \pm 0.14_{\text{stat}} \pm 0.11_{\text{syst}}) \text{ MeV}/c^2$	$(5367.33 \pm 0.16_{\text{stat}} \pm 0.50_{\text{syst}}) \text{ MeV}/c^2$

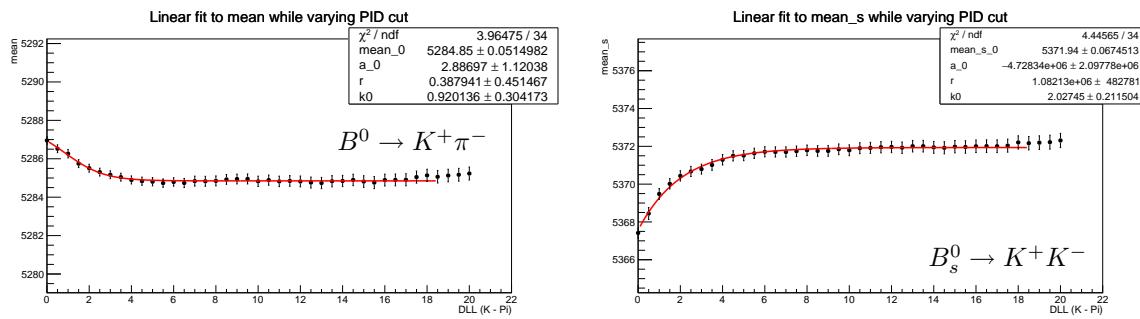


Figure 46: Dependence of the mean on PID cut κ , used to determine the systematic uncertainty, for 2011 data. The line shows the function which is fitted to the data (given in Equation 18).

1307 6.5.2 Invariant mass resolution

1308 The interpolation of the invariant mass resolution between the Charmonium and Bottomo-
1309 nium resonances is used to extrapolate the resolution of the invariant mass peak at the

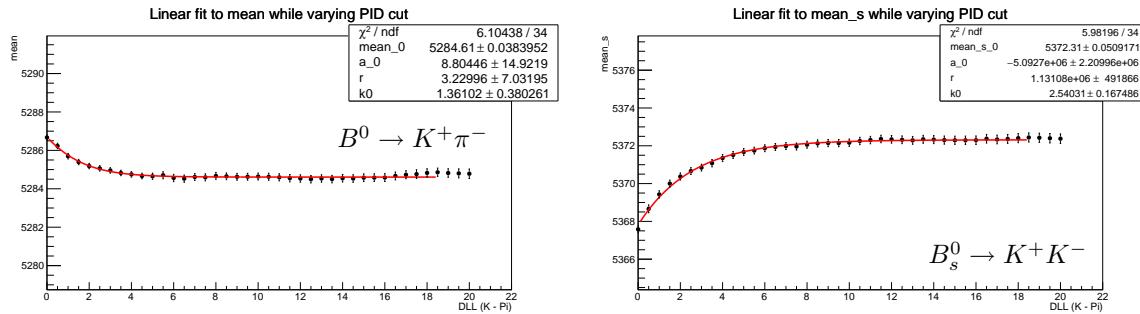


Figure 47: Dependence of the mean on PID cut κ , used to determine the systematic uncertainty, for 2012 data. The line shows the function which is fitted to the data (given in Equation 18).

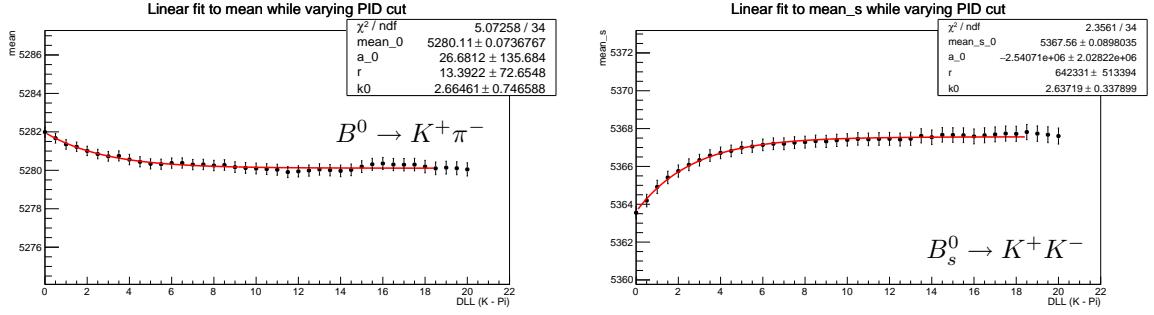


Figure 48: Dependence of the mean on PID cut κ , used to determine the systematic uncertainty, for 2015 data. The line shows the function which is fitted to the data (given in Equation 18).

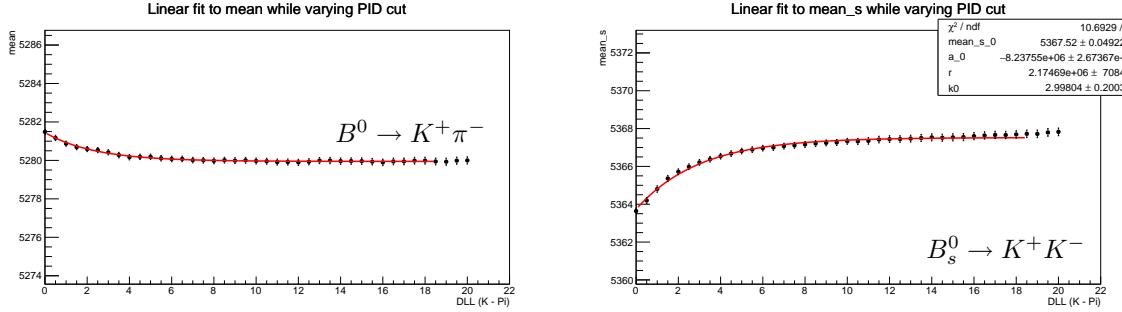


Figure 49: Dependence of the mean on PID cut κ , used to determine the systematic uncertainty, for 2016 data. The line shows the function which is fitted to the data (given in Equation 18).

mass range of B^0 and B_s^0 . Respect to the previous analysis, the combinatorial background in the fit for the extraction of the Υ resonance resolutions is not described anymore with an exponential but with a power-law function, that performs better on a wide mass range.

By analyzing Drell-Yan Monte Carlo the invariant mass resolution as a function of the invariant dimuon mass has been found to be well-described by a power-law function of the form:

$$\sigma_{\mu\mu}(m_{\mu\mu}) = a_0 + a_1 \cdot m_{\mu\mu}^\gamma.$$

Fig. 51 shows the Charmonium and Bottomonium resonances while Fig. 52 shows the interpolation of the invariant mass resolution.

The values for the interpolated mass resolutions at m_{B^0} and $m_{B_s^0}$ are reported in Tab. 51

The systematic uncertainties – summarized in Tab. 52 and Tab. 53 for B^0 and B_s^0 respectively – are analogously determined as in previous analyses. The error asymmetry

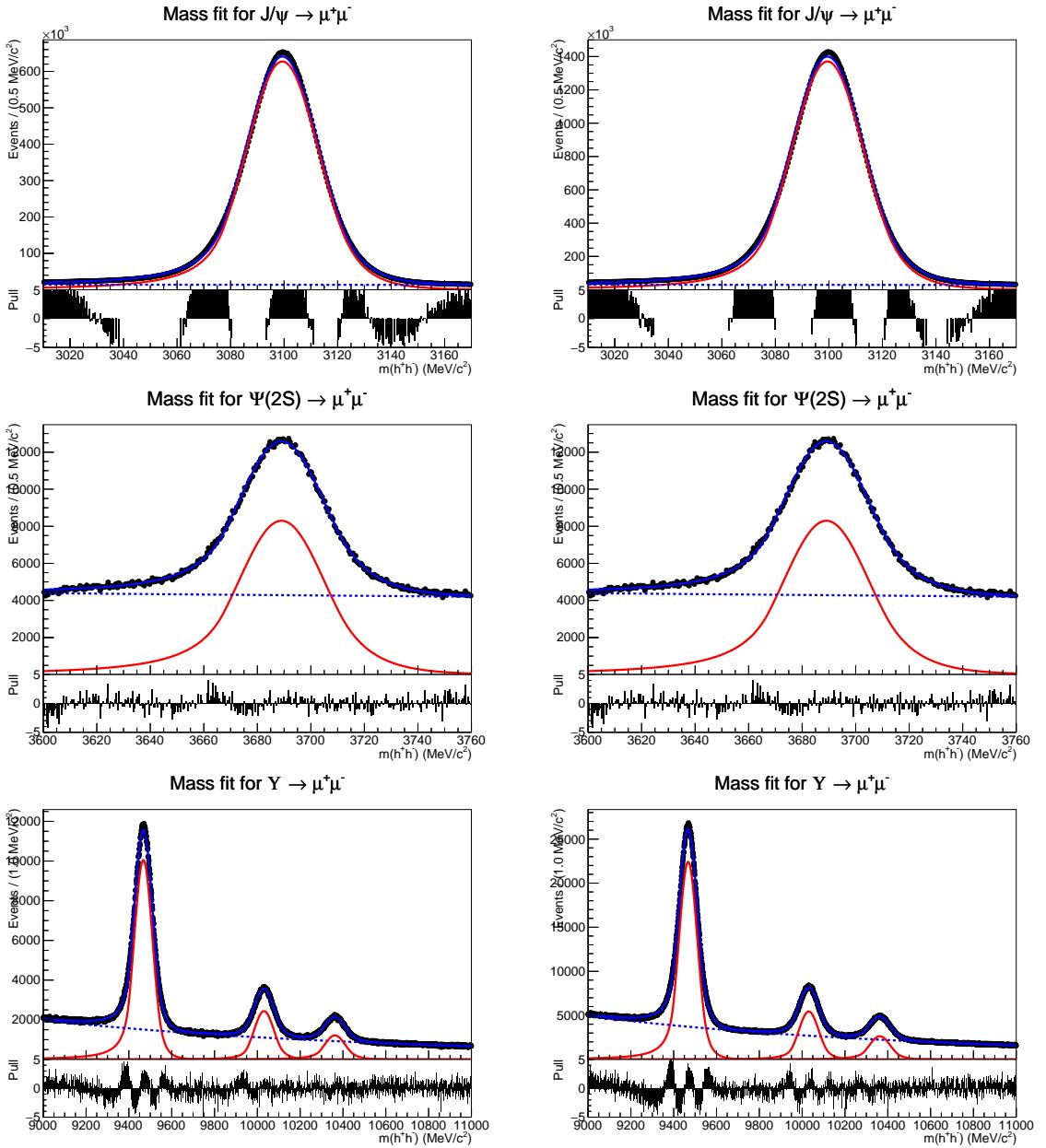


Figure 50: Invariant dimuon mass distribution for, from top to bottom, $J/\psi(1S) \rightarrow \mu^+\mu^-$, $\psi(2S) \rightarrow \mu^+\mu^-$ and $\Upsilon(1,2,3S) \rightarrow \mu^+\mu^-$ (left: 2011 data, right: 2012 data). All resonances are described by Double Sided Crystal Ball functions while the combinatorial background is described by an exponential function.

1320 systematics is not included anymore, because in this case sigma from $\Upsilon(3S) \rightarrow \mu^+\mu^-$,
 1321 with an uncertainty of 0.20 MeV (0.4%), counts as much as sigma from $J/\psi \rightarrow \mu^+\mu^-$,
 1322 with an uncertainty of 0.003 MeV (0.02%).

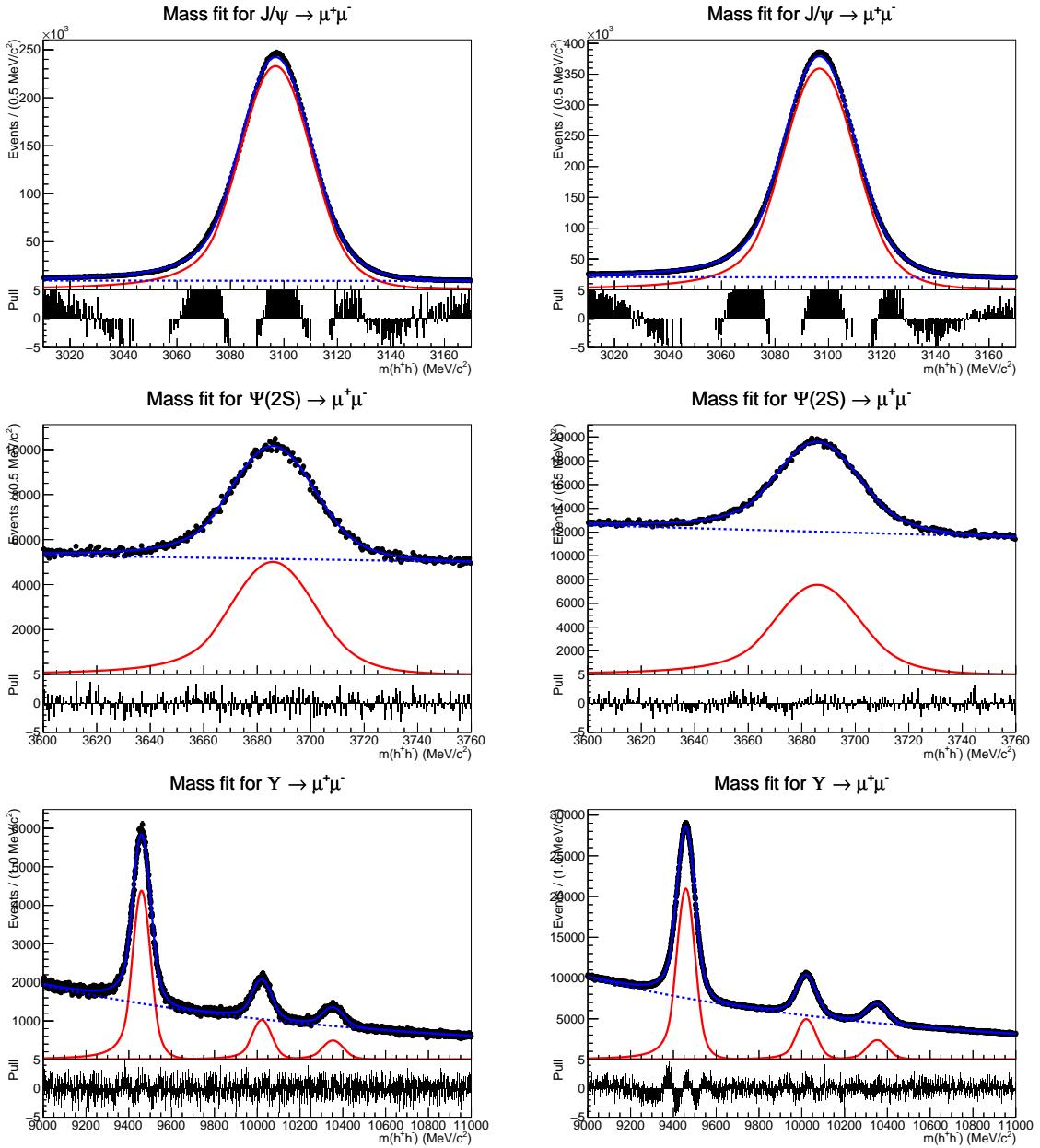


Figure 51: Invariant dimuon mass distribution for, from top to bottom, $J/\psi(1S) \rightarrow \mu^+\mu^-$, $\psi(2S) \rightarrow \mu^+\mu^-$ and $\Upsilon(1, 2, 3S) \rightarrow \mu^+\mu^-$ (left: 2015 data, right: 2016 data). All resonances are described by Double Sided Crystal Ball functions while the combinatorial background is described by an exponential function.

1323 6.5.3 Crystal Ball parameters α and n

1324 The transition point α and the exponent n are determined by smearing gaussianly the
 1325 true invariant dimuon mass distribution such that the resulting invariant mass distribution

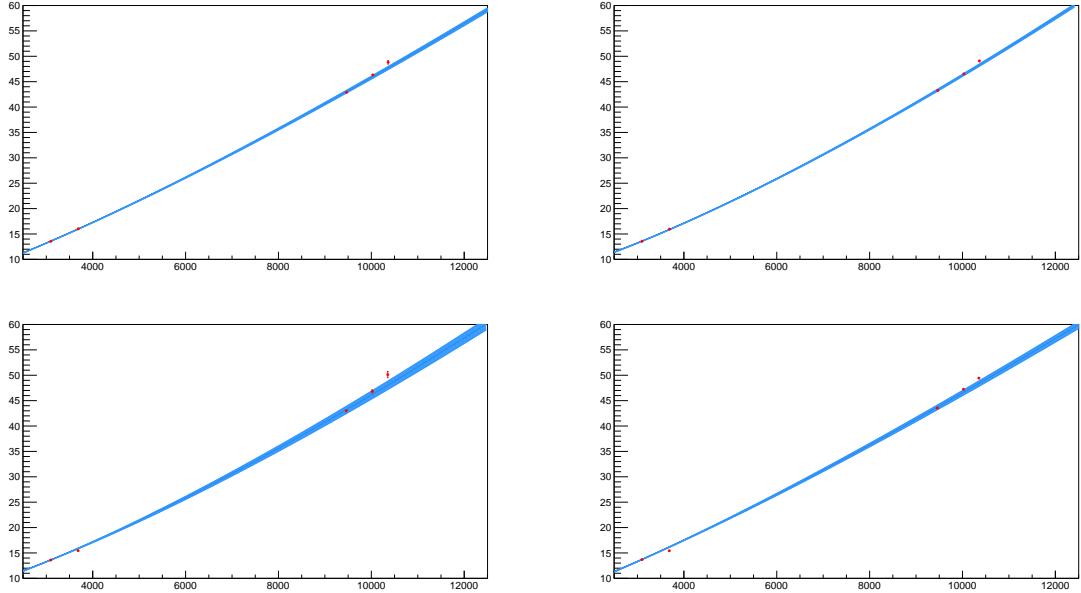


Figure 52: Interpolation of the invariant mass resolution between Charmonium and Bottomonium resonances to the mass of the B^0 and B_s^0 mesons (top left: 2011 data, top right: 2012 data, bottom left: 2015 data, bottom right: 2016 data). The blue band represents the uncertainty at 68% CL on the fitted function.

Table 51: Dimuon invariant mass resolutions for 2011, 2012 and 2015 data from the interpolation from Charmonium and Bottomonium resonances.

Dataset	B^0 resolution	B_s^0 resolution
2011	$(22.84 \pm 0.08_{\text{stat}} \pm 0.36_{\text{syst}}) \text{ MeV}/c^2$	$(23.23 \pm 0.08_{\text{stat}} \pm 0.36_{\text{syst}}) \text{ MeV}/c^2$
2012	$(22.59 \pm 0.06_{\text{stat}} \pm 0.42_{\text{syst}}) \text{ MeV}/c^2$	$(22.98 \pm 0.06_{\text{stat}} \pm 0.42_{\text{syst}}) \text{ MeV}/c^2$
2015	$(22.60 \pm 0.21_{\text{stat}} \pm 0.32_{\text{syst}}) \text{ MeV}/c^2$	$(22.99 \pm 0.22_{\text{stat}} \pm 0.32_{\text{syst}}) \text{ MeV}/c^2$
2016	$(22.44 \pm 0.08_{\text{stat}} \pm 0.50_{\text{syst}}) \text{ MeV}/c^2$	$(22.83 \pm 0.08_{\text{stat}} \pm 0.51_{\text{syst}}) \text{ MeV}/c^2$

1326 has the width determined in Table 51). This distribution is then fitted with a Crystal Ball
 1327 function from which we extract α and n .
 1328 By repeating this procedure several times, we determine the distributions of α and n and
 1329 take their mean as value for the two parameters. The invariant mass resolution used to
 1330 smear the true invariant mass distribution is varied within its uncertainty, such that the
 1331 uncertainty reported for the tail parameters includes both the statistical and systematic
 1332 uncertainties.

Table 52: List of systematic uncertainties on invariant mass resolution for B^0 using the interpolation method.

Systematic	2011 data	2012 data	2015 data	2016 data
Selection cuts	$\pm 0.21 \text{ MeV}/c^2$	$\pm 0.06 \text{ MeV}/c^2$	$\pm 0.05 \text{ MeV}/c^2$	$\pm 0.35 \text{ MeV}/c^2$
Mass window	$\pm 0.27 \text{ MeV}/c^2$			
Fit function of invariant mass	$\pm 0.12 \text{ MeV}/c^2$	$\pm 0.31 \text{ MeV}/c^2$	$\pm 0.17 \text{ MeV}/c^2$	$\pm 0.18 \text{ MeV}/c^2$
Total systematic error	$\pm 0.36 \text{ MeV}/c^2$	$\pm 0.42 \text{ MeV}/c^2$	$\pm 0.32 \text{ MeV}/c^2$	$\pm 0.50 \text{ MeV}/c^2$

Table 53: List of systematic uncertainties on invariant mass resolution for B_s^0 using the interpolation method.

Systematic	2011 data	2012 data	2015 data	2016 data
Selection cuts	$\pm 0.21 \text{ MeV}/c^2$	$\pm 0.06 \text{ MeV}/c^2$	$\pm 0.05 \text{ MeV}/c^2$	$\pm 0.36 \text{ MeV}/c^2$
Mass window	$\pm 0.27 \text{ MeV}/c^2$	$\pm 0.27 \text{ MeV}/c^2$	$\pm 0.27 \text{ MeV}/c^2$	$\pm 0.27 \text{ MeV}/c^2$
Fit function of invariant mass	$\pm 0.12 \text{ MeV}/c^2$	$\pm 0.32 \text{ MeV}/c^2$	$\pm 0.17 \text{ MeV}/c^2$	$\pm 0.19 \text{ MeV}/c^2$
Total systematic error	$\pm 0.636 \text{ MeV}/c^2$	$\pm 0.42 \text{ MeV}/c^2$	$\pm 0.32 \text{ MeV}/c^2$	$\pm 0.51 \text{ MeV}/c^2$

1333

2011 data:

$$\begin{aligned}\alpha_{B^0} &= (2.605 \pm 0.056) \\ n_{B^0} &= (3.066 \pm 0.308) \\ \alpha_{B_s^0} &= (2.668 \pm 0.056) \\ n_{B_s^0} &= (2.604 \pm 0.259)\end{aligned}$$

1334

2012 data:

$$\begin{aligned}\alpha_{B^0} &= (2.625 \pm 0.029) \\ n_{B^0} &= (2.703 \pm 0.136) \\ \alpha_{B_s^0} &= (2.668 \pm 0.031) \\ n_{B_s^0} &= (2.614 \pm 0.137)\end{aligned}$$

1335 **2015 data:**

$$\begin{aligned}\alpha_{B^0} &= (2.724 \pm 0.032) \\ n_{B^0} &= (2.544 \pm 0.139) \\ \alpha_{B_s^0} &= (2.730 \pm 0.033) \\ n_{B_s^0} &= (2.553 \pm 0.143)\end{aligned}$$

1336 **2016 data:**

$$\begin{aligned}\alpha_{B^0} &= (2.744 \pm 0.033) \\ n_{B^0} &= (2.361 \pm 0.134) \\ \alpha_{B_s^0} &= (2.694 \pm 0.046) \\ n_{B_s^0} &= (2.827 \pm 0.229)\end{aligned}$$

1337 **6.5.4 Combination of Run I and Run II parameters**

As the values for all four parameters μ , σ , n and α are in good agreement between the two Run I and two Run II data sets and as the BDT PDF for signal is also calibrated on the full Run I and Run II datasets, for the signal mass PDF the parameter values are averaged.

All the parameters are combined in the following matter: We take the weighted average of the two values where we take $1/\sigma_{\text{stat}}^2$ as weighting factor. As combined statistical uncertainty we take

$$\sigma_{\text{stat;comb}} = \sqrt{\frac{1}{1/\sigma_{\text{stat};1}^2 + 1/\sigma_{\text{stat};2}^2}}$$

1338 while the combined systematic uncertainty is taken as the normal average of the systematic
1339 uncertainties of the two datasets, which follows from the assumption that systematics
1340 uncertainties are fully correlated.

1341

1342

1343 **Run I parameters:**

1344 **Central value μ :**

1345

$$\begin{aligned}\mu_{B^0} &= (5284.73 \pm 0.15_{\text{stat}} \pm 0.27_{\text{syst}}) \text{ MeV}/c^2 \\ \mu_{B_s^0} &= (5372.05 \pm 0.16_{\text{stat}} \pm 0.36_{\text{syst}}) \text{ MeV}/c^2\end{aligned}$$

1346 **Resolution σ :**

1347

$$\begin{aligned}\sigma_{B^0} &= (22.68 \pm 0.05_{\text{stat}} \pm 0.39_{\text{syst}}) \text{ MeV}/c^2 \\ \sigma_{B_s^0} &= (23.07 \pm 0.05_{\text{stat}} \pm 0.39_{\text{syst}}) \text{ MeV}/c^2\end{aligned}$$

1348 **Tail parameters α and n**

1349

$$\begin{aligned}\alpha_{B^0} &= (2.621 \pm 0.026) \\ n_{B^0} &= (2.762 \pm 0.124) \\ \alpha_{B_s^0} &= (2.668 \pm 0.027) \\ n_{B_s^0} &= (2.612 \pm 0.121)\end{aligned}$$

1350 **Run II parameters:**

1351 **Central value μ :**

1352

$$\begin{aligned}\mu_{B^0} &= (5279.95 \pm 0.13_{\text{stat}} \pm 0.08_{\text{syst}}) \text{ MeV}/c^2 \\ \mu_{B_s^0} &= (5367.34 \pm 0.14_{\text{stat}} \pm 0.35_{\text{syst}}) \text{ MeV}/c^2\end{aligned}$$

1353 **Resolution σ :**

1354

$$\begin{aligned}\sigma_{B^0} &= (22.46 \pm 0.08_{\text{stat}} \pm 0.41_{\text{syst}}) \text{ MeV}/c^2 \\ \sigma_{B_s^0} &= (22.85 \pm 0.08_{\text{stat}} \pm 0.42_{\text{syst}}) \text{ MeV}/c^2\end{aligned}$$

1355 **Tail parameters α and n**

1356

$$\begin{aligned}\alpha_{B^0} &= (2.734 \pm 0.023) \\ n_{B^0} &= (2.449 \pm 0.096) \\ \alpha_{B_s^0} &= (2.718 \pm 0.027) \\ n_{B_s^0} &= (2.630 \pm 0.121)\end{aligned}$$

1357 **7 Normalisation**

1358 The signals branching fractions are normalised to $B^+ \rightarrow J/\psi K^+$ (where $J/\psi \rightarrow \mu^+ \mu^-$)
 1359 and $B^0 \rightarrow K^+ \pi^-$ decay channels. An alternative would be to use the $b\bar{b}$ production
 1360 cross-section and the collected luminosity of the sample and express the signal branching
 1361 fraction directly, but given the current precision of the $b\bar{b}$ production cross-section and the
 1362 luminosity measurement, the result has around three times larger uncertainty. Therefore,
 1363 the signals branching fractions are expressed as

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = \underbrace{\frac{\mathcal{B}_{norm}}{N_{norm}} \times \frac{\epsilon_{norm}}{\epsilon_{sig}} \times \frac{f_d}{f_s}}_{\alpha_s} \times N_{B_s^0 \rightarrow \mu^+ \mu^-}, \quad (19)$$

$$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) = \alpha_d \times N_{B^0 \rightarrow \mu^+ \mu^-},$$

1364 where α_s and α_d are the normalisation factors for $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$. These
 1365 factors are calculated separately for each normalisation channel and collision energy, and
 1366 are finally combined.

1367 Among other sources of systematic uncertainties, (19) also avoids the detection effi-
 1368 ciencies common to signal and normalisation channels. This motivates the choice of a
 1369 normalisation channel “as similar as possible” to the signal decays, with common trigger,
 1370 reconstruction, and selection procedures. $B^+ \rightarrow J/\psi K^+$ and $B^0 \rightarrow K^+ \pi^-$ decays were
 1371 chosen because of their relative abundance and well measured branching fractions. In
 1372 addition, the $B^+ \rightarrow J/\psi K^+$ candidates are often triggered by the same muon trigger lines
 1373 than the signal¹². This will result in a very similar trigger efficiency ratio in (19), and allows
 1374 to determine the signal trigger efficiency from the $B^+ \rightarrow J/\psi K^+$ data. The $B^0 \rightarrow K^+ \pi^-$,
 1375 on the other hand, has two particles in the final state and will have reconstruction and
 1376 topological selection efficiencies very similar to the signal decays.

1377 The detection efficiencies in (19) are split according to the detection stages: detector
 1378 acceptance (i.e. generation level), reconstruction and selection, and trigger, such that

$$\frac{\epsilon_{norm}}{\epsilon_{sig}} = \frac{\epsilon_{norm}^{Acc}}{\epsilon_{sig}^{Acc}} \times \frac{\epsilon_{norm}^{RecSel|Acc}}{\epsilon_{sig}^{RecSel|Acc}} \times \frac{\epsilon_{norm}^{Trig|RecSel}}{\epsilon_{sig}^{Trig|RecSel}}, \quad (20)$$

1379 where the efficiency for each subsequent stage is estimated for events that pass the previous
 1380 stages.

1381 The $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ detection efficiencies were studied on simulated
 1382 samples. In the previous analysis they were found to be equal within 1% for all the three

¹²The lines in the first two trigger levels are identical to the the signal decays: L0Muon or L0Dimuon and Hlt1TrackMuon trigger 93% of the selected and triggered $B_s^0 \rightarrow \mu^+ \mu^-$ and 95% of the selected and triggered $B^+ \rightarrow J/\psi K^+$ candidates. If a positive Hlt2DiMuonDecision is required, the fraction is 83% for both. In addition to the Hlt2DiMuonDecision, there are many more muon lines Hlt2 triggering on both decays (e.g. Single muon lines) and raising the percentage on similarly triggered candidates

1383 terms in (20). This analysis is re-optimised for the $B^0 \rightarrow \mu^+ \mu^-$ mode, and the efficiencies
1384 are estimated separately to account for the small differences.

1385 The main uncertainty in the normalisation procedure arises from the hadronisation
1386 fraction ratio f_s/f_d . The hadronisations fraction could be avoided if another B_s^0 decay
1387 was used as a normalisation channel. The possibility of including $B_s^0 \rightarrow J/\psi \phi$ as a
1388 third normalisation channel was studied. The improvement was found negligible due
1389 to the uncertain branching fraction, even after including the newest branching fraction
1390 measurement from BELLE [34]. The $B_s^0 \rightarrow J/\psi \phi$ was used to check the stability of the
1391 relative B_s^0 and B^0 production (f_s/f_d) at 8 and 13 TeV (see Sect. 7.6).

1392 The calculation of the various normalisation factor components are described in the
1393 following sections: the normalisation-channel candidate yields in Sect. 7.1; geometrical
1394 detector acceptance for the signal and normalisation channels in Sect. 7.2; reconstruction
1395 and selection efficiencies for the signal and normalisation channels in Sect. 7.3; trigger
1396 efficiencies for the signal and normalisation channels in Sect. 7.4; the hadronisation fractions
1397 ratio stability in Sect. 7.6; and the normalisation factors are calculated in Sect. 7.7.

1398 The normalisation procedure has previously been described in [6], and in a recent PhD
1399 thesis [35].

1400 7.1 Normalisation channels yields

1401 Normalisation channels candidates and the $B_s^0 \rightarrow \mu^+ \mu^-$ signal candidates are required to
1402 pass identical selection criteria (except for the muon ProbNN requirements, which are
1403 only applied on the signal channels, see Sect. 4). The selected normalisation channels
1404 candidates are thereafter separated from the background with a Maximum Likelihood fit
1405 on the reconstructed B meson mass distribution.

1406 7.1.1 $B^+ \rightarrow J/\psi K^+$ channel

1407 The mass model for the $B^+ \rightarrow J/\psi K^+$ channel uses the B mass reconstructed with
1408 a constraint on the J/ψ mass (B_{JCMass}). The model consists of signal ($B^+ \rightarrow J/\psi K^+$
1409), combinatorial background, and $B^+ \rightarrow J/\psi \pi^+$ candidates wrongly reconstructed as
1410 $B^+ \rightarrow J/\psi K^+$ (no PID cuts are used for $B^+ \rightarrow J/\psi K^+$ selection to reduce the differences
1411 to the signal selection). The signal component shape was studied on the simulated
1412 $B^+ \rightarrow J/\psi K^+$ decays passing the $BKGCAT < 10 || BKGCAT == 50$ requirement
1413 (*matched* in the following). The signal mass distribution was found to be best described
1414 by the ROOFIT Ipatia function. When fitting the data samples, the signal shape was
1415 fluctuated within the Gaussian constraints defined using parameter values and uncertainties
1416 from fits to the matched Monte Carlo samples, separately for each year's conditions (shown
1417 for MC12 in Fig. 53).

1418 The mass shape of the $B^+ \rightarrow J/\psi \pi^+$ background component was determined on a
1419 simulated $B^+ \rightarrow J/\psi \pi^+$ sample, after reconstructing the candidates under $B^+ \rightarrow J/\psi K^+$
1420 hypothesis. The shape is not well described by any parametric forms (Crystal Ball,
1421 Gaussian combinations) and the Gaussian Kernel estimation (RooKeysPdf, mirror = 1,

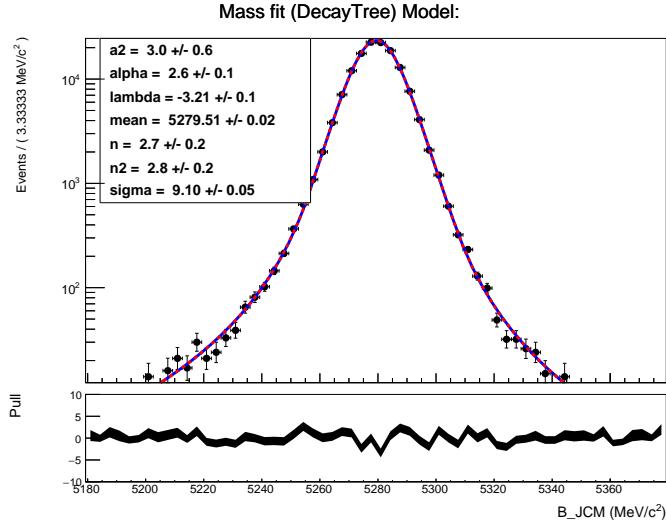


Figure 53: Invariant mass distribution for the simulated signal candidates in the reconstructed, selected and matched MC12 $B^+ \rightarrow J/\psi K^+$ sample. The continuous blue curve denotes the fitted `Ipatia` function. Different simulation versions are used.

1422 rho=1) was used instead. The resulting distribution after the selection (and $B^+ \rightarrow J/\psi \pi^+$
 1423 matching) in shown in Fig. 54. Two shapes are extracted: one on the matched and one
 1424 on the non-matched sample; the shape from the latter is used in an alternative and the
 1425 difference in the signal yield is assined as a systematic uncertainty.

1426 The $B^+ \rightarrow J/\psi \pi^+$ yield in the $B^+ \rightarrow J/\psi K^+$ mass fit is linked to the signal yield using
 1427 the measured ratio of $R^{\pi K} = \frac{\mathcal{B}(B^+ \rightarrow J/\psi \pi^+)}{\mathcal{B}(B^+ \rightarrow J/\psi K^+)} = (3.83 \pm 0.11 \pm 0.07) \times 10^{-2}$ from Ref. [36],
 1428 and correcting it for the ratio of the selection efficiencies:

$$f_\pi = \frac{N(B^+ \rightarrow J/\psi \pi^+)}{N(B^+ \rightarrow J/\psi K^+)} = R^{\pi K} \times \frac{\epsilon^{B^+ \rightarrow J/\psi \pi^+}}{\epsilon^{B^+ \rightarrow J/\psi K^+}} = 2.8 \pm 0.1\%. \quad (21)$$

1429 The fraction is Gaussian-constrained in the fit.

1430 The possible mis-reconstructed background from $B_c^+ \rightarrow J/\psi K^+ K^- \pi^+$ (and $\Lambda_b^0 \rightarrow$
 1431 $p J/\psi K^-$) was studied and found to be small and uniformly distributed in the invariant
 1432 mass, as shown in Fig. 54. No additional component was included to the model as the small
 1433 contribution in the selected sample will be accounted for by the exponential component.

1434 The combinatorial background is described with a single exponential functions. The
 1435 slope of the exponential was left free, as was the yield of the combinatorial background.

1436 The fits to the $B^+ \rightarrow J/\psi K^+$ candidates in the 2011, 2012, and 2015 data are shown
 1437 in Fig. 55. The triggered, reconstructed and selected $B^+ \rightarrow J/\psi K^+$ candidate yields in

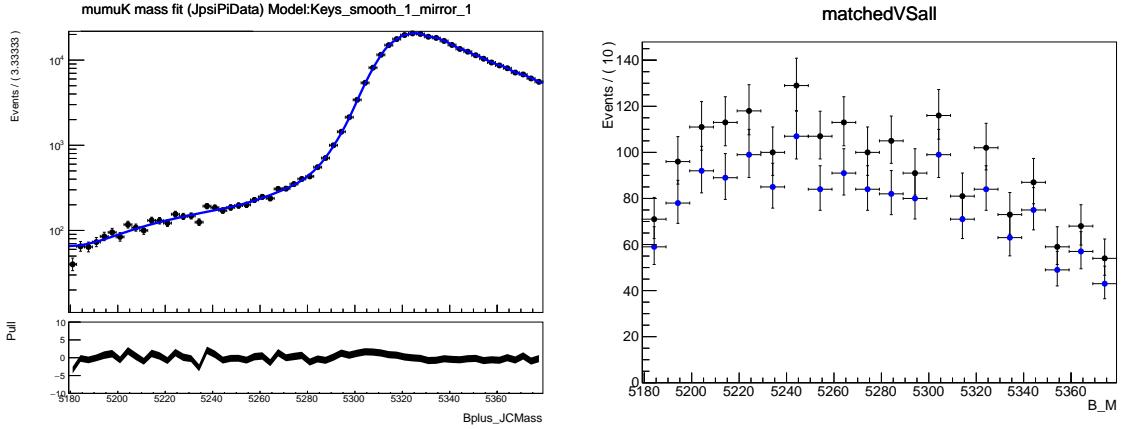


Figure 54: Invariant mass distributions of the simulated background candidates, shown for matched $B^+ \rightarrow J/\psi \pi^+$ candidates (left) and for matched and non-matched $B_c^+ \rightarrow J/\psi K^+ K^- \pi^+$ candidates (right). In both cases, the candidates are reconstructed and selected as $B^+ \rightarrow J/\psi K^+$. The continuous blue curve on the left shown the fitted `RooKeysPdf` function.

1438 the samples are:

$$\begin{aligned}
 N_{B^+ \rightarrow J/\psi K^+}^{2011} &= 347312 \pm 622_{\text{stat.syst}} \pm 121_{\text{match.}}, \\
 N_{B^+ \rightarrow J/\psi K^+}^{2012} &= 774282 \pm 928_{\text{stat.syst}} \pm 54_{\text{match.}}, \\
 N_{B^+ \rightarrow J/\psi K^+}^{2015} &= 166689 \pm 432_{\text{stat.syst}} \pm 19_{\text{match.}}, \\
 N_{B^+ \rightarrow J/\psi K^+}^{2016} &= 684335 \pm 878_{\text{stat.syst}} \pm 22_{\text{match.}},
 \end{aligned} \tag{22}$$

1439 where the first uncertainty terms have been determined using the alternative MLL fit
 1440 model with the $B^+ \rightarrow J/\psi \pi^+$ background shape from a non-matched MC sample, and
 1441 the second terms include all the Gaussian constrained fit model systematics as well as
 1442 statisticsl uncertainties.

1443 The 2011 and 2012 $B^+ \rightarrow J/\psi K^+$ yields have been determined using the Leptonic
 1444 stripping stream, whereas the signal (and $B^0 \rightarrow K^+ \pi^-$) pass the di-muon stream. That
 1445 means there are two sets of ntuples that could contain different numbers of events depending
 1446 on the tuple-production (failed jobs, etc). The luminosity difference between leptonic and
 1447 dimuon streams in 2011 and 2012 (S21r1 and S21) For 2011 data, 1.1% more luminosity is
 1448 included in the Leptonic stream. For 2012 data, 0.57% more luminosity is included in the
 1449 Leptonic stream.

1450 7.1.2 $B^0 \rightarrow K^+ \pi^-$ channel

1451 The $B^0 \rightarrow K^+ \pi^-$ yield is evaluated summing the corrected yield reported in
 1452 Tabs. 42, 43, 44, 45. The sum of the yields used for the normalisation are reported
 1453 in Tab. ??.

1454 The systematic uncertainty are evaluated in the same way as they were for the BDT
 1455 calibration in Sec. Sect. 6.4. An additional systematic uncertainty is evaluated comparing

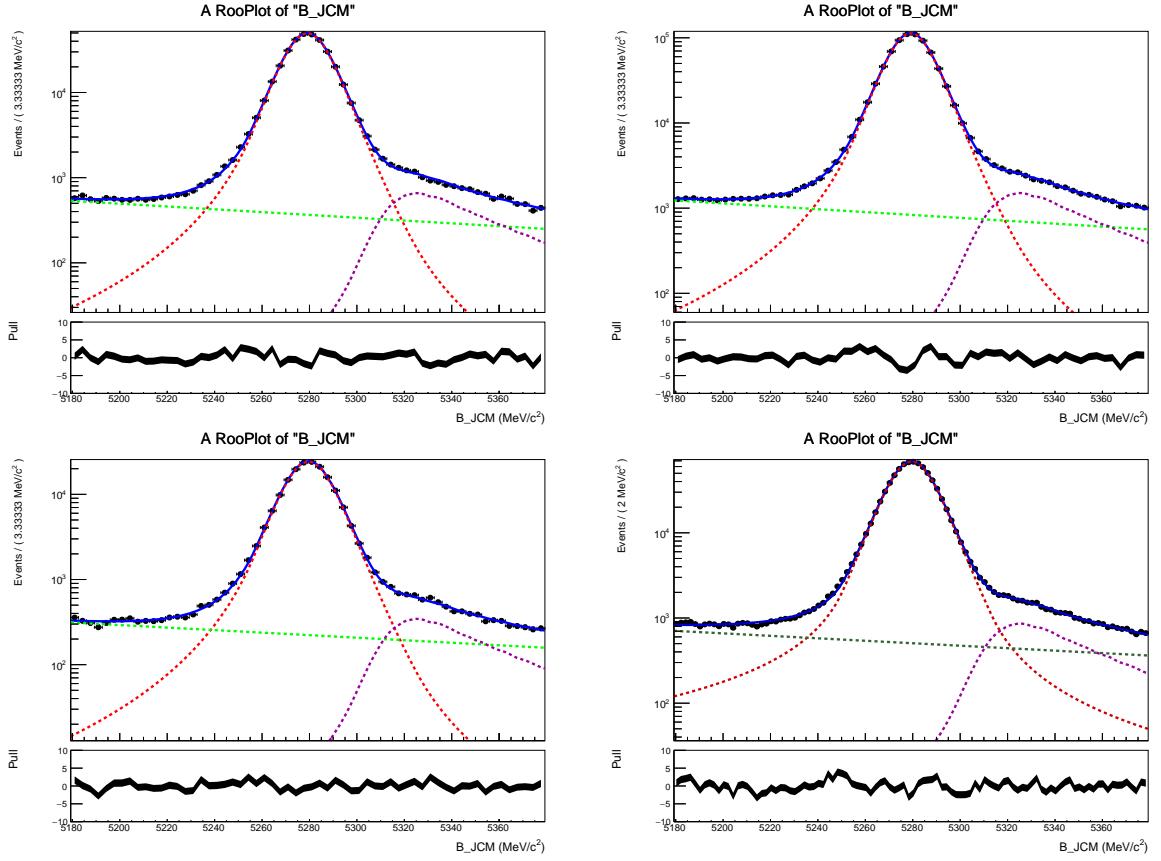


Figure 55: The fitted invariant mass distribution of the $B^+ \rightarrow J/\psi \pi^+$ candidates in 2011 (top left, S21r1), 2012 (top right, S21), 2015 (bottom left, S24) and 2016 (bottom right, S26) data. The $B^+ \rightarrow J/\psi K^+$ component is denoted with a dashed red line, mis-reconstructed $B^+ \rightarrow J/\psi \pi^+$ background component with a dashed magenta line, and the combinatorial background component with a dashed green line. The solid blue line denotes the total likelihood model; the data are shown with black round markers.

Table 54: Luminosities in the Leptonic and Dimuon streams for 2011 and 2012 data.

DataSet	Leptonic stream	Dimuon stream	Ratio
2011 (S21r1)	973.3 pb^{-1}	962.7 pb^{-1}	0.989
2012 (S21)	1984.6 pb^{-1}	1973.2 pb^{-1}	0.994

1456 the $B^0 \rightarrow K^+ \pi^-$ yields in the exclusive and inclusive fits in all the BDT bins excluding
1457 bin 1a and scaled with the full BDT range.

