



Search for the rare decays $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ with 1.02 fb $^{-1}$ at LHCb

C. Adrover¹, J. Albrecht², F. Archilli⁶, M.-O. Bettler^{2,3}, X. Cid Vidal⁵, F. Dettori⁸,
C. Elsasser⁴, J. A. Hernando Morata⁵, G. Lanfranchi⁶, G. Mancinelli¹,
D. Martínez Santos², S. Oggero⁸, M. Palutan⁶, M. Perrin-Terrin¹, N. Sagidova⁷,
A. Sarti⁶, J. Serrano¹, B. Sciascia⁶, Y. Shcheglov⁷, F. Soomro⁶, O. Steinkamp⁴, S. Tolk⁸

¹ CPPM, Aix-Marseille Université, CNRS/IN2P3, Marseille, France.

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

³ Sezione INFN di Firenze, Firenze, Italy.

⁴ Universität Zürich, Zurich, Switzerland.

⁵ Universidad de Santiago de Compostela, Santiago de Compostela, Spain.

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

⁷ Petersburg Nuclear Physics Institute (PNPI), Gatchina, Russia.

⁸ Nikhef, National Institute for Subatomic Physics, Amsterdam, Netherlands.

Abstract

A search for the $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ decays is performed in 1 fb $^{-1}$ of pp collisions at $\sqrt{s} = 7$ TeV collected by the LHCb experiment at the Large Hadron Collider at CERN. For both decays the number of observed events is consistent with expectation from background and standard model signal predictions. The upper limits on their branching fractions are measured to be $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) < 4.5 (3.8) \times 10^{-9}$ and $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) < 10.3 (8.1) \times 10^{-10}$ at 95 % (90 %) confidence level.

¹ Contents

²	1	Introduction	3
³	2	Analysis strategy	3
⁴	3	Data and Monte Carlo Samples	5
⁵	3.1	Data Sample	5
⁶	3.2	Monte Carlo samples	6
⁷	4	Trigger	8
⁸	4.1	Level-0 trigger	10
⁹	4.2	High Level Trigger	11
¹⁰	5	Selection	12
¹¹	5.1	The BDTS discriminant	16
¹²	6	Muon Identification	21
¹³	6.1	MuonID efficiency	21
¹⁴	6.2	Muon hypothesis test (DLL)	27
¹⁵	7	Misidentification rates and background composition	29
¹⁶	7.1	Pion and kaon misidentification probabilities	29
¹⁷	7.2	Double misidentification probability	39
¹⁸	7.3	Peaking background from $B_{(s)}^0 \rightarrow h^+h'^-$ with double misentification: yield, BDT and mass PDFs	40
¹⁹	7.3.1	Relation between the single misidentification probability and double misidentification probability for $B_{(s)}^0 \rightarrow h^+h'^-$ peaking background .	43
²⁰	7.4	Other exclusive backgrounds	46
²¹	7.4.1	Background from $B_c^+ \rightarrow J/\psi(\rightarrow \mu^+\mu^-)\mu^+\nu$	46
²²	7.4.2	Background from $B^+ \rightarrow \pi^+\mu^+\mu^-$	48
²³	7.4.3	Background from $B_s \rightarrow \mu^+\mu^-\gamma$	49
²⁴	8	The Boosted Decision Tree	53
²⁵	8.1	BDT definition and performance	53
²⁶	8.2	Binning of the BDT and Invariant Mass distributions	58
²⁷	9	Calibration of the BDT and the invariant mass	65
²⁸	9.1	Extraction of the BDT for signal with the $B_{(s)}^0 \rightarrow h^+h'^-$ inclusive sample .	65
²⁹	9.2	Invariant mass central values for signal with the $B_{(s)}^0 \rightarrow h^+h'^-$ exclusive samples	78
³⁰	9.3	Invariant mass resolution for signal with the $B_{(s)}^0 \rightarrow h^+h'^-$ exclusive samples	79
³¹	9.4	Calibration of the BDT for signal with the $B_{(s)}^0 \rightarrow h^+h'^-$ exclusive samples	81
³²	9.5	Calibration of the invariant mass for signal using the interpolation method	85

36	9.6 BDT and invariant mass PDFs for combinatorial background	88
37	10 Normalization	93
38	10.1 Ratio of production fractions	93
39	10.2 Branching ratios of the control channels	93
40	10.3 The ratio of reconstruction efficiencies	94
41	10.3.1 Reconstruction efficiencies from MC simulation	94
42	10.3.2 Validation of the ratio of reconstruction efficiencies using data . . .	95
43	10.3.3 Muon detector acceptance and muonID efficiency	96
44	10.4 Ratio of selection efficiencies	99
45	10.4.1 Track parameter smearing	100
46	10.4.2 Selection efficiencies from simulation	100
47	10.4.3 PID efficiency determination	101
48	10.4.4 systematic uncertainties	101
49	10.4.5 Ratio of selection efficiencies	102
50	10.5 Ratio of trigger efficiencies	103
51	10.5.1 Estimation of the trigger efficiencies	103
52	10.5.2 Ratio of trigger efficiency for $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ and $B \rightarrow J/\psi X$ channels	104
53	10.5.3 Global event cuts	105
54	10.6 Number of candidates	106
55	10.6.1 Normalization to $B^+ \rightarrow J/\psi K^+$	106
56	10.6.2 Normalization to $B_s^0 \rightarrow J/\psi (\mu^+ \mu^-) \phi(KK)$	106
57	10.6.3 Normalization to $B^0 \rightarrow K^+ \pi^-$ or $B_{(s)}^0 \rightarrow h^+ h'^-$ TIS	107
58	10.7 Normalization factor	108
59	11 Results: upper limits	109
60	12 Methods for BR evaluation	118
61	12.1 Binned methods	118
62	12.2 Simultaneous fit of mass projections in bins of BDT	119
63	12.3 Two-dimensional unbinned fit	119
64	13 BR results	124
65	13.1 Simultaneous fit of mass projections in bins of BDT	124
66	13.2 Two-dimensional unbinned fit	126
67	14 Conclusions	131
68	A Further cross-checks on signal BDT calibration	132
69	B Appendix B: RICH Calibration	136

70 1 Introduction

71 Measurements of low-energy processes can provide indirect constraints on particles that
72 are too heavy to be produced directly. This is particularly true for Flavour Changing
73 Neutral Current (FCNC) processes which are highly suppressed in the Standard Model
74 (SM) and can only occur through higher-order diagrams. The SM predictions for the
75 branching fractions of the FCNC decays¹ $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ are $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$
76 = $(3.2 \pm 0.2) \times 10^{-9}$ and $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ = $(0.10 \pm 0.01) \times 10^{-9}$ [2]. However, contributions
77 from new processes or new heavy particles can significantly enhance these values. For
78 example, within Minimal Supersymmetric extensions of the SM (MSSM), in the large
79 $\tan \beta$ regime, $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ is found to be approximately proportional to $\tan^6 \beta$ [3],
80 where $\tan \beta$ is the ratio of the vacuum expectation values of the two neutral CP -even
81 Higgs fields. The branching fractions could therefore be enhanced by orders of magnitude
82 for large values of $\tan \beta$.

83 The best published limits from the Tevatron are $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) < 5.1 \times 10^{-8}$
84 at 95% confidence level (CL) by the D0 collaboration using 6.1 fb^{-1} of data [4], and
85 $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) < 6.0 \times 10^{-9}$ at 95% C.L. by the CDF collaboration using 6.9 fb^{-1} of
86 data [5]. In the same dataset the CDF collaboration observes an excess of $B_s^0 \rightarrow \mu^+ \mu^-$
87 candidates compatible with $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (1.8_{-0.9}^{+1.1}) \times 10^{-8}$ and with an upper limit
88 of $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) < 4.0 \times 10^{-8}$ at 95% C.L. . The CMS collaboration has recently
89 published $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) < 1.9 \times 10^{-8}$ at 95% C.L. and $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) < 4.6 \times 10^{-9}$ at
90 95% C.L. using 1.14 fb^{-1} of data [6]. The LHCb collaboration has published the limits
91 [7] $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) < 1.6 \times 10^{-8}$ and $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) < 3.6 \times 10^{-8}$ at 95% C.L. based
92 on about 370 pb^{-1} of integrated luminosity collected in the first half of the 2011 run.
93 A combination of these results with the LHCb limits obtained with the 2010 dataset [8]
94 leads to $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) < 1.4 \times 10^{-8}$ and $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) < 3.2 \times 10^{-9}$ at 95% C.L. :
95 these are the world best limits to date.

96 This document presents an analysis of the data recorded by LHCb in 2011 which
97 correspond to an integrated luminosity of $\sim 1.02 \text{ fb}^{-1}$. This dataset includes the 370
98 pb^{-1} used for the publication of the previous limit [7].

99 2 Analysis strategy

100 Assuming the branching fractions predicted by the SM, and using the $b\bar{b}$ cross-section
101 measured by LHCb in the pseudorapidity interval $2 < \eta < 6$ and integrated over all
102 transverse momenta of $\sigma_{b\bar{b}} = 75 \pm 14 \mu\text{b}$ [11], approximately $11.6 B_s^0 \rightarrow \mu^+ \mu^-$ and 1.3
103 $B^0 \rightarrow \mu^+ \mu^-$ events are expected to be reconstructed and selected in the analysed sample
104 embedded in a large background.

105 The general structure of the analysis is based upon the one described in Ref. [7]. First
106 a very efficient selection removes most of the background while keeping very high efficiency

¹Inclusion of charged conjugated processes is implied throughout.

for signals. The number of observed events then is compared to the number of expected signal and background events in bins of two independent variables, the invariant mass and the output of a multi-variate discriminant, the Boosted Decision Tree (BDT) constructed using the TMVA package [9].

The current analysis has been improved in several points: the selection has been tightened in order to cope with the increased dataset, the binning used for the limit computation has been fully revised and DLL cuts have been introduced in order to reduce the peaking background coming from $B_{(s)}^0 \rightarrow h^+ h'^-$ events with double decays in flight.

The probability for a signal or background event to have a given value of the BDT output is extracted from data using $B_{(s)}^0 \rightarrow h^+ h'^-$ candidates (where $h^{(\prime)}$ can be a pion or a kaon) as signal and sideband $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ candidates as background.

The invariant mass line shape of the signals is described by a Crystal Ball function [10] whose parameters are extracted from data using control samples. The central values of the masses are obtained from $B^0 \rightarrow K^+ \pi^-$ and $B_s^0 \rightarrow K^+ K^-$ samples. The B_s^0 and B^0 mass resolutions are estimated by interpolating the resolution measured on di-muon resonances (J/ψ , $\psi(2S)$ and $\Upsilon(1S, 2S, 3S)$) and cross-checked with a fit to the invariant mass distributions of exclusive $B^0 \rightarrow K^+ \pi^-$ and $B_s^0 \rightarrow K^+ K^-$ decays. The parameter of the radiative tails have been extracted from a fit to the mass distribution of $B_s^0 \rightarrow \mu^+ \mu^-$ simulated events where the resolution has been smeared in order to reproduce the measured one. The central values and the resolutions of the B^0 and B_s^0 masses are used to define the search windows.

The number of expected signal events, for a given branching fraction hypothesis, is obtained by normalizing to channels of known branching fractions: $B^+ \rightarrow J/\psi K^+$, $B_s^0 \rightarrow J/\psi \phi$ and $B^0 \rightarrow K^+ \pi^-$. These channels are selected in a way as similar as possible to the signals in order to minimize the systematic uncertainty related to the different phase space accessible to each final state.

The BDT output and invariant mass distributions for combinatorial background events in the signal regions are obtained using fits of the mass distribution of events in the mass sidebands in bins of the BDT output.

The two-dimensional space formed by the invariant mass and the BDT output is binned. For each bin we count the number of candidates observed in the data, and compute the expected number of signal events and the expected number of background events. The compatibility of the observed distribution of events in all bins with the distribution expected for a given branching fraction hypothesis is computed using the CL_s method [25], which allows a given hypothesis to be excluded at a given confidence level.

In order to avoid unconscious bias the mass region $m_{\mu\mu} = [m(B^0) - 60 \text{ MeV}/c^2, m(B_s^0) + 60 \text{ MeV}/c^2]$ has been blinded until the completion of the analysis².

²The first 370 pb^{-1} have been already used to produce the result published in [7], so strictly speaking, only the remaining 632 pb^{-1} have been blinded.

3 Data and Monte Carlo Samples

3.1 Data Sample

The results described in this note are obtained using the data collected by LHCb at the LHC at a center-of-mass energy of $\sqrt{s} = 7$ TeV between March and October 2011. This dataset corresponds to $\sim 1.02 \text{ fb}^{-1}$ of integrated luminosity.

The LHC machine started the operations in March by reaching very soon a peak luminosity $L \sim 1.6 \times 10^{32} \text{ cm}^{-2} \text{ s}^{-1}$ with 228 bunches (180 bunches colliding in LHCb) and an average number of pp visible interactions per crossing of $\mu \sim 2.5$. After the first 10 pb^{-1} collected by LHCb, the machine moved to the 50 ns bunch scheme and kept increasing the number of bunches by 144 every three fills, by reaching 1380 circulating bunches (1296 colliding bunches in LHCb) beginning of July. Since then the peak luminosity in LHCb was continuously leveled in order not to exceed $3 - 3.5 \times 10^{32} \text{ cm}^{-2} \text{ s}^{-1}$ corresponding to an average $\langle \mu \rangle \sim 1.5$.

The considered data set was taken with a magnetic field both with positive ($B_y > 0$) ($\sim 446 \text{ pb}^{-1}$) and negative ($B_y < 0$) polarity ($\sim 595 \text{ pb}^{-1}$).

In Fig. 1 we show the integrated luminosity and the average value of proton-proton interactions per crossing (μ) as a function of the fill number for the period of data taking considered in this analysis.

The data belong to the Reco12-Stripping17 reconstruction campaign, have been reconstructed with Brunel v41r1 [26], the Condition Data Base [27] (condDB) head-20111111 and the Detector Data Base (DDDb) [28] head-20110914 and analyzed with DaVinci v29r2 [29].

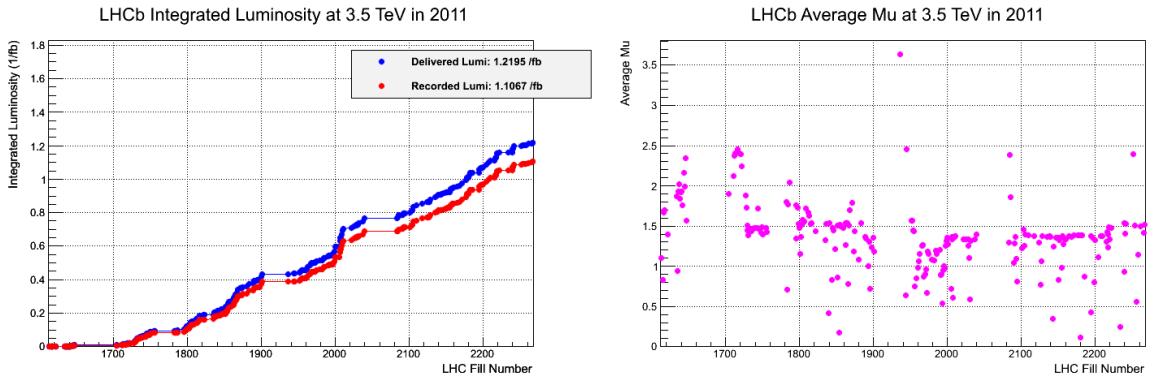


Figure 1: Integrated luminosity (left) and average number of pp -interactions per crossing (right) as a function of the fill number.

167 **3.2 Monte Carlo samples**

168 The Monte Carlo samples used in this analysis belong to the old Monte Carlo production
169 MC10, sim01. These samples In Table 1 the Monte Carlo samples are listed and here
170 below a brief reminder of the generation/reconstruction conditions is reported.

171 The pp interactions have been simulated assuming a beam energy of 3.5 TeV, an average
172 number of interactions per crossing $\nu = 2.5$, which corresponds to an average number of
173 visible interactions per crossing $\mu = 1.75$, and a $\beta^* = 3.5$ m.

174 These samples have been generated using the GAUDI [30] application GAUSS v39r0
175 [31], based on LHCb v26r3, Pythia 6_418.2 [32], PHOTOS 215.2 [33], LHAPDF 5.3.1 [34],
176 HepMC 1.26 [35], and GEANT4 v91r3.p03 [36]. Then they have been reconstructed using
177 Boole v21r9 and Brunel v37r8p5.

178 To save CPU time, cuts are applied at the generation level to enforce particles of
179 interest to be between 10 mrad and 400 mrad. The particles that fulfill this criteria are:
180 the two muons of the signal samples, the two muons of $B_s^0 \rightarrow J/\psi\phi$, $B^+ \rightarrow J/\psi K^+$ and
181 $B_d^0 \rightarrow J/\psi K^{*0}$ control channels and the two hadrons of $B_{(s)}^0 \rightarrow h^+h^-$ decays. For the
182 $b\bar{b} \rightarrow \mu\mu X$ background sample, two muons are required to be in the acceptance regardless
183 of their charge. The generator level efficiencies due to the acceptance cuts are shown
184 in Table 2 [39]. The total number of $b\bar{b} \rightarrow \mu\mu X$ events corresponds to $\sim 570 \text{ pb}^{-1}$ of
185 integrated luminosity, assuming the $b\bar{b}$ cross section measured by LHCb, $\sigma_{b\bar{b}} = 288 \pm 4 \pm 48$
186 μb [40].

187 The trigger has been emulated in *pass-through mode* with the configuration that has
188 been more extensively used during the 2011 data taking, *Trigger Configuration Key* (TCK)
189 0x2d0032 (see Sect. 4). In addition, the same stripping selections (stripping17, see Sect. 5)
190 used for data have been applied. This has been possible as the stripping selections used
191 in 2011 were tighter than the ones used to strip the MC10 samples.

192 A smearing procedure to the first state of the tracks in simulated events has been
193 applied in order to reproduce the impact parameter resolution measured in data. This
194 will be discussed in Section 10.4.1.

Table 1: Monte Carlo samples used in this analysis. The Monte Carlo production version, the magnet polarity, the number of processed events and the number of reconstructed and stripped events are shown in the second, third, fourth and fifth columns, respectively.

Channel	Monte Carlo production	Magnet polarity	Events (processed)	Events (reco and stripped)
Signal: $B_s^0 \rightarrow \mu^+ \mu^-$	MC10sim01	Down / Up	~500 k	~ 180 k
Background: $b\bar{b} \rightarrow \mu\mu, p > 3 \text{ GeV}/c, M_{\mu\mu} > 4.7 \text{ GeV}/c^2$	MC10sim01	Down	50 M	57 k
$b\bar{b} \rightarrow \mu\mu, p > 3 \text{ GeV}/c, M_{\mu\mu} > 4.7 \text{ GeV}/c^2$	MC10sim01	Up	50 M	57 k
Control/normalization channels: $B_s^0 \rightarrow J/\psi \phi$ $B^+ \rightarrow J/\psi K^+$ $B_d^0 \rightarrow J/\psi K^{*0}$ $B^0 \rightarrow K^+ \pi^-$	MC10sim01 MC10sim01 MC10sim01 MC10sim01	Down/Up Down/Up Down/Up Down/Up	— — — —	200 k 365 k 200 k 300 k

Table 2: Generator level efficiency ϵ_{gen} for signals, control channels and background channel. The uncertainty on the quoted numbers is below 0.1 %.

	$B_{(s)}^0 \rightarrow \mu^+ \mu^-$	$B^+ \rightarrow J/\psi K^+$	$B_s^0 \rightarrow J/\psi \phi$	$B_{(s)}^0 \rightarrow h^+ h'^-$	$B_d^0 \rightarrow J/\psi K^{*0}$	$b\bar{b} \rightarrow \mu\mu X$
ϵ_{gen}	17.5%	15.3%	16.0%	17.7%	14.8%	6.1×10^{-4}

195 **4 Trigger**

196 The LHCb trigger scheme is described in [41]. A detailed description of the trigger lines
 197 relevant for the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis is reported in [15]. In this Section we summarize
 198 the main characteristics of the trigger used between March and October 2011.

199 During the 2011 run, the luminosity increased very quickly to the regime value of
 200 $L = 3 - 3.5 \times 10^{32} \text{ cm}^{-2} \text{ s}^{-1}$ and kept constant by continuous levelling (see Fig. 2): about
 201 1 fb^{-1} of data were collected in these conditions.

202 As a consequence of the stability of the peak luminosity, also the trigger thresholds
 203 were kept stable during the overall data taking period, with a L0-output rate exceeding
 204 800 kHz (Fig. 3, left) and a physics trigger HLT output rate of ~ 3 kHz (Fig. 3, right).

205 In LHCb the trigger configuration information is embedded into an hexadecimal word
 206 called *Trigger Configuration Key* (TCK) that identifies the set of trigger decisions, the
 207 algorithms run in the trigger and the cuts applied to trigger the events in a given run.
 208 Table 3 shows the different TCKs used for the 2011 data taking with the corresponding
 209 integrated luminosity and magnet polarity³.

210 Global event cuts (GEC) were introduced at L0, HLT1 and HLT2 level to reduce the
 211 processing time of the HLT. The GEC applied during 2011 run are based on cuts on
 212 the number of IT, OT and SPD hits and are listed in Table 4. From TCK=0x6d0032
 213 onwards, the 10 k OT hits threshold was replaced by a 15 k OT hits threshold with hits
 214 counted inside a time window of $[-8, +56]$ ns with respect to the collision time. This last
 215 change was necessary to cope with the increased number of OT hits in the 75 ns gate due
 216 to the presence of spillover hits during the machine operation with 50 ns bunch spacing
 217 with bunch trains. The SPD hits thresholds depend on the L0 trigger line and are listed
 218 in Table 5.

³The reasons of changing the TCK during the data taking were related to the adjustment of the express and charm lines, to the tests of K_S and electron topological lines, the tuning of the electron and low multiplicity lines.

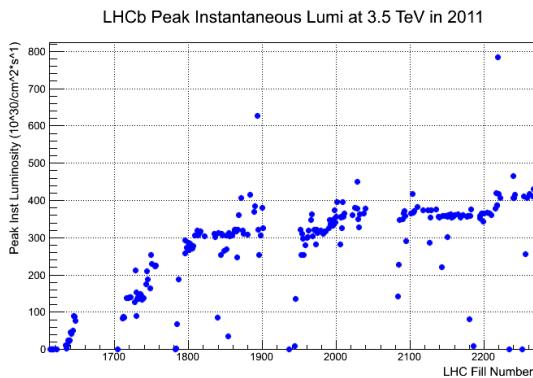


Figure 2: Peak luminosity as a function of the fill number.

Table 3: TCKs used for the 2011 data taking with the corresponding integrated luminosity and magnet polarity.

TCKs	Integrated Luminosity	Magnet Polarity
0x5a0032	39 pb ⁻¹	Up
0x5a0032	29 pb ⁻¹	Down
0x6d0032	100 pb ⁻¹	Down
0x730035	196 pb ⁻¹	134 pb ⁻¹ Up / 62 pb ⁻¹ Down
0x740036	5.2 pb ⁻¹	5.2 pb ⁻¹ Up
0x760037	298.7 pb ⁻¹	107.1 pb ⁻¹ Up / 191.6 pb ⁻¹ Down
0x790037	39.3 pb ⁻¹	39.3 pb ⁻¹ Up
0x790038	363.4 pb ⁻¹	154 pb ⁻¹ Up / 209.4 pb ⁻¹ Down

219 The change in the threshold of the number of OT hits was the only relevant change
 220 in the TCKs used in 2011 run, as far as the trigger lines relevant for the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$
 221 analysis are concerned.

Table 4: Summary of global event cuts (GEC) used in 2011.

trigger level	cut	value
L0	SPD hits	900/600
HLT1	Velo hits	10 k
HLT1	IT hits	3000
HLT1	OT hits	10k / 15 k
HLT2	Velo tracks	350

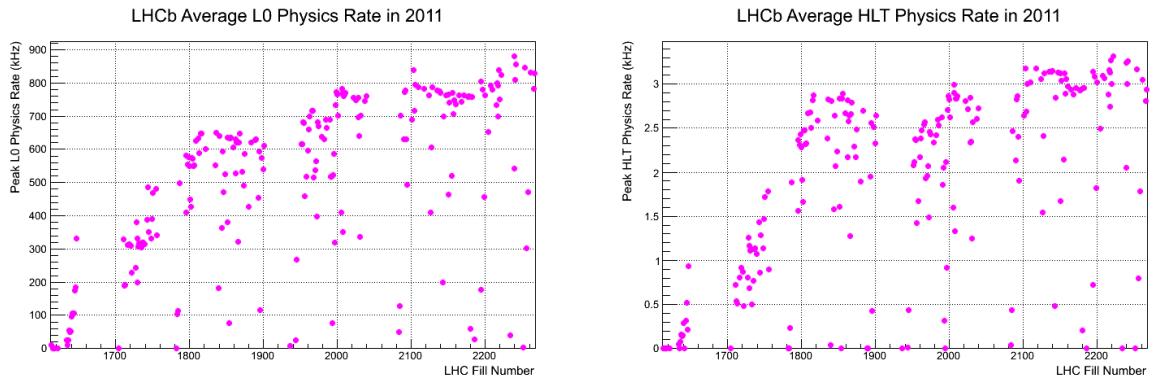


Figure 3: L0 trigger output rate (left) and HLT physics trigger output rate (right) as a function of the fill number.

Table 5: Thresholds of the L0 trigger lines used to trigger the 2011 dataset (TCK=6d0032).

L0 line	TCK = 6d0032
single muon (p_T) (GeV/c) / SPD	1.5/600
di-muon / SPD	$\sqrt{p_{T1} \times p_{T2}} = 1.3/900$
hadron(E_T) (GeV) / SPD	3.5/600

4.1 Level-0 trigger

The L0 trigger is the logical OR of several decisions and the veto of GEC. The decisions are based on the signal measured in several detectors that are associated to individual particles, while the GEC are based on the signatures of the event itself. The L0 decisions (single muon, dimuon, hadron, photon and electron) are described in detail in [41].

In Table 5 the thresholds for the L0 single-muon, di-muon and hadron triggers are shown with the corresponding cut in the SPD multiplicity. All these trigger decisions are relevant for the analysis. The signal and the control channels with $J/\psi(\mu\mu)$ in the final state are mostly triggered by the L0- μ and L0-di μ decisions, while the $B_{(s)}^0 \rightarrow h^+h^-$ channels are triggered mostly by the L0-hadron.

232 4.2 High Level Trigger

233 Events triggered by the L0 trigger are then passed to the EFF where they are further
 234 filtered by the HLT1. First the HLT1 imposes a GEC on the number of hits in the OT
 235 detector, to reject high multiplicity events. Secondly it runs a partial reconstruction to
 236 select one or two tracks with high p_T and IP, on which the decisions are taken.

237 The HLT1 main trigger lines have been described in [15] and are not repeated here.
 238 The most important selection criteria for the present analysis applied at the HLT1 stage
 239 are summarized in Table 6.

Table 6: HLT1 threshold values for the TCK used in 2011 run. Only selected trigger cuts are given.

trigger line	TCK=6d0032
Hlt1SingleMuonNoIP (p_T (GeV/c) / prescaling factor)	1.8 /0.01
Hlt1DiMuon (high mass) (p_T (GeV/c) / $M_{\mu\mu}$ (GeV/ c^2))	0.5/2.5
Hlt1DiMuon (low mass) (p_T (GeV/c) / $M_{\mu\mu}$ (GeV/ c^2) / IP χ^2)	0.5/1/9
Hlt1TrackMuon (p_T / IP (mm) / IP χ^2)	1/0.1/16
Hlt1AllTrack (p_T / IP (mm) / IP χ^2)	1.85/0.1/16

240 Events accepted by the HLT1 are then fully reconstructed with track finding and
 241 fitting algorithms very close to those run offline. As in the 2010 run, a minimal trans-
 242 verse momentum of 500 MeV/c is required for all tracks at reconstruction level for timing
 243 reasons.

244 The main HLT2 trigger for the channels with dimuon in the final state are described
 245 in [22]. The channels containing J/ψ in the final state have been triggered in the first
 246 370 pb $^{-1}$ by the *Hlt2DiMuonJPsi* trigger, which requires two identified muons in a
 247 120 MeV/ c^2 mass window around the J/Ψ mass. This line was then prescaled by a
 248 factor five for the remaining 650 pb $^{-1}$, and the channels containing J/ψ in the final state
 249 were mostly triggered by the *Hlt2DiMuonDetached* line, where a soft cut on the distance
 250 between the secondary and primary vertex divided by its error ($DLS > 3$) has been
 251 added to the previous selection criteria. Therefore the trigger efficiency map used for the
 252 current analysis has been obtained from the J/ψ of the exclusive $B^+ \rightarrow J/\psi K^+$ decay
 253 (Section 10.5.1).

254 The $B_{(s)}^0 \rightarrow \mu^+\mu^-$ signal candidates are selected with an equivalent trigger selection,
 255 *Hlt2DiMuonBmm* but requiring their invariant mass to be above 4.7 GeV/ c^2 . Some extra
 256 J/ψ and $B_{(s)}^0 \rightarrow \mu^+\mu^-$ events are selected by the *topological muon* lines, where only one
 257 muon is requested and combined with a second track.

258 The $B_{(s)}^0 \rightarrow h^+h'^-$ hadronic channels used for calibration and/or normalization are
 259 selected by the *Hlt2Topo2Body* trigger, a generic $B \rightarrow hh + X$ selection and by the
 260 exclusive *Hlt2B2hh* selection [37].

261 5 Selection

262 The aim of the selection is twofold:

- 263 1. reduce the data size to a manageable level by keeping the efficiency on the signal as
 high as possible; the separation between signal and background is then left to the
 main discriminant, the Boosted Decision Tree (BDT) described in Section 8;
- 266 2. deal in similar way signal and control/normalization channels in order to minimize
 the systematic uncertainties in the computation of the normalization factors.

268 These two criteria have been respected for the current analysis, however few changes with
 269 respect to the selection used in the published analysis [7] have been introduced:

- 270 1. the common particles built with DaVinci versions from v29r0 onwards are required
 to have a track $\chi^2/\text{nDoF} < 4$, a transverse momentum $p_T > 250$ MeV/c and an impact
 parameter $\chi^2_{\text{IP}} > 4$. These cuts have been inherited by the particles used in this
 analysis and included in the selection (see Table 7). As in the published analysis, a
 cut on the Kullback-Leibler (KL) distance [38] is used to suppress duplicated tracks
 created by the reconstruction⁴.
- 276 2. The fiducial cuts applied to reject the unphysical signal candidates have been re-
 optimized in order to have an efficiency of $\epsilon = 99.9\%$ on Monte Carlo $B_s^0 \rightarrow \mu^+\mu^-$
 signal events. Decay products of a $B_{(s)}^0 \rightarrow \mu^+\mu^-$ or $B_{(s)}^0 \rightarrow h^+h'^-$ candidates are
 required to have $p < 500$ GeV/c and $p_T < 40$ GeV/c and the proper time of the
 B candidate is required to be $t < 9 \cdot \tau(B_s^0)$, being $\tau(B_s)$ the B_s^0 lifetime [59]. The
 distributions of the p and p_T of one of the two muons and the proper time of the B
 candidate are shown in Fig. 4 for $B_s^0 \rightarrow \mu^+\mu^-$ Monte Carlo events, $b\bar{b} \rightarrow \mu\mu X$ Monte
 Carlo events and background events in data mass sidebands ($[4.9 - 5.0]$ GeV/c² and
 $[5.433 - 6.0]$ GeV/c²).
- 285 The exclusive dimuon production in the process $pp \rightarrow p\mu^+\mu^-p$ is reduced with a
 cut on the transverse momentum of the B candidate $p_T(B) > 500$ MeV/c as in the
 published analysis⁵.
- 288 3. The selection developed for the 370 pb⁻¹ is too loose to reduce efficiently the
 dataset of about 1 fb⁻¹, dominated by the background of the $B_{(s)}^0 \rightarrow h^+h'^-$ inclusive

⁴This procedure compares the parameters and correlation matrices of the reconstructed tracks and where two are found to be similar, in this case with a symmetrized KL divergence less than 5000, only the one with the higher track fit quality is considered.

⁵For a detailed discussion about the exclusive dimuon production see [63], Section 5.1.

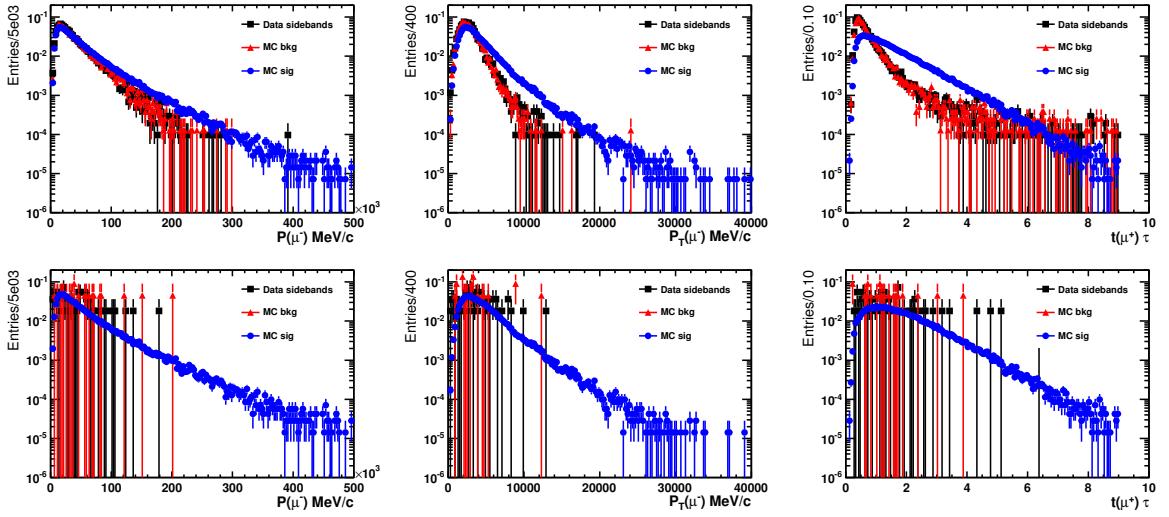


Figure 4: Muon momentum (left), transverse momentum (center) and B proper time (right) distributions for $B_s^0 \rightarrow \mu^+ \mu^-$ Monte Carlo events (blue curve), $b\bar{b} \rightarrow \mu\mu X$ Monte Carlo events (red curve) and $B_s^0 \rightarrow \mu^+ \mu^-$ candidates in data sidebands (black curve). Top: all the BDT range, bottom: only events with $\text{BDT} > 0.5$.

sample. In order to maximize the rejection power while keeping very high efficiency for the signal, a multivariate Boosted Decision Tree discriminant for the selection (BDTS) has been implemented and a loose cut on it has been applied. The definition of the BDTS discriminant is discussed in Section 5.1.

In Table 7 the cuts used to select $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ and $B_{(s)}^0 \rightarrow h^+ h^-$ inclusive decay are shown. The cuts that have been changed with respect to the last published analysis are in bold. The $B_{(s)}^0 \rightarrow h^+ h^-$ inclusive sample is selected as the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ signals (apart the muonID requirement) as it is the main control sample for the extraction of the BDT and the invariant mass PDFs from data. Moreover the exclusive $B^0 \rightarrow K^+ \pi^-$ channel is one of the three normalization channels. As in the published analysis, the signal regions ($\pm 60 \text{ MeV}/c^2$) around the measured B^0 and B_s^0 masses has been blinded until the completion of the analysis.

The two brand new cuts introduced with respect to the past are the BDTS cut and the DLL cuts, listed at the bottom of Table 7. The BDTS cut will be discussed in the following Section while the DLL cut will be discussed in Section 6.2.

For the $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi \phi$ normalization channels, the additional K^\pm candidates are required to pass the same track quality cut ($\chi^2/ndf < 4$). A $\text{IP}\chi^2$ cut is applied on the K^\pm of the $B^+ \rightarrow J/\psi K^+$ decay ($\text{IP}\chi^2 > 25$) and on the two kaons of the $B_s^0 \rightarrow J/\psi \phi$ decay ($\text{IP}\chi^2 > 4, 4$) (Table 8).

For $B_s^0 \rightarrow J/\psi \phi$ candidates, the KK invariant mass is required to be within $\pm 10 \text{ MeV}/c^2$ of the ϕ PDG mass and the K^\pm have to be identified as kaons by the RICH

311 system ($DLL(K - \pi) > 0$). For these normalization channels, the B candidate's IP χ^2
 312 and vertex separation (VDS) requirements are the same as those for the signal selection
 313 (IP $\chi^2 < 25$, VDS>15).

Table 7: Selection for $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ and $B_{(s)}^0 \rightarrow h^+ h'^-$ channels. The cuts changed with respect to the published analysis are in bold.

	Cut	value
		$B_s^0 \rightarrow \mu^+ \mu^-$ and $B_{(s)}^0 \rightarrow h^+ h'^-$
μ / h	track χ^2/ndf	<4 (5)
	IP χ^2	>25
	p_T	> 0.25 and < 40 GeV/ c
	p	< 500 (1000) GeV/ c
	KL	< 5000
	ISMUON	true
$B_{(s)}$	$ M_{hh} - M(B_{(s)}) $	<600 MeV/ c^2
	DOCA	<0.3 mm
	vertex χ^2	<9
	VDS	> 15
	IP χ^2	< 25
	t	< 9 (5) $\cdot \tau(B_s^0)$
	BDTS	> 0.05
	$DLL(K - \pi)$	< 10
	$DLL(\mu - \pi)$	> -5

Table 8: Selections for $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi \phi$ normalization channels. The cuts changed with respect to the published analysis are in bold.

$B^+ \rightarrow J/\psi K^+$			$B_s^0 \rightarrow J/\psi \phi$		
	Cut	value		cut	value
μ	track χ^2/ndf	<4 (5)	μ	track χ^2/ndf	<4 (5)
	IP χ^2	>25		IP χ^2	>25
	ISMUON	true		ISMUON	true
J/ψ	DOCA	<0.3 mm		DOCA	<0.3 mm
	vertex χ^2	<9		vertex χ^2	<9
	Δm	<60 MeV		Δm	<60 MeV
	VDS	>15		VDS	>15
K^\pm	track χ^2/ndf	<5	K^\pm	track χ^2/ndf	<5
	IP χ^2	>25		IP $\chi^2 [K^+, K^-]$	[> 4, > 4]
				$DLL(K - \pi) [K^+, K^-]$	> 0
			ϕ	Δm	<10 MeV
				IP χ^2	> 25
B_u	IP χ^2	<25	B_s	IP χ^2	<25
	Δm	<100 MeV		Δm	<100 MeV
	vertex χ^2	< 45		vertex χ^2	<75
	BDTS	> 0.05		BDTS	>0.05

314 **5.1 The BDTS discriminant**

315 The BDTS discriminant has been found the most efficient way to further reduce back-
316 ground sample while keeping high efficiency on the signal. The variables entering the
317 BDTS are:

- 318 - the impact parameter (IP(B)) and impact parameter χ^2 (IP χ^2 (B)) of the B candi-
319 date;
320 - the χ^2 of the secondary vertex (VCHI2);
321 - the angle between the direction of the momentum of the B candidate and the di-
322 rection defined by the difference between the secondary and the primary vertex
323 (DIRA);
324 - the minimum distance between the two daughter tracks (DOCA);
325 - the minimum impact parameter of the muons with respect any primary vertex
326 (minIP(μ)).

327 Only the variables DOCA and IP(B) are in common with the BDT used in the limit
328 computation (see Section 8).

329 The primary vertex used to compute the IP(B), IP χ^2 (B) and DIRA variables is iden-
330 tified by extrapolating to the beam axis the line defined by the momentum of the B
331 candidate and taking the vertex with the closest distance with respect to the intersection
332 of this line with the beam axis. This vertex is refitted after the removal of the tracks
333 belonging to the B decay chain.

334 In Fig. 5 the BDTS distribution for $B_s^0 \rightarrow \mu^+ \mu^-$ Monte Carlo sample (left) and for
335 dimuon sidebands candidates (right) in data are shown: as for the BDT used for the limit
336 computation, the distribution has been kept uniform between 0 and 1 for signal and, as a
337 consequence, is peaked at zero for background. The background distribution agrees nicely
338 with the expected distribution from $b\bar{b} \rightarrow \mu\mu X$ Monte Carlo events.

339 In Fig. 6 the rejection versus efficiency curve is shown for the $B_s^0 \rightarrow \mu^+ \mu^-$ and
340 $b\bar{b} \rightarrow \mu\mu X$ Monte Carlo samples: a signal efficiency of $\sim 93\%$ can be obtained for a
341 background rejection of $\sim 70\%$ applying a BDTS cut BDTS>0.05.

342 The background events rejected by the BDTS cut are mostly concentrated in the
343 low BDT bins. This is shown in Figure 7 where the BDT distribution in the reduced
344 range between [0–0.25] is shown for $B_s^0 \rightarrow \mu^+ \mu^-$ candidates in the mass sidebands before
345 (black triangles) and after (blue dots) the BDTS cut at 0.05. The BDT distribution for
346 $b\bar{b} \rightarrow \mu\mu X$ Monte Carlo events is also shown overlayed (red squares): the $b\bar{b} \rightarrow \mu\mu X$
347 background reproduces nicely the background in data. This will be further discussed in
348 Section 8.

349 In order to minimize the systematic uncertainty in the normalization factors, the same
350 BDTS cut is also applied to the three normalization channels. For the $B^+ \rightarrow J/\psi K^+$

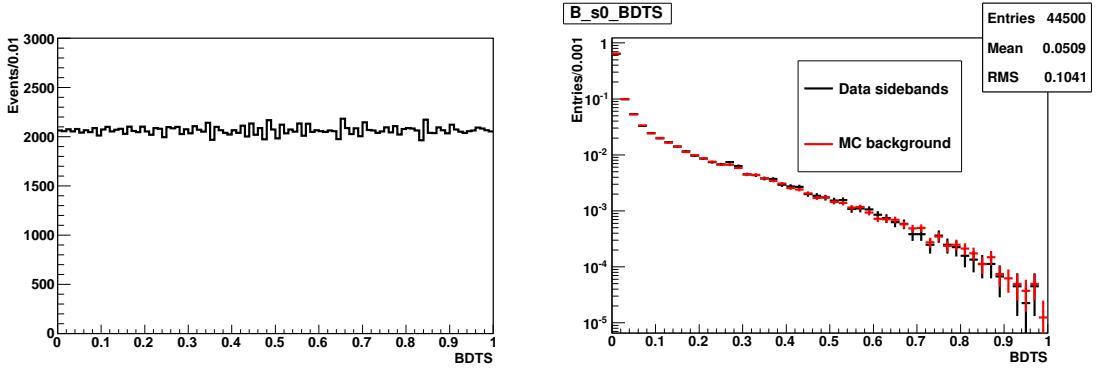


Figure 5: Left: BDTS distribution for $B_s^0 \rightarrow \mu^+\mu^-$ Monte Carlo events; right: BDTS distribution for $B_s^0 \rightarrow \mu^+\mu^-$ sidebands events (black markers) and $b\bar{b} \rightarrow \mu\mu X$ simulated events (red markers).

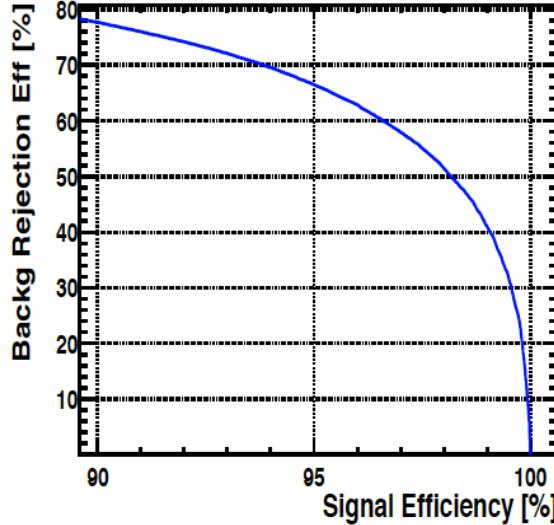


Figure 6: Rejection versus efficiency curve for BDTS built for $B_s^0 \rightarrow \mu^+\mu^-$ and $b\bar{b} \rightarrow \mu\mu X$ Monte Carlo events.

and $B_s^0 \rightarrow J/\psi\phi$ modes the χ^2 of the secondary vertex is substituted by the χ^2 of the J/ψ vertex, the flight distance is computed between the J/ψ vertex and the primary vertex and the DOCA is computed between the two muons from the J/ψ decay. In this way, the distributions of all the variables but minIP, are very similar for $B_s^0 \rightarrow \mu^+\mu^-$, $B_{(s)}^0 \rightarrow h^+h'^-$, $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi\phi$, resulting in a very similar efficiency for signal and normalization channels.

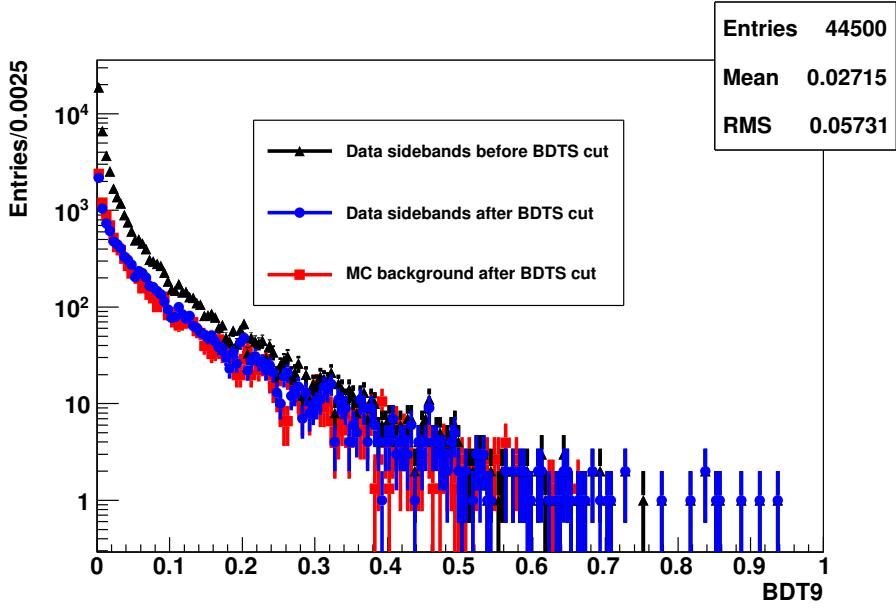


Figure 7: BDT distribution in the reduced range between [0–0.25] is shown for $B_s^0 \rightarrow \mu^+ \mu^-$ candidates in the mass sidebands before (black triangles) and after (blue dots) the BDTS cut at 0.05. The BDT distribution for $b\bar{b} \rightarrow \mu\mu X$ Monte Carlo events is also shown overlayed (red squares).

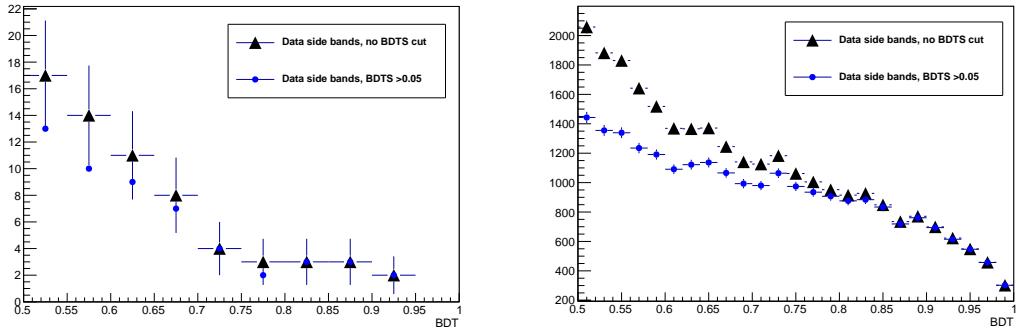


Figure 8: BDT distributions for $BDT = 0.5–1.0$ for dimuon background (left) and $B_{(s)}^0 \rightarrow h^+ h^-$ background (right) in the sidebands before (black) and after (blue) the BDTS cut.

In Fig. 9 the BDTS distribution for $B_s^0 \rightarrow \mu^+ \mu^-$ (top), $B^+ \rightarrow J/\psi K^+$ (center) and $B_s^0 \rightarrow J/\psi \phi$ (bottom) are shown using Monte Carlo samples non-smeared (left), smeared (center) and over-smeared (right). In Table 9 the efficiencies of different BDTS cuts on $B_s^0 \rightarrow \mu^+ \mu^-$, $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi \phi$ are shown for smeared Monte Carlo samples. The efficiency ratios for unsmeared, smeared and oversmeared $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^+ \rightarrow J/\psi K^+$ simulated samples are shown in Fig. 10: they are 1 within 0.4%.

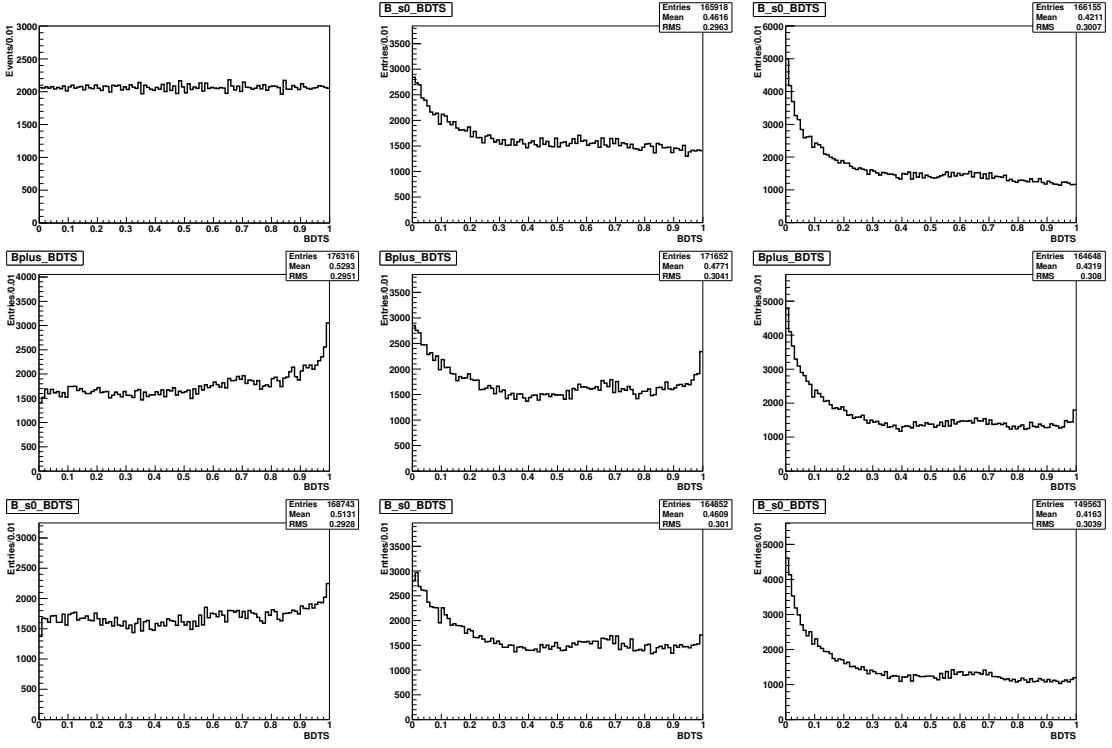


Figure 9: BDTS distribution for $B_s^0 \rightarrow \mu^+ \mu^-$ (top), $B^+ \rightarrow J/\psi K^+$ (center) and $B_s^0 \rightarrow J/\psi \phi$ (bottom) are shown using non-smeared (left), smeared (center) and over-smeared (right) Monte Carlo samples.

363 The BDTS cut has been chosen by looking at the combined performance of the BDTS
 364 and the BDT described in Section 8. The rejection versus efficiency curves for the BDT

Table 9: Efficiencies for different BDTS cuts for $B_s^0 \rightarrow \mu^+ \mu^-$, $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi \phi$ smeared Monte Carlo samples. The efficiency ratios between $B^+ \rightarrow J/\psi K^+$, $B_s^0 \rightarrow J/\psi \phi$, $B^0 \rightarrow K^+ \pi^-$ and $B_s^0 \rightarrow \mu^+ \mu^-$ are reported in the third, fifth and seventh row respectively.

BDTS cut	0.01	0.02	0.03	0.04	0.05	0.08	0.1
$\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}$	0.9828(3)	0.9664(4)	0.9501(5)	0.9354(6)	0.9210(7)	0.8815(8)	0.8570(9)
$\epsilon_{B^+ \rightarrow J/\psi K^+}$ ratio	0.9834(3)	0.9673(4)	0.9516(5)	0.9372(6)	0.9228(6)	0.8833(8)	0.8587(9)
$\epsilon_{B_s^0 \rightarrow J/\psi \phi}$ ratio	1.0006(5)	1.0010(6)	1.0015(8)	1.0019(9)	1.002(1)	1.002(1)	1.002(1)
$\epsilon_{B^0 \rightarrow K^+ \pi^-}$ ratio	0.9830(3)	0.9650(5)	0.9487(5)	0.9328(6)	0.9170(7)	0.8750(8)	0.8495(9)
$\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}$ ratio	1.0001(5)	0.9986(7)	0.9985(8)	0.9972(9)	0.996(1)	0.993(1)	0.991(1)
$\epsilon_{B^0 \rightarrow K^+ \pi^-}$ ratio	0.9825(2)	0.9651(3)	0.9497(4)	0.9350(5)	0.9200(5)	0.8794(6)	0.8538(7)
$\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}$ ratio	0.9996(4)	0.9987(6)	0.9995(7)	0.9995(8)	0.9989(9)	0.998(1)	0.996(1)

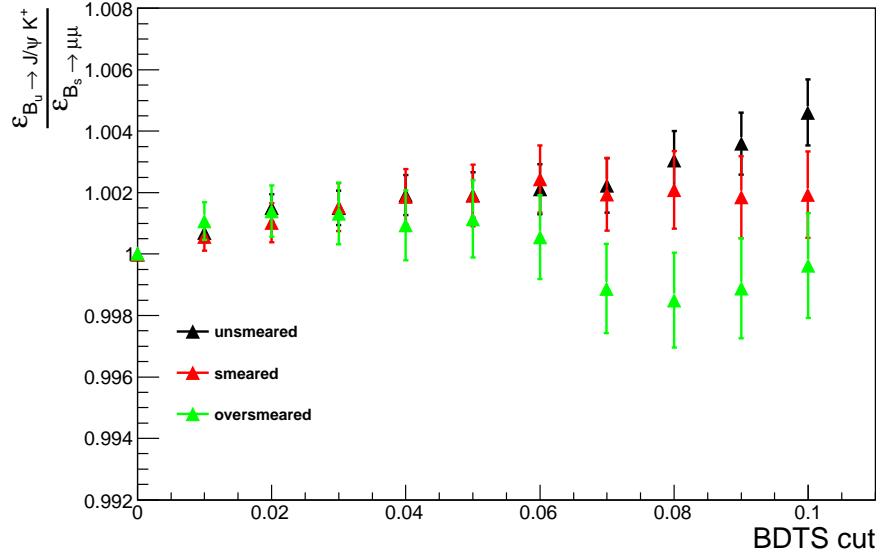


Figure 10: Efficiency ratios as a function of the BDTS cut for unsmeared (black triangles), smeared (red triangles) and over-smeared (light green triangles) $B_s^0 \rightarrow \mu^+\mu^-$ and $B^+ \rightarrow J/\psi K^+$ simulated samples.

as a function of different BDTS cuts are shown in Fig. 11 (left) for $B_s^0 \rightarrow \mu^+\mu^-$ and $b\bar{b} \rightarrow \mu\mu X$ Monte Carlo samples. The curves have been normalized to the number of events before the BDTS cut. Figure 11 (right) shows the zoom in the more sensitive region: the blue line obtained with a BDTS cut $\text{BDTS}>0.05$ shows the best performance.

