ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

January 15, 2018

Search for $B^0_{(s)} \to p\bar{p}p\bar{p}$ decays:
Selection optimization

P. Wampler[1]
Supervisors: F. Blanc[1], P. Marino[1] and E. Rodrigues[2]

[1] *EPFL, Lausanne, Switzerland*
[2] *University of Cincinnati, Cincinnati, OH, United States*

Contents

1 Introduction

Very little is known about the four-body charmless baryonic decays of the B-meson. In particular it is not known whether those decays are suppressed or enhanced as the two-body and three-body decays respectively are. Indeed, a hierarchy in the branching fraction is observed between the two and three body decays. Such a phenomenon is called the multiplicity effect. Figure 1 shows that the two-body decays are suppressed with respect to the three-body.

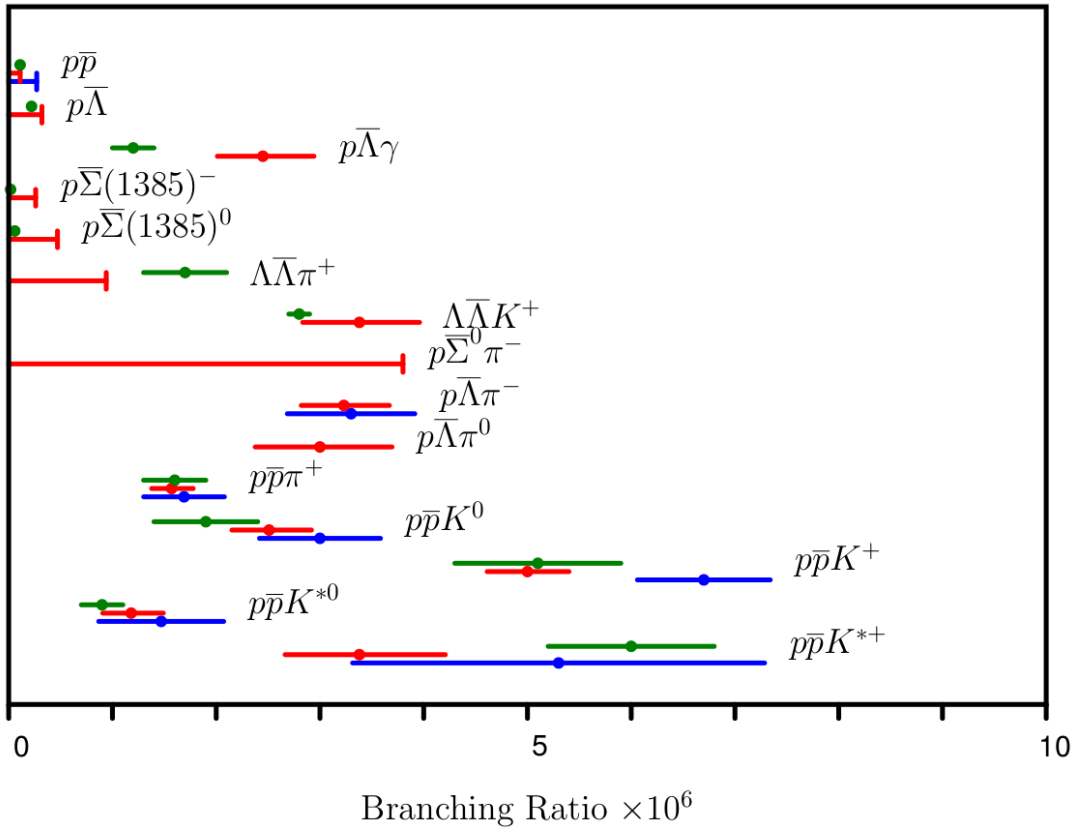

Figure 1: Plot taken from [1]. Experimental results for decays with baryons from BABAR (blue) and Belle (red) and theoretical predictions (green).

The phenomenological explanation of this effect is related to the emission of a gluon for the $p\overline{p}$ pair. More precisely, the two-body baryonic decays are suppressed by the α_s/q^2 term (where α_s is the strong coupling constant and q^2 is the momentum of the exchanged gluon) because the gluon emitted for the final state pair has to carry a large momentum. An example of the necessity of such a gluon can be seen in the Feynman tree diagram of the $B^0 \to p\overline{p}$ decay in Figure 2.

Instead, looking at the Feynman tree diagram of the $B^0 \to p\overline{p}p\overline{p}$ decay[1] in Figure 3, one can see that no hard gluon is necessary. Hence, we expect its branching fraction to be higher than $B^0 \to p\overline{p}$.

[1]Whenever a reference to a decay is made, it recalls implicitly both the decay and its charge conjugated counterpart.

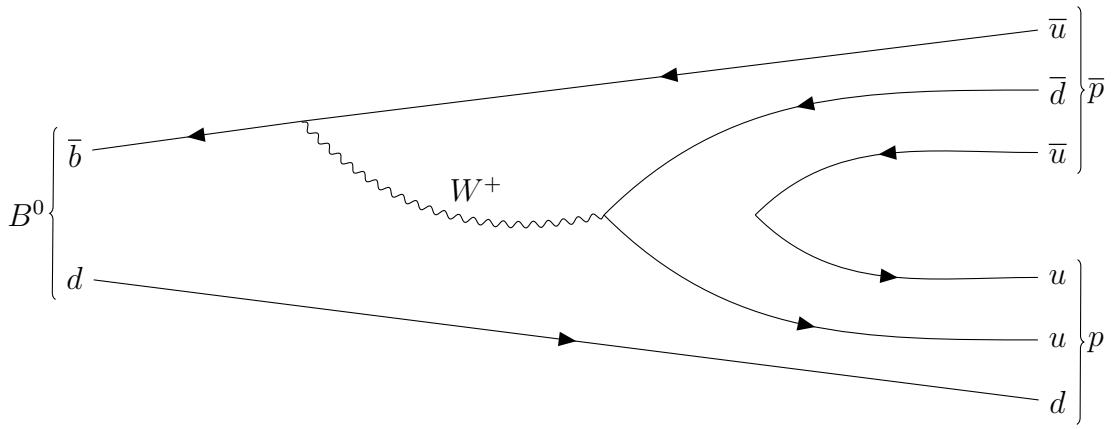

Figure 2: Feynman tree diagram of the $B^0 \to p\overline{p}$.

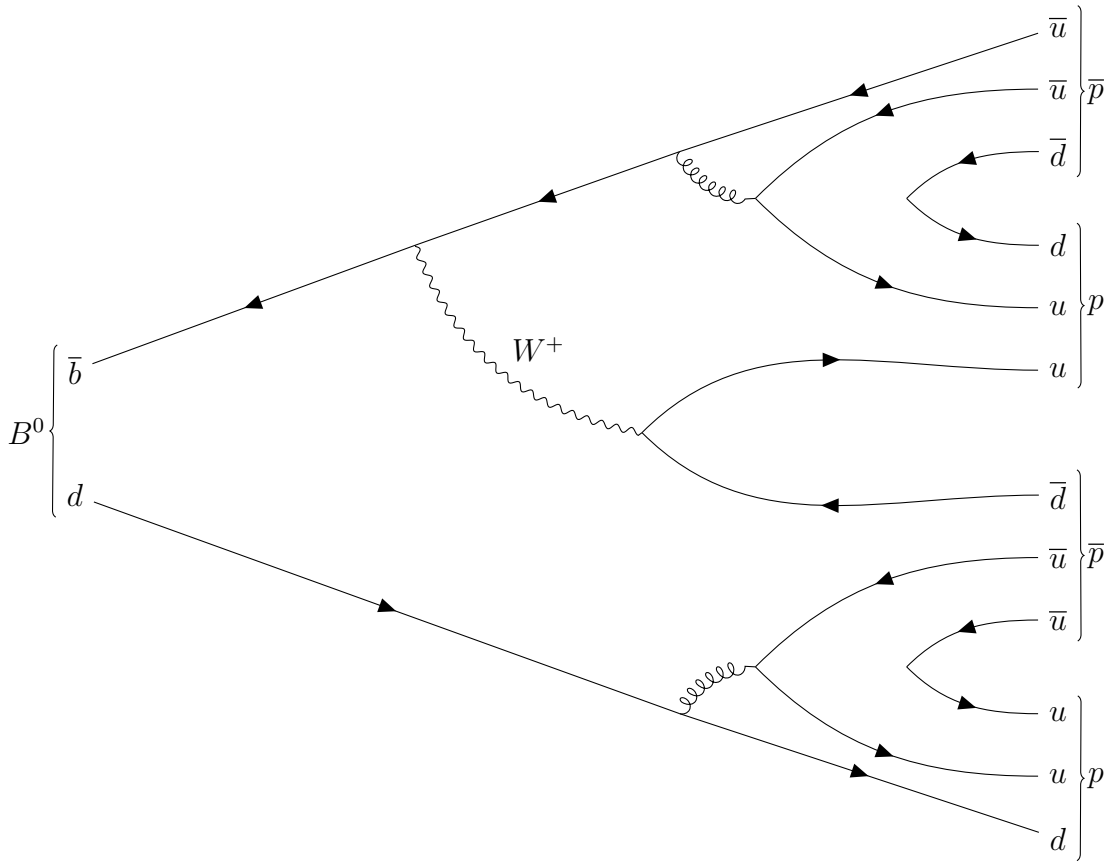

Figure 3: Feynman tree diagram of the $B^0 \to p\overline{p}p\overline{p}$ decay.