1458 The $B^0 \rightarrow K^+ \pi^-$ yields are determined after requiring a TOS decision from the HLT2

1459 B2HH line. This line is chosen because it has a high efficiency on $B^0 \rightarrow K^+ \pi^-$ (at least
1460 70%). The following yields are obtained:

$$\begin{aligned} N_{B^0 \rightarrow K^+ \pi^-}^{2011} &= 6833 \pm 242_{\text{stat.}} \pm 503_{(B \rightarrow h^+ h' - \text{yield})} \pm 341_{(\text{Fit Model})} \pm 99_{(\text{PID Binning})}, \\ N_{B^0 \rightarrow K^+ \pi^-}^{2012} &= 18012 \pm 473_{\text{stat.}} \pm 891_{(B \rightarrow h^+ h' - \text{yield})} \pm 189_{(\text{Fit Model})} \pm 662_{(\text{PID Binning})}, \\ N_{B^0 \rightarrow K^+ \pi^-}^{2015} &= 8552 \pm 305_{\text{stat.}} \pm 306_{(B \rightarrow h^+ h' - \text{yield})} \pm 285_{(\text{Fit Model})} \pm 672_{(\text{PID Binning})}, \\ N_{B^0 \rightarrow K^+ \pi^-}^{2016} &= 28411 \pm 549_{\text{stat.}} \pm 1767_{(B \rightarrow h^+ h' - \text{yield})} \pm 361_{(\text{Fit Model})} \pm 1714_{(\text{PID Binning})} \end{aligned} \quad (23)$$

1461 7.2 Geometrical detector acceptance

1462 In first approximation, the detector acceptance is defined as the fraction of the decays
1463 having all the decay products in the geometrical detector acceptance. The total detector
1464 acceptance, however, will also be affected by the magnetic field and by the interactions
1465 with the detector material; these effects will be evaluated as part of the reconstruction
1466 efficiency in Sect. 7.3.

1467 The geometrical detector acceptances for the signal and normalisation channels have
1468 been estimated with simulated samples, and are listed in Table 55. The decay products
1469 are required to fly in the LHCb detector acceptance, defined by the polar angle in the
1470 range of [10, 400] mrad, which is chosen to be larger than the physical LHCb detector
1471 acceptance in order to allow for the recovery of particles by the magnetic field.

1472 As expected from the kinematic distribution of the final decay products, the geometrical
1473 acceptance efficiencies in Table 55 are similar for the signal and the $B^0 \rightarrow K^+ \pi^-$ decay,
1474 but lower for $B^+ \rightarrow J/\psi K^+$.

1475 7.3 Ratio of reconstruction and selection efficiencies

1476 The reconstruction efficiency is the fraction of decay candidates in the detector acceptance
1477 that are successfully reconstructed; the selection efficiency is the fraction of reconstructed
1478 decay candidates that pass the selection (Sect. 4). Both efficiencies depend on the
1479 characteristics of the decay channel: the number of particles in the final state, their
1480 kinematic distributions, the track finding algorithm efficiency, the particle identification
1481 efficiency, etc.

1482 The determination of the selection and reconstruction efficiency is based both on
1483 simulations and collected data.

1484 The combined effect of reconstruction and selection is first evaluated on simulated
1485 samples, considering only matched candidates in the detector acceptance (see Table 56).
1486 The efficiencies differences between MC11, MC12, and MC15 are not only due to the
1487 different collision energy, but largely also due to different simulation versions. What
1488 matters, is the ratio of the signal and normalisation efficiencies for each year. Therefore,
1489 samples with the same simulation version have been used for each year.

1490 **7.3.1 Corrections to the simulated selection efficiency**

1491 The selection criteria for signal and normalisation channels are kept as similar as possible
1492 (Sect. 4). The selection efficiency is determined partly on simulated samples, and partly on
1493 data. All differences that could lead to wrong efficiency estimates in case the simulation
1494 fails to reproduce the correct selection variable distributions were studied, and, if necessary,
1495 corrected for.

1496 *Tracking efficiency.* The efficiency to reconstruct a particle track is measured on the
1497 data and expressed as a correction to that from the simulation, separately for muons
1498 and hadrons. The individual muon- and hadron-track reconstruction efficiency ratios
1499 are combined according to the final state composition, and multiplied by the simulated
1500 reconstruction efficiencies to obtain the reconstruction efficiency for a given decay channel.

1501 The muon-track reconstruction efficiency is determined on a clean sample of $J/\psi \rightarrow$
1502 $\mu^+ \mu^-$ decays with the tag-and-probe method¹³, described in Ref. [37]; the hadron-track re-
1503 construction efficiency is determined by modifying the measured muon-track reconstruction
1504 efficiency to account for the hadron interactions with the detector material. This is done by
1505 means of simulations. Because of possible inaccuracies in the simulated detector material
1506 budget, the simulation of hadron interactions with the detector material introduces the
1507 largest systematic uncertainty to the selection and reconstruction efficiency.

1508 The reconstruction efficiency depends on the kinematics, which differs from final state
1509 to final state. The muon- and hadron-track efficiencies are determined in bins of track
1510 pseudo-rapidity and momentum. The resulting muon-, kaon-, and pion-tracking efficiency
1511 maps, when applied to the simulated signal and normalisation channel samples, take into
1512 account the momentum and pseudo-rapidity distributions of the final state particles.

1513 The track reconstruction efficiencies measured from data yield the (multiplicative)
1514 correction factors for the different Monte Carlo versions (see Table 57).

1515 The 2012 efficiency maps¹⁴ were used also for 2011. 2015 corrections are calculated
1516 using the latest 2015 maps¹⁵. The 2015 maps are also used for MC16 Sim09a¹⁶. The
1517 uncertainties in Table 57 contain statistical uncertainties, systematic uncertainties arising
1518 from the tag-and-probe procedure ($\pm 0.4\%$ per track in Run 1, 0.8% in Run2) and the
1519 uncertainties arising from the simulation of the hadronic interactions with the detector
1520 material ($\pm 1.1\%$ and per kaon and $\pm 1.4\%$ per pion track).

1521 The choice of the *mass range* ($[4900, 6000] \text{ MeV}/c^2$) might have an effect on the signal
1522 efficiency depending on the spread of the signal events in the di-muon spectrum. The

¹³The tag-and-probe method uses two-prong decays, where one of the decay products, the “tag”, is fully reconstructed as a track, while the particle, the “probe”, is only partially reconstructed, i.e. not using the tracking information with which the tracking efficiency will be evaluated.

¹⁴<https://twiki.cern.ch/twiki/pub/LHCb/TrackingEffStatus2012S20/ratio2012S20.root>,
11/04/2013

¹⁵https://twiki.cern.ch/twiki/pub/LHCb/LHCbTrackingEfficiencies/Ratio_2015_25ns_Long.root,
08/10/2016

¹⁶There is a known issue with Sim09a <https://indico.cern.ch/event/506012/contributions/2025750/attachments/1025750/1713853/>. Due to time constraints, waiting for the Sim09b MC16 is not an option and we are waiting for additional systematics to be assigned due to the problems in the VELO simulation in MC16 Sim09a.

1523 effect of the mass range cut has been previously studied on the simulated signal samples,
 1524 accounting for possible inaccuracies in simulating mass resolution effects [6]. No efficiency
 1525 correction was found necessary for the signal channels. The same procedure was repeated
 1526 for the normalisation channels by investigating the effect of the di-muon mass cut on a
 1527 $B^+ \rightarrow J/\psi K^+$ sample. Again, the correction was found negligible because of the narrow
 1528 J/ψ mass peak.

1529 The IP *distribution* of the simulated samples differs from the distribution observed in
 1530 data. If the differences are large, they could lead to incorrect reconstruction and selection
 1531 efficiency. As described in the previous analysis note [6], we reweighted the simulated IP
 1532 distribution to match the measured IP distribution and studied the effect of reweighting
 1533 on the reconstruction and selection efficiencies. These were found to be dependent on the
 1534 IP distribution reweighting but the effect was shown to cancel in the ratio. No additional
 1535 correction is therefore necessary necessary for the normalisation purposes. Also, the
 1536 simulated distribution of the Kaon IP χ^2 in $B^+ \rightarrow J/\psi K^+$ decay was studied. A cut on
 1537 the kaon IP χ^2 is used only for $B^+ \rightarrow J/\psi K^+$ selection, and therefore is not “balanced”
 1538 in the efficiency ratio. Comparison between the distributions in data and in simulation
 1539 showed no need for an additional correction.

1540 The *decay-time distribution* of the B_s^0 mesons in the Monte Carlo
 1541 $B_s^0 \rightarrow \mu^+ \mu^-$ sample is simulated using a single exponential with a mean B_s^0 life-
 1542 time. The heavy and light B_s^0 meson mass eigenstates have different lifetimes; the relative
 1543 decay-width asymmetry is measured to be non-zero in the B_s^0 system [38, 39]. Because the
 1544 $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ reconstruction and selection (Table 56) depends on the B meson lifetime,
 1545 the signal selection efficiency determined from the simulated $B_s^0 \rightarrow \mu^+ \mu^-$ sample (with a
 1546 zero decay-width asymmetry) must be corrected for [40]. The correction factor for the
 1547 $B_s^0 \rightarrow \mu^+ \mu^-$ selection efficiency is physics-model dependent, and is obtained assuming
 1548 the Standard Model value of $\mathcal{A}_{\Delta\Gamma} = 1$ for $B_s^0 \rightarrow \mu^+ \mu^-$. It means that only the heavy B_s^0
 1549 mass eigenstate decays into two muons. Its lifetime distribution is modelled with a single
 1550 exponential.

1551 As the full detection efficiency in lifetime bins is unchanged and can be determined
 1552 from the simulated sample, the corrected reconstruction and selection efficiency can be
 1553 simply obtained from the exponentials with the SM lifetime:

$$\epsilon_{SM}^{RecSel} = \sum_i^{\tau(B_s)} (e^{-t_i^{low}/\tau_{SM}} - e^{-t_i^{hi}/\tau_{SM}}) \times \epsilon_{MC}^{RecSel}(i). \quad (24)$$

1554 The reconstruction and selection efficiency in each of the 200 lifetime bins is computed
 1555 using the $B_s^0 \rightarrow \mu^+ \mu^-$ MC¹⁷. The resulting multiplicative decay time corrections applied to
 1556 the original MC efficiencies are listed in Table 58.

1557 *Muon identification.* The acceptance of the muon detector and the efficiency of the
 1558 `isMuon` algorithm in identifying true muon tracks is included in the simulated reconstruction
 1559 and selection efficiencies (Table 56). These efficiencies have also been determined using the

¹⁷ A similar check was performed on $B^0 \rightarrow \mu^+ \mu^-$. The MC lifetime is in good agreement with the B^0 lifetime in PDG2016 and no corrections are necessary.

1560 tag-and-probe technique on the $B^+ \rightarrow J/\psi K^+$ sample and included as correction factors
 1561 to the simulated reconstruction and selection efficiencies. The muon detector acceptance
 1562 is also applied to $B^0 \rightarrow K^+\pi^-$, but no correction is included as the simulated efficiency
 1563 has previously been found to be identical to the acceptance efficiency determined on the
 1564 data. The muon identification efficiencies for the signals modes $B_{(s)}^0 \rightarrow \mu^+\mu^-$ also include
 1565 the particle identification requirements (ProbNN). These are not included in the simulated
 1566 reconstruction and selection efficiencies in Table 56, but will be included in the signal
 1567 selection through the correction factors; these factors are computed for each channel as
 1568 ratios between the MC and Data efficiencies in Table 59.

1569 **7.4 Ratio of trigger efficiencies**

1570 The trigger efficiencies are directly determined from data with the TISTOS method, as
 1571 described in [41]. They are calculated for events that have been reconstructed and selected.

1572 The majority of the recorded $B^+ \rightarrow J/\psi K^+$ and $B_{(s)}^0 \rightarrow \mu^+\mu^-$ candidates are triggered
 1573 by the same¹⁸ muon trigger lines (Ch. 3 in [35]). The muon trigger line¹⁹ efficiency is
 1574 measured with the TISTOS method directly from the $B^+ \rightarrow J/\psi K^+$ sample in four-by-
 1575 four grid defined by the p_T and IP of the maximum p_T muon. For the TIS sample we
 1576 consider events with any TIS decision per trigger level, that is not only the muon lines but
 1577 LOGlobalTIS and Hlt1PhysTIS and Hlt2PhysTIS. The trigger efficiency for the signal
 1578 channels $B_{(s)}^0 \rightarrow \mu^+\mu^-$, is found by multiplying the result with the muon spectrum from

¹⁸The lines in the first two trigger levels are identical to the the signal decays: L0Muon or L0Dimuon and Hlt1TrackMuon trigger 93% of the selected $B_s^0 \rightarrow \mu^+\mu^-$ and 95% of the selected $B^+ \rightarrow J/\psi K^+$ candidates. If Hlt2DiMuonDecision line is specified, the fraction is 83% for both. However, there are many more muon lines in Hlt that trigger on both decays (e.g. Single muon lines).

¹⁹Muon lines considered here are: B_L0MuonDecision, B_L0DiMuonDecision, B_Hlt1SingleMuonNoIPDecision, B_Hlt1SingleMuonHighPTDecision, B_Hlt1TrackMuonDecision, B_Hlt1DiMuonLowMassDecision, B_Hlt1DiMuonHighMassDecision, B_Hlt2SingleMuonDecision, B_Hlt2SingleMuonLowPTDecision, B_Hlt2SingleMuonHighPTDecision, B_Hlt2DiMuonDecision, B_Hlt2DiMuonLowMassDecision, B_Hlt2DiMuonJPsiDecision, B_Hlt2DiMuonJPsiHighPTDecision, B_Hlt2DiMuonPsi2SDecision, B_Hlt2DiMuonDetachedDecision, B_Hlt2DiMuonDetachedJPsiDecision, B_Hlt2DiMuonDetachedHeavyDecision, B_Hlt2TriMuonTauDecision, B_Hlt2DiMuonBDecision, B_Hlt2DiMuonZDecision, B_Hlt2DiMuonDYDecision.

1579 simulated $B_s^0 \rightarrow \mu^+ \mu^-$ or $B^0 \rightarrow \mu^+ \mu^-$ decays:

$$\begin{aligned}
\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{TRIG|RecSel}(2011) &= (94.2 \pm 0.7_{stat} \pm 1.9_{syst})\%, \\
\epsilon_{B^0 \rightarrow \mu^+ \mu^-}^{TRIG|RecSel}(2011) &= (94.1 \pm 0.7_{stat} \pm 2.1_{syst})\%, \\
\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{TRIG|RecSel}(2012) &= (94.1 \pm 0.4_{stat} \pm 2.0_{syst})\%, \\
\epsilon_{B^0 \rightarrow \mu^+ \mu^-}^{TRIG|RecSel}(2012) &= (93.9 \pm 0.4_{stat} \pm 2.1_{syst})\%, \\
\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{TRIG|RecSel}(2015) &= (96.1 \pm 0.7_{stat} \pm 3.1_{syst})\%, \\
\epsilon_{B^0 \rightarrow \mu^+ \mu^-}^{TRIG|RecSel}(2015) &= (95.9 \pm 0.7_{stat} \pm 3.2_{syst})\%, \\
\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{TRIG|RecSel}(2016) &= (97.3 \pm 0.3_{stat} \pm 1.3_{syst})\%, \\
\epsilon_{B^0 \rightarrow \mu^+ \mu^-}^{TRIG|RecSel}(2015) &= (97.1 \pm 0.3_{stat} \pm 1.5_{syst})\%,
\end{aligned} \tag{25}$$

1580 where the systematic uncertainty stems from the small fraction of $B_s^0 \rightarrow \mu^+ \mu^-$ events that
1581 are not triggered by the muon trigger lines.

1582 The total trigger efficiency for $B^+ \rightarrow J/\psi K^+$ is the **TISTOS** efficiency determined using
1583 all the trigger lines:

$$\begin{aligned}
\epsilon_{B^+ \rightarrow J/\psi K^+}^{TRIG|RecSel}(2011) &= (87.9 \pm 0.7_{stat} \pm 0.5_{syst})\%, \\
\epsilon_{B^+ \rightarrow J/\psi K^+}^{TRIG|RecSel}(2012) &= (87.5 \pm 0.5_{stat} \pm 0.7_{syst})\%, \\
\epsilon_{B^+ \rightarrow J/\psi K^+}^{TRIG|RecSel}(2015) &= (88.8 \pm 0.8_{stat} \pm 2.1_{syst})\%, \\
\epsilon_{B^+ \rightarrow J/\psi K^+}^{TRIG|RecSel}(2016) &= (90.7 \pm 0.3_{stat} \pm 0.9_{syst})\%,
\end{aligned} \tag{26}$$

1584 where the systematic uncertainty also includes the uncertainty of the **TISTOS** method²⁰.
1585 $B^0 \rightarrow K^+ \pi^-$ candidates are required to be triggered independently of the signal
1586 candidate in the first trigger levels, L0 and HLT1. This reduces the differences between the
1587 signal and the $B^0 \rightarrow K^+ \pi^-$ trigger lines, and, avoiding the HLT2 requirements, provides
1588 enough statistics for the normalisation). The efficiency to trigger independent of the signal
1589 candidate is by definition the same for all B decays²¹. The L0 and HLT1 TIS efficiency
1590 can thus be measured from $B^+ \rightarrow J/\psi K^+$ decays:

²⁰The **TISTOS** bias on MC15 is larger. The trigger conditions have changed and differences are expected (e.g. the TIS efficiency in 2015 is higher than during the previous years). As the reason for the larger bias is unknown, it will be assigned as a systematic. The effect on the total normalisation is very small. The **TISTOS** bias is only used for $B^+ \rightarrow J/\psi K^+$ in the normalisation ratio w.r.t the signal channels trigger efficiencies.

²¹Once the kinematic distribution of the decaying meson is unfolded, as discussed in [41].

$$\begin{aligned}
\epsilon_{L0HLT1}^{TIS}(2011) &= (4.97 \pm 0.04_{stat} \pm 0.03_{syst})\%, \\
\epsilon_{L0HLT1}^{TIS}(2012) &= (5.69 \pm 0.03_{stat} \pm 0.03_{syst})\%, \\
\epsilon_{L0HLT1}^{TIS}(2015) &= (9.21 \pm 0.08_{stat} \pm 0.22_{syst})\%, \\
\epsilon_{L0HLT1}^{TIS}(2016) &= (7.94 \pm 0.03_{stat} \pm 0.08_{syst})\%,
\end{aligned} \tag{27}$$

where the systematic uncertainty from the TISTOS method. The increase in the L0Hlt1 TIS efficiency in 2015 and 2016 was confirmed by using the TISTOS method on the selected $B^0 \rightarrow K^+ \pi^-$ candidates decays in 2012, 2015, 2016 data.

The HLT2 trigger efficiency is estimated from a simulated $B^0 \rightarrow K^+ \pi^-$ sample, for events passing the L0 and HLT1 TIS requirement:

$$\begin{aligned}
\epsilon_{B^0 \rightarrow K^+ \pi^-}^{HLT2B2HHTOS|L0HLT1TIS}(2011) &= (72.31 \pm 0.38_{stat})\%, \\
\epsilon_{B^0 \rightarrow K^+ \pi^-}^{HLT2B2HHTOS|L0HLT1TIS}(2012) &= (75.49 \pm 0.10_{stat})\%. \\
\epsilon_{B^0 \rightarrow K^+ \pi^-}^{HLT2B2HHTOS|L0HLT1TIS}(2015) &= (88.85 \pm 0.11_{stat})\%. \\
\epsilon_{B^0 \rightarrow K^+ \pi^-}^{HLT2B2HHTOS|L0HLT1TIS}(2016) &= (91.20 \pm 0.07_{stat})\%.
\end{aligned} \tag{28}$$

The numbers originate from different simulation versions: Sim08 for 2011 and 2012, and Sim09 for 2015. As these efficiencies enter in ratios together with $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ trigger efficiencies determined from the data, the simulation versions could matter. Even though the effect of different simulation versions on the HLT2 efficiency was found to be small (the MC12 Sim06 MC used for the reconstruction and selection efficiencies gave a very similar result of 89.06(16)%), MC12 with Sim08b was used for consistency.

The TIS efficiency is by definition channel independent. As a crosscheck, the numbers were validated by applying the TISTOS method on the B 2hh candidates in the data. The results confirm the numbers in (28).

As a change w.r.t. the last analysis, we consider here a specific Hlt2 line (B2HH_TOS). Initially, we suspected that the inclusive Hlt2 Phys trigger efficiency does not properly account for the prescales in Run2 data. That turned out not to be the case and the results from the inclusive Phys selection are similar. Nevertheless, focusing on a specific trigger lines is a more reliable practice and the Hlt2B2HHDecision_TOS will be used in the final normalisation factor calculation.

7.5 Normalisation cross-checks

Before calculating the normalisation factors, the detection efficiencies and the measured normalisation channel yields are validated.

The detection efficiencies ratios (without PID cuts for $B_{(s)}^0 \rightarrow \mu^+ \mu^-$) are expected to remain stable over the various data acquisition periods. Small differences can arise in ratios that include decay channels with different numbers of particles in the final states.

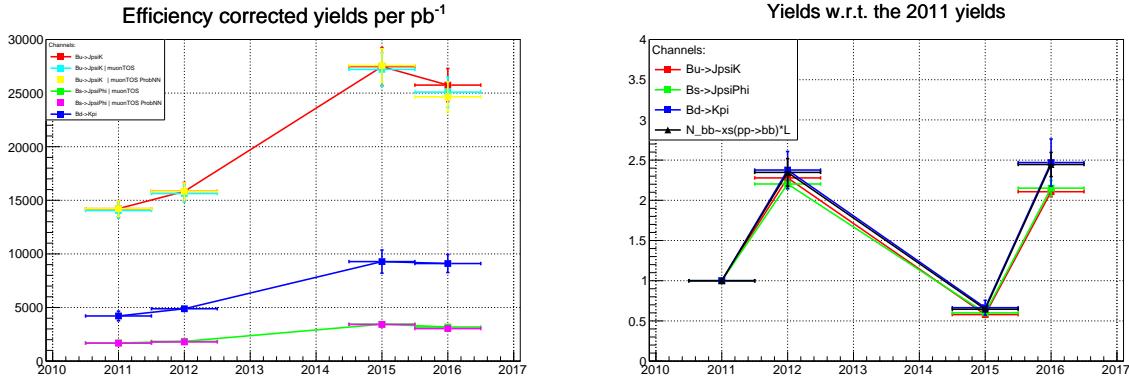


Figure 56: **Left:** Efficiency corrected yields per pb^{-1} for different trigger selections and muon identification criteria in each year. Luminosities are 962.67, 1973.24, 290.24, 1100 for 2011, 2012, 2015 and 2016. A generous systematic of 5% is used for the luminosities. The small discrepancy in 2016 between ProbNN inclusive/exclusive cases, using TrigDec (red) and MuonTOS (yellow) trigger decisions respectively, arises from the different trigger efficiency estimation procedure. Same procedure and MuonTOS requirement leads to a good agreement (yellow VS cyan). The trigger efficiency estimation used on the TrigDec (red) yields relies completely on the data and is used in the normalisation. **Right:** Yields in each year, normalised to the 2011 yields. The number of produced $b\bar{b}$ pairs is estimated using the measured 7 and 13 TeV production cross-sections ($\sigma(pp \rightarrow b\bar{b})$) from [42] and scaling the cross-sections for 8 TeV; these are multiplied with the luminosities for each year. No uncertainty is assigned to the luminosities in ratios.

1617 While generator level efficiency ratios over the years are in an excellent agreement (see
 1618 Table 62), the reconstruction and selection efficiencies in MC15 and MC16 show small
 1619 tensions. The general trend in the reconstruction and selection efficiencies seems to be in
 1620 favour of the signal channels.

1621 The efficiencies for the $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ are very similar, as expected.
 1622 The differences arise mostly from the reconstruction and selection and altogether are on
 1623 the same footing with the statistical uncertainties due to the sample sizes(see Table 61).

1624 The trigger efficiency ratios in Run 2 have changed the most. Especially for $B^0 \rightarrow K^+ \pi^-$
 1625 where the L0 and Hlt1 TIS efficiency is almost twice what is was in Run 1 ((28)).

1626 The efficiency corrected yields per pb^{-1} as well as the yields relative to the 2011 yields
 1627 are shown in Table 60 and in figure Fig. 56.

1628 Total detection efficiencies as well as the mass fits for both normalisation channels
 1629 are verified by comparing the branching fractions of $B^0 \rightarrow K^+ \pi^-$ ($B_s^0 \rightarrow J/\psi \phi$) and

1630 $B^+ \rightarrow J/\psi K^+$ decays in 2011, 2012, 2015 and 2016 data samples:

$$\frac{\mathcal{B}(B^0 \rightarrow K^+ \pi^-)}{\mathcal{B}(B^+ \rightarrow J/\psi K^+)} = \frac{N_{B^0 \rightarrow K^+ \pi^-}}{N_{B^+ \rightarrow J/\psi K^+}} \times \frac{\epsilon_{B^0 \rightarrow K^+ \pi^-}^{Gen}}{\epsilon_{B^0 \rightarrow K^+ \pi^-}^{Gen}} \times \frac{\epsilon_{B^+ \rightarrow J/\psi K^+}^{RecSel}}{\epsilon_{B^0 \rightarrow K^+ \pi^-}^{RecSel}} \times \frac{\epsilon_{B^+ \rightarrow J/\psi K^+}^{Trig}}{\epsilon_{B^+ \rightarrow J/\psi K^+}^{L0Hlt1TIS} \cdot \epsilon_{B^0 \rightarrow K^+ \pi^-}^{HLT2}} \times \frac{f_u}{f_d}, \quad (29)$$

$$\frac{\mathcal{B}(B_s^0 \rightarrow J/\psi \phi)}{\mathcal{B}(B^+ \rightarrow J/\psi K^+)} = \frac{N_{B_s^0 \rightarrow J/\psi \phi}}{N_{B^+ \rightarrow J/\psi K^+}} \times \frac{\epsilon_{B_s^0 \rightarrow J/\psi \phi}^{Gen}}{\epsilon_{B_s^0 \rightarrow J/\psi \phi}^{Gen}} \times \frac{\epsilon_{B^+ \rightarrow J/\psi K^+}^{RecSel}}{\epsilon_{B_s^0 \rightarrow J/\psi \phi}^{RecSel}} \times \frac{\epsilon_{B^+ \rightarrow J/\psi K^+}^{Trig}}{\epsilon_{B^+ \rightarrow J/\psi K^+}^{L0Hlt1TIS} \cdot \epsilon_{B_s^0 \rightarrow J/\psi \phi}^{HLT2}} \times \frac{f_u}{f_s}.$$

1631 The ratios include all the efficiency corrections for the channels and assumes isospin
1632 asymmetry ($f_u = f_d$). The $B^+ \rightarrow J/\psi K^+$ yields in S21r1 (2011) and S21(2012) are
1633 corrected for the small DiMuon/Leptonic stream luminosity differences given in Table 54.
1634 For $B_s^0 \rightarrow J/\psi \phi$, only the events triggered by muon TIS lines are used.

1635 The observed $B_s^0 \rightarrow J/\psi \phi$ and $B^+ \rightarrow J/\psi K^+$ branching fraction ratios in Table 63 are
1636 in agreement with the PDG value. This proves the mass fit and the detection efficiencies
1637 for the main normalisation channel ($B^+ \rightarrow J/\psi K^+$) are understood. The branching
1638 fraction ratio of $B^0 \rightarrow K^+ \pi^-$ and $B^+ \rightarrow J/\psi K^+$ channels is higher than expected. The
1639 discrepancies must arise from the $B^0 \rightarrow K^+ \pi^-$, and most likely, from the PID efficiency.
1640 If the discrepancies are not understood withing the next days we will assign it an hefty
1641 systematic uncertainty to the $B^0 \rightarrow K^+ \pi^-$ yield.

1642 The effect on the combined normalisation factors is small thanks to relatively precise
1643 normalisation constants from $B^+ \rightarrow J/\psi K^+$.

1644 7.6 Ratio of hadronisation fractions

1645 The relative B_s^0 and B^+ production fraction has been previously determined by LHCb
1646 at the proton collision energy of 7 TeV [17]. The stability of the relative B_s^0 and B^+
1647 production at higher collision energies of 8 and 13 TeV will be studied using the efficiency
1648 corrected ratio of $B_s^0 \rightarrow J/\psi \phi$ and $B^+ \rightarrow J/\psi K^+$ candidates relative to their ratio at 7
1649 TeV. In this case the uncertain branching fractions wil not be needed.

1650 The relative B_s^0 and B^+ production, and thus f_s/f_d , is found to be stable (Table 64)
1651 with the increase in collision energy. the f_s/f_d value determined on 7 TeV data can also
1652 be used at 13 TeV data with no additional uncertainty. However, the f_s/f_d value in Run 2
1653 normalisation is multiplied by the observed relative production difference between Run 1
1654 and Run 2:

$$C_{fsfd}^{Run2} = (f_s/f_d)_{13TeV} / (f_s/f_d)_{7+8TeV} = 1.068(46). \quad (30)$$

1655 The corrections for the stability check in Table 64 include the generator level efficiencies
1656 from Table 55, the reconstruction and selection efficiencies from Table 56, the tracking
1657 efficiency corrections from Table 57, the muon acceptance and isMuon corrections from

1658 Table 59 and the muon TOS trigger efficiencies, determined from the $B^+ \rightarrow J/\psi K^+$ sample
 1659 in B -meson phase space bins and multiplied with the simulated $B_s^0 \rightarrow J/\psi \phi$ distribution
 1660 in case of $B_s^0 \rightarrow J/\psi \phi$.

1661 7.7 Normalization factors

1662 The normalisation factors for $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ and also for the exclusive background estimations
 1663 are calculated using the inputs from this section: the measured normalisation channel
 1664 yields from Sect. 7.1 (using B2HHTOS Hlt2 line for $B^0 \rightarrow K^+ \pi^-$), the generator level
 1665 efficiencies from Table 55, the reconstruction and selection efficiencies from Table 56, the
 1666 tracking efficiency corrections from Table 57, the muon acceptance and isMuon corrections
 1667 from Table 59, the trigger efficiencies from Sect. 7.4, decay time corrections from Table 58
 1668 and the hadronisation ratio $f_s/f_d = 0.259 \pm 0.015$ from [17].

1669 The *background* normalisation factors are:

$$\begin{aligned}
 \beta_d^{bkg} &= \frac{N_{B^+ \rightarrow J/\psi K^+}}{\mathcal{B}(B^+ \rightarrow J/\psi K^+)} \times \frac{1}{\epsilon_{B^+ \rightarrow J/\psi K^+}^{Gen} * \epsilon_{B^+ \rightarrow J/\psi K^+}^{RecSel} * \epsilon_{B^+ \rightarrow J/\psi K^+}^{Trig}} \\
 &\quad \times \frac{1}{C_{B^+ \rightarrow J/\psi K^+}^{Track} * C_{B^+ \rightarrow J/\psi K^+}^{MuID}}, \\
 \beta_d^{bkg}(2011) &= (2.191 \pm 0.076) \times 10^{11}, \\
 \beta_d^{bkg}(2012) &= (4.994 \pm 0.174) \times 10^{11}, \\
 \beta_d^{bkg}(2015) &= (1.268 \pm 0.064) \times 10^{11}, \\
 \beta_d^{bkg}(2016) &= (4.617 \pm 0.209) \times 10^{11}.
 \end{aligned} \tag{31}$$

1670 These factors need to be corrected for the differences in the hadronisation factors, in
 1671 case used to calculate the expected yields for background arising from B meson decays
 1672 other than B^0 or B^+ (assuming isospin symmetry). The background normalistion factor
 1673 for total Run 1 set ($\beta_d^{2011+2012} = 7.2(2)$) is lower than the number used in the previous
 1674 LHCb Run 1 analysis ($\beta_d^{2011+2012} = 8.2(3)$). In the current analysis we use 2011 MC to
 1675 estimate precisely the efficiencies for 2011 and also a newer MC12 version (Sim08). The
 1676 reconstruction and selection efficiencies on these samples differ from the previous numbers.
 1677 Note, that this change will be largy mitigated in a ratio with the relevant background
 1678 channel efficiency when estimating the expected background yields.

1679 The signal normalisation factors are defined as:

$$\begin{aligned}
 \alpha_{B_s^0 \rightarrow \mu^+ \mu^-}^{norm} &= \frac{\mathcal{B}_{norm}}{N_{norm}} \times \frac{\epsilon_{norm}^{RecSel}}{\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{RecSel}} \times \frac{1}{C_{B_s^0 \rightarrow \mu^+ \mu^-}^{DecTime}} \times \frac{C_{norm}^{Track}}{C_{B_s^0 \rightarrow \mu^+ \mu^-}^{Track}} \\
 &\quad \times \frac{C_{norm}^{MuID}}{C_{B_s^0 \rightarrow \mu^+ \mu^-}^{MuIDProbNN}} \times \frac{\epsilon_{norm}^{Trig}}{\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{Trig}} \times \frac{f_d}{f_s}, \\
 \alpha_{B^0 \rightarrow \mu^+ \mu^-}^{norm} &= \frac{\mathcal{B}_{norm}}{N_{norm}} \times \frac{\epsilon_{norm}^{RecSel}}{\epsilon_{B^0 \rightarrow \mu^+ \mu^-}^{RecSel}} \times \frac{C_{norm}^{Track}}{C_{B^0 \rightarrow \mu^+ \mu^-}^{Track}} \\
 &\quad \times \frac{C_{norm}^{MuID}}{C_{B^0 \rightarrow \mu^+ \mu^-}^{MuIDProbNN}} \times \frac{\epsilon_{norm}^{Trig}}{\epsilon_{B^0 \rightarrow \mu^+ \mu^-}^{Trig}}.
 \end{aligned} \tag{32}$$

1680 The normalisation factor values are given in Table 65. The systematic uncertainties in
 1681 the tracking efficiency are taken as fully correlated for the two tracks in the signal and
 1682 normalisation channels. Additionally, we treat the signal efficiency as fully correlated
 1683 when combining the normalisation factors for each year.

1684 The numbers for the normalisation factors compare well to the numbers used in the
 1685 analysis of the full Run 1 data. Exact comparison to the Run 1 analysis numbers is
 1686 not attempted: the simulation versions are different, the stripping versions differ, and
 1687 the muon identification has changed. However, the 2012 numbers (which are the most
 1688 comparable between the two analysis) are in agreement within the uncertainties after the
 1689 differences in the muon identification (the most important change) are accounted for:

$$\begin{aligned}
 \alpha_{B_s^0 \rightarrow \mu^+ \mu^-}^{B^+ \rightarrow J/\psi K^+} (2012 \text{ } 3\text{fb ANA}) &= (1.3 \pm 0.1) \times 10^{-10} \times \frac{1.2}{1.02}, \\
 &= (1.5 \pm 0.1) \times 10^{-10}, \\
 \alpha_{B_s^0 \rightarrow \mu^+ \mu^-}^{B^+ \rightarrow J/\psi K^+} (2012) &= (1.59 \pm 0.1) \times 10^{-10}.
 \end{aligned} \tag{33}$$

Table 55: Geometrical detector acceptance, estimated as the fraction of decays contained in the polar angle region of [10, 400] mrad. The numbers in the last column are weighted and average according to the fraction of MagUP and MagDOWN data in each year (Note that for the MC12 MayJune conditions we have the efficiency for the correct event type (11102003) for the $B^0 \rightarrow K^+ \pi^-$. The Sim06 is included only crosscheck, the normalisation will be done using Sim08 for 2012 and MayJune/JunSept mismatch plays no role in the final analysis).

Channel	Conditions	$\epsilon^{Gen, UP}$	$\epsilon^{Gen, DOWN}$	ϵ^{Gen}
$B_s^0 \rightarrow \mu^+ \mu^-$				
MC11	Sim08, Pyth8	(17.41 ± 0.058)%	(17.47 ± 0.058)%	(17.45 ± 0.04)%
MC12 JulSep	Sim06, Pyth6	(17.63 ± 0.093)%	(17.75 ± 0.093)%	(17.69 ± 0.07)%
MC12	Sim08, Pyth8	(18.73 ± 0.033)%	(18.61 ± 0.032)%	(18.67 ± 0.02)%
MC15	Sim09, Pyth8	(19.31 ± 0.053)%	(19.28 ± 0.052)%	(19.29 ± 0.04)%
MC16	Sim09, Pyth8	(19.34 ± 0.066)%	(19.33 ± 0.066)%	(19.33 ± 0.05)%
$B^0 \rightarrow \mu^+ \mu^-$				
MC11	Sim08, Pyth8	(17.75 ± 0.09)%	(17.72 ± 0.09)%	(17.73 ± 0.07)%
MC12 JulSep	Sim06, Pyth6	(17.44 ± 0.06)%	(17.40 ± 0.06)%	(17.42 ± 0.04)%
MC12	Sim08, Pyth8	(18.65 ± 0.05)%	(18.74 ± 0.05)%	(18.69 ± 0.03)%
MC15	Sim09, Pyth8	(19.29 ± 0.06)%	(19.28 ± 0.06)%	(19.28 ± 0.04)%
MC16	Sim09, Pyth8	(19.32 ± 0.07)%	(19.43 ± 0.07)%	(19.37 ± 0.05)%
$B^+ \rightarrow J/\psi K^+$				
MC11	Sim08, Pyth8	(15.45 ± 0.04)%	(15.46 ± 0.04)%	(15.46 ± 0.03)%
MC12 JulSep	Sim06, Pyth6	(15.71 ± 0.08)%	(15.78 ± 0.08)%	(15.74 ± 0.06)%
MC12	Sim08, Pyth8	(16.63 ± 0.07)%	(16.69 ± 0.07)%	(16.66 ± 0.05)%
MC15	Sim09, Pyth8	(17.38 ± 0.05)%	(17.20 ± 0.05)%	(17.28 ± 0.04)%
MC16	Sim09, Pyth8	(17.26 ± 0.07)%	(17.44 ± 0.07)%	(17.34 ± 0.05)%
$B_s^0 \rightarrow J/\psi \phi$				
MC11	Sim08, Pyth8	(16.06 ± 0.07)%	(15.97 ± 0.07)%	(16.01 ± 0.05)%
MC12 JulSep	Sim06, Pyth6	(16.22 ± 0.08)%	(16.33 ± 0.09)%	(16.28 ± 0.06)%
MC12	Sim08, Pyth8	(17.11 ± 0.030)%	(17.17 ± 0.029)%	(17.14 ± 0.02)%
MC15	Sim09, Pyth8	(17.70 ± 0.05)%	(17.73 ± 0.05)%	(17.72 ± 0.04)%
MC16	Sim09, Pyth8	(17.68 ± 0.06)%	(17.75 ± 0.06)%	(17.71 ± 0.05)%
$B^0 \rightarrow K^+ \pi^-$				
MC11	Sim08, Pyth8	(17.73 ± 0.04)%	(17.74 ± 0.03)%	(17.74 ± 0.02)%
MayJune12	Sim06, Pyth6	(17.96 ± 0.09)%	(18.00 ± 0.09)%	(17.98 ± 0.06)%
MC12	Sim08, Pyth8	(19.04 ± 0.075)%	(18.92 ± 0.069)%	(18.98 ± 0.05)%
MC15	Sim09, Pyth8	(19.58 ± 0.05)%	(19.60 ± 0.05)%	(19.59 ± 0.04)%
MC16	Sim09, Pyth8	(19.59 ± 0.07)%	(19.67 ± 0.07)%	(19.63 ± 0.05)%

Table 56: Reconstruction and selection efficiencies for $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ and the normalisation channels, evaluated on the Monte Carlo simulated samples after full reconstruction and selection. J/ψ veto has only been applied on $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow K^+ \pi^-$; BKGCAT matching (incl. radiative decays) is imposed on $B^+ \rightarrow J/\psi K^+$, $B_s^0 \rightarrow J/\psi \phi$, and $B^0 \rightarrow K^+ \pi^-$. $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ MC samples are very clean and need no further matching. The $B^0 \rightarrow K^+ \pi^-$ mass window matches with the mass window used to determine the normalisation $K\pi$ yields, which runs from 5200 to 5850 MeV/ c^2 in the $K^+ \pi^-$ mass hypothesis.

Channel	Conditions	$N_{MCB2XTuple}^{GenInAcc.}$	$N^{RecSel Gen}$	$\epsilon^{RecSel Gen}$
$B_s^0 \rightarrow \mu^+ \mu^-$				
MC11	Sim08, Pyth8	534, 499	180, 540	$(33.78 \pm 0.07)\%$
MC12(JulSep)	Sim06, Pyth6	2, 009, 392	623, 857	$(31.05 \pm 0.03)\%$
MC12	Sim08, Pyth8	2, 080, 642	655, 129	$(31.49 \pm 0.03)\%$
MC15	Sim09, Pyth8	2, 108, 036	582, 012	$(27.61 \pm 0.03)\%$
MC16	Sim09, Pyth8	1, 127, 675	316, 565	$(28.07 \pm 0.03)\%$
$B^0 \rightarrow \mu^+ \mu^-$				
MC11	Sim08, Pyth8	508, 999	171, 101	$(33.62 \pm 0.07)\%$
MC12(JulSep)	Sim06, Pyth6	2, 006, 185	629, 660	$(31.39 \pm 0.03)\%$
MC12	Sim08, Pyth8	498, 027	154, 132	$(30.95 \pm 0.07)\%$
MC15	Sim09, Pyth6	2, 188, 268	592, 842	$(27.09 \pm 0.03)\%$
MC16	Sim09, Pyth8	1, 970, 654	542, 750	$(27.54 \pm 0.03)\%$
$B^+ \rightarrow J/\psi K^+$				
MC11	Sim08, Pyth8	762, 312	136, 728	$(17.94 \pm 0.04)\%$
MC12(JulSep)	Sim06, Pyth6	1, 012, 615	162, 641	$(16.06 \pm 0.04)\%$
MC12	Sim08, Pyth8	5, 047, 318	835, 149	$(16.55 \pm 0.02)\%$
MC15	Sim09, Pyth8	4, 161, 740	557, 633	$(13.40 \pm 0.02)\%$
MC16	Sim09, Pyth8	2, 076, 118	295, 768	$(14.25 \pm 0.02)\%$
$B_s^0 \rightarrow J/\psi \phi$				
MC11	Sim08, Pyth8	315, 999	27, 341	$(8.65 \pm 0.05)\%$
MC12(JulSep)	Sim06, Pyth6	1, 007, 992	75, 000	$(7.44 \pm 0.03)\%$
MC12	Sim08, Pyth8	2, 158, 355	175, 426	$(8.13 \pm 0.02)\%$
MC15	Sim09, Pyth8	2, 115, 485	122, 749	$(6.09 \pm 0.02)\%$
MC16	Sim09, Pyth8	2, 000, 071	136, 904	$(6.85 \pm 0.02)\%$
$B^0 \rightarrow K^+ \pi^-$				
MC11	Sim08, Pyth8	775, 505	180, 611	$(23.29 \pm 0.05)\%$
MC12(JulSep)	Sim06, Pyth6	1, 016, 003	217, 546	$(21.41 \pm 0.04)\%$
MC12(08b+g)	Sim08, Pyth8	10, 543, 584	2, 299, 768	$(21.81 \pm 0.01)\%$
MC15(sim09)	Sim09, Pyth8	4, 050, 901	755, 363	$(18.65 \pm 0.02)\%$
MC16	Sim09, Pyth8	8, 116, 346	1, 554, 248	$(19.15 \pm 0.01)\%$

Table 57: The tracking efficiency corrections to the reconstruction and selection efficiencies ().

$C_{Data/MC}^{Track}$	$B_{(s)}^0 \rightarrow \mu^+ \mu^-$	$B^0 \rightarrow K^+ \pi^-$	$B^+ \rightarrow J/\psi K^+$	$B_s^0 \rightarrow J/\psi \phi$
2011	1.0064(58)	1.010(19)	1.016(13)	1.037(19)
2012(Sim08)	1.0058(58)	1.009(19)	1.015(13)	1.035(19)
2015	1.0152(235)	1.015(29)	1.020(32)	1.025(39)
2016	1.0381(238)	1.036(30)	1.051(33)	1.062(40)

Table 58: The corrections to the reconstruction and selection efficiencies due to the B_s^0 lifetime differences in MC and SM. The correction uncertainties include the SM lifetime uncertainty and the statistical uncertainty arising from determining the full $B_s^0 \rightarrow \mu^+ \mu^-$ selection efficiency in each lifetime bin.

	Sim	$\tau_{MC}(B_s)$	$\tau_{SM}(B_s)$	$\epsilon_{SM}^{RecSel} / \epsilon_{MC}^{RecSel}$
MC11	08	1.503 ps	1.61(1) ps	1.027(3)
MC12	06	1.472 ps	1.61(1) ps	1.038(2)
MC12	08	1.472 ps	1.61(1) ps	1.027(2)
MC15	09	1.512 ps	1.61(1) ps	1.021(2)
MC16	09	1.512 ps	1.61(1) ps	1.022(2)

Table 59: isMuon efficiencies after the full selection for the $B^+ \rightarrow J/\psi K^+$, $B_s^0 \rightarrow J/\psi \phi$ and $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays. Estimated on both MC and Data (using the PIDCalib). The $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ efficiencies on the data also include the ProbNN cut efficiency. The correction factors for the channel selections efficiencies are calculated as the ratio of the Data and MC numbers in this table.

