

Search for the $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ decays with 3 fb^{-1} at LHCb

C. Adrover¹, J. Albrecht², F. Archilli⁶, M.-O. Bettler³, X. Cid Vidal⁵, F. Dettori⁷, C. Elsasser⁴, G. Lanfranchi⁶, G. Mancinelli¹, D. Martínez Santos⁷, A. Mordà¹, S. Oggero⁷, M. Palutan⁶, A. Pellegrino⁷, M. Perrin-Terrin¹, M. Rama⁶, A. Sarti^{6,8}, M. Schlupp², J. Serrano¹, B. Sciascia⁶, F. Soomro⁶, S. Tolk⁷, A. Ustyuzhanin⁹, R. Vazquez Gomez⁶

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

² Fakultät Physik, TU Dortmund, Dortmund, Germany

³ Cavendish Laboratory, University of Cambridge, Cambridge, United Kingdom

⁴ Universität Zürich, Zurich, Switzerland.

⁵ European Organisation for Nuclear Research (CERN), Geneva, Switzerland.

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

⁷ Nikhef, National Institute for Subatomic Physics, Amsterdam, Netherlands.

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

⁹ Imperial College, London & Yandex, Moscow

Abstract

A search for the rare decays $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ is performed using the LHCb experiment. The dataset corresponds to an integrated luminosity of 2 fb^{-1} at $\sqrt{s} = 8 \text{ TeV}$ and 1 fb^{-1} at $\sqrt{s} = 7 \text{ TeV}$. An excess of $B_s^0 \rightarrow \mu^+ \mu^-$ signal candidates with respect to the background expectation is observed with significance of 4.0 standard deviations, while the significance of the $B^0 \rightarrow \mu^+ \mu^-$ decay is 2.0σ . A branching fraction of $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (2.9^{+1.1}_{-1.0}) \times 10^{-9}$ is obtained and an upper limit of $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) < 7.4 \times 10^{-10}$ at 95% confidence level is set. These results are consistent with the Standard Model expectations.

Contents

1	Introduction	4
2	Analysis strategy	5
3	Data and Monte Carlo Samples	7
3.1	Data Sample	7
3.2	Monte Carlo samples	8
3.3	Trigger	10
4	Selection	13
4.1	The BDTS discriminant	14
4.2	Event matching between reco13 and reco14	15
5	Muon Identification	16
6	Misidentification rates and peaking backgrounds	21
6.1	Pion and kaon misidentification probabilities	21
6.2	Proton misidentification probability	26
6.3	Peaking background from $B_{(s)}^0 \rightarrow h^+h^-$ with double misidentification: yield, BDT and mass PDFs	29
6.4	Peaking backgrounds different from $B_{(s)}^0 \rightarrow h^+h^-$	36
6.4.1	Determination of mass shape and BDT fractions	37
6.4.2	$B^0 \rightarrow \pi^-\mu^+\nu_\mu$	38
6.4.3	$B_s^0 \rightarrow K^-\mu^+\nu_\mu$	40
6.4.4	$B^+ \rightarrow \pi^+\mu^+\mu^-$ and $B^0 \rightarrow \pi^0\mu^+\mu^-$	41
6.4.5	$B_c^+ \rightarrow J/\psi\mu^+\nu$	44
6.4.6	$\Lambda_b^0 \rightarrow p\mu^-\nu$	47
6.5	Background from the generic $b\bar{b} \rightarrow \mu^+\mu^-X$	50
6.6	Summary of exclusive backgrounds	53
7	Signal classification	54
7.1	BDT definition	54
7.1.1	Discussion about the BDT choice	62
7.1.2	Performances of the new BDT12 on MC	63
7.1.3	Comparison between BDT9 and BDT12 on data sidebands	68
7.2	BDT calibration	71
7.2.1	$\Delta LL_{K-\pi}$ efficiency evaluation	71
7.2.2	Background from $\Lambda_b^0 \rightarrow ph$	73
7.2.3	BDT calibration with $B_{(s)}^0 \rightarrow h^+h^-$	75
7.2.4	BDT shape correction from DLL cut efficiency	81
7.2.5	BDT shape correction from trigger efficiency	84
7.3	Invariant mass	86

7.3.1	Central value	86
7.3.2	Invariant mass resolution	88
7.3.3	Crystal Ball parameters α and n	96
7.4	Combination between Stripping 20r1 and 20	97
8	Normalization	99
8.1	Acceptance efficiencies from MC simulation	99
8.2	Ratio of reconstruction and selection efficiencies	100
8.2.1	Reconstruction and selection efficiencies from simulation	100
8.2.2	Correction of the tracking efficiency	103
8.2.3	Muon detector acceptance and muonID efficiency	104
8.2.4	Ghost probability efficiency	105
8.2.5	Invariant mass cut efficiency	107
8.2.6	Systematic uncertainties	108
8.2.7	Total ratio of reconstruction and selection efficiencies	109
8.3	Ratio of trigger efficiencies	109
8.3.1	Ratio of trigger efficiency for $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ and $B^+ \rightarrow J/\psi K^+$	109
8.3.2	Determination of ϵ^{TIS} for the $B_{(s)}^0 \rightarrow h^+ h^-$ normalization	111
8.3.3	Ratio of trigger efficiency for $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow K^+ \pi^-$	112
8.4	Ratio of production fractions	112
8.5	Number of candidates	114
8.5.1	Normalization to $B^+ \rightarrow J/\psi K^+$	114
8.5.2	Normalization to exclusive $B^0 \rightarrow K^+ \pi^-$	115
8.6	Normalization factor	115
8.7	Stability of the yields per pb	118
8.8	Cross-check of the ratio of branching fractions of $B^0 \rightarrow K^+ \pi^-$ to $B^+ \rightarrow J/\psi K^+$	119
8.9	Normalization of exclusive backgrounds	120
9	Combinatorial background estimate	121
10	Time dependent effects	126
10.1	B_s^0 - \bar{B}_s^0 Mixing Effect	126
10.2	Model Dependent Normalisation	126
10.3	Model Dependent BDT PDF	128
11	Expected sensitivity with 3 fb^{-1}	131
12	Results	133
A	Trigger line definitions	139

B	Definition of isolations	139
B.1	Track isolation	139
B.2	B isolation or “CDF isolation”	140
C	Event yields in $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ mass regions	140
	References	140

1 Introduction

Measurements of low-energy processes can provide indirect constraints on particles that are too heavy to be produced directly. This is particularly true for Flavour Changing Neutral Current (FCNC) processes which are highly suppressed in the Standard Model (SM) and can only occur through higher-order diagrams. The SM predictions for the branching fractions of the FCNC decays¹ $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ are [2]

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)^{CP} = 3.35 \pm 0.28 \times 10^{-9} \text{ and} \quad (1)$$

$$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)^{CP} = 1.07 \pm 0.05 \times 10^{-10} \quad (2)$$

which are the CP averaged branching fractions. For $B_s^0 \rightarrow \mu^+ \mu^-$ the prediction given in the previous equation has been updated with respect to the one provided in [2] to use the latest published average [3] for the B_s^0 lifetime ($\tau_{B_s^0} = (1.516 \pm 0.011) \text{ ps}$ instead of $\tau_{B_s^0} = (1.466 \pm 0.031) \text{ ps}$).

Contributions from new processes or new heavy particles can significantly enhance these values. For example, within Minimal Supersymmetric extensions of the SM (MSSM), in the large $\tan \beta$ regime, $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ is found to be approximately proportional to $\tan^6 \beta$ [4–6], where $\tan \beta$ is the ratio of the vacuum expectation values of the two neutral CP -even Higgs fields. The branching fractions could therefore be enhanced by orders of magnitude for large values of $\tan \beta$. Other models such as non minimal flavor violating or Littlest Higgs models as well as those with extra dimensions like Randall Sundrum models predict large effects independent of the value of $\tan \beta$ [7–14].

The current most precise result for $B_s^0 \rightarrow \mu^+ \mu^-$ comes from LHCb, which in November 2012 published the first measurement of the branching fraction of this decay with 3.5σ statistical significance, $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (3.2^{+1.5}_{-1.2}) \times 10^{-9}$ [1]. This result, based on 1.0 fb^{-1} recorded at $\sqrt{s} = 7 \text{ TeV}$ in 2011 and 1.1 fb^{-1} recorded at $\sqrt{s} = 8 \text{ TeV}$ in 2012, supersedes the previous stringent upper limit measured by LHCb on 2011 data only, $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) < 4.5 \times 10^{-9}$ at 95% CL [15].

As discussed in detail in Ref. [16, 17], when comparing the experimental branching fraction to its theoretical expectation, the latter has to be corrected for the finite width difference measured in the B_s^0 system. This leads to the definition of a decay time integrated SM branching fraction

$$\begin{aligned} \mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)^{\text{TH}, \langle t \rangle} &= \frac{1 + y_s A_{\Delta \Gamma}}{1 - y_s^2} \times \mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)^{CP} \\ &=_{SM} \frac{1}{1 - y_s} \times \mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)^{CP} \\ &=_{SM} 3.56 \pm 0.30 \times 10^{-9} \end{aligned} \quad (3)$$

where y_s is given as $y_s = \frac{\Delta \Gamma_s}{2 \Gamma_s} = 0.0615 \pm 0.0085$, from [3]. The above decay time integrated branching fraction can be well compared with an experimental result if the analysis does

¹Inclusion of charged conjugated processes is implied throughout.

not depend on the decay time, which is not the case for $B_s^0 \rightarrow \mu^+ \mu^-$, both for what concern the selection efficiency and the signal BDT pdf. In the previous analysis version [18] the latter effects were not accounted for, and we directly compared the above time integrated branching fraction to our fit result, $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (3.2^{+1.5}_{-1.2}) \times 10^{-9}$. A careful evaluation of all of the lifetime dependent biases has been performed in [19]. The lifetime-dependent biases partly depend on the underlying theoretical model, and in the SM hypothesis they produce a shift in our fit result, which now reads $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (2.7^{+1.3}_{-1.0}) \times 10^{-9}$. Given the definitions of the various corrections [19], this has to be directly compared to the CP averaged branching fraction, $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)^{CP} = 3.35 \pm 0.28 \times 10^{-9}$.

This document presents an analysis of the full LHCb data set, consisting of 1.0 fb^{-1} recorded $\sqrt{s} = 7 \text{ TeV}$ in 2011 and 2.0 fb^{-1} recorded $\sqrt{s} = 8 \text{ TeV}$ in 2012. A part from the last fb^{-1} of 2012, all of the other data have been already published in [1, 15, 20]; these data are here reanalysed using the most up-to-date reconstruction version which has been applied uniformly to the whole data set, and improved signal over background rejection criteria.

2 Analysis strategy

Assuming the branching fractions predicted by the SM, and using the $b\bar{b}$ cross-section measured by LHCb in the pseudo rapidity interval $2 < \eta < 6$ and integrated over all transverse momenta of $\sigma_{b\bar{b}} = 75 \pm 14 \mu\text{b}$ [21], approximately 38 $B_s^0 \rightarrow \mu^+ \mu^-$ and 4 $B^0 \rightarrow \mu^+ \mu^-$ events are expected to be reconstructed and selected (the signal efficiency is $\sim 6\%$ and will be described in detail in Section 8) in the analysed sample embedded in a large background.

The general structure of the analysis is based upon the one described in Ref. [22], which is published in Ref. [15]. First a very efficient selection removes most of the background while keeping very high efficiency 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 [23]. The main improvement with respect to past is the use of a factor of ten larger MC background sample to train and test the algorithm. The old background MC sample was indeed equivalent to 0.5 fb^{-1} only, and was one of the factors limiting the BDT performances.

The current analysis is kept identical with respect to the last publication [18]. The treatment of the physical background is also extended to include the $B^0 \rightarrow \pi^- \mu^+ \nu$ and $B^{0(+)} \rightarrow \pi^{0(+)} \mu^+ \mu^-$ channels. These have a significant effect on the left sideband in the mass, which is used to determine the expected background level.

The probability for a signal or background event to have a given value of the BDT output is extracted from data using $B_{(s)}^0 \rightarrow h^+ h^{(\prime)-}$ candidates (where $h^{(\prime)}$ can be a pion or a kaon) as signal and sideband $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ candidates as background. As in [18], we count the $B_{(s)}^0 \rightarrow h^+ h^{(\prime)-}$ yields by fitting data after a PID-based selection, and including

PID efficiency corrections as event weights in the fit.

The invariant mass line shape of the signals is described by a Crystal Ball function [24] 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 parameters 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: $B^+ \rightarrow J/\psi K^+$ 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 less precise $B_s^0 \rightarrow J/\psi \phi$ normalization channel² is used instead to crosscheck the \sqrt{s} dependence of f_s/f_d via the measurement of the ratio of the yields of $B_s^0 \rightarrow J/\psi \phi$ and $B^+ \rightarrow J/\psi K^+$ decays.

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 analysis.

²The $\mathcal{B}(B_s^0 \rightarrow J/\psi \phi)$ is known to $\sim 30\%$.

3 Data and Monte Carlo Samples

3.1 Data Sample

The results described in this note are obtained using the pp collision data collected by LHCb in years 2011 and 2012 at a center-of-mass energy of $\sqrt{s} = 7\text{ TeV}$ and 8 TeV , respectively. This dataset corresponds to 3 fb^{-1} of integrated luminosity, as shown in Fig. 1.

In 2011 the LHC machine started the operations from 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). 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 $\langle \mu \rangle \sim 1.5$.

During 2012, the data taking conditions were very stable, about the first 100 pb^{-1} were collected while the machine ramped up the luminosity to $4 \times 10^{32}\text{ cm}^{-2}\text{ s}^{-1}$, at which the remaining 2 fb^{-1} were taken. The average number of pp visible interactions per crossing was very stable at $\mu \sim 1.6$, as shown in Fig. 1. All data were recorded with a LHC bunch spacing of 50 ns.

The data belong to the Reco14-Stripping20 campaign, and have been analyzed filtering on the DQ flag, in order to ensure that no runs that are flagged as “subdetector bad” are included in the analysis. For 2011 and 2012 data we have the following integrated luminosities and database tags:

- **2011, Reco14-Stripping20r1, 1018 pb^{-1} :**
 $\text{DQflag} = \text{'dq-20130418'}$, $\text{CondDB} = \text{'cond-20121116'}$ and $\text{DDDB} = \text{'dddb-20120831'}$
- **2012, Reco14-Stripping20, 2028 pb^{-1} :**
 $\text{DQflag} = \text{'dq-20121016'}$, $\text{CondDB} = \text{'cond-20121116'}$ and $\text{DDDB} = \text{'dddb-20120831'}$

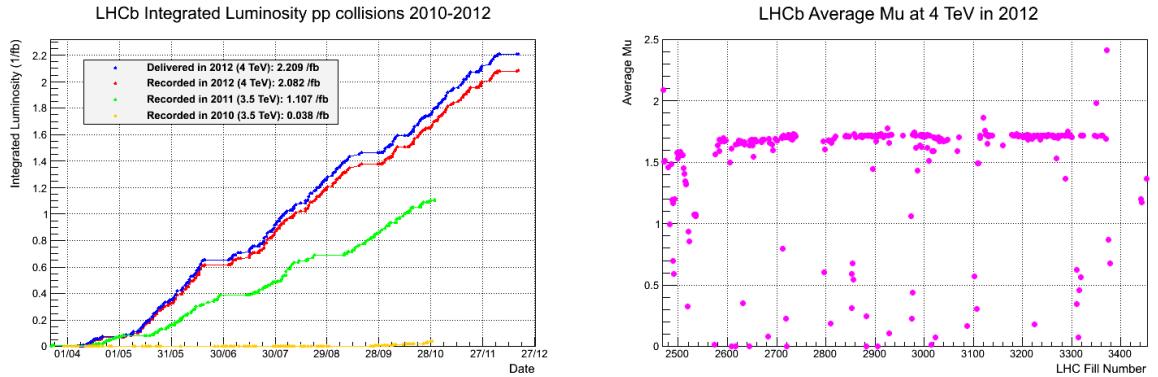


Figure 1: Integrated luminosity in pp collisions in 2010-2012 (left) and average number of visible pp -interactions per crossing in 2012 run as a function of the fill number (right).

'ddb-20120831'

Details about the reconstruction and stripping versions given above are summarized in Ref. [26]. The biggest change in the reconstruction of 2012 (Reco14) vs. 2011 (Reco12) is the new treatment of OT outliers in the track fit. This effectively redefined the meaning of the reduced track fit χ^2 and increased the ghost rate significantly. An updated clone killing strategy was also implemented, which is only of minor significance for this analysis. As consequence of the increased ghost rates, the ghost probability [27] was introduced, which is very powerful in restoring the previous background conditions, especially for $B_{(s)}^0 \rightarrow h^+h^-$.

3.2 Monte Carlo samples

The Monte Carlo samples used in this analysis belong to the 2012 production generated in either May/June 2012 conditions (the inclusive dimuon background) or July/September 2012 conditions (the signal and normalization channels, and the exclusive backgrounds)³. The simulation version used by us is Sim06. The pp interactions have been simulated assuming a beam energy of 4 TeV, an average number of interactions per crossing $\nu = 2.5$, which corresponds to an average number of visible interactions per crossing $\mu = 1.75$, and a $\beta^* = 3.5$ m. All of the relevant MC samples are listed in Table 1, together with the distinctive features and the number of events available.

To save CPU time, cuts are applied at the generation level to enforce particles of interest to be between 10 mrad and 400 mrad. The particles that fulfill this criterion are: the two muons of the signal samples, the two muons of $B_s^0 \rightarrow J/\psi\phi$, $B^+ \rightarrow J/\psi K^+$ control channels and the two hadrons of $B_{(s)}^0 \rightarrow h^+h^-$ decays. For the $b\bar{b} \rightarrow \mu^+\mu^-X$ background sample at least two muons of opposite electric charge are required to be in the acceptance.

All of the samples have been reconstructed with Reco14-Stripping20 model, with the stripping in flagging mode for all channels but for the dimuon inclusive background (EvType 10012009), where the filtering mode has been chosen in order to save disk space. This allowed to reduce by a factor of 10 the amount of events on disk with respect to MC10 production (also listed in Table 1, EvType 10012005), for an integrated luminosity more than 10 times larger: the full MC12 dimuon background sample is equivalent to 7 fb^{-1} . In 2012 dimuon background production (EvType 10012009) we also added further generation cuts with respect to the lower thresholds on muon momenta and invariant mass used in MC10: the requirement for the muons to have opposite charge, an upper threshold on the invariant mass, and cuts on the distance of closest approach and the product of the transverse momenta of the muons (PtProd in the following). All of these cuts, evaluated at the generation level, allow to save a factor of 5 at generation level, while producing a loss of 1.5% of the events passing the stripping. This loss, apparently acceptable, was instead a posteriori the source of a bias in the training of the BDT algorithms, since a

³These production have the following nicknames: Beam4000GeV-MayJune2012-Nu2.5-EmNoCuts and Beam4000GeV-JulSep2012-Nu2.5-EmNoCuts.

sizeable fraction of signal $B_s^0 \rightarrow \mu^+ \mu^-$ events, 20%, was populating the phase space at high values of PtProd left out by our generation cuts. For this reason, a suitable sample of these missing background events has been produced at a second stage (EvType 10012011).

Table 1: Monte Carlo samples used in this analysis. The Monte Carlo production version, the event types, the reconstruction and stripping versions and the number of events per magnet polarity are also shown.

Channel	Monte Carlo production	EvType	Reco/Strip	Evts. per mag. polar.
Signal:				
$B_s^0 \rightarrow \mu^+ \mu^-$	MC2012/Sim06b	13112001	r14/s20flag	1 M
$B_s^0 \rightarrow \mu^+ \mu^-$	MC2012/Sim06b	13112001	r14/s20flag	1 M
Combinatorial background:				
$b\bar{b} \rightarrow \mu\mu X, p > 3 \text{ GeV}/c, M_{\mu\mu} > 4.7 \text{ GeV}/c^2$	MC10/Sim01	10012005	r08/s12flag	50 M
$b\bar{b} \rightarrow \mu^+ \mu^- X, p > 3 \text{ GeV}/c, 4.7 < M_{\mu\mu} < 6.0 \text{ GeV}/c^2, \text{doca} < 0.4 \text{ mm}, 1 < \text{PtProd} < 16(\text{GeV}/c^2)^2$	MC2012/Sim06a	10012009	r13a/s20filt	1.3 M
same as above	MC2012/Sim06b	10012009	r14/s20filt	4 M
$b\bar{b} \rightarrow \mu^+ \mu^- X, p > 3 \text{ GeV}/c, 4.7 < M_{\mu\mu} < 6.0 \text{ GeV}/c^2, \text{doca} < 0.4 \text{ mm}, \text{PtProd} > 16(\text{GeV}/c^2)^2$	MC2012/Sim06b	10012011	r14/s20flag	3.3 M
Control/normalization channels:				
$B_s^0 \rightarrow J/\psi \phi$	MC2012/Sim06b	13144002	r14/s20flag	500k
$B_c^+ \rightarrow J/\psi K^+$	MC2012/Sim06b	12143001	r14/s20flag	500k
$B_c^0 \rightarrow K^+ \pi^-$	MC2012/Sim06b	11102001	r14/s20flag	500k
Exclusive backgrounds				
$B_d^0 \rightarrow \pi^- \mu^+ \nu_\mu$ with $m_{\pi\mu} > 4.5 \text{ GeV}/c^2$	MC2012/Sim06b	11512012	r14/s20flag	3 M
$B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ with $m_{K\mu} > 4.5 \text{ GeV}/c^2$	MC2012/Sim06b	13512011	r14/s20flag	3 M
$B_c^+ \rightarrow \pi^+ \mu^+ \mu^-$ with DiLeptonInAcc.	MC2012/Sim06b	12113023	r14/s20flag	500 k
$B_c^0 \rightarrow \pi^0 \mu^+ \mu^-$ with DiLeptonInAcc.	MC2012/Sim06b	11112401	r14/s20flag	500 k
$\Lambda_b^0 \rightarrow p \mu^- \bar{\nu}_\mu$ with $m_{p\mu} > 4.5 \text{ GeV}/c^2$	MC2012/Sim06b	15512012	r14/s20flag	500 k
$B_c^+ \rightarrow J/\psi (\rightarrow \mu^+ \mu^-) \mu^+ \nu_\mu$ with $m_{\mu\mu} > 4.5 \text{ GeV}/c^2$	MC2012/Sim06b	14543009	r14/s20flag	500 k
$\Lambda_b^0 \rightarrow p \pi^-$	MC2012/Sim06b	15502011	r14/s20flag	1M
$\Lambda_b^0 \rightarrow p K^-$	MC2012/Sim06b	15502001	r14/s20flag	1M

For the signal modes and the control channels, the trigger has been emulated in *flagging mode* with the configuration that has been more extensively used during the 2012 data taking, *Trigger Configuration Key* (TCK) 0x40990042 (see Sect. 3.3).

A smearing procedure to the first state of the tracks in simulated events has been applied in order to reproduce the impact parameter resolution measured in data. This will be discussed in Section 8.2.1.

3.3 Trigger

The LHCb trigger scheme is described in [28]. A detailed description of the trigger lines relevant for the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis is reported in [29]. In this Section we summarize the main characteristics of the trigger used for the data used in this analysis.

In LHCb the trigger configuration information is embedded into an hexadecimal word called *Trigger Configuration Key* (TCK) that identifies the set of trigger decisions, the algorithms run in the trigger and the cuts applied to trigger the events in a given run. Table 2 shows the different TCKs used for the data taking with the corresponding integrated luminosity and magnet polarity.

Table 2: TCKs used with the corresponding integrated luminosity (approximately) and magnet polarity.

TCK	Integrated Luminosity	Magnet Polarity
2012		
0x008c0040	70 pb ⁻¹	Down
0x0094003D	280 pb ⁻¹	Up & Down
0x0097003D	280 pb ⁻¹	Up & Down
0x00990042	620 pb ⁻¹	Up & Down
0x00990044	140 pb ⁻¹	Up & Down
0x009F0045	20 pb ⁻¹	Up
0x00A10044	15 pb ⁻¹	Up
0x00A10045	75 pb ⁻¹	Up
0x00A30044	320 pb ⁻¹	Up & Down
0x00A30046	30 pb ⁻¹	Up & Down
0x00A90046	60 pb ⁻¹	Down
0x00AB0046	55 pb ⁻¹	Down
0x00AC0046	120 pb ⁻¹	Up & Down
2011		
0x5a0032	40 pb ⁻¹	Up
0x5a0032	30 pb ⁻¹	Down
0x6d0032	100 pb ⁻¹	Down
0x730035	200 pb ⁻¹	Up & Down
0x760037	300 pb ⁻¹	Up & Down
0x790037	40 pb ⁻¹	Up
0x790038	360 pb ⁻¹	Up & Down

In the TCKs used in the 2012 data sample, the most significant difference for the signal candidates is the change in L0 thresholds between 0x008c0040 and the other TCKs: in 0x008c0040, the single muon requirement on p_T is 1.48 GeV whereas in the rest of the

dataset, it is 1.76 GeV. The dimuon threshold was also increased from 1.3 GeV to 1.6 GeV. The TCKs 0x0094003D and 0x0097003D differ only in technical aspects. Finally the TCK 0x00990042 has a looser p_T threshold in the generic B selection of HLT1 and a prescale of a factor 5 on the prompt J/ψ trigger in HLT2. The former is of negligible importance for the muon triggered events analyzed here whereas the latter needs to be taken into account for the normalization channels. The changes in 2012 after TCK 0x00990042 are mostly technical nature or concern the adjustment of Charm thresholds. They are transparent to the present analysis.

The dataset used in 2011 is identical to the one discussed in [22], the thresholds applied for this analysis differ only slightly to the ones discussed here for 2012 and will not be repeated here. For more details, please refer to the extended discussion in Ref. [22].

The HLT1 main trigger lines have been described in [29] and are not repeated here. The most important lines for this analysis are the single muon trigger (`Hlt1TrackMuon`) and the two dimuon triggers (`Hlt1DiMuonLowMass` and `Hlt1DiMuonHighMass`).

The main HLT2 trigger for the channels with dimuon in the final state are described in [29]. The channels containing J/ψ in the final state have been triggered in the first 630 pb^{-1} by the `Hlt2DiMuonJPsi` trigger, which requires two identified muons in a $120 \text{ MeV}/c^2$ mass window around the J/ψ mass. This line was then prescaled by a factor five for the remaining 470 pb^{-1} , and the channels containing J/ψ in the final state were mostly triggered by the `Hlt2DiMuonDetached` line, where a soft cut on the significance of the vertex separation ($DLS > 3$) is additionally required.

The $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ signal candidates are selected with an equivalent trigger selection, `Hlt2DiMuonB` but requiring their invariant mass to be above $4.7 \text{ GeV}/c^2$. A small fraction of additional J/ψ and $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ events are selected by the *topological lines*, where the B -decay is only partially reconstructed and in some lines a muon ID requirement is made on one track.

The $B_{(s)}^0 \rightarrow h^+ h^-$ hadronic channels used for calibration and/or normalization are dominantly selected by the `Hlt2Topo2Body` trigger, a generic $B \rightarrow hh + X$ selection and by the exclusive `Hlt2B2hh` selection [30].

Global event cuts (GEC) were introduced at L0, HLT1 and HLT2 level to reduce the processing time of the HLT. The GEC applied during 2012 run are based on cuts on the number of IT, OT and SPD hits and are listed in Table 3. The SPD hits thresholds depend on the L0 trigger line, it is 900 for L0DiMuon triggers and 600 for all other trigger decisions. The dominant L0 trigger lines are L0Muon and L0DiMuon.

Table 3: Summary of global event cuts (GEC) used in 2012.

trigger level	cut	value
L0	SPD hits	900/600
HLT1	Velo hits	10 000
HLT1	IT hits	3 000
HLT1	OT hits	15 000
HLT2	Velo tracks	350

4 Selection

The aim of the selection is twofold:

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 7;
2. treat the signal and control/normalization channels in a similar manner in order to minimize the systematic uncertainties in the computation of the normalization factors.

In Table 4 the cuts used to select the $B_{(s)}^0 \rightarrow \mu^+\mu^-$, $B_{(s)}^0 \rightarrow h^+h^-$ and $B^+ \rightarrow J/\psi K^+$ decays are listed. The selection criteria are unchanged compared to the previous analysis [18].

The selection on the ghost probability has an efficiency of above 99.5% for $B_{(s)}^0 \rightarrow \mu^+\mu^-$ signal candidates and removes a significant part of the background tracks, specially in the $B_{(s)}^0 \rightarrow h^+h^-$ control sample. Its effect is studied in detail in Sec. 8.

The fiducial cuts applied on muon momenta and on the B candidate lifetime to reject the unphysical signal candidates are unchanged, they have an efficiency of $\epsilon = 99.9\%$ on Monte Carlo $B_s^0 \rightarrow \mu^+\mu^-$ signal events.

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 $p_T(B) > 500 \text{ MeV}/c$ as in the previous analysis version. For a detailed discussion about the exclusive dimuon production see Ref. [31], Section 5.1.

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 briefly reminded below, for further discussion refer to Section 5.1 in Ref. [22].

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. The selection for the $B^+ \rightarrow J/\psi K^+$ normalization channel was, as in the previous analyses, optimized to be as close as possible to the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ signal selection such that possible systematic uncertainties cancel in the ratio of selection efficiencies.

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, irrespective of the fact that 2/3 of the data have been unblinded already in a the previous reconstruction version.

Table 4: Selection for $B_{(s)}^0 \rightarrow \mu^+ \mu^-$, $B_{(s)}^0 \rightarrow h^+ h^-$ and $B^+ \rightarrow J/\psi K^+$ channels; DOCA is the instance of closest approach between the two tracks, VDS is the secondary vertex flight distance significance, and DLL is the combined PID likelihood to discriminate different particle hypotheses.

Cut	applied on	value	applied on	value
		$B_s^0 \rightarrow \mu^+ \mu^-$ and $B_{(s)}^0 \rightarrow h^+ h^-$		$B^+ \rightarrow J/\psi K^+$
track χ^2/ndf ghost prob DOCA $IP\chi^2$ p_T p ISMUON	μ / h	<3 < 0.3 <0.3 mm >25 > 0.25 and < 40 GeV/c <500 GeV/c true	μ / h	< 3 <0.3 mm >25 > 0.25 and < 40 GeV/c <500 GeV/c true
vertex χ^2 VDS ΔM	$B_{(s)}$	<9 > 15 $ M(hh, \mu\mu) - m_B < 60 \text{ MeV}/c^2$	J/ψ	<9 > 15 $ M(\mu\mu) - m_{J/\psi} < 60 \text{ MeV}/c^2$
$IP\chi^2$ t BDTS $DLL(K - \pi)$ $DLL(\mu - \pi)$ ΔM	$B_{(s)}$	< 25 < $9 \cdot \tau(B_s^0)$ > 0.05 < 10 > -5	B^+	< 25 < $9 \cdot \tau(B_s^0)$ > 0.05 $ M(J/\psi K) - m_B < 100 \text{ MeV}/c^2$
$p_T (B_s^0)$	$B_s^0 \rightarrow \mu^+ \mu^-$ $B_{(s)}^0 \rightarrow h^+ h^-$	> 0.5 GeV/c		

4.1 The BDTS discriminant

The BDTS discriminant has been found the most efficient way to further reduce background sample while keeping high efficiency on the signal. The variables entering the BDTS are:

- the impact parameter (IP(B)) and impact parameter χ^2 ($IP\chi^2(B)$) of the B candidate;
- the χ^2 of the secondary vertex (VCHI2);
- the angle between the direction of the momentum of the B candidate and the direction defined by the secondary and the primary vertices (DIRA);
- the minimum distance between the two daughter tracks (DOCA);
- the minimum impact parameter of the muons with respect to any primary vertex ($minIP(\mu)$).

Only the variables DOCA and IP(B) are in common with the BDT used in the limit computation (see Section 7).

In order to minimize the systematic uncertainty in the normalization factors, the same BDTS cut is also applied to the three normalization channels. For the $B^+ \rightarrow J/\psi K^+$ mode, 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 $B_s^0 \rightarrow \mu^+ \mu^-$, $B_{(s)}^0 \rightarrow h^+ h^-$ and $B^+ \rightarrow J/\psi K^+$, resulting in a very similar efficiency for signal and normalization channels. The efficiency ratios for unsmeared, smeared and oversmeared $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^+ \rightarrow J/\psi K^+$ simulated samples (see Sec. 8.2.1) agree within 0.4%.

The BDTS cut has not been optimized since Ref. [22] where it was chosen by looking at the combined performance of the BDTS and the BDT; a BDTS cut $BDTS > 0.05$ was found to be optimal.

4.2 Event matching between reco13 and reco14

One delicate point to investigate, is the matching of the selected candidates when passing from the previous reconstruction version (reco12 and reco13) to the one used in this analysis, reco14. To this purpose, a sample of 102 events in the stripping 19 data has been selected in the dimuon mass sidebands⁴, passing all of the analysis cuts and satisfying $BDT > 0.5$. By matching the event and run numbers, 86 events out of 102 have been found in the stripping 20 data. The 16 missing events have been re-reconstructed with reco14 starting from the stripping 19 files, so that their fate could be determined before passing the stripping selection:

- 10 out of 16 dimuon candidates were found to be made with the same couple of tracks as of reco13: 2 of them passed all cuts, but were displaced into the mass blind region, 6 failed the stripping cuts after reco14;
- 6 out of 16 had only one track in common with reco13 selection, and failed the stripping cuts after reco14.

Nothing pathological has been found in the above analysis exercise, and our conclusion is that we do expect on background events a jitter of about 15% due to the change of the reconstruction.

⁴This check was made prior to signal region unblinding

5 Muon Identification

The muon identification procedure (muonID) is a key ingredient of the analysis. For this update of the measurement the same procedure used in 2011 measurement has been applied on 2012 data. For all the details we refer to the 2011 and 2012 notes (Refs. [32] and [18] respectively) and references therein; here we summarize only the muID results relevant for the 3 fb^{-1} analysis update.

Data Samples - We use $JpsiFromBNoPIDNoMip$ and $JpsiKFromBNoPIDNoMip$ calibration lines, with Reco14 and Stripping 20r1 (for the 2011 1 fb^{-1}) or Stripping 20 (for the 2012 2 fb^{-1}). The effect of the different reconstruction is summarized in Fig. 2, where the geometrical acceptance (left panel) and the efficiency (right panel) of the muon system are shown for different reconstruction/stripping versions. The "geometrical acceptance" is the efficiency for a LongTrack to be linearly extrapolated in the Muon System. The "efficiency" tout-court is the efficiency for a track in the MuonSystem to be identified as a muon by the MuonID algorithm. The muon performance is quite stable along the data taking: within 0.5% for the acceptance and within 1% for the efficiency. The stability between different Reco/Stripping versions is very good. In any case, we decided to evaluate again the muID in Stripping 20(r1) data to take into account of kinematic differences possibly induced by the changing on the tracking.

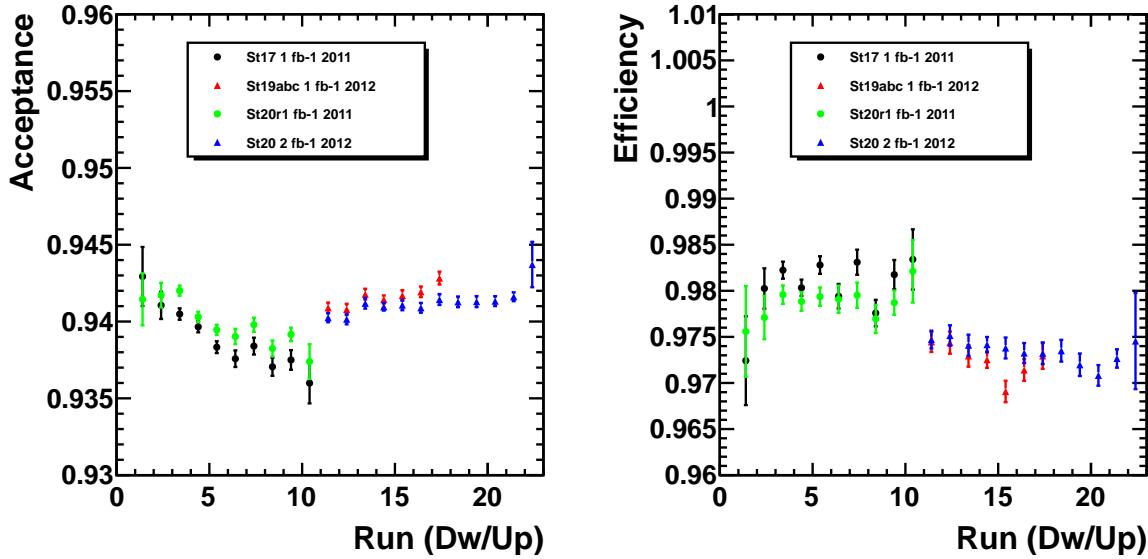


Figure 2: Acceptance (left) and efficiency (right) with TIS-probe trigger unbias, measured as a function of homogeneous (same magnet polarity) run groups along the 2011 and 2012 data taking. Different Reco/Stripping versions are superimposed.

In $B_s^0 \rightarrow \mu^+ \mu^-$ analysis we use the geometrical acceptance to correct the $B_{(s)}^0 \rightarrow h^+ h^-$ event selection and the product acceptance times efficiency (shown in the left panel of Fig. 3) to correct the $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^+ \rightarrow J/\psi K^+$ event selection.

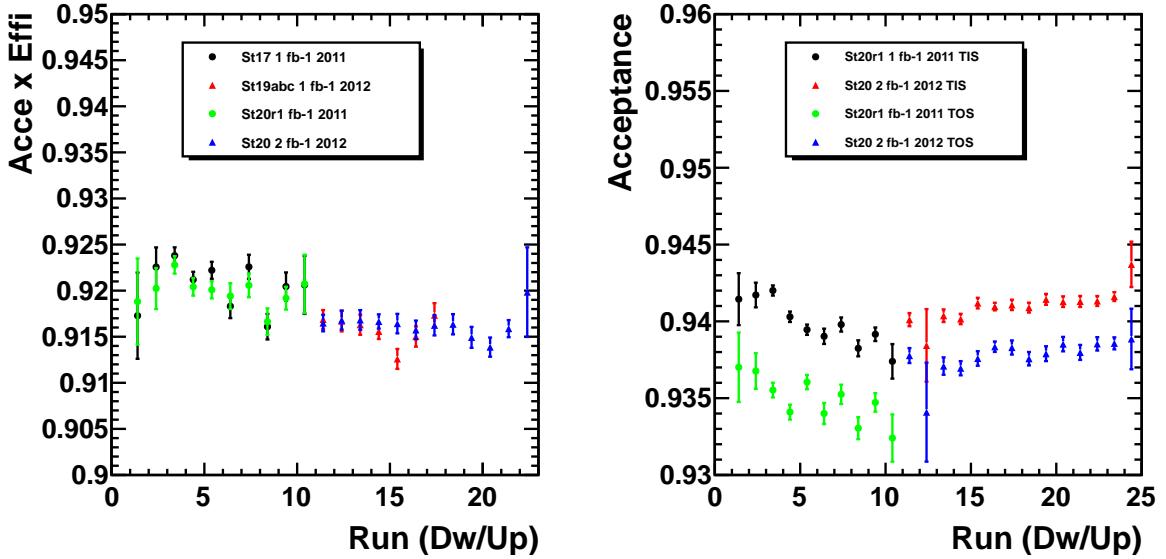


Figure 3: **Left:** Acceptance times efficiency with TIS-probe trigger unbias, measured as a function of homogeneous (same magnet polarity) run groups along the 2011 and 2012 data taking. Different Reco/Stripping versions are superimposed. **Right:** Acceptance with TIS-probe (black and red points) or TOS-tag (green and blue points) trigger unbias applied to the same Reco14/Stripping20 version of 2011 and 2012 data.

Both acceptance and efficiency, show a run-dependent variation. For the efficiency we can use general and reasonable arguments related to the hardware efficiency, even though a systematic study has still to be done. For the acceptance the “culprits” might be the tracking and/or the trigger. Since the seed of the muonID (`MuonAcc + isMuon`) is a LongTrack, changes in the tracking could reflect in the muonID (also in the acceptance). But the comparison done in Fig. 2 between different Reco versions, show that the effect of changing the tracking is at the per mil level on average, and doesn’t explain the run-dependent variation, which remains the same. The trigger enters in the muonID evaluation because of the need of trigger unbias of the data: we use TIS-probe for our efficiency determination and TOS-tag to asses the systematic effect due to this choice. Points on Fig. 2 are all with TIS-probe unbias. Instead, in the right panel of Fig. 3 we superimpose the TIS-probe and TOS-tag points for all 2011 and 2012 with Reco 14 and stripping 20(r1). The average difference between the two is 0.6% in the 2011 and 0.3% in 2012 (both compatible with the spread visible in the left panel of Fig. 2) and with some dependence on the kinematics: using the usual (p, p_T) binning, the bin-by-bin difference can be as high as 2%. In the analysis we account for this TIS-TOS unbias difference, propagating both choices for the muID unbias till to the end and quoting a systematic effect for this: it is included in the values in Table 44. For the run-dependent variation, the plot on Fig. 3 shows that the trend is the same with both trigger unbias choices, and that the TOS-tag amplifies the difference present at each polarity change in 2011. The same difference seems

absent in 2012 (we could invoke the change of beam crossing angle between 2011 and 2012 data taking, in principle done to have “more symmetric” data, but a systematic study on this respect for the muonID has still to be done). So, a possible answer is that the trigger unbias could cause a run-dependent variation of the acceptance. Finally, the spread in the acceptance has a negligible impact on the analysis (0.5% variation for a correction which has a marginal numerical impact on the results). In the previous analysis versions, we used for the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ measurement the efficiency measured in $B \rightarrow J/\psi$ -inclusive events, while the one from $B \rightarrow J/\psi K$ events was used to assess the systematic error. Now, to profit of the new $B \rightarrow J/\psi K$ MC12 sample, we use only the efficiency evaluated in $B \rightarrow J/\psi K$ events. The $B \rightarrow J/\psi$ -inclusive events were used to obtain the PID cut efficiency (see later).

Starting from Ref. [32], we found that the request $DLL(K - \pi) < 10 \&& DLL(\mu - \pi) > -5$ (PID) can reduce by a factor ~ 5 the rate of $B_{(s)}^0 \rightarrow h^+ h^-$ events with double decays in flight. We keep this request also in the present version of the analysis. As in the past, the efficiency of the PID cut has been measured on data using the $B \rightarrow J/\psi$ -inclusive events (*JpsiFromBnoPIDnoMIP* stripping line data). In the left panel of Fig. 4, the efficiency of the PID cut is shown as a function of the data taking period and for the different Reco/Stripping versions. Since the $DLL(K - \pi)$ and $DLL(\mu - \pi)$ variables depend also on RICH and Calorimeter, the effect of the PID cut can considerably vary as a function of the time (e.g. RICH calibration) and of the Reco/Stripping version; for this reason, it is important to measure the efficiency of the PID cut using the same data set and the same version of the data used for the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis.

Folding the measured PID cut efficiency for single muons into the spectrum of selected $B_s^0 \rightarrow \mu^+ \mu^-$ MC events, allows to evaluate the signal efficiency. The correction applied to the MC signal due to the PID cut efficiency can be ideally divided into two steps. The first one is the average correction which is $0.9749 \pm 0.0001_{stat} \pm 0.0027_{syst}$ from 1 fb^{-1} 2011 data and $0.9798 \pm 0.0001_{stat} \pm 0.0028_{syst}$ from 2 fb^{-1} 2012 data. The systematic error includes the effect of the trigger and of the use of a different efficiency (measured in *JpsiKfromB* sample). The second step is to check if this PID correction introduces some distortion on the Invariant Mass and/or the BDT distributions. Already in the 2011 data analysis we checked for these effects. The average PID cut efficiency in each bin of BDT or invariant mass is a weighted average, calculated using the PID cut efficiency measured on data as a function of muon momentum and MC muon momentum spectrum in each bin as weight. For the BDT we decided to add a “shape correction” shown in the right panel of Fig. 5 (from 2012 data): the correction is smaller in the high BDT bins where the muon momenta are higher. The slope of the efficiency as a function of the BDT output has been applied as correction (separately for 2011 and 2012 data) to the BDT PDF extracted from the $B_{(s)}^0 \rightarrow h^+ h^-$ sample, as discussed in Section 7.2.4.

The shape of the PID correction as a function of the mass is shown in the left panel of Fig. 5 (from 2012 data). In this case is not obvious to explain the behavior of the corrections as a function of the mass: each MC signal event is weighted choosing two PID cut efficiency values following the momentum of the two muons; the momentum distribution of the two muons can be different in each mass bin, picking up different PID

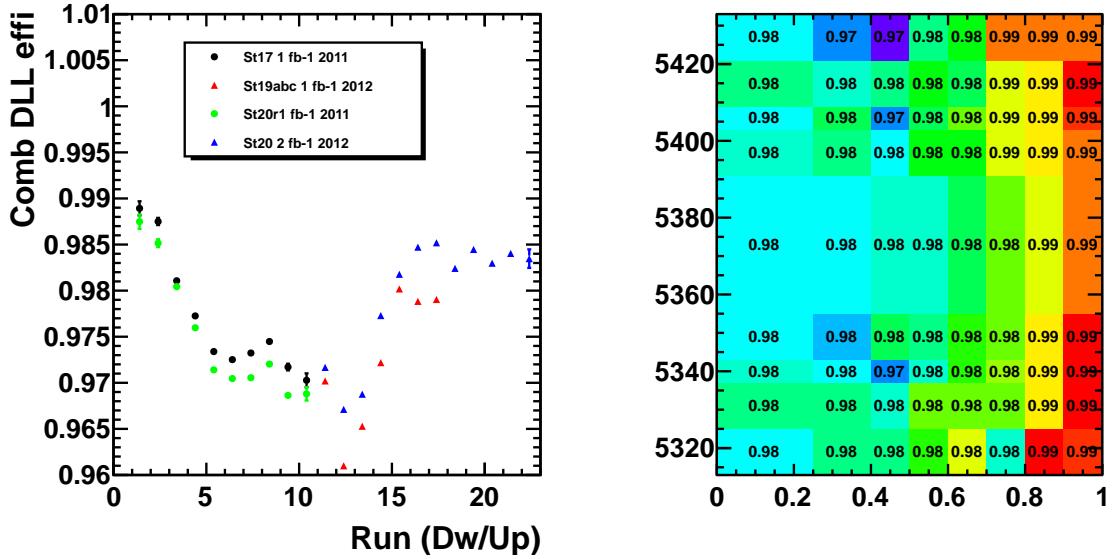


Figure 4: Left, the PID cut efficiency measured as a function of homogeneous (same magnet polarity) run groups along the 2011 and 2012 data taking. Different Reco/Stripping versions are superimposed. Right, PID cut efficiency as a function of the $B_s^0 \rightarrow \mu^+ \mu^-$ invariant mass and the BDT (same binning used in the signal extraction)

correction. The tiny effect of the DLL cuts on the mass resolution (the corrections shown in the left panel of Fig. 5 are well within a spread of 0.25%) has been studied by fitting the invariant mass of $B_s^0 \rightarrow \mu^+ \mu^-$ simulated sample without and with a DLL cut, and has been found negligible (well within the total uncertainty in the mass resolution itself).

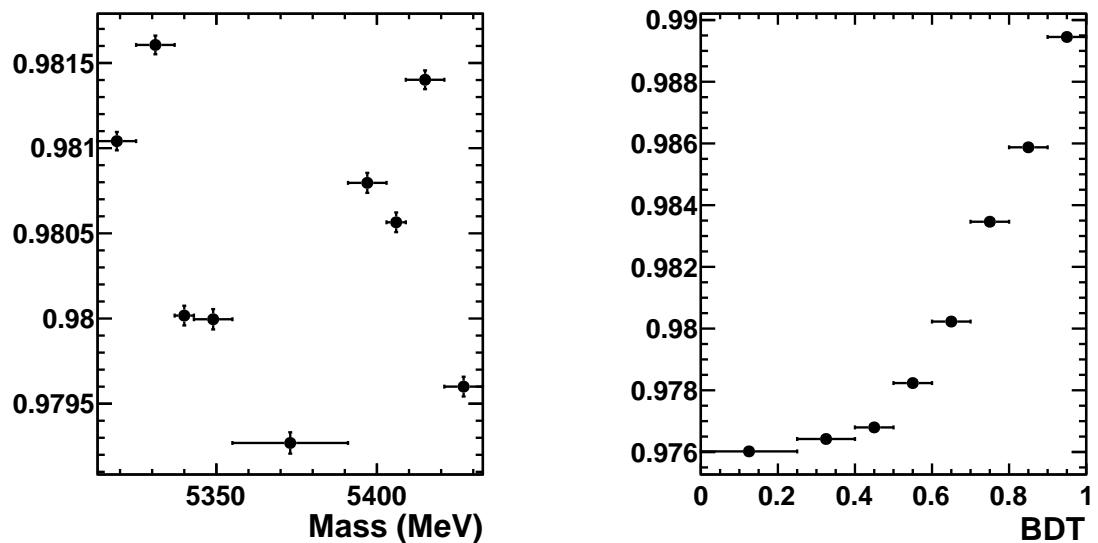


Figure 5: Efficiency of the PID cut when (2012 data) when folded into the p spectrum of selected $B_s^0 \rightarrow \mu^+ \mu^-$ events projected onto the $B_s^0 \rightarrow \mu^+ \mu^-$ invariant mass (left panel) and BDT (right panel) axis (see text for details).

6 Misidentification rates and peaking backgrounds

In this section we discuss the evaluation of various peaking background sources for the search of $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays. Some of these background sources have one or two real muons in the final state, and in the former case one of the hadrons in the decay is misidentified as a muon to mimic a $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ signal. Background from $h^+ h^-$ decays where h represents a k or a π requires both hadrons to be mis-identified.

In the following we discuss the procedure and results for the computation of various $h \rightarrow \mu$ mis-identification probabilities, i.e. the probabilities that each species of hadrons survives the identification cuts, which are the `isMuon` requirement and the DLL⁵ cuts: `DLLK < 10 & DLLmu > -5`. This is followed by their use in the determination of the expected $B \rightarrow h^+ h^-$ background. We then discuss the evaluation of the background expected from semi-leptonic B decays and decays with two real muons in the final state.

6.1 Pion and kaon misidentification probabilities

This section has been updated: it now describes the new misID efficiency evaluation (both method and results) chosen after the study done in the first days of July 2013. Previous methods and choices are documented in the LHCb-ANA-2013-032_v2.2 version of the note.

The pion and kaon misidentification probabilities, $\epsilon(\pi \rightarrow \mu)$ and $\epsilon(K \rightarrow \mu)$, have been measured using a tag a probe technique in a sample of $D^0 \rightarrow K\pi$ from the $D^{*\pm} \rightarrow D^0\pi^\pm$ decays extracted from the line `NoPIDstarWithD02RSKPiLine` of the Stripping20 and Stripping20r1. The samples used here correspond to 2 fb^{-1} and $\sim 800 \text{ pb}^{-1}$ respectively.

To cancel possible biases arising from the trigger, the probe track (kaon or pion) is required to be TIS with respect the L0Global and the Hlt1Phys lines while no requirement on the HLT2 level has been set. This does not bias the results because anyhow these decays are selected by HLT2 lines without PID conditions.

A cut of <0.3 is applied on the `ghostProbability` variable as advised by the tracking group. To further clean the sample, a cut on the D^{*+} and D^0 mass difference, ΔM , has been applied: $144.5 < \Delta M < 146.5$. Finally the track of the D^0 decays that is not used to evaluate the fake rate (the *tag* track), is required to be well identified with a cut on the $DLL(K - \pi)$ ⁶.

The $\pi \rightarrow \mu$ and $K \rightarrow \mu$ fake rates (misID rates), are evaluated by fitting the D^0 mass distribution and extracting the total number of signal and background events in a suitable mass window around the D^0 peak (general information about the method used to measure the efficiency can be found here [33] and here [18]). The signal is fitted with the sum of a

⁵The symbol *DLL* refers to the combined *DLL* unless otherwise stated.

⁶The 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.

Crystal Ball and a Gaussian functions (with the mean constrained to the same value) and the background is fitted with a polynomial. The misID is measured in bins of p and p_T of the probe track.

The pion and kaon fake rates as a function of the momentum and for the four different p_T ranges are shown in Fig. 6 for 2012 data (Stripping20) and in Fig. 7 for 2011 data (Stripping20r1) and the numbers are reported in Tables 5 to 8. The results reported here are the average of the two magnet polarities. Some bins in this kinematic phasespace are empty because there were not enough events in the calibration sample to perform the fit and extract a reliable number for the mis-identification efficiency. In converting the total mis-identification efficiency into an expected number of $B \rightarrow h^+h^-$ events (see Sect. 6.3), when we find a track from $B \rightarrow h^+h^-$ lying in one of the empty bins, we assign to it the mis-identification rate of the next momentum bin, in the same p_T bin.

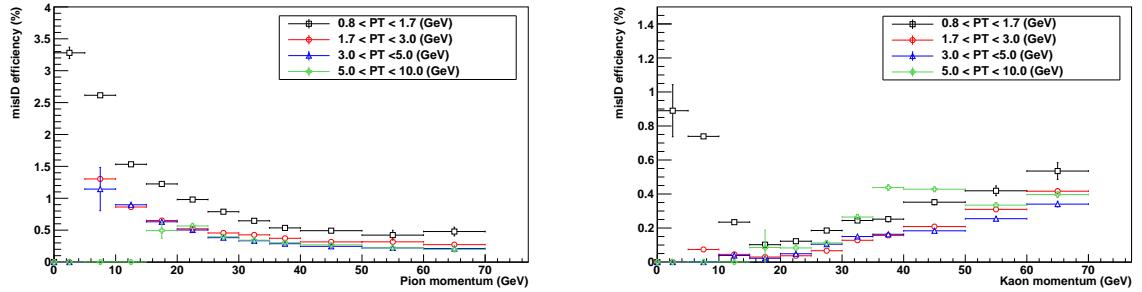


Figure 6: **2012 data Stripping20:** $\pi \rightarrow \mu$ (left) and $K \rightarrow \mu$ (right) fake rates as a function of the momentum of the probe track for four different p_T ranges.

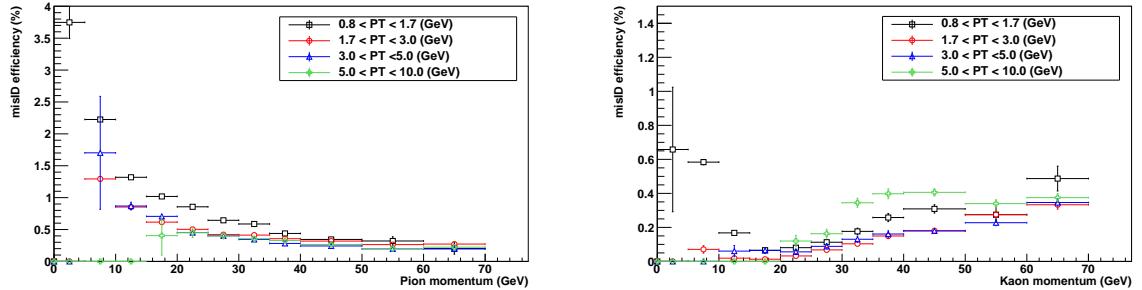


Figure 7: **2011 data Stripping20r1:** $\pi \rightarrow \mu$ (left) and $K \rightarrow \mu$ (right) fake rates as a function of the momentum of the probe track for four different p_T ranges.

To check for the stability and the reliability of the misID rates evaluation, these have been measured also by fitting the D^{*+} mass peak instead of the D^0 mass one, by changing

the background description from polynomial to exponential, and by varying the ΔM cut (no cut and $144 < \Delta M < 147$).

The difference between the D^0 fit (polynomial background) with no δM cut and $144.5 < \Delta M < 146.5$ has been used to assess the systematic error.

Event multiplicity may be different in the signal and calibration samples (trigger, trigger unbias, ...) A good variable to parameterize the event multiplicity that can spoil the RICH information used in the PID variables DLLk and DLLmu, is the number of tracks in the event, in particular the number of long tracks, nLong, being a “long track” the seed for the isMuon algorithm. To check for a “multiplicity effect” we measured the $\varepsilon(K \rightarrow \mu)$ misID rates, for the muonID+PID cut request, in the control sample in bins of nLong defined as $n\text{Long} < 40$, $40 < n\text{Long} < 60$, $60 < n\text{Long} < 80$, and $n\text{Long} > 80$. Then we calculated the average misID weighted according to the nLong distribution of $B^+ \rightarrow J/\psi K^+$ (signal proxy) events. We finally compared the weighted average misID rate with the one measured in the control sample with no nLong binning. The difference between the two misId rate evaluations appear negligible within the errors.