In $B \to Mp\overline{p}$ decays (where M is a meson), the $m(p\overline{p})$ distribution is enhanced at twice the rest mass of the p. This effect is the so-called threshold enhancement. It is related to the absence of the hard gluon mentioned above. Indeed, as the gluon does not have to bear much energy, the protons with low energy are favoured.

Because, we expect a multiplicity effect on the branching fraction of the $B^0 \to p\overline{p}p\overline{p}$ decay, we also expect to see a threshold enhancement in two of the four possible $p\overline{p}$ pairs.

An upper limit on the branching fraction of $B^0 \to p\bar{p}p\bar{p}$ was recently set by BABAR[2]:

$$\mathrm{BF}(B^0 \to p\bar{p}p\bar{p}) < 2 \times 10^{-7} \text{ at } 90\% \text{ C.L.} \quad (1)$$

Thanks to a large B cross-section, a good resolution power and the ability to trigger on displaced vertices, the measurement capabilities of the LHCb experiment motivate the search for the $B^0_{(s)} \to p\bar{p}p\bar{p}$ decays.

The goals of this analysis are: firstly to provide the observation of both the $B^0 \to p\bar{p}p\bar{p}$ and the $B^0_s \to p\bar{p}p\bar{p}$ decays. Secondly to determine their exclusive branching fraction in order to probe the multiplicity effect. Finally, to study their phase space in order to settle on the threshold enhancement question. The selections are optimized for the B^0 meson and will directly be applied to the B^0_s meson.

The expected yields for RunI and II are given in Table 1. The low value is determined using the $B^0 \to p\bar{p}$ branching fraction ($\mathrm{BF} = 0.15 \times 10^{-7}$) because we expect to see the multiplicity effect. The high value is determined using the upper limit on the branching fraction stated above. A rough estimation of the expected yields was computed in the following way[3]:

$$\mathcal{L}_{int} \cdot \sigma(b\bar{b}) \cdot 2 \cdot \mathrm{BF} \cdot \epsilon_{\mathrm{MC}} (\approx 0.8\%).$$

Table 1: Expected $B^0 \to p\bar{p}p\bar{p}$ yields for RunI and RunII. The low value is determined using the $B^0 \to p\bar{p}$ branching fraction and the high value is determined using the upper limit on the branching fraction in equation (1).

Run number	Low Yield	High Yield	$\sigma(b\bar{b})\,[\mu\mathrm{m}]$	$\int \mathcal{L}\,[\mathrm{fb}^{-1}]$
I: $\sqrt{s} \approx 7$ TeV	50	675	72	1.2
II: $\sqrt{s} = 13$ TeV	70	900	144	3

The report presents the selection optimization and is organized as follows: The LHCb detector, the event samples and the blinding procedure are presented in section 2 while the various steps of the event selection and figure-of-merit optimization are presented in section 3.

2 Event samples

2.1 The LHCb detector

The LHCb detector [2,3] is a single-arm forward spectrometer covering the pseudorapidity range $2 < \eta < 5$, designed for the study of particles containing b or c quarks. The detector elements that are particularly relevant to this analysis are: a silicon-strip vertex detector surrounding the pp interaction region that allows b hadrons to be identified from their characteristically long flight distance; a tracking system that provides a measurement of momentum, p, of charged particles; and two ring-imaging Cherenkov detectors that are able

[2] 2017 Hadron Conference

[3] The hadronization factor is assumed to be 1.

to discriminate between different species of charged hadrons. The online event selection is performed by a trigger, which consists of a hardware stage, based on information from the calorimeter and muon systems, followed by a software stage, which applies a full event reconstruction. A loose preselection is then centrally produced to provide the candidates for designated decays. Those samples are the so-called stripping lines.

2.2 Data

The data used in this analysis have been collected during the so-called RunI and RunII periods of operation of the LHC machine in 2012-2012 and 2015-2018, respectively[4]. Table 2 shows the centre-of-mass energies and the corresponding integrated luminosities acquired for both RunI and II. The results presented in this report were only performed on RunI data, RunII samples will be incorporated subsequently.

Table 2: Data samples divided for the year of data-taking.

Run#	Year	$\sqrt{s}\,[\mathrm{TeV}]$	$\int \mathcal{L}\,[\mathrm{fb}^{-1}]$
RunI	2011	7	1
	2012	8	2
RunII	2015	13	0.3
	2016	13	1.7

2.3 Monte Carlo simulation

The simulated samples for the $B^0 \to p\bar{p}p\bar{p}$ decay were produced centrally by the LHCb collaboration, with a uniform 4-body phase space distribution.

Table 3 shows the approximate number of events[5] for both the data and simulated samples as processed by the stripping line (see section 3.2):

Table 3: The number of events in data and simulated samples straight out of the stripping.

Year and Polarity	Data[6][M]	MC[k]
2011 *MagUp*	0.75	9
2011 *MagDown*	1.15	10
2012 *MagUp*	1.95	18
2012 *MagDown*	1.60	17
2015 *MagUp*	0.013	8
2015 *MagDown*	0.019	6
2016 *MagUp*	0.60	26
2016 *MagDown*	0.61	26

[4]RunII is still ongoing.

[5]This corresponds to about 80% of the total sample due to a recent crash of the TIER1 node of the grid (CNAF).

Throughout the optimization of the selection, all simulated events are required to be truth-matched. This means that both the reconstructed B-meson and the final states can be matched with corresponding generated particles.

2.4 Blinding procedure

In order to avoid biasing the event selection, the measurement is performed while blinding the signal region. The nominal masses [4] for the B^0 and the B_s^0 meson are $(5279.63 \pm 0.15)\,\mathrm{MeV}/c^2$ and $(5366.89 \pm 0.19)\,\mathrm{MeV}/c^2$, respectively. The mass resolution obtained from the simulated samples is about $8\,\mathrm{MeV}/c^2$ and is rounded up to $10\,\mathrm{MeV}/c^2$. Selecting a region of 4σ around the signal peak produces as signal regions:

- $[5240, 5320]$ MeV/c^2 for the B^0,
- $[5327, 5407]$ MeV/c^2 for the B_s^0.

Hence, the events in the region $[5240, 5407]$ MeV/c^2 were cut away. However, the small background region between $[5320, 5327]$ MeV/c^2 might be useful to further constrain the background distribution.

3 Event Selection

The various stages of the selection are presented hereafter. All events are selected by specific trigger (section 3.1) and stripping (section 3.2) lines. Those two stages are later referred to as the preselection.

A selection based on particle identification (PID) variables is applied to distinguish signal from background due to *misidentification* (section 3.3).

Furthermore, a multivariate (MVA) classifier trained on topological and kinematical variables is used to separate signal from *combinatorial* background (section 3.4).

Finally the study of a figure-of-merit is performed as a function of both the PID and the MVA selection parameters. From this study an optimal working point is extracted (section 3.5).

3.1 Trigger selection

The trigger lines chosen for this analysis and their associated efficiencies on the Run1 simulated signal samples are detailed in Table 4. The set of lines is typical for analyses of similar hadronic final states.

At the hardware trigger level (L0), an event is required to have a particle from the signal that passes the threshold on the hadronic calorimeter (Trigger On Signal `TOS`) or any threshold if the particle is independent from the signal (Trigger Independent of Signal `TIS`). At the first software trigger level (HLT1), events with a good track quality and a suitable impact-parameter are selected. At the second software trigger level (HLT2), a multivariate algorithm trained on B-meson samples looks for either 2,3 or 4 particles that could add up to be compatible with the B^0 invariant mass.