	$B_s^0 \rightarrow \mu^+ \mu^-$		$B^0 \rightarrow \mu^+ \mu^-$	
	$\epsilon_{MC}^{isMu RecSel}$	$\epsilon_{Data}^{isMuProbNN RecSel}$	$\epsilon_{MC}^{isMu RecSel}$	$\epsilon_{Data}^{isMuProbNN RecSel}$
2011	96.02(5)%	82.3(1.6))%	96.14(5)%	82.3(1.6)%
2012	96.42(2)%	82.3(1.6))%	96.41(2)%	82.3(1.6)%
2015	95.98(3)%	83.5(1.7)%	95.93(3)%	83.5(1.7)%
2016	95.99(3)%	84.3(1.7)%	95.94(3)%	84.3(1.7)%
	$B^+ \rightarrow J/\psi K^+$		$B_s^0 \rightarrow J/\psi \phi$	
	$\epsilon_{MC}^{isMu RecSel}$	$\epsilon_{Data}^{isMu RecSel}$	$\epsilon_{MC}^{isMu RecSel}$	$\epsilon_{Data}^{isMu RecSel}$
2011	94.62(6)%	97.90(4)%	94.48(6)%	97.93(4)%
2012	94.90(2)%	97.76(3)%	95.03(7)%	97.78(3)%
2015	94.85(3)%	97.22(3)%	94.86(6)%	97.25(3)%
2016	94.79(4)%	97.52(3)%	94.82(6)%	97.54(3)%

Table 60: Normalisation and control channels yields per pb^{-1} in different years. Yields are corrected for all the efficiencies and the uncertainties include a 5% uncertainty on the luminosity (5%)

	$B^+ \rightarrow J/\psi K^+$	$B_s^0 \rightarrow J/\psi \phi$	$B^0 \rightarrow K^+ \pi^-$
2011	13923(733)	1657(93)	4211(449)
2012	15477(815)	1800(98)	4883(300)
2015	26716(1719)	3346(215)	9283(1078)
2016	25673(1542)	3170(204)	9103(842)

Table 61: Stability of simulated geometrical detector acceptance, MC reconstruction and selection efficiencies, tracking efficiency and muonID corrected and trigger efficiencies in different MC versions (B2HH_TOS for $B^0 \rightarrow K^+ \pi^-$).

	$\frac{\epsilon_{B^+ \rightarrow J/\psi K^+}^{Gen}}{\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{Gen}}$	$\frac{\epsilon_{B^0 \rightarrow K^+ \pi^-}^{Gen}}{\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{Gen}}$	$\frac{\epsilon_{B_s^0 \rightarrow J/\psi \phi}^{Gen}}{\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{Gen}}$	$\frac{\epsilon_{B^+ \rightarrow J/\psi K^+}^{RecSel}}{\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{RecSel}}$	$\frac{\epsilon_{B^0 \rightarrow K^+ \pi^-}^{RecSel}}{\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{RecSel}}$	$\frac{\epsilon_{B_s^0 \rightarrow J/\psi \phi}^{RecSel}}{\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{RecSel}}$
MC11	0.886(3)	1.017(3)	0.918(4)	0.517(2)	0.671(3)	0.249(2)
MC12	0.890(5)	1.016(5)	0.920(5)	0.517(1)	0.690(1)	0.240(1)
MC12 Sim08	0.892(3)	1.017(3)	0.918(1)	0.512(2)	0.674(2)	0.251(1)
MC15	0.896(3)	1.016(3)	0.919(3)	0.475(1)	0.662(2)	0.216(1)
MC16	0.897(3)	1.016(4)	0.916(4)	0.497(1)	0.668(2)	0.239(1)
	$\frac{\epsilon_{B^+ \rightarrow J/\psi K^+}^{FullRecSel}}{\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{FullRecSel}}$	$\frac{\epsilon_{B^0 \rightarrow K^+ \pi^-}^{FullRecSel}}{\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{FullRecSel}}$	$\frac{\epsilon_{B_s^0 \rightarrow J/\psi \phi}^{FullRecSel}}{\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{FullRecSel}}$	$\frac{\epsilon_{B^+ \rightarrow J/\psi K^+}^{Trig}}{\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{Trig}}$	$\frac{\epsilon_{B^0 \rightarrow K^+ \pi^-}^{Trig}}{\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{Trig}}$	$\frac{\epsilon_{B_s^0 \rightarrow J/\psi \phi}^{Trig}}{\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{Trig}}$
MC11	0.630(15)	0.786(22)	0.311(8)	0.933(22)	0.038(1)	0.844(19)
MC12 Sim08	0.623(15)	0.793(22)	0.312(9)	0.930(23)	0.044(1)	0.844(20)
MC15	0.563(25)	0.760(32)	0.257(12)	0.924(38)	0.085(4)	0.809(28)
MC16	0.589(26)	0.759(32)	0.276(13)	0.932(16)	0.074(1)	0.870(12)

Table 62: Signal channels efficiency ratios at different levels and for different data taking periods. The RecSel efficiency for $B_s^0 \rightarrow \mu^+ \mu^-$ included the decay time correction.

	$\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{Full}$	$\epsilon_{B^0 \rightarrow \mu^+ \mu^-}^{Full}$	$\frac{\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{Gen}}{\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{Gen}}$	$\frac{\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{RecSel}}{\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{RecSel}}$	$\frac{\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{Trig}}{\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{Trig}}$	$\frac{\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{Full}}{\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{Gen}}$
2011	0.0492(15)	0.0483(15)	0.984(4)	1.032(4)	1.001(3)	1.018(44)
2012	0.0488(15)	0.0466(14)	0.999(2)	1.045(4)	1.002(3)	1.046(45)
2015	0.0462(21)	0.0443(21)	1.001(3)	1.041(3)	1.002(5)	1.043(67)
2016	0.0492(17)	0.0472(17)	0.998(4)	1.042(3)	1.002(2)	1.041(50)

Table 63: Normalisation cross-check: estimating the branching fraction ratio between $B^0 \rightarrow K^+ \pi^-$ and $B^+ \rightarrow J/\psi K^+$ from the efficiency corrected yields.

	PDG 2015	2011	2012(Sim08)	2015	2016
$\frac{\mathcal{B}(B^0 \rightarrow K^+ \pi^-)}{\mathcal{B}(B^+ \rightarrow J/\psi K^+)}(B^0 \text{ Hlt2 B2HH})$	0.320(13)	0.340(34)	0.329(24)	0.363(41)	0.357(36)
$\frac{\mathcal{B}(B_s^0 \rightarrow J/\psi \phi)}{\mathcal{B}(B^+ \rightarrow J/\psi K^+)}(B_s^0 \text{ muon TOS})$	0.510(42)	0.467(30)	0.452(28)	0.484(38)	0.476(36)

Table 64: The efficiency corrected $B_s^0 \rightarrow J/\psi \phi$ and $B^+ \rightarrow J/\psi K^+$ yield ratio, normalised to the same ratio in 7 TeV (2011, S21r1) data. The first uncertainty on the $B^+ \rightarrow J/\psi K^+$ yield arises from the model choice, the second is a combinatio of statistical and systematic uncertainty from the caonstrained model parameters. The first uncertainty on the $B_s^0 \rightarrow J/\psi \phi$ yield arises from the possible contribution from the additional resonances and interferences in $\mu^+ \mu^- KK$ spectrum (0.8%), the second and third uncertainties are from the model choice and statistics and model systematics, respectively.

	2011(7 TeV)	2012(8 TeV)
$N(B_s^0 \rightarrow J/\psi \phi)$	$19,190 \pm 154 \pm 298 \pm 166$	$42,103 \pm 337 \pm 577 \pm 256$
$N(B^+ \rightarrow J/\psi K^+)$	$307,856 \pm 994 \pm 661$	$693,930 \pm 27 \pm 878$
$\mathcal{C}_{track}(\frac{B^+ \rightarrow J/\psi K^+}{B_s^0 \rightarrow J/\psi \phi})$	0.980(13)	0.981(13)
$\frac{\epsilon^{\mu TOS}(B^+)}{\epsilon^{\mu TOS}(B_s^0)}$	0.987(11)	1.000(7)
$\frac{f_s/f_d @ X \text{ TeV}}{f_s/f_d @ 7 \text{ TeV}}$	1.00	0.98(3)
	2015(13 TeV)	2016(13 TeV)
$N(B_s^0 \rightarrow J/\psi \phi)$	$8,571 \pm 69 \pm 123 \pm 102$	$37,765 \pm 302 \pm 718 \pm 255$
$N(B^+ \rightarrow J/\psi K^+)$	$141,150 \pm 260 \pm 441$	$614,498 \pm 835 \pm 1716$
$\mathcal{C}_{track}(\frac{B^+ \rightarrow J/\psi K^+}{B_s^0 \rightarrow J/\psi \phi})$	0.995(42)	0.995(42)
$\frac{\epsilon^{\mu TOS}(B^+)}{\epsilon^{\mu TOS}(B_s^0)}$	0.975(12)	0.987(5)
$\frac{f_s/f_d @ X \text{ TeV}}{f_s/f_d @ 7 \text{ TeV}}$	1.05(6)	1.05(6)

Table 65: $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ normalisation factors and expected yields for each year. The expected signal yields are estimated assuming SM branching fractions as well as the latest most precise experimental branching fraction measurements [1]. The correlations in tracking efficiency systematic uncertainty and the hadronisation fraction (f_s/f_d) are accounted for in the combined alpha $\alpha^{comb.}$. The combined numbers for Run 1 and Run 2 additionally account for the normalisation channel branching fraction correlation.

$B_s^0 \rightarrow \mu^+ \mu^-$	$\alpha^{B^+ \rightarrow J/\psi K^+} \times 10^{10}$	$\alpha^{B^0 \rightarrow K^+ \pi^-} \times 10^{10}$	$\alpha^{comb.} \times 10^{10}$	$N_{expected}^{SM}$	$N_{expected}^{LHC}$
2011	3.507(258)	3.376(409)	3.492(252)	10.5 ± 1.0	7.7 ± 2.0
2012	1.550(114)	1.545(154)	1.549(111)	23.6 ± 2.3	17.4 ± 4.6
2015	6.416(564)	5.818(761)	6.296(529)	5.8 ± 0.6	4.3 ± 1.2
2016	1.695(134)	1.527(180)	1.662(125)	22.0 ± 2.2	16.2 ± 4.4
Run 1	1.075(76)	1.060(94)	1.071(72)	34.2 ± 3.2	25.2 ± 6.8
Run 2	1.341(101)	1.210(126)	1.306(95)	28.0 ± 2.7	20.7 ± 5.6
		Total:	0.588(38)	62.2 ± 5.6	45.9 ± 12.3
$B^0 \rightarrow \mu^+ \mu^-$	$\alpha^{B^+ \rightarrow J/\psi K^+} \times 10^{11}$	$\alpha^{B^0 \rightarrow K^+ \pi^-} \times 10^{11}$	$\alpha^{comb.} \times 10^{11}$	$N_{expected}^{SM}$	$N_{expected}^{LHC}$
2011	9.245(427)	8.902(952)	9.206(405)	1.2 ± 0.1	4.2 ± 1.7
2012	4.198(192)	4.186(340)	4.196(179)	2.5 ± 0.2	9.3 ± 3.8
2015	17.329(1.16)	15.713(1.848)	16.998(1.042)	0.6 ± 0.1	2.3 ± 1.0
2016	4.571(250)	4.118(424)	4.479(219)	2.4 ± 0.2	8.7 ± 3.6
Run 1	2.887(117)	2.847(192)	2.877(101)	3.7 ± 0.3	13.6 ± 5.6
Run 2	3.617(179)	3.262(283)	3.521(155)	3.0 ± 0.3	11.1 ± 4.6
		Total:	1.583(44)	6.7 ± 0.6	24.6 ± 10.1

1690 8 Time dependent effects

1691 8.1 Effect of effective lifetime on the distribution of the BDT 1692 response

1693 The output of the BDT is flattened using Monte Carlo simulated $B_s^0 \rightarrow \mu^+ \mu^-$ decays such
1694 that the distribution of the BDT classifier is flat for signal candidates. The consequence of
1695 this is that the expected yield of $B_s^0 \rightarrow \mu^+ \mu^-$ decays in each BDT bin is simply proportional
1696 to the width of the bin.

1697 However, a complication arises in that the simulated decays are generated using the
1698 mean B_s^0 lifetime (at the time of production) as the decay lifetime. The lifetimes used to
1699 generate the various $B_s^0 \rightarrow \mu^+ \mu^-$ Monte Carlo simulation samples are listed in Table 66.
However, as discussed in Section 2.2 and elsewhere the effective lifetime may take a range

Year	Simulation version	Generated lifetime / ps
2011	sim08a	1.503
2012	sim06b	1.472
2012	sim08h	1.510
2015	sim09a	1.512

Table 66: Lifetimes used to generate $B_s^0 \rightarrow \mu^+ \mu^-$ Monte Carlo simulated events.

1700
1701 of values between the lifetimes of the light and heavy mass eigenstates. Specifically, the
1702 effective lifetime is equal to

$$\tau_{\mu^+ \mu^-} = \frac{2\tau_{B_s^0} \mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-} y_s + (1 + y_s^2) \tau_{B_s^0}}{(1 + y_s^2) + \mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-} y_s (1 - y_s^2)} \quad (34)$$

1703 where

$$y_s \equiv \frac{\Gamma_L - \Gamma_H}{\Gamma_L + \Gamma_H} = 0.062 \pm 0.006 \quad (35)$$

1704 and the mean B_s^0 lifetime $\tau_{B_s^0} = 1.511 \pm 0.014$ ps [19]. The parameter $\mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-}$ is unknown
1705 but is expected to be equal to +1 in the Standard Model.

1706 Since the response of the BDT is correlated with the decay time of the candidate
1707 (with longer-lived candidates tending to have larger values of BDT response) the fact that
1708 the simulated events use the mean B_s^0 lifetime may bias the BDT distribution of true
1709 $B_s^0 \rightarrow \mu^+ \mu^-$ candidates in data if the $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime is different from the
1710 one used to generate the simulated events.

1711 To account for this, numerical factors are calculated for each Monte Carlo sample,
1712 which can be used to ‘correct’ the BDT distribution in data for $\mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-} = -1, 0$ and 1
1713 (corresponding to $\tau_{\mu^+ \mu^-} = 1.411$ ps, 1.511 ps and 1.599 ps). Truth-matched simulated
1714 $B_s^0 \rightarrow \mu^+ \mu^-$ decays are used for this purpose. Simulated candidates are required to pass

1715 the full trigger, stripping and pre-selection requirements. Weights are then calculated for
 1716 each candidate according to

$$\omega_i = \frac{\tau_{gen}}{\tau_{\mu^+\mu^-}} e^{-t_i(1/\tau_{\mu^+\mu^-} - 1/\tau_{gen})} \quad (36)$$

1717 where t_i is the reconstructed decay time for that candidate, τ_{gen} is the lifetime with which
 1718 the candidates were generated and $\tau_{\mu^+\mu^-}$ is the effective lifetime calculated from y_s and
 1719 $\tau_{B_s^0}$.

1720 A correction factor is then calculated for each BDT bin according to

$$k = \sum_{i=1}^N \omega_i / N = \frac{\epsilon_{\tau_{\mu^+\mu^-}}}{\epsilon_{\tau_{gen}}} \quad (37)$$

1721 where N is the number of candidates in the bin. These correction factors then represent
 1722 the ratios of the total efficiencies (including reconstruction, trigger and offline selection) of
 1723 a data sample with a mean lifetime corresponding to one of the three different values of
 1724 $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-}$ considered to a data sample with a mean lifetime corresponding to the one used to
 1725 generate the Monte Carlo sample in question. The results are listed below in Tables 67,
 1726 68, 69, 70 and 71.

BDT bin	N	$k(\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} = -1)$	$k(\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} = 0)$	$k(\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} = +1)$
1	39,487	1.006	0.999	0.992
2	23,901	0.988	1.001	1.010
3	16,021	0.979	1.002	1.019
4	16,057	0.975	1.002	1.023
5	16,141	0.969	1.003	1.030
6	16,165	0.961	1.003	1.038
7	16,354	0.943	1.005	1.056
8	16,512	0.902	1.009	1.102
Total	160,639	0.972	1.002	1.026

Table 67: Factors (k) required to correct the BDT distribution of candidates in data due to the difference between the lifetime used to generate simulated decays and various physics scenarios with different values of $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-}$ for 2011 sim08a Monte Carlo.

BDT bin	N	$k(\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} = -1)$	$k(\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} = 0)$	$k(\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} = +1)$
1	161,692	1.002	0.999	0.994
2	85,005	0.989	1.006	1.020
3	55,109	0.984	1.010	1.029
4	52,428	0.980	1.012	1.037
5	50,545	0.976	1.015	1.046
6	48,249	0.970	1.019	1.059
7	46,775	0.958	1.027	1.085
8	42,544	0.930	1.045	1.147
Total	542,349	0.981	1.012	1.036

Table 68: Factors (k) required to correct the BDT distribution of candidates in data due to the difference between the lifetime used to generate simulated decays and various physics scenarios with different values of $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-}$ for 2012 sim06b Monte Carlo.

BDT bin	N	$k(\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} = -1)$	$k(\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} = 0)$	$k(\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} = +1)$
1	144,693	1.005	1.000	0.994
2	84,030	0.986	1.000	1.010
3	57,232	0.978	1.000	1.017
4	57,273	0.973	1.000	1.022
5	57,783	0.967	1.000	1.027
6	57,093	0.958	1.000	1.035
7	58,236	0.939	1.001	1.052
8	55,595	0.894	1.001	1.095
Total	571,940	0.970	1.000	1.025

Table 69: Factors (k) required to correct the BDT distribution of candidates in data due to the difference between the lifetime used to generate simulated decays and various physics scenarios with different values of $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-}$ for 2012 sim08h Monte Carlo.

BDT bin	N	$k(\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} = -1)$	$k(\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} = 0)$	$k(\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} = +1)$
1	123,195	0.999	1.000	0.998
2	75,169	0.982	1.000	1.013
3	50,138	0.977	1.000	1.017
4	50,374	0.971	1.000	1.022
5	50,976	0.965	1.000	1.028
6	50,300	0.956	1.000	1.035
7	51,302	0.934	0.999	1.054
8	51,732	0.884	0.999	1.101
Total	503,187	0.965	1.000	1.028

Table 70: Factors (k) required to correct the BDT distribution of candidates in data due the difference between the lifetime used to generate simulated decays and various physics scenarios with different values of $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-}$ for 2015 sim09a Monte Carlo.

BDT bin	N	$k(\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} = -1)$	$k(\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} = 0)$	$k(\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} = +1)$
1	77351	0.998432	0.999999	0.999096
2	46021	0.980531	0.999825	1.01395
3	30567	0.975213	0.999772	1.01868
4	30463	0.969693	0.999717	1.02345
5	30314	0.962614	0.999647	1.02949
6	29688	0.954673	0.99957	1.03604
7	29648	0.932937	0.999354	1.05489
8	30228	0.883375	0.998841	1.10108
Total	304280	0.964865	0.999667	1.02795

Table 71: Factors (k) required to correct the BDT distribution of candidates in data due the difference between the lifetime used to generate simulated decays and various physics scenarios with different values of $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-}$ for 2016 sim09a Monte Carlo.

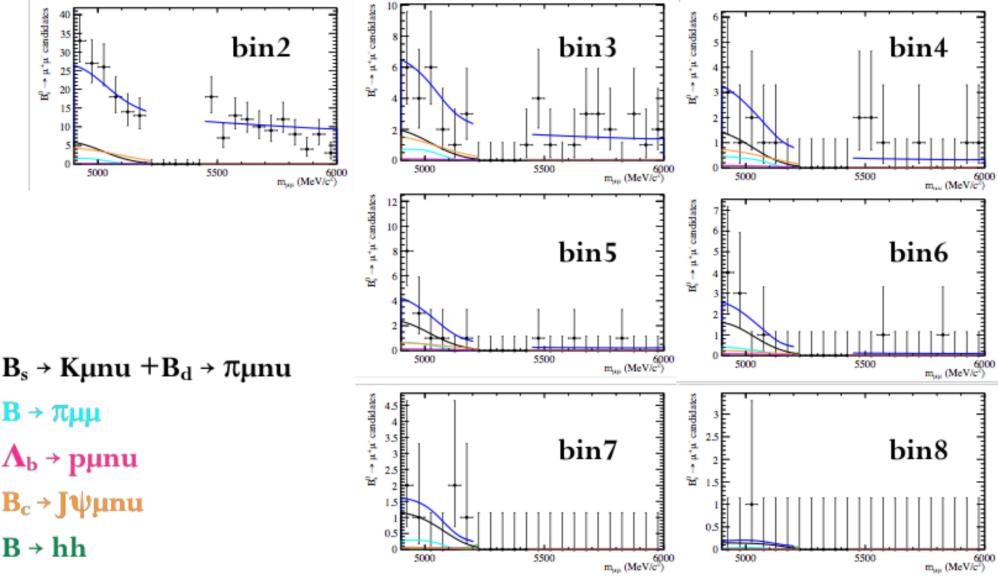


Figure 57: Invariant mass fit of the dimuon sidebands for the 8 BDT bins of run 1 data; full fit model is used, with signal region excluded.

9 BF results

9.1 Toy MC studies

In order to assess the expected sensitivity and the best BDT binning configuration Toy MC studies are performed. These toys are based on the background measured from the dimuon data sidebands, and fitted according to the background model discussed in Sections 5.1 and 5.2. The fit results are shown in figs. 57, 58, 59 for Run 1, 2015 and 2016 data, respectively. Thanks to the excellent BDT performances, in all samples the last two bins have no combinatorial events on the right mass sidebands, which represents of course a problem in constraining this component in the fit. This has been investigated through toys, and using Run1 data (3fb^{-1}). Three different BDT bin configurations are considered:

1. **7-bins**: the usual bin boundaries are considered but excluding the first BDT bin, [0.25,0.4,0.5,0.6,0.7,0.8,0.9,1.0];
2. **5-bins**: the last 3 bins are merged together and the first BDT bin is removed, [0.25,0.4,0.5,0.6,0.7,1.0]; with this configuration 2 events are present in the right sideband of the last bin (see fig. 57);
3. **4-bins**: the last 4 bins are merged together and the first BDT bin is removed,

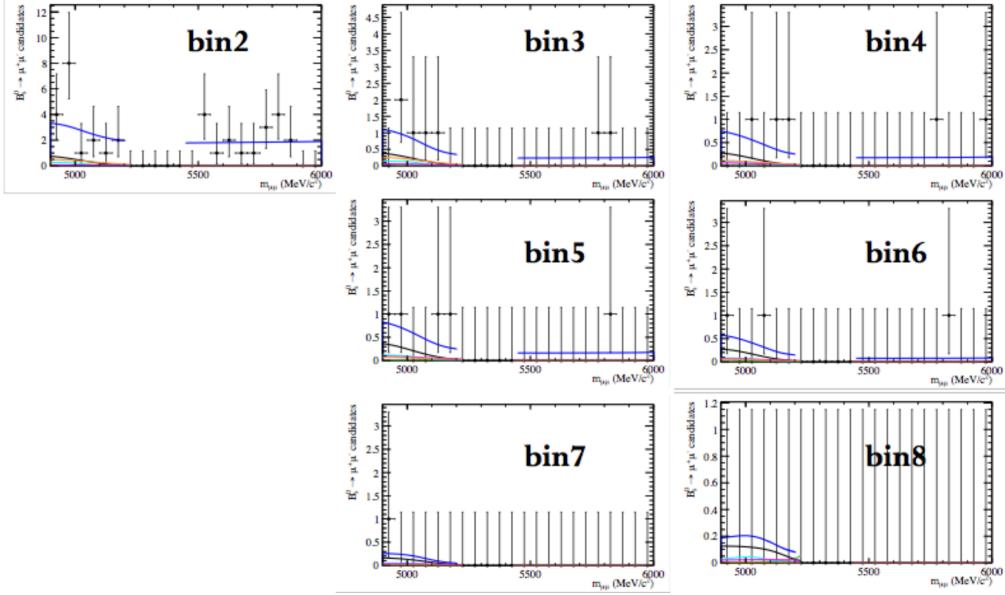


Figure 58: Invariant mass fit of the dimuon sidebands for the 8 BDT bins of 2015 data; full fit model is used, with signal region excluded.

1743 [0.25,0.4,0.5,0.6,1.0]; in this case, 5 events are present in the right sideband of the
 1744 last bin.

1745 A small loss of sensitivity is expected merging the latest bins where the combinatorial
 1746 background is negligible, while an improvement of the fit stability is expected. Since the
 1747 right sideband is populated only by combinatorial background, its yield is assumed to
 1748 be always positive defined. For each configuration $\sim 1k$ toys are generated. The pulls
 1749 for the B_s^0 and B^0 modes are shown in Figs. 60, 61, 62 for the 7-bin, 5-bin and 4-bin
 1750 configurations, respectively.

1751 A $\sim 5\%$ bias in the BF of the B_s^0 mode is observed in the 7-bins and 5-bins. This seems
 1752 to totally disappear in the 4-bins configuration. For each toy the statistical significance
 1753 of the B_s^0 mode is also evaluated using the Wilks' theorem. In Tab. 72 the expected
 1754 significance and the probability of having more than 5σ are listed. The expected significance
 1755 is evaluated as the median value of the statistical significance distribution reported in
 1756 Fig. 63 for each binning configuration.

1757 In the 4-bin configuration the merging of the latest BDT bins seems to produce a
 1758 positive effect on the overall stability of the fit without a significant decrease of sensitivity.
 1759 This is confirmed by looking at the pulls of the combinatorial background for the 5- and
 1760 4-bin configuration, which are shown in fig. 64 and 65, respectively. In the case of the 5-bin
 1761 configuration, the pull of the most significant bin is much more distorted with respect to

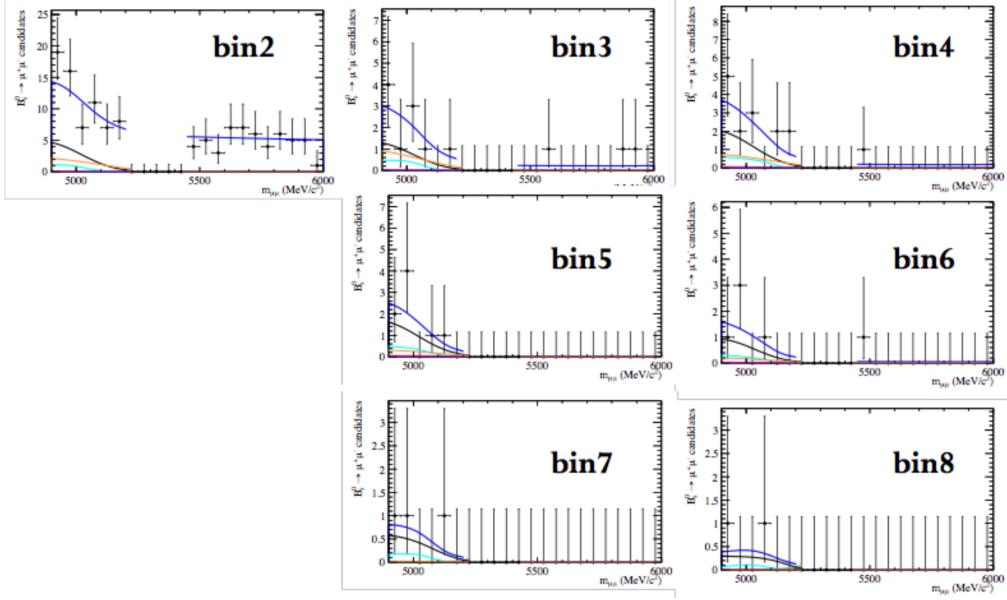


Figure 59: Invariant mass fit of the dimuon sidebands for the 8 BDT bins of 2016 data; full fit model is used, with signal region excluded.

Table 72: Expected significance and observation probability of the $B_s^0 \rightarrow \mu^+\mu^-$ decay mode, Run1 toys.

BDT binning configuration	expected significance	$> 5\sigma$ probability
7-bins	6.5	89.2%
5-bins	6.3	88.3%
4-bins	6.1	83.8%

the 4-bin case. A further reason to go for a 4-bin configuration, is that for 2016 right mass sidebands (fig. 59), one event only is observed in the last 4 BDT bins, so that in this case a 4-bin configuration is mandatory.

Finally, toys have been run on the full dataset, using the 4-bin fit configuration for all samples. The branching fraction pulls are shown in Fig. 66, while the expected sensitivities are shown in Fig. 67, for both B_s^0 and B^0 . While it is very unlikely not to have a 5-sigma observation on the SM $B_s^0 \rightarrow \mu^+\mu^-$ decay, the toys indicate a $\sim 8\%$ probability of having a 3σ evidence for a SM $B^0 \rightarrow \mu^+\mu^-$ decay.

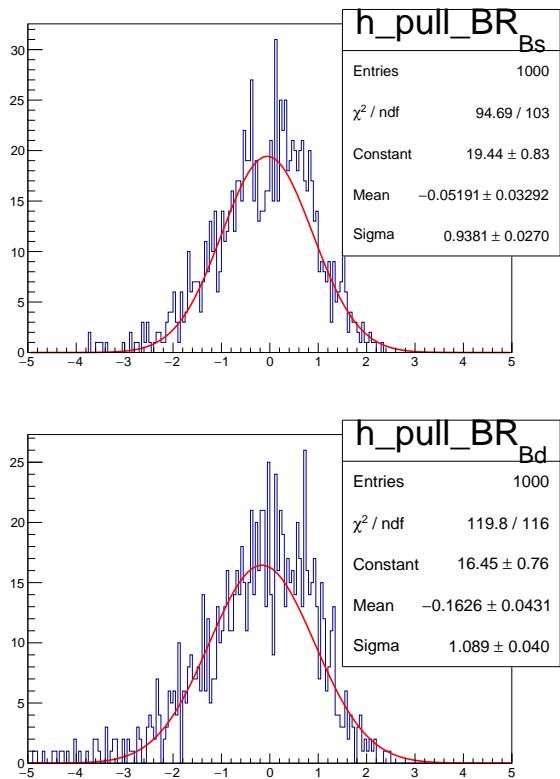


Figure 60: Pulls of $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ (left) and $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ (right) for the 7-bins configuration, Run 1 toys.

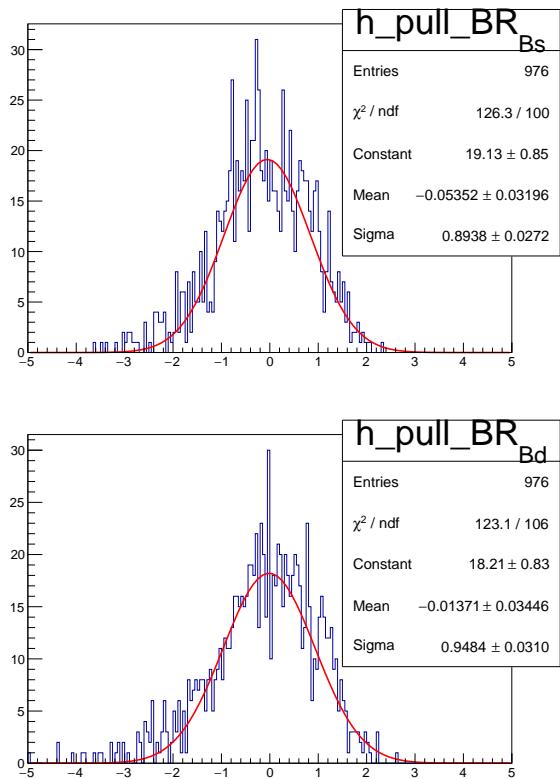


Figure 61: Pulls of $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ (left) and $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ (right) for the 5-bins configuration, Run 1 toys.

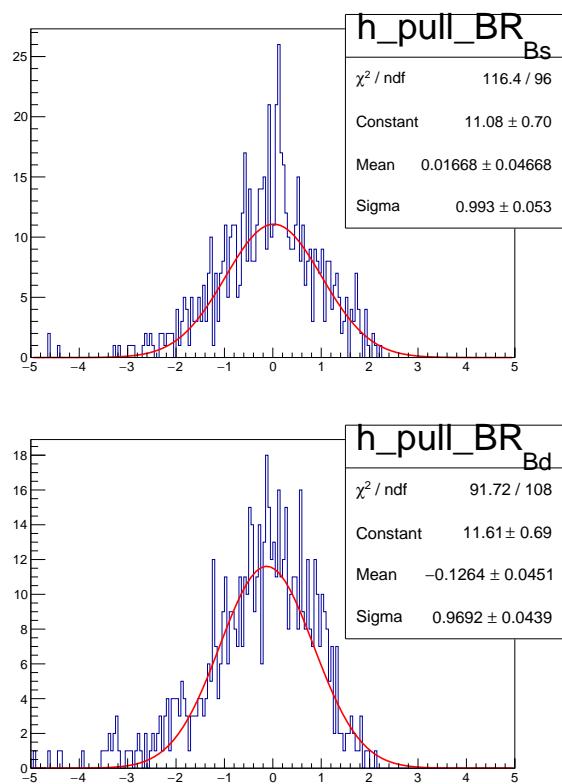


Figure 62: Pulls of $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ (left) and $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ (right) for the 4-bins configuration, Run 1 toys.

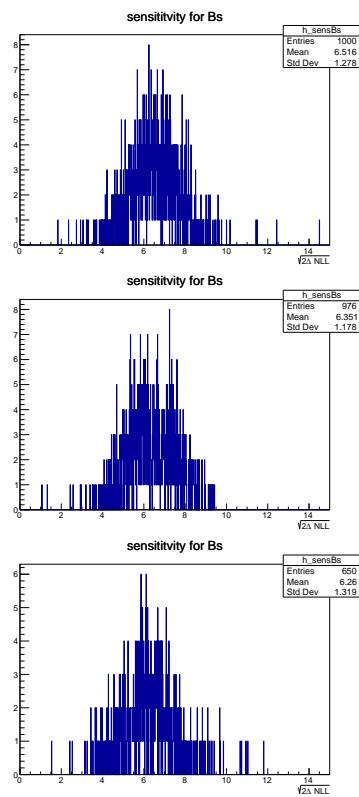


Figure 63: Expected B_s^0 significance for the Run 1 data fitted in 7-bin (top), 5-bin (middle), and 4-bin (bottom) configurations.

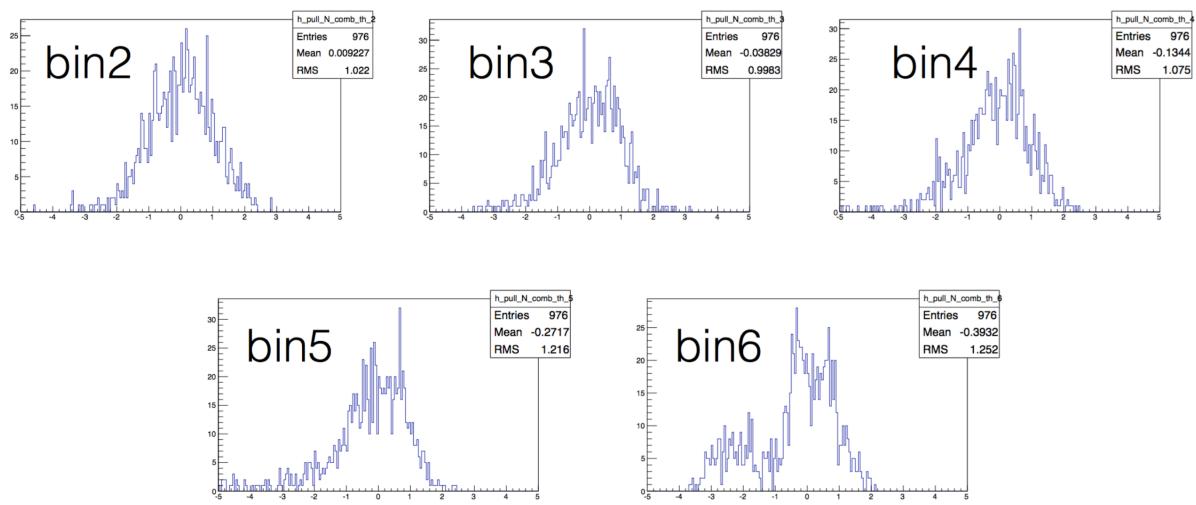


Figure 64: Fit pulls for the combinatorial background from the toy fits on Run 1, in the 5-bin fit configuration; bins are labelled 2 to 6 since the least significant bin is excluded from the fit.

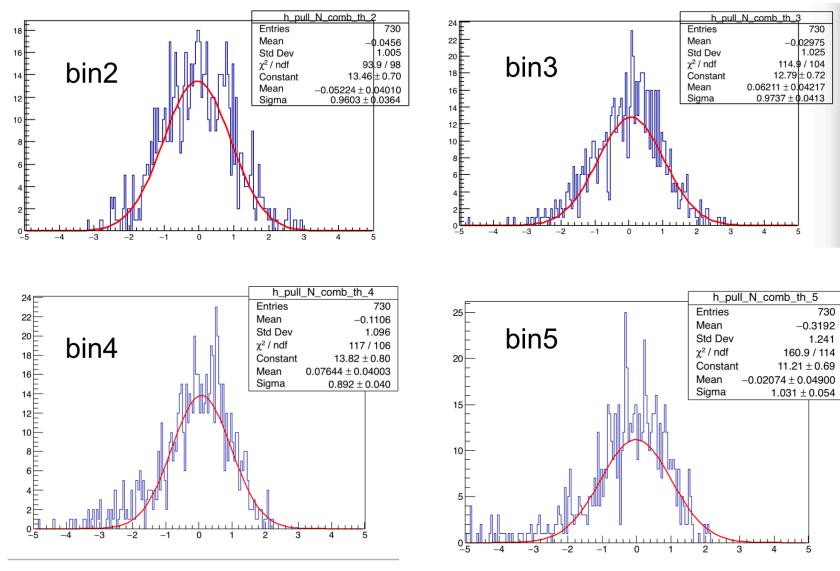


Figure 65: Fit pulls for the combinatorial background from the toy fits on Run 1, in the 4-bin fit configuration; bins are labelled 2 to 5 since the least significant bin is excluded from the fit.

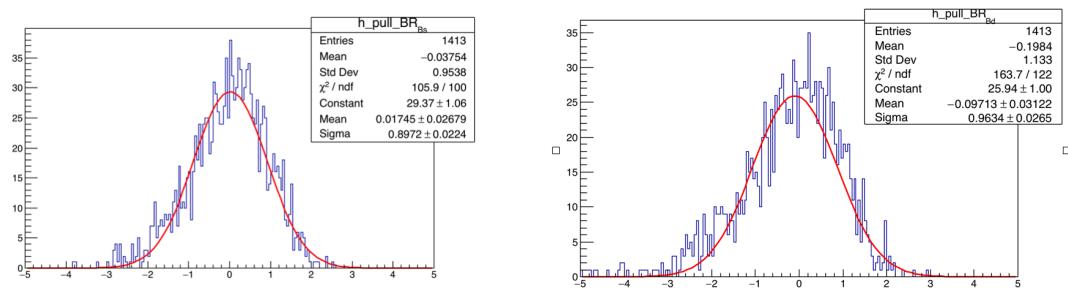


Figure 66: Fit pulls for the B_s^0 (left) and B^0 (right) branching fractions from the toys of the full dataset, fitted in the 4-bin configuration.

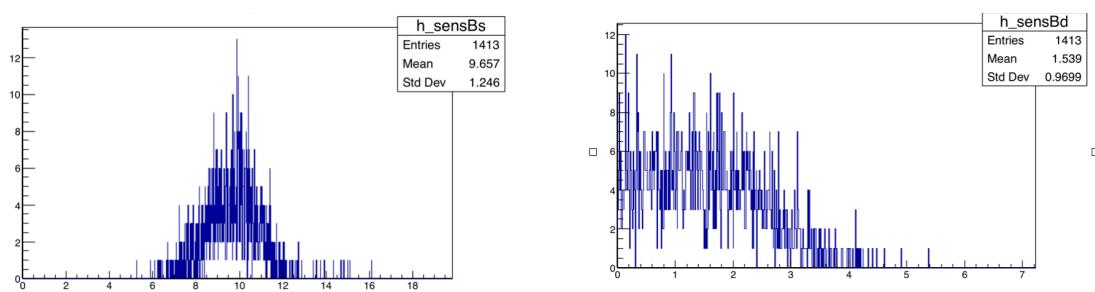


Figure 67: Expected B_s^0 (left) and B^0 (right) sensitivities from the toys of the full dataset, fitted in the 4-bin configuration.

1770 **9.2 Unblinded results**

1771 An unbinned maximum likelihood fit is performed simultaneously to the dimuon mass
1772 distributions in each BDT bin of Run1, 2015, and 2016 datasets. The $B_s^0 \rightarrow \mu^+ \mu^-$
1773 and $B^0 \rightarrow \mu^+ \mu^-$ fraction yields in each BDT bin are constrained to the BDT fractions
1774 determined in Sect. 6.4.2. These fractions and the parameters of the Crystal Ball functions
1775 that describe the mass shapes are Gaussian-constrained according to their expected values
1776 and uncertainties. The combinatorial background in each BDT bin is parameterised with
1777 an exponential function whose slope is common in bins of the same dataset while the yield
1778 is allowed to vary independently. The backgrounds from $B_{(s)}^0 \rightarrow h^+ h^-$, $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$,
1779 $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$, $B^{0(+)} \rightarrow \pi^{0(+)} \mu^+ \mu^-$, $\Lambda_b^0 \rightarrow p \mu^- \nu$ and $B_c^+ \rightarrow J/\psi(\mu^+ \mu^-) \mu^+ \nu_\mu$ shapes are
1780 evaluated in Sect. 5 . Their overall yields as well as the fractions in each BDT bin are bound
1781 with Gaussian constraints according to their expected values. The following result was
1782 found: $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (2.94^{+0.65}_{-0.57} \times 10^{-9}$. Surprisingly, the 2015 signal region contains
1783 no events in the $\text{BDT} > 0.4$ region. Numerous checks have been performed to search for
1784 a possible explanation other than statistical fluctuation. Two checks are discussed in
1785 Secs. 9.3 and 9.4. Furthermore, checks have been performed on the normalisation yields
1786 as a function of the data taking year, documented in Sec. 7.5. The compatibility of the fit
1787 results in different subsamples is documented in Sec. 9.5.

1788 No anomalies have been found. The final branching fraction measurements are de-
1789 termined from a simultaneous fit to the Run 1 and Run 2 datasets, as discussed in
1790 Sec. 9.6.

1791 **9.3 Probability to observe N events in the 2015 signal region**

1792 Using the expected number of candidates in the 2015 signal region (defined as the blinded
1793 region, $m_{\mu\mu} \in [5200, 5450] \text{ MeV}/c^2$) with $\text{BDT} > 0.4$, we compute the probability to observe
1794 0 candidates as found in data. The expected numbers of candidates are $N_{B_s} = 2.52 \pm 0.66$,
1795 $N_{B_d} = 0.36$, $N_{comb} = 2.15 \pm 0.83$ and $N_{excl} = 0.75 \pm 0.23$, where N_{comb} and N_{excl} are the
1796 expected combinatorial and total exclusive background, respectively. In the computation
1797 of the probability, the uncertainties on the expected number of events are treated as
1798 Gaussian. The resulting probability is 0.6%. A similar exercise is repeated for a looser
1799 PID selection where the muon candidates are only required to pass the `isMuon=1`. In
1800 this case 5 events are selected with $\text{BDT} > 0.4$, while the expected number of events are
1801 $N_{B_s} = 2.8 \pm 0.7$, $N_{B_d} = 0.4$, $N_{comb} = 3.6 \pm 1.5$ and $N_{excl} = 5.6 \pm 1.7$. The probability in
1802 this case is 3.7%.

1803 **9.4 Kolmogorov-Smirnov test of the BDT distribution in the
1804 2015 and 2016 datasets**

1805 The BDT distributions for events selected with $m_{\mu^+ \mu^-} \in [5200, 5450] \text{ MeV}/c^2$ in the 2015
1806 and 2016 datasets are shown in Fig. 68 for the standard ProbNN-based selection of
1807 the muon candidates, and for a looser selection where only `isMuon=1` is required. The

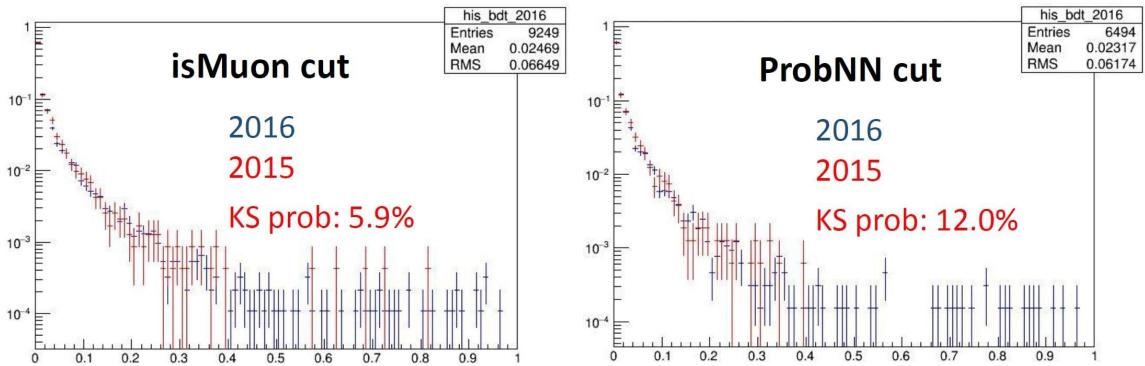


Figure 68: Comparison of the BDT distributions for events selected with $m_{\mu^+\mu^-} \in [5200, 5450] \text{ MeV}/c^2$ in the 2015 and 2016 datasets. Left: isMuon=1 selection; Right: standard ProbNN selection.

1808 Kolmogorov-Smirnov test returns probabilities of 5.9% and 12.0% for the isMuon=1 and
 1809 ProbNN selections, respectively, indicating that the bulks of the BDT distributions in
 1810 the two datasets do not differ much. The test run on the left or right mass sidebands
 1811 returns probabilities ranging between 15% and 70%. On the contrary, the tests performed
 1812 between the 2015 (or 2016) and Run 1 distributions return significantly lower probabilities.
 1813 It should be noted that the test has little sensitivity to differences in the tail of the
 1814 distributions.

