CDF/ANAL/BOTTOM/CDFR/10180
Draft 3.0
August 23, 2010

A measurement of the ratio of the branching fraction of $B_s \to J/\psi K_S$ to $B^0 \to J/\psi K_S$

Olga Norniella and Kevin Pitts
University of Illinois at Urbana-Champaign

Abstract

This note presents a measurement of the ratio of the branching fractions of $B_s \to J/\psi K_S$ to $B^0 \to J/\psi K_S$ based on 5.9 fb^{-1} of data. Using a sample of reconstructed $B^0 \to J/\psi K_S$ collected via the di-muon trigger, decay of $B_s \to J/\psi K_S$ are identified. Fitting the B mass distribution with a binned likelihood function, a sample of 64 $\pm$ 14 reconstructed $B_s \to J/\psi K_S$ are observed with a statistical significance of 7.2 σ. The ratio $N(B_s \to J/\psi K_S)/N(B^0 \to J/\psi K_S)$ is measured to be 0.0108 $\pm$ 0.0019 (stat.) $\pm$ 0.0010 (sys.). This measurement allows a determination of $Br(B_s \to J/\psi K_S)/Br(B^0 \to J/\psi K_S) = 0.041 \pm 0.007$ (stat.)$\pm$ 0.004 (sys.) $\pm$ 0.005 (frag.).

Contents

1 Introduction

The $B_s \to J/\psi K_S$ decay is very similar to the $B^0 \to J/\psi K_S$ transition. The only substantial difference is the $b \to c\bar{c}s$ transition in $B^0 \to J/\psi K_S$ versus the $b \to c\bar{c}d$ transition in $B_s \to J/\psi K_S$. Therefore the $B_s \to J/\psi K_S$ is Cabibbo suppressed in comparison to $B^0 \to J/\psi K_S$.

The purpose of this analysis is to measure the ratio of branching fractions of $B_s \to J/\psi K_S$ to $B^0 \to J/\psi K_S$. Since the two decay modes have identical final states and extremely similar kinematics, the reconstruction efficiency cancels in the ratio and we can use the relation

$$\frac{N(B_s \to J/\psi K_S)}{N(B^0 \to J/\psi K_S)} = \frac{f_s}{f_d} \frac{Br(B_s \to J/\psi K_S)}{Br(B^0 \to J/\psi K_S)} \frac{1}{A_{rel}}, \tag{1}$$

where $N(B_s \to J/\psi K_S)$ and $N(B^0 \to J/\psi K_S)$ are the yields of B_s and B^0 signals respectively, A_{rel} is the relative acceptance of $B^0 \to J/\psi K_S$ to $B_s \to J/\psi K_S$ (see Equation 2) and $\frac{f_s}{f_d}$ is the ratio of fragmentation functions.

$$A_{rel} = \frac{N(B^0 \to J/\psi K_S \; passed)/N(B^0 \to J/\psi K_S \; generated)}{N(B_s \to J/\psi K_S \; passed)/N(B_s \to J/\psi K_S \; generated)}, \tag{2}$$

Assuming that the acceptance and efficiency to reconstruct $J/\psi K_S$ is the same for B^0 and B_s ($A_{rel} \sim 1$) and plugging into Equation 1 the following numbers: $\frac{f_s}{f_d}$=0.28 [1], Br($B^0 \to J/\psi K_S$)=4.4x10^{-4} [1] and Br($B_s \to J/\psi K_S$)=2x10^{-5} [2], the ratio of yields is estimated to be 0.011.

Once the relative ratio of branching fractions is obtained the absolute branching fraction of $B_s \to J/\psi K_0$ can be calculated using Equation 1 and the PDG value of the $B^0 \to J/\psi K_0$ branching ratio. In addition, the observation of the $B_s \to J/\psi K_S$ decay, which is a CP eigenstate, will allow a measurement of its lifetime. This future measurement, not included in this note, will be a direct measurement of $\tau_{Bs\ (Heavy)}$, being $\tau_{Bs\ (Heavy)} > \tau_{Bs\ (Light)}$. Furthermore, the $B_s \to J/\psi K_S$ decay can be use to extract the angle γ of the unitary triangle [2].

2 Measurement

The measurement starts with a sample of reconstructed $B^0 \to J/\psi K_S$, acquired via the di-muon trigger and ntuplized as BStNtuple. After preselection cuts, a Neural Network is used to separate the combinatorial background from the signal. Finally, the $N(B_s)$ yield is determined after fitting the B mass distribution with a binned likelihood function.

2.1 Data sample

The data sample consists of $B^0 \to J/\psi K_S$ events from the J/ψ dataset (di-muon triggers). The sample includes events taken from runs 138425 to 289197. Only runs enclosed in *goodrun_b_bs_nocal_nosvt.list* good run list (version 34) are included. This sample corresponds to data taken through period 28 and an integrated luminosity of 5.9 fb^{-1}.

2.2 Monte Carlo samples

In different steps of this analysis Monte Carlo (MC) samples are used. The B^0, B_s samples are generated using BGen decay model SVS (pseudoscalar to vector pseudoscalar). We explicitly turn off mixing and CP violation in these samples. The Λ_b MC sample is generated using a phase space decay model.

For B^0 MC and B_s MC the events are forced to decay to $J/\psi K_S$, where $J/\psi \to \mu\mu$ and $K_S \to \pi\pi$. Similarly, Λ_b events are forced to decay to $J/\psi\Lambda$, where $J/\psi \to \mu\mu$ and $\Lambda \to \pi p$.

Special attention has been used when $B_s \to J/\psi K_S$ MC was produced, insuring that the sample has been generated with the correct lifetime. Since $B_s \to J/\psi K_S$ is a CP eigenstate (ignoring small CP violation in K_S decays), the MC has been generated with a lifetime of $\tau_{Bs\ (Heavy)}$= 463 μm. This number comes from the PDG evaluation of $\tau_{Bs\ (Heavy)}$ which is 1.543 +0.058 -0.060 ps. This lifetime is the PDG average for the $B_s(Heavy)$ state, which is the longer lived of the $B_{s,H}$ and $B_{s,L}$ weak eigenstates. Supplemental samples were generated for systematic acceptance studies.

All samples are passed through cdfsim, TRGSim++, production and then ntuplized.

2.3 Event Selection

2.3.1 Preselection Cuts

BStNtuple cuts

Data and MC samples are ntuplized as BStNtuple, and in data all the events with at least one $B^0 \to J/\psi K_S$ candidate are selected. Inside the BStNtuple the following selection criteria are applied:

- 3 GeV/c^2 < M_B < 6.7 GeV/c^2
- $\chi^2(B)$ <50
- 2.8 GeV/c^2 < Mass(J/ψ)< 3.75 GeV/c^2
- $\chi^2(J/\psi)$ < 30
- 0.45 GeV/c^2 < Mass(K_S)< 0.55 GeV/c^2

- $\chi^2(K_S) < 20$
- $L_{xy}(K_S) > 0.5$ cm
- $p_T(\pi) > 0.35$ GeV/c (for both pions)

Trigger cuts

The J/ψ dataset includes several trigger paths, each one with specific online requirements. Unfortunately, these online cuts (not implemented in the simulation) affect the final kinematics of the selected B^0 candidates. In order to reduce this trigger bias, a set of offline cuts are also applied. These cuts allow a more "unbiased" comparison between the real data and the simulation.

- $\Delta\phi(\mu\mu) < 2.25$ radians
- $\Delta z(\mu\mu) < 5$ cm
- 2.7 GeV/c^2 < Mass($\mu\mu$)< 4 GeV/c^2
- Charge(μ) * Charge(μ) < 0
- Both μ matched XFT
- $p_T(\mu) > 1.5$ GeV/c (for both muons)

Neural Network optimization cuts

A Neural Network is used in this analysis to remove combinatorial background. One can find more information in the section 2.3.2. Prior the Neural Network training some additional cuts are used in the analysis to improve the Neural Network performance.

- B Fit Probability > 10^{-5}
- Mass($\mu\mu$)< 3.3 GeV/c^2
- $L_{xy}(B)/\sigma > 2$
- Number of COT Hits (axial/stereo) ≥ 10
- $p_T(B) > 4$ GeV/c
- $p_T(\pi) > 0.5$ GeV/c (for both pions)

Cut to remove Λ_b contribution

Finally, $\Lambda_b \to J/\psi\Lambda$, where $\Lambda \to p\pi$, is a background in this analysis when the p is reconstructed as a π. Figure 1 shows the mass distribution of B candidates from B_s MC and Λ_b MC.

In order to suppress the Λ_b contribution, a study comparing different variables for events coming from B_s MC and Λ_b MC has been performed. As shown by Figure 2 the cos (θ_{K_S,π_2}), where π_2 is the π with lower p_T, helps to distinguish B_s events from Λ_b events. The angle is defined in the K_S center of mass frame. As a result, a cut in this angular variable, cos (θ_{K_S,π_2})> -0.75, is applied. This cut decreases the acceptance for Λ_b by a factor of 99.8%, but only 14.2% for B_s. This cut is more efficient in removing Λ_b background than a cut on the reconstructed Λ mass (see A for more details).

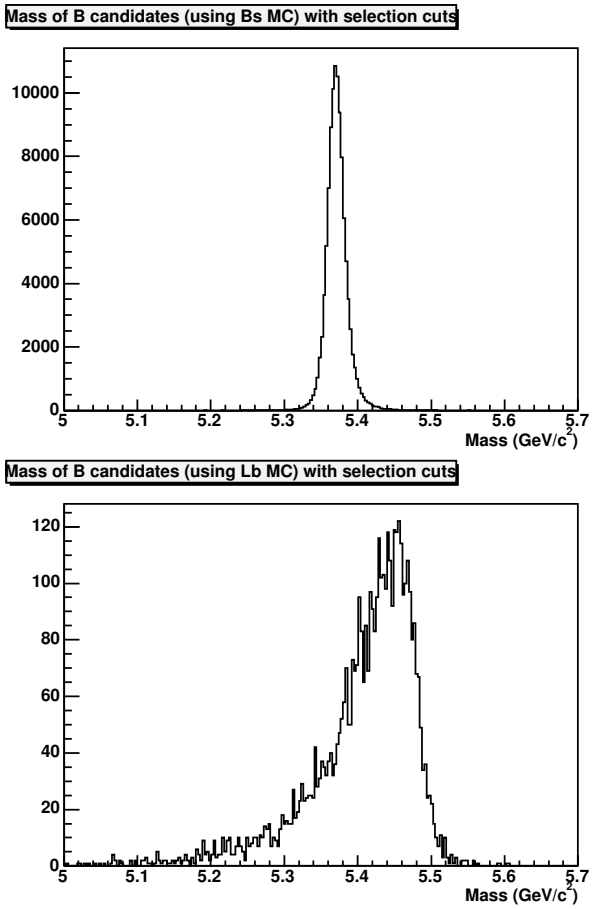

Figure 1: Mass distribution of B candidates from B_s MC (top) and Λ_b MC (bottom).

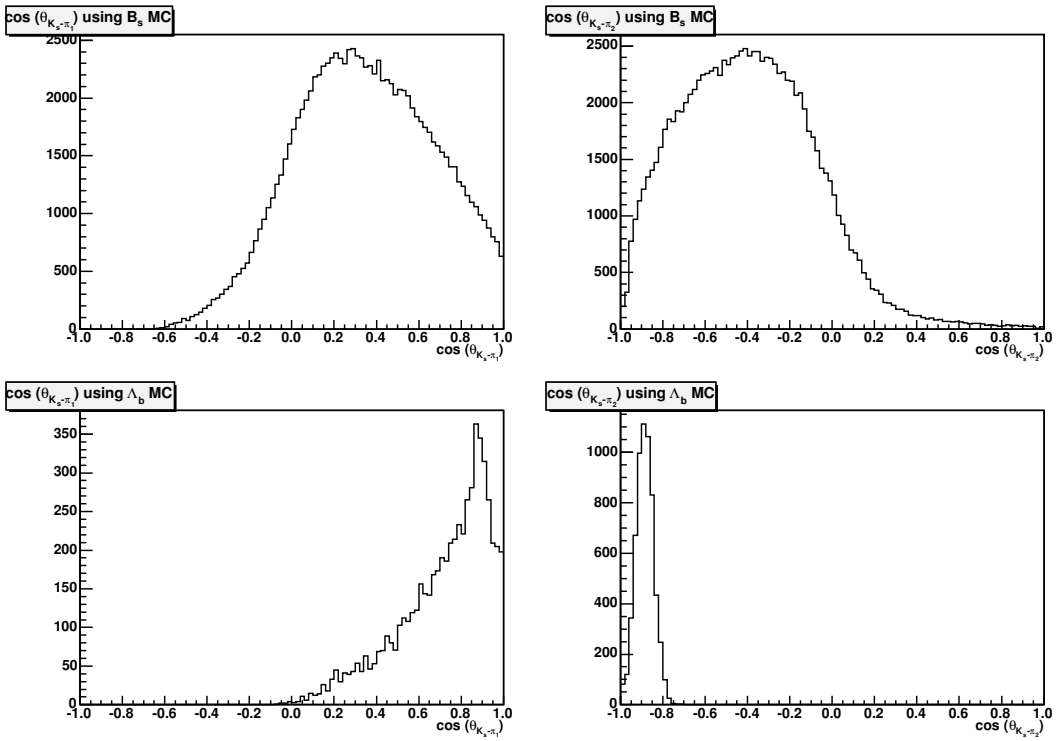

Figure 2: Cosine of the angle between the reconstructed K_S and both π's, where $p_T(\pi_1) > p_T(\pi_2)$. On the top the distributions for the B_s and on the bottom the ones for Λ_b. In both cases, cos (θ_{K_S,π_1}) on the left, and cos (θ_{K_S,π_2}) on the right.

It is important to note that the suppression of this background at this point of the analysis is crucial. As explained in the next section, the training of the Neural Network uses the upper side band from data as the background sample. Events coming from Λ_b, which are located in this upper side band, are topological very similar to signal events. Therefore, keeping them in the background sample could affect the Neural Network performance.

2.3.2 Neural Network

After applying all preselection cuts, an artificial Neural Network is used to do the final signal selection. The Neural Network, which is constructed by means of the NeuroBayes program package [3], is trained to separate B_s events from combinatorial background. In order to train the Neural Network simulated B_s MC events are used as signal and real data from the upper side band, well separated from the signal region in the corresponding invariant mass, as a background data sample. The mass range for signal and background are shown in Figure 3. The mass range for signal is from 5.35 GeV/c^2 to 5.4 GeV/c^2. Events inside the range 5.45 GeV/c^2 to 6 GeV/c^2 are part of the background sample.

Twenty-two variables are chosen as inputs for the Neural Network training. Appendix B includes a specific study to check that all these variable are well simulated in the MC. The input variable list can be found in Table 1. Figures 4 to 6 show the distributions, for signal and background, of the input variables to the Neural Network.

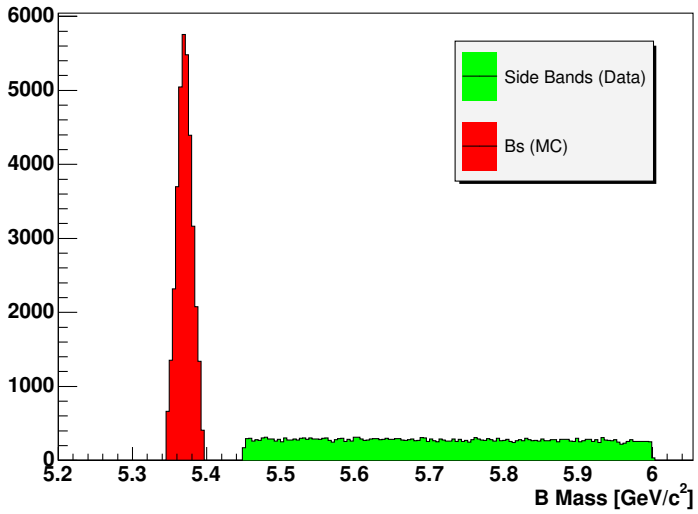

Figure 3: Invariant mass distribution showing the signal and background regions chosen for the Neural Network training

Input variables in the Neural Network			
Particle	Variable Definition	Variable Name	Input Number
B^0	transverse momentum	B p_T	1
	4 track vertex fit χ^2	B Fit χ^2	2
	proper decay length, ct=L_{xy}·Mass/p_T	B ct	3
	impact parameter	B d0	4
K_S	transverse momentum	K_S p_T	5
	invariant mass	K_S Mass	6
	proper decay length	K_S ct	7
	impact parameter	K_S d0	8
J/ψ	invariant mass	J/ψ Mass	9
	proper decay length	J/ψ ct	10
	impact parameter	J/ψ d0	11
	transverse momentum	J/ψ p_T	12
both π's	impact parameter	π_1 d0, π_2 d0	13,14
	transverse momentum	$\pi_1 p_T$, $\pi_2 p_T$	15,16
both μ's	transverse momentum	$\mu_1 p_T$, $\mu_2 p_T$	17,18
	impact parameter	π_1 d0, π_2 d0	19,20
	cosine of the helicity angle in J/ψ CM frame	$cos(\theta_{B,\mu_1})$,$cos(\theta_{B,\mu_2})$	21,22

Table 1: Variables used as input in the Neural Network training.

