CMS Draft Analysis Note

The content of this note is intended for CMS internal use and distribution only

2012/11/09
Head Id: 156816
Archive Id: 157095M
Archive Date: 2012/11/07
Archive Tag: trunk

Search for $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ with the 2011 and 2012 data

D. Kotlinski[1], U. Langenegger[1], J. Lazo-Flores[2], C. Lourenco[3], L. Martini[4], C. Nägeli[1], F. Palla[5], P. Squillacioti[4], A. Starodumov[6], and K. Ulmer[7]

[1] Paul Scherrer Institute, Villigen
[2] University of Nebraska, Lincoln
[3] CERN, Geneva
[4] INFN Pisa and University of Siena
[5] INFN Pisa
[6] ETH Zurich
[7] University of Colorado, Boulder

Abstract

We search for the rare decays $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$. This version documents the search on the 2012 dataset until the technical stop in September 2012. We combine the results with the 2011 dataset. This version uses the 2012-BDT for the analysis.

.tex

Contents

1 Open issues

1.1 Issues as of 2012/10/12

- Issues with (our) trees
    - control sample MC with reduced statistics. This makes the systematics for the control sample somewhat useless.
    - $B^0 \to \mu^+\mu^-$ MC with wrong trigger setup (i.e. useless). This produces the NaN in some tables.
    - $\Lambda_b \to p\mu\nu$ MC with wrong setup (i.e. useless) We replaced it with the 2011 version.
- Fig. 22: The absolute level of the significance should not be taken too seriously (relies on the signal MC luminosity).
- We should meditate on all numbers, and compare with AN-12-238.

1.2 Issues as of 2012/10/24

Compared to the previous version of this document, we realized that there was a gen-level filter on the signal MC (not what was requested). This changes the acceptance down by 40% and affects the expected signal yield. In this version, all numbers are corrected for this.

- we don't yet have the final 2012 muon selection trees ready (the selection is coded and described in the text)
- we don't yet have the final 2012 data book keeping ready (i.e. the 2012 data trees still lack the final validation)
- the trigger efficiency looks somewhat different (lower) between 2012 and the 3e33 menu of 2011. This should be understood.
- since we don't yet have the final trees, the expected sensitivity is not yet computed. Tab. 24 contains the uptodate numbers, however.

2 Introduction

In recent years, the particle physics community has gained new insights into flavor physics and CP violation from the analysis of data from B factories and the Tevatron. The data shows that the Cabibbo-Kobayashi-Maskawa (CKM) matrix is the dominant source of flavor-changing interactions and CP violation. New effects could not (yet) be conclusively established. With the advent of the LHC, the field has entered a new phase in the testing of the flavor sector of the Standard Model (SM), most notably through B_s^0-meson decays. In this respect, two of the most promising channels for detecting signals of new physics (NP) are the rare decays $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ [1].

2.1 Theory

The leptonic decay modes $B^0_{s(d)} \to \ell^+\ell^-$ (where $\ell = e, \mu$)[1] have a highly suppressed rate in the SM. This is because they involve flavor-changing neutral current (FCNC) transitions, $b \to s(d)$, which are forbidden at tree-level and can only proceed through high-order diagrams that are described by electroweak penguin and box diagrams at the one loop level (see Figure 1). In addition, the decays are helicity suppressed by factors of m_ℓ^2 [2].

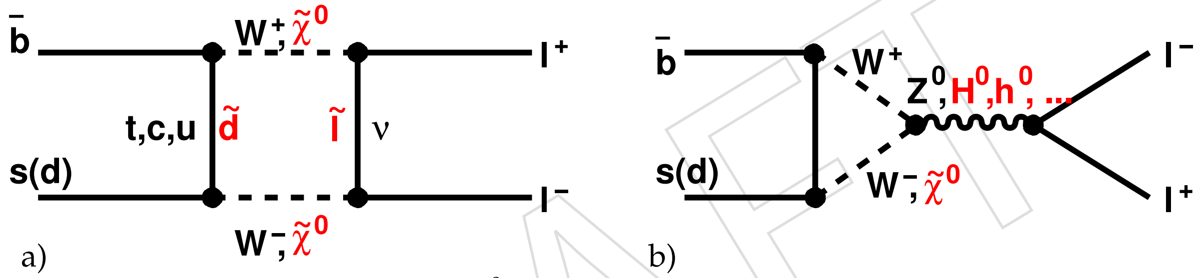

Figure 1: Illustration of the rare decays $B^0_{s(d)} \to \ell^+\ell^-$. In the SM, these decays proceed through $W^\pm$ and Z^0 bosons in a box diagram (a) and Z-penguin (b) interactions. The box diagram is suppressed by a factor of $m_W^2/m_t^2 \approx 0.2$ with respect to the Z-penguin diagram. In SM extensions (*e.g.*, in the MSSM) new particles (*e.g.*, charginos, neutralinos $\tilde{\chi}^0$, Higgs bosons and supersymmetric partners of the quarks and leptons) can contribute to the process and thereby can increase the expected branching fraction by orders of magnitude.

Given that these processes are highly suppressed in the SM, they are potentially sensitive probes for physics beyond the SM. The branching fraction for these decays can be enhanced in NP models, although in most models the rates can also be lowered [3], depending on specific parameters. The branching fraction change arises mostly from scalar or pseudo-scalar couplings which are not helicity suppressed. The decays $B^0 \to \mu^+\mu^-$ provide a unique sensitivity to such couplings.

In the Minimal Supersymmetric extension of the Standard Model (MSSM), the rates are greatly enhanced by large values of $\tan\beta$ [4, 5]. In supersymmetric models with modified minimal flavor violation (MFV), the branching fraction can be increased by up to four orders of magnitude at large $\tan\beta$ [6]. A measurement of both $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays are interesting since they can be enhanced separately even at low $\tan\beta$ in specific models containing leptoquarks [7] and supersymmetric models with non-universal Higgs masses [8].

The branching fraction enhancement for $B_s^0 \to \mu^+\mu^-$ in the MSSM is proportional to $\tan^6\beta$, which provides a certain sensitivity to $\tan\beta$[9]. There has been significant interest, in the past [10], in using the decay mode $B_s^0 \to \mu^+\mu^-$ to "measure" the key parameter $\tan\beta$ of the

[1] Charge conjugation is implied throughout this note; exceptions will be clearly spelled out.

MSSM and to constrain other extensions of the SM. The determination of tan β is difficult—there is not a general technique to measure it at hadron colliders—yet all supersymmetric observables depend on it. It has been shown that with very general assumptions that do not depend on specific models, it is possible to put significant lower (and to a lesser extent also upper) bounds on tan β. Since however, based on very general principles, tanβ is constrained from above [11], already a lower bound on tan β is arguably tantamount to a measurement.

Since the observation of $B_s^0 \to \mu^+\mu^-$ seems imminent (at least if the decay proceeds at the SM level), a renewed theoretical assessement of the SM expecation [12] has taken place and new observables [13] have been found. The effects of radiative corrections, pointed out in Ref. [14], has been taken into account in this analysis and therefore does not lead to a decrease of the physically observable branching fraction.

2.2 Other experiments

The searches for the rare B decays at the $\Upsilon(4S)$ resonance, i.e., the CLEO, Belle, and BABAR experiments, have no sensitivity to B_s^0 decays. However, the CDF and D0 experiments at the Tevatron have sensitivity to the decay $B_s^0 \to \mu^+\mu^-$. The D0 experiment cannot discriminate between the decay $B_s^0 \to \mu^+\mu^-$and $B^0 \to \mu^+\mu^-$because of its limited mass resolution[2]. So far, neither CDF nor D0 have found evidence for the decay. The current best limits from CLEO [15], Belle [16], BABAR [17], CDF [18], D0 [19], and LHCb [20] are shown in Table 1 together with the SM expectation.

Table 1: The expected branching ratios in the Standard Model [2] and the current best upper limits (U.L.) at the 95% C.L. The experimental results are ordered chronologically.

Mode	$B_s^0 \to \mu^+\mu^-$	$B^0 \to \mu^+\mu^-$	$B^0 \to e^+e^-$
SM Expect.	$(3.2 \pm 0.2) \times 10^{-9}$	$(1.0 \pm 0.1) \times 10^{-10}$	$(2.5 \pm 0.1) \times 10^{-15}$
CLEO [15]	-	6.1×10^{-7}	8.3×10^{-7}
BELLE [16]	-	1.6×10^{-7}	1.9×10^{-7}
BABAR [17]	-	6.1×10^{-8}	8.3×10^{-8}
D0 [19]	5.1×10^{-8}	-	-
CDF [21]	4.0×10^{-8}	$6.0(7.6) \times 10^{-9}$	-
CMS [22]	1.6×10^{-8}	$3.7(4.6) \times 10^{-9}$	-
LHCb [23]	1.2×10^{-8}	$2.6(3.2) \times 10^{-9}$	-
CMS [24]	7.7×10^{-9}	$1.8(1.6) \times 10^{-9}$	-
LHCb [25]	4.5×10^{-9}	$1.1(1.4) \times 10^{-9}$	-
ATLAS [26]	2.2×10^{-8}	-	-

The Tevatron has not integrated enough luminosity for the D0 and CDF experiments to measure this process at the SM expectation. Their analyses have been tuned for high efficiency and are limited by backgrounds. The baton has been passed to the LHC experiments CMS and LHCb.

2.3 Analysis Overview

In this (blind) analysis we search simultaneously for the decay $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$. We perform two analyses: (1) a counting experiment in a one-dimensional signal window in the dimuon mass distribution centered on the B_s^0 (B^0) meson mass, and (2) an unbinned maximum likelihood fit to the dimuon mass window.

[2]The signal mass window in D0 is $5.047 < m_{\mu\mu} < 5.622$ GeV.

The goal of the analysis consists of a very strong background reduction while keeping the signal efficiency as high as possible. The background is estimated from the sidebands and from MC simulation for peaking backgrounds (for example $B_s^0 \to KK$ where both kaons are misidentified as muons).

The present analysis uses a relative normalization to the well-measured decay $B^\pm \to J/\psi K^\pm$ to avoid a dependence on the uncertainties of the $b\bar{b}$ production cross section and luminosity measurements (in fact, this analysis is completely independent of the luminosity measurement). Furthermore, many systematic errors cancel to first order when deriving the upper limit normalizing to a similar decay channel measured in data. Choosing a decay channel with a signature similar to the signal decay has the advantage that many systematic errors cancel to first order. We refer to the $B^\pm \to J/\psi K^\pm$ sample as 'normalization sample' below.

The upper limit at the 95% C.L. on the branching fraction is (schematically) determined by

$$
\begin{aligned}
\mathcal{B}(B_s^0 \to \mu^+\mu^-; 95\%\mathrm{C.L.}) &= \frac{N(n_{obs}, n_B, \delta; 95\%\mathrm{C.L.})}{\varepsilon_{B_s^0}\, N^{\mathrm{obs}}_{B_s^0}} = \frac{N(n_{obs}, n_B, \delta; 95\%\mathrm{C.L.})}{\varepsilon_{B_s^0}\, \mathcal{L}\, \sigma(pp \to B_s^0)} \\
&= \frac{N(n_{obs}, n_B, \delta; 95\%\mathrm{C.L.})}{N(B^\pm \to J/\psi K^\pm)} \times \left(\frac{f_u}{f_s}\right) \left(\frac{A^{B^+}}{A^{B_s^0}}\right) \\
&\quad \times \left(\frac{\varepsilon^{B^+}_{\mathrm{trig}}}{\varepsilon^{B_s^0}_{\mathrm{trig}}}\right) \left(\frac{\varepsilon^{B^+}_{\mu}}{\varepsilon^{B_s^0}_{\mu}}\right) \left(\frac{\varepsilon^{B^+}_{\mathrm{analysis}}}{\varepsilon^{B_s^0}_{\mathrm{analysis}}}\right) \\
&\quad \times \mathcal{B}(B^\pm \to J/\psi K^\pm)\mathcal{B}(J/\psi \to \mu^+\mu^-)
\end{aligned}
\tag{1}
$$

where $N(n_{obs}, n_B, \delta; 95\%\mathrm{C.L.})$ is the expected 95% C.L. upper limit on the number of signal decays, for n_{obs} observed events with n_B expected back-ground events and δ is the corresponding error. $N(B^\pm \to J/\psi K^\pm)$ is the number of reconstructed $B^\pm \to J/\psi K^\pm$ candidates, $f_s/f_u = 0.267 \pm 0.021$ (measured by LHCb for $2 < \eta < 5$ [27]) describes the ratio of probabilities for a b-quark hadronizing into a B_s^0 or a B^+ meson, A_{B^+} and $A_{B_s^0}$ are the B^+ and B_s^0 acceptance, $\varepsilon^{B^+}_{\mathrm{trig}}$ and $\varepsilon^{B_s^0}_{\mathrm{trig}}$ are the corresponding trigger efficiencies, $\varepsilon^{B^+}_{\mu}$ and $\varepsilon^{B_s^0}_{\mu}$ are the muon identification efficiency for muons coming form B^+ and B_s^0 decays, $\varepsilon^{B^+}_{\mathrm{analysis}}$ ($\varepsilon^{B_s^0}_{\mathrm{analysis}}$) is the analysis efficiency for signal (normalization) events, and $\mathcal{B}(B^\pm \to J/\psi K^\pm)$ and $\mathcal{B}(J/\psi \to \mu^+\mu^-)$ are the branching fractions for $B^\pm \to J/\psi K^\pm$ and $J/\psi \to \mu^+\mu^-$, respectively.

The background level is very different in the forward direction compared to the barrel. Furthermore, the mass resolution in the CMS detector depends strongly on the pseudorapidity of the reconstructed particles, which will help in distinguishing between $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-$ decays. Therefore we perform the analysis in two 'channels' (barrel and endcap) and combine them for the final result. The exact mathematical relation between signal and normalization for the upper limit determination is described in more detail later.

The determination of the signal efficiency in this analysis depends on Monte Carlo (MC) simulation. Therefore we validate the MC simulation through two samples of fully reconstructed B decays. The decay $B^\pm \to J/\psi K^\pm$ provides a high-statistics sample to allow fine-grained comparisons. The decay $B_s^0 \to J/\psi \phi$ is essential to compare B_s^0 mesons in data and MC simulations and to estimate systematic uncertainties for the analysis efficiency. We refer to the $B_s^0 \to J/\psi \phi$ sample as 'control sample' below.

This analysis is based on charged particles measured with the pixel and strip trackers and the muon system. The analysis is not affected by pileup , because the excellent spatial vertex resolution, obtained with the pixel detector, allows a good separation between different primary vertices.

3 Datasets and Trigger

Table 2 summarizes the names of the official datasets used in the analysis.

Table 2: Dataset names of official samples used in the analysis.

Official MC datasets
signal
/BsToMuMu_BsFilter_8TeV-pythia6-evtgen/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
/BdToMuMu_BdFilter_8TeV-pythia6-evtgen/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
control sample
/BsToJPsiPhi_2K2MuPtEtaFilter_8TeV-pythia6-evtgen/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
/BsToJPsiPhi_2K2MuFilter_8TeV-pythia6-evtgen/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
normalization sample
/BuToJPsiK_K2MuPtEtaEtaFilter_8TeV-pythia6-evtgen/Summer12_DR53X-PU_S10_START53_V7A-v2/AODSIM
/BuToJPsiK_K2MuFilter_8TeV-pythia6-evtgen/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
rare decays
/BsToPiPi_2PiPtEtaFilter_8TeV-pythia6-evtgen/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
/BsToKPi_KPiPtEtaFilter_8TeV-pythia6-evtgen/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
/BsToKMuNu_KMuNuEtaFilter_8TeV-pythia6-evtgen/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
/BsToKK_2KPtEtaFilter_8TeV-pythia6-evtgen/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
/BdToKPi_KPiPtEtaFilter_8TeV-pythia6-evtgen/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
/BdToKK_2KPtEtaFilter_8TeV-pythia6-evtgen/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
/BdToPiPi_2PiPtEtaFilter_8TeV-pythia6-evtgen/Summer12_DR53X-PU_S10_START53_V7A-v2/AODSIM
/LambdaBToPPi_PPiPtEtaFilter_8TeV-pythia6-evtgen/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
/LambdaBToPK_PKPtEtaFilter_8TeV-pythia6-evtgen/Summer12_DR53X-PU_S10_START53_V7A-v2/AODSIM
/BdToPiMuNu_PiMuNuPtEtaFilter_8TeV-pythia6-evtgen/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
/LambdaBToPMuNu_PMuNuPtEtaFilter_8TeV-pythia6-evtgen/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
MuOnia primary dataset
/MuOnia/Run2012A-PromptReco-v1/AOD
/MuOnia/Run2012B-PromptReco-v1/AOD
/MuOnia/Run2012C-PromptReco-v{1,2}/AOD

3.1 Monte Carlo simulation

This analysis uses Monte Carlo (MC) simulation samples to determine the signal and normalization efficiency, and to estimate the peaking background in the signal mass window from rare hadronic decays where both hadrons are misidentified as muons.

The primary MC simulation event samples were generated in the cetrnal production. Table 3 provides a summary of the event samples used in the analysis. The different components are discussed in more detail in the following subsections.

The event generation through minimum-bias processes is very time-consuming, but necessary for this analysis, as isolation variables have been found crucial for background reduction [18]. It is essential to also include gluon splitting and flavor excitation for $b\bar{b}$ production when studying the impact of these variables: The two b-quarks in gluon fusion events tend to be back-to-back, while those from gluon splitting are closer together in phase space; this has strong influences on the hadronic activity around the dimuon direction.

The background sources that mimic the signal topology can be grouped into three categories. First, $q\bar{q}$ events (where $q = b, c$) with $q \to X\mu\bar{\nu}$ (prompt or cascade) decays of both q hadrons. Second, events where a true muon is combined with a hadron misidentified as a muon (punch-through or in-flight decay of a hadron). And finally, rare B^0, B^+, B^0_s and Λ_b decays, mostly from semileptonic decays. For this analysis, we have simulated only the last of the three cases. The other backgrounds are of a combinatorial nature without structure in the dimuon mass distribution. Therefore, we define regions next to the dimuon signal window(s) to determine in data the combinatorial background an interpolate to the signal window(s).

Table 3: Central MC production event samples used in the analysis. The events in the generator-sample file N_{file}, the equivalent integrated luminosity $\mathcal{L}_{\text{gen}}$, and the branching fraction $\mathcal{B}$ is given.

Sample	N_{file}	$\mathcal{L}_{\text{gen}}[\text{fb}^{-1}]$	$\mathcal{B}$	Ref.
$B^0_s \to \mu^+\mu^-(MC)$	2939886	7405.3	$(3.20 \pm 0.19) \times 10^{-9}$	[2]
$B^0 \to \mu^+\mu^-(MC)$	507255	14767.2	$(1.00 \pm 0.10) \times 10^{-10}$	[2]
$B^\pm \to J/\psi K^\pm$ (MC)	4862178	10.3	$(6.00 \pm 0.18) \times 10^{-5}$	[28]
$B^0_s \to J/\psi \phi$ (MC)	1635017	17.5	$(3.20 \pm 1.02) \times 10^{-5}$	[28]
$B^0_s \to K^+K^-$	4104882	16.2	$(2.54 \pm 0.38) \times 10^{-5}$	[29]
$B^0_s \to \pi^+K^-$	593173	15.3	$(5.00 \pm 1.10) \times 10^{-6}$	[29]
$B^0_s \to \pi^+\pi^-$	305086	33.7	$(7.30 \pm 1.39) \times 10^{-7}$	[29]
$B^0 \to K^+K^-$	48205	16.2	$(1.30 \pm 1.00) \times 10^{-7}$	[29]
$B^0 \to K^+\pi^-$	5703439	15.2	$(1.95 \pm 0.06) \times 10^{-5}$	[29]
$B^0 \to \pi^+\pi^-$	1881506	17.7	$(5.11 \pm 0.20) \times 10^{-6}$	[29]
$\Lambda^0_b \to p\pi^-$	175744	9.0	$(3.50 \pm 1.01) \times 10^{-6}$	[28]
$\Lambda^0_b \to pK^-$	318990	11.0	$(5.50 \pm 1.38) \times 10^{-6}$	[28]
$B^0_s \to K^-\mu^+\nu$	5992326	16.8	$(1.40 \pm 0.07) \times 10^{-4}$	[29]
$B^0 \to \pi^-\mu^+\nu$	17498188	13.9	$(1.40 \pm 0.07) \times 10^{-4}$	[29]
$\Lambda^0_b \to p\mu^-\bar{\nu}$	1508130	5.8	$(3.00 \pm 0.99) \times 10^{-4}$	[30]

The rare decays could potentially lead to sizable background contributions which cannot be determined based on data sidebands. Two cases can be distinguished: (1) Peaking backgrounds from rare decays, where a heavy particle decays into a pair of hadrons. Examples for these decays include $B^0_s \to K^+K^-$, $\Lambda_b \to pK^-$. (2) Non-peaking semileptonic backgrounds from rare B^0, B^+, B^0_s, and Λ_b decays. The invariant dimuon mass distribution for these decays is a continuum with an upper edge at the mass of the decaying particle; the finite momentum resolution could lead to events reconstructed in the $B^0_s \to \mu^+\mu^-$ or $B^0 \to \mu^+\mu^-$ signal mass windows. Because (even Cabibbo-suppressed) semileptonic decays have branching fractions several orders of magnitude above $\mathcal{B}(B^0_s \to \mu^+\mu^-)$, this background could be problematic. For each decay channel, events were generated and analyzed without requiring explicit muon identification. The misidentification probability (and muon identification efficiency, when one final state particle is a muon) were applied as weighting factors at the end.

3.2 Data

The data for this analysis were taken in 2012. We use the `PromptReco` processing datasets.

The signal, normalization, and control sample are analyzed in the `MuOnia` primary dataset. These datasets were analyzed with the release `CMSSW_5_3_2_patch1` with global tags `GR_P_V32::All` for all of the `PromptReco` dataset.

For the analysis we use the official JSON file:

`Cert_190456-203002_8TeV_PromptReco_Collisions12_JSON_MuonPhys`

3.3 Trigger

The high-level trigger selection is based on two L3 muons. With increasing luminosity the threshold of the muon $p_\perp$ has been raised and additional requirements (on the dimuon mass, the distance of closest approach between the two muons, and the dimuon $p_\perp$) were implemented. Our trigger selection is summarized in table 4. As can be seen from the table, the data taken in 2012 used 4 trigger selections for the signal decay and also 4 selections for the normalization and control channels.

For the signal the following triggers, optimized for the B_s^0 detection, have been used[3]:

- `HLT_DoubleMu3_4_Dimuon5_Bs_Central_v2`
- `HLT_DoubleMu3_4_Dimuon5_Bs_Central_v3`
- `HLT_DoubleMu3_4_Dimuon5_Bs_Central_v4`
- `HLT_DoubleMu3_4_Dimuon5_Bs_Central_v5`

  The above triggers are required to have two L3 muons having for the subleading muon $p_{\perp_\mu} > 3\,\text{GeV}$ and for the leading muon $p_{\perp_\mu} > 4\,\text{GeV}$, dimuon pair having $p_{\perp dimuon} > 5\,\text{GeV}$ and a mass window of $4.8 - 6.0\,\text{GeV}$. The *Central* part of the trigger name refers to the requirement that the dimuon pair be limited to $|\eta_\mu| < 1.8$.

- `HLT_DoubleMu3p5_4_Dimuon5_Bs_Central_v2`
- `HLT_DoubleMu3p5_4_Dimuon5_Bs_Central_v3`
- `HLT_DoubleMu3p5_4_Dimuon5_Bs_Central_v4`
- `HLT_DoubleMu3p5_4_Dimuon5_Bs_Central_v5`

  The above triggers are required to have two L3 muons having for the subleading muon $p_{\perp_\mu} > 3.5\,\text{GeV}$ and for the leading muon $p_{\perp_\mu} > 4\,\text{GeV}$, dimuon pair having $p_{\perp dimuon} > 5\,\text{GeV}$ and a mass window of $4.8 - 6.0\,\text{GeV}$. The *Central* part of the trigger name refers to the requirement that the dimuon pair be limited to $|\eta_\mu| < 1.8$.

- `HLT_DoubleMu4_Dimuon7_Bs_Forward_v2`
- `HLT_DoubleMu4_Dimuon7_Bs_Forward_v3`
- `HLT_DoubleMu4_Dimuon7_Bs_Forward_v4`
- `HLT_DoubleMu4_Dimuon7_Bs_Forward_v5`

  The above triggers are required to have two L3 muons having for both muons $p_{\perp_\mu} > 4\,\text{GeV}$, dimuon pair having $p_{\perp dimuon} > 7\,\text{GeV}$ and a mass window of $4.8 - 6.0\,\text{GeV}$. The *Forward* part of the trigger name refers to the requirement that the dimuon pair be limited to $|\eta_\mu| < 2.2$.

For the normalization and control channels the following triggers, optimized for the J/ψ detection have been used:

- `HLT_DoubleMu4_Jpsi_Displaced_v9`
- `HLT_DoubleMu4_Jpsi_Displaced_v10`
- `HLT_DoubleMu4_Jpsi_Displaced_v11`
- `HLT_DoubleMu4_Jpsi_Displaced_v12`

  Two L3 muons, each with $p_{\perp_\mu} > 4\,\text{GeV}$ and $|\eta_\mu| < 2.2$ with the dimuon pair having $p_{\perp dimuon} > 6.9\,\text{GeV}$. The vertex fit minimum probability cut is 0.15. The rest of the cuts are: mass window is $2.9 - 3.3\,\text{GeV}$, a lifetime significant of > 3, cosine of the pointing angle > 0.9 and the DCA) of less than 0.5 cm.

The number of primary vertices is shown in Fig. 2 for the signal and normalization samples.

[3]We use units where $c = 1$.

Table 4: HLT paths for the signal decay (top part of the table) and the normalization and control samples (bottom part of the table).

Year	Run range	L [pb^{-1}]	HLT Path	L1 seed
		Signal sample		
2012	190456–194712	2.306	HLT_DoubleMu3_4_Dimuon5_Bs_Central_v2	L1_DoubleMu0er_HighQ
		2.313	HLT_DoubleMu3p5_4_Dimuon5_Bs_Central_v2	L1_DoubleMu3er_HighQ_WdEta22
		2.313	HLT_DoubleMu4_Dimuon7_Bs_Forward_v2	L1_DoubleMu3er_HighQ_WdEta22
	194735–196531	3.296	HLT_DoubleMu3_4_Dimuon5_Bs_Central_v3	L1_DoubleMu0er_HighQ
		3.296	HLT_DoubleMu3p5_4_Dimuon5_Bs_Central_v3	L1_DoubleMu3er_HighQ_WdEta22
		3.296	HLT_DoubleMu4_Dimuon7_Bs_Forward_v3	L1_DoubleMu3er_HighQ_WdEta22
	198049–199608	1.825	HLT_DoubleMu3_4_Dimuon5_Bs_Central_v4	L1_DoubleMu0er_HighQ
		1.825	HLT_DoubleMu3p5_4_Dimuon5_Bs_Central_v4	L1_DoubleMu3er_HighQ_WdEta22
		1.825	HLT_DoubleMu4_Dimuon7_Bs_Forward_v4	L1_DoubleMu3er_HighQ_WdEta22
	199698–203002	5.398	HLT_DoubleMu3_4_Dimuon5_Bs_Central_v5	L1_DoubleMu0er_HighQ
		5.398	HLT_DoubleMu3p5_4_Dimuon5_Bs_Central_v5	L1_DoubleMu3er_HighQ_WdEta22
		5.398	HLT_DoubleMu4_Dimuon7_Bs_Forward_v5	L1_DoubleMu3er_HighQ_WdEta22
		Normalization and control sample		
2012	190456–194712	2.313	HLT_DoubleMu4_Jpsi_Displaced_v9	L1_DoubleMu3er_HighQ_WdEta22
	194735–196531	3.296	HLT_DoubleMu4_Jpsi_Displaced_v10	L1_DoubleMu3er_HighQ_WdEta22
	198049–199608	1.825	HLT_DoubleMu4_Jpsi_Displaced_v11	L1_DoubleMu3er_HighQ_WdEta22
	199698–203002	5.398	HLT_DoubleMu4_Jpsi_Displaced_v12	L1_DoubleMu3er_HighQ_WdEta22

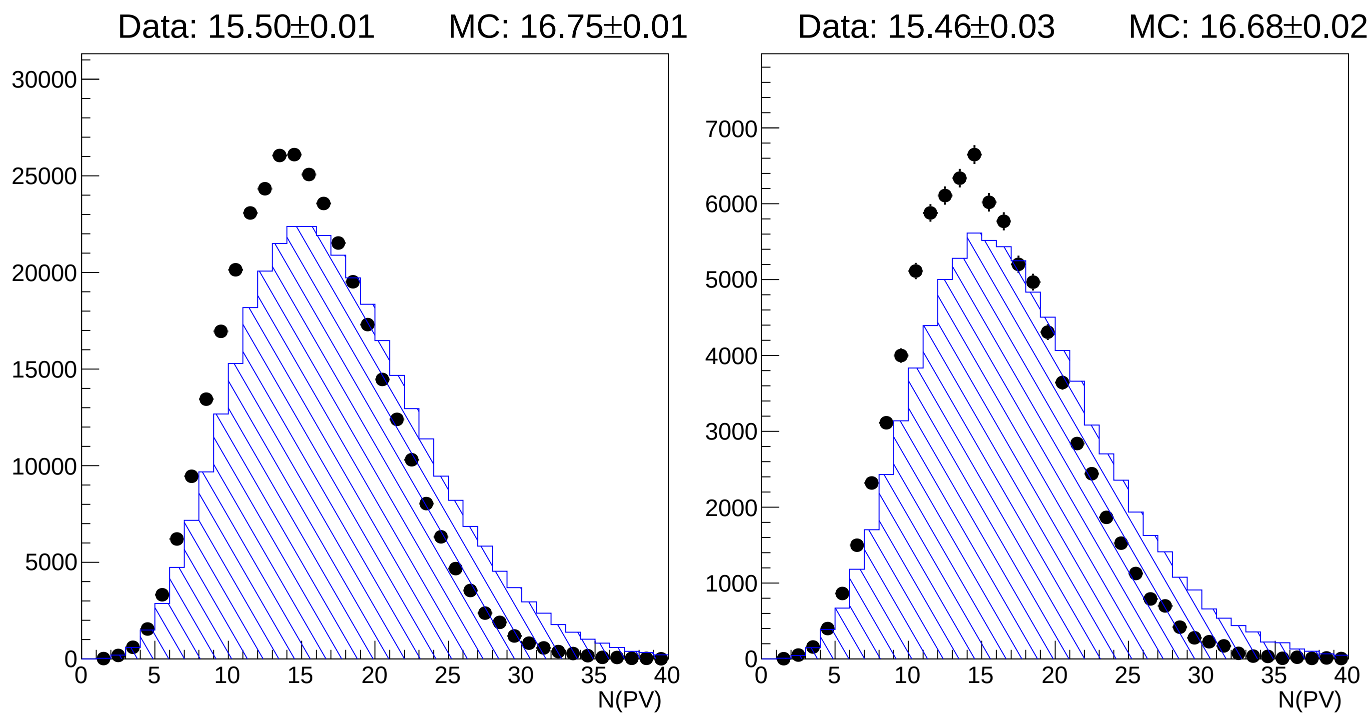

Figure 2: Number of primary vertices in data and MC simulation for $B^\pm \rightarrow J/\psi K^\pm$ candidates in the barrel (left) and the endcap (right). The numbers printed on top of the histograms indicate the mean and its uncertainty for data and MC simulation.

4 Variables

4.1 Summary of variables

The following variables, explained in more detail in the subsequent sections, have been used in the study and (sometimes) training of the BDT (the name in parentheses indicates the name used in the code and appears in TMVA-generated plots):

- $p_{\perp\mu,1}$ (`m1pt`): The $p_\perp$ of the muon with the higher $p_\perp$
- $p_{\perp\mu,2}$ (`m2pt`): The $p_\perp$ of the muon with the lower $p_\perp$
- $\eta_{\mu,1}$ (`m1eta`): The η of the muon with the higher $p_\perp$
- $\eta_{\mu,2}$ (`m2eta`): The η of the muon with the lower $p_\perp$
- $p_{\perp B}$ (`pt`): The $p_\perp$ of the B candidate
- η_B (`eta`): The η of the B candidate
- $\ell_{3D}/\sigma(\ell_{3D})$ (`fls3d`): The flight length significance of the B candidate
- α (`alpha`): The pointing angle of the B candidate (angle between candidate momentum and flight direction in 3D)
- χ^2/dof (`chi2dof`): The vertex fit χ^2 of the B candidate
- I (`iso`): The isolation of the B candidate
- $N_{\mathrm{trk}}^{\mathrm{close}}$ (`closetrk`): The number of tracks in the vicinity of the B decay vertex
- d_{ca}^0 (`docatrk`): The minimum distance of closest approach of a track in the event
- d_{ca} (`maxdoca`): The distance of closest approach between the two muon tracks
- δ_{3D} (`pvip`): The 3D impact parameter of the B candidate with respect to the primary vertex
- $\delta_{3D}/\sigma(\delta_{3D})$ (`pvips`): The significance of the 3D impact parameter of the B candidate with respect to the primary vertex

These variables have been used in the baseline cut-and-count selection [31] and are well described in the MC simulation (see below). The kinematical variables of the muons ($p_{\perp\mu,1}$, $p_{\perp\mu,2}$, $\eta_{\mu,1}$, $\eta_{\mu,2}$) were not included in the training of the BDT.

4.2 Dimuon sample reconstruction

For the offline event selection, variables related to the muons, the primary vertex, and the B_s^0 candidate[4] with its associated secondary vertex are calculated.

4.2.1 Primary Vertex

The primary vertex is determined with the standard algorithm [32] used in CMS. We use `OfflinePrimaryVertices` which we refit without the signal tracks of the candidates.

4.2.2 Muon variables

The inner track of the muon candidate is required to be of `highPurity` quality [33].

Our default muon identification is based on the requirement that a muon candidate satisfy the 'tight' muon selection [34], updated for the new selection in 2012. In particular we require that

- more than one muon stations is matched

[4] We speak of B_s^0 candidate, but this refers also to B_d^0 candidates (in a different mass window).

- the transverse impact parameter with respect to the beam spot is less than 2 mm
- at least one pixel layer has a hit
- the number of tracker layers with hits must be larger than 5.

As a very basic preselection at the EDanalyzer level (superseeded by more stringent requirements for the BDT preselection and the full analysis) we require $p_{\perp_\mu} > 3.0\,\mathrm{GeV}$ and $|\eta_\mu| < 2.4$ for both muons. We label the leading muon (the muon with the higher $p_\perp$) as μ_1 and the subleading muon (the muon with the lower $p_\perp$) as μ_2.

For the analysis we keep candidates with an invariant mass $4.5 < m_{\mu_1\mu_2} < 6.5\,\mathrm{GeV}$, even though the full mass range kept in the trigger path is smaller than this. In this region, we define a blinding window that covers both the $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ signal regions

$$
\begin{aligned}
\text{overall mass window} &: \; 4.900 < m_{\mu_1\mu_2} < 5.900\,\mathrm{GeV} \\
B^0 \to \mu^+\mu^- \text{signal window} &: \; < m_{\mu_1\mu_2} < \;\mathrm{GeV} \\
B_s^0 \to \mu^+\mu^- \text{signal window} &: \; < m_{\mu_1\mu_2} < \;\mathrm{GeV} \\
\text{blinding window} &: \; 5.200 < m_{\mu_1\mu_2} < 5.450\,\mathrm{GeV}.
\end{aligned}
\tag{2}
$$

The sideband, used to study the background in data, is defined as the overall mass window minus the blinding window. The upper and lower sidebands have different size, and are expected to be populated by dimuons of different origin: the upper sideband contains combinatorial and Drell-Yan background, the lower sideband contains additionally background from decays like $B \to \mu^+\mu^- X$ and $B_s^0 \to K\mu\nu$, with the kaon misidentified as muon.

We combine pairs of unlike-charge muons and subject them to a (secondary) vertex fit. We require the dimuon candidate to fulfill $p_{\perp,B} > 7.5\,\mathrm{GeV}$.

4.2.3 Candidate Vertexing

The B decay vertex is determined with the standard CMSSW kinematic vertex fitting package [35]. We apply a geometric constraint, but no mass constraints. To select well-reconstructed vertex fits, we require that the fit returns a valid candidate and that the vertex fit χ^2 is larger than zero[5]. We have also included (for study purposes only) the maximum distance $d^{\max}$ between the B candidate tracks; this variable is strongly correlated to the fit quality given by the $\chi^2/Ndof$.

From the candidate's secondary vertex and its momentum we select a matching primary vertex based on the distance of closest approach along z[6]. This longitudinal impact parameter l_z and its significance $l_z/\sigma(l_z)$ (the value divided by its error) could be used useful to reject outliers. Instead, the 3D impact parameter δ_{3D} and its significance $\delta_{3D}/\sigma(\delta_{3D})$ are used in the event selection. For the selected primary vertex, the average track weight $\langle w \rangle$ is computed as $(Ndof+2)/(2N_{trk}^{PV})$. This quantity should be 'close' to one for good primary vertices [7]. The agreement between data and MC simulation for this variable is not outstanding (see sections 4.3 and 4.4); we apply this selection criterion only very loosely.

We compute the flight distance and its error in in three-dimensional space (ℓ_{3d}) with the standard CMSSW tools. The flight length significance is computed as the ratio of the flight length to its error. We require that the flight length is smaller than 2 cm and its significance is not NaN.

[5] This is an entirely technical sanity check and has no influence on 'valid' well-reconstructed B_s^0 candidates.
[6] This selection is done before the primary vertex is refitted.
[7] Private communication (Wolfram Erdmann).

The pointing angle α_{3D} is the angle in three dimensions between the B_s^0 momentum and the vector from the primary vertex to the B_s^0 secondary vertex.

4.2.4 Isolation

The isolation I is determined from the B_s^0 candidate transverse momentum and other charged tracks in a cone with radius $\Delta R = 0.7$ around the B_s^0 momentum as follows:

$$I = \frac{p_\perp(B_s^0)}{p_\perp(B_s^0) + \sum_{trk} |p_\perp|}, \tag{3}$$

where the sum includes tracks with $p_\perp > 0.9\,\mathrm{GeV}$ that are

- not part of the B_s^0 candidate
- (1) associated to the same primary vertex as the B_s^0 candidate or (2) are not associated to any primary vertex but have a distance of closest approach $d_{ca} < 500\,\mu\mathrm{m}$ to the B_s^0 secondary vertex

In addition to this isolation, we use two further variables to reject candidates arising from partially reconstructed B decays:

- The minimum distance of closest approach of tracks (either associated to no primary vertex or to the same primary vertex as the B candidate) to the candidate vertex, d_{ca}^0.
- The number N_{trk} of close tracks with $d_{ca} < 300\,\mu\mathrm{m}$ and $p_\perp > 0.5\,\mathrm{GeV}$ provides additional rejection power against dimuon backgrounds in data.

4.3 Normalization sample reconstruction

The reconstruction of $B^\pm \to J/\psi K^\pm$ candidates starts from two unlike-sign muons, which are combined and vertexed with a `highPurity` track fulfilling $p_\perp >$ GeV and $|\eta| <$. The distance of closest approach between all pairs among the three tracks is required to be less than 1 mm.

For the analysis we keep candidates with an invariant mass $4.5 < m_{\mu_1\mu_2 K} < 6.5\,\mathrm{GeV}$. In this region, we define the $B^\pm \to J/\psi K^\pm$ sidebands and signal region as follows

$$\begin{aligned} \text{overall mass window} &: \; 4.900 < m_{\mu_1\mu_2 K} < 5.900\,\mathrm{GeV} \\ \text{low sideband} &: \; 5.050 < m_{\mu_1\mu_2 K} < 5.150\,\mathrm{GeV} \\ \text{signal region} &: \; 5.200 < m_{\mu_1\mu_2 K} < 5.350\,\mathrm{GeV} \\ \text{high sideband} &: \; 5.400 < m_{\mu_1\mu_2 K} < 5.500\,\mathrm{GeV}. \end{aligned} \tag{4}$$

These mass regions are the basis to define signal and background regions for the sideband-subtraction used in the comparison of $B^\pm \to J/\psi K^\pm$ decays in data and MC simulation.

All three tracks are used in the vertexing. The two muons of the candidate must have an invariant mass $3.0 < m_{\mu_1\mu_2} < 3.2\,\mathrm{GeV}$. The transverse momentum is required to be larger than 7 GeV.

4.4 Control sample reconstruction

We use $B_s^0 \to J/\psi\phi$ candidates

- to validate the MC simulation of exclusive B_s^0 meson decays in data. Agreement for the decay $B_s^0 \to J/\psi\phi$ will be interpreted as motivation to trust the simulation of $B_s^0 \to \mu^+\mu^-$.

- to estimate the systematic uncertainty of the analysis efficiency for $B_s^0 \to \mu^+\mu^-$, based on the level of agreement between data and MC simulation.

The validation of the MC simulation is especially important for the isolation variable, as the description of the B_s^0 hadronization is based on heuristic models.

The decay $B_s^0 \to J/\psi\phi$ has four final state particles, thus lowering on average the muon transverse momenta compared to the signal decay. Since the muon $p_\perp$ requirements stay at the same numerical values like for the signal decay and the normalization channel, the $p_\perp$ distribution of the B_s^0 candidates is expected to be harder. We use the same trigger path as for the normalization sample.

The reconstruction of $B_s^0 \to J/\psi\phi$ candidates starts from two unlike-sign muons, which are combined and vertexed with two unlike-sign `highPurity` tracks fulfilling $p_\perp >$ GeV and $|\eta| <$. The distance of closest approach between all pairs among the four tracks is required to be less than 1 mm. The two tracks are assumed to be kaons.

For the analysis we keep candidates with an invariant mass $4.5 < m_{\mu_1\mu_2KK} < 6.5\,\mathrm{GeV}$. In this region, we define the $B_s^0 \to J/\psi\phi$ sidebands and signal region as follows

$$\begin{aligned} \text{overall mass window} &: \; 4.900 < m_{\mu_1\mu_2KK} < 5.900\,\mathrm{GeV} \\ \text{low sideband} &: \; 5.100 < m_{\mu_1\mu_2KK} < 5.200\,\mathrm{GeV} \\ \text{signal region} &: \; 5.270 < m_{\mu_1\mu_2KK} < 5.470\,\mathrm{GeV} \\ \text{high sideband} &: \; 5.500 < m_{\mu_1\mu_2KK} < 5.600\,\mathrm{GeV}. \end{aligned} \tag{5}$$

These mass regions are the basis to define signal and background regions for the sideband-subtraction used in the comparison of $B_s^0 \to J/\psi\phi$ decays in data and MC simulation.

All four tracks are used in the vertexing. The two muons of the candidate must have an invariant mass $3.0 < m_{\mu_1\mu_2} < 3.2\,\mathrm{GeV}$. The transverse momentum is required to be larger than 7 GeV. The two kaons must have an invariant mass of $0.995 < m_{KK} < 1.045\,\mathrm{GeV}$ and have $\Delta R < 0.25$ in the $\eta\phi$ plane. This selection was devised, motivated, and illustrated in Ref. [31].

4.5 Variable distributions

In Fig. 3 (4) the distribution of the variables used in the BDT are shown for the barrel (endcap). The signal and background distributions are overlaid.

In Fig. 5 the standard TMVA variable correlation plots for signal MC (even) events are shown. The correponding plots for the odd events are similar and show not statistically significant differences. Fig. 6 shows the same plots for the data sideband background events.

4.6 Variable ranking

In Tab. 5 we provide the ranking of the variables before training of the BDT. It is evident that there are statistical fluctuations in the three datasets (see section 5.1 for the motivation to do this).

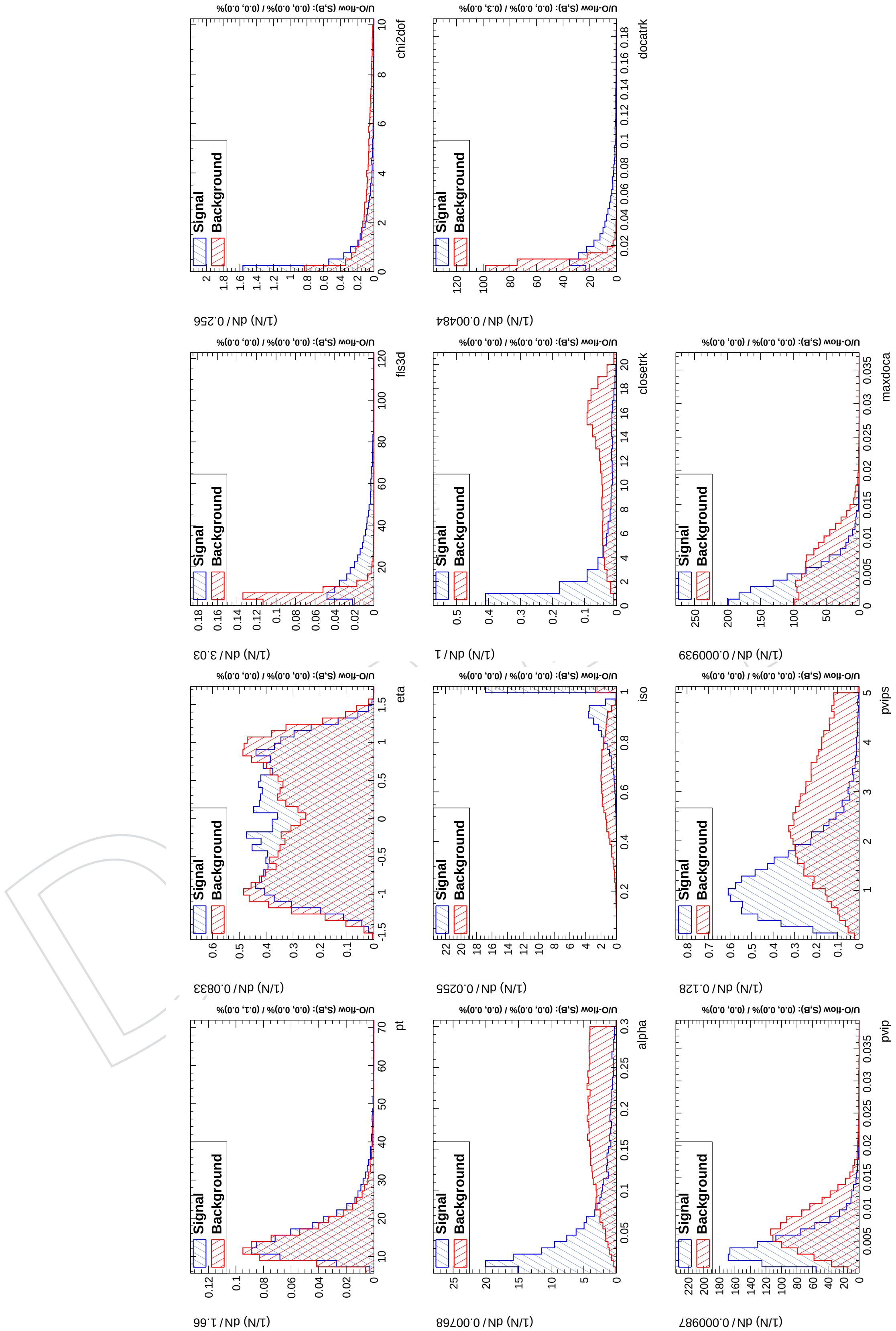

Figure 3: Standard TMVA plot of the input variables for the barrel BDT for signal (blue) and background (red). The background is extracted from data dimuon sidebands.

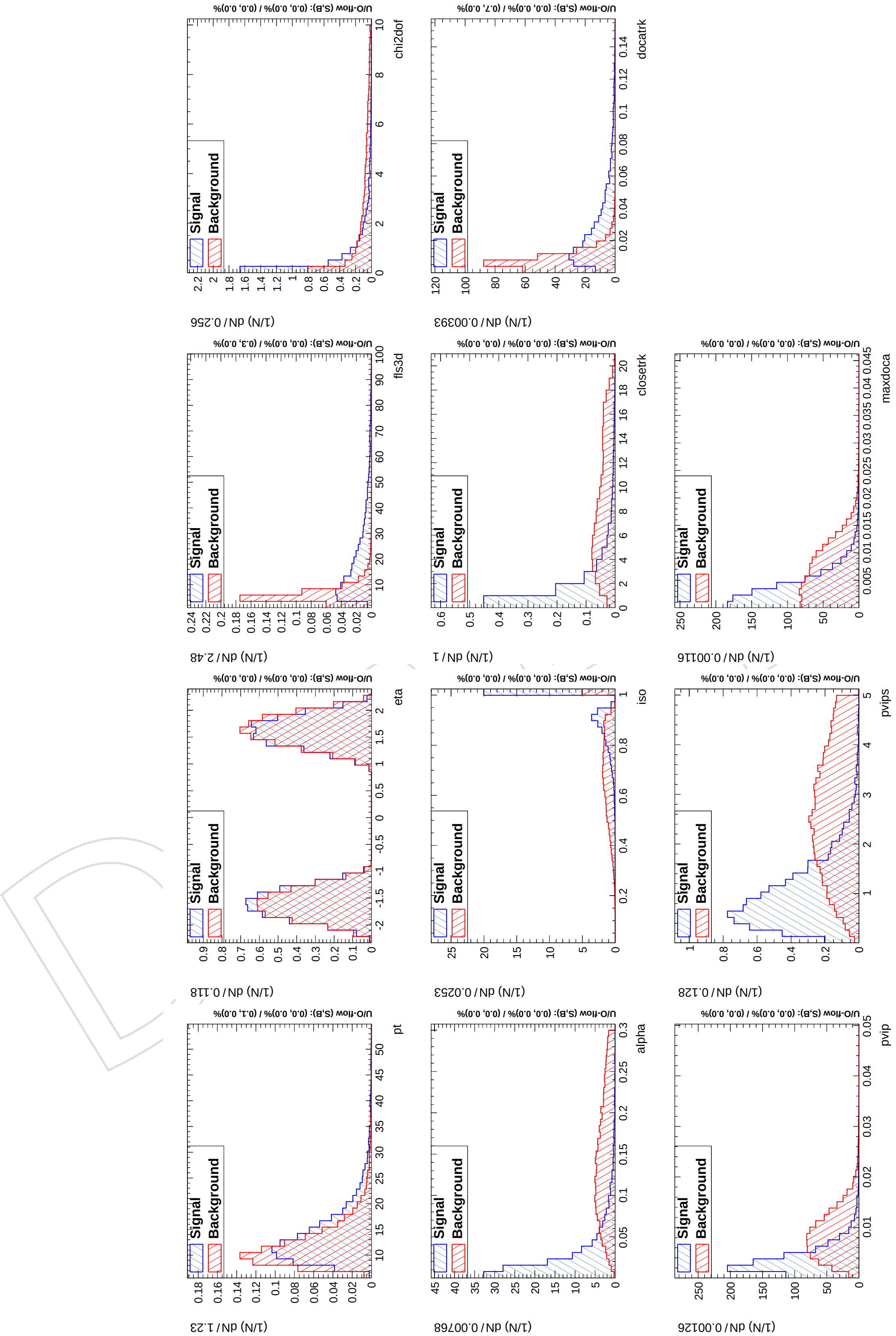

Figure 4: Standard TMVA plot of the input variables for the endcap BDT for signal (blue) and background (red). The background is extracted from data dimuon sidebands.