1815 9.5 Compatibility of fit results from different data samples

1816 As a check, the fit was performed on different samples, and the compatibility of the measured
 1817 branching fraction evaluated. The compatibility of different fit results is evaluated by taking
 1818 $\sqrt{-2\Delta \log L}$, where the likelihoods from the simultaneous fit with common branching
 1819 fractions are compared with fits with independent branching fractions. For 2015, the
 1820 branching fraction is fixed to zero to avoid fit instabilities, while it is left floating for all
 1821 other datasets. The results are summarised in Tables 73 and 74.

Table 73: Fit results for $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-)$ for different data samples and fit configurations.

Data sample + fit configuration	$\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-)(10^{-9})$
Run 1	$1.88^{+0.77}_{-0.62}$
2015	< 2.9 at 95% CL
2016	$5.3^{+1.4}_{-1.2}$
Run 2	$3.95^{+1.1}_{-0.9}$
Run 1 + 2015 + 2016	$2.94^{+0.65}_{-0.57}$
Run 1 + Run 2	$2.82^{+0.63}_{-0.56}$

Table 74: Compatibility of $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ results obtained with the different data samples and fit configurations.

Comparison	Compatibility
Run I vs. 2016	2.3σ
Run I vs. Run II	1.7σ
Run I vs. 2015 vs. 2016	3.5σ

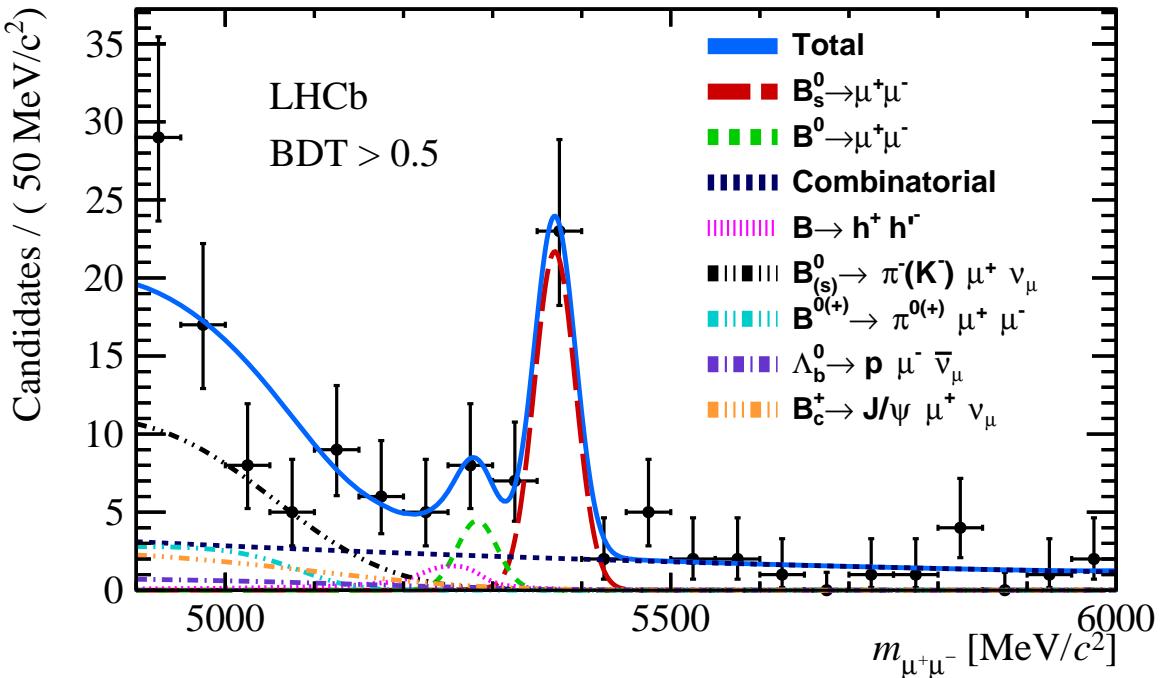


Figure 69: Mass distribution of the selected $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ candidates (black dots) with $BDT > 0.5$. The result of the fit is overlaid and the different components detailed.

1822 9.6 Fit to Run 1 and Run 2 datasets

1823 No problems have emerged in the checks discussed in the previous sections. We interpret
 1824 the depletion of events in the 2015 signal region with $BDT > 0.4$ as the result of a statistical
 1825 fluctuation. For the final result, the 2015 and 2016 datasets are merged and a simultaneous
 1826 fit is performed to the Run 1 and Run 2 datasets in 8 BDT bins (4 per dataset). The fit
 1827 result in the most sensitive region ($BDT > 0.5$) is shown in Figure 69. In addition, the fit
 1828 result per BDT bin is shown in Figure 70.

1829 From this fit, the following branching fraction measurements are obtained:

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (2.82^{+0.63}) \times 10^{-9} \quad (38)$$

$$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) = (1.6^{+1.1}) \times 10^{-10} \quad (39)$$

1830 The significance of the $B_s^0 \rightarrow \mu^+ \mu^-$ signal is 7.8σ , while the significance of the $B^0 \rightarrow \mu^+ \mu^-$
1831 signal is 1.9σ . This is the first observation of $B_s^0 \rightarrow \mu^+ \mu^-$ by a single experiment. As
1832 the $B^0 \rightarrow \mu^+ \mu^-$ significance is only 1.9σ , the CLs method is used to set a limit on the
1833 branching fraction. This limit is found to be $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) < 6.1 \times 10^{-10}$ at 95% CL.

1834 The nominal fit assumes $A_{\Delta\Gamma} = 1$ as predicted by the SM. In order to study the
1835 dependence of the $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction on the $A_{\Delta\Gamma}$ assumption, the fit is
1836 repeated in the hypotheses $A_{\Delta\Gamma} = 0$ and -1 using the corrections discussed in Sec. 8.
1837 The central value of the $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction increases by 4.6% and 10.9%, for
1838 $A_{\Delta\Gamma} = 0, -1$, respectively.

- 1839 • **NOTE 1:** sim09a MC samples have been used in the analysis (with data/MC
1840 tracking efficiencies corrections tuned on sim09a). Checks are planned during the
1841 1st circulation to compute the efficiencies using sim09b MC samples and compare
1842 the results.
- 1843 • **NOTE 2:** The $B^0 \rightarrow \mu^+ \mu^-$ UL estimate is preliminary and will probably change.
1844 The final estimate is planned to be finalised during the paper internal circulation.

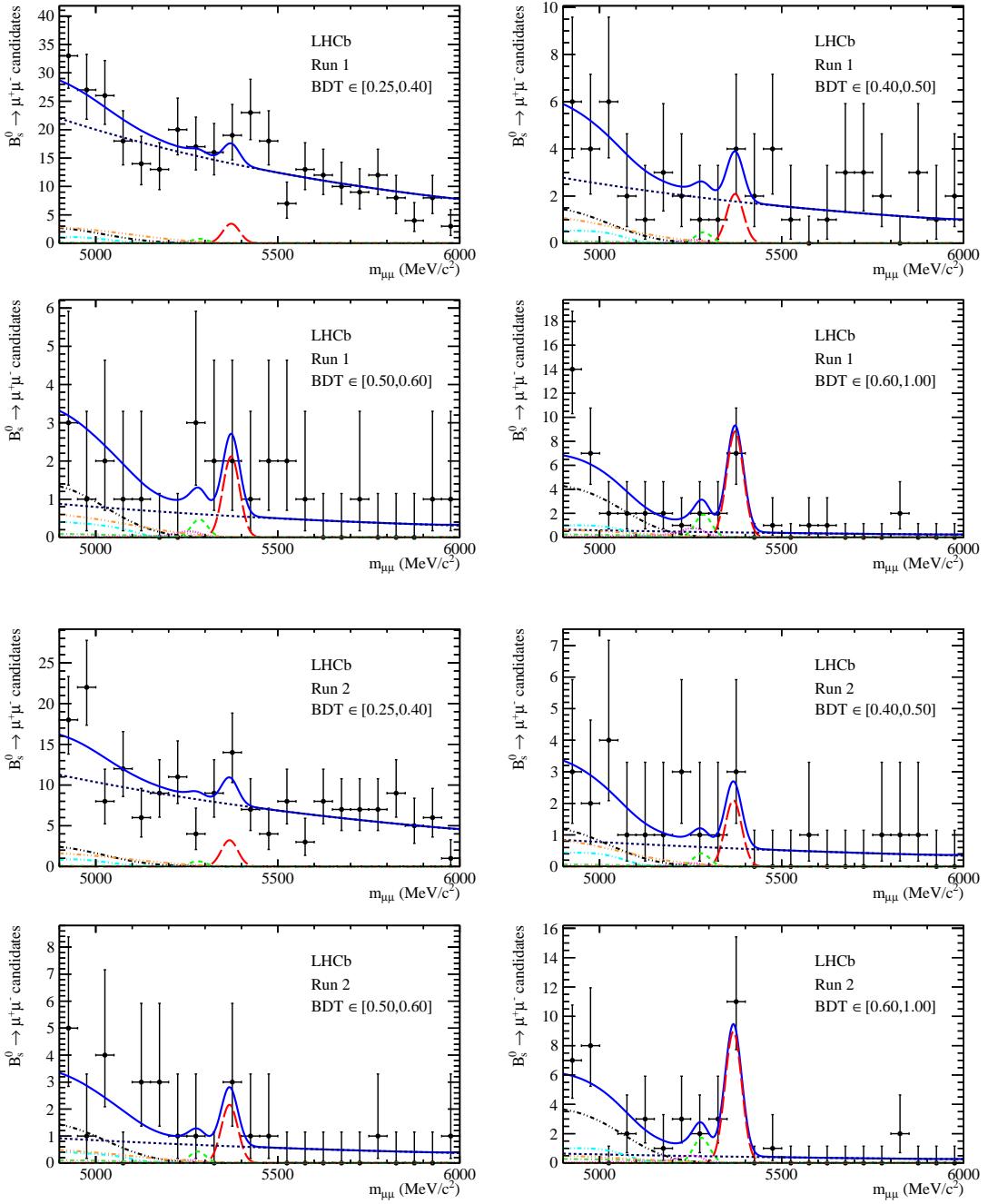


Figure 70: Mass distribution of the selected $B_{(s)}^0 \rightarrow \mu^+\mu^-$ candidates (black dots) in each bins of BDT. The result of the fit is overlaid (blue solid line) and the different components detailed: $B_s^0 \rightarrow \mu^+\mu^-$ (red long dashed), $B^0 \rightarrow \mu^+\mu^-$ (green medium dashed), combinatorial background (blue short dashed), $B_{(s)}^0 \rightarrow h^+h^-$ (magenta dotted), $B^0 \rightarrow \pi^-\mu^+\nu_\mu$ and $B_s^0 \rightarrow K^-\mu^+\nu_\mu$ (black dot-dashed), $B^{0(+)} \rightarrow \pi^{0(+)}\mu^+\mu^-$ (cyan dot-dashed), $B_c^+ \rightarrow J/\psi(\mu^+\mu^-)\mu^+\nu_\mu$ (orange dot-dashed) and $\Lambda_b^0 \rightarrow p\mu^-\nu$ (green dot-dashed).

1845 10 Effective lifetime measurement

1846 This section describes the method used to measure the effective $B_s^0 \rightarrow \mu^+ \mu^-$ lifetime.

1847 10.1 Data sample and selection

1848 The measurement of the effective lifetime is made on the same data sample as the branching
1849 fraction analysis; that is the 2011, 2012, 2015 data samples and part of the 2016 data
1850 sample. This data sample shall henceforth be referred to as the ‘CKM16’ data sample for
1851 the sake of brevity.

1852 The data are selected in a similar way as the branching fraction measurements. The
1853 same stripping and pre-selection requirements are imposed, but different PID, BDT and
1854 trigger requirements are used. Instead of fitting in bins of BDT response (as in the BF
1855 measurement) a single requirement is made on the BDT response and a softer particle
1856 identification (PID) requirement is also imposed. These are described in the following two
1857 sections.

1858 Initially the trigger requirements for the effective lifetime measurement were the same
1859 as those used in the branching fraction measurements where no specific requirements are
1860 imposed. These requirements were used in previous versions of the analysis note. However,
1861 imposing no specific trigger requirements means that candidates could be selected by
1862 trigger decisions for lines relevant for selecting J/ψ or Z candidates but passing neither
1863 TIS nor TOS decisions of these lines. The presence of these events would mean that
1864 the shape of the acceptance function extracted from MC (see Section 10.2) could not be
1865 properly controlled. Table 75 shows the efficiencies for 2016 $B_s^0 \rightarrow \mu^+ \mu^-$ MC events that
1866 pass the pre-selection, BDT, PID and trigger requirements at the L0 and Hlt1 levels of
1867 events that are trigger by Hlt2 lines relevant for J/ψ or Z candidates. A small fraction of
events are passed by trigger decisions of these lines but not by the TIS or TOS decisions.

Trigger lines	Dec	TIS	(Dec && !TOS &&!TIS)
Hlt2DiMuonJPsiDecision	3.31%	0.10 %	3.21 %
Hlt2DiMuonPsi2SDecision	2.14%	0.02 %	2.12%
Hlt2DiMuonJPsiHighPTDecision	3.22 %	0.07 %	3.15 %
Hlt2DiMuonDetachedJPsiDecision	2.19 %	0.04 %	2.15 %
Hlt2DiMuonZDecisionTOB	0.00 %	0.00 %	0.00 %

Table 75: Efficiencies for Hlt2 lines relevant for events including J/ψ or Z candidates for 2016 $B_s^0 \rightarrow \mu^+ \mu^-$ MC events passing the pre-selection, BDT, PID, and trigger requirements at the L0 and Hlt1 levels. Efficiencies for being triggered on signal (TOS) are zero for each line.

1868 Therefore the effective lifetime, $B_s^0 \rightarrow \mu^+ \mu^-$ candidates are required to pass either TIS or
1869 TOS triggers at L0, Hlt1 and Hlt2, specifically (B_L0Global_TIS == 1 || B_L0Global_TOS
1870 == 1) && (B_Hlt1Phys_TIS == 1 || B_Hlt1Phys_TOS == 1) && (B_Hlt2Phys_TIS == 1
1871

1872 || B_Hlt2Phys_TOS == 1). The change in overall selection efficiency for requiring TIS
 1873 or TOS triggers at each level instead of imposing not specific requirements Dec is very
 1874 small, this is shown in Table 76. The efficiencies are computed for events that pass the
 1875 full selection, including PID and BDT requirements.

Year	Efficiency
2011	0.988
2012	0.999
2015	0.997

Table 76: Efficiency of $B_s^0 \rightarrow \mu^+ \mu^-$ MC events passing the full selection, including PID and BDT requirements, and either TIS or TOS triggers at L0, Hlt1 and Hlt2 relative to the events passing the full selection, including PID and BDT requirements, and either Dec triggers at L0, Hlt1 and Hlt2. No efficiency is included for 2016 due to a problem with the Hlt2 trigger decision written to the ntuples such that Hlt2Phys Dec was always equal to one.

1876 For the 2016 Monte Carlo simulated samples there was a problem with the Hlt2 trigger
 1877 decision written to the ntuples such that Hlt2Phys Dec was always equal to one. This
 1878 line is used for studies with $B \rightarrow h^+ h^-$ decays in this analysis, therefore instead of using
 1879 Hlt2Phys in 2016 MC and data requirements were imposed on specific lines for $B \rightarrow h^+ h^-$
 1880 events. The list of trigger lines at Hlt2 used instead of Hlt2Phys for 2016 data and MC
 1881 are given in Table 77.

$B \rightarrow h^+ h^-$ trigger lines
Hlt2Topo2BodyDecision Dec
Hlt2B2HH Lb2PPiDecision Dec
Hlt2B2HH Lb2PKDecision Dec
Hlt2B2HH B2PiPiDecision Dec
Hlt2B2HH B2PiKDecision Dec
Hlt2B2HH B2KKDecision Dec
Hlt2B2HH B2HHDecision Dec

Table 77: Trigger lines that $B \rightarrow h^+ h^-$ candidates are required to pass at Hlt2 for 2016 data (the logical OR of the lines is imposed).

1882 10.1.1 PID requirements

1883 As in the BF analysis, a requirement on the quantity

$$\text{PID}_\mu \equiv \text{ProbNN}_\mu \times (1 - \text{ProbNN}_p) \times (1 - \text{ProbNN}_K) \quad (40)$$

is imposed in order to suppress contamination from exclusive backgrounds, particularly from $B \rightarrow h^+h'^-$, $B^0 \rightarrow \pi^-\mu^+\nu_\mu$, $B_s^0 \rightarrow K^-\mu^+\nu_\mu$ and $\Lambda_b^0 \rightarrow p\mu^-\bar{\nu}_\mu$ decays which involve a hadron being mis-identified as a muon. For 2011, 2012 and 2015 data the MC12TuneV2 ProbNN variables are used, for 2016 MC15TuneV1 ProbNN are chosen as these give better performance.

As can be seen in Figure 71, $B \rightarrow h^+h'^-$ and $\Lambda_b^0 \rightarrow p\mu^-\bar{\nu}_\mu$ decays are the most problematic for the measurement as they lie highest in invariant mass, close to the $B^0 \rightarrow \mu^+\mu^-$ and $B_s^0 \rightarrow \mu^+\mu^-$ peaks. As the branching fraction measurement is particularly targeting the $B^0 \rightarrow \mu^+\mu^-$ decay, it is important to control these backgrounds and so a harder cut is imposed on both muons. However, since these backgrounds hardly enter the $B_s^0 \rightarrow \mu^+\mu^-$ signal region, softer cuts are imposed for the lifetime measurement. The requirements are:

$$\text{PID}_\mu(\text{MC12TuneV2}) > 0.2 \quad (41)$$

$$\text{PID}_\mu(\text{MC15TuneV2}) > 0.4 \quad (42)$$

where the first requirement is imposed on 2011, 2012 and 2015 data and the second on 2016 data. The 2016 requirement was chosen to give similar background rejection as the MC12TuneV2 cut but has a higher efficiency on muons.

Any remaining contamination has a negligible effect on the measurement, as described in Section 10.4.2. This cut has not been optimised in the same way as the BDT requirement (see Section 10.1.2) since PID variables are not well reproduced in Monte Carlo simulations and there are too few events in the high-mass data sideband once the selection has been applied.

10.1.2 BDT requirement optimisation

Rather than performing a fit in bins of BDT output, a simple cut is imposed. This was necessary to ensure a stable fit; a simultaneous fit in bins of BDT was investigated but proved to be unstable and inaccurate.

The requirement on the BDT output was chosen to minimise the statistical uncertainty on the lifetime measurement. Using the toy model described in Appendix A, 10,000 pseudo-experiments were performed for each cut value and the median uncertainty on the lifetime and inverse lifetime was calculated.

The numbers of candidates generated in each pseudo-experiment vary depending upon the BDT requirement. The starting point for this calculation is the number of candidates expected in the CKM16 data sample, calculated using the same method as for the branching fraction analysis but with the softer PID requirements described earlier and a default BDT1flat requirement at 0.55. These numbers are given in Table 78.

In order to calculate the expected yields of $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$ decays at different values of the minimum requirement on BDT1flat it is assumed that these decays are distributed evenly across the BDT1flat range. In the case of combinatorial background the ratio

$$R_\epsilon = \frac{\epsilon(\text{BDT1flat} > X)}{\epsilon(\text{BDT1flat} > 0.55)} \quad (43)$$

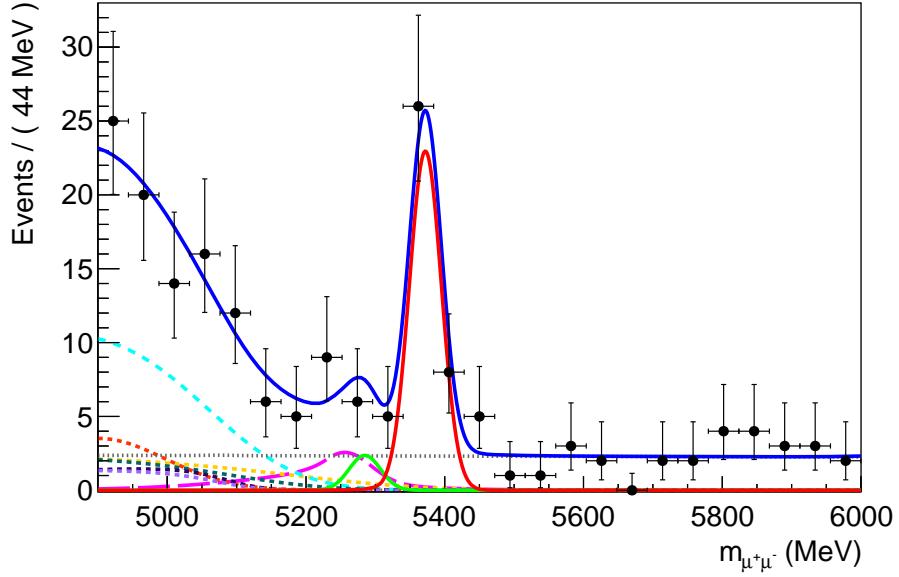


Figure 71: The di-muon invariant mass spectrum generated for a typical toy pseudoexperiment, with the PDFs used to generate the data overlaid. The most important components are as follows: $B_s^0 \rightarrow \mu^+ \mu^-$ solid red line, $B^0 \rightarrow \mu^+ \mu^-$ solid green line, combinatorial background narrow-dashed grey line, $B \rightarrow h^+ h^-$ long-dashed magenta line, $B^0 \rightarrow \pi^0 \mu^+ \mu^-$ medium-dashed cyan line and $\Lambda_b^0 \rightarrow p \mu^- \bar{\nu}_\mu$ short-dashed orange line. The remaining components are $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$, $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$, $B^+ \rightarrow \pi^+ \mu^+ \mu^-$ and $B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu$. The statistics correspond to those expected for the CKM16 data sample with the lifetime PID selection imposed and $\text{BDT1flat} > 0.55$.

where $\epsilon(\text{BDT1flat} > X)$ is the efficiency of $\text{BDT1flat} > X$, is calculated for a range of BDT requirements using $b\bar{b} \rightarrow X \mu^+ \mu^-$ Monte Carlo simulated events. The results are listed in Table 79.

It is also observed that the slope of the exponential mass PDF (λ) used to describe/generate the combinatorial background component varies with BDT requirement. To ensure that the toys accurately reflect the expected distributions in data as the BDT cut is varied, λ was evaluated for a range of BDT requirements using $b\bar{b} \rightarrow X \mu^+ \mu^-$ combinatorial background Monte Carlo simulated events. These values are listed in Table 80.

Decay mode	Expected number of candidates
$B_s^0 \rightarrow \mu^+ \mu^-$	30.94
$B^0 \rightarrow \mu^+ \mu^-$	3.27
Combinatorial background	66.23
Total	100.44

Table 78: The expected number of candidates for each decay mode in the CKM16 data sample corresponding to a $\text{BDT1flat} > 0.55$ and in the mass range 4900-6000 MeV. Numbers correspond to the expected numbers of each decay mode based on SM predictions, with the exception of the combinatorial background which is based on an extrapolation from the 3 fb^{-1} branching fraction analysis, taking into account the new selection efficiency and increased data sample size.

BDT1flat cut	R_ϵ
0.40	8.69
0.45	3.91
0.50	1.91
0.55	1.00
0.60	0.55
0.65	0.32

Table 79: Ratio of the efficiency of a range of BDT1flat requirements to $\text{BDT1flat} > 0.55$ for combinatorial background candidates, calculated using $b\bar{b} \rightarrow X\mu^+\mu^-$ MC simulated events.

BDT cut	$\lambda / \text{c}^2 \text{MeV}^{-1}$
0.40	-0.00114 ± 0.00028
0.45	-0.00129 ± 0.00041
0.50	-0.00132 ± 0.00060
0.55	-0.00004 ± 0.00089
0.60	-0.00000 ± 0.00114
0.65	-0.00024 ± 0.00122

Table 80: Values of the slope of the combinatorial background mass PDF used to generate pseudo-experiments as a function of BDT requirement. The values were determined by performing fits to the mass spectra of $b\bar{b} \rightarrow X\mu^+\mu^-$ Monte Carlo simulated events.

1929 Using these numbers, 10,000 pseuodoexperiments are performed for each BDT cut
 1930 value and the results are presented in Table 81. Each pseudo-experiment is generated with
 1931 a cut of $m_{\mu^+\mu^-} > 5320$ MeV and only $B_s^0 \rightarrow \mu^+\mu^-$ decays and combinatorial background
 1932 are generated. Fits for both $\tau_{\mu^+\mu^-}$ and $\tau_{\mu^+\mu^-}^{-1}$ are performed for each pseudoexperiment
 1933 (see Section 10.3). For each BDT cut the signal significance is evaluated and the values of
 1934 statistical uncertainties and the mean and widths of the pulls for the fit of $\tau_{\mu^+\mu^-}^{-1}$ are given.
 1935 A cut of $BDT > 0.55$ returns the highest signal significance and the lowest statistical
 1936 uncertainties on both the lifetime and the inverse lifetime and is selected as the optimum
 1937 cut.

BDT cut	$S/\sqrt{(S+B)}$	$\sigma(\tau_{\mu^+\mu^-})$ / ps	$\sigma(\tau_{\mu^+\mu^-}^{-1})$ / ps $^{-1}$	$\tau_{\mu^+\mu^-}^{-1}$ pull mean	$\tau_{\mu^+\mu^-}^{-1}$ pull width
0.40	3.87	0.345	0.128	-0.01 ± 0.01	1.020 ± 0.007
0.45	4.51	0.309	0.114	-0.02 ± 0.01	1.014 ± 0.007
0.50	4.85	0.291	0.108	-0.01 ± 0.01	1.029 ± 0.007
0.55	4.94	0.285	0.106	0.00 ± 0.01	1.010 ± 0.007
0.60	4.86	0.297	0.109	-0.02 ± 0.01	0.996 ± 0.007
0.65	4.65	0.309	0.115	-0.01 ± 0.01	1.000 ± 0.007

Table 81: The signal significance of $B_s^0 \rightarrow \mu^+\mu^-$ decays and median uncertainties and mean and width of the pulls of a fit for $\tau_{\mu^+\mu^-}^{-1}$ as a function of BDT requirement from toy MC studies. 10,000 pseudo-experiments were conducted for each cut value, and only $B_s^0 \rightarrow \mu^+\mu^-$ decays and combinatorial background candidates were generated.

1938 10.2 Decay time acceptance

1939 The trigger and selection efficiency varies as a function of the decay time of $B_s^0 \rightarrow \mu^+ \mu^-$
1940 candidates, t , introducing a bias into the decay time distribution. This must be corrected
1941 for in the fit. The decay time efficiency ('acceptance') is modelled using the function:

$$\epsilon(t) = \frac{[a(t - t_0)]^n}{1 + [a(t - t_0)]^n}. \quad (44)$$

1942 The parameters of the function are determined by fits to full LHCb Monte Carlo simulated
1943 $B_s^0 \rightarrow \mu^+ \mu^-$ candidates and then fixed in the lifetime fit to real data. The simulated data
1944 are selected using the same trigger, stripping and offline selection requirements as for the
1945 real data.

1946 Monte Carlo models the data reasonably well for most variables, however there are
1947 significant differences between event multiplicity variables such as the number of SPD hits
1948 or tracks in the event between data and MC. Although these variables are not themselves
1949 directly correlated with decay time, the VELO and long track isolation variables that enter
1950 the BDT classifier are affected by these quantities, and are themselves correlated with
1951 the decay time of the candidate. Therefore, significant differences in event multiplicity
1952 between data and MC may introduce a bias in the way the acceptance is modelled in
1953 Monte Carlo.

1954 To remedy this, the Monte Carlo simulated events are reweighted so that they have
1955 the same distribution of the number of tracks in each event as data. This is done using
1956 $B^0 \rightarrow K^+ \pi^-$ Monte Carlo simulated candidates and $B^0 \rightarrow K^+ \pi^-$ decays in data. The
1957 weights are calculated as follows. Both MC and data $B^0 \rightarrow K^+ \pi^-$ candidates are required
1958 to pass the trigger, stripping and pre-selection. A further PID cut is used to separate
1959 $B^0 \rightarrow K^+ \pi^-$ events from other $B \rightarrow h^+ h^-$ channels;

$$(\text{DLL}k^+ > 10 \quad \& \quad \text{DLL}k^- < -10) \quad || \quad (\text{DLL}k^+ < -10 \quad \& \quad \text{DLL}k^- > 10). \quad (45)$$

1960 The distribution of the number of tracks in Monte Carlo is then plotted, alongside the one
1961 from data extracted using sWeights calculated using a fit to the invariant mass distribution
1962 of $B^0 \rightarrow K^+ \pi^-$ candidates. The resulting distributions for 2011, 2012, 2015 and 2016 are
1963 shown in Figure 72.

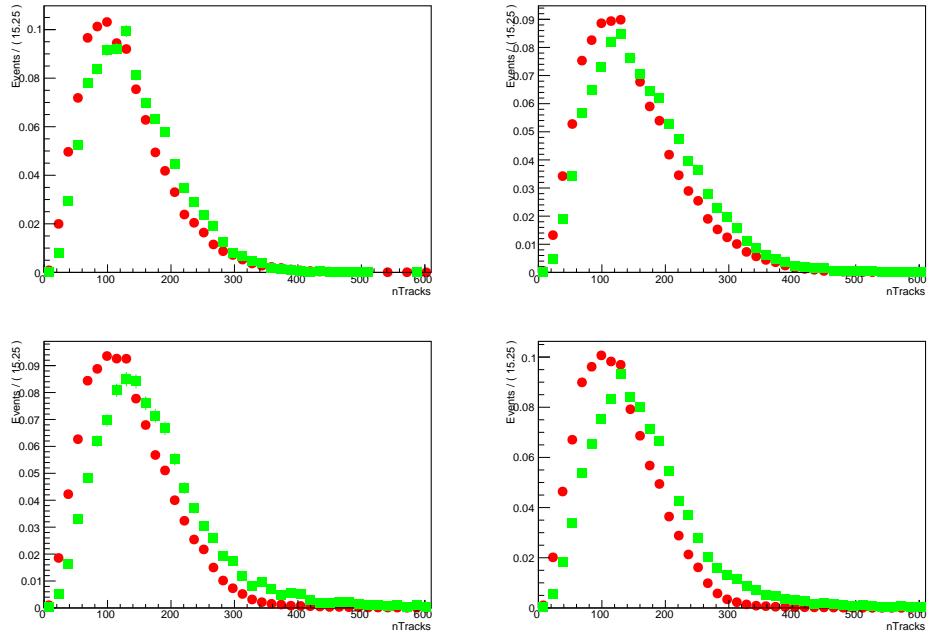


Figure 72: The number of tracks for $B^0 \rightarrow K^+ \pi^-$ candidates for 2011 (top left), 2012 (top right), 2015 (bottom left) and 2016 (bottom right), for Monte Carlo simulated events (red circles) and sWeighted data (green triangles). The histograms have been normalised to have unit area and so the y axis scale is arbitrary.

1964 The simulated events are then given weights equal to the ratio of the normalised
 1965 histogram entries shown in Figure 72 in bins of the number of tracks. Applying these
 1966 weights causes the Monte Carlo to have the same number of tracks distribution as data.
 1967 The effect of this re-weighting on the decay time distributions of simulated $B^0 \rightarrow K^+ \pi^-$
 1968 candidates after the BDT cut and TIS triggers are required at L0, Hlt1 and Hlt2 can be
 1969 seen in Figure 73. We see that the largest shifts due to the weighting are at low decay
 1970 time where the acceptance rises sharply.

1971 These weights calculated from $B^0 \rightarrow K^+ \pi^-$ data and MC are then used to reweight
 1972 $B_s^0 \rightarrow \mu^+ \mu^-$ MC candidates. This assumes that the number of tracks distributions are
 1973 the same for $B^0 \rightarrow K^+ \pi^-$ and $B_s^0 \rightarrow \mu^+ \mu^-$ events. This cannot be tested directly in data,
 1974 as there are too few $B_s^0 \rightarrow \mu^+ \mu^-$ decays, however, a comparison can be made in Monte
 1975 Carlo before any reweighting is applied. Comparisons of the number of tracks between
 1976 $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow K^+ \pi^-$ decays in Monte Carlo before weighting are shown in Figure
 1977 74, and we see that the agreement is rather good although $B_s^0 \rightarrow \mu^+ \mu^-$ decays tend to
 1978 have slightly more tracks in the event.

1979 To re-weight the $B_s^0 \rightarrow \mu^+ \mu^-$ Monte Carlo the data are split into the same bins of the
 1980 number of tracks as $B^0 \rightarrow K^+ \pi^-$ and the same weights are then applied. The effect of
 1981 the re-weighting on the decay time distribution of $B_s^0 \rightarrow \mu^+ \mu^-$ MC is shown in Figure 75.
 1982 Again we see that the effect of the re-weighting is relatively small but is most pronounced

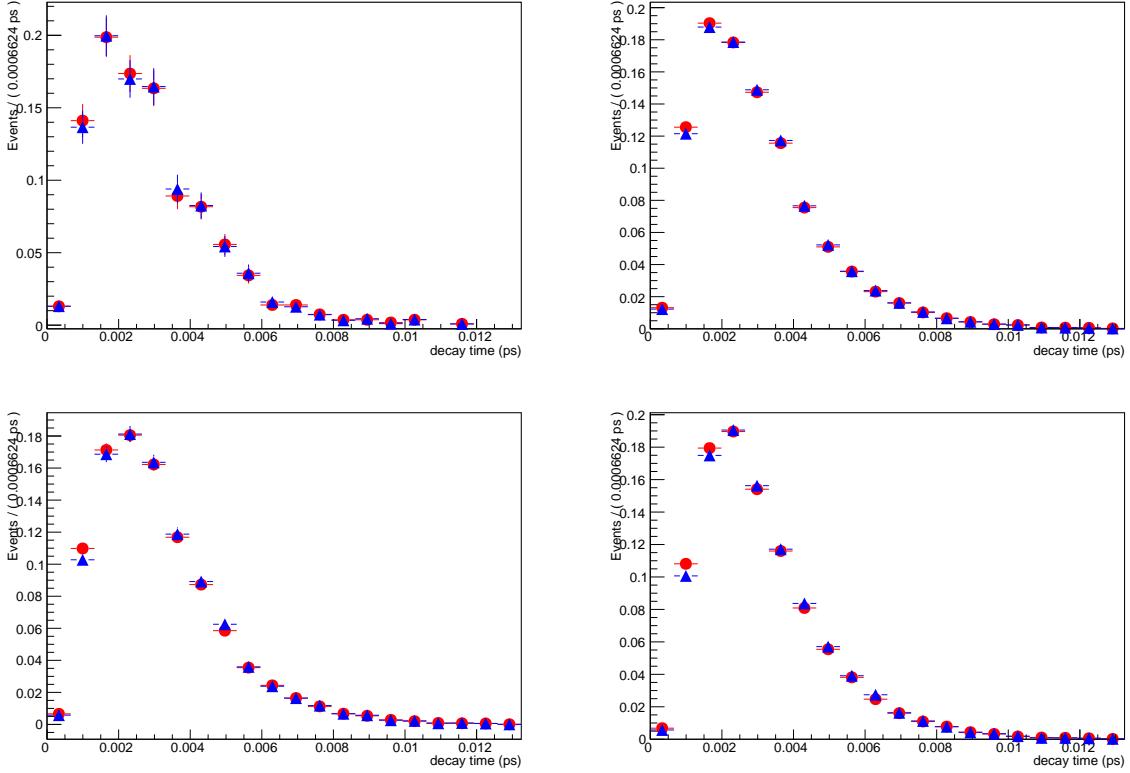


Figure 73: The decay time distributions for $B^0 \rightarrow K^+ \pi^-$ simulated candidates for 2011 (top left), 2012 (top right), 2015 (bottom left) and 2016 (bottom right), before (red circles) and after (blue squares) reweighting by the number of tracks, after the BDT cut and TIS trigger requirements have been applied.

at low decay time.

Since the beam energy and trigger conditions differ between 2011, 2012, 2015 and 2016 data, simulated events generated with 2011, 2012, 2015 and 2016 are mixed together in a cocktail with weights that ensure the same proportions of each year as are present in real data. The weights for each year are

$$\omega_i = \frac{Y_i^{J/\psi\phi} \epsilon_i}{\sum_j Y_j^{J/\psi\phi} \epsilon_j} \frac{\sum_k N_k^{\mu^+\mu^-}}{N_i^{\mu^+\mu^-}} \quad (46)$$

where $i = 2011, 2012, 2015$ or 2016 , $Y_i^{J/\psi\phi}$ is the fitted yield of $B_s^0 \rightarrow J/\psi\phi$ candidates passing the trigger, stripping and pre-selection requirements (but excluding BDT and PID requirements) in real data for each year, ϵ_i is the efficiency of the combined BDT and PID requirements for $B_s^0 \rightarrow \mu^+\mu^-$ truth-matched Monte Carlo simulated candidates with respect to those already passing the reconstruction, trigger, stripping and pre-selection

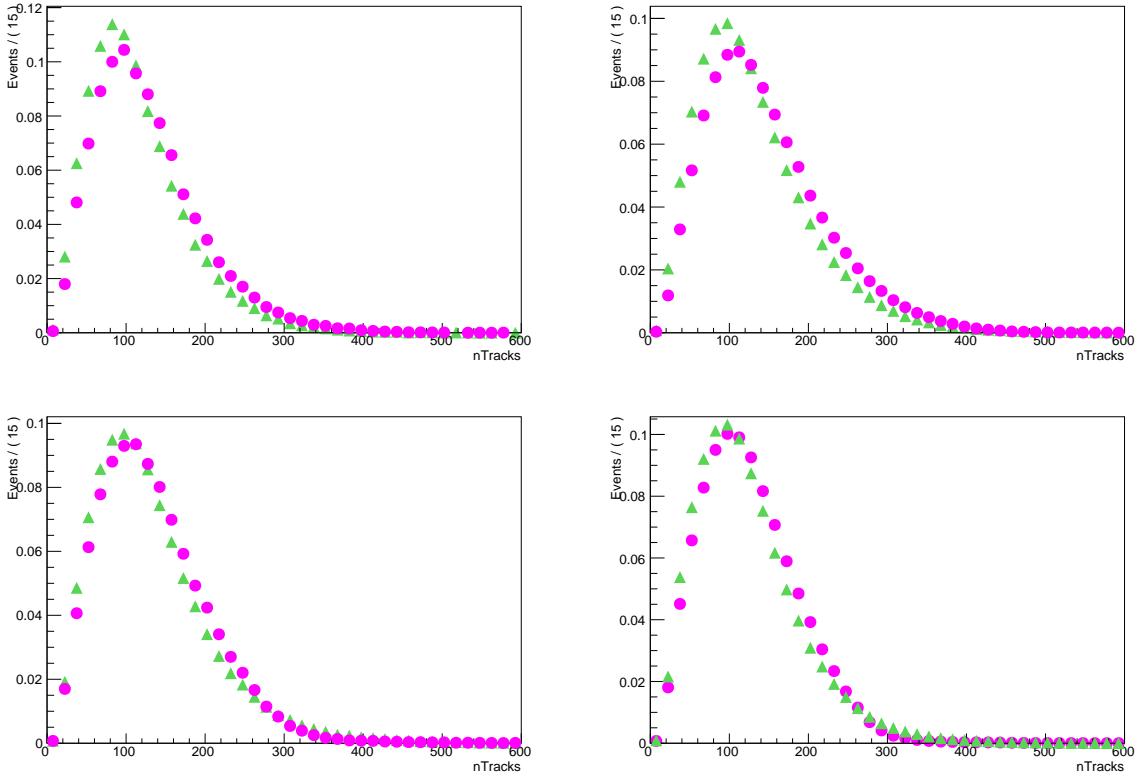


Figure 74: The number of tracks for $B^0 \rightarrow K^+\pi^-$ (green triangles) and $B_s^0 \rightarrow \mu^+\mu^-$ (pink circles) decays in Monte Carlo simulated events after the trigger, stripping and pre-selection but before any BDT cut, for each year: 2011 (top left), 2012 (top right), 2015 (bottom left) and 2016 (bottom right).

1993 requirements and $N_i^{\mu^+\mu^-}$ is the number of $B_s^0 \rightarrow \mu^+\mu^-$ truth-matched Monte Carlo
 1994 simulated candidates passing the reconstruction, trigger, stripping pre-selection, BDT and
 1995 PID requirements in the available simulated samples for each year.

1996 As such $B_s^0 \rightarrow J/\psi\phi$ decays, selected using identical criteria as $B_s^0 \rightarrow \mu^+\mu^-$ decays,
 1997 are used to normalise for the product of the integrated luminosity, B_s^0 production cross
 1998 section, reconstruction, trigger and stripping efficiencies for each year. Table 82 shows the
 1999 weights used to reweight each MC sample, along with the information used to calculate
 2000 the weights.

2001 A fit is then performed to the resulting decay time distribution where the parameters
 2002 of the function t_0 , a and n are allowed to float freely and the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime
 2003 is fixed to a value obtained from a fit to the generator level decay time distribution. Figure
 2004 76 shows the resulting fit, which has accurate pulls and residuals, and the final values of
 2005 the fit parameters are listed in Table 83. The same fit procedure has been applied to the
 2006 MC for each year separately and the results for the acceptance fit parameters are listed in
 2007 Table 84. The parameter values for each year are on the whole, not consistent with each

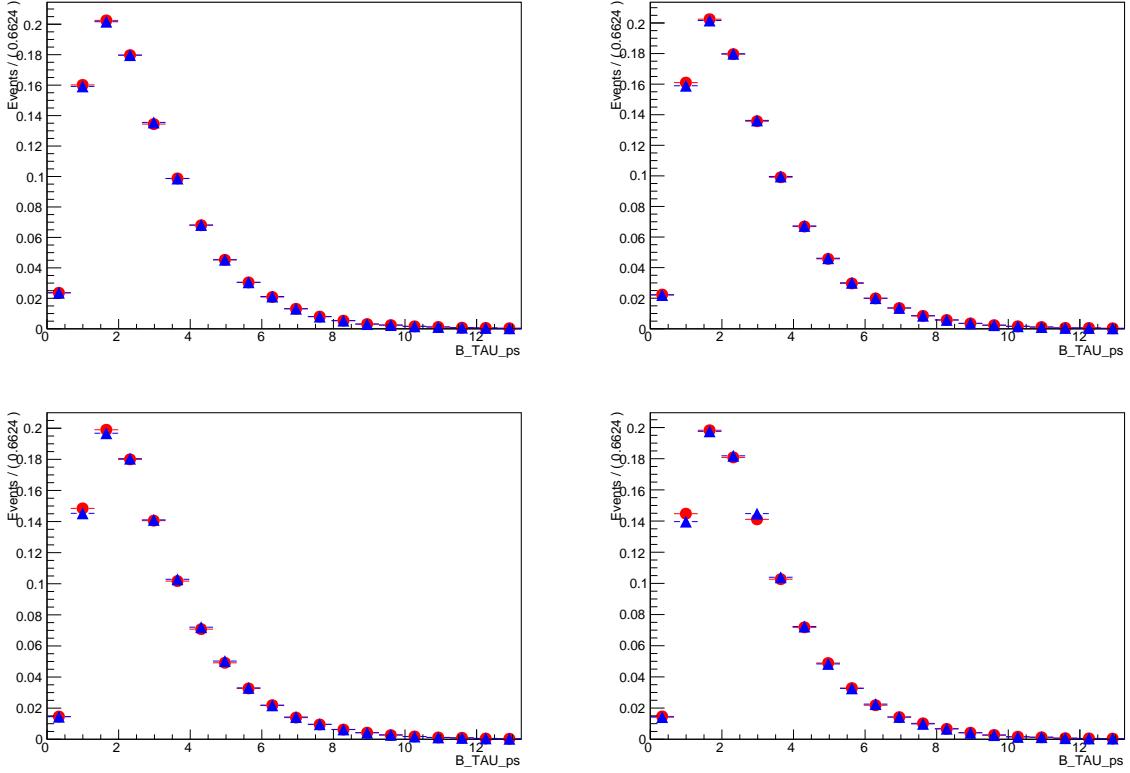


Figure 75: The decay time distributions for $B_s^0 \rightarrow \mu^+ \mu^-$ simulated candidates for 2011 (top left), 2012 (top right), 2015 (bottom left) and 2016 (bottom right), before (red circles) and after (blue squares) reweighting by the number of tracks, after the BDT cut and trigger requirements have been applied.

Year (i)	$Y_i^{J/\psi\phi}$	ϵ_i	$N_i^{\mu^+\mu^-}$	ω_i	$\mathcal{N}_i^{\mu^+\mu^-} \equiv N_i^{\mu^+\mu^-} \omega_i$
2011	19190	0.412	70448	1.72	131364
2012	42103	0.406	254822	1.03	262461
2015	8571	0.410	222820	0.24	53917
2016	37765	0.406	124870	1.88	235218

Table 82: Weights used to re-weight Monte Carlo simulated $B_s^0 \rightarrow \mu^+ \mu^-$ decays to create a cocktail where the proportions of each year are equivalent to in real data. $\mathcal{N}_i^{\mu^+\mu^-}$ is the weighted number of candidates in each MC sample.

²⁰⁰⁸ other.