Table 5: Stripping20, 2 fb^{-1} , $K \rightarrow \mu$ rate (%), `isMuon && DLL`

Momentum (GeV)	$0.8 < PT < 1.7$	$1.7 < PT < 3.0$	$3.0 < PT < 5.0$	$5.0 < PT < 10.0$
$0.0 < P < 5.0$	0.89 ± 0.15	0 ± 0	0 ± 0	0 ± 0
$5.0 < P < 10.0$	0.739 ± 0.0094	0.0741 ± 0.012	0 ± 0	0 ± 0
$10.0 < P < 15.0$	0.235 ± 0.0055	0.045 ± 0.006	0.0386 ± 0.016	0 ± 0
$15.0 < P < 20.0$	0.102 ± 0.0056	0.0295 ± 0.0054	0.0221 ± 0.01	0.0856 ± 0.1
$20.0 < P < 15.0$	0.123 ± 0.0066	0.037 ± 0.0057	0.0488 ± 0.009	0.083 ± 0.016
$25.0 < P < 30.0$	0.185 ± 0.009	0.0666 ± 0.0068	0.105 ± 0.0091	0.112 ± 0.014
$30.0 < P < 35.0$	0.244 ± 0.012	0.128 ± 0.0075	0.15 ± 0.0097	0.265 ± 0.016
$35.0 < P < 40.0$	0.252 ± 0.015	0.157 ± 0.0087	0.163 ± 0.011	0.438 ± 0.017
$40.0 < P < 50.0$	0.352 ± 0.016	0.209 ± 0.0077	0.184 ± 0.008	0.428 ± 0.012
$50.0 < P < 60.0$	0.419 ± 0.029	0.31 ± 0.011	0.255 ± 0.01	0.335 ± 0.013
$60.0 < P < 70.0$	0.535 ± 0.049	0.417 ± 0.016	0.341 ± 0.017	0.397 ± 0.018

 Table 6: Stripping20, $\pi \rightarrow \mu$ rate (%), `isMuon && DLL`

Momentum (GeV)	$0.8 < PT < 1.7$	$1.7 < PT < 3.0$	$3.0 < PT < 5.0$	$5.0 < PT < 10.0$
$0.0 < P < 5.0$	3.28 ± 0.089	0 ± 0	0 ± 0	0 ± 0
$5.0 < P < 10.0$	2.61 ± 0.013	1.3 ± 0.022	1.14 ± 0.34	0 ± 0
$10.0 < P < 15.0$	1.53 ± 0.0096	0.863 ± 0.01	0.898 ± 0.033	0 ± 0
$15.0 < P < 20.0$	1.22 ± 0.01	0.651 ± 0.0083	0.631 ± 0.017	0.494 ± 0.12
$20.0 < P < 15.0$	0.98 ± 0.011	0.521 ± 0.0082	0.503 ± 0.013	0.566 ± 0.043
$25.0 < P < 30.0$	0.789 ± 0.013	0.456 ± 0.0087	0.382 ± 0.012	0.397 ± 0.026
$30.0 < P < 35.0$	0.647 ± 0.016	0.426 ± 0.0094	0.334 ± 0.012	0.343 ± 0.021
$35.0 < P < 40.0$	0.535 ± 0.021	0.373 ± 0.011	0.287 ± 0.012	0.303 ± 0.019
$40.0 < P < 50.0$	0.492 ± 0.021	0.316 ± 0.01	0.246 ± 0.0095	0.269 ± 0.012
$50.0 < P < 60.0$	0.422 ± 0.076	0.317 ± 0.014	0.222 ± 0.012	0.224 ± 0.013
$60.0 < P < 70.0$	0.479 ± 0.068	0.273 ± 0.02	0.212 ± 0.014	0.198 ± 0.016

Table 7: Stripping20r1, $K \rightarrow \mu$ rate (%), `isMuon && DLL`

Momentum (GeV)	$0.8 < PT < 1.7$	$1.7 < PT < 3.0$	$3.0 < PT < 5.0$	$5.0 < PT < 10.0$
$0.0 < P < 5.0$	0.658 ± 0.37	0 ± 0	0 ± 0	0 ± 0
$5.0 < P < 10.0$	0.584 ± 0.016	0.0708 ± 0.022	0 ± 0	0 ± 0
$10.0 < P < 15.0$	0.168 ± 0.0091	0.0191 ± 0.012	0.0606 ± 0.033	0 ± 0
$15.0 < P < 20.0$	0.0658 ± 0.0093	0.013 ± 0.01	0.0642 ± 0.016	0 ± 0
$20.0 < P < 15.0$	0.0808 ± 0.011	0.0322 ± 0.01	0.0562 ± 0.016	0.12 ± 0.034
$25.0 < P < 30.0$	0.113 ± 0.014	0.0677 ± 0.011	0.0886 ± 0.016	0.163 ± 0.026
$30.0 < P < 35.0$	0.177 ± 0.019	0.103 ± 0.013	0.13 ± 0.018	0.345 ± 0.027
$35.0 < P < 40.0$	0.258 ± 0.024	0.15 ± 0.016	0.161 ± 0.019	0.398 ± 0.029
$40.0 < P < 50.0$	0.309 ± 0.025	0.181 ± 0.013	0.179 ± 0.015	0.406 ± 0.021
$50.0 < P < 60.0$	0.275 ± 0.045	0.275 ± 0.019	0.228 ± 0.017	0.34 ± 0.023
$60.0 < P < 70.0$	0.487 ± 0.073	0.333 ± 0.026	0.346 ± 0.021	0.375 ± 0.025

 Table 8: Stripping20r1, $\pi \rightarrow \mu$ rate (%), `isMuon && DLL`

Momentum (GeV)	$0.8 < PT < 1.7$	$1.7 < PT < 3.0$	$3.0 < PT < 5.0$	$5.0 < PT < 10.0$
$0.0 < P < 5.0$	3.75 ± 0.25	0 ± 0	0 ± 0	0 ± 0
$5.0 < P < 10.0$	2.23 ± 0.021	1.29 ± 0.039	1.7 ± 0.89	0 ± 0
$10.0 < P < 15.0$	1.32 ± 0.015	0.853 ± 0.017	0.868 ± 0.06	0 ± 0
$15.0 < P < 20.0$	1.02 ± 0.016	0.616 ± 0.014	0.706 ± 0.03	0.404 ± 0.31
$20.0 < P < 15.0$	0.856 ± 0.018	0.503 ± 0.014	0.452 ± 0.022	0.451 ± 0.075
$25.0 < P < 30.0$	0.645 ± 0.02	0.42 ± 0.014	0.402 ± 0.02	0.399 ± 0.047
$30.0 < P < 35.0$	0.587 ± 0.03	0.413 ± 0.015	0.345 ± 0.02	0.356 ± 0.036
$35.0 < P < 40.0$	0.438 ± 0.031	0.356 ± 0.018	0.282 ± 0.02	0.328 ± 0.031
$40.0 < P < 50.0$	0.345 ± 0.033	0.319 ± 0.015	0.242 ± 0.017	0.265 ± 0.02
$50.0 < P < 60.0$	0.32 ± 0.066	0.263 ± 0.022	0.197 ± 0.022	0.2 ± 0.023
$60.0 < P < 70.0$	0.203 ± 0.089	0.272 ± 0.037	0.195 ± 0.024	0.223 ± 0.025

6.2 Proton misidentification probability

The proton misidentification probability, $\epsilon(p \rightarrow \mu)$, has been measured using a sample of $\Lambda \rightarrow p\pi$ decays from the Stripping20 and Stripping20r1 samples. The procedure is similar to that used for the evaluation of the kaon and pion fake rates: the probe track is required to be TIS with respect to L0Global and HLT1Phys lines. The mass distribution of the $\Lambda \rightarrow p\pi$ candidates is fitted with a double Gaussian function (for the signal) and a polynomial function (for the background) in bins of p and p_T of the proton. For the highest p_T bin, the signal is obtained by subtracting from the number of events in the signal mass region the background events in the same region extrapolated from an exponential fit to the Λ mass sidebands.

The $\epsilon(p \rightarrow \mu)$ values after the IsMuon and DLL requirements are shown in Figure 8 and the numbers reported in Tables 9 and 10.

It is worth to add here a note on the size of the misID of the different particles. For muonID only, the proton misID rate is on average smaller than the misID for pions or kaons. This difference is concentrated at low momentum ($p < 40$ GeV) while at high momentum the difference is less pronounced with respect to the kaons and negligible with respect to the pions. The PID cut significantly reduces the misID rate of kaons, while it is less effective on pions and even less on protons. The net effect is to lower the kaon misID rate, making it quite similar to the proton one.

The measured $\epsilon(p \rightarrow \mu)$ misID will be used in Section 6.4.6 to evaluate the yield of the $\Lambda_b \rightarrow p\mu\nu$ events, with the proton misidentified as a muon, in the low dimuon mass sideband. Since the flight distance (FD) of the Λ_b is considerably shorter than the one of Λ (the reconstructed FD distribution of the first one dies around 40 mm, where the second one starts to rise), to safely use the misID measured on the Λ control sample to correct the proton from the Λ_b , we have to check if the misID depends on the flight distance. Given the limited statistics of the control sample, we divided it into two parts, defined by $FD_\Lambda < 200$ mm and $FD_\Lambda > 200$ mm respectively (in fact the first bin is $\sim 50 < FD_\Lambda < 200$ mm), and we measured the proton misID in the two samples⁷. In each (p, p_T) bin where the comparison is possible⁸, the misID measured in the low and high FD_Λ bin agree within the errors, and agree with the one measured in the whole sample. Moreover, within the limited statistical accuracy, there it doesn't seem to be a trend as a function of the FD_Λ value. To the extent of the available control sample statistics, we can assume that the measured $\epsilon(p \rightarrow \mu)$ misID doesn't depend on the Λ decay point.

⁷for this tests we used the linear fit to side bands instead of the full fit

⁸some (p, p_T) bin has no enough statistic left

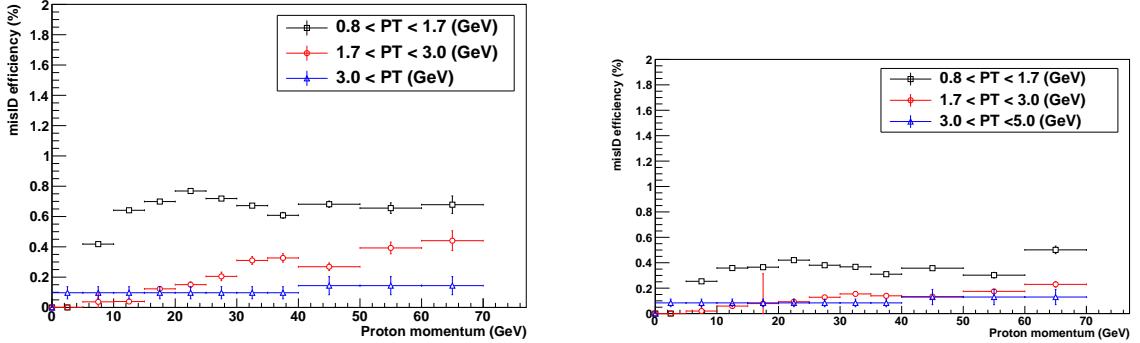


Figure 8: The $\epsilon(p \rightarrow \mu)$ mis-identification rate for Stripping20 (2012 data, left) and Stripping20r1 (2011 data, right) as a function of the proton momentum.

Momentum (GeV)	$0.8 < PT < 1.7$	$1.7 < PT < 3.0$	$PT > 3.0$
$0.0 < P < 5.0$	—	—	
$5.0 < P < 10.0$	0.418 ± 0.012	-	
$10.0 < P < 15.0$	0.641 ± 0.0096	0.039 ± 0.014	
$15.0 < P < 20.0$	0.699 ± 0.01	0.123 ± 0.015	
$20.0 < P < 25.0$	0.769 ± 0.012	0.151 ± 0.018	0.089 ± 0.048
$25.0 < P < 30.0$	0.719 ± 0.012	0.205 ± 0.028	
$30.0 < P < 35.0$	0.672 ± 0.016	0.31 ± 0.026	
$35.0 < P < 40.0$	0.608 ± 0.02	0.327 ± 0.028	
$40.0 < P < 50.0$	0.681 ± 0.021	0.269 ± 0.025	
$50.0 < P < 60.0$	0.656 ± 0.035	0.393 ± 0.039	0.151 ± 0.084
$60.0 < P < 70.0$	0.678 ± 0.058	0.441 ± 0.066	

Table 9: Proton to muon misidentification probability (in %) for the combined request isMuon && PID cut, as a function of the proton p and p_T for Stripping20 sample (2012 data).

Momentum (GeV)	$0.8 < PT < 1.7$	$1.7 < PT < 3.0$	$PT > 3.0$
$0.0 < P < 5.0$	-	-	0.084 ± 0.030
$5.0 < P < 10.0$	0.254 ± 0.0074	0.0198 ± 0.033	
$10.0 < P < 15.0$	0.358 ± 0.0051	0.0589 ± 0.0088	
$15.0 < P < 20.0$	0.365 ± 0.0053	0.0785 ± 0.23	
$20.0 < P < 25.0$	0.42 ± 0.0059	0.0932 ± 0.0075	
$25.0 < P < 30.0$	0.38 ± 0.0062	0.128 ± 0.0097	
$30.0 < P < 35.0$	0.368 ± 0.0078	0.154 ± 0.011	
$35.0 < P < 40.0$	0.31 ± 0.011	0.14 ± 0.012	
$40.0 < P < 50.0$	0.357 ± 0.0094	0.133 ± 0.01	
$50.0 < P < 60.0$	0.302 ± 0.022	0.175 ± 0.013	0.130 ± 0.057
$60.0 < P < 70.0$	0.501 ± 0.03	0.23 ± 0.022	

Table 10: Proton to muon misidentification probability (in %) for the combined request `isMuon && PID` cut, as a function of the proton p and p_T for Stripping20r1 sample (2011 data)

6.3 Peaking background from $B_{(s)}^0 \rightarrow h^+h^-$ with double misidentification: yield, BDT and mass PDFs

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 $B \rightarrow h^+h'$ decays (MC10, stripping17) selected as described in Section 4.

The convolution of the pion and kaon fake rate curves with the p and p_T spectrum of selected $B \rightarrow h^+h'$ decays has been done with a toy technique that takes properly into account the uncertainties on the fake rate in each of the 4×11 p and p_T bins and the correlations between p and p_T of both tracks.

The average double misID probability for Stripping20 and Stripping20r1 samples for the four exclusive $B^0 \rightarrow K^+\pi^-$, $B^0 \rightarrow \pi^+\pi^-$, $B_s^0 \rightarrow \pi^+K^-$ and $B_s^0 \rightarrow K^+K^-$ decays is shown in Table 11, after `isMuon` and the DLL selection. (the chosen combination of DLL cuts $DLL(K - \pi) < 10$ and $DLL(\mu - \pi) > -5$ selects $B_s^0 \rightarrow \mu^+\mu^-$ with high efficiency, ~ 0.98 see Section 5, and reduces the double misID rate by more than a factor of 5). The average double misID for the inclusive $B_{(s)}^0 \rightarrow h^+h^-$ decays, $\epsilon_{hh \rightarrow \mu\mu}$, is also shown in last column. This has been obtained by weighting the values obtained for exclusive decays according to their relative production rates⁹. The average double misID probability has been evaluated also using the misID rates obtained without cutting on ΔM (see Sect. 6.1). The latter evaluation differs by 8% and 10% for Stripping20 and Stripping20r1 respectively, and the difference will be assigned as systematic error on the double misID rate. A further 6% will be assigned to the Stripping20r1 due to difference between different trigger TIS conditions.

Table 11: Double misID probability, in units of 10^{-4} for different stripping samples

	$B^0 \rightarrow \pi\pi$	$B_s^0 \rightarrow KK$	$B^0 \rightarrow K\pi$	$B_s^0 \rightarrow K\pi$	Total($\epsilon_{h^+h' \rightarrow \mu\mu}$)
S20 (MD)	0.21 ± 0.0058	0.063 ± 0.0022	0.13 ± 0.0032	0.12 ± 0.0038	0.12 ± 0.0039
S20 (MU)	0.2 ± 0.0049	0.066 ± 0.002	0.13 ± 0.0029	0.12 ± 0.0034	0.12 ± 0.0037
S20 (2 fb^{-1})	0.21 ± 0.0036	0.064 ± 0.0017	0.13 ± 0.0023	0.12 ± 0.0027	0.12 ± 0.0036
S20r1 (MD)	0.18 ± 0.0073	0.056 ± 0.0029	0.11 ± 0.0043	0.11 ± 0.0049	0.11 ± 0.0039
S20r1 (MU)	0.21 ± 0.0087	0.052 ± 0.0033	0.11 ± 0.0053	0.11 ± 0.0059	0.11 ± 0.0046
S20r1(800 pb^{-1})	0.19 ± 0.0055	0.054 ± 0.0021	0.11 ± 0.0033	0.11 ± 0.0037	0.11 ± 0.0037

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

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

⁹ The following branching fractions are assumed: $\mathcal{B}(B^0 \rightarrow K^+\pi^-) = (1.95 \pm 0.06)10^{-5}$, $\mathcal{B}(B^0 \rightarrow \pi^+\pi^-) = (5.13 \pm 0.24)10^{-6}$ [34], $\mathcal{B}(B_s^0 \rightarrow \pi^+K^-) = (5.05 \pm 1.0)10^{-6}$, $\mathcal{B}(B_s^0 \rightarrow K^+K^-) = (3.5 \pm 0.28)10^{-5}$ and $f_s/f_d = 0.259 \pm 0.016$ [35]. Note that the uncertainty on f_s/f_d has been revised to be ± 0.015 in the mean time. The rest of this document uses the correct uncertainty, only this part uses ± 0.016 .

where N_{hh}^{TIS} is the number of $B_{(s)}^0 \rightarrow h^+h^-$ TIS events, ϵ^{TIS} is the L0 and HLT1 TIS efficiency for $B_{(s)}^0 \rightarrow h^+h^-$ events, $\epsilon^{\text{HLT2},MC}$ is the HLT2 efficiency for $B_{(s)}^0 \rightarrow h^+h^-$ events and $\epsilon_{B_s^0 \rightarrow \mu^+\mu^-}^{\text{TRIG|SEL}}$ is the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ trigger efficiency.

The number of $B_{(s)}^0 \rightarrow h^+h^-$ TIS events are extracted as described in Section 7.2.3. Table 12 shows the factors entering the Eq. 4. As a result, the expected background events due to the double misID of $B_{(s)}^0 \rightarrow h^+h^-$ events is $N(B_{(s)}^0 \rightarrow h^+h^- \rightarrow \mu\mu) = 15.2 \pm 1.2$ in the Stripping20 sample of 2 fb^{-1} and 6.4 ± 0.6 in the Stripping20r1 sample of 1 fb^{-1} .

Parameter	Stripping20 (2012 data)	Stripping20r1 (2011 data)
$B_{(s)}^0 \rightarrow h^+h^-$ yield	49653 ± 507	20143 ± 572
$\epsilon_{B_{(s)}^0 \rightarrow \mu^+\mu^-}^{\text{TRIG SEL}}$	$(92.4 \pm 0.3_{\text{stat}} \pm 1.9_{\text{syst}})\%$	$(92.1 \pm 0.5_{\text{stat}} \pm 1.6_{\text{syst}})\%$
ϵ^{TIS}	$(5.92 \pm 0.04_{\text{stat}} \pm 0.4_{\text{syst}})\%$	$(5.05 \pm 0.04_{\text{stat}} \pm 0.4_{\text{syst}})\%$
$\epsilon^{\text{HLT2},MC}$	$(91.6 \pm 0.2)\%$	$(91.5 \pm 0.3)\%$
$\epsilon_{hh \rightarrow \mu\mu}$	$(0.12 \pm 0.01) \times 10^{-4}$	$(0.11 \pm 0.01) \times 10^{-4}$
$N(B_{(s)}^0 \rightarrow h^+h^- \rightarrow \mu\mu)$	combined dataset 14.6 ± 1.3	

Table 12: Numbers entering into the computation of the $B_{(s)}^0 \rightarrow h^+h^- \rightarrow \mu\mu$ peaking background: comparison 2012 and 2011.

A second estimation of the peaking background is obtained using only data as a cross-check. The average single misID rate per hadron is computed in a sample of $B_{(s)}^0 \rightarrow h^+h^-$ LOTIS from events passing Hlt1AllL0Track, in which one of the hadrons has been misidentified as muon, with the resulting value being squared to obtain the double misID rate. Details are reported in Ref. [18], the conclusion being that the $B_{(s)}^0 \rightarrow h^+h^- \rightarrow \mu\mu$ rate from this method is compatible with the one described above, but with much larger uncertainty. This is due to the poor quality of the fit made to the $B \rightarrow h^+h'^-$ TIS sample with one hadron mis-identified as a muon.

Mass PDF of doubly misidentified $B \rightarrow h^+h'^-$ events

The mass lineshape of this background is obtained by emulating the double decay in flight by smearing the track momenta of the two hadrons according to a double Gaussian function that takes into account the missing momentum due to the neutrino and by rebuilding the mass of the mother particle.

This function is built using simulated $B \rightarrow h^+h'^-$ events where only one of the two hadrons decays in flight by fitting the distribution $(p_{\text{reco}} - p_{\text{true}})/p_{\text{true}}$ where p_{reco} is the reconstructed momentum of the track that has decayed in flight and p_{true} is its true momentum.

The above function is fitted separately for kaon and pion tracks, Fig. 9, in each case asking for the kaon or pion tracks to pass the IsMuon=1 requirement. The result of the fit are shown in Table 13.

parameter	value
f_π	0.9442 ± 0.0471
$\bar{x}_{1,\pi}$	0.0004 ± 0.00016
$\sigma_{1,\pi}$	0.0005 ± 0.0002
$\bar{x}_{2,\pi}$	-0.031 ± 0.002
$\sigma_{2,\pi}$	0.030 ± 0.001
f_K	0.9821 ± 0.0281
$\bar{x}_{1,K}$	0.00009 ± 0.00011
$\sigma_{1,\pi}$	0.00449 ± 0.00009
$\bar{x}_{2,K}$	-0.003 ± 0.003
$\sigma_{2,K}$	0.005 ± 0.003

Table 13: Parameters of the double Gaussian used to fit the distributions $(p_{\text{reco}} - p_{\text{true}})/p_{\text{true}}$. f represents the fraction of the first gaussian in the total PDF, \bar{x} represents the mean and σ the standard deviation of the gaussian. The subscript 1 denotes the major gaussian and 2 denotes the minor gaussian. The subscript K and π denotes the particle to which the gaussian corresponds.

This method is used because in the momentum regime of kaons and pions from $B \rightarrow h^+h'^-$ decays, the dominant mechanism of hadron to muon misidentification is the in flight decay of the hadron into muon and neutrino. It is assumed that the decay in flight happens outside the LHCb vertex locator (VELO); hence it does not affect the opening angle of the tracks. We tested this assumption (see Fig.10) and found it to be reasonable.

The lineshape is used to evaluate the efficiency of the signal mass window used in the analysis. The systematic uncertainty is evaluated by “over” or “under” smearing the track in question, recomputing the mass distribution of the mother particle and then the mass window efficiency, as shown in Fig. 11.

To test the validity of this method, a special MC sample of $B^0 \rightarrow K^+\pi^-$ events was generated where only the $B^0 \rightarrow K^+\pi^-$ events in which both hadrons have decayed in

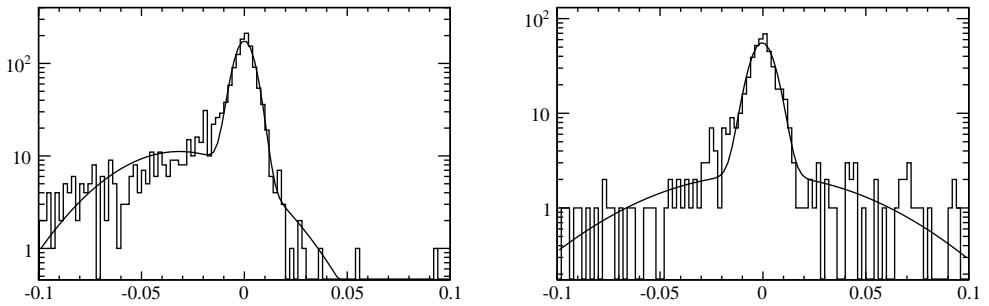


Figure 9: Distributions of $p_{\text{reco}} - p_{\text{true}}/p_{\text{true}}$ for pions (left) and kaons (right) from a simulated $B_{(s)}^0 \rightarrow h^+h^-$ sample.

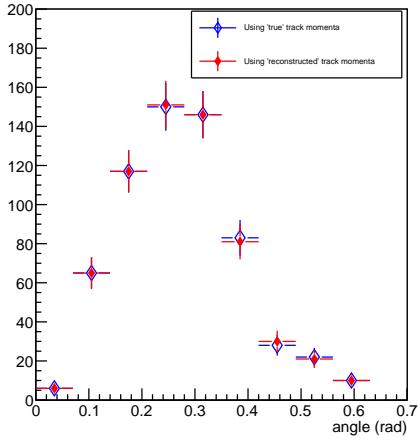


Figure 10: Opening angle of the daughter tracks for $B_{(s)}^0 \rightarrow h^+h^-$ events with and without double decays in flight.

flight before the LHCb muon chambers have been kept (hereafter called DIF sample). This sample is very slow to produce because the probability for a kaon and pion to decay in flight before the muon chambers is less than 1% for most of the momentum range of $B \rightarrow h^+h^-$ decays. Therefore we used this special sample only to validate the smearing procedure outlined above.

The invariant mass for a monte carlo sample $B^0 \rightarrow K^+\pi^-$ has been recomputed after smearing the tracks of the daughters and compared to the mass obtained with the (limited) MC $B^0 \rightarrow K^+\pi^-$ sample with double decays in flight, shown in Fig. 12.

Given the fact that the difference of the fractions of the two distributions (smeared and DIF) in the B^0 and B_s^0 mass windows is less than the systematic uncertainty as

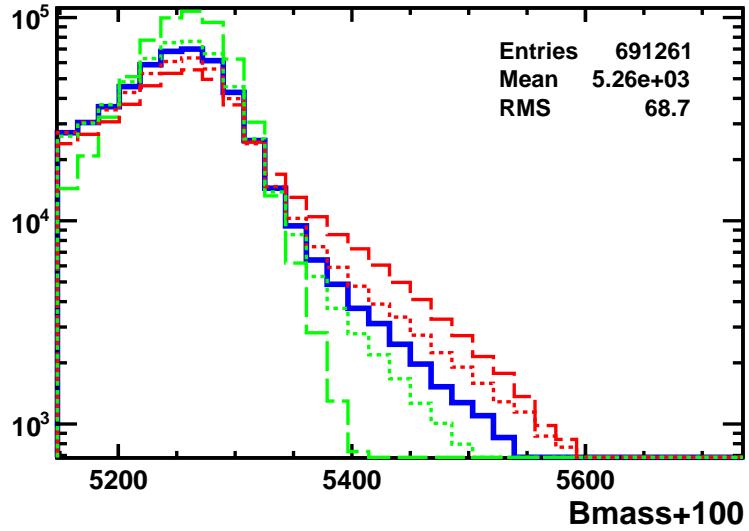


Figure 11: Invariant mass distribution calculated for $B_{(s)}^0 \rightarrow h^+h^-$ with double misID, evaluated as described in the text: blue curve represents the central value, red and green curves the error band. Short dashed curves show how the central curves are modified by moving the track parameters 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).

described above, we assume that the smearing method produces the mass lineshape of $B \rightarrow h^+h' \rightarrow \mu\mu$ to a good approximation. More details of the smearing procedure and systematic uncertainty evaluation can be found in Ref. [36].

In this approximation, 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.

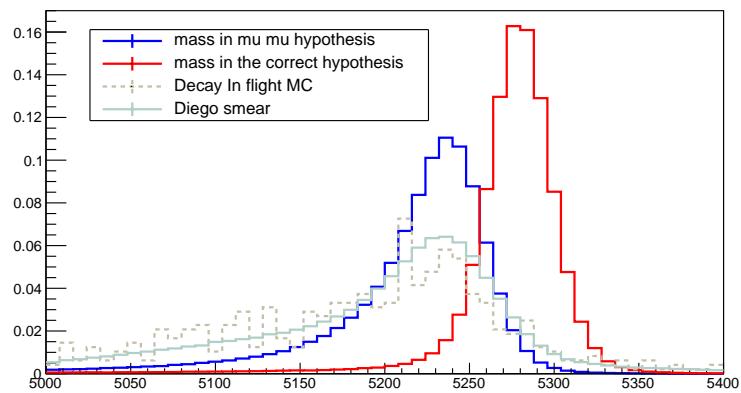


Figure 12: Invariant mass distribution of $B^0 \rightarrow K^+\pi^-$ monte carlo events in the correct mass hypothesis ($K\pi$) (red line), in the $\mu\mu$ hypothesis (blue line), in the $\mu\mu$ hypothesis after the smearing procedure (light blue line) and for double decays in flight events (gray dotted line).

BDT shape for doubly misidentified $B_{(s)}^0 \rightarrow h^+h^-$ events

To take into account the dependence of double misID from the BDT bins, we evaluated it as a function of BDT bin, $\epsilon_{hh \rightarrow \mu\mu}(i)$, and we defined for each BDT bin a fractional 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 given above. The values of $f_{misID}(i)$ are given in Table 14, for each decay separately and for the weighted average, after the `isMuon` and `DLL` cuts for both samples under study. The number of $B_{(s)}^0 \rightarrow h^+h^- \rightarrow \mu\mu$ events in each BDT bin is then computed using Eq. 4.

BDT range	$B^0 \rightarrow \pi\pi$	$B_s^0 \rightarrow KK$	$B^0 \rightarrow K\pi$	$B_s^0 \rightarrow K\pi$	Total($\epsilon_{hh \rightarrow \mu\mu}$)
Stripping20 (2012 data)					
[0-1.0]	0.21 ± 0.0036	0.064 ± 0.0017	0.13 ± 0.0023	0.12 ± 0.0027	0.12 ± 0.0036
[0-0.25]	0.27 ± 0.0029	0.05 ± 0.0011	0.12 ± 0.0019	0.12 ± 0.0019	0.12 ± 0.0042
[0.25-0.4]	0.26 ± 0.0033	0.056 ± 0.0012	0.13 ± 0.002	0.13 ± 0.0022	0.13 ± 0.0043
[0.4-0.5]	0.24 ± 0.0036	0.061 ± 0.0015	0.14 ± 0.0025	0.13 ± 0.0026	0.13 ± 0.0042
[0.5-0.6]	0.22 ± 0.0041	0.067 ± 0.0018	0.14 ± 0.0025	0.13 ± 0.0031	0.13 ± 0.0039
[0.6-0.7]	0.21 ± 0.0044	0.069 ± 0.0019	0.14 ± 0.0028	0.13 ± 0.0031	0.13 ± 0.0036
[0.7-0.8]	0.17 ± 0.0041	0.072 ± 0.0022	0.14 ± 0.0029	0.13 ± 0.0033	0.13 ± 0.0035
[0.8-0.9]	0.16 ± 0.004	0.073 ± 0.0024	0.12 ± 0.0029	0.12 ± 0.0032	0.12 ± 0.003
[0.9-1.0]	0.14 ± 0.0037	0.07 ± 0.0023	0.11 ± 0.0027	0.11 ± 0.0031	0.1 ± 0.0025
Stripping20r1 (2011 data)					
[0-0.25]	0.24 ± 0.0046	0.039 ± 0.0014	0.1 ± 0.0027	0.099 ± 0.0026	0.11 ± 0.004
[0.25-0.4]	0.23 ± 0.0051	0.045 ± 0.0016	0.11 ± 0.003	0.11 ± 0.003	0.11 ± 0.0043
[0.4-0.5]	0.22 ± 0.0055	0.05 ± 0.0019	0.12 ± 0.0036	0.12 ± 0.0035	0.12 ± 0.0043
[0.5-0.6]	0.2 ± 0.0061	0.055 ± 0.0023	0.12 ± 0.0035	0.12 ± 0.0042	0.12 ± 0.0041
[0.6-0.7]	0.19 ± 0.0066	0.058 ± 0.0024	0.12 ± 0.0039	0.12 ± 0.0042	0.12 ± 0.0039
[0.7-0.8]	0.16 ± 0.0063	0.062 ± 0.0029	0.13 ± 0.004	0.12 ± 0.0046	0.12 ± 0.0039
[0.8-0.9]	0.15 ± 0.0062	0.064 ± 0.0031	0.11 ± 0.004	0.11 ± 0.0045	0.11 ± 0.0034
[0.9-1.0]	0.14 ± 0.006	0.063 ± 0.0031	0.1 ± 0.0039	0.1 ± 0.0045	0.098 ± 0.0031

Table 14: Double misID probability in units of 10^{-4} as a function of the BDT bin for the four exclusive $B_{s,d} \rightarrow KK, K\pi, \pi\pi$ samples separately and the for the inclusive $B_{(s)}^0 \rightarrow h^+h^-$ sample.

6.4 Peaking backgrounds different from $B_{(s)}^0 \rightarrow h^+h^-$

Besides $B_{(s)}^0 \rightarrow h^+h^-$, some other exclusive B -meson decays can modify the background expectation of $B_{s,d}^0 \rightarrow \mu^+\mu^-$ if their shape differs from the exponential used to fit the combinatorial background. Not only their influence in the mass windows need to be investigated, but also the complete pdf in the full mass range, to avoid background overestimation. In this section the treatment of various exclusive decays with one or two real muons in the final state is described.

In order to estimate the contribution of each of these decays, a similar procedure has been adopted for each of them. The number of expected candidates has been estimated by normalising to the $B^+ \rightarrow J/\psi K^+$ channel:

$$\begin{aligned} N^{Exp} &= N(B^+ \rightarrow J/\psi K^+) \frac{f_x}{f_u} \frac{\mathcal{B}_x}{\mathcal{B}(B^+ \rightarrow J/\psi K^+)} \frac{\epsilon_x^{tot}}{\epsilon^{tot}(B^+ \rightarrow J/\psi K^+)} \\ &= \beta_x \cdot \epsilon_x^{tot} \cdot \mathcal{B}_x \end{aligned} \quad (5)$$

where x indicates a specific background channel. This procedure and the calculation of the normalisation parameters β_x is described in §8.9.

The total efficiency (ϵ_x^{tot}) for the given channel was estimated from MC simulations. For the channels that require at least one mis-identification in order to fake the signal the estimation of the mis-identification efficiency has been done with a dedicated method (similar to the one used for two-body hadronic decays) described below.

For each channel the total number of expected events has been estimated in the dimuon invariant mass region [4900, 6000] MeV/ c^2 and in the B^0 and B_s^0 search windows. Furthermore expectations in bins of BDT have been computed. In the cases in which the expected number of events in the most sensitive bin was not negligible the channel was included as separate component in the final estimations. Therefore PDFs of the invariant mass distributions were measured in MC simulations.

For these channels the same efficiency is assumed for 2011 and 2012 data samples with the exception of the mis-identification maps which are measured in their respective data samples.

MisID method

In order to increase the statistics used to determine the total efficiency, but also in order to employ data measurements of the particle mis-identification efficiencies, for background channels with at least one particle mis-identified the following procedure was adopted.

MC candidates were selected using the $B_{(s)}^0 \rightarrow h^+h^-$ selection: this is identical to the $B_s^0 \rightarrow \mu^+\mu^-$ but without the **IsMuon** requirement (and with a smaller invariant mass cut, irrelevant at this stage of analysis). No trigger requirement was applied at this stage;

the trigger efficiency was estimated instead only from the candidates selected with the $B_s^0 \rightarrow \mu^+ \mu^-$ selection and applied separately.

Each candidate was then weighted according to the mis-identification probability (after `IsMuon` and the ΔLL cuts) as measured in data for kaons, pions (§6.1) and protons (§6.2). The weight was obtained from the maps binned in momentum and p_T . Some of the maps have empty bins due to lack of statistics, for these bins the weight of the closest non-empty bin was used; the difference with respect to the case of weights left at one was taken as systematic uncertainty.

In the channels with one real muon (like $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$) the muon candidate was required to pass the standard identification criteria.

Monte Carlo truth matching

All the considered channels are simulated within normal pp events. Therefore some of the selected candidates can actually come from combinations of one of the channel tracks with other tracks of the event or even completely from random combinations of un-related tracks. This kind of candidates is already taken into account in the combinatorial background; therefore in order not to overestimate the number of background coming from specific channels we require the MC truth matching of the candidates.

Each track is considered as matched if it satisfies one of the following conditions:

1. it's true ID is the correct hadron ID and it comes from the correct *mother* (e.g. for $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$ the absolute pion true ID must be 211 and it must come from the B^0);
2. it's true ID is the one of a muon (13) and it's *mother* is the correct hadron ID and it's *grand-mother* is the correct one;
3. it's true ID is the correct hadron ID and it comes from an hadron of same ID which then comes from the correct *grand-mother*.

Case 1 takes into account the standard situation when most of the track is associated with the hadron. Case 2 takes into account that in part of the decays in flight most of the track is associated with the muon. Case 3 is just a feature of the association procedure in cases of pion scattering with material.

6.4.1 Determination of mass shape and BDT fractions

The backgrounds that are not negligible have to be accounted for in the background PDFs, both for the fit to the mass sidebands and for the branching ratio limit computation in the signal mass windows. For this, both the mass shape and the fraction of events expected per BDT bin has to be determined. Due to the aforementioned correlations between mass and BDT the invariant mass shape depends on the BDT bin. The PDF distribution is assumed to be the same among the BDT bins while the parameters of the distribution are determined bin-by-bin.

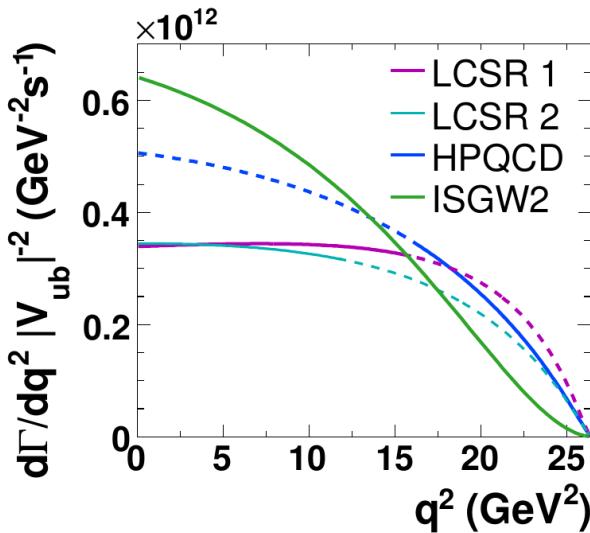


Figure 13: Di-lepton invariant mass squared for $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$ events, different models. The model we use is ISGW2. See Ref. [39] and references therein for details.

The model used to fit the mass in bins of BDT is the *RooPhysBkg* model, developed for the LHCb $B \rightarrow hh$ and documented in the LHCb roadmap [37].

In the final fit these channels are therefore implemented as separate PDFs in the different BDT bins with a common normalisation given by the number of expected events and a bin dependent scaling fraction; these fractions contain only the multinomial error of the distribution in bins and are constrained to sum up at 1. The normalisation uncertainty is instead all included in the total expected yield.

6.4.2 $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$

The semileptonic decay $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$, with a branching fraction of $(1.44 \pm 0.05) \cdot 10^{-4}$ [38], can contribute to the backgrounds of $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ when the pion is misidentified as muon. The invariant mass of the dimuon candidate is shifted to the left due to the missing neutrino and for the same reason the BDT is shifted to low values: however these two requirements are correlated as more “pointing” decays have also less missing mass, therefore the contribution and the invariant mass distribution of this decay have to be carefully estimated in bins of BDT.

A MC sample of about 6 millions events equally split in the two magnetic polarities has been used to study this channel. This sample was produced with the physics model “ISGW2”, which shows rather good agreement with other potential physics models (see Fig. 13). The sample was produced with a cut at generator level requiring the $\pi\mu$ invariant mass to satisfy $m(\pi\mu) > 4500$ MeV/ c^2 . This sample corresponds to approximately 190 M events produced with standard DecProdCut, i.e. only acceptance. (See Table 1 for details

on the sample production). A generation efficiency of 0.00645 ± 0.00004 was measured (including acceptance and the invariant mass cut).

The trigger efficiency for this channel was measured to be 0.857 ± 0.011 while the total selection efficiency including particle identification was estimated to be $(1.9395 \pm 0.0001) \cdot 10^{-4}$ in the invariant mass range [4900-6000] MeV/ c^2 . The number of expected events for the total 2011+2012 sample is:

$$N_{exp}^{2011+2012} = 114.4 \pm 6$$

The full breakdown of the numbers is reported in Table 15.

The expected events in bins of BDT are shown in Table 16 together with the corresponding fractions. The invariant mass distributions in bins of BDT have been fitted with a RooPhysBkg distribution: the plots are shown in Fig. 14 for the combined sample.

Table 15: $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$ expected yields and efficiencies, for the combined sample.

β	$8.19419e + 11 \pm 3.3052e + 10$
Branching fraction (\mathcal{B})	$0.000144 \pm 5e - 06$
Normalised total decays ($\beta \cdot \mathcal{B}$)	$1.17996e + 08 \pm 6.28004e + 06$
ϵ_{gen}	$0.00645226 \pm 3.67177e - 05$
ϵ_{trig}	0.85687 ± 0.0108179
ϵ_{sel} (up to mass cut)	$0.000175312 \pm 4.00934e - 07$
	Efficiency
After stripping	$1.41733e - 06 \pm 1.98142e - 08$
After fid. cuts	$1.19852e - 06 \pm 1.67785e - 08$
Mass in [4900,6000] MeV/ c^2	$9.69255e - 07 \pm 1.36042e - 08$
In B_s^0 masswin	$4.11005e - 11 \pm 1.37163e - 11$
In B^0 masswin	$6.80063e - 09 \pm 2.0519e - 10$
	Events
After stripping	167.239 ± 9.20281
After fid. cuts	141.42 ± 7.78274
Mass in [4900,6000] MeV/ c^2	114.369 ± 6.29507
In B_s^0 masswin	0.0048497 ± 0.00163893
In B^0 masswin	0.802449 ± 0.0490937

Table 16: $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$ BDT fractions and expected yields per BDT bin.

BDT bin	N_{exp}	fraction
1	42.9791 ± 2.3769	0.375795 ± 0.00201952
2	21.5825 ± 1.1989	0.18871 ± 0.00141381
3	12.355 ± 0.690664	0.108028 ± 0.00105513
4	10.7243 ± 0.600404	0.0937695 ± 0.00095977
5	9.45185 ± 0.529989	$0.0826438 \pm 0.000884386$
6	8.19143 ± 0.459544	$0.0716231 \pm 0.000776908$
7	6.02384 ± 0.339614	$0.0526704 \pm 0.000642686$
8	3.06051 ± 0.175052	$0.0267601 \pm 0.000416164$
7 & 8	9.08436 ± 0.50763	$0.0794305 \pm 0.000765662$

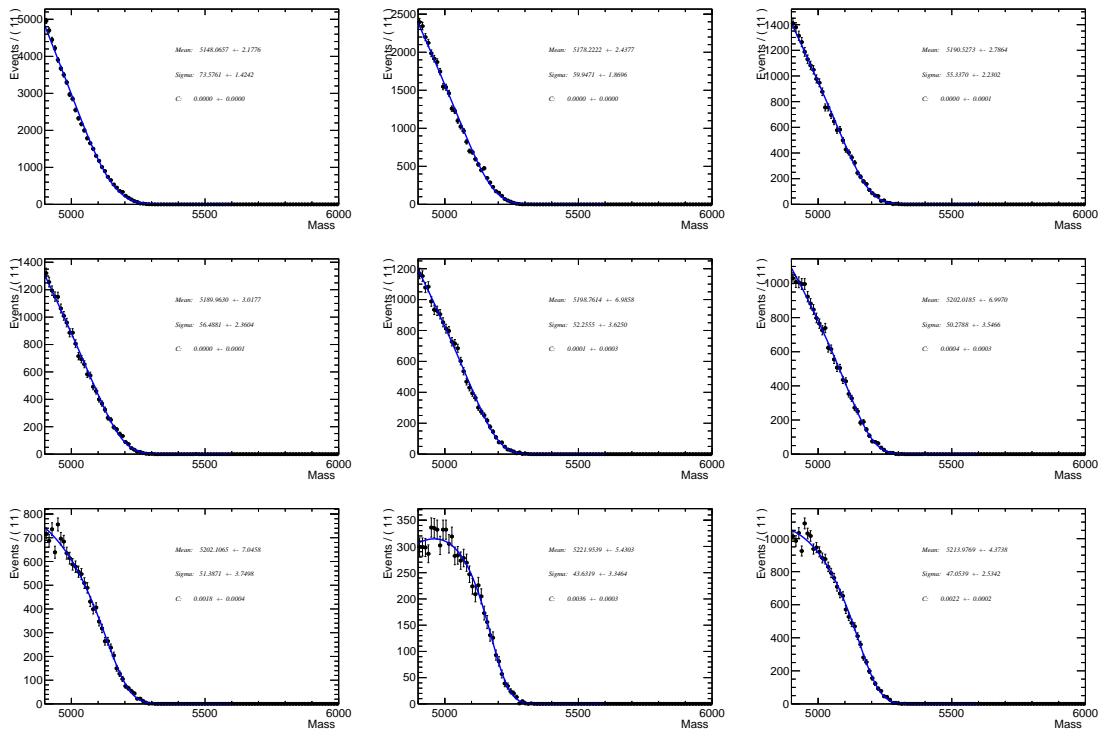


Figure 14: Invariant mass distributions and RooPhysBkg fits to the $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$ channel weighted according to the combined dataset misidentification efficiency.

6.4.3 $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$

Similarly to $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$, also the semileptonic decay $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ can represent a significant peaking background for the analysis, if the kaon is misidentified as muon. The branching fraction of this decay is of the same order of the branching fraction of $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$, and has been recently computed to be $(1.27 \pm 0.49) \cdot 10^{-4}$ [40]; we take this estimate in absence of experimental results. Nevertheless, we expect a smaller contribution from $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ due to the larger mass shift and lower fragmentation fraction of the B_s^0 .

A sample of about 6 million Monte Carlo events of $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ was used, produced requiring the $K\mu$ invariant mass to satisfy $m(K\mu) > 4500$ MeV/ c^2 . (See Table 1 for details on the sample production). A generation efficiency of 0.00880 ± 0.00005 was measured (including acceptance and the invariant mass cut).

The trigger efficiency was measured to be $\epsilon^{Trig|sel} = 0.777 \pm 0.018$.

Weighting this sample with the particle mis-identification efficiencies computed in §6.1, the expected number of events after full selection in the invariant mass range [4900-6000] MeV/ c^2 is:

$$N_{exp}^{2011+2012} = 10.2 \pm 4.0$$

The full breakdown of the numbers is reported in Table 17.

The expected events in bins of BDT are shown in Table 18 together with the corresponding fractions. The invariant mass distributions in bins of BDT have been fitted with a RooPhysBkg distribution: the plots are shown in Fig. 15 for the combined sample.

Table 17: $B_s^0 \rightarrow K^-\mu^+\nu_\mu$ expected yields and efficiencies, for the full data sample.

β	$2.12229e + 11 \pm 1.5658e + 10$
Branching fraction (\mathcal{B})	$0.000127 \pm 4.9e - 05$
Normalised total decays ($\beta \cdot \mathcal{B}$)	$2.69531e+07 \pm 1.05877e+07$
ϵ_{gen}	$0.0088 \pm 5e - 05$
ϵ_{trig}	0.776735 ± 0.0180378
ϵ_{sel} (up to mass cut)	$5.55827e - 05 \pm 1.34023e - 07$
Efficiency	
After stripping	$5.34202e - 07 \pm 1.28178e - 08$
After fid. cuts	$4.57411e - 07 \pm 1.09813e - 08$
Mass in [4900,6000] MeV/ c^2	$3.79923e - 07 \pm 9.12909e - 09$
In B_s^0 masswin	$6.13852e - 10 \pm 4.28611e - 11$
In B^0 masswin	$9.74026e - 09 \pm 2.80173e - 10$
Events	
After stripping	14.3984 ± 5.6665
After fid. cuts	12.3287 ± 4.85195
Mass in [4900,6000] MeV/ c^2	10.2401 ± 4.03001
In B_s^0 masswin	0.0165452 ± 0.00660113
In B^0 masswin	0.262531 ± 0.103403

Table 18: $B_s^0 \rightarrow K^-\mu^+\nu_\mu$ BDT fractions and expected yields per BDT bin.

BDT bin	N_{exp}	fraction
1	3.2481 ± 1.27842	0.317194 ± 0.00173252
2	1.72879 ± 0.680489	0.168825 ± 0.00125924
3	1.07261 ± 0.422249	0.104746 ± 0.000985416
4	1.04855 ± 0.41278	0.102397 ± 0.000971286
5	1.03218 ± 0.406337	0.100798 ± 0.000961143
6	0.976945 ± 0.384597	$0.0954037 \pm 0.000932054$
7	0.746196 ± 0.293782	0.0728699 ± 0.00080375
8	0.386737 ± 0.152311	$0.0377668 \pm 0.000565983$
7 & 8	1.13293 ± 0.445981	0.110637 ± 0.000983031

6.4.4 $B^+ \rightarrow \pi^+\mu^+\mu^-$ and $B^0 \rightarrow \pi^0\mu^+\mu^-$

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

The $B^+ \rightarrow \pi^+\mu^+\mu^-$ has been observed by the LHCb collaboration with a branching fraction [41]:

$$\mathcal{B}(B^+ \rightarrow \pi^+\mu^+\mu^-) = (2.3 \pm 0.6(\text{stat.}) \pm 0.1(\text{syst.})) \cdot 10^{-8},$$

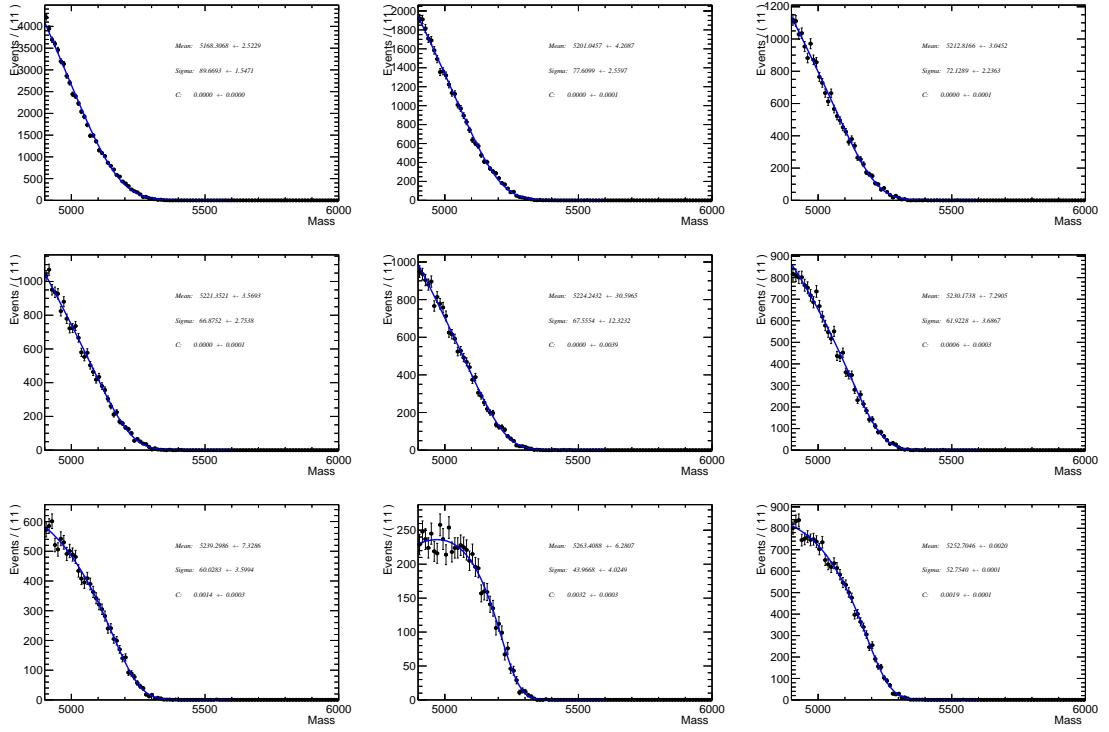


Figure 15: Invariant mass distributions and RooPhysBkg fits to the $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ channel weighted according to the combined dataset misidentification efficiency.

while the $B^0 \rightarrow \pi^0 \mu^+ \mu^-$ has not been observed so far, but is expected with a similar rate. As far as their contribution as $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ background is concerned, these two modes are indistinguishable and will be therefore treated together.

In order to have a value for the $B^0 \rightarrow \pi^0 \mu^+ \mu^-$ branching fraction we use the predicted branching fractions of the two modes in Ref. [40], namely $\mathcal{B}(B^+ \rightarrow \pi^+ \mu^+ \mu^-) = 1.95^{+0.61}_{-0.48}$ and $\mathcal{B}(B^0 \rightarrow \pi^0 \mu^+ \mu^-) = 0.91^{+0.33}_{-0.28}$ and calculate the ratio of branching fractions to be:

$$\frac{\mathcal{B}(B^0 \rightarrow \pi^0 \mu^+ \mu^-)}{\mathcal{B}(B^+ \rightarrow \pi^+ \mu^+ \mu^-)} = 0.47^{+0.22}_{-0.18} ; \quad (6)$$

while as absolute scale we take the experimentally measured one.

An MC sample of 1 M events from the MC2012 productions is used to assess the properties of this channel as background. At generation both leptons are required to be in the detector acceptance leading to a generation efficiency of $\epsilon_{gen} = 0.25075 \pm 0.0007575$.

The total efficiency is measured to be $\epsilon_{tot} = (1.011 \pm 0.016) \cdot 10^{-3}$. When normalising to $B^+ \rightarrow J/\psi K^+$, the expected number of $B^{0(+)} \rightarrow \pi^{0(+)} \mu^+ \mu^-$ events in the invariant mass range [4900, 6000] MeV/c^2 is

$$N_{exp}^{2011+2012} = 28.01^{+8.51}_{-8.16}$$

for the combined 2011 and 2012 dataset. Table 19 summarizes the input values, cumulative efficiencies and expected events, which are scaled according to Eq. (6).

Table 19: $B^{0(+)} \rightarrow \pi^{0(+)}\mu^+\mu^-$ expected yields and efficiencies, for the combined 2011 and 2012 data sample.

β	$8.19419e + 11 \pm 3.3052e + 10$
Branching fraction (\mathcal{B})	$2.3e - 08 \pm 6e - 09$
Normalised total decays ($\beta \cdot \mathcal{B}$)	$18846.6 + - 4974.94$
ϵ_{gen}	0.25075 ± 0.0007575
ϵ_{trig}	0.921533 ± 0.00403786
ϵ_{sel} (up to mass cut)	$0.00403219 \pm 6.2945e - 05$
	Efficiency
After stripping	$0.00814571 \pm 5.05492e - 05$
After fid. cuts	$0.00508826 \pm 3.83341e - 05$
Mass in [4900,6000] MeV/ c^2	$0.00101107 \pm 1.60763e - 05$
In B_s^0 masswin	0 ± 0
In B^0 masswin	0 ± 0
	Events
After stripping	$225.673^{+68.4934}_{-65.683}$
After fid. cuts	$140.968^{+42.789}_{-41.0336}$
Mass in [4900,6000] MeV/ c^2	$28.0113^{+8.5115}_{-8.16308}$
In B_s^0 masswin	0^{+0}_{-0}
In B^0 masswin	0^{+0}_{-0}

The expected total number of $B^{0(+)} \rightarrow \pi^{0(+)}\mu^+\mu^-$ events divided in BDT bins and the corresponding fractions are shown in Table 20.

Table 20: $B^{0(+)} \rightarrow \pi^{0(+)}\mu^+\mu^-$ BDT fractions and expected yields per BDT12 bin.

BDT bin	N_{exp}	fraction
1	$10.6534^{+3.24835}_{-3.11632}$	0.380326 ± 0.00962312
2	$5.16984^{+1.58209}_{-1.51826}$	0.184563 ± 0.00670363
3	$2.91912^{+0.898156}_{-0.862316}$	0.104212 ± 0.00503729
4	$2.9123^{+0.896083}_{-0.860329}$	0.103969 ± 0.0050314
5	$2.11431^{+0.653582}_{-0.62775}$	0.0754809 ± 0.00428703
6	$2.08021^{+0.643218}_{-0.61781}$	0.0742635 ± 0.00425231
7	$1.405^{+0.438001}_{-0.420985}$	0.0501583 ± 0.00349469
8	$0.757061^{+0.241002}_{-0.232031}$	0.027027 ± 0.00256529
7 & 8	$2.16206^{+0.668091}_{-0.641665}$	0.0771853 ± 0.00433516

Based on this we have considered this background channels to be not negligible and therefore we included it in the final fit. In order to do this we need the mass shape parameters. We have therefore fitted the dimuon mass distribution from the $B^+ \rightarrow \pi^+\mu^+\mu^-$ MC sample in BDT bins with the RooPhysBkg distributions as shown in Fig. 16 for MC2012. We widen the fit range to [4200,6000] MeV in order to increase the fit stability. The parameters that define the dimuon mass distributions are assumed to be the same for $B^+ \rightarrow \pi^+\mu^+\mu^-$ and $B^0 \rightarrow \pi^0\mu^+\mu^-$ decays.