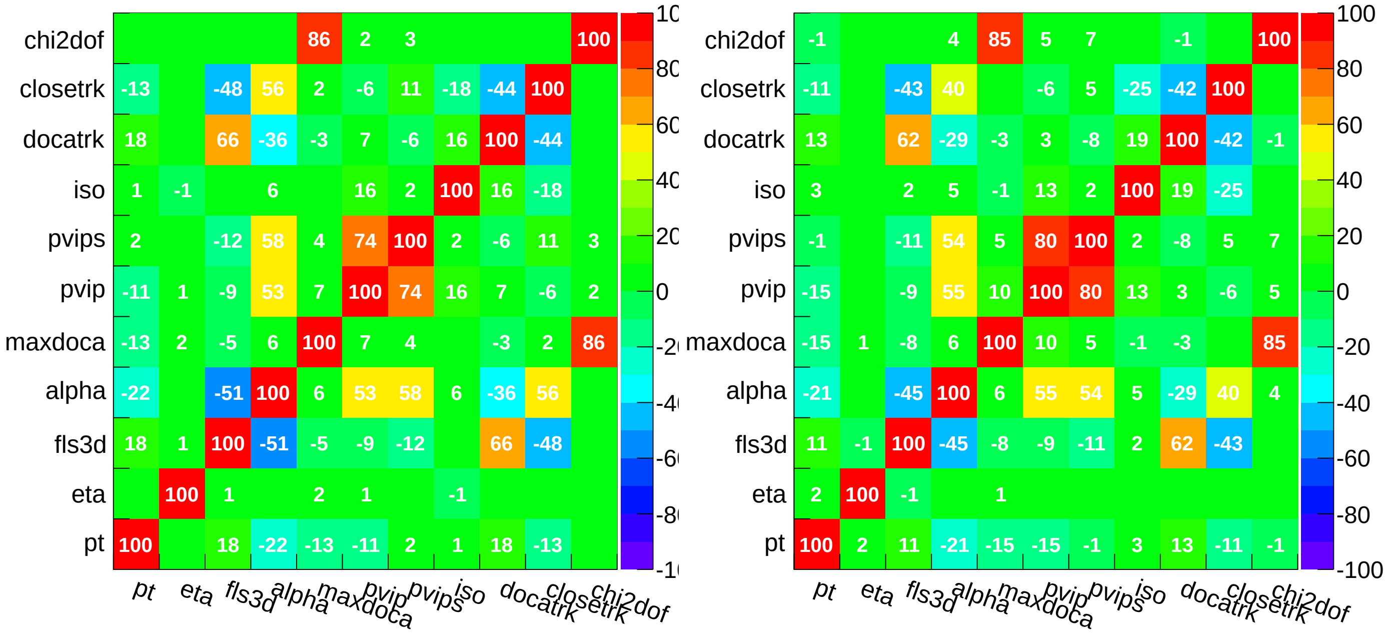

Figure 5: Correlation matrix for signal events in the barrel (left) and the endcap (right).

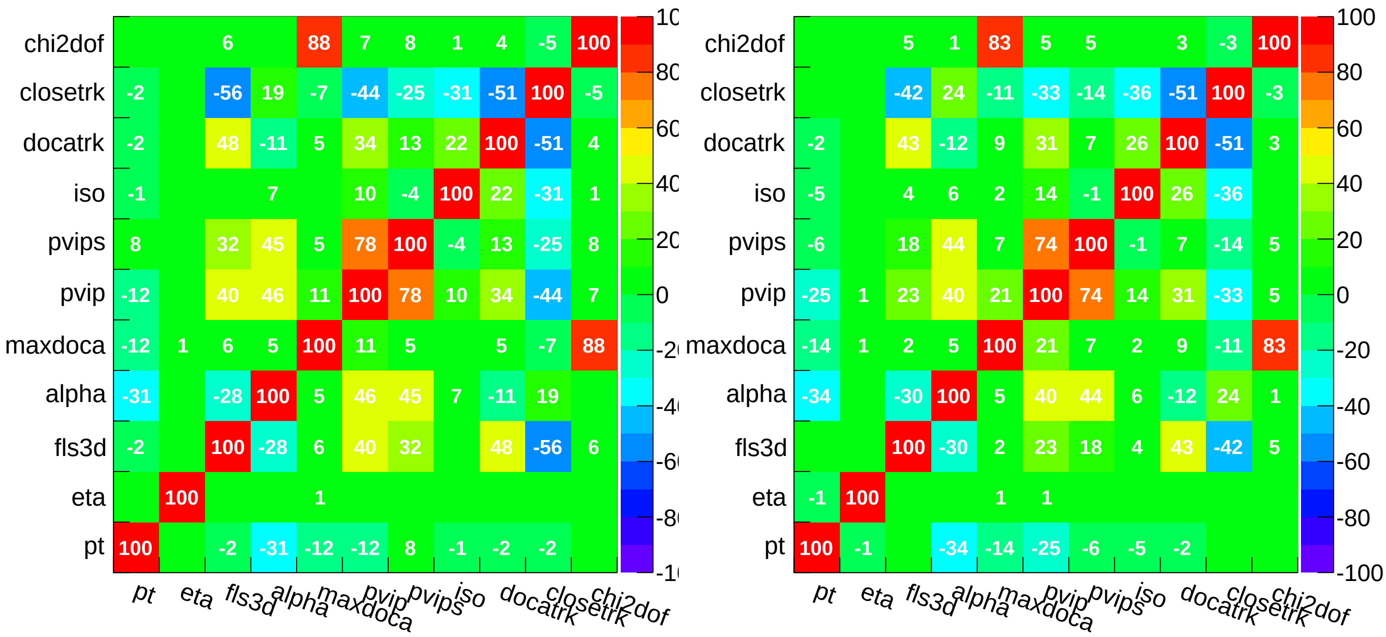

Figure 6: Correlation matrix for background events in the barrel (left) and the endcap (right). The background events are extracted from data dimuon sidebands.

Table 5: Variable ranking for events of the three different event samples.

Barrel						Endcap					
0		1		2		0		1		2	
Variable	Weight	Variable	Weight	Variable	Weight	Variable	Weight	Variable	Weight	Variable	Weight
$N_{\mathrm{trk}}^{\mathrm{close}}$	0.4932	α	0.4852	$N_{\mathrm{trk}}^{\mathrm{close}}$	0.4956	α	0.5851	α	0.5876	α	0.5858
α	0.4846	$N_{\mathrm{trk}}^{\mathrm{close}}$	0.4832	α	0.4928	$N_{\mathrm{trk}}^{\mathrm{close}}$	0.4357	$N_{\mathrm{trk}}^{\mathrm{close}}$	0.4424	$N_{\mathrm{trk}}^{\mathrm{close}}$	0.4342
$\ell_{3D}/\sigma(\ell_{3D})$	0.4012	$\ell_{3D}/\sigma(\ell_{3D})$	0.3975	$\ell_{3D}/\sigma(\ell_{3D})$	0.4029	$\delta_{3D}/\sigma(\delta_{3D})$	0.4149	$\delta_{3D}/\sigma(\delta_{3D})$	0.4267	$\delta_{3D}/\sigma(\delta_{3D})$	0.4223
d_{ca}^{0}	0.3759	d_{ca}^{0}	0.3730	d_{ca}^{0}	0.3793	$\ell_{3D}/\sigma(\ell_{3D})$	0.4026	$\ell_{3D}/\sigma(\ell_{3D})$	0.3998	$\ell_{3D}/\sigma(\ell_{3D})$	0.3979
$\delta_{3D}/\sigma(\delta_{3D})$	0.3120	$\delta_{3D}/\sigma(\delta_{3D})$	0.3172	$\delta_{3D}/\sigma(\delta_{3D})$	0.3210	d_{ca}^{0}	0.3196	d_{ca}^{0}	0.3324	d_{ca}^{0}	0.3228
I	0.3051	I	0.3118	I	0.3000	δ_{3D}	0.3068	δ_{3D}	0.3145	δ_{3D}	0.3142
δ_{3D}	0.1540	δ_{3D}	0.1630	δ_{3D}	0.1633	I	0.2406	I	0.2406	I	0.2412
d_{ca}	0.1256	d_{ca}	0.1235	d_{ca}	0.1245	d_{ca}	0.1566	d_{ca}	0.1589	d_{ca}	0.1509
χ^2/dof	0.1201	χ^2/dof	0.1224	χ^2/dof	0.1193	χ^2/dof	0.1414	χ^2/dof	0.1449	χ^2/dof	0.1361
η_B	0.0147	η_B	0.0161	η_B	0.0165	$p_{\perp B}$	0.0429	$p_{\perp B}$	0.0413	$p_{\perp B}$	0.0397
$p_{\perp B}$	0.0102	$p_{\perp B}$	0.0107	$p_{\perp B}$	0.0103	η_B	0.0041	η_B	0.0027	η_B	0.0047

5 Boosted Decision Tree

A boosted decision tree (BDT) is a multivariate analysis technique that has been widely employed in high-energy physics since a few years. It is basically a (large) 'forest' of decision trees, where the weighted output of all trees is combined into one response, which we shall abbreviate as b. The distribution of b depends on the structure of the BDT, which in turn is determined by the BDT parameters (number of layers, number trees, weighting factor, etc.) and the training sample. Different BDTs will deliver different b distributions even when using the same set of parameters and statistically consistent training samples[8].

A few parameters control the structure and behavior of a BDT

- `Ntrees`: Number of trees in the forest
- `NNodesmax`: Maximum number of nodes in the decision tree. At each node, the split is determined by finding the variable and its cut value providing the best separation between signal and background. The separation is measured by the Gini index, $p(1-p)$, where p is the purity of the node given by the ratio of the sum of the signal weights to all weights.
- `nEventsMin`: Minimum number of events required in a leaf node (a node at the end of a decision flow).
- `nCuts`: Number of steps during the node cut optimization
- `MaxDepth`: Maximum allowed depth of the decision tree
- `AdaBoost` β: Events misclassified in tree $n-1$ are given a different (higher) event weight in the training of tree n. The original weights are multiplied by a common boost weight $\alpha = [(1-f)/f]^{\beta}$, where f is the misclassification rate of tree $n-1$. The weights are (re)normalized such that the sum of weights remains constant. The resulting event classification response is given by

  $$b = \frac{1}{N_{\text{all trees}}} \sum_{i \in \text{all trees}} \ln(\alpha_i) h_i(\vec{x}),$$

  where $h_i(\vec{x})$ is the output of tree i using the input variables $\vec{x}$. Alternative boosting schemes have been studied.

In addition to the setup described above, other choices are possible. However, the differences are not expected to be significant (e.g. using a separation criterion other than Gini index or using bagged decision trees instead of boosting).

5.1 Event sample splitting

It is absolutely imperative to avoid any possible bias when searching for a rare decay with a BDT. It would be best to have train the BDT with a background sample obtained from MC simulation. In such a way, there would be no bias with regard to the variables chosen for the analysis nor in the specific events. Within the CMS collaboration, simulation of the generic dimuon background is not possible because of the large (effective) branching fraction[9] Therefore we use the data dimuon mass sideband events from which we select events to train the BDT.

[8] This can be verified quickly by running toy examples slighly modifying the tutorial macros distributed with TMVA.

[9] It is possible for the LHCb collaboration.

To avoid a possible bias when using background events from data we split the data event sample into three subsets depending on the 'event type'. We define the event type by the remainder of the event number divided by three, i.e. `type = iEvent.id().event()%3`. To be very specific, we work per channel with three BDTs, depending on the event type

- events of type 0: analyzed by BDT0, trained on type-1 events, tested on type-2 events
- events of type 1: analyzed by BDT1, trained on type-2 events, tested on type-0 events
- events of type 2: analyzed by BDT2, trained on type-0 events, tested on type-1 events

This division of the event sample into different types based on the event number yields an unbiased splitting. The event numbers per type for signal and background events are provided in Tab. 6. These numbers are after preselection (muon identification, HLT passed, and the removal of outliers).

Table 6: Number of events per type for signal and background events in the barrel and endcap.

Sample	Type 0	Type 1	Type 2
Signal barrel	10870	10728	10767
Signal endcap	6478	6354	6428
Background barrel	19585	19375	19406
Background endcap	22599	22403	22735

To ensure that the sample splitting into events of different type does not introduce a hidden problem, we compare the variable distributions for the different event types against each other. This is illustrated in figs. 7 and 8 for the barrel BDT in background data sidebands and signal MC, respectively. The Kolmogorov-Smirnov comparisons, printed on top of the plots, indicates no difference between the distributions: The first number compares the distribution for events of type 0 and 1, the second for events of type 1 and 2, and the third number for events of type 2 and 0. In Figs. 9 and 10 the corresponding plots are shown for the endcap BDT.

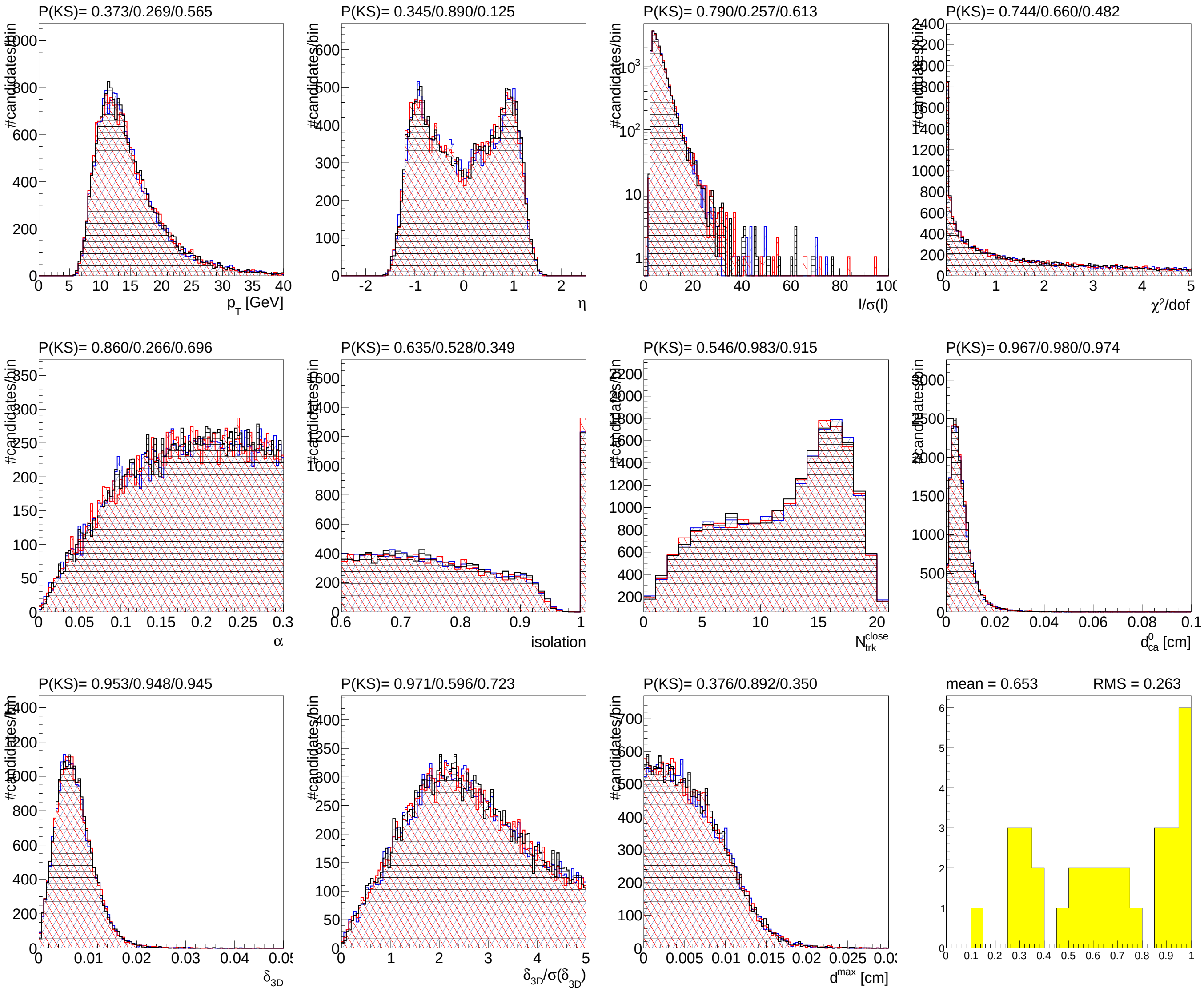

Figure 7: Overlay of BDT training variable distributions in data sideband background for events of the three subsets in the barrel. The plot on the bottom right summarizes all KS probabilities.

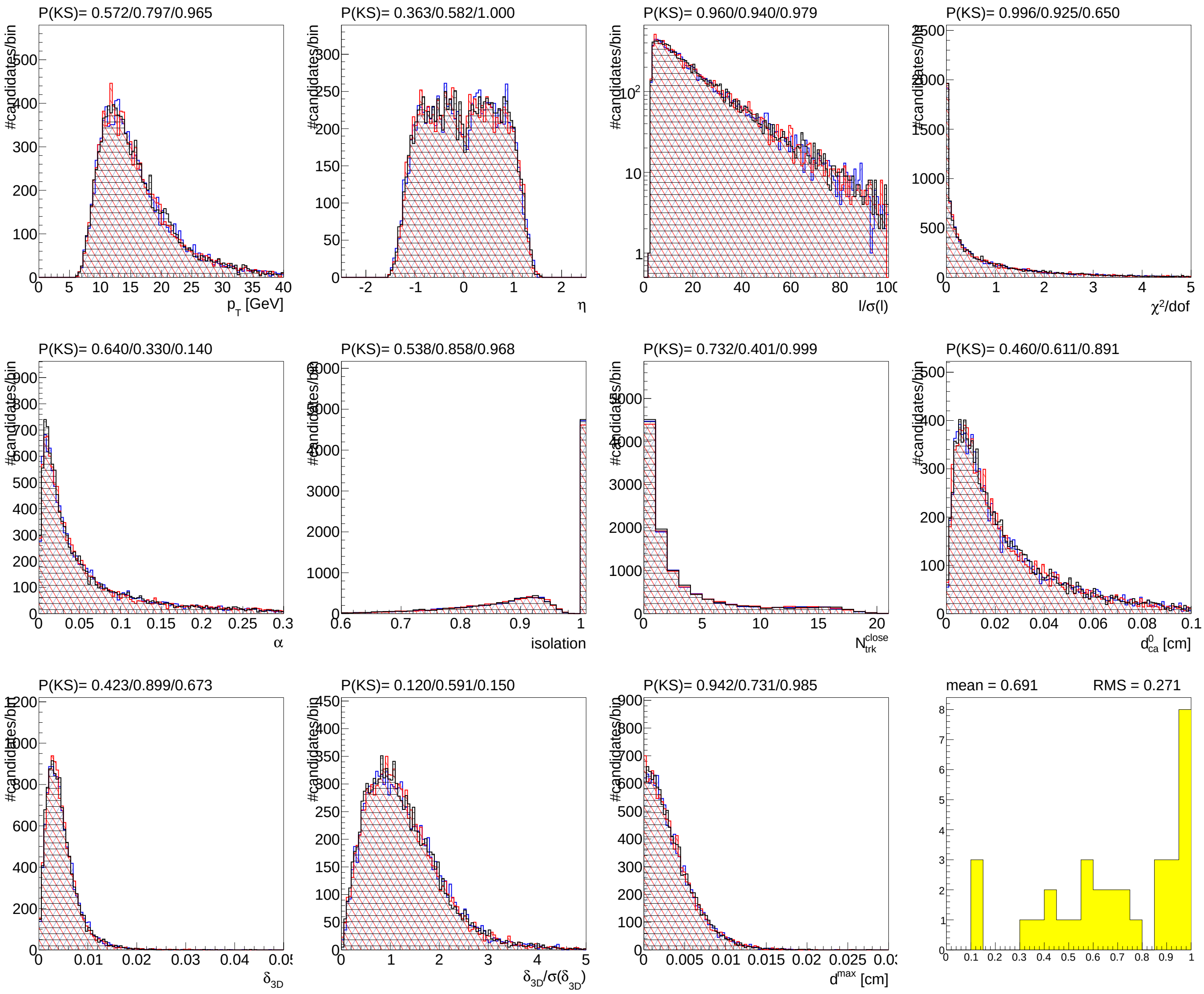

Figure 8: Overlay of BDT training variable distributions in signal MC for events of the three subsets in the barrel. The plot on the bottom right summarizes all KS probabilities.

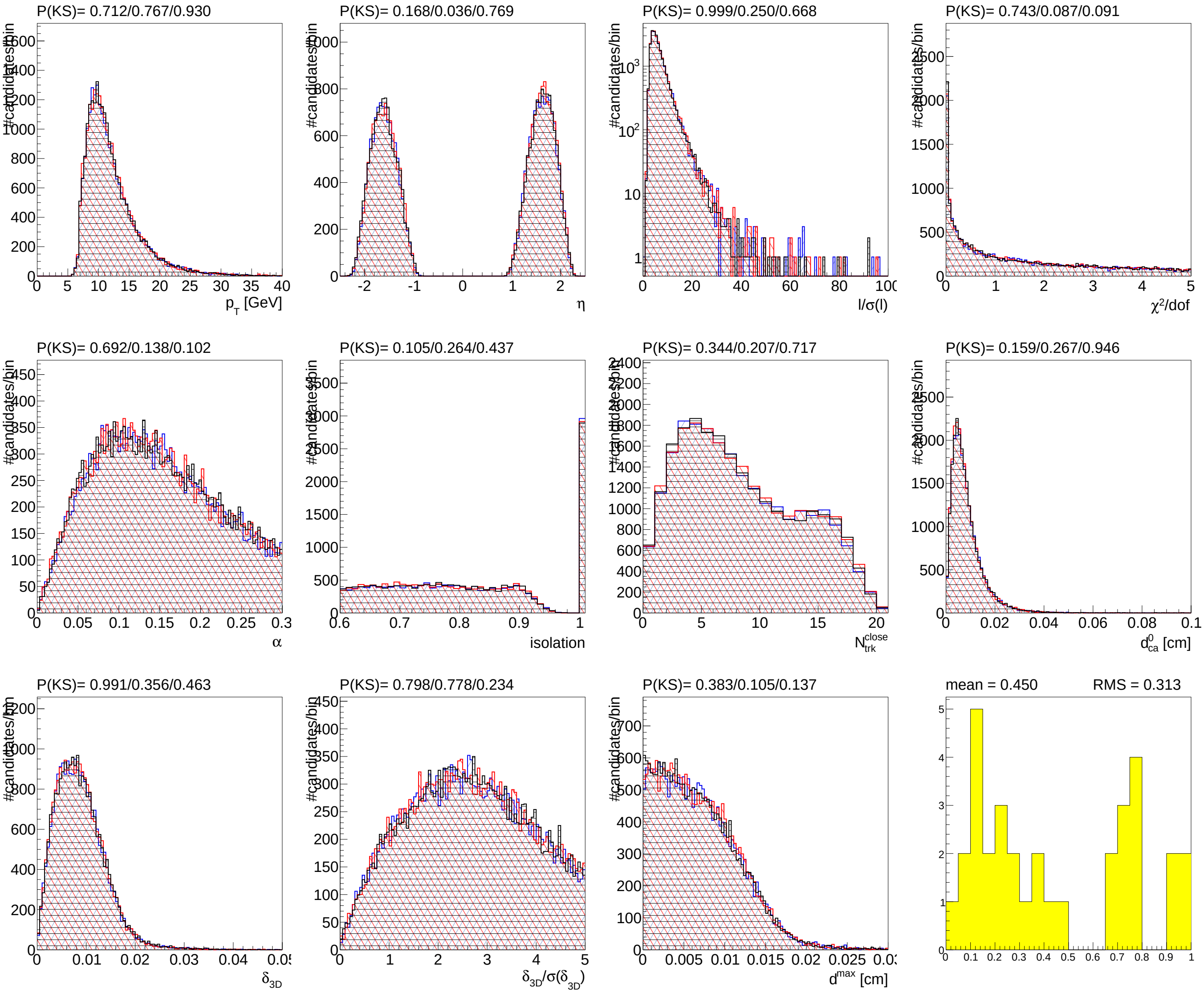

Figure 9: Overlay of BDT training variable distributions in data sideband background for events of the three subsets in the endcap. The plot on the bottom right summarizes all KS probabilities.

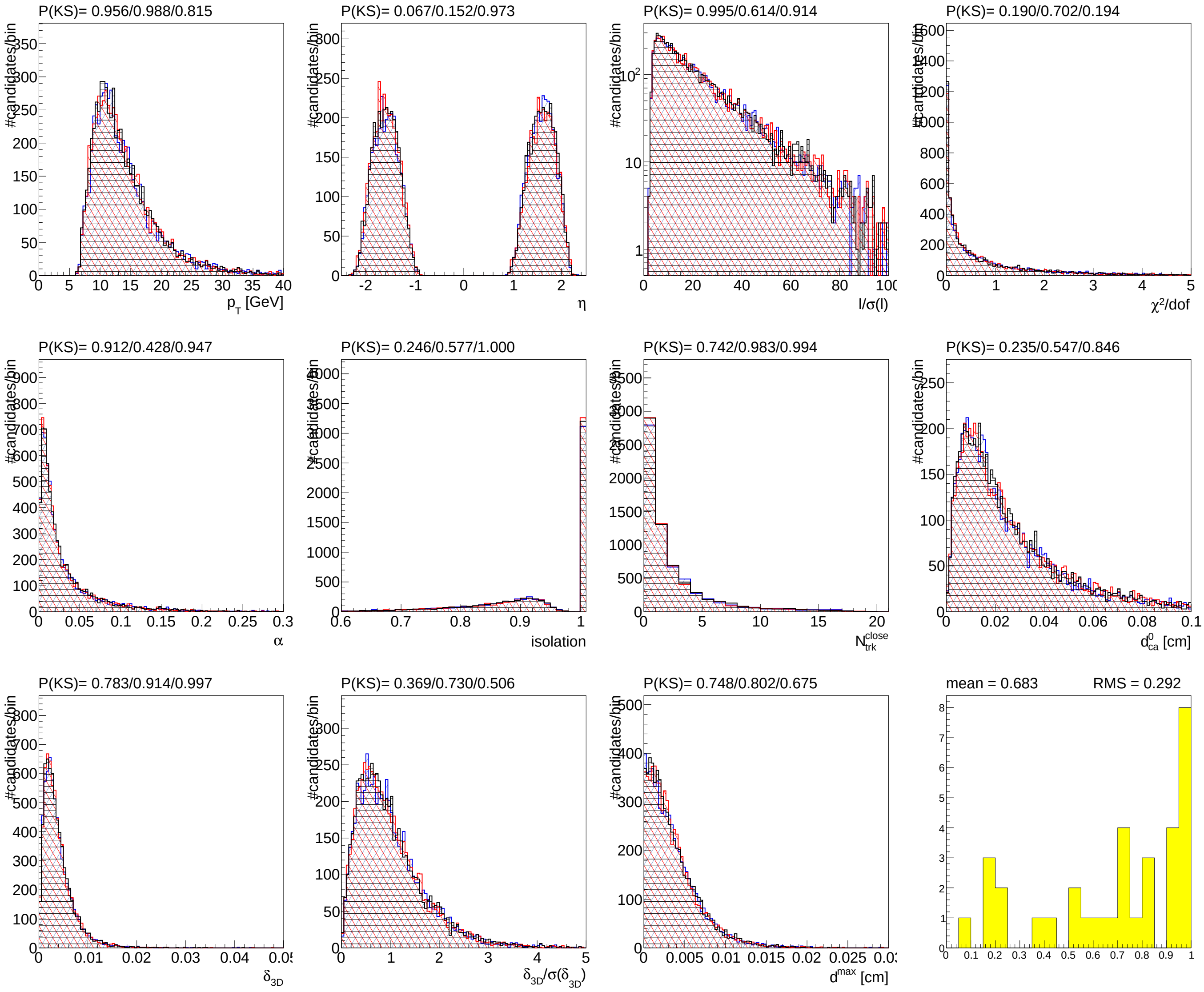

Figure 10: Overlay of BDT training variable distributions in signal MC for events of the three subsets in the endcap. The plot on the bottom right summarizes all KS probabilities.

5.2 Preselection

A preselection, applied for the training samples, is required to remove outliers in the BDT input distributions that could distort the BDT architecture and lead to too coarse selection criteria within a specific decision tree. The preselection also has a considerable influence on the BDT performance and characteristics. In Tab. 7 the preselection requirements applied are summarized.

Table 7: Preselection for the BDT training.

Variable	minimum	maximum	unit
$p_{\perp B}$	5.00	9999.00	GeV
$p_{\perp \mu,1}$	4.00	999.00	GeV
$p_{\perp \mu,2}$	4.00	999.00	GeV
ℓ_{3D}	-	2.00	cm
$\ell_{3D}/\sigma(\ell_{3D})$	0.00	120.00	
χ^2/dof	-	10.00	
δ_{3D}	-	0.10	cm
$\delta_{3D}/\sigma(\delta_{3D})$	-	5.00	
d_{ca}	-	0.10	cm
α	-	0.30	
$N_{\mathrm{trk}}^{\mathrm{close}}$	-	21.00	
d_{ca}^{0}	-	0.25	
I	0.00	-	

5.3 Generic BDT characterization

The values shown in Tab. 8 represent the setup parameters for the barrel and endcap BDT that offered the best significance, defined as $S_{\max} \equiv S/\sqrt{S+B}$, for a specific choice of the selection requirement on the BDT reponse $b > B_{\max}$.

Table 8: BDT Parameters

Parameter	Barrel	Endcap	TMVA default
Ntrees	800	800	200
NEvents	50	50	max(20,NEvtsTrain/NVar^2/10)
Ncuts	20	20	20
MaxDepth	2	2	3
NNodesMax	5	5	1000000
Beta	1.00	1.00	1

As shown in Fig. 11 the BDT converges well within the first 100 decision trees for both the boost weight and the error fraction.

In Fig. 12 the number of 'intermediate' nodes is shown. Trees with 'zero' nodes have been pruned from a decision tree that resulted in two nodes being classified as the same category (i.e. useless). In the next step of the BDT growing, the events are reweighted and the resulting tree is again a 'normal' tree.

In Fig. 13 (14) the cut values for all variables used in the barrel (endcap) BDT are shown.

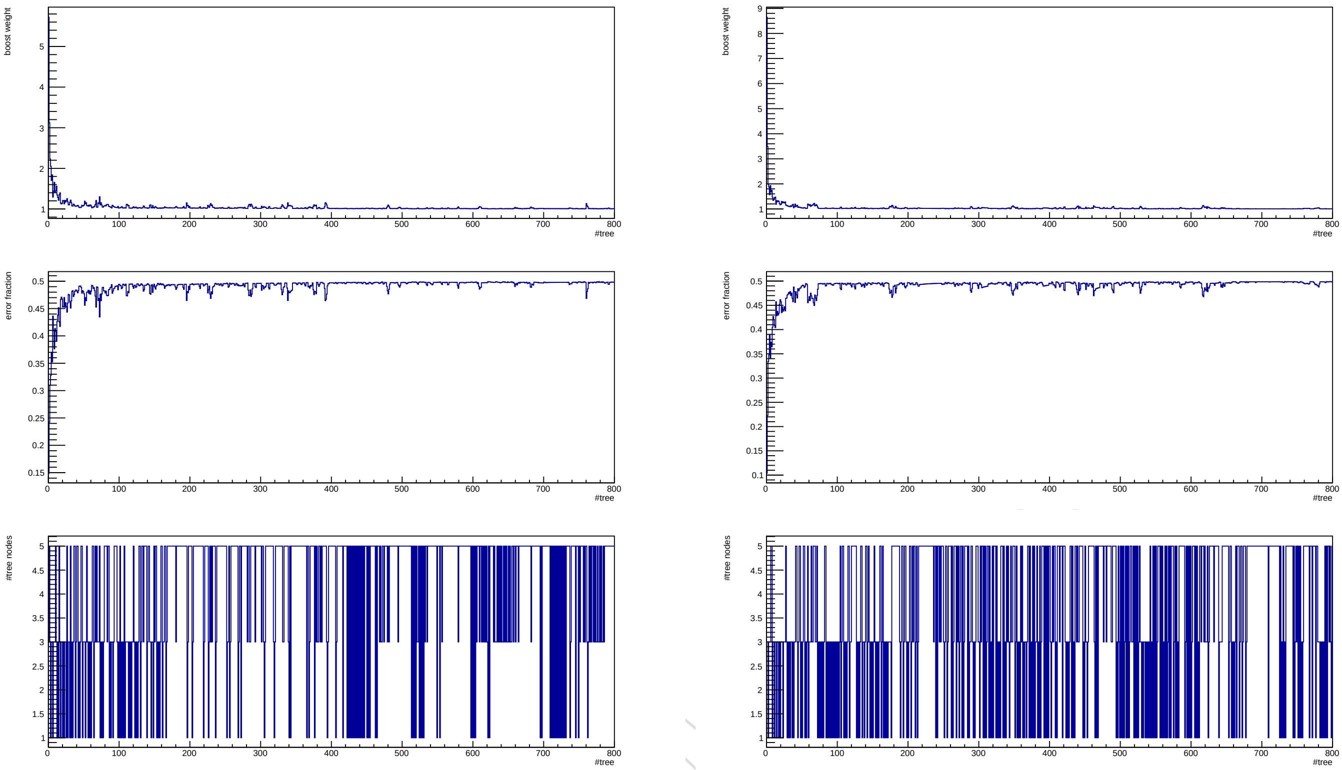

Figure 11: TMVA BDT charaterization plots for the barrel (left) and the endcap (right). Shown versus the tree number is the boost weight (top) and the event misclassification rate (middle), and the number of nodes before pruning (bottom).

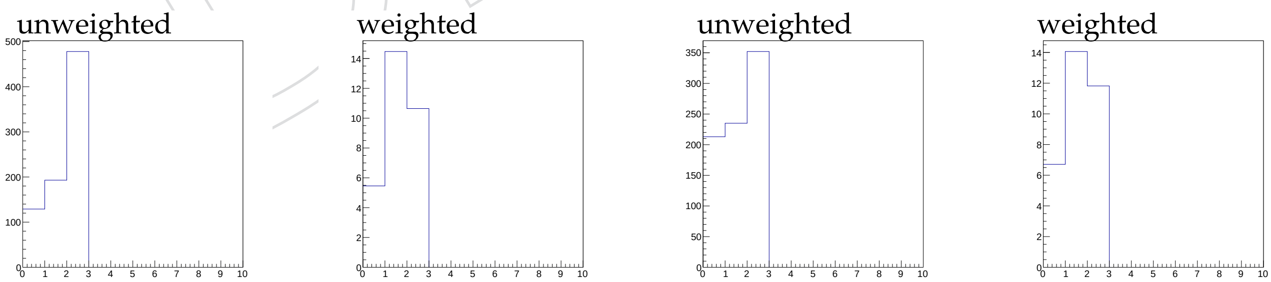

Figure 12: The number of intermediate nodes per decision tree in the barrel (two left plots) and endcap (two right plots). For each channel, the plot on the left shows the unweighted nodes count and the plot on the right shows the weighted node count. The weight corresponds to the weight of the decision tree.

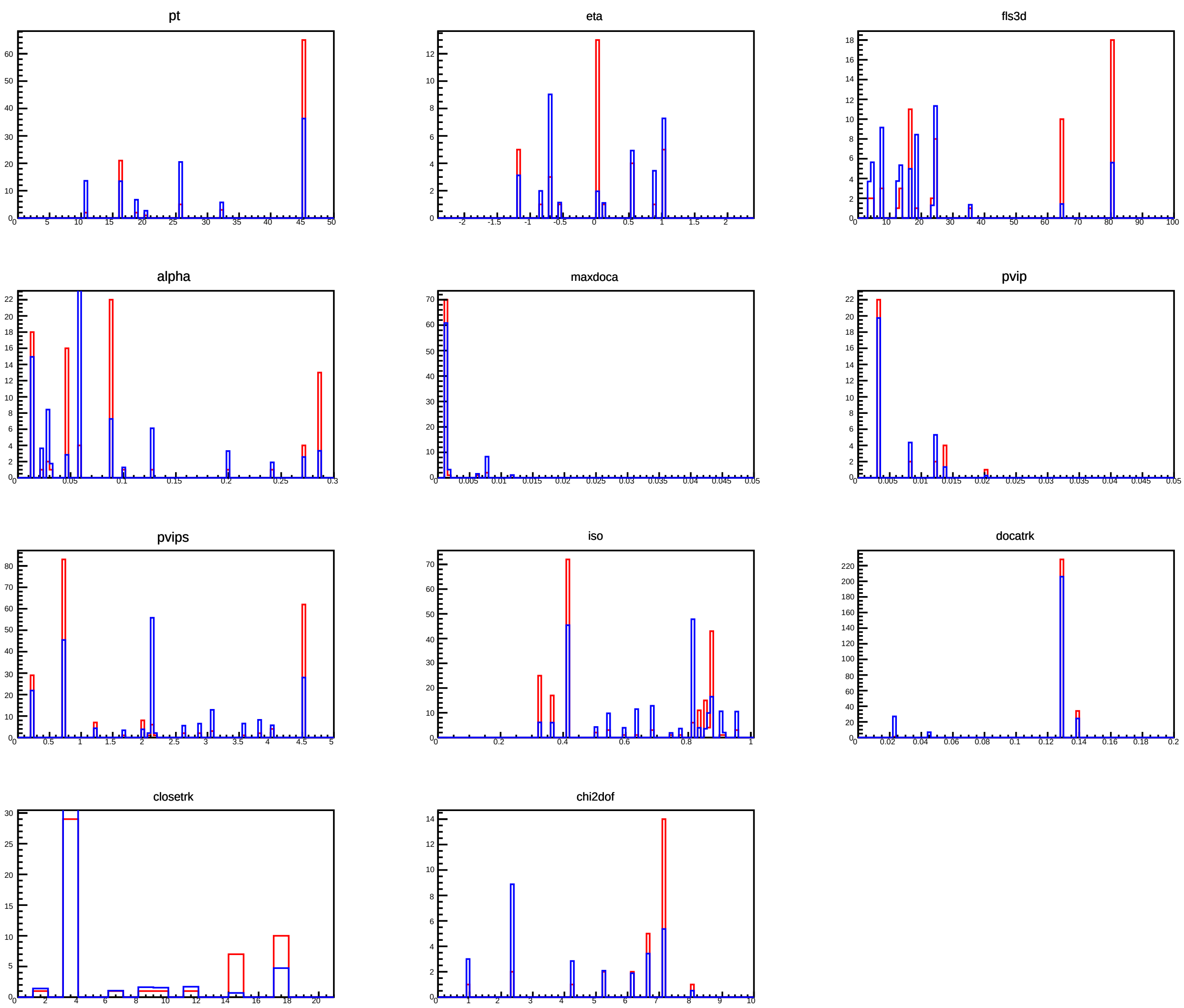

Figure 13: The cuts applied on the variables in the BDT of the barrel.

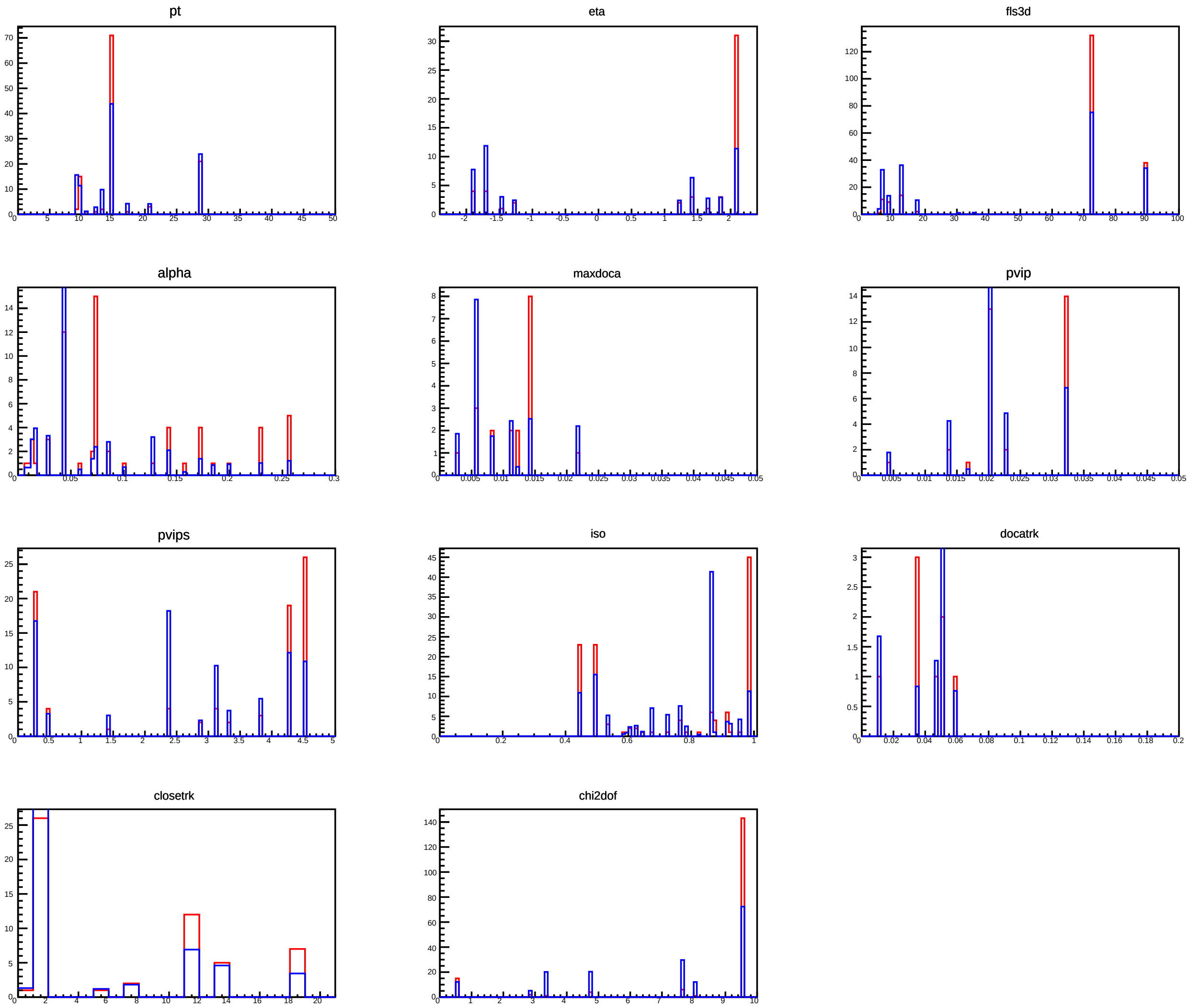

Figure 14: The cuts applied on the variables in the BDT of the endcap.

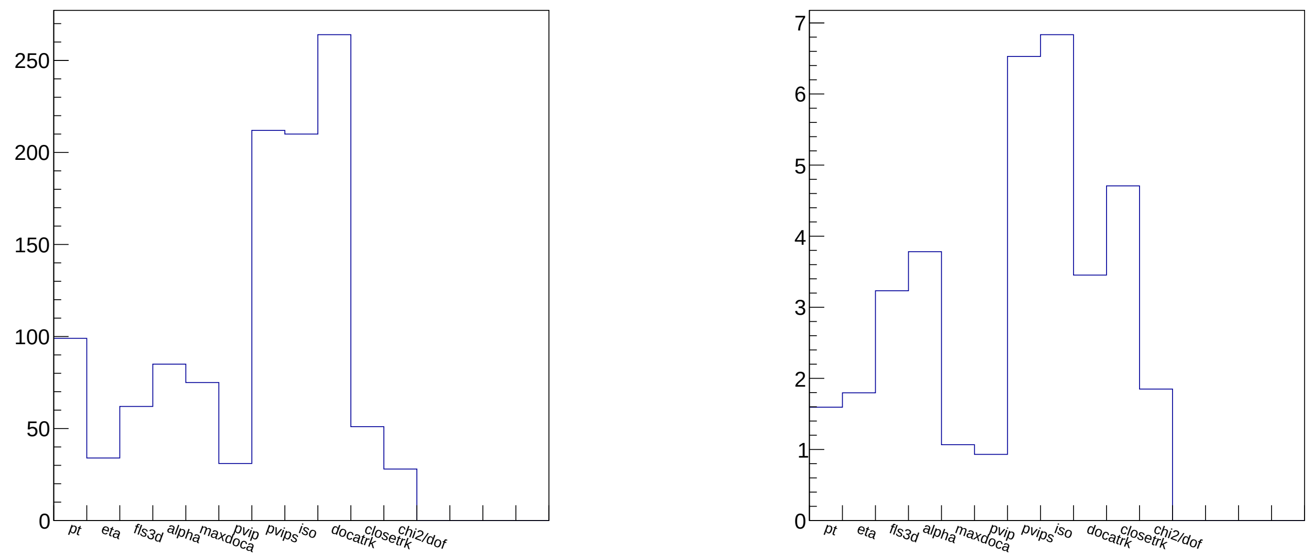

Figure 15: The cuts applied on the variables in the BDT of the barrel. The plot of the left shows the unweighted distribution, the plot on the right shows the distribution weighted with the decision tree weight.

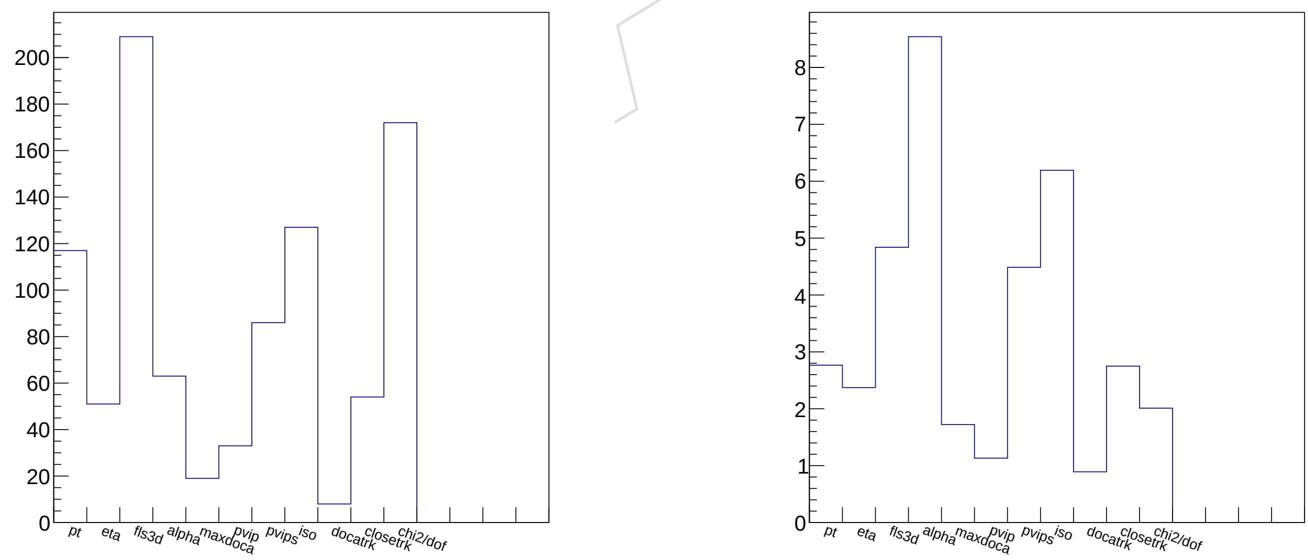

Figure 16: The cuts applied on the variables in the BDT of the endcap. The plot of the left shows the unweighted distribution, the plot on the right shows the distribution weighted with the decision tree weight.

In Fig. 15 (16) we show how often a variable is used in the barrel (endcap) BDT.

In Fig. 17 (18) the standard control plots of TMVA with a linear (logarithmic) scale are shown that check against overtraining of the BDT in the barrel and the endcap. The BDT response distributions for training and control samples are compared with the Kolmogorov-Smirnov test for consistency. The results of these tests are printed on top of the figures. All tests are passed with high probability.

In Fig. 19 the BDT response distributions, obtained on the overall event sample (all events, irrespective of their types as defined in subsection 5.1), are overlayed to compare the different BDT response. This plot is for illustration only, as in the real analysis the BDTs are applied to different event sample. The background distributions are well aligned, but in the signal some misalignment on the high side of the distribution is observed.

In Fig. 20 the receiver operating characteristic (ROC) curves are shown. This curve shows the background rejection power against the signal efficiency. The operating point of the analysis, where the expected significance is maximized, is indicated by the solid cross.

In Fig. 21 the number of background events vs the requirement on the BDT response is overlayed with the signal efficiency. This figure provides an illustration how fast the background event yield in the sidebands rises when the BDT response requirement is changed.

In Tab. 9 the number of background events is tabulated together with the signal efficiency for a few (quasi-) randomly selected operating points.

In Fig. 22 the figure of merit $S/\sqrt{S+B}$ is shown vs the BDT response requirement in the barrel and the endcap. The expected signal is based on the MC simulation luminosity (in the final analysis the expected signal is normalized to the reconstructed $B^{\pm} \to J/\psi K^{\pm}$ yield). The expected background is determined in two ways:

- The high and low sidebands are interpolated, assuming a flat behavior for the combinatorial background
$$B = \frac{\Delta m_{\text{signalbox}}}{\Delta m_{\text{sideband}}} N_{\text{sideband}}$$
The disadvantages of this method are the assumption of a flat background and the ignorance of the rare semileptonic B decays.
- The combinatorial background is extrapolated using only the high sideband. The peaking background in the signal box is assumed to be 7% of the expected signal (this is based on the expectation the full analysis, normalizing the rare backgrounds and signal expectation to the reconstructed $B^{\pm} \to J/\psi K^{\pm}$ yields).