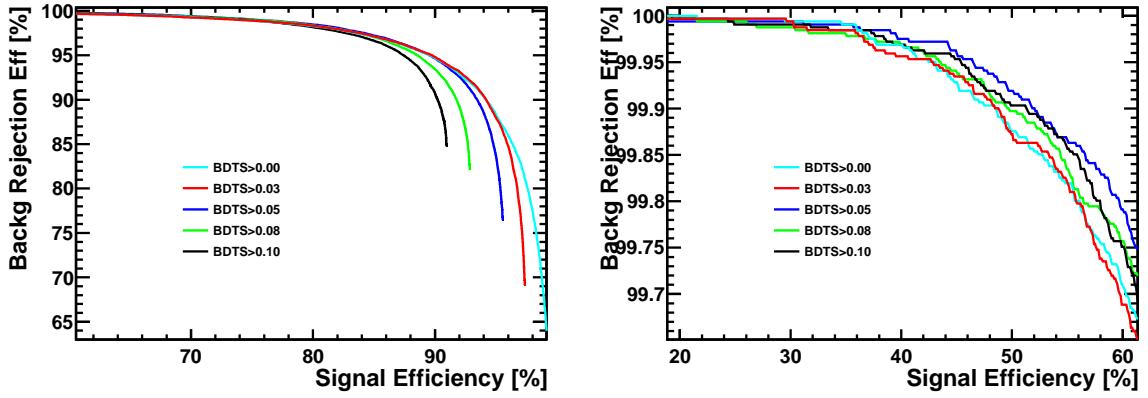


Figure 11: BDT roc curves for different BDTS cuts. Left: full range, right: zoomed plot.

369

6 Muon Identification

370 The muon identification procedure (muonID) is a key ingredient of the analysis. It must
371 be very efficient on $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ signals and has to reduce all the sources of background
372 not made up of real muons. In addition, the absolute value of the muonID efficiency
373 is required when we normalize to $B^0 \rightarrow K^+ \pi^-$ events, while only the ratio of muonID
374 efficiencies is necessary when $B^+ \rightarrow J/\psi K^+$ or $B_s^0 \rightarrow J/\psi \phi$ are used as normalization
375 channels.

376 The muonID and the details of the procedures used to measure its performance with
377 data are described in Ref. [46], where all the results on muon identification efficiency mea-
378 sured on a sample of about 300 pb^{-1} acquired in the first months of 2011 data taking, are
379 presented. The results updated with $\sim 966 \text{ pb}^{-1}$ of the stripping17 (reco12) production
380 are presented in this Section.

381 The usual tag-and-probe method is applied using $J/\psi \rightarrow \mu\mu$ candidates, where one of
382 the muons (μ_{tag}) is identified with $\text{IsMuon}=1$ while the second muon (μ_{probe}) is selected
383 without using any information from the muon system. This second muon is used to
384 estimate the geometrical acceptance of the muon detector relative to tracking chambers,
385 the efficiency of the IsMuon request and the performance of the muon hypothesis test.
386 The muon or non-muon hypothesis is based on how the hits in the muon chambers are
387 aligned with respect to the extrapolation of the tracks from the tracking system. This
388 information is combined together with the informations coming from calorimeters and
389 RICH detectors into a Global Likelihood (DLL) that is used to perform an hypothesis
390 test.

391 The larger dataset ($\sim 1 \text{ fb}^{-1}$) with respect to the analysis based on 370 pb^{-1} allows
392 to study the dependence of the muonID efficiency not only as a function of p but also
393 as a function of p_T of the probe muon. The parameterization in p, p_T allows to reduce
394 the dominating systematic uncertainty in the muonID efficiency determination in the old
395 analysis that was due to the different p_T phase space covered by the calibration samples.

396

6.1 MuonID efficiency

397 The muonID efficiency is defined as the number of muons reconstructed as tracks surviving
398 the *IsMuon* requirement with respect to the number of total muons reconstructed as
399 tracks. This efficiency can be factorized into two parts:

- 400 1. the first contribution is the geometrical acceptance, α_μ , i.e. the request that the
401 reconstructed muon track points into the muon detector;
- 402 2. the second contribution is related to the efficiency of finding hits inside the FOI in
403 the muon stations for tracks pointing into the muon detector, ϵ_μ .

404 **Data Samples** - The same stripping lines used for the 370 pb^{-1} analysis have been
405 used: they are dedicated *tag-and-probe* selections developed in the framework of the *Cal-*
406 *ibration* stream to select J/ψ from b decays. In particular we used $b \rightarrow J/\psi X$ events

407 (2-body sample) from the *JpsiFromBNoPIDNoMip* calibration line and $B^+ \rightarrow J/\psi K^+$
 408 events (3-body sample) from the *JpsiKFromBNoPIDNoMip* calibration line. As in the
 409 previous analysis, in order to perform the measurement of the muon detector acceptance,
 410 in both lines the request that the μ_{probe} is in muon acceptance was registered but not
 411 enforced and the request that the μ_{probe} releases in the electromagnetic and hadronic
 412 calorimeters an energy compatible with a minimum ionizing particle removed. The re-
 413 sults obtained with the 2-body sample are then compared with the ones obtained with
 414 a 3-body sample and the observed difference added as a systematic uncertainty due to
 415 the choice of the sample. In the 370 pb^{-1} analysis this was the dominant systematic
 416 uncertainty.

417 **Selection -** The “tag-and-probe” selection used in muonID efficiency measurement
 418 is summarized in Table 10. It did not change with respect to the previous analysis.

Table 10: Selection cuts for the J/ψ , the μ_{tag} and the μ_{probe} , used to define the samples for α_μ and ϵ_μ measurements.

μ_{probe} and μ_{tag}	two long tracks with opposite charge $\chi^2_{\text{trk}}/\text{nDoF} < 3$ $p > 3 \text{ GeV}/c$ $p_T > 800 \text{ MeV}/c$ $\text{IP}\chi^2 > 25$
J/ψ	$\chi^2_{\text{vertex}}/\text{nDoF} < 8$ vertex distance significance > 15 $ M_{\mu\mu} - M_{J/\Psi} < \pm 200 \text{ MeV}/c^2$
μ_{tag} only	$p_T > 1.5 \text{ GeV}/c$ $p > 6 \text{ GeV}/c$ $\text{IP} > 120 \mu\text{m}$ $\text{MuAcc} = 1$ $\text{IsMuon} = 1$
μ_{probe} only	$\text{IP} > 50 \mu\text{m}$ $\text{MuAcc}=1$ (ϵ_μ evaluation only)

419 For the *JpsiKfromB* line only, a further track with $\text{IP}\chi^2 > 25$ is required, and the
 420 $\mu\mu K$ invariant mass must be in a 500 MeV/c window around M_{B^+} .

421 **Trigger -** The μ_{probe} is required not to fire the trigger at any level⁶ (“TIS event with
 422 respect to μ_{probe} ”, “TIS-unbias” in the following). To cross-check the results obtained
 423 with this condition, a different sample is used forcing μ_{tag} to satisfy the trigger by itself
 424 (“TOS event with respect to μ_{tag} ” or “TOS-unbias” in the following.)⁷

⁶A TIS event with respect to μ_{probe} is defined by the following condition: $[L0Global \times Hlt1Global \times Hlt2Global]_{\text{TIS}(\mu_{\text{probe}})} = \text{true}$.

⁷A TOS event with respect to μ_{tag} is defined by the following specifications: $[L0SingleMuon \times (Hlt1SingleMuonNoIP + Hlt1Track) \times Hlt2SingleMuon]|_{\text{TOS}(\mu_{\text{tag}})} = \text{true}$.

425 The difference between the results obtained with the two samples is added as systematic
426 uncertainty due to the trigger, as in the previous analysis.

427 **Method to measure the efficiency and background subtraction** - The values
428 of the muonID acceptance, α_μ , and the muonID efficiency given acceptance, ϵ_μ , are
429 extracted from the above samples of unbiased μ_{probe} , properly selecting the events around
430 the J/ψ mass peak. Since the average muonID acceptance or efficiency measured in a
431 given mass window includes contributions from both muons from J/ψ and spurious tracks,
432 a background subtraction procedure has to be applied. Two different methods have been
433 used to evaluate α_μ and ϵ_μ , and two different approaches have been used to subtract
434 the background. This translates into four possible estimates. All the details about the
435 methods are described in [46]. The reference method is the one called $W11$ in [46]. The
436 other methods are used to evaluate the systematic uncertainty on α_μ and ϵ_μ due to the
437 method.

438 **Results** - Figures 12–13–14 show the acceptance, efficiency and the convolution of
439 the acceptance and efficiency curves as a function of the momentum of μ_{probe} (left) and
440 the 2D maps (right) as a function of the p and p_T of the probe muon.

441 In the left plots the results obtained with the 2-body sample are compared with the
442 ones obtained with the 3-body sample and with the corresponding curves from simulated
443 events. In Fig. 15 the efficiency curve obtained with the 2-body sample in 966 pb^{-1} of
444 Stripping 17 is compared with the one obtained with 325 pb^{-1} of Stripping 13b. The two
445 curves are in agreement within 0.5%: this demonstrates the stability of the response of
446 the muon system performance during the full data taking period.

447 The acceptance and efficiency 2D maps are used in the computation of the normalization
448 factors, as explained in Sect. 10.

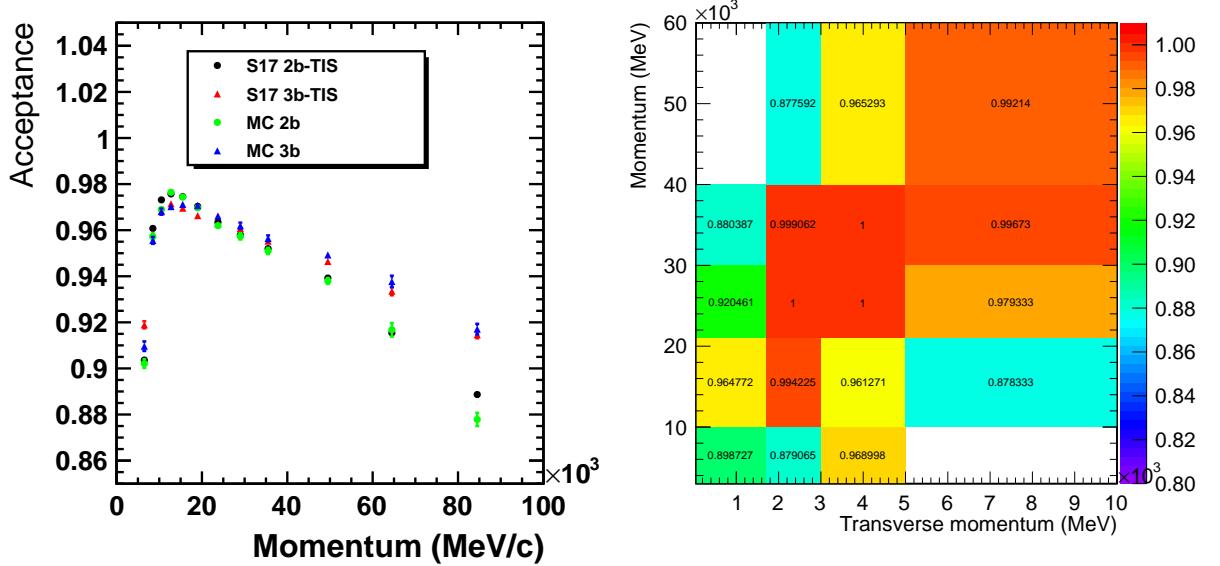


Figure 12: Acceptance of the muon system as a function of momentum of the probe muon for 2-body ($b \rightarrow J/\psi X$) and 3-body (J/ψ from $B^+ \rightarrow J/\psi K^+$) samples. Data are compared with Monte Carlo predictions.

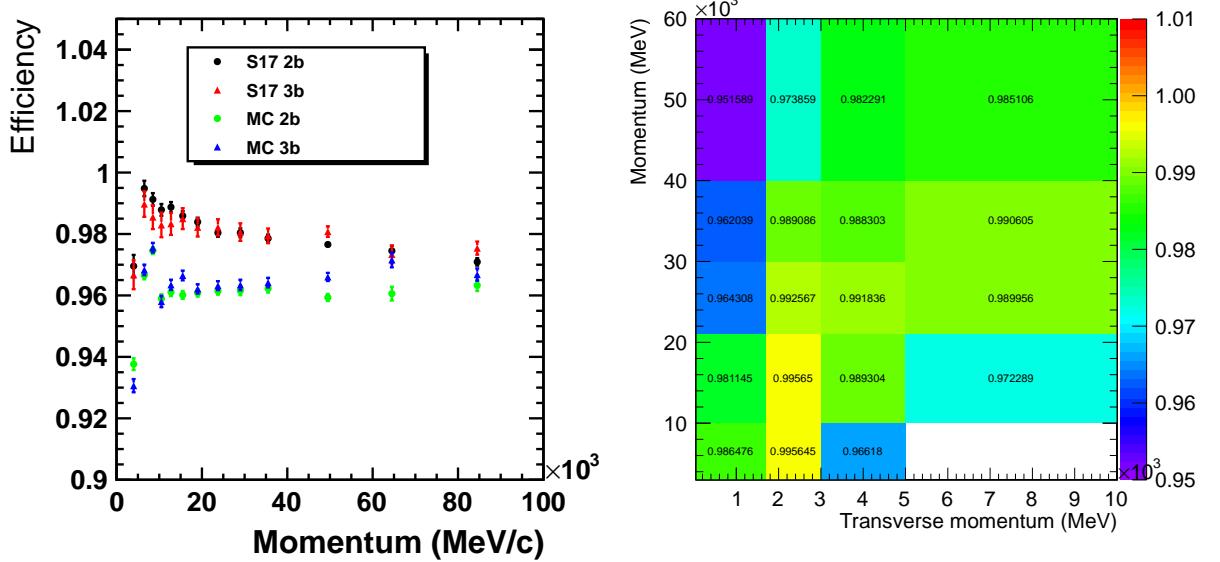


Figure 13: Efficiency of the $\text{IsMuon}=1$ requirement as a function of momentum of the probe muon for 2-body ($b \rightarrow J/\psi X$) and 3-body (J/ψ from $B^+ \rightarrow J/\psi K^+$) samples. Data are compared with Monte Carlo predictions.

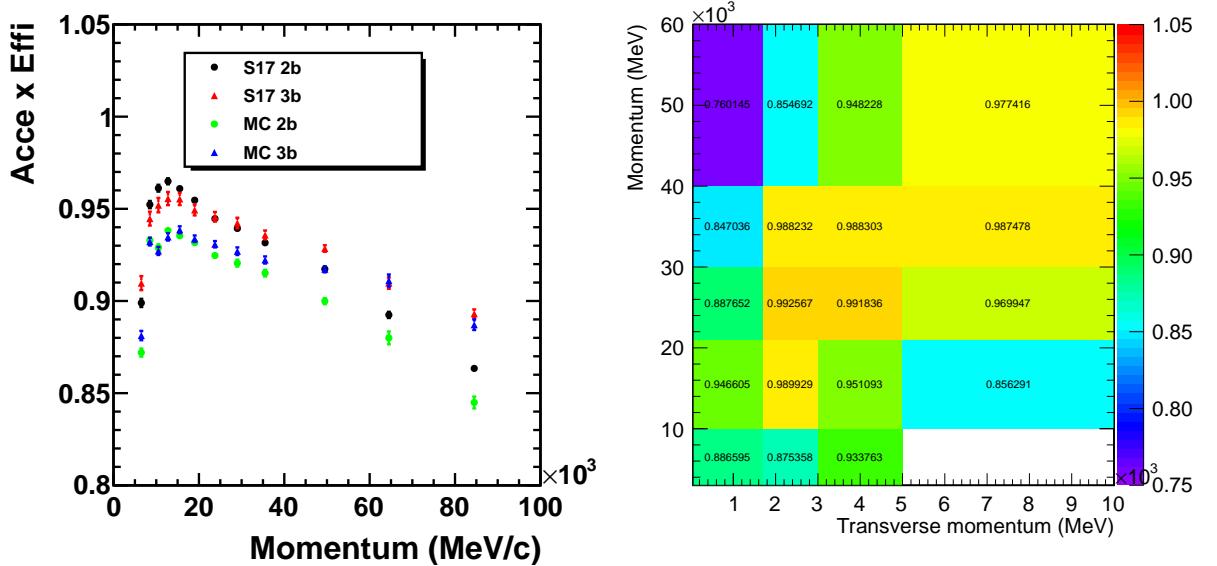


Figure 14: Convolution of the acceptance and IsMuon=1 requirements as a function of momentum of the probe muon for 2-body ($b \rightarrow J/\psi X$) and 3-body (J/ψ from $B^+ \rightarrow J/\psi K^+$) samples. Data are compared with Monte Carlo predictions.

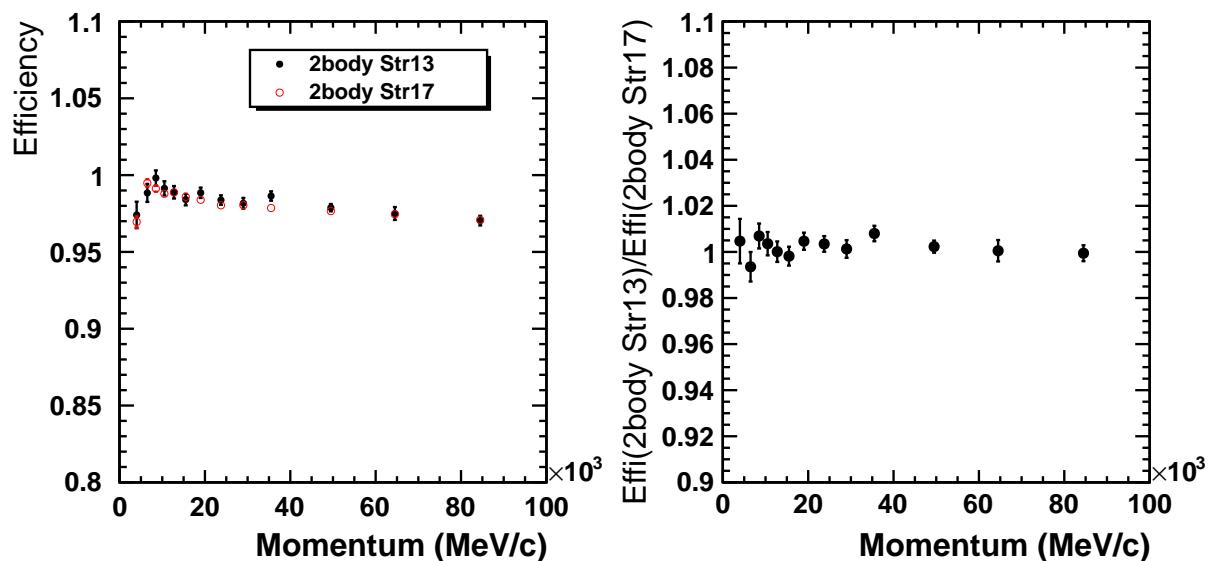


Figure 15: Efficiency of the IsMuon=1 requirement for the probe muon of a J/ψ sample in the stripping 13b and Stripping 17 production. Right: ratios of the two efficiency curves.

449 6.2 Muon hypothesis test (DLL)

450 It will be shown in Section 7.2 that the simultaneous application of cuts in the $DLL(K - \pi)$
 451 and $DLL(\mu - \pi)$ variables can considerably reduce the number of double misidentified
 452 $B_{(s)}^0 \rightarrow h^+h'^-$ events. In particular, the request $DLL(K - \pi) < 10 \& DLL(\mu - \pi) > -5$
 453 can reduce by a factor ~ 5 the rate of $B_{(s)}^0 \rightarrow h^+h'^-$ events with double decays in flight (Table 11).

455 The efficiency loss for $B_s^0 \rightarrow \mu^+\mu^-$ signal has been evaluated by measuring in bins
 456 of p of the probe muon from a $J/\psi \rightarrow \mu\mu$ sample the efficiency of this set of DLL cuts
 457 (ϵ_{DLLcut}^{data}) and then folding it into the spectrum of selected $B_s^0 \rightarrow \mu^+\mu^-$ events. Figure 16
 458 (left) shows the efficiency of the DLL cuts $DLL(K - \pi) < 10$ and $DLL(\mu - \pi) > -5$
 459 measured on a sample of $J/\psi \rightarrow \mu\mu$ on data and on Monte Carlo as a function of the
 460 momentum of μ -probe. The difference between data and Monte Carlo is fully due to the
 461 RICH $DLL(K - \pi)$ cut, since the efficiency for $DLL(\mu - \pi) > -5$ agrees nicely with the
 462 simulation (Fig. 17).

463 The efficiency on a $B_s^0 \rightarrow \mu^+\mu^-$ sample has been evaluated to be 97.3%. The same
 464 efficiency has been calculated in bins of the invariant mass-BDT plane, as shown in the
 465 right panel of Fig. 16. The efficiency is almost flat as a function of the invariant mass,
 466 while a $\sim 1.5\%$ slope is present when studied as a function of the BDT output; this a
 467 correction has been applied to the BDT PDF extracted from the $B_{(s)}^0 \rightarrow h^+h'^-$ sample,
 468 as discussed in Section 9.1.

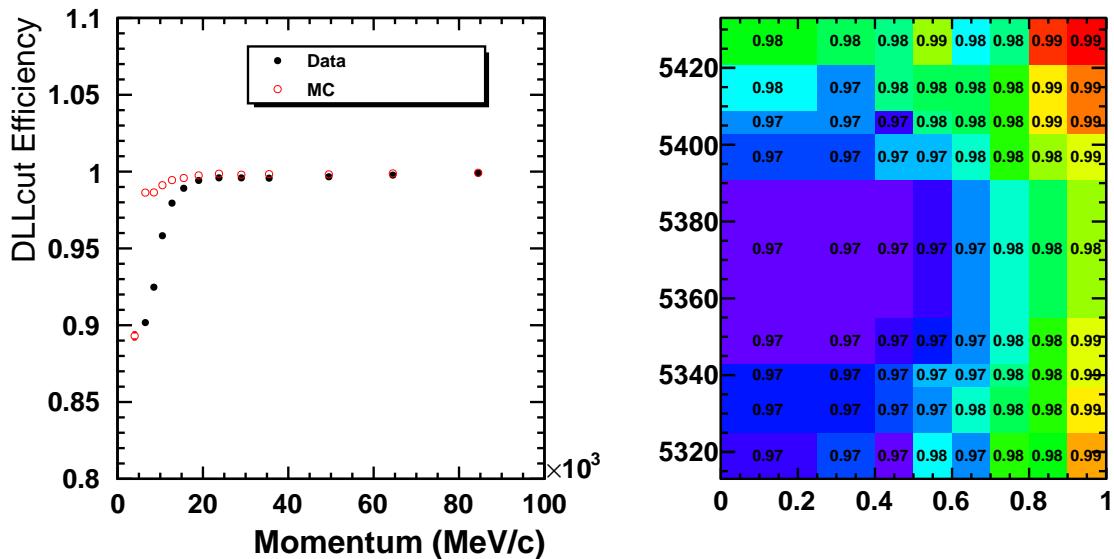


Figure 16: Left: efficiency of the DLL cut $DLL(K - \pi) < 10 \& DLL(\mu - \pi) > -5$ measured on data and MC samples as a function of the momentum of μ -probe, after the request IsMuon=1. Right: efficiency of the same DLL cut when folded into the p, p_T spectrum of selected $B_s^0 \rightarrow \mu^+\mu^-$ events as a function of the $B_s^0 \rightarrow \mu^+\mu^-$ invariant mass and BDT.

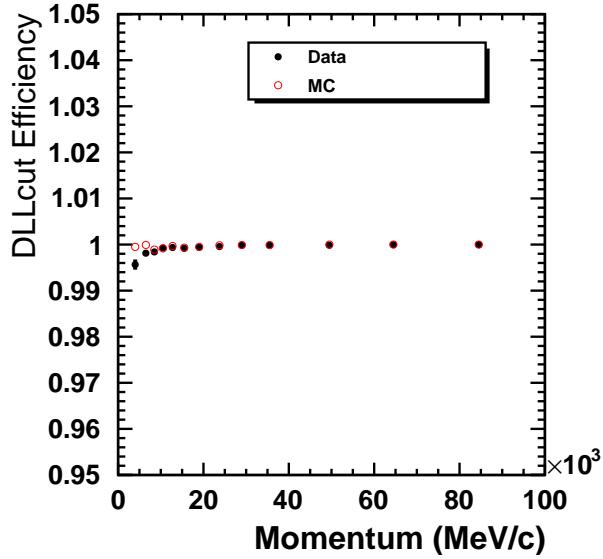


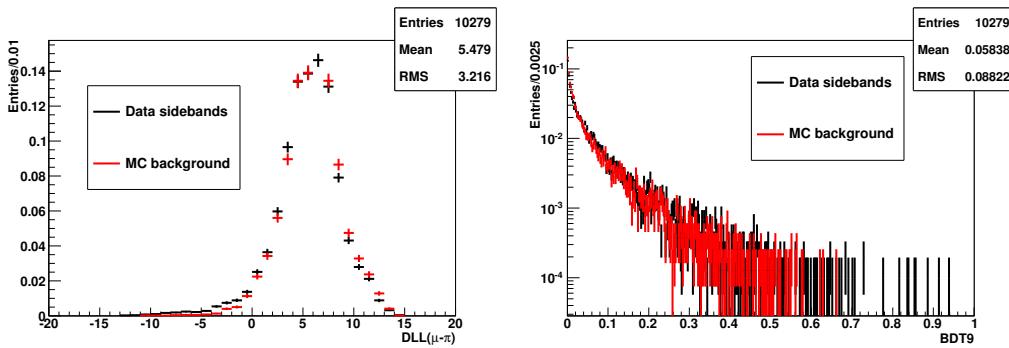
Figure 17: Efficiency of the DLL cut $DLL(\mu - \pi) > -5$ measured on data and MC samples as a function of the momentum of μ -probe, after the request IsMuon=1.

469 The tiny effect of the DLL cut on the mass resolution has been studied by fitting the
 470 invariant mass of $B_s^0 \rightarrow \mu^+ \mu^-$ simulated sample without and with a DLL cut: when the
 471 DLL cut is applied each event is weighted by the efficiency corrections evaluated in data.
 472 The resolution of the invariant mass is larger by 0.2 MeV/ c^2 for the corrected sample,
 473 bias which is well withing the total uncertainty in the mass resolution itself quoted in
 474 Section 9.5.

475 7 Misidentification rates and background composition

477 The number of combinatorial background events in the search windows is estimated by
 478 interpolating from the mass sidebands, therefore the knowledge of the exact background
 479 composition is not required.

480 As in the published analysis, the combinatorial background is fully dominated by the
 481 $b\bar{b} \rightarrow \mu\mu X$ component: this is shown in Fig. 18 where the $DLL(\mu - \pi)$ ⁸ and the BDT
 482 distributions obtained with background events in data sidebands are compared with the
 same distributions obtained from simulated $b\bar{b} \rightarrow \mu\mu X$ events.



483 Figure 18: $DLL(\mu - \pi)$ and BDT distributions obtained with background events in data sidebands and simulated $b\bar{b} \rightarrow \mu\mu X$ events.

484 Among the peaking backgrounds, the most dangerous one is represented by the $B_s^0 \rightarrow$
 485 $h^+ h'^-$ decays (with $h^{(\prime)}$ being a kaon or a pion) where both hadrons are misidentified as
 486 muons: in fact, these decays are topologically identical to the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ signals and
 487 peak in the search window.

488 7.1 Pion and kaon misidentification probabilities

489 The $\epsilon(\pi \rightarrow \mu)$ and $\epsilon(K \rightarrow \mu)$ fake rates have been measured using a sample of $D^0 \rightarrow K\pi$
 490 from the $D^* \rightarrow D^0\pi$ decays extracted from $\sim 900 \text{ pb}^{-1}$ of the line *NoPIDstar-
 491 WithD02RSKPiLine* of the Stripping 17 production.

492 Two different approaches have been used to cancel possible biases arising from the
 493 trigger:

- 494 • the kaon or pion probe track is required to be TIS with respect the L0Global and
 495 Hlt1Phys lines (“Hlt1TIS” in the following), a similar requirement on Hlt2Phys line
 496 leaving the result unaffected;

⁸ DLL means combined DLL unless otherwise stated.

- 497 ● the kaon or pion probe track is required to be TIS with respect the L0Global, while
 498 a Hlt1AllL0Track decision is required to the event (“TrackAllDec” in the following),
 499 which, in the case of events with hadron decays, gives sufficient unbias with respect
 500 to the muon Hlt1 lines.

501 The comparison between the two unbias strategies is used as a systematic cross-check,
 502 the first being more accurate from a pure trigger unbias strategy perspective, but also
 503 providing a less abundant sample of events.

504 In order to cleanup the sample after the stripping selection, the track of the D^0 decays
 505 that is not used to evaluate the fake rate is required to be well identified with a cut on
 506 the $DLL(K - \pi)$ ⁹. To evaluate the $\pi \rightarrow \mu$ and $K \rightarrow \mu$ fake rates two methods have been
 507 used:

508 - **method n.1:**

509 $D^0 \rightarrow K\pi$ candidates are selected within a window of $\pm 20 \text{ MeV}/c^2$ around the mass
 510 peak and requiring that the π or the K passes the IsMuon=1 condition. The fake
 511 rate measured in a given mass window (ϵ_{S+B}) is related to the actual fake rate ϵ_S
 512 by the following equation:

$$\epsilon_{S+B} = \frac{S}{S+B}\epsilon_S + \frac{B}{S+B}\epsilon_B \quad (1)$$

513 where S and B are the number of signal and background events in a $\pm 20 \text{ MeV}/c^2$
 514 mass window around the peak, which are extracted by fitting with a Crystal Ball
 515 the signal and with a polynomial function the background. The fake rate for the
 516 background ϵ_B is evaluated in the mass sidebands ($|m - m_{D^0}| > 40 \text{ MeV}/c^2$). The
 517 above relation becomes:

$$\epsilon_S = \frac{B+S}{S} \left(\epsilon_{S+B} - \frac{B}{S+B}\epsilon_B \right) \quad (2)$$

518 solving for ϵ_S . The procedure is repeated by dividing the D^0 sample in 11 bins in p
 519 and 4 bins in p_T of the track used to evaluate the fake rates (probe track). In Fig. 19
 520 the mass distributions corresponding to 10 p -bins for a p_T range $1700 < p_T < 3000$
 521 MeV/ c of the probe track are shown.

522 - **method n.2:**

523 The mass distributions of D^0 candidates are obtained by dividing the sample in 11
 524 bins in p and 4 bins in p_T of probe track as in the previous method. Each distribution
 525 is fitted with a double Gaussian for the signal and an exponential function for the
 526 background (an alternative fit model with a single Gaussian for the signal and a

⁹The kaon (pion) is required to pass the cut $DLL(K - \pi) > 10$ ($DLL(K - \pi) < 0$) when the $\epsilon(\pi \rightarrow \mu)$ ($\epsilon(K \rightarrow \mu)$) fake rate is measured.

527 3rd order Chebichev polynomial for the background has been also used). The probe
 528 track is required to satisfy the condition IsMuon = 0, each distribution is fitted
 529 and the number of signal candidates $N_S(\text{IsMuon}=0)$ is obtained. The fits are then
 530 repeated by requiring the probe track to satisfy IsMuon=1 and $N_S(\text{IsMuon}=1)$ is
 531 extracted. Finally the fake rate is obtained via the relation:

$$\epsilon_s = \frac{N_S(\text{IsMuon} = 1)}{N_S(\text{IsMuon} = 0) + N_S(\text{IsMuon} = 1)} \quad (3)$$

532 The fits for 10 bins in p of the probe track for the range $1700 < p_T < 3000$ MeV/ c
 533 are shown in Fig. 20 for IsMuon=0 and in Fig. 21 for IsMuon=1.

534 This method has been also implemented, for systematic evaluation, by changing
 535 the strategy for signal counting (in both Ismuon=0 and Ismuon=1 hypotheses) to
 536 a background extrapolation under the peak via and exponential function. Results
 537 will be given in the following.

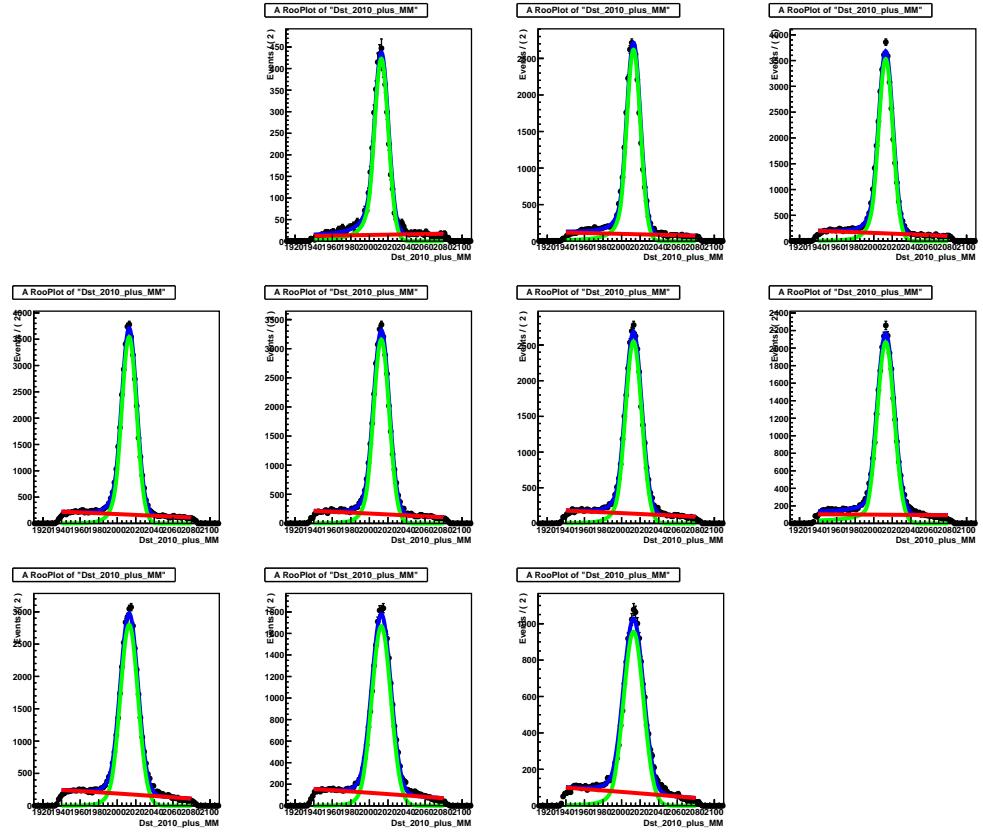


Figure 19: Mass distributions of the $D^0 \rightarrow K\pi$ decay for 10 p bins and for the p_T range 1700 MeV/ c < p_T 3000 MeV/ c of the probe track used to estimate the fake rate $\epsilon(K \rightarrow \mu)$ with method 1.

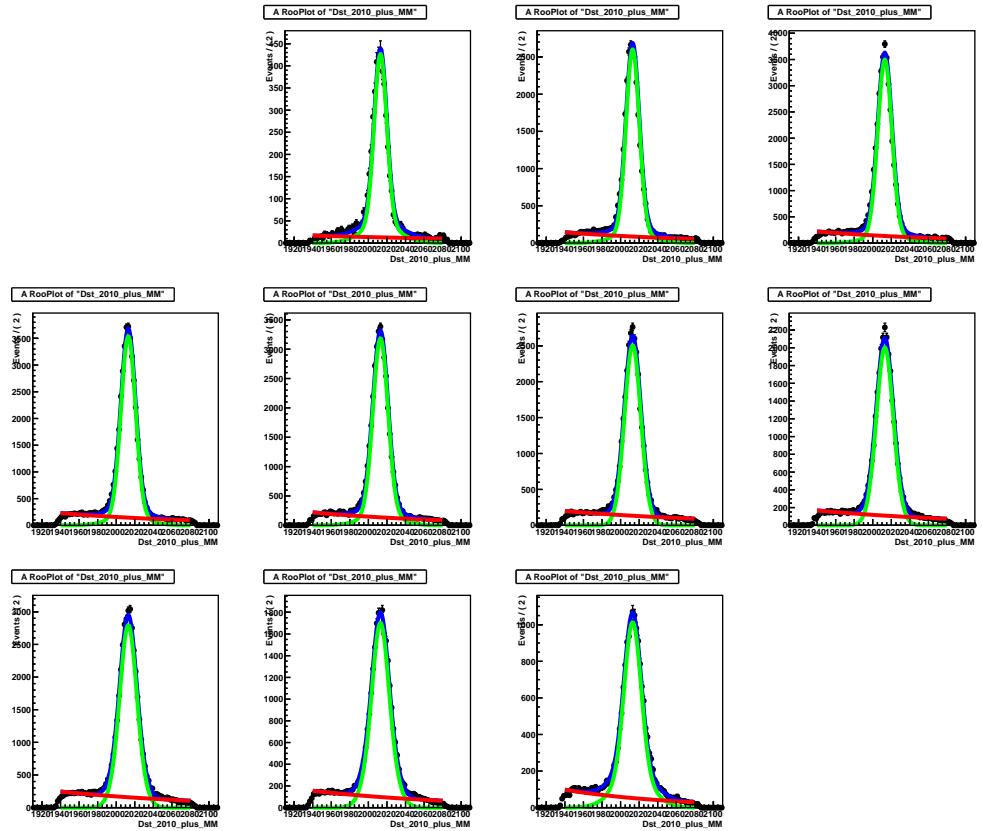


Figure 20: Mass distributions of the $D^0 \rightarrow K\pi$ decay for 10 p bins and for the p_T range 1700 MeV/ c < p_T 3000 MeV/ c of the probe track used to estimate the fake rate $\epsilon(K \rightarrow \mu)$ with method 2. The probe track is required to satisfy the condition IsMuon=0.

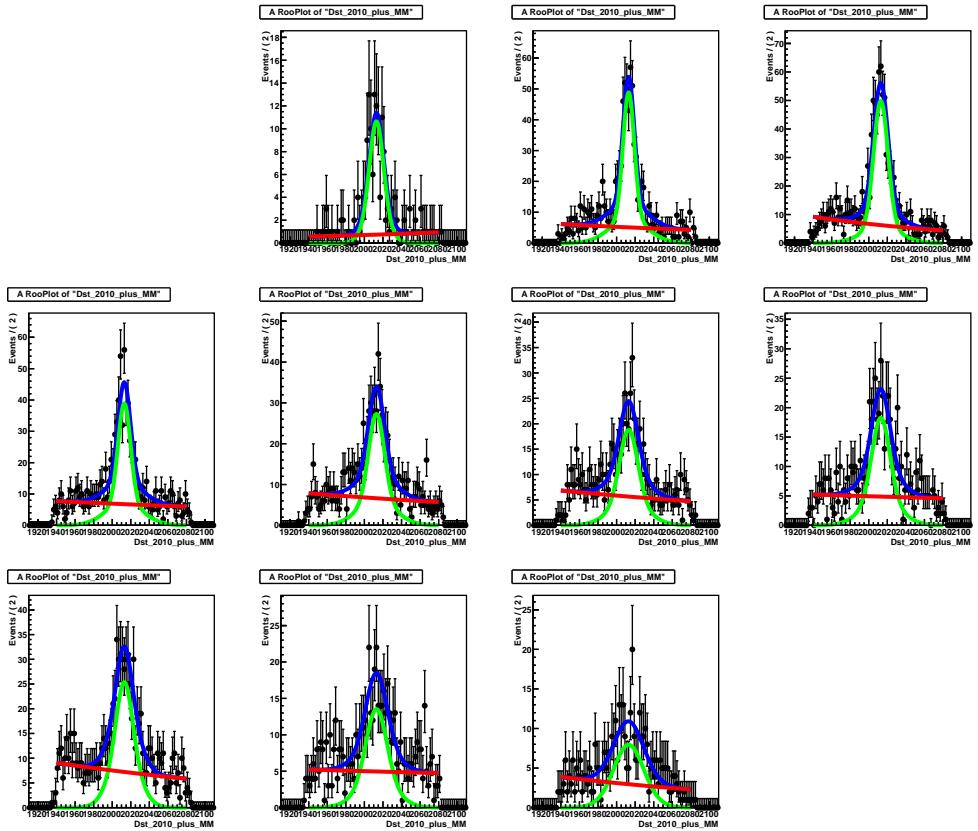


Figure 21: Mass distributions of the $D^0 \rightarrow K\pi$ decay for 10 p bins and for the p_T range 1700 MeV/ c < p_T 3000 MeV/ c of the probe track used to estimate the fake rate $\epsilon(K \rightarrow \mu)$ with method 2. The probe track is required to satisfy the condition IsMuon=1.

538 The fake kaon misID probability evaluated with method 1 in Hlt1TIS hypothesis
539 (Hlt1TIS sample in the following) as a function of probe track momentum for the four
540 p_T bins is shown in Fig. 22 (top); the same probability computed with method 2 in
541 TrackAllDec hypothesis (TrackAllDec sample in the following) is shown in the middle,
542 and the ratio between the two in the bottom. The same results are shown for pion in
543 Fig. 23. The two methods are in good agreement. A systematic error will be assessed by
544 using both set of curves to compute the double misID background.

545 In order to reduce the number of double misidentified $B_{(s)}^0 \rightarrow h^+h'^-$ events, cuts on
546 the $DLL(K - \pi)$ and $DLL(\mu - \pi)$ can be applied on top of the IsMuon requirement¹⁰. In
547 Section 6 it has been demonstrated that the combination of cuts $DLL(K - \pi) < 10$ and
548 $DLL(\mu - \pi) > -5$ selects $B_s^0 \rightarrow \mu^+\mu^-$ with high efficiency, ~ 0.97 . The same selection,
549 applied to kaons and pions, reduces the misID probabilities by large factors, as shown in
550 Fig. 24.

¹⁰The symbol DLL refers to the combined DLL unless otherwise stated.

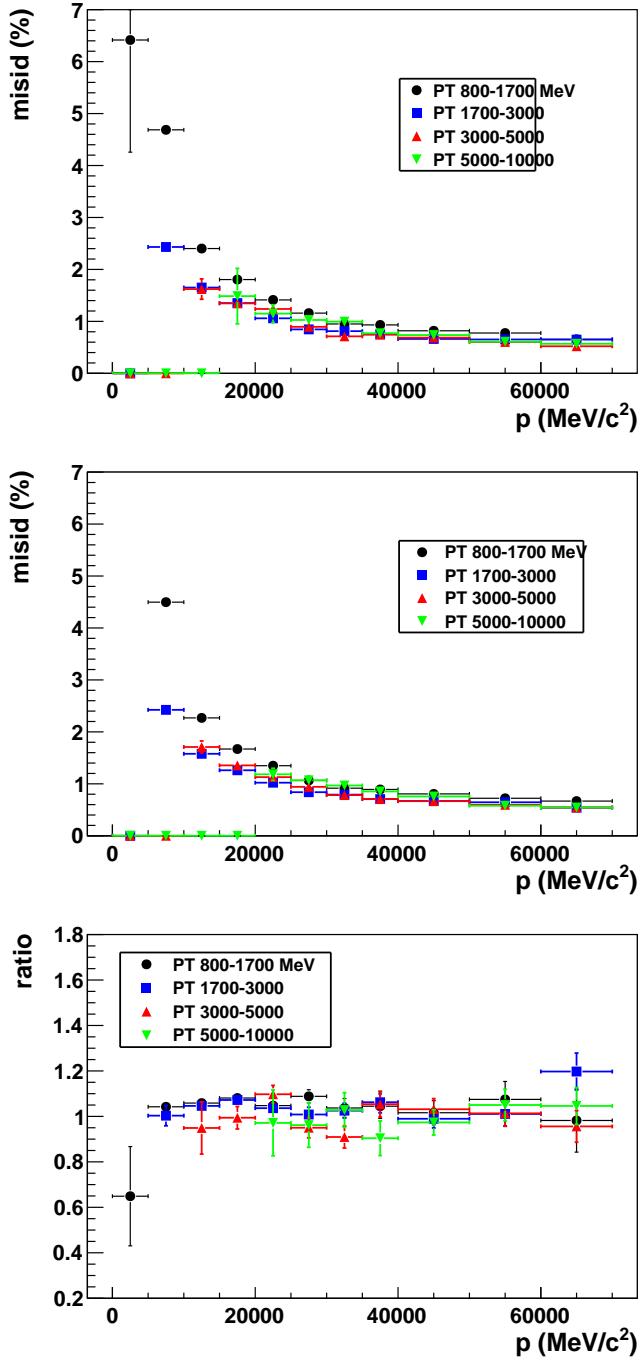


Figure 22: Kaon misid vs momentum for the four p_T bins of the probe track: Hlt1TIS sample (top), TrackAllDec sample (middle), and the ratio between the two.

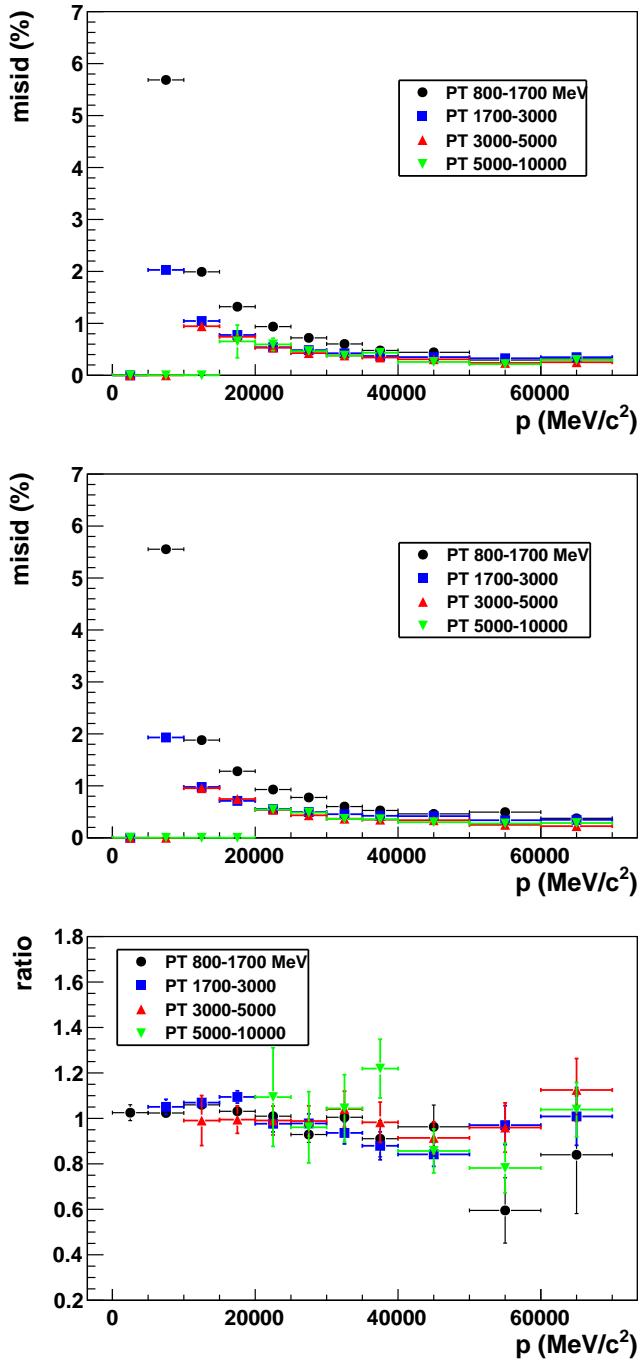


Figure 23: Pion misid vs momentum for the four p_T bins of the probe track: Hlt1TIS sample (top), TrackAllDec sample (middle), and the ratio between the two.

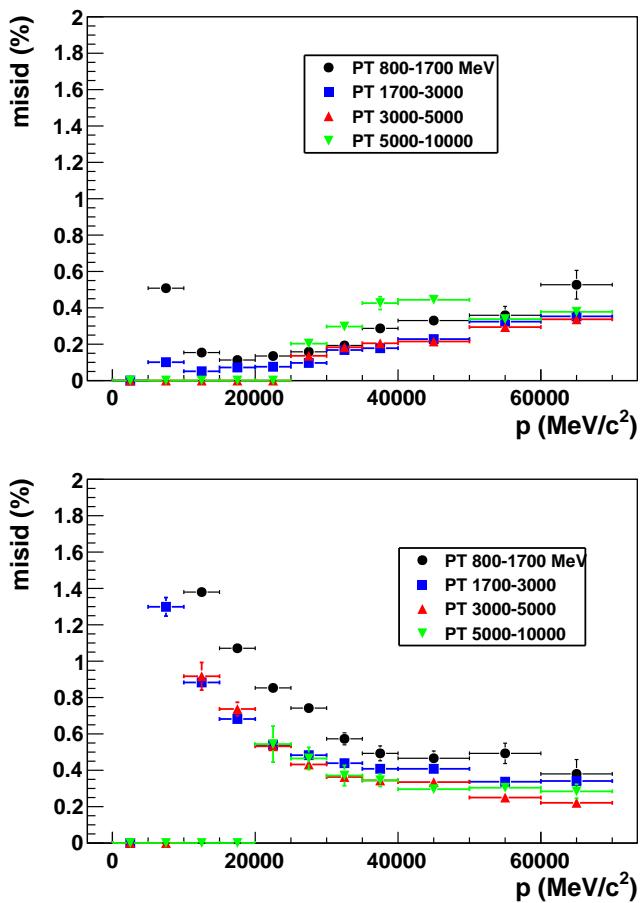


Figure 24: Kaon (top) and pion (bottom) misid probabilities vs momentum for the four p_T bins of the probe track after *DLL* selection (TrackAllDec sample)

Table 11: Average double misID probability for exclusive $B_{(s)}^0 \rightarrow h^+h'^-$ decays and for the inclusive combination of them, for the two series of single particle misID probabilities available.

dataset	$B_s^0 \rightarrow K^+K^-$ 10^{-4}	$B_s^0 \rightarrow \pi^+K^-$ 10^{-4}	$B^0 \rightarrow K^+\pi^-$ 10^{-4}	$B^0 \rightarrow \pi^+\pi^-$ 10^{-4}	$B_{(s)}^0 \rightarrow h^+h'^-$ 10^{-4}
Hlt1TIS	1.19 ± 0.05	0.82 ± 0.05	0.84 ± 0.05	0.52 ± 0.03	0.88 ± 0.04
TrackAllDec	1.16 ± 0.04	0.79 ± 0.05	0.81 ± 0.04	0.50 ± 0.03	0.86 ± 0.03
$DLL(K - \pi) < 10$ and $DLL(\mu - \pi) > -5$					
Hlt1TIS	0.065 ± 0.006	0.146 ± 0.008	0.146 ± 0.008	0.334 ± 0.014	0.153 ± 0.008
trackAllDec	0.066 ± 0.005	0.148 ± 0.007	0.148 ± 0.007	0.323 ± 0.012	0.152 ± 0.007

7.2 Double misidentification probability

The double fake rate has been estimated by convoluting the kaon and pion misID curves given above with the momentum and p_T spectrum of the two hadrons of MC $B_{(s)}^0 \rightarrow h^+h'^-$ decays, selected as described in Section 5. This has been done with a toy technique that takes properly into account the uncertainties on the fake rate in each of the 4×11 bins and the correlations between p and p_T of both tracks. The average double misID probability for the exclusive $B_{(s)}^0 \rightarrow h^+h'^-$ decays is shown in Table 11 before and after the DLL selection described above, and separately for the Hlt1TIS and TrackAllDec samples.

The average double misID for inclusive $B_{(s)}^0 \rightarrow h^+h'^-$ decays, $\epsilon_{hh \rightarrow \mu\mu}$, is also shown in Table 11, and it has been obtained by weighting the values obtained for exclusive decays according to their relative production rate ¹¹.

Before DLL cuts, a double misID rate $\epsilon_{hh \rightarrow \mu\mu} = (8.6 \pm 0.34_{\text{stat}} \pm 0.30_{\text{syst}}) \times 10^{-5}$ is measured, where the syst comes from the difference between the results given in Table 11, plus an additional contribution coming from fit models. To compare with the result given in the previous analysis [22], $\epsilon_{hh \rightarrow \mu\mu} = (7.5 \pm 0.5) \times 10^{-5}$, we also evaluated the double misID with the same selection, that is $GL > 0.5$ and $p > 5$ GeV/ c^2 , which gives a double misID $\epsilon_{hh \rightarrow \mu\mu} = (7.38 \pm 0.25) \times 10^{-5}$.

After DLL cuts, the double misID rate is reduced by more than a factor of 5, giving $\epsilon_{hh \rightarrow \mu\mu} = (1.52 \pm 0.07_{\text{stat}} \pm 0.07_{\text{syst}}) \times 10^{-5}$.

¹¹The following branching fractions are assumed: $\mathcal{B}(B^0 \rightarrow K^+\pi^-) = (1.95 \pm 0.06)10^{-5}$, $\mathcal{B}(B^0 \rightarrow \pi^+\pi^-) = (5.13 \pm 0.24)10^{-6}$, $\mathcal{B}(B_s^0 \rightarrow \pi^+K^-) = (5.05 \pm 1.0)10^{-6}$, $\mathcal{B}(B_s^0 \rightarrow K^+K^-) = (3.5 \pm 0.7)10^{-5}$ and $f_s/f_d = 0.267 \pm 0.021$.

570 **7.3 Peaking background from $B_{(s)}^0 \rightarrow h^+h'^-$ with double misen-**
 571 **tification: yield, BDT and mass PDFs**

The number of $B_{(s)}^0 \rightarrow h^+h'^-$ double misidentified events is evaluated as:

$$N_{B_{(s)}^0 \rightarrow h^+h'^- \rightarrow \mu\mu}^{\text{TRIG|SEL}} = \epsilon_{B_{(s)}^0 \rightarrow \mu^+\mu^-}^{\text{TRIG|SEL}} \frac{N_{hh}^{\text{TIS}}}{\epsilon^{\text{TIS}} \epsilon^{\text{HLT2},MC}} \epsilon_{hh \rightarrow \mu\mu} \quad (4)$$

572 where N_{hh}^{TIS} is the number of $B_{(s)}^0 \rightarrow h^+h'^-$ TIS events¹², ϵ^{TIS} is the L0 and HLT1 TIS
 573 efficiency, $\epsilon^{\text{HLT2},MC}$ is the HLT2 TOS efficiency and $\epsilon_{B_{(s)}^0 \rightarrow \mu^+\mu^-}^{\text{TRIG|SEL}}$ is the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ trigger
 574 efficiency. These efficiencies are estimated in Section 10.

575 As a result, the expected background events due to double misID of $B_{(s)}^0 \rightarrow h^+h'^-$
 576 events before *DLL* cuts is (30.4 ± 4.2) in the whole BDT range, for an integrated luminosity
 577 of 1.02 fb^{-1} . This has to be compared to 9.5 ± 1.7 on 0.37 fb^{-1} . After the *DLL* cuts the
 578 double misID events are estimated to be (5.4 ± 0.7) .

579 A second estimation of the peaking background is obtained from data as a cross-
 580 check. The average single misID rate per hadron has been computed in a sample of
 581 $B_{(s)}^0 \rightarrow h^+h'^-$ L0Tis from events passing Hlt1AllL0Track, in which one of the hadrons has
 582 been misidentified as muon, and we square that value to obtain the double misID rate.
 583 Events are selected with $\text{BDT} > 0.5$. Muon candidates are required to pass the DLL
 584 cuts. We used the $B_{(s)}^0 \rightarrow h^+h'^-$ inclusive fit to the invariant mass described in Sec. 10.
 585 The estimated single muonID fake rate per hadron ($\epsilon_{h \rightarrow \mu}$) is obtained by inverting the
 586 formula:

$$p_{h\mu/\mu h} = 2 \times \epsilon_{h \rightarrow \mu} \times (1 - \epsilon_{h \rightarrow \mu}) \sim 2 \times \epsilon_{h \rightarrow \mu} \quad (5)$$

587 where $p_{h\mu/\mu h}$ is the probability to have at least an hadron misidentified as muon:

$$p_{h\mu/\mu h} = \frac{N_{h\mu/\mu h}}{N_{h\mu/\mu h} + N_{hh}}. \quad (6)$$

588 and where $N_{h\mu/\mu h}$ is the number of $B_{(s)}^0 \rightarrow h^+h'^-$ events with an hadron misidentified
 589 as muon and N_{hh} is the number of $B_{(s)}^0 \rightarrow h^+h'^-$ events with no hadron misidentified as
 590 muon. Both N_{hh} and $N_{h\mu/\mu h}$ are extracted from a fit of the invariant mass distribution
 591 of $B_{(s)}^0 \rightarrow h^+h'^-$ which are L0Tis and come from events passing Hlt1AllL0Track in which
 592 none of the hadrons has been misidentified (Fig. 25, right) or only one of the hadrons has
 593 been misidentified (Fig. 25, left). Two methods to extract the number of $B_{(s)}^0 \rightarrow h^+h'^-$
 594 events are used: the 1D fit used in the BDT calibration (Fig. 25, top) and computing the
 595 events above the background, modeled as an exponential (Fig. 25, bottom); both methods
 596 give compatible numbers.

¹²The inclusive $B_{(s)}^0 \rightarrow h^+h'^-$ TIS events are evaluated with track in the geometrical acceptance of the muon detector, matching the definition of the above misID probabilities.

