LHCb-INT-2012-025
18 September 2012

Ghost probability: an efficient tool to remove background tracks

J. Brehmer[1], J. Albrecht[1], P. Seyfert[2]

[1] *CERN, Geneva, Switzerland*
[2] *Physikalisches Institut, Ruprecht-Karls-Universität Heidelberg, Heidelberg, Germany*

Abstract

A multivariate classifier separating misreconstructed tracks (ghost tracks) from good tracks is presented. This classifier is named ghost probability and combines information from different stages of the track reconstruction and from global event properties. With an efficiency loss of less than 0.5% per track, 70% of the ghost tracks can be removed on Monte-Carlo samples. This good performance on simulation is validated on different data samples recorded in 2012. The ghost probability is already implemented in the selections of a large set of physics analyses.

Contents

1 Introduction

The track reconstruction algorithms of LHCb have been tuned for an optimal signal efficiency, allowing for a significant rate of fakes ("ghost tracks"). In the current reconstruction (`Reco13` – `Reco13d`), the average ghost rate for long tracks is 27%, whereas for tracks identified as muons it is 17%. Both these ghost rates have been measured on simulated events (MC/2012) and are known to be higher in data.

The most common criterium to clean up the reconstructed tracks is the quality as measured in the Kalman filter based track fit (typically the fit χ^2/ndf or prob(χ^2) are used). This criterium, however, introduces significant inefficiencies when a high purity is required. A more efficient way to achieve high purities is presented here: a multivariate classifier, the ghost probability (GP), which is based on a neural network implemented in the ROOT framework TMVA [1] and trained on a Monte-Carlo (MC) sample based on 2012 conditions. 22 different tracking variables are combined into one value between 0 and 1 for each track, where 0 is most signal-like and 1 is most ghost-like.

The ghost probability developed in this note is based on a similar quantity that has been developed in 2009 [2]. While the structure of this old classifier is similar to the one presented here, the set of input variables was re-evaluated to optimally exploit the track properties as given in the current track reconstruction software and the more realistic detector simulation. The performance of the updated GP classifier is compared to the MC09 version as well as to other measures of the track quality.

This note is organised as follows. In section 2, the input variables to the classifier for fitted long tracks are presented and its architecture is discussed. The performance on MC samples of long tracks is evaluated and compared to that of other classifiers in section 3. Section 4 contains a first validation on LHCb data from different working groups. In section 5, the extension of this classifier for long tracks to other track types is presented. We give our conclusions in section 6.

2 The ghost probability classifier for long tracks

2.1 Input variables

The input to the ghost probability classifier was defined with an inclusive J/ψ sample generated under the MC2012 / April conditions[1]. Tracks passing the criteria required for StdLoosePions (track $\chi^2/ndf < 3$, IP $\chi^2 > 4$, $p_T > 250$ MeV, Kullback-Lieber clone distance = 5000) were used to optimize the GP for generic signal candidates from B or D decays. Samples of good tracks ($\sim$ 1.4 million), which could be associated to a generated particle, and "ghost tracks" ($\sim$ 400 000), for which this association failed, were used.

An optimal set of discriminating variables was chosen so that a neural network using these variables has a maximal separation power, i. e. cuts on the resulting classifier are

[1]The sample was simulated with `Sim06` (with Nu = 2.0), digitized with `Digi12` (without spillover), reconstructed with `Reco13a`, and analysed with `DaVinci v32r1p2`.

highly efficient on the good tracks, while rejecting as many ghost tracks as possible. The input variables for the GP classifier were restricted to be tracking variables, potentially discriminating PID information was not used. It can be utilized in a different classifier which will be trained at a later stage.