In Fig. 23 the BDT response value b is overlayed for data and MC simulation for all samples. The MC simulation sample follow the distributions in data well. The difference is used to estimate the systematic uncertainty on the analysis efficiency, see section 8.

5.4 BDT ranking of variables

In Tab. 10 the variables used in the BDT are shown, together with their weight. The ranking and weight is extracted directly from TMVA.

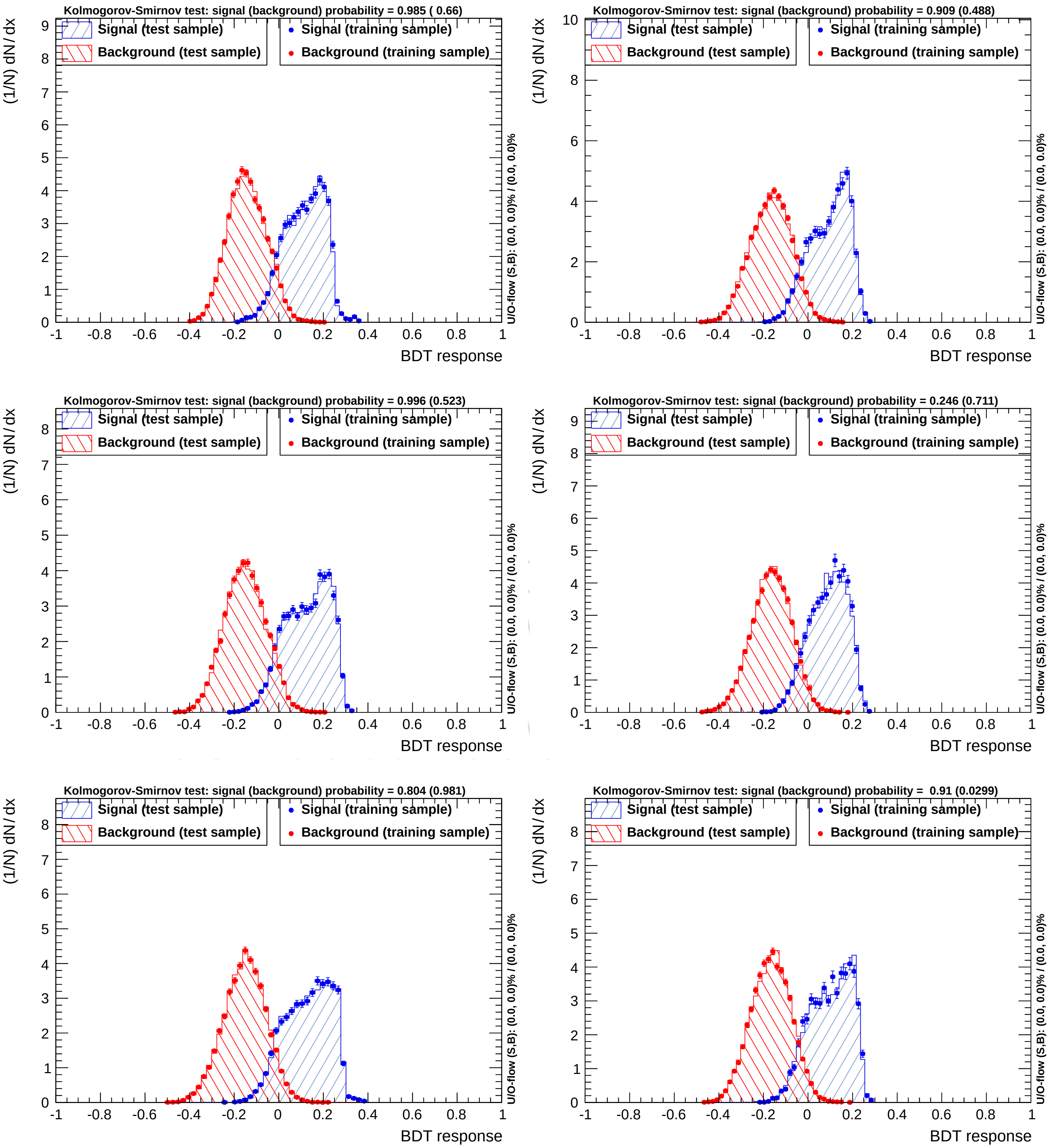

Figure 17: TMVA overtraining control plot for the barrel (left) and the endcap (right), for events of type 0 . . . 2 from the the top to the bottom.

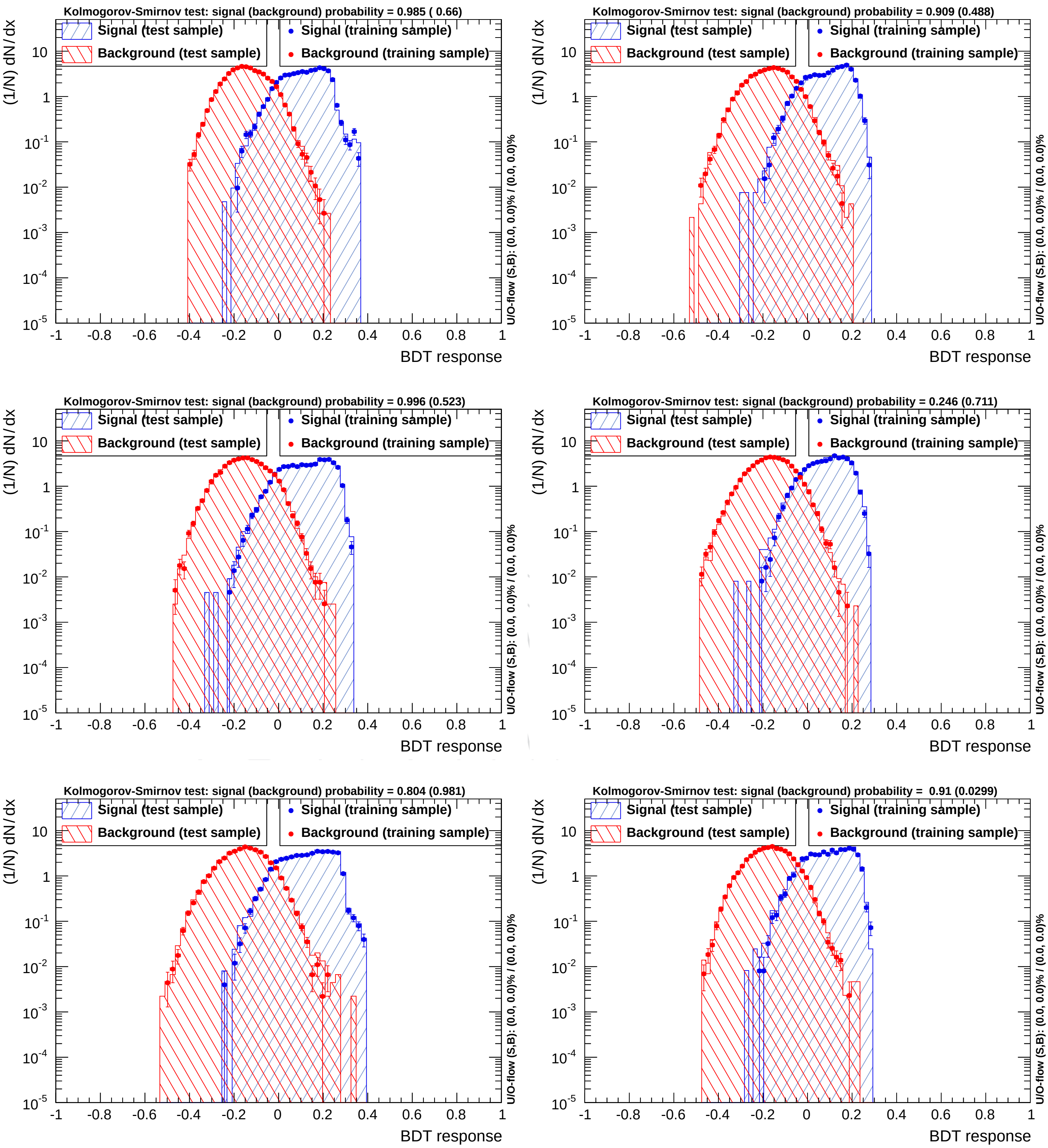

Figure 18: TMVA overtraining control plot for the barrel (left) and the endcap (right), for events of type 0 . . . 2 from the the top to the bottom. This Figure shows on a logarithmic scale the same plots as Fig. 17.

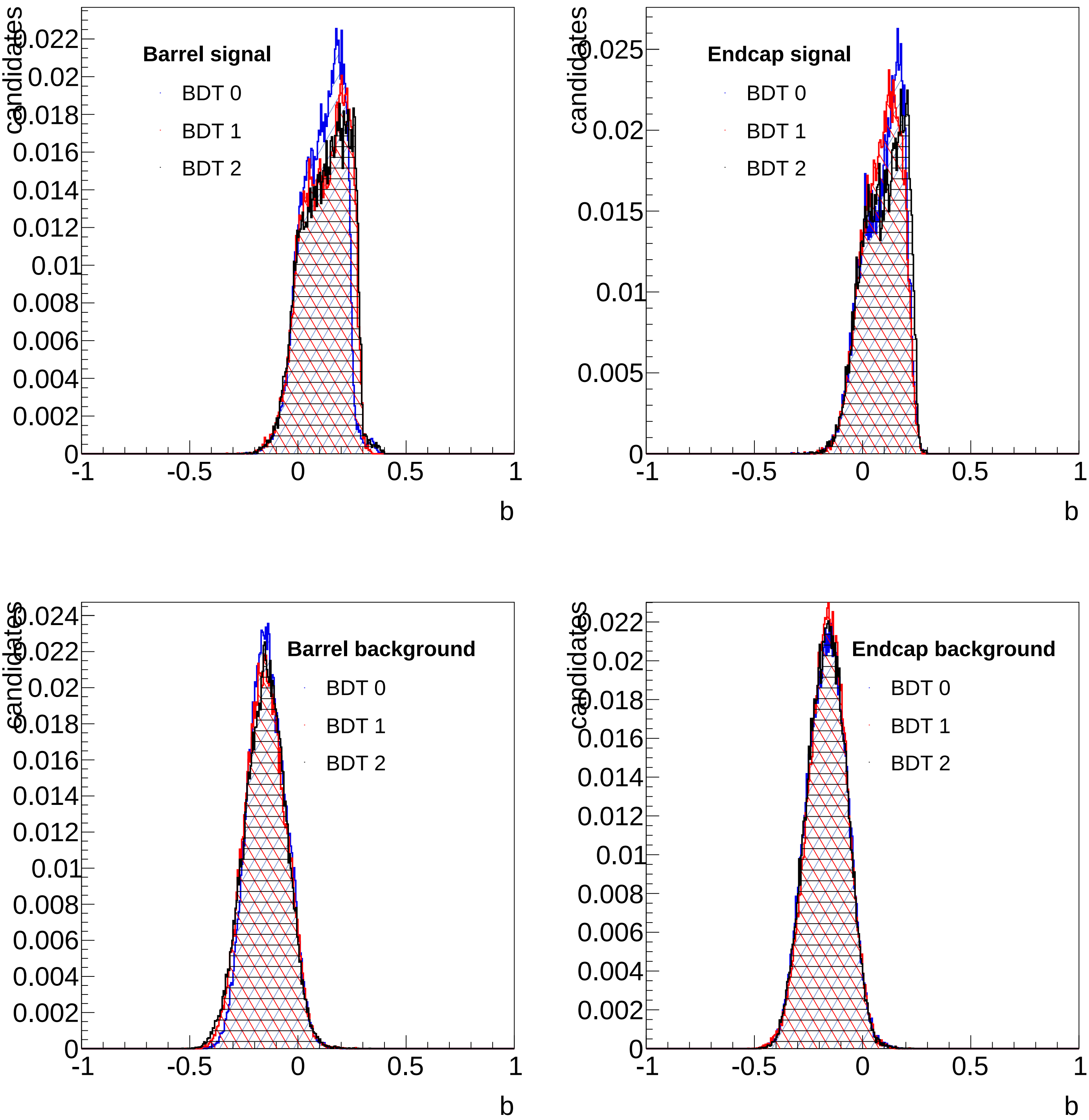

Figure 19: Overlays of the BDT response distributions on the overall event sample.

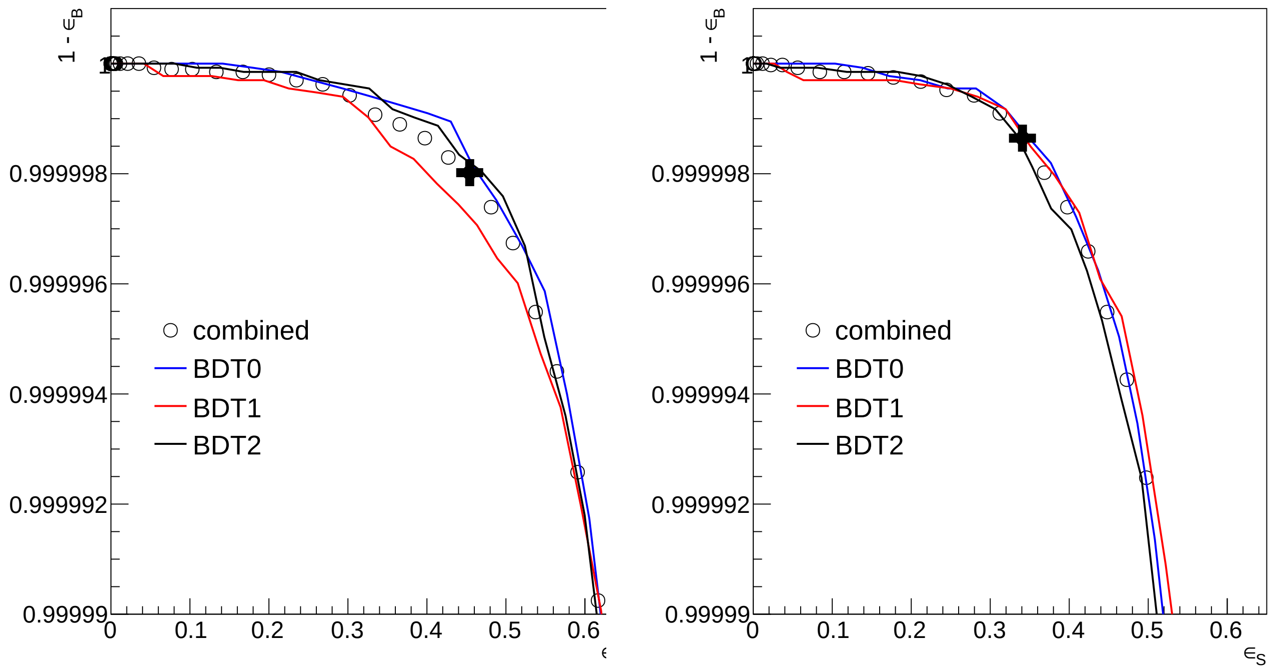

Figure 20: Receiver operating characteristic curve for the barrel (left) and endcap (right) BDTs.

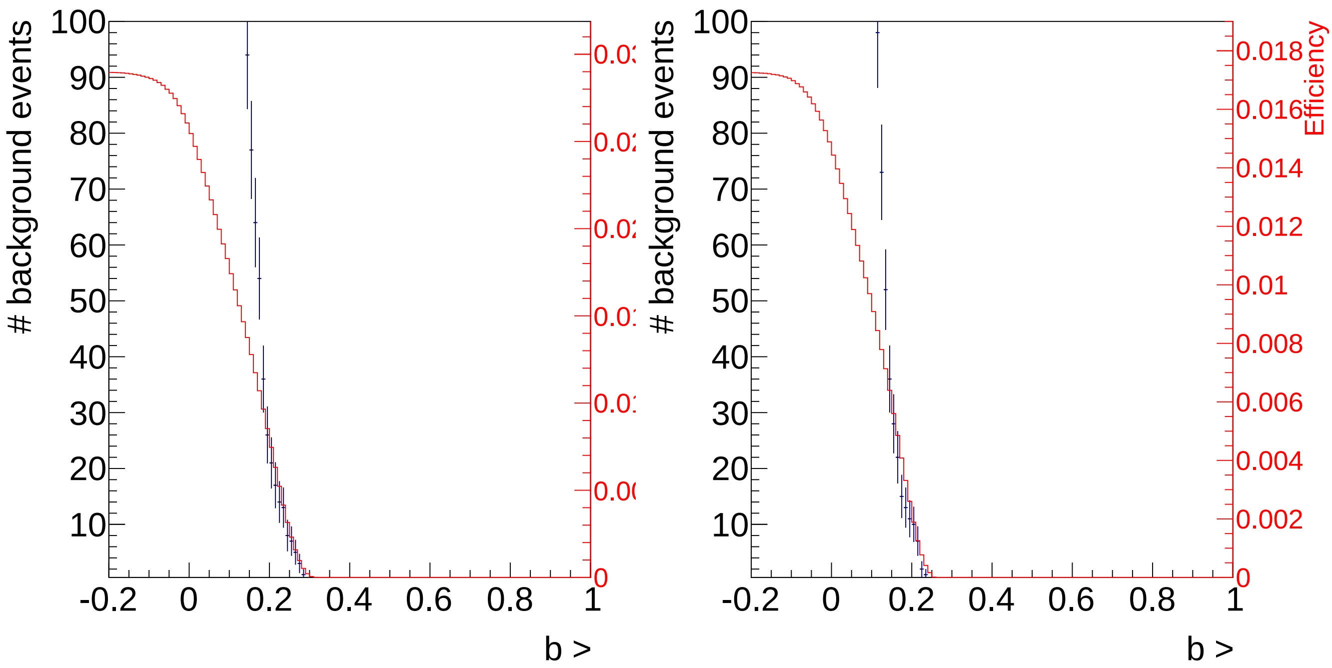

Figure 21: Number of background events vs the requirement on the BDT response (solid markers) and signal efficiency (open red histogram) in the barrel (left) and the endcap (right). The signal efficiency includes the analysis, muon identification, and HLT efficiency. Note that the background events are correlated from bin to bin; the error bars do not reflect that.

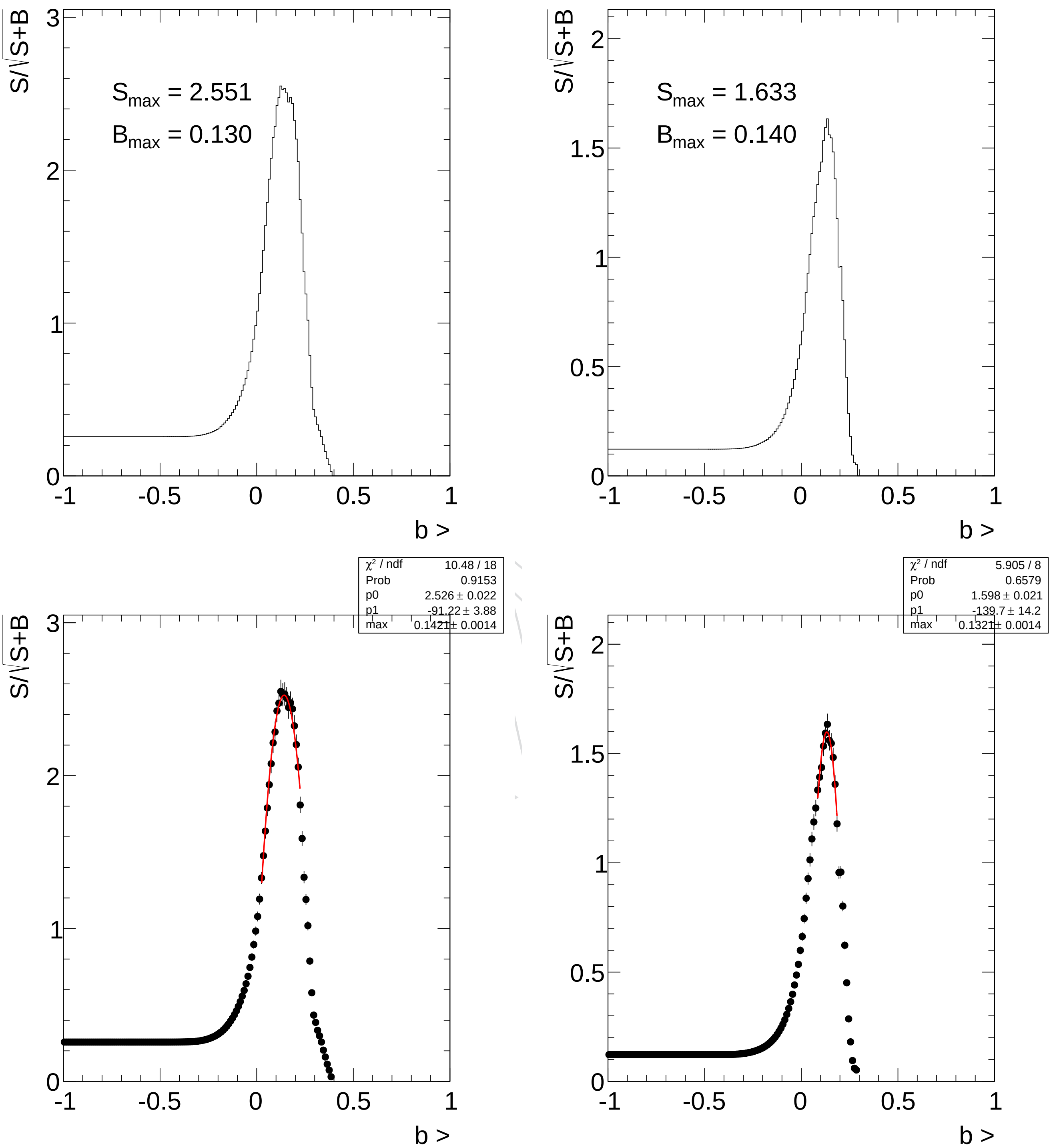

Figure 22: BDT selection: Scan of $S/\sqrt{S+B}$ vs. the minimum requirement on the BDT response for the barrel (left) and endcap (right). The plots on the bottom show the same scans, with a parabola fitted to the maximum. The fit parameter 'max' indicates the apex.

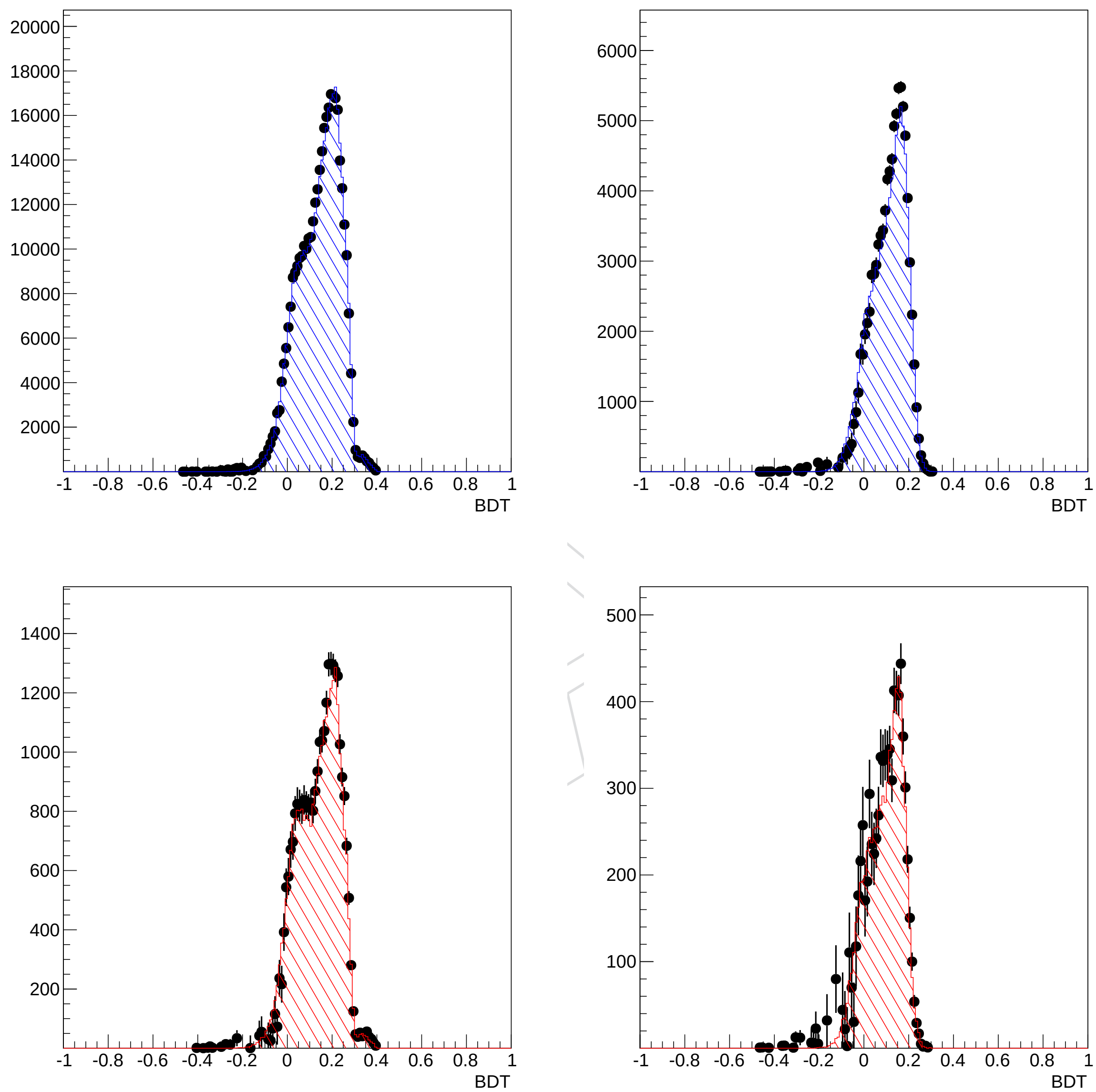

Figure 23: Overlay of b of normalization sample (top) and control sample (bottom) in the barrel (left) and endcap (right).

Table 9: Selection efficiency and number of sideband background events vs b selection requirement. The baseline (cut-n-count) selection of Ref. [31] has $\varepsilon = 0.012$ ($N_{\mathrm{bg}}^{\mathrm{obs}} = 6$) in the barrel and $\varepsilon = 0.0075$ ($N_{\mathrm{bg}}^{\mathrm{obs}} = 7$) in the endcap, where the efficiency includes the analysis selection, the muon trigger and the muon identification requirements.

$b >$	$N_{\mathrm{bg}}^{\mathrm{obs}}$	ε	$b >$	$N_{\mathrm{bg}}^{\mathrm{obs}}$	ε
−1.00	79006	0.029	−1.00	78645	0.017
−0.02	6470	0.027	−0.01	2937	0.015
0.01	3145	0.025	0.03	885	0.013
0.04	1313	0.022	0.07	267	0.011
0.06	739	0.021	0.10	122	0.009
0.09	308	0.018	0.13	52	0.007
0.11	184	0.017	0.16	22	0.005
0.13	113	0.015	0.19	11	0.003
0.15	77	0.013	0.22	2	0.001
0.17	54	0.011			
0.19	26	0.009			
0.21	17	0.006			
0.23	13	0.004			
0.25	7	0.002			
0.28	1	0.001			

Table 10: BDT Variable ranking

Barrel						Endcap					
0		1		2		0		1		2	
Variable	Weight	Variable	Weight	Variable	Weight	Variable	Weight	Variable	Weight	Variable	Weight
I	0.1501	I	0.1569	I	0.1508	I	0.1460	I	0.1279	χ^2/dof	0.1983
$\delta_{3D}/\sigma(\delta_{3D})$	0.1152	$p_{\perp B}$	0.1239	$\delta_{3D}/\sigma(\delta_{3D})$	0.1255	$N_{\text{trk}}^{\text{close}}$	0.1382	$\delta_{3D}/\sigma(\delta_{3D})$	0.1202	$\ell_{3D}/\sigma(\ell_{3D})$	0.1339
$\ell_{3D}/\sigma(\ell_{3D})$	0.1036	$\ell_{3D}/\sigma(\ell_{3D})$	0.1223	d_{ca}^0	0.1238	χ^2/dof	0.1138	$N_{\text{trk}}^{\text{close}}$	0.1057	I	0.1161
η_B	0.1006	$\delta_{3D}/\sigma(\delta_{3D})$	0.1046	α	0.1108	$\delta_{3D}/\sigma(\delta_{3D})$	0.1051	η_B	0.1012	$p_{\perp B}$	0.0981
d_{ca}^0	0.1003	$N_{\text{trk}}^{\text{close}}$	0.1000	$p_{\perp B}$	0.1065	α	0.0948	δ_{3D}	0.0975	α	0.0981
α	0.0976	α	0.0895	d_{ca}	0.0782	$\ell_{3D}/\sigma(\ell_{3D})$	0.0820	d_{ca}	0.0974	$\delta_{3D}/\sigma(\delta_{3D})$	0.0847
$p_{\perp B}$	0.0936	η_B	0.0662	$\ell_{3D}/\sigma(\ell_{3D})$	0.0748	$p_{\perp B}$	0.0709	$\ell_{3D}/\sigma(\ell_{3D})$	0.0866	$N_{\text{trk}}^{\text{close}}$	0.0777
d_{ca}	0.0695	d_{ca}^0	0.0661	$N_{\text{trk}}^{\text{close}}$	0.0662	d_{ca}	0.0694	α	0.0792	η_B	0.0744
δ_{3D}	0.0609	d_{ca}	0.0628	δ_{3D}	0.0611	δ_{3D}	0.0689	$p_{\perp B}$	0.0759	δ_{3D}	0.0649
$N_{\text{trk}}^{\text{close}}$	0.0572	δ_{3D}	0.0550	η_B	0.0549	η_B	0.0672	d_{ca}^0	0.0549	d_{ca}	0.0312
χ^2/dof	0.0515	χ^2/dof	0.0529	χ^2/dof	0.0474	d_{ca}^0	0.0437	χ^2/dof	0.0535	d_{ca}^0	0.0226

5.5 BDT pileup studies

In Fig. 24 the mean BDT response value vs the number of primary vertices is shown without any preselection. This distribution is dominated by background which does not enter the final selection.

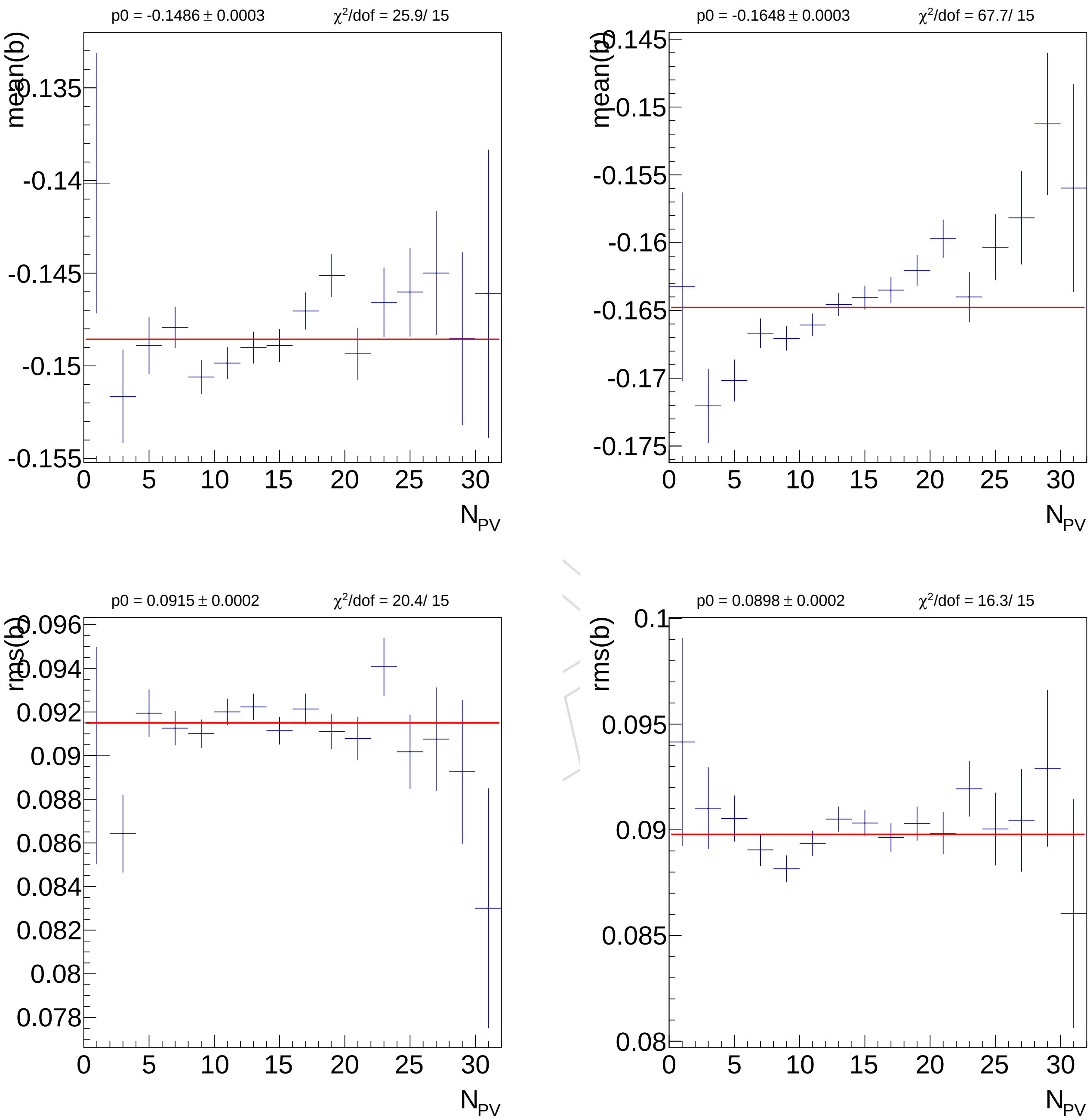

Figure 24: Profile histogram showing the mean (top row) and RMS (bottom row) of b vs. the number of primary vertices in the data dimuon sidebands for the barrel (left) and endcap (right). The line indicates a fit with a constant. No preselection is applied.

In Fig. 26 the mean BDT response value vs the number of primary vertices is shown for a preselection of $b > 0$. This plot concentrates on the more interesting region for the analysis and demonstrates a fair pileup independence. A constant is fitted to the data. In Fig. 27 a linear

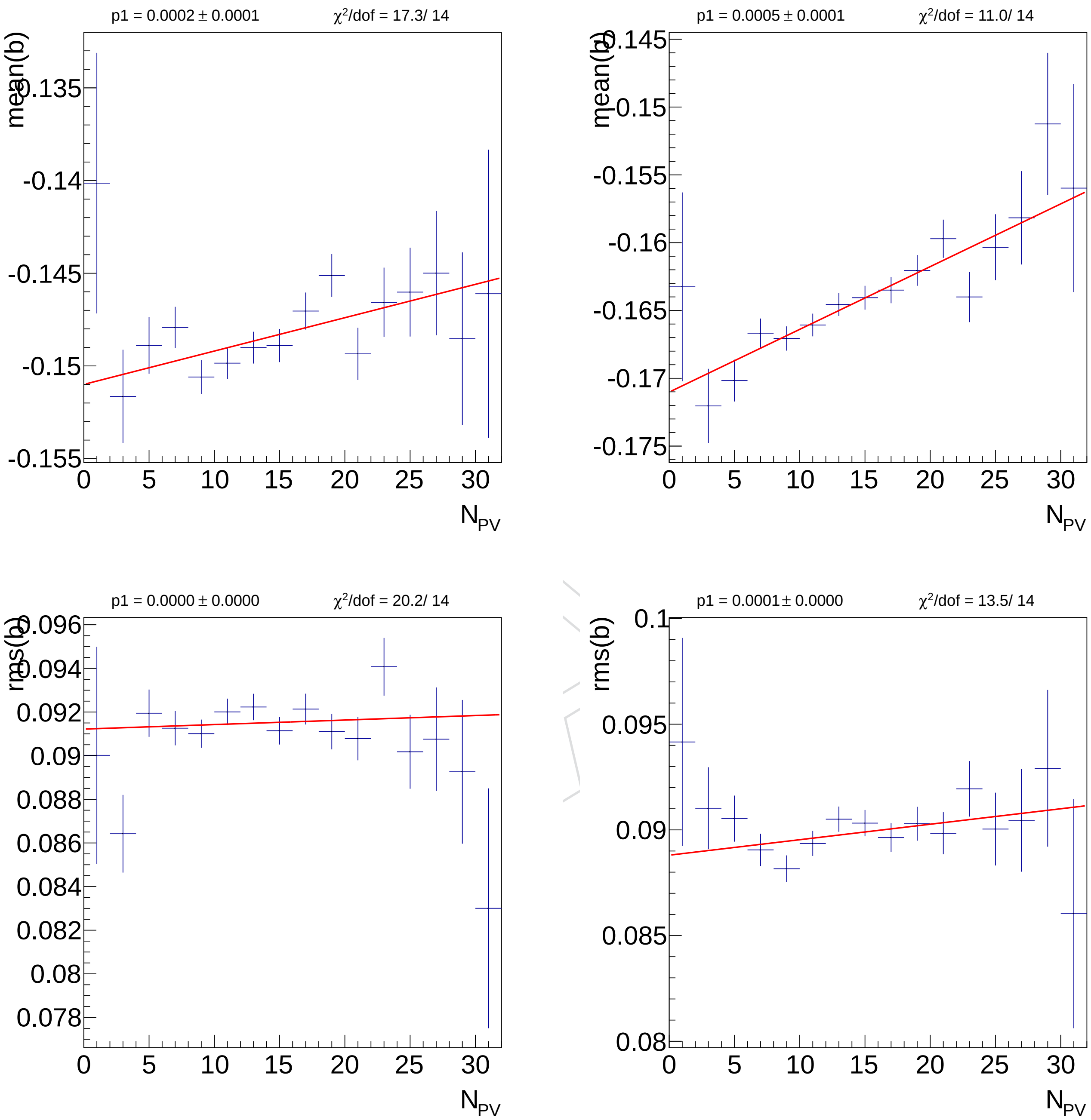

Figure 25: Profile histogram showing the mean (top row) and RMS (bottom row) of b vs. the number of primary vertices in the data dimuon sidebands for the barrel (left) and endcap (right). The line indicates a fit with a linear function. No preselection is applied.

504 function is fitted to the data.

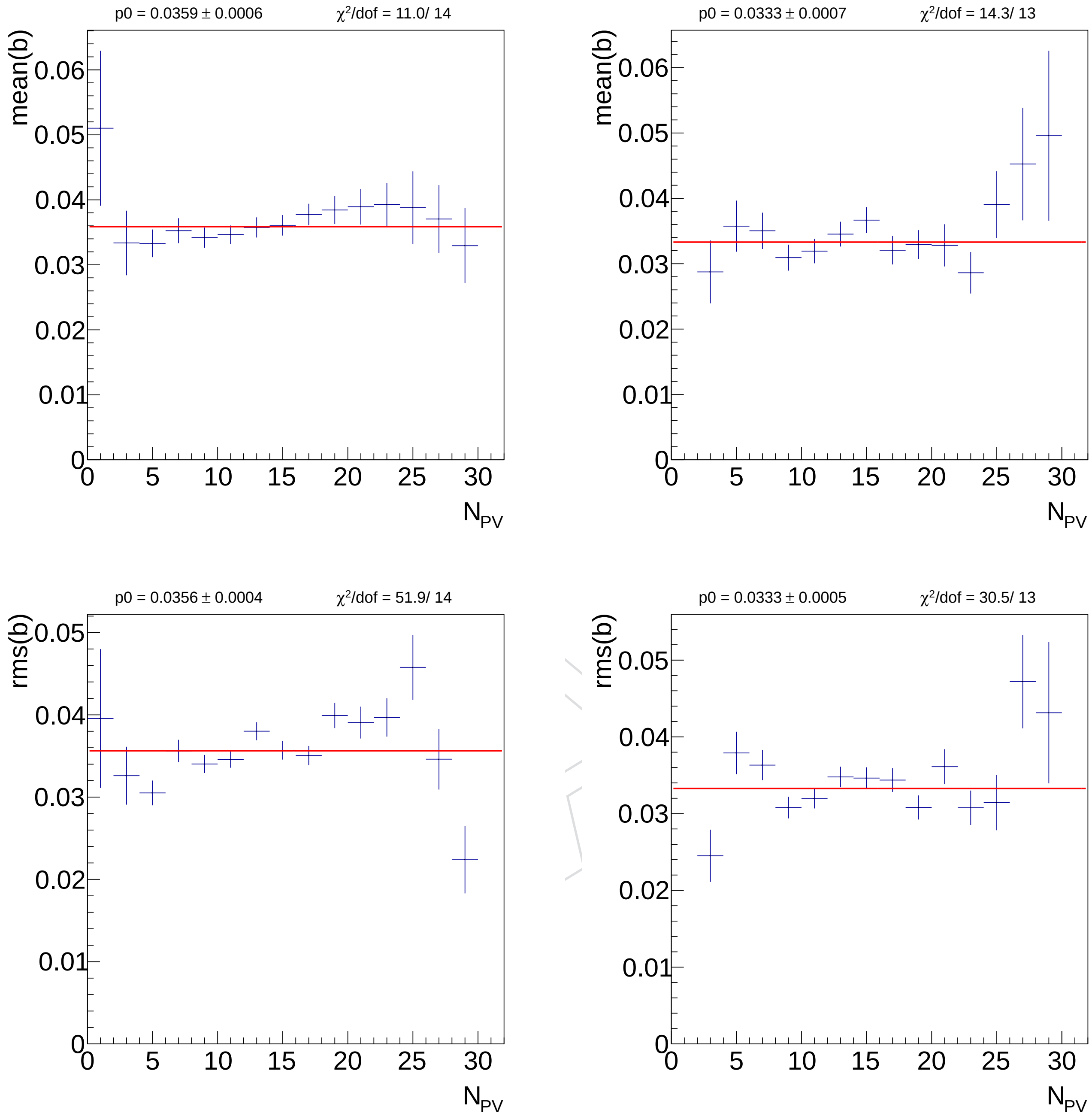

Figure 26: Profile histogram showing the mean (top row) and RMS (bottom row) of b vs. the number of primary vertices in the data dimuon sidebands for the barrel (left) and endcap (right). The line indicates a fit with a constant. In contrast to Fig. 24 a preselection of $b > 0$ has been applied here.

505 ## 5.6 BDT mass dependence studies

506 In Fig. 30 (barrel) and 31 (endcap) we show the effects of $b > b_{\text{min}}$ requirements separately
507 for the low and the high dimuon mass sidebands. The 'selection efficiency' (or background
508 rejection) is seen to be very similar in the low and high sidebands.

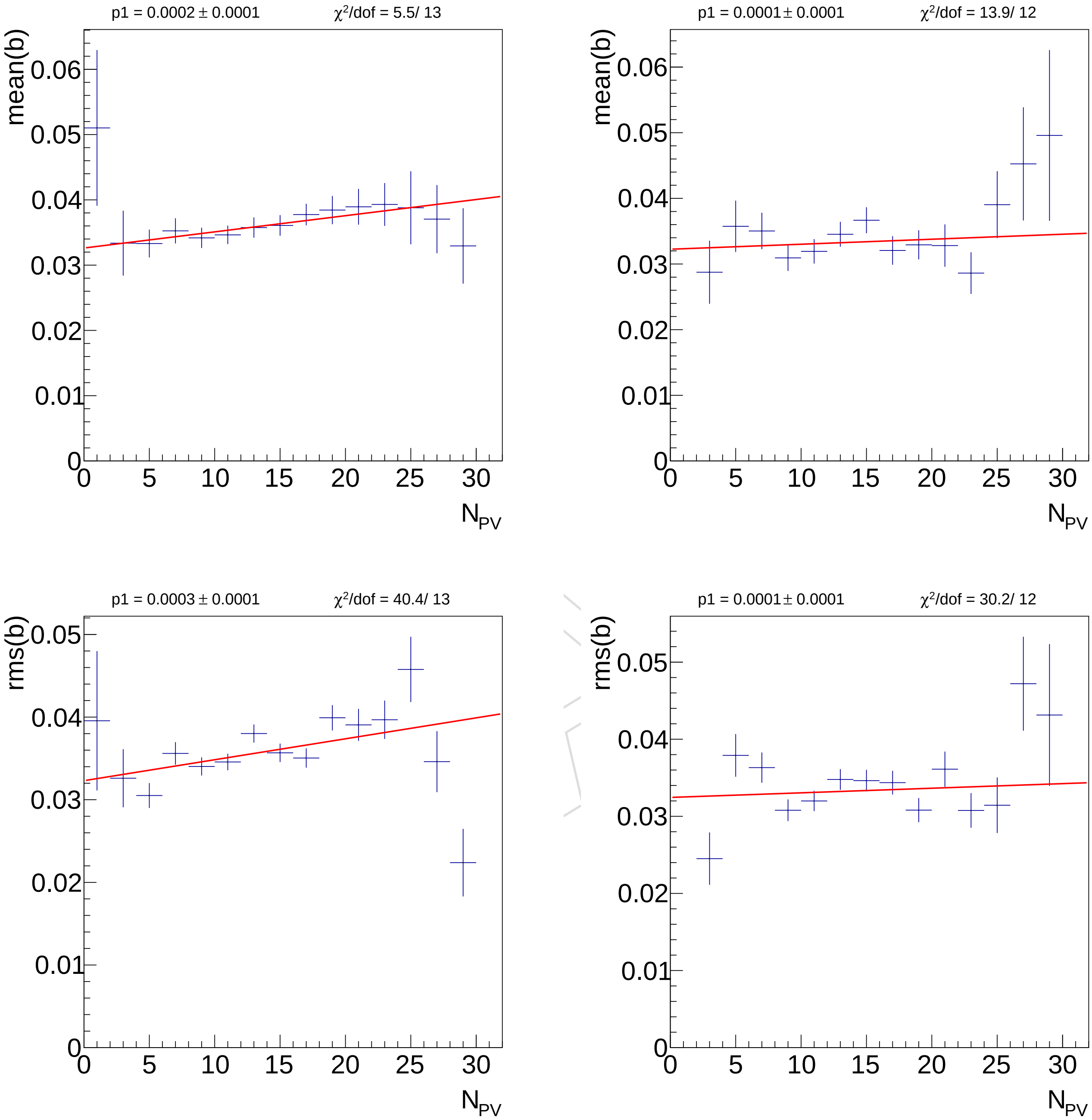

Figure 27: Profile histogram showing the mean (top row) and RMS (bottom row) of b vs. the number of primary vertices in the data dimuon sidebands for the barrel (left) and endcap (right). The line indicates a fit with a linear function. In contrast to Fig. 25 a preselection of $b > 0$ has been applied here.

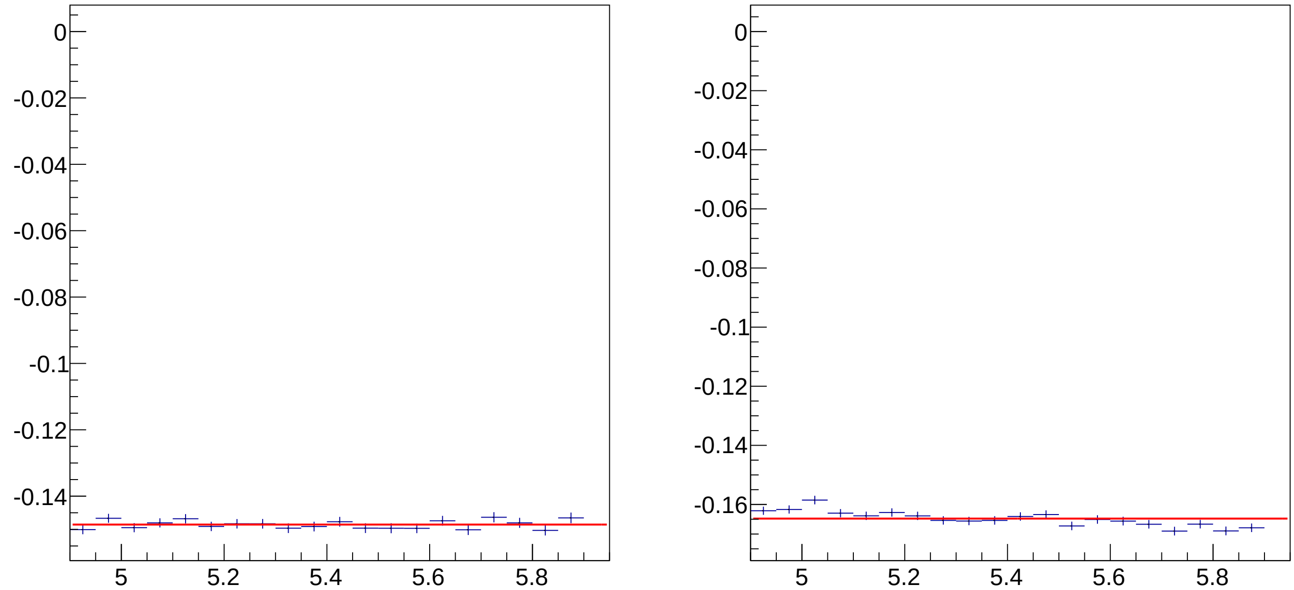

Figure 28: Profile histogram showing the mean of b vs. the dimuon invariant mass in the data dimuon sidebands for the barrel (left) and endcap (right). No preselection ius applied.