The efficiencies on Table 4 show that more than half of the simulated signal is lost through L0 and a quarter is lost through HLT1 and HLT2 combined. The total efficiency is not so far from the one for the $B^0 \to p\overline{p}$ analysis [5] ($\epsilon \approx 38\%$). Expecting the multiplicity

effect and given that $B^0 \to p\bar{p}$ observation (5.3σ) was performed with RunI data only, the comparable efficiency indicates a large probability of observation of the $B^0 \to p\bar{p}p\bar{p}$ decay.

Table 4: Trigger requirements applied to all event samples. The associated efficiencies are given for all RunI simulated samples as a whole. All efficiencies are normalized to the number of events that passed the previous trigger level.

Level	Line	MC Efficiency[%]
L0	L0Global_TIS	27.29
	L0HadronDecision_TOS	21.84
	OR	41.44
HLT1	Hlt1TrackAllL0Decision_TOS	86.61
HLT2	Hlt2Topo2BodyBBDTDecision_TOS	53.30
	Hlt2Topo3BodyBBDTDecision_TOS	38.91
	Hlt2Topo4BodyBBDTDecision_TOS	50.86
	OR	70.90
Total		25.45

3.2 Stripping Selection

The fully reconstructed events by the software trigger stage are selected under specific criteria. The stripping selection details are given in Table 5.

A selection on each final state proton is performed, requiring good track and particle identification. Moreover, a large impact parameter χ^2 is required. The χ^2 comes from the fit of tracks to determine the primary vertex. If a proton track from the $B^0_{(s)} \to p\bar{p}p\bar{p}$ decay is included, the χ^2 is larger because its track does not come from the primary vertex but from a displaced one at which the $B^0_{(s)}$ has decayed.

The final states as a whole are cut on the invariant mass and the distance of closest approach among the them.

The $B^0_{(s)}$ candidate is required to have good topological parameters, a large enough decay time and a small impact parameter χ^2. For a true $B^0_{(s)}$ candidate, the impact parameter χ^2 is expected be to be small since the meson is produced at the primary vertex.

The invariant mass distribution straight out of the stripping for both the data and the simulated a signal sample can be seen on Figure 4.

Table 5: List of stripping selection criteria for the signal decays.

PROTON SELECTION	
Track fit χ^2/nDoF	< 4
Track ghost probability	< 0.3
$\mathrm{DLL}_{p\pi}$	> 0
DLL_{pK}	> -5
Impact parameter χ^2	> 25
$p\bar{p}p\bar{p}$ SELECTION	
Invariant mass	400 MeV around 5323 MeV
Distance of closest approach between any two daughters	< 0.3mm
$B^0_{(s)}$ SELECTION	
Vertex fit χ^2/nDoF	< 9
Cosine of the angle between the particle momentum and its direction of flight	> 0
Decay time	> 1.0ps
Impact parameter χ^2	< 25
χ^2 separation from related PV	> 225

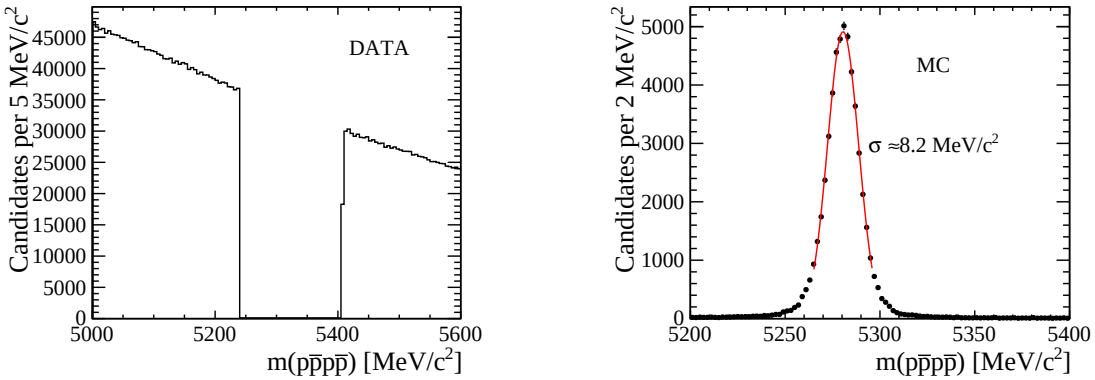

Figure 4: Invariant mass distribution straight out of the stripping. On the left is the data and on the right is the simulated sample fitted with a simple Gaussian, both plots use the samples associated to the 2012 magnet polarity up configuration.

3.3 PID Selection

In order to discard events that result from misidentification (*e.g.* $B^0 \to p\bar{p}hh'$ reconstructed as $B^0 \to p\bar{p}p\bar{p}$), a selection on a PID variable is performed. The variable used is `ProbNN_p`. It is the output of a neural network. Using as input PID information from the subdetector, the network is trained with simulated samples of inclusive B events. Its output is a single probability value for the proton hypothesis. The distribution of the `ProbNN_p` for all final protons is shown on Figure 5.

A study of the PID cut efficiency was done for the sideband data and two samples simulated by the LHCb collaboration, $B^0 \to p\bar{p}KK$ and $B^0 \to p\bar{p}K\pi$. Such decays could potentially contribute to the signal region if reconstructed as $B^0 \to p\bar{p}p\bar{p}$. However, a toy simulation shows that almost no events from those two modes could sneak in our signal regions. Figure 8 shows how far those misidentified reconstructions would fall from both of the signal regions. This allows to safely neglect those sources of background and concentrate on removing the background as given by the sidebands on the left of Figure 4.

Both a cut on the `ProbNN_p` of each proton and a cut on the *product* of the `ProbNN_p` of each proton were tried. A comparison of Figures 6 and 7 shows that cutting on the product is more efficient than cutting on each `ProbNN_p` separately. This can be understood in the following way: With a cut applied to each `ProbNN_p` separately, it makes no difference if the threshold is easily passed or not. Hence, an event with very high `ProbNN_p` protons has the same probability to pass the cut as an event with `ProbNN_p` just above the threshold. In contrast, the cut on the product of `ProbNN_p` exploits the full `ProbNN_p` value of each proton which allows to better distinguish signal from background.

Figure 9 shows the number of expected events in the B^0 signal region as a function of the cut on the product of all `ProbNN_p` for both the simulated signal sample and the sideband data. The PID cut allows to reduce drastically the background contained in the sidebands while preserving most of the signal. This indicates that this background is mainly composed of π and K, as expected.

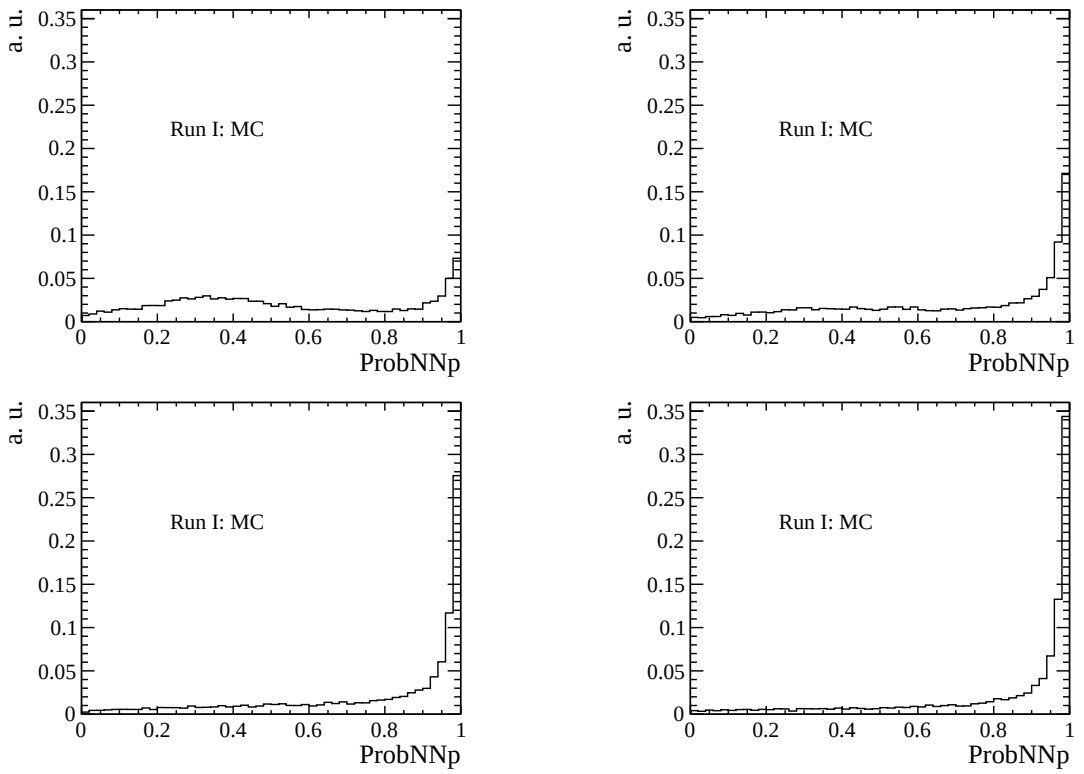

Figure 5: `ProbNN_p` distribution of each final state proton (ordered by increasing momentum from the top left to the bottom right).