²⁰⁰⁹ An alternative acceptance functional form was also tested

$$\epsilon(t) = \frac{[a(t - t_0)]^n}{1 + [a(t - t_0)]^n} e^{-\delta \Gamma t}, \quad (47)$$

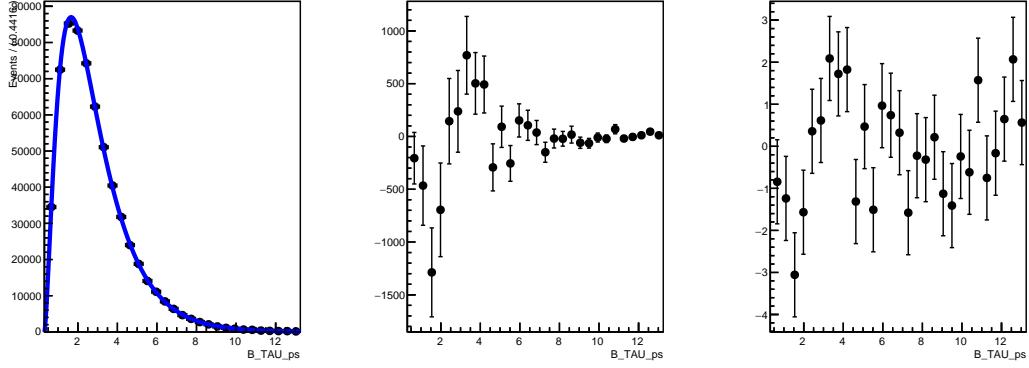


Figure 76: Fit for the acceptance function parameters made to truth-matched $B_s^0 \rightarrow \mu^+ \mu^-$ simulated candidates (left) and the resulting residual (centre) and pull (right) distributions. The data shown are a weighted cocktail of Monte Carlo simulated candidates generated with 2011, 2012, 2015 and 2016 conditions.

Parameter	Value
t_0	$0.313 \pm 0.007 \text{ ps}$
a	$0.574 \pm 0.011 \text{ ps}^{-1}$
n	1.485 ± 0.030

Table 83: Values of the parameters of the acceptance function used to model decay time-dependent trigger and selection efficiencies determined from a fit to truth-matched $B_s^0 \rightarrow \mu^+ \mu^-$ simulated candidates.

in which $\delta\Gamma$ takes into account acceptance effects at upper lifetimes. However using this acceptance function to fit the MC, returned values for $\delta\Gamma$ at or extremely close to zero showing that at large lifetimes the acceptance function is flat with no additional acceptance effects. Furthermore the flat upper lifetime acceptance is illustrated in Figure 77 which shows a histogram of the selection efficiency of 2011, 2012, 2015 and 2016 MC as a function of decay time.

Parameter	2011	2012	2015	2016
t_0 / ps	0.301 ± 0.010	0.313 ± 0.004	0.356 ± 0.007	0.325 ± 0.016
a / ps^{-1}	0.571 ± 0.018	0.571 ± 0.010	0.567 ± 0.013	0.585 ± 0.020
n	1.42 ± 0.04	1.408 ± 0.019	1.453 ± 0.030	1.582 ± 0.065

Table 84: Values of the parameters of the acceptance function used to model decay time-dependent trigger and selection efficiencies determined from a fit to truth-matched $B_s^0 \rightarrow \mu^+ \mu^-$ simulated candidates.

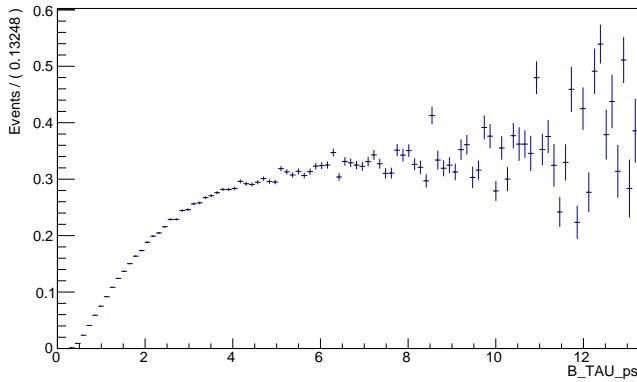


Figure 77: Histogram of the selection efficiency of the weighted combination of 2011, 2012, 2015 and 2016 MC as a function of decay time.

10.3 Mass and decay time fit

As described in Section 2.2 the fit for $\tau_{\mu^+ \mu^-}$ is performed in two stages: first an extended unbinned maximum-likelihood fit to the di-muon invariant mass spectrum is used to calculate *sWeights*, and then the effective lifetime is extracted using an unbinned maximum-likelihood fit to the decay time distribution of weighted candidates.

As described in [1] the mass spectrum potentially contains both $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ candidates, as well as a number of backgrounds. The backgrounds that contribute significantly are: candidates formed from the combination of two random tracks (combinatorial background), $B \rightarrow h^+ h^-$ decays where both hadrons are misidentified as muons, and a number of partially reconstructed decays. The partially reconstructed decays that contribute significantly in the range $4.9 - 6.0$ GeV are $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$, $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ and $\Lambda_b^0 \rightarrow p \mu^- \bar{\nu}_\mu$, which have a neutrino in the final state and a hadron misidentified as a muon, as well as $B^0 \rightarrow \pi^0 \mu^+ \mu^-$ and $B^+ \rightarrow \pi^+ \mu^+ \mu^-$ decays, which have two muons in the final state and a hadron that is not reconstructed.

For the branching fraction measurement each background is modelled using a probability distribution function (PDF) with the yields of the $B \rightarrow h^+ h^-$ and partially reconstructed backgrounds constrained to the expected yields derived from branching

2033 fraction measurements, theoretical predictions and simulated events.

2034 However, for a lifetime measurement this approach may not necessarily yield the
2035 smallest uncertainty on the measured lifetime. A series of studies were performed using a
2036 toy model to determine the optimum choice of mass fit.

2037 10.3.1 Correlation of mass and decay time

2038 The use of the di-muon invariant mass spectrum for calculating *sWeights* to fit the signal
2039 weighted decay time distribution, requires that the mass and decay time of candidate
2040 events are not correlated. The correlation between the mass and decay time of candidate
2041 events has been calculated for $B_s^0 \rightarrow \mu^+\mu^-$ signal MC and $B_s^0 \rightarrow \mu^+\mu^-$ data in right
2042 mass sideband, for 2011, 2012 and 2015. All selection requirements are applied to the
2043 $B_s^0 \rightarrow \mu^+\mu^-$ signal MC before the correlation is computed, however for the $B_s^0 \rightarrow \mu^+\mu^-$
2044 data all selection requirements except the BDT cut are applied before computing the
2045 correlation because the BDT requirement removes the majority of events. The results
2046 in Table 85 show that the correlation between mass and decay time is very small and
2047 therefore the di-muon invariant mass spectrum can be used to calculate *sWeights* to fit
the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime.

Year	$B_s^0 \rightarrow \mu^+\mu^-$ MC correlation	$B_s^0 \rightarrow \mu^+\mu^-$ data correlation
2011	0.028	-0.001
2012	0.025	-0.007
2015	0.023	-0.006

Table 85: Correlation between mass and decay time for candidate events from $B_s^0 \rightarrow \mu^+\mu^-$ signal MC and $B_s^0 \rightarrow \mu^+\mu^-$ data in right mass sideband (above di-muon invariant mass of 5447 MeV/c²) for 2011, 2012 and 2015.

2048

2049 10.3.2 Mass PDFs

2050 The mass probability distribution functions (PDFs) used to model $B_s^0 \rightarrow \mu^+\mu^-$ decays are
2051 determined in the same manner as for the branching fraction measurement. As we will
2052 see, the final mass fit only includes a $B_s^0 \rightarrow \mu^+\mu^-$ and combinatorial component, which
2053 are modelled using a Crystal Ball function [43] and exponential function respectively. All
2054 other channels are ignored and any contamination is treated as a systematic uncertainty,
2055 although PDFs for these modes are included in studies using toy pseudoexperiments (see
2056 Appendix A). Table 86 lists the values of the parameters of the $B_s^0 \rightarrow \mu^+\mu^-$ Crystal Ball
2057 function, which are fixed in the fit. The slope parameter of the combinatorial background
2058 exponential is allowed to float freely.

Decay mode	Mass PDF	Parameters
$B_s^0 \rightarrow \mu^+ \mu^-$	Crystal Ball	$m = 5372.05$ MeV $\sigma = 23.07$ MeV $\alpha = 2.668$ $n = 2.612$

Table 86: $B_s^0 \rightarrow \mu^+ \mu^-$ mass PDF parameters used in the fit. All parameters of the Crystal Ball are fixed.

10.3.3 Decay time fit

The mass fit is used to calculate sWeights for each candidate, which can then be used to extract the decay time distribution of $B_s^0 \rightarrow \mu^+ \mu^-$ signal events. However, the raw sWeights cannot be used directly in the likelihood as their normalisation properties do not produce the correct statistical uncertainty on the measured lifetime. The sWeights are therefore re-normalised according to

$$\omega'_i = \omega_i \frac{\sum_j \omega_j}{\sum_k \omega_k^2} \quad (48)$$

which gives the events the correct statistical weighting in the likelihood, as explained in [44].

10.3.4 The trouble with τ (at low statistics)

During studies with the toy model described above it was discovered that the result of the effective lifetime fit was slightly biased. This bias appeared independent of which backgrounds are generated and fitted for and of the choice of mass fit range. To better understand this effect a series of very simple toy studies were conducted where only $B_s^0 \rightarrow \mu^+ \mu^-$ decays and combinatorial background candidates were generated and fitted for. The mass range used for this study is 4900 – 6000 MeV. For these toy studies the statistical measure used to determine whether the fit returns an unbiased estimator and accurate uncertainties is the pull, defined using symmetric errors. If the fit is unbiased then the pull should follow a Gaussian distribution with mean zero and width one.

Figure 78 shows the pull distribution of $\tau_{\mu^+ \mu^-}$ for 10,000 pseudoexperiments with statistics equivalent to the CKM16 data sample. The distribution is non-Gaussian, with a tail in the negative region and a fitted width different from one. The deviation of the $\tau_{\mu^+ \mu^-}$ pull distribution from a Gaussian was much greater during the development of this analysis when a dataset smaller than the CKM16 dataset was being used. Figure 79 shows the same distribution along with those for pseudoexperiments generated with cumulative Run II and Run III statistics of 5fb^{-1} and 50 fb^{-1} respectively. As the number of events

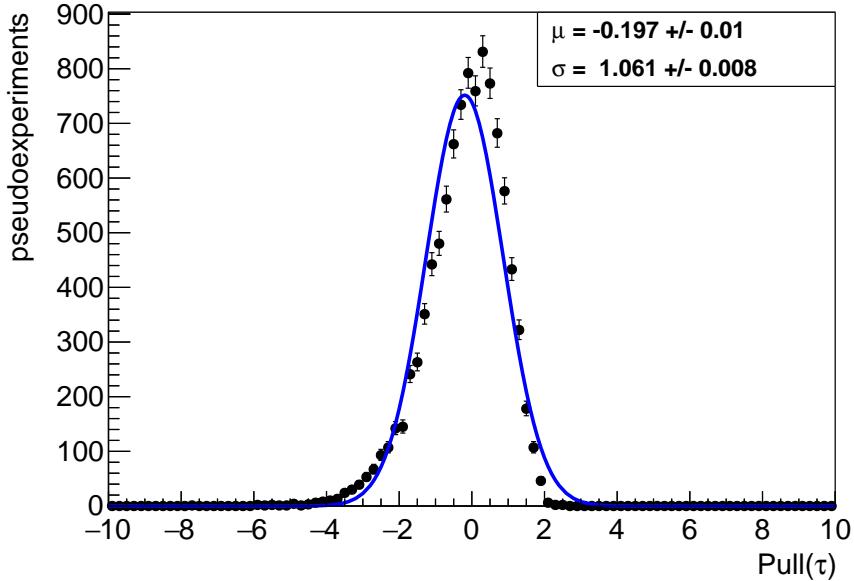


Figure 78: The pull of the effective lifetime measured for 10,000 pseudoexperiments corresponding to CKM16 statistics, generated including only $B_s^0 \rightarrow \mu^+ \mu^-$ decays and combinatorial backgrounds. The data have been fitted with a Gaussian distribution.

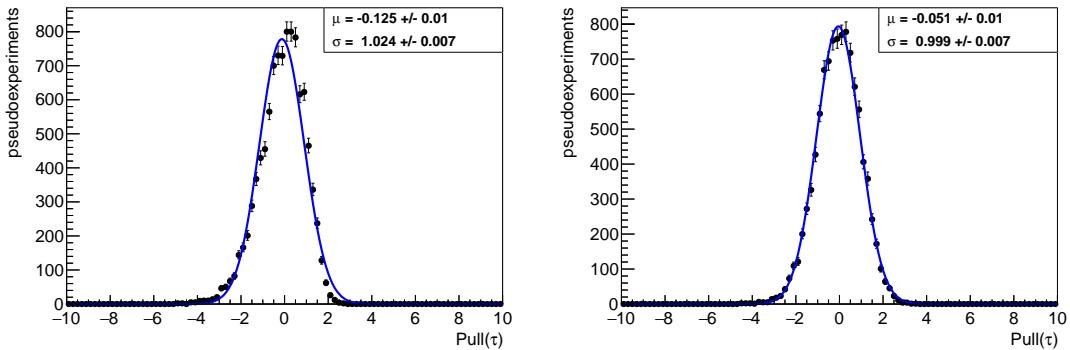


Figure 79: The pulls of the effective lifetime measured for 10,000 pseudoexperiments corresponding to Run II (left) and Run III statistics (right). Pseudoexperiments are generated including only $B_s^0 \rightarrow \mu^+ \mu^-$ decays and combinatorial backgrounds. The pull distributions have been fitted with Gaussian functions.

generated in each pseudoexperiment is increased, the bias disappears, with the fitted means and widths moving towards zero and one respectively.

Examination of the projected log-likelihood as a function of $\tau_{\mu^+ \mu^-}$ reveal the likely cause of the problem. Figure 80 shows that as $\tau_{\mu^+ \mu^-}$ approaches zero the log-likelihood diverges followed by a sharp discontinuity. With a little thought the reason for this is

2089 clear. The decay time distribution is given by:

$$N(t, \tau) = N_0 e^{-t/\tau}. \quad (49)$$

2090 As τ approaches the origin from the positive side $N(t, \tau)$ approaches zero at all values
 2091 of decay time, t . On passing through the origin $N(t, \tau)$ goes from zero to $+\infty$, and this
 2092 discontinuity is reflected in the likelihood profile. In the case of small statistics (as in
 2093 Run I) the fitted value of $\tau_{\mu^+\mu^-}$ is within a few standard deviations (in the case of this
 2094 simple toy study around 0.7 ps) of this discontinuity at $\tau = 0$, which creates a bias in the
 2095 estimate of the uncertainty on $\tau_{\mu^+\mu^-}$.

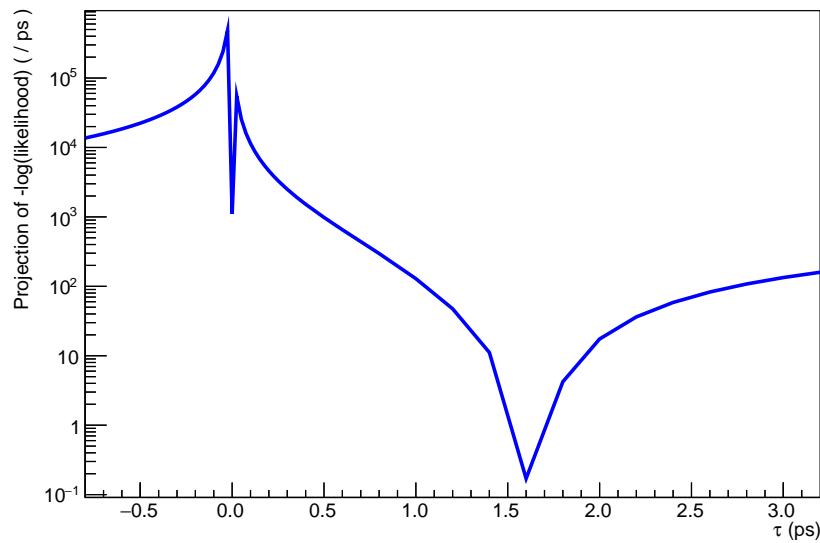


Figure 80: The projected negative log-likelihood of the lifetime fit as a function of $\tau_{\mu^+\mu^-}$.

2096 Therefore it is not clear from the pull distributions alone that the uncertainties returned
 2097 by a fit for $\tau_{\mu^+\mu^-}$ at low statistics can be interpreted in the usual way. One solution to
 2098 this problem is to fit for $\tau_{\mu^+\mu^-}^{-1}$, whose log-likelihood profile is well behaved at all values
 2099 as shown by Figure 81. The same toy studies as performed for $\tau_{\mu^+\mu^-}$ show that $\tau_{\mu^+\mu^-}^{-1}$ is
 2100 unbiased and has correctly estimated uncertainties regardless of statistics, as shown in
 2101 Figure 82.

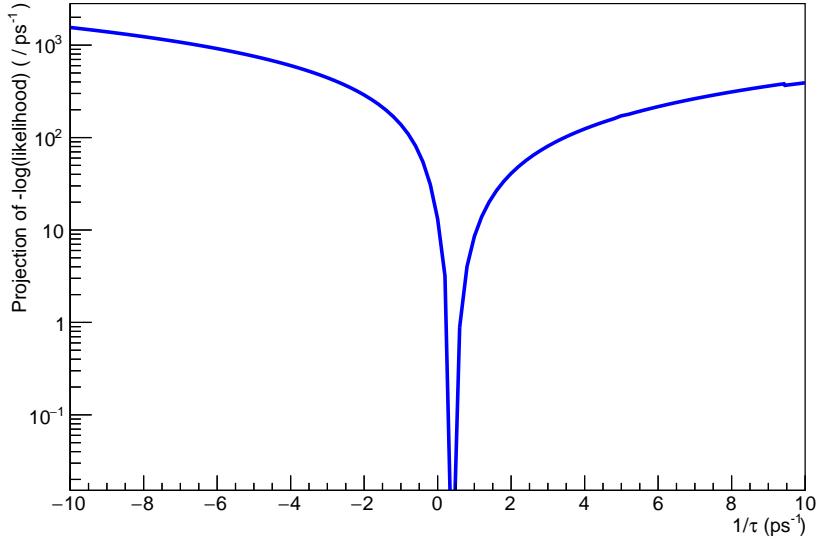


Figure 81: The projected negative log-likelihood of the lifetime fit as a function of $\tau_{\mu^+\mu^-}^{-1}$.

2102 Although the pull distribution for fitting for $\tau_{\mu^+\mu^-}$ is worse than fitting for $\tau_{\mu^+\mu^-}^{-1}$, the
 2103 coverage of the statistical uncertainties returned by the fits for $\tau_{\mu^+\mu^-}$ and $\tau_{\mu^+\mu^-}^{-1}$ offers
 2104 an alternative test to check whether the uncertainties returned by a fit for $\tau_{\mu^+\mu^-}$ can be
 2105 interpreted in the usual way. The results are presented in Table 87. While not perfect,
 2106 the coverage for $\tau_{\mu^+\mu^-}$ is actually reasonably good and the bias in the pull distribution
 2107 when fitting for $\tau_{\mu^+\mu^-}$ in Fig. 78 is small compared to the statistical uncertainty expected
 2108 for this measurement which is discussed in Section 10.3.5. Therefore with the expected
 2109 statistics for the CKM16 dataset fitting for either $\tau_{\mu^+\mu^-}$ or $\tau_{\mu^+\mu^-}^{-1}$ would be reasonable.
 2110 For the following sections of the note, fitting for both $\tau_{\mu^+\mu^-}$ and $\tau_{\mu^+\mu^-}^{-1}$ shall be considered
 2111 and this issue will be discussed further in Section 10.4.1.

	$\tau_{\mu^+\mu^-}$	$\tau_{\mu^+\mu^-}^{-1}$	Gaussian
1σ	68.52%	68.67%	68.27%
2σ	93.82%	95.29%	95.45 %
3σ	98.54%	99.66%	99.73 %

Table 87: The percentage of fitted values of $\tau_{\mu^+\mu^-}$ and $\tau_{\mu^+\mu^-}^{-1}$ falling within 1, 2 and 3 standard deviations of the input value calculated from 10,000 pseudoexperiments at CKM16 statistics. The intervals for a Gaussian distribution are listed for comparison.

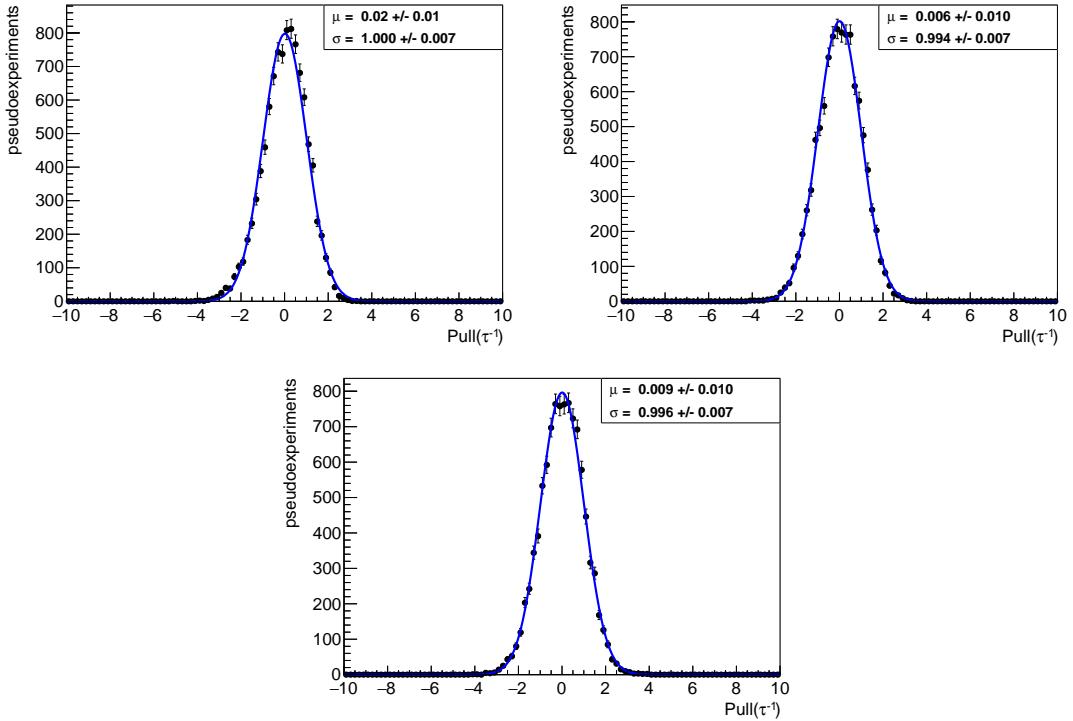


Figure 82: The pulls of the inverse effective lifetime measured for 10,000 pseudoexperiments corresponding to Run I (top-left), Run II (top-right) and Run III statistics (bottom). Pseudoexperiments are generated including only $B_s^0 \rightarrow \mu^+ \mu^-$ decays and combinatorial backgrounds. The pull distributions have been fitted with Gaussian functions.

10.3.5 Optimising the fit

The toy model described in Appendix A was used to determine the optimum configuration of the mass fit. A range of options were investigated, varying the mass range of the fit and the components included in the likelihood.

Looking at Figure 102 we can see that the only peaking backgrounds that contribute significantly in the $B_s^0 \rightarrow \mu^+ \mu^-$ signal region are $B^0 \rightarrow \mu^+ \mu^-$, $B \rightarrow h^+ h'^-$ and $\Lambda_b^0 \rightarrow p \mu^- \bar{\nu}_\mu$ decays, and this informs the fitting options investigated. A full list of configurations investigated using the toy model are listed in Table 88.

In total, 1000 pseudoexperiments were conducted for each fit configuration with statistics corresponding to the CKM16 data sample. The metrics used to determine the choice of fit are the means and widths of the pulls of the free parameters of the fit (the $B_s^0 \rightarrow \mu^+ \mu^-$ and combinatorial background yields, the slope of the combinatorial background mass PDF and the inverse lifetime) and the median uncertainties on $\tau_{\mu^+ \mu^-}^{-1}$ and $\tau_{\mu^+ \mu^-}$. For these studies the $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime is set to τ_H , i.e. the SM value. The results of these studies are listed in Table 89.

Fit 9, which has a cut of $m_{\mu^+ \mu^-} > 5320$ MeV and fits only for $B_s^0 \rightarrow \mu^+ \mu^-$ decays and

	Mass range / GeV	Individual peaking background components fitted for	Yields fixed?
Fit 1	4.90 – 6.00	$B^0 \rightarrow \mu^+ \mu^-$ $B \rightarrow h^+ h'^-, B \rightarrow h \mu^+ \nu_\mu, B \rightarrow \pi \mu^+ \mu^-, \Lambda_b^0 \rightarrow p \mu^- \bar{\nu}_\mu$	No Yes
Fit 2	5.15 – 6.00	$B^0 \rightarrow \mu^+ \mu^-$ $B \rightarrow h^+ h'^-, B \rightarrow h \mu^+ \nu_\mu, B \rightarrow \pi \mu^+ \mu^-, \Lambda_b^0 \rightarrow p \mu^- \bar{\nu}_\mu$	No Yes
Fit 3	5.20 – 6.00	$B^0 \rightarrow \mu^+ \mu^-$ $B \rightarrow h^+ h'^-, B \rightarrow h \mu^+ \nu_\mu, B \rightarrow \pi \mu^+ \mu^-, \Lambda_b^0 \rightarrow p \mu^- \bar{\nu}_\mu$	No Yes
Fit 4	5.20 – 6.00	$B^0 \rightarrow \mu^+ \mu^-$ $B \rightarrow h^+ h'^-, \Lambda_b^0 \rightarrow p \mu^- \bar{\nu}_\mu$	No Yes
Fit 5	5.20 – 6.00	$B^0 \rightarrow \mu^+ \mu^-$	No
Fit 6	5.20 – 6.00	None	-
Fit 7	5.25 – 6.00	None	-
Fit 8	5.30 – 6.00	None	-
Fit 9	5.32 – 6.00	None	-

Table 88: Mass fit configurations investigated. All options include $B_s^0 \rightarrow \mu^+ \mu^-$ and combinatorial background components with yields that are allowed to float freely.

2128 combinatorial background gives the lowest uncertainties on $\tau_{\mu^+ \mu^-}^{-1}$ and $\tau_{\mu^+ \mu^-}$ as well as the
 2129 most accurate pulls and is therefore selected as the fit to be used for the measurement.
 2130 Figure 83 shows the distribution of uncertainties for $\tau_{\mu^+ \mu^-}$ and $\tau_{\mu^+ \mu^-}^{-1}$ for this choice of
 2131 fit. It should be noted that while the studies return expected median uncertainties of
 2132 $\sigma(\tau_{\mu^+ \mu^-}) = 0.28 \text{ ps}$ and $\sigma(\tau_{\mu^+ \mu^-}^{-1}) = 0.11 \text{ ps}^{-1}$ there is a wide spread in uncertainties
 2133 and the actual fit to data may return uncertainties anywhere in the range $0.1 - 0.8 \text{ ps}$ or
 2134 $0.07 - 0.2 \text{ ps}^{-1}$. Some example mass and decay time fits from a single pseudoexperiment
 2135 using Fit 9 settings are shown in Figure 84.

Fit	$N(B_s^0 \rightarrow \mu^+ \mu^-)$		N(C.Bkg.)		$\lambda(C.Bkg.)$		$\tau_{\mu^+ \mu^-}$	$\tau_{\mu^+ \mu^-}^{-1}$		
	mean	width	mean	width	mean	width	σ / ps	mean	width	σ / ps^{-1}
1	-0.06	1.02	-0.16	1.10	0.00	1.00	0.38	0.04	0.94	0.14
2	-0.08	1.02	-0.12	1.06	0.06	0.99	0.39	0.07	0.93	0.15
3	-0.11	1.02	-0.11	1.03	0.01	1.01	0.39	0.06	0.94	0.15
4	-0.15	1.02	0.04	1.02	-0.12	1.04	0.33	-0.06	1.04	0.12
5	-0.31	1.03	0.86	0.98	-1.14	1.18	0.32	-0.07	0.98	0.11
6	-0.57	1.06	1.77	0.91	-2.35	1.10	0.33	-0.14	0.98	0.12
7	-0.41	1.07	1.24	0.95	-1.82	1.15	0.32	-0.12	1.03	0.12
8	-0.07	1.04	0.35	1.00	-0.43	1.09	0.30	-0.08	0.98	0.11
9	0.04	1.01	0.07	1.03	0.09	1.02	0.28	0.02	1.00	0.11

Table 89: Results of the studies of the various fit configurations. The means and widths of the pull distributions are determined by fitting Gaussian functions to the pull distributions. The uncertainties on the pull means and widths are 0.03 and 0.02 respectively in all cases, except for Fit 9 where they are both 0.01.

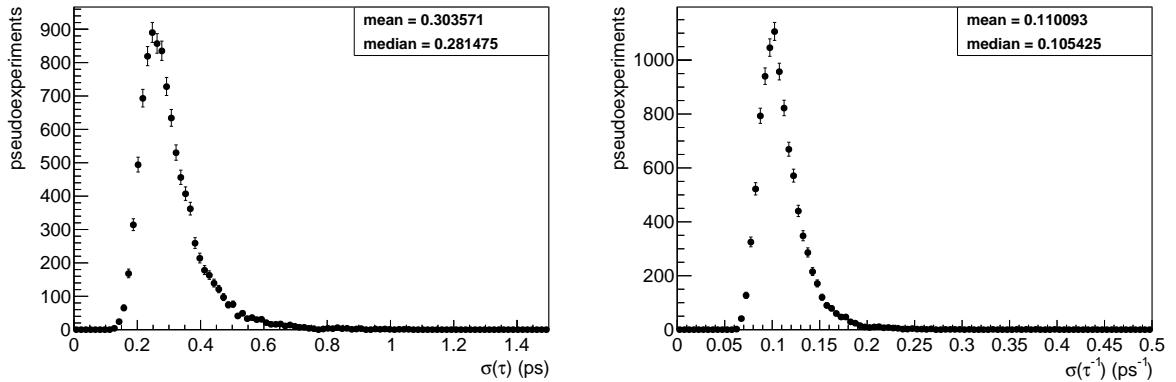


Figure 83: The distribution of the statistical uncertainties on the $B_s^0 \rightarrow \mu^+ \mu^-$ lifetime (left) and inverse lifetime (right) for Fit 9, from 10,000 pseudoexperiments.

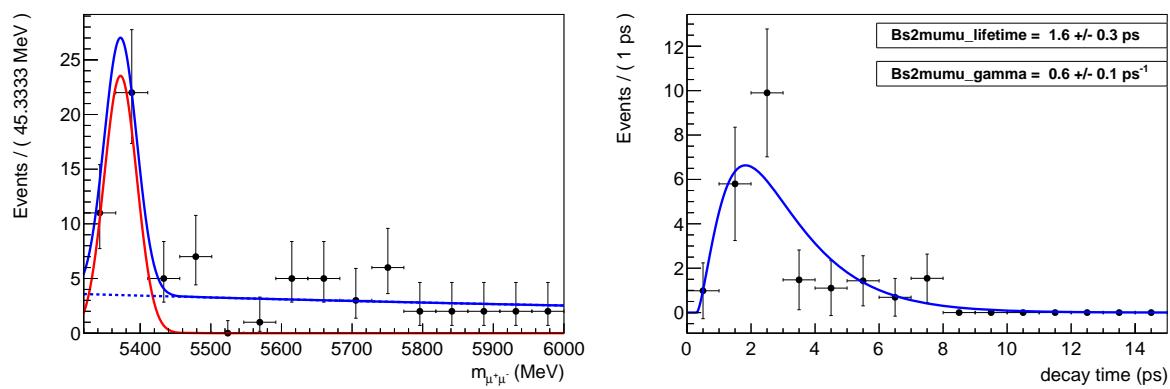


Figure 84: Example of mass (left) and decay time (right) fits for the inverse $B_s^0 \rightarrow \mu^+ \mu^-$ lifetime for Fit 9, from a single pseudo-experiment.

2136 **10.4 Systematic uncertainties**

2137 **10.4.1 Fit accuracy**

2138 A number of checks are performed to ensure that the fit returns an unbiased estimate of the
 2139 (inverse) lifetime. The value of the effective lifetime used to generate the events is varied
 2140 in order to check that the pull does not look unbiased simply because the signal events
 2141 have a similar decay time distribution to the background. Batches of pseudoexperiments
 2142 are generated with the lifetime set to the lifetimes of the light and heavy eigenstates of the
 2143 B_s^0 meson and the mean lifetime of the B_s^0 meson. Only combinatorial background and
 2144 $B_s^0 \rightarrow \mu^+ \mu^-$ candidates are generated so that any effects are not masked by interference
 2145 from exclusive backgrounds, which are considered in more detail in Section 10.4.2. The
 2146 results are listed in Table 90. The mean and width of the pull distributions for each input
 2147 lifetime are consistent with 0 and 1 respectively, demonstrating that the fit returns an
 2148 accurate estimate of the inverse lifetime and its uncertainty, independent of the value of
 2149 the inverse lifetime.

Input	$N(B_s^0 \rightarrow \mu^+ \mu^-)$		$N(\text{C.Bkg.})$		$\lambda(\text{C.Bkg.})$		$\tau_{\mu^+ \mu^-}$	$\tau_{\mu^+ \mu^-}^{-1}$		
	mean	width	mean	width	mean	width	σ / ps	mean	width	σ / ps^{-1}
τ_{B_s}	-0.09	1.00	-0.07	1.03	-0.01	1.00	0.26	-0.02	1.01	0.11
τ_L	-0.09	1.01	-0.07	1.03	-0.01	0.98	0.24	-0.01	1.01	0.12
τ_H	-0.08	1.01	-0.08	1.03	0.00	1.00	0.29	0.00	1.01	0.11

Table 90: Results of studies using Fit 9 (with $m_{\mu^+ \mu^-} > 5320$ MeV) and varying the lifetime used to generate $B_s^0 \rightarrow \mu^+ \mu^-$ events. The means and widths of the pull distributions are determined by fitting Gaussian functions to the pull distributions. The uncertainties on the pull means and widths are approximately 0.01 in all cases. 10,000 pseudoexperiments are generated for each input lifetime, and only $B_s^0 \rightarrow \mu^+ \mu^-$ and combinatorial background candidates are generated.

2150 However, we see that for all input lifetimes the means of the pulls of the $B_s^0 \rightarrow \mu^+ \mu^-$
 2151 yields are biased. At first glance this result would suggest that the mass fit does not
 2152 accurately model the mass distribution, however, the fact that the fit returns the correct
 2153 value of the inverse lifetime contradicts this assumption.

2154 The value for N_{true} is fixed for each decay mode / background, however, the actual
 2155 number of events (N_{actual}) that are generated in each pseudoexperiment fluctuates according
 2156 to a Poisson distribution. This is done since an extended maximum-likelihood fit is
 2157 performed with the total number of events included as a free-parameter of the fit and as
 2158 such the number of events generated must be allowed to fluctuate. At high statistics the
 2159 Poisson distribution of N_{actual} becomes a good approximation to a Gaussian distribution,
 2160 but this is a poor approximation at low statistics. Since the uncertainty on N_{meas} is
 2161 proportional to $\sqrt{N_{actual}}$ this means that the uncertainties are also not Gaussian-distributed

at low statistics. Consequently the mean of the pull distribution is shifted from zero. This does not indicate a problem with the estimate of the yields of each mode, but with the interpretation of the uncertainties on the yields as being Gaussian-distributed.

If this interpretation is correct, then the means of the pull distributions of the yield variables should approach zero as the number of events generated in each pseudoexperiment is increased. Figure 85 shows the pull distributions of the fitted yields of $B_s^0 \rightarrow \mu^+ \mu^-$ decays at Run I, Run II and Run III statistics, and clearly demonstrates that as statistics are increased the pull mean moves towards zero. Figure 86 shows the ‘fractional bias’, defined as $(N_{\text{meas}} - N_{\text{actual}}) / N_{\text{actual}}$, of $B_s^0 \rightarrow \mu^+ \mu^-$ decays and combinatorial background at CKM16 statistics. The mean bias is around 0.5% of the yield, demonstrating that the fit does indeed return an accurate estimate of the yields of each mode, despite the shift in the means of the pull distributions.

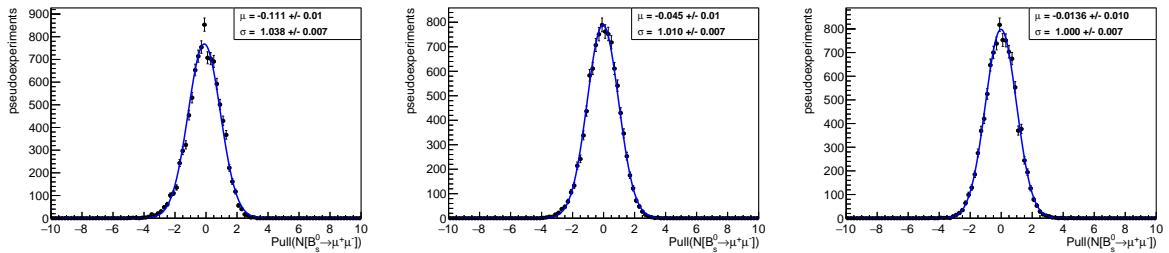


Figure 85: The pulls of the yields of $B_s^0 \rightarrow \mu^+ \mu^-$ decays for 10,000 pseudoexperiments corresponding to Run I + 2015 statistics (left), full Run II statistics (centre) and Run III statistics (right). Pseudoexperiments are generated including only $B_s^0 \rightarrow \mu^+ \mu^-$ decays and combinatorial backgrounds. The pull distributions have been fitted with Gaussian functions.

These checks having been made, it remains to evaluate the size of any underlying systematic bias in the fit for the inverse lifetime. Figure 87 shows the pull distributions of the lifetime and the inverse lifetime for CKM16 statistics. These distributions were calculated for 10,000 pseudoexperiments using the default fit but without any peaking backgrounds. For the inverse lifetime the pull mean is consistent with zero and the width is consistent with one. Any residual bias is below 1% of the statistical uncertainty and is therefore negligible.

For the effective lifetime itself the mean of the pull and width deviate significantly from zero and one, which is expected from the discussion in Section 10.3.4. This raises a question as to which quantity should be presented as the final result, and how the results should be presented. The inverse lifetime is clearly well behaved statistically and therefore more straightforward to present, however the effective lifetime itself is of more interest from a physics perspective.

To aid interpretation the coverage of the measurement uncertainty of both $\tau_{\mu^+ \mu^-}$ and $\tau_{\mu^+ \mu^-}^{-1}$ has been checked using the same toy studies. The results are presented in Table 91 and it seems that although the pulls for the lifetime are asymmetric the coverage of

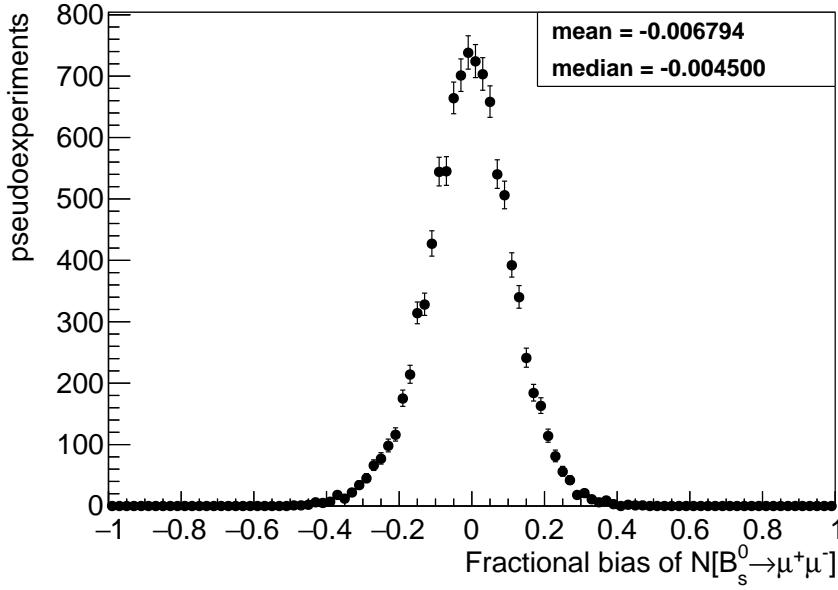


Figure 86: The fractional bias of the yields of $B_s^0 \rightarrow \mu^+ \mu^-$ decays for 10,000 pseudoexperiments corresponding to CKM16 statistics. Pseudoexperiments are generated including only $B_s^0 \rightarrow \mu^+ \mu^-$ decays and combinatorial background.

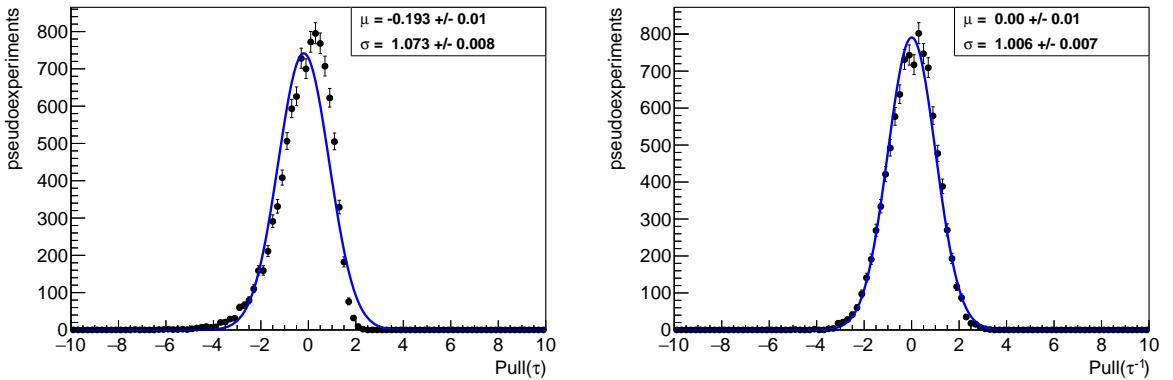


Figure 87: The pulls of the effective lifetime (left) and the inverse effective lifetime (right) measured for 10,000 pseudoexperiments corresponding to CKM16 statistics, generated including only $B_s^0 \rightarrow \mu^+ \mu^-$ decays and combinatorial backgrounds. The data have been fitted with Gaussian distributions.

the uncertainties is reasonably good, particularly at the one sigma level, therefore for the expected statistics in the CKM16 dataset it is reasonable to fit for $\tau_{\mu^+ \mu^-}$. Figure 88 shows the difference between the measured and true effective lifetime for 10,000 pseudoexperiments for our expected statistics. The mean of the distribution is 0.02 ps, which would be taken as the systematic uncertainty in the event that the lifetime is presented as the result rather

2195 than the inverse lifetime.

2196 The observed number of events detailed in Section 10.5 differs from the expected
 2197 number of events for the CKM16 dataset. The systematic uncertainty for the fit accuracy
 2198 has been re-evaluated for the observed number of signal and background events and is
 now 0.03 ps.

	$\tau_{\mu^+\mu^-}$	$\tau_{\mu^+\mu^-}^{-1}$	Gaussian
1σ	68.29%	67.39%	68.27%
2σ	93.65%	95.49%	95.45 %
3σ	98.50%	99.67%	99.73 %

Table 91: The percentage of fitted values of $\tau_{\mu^+\mu^-}$ and $\tau_{\mu^+\mu^-}^{-1}$ falling within 1, 2 and 3 standard deviations of the input value calculated from 10,000 pseudoexperiments at CKM16 statistics, in the mass range 5320-6000 MeV. The intervals for a Gaussian distribution are listed for comparison.

2199

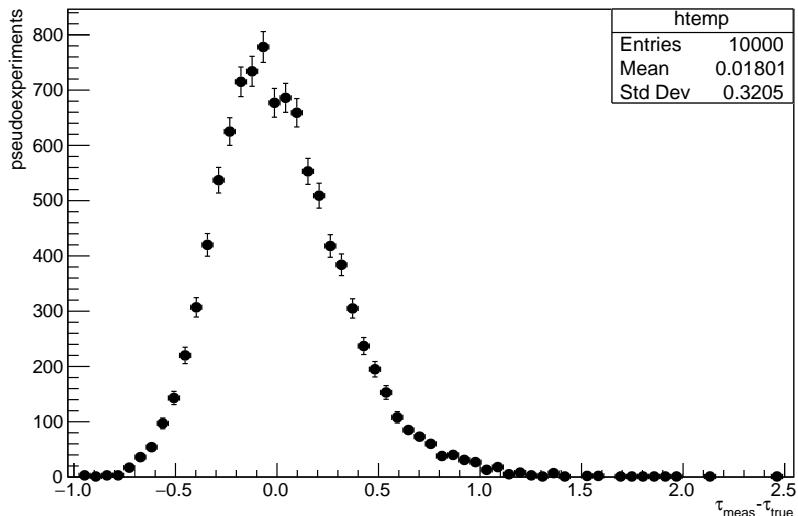


Figure 88: The difference between the fitted value and true value of the effective lifetime, measured for 10,000 pseudoexperiments corresponding to CKM16 statistics, generated including only $B_s^0 \rightarrow \mu^+\mu^-$ decays and combinatorial backgrounds, in the mass range 5320-6000 MeV.