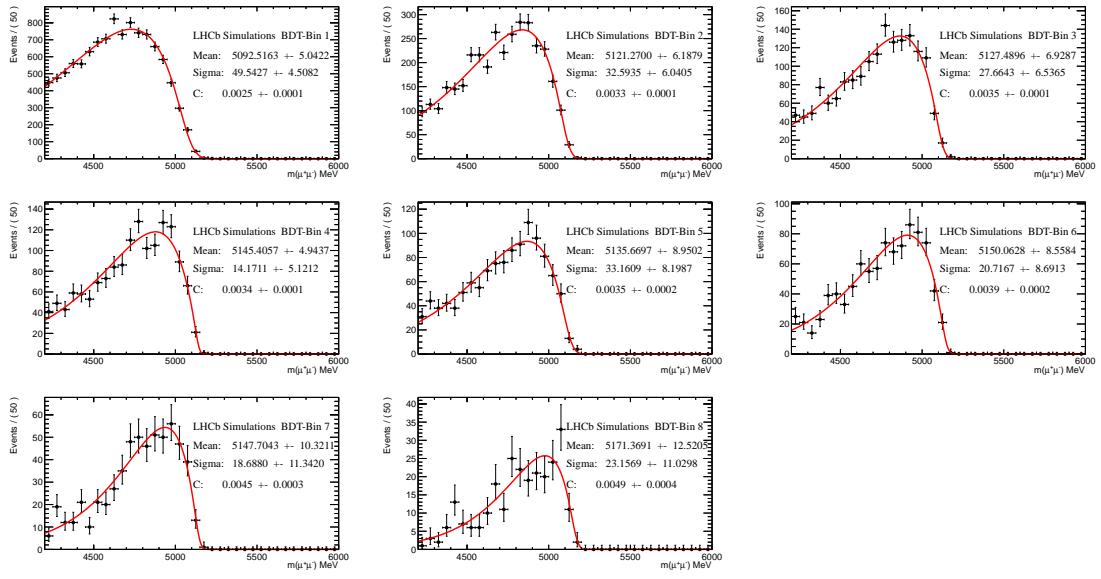


Figure 16: Invariant mass distributions and RooPhysBkg fits to the $B^+ \rightarrow \pi^+ \mu^+ \mu^-$ channel for each BDT bin.

6.4.5 $B_c^+ \rightarrow J/\psi \mu^+ \nu$

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

The $B_c^+ \rightarrow J/\psi \mu^+ \nu$ decay, with $J/\psi \rightarrow \mu^+ \mu^-$, could be a background due to the good vertex of the muon from the semileptonic decay and the oppositely charged one from the J/ψ . Clearly pointing and muon isolation of such a decay will be in general background-like but a detailed study was considered necessary.

The absolute branching fraction of this decay is not known. But as we normalise directly to $B^+ \rightarrow J/\psi K^+$, we use a measurement by the CDF collaboration [42]

$$\begin{aligned} \mathcal{R} &= \frac{\sigma(B_c^+) \mathcal{B}(B_c^+ \rightarrow J/\psi \ell \nu X)}{\sigma(B^+) \mathcal{B}(B^+ \rightarrow J/\psi K^+)} \\ &= 0.132^{+0.041}_{-0.037} (\text{stat}) \pm 0.031 (\text{sys})^{+0.032}_{-0.020} (\text{lifetime}) \\ &= 0.132^{+0.051}_{-0.052} \end{aligned}$$

to determine the number of expected events from $B_c^+ \rightarrow J/\psi \mu^+ \nu$ as:

$$N_{\text{exp}} = \frac{1}{2} \cdot \mathcal{R} \cdot \beta \cdot \mathcal{B}(B^+ \rightarrow J/\psi K^+) \cdot \mathcal{B}(J/\psi \rightarrow \mu^+ \mu^-) \cdot \varepsilon \quad (7)$$

where β is the already mentioned normalisation factor. The factor 1/2 comes from the fact that CDF measured the branching fraction regardless of the lepton being a muon

or an electron. In absence of further information we follow the authors of that paper in considering the value to be split half and half between the two leptons.

The efficiency of this channel has been measured in Monte Carlo simulations. A total efficiency of $\varepsilon = (3.33 \pm 0.05) \cdot 10^{-5}$ was measured for the full BDT range and the invariant mass range [4900, 6000] MeV/ c^2 . Table 21 summarises the input values, cumulative efficiencies and expected events. The number of expected events for the combined 2011 and 2012 dataset is:

$$N_{exp}^{2011+2012} = 108.3^{+42.3}_{-43.1} .$$

The distribution of the expected events in BDT bins is reported in Table 22 with the correspondent fractions. Since in the most sensitive part of the BDT distribution this background appears to be negligible we do not include it as a separate PDF in the final fit.

Table 21: $B_c^+ \rightarrow J/\psi \mu^+ \nu$ expected yields and efficiencies, for the combined 2011 and 2012 data sample.

	Efficiency	Events
β	$8.19419e + 11 \pm 3.3052e + 10$	
Branching fraction (\mathcal{B})	$3.96859e - 06^{+1.53961e-06}_{-1.56955e-06}$	
Normalised total decays ($\beta \cdot \mathcal{B}$)	$3.25194e+06 + 1.26838e+06 - 1.29279e+06$	
ϵ_{gen}	$0.00286931 \pm 2.8693e - 05$	
ϵ_{trig}	0.946182 ± 0.00201795	
ϵ_{sel} (up to mass cut)	$0.0116068 \pm 0.000106084$	
After stripping	$0.000263395 \pm 2.7588e - 06$	$856.543^{+334.206}_{-340.633}$
After fid. cuts	$0.000122927 \pm 1.3573e - 06$	$399.75^{+155.981}_{-158.98}$
Mass in [4900,6000] MeV/ c^2	$3.33036e - 05 \pm 4.51181e - 07$	$108.301^{+42.2672}_{-43.0796}$
In B_s^0 masswin	$9.0915e - 07 \pm 5.13891e - 08$	$2.9565^{+1.1652}_{-1.18716}$
In B^0 masswin	$2.90759e - 06 \pm 9.49796e - 08$	$9.45531^{+3.70085}_{-3.77158}$

Table 22: Expected background from $B_c^+ \rightarrow J/\psi \mu^+ \nu$ events (N_{exp}) and fractional yield (f) per BDT12 bin for the combined 2011 and 2012 data sample.

BDT bin	N_{exp}	fraction
1	$79.4648^{+31.0248}_{-31.6206}$	0.733738 ± 0.00786288
2	$14.5278^{+5.6815}_{-5.79026}$	0.134142 ± 0.00336197
3	$5.95895^{+2.33729}_{-2.38177}$	0.0550219 ± 0.00215317
4	$4.03347^{+1.58581}_{-1.61584}$	0.037243 ± 0.00177147
5	$2.327^{+0.919785}_{-0.937024}$	0.0214863 ± 0.00134553
6	$1.45095^{+0.577843}_{-0.588513}$	0.0133974 ± 0.00106248
7	$0.456275^{+0.189403}_{-0.192624}$	$0.00421301 \pm 0.00059581$
8	$0.0821294^{+0.0421529}_{-0.0426232}$	$0.000758342 \pm 0.000252781$
7 & 8	$0.538404^{+0.221508}_{-0.225343}$	$0.00497135 \pm 0.000647215$

6.4.6 $\Lambda_b^0 \rightarrow p\mu^-\nu$

The $\Lambda_b^0 \rightarrow p\mu^-\nu$ can fake the signal if the proton is mis-identified as a muon. This can happen due to noise or *punch-through* muons from the proton shower in the calorimeters. The branching fraction of this decay is largely unknown, we use the theoretical estimate given in Ref. [43]:

$$\mathcal{B}(\Lambda_b^0 \rightarrow p\mu^-\nu) = 3.3_{-1.2}^{+1.5} \cdot 10^{-4} \times \left(\frac{V_{ub}}{3.5 \cdot 10^{-3}} \right)^2 \quad (8)$$

and update this prediction with the latest averages of V_{ub} and of the Λ_b^0 lifetime from Ref. [38]. This results in a $\mathcal{B}(p)$ prediction of

$$\mathcal{B}(\Lambda_b^0 \rightarrow p\mu^-\nu) = (4.75 \pm 2.11) \cdot 10^{-4}. \quad (9)$$

It has to be noted that this value is significantly higher than the previously used value $(1.59 \pm 0.84) \cdot 10^{-4}$, which came from an outdated paper [44].

The authors of a very recent paper [45] have performed a Lattice QCD calculation of the form factors of the $\Lambda_b^0 \rightarrow p\mu^-\nu$ decay, giving a partial width of:

$$\frac{1}{|V_{ub}|^2} \int_{14\text{GeV}^2}^{q_{max}^2} \frac{d\Gamma(\Lambda_b^0 \rightarrow p\mu^-\nu)}{dq^2} dq^2 = 15.3 \pm 2.4 \pm 3.4 \text{ps}^{-1} \quad (10)$$

which can be translated (always using the above mentioned $|V_{ub}|$ and τ_{Λ_b} values) into a branching fraction of $(3.75 \pm 1.20) \cdot 10^{-4}$ which however is limited to the kinematic region $14\text{GeV}^2 < q^2 < q_{max}$. This prediction can thus be considered as lower bound for the total branching fraction. We therefore consider the value given in Eq. (9) to estimate the contribution of this decay.

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

$$r_\Lambda \equiv \frac{f_\Lambda}{f_u + f_d} = (0.404 \pm 0.017 \pm 0.027 \pm 0.105) \times [1 - (0.031 \pm 0.003 \pm 0.003) \times p_T(\text{GeV})] \quad (11)$$

and assuming $f_d = f_u$ we compute $\beta_\Lambda = \beta \cdot 2r_\Lambda$. However given the p_T dependence of the r_Λ measurement we assign a per event weight given by $2 * r_\Lambda(\tilde{p}_T)$ (multiplied by the already mentioned misID probability) and use simply β to normalise. Here there is a caveat: \tilde{p}_T is the transverse momentum of the dimuon combination formed by the muon and the proton in this decay, so it is an approximation of the transverse momentum of the Λ_b .

The $\Lambda_b^0 \rightarrow p\mu^-\nu$ has been studied with a sample of 1M events generated in a “Phase space” distribution. The events have been generated requiring $m(p\mu) > 4500$ MeV/ c^2 at generator level. The total generator level efficiency is measured to be 0.02909 ± 0.00012^{10} .

¹⁰The MC production used here is produced with simulation conditions “Sim06b”. For this configuration the standard LHCb calculation of the generation efficiency was affected by a bug; the “signal counters” efficiency had to be used instead. [46]

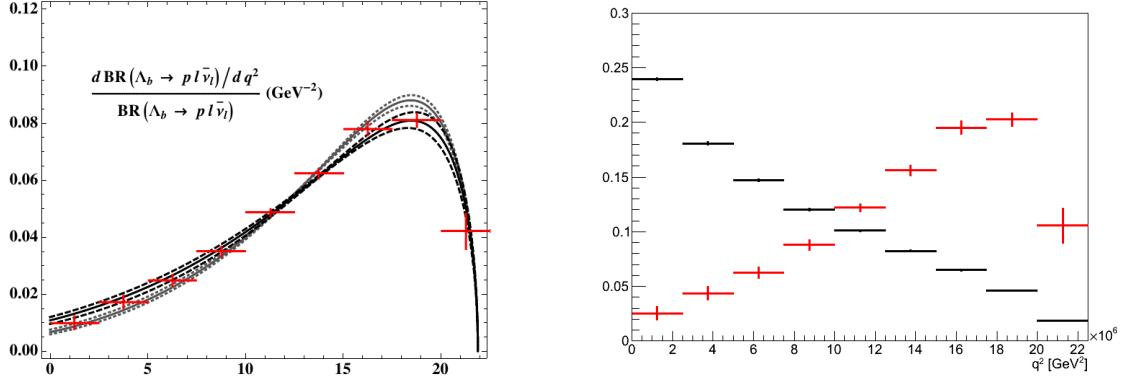


Figure 17: Di-lepton invariant mass squared of $\Lambda_b^0 \rightarrow p\mu^-\nu$ events. (left) Theoretical prediction using two different interpolating currents for the Λ_b (axial-vector and pseudo scalar) [47], superimposed in red is the binned approximation used for the reweighting. (right) Binned theoretical prediction prediction (red) shown together with generator level distribution obtained from the “phase space” sample (black).

The equivalent luminosity of the considered sample is about 0.4 fb^{-1} , the trigger efficiency was measured to be 0.74 ± 0.03 .

The predicted number of $\Lambda_b^0 \rightarrow p\mu^-\nu$ events mis-reconstructed as $B_s^0 \rightarrow \mu^+\mu^-$ candidates depends strongly on the distribution of the decay particles. The distribution of the di-lepton invariant mass squared ($q^2 = m_{\mu\nu}^2$), as predicted from Ref. [47], is shown in Fig. 17 (left). It is extracted in 5 bins of q^2 , which are overlaid to the theoretical prediction. This predicted q^2 distribution differs significantly from the one generated using the “phase space” model: the parameterization is shown in Fig. 17 (right), together with the generator level q^2 distribution of “phase space” events. The ratio between the “phase space” and the theoretical in each q^2 bin defines a weight. These weights are given in Tab. 23, they are used to reweight the events in the “phase space” full simulation.

The number of expected events, determined from the “phase space” sample, after reweighting in q^2 and after the full selection and trigger, is:

$$N_{exp}^{2011+2012} = 67.6 \pm 30.3$$

The full breakdown of the numbers is reported in Table 24. For the way this particular background has been reweighted the shown total efficiencies include also the lambda hadronisation fraction.

A comment on $\Lambda_b^0 \rightarrow p\mu^-\nu$. The estimated numbers for this channel are considerably higher than in the past analysis. This is due to various effects which we explain in the

Table 23: Weights to be applied to the “phase space” sample (PS) $n_{corr} = n_{PS}/weight$.

$q^2[\text{GeV}^2]$	weight
[0 – 2.5]	9.5 ± 2.4
[2.5 – 5]	4.15 ± 0.60
[5 – 7.5]	2.35 ± 0.20
[7.5 – 10]	1.37 ± 0.080
[10 – 12.5]	0.829 ± 0.027
[12.5 – 15]	0.528 ± 0.018
[15 – 17.5]	0.333 ± 0.012
[17.5 – 20]	0.227 ± 0.008
[20 – 22.5]	0.176 ± 0.027

following.

1. As mentioned the branching fraction is about 3 times higher than in the previous analysis.
2. The BDT shape of the signal is different between the MC simulation we use in this analysis and the previous one (MC11). This is due to a bug present in the previous version of the simulation for all hadron productions. This bug was biasing in the Λ_b decay products and was corrected by changing the production from “repeated hadronisation” to “signal plain” [46, 48].

The $\Lambda_b^0 \rightarrow p\mu^-\nu$ component, with its normalization and mass pdf will be inserted in the sideband interpolation, as a systematic crosscheck.

Table 24: $\Lambda_b^0 \rightarrow p\mu^-\nu$ expected yields and efficiencies (including hadronisation fraction), for the full dataset. The samples have been reweighted to match the predicted q^2 distribution.

	Efficiency	Events
β	$8.19419e + 11 \pm 3.3052e + 10$	
Branching fraction (\mathcal{B})	0.000475 ± 0.000211	
Normalised total decays ($\beta \cdot \mathcal{B}$)	$3.89224e + 08 \pm 1.73609e + 08$	
ϵ_{gen}	$0.0290911 \pm 0.000117567$	
ϵ_{trig}	0.745283 ± 0.0299241	
ϵ_{sel} (up to mass cut)	$8.01576e - 06 \pm 4.23253e - 08$	
After stripping	$2.36673e - 07 \pm 9.61411e - 09$	92.1187 ± 41.2585
After fid. cuts	$1.95882e - 07 \pm 7.96647e - 09$	76.2419 ± 34.1479
Mass in [4900,6000] MeV/ c^2	$1.7379e - 07 \pm 7.07296e - 09$	67.6434 ± 30.2969
In B^0 masswin	$1.18387e - 08 \pm 5.18891e - 10$	4.6079 ± 2.0652
In B_s^0 masswin	$2.06724e - 08 \pm 8.77522e - 10$	8.04618 ± 3.60512

Table 25: $\Lambda_b^0 \rightarrow p\mu^-\nu$ BDT fractions and expected yields per BDT bin.

BDT bin	N_{exp}	fraction
1	30.9497 ± 13.8664	0.457542 ± 0.00509792
2	12.6781 ± 5.68197	0.187426 ± 0.0029678
3	6.79066 ± 3.04469	0.100389 ± 0.00206636
4	5.44686 ± 2.44236	0.0805231 ± 0.00171614
5	4.52367 ± 2.02859	0.0668752 ± 0.0014821
6	3.65347 ± 1.63849	0.0540107 ± 0.00123508
7	2.42853 ± 1.08941	0.0359019 ± 0.00089879
8	1.17238 ± 0.526494	$0.0173317 \pm 0.000566565$
7 & 8	3.60091 ± 1.61441	0.0532337 ± 0.00106246

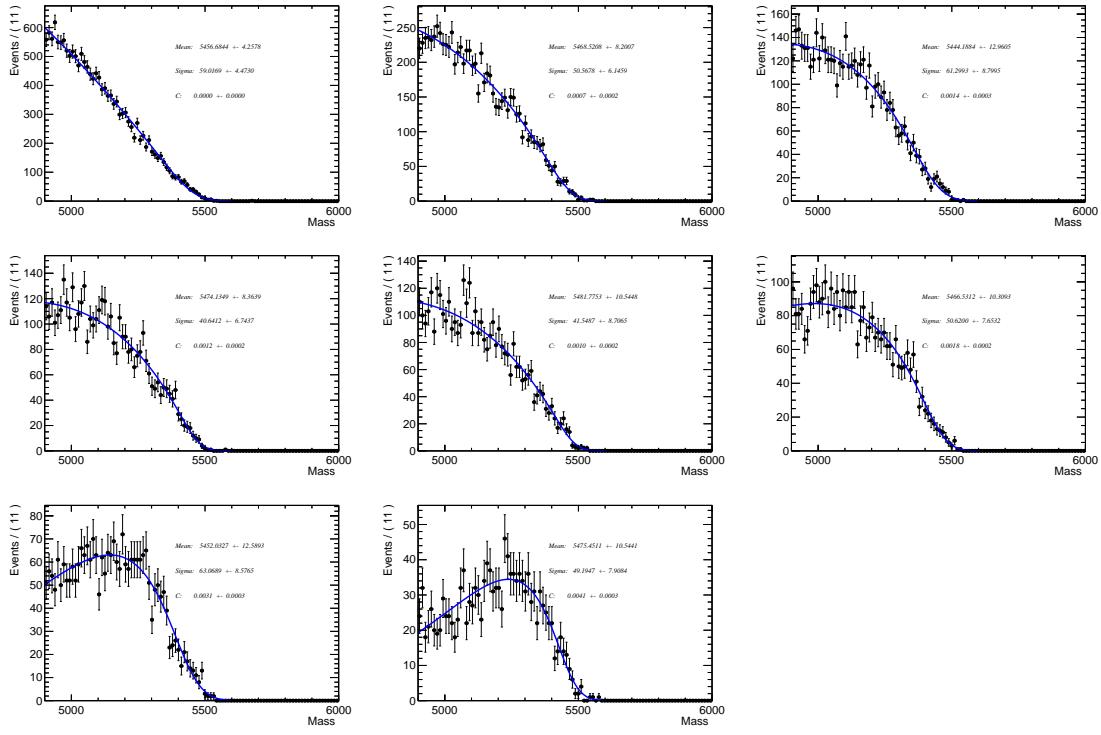


Figure 18: Invariant mass distributions and RooPhysBkg fits to the $\Lambda_b^0 \rightarrow p\mu^-\nu$ channel for each BDT bin for the combined dataset.

6.5 Background from the generic $b\bar{b} \rightarrow \mu^+\mu^-X$

The $b\bar{b} \rightarrow \mu^+\mu^-X$ Monte Carlo sample used for the BDT training has been used to investigate the presence of additional exclusive backgrounds with dimuon in the final state. The sample belongs to the MC12 production and corresponds to about 7 fb^{-1} of integrated luminosity at a pp center-of-mass energy of $\sqrt{s} = 8 \text{ TeV}$. It has been reconstructed with the Reco13a and Reco14 Stripping20 models, with the stripping in filtering mode. For

more details see the discussion in Sec. 3.1 and Table 1 (EvType 10012009). The dimuon invariant mass is required to be above $4700 \text{ MeV}/c^2$.

The selected candidates are classified according to their Monte Carlo information as $b\bar{b} \rightarrow \mu^+\mu^-X$ (two muons originating from two different b -quarks) or as $b \rightarrow \mu^+\mu^-X$ (two muons originating from the same b quark). In both cases the muons can either originate directly from the hadron containing the b quark or from intermediate resonances, mostly hadrons containing a c -quark. Three other categories are identified based on the information from the MC truth: NullId, where at least one of the two muon candidates does not have the MC truth information; NotAMuon, where at least one of the two muon candidates is not a real muon according to the MC truth; Other, which includes the remaining events and is mostly composed of candidates with at least one muon originating from the fragmentation of a charm quark¹¹. Table 26 summarizes the composition of the selected sample. The $b \rightarrow \mu^+\mu^-X$ component is about 0.4% of the $b\bar{b} \rightarrow \mu^+\mu^-X$ one for a

Table 26: Breakdown of the components in the generic MC sample for two range of the dimuon invariant mass.

channel	$m_{\mu\mu} > 4700 \text{ MeV}/c^2$	$m_{\mu\mu} = [4900, 6000] \text{ MeV}/c^2$
$b\bar{b} \rightarrow \mu^+\mu^-X$	103277	82414
$b \rightarrow \mu^+\mu^-X$	427	74
NullId	784	618
NotAMuon	163	118
Other	4322	3357
Total	108973	86581

dimuon mass above $4700 \text{ MeV}/c^2$ and it decreases to about 0.1% when the dimuon mass range is restricted to $[4900, 6000] \text{ MeV}/c^2$.

For candidates belonging to the $b \rightarrow \mu^+\mu^-X$ category, the mother, grand mother and grand grand mother of the two muons have been identified to characterize the decay process. The results are shown in Table 27. In the mass range $[4900, 6000] \text{ MeV}/c^2$ the sample is fully dominated by the $B_c \rightarrow J/\psi(\mu X)\mu X$ decay mode. Table 28 shows the seven largest components of the $b\bar{b} \rightarrow \mu^+\mu^-X$ background category in the $[4900, 6000] \text{ MeV}/c^2$ mass range.

¹¹Since the MC truth information of the muon candidate ancestors stored in the ntuples is limited to the mother, grand mother and grand grand mother, this category also contains very few events where both muons come from b decays but at least one b is beyond the grand grand mother relationship.

Table 27: Breakdown of the components in the $B \rightarrow \mu^+ \mu^- X$ sample for two ranges of the dimuon invariant mass.

channel	$m_{\mu\mu} > 4700 \text{ MeV}/c^2$	$m_{\mu\mu} = [4900, 6000] \text{ MeV}/c^2$
$B_c \rightarrow J/\Psi(\mu X) \mu X$	149	71
$B_s \rightarrow D_s(\mu X) \mu X$	139	3
$B^+ \rightarrow \mu \mu X$	52	0
$B^0 \rightarrow \mu \mu X$	43	0
$B_s \rightarrow D_s^*(D_s(\mu X) \gamma/\pi^0) \mu X$	28	0
$B^0 \rightarrow D^-(\mu X) \mu X$	13	0
$B^+ \rightarrow \bar{D}^0(\mu X) \mu X$	1	0
$B^0 \rightarrow \pi^-(\mu) \mu X$	1	0
$\Upsilon(1S) \rightarrow \tau(\mu) \tau(\mu)$	1	0
Total	427	74

Table 28: Main components of the $b\bar{b} \rightarrow \mu^+ \mu^- X$ background category.

channel	$m_{\mu\mu} = [4900, 6000] \text{ MeV}/c^2$
$B^- \rightarrow \mu X + B^0 \rightarrow \mu X$	$26.0 \pm 0.2\%$
$B^- \rightarrow \mu X + B^+ \rightarrow \mu X$	$16.2 \pm 0.1\%$
$B^0 \rightarrow \mu X + \bar{B}^0 \rightarrow \mu X$	$11.3 \pm 0.1\%$
$\Lambda_b \rightarrow \mu X + B^+ \rightarrow \mu X$	$6.2 \pm 0.1\%$
$\Lambda_b \rightarrow \mu X + B^0 \rightarrow \mu X$	$5.2 \pm 0.1\%$
$B^- \rightarrow \mu X + B_s^0 \rightarrow \mu X$	$4.0 \pm 0.1\%$
$B_s^0 \rightarrow \mu X + \bar{B}^0 \rightarrow \mu X$	$3.8 \pm 0.1\%$
other	27.3%
Total $b\bar{b} \rightarrow \mu\mu X$ events	82414

6.6 Summary of exclusive backgrounds

The number of events for all the exclusive decays described in this section estimated in the mass range $[4900 - 6000] \text{ MeV}/c^2$ and for BDT above 0.8 are listed in Tab. 29.

Table 29: Number of events expected in the 2011 and 2012 data samples for all the dominant exclusive background sources estimated in the mass range $[4900-6000] \text{ MeV}/c^2$, for BDT above 0.8.

	2011 + 2012, 3 fb^{-1}
$B^0 \rightarrow \pi^- \mu^+ \nu_\mu$	9.1 ± 0.5
$B_s^0 \rightarrow K^- \mu^+ \nu_\mu$	1.1 ± 0.4
$\Lambda_b^0 \rightarrow p \mu^- \nu$	3.6 ± 1.6
$B_c^+ \rightarrow J/\psi \mu^+ \nu$	0.5 ± 0.2
$B^{+(0)} \rightarrow \pi^{+(0)} \mu^+ \mu^-$	2.2 ± 0.6
$B_{(s)}^0 \rightarrow h^+ h^- \text{ misID}$	2.6 ± 0.3

7 Signal classification

7.1 BDT definition

As in the previous analysis of 2011 and half of 2012 data, we use a multivariate classifier to separate our signal candidate from the background even further. A Boosted Decision Tree (BDT) was constructed using the TMVA package [23]. It has been designed to select two body decays of b hadrons with respect to the background coming from semileptonic $b\bar{b} \rightarrow \mu^+\mu^-X$ decays. The number of variables present in the BDT discriminant is twelve, compared to the nine present in the previous analyses with 370 pb^{-1} , 1 fb^{-1} and $1\text{ fb}^{-1}+1.1\text{ fb}^{-1}$ [18, 22, 31]. The twelve topological and kinematical variables used in the BDT are:

- B meson proper time (t),
- minimum impact parameter significance of the muons ($IPS(\mu)$),
- the impact parameter of the B ($IP(B)$),
- distance of closest approach between the two muons ($DOCA$),
- the isolation of the two muons with respect to any other track in the event ($I(\mu)$),
- the transverse momentum of the B meson ($p_T(B)$),
- the cosine of the angle between the muon momentum in the dimuon rest frame and the vector perpendicular to the B momentum and the beam axis ($\cos P$),
- the B isolation based on the CDF definition ($I(B)$),
- angle between the B candidate's momentum and the thrust momentum of the B , defined as the sum of momenta of all the long tracks coming from the B PV and excluding those coming from long lived particles. If no such tracks are available, the variable is set to 0 (other B angle),
- angle between the direction of the positive muon candidate in the rest frame of the B and the thrust momentum in the B rest frame (B boost),
- absolute value of the difference between the pseudorapidity of the two muon candidates ($\Delta\eta$),
- absolute value of the difference between the spherical ϕ coordinate of the two muon candidates ($\Delta\phi$).

As shown in Section 4, only the variables $DOCA$ and $IP(B)$ are in common with the BDTS used in the selection. The exact definitions of the isolation variables is given in App. B.

The BDT training has been done using $B_s^0 \rightarrow \mu^+ \mu^-$ and $b\bar{b} \rightarrow \mu\mu X$ simulated events produced in 2012.

The agreement in all the variables between data sidebands and MC $b\bar{b} \rightarrow \mu^+ \mu^- X$ decays has been checked for the two data reconstructions i.e. Reco 13 with Stripping 19 and Reco 14 with Stripping 20, and with two different MC versions, i.e. MC10 and MC12. The comparison of all twelve variables is shown in Fig. 19, Fig. 20 and Fig. 21. The overall agreement between data and simulation is rather good, with the exception of the isolation variables, and to some extent the $p_T(B)$. Moreover, there is no appreciable difference in the data/simulation comparison between stripping 19 and stripping 20 data. In terms of signal the comparison is done with sideband-subtracted $B_{(s)}^0 \rightarrow h^+ h^-$ data which is compared for Reco 14 Stripping 20 and Reco 14 Stripping 20r1 to MC10 and MC12 simulation (cf. Fig. 22 to 24). Overall there is a good agreement between data and simulation. The only exception are again the isolation variables where the different running conditions between data and simulation (higher average number of collisions) lead to a worse isolation in data.

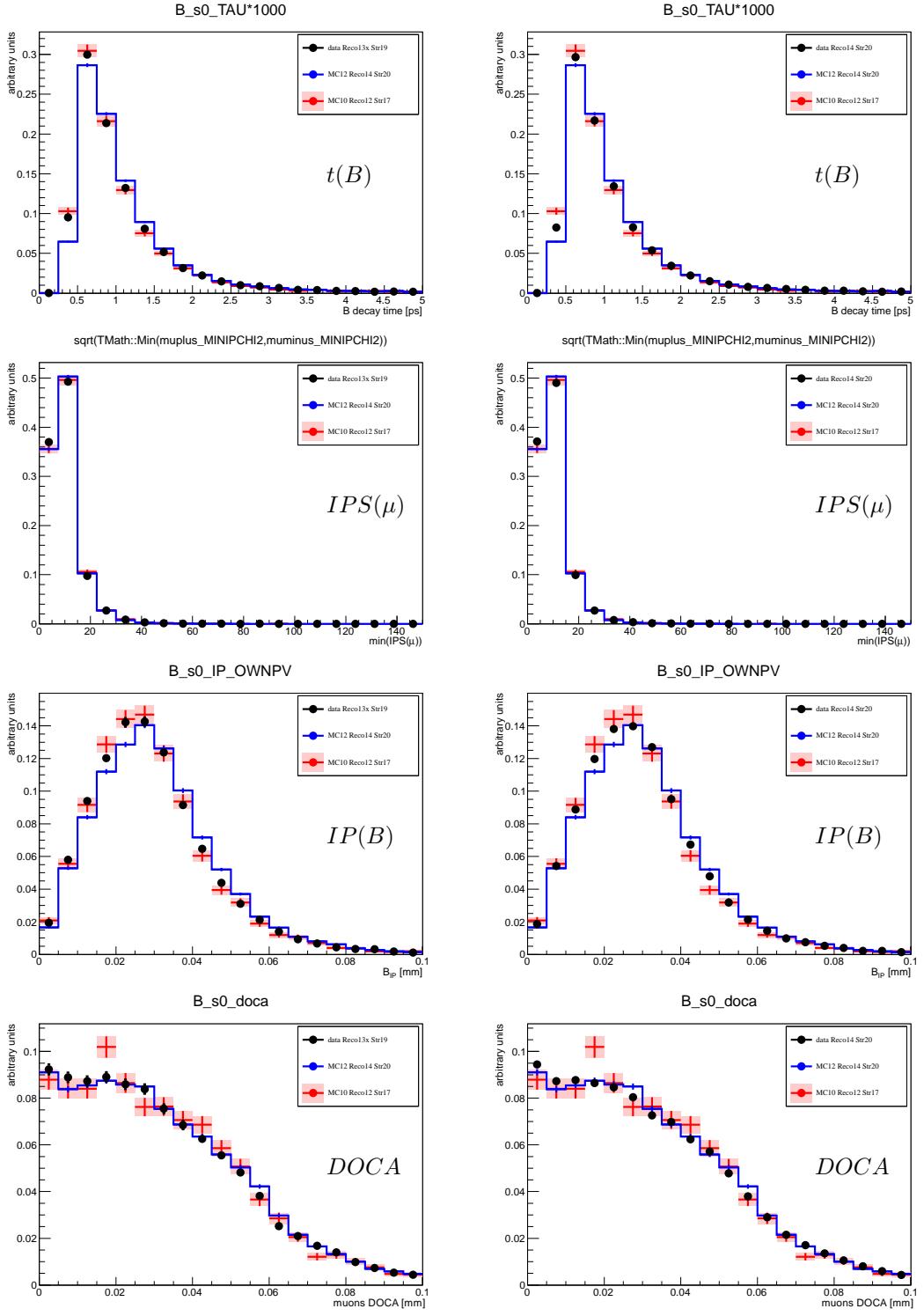


Figure 19: Comparison of BDT input variables between data sidebands and MC $b\bar{b} \rightarrow \mu^+\mu^-X$ decays for Reco13 Stripping19 data (left) and Reco14 Stripping20 (right). The MC10 (red) and MC12 (blue) versions of the simulation are also shown in both cases. The data is required to have passed any physics trigger while the MC is required to have passed any muon trigger.

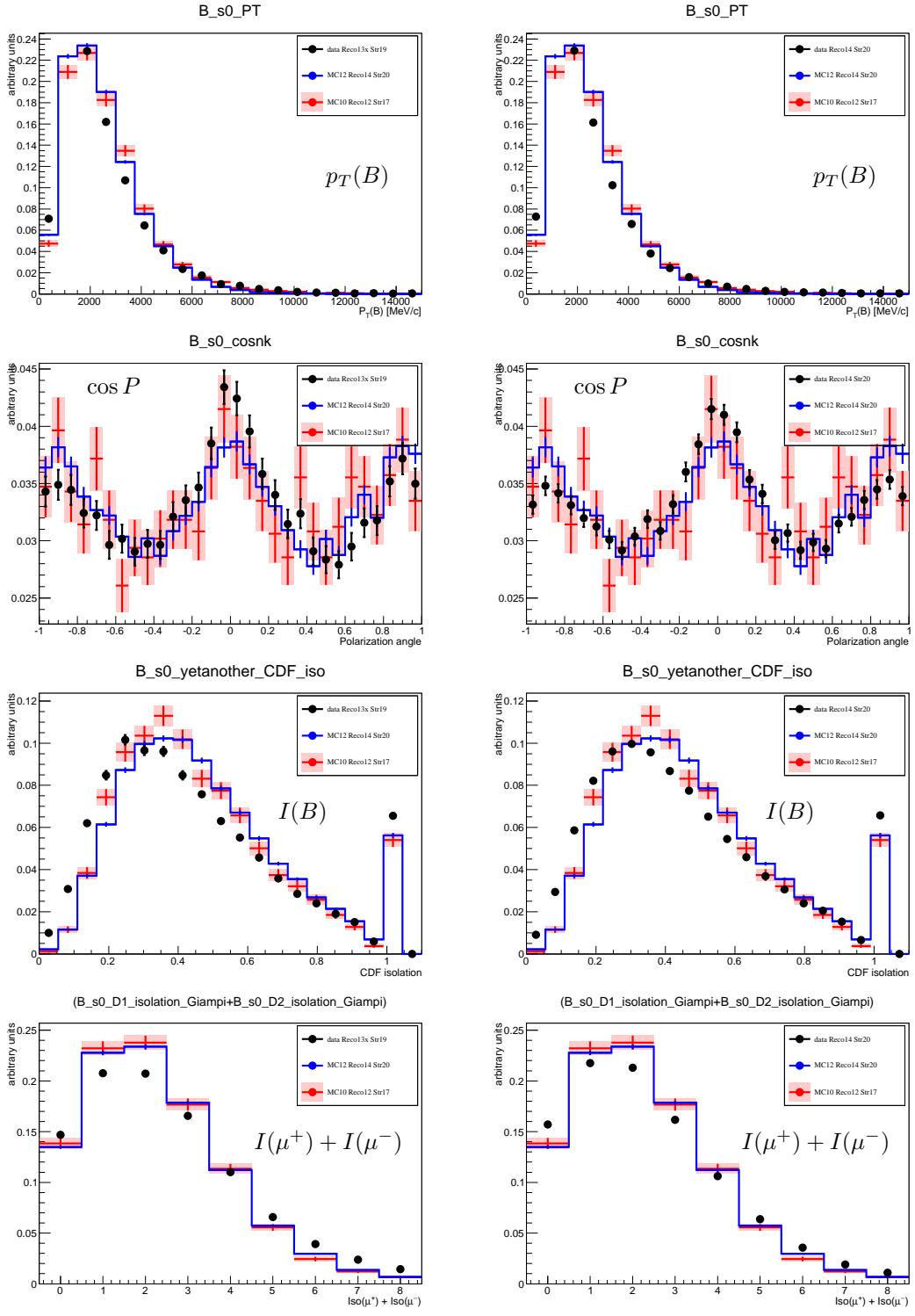


Figure 20: Comparison of BDT input variables between data sidebands and MC $b\bar{b} \rightarrow \mu^+\mu^-X$ decays for Reco13 Stripping19 data (left) and Reco14 Stripping20 (right). The MC10 (red) and MC12 (blue) versions of the simulation are also shown in both cases. The data is required to have passed any physics trigger while the MC is required to have passed any muon trigger.

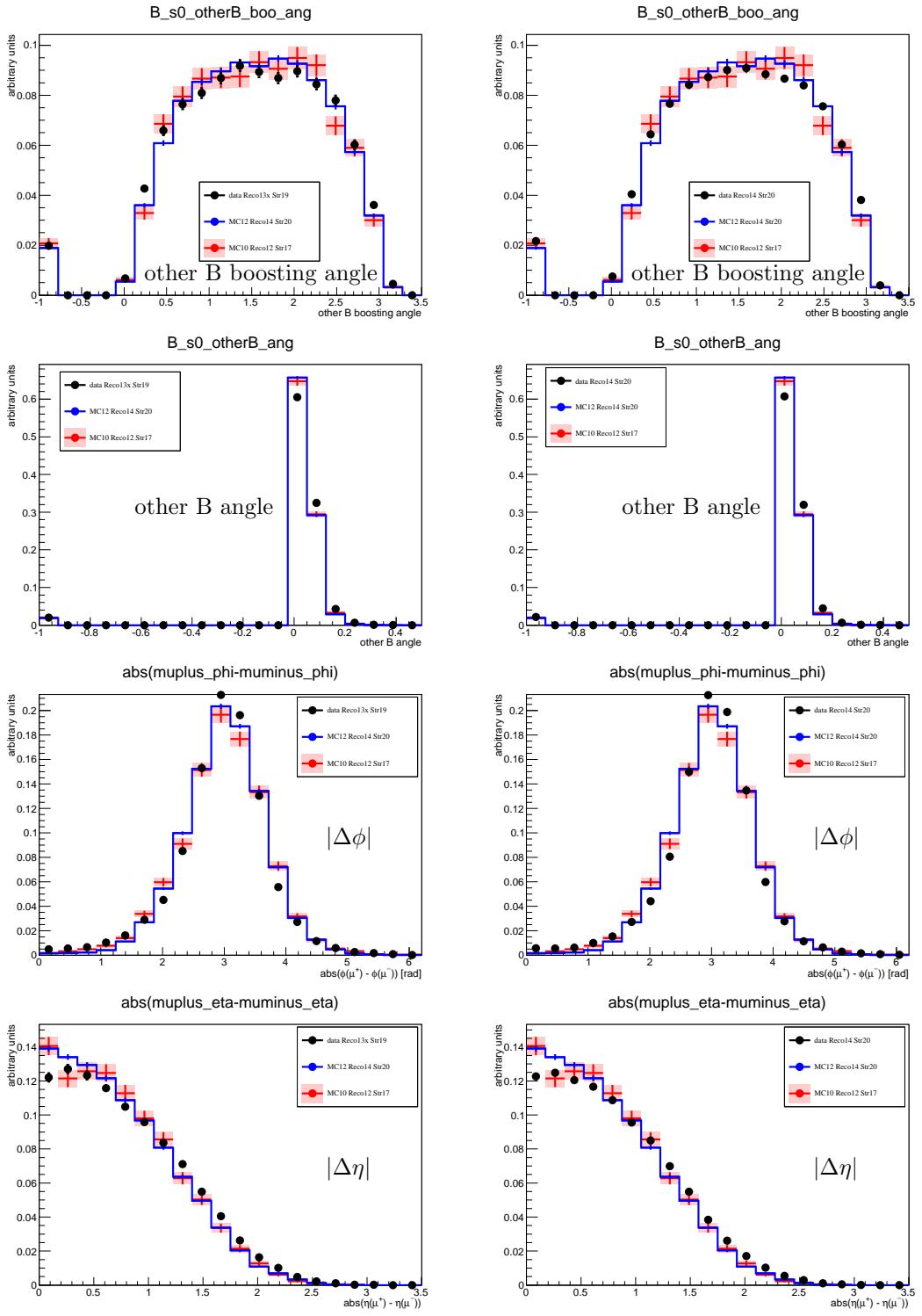


Figure 21: Comparison of BDT input variables between data sidebands and MC $b\bar{b} \rightarrow \mu^+\mu^-X$ decays for Reco13x Stripping19 data (left) and Reco14 Stripping20 (right). The MC10 (red) and MC12 (blue) versions of the simulation are also shown in both cases. The data is required to have passed any physics trigger while the MC is required to have passed any muon trigger.

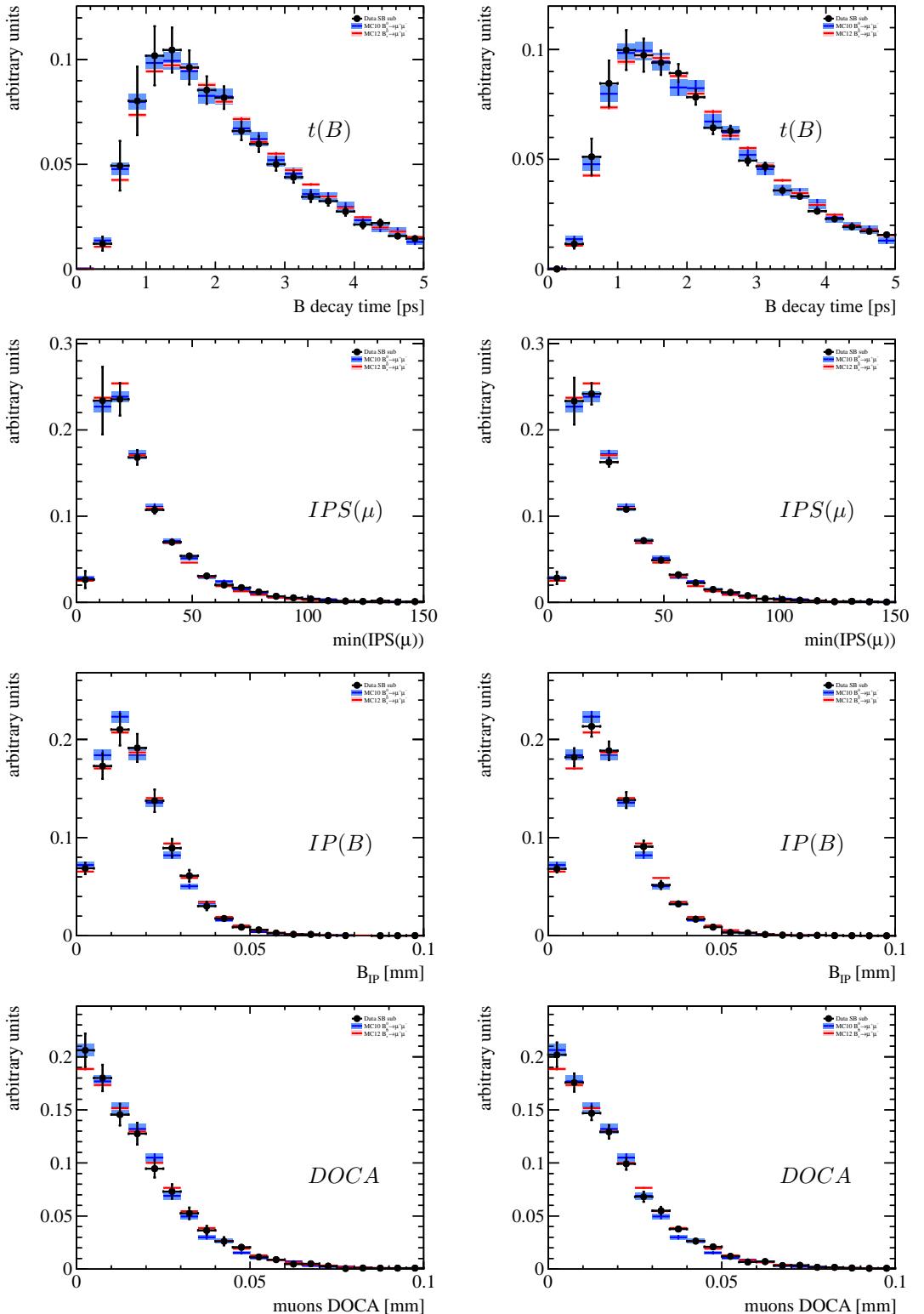


Figure 22: Comparison of BDT input variables between sideband subtracted $B_{(s)}^0 \rightarrow h^+h^-$ Reco 14 Stripping 20r1 data (left) and Reco 14 Stripping 20 (right) data and $B_s^0 \rightarrow \mu^+\mu^-$ MC. Both MC versions MC10 (blue) and MC12 (red) are shown. For MC as well as data $BDTS > 0.5$ and L0GlobalTIS as well as Hlt1PhysTIS is required.

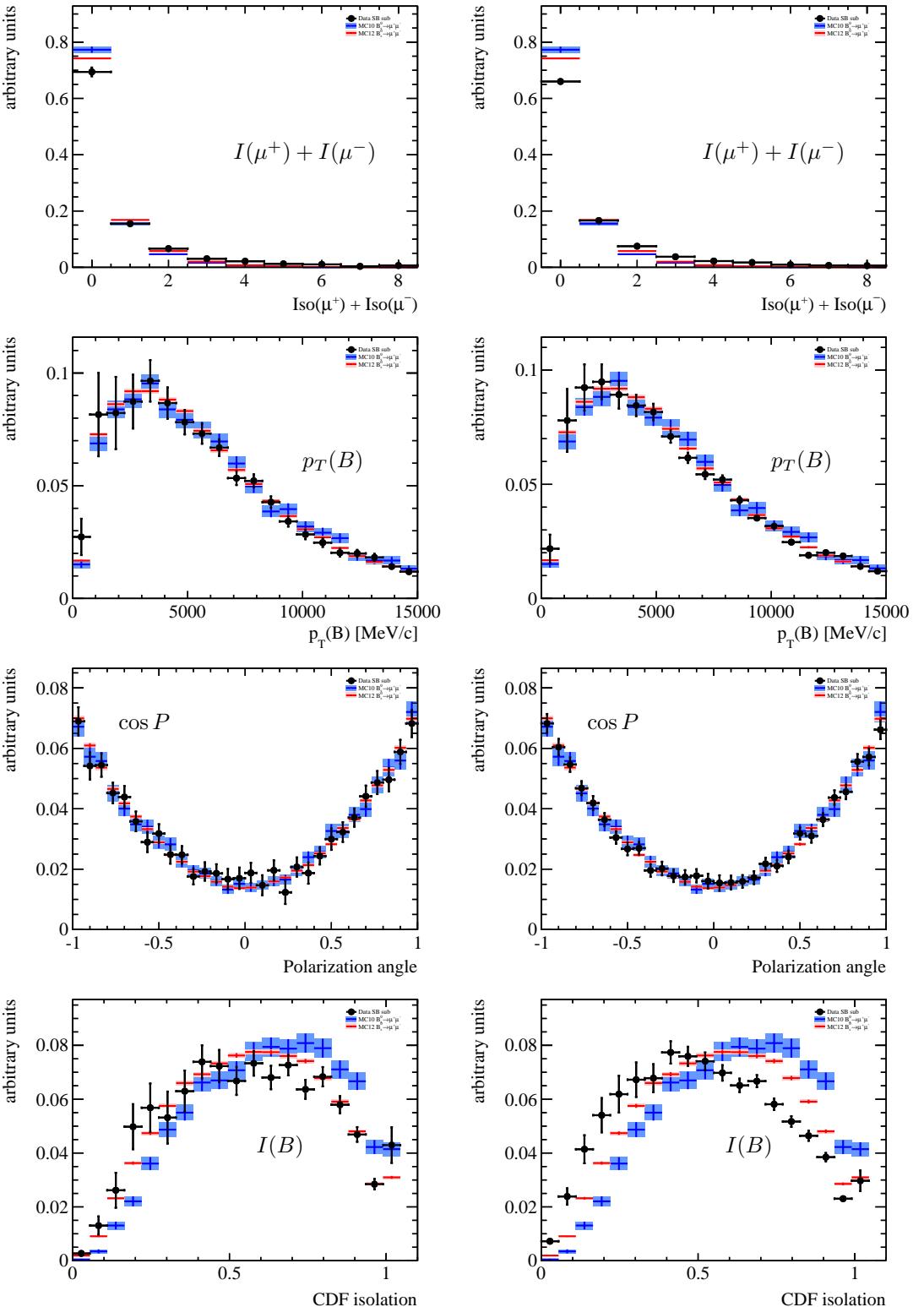


Figure 23: Comparison of BDT input variables between sideband subtracted $B_{(s)}^0 \rightarrow h^+h^-$ Reco 14 Stripping 20r1 data (left) and Reco 14 Stripping 20 (right) data and $B_s^0 \rightarrow \mu^+\mu^-$ MC. Both MC versions MC10 (blue) and MC12 (red) are shown. For MC as well as data BDTS > 0.5 and L0GlobalTIS as well as Hlt1PhysTIS is required.

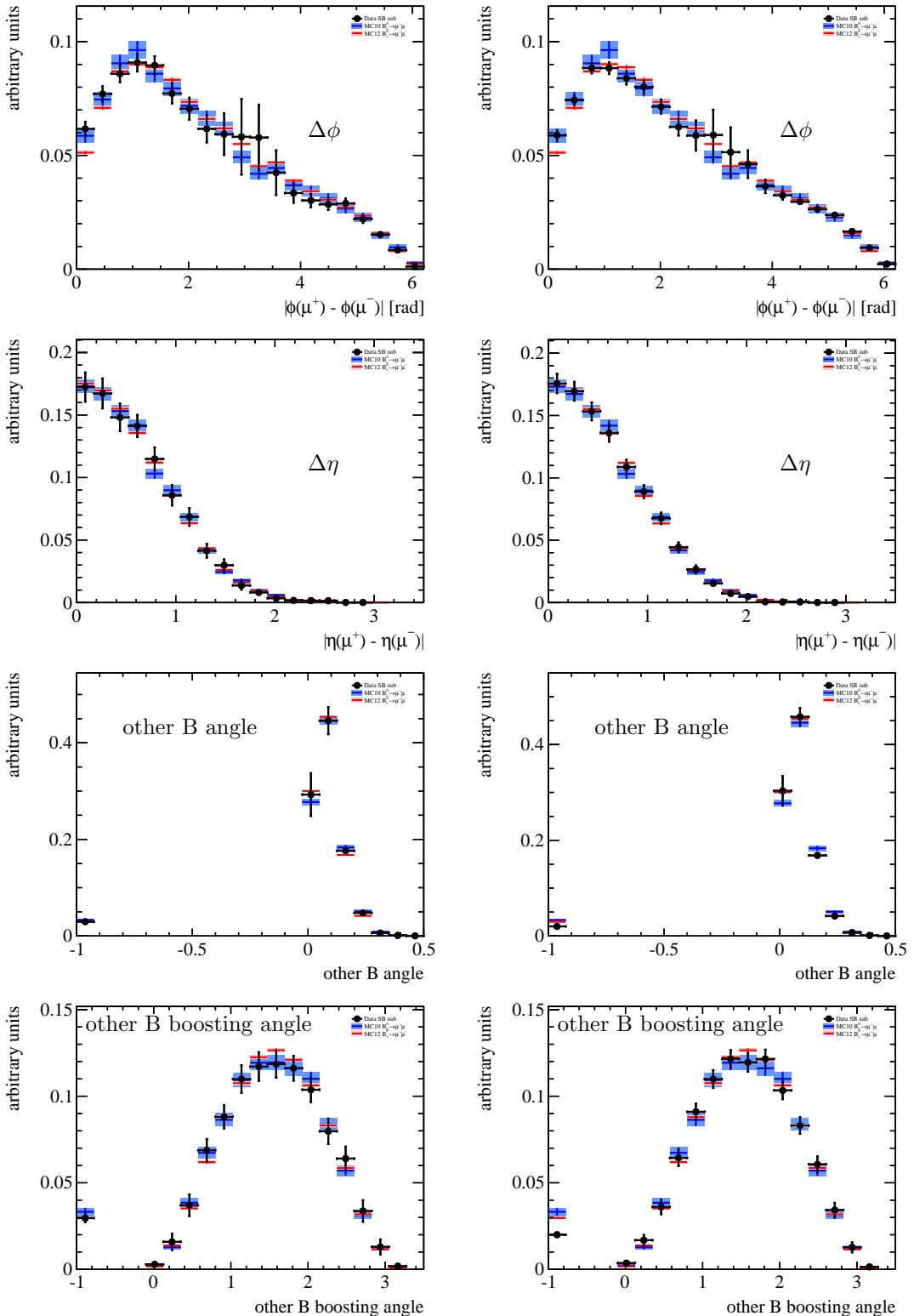


Figure 24: Comparison of BDT input variables between sideband subtracted $B_{(s)}^0 \rightarrow h^+h^-$ Reco 14 Stripping 20r1 data (left) and Reco 14 Stripping 20 (right) data and $B_s^0 \rightarrow \mu^+\mu^-$ MC. Both MC versions MC10 (blue) and MC12 (red) are shown. For MC as well as data $\text{BDTS} > 0.5$ and L0GlobalTIS as well as Hlt1PhysTIS is required.

7.1.1 Discussion about the BDT choice

This set of 12 input variables contains 8 of the 9 input variables of the BDT operator used for the last analysis (the minimum transverse momentum of the two μ (" $p_{T,min}(\mu)$ ") has been removed) with in addition the last four new variables.

To improve the performances of the MVA classifier used for the last analysis, an optimization of the tuning parameters of the old BDT classifier with 9 variables and of a classifier with 13 input variables (with the aforementioned 12 variables plus the $p_{T,min}(\mu)$) has been firstly tried. Even if the optimized operators perform better than the "old" BDT9, the output of the new BDTs is strongly correlated with the invariant mass of the two μ in the background sample. This correlation increases as the value of the N_{Depth}^{max} parameter increases (which at the same time gives better performances of the classifier), as can be seen from Figure 25.

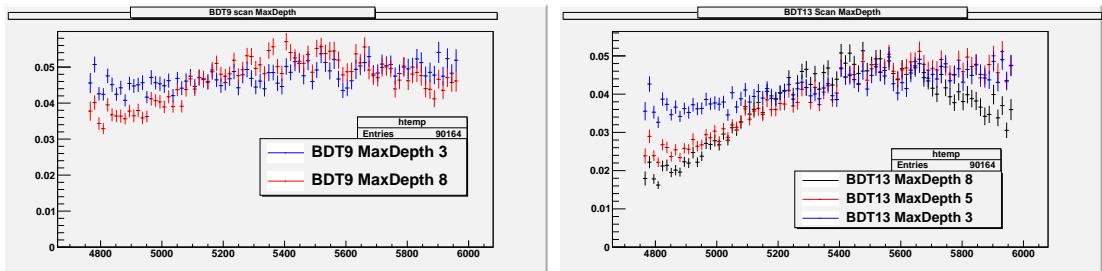


Figure 25: Left: average value of the BDT output as a function of the invariant mass of the two μ in the background sample and for a BDT classifier with the 9 input variables of the last BDT with two different values of the N_{Depth}^{max} tuning parameter: 3 (blue) and 8 (red); all the other tuning parameters are equals. Right: average value of the BDT output as a function of the invariant mass of the two μ in the background sample for a BDT classifier with 13 input variables (the 12 used for this analysis plus the $p_{T,min}(\mu)$) for three different values of N_{Depth}^{max} : 3 (blue), 5 (red) and 8 (black); all the other tuning parameters are equals

This correlation has a linear and *peaking* dependence on the invariant mass in the signal mass region. Such a peaking correlation could create a false peak under the signal mass region and hence can cause an underestimation of the background in the signal region when we extrapolate the mass shape of the combinatorial background from the mass sidebands. A linear correlation is less dangerous since has been checked that the fit procedure can deal properly with this bias up to correlation values of $\sim 10\%$.

This peaking correlation can be cured if the $p_{T,min}(\mu)$ variable is removed by the set of the BDT input variables. This is shown in the plots in Figure 26. In addition, also the effects of other input variables has been checked. As the plots in Figures 27 and 28 show, the peaking correlation still remains if the $p_{T,min}(\mu)$ is among the input variables.

It's worth to remark that this peaking correlation is due both to the algorithm we use (the BDT with high N_{Depth}^{max}) and to the correlating variable $p_{T,min}(\mu)$. Indeed such a

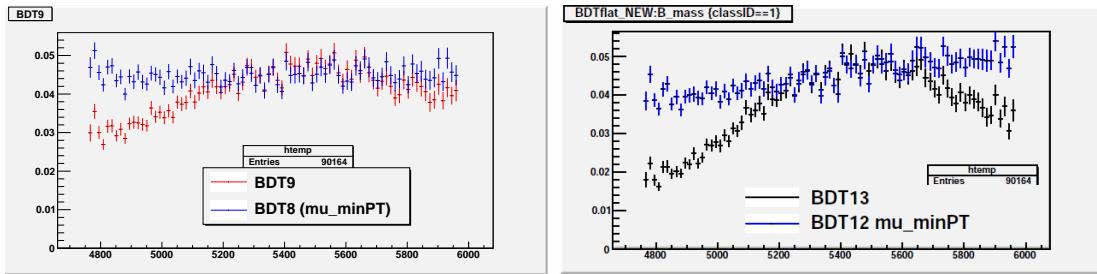


Figure 26: Left: average value of the BDT output as a function of the invariant mass of the two μ for the background sample and for a BDT classifier with the 9 input variables of the last BDT (red) and 8 input variables (blue) where the $p_{T,min}(\mu)$ has been removed. Right: average value of the BDT output as a function of the invariant mass of the two μ for the background sample and for a BDT classifier with the 12 input variables used in this analysis (blue) and 13 input variables (black) where the $p_{T,min}(\mu)$ has been added to the previous set

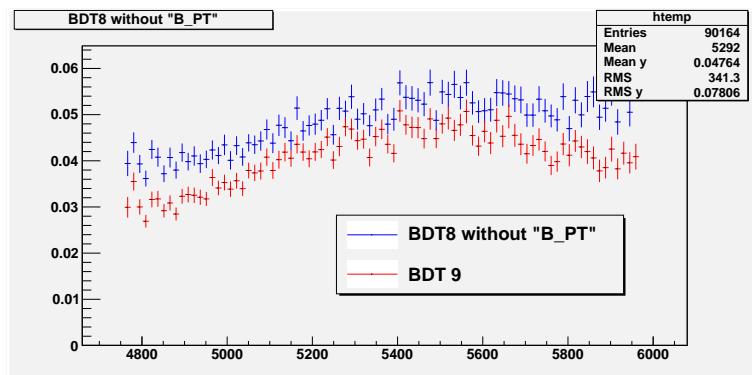


Figure 27: Average value of the BDT output as a function of the invariant mass of the two μ for the background sample and for a BDT classifier with the 9 input variables of the last analysis (red) and with 8 input variables where the transverse momentum of the B candidate has been removed (blue)

correlation is also present if we use the MatrixNet (MN) classifier (which is also based on BDT algorithm) with 13 input variables (Figure 29), but is absent if we use a different algorithm like a Neural Network classifier (MLP), as shown in Figure 30.