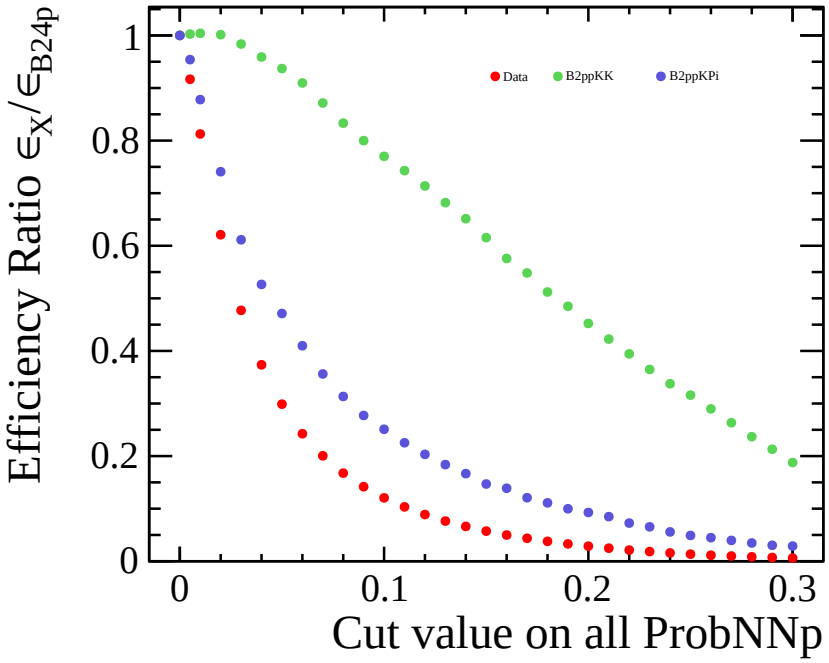

Figure 6: Cut efficiencies of $B^0 \to p\bar{p}KK$ and $B^0 \to p\bar{p}K\pi$ simulated samples and sideband data relative to the $B^0 \to p\bar{p}p\bar{p}$ simulated sample. The same cut is performed on the `ProbNN_p` of each final state proton.

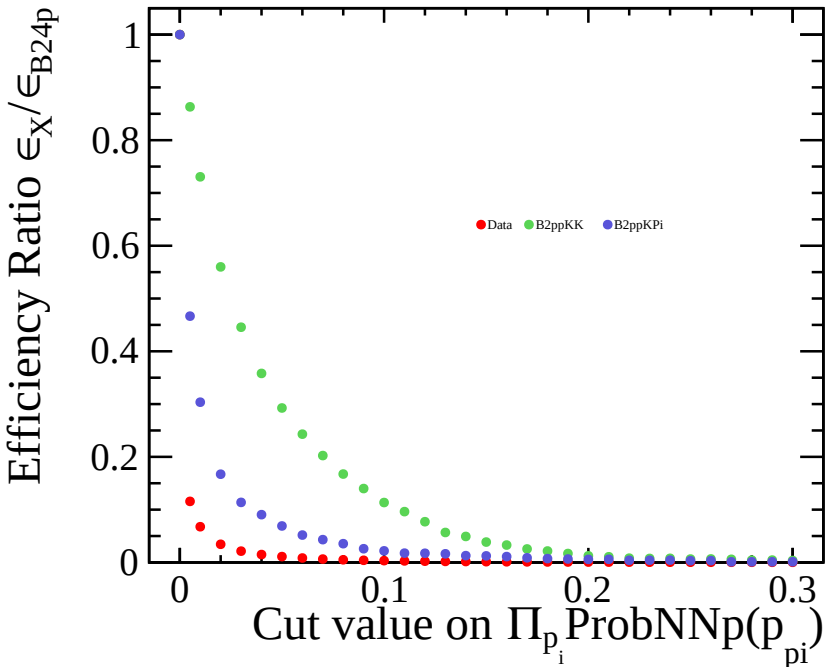

Figure 7: Cut efficiencies of $B^0 \to p\bar{p}KK$ and $B^0 \to p\bar{p}K\pi$ simulated samples and sideband data relative to the $B^0 \to p\bar{p}p\bar{p}$ simulated sample. The cut is performed on the *product* of `ProbNN_p` of all final state protons.

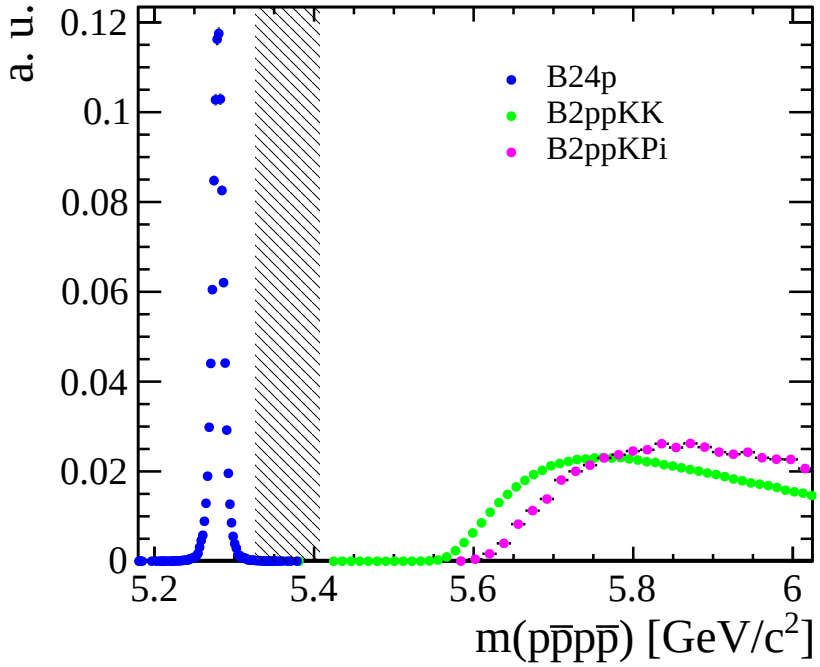

Figure 8: Simulated samples showing where would fall the $B^0 \to p\bar{p}KK$ and $B^0 \to p\bar{p}K\pi$ if they were reconstructed as $B^0 \to p\bar{p}p\bar{p}$. The gray zone is the signal region for the $B_s^0 \to p\bar{p}p\bar{p}$.