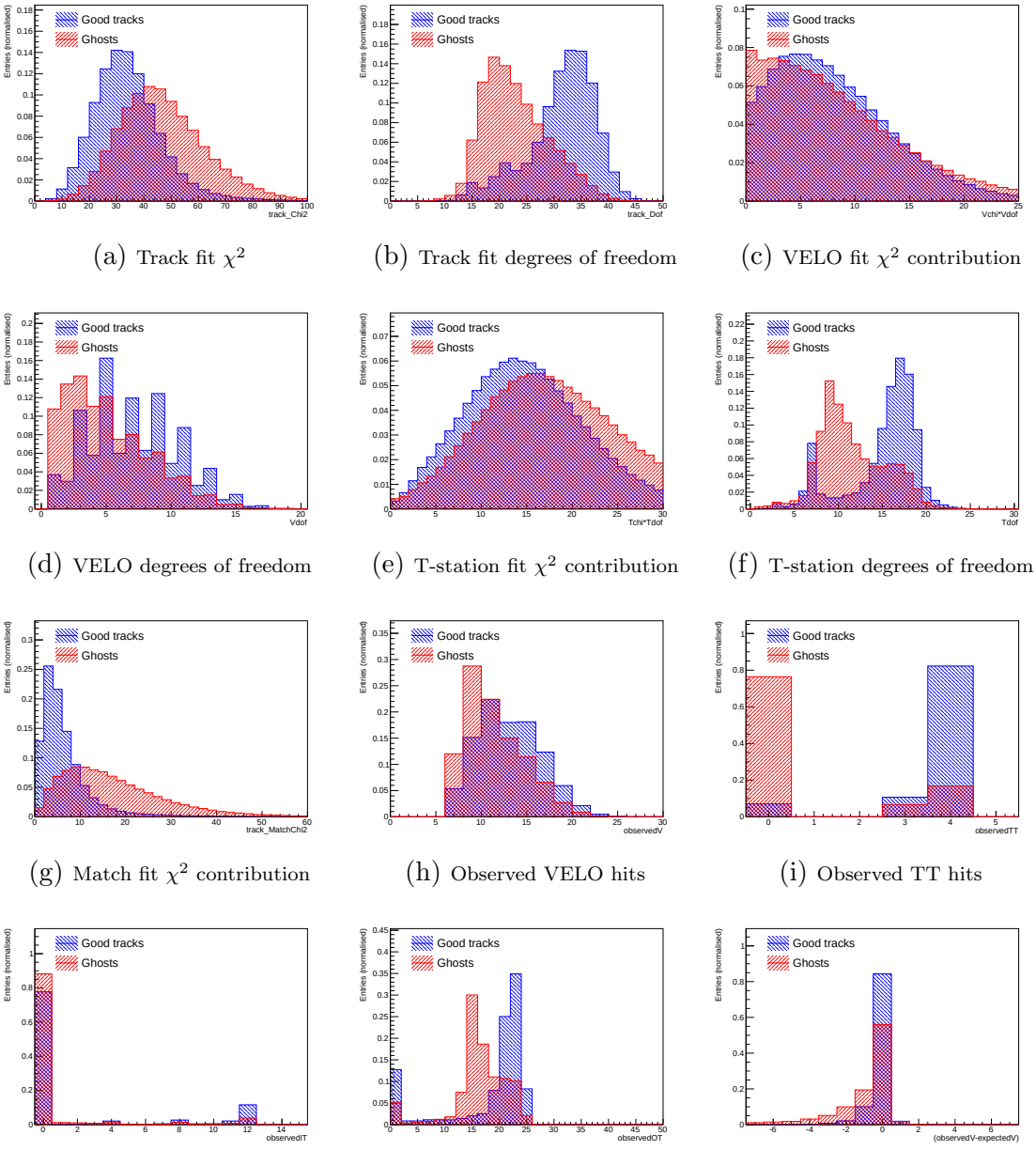
A set of 22 input variables for the neural network was found to be optimal:

- the total track fit χ^2 and the corresponding number of degrees of freedom,
- the track fit χ^2 contribution from the VELO and T-station segments and the corresponding numbers of degrees of freedom,
- the track fit χ^2 contribution from the match between VELO and T-station segments,
- the number of observed hits assigned to the track in each of the VELO, TT, IT, and OT detectors,
- the excess of the observed number of hits over the number of hits expected from geometry in each of the VELO, TT, IT, and OT,
- the sum of the number of OT hits that were rejected by the track fit ("outlier") and the number of OT hits where the measured drift time was inconsistent with the track hypothesis and therefore not used for the fit ("bad drift time"),
- the total hit multiplicities in the VELO, TT, IT, and OT,
- the pseudorapidity η and
- the transverse momentum p_T.