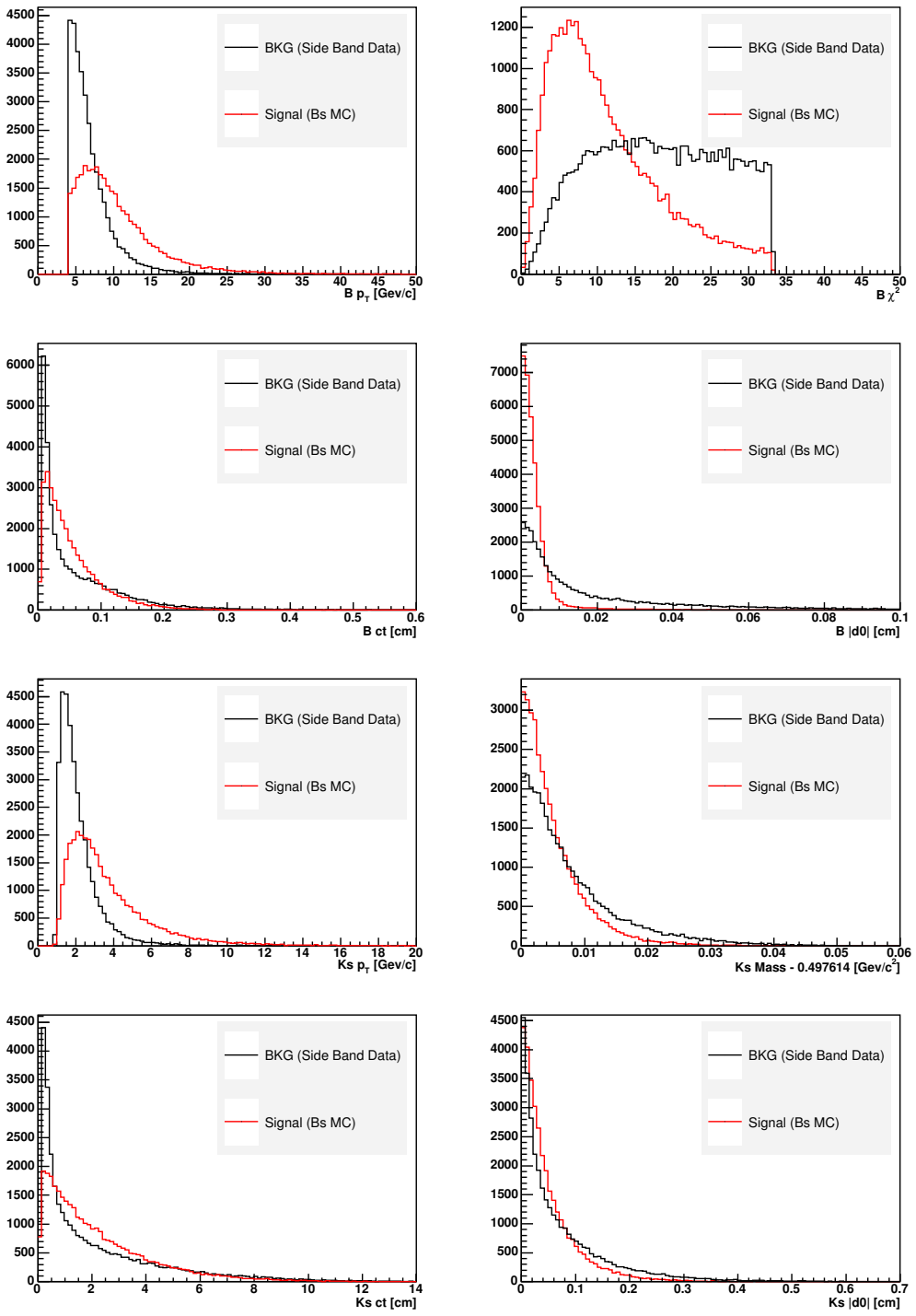

Figure 4: Distributions of the input variables to the Neural Network. The black histograms are sideband data and the red histograms are B_s signal MC.

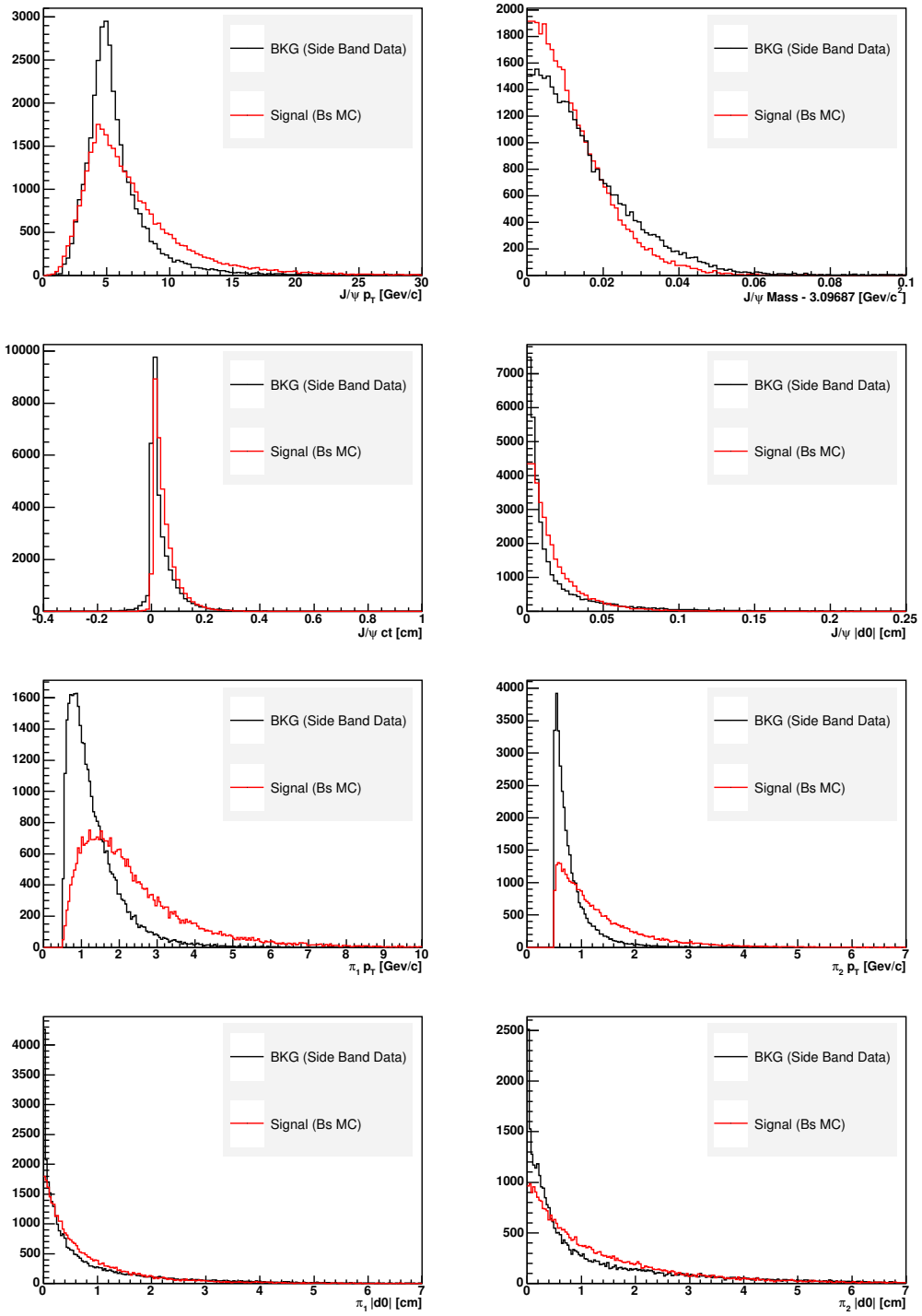

Figure 5: Distributions of the input variables to the Neural Network. The black histograms are sideband data and the red histograms are B_s signal MC.

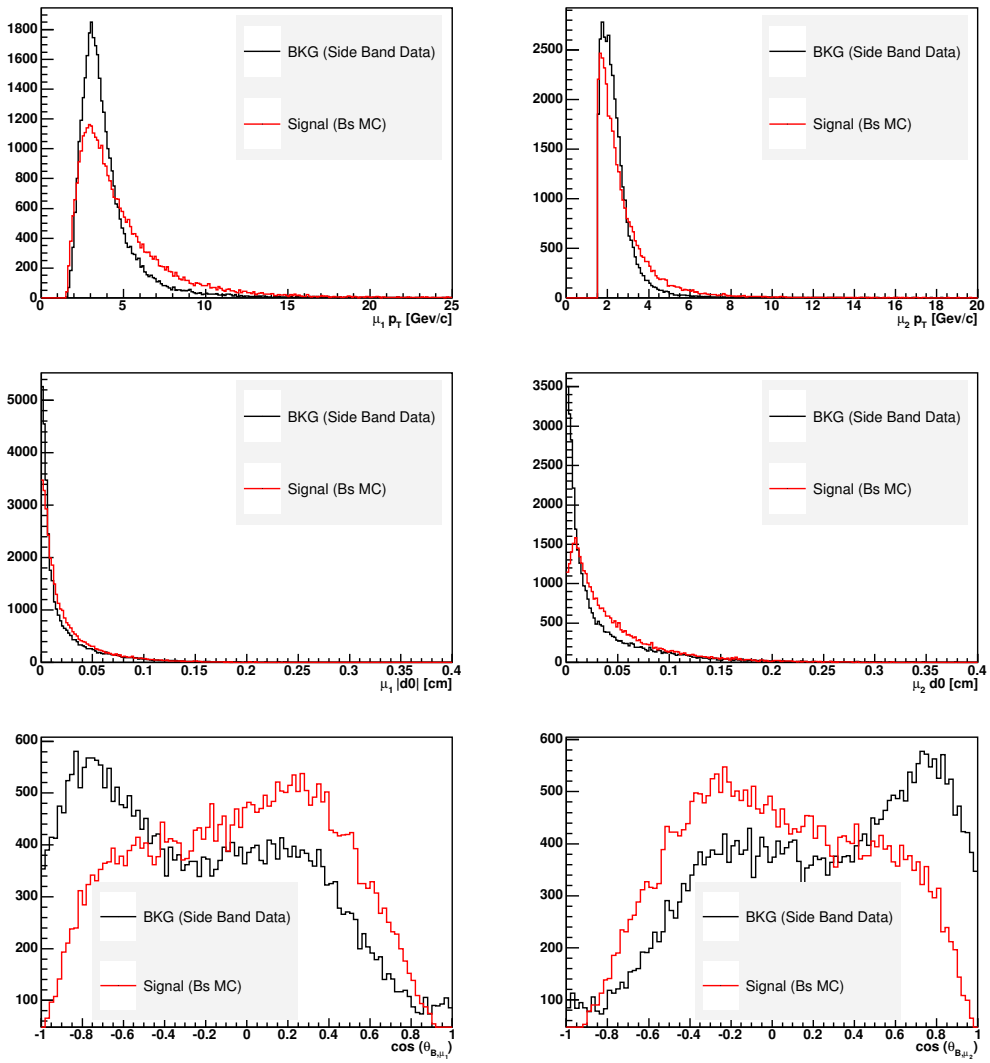

Figure 6: Distributions of the input variables to the Neural Network. The black histograms are sideband data and the red histograms are B_s signal MC.

The training is performed using the same number of signal and background events. The results of the Neural Network training (after 20 iterations) are shown in Figures 7 to 9. In particular, Figure 7 shows that the Neural Network achieves strong discrimination between signal and background. Table 2 shows the input variables list again, but this time sorted by significance. Appendix C summarizes different studies that have been carried out to verify that the Neural Network is not biased.

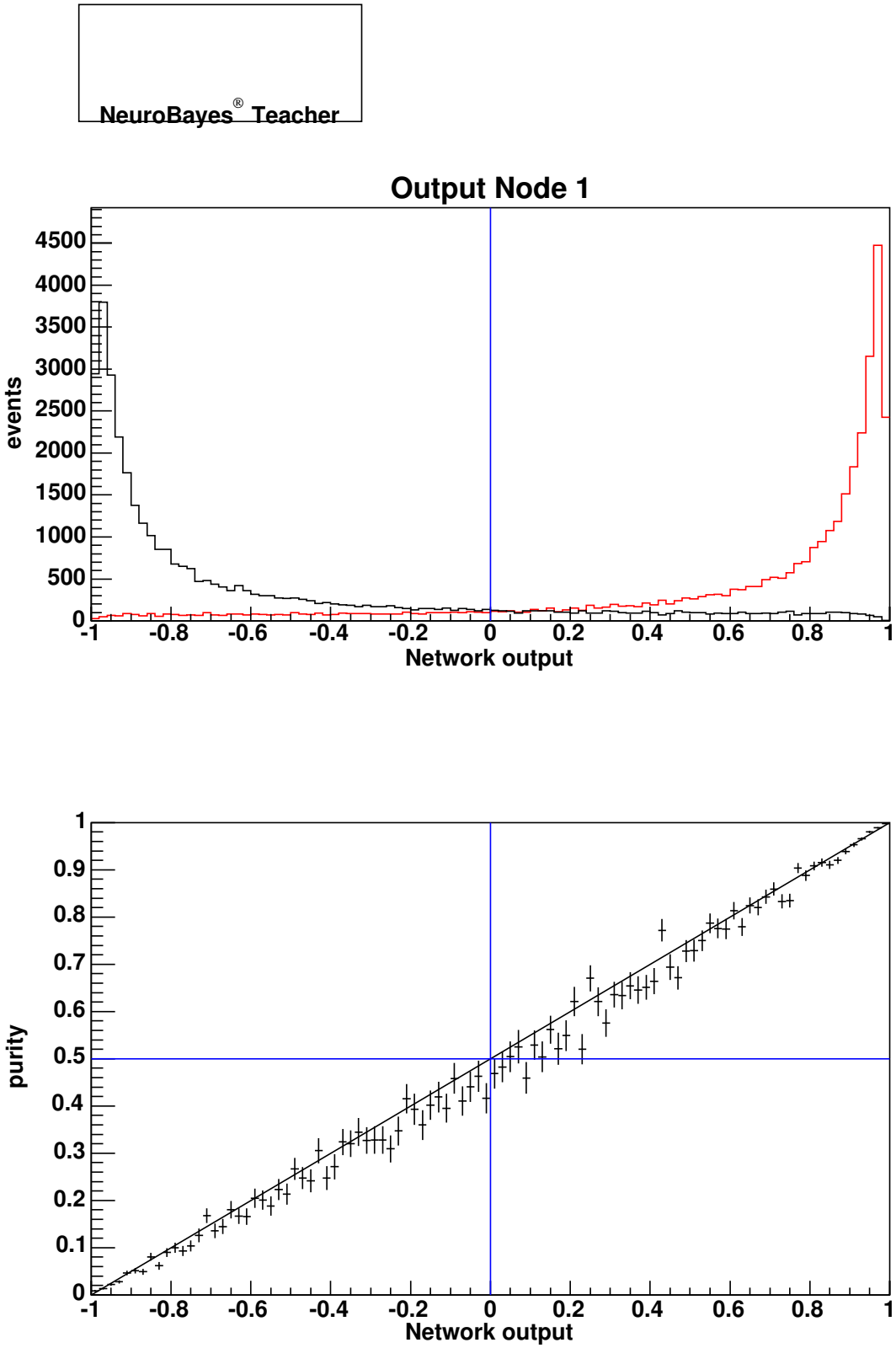

Figure 7: Neural Network output (top), where the red histogram is signal MC and the black one is sideband data, and purity (bottom).