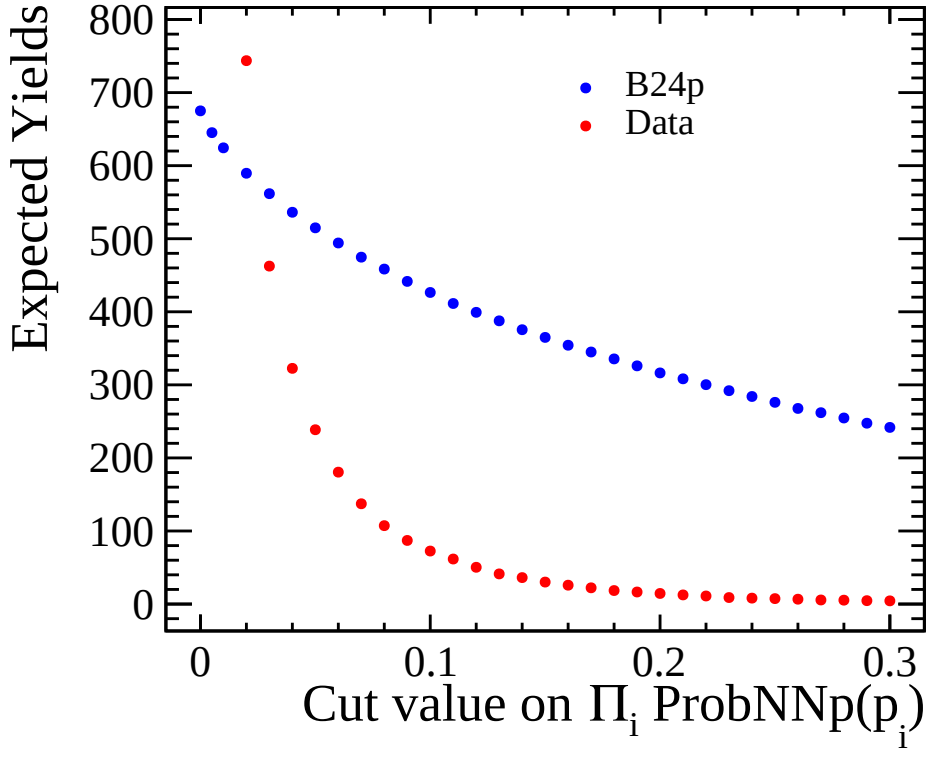

Figure 9: Expected yields in the signal region after the PID cut on the product of `ProbNN_p` of all protons as a function of the cut value.

3.4 MVA selection

A multivariate classifier is a performant tool to distinguish signal from background through the use of reasonably chosen variables and their correlation. Because of its robustness, we chose to use the boosted decision tree (BDT) [6, 7] implemented in the TMVA package [8]. It was trained with the full RunI simulated samples for the signal and the full RunI sideband data for the background, both after the application of the PID cut (product of `ProbNN_p` > 0.045).

The variables used are mainly topological and kinematical, Table 6 gives a list of them. They were chosen because of their signal/background separating power and their very low correlation with the invariant mass. Their distribution and linear correlation can be seen in Figures 10 and 11 respectively. Figure 12 shows that no overtraining occurred. Indeed, one can see that the BDT picked almost no statistical fluctuation from the signal training sample and very little from the background training sample. Figure 13 shows the receiver operating characteristic (ROC) curve from the application of the BDT to the test samples. The classifier is able to reject a large proportion of background while retaining much of the signal.

Instead of separating with respect to the year and magnet polarity samples, the BDT was trained and tested on the whole RunI data. To cross-check that this simplification did not introduce any bias, the training has been repeated separately on each sub-sample (as present on Table 3) and the obtained ROC curves have been compared as shown in Figure 14. No significant differences are observed. Indeed, the four ROC curves are compatible within the statistical fluctuations.

In order to show that no peak is artificially created by the application of the classifier, the BDT selection was applied to the wrong sign data. This sample is a set of events for which the total charge of the final states protons does not add up to zero. Those events were reconstructed with at least one wrong proton such that the conservation of charge is violated. Hence, the wrong sign sample bears no physical events. The invariant mass distribution of the wrong sign data before and after the action of the BDT is shown in Figures 15 and 16 respectively, no artificially created peaks are observed.

Table 6: List of the variables passed to the BDT.

Variable	Definition
`B_log_IPCHI2_OWNPV`	Impact parameter χ^2 of the B^0 candidate
`DTF_PV_CHI2_nDOF`	χ^2 per degree of freedom from the fit of the decay tree including the primary vertex
`B_DIRA_OWNPV`	Cosine of the angle between the direction of flight and the momentum of the B^0 candidate
`B_P`	Momentum of the B^0 candidate
`B_FD_OWNPV`	Flight distance of the B^0 candidate from the primary vertex

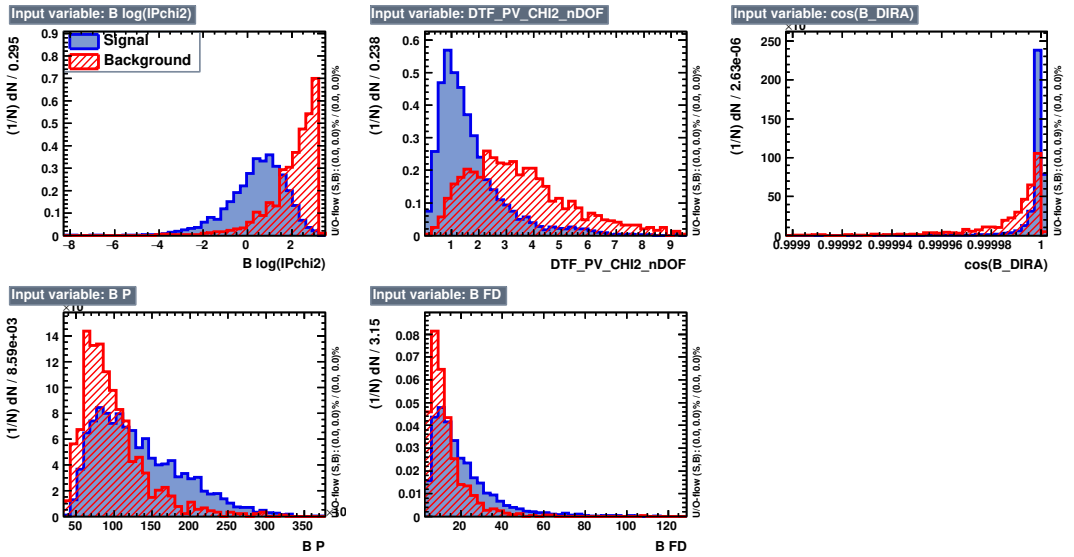

Figure 10: Distribution of the variables from the training sample used in the BDT. The variables are ordered in decreasing separation power from the top left to the bottom right. The signal is shown in blue whereas the background is shown is red.

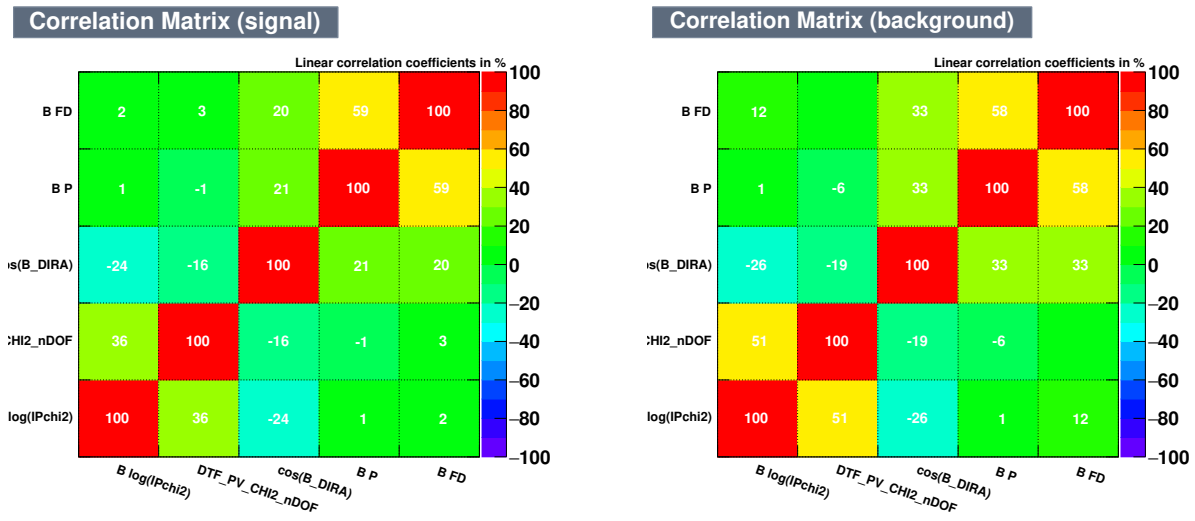

Figure 11: Linear correlation between the variables used in the BDT.

