

LHCb Turbo++ Slimming: Inclusive B Hadron Decay Product Selection Summer Student Report

David Bogensberger

Supervisor: Sascha Stahl

16.09.2016

Abstract

My summer student project focused on reducing the turbo++ output bandwidth for B meson decays at LHCb. The aim is to discard as many of the background tracks as possible, while still retaining almost all of the signal events. I analysed the ability to distinguish signal from background tracks, by considering four different B meson decays, while using signal definitions for just one particular decay, as well as for a general type of decay. I added tracks to a triggered turbo candidate, to reconstruct the B meson. I used a boosted decision tree and an artificial neural network to classify events, and compared the resulting ROC curves.

1 Introduction

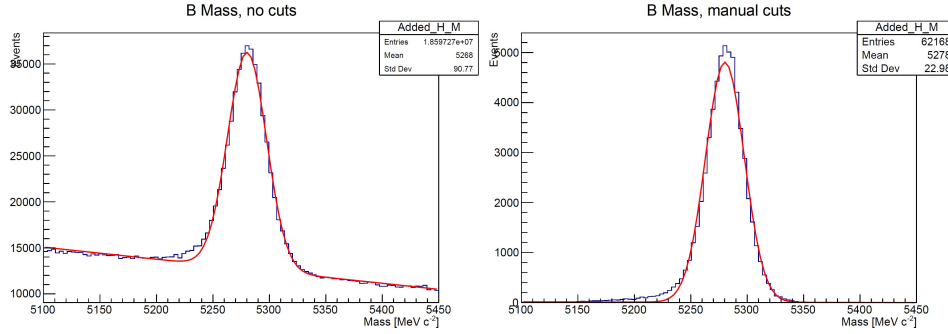
Turbo and turbo++ prepare the data from the LHCb high level trigger for physics analysis. Turbo only accepts triggered candidates, while turbo++ saves all reconstructed tracks. As a result of this, turbo++ gives us a greater freedom for analysis, as it allows us to add interesting tracks to turbo triggered candidates, in order to reconstruct their mother particle. However, turbo++ has a huge bandwidth, most of which is due to unrelated tracks. This is an unnecessary waste of storage space. The turbo++ bandwidth should therefore be slimmed down to a level in between that of turbo and turbo++, so that almost all interesting and potentially relevant tracks that would not have been saved by turbo are still retained, but most irrelevant tracks are discarded. My project was related to this on this effort. I analysed the ability to distinguish signal from background tracks, thereby reducing the turbo++ bandwidth. Throughout my project, I used 2016 data from the DiMuonJPsiTurbo line with PersistReco.

2 Cuts to crystallize B meson mass peaks

I added a hadron to a triggered J/ψ candidate, in order to reconstruct the B meson from the $B \rightarrow J/\psi K$ decay. As can be seen in figure 1(a), the peak corresponding to the reconstructed B meson mass lies on top of a locally linear mass distribution of irrelevant tracks. In order to get rid of as much of this background as possible, while trying to retain the peak distribution. I applied cuts to the following variables:

- Impact parameter (IP)
- Minimum impact parameter chi squared (MINIPCHI2)
- Vertex chi squared (VERTEXCHI2)
- Kaon probability (ProbNNk)

I used the shape of the distributions of these variables, and a knowledge of which region signal events should predominantly belong to, in order to guess reasonable cut values for them. The results of this can be seen in figure 1(b).



(a) B mass distribution without any cuts (b) B mass distribution with manual cuts

Figure 1: B mass distribution

By applying cuts on these variables, I was able to remove almost all of the background. But in doing so, I also removed most of the events in the peak as well. I attempted to optimize the ratio $\frac{S}{\sqrt{S+B}}$, where S is the number of signal tracks, and B is the number of background tracks. By modifying the cut values used, I managed to improve this ratio, but not significantly.

In order to improve the efficiency of this analysis, I merged the relevant branches of the ntuples I was using, and placed them into a new tree. Only events whose invariant B meson mass fell within a certain range, were accepted. I also improved the results and efficiency, by ensuring that for each J/ψ meson, only one added hadron would be accepted.