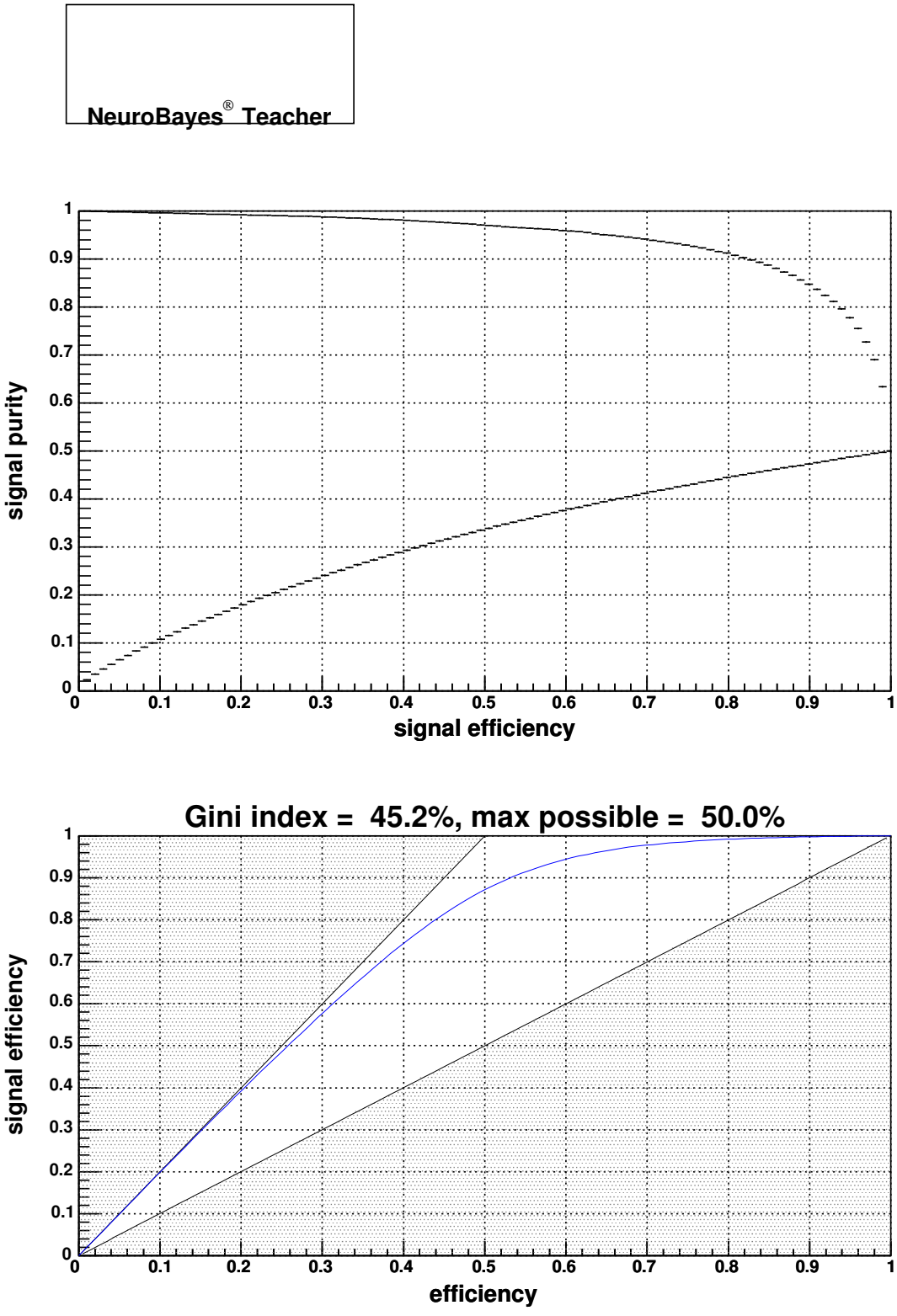

Figure 8: Signal Purity versus Signal Efficiency (top) and Signal Efficiency versus Efficiency (bottom).

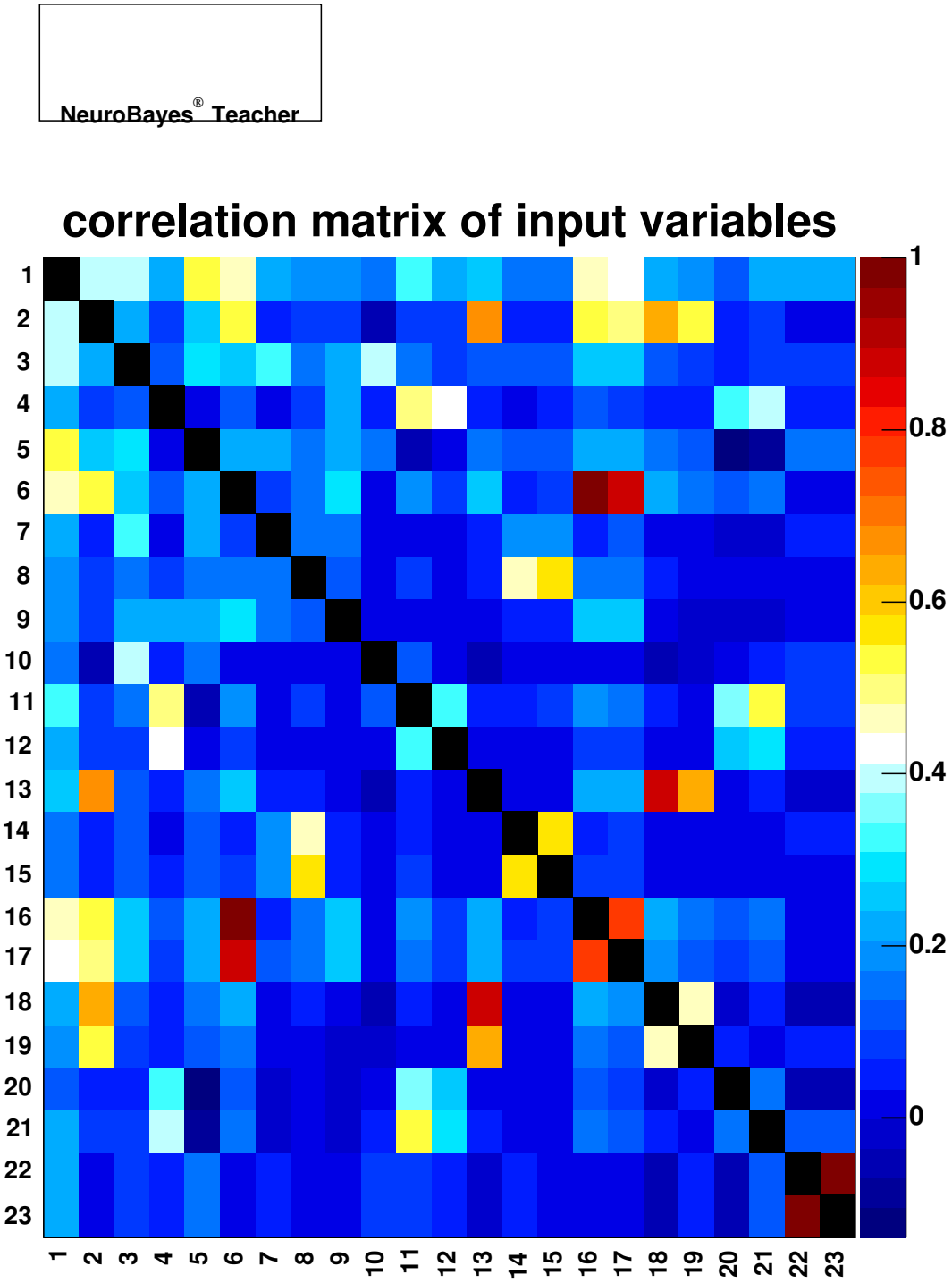

Figure 9: Correlation between the different input variables.The numbers in this figure are input variables as listed in Table 1. It is important to note that there are 23 variables instead of 22 because the first variable refers to the Neural Network output therefore everything is shifter one unit.

sorted by significance	variable name	additional significance	only this variable	loss when removed	global correlation to others (%)
1	B $\|d0\|$	119.70	119.70	80.49	47.9
2	K_S p_T	77.43	103.27	4.20	99.8
3	J/ψ ct	60.51	68.57	23.75	66.8
4	B Fit χ^2	35.55	89.76	25.35	58.7
5	$cos(\theta_{B,\mu_1})$	22.56	46.09	0.00	100.0
6	J/ψ $\|d0\|$	21.17	49.65	13.92	47.2
7	B p_T	19.10	88.25	8.96	82.1
8	μ_2 $\|d0\|$	15.82	51.87	15.96	59.6
9	π_1 $\|d0\|$	15.01	37.35	6.88	61.0
10	μ_1 $\|d0\|$	13.06	28.88	12.52	46.9
11	K_S Mass	9.75	49.04	9.48	42.3
12	K_S ct	6.59	44.29	4.89	61.5
13	μ_1 p_T	5.48	48.98	2.15	90.9
14	K_S $\|d0\|$	4.31	40.15	4.58	45.6
15	μ_2 p_T	2.98	41.62	1.97	71.2
16	π_2 p_T	2.88	92.46	1.56	98.4
17	π_2 $\|d0\|$	2.71	37.51	2.27	67.5
18	B ct	2.46	52.67	1.95	64.1
19	π_1 p_T	1.20	100.76	0.00	99.7
20	J/ψ Mass	0.41	33.50	0.00	45.5
21	J/ψ p_T	0.03	52.96	0.00	93.2
22	$cos(\theta_{B,\mu_2})$	0.00	45.89	0.00	100.0

Table 2: Variables used as input in the Neural Network training sorted by significance. The four last variables are not kept by the Neural Network.

2.4 Signal optimization

In order to choose a value for a cut on the Neural Network output, one needs to decide upon an appropriate figure of merit. The figure of merit chosen for optimization is $S/(1.5+\sqrt{B})$. This quantity is well accepted for signal discovery as described in [4]. For the signal sample $B_s \to J/\psi K_S$ MC is used. The events selected as signal, S, are the ones in the reconstructed mass range 5.35 GeV/c^2 $< M_B <$ 5.4 GeV/c^2. For the background sample, B, an upper sideband in the $B^0 \to J/\psi K_S$ reconstructed invariant mass plot is used. In this latter case, the range is 5.43 GeV/c^2 $< M_B <$ 5.48 GeV/c^2. Figure 10 shows $S/(1.5+\sqrt{B})$ versus the Neural Network output. This figure of merit suggests a cut in the Neural Network response of 0.88. Note that the points move around for Neural Network response>0.7 due to low statistics on the upper sideband. As illustrated in Figure 11 the Neural Network efficiency at this point is close to 45% for signal and about 1% for background.

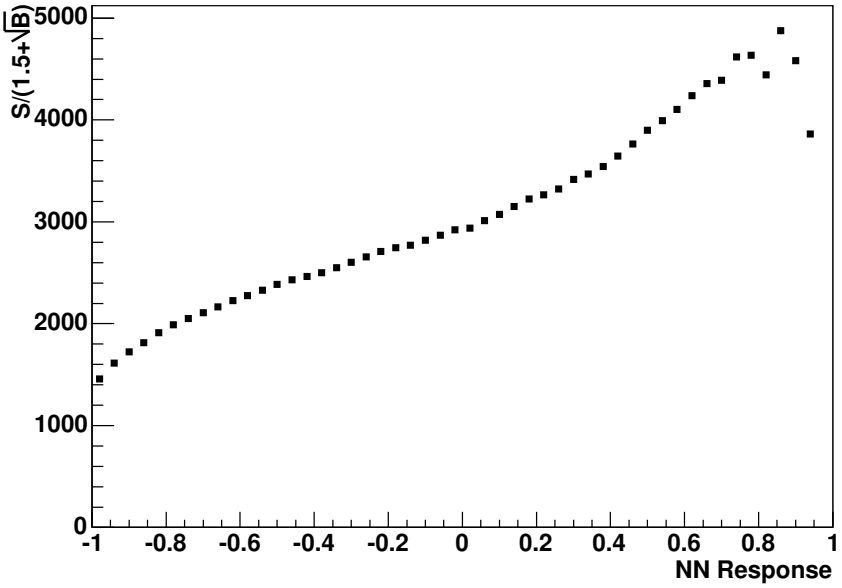

Figure 10: Figure of merit $S/(1.5+\sqrt{B})$ as a function of Neural Network response.

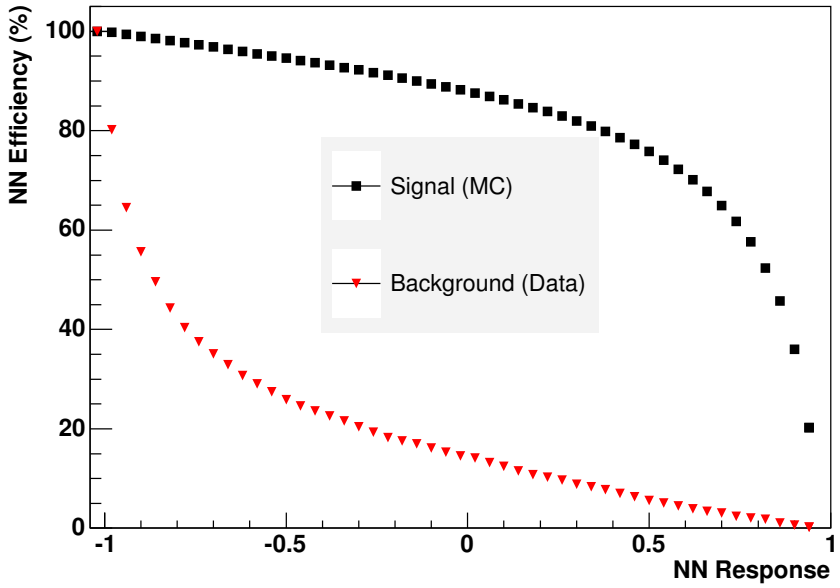

Figure 11: Neural Network efficiency versus Neural Network response.

The invariant B mass distribution is shown in Figure 12 before and after the optimization cut.

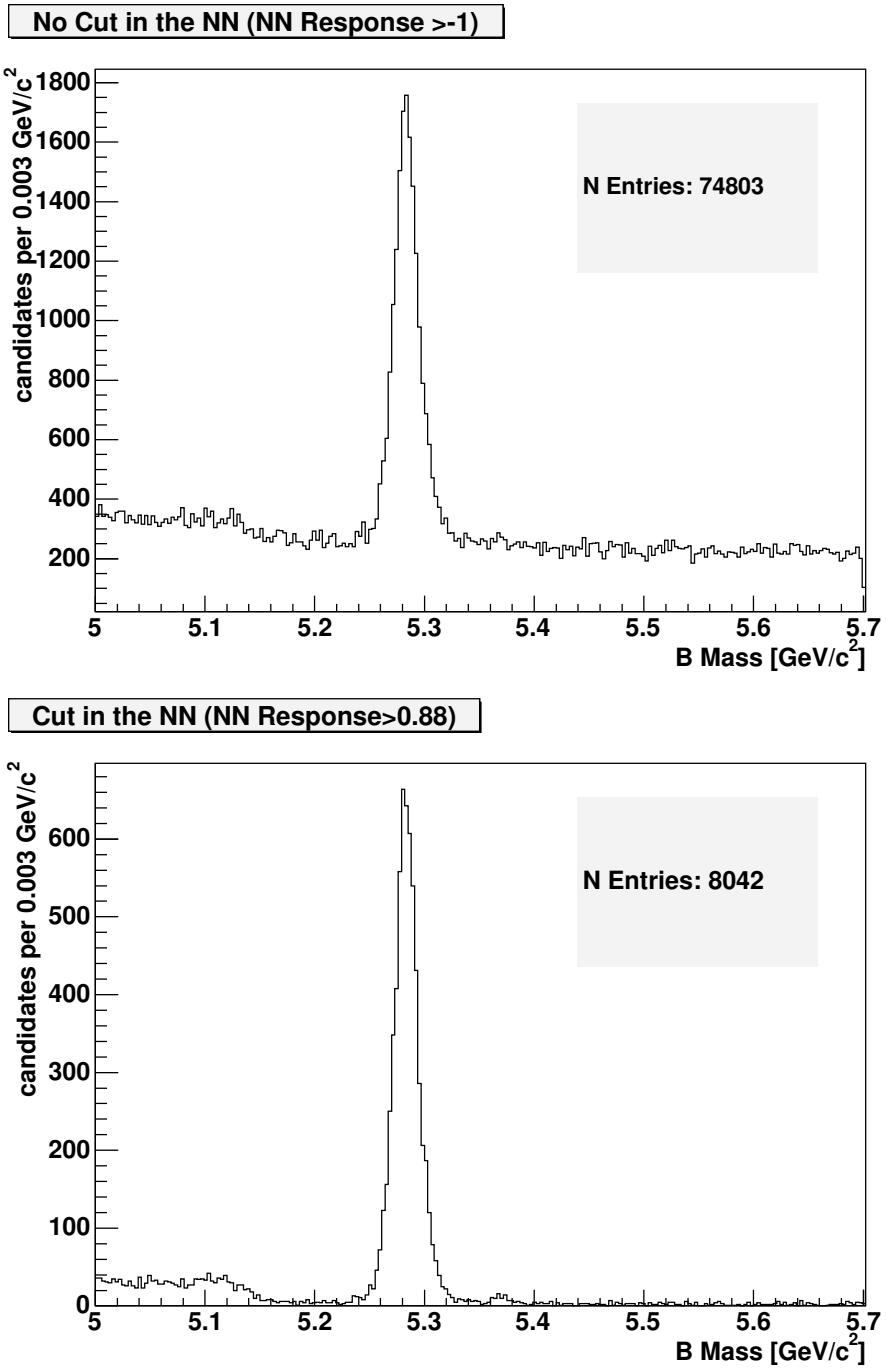

Figure 12: Invariant B mass distribution before (top) and after (bottom) the cut in the Neural Network response.