These variables contain separation power between good tracks and ghosts, and removing any of these input variables leads to a decline of the performance of the classifier. The distribution of these quantities for good tracks and ghosts is displayed in Fig. 1 and 2. Of these input variables, the total track fit χ^2 and the excess of the observed over the expected hits in the different detectors are most relevant to discriminate good tracks from ghosts. A classifier only using these five variables and the pseudorapidity as input already gives a good performance and should be considered if a simpler neural network with only six input parameters is needed.

For some of the variables, alternative parameterisations have been examined. Instead of the separate use of χ^2 and number of degrees of freedom, the use of reduced χ^2 values or prob(χ^2) values has been tested and leads to a worse performance. Similarly, using only one total hit multiplicity instead of four values for the different tracking systems, or using the expected number of hits instead of the excess of observed hits over expected hits decreases the separation power of the GP classifier.

Other quantities that have been tested as input variables include the type of the track reconstruction algorithm (`PatForward` or `PatMatch`), the hit multiplicity in the SPD, the polar angle φ and the track position (x, y) at the T stations. Including any or all of these has been found not to lead to a gain in performance, hence these variables are not used in the classifier.



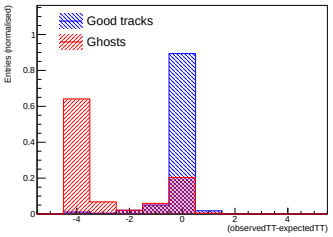
(a) Track fit χ^2 (b) Track fit degrees of freedom (c) VELO fit χ^2 contribution

(d) VELO degrees of freedom (e) T-station fit χ^2 contribution (f) T-station degrees of freedom

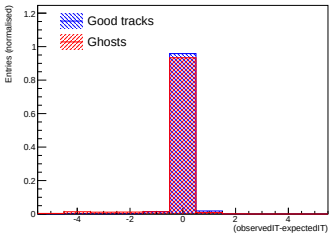
(g) Match fit χ^2 contribution (h) Observed VELO hits (i) Observed TT hits

(j) Observed IT hits (k) Observed OT hits (l) VELO hits − expectation

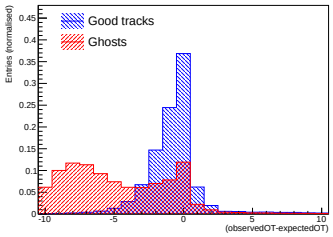
Figure 1: Distribution of input variables to the classifier for good tracks (blue) and ghosts (red) in the training sample.