7.1.2 Performances of the new BDT12 on MC

The new BDT12, as the the BDT classifier used in the previous analysis (BDT9), has been trained using exclusively Monte Carlo samples. In Table 30 the MC samples used for the training and the test of the classifier are listed. The test sample is used for optimization and correlation studies.

The luminosity scale factor between the two background samples used in the training process is taken into account by scaling the $b\bar{b} \rightarrow \mu^+\mu^-X$ samples produced in the region of high PtProd by a factor 1.8.

The performances of the trained BDT are evaluated by using a Figure Of Merit (FOM).

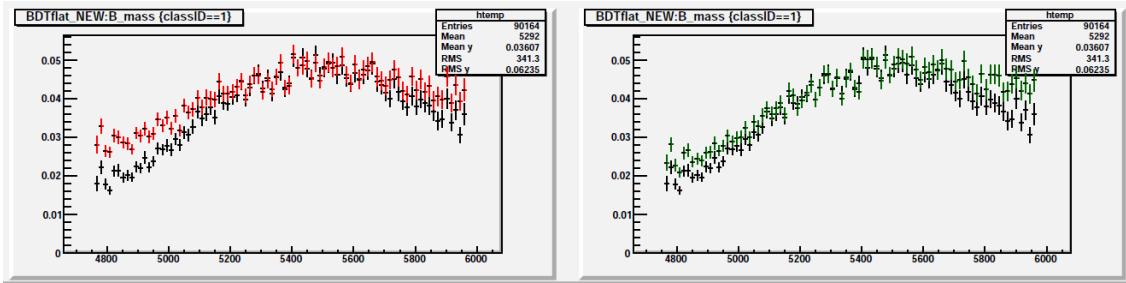


Figure 28: Left: average value of the BDT output as a function of the invariant mass of the two μ for the background sample and for a BDT classifier with the 13 input variables (black) and with 12 input variables where the $\Delta\eta$ variable has been removed (red). Right: average value of the BDT output as a function of the invariant mass of the two μ for the background sample and for a BDT classifier with the 13 input variables (black) and with 12 input variables where the $\Delta\phi$ variable has been removed (green)

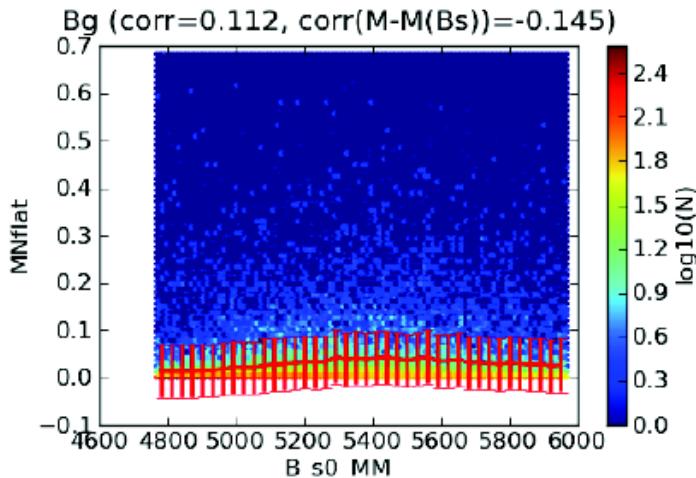


Figure 29: Average value of the MatrixNet classifiers with 13 input variables output as a function of the invariant mass of the two μ for the background sample

In the past we used the integral of the ROC (Receiver Operating Characteristic) curve as a FOM, but this doesn't take into account the importance of the sensitive bins. In order to overcome this problem a new FOM has been implemented including the sensitivity in each bin of the BDT9. The FOM evaluation can be summarized as follows:

- for each bin of the BDT under optimization we compute the expected number of background (n_b) and signal+background (n_{s+b}) events, from the efficiency which is evaluated on the signal and background test samples, and the total number of signal and background events expected in the 2012 1.1 fb^{-1} analysis;

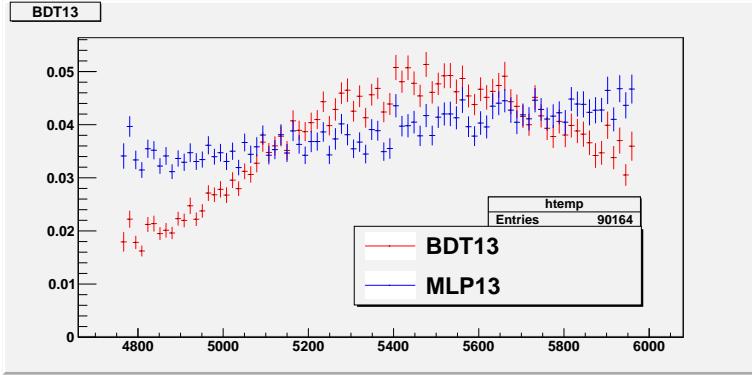


Figure 30: Average value of two different MVA classifiers output as a function of the invariant mass of the two μ for the background sample: BDT (red) and NeuralNetwork (blue) with the same 13 input variables)

- assuming these numbers fluctuating with a Poissonian probability we compute the following quantity for the i -th BDT bin:

$$\Delta LQ_i = 2 \cdot \log \left(\frac{\frac{Poiss(n_b, n_b)}{Poiss(n_{s+b}, n_b)}}{\frac{Poiss(n_b, n_{s+b})}{Poiss(n_{s+b}, n_{s+b})}} \right);$$

- the FOM is then given by the sum over all the BDT bins of ΔLQ_i .

Higher values of the FOM correspond to higher performances in discriminating background, especially in the most sensitive bins.

TMVA root package permits to modify the training parameters of the chosen classifier. These parameters allow to tweak the classifier performances and also avoid the overtrainig regime. In order to get the best performances we tested several training parameter sets and evaluated the FOM each time. The parameters modified in the optimization process are:

- **NTrees**: number of decision trees in the forest;
- **EvtsMin**: minimum number of events required in a leaf node;
- **MaxDepth**: maximum allowed depth of the decision tree;
- **nCuts**: number of steps during node cut optimization;
- **AdaBoostBeta**: parameter for AdaBoost algorithm.

Each training process required about 10 minutes to run. So that the training algorithm was run in parallel for several configurations at the same time. In Table 31 the

Table 30: MC Samples used for the training and the performance studies of the new BDT classifier.

Channel	Monte Carlo production	Number of events after all cuts	weight
Training			
$B_s^0 \rightarrow \mu^+ \mu^-$ $b\bar{b} \rightarrow \mu^+ \mu^- X, p > 3 \text{ GeV}/c,$ $4.7 < M_{\mu\mu} < 6.0 \text{ GeV}/c^2,$ $\text{doca} < 0.4 \text{ mm},$ $1 < \text{PtProd} < 16(\text{GeV}/c^2)^2$	MC2012/Sim06b MC2012/Sim06b	71939 70516	1.0 1.0
$b\bar{b} \rightarrow \mu^+ \mu^- X, p > 3 \text{ GeV}/c,$ $4.7 < M_{\mu\mu} < 6.0 \text{ GeV}/c^2,$ $\text{doca} < 0.4 \text{ mm},$ $\text{PtProd} > 16(\text{GeV}/c^2)^2$	MC2012/Sim06b	371	1.8
Test/Optimization			
$B_s^0 \rightarrow \mu^+ \mu^-$ $b\bar{b} \rightarrow \mu^+ \mu^- X, p > 3 \text{ GeV}/c,$ $4.7 < M_{\mu\mu} < 6.0 \text{ GeV}/c^2,$ $\text{doca} < 0.4 \text{ mm},$ $1 < \text{PtProd} < 16(\text{GeV}/c^2)^2$	MC2012/Sim06b MC2012/Sim06b	143810 25669	1.0 1.0
$b\bar{b} \rightarrow \mu^+ \mu^- X, p > 3 \text{ GeV}/c,$ $4.7 < M_{\mu\mu} < 6.0 \text{ GeV}/c^2,$ $\text{doca} < 0.4 \text{ mm},$ $\text{PtProd} > 16(\text{GeV}/c^2)^2$	MC2012/Sim06b	236	1.0

Table 31: Training parameter configurations with a FOM value greater than 32. The line highlighted in red is chosen as configuration for the new BDT.

NTrees	EvtsMin	MaxDep	nCuts	AdaBoostBeta	FOM
1000	2500	4	40	1	32.041999
1000	4500	6	30	0.5	32.595298
1000	4500	6	45	0.2000000	32.275600
1100	4500	6	30	0.5	32.138999
700	2500	4	40	1	32.4445
700	4500	6	30	0.5	32.081401
800	2500	4	40	1	32.066501
800	4500	7	45	0.3000000	32.154201
900	2500	4	40	1	32.063999
900	4500	6	45	0.2000000	32.267700

Table 32: Correlation values of the BDT with $m_{\mu\mu}$ and $|m_{\mu\mu} - M_{\text{PDG}}|$ in the whole mass range [4800-6000]. For signal the correlations are evaluated also in the ± 60 MeV/ c^2 mass range.

Sample	Correlation with $m_{\mu\mu}$	Correlation with $ m_{\mu\mu} - M_{\text{PDG}} $
$B_s^0 \rightarrow \mu^+ \mu^-$	5.3% (0.32%)	5.0% (0.26%)
$b\bar{b} \rightarrow \mu^+ \mu^- X$	1.2%	0.02%

training parameter sets are listed, which correspond to a FOM greater than 32. The best configuration, chosen for the new BDT, is reported in red in the same table.

The FOM of the new BDT, 32.6, has to be compared with that of the BDT9 which is equal to 18.6, as evaluated on the same MC test samples.

In Fig. 31 the ROC curves for the new BDT12 and the old BDT9 are shown. A large improvement is observed at low signal efficiency, which reflects directly into an improvement on the sensitivity of last BDT bins.

As reported in the previous sec. 7.1.1, the correlation with the invariant mass ($m_{\mu\mu}$) is also considered in the choice of the BDT classifier. A small correlation between mass and BDT prevents from underestimating the background in the signal region, and allows to factorize the signal pdf into a mass pdf and a BDT pdf. In Fig. 32 the BDT12 is shown as a function of the mass for MC signal and background events. The values of the correlation, both with respect to the mass and to the mass difference, $|m_{\mu\mu} - M_{\text{PDG}}|$ (M_{PDG} is the B_s^0 mass value as reported in the PDG) are listed in Table 32. The latter is very sensitive with respect to possible *peaking* dependence of the BDT12 output with the invariant mass.

The results show a negligible correlation for $B_s^0 \rightarrow \mu^+ \mu^-$ simulated events in the restricted signal window; the correlation observed for signal in the low mass region is due to the radiative tail. The correlation for $b\bar{b} \rightarrow \mu^+ \mu^- X$ events is negligible.

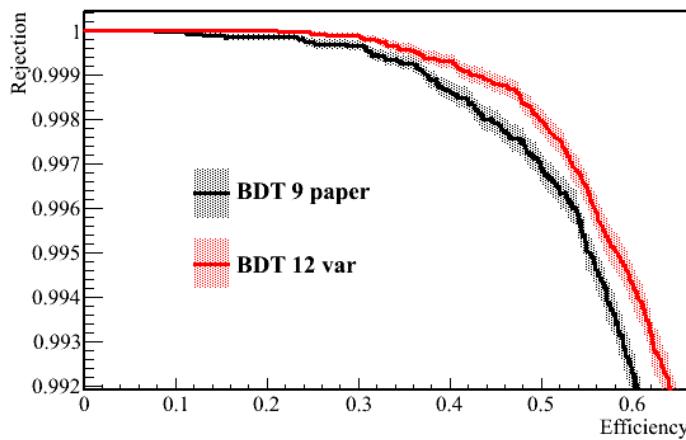


Figure 31: ROC curves for the BDT9 (black) and the new BDT12 (red)

To estimate the effect of this tiny linear correlation for background, we estimated the number of events in the B_s signal region ($m_{B_s} - 60 \text{ MeV}/c^2 \leq m_{\mu\mu} \leq m_{B_s} + 60 \text{ MeV}/c^2$) for each BDT bin fitting the MC background sidebands and compare it with the actual yields in the signal region. For this cross-check the test $b\bar{b} \rightarrow \mu^+\mu^-X$ MC samples reported in Tab. 30 have been used and three different models have been considered to fit the invariant mass:

- a single exponential PDF fitted to the mass range [4900-5000] MeV/c^2 & [5433-5973] MeV/c^2 ;
- a double exponential PDF fitted to the mass range [4773-5000] MeV/c^2 & [5433-5973] MeV/c^2 ;
- a single exponential PDF fitted to the mass range [5433-5973] MeV/c^2

In Fig. 33 only the first six bins are shown. The last two bins are not fitted because of the lack of statistics. In Table 33 the number of events in the B_s signal region by fitting the invariant mass distribution of $b\bar{b} \rightarrow \mu^+\mu^-X$ from the MC sample and the MC yields for the first six BDT bins are reported.

All the BDT bins show a good agreement between the background extrapolations and the MC yields; for the fourth bin the agreement is within 1.2σ , including the fluctuation of the observed events.

7.1.3 Comparison between BDT9 and BDT12 on data sidebands

The correlation between the outputs of the BDT9 and BDT12 has been checked in the sidebands of the full data sample after all the selection cuts. Figure 34 shows the comparison between the two BDT outputs (black points) with the profile distribution of the BDT values (blue points), and the curve with unit slope (red) superimposed. As the

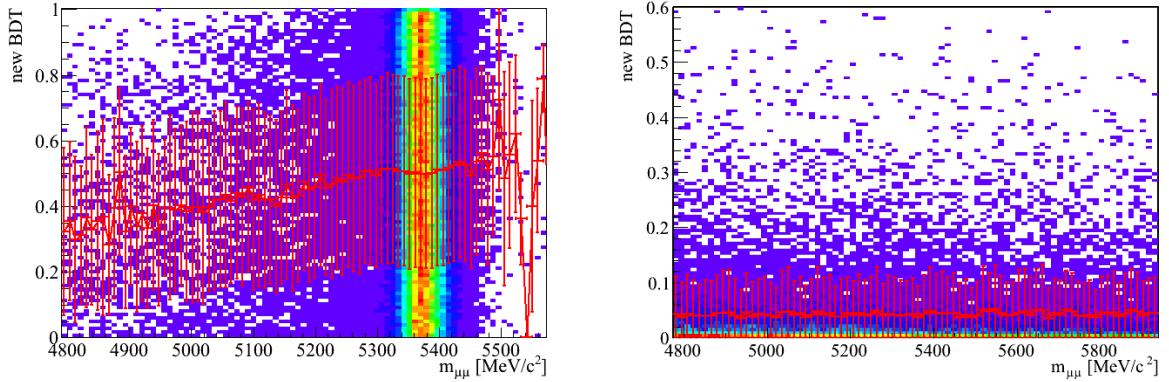


Figure 32: BDT12 output versus the invariant mass $m_{\mu\mu}$ for simulated $B_s^0 \rightarrow \mu^+\mu^-$ (left) and $b\bar{b} \rightarrow \mu^+\mu^-X$ (right); average profile histogram is also provided (red points).

blue points are all below the red curve, this means that the BDT12 returns a lower output than the BDT9 for the same background event. This translates into an increase of the sensitivity of this new operator with respect to the old one.

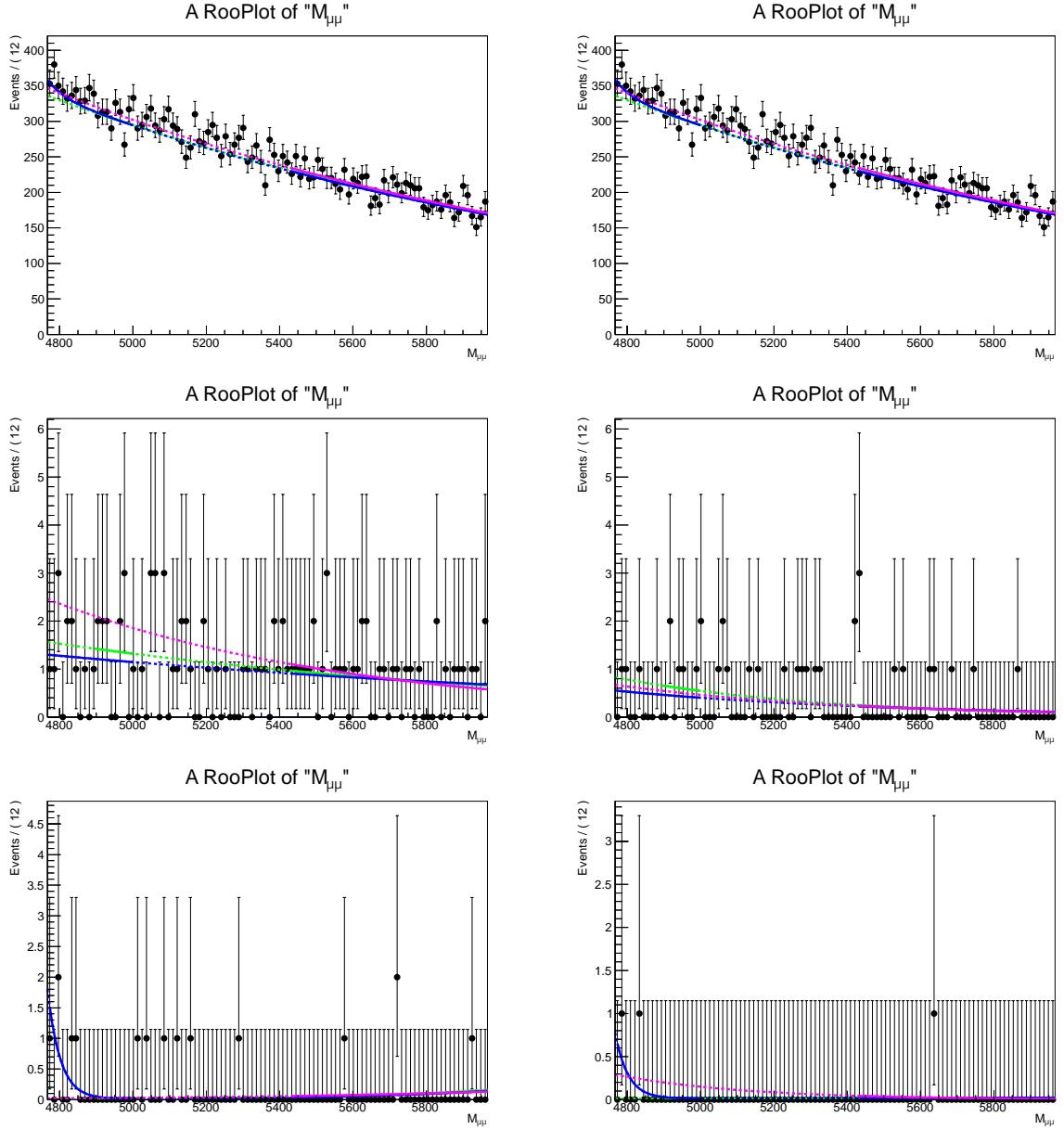


Figure 33: Invariant mass distribution of $b\bar{b} \rightarrow \mu^+\mu^- X$ from the MC sample in the first 6 bins of new BDT. Three different PDF are used for the fit: a single exponential fitted to the mass range [4900-5000] MeV/ c^2 & [5433-5973] MeV/ c^2 (green), a double exponential fitted to the mass range [4773-5000] MeV/ c^2 & [5433-5973] MeV/ c^2 (blue), a single exponential fitted to the mass range [5433-5973] MeV/ c^2 .

Table 33: Number of events obtained by interpolating the events in $b\bar{b} \rightarrow \mu^+\mu^-X$ MC sidebands and the MC yields for each BDT bin.

bin	Single Exponential	Double Exponential	Single Exponential on Right Sideband	MC yields
1	2434 ± 25	2442 ± 25	2453 ± 58	2453
2	49.0 ± 3.5	47.4 ± 2.8	51 ± 8	45
3	10.3 ± 1.6	9.7 ± 1.2	12.2 ± 4.4	10
4	2.8 ± 0.8	2.4 ± 0.6	2.6 ± 2.1	6
5	0.36 ± 0.38	0.36 ± 0.37	0.53 ± 0.68	0
6	0.16 ± 0.22	0.57 ± 0.35	0.52 ± 0.93	0
7	//	//	//	0
8	//	//	//	0

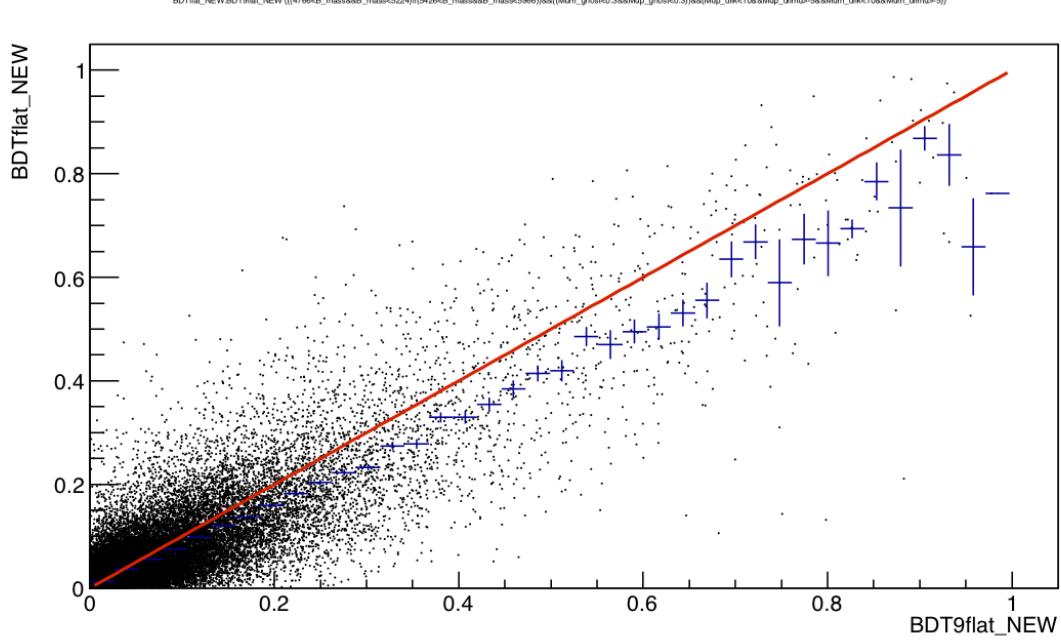


Figure 34: Comparison between the BDT9 output and BDT12 output for the data sidebands after the selection. The red solid line shows the BDT09 equal to BDT12 curve, and the blue points represent the profile distribution i.e. the average of the BDT12 values for each BDT09 value.

7.2 BDT calibration

7.2.1 $\Delta LL_{K-\pi}$ efficiency evaluation

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.

To calibrate the BDT for signal we are relying on the exclusive channels $B \rightarrow h^+h^-$ being able to act as a proxy for $B_{(s)}^0 \rightarrow \mu^+\mu^-$ as they are also two-body decays of a neutral B meson.

The basic idea is to distinguish the different channels and therefore to identify the correct mass hypothesis of the final state particles by applying cuts on the $\Delta LL_{K-\pi}$ variable of the final state particles.

These cuts can induce biases as the separation power of the $\Delta LL_{K-\pi}$ variable is highly dependent on the kinematics of the final state particle and the multiplicity of the event. We correct for these biases by taking the efficiency of the $\Delta LL_{K-\pi}$ cuts as a function of the momentum of the particle, its pseudo rapidity and the number of *best tracks* (long tracks with clone removal) in the event into account.

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

- p [5,9.3,15.6,17.675,20,23,26,29.65,30,35,40,45,50,55,60,65,70,75,80,85,90,95,100,125,150,200] GeV/c (25 bins)
- η [1.5,2.5,3.0,3.5,4.0,5.0] (5 bins)
- nTracks [0,50,100,200,400,800] (5 bins)

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

So the tool returns three-dimensional maps of the efficiencies for kaons as well as for pions and also for different cuts on $\Delta LL_{K-\pi}$.

We use the channel $B^\pm \rightarrow J/\psi K^\pm$ to evaluate the output of the PIDcalibTool. We fold the efficiency table determined by the PIDcalibTool and the normalized distribution of p and η of the kaon as well as number of best tracks in the $B^+ \rightarrow J/\psi K^+$ sample described in Sec. 4 with additional cuts of $BDTS > 0.25$ and $|m_{\mu\mu} - m_{J/\psi}| < 60$ MeV/ c^2 . By integrating the resulting distribution we get the average efficiency $\langle \varepsilon_{\text{kaon}} \rangle$.

This is done with efficiency tables for $\Delta LL_{K-\pi} > \kappa$ where $\kappa \in [-10, 20]$ in steps of 0.5. Events where the kaon is outside the three-dimensional parameter space of the tables are not taken into account. This is a fraction of on average 4% of the signal events in $B^\pm \rightarrow J/\psi K^\pm$.

To compare we fit the $m_{K\mu\mu}$ mass distribution of the same $B^\pm \rightarrow J/\psi K^\pm$ sample by the

sum of two Crystal Ball functions – one with its tail to the left, the other with its tail to the right – (for the B^\pm signal) and an exponential function (for the combinatorial background). The efficiency is estimated from the ratio of signal events with and without the $\Delta LL_{K-\pi}$ cut extracted from the fit:

$$\varepsilon_{\text{fit}} = \frac{N_{\text{sig,cut}}}{N_{\text{sig}}}$$

The signal fit includes also $B^\pm \rightarrow J/\psi\pi^\pm$ events ($\frac{\mathcal{B}(B^\pm \rightarrow J/\psi\pi^\pm)}{\mathcal{B}(B^\pm \rightarrow J/\psi K^\pm)} = 0.048$) and therefore $\langle \varepsilon_{\text{fit}} \rangle$ can be seen as

$$\varepsilon_{\text{fit}} = \frac{\langle \varepsilon_{\text{kaon}} \rangle \cdot \mathcal{B}(B^\pm \rightarrow J/\psi K^\pm) + \langle \varepsilon_{\text{pion}} \rangle \cdot \mathcal{B}(B^\pm \rightarrow J/\psi\pi^\pm)}{\mathcal{B}(B^\pm \rightarrow J/\psi K^\pm) + \mathcal{B}(B^\pm \rightarrow J/\psi\pi^\pm)}$$

We correct therefore ε_{fit} by calculating the efficiency $\langle \varepsilon_{\text{pion}} \rangle$ for pion by the PIDcalibTool in the same manner as it is done for $\langle \varepsilon_{\text{kaon}} \rangle$ and calculating the efficiency $\varepsilon_{\text{corr}}$ corrected for the π contribution which can be compared to $\langle \varepsilon_{\text{kaon}} \rangle$:

$$\varepsilon_{\text{corr}} = \varepsilon_{\text{fit}} + (\varepsilon_{\text{fit}} - \langle \varepsilon_{\text{pion}} \rangle) \frac{\mathcal{B}(B^\pm \rightarrow J/\psi\pi^\pm)}{\mathcal{B}(B^\pm \rightarrow J/\psi K^\pm)}$$

Fig. 35 shows the different efficiencies as a function of the $\Delta LL_{K-\pi}$ cut. The gray band at the bottom shows the difference $\Delta\varepsilon = \langle \varepsilon_{\text{kaon}} \rangle - \varepsilon_{\text{corr}}$. The maximal deviation for $\Delta LL_{K-\pi}$ cut between -10 and 20 is – depending on the Stripping and polarity – between 1% and 2%. The difference $\Delta\varepsilon$ as a function of the $\Delta LL_{K-\pi}$ cut is later used to estimate the systematics in the BDT calibration.

A second validation has been performed on the decay $B_s^0 \rightarrow J/\psi\phi$.

We calculate the efficiency for signal events defined as events with $\text{BDTS} > 0.05$ and lying in the invariant mass range $m_{\mu\mu KK} \in [5310, 5430] \text{ MeV}/c^2 \times m_{\mu\mu} \in [3050, 3150] \text{ MeV}/c^2 \times m_{KK} \in [1014.5, 1024.5] \text{ MeV}/c^2$ by taking the average of the product of the $\Delta LL_{K-\pi}$ efficiencies for the two kaons

$$\langle \varepsilon_{\text{Tool}} \rangle = \frac{\sum_i^{N_{\text{evt}}} \varepsilon(p, \eta, \text{nTr}|h_i^+) \cdot \varepsilon(p, \eta, \text{nTr}|h_i^-)}{N_{\text{evt}}}$$

We compare this value again to the signal yield of $B_s^0 \rightarrow J/\psi\phi$ with and without the $\Delta LL_{K-\pi}$ cut on the two kaons. These yields are extracted from a three-dimensional fit to $m_{KK\mu\mu}$, m_{KK} and $m_{\mu\mu}$ to separate combinatorial background and background from $B_s^0 \rightarrow J/\psi h^+ h^-$ from signal. In Fig. 36 we see the comparison between the calculation from the signal yields and from the PIDcalibTool. They are in a good agreement, but the errors are rather large due to the limited statistics. As this cross check is limited by the statistics of the sample, it is not used to estimate an additional systematic uncertainty.

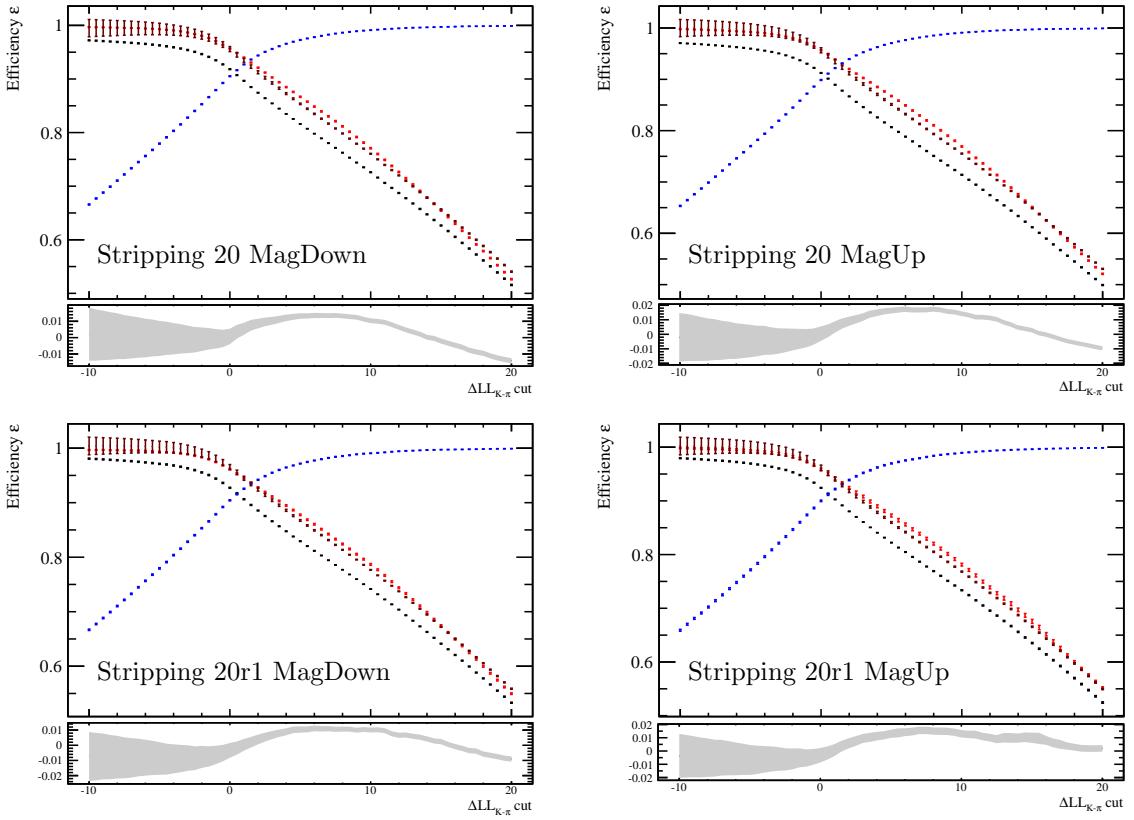


Figure 35: Comparison of $\Delta LL_{K-\pi}$ efficiency between output of PIDcalibTool and signal yields extracted from fit evaluated on $B^\pm \rightarrow J/\psi K^\pm$; Red: $\langle \varepsilon_{\text{kaon}} \rangle$, Black: ε_{fit} , Blue: $1 - \langle \varepsilon_{\text{pion}} \rangle$, Dark red: $\varepsilon_{\text{corr}}$; The grey band shows the difference between $\langle \varepsilon_{\text{kaon}} \rangle$ and $\varepsilon_{\text{corr}}$ with its error

7.2.2 Background from $\Lambda_b^0 \rightarrow ph$

As we do not use the PID information to separate kaons or pions from protons there will be a background component originating from the two-body decay $\Lambda_b^0 \rightarrow ph$ which will be present in the right sideband of the invariant mass distribution.

The description of this background component is taken from $\Lambda_b^0 \rightarrow p\pi^-$ and $\Lambda_b^0 \rightarrow pK^-$ MC with the same number of generated MC events. These truth-matched MC samples are created with the same selection cuts as the $B_{(s)}^0 \rightarrow h^+h^-$ data sample.

We assign to each event the mass hypothesis m_{hh} according to the true type of h being a kaon or pion. The proton is treated as a kaon if $\Delta LL_{K-\pi} > 0$ and else as a pion. The description of the PID in MC is not in perfect agreement with the one in data. Therefore we checked a possible systematic of this fact by changing the threshold to identify it as a kaon or pion to $\Delta LL_{K-\pi} > 2$ and $\Delta LL_{K-\pi} > -2$ with no significant change in the final description of the invariant mass distribution.

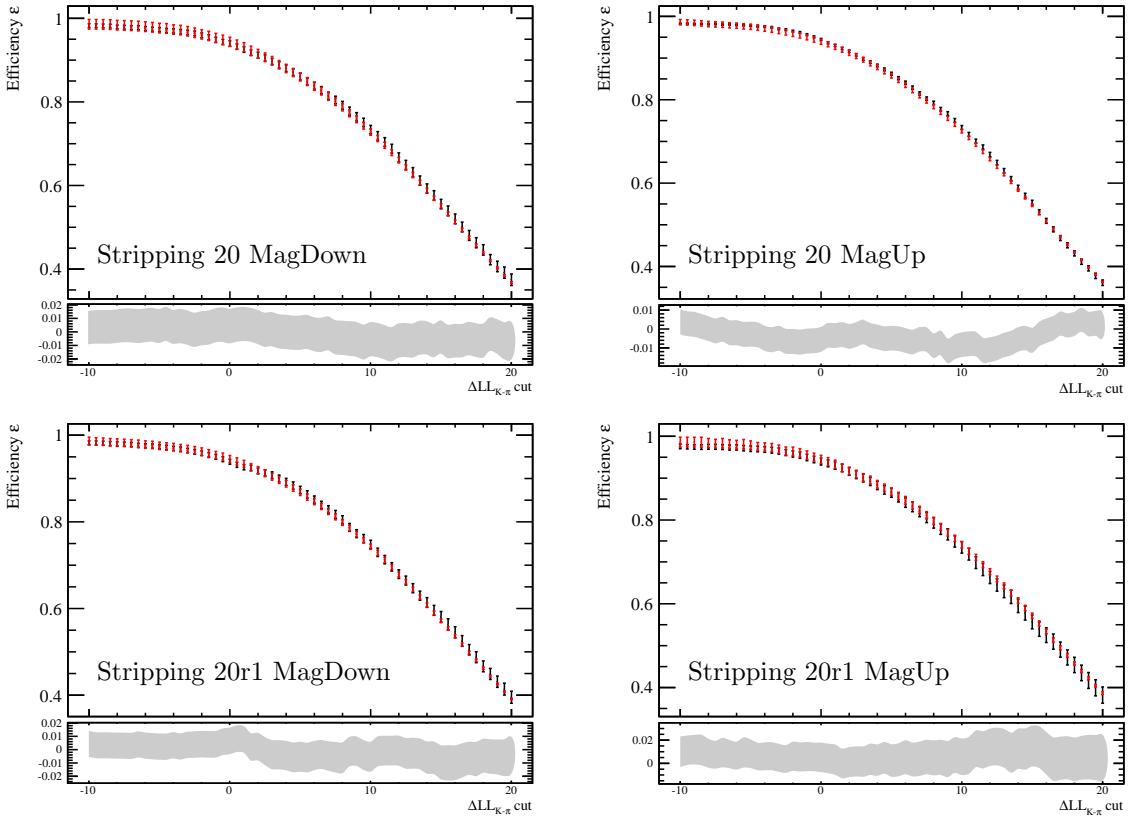


Figure 36: Comparison of $\Delta LL_{K-\pi}$ efficiency between output of PIDcalibTool and signal yields extracted from fit evaluated on $B_s^0 \rightarrow J/\psi\phi$; Red: $\langle \varepsilon_{\text{tool}} \rangle$, Black: $\langle \varepsilon_{\text{fit}} \rangle$; The grey band shows the difference between the two and its error

The events are weighted by

$$w = \frac{\mathcal{B}(\Lambda_b^0 \rightarrow ph)}{\mathcal{B}(\Lambda_b^0 \rightarrow p\pi) + \mathcal{B}(\Lambda_b^0 \rightarrow pK)}$$

to take differences in the branching fraction into account. The branching fractions are taken from PDG [38] with the values $\mathcal{B}(\Lambda_b^0 \rightarrow p\pi) = 3.5 \times 10^{-6}$ and $\mathcal{B}(\Lambda_b^0 \rightarrow pK) = 5.5 \times 10^{-6}$.

The resulting invariant mass distribution is then fitted with the sum of two Crystal Ball functions (one with a tail to the left, one with a tail to the right) as PDF where the two components have the same mean μ , but a different width $\sigma_{1,2}$. In total there are eight parameters (μ , $\sigma_{1,2}$, $\alpha_{1,2}$: transition points of the two tails, $n_{1,2}$: exponents of the two tails, f : relative fraction of the left-sided Crystal Ball function to the total PDF) which are extracted from the fit. Fig. 37 shows the invariant mass distribution with the fitted PDF. We see a very good description of the distribution.

In the later fits of the $B_{(s)}^0 \rightarrow h^+h^-$ mass distribution all parameters are fixed to

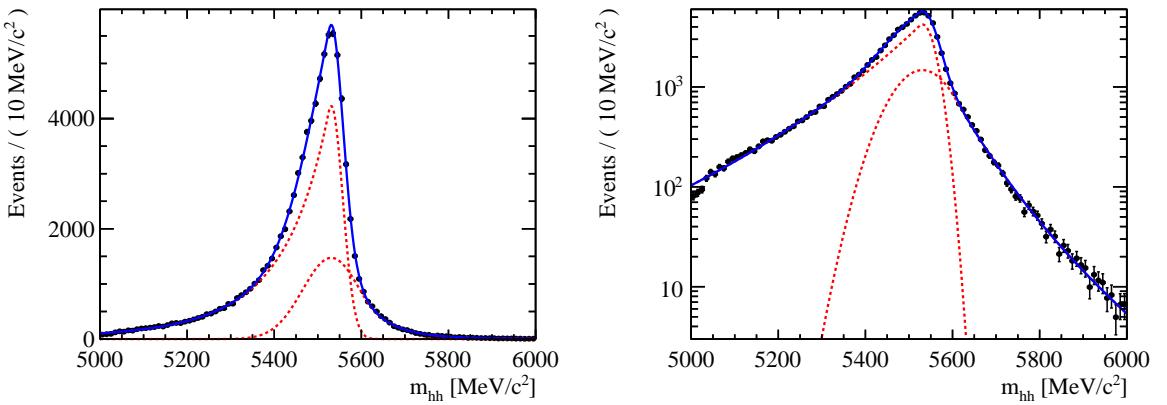


Figure 37: Invariant mass distribution of $\Lambda_b^0 \rightarrow pK$ and $\Lambda_b^0 \rightarrow p\pi$ from MC with a wrong mass hypothesis for the proton. The distribution is described by the sum of two Crystal Ball functions (one with a tail to the left, one with a tail to the right) with common mean, but different width.

the values extracted from this fit, except the mean which is allowed to vary within ± 10 MeV/ c^2 around the values determined from this fit. This should talk differences in the uncalibrated mass scale between data and MC into account.

7.2.3 BDT calibration with $B_{(s)}^0 \rightarrow h^+h^-$

The sample used to calibrate the BDT is the $B_{(s)}^0 \rightarrow h^+h^-$ sample defined in Sec. 4 with the following additional cuts:

- $0.5 \text{ GeV}/c < p_T(h) < 40 \text{ GeV}/c$
- both hadrons in the Muon acceptance
- L0Global TIS
- Hlt1Physics TIS

where the first one is applied to have an agreement between the p_T range of the hadrons in the $B_{(s)}^0 \rightarrow h^+h^-$ sample and in the sample used to calculate the $\Delta LL_{K-\pi}$ efficiency. As there is a good agreement of the invariant mass distribution for $B_{(s)}^0 \rightarrow h^+h^-$ between Stripping 20 (2012 data) and Stripping 20r1 (2011 data), we are performing a combined calibration of the BDT.

The total sample is divided into the eight bins in BDT: $[0.00, 0.25, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, 1.00]$. For each bin the invariant mass distribution of the two hadrons $m(hh)$ is created (cf. Fig. 38). We refer by m_{hh} henceforth to invariant mass values where the mass hypothesis for hadron (pions or kaons) are chosen by PID requirements. We identify a hadron as kaon if $\Delta LL_{K-\pi} > \kappa$ and as a pion if

$-\Delta LL_{K-\pi} > \kappa$ (if $|\Delta LL_{K-\pi}| < \kappa$ the event is rejected) and use the corresponding mass hypothesis of the two hadrons for the particular event.

We do not choose a specific value of κ , but consider values of κ between 0 and 10 in steps of 0.5.

The possible bias by the cuts on $\Delta LL_{K-\pi}$ is corrected by weighting each event by

$$w = \frac{1}{\varepsilon_{\Delta LL}(p, \eta, \text{nTr}|h^+, \kappa) \cdot \varepsilon_{\Delta LL}(p, \eta, \text{nTr}|h'^-, \kappa)} \quad (12)$$

where the efficiencies of the $\Delta LL_{K-\pi}$ cuts on pions and kaons are determined as described above by the PIDcalibTool in the same three-dimensional parameter space for the different cuts. $\varepsilon_{\Delta LL}(p, \eta, \text{nTr}|h, \kappa)$ is meant in case of a kaon as the $\Delta LL_{K-\pi}$ efficiency for $\Delta LL_{K-\pi} > \kappa$ while it is meant in case of a pion to be the $\Delta LL_{K-\pi}$ efficiency for $\Delta LL_{K-\pi} < -\kappa$.

The efficiency values $\varepsilon_{\Delta LL}(p, \eta, \text{nTr}|h, \kappa)$ are separately used for Stripping 20 and Stripping 20r1.

Events where one or both of the two hadrons lie outside the three-dimensional binning scheme are rejected. In addition also events where at least one of the hadrons has a $\Delta LL_{K-\pi}$ efficiency smaller than 2 % or the relative error on the $\Delta LL_{K-\pi}$ efficiency is larger than 50 %, are rejected to avoid the dominance of single events with a large weight. This rejection affects less than 2 % of the events and is as discussed later taken as a systematic uncertainty.

The resulting invariant mass distributions of the weighted events are fitted by two times (describing the B^0 and the B_s^0) the sum of two Crystal Ball functions (one with a tail to the left, one with a tail to the right) with common width (σ). Further we use an exponential function to describe the combinatorial background, the distribution discussed in Sec. 7.2.2 to describe possible contributions from $\Lambda_b \rightarrow p\pi/K$ and the RooPhysBkg distribution (Eq. (13)) to describe partially reconstructed B decays.

The RooPhysBkg distribution is defined as

$$\begin{aligned} f(m; m_0, c_p, \sigma_p) = & N \cdot \int_{-\infty}^{\infty} m' \left(1 - \frac{m'^2}{m_0^2}\right) \Theta(m_0 - m') \exp(-c_p \cdot m') \\ & \cdot \frac{1}{\sqrt{2\pi}\sigma_p} \exp\left(-\frac{(m - m')^2}{2\sigma_p^2}\right) dm' \end{aligned} \quad (13)$$

where Θ is the Heaviside theta-function and N is a normalization factor.

The number of B^0 and B_s^0 in the i -th BDT bin $N_{B_{(s)}^0, i}$ events is taken from the fit. For each value of κ we calculate the fraction of $B_{(s)}^0 \rightarrow h^+h^-$ events in the i -th BDT bin by

$$r_i(\kappa) = \frac{N_{B_{(s)}^0, i}(\kappa)}{\sum_j N_{B_{(s)}^0, j}(\kappa)} \quad (14)$$

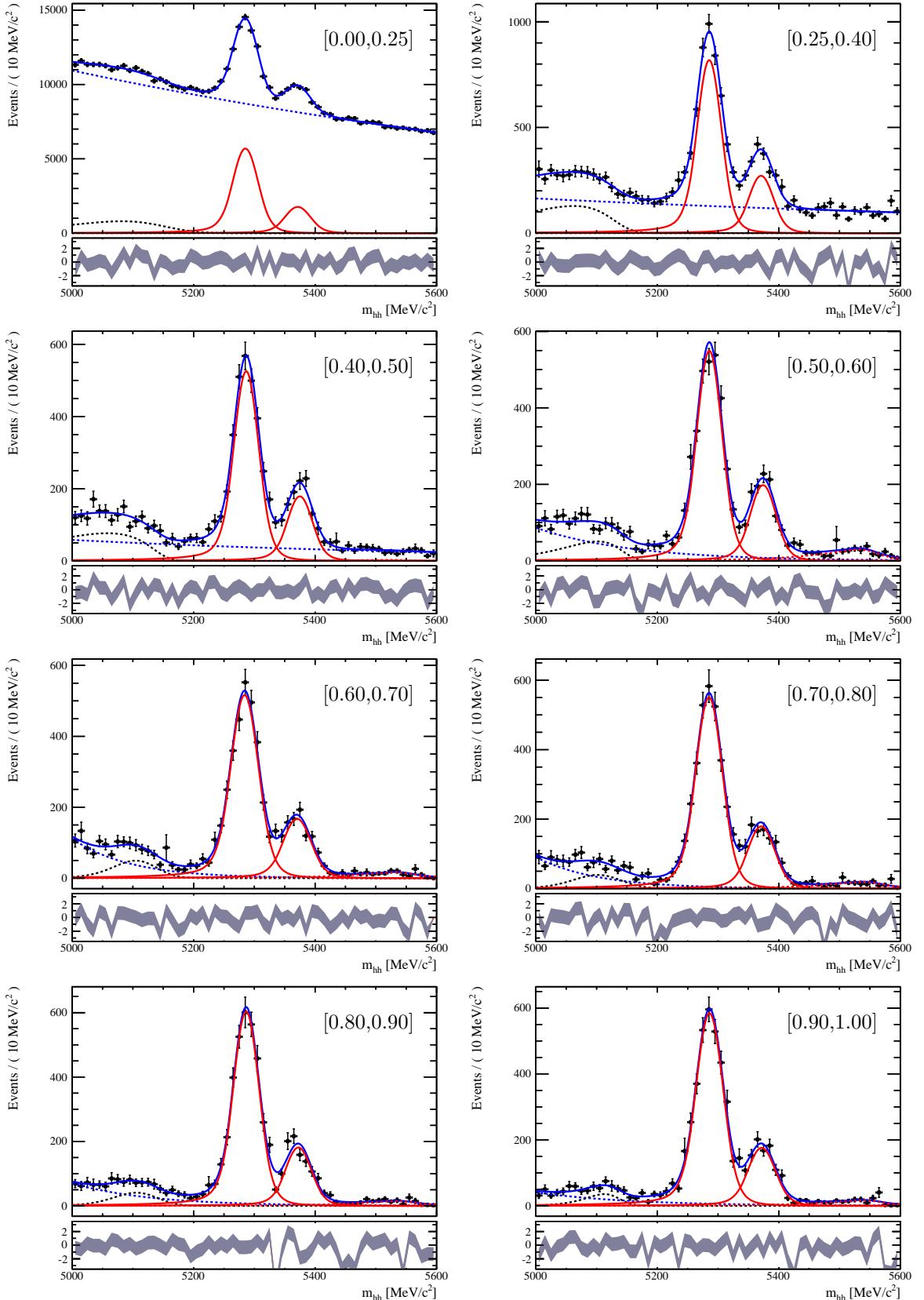


Figure 38: Invariant mass distributions of $B_{(s)}^0 \rightarrow h^+h^-$ from 2011 and 2012 data combined with $\Delta LL_{K-\pi}$ efficiency correction in different BDT bins for $\Delta LL_{K-\pi}$ cut value $\kappa = 5$. Each event in the distributions is weighted according to (12).

The red solid line shows the B^0 and B_s^0 signals, the red dashed one the one from $\Lambda_b^0 \rightarrow pK/\pi$ where the proton is misidentified as a kaon or pion. The combinatorial background is shown by the blue dashed line, while the physical background (partially-reconstructed) is shown by the black solid line.

Fig. 39 (left) shows the $r_i(\kappa)$ for the different BDT bins and cut values. We see that the fraction per bin is stable over the range of $\Delta LL_{K-\pi}$ cuts. The first bin (red data points) shows a larger variation. This is caused by the additional background in this bin causing larger uncertainties on the yield. But the larger background is also leading to a smaller correlation between the data points for different $\Delta LL_{K-\pi}$ cuts as the sample size is more reduced by the $\Delta LL_{K-\pi}$ cuts.

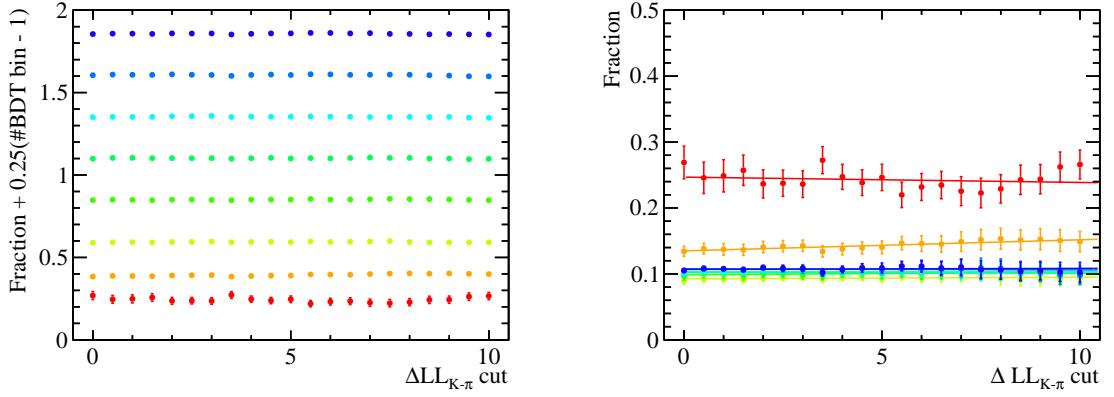


Figure 39: Fraction r_i of signal events from $B^0_{(s)} \rightarrow h^+h^-$ in the i -th BDT bin as a function of the $\Delta LL_{K-\pi}$ cut Left: For better visualization the values are shifted by $0.25 \cdot (i - 1)$; Right: Data points for r_i are fitted with a linear function to detect systematic drifts in the fraction over the $\Delta LL_{K-\pi}$ range as described below.

For the signal BDT PDF we calculate per BDT bin the weighted average \bar{r}_i over the data points for the whole range of $\Delta LL_{K-\pi}$ cuts where we take the correlation between $r_i(\kappa)$ for different κ values into account:

$$\begin{aligned}\bar{r}_i &= \sigma_{\bar{r}_i}^2 \sum_a c_{i;a,b}^{-1} r_{i,a} \\ \sigma_{\bar{r}_i}^2 &= \frac{1}{\sum_{a,b} c_{i;a,b}^{-1}}\end{aligned}$$

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

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

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

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

$$\begin{aligned}c_{i;a,b} &= \sigma_{r_{i,a}} \cdot \sigma_{r_{i,b}} \cdot \rho_{i,a,b} \\ \text{with } \sigma_{r_{i,a}} &: \text{statistical error on } r_{i,a} \text{ propagated from (14)}\end{aligned}$$

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

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

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

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

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

Fig. 40 shows the signal BDT PDF with the statistical uncertainty ($\sigma_{\bar{r}_i}$) as red band and the systematic uncertainty as grey band for the different BDT bins. The results are summarized in Tab. 34.

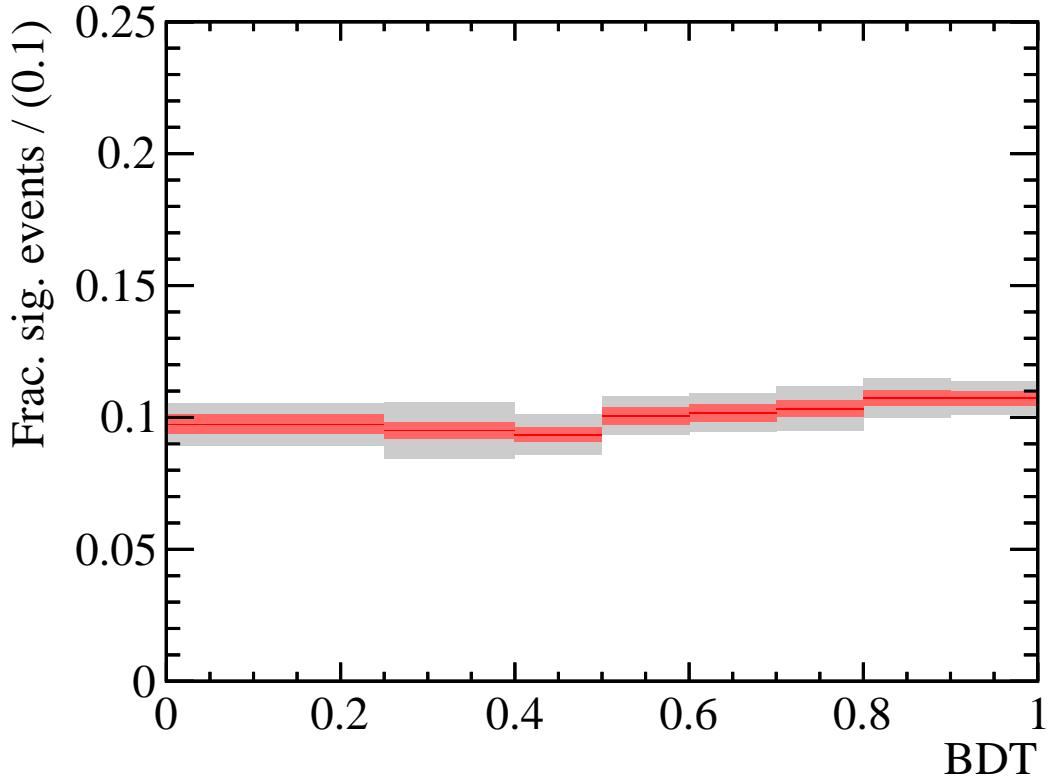


Figure 40: Signal PDF of the BDT. The red band shows the statistical uncertainty, the grey one the systematic one. Values are normalized to a bin size of 0.1.

Table 34: Summary of signal BDT PDF (value with statistical and systematic uncertainty)

BDT bin	value	stat.	syst.
[0.00, 0.25]	0.2437	0.0095	0.0105
[0.25, 0.40]	0.1425	0.0038	0.0115
[0.40, 0.50]	0.0935	0.0028	0.0050
[0.50, 0.60]	0.1006	0.0034	0.0040
[0.60, 0.70]	0.1018	0.0033	0.0039
[0.70, 0.80]	0.1033	0.0031	0.0054
[0.80, 0.90]	0.1072	0.0029	0.0046
[0.90, 1.00]	0.1074	0.0027	0.0037

The following effects have been taken into account when calculating the systematic uncertainty:

1. **Consistency of the data points along the $\Delta LL_{K-\pi}$ range:** We perform a linear χ^2 fit to the data points in the $\Delta LL_{K-\pi}$ range for each BDT bin taking the correlations into account (cf. Fig. 39 (right)). From this fit we calculate the extrapolated value for $\kappa = 0$ r_{inter} . If this value is significant different from the baseline value \bar{r} , we take the difference as systematic uncertainty. Significant means that $|r_{\text{inter}} - \bar{r}| > \sqrt{\sigma_{\text{inter}}^2 + \sigma_{\bar{r}}^2}$ where σ_{inter} is the uncertainty extracted from the fit on the extrapolated value.
2. **Binning scheme of $\Delta LL_{K-\pi}$ efficiency determination:** The whole determination of \bar{r}_i is repeated with using a finer binning scheme:
 - p : [5,9.3,12,15.6,17.675,20,21.5,23,24.5,26,27.5,29.65,30,32.5,35,37.5,40,42.5,45,47.5, 50,52.5,55,57.5,60,62.5,65,67.5,70,72.5,75,77.5,80,82.5,85,87.5,90,92.5,95,97.5, 100,110,125,135,150,170,200] GeV/c (46 bins)
 - η : [1.5,2.0,2.5,3.0,3.5,4.0,4.5,5.0] (7 bins)
 - nTracks: [0,50,75,100,150,200,400,800] (7 bins)

to calculate the $\Delta LL_{K-\pi}$ efficiency. We take the difference of \bar{r}_i as systematic uncertainty.

