
Background studies for $B^0 \rightarrow \mu^+ \mu^-$ analysis optimization

F. Archilli¹, M. Palutan², M. Rama³, A. Sarti^{2,4}, B. Sciascia², R. Vazquez Gomez²

¹ European Organisation for Nuclear Research (CERN), Geneva, Switzerland.

² INFN, Laboratori Nazionali di Frascati, Italy.

³ INFN, Pisa, Italy.

⁴ Dipartimento di Scienze di Base e Applicate per l'Ingegneria, “La Sapienza” Università di Roma, Italy.

Abstract

After the recent observation of $B_s^0 \rightarrow \mu^+ \mu^-$ decay, attention is focussing now on the optimization of the sensitivity for the $B^0 \rightarrow \mu^+ \mu^-$ decay mode. On this respect, the main issues are the control of the systematics in the determination of the peaking background, and the optimization of the rejection cuts for both the peaking background and the combinatorial background. In this document we review the recent progresses we made along the lines described above, consisting of an improved determination of the hadron PID misidentification probabilities, the usage of a new PID selection in the rejection of peaking background, and the introduction of more performant isolation variables in the rejection of the combinatorial background. We conclude by giving estimates of the impact of these improvements on the $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ expected sensitivities.

Contents

1	Introduction	1
2	Background overview	1
3	Muon isolation studies	4
3.1	Cut-based muon isolation	4
3.2	BDT muon isolation <i>isoBDT</i>	5
3.2.1	Discriminating variables	5
3.2.2	Training of the BDT	7
3.2.3	isoBDT of the event	7
3.3	Performance comparison: isoBDT vs cut-based isolation	10
3.4	Performance comparison: isoBDT used in global BDT	13
3.5	Prospects	17
4	PID efficiency from data	20
4.1	Muons	21
4.2	Pions and kaons	22
4.3	Protons	25
4.4	Summary on PID efficiency evaluation	29
5	PID selection and exclusive background estimate	31
5.1	BDT PDFs	33
5.2	Mass PDFs	35
6	Combinatorial background estimate	39
7	Sensitivity studies	42
7.1	Limit case: $B_{(s)}^0 \rightarrow h^+ h^{(\prime)-}$ background reduction	43
7.2	Limit case: combinatorial background reduction	45
8	Conclusions	47
	References	48

¹ 1 Introduction

² The $B_s^0 \rightarrow \mu^+ \mu^-$ decay has been recently observed with a significance of 6.2σ through the
³ combined analysis of LHCb and CMS data of RUN I [1]. The measured BR, $2.8_{-0.6}^{+0.7} \times 10^{-9}$,
⁴ is in agreement with the SM prediction, $(3.66 \pm 0.23) \times 10^{-9}$ [2]. With the same data,
⁵ an evidence of the $B^0 \rightarrow \mu^+ \mu^-$ decay with 3.2σ significance was observed (1.8σ and 2.6σ
⁶ from LHCb and CMS data, respectively). The measured BR is in this case $3.9_{-1.4}^{+1.6} \times 10^{-10}$,
⁷ which is 2.2σ above the SM prediction, $(1.06 \pm 0.09) \times 10^{-10}$ [2].

⁸ We remind here that a departure from the SM prediction on the B_d/B_s ratio would fal-
⁹ sify the Minimal Flavour Violation (MFV) hypothesis, consisting of identifying the flavour
¹⁰ symmetry and symmetry breaking structure of the SM and enforcing it also for physics
¹¹ beyond the SM (BSM). The latter hypothesis is indeed an excellent phenomenological
¹² explanation for the lack of BSM physics signals in the flavor physics data so far. For the
¹³ above reasons, the clarification of the experimental picture on the $B^0 \rightarrow \mu^+ \mu^-$ decay rate
¹⁴ is the challenge for this analysis during LHC RUN II. In this document, we review the
¹⁵ studies we made during the past months towards a more performant $B^0 \rightarrow \mu^+ \mu^-$ analysis,
¹⁶ and which would hardly fit in the official note of the next analysis round.

¹⁷ For LHCb, the analysis approach used so far had the primary goal of maximizing the
¹⁸ acceptance on $B_s^0 \rightarrow \mu^+ \mu^-$ signal, so that only loose PID cuts have been applied against
¹⁹ $B_{(s)}^0 \rightarrow h^+ h^{(\prime)-}$ peaking background ($h, h' = K, \pi$), which mostly affects the B^0 region.
²⁰ This implies that $B^0 \rightarrow \mu^+ \mu^-$ sensitivity could really benefit from an optimized PID
²¹ selection strategy. We review the PID selection strategy in Sect. 5.

²² The other important background source in B^0 signal region is represented by dimuon
²³ combinatorial events. In this case, a better rejection power essentially rests on the
²⁴ improvement of the muon isolation algorithms, which in the published analysis [3] were
²⁵ defined by using rectangular cuts on the relevant variables. Progresses in the definition of
²⁶ the isolation algorithms are discussed in Sect. 3.

²⁷ Not only the background rejection is relevant to the determination of the B^0 sensitivity,
²⁸ but also systematics can play a crucial role, and especially the ones affecting the peaking
²⁹ background yield. The latter is determined from the observed number of exclusive
³⁰ $B_{(s)}^0 \rightarrow h^+ h^{(\prime)-}$ events combined with the $\pi \rightarrow \mu$ and $K \rightarrow \mu$ misID probabilities, as
³¹ measured from data control samples. In Sect. 4 we describe the improvements achieved
³² in the extraction of misID probabilities from data, and their impact on the background
³³ determination.

³⁴ Finally, in Sect. 7, we discuss the impact of our achievements on the signal sensitivity,
³⁵ for both B_s^0 and B^0 , as determined from toy MC studies.

³⁶ 2 Background overview

³⁷ The background yield leaking into B_s^0 and B^0 mass windows is dominated by combinatorics
³⁸ of two muons coming from different B hadrons, with the $B_{(s)}^0 \rightarrow h^+ h^{(\prime)-}$ and $\Lambda_b^0 \rightarrow p \mu^- \nu$
³⁹ misID backgrounds also playing a role, especially in the B^0 window. This is shown in

Fig. 1, where the fit to the 3 fb^{-1} data sample is shown, for BDT above 0.7. Many

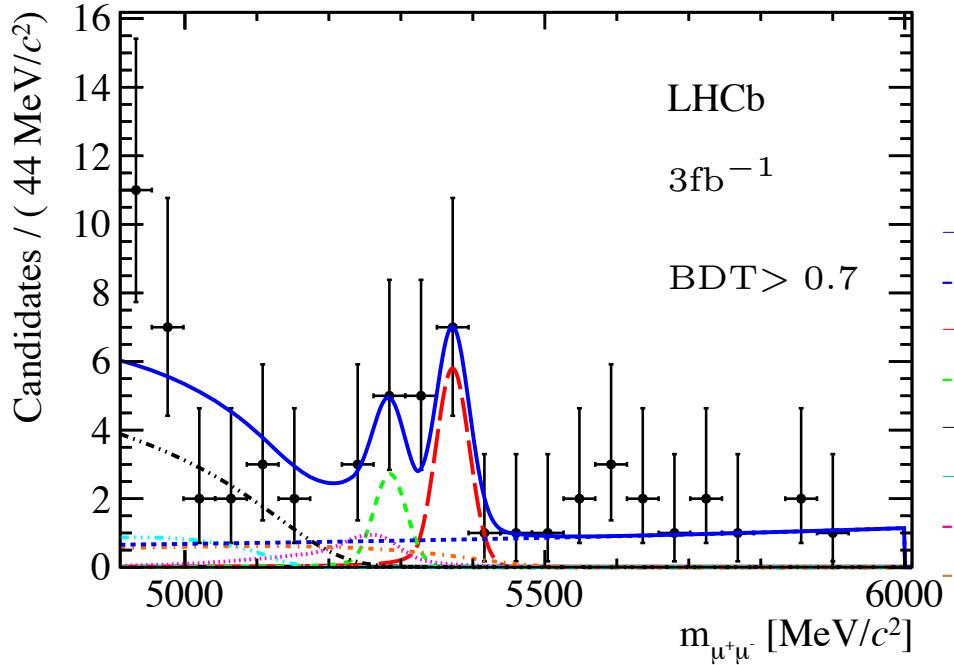


Figure 1: Invariant mass distribution of the selected signal candidates (black dots) with $\text{BDT} > 0.7$. The result of the fit is overlaid (blue solid line) and the different components detailed: $B_s^0 \rightarrow \mu^+\mu^-$ (red long dashed), $B^0 \rightarrow \mu^+\mu^-$ (green medium dashed), $B_{(s)}^0 \rightarrow h^+h^-$ (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), $\Lambda_b^0 \rightarrow p\mu^-\nu$ (orange dot-dashed), and the combinatorial background (blue medium dashed).

40
 41 exclusive B decay channels have been considered which may pollute the mass sidebands,
 42 and thus affect the combinatorial background estimate. A summary of the estimates for
 43 the exclusive background made at the time of the 3 fb^{-1} data publication [3] is shown in
 44 Table 1, taken from Ref. [4], where the expected yields for all of the relevant channels in
 45 the mass range [4900-6000] MeV/c^2 , and for BDT above 0.8 are listed.

46 The following strategy has been adopted in Ref. [4] to treat the various exclusive
 47 background components in the fit:

- 48 • The $B^0 \rightarrow \pi^-\mu^+\nu_\mu$, $B_{(s)}^0 \rightarrow h^+h^{(-)}$ and $B^{0(+)} \rightarrow \pi^{0(+)}\mu^+\mu^-$ components have been
 49 included as separate PDFs, with normalization fluctuating in the fit according to the
 50 estimated errors; these decays sources represent $\sim 80\%$ of the exclusive background
 51 in the considered mass and BDT range, and moreover they are all measured decay
 52 channels.
- 53 • The $\Lambda_b^0 \rightarrow p\mu^-\nu$ gives a sizable contribution to the yield, but is largely unknown.
 54 In Ref [3] it has been only treated as a source of systematic, consisting of excluding

Table 1: 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] MeV/ c^2 , for BDT above 0.8. Taken from Ref. [4].

	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

55 it from the baseline fit, and adding the difference in the BRs obtained with this
 56 component as a systematic error¹. The latter choice has been reversed in Ref. [1],
 57 where the $\Lambda_b^0 \rightarrow p \mu^- \nu$ background source has been treated independently.

- 58 • The $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$, which is less relevant and doesn't give appreciable contribution
 59 in the signal region, is accounted for by adding the expected number of events to the
 60 $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$, which has a very similar mass shape, and summing in quadrature
 61 the quoted uncertainty on the normalization.
- 62 • The $B_c^+ \rightarrow J/\psi \mu^+ \nu$ has been neglected, since its contribution is small compared
 63 to the other channels, and moreover its mass shape is well approximated by an
 64 exponential, like the combinatorial, within the expected statistics.

65 For all of the exclusive channels but for $B_{(s)}^0 \rightarrow h^+ h^{(\prime)-}$, the normalisation is computed
 66 relative to the $B^+ \rightarrow J/\psi K^+$, according to their measured (or predicted) BRs, and using
 67 the MC efficiencies for all selection steps but for the PID cuts, which are evaluated instead
 68 from data control samples. The trigger efficiency is also taken from MC, after all selection
 69 cuts, including PID, are applied. Finally, the mass and BDT PDFs are extracted from
 70 MC, with inclusion of event-by-event PID weights.

71 The $B_{(s)}^0 \rightarrow h^+ h^{(\prime)-}$ are determined by measuring the yields of the exclusive decays
 72 directly on data, and combining them with the $\pi \rightarrow \mu$ and $K \rightarrow \mu$ misID probabilities,
 73 as measured from data control samples. The mass PDFs for $B_{(s)}^0 \rightarrow h^+ h^{(\prime)-}$ with double
 74 misID are computed from MC, and implementing a smearing procedure which takes into
 75 account the effect of decays in flight.

76 Finally, the combinatorial background is parameterized with an exponential function,
 77 the parameters of which are free to vary in the fit.

¹ This choice has been done before the unblinding, not to bias the sensitivity.

78 **3 Muon isolation studies**

79 One of the most powerful tool to distinguish signal from background exploits the concept
80 of *muon isolation*, that is the property of the muons from $B \rightarrow \mu\mu$ to be relatively isolated
81 with respect to the other tracks of the events. In this section we present the implementation
82 and the performance studies of a new muon isolation for $B \rightarrow \mu\mu$ developed after the
83 latest measurement published on the Run1 dataset. In Sect. 3.1 we briefly describe the
84 “cut-based” isolation used in the last publication. In Sect. 3.2 we describe the new isolation,
85 that we call *isoBDT*. Next we compare the performance of the isoBDT and the cut-based
86 isolation.

87 **3.1 Cut-based muon isolation**

88 The muon isolation used in the 2013 $B \rightarrow \mu\mu$ measurement is described in detail in [5].
89 The isolation of a given muon candidate from $B \rightarrow \mu\mu$ is defined as the number of long
90 tracks² *close* to the muon, where *close* is specified by a set of rectangular cuts on eight
91 variables grouped below in two sets, labeled as “set A” and “set B”.

92

93 **set A:**

- 94
 - trk_ips: the minimum of the square root of the impact parameter χ^2 of the track
95 with respect to any primary vertex (PV) of the event
 - trk_pvdis: the signed distance between the (track,muon) vertex and the PV
 - trk_svdis: the signed distance between the (track,muon) vertex and the $B \rightarrow \mu\mu$
96 vertex
 - trk_doca: the DOCA between the track and the muon
 - trk_angle: the angle between the track and the muon
- 97
 - $f_c = \frac{|\vec{P}_\mu + \vec{P}_{trk}| \sin(\alpha^{\mu+trk,PV})}{|\vec{P}_\mu + \vec{P}_{trk}| \sin(\alpha^{\mu+trk,PV}) + P_{T_\mu} + P_{T_{trk}}}$, where $\alpha^{\mu+trk,PV}$ is the angle between the sum
98 of the muon and track momenta (\vec{P}_μ , \vec{P}_{trk}), and the direction defined by the PV and
99 the (track,muon) vertex [5]. $f_c \rightarrow 0$ if the track+muon system originates from the
100 PV.

101 **set B:**

- 102
 - trk_gho: track ghost probability
 - trk_chi: track χ^2

103 ²Excluding the other muon candidate

	trk_angle	f_c	trk_doca	trk_ips	trk_pvdis	trk_svdis	trk_gho	trk_chi
sum_isolation	< 0.27	< 0.60	< 0.13	> 3	[0.5, 40]	[-0.15, 30]	—	—
sum_isolation_tc	< 0.27	< 0.60	< 0.13	> 3	[0.5, 40]	[-0.15, 30]	< 0.3	< 3
sum_isolation_new1	< 0.32	< 0.62	< 0.125	> 2.75	> 0.7	> -0.35	< 0.35	< 4
sum_isolation_new2	< 0.27	< 0.60	< 0.13	> 3	[0.5, 40]	[-0.15, 30]	—	< 5
sum_isolation_newD0	< 0.13	< 0.67	< 0.13	> 2.6	> -2.2	> -0.8	< 0.6	< 5

Table 2: Selection requirements for different versions of the cut-based muon isolation [5]. See Sect. 3.1.

108 A number of versions of the cut-based muon isolation were considered [5], differing by the
109 requirements imposed on the set of variables. Their definition is summarized in Table 2.

110 The muon isolation of the event is defined as the sum of the isolations of the two
111 muons. Fig. 2 shows an example of the cut-based isolation distributions for signal and
112 combinatorial background events, taken from ref. [5]. The signal is peaked at zero, while
113 the background has a long tail at larger values. From now on with “cut-based muon
114 isolation” we will refer to the isolation of the event unless otherwise specified.

115 From the physics point of view the muon isolation defined above is sensitive to the
116 number of reconstructed tracks belonging to the same decay tree of the muon candidate.
117 These tracks are labeled as “non-isolating”, as opposed to the “isolating” tracks which do
118 not share an ancestor with the muon [5]. We will keep using this notation in the following
119 sections. The non-isolating tracks are more likely to be close to the muon candidate
120 compared to the other tracks of the event. If the $B \rightarrow \mu\mu$ candidate is signal there are
121 no other tracks sharing the same decay tree of the two muons and the isolation peaks at
122 zero. In combinatorial background, where most of the muon candidates originate from
123 semileptonic decays, the tracks belonging to the same decay tree of the muons are likely
124 to be closer to them and consequently on average the isolation variable is larger.

125

126 3.2 BDT muon isolation *isoBDT*

127 As an alternative approach, the non-isolating and isolating tracks can be discriminated
128 using a multivariate classifier taking as input the variables used to define the cut-based
129 isolation discussed in Sect. 3.1, or others. This strategy has been adopted in the LHCb
130 search for $\tau \rightarrow \mu\mu\mu$ [6, 7]. In this section we describe the implementation of a BDT-based
131 muon isolation for $B \rightarrow \mu\mu$, which is called *isoBDT* in the following [8].

132 3.2.1 Discriminating variables

133 We consider four set of variables as input to the BDT classifier. The sets A and B have
134 been defined in Sect. 3.1 and include the variables used to define the cut-based muon
135 isolations listed in Table 2. In addition, we consider the sets “C” and “D” defined below.

136 **set C:**

- 137 • trk_dphi: difference between the azimuthal angles of track and muon (absolute
138 value)
- 139 • trk_deta: difference between the pseudorapidities of track and muon (absolute value)
- 140 • trk_pt: transverse momentum of the track

141 **set D:**

- 142 • trk_ch: relative charge of the track with respect to the muon candidate charge
- 143 • trk_nnmu, trk_nnpi, trk_nnk: ProbNNmu, ProbNNpi and ProbNNk of the track,
144 respectively

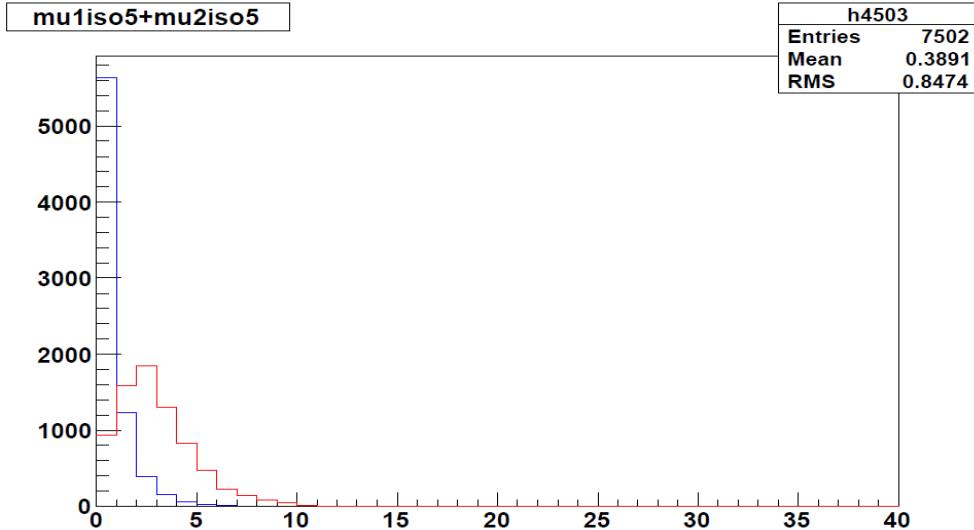


Figure 2: Example of cut-based muon isolation distribution of signal (blue) and combinatorial background (red) [5].

145 The distributions of the variables in set A, B and C are reported in Figs. 3-5. Each
146 plot compares the distribution of non-isolating and isolating tracks in generic MC, and
147 isolating tracks in $B \rightarrow \mu\mu$ MC. Here and in the following we use “generic MC” to indicate
148 simulated events of the kind $b\bar{b} \rightarrow \mu^+\mu^- X$.

149 It is worth to spend a few words on the PID-related variables of set D. The charge
150 and flavour distributions of the non-isolating tracks have distinctive features originating
151 from the decay tree to which they belong. Most of muon candidates in the combinatorial

152 background are muons from semileptonic B decays (see Table 28 in [4]). Let us consider a
153 “typical” \bar{b} semileptonic decay:

$$\begin{aligned}\bar{b} &\rightarrow \bar{c}W^+ \\ W^+ &\rightarrow \mu^+ X \\ \bar{c} &\rightarrow \bar{s}W^- \\ \bar{s} &\rightarrow K^+ X \\ W^- &\rightarrow \mu^- X \text{ or } \pi^- X.\end{aligned}\tag{1}$$

154 From eq. 1 one expects that among the non-isolating tracks associated to the selected muon
155 there are more same-charge kaons than opposite-charge kaons, and more opposite-charge
156 muons/pions than same-sign muons/pions. This is confirmed in generic MC, as shown
157 in Fig. 6. The left plot shows the non-isolating track population associated to a selected
158 positive muon. All charges flip sign in case of a selected negative muon. The right plot
159 shows the same distributions for the isolating tracks. The different pattern can be exploited,
160 in principle, to further distinguish isolating and non-isolating tracks. Fig. 7 shows the
161 distribution of trk_ch , trk_nnk and trk_nnpi for non-isolating and isolating tracks in generic
162 MC, and isolating tracks in $B \rightarrow \mu\mu$ MC.