Figure 13 shows the invariant B mass distributions for different optimization cuts (from 0.72 to 0.96). In all the distributions one can clearly see the B_s peak, which proves that the analysis is not extremely sensitive to the Neural Network cut. In other words, the figure shows that the result is very robust.

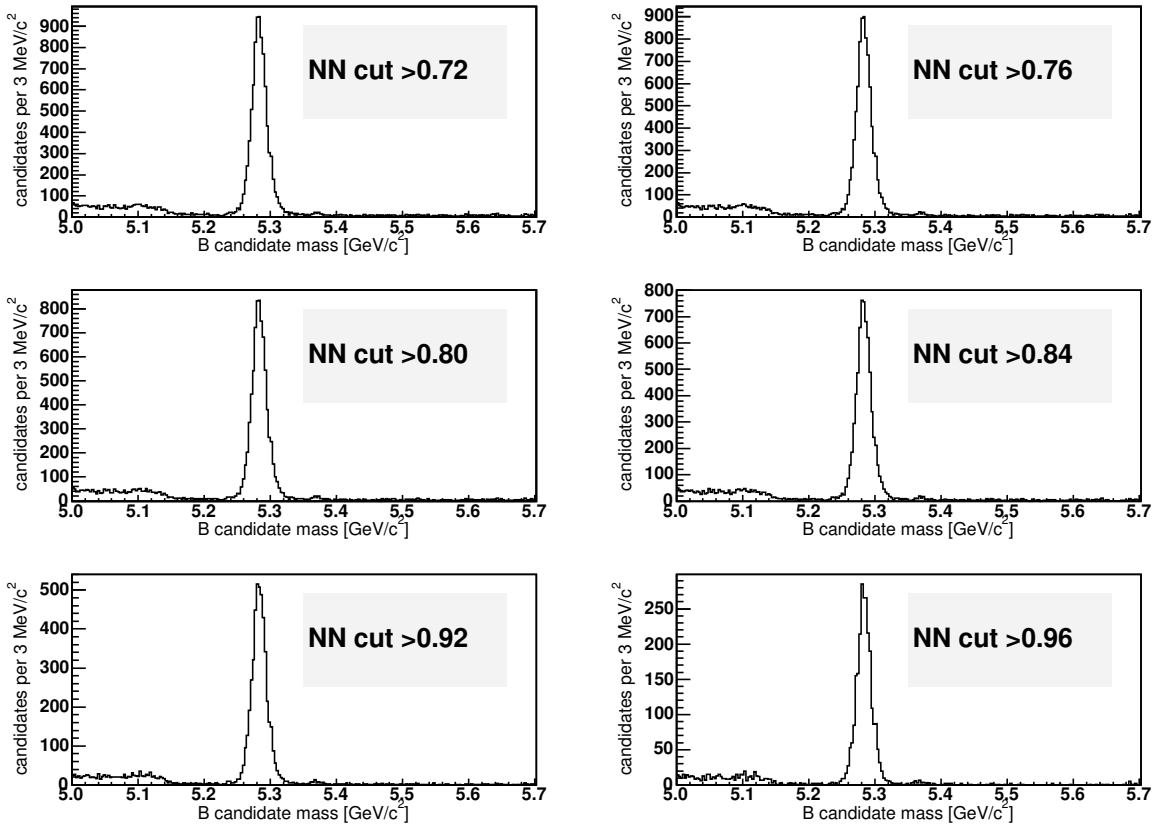

Figure 13: Invariant B mass distribution for different Neural Network response cuts.

2.5 Backgrounds, signals and Fit

For the purpose of extracting the yields of $B^0 \to J/\psi K_S$ and $B_s \to J/\psi K_S$ signals in the invariant mass distribution, an accurate modeling for signals and backgrounds is needed. This section summarizes the different signals and background contributions together with their modeling for the final fit.

2.5.1 Backgrounds

This analysis has the advantage that in the decay products there is a K_S. Because of the long K_S lifetime the final sample is cleaner than other analysis, such as $B_s \to J/\psi K^*$ decay [5]. But there are several backgrounds in the $B^0 \to J/\psi K_S$ sample that must be accounted for.

Λ_b background

As it was discussed in section 2.3, the $\Lambda_b \to J/\psi\Lambda$ contribution, where $\Lambda \to p\pi$, is a background in this analysis when the p is reconstructed as a π. A specific angular cut was designed to remove the majority of this contribution. To estimate the number of Λ_b remaining in the final invariant mass distribution, relative to the number of B^0, the following equation is used

$$\frac{N(\Lambda_b \to J/\psi\Lambda)}{N(B^0 \to J/\psi K_S)} = \frac{f_{\Lambda_b}}{f_d}\frac{Br(\Lambda_b \to J/\psi\Lambda)}{Br(B^0 \to J/\psi K_S)}\frac{1}{A_{rel}}, \tag{3}$$

where

$$A_{rel} = \frac{N(B^0 \to J/\psi K_S\ passed)/N(B^0 \to J/\psi K_S\ generated)}{N(\Lambda_b \to J/\psi\Lambda\ passed)/N(\Lambda_b \to J/\psi\Lambda\ generated)}. \tag{4}$$

The A_{rel} value calculated using MC is 13239. This number is big because the Λ_b acceptance is very small. The rest of values are extracted from a CDF analysis [6]: $\frac{f_{\Lambda_b}}{f_d} \cdot \frac{Br(\Lambda_b \to J/\psi\Lambda)}{Br(B^0 \to J/\psi K_S)} = 0.27$. Plugging all these values in Equation 3, $\frac{N(\Lambda_b \to J/\psi\Lambda)}{N(B^0 \to J/\psi K_S)}$ is calculated to be $2 \cdot 10^{-5}$. Therefore, this background contribution in the final invariant mass distribution is considered negligible.

Combinatorial background

We must take into account the combinatorial background contribution, resulting from B^0 being reconstructed from J/ψ candidates and two random tracks. This combinatorial background is modeled in the final fit with an exponential function,

$$f_{comb}(x) = N_0 \cdot e^{C_0 x}. \tag{5}$$

In the final fit, the fraction of combinatorial background events and the decay constant, C_0, are allowed to float.

Partially reconstructed background

A separate background that emerges is partially reconstructed B-hadrons where a five-body decay occurs and a π is not reconstructed. This background is fitted with an ARGUS function, see Equation 6, convoluted with a Gaussian of mean zero and width 12 MeV/c^2 [7]. In the ARGUS function N_1 is the normalization and m_0 is the mass cut off. All the parameters are allowed to float in the final fit.

$$f_{ARGUS}(x) = N_1 \cdot \sqrt{1 - \frac{x^2}{m_0^2}} \cdot e^{-C_1 \frac{x^2}{m_0^2}} \tag{6}$$

2.5.2 Signals

The signal contributions are modeled with three Gaussian template obtained from fits to B^0 MC. The relative contributions, means and widths from each Gaussian are fixed in the final fit. The B^0 MC fitted template and residuals appears in Figure 14 and the fit parameters in Table 3. The residual is defined as the bin content minus fit value divided by bin content uncertainty.

$$f_{B^0} = N_{B^0} \cdot \left(\frac{f_1}{\sigma_1\sqrt{2\pi}} e^{-(x-\mu_1)^2/2\sigma_1^2} + \frac{f_2}{\sigma_2\sqrt{2\pi}} e^{-(x-\mu_2)^2/2\sigma_2^2} + \frac{f_3}{\sigma_3\sqrt{2\pi}} e^{-(x-\mu_3)^2/2\sigma_3^2}\right) \tag{7}$$

Parameter	value
f_1 (1st contribution)	0.8054 ± 0.0005
μ_1 (1st mean)	5.2831 ± 0.0001 GeV/c^2
σ_1 (1st width)	(9.20 ± 0.05) · 10^{-3} GeV/c^2
f_2 (2nd contribution)	0.1899 ± 0.0005
μ_2 (2nd mean)	5.2873 ± 0.0003 GeV/c^2
σ_2 (2nd width)	(1.93 ± 0.02) · 10^{-2} GeV/c^2
f_3 (3rd contribution)	0.010 ±0.006
μ_3 (3rd mean)	5.289 ±0.007 GeV/c^2
σ_3 (3rd width)	(9.20 ± 0.54) · 10^{-2} GeV/c^2

Table 3: Template parameters for $B^0 \to J/\psi K_S$

The $B_s \to J/\psi K_S$ template used in the final fit is identical to $B^0 \to J/\psi K_S$, except for a shift of 86.8 MeV/c^2 in the mean value of the three Gaussians. This value corresponds

to the PDG mass difference between B_s and B^0. It is important to note that the MC generally underestimates the widths of the mass distribution. Therefore, the Gaussian widths of the two narrowest Gaussians are multiplied by a scale factor, s_{width}, which is allowed to float in the final fit. The scale factor is not applied to the third Gaussian since it is not expected to be governed by detector resolution effects as the other two. Moreover, a mass shift, m_{shift}, is added to the means of all Gaussians templates to account for a possible mass mismodeling in the MC.

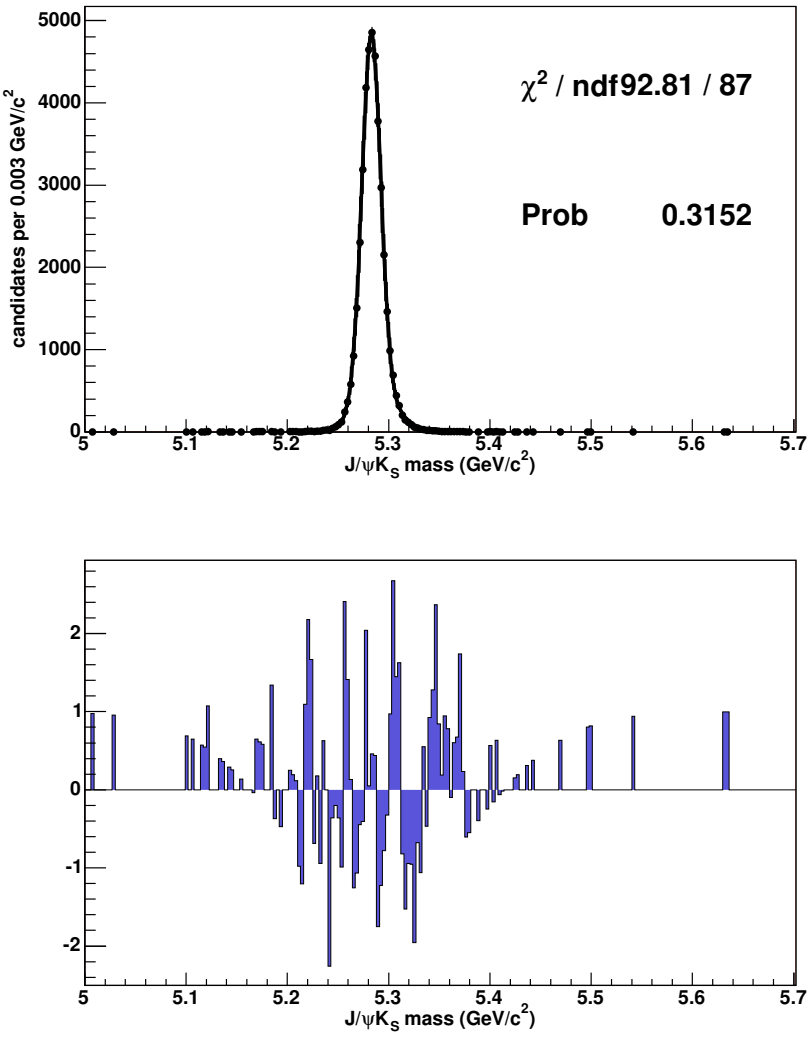

Figure 14: Invariant mass distribution fitted with three Gaussians (top) and residuals (bottom) of $B^0 \to J/\psi K_S$ signal candidates from MC.

2.5.3 Binned Log Likelihood Fit

A binned log likelihood fit is performed using the templates for $B^0 \to J/\psi K_S$ and $B_s \to J/\psi K_S$ signals and the functions described above. The symbolic form of the likelihood is written below:

$$\mathcal{L} = \prod_{i=1}^{Nevents} \mathcal{L}_i, \tag{8}$$

where $\mathcal{L}_i$ is the likelihood of the i^{th} event, and the index i runs over the events with $5.0 < m_{J/\Psi K_S} < 5.7$ GeV/c^2. $\mathcal{L}_i$ has the form:

$$\mathcal{L}_i = b \cdot \mathcal{L}^{bckg} + (1-b) \cdot \mathcal{L}^{sig} \tag{9}$$

where the index *sig* and *bckg* labels the part of the function that described the signal and background respectively; b is the fraction of the background events and $1-b$ is the fraction of B^0 and B_s events. The likelihood of the signal and background events, using the functions described before, can be written as $\mathcal{L}^{sig}() = f_{B^0} + f_{B^s}$ and $\mathcal{L}^{bckg}() = f_{comb} + f_{ARGUS}$.

In the final fit there are 27 parameters of which 19 are fixed. In addition to the parameters mentioned above, the total $N_{candidates}$ from all contributions is fixed to the number of entries in the histogram. The seven parameters allowed to float are the following:

- $f_{N(B^0)+N(B_s)}$: Fraction of candidates that are $B^0 \to J/\psi K_S$ and $B_s \to J/\psi K_S$,
- $\frac{N_{B_s}}{N_{B^0}}$: Ratio of the number of $B_s \to J/\psi K_S$ to $B^0 \to J/\psi K_S$,
- $f_{comb.bkg.}$: Fraction of candidates that are combinatorial background,
- C_0: The exponential decay constant for the combinatorial background modeling,
- C_1: The exponential decay constant for the ARGUS function,
- m_0: Mass cut off for the ARGUS function,
- s_{width}: The Gaussian scale factor,
- m_{shift}: The mass shift added to the means of all Gaussian templates.

The mass distribution in data, the final fit, and the residuals appear in Figure 15. Table 4 shows the values of all the floating parameters after the fit.
Figure 16 shows the final fit including the different contributions. An expanded view of the fit is shown in Figure 17. A plot of the residual population fitted with a Gaussian is in Figure 18. In addition, Table 5 shows the correlation matrix for the fit.

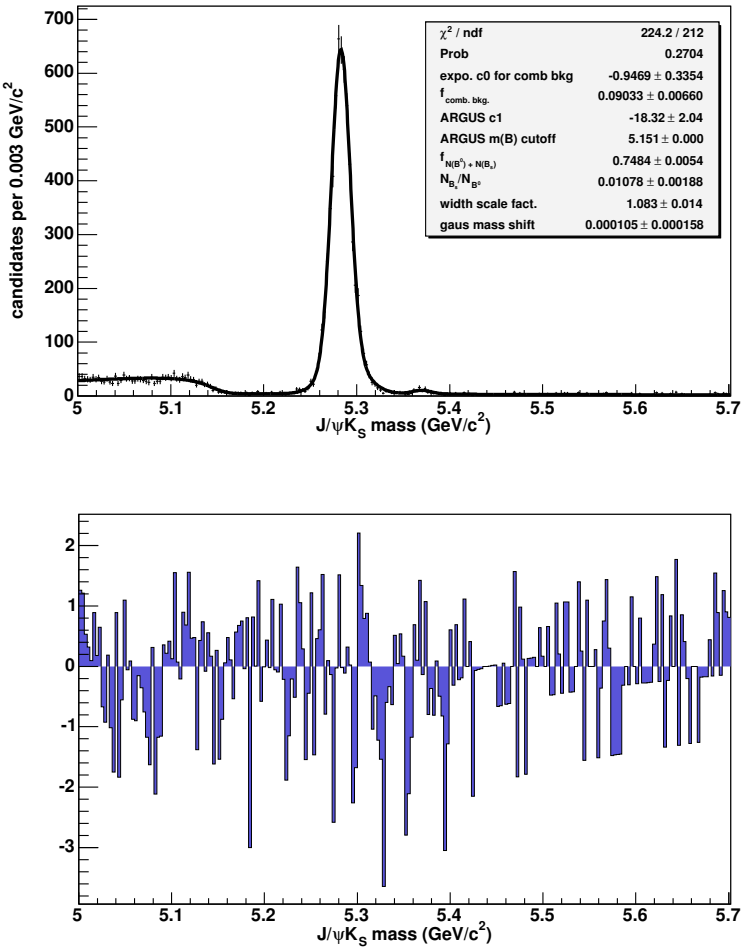

Figure 15: Mass distribution overlayed with the binned likelihood fit (top) and the residuals (bottom).

Parameter	value
$f_{N(B^0)+N(B_s)}$	0.748 ± 0.005
$\frac{N_{B_s}}{N_{B^0}}$	0.0108 ± 0.0019
$f_{comb.bkg.}$	0.090 ± 0.007
C_0	-0.95 ± 0.34 GeV/c^2
C_1	-18.4 ± 2.1 GeV/c^2
m_0	5.15070 ± 0.00004 GeV/c^2
s_{width}	1.08 ± 0.01
m_{shift}	(1 ± 2)· 10^{-4} GeV/c^2

Table 4: Values of the floating parameters after the fit.