3. **Uncertainty on the $\Delta LL_{K-\pi}$ efficiency:** We take the difference between the efficiency calculated with the PIDcalibTool and from the fitted yields as a function of the $\Delta LL_{K-\pi}$ cut ($\Delta \varepsilon(\kappa)$) and add it to the efficiency when calculating the weighting:

$$w = \frac{1}{(\varepsilon_{\Delta LL}(p, \eta, \text{nTr}|h^+, \kappa) + \Delta \varepsilon(\kappa)) \cdot (\varepsilon_{\Delta LL}(p, \eta, \text{nTr}|h'^-, \kappa) + \Delta \varepsilon(\kappa))}$$

and recalculate the signal PDF with this adapted weighting.

The difference to the baseline PDF in Tab. 34 is taken as systematic uncertainty.

4. **Skipped events:** We assume that all the events we have rejected due to the too small $\Delta LL_{K-\pi}$ efficiency ($< 2\%$) or the too high relative uncertainty on the efficiency ($> 50\%$) have been signal events and recalculate the signal PDF and take the difference to the baseline one. The ratio of skipped events is similar over the different BDT bins.
5. **Fit model of m_{hh} distribution:** We repeated the analysis with a modified fit model: We described the combinatorial background by two exponential functions instead of one and used the sum of the Gaussians with common mean but different width to describe the B^0 and the B_s^0 signal.

The invariant mass distributions together with modified fit model are shown in Fig. 41.

The difference between the values extracted with the baseline fit model and this modified fit model are taken as systematic uncertainty.

The total systematic uncertainty is calculated by taking the square root of the sum of the squared individual systematic uncertainties. Tab. 35 summarizes the systematic uncertainties for the different BDT bins.

Table 35: Summary of systematic uncertainties on the signal BDT PDF for 2011 and 2012 data combined

BDT bin	Consistency	Binning scheme	$\Delta LL_{K-\pi}$	Skipped events	Fit model	Total
[0.00, 0.25]	0.0053	0.0035	0.0040	0.0053	0.0050	0.0105
[0.25, 0.40]	0.0093	0.0024	0.0041	0.0023	0.0039	0.0115
[0.40, 0.50]	0.0000	0.0022	0.0032	0.0012	0.0029	0.0050
[0.50, 0.60]	0.0000	0.0016	0.0017	0.0010	0.0031	0.0040
[0.60, 0.70]	0.0000	0.0015	0.0024	0.0015	0.0023	0.0039
[0.70, 0.80]	0.0028	0.0025	0.0027	0.0014	0.0024	0.0054
[0.80, 0.90]	0.0032	0.0014	0.0013	0.0021	0.0017	0.0046
[0.90, 1.00]	0.0000	0.0009	0.0023	0.0018	0.0021	0.0037

We see that the systematic uncertainty is the highest for the second BDT bin which is driven by the uncertainty associated to the significant drift over the considered $\Delta LL_{K-\pi}$ cut range being 0.0093.

7.2.4 BDT shape correction from DLL cut efficiency

As shown in Sections 5 and 6, the double misID background rejection greatly benefits from the introduction of a loose DLL selection cut ($DLL(K - \pi) < 10 \&\& DLL(\mu - \pi) > -5$, PID cut), at a cost of $\sim 3\%$ loss of signal events, as described in Sect. 5. Though the effect is small compared to other effects previously discussed, nevertheless we studied the behaviour of the DLL cut efficiency on signal, ϵ_{PID} , as a function of BDT in order to

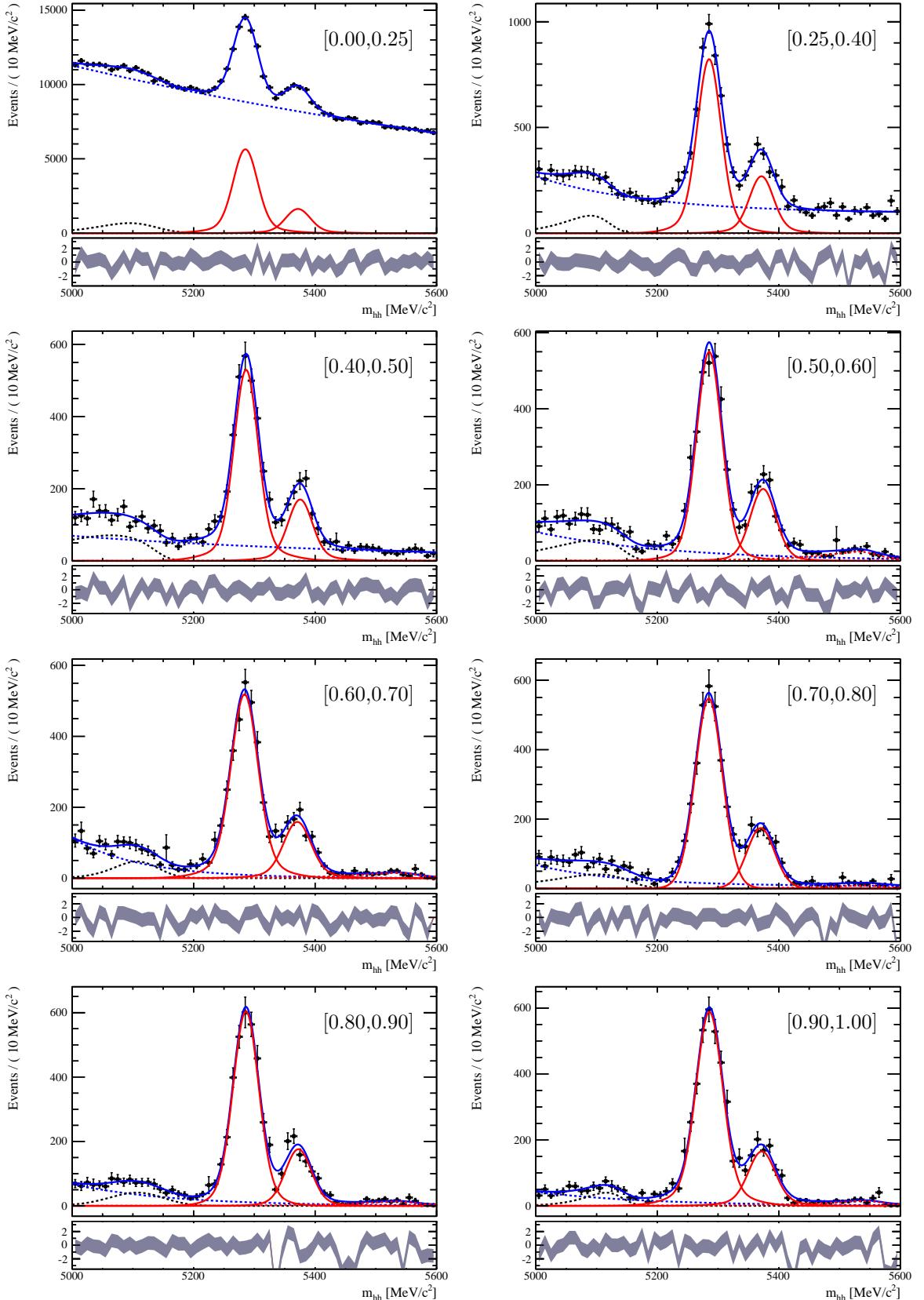


Figure 41: Invariant mass distributions of $B_{(s)}^0 \rightarrow h^+h^-$ from 2011 and 2012 data combined with $\Delta LL_{K-\pi}$ efficiency correction in different BDT bins for $\Delta LL_{K-\pi}$ cut value $\kappa = 5$. Each event in the distributions is weighted according to (12).

The fitted mass model is modified with respect to the baseline one used in Fig. 38: The B^0 and B_s^0 signals, which are here both described as the sum of two Gaussians with common mean, but different width. The combinatorial background shown by the blue dashed line is described by the sum of two exponential distributions.

be able to reproduce it on the BDT signal PDF. The results are listed for the 2011, the 2012 and the combined 2011+2012 data sets in Table 36 and plotted as a function of the BDT12 bin in Fig. ?? for the combined data set only. The observed bias (*i.e.* the efficiency per bin normalized to its average value) has been introduced in the PDF definition of the signal BDT.

Table 36: PID cut efficiency as a function of the BDT12 bins. For 2011, 2012, and 2011+2012 data sets.

BDT bin	$\epsilon_{PID,2011}$ [%]	$\epsilon_{PID,2012}$ [%]	$\epsilon_{PID,AV}$ [%]
0.0-0.25	97.13 ± 0.00	97.71 ± 0.00	97.55 ± 0.00
0.25-0.4	97.16 ± 0.00	97.73 ± 0.00	97.57 ± 0.00
0.4-0.5	97.21 ± 0.00	97.78 ± 0.00	97.62 ± 0.00
0.5-0.6	97.23 ± 0.00	97.78 ± 0.00	97.63 ± 0.00
0.6-0.7	97.45 ± 0.00	97.95 ± 0.00	97.81 ± 0.00
0.7-0.8	97.83 ± 0.00	98.23 ± 0.00	98.12 ± 0.00
0.8-0.9	98.01 ± 0.00	98.37 ± 0.00	98.27 ± 0.00
0.9-1.0	98.57 ± 0.00	98.79 ± 0.00	98.73 ± 0.00

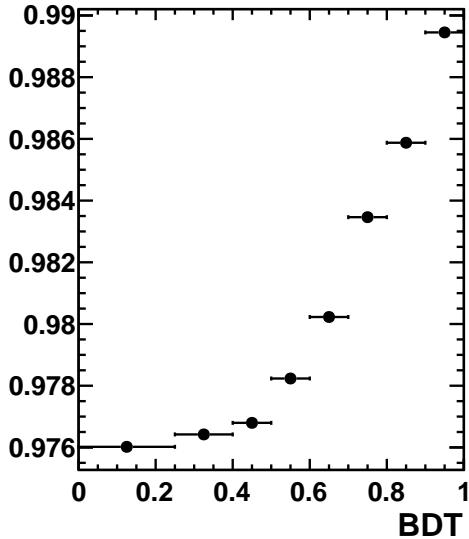


Figure 42: Efficiency of the PID cut (2011 and 2012 data) when folded into the p spectrum of selected $B_s^0 \rightarrow \mu^+\mu^-$ events projected onto the BDT12 axis.

7.2.5 BDT shape correction from trigger efficiency

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:

$$\frac{\epsilon_{\text{TIS}}(L0 \times \text{HLT1})_{hh} \times \epsilon_{\text{trigger}}(\text{HLT2})}{\epsilon_{\text{trigger}}(L0 \times \text{HLT1} \times \text{HLT2})_{\mu\mu}} \quad (15)$$

where the numerator is the efficiency correction to be applied to $B_{(s)}^0 \rightarrow h^+h^-$ TIS events to get a trigger unbiased sample, while the denominator is the efficiency correction to be applied to the unbiased sample to emulate the muon trigger bias on $B_{(s)}^0 \rightarrow \mu^+\mu^-$ signal events.

The evaluation of the above factors has been performed using $B^0 \rightarrow K^+\pi^-$ and $B_{(s)}^0 \rightarrow \mu^+\mu^-$ Monte Carlo events. The trigger bias as a function of the BDT output is shown in Fig. 43, on top plot for the hadronic triggers, *i.e.* the numerator of Eq. 15, and on bottom plot for muon triggers, *i.e.* the denominator of Eq. 15. Both the hadronic and muon trigger biases have been evaluated on offline selected events. The results are shown in Table 37 as a function of the BDT bins.

Table 37: Trigger corrections as a function of the BDT bins.

BDT bin	2011 [%]	2012 [%]	average [%]
0.0-0.25	5.9 ± 0.1	5.9 ± 0.1	5.9 ± 0.1
0.25-0.40	5.9 ± 0.2	5.6 ± 0.1	5.7 ± 0.1
0.40-0.50	5.9 ± 0.2	5.7 ± 0.1	5.8 ± 0.1
0.5-0.6	6.5 ± 0.2	5.6 ± 0.1	5.9 ± 0.1
0.6-0.7	6.1 ± 0.2	6.0 ± 0.2	6.0 ± 0.2
0.7-0.8	6.7 ± 0.2	6.3 ± 0.2	6.4 ± 0.2
0.8-0.9	6.7 ± 0.2	6.5 ± 0.2	6.6 ± 0.2
0.9-1.0	6.9 ± 0.2	6.5 ± 0.2	6.6 ± 0.2

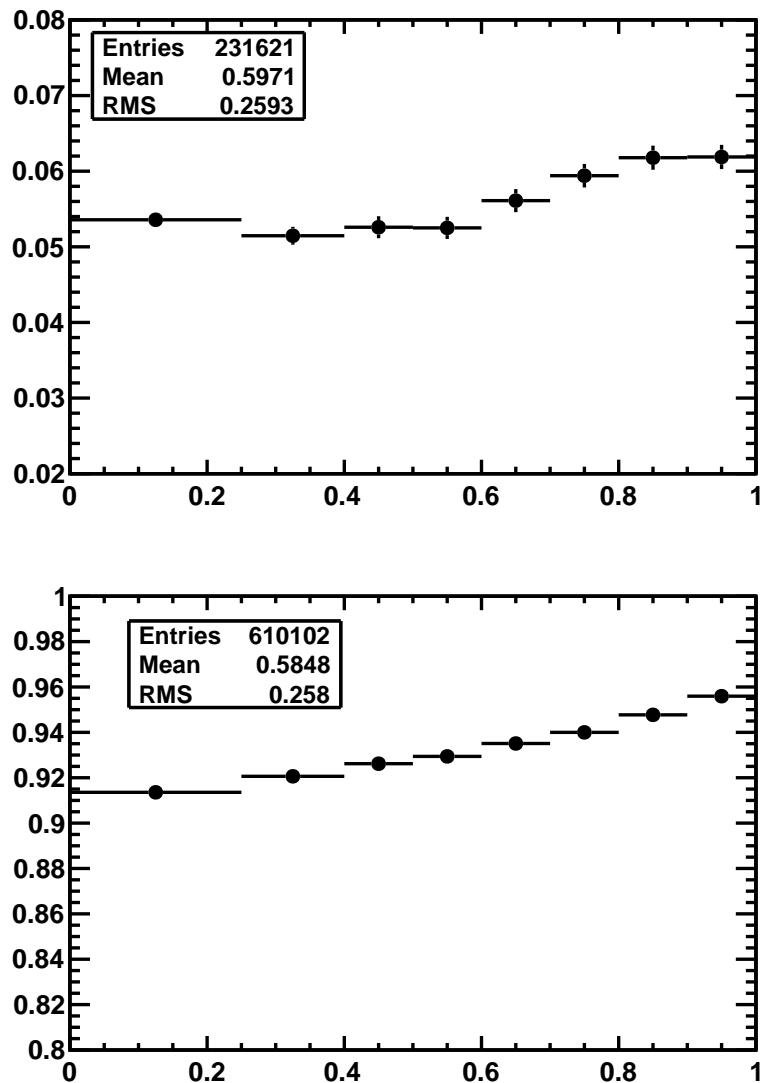


Figure 43: Trigger bias as a function of the BDT output. Top: bias due to the hadronic trigger. Bottom: bias due to the muon trigger.

7.3 Invariant mass

The invariant mass distribution for the signal is described by a Crystal Ball function. The way to determine the different parameters of the function are the same as in the previous analysis [22]. Therefore we only stress differences between the recent analysis and the present one.

The Crystal Ball function is characterized by the following parameters:

- Central value μ
- Mass resolution σ
- Transition point α
- Exponent n

The determination of them is described in the following subsection.

7.3.1 Central value

The central value for the invariant mass distribution is estimated from $B_{(s)}^0 \rightarrow h^+h^-$ in the same way as in the previous $B_s^0 \rightarrow \mu^+\mu^-$ analysis [18]. We require on the candidates a BDTS > 0.1 cut and separate the four mass hypothesis $m_{\pi\pi}$, $m_{K\pi}$, $m_{\pi K}$ and m_{KK} by requiring for

- a kaon: $\Delta LL_{K-\pi} > 10$ and $\Delta LL_{K-p} > 2$
- a pion: $\Delta LL_{K-\pi} < -10$ and $\Delta LL_{\pi-p} > 2$.

Figs 44 and 45 show the invariant mass distribution for the four mass hypotheses ($m_{\pi\pi}$, $m_{K\pi}$, $m_{\pi K}$ and m_{KK}) from the selected events in the $B_{(s)}^0 \rightarrow h^+h^-$ sample. Tab. 38 and 39 summarize the results for the central values. In case of m_{B^0} the final value is calculated as the weighted average:

$$m_{B^0} = \frac{\sum_{i \in \{\pi\pi, \pi K, K\pi\}} m_{B_i^0} / \sigma_{B^0, i, \text{stat}}^2}{\sum_{i \in \{\pi\pi, \pi K, K\pi\}} 1 / \sigma_{B^0, i, \text{stat}}^2}$$

with

$$\begin{aligned} \sigma_{m_{B^0, \text{stat}}}^2 &= \frac{1}{\sum_{i \in \{\pi\pi, \pi K, K\pi\}} 1 / (\sigma_{B^0, i, \text{stat}}^2)} \\ \sigma_{m_{B^0, \text{syst}}} &= \sqrt{\sum_{i \in \{\pi\pi, \pi K, K\pi\}} \sigma_{B^0, i, \text{syst}}^2 / 3} \end{aligned}$$

where we take the average of the systematic uncertainties as they are highly correlated among the three channels.

Tab. 40 lists the different systematic uncertainties estimated for the BDTS > 0.1 cut as well as the $\Delta LL_{K-\pi}$ and $\Delta LL_{p-K,\pi}$.

They are estimated in the following manner:

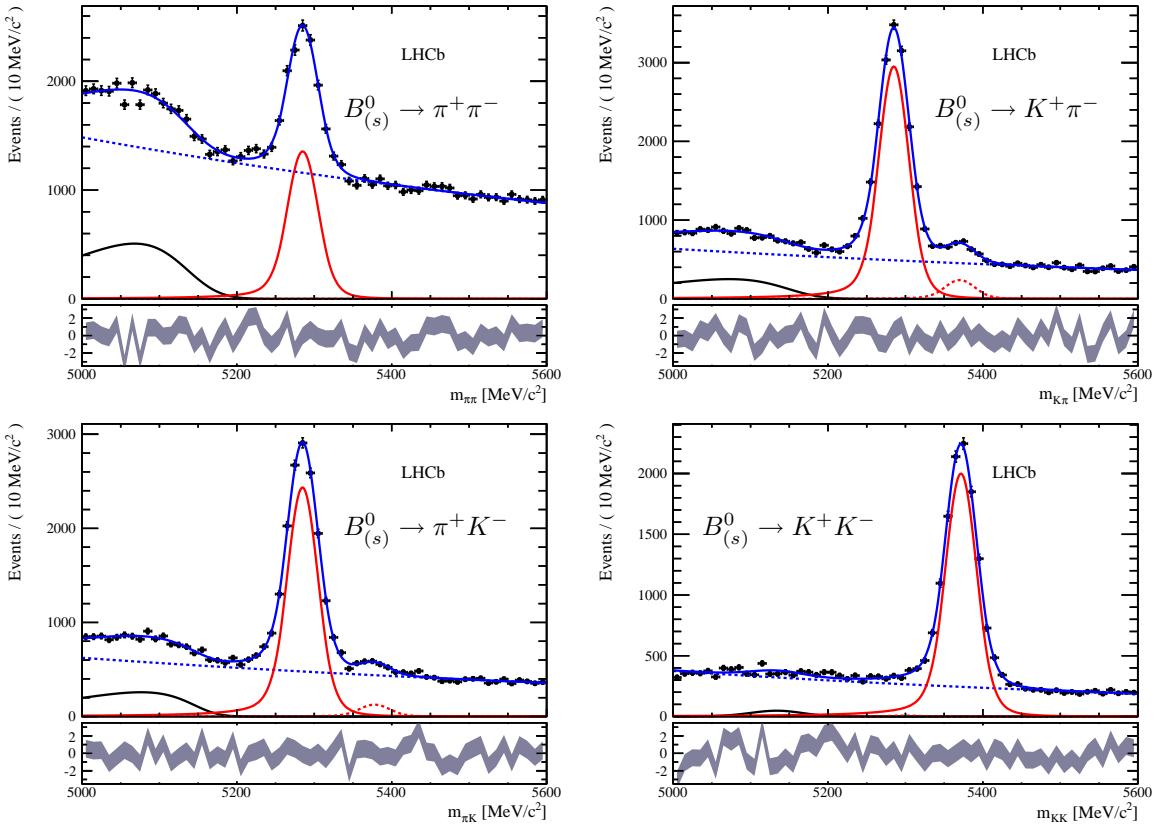


Figure 44: Invariant mass distribution of $B_{(s)}^0 \rightarrow h^+h^-$ for Stripping 20r1 separated into the different decay channels. These distributions are used for the determination of the central value of the invariant mass distribution for $B_{(s)}^0 \rightarrow \mu^+\mu^-$. The full red lines indicates the dominant signal model, the dashed red the sub-dominant (e.g. upper right: full is $B^0 \rightarrow K^+\pi^-$ and dashed is $B_s^0 \rightarrow K^+\pi^-$). The black curve to the left indicated partially reconstructed background.

- BDTS: We repeat the analysis with BDTS cuts of 0.2, 0.15 and 0.05. We perform a linear regression of the four data points (i.e. including the one for 0.1) and take the change of the linear function of the range [0.05,0.2] as systematic uncertainty.
- $\Delta LL_{K-\pi}$: We repeat the analysis with $\Delta LL_{K-\pi}$ cuts for 8, 9, 11 and 12 (or -8, -9, -11, -12 for pions respectively). We perform again a linear regression of the five data points and take the change of the linear function of the range [8,12] as systematic uncertainty.
- $\Delta LL_{K/\pi-p}$: The same is done for the cuts to reject protons with cut values of 0, 1, 3 and 4. The total systematic uncertainty in Tab. 40 assigned to the ΔLL cuts is the square root of the squared sum of the uncertainties from $\Delta LL_{K-\pi}$ and $\Delta LL_{K/\pi-p}$.

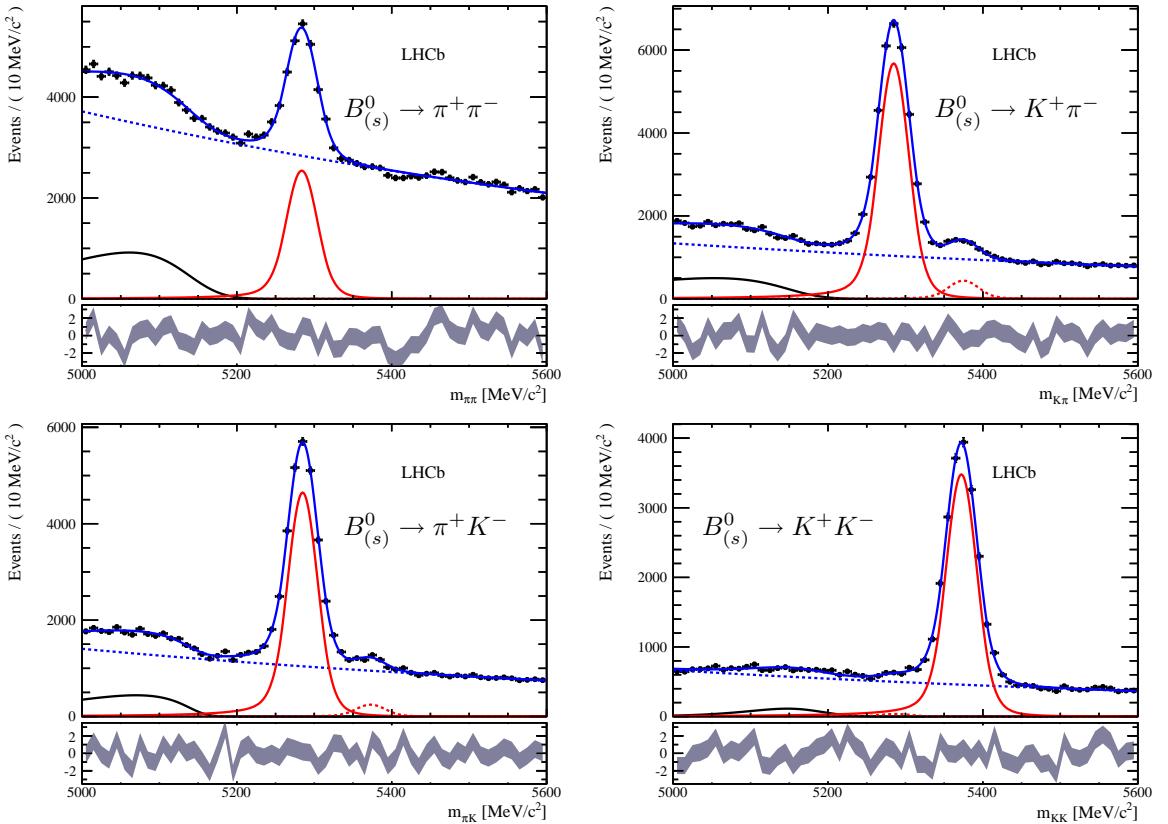


Figure 45: Invariant mass distribution of $B_{(s)}^0 \rightarrow h^+ h^-$ for Stripping 20 separated into the different decay channels. These distributions are used for the determination of the central value of the invariant mass distribution for $B_{(s)}^0 \rightarrow \mu^+ \mu^-$. The full red lines indicates the dominant signal model, the dashed red the sub-dominant (e.g. upper right: full is $B^0 \rightarrow K^+ \pi^-$ and dashed is $B_s^0 \rightarrow K^+ \pi^-$). The black curve to the left indicated partially reconstructed background.

The central values are as in previous reconstruction versions for Stripping 20 and 20r1 about 0.1 % above the nominal value for the B^0 and B_s^0 mass. This difference is in good agreement with the measurement of other resonances from two body decays (cf. Fig. 46).

7.3.2 Invariant mass resolution

The invariant mass resolution is estimated by two methods. One uses the interpolation of the invariant mass resolution of Charmonium and Bottomonium decays, the second one exclusive $B_{(s)}^0 \rightarrow h^+ h^-$ decays.

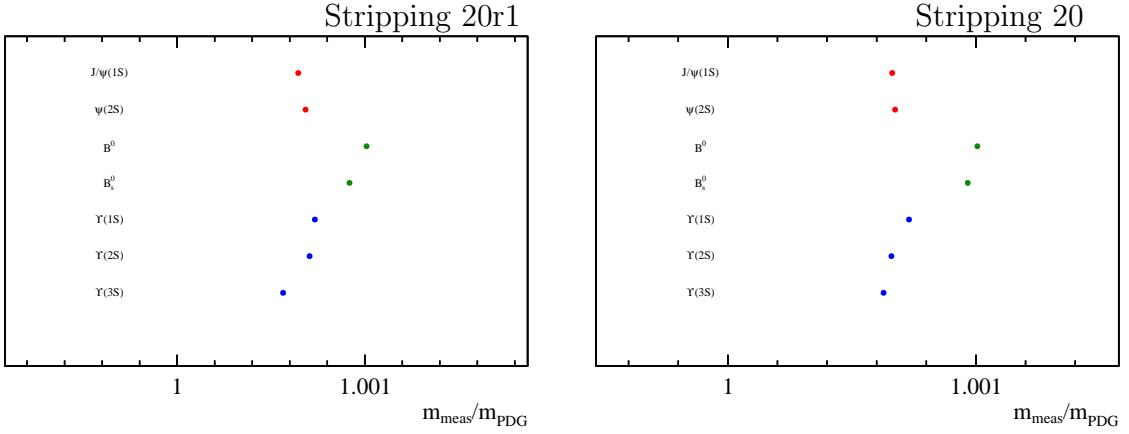


Figure 46: Comparison of measured to nominal central value of several resonances

Table 38: Summary of central value for the invariant dimuon mass distribution for the Stripping 20r1.

Channel	value with stat. & syst. uncertainties
$B^0 \rightarrow \pi^+ \pi^-$	$(5284.99 \pm 0.52_{\text{stat}} \pm 0.24_{\text{syst}}) \text{ MeV}/c^2$
$B^0 \rightarrow K^+ \pi^-$	$(5284.98 \pm 0.24_{\text{stat}} \pm 0.16_{\text{syst}}) \text{ MeV}/c^2$
$B^0 \rightarrow \pi^+ K^-$	$(5284.79 \pm 0.27_{\text{stat}} \pm 0.21_{\text{syst}}) \text{ MeV}/c^2$
$B_s^0 \rightarrow K^+ K^-$	$(5371.70 \pm 0.26_{\text{stat}} \pm 0.16_{\text{syst}}) \text{ MeV}/c^2$
m_{B^0}	$(5284.91 \pm 0.17_{\text{stat}} \pm 0.19_{\text{syst}}) \text{ MeV}/c^2$
$m_{B_s^0}$	$(5371.70 \pm 0.26_{\text{stat}} \pm 0.16_{\text{syst}}) \text{ MeV}/c^2$

Interpolation

As in the previous analysis we use the interpolation of the invariant mass resolution between the Charmonium and Bottomonium resonances. There has been no modification in the interpolation method with respect to the last analysis round.

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

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

Fig. 47 shows the Charmonium and Bottomonium resonances while Fig. 48 shows the interpolation of the invariant mass resolution.

The values for the interpolated mass resolutions at m_{B^0} and $m_{B_s^0}$ are:

Table 39: Summary of central value for the invariant dimuon mass distribution for the Stripping 20.

Channel	value with stat. & syst. uncertainties
$B^0 \rightarrow \pi^+ \pi^-$	$(5284.78 \pm 0.57_{\text{stat}} \pm 0.24_{\text{syst}}) \text{ MeV}/c^2$
$B^0 \rightarrow K^+ \pi^-$	$(5284.85 \pm 0.17_{\text{stat}} \pm 0.21_{\text{syst}}) \text{ MeV}/c^2$
$B^0 \rightarrow \pi^+ K^-$	$(5284.96 \pm 0.19_{\text{stat}} \pm 0.22_{\text{syst}}) \text{ MeV}/c^2$
$B_s^0 \rightarrow K^+ K^-$	$(5371.96 \pm 0.22_{\text{stat}} \pm 0.22_{\text{syst}}) \text{ MeV}/c^2$
m_{B^0}	$(5284.89 \pm 0.12_{\text{stat}} \pm 0.22_{\text{syst}}) \text{ MeV}/c^2$
$m_{B_s^0}$	$(5371.96 \pm 0.22_{\text{stat}} \pm 0.22_{\text{syst}}) \text{ MeV}/c^2$

Table 40: Summary of systematic uncertainty on the central value determination

Channel	Stripping 20r1		Stripping 20	
	ΔLL cuts	BDTS cut	ΔLL cuts	BDTS cut
$B^0 \rightarrow \pi^+ \pi^-$	$\pm 0.22 \text{ MeV}/c^2$	$\pm 0.10 \text{ MeV}/c^2$	$\pm 0.24 \text{ MeV}/c^2$	$\pm 0.12 \text{ MeV}/c^2$
$B^0 \rightarrow K^+ \pi^-$	$\pm 0.14 \text{ MeV}/c^2$	$\pm 0.07 \text{ MeV}/c^2$	$\pm 0.19 \text{ MeV}/c^2$	$\pm 0.11 \text{ MeV}/c^2$
$B^0 \rightarrow \pi^+ K^-$	$\pm 0.19 \text{ MeV}/c^2$	$\pm 0.08 \text{ MeV}/c^2$	$\pm 0.15 \text{ MeV}/c^2$	$\pm 0.10 \text{ MeV}/c^2$
$B_s^0 \rightarrow K^+ K^-$	$\pm 0.12 \text{ MeV}/c^2$	$\pm 0.11 \text{ MeV}/c^2$	$\pm 0.18 \text{ MeV}/c^2$	$\pm 0.10 \text{ MeV}/c^2$

Stripping 20r1

$$\begin{aligned}\sigma_{B^0} &= (23.11 \pm 0.22_{\text{stat}} \pm 0.49_{\text{syst}}) \text{ MeV}/c^2 \\ \sigma_{B_s^0} &= (23.52 \pm 0.23_{\text{stat}} \pm 0.54_{\text{syst}}) \text{ MeV}/c^2\end{aligned}$$

Stripping 20

$$\begin{aligned}\sigma_{B^0} &= (23.21 \pm 0.16_{\text{stat}} \pm 0.51_{\text{syst}}) \text{ MeV}/c^2 \\ \sigma_{B_s^0} &= (23.63 \pm 0.17_{\text{stat}} \pm 0.54_{\text{syst}}) \text{ MeV}/c^2\end{aligned}$$

The systematic uncertainties – summarized in Tab. 41 – are analogously determined as in previous analyses. The only difference is the missing of the systematic uncertainty due to the momentum weighting as this is not used anymore. Also the fit functions for the invariant dimuon mass distributions cause (mainly due to the η_μ cut removing as described some acceptance effects) smaller systematic uncertainties. This leads to the overall reduction of the systematic uncertainty by one third.

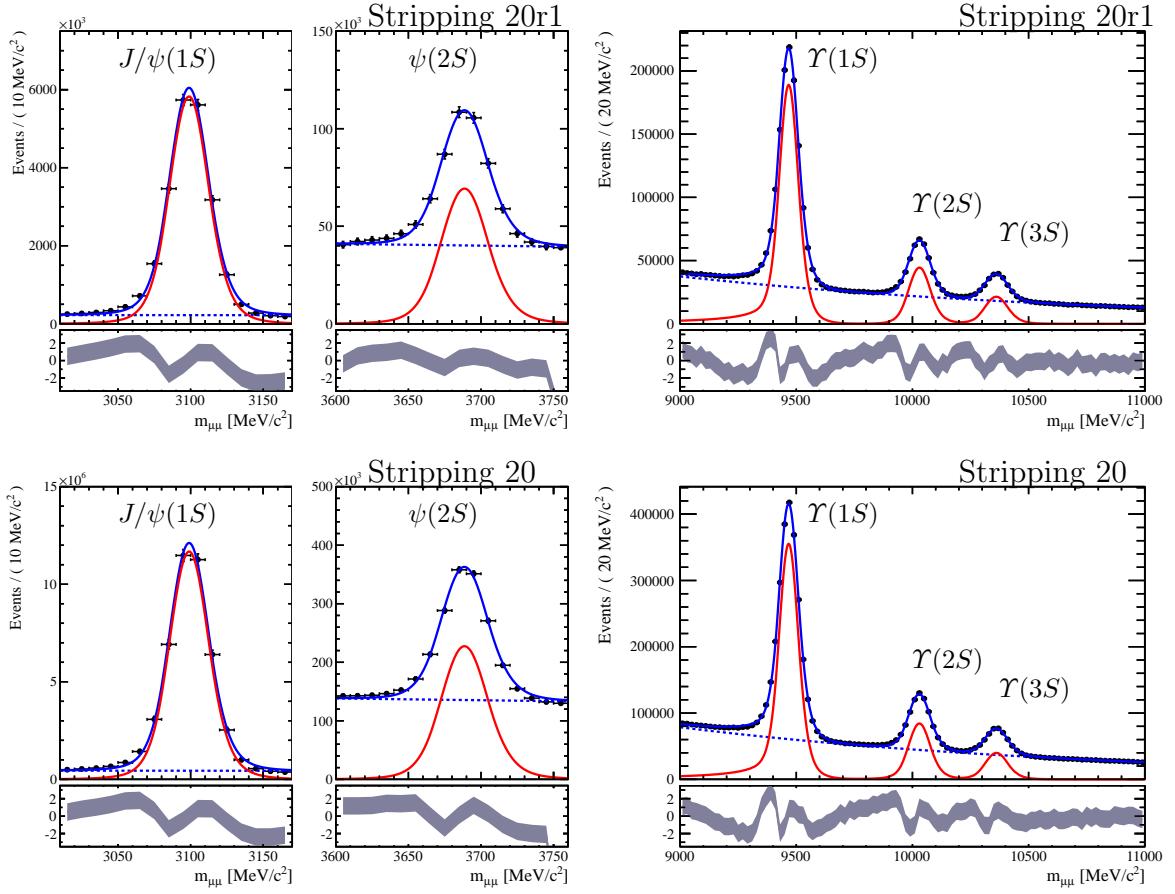


Figure 47: Invariant dimuon mass distribution for $J/\psi(1S) \rightarrow \mu^+\mu^-$, $\psi(2S) \rightarrow \mu^+\mu^-$ and $\Upsilon(1, 2, 3S) \rightarrow \mu^+\mu^-$ (top: 2011 data, bottom: 2012 data). All resonances are described by the sum of two Crystal Ball functions (one with its tail to the left, the other with its tail to the right) while the combinatorial background is described by an exponential function.

$$B_{(s)}^0 \rightarrow h^+h^-$$

The second method for the estimation of the invariant mass resolution is done with a $B_{(s)}^0 \rightarrow h^+h^-$ sample which is defined in the same manner as the one for the BDT calibration (cf. Sec. 7.2.3) without the TIS requirements.

We also determine in the same manner the invariant mass distribution of the events where we weight them also by

$$w = \frac{1}{\varepsilon_{\Delta LL}(p, \eta, nTr|h^+, \kappa) \cdot \varepsilon_{\Delta LL}(p, \eta, nTr|h'^-, \kappa)}$$

using the same efficiency tables.

The invariant mass distribution (cf. Fig. 49) is fitted by the same fit model as in the BDT calibration except that we do not use a combined width for B^0 and B_s^0 , but fix the width

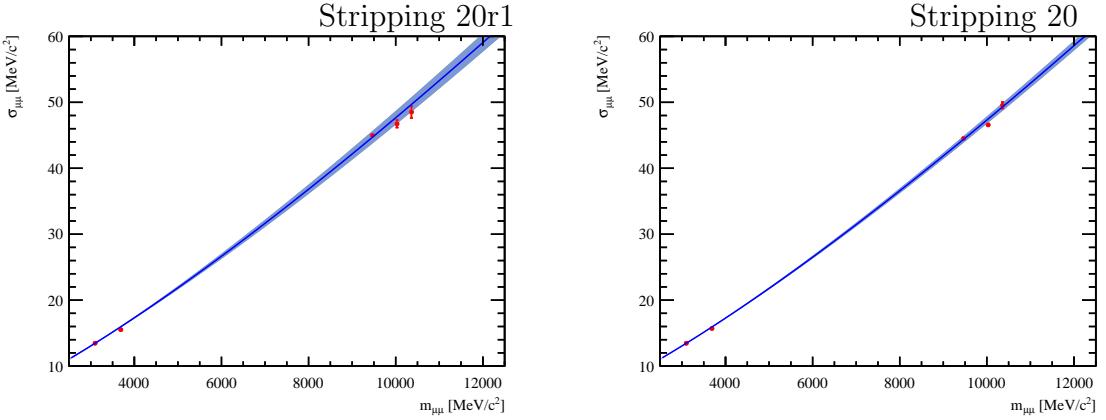


Figure 48: Interpolation of the invariant mass resolution between Charmonium and Bottomonium resonances to the mass of the B^0 and B_s^0 mesons (left: 2011 data, right: 2012 data)

Table 41: List of systematic uncertainties for the interpolation method.

Systematic	Stripping 20r1		Stripping 20	
	B^0	B_s^0	B^0	B_s^0
Selection cuts	$\pm 0.20 \text{ MeV}/c^2$	$\pm 0.22 \text{ MeV}/c^2$	$\pm 0.19 \text{ MeV}/c^2$	$\pm 0.23 \text{ MeV}/c^2$
Error asymmetry	$\pm 0.17 \text{ MeV}/c^2$	$\pm 0.28 \text{ MeV}/c^2$	$\pm 0.21 \text{ MeV}/c^2$	$\pm 0.23 \text{ MeV}/c^2$
Mass window	$\pm 0.28 \text{ MeV}/c^2$			
Fit function of invariant mass	$\pm 0.30 \text{ MeV}/c^2$	$\pm 0.29 \text{ MeV}/c^2$	$\pm 0.32 \text{ MeV}/c^2$	$\pm 0.33 \text{ MeV}/c^2$
Total systematic error	$\pm 0.49 \text{ MeV}/c^2$	$\pm 0.54 \text{ MeV}/c^2$	$\pm 0.51 \text{ MeV}/c^2$	$\pm 0.54 \text{ MeV}/c^2$

of $B_s^0 \sigma_s$ to the one of $B^0 \sigma$ times the ratio of the invariant mass resolutions determined by the interpolation method:

$$\begin{aligned} \sigma_s(\kappa) &= \beta_{\text{reso}} \cdot \sigma(\kappa) \\ \text{with } \beta_{\text{reso}} &= \frac{\sigma_{B_s^0}}{\sigma_{B^0}} = 1.018 \text{ for 2011} \\ \beta_{\text{reso}} &= \frac{\sigma_{B_s^0}}{\sigma_{B^0}} = 1.018 \text{ for 2012} \end{aligned}$$

We neglect the uncertainty on β_{reso} as the uncertainties on $\sigma_{B_s^0}$ and σ_{B^0} are very strongly correlated.

We determine the width of the B^0 resonance for $\Delta LL_{K-\pi}$ cut values κ as they are used in the BDT calibration in the range $\kappa \in [0, 20]$ in 0.5 steps.

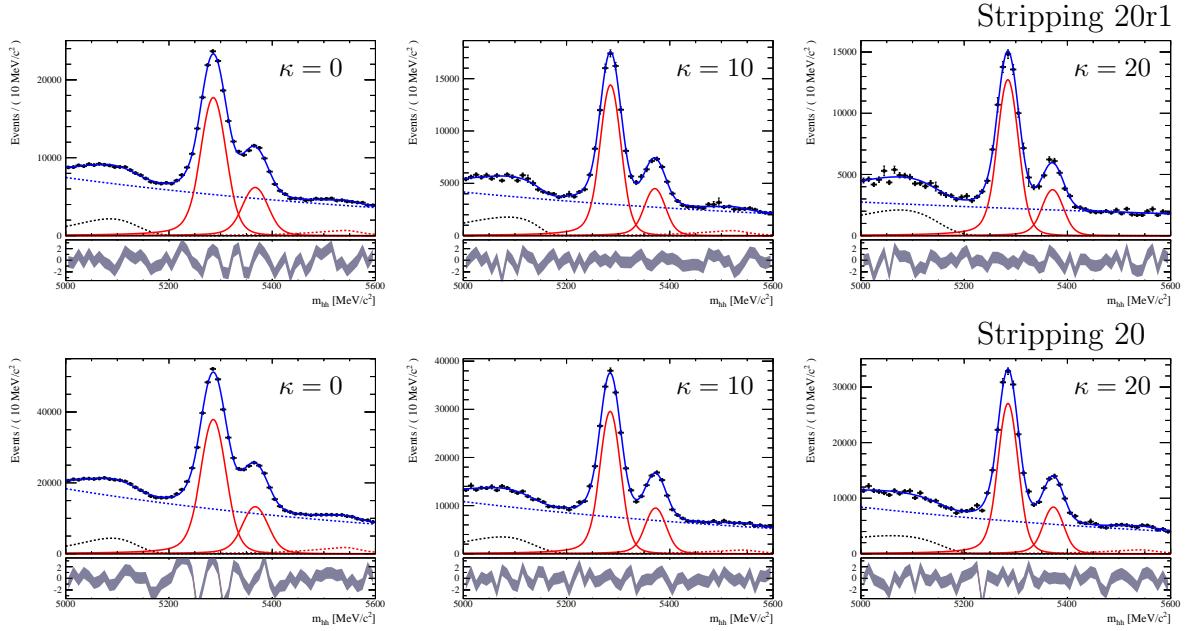


Figure 49: Invariant mass distribution of $B_{(s)}^0 \rightarrow h^+h^-$ used to determine the invariant mass resolution for different $\Delta LL_{K-\pi}$ cut values κ (top: 2011 data, bottom: 2012 data). Each event in the distributions is weighted according to (12).

As resolution for the invariant mass for B^0 we extract the width of the corresponding signal. This width is diluted by a mis-ID component $\pi \leftrightarrow K$ of the final state particles. Therefore we fit the width of the peaks as a function of the $\Delta LL_{K-\pi}$ cut (cf. Fig. 50) by the function

$$\sigma(\kappa) = \sigma_0 + \frac{a_0}{1 + r \cdot \exp(\kappa/\kappa_0)}. \quad (16)$$

The first component is the width of the resonances corrected for the mis-ID component while the second term describes the magnitude of the mis-ID component as a function of the $\Delta LL_{K-\pi}$ cut taking into account the definition of $\Delta LL_{K-\pi}$ as a log-likelihood. Therefore the probability to misidentify a particle should be $p(\kappa) = (1 + \exp(\kappa))^{-1}$. The fit parameters r and κ_0 take the imperfection of the definition of $\Delta LL_{K-\pi}$ into account. The fit range is restricted (to $\kappa \in [0, 13]$ for 2011 data and to $\kappa \in [0, 17]$ for 2012 data) as for tighter $\Delta LL_{K-\pi}$ the number of events with at least one hadron with high p is very small. This makes it impossible to correct the invariant mass distribution by the weighting. This is in Fig. 50 on the left hand side visible by the drop in $\sigma(\kappa)$ at ~ 15 . From the fit we get:

2011

$$\sigma_0 = (21.66 \pm 0.18) \text{ MeV}/c^2$$

2012

$$\sigma_0 = (21.57 \pm 0.10) \text{ MeV}/c^2$$

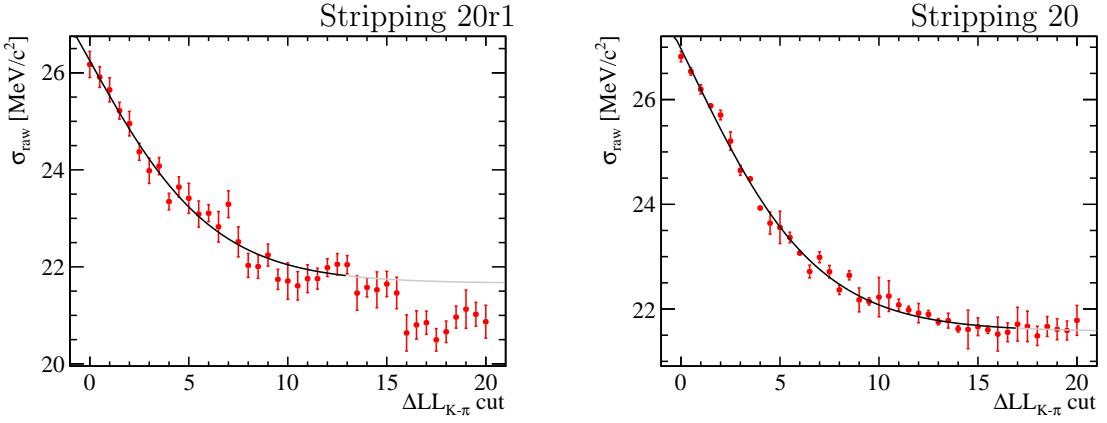


Figure 50: Width of the B^0 peak as a function of the $\Delta LL_{K-\pi}$ (left: 2011 data, right: 2012 data). The line indicates the fitted function (16).

which has to be further corrected to incorporate the restriction of the phase space by the efficiency tables (i.e. we reject all events including a hadron with $p > 200 \text{ GeV}/c$). This has been done in two independent ways by measuring the width of the dimuon invariant mass distribution in $B_s^0 \rightarrow \mu^+ \mu^-$ Monte Carlo and take the ratio of the width with the phase space restriction to the width without phase space restriction as the correction γ .

As second method we apply the interpolation method (described above in this subsection) with and without phase space restriction and calculate the correction γ as ratio between to the resulting values for the invariant mass resolution for B_s^0 :

2011

$$\begin{aligned}
 B_s^0 \rightarrow \mu^+ \mu^- \text{ Monte Carlo: } \gamma &= (1.030 \pm 0.006) \\
 \text{Interpolation } B_s^0: \gamma &= (1.031 \pm 0.013) \\
 \text{Interpolation } B^0: \gamma &= (1.032 \pm 0.014) \\
 \text{Combination: } \gamma &= (1.030 \pm 0.005)
 \end{aligned}$$

2012

$$\begin{aligned}
 B_s^0 \rightarrow \mu^+ \mu^- \text{ Monte Carlo: } \gamma &= (1.030 \pm 0.006) \\
 \text{Interpolation } B_s^0: \gamma &= (1.032 \pm 0.012) \\
 \text{Interpolation } B^0: \gamma &= (1.029 \pm 0.010) \\
 \text{Combination: } \gamma &= (1.030 \pm 0.005)
 \end{aligned}$$

After the correction of σ_0 by $\sigma_{B^0} = \gamma \cdot \sigma_0$ we get the final result of

2011

$$\sigma_{B^0} = (22.30 \pm 0.19_{\text{stat}} \pm 0.74_{\text{syst}}) \text{ MeV}/c^2$$

2012

$$\sigma_{B^0} = (22.22 \pm 0.10_{\text{stat}} \pm 0.70_{\text{syst}}) \text{ MeV}/c^2$$

while for $\sigma_{B_s^0} = \gamma \cdot \beta_{\text{reso}} \cdot \sigma_0$ we get

2011

$$\sigma_{B_s^0} = (22.71 \pm 0.19_{\text{stat}} \pm 0.76_{\text{syst}}) \text{ MeV}/c^2$$

2012

$$\sigma_{B_s^0} = (22.62 \pm 0.11_{\text{stat}} \pm 0.71_{\text{syst}}) \text{ MeV}/c^2.$$

The systematic uncertainties are determined in the following manner:

1. The propagated error of the phase space correction of ± 0.009 :

2011: $\pm 0.11 \text{ MeV}/c^2$

2012: $\pm 0.11 \text{ MeV}/c^2$

2. **The fit range of $\sigma(\kappa)$:** To estimate the systematic uncertainty by the choice on the fit range of $\sigma(\kappa)$ we decided to change the upper limit of the fit range by ± 0.5 and repeat the extraction of the resolution with those modified fit ranges (2011: [0, 16.5] or [0, 17.5], respectively; 2012: [0, 12.5] or [0, 13.5], respectively). The assigned systematic uncertainty is the maximal difference of the resolution with the modified fit ranges to the resolution determined with the nominal fit ranges. **2011:** $\pm 0.25 \text{ MeV}/c^2$
2012: $\pm 0.02 \text{ MeV}/c^2$

3. **Fit function $\sigma(\kappa)$:** We check the functional form of $\sigma(\kappa)$ by adding a third term $b_0/(1 + r \exp(\kappa/\kappa_0))^2$ which takes double mis-ID $\pi \leftrightarrow K$ into account and take the difference between the result with and without the extra term as systematic uncertainty:

2011: $\pm 0.30 \text{ MeV}/c^2$

2012: $\pm 0.23 \text{ MeV}/c^2$

4. **Fit function of m_{hh} :** We modify the fit function describing the invariant mass of the two hadrons by letting the tail parameters floating (in the fit for the result they are fixed to values determined from Monte Carlo) in the range $\alpha_{l,r} \in [\pm 1, \pm 5]$ (left transition point + sign, right transition point - sign) and $n_{l,r} \in [0.75, 2]$. The systematics is again the difference between the resolution determined with this modification and the original value:

2011: $\pm 0.02 \text{ MeV}/c^2$

2012: $\pm 0.03 \text{ MeV}/c^2$

5. **Binning scheme:** The systematic uncertainty caused by the binning scheme used to determine the $\Delta LL_{K-\pi}$ is as in the BDT calibration estimated by using the efficiency table based on the $46 \times 7 \times 7$ binning scheme, recalculate the resolution and taking the difference to the original resolution:

2011: $\pm 0.42 \text{ MeV}/c^2$

2012: $\pm 0.39 \text{ MeV}/c^2$

6. **Efficiency:** The systematic uncertainty on the $\Delta LL_{K-\pi}$ efficiency is taken into account by rerunning the determination as in case of the BDT calibration by adding $\Delta\varepsilon(\kappa)$ when calculating the weighting:

$$w = \frac{1}{(\varepsilon_{\Delta LL}(p, \eta, \text{nTr}|h^+, \kappa) + \Delta\varepsilon(\kappa)) \cdot (\varepsilon_{\Delta LL}(p, \eta, \text{nTr}|h'^-, \kappa) + \Delta\varepsilon(\kappa))}$$

and taking the difference between the resolution determined with this and the one determined the usual weighting as systematics:

2011: $\pm 0.50 \text{ MeV}/c^2$

2012: $\pm 0.52 \text{ MeV}/c^2$

The total systematic uncertainty is taken as the square root of the sum of the squared single systematics being

2011: $\pm 0.74 \text{ MeV}/c^2$

2012: $\pm 0.70 \text{ MeV}/c^2$.

This value is for B_s^0 also scaled with β_{reso} .

Combination

Combining the two results for the invariant mass resolution by taking the weighted average we get:

Stripping 20r1:

$$\sigma_{B^0} = (22.84 \pm 0.14_{\text{stat}} \pm 0.41_{\text{syst}}) \text{ MeV}/c^2 \quad (17)$$

$$\sigma_{B_s^0} = (23.22 \pm 0.18_{\text{stat}} \pm 0.44_{\text{syst}}) \text{ MeV}/c^2 \quad (18)$$

Stripping 20:

$$\sigma_{B^0} = (22.83 \pm 0.08_{\text{stat}} \pm 0.43_{\text{syst}}) \text{ MeV}/c^2 \quad (19)$$

$$\sigma_{B_s^0} = (23.24 \pm 0.09_{\text{stat}} \pm 0.43_{\text{syst}}) \text{ MeV}/c^2 \quad (20)$$

which are the values used in the extraction of the result.

7.3.3 Crystal Ball parameters α and n

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

has the width determined in Eqs (17), (18), (19) and (20). This distribution is then fitted with a Crystal Ball function from which we extract α and n .

By repeating this procedure several times we determine the distributions of α and n and take their mean as value for the two parameters. The statistical uncertainty is the uncertainty on the mean. The uncertainty on the invariant mass resolution used to smear the true invariant mass distribution is taken as a systematic uncertainty.

Stripping 20r1:

$$\begin{aligned}\alpha_{B^0} &= (2.066 \pm 0.013_{\text{stat}} \pm 0.010_{\text{syst}}) \\ n_{B^0} &= (1.112 \pm 0.030_{\text{stat}} \pm 0.045_{\text{syst}}) \\ \alpha_{B_s^0} &= (2.065 \pm 0.010_{\text{stat}} \pm 0.010_{\text{syst}}) \\ n_{B_s^0} &= (1.119 \pm 0.025_{\text{stat}} \pm 0.043_{\text{syst}})\end{aligned}$$

Stripping 20:

$$\begin{aligned}\alpha_{B^0} &= (2.065 \pm 0.012_{\text{stat}} \pm 0.009_{\text{syst}}) \\ n_{B^0} &= (1.114 \pm 0.029_{\text{stat}} \pm 0.028_{\text{syst}}) \\ \alpha_{B_s^0} &= (2.065 \pm 0.009_{\text{stat}} \pm 0.012_{\text{syst}}) \\ n_{B_s^0} &= (1.122 \pm 0.023_{\text{stat}} \pm 0.037_{\text{syst}})\end{aligned}$$

7.4 Combination between Stripping 20r1 and 20

As the values for all four parameters μ , σ , n and α are in good agreement between the two data sets and as we have also calibrated the BDT PDF for signal in a combined approach between Stripping 20r1 and 20, we decided to use averaged values for the signal mass PDF. All the parameters are combined in the following manner: We take the weighted average of the two values where we take $1/\sigma_{\text{stat}}^2$ as weighting factor. As combined statistical uncertainty we take

$$\sigma_{\text{stat;comb}} = \sqrt{\frac{1}{1/\sigma_{\text{stat;s20r1}}^2 + 1/\sigma_{\text{stat;s20}}^2}}$$

while the combined systematic uncertainty is taken as the normal average of the systematic uncertainties of Stripping 20r1 and 20.

Central value μ :

$$\begin{aligned}\mu_{B^0} &= (5284.90 \pm 0.10_{\text{stat}} \pm 0.20_{\text{syst}}) \text{ MeV}/c^2 \\ \mu_{B_s^0} &= (5371.85 \pm 0.17_{\text{stat}} \pm 0.19_{\text{syst}}) \text{ MeV}/c^2\end{aligned}$$

Resolution σ :

$$\begin{aligned}\sigma_{B^0} &= (22.83 \pm 0.07_{\text{stat}} \pm 0.42_{\text{syst}}) \text{ MeV}/c^2 \\ \sigma_{B_s^0} &= (23.24 \pm 0.08_{\text{stat}} \pm 0.44_{\text{syst}}) \text{ MeV}/c^2\end{aligned}$$

As the tail parameters n and α are also compatible between the signal PDF for B^0 and B_s^0 , we have also decided to have a single value n and α each. These are estimated in the same manner as described above using the four values from Stripping 20r1 and 20.