163 3.2.2 Training of the BDT

164 A Boosted Decision Tree classifier³ is used to separate the categories composed of non-
165 isolating tracks in generic MC and isolating tracks in signal $B_s \rightarrow \mu\mu$ MC. Instead of the
166 latter, the dataset of isolating tracks in generic MC was also tested but showed a worse
167 performance.

168 Four sets of variables are compared: set A, A+B, A+B+C and A+B+C+D (definition
169 in Sect. 3.1 and 3.2.1). The initial set of BDT configuration parameters was:

```
170 ntrees=200
171 nevtmin=2500
172 maxdepth=3
173 ncuts=30
174 adaboostbeta=0.5
```

175 These values have changed to improve the performance as discussed in Sect. 3.3. Fig. 8
176 shows the outcome of the BDT training for the sets of variables A and A+B+C.

177 3.2.3 isoBDT of the event

178 The resulting BDT classifies each longtrack of the event with two values, the BDT outputs
179 of the track with respect to the μ^+ and the μ^- candidates. We refer to these values as
180 $\text{iso1}(\text{track})$ and $\text{iso2}(\text{track})$. Several criteria have been considered to classify the whole
181 event with one number, which will be called isoBDT from now on:

³The TMVA toolkit was used to train the BDT.

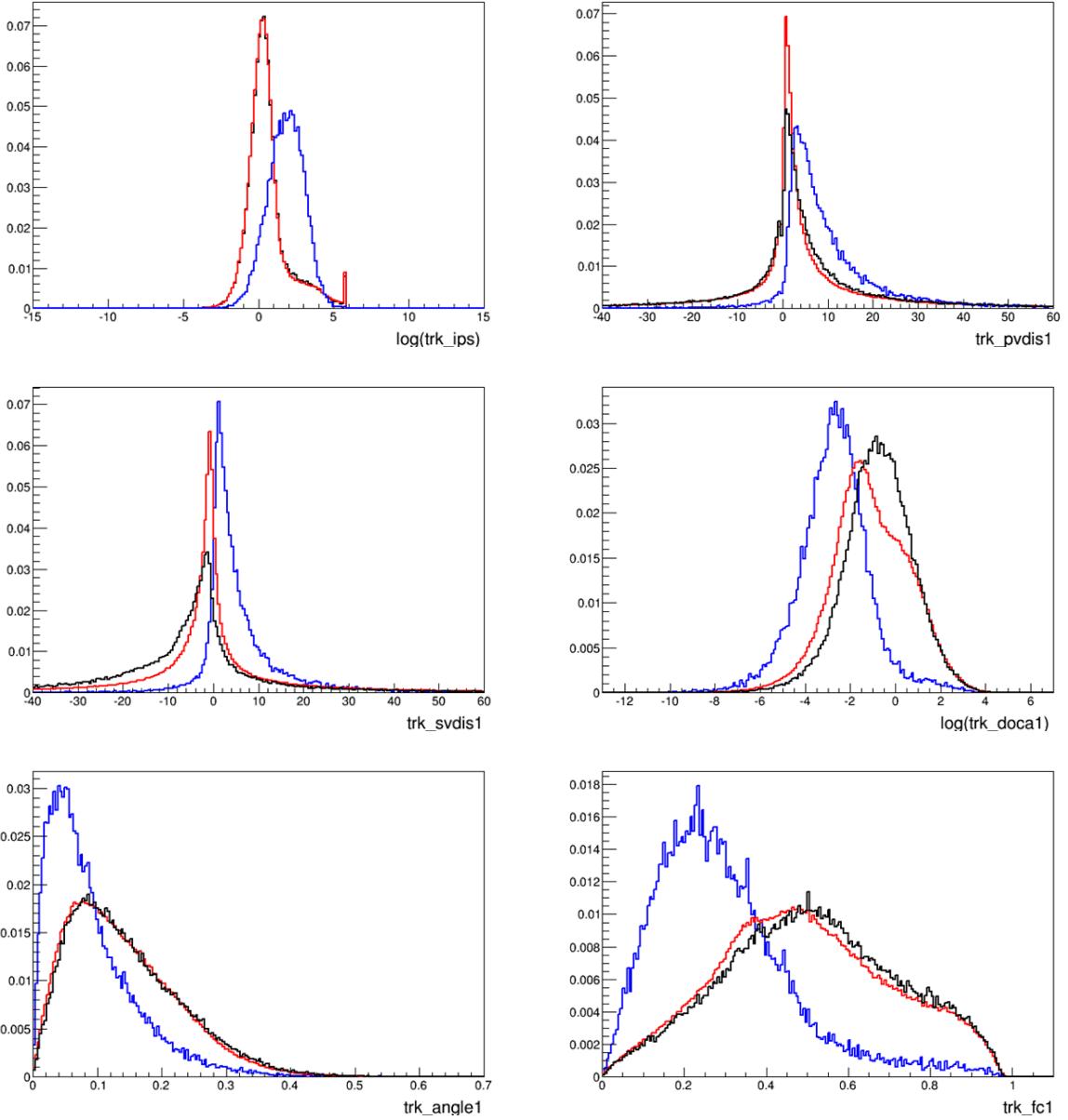


Figure 3: Distributions of the variables belonging to “set A” defined in Sect. 3.1. Blue: non-isolating tracks in generic MC; Black: isolating tracks in signal MC; Red: isolating tracks in generic MC.

- 182 ● iso_max2: maximum value of $\text{iso1}(\text{track}_i) + \text{iso2}(\text{track}_j)$, where track_i and track_j
183 are not required to be the same track
- 184 ● iso_ave2: mean of the two largest iso1 plus mean of the two largest iso2
- 185 ● iso_ave3: mean of the three largest iso1 plus mean of the three largest iso2

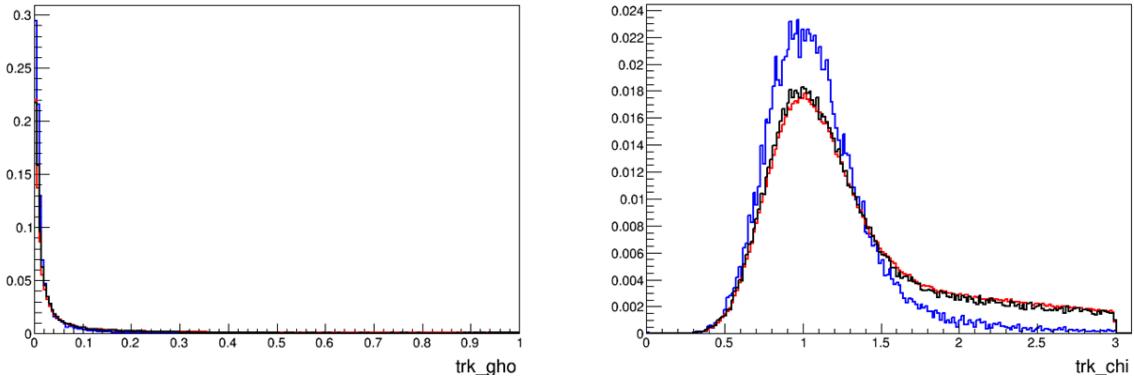


Figure 4: Distributions of the variables belonging to “set B” defined in Sect. 3.1. Blue: non-isolating tracks in generic MC; Black: isolating tracks in signal MC; Red: isolating tracks in generic MC.

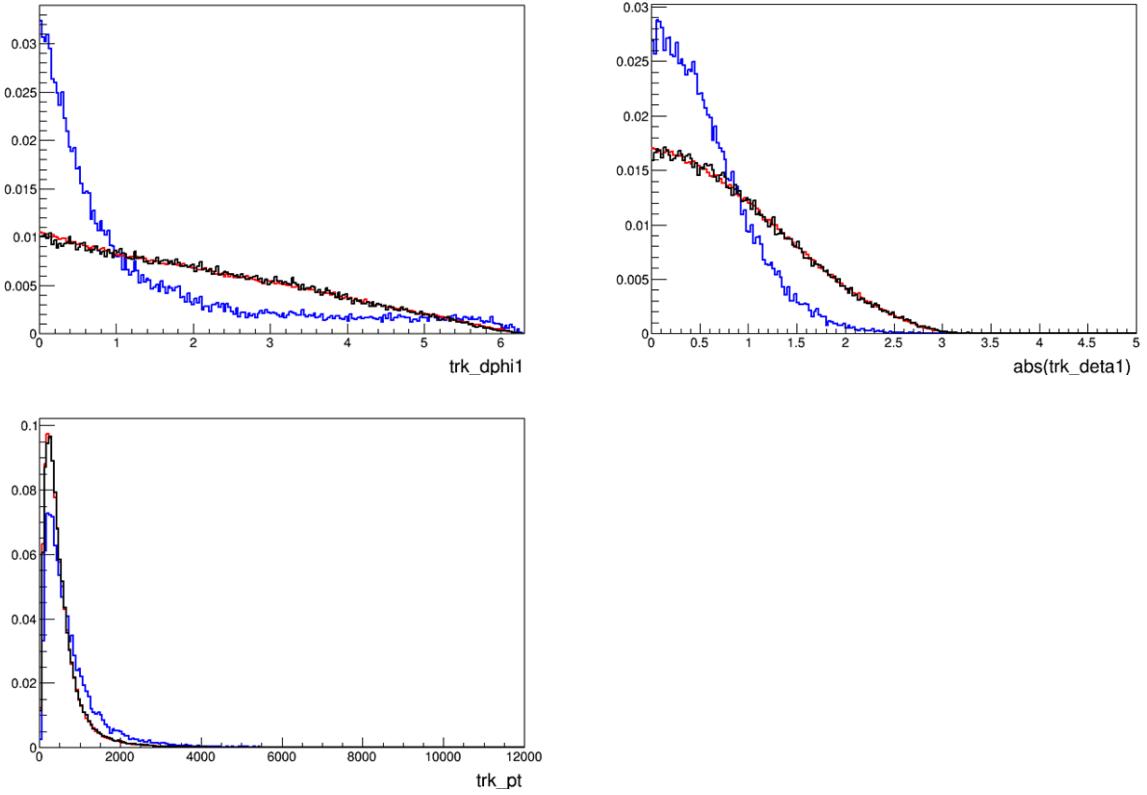


Figure 5: Distributions of the variables belonging to “set C” defined in Sect. 3.2.1. Blue: non-isolating tracks in generic MC; Black: isolating tracks in signal MC; Red: isolating tracks in generic MC.

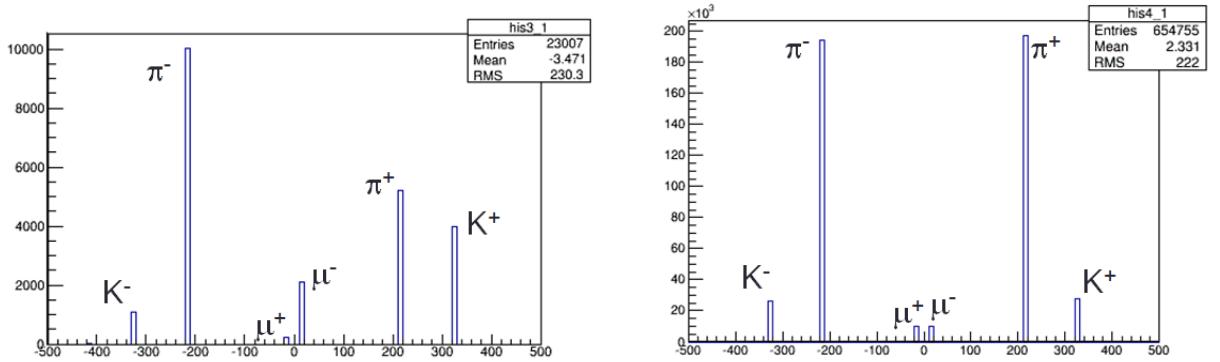


Figure 6: Track population in generic MC, split in non-isolating tracks (left) and isolating tracks (right). The track charge is relative to the muon charge. The track type (K, π, μ) is determined from the MC truth information.

- 186 • iso_ave4: mean of the four largest iso1 plus mean of the four largest iso2

187 The performance of these criteria are compared in the following sections. Other choices
 188 have been tested but showed a lower performance: iso1(track.i)+iso2(track.i) (the track
 189 is constrained to be the same in iso1 and iso2) and the fraction of tracks with $BDT > x$
 190 ($x = 0, 0.1, 0.2 \dots$).

191 Fig. 9 shows the resulting distributions of the isoBDT (set A) for $B_s \rightarrow \mu\mu$ signal
 192 events and generic MC events.

193 3.3 Performance comparison: isoBDT vs cut-based isolation

194 In this section we compare the performance of a number of versions of isoBDT defined in
 195 Sect. 3.2.3 with five versions of the cut-based isolation defined in Table 2. As figure of
 196 merit the ROC curve is used, which displays the background rejection power as a function
 197 of the signal efficiency. Fig. 10 compares the resulting ROCs obtained using $B_s \rightarrow \mu\mu$
 198 signal MC and generic MC (i.e. combinatorial background). The ROCs for the cut-based
 199 isolations are composed of a limited number of points due to the discrete nature of their
 200 distributions (see Fig. 2). The plots differ by the set of input variables used in the isoBDT
 201 definition or by the values of the BDT configuration parameters, and are displayed in
 202 order of increasing performance. A description follows, with the labels a)–h) referring to
 203 the plots of Fig. 10 from top-left down to bottom-right:

- 204 a) The input variables for the isoBDT are those in set A (Sect. 3.1). The training
 205 datasets are non-isolating tracks from generic MC and isolating tracks from generic
 206 MC. The BDT configuration parameters are the ones listed in Sect. 3.2.2.

 207 b) As a) with the dataset of isolating tracks from generic MC replaced with isolating
 208 tracks from signal $B_s \rightarrow \mu\mu$ MC.

 209 c) As b) using sets A+B instead of A.

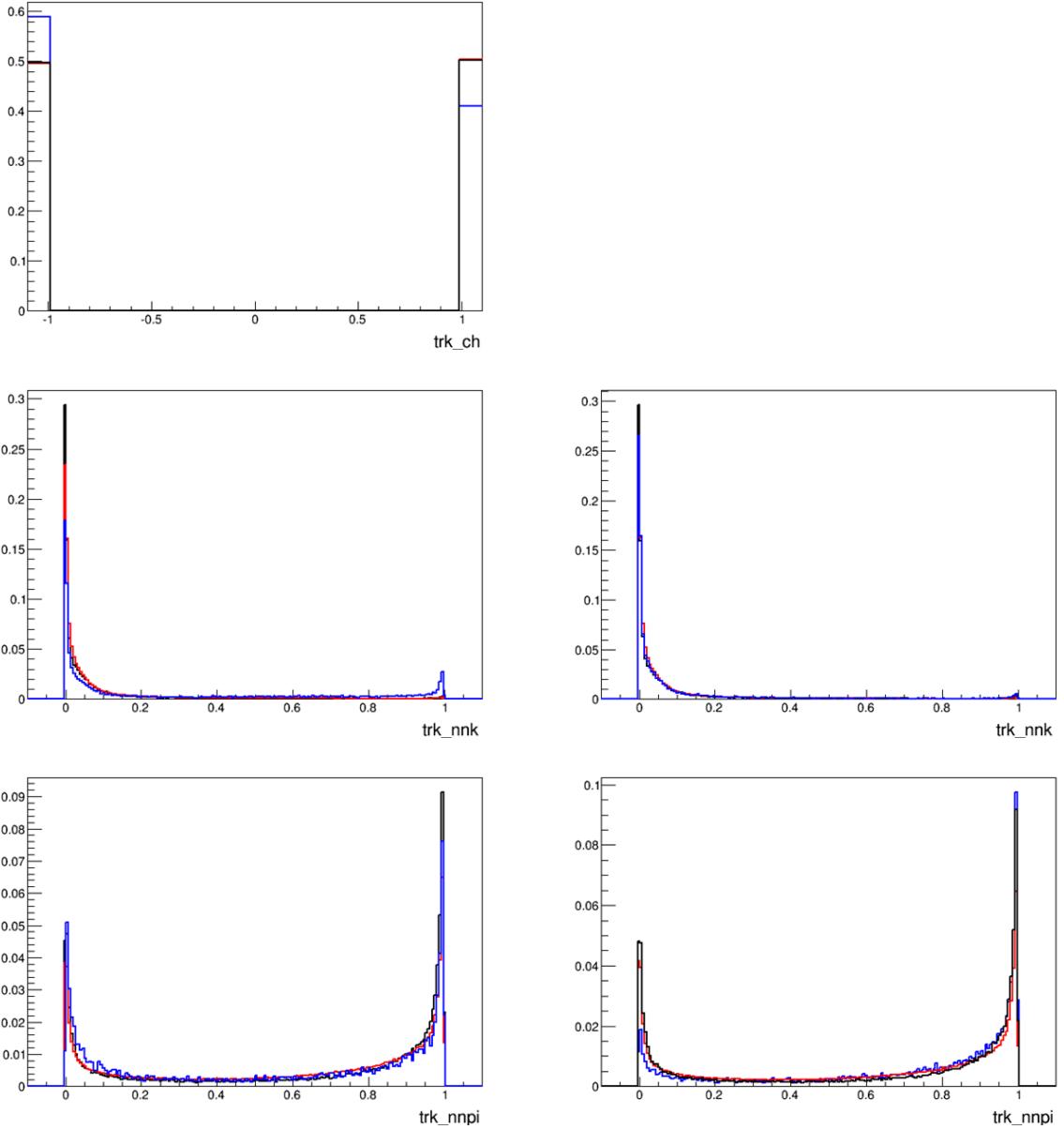


Figure 7: Distributions of variables related to set “D” defined in Sect. 3.2.1. Top: trk_ch . Middle: ProbNNk for $\text{trk_ch} > 0$ (left) and $\text{trk_ch} < 0$ (right). Bottom: ProbNNpi for $\text{trk_ch} > 0$ (left) and $\text{trk_ch} < 0$ (right). Blue: non-isolating tracks in generic MC; Black: isolating tracks in signal MC; Red: isolating tracks in generic MC.

- 210 d) As c) using sets A+B+C instead of A+B.
- 211 e) As d) changing the BDT parameter `ntrees` from 200 to 600.
- 212 f) As e) using sets A+B+C+D instead of A+B+C.

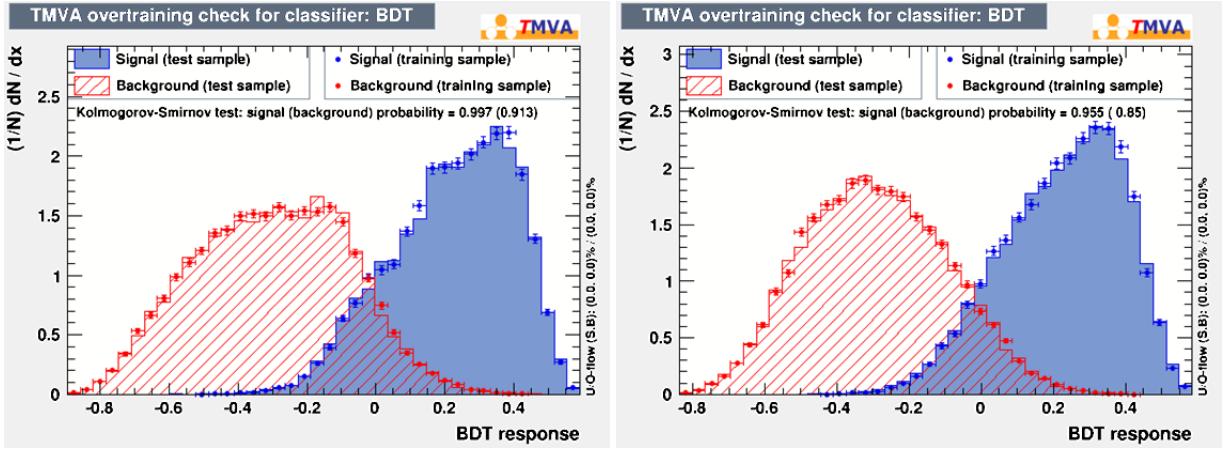


Figure 8: Results of the muon isolation BDT training: superposition of training and test samples for “signal” (non-isolating tracks from generic MC) and “background” (isolating tracks from $B_s \rightarrow \mu\mu$ MC) for the set of variables A (left) and A+B+C (right).