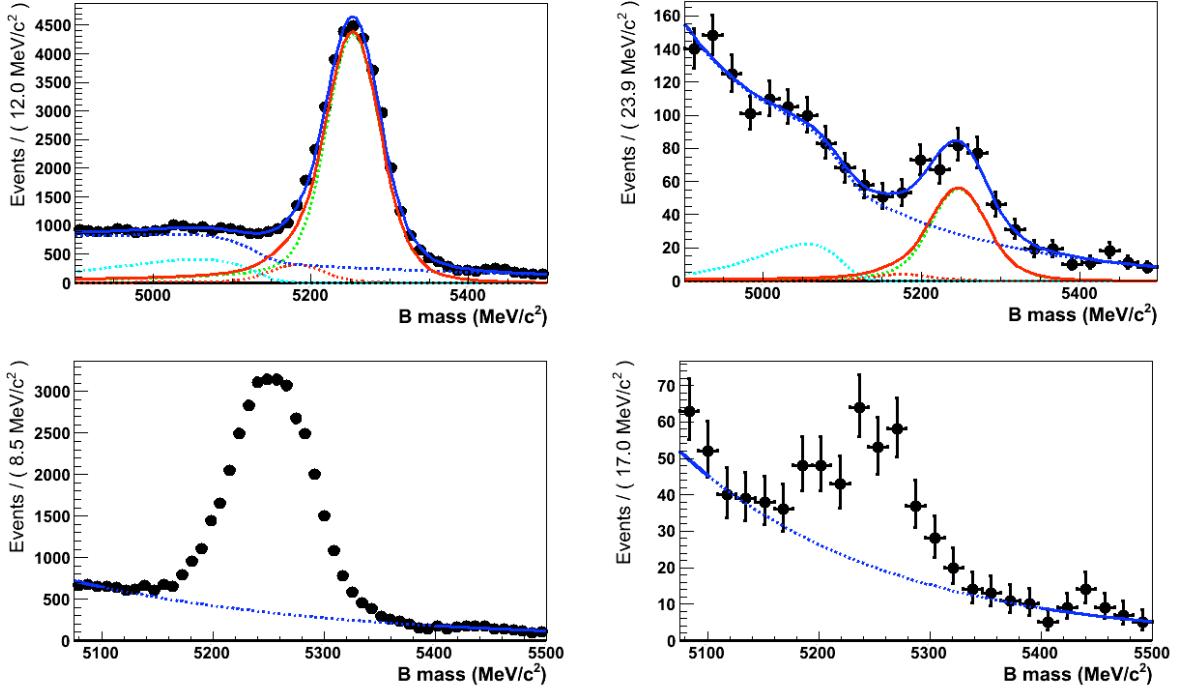


Figure 25: Left: invariant mass distribution of $B_{(s)}^0 \rightarrow h^+h'^-$ events which none of the hadrons have been identified as muon; Right: invariant mass distribution of $B_{(s)}^0 \rightarrow h^+h'^-$ events in which only one of the hadrons has been identified as muon; Top: Results of the 1D fit (same fit used for the estimation of the BDT pdf of the signal), Bottom: Fit to the background only (fit to an exponential). In all cases, B candidates are L0Tis and from events passing Hlt1AllL0Track and the muon candidate is required to pass the DLL cuts

597 The single hadron misID rate for DLL cuts (with $BDT > 0.5$) is: $\epsilon_{h \rightarrow \mu} = (0.33 \pm$
 598 $0.05_{stat} \pm 0.10_{syst}) \times 10^{-2}$. The systematic error is due to the estimation of the background
 599 contamination under the mass peak. The stability of the misID (with DLL cuts) rate vs
 600 time has been checked in two different run periods. The double misID rate (with DLL
 601 cuts) is $(1.11 \pm 0.31_{stat} \pm 0.57_{syst}) 10^{-5}$, which is in agreement with the value of misID rate
 602 obtained with the method described above in the same range of BDT, $(1.44 \pm 0.09) 10^{-5}$.

603 To take into account the dependence of double misID from the BDT bins, we evaluated
 604 it as a function of BDT bin, $\epsilon_{hh \rightarrow \mu\mu}(i)$, and we defined for each BDT bin a fractional
 605 correction as $f_{misID}(i) = \epsilon_{hh \rightarrow \mu\mu}(i)/\epsilon_{hh \rightarrow \mu\mu}$, where $\epsilon_{hh \rightarrow \mu\mu}$ is the average double misID
 606 given above; the values of $f_{misID}(i)$ are given in Table 7.3, before and after the *DLL* cuts;
 607 the values are given without errors, since they are mostly correlated among the bins. This
 608 correction is then introduced in Eq. 4, as well as the dependence of the trigger factors as
 609 a function of the BDT bin (also discussed in Sec. 9.1).

610 To estimate the number of the peaking background events that are in the search

BDT bin	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
$f_{\text{misID}}(i)$	1.17	1.15	1.12	1.07	0.99	0.86	0.74	0.66
$DLL(K - \pi) < 10$ and $DLL(\mu - \pi) > -5$								
$f_{\text{misID}}(i)$	1.04	1.07	1.06	1.07	1.04	0.95	0.87	0.81

Table 12: Ratio between double misID computed in each BDT bin and its average value, before and after the DLL cuts

611 window, the invariant mass lineshape of $B_{(s)}^0 \rightarrow h^+h'^-$ with a double decay in flight
 612 has also been evaluated, with the same method as described in the previous analysis
 613 note [22]. The slope of the tracks at the origin (which defines the angle between the
 614 tracks) is assumed to be unaffected by the decay in flight while the momentum used for
 615 the invariant mass computation is corrected using Monte Carlo simulation. The difference
 616 of the momentum of the hadron before and after the decay in flight is obtained by MC
 617 and applied to data. The invariant mass distribution for $B_{(s)}^0 \rightarrow h^+h'^-$ with double decay
 618 in flight is obtained by applying the MC momentum corrections to both hadrons, and is
 shown in Fig. 26.

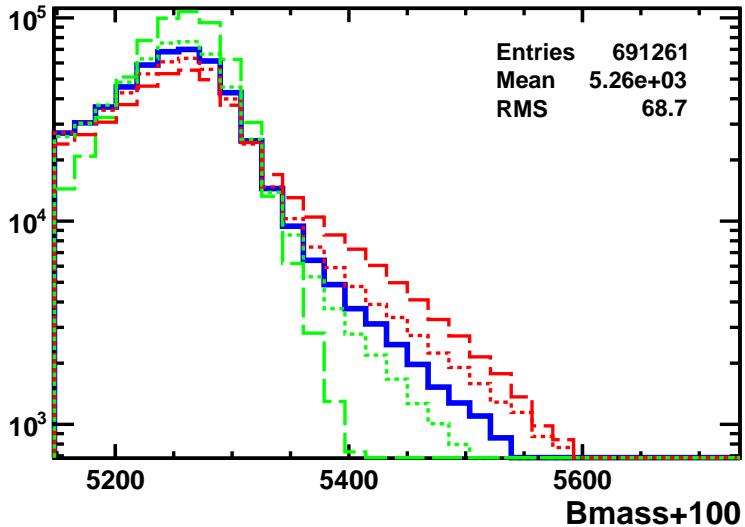


Figure 26: Invariant mass distribution calculated for $B_{(s)}^0 \rightarrow h^+h'^-$ with double misID, evaluated as described in the text: blue curve represents the central value, red and green curves the error band. Short dashed curves show how the central curves are modified by moving the track parameters by $\pm 1\sigma$ with respect to the central values. Long dashed curves represent the more conservative cases: no kink at all, just wrong mass hypothesis and mass resolution taken from data (optimistic case) and kink as big as for kaons applied also to pions (pessimistic case).

With this method, the fraction of double misID $B_{(s)}^0 \rightarrow h^+h'^-$ that are in the search windows is found to be $(8.8^{+3.0}_{-2.1})\%$ in ± 60 MeV/ c^2 around the B_s mass and $(48.0^{+20}_{-8})\%$ in ± 60 MeV/ c^2 around the B_d mass.

7.3.1 Relation between the single misidentification probability and double misidentification probability for $B_{(s)}^0 \rightarrow h^+h'^-$ peaking background

In this Section we investigate the possibility that the two misidentified hadrons from a $B_{(s)}^0 \rightarrow h^+h'^-$ decay are correlated via the sharing of hits in the muon system. This test is performed using exclusive $B_{(s)}^0 \rightarrow h^+h'^-$ Monte Carlo samples from the MC10 and MC11 productions. In Table 13 for each exclusive decay we show from column 2 to column 7 in order: the total number of events after the selection, the number of events where the positive track (π^+ or K^+) has been misidentified as a muon, the single misid efficiency, the double misid efficiency computed as the square of the single misID, the number of double misidentified events and the double misID efficiency directly extracted from the sample. From the comparison between the double misID rate computed as the square of the single misID and the double misID computed directly on the track pair, we conclude that the two hadrons misidentified are not correlated each other.

This is also proved by the Fig. 27 that shows the NShared variable for the two hadrons misidentified as muons after IsMuon=1 requirement: the NShared variable for both hadrons is peaked at zero, hence the hadrons are not misidentified for a sharing of hits in the muon system.

This study has been repeated including the effect of the DLL cuts discussed in the previous sections, $DLL(K - \pi) < 10 \& DLL(\mu - \pi) > -5$. The single misID results are reported in Table 14 for $B_s^0 \rightarrow K^+K^-$ and $B^0 \rightarrow \pi^+\pi^-$ samples of the MC10 and MC11 Monte Carlo productions. The single misID results obtained with the two Monte Carlo productions are then averaged in Table 15 and the double misID obtained as the product of the single misID is compared with the double misID obtained directly from the same dataset. The two results are in nice agreement, proving that the two hadrons do not interfere each other.

Table 13: MC11 $B_{(s)}^0 \rightarrow h^+h'^-$ exclusive samples, from column 2 to seven: total number of events after the selection, number of events where the positive track (π^+ or K^+) has been misidentified as a muon, single misid efficiency, double misid efficiency computed as the square of the single misID, number of double misidentified events and the double misID efficiency directly extracted from the sample.

channels	N_{tot}	$N_{(\pi^+, K^+) \rightarrow \mu^+}$	$\epsilon_{\text{single misID}}$ [%]	$\epsilon_{\text{single misID}}^2$ 10^{-4}	$N_{\text{double misID}}$	$\epsilon_{\text{double misID}}$ [%]
$B^0 \rightarrow K^+\pi^-$	584001	6071	1.04 ± 0.01	1.08 ± 0.02	53	0.9 ± 0.1
$B^0 \rightarrow \pi^+\pi^-$	569736	4563	0.80 ± 0.01	0.64 ± 0.02	40	0.7 ± 0.1
$B_s^0 \rightarrow K^+K^-$	533059	6363	1.19 ± 0.02	1.42 ± 0.05	76	1.4 ± 0.2
$B_s^0 \rightarrow \pi^+K^-$	141097	1553	1.10 ± 0.03	1.21 ± 0.07	12	0.9 ± 0.2

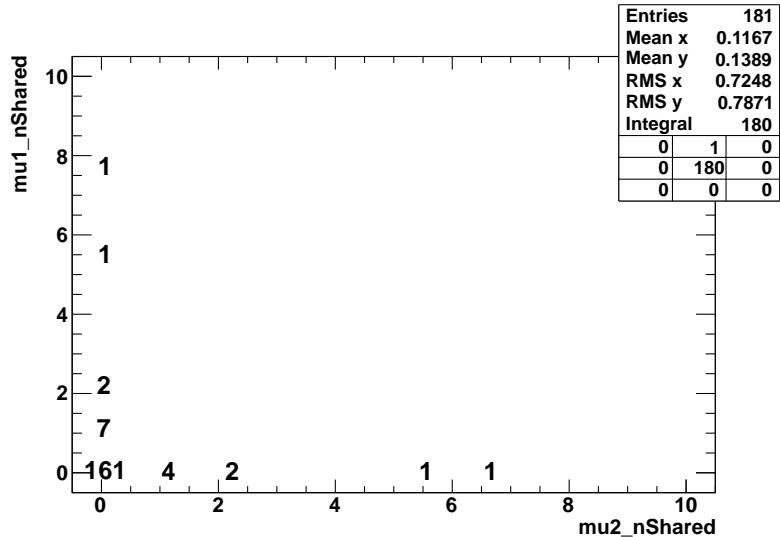


Figure 27: MC11 $B_{(s)}^0 \rightarrow h^+h'^-$ exclusive samples: NShared variables for the two hadrons misidentified as muons after IsMuon=1 requirement.

Table 14: Single misID for $B_s^0 \rightarrow K^+K^-$ and $B^0 \rightarrow \pi^+\pi^-$ samples of MC10 and MC11 productions.

channel	MC production	$\epsilon_{(\pi^+, K^+) \rightarrow \mu^+}$	$\epsilon_{(\pi^-, K^-) \rightarrow \mu^-}$
$B^0 \rightarrow \pi^+\pi^-$	MC10	$(0.655 \pm 0.015)\%$	$(0.652 \pm 0.015)\%$
	MC11	$(0.657 \pm 0.010)\%$	$(0.649 \pm 0.010)\%$
$B_s^0 \rightarrow K^+K^-$	MC10	$(0.251 \pm 0.010)\%$	$(0.232 \pm 0.009)\%$
	MC11	$(0.225 \pm 0.006)\%$	$(0.217 \pm 0.006)\%$

Table 15: $B_s^0 \rightarrow K^+K^-$ and $B^0 \rightarrow \pi^+\pi^-$ exclusive MC11 samples: double misID obtained as the product of the single misID and double misID obtained directly from the samples.

channel	Single misID (positive particle)	Single misID (negative particle)	Product single misID	Double misID
$B^0 \rightarrow \pi^+\pi^-$	$(0.656 \pm 0.009)\%$	$(0.650 \pm 0.009)\%$	$(4.265 \pm 0.079) \cdot 10^{-5}$	$(4.558 \pm 0.721) \cdot 10^{-5}$
$B_s^0 \rightarrow K^+K^-$	$(0.234 \pm 0.005)\%$	$(0.222 \pm 0.005)\%$	$(0.518 \pm 0.017) \cdot 10^{-5}$	$(0.730 \pm 0.298) \cdot 10^{-5}$

Table 16: Event statistics obtained on the $B_c^+ \rightarrow J/\psi(\rightarrow \mu^+\mu^-)\mu^+\nu$ Monte Carlo sample. Listed efficiencies include trigger and stripping, while the acceptance efficiency is already included in the initial sample.

	N	ε
Tot Events	1007188	
After stripping	4315	0.428%
of which MC associated	4216	0.419%
BDTS and fiducial cuts	1917	0.190%
of which MC associated	1900	0.188%
in Bs mass window (5358 ± 60) MeV	10	$9.43 \cdot 10^{-6}$
in Bd mass window (5272 ± 60) MeV	21	$2.085 \cdot 10^{-5}$
in sidebands range [4900 - 6000] MeV	263	$2.611 \cdot 10^{-4}$

647 7.4 Other exclusive backgrounds

648 7.4.1 Background from $B_c^+ \rightarrow J/\psi(\rightarrow \mu^+\mu^-)\mu^+\nu$

649 Among the various exclusive backgrounds the $B_c^+ \rightarrow J/\psi(\rightarrow \mu^+\mu^-)\mu^+\nu$ appears to be
650 one of the most dangerous. The muon from the B_c^+ semileptonic decay and the oppositely
651 charged muon from the J/ψ can be combined together to form a fake $B_s^0 \rightarrow \mu^+\mu^-$ decay.
652 While all the other B semileptonic decays, due to the missing particles, fall outside the
653 signal region, the large B_c^+ mass can lead the dimuon to lie in it. Moreover, the two
654 muons come exactly from the same decay vertex so that the combination cannot be
655 discarded by the vertex fit or the DOCA between the tracks. Furthermore, this decay
656 could influence the shape of the left sideband leading to an incorrect estimate of the
657 combinatorial background. For these reasons a study has been done with a Monte Carlo
658 sample of this decay.

659 The different signal selection steps have been applied to this sample. The statistics
660 obtained in the various steps is reported in Table 16.

The number of expected fake $B_s^0 \rightarrow \mu^+\mu^-$ events coming from $B_c^+ \rightarrow J/\psi(\rightarrow \mu^+\mu^-)\mu^+\nu$ is calculated as follows:

$$N^{exp} = \mathcal{L} \cdot \sigma(pp \rightarrow b\bar{b}) \cdot 2 \cdot f_c \cdot \mathcal{B}(B_c^+ \rightarrow J/\psi\mu^+\nu) \cdot \mathcal{B}(J/\psi \rightarrow \mu^+\mu^-) \cdot \varepsilon \quad (7)$$

where the efficiency includes reconstruction and selection (from Table 16) and the acceptance efficiency, which has been measured from the Monte Carlo generator to be $\varepsilon_{acc} = 0.1389 \pm 0.0003$. The considered decay branching ratio multiplied by the fragmentation fraction $b \rightarrow B_c$ has been measured by the CDF collaboration in an inclusive way and considering both a muon and an electron in the final state [74]; the measured value is:

$$f_c \cdot \mathcal{B}(B_c^+ \rightarrow J/\psi\ell^+\nu X) = 5.2_{-2.1}^{+2.4} \cdot 10^{-5}$$

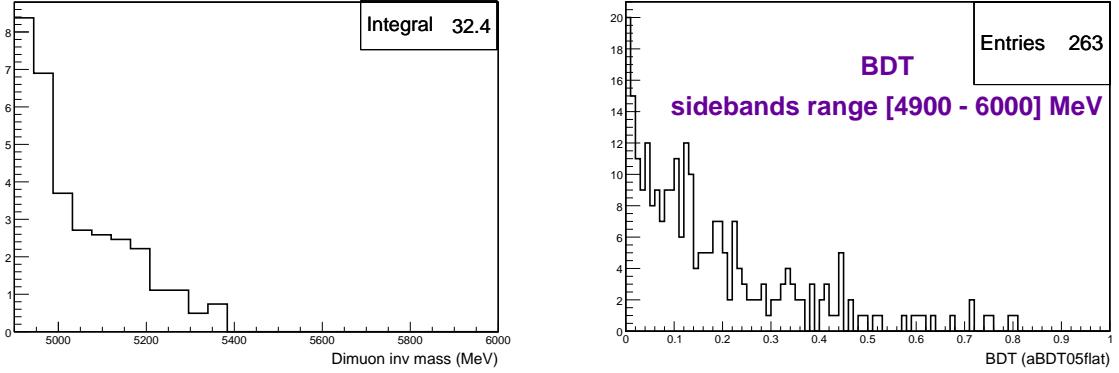


Figure 28: The invariant mass distribution of fake $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ candidates from $B_c^+ \rightarrow J/\psi(\rightarrow \mu^+ \mu^-) \mu^+ \nu$ scaled to the expected statistics as obtained in Monte Carlo simulations is shown in (a) and the BDT distribution obtained from the same events (but not scaled) is shown in (b). See text for a detailed explanation of these plots.

in the absence of more information, we follow the authors of this measurement, applying a multiplicative factor 1/2 for the muon channel, and we consider negligible the non exclusive decays, all this leading to:

$$f_c \cdot \mathcal{B}(B_c^+ \rightarrow J/\psi \mu^+ \nu) = 0.5 \cdot 5.2^{+2.4}_{-2.1} \cdot 10^{-5} .$$

Considering the cross-section $\sigma(pp \rightarrow b\bar{b}) = (284 \pm 20 \pm 49)\mu\text{b}$ measured by LHCb[11] and the branching ratio $\mathcal{B}(J/\psi \rightarrow \mu^+ \mu^-) = (5.93 \pm 0.06) \cdot 10^{-2}$ [59], we have computed a total expected number of background events of 234 ± 117 in the full stripping mass region; this number falls to 32 ± 16 inside the mass region [4900, 6000] MeV. Inside the B_s^0 and B_d^0 mass windows we estimate 1.23 ± 0.73 and 2.59 ± 1.41 events respectively. Note that these estimates are ultimately limited by the MC statistics (which corresponds approximately to about 10fb^{-1}) but even more by the error on B_c^+ production and branching ratio. The invariant mass distribution of the fake $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ candidates from $B_c^+ \rightarrow J/\psi(\rightarrow \mu^+ \mu^-) \mu^+ \nu$, scaled at the expected value, is shown in Fig. 28 (a).

The distribution of the BDT variable of the selected events in the invariant mass region [4900, 6000] is shown, not normalized, in Fig. 28(b). As it can be seen this distribution peaks at low values of BDT so that the expected events in the B_s^0 and B_d^0 mass windows are likely to lie in the low BDT region. Taking a BDT value larger than 0.8, only one MC event is left in the B_d^0 signal region, leading to an expected number of events of 0.12 ± 0.13 , while no events are left in the B_s^0 mass window, leading to negligible expected number of events.

Therefore we consider the $B_c^+ \rightarrow J/\psi(\rightarrow \mu^+ \mu^-) \mu^+ \nu$ decay as neglectable in the signal region, for the present statistics. Finally, its invariant mass distribution in the sidebands exponentially distributed, so that this decay is largely included in the combinatorial background extrapolation.

681 **7.4.2 Background from $B^+ \rightarrow \pi^+ \mu^+ \mu^-$**

682 The decay $B^+ \rightarrow \pi^+ \mu^+ \mu^-$ is a very rare decay recently discovered by LHCb [65], with
 683 a branching fraction $\mathcal{B}(B^+ \rightarrow \pi^+ \mu^+ \mu^-) = (2.4 \pm 0.6 \pm 0.2) \cdot 10^{-8}$. The two muons from
 684 this decay could be a background to the $B^0 \rightarrow \mu^+ \mu^-$ and $B_s^0 \rightarrow \mu^+ \mu^-$ decays. While,
 685 due to the missing pion, the $\mu\mu$ invariant mass will not be in the signal windows, it could
 686 influence the evaluation of the combinatorial background.

The influence of the $B^+ \rightarrow \pi^+ \mu^+ \mu^-$ to this analysis was studied with MC simulations (MC11 production). The signal decay is generated according to the BTOSLLBALL parametrisation. The total number of expected events as background for $B_s^0 \rightarrow \mu^+ \mu^-$ is:

$$N_{B^+ \rightarrow \pi^+ \mu^+ \mu^-}^{bkg} = \mathcal{L} \cdot 2 \cdot \sigma(pp \rightarrow b\bar{b}) \cdot f_c \cdot \mathcal{B}(B^+ \rightarrow \pi^+ \mu^+ \mu^-) \cdot \varepsilon \quad (8)$$

Assuming the already mentioned luminosity and $b\bar{b}$ production cross-section, and considering the B^+ production fraction from PDG, $f_u = (40.3 \pm 1.1)$, the total number of produced N_{B^+} is:

$$N_{B^+} = 2.335 \cdot 10^{11}$$

The efficiencies for $B^+ \rightarrow \pi^+ \mu^+ \mu^-$ as obtained from MC are reported in Table 17. In summary, after all the cuts (trigger, stripping, fiducial cuts, BDTS and μ -ID) we expect

$$N_{B^+}^{bkg} = N_{B^+} \cdot \mathcal{B}(B^+ \rightarrow \pi^+ \mu^+ \mu^-) \cdot \varepsilon = 12.7$$

687 events due to $B^+ \rightarrow \pi^+ \mu^+ \mu^-$ in the full invariant mass region and full BDT range. In
 688 the B_s^0 signal mass window no MC events are found, while in B_d^0 we expect about 0.001
 689 events.

690 With present statistics also the events in the sidebands are negligible (and included in
 691 the combinatorial background) however with higher statistics it could be necessary to take
 692 this background into account. In fact some of the events peak around $m_{\mu\mu} = 5030$, which
 693 correspond to the ones in which the pion is created at rest in the B^+ frame. The invariant
 694 mass distribution for the different analysis steps is shown in Fig. 30(a), not scaled to the
 695 expected values, while in Figure 30(b) the distribution of $m_{\mu\mu}$ scaled to the number of
 696 expected events is shown as obtained after all cuts (but for the full BDT range).

697 The $B^+ \rightarrow \pi^+ \mu^+ \mu^-$ distribution as a function of BDT is shown in Fig. 29(a) not scaled
 698 to the expected events. The $B^+ \rightarrow \pi^+ \mu^+ \mu^-$ events are divided in two components: events
 699 which are in the tail of the $\mu\mu$ invariant mass behave as combinatorial background while
 700 the ones which accumulate around $m_{\mu\mu} \sim 5030 MeV/c^2$ behave as a peaking background.
 701 This can be seen in Fig. 29(b) where the two dimensional distribution of these events
 702 versus dimuon invariant mass and BDT is shown. For $BDT > 0.5$ we expect a total of
 703 2.87 events of which 0.229 in the sideband region $[4900, 5000] MeV/c^2$. No MC events are
 704 found in the signal invariant mass regions after this cut.

705 As a final note, the isospin-conjugate mode $B^0 \rightarrow \pi^0 \mu^+ \mu^-$ has also to be considered;
 706 the branching fraction should be approximately the same of the charged one and so should
 707 be the efficiencies. Therefore we can assume roughly a factor two for the numbers listed
 708 in this paragraph and still this background is clearly negligible for the present statistics.

Table 17: Efficiencies for the $B^+ \rightarrow \pi^+\mu^+\mu^-$ selected as $B_s^0 \rightarrow \mu^+\mu^-$. Note: the stripping numbers include also acceptance and reconstruction.

	N	ε	ε_{rel}
Tot	766200		
Stripping	19157	0.0250026	
Trigger	17847	0.0232929	0.928612
Fiducial cuts	17425	0.0227421	0.976355
BDTS	11506	0.015017	0.660316
DLL cuts	11392	0.0148682	0.990092
$BDT > 0.5$	2583	0.00337118	0.226738

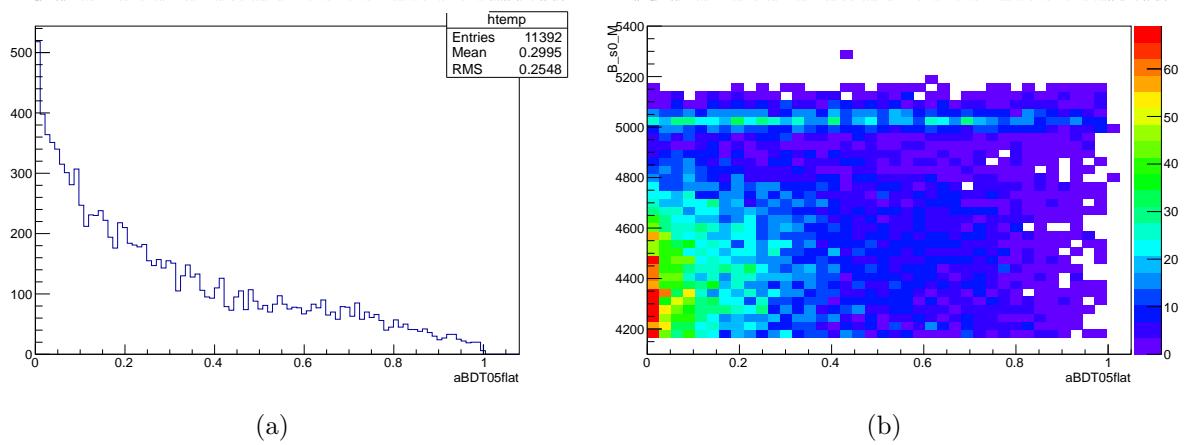


Figure 29: BDT distribution for $B^+ \rightarrow \pi^+\mu^+\mu^-$ after all cuts (not normalised) Bi-dimensional distribution of the $B^+ \rightarrow \pi^+\mu^+\mu^-$ MC events as a function of the dimuon invariant mass and the BDT.

7.4.3 Background from $B_s \rightarrow \mu^+\mu^-\gamma$

A possible background contribution from $B_s \rightarrow \mu^+\mu^-\gamma$ has also been considered for this analysis. In this case, the categorization of this decay as “background” depends on the origin of the photon at Feynman diagram level. In this way, three different contributions can be accounted for: photons from initial state radiation (ISR), photons from final state radiation (FSR) and photons coming from an interference between FSR and ISR. ISR and the interference between ISR and FSR can be considered as a background for $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$, while FSR is essentially part of the signal being looked for.

It is also remarkable that the importance of this background will be very dependent on the branching ratio of $B_s \rightarrow \mu^+\mu^-\gamma$ and on the mass range in which $B_s \rightarrow \mu^+\mu^-\gamma$ lies. For the former, the branching ratio is not known, but is currently thought to be ~ 10

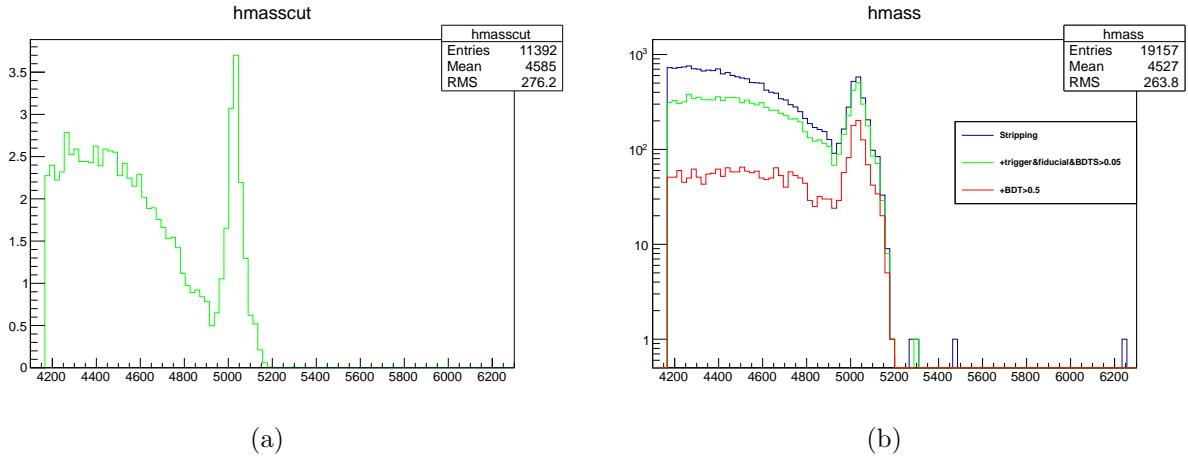


Figure 30: Di-muon invariant mass distribution from $B^+ \rightarrow \pi^+ \mu^+ \mu^-$ events normalised to the expected events

times larger than that of $B_s^0 \rightarrow \mu^+ \mu^-$ [75]. For the latter, it will vary with the momentum spectrum of the photon, which is difficult to parametrize, and also depending on the origin of the photon (ISR, FSR or interference). For this study, a new parametrization for the photons prepared by N. Nikitin for the MC11 simulation has been used. It must be said that this MC11 simulation includes, at the same time, all the three mentioned contributions for the photon origin, so that all the results shown are considered to be upper limits for this exclusive background computation, since the contribution from FSR should not be taken into account.

In order to estimate the contribution of this background, two samples of MC11 simulation have been employed. For $B_s \rightarrow \mu^+ \mu^- \gamma$, as just explained, a set of specially generated 1.66 M events has been used. At the same time, 0.21 M $B_s^0 \rightarrow \mu^+ \mu^-$ events have also been studied. With this, the reconstruction, selection and trigger efficiencies have been obtained purely on MC. For the selection, the cuts explained in section 5 have been applied. For the trigger, L0, Hlt1Physics and Hlt2Physics decisions were required. On top of this, the focus has been put in last two BDT bins ($BDT > 0.8$). The BDT is a MVA variable used to classify the candidates which pass all the previous selections. Signal is expected to be flat in BDT while combinatorial background is expected to peak at 0. This will be seen in detail in section 8. The mass spectrum of the simulated $B_s \rightarrow \mu^+ \mu^- \gamma$ signal after the mentioned requirements is shown in figure 31 for the B_s^0 and B^0 mass regions. The total efficiencies obtained for $B_s^0 \rightarrow \mu^+ \mu^-$, $B^0 \rightarrow \mu^+ \mu^-$ and $B_s \rightarrow \mu^+ \mu^- \gamma$ in both mass regions are shown in table 18, together with the corresponding ratios. The generator efficiencies are considered to be the same for all the channels. For $B^0 \rightarrow \mu^+ \mu^-$, the efficiency has been considered the same as for $B_s^0 \rightarrow \mu^+ \mu^-$.

These efficiency ratios can now be used to obtain the expected yields of $B_s \rightarrow \mu^+ \mu^- \gamma$ in the B^0 and B_s^0 mass regions after assuming a particular branching ratio for this decay. In this sense, it must be said that the number of SM $B_s^0 \rightarrow \mu^+ \mu^-$ ($B^0 \rightarrow \mu^+ \mu^-$) candidates

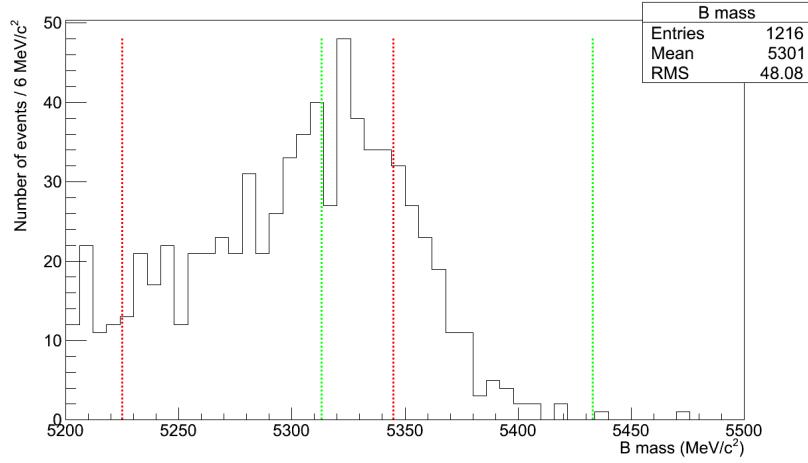


Figure 31: The invariant mass distribution of fake $B_{d,s}^0 \rightarrow \mu^+\mu^-$ candidates from $B_s \rightarrow \mu^+\mu^-\gamma$ after selection cuts, trigger and $\text{BDT}>0.8$ (see text for details). Red vertical bars bound the B^0 mass region, while green vertical bars bound the B_s^0 mass region.

Table 18: Total efficiencies of $B^0 \rightarrow \mu^+\mu^-$, $B_s^0 \rightarrow \mu^+\mu^-$ and $B_s \rightarrow \mu^+\mu^-\gamma$ after selection, trigger and $\text{BDT}>0.8$ in B_s^0 and B^0 mass regions (see text for details). The generator efficiency of the three channels is considered to be the same. For $B^0 \rightarrow \mu^+\mu^-$, the same efficiency than $B_s^0 \rightarrow \mu^+\mu^-$ is assumed. The efficiency ratios are also shown.

	B_s^0	B^0
$\epsilon(B_s \rightarrow \mu^+\mu^-\gamma)$	$(0.0197 \pm 0.0011)\%$	$(0.0325 \pm 0.0014)\%$
$\epsilon(B_{d,s} \rightarrow \mu\mu)$	$(6.7475 \pm 0.0547)\%$	—
Efficiency ratio	0.0029 ± 0.0002	0.0048 ± 0.0002

expected in the region of $\text{BDT}>0.8$ in the currently analyzed data sample is ~ 2.0 (~ 0.2). These numbers come from the use of the normalization factors that will be obtained in section 10. Using now the mentioned yields, the upper limit for the expected $B_s \rightarrow \mu^+\mu^-\gamma$ background in the B_s^0 mass region is given by:

$$N(B_s \rightarrow \mu^+\mu^-\gamma) < \frac{\epsilon(B_s \rightarrow \mu^+\mu^-\gamma)}{\epsilon(B_s^0 \rightarrow \mu^+\mu^-)} \times \frac{\mathcal{B}(B_s \rightarrow \mu^+\mu^-\gamma)}{\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-)} \times N(B_s^0 \rightarrow \mu^+\mu^-) \quad (9)$$

with a very similar equation for the upper limits in the B^0 mass region. With all this, and assuming $\mathcal{B}(B_s \rightarrow \mu^+\mu^-\gamma) = 3.2 \times 10^{-8}$, the upper limits are of 0.058 (0.096) candidates in the B_s^0 (B^0) mass region. These upper limits are smaller than the peaking background (as seen previously in this section) and almost negligible when compared to

754 the combinatorial background extrapolated from the sidebands (as it will be seen in section
755 9). In fact, these limits are also small when compared to the errors in the combinatorial
756 background expectations, so any potential contribution would be well covered by them.

757 In summary, the $B_s \rightarrow \mu^+ \mu^- \gamma$ background has been checked to be under control for
758 this level of statistics, even when its contribution may become more relevant in future
759 searches, particularly for the $B^0 \rightarrow \mu^+ \mu^-$ case.

760 8 The Boosted Decision Tree

761 The number of observed events after the selection is compared to the number of expected
762 signal and background events in bins of two independent variables, the invariant mass and
763 the output of a multivariate discriminant, the Boosted Decision Tree (BDT) constructed
764 using the TMVA package [9]. The same BDT used for the analysis of the 370 pb^{-1} dataset
765 is used in the current analysis. Several other multivariate discriminants have been tested
766 but none of them have shown better performance than the old one. The binning of the
767 BDT and invariant mass distributions has been, instead, fully reoptimized. This will be
768 discussed in Section 8.2.

769 8.1 BDT definition and performance

770 The BDT operator used in the current analysis is built with the nine variables used for
771 the 370 pb^{-1} analysis: the B proper time (t), minimum impact parameter significance
772 of the muons ($\text{IPS}(\mu)$), the impact parameter of the B ($\text{IP}(B)$), the distance of closest
773 approach between the two muons (DOCA), the isolation of the two muons with respect
774 to any other track in the event (μ isolation, $I(\mu)$), the transverse momentum of the B
775 ($p_T(B)$), the cosine of the angle between the muon momentum in the B rest frame and
776 the vector perpendicular to the B momentum and the beam axis ($\cos P$), the B isolation
777 (CDF definition [16]) $I(B)$ and the minimum $p_{T,\min}(\mu)$ of the two muons. The description
778 of the nine variables is done in details in [23]. Only two variables (DOCA and $\text{IP}(B)$)
779 are in common with the BDTS discriminant (Section 5.1).

780 As in published analysis, the BDT training has been done using $B_s^0 \rightarrow \mu^+ \mu^-$ and
781 $b\bar{b} \rightarrow \mu\mu X$ simulated events which have passed the cuts listed in Table 7 including the
782 BDTS cut but excluding the DLL cuts.

783 Differently from the published analysis, the simulated events used for the training are
784 required also to pass the trigger at each level¹³. This allows to train the BDT with events
785 covering a similar phase space as real data: in fact the trigger efficiency on $B_s^0 \rightarrow \mu^+ \mu^-$
786 selected events is $\sim 87\%$ while on $b\bar{b} \rightarrow \mu\mu X$ selected events is $\sim 85\%$, hence the trigger
787 rejects about 15 % of both samples.

788 Half of the samples of simulated events is used to train the BDT and the second
789 half is used to measure its performance. Then the two halves are swapped (the half
790 previously used to train is used to test and vice-versa) and the process is repeated. The
791 total performance is then obtained by merging the two datasets. In this way the full
792 Monte Carlo statistics can be used to train the discriminant and measure the performance
793 without risking to do an overtraining. As in the published analysis no data are used to
794 train or test the BDT in order to avoid any bias in the result.

795 The background rejection versus signal efficiency curves are shown in Fig. 32 for dif-
796 ferent BDTS cuts applied at the selection level: the curves are zoomed in the region that

¹³The condition $\text{L0GlobalPhys} \& \text{HLT1GlobalPhys} \& \text{HLT2GlobalPhys} = \text{TRUE}$ has to be satisfied.

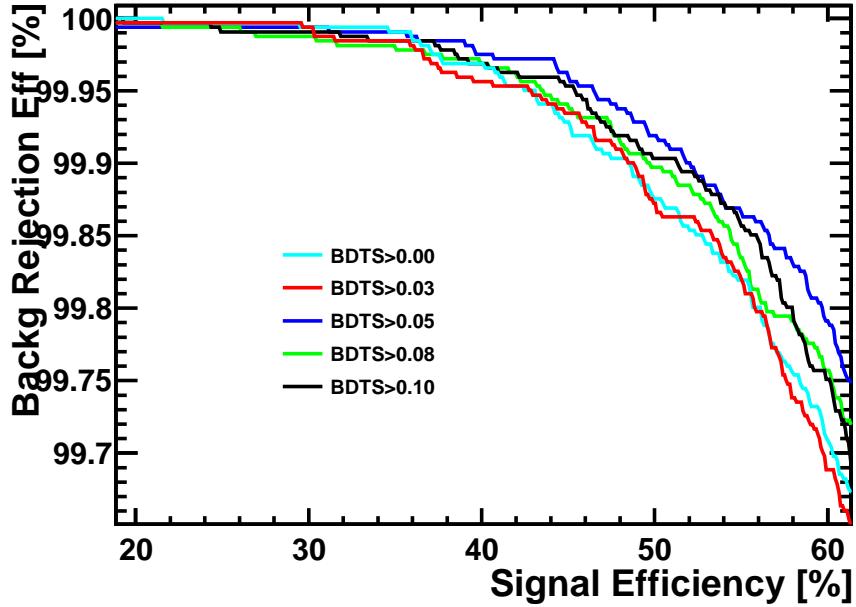


Figure 32: BDT performance: rejection versus efficiency curves for different BDTS cuts for $B_s^0 \rightarrow \mu^+ \mu^-$ and $b\bar{b} \rightarrow \mu\mu X$ Monte Carlo events.

gives most of the sensitivity. As already discussed in Section 5, the curve used in the current analysis is the one corresponding to BDTS> 0.05.

Two other combinations of variables entering the BDT have been tested, one with 13 variables (BDT13) and the other with 19 variables (BDT19). The list of variables of these two new combinations (BDT13 and BDT19) is shown here below together with the one with 9 variables used as baseline. The variables in bold are in common with the BDTS.

- **BDT or BDT9 (baseline):** $IP(B)$, $t(B)$, **DOCA**, $I(B)$, $I(\mu_1) + I(\mu_2)$, $\min IP\chi^2(\mu)$, $p_T(B)$, $\cos P$, $p_{T,min}(\mu)$;
- **BDT13:** $IP\chi^2(B)$, $t(B)$, **DOCA**, $I(B)$, $I(\mu_1)$, $I(\mu_2)$, $\chi^2/nDoF$ of the secondary vertex ($\chi^2(B)/nDoF$), cosine of the angle formed by the direction of the B momentum and the direction defined by the primary and secondary vertices (**DIRA**), B flight distance (**BFD**) and B flight distance significance (**BFDS**), $\min IP(\mu)$, $\cos(P)$;
- **BDT19:** $IP\chi^2(B)$, $t(B)$, **DOCA**, $I(B)$, $I(\mu_1)$, $I(\mu_2)$, $\chi^2(B)/nDoF$, **DIRA**, **BFD**, **BFDS**, $\min IP(\mu)$, $\cos(P)$, azimuthal angle between the two muons (ϕ), the polar angle between the two muons (θ), the angle between the μ^+ direction in the rest frame of the B candidate and the direction of the *thrust momentum*¹⁴ also boosted

¹⁴The *thrust* momentum should approximate the direction of the other B hadron for a signal event. It is defined as the sum of the momenta of all long tracks with IPS with respect to the PV associated to the B candidate between 4 and 40, $200 \text{ MeV}/c < p_T < 2000 \text{ MeV}/c$, $p < 30 \text{ GeV}/c$

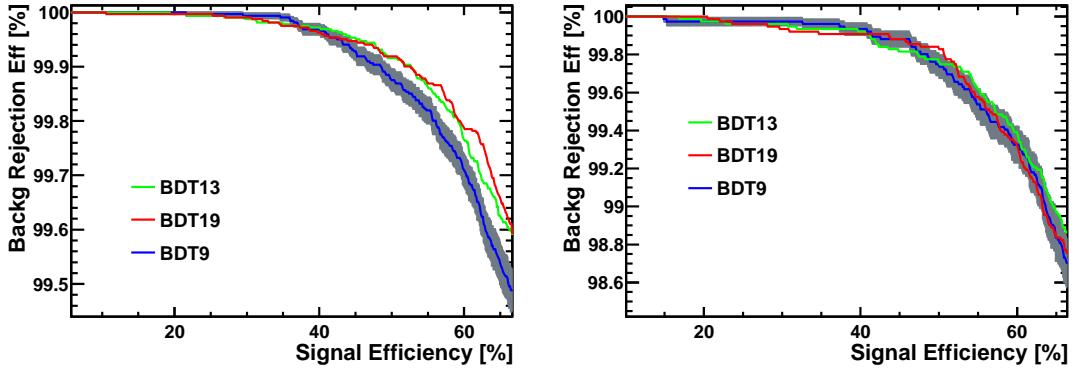


Figure 33: Rejection versus efficiency curves for the three different combinations of variables entering the BDT without (left) and with (right) the BDTS cut $BDTS > 0.05$.

in the same frame (*otherBboost*), the angle between the B candidate momentum and the direction of the trust angle (*otherBangle*), $IP(B)$, $p_T(\mu)$.

The rejection versus efficiency curves for the three different combinations of variables entering the BDT are shown in Fig. 33 without (left) and with (right) the BDTS cut $BDTS > 0.05$: without applying the BDTS cut in the selection, BDT13 and BDT19 perform better than the baseline BDT. However, when the BDTS cut is applied, the three configurations are equivalent.

This is due to the fact that most of the new variables introduced in the BDT13/BDT19 overlaps with the variables entering the BDTS¹⁵ and therefore most of the gain is already achieved at the previous stage. This is not true for the baseline BDT that really profit from the discrimination introduced by the BDTS.

Figure 34 shows the correlation matrix among the nine variables entering the BDT for signal (left) and background (right) MC events. Figures 35 show the distributions of the height out of the nine variables in input to the BDT for signal $B_s^0 \rightarrow \mu^+ \mu^-$ and $b\bar{b} \rightarrow \mu\mu X$ Monte Carlo events and for dimuon background events from data sidebands. The distributions agree nicely and prove that the dominant background component is the $b\bar{b} \rightarrow \mu\mu X$ one.

¹⁵The variables entering the BDTs are: $IP(B)/\sigma(IP)$, DOCA, $\chi^2/nDoF(B)$, DIRA, $IPS(\mu)$ and $IP(B)$.

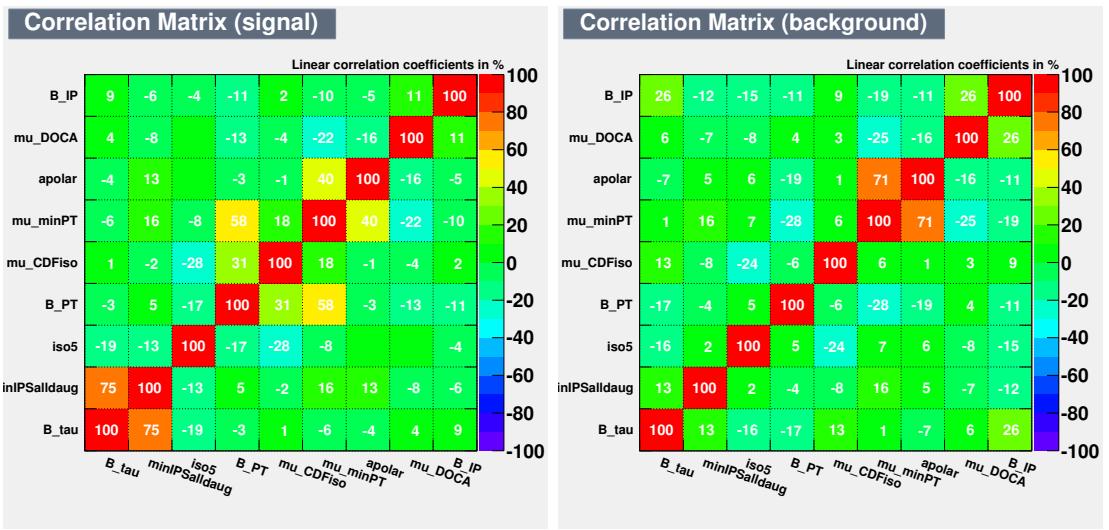


Figure 34: Correlation matrix among the nine variables entering the BDT definition for signal (left) and $b\bar{b} \rightarrow \mu\mu X$ (right) MC events.

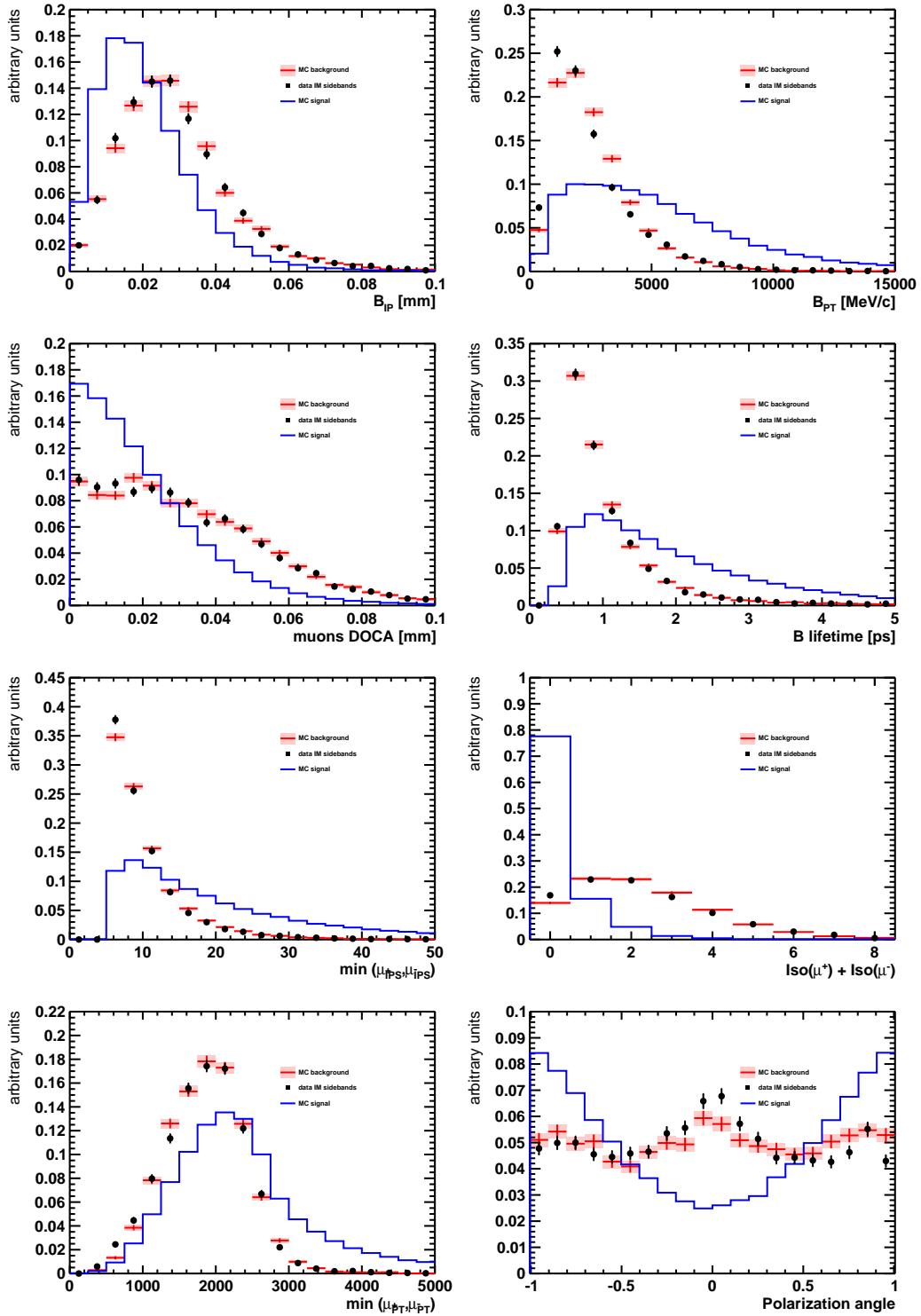


Figure 35: Distribution, for signal and background, of the variables used in the BDT definition. Blue (black): MC $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ signal. Red (grey): MC $b\bar{b} \rightarrow \mu\mu X$.

831 **8.2 Binning of the BDT and Invariant Mass distributions**

832 The extraction of the limit is performed by doing a N-counting experiment, in which the
833 informations given by the various bins in the BDT and invariant mass distributions are
834 combined together using a modified frequentist approach. Each bin can be considered as
835 an independent experiment. The number and the size of the bins used in the analysis of
836 the 370 pb^{-1} dataset were the same optimized for the analysis based on 37 pb^{-1} in 2010:
837 the BDT PDF was divided into four equally spaced bins and the mass PDF was divided
838 into six equally spaced bins.

839 The aim of this Section is to investigate which is the binning that maximizes the
840 sensitivity for the current analysis. This study is described in detail in [13]. Note that
841 the central value (i.e. most probable) of the extracted number of observed signal events
842 do not depend on the binning, however its uncertainty, and therefore the upper limits,
843 will depend on it.

844 The compatibility of the distribution of events inside each bin of the search window
845 in the invariant mass-BDT plane with a given branching fraction hypothesis is evaluated
846 using the CL_s method. This method provides three estimators: CL_{s+b} , a measure of the
847 compatibility of the observed distribution with the signal-plus-background hypothesis,
848 CL_b , a measure of the compatibility with the background-only hypothesis and CL_s , a
849 measure of the compatibility of the observed distribution with the signal-plus-background
850 hypothesis normalized to the background-only hypothesis. These estimators are related
851 to the classifier used to do the test statistic Q defined as:

$$Q = \prod_{i=1}^N P(d_i, s_i(\mathcal{B}) + b_i) / P(d_i, b_i) \quad (10)$$

852 where the products runs over the number of bins, P is the Poisson distribution with
853 expected values $s_i + b_i$ and b_i respectively, s_i and b_i denote the expected number of
854 signal and background events in bin i and d_i is the number of observed events in bin i .
855 The number of expected signal events depends on the assumed branching fraction and
856 integrated luminosity, the number of expected background events and observed events
857 depend on the integrated luminosity.

858 The distributions of Q for the signal-plus-background hypothesis (Q_{s+b}) and
859 background-only (Q_b) hypothesis are shown in Fig. 36 (left). The quantity CL_{s+b} is
860 the integral of the Q_{s+b} distribution from the observed value upwards, hence represents
861 the probability that another experiment gives a lower likelihood than the observed one
862 under the hypothesis of signal-plus-background. CL_{s+b} is a measure of the incompatibility
863 with the signal-plus-background hypothesis. The quantity CL_b is the integral of the Q_b
864 distribution from the observed value upwards, hence the quantity $1-\text{CL}_b$ represents the
865 probability that another experiment gives a lower likelihood than the observed one under
866 the hypothesis of background-only. $1-\text{CL}_b$ is a measure of the incompatibility with the
867 background hypothesis.

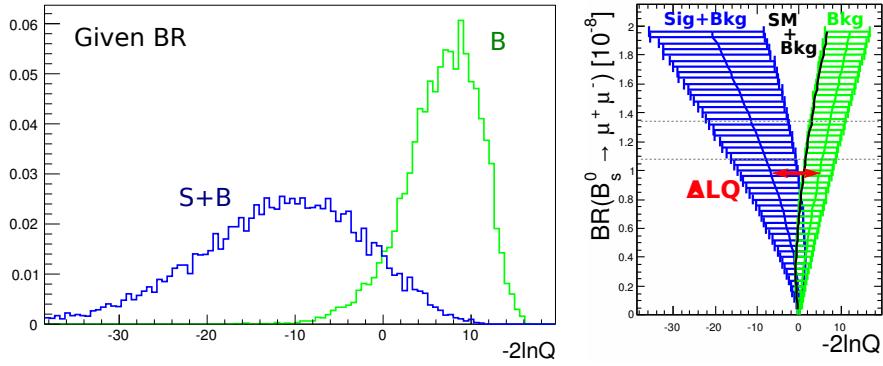


Figure 36: Left: distributions of $-2\ln Q$ for signal-plus-background hypothesis (Q_{s+b}) and background-only (Q_b) hypothesis. Right: separation of the medians of the two distributions as a function of the $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction.

868 The best sensitivity corresponds to the largest separation of the two distributions for
 869 a given signal yield (hence a given branching fraction hypothesis). The separation is
 870 quantified by the variable:

$$\Delta LQ = 2 \ln Q_{s+b}^{\text{med}} - 2 \ln Q_b^{\text{med}} \quad (11)$$

871 where Q_{s+b}^{median} and Q_b^{median} are the medians of the Q distributions calculated in the
 872 signal plus background and background-only hypotheses. The separation between the
 873 two distributions as a function of the branching fraction hypothesis is shown graphically
 874 in Fig. 36 (right): for a given branching fraction, the error around the median gives the
 875 68% containment of possible results. Intuitively the more the two bands are separated
 876 the larger is the sensitivity.