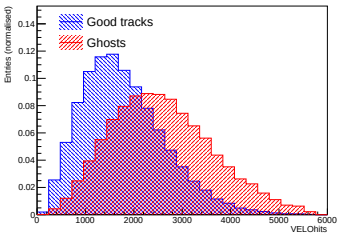
(a) TT hits − expectation



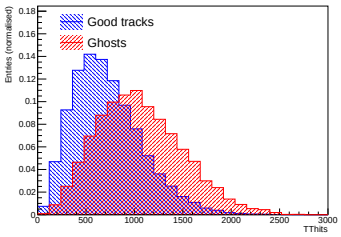
(b) IT hits − expectation



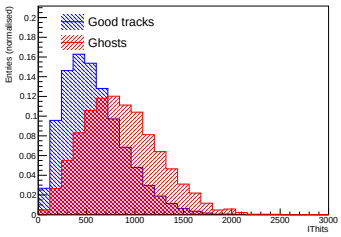
(c) OT hits − expectation



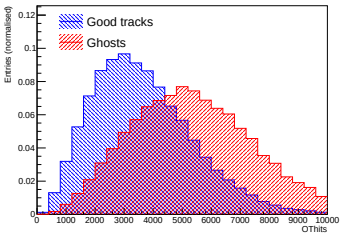
(d) VELO total hit multiplicity



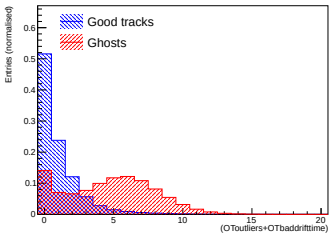
(e) TT total hit multiplicity



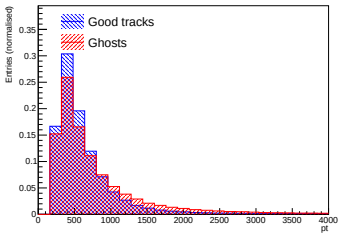
(f) IT total hit multiplicity



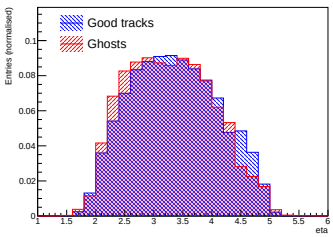
(g) OT total hit multiplicity



(h) OT outliers + bad drift times



(i) Transverse momentum



(j) Pseudorapidity

Figure 2: Distribution of input variables to the classifier for good tracks (blue) and ghosts (red) in the training sample (continued).

2.2 Architecture of the classifier

The ghost probability classifier is based on a multilayer perceptron (MLP) neural network. It is implemented using the ROOT framework TMVA [1]. Other classifier types, including boosted decision trees, have been tested, but did not give a better performance. The classifier was trained on an inclusive J/ψ MC sample as described in the previous section.

The ghost probability classifier returns values in the range $[0, 1]$ in such a way that the output values for the ghost tracks of the training sample are evenly distributed. In this way, a cut on the classifier at a value x leaves a fraction x of the ghost tracks.[2] This flat output was realised by transforming the original output of the neural network with a lookup table based on the Brunel class `Rich::TabulatedFunction1D` [3]. It was checked that this flattening procedure leaves the discrimination power of the classifier unchanged. The distribution of classifier values for good tracks and ghosts of the MC training sample is shown in Fig. 3.

2.3 Implementation in the LHCb software

The ghost probability classifier for long tracks is part of the `Tr/TrackTools` package from version v4r18 onwards and used by the `TrackAddNNGhostId` algorithm. This is included in the `Reco13e` and `Reco14` reconstruction sequence. The classifiers for the other track types will be part of `TrackTools v4r19`, which can be used to evaluate the ghost probability in `DaVinci` either by running the `TrackAddNNGhostId` algorithm or using the `TrackNNGhostId` tool directly.