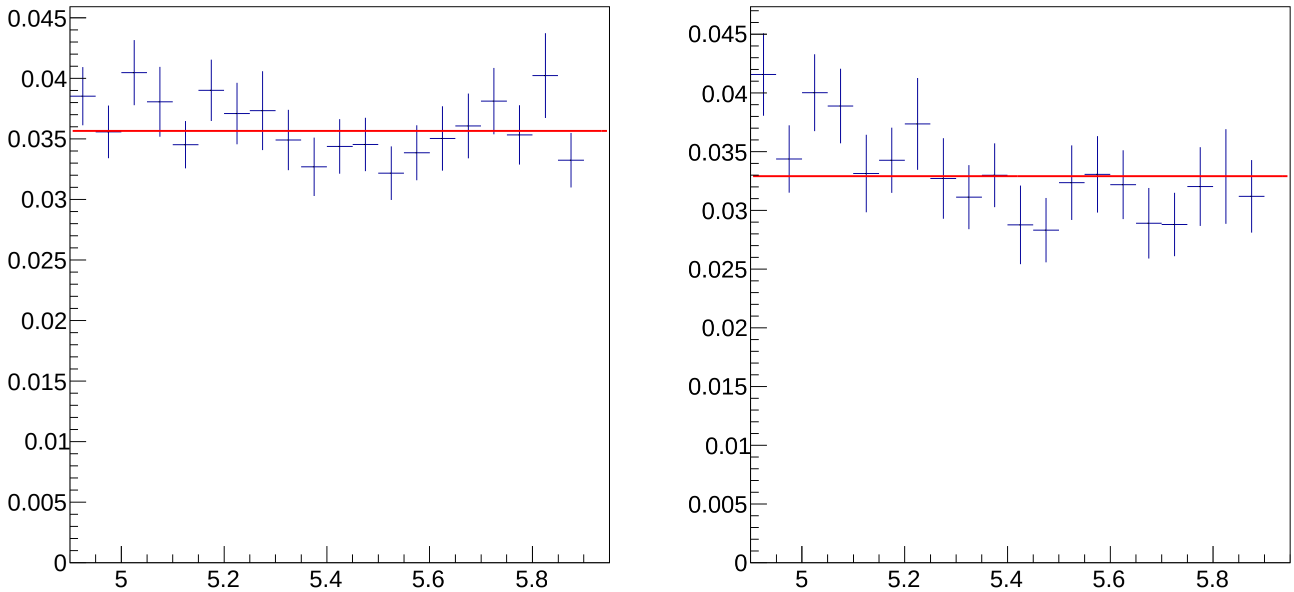

Figure 29: Profile histogram showing the mean of b vs. the dimuon invariant mass in the data dimuon sidebands for the barrel (left) and endcap (right). In contrast to Fig. 28 a preselection of $b > 0$ has been applied here.

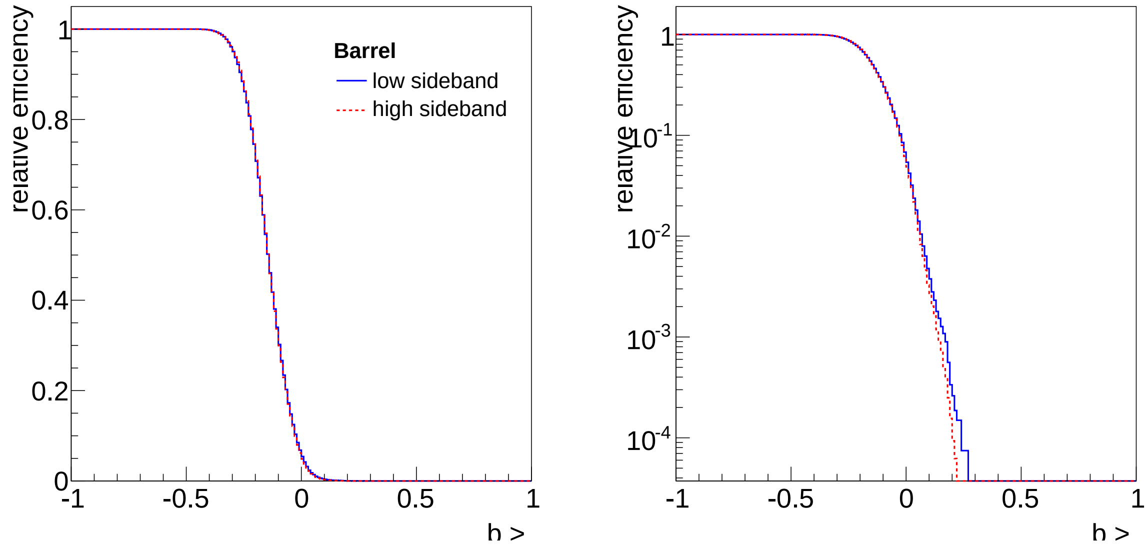

Figure 30: ‘Selection efficiency’ in data sidebands vs. $b > b_{\rm min}$ selection requirement in the barrel for the low and high dimuon mass sideband on a linear (left) and logarithmic (right) scale.

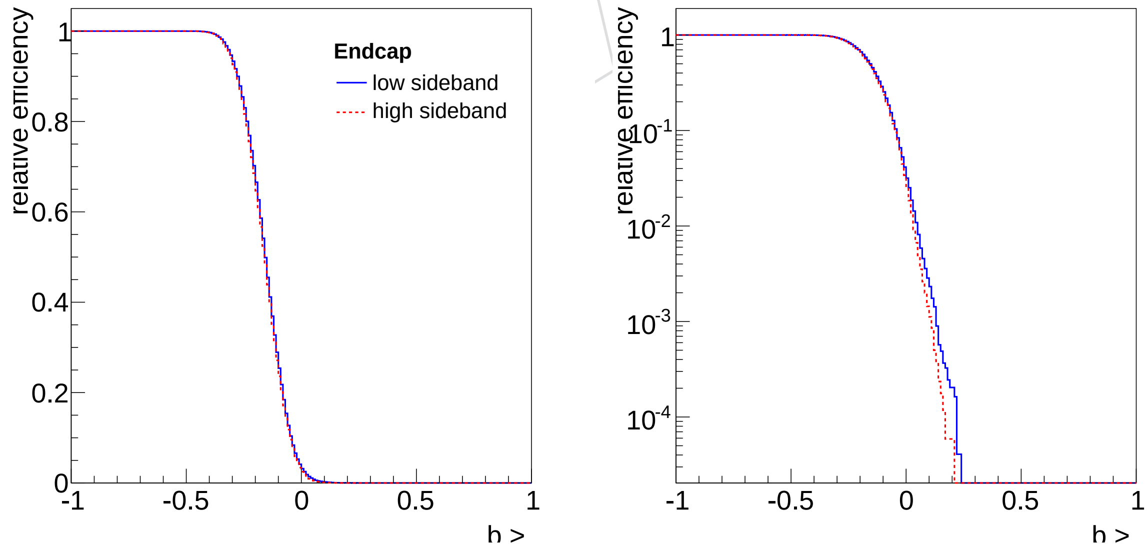

Figure 31: ‘Selection efficiency’ in data sidebands vs. $b > b_{\rm min}$ selection requirement in the endcap for the low and high dimuon mass sideband on a linear (left) and logarithmic (right) scale.

In Fig. 32 (33) the BDT response for three different (hypothetical) B_s^0 masses are shown in the barrel (endcap).

In Fig. 34 (35) the mass dependence of specific BDT response selections are shown in the barrel (endcap).

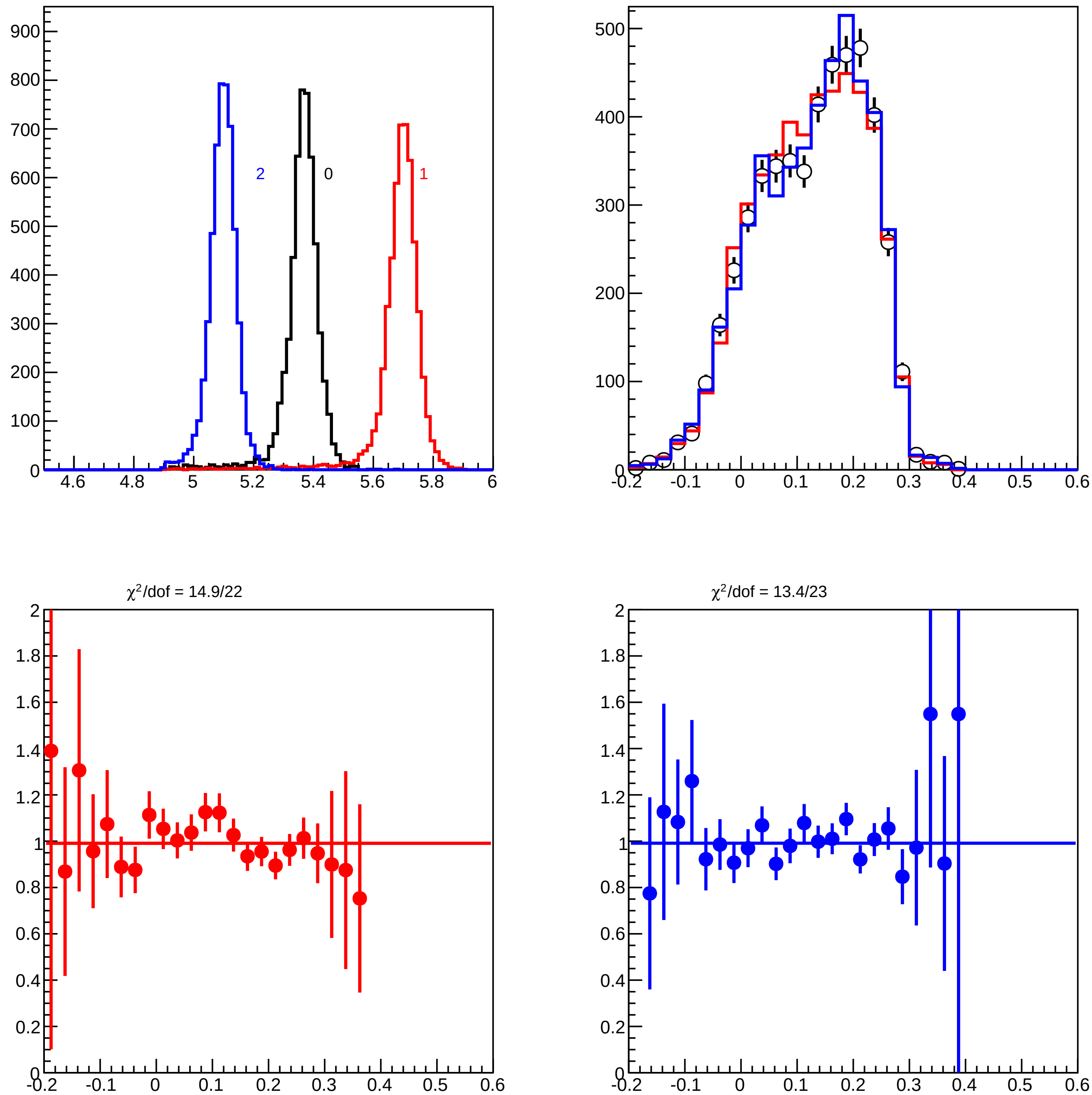

Figure 32: Study in the barrel for three different (hypothetical) B_s^0 masses (top left) of the BDT response (top right). The open markers show the BDT response for the normal B_s^0 mass, the colored histograms correspond to the respective shifted masses. In the middle row, the ratio of hypothetical B_s^0 (1) to the known B_s^0 (left) and ratio of the hypothetical B_s^0 (2) to the known B_s^0 (right). The line illustrates a constant fit to the points. In the bottom row, the line illustrates a linear fit to the points.

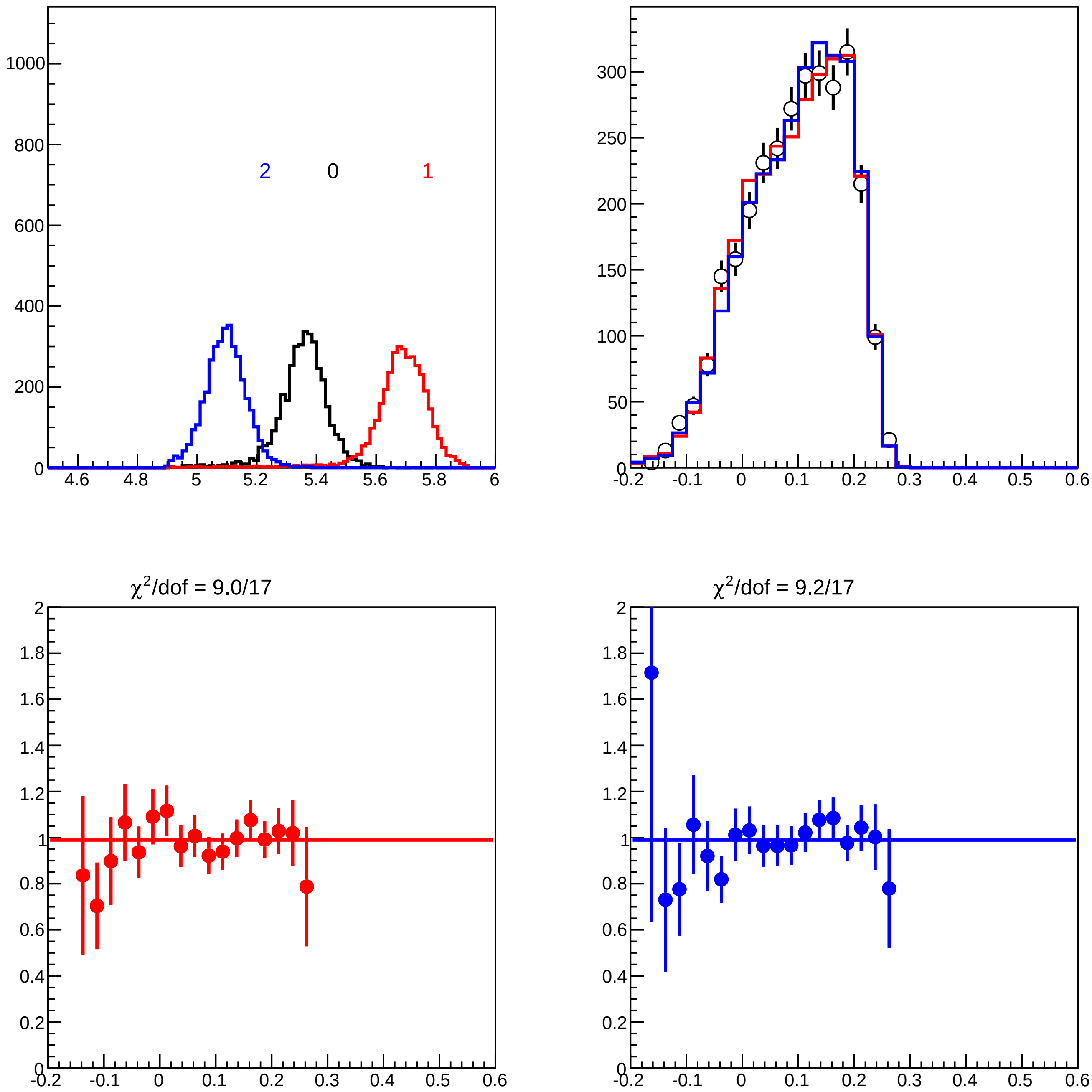

Figure 33: Study in the endcap for three different (hypothetical) B_s^0 masses (top left) of the BDT response (top right). The open markers show the BDT response for the normal B_s^0 mass, the colored histograms correspond to the respective shifted masses. In the middle row, the ratio of hypothetical B_s^0 (1) to the known B_s^0 (left) and ratio of the hypothetical B_s^0 (2) to the known B_s^0 (right). The line illustrates a constant fit to the points. In the bottom row, the line illustrates a linear fit to the points.

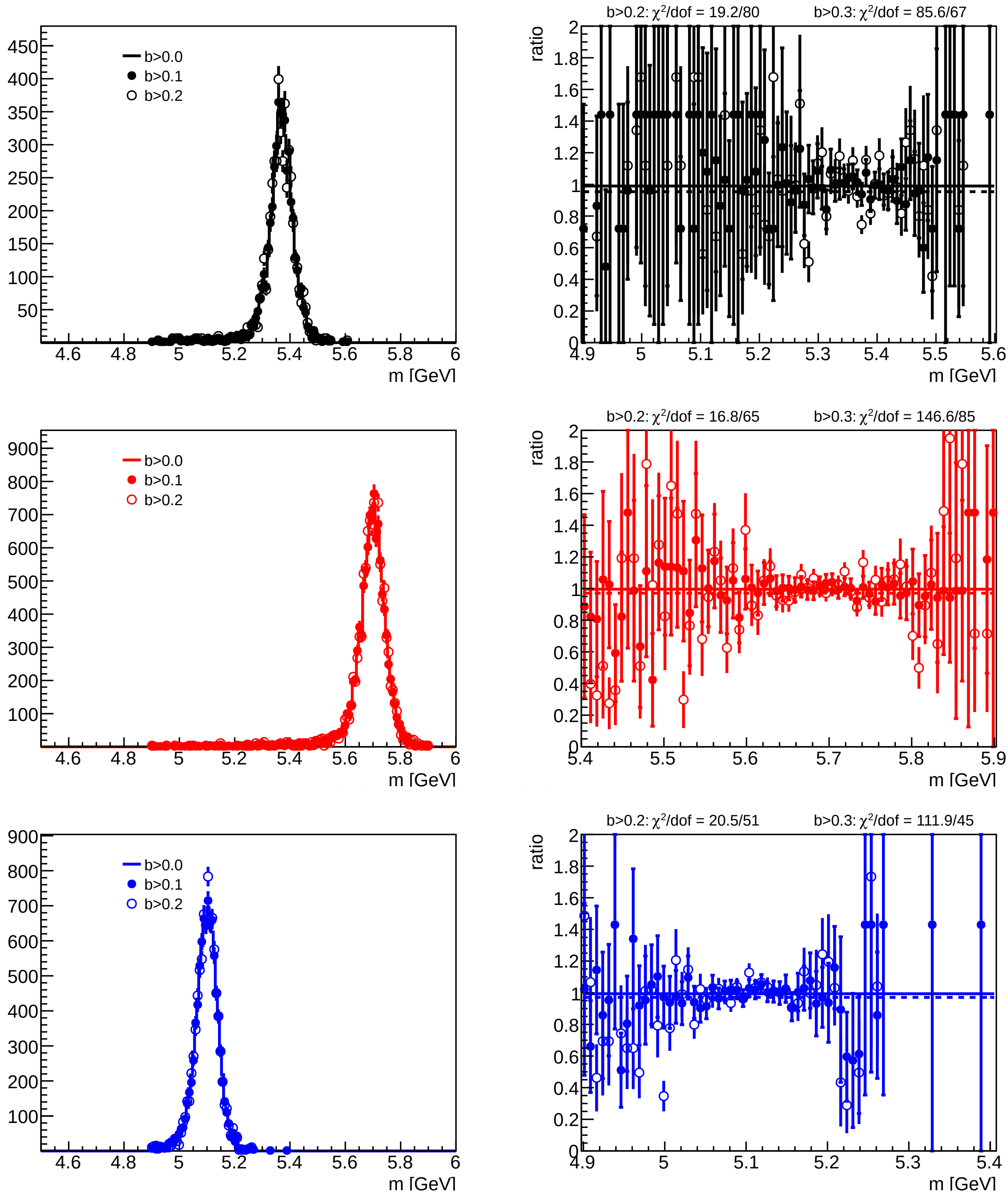

Figure 34: Study in the barrel for three different (hypothetical) B_s^0 masses (top left) of the mass distribution for different BDT response selections ($b > 0.0$, $b > 0.2$, and $b > 0.3$). The left column shows the overlayed mass distributions for the three different B_s^0 masses. The right column shows the ratio of the two tightened selections to the base selection, respectively. The numbers on top of the figures indicate the χ^2/dof when fitting a constant to the ratios.

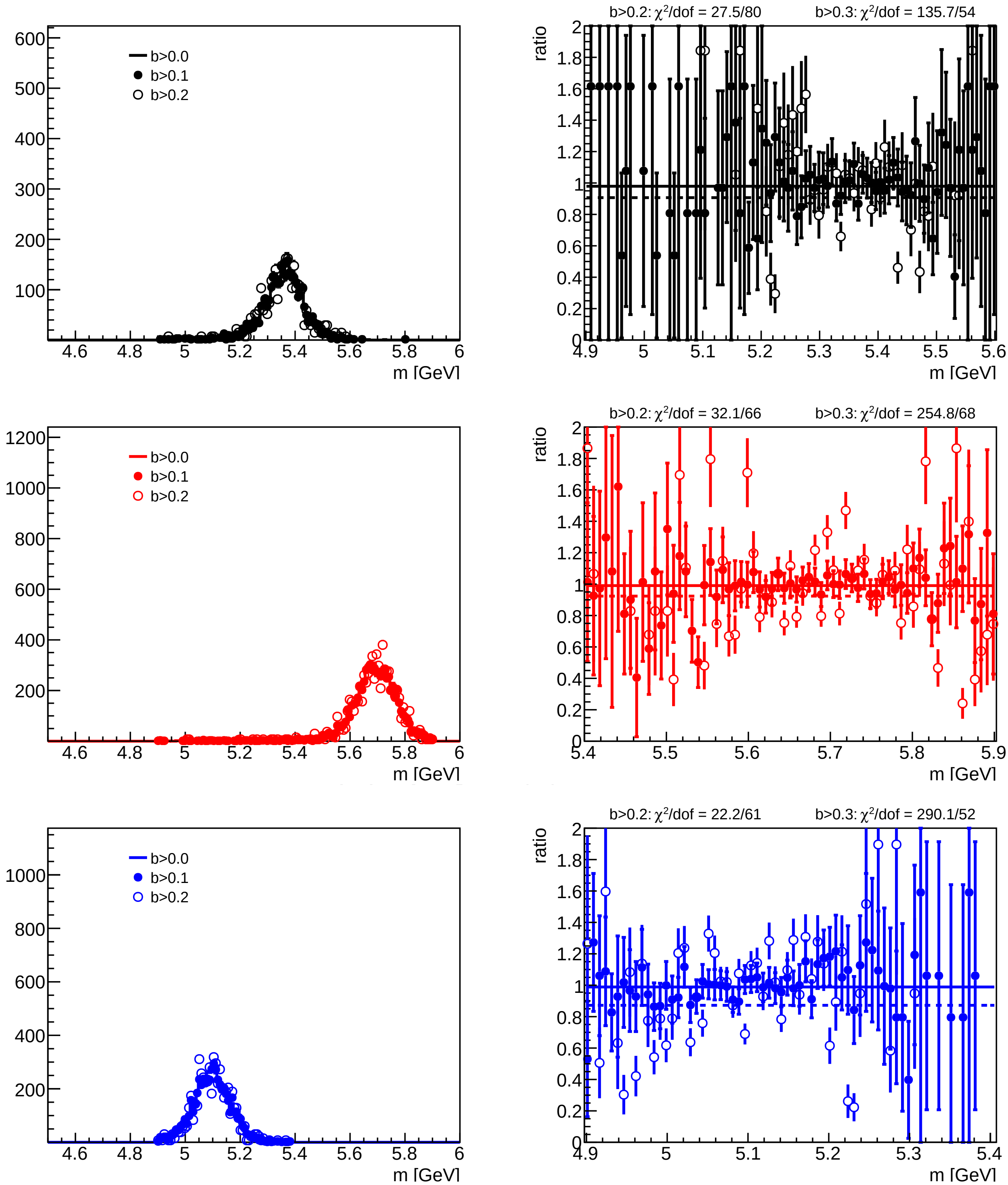

Figure 35: Study in the endcap for three different (hypothetical) B_s^0 masses (top left) of the mass distribution for different BDT response selections ($b > 0.0$, $b > 0.2$, and $b > 0.3$). The left column shows the overlayed mass distributions for the three different B_s^0 masses. The right column shows the ratio of the two tightened selections to the base selection, respectively. The numbers on top of the figures indicate the χ^2/dof when fitting a constant to the ratios.

6 Comparison of Data and MC simulation

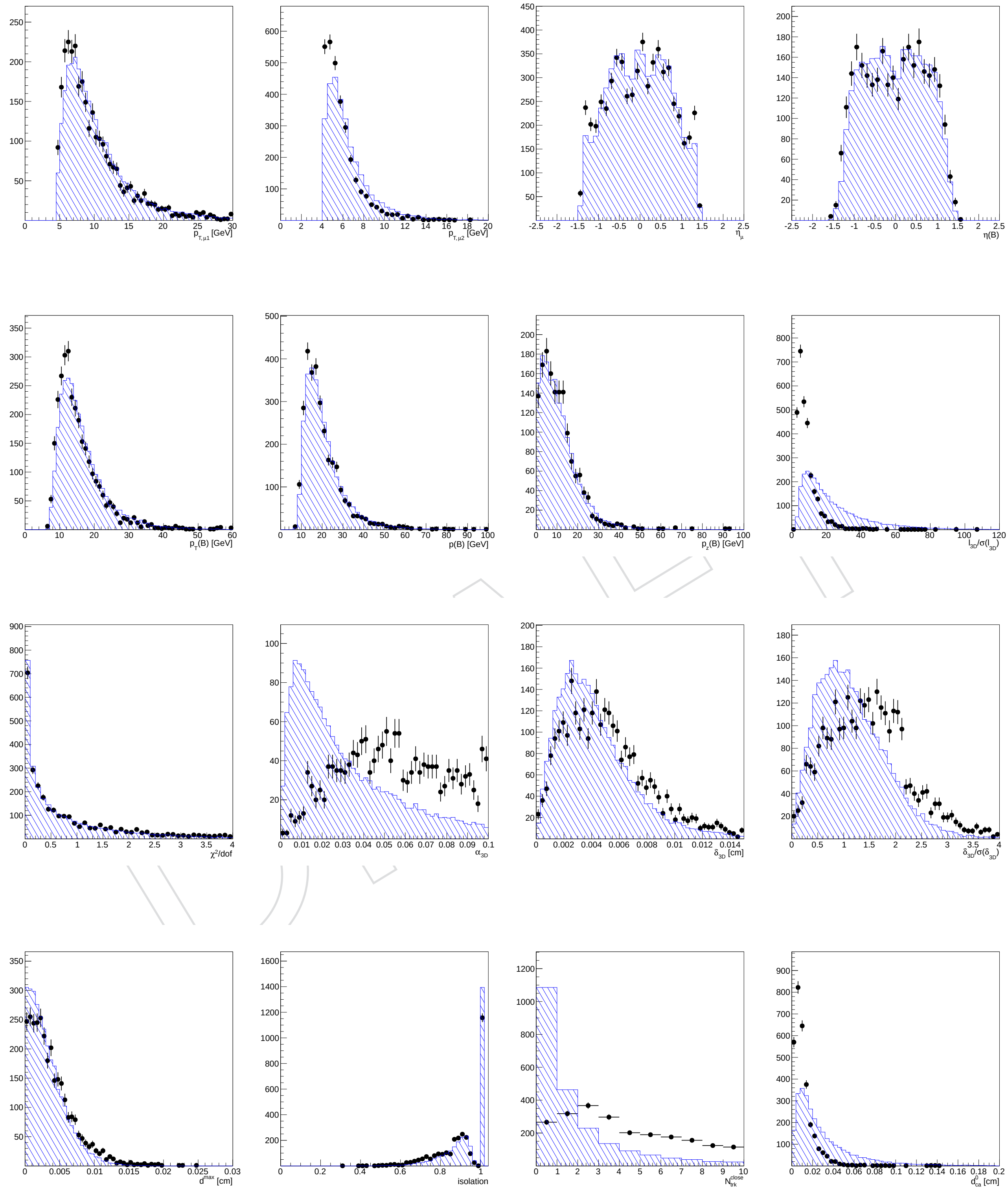

Figure 36: Comparison of barrel dimuon candidate distributions in sideband background data (solid markers) and signal MC simulation (hatched histogram). For each figure, the selection criteria $b > 0$ is applied. The MC simulation histogram is normalized to the same number of events like the data histogram.

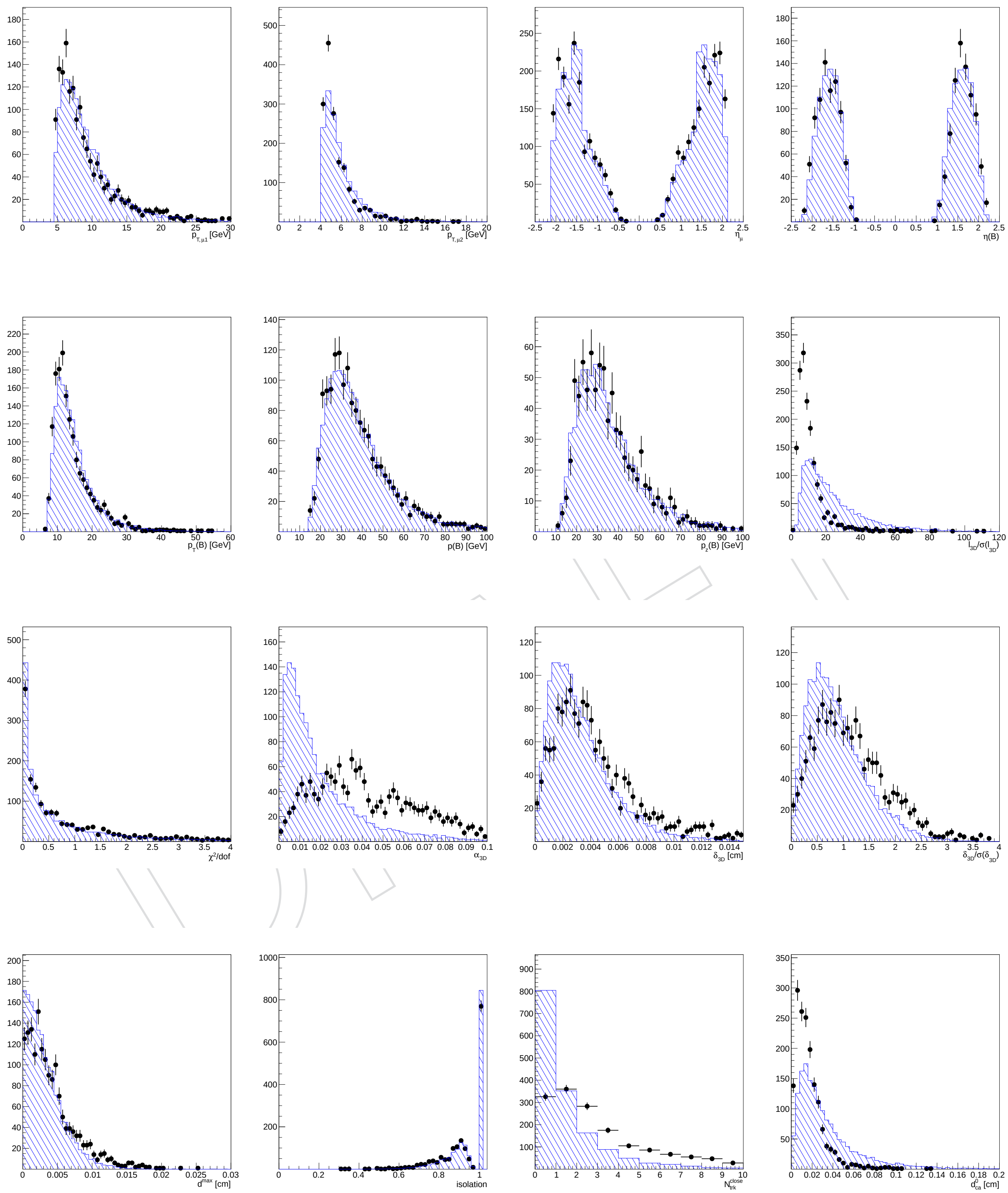

Figure 37: Comparison of endcap dimuon candidate distributions in sideband background data (solid markers) and signal MC simulation (hatched histogram). For each figure, the selection criteria $b > 0$ is applied. The MC simulation histogram is normalized to the same number of events like the data histogram.

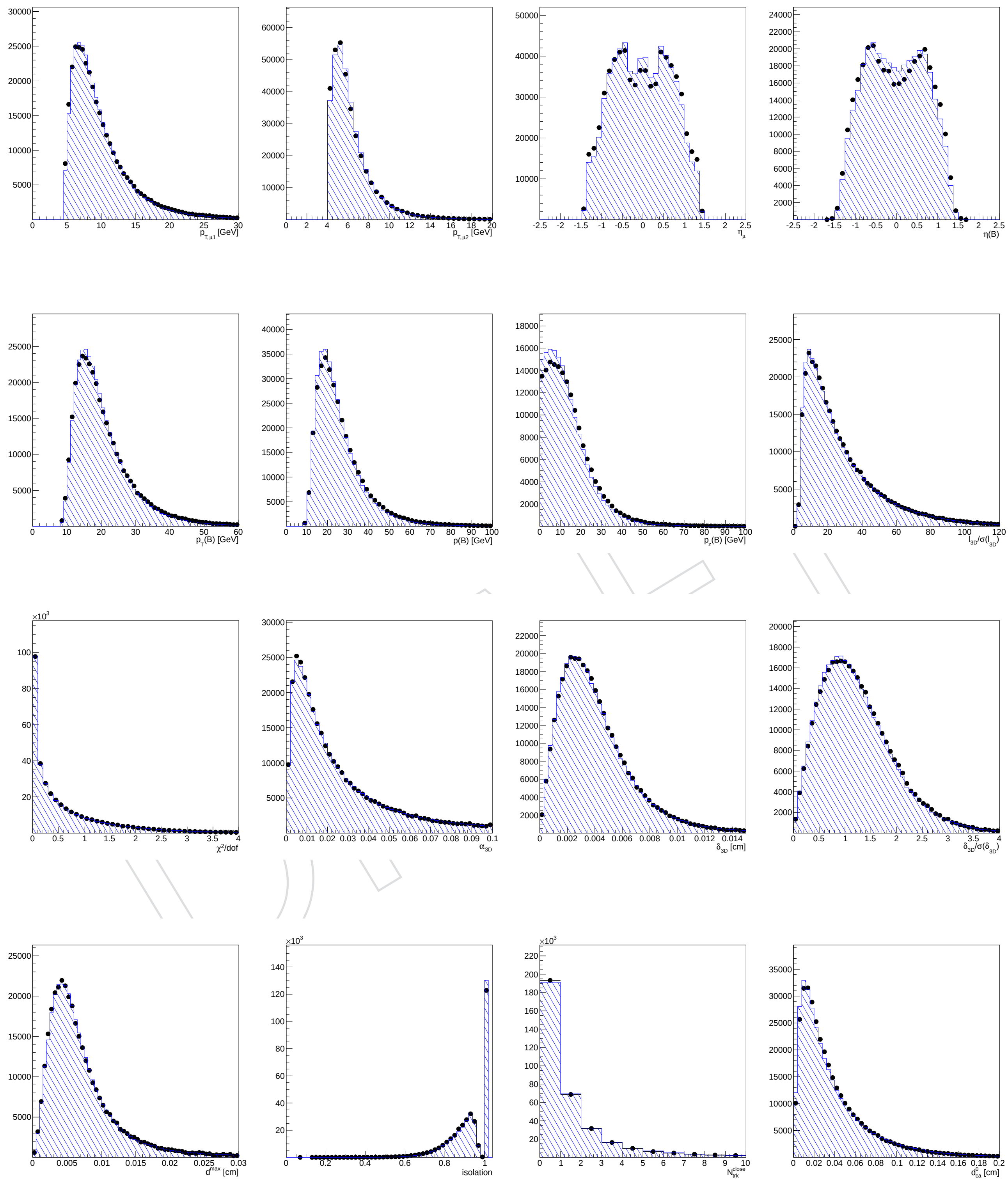

Figure 38: Comparison of barrel $B^{\pm} \to J/\psi K^{\pm}$ candidate distributions in data (solid markers) and MC simulation (hatched histogram). For each figure, the selection criteria $b > 0$ is applied. The MC simulation histogram is normalized to the same number of events like the data histogram.

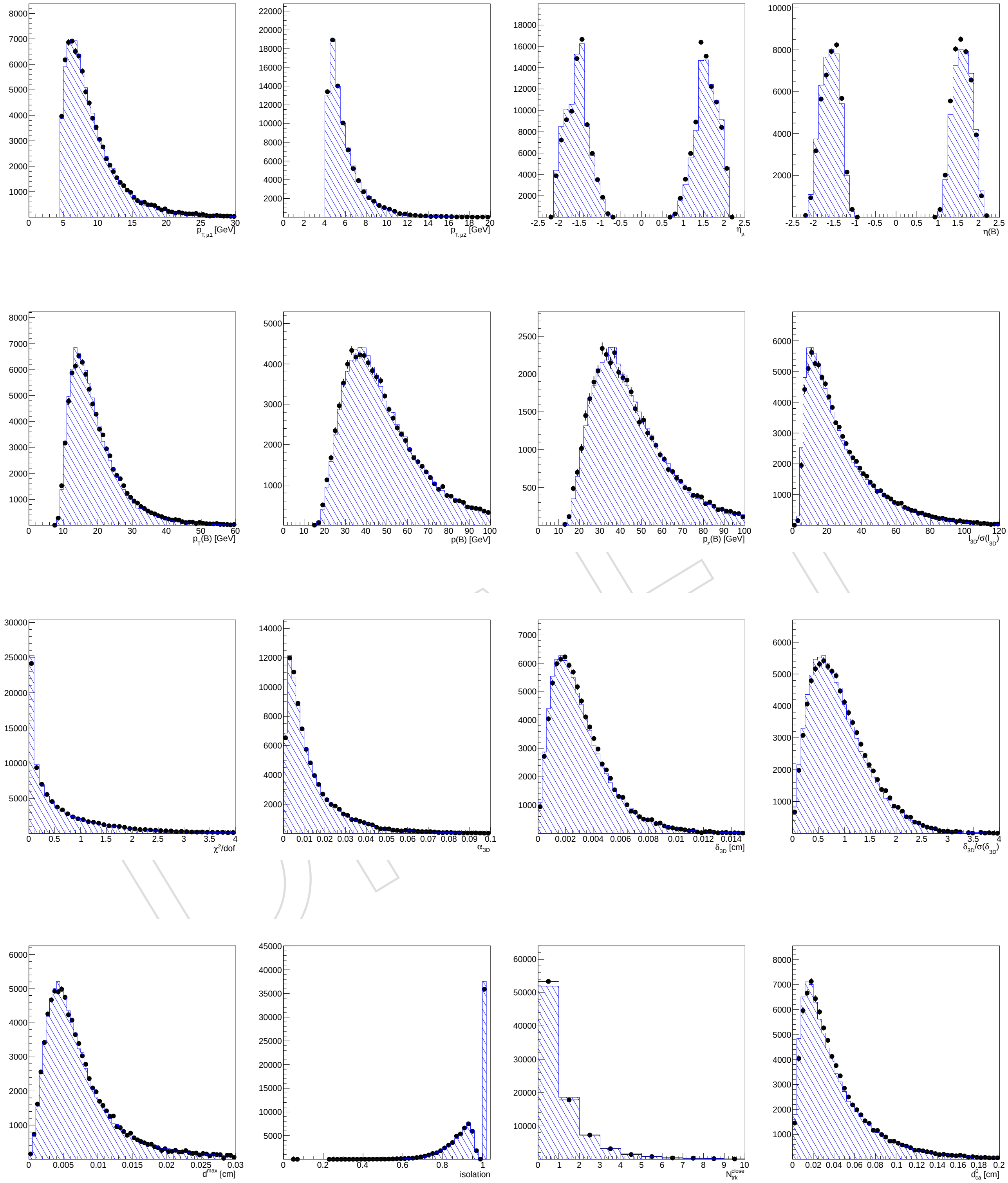

Figure 39: Comparison of endcap $B^{\pm} \to J/\psi K^{\pm}$ candidate distributions in data (solid markers) and MC simulation (hatched histogram). For each figure, the selection criteria $b > 0$ is applied. The MC simulation histogram is normalized to the same number of events like the data histogram.

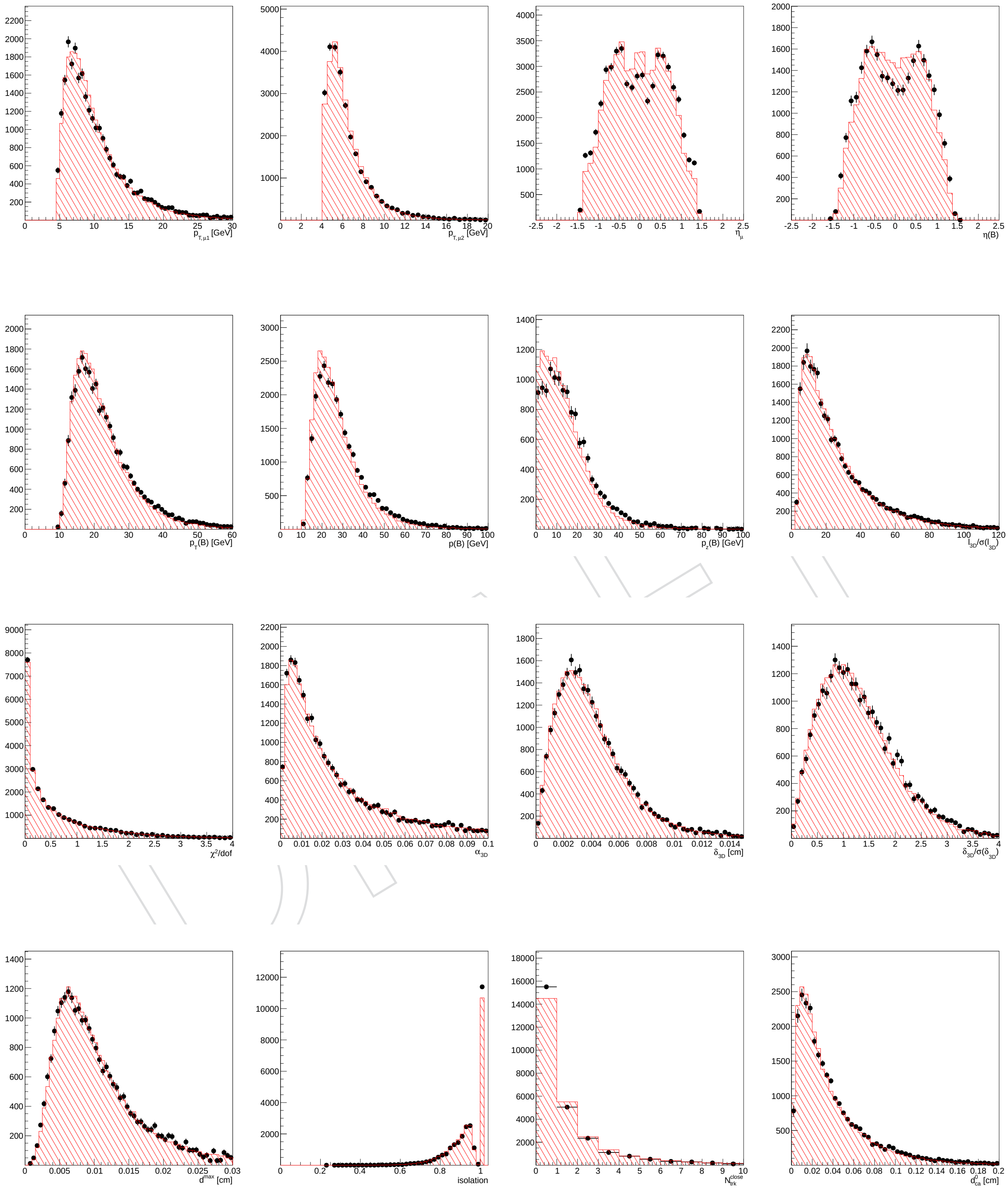

Figure 40: Comparison of barrel $B_s^0 \to J/\psi\,\phi$ candidate distributions in data (solid markers) and MC simulation (hatched histogram). For each figure, the selection criteria $b > 0$ is applied. The MC simulation histogram is normalized to the same number of events like the data histogram.

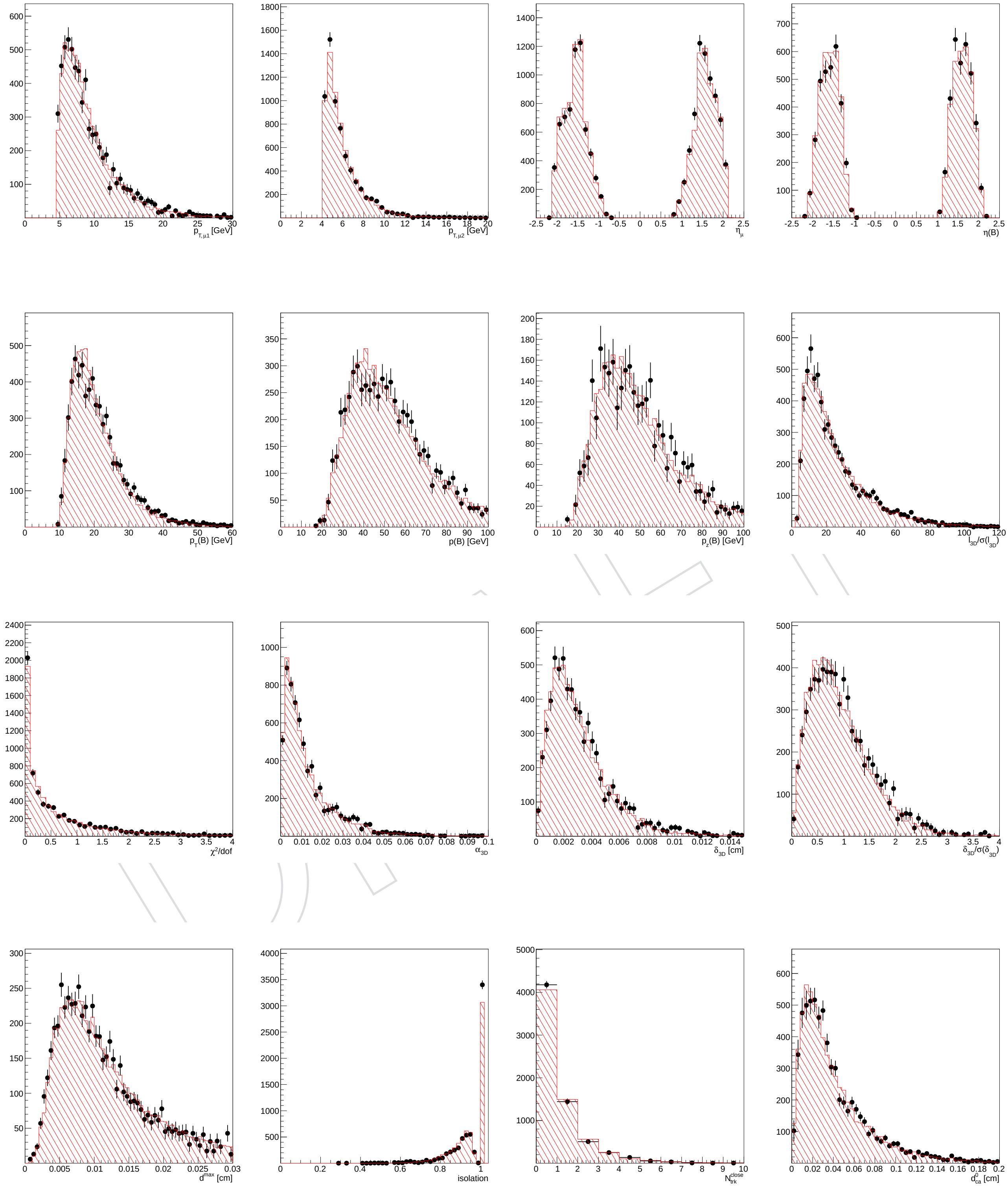

Figure 41: Comparison of endcap $B_s^0 \to J/\psi\phi$ candidate distributions in data (solid markers) and MC simulation (hatched histogram). For each figure, the selection criteria $b > 0$ is applied. The MC simulation histogram is normalized to the same number of events like the data histogram.

7 Muon identification and trigger efficiency

Note: The tag-and-probe tables used here are still from 2011. The MC efficiencies have been determined on the 2012 MC samples.

7.1 Comparison of the tag-and-probe efficiency with dimuon efficiency

To compare the tag-and-probe single-muon efficiency to the MC simulation muon efficiency, the single-muon efficiencies have to be combined

$$\varepsilon^{TNP} = \varepsilon^{TNP}(p_{\perp\mu1}, \eta_{\mu1}) \times \varepsilon^{TNP}(p_{\perp\mu2}, \eta_{\mu2}). \tag{6}$$

It is not clear to what precision such a factorizing approach holds. This is investigated in more details in this section. We define the ratio ρ_{TNP} as

$$\rho_{TNP} = \frac{\varepsilon(\text{dimuon})}{\langle \varepsilon^{TNP}(p_{\perp\mu1}, \eta_{\mu1}) \times \varepsilon^{TNP}(p_{\perp\mu2}, \eta_{\mu2}) \rangle}. \tag{7}$$

where the two single-muon efficiencies are computed in the Monte Carlo with the tag-and-probe procedure and the values are taken on a per-event basis contributing to the chosen dimuon $(p_T, |y|)$ bin. The numerator ε(dimuon is the dimuon MC efficiency. In Tab. 12 the ρ factors are provided for the signal and normalization samples.

Table 11: Trigger and muon identification efficiencies for signal and normalization, split into barrel (B) and endcap (E).

	trigger efficiency				muon identification			
	Signal		Normalization		Signal		Normalization	
$p_\perp >$	ε_B	ε_E	ε_B	ε_E	ε_B	ε_E	ε_B	ε_E
4.0 GeV	0.615	0.434	0.526	0.365	0.676	0.816	0.728	0.731
5.0 GeV	0.746	0.471	0.656	0.399	0.731	0.842	0.799	0.743
6.0 GeV	0.796	0.484	0.697	0.413	0.740	0.846	0.805	0.734
7.0 GeV	0.810	0.494	0.706	0.418	0.739	0.834	0.791	0.711
8.0 GeV	0.814	0.493	0.706	0.422	0.724	0.817	0.767	0.683
9.0 GeV	0.805	0.490	0.702	0.421	0.715	0.798	0.740	0.649
analysis	0.620	0.447	0.532	0.375	0.679	0.825	0.735	0.738

This analysis depends on the ratio of efficiencies for the signal and normalization sample. This ratio can be obtained for the efficiencies obtained by the tag-and-probe method or in MC simulation; we summarize the results in Tab. 13.

7.2 Study of the HLT efficiency

The HLT efficiency in 2012 is different from the 2011 version. In this section we summarize our studies. Figs. 42–45 show the HLT trigger efficiency for 2011 and 2012 overlayed for the signal and normalization in the barrel and the endcap. The events are selected as passing the BDT response criterion $b > 0.140$ ($b > 0.130$) in the barrel (endcap).

A few observations on Figs. 42–45

- In the barrel, the turn-on curve with respect to the muon $p_\perp$ is broader in 2012 than it used to be in 2011. This is also true (by consequence) for the B candidate $p_\perp$.