- 213 g) As f) changing the BDT parameter `maxdepth` from 3 to 4.
- 214 h) As g) changing the BDT parameter `adaboostbeta` from 0.50 to 0.75.

215 The isoBDT performance on MC increases as more input variables are used in the
 216 training. This is consistent with the fact that all variables in sets A-B-C-D show some
 217 discrimination power (not 100% correlated), as shown in Figs. 3-7.

218 The next step is to perform the same kind of comparison when the ROC is built using
 219 the right data sidebands instead of the generic MC. The results are summarized in Fig. 11
 220 and show two important points. First, the better performance of the isoBDT compared
 221 to the cut-based isolation is evident also in this case. However, it worsens when set B
 222 is included in the isoBDT training and it does not improve even when set D is added.
 223 The degradation may be explained with the poor data-MC agreement of the `trk_chi` and
 224 `trk_gho` variables composing set B, as shown in Fig. 12⁴. In principle the distributions
 225 in generic MC and right data sidebands should be quite similar, but in fact they are
 226 significantly different. The agreement is poor also for the `ProbNNpi` variable (and in
 227 general for the other ProbNN quantities of set D) and this might explain why the ROC
 228 does not improve when set D is included. It is worth to note that the agreement is not
 229 very good also for `trk_ips`. A preliminary attempt has been made to calibrate the `trk_chi`
 230 and `trk_gho` distributions of isolating and non-isolating tracks using ad hoc data control
 231 samples. However, the results have not been satisfactory. More investigation is required,
 232 both to better understand the dependence of the isoBDT performance in data as a function
 233 of the input variables, and to optimize the performance in data using control samples.

⁴Note that in contrast to what has been shown in Figs. 3-7, these plots compare the distributions of the *sum* of isolating and non-isolating longtracks. The concept of isolating/non-isolating is intrinsically related to the MC truth information and hence this separation cannot be realized in real data.

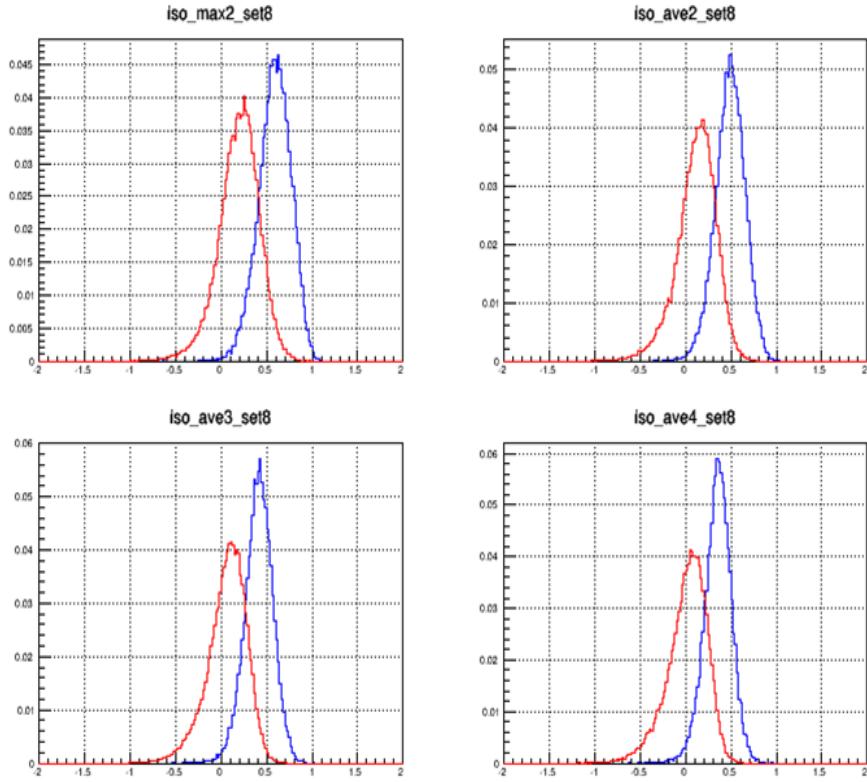


Figure 9: Signal (red) and background (blue) distributions of four versions of the isoBDT trained on the set of variables A (Sect. 3.2.1). The four versions are defined in Sect. 3.2.3: iso_max2 (top-left), iso_ave2 (top-right), iso_ave3 (bottom-left) and iso_ave4 (bottom-right). .

234 3.4 Performance comparison: isoBDT used in global BDT

235 In this section we compare the performance of the global BDT used in the 2013 mea-
236 surement [4] as a function of the muon isolation and of the training sample. Ten muon
237 isolations are considered:

- 238 ● 2 cut-based isolations: sum_isolation (aka *iso5*) and sum_isolation_new1 (aka *iso1*),
239 defined in Table 2
- 240 ● 8 isoBDT isolations: iso_max2 and iso_ave2, each with 4 different sets of variables:
 - 241 – *set1*: set A [the 6 input variables used in sum_isolation]
 - 242 – *set3*⁵: set A+B+C+D [expected to be the best in MC]
 - 243 – *set4*: set A+C [expected to be the best in data]
 - 244 – *set5*: set A+C+D [as set4 + PID]

⁵The reason why *set2* is missing is simply 'historical'.

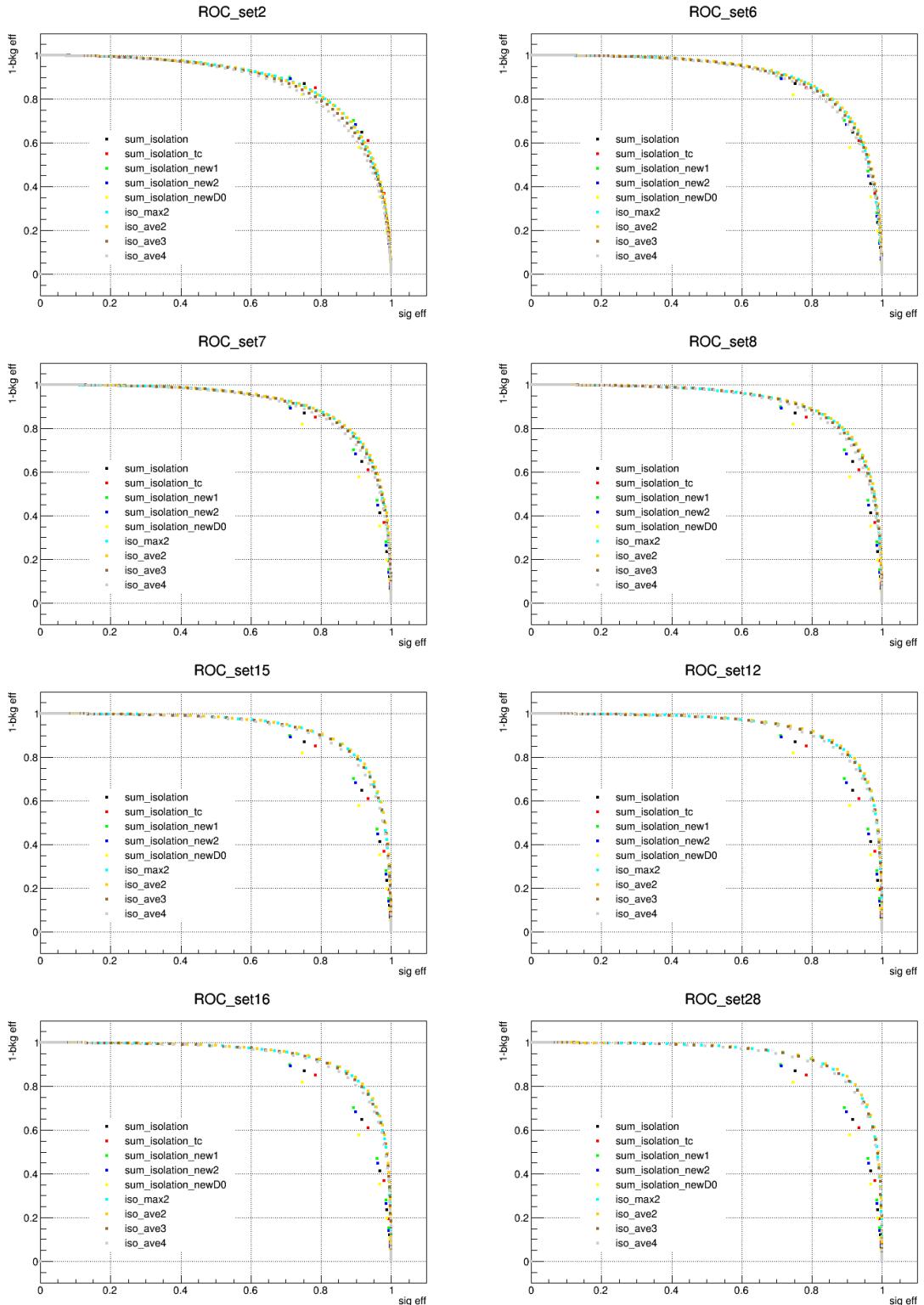


Figure 10: Comparison of the ROCs of the isoBDT and the cut-based isolations, evaluated using signal MC and generic MC. The different versions of isoBDT are discussed in Sect. 3.3 while the cut-based isolations are defined in Table 2

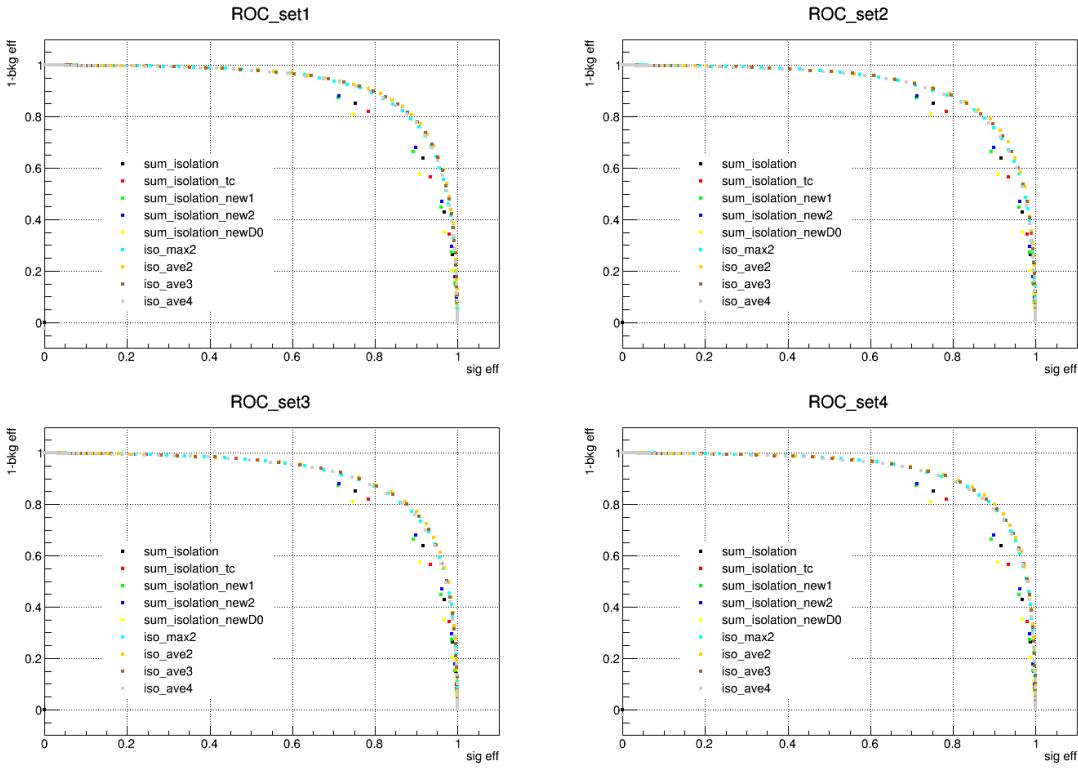


Figure 11: Comparison of the ROCs of the isoBDT and the cut-based isolations evaluated using signal MC and right data sidebands. Set of variables A (top-left), A+B+C (top-right), A+B+C+D (bottom-left), A+C+D (bottom-right). The performance on data decreases when set B is used, as discussed in Sect. 3.3.

The datasets used to evaluate the performance are listed below. All numbers are given after having applied the selection used in the 2013 measurement. The mass acceptance region is $m(\mu\mu) \in [4.9, 6.0]$ GeV. The data sidebands are defined as the mass regions outside the blind signal window $m(\mu\mu) \in [m(B_d) - 80 \text{ MeV}, m(B_s) + 80 \text{ MeV}]$.

- $B_s \rightarrow \mu\mu$ MC: 3 samples of 70×10^3 events each
- generic MC:⁶
 - 71×10^3 (low p_T product) + 400 (high p_T product) events used for training
 - 28×10^3 (low p_T product) + 400 (high p_T product) events used for test
- data sidebands: 42×10^3 events, corresponding to 3 fb^{-1}

⁶“low p_T product” indicates the requirement: $p_T(\mu^+)p_T(\mu^-) < (4 \text{ GeV})^2$. “high p_T product” indicates $p_T(\mu^+)p_T(\mu^-) > (4 \text{ GeV})^2$.

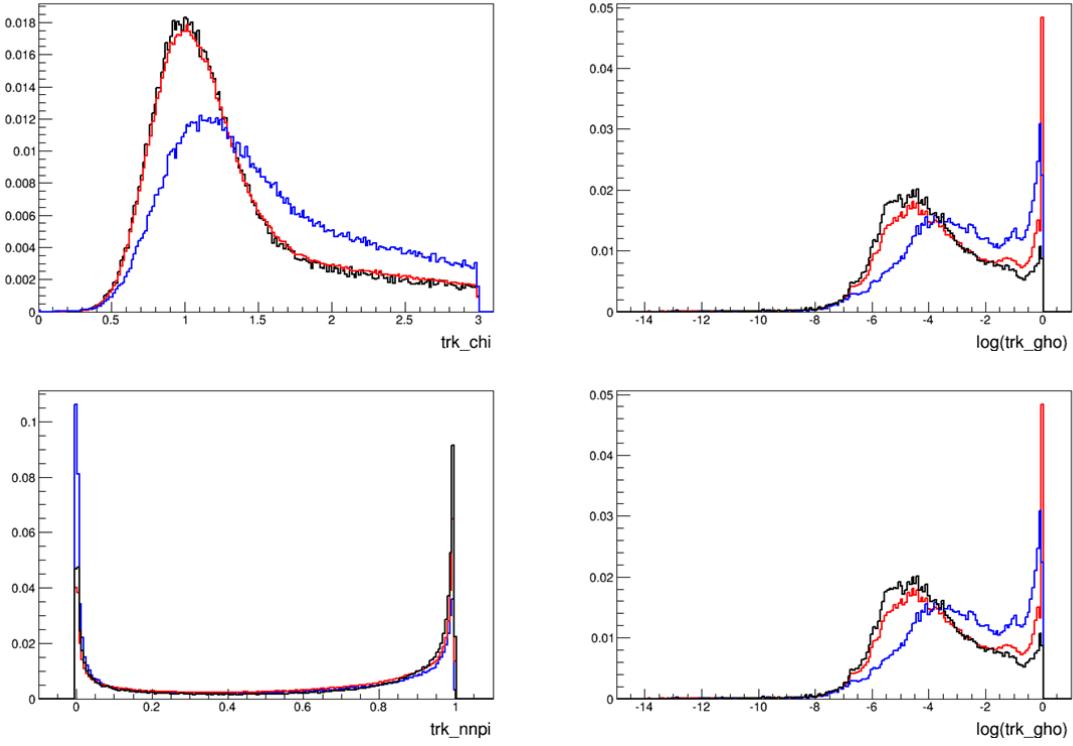


Figure 12: Comparison of the trk_chi (top-left), $\log(\text{trk_gho})$ (top-right), trk_nnpi (bottom-left) and $\log(\text{trk_gho})$ (bottom-right) distributions in generic MC (red), right data sidebands (blue) and signal MC (black). The distributions of data right sidebands and generic MC should be similar but in fact they are significantly different.

The results are summarized in Figs. 13 and 14. In each figure the plots in the top, middle and bottom rows show the ROC in the full signal efficiency range, in the region $\epsilon_{\text{sig}} \in [0.59, 0.75]$ ⁷ and in the most sensitive region, respectively.

Let us first consider the ROCs evaluated on the right data sidebands with BDT12 trained using generic MC, shown in the left column of Fig. 13. From the middle plot one can see that in the 60-70% signal efficiency region the best performance is achieved by the set4 isoBDT configuration, as expected, achieving a 35-40% background rejection increase compared to the cut-based isolation. The 'max2' and 'ave2' criteria give similar results, with a slight preference for the former. However, the situation is less clear in the most sensitive BDT region with ($\epsilon_{\text{sig}} < 0.5$), where it is not possible to identify a configuration which is significantly better than the others.

Similar conclusions are drawn when the BDT is trained using 1/3 of the right data sidebands and the ROC is evaluated on the right data sidebands (Fig. 13, right column) or

⁷The notation $\epsilon_{\text{sig}} \in [0.59, 0.75]$ used here and elsewhere is a compact way to indicate the region $0.25 < \text{BDT12}_{\text{flat}} < 0.41$, where $\text{BDT12}_{\text{flat}}$ is the output of the global BDT defined to be flat between 0 and 1 for $B \rightarrow \mu\mu$ signal and strongly peaking at 0 for combinatorial background.

267 when the BDT is trained using 5/11 of the right data sidebands and the ROC is evaluated
268 on the left+right data sidebands (Fig. 14, left column). Additional comparisons, with
269 similar results, can be found in ref. [8].

270 When the generic MC is used to both train the BDT and evaluate the ROC (Fig. 14,
271 right column) the best performance is given by the set3 isoBDT configuration, as expected
272 from the discussion in Sect. 3.3. The amount of background in the most sensitive BDT
273 region is significantly smaller than in previous cases, but again the different isolations are
274 statistically compatible.

275 3.5 Prospects

276 The results discussed in Sect. 3.4 indicate that the isoBDT performs significantly better
277 than the cut-based muon isolation used in the past $B \rightarrow \mu\mu$ LHCb measurement. In
278 the intermediate BDT region, corresponding to $\epsilon_{sig} \in [0.59, 0.75]$, a background rejection
279 increase of 35-40% is observed. On the other hand, a similar performance is not observed
280 in the most sensitive BDT region. A detailed study of the background properties in this
281 region is planned.

282 At least two other isolation strategies are currently under investigation, one based
283 on the vertex information between the tracks of the event [9], the other exploiting the
284 reconstruction of jets [10]. Eventually the different approaches will have to be combined
285 or merged together to maximize the analysis sensitivity.