2200 10.4.2 Contamination from exclusive backgrounds

2201 Since the fit only includes PDFs for $B_s^0 \rightarrow \mu^+\mu^-$ decays and combinatorial background the
 2202 presence of additional exclusive backgrounds in the mass fit range may introduce a bias

to the measured lifetime. Exclusive backgrounds in between 4900 and 6000 MeV include $B^0 \rightarrow \mu^+ \mu^-$, $B \rightarrow h^+ h^-$, $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$, $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$, $B^0 \rightarrow \pi^0 \mu^+ \mu^-$, $B^+ \rightarrow \pi^+ \mu^+ \mu^-$, $\Lambda_b^0 \rightarrow p \mu^- \bar{\nu}_\mu$ and $B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu$ decays. As discussed in Section 10.3 the mass of any B_s^0 meson candidate is required to be greater than 5320 MeV and as can be seen in Figure 102 this has the effect of removing almost all of the exclusive backgrounds from the data sample.

The quantity of any contamination after this mass cut is estimated by integrating the exclusive background mass PDFs over the fit range 5320-6000 MeV normalised to an integral over the full mass range 4900-6000 MeV. The estimated remaining contamination is given in Table 92.

Decay mode	Number of candidates 4900-6000 MeV	Number of candidates 5320-6000 MeV
$B^0 \rightarrow \mu^+ \mu^-$	3.3 (SM) or 12.2 (WA)	0.2 (SM) or 0.7 (WA)
$B \rightarrow h^+ h^-$	9.7	0.9
$B^0 \rightarrow \pi^- \mu^+ \nu_\mu$	40.5	0.1
$B_s^0 \rightarrow K^- \mu^+ \nu_\mu$	9.1	0.0
$B^0 \rightarrow \pi^0 \mu^+ \mu^-$	4.9	0.0
$B^+ \rightarrow \pi^+ \mu^+ \mu^-$	6.0	0.0
$\Lambda_b^0 \rightarrow p \mu^- \bar{\nu}_\mu$	13.3	0.6
$B_c^+ \rightarrow J/\psi \mu^+ \nu_\mu$	9.7	0.0
$B_s^0 \rightarrow \mu^+ \mu^-$	30.9 (SM) or 22.8 (WA)	30.5 (SM) or 22.5 (WA)
Combinatorial background	66.2	40.6

Table 92: The estimated yields of each exclusive background in the CKM16 data sample after the full selection and in the full mass range and the fit mass range (5320-6000 MeV). For $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ decays, yields assuming the Standard Model (SM) and the world average (WA) branching fractions are presented.

To estimate the effect of these contaminations on the measured (inverse) lifetime, pseudoexperiments are produced with exclusive backgrounds generated according to the numbers in Table 92 with the yields fluctuated according to a Poisson distribution (the decays with zero yields are fluctuated around their small but non-zero values below the first decimal point). Two sets of 10,000 pseudoexperiments are produced, one with the $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ branching fractions set to their Standard Model values and the other set to the current world average values.

The resulting pull distributions can be found in Figure 89. The largest shift in the pull mean is introduced in the case of SM branching fractions, with the mean shifted from zero to 0.025. Assuming the median uncertainty on the inverse lifetime, this shift corresponds to a systematic shift due to neglecting the peaking backgrounds in the fit of $\pm 0.003 \text{ ps}^{-1}$ in a fit for $\tau_{\mu^+ \mu^-}^{-1}$ and 0.007 ps in a fit for $\tau_{\mu^+ \mu^-}$.

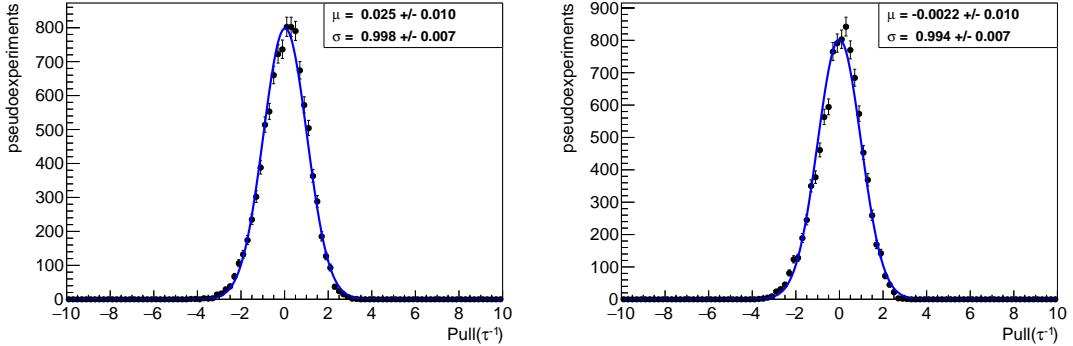


Figure 89: The pulls of the inverse effective lifetime measured for 10,000 pseudoexperiments corresponding to CKM16 statistics, generated including all exclusive backgrounds. The Standard Model branching fractions for $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ are used in the left plot, the world average branching fractions in the right plot. The data have been fitted with Gaussian distributions.

10.4.3 Effect of the production asymmetry

As discussed in the introduction, the effective lifetime is defined as the mean decay time of an unbiased sample of $B_s^0 \rightarrow \mu^+ \mu^-$ decays;

$$\tau_{\mu^+ \mu^-} \equiv \frac{\int_0^\infty t \langle \Gamma(B_s^0(t) \rightarrow \mu^+ \mu^-) \rangle dt}{\int_0^\infty \langle \Gamma(B_s^0(t) \rightarrow \mu^+ \mu^-) \rangle dt} \quad (50)$$

where the untagged decay rate is

$$\Gamma(B_s(t) \rightarrow \mu^+ \mu^-) = \Gamma(B_s^0(t) \rightarrow \mu^+ \mu^-) + \Gamma(\bar{B}_s^0(t) \rightarrow \mu^+ \mu^-). \quad (51)$$

This definition assumes that B_s^0 and \bar{B}_s^0 mesons are produced in equal quantities. However, since the LHC is a proton-proton collider the production of the production asymmetry of B_s mesons, defined as

$$A_p \equiv \frac{\sigma(B_s^0) - \sigma(\bar{B}_s^0)}{\sigma(B_s^0) + \sigma(\bar{B}_s^0)} \quad (52)$$

is non-zero, where $\sigma(B_s^0)$ and $\sigma(\bar{B}_s^0)$ are the production cross-sections for B_s^0 and \bar{B}_s^0 mesons. The production asymmetry for B_s^0 mesons has been measured at LHCb at a centre of mass energy of 7 TeV as $A_p = (1.09 \pm 2.61 \pm 0.66)\%$ [45].

The effect of A_p is calculated as follows. Following the notation used in [4] the

²²³⁶ flavour-specific decay rates are

$$\begin{aligned}\Gamma(B_s^0(t) \rightarrow \mu^+ \mu^-) &= \frac{1}{2} N_{\mu^+ \mu^-} |A_{\mu^+ \mu^-}|^2 (1 + |\xi_\lambda|^2) e^{-\Gamma_s t} \left\{ \cosh \frac{\Delta \Gamma_s t}{2} + C_\lambda \cos(\Delta m_s t) \right. \\ &\quad \left. + \mathcal{A}_{\Delta \Gamma}^{\mu^+ \mu^-} \sinh \frac{\Delta \Gamma_s t}{2} + S_\lambda \sin(\Delta m_s t) \right\},\end{aligned}\quad (53)$$

$$\begin{aligned}\Gamma(\bar{B}_s^0(t) \rightarrow \mu^+ \mu^-) &= \frac{1}{2} N_{\mu^+ \mu^-} |A_{\mu^+ \mu^-}|^2 (1 + a)(1 + |\xi_\lambda|^2) e^{-\Gamma_s t} \left\{ \cosh \frac{\Delta \Gamma_s t}{2} - C_\lambda \cos(\Delta m_s t) \right. \\ &\quad \left. + \mathcal{A}_{\Delta \Gamma}^{\mu^+ \mu^-} \sinh \frac{\Delta \Gamma_s t}{2} - S_\lambda \sin(\Delta m_s t) \right\}\end{aligned}\quad (54)$$

²²³⁷ so the total decay rate including the effect of production asymmetry is then

$$\begin{aligned}\Gamma(B_s(t) \rightarrow \mu^+ \mu^-) &= \left(\frac{1 + A_p}{2} \right) \Gamma(B_s^0(t) \rightarrow \mu^+ \mu^-) + \left(\frac{1 - A_p}{2} \right) \Gamma(\bar{B}_s^0(t) \rightarrow \mu^+ \mu^-) \\ &= \frac{1}{2} N_{\mu^+ \mu^-} |A_{\mu^+ \mu^-}|^2 (1 + |\xi_\lambda|^2) e^{-\Gamma_s t} \left\{ \cosh \frac{\Delta \Gamma_s t}{2} + \mathcal{A}_{\Delta \Gamma}^{\mu^+ \mu^-} \sinh \frac{\Delta \Gamma_s t}{2} \right. \\ &\quad \left. + A_p [C_\lambda \cos(\Delta m_s t) + S_\lambda \sin(\Delta m_s t)] \right\} + \mathcal{O}(a).\end{aligned}\quad (55)$$

²²³⁸ Using $\Delta \Gamma_s = \Gamma_L - \Gamma_H$ and $\Gamma_s = (\Gamma_L + \Gamma_H)/2$ and neglecting terms of order a , this becomes

$$\begin{aligned}\Gamma(B_s(t) \rightarrow \mu^+ \mu^-) &\simeq \frac{1}{4} N_{\mu^+ \mu^-} |A_{\mu^+ \mu^-}|^2 (1 + |\xi_\lambda|^2) \left\{ (1 - \mathcal{A}_{\Delta \Gamma}^{\mu^+ \mu^-}) e^{-\Gamma_L t} + (1 + \mathcal{A}_{\Delta \Gamma}^{\mu^+ \mu^-}) e^{-\Gamma_H t} \right. \\ &\quad \left. + 2A_p e^{-\Gamma_s t} [C_\lambda \cos(\Delta m_s t) + S_\lambda \sin(\Delta m_s t)] \right\}.\end{aligned}\quad (56)$$

²²³⁹ Inspecting Equation 56, the production asymmetry introduces an extra oscillatory term and
²²⁴⁰ by letting $A_p = 0$ the usual double exponential form of the proper decay time distribution
²²⁴¹ is recovered.

²²⁴² Using the results from integration by parts

$$\int_0^\infty e^{-\Gamma_s t} \cos(\Delta m_s t) dt = \frac{\Gamma_s}{\Delta m_s^2 + \Gamma_s^2} \quad (57)$$

$$\int_0^\infty e^{-\Gamma_s t} \sin(\Delta m_s t) dt = \frac{\Delta m_s}{\Delta m_s^2 + \Gamma_s^2} \quad (58)$$

$$\int_0^\infty t \cdot e^{-\Gamma_s t} dt = \frac{1}{\Gamma_s^2} \quad (59)$$

$$\int_0^\infty t \cdot e^{-\Gamma_s t} \cos(\Delta m_s t) dt = \frac{\Gamma_s^2 - \Delta m_s^2}{(\Delta m_s^2 + \Gamma_s^2)^2} \quad (60)$$

$$\int_0^\infty t \cdot e^{-\Gamma_s t} \sin(\Delta m_s t) dt = \frac{2\Gamma_s \Delta m_s}{(\Delta m_s^2 + \Gamma_s^2)^2} \quad (61)$$

and defining $N' \equiv \frac{1}{4}N_{\mu^+\mu^-}|A_{\mu^+\mu^-}|^2(1 + |\xi_\lambda|^2)$ it can be shown that

$$\int_0^\infty t \cdot \Gamma(B_s(t) \rightarrow \mu^+\mu^-) dt = N' \left\{ \frac{1 - \mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-}}{\Gamma_L^2} + \frac{1 + \mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-}}{\Gamma_H^2} + 2A_p \left[C_\lambda \frac{\Gamma_s^2 - \Delta m_s^2}{(\Delta m_s^2 + \Gamma_s^2)^2} + S_\lambda \frac{2\Gamma_s \Delta m_s}{(\Delta m_s^2 + \Gamma_s^2)^2} \right] \right\} \quad (62)$$

and

$$\int_0^\infty \Gamma(B_s(t) \rightarrow \mu^+\mu^-) dt = N' \left\{ \frac{1 - \mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-}}{\Gamma_L} + \frac{1 + \mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-}}{\Gamma_H} + 2A_p \left[C_\lambda \frac{\Gamma_s}{\Delta m_s^2 + \Gamma_s^2} + S_\lambda \frac{\Delta m_s}{\Delta m_s^2 + \Gamma_s^2} \right] \right\}. \quad (63)$$

The effective lifetime in the presence of a production asymmetry may now be calculated. In order to obtain an upper limit on the size of any such effect we choose $A_p = 0.040$ (1σ larger the measured value at LHCb) and the extreme value $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} = 0.0$. Since $(\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-})^2 + (C_\lambda)^2 + (S_\lambda)^2 = 1$ this choice requires that $(C_\lambda)^2 + (S_\lambda)^2 = 1$. It is assumed that $C_\lambda = S_\lambda = \sqrt{0.5}$. Finally the PDG [19] values of $\Gamma_L = 0.703 \text{ ps}^{-1}$, $\Gamma_H = 0.621 \text{ ps}^{-1}$, $\Gamma_s = 0.662 \text{ ps}^{-1}$ and $\Delta m_s = 17.757 \text{ ps}^{-1}$ are used.

The effective lifetime in the presence of a production asymmetry is found to be 1.520 ps. Setting $A_p = 0$ this becomes 1.522 ps. The difference is 0.002 ps, corresponding to a systematic shift in the lifetime of 0.002 ps and in the inverse lifetime of 0.001 ps^{-1} , this is assigned as the systematic uncertainty.

10.4.4 The effect of decay time acceptance on the mix of light and heavy mass eigenstates

In the SM, the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime is that of the heavy mass eigenstate with $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} = 1.0$. However, in the presence of new physics there may also be a contribution from the light mass eigenstate:

$$\Gamma(B_s(t) \rightarrow \mu^+\mu^-) = N_{\mu^+\mu^-} \frac{|A_{\mu^+\mu^-}|^2}{2} (1 + |\xi_\lambda|^2) \left[(1 - \mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-}) e^{-\Gamma_L t} + (1 + \mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-}) e^{-\Gamma_H t} \right]. \quad (64)$$

Since the trigger and offline selections tend to reject short-lived particles, the light mass component efficiency (to be triggered and selected) is less than that of the heavy mass component, potentially disguising evidence of new physics. In order to estimate the size of such an effect, a check has been performed using a toy Monte Carlo simulation where the $B_s^0 \rightarrow \mu^+\mu^-$ decay mode is modelled with $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} = 0$ so that the decay is described by an equal mixture of the light and heavy eigenstates. The acceptance function has the same form and parameter values as the one used in the analysis. A million events are generated with no acceptance and fitted with a single exponential function, followed by a further million including an acceptance function, and fitted with a single exponential multiplied

2269 by the acceptance function. The acceptance parameters for the $B_s^0 \rightarrow \mu^+ \mu^-$ acceptance
 2270 from MC given in Table 83 are used. The results are shown in Figure 90 and the shift
 2271 in the measured values of $\tau_{\mu^+ \mu^-}^{-1}$ is 0.007 ps⁻¹ and the shift in $\tau_{\mu^+ \mu^-}$ is 0.018 ps, which is
 2272 conservatively assigned as the systematic uncertainty.

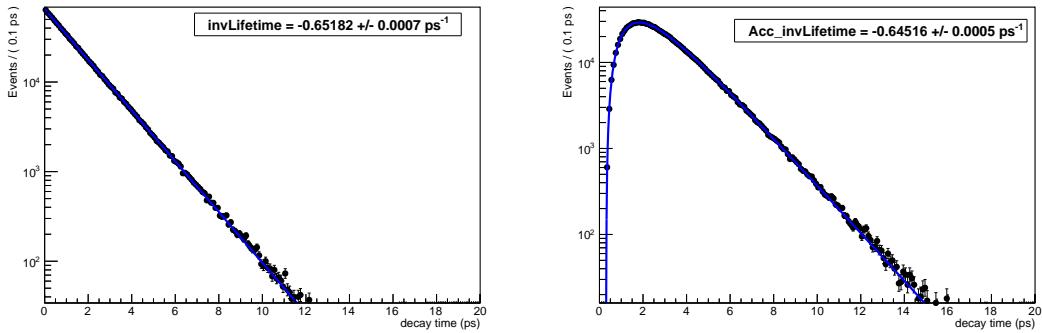


Figure 90: Decay time distribution of one million $B_s^0 \rightarrow \mu^+ \mu^-$ candidates generated with $\mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-} = 0.0$, before (left) and after (right) the effects of decay time acceptance for parameters in Table 83. The left plot has been fitted with a single exponential function, the right plot with a single exponential function multiplied by the acceptance. The fitted value of the inverse lifetime ($\tau_{\mu^+ \mu^-}^{-1}$) multiplied by -1 is displayed on each plot.

2273 10.4.5 Mass model

2274 Toy studies were used to check the sensitivity of the result to the PDF used to describe
 2275 the mass distribution of $B_s^0 \rightarrow \mu^+ \mu^-$ decays. Events were generated using the standard
 2276 Crystal Ball function using the parameters given in Table 86, which correspond to the
 2277 shape parameters calculated for Run I in Section 6.5.4, but fitted with the values of the
 2278 parameters calculated for Run II.

2279 In the toy studies only the combinatorial background is generated to keep mass model
 2280 effects separate from the bias introduced by the presence of exclusive backgrounds that are
 2281 not included in the mass fit. As can be seen from Fig. 91, the effect on the inverse lifetime
 2282 pull is negligible. Overlaying the pulls of the $B_s^0 \rightarrow \mu^+ \mu^-$ yields in Fig. 91, we see that the
 2283 effect of this change is indeed tiny. Consequently no systematic uncertainty is assigned.

2284 10.4.6 Decay time acceptance

2285 Determining the acceptance function from a fit to simulated candidates assumes that the
 2286 simulation models data accurately. To test this assumption, the same method is used to
 2287 measure the already precisely measured lifetime of the $B^0 \rightarrow K^+ \pi^-$ decay, which is equal
 2288 to the mean B^0 lifetime.

2289 $B^0 \rightarrow K^+ \pi^-$ candidates are selected from the combined 2011, 2012, 2015 and 2016
 2290 data sample with the same stripping, pre-selection and BDT requirements as $B_s^0 \rightarrow \mu^+ \mu^-$

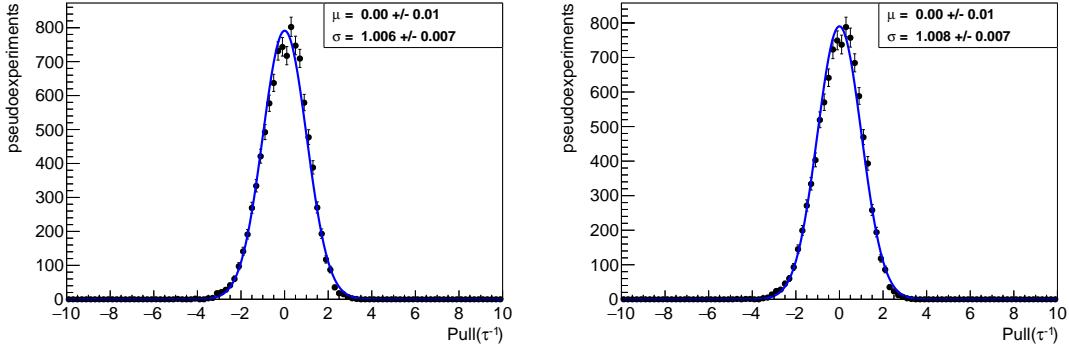


Figure 91: Pull distributions of the inverse lifetime ($\tau_{\mu^+\mu^-}^{-1}$) where $B_s^0 \rightarrow \mu^+\mu^-$ candidates are generated and fitted with the Run I Crystal Ball parameters (left) and generated with the Run I parameters but fitted with the Run II parameters (right). 10,000 pseudoexperiments were generated in each case.

2291 decays. Candidates are required to be TIS with respect to L0, Hlt1 and Hlt2 to avoid
 2292 significant differences with the $B_s^0 \rightarrow \mu^+\mu^-$ triggers, which are themselves almost entirely
 2293 lifetime-unbiased. The $B_s^0 \rightarrow \mu^+\mu^-$ PID requirement on the daughters is replaced (for the
 2294 obvious reason that it would reject all candidates) with the following requirement

$$(\text{DLL}k^+ > 10 \quad \& \quad \text{DLL}k^- < -10) \quad || \quad (\text{DLL}k^+ < -10 \quad \& \quad \text{DLL}k^- > 10) \quad (65)$$

2295 where $\text{DLL}k^{+/-}$ is the difference between the kaon and pion log-likelihoods for the pos-
 2296 tively/negatively charged daughters.

2297 $B^0 \rightarrow K^+\pi^-$ candidates are reconstructed under a $K\pi$ mass hypothesis, with the kaon
 2298 mass assigned to the daughter with the highest value of $\text{DLL}k$. A extended, unbinned
 2299 maximum-likelihood fit is made to the invariant mass distribution of selected candidates,
 2300 with $B^0 \rightarrow K^+\pi^-$ and $B_s^0 \rightarrow \pi^+K^-$ decays modelled using double Crystal Ball functions
 2301 and the combinatorial background with an exponential function. The shape parameters
 2302 of the double Crystal Ball functions (σ_1 , σ_2 , α_1 , α_2 , n_1 , n_2 and f_1) are the same for
 2303 both decays and all are fixed to values obtained using fits to truth-matched Monte Carlo
 2304 simulated $B^0 \rightarrow K^+\pi^-$ decays, with the exception of σ_1 and σ_2 , which are allowed to
 2305 float to account for differences between mass resolution in data and MC. The difference
 2306 between the means of the double Crystal Ball function, $\Delta m_{12} \equiv m_1 - m_2$, is also fixed
 2307 from the fit to simulated events. The difference between the values of m_1 for $B^0 \rightarrow K^+\pi^-$
 2308 and $B_s^0 \rightarrow \pi^+K^-$ decays, $\Delta m_{sd} \equiv m_1(B_s^0) - m_1(B^0)$, is fixed to the $B_s^0 - B^0$ mass
 2309 difference from the PDG. The value of $m_1(B^0)$, the slope of the combinatorial background
 2310 exponential and the yield of each mode is allowed to float freely in the fit. The resulting
 2311 invariant mass distribution and mass PDF is shown in Figure 92.

2312 This mass fit is used to calculate sWeights and the sWeighted decay time distribution
 2313 for $B^0 \rightarrow K^+\pi^-$ candidates is extracted. The acceptance function for $B^0 \rightarrow K^+\pi^-$ decays
 2314 is determined using a fit to weighted Monte Carlo using the method described in Section
 2315 10.2. This fit is shown in Figure 93.

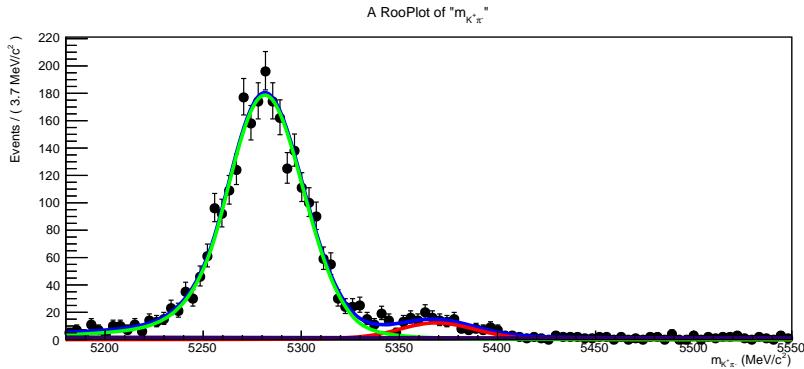


Figure 92: Extended unbinned maximum likelihood fit to the invariant mass distribution of $B^0 \rightarrow K^+\pi^-$ candidates from the CKM16 data sample, which is then used to extract sWeights. The PDF contains a $B^0 \rightarrow K^+\pi^-$ component (green), $B_s^0 \rightarrow \pi^+K^-$ component (red) and a combinatorial background component (dark blue).

The resulting acceptance function is then used to perform a fit to the sWeighted decay time distribution of real $B^0 \rightarrow K^+\pi^-$ decays from the CKM16 data sample and the effective lifetime is extracted and compared to the PDG value. The fit for the $B^0 \rightarrow K^+\pi^-$ lifetime is shown in Figure 94 and the fitted value of the lifetime is

$$\tau(B^0 \rightarrow K^+\pi^-) = 1.52 \pm 0.03 \text{ ps}, \quad (66)$$

$$1/\tau(B_s^0 \rightarrow \mu^+\mu^-) = 0.658 \pm 0.012 \text{ ps}^{-1} \quad (67)$$

which is consistent with the PDG value of $\tau(B^0) = 1.520 \pm 0.004 \text{ ps}$. A systematic uncertainty due to the agreement between data and simulation in the acceptance calculation is therefore assigned as the statistical precision of the measurement of the B^0 lifetime, therefore 0.03 ps for $\tau_{\mu^+\mu^-}$ and 0.012 ps⁻¹ for $\tau_{\mu^+\mu^-}^{-1}$.

The method to find the $B_s^0 \rightarrow \mu^+\mu^-$ acceptance outlined in Section 10.2 relies on reweighting the number of tracks in simulated $B_s^0 \rightarrow \mu^+\mu^-$ events using weights taken from a comparison between $B^0 \rightarrow K^+\pi^-$ MC and data events. To test whether weights taken from $B^0 \rightarrow K^+\pi^-$ can be used for other decays, the same method has been used to measure the $B_s^0 \rightarrow K^+K^-$ effective lifetime.

$B_s^0 \rightarrow K^+K^-$ candidates are selected in MC and data using the same stripping, pre-selection and BDT requirements as $B_s^0 \rightarrow \mu^+\mu^-$ decays. Only data from 2012 and 2015 are used because the BDT has been computed only for 2012 and 2015 $B_s^0 \rightarrow K^+K^-$ MC. Candidates are required to be TIS with respect to L0, Hlt1 and Hlt2 and both daughters are required to pass the requirement DLLk > 10.

$B_s^0 \rightarrow K^+K^-$ candidates are reconstructed under a KK mass hypothesis and an extended, unbinned maximum-likelihood fit is made to the invariant mass distribution of selected candidates, with $B_s^0 \rightarrow K^+K^-$ modelled using the same double Crystal Ball function as used for $B_s^0 \rightarrow \pi^+K^-$ events and the combinatorial background with an exponential function. The resulting invariant mass distribution and mass PDF is shown

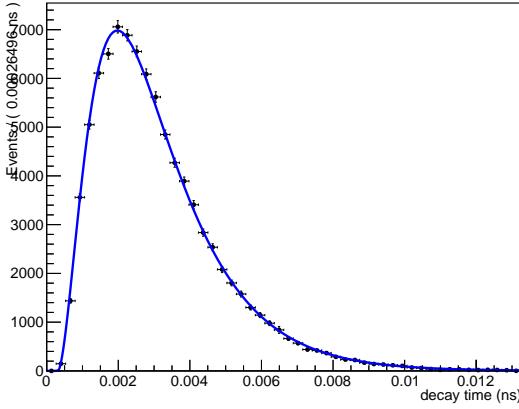


Figure 93: The fit to weighted $B^0 \rightarrow K^+ \pi^-$ Monte Carlo simulated events used to determine the acceptance function. The candidates are weighted according to the number of tracks in each event and by year (2011, 2012, 2015 and 2016) in order to produce a sample that closely models the real data.

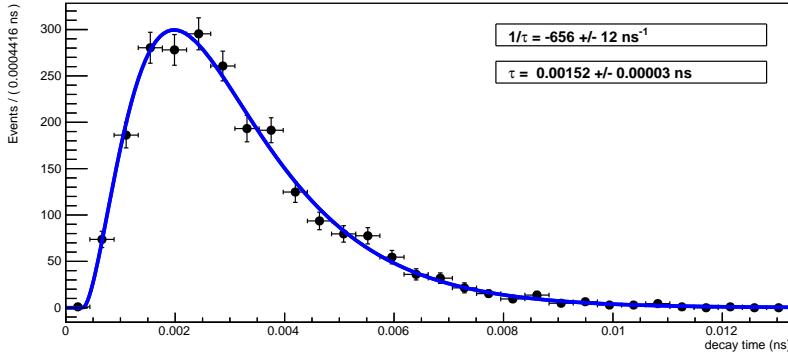


Figure 94: Unbinned maximum likelihood fit to the sWeighted decay time distribution of $B^0 \rightarrow K^+ \pi^-$ candidates from the CKM16 data sample used to determine the $B^0 \rightarrow K^+ \pi^-$ lifetime.

in Figure 95. This mass fit is used to calculate sWeights and the sWeighted decay time distribution for $B_s^0 \rightarrow K^+ K^-$ candidates is extracted.

The acceptance function for $B_s^0 \rightarrow K^+ K^-$ decays is determined using a fit to weighted Monte Carlo, as described in Section 10.2. This fit is shown in Figure 96. The acceptance function is then used to fit the sWeighted decay time distribution of real $B_s^0 \rightarrow K^+ K^-$ decays from the 2012 and 2015 data, the effective lifetime is extracted and compared to the SM prediction of $\tau_{B_s^0 \rightarrow K^+ K^-} = 1.395 \pm 0.020$ ps from [46]. The fit for the $B_s^0 \rightarrow K^+ K^-$ lifetime is shown in Figure 97 and the fitted value of the lifetime is

$$\tau(B_s^0 \rightarrow K^+ K^-) = 1.39 \pm 0.06 \text{ ps}, \quad (68)$$

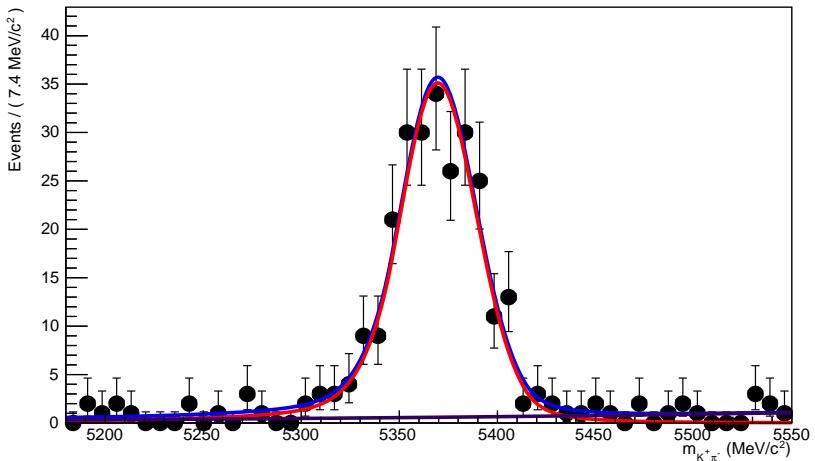


Figure 95: Extended unbinned maximum likelihood fit to the invariant mass distribution of $B_s^0 \rightarrow K^+K^-$ candidates from the 2012 and 2015 data sample, which is then used to extract sWeights. The PDF contains a $B_s^0 \rightarrow K^+K^-$ component (red) and a combinatorial background component (dark blue).

which is consistent with the predicted value.

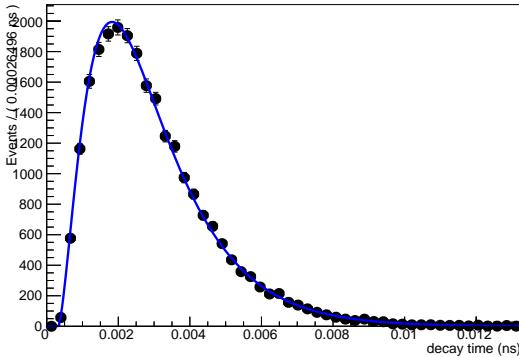


Figure 96: The fit to weighted $B_s^0 \rightarrow K^+K^-$ Monte Carlo simulated events used to determine the acceptance function. The candidates are weighted according to the number of tracks in each event and by year (2012 and 2015) in order to produce a sample that closely models the real data.

10.4.7 Combinatorial background decay time model

During the review it was discovered (thanks to Christoph Langenbruch) that the pull distributions of the inverse lifetime can be biased if the decay time distribution of the

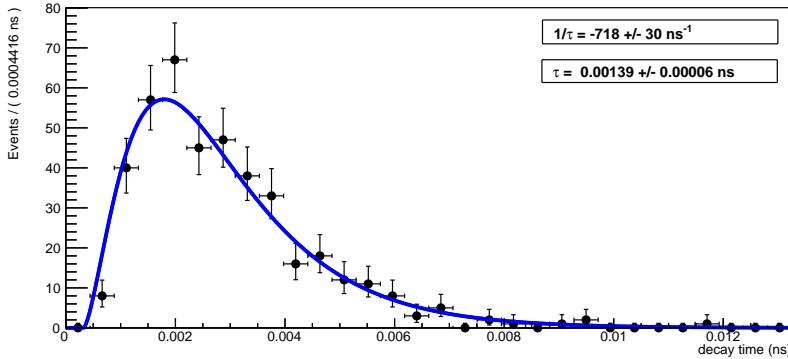


Figure 97: Unbinned maximum likelihood fit to the sWeighted decay time distribution of $B_s^0 \rightarrow K^+K^-$ candidates from the 2012 and 2015 data sample used to determine the $B_s^0 \rightarrow K^+K^-$ lifetime.

combinatorial background has a much longer mean decay time than $B_s^0 \rightarrow \mu^+\mu^-$ decays. This is an unavoidable statistical consequence of the use of sWeights to weight the decay time fit, rather than defect in the fitting procedure itself. The upper decay time cut imposed in the pre-selection goes a long way to mitigate any effect on the final result by rejecting any potential long-lived background. This section describes the method use to estimate the size of any residual bias.

The decay time distribution of the di-muon combinatorial background is *a priori* unknown, and too few events survive the full selection to allow it to be determined from the data sidebands. This is also true to Monte Carlo simulated $b\bar{b} \rightarrow X\mu^+\mu^-$ events. Instead, the distribution used in the toy studies is taken from $B^0 \rightarrow K^+\pi^-$ data sidebands, which are fitted by the acceptance function described in Section 10.2 multiplied by the sum of two exponential functions with different mean lifetimes.

The data selected are required to pass the $B^0 \rightarrow K^+\pi^-$ selection described in Section 10.4.6 and are required to have invariant masses in the range 5600-6000 MeV. The decay time distribution is fitted with the function described above and the results are given in Table 93.

This clearly assumes that $B^0 \rightarrow K^+\pi^-$ combinatorial background is a good model for di-muon combinatorial background. Given the low statistics problem it is not possible to test this assumption with the current selection, however it is possible to compare the decay time distributions of $B^0 \rightarrow K^+\pi^-$ and $B_s^0 \rightarrow \mu^+\mu^-$ combinatorial background in data for lower values of the multivariate classifier.

Table 94 shows the mean decay times of $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow K^+\pi^-$ combinatorial background candidates in bins of BDT output. In the first few BDT bins the mean decay times agree with each other reasonably well, with the $B^0 \rightarrow K^+\pi^-$ background generally having a larger mean decay time than the $B_s^0 \rightarrow \mu^+\mu^-$ background. Since it is the long-lived background that is potentially biasing this suggests that using $B^0 \rightarrow K^+\pi^-$ background as a proxy for $B_s^0 \rightarrow \mu^+\mu^-$ background is a conservative assumption.

Parameter	Value
a	$1.45 \pm 0.12 \text{ ps}^{-1}$
n	1.92 ± 0.17
t_0	0.290 ps
Γ_S	$0.77 \pm 0.17 \text{ ps}^{-1}$
Γ_L	$0.05 \pm 0.05 \text{ ps}^{-1}$
f_L	0.032 ± 0.027

Table 93: Results of the fit to the decay time distribution of the $B^0 \rightarrow K^+ \pi^-$ sidebands in data.

BDT bin	$B_s^0 \rightarrow \mu^+ \mu^-$		$B^0 \rightarrow K^+ \pi^-$	
	mean decay time / ps	Number of candidates	mean decay time / ps	Number of candidates
1	1.178 ± 0.005	50,695	1.124 ± 0.001	964,502
2	1.936 ± 0.098	244	2.394 ± 0.022	8,838
3	2.570 ± 0.327	46	2.781 ± 0.051	2,373
4	2.210 ± 0.361	17	3.023 ± 0.076	1,125
5	2.582 ± 1.103	4	3.417 ± 0.112	655
6	2.540 ± 0.390	3	3.978 ± 0.187	313
7	2.868 ± 1.048	2	4.626 ± 0.363	109
8	-	0	5.706 ± 0.683	35

Table 94: Mean decay times of $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow K^+ \pi^-$ combinatorial background candidates from data upper mass sidebands in bins of BDT output.

To evaluate the size of any systematic bias due to uncertainty on the decay time distribution of combinatorial background candidates, two sets of 10,000 pseudoexperiments are carried out; one set using the values of the parameters listed in Table 93 and one where the lifetimes of the two exponentials and the fraction of the long-lived component are increased by one standard deviation. This is done since the effect is more pronounced the longer the mean lifetime of the background. The resulting pull distributions are shown in Figure 98. We see that the mean bias does not change and the width of the pull increases by 0.008. Since the bias in the width is the larger effect we conservatively take this value multiplied by the median statistical uncertainty on the inverse lifetime (0.11 ps^{-1}), which is equal to 0.0009 ps^{-1} .

10.4.8 Incorrect assignment of primary vertices cross check

Measuring the $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime accurately relies on the B_s^0 candidate being assigned to the correct primary vertex in the event, incorrect assignment would lead to the

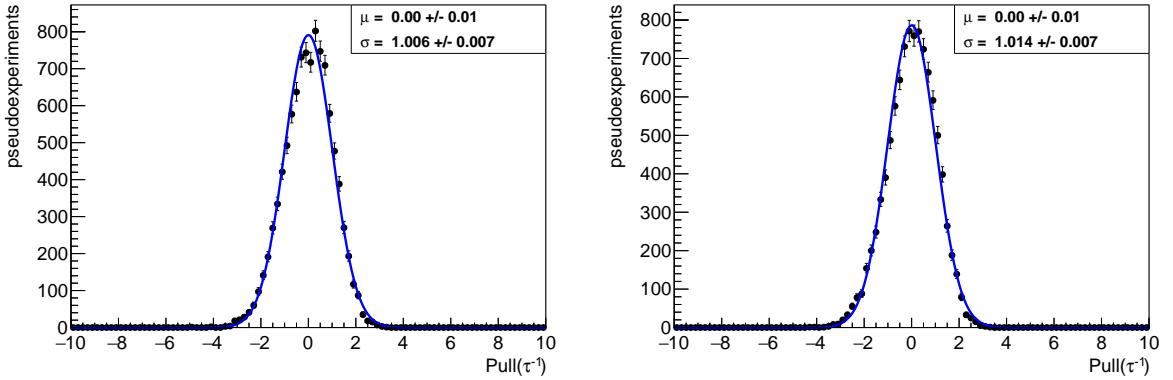


Figure 98: Pull distributions of the inverse lifetime produced using the standard background decay time model (left) in Table 93 (left) and with the lifetimes and fraction of the long-lived component increased by 1 standard deviation (right).

wrong value for the decay time of an event. In the papers in Refs. [47, 48], the detector resolution is modelled by the sum of two Gaussians and a third Gaussian is added to resolution to describe the number of events (< 1%) with incorrectly assigned PVs, the decay time model for the $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime does not include either of these resolution effects. We use a similar model to the papers [47, 48] to check that the impact of events of the incorrectly assigned PVs on the $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime and any detector resolution effects the acceptance does not model are negligible.

The weighted $B_s^0 \rightarrow \mu^+ \mu^-$ MC used to compute the acceptance function parameters is used to determine the decay time resolution from the detector resolution and incorrectly assigned PVs. The difference between the reconstructed decay time and the true decay time of each event is computed for events that pass the full selection, the resulting distribution is fitted with the sum of three Gaussians. The Gaussians have the same mean value, which is allowed to float in the fit along with the widths and fraction of each Gaussian entering in the sum. The fit parameters are shown in Table 95 and the fit in Figure 99. The fit results have a similar form to those used in [47, 48], where the detector resolution is modelled with two narrow Gaussians and the Gaussian for incorrectly assigned PVs is broader and describes a small fraction of events.

We assume that MC gives good estimate of the number of incorrectly assigned PVs, as a check Figure 100 shows the number of $B^0 \rightarrow K^+ \pi^-$ events passing the selection outlined in Section 10.4.6 for MC and sWeighted data for each year. On average there are more PVs per event in MC compared to data therefore using MC would give an overestimation of the number of incorrectly assigned PVs expected in data.

The following model is used to generate 1 million events

$$\epsilon(t)[\mathcal{R}(t) \otimes e^{-t/\tau}] \quad (69)$$

where $\epsilon(t)$ is the acceptance function with parameters given in Table 83 and $\mathcal{R}(t)$ is the resolution function, the sum of the three Gaussians from the MC fit. Events are generated

Parameter	Fit value
μ (ps)	0.00063 ± 0.00005
$\sigma_1(ps)$	5.62 ± 0.07
f_1	0.006
$\sigma_2(ps)$	0.0573 ± 0.0003
f_2	0.313
$\sigma_3(ps)$	0.0294 ± 0.0001
f_3	0.681

Table 95: Parameters from fit to the difference between the reconstructed decay time and the true decay time of each event that passes the full selection for $B_s^0 \rightarrow \mu^+ \mu^-$ MC. The mean used for all Gaussians is μ and σ_i are the widths of each Gaussian which make up a fraction f_i for the total sum.

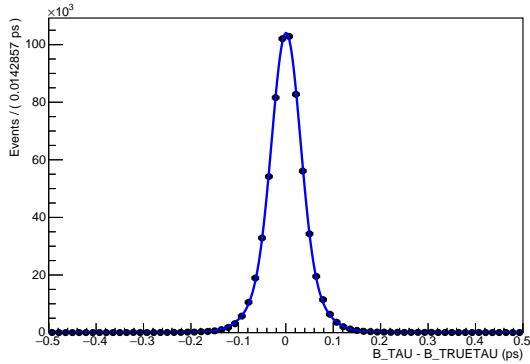


Figure 99: Fit of the sum of 3 Gaussian functions with a common mean to the difference between the reconstructed decay time and the true decay time of each event that passes the full selection for $B_s^0 \rightarrow \mu^+ \mu^-$ MC.

assuming the $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime is equal to the lifetime of the heavy B_s^0 mass eigenstate. A fit to the generated events is performed to extract the lifetime, in this fit the resolution is not included and only the acceptance function and the exponential for the lifetime. The results from the fit are $\tau(B_s^0) = 1.6098 \pm 0.0014$ ps, which is consistent with the lifetime of generate events. The difference between the lifetime used to generate events and the fitted value is 0.0002 ps, a factor of 10 smaller than the lowest systematic uncertainty, this cross check shows that the presence of incorrectly assigned PVs or detector resolution effects that are not included in the acceptance function have a negligible effect on the $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime.

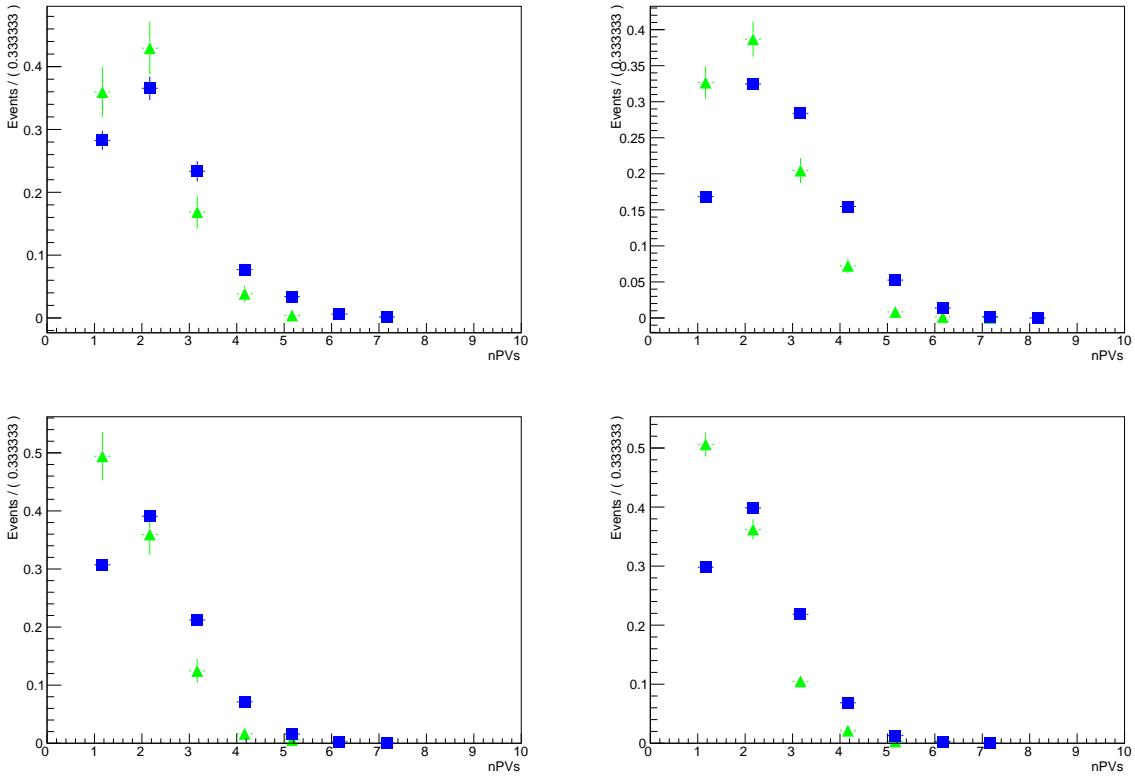


Figure 100: The distributions for the number of primary vertices in an event $B^0 \rightarrow K^+ \pi^-$ data (green triangles) and simulated candidates (blue squares) for 2011 (top left), 2012 (top right), 2015 (bottom left) and 2016 (bottom right), for events that pass the selection in Section 10.4.6.