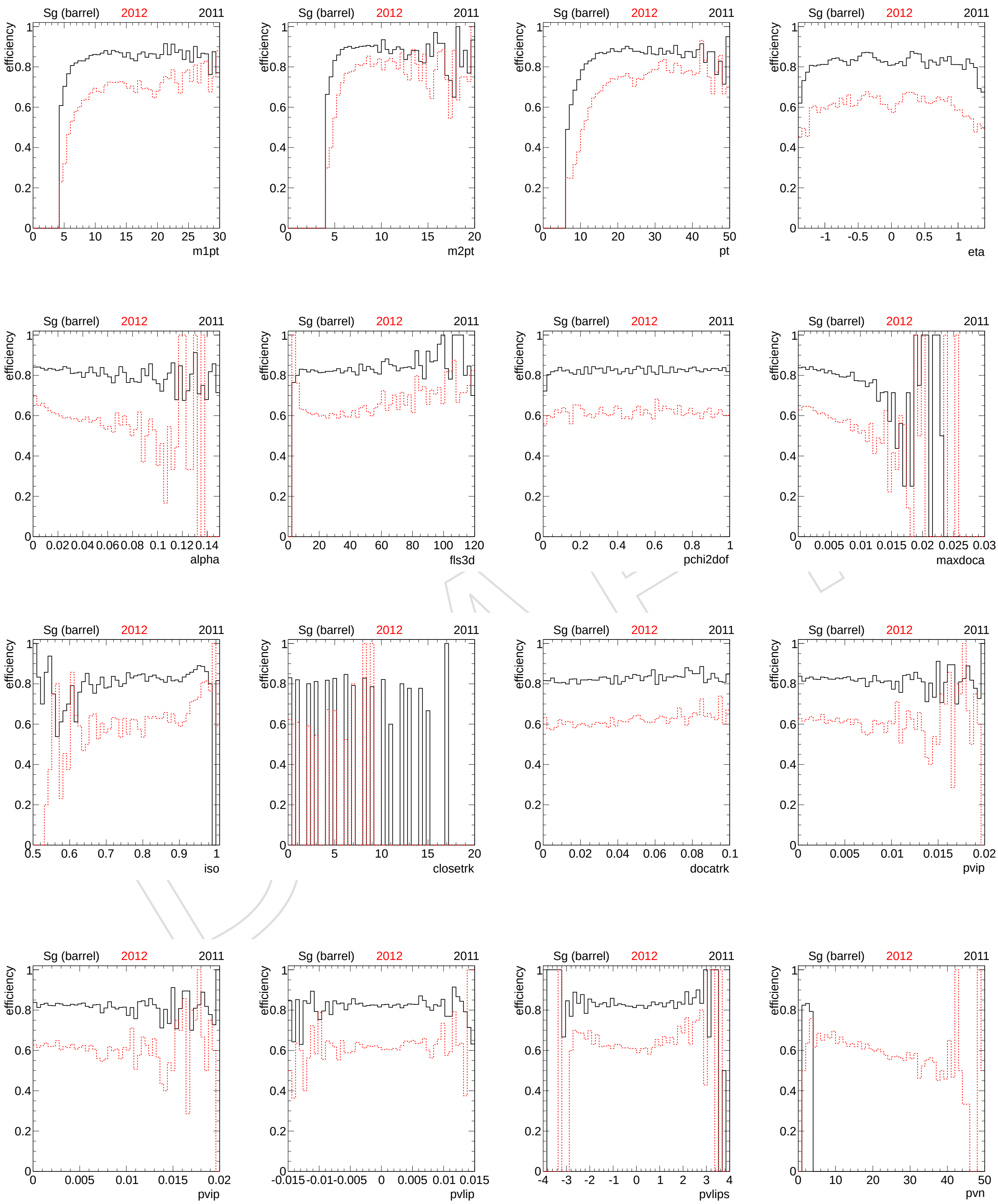

Figure 42: Comparison of barrel $B_s^0 \to \mu^+\mu^-$ MC trigger efficiency in 2012 (red dashed) and 2011 (black solid).

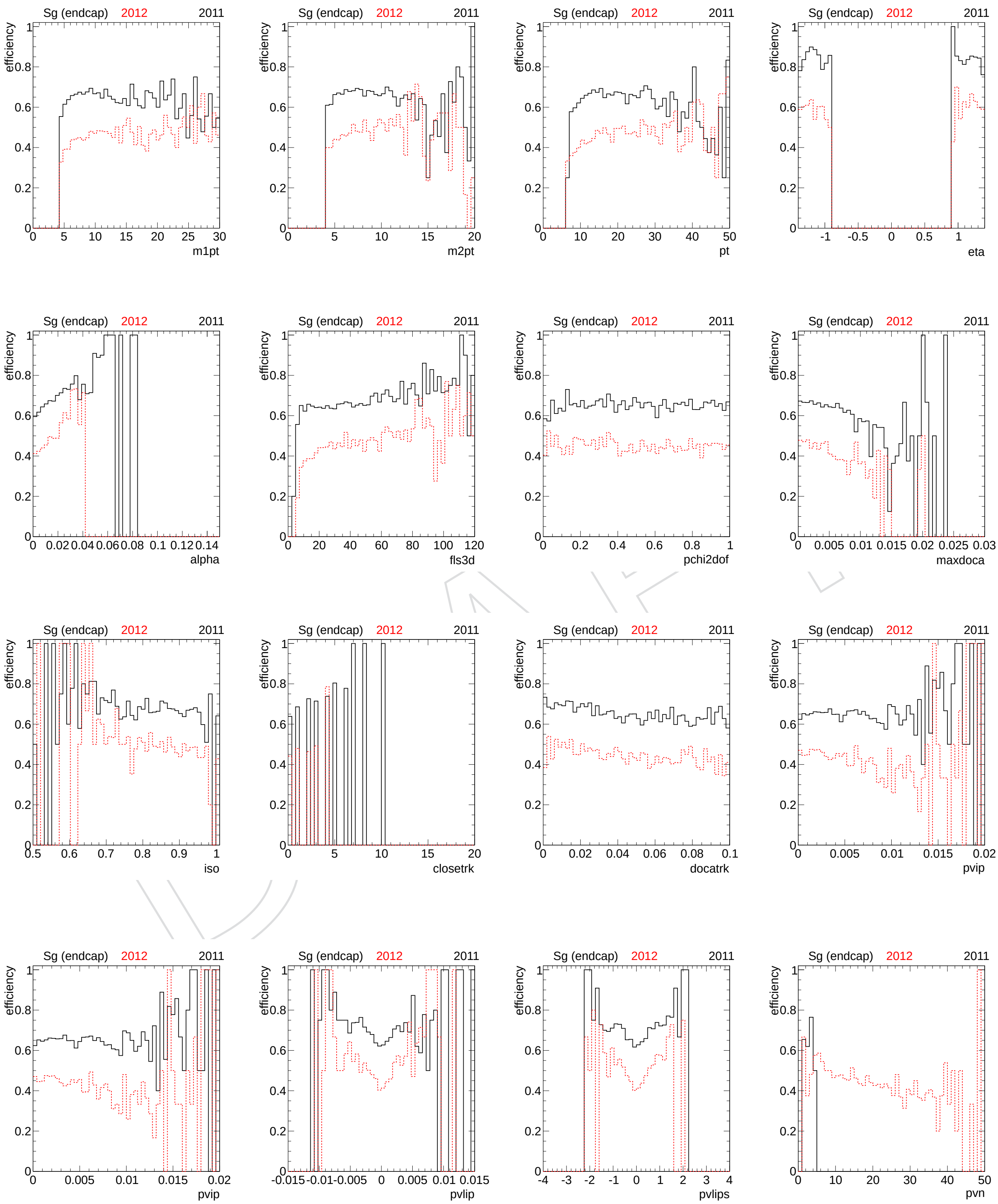

Figure 43: Comparison of endcap $B_s^0 \to \mu^+\mu^-$ MC trigger efficiency in 2012 (red dashed) and 2011 (black solid).

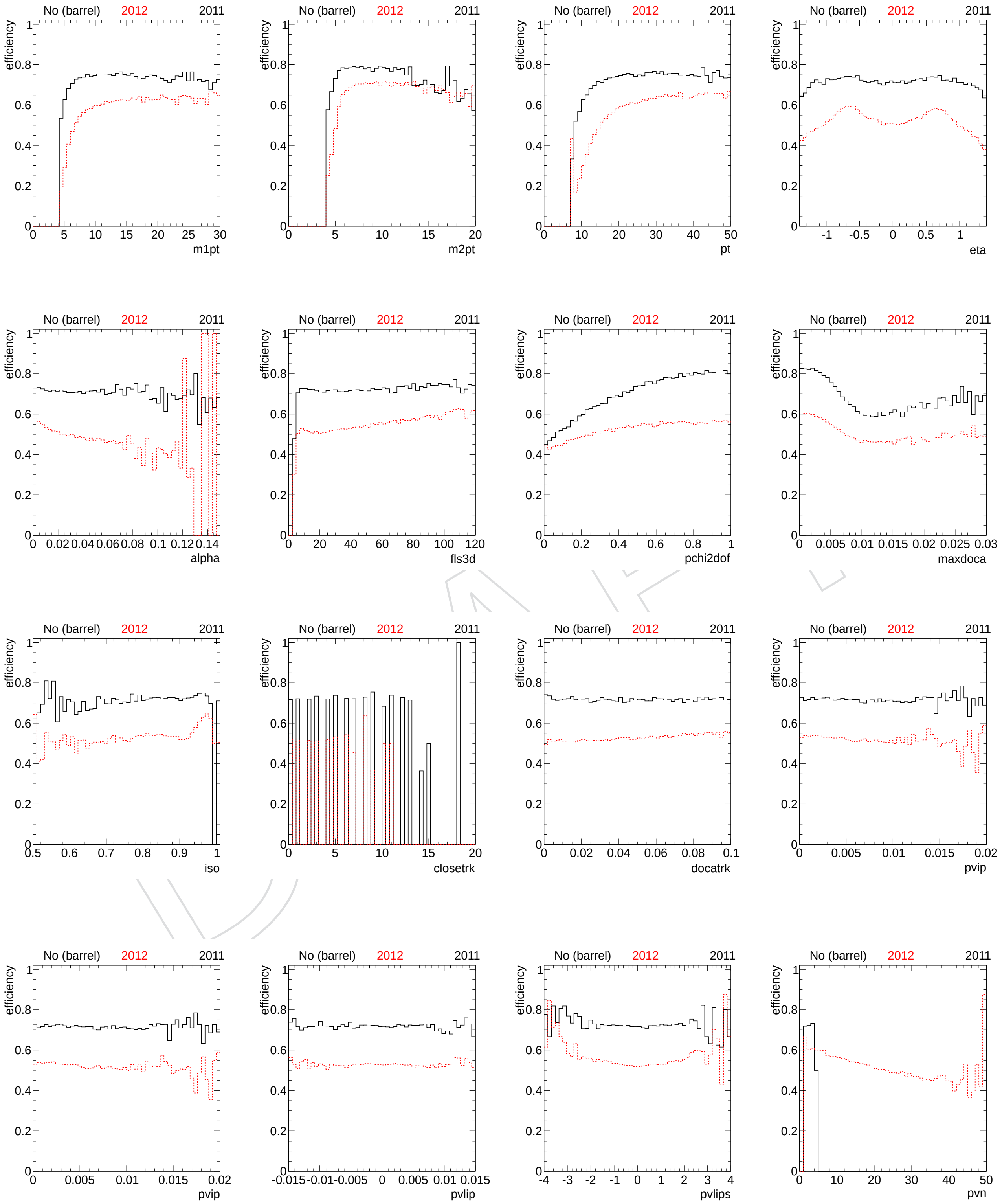

Figure 44: Comparison of barrel $B^{\pm} \to J/\psi K^{\pm}$ MC trigger efficiency in 2012 (red dashed) and 2011 (black solid).

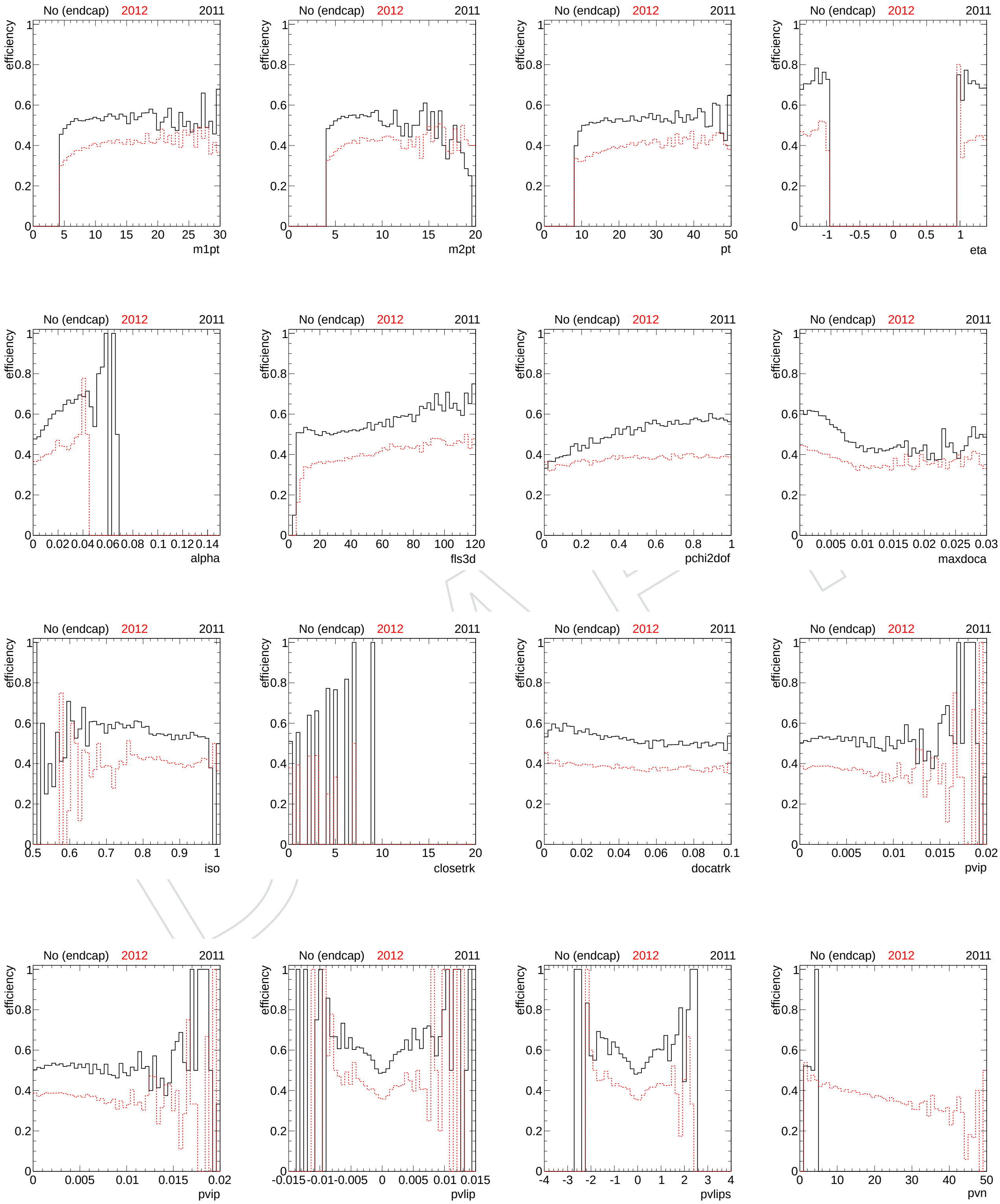

Figure 45: Comparison of endcap $B^{\pm} \to J/\psi K^{\pm}$ MC trigger efficiency in 2012 (red dashed) and 2011 (black solid).

Table 12: Trigger and muon identification efficiency ρ factors for signal and normalization, split into barrel (B) and endcap (E). The 'cowboy' veto is not applied to the normalization sample.

	trigger efficiency				muon identification			
	Signal		Normalization		Signal		Normalization	
$p_{\perp} >$	ρ_B	ρ_E	ρ_B	ρ_E	ρ_B	ρ_E	ρ_B	ρ_E
4.0 GeV	0.736	0.584	0.639	0.514	0.868	0.987	0.953	0.885
5.0 GeV	0.814	0.618	0.717	0.538	0.840	0.988	0.921	0.871
6.0 GeV	0.852	0.640	0.747	0.552	0.833	0.988	0.908	0.855
7.0 GeV	0.862	0.661	0.752	0.556	0.832	0.981	0.890	0.832
8.0 GeV	0.864	0.661	0.749	0.559	0.817	0.974	0.866	0.808
9.0 GeV	0.856	0.654	0.746	0.557	0.808	0.969	0.837	0.781
analysis	0.737	0.595	0.641	0.522	0.865	0.988	0.949	0.883

Table 13: Ratio of trigger and muon identification efficiency between signal and normalization, split into barrel and endcap. The TNP numbers are *not* corrected with the ρ factors.

	$\varepsilon_{trig}^{B_s^0}/\varepsilon_{trig}^{B^+}$(barrel)			$\varepsilon_{trig}^{B_s^0}/\varepsilon_{trig}^{B^+}$(endcap)		
$p_{\perp} >$	TNP	TNP-MC	MC	TNP	TNP-MC	MC
4.0 GeV	1.012 ± 0.001	1.015 ± 0.001	1.169 ± 0.006	1.047 ± 0.001	1.049 ± 0.002	1.191 ± 0.011
5.0 GeV	1.000 ± 0.000	1.003 ± 0.000	1.137 ± 0.005	1.031 ± 0.002	1.028 ± 0.002	1.182 ± 0.013
6.0 GeV	0.999 ± 0.000	1.001 ± 0.000	1.142 ± 0.006	1.018 ± 0.003	1.011 ± 0.003	1.172 ± 0.017
7.0 GeV	0.999 ± 0.000	1.001 ± 0.000	1.147 ± 0.007	1.002 ± 0.004	0.995 ± 0.004	1.182 ± 0.022
8.0 GeV	0.999 ± 0.001	1.000 ± 0.000	1.154 ± 0.009	0.993 ± 0.005	0.988 ± 0.006	1.167 ± 0.028
9.0 GeV	0.999 ± 0.001	1.000 ± 0.001	1.148 ± 0.012	0.995 ± 0.007	0.992 ± 0.007	1.165 ± 0.035
analysis	1.009 ± 0.001	1.012 ± 0.001	1.165 ± 0.006	1.043 ± 0.002	1.043 ± 0.002	1.190 ± 0.011

	$\varepsilon_{\mu}^{B_s^0}/\varepsilon_{\mu}^{B^+}$ (barrel)			$\varepsilon_{\mu}^{B_s^0}/\varepsilon_{\mu}^{B^+}$ (endcap)		
$p_{\perp} >$	TNP	TNP-MC	MC	TNP	TNP-MC	MC
4.0 GeV	1.008 ± 0.001	1.018 ± 0.001	0.928 ± 0.003	0.998 ± 0.001	1.000 ± 0.001	1.116 ± 0.004
5.0 GeV	0.999 ± 0.000	1.003 ± 0.000	0.915 ± 0.004	0.996 ± 0.001	0.999 ± 0.001	1.133 ± 0.005
6.0 GeV	0.998 ± 0.000	1.001 ± 0.000	0.919 ± 0.005	0.996 ± 0.001	0.997 ± 0.001	1.152 ± 0.006
7.0 GeV	0.997 ± 0.001	1.000 ± 0.000	0.935 ± 0.006	0.992 ± 0.002	0.995 ± 0.001	1.173 ± 0.009
8.0 GeV	0.998 ± 0.001	1.001 ± 0.000	0.944 ± 0.008	0.986 ± 0.002	0.992 ± 0.002	1.196 ± 0.012
9.0 GeV	0.999 ± 0.001	1.000 ± 0.000	0.966 ± 0.010	0.985 ± 0.003	0.991 ± 0.002	1.229 ± 0.017
analysis	1.006 ± 0.001	1.014 ± 0.001	0.924 ± 0.003	0.998 ± 0.001	0.999 ± 0.001	1.119 ± 0.004

- There are various quite strong dependencies on variables that are not obvious, e.g., pointing angle α, 3D flight length signficance (fls3d), etc.
- The HLT shows a dependence on the isolation in 2012 which was not present in 2011.
- There is a pronounced dependence of the HLT efficiency on the number of primary vertices.
- In 2012, the endcap trigger was changed with respect to the 2011 version: The η range was restricted, but the threshold were lowered between $1.4 < |\eta| < 1.8$ (albeit with an apparent broadened threshold, see first item in this list). We have not investigated where the absolute efficiency level would move for the endcap, given these changes.

The analysis depends only on the ratio of trigger efficiency between the signal and the normalization mode. Fig. 46 illustrates the ratio vs. the number of primary vertices. It is observed that this ratio does not show a significant slope.

With the different mean primary vertex distributions in data and MC simulation (shown in Fig. 2) we derive a trigger efficiency ratio difference between data and MC simulation of less

Table 14: Ratio of trigger and muon identification efficiency between signal and normalization, split into barrel and endcap. The TNP numbers are corrected with the ρ factors.

	$\varepsilon^{B^0_s}_{trig}/\varepsilon^{B^+}_{trig}$(barrel)			$\varepsilon^{B^0_s}_{trig}/\varepsilon^{B^+}_{trig}$(endcap)		
$p_\perp >$	TNP	TNP-MC	MC	TNP	TNP-MC	MC
4.0 GeV	1.166 ± 0.001	1.169 ± 0.001	1.169 ± 0.006	1.190 ± 0.003	1.191 ± 0.003	1.191 ± 0.011
5.0 GeV	1.134 ± 0.000	1.137 ± 0.000	1.137 ± 0.005	1.186 ± 0.004	1.182 ± 0.004	1.182 ± 0.013
6.0 GeV	1.140 ± 0.000	1.142 ± 0.000	1.142 ± 0.006	1.180 ± 0.005	1.172 ± 0.006	1.172 ± 0.017
7.0 GeV	1.145 ± 0.001	1.147 ± 0.000	1.147 ± 0.007	1.191 ± 0.007	1.182 ± 0.008	1.182 ± 0.022
8.0 GeV	1.152 ± 0.001	1.154 ± 0.001	1.154 ± 0.009	1.173 ± 0.010	1.167 ± 0.011	1.167 ± 0.028
9.0 GeV	1.146 ± 0.001	1.148 ± 0.001	1.148 ± 0.012	1.168 ± 0.012	1.165 ± 0.014	1.165 ± 0.035
analysis	1.162 ± 0.001	1.165 ± 0.001	1.165 ± 0.006	1.190 ± 0.003	1.190 ± 0.003	1.190 ± 0.011

	$\varepsilon^{B^0_s}_{\mu}/\varepsilon^{B^+}_{\mu}$ (barrel)			$\varepsilon^{B^0_s}_{\mu}/\varepsilon^{B^+}_{\mu}$ (endcap)		
$p_\perp >$	TNP	TNP-MC	MC	TNP	TNP-MC	MC
4.0 GeV	0.918 ± 0.001	0.928 ± 0.001	0.928 ± 0.003	1.113 ± 0.001	1.116 ± 0.001	1.116 ± 0.004
5.0 GeV	0.911 ± 0.000	0.915 ± 0.000	0.915 ± 0.004	1.130 ± 0.001	1.133 ± 0.001	1.133 ± 0.005
6.0 GeV	0.916 ± 0.000	0.919 ± 0.000	0.919 ± 0.005	1.151 ± 0.001	1.152 ± 0.001	1.152 ± 0.006
7.0 GeV	0.932 ± 0.001	0.935 ± 0.000	0.935 ± 0.006	1.169 ± 0.002	1.173 ± 0.001	1.173 ± 0.009
8.0 GeV	0.942 ± 0.001	0.944 ± 0.000	0.944 ± 0.008	1.188 ± 0.003	1.196 ± 0.002	1.196 ± 0.012
9.0 GeV	0.964 ± 0.001	0.966 ± 0.001	0.966 ± 0.010	1.221 ± 0.004	1.229 ± 0.003	1.229 ± 0.017
analysis	0.917 ± 0.001	0.924 ± 0.001	0.924 ± 0.003	1.117 ± 0.001	1.119 ± 0.001	1.119 ± 0.004

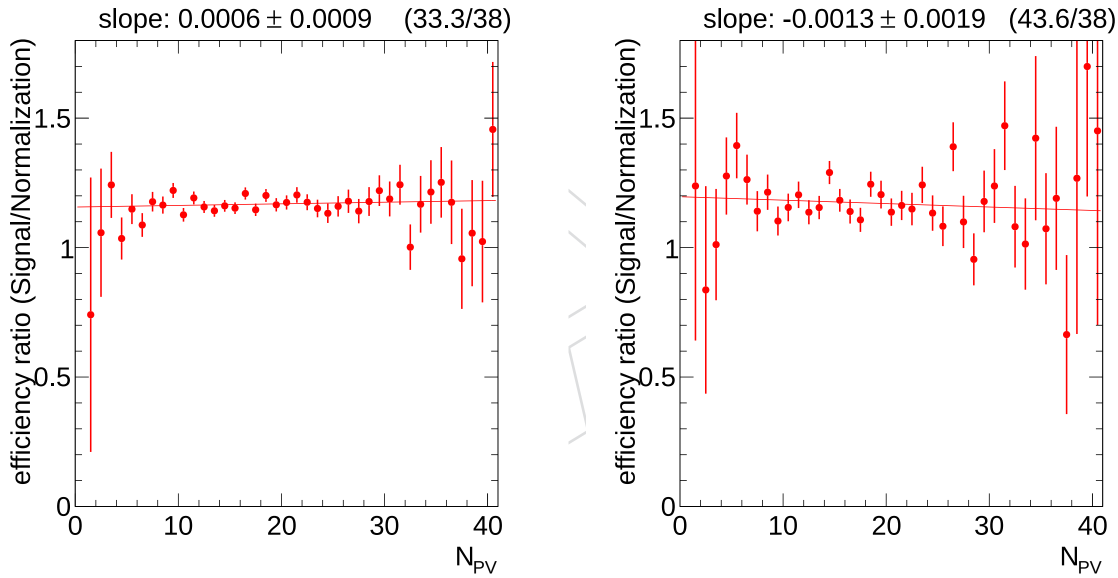

Figure 46: Ratio of the HLT efficiency for the $B^0_s \to \mu^+\mu^-$ signal and the $B^\pm \to J/\psi K^\pm$ normalization in the barrel (left) and endcap (right) vs. the number of primary vertices. A straight line is fitted to the data. The ratio is independent of pileup. The numbers in parentheses give the χ^2/dof for the fit.

than 0.2%.

7.3 Determination of muon misidentification with non-muon primary datasets

In order to estimate the muon misidentification rate we have processed three primary 2011 datasets which are independent of muon triggers, as detailed in Tab. 2. In these files we reconstruct the decay chain $D^* \to D^0\pi(slow) \to K\pi\pi(slow)$. The reconstruction followed the following steps:

- Two opposite charge hadron tracks above $p_\perp > 4\,\text{GeV}$ are combined to a vertex to form a D^0 candidate. The D^0 candidate is than combined with a low momentum $p_\perp > 0.4\,\text{GeV}$ track (called slow π) to another vertex to form the D^* candidate. This is performed with the sequential vertex fitter using in addition loose mass cuts of $\pm 100\,\text{MeV}$ on the D^0 mass and $\pm 30\,\text{MeV}$ on the mass difference. There is also a 0.3 cut on the maximum angle between the reconstructed D^* and the slow π directions.
- In further selection we apply tighter cuts on the slow pion ($p_\perp > 0.5\,\text{GeV}$), on the D^* candidate ($p_\perp > 6.0\,\text{GeV}$) , $\chi^2 < 2$ on the vertex fit, minimum flight significance of 2 for the D^0, maximum angle of 0.08 between the slow π and the D^*, maximum angle pf 0.2 between the D^0 momentum and its flight direction obtained from the vertex fits and tighter mass cuts of $\pm 40\,\text{MeV}$ and $\pm 45\,\text{MeV}$ on the D^* and D^0 respectively.
- The π and K identification was done by requiring that the π charge is equal to the slow π charge.

An example of the D^*D^0 mass difference distributions are shown in Fig. 47. The left-upper plot shows all events from the HT dataset which pass the selection cuts mentioned above. The right-upper plot shows events where a pion has been misidentified as a muon, the left-lower plot is the same but for kaons. Finally the right-lower plot is for pure events where neither the pion nor the kaon has been misidentified. The peaks in the mass difference distributions have been fitted by a single Gaussian using a Fermi distribution to parametrized the background.

The final results are summarized in Tab. 15, where the pion and kaon misidentification rates are shown for different data sets and different selections.

Table 15: A summary of the muon misidentification rates η for pion and kaons shown for different data sets and different selections. The top part of the table provides the numbers for the 'tight muon' selection [34] as used in the analysis, the bottom part of the table shows more 'differential' studies. The rows called "All" combine the three data sets in order to increase statistics.

Data set	pions	kaons
HT	$(0.97 \pm 0.24) \times 10^{-3}$	$(0.95 \pm 0.22) \times 10^{-3}$
All	$(1.04 \pm 0.20) \times 10^{-3}$	$(1.04 \pm 0.19) \times 10^{-3}$
All kink	$(1.04 \pm 0.20) \times 10^{-3}$	$(0.93 \pm 0.18) \times 10^{-3}$
All positive charge	$(1.10 \pm 0.28) \times 10^{-3}$	$(1.62 \pm 0.29) \times 10^{-3}$
All negative charge	$(1.00 \pm 0.28) \times 10^{-3}$	$(0.55 \pm 0.26) \times 10^{-3}$
All $\eta < 1.4$	$(1.12 \pm 0.20) \times 10^{-3}$	$(1.18 \pm 0.20) \times 10^{-3}$
All $\eta > 1.4$	$(1.12 \pm 0.73) \times 10^{-3}$	$(0.63 \pm 0.73) \times 10^{-3}$
All $p_\perp < 7.5\,\text{GeV}$	$(0.67 \pm 0.21) \times 10^{-3}$	$(0.88 \pm 0.24) \times 10^{-3}$
All $p_\perp > 7.5\,\text{GeV}$	$(2.44 \pm 0.58) \times 10^{-3}$	$(1.73 \pm 0.49) \times 10^{-3}$

In the first two rows we show the results used for the rare background estimation. Further results for different selections are shown for the combined data sets. A few explanations on the rows:

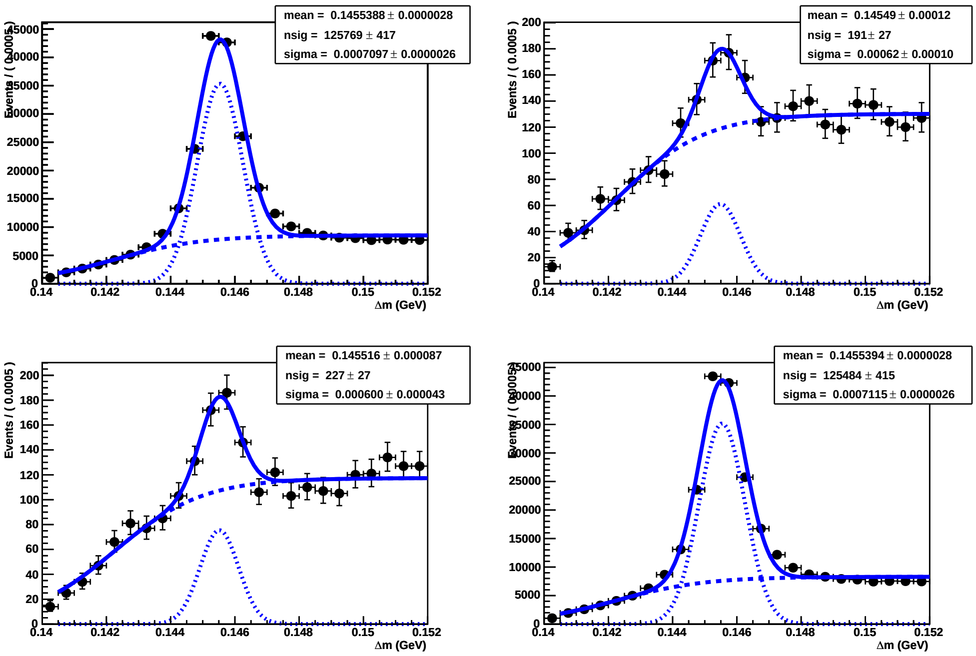

Figure 47: Mass difference distribution between the D^* and D^0 mesons obtained from the HT data set. See text for more detailed explanation.

- The "All kink" applies the "kink" finder cut of $\chi^2 < 15$ for the muon track.
- The "All positive charge" and "All negative charge" indicates the charge of pions and kaons.
- The $p_\perp$ and η selections include cuts on the pions and kaon as specified.

The mis-identification ratios in Tab. 15 are consistent with those reported in CMS Note [36], when those authors apply similar selection criteria in similar hadronic environments.

Within statistical uncertainties the rates provided in Tab. 15 are consistently described by single numbers covering the entire $p_\perp$ and η spectrum. We use $\eta = 10^{-3}$ for the pion and kaon misidentification rate.

For the proton misidentification rate we use $\eta = 5 \times 10^{-4}$.

The uncertainties on the misidentication rates are estimated at 25%.

8 Systematic Uncertainties and Crosschecks

This section describes the systematic studies and crosschecks that were done to estimate and corroborate the systematic errors of the analysis.

8.1 Crosscheck: Normalization yield for different run ranges

With time various conditions related to the data taking might change, e.g. LHC luminosity and trigger selection. Our event selection efficiency might depend on these conditions. In order to check this the run period for which the data is presented here has been split into 6 periods of roughly with equal luminosity, about 2 fb^{-1} each. As an additional check the run period has been split into the 4 HLT periods spaning the 5E33 and 7E33 menus, as explained in section 3.3 the triggers used for this anaysis have not changed during 2012, therefore we do not expect any changed in the yield rate. The data included here are for until the technical stop number 3 which was in September.

The results are summarized in Fig. 48 for the $B^{\pm} \to J/\psi K^{\pm}$ decay yield, Fig. 49 for the $B^0_s \to J/\psi \phi$ decay yield and Fig. 50 for the $B^0_s \to \mu^+\mu^-$ signal yield. The yields are always normalized to 1 pb^{-1}. In addition to the full acceptance yield (black squares) the plots also show barrel yields for $\eta < 1.4$. For the normalization and control channel yields we fit the signal mass peak with a single Gaussian after subtracting the background. For the $B^0_s \to \mu^+\mu^-$ we count the number of events in the mass window from 4.8 GeV to 6.0 GeV keeping the signal window (5.2 GeV-5.45 GeV) blank. In order to increase the number of events the standard isolation cuts were removed. The main feature of all data points is that there is a linear rise of the normalized yields with the run number. This is presently not understood. The reason that the last poinst are always low is that some of the data which corresponds to this running period have not been processed yet.

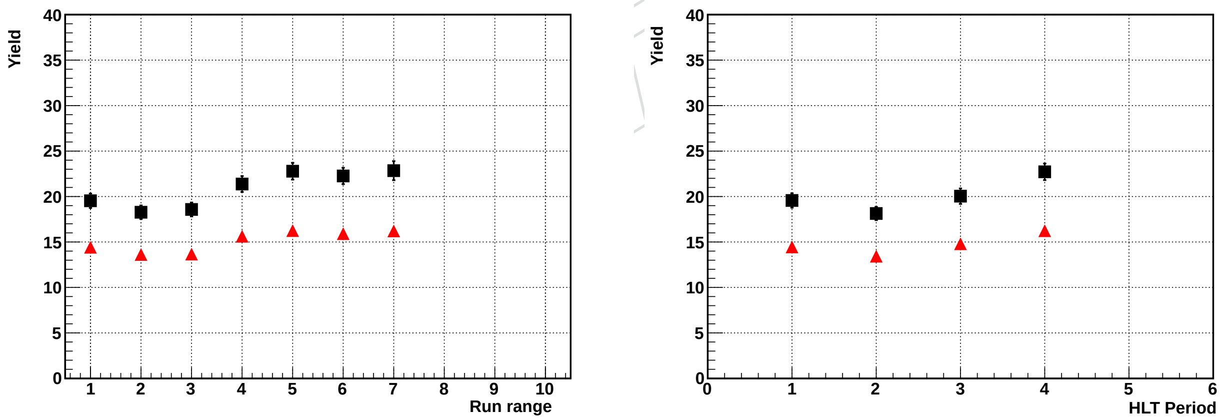

Figure 48: Measured yields for $B^{\pm} \to J/\psi K^{\pm}$ data split into 6 run periods of roughly equal luminosity (left plot) and into 4 HLT periods (right plot). The yield for each period is normalized to a luminosity of 1 pb^{-1}. The square (black) symbols show the full acceptance range and the triangles(red) show the yields for $\eta < 1.4$.

As an additional check we have divided the $B^0_s \to \mu^+\mu^-$ signal yields from Fig. 50 with the normalisation $B^{\pm} \to J/\psi K^{\pm}$ yields from Fig. 48. The resulting plots are shown in Fig. 51 with a 0-order polynomial fit included. These ratios are constant as shown by the fit χ^2 value and the low error on the average.

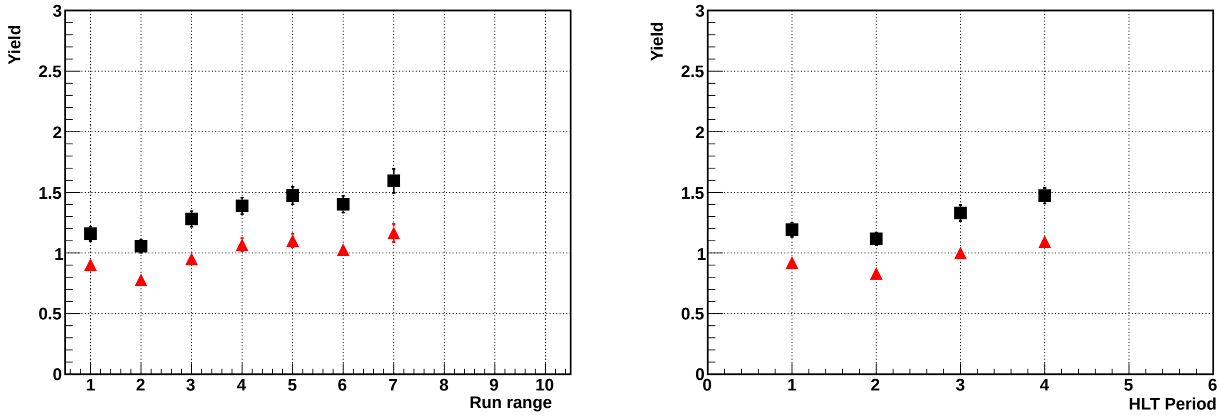

Figure 49: Similar to Fig. 48 but for $B_s^0 \to J/\psi\phi$ yields.

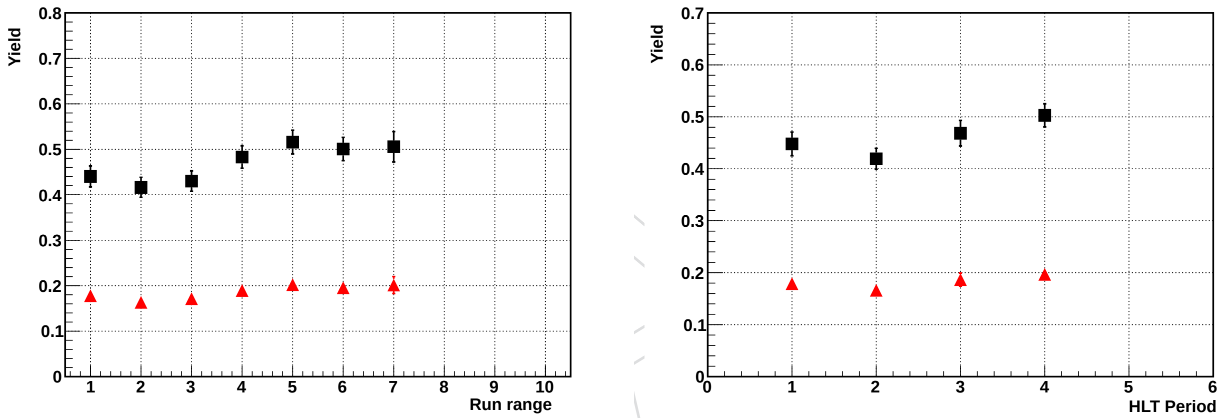

Figure 50: Similar to Fig. 48 but for $B_s^0 \to \mu^+\mu^-$ yields.

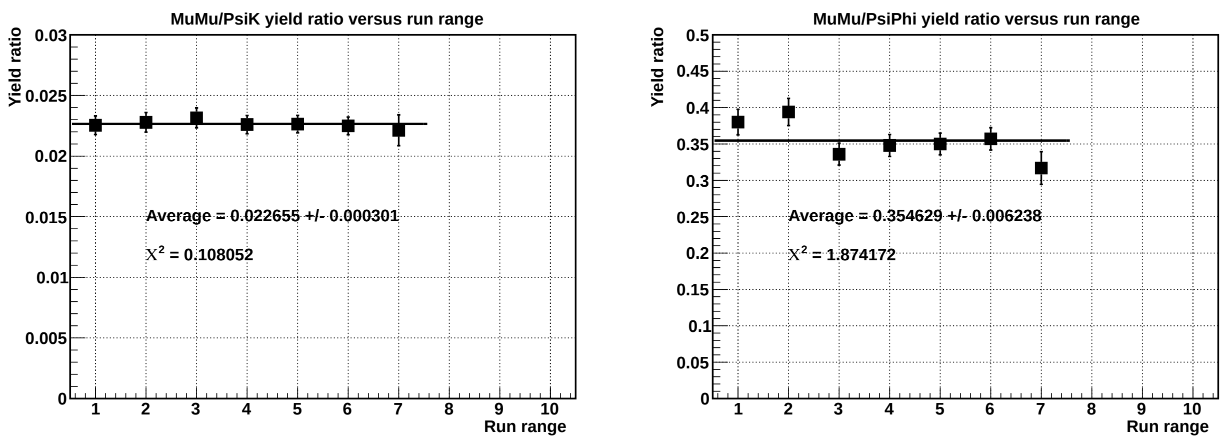

Figure 51: The ratio of $B_s^0 \to \mu^+\mu^-$ yields divided by $B^\pm \to J/\psi K^\pm$ yields on the right plot, and the ratio of $B_s^0 \to \mu^+\mu^-$ yields divided by $B_s^0 \to J/\psi\phi$ yields in the left plot.

8.2 BDT selection efficiency uncertainty

A preliminary systematic error is studied for the BDT selection. Starting from the sideband subtracted BDT response distribution for the normalization and control sample (figure 23), we evaluate the systematic error, summarized in Tab. 16, as follows.

- To account for systematic uncertainties in the normalization efficiency between data and MC simulation, we use the different selection efficiency for $b > 0.140$ ($b > 0.130$) of the $B^{\pm} \to J/\psi K^{\pm}$ sample in the barrel (endcap). Fig. 52 illustrates the distributions.
- We use the efficiency difference between the signal MC simulation and the $B_s^0 \to J/\psi \phi$ MC simulation to account for differences between the control sample and the signal. Fig. 53 illustrates the distributions.
- To account for systematic uncertainties in the signal efficiency between data and MC simulation, we use the different selection efficiency for $b >$ of the $B_s^0 \to J/\psi \phi$ sample. Fig. 54 illustrates the distributions.

Figure 52 shows the distribution of the cumulative number of event cut as a function of the BDT value for the normalization sample, Fig. 53 for the control sample.

Table 16: Relative systematic error for the BDT selection efficiency. The bottom row is an alternative to the (quadratic) sum of the two rows in the middle of the table.

Type	Barrel	Endcap
Normalization sample data/MC	0.010	0.044
Control sample data/MC	0.035	0.008
Signal/Control sample MC	0.093	0.117
MC Signal/Data control sample	0.127	0.109

It is evident that the systematic uncertainty, evaluated from the difference of the control sample to the signal sample, is quite high. It is not clear that this estimate is a good estimator of the systematic uncertainty. More work is underway.

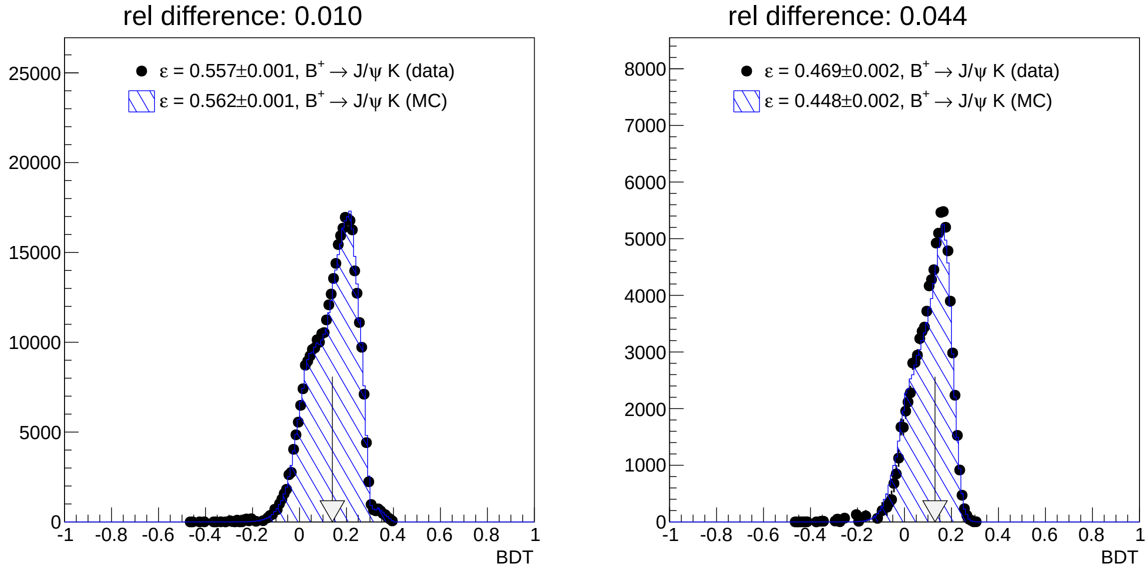

Figure 52: The BDT response b for $B^{\pm} \to J/\psi K^{\pm}$ for the barrel (left) and the endcap (right). The data is shown in black solid markers, the MC simulation in hatched histogram. The numbers printed onto the plot provide the efficiency for the requirement $b > 0.140$ ($b > 0.140$) in the barrel (endcap).

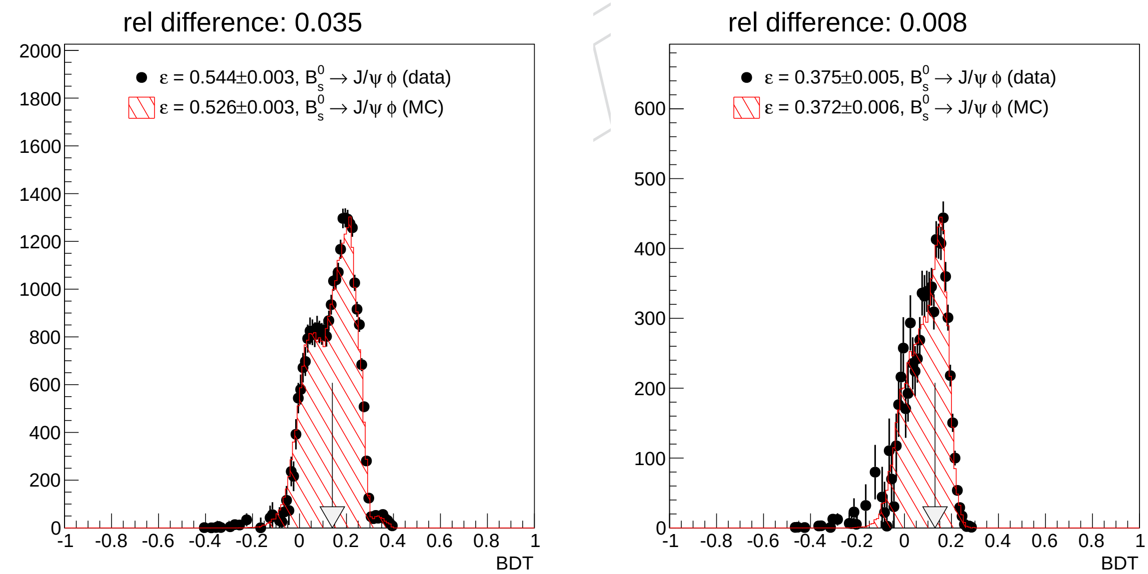

Figure 53: The BDT response b for $B_s^0 \to J/\psi\phi$ for the barrel (left) and the endcap (right). The data is shown in black solid markers, the MC simulation in hatched histogram. The numbers printed onto the plot provide the efficiency for the requirement $b > 0.140$ ($b > 0.140$) in the barrel (endcap)

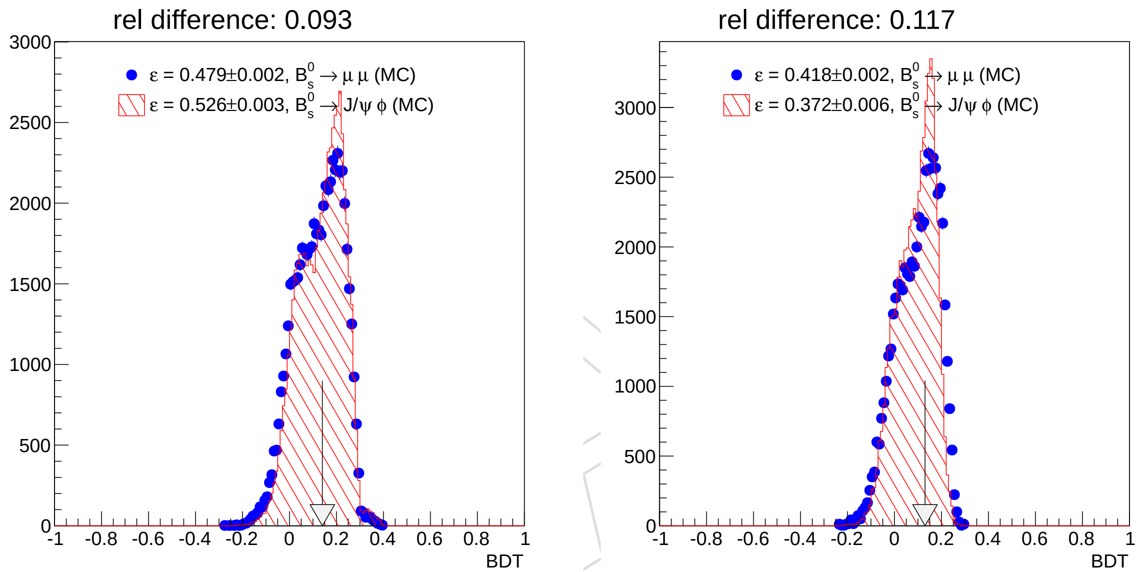

Figure 54: The BDT response b for $B_s^0 \to \mu^+\mu^-$ and $B_s^0 \to J/\psi\phi$ for the barrel (left) and the endcap (right). The numbers printed onto the plot provide the efficiency for the requirement $b > 0.140$ ($b > 0.140$) in the barrel (endcap)

9 Binned result extraction

9.1 Statistical Model

After all selection requirements, the candidate masses are filled into a histogram covering the full mass range as defined in Eq. 2. Within this histogram, three regions are defined:

- the $B_s^0 \to \mu^+\mu^-$ signal window for the barrel and one for the endcaps,
- the $B^0 \to \mu^+\mu^-$ signal window for the barrel and one for the endcaps,
- the background region, consisting of the complement of the above.