Afterwards, I combined exactly two hadrons that passed more lenient cuts, with the corresponding J/ψ meson. In this way, I was able to recon-

struct the B_s meson, and its mass peak, as $B_s \rightarrow J/\psi \phi$, where $\phi \rightarrow K^+ K^-$ almost instantaneously.

3 Using TMVA to optimize background rejection at high signal efficiency

In order to vastly improve my analysis, I used the Toolkit for Multivariate Data Analysis with ROOT (TMVA). By using cuts on the classifier output of TMVA, I was able to achieve a significantly better distinction of signal and background events. I first used a rectangular cut optimization, which already worked a lot better than my manual cuts. But it wasn't good enough, so I used a boosted decision tree (BDT) and an artificial neural network (ANN) instead.

The data has to be split into signal and background, so that TMVA can try to distinguish them. I considered two different signal definitions. At first I used a signal definition to classify the data for just that one particular decay, $B \rightarrow J/\psi K$. To do that, I used the following definition:

- The added hadron and the triggered J/ψ both had to have a background category (BKGCAT) of less than, or equal to 50.
- The KEY variable was used to ensure that the added hadron had the same mother, grandmother and great-grandmother particle as the triggered J/ψ .
- The identity (ID) of the mother particle of the added hadron and the J/ψ had to correspond to a B meson

And all the tracks that did not satisfy these criteria were saved in the background ntuple. But as I ultimately wanted to create classifiers that did not just work for one particular decay, but a whole range of similar decays, I had to use a more general signal definition. As the last two conditions only apply to the $B \rightarrow J/\psi K$ decay, I only used the BKGCAT condition to define the signal for a group of decays.

The variables that are to be used by TMVA to distinguish signal from background also need to be chosen carefully. Sufficiently many variables should be used to avoid overtraining, and to give the program sufficient freedom to use the most suitable variable at each node or neuron. But variables whose distribution is almost identical for signal and background events are not useful for the TMVA analysis. I therefore chose the following set of variables to be used, as almost all of them show a clear distinction in the signal and background distributions. The variables mainly belong to two groups: the kinematic variables, and the geometric variables.

- Transverse momentum (PT)
- Corrected mass (CORRM). I only used this for the data from the topological trigger.
- Impact parameter (IP)
- Impact parameter chi squared (IPCHI2)
- Minimum impact parameter chi squared (MINIPCHI2)
- Vertex chi squared (VERTEXCHI2)
- Direction angle (DIRA)
- Number of primary vertices (nPV)
- SAMEPV
- GHOST

In figure 2, I am comparing the transverse momentum distribution of signal and background for the two signal definitions I used. The figure indicates that the specific signal definition allows a better distinction of signal from background. This also applies to the other variables I used in TMVA. Therefore, it is understandable that the background rejection at a certain signal efficiency is significantly less for the general than for the specific decay.

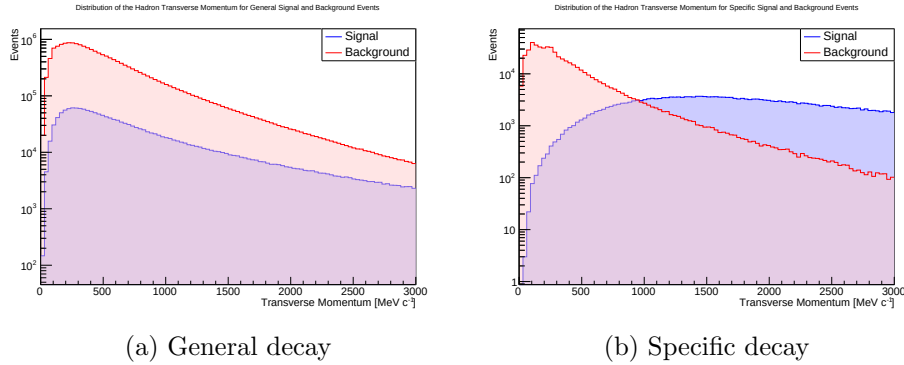


Figure 2: Signal and background transverse momentum distribution

4 $B \rightarrow J/\psi K$ Results

At first, I compared the ability of the BDT and the ANN to classify signal and background events, for both definitions of signal that I used. As expected from the above distributions, the classifier for the specific decay

works better than the one for the general decay. But the general decay classifier is a lot more useful, when considering a whole group of decays.