877 A (coarse) application of the Wilk's theorem [14] leads to interpret ΔLQ as a χ^2
 878 that measures the compatibility of the signal-plus-background distribution with the
 879 background-only hypothesis. This χ^2 can be read as a statistical significance: for ex-
 880 ample, a $\Delta LQ = 9$ for a given branching fraction hypothesis means that the observations
 881 are not compatible with the background-only hypothesis at 3σ .

882 If the number of bins is increased the statistical error will in principle decrease. Note
 883 however that the sensitivity does not improve when the number of expected events is
 884 smaller than 1 in each bin of the sensitive region. Moreover the binned BDT must be
 885 calibrated using control samples hence each BDT bin has to be large enough to contain a
 886 reasonable amount of events. The choice of the binning has to take into account all these
 887 constraints.

888 The choice of the binning is made out of a comparison of many set of bins on their
 889 ΔLQ . A set of N boundaries is fixed for each variable, defining for each of them a grid.
 890 For all possible $\binom{N}{n_b}$ sets of n_b -bins that can be made on this grid, the separation ΔLQ
 891 between Q_{s+b}^{med} and Q_b^{med} is computed. For this given number of bins n_b , the configuration

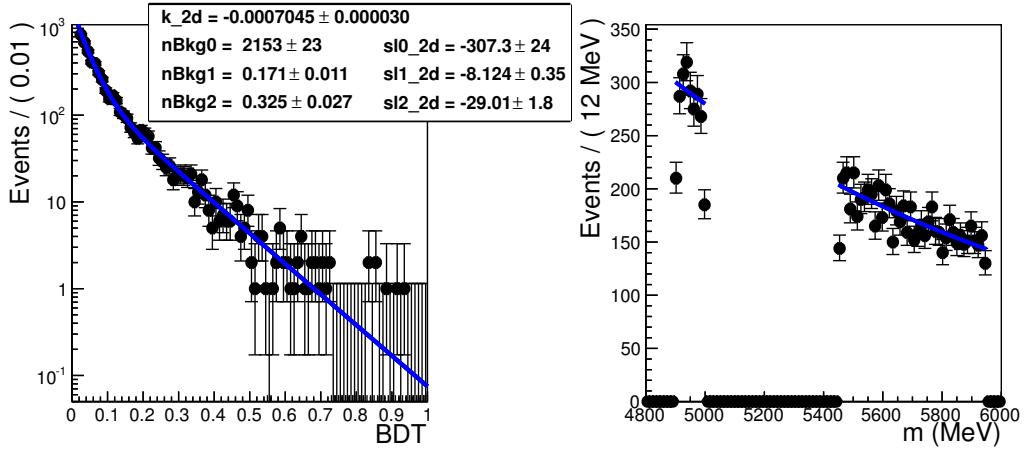


Figure 37: Simultaneous fit of the BDT and mass distributions of events in the mass sidebands [4.9-5.0] GeV/c^2 and [5.45-5.95] GeV/c^2 .

which gives the largest ΔLQ is chosen. The process is then repeated by increasing the number of bins.

The binning is optimized for the $B_s^0 \rightarrow \mu^+ \mu^-$ decay, assuming an integrated luminosity of 1 fb^{-1} and assuming the $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ from Standard Model predictions. The BDT PDF is assumed to be flat for the signal, the mass PDF for signal is a Crystal Ball whose parameters will be discussed in Sections 9.2-9.5, the number of generated events is linked to a given branching fraction via the normalization factor discussed in Section 10. The BDT and mass PDFs of the background are obtained by fitting simultaneously the BDT and mass distributions of events in the mass sidebands [4.9-5.0] GeV/c^2 and [5.45-5.95] GeV/c^2 . The simultaneous fit to the mass and BDT distribution of the background is shown in Fig. 37. The BDT is fitted with the sum of three exponential functions while the mass with a single exponential.

Possible biases coming from the small number of events in the high BDT range are checked by using the fitted function to generate toy monte carlo and repeating the fit. In Fig. 38 the pull distributions of the first (left) and the second (center) slope of the exponential function used to fit the BDT distribution and of the slope of the exponential function used to fit the mass distribution (right) are shown. The good quality of the pulls shows that the fit is stable against statistical fluctuations in the high BDT range.

The binning optimization is done in two steps: first the mass binning is optimized for events with BDT between 0.8 and 1.0. This optimization is done on a grid of 6 MeV wide bins. For a fixed number of bins, $O(2000\text{--}3000)$ different configurations are tested and the pattern which gives the largest ΔLQ is recorded. In Fig. 39 (top) the best patterns corresponding to a given number of bins are shown together with the corresponding value of ΔLQ .

The ΔLQ value as a function of the number of mass bins is shown in Fig. 40 After 9

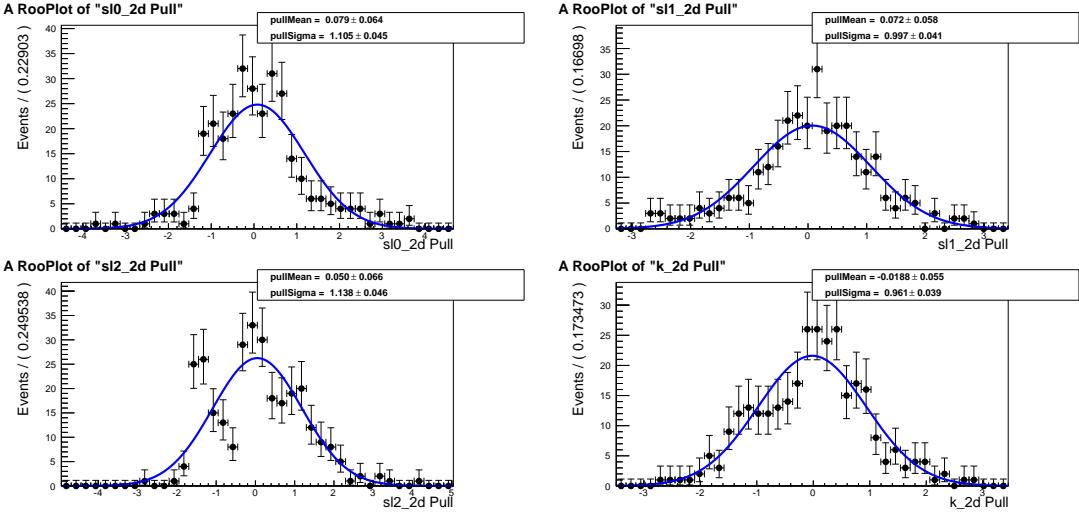


Figure 38: Pulls of the first exponential slope of the BDT PDF (left), second exponential slope of the BDT PDF (center) and exponential slope of the mass PDF (right).

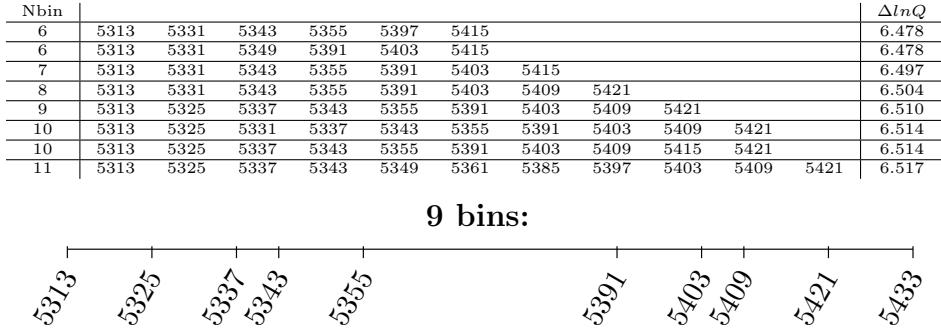


Figure 39: .

bins the gain is very little.

The configuration with 9 bins that gives the larger ΔLQ is shown graphically in Fig. 39 (bottom): the binning is finer when the gradient of the mass distribution is larger.

The optimal mass binning is then used to find the best binning configuration for the BDT. The optimization procedure is performed for 5,6,7,8 and 9 bins configurations using a grid of 0.05 granularity. The best patterns relative to a given number of bins are shown in Table 20 with the relative ΔLQ value. The ΔLQ as a function of the number of bins is shown in Fig. 41. The optimal 8 bins set, reaches a ΔLQ value greater than 99% of the maximum possible value evaluated by fitting the point with an exponential function.

The best bin configuration is shown in Table 19 both for mass and BDT.

Table 20 shows the number of expected signal events and background events for each BDT bin evaluated assuming the conditions described above. These numbers can be compared with the ones obtained in Section 9 using the full calibration procedure.

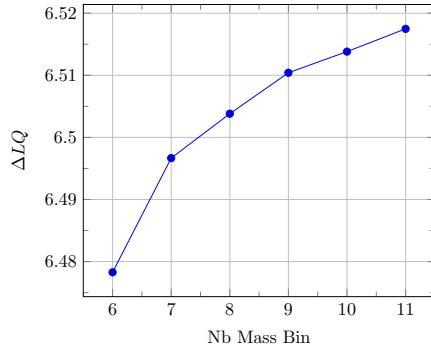


Figure 40: ΔLQ value (left) as a function of the number of mass bins.

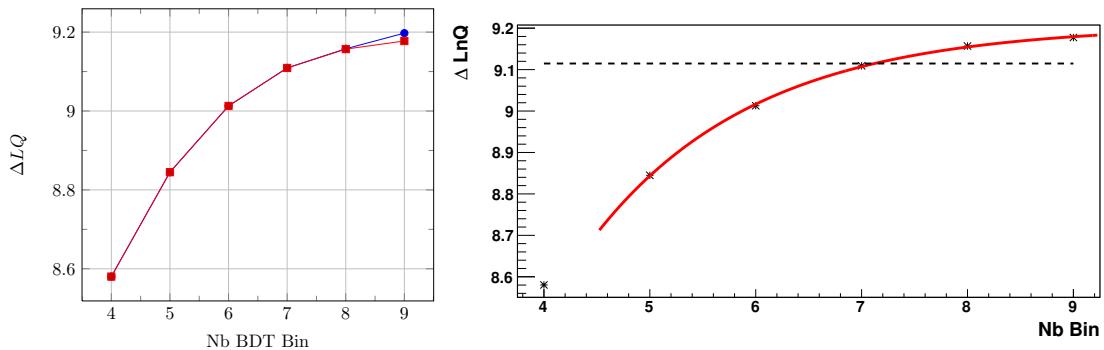


Figure 41: (Right) Evolution of the separation ΔLQ with the number of BDT bins. For the blue (circle markers) curve, no constrain on the bin width is imposed while for the red (squared markers) a minimum bin width of 0.1 is imposed. The latter curve is fitted (left) with an exponential function. The dashed line represent 99% of the maximum of this function

930 In Table 21 the ΔLQ obtained with the new binning (first row) is compared with the
 931 one obtained with the old binning configurations either in mass or BDT or both: the
 932 new binning brings a 16% improvement in ΔLQ with respect to the old one and this
 933 improvement mainly comes from the BDT part.

The contribution of each bin to the analysis sensitivity can be evaluated using the additivity of ΔLQ :

$$\Delta LQ = \sum_{bin} \Delta LQ_i$$

Table 19: Best bin configurations for mass and BDT.

Bins	bin 1	bin 2	bin 3	bin 4	bin 5	bin 6	bin 7	bin 8	bin 9
Mass	5313–5325	5325–5337	5337–5343	5343–5355	5355–5391	5391–5403	5403–5409	5409–5421	5421–5433
BDT	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	

Table 20: Number of expected signal events and background events for each BDT bin.

Bins	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
N_{signal}	2.581	1.548	1.032	1.032	1.032	1.032	1.032	1.032
$N_{\text{background}}$	2061	65.45	15.11	6.704	2.975	1.320	0.586	0.26

Table 21: ΔLQ obtained with the new binning (first row) compared with the ΔLQ obtained using the old mass binning or the old BDT binning or both.

Mass binning	BDT binning	ΔLQ
New (9 bins)	New (8 bins)	9.16
Old (6 bins)	New (8 bins)	9.07
New (9 bins)	Old (4 bins)	8.12
Old (6 bins)	Old (4 bins)	8.04

where ΔLQ_i is the ΔLQ in bin i . Figure 42 shows the contribution that each bin brings to the total ΔLQ and their ranking with respect to this criteria. Table 22 shows the contributions of each BDT bin if the mass bins are added up: more than 85 % of the total sensitivity is contained in the BDT bins above 0.7 while the first bin contributes to less than 0.1%. Figure 43 shows how ΔLQ evolves when bins are added from the more sensitive to the less sensitive one.

Table 22: Absolute (second column) and relative (third column) contribution to the total ΔLQ for each BDT bin.

BDT bin	ΔLQ_i	$\Delta LQ_i/\Delta LQ$ (%)
0 – 0.25	0.00899	0.0981
0.25 - 0.4	0.1	1.09
0.4 – 0.5	0.186	2.03
0.5–0.6	0.396	4.33
0.6 – 0.7	0.797	8.70
0.7 – 0.8	1.48	16.1
0.8 – 0.9	2.47	26.9
0.9 – 1.0	3.73	40.7

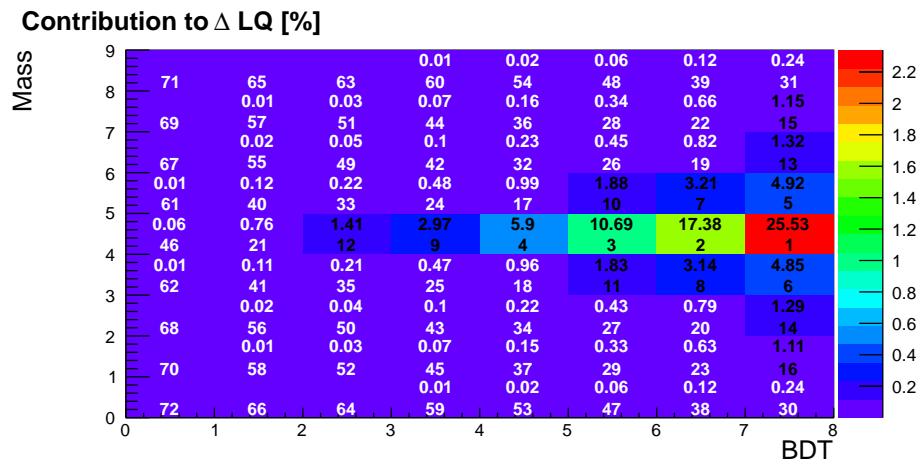


Figure 42: Contribution of each bin to the total ΔLQ and their ranking.

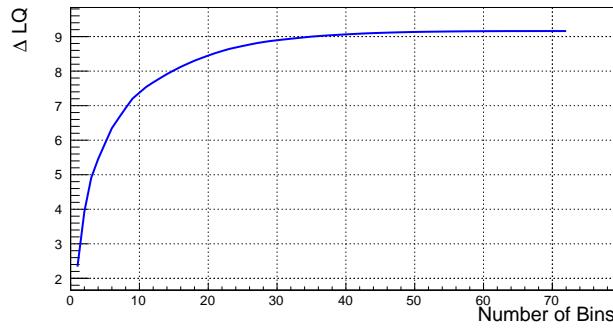


Figure 43: ΔLQ evolution when the bins are added up from the more sensitive to the less sensitive one.

9 Calibration of the BDT and the invariant mass

The BDT discriminant is trained using Monte Carlo samples ($B_{(s)}^0 \rightarrow \mu^+ \mu^-$ for signals and $b\bar{b} \rightarrow \mu\mu X$ for background) and its PDF for signal and background events is extracted from data.

The BDT PDF for the signals is extracted as in the past analyses using TIS events from the inclusive $B_{(s)}^0 \rightarrow h^+ h'^-$ sample (see Section 9.1), while the exclusive $B_{(s)}^0 \rightarrow KK, K\pi, \pi\pi$ samples are used to cross check the results.

The mass PDFs for the signals are parameterized as two Crystal Ball functions:

- the central values are extracted from the exclusive $B^0 \rightarrow K^+ \pi^-$ and $B_s^0 \rightarrow K^+ K^-$ samples (Section 9.2);
- the resolutions are measured by interpolating at the B^0 and B_s^0 mass point the resolutions measured with the $J/\psi, \Psi(2S), \Upsilon(1S, 2S, 3S)$ dimuon resonances (Section 9.5) and cross-checked by fitting the mass lineshape of the exclusive $B_{(s)}^0 \rightarrow h^+ h'^-$ decays;
- the turning point of the radiative tail is obtained by fitting the mass lineshape of $B_s^0 \rightarrow \mu^+ \mu^-$ Monte Carlo events smeared in such a way to recover the mass resolution measured in data.

The BDT and invariant mass PDFs for the combinatorial background are extracted from of a fit of the dimuon mass sidebands in each BDT bin as in the previous analyses (Section 9.6).

The $B_{(s)}^0 \rightarrow h^+ h'^-$ peaking background yield with double misidentified hadrons has been evaluated in Sect.7. Its mass lineshape is obtained from a Monte Carlo sample of double misidentified $B_{(s)}^0 \rightarrow h^+ h'^-$ events and normalized to the number of events expected in the two search windows. The BDT PDF of the double misidentified $B_{(s)}^0 \rightarrow h^+ h'^-$ is evaluated by convoluting the fake rate $\epsilon(K \rightarrow \mu)$ and $\epsilon(\pi \rightarrow \mu)$ probability extracted from the $D^0 \rightarrow K\pi$ decay to the (p, p_T) spectrum of selected $B_{(s)}^0 \rightarrow h^+ h'^-$ events in a given BDT bin. These two PDFs have been already discussed in Section 7.3.

9.1 Extraction of the BDT for signal with the $B_{(s)}^0 \rightarrow h^+ h'^-$ inclusive sample

The BDT PDF for the $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ signals has been extracted from data using a sample of *Triggered Independently of the Signal* (TIS) $B_{(s)}^0 \rightarrow h^+ h'^-$ inclusive events in different BDT bins and extracting the signal yield in each bin. This has been performed by a two-dimensional fit of the invariant mass lineshape (in $\mu\mu$ hypothesis) and of the momentum unbalance of the decay, $\beta = (p^+ - p^-)/(p^+ + p^-)$, where p^\pm are

974 the momenta of the tracks with positive/negative charge. The role of β in the fit is to
975 disentangle among the different $B_{(s)}^0 \rightarrow h^+h'^-$ decay modes.

976 As in the published analysis, only TIS events at the L0 and HLT1 level have been
977 considered, while the bias introduced by the HLT2 trigger level is computed from Monte
978 Carlo and added as a correction to the BDT shape.

979 The invariant mass line shape of $B_{(s)}^0 \rightarrow h^+h'^-$ decays has been presented in detail
980 elsewhere [68] and summarized in [22]. The radiative tail due to the photon emission is
981 also taken into account in the fit.

982 The background is composed by a combinatorial component and a “physical back-
983 ground” due to the partial reconstruction of three-body B meson decays, like $B^0 \rightarrow \rho^\pm\pi^\mp$,
984 in which a $\pi^+\pi^-$ pair mimics the $B_{(s)}^0 \rightarrow h^+h'^-$ signal. Due to the missing π^0 , the $\pi^+\pi^-$
985 invariant mass distribution is kinematically limited to the region below 5.14 GeV/c² (i.e.
986 about $m_{B^0} - m_{\pi^0}$), apart from experimental resolution effects which might lead to higher
987 values. For this reason, this background component mainly affects the left tail of the signal
988 mass distribution. Other relevant decay modes belonging to this category are $B \rightarrow \rho K$
989 and $B \rightarrow K^*\pi$, as well as decays involving other intermediate resonances. Also, a small
990 component due to $\Lambda_b^0 \rightarrow pK^-$ and $\Lambda_b^0 \rightarrow p\pi^-$ is present in the right sideband. Differ-
991 ent background models used to fit both the physical and combinatorial background are
992 described in [22] and not repeated here.

993 The outputs of the fit are:

- 994 - the number of $N(B_{(s)}^0 \rightarrow h^+h'^-)$;
- 995 - the fraction of the radiative tail (f_{rad});
- 996 - the yield (N_{comb}) of the combinatorial background;
- 997 - the exponential slope (τ_k) of the combinatorial background;
- 998 - the yield (N_{phys}) and the coefficient (c) of the Argus function which describes the
999 physical background.

1000 The B^0 and B_s^0 central values and the mass resolutions have been fixed to the values
1001 obtained from the exclusive samples and from the interpolation method, respectively
1002 (Section 9.2 and Section 9.5).

1003 To allow better constraining, the relative yield of the different 2-body decays with
1004 respect to $B^0 \rightarrow K\pi$ is fixed in the fit to the PDG values. The related uncertainty
1005 which is introduced in the fit is not large compared to the systematic uncertainties in the
1006 signal yield extraction which are caused by the parameterization of the mass shape of the
1007 physical background, and of the β shape of the combinatorial background. This second
1008 effect is the most important, since the combinatorial background largely dominates on the
1009 physical background. The effect on the signal counts is further complicated by the fact

1010 that the physical background is strongly anti-correlated with respect to the combinatorial
 1011 background on the left mass sideband.

1012 The current dataset of 1 fb^{-1} allows to extract the β PDFs for the combinatorial back-
 1013 ground from the upper and physical background-free mass sideband, $m_{\mu\mu} > 5700 \text{ MeV}/c^2$.
 1014 This part of the sideband is then not considered in the fit to the mass lineshape. This
 1015 was not possible in the analysis of 370 pb^{-1} due to lack of statistics; in fact, at that time,
 1016 40% of the full $B_{(s)}^0 \rightarrow h^+h'^-$ sample was dedicated to calibrate the β functions using
 1017 events in the upper sideband, above $5.4 \text{ GeV}/c^2$. The β function distribution above 5.7
 1018 GeV/c^2 is in nice agreement with the Monte Carlo predictions in the full mass range (see
 1019 Fig. 47).

1020 Only for the fit to the total number of events, additional information is used for the
 1021 exponential slope of the combinatorial background. More specifically, we fit the value of
 1022 the slope from the events rejected by the BDTS cut, which are fully dominated by the
 1023 combinatorial background (see Fig. 44), and we include it as a constrain for the fit to
 1024 the event with BDTS above cut. In applying the constrain, we allow for a 15% uncer-
 1025 tainty which parameterizes the bias induced by the BDTS selection on the combinatorial
 background slope, as estimated from MC.

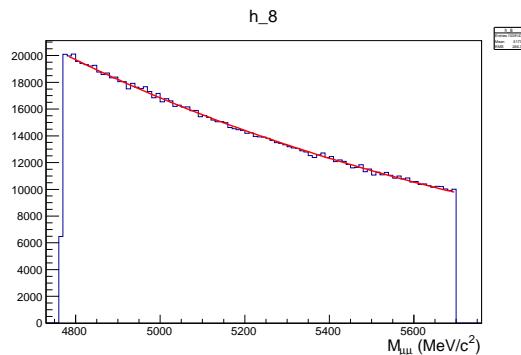


Figure 44: Exponential fit to the events with $\text{BDTS} < 0.05$, data.

1026
 1027 As for the published analysis, the number of $B_{(s)}^0 \rightarrow h^+h'^-$ events in the first BDT
 1028 bin is evaluated by subtracting to the total yield of $B_{(s)}^0 \rightarrow h^+h'^-$ TIS events the yields
 1029 obtained in the seven higher BDT bins. No clean-up cuts have been applied this time to
 1030 the total sample, as the cut $\text{BDTS} > 0.05$ removes already about 70% of the $B_{(s)}^0 \rightarrow h^+h'^-$
 1031 background (see Section 5). For the same reason in the present analysis the yield of the
 1032 first BDT bin can be also independently fitted, and this new feature will be used as a
 1033 cross-check.

1034 The fit projections for mass and β for the total $B_{(s)}^0 \rightarrow h^+h'^-$ sample and for the BDT
 1035 bins above 0.25 are shown in Figures 45 and 46. In Table 23 the corresponding numbers
 1036 of fitted $B_{(s)}^0 \rightarrow h^+h'^-$ events ($N_{B_{(s)}^0 \rightarrow h^+h'^-}$), the number of physical background events
 1037 (N_{phys}) and combinatorial background events (N_{comb}) are listed.

Table 23: Fitted values of $N_{B_{(s)}^0 \rightarrow h^+h'^-}$, N_{phys} and N_{comb} : total yield and BDT bins 2-7.

BDT range	$N_{B_{(s)}^0 \rightarrow h^+h'^-}$	N_{phys}	N_{comb}
0 – 1	16734 ± 757	12115 ± 1610	$(304 \pm 2) \times 10^3$
0.25 – 0.4	1877 ± 88	1575 ± 123	7326 ± 184
0.4 – 0.5	1486 ± 57	1006 ± 63	2128 ± 87
0.5 – 0.6	1447 ± 53	810 ± 57	1492 ± 74
0.6 – 0.7	1729 ± 52	954 ± 46	767 ± 47
0.7 – 0.8	1846 ± 52	1038 ± 45	482 ± 38
0.8 – 0.9	2133 ± 53	923 ± 42	145 ± 25
0.9 – 1.0	2245 ± 53	663 ± 37	34 ± 13

From the total $B_{(s)}^0 \rightarrow h^+h'^-$ yield and the BDT bins above 0.25 we evaluate 3973 events in the first BDT bin, which has to be compared to the 4116 events we get from a direct fit; this result points to a $\sim 4\%$ systematic shift on the first bin yield ($\sim 1\%$ on the total yield).

Further systematic checks have been performed on the fit for the total $B_{(s)}^0 \rightarrow h^+h'^-$ yield:

- To account for the small discrepancy observed at β around zero, which is due to the combinatorial background, we tried two alternative approaches for the parameterization of its β PDF: a polynomial; and the same strategy as for the published paper, consisting of taking the β PDF from a wider sideband, $m_{\mu\mu} > 5400$ MeV/ c^2 , on 40% of the available statistics, while fitting the rest to get the signal yield. The maximum spread observed is 5%, which is included in the systematic, too.
- The stability of the result was tried moving the BDTS cut from BDTS > 0.05 ($\epsilon_{\text{sig}} = 0.92$) to BDT > 0.03 ($\epsilon_{\text{sig}} = 0.92$) and BDT > 0.1 ($\epsilon_{\text{sig}} = 0.85$), which allows to change the combinatorial background by as much as factor of two. The maximum spread observed in the signal yield, after having corrected for MC efficiencies, is in this case 4%.

Taking into account the above systematic errors, the total yield of $B_{(s)}^0 \rightarrow h^+h'^-$ events is $N_{B_{(s)}^0 \rightarrow h^+h'^-} = (16734 \pm 1312)$.

The $B_{(s)}^0 \rightarrow h^+h'^-$ total yield is used not only in the BDT calibration but also in the computation of the normalization factor for the $B^0 \rightarrow K^+\pi^-$ channel (see Sect. 10) and in the evaluation of the total yield of the $B_{(s)}^0 \rightarrow h^+h'^-$ events with double misID (see Sect. 7).

In order to cross check the results, two independent 1D fits to the mass distribution have been performed on the same dataset by two different groups (CERN and Marseille).

1063 In these fits the signal line shape is parameterized by either a double Crystal Ball (CERN
 1064 fit, described in Appendix D of [22]) or a single Crystal Ball with radiative tail parameters
 1065 fixed from simulation and resolution fixed from data (Marseille fit). In both cases the
 1066 combinatorial and physical backgrounds are parameterized by an exponential and Argus
 1067 functions, respectively. As a result, a 5% maximum spread is observed on the total
 1068 $B_{(s)}^0 \rightarrow h^+h^-$ yield between these fits and the 2D fit described above, falling within the
 1069 quoted systematic uncertainty.

1070 The fits to the invariant mass distributions for all BDT bins are shown in Fig. 48 for
 1071 the CERN fit.

1072 In Table 24 the fractional yields for all of the eight BDT bins are shown, with the 2D
 1073 and 1D fit approaches. The bin-by-bin maximum difference is included as a systematic
 error in the evaluation of signal BDT PDF.

Table 24: Fractional yields for the eight BDT bins of the present analysis: 2D fit and 1D fit results, with statistical errors.

BDT bin	$B_{(s)}^0 \rightarrow h^+h^-$ 2D-fit (LNF) [%]	$B_{(s)}^0 \rightarrow h^+h^-$ 1D-fit (CPPM) [%]	B_{2hh} 1D-fit (CERN) [%]
0 – 0.25	23.7 ± 3.6	21.1 ± 3.7	24.4 ± 3.4
0.25 – 0.4	11.2 ± 0.7	12.5 ± 0.9	12.6 ± 0.7
0.4 – 0.5	8.9 ± 0.5	9.5 ± 0.7	9.8 ± 0.7
0.5 – 0.6	8.6 ± 0.5	9.1 ± 0.6	8.9 ± 0.5
0.6 – 0.7	10.3 ± 0.6	9.8 ± 0.6	10.0 ± 0.5
0.7 – 0.8	11.0 ± 0.6	11.1 ± 0.6	10.8 ± 0.6
0.8 – 0.9	12.7 ± 0.7	12.7 ± 0.7	12.2 ± 0.6
0.9 – 1.0	13.4 ± 0.7	14.2 ± 0.7	11.5 ± 0.7

1074 Further studies have been performed by using the CERN 1D fit under different mass
 1075 hypotheses, and applying the HLT2 TIS unbias, too. The results, which confirm the size
 1076 of the above systematic effects, are presented in Appendix A.

As in the previous analysis, in order to keep a reasonable dataset for the calibration,
 the $B_{(s)}^0 \rightarrow h^+h^-$ events are required to be TIS events only at L0 and HLT1 stages and
 then corrected for possible HLT2 trigger biases. Therefore the BDT response calibrated
 with $B_{(s)}^0 \rightarrow h^+h^-$ TIS events has to be divided by the following correction factor:

$$\frac{\epsilon_{\text{TIS}}(L0 \times \text{HLT1})_{hh} \times \epsilon_{\text{trigger}}(\text{HLT2})}{\epsilon_{\text{trigger}}(L0 \times \text{HLT1} \times \text{HLT2})_{\mu\mu}} \quad (12)$$

1078 where the numerator is the efficiency correction to be applied to $B_{(s)}^0 \rightarrow h^+h^-$ TIS events
 1079 to get a trigger unbiased sample, while the denominator is the efficiency correction to be

1080 applied to the unbiased sample to emulate the muon trigger bias on $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ signal
 1081 events.

1082 The evaluation of the above factors has been performed using $B^0 \rightarrow K^+ \pi^-$ and
 1083 $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ Monte Carlo events where the TCK 6d0032 has been emulated. The trigger
 1084 bias as a function of the BDT output is shown in Fig. 49, on top plot for the hadronic
 1085 triggers, *i.e.* the numerator of Eq. 12, and on bottom plot for muon triggers, *i.e.* the
 1086 denominator of Eq. 12. Both the hadronic and muon trigger biases has been evaluated on
 offline selected events. The results are shown in Table 25 as a function of the BDT bins.

Table 25: Trigger corrections as a function of the BDT bins.

BDT bin	Trigger correction [%]
0.0-0.25	5.4±0.1
0.25-0.40	5.2 ±0.1
0.40-0.50	5.2±0.1
0.5-0.6	5.7±0.2
0.6-0.7	5.9±0.2
0.7-0.8	6.1±0.2
0.8-0.9	6.2±0.2
0.9-1.0	6.3±0.2

1087
 1088 As shown in Sections 6 and 7, the double misID background rejection greatly benefits
 1089 from the introduction of a loose DLL selection cut, at a cost of $\sim 3\%$ loss of signal events.
 1090 Though the effect is small compared to other effects previously discussed, nevertheless we
 1091 studied the behaviour of the DLL cut efficiency on signal, ϵ_{DLL} , as a function of BDT, in
 1092 order to be able to reproduce it on the BDT signal PDF. The efficiency has been evaluated
 1093 by convoluting the muon DLL efficiencies from data control samples with the MC signal
 1094 events (Sec. 6). The results are listed in Table 26, and shown in Fig. 50. The observed
 1095 bias (*i.e.* the efficiency per bin normalized to its average value) has been introduced in
 1096 the definition of the signal BDT PDF.

1097 Finally, in Fig. 51, the calibrated BDT shape is shown before (left) and after (right)
 1098 the corrections are applied.

Table 26: DLL cut efficiency as a function of the BDT bins.

BDT bin	ϵ_{DLL} [%]
0.0-0.25	96.86 ± 0.01
0.25-0.4	96.86 ± 0.01
0.4-0.5	96.93 ± 0.01
0.5-0.6	97.10 ± 0.01
0.6-0.7	97.40 ± 0.01
0.7-0.8	97.76 ± 0.01
0.8-0.9	98.12 ± 0.01
0.9-1.0	98.54 ± 0.01

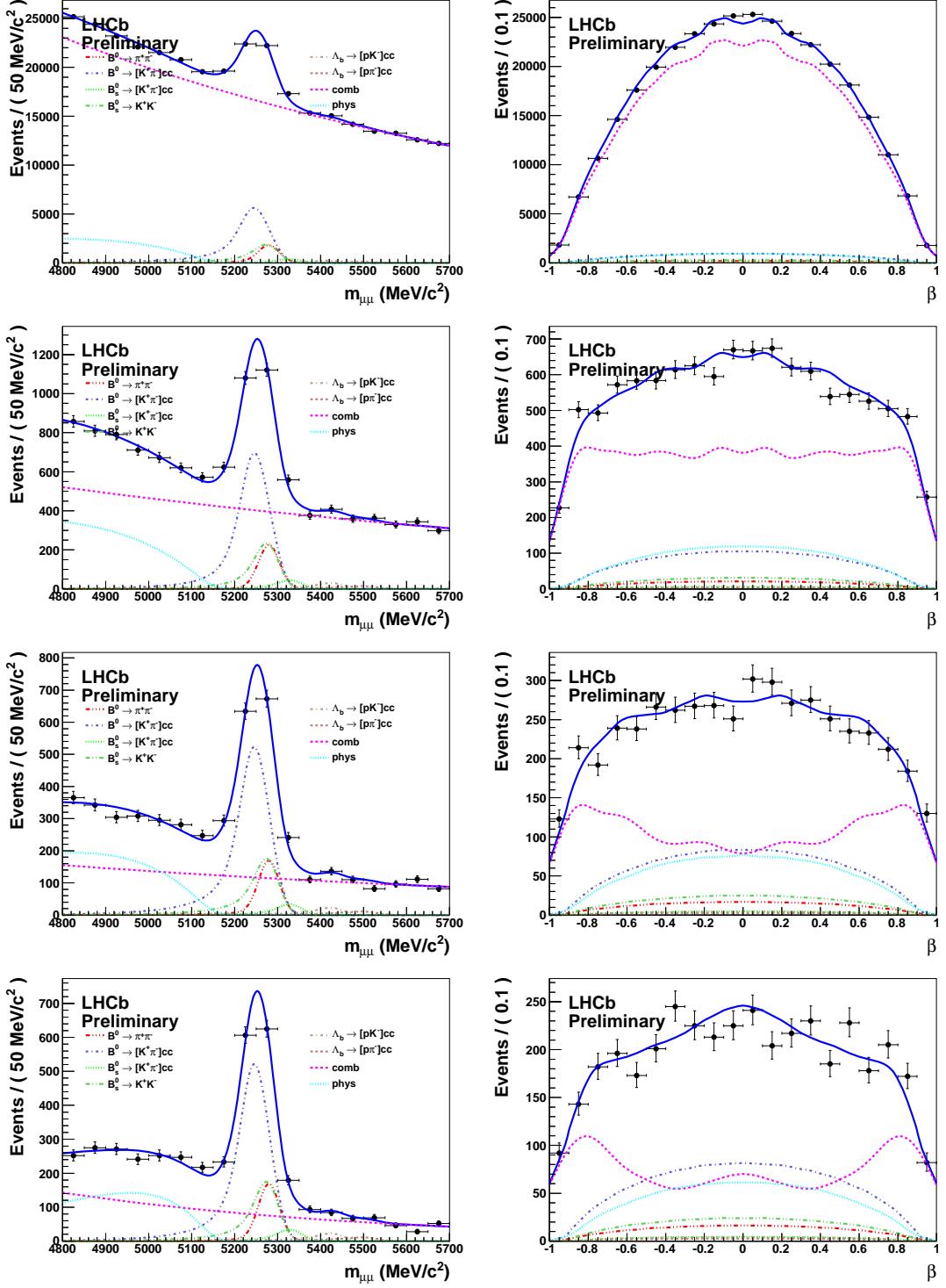


Figure 45: Fit projections on mass (left) and β (right): whole BDT range, and BDT bins 2-3-4.

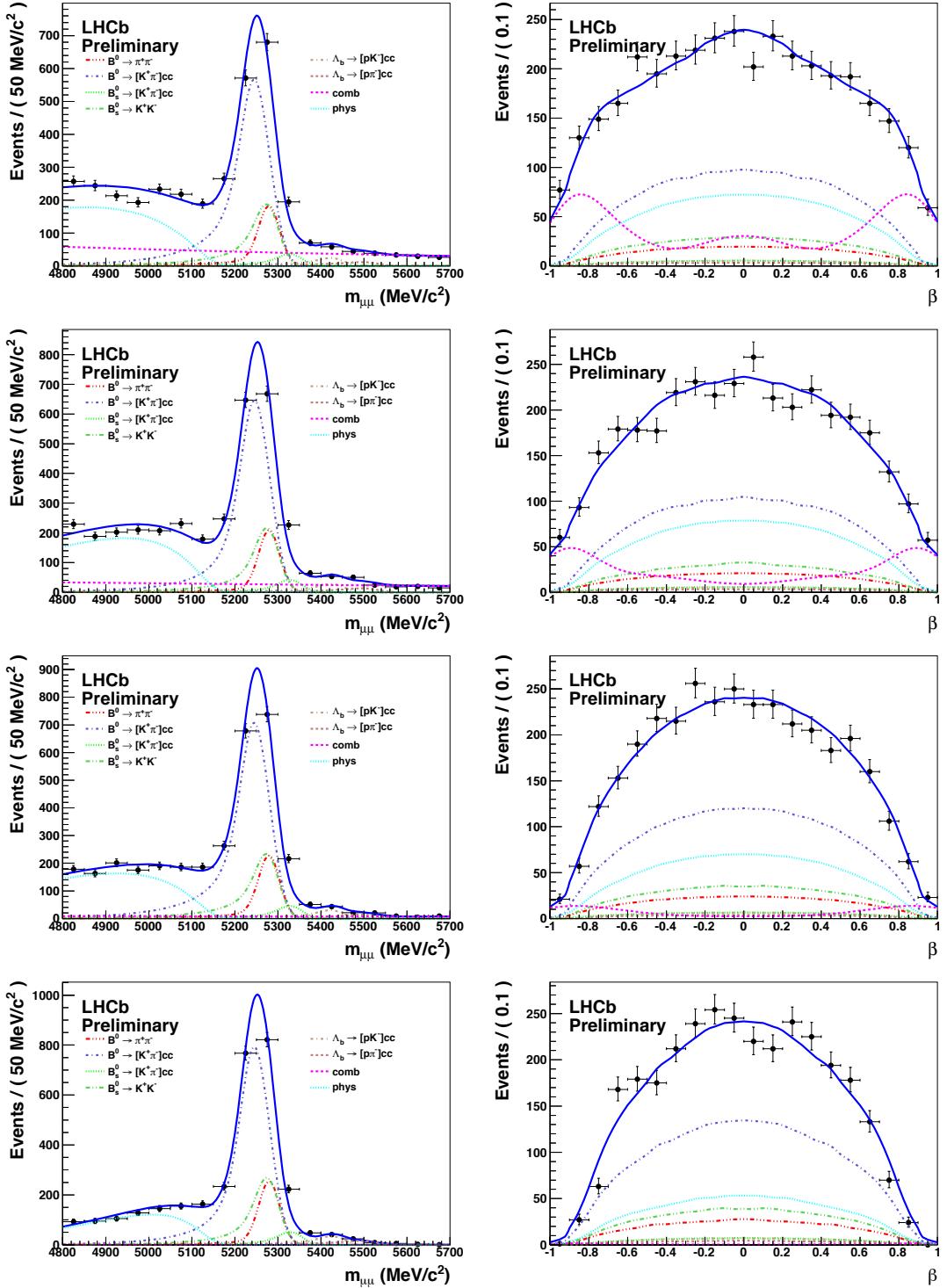


Figure 46: Fit projections on mass (left) and β (right): BDT bins 5-6-7-8.

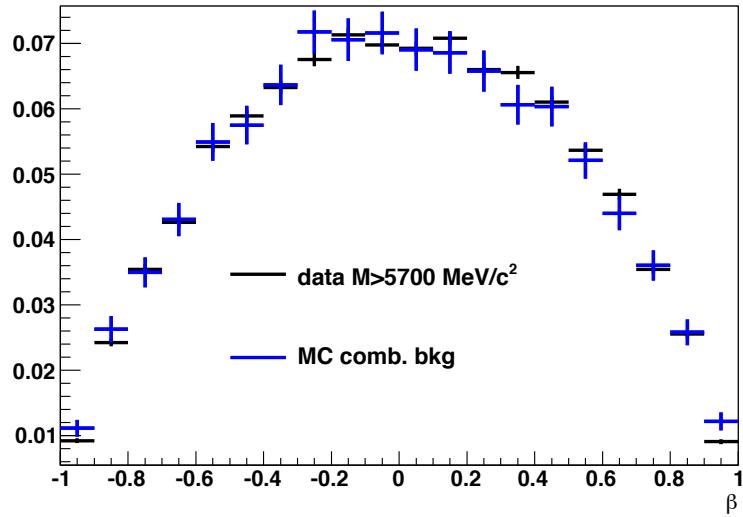


Figure 47: Distribution of the β function: data for $m_{\mu\mu} > 5.7$ GeV/ c^2 (blue markers) are compared with Monte Carlo simulation $b\bar{b} \rightarrow \mu\mu X$ (black markers).

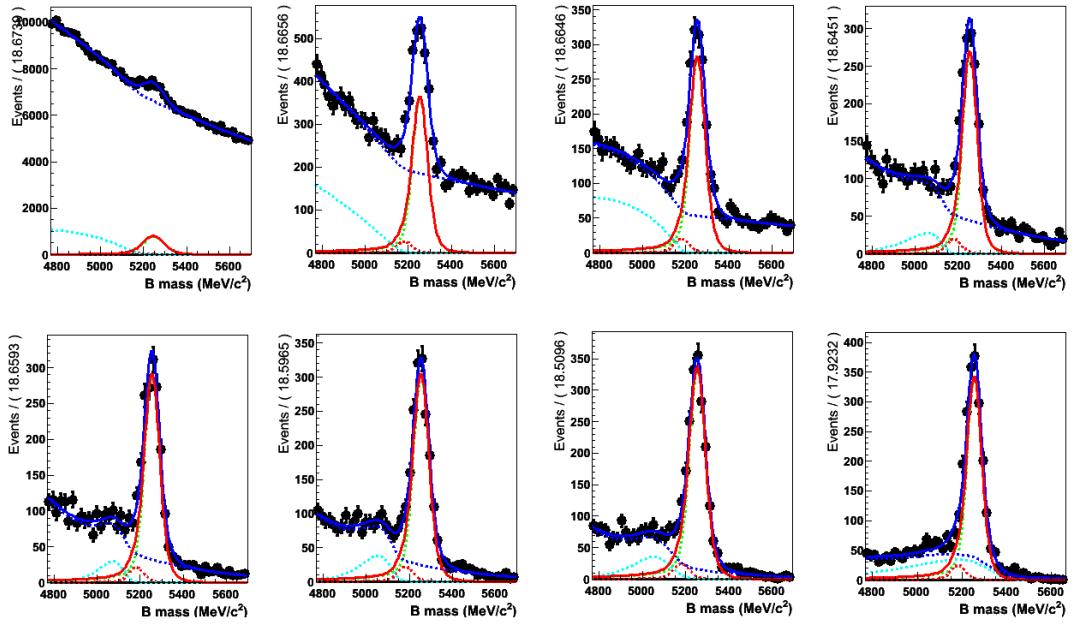


Figure 48: Mass fit results for all of the BDT bins (CERN).

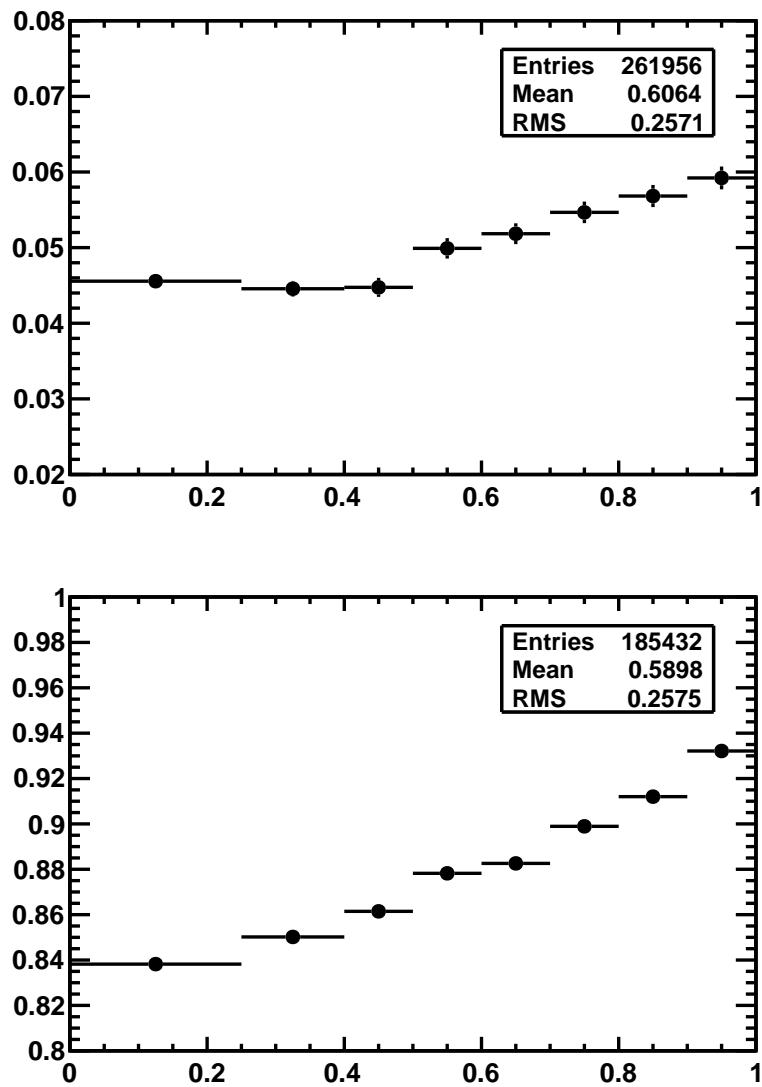


Figure 49: Trigger bias as a function of the BDT output. Top: bias due to the hadronic trigger. Bottom: bias due to the muon trigger.

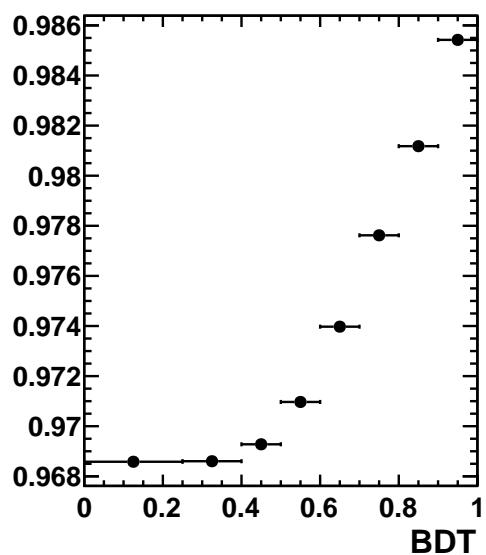


Figure 50: DLL cuts efficiency on signal, as a function of BDT bins.

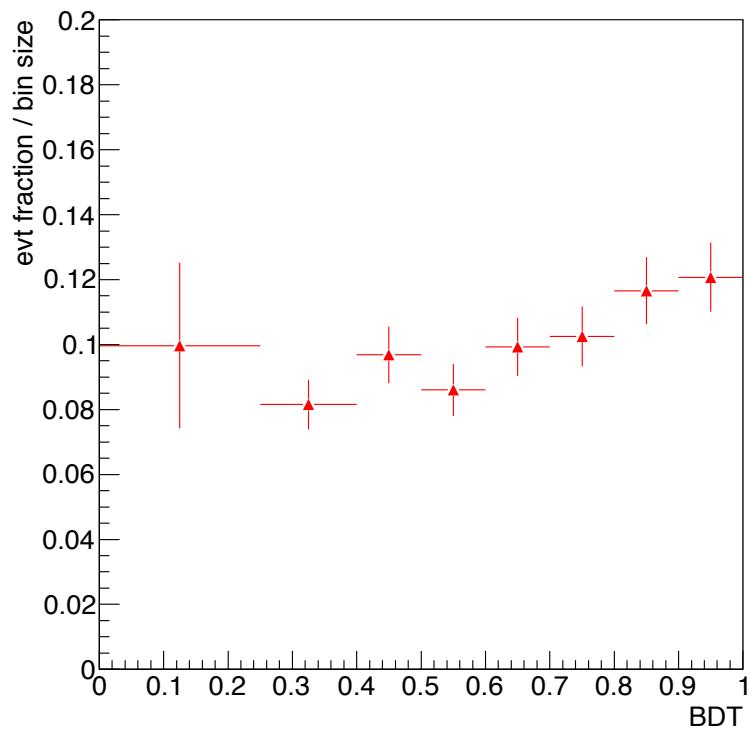
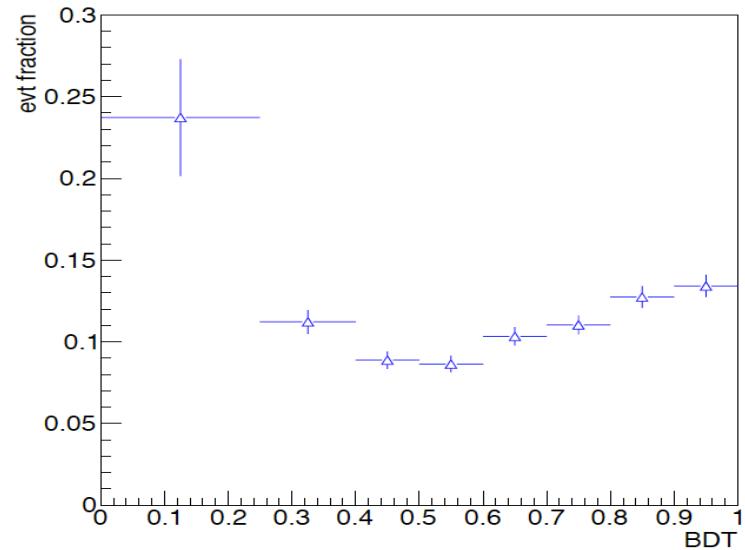


Figure 51: Calibrated BDT shape before (top) and after (bottom) the trigger corrections and the normalization to the bin size.

1099 **9.2 Invariant mass central values for signal with the $B_{(s)}^0 \rightarrow h^+h'^-$**
 1100 **exclusive samples**

1101 The exclusive $B_{s,d} \rightarrow \pi K, \pi\pi, KK, K\pi$ channels are used in 2011 data to estimate the
 1102 BDT pdf and the B_s^0 and B^0 masses. This was not possible in the 2010 analysis as
 1103 the RICH efficiency map was not still calibrated. However, due to a still not yet fully
 1104 understood control of the RICH calibration (see the discussion reported in detail in Ap-
 1105 pendix B) the exclusive channels are used to estimate the central mass values and only
 1106 to cross-check the mass resolution and the BDT calibration.

1107 The information of the Cherenkov angle and the number of photo-electrons produced
 1108 by a given particle allows to perform an PID hypothesis test, via a delta-log-likelihood
 1109 (DLL). In the kaon and pion hypothesis, a $DLL_{K-\pi}$ is built and a particle is identified
 1110 with a certain probability $P(k)$ as a kaon if $DLL_{K-\pi} > k$ or as a pion if $DLL_{K-\pi} < -k$.

1111 This RICH efficiency map has been evaluated for the 2011 data using the $D^* \rightarrow$
 1112 $D^0(K\pi)\pi$ calibration sample. For each $DLL_{K-\pi}$ cut, an efficiency value is provided for
 1113 each bin of the 3-dimensional phase space defined by the three axes ($p, \eta, N_{\text{tracks}}$). The
 1114 efficiency map used in this analysis is built with reco12 and is divided in the following
 1115 bins:

1116 **p binning [GeV/c]** : 2, 5, 9.3, 15.6, 19, 24.4, 29.8, 35.2, 40.6, 46, 51.4, 56.8, 62.2,
 1117 67.6, 73, 78.4, 83.8, 89.2, 94.6, 100, 130, 200.

1118 **η binning** : 1.5, 2.5, 3.5, 4.5, 5.5.

1119 **N_{tracks} binning**: 0, 55, 200, 300, 500.

1120 Exclusive $B_{(s)}^0 \rightarrow h^+h'^-$ decays are extracted from $B_{(s)}^0 \rightarrow h^+h'^-$ inclusive TIS events,
 1121 where a loose cut in the BDT (BDT > 0.2) is applied to cleanup part of combinatorial
 1122 background¹⁶. A particle is consider a kaon if $DLL_{K-\pi} > 10$ (and $DLL_{\pi-p} > 0$) and a
 1123 pion if $DLL_{K-\pi} < -10$ (and $DLL_{\pi-p} > 0$). From the original $B_{(s)}^0 \rightarrow h^+h'^-$ sample, four
 1124 disjunctive sets of events for different mass hypotheses $\pi^+\pi^-$, $K^+\pi^-$, π^+K^- and K^+K^-
 1125 are built-up. The distribution of the $B_{(s)}^0 \rightarrow h^+h'^-$ exclusive decays mass is obtained
 1126 weighting each event by the DLL-cut efficiency. The weight used is given by:

$$w(k) = \frac{1}{\epsilon_h(p(h^+)), \eta(h^+), N_{\text{tracks}} | | DLL_{K-\pi}| > k} \frac{1}{\epsilon_h(p(h^-)), \eta(h^-), N_{\text{tracks}} | | DLL_{K-\pi}| > k} \quad (13)$$

1127 The mass distributions are shown in Fig. 52 for the K, π hypothesis (left: π^+K^- ,
 1128 right: $K^+\pi^-$) , and in Fig. 53 left for the K^+K^- hypothesis and Fig. 53 right for the

¹⁶A BDT cut does not affect the mass resolution as has been proven that the BDT output and mass resolution are fully uncorrelated, see Section 7.3 of ANA-NOTE-078-2011

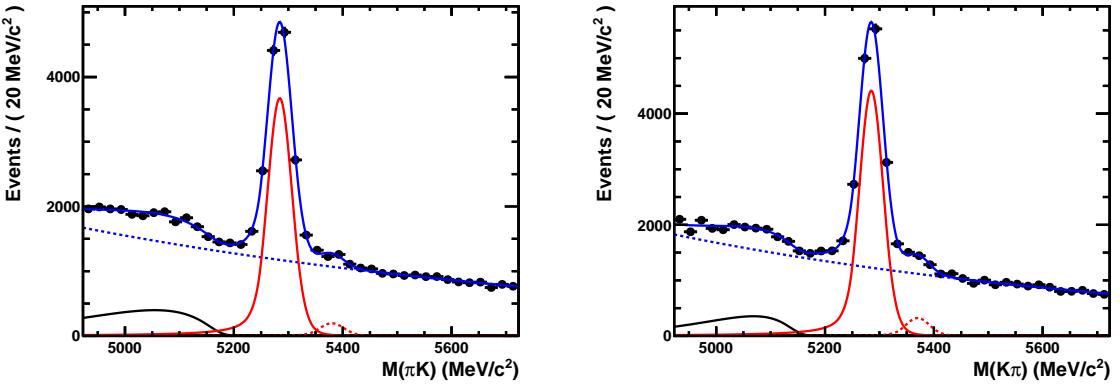


Figure 52: Mass distributions for $K\pi$ combinations (left: π^+K^- , right: $K^+\pi^-$).

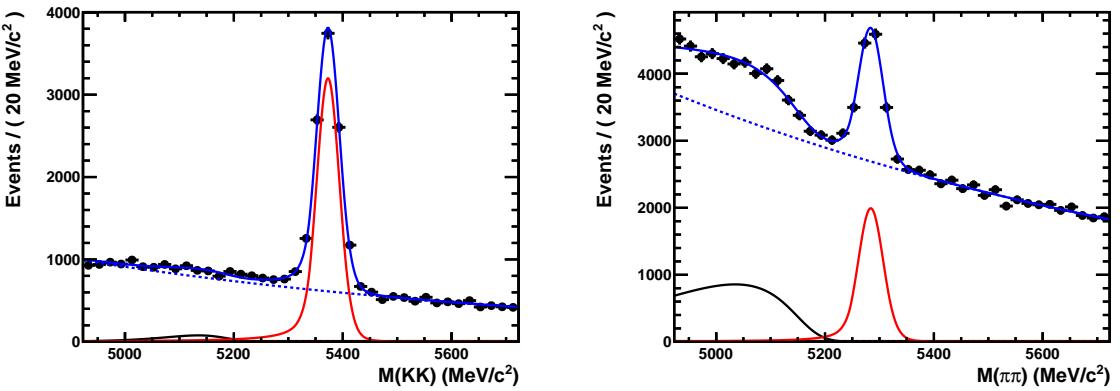


Figure 53: Mass distributions for K^+K^- (right) and $\pi^+\pi^-$ (right) combinations.