The number of entries in each window is considered as a random variable satisfying Poissonian statistics. The variables are denoted by

N_s^B	Number of candidates in $(B_s^0 \to \mu^+\mu^-)$-signal window and barrel region,
N_d^B	Number of candidates in $(B^0 \to \mu^+\mu^-)$-signal window and barrel region,
N_b^B	Number of candidates in background window and barrel region.

The variables N_s^E, N_d^E, N_b^E are defined with events in the endcap region in an analogous way. For simplicity, all considerations below are written for the barrel histogram although they are also equally valid for the endcap histogram.

The background observable is modeled as

$$N_b^B \quad \sim \quad \mathrm{Pois}(\nu_b^B + \nu_{b,\mathrm{rare}}^B),$$

where the expected number of combinatorial background events ν_b^B enters the model as a nuisance parameter and will be determined from the sideband. $\nu_{b,\mathrm{rare}}^B$ is the expected number of rare (peaking and semileptonic) background events in the sideband.

The ratio of combinatorial background in the signal windows with respect to the sideband window is given by the scale factor τ_s^B and τ_d^B. In total the Poissonian for the signal windows are of the following form

$$\begin{aligned} N_s^B &\sim \mathrm{Pois}(\tau_s^B \nu_b^B + \nu_{s,\mathrm{rare}}^B + P_{ss}^B \mu_s \nu_s^B + P_{sd}^B \mu_d \nu_d^B) \quad \text{and} \\ N_d^B &\sim \mathrm{Pois}(\tau_d^B \nu_b^B + \nu_{d,\mathrm{rare}}^B + P_{ds}^B \mu_s \nu_s^B + P_{dd}^B \mu_d \nu_d^B). \end{aligned}$$

The following additional notation is used for $i, j \in \{s, d\}$:

P_{ij}^B	Probability for a reconstructed $B_j^0 \to \mu\mu$ decay to be in $(B_i^0 \to \mu\mu)$-signal window.
μ_i	Signal strength of $B_i^0 \to \mu\mu$, that is the ratio of true branching ratio to SM branching ratio.

The expected number of reconstructed decays assuming SM is

$$\nu_i = \frac{\mathcal{B}^{\mathrm{SM}}(B_i^0 \to \mu\mu)}{\mathcal{B}(B^\pm \to J/\psi K^\pm)} \frac{f_s}{f_u} \frac{A_{B_s^0}}{A_{B^\pm}} \frac{\varepsilon_{\mathrm{trig}}^{B_s^0}}{\varepsilon_{\mathrm{trig}}^{B^+}} \frac{\varepsilon_{\mu}^{B_s^0}}{\varepsilon_{\mu}^{B^+}} \frac{\varepsilon_{\mathrm{analysis}}^{B_s^0}}{\varepsilon_{\mathrm{analysis}}^{B^+}} N^{\mathrm{obs}}(B^\pm \to J/\psi K^\pm). \tag{8}$$

Our total model thus consists of six Poissonian observables $(N_s^E, N_s^B, N_d^E, N_d^B, N_b^E, N_b^B)$.

The constraints for the combinatorial background and crossfeed is given by the Poissonians. All other constraints of the nuisance parameters are implemented as bifurcated Gaussians. The total likelihood function for each 2011 and 2012 measurement is given by

$$L(N^{B,E}_{s,d,b}; \mu_s, \mu_d, \nu) = \left(\prod_{\substack{k\in\{s,d,b\} \\ \ell\in\{B,E\}}} \mathrm{Pois}(N^\ell_k) \right) \times \left(\prod_k \mathrm{BifurGauss}(\nu_k) \right),$$

where ν stand for all nuisance parameters entering the model.

The total likelihood function to combine 2011 and 2012 measurement is simply given by the product of twice the function above.

$$L_{\mathrm{comb}} = L_{2011} \cdot L_{2012}$$

9.2 Upper Limit

The upper limit is computed using the CL_s method [37]. When computing the upper limit for $B^0_s \to \mu^+\mu^-$ ($B^0 \to \mu^+\mu^-$), the branching ration of $B^0 \to \mu^+\mu^-$ ($B^0_s \to \mu^+\mu^-$) is assumed unknown and enters the computation as a nuisance parameter as well. The discussion below is written only for $B^0_s \to \mu^+\mu^-$ but also holds for $B^0 \to \mu^+\mu^-$ analogously.

To test wether a fixed μ_s is in the confidence interval, a test statistic has to be chosen. We use the ratio of profiled likelihoods test statistic

$$q_{\mathrm{ul}} = \frac{L(N^{B,E}_{s,d,b}; \mu_s, \hat{\hat{\nu}}(\mu_s))}{L(N^{B,E}_{s,d,b}; 0, \hat{\hat{\nu}}(0))}, \tag{9}$$

where the double hat means the conditional maximum likelihood estimate for a given fixed parameter of interest. Furthermore, to compute the CL_{s+b} (CL_b) value of the given observation under the hypothesis μ_s, we need the pdf of the given test statistic assuming μ_s ($\mu_s = 0$). Toys are generated randomizing all nuisance parameters according to their constraint function (Frequentist-Bayesian Hybrid). For each toy experiment, the value observed in the sideband is kept fix at the actual observed value.

9.3 Significance

To test the significance of the result, one needs to compute a p-value for the given observation. As there is only the "background only" hypothesis, the ratio as above cannot be defined. Hence for significance computation we use the profiled likelihood ratio given by

$$q_{\mathrm{sig}} = \frac{L(N^{B,E}_{s,d,b}; \mu_s = 0, \hat{\hat{\nu}}(0))}{L(N^{B,E}_{s,d,b}; \hat{\mu}_s, \hat{\nu})}. \tag{10}$$

Toy generation to sample the pdf of q_{sig} is done analogously to the upper limit computation. The p-value can then be converted to the significance Z accoring to

$$Z = \sqrt{2} \cdot \mathrm{erf}^{-1}(1 - 2p).$$

9.4 Branching ratio extraction

The two sided interval is built using the all branching ratio values with a p-value of more than 32%. To test the p-value of a given branching ratio μ_s, the profiled likelihood ratio

$$q_{\mu_s} = \frac{L(N^{B,E}_{s,d,b}; \mu_s, \hat{\hat{\nu}}(\mu_s))}{L(N^{B,E}_{s,d,b}; \hat{\mu}_s, \hat{\nu})}. \tag{11}$$

is used again as test statistic. Toy generation analogously. The reported observed value is obtained from a maximum likelihood fit of the model to the observed values.

10 Unbinned maximum likelihood fit

The unblinded signal yield will be extracted with an Unbinned Maximum Likelihood fit (UML) on the invariant mass distribution.

10.1 The probability distribution function for each channel

There are four main contributions to the mass yield, and for each of them a different probability density function (pdf) has been chosen and studied. The four contributions are the **two signals** B_s^0 and B^0, the **combinatorial** background and the **rare** background, divided in the **semileptonic** and the **peaking** ones.

Every source is fitted to MC simulation events and every parameter is fixed to the fit result. Only the number of events of each contribution is let floating, to be fitted to the unblinded event yield.

Thus the final pdf is:

$$L = N_{B_s^0}F_{B_s^0} + N_{B^0}F_{B^0} + N_{\mathrm{comb}}F_{\mathrm{comb}} + N_{\mathrm{peak}}F_{\mathrm{peak}} + N_{\mathrm{semi}}F_{\mathrm{semi}} \tag{12}$$

where N_i is the number of events and F_i is the normalized pdf for each contribution i.

A per-event error (PEE) pdf has been implemented. In this case the widths of the B_s^0 and B^0 mesons are not constant and fixed to the MC simulation, but rather they are taken from each event. The width is taken as a function of the candidate η, and is defined as the sigma parameter of the Gaussian fitted on each η bin. The resulting width distribution has been fitted on a sixth degree polynomial (Fig. 55).

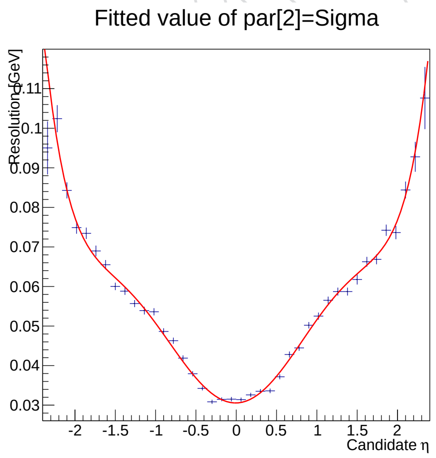

Figure 55: Resolution versus candidate pseudorapidity. A sixth degree polynomial fit is superimposed.

With this method the pdf for each signal is simply a single Crystal Ball, multiplied by the pdf of its mass resolution M^R :

$$F_{B_s^0} = CB(m|\mu_{B_s^0}, k_{B_s^0} \times \sigma(\eta), \alpha_{B_s^0}, n_{B_s^0}) \times M^R_{B_s^0} \tag{13}$$

for the B_s^0 meson, and similar for B^0:

$$F_{B^0} = CB(m|\mu_{B^0}, k_{B^0} \times \sigma(\eta), \alpha_{B^0}, n_{B^0}) \times M^R_{B^0} \tag{14}$$

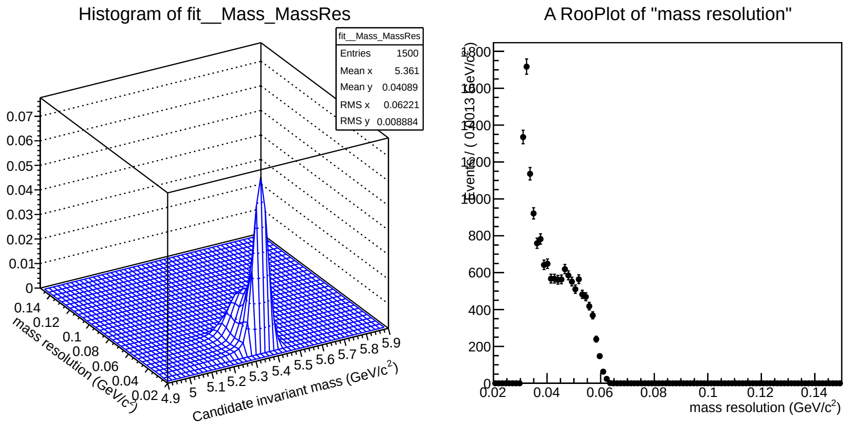

Figure 56: Mass resolution versus invariant mass for B_s^0 in the barrel (left) for 2012 data. Projection of the mass resolution (right).

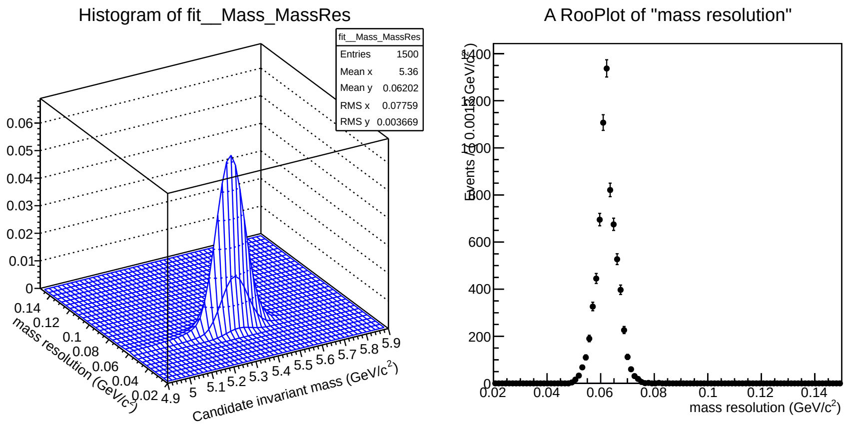

Figure 57: Mass resolution versus invariant mass for B_s^0 in the endcap (left) for 2012 data. Projection of the mass resolution (right).

The mass resolution is a conditional observable, that is, the fit must not integrate over it. The parameter k_i is a global scale factor that takes into account a possible improper error estimate.

Figures 56 and 57 show the mass resolution distributions for the B_s^0 signal in barrel and endcap for 2012 data. Fig. 58 and Fig. 59 show the fitted invariant mass distribution of B_s^0 and B_s^0 for 2012. The shapes for the 2012 are similar.

For the combinatorial background we assume a flat probability and the pdf is uniform:

$$F_{\text{comb}} = U(m) \times M^R_{\text{comb}} \tag{15}$$

The peaking background is fitted with the sum of a Gaussian and a Crystal Ball sharing the same mean. The width of this pdf will be larger than the ones of $B^0_{(s)}$, because there are many contributions: $B_s^0 \to KK$, $B^0 \to \pi\pi$, ...

The semileptonic background has a decreasing shape. The chosen pdf is an exponential multiplied by a 3^{th}-degree Chebychev polynomial, for robustness and fit stability reasons.

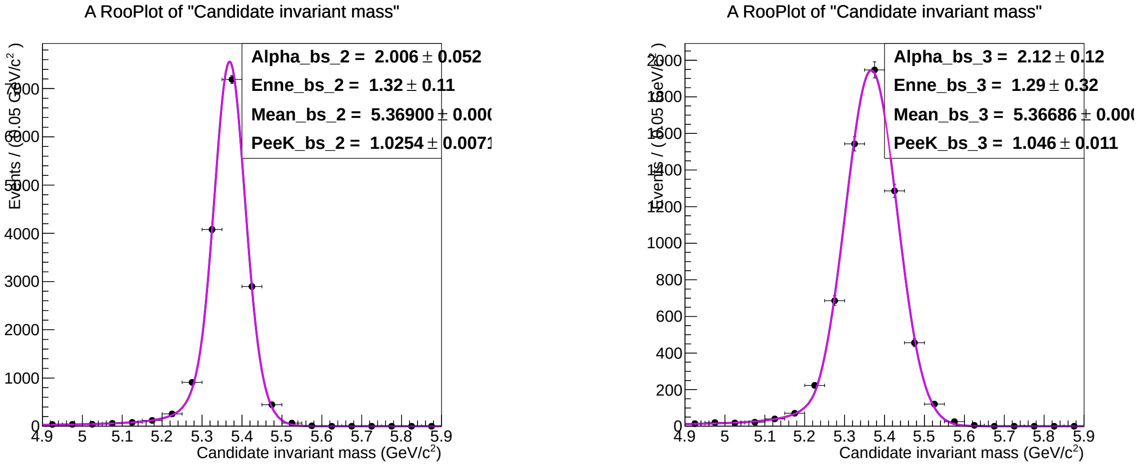

Figure 58: MC invariant mass distributions and fit results for B_s^0 barrel (*left*) and endcap (*right*), with the per-event error method for 2012 data.

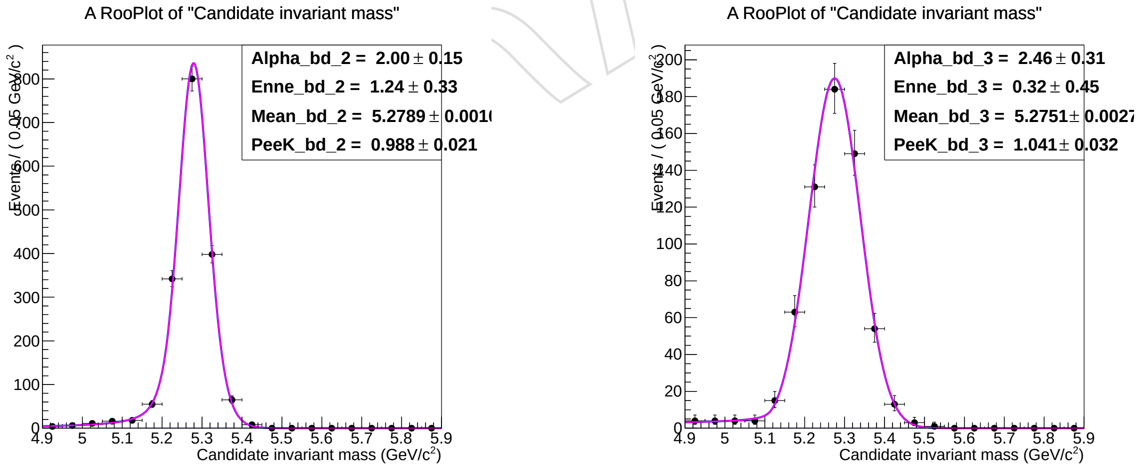

Figure 59: MC invariant mass distributions and fit results for B^0 barrel (*left*) and endcap (*right*), with the per-event error method for 2012 data.

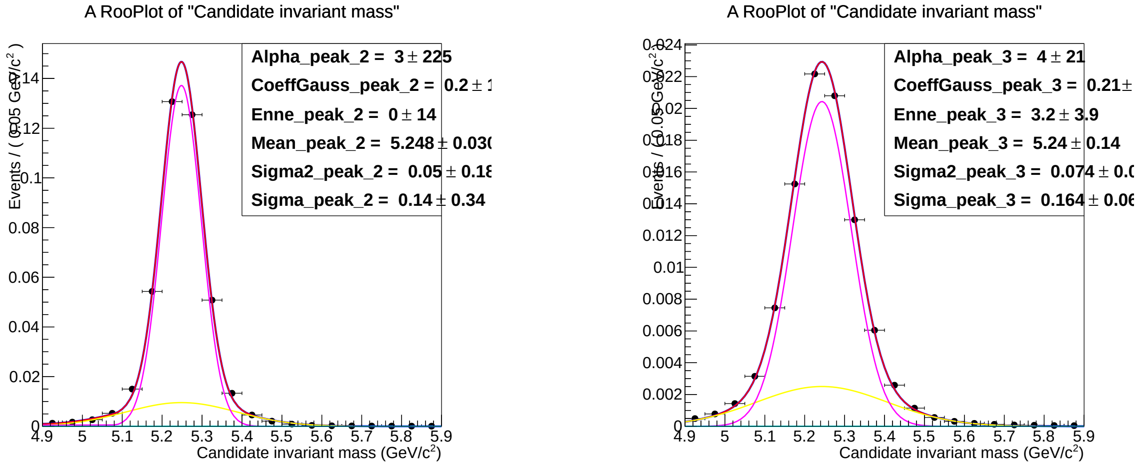

Figure 60: MC invariant mass distributions and fit results for the peaking background in barrel (left) and endcap (right)

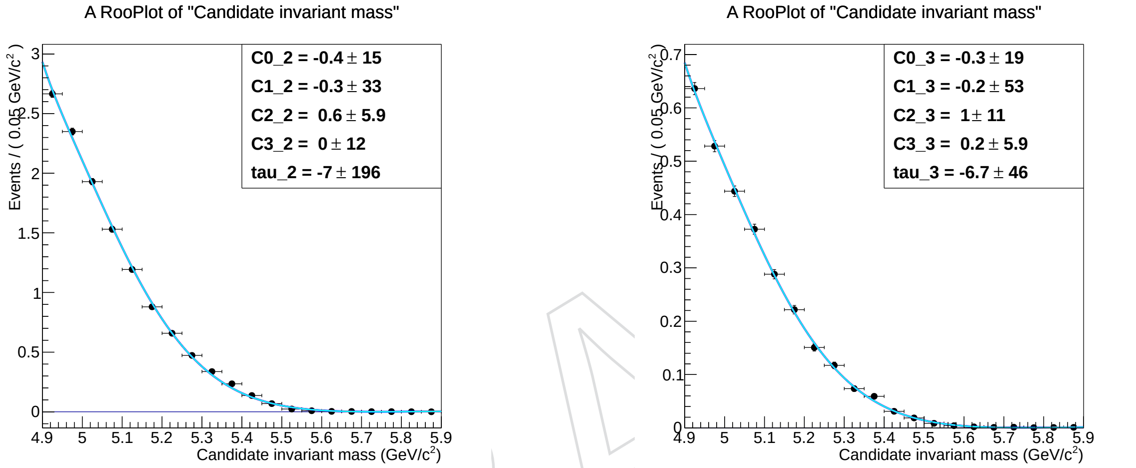

Figure 61: MC invariant mass distributions and fit results for the semileptonic background in barrel (left) and endcap (right).

Thus the rare background pdfs are

$$F_{\text{peak}} = f^{G}_{\text{peak}} G(m|\mu_{\text{peak}}, \sigma^{G}_{\text{peak}}) + (1 - f^{G}_{\text{peak}}) CB(m|\mu_{\text{peak}}, \sigma^{CB}_{\text{peak}}, \alpha_{\text{peak}}, n_{\text{peak}}) \times M^{R}_{\text{peak}} \tag{16}$$

$$F_{\text{semi}} = E(m|\tau) \times C(m|C_0, C_1, C_2, C_3) \times M^{R}_{\text{semi}} \tag{17}$$

The results of the fits are shown in Fig. 60 and in Fig. 61.

The number N_{peak} of peaking decays is constrained to the normalization decay $B^+ \to J/\psi K^+$.

10.2 The total pdf

Since there are four statistically independent channels (the two $|\eta|$ regions for the two data-taking years) the final pdf contains 16 parameters (the number of peaking background is fixed) and it is the product of four likelihoods of type Eq. 12:

$$L_{\text{tot}} = \prod_{i=0}^{3} L_i \tag{18}$$

where the index i stands for the channels shown in Table 17.

Table 17: index label of each channel

Index	Channel
0	barrel 2011
1	endcap 2011
2	barrel 2012
3	endcap 2012

With this pdf it is possible to extract one single branching fraction: in this way $N_{B^0_s,i}$ becomes a linear function of the branching fraction $BF_{B^0_s}$: $N_{B^0_s,i} = K_{B^0_s} \times BF_{B^0_s}$. The constant of proportionality is

$$K_{B^0_s,i} = (N_{B_u,i}) \times \left(\frac{f_s}{f_u}\right) \times \left(\frac{\epsilon_{B^0_s,i}}{\epsilon_{B_u,i}}\right) \times \left(\frac{1}{\mathrm{BR(Bu2JpsiK)} \cdot \mathrm{BR(Jpsi2MuMu)}}\right) \tag{19}$$

The same is true for the B^0 signal:

$$K_{B^0,i} = (N_{B_u,i}) \times \left(\frac{\epsilon_{B^0,i}}{\epsilon_{B_u,i}}\right) \times \left(\frac{1}{\mathrm{BR(Bu2JpsiK)} \cdot \mathrm{BR(Jpsi2MuMu)}}\right) \tag{20}$$

Thus, Eq. 12 becomes:

$$L_i = BF_{B^0_s} K_{B^0_s,i} F_{B^0_s} + BF_{B^0} K_{B^0,i} + N_{\mathrm{comb}} F_{\mathrm{comb}} + N_{\mathrm{semi}} F_{\mathrm{semi}} + N_{\mathrm{peak}} F_{\mathrm{peak}} \tag{21}$$

10.2.1 Constraints

The constants K_i, shown in the previous section, are products of parameters P that have systematic uncertainties σ. We add these systematic errors as Gaussian nuisance parameters to the likelihood. The Gaussian parameters are constants: the mean is equal to the estimator and the sigma is set to its uncertainty. Thus for each parameter P of Eqs. 19,20:

$$P \sim Gauss(P, P_0, \delta_P^{syst}) \tag{22}$$

The the total likelihood becomes:

$$L_{\mathrm{tot}} = \prod_{i=0}^{3} L_i L_{\mathrm{constr},i} \tag{23}$$

10.3 MC study of the final pdf

The final pdf has been studied with MC toy experiments. The assumed integrated luminosity is equivalent to 2011 + 2012 data. For each of the two η channels, 10 000 experiments have been run with a random number of events generated with the Poissonian means of the expected numbers of events for each yield contribution. Then, each experiment was fitted with the likelihood of Eq. 12.

Figure 62 shows the B^0_s yield, the statistical error on the B^0_s yield and the pull distribution, defined as $p = (N_{\mathrm{fit}} - \mu)/N_{\mathrm{err}}$ for the barrel. The pull has been fitted with a Gaussian. The pull has the mean slightly shifted to left since the left tail of the CB of the B^0_s shares the same region of the B^0 pdf and it is worse for the endcap channel, for the worse resolution. Indeed the two signals can have a negative correlation (Fig. 63). This effect is purely statistical. Generating toy experiments with equivalent integrated luminosities of 40 fb^{-1} and 80 fb^{-1}, the pull distributions of Fig. 64 are obtained. The bias in these cases is sensibly smaller.

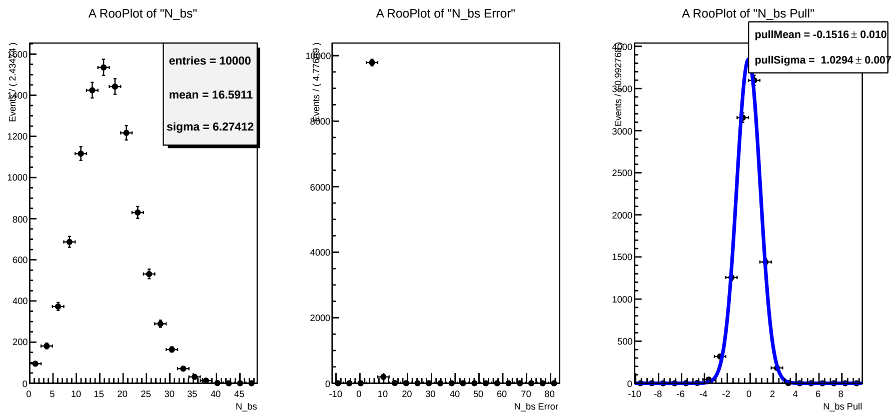

Figure 62: B_s^0 yield on the left, error on the B_s^0 yield in the middle and pull distribution on the right.

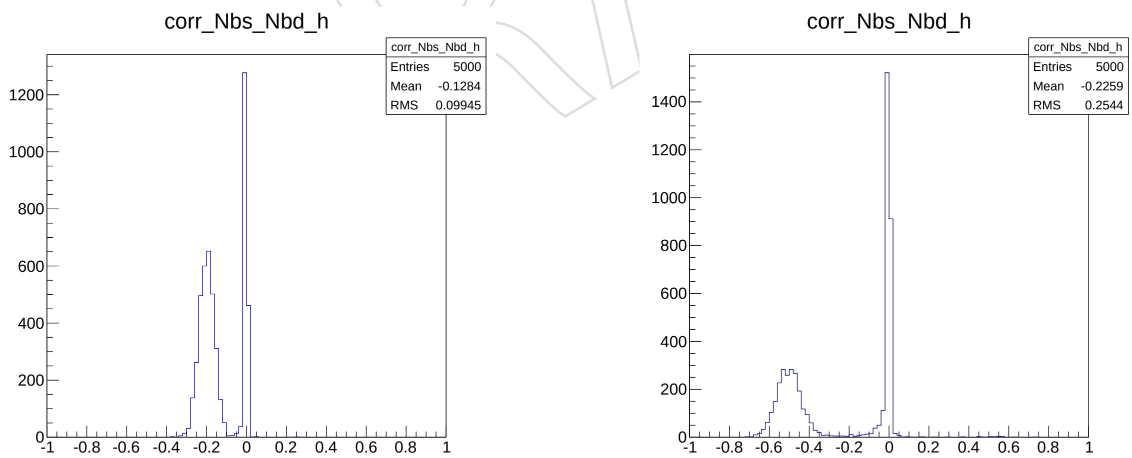

Figure 63: $B^0 - B_s^0$ correlation for the barrel (*left*) and the endcap (*right*) channels, for each toy experiment.

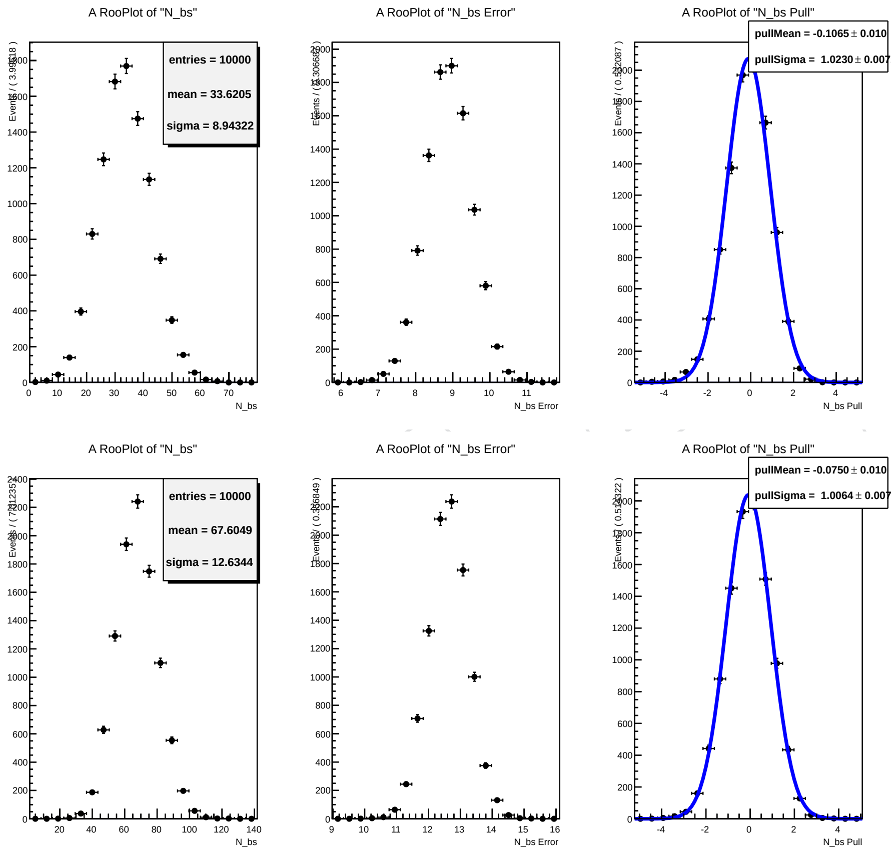

Figure 64: B_s^0 yield, statistical error on the B_s^0 yield and pull distribution for 40 fb^{-1} (top) and 80 fb^{-1} (bottom).

10.4 Systematics associated to the pdf shapes

To evaluate the systematics associated to the pdf shapes for both the B_s^0 branching fraction and B_s^0 significance we varied the shapes in the following way:

- We change the pdf signal shape to a non-per-event error shape (a Crystal Ball + Gaussian). The systematics on the B_s^0 branching fraction is 1.% and the B_s^0 significance obtained is 3.407.
- The degree of the polynomial of the semileptonic pdf is somehow arbitrary. We changed the degree to 4 and we fitted only with a simple exponential. The maximum systematics that this involves on the B_s^0 branching fraction is 3%. The minimum B_s^0 significance obtained is 3.456.
- We change the relative branching fractions of the semileptonic backgrounds and the misid efficiencies one sigma from their mean value. The maximum systematics obtained on the B_s^0 branching fraction is 2% and the minimum B_s^0 significance obtained is 3.366.
- We doubled the peaking background contribution obtaining a systematics of 0.7% on the B_s^0 branching fraction and a B_s^0 significance of 3.428.

The minimum significance obtained from these studies is 3.366.

10.5 Significance calculation

For the B_s^0 significance, the null hypothesis and the alternative hypothesis are the following:

- **Null Hypothesis**: $\mathrm{BF}_{\mathrm{B_s^0}} = 0$.
- **Alternative Hypothesis**: $\mathrm{BF}_{\mathrm{B_s^0}}$ is let floating.

In the fitting, the B^0 branching fraction can be left floating (i) or set to the SM prediction (ii). The B_s^0 significance will be evaluated in both cases.

The methods used here for the extraction of the significance are the same as the ones described in Sec. 9: We use a hybrid profile likelihood test statistics. Our priors for the nuisance parameters $N_{i,j}$ (where i is the channel of Tab. 17 and j can be $j = BF_{B_s^0}$, semi, comb) are Gaussians with mean and width equal to the fit results:

$$N_{i,j} \sim Gauss(N_{i,j}, N_{i,j,0}, \delta_{i,j}^{stat}) \tag{24}$$

500 toys have been generated evaluating the significance defined as the ratio of the two likelihood hypotheses. The B_s^0 mean significance expected is around 3.5 Gaussian sigmas (Fig. 65).

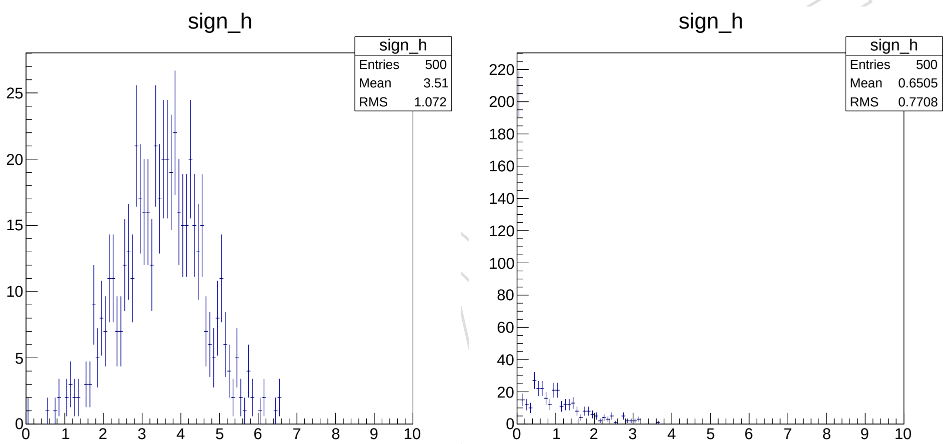

Figure 65: Statistical significance, evaluated with a likelihood ratio test, in Gaussian sigmas, for the total integrated luminosity, for the B_s^0 hypothesis (left), and the B^0 hypothesis (right)

11 Results

We require $b > 0.140$ ($b > 0.130$) for the BDT selection in the barrel (endcap).

11.1 Background determination

11.1.1 Rare peaking backgrounds

Fig. 66 shows the distribution of rare backgrounds in the barrel and endcap channels. Table 18 summarizes the relevant numbers for the analysis. The complete details can be found in Tab. 21.

The total contribution of the rare peaking backgrounds are normalized to the reconstructed B^+ yield in data using (for example in the decay $Y \to X$)

$$N(X) = \frac{\mathcal{B}(Y \to X)}{\mathcal{B}(B^+ \to \mu^+\mu^-K^+)} \frac{f_Y}{f_u} \frac{A(X)\varepsilon_{\text{ana}}(X)}{A(B^+)\varepsilon_{\text{ana}}(B^+)} N_{\text{obs}}(B^+), \tag{25}$$

where A and ε_{ana} are the acceptance and analysis efficiency, respectively. The resulting event yields are weighted with the corresponding muon misidentification rate f (or muon identification efficiency ε_μ) and the (signal) trigger efficiency $\varepsilon_{\text{trig}}$.

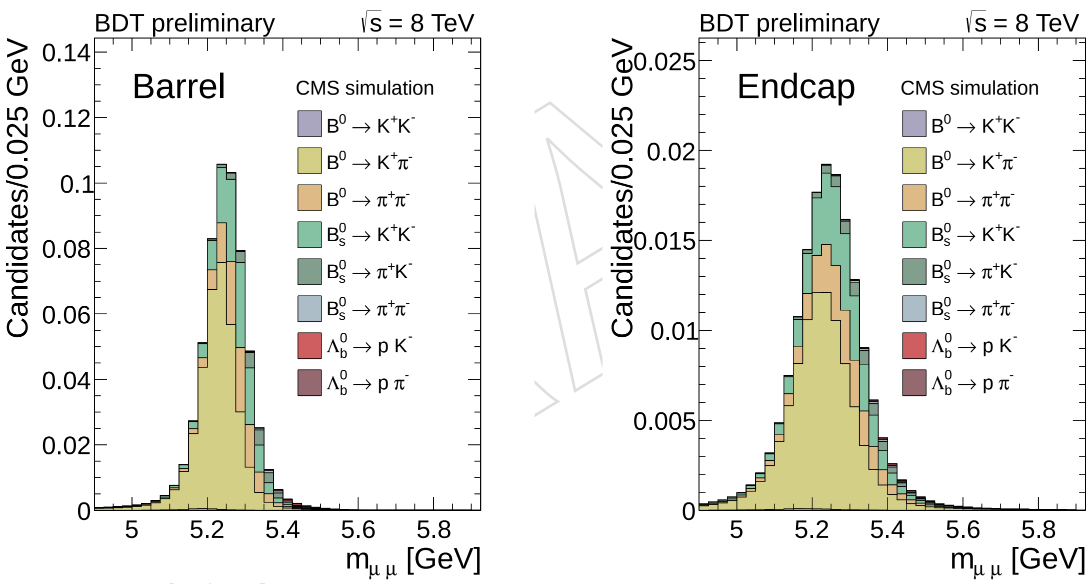

Figure 66: Invariant mass distribution of rare peaking backgrounds in the barrel (left) and the endcap channel (right).

Table 18: Expected total contribution of rare hadronic two-body B decays to the B_s^0 and B^0 signal mass window plus the low and high sidebands. The errors include the quadratic sum of the uncertainties from the production cross section, the muon misidentification, and the branching fractions. This table provides the sums of the rare hadronic decays in Tab. 21.

Channel	low sideband	B^0 window	B_s^0 window	high sideband
Barrel	0.116 ± 0.054	0.371 ± 0.141	0.099 ± 0.028	0.005 ± 0.002
Endcap	0.047 ± 0.021	0.072 ± 0.027	0.036 ± 0.011	0.004 ± 0.001

11.1.2 Combinatorial and rare semileptonic backgrounds

The non-peaking background has two components:

- Combinatorial background arising from two muons (real or misidentified hadrons) originating from separate particle decays. We estimate this from the high sideband (where only very little contribution from the second component is expected) with a constant function.
- non-peaking rare semileptonic decays. The shape of this is estimated using the MC simulation expectation using the sum of the three components shown in Fig. 67. Since this composition is missing some decays (e.g. $B^+ \to \rho\mu^+\nu$), we let the normalization float and scale the histogram in such a way that the sum of the combinatorial background and the rare semileptonic B decays account for the observed events in the low sideband.

  The systematic uncertainty of the non-peaking component is estimated by the difference when using an exponential fit to the data sidebands.

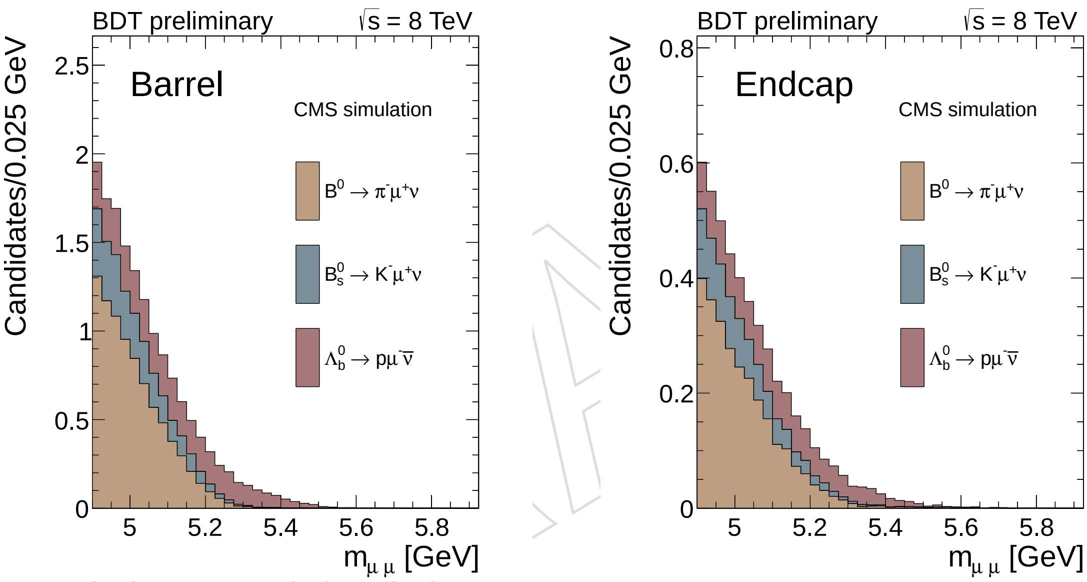

Figure 67: Invariant mass distribution of semileptonic rare B decays in the barrel (left) and the endcap channel (right).

In Fig. 68 we show the dimuon invariant mass distribution in the barrel and endcap channels, together with the fitted combinatorial and rare semileptonic backgrounds.

Table 19: Expected total contribution of rare semileptonic B decays to the B_s^0 and B^0 signal mass window plus the low and high sidebands. The errors include the quadratic sum of the uncertainties from the production cross section, the muon misidentification, and the branching fractions.

Channel	low sideband	B^0 window	B_s^0 window	high sideband
Barrel	13.475 ± 4.509	0.912 ± 0.665	0.480 ± 0.472	0.080 ± 0.068
Endcap	4.168 ± 1.386	0.321 ± 0.187	0.165 ± 0.137	0.045 ± 0.030

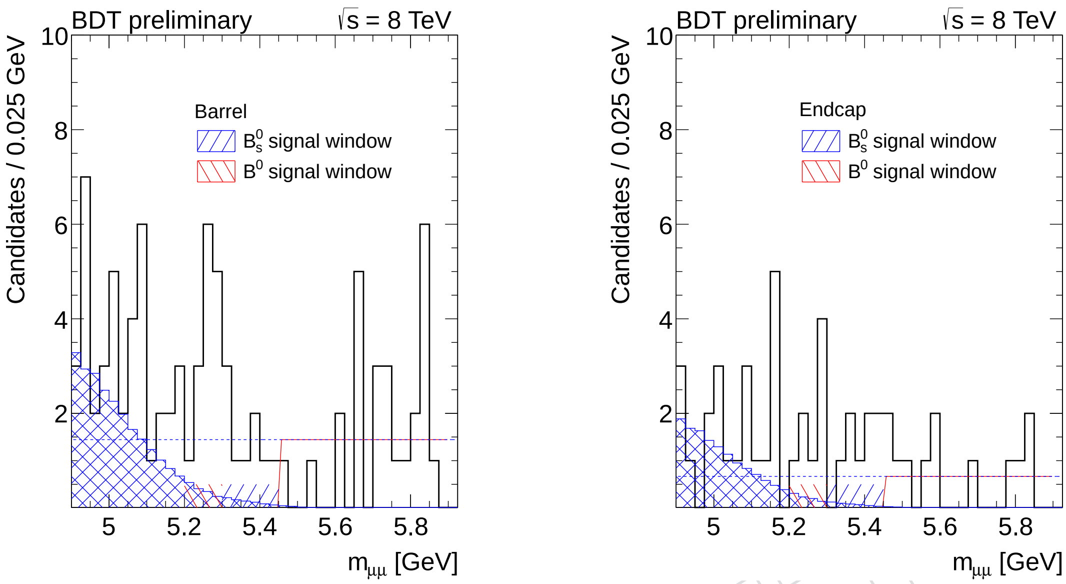

Figure 68: Dimuon invariant mass distribution with the non-peaking background components in the barrel (left) and the endcap channel (right). The combinatorial background has been fit with a constant to the high sideband, the rare semileptonic background is scaled such that the sum of rare semileptonic and combinatorial background in the low sideband sum up to the observed number of events in the low sideband.

Table 20: Combinatorial (top part) and scaled rare semileptonic (middle part) backgrounds in the B_s^0 and B^0 signal mass window plus the low and high sidebands. The bottom part summarizes the sum of the two. By construction, the lower sideband sum agrees with the observed counts in the low sideband. The errors include the quadratic sum of the uncertainties from the production cross section, the muon misidentification, and the branching fractions.

Channel	low sideband	B^0 window	B_s^0 window	high sideband
Comb. Barrel	17.333 ± 3.508	5.778 ± 1.169	8.667 ± 1.754	26.000 ± 5.262
Comb. Endcap	8.000 ± 2.344	2.667 ± 0.781	4.000 ± 1.172	12.000 ± 3.516
scaled sl. Barrel	22.667 ± 4.509	1.534 ± 0.665	0.807 ± 0.472	0.807 ± 0.068
scaled sl. Endcap	13.000 ± 1.386	0.880 ± 0.187	0.463 ± 0.137	0.463 ± 0.030
Total Barrel	40.000 ± 9.048	7.312 ± 1.581	9.474 ± 1.917	26.135 ± 5.126
Total Endcap	21.000 ± 6.219	3.546 ± 1.041	4.463 ± 1.296	12.077 ± 3.487
Observed barrel	40			26
Observed endcap	21			12

Table 21: Expected contributions of rare hadronic B decays to the B_s^0 and B^0 signal mass window plus the low and high sidebands. The errors include the quadratic sum of the uncertainties from the production cross section, the muon misidentification, and the branching fractions. (In case you wonder: we do realize that the number of digits can be critized, but this is a computer-generated table and there are more important aspects to finish in the analysis than to address this minor point.)

	Barrel			Endcap		
Channel	low sideband	B^0	B_s^0	low sideband	B^0	B_s^0
$B^0 \to K^+K^-$	0.002 ± 0.002	0.000739 ± 0.000706	0.000033 ± 0.000031	0.000 ± 0.000	0.000206 ± 0.000196	0.000041 ± 0.000040
$B^0 \to K^+\pi^-$	0.095 ± 0.054	0.229233 ± 0.129856	0.023205 ± 0.013145	0.035 ± 0.020	0.042678 ± 0.024176	0.014422 ± 0.008170
$B^0 \to \pi^+\pi^-$	0.008 ± 0.004	0.056850 ± 0.032240	0.023816 ± 0.013506	0.005 ± 0.003	0.010748 ± 0.006095	0.007509 ± 0.004258
$B^0 \to \pi^-\mu^+\nu$	8.135 ± 3.279	0.190257 ± 0.076695	0.020339 ± 0.008199	2.523 ± 1.017	0.105978 ± 0.042721	0.022788 ± 0.009186
$B_s^0 \to K^+K^-$	0.010 ± 0.006	0.076903 ± 0.045007	0.031264 ± 0.018297	0.006 ± 0.004	0.015805 ± 0.009249	0.010051 ± 0.005882
$B_s^0 \to \pi^+K^-$	0.001 ± 0.001	0.006836 ± 0.004149	0.014852 ± 0.009014	0.001 ± 0.000	0.002041 ± 0.001239	0.003194 ± 0.001939
$B_s^0 \to \pi^+\pi^-$	0.000 ± 0.000	0.000286 ± 0.000171	0.002718 ± 0.001622	0.000 ± 0.000	0.000151 ± 0.000090	0.000543 ± 0.000324
$B_s^0 \to K^-\mu^+\nu$	2.570 ± 1.036	0.095147 ± 0.038355	0.011893 ± 0.004794	0.806 ± 0.325	0.043026 ± 0.017344	0.012201 ± 0.004919
$\Lambda_b^0 \to pK^-$	0.000 ± 0.000	0.000203 ± 0.000225	0.001925 ± 0.002129	0.000 ± 0.000	0.000088 ± 0.000098	0.000413 ± 0.000456
$\Lambda_b^0 \to p\pi^-$	0.000 ± 0.000	0.000083 ± 0.000093	0.000703 ± 0.000784	0.000 ± 0.000	0.000038 ± 0.000042	0.000181 ± 0.000202
$\Lambda_b^0 \to p\mu^-\bar{\nu}$	2.770 ± 2.916	0.626572 ± 0.659808	0.447672 ± 0.471418	0.839 ± 0.883	0.171960 ± 0.181082	0.129859 ± 0.136747
Total	13.591 ± 4.510	1.283 ± 0.680	0.578 ± 0.472	4.215 ± 1.386	0.393 ± 0.189	0.201 ± 0.138

11.2 Measurement of $\mathcal{B}(B_s^0 \to J/\psi\,\phi)/\mathcal{B}(B^\pm \to J/\psi\,K^\pm)$

11.2.1 Integrated over $p_\perp(B)$

In Fig. 69 we show the invariant mass distribution of the normalization channel in the barrel and endcap channels in the 2011 dataset of fb^{-1}. We fit a double (single) Gaussian for the signal in the barrel (endcap). The background (combinatorial and partially reconstructed B decays) is modeled by the sum of an error function and a exponential function for both barrel and endcap. The decays $B^\pm \to J/\psi\,\pi^\pm$ are modeled with a Landau function, where the width and peak position are fixed to the expectation from the MC simulation. The normalization is constrained relative to the $B^\pm \to J/\psi\,K^\pm$ yield to be consistent with the expectation [28].

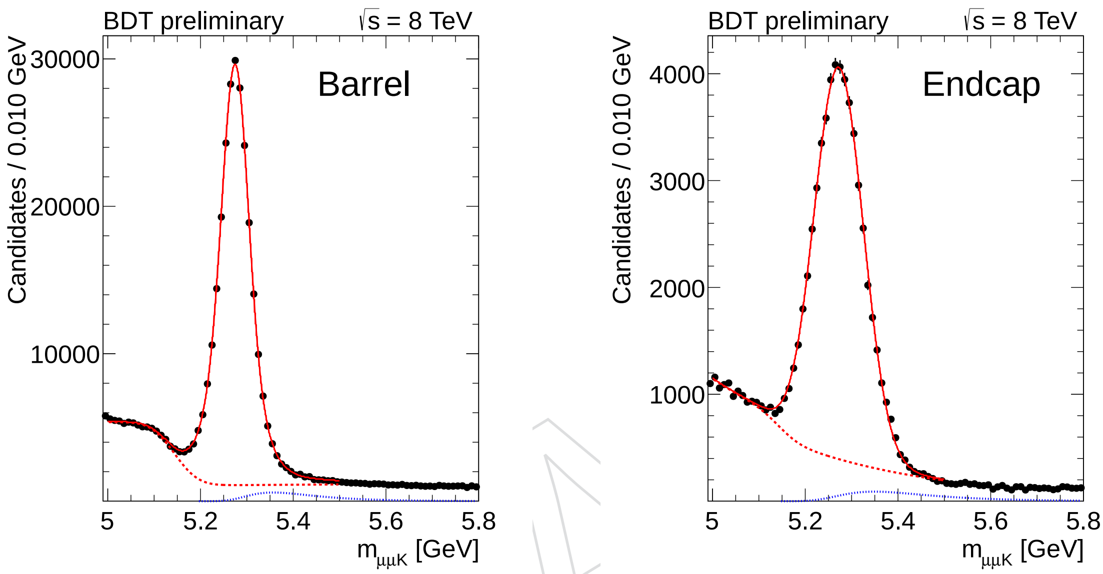

Figure 69: $B^\pm \to J/\psi\,K^\pm$ invariant mass distribution in the barrel (left) and the endcap channel (right). The red dashed line indicates the function used to describe the combinatorial background and the contribution from partially reconstructed B decays. The blue dotted line indicates the component from $B^\pm \to J/\psi\,\pi^\pm$.