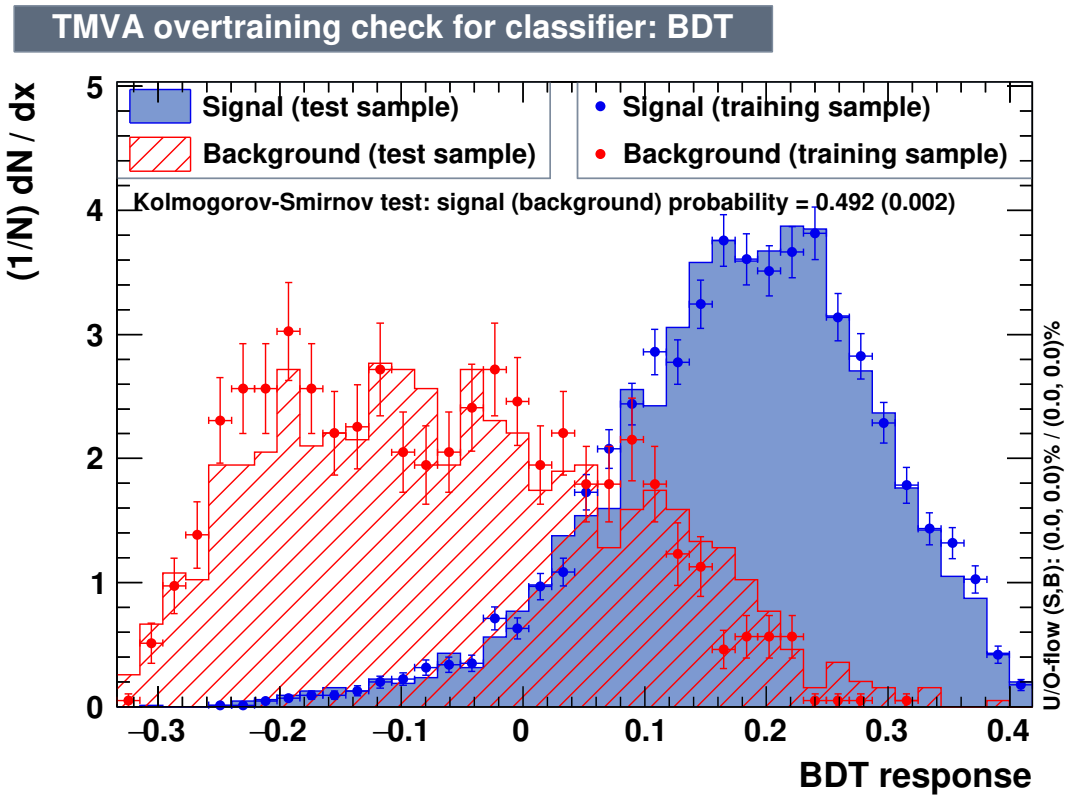

Figure 12: Distribution of the BDT output for both the training and the testing samples

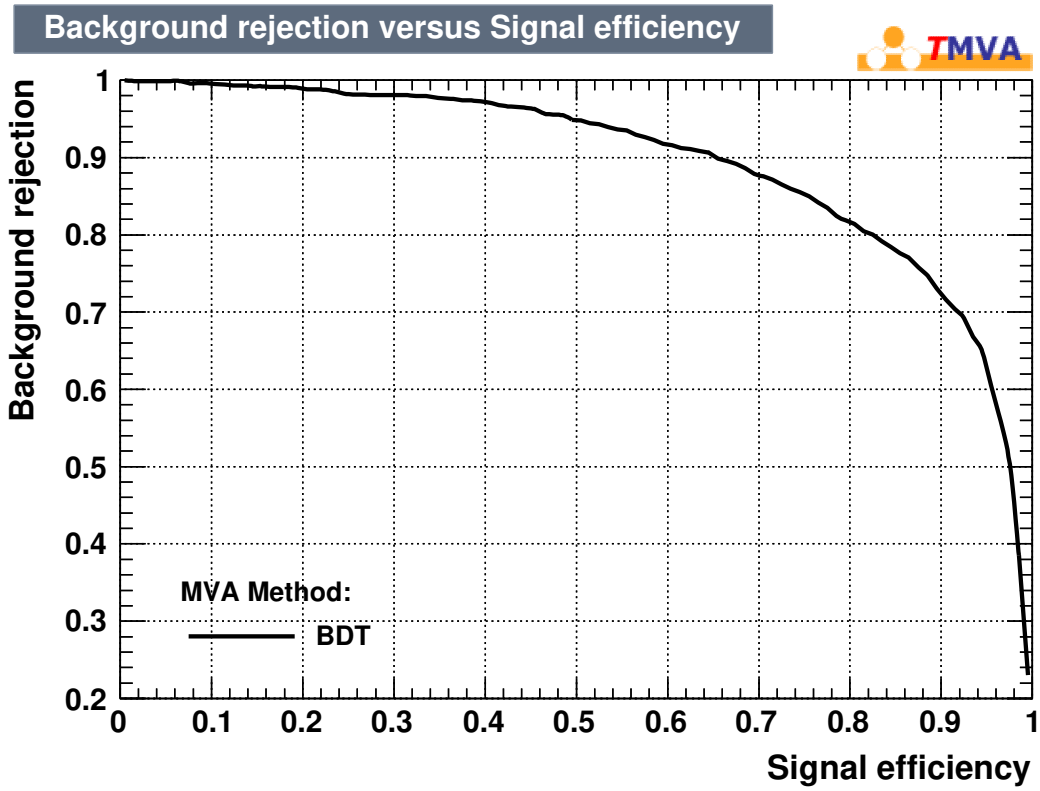

Figure 13: The ROC curve from the application of the BDT to the test samples.

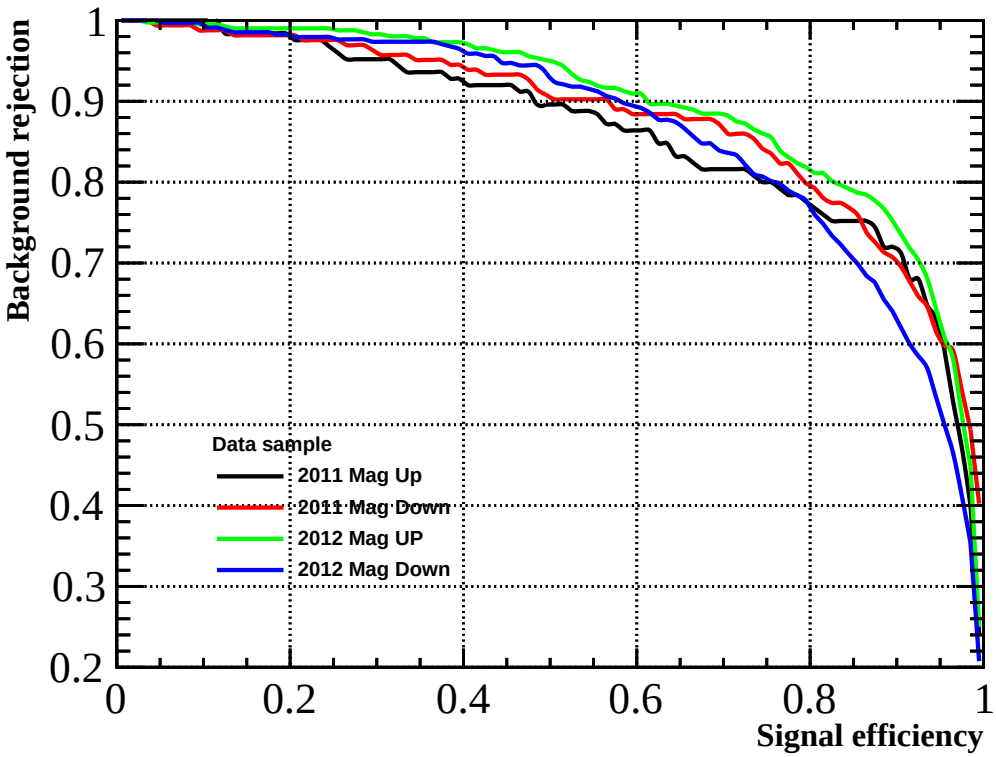

Figure 14: Superposition of the ROC curve for all sub-samples of RunI.