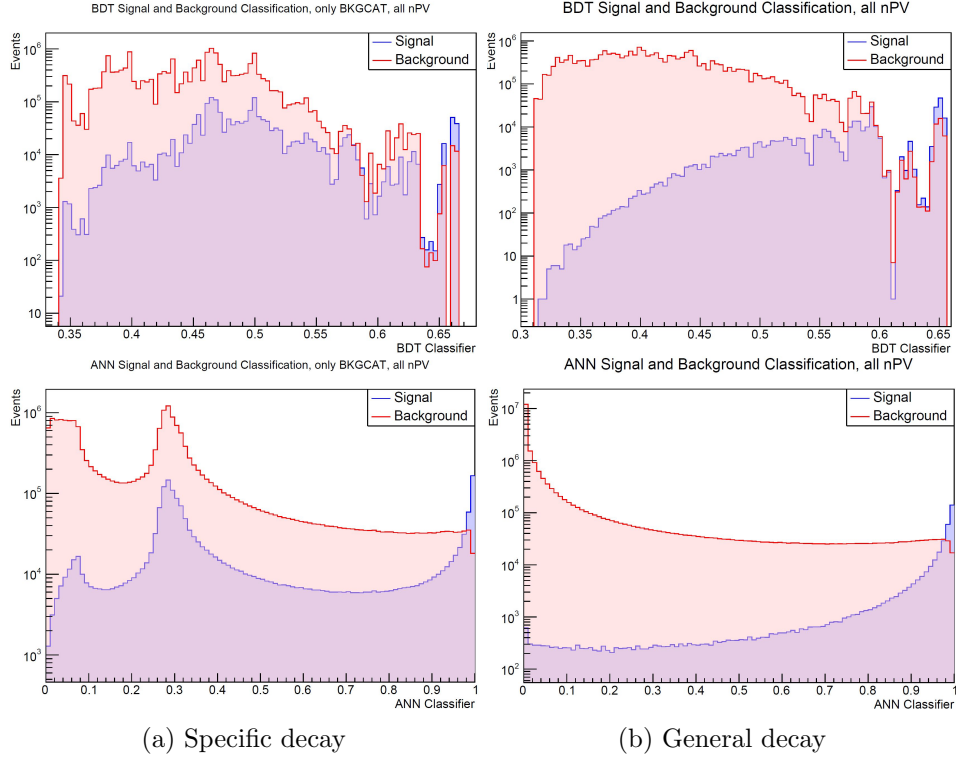


Figure 3: ANN and BDT classifiers

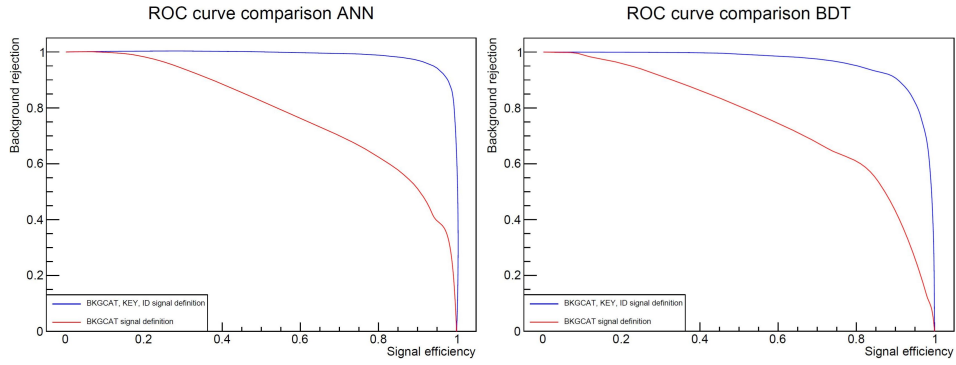
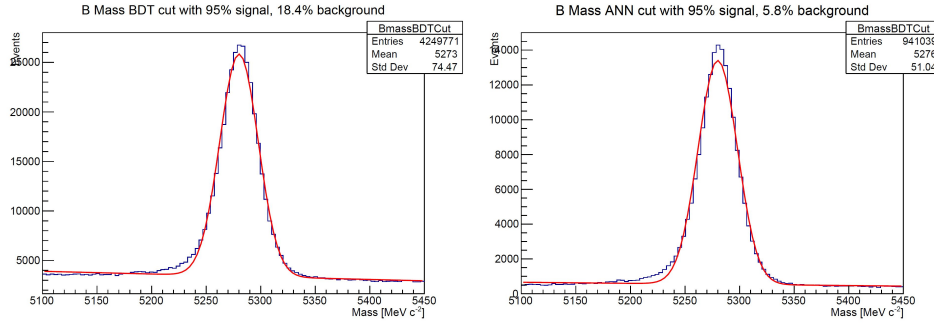