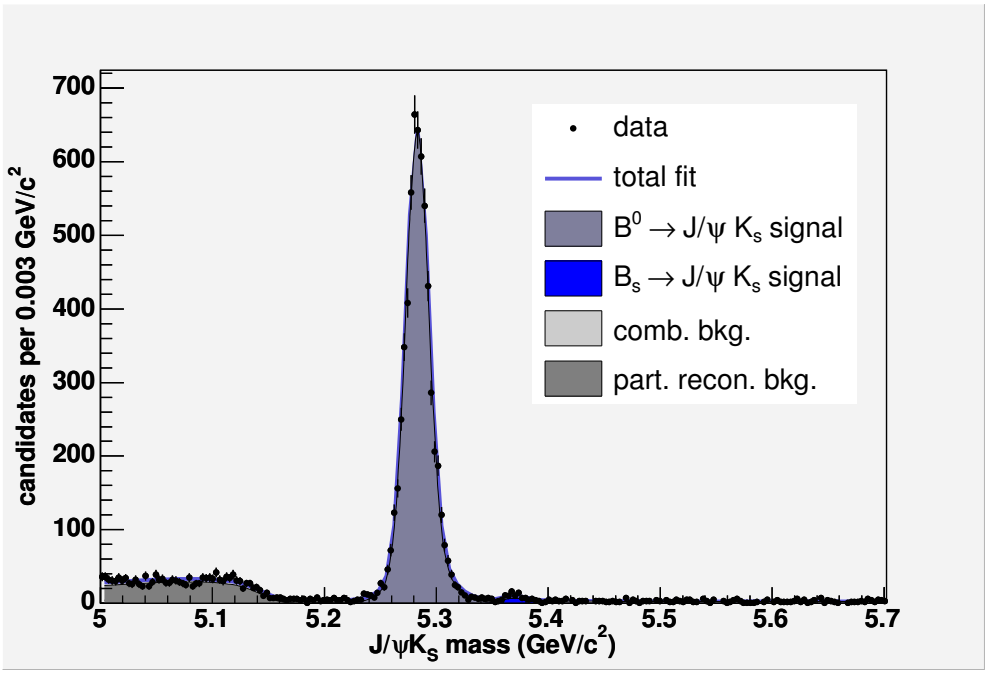

Figure 16: Mass distribution and fit.

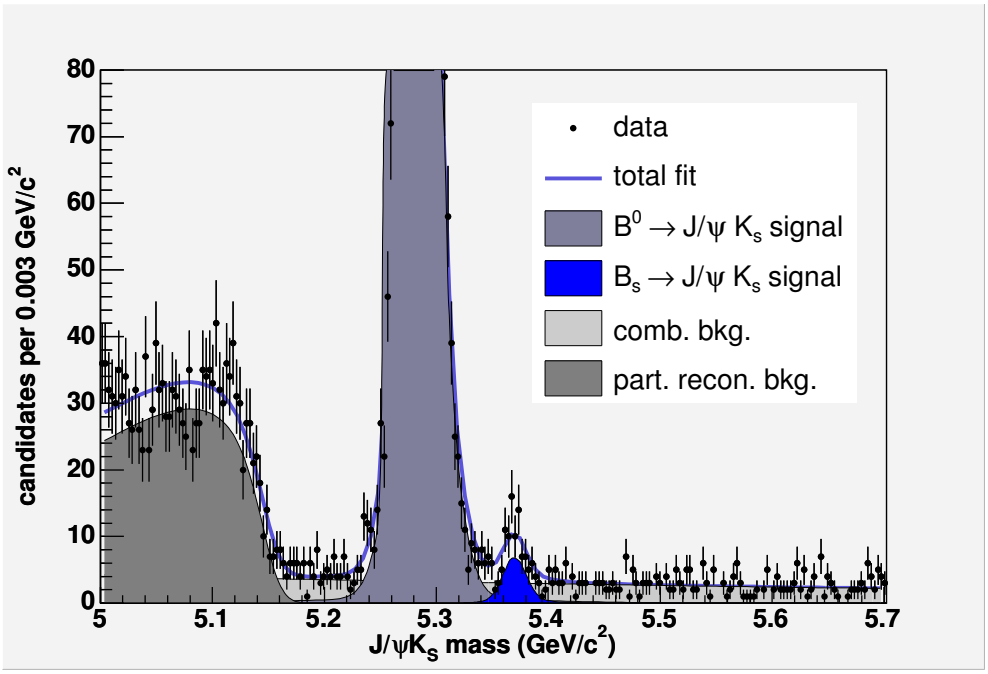

Figure 17: Mass distribution and fit enlarged in the signal region.

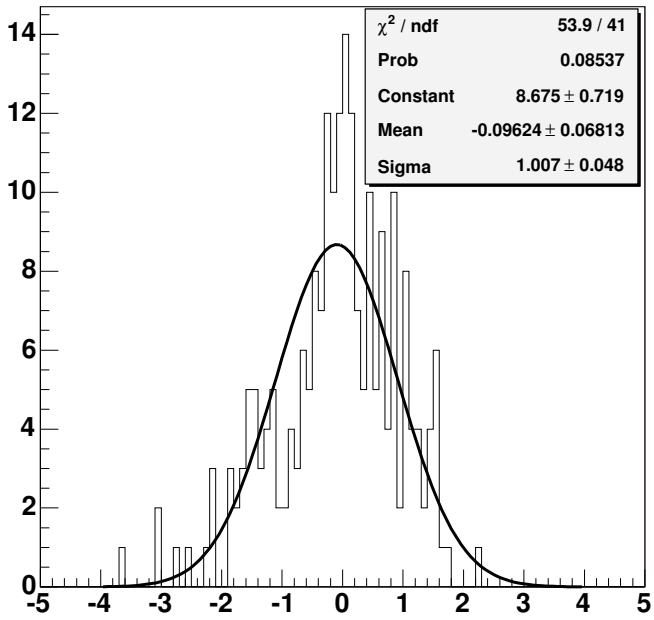

Figure 18: Residual population overlayed with a Gaussian fit

	C_0	$f_{comb.bkg}$	C_1	m_0	$f_{N(B^0)+N(B_s)}$	$N(B_s)/N(B^0)$	s_{width}	m_{shift}
C_0	1	-0.67	0.03	-0.12	0.22	0.12	0.13	0.03
$f_{comb.bkg}$	-0.67	1	0.03	-0.12	-0.53	-0.21	-0.22	0.004
C_1	0.03	0.03	1	-0.22	0.02	0.009	-0.04	0.03
m_0	-0.12	-0.12	-0.22	1	-0.14	0.11	0.24	-0.17
$f_{N(B^0)+N(B_s)}$	0.22	-0.53	0.02	-0.14	1	0.16	0.09	0.02
$N(B_s)/N(B^0)$	0.12	0.21	0.009	0.11	0.16	1	0.01	0.02
s_{width}	0.13	-0.22	-0.04	0.24	0.09	0.01	1	-0.11
m_{shift}	0.03	0.004	0.03	-0.17	0.02	0.02	-0.11	1

Table 5: Correlation matrix for the fit.

A toy MC test of the fitter has been done. Figure 19 show the $N(B_s)/N(B^0)$ with 1000 toys and the pull distribution. Results from the toys are compatible with the fit results.

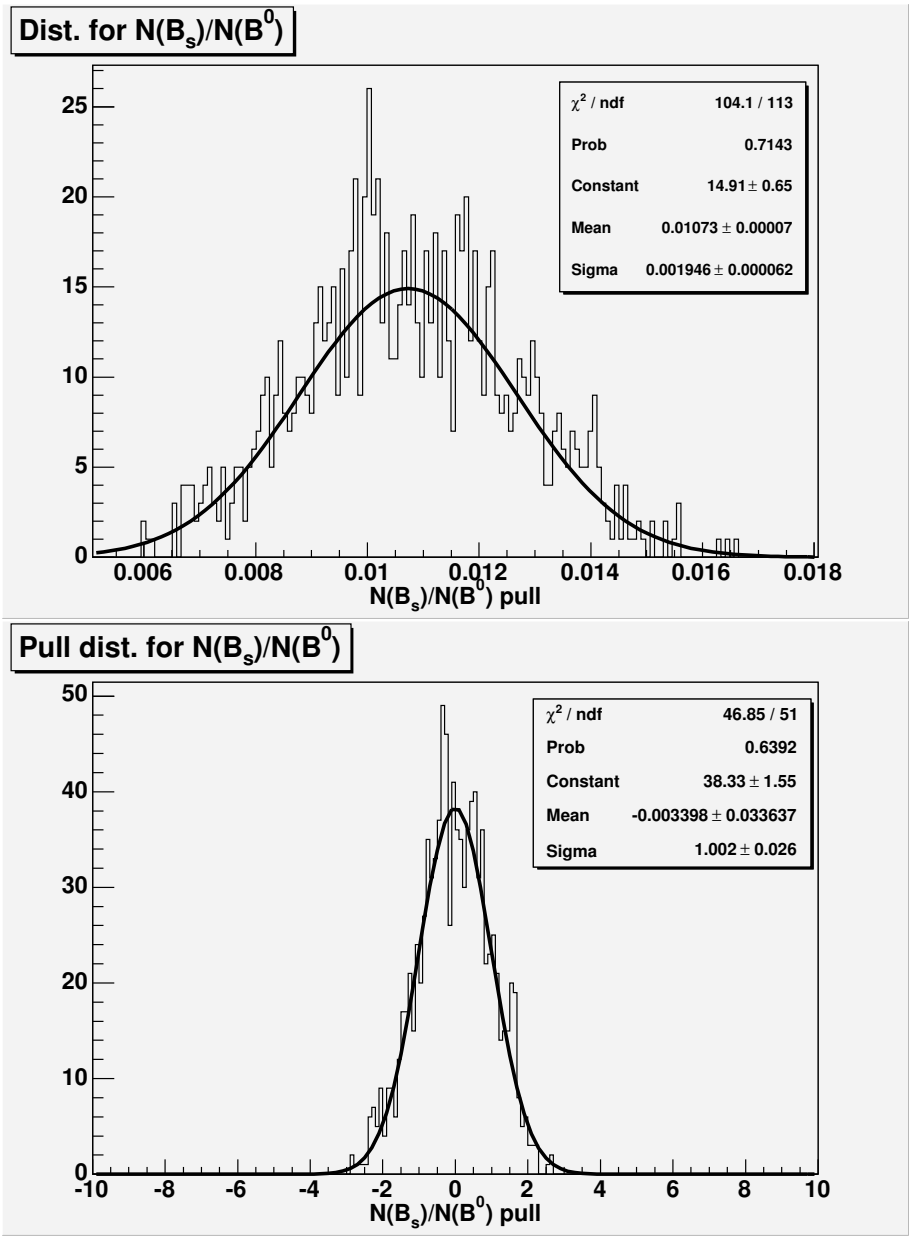

Figure 19: $N(B_s)/N(B^0)$ values from the 1000 toys (top) and Pull distribution for the $N(B_s)/N(B^0)$ variable (bottom).

In Figure 20 appears the likelihood profile close to the minimum. For the scan, the fit is repeated several times after fixing the parameter $N(B_s)/N(B^0)$ to different values. A standard fit check, using MINOS technique, has been done. Using this technique, the value and positive (negative) uncertainty of the parameter $N(B_s)/N(B^0)$ are 0.0108 +0.00195 (-0.00183).

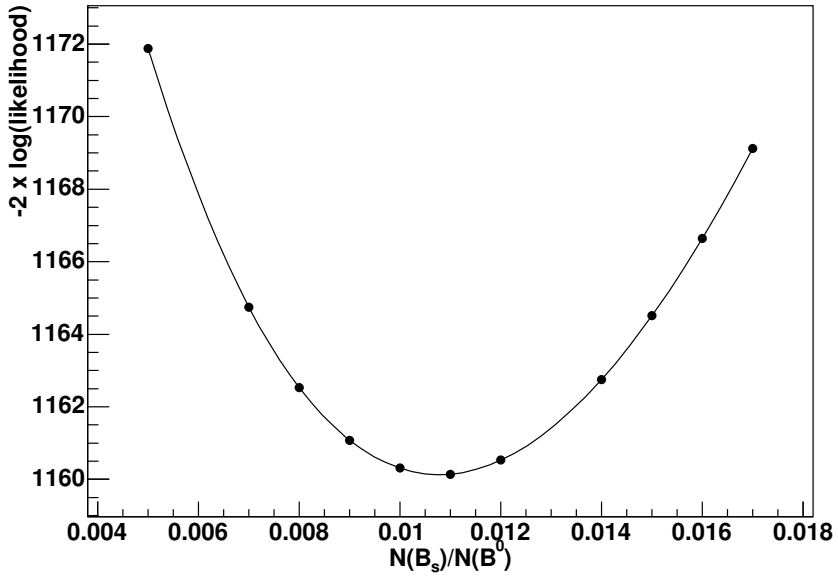

Figure 20: A scan of -2 x log(likelihood) for the fit parameter $N(B_s)/N(B^0)$.

The yields of the $B^0 \to J/\psi K_S$ and $B_s \to J/\psi K_S$ signal are determined to be 5954 $\pm$ 79 and 64 $\pm$ 14, respectively. The value of ratio of $N(B_s \to J/\psi K_S)/N(B^0 \to J/\psi K_S)$ extracted from the fit is 0.0108 $\pm$ 0.0019 (stat.).

2.5.4 Fit fixing the combinatorial background contribution

In the $B_s \to J/\psi K^*$ analysis [5], an excess was observed in the B mass region in the range 5.12 GeV/c^2< M_B <5.25 GeV/c^2 when the combinatorial contribution was fixed before the final fit. With the purpose of check if this phenomenon also occurs in this analysis, a fit fixing the combinatorial contribution is done.
The C^0 parameter and the fraction of combinatorial background are fixed after fitting the invariant mass contribution, over the range 5.4 GeV/c^2< M_B < 5.7 GeV/c^2, with an exponential function. Figure 21 shows the result of the fit and Table 6 summarizes the values of combinatorial background parametrization.

Parameter	value
N_0	643 $\pm$ 103
C_0	0.29 $\pm$ 0.71 (GeV/c^2)$^{-1}$

Table 6: Values of combinatorial background parametrization

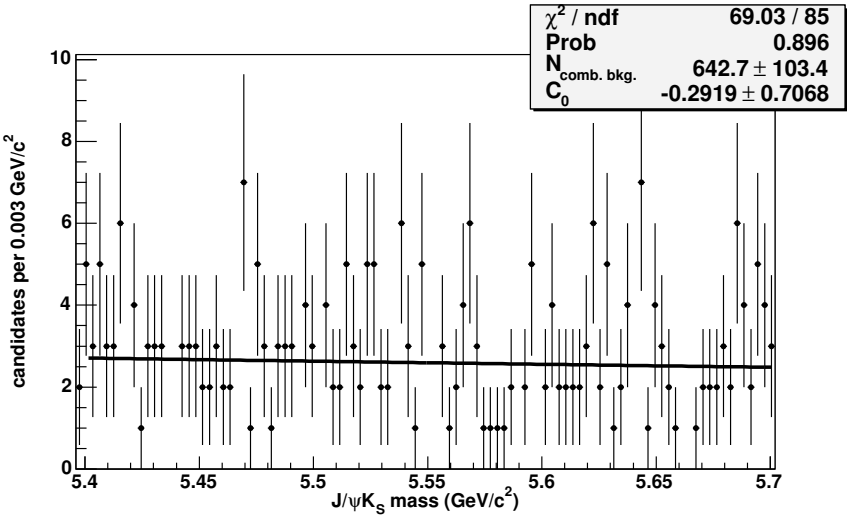

Figure 21: Fit over the range 5.4 GeV/c^2< M_B < 5.7 GeV/c^2 in the mass distribution to determine the shape and contribution of the combinatorial background modeled with an exponential function.

As done previously, a binned log likelihood fit is performed using the templates for signals, an exponential for the combinatorial background and the ARGUS function for the partially reconstructed contributions. As shown in Figures 22, 23 and 24 no excess is observed at the mass range 5.12 GeV/c^2< M_B < 5.25 GeV/c^2. Table 7 shows the values of all the floating parameters after the fit. In this case, the value of the ratio of $N(B_s \to J/\psi K_S)/N(B^0 \to J/\psi K_S)$ is 0.0114 ± 0.0018 (stat.), a variation of 0.0006 with respect to the default fit. This is included as a systematic uncertainty.

Parameter	value
$f_{N(B^0)+N(B_s)}$	0.752 ± 0.004
$\frac{N_{B_s}}{N_{B^0}}$	0.011 ± 0.002
C_1	-17.2 ± 1.9 GeV/c^2
m_0	5.15240 ± 0.00003 GeV/c^2
s_{width}	1.09 ± 0.01
m_{shift}	(1 ± 2)· 10^{-4} GeV/c^2

Table 7: Values of the floating parameters after the fit where the combinatorial background contribution is fixed.