1129 $\pi^+\pi^-$ one. The mass distribution is fit with a double-sided Crystal Ball function for the
 1130 signal, an Argus function for the physical background and an exponential function for
 1131 the combinatorial background. The values of the mass average for the four combinations
 1132 are listed in Table 27. Notice that there is a bias of 0.1 % above the PDG values. The
 1133 systematic uncertainties associated to the mass average are listed in Table 28. They have
 1134 been estimated varying the PID and BDTS cuts and performing the fit fixing or not the
 1135 tail parameter to the MC value.

1136 9.3 Invariant mass resolution for signal with the $B_{(s)}^0 \rightarrow h^+h^-$ 1137 exclusive samples

1138 For a given $DLL(K - \pi) > k$ cut, a particle is identified as a kaon if $DLL(K - \pi) > k$ or
 1139 as a pion if $-DLL(K - \pi) > k$. From the original sample, four disjunctive sets of events
 1140 for different mass hypotheses $\pi^+\pi^-$, $K^+\pi^-$, π^+K^- and K^+K^- are then built-up. To each
 1141 event the corresponding mass hypothesis is assigned and the event is weighted by:

Table 27: Mass of the exclusive $B_{(s)}^0 \rightarrow h^+ h'^-$ samples (in units of MeV/c^2).

$B^0 \rightarrow \pi^+ \pi^-$	$5283.98 \pm 0.63_{(\text{stat})} \pm 0.21_{(\text{syst})}$
$B^0 \rightarrow K^+ \pi^-$	$5285.27 \pm 0.29_{(\text{stat})} \pm 0.32_{(\text{syst})}$
$B^0 \rightarrow \pi^+ K^-$	$5284.31 \pm 0.32_{(\text{stat})} \pm 0.25_{(\text{syst})}$
B^0 combined	$5284.63 \pm 0.20_{(\text{stat})} \pm 0.27_{(\text{syst})}$
$B_s^0 \rightarrow K^+ K^-$	$5372.96 \pm 0.32_{(\text{stat})} \pm 0.28_{(\text{syst})}$

Table 28: Systematic errors in the determination of the invariant mass average value from exclusive $B_{(s)}^0 \rightarrow h^+ h'^-$ samples (in unit of MeV/c^2).

	PID	BDTS	fit model
$B^0 \rightarrow K^+ \pi^-$	0.10	0.25	0.17
$B^0 \rightarrow \pi^+ K^-$	0.04	0.19	0.15
$B^0 \rightarrow \pi^+ \pi^-$	0.05	0.18	0.10
$B_s^0 \rightarrow K^+ K^-$	0.05	0.24	0.13

$$w = \frac{1}{\epsilon_h(p(h^+)), \eta(h^+), N_{\text{tracks}} | | DLL(K - \pi)| > k \cdot \epsilon_h(p(h^-)), \eta(h^-), N_{\text{tracks}} | | DLL(K - \pi)| > k} \quad (14)$$

Finally the mass distributions of the four disjunctive $K\pi$, KK , $\pi\pi$ and πK datasets are merged (see Fig. 54, left).

The widths of the two peaks are determined by two factors:

1. the intrinsic mass resolution of the B^0 and B_s^0 two body decays which is constant with respect to the DLL cut once corrected for the RICH efficiency;
2. the contribution of mis-identified $K \rightarrow \pi$ and $\pi \rightarrow K$ which is not constant and decreases for tighter DLL cuts.

The mass distributions is then fit with two Crystal Ball functions for the signal, a Gaussian distribution with $\sigma > 30$ MeV/c^2 for the $\Lambda_b \rightarrow ph$ contribution visible in the upper sideband, the Argus function for the physical background in the lower sideband and an exponential function for the combinatorial background. The σ of the two Crystal Ball functions (not anylonger constrained to be the same as it was for the 370 pb^{-1} analysis as now we have much more data) are then plot as a function of the DLL cut k (Fig. 54) (left) and fit with the function:

$$\sigma_{\text{IM}}(k) = \sigma_{\text{unbiased}} + \frac{\sigma_{\text{mis-ID}}}{1 + re^{k/k^0}} \quad (15)$$

1156 where σ_{unbiased} is the intrinsic mass resolution, efficiency-corrected, for a $B_{d,s}$ two body
 1157 decay and $\sigma_{\text{mis-ID}}/(1 + r \cdot e^{k/k^0})$ is the contribution to the width given by misidentified
 1158 particles. The results obtained separately for B^0 and B_s^0 are shown in Fig. 54 (right). The
 1159 drop-off of the black points is due to a lack of $B^0 \rightarrow K^+\pi^-$ events in the high p range is
 1160 now at $|DLL(K - \pi)| > 8$ (with 370 pb^{-1} this was at $|DLL(K - \pi)| > 6$)

1161 The re-weighting of Eq. 14 is only possible for tracks with $p \in [2, 200] \text{ GeV}/c$ (in 370
 1162 pb^{-1} it was $[2, 150] \text{ GeV}/c$) and $\eta \in [1.5, 5.5]$ as the calibration map provided by the RICH
 1163 group covers only this phase space. Hence to the resolution obtained from $B^0 \rightarrow K^+\pi^-$
 1164 events correction factor has to be applied to account for this reduced phase space not
 1165 present in the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ selection.

1166 The correction factor γ is evaluated by fitting the invariant mass distribution of a pure
 1167 $B^0 \rightarrow K^+\pi^-$ Monte Carlo sample with and without the cuts in p and η . The result is
 1168 $\gamma = 1.023 \pm 0.009$. Putting all together the B^0 and B_s^0 mass resolutions are:

$$\begin{aligned}\sigma_{\text{IM}}^{\text{corr}}(B^0 \rightarrow \mu^+\mu^-) &= \gamma \cdot \sigma_{\text{unbiased}} = 24.1 \pm 0.1 \pm 0.8 \text{ MeV}/c^2 \\ \sigma_{\text{IM}}^{\text{corr}}(B_s^0 \rightarrow \mu^+\mu^-) &= \gamma \cdot \sigma_{\text{unbiased}} = 24.7 \pm 0.2 \pm 1.0 \text{ MeV}/c^2\end{aligned}\quad (16)$$

and are therefore also in good agreement with the interpolation method.

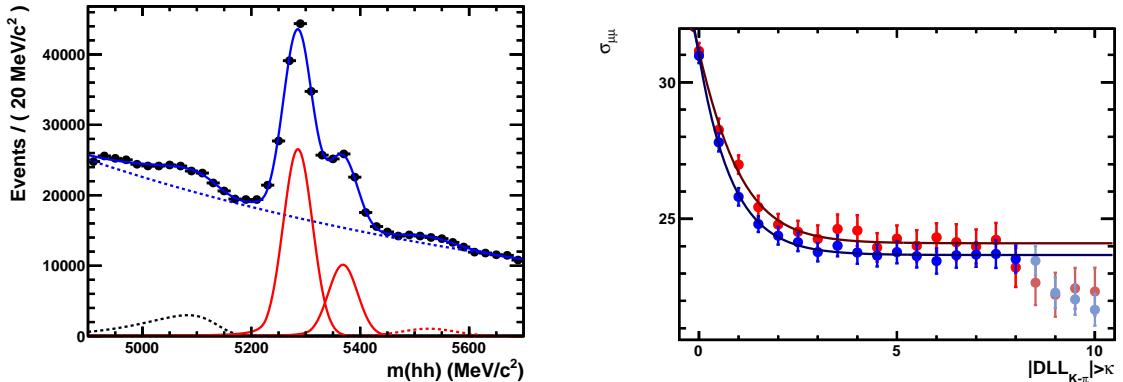


Figure 54: Left: Fit to the invariant mass distribution in the $K^+\pi^-$, $\pi^+\pi^-$, π^+K^- and K^+K^- hypotheses for a given $|DLL(K - \pi)| > k$ cut. Right: mass resolutions for B^0 and B_s^0 as a function of the DLL cut.

1169

1170 9.4 Calibration of the BDT for signal with the $B_{(s)}^0 \rightarrow h^+h'^-$ 1171 exclusive samples

1172 The exclusive $B_{(s)}^0 \rightarrow h^+h'^-$ decays are also used to estimate the BDT pdf. $B_{(s)}^0 \rightarrow h^+h'^-$
 1173 candidates are required to be TIS and to pass the selection described in Appendix A of
 1174 Ref. [22]. We identify kaons or pions for given κ cut, that is, for a Kaon we require a

1175 $DLL_{K-\pi}(K) > \kappa$ and for a pion $DLL_{K-\pi}(\pi) < -\kappa$. Selected $B_{(s)}^0 \rightarrow h^+h'^-$ exclusive
 1176 candidates are then weighted by the RICH PID efficiency (Eq. 13). The $B_{(s)}^0 \rightarrow h^+h'^-$
 1177 exclusive or inclusive mass distributions are then obtained for a given κ cut.

1178 The invariant mass shapes for the four combination $\pi K, \pi\pi, KK$ and $K\pi$ combined,
 1179 for a $\kappa = 0$ cut are shown in Fig. 55 for the full BDT range (top-left) and for the rest
 1180 of the BDT bins except the 1st one (top-left plot corresponds to BDT range [0.25,0.4],
 1181 second bin, bottom-right plot corresponds to BDT range [0.9,1.]). The mass distribution
 1182 is fit to a two Crystal Ball functions for the signal, a Gaussian distribution with $\sigma > 30$
 1183 MeV/ c^2 for the $\Lambda_b \rightarrow ph$ contribution visible in the upper sideband, an Argus function
 1184 for the physical background in the lower sideband and an exponential function for the
 1185 combinatorial background. The σ of the two Crystal Ball functions are constrained to be
 1186 the same.

1187 In order to check that the RICH PID efficiency corrections have been properly applied,
 1188 the fraction of corrected events yield obtained from the fit is plot as a function of κ . The
 1189 results are shown in Fig. 56. No dependence of the corrected yield with the κ cut has
 1190 been observed.

1191 The results of the BDT pdf estimation (in %) are shown in Table 29. The first column
 1192 correspond to the result obtained from the combination of $B_{(s)}^0 \rightarrow h^+h'^-$ exclusive decays,
 1193 second column to the one obtained with $B^0 \rightarrow K^+\pi^-$ decays. The third column is the
 1194 result obtained with the 2-D fit to $B_{(s)}^0 \rightarrow h^+h'^-$ inclusive events (from Sec. ??). The
 1195 BDT pdf obtained with $B_{(s)}^0 \rightarrow h^+h'^-$ exclusive decays has been used only as a cross-check
 1196 to the one obtained with the 2-D fit to $B_{(s)}^0 \rightarrow h^+h'^-$ inclusive decays.

Table 29: BDT pdf (in %) estimation using the combination of $B_{(s)}^0 \rightarrow h^+h'^-$ exclusive channels, $B^0 \rightarrow K^+\pi^-$ decays and the pdf estimation using the 2D fit to $B_{(s)}^0 \rightarrow h^+h'^-$ inclusive decays (from Sec 9)

BDT range	$B_{(s)}^0 \rightarrow h^+h'^-$	$B^0 \rightarrow K^+\pi^-$	$B_{(s)}^0 \rightarrow h^+h'^-$ 2D fit
0 – 0.25	26.3 ± 3.2	26.9 ± 3.2	23.7 ± 3.6
0.25 – 0.4	13.1 ± 0.8	13.6 ± 1.9	11.2 ± 0.7
0.4 – 0.5	9.9 ± 0.5	9.7 ± 1.1	8.9 ± 0.5
0.5 – 0.6	9.1 ± 0.6	8.9 ± 1.0	8.6 ± 0.5
0.6 – 0.7	9.8 ± 0.5	9.4 ± 1.0	10.3 ± 0.6
0.7 – 0.8	9.8 ± 0.5	9.7 ± 1.1	11.0 ± 0.6
0.8 – 0.9	11.1 ± 0.6	10.7 ± 1.4	12.7 ± 0.7
0.9 – 1.0	11.0 ± 0.8	11.1 ± 1.4	13.4 ± 0.7

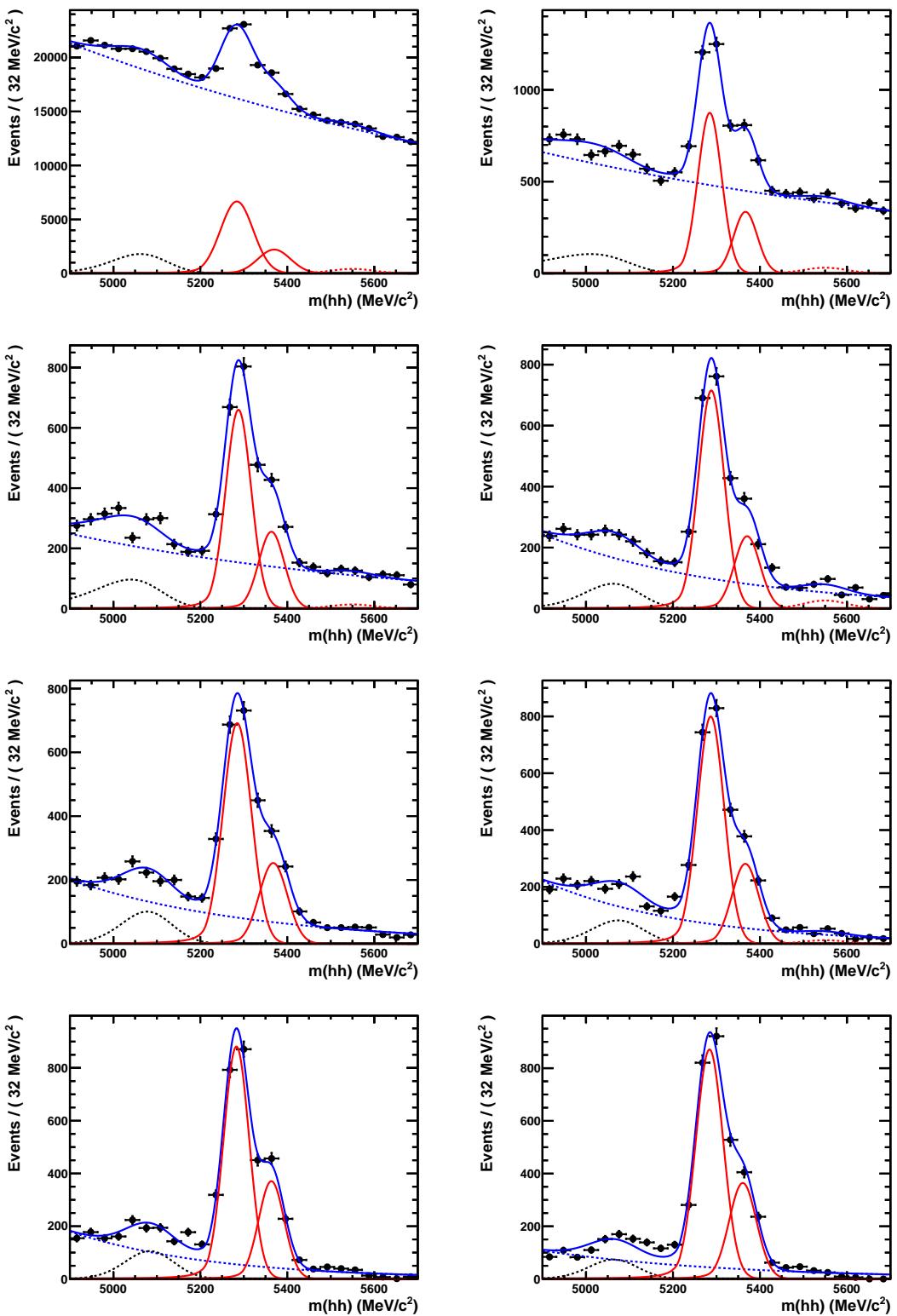


Figure 55: Invariant combined $B_{(s)}^0 \rightarrow h^+h^-$ mass shape for the four combinations ($\pi K, \pi\pi, KK$ and $K\pi$) with $\kappa = 0$, for the full range of BDT (top-left) and for the BDT bins (except the 1st one): BDT $[0.25, 4]$ (top-right), then in order BDT $[0.4, 0.5], [0.5, 0.6], [0.6, 0.7], [0.7, 0.8]$, and BDT⁸³ range $[0.9, 1.]$ (bottom-right).

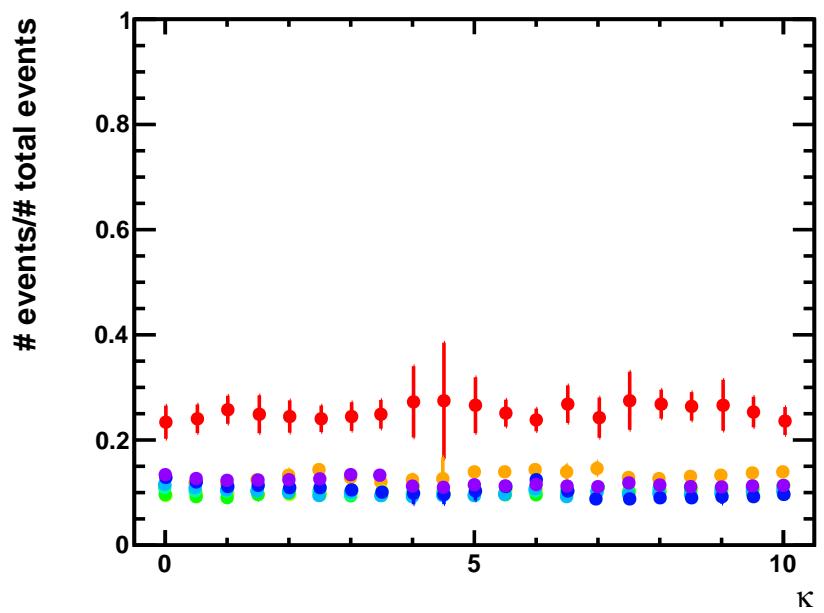


Figure 56: BDT pdf estimation using $B_{(s)}^0 \rightarrow h^+h'^-$ exclusive events as a function of the κ cut for all BDT bins. Different colours correspond to different bins. The red markers correspond to the first bin.

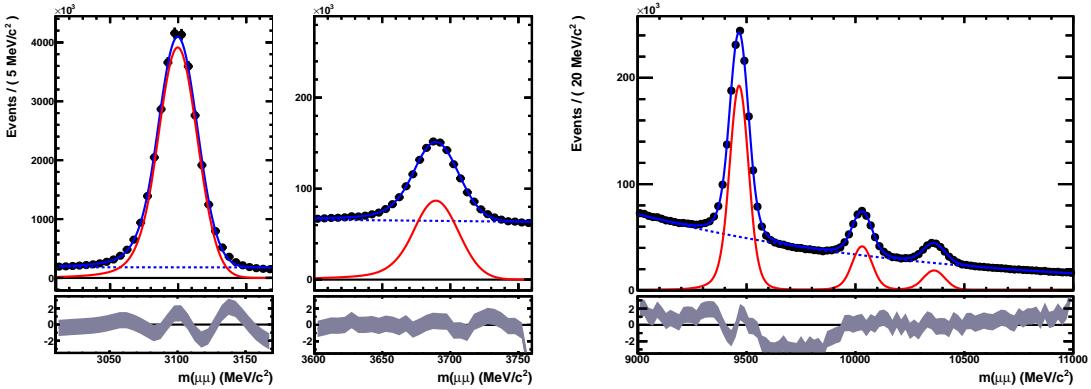


Figure 57: Dimuon invariant mass spectrum between (2.9 - 3.9) GeV/c^2 (left) and (9-11) GeV/c^2 (right) range.

1197 9.5 Calibration of the invariant mass for signal using the interpolation method 1198

1199 The $m_{B_s^0}$ and m_{B^0} mass resolutions are extracted from data by performing an interpolation
1200 from the measured resolutions of Charmonium and Bottomonium resonances decaying
1201 significantly into two muons (i.e. J/ψ , $\psi(2S)$ and $\Upsilon(1S)$, $\Upsilon(2S)$, $\Upsilon(3S)$).

1202 The Υ resonances have been extracted from the *Dimuon High Mass* stripping line, the
1203 J/ψ from the *Jpsi2MuMuLine* (that selects the mass region $|m_{\mu\mu} - m_{J/\psi}| < 80 \text{ MeV}/c^2$)
1204 and the $\Psi(2S)$ from the *Psi2MuMuLine* (that selects the mass region $|m_{\mu\mu} - m_{\Psi(2S)}| <$
1205 $100 \text{ MeV}/c^2$). The *Dimuon Low Mass* stripping line that was used in the previous analysis
1206 could not be used because it was prescaled to zero in Stripping 17. The same additional
1207 cuts used in the previous analysis, $p_T(\mu) > 1 \text{ GeV}$, $\chi^2_{\text{track}}/N_{\text{DOF}} < 5$ and $\chi^2_{\text{PV}}/N_{\text{DOF}} <$
1208 10 have been added to the stripping selections to clean up the peaks.

1209 As in the published analysis, the events in the data samples are weighted such that
1210 the momentum spectra of the Charmonium and Bottomonium resonances are adjusted to
1211 the momentum spectrum of selected $B_s^0 \rightarrow \mu^+ \mu^-$ simulated events from Stripping 17.

1212 The mass lineshapes for charmonium and bottomonium resonances are shown in
1213 Fig. 57. The resonances are fitted with a double-sided Crystal Ball (being the tail parameters
1214 fixed by Monte Carlo) with common resolution. The background is fitted with an
1215 exponential function.

Differently from the published analysis where a linear interpolation of the invariant mass resolution was used, the interpolation has been done using a power-law function:

$$\sigma_{\mu\mu}(m_{\mu\mu}) = a_0 + a_1 \cdot m_{\mu\mu}^\gamma \quad \gamma = 1.37 \pm 0.08 \quad (17)$$

1216 In fact, the long range characteristics studied in simulated Drell-Yan events was found
1217 to be not linear. The behaviour of the mass resolution in the extended mass range from
1218 the J/ψ to the Z^0 is shown in Fig. 59 (left): the power-law function (blue curve) nicely
1219 predicts the measured Z^0 mass resolution within $0.75 \cdot \sigma$.

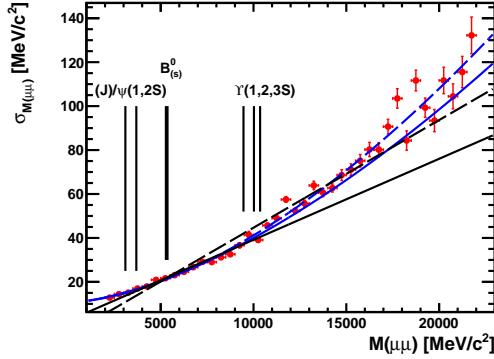


Figure 58: MC10 Drell-Yan Monte Carlo: Red markers: standard deviation of $m_{\text{true}} - m_{\text{rec}}$. Black line is linear fit, blue line the power-law function used in the interpolation method, solid blue line is the fit for [3–11] GeV/c^2 , dashed blue line for the whole range.

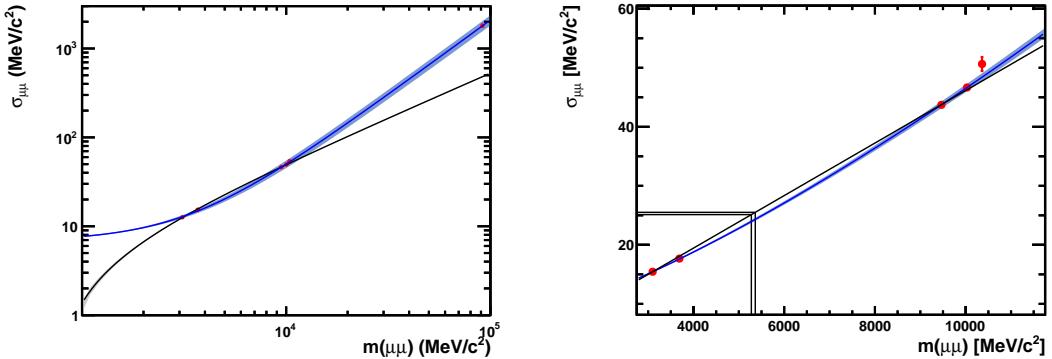


Figure 59: Left: behaviour of mass resolution in the mass range from J/ψ to the Z^0 with power-law (blue) and linear function (black) superimposed; right: zoom between J/ψ and $\Upsilon(3S)$.

1220 The power-law behaviour is also confirmed by a study in the Monte Carlo Drell-Yan
 1221 processes (see Fig. 58).

1222 The zoomed range between J/ψ and $\Upsilon(3S)$ is shown in Fig. 59 (right) where both
 1223 the power-law and the linear fits are performed: the fit with a power-law function gives
 1224 a $\chi^2/ndof = 3.7/2$ while the fit with a linear function $\chi^2/ndof = 16.3/3$. The power-
 1225 law function clearly reproduces better the data also in the restricted range between J/ψ
 1226 and $\Upsilon(3S)$. In the last row the results obtained in the 0.37 fb^{-1} dataset are shown for
 1227 comparison. In Table 31 the breakdown of the different components of the systematic
 1228 errors is detailed for the two fits.

Finally the $m_{B_s^0}$ and m_{B^0} mass resolutions are as follows:

$$\begin{aligned}\sigma(B_s^0) &= (24.8 \pm 0.3_{\text{stat}} \pm 0.7_{\text{syst}}) \text{ MeV}/c^2 \\ \sigma(B^0) &= (24.3 \pm 0.3_{\text{stat}} \pm 0.6_{\text{syst}}) \text{ MeV}/c^2\end{aligned}\quad (18)$$

Table 30: Measured B_s^0 and B^0 mass resolutions.

	$\sigma(m_{B_s^0})$ [MeV/ c^2]	$\sigma(m_{B^0})$ [MeV/ c^2]	$\chi^2/ndof$
power-law (1 fb $^{-1}$)	$24.8 \pm 0.3_{\text{stat}} \pm 0.7_{\text{syst}}$	$24.3 \pm 0.3_{\text{stat}} \pm 0.6_{\text{syst}}$	3.7/2
linear (1 fb $^{-1}$)	$25.5 \pm 0.03_{\text{stat}} \pm 0.9_{\text{syst}}$	$25.1 \pm 0.03_{\text{stat}} \pm 0.8_{\text{syst}}$	16.3/3
linear (0.37 fb $^{-1}$)	$24.6 \pm 0.2_{\text{stat}} \pm 1.0_{\text{syst}}$	$24.3 \pm 0.2_{\text{stat}} \pm 1.0_{\text{syst}}$	—

Table 31: Mass resolution: breakdown of various contributions to the systematic error.

Systematics	B^0	B_s^0
Selection cuts	± 0.22 MeV/ c^2	± 0.18 MeV/ c^2
Momentum weighting	± 0.25 MeV/ c^2	± 0.25 MeV/ c^2
Error asymmetry	± 0.21 MeV/ c^2	± 0.26 MeV/ c^2
Mass window	± 0.32 MeV/ c^2	± 0.32 MeV/ c^2
Fit function of invariant mass	± 0.39 MeV/ c^2	± 0.46 MeV/ c^2
Total systematic error	± 0.64 MeV/ c^2	± 0.69 MeV/ c^2

¹²²⁹ where the systematic uncertainties have been added in quadrature.

1230 **9.6 BDT and invariant mass PDFs for combinatorial back-**
1231 **ground**

1232 The BDT and invariant mass PDFs for the combinatorial background inside the signal
1233 regions are extracted from data by interpolating the number of expected events using the
1234 invariant mass sidebands for each BDT bin. In Fig. 60 the invariant mass distribution for
1235 events in each BDT bin is shown and the single exponential function used to interpolate
1236 the expected number of background events in the signal regions is superimposed.

1237 The sidebands considered for the fit are defined by $m_{\mu\mu} = [4900\text{--}5000] \text{ MeV}/c^2$ and
1238 $m_{\mu\mu} = [5433\text{--}6000] \text{ MeV}/c^2$. The boundaries of in the lower sidebands are chosen to
1239 exclude mass regions polluted by background sources other than combinatorial: cascading
1240 $b \rightarrow c\mu \rightarrow \mu\mu X$ decays below $4900 \text{ MeV}/c^2$ and $B_{(s)}^0 \rightarrow h^+h^-$ events with both hadrons
1241 misidentified as muons above $5000 \text{ MeV}/c^2$. The former contribution does not reach the
1242 signal region and consequently the fit remain unaffected by it. The latter contribution
1243 is modeled independently and should not affect the fit of the combinatorial background
1244 either. The lower boundary of the upper sideband is defined by the measured B_s^0 mass
1245 plus $60 \text{ MeV}/c^2$. The same rational was used in the two previous publications.

1246 The expected number of combinatorial background events in the B^0 and B_s^0 mass
1247 windows as well as in the full blinded region (more precisely the region defined as the
1248 union of the B_s^0 and the B^0 signal region) and the indices of the exponential functions
1249 are shown in Table 32, column labeled 'blind'. Note that due to the insufficient number
1250 of events in the sidebands, the fit of the last bin of BDT would be meaningless. Thus
1251 the a fit is performed to the events with $0.8 < \text{BDT} < 1$. (two last bins together). The
1252 index of the resulting exponential is then used as constant input to fit the two last bins
1253 independently.

1254 In Table 32 the results obtained with a double exponential function fitted to the
1255 extended invariant mass range $[4173\text{--}5000] \text{ MeV}/c^2$ and $[5433\text{--}6573] \text{ MeV}/c^2$ and with a
1256 linear fit in the nominal mass sidebands are also shown.

1257 A systematic uncertainty is introduced if the yields in the blinded regions differ by
1258 more than 1σ between the fit models. This is the case for the three first bins and the
1259 additional relative systematic are 1.8 %, 17.6 % and 21.8 %, respectively.

1260 In order to compare with the published analysis based on the 0.37 fb^{-1} dataset, the
1261 fit of the background has been repeated with the old BDT binning, given by four equally
1262 spaced bins. The fit has been performed using the same single exponential function and
1263 the same mass sidebands described above. The fits to the invariant mass spectrum divided
1264 in BDT bins are shown in Fig. 61. The expected number of background events in the B_s^0
1265 search window, in the B^0 search window and in the blinded region are shown in Table 33.
1266 In Table 34 are reported the results found for the 0.37 fb^{-1} dataset. If we rescale the
1267 numbers in Table 34 by the ratio of luminosities, $1025/370 = 2.77$, in the current analysis
1268 we see a dramatic background reduction in the first three BDT bins with respect to the
1269 projections from the 0.37 fb^{-1} analysis. This reduction is mostly due to the introduction

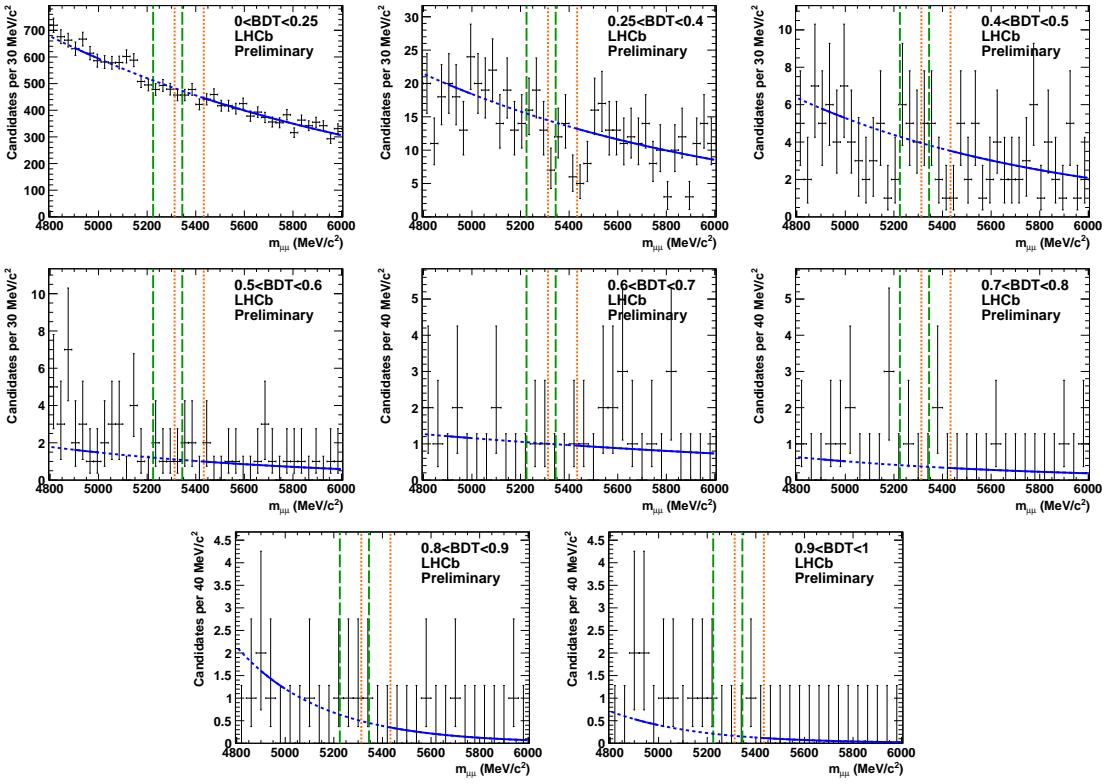


Figure 60: Distribution of the $\mu^+\mu^-$ invariant mass for events in each BDT output bin. The curve shows the model used to fit the sidebands and extract the expected number of combinatorial background events in the B_s^0 and B^0 signal regions, delimited by the vertical dotted orange and dashed green lines respectively. Only events in the region in which the line is solid have been considered in the fit.

of the BDTS, and in the low BDT bins, to the introduction of the DLL cuts that reduce all the background components not represented by real muons. The background expectations in the last BDT bin is, instead, in agreement with the projections within the (large!) uncertainties: this is expected since neither the BDTS or the DLL cuts affect events with very high BDT values.

The final shapes for signal and background BDT pdf's are shown in Fig. 62.

Table 32: Expected yields in the B_s^0 and B^0 mass windows as well as in the union of those for different fit models and per bin of BDT. The indices of the exponential function is shown in the case of the single exponential model.

Bin	$k [10^{-4}]$	exponential in [4900,5000] and [5433,6000] GeV			double exponential in [4173,5000] and [5433,6573] GeV Union of search windows
		B_s	B_0	Union of search windows	
1	$-6.62^{+0.30}_{-0.30}$	1877^{+22}_{-21}	1990^{+21}_{-21}	3356^{+35}_{-35}	3295^{+23}_{-23}
2	$-7.7^{+1.8}_{-1.7}$	$55.9^{+3.8}_{-3.6}$	$59.9^{+3.7}_{-3.6}$	$100.5^{+6.3}_{-6.0}$	$82.9^{+3.8}_{-3.7}$
3	$-9.3^{+3.4}_{-3.3}$	$15.1^{+2.0}_{-1.8}$	$16.4^{+2.0}_{-1.9}$	$27.3^{+3.4}_{-3.1}$	$21.5^{+2.0}_{-1.9}$
4	$-9.2^{+6.5}_{-6.2}$	$4.25^{+1.08}_{-0.93}$	$4.61^{+1.11}_{-0.95}$	$7.7^{+1.8}_{-1.6}$	$7.7^{+1.2}_{-1.1}$
5	$-4.6^{+8.0}_{-7.4}$	$2.96^{+0.94}_{-0.80}$	$3.08^{+0.87}_{-0.73}$	$5.3^{+1.5}_{-1.2}$	$4.50^{+0.82}_{-0.72}$
6	-10^{+13}_{-12}	$1.07^{+0.58}_{-0.43}$	$1.17^{+0.61}_{-0.45}$	$1.95^{+1.01}_{-0.75}$	$1.37^{+0.39}_{-0.27}$
7	$-28.6^{+9.6}_{-10.4}$	$1.26^{+0.59}_{-0.45}$	$1.63^{+0.76}_{-0.58}$	$2.51^{+1.17}_{-0.89}$	$1.52^{+0.54}_{-0.40}$
8	$-28.6^{+9.6}_{-10.4}$	$0.42^{+0.37}_{-0.23}$	$0.54^{+0.48}_{-0.30}$	$0.84^{+0.74}_{-0.46}$	$0.45^{+0.25}_{-0.13}$

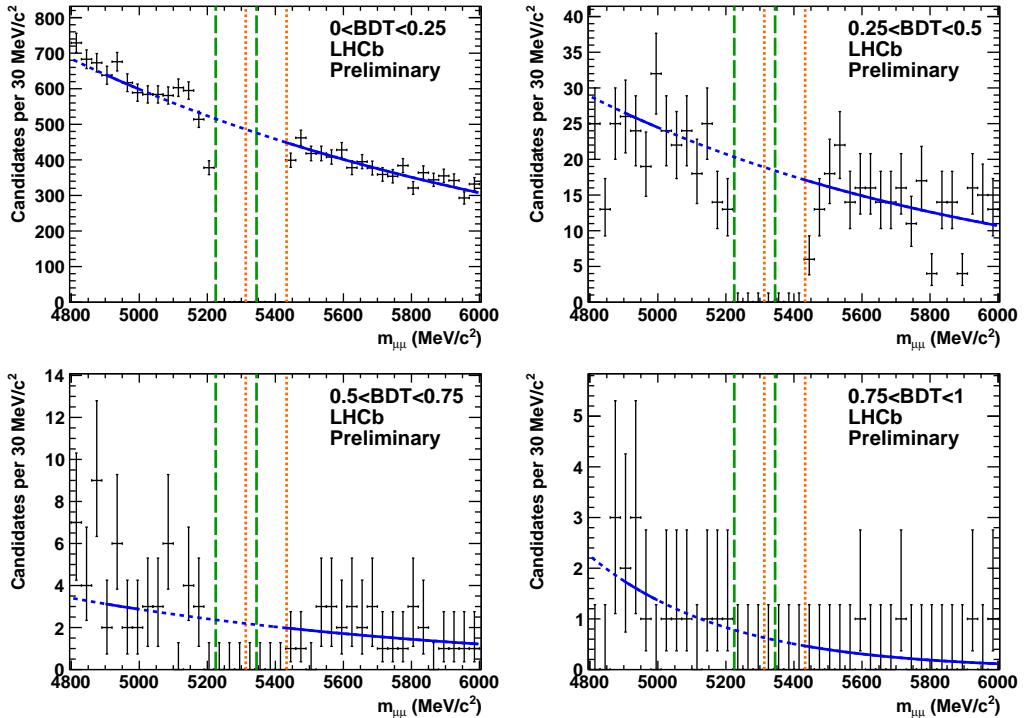


Figure 61: Distribution of the $\mu^+\mu^-$ invariant mass for events in each BDT bin for the binning used in the 0.37 fb^{-1} analysis. The curve shows the model used to fit the sidebands and extract the expected number of combinatorial background events in the B_s^0 and B^0 signal regions, delimited by the vertical dotted orange and dashed green lines respectively. Only events in the region in which the line is solid have been considered in the fit.

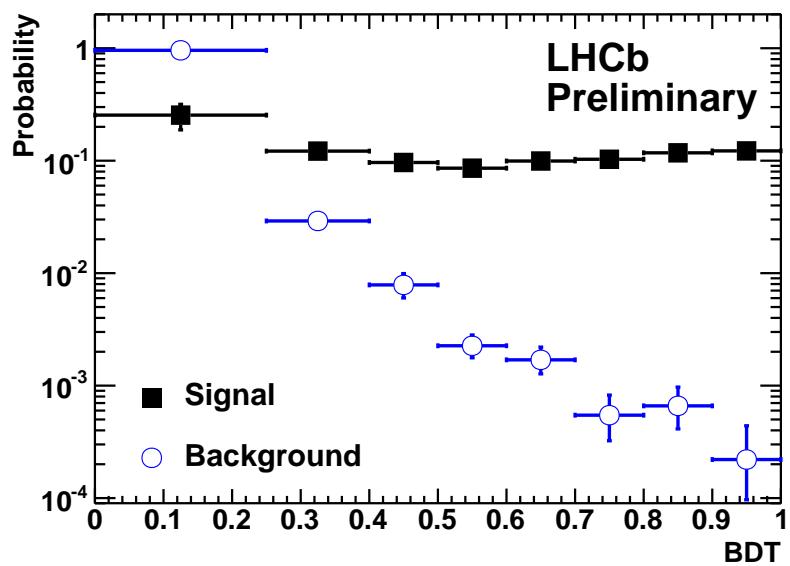


Figure 62: BDT pdf's for signal and background.

Table 33: Analysis with 1 fb^{-1} with the old binning: index of the exponential and number of expected background events in B_s^0 and B^0 mass window, as well as in the full blinded region for the single exponential function fitted to the mass sidebands for the four BDT bins

Bin	expo, 4.9-5, right			
	$k [10^{-4}]$	B_s	B_0	blind
1	$-6.67^{+0.30}_{-0.30}$	1888^{+22}_{-22}	2003^{+21}_{-21}	3378^{+35}_{-35}
2	$-8.3^{+1.5}_{-1.5}$	$72.7^{+4.3}_{-4.1}$	$78.2^{+4.3}_{-4.1}$	$131.1^{+7.2}_{-6.9}$
3	$-8.6^{+4.6}_{-4.4}$	$8.4^{+1.5}_{-1.4}$	$9.1^{+1.5}_{-1.4}$	$15.2^{+2.5}_{-2.3}$
4	$-24.8^{+8.5}_{-8.8}$	$2.20^{+0.79}_{-0.64}$	$2.73^{+0.96}_{-0.78}$	$4.3^{+1.5}_{-1.2}$

Table 34: Analysis of 0.37 fb^{-1} : index of the exponential and number of expected background events in B_s^0 and B^0 mass window, as well as in the full blinded region for the single exponential function fitted to the mass sidebands for the four BDT bins.

Bin	expo, 4.9-5, right				Observed
	$k [10^{-4}]$	B_s	B_0	blind	
1	$-7.59^{+0.23}_{-0.23}$	3326^{+29}_{-29}	3550^{+28}_{-28}	5902^{+47}_{-46}	5672
2	$-4.5^{+2.1}_{-2.1}$	$41.0^{+3.3}_{-3.2}$	$42.6^{+3.0}_{-2.8}$	$71.7^{+5.1}_{-4.8}$	73
3	$-11.3^{+5.0}_{-4.9}$	$6.9^{+1.4}_{-1.2}$	$7.6^{+1.4}_{-1.3}$	$12.4^{+2.3}_{-2.1}$	10
4	-12^{+17}_{-16}	$0.65^{+0.47}_{-0.32}$	$0.71^{+0.50}_{-0.34}$	$1.17^{+0.81}_{-0.55}$	7

1276 10 Normalization

1277 To estimate the signal branching ratio, we normalize the number of observed signal events
 1278 to the number of events of a calibration channel of a well known branching ratio. We
 1279 have used three complementary normalization channels: $B^+ \rightarrow J/\psi(\mu^+\mu^-)K^+$, $B_s^0 \rightarrow$
 1280 $J/\psi(\mu^+\mu^-)\phi(KK)$ and $B^0 \rightarrow K^+\pi^-$. The first two decays have similar trigger and
 1281 muon identification efficiencies as the signal but different number of particles in the final
 1282 state, while the third channel has very similar topology but very different trigger.

1283 The selection of the normalization channels was described in Sec. 5, it has been de-
 1284 signed to be very similar to the selection of the signal events, such that the systematic
 1285 uncertainties cancel in the ratio of efficiencies and the knowledge of the absolute integrated
 1286 luminosity and the $b\bar{b}$ cross-section are not needed.

To translate of the number of observed events into a branching ratio we use the following equation:

$$1287 \text{BR} = \text{BR}_{\text{cal}} \times \frac{\epsilon_{\text{cal}}^{\text{REC}} \epsilon_{\text{cal}}^{\text{SEL|REC}} \epsilon_{\text{cal}}^{\text{TRIG|SEL}}}{\epsilon_{\text{sig}}^{\text{REC}} \epsilon_{\text{sig}}^{\text{SEL|REC}} \epsilon_{\text{sig}}^{\text{TRIG|SEL}}} \times \frac{f_{\text{cal}}}{f_{B_q^0}} \times \frac{N_{B_q^0 \rightarrow \mu^+\mu^-}}{N_{\text{cal}}} = \alpha_{\text{cal}} \times N_{B_q^0 \rightarrow \mu^+\mu^-}, \quad (19)$$

1287 where $f_{B_q^0}$ and f_{cal} are the probabilities that a b -quark fragments into a B_q^0 and into the
 1288 b -hadron relevant for the chosen calibration mode. BR_{cal} is the branching ratio and N_{cal}
 1289 is the number of selected events of the calibration channel. The efficiency is separated in
 1290 three factors: ϵ^{REC} is the efficiency to reconstruct all the tracks of the decay including
 1291 the geometrical acceptance of the detector; $\epsilon^{\text{SEL|REC}}$ is the efficiency to select the events
 1292 which have been reconstructed; $\epsilon^{\text{TRIG|SEL}}$ is the efficiency of the trigger on reconstructed
 1293 and selected events. The sub-indexes indicate if the efficiency refers to the signal (sig)
 1294 or the calibration channel (cal). Finally, α_{cal} is the normalization factor (or single event
 1295 sensitivity) and $N_{B_q^0 \rightarrow \mu^+\mu^-}$ the number of observed signal events.

1296 In this Section we discuss the estimation of the factors that enter in the computation
 1297 of α_{cal} .

1298 10.1 Ratio of production fractions

1299 The ratio $f_{B^0}/f_{B_s^0} = f_{B^+}/f_{B_s^0}$ has been measured at the Z resonance at LEP to be $3.91 \pm$
 1300 0.35 and at Tevatron to be 2.75 ± 0.40 . In the previous analysis note [71] we used the
 1301 world average 3.71 ± 0.47 [24].

1302 LHCb has recently determined the fragmentation fraction f_s/f_d in two different ways:
 1303 using the relative abundance of $B_s^0 \rightarrow D_s^-\pi^+$, $B^0 \rightarrow D^-K^+$ and $B^0 \rightarrow D^-\pi^+$ [58]
 1304 and using semileptonic $B \rightarrow DX$ decays [57]. The average of the two LHCb results is
 1305 $f_s/f_d = 0.267^{+0.021}_{-0.020}$ [56]. This number will be used in the limit computation.

1306 10.2 Branching ratios of the control channels

1307 Table 35 shows the values of the \mathcal{B} of the three control channels. The error on the \mathcal{B} of
 1308 $B_s^0 \rightarrow J/\psi(\mu^+\mu^-)\phi(KK)$ is 26% and is the dominant uncertainty when this channel is

1309 used for the normalization.

Table 35: Branching ratios of channels used for the normalization

$B^+ \rightarrow J/\psi K^+$	$B_s^0 \rightarrow J/\psi \phi$	$B^0 \rightarrow K^+ \pi^-$
$(6.01 \pm 0.21) \times 10^{-5}$	$(3.4 \pm 0.9) \times 10^{-5}$	$(1.94 \pm 0.06) \times 10^{-5}$

1310 10.3 The ratio of reconstruction efficiencies

1311 The acceptance and reconstruction efficiencies depends on the phase space of the final
1312 decay products, the acceptance of the detector and the efficiency of track finding algo-
1313 rithms. We compute the ratio of acceptance and reconstruction efficiencies using the MC
1314 simulation. The differences between the simulation and data are treated as systematic
1315 uncertainties. In the following we describe the factors that enters in the estimation of
1316 ϵ^{REC} .

1317 10.3.1 Reconstruction efficiencies from MC simulation

1318 The reconstruction efficiency factorizes in two parts: the detector acceptance (or gener-
1319 ation efficiency) ϵ_{gen} and the reconstruction efficiency $\epsilon_{\text{reco/gen}}$. The detector acceptance
1320 is defined as the fraction of the tracks in the final state that are inside the LHCb accep-
1321 tance¹⁷. The reconstruction efficiency is defined as the efficiency to reconstruct all the
1322 tracks in the final state.

1323 The generation acceptance and reconstruction efficiency, and its product are summa-
1324 rized in Table 36. They have been computed using MC simulation, and its uncertainty is
1325 smaller than 0.1%.

Table 36: Generator level efficiency ϵ_{gen} and reconstruction efficiencies $\epsilon_{\text{reco/gen}}$ for signal and control channels. The muonID efficiency (for $B_{(s)}^0 \rightarrow h^+ h^-$ acceptance) is taken into account in $\epsilon_{\text{reco/gen}}$. The uncertainty on the quoted numbers is below 0.1%.

	$B_s^0 \rightarrow \mu^+ \mu^-$	$B^+ \rightarrow J/\psi K^+$	$B_s^0 \rightarrow J/\psi \phi$	$B_{(s)}^0 \rightarrow h^+ h^-$	$B_d^0 \rightarrow J/\psi K^{*0}$
ϵ_{gen}	17.5%	15.3%	16.0%	17.7%	14.8%
$\epsilon_{\text{reco/gen}}$	58.3%	38.6%	25.0%	44.5%	24.8%
total efficiency					
$\epsilon_{\text{reco/gen}} \times \epsilon_{\text{gen}}$	10.2%	5.9%	4.0%	7.9%	3.7%

1326 When normalizing to $B_{(s)}^0 \rightarrow h^+ h^-$, we require that the two hadrons are within the
1327 muon detector acceptance to minimize the differences between the signal and this control

¹⁷The acceptance is defined by the interval [10,400] mrad in the polar angle. The limit at 400 mrad is a bit larger than the actual detector acceptance (~ 330 mrad) to avoid losses of events due to the magnetic field.

channel. The uncertainties associated to the muon system acceptance will be treated together with the muonID efficiency and described later in this subsection.

The ratio of the reconstruction efficiency depends on the reconstruction efficiency of an extra track (the kaon) when the normalization is done with the $B^+ \rightarrow J/\psi K^+$ channel, of two extra tracks when $B_s^0 \rightarrow J/\psi \phi$ is used and on the different phase space of the muons in the final state for signal and $J/\Psi \rightarrow \mu^+ \mu^-$.

We determine the track reconstruction efficiency from the simulation and correct it using the tracking efficiency map¹⁸ as provided by the tracking group [55]. The correction factors depend on the pseudo-rapidity and the momentum of the tracks. The correction factor c to the ratio of reconstruction efficiencies taken from the simulation is then:

$$\begin{aligned} c\left(\frac{B^+ \rightarrow J/\psi K^+}{B_s^0 \rightarrow \mu^+ \mu^-}\right) &= 1.002 \pm 0.003 \pm 0.004 \pm 0.004 \pm 0.014, \\ c\left(\frac{B_s^0 \rightarrow J/\psi \phi}{B_s^0 \rightarrow \mu^+ \mu^-}\right) &= 1.022 \pm 0.005 \pm 0.008 \pm 0.008 \pm 0.027, \\ c\left(\frac{B_{(s)}^0 \rightarrow h^+ h'^-}{B_s^0 \rightarrow \mu^+ \mu^-}\right) &= 0.998 \pm 0.003, \end{aligned}$$

where the first error is the statistical error from the tracking efficiency map. The second is the correlated systematic error for the mapping method, the third accounts for the different reconstruction versions used in this analysis (reco12) and in the map (reco10). Finally the fourth accounts for uncertainty in the hadronic interactions with material [55]. They have to be added linearly for each additional track in the ratio. These factors are applied as correction to the numbers given in Table 36 for the computation of the normalization factor.

10.3.2 Validation of the ratio of reconstruction efficiencies using data

The signal $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decay and the control channel $B^+ \rightarrow J/\psi K^+$ differ by one track in the final state. The acceptance and reconstruction efficiency for this extra-track can be probed by using the ratio of events of $B^+ \rightarrow J/\psi K^+$ with respect $B_d^0 \rightarrow J/\psi K^{*0}$, which contains four tracks in the final state. The ratio of reconstruction efficiencies can be written as:

$$\frac{\epsilon^{\text{REC}}(B^+ \rightarrow J/\psi K^+)}{\epsilon^{\text{REC}}(B_s^0 \rightarrow \mu^+ \mu^-)} \simeq \frac{\epsilon^{\text{REC}}(B_d^0 \rightarrow J/\psi K^{*0})}{\epsilon^{\text{REC}}(B^+ \rightarrow J/\psi K^+)}. \quad (20)$$

The procedure to estimate that ratio was described in the previous analysis note [?]. We are using here the same procedure. The selection cuts for $B_d^0 \rightarrow J/\psi K^{*0}$ and $B^+ \rightarrow J/\psi K^+$ selections can be found in Table 37. To eliminate the small contamination of $\phi \rightarrow K^+ K^-$ decays that pollute the K^* peak, the K^* mass is further constrained. Fig. 63 shows the fit to the invariant mass of the $B^+ \rightarrow J/\psi K^+$ and $B_d^0 \rightarrow J/\psi K^{*0}$ distributions.

¹⁸The map has recently been updated by the tracking group and gives the correction of MC10 vs. Stripping13b data. For Stripping 17 data, which is used in this analysis, the tracking group recommends to add an additional uncertainty of 0.4%.

Table 37: Selection of $B^+ \rightarrow J/\psi K^+$ and $B_d^0 \rightarrow J/\psi K^{*0}$ used to obtain the ratio of reconstruction efficiencies. Note that the selections have been retuned with respect to the previous analysis note.

	Cut
$B^+ \rightarrow J/\psi K^+, B_d^0 \rightarrow J/\psi K^{*0}$	VDS(J/ψ) > 25 PIDk > 0
$B_d^0 \rightarrow J/\psi K^{*0}$	IPS(K^*) > 5 IPS(K^\pm) > 2 IPS(π^\pm) > 2 $\Delta m(K^*) < 40 \text{ MeV}/c^2$ vertex $\chi^2(K^*) < 9$

For the simulation, we have used a $b\bar{b} \rightarrow \mu^+\mu^-X$ sample to which we have applied the same selection than to the data. The simulated sample has been prepared as described in Sec. 10.4.1. We have obtained the following number of events:

$$N_{\text{data}}(B^+ \rightarrow J/\psi K^+) = 291\,632 \pm 560 \quad N_{\text{MC}}(B^+ \rightarrow J/\psi K^+) = 2653 \pm 60 \\ N_{\text{data}}(B_d^0 \rightarrow J/\psi K^{*0}) = 84184 \pm 297 \quad N_{\text{MC}}(B_d^0 \rightarrow J/\psi K^{*0}) = 802 \pm 29$$

From these values, we estimate the ratio data/MC of the ratio of the reconstruction efficiencies of:

$$\frac{\epsilon_{B^+ \rightarrow J/\psi K^+}^{\text{REC,data}} / \epsilon_{B_d^0 \rightarrow J/\psi K^{*0}}^{\text{REC,data}}}{\epsilon_{B^+ \rightarrow J/\psi K^+}^{\text{REC,MC}} / \epsilon_{B_d^0 \rightarrow J/\psi K^{*0}}^{\text{REC,MC}}} = 1.032 \pm 0.04 \pm 0.06 \quad (21)$$

where the first error is due to the statistics and it is dominated by the number of MC events, the second is the uncertainty associated with the \mathcal{B} of $B_d^0 \rightarrow J/\psi K^{*0}$ and $B^+ \rightarrow J/\psi K^+$. This estimation from the data validates the use of the reconstruction efficiency from simulation.

10.3.3 Muon detector acceptance and muonID efficiency

In Sect. 6 the muon detector acceptance (acc_μ) and the muonID efficiency (ϵ_μ) have been determined as a function of p and p_T using control samples.

The acceptance×efficiency 2D map in p and p_T bins obtained both in data and in simulated events are then folded into the p, p_T spectrum of the muons from reconstructed and selected $B_s^0 \rightarrow \mu^+\mu^-$, $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi \phi$ Monte Carlo events and the average values $\langle \epsilon_\mu \times acc_\mu \rangle_{\text{data}}$ and $\langle \epsilon_\mu \times acc_\mu \rangle_{\text{MC}}$ are obtained. The ratio $C_{\mu ID}(B_s^0 \rightarrow \mu^+\mu^-, B^+ \rightarrow J/\psi K^+, B_s^0 \rightarrow J/\psi \phi) = \langle \epsilon_\mu \times acc_\mu \rangle_{\text{data}} / \langle \epsilon_\mu \times acc_\mu \rangle_{\text{MC}}$ is then used to correct the muonID efficiency evaluated on Monte Carlo for a given channel. For the $B_{(s)}^0 \rightarrow h^+h'^-$ sample, the selection efficiency obtained in Monte Carlo is corrected only for the ratio $C_{\mu ID}(B_{(s)}^0 \rightarrow h^+h'^-) = \langle acc_\mu \rangle_{\text{data}} / \langle acc_\mu \rangle_{\text{MC}}$.