Figure 4: ROC curves

After establishing the classifier, I created receiver operating characteristic (ROC) curves to have a better comparison of the results. These can be seen in figure 4. At 95% signal efficiency, the specific decay signal analysed by the ANN achieved a background rejection of 94.2%. This corresponds to a reduction of the output bandwidth by 92.3%. In comparison, the BDT achieved 81.6% background rejection at 95% signal efficiency. This corresponds to a reduction of the output bandwidth by 80%. The general decays are obviously less capable of reducing the bandwidth at a fixed signal efficiency by as much. For the ANN, at 95% signal efficiency, the bandwidth can be reduced by 33.5%. The BDT can reduce it by 26.4%.



5 Topological Trigger Results

Next I used data from the topological trigger, for four different decays: $B \rightarrow J/\psi K$, $B_s \rightarrow D_s \mu \nu_\mu$, $B \rightarrow D_0 \mu \nu_\mu X$, and $B \rightarrow D \mu \nu_\mu X$. The topological trigger performs inclusive selection of B hadron decays by selecting two, three and four body signatures. As before, I added interesting tracks to the two, three, or four body candidate, to thereby reconstruct the B meson. At first, I used the general signal definition for all of these decays. I produced the ROC curves to analyse how much background rejection could be achieved at very high signal efficiency, and compared the results for two, three and four body decay modes. As the J/ψ usually decays almost instantaneously into two daughter particles, such as muons: $J/\psi \rightarrow \mu^+ \mu^-$, the detectors measure three outgoing particles from the same primary vertex. As I am adding one particle to the triggered particles from the topological trigger, I expected to see the best background rejection at a specified signal efficiency for the two body channel. The results didn't highlight this mode as clearly as I had expected, but at high signal efficiency it does dominate. At 95% signal efficiency, the background rejection was 41%, so that the bandwidth reduction was about 38.1%.

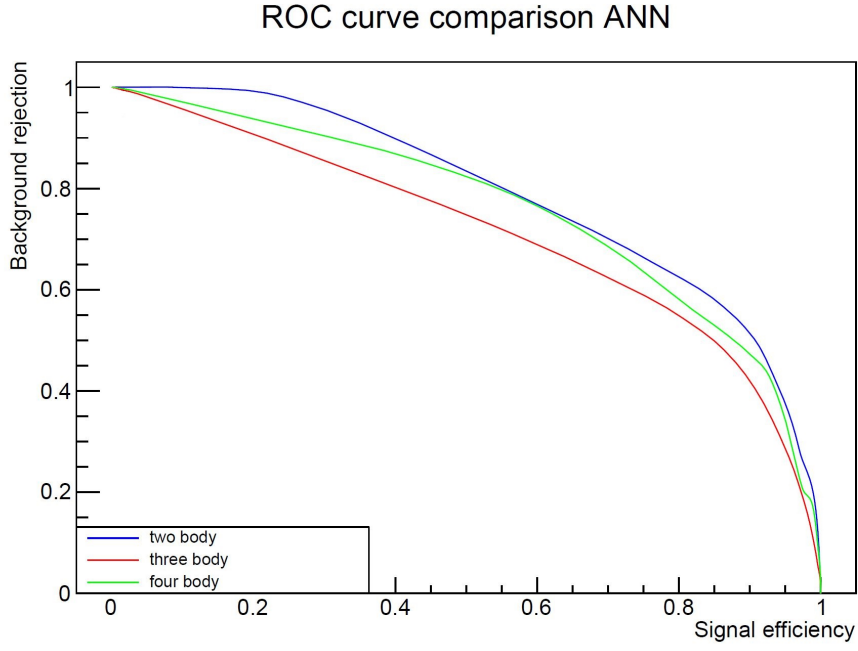


Figure 6: ROC curve for the $B \rightarrow J/\psi K$ decay, using data from the topological trigger