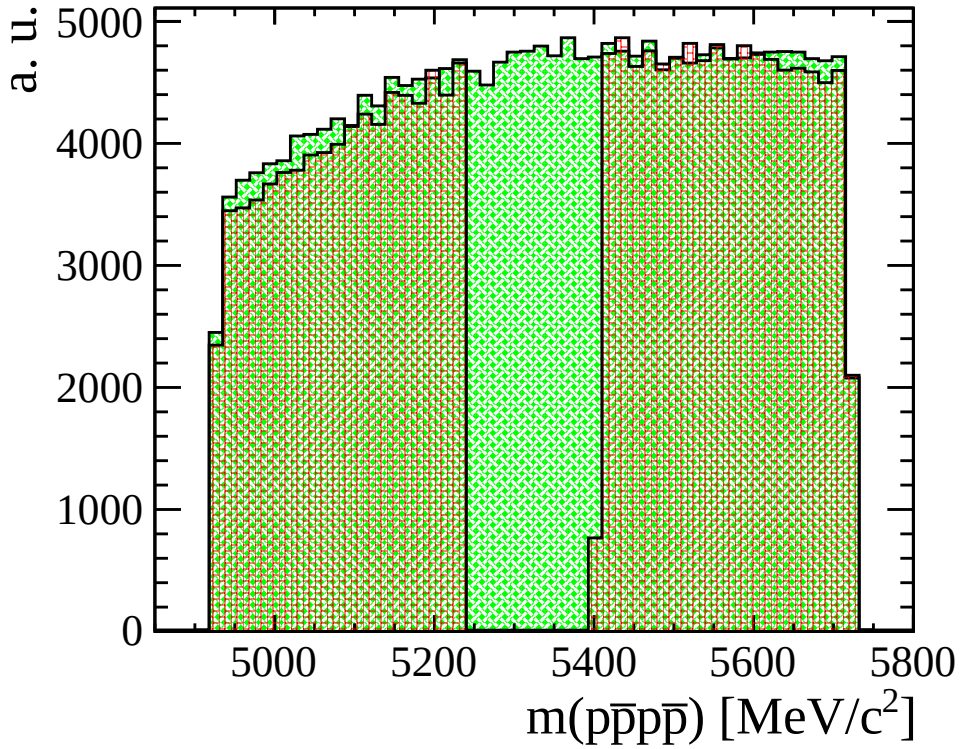

Figure 15: Invariant mass distribution of the data (in red bricks) and the wrong sign data (in green fabric) just after the preselection and *before* applying the BDT.

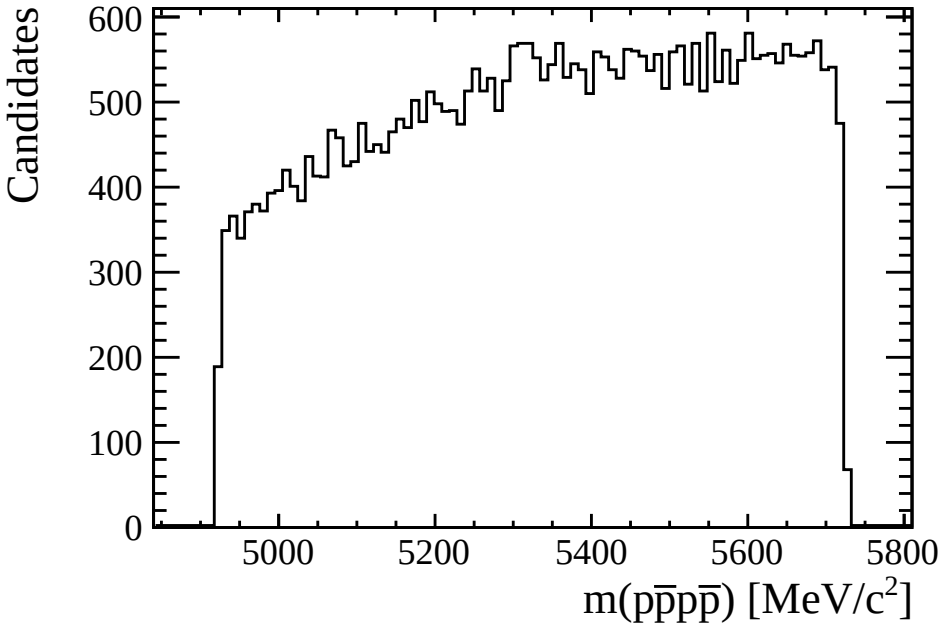

Figure 16: Invariant mass distribution of the wrong sign data *after* applying the BDT but without any PID cut.

3.5 Working point

The effect of both the PID and the BDT cuts have been studied separately but it remains to optimize the two cuts with respect to a figure-of-merit. The chosen figure-of-merit is the significance defined as:

$$\frac{S}{\sqrt{S+B}} \tag{2}$$

where S and B are the yields of respectively the signal and the background.

In this study S and B are the expected yields in the B^0 signal region after the application of the PID and the BDT selection. Figure 17 shows the significance as a function of both the cut on the PID and the BDT output. A zoom on the high region is shown in Figure 18. We can see that the optimal pair of cut values is approximately (PID,BDT)= $(0.045, 0)$, which yields a significance of 20.35.

Figures 19 and 20 show the invariant mass after both the PID and the BDT cuts at the optimal values. The efficiencies of the simulated samples and the background rejection for each selection stage are shown in Table 7. The sideband data is mostly rejected *via* the trigger and the PID selections. Nonetheless, the multivariate classifier is useful to further distinguish the signal from the background. The simulated signal events were mainly preserved by the PID and the classifier selection. Their low efficiency is mostly due to the trigger requirements.

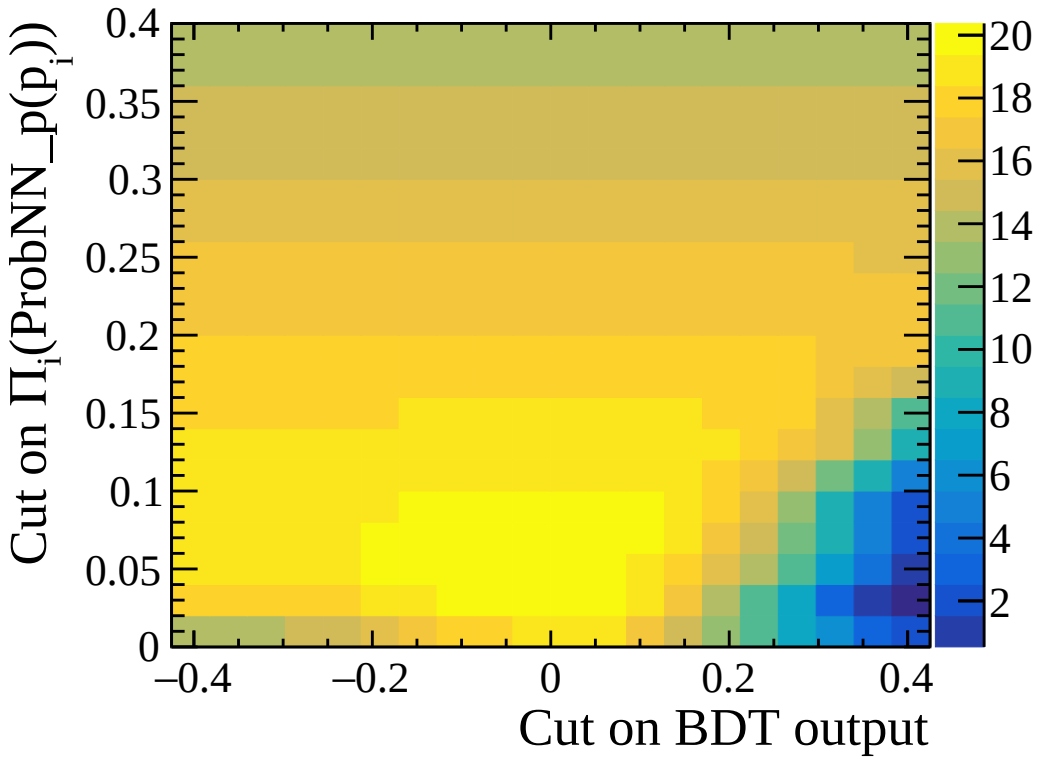

Figure 17: The significance as a function of the cut on the product of `ProbNN_p` (PID) and the cut on the BDT output.