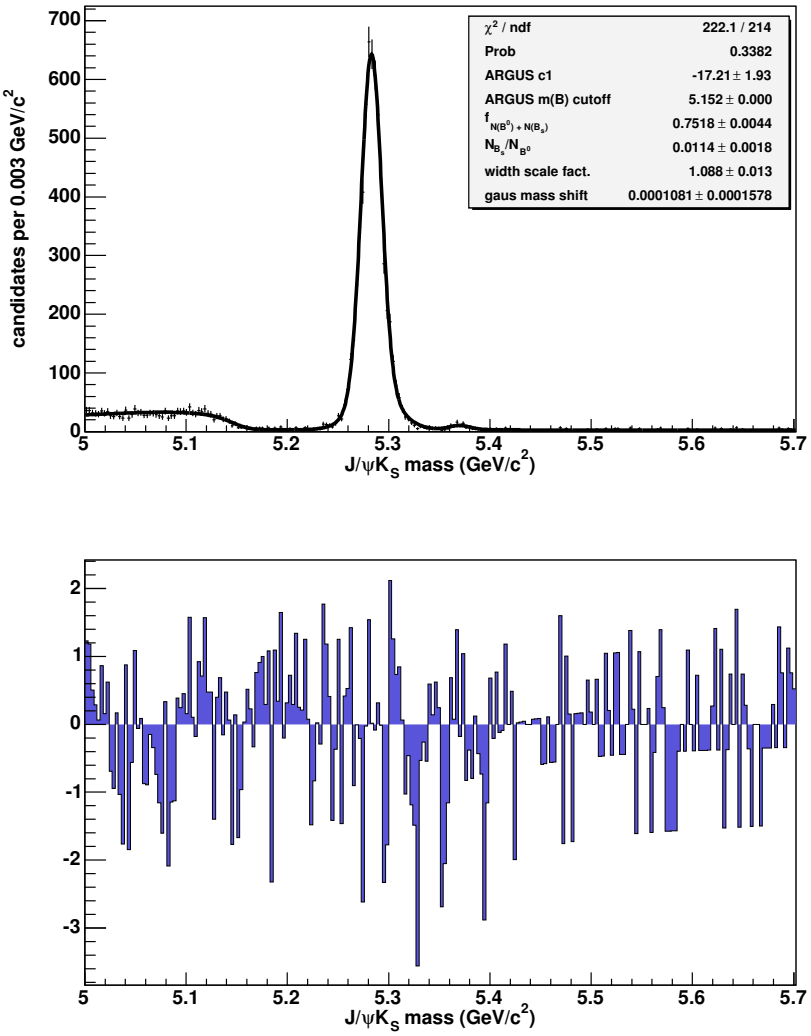

Figure 22: Mass distribution and fit with the combinatorial background contribution fixed (top) and the residuals (bottom).

2.6 Statistical significance of the $B_s \to J/\psi K_S$

The statistical significance of the $B_s \to J/\psi K_S$ signal is determined by fitting the mass distribution without the $B_s \to J/\psi K_S$ contribution. The $\Delta log(likelihood)$ between the two hypotheses is 52.7. With the amount of $B_s \to J/\psi K_S$ being the only degree of freedom separating the two hypothesis, a $\Delta\chi^2$ distribution gives a p-value of $3.85{\cdot}10^{-13}$ or 7.2 σ.

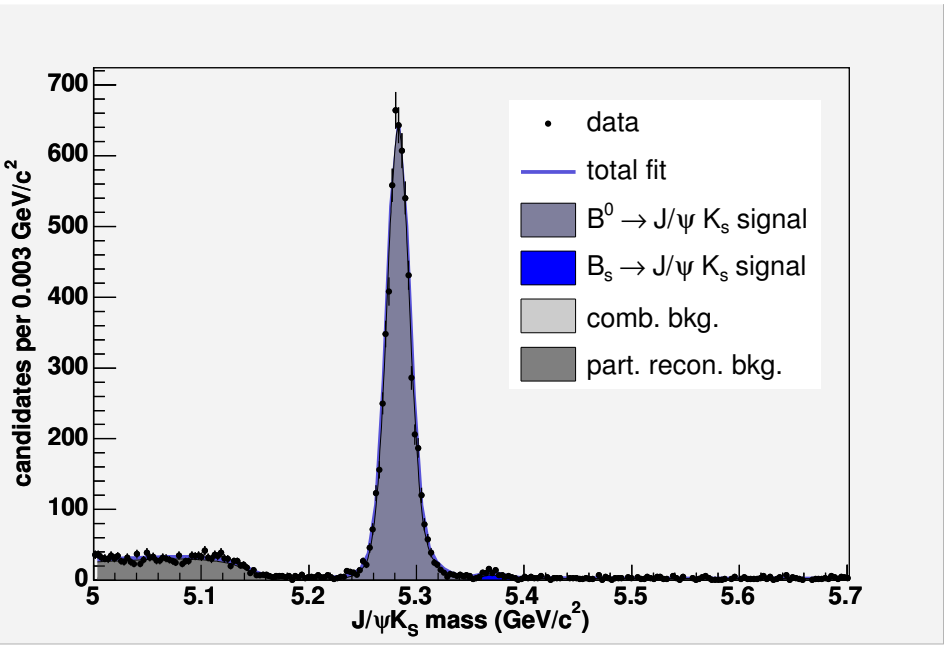

Figure 23: Mass distribution and fit with the combinatorial background contribution fixed.

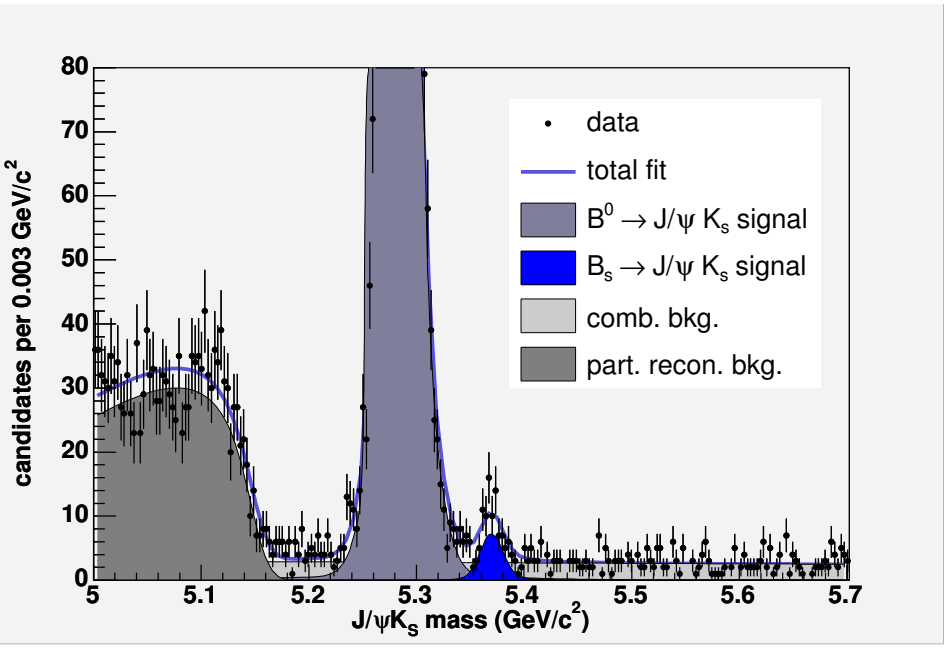

Figure 24: Mass distribution and fit with the background contribution fixed enlarged in the signal region.

2.7 Systematic uncertainties

The different sources of systematic uncertainties, which can influence the measured ratio of $N(B_s \to J/\psi K_S)/N(B^0 \to J/\psi K_S)$, have been studied.

- Modeling of the B^0 and B_s signals: The modeling of the B^0 and B_s signal peaks can influence the measurement of the ratio. The widths of the three Gaussians are allowed to float in the final fit. This change results in a shift of 0.0005 in $N(B_s \to J/\psi K_S)/N(B^0 \to J/\psi K_S)$. Therefore, a systematic uncertainty of $\pm$ 0.0005 is assigned to the ratio value to account for the signal model.

- Modeling of the combinatorial background: The shape of the combinatorial background is another source of systematic uncertainty to consider. In this case, a first order polynomial is used instead of an exponential. An additional systematic uncertainty of ± 0.0006 is included in the final measurement to take into account this effect. An additional systematic uncertainty of ± 0.0006 is included to take into account the difference coming from fixing this contribution before the fit.

- Mass difference between B_s and B^0: The PDG mass difference between B_s and B^0, 86.9 MeV/c^2, was used in the B_s template. This value has a ± 0.7 MeV/c^2 uncertainty. Therefore, the mass shift was varied within its uncertainty leading to an average tiny change in the ratio of $1.3 \cdot 10^{-5}$.

The different contributions are added in quadrature resulting in a total systematic uncertainty of ± 0.0010 for $N(B_s \to J/\psi K_S)/N(B^0 \to J/\psi K_S)$.

3 Results

The final measured value of the ratio is

$$N(B_s \to J/\psi K_S)/N(B^0 \to J/\psi K_S) = 0.0108 \pm 0.0019(stat.) \pm 0.0010(sys.).$$

This measurement can be used to determine the value of the ratio of branching fractions of $B_s \to J/\psi K_S$ to $B^0 \to J/\psi K_S$ using Equation 1.
The first step to determine the $Br(B_s \to J/\psi K_S)/Br(B^0 \to J/\psi K_S)$ is to calculate the relative acceptance of $B^0 \to J/\psi K_S$ to $B_s \to J/\psi K_S$. A MC sample of $B^0 \to J/\psi K_S$ and $B_s \to J/\psi K_S$ is used to extract A_{rel} as follows

$$A_{rel} = \frac{N(B^0 \to J/\psi K_S \; passed)/N(B^0 \to J/\psi K_S \; generated)}{N(B_s \to J/\psi K_S \; passed)/N(B_s \to J/\psi K_S \; generated)}, \tag{10}$$

where the number of *passed* candidates is simply the number that passed the event selection criteria described in section 2.3 and the number of *generated* is number of candidates generated by the MC. The value is determined to be $A_{rel} = 1.012 \pm 0.010$,

assuming binomial statistics to calculate the uncertainty. The uncertainty on the acceptances was then propagated through using Gaussian uncertainties for A_{rel} and added in as a systematic uncertainty for $Br(B_s \to J/\psi K_S)/Br(B^0 \to J/\psi K_S)$.
Different sources of systematics uncertainties for A_{rel} are being evaluated.

- Lifetime used for the B^0 and B_s MC: Different MC samples have been generated modifying the lifetime 1σ up and down to their PDG values. The lifetime values used are the following: $c\tau_{B^0}$(default)=457.5 μm, $c\tau_{B^0}(+1\sigma)$=460 μm, $c\tau_{B^0}(-1\sigma)$=454.8 μm, $c\tau_{B_s}$(default)=463 μm, $c\tau_{B_s}(+1\sigma)$=481 μm and $c\tau_{B_s}(-1\sigma)$=445 μm. The acceptance has been calculated nine different times taking into account all the variations. The maximum deviation is 0.028 and this value has been taken has a systematic uncertainty.

- p_T spectrum used for the B^0 and B_s MC: The default MC samples are generated using a p_T spectrum which from the NLO prediction calculation [8]. Additional samples are produced using the a p_T spectrum measured in $B \to J/\psi X$. The value of the A_{rel} varies 0.032 using these additional samples and this is added as systematic uncertainty.

With the values of A_{rel}, the measurement of the $f_s Br(B_s \to J/\psi K_S)/f_d Br(B^0 \to J/\psi K_S)$ is made to be:

$$f_s Br(B_s \to J/\psi K_S)/f_d Br(B^0 \to J/\psi K_S) = 0.0109 \pm 0.0019(stat.) \pm 0.0011(sys.)$$

To determine f_s/f_d , the most recent CDF measurement [9] of $f_s/(f_u+f_d)\cdot Br(D_s \to \phi\pi)$ is combined with the actual PDG value for $Br(D_s \to \phi\pi)$. With the input of $f_s/f_d = 0.269 \pm 0.033$, the ratio of branching fractions to the reference B^0 decays are:

$$Br(B_s \to J/\psi K_S)/Br(B^0 \to J/\psi K_S) = 0.041 \pm 0.007(stat.) \pm 0.004(sys.) \pm 0.005(frag.)$$

The PDG values for $Br(B_s \to J/\psi K_S)$ is used to calculate the absolute branching fractions:

$$Br(B_s \to J/\psi K^0) = (3.5 \pm 0.6(stat.) \pm 0.4(sys.) \pm 0.4(frag.) \pm 0.1(PDG)) \cdot 10^{-5}$$

Table 8 summarizes all the values of the quantities that go into the final result together with the final results themselves.

4 Summary

A sample of reconstructed $B^0 \to J/\psi K_S$ collected via the di-muon trigger is used to measure the ratio of branching fractions of $B_s \to J/\psi K_S$ to $B^0 \to J/\psi K_S$. This ratio is measured to be 0.0108 ± 0.0019 (stat.) ± 0.0010 (sys.) allowing a determination of $Br(B_s \to J/\psi K_S)/Br(B^0 \to J/\psi K_S)$ = 0.041 ± 0.007 (stat.) ± 0.004(sys.) ± 0.005(frag.)

N(B_s)	64 ± 14
N(B^0)	5954 ± 79
$\frac{N(B_s \to J/\psi K_S)}{N(B^0 \to J/\psi K_S)}$	0.0108 ± 0.0019 (stat.) ± 0.0010 (sys.)
A_{rel}	1.012 ± 0.010 (stat.) ± 0.042 (sys.)
$\frac{f_s Br(B_s \to J/\psi K_S)}{f_d Br(B^0 \to J/\psi K_S)}$	0.0109 ± 0.0019 (stat.) ± 0.0011 (sys.)
f_s/f_d (from CDF measurement)	0.269 ± 0.033
$\frac{Br(B_s \to J/\psi K_S)}{Br(B^0 \to J/\psi K_S)}$	0.041 ± 0.007 (stat.) ± 0.004 (sys.) ± 0.005 (frag.)
$Br(B^0 \to J/\psi K^0)$ (from PDG)	(8.71 ± 0.32)· 10^{-4}
$Br(B_S \to J/\psi K^0)$	(3.5± 0.6 (stat.) ± 0.4 (sys.) ± 0.4 (frag.) ± 0.1 (PDG)) · 10^{-5}

Table 8: All the quantities that go into the final result together with the final results themselves.

A Studies to remove Λ_b background

To remove the Λ_b background a different was tested. In this case a cut on the reconstructed K_S candidate mass using the Λ hypothesis is applied. Figure 25 shows the mass distribution of Λ candidates reconstructed from B_s MC and Λ_b MC.

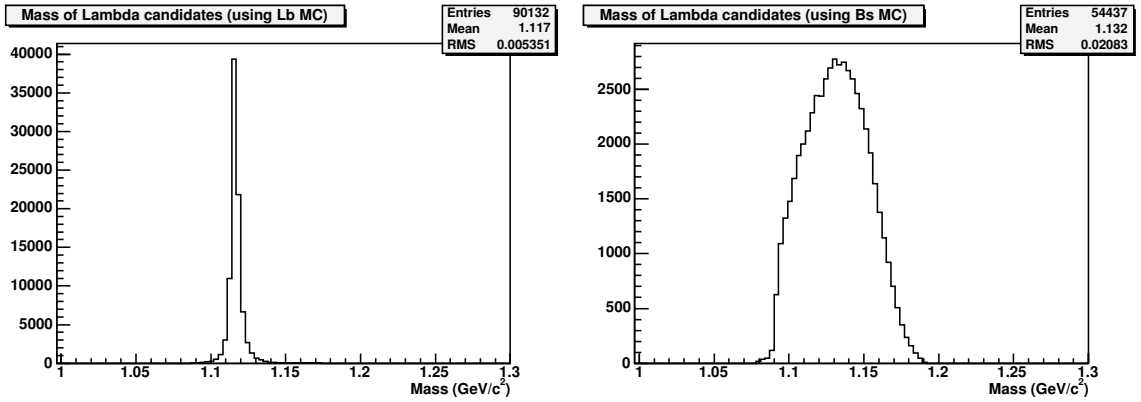

Figure 25: Mass distribution of Λ candidates from Λ_b MC (left) and B_s MC (right).

Figure 26 shows the same distributions after applying a cut at 1.11 GeV/$c^2 \leq$ Mass(Λ) $\leq$ 1.14 GeV/c^2. This cut decreases the acceptance for Λ_b by a factor of 98.9% and 57.2% for B_s. Therefore this cut is less efficient in removing Λ_b background than the cut at cos (θ_{K_S,π_2})> -0.75.