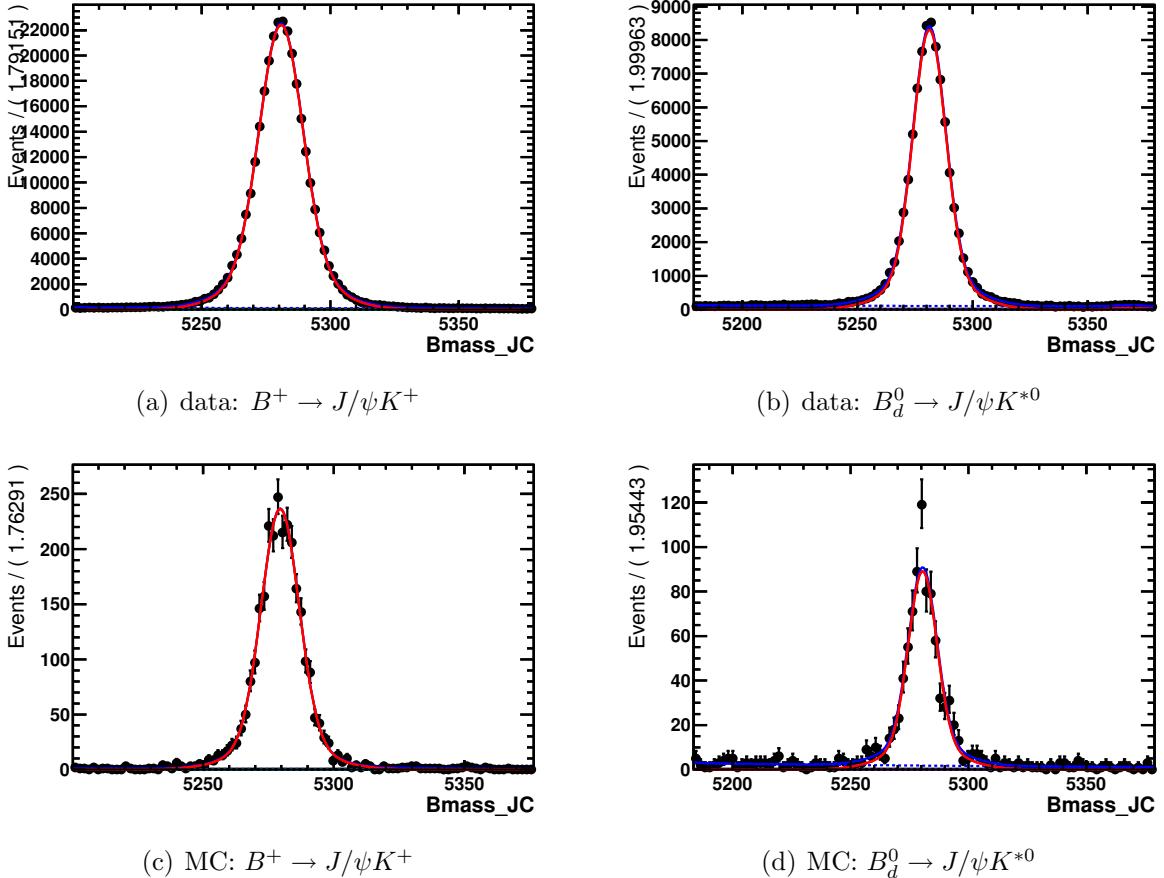


Figure 63: Invariant mass distributions of $B^+ \rightarrow J/\psi K^+$ (left) and $B_d^0 \rightarrow J/\psi K^{*0}$ (right) candidates after the selection described in the text. FIXME: correct Bu data!

1357 The corrections to be applied to the muonID efficiencies determined on simulated
 1358 events are listed in Table 38 for $B_s^0 \rightarrow \mu^+ \mu^-$, $B^+ \rightarrow J/\psi K^+ / B_s^0 \rightarrow J/\psi \phi$ and $B_{(s)}^0 \rightarrow$
 1359 $h^+ h'^-$ events: the corrections are obtained from the $b \rightarrow J/\psi X$ (2-body) calibration
 1360 sample and from events where the μ_{probe} is TIS with respect to the trigger (first row of
 1361 Table 38). Systematic uncertainties on these corrections are then obtained by comparing
 1362 the results using a different calibration sample (J/ψ from $B^+ \rightarrow J/\psi K^+$) and requiring
 1363 the μ_{tag} to be TOS with respect to the trigger. For completeness, in the last three rows
 1364 of the same Table 38, we report the corrections obtained by using the 1D curves $\epsilon_\mu \times acc_\mu$
 1365 or the acc_μ parameterized as a function of p or p_T but not of both. This is done in order
 1366 to compare the new results to the ones obtained with the 370 pb^{-1} dataset where, due to
 1367 the limited statistics, only the 1D parameterizations were performed.

The ratios of these corrections are used in the computation of the normalization factors.
 In the case of the ratio between the $B^+ \rightarrow J/\psi K^+$, $B_s^0 \rightarrow J/\psi (\mu^+ \mu^-) \phi(KK)$ and $B_{(s)}^0 \rightarrow$

$\mu^+\mu^-$ we get:

$$\frac{C_{\mu ID}(B \rightarrow J/\psi X)}{C_{\mu ID}(B_{(s)}^0 \rightarrow \mu^+\mu^-)} = 1.0108 \pm 0.0011_{stat} \pm 0.0004_{sample} \pm 0.0073_{trigger}, \quad (22)$$

where the differences between data and MC largely cancel in the ratio between $B_{(s)}^0 \rightarrow \mu^+\mu^-$ and $J/\psi \rightarrow \mu^+\mu^-$.

In the case of the ratio between $B_{(s)}^0 \rightarrow h^+h'^-$ and $B_{(s)}^0 \rightarrow \mu^+\mu^-$ (we remind here that the muon detector acceptance is also required for $B_{(s)}^0 \rightarrow h^+h'^-$), with the same technique we get:

$$\frac{C_{\mu ID}(B_{(s)}^0 \rightarrow h^+h'^-)}{C_{\mu ID}(B_{(s)}^0 \rightarrow \mu^+\mu^-)} = 0.9698 \pm 0.0006_{stat} \pm 0.0042_{sample} \pm 0.0032_{trigger}, \quad (23)$$

which is entirely dominated by the remaining data-MC differences in the muonID efficiencies for $B_{(s)}^0 \rightarrow \mu^+\mu^-$.

These results can be compared with the same ratios obtained with 370 pb^{-1} :

$$\frac{C_{\mu ID}(B \rightarrow J/\psi X)}{C_{\mu ID}(B_{(s)}^0 \rightarrow \mu^+\mu^-)} = 0.997 \pm 0.002_{stat} \pm 0.013_{syst}, \quad (24)$$

and

$$\frac{C_{\mu ID}(B_{(s)}^0 \rightarrow h^+h'^-)}{C_{\mu ID}(B_{(s)}^0 \rightarrow \mu^+\mu^-)} = 0.971 \pm 0.002_{stat} \pm 0.019_{syst}, \quad (25)$$

The results obtained with 1 fb^{-1} agree nicely with the ones obtained with 370 pb^{-1} and have a smaller systematic uncertainty. In fact the corrections in the previous analysis were obtained using only 1D parameterizations of the acceptance and efficiency and therefore their systematic uncertainty was dominated by the choice of the parameterization (p or p_T) and/or by the choice of the sample used to calibrate the curves. The 2D parameterization allows to eliminate this source of error (0.04%) for $C_{\mu ID}(B \rightarrow J/\psi X)/C_{\mu ID}(B_{(s)}^0 \rightarrow \mu^+\mu^-)$ and to reduce it (0.4%) for $C_{\mu ID}(B_{(s)}^0 \rightarrow h^+h'^-)/C_{\mu ID}(B_{(s)}^0 \rightarrow \mu^+\mu^-)$.

A further effect has been taken into account in the case of normalization to $B_{(s)}^0 \rightarrow h^+h'^-$, which is the interaction of the hadrons with the detector material. We estimated in the previous analysis note that this effect introduces an extra uncertainty in the tracking efficiency reconstruction of 1% per track.

Finally, since the DLL cut $DLL(K - \pi) < 10 \& DLL(\mu - \pi) > -5$ has been used to reduce the rate of peaking background, we have to correct MC in BDT bins with the efficiency evaluated on data (as explained in Sect. ??). Fig. 64 shows the applied bin-by-bin correction.

Table 38: .

sample	trigger	$C_{\mu ID}(B_s^0 \rightarrow \mu^+ \mu^-)$	$C_{\mu ID}(B^+ \rightarrow J/\psi K^+)$ or $C_{\mu ID}(B_s^0 \rightarrow J/\psi \phi)$	$C_{\mu ID}(B_{(s)}^0 \rightarrow h^+ h'^{-})$
p, p_T	$b \rightarrow J/\psi X$ (2-body)	TIS-probe	1.0319(12)	1.0006(8)
p, p_T	$B^+ \rightarrow J/\psi K^+$ (3-body)	TIS-probe	1.0254(16)	0.9987(9)
p, p_T	$b \rightarrow J/\psi X$ (2-body)	TOS-tag	1.0228(14)	0.9952(8)
p_T	$b \rightarrow J/\psi X$ (2-body)	TIS-probe	1.0284(12)	1.0001(8)
p	$b \rightarrow J/\psi X$ (2-body)	TIS-probe	1.0466(19)	1.0062(15)
p	$B^+ \rightarrow J/\psi K^+$ (3-body)	TIS-probe	1.0313(15)	0.9980(8)

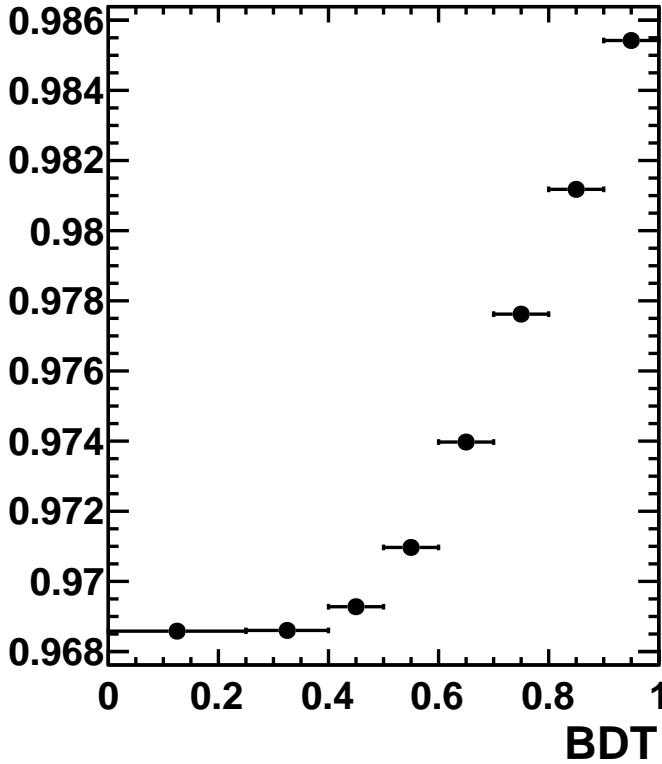


Figure 64: $B_s^0 \rightarrow \mu^+ \mu^-$ MC efficiency for the DLL cut $DLL(K - \pi) < 10 \& DLL(\mu - \pi) > -5$ calculated using ϵ_{DLLcut}^{data} from data.

10.4 Ratio of selection efficiencies

The selection criteria for signal and normalization channels were described in Sect. 5. The selection is identical to the one used in the previous analysis note, except for the $B_s^0 \rightarrow J/\psi (\mu^+ \mu^-) \phi(KK)$ where a PID requirement has been imposed to the kaons of the ϕ . The selection efficiencies are determined using MC simulation and then cross-checked

1394 with data. The main difference between the MC and the data is in the IP distribution.
 1395 In order to minimize such a difference, we have applied a technique developed by the
 1396 tracking group to smear on the MC sample the parameters of the tracks at the origin,
 1397 and thus reproduce the observed IP resolution in data. The selection efficiencies are then
 1398 determined from the smeared MC.

1399 As the uncertainties cancel in the ratio of efficiencies, we have not assigned systematic
 1400 uncertainties in the ratio of selection efficiencies for the cuts that are identical between the
 1401 signal and the control channel. But a systematic error on the ratio of selection efficiencies
 1402 has been estimated on the cuts that are different between the signal and the control
 1403 channel.

1404 **10.4.1 Track parameter smearing**

1405 The tracking group has proposed a method [52] to adjust the MC precision of the first
 1406 state of a track to the one observed in data. The width of the IP distribution of minimum
 1407 bias tracks is parametrized as a function of p_T and ϕ (the angle of the track in the
 1408 transverse plane with respect the x direction) for both data and MC. The difference of
 1409 both distributions is then used as width for a Gaussian smearing of the MC. This method
 1410 has been applied to the MC tracks of the B decay products, and all the quantities of the
 1411 selection have been recomputed.

1412 **10.4.2 Selection efficiencies from simulation**

1413 The selection efficiencies have been computed using the MC in three different scenarios:
 1414 the original (unsmeared) MC, the smeared MC and a MC where the smearing is over-
 1415 done by 50% to estimate possible residual effects. Notice that the selection efficiency is
 1416 determined on events where all signal tracks have been reconstructed.

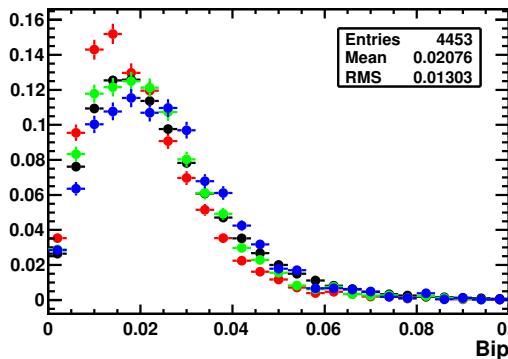
1417 The biggest difference in the selection efficiencies for the $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow$
 1418 $J/\psi \phi$ normalization channels comes from additional IP χ^2 requirements on the extra tracks
 1419 in the normalization channel, and, to a smaller extend, from the different kinematics of
 1420 the muons from the J/ψ and the B_s^0 .

Table 39: Selection efficiencies for signal and control channels. The efficiencies include the efficiency of hte fiducial cuts as well as the BDTS cut efficiencies.

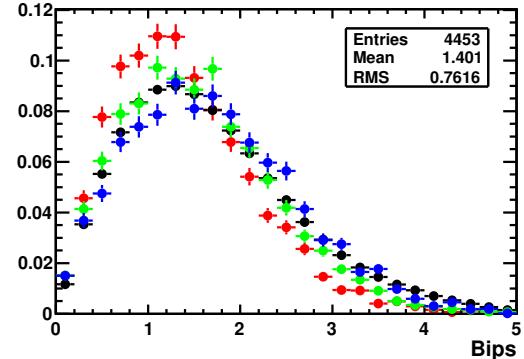
channel	normal	smeared	50% over smeared
$B_s^0 \rightarrow \mu^+ \mu^-$	49.8%	47.2%	44.0%
$B^0 \rightarrow K^+ \pi^-$	55.1%	52.5%	48.8%
$B^+ \rightarrow J/\psi K^+$	41.7%	39.5%	36.3%
$B_s^0 \rightarrow J/\psi (\mu^+ \mu^-) \phi(KK)$	32.3%	30.4%	28.1%

1421 For the $B_{(s)}^0 \rightarrow h^+h'^-$ normalization channel, the selection efficiency is higher than for
 1422 the signal as the tight (60 MeV) mass window is not applied to the $B_{(s)}^0 \rightarrow h^+h'^-$ channel.

1423 To validate the smearing procedure, the selection variables can be compared in data
 1424 and MC. The channel $B^+ \rightarrow J/\psi K^+$ is best suitable for this as it is almost background
 1425 free and has a very number of events.



(a) $B^+ \rightarrow J/\psi K^+$: Bip



(b) $B^+ \rightarrow J/\psi K^+$: Bips

Figure 65: In all figures: black : data, red: unsmeared MC, green: smeared MC, blue: oversmeared MC. It is seen that the three MC samples cover the whole range of the data distribution.

1426 10.4.3 PID efficiency determination

1427 The RICH PID efficiency for $B_s^0 \rightarrow J/\psi \phi$ is not included in Table 39 as it is known
 1428 that the simulation does not describe the PID efficiency in data correctly. Instead, it is
 1429 measured directly on data using two different tight selection criteria (additional criteria:
 1430 $\tau > 1$ ps, B vertex $\chi^2 < 25$ or all track IPS > 9 , B VDS > 25). It has been verified on
 1431 MC that the PID efficiency after the tight selection is within 1% identical to the original
 1432 RICH PID.

The PID efficiency of $DLL(K - \pi) > 0$ for both kaons has been determined to be

$$\epsilon = (91.4 \pm 0.9_{\text{stat}} \pm 1.0_{\text{syst}})\% \quad (26)$$

1433 efficient and 1% additional systematic uncertainty is added to account for selection de-
 1434 pendent PID effects.

1435 10.4.4 systematic uncertainties

1436 As the selection efficiencies enter in the normalization formula as ratios, the systematic
 1437 effects affecting equally the signal and the normalization channel cancel.

1438 - Unbalanced selection criteria

1439 The effect of the unbalanced selection criteria of the B^+ (IPS of the kaon) has been
 1440 studied using a PID based selection for $B^+ \rightarrow J/\psi K^+$, as in the previous analysis
 1441 note. The MC predicted value have found to be in good agreement with the data.

1442 - **Vertex χ^2**

1443 The efficiency of the requirement on the vertex quality (χ^2) has been compared
 1444 between data and simulation and has been found in good agreement.

1445 - **Clone rejection**

In principle, the requirement of the Kullback-Liebler distance larger than 5000 could introduce a difference between the signal and the three normalization channels. However, this has been studied by reducing the data sample to events with exactly one candidate and measuring the efficiency of this selection directly:

$$B^+ \rightarrow J/\psi K^+ \text{ data} : 99.7 \pm 0.6\%, \quad (27)$$

$$B_s^0 \rightarrow J/\psi \phi \text{ data} : 99.1 \pm 1.9\% \quad (28)$$

which compares very well to the efficiencies measured in the simulation:

$$B^+ \rightarrow J/\psi K^+ \text{ MC} : 99.88\%, \quad (29)$$

$$B_s^0 \rightarrow J/\psi \phi \text{ MC} : 99.6\%. \quad (30)$$

1446 As no additional difference between the signal and the normalization channels was
 1447 found, no additional systematic uncertainties have been added to the ratio of selection
 1448 efficiencies.

1449 **10.4.5 Ratio of selection efficiencies**

The total selection efficiency for the signal and control channels is given in Table 39. The ratio of selection efficiencies for the three normalization channels is determined from the smeared MC, the values are:

$$\epsilon_{B^+ \rightarrow J/\psi K^+}^{\text{SEL|REC}} / \epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{\text{SEL|REC}} = 0.836 \pm 0.01 \quad (31)$$

$$\epsilon_{B_s^0 \rightarrow J/\psi (\mu^+ \mu^-) \phi (KK)}^{\text{SEL|REC}} / \epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{\text{SEL|REC}} = 0.589 \pm 0.02 \quad (32)$$

$$\epsilon_{B^0 \rightarrow K^+ \pi^-}^{\text{SEL|REC}} / \epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{\text{SEL|REC}} = 1.11 \pm 0.01 \quad (33)$$

$$(34)$$

1450 where the errors are from MC statistics.

1451 The absolute selection efficiencies between the unsmeared MC and the over-smeared
 1452 sample vary by 5 – 7%, depending on the channel. However, the ratio between signal and
 1453 all three normalization channels stays constant within 1%, because the efficiency change
 1454 in the signal is cancelled by the corresponding change in the normalization channel.

1455 10.5 Ratio of trigger efficiencies

1456 The trigger efficiencies are estimated using the TISTOS method [53]. As in the previous
 1457 analysis note, for the control samples, we can apply directly this method on the data
 1458 while for the $B_s^0 \rightarrow \mu^+ \mu^-$ we do it in two steps. First, we compute the efficiency with the
 1459 TISTOS method on $J/\psi \rightarrow \mu\mu$ events selected by the detached J/ψ selection, in bins of
 1460 the p_T and IP. Then we reweight this efficiency using the $B_s^0 \rightarrow \mu^+ \mu^-$ spectra taken from
 1461 the MC. Only the muon trigger lines are used in this process (see Ref.[22] for detailed
 1462 discussion).

1463 The trigger lines that select the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ channel are the same lines which select
 1464 the normalization channels containing muons in the final state, $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow$
 1465 $J/\psi \phi$. In this case the ratio of trigger efficiencies is not expected to be very different from
 1466 unity, being the differences manly due to the different phase space covered by the muons
 1467 in the final state.

1468 On the contrary we will expect large correction in the case of the normalization with
 1469 a fully hadronic channel as the $B^0 \rightarrow K^+ \pi^-$. But, as it will be described later, some
 1470 factors cancel in the ratio and finally we normalize to $B_{(s)}^0 \rightarrow h^+ h^-$ TIS events and what
 1471 enter in the normalization is the trigger TIS efficiency.

1472 10.5.1 Estimation of the trigger efficiencies

1473 Events can be classified into two categories [53]: TIS (Trigger Independent of Signal): the
 1474 event would also have been triggered without the signal under study; TOS (Trigger On
 1475 Signal): the signal under study is sufficient to trigger the event. Note that an event can
 1476 also be TIS and TOS simultaneously (TIS&TOS), or neither TOS nor TIS. The LHCb
 1477 trigger system records all the information needed for such a classification.

The overall trigger efficiency on such events can be expressed as [54]:

$$\epsilon^{\text{TRIG}} = \frac{N^{\text{TRIG}}}{N^{\text{SEL}}} = \frac{N^{\text{TIS}}}{N^{\text{SEL}}} \frac{N^{\text{TRIG}}}{N^{\text{TIS}}} = \epsilon^{\text{TIS}} \frac{N^{\text{TRIG}}}{N^{\text{TIS}}} \quad (35)$$

1478 Both N^{TRIG} and N^{TIS} are observable quantities, ϵ^{TIS} is the efficiency to trigger without
 1479 any information from the signal.

1480 As the TIS events are mainly triggered by the decay products of the other b quark in
 1481 the event, the TIS efficiency ϵ^{TIS} can slightly depend on the p_T of the B signal via the $b\bar{b}$
 1482 correlation. However, as the signal and the normalization channels were selected in the
 1483 same way (see Sect. 5), the ϵ^{TIS} is independent of the considered decay and cancels in the
 1484 ratio.

The most precise measurement of ϵ^{TIS} can be done in the $B^+ \rightarrow J/\psi K^+$ channel, it
 is found to be

$$\epsilon^{\text{TIS}}(\text{L0} \times \text{HLT1}) = (5.13 \pm 0.9_{\text{stat}} \pm 0.4_{\text{syst}})\%, \quad (36)$$

1485 the first error is statistical and it is dominated by the number of TIS events. The system-
 1486 atic error has been evaluated in MC from the difference of the trigger efficiencies between
 1487 the true value and the one obtained from the TISTOS method.

1488 The requirement that events are TIS respect HLT2 reduces drastically the sample.
 1489 Nevertheless we are able to compute the total TIS efficiency: $\epsilon^{\text{TIS}}(\text{L0} \times \text{HLT1} \times \text{HLT2}) =$
 1490 $(1.86 \pm 0.07)\%$. As we did in the previous analysis note, we consider $\epsilon^{\text{TIS}}(\text{L0} \times \text{LT1})$ and
 1491 use when needed, the HLT2 efficiencies from the MC and assign a systematic uncertainty
 1492 to it.

1493 10.5.2 Ratio of trigger efficiency for $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ and $B \rightarrow J/\psi X$ channels

1494 *Main difference wrt 0.37 fb-1: efficiency map built not as function of $J/\psi(pT)$ and max
 1495 $IP(\text{muons})$ but as a function of the max pT and maxIP of the muons. This is due to the
 1496 change in the calibration sample from $b\bar{c}J/\psi X$ to $B u J/\psi K$*

The trigger efficiency for the $B \rightarrow J/\psi X$ channels estimated using the TISTOS method is:

$$\epsilon_{J/\psi}^{\text{TRIG|SEL}} = (87.2 \pm 0.4_{\text{stat}} \pm 3.8_{\text{syst}})\%, \quad (37)$$

1497 The systematic error is the combination of two errors: one associated to the TISTOS
 1498 method precision obtained from MC (3%) and the second one due to the 2.3% of $B \rightarrow$
 1499 $J/\psi X$ events that are triggered not using the muon triggers.

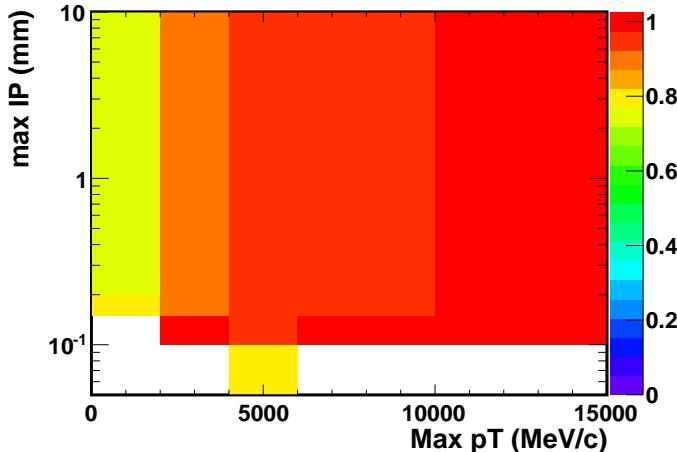


Figure 66: Trigger efficiency map as a function of the max p_T and maxIP of the muons from $B \rightarrow J/\psi X$ events.

1500 To estimate the trigger efficiency for the signal, we have computed an efficiency map
 1501 (see Fig. 66) as a function of the largest p_T and largest IP of the muons from the $B^+ \rightarrow$
 1502 $J/\psi K^+$ detached selection.

Several variables have been tested and we found that these ones give the smallest difference between the TISTOS method and the true efficiency. The trigger decision used to compute this map are listed in Table 40. For the TIS events, all physics lines are used,

see Tab. 41. We then apply this efficiency map to the muon spectrum of the $B_s^0 \rightarrow \mu^+ \mu^-$ MC sample. The estimated trigger efficiency for $B_s^0 \rightarrow \mu^+ \mu^-$ events is:

$$\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{\text{TRIG|SEL}} = (91.4 \pm 0.4_{\text{stat}} \pm 3.9_{\text{syst}})\%, \quad (38)$$

1503 The systematic error is the combination of two errors: one associated to the TISTOS
 1504 method applied to $B_s^0 \rightarrow \mu^+ \mu^-$ (3%) and the second one due to the 2.5% of $B_s^0 \rightarrow \mu^+ \mu^-$
 1505 events that are triggered not using the muon triggers.

The ratio of trigger efficiencies between the signal and the $B^+ \rightarrow J/\psi K^+$ and the $B_s^0 \rightarrow J/\psi \phi$ normalization channels is then computed as the ratio of the two efficiencies:

$$\frac{\epsilon_{B^+ \rightarrow J/\psi K^+}^{\text{TRIG|SEL}}}{\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{\text{TRIG|SEL}}} = \frac{\epsilon_{B_s^0 \rightarrow J/\psi(\mu^+ \mu^-) \phi(KK)}^{\text{TRIG|SEL}}}{\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{\text{TRIG|SEL}}} = (95.4 \pm 0.4_{\text{stat}} \pm 1_{\text{syst}})\%, \quad (39)$$

1506 where the systematic error has been evaluated from MC. It is small because the bias due
 1507 to the TISTOS method is the same for the two channels and cancels in the ratio.

Table 40: Trigger decisions used to compute the trigger efficiency map discussed in the text.

L0Muon, L0Dimuon
Hlt1SingleMuonNoIP, Hlt1TrackAllL0, Hlt1TrackMuon,
Hlt1DiMuonHighMass, Hlt1DiMuonLowMass, Hlt1SingleMuonHighPT
Hlt2DiMuonB, Hlt2DiMuon, Hlt2DiMuonDetached,
Hlt2DiMuonDetachedHeavy, Hlt2DiMuonDetachedJPsi,
Hlt2SingleMuonHighPT, Hlt2SingleMuon, Hlt2TopoMu2BodyBBDT,
Hlt2TopoMu3BodyBBDT, Hlt2TopoMu4BodyBBDT

Table 41: Trigger decisions used to compute the TIS efficiency.

L0	L0.*Decision
HLT1	Routing bit 46 (Hlt1Physics) 'Hlt1(?!ODIN)(?!L0)(?!Lumi)(?!Tell1)(?!MB)(?!NZS)(?!Velo)(?!BeamGas)(?!Incident).*Decision'
HLT2	Routing bit 77 (Hlt2Physics) 'Hlt2(?Forward)(?!DebugEvent)(?!Express)(?!Transparent)(?!PassThrough).*Decision'

1508 10.5.3 Global event cuts

1509 The effect of GECs has been evaluated repeating the efficiency evaluation using only the
 1510 TIS L0DiMuon, as it is the only L0 line with a loose SPD multiplicity cut. The absolute

efficiency values for J/ψ and $B_s^0 \rightarrow \mu^+ \mu^-$ go down by about 4% but the ratio remains constant within 0.1%. It can hence be concluded that the GEC do not introduce an additional uncertainty.

10.6 Number of candidates

10.6.1 Normalization to $B^+ \rightarrow J/\psi K^+$

Fig. 67 shows the invariant mass distribution of the events passing the $B^+ \rightarrow J/\psi K^+$ selection and the fit to the signal and background. The signal distribution is modeled with a double Gaussian function, while the background is modeled with two functions: an exponential for the combinatorial background, and for the physical background a Gaussian on the right. This one is associated to $B^+ \rightarrow J/\psi \pi^+$ events. The number of signal events after the selection and a BDTS cut of 0.05 is: $N(B^+ \rightarrow J/\psi K^+) = 340\,129 \pm 640$. We have assigned a systematic error of 1.3% due to the differences between the result obtained with the fit and the number of candidates obtained after background subtraction. This number of signal events contains 0.1% duplicated candidates.

10.6.2 Normalization to $B_s^0 \rightarrow J/\psi (\mu^+ \mu^-) \phi(KK)$

Fig. 68 shows the invariant mass distribution of the events passing the $B_s^0 \rightarrow J/\psi (\mu^+ \mu^-) \phi(KK)$ selection and the fit to the signal and background. The signal distribution is modeled with a double Gaussian function, while the combinatorial background is modeled with an exponential. The number of signal events after the selection and a BDTS cut of 0.05 is: $N(B^+ \rightarrow J/\psi K^+) = 19\,035 \pm 139$. The fit result has been compared to the number of candidates obtained after background subtraction and a 0.4% systematic

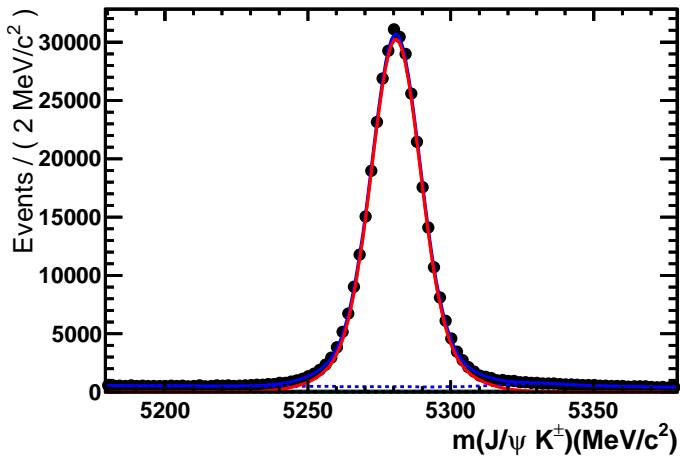


Figure 67: Invariant mass distribution of the $B^+ \rightarrow J/\psi K^+$ candidates after the stripping selection.

1532 uncertainty is added to account for the differences. The fraction of duplicated candidates
1533 in this signal is 0.2%.

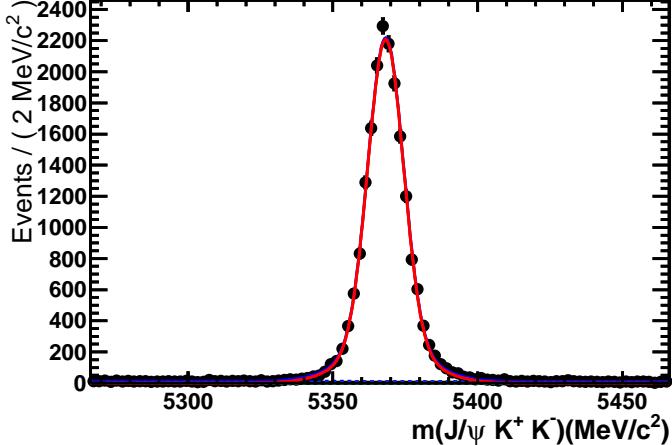


Figure 68: Invariant mass distribution of the $B_s^0 \rightarrow J/\psi \phi$ candidates after the selection.

1534 10.6.3 Normalization to $B^0 \rightarrow K^+\pi^-$ or $B_{(s)}^0 \rightarrow h^+h'^-$ TIS

The $\alpha_{B^0 \rightarrow K^+\pi^-}$ factor can be written in the following way:

$$\alpha_{B^0 \rightarrow K^+\pi^-} = \mathcal{B}(B^0 \rightarrow K^+\pi^-) \times \frac{f_d}{f_s} \frac{\epsilon_{hh}^{REC}}{\epsilon_s^{REC}} \frac{\epsilon_{hh}^{SEL/REC}}{\epsilon_s^{SEL/REC}} \frac{\epsilon_{hh}^{TRIG/SEL}}{\epsilon_s^{TRIG/SEL}} \frac{1}{f_{B^0 \rightarrow K^+\pi^-} N_{hh}} \quad (40)$$

1535 where N_{hh} is the number of $B_{(s)}^0 \rightarrow h^+h'^-$ events and $f_{B^0 \rightarrow K^+\pi^-}$ is the fraction of the
1536 $B^0 \rightarrow K^+\pi^-$ in the inclusive sample. We assume that the reconstruction, selection,
1537 and trigger efficiencies are the same for all the $B_{(s)}^0 \rightarrow h^+h'^-$ modes. This fraction was
1538 measured in the previous analysis note to be: $f_{B^0 \rightarrow K^+\pi^-} = 0.605 \pm 0.027$.

The $\epsilon_{hh}^{TRIG/SEL}$ can be written as:

$$\epsilon_{hh}^{TRIG/SEL} = \epsilon_{hh}^{TIS/SEL} \frac{N_{hh}}{N_{hh}^{TIS}} \quad (41)$$

Therefore, replacing Eq. 41 in Eq. 40, N_{hh} cancels out. This means that the normalization to a given channel, when the trigger efficiency is measured in the same channel, becomes equivalent to a normalization to TIS events:

$$\alpha = \mathcal{B}(B^0 \rightarrow K^+\pi^-) \times \frac{f_N}{f_{sig}} \frac{\epsilon_{hh}^{REC}}{\epsilon_s^{REC}} \frac{\epsilon_{hh}^{SEL/REC}}{\epsilon_s^{SEL/REC}} \frac{\epsilon_{hh}^{TIS/SEL}}{\epsilon_s^{TRIG/SEL}} \frac{1}{f_{B^0 \rightarrow K^+\pi^-} N_{hh}^{TIS}} \quad (42)$$

$\epsilon^{TIS}(L0 \times HLT1)$ was computed in Sec. 10.5.1. We estimate from the MC the HLT2 efficiency on L0 and HLT1 $B_{(s)}^0 \rightarrow h^+h^-$ TIS events to be $\epsilon^{HLT2} = 83.6\%$, this gives: $\epsilon_{hh}^{TIS/SEL} = (4.4 \pm 0.3)\%$. The ratio of trigger efficiencies is:

$$\frac{\epsilon_{hh}^{TIS/SEL}}{\epsilon_{B_s^0 \rightarrow \mu^+\mu^-}^{TRIG|SEL}} = (5.13 \pm 0.09 \pm 0.4)\% \quad (43)$$

1539 The total number of $B_{(s)}^0 \rightarrow h^+h^-$ has been computed when calibrating the BDT, and
 1540 found to be: $N_{hh} = 16734 \pm 757$.

1541 10.7 Normalization factor

1542 The normalization factor α_{cal} for the different control channels and the factors that enters
 1543 in its calculation, that has been presented along this Section, are listed in Table 42. To
 1544 obtain α_{cal} we have used: $f_d/f_s = 3.75 \pm 0.29$ [56].

Table 42: Summary of the factors and their uncertainty entering in the normalization for the three normalization channels considered.

\mathcal{B}	$\frac{\epsilon_{cal}^{REC}\epsilon_{cal}^{SEL REC}}{\epsilon_{sig}^{REC}\epsilon_{sig}^{SEL REC}}$ ($\times 10^{-5}$)	$\frac{\epsilon_{cal}^{TRIG SEL}}{\epsilon_{sig}^{TRIG SEL}}$	N_{cal}	$\alpha_{B_d \rightarrow \mu^+\mu^-}^{cal}$ ($\times 10^{-11}$)	$\alpha_{B_s \rightarrow \mu^+\mu^-}^{cal}$ ($\times 10^{-10}$)
$B^+ \rightarrow J/\psi K^+$	6.01 ± 0.21	0.502 ± 0.013	0.954 ± 0.022	$340\,129 \pm 4468$	8.464 ± 0.433
$B_s^0 \rightarrow J/\psi \phi$	3.4 ± 0.9	0.245 ± 0.011	0.954 ± 0.022	$19\,035 \pm 158$	11.13 ± 3.124
$B^0 \rightarrow K^+\pi^-$	1.94 ± 0.06	0.857 ± 0.028	0.0469 ± 0.0034	$10\,124 \pm 916$	7.709 ± 0.957

1545 A weighted average taking the tracking and trigger uncertainties to be correlated
 1546 between the two J/ψ normalization channels, and the uncertainty on f_d/f_s to be correlated
 1547 between the $B^+ \rightarrow J/\psi K^+$ and $B^0 \rightarrow K^+\pi^-$, gives:

$$\alpha_{B_s^0 \rightarrow \mu^+\mu^-} = (3.19 \pm 0.28) \times 10^{-10}, \quad (44)$$

$$\alpha_{B^0 \rightarrow \mu^+\mu^-} = (8.38 \pm 0.39) \times 10^{-11}, \quad (45)$$

1548 which are the normalization factors used in the computation of the limits.

1549 11 Results: upper limits

1550 The expected number of combinatorial background events, peaking background events and
1551 signal events assuming the SM branching ratios, together with the number of observed
1552 events, are shown in Table 45 for $B_s^0 \rightarrow \mu^+ \mu^-$ analysis and in Table 46 for $B^0 \rightarrow \mu^+ \mu^-$
1553 analysis for each of the 72 bins (8 bins in BDT and 9 bins in invariant mass) that de-
1554 fine the signal regions. The uncertainties in the signal and background likelihoods and
1555 normalization factors are used to compute the uncertainties in the background and signal
1556 predictions.

1557 The distribution of events in the invariant mass vs BDT plane can be seen in Fig. 69.
1558 The distribution of the invariant mass in the eight BDT bins is shown in Fig. 70 for
1559 $B_s^0 \rightarrow \mu^+ \mu^-$ candidates and in Fig. 71 for $B^0 \rightarrow \mu^+ \mu^-$ candidates. The same distributions
1560 obtained by integrating the five most significant BDT bins and with equally spaced mass
1561 bins are shown in Fig. 72. In the limit computation the cross-feed of $B_s^0 \rightarrow \mu^+ \mu^-$ ($B^0 \rightarrow$
1562 $\mu^+ \mu^-$) events in the B^0 (B_s^0) mass window has been taken into account assuming the SM
1563 rates.

1564 The compatibility of the distribution of events in the invariant mass *vs* BDT plane with
1565 a given branching ratio hypothesis is evaluated using the CL_s method [25]. This provides
1566 two estimators: CL_s is a measure of the compatibility of the observed distribution with the
1567 signal hypothesis, while CL_b is a measure of the compatibility with the background-only
1568 hypothesis.

1569 For the $B_s^0 \rightarrow \mu^+ \mu^-$ decay, the distribution of expected CL_s values is shown as dashed
1570 (black) lines in Fig. 73 under the hypothesis to observe background-only (left) or a combi-
1571 nation of background-plus-SM events (right). The green shaded areas cover the region of
1572 $\pm 1\sigma$ of compatible observations. The observed CL_s as a function of the assumed branch-
1573 ing ratio is shown as a dotted (blue) line. The expected and observed limits at 90 % and
1574 95 % C.L. are listed in Table 43, together with the observed CL_b . In the same Table
1575 the results obtained with the 0.37 fb^{-1} sample are shown as reference, together with the
1576 projections extrapolated from the 0.37 fb^{-1} analysis for the 1 fb^{-1} data sample. For the
1577 $B_s^0 \rightarrow \mu^+ \mu^-$ search window, the comparison of the distributions of observed events and
1578 expected background events results in a p-value ($1 - CL_b$) of 18%. The probability that the
1579 observed events are compatible with the sum of expected background events and signal
1580 events according to the SM rate (CL_s) is $\sim 17\%$.

1581 For the $B^0 \rightarrow \mu^+ \mu^-$ decay, the distribution of expected (under the hypothesis to
1582 observe background-only events) and observed CL_s is shown in Fig. 74; the limits are
1583 listed in Table 44. For B^0 , the comparison between data and expected background results
1584 in a p-value of 60%.

1585 The performances of the present analysis can be compared with the projections from
1586 the 0.37 fb^{-1} dataset. The expected limits (computed in background only and background
1587 plus SM hypotheses) as a function of the integrated luminosity are shown in Fig. 75; the
1588 BR value for which a 3σ evidence would have been expected is shown instead in Fig. ??).

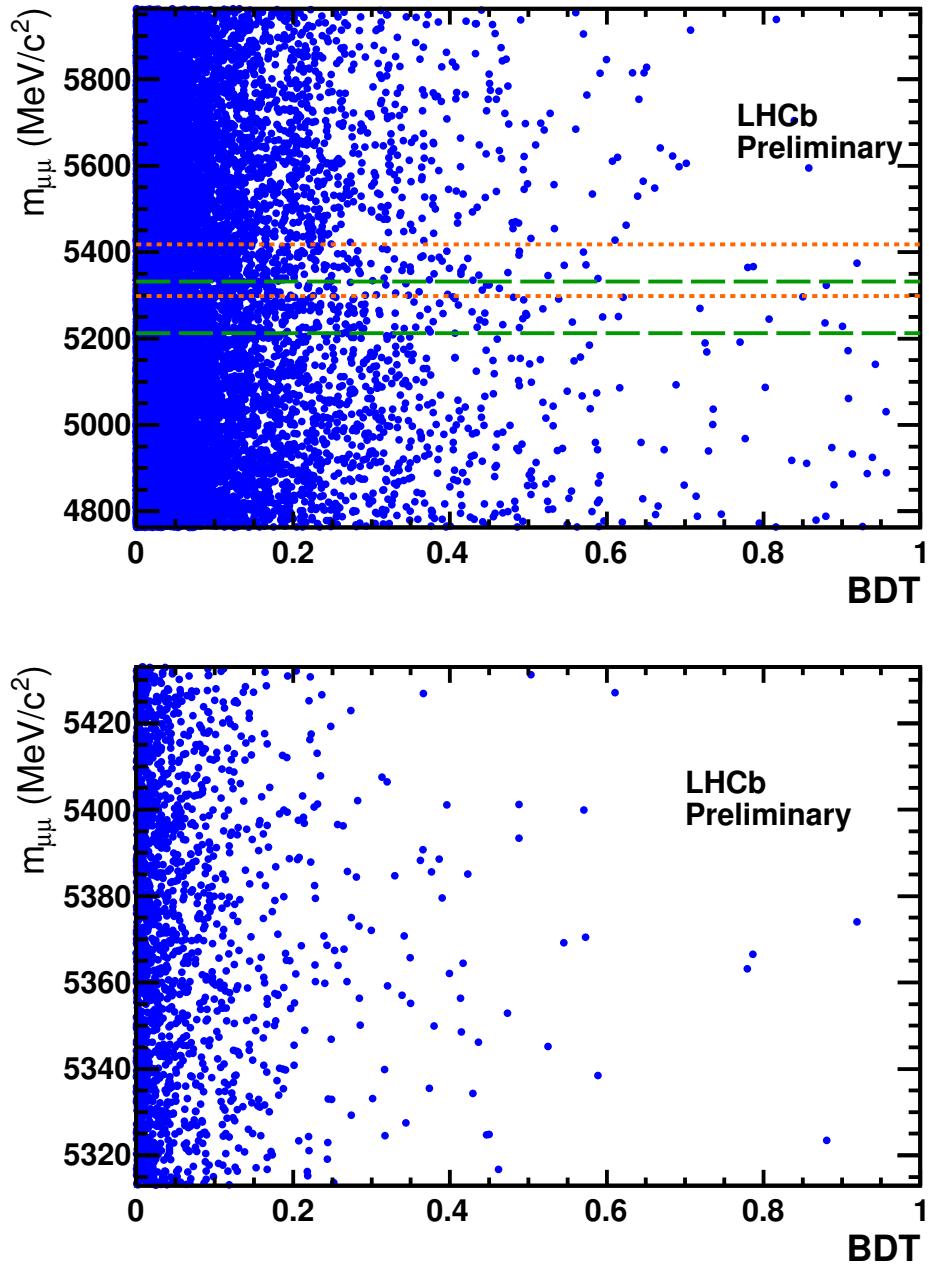


Figure 69: Unblinded data: 2D plot of mass versus BDT. Top: orange short-dashed (green long-dashed) lines indicate the $\pm 60 \text{ MeV}/c^2$ search window around the B_s^0 (B^0). Bottom: zoom in the B_s^0 search window.

1589 From the comparison of the expected limits in the background-only hypothesis evaluated
1590 for 0.37 fb^{-1} and for 1 fb^{-1} we see that they scale as the luminosity: this behaviour was
1591 expected only in absence of background as the normalization factors scale as $\sim 1/L$ This

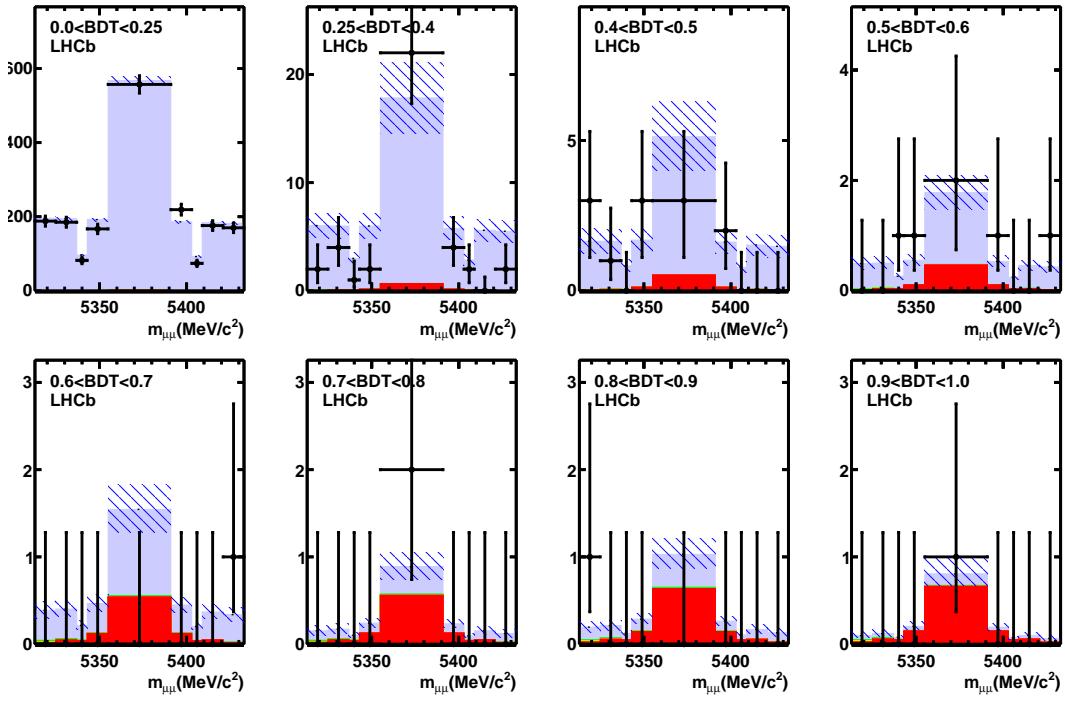


Figure 70: $B_s^0 \rightarrow \mu^+ \mu^-$: distribution of selected dimuon events in the invariant mass plane for the eight BDT bins. The black dots are data, the light blue histogram shows the contribution of the combinatorial background, the green histogram shows the contribution of the $B_{(s)}^0 \rightarrow h^+ h^-$ background and the red filled histogram the contribution of $B_s^0 \rightarrow \mu^+ \mu^-$ signal events according to the SM rate. The hatched area depicts the uncertainty on the sum of the expected contributions.

Table 43: Expected and observed limits for the $B_s^0 \rightarrow \mu^+ \mu^-$ channel, for 1 fb^{-1} and 0.37 fb^{-1} datasets. The projections from the 0.37 fb^{-1} are also shown.

$B_s^0 \rightarrow \mu^+ \mu^-$	at 90% CL	at 95% CL	CL_b
expected limits: bkg-only, 1 fb^{-1}	2.8×10^{-9}	3.4×10^{-9}	—
expected limits: bkg-only, 0.37 fb^{-1}	8×10^{-9}	10×10^{-9}	—
projections from 0.37 fb^{-1}	—	$(5 \pm 2) \times 10^{-9}$	—
expected limits: bkg+SM, 1 fb^{-1}	6.3×10^{-9}	7.3×10^{-9}	—
expected limits: bkg+SM, 0.37 fb^{-1}	11×10^{-9}	14×10^{-9}	—
projections from 0.37 fb^{-1}	—	$(8.5 \pm 2) \times 10^{-9}$	—
observed limits, 1 fb^{-1}	3.8×10^{-9}	4.5×10^{-9}	0.82
observed limits, 0.37 fb^{-1}	13×10^{-9}	16×10^{-9}	0.95

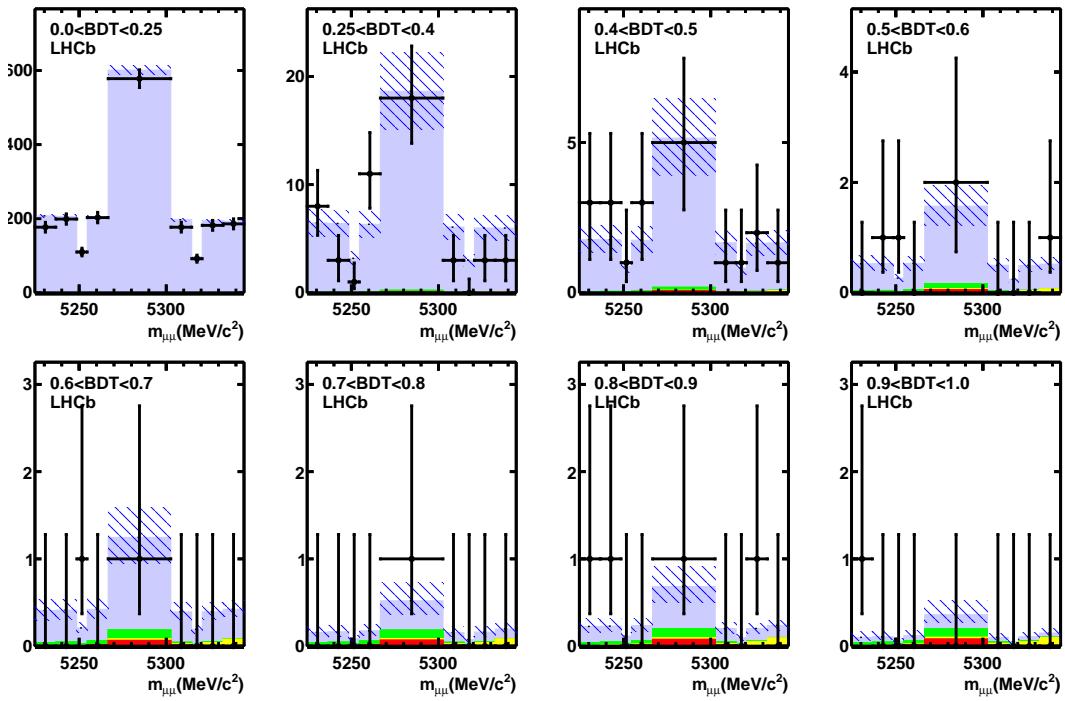


Figure 71: $B^0 \rightarrow \mu^+ \mu^-$: distribution of selected dimuon events in the invariant mass plane for the eight BDT bins. The black dots are data, the light blue histogram shows the contribution of the combinatorial background, the green histogram shows the contribution of the $B_{(s)}^0 \rightarrow h^+ h^-$ background and the yellow histogram shows the cross-feed of $B_s^0 \rightarrow \mu^+ \mu^-$ events in the B^0 mass window assuming the SM rate. The hatched area depicts the uncertainty on the sum of the expected contributions.

Table 44: Expected and observed limits for the $B^0 \rightarrow \mu^+ \mu^-$ channel for 1 fb^{-1} and 0.37 fb^{-1} datasets.

$B^0 \rightarrow \mu^+ \mu^-$	at 90% CL	at 95% CL	CL _b
expected limits: bkg-only, 1 fb^{-1}	0.91×10^{-9}	1.1×10^{-9}	—
expected limits: bkg-only, 0.37 fb^{-1}	2.5×10^{-9}	3.2×10^{-9}	—
observed limits, 1 fb^{-1}	0.81×10^{-9}	1.0×10^{-9}	0.40
observed limits, 0.37 fb^{-1}	3.0×10^{-9}	3.6×10^{-9}	0.68

1592 result has been obtained thanks to the improvement of the analysis.

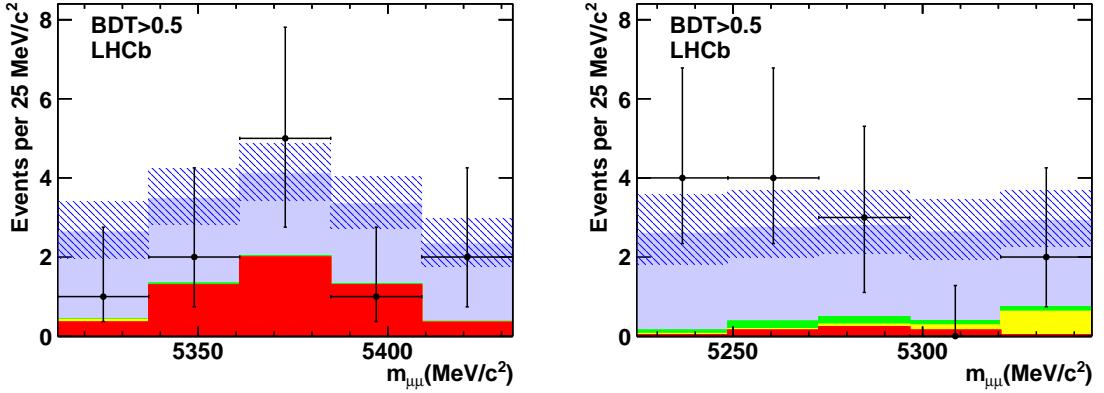


Figure 72: $B_s^0 \rightarrow \mu^+ \mu^-$ (left) and $B^0 \rightarrow \mu^+ \mu^-$ (right) distribution of selected dimuon events in the invariant mass plane for events with $\text{BDT} > 0.5$. The black dots are data, the light blue histogram shows the contribution of the combinatorial background, the green histogram shows the contribution of the $B_{(s)}^0 \rightarrow h^+ h^-$ background and the red filled histogram the contribution of signal events according to the SM rate. The hatched area depicts the uncertainty on the sum of the expected contributions; differently from what used for the limit evaluation, equally spaced mass bins are shown.