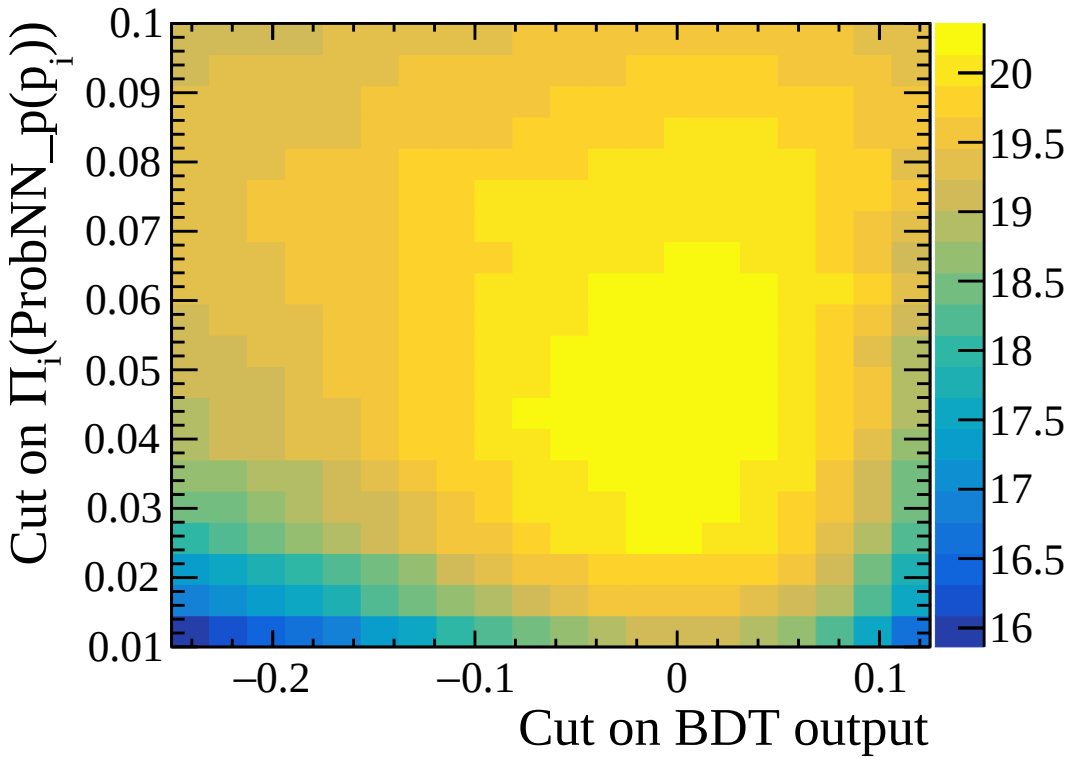

Figure 18: The significance as a function of the cut on the product of `ProbNN_p` (PID) and the cut on the BDT output.

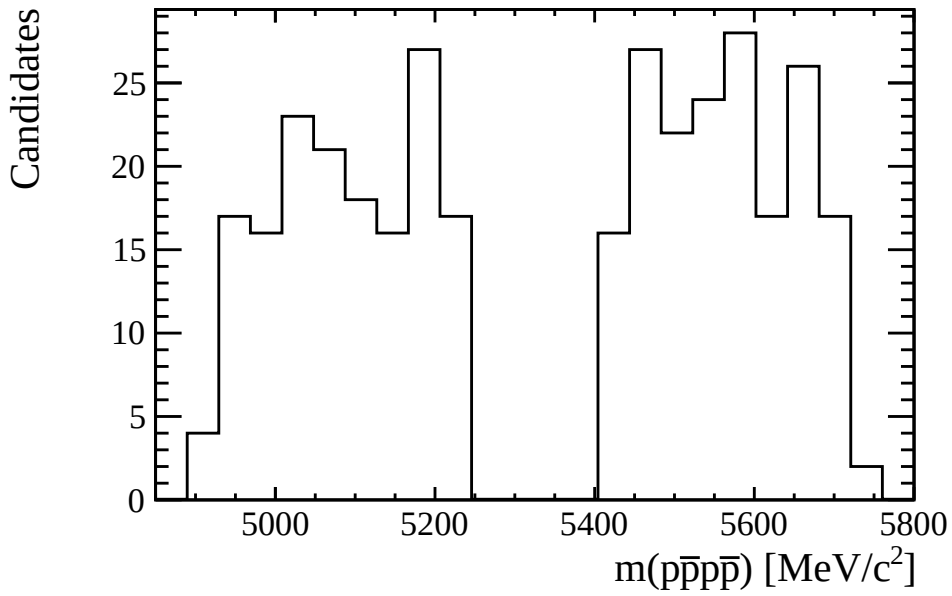

Figure 19: The invariant mass distribution of the sideband data after the application of the PID and the BDT cuts.

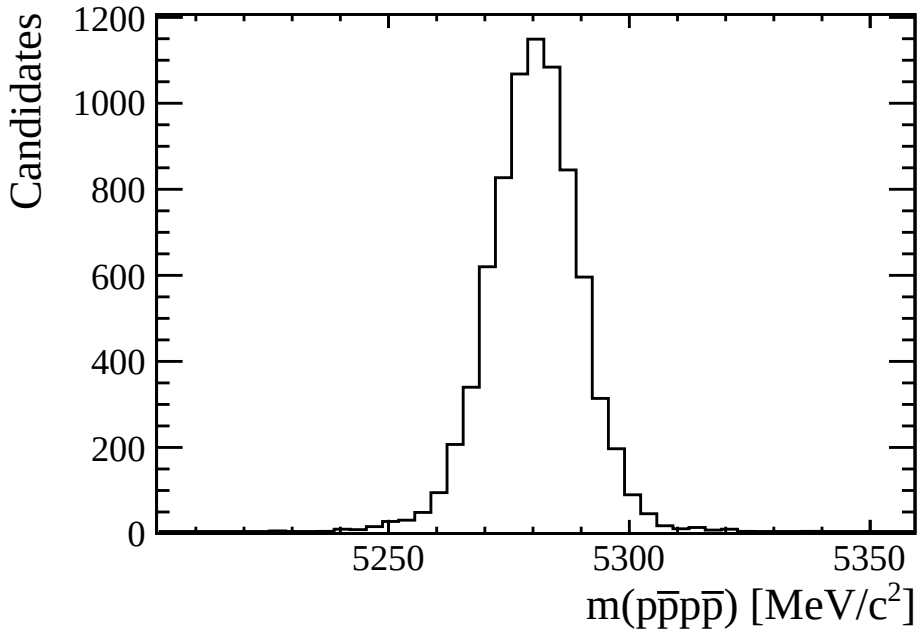

Figure 20: The invariant mass distribution of the $B^0 \to p\bar{p}p\bar{p}$ simulated sample after the application of the PID and the BDT cuts.

Table 7: Efficiencies of the various selection stages for both the simulated RunI sample and the RunI sideband data. The efficiencies are calculated with respect to the previous stage.

Selection Stage	background rejection [%]	ϵ signal from MC [%]
True-id	-	89.22
Trigger	96.07	25.45
PID	99.37	70.33
BDT	70.99	90.40
Total	99.9932	13.35

4 Conclusion

We have put into place a selection using adequate trigger requirements, a cut on a PID variable and used the BDT multivariate classifier. The PID selection was found to be extremely efficient at removing background. The BDT was useful to further distinguish signal from background. A two-dimensional optimization of the cut values on the PID and the BDT output was performed. At the optimal working point, the overall selection yields a significance of ≈ 20. Moreover, it was shown that very little contamination is expected from the $B^0 \to p\bar{p}hh'$ modes.

We have now confidence that the samples should be clean of most of the background. For the master's thesis, we will start with the mass fit and the study of the normalization mode ($B^0 \to p\bar{p}KK$). RunII data will also be incorporated.

References

[1] H.-Y. Cheng and J. G. Smith, *Charmless Hadronic B-Meson Decays*, Ann. Rev. Nucl. Part. Sci. **59** (2009) 215, arXiv:0901.4396.

[2] LHCb collaboration, A. A. Alves Jr. *et al.*, *The LHCb detector at the LHC*, JINST **3** (2008) S08005.

[3] LHCb collaboration, R. Aaij *et al.*, *LHCb detector performance*, Int. J. Mod. Phys. **A30** (2015) 1530022, arXiv:1412.6352.

[4] Particle Data Group, C. Patrignani *et al.*, *Review of particle physics*, Chin. Phys. **C40** (2016) 100001, and 2017 update.

[5] J. Beddow *et al.*, *Analysis of $B^0 \to p\bar{p}$ and $B_s^0 \to p\bar{p}$ with the Run I Data Sample*, .

[6] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and regression trees*, Wadsworth international group, Belmont, California, USA, 1984.

[7] Y. Freund and R. E. Schapire, *A decision-theoretic generalization of on-line learning and an application to boosting*, J. Comput. Syst. Sci. **55** (1997) 119.

[8] A. Hoecker *et al.*, *TMVA: Toolkit for Multivariate Data Analysis*, PoS **ACAT** (2007) 040, arXiv:physics/0703039.