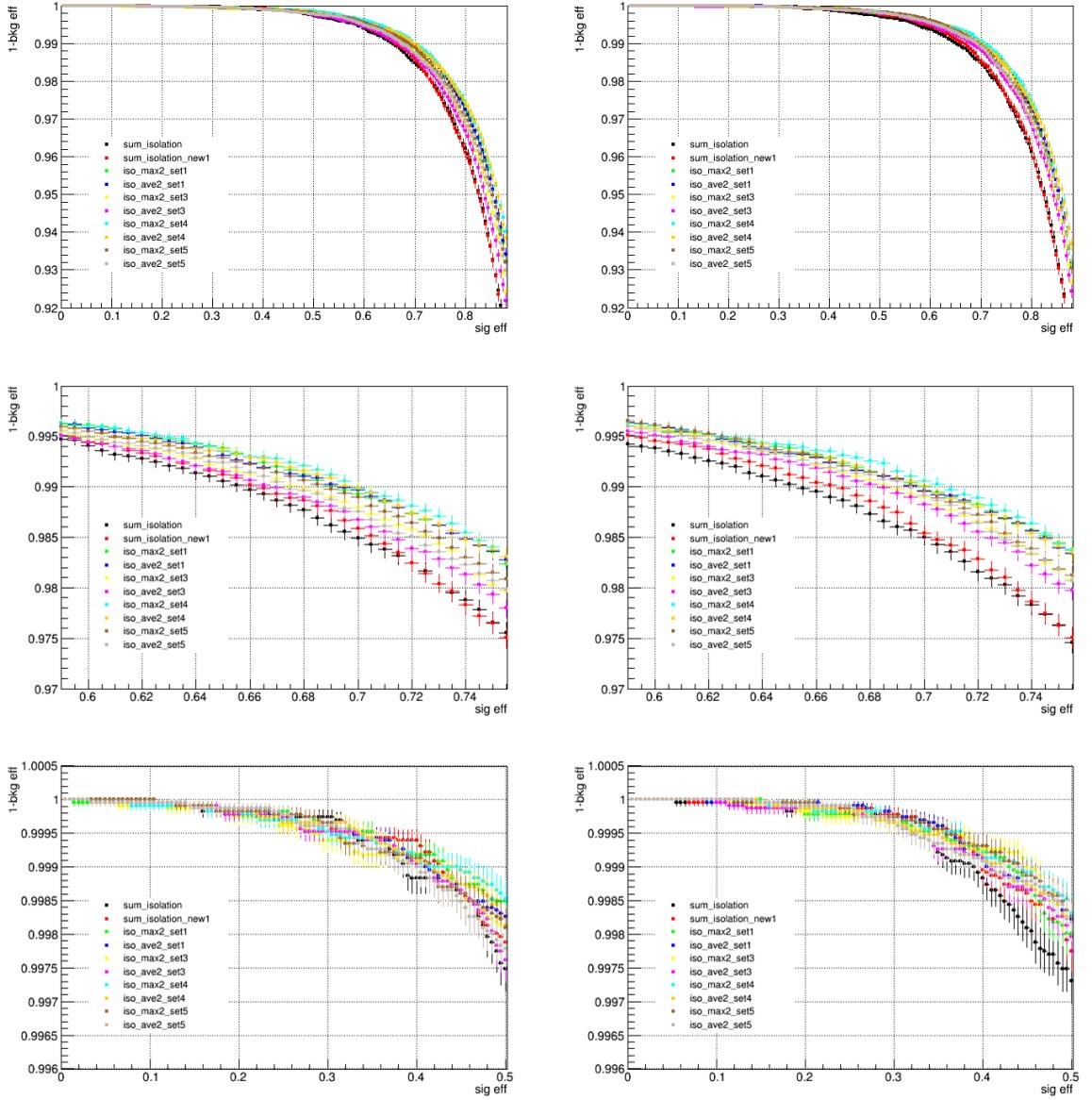


Figure 13: ROC of BDT12 with different muon isolations. Left column: BDT12 trained with generic MC and ROC evaluated on the right sidebands. Right column: BDT12 trained with 1/3 of the right data sidebands and ROC evaluated on the right data sidebands. Top row: full range; middle row: zoom in $\epsilon_{\text{sig}} \in [0.59, 0.65]$; bottom row: zoom in the most sensitive region. Details in Sect. 3.4.

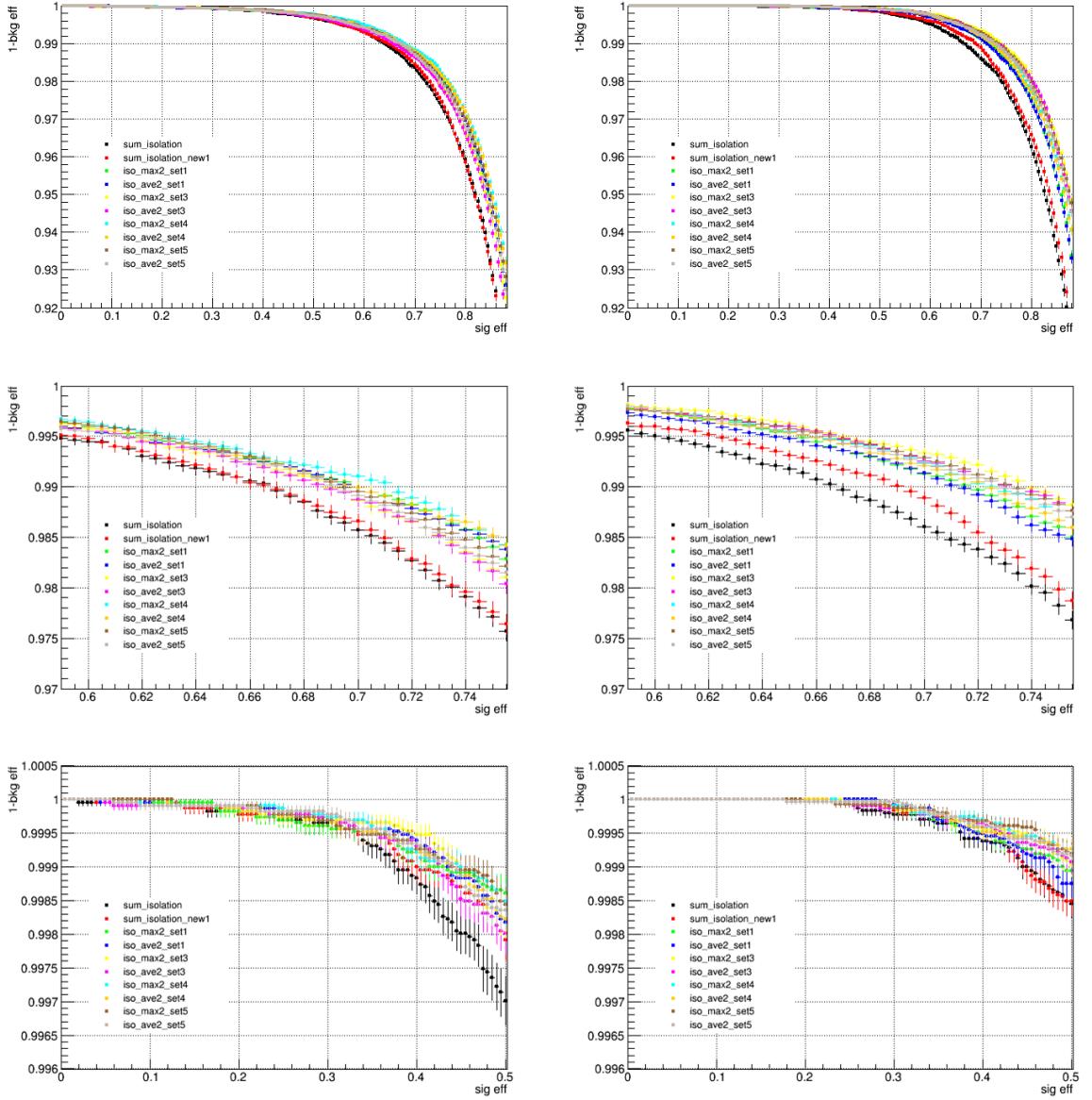


Figure 14: ROC of BDT12 with different muon isolations. Left column: BDT12 trained using 5/11 of the left+right data sidebands and ROC evaluated on the left+right data sidebands. Right column: BDT12 trained on generic MC and ROC evaluated on generic MC. Top row: full range; middle row: zoom in $\epsilon_{sig} \in [0.59, 0.65]$; bottom row: zoom in the most sensitive region. Details in Sect. 3.4.

286 4 PID efficiency from data

287 Particle identification performance can be measured from data using the *tag-and-probe*
 288 method on suitable control samples: $J/\psi \rightarrow \mu^+ \mu^-$ from b decays are used for muons (from
 289 `JpsiFromBNOPIDNoMip` stripping line), $D^0 \rightarrow K\pi$ from the $D^* \rightarrow D^0\pi$ decays are used
 290 for pions and kaons (`NoPIDDstarWithD02RSKPi` stripping line), and finally $\Lambda \rightarrow p\pi$ decays
 291 for protons (`LamOLLLine1V0ForPID` stripping line).

292 In particular for the $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ analysis we need to evaluate the muon ID efficiency,
 293 $\epsilon(\mu \rightarrow \mu)$, and the probability that a hadron is misidentified as a muon: $\epsilon(\pi \rightarrow \mu)$,
 294 $\epsilon(K \rightarrow \mu)$, and $\epsilon(p \rightarrow \mu)$.

295 The control samples are selected without using any PID information, at least for the
 296 probe track of each decay. To cancel out possible biases arising from the trigger, special
 297 attention is put in making the probe track *unbiased* with respect to the trigger itself.

298 Selected control samples contain also a fraction of background events that have to be
 299 subtracted to measure the PID efficiency. The subtraction is done applying the *fit-and-*
 300 *count* method or the *sPlot* technique [11], both implying a fit to the relevant invariant
 301 mass distributions. The latter is used in the *PIDCalib* package [12] and will not be further
 302 described here. The former can be implemented in different ways, depending on the
 303 method used to evaluated the $\epsilon(x \rightarrow \mu)$ efficiency and on the technique applied to subtract
 304 the background.

305 In a first method, the total PID efficiency, ϵ_{S+B} , measured in a given mass window
 306 around the peak, is related to the actual efficiency, ϵ_S , by the relation: $\epsilon_{S+B} = \frac{S}{S+B}\epsilon_S +$
 307 $\frac{B}{S+B}\epsilon_B$, where S and B are the number of signal and background events in the window
 308 around the appropriate invariant mass peak. Solving for ϵ_S the above relation becomes
 309 $\epsilon_S = \frac{B+S}{S} \left(\epsilon_{S+B} - \frac{B}{S+B}\epsilon_B \right)$, indicated as ϵ_α in the following.

310 In an alternative method the control sample events are splitted into two subsamples
 311 depending on whether the probe track satisfies, $N1$, or not satisfies, $N0$, the requested
 312 PID selection (i.e. `isMuon==1`), then the efficiency is obtained via the relation $\epsilon_S = \frac{S1}{S1+S0}$,
 313 indicated as ϵ_β in the following, where $S1$ and $S0$ are the number of signal events in the
 314 $N1$ and $N0$ samps, respectively.

315 S and B in the first method, and $S1$ and $S0$ in the second, can be evaluated with
 316 two different approaches: a full signal plus background fit to the whole invariant mass
 317 spectrum, or by subtracting from the number of events in a suitable signal mass window
 318 the background events in the same region extrapolated from a fit to the mass sidebands.

319 All this translates into four possible estimates; different choices have been done along
 320 the years for each $\epsilon(x \rightarrow \mu)$ evaluation that will be described in the following sections.

321 Using ϵ_α or ϵ_β method is equivalent in case of good quality fits. The agreement between
 322 the two methods has been quantitatively checked for the muon data fit [13] where the
 323 difference was $\sim 0.5\%$. In case of fit imperfections, these cancel out at the first level in ϵ_β
 324 method, while the ϵ_α method is less protected against a bad quality fit. A second issue
 325 with the ϵ_α method can arise from the assumption that the background efficiency under
 326 the mass peak is the same as the one in the sidebands.

327 In this work of analysis optimization, we choose as the best option for all the PID

328 performance evaluations the method ϵ_β where $S1$ and $S0$ are determined from a full
 329 signal plus background fit to the separated m_{N1} and m_{N0} invariant mass distributions,
 330 respectively.

331 The kinematic distribution of the probe track in the control samples is in principle
 332 different from the one belonging to the decay channels of interest ($B_{d,s}^0 \rightarrow \mu^+ \mu^-$, $B_{(s)}^0 \rightarrow$
 333 $h^+ h^{(\prime)-}$, $\Lambda_b^0 \rightarrow p \mu^- \nu$, $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$, $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$, and $B^{0(+)} \rightarrow \pi^{0(+)} \mu^+ \mu^-$). To take
 334 this into account, each $\epsilon(x \rightarrow \mu)$ is evaluated in 11 bins in momentum⁸, p , and 4 bins in
 335 transverse momentum⁹, p_T , of the probe track, and the correction to each decay channel
 336 is estimated convoluting event-by-event the measured $\epsilon(x \rightarrow \mu)$ curves with the p and p_T
 337 spectra of the relevant tracks of the selected MC samples.

338 4.1 Muons

339 A detailed study of $\epsilon(\mu \rightarrow \mu)$ performance evaluation has been done using the first 300
 340 pb^{-1} of data acquired in 2011, and it is documented elsewhere [13]. The muonID used
 341 in the latest $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ results [3, 14, 15] has been calculated applying the ϵ_β method
 342 and by a linear fit of the background on the sidebands of the $J/\psi \rightarrow \mu^+ \mu^-$ peak of the
 343 separated m_{N1} and m_{N0} invariant mass distributions. During the last year, we continue
 344 to apply the ϵ_β method but we moved to the signal plus background full fit of the same
 345 m_{N1} and m_{N0} invariant mass distributions, by fitting with a Crystal Ball the signal and
 346 with an exponential function the background; a typical fit is shown in Fig. 15. Although
 347 results of the two approaches are in agreement within the errors [13], the latter is more
 348 robust in case where a very few background events are present in the sidebands.

349 The unbias with respect to the trigger has been implemented requiring the probe muon
 350 to be independent of the trigger (TIS) at all level, specifically to satisfy the

```
L0Global_TIS && Hlt1Phys_TIS && Hlt2Phys_TIS
```

351 condition. Alternatively, to size the systematic effect possibly introduced by this choice,
 352 the tag muon can be forced to satisfy the trigger by itself (TOS), specifically we required
 353 to the track

```
L0MuonDec_TOS &&  

  354 (Hlt1SingleMuonHighPTDec_TOS || Hlt1TrackMuonDec_TOS || Hlt1TrackAllL0Dec_TOS)  

  355 && (Hlt2SingleMuonDec_TOS || Hlt2SingleMuonHighPTDec_TOS).
```

356 For the muonID, the agreement between our fits-and-count results and the ones from
 357 PIDCalib has always been good within the errors. The new muon data sets in preparation
 358 in PIDCalib [12] will implement our choices for both the fit model and the trigger unbias
 359 requests.

⁸ $p = (0,5), (5,10), (10,15), (15,20), (20,25), (25,30), (30,35), (35,40), (40,50), (50,60)$, and >60 GeV
⁹ $p_T = (0.8,1.7), (1.7,3), (3,5)$, and $(5,10)$ GeV

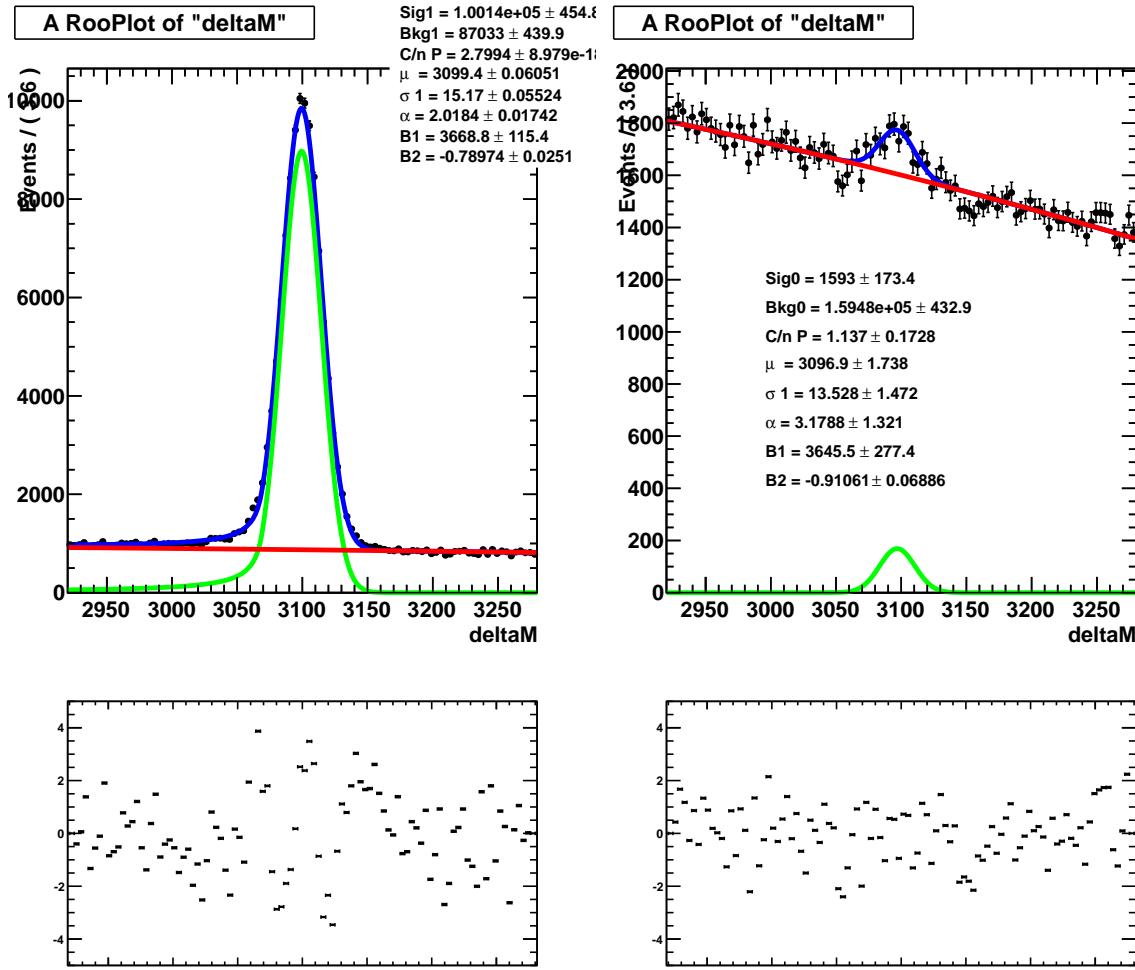


Figure 15: Example of muon fits for events in $p > 60$ GeV and $p_T = (3,5)$ GeV satisfying the $isMuon == 1$ PID selection, in the left panel, and $isMuon == 0$, in the right one (deltam label is a typo, while the fitted distributions are correctly the $\mu\mu$ invariant mass).

360 4.2 Pions and kaons

361 To estimate the $\epsilon(\pi \rightarrow \mu)$ and $\epsilon(K \rightarrow \mu)$ probabilities, many choices have been done along
 362 the years, which will not be described here. We start this section in summarizing what
 363 has been done for the last published result [3, 4]. The ϵ_α method was used to evaluate
 364 the misID probability, and the D^0 invariant mass distribution of $D^0 \rightarrow K^- \pi^+$ from the
 365 $D^{*\pm} \rightarrow D^0 \pi^\pm$ decays was fitted with the sum of a Crystal Ball and a Gaussian function
 366 (with the mean constrained to the same value) and the background was fitted with a
 367 polynomial.

368 To clean the sample after the stripping selection, a cut on the difference between D^{*+}
 369 and D^0 masses, δM , has been applied: $144.5 < \delta M < 146.5$ MeV. The difference between

misID probability evaluated applying or not applying the δM cut was used to assess the systematic error. Averaged over the four $B_{(s)}^0 \rightarrow h^+h^-$ modes, the quoted systematic error was 4% (5%) per particle for 2012 (2011) data.

In addition, the tagging track of the D^0 decays was required to be positively identified with a cut on the $DLL(K-\pi)$: the kaon (pion) is required to pass the cut $DLL(K-\pi) > 10$ ($DLL(K-\pi) < 0$) when the $\epsilon(\pi \rightarrow \mu)$ ($\epsilon(K \rightarrow \mu)$) misID probability is measured.

Finally, to cancel out possible biases arising from the trigger, the probe track (kaon or pion) was required to be TIS with respect to 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.

In this work, we critically reviewed all the assumptions made in the past. As a figure of merit we chose the variation of the $\epsilon(\pi \rightarrow \mu)$ or $\epsilon(K \rightarrow \mu)$ when averaged over the $B_{(s)}^0 \rightarrow h^+h^-$ spectra.

- **Method.** Applying the ϵ_β method to the same D^0 invariant mass distribution, increases the average $\epsilon(\pi \rightarrow \mu)$ pion misID of 23% relatively to what we evaluated with the ϵ_α method; for the average $\epsilon(K \rightarrow \mu)$ kaon misID, the increase is 10%. Such a difference when the two methods are applied, is mainly due to the different background composition in the sidebands and below the D^0 peak, which introduces a bias when the ϵ_α method is used.
- **δM cut.** Repeating with ϵ_β method the evaluation of the systematic check related to the δM , we find -6% variation for pions and 6% for kaons, which translate in a $O(1\%)$ difference for the average over the four $B_{(s)}^0 \rightarrow h^+h^-$ modes.
- **Stripping selection.** The selection applied in `NoPIDstarWithD02RSKPi` stripping line, includes different “mass vetoes” to remove background events from $D^0 \rightarrow KK$, $D^0 \rightarrow \pi\pi$, and $D^0 \rightarrow K^+\pi^-$ decays. These vetoes introduce distortions in one or both the D^0 sidebands, depending on the (p,p_T) range of the probe kaon/pion, spoiling the fit quality. To size the effect of this “sculpture”, we compare results from the whole mass range fit and from the (1823,1900) MeV mass range fit (not affected by the Stripping selection) when the ϵ_β method is applied; the relative change is -0.6% for pions and of -4% for kaons.
- **PID selection of the tag.** To reduce at minimum possible correlations introduced by the RICH and to gain statistics for the control sample, we decided to remove the PID selection applied to the tag track. This implies a 2.8% change for the pion misID and 1.2% for the kaons (when the ϵ_β method is used).
- **Trigger unbias.** Finally, we checked for the effect of applying a trigger unbias also at Hlt2 level. Requiring the probe track to satisfy `L0Global_TIS && Hlt1Phys_TIS && Hlt2Phys_TIS` instead of `L0Global_TIS && Hlt1Phys_TIS` only, the relative change was -0.8% for pion misID and 1.4% for the kaon one (when the ϵ_β method is used). This tiny difference allows to measure the misID using the “Hlt1 sample”

sample, ~ 10 times more abundant than the “Hlt2 sample”, a fundamental bonus since we need to apply harder PID cuts with respect to the past. An $O(\%)$ systematic error can be determined by the comparison of misID as measured from “Hlt1” and “Hlt2” samples.