10.4.9 Total systematic uncertainty

The full list of significant systematic uncertainties are listed in Table 96. Adding the systematic uncertainties in quadrature the total systematic uncertainty for $\tau_{\mu^+\mu^-}^{-1}$ is $\sigma(\text{syst.}) = 0.014 \text{ ps}^{-1}$, which is approximately 13% of the median expected statistical uncertainty and the total systematic uncertainty for $\tau_{\mu^+\mu^-}$ is $\sigma(\text{syst.}) = 0.05 \text{ ps}$, which is 18% of the median expected statistical uncertainty.

Source	Uncertainty on $\tau_{\mu^+\mu^-}^{-1}$ / ps $^{-1}$	Uncertainty on $\tau_{\mu^+\mu^-}$ / ps
Fit accuracy	negligible	0.03
Exclusive backgrounds	0.003	0.007
Production asymmetry	0.001	0.002
Mix of eigenstates	0.007	0.018
Decay time acceptance	0.012	0.03
Background decay time model	0.0009	0.002
Total	0.014	0.05

Table 96: List of all systematic uncertainties for the $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime. Uncertainties have been converted between the lifetime and its inverse using the propagation of uncertainties formula.

10.5 Results

The unblinded results for the fit to the $B_s^0 \rightarrow \mu^+\mu^-$ invariant mass distribution and weighted decay time are shown in Figure 101 and the yields for $B_s^0 \rightarrow \mu^+\mu^-$ and the combinatorial background are shown in Table 97. The measured $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime is

$$\tau(B_s^0 \rightarrow \mu^+\mu^-) = 2.04 \pm 0.44 \pm 0.05 \text{ ps} \quad (70)$$

$$1/\tau(B_s^0 \rightarrow \mu^+\mu^-) = 0.489 \pm 0.117 \pm 0.014 \text{ ps} \quad (71)$$

where the uncertainties are the statistical and systematic uncertainties, respectively. The result is within 1σ of the SM prediction of the lifetime of the heavy B_s^0 mass eigenstate and within 1.5σ of the lifetime for the light B_s^0 mass eigenstate.

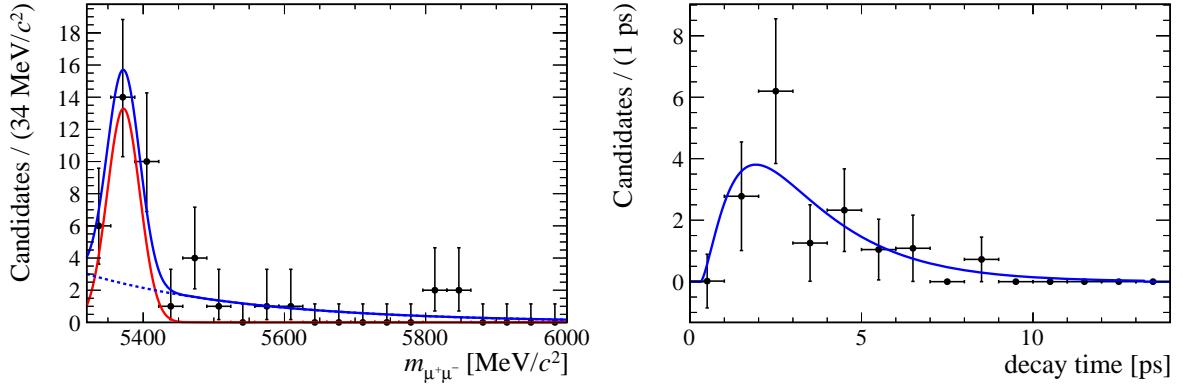


Figure 101: Unblinded fits for to the $B_s^0 \rightarrow \mu^+\mu^-$ invariant mass (left) and decay time (right). The combinatorial background is shown by the blue dashed line and $B_s^0 \rightarrow \mu^+\mu^-$ by the red solid line in the $B_s^0 \rightarrow \mu^+\mu^-$ invariant mass fit.

Component	Yield
$B_s^0 \rightarrow \mu^+ \mu^-$	22 ± 6
Combinatorial background	20 ± 6

Table 97: $B_s^0 \rightarrow \mu^+ \mu^-$ and combinatorial background yields for a fit to the $B_s^0 \rightarrow \mu^+ \mu^-$ invariant mass distribution used to compute sWeights to measure the $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime.

2441 The observed yields for both $B_s^0 \rightarrow \mu^+ \mu^-$ and the combinatorial background are smaller
 2442 than expected, therefore toy studies were preformed to check whether the coverage of the
 2443 statistical uncertainty for $\tau_{\mu^+ \mu^-}$ is still reasonable for the observed number of events. The
 2444 toy studies for the coverage check performed in Section 10.3.4 were repeated but changing
 2445 the expected $B_s^0 \rightarrow \mu^+ \mu^-$ and the combinatorial background yields to the observed values.
 2446 The contributions from $B^0 \rightarrow \mu^+ \mu^-$, $B \rightarrow h^+ h'^-$ and the exclusive backgrounds remain at
 2447 the expected value. The results are presented in Table 98. The coverage for both $\tau_{\mu^+ \mu^-}$
 2448 and $\tau_{\mu^+ \mu^-}^{-1}$ is good and extremely similar coverage quoted in Table 87. Therefore since the
 2449 $\tau_{\mu^+ \mu^-}$ is the more interesting from a physics perspective and the coverage of the measured
 2450 uncertainty is reasonable good, the lifetime can be presented as the result rather than the
 2451 inverse lifetime.

	$\tau_{\mu^+ \mu^-}$	$\tau_{\mu^+ \mu^-}^{-1}$	Gaussian
1σ	68.83%	67.76%	68.27%
2σ	93.11%	95.55%	95.45 %
3σ	97.92%	99.67%	99.73 %

Table 98: The percentage of fitted values of $\tau_{\mu^+ \mu^-}$ and $\tau_{\mu^+ \mu^-}^{-1}$ falling within 1, 2 and 3 standard deviations of the input value calculated from 10,000 pseudoexperiments at the $B_s^0 \rightarrow \mu^+ \mu^-$ and the combinatorial background yields and expected $B^0 \rightarrow \mu^+ \mu^-$, $B \rightarrow h^+ h'^-$ and the exclusive background yields. The intervals for a Gaussian distribution are listed for comparison.

2452 11 Conclusions

2453 In summary, a search for the rare decays $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ is performed
2454 in pp collision data corresponding to a total integrated luminosity of 4.4 fb^{-1} . The
2455 $B_s^0 \rightarrow \mu^+ \mu^-$ signal is seen with a significance of 7.8σ and provides the first observation
2456 of this decay from a single experiment. The time-integrated $B_s^0 \rightarrow \mu^+ \mu^-$ branching
2457 fraction is measured to be $(2.8 \pm 0.6) \times 10^{-9}$. It is the most precise measurement of
2458 this quantity to date. The first measurement of the $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime,
2459 $\tau(B_s^0 \rightarrow \mu^+ \mu^-) = 2.04 \pm 0.44 \pm 0.05 \text{ ps}$, is presented. No evidence of $B^0 \rightarrow \mu^+ \mu^-$ signal is
2460 found and the upper limit $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) < 6.1 \times 10^{-10}$ at 95 % CL is set. The results are
2461 in agreement with the SM prediction and set tighter constraints on possible new physics
2462 contributions to these decays

2463 **Appendix**

2464 **A Description of the toy model used for the lifetime
2465 measurement**

2466 **Toy description needs updating as we have now switched to use the same mass
2467 PDFs as the branching fraction analysis.** A simple toy model is used to conduct
2468 pseudoexperiments in order to optimise the selection, determine the best choice of fit
2469 procedure as well as for evaluating systematic uncertainties. This is therefore described
2470 before the rest of the analysis. The process is:

- 2471 1. Construct mass and decay time PDFs for the relevant particle decays (and back-
2472 grounds).
- 2473 2. Generate a data sample, including both mass and decay time, from these PDFs with
2474 statistics corresponding to those collected by LHCb in Run I and 2015.
- 2475 3. Perform a fit to the resulting invariant mass distribution and calculate the sWeights.
- 2476 4. Fit the weighted decay time distribution and extract an estimate of the effective
2477 lifetime.
- 2478 5. Repeat from step 2.

2479 A large number of pseudoexperiments are conducted for each set of conditions and fit
2480 procedures under study, and the fitted values of particle yields and lifetimes are compared
2481 to their inputs to check the validity of the fit. The number of each component generated
2482 in each pseudoexperiment is allowed to fluctuate around the input value according to a
2483 Poisson distribution.

2484 The mass and decay time PDFs used to generate each mode are listed in Tables 99
2485 and 100 ²² and the mass spectrum of a typical Run I + 2015 pseudoexperiment is shown in
2486 Figure 102.

²²For peaking backgrounds with mass distributions modelled using the sum of two Crystal Ball functions the fraction parameter f_1 is defined in the mass range 4.9 – 6.0 GeV; if a narrower mass range is used in the fit then the appropriate fractions are recalculated on the fly.

Decay mode	Mass PDF	Parameters	Decay time PDF	Parameters
$B_s^0 \rightarrow \mu^+ \mu^-$	Crystal Ball	$m = 5371.85$ MeV $\sigma = 23.2334$ MeV $\alpha = 2.06505$ $n = 1.11846$	Exponential	$\tau_{B_s} = 1.512$ ps $\tau_L = 1.405$ ps $\tau_H = 1.661$ ps
$B^0 \rightarrow \mu^+ \mu^-$	Crystal Ball	$m = 5284.9$ MeV $\sigma = 22.838$ MeV $\alpha = 2.06505$ $n = 1.11846$	Exponential	$\tau = 1.519$ ps

Table 99: Mass and decay time PDFs used to generate $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ decays in the toy Monte Carlo pseudoexperiments.

Decay mode	Mass PDF	Parameters	Decay time PDF	Parameters
Combinatorial background	Exponential	$\lambda = [-0.0014 - 0.0] \text{ eV}^{-1}$ (depending on BDT cut)	Sum of three exponentials	$\tau_1 = 0.3 \text{ ps}$ $\tau_2 = 1.0 \text{ ps}$ $\tau_3 = 2.0 \text{ ps}$ $f_1 = 0.85$ $f_2 = 0.05$
$B \rightarrow h^+ h^-$ mis-ID background	Sum of two Crystal Ball functions	$m_1 = 5256.99 \text{ MeV}$ $\sigma_1 = 36.6098 \text{ MeV}$ $\alpha_1 = 0.802753$ $n_1 = 2.22626$ $m_2 = 5199.68 \text{ MeV}$ $\sigma_2 = 127.134 \text{ MeV}$ $\alpha_2 = -1.89764$ $n_2 = 2.57117$ $f_1 = 0.602049$	Exponential	$\tau = 1.504 \text{ ps}$
$B^0 \rightarrow \pi^- \mu^+ \nu_\mu$ and $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$	Sum of two Crystal Ball functions	$m_1 = 4982.08 \text{ MeV}$ $\sigma_1 = 87.8604 \text{ MeV}$ $\alpha_1 = 0.0305266$ $n_1 = 78.9461$ $m_2 = 5132.18 \text{ MeV}$ $\sigma_2 = 55.9449 \text{ MeV}$ $\alpha_2 = -4.85638$ $n_2 = 2.64781$ $f_1 = 0.775402$	Exponential	$\tau = 1.216 \text{ ps}$
$B^0 \rightarrow \pi^0 \mu^+ \mu^-$ and $B^+ \rightarrow \pi^+ \mu^+ \mu^-$	Sum of two Crystal Ball functions	$m_1 = 4996.57 \text{ MeV}$ $\sigma_1 = 59.6921 \text{ MeV}$ $\alpha_1 = 0.310876$ $n_1 = 99.9999$ $m_2 = 5086.38 \text{ MeV}$ $\sigma_2 = 41.0478 \text{ MeV}$ $\alpha_2 = -6.07878$ $n_2 = 10.3514$ $f_1 = 0.72333$	Exponential	$\tau = 1.247 \text{ ps}$
$\Lambda_b^0 \rightarrow p \mu^- \bar{\nu}_\mu$	Sum of two Crystal Ball functions	$m_1 = 5211.67 \text{ MeV}$ $\sigma_1 = 98.3845 \text{ MeV}$ $\alpha_1 = 0.187159$ $n_1 = 99.9982$ $m_2 = 5327.94 \text{ MeV}$ $\sigma_2 = 63.8051 \text{ MeV}$ $\alpha_2 = -5.45493$ $n_2 = 76.3724$ $f_1 = 0.818301$	Exponential	$\tau = 1.012 \text{ ps}$

Table 100: Mass and decay time PDFs used to generate backgrounds in the toy Monte Carlo pseudoexperiments. The parameters of the decay time PDFs for each mode are determined from fits to full LHCb MC simulated events. The slope (λ) parameter of the combinatorial background mass PDF varies depending on the BDT cut imposed in the selection.

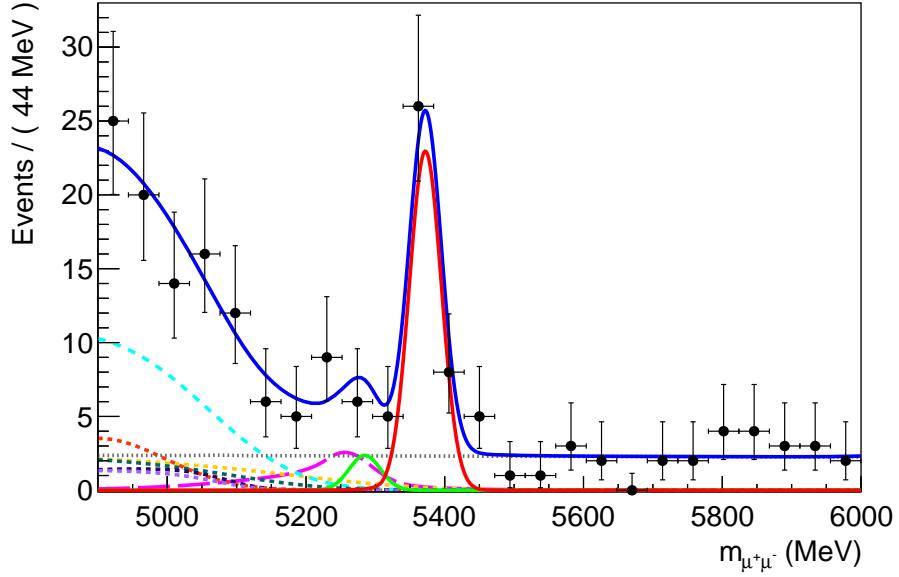


Figure 102: The di-muon invariant mass spectrum generated for a typical toy pseudoexperiment, with the PDFs used to generate the data overlaid. The components are as follows: $B_s^0 \rightarrow \mu^+ \mu^-$ solid red line, $B^0 \rightarrow \mu^+ \mu^-$ solid green line, combinatorial background dashed blue line, $B \rightarrow h^+ h^-$ dashed magenta line, $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$ and $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ dashed dark violet line, $B^0 \rightarrow \pi^0 \mu^+ \mu^-$ and $B^+ \rightarrow \pi^+ \mu^+ \mu^-$ dashed cyan line and $\Lambda_b^0 \rightarrow p \mu^- \bar{\nu}_\mu$ dashed orange line. The statistics correspond to those expected for the combined 2011, 2012 and 2015 data.

2487 The effect of the detector and selection on the decay time distributions of the various
 2488 modes is modelled using the function

$$\epsilon(t) = \frac{[a(t - t_0)]^n}{1 + [a(t - t_0)]^n} \quad (72)$$

2489 where t_0 , a and n have the same values as used in the fit to data (see Section 10.2). This
 2490 acceptance curve is used for all modes, such that the final decay time PDF is simply the
 2491 acceptance curve multiplied by the decay time PDF of the specific mode. This function
 2492 is also used to fit the toy data. Figure 103 shows the resulting decay time PDF used to
 2493 generate $B_s^0 \rightarrow \mu^+ \mu^-$ candidates, with $\tau_{\mu^+ \mu^-} = \tau_H$.

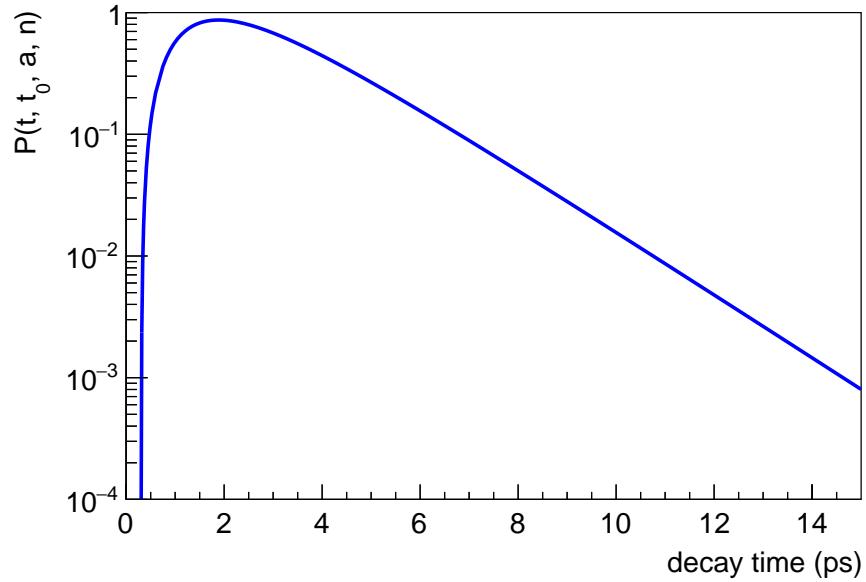


Figure 103: The decay time PDF used to generate $B_s^0 \rightarrow \mu^+ \mu^-$ candidates in pseudoexperiments, including the effects of detector, trigger and selection efficiency.

2494 The same mass PDFs are used to fit and generate the data. In the minimal fit only
 2495 $B_s^0 \rightarrow \mu^+ \mu^-$ and combinatorial background components are included; in the most complex
 2496 fit a $B^0 \rightarrow \mu^+ \mu^-$ component and all background components are also included. In all cases
 2497 the shape parameters of the various mass PDFs are fixed to the values with which the
 2498 data were generated and the yields of peaking backgrounds are also fixed. Only the yields
 2499 of $B_s^0 \rightarrow \mu^+ \mu^-$, $B^0 \rightarrow \mu^+ \mu^-$ and combinatorial background are allowed to float freely, along
 2500 with the slope (λ) parameter of the combinatorial background mass PDF.

2501 The resulting sWeighted decay time distribution for $B_s^0 \rightarrow \mu^+ \mu^-$ decays is then
 2502 fitted using a simple exponential function multiplied by the acceptance function given in
 2503 Equation 72.

2504 **B Distributions of the long track isolation input variables**

2505

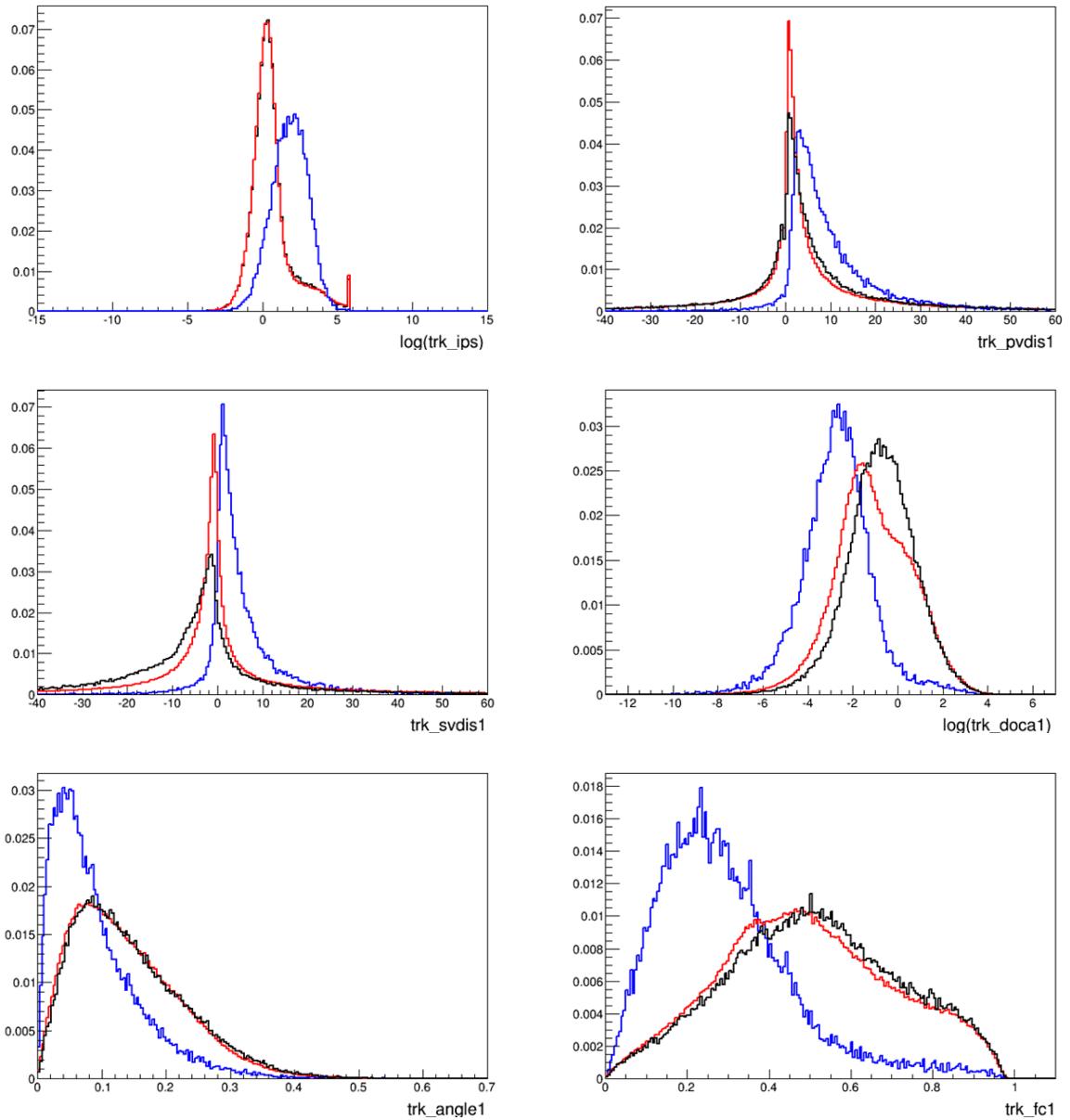


Figure 104: Distributions of the variables belonging to “set A” defined in Sec. 6.1.1. Blue: non-isolating tracks in generic MC; Black: isolating tracks in signal MC; Red: isolating tracks in generic MC (not used for the isoBDT training).

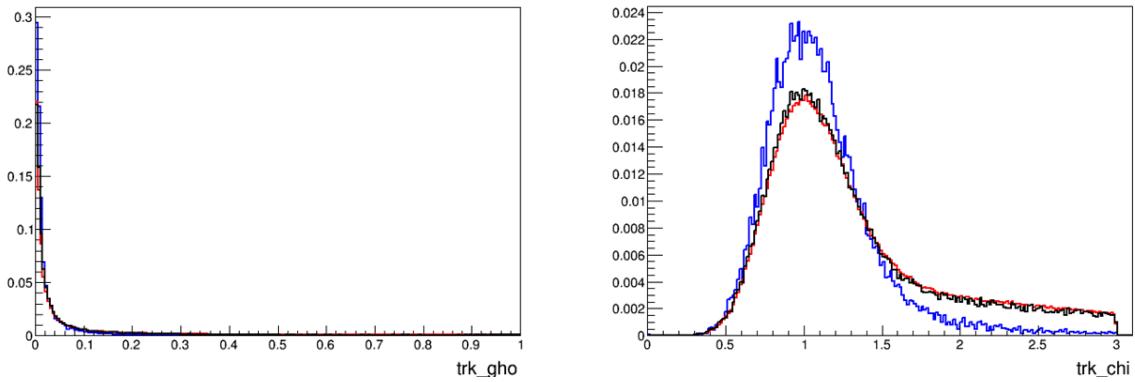


Figure 105: Distributions of the variables belonging to “set B” defined in Sec. 6.1. Blue: non-isolating tracks in generic MC; Black: isolating tracks in signal MC; Red: isolating tracks in generic MC (not used for the isoBDT training).

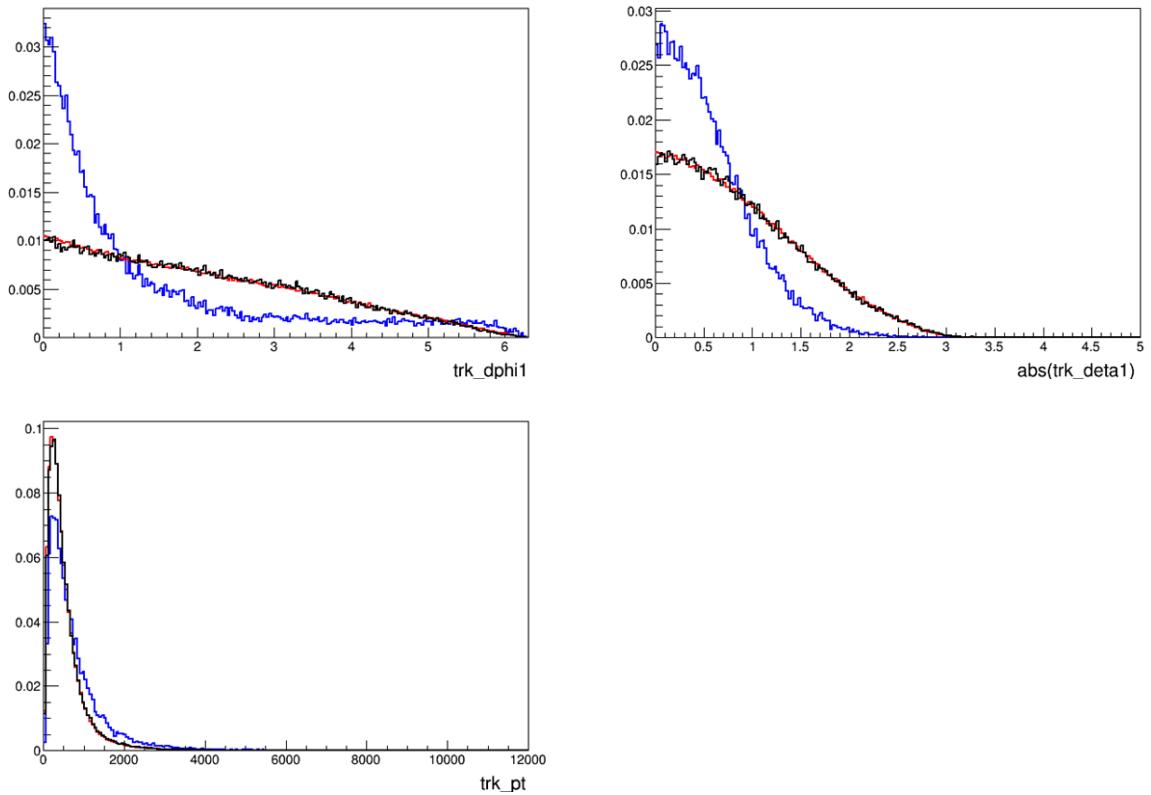


Figure 106: Distributions of the variables belonging to “set C” defined in Sec. 6.1. Blue: non-isolating tracks in generic MC; Black: isolating tracks in signal MC; Red: isolating tracks in generic MC (not used for the isoBDT training).

2506 **C Definition of BDT 2013 input variables**

- 2507 • Cut-based long track isolation (see Sec. 6.1.1).
- 2508 • B_yet_another_CDF_iso: the B isolation based on the CDF definition.
- 2509 • mu_DeltaEta: absolute value of the difference between the pseudorapidity of the two
2510 muon candidates.
- 2511 • mu_AbsPhi: absolute value of the difference between the azimuthal angles of the
2512 two muon candidates.
- 2513 • mu_MINIPS: smallest value among the muon impact parameter significance of two
2514 muons with respect to the primary vertex associated to the $B \rightarrow \mu\mu$ candidate.
- 2515 • B_TAU_ps: B meson proper time.
- 2516 • B_IP_OWNPV: the impact parameter of the B :
- 2517 • B_doca: distance of closest approach between the two muons.
- 2518 • B_PT: the transverse momentum of the B meson.
- 2519 • B_cosnk: the cosine of the angle between the muon momentum in the dimuon rest
2520 frame and the vector perpendicular to the B momentum and the beam axis.
- 2521 • B_otherB_ang: angle between the B candidates's momentum and the thrust momen-
2522 tum of the B , defined as the sum of momenta of all the long tracks coming from the
2523 B PV and excluding those coming from long lived particles. If no such tracks are
2524 available, the variable is set to 0.
- 2525 • B_otherB_boo_ang: angle between the direction of the positive muon candidate in
2526 the rest frame of the B and the thrust momentum in the B rest frame.

2527 **D Distribution of BDT input variables in MC train-**
2528 **ing sample, more recent MC and data**

2529 In Figs 107 and 108 we show the distributions of the BDT and its input variables in the
2530 MC sample used to train the BDT ($B_s \rightarrow \mu\mu$ MC12 sim06 and $b\bar{b} \rightarrow \mu\mu X$ MC sim06),
2531 in the right mass data sideband (Run1) and in a more recent version of $B_s \rightarrow \mu\mu$ MC
2532 (sim08).

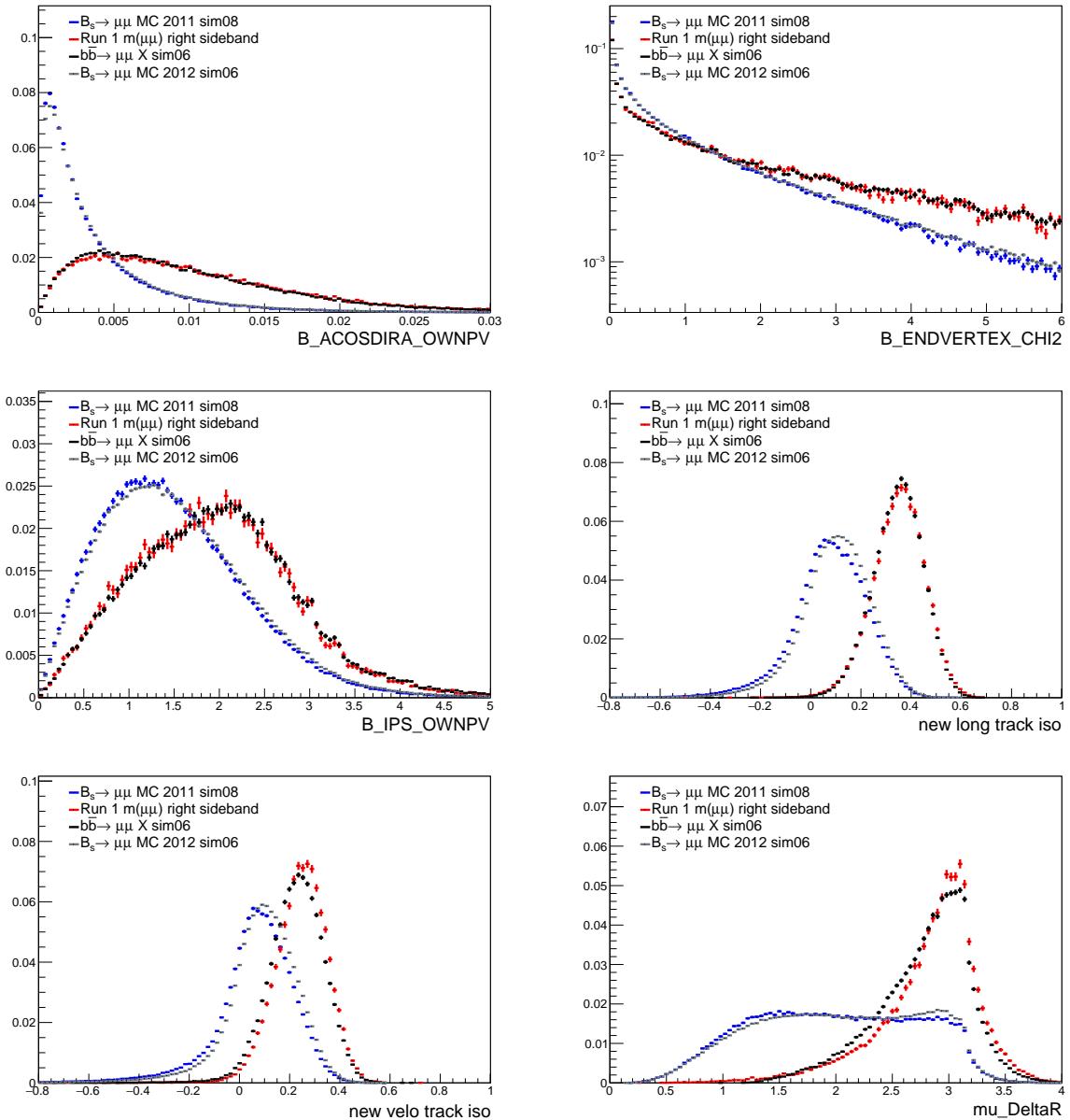


Figure 107: Distribution of 6 out of 7 BDT input variables in the MC sample used to train the BDT ($B_s \rightarrow \mu\mu$ MC12 sim06 and $b\bar{b} \rightarrow \mu\mu X$ MC sim06), in the right mass data sideband (Run1) and in a more recent version of $B_s \rightarrow \mu\mu$ MC (sim08)

2533 E Comparison of BDT training strategies

2534 Some time ago a study was performed to evaluate the BDT performance on the mass
2535 right data sideband as a function of the BDT training strategy. The BDT²³ was trained

²³An old version based on 12 variables, as the one used in 2013 except for a different track isolation.

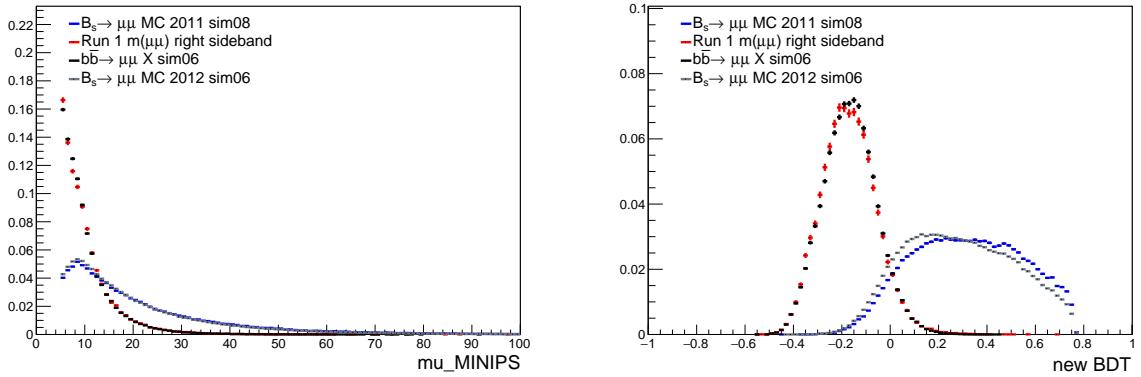


Figure 108: Distribution of the μ_{MINIPS} BDT input variable and of the resulting BDT in the MC sample used to train the BDT ($B_s \rightarrow \mu\mu$ MC12 sim06 and $b\bar{b} \rightarrow \mu\mu X$ MC sim06), in the right mass data sideband (Run1) and in a more recent version of $B_s \rightarrow \mu\mu$ MC (sim08). The distributions of the other 6 input variables are shown in Fig. 107.

2536 using $B_s \rightarrow \mu\mu$ MC12 sim06 events as signal sample and one of the following options as
 2537 background sample:

- 2538 • $b\bar{b} \rightarrow \mu\mu X$ MC sim06 events
- 2539 • *1/3 of randomly selected events in the right mass data sideband.* The right mass
 2540 data sideband was divided in three equal-size independent subsets. The first one
 2541 was used as training sample and the BDT output was tested on the second subset.
 2542 Finally, the BDT was used in the third subset. The procedure was applied iteratively
 2543 to the three subsets. As a result the whole data sample had the BDT trained on a
 2544 statistically independent sample corresponding to 1/3 of the overall sample size.
- 2545 • *5/11 of randomly selected events in the right mass data sideband.* As in the previous
 2546 bullet, but the right mass data sideband was trained on a 5/11 subset, tested on a
 2547 different 5/11 subset and finally applied to the remaining 1/11 subset. The procedure
 2548 was applied iteratively 11 times.
- 2549 • *1/3 or 5/11 of randomly selected events in the left+right mass data sideband.* The
 2550 same procedure as described in the previous two bullets was applied, but the training
 2551 subsets were composed of both right and left mass data sidebands.

2552 The outcome of the test is shown in fig. 109. The BDT rejection power in the five scenarios
 2553 looks quite similar. In the most sensitive region the performance of the BDT trained on
 2554 $b\bar{b} \rightarrow \mu\mu X$ MC seems slightly lower, but still consistent within the statistical uncertainty.
 2555 Hence, it was decided to keep using the MC background sample for training. The study
 2556 may be repeated in a next round of the analysis using the new BDT and on the available
 2557 Run 1 + Run 2 datasets.

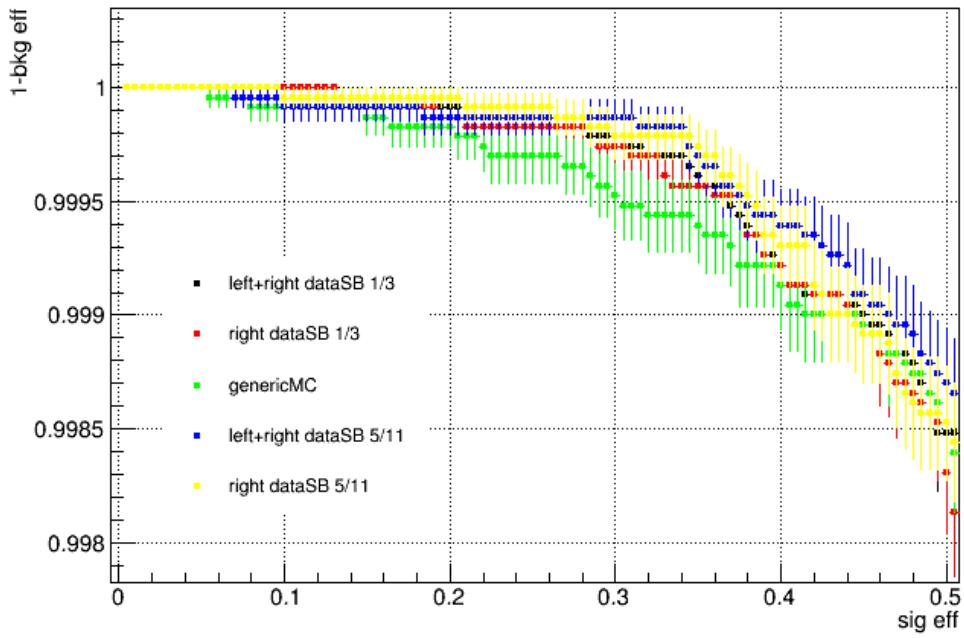
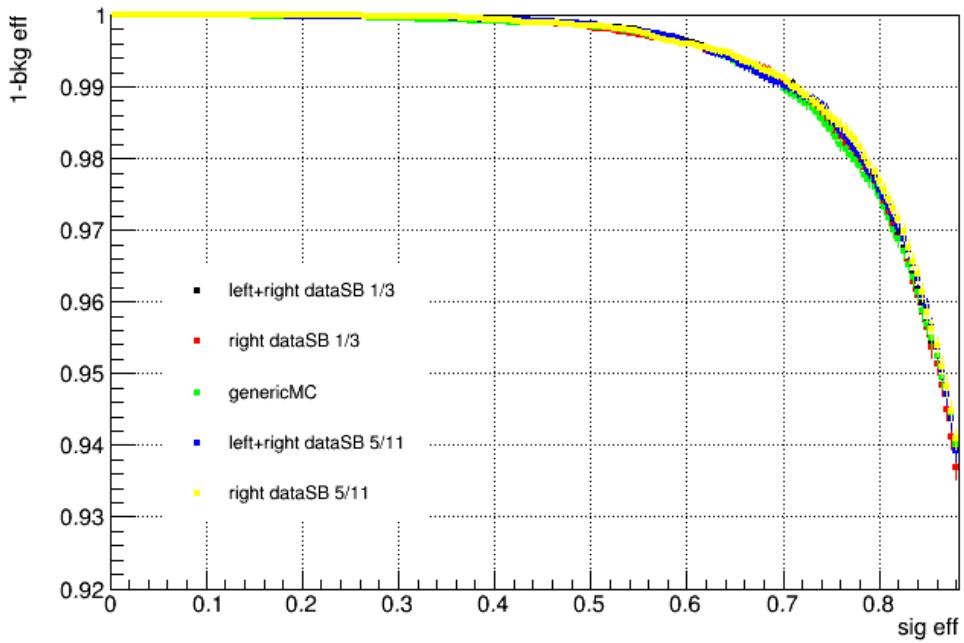


Figure 109: ROC of an old version of the BDT as a function of the BDT training sample. The signal training sample is always $B_s \rightarrow \mu\mu$ MC12 sim06 while the background training sample varies: 1/3 or 5/11 of the left+right mass data sidebands (black and blue, respectively), 1/3 or 5/11 of the right data sidebands (red and yellow, respectively) and $b\bar{b} \rightarrow \mu\mu X$ MC sim06 (green). The ROC is always evaluated on $B_s \rightarrow \mu\mu$ MC12 sim06 and right mass data sideband events. The bottom plot is a zoom of the top plot in the most sensitive region.

2558 **F Comparison of BDT input variables in Run I, 2015**
2559 **and 2016 data**

2560 To investigate the lack of events in the upper BDT region ($BDT > 0.4$) for 2015 data, the
2561 variables used as input for the BDT, were compared in Run I, 2015 and 2016 data. Four
2562 cases were examined:

- 2563 1. Events with mass between 5200 and 5600 MeV/ c^2 for $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ (with the full
2564 PID selection).
- 2565 2. Events with mass between 5200 and 5600 MeV/ c^2 for $B_{(s)}^0 \rightarrow h^+ h^-$ (without applying
2566 any PID selection).
- 2567 3. Events with mass between 4900 and 6000 MeV/ c^2 for $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ (with the full
2568 PID selection).
- 2569 4. Events with mass between 4900 and 6000 MeV/ c^2 for $B_{(s)}^0 \rightarrow h^+ h^-$ (without applying
2570 any PID selection).

2571 In the following, only cases 1 and 2 are discussed, but similar results were obtained for
2572 cases 3 and 4. In Figures 110 and 111, the distributions per dataset (Run I, 2015 and
2573 2016) for case 1 and 2 respectively are shown. Additionally, the distributions with for
2574 the upper BDT region ($BDT > 0.4$) are shown with the same normalisation per dataset.
2575 From these figures, it is concluded that the different BDT input variable distributions
2576 agree, also in the high BDT region.

2577 G Trigger BDT pdf's for signal and double misID

2578 In the following Tables 101, 102, and 103, we list the values of $\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{\text{TRIG|SEL}}$, $\epsilon^{\text{TIS}} \cdot \epsilon^{\text{HLT2,MC}}$
 2579 and their ratio normalized on the BDT range, as a function of BDT, extracted from 2012,
 2580 2015 and 2016 simulations.