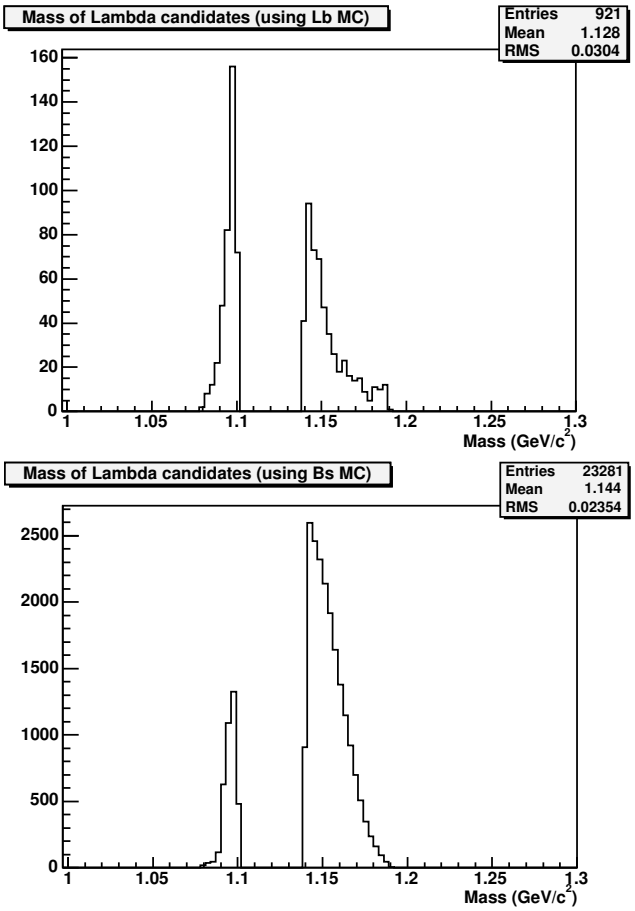

Figure 26: Mass distribution of Λ candidates from Λ_b MC (top) and B_s MC (bottom) after applying a cut at 1.11 GeV/$c^2 \leq$ Mass(Λ) $\leq$ 1.14 GeV/c^2.

B Studies to check the modeling of the Signal

To achieve the optimal Neural Network performance it is important that the input distributions from the signal MC are well-modeled. In order to study the signal modeling B^0 signal events are used. B^0 events have the advantage that they are abundant in data which allows a direct MC-to-data comparison to validate the simulation. Moreover, differences between B^0 MC and B_s MC are not expected to be large. Figures 27 to 29 show the input distribution for data and MC. It is important to point out that in all the distributions for data, the upper sideband has been used to subtract the background contribution. Overall, there is a reasonably good agreement between B^0 data and B^0 MC. Systematic uncertainties are added to the A_{rel} calculation to account for the mismodeling in the lifetime and p_T distributions.

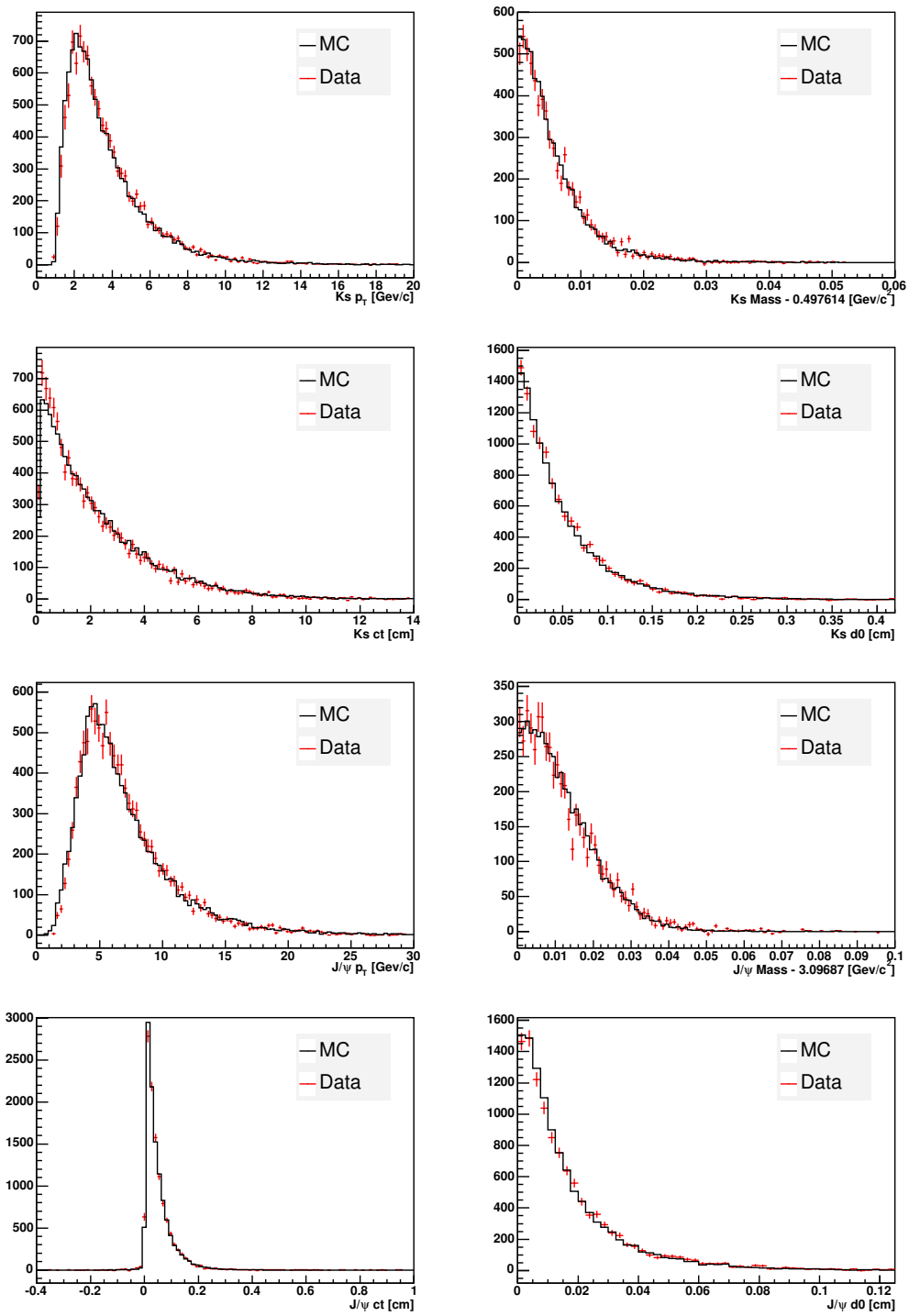

Figure 27: Comparison of data and MC distributions for the input variables to the Neural Network. The black histograms are B^0 MC and the red points are the B^0 data.

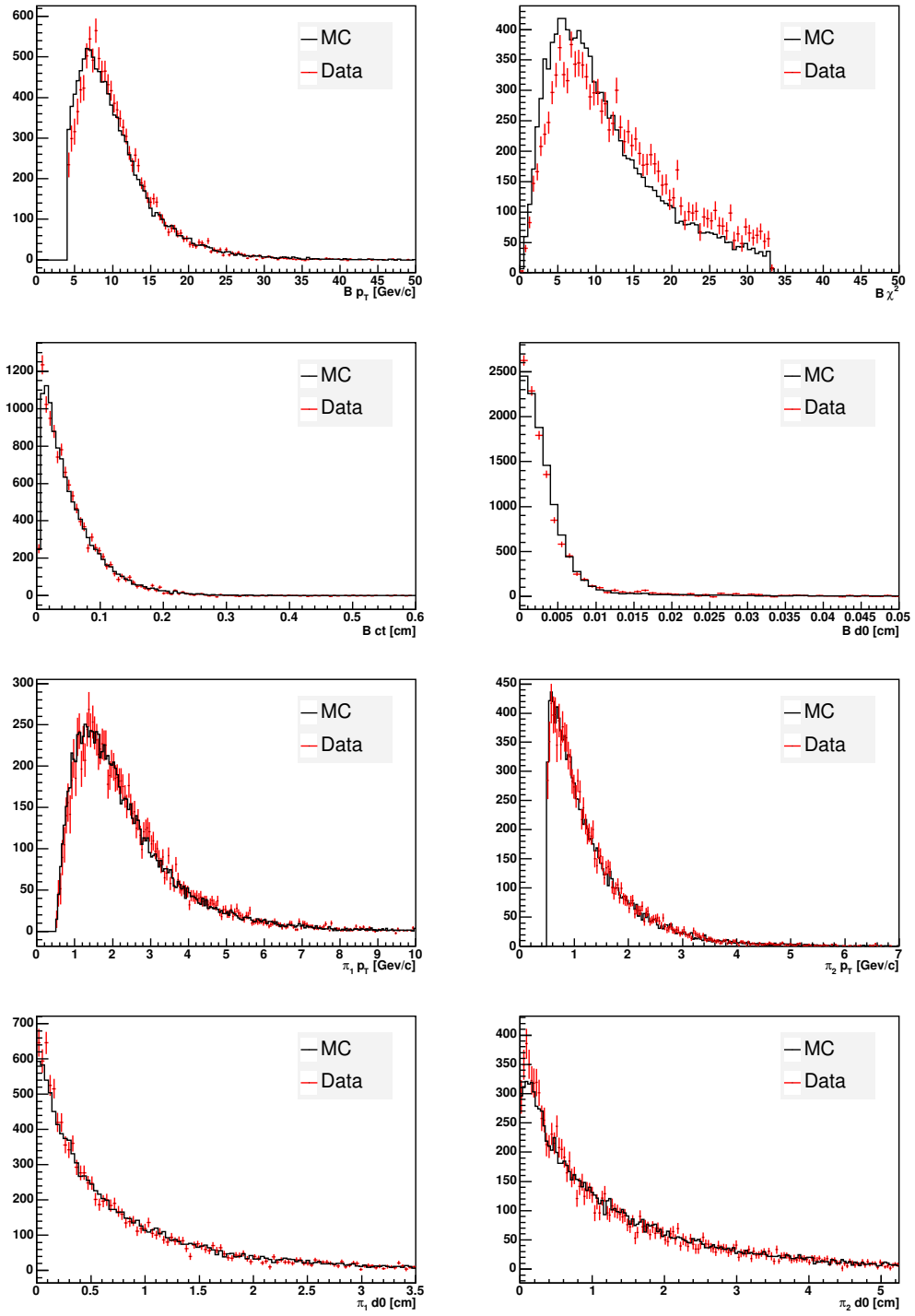

Figure 28: Comparison of data and MC distributions for the input variables to the Neural Network. The black histograms are B^0 MC and the red points are the B^0 data.

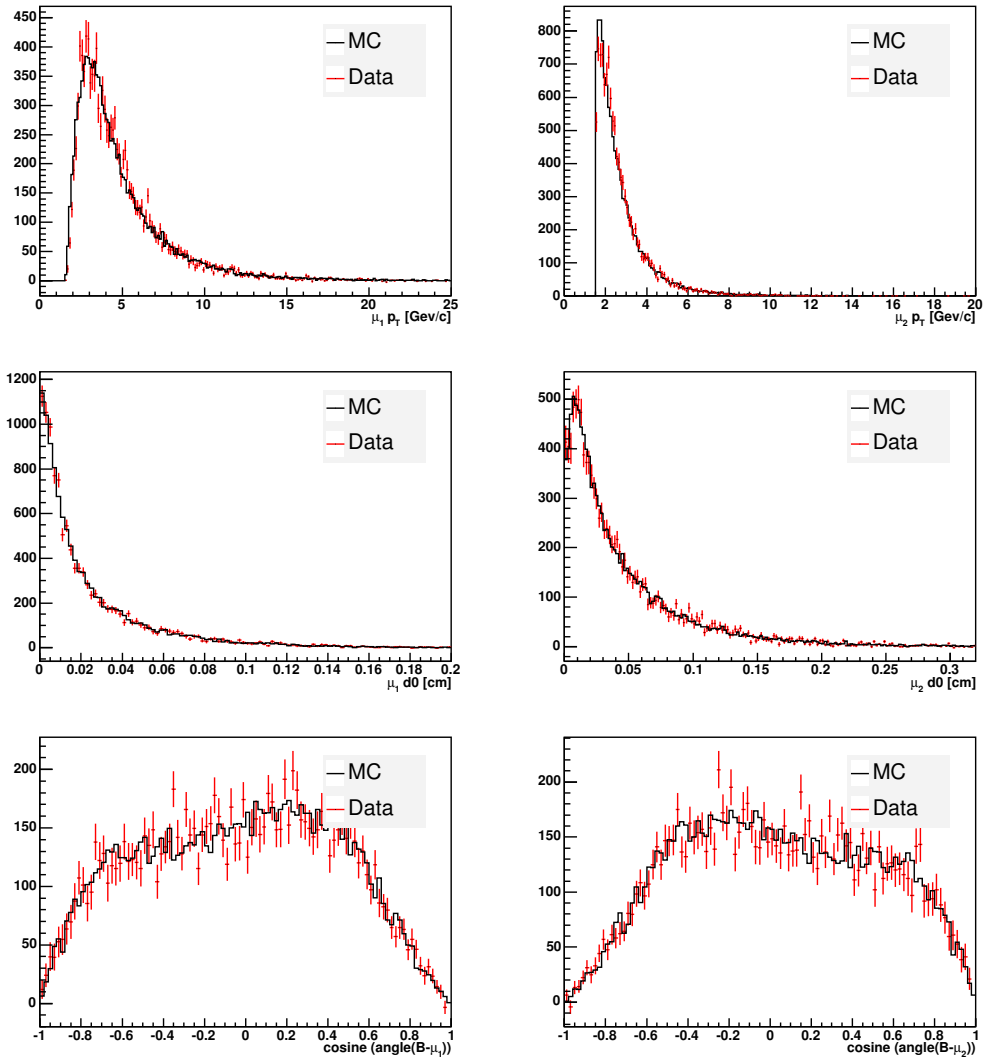

Figure 29: Comparison of data and MC distributions for the input variables to the Neural Network. The black histograms are B^0 MC and the red points are the B^0 data.

C Studies to check the Neural Network performance

Different cross checks has been done to demonstrate that the variables used in the Neural Network training do not sculpt the B invariant mass distribution.
One of the checks is to divide the sideband regions into two parts. One part is used to train the Neural Network as background and the second sideband region is used to train the Neural Network as signal. The object of this exercise is to see if the Neural Network is able to differentiate between different mass regions and preferentially select events from the sidebands regions that were trained as signal, thus sculpting the background shape. The result is shown in Figures 30 and 31. The first plot shows the Neural Network output distribution for "signal" and background. In this case "signal" is sideband events trained as signal. There is a slight separation between the two samples due to the J/ψ p_T and B p_T.

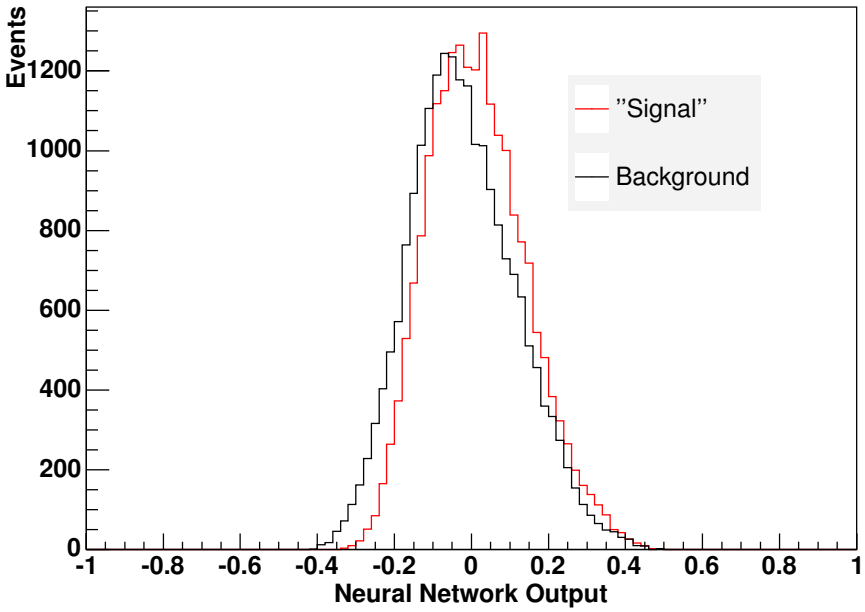

Figure 30: Neural Network Output distribution for background (black) and background trained as "signal" (red).

The second plot shows the B invariant mass distribution as a function of Neural Network output cuts. The sideband events within 5.45 GeV/c^2< M_B <5.8 GeV/c^2 are used as signal in the training and events within 5.8 GeV/c^2< M_B <6.15 GeV/c^2 are background. The result of this study shows that the mass distributions are smooth across the training boundaries and no evidence of sculpting is observed. It is important to stress that while there is a smooth change in efficiency, it is not capable of creating a peak.