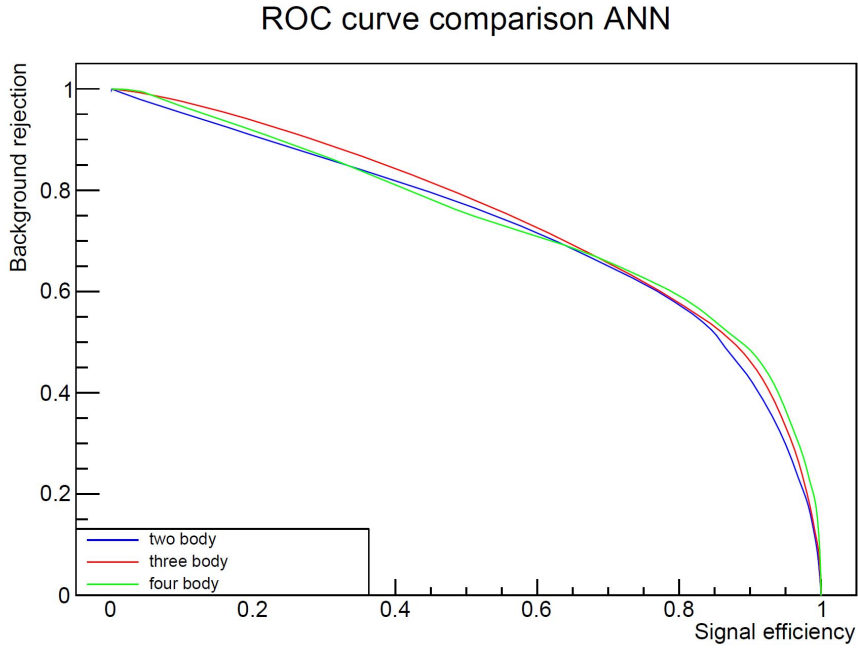


Figure 7: ROC curve for the $B_s \rightarrow D_s \mu \nu_\mu$ decay, using data from the topological trigger

I then used data corresponding to the $B_s \rightarrow D_s \mu \nu_\mu$ decay, once again split into two, three and four body decay modes. This decay produces 4 measurable particles, as $D_s \rightarrow K K \pi$, and the ν_μ is not measured by the detectors. Therefore it could be expected that the signal and background distinction works best for a three body decay mode in the topological trigger. But the results for two, three and four body modes are almost identical, as can be seen in figure 7.

The neutrino produced in this decay is not measured by the detectors. Another, potentially more important factor that led to these results is the relatively long lifetime of the D_s meson. As a result, the secondary vertex of its decay into two Kaons and a Pion is displaced from the primary vertex of the $B_s \rightarrow D_s \mu \nu_\mu$ decay, by a measurable distance. As a result, the BKGCAT distinction of input signal and background in the TMVA probably didn't work as well as for the previously studied decay.

For the last two decays I considered, $B \rightarrow D_0 \mu \nu_\mu X$, and $B \rightarrow D \mu \nu_\mu X$, I was only able to analyse the two and three body modes. For these decays, a slightly modified BDT achieved the best background rejection at a given signal efficiency, as can be seen in figure 10. Figure 8 indicates that both two and three body decay modes performed similarly for these two decays.