From the above it seems that to apply the ϵ_β method, using fits to the D^0 invariant mass distributions to evaluated the number of signal events could be a good choice. However, the need of restricting the fit range in the invariant mass region not affected by the mass vetoes [16], makes the D^0 sidebands too small for a reliable and stable fit of the background, while this effect becomes negligible in the $(m_{D^0}, \delta M)$ plane being the δM axis unaffected by the stripping selection. Moreover, a bidimensional fit helps in keeping under control a background component almost flat in D^0 and peaking in δM . For this we move from the 1D fit of the m_{D^0} distribution, to a bidimensional fit of the $(m_{D^0}, \delta M)$ plane.

Description of 2D plane. The two dimensional space of the D^0 invariant mass and the δM variables (see Fig. 16) provides an efficient discrimination between signal and backgrounds. The signal features a narrow peak in both mass distributions and it is fitted using a double Gaussian with the mean constrained to the same value for both the m_{D^0} and the δM axis. A first background component consists of multibody partially reconstructed charm decays, which have a smooth distributions in m_{D^0} and a peaking structure in δM ; this component is fitted with a first order polynomial in m_{D^0} and with a double Gaussian (same mean) in δM . A second background component, which comes from properly-reconstructed D^0 decays associated with a random pion, is indistinguishable from the signal in m_{D^0} , but features a smoothly-growing shape in δM ; it is fitted with a double Gaussian (same mean) in m_{D^0} and a threshold function in δM . Finally, random three track combinations that accidentally meet the selection requirements show a smooth distribution in m_{D^0} , and the same square-root-like shape as before in δM ; it is described by a polynomial in m_{D^0} and a threshold function in δM .

2D fit: implementation and results. To fit the threshold trend, we implemented the *RooThreshold* function [16] $\sqrt{x/m_\pi - 1} e^{-c*(x/m_\pi)}$ which, in any configuration we checked, describes data better than the usual *RooDstD0BG* function. The bidimensional $(m_{D^0}, \delta M)$ distribution in each (p, p_T) bin has been fitted with the four component described above. A typical fit result is shown in Fig. 17 for the pions, and in Fig. 18 for the kaons.

The four components fit has been also implemented as better option in PIDCalib (previously a three components fit was performed).

Including all the effects mentioned above, in [17] the contribution of $B_{(s)}^0 \rightarrow h^+ h^-$ background was underestimated by 35%, being the method applied to determine the efficiency the main responsible of this variation.

The reliability of the present choice of a bidimensional fit to the $(m_{D^0}, \delta M)$ distribution to extract the number of signal events, has been checked comparing the misID probabilities evaluated in that way, with the ones from the D^0 mass fit in the $(1823, 1900)$ MeV mass range (not affected by the stripping selection) and when the $144.5 < \delta M < 146.5$ cut is applied (in both cases the ϵ_β method has been used). The misID probabilities are shown for both choices in Fig. 19 and Fig. 20 for pion and kaon, respectively. Using the misID

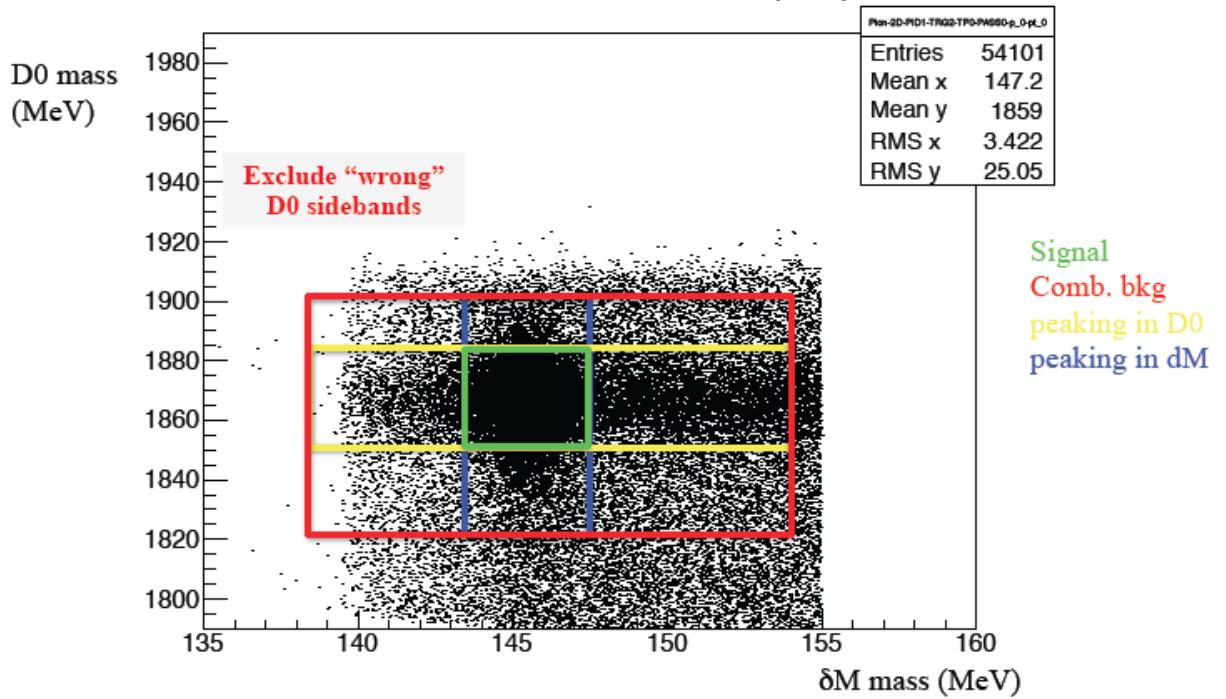


Figure 16: Scatter plot of D^0 vs δM mass distribution. Colors schematically indicate the different components described in the text.

probabilities from the 2D fit introduces a 3% change for both pions and kaons, with respect to the 1D fit, when averaged over the $B_{(s)}^0 \rightarrow h^+h^-$ kinematics.

4.3 Protons

The proton misidentification probability, $\epsilon(p \rightarrow \mu)$, has been measured using a sample of $\Lambda \rightarrow p\pi$ decays. The mass distribution of the $\Lambda \rightarrow p\pi$ candidates was fitted with a double Gaussian function with the mean constrained to the same value, for the signal, and a polynomial function for the background, and the ϵ_a method was used. Due to the low statistics of available samples, the two highest p_T bins were merged in a unique bin (3-10 GeV) and only two p bins were used ($p < 40$ GeV and $p > 40$ GeV). In these bins the number of signal events was 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.

Since the flight distance (FD) of the Λ_b is considerably shorter than the one of Λ , to safely use the misID measured on the Λ control sample to correct the proton from the Λ_b , we checked for a possible misID dependence on the flight distance. To the extent of the available control sample statistics, we found the measured $\epsilon(p \rightarrow \mu)$ to be independent from the Λ decay point.

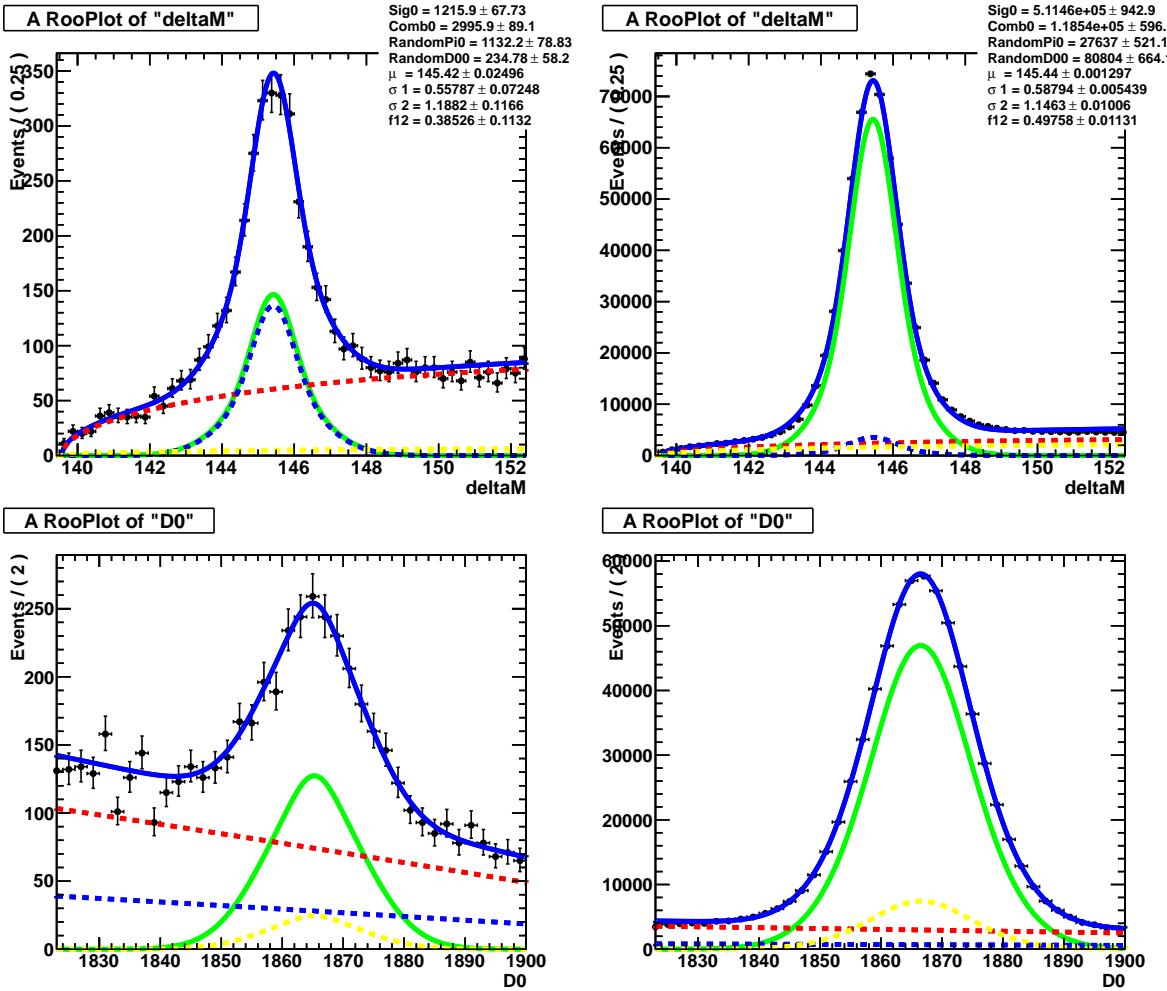


Figure 17: Example of pion fits for events in $p > 60$ GeV and $p_T = (3,5)$ GeV for the $isMuon == 1$ PID selection, in the left panels and $isMuon == 0$ in the right ones. Upper panels show the result of the 2D ($m_{D^0}, \delta M$) fit projected on δM axis; right panels show the projection on to m_{D^0} axis.

In the analysis optimization, we decided to use the ϵ_β method also for the protons, which is more robust when a few background events are present as in many proton fits. Besides this, we started also to use a new proton sample made available in the meanwhile¹⁰, which allows a better kinematic coverage at high p . A typical fit of the $\Lambda \rightarrow p\pi$ peak is shown in Fig. 21.

In the present work, we found that not including the Hlt2 unbias implies a $\epsilon(p \rightarrow \mu)$ misID overestimate. Quantitatively the average proton misID integrated over the Λ spectra is 1.474% when applying L0Global_TIS && Hlt1Phys_TIS trigger unbias, while becomes

¹⁰The Stripping LamOLLLine2V0ForPID line, which select protons with $p > 40$ GeV and applies a prescale factor equal to 1, instead of the 0.14 applied in the StrippingLamOLLLine1V0ForPID line

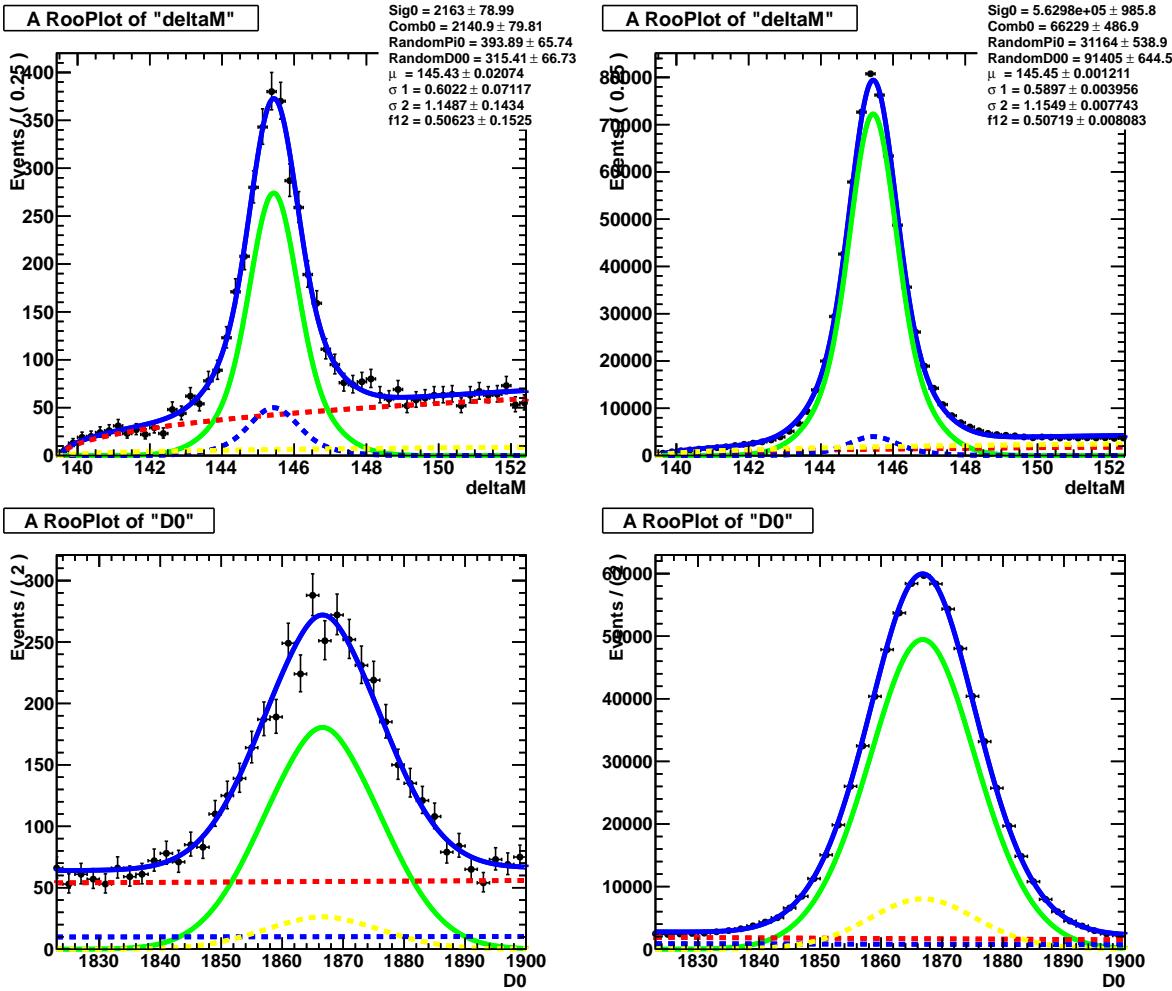


Figure 18: Example of kaon fits for events in $p > 60$ GeV and $p_T = (3,5)$ GeV for the $isMuon == 1$ PID selection, in the left panels and $isMuon == 0$ in the right ones. Upper panels show the result of the 2D ($m_{D^0}, \delta M$) fit projected on δM axis; right panels show the projection on to m_{D^0} axis.

476 1.177% applying the L0Global_TIS && Hlt1Phys_TIS && Hlt2Phys_TIS request.

477 Investigating the origin of this difference, three Hlt2 lines, largely uncorrelated, were
 478 found relevant: Topo (~41%), Charm (~38%), and SingleMuon (~15%). While the Topo
 479 and SingleMuon lines slightly modify the proton misID, surprisingly the Charm line has a
 480 large impact on proton misID: asking Hlt2Charm_Dec to be 0 or 1, the misID integrated
 481 over the Λ spectrum changes from 1.70% to 1.12%. The same numbers for the other two
 482 lines are 1.50% and 1.39% when asking Hlt2Topo_Dec to be 0 or 1, respectively, and 1.50%
 483 and 1.49% when selecting Hlt2SingleMu_Dec equals 0 or 1, respectively. While it is not
 484 completely understood the correlation between the Hlt2 line selection and the $\epsilon(p \rightarrow \mu)$
 485 misID probability increase, we conservatively choose to ask for the trigger unbias also at

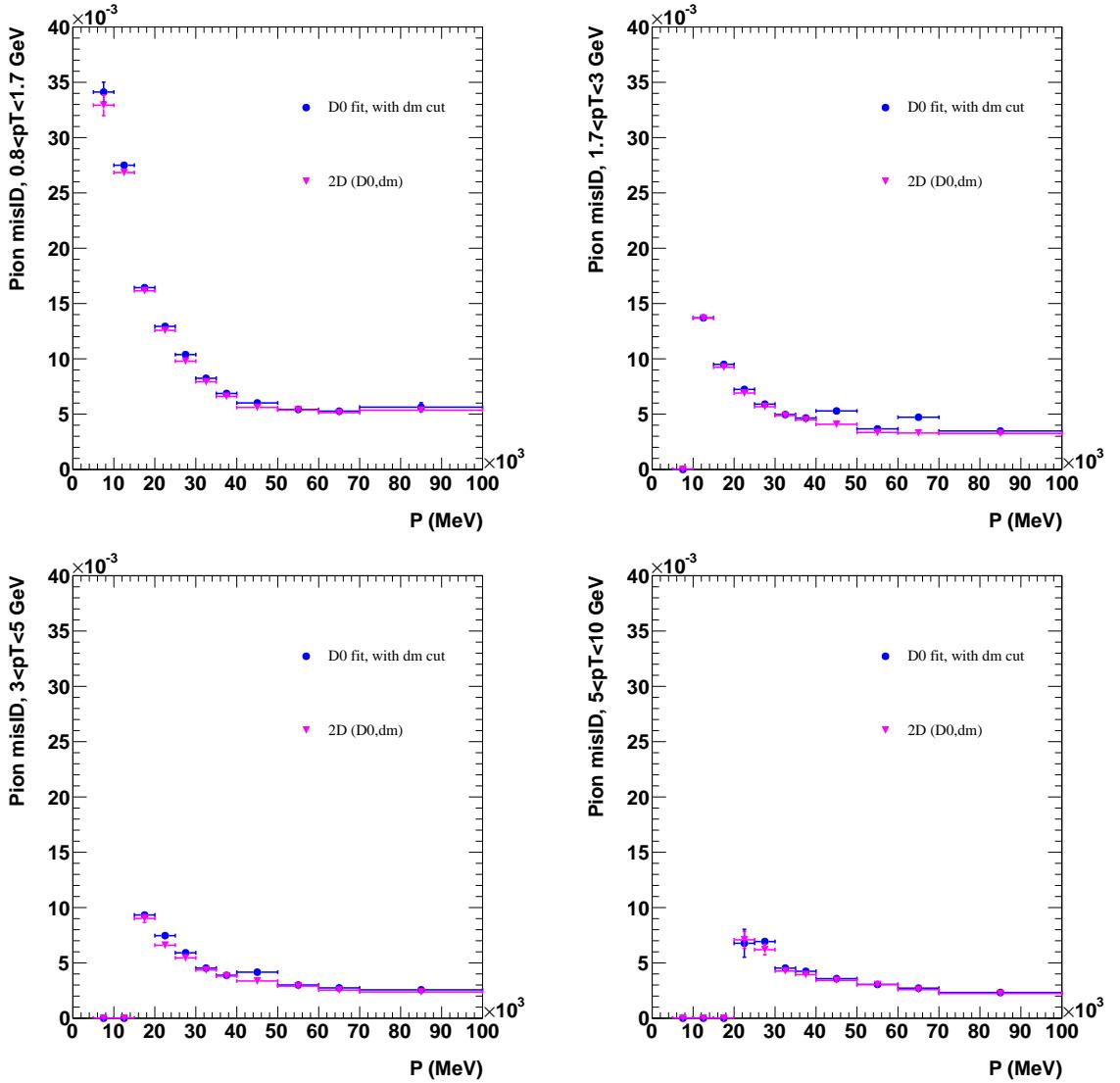


Figure 19: Comparison of the pion misID probability vs momentum for the four p_T bins of the probe track, evaluated with a fit to the D^0 mass distribution or with a four components fit of the $(m_{D^0}, \delta M)$ plan.