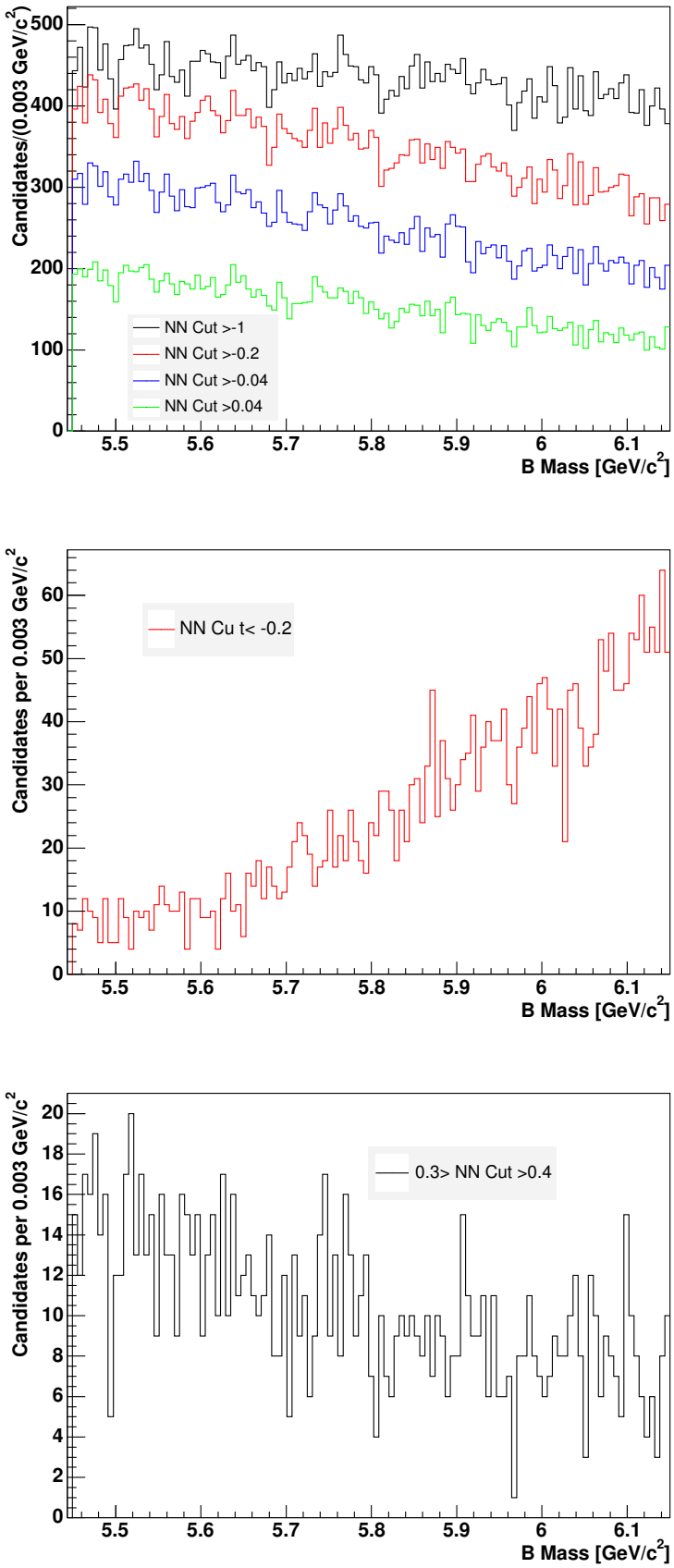

Figure 31: Invariant mass distribution for different Neural Network response cuts.

Another of the checks is to compare the Neural Network output for two different trainings. In one of the trainings the upper sideband is used while the other one uses the lower sideband. The lower sideband includes events within 5.16 GeV/c^2< M_B <5.22 GeV/c^2 and the upper one events within 5.45 GeV/c^2 < M_B <6 GeV/c^2. Figure 32 shows that the behavior of the shape of the Neural Network output is very similar in the upper and lower sidebands. It is important to point out that the number of events used in both trainings is different, approximately 34K in the upper case and 4K in lower one). This result supports the idea the upper sideband can be used to extrapolate the combinatorial background contribution in the signal region and in the lower sideband.

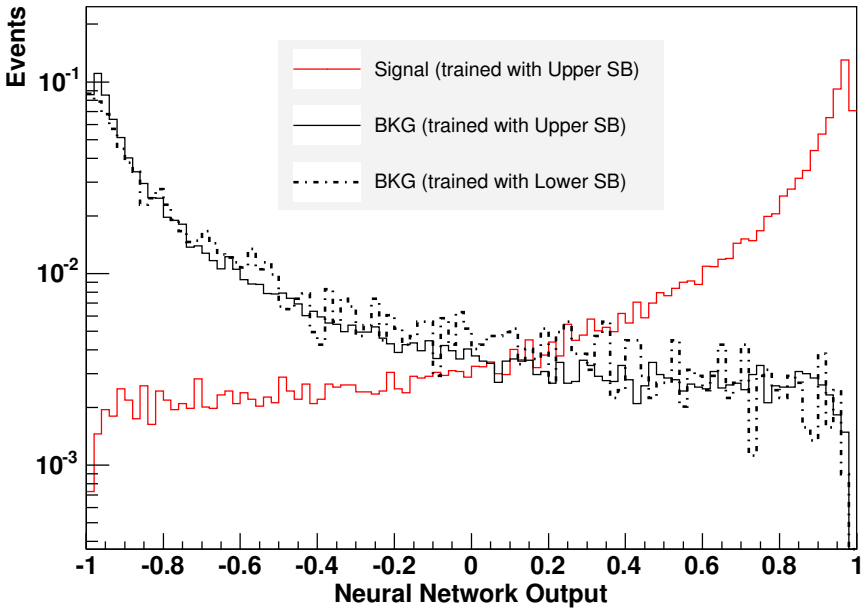

Figure 32: Neural Network Output distributions. The solid line is the result using the upper sideband and the dashed line is the output for the lower sideband.

An additional check is to train the Neural Network with a new B_s MC, generated with B_s mass set to 5.6 GeV/c^2. This mass is far apart from the one used in the default MC. For this test, the mass range for signal is from 5.58 GeV/c^2 to 5.63 GeV/c^2 and events inside the range 5.45 GeV/c^2 to 6 GeV/c^2 are part of the background sample. Figure 33 shows the mass range for signal and background for this check.
The object of this exercise is to see if the Neural Network is bias because it uses the mass of the signal events to differentiate between signal and background. In this case, this new Neural Network will show a much smaller efficiency, compared to the default one, when it is used in data. The results of the Neural Network are shown in Figure 34 and 35. These two plots show that this Neural Network is capable of identifying B^0

and B_s events as signal with similar efficiency than the default one. Therefore this study shows that Neural Network is not using the mass of the events to discriminate B_s events from combinatorial background, in oder words, the Neural Network is not biased.

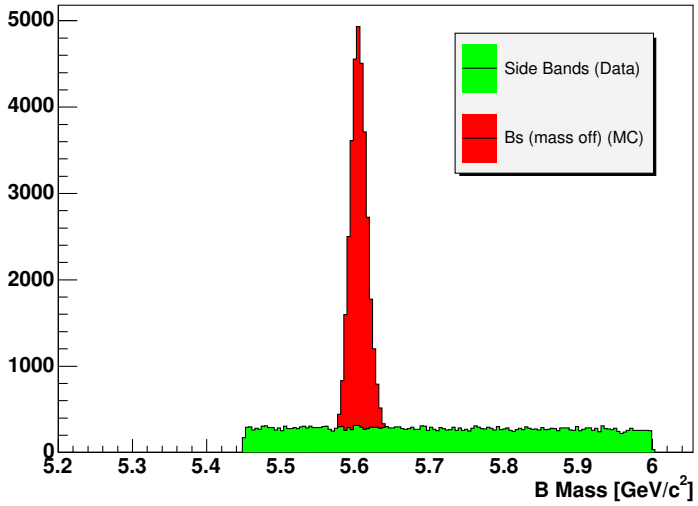

Figure 33: Invariant mass distribution showing the signal and background regions chosen for the Neural Network training in this check.

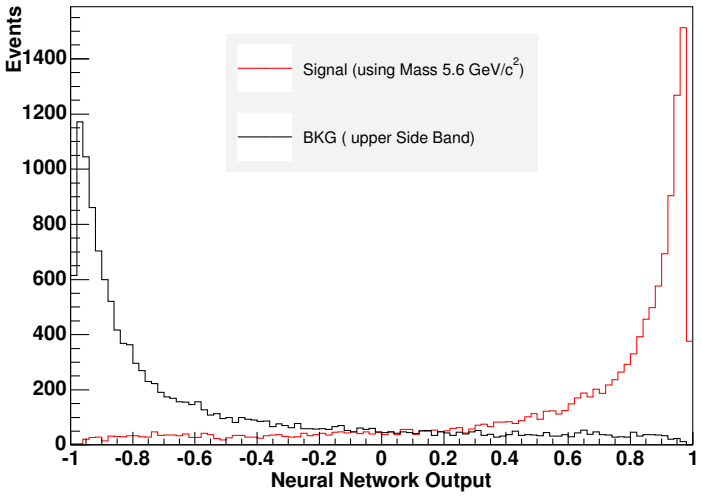

Figure 34: Neural Network output where the red histogram is signal MC and the black one is sideband data in this check.

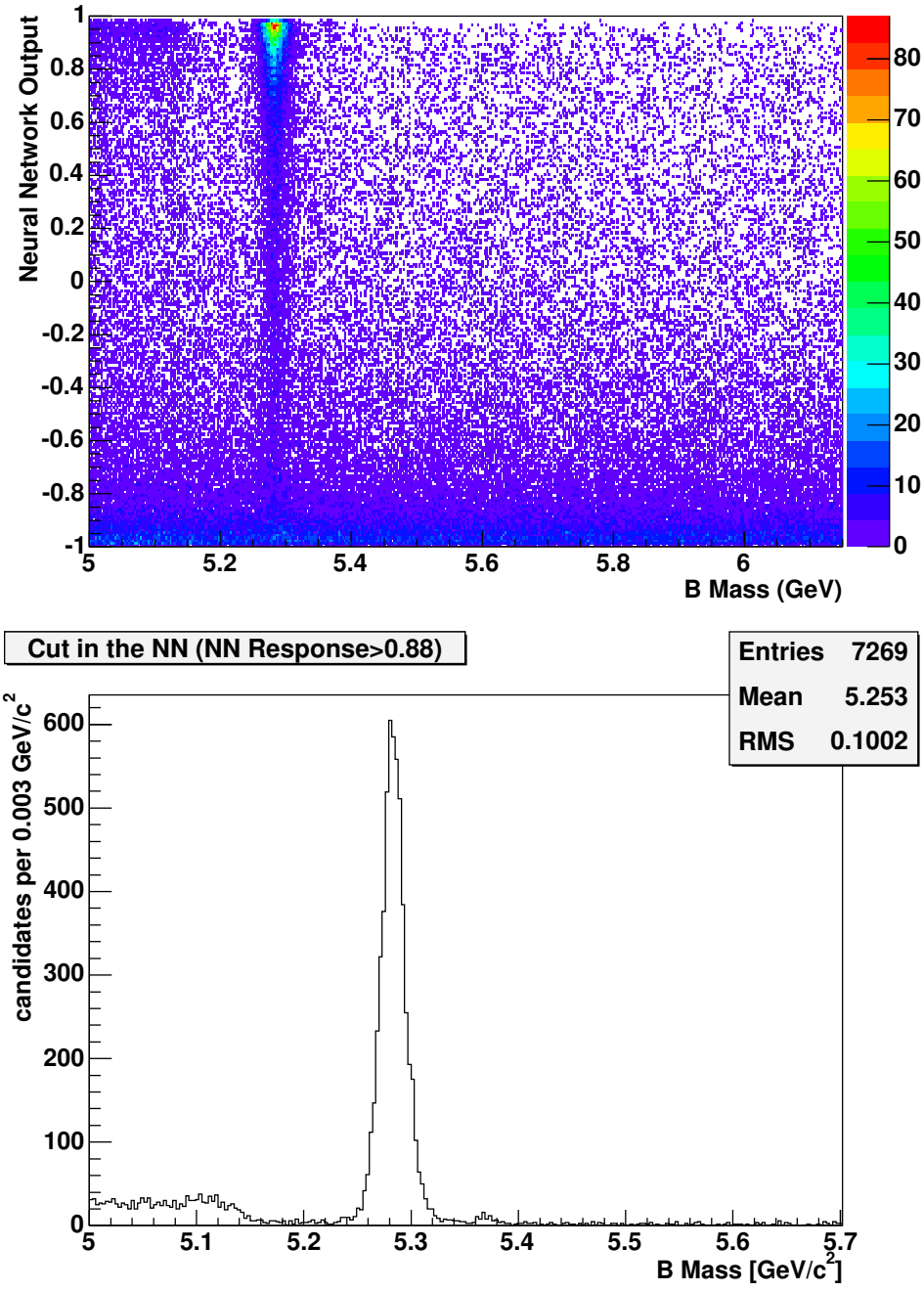

Figure 35: Neural Network output vs the invariant B mass (top) and invariant B mass distribution after the cut in the Neural Network response at 0.88 (bottom).

D Studies to check the consistency of A_{rel} across different trigger selections

A study has been carried out to check that the A_{rel} value is independent of the kinematics of the trigger selection. Seven triggers are selected and A_{rel} is calculated for each one. The trigger selection and the value of A_{rel} are as follow:

- L2_CMU1.5_PT1.5_&_CMX1.5_PT2_DPHI120_OPPQ
  Trigger Cuts: $0 \leq \Delta\phi(\mu\mu) \leq 120$, $M_T \leq 20$ GeV/c^2, μ_1 is CMU and $p_T \geq 1.5$ GeV/c, μ_2 is CMX and $p_T \geq 2$ GeV/c, μ opposite charge and both μ match XFT.
  Relative acceptance is 1.018 ± 0.020.

- L2_TWO_CMU1.5_PT1.5_DPHI120_OPPQ
  Trigger Cuts: $0 \leq \Delta\phi(\mu\mu) \leq 120$, $M_T \leq 20$ GeV/c^2, μ_1 is CMU and $p_T \geq 1.5$ GeV/c, μ_2 is CMU and $p_T \geq 1.5$ GeV/c, μ opposite charge and both μ match XFT.
  Relative acceptance is 1.015 ± 0.016.

- L2_CMU1.5_PT1.7_CMU1.5_PT3_1.7MT7
  Trigger Cuts: $0 \leq \Delta\phi(\mu\mu) \leq 120$, 1.7 GeV/$c^2 \geq M_T \geq 7$ GeV/c^2, μ_1 is CMU and $p_T \geq 3.04$ GeV/c, μ_2 is CMU and $p_T \geq 1.74$ GeV/c and both μ match XFT.
  Relative acceptance is 1.029 ± 0.024.

- L2_TWO_CMUP6_PT4: Trigger Cuts: $0 \leq \Delta\phi(\mu\mu) \leq 180$, μ_1 is CMUP and $p_T \geq 4$ GeV/c, μ_2 is CMUP and $p_T \geq 4$ GeV/c and both μ match XFT.
  Relative acceptance is 0.963 ± 0.044.

- L2_TWO_CMU1.5_PT2_D80_DPHI120_OPPQ
  Trigger Cuts: $0 \leq \Delta\phi(\mu\mu) \leq 120$, $M_T \leq 20$ GeV/c^2, μ_1 is CMU and $p_T \geq 2$ GeV/c, μ_2 is CMU and $p_T \geq 2$ GeV/c, μ opposite charge and both μ match XFT, J/ψ d0 $\leq 80\mu$m and J/ψ $\chi^2 \leq 15$.
  Relative acceptance is 0.991 ± 0.022.

- L2_CMUP1.5_PT3_&_CMU1.5_PT1.5_DPS
  Trigger Cuts: $0 \leq \Delta\phi(\mu\mu) \leq 180$, μ_1 is CMUP and $p_T \geq 3.04$ GeV/c, μ_2 is CMU and $p_T \geq 1.5$ GeV/c and both μ match XFT.
  Relative acceptance is 1.015 ± 0.019.

- L2_CMUP1.5_PT3_&_CMX1.5_PT2_CSX_DPS
  Trigger Cuts: $0 \leq \Delta\phi(\mu\mu) \leq 180$, μ_1 is CMUP and $p_T \geq 3.04$ GeV/c, μ_2 is CMX and $p_T \geq 2$ GeV/c and both μ match XFT.
  Relative acceptance is 1.026 ± 0.032.

This study shows that the relative acceptance is insensitive to trigger mixture.

References

[1] C. Amsler et al., *Phyis. Lett, B 667, 1*, (2008) 2

[2] R.Fleischer, *arXiv:hep-ph/9903455v1*, (1999) 2

[3] M. Feindt, *arXiv:physics/0402093*, (2004) 6

[4] G. Punzi, *arXiv:physics/030863*, 14

[5] B.Carls et al, *CDF note 9895*, (2010) 18, 26

[6] F. Abe et at, *Phys. Rev D55, 1142*, (1997) 18

[7] S. Donati et al, CDF note 8464, (2008) 19

[8] P. Nason, S. Dawson and R.K. Ellis, Nucl. Phys. **B 303** (1988) 607; Nucl. Phys. **B327** (1989) 49. 31

[9] T. Aaltonen et al., *Phys.Rev. D77, 072003*, (2008) 31