Tail parameters α and n

$$\begin{aligned}\alpha &= (2.065 \pm 0.005_{\text{stat}} \pm 0.010_{\text{syst}}) \\ n &= (1.118 \pm 0.013_{\text{stat}} \pm 0.038_{\text{syst}})\end{aligned}$$

8 Normalization

To estimate the signal branching ratio, we normalize the number of observed signal events to the number of events of a calibration channel of a well known branching ratio. We have used two complementary normalization channels: $B^+ \rightarrow J/\psi(\mu^+\mu^-)K^+$ and $B^0 \rightarrow K^+\pi^-$. The first decay has similar trigger and muon identification efficiencies as the signal but different number of particles in the final state, while the second channel has a similar topology but a different trigger selection.

The selection of the normalization channels was described in Sec. 4, it has been designed to be very similar to the selection of the signal events, such that the systematic uncertainties cancel in the ratio of efficiencies and the knowledge of the absolute integrated luminosity and the $b\bar{b}$ cross-section are not needed.

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

$$\text{BR} = \text{BR}_{\text{cal}} \times \frac{\epsilon_{\text{cal}}^{\text{GEN}} \epsilon_{\text{cal}}^{\text{SEL&REC|GEN}} \epsilon_{\text{cal}}^{\text{TRIG|SEL}}}{\epsilon_{\text{sig}}^{\text{GEN}} \epsilon_{\text{sig}}^{\text{SEL&REC|GEN}} \epsilon_{\text{sig}}^{\text{TRIG|SEL}}} \times \frac{f_{\text{cal}}}{f_{B_q^0}} \times \frac{N_{B_q^0 \rightarrow \mu^+\mu^-}}{N_{\text{cal}}} = \alpha_{\text{cal}} \times N_{B_q^0 \rightarrow \mu^+\mu^-}, \quad (21)$$

where $f_{B_q^0}$ and f_{cal} are the probabilities that a b -quark fragments into a B_q^0 and into the b -hadron relevant for the chosen calibration mode. BR_{cal} is the branching ratio and N_{cal} is the number of selected events of the calibration channel. The efficiency is separated in three factors: ϵ^{GEN} is the efficiency to that all final state tracks pass the generator level cuts¹² $\epsilon^{\text{SEL&REC|GEN}}$ is the efficiency to reconstruct and select the events; $\epsilon^{\text{TRIG|SEL}}$ is the efficiency of the trigger on reconstructed and selected events. The sub-indexes indicate if the efficiency refers to the signal (sig) or the calibration channel (cal). Finally, α_{cal} is the normalization factor (or single event sensitivity) and $N_{B_q^0 \rightarrow \mu^+\mu^-}$ the number of observed signal events. In this Section we discuss the estimation of the factors that enter in the computation of α_{cal} .

The branching ratios of the considered normalisation channels are respectively $\mathcal{B}(B^+ \rightarrow J/\psi K^+) = (6.025 \pm 0.205) \cdot 10^{-5}$ (where the $J/\psi \rightarrow \mu^+\mu^-$ branching fraction is also included) and $\mathcal{B}(B^0 \rightarrow K^+\pi^-) = (1.94 \pm 0.06) \cdot 10^{-5}$ [38].

8.1 Acceptance efficiencies from MC simulation

The reconstruction efficiency factorizes in two parts: the detector acceptance (or generation efficiency) ϵ_{gen} and the reconstruction efficiency $\epsilon_{\text{reco/gen}}$. The detector acceptance is defined as the fraction of the tracks in the final state that are inside the LHCb acceptance¹³. The reconstruction efficiency is defined as the efficiency to reconstruct all the tracks in the final state.

¹²Angle between 10 and 400 mrad.

¹³The 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.

The generation acceptances are summarised in Table 42. They have been computed using MC simulation, and its uncertainty is smaller than 0.1%.

Table 42: Generator level efficiency ϵ_{gen} for signal and control channels.

	$B_s^0 \rightarrow \mu^+ \mu^-$	$B^+ \rightarrow J/\psi K^+$	$B^0 \rightarrow K^+ \pi^-$
ϵ_{gen}	$(17.75 \pm 0.09)\%$	$(15.78 \pm 0.08)\%$	$(17.73 \pm 0.09)\%$

When normalising to $B^0 \rightarrow K^+ \pi^-$, we require that the two hadrons are within the muon detector acceptance to minimize the differences between the signal and this control channel. The uncertainties associated to the muon system acceptance will be treated together with the muonID efficiency and described later in this section.

8.2 Ratio of reconstruction and selection efficiencies

The acceptance and reconstruction efficiencies depend on the phase space of the final decay products, the acceptance of the detector and the efficiency of track finding algorithms. We compute the ratio of acceptance and reconstruction efficiencies using the MC simulation. The differences between the simulation and data are treated as systematic uncertainties.

The selection criteria for signal and normalization channels were described in Sect. 4. The selection efficiencies are determined using MC simulation and then cross-checked with data.

In this section we discuss all the factors and corrections that enter in this ratio. The acceptance of the detector is measured as the generation efficiency of the MC (§8.1).

For this round of the analysis we decided to group the selection and reconstruction efficiencies and to estimate them together in order to avoid the requirement of MC truth matching necessary to estimate the reconstruction efficiency alone. The MC matching procedure has some inefficiency that has to be corrected for, while we prefer to directly avoid this step (§8.2.1).

8.2.1 Reconstruction and selection efficiencies from simulation

The reconstruction and selection efficiencies are determined using MC simulation and then cross-checked with data in various points. As already said we estimate the final efficiency altogether, instead of the reconstruction and selection efficiencies separately in order to avoid the MC truth matching procedure and its inefficiency. The number of selected and reconstructed events is estimated from a fit of the final MC invariant mass distribution: this is necessary in order to subtract the very little background present. For the $B_s^0 \rightarrow \mu^+ \mu^-$ channel no background is present in the signal MC so that we just take the number of events.

Impact parameter smearing. The main difference between the MC and the data is in the IP distribution. In order to cross-check such a difference, we have applied a technique developed by the tracking group to smear on the MC sample the parameters of the tracks at the origin, and thus reproduce the observed IP resolution in data. We have considered MC samples treated with different values of the impact parameter smearing. The default value was optimised to match the data with MC11 and was not reoptimised this year. We control this smearing via a “smearing scale” which is just a multiplicative factor to this default value. Comparing data and MC, we have found the best smearing value at a scale of [0.4-0.5]: in Fig. 51 we show the Kolmogorov-Smirnov test distance between the sideband subtracted data and MC distributions of the impact parameter χ^2 of $B^+ \rightarrow J/\psi K^+$ candidates. As it can be seen the lowest value is obtained for a smearing scale of about 0.4 (For comparison the smallest value for old processing of data is shown in the third plot to be reproduced by an additional smearing scale of zero, sign that the new MC has been improved to reproduce the impact parameter resolution in data, at least at the level it was known with the old processing).

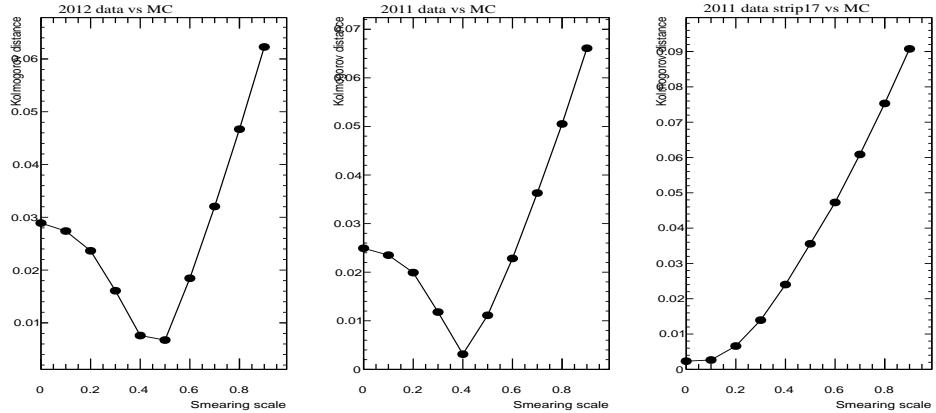


Figure 51: Kolmogorov-Smirnov distance between the sideband subtracted data and MC distributions of the impact parameter χ^2 of $B^+ \rightarrow J/\psi K^+$ candidates for 2012 data and 2011 data reprocessed and for 2011 data with old processing.

The overall efficiency in MC can vary of about 7.5% between unsmeared and smeared samples; however the ratio of efficiencies stays constant within 0.1% over a smearing scale from 0 to 1.1. The absolute efficiencies and their ratio are shown in Fig. 52.

In the previous analysis rounds it was considered as central value for the absolute efficiencies the smeared one, and over-smeared and unsmeared were taken as systematics. In this round we have studied the modification of the distribution of other variables (apart from the IP) when applying the smearing procedure. We have found that some variables ($FD\chi^2$, DOCA, BDTS) were also modified by the smearing and tended to be more different from data the larger was the smearing scale. We show this in Fig. 53, by comparing the

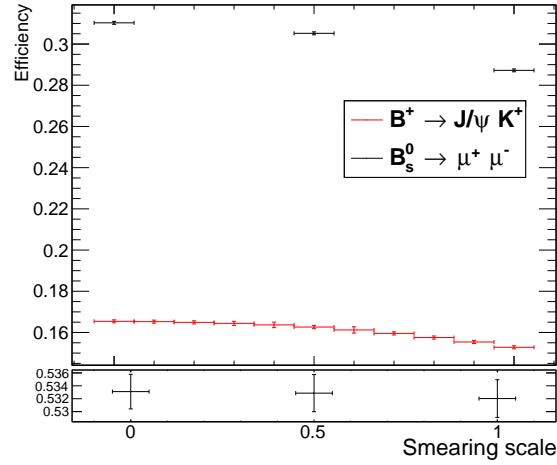


Figure 52: Reconstruction and selection efficiency for $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow \mu^+ \mu^-$ as a function of the impact parameter smearing scale. The ratio, shown below, is constant as a function of this scale.

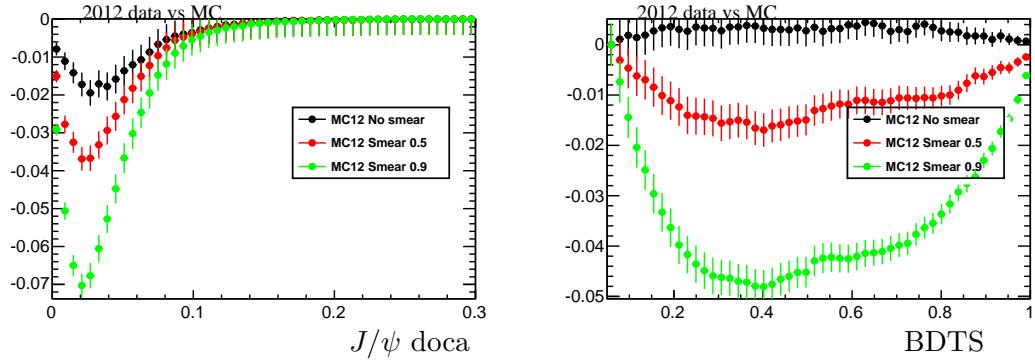


Figure 53: Difference between the cumulative distributions of the sideband subtracted histograms of $B^+ \rightarrow J/\psi K^+$ candidates for 2012 data and three different MC smearing scales.

cumulative distributions of the sideband subtracted histograms of $B^+ \rightarrow J/\psi K^+$ of 2012 data and three different smearing scales¹⁴. For this reason we have decided to use the unsmeared (normal) MC as baseline for all the efficiency studies and use the smeared one only as cross-check. It has to be noticed that in every part of the analysis, with the exception of the normalisation of exclusive channels, only efficiency ratios are employed, so that as shown the smearing scale is irrelevant.

¹⁴This is nothing else than a different way of performing the Kolmogorov-Smirnov test.

Table 43: Reconstruction and selection efficiencies for signal and control channels. The efficiencies include the efficiency of the fiducial cuts as well as the BDTS cut efficiencies, but do not include the GhostProb cut efficiency evaluated separately.

channel	ε
$B_s^0 \rightarrow \mu^+ \mu^-$	$(31.42 \pm 0.05)\%$
$B^0 \rightarrow K^+ \pi^-$	$(25.80 \pm 0.04)\%$
$B^+ \rightarrow J/\psi K^+$	$(16.51 \pm 0.06)\%$

Reconstruction and selection efficiencies are shown in Table 43 where the uncertainties are only from the MC statistics.

As the they cancel in the ratio of efficiencies, we have not assigned systematic uncertainties for the cuts that are identical between the signal and the control channel. Systematic studies on the ratio of selection efficiencies has been estimated on the cuts that are different between the signal and the control channel. The biggest difference in the selection efficiencies for the $B^+ \rightarrow J/\psi K^+$ normalization channel comes from additional IP χ^2 requirements on the extra tracks in the normalization channel, and, to a smaller extent, from the different kinematics of the muons from the J/ψ and the B_s^0 .

8.2.2 Correction of the tracking efficiency

The ratio of the reconstruction efficiencies depends on the reconstruction efficiency of an extra track (the kaon) when the normalization is done with the $B^+ \rightarrow J/\psi K^+$ channel 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 map as provided by the tracking group [49]. 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:

$$c \left(\frac{B^+ \rightarrow J/\psi K^+}{B_s^0 \rightarrow \mu^+ \mu^-} \right) = 0.991 \pm 0.007 \pm 0.004 \pm 0.013$$

$$c \left(\frac{B^0 \rightarrow K^+ \pi^-}{B_s^0 \rightarrow \mu^+ \mu^-} \right) = 0.9996 \pm 0.0059 \pm 0.004 \pm 0.027$$

where the first error is the statistical error from the tracking efficiency map. The second is the correlated systematic error for the mapping method and the third accounts for the uncertainty in the hadronic interactions with material [49]. 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 43 for the computation of the normalization factor.

At the same time we also calculated this correction factor for the ratio of efficiencies of the $B^+ \rightarrow J/\psi K^+$ to the $B_s^0 \rightarrow J/\psi \phi$ channel. While this last channel is not used in the

normalisation explicitly, it is used in order to cross-check the ratio of B meson production fractions as described in Sec. 8.4. The obtained value for this correction is

$$c \left(\frac{B_s^0 \rightarrow J/\psi \phi}{B^+ \rightarrow J/\psi K^+} \right) = 0.992 \pm 0.019 \pm 0.004 \pm 0.013 . \quad (22)$$

8.2.3 Muon detector acceptance and muonID efficiency

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, both on 2011, 2012 data, and in simulated events.

The data over MC ratio of the acceptance×efficiency 2D map in p and p_T bins has been folded into the p, p_T spectrum of the muons from reconstructed and selected $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^+ \rightarrow J/\psi K^+$ Monte Carlo events. The average values $C_{\mu ID}(B_s^0 \rightarrow \mu^+ \mu^-, B^+ \rightarrow J/\psi K^+, B_s^0 \rightarrow J/\psi \phi) = <\epsilon_\mu \times acc_\mu>_{\text{data}} / <\epsilon_\mu \times acc_\mu>_{\text{MC}}$ are 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 $C_{\mu ID}(B_{(s)}^0 \rightarrow h^+ h^-) = <acc_\mu>_{\text{data}} / <acc_\mu>_{\text{MC}}$.

The corrections to be applied to the muonID efficiencies determined on simulated events are listed in Table 44 for $B_s^0 \rightarrow \mu^+ \mu^-$, $B^+ \rightarrow J/\psi K^+$ and $B_{(s)}^0 \rightarrow h^+ h^-$ events, separately for 2011 and 2012 data.

Table 44: Correction factors to be applied to the muon Id as determined from MC.

data set	$C_{\mu ID}(B_s^0 \rightarrow \mu^+ \mu^-)$	$C_{\mu ID}(B^+ \rightarrow J/\psi K^+)$	$C_{\mu ID}(B_{(s)}^0 \rightarrow h^+ h^-)$
2011	$1.0077(15)_{\text{stat}}(41)_{\text{syst}}$	$1.0104(17)_{\text{stat}}(32)_{\text{syst}}$	$1.0011(8)_{\text{stat}}(7)_{\text{syst}}$
2012	$1.0109(13)_{\text{stat}}(152)_{\text{syst}}$	$1.0068(14)_{\text{stat}}(14)_{\text{syst}}$	$0.9997(8)_{\text{stat}}(21)_{\text{syst}}$

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

$$\frac{C_{\mu ID}(B^+ \rightarrow J/\psi K^+)}{C_{\mu ID}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)}(2011) = 1.0027 \pm 0.0013_{\text{stat}} \pm 0.0072_{\text{syst}} \quad (23)$$

and

$$\frac{C_{\mu ID}(B^+ \rightarrow J/\psi K^+)}{C_{\mu ID}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)}(2012) = 0.9959 \pm 0.0011_{\text{stat}} \pm 0.0140_{\text{syst}}, \quad (24)$$

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:

$$\frac{C_{\mu ID}(B_{(s)}^0 \rightarrow h^+ h^-)}{C_{\mu ID}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)}(2011) = 0.9935 \pm 0.0010_{\text{stat}} \pm 0.0034_{\text{syst}} \quad (25)$$

and

$$\frac{C_{\mu ID}(B_{(s)}^0 \rightarrow h^+h^-)}{C_{\mu ID}(B_{(s)}^0 \rightarrow \mu^+\mu^-)}(2012) = 0.9889 \pm 0.0008_{stat} \pm 0.0130_{syst}, \quad (26)$$

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

Finally, since the PID cut $DLL(K - \pi) < 10 \& DLL(\mu - \pi) > -5$ has been used to reduce the rate of peaking background, the MC has to be corrected for both the loss of the events and the BDT shape due to the cut itself. The latter has been taken into account as described in section 7.2.4. The efficiency correction has been already detailed in section 5 and it is $0.9749 \pm 0.0001_{stat} \pm 0.0027_{syst}$ for 2011 data and $0.9798 \pm 0.0001_{stat} \pm 0.0028_{syst}$ for 2012 data; these contributions have to be added to the one due to the isMuon request.

8.2.4 Ghost probability efficiency

In order to assess the efficiency of the ghost probability (GhostProb) cut applied to $B_s^0 \rightarrow \mu^+\mu^-$ events (GhostProb < 3), this was studied on $B^+ \rightarrow J/\psi K^+$ event data by applying the cut to the two muons from the J/ψ .

In Fig. 54 and 55 the efficiency of the GhostProb cut applied to the single muon is shown as a function of p_T and p_Z for both 2011 and 2012 data. Fitting the data with a constant and with a first order polynomial, a slight dependence on the muon momentum has been found. Note that the fit is only shown for illustrative purpose and not used in correcting for the dependence.

The dependence on the muon kinematics has been accounted for by weighting the single muon GhostProb cut efficiency map in p_T and p_Z (gotten from $B^+ \rightarrow J/\psi K^+$) with the fraction of MC signal events falling in each bin. The sum of the weights multiplied by the efficiencies in each bin is thereafter squared to get the GhostProb cut efficiency for the $B_s^0 \rightarrow \mu^+\mu^-$ events. The measured efficiency for the two muons in the signal channel is: $(99.70 \pm 0.01)\%$ in 2011 and. $(99.59 \pm 0.01)\%$ in 2012.

The muon GhostProb cut efficiency has been also measured with the same tag and probe technique and in the same samples employed for the evaluation of the muonID efficiency (see Sect. 5). In particular, using *JpsiFromBinclusive* sample, we found a difference with respect to the previous method of 0.04% for 2011 and 0.09% for 2012 data respectively. These residual differences have been taken as the systematic error due to the choice of the method and of the sample used for the GhostProb cut efficiency determination.

The GhostProb cut efficiency for the normalisation channel $B^+ \rightarrow J/\psi K^+$ could be measured directly on data, however since this would imply multiplying and dividing with the same number it was decided to use directly the yield after fiducial cuts, without GhostProb cut applied, and correct for the measured signal candidate yield for the GhostProb cut efficiency.

The efficiency for the normalisation channel $B^0 \rightarrow K^+\pi^-$ has also been estimated on the $B^+ \rightarrow J/\psi K^+$. We estimated the efficiency of the GhostProb cut applied to the kaon on data, which yields: $(99.53 \pm 0.01)\%$. For the $B^0 \rightarrow K^+\pi^-$ we assume that

kaon and pion efficiency to be equal and therefore take the square of this efficiency, *i.e.* $(99.057 \pm 0.015)\%$.

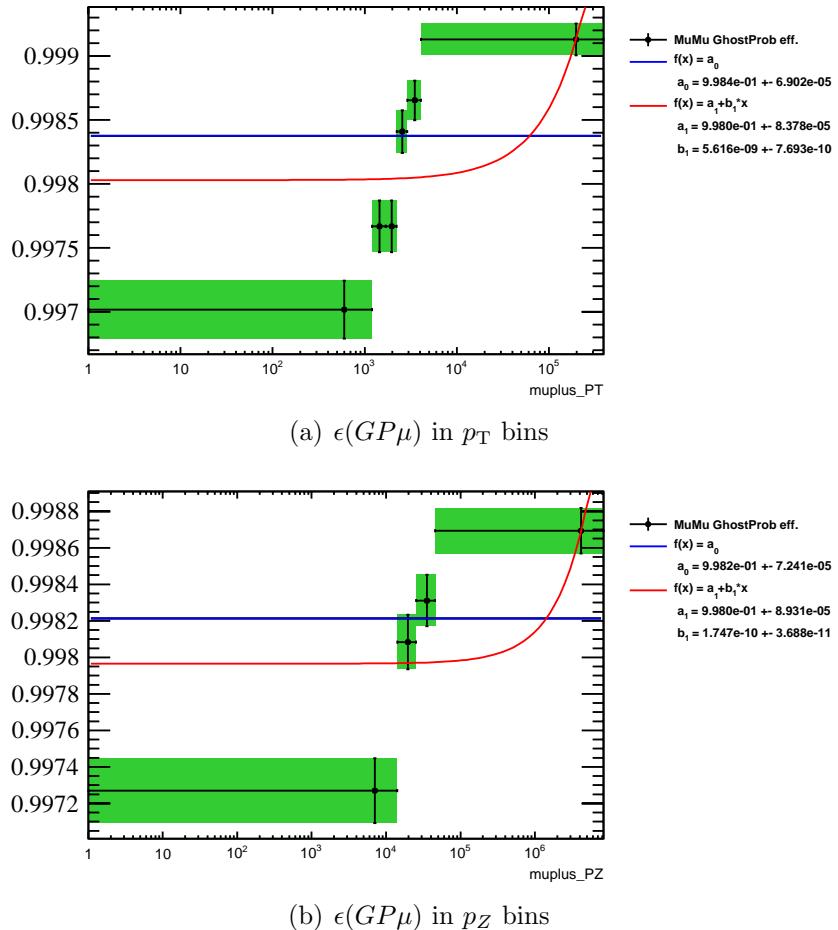


Figure 54: The $GhostProb < 0.3$ cut efficiency on a single muon from $B^+ \rightarrow J/\psi K^+$ in 2011 data.

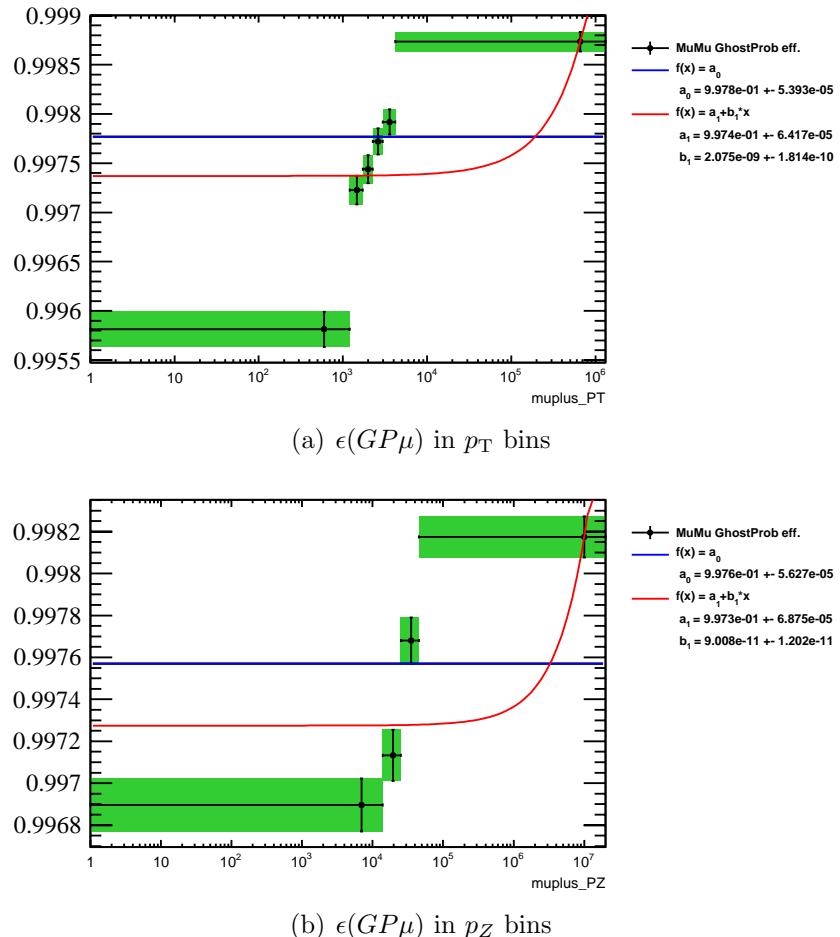


Figure 55: The *GhostProb* < 0.3 cut efficiency on a single muon from $B^+ \rightarrow J/\psi K^+$ in 2012 data.

8.2.5 Invariant mass cut efficiency

When estimating the number of signal events with a fit to the invariant mass distribution the invariant mass is considered in the range [4900, 6000] MeV/ c^2 . On the other hand when calculating the expected number of events per bin of invariant mass and BDT for the CLs limit estimate, the invariant mass is restricted to a ± 60 MeV/ c^2 around the B_s^0 (B^0) mass. The efficiency for these two different cuts has to be included in the selection efficiency. The efficiencies for these cuts were estimated in MC simulation, centering the mass window on the MC input values.

The values for the $B_s^0 \rightarrow \mu^+ \mu^-$ decay are:

$$\varepsilon_{[4900,6000] \text{ MeV}/c^2}^{B_s^0 \rightarrow \mu^+ \mu^-} = 99.22\% \quad (27)$$

$$\varepsilon_{B_s^0 MW}^{B_s^0 \rightarrow \mu^+ \mu^-} = 91.94\% \quad (28)$$

$$\varepsilon_{B_s^0 MW}^{B_s^0 \rightarrow \mu^+ \mu^-} = 14.38\% \quad (29)$$

where the latter efficiency is the percentual leakage of $B_s^0 \rightarrow \mu^+ \mu^-$ into the $B^0 \rightarrow \mu^+ \mu^-$ mass window. The values for the B^0 case are instead:

$$\varepsilon_{[4900,6000] \text{ MeV}/c^2}^{B^0 \rightarrow \mu^+ \mu^-} = 98.95\% \quad (30)$$

$$\varepsilon_{B^0 MW}^{B^0 \rightarrow \mu^+ \mu^-} = 92.45\% \quad (31)$$

$$\varepsilon_{B_s^0 MW}^{B^0 \rightarrow \mu^+ \mu^-} = 11.19\% \quad (32)$$

Also in this case there is a certain leakage, but it is not influent for the final results given the $B^0 \rightarrow \mu^+ \mu^-$ branching fraction.

A correction has to be applied to this efficiency in order to take into account the different resolution and possible different central value of the signal in data and MC. The effect of these difference on the invariant mass cut efficiency has been estimated by applying an additional smearing to the mass of the $B_s^0 \rightarrow \mu^+ \mu^-$ MC sample. The size of the additional smearing σ_b was chosen such that $\sigma'_{MC} = \sigma_{data} = \sqrt{\sigma_{MC}^2 + \sigma_b^2}$. Toy Monte Carlo simulations were then generated letting the smearing vary according to a Gaussian centered in σ_b and wide as the error propagated from the data and MC resolutions. The final correction to be multiplied to the efficiency in the invariant mass window was 0.99305 ± 0.00058 for the 2011 sample and 0.99260 ± 0.00058 for the 2012 one. No correction was necessary for the cut in the $[4900, 6000]$ MeV/c^2 range.

The same procedure was repeated for the mass cut on the J/ψ of the $B^+ \rightarrow J/\psi K^+$ control channel but the correction was found negligible, owing to the much smaller Data-MC mass resolution difference and the wider mass window compared to the resolution.

8.2.6 Systematic uncertainties

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

- The effect of the cut on the $\text{IP}\chi^2$ of the kaon which is not balanced in the signal selection was studied in the past and the MC predicted efficiencies were found in agreement with data.
- The correction for the invariant mass cut was calculated also for the $B^+ \rightarrow J/\psi K^+$ channel: in this channel we apply a cut of ± 60 MeV around the J/ψ mass. However since the resolution of the J/ψ is smaller than for the B_s^0 and the agreement of data with MC is better this correction was found negligible at the level of 0.01%.

8.2.7 Total ratio of reconstruction and selection efficiencies

The reconstruction and selection efficiency for the signal and control channels is given in Table 43. Including the acceptance efficiency, the mentioned corrections and the GhostProb cut efficiency (evaluated separately in §8.2.4), the ratios of total reconstruction and selection efficiencies are:

$$\epsilon_{B^+ \rightarrow J/\psi K^+}^{\text{SEL\&REC}} / \epsilon_{B_{(s)}^0 \rightarrow \mu^+ \mu^-}^{\text{SE\&REC}} = 0.473 \pm 0.011 \quad (0.478 \pm 0.012) \quad (33)$$

$$\epsilon_{B^0 \rightarrow K^+ \pi^-}^{\text{SEL\&REC}} / \epsilon_{B_{(s)}^0 \rightarrow \mu^+ \mu^-}^{\text{SEL\&REC}} = 0.840 \pm 0.026 \quad (0.847 \pm 0.024) \quad (34)$$

for the 2012 (2011) sample.

The absolute efficiencies between the unsmeared MC and the over-smeared sample vary by 8-9%, depending on the channel. However, the ratio between signal and the two normalization channels stays constant within less than 1%, because the efficiency change in the signal is cancelled by the corresponding change in the normalization channel.

8.3 Ratio of trigger efficiencies

As in the previous analysis note, the trigger efficiencies are estimated using the TISTOS method [50]. For the $B^+ \rightarrow J/\psi K^+$ control channel, this measurement can be performed directly on the data. For the $B_s^0 \rightarrow \mu^+ \mu^-$ signal, a map from detached $J/\psi \rightarrow \mu^+ \mu^-$ is created in the maximum p_T and IP of the muons which is then applied to the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ spectrum taken from the simulation. Only muon trigger lines are used in this process, the fraction of events which are selected by non-muon triggers is taken as additional uncertainty, see Ref. [36] for detailed discussion.

The trigger lines that select the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ channel are the same lines which select the normalization channels containing muons in the final state, $B^+ \rightarrow J/\psi K^+$. Hence, the trigger efficiency ratio is dominated by the different phase space covered by the muons in the final state.

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

The trigger efficiency for the $B^+ \rightarrow J/\psi K^+$ channels estimated using the TISTOS method for the 2011 data is:

$$\epsilon_{J/\psi}^{\text{TRIG|SEL}} = (88.0 \pm 0.5_{\text{stat}} \pm 2.5_{\text{syst}})\%, \quad (35)$$

where the systematic error is the combination of two errors: one associated to the TISTOS method precision obtained from MC (2.7%) and the second one due to the 1.7% of $B^+ \rightarrow J/\psi K^+$ events that are triggered not using the muon triggers.

The trigger efficiency for the $B^+ \rightarrow J/\psi K^+$ channels estimated using the TISTOS method for the 2012 data is:

$$\epsilon_{J/\psi}^{\text{TRIG|SEL}} = (86.6 \pm 0.3_{\text{stat}} \pm 2.3_{\text{syst}})\%, \quad (36)$$

where the systematic error is the combination of two errors: one associated to the TISTOS method precision obtained from MC (1.5%) and the second one due to the 1.8% of $B^+ \rightarrow J/\psi K^+$ events that are triggered not using the muon triggers.

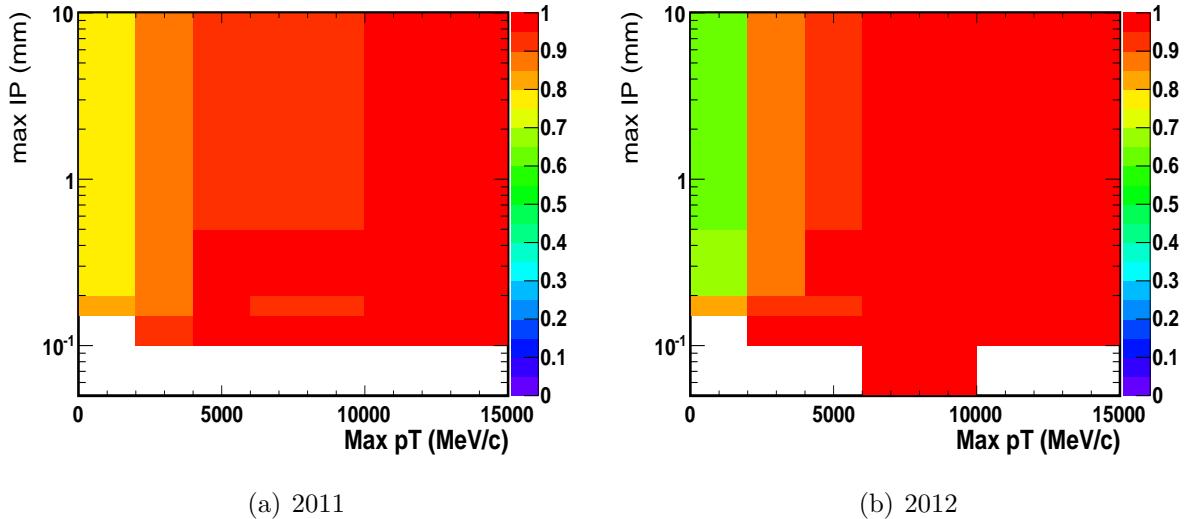


Figure 56: Trigger efficiency map as determined from data as a function of the max p_T and max_{IP} of the muons from $B^+ \rightarrow J/\psi K^+$ events.

To estimate the trigger efficiency for the signal, we have computed an efficiency map (see Fig. 56) as a function of the largest p_T and largest IP of the muons from the $B^+ \rightarrow J/\psi K^+$ 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 App. A, Tab. 56. For the TIS events, all physics lines are used, see App. A, Tab. 57. 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 2011 $B_s^0 \rightarrow \mu^+ \mu^-$ events is:

$$\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{\text{TRIG|SEL}} = (92.1 \pm 0.5_{\text{stat}} \pm 1.6_{\text{syst}})\%, \quad (37)$$

where the systematic error is again the combination of two errors: one associated to the TISTOS method applied to $B_s^0 \rightarrow \mu^+ \mu^-$ (0.1%) and the second one due to the 1.6% of $B_s^0 \rightarrow \mu^+ \mu^-$ events that are triggered not using the muon triggers.

The estimated trigger efficiency for 2012 $B_s^0 \rightarrow \mu^+ \mu^-$ events is:

$$\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{\text{TRIG|SEL}} = (92.4 \pm 0.3_{\text{stat}} \pm 1.9_{\text{syst}})\%, \quad (38)$$

where the systematic error is again the combination of two errors: one associated to the TISTOS method applied to $B_s^0 \rightarrow \mu^+ \mu^-$ (0.8%) and the second one due to the 1.7% of $B_s^0 \rightarrow \mu^+ \mu^-$ events that are triggered not using the muon triggers.

We also evaluate this efficiencies on the merged 2011 and 2012 dataset. We obtain:

$$\epsilon_{J/\psi}^{\text{TRIG|SEL}} = (87.0 \pm 0.3_{\text{stat}} \pm 2.4_{\text{syst}})\%, \quad (39)$$

$$\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{\text{TRIG|SEL}} = (92.4 \pm 0.2_{\text{stat}} \pm 1.8_{\text{syst}})\%. \quad (40)$$

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

$$2011 \text{ data : } \frac{\epsilon_{B^+ \rightarrow J/\psi K^+}^{\text{TRIG|SEL}}}{\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{\text{TRIG|SEL}}} = (95.8 \pm 0.7_{\text{stat}} \pm 1.9_{\text{syst}})\% \quad (41)$$

$$2012 \text{ data : } \frac{\epsilon_{B^+ \rightarrow J/\psi K^+}^{\text{TRIG|SEL}}}{\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{\text{TRIG|SEL}}} = (93.7 \pm 0.5_{\text{stat}} \pm 2.5_{\text{syst}})\%, \quad (42)$$

$$2011 + 2012 \text{ data : } \frac{\epsilon_{B^+ \rightarrow J/\psi K^+}^{\text{TRIG|SEL}}}{\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{\text{TRIG|SEL}}} = (94.3 \pm 0.4_{\text{stat}} \pm 2.3_{\text{syst}})\%, \quad (43)$$

where the systematic error has been evaluated from MC.

8.3.2 Determination of ϵ^{TIS} for the $B_{(s)}^0 \rightarrow h^+ h^-$ normalization

The TIS efficiency is needed for the $B_{(s)}^0 \rightarrow h^+ h^-$ normalization and for the determination of the $B_{(s)}^0 \rightarrow h^+ h^-$ double-misID backgrounds (see Eq. ??). The most precise measurement of ϵ^{TIS} can be done in the $B^+ \rightarrow J/\psi K^+$ channel by binning the B meson phase space in p_T, p_Z and combining the efficiencies. This procedure is described explicitly in Ref. [51].

For 2011 data we find

$$\epsilon^{\text{TIS}}(\text{L0} \times \text{HLT1}) = (5.018 \pm 0.04_{\text{stat}} \pm 0.04_{\text{syst}})\%, \quad (44)$$

where the first error is statistical and it is dominated by the number of TIS events.

For 2012 data we find instead

$$\epsilon^{\text{TIS}}(\text{L0} \times \text{HLT1}) = (5.592 \pm 0.03_{\text{stat}} \pm 0.04_{\text{syst}})\%, \quad (45)$$

Choosing 4 bins in p_Z and 9 bins in p_T has been shown on MC to give the smallest difference between the true value and the one obtained from the TISTOS method. The remaining relative bias of 1% is assigned as a systematic error to the efficiency evaluated on the data.

Asking for also HLT2 trigger level reduces the statistics considerably. For this reason, as we did in the previous analysis note, we consider $\epsilon^{\text{TIS}}(\text{L0} \times \text{HLT1})$ and use the HLT2 efficiencies from the MC and assign a systematic uncertainty to it.

8.3.3 Ratio of trigger efficiency for $B_{(s)}^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow K^+\pi^-$

The ratio of trigger efficiencies of $B^0 \rightarrow K^+\pi^-$ versus the signal $B_{(s)}^0 \rightarrow \mu^+\mu^-$ is given by

$$\frac{\epsilon^{\text{TIS}}(\text{L0} \times \text{HLT1}) \times \epsilon_{B^0 \rightarrow K^+\pi^-}^{\text{HLT2|L0HLT1TIS}}}{\epsilon_{B_s^0 \rightarrow \mu^+\mu^-}^{\text{TRIG|SEL}}} . \quad (46)$$

The HLT2 trigger efficiency, with respect to L0 and HLT1 TIS events of the $B^0 \rightarrow K^+\pi^-$ channel is determined from MC simulations and is

$$\epsilon_{B^0 \rightarrow K^+\pi^-}^{\text{HLT2|L0HLT1TIS}} = 91.5 \pm 0.3\% \quad (47)$$

for 2011 and

$$\epsilon_{B^0 \rightarrow K^+\pi^-}^{\text{HLT2|L0HLT1TIS}} = 91.6 \pm 0.2\% \quad (48)$$

for 2012. The total ratio of trigger efficiencies is therefore:

$$\frac{\epsilon_{B^0 \rightarrow K^+\pi^-}^{\text{TRIG|SEL}}}{\epsilon_{B_s^0 \rightarrow \mu^+\mu^-}^{\text{TRIG|SEL}}} = 0.0587 \pm 0.005 \quad (49)$$

and

$$\frac{\epsilon_{B^0 \rightarrow K^+\pi^-}^{\text{TRIG|SEL}}}{\epsilon_{B_s^0 \rightarrow \mu^+\mu^-}^{\text{TRIG|SEL}}} = 0.0515 \pm 0.005 \quad (50)$$

for the 2012 and 2011 samples respectively.

8.4 Ratio of production fractions

LHCb has determined the fragmentation fraction f_s/f_d in two different ways: using the relative abundance of $B_s^0 \rightarrow D_s^-\pi^+$, $B^0 \rightarrow D^-K^+$ and $B^0 \rightarrow D^-\pi^+$ [52] and using semileptonic $B \rightarrow DX$ decays [53]. In the previous analysis, we used the average of the two LHCb results of $f_s/f_d = 0.267^{+0.021}_{-0.020}$ [53]. Recently, the hadronic measurement has been updated [54] and the combination is updated to [35]

$$f_s/f_d = 0.259 \pm 0.015 , \quad (51)$$

which is the combination of the updated hadronic measurement with the existing semileptonic measurement. This update shows also a dependence of f_s/f_d on the B meson p_T at the level of three sigma. As the BDT is correlated with the p_T (B), it needs to be tested if this dependence introduces a BDT dependence of f_s/f_d . The average p_T of the hadronic measurement is 10.4 GeV whereas the average p_T from MC $B_s^0 \rightarrow \mu^+\mu^-$ candidates varies between 3 GeV (bin 1) and 9 GeV (bin 8). The variation in f_s/f_d corresponding to a variation of the mean p_T from 10.4 GeV to 3 GeV is 0.02, corresponding to 1 sigma of Eq. 51. Therefore, the central value from the combination is used without further correction to account for the p_T dependence of f_s/f_d .

The LHC center-of-mass energy for the dataset used to produce these numbers in 2011 was $\sqrt{s} = 7$ TeV. In 2012, the energy was increased to $\sqrt{s} = 8$ TeV. The effect of the energy increase on the f_s/f_d has been studied by an analysis of the the yield ratio between channels involving relevant meson decays, i.e $B_s^0 \rightarrow J/\psi \phi$ and $B^+ \rightarrow J/\psi K^+$.

The trigger efficiencies in the two control channels are different. In order to deal with the difference, only events triggered by specific muon trigger lines (see Table 45) were considered. The J/ψ in $B_s^0 \rightarrow J/\psi \phi$ and $B^+ \rightarrow J/\psi K^+$ has been shown to possess very similar momentum and transverse momentum distribution, leading to an identical muon trigger efficiency for both modes.

Table 45: Muon trigger lines used for the determination of the hadronization fractions.

L0	<code>LOMuonDecision_TOS</code>
Hlt1	<code>Hlt1DiMuonHighMassDecision_TOS</code> OR <code>Hlt1TrackMuonDecision_TOS</code>
Hlt2	<code>Hlt2DiMuonDetachedJPsiDecision_TOS</code>

The samples of 2012 data were cleaned further by removing the possible ghost tracks with a `GhostProb<0.3` cut and requiring one of the kaons to pass the `PIDK>0` cut. The data from 2011 and 2012 are known to have different tracking and `PIDK` cut efficiencies, which are accounted for. In addition, the `GhostProb<0.3` cut efficiency difference for $B_s^0 \rightarrow J/\psi \phi$ and $B^+ \rightarrow J/\psi K^+$ in 2012 has to be considered.

The cut efficiencies and yields are given in Tab. 46. Both, `PIDK` and `GhostProb` cut efficiencies were estimated on the data, analogously to the efficiencies discussed in Sec. 8.2.4. The efficiency correction for the tracking efficiency is, as discussed in Sec. 8.2, taken from a map provided by the tracking group.

Table 46: Efficiencies and yields for the $B_s^0 \rightarrow J/\psi\phi$ and $B^+ \rightarrow J/\psi K^+$

	2011	2012
$n(B^+)$	$254\,675 \pm 516$	$322\,446 \pm 582$
$n(B_s^0)$	$15\,968 \pm 128$	$20\,519 \pm 145$
	$\epsilon^{Bu/Bs}$	$\epsilon^{Bu/Bs}$
PIDK>0	0.987 ± 0.020	0.983 ± 0.006
GhostProb<0.4373	-	1.014 ± 0.001
Tracking efficiency	Stripping 19 0.9775 ± 0.015	Stripping 19a 1.008 ± 0.019
		1.0001 ± 0.013

The ratio of the $B_s^0 \rightarrow J/\psi\phi$ and $B^+ \rightarrow J/\psi K^+$ signal events in 2011 and 2012 is shown in Fig. 57. The change between the two years is

$$\frac{N^{2012}(B_s^0 \rightarrow J/\psi\phi)/N^{2012}(B^+ \rightarrow J/\psi K^+)}{N^{2011}(B_s^0 \rightarrow J/\psi\phi)/N^{2011}(B^+ \rightarrow J/\psi K^+)} = 1.049 \pm 0.033, \quad (52)$$

which is compatible with a stable value for the hadronization fractions. f_s/f_d estimated with $\sqrt{s} = 7$ TeV (Eq. 51) will then be used in the analysis of 2012 $\sqrt{s} = 8$ TeV data.

8.5 Number of candidates

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

Fig. 58 (a) 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 Crystal Ball function, while the background is modeled with two functions: an exponential for the combinatorial background, and for the physical background a Crystal Ball function 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: $N(B^+ \rightarrow J/\psi K^+) = 355\,232 \pm 608$ in 2011 and $N(B^+ \rightarrow J/\psi K^+) = 761\,122 \pm 891$ in 2012 datasample.

We have assigned a systematic error of 0.3% due to the differences between the result obtained with the fit and the number of candidates obtained after background subtraction. This difference has remained the same after changing from Double Gaussian to Double Crystal Ball in our signal model. The number of signal events contains 0.1% duplicated candidates.

As a cross-check, we have also estimated the number of events from the full sample. The estimated number is $1\,115\,854 \pm 1079$ in excellent agreement with the sum of the two samples ($355\,232(2011) + 761\,122(2012) = 1116354.0$)

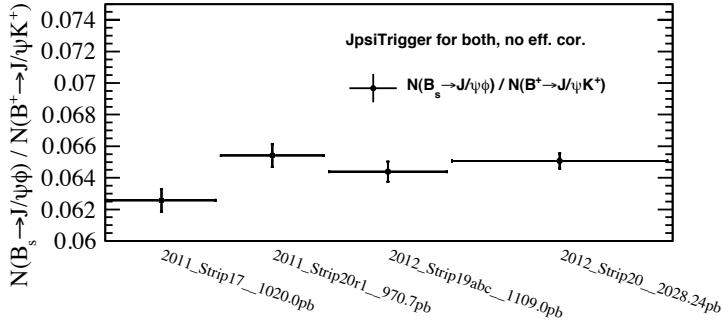


Figure 57: The ratio of $B_s^0 \rightarrow J/\psi\phi$ and $B^+ \rightarrow J/\psi K^+$ signal events in 2011 $\sqrt{s} = 7$ TeV and 2012 $\sqrt{s} = 8$ TeV data, corrected for the relative PID, tracking and ghost probability efficiencies.

8.5.2 Normalization to exclusive $B^0 \rightarrow K^+\pi^-$

The second normalisation channel used in this analysis is $B^0 \rightarrow K^+\pi^-$. To separate the $B^0 \rightarrow K^+\pi^-$ candidates from the inclusive $B_{(s)}^0 \rightarrow h^+h^-$ sample, the same method as discussed in Sec. 7.2.1 is used. The distribution of events is shown in Fig. 60. The total number of $B^0 \rightarrow K^+\pi^-$ candidates was found to be:

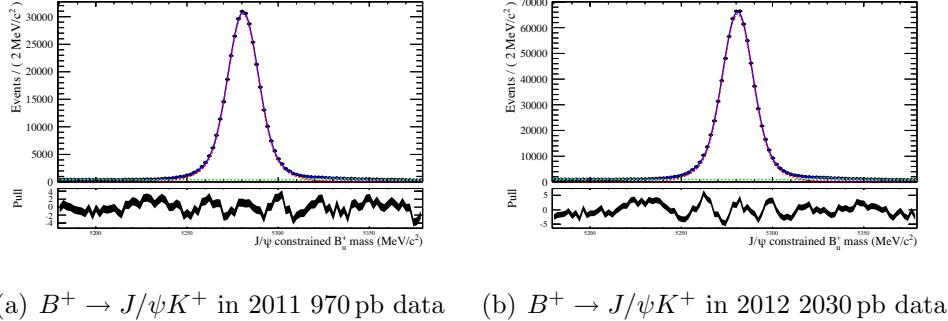
2011: $N(B^0 \rightarrow K^+\pi^-) = 10'809 \pm 439$

2012: $N(B^0 \rightarrow K^+\pi^-) = 26'749 \pm 447$.

8.6 Normalization factor

The normalization factor α_{cal} for the different control channels and the factors that enters in its calculation, that has been presented along this Section, are listed in Table 47.

Weighted averages taking the uncertainty on f_d/f_s to be correlated between the $B^+ \rightarrow J/\psi K^+$ and $B^0 \rightarrow K^+\pi^-$ are reported in Table 48 for the 2011 and 2012 datasets. We report both the values with and without the mass window cut (MW) ($\Delta m < 60$ MeV/ c^2) in the $B_q^0 \rightarrow \mu^+\mu^-$ signal, and after the proper time acceptance, as discussed in Sec. 10.2 which are used for the computation of the limits and for the final fit respectively.



(a) $B^+ \rightarrow J/\psi K^+$ in 2011 970 pb data (b) $B^+ \rightarrow J/\psi K^+$ in 2012 2030 pb data

Figure 58: Invariant mass distribution of the $B^+ \rightarrow J/\psi K^+$ candidates after the stripping selection in reprocessed 2011(left) and 2012 (right) data

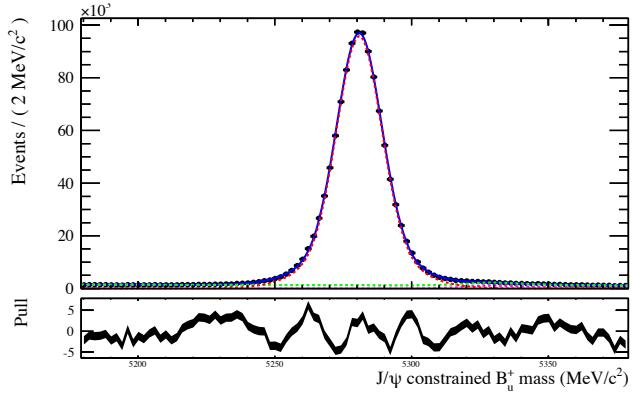


Figure 59: Invariant mass distribution of the $B^+ \rightarrow J/\psi K^+$ candidates in the full data sample.

The normalisation factor for the combined dataset is calculated as follows:

$$N_{B_q^0 \rightarrow \mu^+ \mu^-} = N_{B_q^0 \rightarrow \mu^+ \mu^-}^{2011} + N_{B_q^0 \rightarrow \mu^+ \mu^-}^{2012} = \left(\frac{1}{\alpha^{2011}} + \frac{1}{\alpha^{2012}} \right) \cdot \mathcal{B}(B_q^0 \rightarrow \mu^+ \mu^-) \equiv \frac{1}{\alpha} \cdot \mathcal{B}(B_q^0 \rightarrow \mu^+ \mu^-) \quad (53)$$

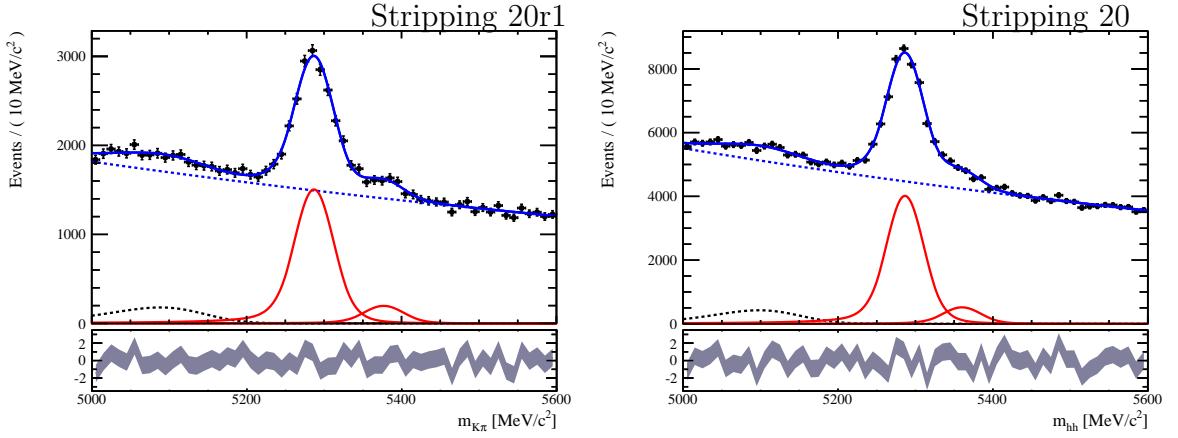


Figure 60: Invariant mass distribution of $B_{(s)}^0 \rightarrow K^\pm \pi^\mp$ candidates (left: 2011 data, right: 2012 data). The yield for $B^0 \rightarrow K^\pm \pi^\mp$ is extracted from a binned maximum-likelihood fit where the two red distribution are describing the $B^0 \rightarrow K^\pm \pi^\mp$ and $B_s^0 \rightarrow K^\pm \pi^\mp$ component, the blue dashed line the combinatorial background and the black dashed line the background coming from partially reconstructed B decays.

Table 47: Summary of the factors and their uncertainty entering in the normalization for the 2011 and 2012 sample for the two normalisation channels. The value of $f_d/f_s = 0.259 \pm 0.015$ is used, as discussed in Sec. 8.4.

(a) 2012 sample						
\mathcal{B}	$\frac{\epsilon_{\text{cal}}^{\text{REC}} \epsilon_{\text{cal}}^{\text{SEL REC}}}{\epsilon_{\text{sig}}^{\text{REC}} \epsilon_{\text{sig}}^{\text{SEL REC}}}$ $(\times 10^{-5})$	$\frac{\epsilon_{\text{cal}}^{\text{TRIG SEL}}}{\epsilon_{\text{sig}}^{\text{TRIG SEL}}}$	N_{cal}	$\alpha_{B_d \rightarrow \mu^+ \mu^-}^{\text{cal}}$ $(\times 10^{-11})$	$\alpha_{B_s \rightarrow \mu^+ \mu^-}^{\text{cal}}$ $(\times 10^{-10})$	
$B^+ \rightarrow J/\psi K^+$	6.025 ± 0.205	0.473 ± 0.011	0.937 ± 0.030	761122 ± 2451	3.51 ± 0.18	1.35 ± 0.11
$B^0 \rightarrow K^+ \pi^-$	1.94 ± 0.06	0.840 ± 0.026	0.0587 ± 0.0024	26749 ± 447	3.57 ± 0.30	1.38 ± 0.15
(b) 2011 sample						
\mathcal{B}	$\frac{\epsilon_{\text{cal}}^{\text{REC}} \epsilon_{\text{cal}}^{\text{SEL REC}}}{\epsilon_{\text{sig}}^{\text{REC}} \epsilon_{\text{sig}}^{\text{SEL REC}}}$ $(\times 10^{-5})$	$\frac{\epsilon_{\text{cal}}^{\text{TRIG SEL}}}{\epsilon_{\text{sig}}^{\text{TRIG SEL}}}$	N_{cal}	$\alpha_{B_d \rightarrow \mu^+ \mu^-}^{\text{cal}}$ $(\times 10^{-11})$	$\alpha_{B_s \rightarrow \mu^+ \mu^-}^{\text{cal}}$ $(\times 10^{-10})$	
$B^+ \rightarrow J/\psi K^+$	6.025 ± 0.205	0.478 ± 0.009	0.955 ± 0.020	355232 ± 1227	7.74 ± 0.36	2.99 ± 0.23
$B^0 \rightarrow K^+ \pi^-$	1.94 ± 0.06	0.847 ± 0.024	0.0501 ± 0.0212	10809 ± 439	7.62 ± 0.77	2.94 ± 0.35

Table 48: Weighted averages of the normalisation factors. For the ones for $B_s^0 \rightarrow \mu^+ \mu^-$ the f_d/f_s uncertainty is treated as correlated. The average normalization factors are corrected for the proper time acceptance, as discussed in Sec. 10.2.

	$\alpha_{B_s^0 \rightarrow \mu^+ \mu^-}$ (10^{-11})	$\alpha_{B_s^0 \rightarrow \mu^+ \mu^-}$ (10^{-10})		
	w/ MW cut	w/o MW cut	w/ MW cut	w/o MW cut
2011	8.33 ± 0.36	7.61 ± 0.32	3.12 ± 0.23	2.85 ± 0.21
2012	3.81 ± 0.17	3.47 ± 0.15	1.43 ± 0.11	1.30 ± 0.10

In order to take into account the correlated input parameters, the actual formula is

$$\alpha_{\text{cal}} = \mathcal{B}_{\text{cal}} \cdot \frac{f_{\text{cal}}}{f_{B_q^0}} \cdot \left(\frac{\varepsilon_{\text{cal}}^{2011}}{\varepsilon_{\text{sig}}^{2011} \cdot N_{\text{cal}}^{2011}} + \frac{\varepsilon_{\text{cal}}^{2012}}{\varepsilon_{\text{sig}}^{2012} \cdot N_{\text{cal}}^{2012}} \right)^{-1} \quad (54)$$

The values of the obtained normalisation parameters (excluding the signal invariant mass cut, but including the proper time acceptance effect) are:

$$\alpha_s = (8.93 \pm 0.62) \cdot 10^{-11} \quad (55)$$

$$\alpha = (2.38 \pm 0.09) \cdot 10^{-11} . \quad (56)$$

this translates in a number of expected signal events for the Standard Model branching fractions of:

$$N_{B_s^0 \rightarrow \mu^+ \mu^-}^{\text{exp}} = 39.5 \pm 4.2 \quad (57)$$

$$N_{B^0 \rightarrow \mu^+ \mu^-}^{\text{exp}} = 4.5 \pm 0.4 \quad (58)$$

in the full BDT12 range and in [4900, 6000] MeV/ c^2 .