486 the Hlt2 level.

487 For what concerns the impact of Λ_b background to $B_{d,s}^0 \rightarrow \mu^+ \mu^-$ measurement, when
488 the misID evaluated including the Hlt2 unbias is convoluted with the (p, p_T) spectrum of
489 the protons from $\Lambda_b^0 \rightarrow p \mu^- \nu$, we found that in the past we overestimated by 25% the
490 $\Lambda_b^0 \rightarrow p \mu^- \nu$ contribution itself.

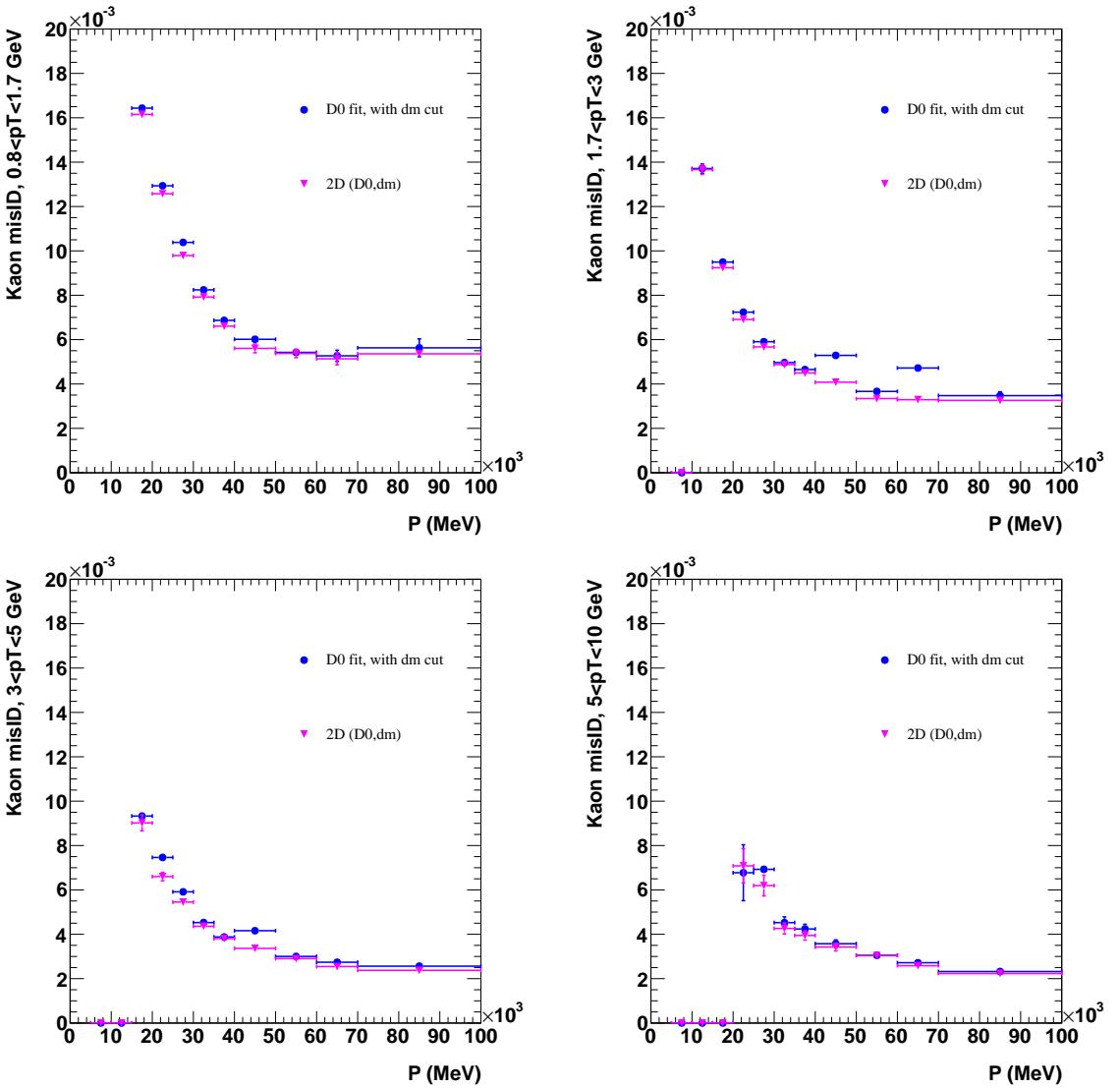


Figure 20: Comparison of the kaon misID probability vs momentum for the four p_T bins of the probe track, evaluated with a fit to the D^0 mass distribution or with a four components fit of the $(m_{D^0}, \delta M)$ plan.

4.4 Summary on PID efficiency evaluation

In this work the evaluation of PID performance from data control samples has been critically reviewed, specifically for what concerns the $\epsilon(x \rightarrow \mu)$ efficiency. Most of our findings have been included in *PIDCalib* [12] and are widely available through it. For the muons, the main novelty is the inclusion of two trigger unbias strategies. For pions and kaons, the possibility of applying a trigger unbias only at Hlt1 level gives ten times more statistics, important to pursue systematics studies; also, the bidimensional four

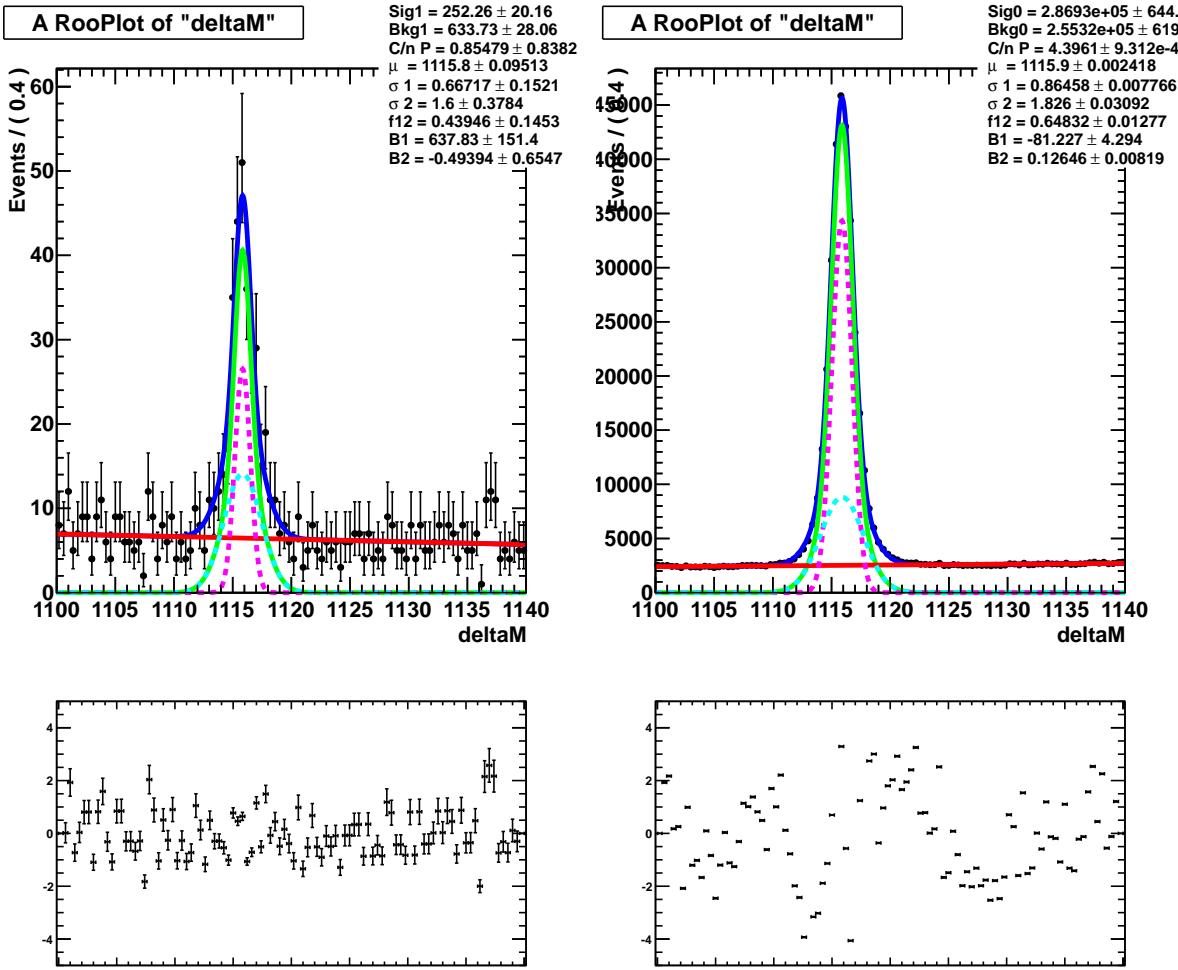


Figure 21: Example of proton fits for events in $p = (20,25)$ GeV and $p_T = (1.7,3)$ GeV for the $isMuon == 1$ PID selection, in the left panel and $isMuon == 0$ in the right one (deltam label is a typo and the fitted distribution is correctly the $\pi\pi$ invariant mass).

498 components fit to the $(m_{D^0}, \delta M)$ plane, is the best option to be used. Finally for the
 499 protons, differently from pions and kaons, we confirm instead the need of a trigger unbias
 500 at Hlt2 level, since the Hlt2 selections apparently introduce appreciable biases in the misID
 501 evalutation.

5 PID selection and exclusive background estimate

As discussed in section 2, the B^0 analysis performance depends critically on the ability to reject the $B_{(s)}^0 \rightarrow h^+h^{(\prime)-}$ events under the signal peak; it can also benefit from a better rejection of the semileptonic modes polluting the sidebands ($B^0 \rightarrow \pi^-\mu^+\nu_\mu$, $B_s^0 \rightarrow K^-\mu^+\nu_\mu$, $B^{0(+)} \rightarrow \pi^{0(+)}\mu^+\mu^-$ and $\Lambda_b^0 \rightarrow p\mu^-\nu$), since in this region there's a strong correlation with the combinatorial background. In addition, the $\Lambda_b^0 \rightarrow p\mu^-\nu$ decay, even if suppressed with respect to $B_{(s)}^0 \rightarrow h^+h^{(\prime)-}$, is also leaking into the signal region.

In the last analysis version [4], the PID selection after `isMuon` consisted of requiring to both tracks $DLL(K - \pi) < 10$ and $DLL(\mu - \pi) > -5$, which ensured enough rejection on $B_{(s)}^0 \rightarrow h^+h^{(\prime)-}$ (a factor of ~ 5 with respect to `isMuon` only) while keeping a very high B_s^0 signal efficiency, ~ 0.95 . For B^0 , instead, tighter selection strategies could improve the final analysis sensitivity.

We'll discuss here the results obtained by using the `NNpid` [18] variables combined as

$$ProbNN = ProbNN_\mu \times (1 - ProbNN_K) \times (1 - ProbNN_p),$$

with different thresholds: $ProbNN > 0.1, 0.2, 0.3, 0.4, 0.5$, referred to as $ProbNN_{1,2,3,4,5}$, respectively.

Besides this, a slightly different version of the $ProbNN$ operator has been tried, by dropping the rightmost proton term in the previous definition:

$$ProbNN_{nop} = ProbNN_\mu \times (1 - ProbNN_K),$$

with the same threshold definitions. The $ProbNN_{nop}$ operator is expected to have very similar performances on all of the exclusive backgrounds, but for a sensible increase of the $\Lambda_b^0 \rightarrow p\mu^-\nu$ background, and a slightly higher signal efficiency.

In Fig. 22 the yields for $B_{(s)}^0 \rightarrow h^+h^{(\prime)-}$ and $\Lambda_b^0 \rightarrow p\mu^-\nu$ backgrounds as a function of the signal efficiencies for all of the above PID selections are shown, together with the results of the DLL selection. The same information is listed in Tab. 3 for all background sources, for DLL and $ProbNN_{2,3,4}$ selections only. The yields are evaluated in the mass range [4900-6000] MeV/c² and in the full BDT range. This has to be considered carefully when comparing among each other the different background sources: the semileptonic B decays, for instance, contribute only to the left mass sideband; on the contrary, the $B_{(s)}^0 \rightarrow h^+h^{(\prime)-}$ is peaking in the signal region and has a signal-like BDT distribution, as will be discussed later in this section.

The normalization factors for the semileptonic modes, and the number of TIS $B_{(s)}^0 \rightarrow h^+h^-$ events are taken from Ref [4]; the PID selection efficiency has been recomputed in each case by convoluting the pion, kaon and proton misID probabilities discussed in Sec. 4 with the proper MC spectra, evaluated after all selection cuts but the trigger and the PID (including `isMuon`) for the hadrons; finally, the trigger efficiencies have been computed by MC after the whole selection chain is applied, including PID requests, as in Ref. [4].

A few comments are in order:

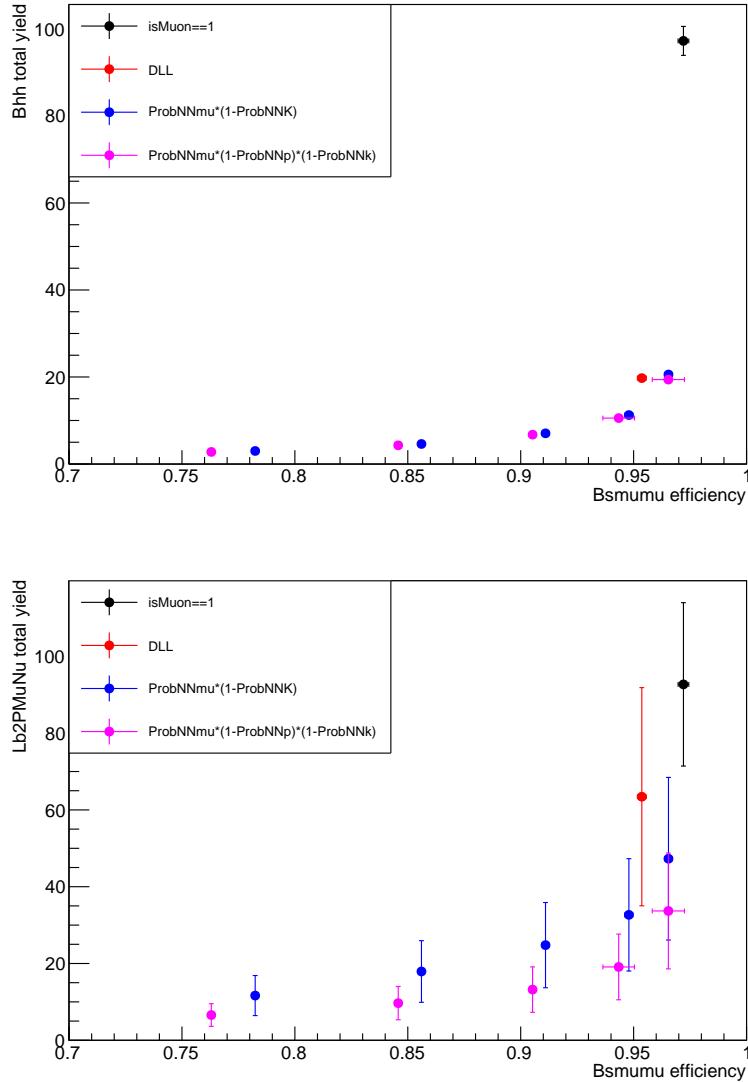


Figure 22: Yields (for 3 fb^{-1}) as a function of signal efficiencies for $B_{(s)}^0 \rightarrow h^+ h^{(-)}$ (left) and $\Lambda_b^0 \rightarrow p \mu^- \nu$ (right): results from $ProbNN_{1,2,3,4,5}$ (purple points) and from $ProbNN_{nop1,2,3,4,5}$ (blue points) selections are shown from right to left in the horizontal axis; the results with $isMuon$ only (black point) and after the DLL selection (red point) are also shown.

- 535 ● When comparing to DLL, the $ProbNN_4$ selection improves by a factor of ~ 5 on
 536 $B_{(s)}^0 \rightarrow h^+ h^{(-)}$, with a moderate signal loss, $\sim 12\%$. The overall sensitivity gain will
 537 depend of course also on the BDT shape of the events, which will be discussed later.
 538 ● For $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$ and $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$, the improvement for $ProbNN_4$ reduces to
 539 a factor of ~ 2 and ~ 3 , respectively, since only one hadron is present in the final

Table 3: Background event yields estimated for 3 fb^{-1} in the mass range [4900-6000] MeV/ c^2 , and integrating over the full BDT range, for different PID selection strategies; the corresponding signal selection efficiency, given dimuon stripping selection, is given in the bottom row.

	DLL	$ProbNN_2$	$ProbNN_3$	$ProbNN_4$
$B_{(s)}^0 \rightarrow h^+ h^{(')-}$	19.7 ± 0.7	10.5 ± 0.4	6.7 ± 0.2	4.3 ± 0.2
$B_s^0 \rightarrow \pi^- \mu^+ \nu_\mu$	125.9 ± 7.1	89.7 ± 5.0	77.1 ± 4.3	63.2 ± 3.5
$B_s^0 \rightarrow K^- \mu^+ \nu_\mu$	13.6 ± 5.4	10.5 ± 4.1	7.1 ± 2.8	4.8 ± 1.9
$\Lambda_b^0 \rightarrow p \mu^- \nu$	63.4 ± 28.4	19.1 ± 8.6	13.2 ± 5.9	9.7 ± 4.3
$B^{0(+)} \rightarrow \pi^{0(+)} \mu^+ \mu^-$	29.1 ± 7.7	28.8 ± 7.6	27.3 ± 7.2	25.5 ± 6.7
Signal effi	95.4%	94.3%	90.6%	84.6%

state. The reduction observed for $B^{0(+)} \rightarrow \pi^{0(+)} \mu^+ \mu^-$, $\sim 12\%$, is instead the same dimuon loss as observed for signal.

- For $\Lambda_b^0 \rightarrow p \mu^- \nu$ a factor of ~ 6 improvement is observed, which reflects the fact our DLL selection used no proton information.
- For DLL cut, the present estimates exceed the ones in Ref [4] by 35%, 33% and 10% for $B_{(s)}^0 \rightarrow h^+ h^{(')-}$, $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ and $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$, respectively, which reflect the systematic increase observed in the pion and kaon misID probabilities, discussed in Sec. 4. In the case of $\Lambda_b^0 \rightarrow p \mu^- \nu$, instead, we observe a reduction of yield by 25%, again due to the observed change in the single proton misID ¹¹.

5.1 BDT PDFs

Besides the event yields, the complete characterization of the exclusive backgrounds (both peaking and semileptonic) needs the mass and BDT PDFs to be determined.

For BDT, the event fractions populating each of the analysis bins are evaluated by convoluting the MC spectra, after all selection steps but for PID requests and trigger, with the proper hadron misID probabilities. This is the same procedure as followed in Ref. [4], now repeated with the new PID weights. We remind here that the chosen BDT intervals for this analysis are $0 - 0.25 - 0.4 - 0.5 - 0.6 - 0.7 - 0.8 - 0.9 - 1.0$, so that a flat BDT distribution results into fractions of 0.25 and 0.15 in the first two bins, respectively, and a fraction of 0.1 in the following six bins.

The results are shown in Fig. 23, Fig. 24 and Fig. 25 for $B_{(s)}^0 \rightarrow h^+ h^{(')-}$, $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$ and $\Lambda_b^0 \rightarrow p \mu^- \nu$, respectively, after DLL and $ProbNN_4$ selections.

¹¹We use here a $\Lambda_b^0 \rightarrow p \mu^- \nu$ MC sample generated according to LCSR, while a MC phase space with offline corrections applied to the q^2 spectrum was used in Ref [4]: this has no appreciable consequences on the yield.