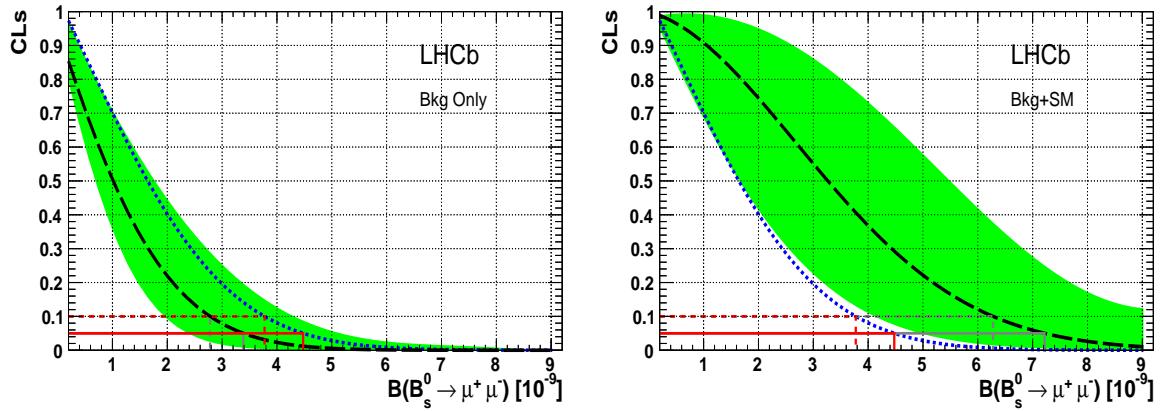


Figure 73: $B_s^0 \rightarrow \mu^+ \mu^-$: expected CL_s (dashed black line) under the hypothesis to observe background-only (left) and a combination of background-plus-signal events according to the SM rate (right), with green area covering the region of $\pm 1\sigma$ of compatible observations; in both plots the observed CL_s is given by the blue dotted line; the expected (observed) upper limits at 90% and 95% C.L. are also shown as dashed and solid grey (red) lines.

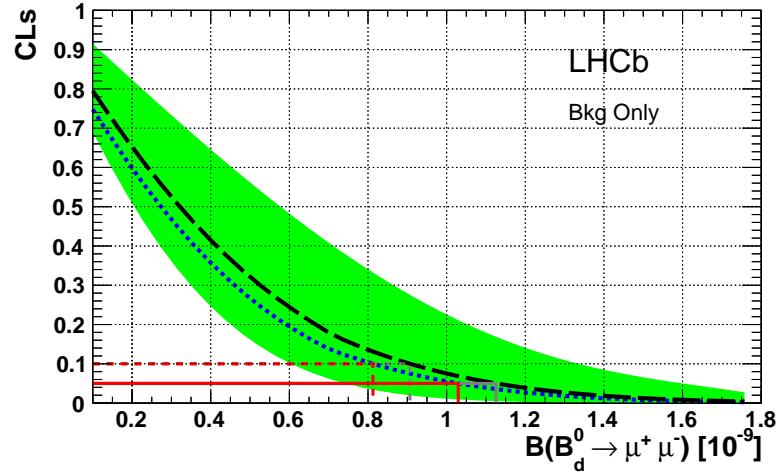


Figure 74: $B^0 \rightarrow \mu^+ \mu^-$: expected CL_s (dashed black line) under the hypothesis to observe background-only, with green area covering the region of $\pm 1\sigma$ of compatible observations; the observed CL_s is given by the blue dotted line; the expected (observed) upper limits at 90% and 95% C.L. are also shown as dashed and solid grey (red) lines.

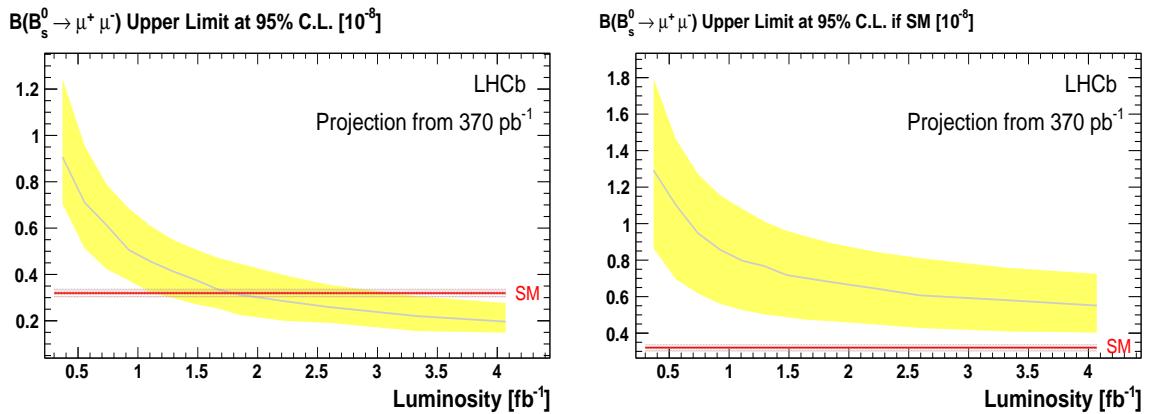


Figure 75: $B_s^0 \rightarrow \mu^+ \mu^-$: projections from the 0.37 fb^{-1} for the $B_s^0 \rightarrow \mu^+ \mu^-$ expected limit in the background-only (left) and background-plus-SM (right) hypotheses.

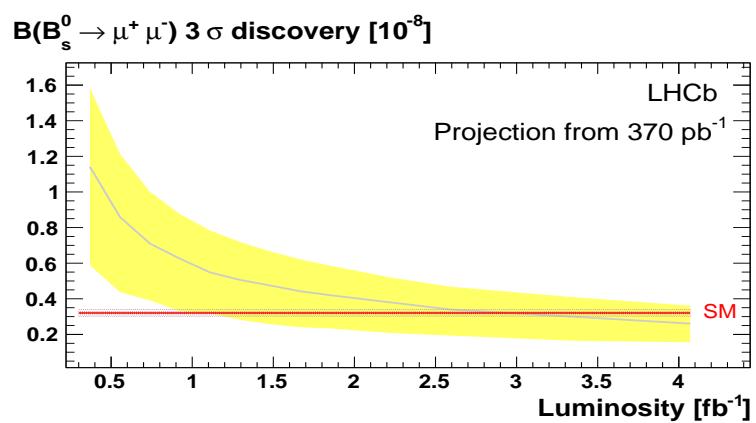


Figure 76: $B_s^0 \rightarrow \mu^+ \mu^-$: projections from the 0.37 fb^{-1} for the $B_s^0 \rightarrow \mu^+ \mu^-$ 3σ evidence.

Table 45: $B_s^0 \rightarrow \mu^+ \mu^-$: Expected combinatorial background events, expected peaking ($B_{(s)}^0 \rightarrow h^+ h^-$) background events and expected signal events assuming the SM branching fraction prediction.

Invariant mass [MeV/ c^2]		BDT							
		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
5312 – 5324	Exp. comb. bkg	195.7 ^{+4.1} _{-4.1}	6.0 ^{+1.2} _{-1.1}	1.61 ^{+0.41} _{-0.40}	0.45 ^{+0.12} _{-0.11}	0.345 ^{+0.105} _{-0.094}	0.110 ^{+0.063} _{-0.049}	0.147 ^{+0.065} _{-0.056}	0.050 ^{+0.048} _{-0.027}
	Exp. peak. bkg	0.035 ^{+0.016} _{-0.012}	0.0177 ^{+0.0058} _{-0.0043}	0.0139 ^{+0.0044} _{-0.0033}	0.0125 ^{+0.0039} _{-0.0028}	0.0140 ^{+0.0043} _{-0.0032}	0.0132 ^{+0.0041} _{-0.0030}	0.0137 ^{+0.0042} _{-0.0032}	0.0132 ^{+0.0044} _{-0.0033}
	Exp. signal	0.050 ^{+0.014} _{-0.015}	0.0240 ^{+0.0046} _{-0.0040}	0.0191 ^{+0.0033} _{-0.0029}	0.0170 ^{+0.0025} _{-0.0021}	0.0197 ^{+0.0028} _{-0.0024}	0.0204 ^{+0.0027} _{-0.0024}	0.0233 ^{+0.0033} _{-0.0028}	0.0241 ^{+0.0049} _{-0.0043}
	Observed	188	2	3	0	0	0	1	0
5324 – 5336	Exp. comb. bkg	194.2 ^{+4.1} _{-4.0}	5.9 ^{+1.1} _{-1.1}	1.59 ^{+0.41} _{-0.40}	0.45 ^{+0.11} _{-0.11}	0.341 ^{+0.102} _{-0.091}	0.109 ^{+0.060} _{-0.048}	0.142 ^{+0.062} _{-0.055}	0.049 ^{+0.046} _{-0.026}
	Exp. peak. bkg	0.0237 ^{+0.0093} _{-0.0083}	0.0120 ^{+0.0028} _{-0.0024}	0.0094 ^{+0.0020} _{-0.0018}	0.0085 ^{+0.0016} _{-0.0014}	0.0095 ^{+0.0018} _{-0.0016}	0.0090 ^{+0.0016} _{-0.0015}	0.0094 ^{+0.0018} _{-0.0016}	0.0090 ^{+0.0022} _{-0.0019}
	Exp. signal	0.120 ^{+0.033} _{-0.035}	0.0579 ^{+0.0103} _{-0.0092}	0.0459 ^{+0.0073} _{-0.0065}	0.0409 ^{+0.0053} _{-0.0047}	0.0474 ^{+0.0060} _{-0.0053}	0.0492 ^{+0.0058} _{-0.0052}	0.0560 ^{+0.0070} _{-0.0062}	0.0582 ^{+0.0111} _{-0.0100}
	Observed	185	4	1	0	0	0	0	0
5336 – 5342	Exp. comb. bkg	96.5 ^{+2.0} _{-2.0}	2.94 ^{+0.56} _{-0.56}	0.79 ^{+0.20} _{-0.20}	0.223 ^{+0.055} _{-0.053}	0.169 ^{+0.051} _{-0.044}	0.054 ^{+0.029} _{-0.024}	0.069 ^{+0.030} _{-0.027}	0.024 ^{+0.022} _{-0.013}
	Exp. peak. bkg	0.0092 ^{+0.0038} _{-0.0033}	0.0046 ^{+0.0012} _{-0.0010}	0.00365 ^{+0.00090} _{-0.00077}	0.00328 ^{+0.00075} _{-0.00063}	0.00368 ^{+0.00083} _{-0.00070}	0.00347 ^{+0.00077} _{-0.00064}	0.00361 ^{+0.00081} _{-0.00068}	0.00347 ^{+0.00095} _{-0.00082}
	Exp. signal	0.103 ^{+0.028} _{-0.030}	0.0494 ^{+0.0083} _{-0.0076}	0.0392 ^{+0.0059} _{-0.0053}	0.0349 ^{+0.0042} _{-0.0038}	0.0404 ^{+0.0047} _{-0.0042}	0.0419 ^{+0.0045} _{-0.0040}	0.0478 ^{+0.0054} _{-0.0048}	0.0496 ^{+0.0090} _{-0.0083}
	Observed	82	1	0	1	0	0	0	0
5342 – 5354	Exp. comb. bkg	191.8 ^{+4.0} _{-4.0}	5.8 ^{+1.1} _{-1.1}	1.56 ^{+0.40} _{-0.39}	0.44 ^{+0.11} _{-0.10}	0.336 ^{+0.099} _{-0.088}	0.108 ^{+0.056} _{-0.047}	0.135 ^{+0.058} _{-0.053}	0.047 ^{+0.043} _{-0.025}
	Exp. peak. bkg	0.0136 ^{+0.0080} _{-0.0055}	0.0070 ^{+0.0032} _{-0.0022}	0.0055 ^{+0.0024} _{-0.0017}	0.0050 ^{+0.0021} _{-0.0015}	0.0056 ^{+0.0023} _{-0.0017}	0.0054 ^{+0.0021} _{-0.0017}	0.0055 ^{+0.0022} _{-0.0017}	0.0052 ^{+0.0024} _{-0.0017}
	Exp. signal	0.312 ^{+0.084} _{-0.089}	0.150 ^{+0.025} _{-0.023}	0.119 ^{+0.018} _{-0.016}	0.106 ^{+0.012} _{-0.011}	0.122 ^{+0.014} _{-0.012}	0.127 ^{+0.013} _{-0.012}	0.145 ^{+0.016} _{-0.014}	0.150 ^{+0.027} _{-0.025}
	Observed	167	2	3	1	0	0	0	0
5354 – 5390	Exp. comb. bkg	566 ⁺¹² ₋₁₂	17.2 ^{+3.3} _{-3.3}	4.6 ^{+1.2} _{-1.1}	1.30 ^{+0.31} _{-0.30}	0.99 ^{+0.28} _{-0.26}	0.32 ^{+0.15} _{-0.14}	0.38 ^{+0.16} _{-0.15}	0.133 ^{+0.121} _{-0.073}
	Exp. peak. bkg	0.021 ^{+0.028} _{-0.011}	0.0137 ^{+0.0093} _{-0.0079}	0.0113 ^{+0.0066} _{-0.0067}	0.0110 ^{+0.0049} _{-0.0068}	0.0125 ^{+0.0054} _{-0.0077}	0.0121 ^{+0.0047} _{-0.0076}	0.0124 ^{+0.0052} _{-0.0077}	0.0098 ^{+0.0075} _{-0.0055}
	Exp. signal	1.37 ^{+0.37} _{-0.39}	0.66 ^{+0.11} _{-0.10}	0.523 ^{+0.077} _{-0.070}	0.466 ^{+0.055} _{-0.049}	0.539 ^{+0.061} _{-0.055}	0.559 ^{+0.059} _{-0.052}	0.638 ^{+0.071} _{-0.062}	0.66 ^{+0.12} _{-0.11}
	Observed	557	22	3	2	0	2	0	1
5390 – 5402	Exp. comb. bkg	185.8 ^{+3.8} _{-3.8}	5.6 ^{+1.1} _{-1.1}	1.50 ^{+0.37} _{-0.37}	0.424 ^{+0.098} _{-0.097}	0.323 ^{+0.090} _{-0.081}	0.103 ^{+0.048} _{-0.043}	0.117 ^{+0.055} _{-0.049}	0.042 ^{+0.037} _{-0.024}
	Exp. peak. bkg	0.0034 ^{+0.0091} _{-0.0022}	0.0035 ^{+0.0024} _{-0.0028}	0.0029 ^{+0.0017} _{-0.0023}	0.0028 ^{+0.0013} _{-0.0023}	0.0032 ^{+0.0014} _{-0.0026}	0.0031 ^{+0.0012} _{-0.0026}	0.0032 ^{+0.0013} _{-0.0026}	0.0025 ^{+0.0019} _{-0.0020}
	Exp. signal	0.311 ^{+0.084} _{-0.089}	0.150 ^{+0.025} _{-0.023}	0.119 ^{+0.018} _{-0.016}	0.106 ^{+0.012} _{-0.011}	0.122 ^{+0.014} _{-0.012}	0.127 ^{+0.013} _{-0.012}	0.145 ^{+0.016} _{-0.014}	0.150 ^{+0.027} _{-0.025}
	Observed	219	4	2	1	0	0	0	0
5402 – 5408	Exp. comb. bkg	92.3 ^{+1.9} _{-1.9}	2.78 ^{+0.53} _{-0.53}	0.74 ^{+0.18} _{-0.18}	0.210 ^{+0.048} _{-0.048}	0.160 ^{+0.045} _{-0.039}	0.051 ^{+0.023} _{-0.021}	0.057 ^{+0.028} _{-0.024}	0.021 ^{+0.018} _{-0.012}
	Exp. peak. bkg	0.0015 ^{+0.0041} _{-0.0010}	0.0016 ^{+0.0011} _{-0.0013}	0.00129 ^{+0.00076} _{-0.00108}	0.00126 ^{+0.00056} _{-0.00107}	0.00142 ^{+0.00062} _{-0.00121}	0.00138 ^{+0.00054} _{-0.00118}	0.00141 ^{+0.00059} _{-0.00120}	0.00112 ^{+0.00085} _{-0.00092}
	Exp. signal	0.103 ^{+0.028} _{-0.030}	0.0494 ^{+0.0084} _{-0.0075}	0.0391 ^{+0.0059} _{-0.0053}	0.0349 ^{+0.0047} _{-0.0037}	0.0404 ^{+0.0042} _{-0.0042}	0.0419 ^{+0.0045} _{-0.0040}	0.0478 ^{+0.0054} _{-0.0048}	0.0496 ^{+0.0091} _{-0.0082}
	Observed	74	2	0	0	0	0	0	0
5408 – 5420	Exp. comb. bkg	183.6 ^{+3.7} _{-3.7}	5.5 ^{+1.0} _{-1.0}	1.48 ^{+0.37} _{-0.36}	0.418 ^{+0.095} _{-0.094}	0.318 ^{+0.089} _{-0.078}	0.101 ^{+0.046} _{-0.042}	0.110 ^{+0.056} _{-0.046}	0.040 ^{+0.035} _{-0.023}
	Exp. peak. bkg	0.0029 ^{+0.0079} _{-0.0021}	0.0030 ^{+0.0020} _{-0.0026}	0.0025 ^{+0.0015} _{-0.0021}	0.0024 ^{+0.0011} _{-0.0021}	0.0027 ^{+0.0012} _{-0.0024}	0.0027 ^{+0.0010} _{-0.0023}	0.0027 ^{+0.0011} _{-0.0024}	0.0022 ^{+0.0016} _{-0.0018}
	Exp. signal	0.121 ^{+0.033} _{-0.035}	0.0580 ^{+0.0103} _{-0.0092}	0.0459 ^{+0.0073} _{-0.0065}	0.0409 ^{+0.0053} _{-0.0047}	0.0474 ^{+0.0060} _{-0.0054}	0.0491 ^{+0.0058} _{-0.0052}	0.0560 ^{+0.0069} _{-0.0062}	0.0581 ^{+0.0111} _{-0.0099}
	Observed	176	0	0	0	0	0	0	0
5420 – 5432	Exp. comb. bkg	182.1 ^{+3.7} _{-3.7}	5.5 ^{+1.0} _{-1.0}	1.46 ^{+0.36} _{-0.36}	0.413 ^{+0.093} _{-0.093}	0.317 ^{+0.087} _{-0.077}	0.100 ^{+0.044} _{-0.042}	0.106 ^{+0.056} _{-0.044}	0.039 ^{+0.033} _{-0.022}
	Exp. peak. bkg	0.0024 ^{+0.0067} _{-0.0018}	0.0026 ^{+0.0017} _{-0.0022}	0.0021 ^{+0.0012} _{-0.0018}	0.00207 ^{+0.00092} _{-0.00182}	0.0023 ^{+0.0010} _{-0.0021}	0.00227 ^{+0.00088} _{-0.00200}	0.00232 ^{+0.00097} _{-0.00204}	0.0018 ^{+0.0014} _{-0.0016}
	Exp. signal	0.048 ^{+0.014} _{-0.014}	0.0231 ^{+0.0047} _{-0.0040}	0.0183 ^{+0.0034} _{-0.0030}	0.0163 ^{+0.0026} _{-0.0023}	0.0189 ^{+0.0030} _{-0.0026}	0.0196 ^{+0.0029} _{-0.0026}	0.0224 ^{+0.0034} _{-0.0030}	0.0231 ^{+0.0050} _{-0.0043}
	Observed	170	2	0	1	1	0	0	0

Table 46: $B^0 \rightarrow \mu^+ \mu^-$: Expected combinatorial background events, expected peaking ($B_{(s)}^0 \rightarrow h^+ h^-$) background events, expected cross-feed events from $B_s^0 \rightarrow \mu^+ \mu^-$ assuming the SM branching fraction and expected $B^0 \rightarrow \mu^+ \mu^-$ signal events assuming the SM branching fraction.

Invariant mass [MeV/ c^2]		BDT							
		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
5224 – 5236	Exp. comb. bkg	207.6 ^{+4.5} _{-4.5}	6.4 ^{+1.3} _{-1.2}	1.74 ^{+0.47} _{-0.45}	0.49 ^{+0.14} _{-0.13}	0.37 ^{+0.13} _{-0.11}	0.121 ^{+0.074} _{-0.059}	0.187 ^{+0.087} _{-0.067}	0.063 ^{+0.061} _{-0.033}
	Exp. peak. bkg	0.084 ^{+0.043} _{-0.031}	0.042 ^{+0.017} _{-0.010}	0.0333 ^{+0.0131} _{-0.0079}	0.0297 ^{+0.0017} _{-0.0065}	0.0334 ^{+0.0130} _{-0.0073}	0.0315 ^{+0.0122} _{-0.0067}	0.0328 ^{+0.0128} _{-0.0070}	0.0317 ^{+0.0129} _{-0.0081}
	Exp. Cross-feed	0.0064 ^{+0.0018} _{-0.0019}	0.00310 ^{+0.00056} _{-0.00050}	0.00246 ^{+0.00040} _{-0.00036}	0.00219 ^{+0.00030} _{-0.00026}	0.00253 ^{+0.00034} _{-0.00030}	0.00263 ^{+0.00033} _{-0.00029}	0.00300 ^{+0.00039} _{-0.00034}	0.00311 ^{+0.00060} _{-0.00055}
	Exp. signal	0.0055 ^{+0.0017} _{-0.0016}	0.00266 ^{+0.00045} _{-0.00047}	0.00211 ^{+0.00039} _{-0.00034}	0.00188 ^{+0.00030} _{-0.00026}	0.00218 ^{+0.00034} _{-0.00030}	0.00226 ^{+0.00034} _{-0.00030}	0.00257 ^{+0.00040} _{-0.00035}	0.00266 ^{+0.00057} _{-0.00050}
Observed		177	8	3	0	0	0	1	1
5236 – 5248	Exp. comb. bkg	205.9 ^{+4.5} _{-4.5}	6.3 ^{+1.2} _{-1.2}	1.72 ^{+0.46} _{-0.44}	0.49 ^{+0.14} _{-0.13}	0.37 ^{+0.12} _{-0.11}	0.119 ^{+0.073} _{-0.058}	0.181 ^{+0.084} _{-0.065}	0.061 ^{+0.059} _{-0.032}
	Exp. peak. bkg	0.096 ^{+0.054} _{-0.036}	0.049 ^{+0.022} _{-0.013}	0.038 ^{+0.017} _{-0.010}	0.0346 ^{+0.0147} _{-0.0091}	0.039 ^{+0.016} _{-0.010}	0.0371 ^{+0.0149} _{-0.0100}	0.038 ^{+0.016} _{-0.010}	0.036 ^{+0.017} _{-0.010}
	Exp. Cross-feed	0.0073 ^{+0.0021} _{-0.0021}	0.00351 ^{+0.00063} _{-0.00057}	0.00279 ^{+0.00045} _{-0.00040}	0.00248 ^{+0.00034} _{-0.00030}	0.00287 ^{+0.00038} _{-0.00033}	0.00298 ^{+0.00037} _{-0.00033}	0.00340 ^{+0.00044} _{-0.00039}	0.00353 ^{+0.00068} _{-0.00062}
	Exp. signal	0.0136 ^{+0.0040} _{-0.0040}	0.0065 ^{+0.0013} _{-0.0011}	0.00519 ^{+0.00092} _{-0.00082}	0.00463 ^{+0.00070} _{-0.00063}	0.00537 ^{+0.00079} _{-0.00072}	0.00556 ^{+0.00079} _{-0.00071}	0.00634 ^{+0.00093} _{-0.00083}	0.0066 ^{+0.0014} _{-0.0012}
Observed		199	3	3	1	0	0	1	0
5248 – 5254	Exp. comb. bkg	102.3 ^{+2.2} _{-2.2}	3.15 ^{+0.61} _{-0.60}	0.85 ^{+0.23} _{-0.22}	0.241 ^{+0.068} _{-0.062}	0.182 ^{+0.061} _{-0.053}	0.059 ^{+0.036} _{-0.029}	0.088 ^{+0.041} _{-0.032}	0.030 ^{+0.029} _{-0.015}
	Exp. peak. bkg	0.051 ^{+0.030} _{-0.019}	0.0259 ^{+0.012} _{-0.0074}	0.0204 ^{+0.0093} _{-0.0058}	0.0186 ^{+0.0089} _{-0.0052}	0.0209 ^{+0.0088} _{-0.0059}	0.0200 ^{+0.0080} _{-0.0057}	0.0206 ^{+0.0086} _{-0.0058}	0.0193 ^{+0.0093} _{-0.0057}
	Exp. Cross-feed	0.0040 ^{+0.0011} _{-0.0012}	0.00195 ^{+0.00035} _{-0.00032}	0.00155 ^{+0.00025} _{-0.00022}	0.00138 ^{+0.00019} _{-0.00016}	0.00159 ^{+0.00021} _{-0.00019}	0.00165 ^{+0.00021} _{-0.00018}	0.00188 ^{+0.00024} _{-0.00021}	0.00196 ^{+0.00038} _{-0.00034}
	Exp. signal	0.0119 ^{+0.0034} _{-0.0035}	0.00573 ^{+0.00109} _{-0.00096}	0.00454 ^{+0.00078} _{-0.00070}	0.00405 ^{+0.00059} _{-0.00053}	0.00470 ^{+0.00066} _{-0.00061}	0.00487 ^{+0.00065} _{-0.00059}	0.00555 ^{+0.00077} _{-0.00069}	0.0057 ^{+0.0012} _{-0.0010}
Observed		110	1	1	1	1	0	0	0
5254 – 5266	Exp. comb. bkg	203.5 ^{+4.4} _{-4.4}	6.3 ^{+1.2} _{-1.2}	1.69 ^{+0.45} _{-0.43}	0.48 ^{+0.13} _{-0.12}	0.36 ^{+0.12} _{-0.10}	0.115 ^{+0.073} _{-0.055}	0.172 ^{+0.079} _{-0.062}	0.058 ^{+0.056} _{-0.030}
	Exp. peak. bkg	0.104 ^{+0.062} _{-0.040}	0.053 ^{+0.025} _{-0.016}	0.042 ^{+0.019} _{-0.012}	0.038 ^{+0.016} _{-0.011}	0.043 ^{+0.018} _{-0.013}	0.041 ^{+0.017} _{-0.012}	0.042 ^{+0.018} _{-0.012}	0.040 ^{+0.019} _{-0.012}
	Exp. Cross-feed	0.0091 ^{+0.0026} _{-0.0026}	0.00439 ^{+0.00079} _{-0.00071}	0.00348 ^{+0.00057} _{-0.00050}	0.00310 ^{+0.00042} _{-0.00037}	0.00359 ^{+0.00047} _{-0.00042}	0.00373 ^{+0.00046} _{-0.00041}	0.00425 ^{+0.00055} _{-0.00048}	0.00441 ^{+0.00085} _{-0.00077}
	Exp. signal	0.037 ^{+0.011} _{-0.011}	0.0177 ^{+0.0033} _{-0.0029}	0.0140 ^{+0.0024} _{-0.0021}	0.0125 ^{+0.0018} _{-0.0016}	0.0145 ^{+0.0020} _{-0.0019}	0.0151 ^{+0.0020} _{-0.0018}	0.0172 ^{+0.0023} _{-0.0021}	0.0178 ^{+0.0035} _{-0.0032}
Observed		203	11	3	0	0	0	0	0
5266 – 5302	Exp. comb. bkg	601 ⁺¹³ ₋₁₃	18.4 ^{+3.6} _{-3.5}	5.0 ^{+1.3} _{-1.3}	1.40 ^{+0.38} _{-0.35}	1.06 ^{+0.34} _{-0.30}	0.33 ^{+0.21} _{-0.15}	0.48 ^{+0.22} _{-0.18}	0.163 ^{+0.158} _{-0.086}
	Exp. peak. bkg	0.256 ^{+0.143} _{-0.096}	0.130 ^{+0.059} _{-0.036}	0.102 ^{+0.045} _{-0.027}	0.092 ^{+0.039} _{-0.024}	0.104 ^{+0.043} _{-0.027}	0.099 ^{+0.040} _{-0.026}	0.102 ^{+0.042} _{-0.027}	0.097 ^{+0.044} _{-0.027}
	Exp. Cross-feed	0.043 ^{+0.012} _{-0.012}	0.0207 ^{+0.0037} _{-0.0033}	0.0164 ^{+0.0026} _{-0.0024}	0.0146 ^{+0.0020} _{-0.0017}	0.0169 ^{+0.0022} _{-0.0020}	0.0175 ^{+0.0021} _{-0.0019}	0.0200 ^{+0.0026} _{-0.0022}	0.0208 ^{+0.0040} _{-0.0036}
	Exp. signal	0.165 ^{+0.048} _{-0.049}	0.080 ^{+0.015} _{-0.013}	0.0631 ^{+0.0106} _{-0.0096}	0.0563 ^{+0.0080} _{-0.0072}	0.0652 ^{+0.0091} _{-0.0083}	0.0676 ^{+0.0089} _{-0.0081}	0.0771 ^{+0.0105} _{-0.0095}	0.080 ^{+0.016} _{-0.014}
Observed		578	18	5	2	1	1	1	0
5302 – 5314	Exp. comb. bkg	197.1 ^{+4.1} _{-4.1}	6.0 ^{+1.2} _{-1.1}	1.62 ^{+0.42} _{-0.41}	0.46 ^{+0.12} _{-0.11}	0.348 ^{+0.108} _{-0.096}	0.111 ^{+0.065} _{-0.050}	0.152 ^{+0.067} _{-0.057}	0.051 ^{+0.049} _{-0.027}
	Exp. peak. bkg	0.048 ^{+0.021} _{-0.017}	0.0242 ^{+0.0073} _{-0.0049}	0.0190 ^{+0.0055} _{-0.0035}	0.0169 ^{+0.0049} _{-0.0032}	0.0190 ^{+0.0055} _{-0.0029}	0.0179 ^{+0.0052} _{-0.0026}	0.0186 ^{+0.0054} _{-0.0028}	0.0181 ^{+0.0056} _{-0.0038}
	Exp. Cross-feed	0.0277 ^{+0.0078} _{-0.0080}	0.0134 ^{+0.0024} _{-0.0021}	0.0106 ^{+0.0017} _{-0.0015}	0.0094 ^{+0.0013} _{-0.0011}	0.0109 ^{+0.0014} _{-0.0013}	0.0113 ^{+0.0014} _{-0.0012}	0.0129 ^{+0.0017} _{-0.0014}	0.0134 ^{+0.0026} _{-0.0023}
	Exp. signal	0.037 ^{+0.011} _{-0.011}	0.0177 ^{+0.0033} _{-0.0029}	0.0140 ^{+0.0024} _{-0.0021}	0.0125 ^{+0.0018} _{-0.0016}	0.0145 ^{+0.0020} _{-0.0018}	0.0151 ^{+0.0020} _{-0.0018}	0.0172 ^{+0.0024} _{-0.0021}	0.0178 ^{+0.0035} _{-0.0032}
Observed		177	3	1	0	0	0	0	0
5314 – 5320	Exp. comb. bkg	97.9 ^{+2.0} _{-2.0}	2.99 ^{+0.57} _{-0.57}	0.81 ^{+0.21} _{-0.20}	0.227 ^{+0.058} _{-0.055}	0.173 ^{+0.053} _{-0.047}	0.055 ^{+0.031} _{-0.025}	0.074 ^{+0.033} _{-0.028}	0.025 ^{+0.024} _{-0.013}
	Exp. peak. bkg	0.0175 ^{+0.0068} _{-0.0059}	0.0088 ^{+0.0021} _{-0.0016}	0.0069 ^{+0.0015} _{-0.0012}	0.00620 ^{+0.00118} _{-0.00094}	0.0070 ^{+0.0013} _{-0.0010}	0.00657 ^{+0.00115} _{-0.00095}	0.00683 ^{+0.00126} _{-0.00100}	0.0066 ^{+0.0016} _{-0.0013}
	Exp. Cross-feed	0.0219 ^{+0.0063} _{-0.0064}	0.0106 ^{+0.0020} _{-0.0017}	0.00884 ^{+0.0014} _{-0.0012}	0.00745 ^{+0.00098} _{-0.00093}	0.00868 ^{+0.0012} _{-0.0011}	0.00900 ^{+0.0012} _{-0.0011}	0.0102 ^{+0.0014} _{-0.0012}	0.0106 ^{+0.0021} _{-0.0019}
	Exp. signal	0.0119 ^{+0.0034} _{-0.0035}	0.00573 ^{+0.00108} _{-0.00096}	0.00454 ^{+0.00078} _{-0.00070}	0.00405 ^{+0.00059} _{-0.00052}	0.00469 ^{+0.00066} _{-0.00061}	0.00487 ^{+0.00066} _{-0.00060}	0.00555 ^{+0.00078} _{-0.00069}	0.0057 ^{+0.0012} _{-0.0010}
Observed		92	0	1	0	0	0	0	0
5320 – 5332	Exp. comb. bkg	194.7 ^{+4.1} _{-4.0}	5.9 ^{+1.1} _{-1.1}	1.60 ^{+0.41} _{-0.40}	0.45 ^{+0.11} _{-0.11}	0.343 ^{+0.102} _{-0.092}	0.110 ^{+0.061} _{-0.048}	0.144 ^{+0.063} _{-0.055}	0.049 ^{+0.047} _{-0.026}
	Exp. peak. bkg	0.0257 ^{+0.0104} _{-0.0087}	0.0130 ^{+0.0032} _{-0.0028}	0.0103 ^{+0.0024} _{-0.0021}	0.0093 ^{+0.0019} _{-0.0019}	0.0104 ^{+0.0021} _{-0.0021}	0.0099 ^{+0.0019} _{-0.0020}	0.0102 ^{+0.0020} _{-0.0021}	0.0097 ^{+0.0025} _{-0.0021}
	Exp. Cross-feed	0.089 ^{+0.025} _{-0.026}	0.0427 ^{+0.0079} _{-0.0070}	0.0339 ^{+0.0056} _{-0.0050}	0.0302 ^{+0.0042} _{-0.0037}	0.0350 ^{+0.0048} _{-0.0042}	0.0363 ^{+0.0046} _{-0.0041}	0.0414 ^{+0.0055} _{-0.0049}	0.0429 ^{+0.0084} _{-0.0075}
	Exp. signal	0.0136 ^{+0.0040} _{-0.0040}	0.0065 ^{+0.0013} _{-0.0011}	0.00519 ^{+0.00093} _{-0.00081}	0.00463 ^{+0.00070} _{-0.00062}	0.00537 ^{+0.00079} _{-0.00072}	0.00557 ^{+0.00079} _{-0.00071}	0.00635 ^{+0.00083} _{-0.00083}	0.0066 ^{+0.0014} _{-0.0012}
Observed		182	3	2	0	0	0	1	0
5332 – 5344	Exp. comb. bkg	193.2 ^{+4.0} _{-4.0}	5.9 ^{+1.1} _{-1.1}	1.58 ^{+0.40} _{-0.39}	0.45 ^{+0.11} _{-0.11}	0.339 ^{+0.101} _{-0.089}	0.109 ^{+0.058} _{-0.048}	0.140 ^{+0.060} _{-0.054}	0.048 ^{+0.045} _{-0.026}
	Exp. peak. bkg	0.0172 ^{+0.0092} _{-0.0064}	0.0089 ^{+0.0034} _{-0.0027}	0.0071 ^{+0.0026} _{-0.0021}	0.0064 ^{+0.0022} _{-0.0019}	0.0071 ^{+0.0025} _{-0.0022}	0.0067 ^{+0.0023} _{-0.0021}	0.0070 ^{+0.0024} _{-0.0020}	0.0067 ^{+0.0026} _{-0.0020}
	Exp. Cross-feed	0.192 ^{+0.052} _{-0.055}	0.093 ^{+0.016} _{-0.014}	0.073 ^{+0.011} _{-0.010}	0.0653 ^{+0.0079} _{-0.0071}	0.0756 ^{+0.0090} _{-0.0087}	0.0785 ^{+0.0086} _{-0.00}		

1593 12 Methods for BR evaluation

1594 Different treatments of the data are possible to the purpose of measuring a $\text{BR}(B_s^0 \rightarrow \mu^+ \mu^-)$:
 1595 a comparison between data and signal and background expectations in bins of mass
 1596 and BDT, as the logical continuation of the upper limit evaluation, a likelihood fit to
 1597 the unbinned mass projections simultaneously in the 8 different BDT bins, and a two-
 1598 dimensional unbinned fit in the mass-BDT plane.

1599 All methods above will be briefly discussed in the following, and applied to a fake-
 1600 unblinded data sample. This has been obtained by merging the data sidebands of the
 1601 1 fb^{-1} sample with a toy-MC sample for the region $[5000 - 5418] \text{ MeV}/c^2$, generated in the
 1602 SM hypothesis ($10 B_s^0 \rightarrow \mu^+ \mu^-$ events). The combinatorial background parameterization
 1603 used for this generation has been an exponential pdf for the mass, and a triple exponential
 1604 for the BDT. For the double misID background the evaluation of Sec.7 has been used.
 1605 The generated sample corresponds to a CL_b in the range $0.982 - 0.989$, which is close to
 1606 a 3σ evidence.

1607 12.1 Binned methods

1608 A possible option to define a BR interval is by means of the CL_{s+b} . The main advantage
 1609 of this approach is the use of the same inputs as for the upper limit calculation, and in
 1610 particular the full treatment of systematic errors on the nuisance parameters.

1611 When using CL_{s+b} , the BR median value is defined by $\text{CL}_{s+b} = 0.5$, while the (negative)
 1612 derivative of CL_{s+b} with respect to BR, $-\text{dCL}_{s+b}/\text{dBR}$, can provide an estimator of the
 1613 BR pdf. The result we obtain on the fake-unblinded data sample is shown in Fig. 77
 (left). The value that maximizes $-\text{dCL}_{s+b}/\text{dBR}$ is $\text{BR} \sim 3.3 \times 10^{-9}$.

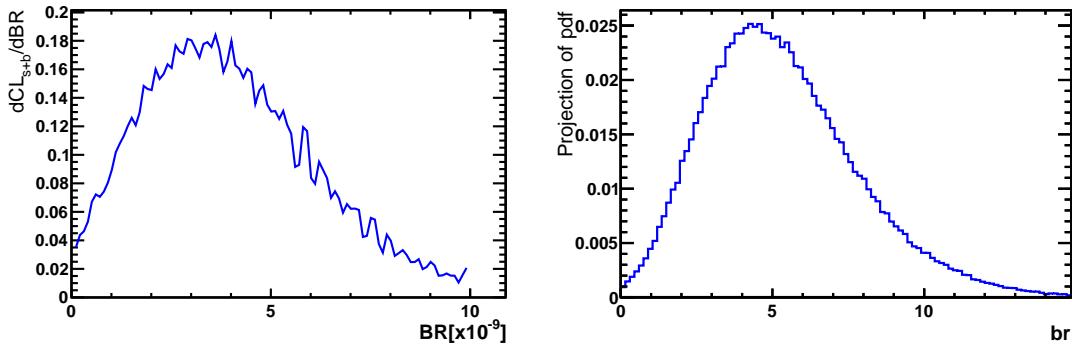


Figure 77: Fake-unblinded data sample: BR pdf from the computation of $-\text{dCL}_{s+b}/\text{dBR}$ (left), and from the evaluation of Bayesian posterior (right).

1614
 1615 From the same signal and background binned expectation values used as input to the
 1616 CL_{s+b} , it is also possible to extract a bayesian estimate of the BR. The BR pdf obtained
 1617 in this case is shown in the right panel of Fig. 77 , from which we get $\text{BR} = 4.4^{+2.8}_{-2.2} \times 10^{-9}$.

1618 **12.2 Simultaneous fit of mass projections in bins of BDT**

1619 A natural evolution of the above binned approach is an unbinned likelihood fit to the
1620 mass projections in the 8 BDT bins. In this case, we still profit of most of the existing
1621 work about calibration and systematic studies. The fit is performed simultaneously in
1622 the 8 BDT bins, with the signal constrained through the BDT fractions calibrated on
1623 the $B_{(s)}^0 \rightarrow h^+h^-$ sample (Sec. 9). More specifically, the BDT and mass signal pdf's
1624 are the ones used in the upper limit evaluation. The same occurs for the misidentified
1625 $B_{(s)}^0 \rightarrow h^+h^-$ events, with BDT and mass pdf's as described in Sec. 7. The uncertainties
1626 on the parameters defining these pdf's are included in the fit as gaussian constraints.

1627 The combinatorial background is instead evaluated for each BDT bin by including
1628 the mass sidebands in the fit range. The mass pdf assumed for this component is an
1629 exponential. As in the upper limit evaluation, the exponential slopes are left free between
1630 the different BDT bins, a part from the last two bins which are constrained to the same
1631 slope.

1632 The BR is extracted directly from the fit, by defining $\text{BR}/\alpha_{B_s^0 \rightarrow \mu^+\mu^-}$ as the normal-
1633 ization for the signal pdf. The uncertainty on $\alpha_{B_s^0 \rightarrow \mu^+\mu^-}$ is also included in the fit as a
1634 gaussian constraint.

1635 The mass fits for all of the 8 BDT bins are shown in Fig. 78. The resulting profile like-
1636 lihood scan is shown in Fig. 79, from which we obtain $\text{BR} = 4.2_{-1.9}^{+2.5} \times 10^{-9}$ (corresponding
1637 to $N(B_s^0 \rightarrow \mu^+\mu^-) = 13.1_{-6.0}^{+7.5}$).

1638 The relative weight of the systematics in the final error is tiny, as demonstrated by
1639 looking at the BR result without allowing fluctuations of the nuisance parameters in the
1640 fit: $\text{BR} = 4.1_{-1.9}^{+2.3} \times 10^{-9}$.

1641 To validate the fit, a set of ~ 1000 toys have been simulated. The results, shown in
1642 Fig. 80, clearly demonstrate the absence of appreciable biases both in the central value,
1643 and in the fit error.

1644 **12.3 Two-dimensional unbinned fit**

1645 The above results can be compared to a two-dimensional unbinned likelihood fit in the
1646 mass-BDT plane. The pdf's for signal and $B_{(s)}^0 \rightarrow h^+h^-$ misidentified events are defined
1647 as in the previous cases. For the combinatorial background a parameterization is needed to
1648 define the BDT pdf. In this case, a superposition of three exponential has been assumed,
1649 which is the same used in the generation of the events in the blinded mass region. This
1650 fit has to be considered therefore as an optimal reference point.

1651 To validate the fit, a set of 1000 toys have been simulated, with 10 $B_s^0 \rightarrow \mu^+\mu^-$ signal
1652 events superimposed to 15090 $\mu\mu$ combinatorial background events, as expected in the
1653 data unblinded region. The results, shown in Fig. 81, clearly demonstrate the absence of
1654 appreciable biases both in the central value, and in the fit error.

1655 The fit projections on mass and BDT for the fake-unblinded data sample are shown in
1656 Fig. 82. The systematic uncertainties on the nuisance parameters have been incorporated
1657 by repeating the fit after having fluctuated these parameters according to their measured

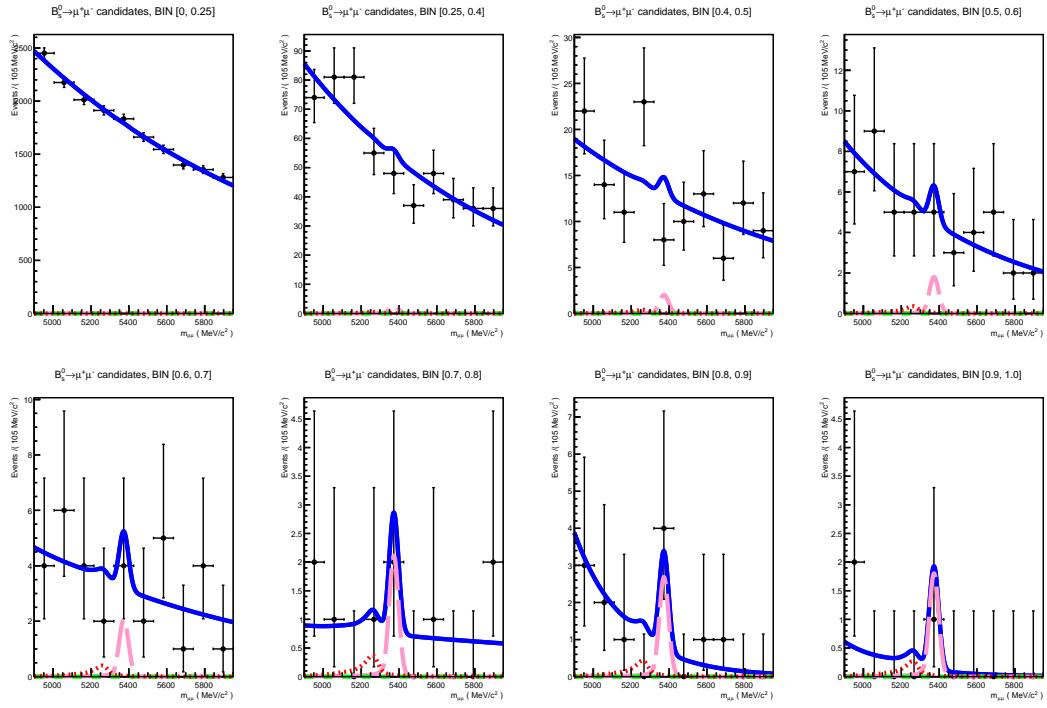


Figure 78: Likelihood mass fit in the 8 BDT bins, for the fake-unblinded sample: blue line is the total fit result, pink is $B_s^0 \rightarrow \mu^+\mu^-$ signal, red is $B_{(s)}^0 \rightarrow h^+h^-$ misID.

:

uncertainties. As a result, we obtain $N(B_s^0 \rightarrow \mu^+\mu^-) = 12.4^{+6.9}_{-5.4} \pm 3.1$, where the first error is statistical and the second systematic. The quoted error does not include an additional systematic contribution from the parameterization of the BDT background pdf, which has to be included when fitting the real unblinded sample. This has been evaluated to be ± 4 events.

This number of events corresponds to a $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-) = (4.0^{+2.2}_{-1.8} \pm 1.0) \times 10^{-9}$, in very good agreement with the simultaneous fit.

The fact that the simultaneous fit, which is binned in BDT, has the same accuracy as the unbinned one, can be interpreted as an optimal binning being reached, with a consequent negligible loss of information (and smaller systematics). For this reason, we plan to give the final result on the BR measurement by using the simultaneous fit approach.

In Table 47 the results from the three methods described above are summarized. In bold the method we plan to use for the final result.

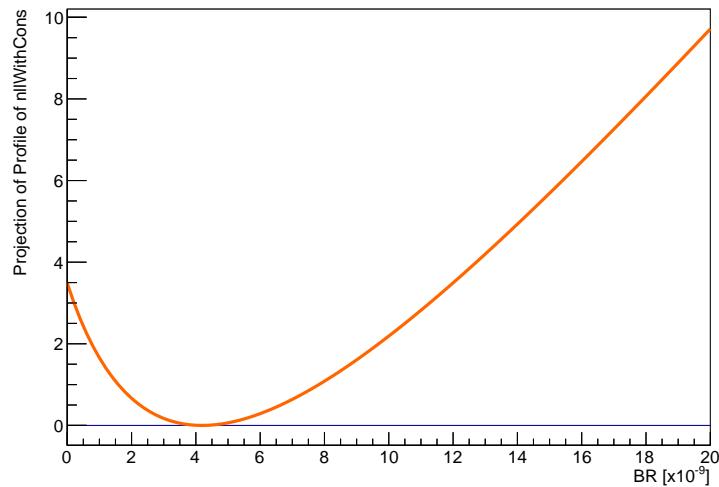


Figure 79: Profile likelihood scan corresponding to the simultaneous fit of the 8 BDT mass projections.

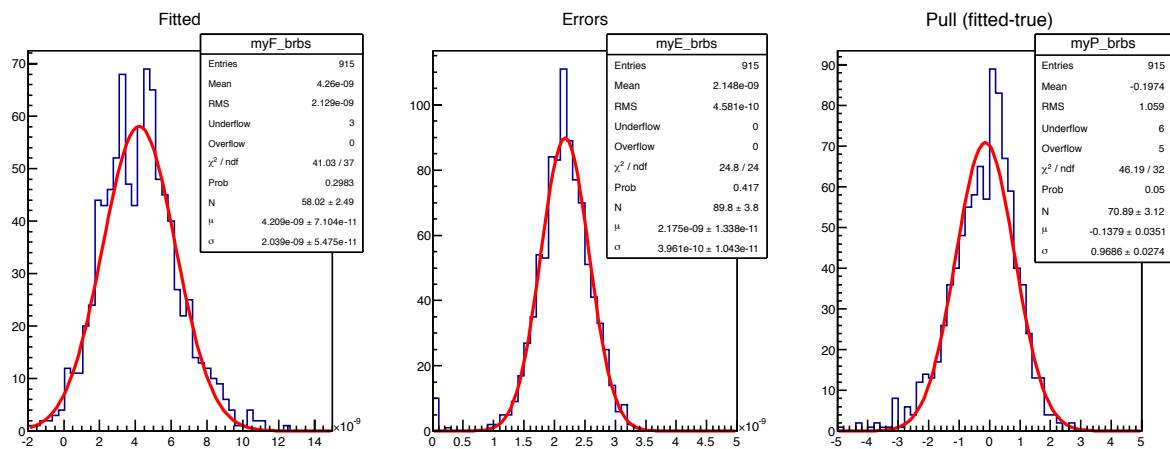


Figure 80: Toy validation for the simultaneous fit: $B_s^0 \rightarrow \mu^+ \mu^-$ yield, fit error and pull.

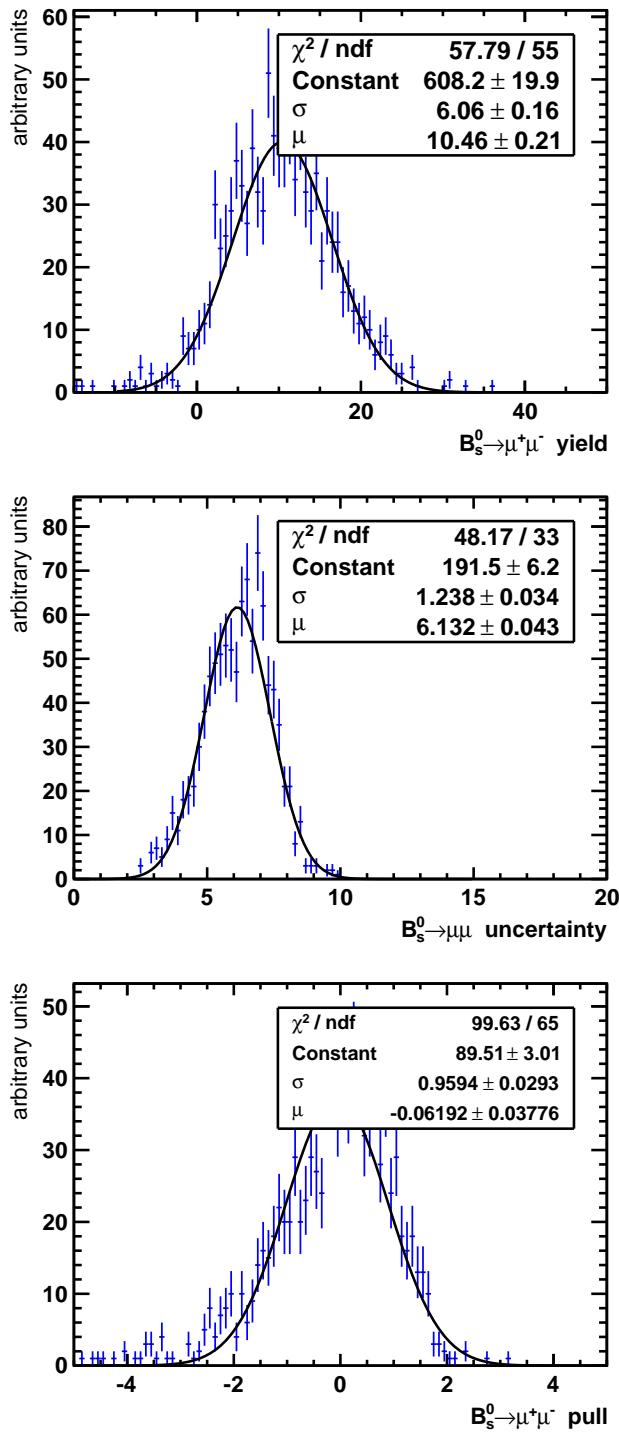


Figure 81: Toy validation for the two-dimensional fit: $B_s^0 \rightarrow \mu^+\mu^-$ yield, fit error and pull.

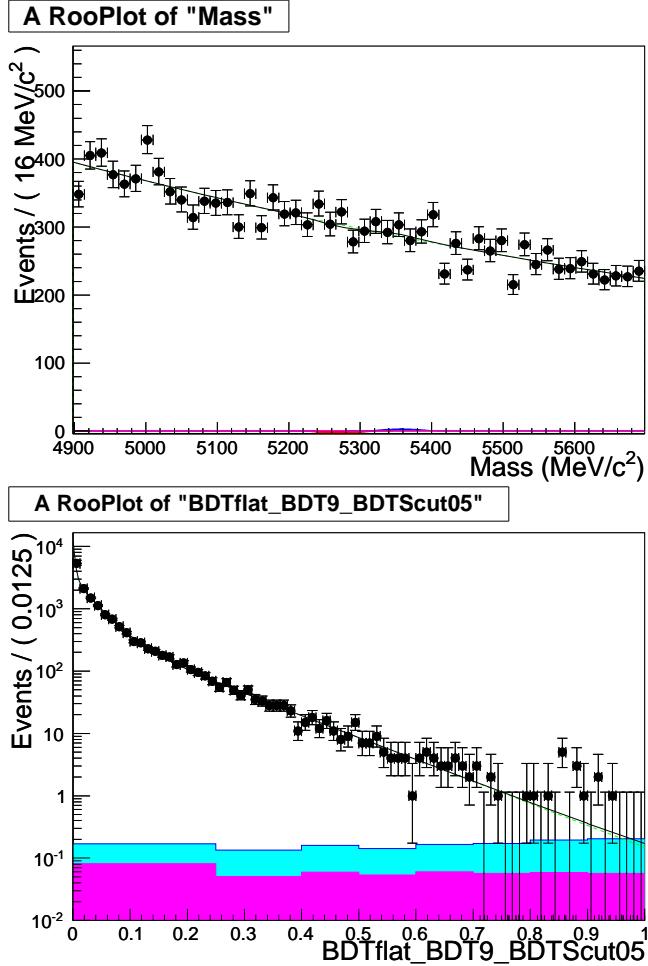


Figure 82: Two-dimensional fit results on the fake-unblinded data sample: mass projection (top), and BDT projection (bottom); cyan is $B_s^0 \rightarrow \mu^+ \mu^-$, purple is $B_{(s)}^0 \rightarrow h^+ h'^-$ misID.

Table 47: $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ extracted from the fake data sample with the three methods discussed in the text. The method in bold is the one we plan to use for the final result.

method	$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$
profile likelihood with simultaneous fit in 8 BDT bins	$(4.2^{+2.5}_{-1.9}) \times 10^{-9}$
unbinned likelihood fit	$(4.0^{+2.2}_{-1.8} \pm 1.0) \times 10^{-9}$
bayesian posterior	$(4.4^{+2.8}_{-2.2}) \times 10^{-9}$

1671 13 BR results

1672 In this section the BR results will be given for the unblinded data sample, obtained with
 1673 the simultaneous fit and the two-dimensional unbinned fit.

1674 13.1 Simultaneous fit of mass projections in bins of BDT

1675 The simultaneous fit approach has been repeated on the unblinded data sample, by al-
 1676 lowing both B_s^0 and B^0 signals floating freely during the minimization of the likelihood.
 1677 The mass fits for the 8 BDT bins are shown in Fig. 83, while the profile likelihood for
 1678 B_s^0 is shown on Fig. 84. The fitted BRs are $\text{BR}(B_s^0 \rightarrow \mu^+ \mu^-) = (0.79^{+1.68}_{-1.17}) \times 10^{-9}$, and
 1679 $\text{BR}(B^0 \rightarrow \mu^+ \mu^-) = (0.06^{+5.46}_{-4.27}) \times 10^{-10}$. The quoted errors include both the statistical fluc-
 1680 tuations and the systematic uncertainties from the knowledge of the nuisance parameters,
 1681 which are included in the fit through gaussian constraints.

1682 An additional systematic error of $\pm 0.46 \times 10^{-9}$ and $\pm 1.95 \times 10^{-10}$ has to be added in
 1683 quadrature to the B_s^0 and B^0 results, respectively. This has been evaluated by taking the
 1684 difference in the fit results using a double exponential for the combinatorial background
 1685 mass pdf instead of a single exponential.