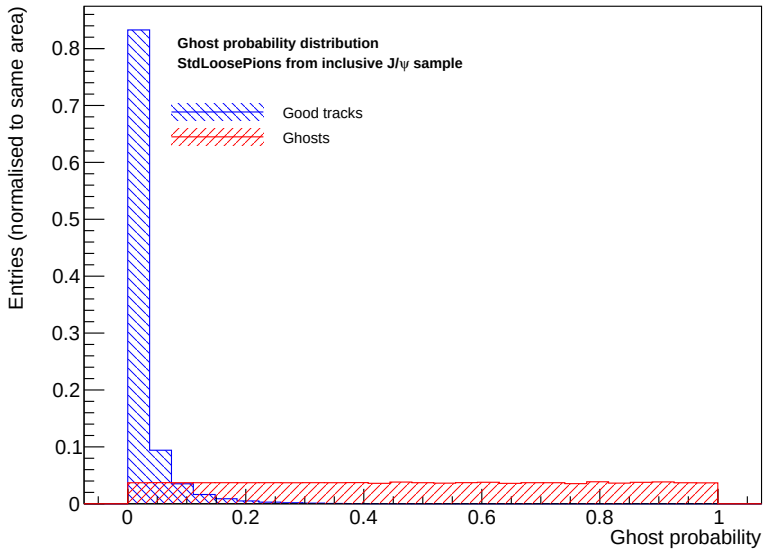
3 Classifier performance

3.1 Discrimination power

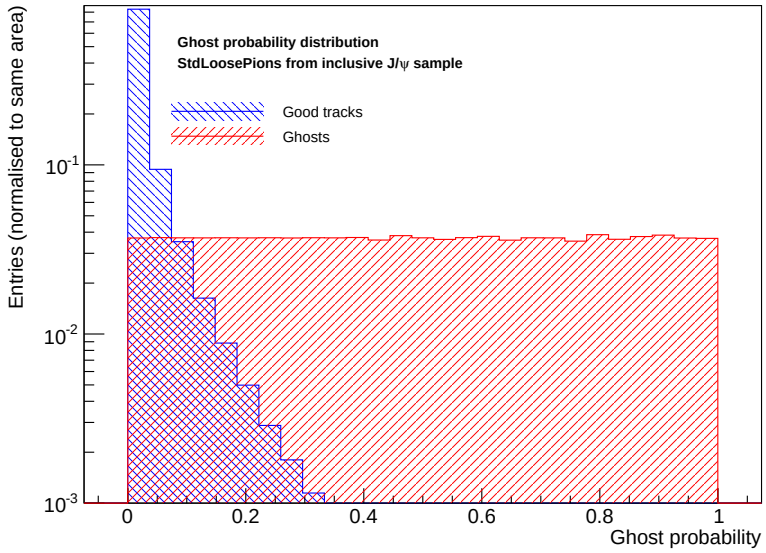
The performance of the classifier was evaluated on different MC samples, including StdLoosePion tracks from an inclusive B sample based on 2012 conditions, the inclusive J/ψ sample[3] and StdAnyNoPIDsPion tracks from an MC11 minimum bias sample. For a given sample, the good track (signal) efficiency and the ghost (background) rejection of different GP cuts were calculated. The resulting receiver operating characteristics curve (ROC curve) was compared to that of other track quality variables, such as the old ghost probability variable from 2009, the χ^2/ndf of the track fit, the track likelihood [4] and the ANNghost variable [5]. Alternative versions of the GP classifier have been constructed and evaluated, including a classifier without the kinematic input variables p_T and η and a classifier trained on StdLoosePions from an inclusive B MC sample, also based on April 2012 conditions. Fig. 4 shows the ROC curves of different classifiers on the J/ψ sample. The evaluation on other samples yields comparable results.

[2] Of course, this statement is only exact for the training sample. The GP distribution for ghosts in data will not be completely flat, see for instance Fig. 7.

[3] The sample has been split up in two equal halves for training and evaluation.



(a) On a linear scale



(b) On a logarithmic scale

Figure 3: Distribution of ghost probability output values for good tracks and ghosts on long tracks.