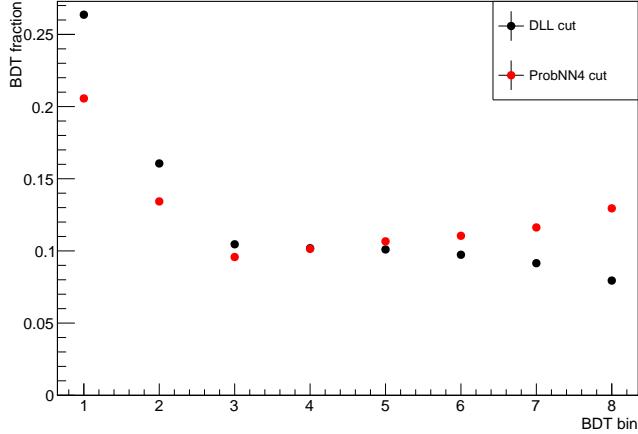


Figure 23: BDT fractions for $B^0_{(s)} \rightarrow h^+h^{(')-}$ events after DLL (black points) and ProbNN4 (red points) selection.

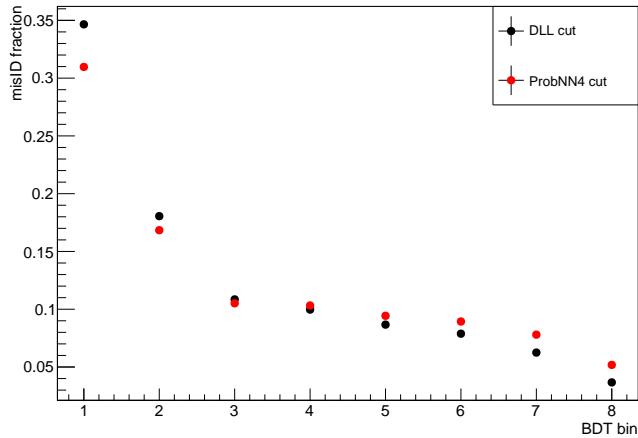


Figure 24: BDT fractions for $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$ events after DLL (black points) and ProbNN4 (red points) selection.

For $B^0_{(s)} \rightarrow h^+h^{(')-}$ the BDT distribution is quite flat, with a residual significant slope at high BDT, which is negative for DLL and positive for $ProbNN_4$. This behaviour is correlated with the slope of the PID efficiency as a function of hadron p_T , which is negative for DLL and positive for $ProbNN_4$. The observed BDT PDF behaviour tends to spoil part of the gain in $B^0_{(s)} \rightarrow h^+h^{(')-}$ background reduction for $ProbNN_4$ with respect to DLL, which is in fact ~ 5 on the total yield while decreasing to ~ 3 in the most significant BDT bin.

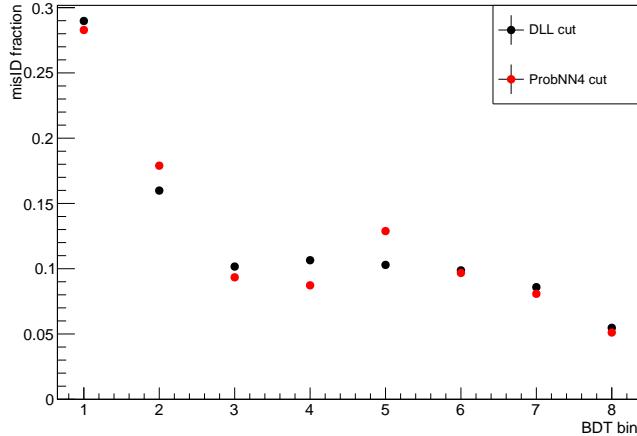


Figure 25: BDT fractions for $\Lambda_b^0 \rightarrow p\mu^-\nu$ events after DLL (black points) and ProbNN4 (red points) selection.

For semileptonic $B^0 \rightarrow \pi^-\mu^+\nu_\mu$ and $\Lambda_b^0 \rightarrow p\mu^-\nu$ backgrounds, a moderate migration towards lower values of BDT is observed, which reflects into a drop in the last BDT bin compensated by an increase of the first BDT bin. For these decays, no significant differences are observed between DLL and ProbNN selections.

5.2 Mass PDFs

For mass PDF, the treatment between semileptonic and peaking background is different:

- For the semileptonic decays, MC events in each of the BDT bins are considered after the full selection but for PID requests on the hadron and for the trigger. PID weights are applied to the hadron tracks on a event-by-event basis. The result is shown in Fig. 26 for $B^0 \rightarrow \pi^-\mu^+\nu_\mu$, and in Fig. 27 for $\Lambda_b^0 \rightarrow p\mu^-\nu$. In both cases, a significant dependence on the BDT bin is observed. Differently from Ref. [4], the mass PDFs are extracted from the MC spectra via a Rookey interpolation: this appears a more flexible solution, when a number of different exclusive channels have to be treated at the same time, and in presence of many different analyses flavours. For the moment no additional systematic errors are added to the mass PDFs, which would require more study.
- For the $B_{(s)}^0 \rightarrow h^+h^{(\prime)-}$, we start from MC events without IsMuon and trigger requests, and we apply a momentum smearing procedure, tuned on MC for kaons and pions separately, which accounts for the hadron decay in flight. This procedure overcomes the obvious lack of MC statistics which would occur when requiring both hadrons to satisfy the IsMuon algorithm. On top of the momentum smearing, event-by-event PID weights are also applied, as for the semileptonic decays. The mass

590 spectra are calculated separately for the $B^0 \rightarrow K^+ \pi^-$, $B^0 \rightarrow K^+ K^-$, $B_s^0 \rightarrow K^+ K^-$
 591 and $B_s^0 \rightarrow \pi^+ K^-$ decay modes, and the results are then combined according to their
 592 BRs and average PID efficiencies. The results are shown in Fig. 28. In the same plot
 593 is also superimposed the mass spectrum used so far in Ref. [4], which comes from a
 594 similar momentum smearing procedure. The observed differences come mainly from
 595 the PID weighting procedure, the expected bias on the B^0 BR being however less
 596 than 5%.

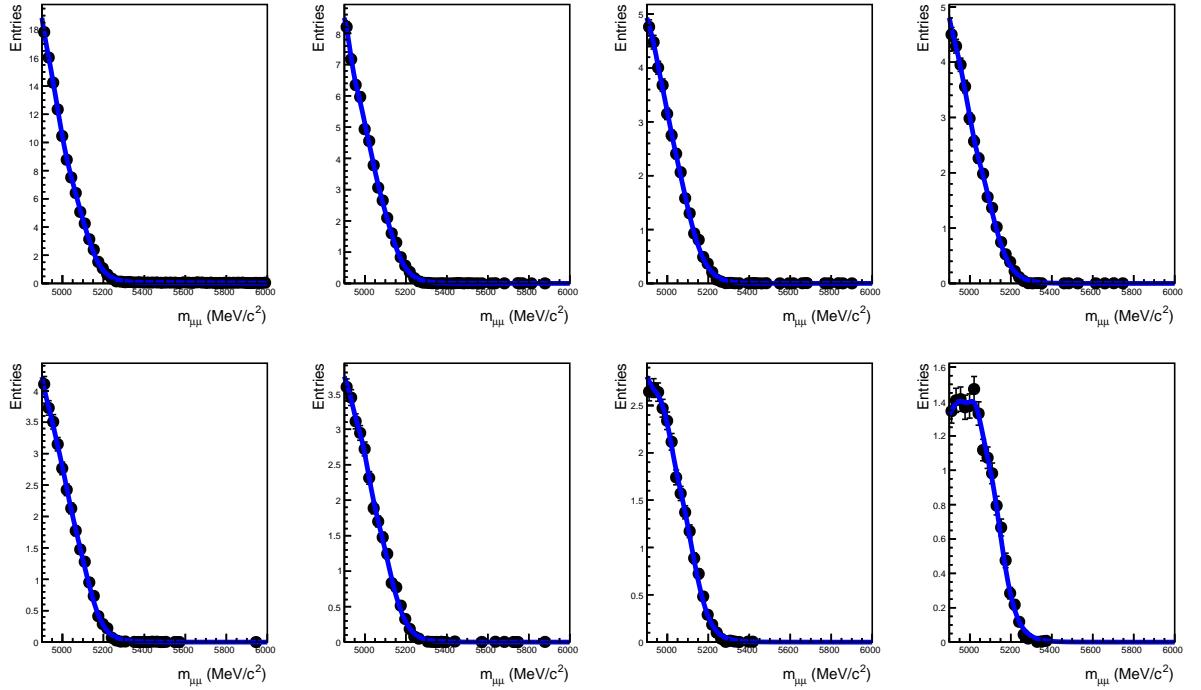


Figure 26: $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$ candidates invariant mass. Event-by-event weights are applied to account for the PID selection (*ProbNN*₄ in this example).

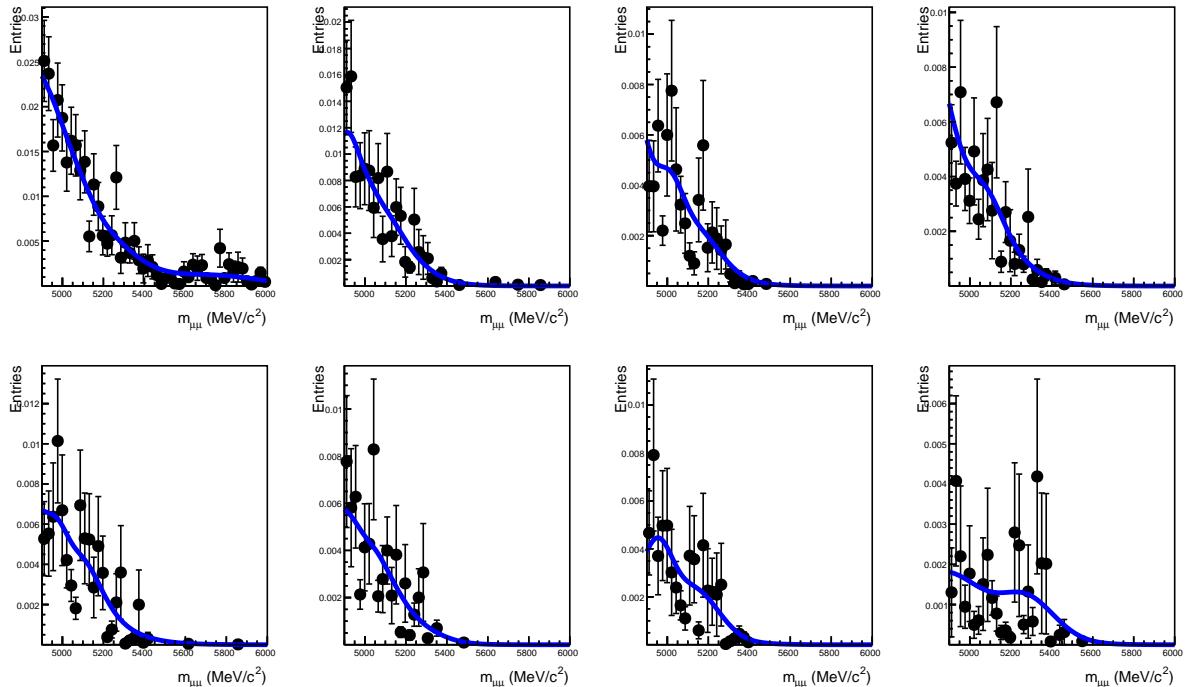


Figure 27: $\Lambda_b^0 \rightarrow p\mu^-\nu$ candidates invariant mass. Event-by-event weights are applied to account for the PID selection (*ProbNN₄* in this example).

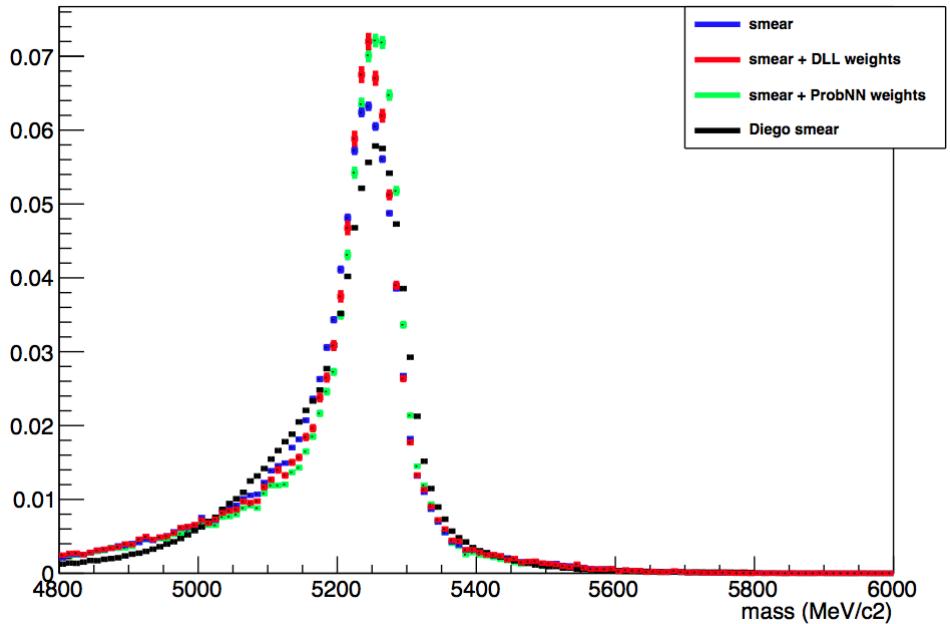


Figure 28: $B_{(s)}^0 \rightarrow h^+ h^{(\prime)-}$ inclusive mass spectrum, with double hadron misID simulated via a momentum smearing (blue), with additional event-by-event DLL (red) and ProbNN (green) weights; the spectrum used in Ref. [4] is also shown (black).

597 6 Combinatorial background estimate

598 The combinatorial background, which is needed to compute the analysis sensitivity,
 599 is evaluated by interpolating the mass sidebands, for each BDT bin separately. The
 600 interpolation is performed after all selection cuts have been applied, including PID requests.
 601 The invariant mass distribution is fitted in the range $4900 < m_{\mu\mu} < 5224.9 \text{ MeV}/c^2$ and
 602 $5431.9 < m_{\mu\mu} < 6000 \text{ MeV}/c^2$, assuming a single exponential function for the combinatorial
 603 background. The lower boundary at $4900 \text{ MeV}/c^2$ was chosen to exclude background
 604 sources like cascading $b \rightarrow c\mu X \rightarrow \mu\mu X$, while the intermediate mass values are set at
 605 $m(B^0) - 60 \text{ MeV}/c^2$ and $m(B_s^0) + 60 \text{ MeV}/c^2$ respectively, which corresponds to more than
 606 2 times the mass resolution. All of the exclusive background sources populating the
 607 sidebands, including leak from the $B_{(s)}^0 \rightarrow h^+ h^{(\prime)-}$, are included in the fit. The background
 608 yields, as well as the mass and BDT PDFs, are given in Sec. 5.

609 The combinatorial background evaluated in the mass range $4900 < m_{\mu\mu} < 6000 \text{ MeV}/c^2$
 610 is listed in Table 4, for different analysis scenarios: a BDT12 with DLL selection, which is
 611 the published analysis [3], and a BDT12_{iso} built with isoBDT in place of the standard
 612 isolation algorithm, as described in Sect. 3.4. In this case, we try all possible PID selections
 613 explored in Sect. 5. Sideband fits are shown in Fig. 29 for BDT12, and for BDT12_{iso} with
 614 DLL and ProbNN4.

Table 4: Combined 2011+2012 (3 fb^{-1}) dataset, s20r1+s20: extrapolated combinatorial background in the mass range $4900 < m_{\mu\mu} < 6000 \text{ MeV}/c^2$ for BDT12 with DLL selection, and BDT12_{iso} with DLL and ProbNN selections.

BDT range	BDT12	BDT12 _{iso}	BDT12 _{iso}		
	DLL	DLL	ProbNN ₂	ProbNN ₃	ProbNN ₄
[0.00, 0.25]	54336 ± 234	54906 ± 235	51788 ± 228	48322 ± 221	43777 ± 210
[0.25, 0.40]	1094 ± 34	670 ± 27	618 ± 26	576 ± 25	519 ± 23
[0.40, 0.50]	234 ± 16	159 ± 14	145 ± 13	129 ± 12	118 ± 12
[0.50, 0.60]	108 ± 11	57 ± 9	54 ± 8	50 ± 8	46 ± 8
[0.60, 0.70]	$49.2^{+8.4}_{-7.8}$	$24.0^{+5.8}_{-5.0}$	$23.7^{+5.7}_{-5.0}$	$24.1^{+5.7}_{-5.0}$	$20.4^{+5.4}_{-4.6}$
[0.70, 0.80]	$20.7^{+5.7}_{-4.9}$	$15.0^{+5.3}_{-4.5}$	$16.7^{+5.5}_{-4.7}$	$16.0^{+5.2}_{-4.4}$	$14.2^{+5.0}_{-4.2}$
[0.80, 0.90]	$12.8^{+4.6}_{-3.9}$	$6.9^{+3.7}_{-2.8}$	$5.8^{+3.5}_{-2.7}$	$3.2^{+3.0}_{-2.0}$	$3.9^{+3.2}_{-2.2}$
[0.90, 1.00]	$3.9^{+3.2}_{-2.4}$	$4.2^{+3.4}_{-2.5}$	$6.0^{+3.6}_{-2.8}$	$5.4^{+3.6}_{-2.8}$	$5.5^{+3.4}_{-2.7}$

615 A couple of points are in evidence:

- 616 • Passing from the BDT12 to BDT12_{iso}, there's a clear migration towards BDT values
 617 below 0.25: in particular, if we consider *e.g.* the region above 0.5, the combinatorial
 618 background is reduced by $\sim 40\%$;
- 619 • Moving towards a tighter PID selection also rejects some combinatorial background,
 620 more or less flat in BDT: *e.g.* passing from DLL to ProbNN₄ the event reduction is

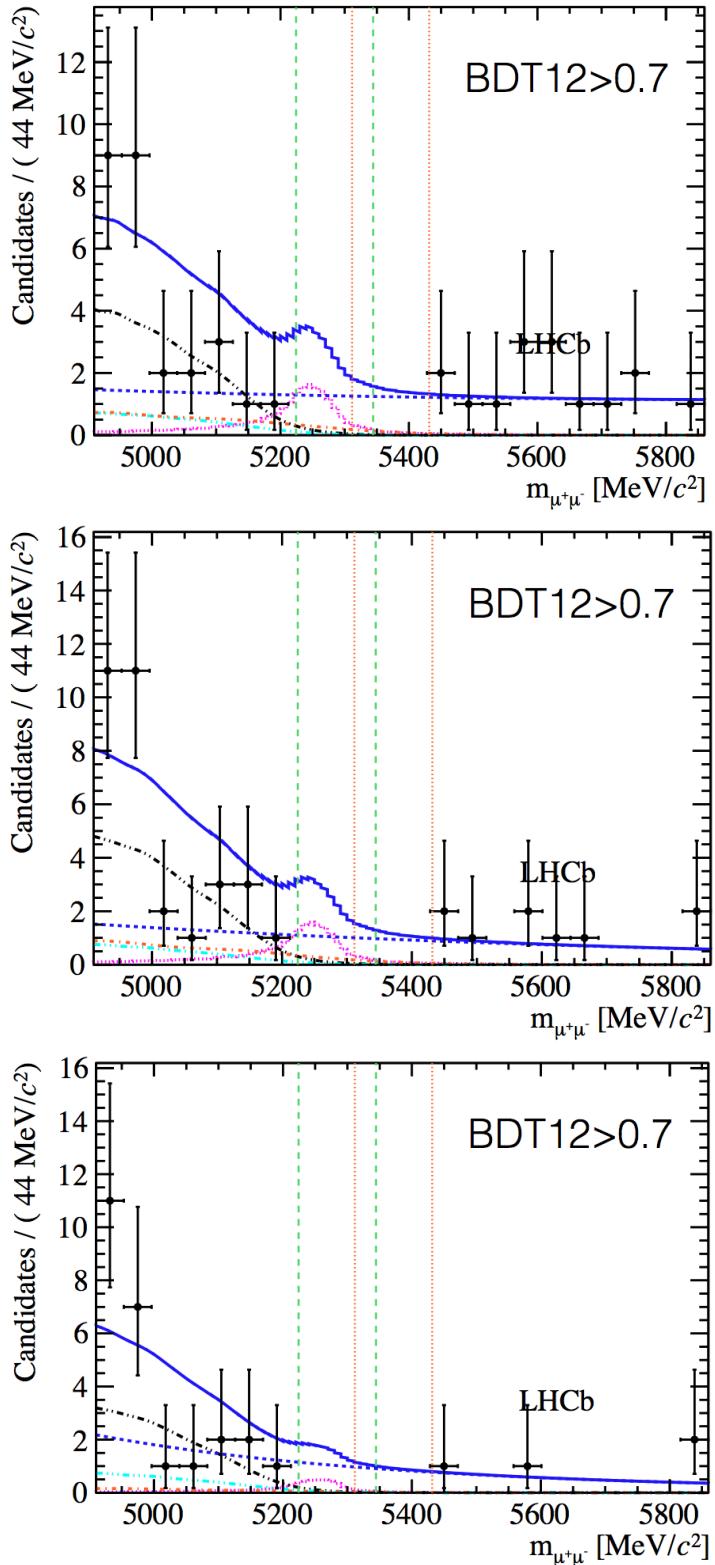


Figure 29: Sideband mass fits for combinatorial background interpolation for $\text{BDT} > 0.7$: BDT12-DLL (top), BDT12_{iso}-DLL (medium) and BDT12_{iso}-ProbNN4 (bottom); $B_{(s)}^0 \rightarrow h^+h^-$ (magenta dotted), $B^0 \rightarrow \pi^-\mu^+\nu_\mu$ and $B_s^0 \rightarrow K^-\mu^+\nu_\mu$ (black dot-dashed), $B^{0(+)} \rightarrow \pi^{0(+)}\mu^+\mu^-$ (light blue dot dashed), $\Lambda_b^0 \rightarrow p\mu^-\nu$ (orange dot-dashed), and the combinatorial background (blue medium dashed).