In Fig. 70 the invariant mass distribution of the control sample in the barrel and endcap channels.We fit a double (single) Gaussian for the signal in the barrel (endcap). The background is modeled by and exponential function plus a Gaussian component for $B_d^0 \to J/\psi\,K^{*0}$, where the shape is taken from a fit to $B_d^0 \to J/\psi\,K^{*0}$ MC and the normalization is floated in the fit.

We use Eq. 1 to determine the results in the barrel and endcap channels

$$\begin{aligned}
\mathcal{B}(B_s^0 \to J/\psi(\mu^+\mu^-)\phi(KK)) &= 0.000022 \pm 0.000001 \pm 0.000002 \qquad \text{(barrel)}\\
\mathcal{B}(B_s^0 \to J/\psi(\mu^+\mu^-)\phi(KK)) &= 0.000028 \pm 0.000001 \pm 0.000002 \qquad \text{(endcap)}
\end{aligned}$$

The first error is statistical only, obtained from the fitted yields for $B^\pm \to J/\psi\,K^\pm$ and $B_s^0 \to J/\psi\,\phi$. The second error includes the systematic error on the fitted yields.

11.2.2 Differential in $p_\perp(B)$

To test a possible dependence of f_s/f_u on the $p_\perp$ of the B meson, we perform a measurement of the ratio of $B^\pm \to J/\psi\,K^\pm$ and $B_s^0 \to J/\psi\,\phi$ differentially in B $p_\perp$. Four bins in B $p_\perp$ are considered: < 16 GeV, $16 - 20$ GeV, $20 - 25$ GeV and > 25 GeV. In each bin, we fit for the yield

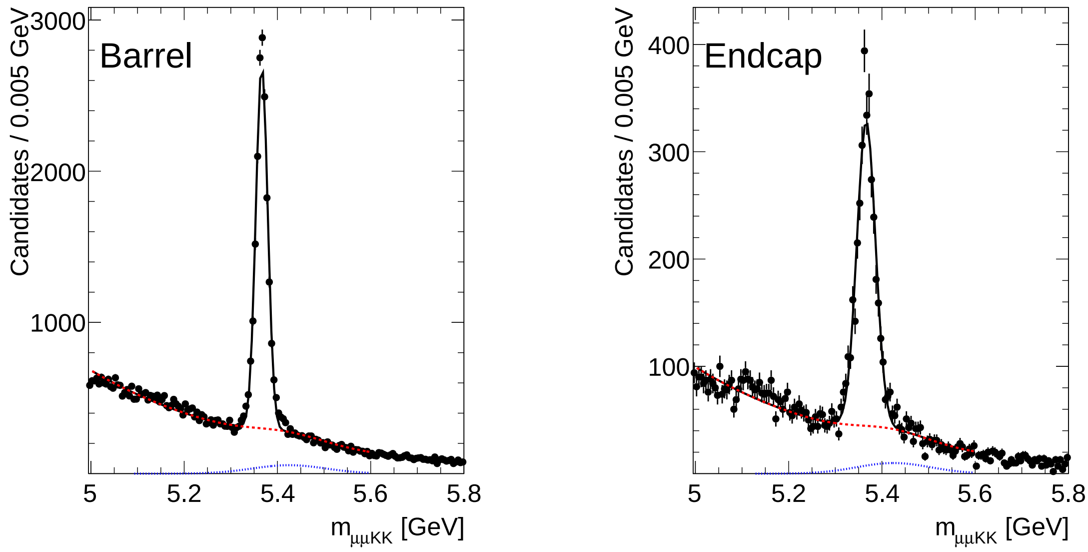

Figure 70: $B_s^0 \to J/\psi\phi$ invariant mass distribution in the barrel (left) and the endcap channel (right). The red dashed line indicates the function used to describe the combinatorial background and the contribution from partially reconstructed B decays. The blue dotted line indicates the component from $B_d^0 \to J/\psi K^{*0}$.

Table 22: BDT selection efficiency and number of observed events for the control sample and the normalization sample. The errors are statistical only.

Variable	$B_s^0 \to J/\psi\phi$ Barrel	$B^\pm \to J/\psi K^\pm$ Barrel	$B_s^0 \to J/\psi\phi$ Endcap	$B^\pm \to J/\psi K^\pm$ Endcap
Acceptance	0.1021 ± 0.0003	0.157 ± 0.000	0.0702 ± 0.0003	0.106 ± 0.000
$\varepsilon_{\text{analysis}}$	0.0181 ± 0.0005	0.0187 ± 0.0003	0.0078 ± 0.0006	0.0093 ± 0.0003
ε_μ^{MC}	0.7365 ± 0.0012	0.735 ± 0.001	0.7353 ± 0.0022	0.738 ± 0.001
ε_μ^{MC-TNP}	0.7789 ± 0.0003	0.775 ± 0.000	0.8345 ± 0.0003	0.836 ± 0.000
ε_μ^{TNP}	0.7888 ± 0.0002	0.787 ± 0.000	0.7797 ± 0.0003	0.781 ± 0.000
$\varepsilon_{\text{trig}}^{MC}$	0.5390 ± 0.0015	0.532 ± 0.001	0.3820 ± 0.0029	0.375 ± 0.001
$\varepsilon_{\text{trig}}^{MC-TNP}$	0.8351 ± 0.0003	0.831 ± 0.000	0.7294 ± 0.0013	0.719 ± 0.001
$\varepsilon_{\text{trig}}^{TNP}$	0.7891 ± 0.0003	0.786 ± 0.000	0.7369 ± 0.0013	0.728 ± 0.001
N_{obs}	15456 ± 531	241967 ± 667	3168 ± 70	46855 ± 241

of normalization and control sample events and correct the yields by the efficiencies found in each bin. The fits are as described for the integrated ratio results above. The PDG values for the normalization and control sample branching fractions are used to compute a value for f_s/f_u for each bin. The results, shown in Table 23, are found to be consistent with flat across the $p_\perp$ bins for both barrel and endcap. No additional systematic is assigned for this effect.

Figure 71 illustrates the numbers and fits a constant and a straight line to the measurements. There is no significant effect observable

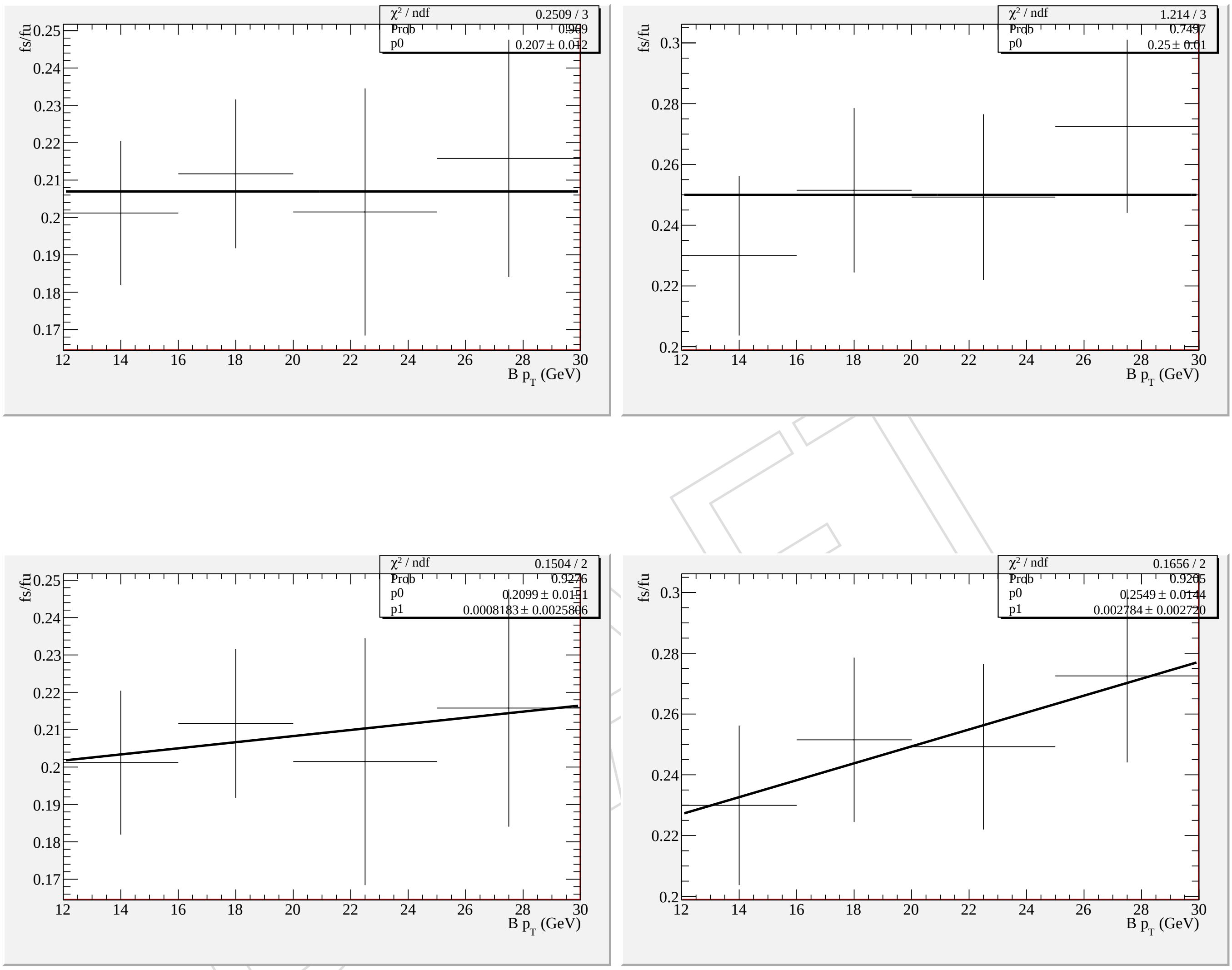

Figure 71: Display of the numbers in Tab. 23 together with fits of a constant function (top row) and a straight line (bottom row) to the measurements. The barrel is in the left column, the endcap in the right endcap.

Table 23: The yields and efficiencies are shown for the normalization and control samples in bins in $Bp_\perp$. The PDG branching fractions are used to compute f_s/f_u in each bin. The uncertainties shown are statistical and systematic, respectively. The same fractional systematic uncertainties determined for the integrated fits are applied to the binned fits. No attempt to distinguish between effects correlated and uncorrelated between the bins has been made. The 23% uncertainty on the control sample branching fraction is also not shown. Note that the results of this study are shown for a slightly out of date version of the BDT, and thus the results differ from those shown in Table 22.

Barrel					
$Bp_\perp$	Integrated	$< 16\,\mathrm{GeV}$	$16-20\,\mathrm{GeV}$	$20-25\,\mathrm{GeV}$	$> 25\,\mathrm{GeV}$
$B^\pm$ yield	$192562 \pm 468 \pm 9600$	$54728 \pm 259 \pm 2700$	$44752 \pm 224 \pm 2200$	$39563 \pm 212 \pm 2000$	$54801 \pm 260 \pm 2700$
B_s^0 yield	$13336 \pm 146 \pm 670$	$2217 \pm 62 \pm 110$	$3015 \pm 68 \pm 150$	$3028 \pm 68 \pm 150$	$5118 \pm 86 \pm 250$
$B^\pm$ eff. ($\times 10^{-6}$)	$753 \pm 2 \pm 30$	$219 \pm 1 \pm 9$	$184 \pm 1 \pm 7$	$154 \pm 1 \pm 6$	$196 \pm 1 \pm 8$
B_s^0 eff. ($\times 10^{-6}$)	$469 \pm 2 \pm 23$	$84 \pm 1 \pm 3$	$112 \pm 1 \pm 4$	$112 \pm 1 \pm 4$	$162 \pm 1 \pm 6$
f_s/f_u	$0.212 \pm 0.004 \pm 0.019$	$0.201 \pm 0.005 \pm 0.018$	$0.212 \pm 0.006 \pm 0.019$	$0.201 \pm 0.006 \pm 0.018$	$0.216 \pm 0.005 \pm 0.019$
Endcap					
$Bp_\perp$	Integrated	$< 16\,\mathrm{GeV}$	$16-20\,\mathrm{GeV}$	$20-25\,\mathrm{GeV}$	$> 25\,\mathrm{GeV}$
$B^\pm$ yield	$33968 \pm 198 \pm 1700$	$12643 \pm 121 \pm 630$	$8334 \pm 97 \pm 420$	$6208 \pm 85 \pm 310$	$6969 \pm 89 \pm 350$
B_s^0 yield	$2045 \pm 53 \pm 102$	$439 \pm 25 \pm 22$	$547 \pm 26 \pm 27$	$479 \pm 24 \pm 24$	$626 \pm 27 \pm 32$
$B^\pm$ eff. ($\times 10^{-6}$)	$305 \pm 2 \pm 12$	$116 \pm 1 \pm 4$	$78 \pm 1 \pm 3$	$56 \pm 1 \pm 2$	$55 \pm 1 \pm 2$
B_s^0 eff. ($\times 10^{-6}$)	$128 \pm 3 \pm 5$	$32 \pm 1 \pm 1$	$35 \pm 1 \pm 1$	$31 \pm 1 \pm 1$	$31 \pm 1 \pm 1$
f_s/f_u	$0.245 \pm 0.008 \pm 0.023$	$0.220 \pm 0.013 \pm 0.020$	$0.251 \pm 0.013 \pm 0.023$	$0.251 \pm 0.014 \pm 0.023$	$0.265 \pm 0.014 \pm 0.024$

11.3 Upper limit expectation with the BDT selection

In Fig. 72 we show the (unblinded) dimuon invariant mass distribution in the barrel and endcap channels.

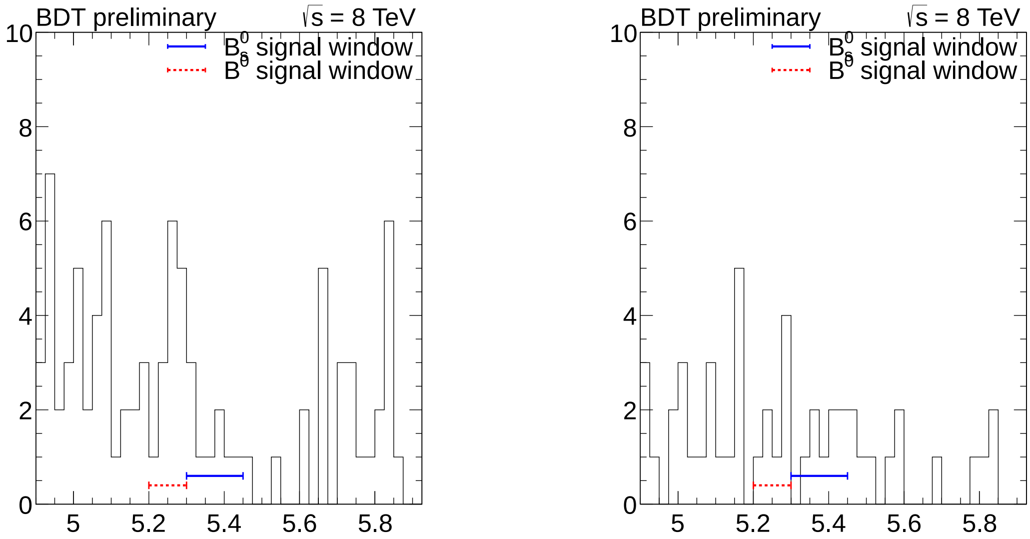

Figure 72: Dimuon invariant mass distribution in the barrel (left) and the endcap channel (right).

In Tab. 24 we summarize all numbers relevant for the extraction of the upper limit.

Tab. 25 summarizes the equivalent numbers for the normalization sample, selected by the BDT.

Expected limits are computed as described in section 9 using the CL$_s$ technique. Expected results for $B_s^0 \to \mu^+\mu^-$:

$$6.7^{+2.5}_{-2.1} \times 10^{-9} \qquad \text{SM}$$
$$3.4^{+1.6}_{-1.1} \times 10^{-9} \qquad \text{Background only}$$

Expected results for $B^0 \to \mu^+\mu^-$:

$$1.0^{+0.5}_{-0.3} \times 10^{-9} \qquad \text{SM}$$
$$8.1^{+4.0}_{-2.6} \times 10^{-10} \qquad \text{Background only}$$

The expected two sided 1σ interval from the binned analysis is

$$3.2^{+1.4}_{-1.7} \times 10^{-9} \qquad \text{SM } (B_s^0 \to \mu^+\mu^-)$$
$$1.0^{+3.8}_{-1.0} \times 10^{-10} \qquad \text{SM } (B^0 \to \mu^+\mu^-)$$

The expected p value for $B_s^0 \to \mu^+\mu^-$ ($B^0 \to \mu^+\mu^-$) is $2.6^{+14.6}_{-2.4} \times 10^{-2}$ ($3.4^{+5.7}_{-2.6} \times 10^{-1}$) corresponding to $1.9^{+1.0}_{-1.0}$ ($0.4^{+1.0}_{-1.8}$) sigmas. Probability to observe 3σ is 14% (1%).

11.4 Results with binned analysis

The observation is evaluated using the binned techniques to give the following upper limits

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-) \leq 2.8 \times 10^{-9} \qquad \text{at 95\% C.L.}$$
$$\mathcal{B}(B^0 \to \mu^+\mu^-) \leq 2.0 \times 10^{-9} \qquad \text{at 95\% C.L.}$$

Table 24: Efficiency and background expectations and observations. The errors are the combined statistical and systematic errors. $N^{\text{exp}}_{\text{signal}}$ is the expected signal yield in the respective signal windows. $N^{\text{exp}}_{\text{cross-feed}}$ describes the cross feed of B^0_s into the B^0 signal window and vice versa. $N^{\text{exp}}_{\text{non-peak. bg}}$ combines rare semileptonic and combinatorial backgrounds. $N^{\text{exp}}_{\text{peak.bg}}$ summarizes the hadronic two-body peaking backgrounds. $N^{\text{exp}}_{\text{all bg}}$ sums up all backgrounds. $N^{\text{exp}}_{\text{total}}$ is the total of background and signal (plus cross feed). $N^{\text{obs}}_{\text{sidebands}}$ is the observed event count in the sideband, excluding the signal regions. The number of events N_{obs} is observed in the respective signal boxes.

Variable	$B^0 \to \mu^+\mu^-$ Barrel	$B^0_s \to \mu^+\mu^-$ Barrel	$B^0 \to \mu^+\mu^-$ Endcap	$B^0_s \to \mu^+\mu^-$Endcap
Acceptance	0.237 ± 0.008	0.237 ± 0.008	0.218 ± 0.011	0.218 ± 0.011
$\varepsilon_{\text{analysis}}$	0.033 ± 0.001	0.032 ± 0.001	0.019 ± 0.001	0.019 ± 0.001
ε_μ^{MC}	0.690 ± 0.029	0.679 ± 0.027	0.813 ± 0.066	0.826 ± 0.066
ε_μ^{MC-TNP}	0.784 ± 0.031	0.785 ± 0.031	0.835 ± 0.067	0.835 ± 0.067
ε_μ^{TNP}	0.790 ± 0.032	0.792 ± 0.032	0.776 ± 0.062	0.779 ± 0.062
$\varepsilon_{\text{trig}}^{MC}$	0.619 ± 0.021	0.620 ± 0.019	0.432 ± 0.029	0.447 ± 0.027
$\varepsilon_{\text{trig}}^{MC-TNP}$	0.840 ± 0.025	0.841 ± 0.025	0.748 ± 0.045	0.750 ± 0.045
$\varepsilon_{\text{trig}}^{TNP}$	0.793 ± 0.024	0.794 ± 0.024	0.758 ± 0.046	0.759 ± 0.046
ε_{tot}	0.0033 ± 0.0002	0.0031 ± 0.0002	0.0014 ± 0.0002	0.0015 ± 0.0002
$N^{\text{exp}}_{\text{signal}}$	0.955 ± 0.096	9.851 ± 1.478	0.260 ± 0.026	3.314 ± 0.497
$N^{\text{exp}}_{\text{cross-feed}}$	0.838 ± 0.126	0.384 ± 0.038	0.653 ± 0.098	0.172 ± 0.017
$N^{\text{exp}}_{\text{non-peak. bg}}$	7.312 ± 1.581	9.474 ± 1.917	3.546 ± 1.041	4.463 ± 1.296
$N^{\text{exp}}_{\text{peak.bg}}$	0.371 ± 0.141	0.099 ± 0.028	0.072 ± 0.027	0.036 ± 0.011
$N^{\text{exp}}_{\text{all bg}}$	7.683 ± 1.587	9.572 ± 1.917	3.618 ± 1.041	4.499 ± 1.296
$N^{\text{exp}}_{\text{total}}$	9.476 ± 1.868	19.808 ± 2.421	4.531 ± 1.163	7.985 ± 1.388
$N^{\text{obs}}_{\text{sidebands}}$	66		33	
N_{obs}	15	9	8	8

Table 25: Selection efficiency and number of observed events for the normalization sample. The errors are the combined statistical and systematic errors.

Variable	$B^\pm \to J/\psi K^\pm$ Barrel	$B^\pm \to J/\psi K^\pm$ Endcap
Acceptance	0.157 ± 0.005	0.106 ± 0.005
$\varepsilon_{\text{analysis}}$	0.0187 ± 0.0011	0.0093 ± 0.0006
ε_μ^{MC}	0.735 ± 0.029	0.738 ± 0.059
ε_μ^{MC-TNP}	0.775 ± 0.031	0.836 ± 0.067
ε_μ^{TNP}	0.787 ± 0.031	0.781 ± 0.062
$\varepsilon_{\text{trig}}^{MC}$	0.532 ± 0.016	0.375 ± 0.023
$\varepsilon_{\text{trig}}^{MC-TNP}$	0.831 ± 0.000	0.719 ± 0.001
$\varepsilon_{\text{trig}}^{TNP}$	0.786 ± 0.024	0.728 ± 0.044
ε_{tot}	0.00094 ± 0.00008	0.00022 ± 0.00003
N_{obs}	241967 ± 12116	46855 ± 2355

The p-value of the $B_s^0 \to \mu^+\mu^-$ ($B^0 \to \mu^+\mu^-$) signal assuming unknown $B^0 \to \mu^+\mu^-$ ($B_s^0 \to \mu^+\mu^-$) signal is 5.5×10^{-1} (5.1×10^{-3}) corresponding to -0.1 (2.6) sigmas.

If the observed excess is assumed from a signal, the branching fraction can be measured to be (1σ uncertainty)

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (1.7^{+795349.7}_{-1.7}) \times 10^{-15}$$
$$\mathcal{B}(B^0 \to \mu^+\mu^-) = (9.7^{+5.0}_{-4.5}) \times 10^{-10}$$

The corresponding confidence level plots are shown in figure 73 ($B_s^0 \to \mu^+\mu^-$) and figure 74.

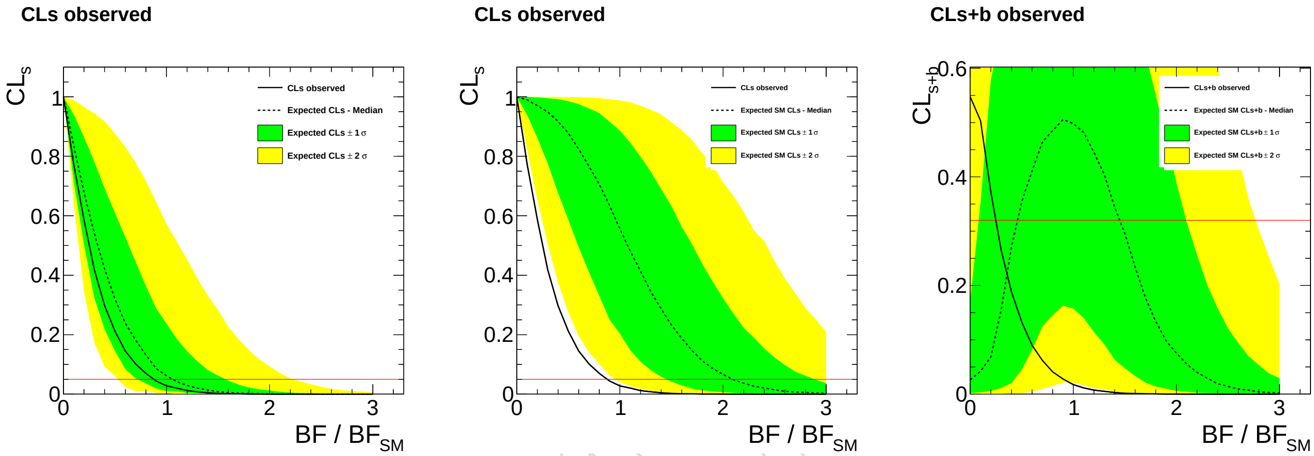

Figure 73: $B_s^0 \to \mu^+\mu^-$ confidence level plots. Left two plots show the observed CL_s curve with expected curve from background only model (left) and SM (middle). Two sided interval confidence level shown on the right

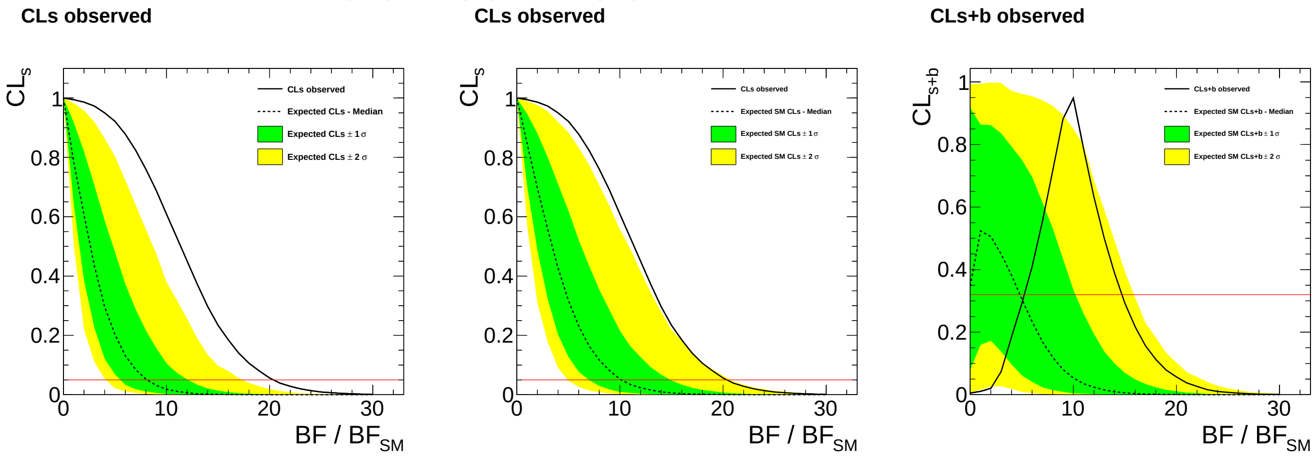

Figure 74: $B^0 \to \mu^+\mu^-$ confidence level plots. Left two plots show the observed CL_s curve with expected curve from background only model (left) and SM (middle). Two sided interval confidence level shown on the right

A comparison between the expected results and the fitted results is given in table 26

Table 26: Expected yields compared to fitted yields in the binned analysis

Variable	$B^0 \to \mu^+\mu^-$ Barrel	$B_s^0 \to \mu^+\mu^-$ Barrel	$B^0 \to \mu^+\mu^-$ Endcap	$B_s^0 \to \mu^+\mu^-$ Endcap
N^{exp}_{signal}	0.955 ± 0.096	9.851 ± 1.478	0.260 ± 0.026	3.314 ± 0.497
$N^{exp}_{non-peak.\ bg}$	7.312 ± 1.581	9.474 ± 1.917	3.546 ± 1.041	4.463 ± 1.296
$N^{exp}_{peak.bg}$	0.371 ± 0.141	0.099 ± 0.028	0.072 ± 0.027	0.036 ± 0.011
N^{fit}_{signal}	7.928	0.000	2.219	0.000
$N^{fit}_{comb.bg}$	7.010	8.365	4.069	5.115
$N^{fit}_{peak.bg}$	0.371	0.098	0.072	0.036

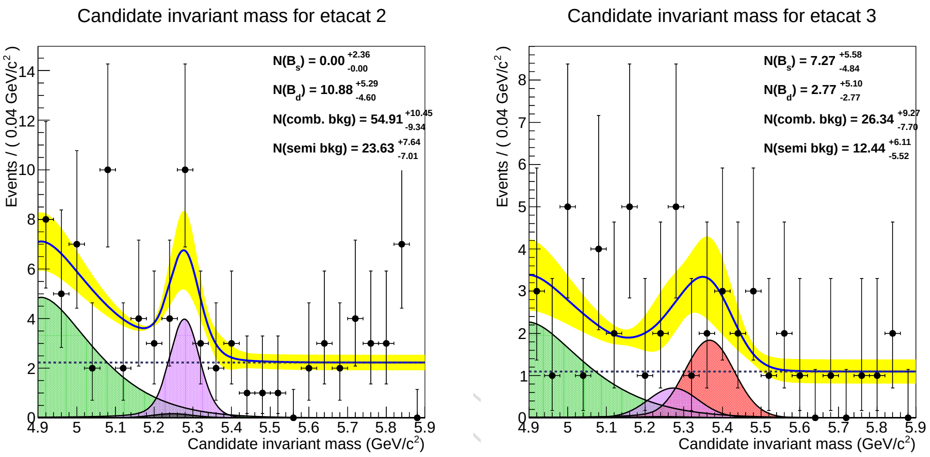

Figure 75: UML results for the barrel (left) and the endcap (right) categories.

11.5 Results with the UML

The UML fit has been performed on the unblinded 2012 data. Results of the fitting are shown in Fig. 75. Putting a common branching fraction between the two categories (barrel and endcap) we obtain instead results shown in Fig. 76. In this case, the fitted branching fractions are:

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (0.0488^{+9.97}_{-0.0488}) \times 10^{-9}$$
$$\mathcal{B}(B^0 \to \mu^+\mu^-) = (1.95^{+0.773}_{-0.773}) \times 10^{-9}$$

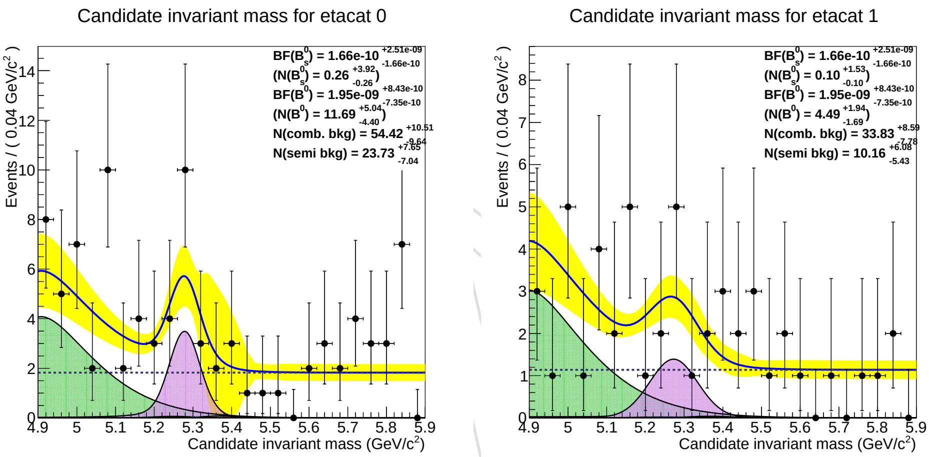

Figure 76: UML results for the barrel (left) and the endcap (right) categories.

12 Studies after unblinding

12.1 Hadronic two-body B decays

There is an irreducible background underneath the B_s^0 and even more so the B^0 signal windows from hadronic two-body B decays. These decays are not rejected by the BDT, as they have the same final state topology as the signal. The only handle against them is that their final state particles started out as hadrons. In Fig. 77 (78) we illustrate the data in the barrel (endcap) with the arbitrarily scaled mass shapes of the signal and the three dominant backgrounds: $B^0 \to K^+\pi^-$, $B^0 \to \pi^+\pi^-$, and $B_s^0 \to K^+K^-$ (see Tab. 21). The mass shifts due to the wrong mass hypotheses are visible, especially for $B^0 \to K\pi$.

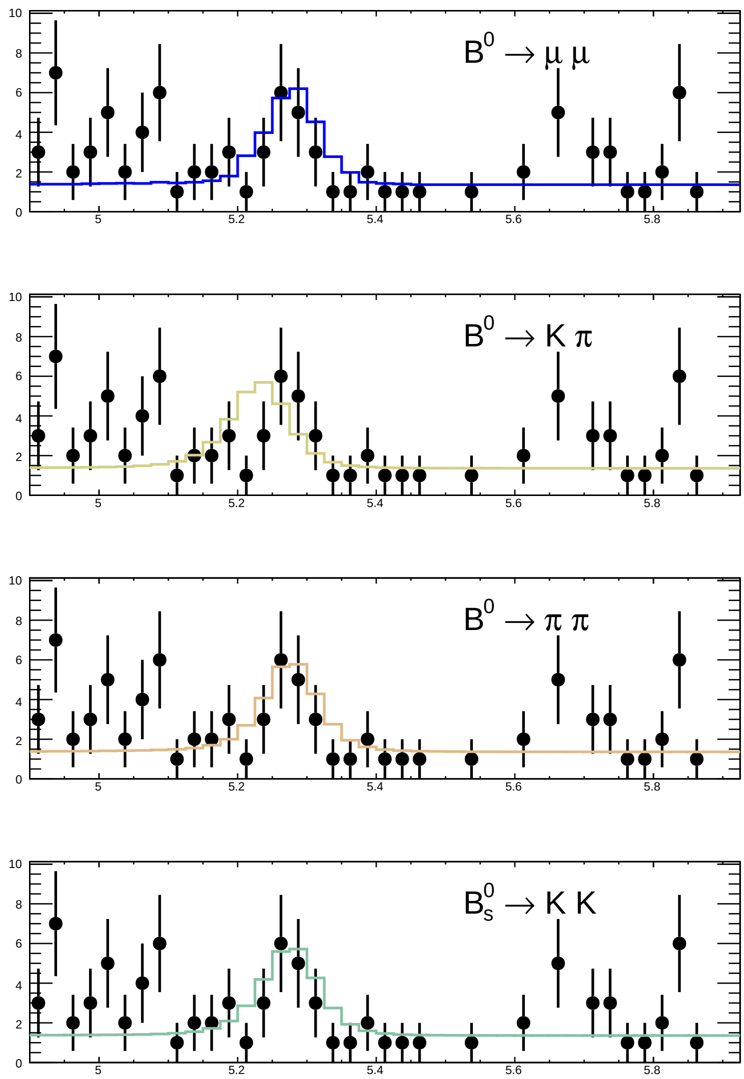

Figure 77: Dimuon mass distributions in the barrel for data (black solid markers) and various rare decays. The peaking MC shapes have been scaled by an arbitrary amount, the horizontal component correponds to a fit to the data with $m > 5.4\,\text{GeV}$.

In Tables 24 and 21 the details for the (peaking) background estimation are provided. They are based on the measured branching fractions given in Tab. 3 and the hadron fake rates discussed in section 7.3.

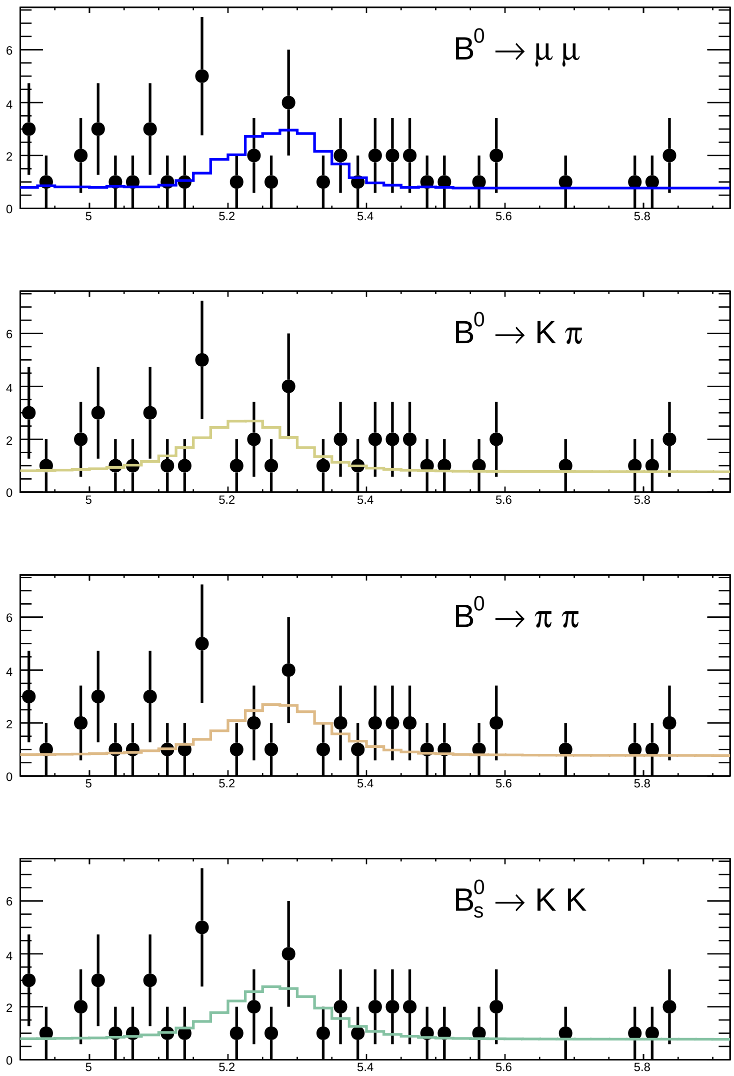

Figure 78: Dimuon mass distributions in the endcap for data (black solid markers) and various rare decays. The peaking MC shapes have been scaled by an arbitrary amount, the horizontal component correponds to a fit to the data with $m > 5.4\,\mathrm{GeV}$.

It is an interesting exercise to see how much off the hadron fake rate must be to account for the peak in the B^0 window. Attributing the entire excess in the barrel B^0 window to an underestimated hadron fake rate would require a misidentification rate of roughly 3.6×10^{-3} (instead of the 1.0×10^{-3} we are currently using and which is already somewhat higher than the Muon POG estimate of 6×10^{-4}; this difference is covered by the systematic uncertainty we assume for the fake rate). The deficit in the barrel B_s^0 window increases slightly. The background-only expectation (without any $B_s^0 \to \mu^+\mu^-$ contriution) would be one count above the observation (not a problem, of course).

Applying such a fake rate also to the rare semileptonic B decays implies that they have to be scaled down by a factor 2.4 in the barrel to not saturate the event count when combined with the (flat) combinatorial background.

To attribute the excess in the endcap onto the hadron fake rate would imply an 7-fold increase of the hadron fake rate. This could be accomodated by the observations.

Note: we will add soon the measured fake kaons in the $B^\pm \to J/\psi K^\pm$ sample to the AN.

12.2 Mass scale

As a first cross check we have fitted the mass distributions of J/ψ and $B^\pm \to J/\psi K^\pm$ and compare the fitted peak positions in low-pileup ($N_{PV} < 6$ and high-pileup ($N_{PV} > 10$) runs, as shown in Fig. 79.

From the spread we conclude that the effect of pileup on the mass scale is of the order 10^{-4}. This is much less than would be required to shift a $B_s^0 \to \mu^+\mu^-$ signal into the B^0 signal window.

Note: We will add soon the plots where we fit the J/ψ and the $\Upsilon(1S)$ peaks as a function of primary vertex multiplicity. This allows a proper interpolation of dimuon mass fits.

12.3 Changing Tracker Alignment

The events that passed the bdt cut at 0.13 have been re-analized with a different alignment from the Tracker DPG, tag to

```
GR_R_53_V16::All
```

and re-reconstruction with

```
CMSSW_5_3_4_patch2
```

. The new mass distributions are shown below(red new, blue old). The shift is present, a few events move from the B0 to the Bs mass region. The event by event mass shifts in the run periods (2012A, 2012B) and 2012C are plotted in Figs. 80 and 81. Overall there is no indication of a global shift, even if there are few events with large shifts.

12.4 B candidate variables in B^0 mass window

Figure 82 shows a comparison of several variables between events selected in the B^0 mass window and those outside the mass window. All events plotted pass the BDT selection for the barrel region. These variables were chosen based on their potential sensitivity to differences between prompt muons and those from decay-in-flight hadrons. No clear differences in the distributions are observed.

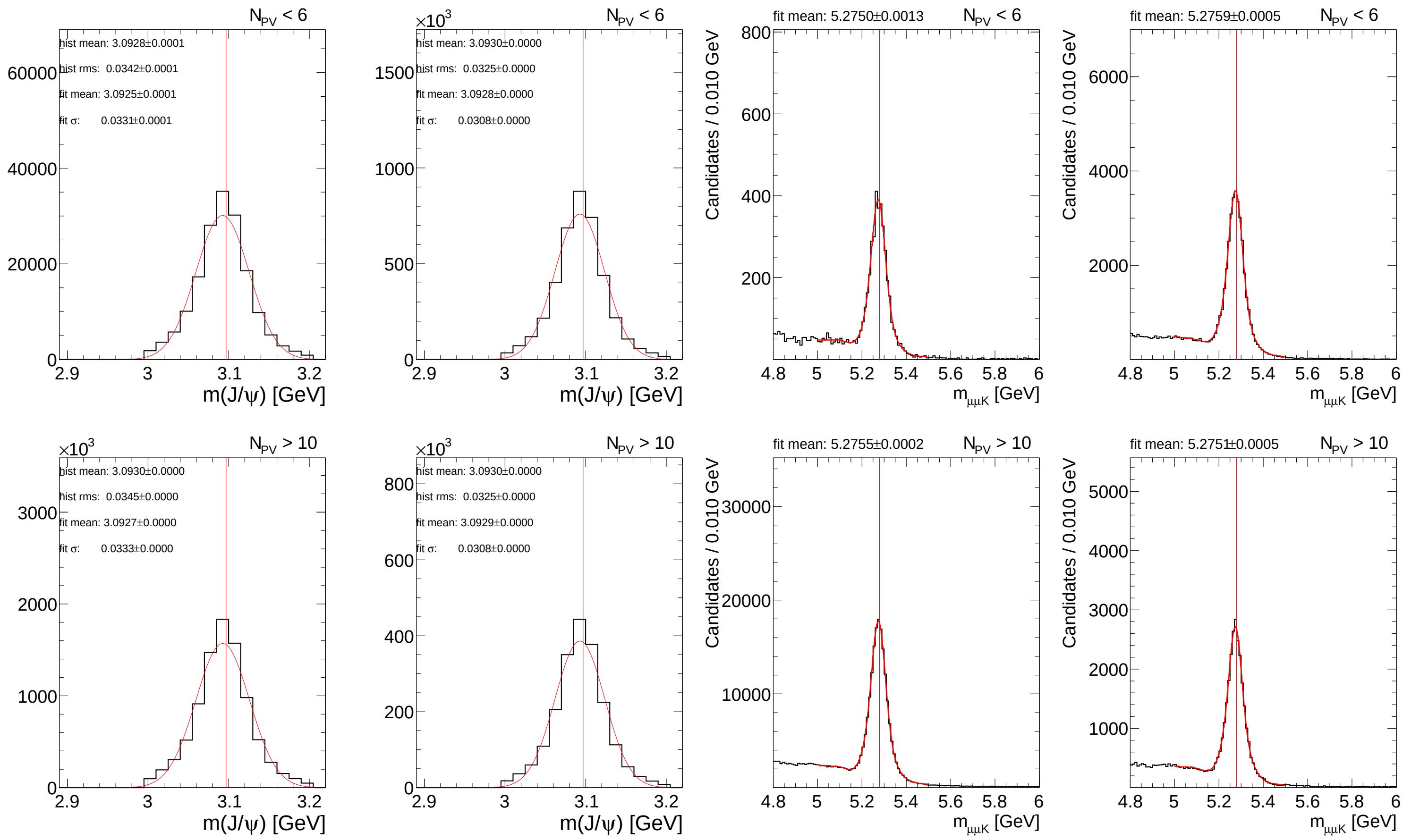

Figure 79: Example mass peaks for low pileup (top) and high pileup (bottom). The columns show J/ψ in 2012 (far left) and in 2011 (middle left), and $B^{\pm} \to J/\psi K^{\pm}$ in 2012 (middle right) and in 2011 (far right). The red line indicates the PDG expectation.

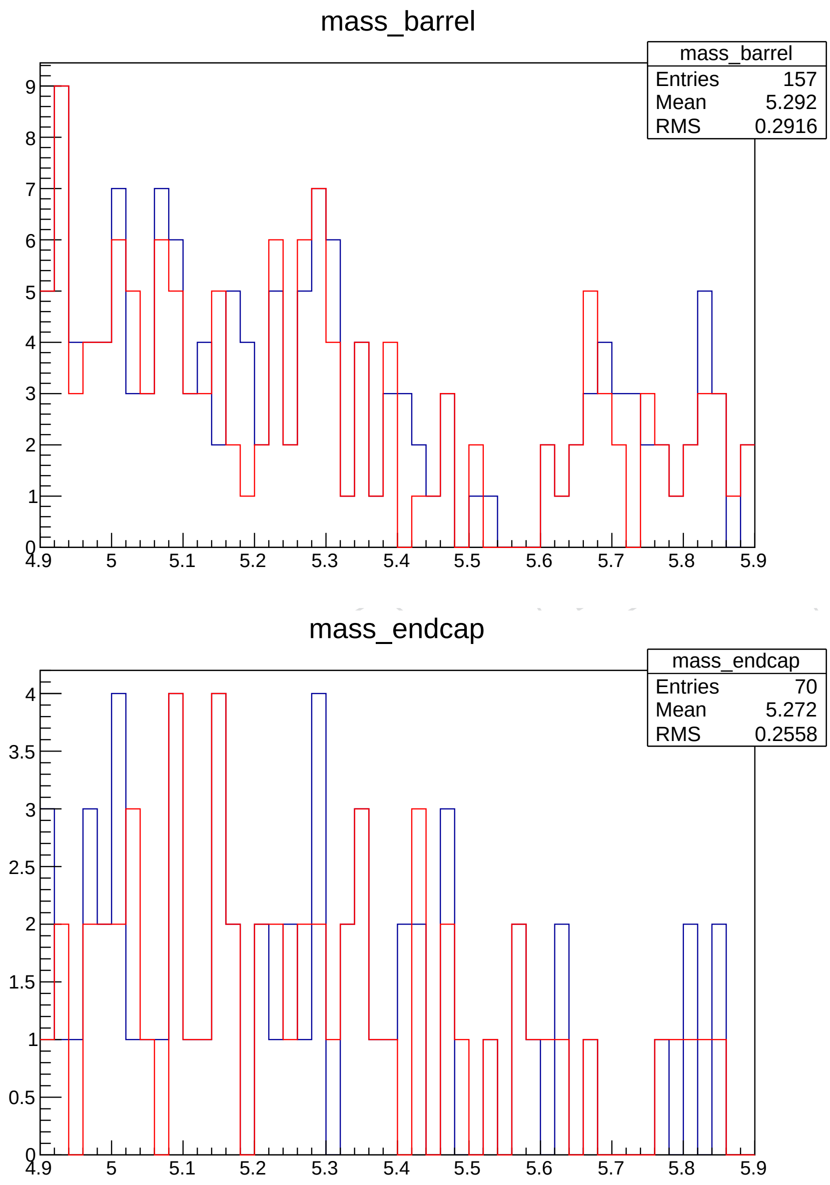

Figure 80: Comparison of distributions for events selected by the BDT in the barrel and the endcap regions. Red histogram show the new alignment, blue the old.

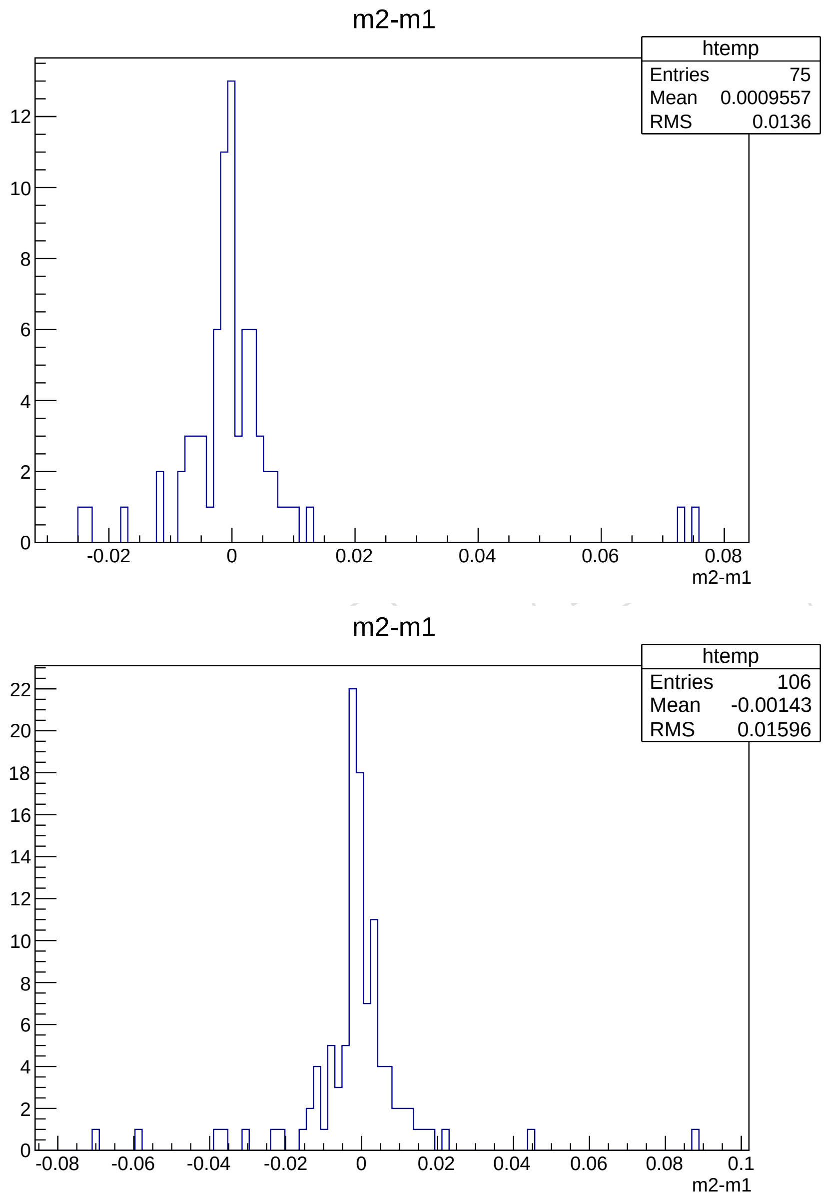

Figure 81: Event by event mass shift (new-old) in the 2012AB (top) and 2012C (bottom) run periods.