8.7 Stability of the yields per pb

The stability of the yields of $B^+ \rightarrow J/\psi K^+$ candidates per pb $^{-1}$ is checked along the data taking after having corrected for the TISTOS trigger efficiency per each sample.

Extra checks were performed on the 2011 Strip20r1 subsample yields as the 4th point (2011 Strip20r1 D) exceeds the average. The yields per pb $^{-1}$ for the smallest samples available (12 files) within the Stripping 20r1 nTuple revealed an effect of grouping - by chance the files with slight upward fluctuation in their yields have been grouped together in the 4th point. The individual 12 yields from the files agree to each other within 1σ .

Note that the samples are not samples in time but arbitrary collection of runs. We also checked for correlation between the higher yields/ pb $^{-1}$ and multiplicity variables, none was found.

Fig. 61 shows the trigger efficiency corrected yield per pb. The average yields for 2011 and 2012 data are indicated by coloured lines.

8.8 Cross-check of the ratio of branching fractions of $B^0 \rightarrow K^+ \pi^-$ to $B^+ \rightarrow J/\psi K^+$

As additional cross-check on the normalisation we computed the ratio of branching fractions between the two normalisation channels, *i.e.*

$$\frac{\mathcal{B}(B^0 \rightarrow K^+ \pi^-)}{\mathcal{B}(B^+ \rightarrow J/\psi K^+)} = \frac{N_{B^0 \rightarrow K^+ \pi^-}}{N_{B^+ \rightarrow J/\psi K^+}} \cdot \frac{\epsilon_{B^+ \rightarrow J/\psi K^+}^{\text{SEL\&REC}}}{\epsilon_{B^0 \rightarrow K^+ \pi^-}^{\text{SEL\&REC}}} \frac{\epsilon_{B^+ \rightarrow J/\psi K^+}^{\text{TRIG|SEL}}}{\epsilon_{B^0 \rightarrow K^+ \pi^-}^{\text{TRIG|SEL}}}$$

The ratio of selection and reconstruction efficiencies of the two channels was calculated to be 0.604 ± 0.013 , while the trigger efficiency ratio is 16.24 ± 0.18 . Being this only a cross-check not all the systematic uncertainties have been properly recomputed and accounted for in this ratio. Therefore the total error is a bit underestimated. The measured value of the ratio of branching ratios is:

$$r^{2012} = \frac{\mathcal{B}(B^0 \rightarrow K^+ \pi^-)}{\mathcal{B}(B^+ \rightarrow J/\psi K^+)} = 0.316 \pm 0.024$$

$$r^{2011} = \frac{\mathcal{B}(B^0 \rightarrow K^+ \pi^-)}{\mathcal{B}(B^+ \rightarrow J/\psi K^+)} = 0.325 \pm 0.030$$

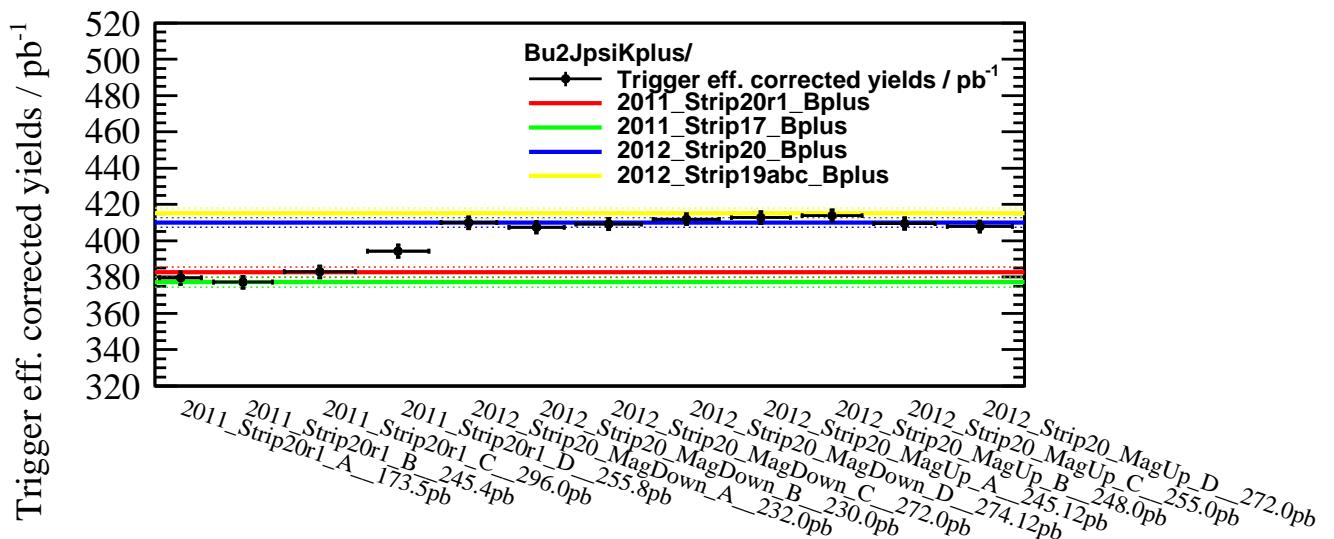


Figure 61: Trigger efficiency corrected signal yields per pb for $B^+ \rightarrow J/\psi K^+$ candidates after the stripping selection; coloured lines indicate the average yields for each sample and reconstruction: red line is stripping20r1, green line is stripping17, blue is stripping 20 and yellow is stripping 19.

while the ratio of the PDG values is

$$\frac{\mathcal{B}^{\text{PDG}}(B^0 \rightarrow K^+ \pi^-)}{\mathcal{B}^{\text{PDG}}(B^+ \rightarrow J/\psi K^+)} = 0.322 \pm 0.015$$

so that good compatibility is found within the uncertainties.

In order to compare also the stability of our measurements we can consider the ratio of this ratio for the two years, which yields:

$$r^{2011}/r^{2012} = 1.03 \pm 0.12 .$$

8.9 Normalization of exclusive backgrounds

In order to have precise estimates of the number of expected background events, we have decided to normalise the considered channels directly with data using $B^+ \rightarrow J/\psi K^+$ events; the only exception being the $B_{(s)}^0 \rightarrow h^+ h^-$ peaking background which is normalised to $B^0 \rightarrow K^+ \pi^-$ events. In this way we are not limited by the uncertainties on luminosity and cross-section.

The number of expected events for a given channel is therefore:

$$\begin{aligned} N_{\text{exp}}^x &= N^{B^+ \rightarrow J/\psi K^+} \frac{f_x}{f_u} \frac{\mathcal{B}(x)}{\mathcal{B}(B^+ \rightarrow J/\psi K^+)} \frac{\varepsilon_{\text{tot}}^x}{\varepsilon_{\text{tot}}^{B^+ \rightarrow J/\psi K^+}} \\ &= \beta_x \cdot \varepsilon_{\text{tot}}^x \cdot \mathcal{B}(x) \end{aligned}$$

where β_x , defined as:

$$\beta_x = \frac{N^{B^+ \rightarrow J/\psi K^+}}{\mathcal{B}(B^+ \rightarrow J/\psi K^+) \cdot \varepsilon_{\text{tot}}^{B^+ \rightarrow J/\psi K^+}} \cdot \frac{f_x}{f_u}$$

is nothing else than the yield of B_x produced in LHCb in the given luminosity. Using the input numbers already calculated for the α parameter we get the following values:

$$\beta^{2011} = (2.56 \pm 0.12) \cdot 10^{11} \quad (59)$$

$$\beta^{2012} = (5.64 \pm 0.25) \cdot 10^{11} \quad (60)$$

where β_s is calculated from β and the average value of f_s/f_d (always assuming $f_d = f_u$).

In order to take into account the correlated parameters, the normalisation for the combined dataset is calculated as follows:

$$\beta_x = \frac{1}{\mathcal{B}(B^+ \rightarrow J/\psi K^+)} \left(\frac{N_{2011}^{B^+ \rightarrow J/\psi K^+}}{\varepsilon_{2011}^{B^+ \rightarrow J/\psi K^+}} + \frac{N_{2012}^{B^+ \rightarrow J/\psi K^+}}{\varepsilon_{2012}^{B^+ \rightarrow J/\psi K^+}} \right) \cdot \frac{f_x}{f_u} , \quad (61)$$

and the obtained values are:

$$\beta = (8.19 \pm 0.33) \cdot 10^{11} \quad (62)$$

$$\beta_s = (2.12 \pm 0.16) \cdot 10^{11} . \quad (63)$$

9 Combinatorial background estimate

The background yield leaking into B_s^0 and B^0 mass windows is dominated by combinatorics of two muons coming from different B hadrons, with the $B_{(s)}^0 \rightarrow h^+h^{(\prime)-}$ and $\Lambda_b^0 \rightarrow p\mu^-\nu$ misID backgrounds also playing a role, especially in the B^0 window. The yield of the combinatorial background in the signal window is evaluated by interpolating the mass sidebands, for each BDT bin separately.

To this purpose, the invariant mass distribution has been fitted in the range $4900 < m_{\mu\mu} < 5224.9$ MeV/ c^2 and $5431.9 < m_{\mu\mu} < 6000$ MeV/ c^2 , assuming a single exponential function for the combinatorial background. The lower boundary at 4900 MeV/ c^2 was chosen to exclude background sources like cascading $b \rightarrow c\mu X \rightarrow \mu\mu X$, while the intermediate mass values are set at $m(B^0) - 60$ MeV/ c^2 and $m(B_s^0) + 60$ MeV/ c^2 respectively, which corresponds to more than 2 times the mass resolution. For both 2011 and 2012 data we used the same 8 BDT binning scheme optimised in Ref. [22], with bin edges defined as follows: $0. - 0.25 - 0.4 - 0.5 - 0.6 - 0.7 - 0.8 - 0.9 - 1.0$.

Since the previous version of the analysis, Ref. [18] many exclusive B decay channels have been considered which may pollute the mass sidebands, and thus affect the combinatorial background estimate. The results of our estimates are shown in Table 29 of Sec. 6.6, which summarizes the expected yields for all of the relevant channels in the mass range [4900-6000] MeV/ c^2 , and for BDT above 0.8. The following strategy has been adopted to treat the various exclusive background components in the combinatorial background interpolation:

- The $B^0 \rightarrow \pi^-\mu^+\nu_\mu$, $B_{(s)}^0 \rightarrow h^+h^{(\prime)-}$ and $B^{0(+)} \rightarrow \pi^{0(+)}\mu^+\mu^-$ components have been included as separate pdfs in the basic version of the background interpolation (“fit1” in the following): they represent 81% of the exclusive background in the considered mass and BDT range, and moreover they are all measured decay channels; all of the other background components are absorbed in the combinatorial background normalization, which is a free parameter of the fit;
- The $\Lambda_b^0 \rightarrow p\mu^-\nu$ gives a sizable contribution to the yield, but is largely unknown. It has been added as a systematic check, consisting of performing the fit with this component (“fit2” in the following) and comparing the results for the total background yield in the signal mass window with the previous case;
- Starting from fit2 configuration, we also accounted for the $B_s^0 \rightarrow K^-\mu^+\nu_\mu$, which is less relevant and doesn’t give appreciable contribution in the signal region: in this case we simply added the expected number of events of $B_s^0 \rightarrow K^-\mu^+\nu_\mu$ to the $B^0 \rightarrow \pi^-\mu^+\nu_\mu$, which has a very similar mass shape, and summing in quadrature the quoted uncertainty on the normalization (“fit3” in the following).
- The $B_c^+ \rightarrow J/\psi\mu^+\nu$ has been neglected, since its contribution is small compared to the other channels, and moreover its mass shape is well approximated by an exponential, like the combinatorial, within the expected statistics.

The expected yields in the combined 2011+2012 data sample, and the mass and BDT PDFs for all the background sources above are discussed in detail in Sec. 6.4.2 ($B^0 \rightarrow \pi^- \mu^+ \nu_\mu$), Sec. 6.4.3 ($B_s^0 \rightarrow K^- \mu^+ \nu_\mu$), Sec. 6.4.4 ($B^{0(+)} \rightarrow \pi^{0(+)} \mu^+ \mu^-$), Sec. 6.3 ($B_{(s)}^0 \rightarrow h^+ h^{(')-}$), and Sec. 6.4.6 ($\Lambda_b^0 \rightarrow p \mu^- \nu$).

For all of the exclusive channels, the normalisation and the BDT fractions are fluctuated in the fit according to their total uncertainties. For technical reasons, the parameters describing the mass PDF (RooPhysBkg) are kept fixed and the fit is then repeated with parameters varied according to their uncertainites to compute the systematics from the mass shapes. The combinatorial background is parameterized with an exponential function, the parameters of which are free to vary in the fit.

The fits to the mass spectra of the 8 BDT bins of the combined 2011+2012 data are shown in Fig. 62 and for the Fig. 63, for fit1 and fit2 configurations respectively. The expected numbers of combinatorial background events in the range $5224.9 < m_{\mu\mu} < 5431.9$ MeV/ c^2 , corresponding to the union of B_s^0 and B^0 search windows, are listed in Table 49, with their symmetric errors¹⁵. From the above results, we can conclude that the

Table 49: Extrapolated combinatorial background events (and $\Lambda_b^0 \rightarrow p \mu^- \nu$ event if fitted as a separate component) in the mass range $5224.9 < m_{\mu\mu} < 5431.9$ MeV/ c^2 , for the combined 2011+2012 dataset, 3 fb; fit1: without $\Lambda_b^0 \rightarrow p \mu^- \nu$ and $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$, fit2: with $\Lambda_b^0 \rightarrow p \mu^- \nu$ as separate component.

BDT range	fit1 N_{comb}	fit2	
		N_{comb}	$N_{\Lambda_b^0 \rightarrow p \mu^- \nu}$
[0.00, 0.25]	10915 ± 53	10910 ± 53	4.0
[0.25, 0.40]	214.0 ± 7.7	211.9 ± 7.8	1.9
[0.40, 0.50]	49.4 ± 3.8	48.3 ± 3.9	1.1
[0.50, 0.60]	21.2 ± 2.6	20.4 ± 2.6	0.9
[0.60, 0.70]	8.4 ± 2.0	7.6 ± 2.2	0.8
[0.70, 0.80]	3.8 ± 1.5	3.5 ± 1.5	0.7
[0.80, 0.90]	1.4 ± 0.8	1.2 ± 0.8	0.5
[0.90, 1.00]	0.4 ± 0.4	0.2 ± 0.4	0.3

introduction of $\Lambda_b^0 \rightarrow p \mu^- \nu$ does not alter significantly the background estimate in signal window: indeed the N_{comb} from fit1 agrees with $N_{comb} + N_{\Lambda_b^0 \rightarrow p \mu^- \nu}$ from fit2 within less than half a sigma¹⁶. The introduction of $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ (fit3) has a negligible impact on the results. For this reason, we choose fit1 as our baseline for combinatorial background

¹⁵For the evaluation of the CL_s confidence levels we use instead the number of events in the sidebands, fluctuated according to Poisson, the fitted fraction of combinatorial events in the sidebands, with its asymmetric error, and the fitted exponential slope, with its asymmetric error.

¹⁶The $B_{(s)}^0 \rightarrow h^+ h^{(')-}$, the only other bkg leaking in the signal region, is always present as independent component in the fit.

interpolation. The systematic error from the exclusive background mass shapes have been found negligible for the two most sensitive bins, as measured by independent fits with parameters of the RooPhysBkg pdf varied according to their expected errors. The systematic error from the exclusive background normalization and the BDT pdf is instead accounted for by fluctuating these parameters in the fit with gaussian constraints.

The results from fit1 is also compared to what obtained on the same data set (s20r1+s20) with the old BDT9, and in the previous versions of the analysis, Ref. [22] and Ref. [18], for 2011 and 2012 data separately. From the above results, we can appreciate an improvement

Table 50: Extrapolated combinatorial background events obtained in the mass range $5224.9 < m_{\mu\mu} < 5431.9$ MeV/ c^2 for different BDT and reconstructions versions; three set of results are reported: combined 2011+2012 (3 fb^{-1}) dataset s20r1+s20 with BDT12, 2011 (1 fb^{-1}) s20r1 and 2012 (2 fb^{-1}) s20 with BDT9, and published results on 2011 (1 fb^{-1}) s17 and 2012 (1.1 fb^{-1}) s19 with BDT9; for stripping 19, the bins above 0.8 were merged.

BDT range	BDT12		BDT9		BDT9				
	s20r1+s20	3 fb^{-1}	s20r1	1 fb^{-1}	s20	2 fb^{-1}	s17	1 fb^{-1}	s19
[0.00, 0.25]	10915 ± 53		3150 ± 30		7516 ± 46		3375 ± 30		4173 ± 33
[0.25, 0.40]	214.0 ± 7.7		74.0 ± 4.7		186.8 ± 7.4		99.6 ± 5.2		100.1 ± 5.2
[0.40, 0.50]	49.4 ± 3.8		16.9 ± 2.4		33.4 ± 3.3		21.5 ± 2.5		23.2 ± 2.5
[0.50, 0.60]	21.2 ± 2.6		6.9 ± 1.6		17.8 ± 2.5		7.5 ± 1.5		7.9 ± 1.5
[0.60, 0.70]	8.4 ± 2.0		4.1 ± 1.3		6.8 ± 1.7		3.0 ± 1.1		3.6 ± 1.0
[0.70, 0.80]	3.8 ± 1.5		0.8 ± 0.6		2.5 ± 1.3		1.6 ± 0.9		0.9 ± 0.7
[0.80, 0.90]	1.4 ± 0.8		0.6 ± 0.5		1.0 ± 0.7		1.3 ± 0.7		
[0.90, 1.00]	0.4 ± 0.4		0.5 ± 0.4		0.35 ± 0.36		0.8 ± 0.7		0.6 ± 0.4

in background rejection in the last 2 bins when using the BDT12 with respect to what obtained with BDT9.

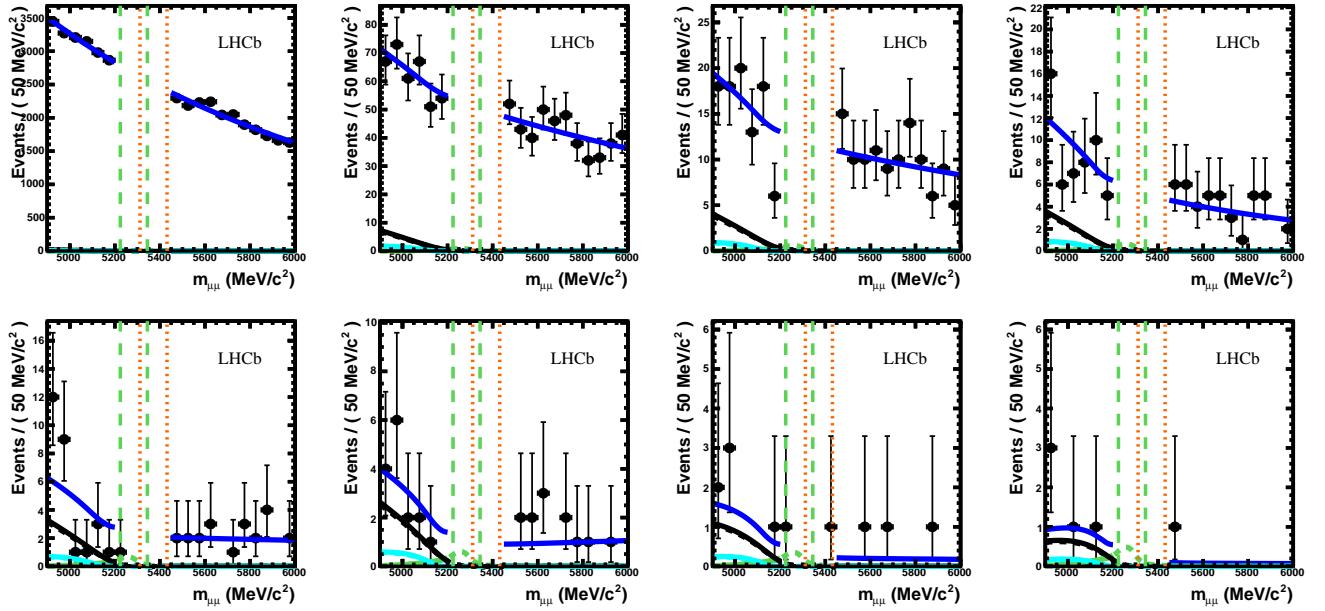


Figure 62: 2011+2012 data, stripping 20r1+20: fit to the dimuon mass sidebands in 8 bins of BDT. Dots are data, black line is $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$, cyan is $B^{0(+)} \rightarrow \pi^{0(+)} \mu^+ \mu^-$, green is $B_{(s)}^0 \rightarrow h^+ h^{(-)}$ misID and blue is total fit. Vertical orange (green) dashed lines indicate the $B_s^0 \rightarrow \mu^+ \mu^-$ ($B^0 \rightarrow \mu^+ \mu^-$) search windows excluded from the background estimation fit.

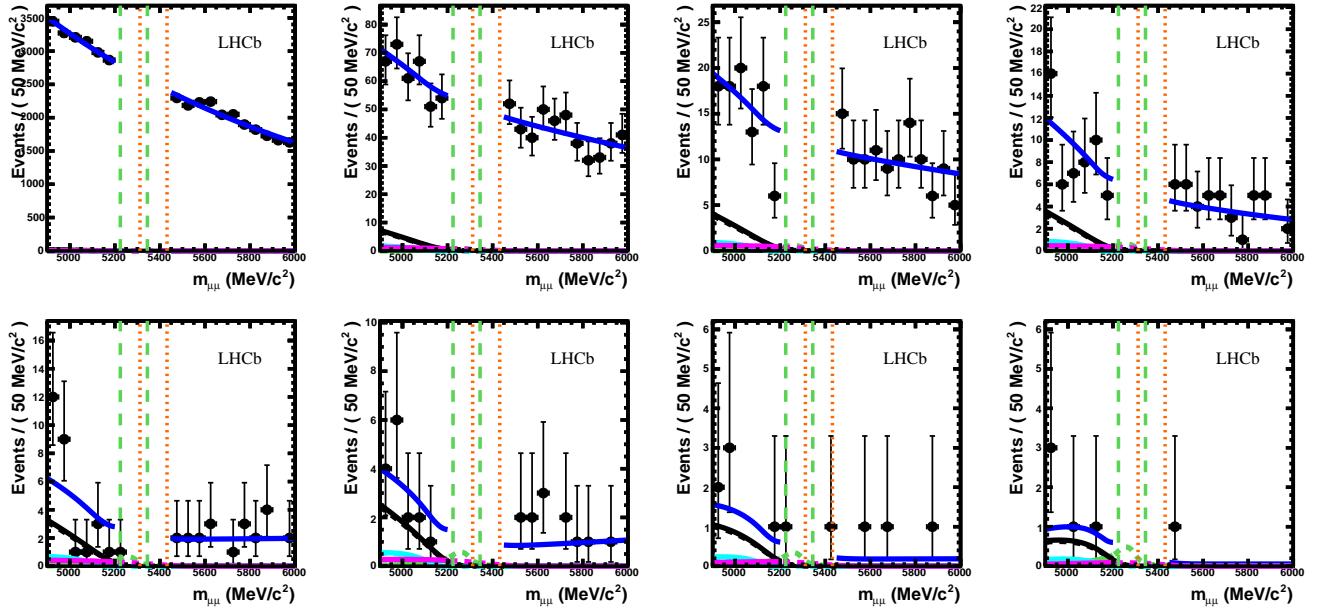


Figure 63: 2011+2012 data, stripping 20r1+20: fit to the dimuon mass sidebands in 8 bins of BDT. Dots are data, black line is $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$, cyan is $B^{0(+)} \rightarrow \pi^{0(+)} \mu^+ \mu^-$, green is $B_{(s)}^0 \rightarrow h^+ h^{(-)}$ misID, purple is the $\Lambda_b^0 \rightarrow p \mu^- \nu$ and blue is total fit. Vertical orange (green) dashed lines indicate the $B_s^0 \rightarrow \mu^+ \mu^-$ ($B^0 \rightarrow \mu^+ \mu^-$) search windows excluded from the background estimation fit.

10 Time dependent effects

10.1 B_s^0 - \bar{B}_s^0 Mixing Effect

The decay time distribution of the $B_s^0 \rightarrow \mu^+ \mu^-$ is in fact the sum of two exponential functions accounting for the fact that the heavy and light B_s^0 mass eigenstates have different decay widths. The relative fraction of the two exponential functions is not known and depends on the model considered.

The subtlety arises in the analysis when selection criteria involving the candidate decay time are applied. For these selections the signal efficiency is computed with simulated data which were produced with some hypothesis for the decay time distribution. We choose to parametrise this distribution with y_s and $\mathcal{A}_{\Delta\Gamma}$ defined as:

$$\begin{aligned} y_s &= \frac{\Gamma_L - \Gamma_H}{\Gamma_L + \Gamma_H} \\ \mathcal{A}_{\Delta\Gamma} &= \frac{\Gamma_{B_{s,H}^0 \rightarrow \mu^+ \mu^-} - \Gamma_{B_{s,L}^0 \rightarrow \mu^+ \mu^-}}{\Gamma_{B_{s,H}^0 \rightarrow \mu^+ \mu^-} + \Gamma_{B_{s,L}^0 \rightarrow \mu^+ \mu^-}}. \end{aligned} \quad (64)$$

y_s has been measured experimentally ($y_s = 0.0615 \pm 0.0085$) in many channels [3], however $\mathcal{A}_{\Delta\Gamma}$ is channel and model dependent. In the SM for $B_s^0 \rightarrow \mu^+ \mu^-$, $\mathcal{A}_{\Delta\Gamma} = 1$, [55] hence the $B_s^0 \rightarrow \mu^+ \mu^-$ decay time distribution is a single exponential distribution of slope Γ_H .

Hence the time integrated efficiency,

$$\epsilon = \frac{\int_0^\infty \Gamma(B_s^0(t) \rightarrow \mu^+ \mu^-, \mathcal{A}_{\Delta\Gamma}, y_s) \epsilon(t) dt}{\int_0^\infty \Gamma(B_s^0(t) \rightarrow \mu^+ \mu^-, \mathcal{A}_{\Delta\Gamma}, y_s) dt}, \quad (65)$$

depends on this hypothesis. Note however that the decay time dependent efficiency $\epsilon(t)$, computed with the simulated data is not affected by this hypothesis.

As a result all the signal time integrated efficiencies of selections based on decay time criteria are model dependent and so are the final analysis results. In the two next Sections we describe in details where the effect explained previously arises.

10.2 Model Dependent Normalisation

The first place where this effect arises is in the initial selection described in Section 4. As shown in Tab. 4, the signal selection involves variables highly correlated with the candidate decay time like the impact parameter of the candidate tracks. Hence the selection is decay time dependent and biases the signal decay time distribution has shown in Fig. 64.

In the MC this decay time distribution is in fact a single exponential with a mean life time of 1.469141 ps. In order to translate the time integrated efficiency obtained with these data to another model (parametrised by y_s and $\mathcal{A}_{\Delta\Gamma}$), the following factor is defined:

$$\begin{aligned} \delta_\epsilon &= \frac{\epsilon^{\mathcal{A}_{\Delta\Gamma}, y_s}}{\epsilon^{MC}} \\ &= \frac{\int_0^\infty \Gamma(B_s^0(t) \rightarrow \mu^+ \mu^-, \mathcal{A}_{\Delta\Gamma}, y_s) \epsilon(t) dt}{\int_0^\infty \Gamma(B_s^0(t) \rightarrow \mu^+ \mu^-, \mathcal{A}_{\Delta\Gamma}, y_s) dt} \times \frac{\int_0^\infty e^{-\Gamma_{MC} t} dt}{\int_0^\infty e^{-\Gamma_{MC} t} \epsilon(t) dt}. \end{aligned} \quad (66)$$

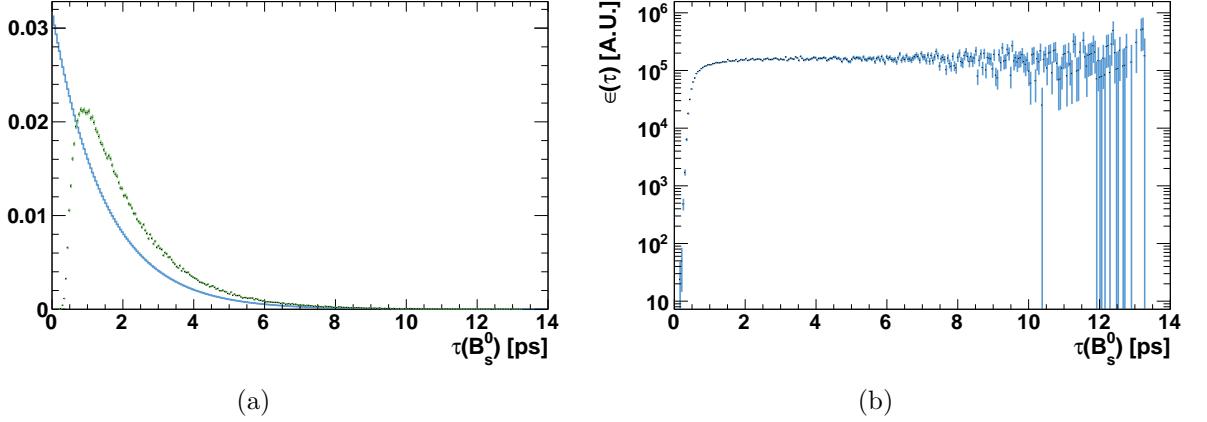


Figure 64: MC simulation: (a) Proper time distribution before (blue) and after (green) the selection and (b) the resulting decay time dependent acceptance shape (arbitrary units).

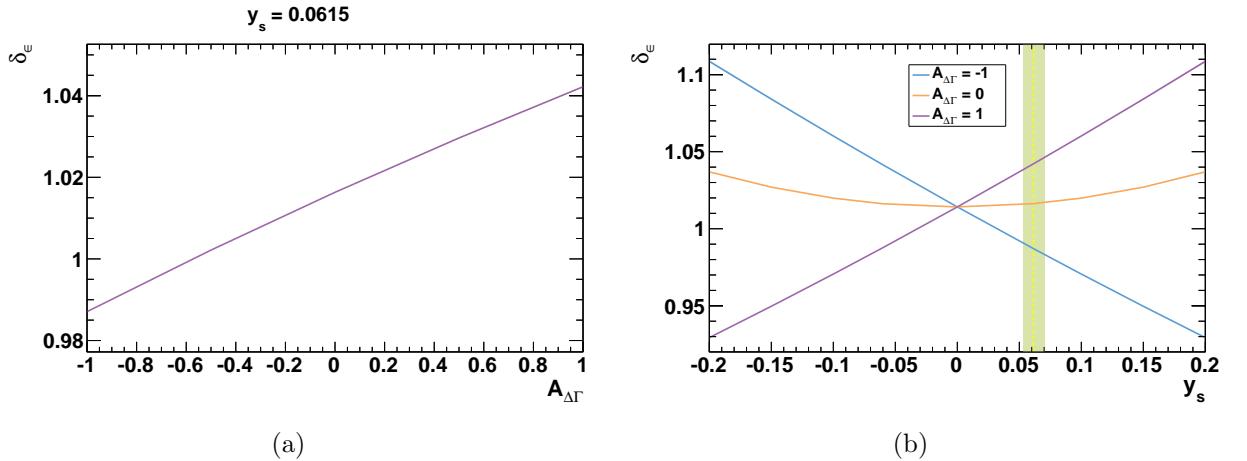


Figure 65: (a) Efficiencies bias as a function of $A_{\Delta\Gamma}$ for the experimental value of $y_s = 0.0615 \pm 0.0085$ [3] and (b) as a function of y_s for different $A_{\Delta\Gamma}$ values. The vertical green band is the experimental averaged y_s value and its uncertainty. The B_s^0 mean life time value used is $\tau_{B_s^0} = 1.516 \pm 0.011$ ps.

This correction factor is shown for $B_s^0 \rightarrow \mu^+ \mu^-$ as a function of y_s and $A_{\Delta\Gamma}$ in Fig. 65. For the SM ($A_{\Delta\Gamma}=1$, $y_s=0.0615 \pm 0.0085$, $\tau_{B_{s,H}^0} = 1.615 \pm 0.021$ ps) the correction factor is:

$$\delta_\epsilon - 1 = (+4.57 \pm 0.02)\%. \quad (67)$$

The uncertainties on this correction are only due to the statistics of the simulated data sample used to compute the decay time dependent efficiency, and they largely cancel since the numerator and denominator in Eq. 66 are highly correlated.

The efficiency for $B^0 \rightarrow \mu^+ \mu^-$ was assumed in the previous analyses to be identical to the one for $B_s^0 \rightarrow \mu^+ \mu^-$. It should therefore also be corrected. Since the life time value used in the simulation is closer from the B^0 ($\tau_{B^0} = 1.519 \pm 0.007$) than for the $B_{s,H}^0$, the B^0 correction factor in the SM is smaller:

$$\delta_\epsilon - 1 = (+1.50 \pm 0.006)\% \quad (68)$$

Note that this correction is not model dependent since the width difference in the B^0 mode is negligible.

For both mode this correction should be included in the analysis, by multiplying the denominator of the normalization factor (see Eq. 21) by δ_ϵ .

10.3 Model Dependent BDT PDF

The effect described in Section 10.1 also enters in the BDT PDF. The BDT uses the candidate decay time as input variable, hence the $B_s^0 \rightarrow \mu^+ \mu^-$ signal decay time distribution in each BDT bin is biased. If this bias was identical in all bins, the PDF would not be affected. However as combinatorial background resemble more short-lived than long-lived signal candidates, the higher the BDT the larger the mean decay time in the BDT bin, as shown in Fig. 66.

Technically in the previous analyses, the BDT PDF is obtained with a calibration method which assumes that the BDT PDF is the same for the $B_{(s)}^0 \rightarrow h^+ h^-$ control channel (cc) and $B_s^0 \rightarrow \mu^+ \mu^-$:

$$\frac{N_{\text{sig}}^i}{N_{\text{sig}}^{\text{tot}}} = \frac{N_{\text{cc}}^i}{N_{\text{cc}}^{\text{tot}}}. \quad (69)$$

This relation is in fact just an approximation which neglects the fact that the signal and the control channels have different decay time distributions and so different time integrated efficiencies for a BDT bin. Without this assumption the BDT PDF expression is

$$\frac{N_{\text{sig}}^{\text{exp},i}}{N_{\text{sig}}^{\text{exp}}} = \frac{N_{\text{cc}}^{\text{exp},i}}{N_{\text{cc}}^{\text{exp}}} \times \frac{\epsilon_{\text{sig}}^i(\mathcal{A}_{\Delta\Gamma}, y_s)}{\epsilon_{\text{cc}}^i} \times \frac{\epsilon_{\text{cc}}}{\epsilon_{\text{sig}}(\mathcal{A}_{\Delta\Gamma}, y_s)}, \quad (70)$$

and is again model dependent. The correction to be applied to the PDF obtained with Eq. 69 is

$$\delta_{\text{PDF}}^i = \frac{\epsilon_{\text{sig}}^i(\mathcal{A}_{\Delta\Gamma}, y_s)}{\epsilon_{\text{cc}}^i} \times \frac{\epsilon_{\text{cc}}}{\epsilon_{\text{sig}}(\mathcal{A}_{\Delta\Gamma}, y_s)}, \quad (71)$$

and is shown in Fig. 67 as a function of $\mathcal{A}_{\Delta\Gamma}$ and y_s for $B^0 \rightarrow h^+ h^-$ as control channel. In the analysis, $B_s^0 \rightarrow h^+ h^-$ control channels are also used but these modes are three times less abundant than $B^0 \rightarrow h^+ h^-$. We therefore neglect them in the PDF correction. We recommend to decide whether or not to include these channels in the calibration process

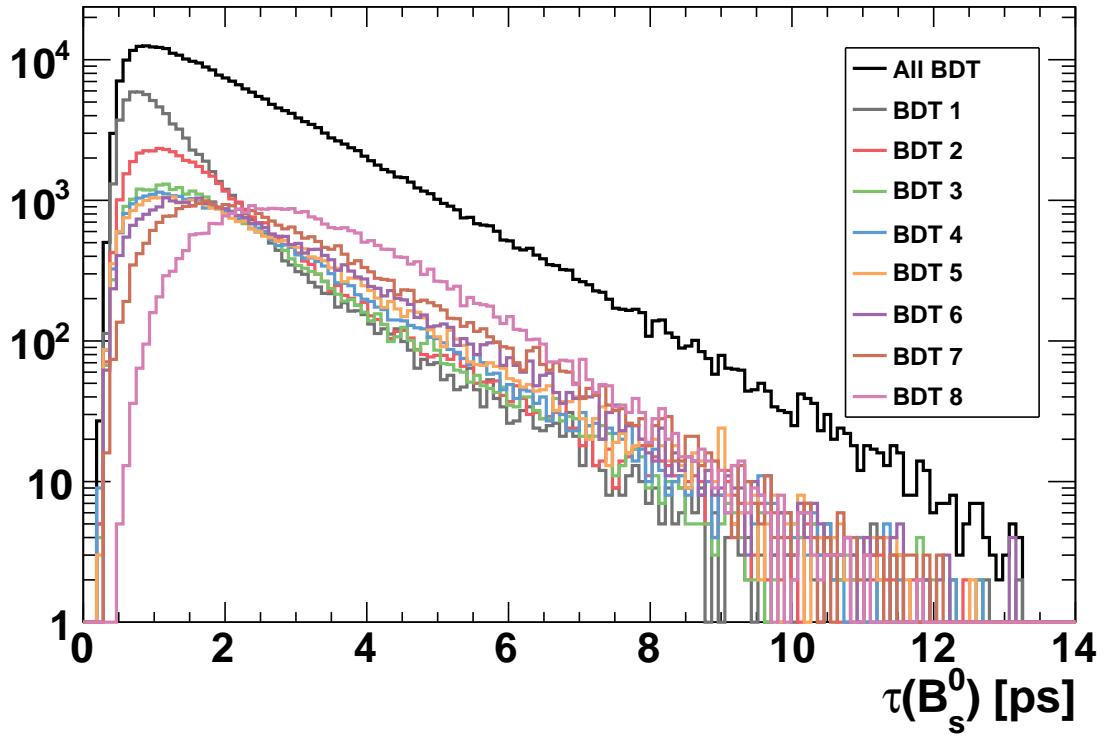


Figure 66: Decay time distributions in each BDT bin and in all BDT bins.

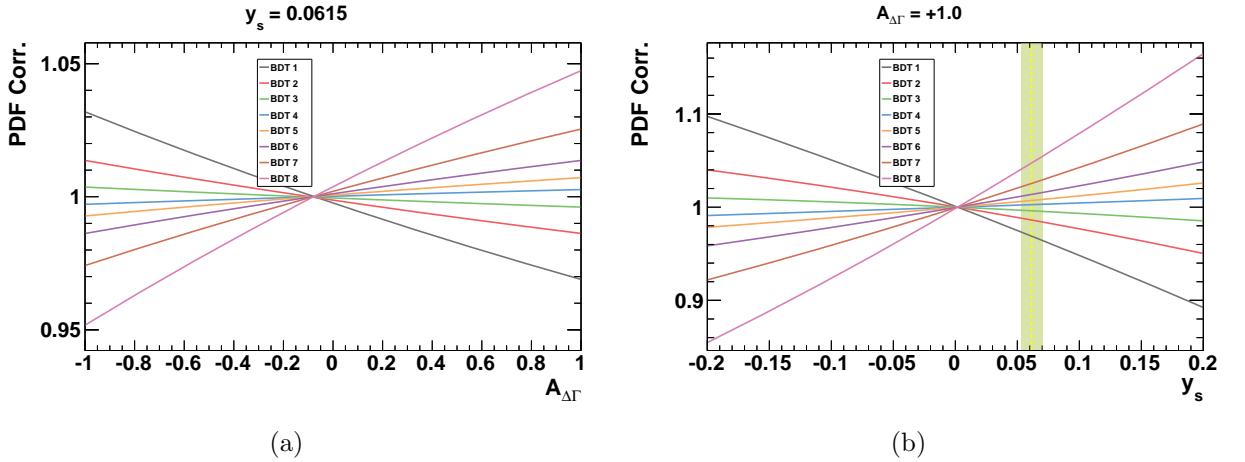


Figure 67: (a) Correction to be applied to BDT PDF as a function of $A_{\Delta\Gamma}$ for the experimental average y_s value [3] and (b) as a function of y_s for the SM $A_{\Delta\Gamma}$ value (+1). The vertical green band is the experimental average y_s value and its uncertainty.

Table 51: Corrections to the BDT PDF in each bin in the SM ($\mathcal{A}_{\Delta\Gamma} = 1$ and $y_s = 0.0615 \pm 0.0085$) and contribution from each bin to the total sensitivity [57].

Bin	PDF Correction $\delta_{PDF}^i - 1$ (%)	Sensitivity Contribution %
1	-3.1061 ± 0.0196	0.0981
2	-1.3778 ± 0.0290	1.09
3	-0.3887 ± 0.0392	2.03
4	$+0.2701 \pm 0.0423$	4.33
5	$+0.7193 \pm 0.0447$	8.70
6	$+1.3650 \pm 0.0457$	16.1
7	$+2.5423 \pm 0.0463$	26.9
8	$+4.7365 \pm 0.0433$	40.7
7+8	$+3.6464 \pm 0.0343$	67.6

Table 52: Fit results with and without the BDT PDF correction.

	$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ [10 ⁻⁹]	Bias	
	Nominal PDF	Corrected PDF	%
2011	$1.35^{+1.67}_{-1.19}$	$1.31^{+1.62}_{-1.15}$	-2.8
2012	$5.05^{+2.44}_{-1.95}$	$4.91^{+2.39}_{-1.91}$	-3.4
2011-12	$3.18^{+1.44}_{-1.19}$	$3.07^{+1.41}_{-1.15}$	-3.3

of future analyses, by comparing the gain these additional events bring to the statistical uncertainties to the additional systematics they add.

The correction to be applied for the SM are reported for each bin in Tab. 51. The effect of these corrections on the previous final results [56] cannot be derived directly has each bin has a different sensitivity. Hence Tab. 51 reports also the contribution each bin brings¹⁷ to the total analysis sensitivity [57]. Weighting the corrections by these numbers, the shift of the final branching ratio results of the previous analysis due to the PDF can be estimated at 3.5%.

A more correct evaluation of this shift was performed by comparing the fit results with and without the corrections which are reported in Tab. 52. With this method the global shift due to the PDF is estimated at 3.3% which is consistent with the previous estimate.

For $B^0 \rightarrow \mu^+ \mu^-$ no correction has to be applied since the control channel is also a B^0 decay and has therefore the same decay rate as the signal one.

¹⁷In the previous analysis bin 7 and 8 were merged.

11 Expected sensitivity with 3 fb^{-1}

The compatibility of the distribution of the observed events with that expected for a given branching fraction hypothesis is computed using the CL_s method [25]. This provides CL_{s+b} , a measurement of the compatibility of the observed distribution with the signal plus background hypothesis, and CL_b , a measurement of the compatibility with the background-only hypothesis. As input to this computation we use the expected number of combinatorial background events, peaking background events, and signal fractions according to the BDT fractions together with the number of observed events for each of the 72 bins, 8 bins in BDT and 9 bins in invariant mass (defined in appendix C, Tables 58 and 59 for B_s^0 and B^0 respectively).

The expected sensitivity for observing a $B_s^0 \rightarrow \mu^+ \mu^-$ signal is measured by the p-value $1 - \text{CL}_b$ ¹⁸ evaluated by generating the number of observed events through toys, according to background expectation and to signal SM time-integrated \mathcal{B} , $(3.56 \pm 0.29) \times 10^{-9}$. More precisely, $1 - \text{CL}_b$ is the integral of the tail of the background only $-2\ln Q$ distribution from the left up to the value of the observed $-2\ln Q$.

The systematic uncertainties on the signal predictions in each bin are computed by fluctuating the mass parameters (see Sec. 7.3), the BDT fractional yields (Sec. 7.2) and the normalization factors (Sec. 8) within Gaussian distributions defined by their associated uncertainties. The systematic uncertainty on the estimated number of combinatorial background events is computed by fluctuating within a Poissonian distribution the number of events measured in the sidebands, and by varying within 1σ the value of the exponent.

For each pseudo experiment a value of CL_b is extracted, from which the significance is computed. The distribution of the expected sensitivity obtained with 14M toys for the analysis of the 3 fb^{-1} sample and based on BDT12 is shown in Fig. 68. For sensitivities higher than 5 discretization problems start to be visible, since the pseudo experiments are sampling in this case the very last portion of the background-only $2\ln Q$ distribution. In particular, the peak at 6 contains all the SM pseudo experiments for which the $-2\ln Q$ falls outside the background-only distribution.

The expected sensitivity figure is defined as the median value of the distribution, and is equal to $4.9996^{+0.0835}_{-0.0587}$. The quoted errors are due to the finite toys statistics, which corresponds to 8 toys in the tail for a sensitivity of 5. The above number has to be intended as giving a 50% probability for the observed sensitivity of a SM signal to be greater equal than the quoted value.¹⁹.

The above result has to be compared with the expected sensitivity in november analysis based on BTD9, 3.7σ . Accounting for the increase of integrated luminosity, the above number can be extrapolated to $\sim 4.4\sigma$. Several factors affecting the sensitivity changed since the previous analysis version:

- The reconstruction, which obviously has an impact on the dimuon background: by looking at the background estimated in the 3 most significant bins (tab. 50),

¹⁸The significance is obtained as `RooStats::PValueToSignificance(0.5*(1-CLb))`

¹⁹From the same toy study, it can also be evaluated a 83.7% probability of a sensitivity above 4

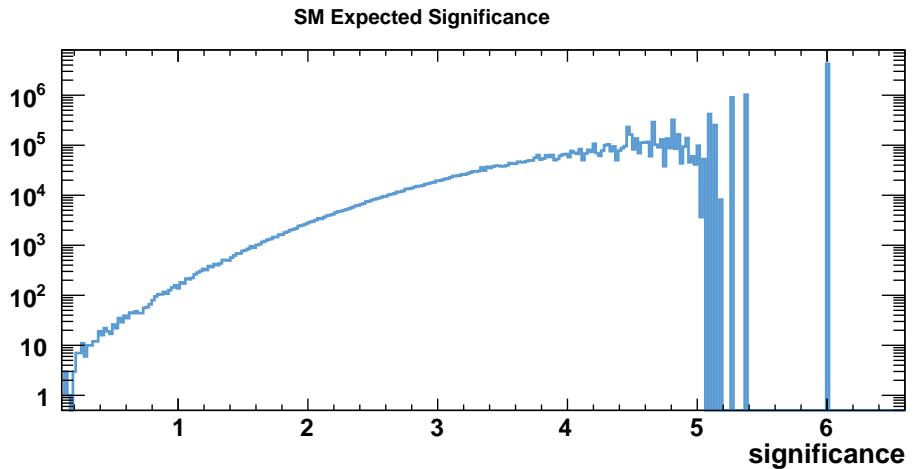


Figure 68: Expected sensitivity assuming a SM \mathcal{B} for 2011+2012 data.

we observe that a lower background is expected in s20r1 sample with respect to s17, while the opposite seems occurring for the s20 sample with respect to s19. In the latter case, we have also to take into account the factor 1.8 in the integrated luminosity; looking at table 50 and summing up the background expectations of the last 3 BDT9 bins for s20r1 and s20 and for s17 and s19 (scaled to 2 fb^{-1}), we get ~ 5.8 for the new reconstruction and ~ 6.4 for the old one. Taking into account the fact that one third of the sample is fully uncorrelated, then the two numbers are in agreement.

- 2011 and 2012 samples are merged in the present analysis, while they were treated separately in november: this was the best choice in the past version of the analysis, since the signal over background ratio was rather different for the two samples; in the present reconstruction, we observe similar background rates for 2011 and 2012 data (see tab. 50), so that no degradation of the sensitivity is expected when merging the two samples.
- The binning scheme: in november analysis we used 8 BDT bins for s17 data, and 7 BDT bins for s19 data, while we're using 8 bins now; the use of 7 bins for s19 was dictated by the absence of events for $\text{BDT} > 0.9$ in the mass sidebands, thus making impossible the background interpolation. In general, the optimal binning depends on a trade-off between the precision obtained on the background estimate and the signal over background ratio in each bin. For the present data reconstruction we could not complete a full toy study to validate the choice of using the 8 bins configuration which was optimized for the 1 fb^{-1} analysis of 2011 data. We however cross-checked that the separation of the background and background plus signal test statistics is larger for the 8 bins than the 7 bins scheme, thus the 8 bins is expected to perform better than the 7 bins.
- The exclusive background yields in the mass sidebands, which are anticorrelated with

respect to the combinatorial background estimated in the signal mass window: with respect to november analysis, a $\sim 15\%$ decrease of the $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$ and a $\sim 25\%$ decrease of the $B \rightarrow h^+ h'^-$ was estimated due to systematic effects discovered in the evaluation of the pion and kaon muon misidentification probabilities. The effect on the combinatorial background has been quantified into a $\sim 15\%$ background increase in the three most significant bins, with a consequent small degradation of the sensitivity.

- The introduction of the time dependent corrections, which reflect into a modification of the signal BDT pdf: as a result, the expected signal yield is enhanced by $\sim 4\%$ in the two most significant BDT bins, which account for $\sim 70\%$ of the total sensitivity.
- The value of the SM branching ratio used to generate the signal expectations: we passed from $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (3.54 \pm 0.29) \times 10^{-9}$ to $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (3.56 \pm 0.29) \times 10^{-9}$, which corresponds to a tiny increase.
- The change in BDT operator: we do expect the choice of BDT12 instead of BDT9 to improve the analysis performances, since a decrease of the background is observed in the mass sidebands (see Fig.34 and tab 50).

In conclusion, we do expect from the present dataset and the present analysis performance an improvement with respect to november result, and the main source of improvement seems the change in BDT algorithm.

12 Results

The unblinded distribution of events in the invariant mass–BDT plane is shown in Fig. 69 for the full dataset.

The number of observed events and the SM signal and background expectations are listed for all of the 8 BDT bins and integrating in the whole mass signal window in Tables 53 and 54, for $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ respectively (the yields per mass bin are given instead in Appendix C).

The comparison of the distributions of the observed events and the expected background events results in a p-value ($1 - CL_b$) of 5.6 % for the $B^0 \rightarrow \mu^+ \mu^-$ decay, computed at the branching fraction value corresponding to $CL_{s+b} = 0.5$. The number of $B^0 \rightarrow \mu^+ \mu^-$ candidates is compatible with background expectation and therefore an upper limit on its branching fraction is set with the CL_s method. The observed and expected CL_s values are shown in Fig. 70 for the $B^0 \rightarrow \mu^+ \mu^-$ channel, as a function of the assumed branching fraction. The corresponding observed and expected limits for $B^0 \rightarrow \mu^+ \mu^-$ at 90 % and 95 % C.L. are shown in Table 55. The expected limits are computed allowing the presence of $B^0 \rightarrow \mu^+ \mu^-$ events according to the SM branching fraction, and including cross-feed from $B_s^0 \rightarrow \mu^+ \mu^-$, and in background only hypothesis.

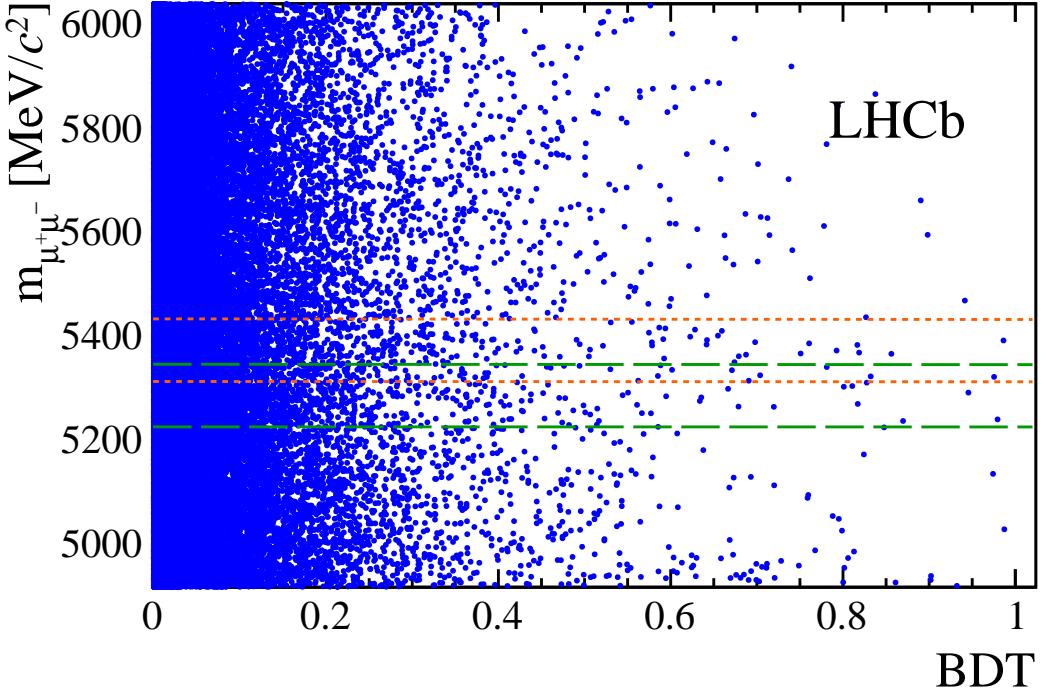


Figure 69: Dimuon mass versus BDT for selected candidates; orange short-dashed (green long-dashed) lines indicate the $\pm 60 \text{ MeV}/c^2$ search window around the B_s^0 (B^0) mass.

Table 53: Expected combinatorial background, $B \rightarrow h^+h'^-$ background, and signal events assuming SM predictions, together with the number of observed events in $B_s^0 \rightarrow \mu^+\mu^-$ mass region, in bins of BDT for the full 3 fb^{-1} sample.

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
Exp. comb. bkg	6138^{+114}_{-112}	$121.6^{+4.8}_{-4.7}$	$28.2^{+2.2}_{-2.1}$	$11.9^{+1.5}_{-1.4}$	$4.77^{+1.11}_{-0.95}$	$2.17^{+0.79}_{-0.65}$	$0.79^{+0.48}_{-0.34}$	$0.29^{+0.32}_{-0.18}$
5311 – 5431	Exp. peak. bkg	$0.330^{+0.105}_{-0.089}$	$0.210^{+0.068}_{-0.058}$	$0.140^{+0.045}_{-0.038}$	$0.148^{+0.048}_{-0.040}$	$0.147^{+0.047}_{-0.040}$	$0.140^{+0.045}_{-0.038}$	$0.130^{+0.042}_{-0.035}$
	Exp. signal	$8.78^{+1.09}_{-0.99}$	$5.40^{+0.75}_{-0.67}$	$3.52^{+0.46}_{-0.41}$	$3.75^{+0.47}_{-0.43}$	$3.76^{+0.47}_{-0.43}$	$3.61^{+0.46}_{-0.42}$	$3.68^{+0.46}_{-0.42}$
	Observed	5885	135	18	16	13	5	4
								2

For the $B_s^0 \rightarrow \mu^+\mu^-$, the comparison between observed data and the expected background results into a p-value ($1 - \text{CL}_b$) of 1.4×10^{-4} , corresponding to a significance of $3.803^{+0.029}_{-0.026}$, where the errors reflect the finite number of toys.

A simultaneous unbinned maximum-likelihood fit to the mass projections in the 8 BDT bins has been performed on the full mass range to extract the $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$ branching fractions. The background PDFs include an exponential for the combinatorial,

Table 54: Expected combinatorial background, $B \rightarrow h^+h^-$ background, cross-feed, and signal events assuming SM predictions, together with the number of observed events in $B^0 \rightarrow \mu^+\mu^-$ mass region, in bins of BDT for the full 3 fb^{-1} sample.