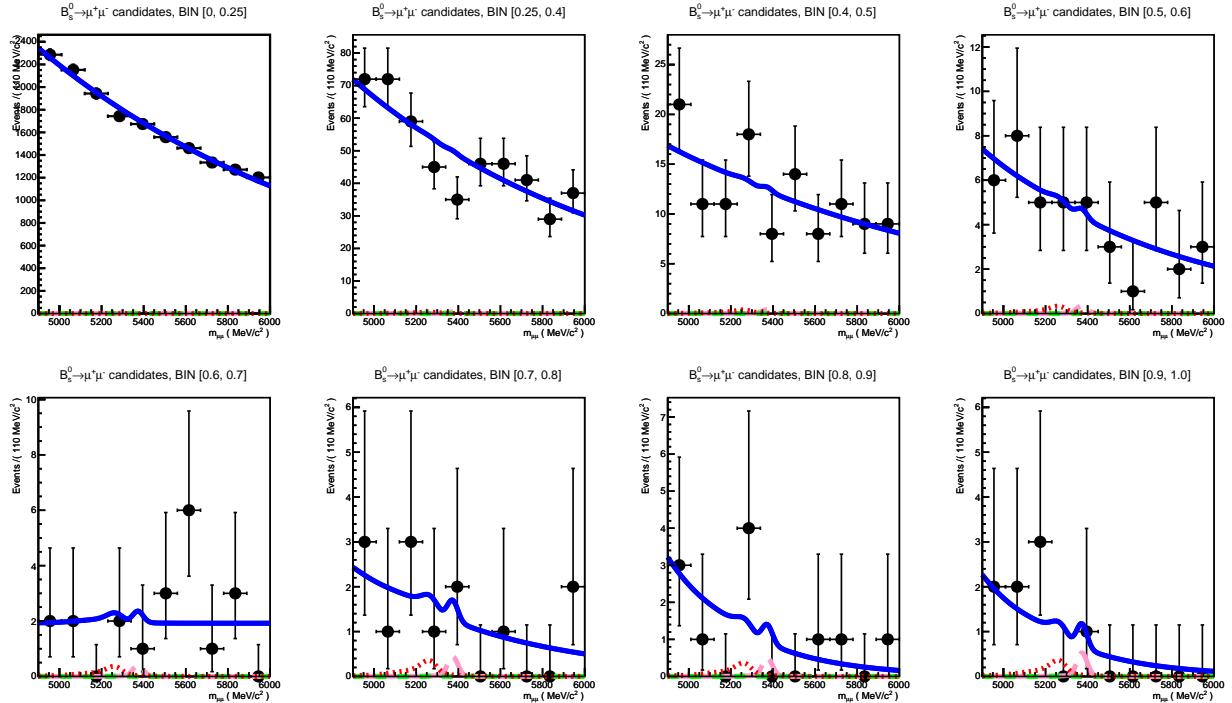


Figure 83: Likelihood mass fit in the 8 BDT bins, for the unblinded sample: blue line is the total fit result, pink is $B_s^0 \rightarrow \mu^+ \mu^-$ signal, red is $B^0 \rightarrow h^+ h'$ misID.

A RooPlot of "BRs"

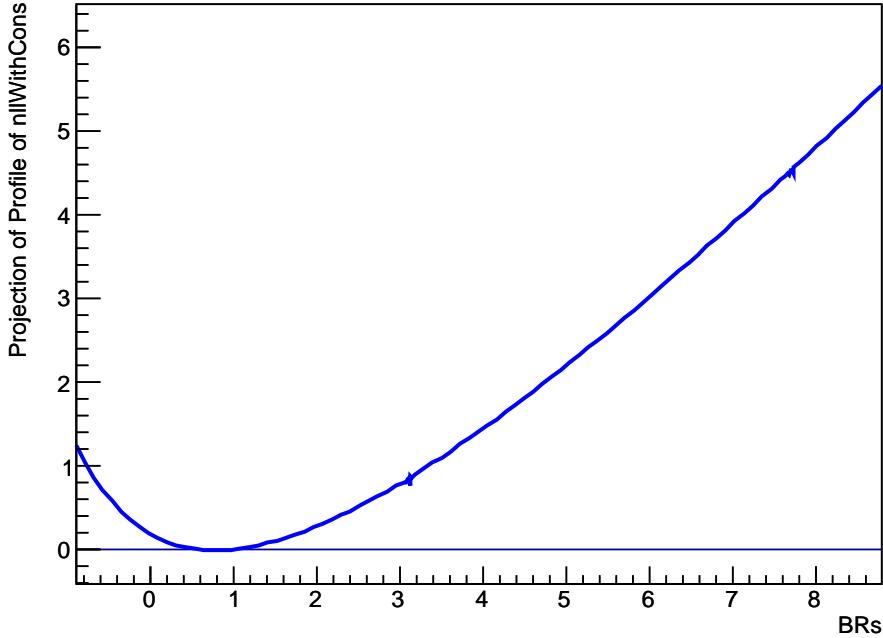


Figure 84: Unblinded data sample: likelihood profile obtained with the simultaneous fit.

1686 A toy MC study has been performed to check for biases and error coverage. The results
 1687 are shown in Fig 85. As expected for small numbers, a non gaussian tail is observed, which
 slightly biases the pull average and sigma at the level of $\sim 13\%$ and $\sim 5\%$, respectively.

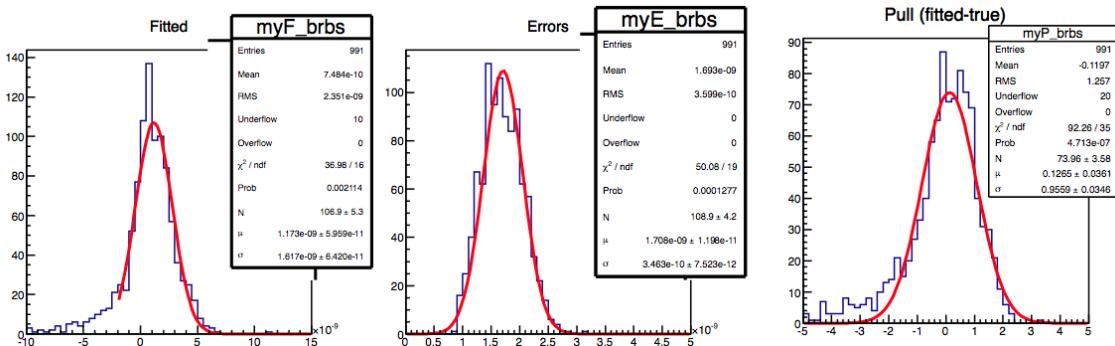


Figure 85: MC toy study for the B_s^0 simultaneous fit on unblinded data sample: distribution of fitted BRs, errors and pulls.

1688
 1689 The above MC toy study has been also used to cross-check the compatibility of the
 1690 simultaneous fit result with the CL_s confidence intervals reported in Sec.11. To this

purpose, we took the toy MC distribution of BR fitted values above zero, and we evaluated the fractional integral up to the SM value, $\sim 84\%$, and the BR values corresponding to 90% and 95% confidence levels, 3.8×10^{-9} and 4.4×10^{-9} , respectively. When adding a gaussian smearing to account for the additional systematic error from the combinatorial background pdf, the SM confidence level becomes 82.4 %, and the estimated 90 % and 95 % CL limits become 3.9×10^{-9} and 4.7×10^{-9} , respectively. These results are in good agreement with what obtained by the CL_s method.

13.2 Two-dimensional unbinned fit

The results with the two-dimensional unbinned fit on the unblinded data sample are reported as well. The fit has been performed on the whole BDT range and removing the first BDT bin fixing all the nuisance parameters. The fitted BRs on the whole range are $BR(B_s^0 \rightarrow \mu^+ \mu^-) = (1.44^{+1.73}) \times 10^{-9}$, and $BR(B^0 \rightarrow \mu^+ \mu^-) = (0.30^{+0.60}) \times 10^{-9}$. The fitted BRs when excluding the first BDT bin are $BR(B_s^0 \rightarrow \mu^+ \mu^-) = (1.47^{+1.76}) \times 10^{-9}$, and $BR(B^0 \rightarrow \mu^+ \mu^-) = (0.26^{+0.57}) \times 10^{-9}$. In all cases, the quoted errors include both the statistical fluctuations and the systematic uncertainties from the knowledge of the nuisance parameters, which are included in the fit through gaussian constraints.

The contribution of the systematic on the previously quoted uncertainties are depicted in tables 48 and 49. The values in the first quoted on the first column correspond to the systematic uncertainty when all the nuisance parameters are freed. On the second column we give the results when all the parameters but those that define the BDT for the background are freed. In the third column we specify the systematic uncertainty that arises when using a different model for the background BDT. Finally, on the last column we quote the systematic uncertainty given the different misID mass shapes. The final systematic uncertainty comes from the first and the last columns. Only in the case that the difference between the fit result using 3 exponentials or using the kernel estimator exceeds the single contribution from the background BDT nuisance parameters, the values on the second and third columns are used instead of the first one. Tables ?? and ??, give the results obtained with three different background models: 3 exponentials, RooKeysPdf and binned pdf (same 8 bins as signal). On the same tables, the systematic uncertainties assigned to the four models are reported as well.

Table 48: Systematic uncertainties of $B(B_s^0 \rightarrow \mu^+ \mu^-)$ and $B^0 \rightarrow \mu^+ \mu^-$ in the whole BDT range.

	All	without bkg BDT	Diff Rookeys	misID mass
$BR(B_s^0 \rightarrow \mu^+ \mu^-)$	$+0.545$ -0.456	$+0.136$ -0.202	$+0.000$ -0.730	$+0.010$ -0.030
$BR(B^0 \rightarrow \mu^+ \mu^-)$	$+0.277$ -0.190	$+0.056$ -0.055	$+0.000$ -0.290	$+0.030$ -0.010

The plots of the fit on the BDT axis and invariant mass axis are shown if 86.

1722 To check the stability of the fit we perform a set of 1000 simulated toys, where the
 1723 obtained fitted yields are required. The results, shown in Fig. 87, reflect that the errors
 1724 are correctly estimated and non-Gaussian. From the pulls, we estimate an small bias in
 1725 the B_s^0 yield of the order of 0.1 statistical sigma. This MC toy study is used to compute
 1726 the exclusion of BR ($B_s^0 \rightarrow \mu^+ \mu^-$) at a certain C.L. The obtained values are reported in
 1727 table 52. For comparison purposes, the CLs values are given in the same table reflecting
 1728 the agreement between them.

Table 49: Systematic uncertainties of $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ and $B^0 \rightarrow \mu^+ \mu^-$ excluding the first BDT bin.

	All	without bkg BDT	Diff Rookeys	misID mass
$\text{BR}(B_s^0 \rightarrow \mu^+ \mu^-)$	+0.340 -0.383	+0.134 -0.115	+0.000 -0.274	+0.016 -0.054
$\text{BR}(B^0 \rightarrow \mu^+ \mu^-)$	+0.149 -0.145	+0.037 -0.036	+0.000 -0.162	+0.034 -0.018

Table 50: Corresponding fit yields obtained with three different models: first row, 3 exponentials; second row, kernel estimator; third row, binned pdf with the same 8 bins as signal. The second part quotes the systematic uncertainties splitted in four different sources. All results are obtained fitting the whole BDT range.

	Yield $B_s^0 \rightarrow \mu^+ \mu^-$	Yield $B^0 \rightarrow \mu^+ \mu^-$
3 exponentials	$4.51^{+5.14}_{-3.51}$	$3.63^{+6.32}_{-4.48}$
RooKeysPdf	$2.23^{+5.12}_{-3.52}$	$0.13^{+5.91}_{-4.01}$
Binned	$4.13^{+5.17}_{-3.57}$	$0.86^{+6.32}_{-4.56}$
Syst. (signal mass)	$+0.096$ -0.066	$+0.227$ -0.166
Syst. (background mass)	$+0.025$ -0.005	$+0.154$ -0.057
Syst. (TIS)	$+0.384$ -0.397	$+0.495$ -0.621
Syst. (background BDT)	$+1.615$ -1.385	$+3.351$ -2.150

Table 51: Corresponding fit yields obtained with three different models: first row, 3 exponentials; second row, kernel estimator; third row, binned pdf with the same 8 bins as signal. The second part quotes the systematic uncertainties splitted in four different sources. All results are obtained fitting after removing the first BDT range.

	Yield $B_s^0 \rightarrow \mu^+ \mu^-$	Yield $B^0 \rightarrow \mu^+ \mu^-$
3 exponentials	$4.62^{+5.41}_{-3.72}$	$3.06^{+6.55}_{-4.65}$
RooKeysPdf	$3.76^{+5.50}_{-3.86}$	$1.15^{+6.33}_{-4.34}$
Binned	$4.18^{+5.31}_{-3.63}$	$0.38^{+6.50}_{-4.60}$
Syst. (signal mass)	$+0.093$ -0.083	$+0.251$ -0.171
Syst. (background mass)	$+0.008$ -0.014	$+0.061$ -0.050
Syst. (TIS)	$+0.194$ -0.179	$+0.373$ -0.333
Syst. (background BDT)	$+1.093$ -1.149	$+2.641$ -1.853

Table 52: Exclusion limits.

C.L.	2D fit	CLs (no syst.)
84%	3.2×10^{-9}	3.2×10^{-9}
90%	3.7×10^{-9}	3.7×10^{-9}
95%	4.4×10^{-9}	4.3×10^{-9}

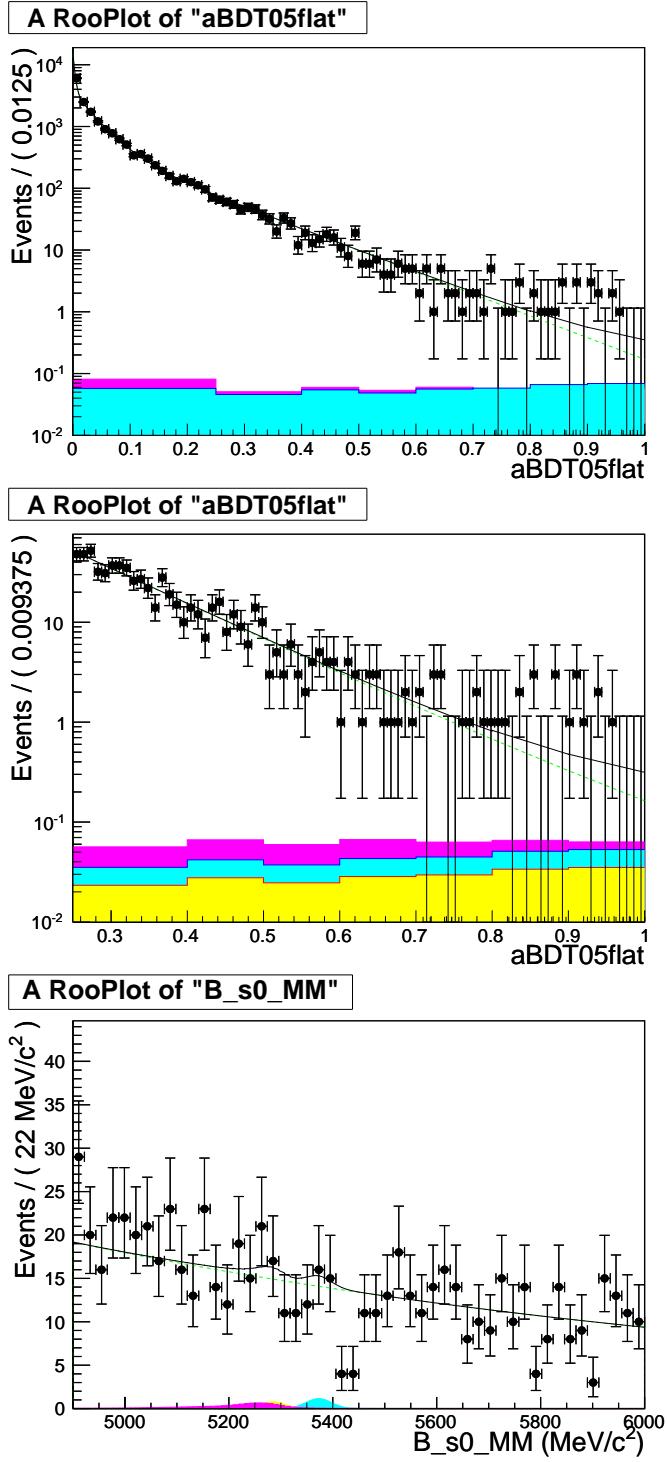


Figure 86: The top plot corresponds to the projection of the fit in the BDT axis using the whole BDT range. The middle and the bottom plots represent the projections of the fit when excluding $\text{BDT} < 0.25$ for BDT and invariant mass respectively. The blue area corresponds to the $B_s^0 \rightarrow \mu^+ \mu^-$ yield, the yellow area to the $B^0 \rightarrow \mu^+ \mu^-$ yield and the pink area to the misID yield.

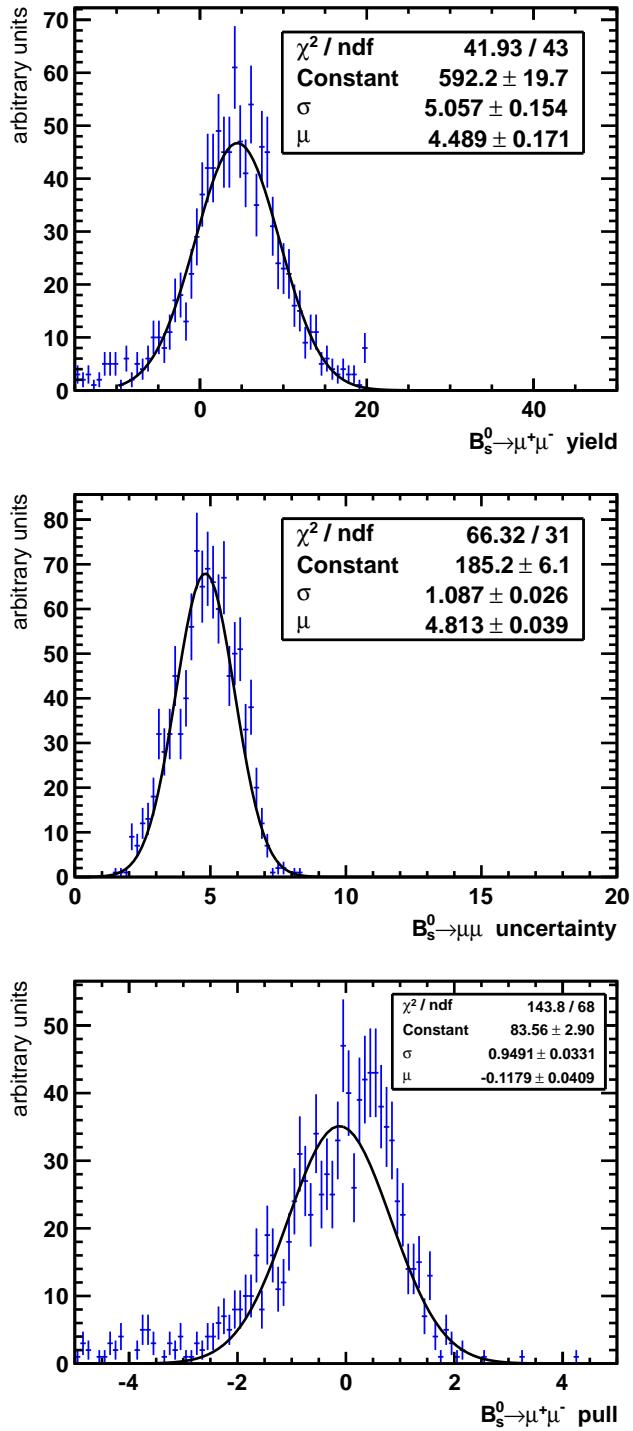


Figure 87: Toy validation for the two-dimensional fit on the unblinded data: $B_s^0 \rightarrow \mu^+\mu^-$ yield, fit error and pull.

₁₇₂₉ **14 Conclusions**

₁₇₃₀ A search for the decays $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ is performed with about 1 fb^{-1}
₁₇₃₁ of pp collisions at $\sqrt{s} = 7 \text{ TeV}$ collected by the LHCb experiment at the Large Hadron
₁₇₃₂ Collider at CERN. The observed event yields are consistent with both the background
₁₇₃₃ only hypothesis and the background plus $B_s^0 \rightarrow \mu^+ \mu^-$ (SM) hypothesis at $\sim 1\sigma$ level. We
₁₇₃₄ measure the following upper limits: $\text{BR}(B_s^0 \rightarrow \mu^+ \mu^-) < 4.5 \times 10^{-9}$ and $\text{BR}(B^0 \rightarrow \mu^+ \mu^-)$
₁₇₃₅ $< 10.3 \times 10^{-10}$ at 95% C.L. These are the lowest limits measured to date for these
₁₇₃₆ modes.

1737 **Appendix**

1738 **A Further cross-checks on signal BDT calibration**

1739 Here we present further results from the CERN 1D fit on $B_{(s)}^0 \rightarrow h^+h^-$ TIS events,
 1740 obtained under different mass hypotheses, and for different trigger unbias conditions.

1741 In Figs. 88 and 89 the mass fits are shown for L0-HLT1 TIS events, under $K\pi$ and KK
 1742 mass hypotheses, respectively. The corresponding BDT fractions are shown in Table 53.
 The results are in good agreement with what shown in Sec. 9.

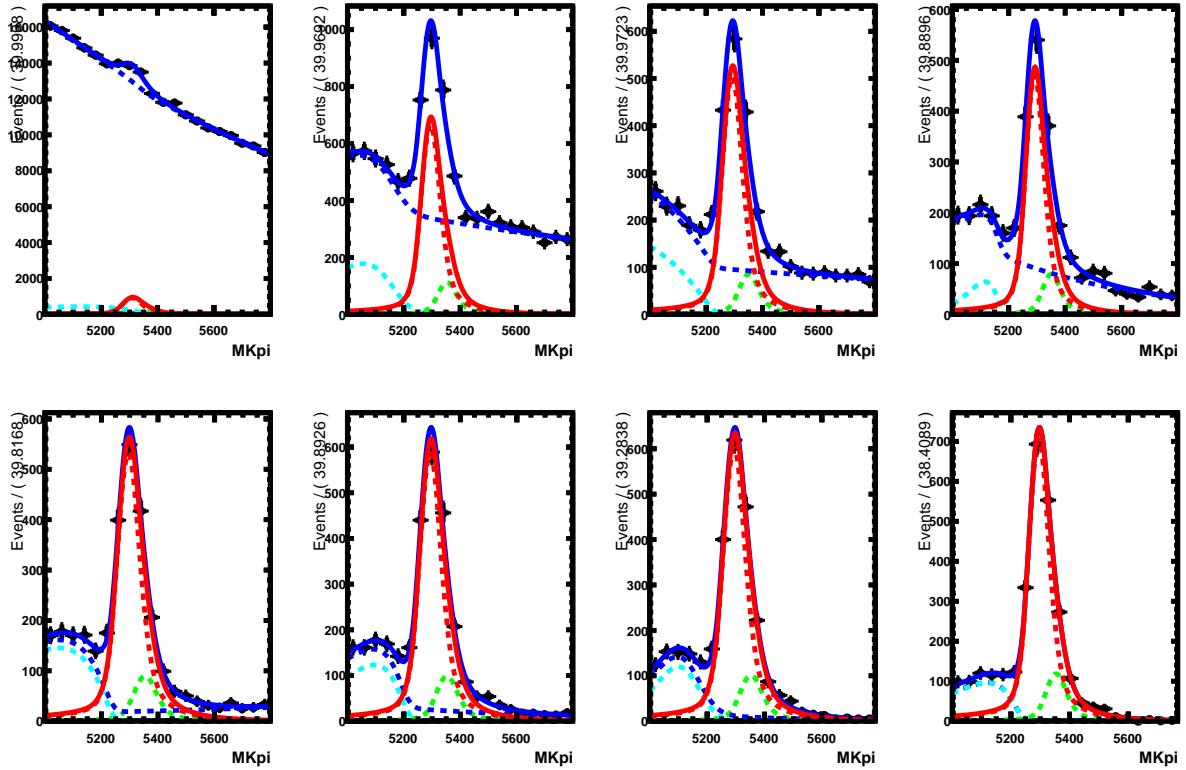


Figure 88: L0-HLT1 TIS events: mass fit results for all of the BDT bins (CERN), under $m_{K\pi}$ hypothesis.

1743 In Figs. 90 and 91 the mass fits are shown for L0-HLT1-HLT2 TIS events, under $K\pi$
 1744 and KK mass hypotheses, respectively. The corresponding BDT fractions are shown in
 1745 Table 53, including the $\mu\mu$ mass hypothesis, too. Again, the results are in good agreement
 1746

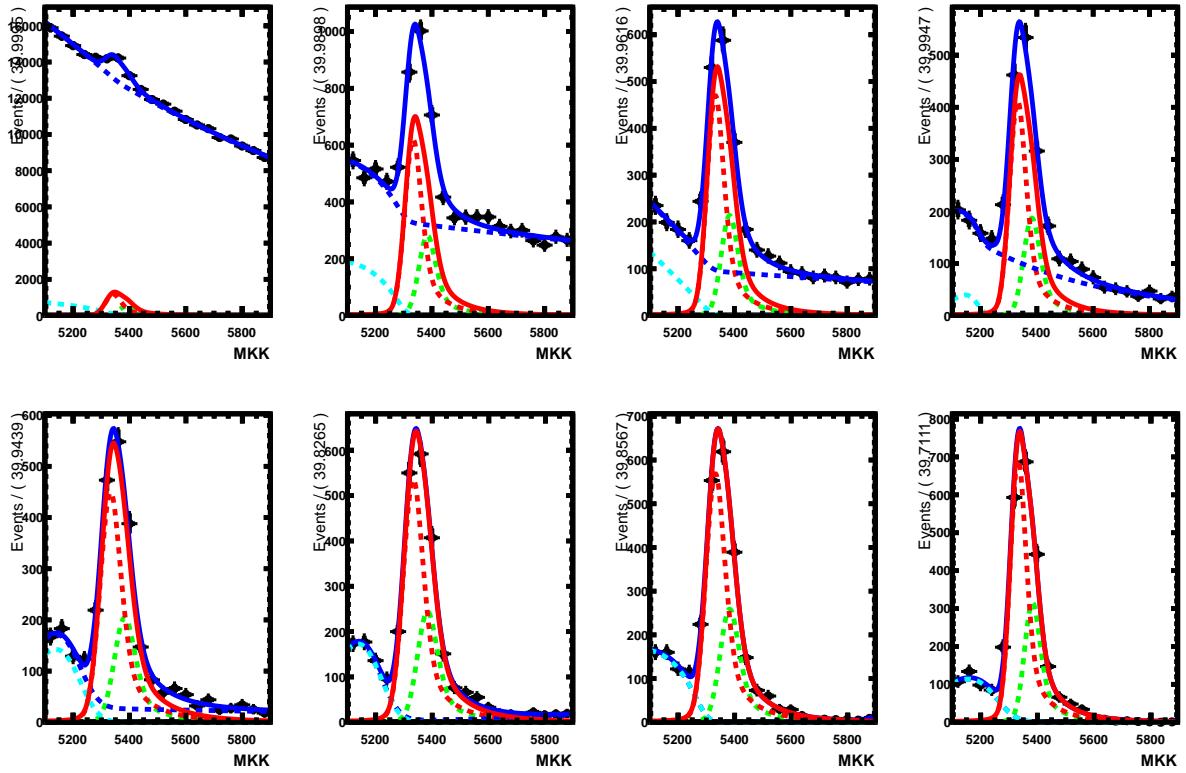


Figure 89: L0-HLT1 TIS events: mass fit results for all of the BDT bins (CERN), under m_{KK} hypothesis.

Table 53: BDT fractions for $B_{(s)}^0 \rightarrow h^+h'^-$ events, fitted under different mass hypotheses and trigger unbias conditions.

BDT range	$K\pi$	KK	$\mu\mu$ HLT2 TIS	$K\pi$ HLT2 TIS	KK HLT2 TIS
0 – 1	17.8	21.6	23.1	21.2	16.8
0.25 – 0.4	12.5	12.2	9.9	10.3	10.6
0.4 – 0.5	10.0	9.3	8.8	9.4	10.1
0.5 – 0.6	8.5	7.9	9.1	8.6	9.0
0.6 – 0.7	11.8	10.9	10.5	11.2	12.3
0.7 – 0.8	12.4	12.5	10.7	11.4	13.0
0.8 – 0.9	13.1	12.7	13.7	14.3	15.0
0.9 – 1.0	14.1	12.8	14.3	13.6	13.3

¹⁷⁴⁷ with what shown in Sec. 9, which is reasonable, since the effect of HLT2 trigger efficiency
¹⁷⁴⁸ is expected to be small.

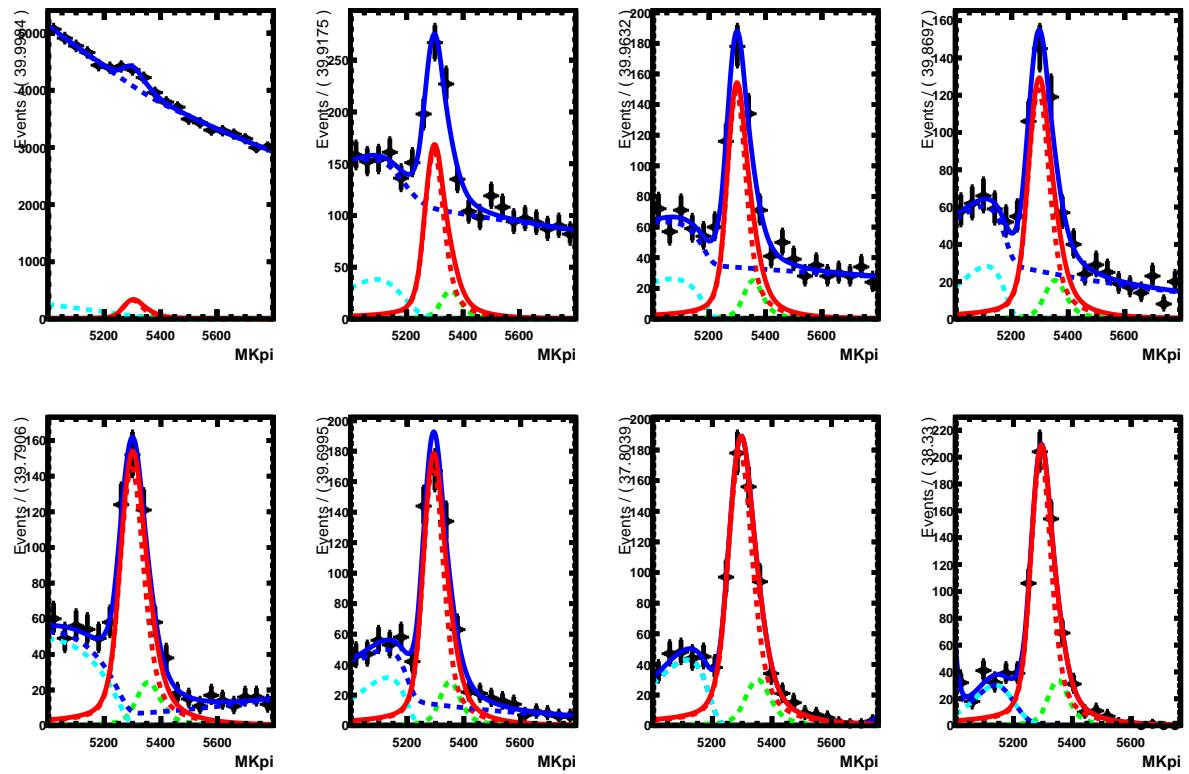


Figure 90: L0-HLT1-HLT2 TIS events: mass fit results for all of the BDT bins (CERN), under $m_{K\pi}$ hypothesis.

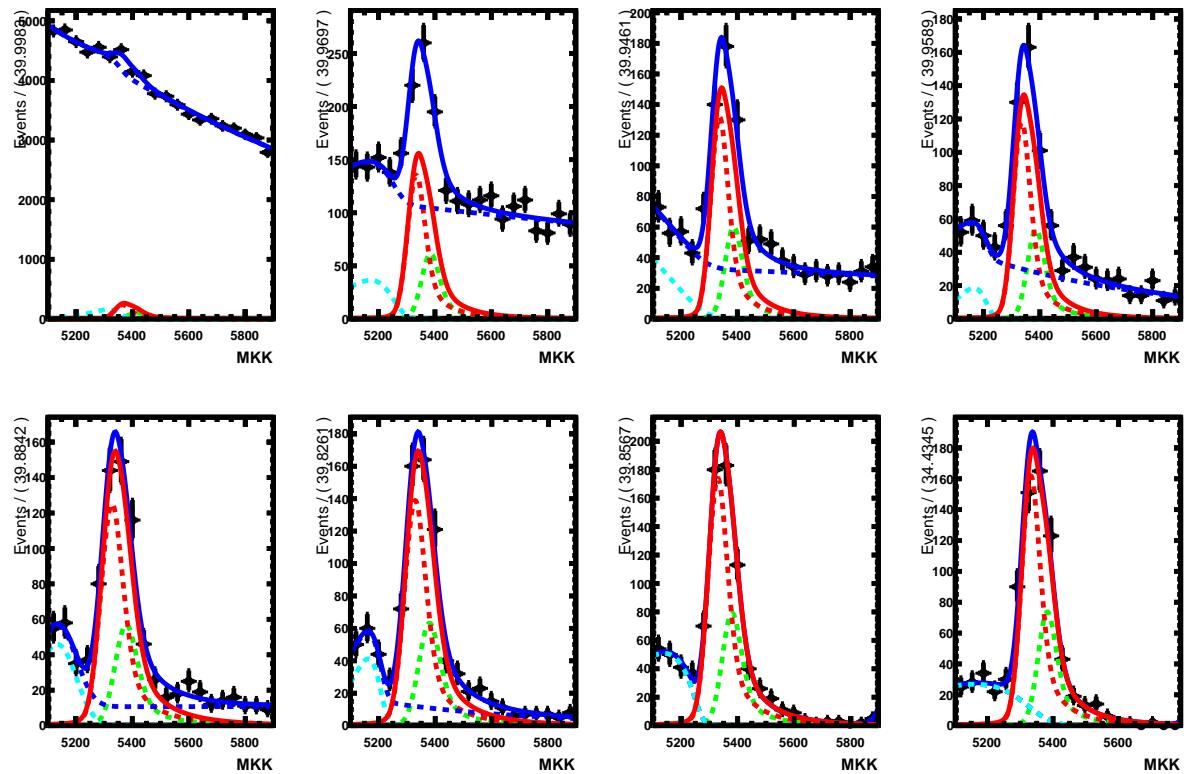


Figure 91: L0-HLT1-HLT2 TIS events: mass fit results for all of the BDT bins (CERN), under m_{KK} hypothesis.

¹⁷⁴⁹ B Appendix B: RICH Calibration

¹⁷⁵⁰ The problem with the RICH calibration is the following: if we use the efficiency tables
¹⁷⁵¹ provided by the RICH calibration group that give the efficiency corresponding to a given
¹⁷⁵² *DLL* cut in bins of p, p_T and track multiplicity in the event, we dont recover the efficiency
¹⁷⁵³ measured “by hands” with a simple counting method.

¹⁷⁵⁴ This test has been performed using a $B^+ \rightarrow J/\psi K^+$ sample. The efficiency of a given
¹⁷⁵⁵ $DLL(K - \pi)$ cut on the kaon has been computed in two ways:

¹⁷⁵⁶ 1. **tool method:** a $DLL(K - \pi)$ cut is applied to the kaon and the signal yield N_{cut}
¹⁷⁵⁷ of the $B^+ \rightarrow J/\psi K^+$ sample is determined via a fit to the invariant mass with a
¹⁷⁵⁸ double Crystal Ball for the signal and an exponential function for the background.
¹⁷⁵⁹ Then every event is weighted for the PID efficiency given by the RICH efficiency
¹⁷⁶⁰ map, the fit is repeated and the signal yield efficiency-corrected N_{corr} is determined.
¹⁷⁶¹ The efficiency if the DLL cut is defined as $\epsilon_{\text{tool}} = N_{\text{cut}}/N_{\text{corr}}$.

¹⁷⁶² 2. **counting method:** The efficiency ϵ_{cut} is simply determined by fitting the invariant
¹⁷⁶³ mass and extracting the yield before and after a $DLL(K - \pi)$ cut.

¹⁷⁶⁴ The results are shown in Fig. 92: ϵ_{cut} and ϵ_{tool} (left) and their difference (right) are
¹⁷⁶⁵ shown as a function of the DLL cut: a fixed difference of 4% and a variable differnce of
¹⁷⁶⁶ 0.4% are observed.

¹⁷⁶⁷ This discrepancy is even larger when decays containing two kaons in the final state
¹⁷⁶⁸ are considered, as in this case the RICH efficiency map has to take into account the
¹⁷⁶⁹ correlation between the two particles. The results obtained with a $B_s^0 \rightarrow J/\psi \phi$ sample
¹⁷⁷⁰ are shown in Fig. 93: in this case the difference between ϵ_{cut} (black markers) and ϵ_{tool}
¹⁷⁷¹ (red markers) is not constant but depends on the DLL cut applied.

¹⁷⁷² In Fig. 93 (left) other three curves are shown:

- ¹⁷⁷³ - light red: $\epsilon_{B_s^0 \rightarrow J/\psi \phi} = \epsilon_K^2(1 - \rho) + \epsilon_K$ with ϵ_K obtained from the RICH tool from the
¹⁷⁷⁴ $B^+ \rightarrow J/\psi K^+$ decay;
- ¹⁷⁷⁵ - light black: $\epsilon_{B_s^0 \rightarrow J/\psi \phi} = \epsilon_K^2(1 - \rho) + \epsilon_K$ with ϵ_K obtained from the cut-based technique
¹⁷⁷⁶ from the $B^+ \rightarrow J/\psi K^+$ decay;
- ¹⁷⁷⁷ - ligh brown: $\epsilon_{B_s^0 \rightarrow J/\psi \phi} = \epsilon_K^2$.

¹⁷⁷⁸ While the light brown curve is completely off showing that the correlations between the
¹⁷⁷⁹ two kaons cannot be neglected, once again we find that the efficiencies evaluated with the
¹⁷⁸⁰ rich efficiency map do not follow the expected behaviour. This mismatch will be further
¹⁷⁸¹ investigated with the RICH calibration group.

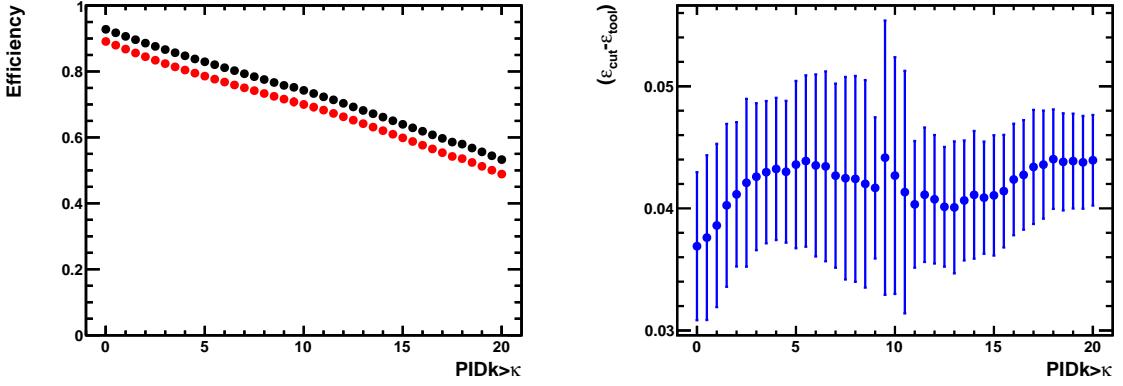


Figure 92: $B^+ \rightarrow J/\psi K^+$. Left: ϵ_{cut} (black markers) and ϵ_{tool} (red markers) as a function of the DLL cut. Right: $\epsilon_{\text{cut}} - \epsilon_{\text{tool}}$ as a function of the DLL cut. An almost constant difference of 4% with a variable difference of 0.4% are observed.

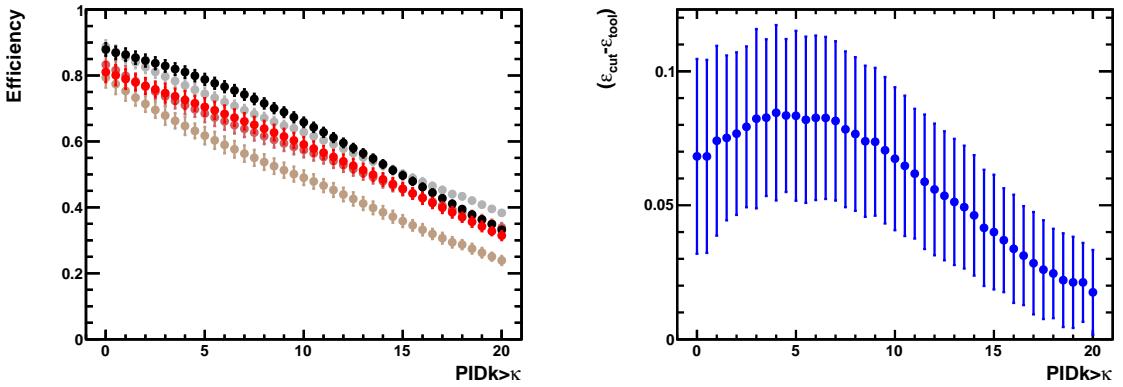


Figure 93: $B_s^0 \rightarrow J/\psi \phi$. Left: ϵ_{cut} (black markers), ϵ_{tool} (red markers) as a function of the DLL cut. Right: difference between ϵ_{rmcut} and ϵ_{tool} as function of the DLL cut. For the other curves see text.

1782 References

- 1783 [1] C. Bobeth et al., “Analysis of neutral Higgs-boson contributions to the decays $B_s^0 \rightarrow$
1784 l^+l^- and $B^0 \rightarrow Kl^+l^-$ ”, Phys. Rev. **D64** 074014 (2001).
- 1785 [2] A.J. Buras, M. V. Carlucci, S. Gori and G. Isidori, “Higgs-mediated FCNCs: Nat-
1786 ural Flavour Conservation vs. Minimal Flavour Violation”, JHEP 1010 (2010) 009,
1787 [arXiv:1005.5310]; A. J. Buras, “Minimal Flavour Violation and Beyond: Towards a
1788 Flavour Code for Short Distance Dynamics”, Acta Phys. Polon. **B41** (2010) 2487-
1789 2561 [arXiv:1012.1447].
- 1790 [3] L. J. Hall, R. Rattazzi and U. Sarid, “The top quark mass in supersymmetric SO(10)
1791 unification” Phys. Rev. D **50** (1994) 7048;
1792 C. Hamzaoui, M. Pospelov and M. Toharia, “Higgs-mediated FCNC in supersym-
1793 metric models with large $\tan\beta$ ”, Phys. Rev. D **59** (1999) 095005;
1794 K.S. Babu and C.F. Kolda, “Higgs-mediated $B_{s,d} \rightarrow \mu^+\mu^-$ in minimal supersymme-
1795 try”, Phys. Rev. Lett. **84** (2000) 228.
- 1796 [4] V. Abazov *et al.* [D0 Collaboration], “Search for the rare decay $B_s^0 \rightarrow \mu^+\mu^-$ ”, Phys.
1797 Lett. B **693** (2010) 539 [arXiv:1006.3469].
- 1798 [5] T. Aaltonen *et al.* [CDF collaboration], “Search for $B^0 \rightarrow \mu^+\mu^-$ and $B_s^0 \rightarrow \mu^+\mu^-$
1799 Decays with CDF II”, Phys. Rev. Lett. **107** (2011) 191801, [arXiv:1107.2304].
- 1800 [6] CMS Collaboration, S. Chatrchyan *et al.*, “Search for $B^0 \rightarrow \mu^+\mu^-$ and $B_s^0 \rightarrow$
1801 $\mu^+\mu^-$ decays in pp Collisions at 7 TeV”, Phys. Rev. Lett. **107** (2011) 191802,
1802 [arXiv:1107.5834].
- 1803 [7] , R. Aaij *et al.* [LHCb Collaboration], “Search for the rare decays $B^0 \rightarrow \mu^+\mu^-$ and
1804 $B_s^0 \rightarrow \mu^+\mu^-$ ”, arXiv: 1112.1600, accepted by Phys. Lett. B.
- 1805 [8] LHCb collaboration, R. Aaij *et al.*, “Search for the Rare Decays $B^0 \rightarrow \mu^+\mu^-$ and
1806 $B_s^0 \rightarrow \mu^+\mu^-$ ”, Phys.Lett. B699 (2011) 330–340, [arXiv:1103.2465].
- 1807 [9] TMVA, Toolkit for Multi Variate analysis with ROOT, <http://tmva.sourceforge.net>.
- 1808 [10] T. Skwarnicki, A Study of the Radiative Cascade Transitions Between the Υ and Υ'
1809 resonances. PhD thesis, DESY F31-86-02, 1986. Appendix E.
- 1810 [11] R. Aaij *et al.*, [LHCb Collaboration], “Measurement of $\sigma(pp \rightarrow b\bar{b} X)$ at $\sqrt{s}=7$
1811 TeV in the forward region”, Physics Letters B **694** (2010) 209.
- 1812 [12] O.Buchmueller *et al.*, “Likelihood functions for supersymmetric observables in fre-
1813 quentist analyses of the CMSSM and NUHM1”, arXiv:0907.5568.
- 1814 [13] M. Perrin Terrin, G. Mancinelli “Optimization of the binning of the discriminating
1815 variables used in the computation of the $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-)$ upper limits with the
1816 modified frequentist approach”, LHCb-INT-2012-003

- 1817 [14] S.S.Wilk, “The large sample distribution of the likelihood ratio for testing composite
1818 hypotheses” *The Annals of Mathematical Statistics* 9 (1938) 60-62.
- 1819 [15] R. Aaij and J. Albrecht, “Muon triggers in the High Level Trigger of LHCb”, LHCb-
1820 PUB-2011-017
- 1821 [16] T. Aaltonen *et al.* [CDF Collaboration], “Search for $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$
1822 decays with CDF-II” , arXiv:0508036v3 [hep-ex] (2011).
- 1823 [17] C. Adrover et al., *Measurement of the BR of $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ decays
1824 using a maximum likelihood fit*, LHCb-INT-2011-023.
- 1825 [18] Y. Shcheglov, *Background Suppression Study with the BDT method*,
1826 <https://indico.cern.ch/conferenceDisplay.py?confId=133530>.
- 1827 [19] The CDF Collaboration, T. Aaltonen et al., “Search for $B_s^0 \rightarrow \mu^+ \mu^-$ and $B_d^0 \rightarrow \mu^+ \mu^-$
1828 Decays in 3.7 fb^{-1} of $p\bar{p}$ Collisions with CDF II”, CDF Public Note 9892.
- 1829 [20] Daejung Kong, *Rare Decays at Tevatron*, talk given at Beauty 2011, Amsterdam,
1830 April 4-8, 2011, <http://www.beauty2011.nikhef.nl/index.html>.
- 1831 [21] The DØ Collaboration, V. Abazov et al., “Search for the rare decay $B_s^0 \rightarrow \mu^+ \mu^-$ ”,
1832 Phys. Lett., **B693** (2010) 539, arXiv:1006.3469 [hep-ex].
- 1833 [22] C. Adrover et al., “Search for the rare decays $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ with the LHCb experi-
1834 ment”, LHCb-ANA-2011-007.
- 1835 [23] C. Adrover et al., “Search for the rare decays $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ with the
1836 LHCb experiment”, LHCb-ANA-2011-078.
- 1837 [24] The Heavy Flavor Averaging Group, D. Asner, “Averages of b -hadron, c -
1838 hadron, and τ -lepton Properties”, arXiv:1010.1589 [hep-ex]. Updates available at
1839 http://www.slac.stanford.edu/xorg/hfag/osc/end_2009.
- 1840 [25] A. Read, “Presentation of Search Results: The CL_s Technique”, J. Phys. **G28** (2002)
1841 2693.
- 1842 [26] “BRUNEL, The LHCb Event Reconstruction program”,
1843 <http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/brunel/>.
- 1844 [27] “CondDB, The Condition Data Base for the LHCb experiment”,
1845 <https://twiki.cern.ch/twiki/bin/view/LHCb/CondDBHowTo>.
- 1846 [28] “DDDB, The LHCb Detector Description”,
1847 <https://lhcb-comp.web.cern.ch/lhcb-comp/Frameworks/DetDesc/default.htm>.
- 1848 [29] “DAVINCI, The Physics Analysis Software for the LHCb experiment”,
1849 <http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/davinci/>.

- 1850 [30] P. Mato, “GAUDI-Architecture design document”, LHCb note 98-064.
- 1851 [31] “GAUSS, the LHCb Simulation Program”,
 1852 <http://cern.ch/LHCb-release-area/DOC/gauss>.
- 1853 [32] T. Sjöstrand, S. Mrenna and P. Skands, “PYTHIA 6.4: Physics and manual”, JHEP
 1854 **05** (2006) 026, [hep-ph/0603175].
- 1855 [33] “PHOTOS Monte Carlo for radiative corrections in particle decays”,
 1856 <http://heptools.inp.demokritos.gr/node/1021>.
- 1857 [34] “LHAPDF, Les Houches Accord PDF Interface”,
 1858 <http://projects.hepforge.org/lhapdf/>.
- 1859 [35] “HepMC - A C++ Event Record for Monte Carlo Generators”,
 1860 <https://savannah.cern.ch/projects/hepmc>.
- 1861 [36] “GEANT 4, toolkit for the simulation of the passage of particles through matter”,
 1862 Nucl. Instr. and Meth. in Phys. Res. **A506** (2003) 250, Nuclear Science 53 No. 1,
 1863 270 (2006)
 1864 Home page: <http://www.geant4.org/geant4/>.
- 1865 [37] M. Williams, V. Gligorov, *The HLT2 topological trigger*, LHCb-INT-xxx, note in
 1866 preparation.
- 1867 [38] S. Kullback and R. A. Leibler, *On Information and Sufficiency*, Ann. Math. Statist.
 1868 **22** (1951) 79-86; S. Kullback, *Letter to Editor: the Kullback-Leibler Distance*, The
 1869 American Statistician **41** (1987) 340. The use of the Kullback-Leibler distance at
 1870 LHCb is described in M. Needham *Clone Track identification using the Kullback-*
 1871 *Liebler distance*, LHCb-2008-002.
- 1872 [39] The acceptance efficiencies corresponding to the
 1873 quoted Monte Carlo production are listed in:
<http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/STATISTICS/MC10STAT/RD->
- 1875 [40] the LHCb Collaboration, “Measurement of J/ψ production in pp collisions at $\sqrt{s} =$
 1876 7 TeV”, Eur. Phys. J. C71 (2011), 1645.
- 1877 [41] The LHCb Collaboration, “The LHCb detector at LHC”, Journal of Instrumentation
 1878 (JINST) **3** (2008) S08005.
- 1879 [42] E. van Herwijnen, “The LHCb Trigger”, LHCb-PROC-2010-055.
- 1880 [43] G. Lanfranchi et al., “The Muon Identification procedure of the LHCb experiment
 1881 for the first data”, LHCb-PUB-2009-013.
- 1882 [44] A. Sarti et al., “Calibration strategy and efficiency measurement of the muon iden-
 1883 tification procedure at LHCb”, LHCb-PUB-2010-002.

- 1884 [45] A. Sarti et al., “Performance of the LHCb Muon identification procedure with the
 1885 first 30 pb^{-1} of data”, LHCb-INT-2011-xx (in preparation).
- 1886 [46] K. Akiba et al., “Results on Muon identification efficiency with 2011 data at LHCb”,
 1887 LHCb-INT-2011-045.
- 1888 [47] J. Albrecht et al., “Updated measurement of the J/Ψ production cross section in
 1889 LHCb”, LHCb-ANA-2010-012.
- 1890 [48] M. Calvi et al., “Lifetime unbiased selection of $B_s^0 \rightarrow J/\psi\phi$ and related control
 1891 channels”, LHCb-2009-025.
- 1892 [49] S. Hansmann-Menzemer, “Status of tracking”, presented at LHCb week 30.06.2010
- 1893 [50] The LHCb Collaboration, “Measurement of prompt and K_s^0 production ratios in
 1894 inelastic non-diffractive pp collisions at $\sqrt{s} = 0.9$ and 7 TeV”, LHCb-CONF-2010-11
- 1895 [51] The LHCb Collaboration, ”Central Exclusive Dimuon Production at $\sqrt{s}=7 \text{ TeV}$ ”,
 1896 LHCb-CONF-2011-022
- 1897 [52] K. Akiba et al., ”Measurement of Δm_d in the decay $B_d \rightarrow D(K\pi\pi)\pi$ and calibration
 1898 of the combined opposite side taggers”, LHCb-ANA-2011-004
 1899 Dedicated note in preparation: S. Stahl, “Smearing of track parameters to account
 1900 for resolution differences in data and Monte Carlo”, LHCb-INT-2011-027
- 1901 [53] J.A. Hernando Morata et al., “Measurement of trigger efficiencies and biases”, LHCb-
 1902 2008-073.
- 1903 [54] Chapter 5 in “Roadmap for selected key measurements of LHCb”, arXiv:0912.4179
 1904 [hep-ex].
- 1905 [55] <https://twiki.cern.ch/twiki/bin/view/LHCb/LHCbTrackingEfficiencies>
- 1906 [56] fsfd LHCb average
- 1907 [57] M. Artuso et al., “Measurement of b -hadron production fractions in 7 TeV centre-of-
 1908 mass energy pp collisions”, LHCb-ANA-2010-015, in preparation.
- 1909 [58] L. Carson et al., “Measurement of the relative yields of the decay modes $B_{d,s} \rightarrow D_{d,s}\pi$
 1910 and $B_{d,s} \rightarrow D_{d,s}K$ and determination of f_d/f_s for 7 TeV pp collisions”, LHCb-ANA-
 1911 2010-010, in preparation.
- 1912 [59] The Particle Data Group, K. Nakamura et al., “Review of Particle Physics”, J. Phys.
 1913 G: Nucl. Part. Phys. **37** (2010) 075021.
- 1914 [60] The BELLE Collaboration, R. Louvot, “ $\Upsilon(5S)$ Results at BELLE”, arXiv:0905.4345
 1915 [hep-ex].

- 1916 [61] D. Karlen, “Using Projections and Correlations to Approximate Probability Distrib-
1917 utions”, Comp. Phys. **12** (1998) 380.
- 1918 [62] D. Martínez Santos, J. A. Hernando and F. Teubert, “LHCb Potential to Mea-
1919 sure/Exclude the Branching Ratio of the Decay $B_s^0 \rightarrow \mu^+ \mu^-$ ”, LHCb-2007-033.
- 1920 [63] C. Adrover et al., *Search for the rare decays $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ with 370 pb⁻¹ at LHCb*,
1921 LHCb-ANA-2011-078.
- 1922 [64] D. Martínez Santos, “Study of the very rare decay $B_s^0 \rightarrow \mu^+ \mu^-$ in LHCb”, CERN-
1923 THESIS-2010-068.
- 1924 [65] G. Ciezarek et al., *The first observation of $B^+ \rightarrow p\mu^+\mu^-$* . LHCb-CONF-2012-006.
- 1925 [66] R. Brun, F. Rademakers, “ROOT: An Object Oriented Data Analysis Framework”,
1926 Nucl. Instrum. Meth. **A389** (1997) 81.
- 1927 [67] P. Speckmayer et al., “The Toolkit for Multivariate Data Analysis TMVA 4”, J.
1928 Phys. Conf. Ser. **219** (2010) 032057.
- 1929 [68] A. Carbone et al., “Invariant mass line shape of $B \rightarrow h^+ h^-$ decays at LHCb”, LHCb-
1930 PUB-2009-031.
- 1931 [69] R.L. Gluckstern, “Uncertainties in track momentum and direction, due to multiple
1932 scattering and measurement errors”, Nucl. Instr. and Meth. **24** (1963) 381.
- 1933 [70] T. Junk, “Confidence Level Computation for Combining Searches with Small Statis-
1934 tics”, Nucl. Instrum. Meth. **A434** (1999) 435, hep-ex/9902006.
- 1935 [71] C. Aldrover et al., “Search for the rare decays $B_q^0 \rightarrow \mu\mu^+\mu^-$ with the LHCb exper-
1936 iment”, LHCb-ANA-2011-007, March 2, 2011.
- 1937 [72] W. Verkerke, D. Kirkby, “The ROOFIT Toolkit for Data Modeling”,
1938 arXiv:physics/0306116
- 1939 [73] P. Koppenburg, “Selection of $B_u \rightarrow llK$ at LHCb and Sensitivity to R_K ”, CERN-
1940 LHCb - 2007-034.
- 1941 [74] F. Abe et al. [CDF Collaboration], “Observation of the B_c meson in $p\bar{p}$ collisions at
1942 $\sqrt{s} = 1.8$ TeV,” Phys. Rev. Lett. **81** (1998) 2432 [hep-ex/9805034].
- 1943 [75] T.M. Aliev et al., “Rare radiative $B \rightarrow \tau^+\tau^-\gamma$ decay,” Phys. Lett. B **424** (1998) 175
1944 [hep-ph/9710304].