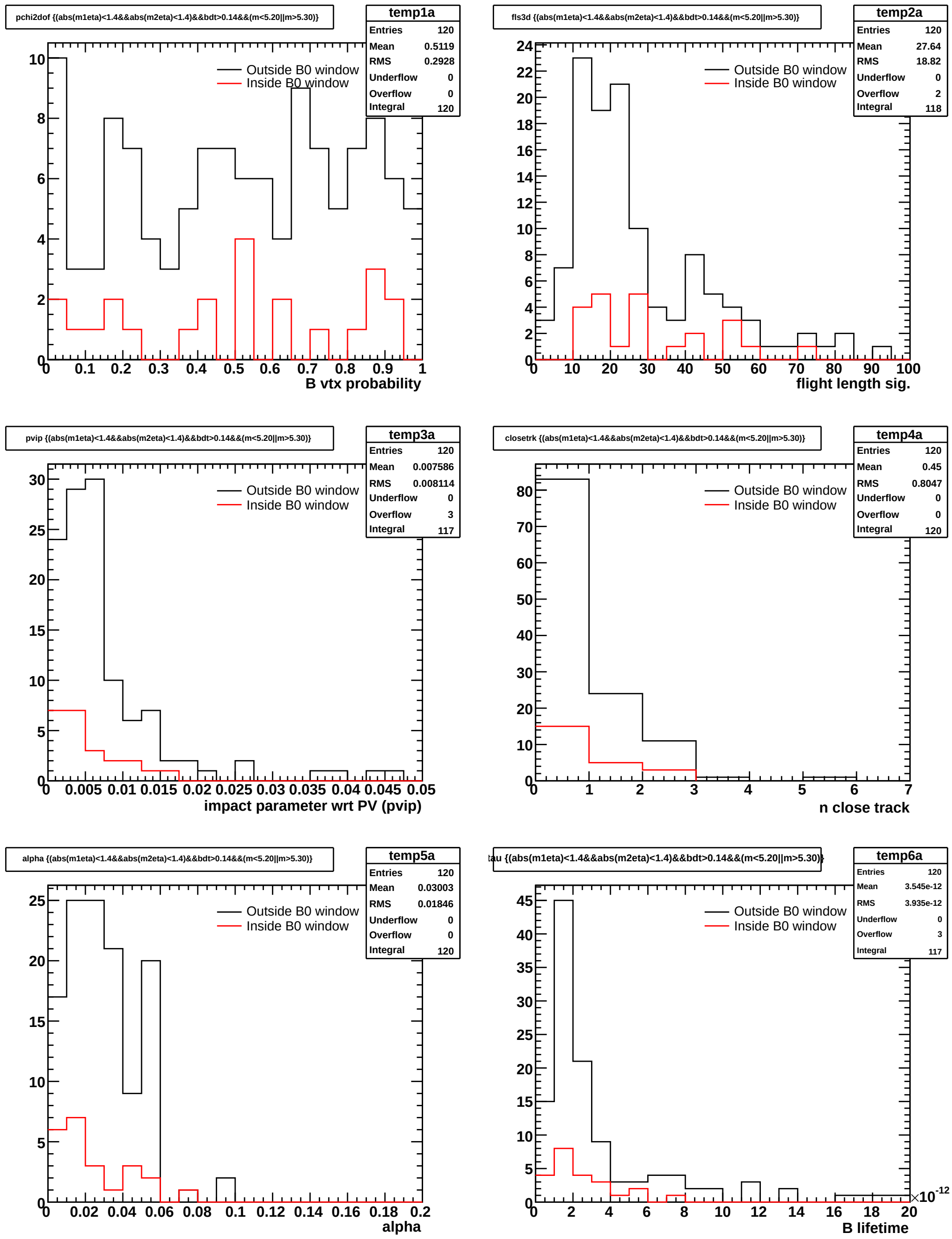

Figure 82: Comparison of distributions for events selected by the BDT in the barrel region from the B^0 mass window (red) with those outside the B^0 mass window (black). The B vertex probability (upper left), B flight length significance (upper right), B impact parameter (middle left), N_{trk}^{close} (middle right), α_{3D} (bottom left), and B lifetime (bottom right) are shown.

12.5 Track variables in B^0 mass window

Figure 83 shows a comparison of several silicon track variables between events selected in the B^0 mass window and those outside the mass window. All events plotted pass the BDT selection for the barrel region. No clear differences in the distributions are observed.

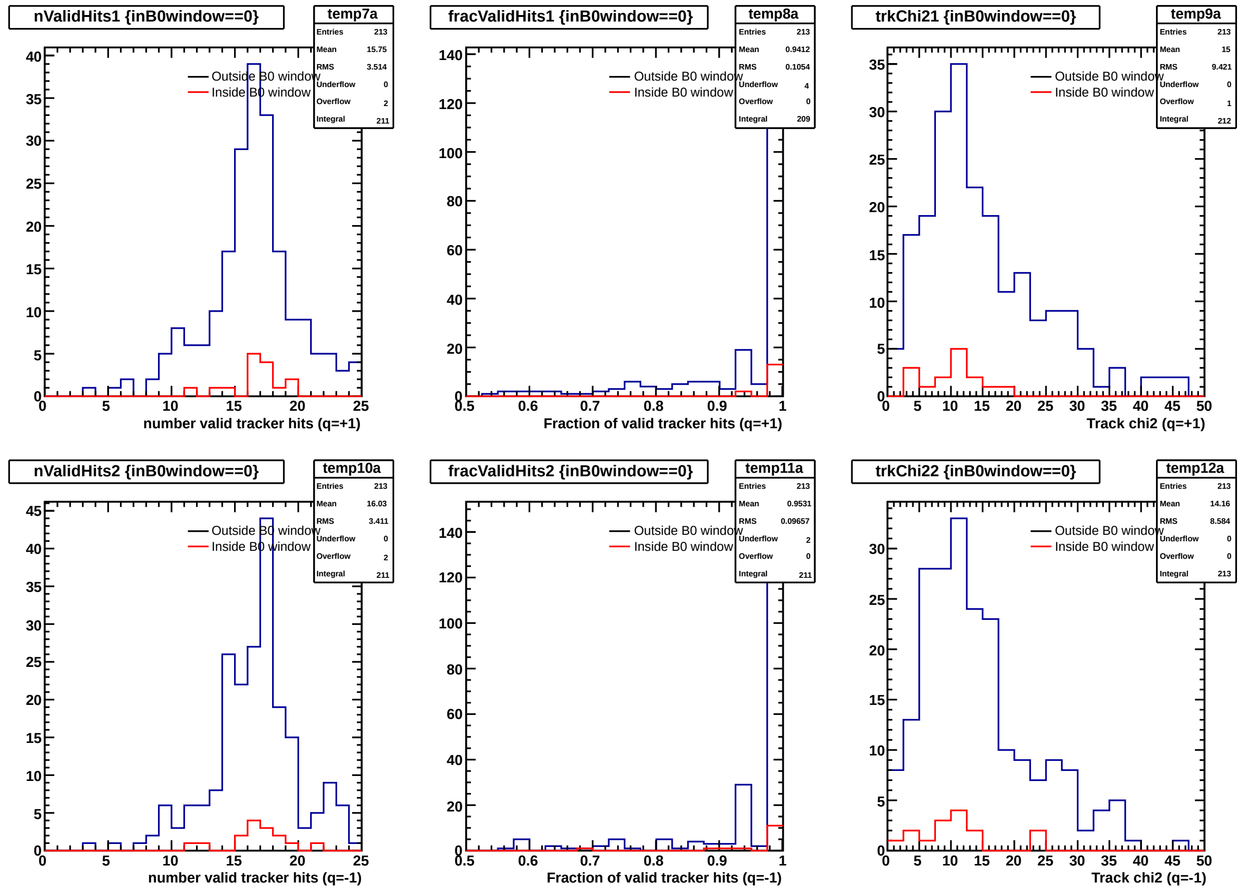

Figure 83: Comparison of distributions for events selected by the BDT in the barrel region from the B^0 mass window (red) with those outside the B^0 mass window (black). The follow silicon track quantities are compared for positive charged tracks (top) and negative charge tracks (bottom): number of valid hits on track (left), fraction of hits on track that are valid (middle), and the track χ^2 (right).

12.6 Cut ’n count analysis

We have used the selection criteria of the cut ’n count analysis, presented at Moriond 2012, to select candidates in the 2012 dataset. In Tab. 27 we summarize that selection criteria applied.

In Tab. 28 we summarize all the numbers relevant for the determination of the upper limit.

The unblinded mass distributions for the cut ’n count analysis are shown in Fig. 84.

With these results, one can determine the branching fractions in the binned analysis as

Table 27: Selection criteria for the cut 'n count analysis in the barrel and endcap.

Variable	Barrel	Endcap	units
$p_{\perp\mu,1} >$	4.5	4.5	GeV
$p_{\perp\mu,2} >$	4.0	4.2	GeV
$p_{\perp B} >$	6.5	8.5	GeV
$\langle w \rangle >$	0.6	0.6	
$\delta_{3D} <$	0.008	0.008	cm
$\delta_{3D}/\sigma(\delta_{3D}) <$	2.000	2.000	
$\alpha <$	0.050	0.030	rad
$\chi^2/dof <$	2.2	1.8	
$\ell_{3d}/\sigma(\ell_{3d}) >$	13.0	15.0	
$I >$	0.80	0.80	
$d^0_{ca} >$	0.015	0.015	cm
$N_{trk} <$	2	2	tracks

Table 28: Efficiency and background expectations and observations. The errors are the combined statistical and systematic errors. $N^{\text{exp}}_{\text{signal}}$ is the expected signal yield in the respective signal windows. $N^{\text{exp}}_{\text{cross-feed}}$ describes the cross feed of B^0_s into the B^0 signal window and vice versa. $N^{\text{exp}}_{\text{non-peak. bg}}$ combines rare semileptonic and combinatorial backgrounds. $N^{\text{exp}}_{\text{peak.bg}}$ summarizes the hadronic two-body peaking backgrounds. $N^{\text{exp}}_{\text{all bg}}$ sums up all backgrounds. $N^{\text{exp}}_{\text{total}}$ is the total of background and signal (plus cross feed). $N^{\text{obs}}_{\text{sidebands}}$ is the observed event count in the sideband, excluding the signal regions. The number of events N_{obs} is observed in the respective signal boxes.

Variable	$B^0 \to \mu^+\mu^-$ Barrel	$B^0_s \to \mu^+\mu^-$ Barrel	$B^0 \to \mu^+\mu^-$ Endcap	$B^0_s \to \mu^+\mu^-$ Endcap
Acceptance	0.237 ± 0.008	0.237 ± 0.008	0.218 ± 0.011	0.218 ± 0.011
$\varepsilon_{\text{analysis}}$	0.019 ± 0.001	0.019 ± 0.001	0.013 ± 0.001	0.013 ± 0.000
ε^{MC}_{μ}	0.683 ± 0.029	0.683 ± 0.028	0.803 ± 0.065	0.813 ± 0.065
$\varepsilon^{MC-TNP}_{\mu}$	0.790 ± 0.032	0.791 ± 0.032	0.835 ± 0.067	0.834 ± 0.067
ε^{TNP}_{μ}	0.793 ± 0.032	0.793 ± 0.032	0.770 ± 0.062	0.774 ± 0.062
$\varepsilon^{MC}_{\text{trig}}$	0.629 ± 0.023	0.624 ± 0.019	0.436 ± 0.030	0.439 ± 0.027
$\varepsilon^{MC-TNP}_{\text{trig}}$	0.844 ± 0.025	0.845 ± 0.025	0.739 ± 0.045	0.743 ± 0.045
$\varepsilon^{TNP}_{\text{trig}}$	0.798 ± 0.024	0.799 ± 0.024	0.748 ± 0.045	0.750 ± 0.045
ε_{tot}	0.0019 ± 0.0001	0.0018 ± 0.0001	0.0010 ± 0.0001	0.0010 ± 0.0001
$N^{\text{exp}}_{\text{signal}}$	0.534 ± 0.053	5.520 ± 0.828	0.178 ± 0.018	2.114 ± 0.317
$N^{\text{exp}}_{\text{cross-feed}}$	0.458 ± 0.069	0.225 ± 0.022	0.412 ± 0.062	0.114 ± 0.011
$N^{\text{exp}}_{\text{non-peak. bg}}$	2.036 ± 0.791	2.797 ± 1.015	2.284 ± 0.870	2.701 ± 1.024
$N^{\text{exp}}_{\text{peak.bg}}$	0.202 ± 0.077	0.054 ± 0.015	0.047 ± 0.017	0.024 ± 0.007
$N^{\text{exp}}_{\text{all bg}}$	2.238 ± 0.795	2.851 ± 1.015	2.331 ± 0.871	2.725 ± 1.024
$N^{\text{exp}}_{\text{total}}$	3.230 ± 1.082	8.596 ± 1.310	2.921 ± 0.970	4.953 ± 1.072
$N^{\text{obs}}_{\text{sidebands}}$	17		22	
N_{obs}	7	5	5	5

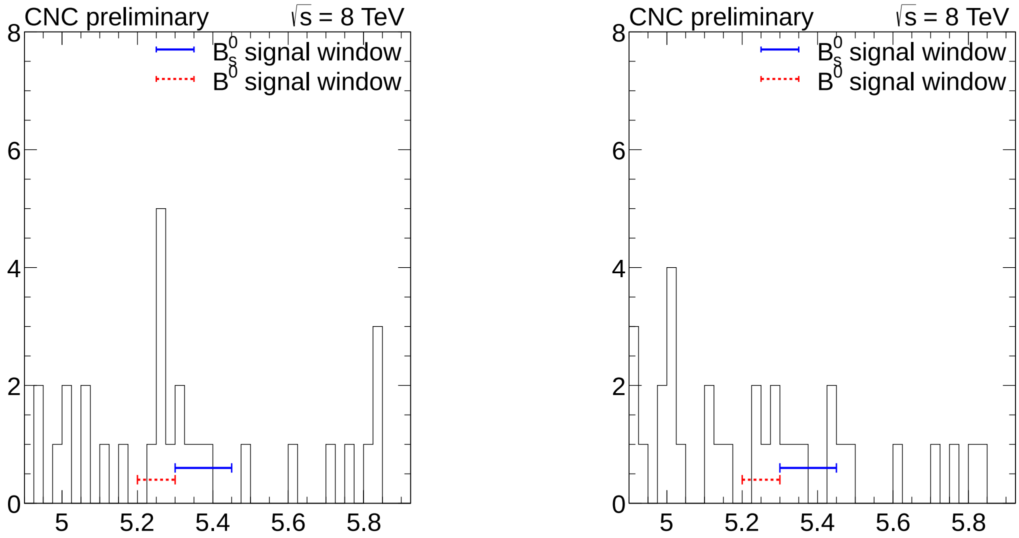

Figure 84: Dimuon invariant mass distribution in the barrel (left) and the endcap channel (right) with the cut 'n count analysis.

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (0.27^{+1.91}_{-0.27}) \times 10^{-9}$$
$$\mathcal{B}(B^0 \to \mu^+\mu^-) = (1.2^{+0.6}_{-0.6}) \times 10^{-9}$$

The upper limit at 95% C.L. is

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-) < 4.3 \times 10^{-9}$$
$$\mathcal{B}(B^0 \to \mu^+\mu^-) < 2.4 \times 10^{-9}$$

The significance of the $B_s^0 \to \mu^+\mu^-$ signal is 0.2σ, the significance of the $B^0 \to \mu^+\mu^-$ signal is at 2.7σ.

To check the stability of the excess, we have loosened

alpha < 0.06&&fls3d > 12&&chi2/dof < 2.5&&iso > 0.7&&closetrk < 3

and tightened

alpha < 0.05&&fls3d > 15&&chi2/dof < 2&&iso > 0.9&&closetrk < 3&&docatrk > 0.02&&pvips < 2&&pv

the selection in Fig. 85. The only purpose of these plots is to show that there seems to be a persistent excess in the B^0 window also in the cut 'n count analysis and that this excess is not an artifact of the BDT selection.

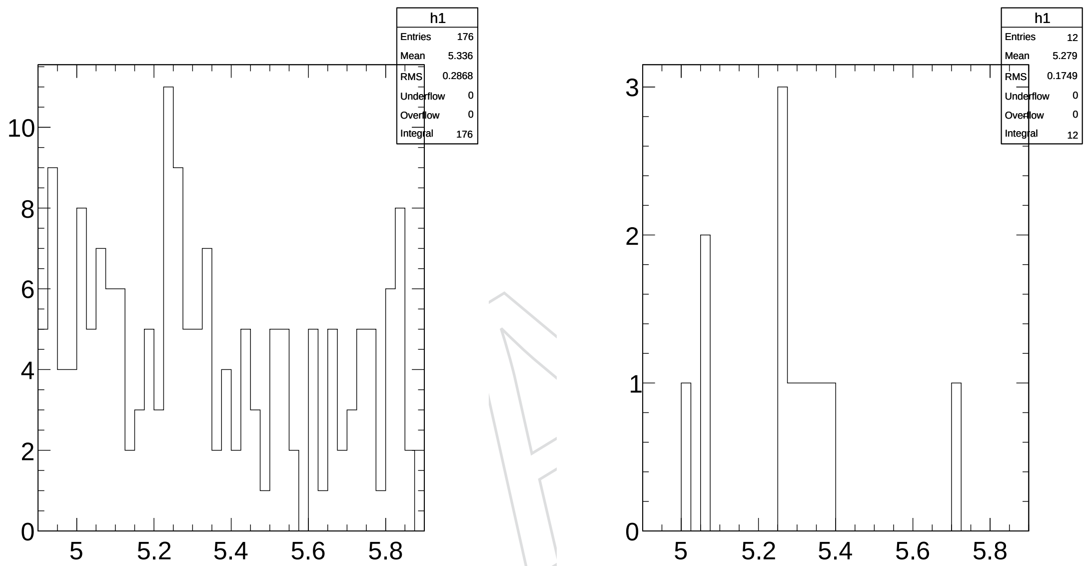

Figure 85: Dimuon invariant mass distribution with tight (left) and loose (right) cut ′n count selections. These plots are just illustrative examples of the excess when varying the selections.

882 # 13 Combination results

883 In this section we combine the results of the 2011 analysis [38] with this 2012 analysis.

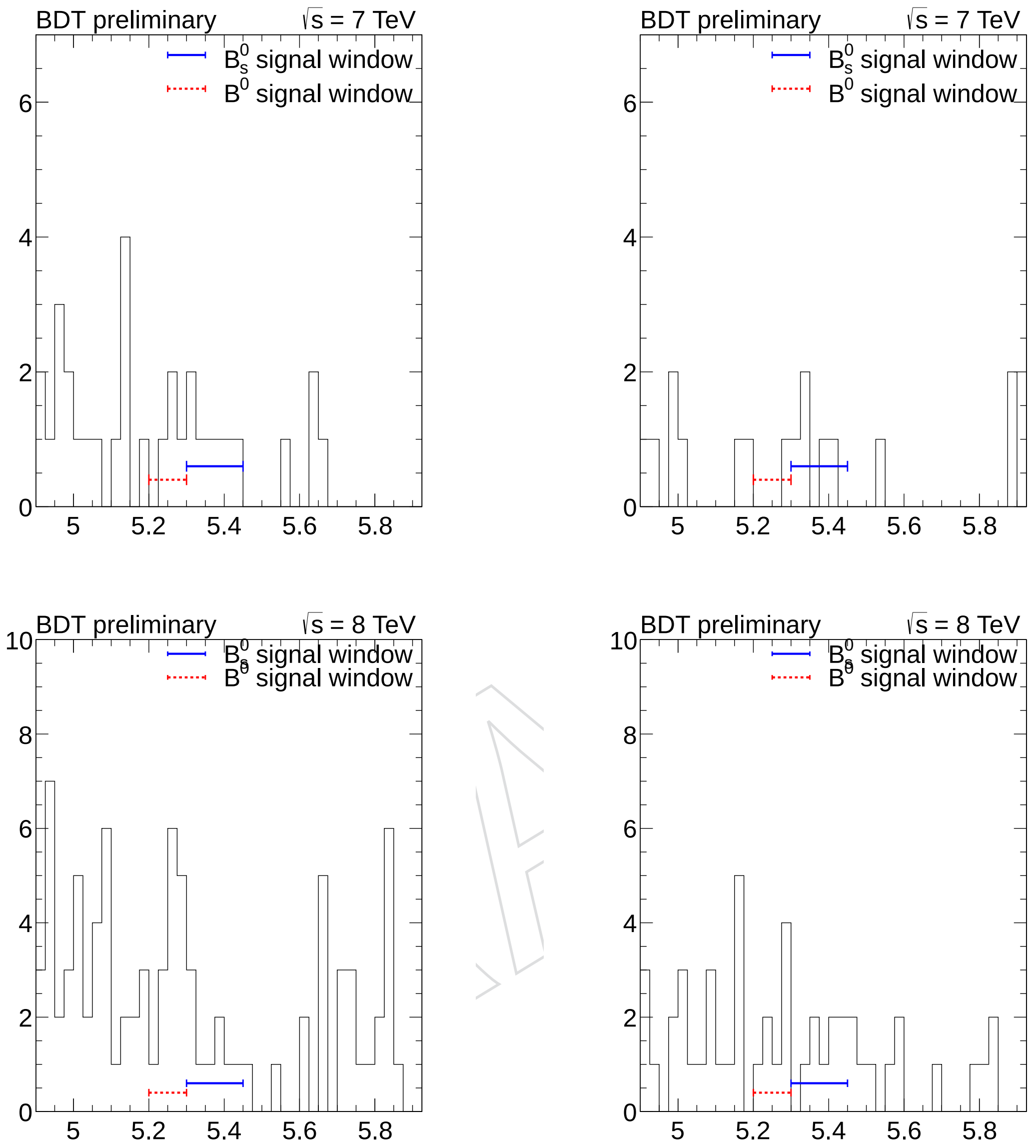

Figure 86: Dimuon invariant mass distribution in the barrel (left) and the endcap channel (right). The top row shows 2011, the bottom row shows 2012.

884 In Tab. 29 we summarize all numbers relevant for the extraction of the upper limit.

885 ## 13.1 Combined upper limit expectation with the BDT selection

Expected limits are computed as described in section 9 using the CL_s technique. Expected results for $B_s^0 \to \mu^+\mu^-$:

$6.0^{+1.8}_{-1.7} \times 10^{-9}$	SM
$2.4^{+1.1}_{-0.8} \times 10^{-9}$	Background only

Table 29: Summary event expectation from Tab. 24 from 2012 (top) and 2011 (bottom).

Variable	$B^0 \to \mu^+\mu^-$ Barrel	$B_s^0 \to \mu^+\mu^-$ Barrel	$B^0 \to \mu^+\mu^-$ Endcap	$B_s^0 \to \mu^+\mu^-$Endcap
2012				
$N^{\text{exp}}_{\text{signal}}$	0.955 ± 0.096	9.851 ± 1.478	0.260 ± 0.026	3.314 ± 0.497
$N^{\text{exp}}_{\text{cross-feed}}$	0.838 ± 0.126	0.384 ± 0.038	0.653 ± 0.098	0.172 ± 0.017
$N^{\text{exp}}_{\text{non-peak. bg}}$	7.312 ± 1.581	9.474 ± 1.917	3.546 ± 1.041	4.463 ± 1.296
$N^{\text{exp}}_{\text{peak.bg}}$	0.371 ± 0.141	0.099 ± 0.028	0.072 ± 0.027	0.036 ± 0.011
$N^{\text{exp}}_{\text{all bg}}$	7.683 ± 1.587	9.572 ± 1.917	3.618 ± 1.041	4.499 ± 1.296
$N^{\text{exp}}_{\text{total}}$	9.476 ± 1.868	19.808 ± 2.421	4.531 ± 1.163	7.985 ± 1.388
$N^{\text{obs}}_{\text{sidebands}}$	66		33	
N_{obs}	15	9	8	8
2011				
$N^{\text{exp}}_{\text{signal}}$	0.400 ± 0.040	4.435 ± 0.665	0.120 ± 0.012	1.407 ± 0.211
$N^{\text{exp}}_{\text{cross-feed}}$	0.341 ± 0.051	0.161 ± 0.016	0.305 ± 0.046	0.075 ± 0.008
$N^{\text{exp}}_{\text{non-peak. bg}}$	2.463 ± 0.493	2.486 ± 0.497	1.216 ± 0.243	1.402 ± 0.280
$N^{\text{exp}}_{\text{peak.bg}}$	0.156 ± 0.038	0.045 ± 0.008	0.030 ± 0.007	0.016 ± 0.003
$N^{\text{exp}}_{\text{all bg}}$	2.619 ± 0.494	2.531 ± 0.497	1.246 ± 0.243	1.418 ± 0.280
$N^{\text{exp}}_{\text{total}}$	3.361 ± 0.804	7.127 ± 0.831	1.671 ± 0.426	2.900 ± 0.351
$N^{\text{obs}}_{\text{sidebands}}$	21		10	
N_{obs}	4	7	1	5

Expected results for $B^0 \to \mu^+\mu^-$:

$$7.6^{+3.4}_{-2.3} \times 10^{-10} \qquad \text{SM}$$
$$6.4^{+2.8}_{-2.0} \times 10^{-10} \qquad \text{Background only}$$

The expected two sided 1σ interval from the binned analysis is

$$3.2^{+1.1}_{-1.3} \times 10^{-9} \qquad \text{SM } (B_s^0 \to \mu^+\mu^-)$$
$$1.0^{+2.8}_{-1.0} \times 10^{-10} \qquad \text{SM } (B^0 \to \mu^+\mu^-)$$

The expected p value for $B_s^0 \to \mu^+\mu^-$ ($B^0 \to \mu^+\mu^-$) is $3.0^{+39.5}_{-2.9} \times 10^{-3}$ ($3.5^{+4.6}_{-2.7} \times 10^{-1}$) corresponding to $2.7^{+1.0}_{-1.0}$ ($0.4^{+1.0}_{-1.3}$) sigmas. Probability to observe 3σ is 41% (0%).

13.2 Combined results with binned analysis

The observation is evaluated using the binned techniques to give the following upper limits

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-) \leq 3.6 \times 10^{-9} \qquad \text{at 95\% C.L.}$$
$$\mathcal{B}(B^0 \to \mu^+\mu^-) \leq 1.4 \times 10^{-9} \qquad \text{at 95\% C.L.}$$

The p-value of the $B_s^0 \to \mu^+\mu^-$ ($B^0 \to \mu^+\mu^-$) signal assuming unknown $B^0 \to \mu^+\mu^-$ ($B_s^0 \to \mu^+\mu^-$) signal is 1.3×10^{-1} (1.7×10^{-2}) corresponding to 1.1 (2.1) sigmas.

If the observed excess is assumed from a signal, the branching fraction can be measured to be (1σ uncertainty)

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (1.2^{+1.3}_{-0.9}) \times 10^{-9}$$
$$\mathcal{B}(B^0 \to \mu^+\mu^-) = (6.9^{+4.2}_{-3.7}) \times 10^{-10}$$

The corresponding confidence level plots are shown in figure 87 ($B_s^0 \to \mu^+\mu^-$) and figure 88.

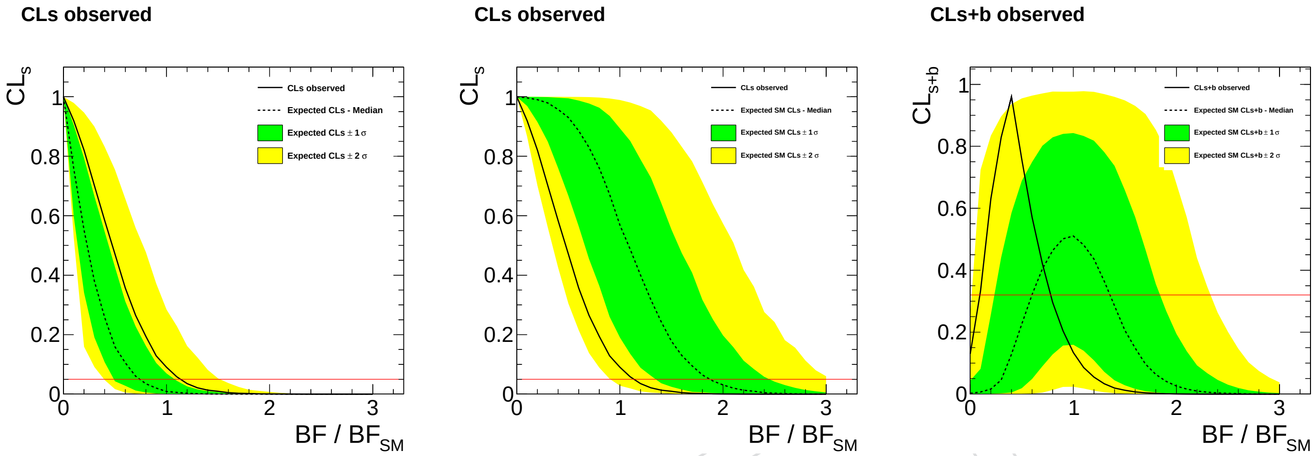

Figure 87: $B_s^0 \to \mu^+\mu^-$ confidence level plots. Left two plots show the observed CL_s curve with expected curve from background only model (left) and SM (middle). Two sided interval confidence level shown on the right

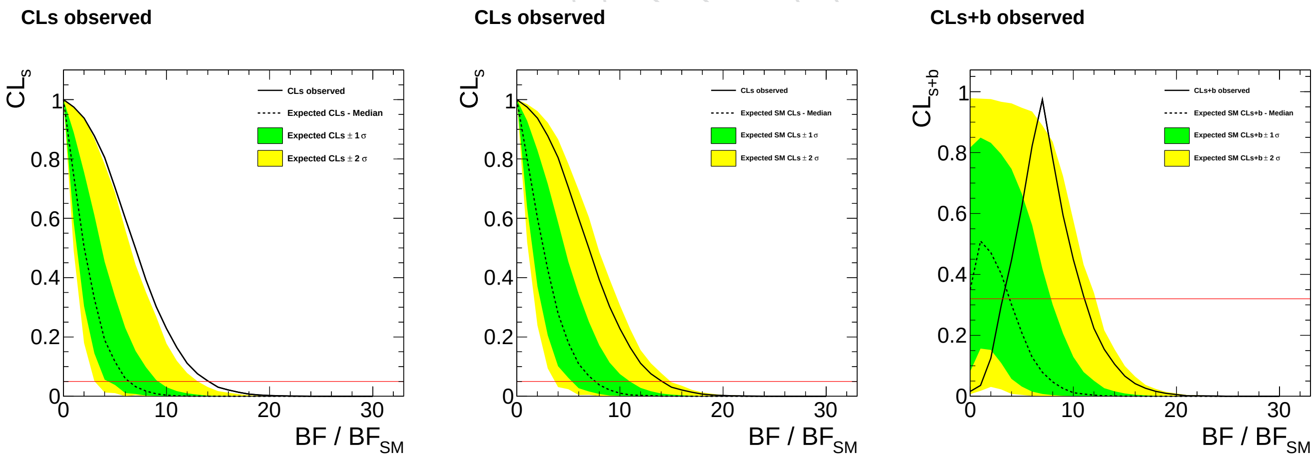

Figure 88: $B^0 \to \mu^+\mu^-$ confidence level plots. Left two plots show the observed CL_s curve with expected curve from background only model (left) and SM (middle). Two sided interval confidence level shown on the right

.

A comparison between the expected results and the fitted results is given in table 30

13.3 Results with the UML

The UML fit has been performed on the combined unblinded 2011 + 2012 data, putting a common branching fraction between all the categories (Fig. 89).

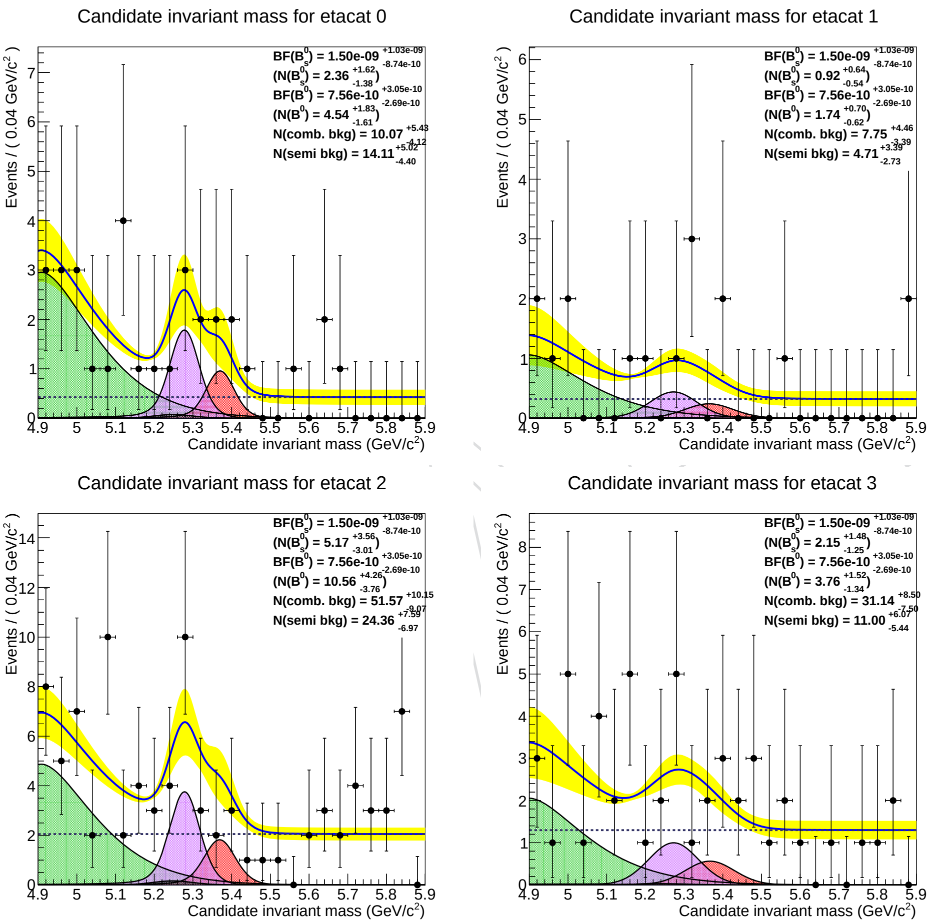

Figure 89: UML results for the barrel (left) and the endcap (right) categories.

Table 30: Expected yields compared to fitted yields in the binned analysis. **FixMe: expected numbers are 2012 only**

Variable	$B^0 \to \mu^+\mu^-$ Barrel	$B_s^0 \to \mu^+\mu^-$ Barrel	$B^0 \to \mu^+\mu^-$ Endcap	$B_s^0 \to \mu^+\mu^-$Endcap
$N^{\mathrm{exp}}_{\mathrm{signal}}$	0.955 ± 0.096	9.851 ± 1.478	0.260 ± 0.026	3.314 ± 0.497
$N^{\mathrm{exp}}_{\mathrm{non-peak.\ bg}}$	7.312 ± 1.581	9.474 ± 1.917	3.546 ± 1.041	4.463 ± 1.296
$N^{\mathrm{exp}}_{\mathrm{peak.bg}}$	0.371 ± 0.141	0.099 ± 0.028	0.072 ± 0.027	0.036 ± 0.011
$N^{\mathrm{fit}}_{\mathrm{signal}}$	8.458	5.050	2.435	1.698
$N^{\mathrm{fit}}_{\mathrm{comb.bg}}$	9.669	10.516	5.392	6.495
$N^{\mathrm{fit}}_{\mathrm{peak.bg}}$	0.535	0.143	0.104	0.052

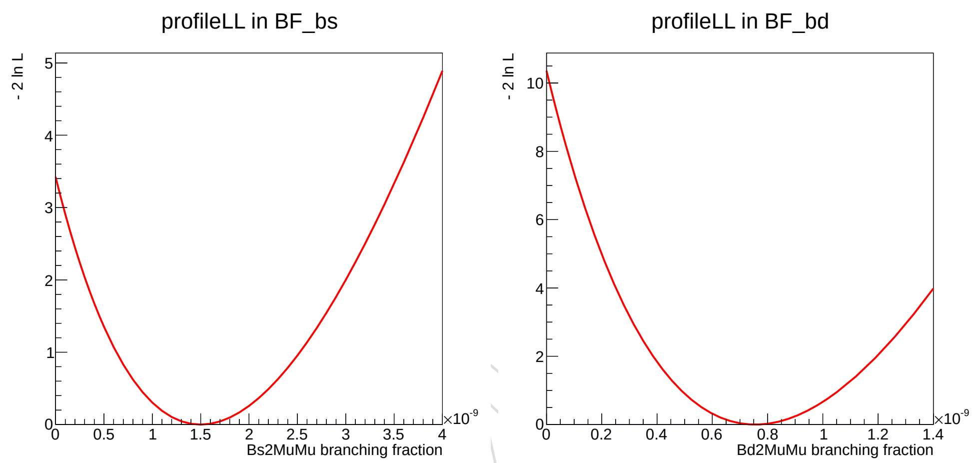

Figure 90: Profile likelihood scans versus the B_s^0 (left) and the B^0 (right) branching fractions.

The profile likelihood scans versus the branching fractions are in Fig. 90.

We obtain the following results for the measured branching fractions:

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (1.50^{+1.03}_{-0.874}) \times 10^{-9}$$
$$\mathcal{B}(B^0 \to \mu^+\mu^-) = (7.56^{+3.05}_{-2.69}) \times 10^{-10}$$

The significance for the two alternative hypotheses, taking into account all the systematic uncertainties (see below), are

$$sign(B_s^0 \to \mu^+\mu^-) = 1.3\sigma$$
$$sign(B^0 \to \mu^+\mu^-) = 2.3\sigma$$

The significance of the B^0 result above the SM prediction (i.e. assuming the SM B^0 branching fraction as the null hypothesis) is

$$sign(B^0 \to \mu^+\mu^-)_{SM} = 2.17\sigma \tag{26}$$

Table 31: Final invariant mass yields evaluated with the UML

Channel: barrel 2011					
Variable	low SB	B^0 window	B_s^0 window	high SB	**all**
$N_{B_s^0}$	0.06	0.24	1.90	0.16	2.36
N_{B^0}	0.51	2.50	1.52	0.00	4.54
N_{peak}	0.05	0.16	0.04	0.00	0.25
N_{semi}	12.84	0.83	0.37	0.07	14.11
N_{comb}	3.02	1.01	1.51	4.53	10.07
N_{all}	16.49	4.73	5.35	4.77	31.34
Channel: endcap 2011					
Variable	low SB	B^0 window	B_s^0 window	high SB	**all**
$N_{B_s^0}$	0.02	0.11	0.74	0.05	0.92
N_{B^0}	0.18	1.05	0.51	0.00	1.74
N_{peak}	0.02	0.03	0.02	0.00	0.07
N_{semi}	4.21	0.31	0.15	0.04	4.71
N_{comb}	2.33	0.78	1.16	3.49	7.75
N_{all}	6.76	2.27	2.59	3.58	15.20
Channel: barrel 2012					
Variable	low SB	B^0 window	B_s^0 window	high SB	**all**
$N_{B_s^0}$	0.14	0.53	4.17	0.32	5.17
N_{B^0}	1.01	6.10	3.44	0.00	10.56
N_{peak}	0.12	0.37	0.10	0.00	0.59
N_{semi}	22.16	1.44	0.64	0.12	24.36
N_{comb}	15.47	5.16	7.74	23.21	51.57
N_{all}	38.91	13.59	16.09	23.66	92.25
Channel: endcap 2012					
Variable	low SB	B^0 window	B_s^0 window	high SB	**all**
$N_{B_s^0}$	0.05	0.24	1.73	0.13	2.15
N_{B^0}	0.45	2.18	1.13	0.00	3.76
N_{peak}	0.05	0.07	0.04	0.00	0.16
N_{semi}	9.83	0.72	0.36	0.10	11.00
N_{comb}	9.34	3.11	4.67	14.01	31.14
N_{all}	19.72	6.32	7.93	14.24	48.20

The significance of the exclusion of the SM B_s^0 prediction (i.e. assuming the SM B_s^0 branching fraction as the null hypothesis) is

$$sign(B_s^0 \to \mu^+\mu^-)_{SM} = 2.6\sigma \tag{27}$$

In Table 31 there are all the fitted yields subdivided in the single contributions and mass regions.

13.3.1 uml systematic uncertainties

To evaluate the systematics associated to the pdf shapes for B_s^0 and B^0 branching fractions and for the B_s^0 and B^0 significances we varied the shapes in the following way:

- We change the pdf signal shape to a non-per-event error shape (a Crystal Ball + Gaussian). The systematics on the B_s^0 branching fraction is 17% and on the B^0 branching fraction is 5%.

- The degree of the polynomial of the semileptonic pdf is somehow arbitrary. We changed the degree to 4 and we fitted only with a simple exponential. The maximum systematics that this involves on the B_s^0 branching fraction is 5% and on the B^0 branching fraction is 3.5%.
- We change the relative branching fractions of the semileptonic backgrounds and the misid efficiencies in the most conservative way. We fixed the $\Lambda_b^0 \to p\mu^-\bar{\nu}$ branching ratio up one sigma, where the sigma is the sum in quadrature of the error on the branching fraction and the error on the proton misid. We also fixed the $B^0 \to \pi^-\mu^+\nu$ and $B_s^0 \to K^-\mu^+\nu$ branching ratios down one sigma, also in this case considering the kaon and pion misid. The systematics obtained on the B_s^0 branching fraction is 9% and on the B^0 branching fraction is 13%.
- We doubled the peaking background contribution obtaining a systematics of 0.7% on the B_s^0 branching fraction and a systematics of 4.4% on the B^0 branching fraction

References

[1] A. J. Buras, "Flavour Theory and the LHC Era", arXiv:1009.1303.

[2] M. Artuso et al., "B, D and K decays", *Eur. Phys. J.* **C57** (2008) 309–492, doi:10.1140/epjc/s10052-008-0716-1, arXiv:0801.1833.

[3] J. R. Ellis, J. S. Lee, and A. Pilaftsis, "B-Meson Observables in the Maximally CP-Violating MSSM with Minimal Flavour Violation", *Phys. Rev.* **D76** (2007) 115011, doi:10.1103/PhysRevD.76.115011, arXiv:0708.2079.

[4] S. R. Choudhury, A. S. Cornell, N. Gaur et al., "Signatures of new physics in dileptonic B-decays", *Int. J. Mod. Phys.* **A21** (2006) 2617–2634, doi:10.1142/S0217751X06029491, arXiv:hep-ph/0504193.

[5] J. Parry, "Lepton flavor violating Higgs boson decays, $\tau \to \mu\gamma$ and $B_s \to \mu^+\mu^-$ in the constrained MSSM+NR with large tan beta", *Nucl.Phys.* **B760** (2007) 38–63, doi:10.1016/j.nuclphysb.2006.10.011, arXiv:hep-ph/0510305.

[6] C. Bobeth, T. Ewerth, F. Kruger et al., "Enhancement of $B(\bar{B}_d \to \mu^+\mu^-) / B(\bar{B}_s \to \mu^+\mu^-)$ in the MSSM with minimal flavor violation and large tan beta", *Phys. Rev.* **D66** (2002) 074021, doi:10.1103/PhysRevD.66.074021, arXiv:hep-ph/0204225.

[7] S. Davidson and S. Descotes-Genon, "Minimal Flavour Violation for Leptoquarks", *JHEP* **11** (2010) 073, doi:10.1007/JHEP11(2010)073, arXiv:1009.1998.

[8] J. R. Ellis, K. A. Olive, Y. Santoso et al., "On $B_s \to \mu^+\mu^-$ and cold dark matter scattering in the MSSM with non-universal Higgs masses", *JHEP* **05** (2006) 063, doi:10.1088/1126-6708/2006/05/063, arXiv:hep-ph/0603136.

[9] A. Dedes, J. Rosiek, and P. Tanedo, "Complete One-Loop MSSM Predictions for $B^0 \to l^+l'^-$ at the Tevatron and LHC", *Phys. Rev.* **D79** (2009) 055006, doi:10.1103/PhysRevD.79.055006, arXiv:0812.4320.

[10] S. Baek, "A two-loop contribution to $B_s \to \mu^+\mu_-$ at large $tan\beta$ in the MSSM", *Phys. Lett.* **B595** (2004) 461–468, doi:10.1016/j.physletb.2004.05.071, arXiv:hep-ph/0406007.

[11] F. Zwirner, "The quest for low-energy supersymmetry and the role of high-energy e+ e- colliders", arXiv:hep-ph/9203204.

[12] K. De Bruyn, R. Fleischer, R. Knegjens et al., "Branching Ratio Measurements of B_s Decays", *Phys.Rev.* **D86** (2012) 014027, doi:10.1103/PhysRevD.86.014027, arXiv:1204.1735.

[13] K. De Bruyn, R. Fleischer, R. Knegjens et al., "Probing New Physics via the $B^0_s \to \mu^+\mu^-$ Effective Lifetime", *Phys.Rev.Lett.* **109** (2012) 041801, doi:10.1103/PhysRevLett.109.041801, arXiv:1204.1737.

[14] A. J. Buras, J. Girrbach, D. Guadagnoli et al., "On the Standard Model prediction for BR(Bs,d to mu+ mu-)", arXiv:1208.0934.

[15] CLEO Collaboration, "Search for decays of B^0 mesons into pairs of leptons: $B^0 \to e^+e^-$, $B^0 \to \mu^+\mu^-$ and $B^0 \to e^\pm\mu^\mp$", *Phys. Rev.* **D62** (2000) 091102, doi:10.1103/PhysRevD.62.091102, arXiv:hep-ex/0007042.

[16] BELLE Collaboration, “Search for $B^0 \to \ell^+\ell^-$ at BELLE”, *Phys. Rev.* **D68** (2003) 111101, doi:10.1103/PhysRevD.68.111101, arXiv:hep-ex/0309069.

[17] BaBar Collaboration, “Search for decays of B^0 mesons into e^+e^-, $\mu^+\mu^-$, and $e^\pm\mu^\mp$ final states”, *Phys. Rev.* **D77** (2008) 032007, doi:10.1103/PhysRevD.77.032007, arXiv:0712.1516.

[18] CDF Collaboration, “Search for $B_s^0 \to \mu^+\mu^-$ and $B_d^0 \to \mu^+\mu^-$ decays with 2 fb^{-1} of $p\bar{p}$ collisions”, *Phys. Rev. Lett.* **100** (2008) 101802, doi:10.1103/PhysRevLett.100.101802, arXiv:0712.1708.

[19] D0 Collaboration, “Search for the rare decay $B_s^0 \to \mu^+\mu^-$”, *Phys. Lett.* **B693** (2010) 539–544, doi:10.1016/j.physletb.2010.09.024, arXiv:1006.3469.

[20] LHCb Collaboration, “Search for the rare decays $B_s^0 \to \mu^+\mu^-$ and $B_d^0 \to \mu^+\mu^-$”, *Phys. Lett.* **B699** (2011) 330–340, doi:10.1016/j.physletb.2011.04.031, arXiv:1103.2465.

[21] CDF Collaboration Collaboration, “Search for $B_s \to \mu^+\mu^-$ and $B_d \to \mu^+\mu^-$ Decays with CDF II”, *Phys.Rev.Lett.* **107** (2011) 191801, doi:10.1103/PhysRevLett.107.191801, arXiv:1107.2304.

[22] CMS Collaboration Collaboration, “Search for B_s^0 and B^0 to dimuon decays in pp collisions at 7 TeV”, *Phys.Rev.Lett.* **107** (2011) 191802, arXiv:1107.5834.

[23] LHCb Collaboration Collaboration, “Search for the rare decays $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$”, arXiv:1112.1600.

[24] CMS Collaboration Collaboration, “Search for $B_s^0 to \mu^+\mu^-$ and $B^0 to \mu^+\mu^-$ decays”, *JHEP* **1204** (2012) 033, doi:10.1007/JHEP04(2012)033, arXiv:1203.3976.

[25] LHCb collaboration Collaboration, “Strong constraints on the rare decays $B_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$”, *Phys.Rev.Lett.* **108** (2012) 231801, doi:10.1103/PhysRevLett.108.231801, arXiv:1203.4493.

[26] ATLAS Collaboration Collaboration, “Search for the decay $B_s^0 \to \mu^+\mu^-$ mu with the ATLAS detector”, *Phys.Lett.* **B713** (2012) 387–407, doi:10.1016/j.physletb.2012.06.013, arXiv:1204.0735.

[27] LHCb Collaboration, “Measurement of b hadron production fractions in 7 TeV pp collisions”, *Phys. Rev. D* **85** (2012) 032008, doi:10.1103/PhysRevD.85.032008, arXiv:1111.2357v2.

[28] Particle Data Group Collaboration, “Review of Particle Physics (RPP)”, *Phys.Rev.* **D86** (2012) 010001, doi:10.1103/PhysRevD.86.010001.

[29] Heavy Flavor Averaging Group Collaboration, “Averages of b-hadron, c-hadron, and τ-lepton properties as of early 2012”, arXiv:1207.1158.

[30] Z. Ligeti (2012). Private communication.

[31] CMS Collaboration, “Search for the $B_{s(d)} \to \mu^+\mu^-$ decay with the 2011 data”, *CMS AN-11-380* (2011).

[32] W. Erdmann, "Offline Primary Vertex Reconstruction with Deterministic Annealing Clustering", *CMS NOTE 2011-014* (2011).

[33] CMS Collaboration, "Tracking and Vertexing Results from First Collisions", *CMS PAS TRK-10-001* (2010).

[34] "Baseline muon selections",. Muon POG, https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideMuonId.

[35] K. Prokofiev and T. Speer, "A kinematic and a decay chain reconstruction library",. Prepared for Computing in High-Energy Physics (CHEP '04), Interlaken, Switzerland, 27 Sep - 1 Oct 2004.

[36] CMS Collaboration, "Performance of CMS muon identification in pp collisions at $\sqrt{s}$ = 7 TeV", *CMS PAPER MUO-10-004* (2011).

[37] A. L. Read, "Modified frequentist analysis of search results (the CL_s method)", *CERN-OPEN-2000-205* (2000).

[38] CMS Collaboration, "Search for the $B_{s(d)} \to \mu^+\mu^-$ decay with a BDT in the 2011 data", *CMS AN-12-238* (2011).