₆₂₁ $\sim 20\%$; this is partly due to the loss in efficiency for true dimuon states, which is
₆₂₂ $\sim 12\%$ in this case, and partly to rejection of misidentified hadrons.

623 7 Sensitivity studies

624 In this section we will show the impact of the optimization studies about BDT and PID
625 selection of Sect. 3 and Sect. 5, respectively, on B^0 and B_s^0 signal sensitivities. To this
626 purpose, we performed toy MC simulations, with the following inputs:

- 627 • B^0 and B_s^0 signal extracted according to SM rates [2], and with Crystal Ball PDFs
628 as described in Ref. [4].
- 629 • The exclusive background sources generated according to the yields and PDFs
630 estimated in Sect. 5, for each PID cut. As usual, the computation assumes no
631 contribution from $B_c^+ \rightarrow J/\psi\mu^+\nu$ and treats $B^0 \rightarrow \pi^-\mu^+\nu_\mu$ and $B_s^0 \rightarrow K^-\mu^+\nu_\mu$ as
632 a single component in the fits.
- 633 • Combinatorial background extracted according to the yields as interpolated from
634 the data sidebands, as discussed in Sect. 6; the interpolation is repeated for all of
635 the PID cuts considered.

636 The sensitivity corresponding to each single toy is extracted with Wilks' theorem by fitting
637 the generated dataset. The median value of the sensitivity distribution from a sample of
638 $\sim 2k$ toys gives the expected sensitivity for SM signals in a given analysis scenario.

639 In the standard analysis scenario, BDT12 with cut-based muon isolation and DLL cuts,
640 we get from this method an expected sensitivity of 0.84 ± 0.02 for B^0 and 4.81 ± 0.05
641 for B_s^0 , respectively. The quoted uncertainty is due to the toy MC statistics, and has
642 been evaluated by repeating 10 times the toy study for this configuration, each time with
643 2k events, and taking the RMS of the median sensitivities. This uncertainty turns out
644 to be more relevant for B^0 , as expected. For this reason, and also to better ensure the
645 validity of the asymptotic conditions in which the Wilks' theorem holds, we decided to
646 perform the toy studies for B^0 at 10 times the luminosity corresponding to our dataset,
647 $10 \times 3 \text{ fb}^{-1}$. This assumption does not introduce a bias in the result, as has been verified
648 on B_s^0 , for which the scaling of the sensitivities in different analysis conditions stays equal
649 independently of the luminosity factor.

650 Starting from the published analysis result (BDT12 & DLL), several scenarios have
651 been compared which make use of the BDT12 built with isoBDT, BDT12_{iso} , in place of
652 the standard isolation algorithm, and exploring different PID selections: the standard
653 DLL selection, and the new $ProbNN_{1,2,3,4,5}$ selections (see Sect. 5 for the definitions of
654 these cuts). In order to verify the impact of the proton PID, three more selections without
655 the proton PID cut have been also tested, which are $ProbNN_{nop3,4,5}$ (see Sect. 5). The
656 expected sensitivity results at SM rates for B^0 (at 30 fb^{-1}) and for B_s^0 (at 3 fb^{-1}) are listed
657 in Table 5. We remind that these results do not take into account the observed yield in
658 the 3 fb^{-1} analysis.

659 For both B^0 and B_s^0 a clear sensitivity improvement is observed, which is 18% and 12%,
660 respectively, as compared to the present analysis performance. The improvement takes
661 contribution from both the new PID selection and the BDT12_{iso} , the PID being more
662 relevant for B^0 with respect to B_s^0 . This is due to the different background composition,

Table 5: Expected sensitivities for B^0 and B_s^0 signals at SM rates, for 30 fb^{-1} and 3 fb^{-1} integrated luminosity, respectively; both absolute sensitivities and relative values with respect to the standard analysis are listed. The listed values do not account for the observed yields in the 3 fb^{-1} data sample.

	$B^0, 30\text{ fb}^{-1}$		$B_s^0, 3\text{ fb}^{-1}$	
BDT12 & DLL	1.77		4.81	
BDT12 _{iso} & DLL	1.85	1.05	5.20	1.07
BDT12 _{iso} & <i>ProbNN</i> ₁	1.87	1.06	5.23	1.08
BDT12 _{iso} & <i>ProbNN</i> ₂	1.94	1.10	5.37	1.11
BDT12 _{iso} & <i>ProbNN</i> ₃	2.08	1.18	5.43	1.12
BDT12 _{iso} & <i>ProbNN</i> ₄	2.06	1.16	5.28	1.09
BDT12 _{iso} & <i>ProbNN</i> ₅	2.04	1.15	4.95	1.02
BDT12 _{iso} & <i>ProbNNnop</i> ₃	2.08	1.18	5.42	1.12
BDT12 _{iso} & <i>ProbNNnop</i> ₄	2.06	1.16	5.29	1.09
BDT12 _{iso} & <i>ProbNNnop</i> ₅	1.96	1.11	5.06	1.05

which is equally shared among combinatorial and peaking background for B^0 , while it is dominated by combinatorial for B_s^0 . If we compare the *ProbNN*₃ selection with the corresponding one without the proton PID cut, *ProbNNnop*₃, we don't observe any improvement. This is due to the fact that already after *ProbNNnop*₃ selection the $\Lambda_b^0 \rightarrow p\mu^-\nu$ background becomes subdominant with respect to the combinatorial background ¹².

To relate further sensitivity improvements to the corresponding background reductions, some interesting limit cases have been studied:

- any residual $B_{(s)}^0 \rightarrow h^+h^{(\prime)-}$ background is removed without introducing any further reduction on the signal efficiency,
- the combinatorial background is further reduced in the sensitive region at high BDT, with an hypothetical more powerful multivariate classifier.

In the following two sections the results of these studies are shown.

7.1 Limit case: $B_{(s)}^0 \rightarrow h^+h^{(\prime)-}$ background reduction

The sensitivity obtained by rejecting any residual $B_{(s)}^0 \rightarrow h^+h^{(\prime)-}$ background without introducing further reduction on the signal efficiency is evaluated for two different con-

¹²A small effect of the $\Lambda_b^0 \rightarrow p\mu^-\nu$ is still visible on *ProbNN*₅: while the B_d channel can profit of a further reduction of this background, thus preferring the *ProbNN*₅ selection, the B_S is mostly unaffected, due to the rapidly falling mass spectrum, and thus prefers the *ProbNNnop*₅ selection, which is slightly more efficient.

679 configurations: the standard DLL selection and a mixed configuration (Mixed). In this case
 680 the signal and the exclusive backgrounds are taken from $ProbNN_2$ selection, while the
 681 combinatorial background is taken from $ProbNN_4$ and rescaled back to muon efficiency of
 682 $ProbNN_2$ selection. This represents an optimal configuration for this study, since most of
 683 the signal is retained, $\sim 99\%$ with respect to DLL, while the residual misidentified hadron
 684 component in the combinatorial background is completely removed.

685 The B^0 expected sensitivities for the DLL and Mixed scenarios with and without the
 686 $B_{(s)}^0 \rightarrow h^+h^{(\prime)-}$ background are listed in Table 6. In the same Table the best PID selection
 687 as determined above is reported, $BDT12_{iso}$ & $ProbNN_3$, and the standard analysis, $BDT12$
 & DLL , to which all numbers are referred.

Table 6: Expected sensitivities for B^0 signal at SM rate, for 30 fb^{-1} ; both absolute sensitivities
 and relative values with respect to the standard analysis, given in the 1st row, are listed. The
 Mixed selection is explained in the text.

	$B^0, 30\text{ fb}^{-1}$	
$BDT12$ & DLL	1.77	
$BDT12_{iso}$ & $ProbNN_3$	2.08	1.18
$BDT12_{iso}$ & DLL	1.85	1.05
$BDT12_{iso}$ & DLL & No $B_{(s)}^0 \rightarrow h^+h^{(\prime)-}$	2.18	1.23
Mixed	2.00	1.13
Mixed & No $B_{(s)}^0 \rightarrow h^+h^{(\prime)-}$	2.35	1.27

688 By comparing the performances of Mixed scenario without $B_{(s)}^0 \rightarrow h^+h^{(\prime)-}$ with respect
 689 to $ProbNN_3$, we can conclude that we could in principle gain $\sim 10\%$ more sensitivity by
 690 improving the PID selection with respect to present best performances. The impact of
 691 removing $B_{(s)}^0 \rightarrow h^+h^{(\prime)-}$ events from DLL selection is slightly lower, due to the larger
 692 contribution from other background sources.

693 The above result shows that the analysis could largely profit from a further optimization
 694 of the PID. Two approaches are under investigation. The first is to use the $NNpid$
 695 outputs for all particle species as an input to a “PID-BDT”, to be trained on the proper
 696 $B_{(s)}^0 \rightarrow h^+h^{(\prime)-}$ phase space; the $PID-BDT$ output could then be used for a cut-based PID
 697 selection, or as a third analysis likelihood, besides the $m_{\mu\mu}$ and the BDT, in a three-
 698 dimensional binned space. The second approach, consists of exploiting at maximum the
 699 ANN-PID tool developed to produce the $NNpid$ variables, and to explore its full potential
 700 for the $B_{d,s}^0 \rightarrow \mu^+\mu^-$. Together with the author of [18], we are studying the case in which
 701 instead of training the neural network with a generic blend of B decays MC samples, more
 702 specific MC samples are feed into the NN for training. Specifically, in our case we are using
 703 muons from $B_{d,s}^0 \rightarrow \mu^+\mu^-$, and pions and kaons from $B_{(s)}^0 \rightarrow h^+h^-$, while for protons the
 704 generic blending used for the ANN-PID tool is well suited for our purpose.

706 **7.2 Limit case: combinatorial background reduction**

707 Another important limit case comes from the asymptotic reduction of the combinatorial
 708 background at high BDT. The PDF of the BDT is varied in order to study the relation
 709 between the background reduction and the sensitivity to B_s and B_d signals. The BDT PDF
 710 is described by two terms: a leading behavior expressed with two exponential functions,
 711 and a uniform term representing the fractional component of events misidentified as signal.
 712 The change in performances of the BDT can be emulated changing the parameters of its
 713 PDF. Some configurations are shown in Fig. 30. Starting from the PDF evaluated from
 714 data sidebands after the $ProbNN_4$ selection (red curve), a reduction of the combinatorial
 715 background is simulated increasing the slopes of the exponential PDFs by 25% (magenta,
 716 conf1 in the following), and 50% (blue, conf2 in the following). The corresponding
 717 background reduction for BDT above 0.5 is $\sim 50\%$ and $\sim 75\%$ for conf1 and conf2,
 718 respectively.

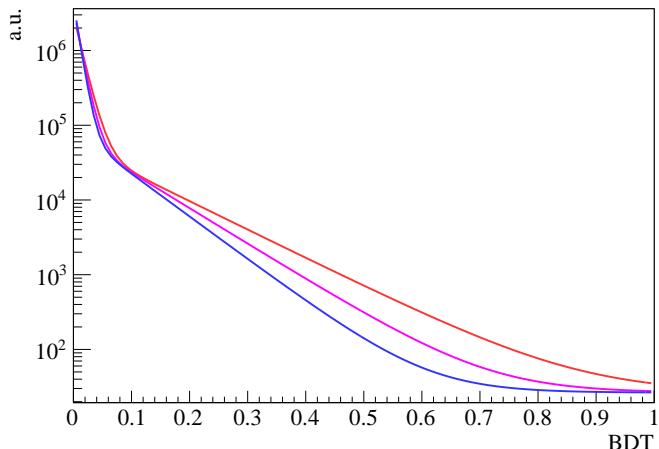


Figure 30: Combinatorial background BDT PDF approximated as a double exponential plus a constant term: red line corresponds to the $ProbNN_4$ selection; magenta and blue lines correspond to a 25% and 50% increase of the exponential slopes, respectively.

719 The B^0 expected sensitivities for conf1 and conf2 scenarios are listed in Table 7. In
 720 the same Table the $ProbNN_4$ sensitivity is reported, since it is used as a starting point
 721 for the combinatorial background scaling, and the standard analysis result, to which all
 722 numbers are referred. By comparing the performances of conf1 and conf2 scenarios with
 723 respect to $ProbNN_4$, we can conclude that a $\sim 20 - 30\%$ sensitivity gain can be obtained
 724 by reducing the combinatorial background in the sensitive region by 50 – 75%.

Table 7: Expected sensitivities for B^0 signals at SM rate, for 30 fb^{-1} ; both absolute sensitivities and relative values with respect to the standard analysis, given in the 1st row, are listed. The conf1 and conf1 analysis scenarios are explained in the text.

	$B^0, 30\text{ fb}^{-1}$	
BDT12 & DLL	1.77	
BDT12 _{iso} & $ProbNN_4$	2.06	1.16
conf1	2.44	1.38
conf2	2.68	1.51

725 8 Conclusions

726 In this document recent studies about background in the $B^0 \rightarrow \mu^+ \mu^-$ analysis have been
727 presented:

- 728 • The determination of the hadron misID probabilities from data control samples,
729 which are crucial for the evaluation of the peaking and exclusive backgrounds, has
730 been reviewed, and significant systematic effects have been discovered and fixed.
- 731 • The PID selection strategy has been revised, with the purpose of reducing the
732 peaking background under B^0 at a cost of a moderate signal loss; the usage of NNpid
733 variables [18] has been explored, using a linear combination of muon, kaon and
734 proton NNpid outputs, and several PID cut values have been tried: as a result, on
735 $B_{(s)}^0 \rightarrow h^+ h^{(\prime)-}$ events reduction factors of 2.9 and 4.5 with respect to DLL selection
736 are obtained with 5% and 12% signal losses, respectively. Further optimization
737 studies on PID selection are ongoing, which include the definition of a non-linear
738 combination of the NNpid outputs optimized on the $B_{(s)}^0 \rightarrow h^+ h^{(\prime)-}$ phase space,
739 and an ad-hoc retuning of the single NNpid variables (together with the author of
740 Ref. [18]).
- 741 • The muon isolation, which is crucial to the rejection of the combinatorial background,
742 has been also revised, by implementing a BDT-based algorithm (isoBDT); the
743 isoBDT performs significantly better than the cut-based muon isolation used in the
744 published analysis: having introduced the isoBDT into the global event BDT, a
745 background rejection increase of 35-40% is observed in the intermediate BDT region.
746 At least two other isolation strategies are currently under investigation by different
747 authors, one based on the vertex information between the tracks of the event [9], the
748 other exploiting the reconstruction of jets [10]. Eventually the different approaches
749 will have to be combined or merged together to maximize the analysis sensitivity.

750 The expected sensitivity at SM rates for B^0 and B_s^0 has been estimated from toy MC to
751 compare different analysis scenarios: the best result consists of an improvement of 18% on
752 B^0 sensitivity and of 12% on B_s^0 sensitivity, obtained with isoBDT and with an NNpid
753 selection implying a 5% signal loss with respect to the published analysis.

754 Dedicated MC toy studies indicate a residual margin of improvement on B^0 sensitivity
755 of $\sim 10\%$ from PID selection optimization, and of $\sim 20\%$ from a further factor of
756 two in the rejection of combinatorial background, with respect to what obtained with
757 isoBDT. The above figures can help to understand which are the limits of possible analysis
758 improvements, which is no guarantee of future results. On the other hand, many aspects of
759 the PID selection and the isolation algorithms present interesting handles which is worth
760 investigating.

761 References

- 762 [1] CMS and LHCb collaborations, V. Khachatryan *et al.*, *Observation of the rare*
763 *$B_s^0 \rightarrow \mu^+ \mu^-$ decay from the combined analysis of CMS and LHCb data*, LHCb-
764 PAPER-2014-049, arXiv:1411.4413, submitted to Nature.
- 765 [2] C. Bobeth *et al.*, *$B_{s,d} \rightarrow l^+ l^-$ in the Standard Model with Reduced Theoretical*
766 *Uncertainty*, Phys. Rev. Lett. **112** (2014) 101801, arXiv:1311.0903.
- 767 [3] LHCb collaboration, R. Aaij *et al.*, *Measurement of the $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction*
768 *and search for $B^0 \rightarrow \mu^+ \mu^-$ decays at the LHCb experiment*, Phys. Rev. Lett. **111**
769 (2013) 101805, arXiv:1307.5024.
- 770 [4] C. Adrover *et al.*, *Search for the $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ decays with 3 fb^{-1} at*
771 *LHCb*, LHCb-ANA-2013-032.
- 772 [5] G. Mancinelli and J. Serrano, *Study of muon isolation in the $B_s \rightarrow \mu\mu$ channel*,
773 LHCb-INT-2010-011.
- 774 [6] L. Gavardi, *Search for lepton flavour violation in τ decays at the LHCb experiment*,
775 CERN-THESIS-2013-259.
- 776 [7] J. Albrecht *et al.*, *Search for lepton flavour violating decay $\tau \rightarrow \mu\mu\mu$* , LHCb-ANA-
777 2014-005.
- 778 [8] M. Rama *et al.*, *BDT studies for $B \rightarrow \mu\mu$* , <https://indico.cern.ch/event/291729/contribution/5/material/slides/0.pdf>.
- 780 [9] A. Morda and G. Mancinelli, *ZVTop isolation and BDT selection*, <https://indico.cern.ch/event/291729/contribution/6/material/slides/0.pdf>.
- 782 [10] X. Cid Vidal, *Jets applied to $B \rightarrow \mu\mu$* , <https://indico.cern.ch/event/291735/contribution/0/material/slides/0.pdf>.
- 784 [11] M. Pivk and F. R. Le Diberder, *sPlot: a statistical tool to unfold data distributions*,
785 Nucl. Instrum. Meth. **A555** (2005) 356, arXiv:physics/0402083.
- 786 [12] *PIDCalib*, <https://twiki.cern.ch/twiki/bin/view/LHCb/PIDCalibPackage>.
- 787 [13] K. Akiba *et al.*, *Results on muon identification efficiency with 2011 data at lhcb*,
788 LHCb-INT-2011-045.
- 789 [14] LHCb Collaboration, R. Aaij *et al.*, *Strong constraints on the rare decays $B_s \rightarrow \mu^+ \mu^-$*
790 *and $B^0 \rightarrow \mu^+ \mu^-$* , Phys. Rev. Lett. **108** (2012) 231801, arXiv:1203.4493.
- 791 [15] LHCb Collaboration, R. Aaij *et al.*, *First Evidence for the Decay $B_s^0 \rightarrow \mu^+ \mu^-$* , Phys.
792 Rev. Lett. **110** (2013) 021801, arXiv:1211.2674.

- 793 [16] M. Palutan *et al.*, *Selection of charged kaons and pions and PIDCalib validation*,
794 <https://indico.cern.ch/event/321939/contribution/0/material/slides/0.pdf>.
- 796 [17] LHCb collaboration, R. Aaij *et al.*, *Measurement of the $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction*
797 *and search for $B^0 \rightarrow \mu^+ \mu^-$ decays at the LHCb experiment*, Phys. Rev. Lett. **111**
798 (2013) 101805, [arXiv:1307.5024](https://arxiv.org/abs/1307.5024).
- 799 [18] C. R. Jones, *ANNPID variables*, <https://indico.cern.ch/event/226062/contribution/1>.