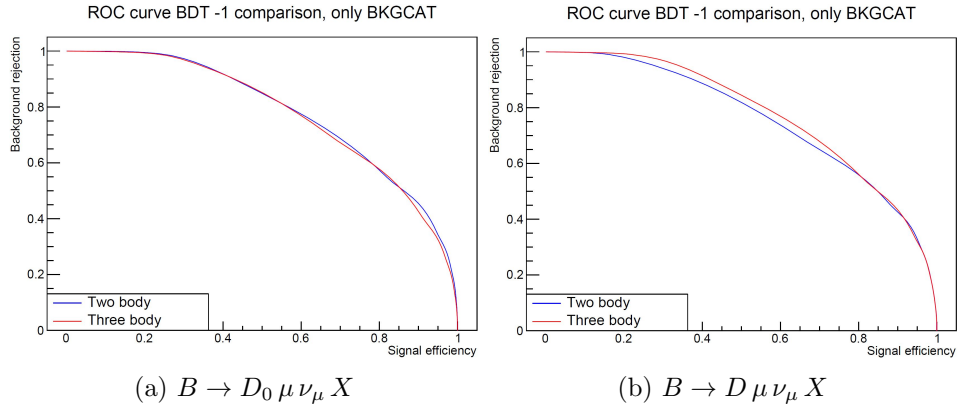


Figure 8: Two and three body ROC curves

For the three decays $B_s \rightarrow D_s \mu \nu_\mu$, $B \rightarrow D_0 \mu \nu_\mu X$, and $B \rightarrow D \mu \nu_\mu X$, I also used a signal definition focused on that particular decay. As the D_s , D_0 and D mesons usually decay to $K^+ K^-$ and some other particles, I used a signal definition that required the two kaons to have had the same mother, grandmother, and great grandmother particles. I also required their mother to have been a D_s , D_0 , or D meson, dependent on the decay I was considering. I also ensured that their grandmother was a B or B_s meson.

However, only very few tracks managed to pass all these criteria, to get identified as being part of the signal. TMVA therefore struggled to distinguish signal from background. So the more general signal definition

based only on BKGCAT actually achieved a better distinction of signal and background, as can be seen in figure 9.

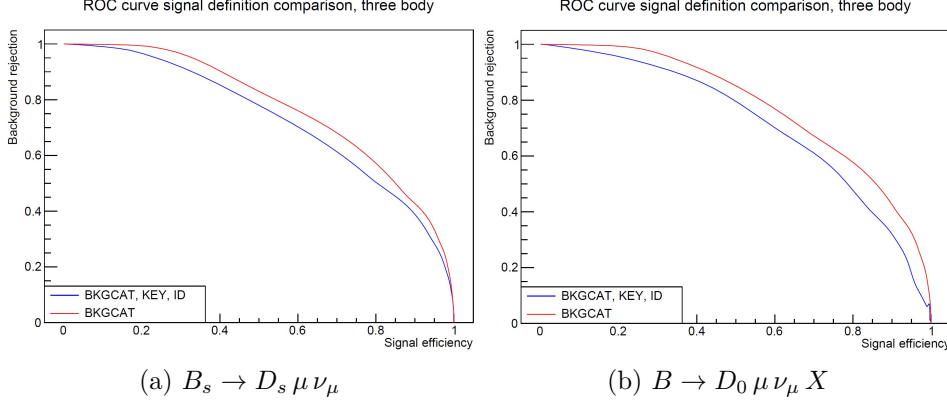


Figure 9: Comparison of signal definitions

I also compared the different TMVA methods that I used to classify events for these last three decays. By setting $nCuts = -1$ in the options of the BDT, I was able to achieve the best performance. In order to get the best results possible, all of these last few ROC curves were produced with this BDT option, and the general signal definition.