Table 101: Pdf for $\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{\text{TRIG|SEL}} / (\epsilon^{\text{TIS}} \cdot \epsilon^{\text{HLT2,MC}})$ as a function of BDT bin, as extracted from $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow K^+ \pi^-$ MC12 samples, given the full $B_s^0 \rightarrow \mu^+ \mu^-$ and $B_{(s)}^0 \rightarrow h^+ h^-$ selection;

BDT range	$\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{\text{TRIG SEL}} (\%)$	$\epsilon^{\text{TIS}} \cdot \epsilon^{\text{HLT2,MC}} (\%)$	pdf
[0-0.25]	94.22 ± 0.07	7.55 ± 0.04	0.939 ± 0.005
[0.25-0.4]	95.11 ± 0.09	6.70 ± 0.05	1.069 ± 0.008
[0.4-0.5]	95.87 ± 0.11	6.69 ± 0.06	1.079 ± 0.009
[0.5-0.6]	96.31 ± 0.10	6.82 ± 0.06	1.063 ± 0.009
[0.6-0.7]	96.93 ± 0.10	6.94 ± 0.06	1.051 ± 0.009
[0.7-0.8]	97.55 ± 0.09	7.28 ± 0.06	1.009 ± 0.008
[0.8-0.9]	98.10 ± 0.08	7.81 ± 0.06	0.945 ± 0.007
[0.9-1.0]	98.67 ± 0.07	8.24 ± 0.06	0.901 ± 0.007

Table 102: Pdf for $\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{\text{TRIG|SEL}} / (\epsilon^{\text{TIS}} \cdot \epsilon^{\text{HLT2,MC}})$ as a function of BDT bin, as extracted from $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow K^+ \pi^-$ MC15 samples, given the full $B_s^0 \rightarrow \mu^+ \mu^-$ and $B_{(s)}^0 \rightarrow h^+ h^-$ selection;

BDT range	$\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{\text{TRIG SEL}} (\%)$	$\epsilon^{\text{TIS}} \cdot \epsilon^{\text{HLT2,MC}} (\%)$	pdf
[0-0.25]	91.90 ± 0.13	9.17 ± 0.09	0.987 ± 0.010
[0.25-0.4]	93.28 ± 0.15	8.30 ± 0.12	1.106 ± 0.015
[0.4-0.5]	94.16 ± 0.18	8.70 ± 0.15	1.066 ± 0.018
[0.5-0.6]	94.67 ± 0.17	9.15 ± 0.15	1.019 ± 0.017
[0.6-0.7]	95.01 ± 0.17	9.32 ± 0.15	1.004 ± 0.016
[0.7-0.8]	95.82 ± 0.15	9.65 ± 0.15	0.978 ± 0.016
[0.8-0.9]	96.66 ± 0.14	10.31 ± 0.16	0.923 ± 0.014
[0.9-1.0]	97.51 ± 0.12	10.83 ± 0.16	0.886 ± 0.013

Table 103: Pdf for $\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{\text{TRIG|SEL}} / \epsilon^{\text{TIS}} \cdot \epsilon^{\text{HLT2,MC}}$ as a function of BDT bin, as extracted from $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow K^+ \pi^-$ MC16 samples, given the full $B_s^0 \rightarrow \mu^+ \mu^-$ and $B_{(s)}^0 \rightarrow h^+ h^-$ selection;

BDT range	$\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{\text{TRIG SEL}} (\%)$	$\epsilon^{\text{TIS}} \cdot \epsilon^{\text{HLT2,MC}} (\%)$	pdf
[0-0.25]	96.20 ± 0.09	9.55 ± 0.09	0.966 ± 0.010
[0.25-0.4]	96.86 ± 0.11	8.63 ± 0.12	1.077 ± 0.015
[0.4-0.5]	97.39 ± 0.12	8.75 ± 0.14	1.068 ± 0.018
[0.5-0.6]	97.40 ± 0.12	8.67 ± 0.14	1.077 ± 0.018
[0.6-0.7]	97.73 ± 0.11	9.11 ± 0.15	1.029 ± 0.017
[0.7-0.8]	98.00 ± 0.11	9.50 ± 0.15	0.989 ± 0.016
[0.8-0.9]	98.26 ± 0.10	10.12 ± 0.16	0.931 ± 0.014
[0.9-1.0]	98.69 ± 0.08	10.80 ± 0.16	0.876 ± 0.013

2581 **H $\pi - \mu$ and $K - \mu$ fits for 2015 and 2016 data**

2582 The mass fits on 2015 data for the $\pi - \mu$ and $K - \mu$ selections are shown in Figs. 112
 2583 and 113, respectively; the mass fits on 2016 data for the $\pi - \mu$ and $K - \mu$ selections are
 2584 shown in Figs. 114 and 115, respectively. The contribution from $B_{(s)}^0 \rightarrow h^+ h^-$ with a single
 2585 misID appears as a very clear peak in these spectra, with semileptonic $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$ and
 2586 $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ contributing as a shoulder on the left side. Data in the first BDT bin,
 2587 [0 – 0.25], are not considered by this method, as they’re fully dominated by combinatorial
 2588 background.

2589 **I $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$ mass pdf’s with FF weights**

2590 The $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$ mass spectra from MC in bins of BDT are shown in Fig.. 116 with and
 2591 without per-event weights to correct the q^2 spectrum according to Ref. [20]. No effect from
 2592 the FF weights is visible.

2593 **J J/ψ veto**

2594 Figure 117 shows the distribution of the events rejected by the J/ψ veto in Run1 data, in
2595 the mass vs BDT plane. 675 out of 64k events are rejected. Five events have BDT>0.25.
2596 The veto efficiency on $B_s \rightarrow \mu\mu$ signal events (selected without specific trigger requirements)
2597 is about 99.8%. The impact of the veto on the signal efficiency is negligible, but also the
2598 impact on background is quite limited. This is due to the long track isolation variable,
2599 which classifies as background-like most of the events with an additional muon forming a
2600 real J/ψ with one of the B candidate tracks.

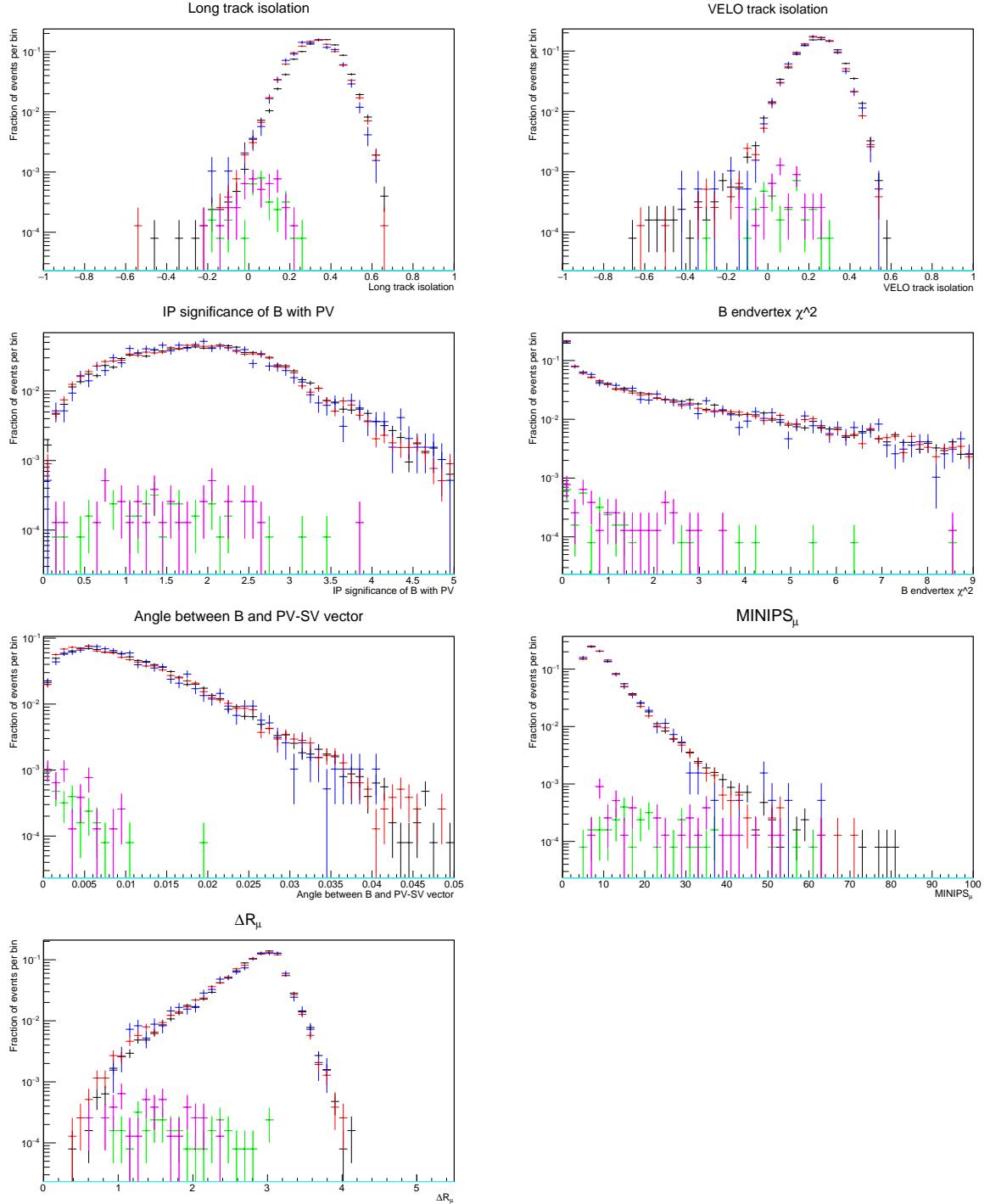


Figure 110: BDT input variable distributions for Run I (black), 2015 (blue) and 2016 (red) for case 1 (muons with PID selection from $5200 < m(\mu\mu) < 5600$ MeV). The same distributions with the extra requirement $BDT > 0.4$ are shown for Run I (green), 2015 (cyan) and 2016 (purple).

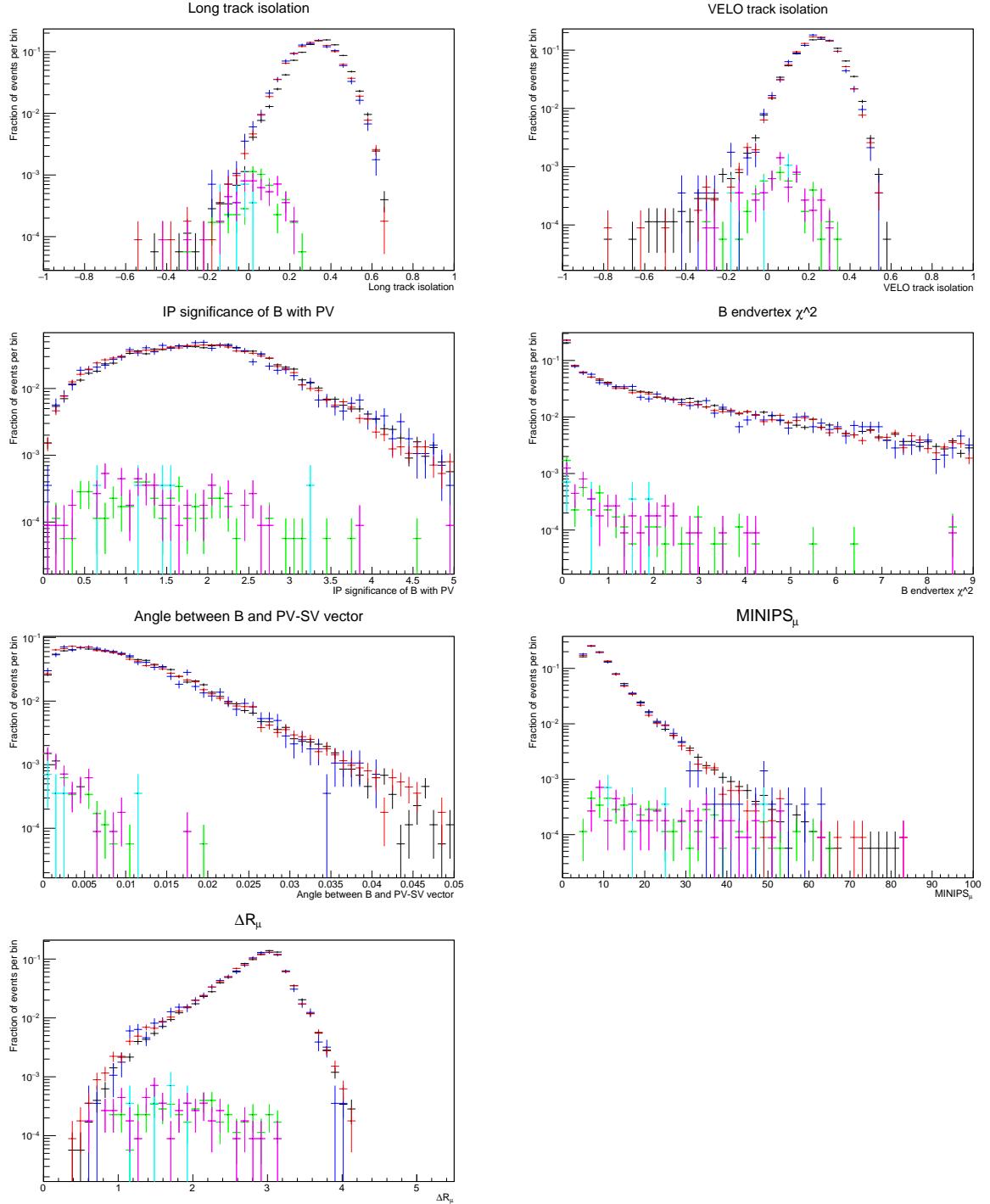


Figure 111: BDT input variable distributions for Run I (black), 2015 (blue) and 2016 (red) for case 2 (muons without any PID selection from $5200 < m(\mu\mu) < 5600$ MeV). The same distributions with the extra requirement $BDT > 0.4$ are shown for Run I (green), 2015 (cyan) and 2016 (purple).

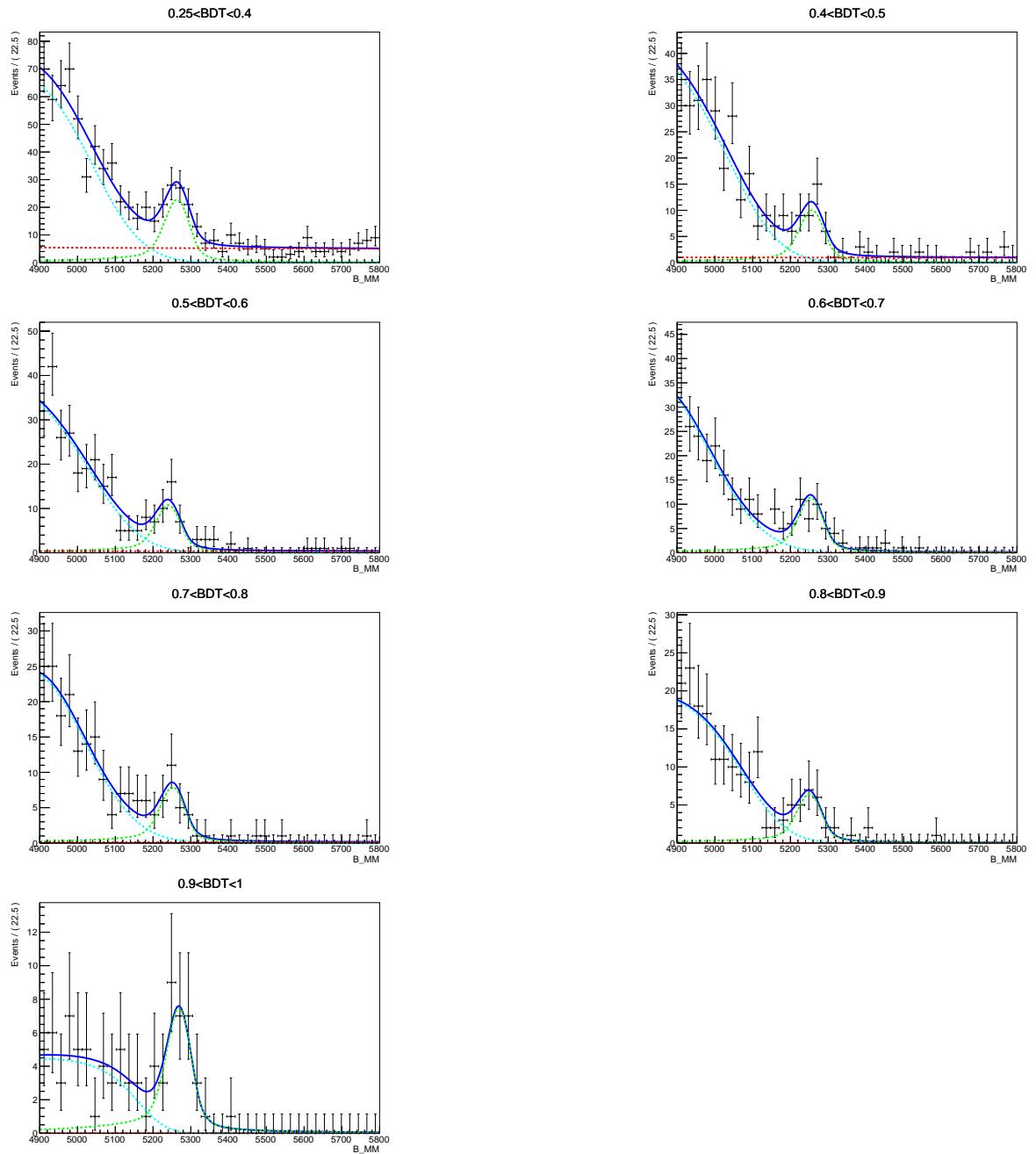


Figure 112: Fit to 2015 data under $\pi - \mu$, for BDT bins > 0.25 ; the 3 components of the fit are: $B_{(s)}^0 \rightarrow h^+ h^-$ (green), semileptonic (cyan), and combinatorial (red).

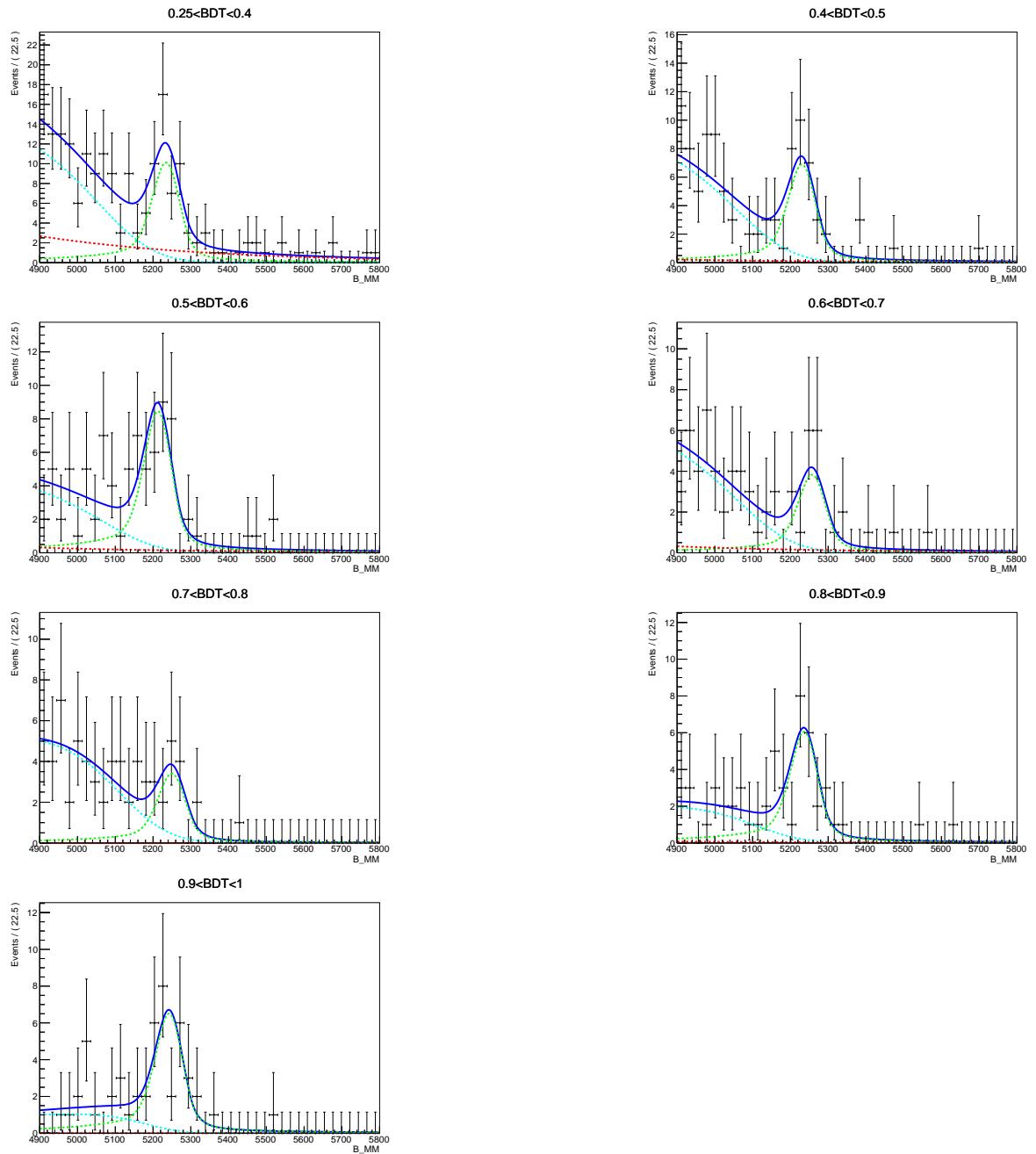


Figure 113: Fit to 2015 data under $K - \mu$, for BDT bins > 0.25 ; the 3 components of the fit are: $B_{(s)}^0 \rightarrow h^+h^-$ (green), semileptonic (cyan), and combinatorial (red).

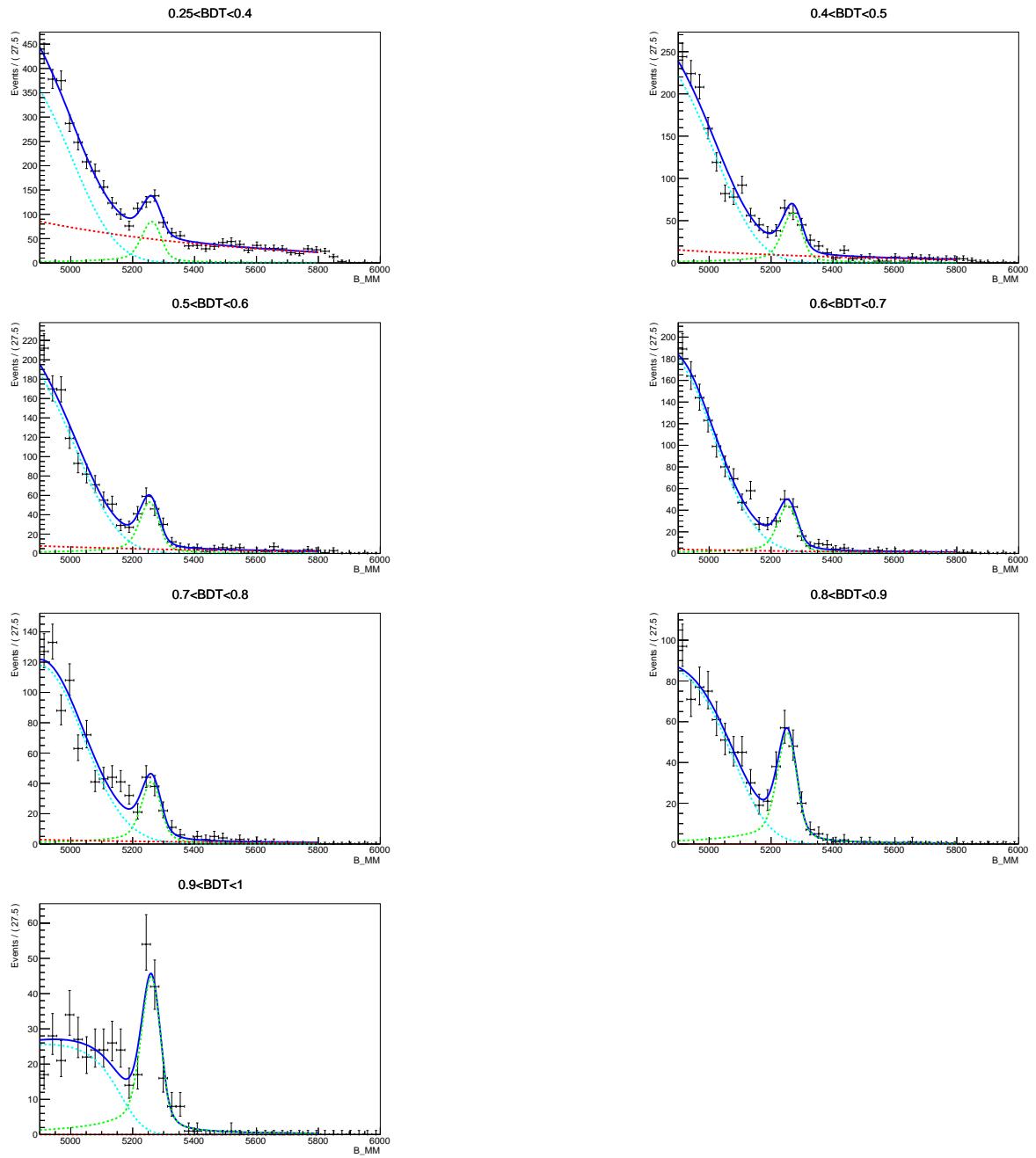


Figure 114: Fit to 2016 data under $\pi - \mu$, for BDT bins > 0.25 ; the 3 components of the fit are: $B_{(s)}^0 \rightarrow h^+h^-$ (green), semileptonic (cyan), and combinatorial (red).

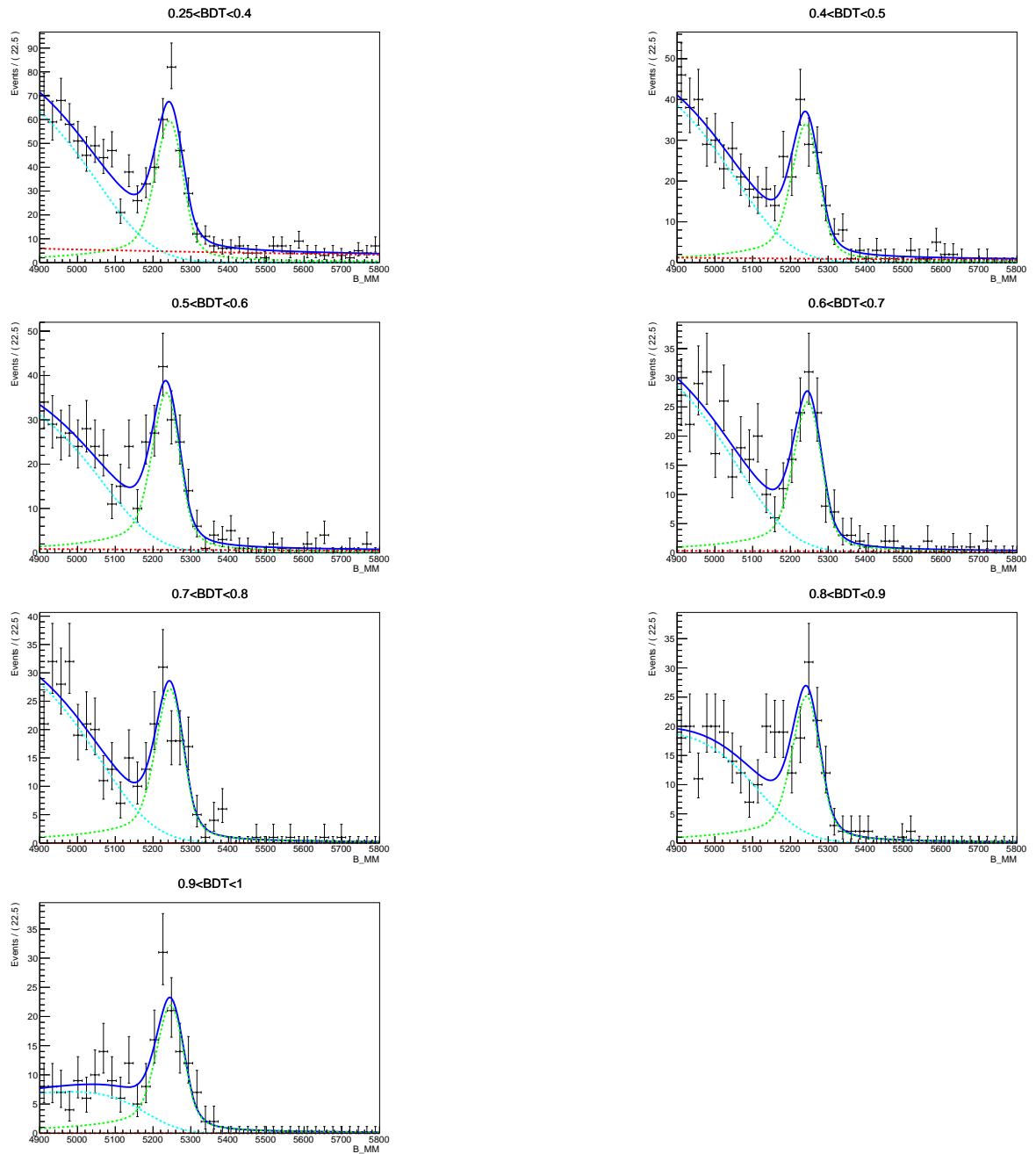


Figure 115: Fit to 2016 data under $K - \mu$, for BDT bins > 0.25 ; the 3 components of the fit are: $B_{(s)}^0 \rightarrow h^+h^-$ (green), semileptonic (cyan), and combinatorial (red).

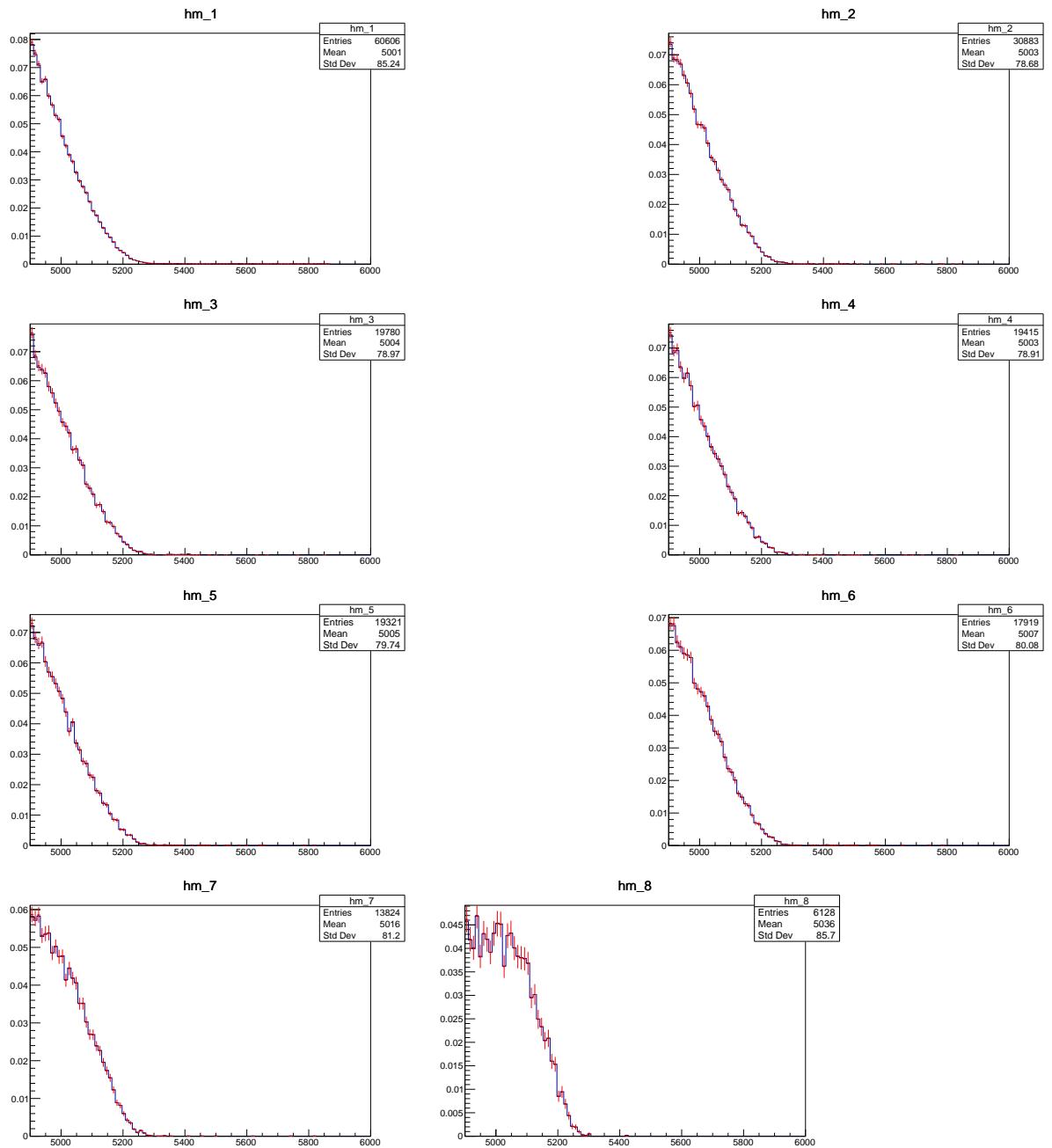


Figure 116: MC $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$ mass spectra in bins of BDT with (red) and without (blue) FF weights.

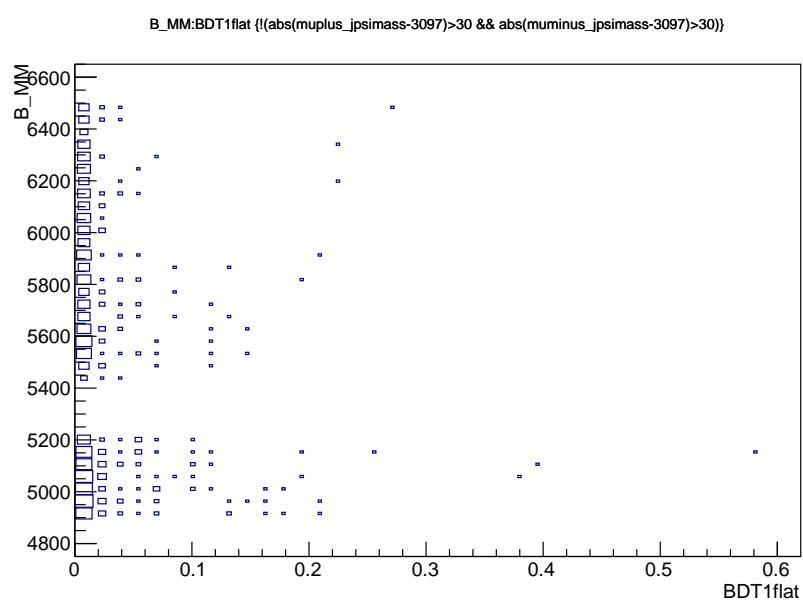


Figure 117: Dimuon invariant mass as a function of BDT for the events selected in Run1 but rejected by the J/ψ veto. The mass signal region is blinded.

2601 References

- 2602 [1] LHCb, CMS, 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.
- 2603 [2] A. J. Buras, J. Gиррбах, D. Guadagnoli, and G. Isidori, *On the Standard Model prediction for $BR(B_{s,d}^0 \rightarrow \mu^+ \mu^-)$* , Eur. Phys. J. **C72** (2012) 2172, arXiv:1208.0934.
- 2604 [3] A. J. Buras, R. Fleischer, J. Gиррбах, and R. Knegjens, *Probing New Physics with the $B^0 \rightarrow \mu^+ \mu^-$ Time-Dependent Rate*, JHEP **07** (2013) 77, arXiv:1303.3820.
- 2605 [4] K. D. Bruyn *et al.*, *Probing New Physics via the $B_s \rightarrow \mu^+ \mu^-$ Effective Lifetime*, Phys. Rev. Lett. **109** (2012) 041801, arXiv:1204.1737.
- 2606 [5] C. Bobeth *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.
- 2607 [6] C. Adrover *et al.*, *Search for the $B_S^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ decays with 3 fb^{-1} at LHCb*, LHCb-ANA-2013-032.
- 2608 [7] Heavy Flavor Averaging Group, Y. Amhis *et al.*, *Averages of b-hadron, c-hadron, and τ -lepton properties as of early 2012*, arXiv:1207.1158, updated results and plots available at: <http://www.slac.stanford.edu/xorg/hfag/>.
- 2609 [8] M. Pivk and F. R. Le Diberder, *sPlot: a statistical tool to unfold data distributions*, Nucl. Instrum. Meth. **A555** (2005) 356, arXiv:physics/0402083.
- 2610 [9] H. Evans, H. Cliff, and M.-O. Bettler, *Lines for $B_{s,d}^0 \rightarrow \mu^+ \mu^-$ in Stripping 21*, Tech. Rep. LHCb-INT-2015-008. CERN-LHCb-INT-2015-008, CERN, Geneva, Mar, 2015.
- 2611 [10] C. Adrover *et al.*, *Search for the rare decays $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ with 1.02 fb^{-1} at LHCb*, LHCb-ANA-2011-102.
- 2612 [11] F. Archilli *et al.*, *Performance of the muon identification at LHCb*, JINST **8** (2013) P10020, arXiv:1306.0249.
- 2613 [12] C.R.Jones, *ANNPID variables*, <https://indico.cern.ch/event/226062/contribution/1>.
- 2614 [13] F. Archilli *et al.*, *Background studies for $B^0 \rightarrow \mu^+ \mu^-$ analysis optimization*, Tech. Rep. LHCb-INT-2014-047. CERN-LHCb-INT-2014-047, CERN, Geneva, Nov, 2014.
- 2615 [14] L. Anderlini *et al.*, *The PIDCalib package*, Tech. Rep. LHCb-PUB-2016-021. CERN-LHCb-PUB-2016-021, CERN, Geneva, Jul, 2016.
- 2616 [15] O. Lupton, L. Anderlini, B. Sciascia, and V. Gligorov, *Calibration samples for particle identification at LHCb in Run 2*, Tech. Rep. LHCb-PUB-2016-005. CERN-LHCb-PUB-2016-005, CERN, Geneva, Mar, 2016.

- 2634 [16] Particle Data Group, K. A. Olive *et al.*, *Review of Particle Physics*, Chin. Phys. **C38**
 2635 (2014) 090001.
- 2636 [17] R. Aaij *et al.*, *Updated average f_s/f_d b-hadron production fraction ratio for 7 TeV pp*
 2637 *collisions*, LHCb-CONF-2013-011.
- 2638 [18] ARGUS, H. Albrecht *et al.*, *Search for hadronic $b \rightarrow u$ decays*, Phys. Lett. **B241** (1990)
 2639 278.
- 2640 [19] Particle Data Group, J. Beringer *et al.*, *Review of particle physics*, Phys. Rev. **D86**
 2641 (2012) 010001.
- 2642 [20] J. M. Flynn *et al.*, *$B \rightarrow \pi \ell \nu$ and $B_s \rightarrow K \ell \nu$ form factors and $|V_{ub}|$ from 2+1-flavor*
 2643 *lattice QCD with domain-wall light quarks and relativistic heavy quarks*, Phys. Rev.
 2644 **D91** (2015), no. 7 074510, [arXiv:1501.0537](https://arxiv.org/abs/1501.0537).
- 2645 [21] C. M. Bouchard *et al.*, *$B_s \rightarrow K \ell \nu$ form factors from lattice QCD*, Phys. Rev. **D90**
 2646 (2014) 054506, [arXiv:1406.2279](https://arxiv.org/abs/1406.2279).
- 2647 [22] S. Aoki *et al.*, *Review of lattice results concerning low-energy particle physics*,
 2648 [arXiv:1607.0029](https://arxiv.org/abs/1607.0029).
- 2649 [23] LHCb, R. Aaij *et al.*, *First measurement of the differential branching fraction and CP*
 2650 *asymmetry of the $B^\pm \rightarrow \pi^\pm \mu^+ \mu^-$ decay*, JHEP **10** (2015) 034, [arXiv:1509.0041](https://arxiv.org/abs/1509.0041).
- 2651 [24] W.-F. Wang and Z.-J. Xiao, *The semileptonic decays $B/B_s \rightarrow (\pi, K)(\ell^+ \ell^-, \ell \nu, \nu \bar{\nu})$*
 2652 *in the perturbative QCD approach beyond the leading-order*, Phys. Rev. **D86** (2012)
 2653 114025, [arXiv:1207.0265](https://arxiv.org/abs/1207.0265).
- 2654 [25] LHCb, R. Aaij *et al.*, *Determination of the quark coupling strength $|V_{ub}|$ using baryonic*
 2655 *decays*, Nature Phys. **11** (2015) 743, [arXiv:1504.0156](https://arxiv.org/abs/1504.0156).
- 2656 [26] LHCb collaboration, R. Aaij *et al.*, *Measurement of b hadron production fractions in*
 2657 *7 TeV pp collisions*, Phys. Rev. **D85** (2012) 032008, [arXiv:1111.2357](https://arxiv.org/abs/1111.2357).
- 2658 [27] LHCb, R. Aaij *et al.*, *Measurements of B_c^+ production and mass with the $B_c^+ \rightarrow J/\psi \pi^+$*
 2659 *decay*, Phys. Rev. Lett. **109** (2012) 232001, [arXiv:1209.5634](https://arxiv.org/abs/1209.5634).
- 2660 [28] LHCb, R. Aaij *et al.*, *Measurement of the ratio of B_c^+ branching fractions to $J/\psi \pi^+$*
 2661 *and $J/\psi \mu^+ \nu_\mu$* , Phys. Rev. **D90** (2014), no. 3 032009, [arXiv:1407.2126](https://arxiv.org/abs/1407.2126).
- 2662 [29] M. Mancinelli and J. Serrano, *Study of Muon Isolation in the $B_s^0 \rightarrow \mu^+ \mu^-$ Channel*,
 2663 LHCb-INT-2010-011.
- 2664 [30] M. e. a. Rama, *BDT studies for $B \rightarrow \mu^+ \mu^-$* , March, 2014. Very Rare Decays meeting,
 2665 <https://indico.cern.ch/event/291729/>.

- 2666 [31] A. Mordà, *Rare dileptonic $B0(s)$ meson decays at LHCb*, PhD thesis, Marseille, CPPM,
2667 2015.
- 2668 [32] X. Cid Vidal, *Using jets to identify semi-leptonic B decays*, LHCb-INT-2015-022.
- 2669 [33] A. Carbone, F. Ferrari, S. Perazzini, and V. Vagnoni, *Search for the $B^0 \rightarrow K^+K^-$*
2670 *decay with $\int \mathcal{L} dt = 3 fb^{-1}$* .
- 2671 [34] Belle, F. Thorne *et al.*, *Measurement of the decays $B_s^0 \rightarrow J/\psi\phi(1020)$, $B_s^0 \rightarrow$*
2672 *$J/\psi f'_2(1525)$ and $B_s^0 \rightarrow J/\psi K^+K^-$ at Belle*, Phys. Rev. **D88** (2013), no. 11 114006,
2673 [arXiv:1309.0704](https://arxiv.org/abs/1309.0704).
- 2674 [35] S. Tolk, *Discovery Of Rare B decays. First observation of the $B_s \rightarrow \mu^+\mu^-$ decays, first*
2675 *evidence of the $B^0 \rightarrow \mu^+\mu^-$ decays.*, PhD thesis, Groningen U., Dec, 2015, Presented
2676 08 Apr 2016.
- 2677 [36] LHCb collaboration, R. Aaij *et al.*, *Measurements of the branching fractions and CP*
2678 *asymmetries of $B^+ \rightarrow J/\psi\pi^+$ and $B^+ \rightarrow \psi(2S)\pi^+$ decays*, Phys. Rev. (2012) 091105,
2679 [arXiv:1203.3592](https://arxiv.org/abs/1203.3592).
- 2680 [37] LHCb, R. Aaij *et al.*, *Measurement of the track reconstruction efficiency at LHCb*,
2681 JINST **10** (2015), no. 02 P02007, [arXiv:1408.1251](https://arxiv.org/abs/1408.1251).
- 2682 [38] LHCb, R. Aaij *et al.*, *Measurement of CP violation and the B_s^0 meson decay width*
2683 *difference with $B_s^0 \rightarrow J/\psi K^+K^-$ and $B_s^0 \rightarrow J/\psi\pi^+\pi^-$ decays*, Phys. Rev. **D87** (2013),
2684 no. 11 112010, [arXiv:1304.2600](https://arxiv.org/abs/1304.2600).
- 2685 [39] LHCb, R. Aaij *et al.*, *Determination of the sign of the decay width difference in the*
2686 *B_s system*, Phys. Rev. Lett. **108** (2012) 241801, [arXiv:1202.4717](https://arxiv.org/abs/1202.4717).
- 2687 [40] M. Perrin-Terrin, *Searches for B meson decays to purely leptonic final states*, PhD
2688 thesis, Aix-Marseille U., Jul, 2013, Presented 23 Sep 2013.
- 2689 [41] S. Tolk, J. Albrecht, and F. Dettori, *Measuring the trigger efficiency in data with the*
2690 *tistos method*, LHCb-INT-2013-038.
- 2691 [42] G. Cowan *et al.*, *Measurement of b quark production in 7 and 13TeV pp collisions*, .
- 2692 [43] T. Skwarnicki, *A study of the radiative cascade transitions between the Upsilon-prime*
2693 *and Upsilon resonances*, PhD thesis, Institute of Nuclear Physics, Krakow, 1986,
2694 DESY-F31-86-02.
- 2695 [44] Y. Xie, *sfit: concept, implementation and possible pitfalls*, January, 2013.
2696 <https://indico.cern.ch/event/201535/>.
- 2697 [45] LHCb, R. Aaij *et al.*, *Measurement of the $\bar{B}^0 - B^0$ and $\bar{B}_s^0 - B_s^0$ production asymmetries*
2698 *in pp collisions at $\sqrt{s} = 7$ TeV*, Phys. Lett. **B739** (2014) 218, [arXiv:1408.0275](https://arxiv.org/abs/1408.0275).

- 2699 [46] LHCb collaboration, R. Aaij *et al.*, *Effective lifetime measurements in the $B_s^0 \rightarrow$*
2700 *K^+K^- , $B^0 \rightarrow K^+\pi^-$ and $B_s^0 \rightarrow \pi^+K^-$ decays*, Phys. Lett. B **736** (2014) 446.
- 2701 [47] LHCb, R. Aaij *et al.*, *First study of the CP -violating phase and decay-width difference*
2702 *in $B_s^0 \rightarrow \psi(2S)\phi$ decays*, Phys. Lett. B**762** (2016) 253, [arXiv:1608.0485](https://arxiv.org/abs/1608.0485).
- 2703 [48] LHCb, R. Aaij *et al.*, *Measurement of CP violation in $B^0 \rightarrow J/\psi K_S^0$ decays*, Phys.
2704 Rev. Lett. **115** (2015), no. 3 031601, [arXiv:1503.0708](https://arxiv.org/abs/1503.0708).