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 – 5344	Exp. comb. bkg	6520^{+119}_{-117}	$127.0^{+5.2}_{-5.0}$	$29.4^{+2.4}_{-2.3}$	$12.8^{+1.7}_{-1.5}$	$4.9^{+1.2}_{-1.1}$	$2.14^{+0.88}_{-0.70}$	$0.82^{+0.53}_{-0.37}$	$0.29^{+0.35}_{-0.19}$
	Exp. peak. bkg	$1.97^{+0.64}_{-0.47}$	$1.25^{+0.41}_{-0.31}$	$0.83^{+0.27}_{-0.20}$	$0.88^{+0.29}_{-0.21}$	$0.88^{+0.28}_{-0.21}$	$0.83^{+0.27}_{-0.20}$	$0.77^{+0.25}_{-0.18}$	$0.66^{+0.21}_{-0.16}$
	Exp. Cross-feed	$1.38^{+0.18}_{-0.16}$	$0.85^{+0.12}_{-0.11}$	$0.554^{+0.075}_{-0.067}$	$0.590^{+0.078}_{-0.070}$	$0.591^{+0.076}_{-0.070}$	$0.567^{+0.077}_{-0.069}$	$0.579^{+0.076}_{-0.069}$	$0.595^{+0.077}_{-0.069}$
	Exp. signal	$0.99^{+0.12}_{-0.11}$	$0.610^{+0.081}_{-0.075}$	$0.398^{+0.049}_{-0.046}$	$0.424^{+0.050}_{-0.047}$	$0.425^{+0.050}_{-0.047}$	$0.408^{+0.050}_{-0.047}$	$0.416^{+0.049}_{-0.046}$	$0.428^{+0.050}_{-0.046}$
Observed		6280	127	30	16	7	3	6	3

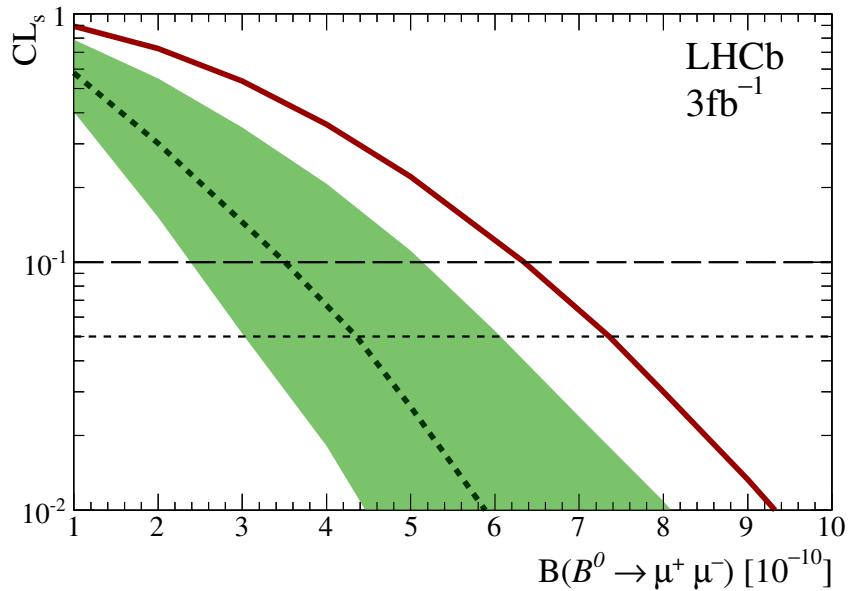


Figure 70: CL_s as a function of the assumed $B^0 \rightarrow \mu^+\mu^-$ branching fraction. The dashed curve is the median of the expected CL_s distribution for background-only hypothesis. The shaded yellow area covers, for each branching fraction value, 34 % of the expected CL_s distribution on each side of its median. The solid red curve is the observed CL_s .

and the PDFs discussed in Sec. 9 for $B_{(s)}^0 \rightarrow h^+h^-$, $B^0 \rightarrow \pi^-\mu^+\nu_\mu$, $B_s^0 \rightarrow K^-\mu^+\nu_\mu$ and $B^{0(+)} \rightarrow \pi^{0(+)}\mu^+\mu^-$. The fit free parameters are the $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$ branching fractions, and the combinatorial background slope and normalizations; the nuisance parameters are instead the normalization factors for the signal and the exclusive backgrounds, the mass and BDT PDFs parameters for the signal, and the BDT PDFs parameters for the exclusive backgrounds, which are left to vary in the fit according to their uncertainties. The systematic error induced by the mass PDFs of the exclusive

Table 55: Observed limits on the $B^0 \rightarrow \mu^+ \mu^-$ branching fraction and corresponding expected limits for the background only and background plus SM signal hypotheses.

Limit	at 90 % C.L.	at 95 % C.L.
Exp. bkg+SM	4.5×10^{-10}	5.4×10^{-10}
Exp. bkg	3.5×10^{-10}	4.4×10^{-10}
Observed	6.3×10^{-10}	7.4×10^{-10}

backgrounds has been evaluated from the shift in the result when repeating the fit with parameters varied according to their uncertainty. The fit mass projections for the 8 BDT bins separately are shown in Fig. 71, while the fit mass projection for BDT above 0.7 is shown in Fig. 72.

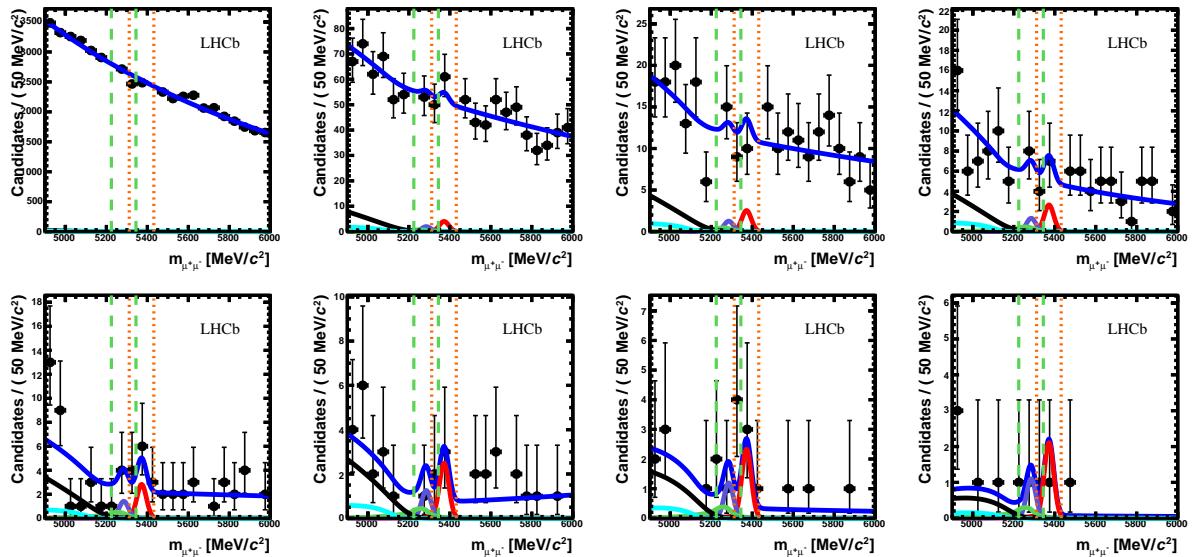


Figure 71: Invariant mass distribution of the selected signal candidates (black dots) for the 8 BDT bins separately. The result of the fit is overlaid (blue solid line) and the different components detailed: $B_s^0 \rightarrow \mu^+ \mu^-$ (red), $B^0 \rightarrow \mu^+ \mu^-$ (purple), $B_{(s)}^0 \rightarrow h^+ h^-$ (green), $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$ and $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ (black), $B^{0(+)} \rightarrow \pi^{0(+)} \mu^+ \mu^-$ (cyan).

The fit gives:

$$\begin{aligned}\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) &= (2.87^{+1.11}_{-0.95}(\text{likelihood})^{+0.05}_{-0.06}(\text{syst})) \times 10^{-9}, \\ \mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) &= (3.74^{+2.42}_{-2.05}(\text{likelihood})^{+0.34}_{-0.38}(\text{syst})) \times 10^{-10},\end{aligned}$$

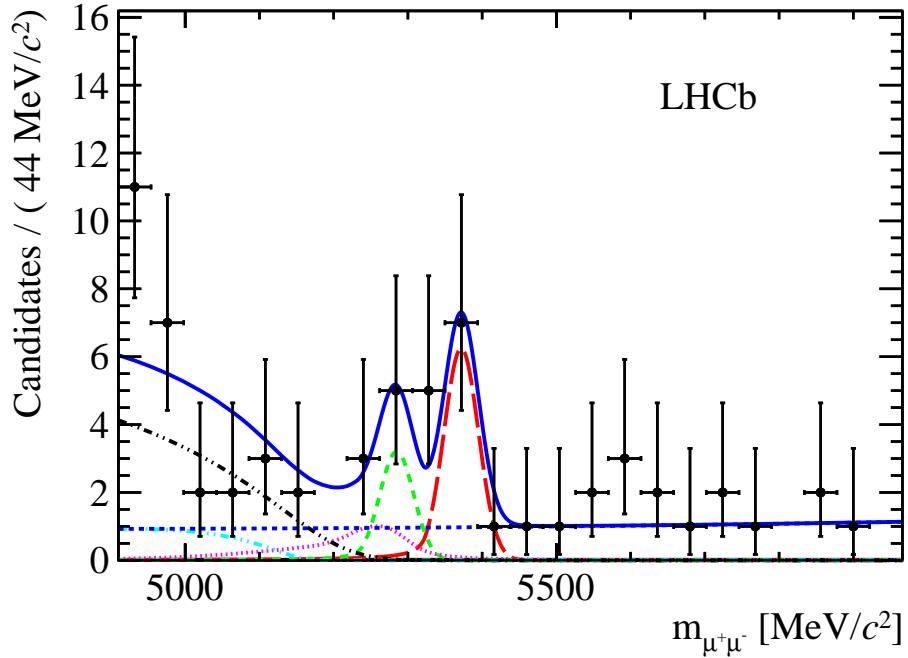


Figure 72: Invariant mass distribution of the selected signal candidates (black dots) with $\text{BDT} > 0.7$. The result of the fit is overlaid (blue solid line) and the different components detailed: $B_s^0 \rightarrow \mu^+ \mu^-$ (red long dashed), $B^0 \rightarrow \mu^+ \mu^-$ (green medium dashed), $B_{(s)}^0 \rightarrow h^+ h^-$ (magenta dotted), $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$ and $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ (black dot-dashed), $B^{0(+)} \rightarrow \pi^{0(+)} \mu^+ \mu^-$ (light blue dot dashed), and the combinatorial background (blue medium dashed).

where the central value are extracted from the maximum of the logarithm of the profile likelihood, the first error reflects the interval corresponding to a change of 0.5 with respect to the maximum of the log-likelihood (see likelihood profile in Fig.73), and the second error reflects the impact on the result by repeating the fit with inclusion of a $\Lambda_b^0 \rightarrow p \mu^- \nu$ among the considered exclusive backgrounds, and from the variation of the exclusive background mass PDFs. For the sake of publication the above results have been rewritten as

$$\begin{aligned}\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) &= (2.9_{-1.0}^{+1.1}(\text{stat})_{-0.1}^{+0.3}(\text{syst})) \times 10^{-9} \\ \mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) &= (3.7_{-2.1}^{+2.4}(\text{stat})_{-0.4}^{+0.6}(\text{syst})) \times 10^{-10}\end{aligned}$$

where the statistical uncertainty reflects the interval corresponding to a change of 0.5 in the log-likelihood after fixing all the nuisance parameters to their expected values, and the systematic uncertainty is obtained by subtracting in quadrature the statistical uncertainty from the likelihood uncertainty of the previous equation, with the additional systematic uncertainty from the exclusive backgrounds added in quadrature.

The value of $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ is in agreement with the time-integrated SM prediction,

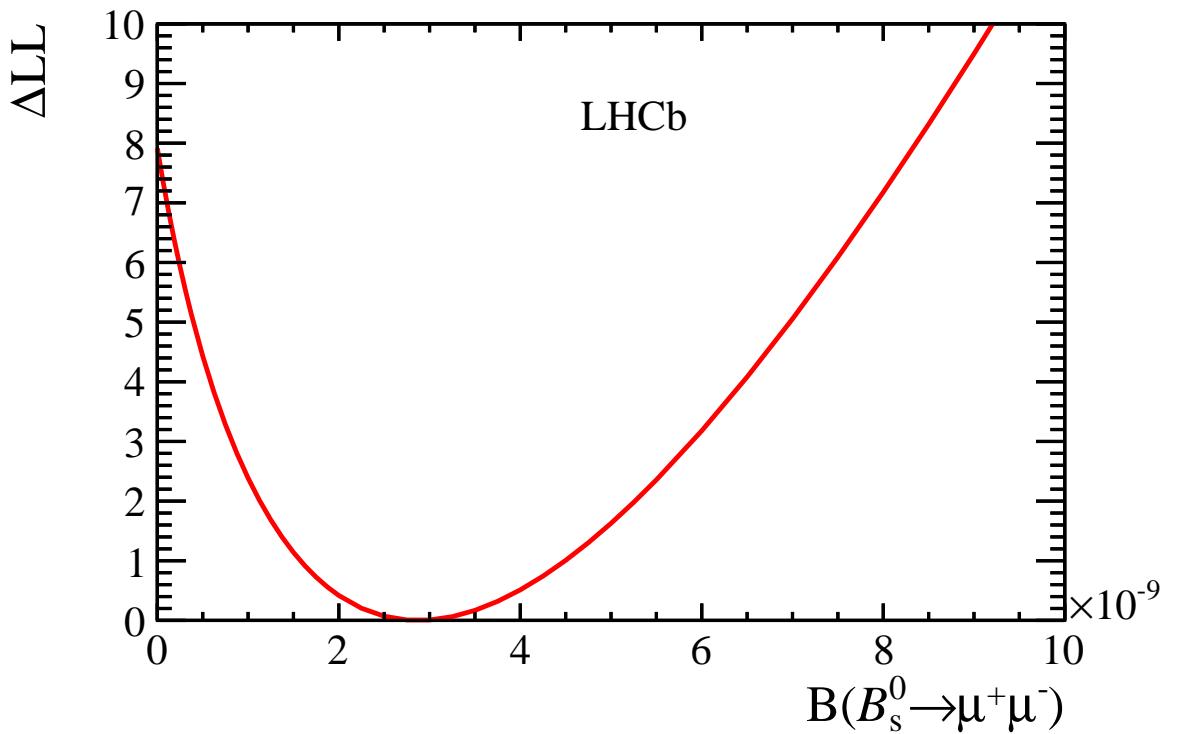


Figure 73: BR fit likelihood profile scan as a function of $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-)$.

$\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-) = (3.56 \pm 0.30) \times 10^{-9}$, Eq. 3. The signal significance obtained from the difference in log-likelihood between signal and null hypotheses corresponds to a 4.0σ evidence of $B_s^0 \rightarrow \mu^+\mu^-$.

Appendix

A Trigger line definitions

Table 56: 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 57: 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'

B Definition of isolations

Definitions copied from previous analysis notes.

B.1 Track isolation

For each of the muon candidates, a search is performed for long tracks (traversing all tracking detectors, with an impact parameter with respect to any primary vertex greater than 3 in units of significance, and of course excluding the other muon candidate), that

can make a “good” vertex with the muon candidate. This requires the muon candidate and the given track to: make an angle < 0.27 rad, have a DOCA $< 130 \mu\text{m}$, and make a vertex whose distance PVdis (SVdis) to the primary (secondary) vertex satisfies: $4 \text{ cm} > \text{PVdis} > 0.5 \text{ cm}$ ($30 \text{ cm} > \text{SVdis} > -0.15 \text{ cm}$). The variables PVdis and SVdis are signed according to $z_{vtx}(mu, tr) - z_{PV}$ (or SV). Furthermore, if we define $\alpha^{\mu+tr, PV}$ as the angle between the sum of the momenta of the muon and the extra track and the direction defined by the PV and the vertex reconstructed using the muon and the extra track candidates, then the sum of the momenta is required to satisfy

$$\frac{\left| \vec{P}_\mu + \vec{P}_{tr} \right| \cdot \sin(\alpha^{\mu+tr, PV})}{\left| \vec{P}_\mu + \vec{P}_{tr} \right| \cdot \sin(\alpha^{\mu+tr, PV}) + P_{T\mu} + P_{Ttr}} < 0.6 \quad (72)$$

where $P_{T\mu}$ and P_{Ttr} are the transverse momentum (with respect to the beam line) of the muon candidate and the extra track. The number of tracks that satisfy these conditions is used as a discriminating variable for each of the muon candidates.

B.2 B isolation or “CDF isolation”

The B isolation or “CDF isolation” (CDF definition [58]) I_{CDF} is defined as follows:

$$I_{\text{CDF}} = \frac{p_T(B)}{p_T(B) + \sum_{\text{tracks}} p_T(\text{tracks})} \quad (73)$$

where $p_T(B)$ is the B transverse momentum and the tracks used in the summation are those, excluding the candidates muons, for which $\sqrt{\delta\eta^2 + \delta\phi^2} < 1.0$, with $\delta\eta$ and $\delta\phi$ denoting respectively the difference in the pseudorapidity and of the ϕ coordinate between the track and the B candidate.

C Event yields in $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$ mass regions

References

- [1] LHCb Collaboration, R. Aaij *et al.*, *First evidence for the decay $B_s \rightarrow \mu^+\mu^-$* , Phys. Rev. Lett. **110** (2013) 021801, arXiv:1211.2674.
- [2] A. J. Buras, J. Giriach, D. Guadagnoli, and G. Isidori, *On the standard model prediction for $\mathcal{B}(B_{s,d}^0 \rightarrow \mu^+\mu^-)$* , Eur. Phys. J. **C72** (2012) 2172, arXiv:1208:0934.
- [3] HFAG. online update http://www.slac.stanford.edu/xorg/hfag/osc/PDG_2013/.
- [4] K. Babu and C. F. Kolda, *Higgs mediated $B^0 \rightarrow \mu^+\mu^-$ in minimal supersymmetry*, Phys. Rev. Lett. **84** (2000) 228, arXiv:hep-ph/9909476.

- [5] C. Hamzaoui, M. Pospelov, and M. Toharia, *Higgs mediated FCNC in supersymmetric models with large tan Beta*, Phys. Rev. **D59** (1999) 095005, arXiv:hep-ph/9807350.

Table 58: Expected combinatorial background, $B \rightarrow h^+h^-$ background, and signal events assuming SM predictions, together with the number of observed events in $B_s^0 \rightarrow \mu^+\mu^-$ mass region, in bins of mass and BDT for the full 3 fb^{-1} sample.

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	
5311 – 5323	Exp. comb. bkg	637^{+12}_{-12}	$12.49^{+0.50}_{-0.49}$	$2.89^{+0.23}_{-0.22}$	$1.25^{+0.16}_{-0.14}$	$0.48^{+0.12}_{-0.10}$	$0.215^{+0.083}_{-0.068}$	$0.081^{+0.051}_{-0.036}$	$0.029^{+0.033}_{-0.019}$
	Exp. peak. bkg	$0.087^{+0.030}_{-0.026}$	$0.056^{+0.019}_{-0.017}$	$0.037^{+0.013}_{-0.011}$	$0.039^{+0.014}_{-0.012}$	$0.039^{+0.013}_{-0.012}$	$0.037^{+0.013}_{-0.011}$	$0.034^{+0.012}_{-0.010}$	$0.0294^{+0.0101}_{-0.0088}$
	Exp. signal	$0.147^{+0.020}_{-0.018}$	$0.091^{+0.013}_{-0.012}$	$0.0590^{+0.0084}_{-0.0074}$	$0.0629^{+0.0087}_{-0.0077}$	$0.0629^{+0.0086}_{-0.0077}$	$0.0605^{+0.0085}_{-0.0075}$	$0.0617^{+0.0084}_{-0.0075}$	$0.0635^{+0.0085}_{-0.0077}$
	Observed	594	16	1	2	1	0	1	1
5323 – 5335	Exp. comb. bkg	632^{+12}_{-12}	$12.41^{+0.50}_{-0.48}$	$2.87^{+0.23}_{-0.22}$	$1.23^{+0.16}_{-0.14}$	$0.483^{+0.116}_{-0.099}$	$0.216^{+0.082}_{-0.067}$	$0.080^{+0.050}_{-0.035}$	$0.029^{+0.032}_{-0.018}$
	Exp. peak. bkg	$0.060^{+0.018}_{-0.015}$	$0.0385^{+0.0118}_{-0.0095}$	$0.0256^{+0.0077}_{-0.0063}$	$0.0270^{+0.0082}_{-0.0066}$	$0.0269^{+0.0081}_{-0.0065}$	$0.0256^{+0.0077}_{-0.0062}$	$0.0237^{+0.0072}_{-0.0057}$	$0.0203^{+0.0061}_{-0.0049}$
	Exp. signal	$0.364^{+0.049}_{-0.044}$	$0.224^{+0.033}_{-0.030}$	$0.146^{+0.020}_{-0.018}$	$0.156^{+0.021}_{-0.019}$	$0.156^{+0.021}_{-0.019}$	$0.150^{+0.021}_{-0.019}$	$0.153^{+0.021}_{-0.019}$	$0.157^{+0.021}_{-0.019}$
	Observed	597	8	0	0	2	1	0	0
5335 – 5341	Exp. comb. bkg	$313.9^{+5.8}_{-5.7}$	$6.18^{+0.25}_{-0.24}$	$1.43^{+0.11}_{-0.11}$	$0.612^{+0.077}_{-0.070}$	$0.241^{+0.057}_{-0.049}$	$0.108^{+0.041}_{-0.033}$	$0.040^{+0.025}_{-0.017}$	$0.0145^{+0.0161}_{-0.0092}$
	Exp. peak. bkg	$0.0234^{+0.0061}_{-0.0049}$	$0.0149^{+0.0040}_{-0.0032}$	$0.0099^{+0.0026}_{-0.0021}$	$0.0105^{+0.0028}_{-0.0022}$	$0.0104^{+0.0027}_{-0.0022}$	$0.0099^{+0.0026}_{-0.0021}$	$0.0092^{+0.0024}_{-0.0019}$	$0.0079^{+0.0020}_{-0.0016}$
	Exp. signal	$0.333^{+0.042}_{-0.039}$	$0.205^{+0.029}_{-0.026}$	$0.134^{+0.018}_{-0.016}$	$0.142^{+0.018}_{-0.017}$	$0.143^{+0.018}_{-0.017}$	$0.137^{+0.018}_{-0.016}$	$0.140^{+0.018}_{-0.016}$	$0.144^{+0.018}_{-0.016}$
	Observed	314	4	2	1	0	1	0	0
5341 – 5355	Exp. comb. bkg	624^{+12}_{-11}	$12.31^{+0.49}_{-0.47}$	$2.85^{+0.22}_{-0.21}$	$1.21^{+0.15}_{-0.14}$	$0.480^{+0.114}_{-0.097}$	$0.216^{+0.081}_{-0.066}$	$0.080^{+0.049}_{-0.035}$	$0.029^{+0.032}_{-0.018}$
	Exp. peak. bkg	$0.0381^{+0.0088}_{-0.0076}$	$0.0243^{+0.0058}_{-0.0050}$	$0.0162^{+0.0038}_{-0.0033}$	$0.0171^{+0.0040}_{-0.0034}$	$0.0170^{+0.0039}_{-0.0034}$	$0.0162^{+0.0038}_{-0.0032}$	$0.0150^{+0.0035}_{-0.0030}$	$0.0128^{+0.0030}_{-0.0025}$
	Exp. signal	$1.07^{+0.13}_{-0.12}$	$0.658^{+0.091}_{-0.082}$	$0.429^{+0.056}_{-0.051}$	$0.457^{+0.058}_{-0.052}$	$0.458^{+0.057}_{-0.052}$	$0.440^{+0.057}_{-0.052}$	$0.449^{+0.056}_{-0.051}$	$0.461^{+0.056}_{-0.052}$
	Observed	610	14	3	1	1	0	0	0
5353 – 5389	Exp. comb. bkg	1841^{+34}_{-34}	$36.5^{+1.4}_{-1.4}$	$8.45^{+0.66}_{-0.63}$	$3.57^{+0.45}_{-0.41}$	$1.43^{+0.33}_{-0.28}$	$0.65^{+0.24}_{-0.19}$	$0.24^{+0.15}_{-0.10}$	$0.086^{+0.095}_{-0.055}$
	Exp. peak. bkg	$0.073^{+0.023}_{-0.020}$	$0.046^{+0.015}_{-0.013}$	$0.0308^{+0.0097}_{-0.0084}$	$0.0326^{+0.0103}_{-0.0088}$	$0.0324^{+0.0102}_{-0.0087}$	$0.0308^{+0.0098}_{-0.0083}$	$0.0286^{+0.0090}_{-0.0077}$	$0.0244^{+0.0077}_{-0.0066}$
	Exp. signal	$4.97^{+0.62}_{-0.56}$	$3.06^{+0.42}_{-0.38}$	$1.99^{+0.26}_{-0.24}$	$2.12^{+0.27}_{-0.24}$	$2.13^{+0.27}_{-0.24}$	$2.04^{+0.26}_{-0.24}$	$2.08^{+0.26}_{-0.24}$	$2.14^{+0.26}_{-0.24}$
	Observed	1773	44	8	6	4	3	3	0
5389 – 5401	Exp. comb. bkg	603^{+11}_{-11}	$12.02^{+0.47}_{-0.46}$	$2.79^{+0.22}_{-0.21}$	$1.16^{+0.15}_{-0.13}$	$0.474^{+0.109}_{-0.093}$	$0.218^{+0.078}_{-0.063}$	$0.079^{+0.048}_{-0.034}$	$0.029^{+0.031}_{-0.018}$
	Exp. peak. bkg	$0.0163^{+0.0063}_{-0.0054}$	$0.0104^{+0.0041}_{-0.0035}$	$0.0069^{+0.0027}_{-0.0023}$	$0.0073^{+0.0029}_{-0.0024}$	$0.0073^{+0.0028}_{-0.0024}$	$0.0069^{+0.0027}_{-0.0023}$	$0.0064^{+0.0025}_{-0.0021}$	$0.0055^{+0.0021}_{-0.0018}$
	Exp. signal	$1.07^{+0.13}_{-0.12}$	$0.658^{+0.091}_{-0.082}$	$0.429^{+0.056}_{-0.051}$	$0.457^{+0.058}_{-0.052}$	$0.458^{+0.057}_{-0.052}$	$0.440^{+0.057}_{-0.052}$	$0.449^{+0.056}_{-0.051}$	$0.461^{+0.056}_{-0.052}$
	Observed	591	21	2	1	3	0	0	1
5401 – 5407	Exp. comb. bkg	$299.8^{+5.6}_{-5.5}$	$5.98^{+0.23}_{-0.23}$	$1.39^{+0.11}_{-0.10}$	$0.578^{+0.073}_{-0.067}$	$0.236^{+0.054}_{-0.046}$	$0.109^{+0.039}_{-0.032}$	$0.039^{+0.024}_{-0.017}$	$0.0143^{+0.0156}_{-0.0090}$
	Exp. peak. bkg	$0.0070^{+0.0029}_{-0.0024}$	$0.0044^{+0.0019}_{-0.0015}$	$0.0030^{+0.0012}_{-0.0010}$	$0.0031^{+0.0013}_{-0.0011}$	$0.0031^{+0.0013}_{-0.0011}$	$0.0030^{+0.0012}_{-0.0010}$	$0.00274^{+0.00116}_{-0.00093}$	$0.00234^{+0.00099}_{-0.00080}$
	Exp. signal	$0.333^{+0.042}_{-0.039}$	$0.205^{+0.029}_{-0.026}$	$0.134^{+0.018}_{-0.016}$	$0.142^{+0.018}_{-0.017}$	$0.143^{+0.018}_{-0.017}$	$0.137^{+0.018}_{-0.016}$	$0.140^{+0.018}_{-0.016}$	$0.144^{+0.018}_{-0.016}$
	Observed	295	12	0	1	0	0	0	0
5407 – 5419	Exp. comb. bkg	596^{+11}_{-11}	$11.92^{+0.47}_{-0.46}$	$2.76^{+0.22}_{-0.21}$	$1.15^{+0.15}_{-0.13}$	$0.472^{+0.109}_{-0.093}$	$0.218^{+0.078}_{-0.063}$	$0.078^{+0.047}_{-0.033}$	$0.029^{+0.031}_{-0.018}$
	Exp. peak. bkg	$0.0131^{+0.0057}_{-0.0045}$	$0.0083^{+0.0037}_{-0.0029}$	$0.0055^{+0.0024}_{-0.0019}$	$0.0059^{+0.0026}_{-0.0020}$	$0.0058^{+0.0026}_{-0.0020}$	$0.0056^{+0.0024}_{-0.0019}$	$0.0051^{+0.0023}_{-0.0018}$	$0.0044^{+0.0019}_{-0.0015}$
	Exp. signal	$0.364^{+0.049}_{-0.044}$	$0.224^{+0.033}_{-0.030}$	$0.146^{+0.020}_{-0.018}$	$0.156^{+0.021}_{-0.019}$	$0.156^{+0.021}_{-0.019}$	$0.150^{+0.021}_{-0.019}$	$0.153^{+0.021}_{-0.019}$	$0.157^{+0.021}_{-0.019}$
	Observed	578	8	2	2	2	0	0	0
5419 – 5431	Exp. comb. bkg	591^{+11}_{-11}	$11.85^{+0.47}_{-0.45}$	$2.75^{+0.22}_{-0.21}$	$1.13^{+0.14}_{-0.13}$	$0.470^{+0.108}_{-0.092}$	$0.219^{+0.078}_{-0.063}$	$0.078^{+0.047}_{-0.033}$	$0.028^{+0.031}_{-0.018}$
	Exp. peak. bkg	$0.0116^{+0.0047}_{-0.0041}$	$0.0074^{+0.0031}_{-0.0026}$	$0.0049^{+0.0020}_{-0.0017}$	$0.0052^{+0.0022}_{-0.0018}$	$0.0052^{+0.0021}_{-0.0018}$	$0.0049^{+0.0020}_{-0.0017}$	$0.0046^{+0.0019}_{-0.0016}$	$0.0039^{+0.0016}_{-0.0014}$
	Exp. signal	$0.128^{+0.020}_{-0.018}$	$0.079^{+0.013}_{-0.012}$	$0.0514^{+0.0083}_{-0.0073}$	$0.0548^{+0.0086}_{-0.0076}$	$0.0548^{+0.0086}_{-0.0076}$	$0.0527^{+0.0085}_{-0.0074}$	$0.0538^{+0.0084}_{-0.0076}$	$0.0553^{+0.0085}_{-0.0076}$
	Observed	533	8	0	2	0	0	0	0

- [6] C.-S. Huang, W. Liao, Q.-S. Yan, and S.-H. Zhu, *$B_s \rightarrow \text{lepton} + \text{lepton}$ - in a general 2 HDM and MSSM*, Phys. Rev. **D63** (2001) 114021, [arXiv:hep-ph/0006250](#).
- [7] M. Blanke *et al.*, *Rare and CP-Violating K and B Decays in the Littlest Higgs Model with T-Parity*, JHEP **0701** (2007) 066, [arXiv:hep-ph/0610298](#).
- [8] M. Blanke *et al.*, *Rare K and B Decays in a Warped Extra Dimension with Custodial Protection*, JHEP **0903** (2009) 108, [arXiv:0812.3803](#).
- [9] W. Liu, C.-X. Yue, and H.-D. Yang, *Rare decays $B_s \rightarrow \ell^+ \ell^-$ and $B \rightarrow K \ell^+ \ell^-$ in the topcolor-assisted technicolor model*, Phys. Rev. **D79** (2009) 034008, [arXiv:0901.3463](#).
- [10] M. Bauer, S. Casagrande, U. Haisch, and M. Neubert, *Flavor Physics in the Randall-Sundrum Model: II. Tree-Level Weak-Interaction Processes*, JHEP **1009** (2010) 017, [arXiv:0912.1625](#).
- [11] A. J. Buras *et al.*, *Lepton Flavour Violation in the Presence of a Fourth Generation of Quarks and Leptons*, JHEP **1009** (2010) 104, [arXiv:1006.5356](#).
- [12] A. J. Buras and J. Girrbach, *BSM models facing the recent LHCb data: A First look*, [arXiv:1204.5064](#).
- [13] P. H. Chankowski and J. Rosiek, *Supersymmetry (at large tan beta) and flavor physics*, Acta Phys. Polon. **B33** (2002) 2329, [arXiv:hep-ph/0207242](#).
- [14] R. Arnowitt, B. Dutta, T. Kamon, and M. Tanaka, *Detection of $bs+$ at the tevatron runii and constraints on the susy parameter space*, Physics Letters B **538** (2002), no. 12 121 .
- [15] LHCb collaboration, R. Aaij *et al.*, *Strong constraints on the rare decays $B_s \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$* , Phys. Rev. Lett. **108**, **231801** (2012) , [arXiv:1203.4493](#).
- [16] K. de Bruyn *et al.*, *A New Window for New Physics in $B_s^0 \rightarrow \mu^+ \mu^-$* , [arXiv:1204.1737](#).
- [17] K. de Bruyn *et al.*, *On Branching Ratio Measurements of B_s Decays*, [arXiv:1204.1735](#).
- [18] C. Adrover *et al.*, *Search for the rare decays $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ with 1 fb^{-1} of 2012 data*, LHCb-ANA-2012-081.
- [19] M. Perrin-Terrin, G. Mancinelli, F. Dettori, and J. Serrano, *Towards model independent results for $B_s^0 \rightarrow \mu^+ \mu^-$* , LHCb-INT-2013-012.
- [20] LHCb collaboration, R. Aaij *et al.*, *Search for the rare decays $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$* , Phys. Lett. **B708** (2012) 55, [arXiv:1112.1600](#).

- [21] LHCb collaboration, R. Aaij *et al.*, *Measurement of $\sigma(pp \rightarrow b\bar{b}X)$ at $\sqrt{s} = 7$ TeV in the forward region*, Phys. Lett. **B694** (2010) 209, [arXiv:1009.2731](https://arxiv.org/abs/1009.2731).
- [22] C. Adrover *et al.*, *Search for the rare decays $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$ with 1.02 fb^{-1} at LHCb*, LHCb-ANA-2011-102.
- [23] A. Hoecker *et al.*, *TMVA: Toolkit for Multivariate Data Analysis*, PoS **ACAT** (2007) 040, [arXiv:physics/0703039](https://arxiv.org/abs/physics/0703039).
- [24] T. Skwarnicki, *A Study of the Radiative Cascade Transitions Between the Υ and Υ' resonances.*, PhD thesis, DESY F31-86-02, 1986. Appendix E.
- [25] A. L. Read, *Presentation of search results: the $cl\ s$ technique*, Journal of Physics G: Nuclear and Particle Physics **28** (2002), no. 10 2693.
- [26] P. Seyfert, <https://twiki.cern.ch/twiki/bin/view/LHCb/RecoXX>, .
- [27] J. Brehmer, J. Albrecht, and P. Seyfert, *Ghost probability: an efficient tool to remove background tracks*, Tech. Rep. LHCb-INT-2012-025. CERN-LHCb-INT-2012-025, CERN, Geneva, Sep, 2012.
- [28] LHCb trigger group, *LHCb trigger paper*, to be submitted soon.
- [29] R. Aaij and J. Albrecht, *Muon triggers in the High Level Trigger of LHCb*, Tech. Rep. LHCb-PUB-2011-017. CERN-LHCb-PUB-2011-017, CERN, Geneva, Sep, 2011.
- [30] V. Gligorov, C. Thomas, and M. Williams, *The HLT inclusive B triggers*, Tech. Rep. LHCb-PUB-2011-016. CERN-LHCb-PUB-2011-016. LHCb-INT-2011-030, CERN, Geneva, Sep, 2011. LHCb-INT-2011-030.
- [31] C. Adrover *et al.*, *Search for the rare decays $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$ with 370 pb^{-1} at LHCb*, LHCb-ANA-2011-078.
- [32] C. Adrover *et al.*, *Searches for $B_s \rightarrow \mu\mu$ and $B_d \rightarrow \mu\mu$ in 370 pb^{-1} at LHCb*, LHCb-ANA-2011-078.
- [33] F. Archilli *et al.*, *Performance of the muon identification at LHCb*, [arXiv:1306.0249](https://arxiv.org/abs/1306.0249), submitted to JINST.
- [34] Particle Data Group, K. Nakamura *et al.* J. Phys. G **37** (2010) 075021.
- [35] LHCb collaboration, *Average f_s/f_d b-hadron production fraction ratio for 7 TeV pp collisions*, Jul, 2013.
- [36] C. Adrover *et al.*, *Search for the rare decays $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$ with 1.02 fb^1 at LHCb*, LHCb-ANA-2011-102.
- [37] LHCb Collaboration, B. Adeva *et al.*, *Roadmap for selected key measurements of LHCb*, [arXiv:0912.4179](https://arxiv.org/abs/0912.4179).

- [38] Particle Data Group, J. Beringer *et al.*, *Review of particle physics*, Phys. Rev. **D86** (2012) 010001.
- [39] BaBar Collaboration, P. del Amo Sanchez *et al.*, *Study of $B \rightarrow \pi \ell \nu$ and $B \rightarrow \rho \ell \nu$ Decays and Determination of $|V_{ub}|$* , Phys. Rev. **D83** (2011) 032007, [arXiv:1005.3288](https://arxiv.org/abs/1005.3288).
- [40] W.-F. Wang and Z.-J. Xiao, *The semileptonic decays $B/B_s \rightarrow (\pi, K)(l^+ l^-, l \nu, \nu \bar{\nu})$ in the perturbative QCD approach beyond the leading-order*, [arXiv:1207.0265](https://arxiv.org/abs/1207.0265).
- [41] LHCb Collaboration, R. Aaij *et al.*, *First observation of the decay $B^+ \rightarrow \pi^+ \mu^+ \mu^-$* , JHEP **1212** (2012) 125, [arXiv:1210.2645](https://arxiv.org/abs/1210.2645).
- [42] CDF Collaboration, F. Abe *et al.*, *Observation of the B_c meson in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV*, Phys. Rev. Lett. **81** (1998) 2432, [arXiv:hep-ex/9805034](https://arxiv.org/abs/hep-ex/9805034).
- [43] Y.-M. Wang, Y.-L. Shen, and C.-D. Lu, *Lambda(b) p, Lambda transition form factors from QCD light-cone sum rules*, Phys. Rev. **D80** (2009) 074012, [arXiv:0907.4008](https://arxiv.org/abs/0907.4008).
- [44] A. Datta, *Semileptonic decays of Lambda(c) and Lambda(b) baryons involving heavy to light transitions and the determination of $-V(ub)$* , [arXiv:hep-ph/9504429](https://arxiv.org/abs/hep-ph/9504429).
- [45] W. Detmold, C. J. D. Lin, S. Meinel, and M. Wingate, $\Lambda_b \rightarrow pl^- \bar{\nu}$ form factors from lattice QCD with static b quarks, [arXiv:1306.0446](https://arxiv.org/abs/1306.0446).
- [46] P. Robbe. Communication at the LHCb Simulations applications mailing list (23/04/2013).
- [47] A. Khodjamirian, C. Klein, T. Mannel, and Y.-M. Wang, *Form Factors and Strong Couplings of Heavy Baryons from QCD Light-Cone Sum Rules*, JHEP **1109** (2011) 106, [arXiv:1108.2971](https://arxiv.org/abs/1108.2971).
- [48] M. Adinolfi, *Recent observation on production tools types and baryons*, <https://indico.cern.ch/contributionDisplay.py?contribId=2&confId=208314>.
- [49] LHCb tracking group, <https://twiki.cern.ch/twiki/bin/view/LHCb/LHCbTrackingEfficiencies>, .
- [50] J. H. Morata *et al.*, *Measurement of trigger efficiencies and biases*, LHCb-2008-073.
- [51] S. Tolk, J. Albrecht, and F. Dettori, *Measuring the trigger efficiency in data with the tistos method*, LHCb-INT-2013-038.
- [52] LHCb Collaboration, R. Aaij *et al.*, *Determination of f_s/f_d for 7 TeV pp collisions and a measurement of the branching fraction of the decay $B_d \rightarrow D^- K^+$* , Phys. Rev. Lett. **107** (2011) 211801, [arXiv:1106.4435](https://arxiv.org/abs/1106.4435).
- [53] LHCb collaboration, R. Aaij *et al.*, *Measurement of b hadron production fractions in 7 TeV pp collisions*, Phys. Rev. **D85** (2012) 032008, [arXiv:1111.2357](https://arxiv.org/abs/1111.2357).

- [54] LHCb collaboration, R. Aaij *et al.*, *Measurement of the ratio of fragmentation functions f_s/f_d and the dependence on B meson kinematics*, arXiv:1301.5286, submitted to JHEP.
- [55] K. D. Bruyn *et al.*, *Probing New Physics via the $B_s \rightarrow \mu^+ \mu^-$ Effective Lifetime*, Phys. Rev. Lett. **109** (2012) 041801, arXiv:1204.1737.
- [56] LHCb collaboration, R. Aaij *et al.*, *First evidence for the decay $B_s^0 \rightarrow \mu^+ \mu^-$* , Phys. Rev. Lett. **110** (2013) 021801, arXiv:1211.2674.
- [57] M. Perrin-Terrin and G. Mancinelli, *Optimisation of the binning of the discriminating variables used in the computation of $b_s^0 \rightarrow \mu^+ \mu^-$ upper limits with the modified frequentist approach*, Tech. Rep. LHCb-INT-2012-003, CERN, Geneva, Feb, 2012.
- [58] CDF Collaboration, A. Abulencia *et al.*, *Search for $B_s \rightarrow \mu^+ \mu^-$ and $B_d \rightarrow \mu^+ \mu^-$ decays in $p\bar{p}$ collisions with CDF II*, Phys. Rev. Lett. **95** (2005) 221805, arXiv:hep-ex/0508036.

Table 59: Expected combinatorial background, $B \rightarrow h^+ h^-$ background, cross-feed, and signal events assuming SM predictions, together with the number of observed events in $B^0 \rightarrow \mu^+ \mu^-$ mass region, in bins of mass and BDT for the full 3 fb^{-1} sample.

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	677^{+12}_{-12}	$13.03^{+0.56}_{-0.54}$	$3.01^{+0.26}_{-0.25}$	$1.35^{+0.18}_{-0.16}$	$0.50^{+0.13}_{-0.12}$	$0.213^{+0.095}_{-0.075}$	$0.083^{+0.058}_{-0.040}$	$0.030^{+0.037}_{-0.019}$
	Exp. peak. bkg	$0.249^{+0.079}_{-0.059}$	$0.159^{+0.052}_{-0.038}$	$0.105^{+0.034}_{-0.025}$	$0.112^{+0.036}_{-0.026}$	$0.111^{+0.035}_{-0.026}$	$0.106^{+0.034}_{-0.025}$	$0.098^{+0.031}_{-0.023}$	$0.084^{+0.026}_{-0.020}$
	Exp. Cross-feed	$0.0234^{+0.0030}_{-0.0027}$	$0.0144^{+0.0020}_{-0.0018}$	$0.0094^{+0.0013}_{-0.0011}$	$0.0100^{+0.0013}_{-0.0012}$	$0.0100^{+0.0013}_{-0.0012}$	$0.0096^{+0.0013}_{-0.0012}$	$0.0098^{+0.0013}_{-0.0011}$	$0.0101^{+0.0013}_{-0.0012}$
	Exp. signal	$0.0160^{+0.0020}_{-0.0019}$	$0.0098^{+0.0014}_{-0.0013}$	$0.00640^{+0.00085}_{-0.00079}$	$0.00682^{+0.00088}_{-0.00081}$	$0.00683^{+0.00087}_{-0.00081}$	$0.00656^{+0.00086}_{-0.00080}$	$0.00669^{+0.00086}_{-0.00080}$	$0.00688^{+0.00087}_{-0.00080}$
	Observed	685	16	2	3	0	0	1	0
5236 – 5248	Exp. comb. bkg	671^{+12}_{-12}	$12.96^{+0.55}_{-0.53}$	$2.99^{+0.25}_{-0.24}$	$1.33^{+0.18}_{-0.16}$	$0.50^{+0.13}_{-0.11}$	$0.213^{+0.093}_{-0.074}$	$0.083^{+0.057}_{-0.039}$	$0.030^{+0.036}_{-0.019}$
	Exp. peak. bkg	$0.274^{+0.080}_{-0.063}$	$0.174^{+0.052}_{-0.041}$	$0.116^{+0.034}_{-0.027}$	$0.123^{+0.036}_{-0.028}$	$0.122^{+0.036}_{-0.028}$	$0.116^{+0.034}_{-0.027}$	$0.108^{+0.032}_{-0.025}$	$0.092^{+0.027}_{-0.021}$
	Exp. Cross-feed	$0.0264^{+0.0034}_{-0.0030}$	$0.0163^{+0.0023}_{-0.0021}$	$0.0106^{+0.0014}_{-0.0013}$	$0.0113^{+0.0015}_{-0.0013}$	$0.0113^{+0.0015}_{-0.0013}$	$0.0109^{+0.0014}_{-0.0013}$	$0.0111^{+0.0014}_{-0.0013}$	$0.0114^{+0.0014}_{-0.0013}$
	Exp. signal	$0.0395^{+0.0050}_{-0.0047}$	$0.0243^{+0.0035}_{-0.0032}$	$0.0159^{+0.0021}_{-0.0020}$	$0.0169^{+0.0022}_{-0.0020}$	$0.0169^{+0.0022}_{-0.0020}$	$0.0163^{+0.0021}_{-0.0020}$	$0.0166^{+0.0021}_{-0.0020}$	$0.0170^{+0.0022}_{-0.0020}$
	Observed	626	12	4	1	0	0	0	1
5248 – 5254	Exp. comb. bkg	$333.4^{+6.1}_{-6.0}$	$6.45^{+0.27}_{-0.26}$	$1.49^{+0.12}_{-0.12}$	$0.661^{+0.087}_{-0.079}$	$0.247^{+0.064}_{-0.056}$	$0.107^{+0.046}_{-0.037}$	$0.041^{+0.028}_{-0.019}$	$0.0148^{+0.0179}_{-0.0096}$
	Exp. peak. bkg	$0.144^{+0.044}_{-0.034}$	$0.092^{+0.029}_{-0.022}$	$0.061^{+0.019}_{-0.014}$	$0.064^{+0.020}_{-0.015}$	$0.064^{+0.020}_{-0.015}$	$0.061^{+0.019}_{-0.014}$	$0.056^{+0.018}_{-0.013}$	$0.048^{+0.015}_{-0.011}$
	Exp. Cross-feed	$0.0146^{+0.0019}_{-0.0017}$	$0.0090^{+0.0013}_{-0.0012}$	$0.00587^{+0.00079}_{-0.00071}$	$0.00626^{+0.00082}_{-0.00074}$	$0.00627^{+0.00080}_{-0.00073}$	$0.00602^{+0.00080}_{-0.00072}$	$0.00614^{+0.00079}_{-0.00072}$	$0.00631^{+0.00080}_{-0.00072}$
	Exp. signal	$0.0369^{+0.0044}_{-0.0042}$	$0.0227^{+0.0031}_{-0.0028}$	$0.0148^{+0.0019}_{-0.0017}$	$0.0158^{+0.0019}_{-0.0018}$	$0.0158^{+0.0019}_{-0.0018}$	$0.0152^{+0.0019}_{-0.0018}$	$0.0155^{+0.0019}_{-0.0018}$	$0.0159^{+0.0019}_{-0.0018}$
	Observed	336	5	2	3	0	0	0	0
5254 – 5266	Exp. comb. bkg	663^{+12}_{-12}	$12.84^{+0.54}_{-0.52}$	$2.97^{+0.25}_{-0.24}$	$1.31^{+0.17}_{-0.16}$	$0.49^{+0.13}_{-0.11}$	$0.214^{+0.091}_{-0.072}$	$0.082^{+0.055}_{-0.038}$	$0.030^{+0.035}_{-0.019}$
	Exp. peak. bkg	$0.290^{+0.093}_{-0.069}$	$0.185^{+0.061}_{-0.045}$	$0.123^{+0.040}_{-0.029}$	$0.130^{+0.042}_{-0.031}$	$0.129^{+0.042}_{-0.031}$	$0.123^{+0.040}_{-0.029}$	$0.114^{+0.037}_{-0.027}$	$0.097^{+0.031}_{-0.023}$
	Exp. Cross-feed	$0.0329^{+0.0042}_{-0.0038}$	$0.0202^{+0.0029}_{-0.0026}$	$0.0132^{+0.0018}_{-0.0016}$	$0.0141^{+0.0018}_{-0.0017}$	$0.0141^{+0.0018}_{-0.0016}$	$0.0135^{+0.0018}_{-0.0016}$	$0.0138^{+0.0018}_{-0.0016}$	$0.0142^{+0.0018}_{-0.0016}$
	Exp. signal	$0.121^{+0.014}_{-0.013}$	$0.0741^{+0.0098}_{-0.0092}$	$0.0484^{+0.0060}_{-0.0056}$	$0.0515^{+0.0062}_{-0.0057}$	$0.0516^{+0.0062}_{-0.0056}$	$0.0495^{+0.0061}_{-0.0057}$	$0.0506^{+0.0060}_{-0.0056}$	$0.0520^{+0.0060}_{-0.0056}$
	Observed	674	8	4	1	1	1	0	0
5266 – 5302	Exp. comb. bkg	1955^{+36}_{-35}	$38.1^{+1.6}_{-1.5}$	$8.81^{+0.72}_{-0.69}$	$3.85^{+0.49}_{-0.45}$	$1.47^{+0.36}_{-0.32}$	$0.64^{+0.26}_{-0.21}$	$0.25^{+0.16}_{-0.11}$	$0.088^{+0.104}_{-0.056}$
	Exp. peak. bkg	$0.70^{+0.22}_{-0.16}$	$0.45^{+0.14}_{-0.11}$	$0.296^{+0.095}_{-0.069}$	$0.313^{+0.100}_{-0.073}$	$0.311^{+0.099}_{-0.072}$	$0.296^{+0.095}_{-0.069}$	$0.275^{+0.088}_{-0.064}$	$0.235^{+0.074}_{-0.054}$
	Exp. Cross-feed	$0.152^{+0.020}_{-0.018}$	$0.094^{+0.013}_{-0.012}$	$0.0611^{+0.0082}_{-0.0074}$	$0.0651^{+0.0085}_{-0.0077}$	$0.0652^{+0.0084}_{-0.0077}$	$0.0625^{+0.0084}_{-0.0075}$	$0.0639^{+0.0083}_{-0.0075}$	$0.0656^{+0.0083}_{-0.0075}$
	Exp. signal	$0.568^{+0.067}_{-0.063}$	$0.350^{+0.046}_{-0.043}$	$0.228^{+0.028}_{-0.026}$	$0.243^{+0.029}_{-0.027}$	$0.243^{+0.029}_{-0.027}$	$0.234^{+0.029}_{-0.027}$	$0.238^{+0.029}_{-0.027}$	$0.245^{+0.029}_{-0.027}$
	Observed	1897	42	10	4	3	0	2	1
5302 – 5314	Exp. comb. bkg	641^{+12}_{-11}	$12.55^{+0.51}_{-0.49}$	$2.90^{+0.23}_{-0.22}$	$1.26^{+0.16}_{-0.15}$	$0.49^{+0.12}_{-0.10}$	$0.215^{+0.085}_{-0.069}$	$0.081^{+0.051}_{-0.036}$	$0.029^{+0.034}_{-0.019}$
	Exp. peak. bkg	$0.131^{+0.042}_{-0.032}$	$0.084^{+0.027}_{-0.021}$	$0.056^{+0.018}_{-0.013}$	$0.059^{+0.019}_{-0.014}$	$0.058^{+0.019}_{-0.014}$	$0.056^{+0.018}_{-0.013}$	$0.052^{+0.016}_{-0.012}$	$0.044^{+0.014}_{-0.011}$
	Exp. Cross-feed	$0.104^{+0.013}_{-0.012}$	$0.0643^{+0.0091}_{-0.0082}$	$0.0419^{+0.0056}_{-0.0051}$	$0.0447^{+0.0058}_{-0.0052}$	$0.0448^{+0.0057}_{-0.0052}$	$0.0430^{+0.0057}_{-0.0051}$	$0.0438^{+0.0056}_{-0.0051}$	$0.0451^{+0.0056}_{-0.0051}$
	Exp. signal	$0.121^{+0.014}_{-0.013}$	$0.0741^{+0.0099}_{-0.0092}$	$0.0484^{+0.0060}_{-0.0056}$	$0.0515^{+0.0062}_{-0.0057}$	$0.0516^{+0.0062}_{-0.0057}$	$0.0495^{+0.0061}_{-0.0057}$	$0.0505^{+0.0060}_{-0.0056}$	$0.0520^{+0.0060}_{-0.0057}$
	Observed	585	14	2	1	1	0	2	0
5314 – 5320	Exp. comb. bkg	$318.5^{+5.8}_{-5.7}$	$6.25^{+0.25}_{-0.24}$	$1.45^{+0.11}_{-0.11}$	$0.624^{+0.079}_{-0.072}$	$0.243^{+0.058}_{-0.051}$	$0.108^{+0.042}_{-0.034}$	$0.040^{+0.025}_{-0.018}$	$0.0146^{+0.0168}_{-0.0093}$
	Exp. peak. bkg	$0.049^{+0.018}_{-0.013}$	$0.0312^{+0.0115}_{-0.0083}$	$0.0207^{+0.0076}_{-0.0054}$	$0.0219^{+0.0080}_{-0.0057}$	$0.0218^{+0.0079}_{-0.0057}$	$0.0208^{+0.0076}_{-0.0054}$	$0.0192^{+0.0070}_{-0.0050}$	$0.0164^{+0.0059}_{-0.0043}$
	Exp. Cross-feed	$0.0623^{+0.0098}_{-0.0087}$	$0.0383^{+0.0065}_{-0.0057}$	$0.0250^{+0.0041}_{-0.0036}$	$0.0266^{+0.0043}_{-0.0038}$	$0.0266^{+0.0042}_{-0.0038}$	$0.0256^{+0.0042}_{-0.0037}$	$0.0261^{+0.0042}_{-0.0037}$	$0.0268^{+0.0043}_{-0.0037}$
	Exp. signal	$0.0369^{+0.0044}_{-0.0041}$	$0.0227^{+0.0031}_{-0.0028}$	$0.0148^{+0.0019}_{-0.0017}$	$0.0158^{+0.0019}_{-0.0018}$	$0.0158^{+0.0019}_{-0.0018}$	$0.0152^{+0.0019}_{-0.0018}$	$0.0155^{+0.0019}_{-0.0017}$	$0.0159^{+0.0019}_{-0.0018}$
	Observed	288	8	0	0	0	0	0	0
5320 – 5332	Exp. comb. bkg	633^{+11}_{-11}	$12.44^{+0.50}_{-0.48}$	$2.88^{+0.23}_{-0.22}$	$1.24^{+0.16}_{-0.14}$	$0.48^{+0.11}_{-0.10}$	$0.216^{+0.083}_{-0.067}$	$0.080^{+0.050}_{-0.036}$	$0.029^{+0.033}_{-0.018}$
	Exp. peak. bkg	$0.075^{+0.031}_{-0.021}$	$0.048^{+0.020}_{-0.014}$	$0.0318^{+0.013}_{-0.0091}$	$0.0337^{+0.0141}_{-0.0096}$	$0.0335^{+0.0140}_{-0.0091}$	$0.0319^{+0.0134}_{-0.0091}$	$0.0296^{+0.0124}_{-0.0084}$	$0.0253^{+0.0105}_{-0.0072}$
	Exp. Cross-feed	$0.289^{+0.040}_{-0.036}$	$0.178^{+0.027}_{-0.024}$	$0.116^{+0.017}_{-0.015}$	$0.124^{+0.017}_{-0.016}$	$0.124^{+0.017}_{-0.016}$	$0.119^{+0.017}_{-0.015}$	$0.121^{+0.017}_{-0.015}$	$0.125^{+0.017}_{-0.015}$
	Exp. signal	$0.0395^{+0.0050}_{-0.0047}$	$0.0243^{+0.0035}_{-0.0031}$	$0.0159^{+0.0021}_{-0.0020}$	$0.0169^{+0.0022}_{-0.0020}$	$0.0169^{+0.0022}_{-0.0020}$	$0.0162^{+0.0022}_{-0.0020}$	$0.0166^{+0.0021}_{-0.0020}$	$0.0170^{+0.0022}_{-0.0020}$
	Observed	614	12	1	1	0	1	1	1
5332 – 5344	Exp. comb. bkg	628^{+11}_{-11}	$12.36^{+0.49}_{-0.48}$	$2.86^{+0.22}_{-0.22}$	$1.22^{+0.15}_{-0.14}$	$0.482^{+0.113}_{-0.108}$	$0.216^{+0.082}_{-0.067}$	$0.080^{+0.049}_{-0.035}$	$0.029^{+0.033}_{-0.018}$
	Exp. peak. bkg	$0.056^{+0.027}_{-0.017}$	$0.035^{+0.017}_{-0.011}$	$0.0235^{+0.0113}_{-0.0074}$	$0.0249^{+0.0120}_{-0.0078}$	$0.0248^{+0.0119}_{-0.0077}$	$0.0236^{+0.0114}_{-0.0074}$	$0.0218^{+0.0105}_{-0.0068}$	$0.0187^{+0.0089}_{-0.0058}$
	Exp. Cross-feed	$0.674^{+0.086}_{-0.078}$	$0.415^{+0.059}_{-0.053}$	$0.271^{+0.036}_{-0.033}$	$0.289^{+0.037}_{-0.034}$	$0.289^{+0.037}_{-0.034}$	$0.277^{+0.037}_{-0.033}$	$0.283^{+0.037}_{-0.033}$	$0.291^{+0.037}_{-0.033}$
	Exp. signal	$0.0134^{+0.0020}_{-0.0018}$	$0.0082^{+0.0013}_{-0.0012}$	$0.00538^{+0.00084}_{-0.00075}$	$0.00573^{+0.00087}_{-0.00079}$	$0.00573^{+0.00087}_{-0.00078}$	$0.00551^{+0.00085}_{-0.00077}$	$0.00562^{+0.00085}_{-0.00078}$	$0.00578^{+0.00086}_{-0.00078}$
	Observed	575	10	5	2	2	1	0	0