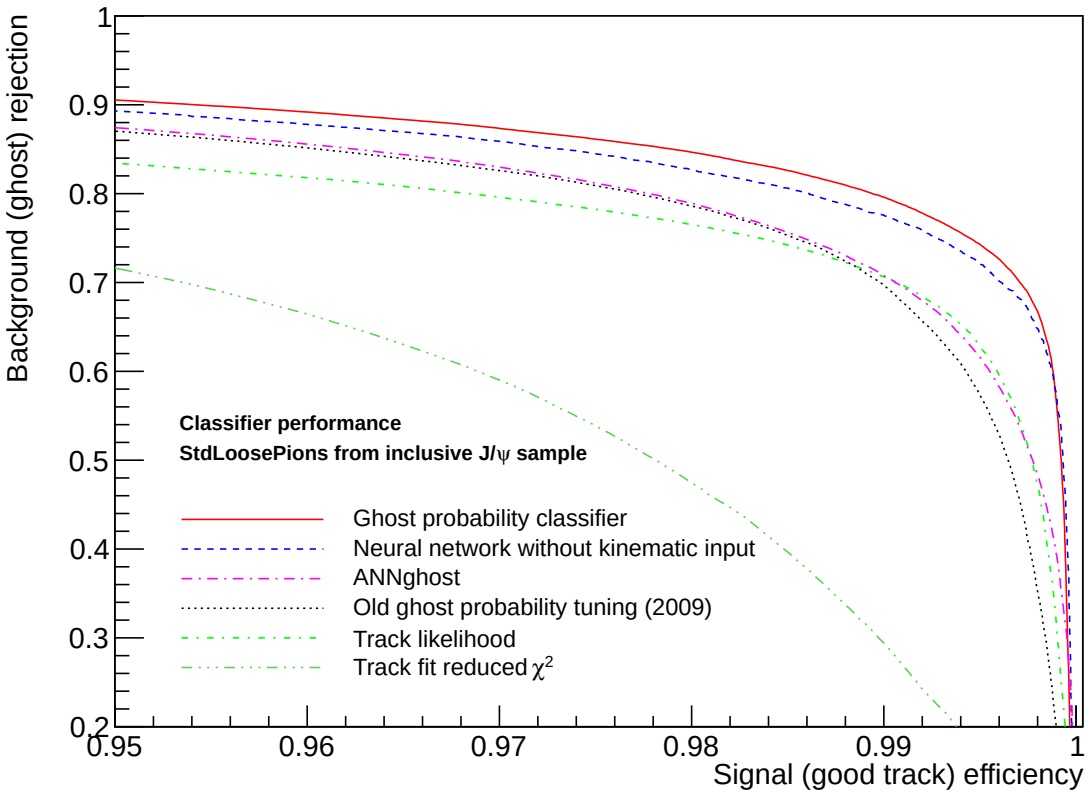


Figure 4: ROC curve displaying the signal efficiencies and background rejection of cuts on different classifiers. The ghost probability presented here is displayed in red. Cutting on it is more efficient than cutting on any of the other quantities.

We found that cutting on the ghost probability classifier gives a very good performance. On the inclusive J/ψ sample, the cut $GP \leq 0.3$ removes 70% of the ghost tracks, but leads to a signal loss of only 0.29% per track. The high signal efficiency and significant ghost reduction mean that this cut is very well suited for analysis purposes. The looser cut $GP \leq 0.5$ leads to a ghost rejection of 50% and a negligable signal loss of 0.08% per track.

In comparison the other classifiers perform worse. For the same signal efficiency, cuts on the track likelihood or the track fit reduced χ^2 give a much lower ghost rejection rate, see Fig. 4. The GP classifier is also an improvement over the old ghost probability variable from 2009. The ANNghost variable uses this old ghost probability classifier in combination with PID information. It will give a better performance once it is retuned to 2012 conditions and makes use of the GP classifier presented here.

3.2 Bias effects

In addition to the separation power tests, it was checked whether cutting on GP introduces bias effects. For the cuts $GP \leq 0.3$ and $GP \leq 0.5$, the signal efficiencies in different bins of the transverse momentum p_T, the total hit multiplicity in the SPD, the pseudorapidity η and the polar angle φ were calculated. The results are shown in Fig. 5. The signal efficiency for the GP cuts is high ($> 99\%$) over the whole range of all the variables analysed. It decreases slightly with an increasing total hit multiplicity. Not surprisingly, the ghost rejection also depends on these variables; the corresponding plots can be found in appendix A.

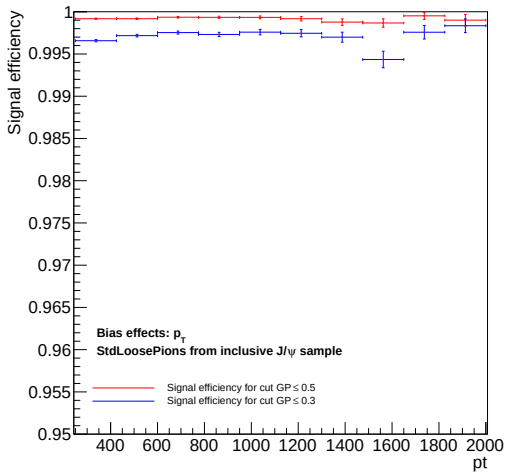
The same procedure was performed for cuts on the track fit reduced χ^2 and a version of the GP classifier that does not use the kinematic variables p_T and η as input. These cuts were chosen to give the same signal efficiency as the cut $GP \leq 0.3$, but yield a significantly lower ghost reduction rate. It was found that the signal efficiency of these other classifiers is not perfectly flat in the analysed variables either. In fact, the bias effects seem to be less pronounced for the ghost probability classifier than for the other quantities. Notably, the neural network without kinematic input variables does not perform better in this regard and leads to a stronger η bias than the ghost probability cut.