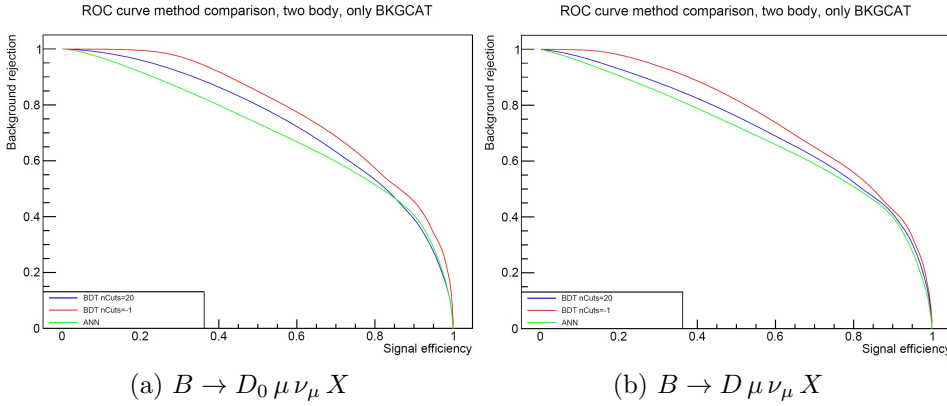


Figure 10: Comparison of TMVA methods

Finally, I tested the ability to use the classifier trained for one decay, on other decays. I used the $B \rightarrow D \mu \nu_\mu X$ decay data, and evaluated the ROC curve when using its own classifier, as well as the one I obtained for the other two decays. I only used BKGCAT to define the signal, as any of the other definitions are specific to one particular decay. As can be seen in figure 11, the ROC curves are almost identical. This is a great result, as it means that the turbo++ bandwidth can be slimmed down for a large group of decays

with just one classifier.

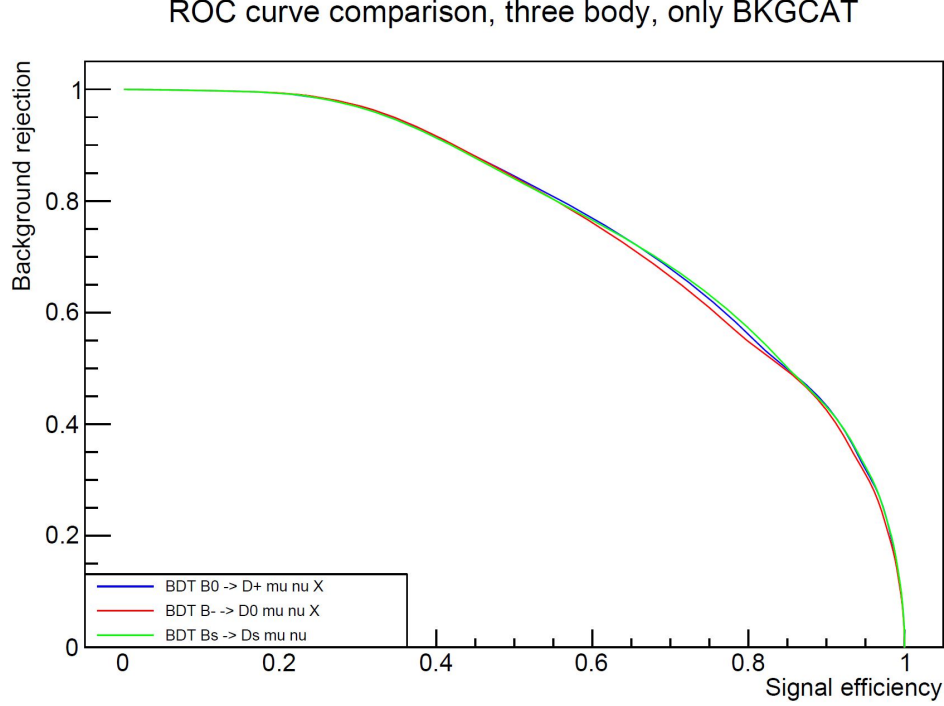


Figure 11: ROC curve of $B \rightarrow D \mu \nu_\mu X$ data, using its own BDT classifier, as well as classifiers of two other decays

6 Conclusion

In conclusion, turbo++ can be slimmed down significantly by using a TMVA classifier to distinguish signal from background events. The classifier can be trained for one particular decay, by defining the signal appropriately. The signal definition can however also be specified to apply for any similar decay, by only using the BKGCAT variable. Given enough signal data, at 95% signal efficiency, the background rejection for the specific $B \rightarrow J/\psi K$ decay was found to be 94.2%, resulting in a reduction of the turbo++ bandwidth by 92.3%. But for the more general signal definition based only on BKGCAT, the results are significantly worse, but are still a clear improvement compared to the unslimmed turbo++ stream. At 95% signal efficiency, the bandwidth reduction is usually about 33%.

BKGCAT works well, except when one of the B meson decay products travels a measurable distance before decaying. But by only using BKGCAT to separate signal from background tracks, the classifier produced by TMVA can be applied to a large group of other decays, without a loss of bandwidth reduction at a specific signal efficiency.