3.3 Performance for different particles and different ghost categories

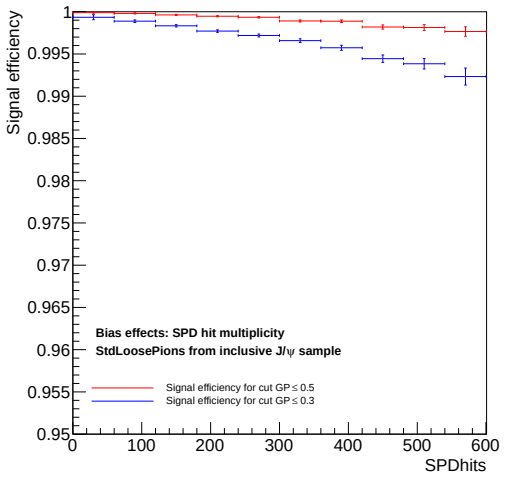
It was checked how efficient the cuts $GP \leq 0.3$ and $GP \leq 0.5$ are for different particles and different categories of ghosts. The results can be seen in Fig. 6. These cuts give approximately the same efficiency for muons, pions, kaons and protons, but lead to a lower efficiency for electrons. This is not surprising, as electron tracks are less clean due to the large amount of bremsstrahlung.

Ghosts can be classified in different categories, see [6]. Four such categories are relevant: ghosts from hadronic interaction of particles with the detector, misreconstructed VELO segments, misreconstructed T-station segments and ghosts from the mismatch of a VELO segment belonging to one particle with the T-station hits caused by another particle. Before the cut on the ghost probability, the dominant ghost source are hadronic interactions, followed by the mismatch between VELO and T-station segments and T-station misreconstructions. However, ghosts from hadronic interactions and T-station ghosts are greatly reduced by cutting on the ghost probability, hence the largest fraction of the remaining ghosts is of the mismatch type.

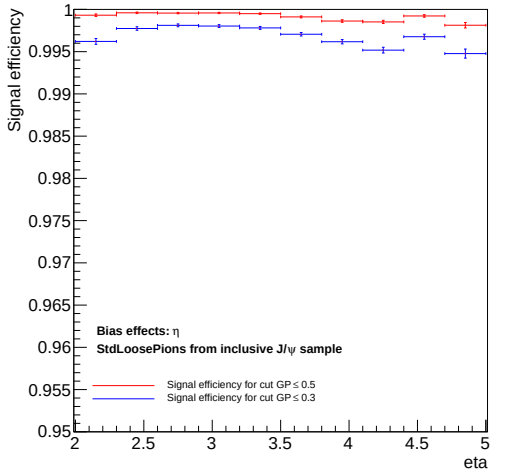
Comparing this cut on the ghost probability to cuts on the old tuning of the ghost probability or on the track fit reduced χ^2 with the same signal efficiency, it was found that the ghost probability reduced ghosts more strongly than the other classifiers in each ghost category.



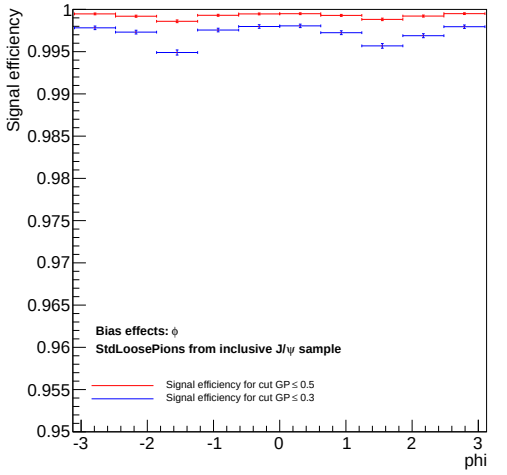
(a) As a function of p_T



(b) As a function of the total hit multiplicity in the SPD



(c) As a function of η

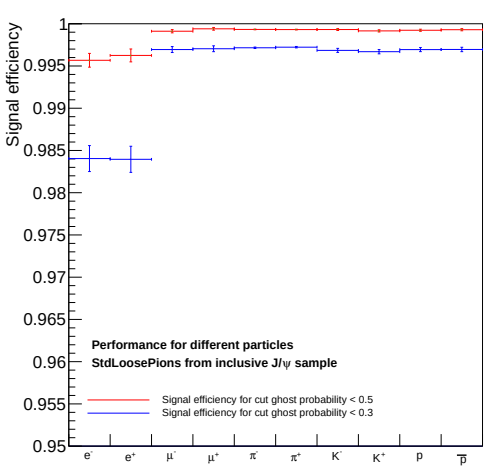


(d) As a function of φ

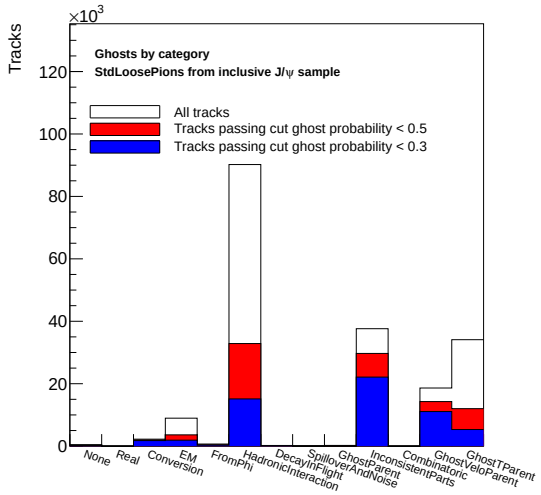
Figure 5: Signal efficiency as a function of different variables for the cuts $GP \leq 0.5$ and $GP \leq 0.3$ on StdLoosePions from an inclusive J/ψ MC sample under 2012 conditions. Note the strongly suppressed scale on the y axis. The signal efficiencies are high over the whole range of these variables, and the bias effect is less pronounced than for other classifiers.

4 Physics validation

The GP classifier was evaluated on 2012 LHCb data samples from the $B_s^0 \to \mu^+\mu^-$ working group, which contain roughly 445 pb^{-1} of data recorded until August 2012. The selections of the $B_s^0 \to \mu^+\mu^-$ control channels $B^+ \to J/\psi\, K^+$, $B \to hh$ and $B_s^0 \to J/\psi\, \phi$ were used,



(a) Cut efficiencies for different particles



(b) Distribution of ghost categories before and after ghost probability cuts

Figure 6: Performance of the cuts $GP \leq 0.5$ and $GP \leq 0.3$ for tracks from different MC particle and ghost types.

details can be found in [7].

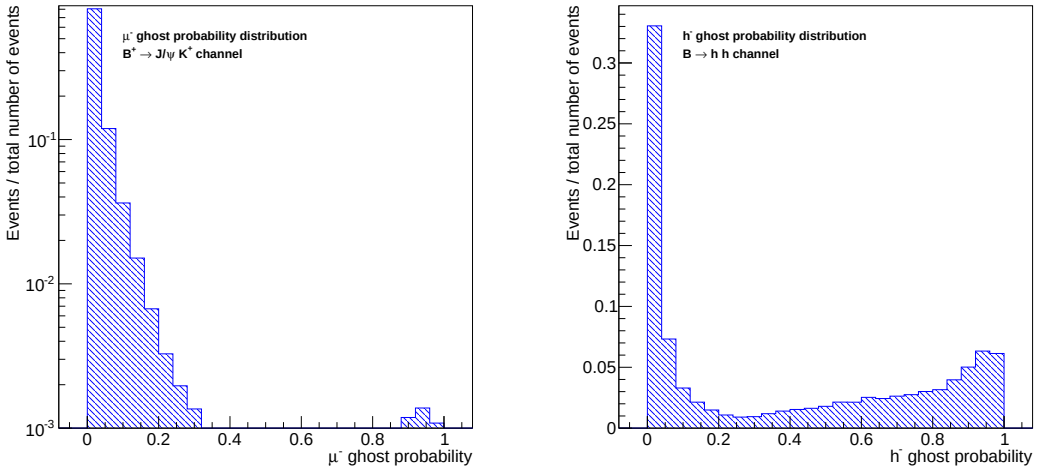
4.1 Signal efficiencies at the $B^+ \to J/\psi\, K^+$ resonance

The $B^+ \to J/\psi\, K^+$ channel contains a very clean resonance peak with no significant background, see Fig. 8. This is reflected in the GP distribution of the tracks, which is almost fully contained in the range [0, 0.3]. As an example the GP distribution of the μ^- track is shown in Fig. 7a.

In order to evaluate the signal efficiency of ghost probability cuts, the data after offline data selection, but before ghost probability cuts was compared to the data surviving the cut $GP \leq 0.3$ on the μ^- track, on the μ^+ track, on the K^+ track and on all tracks, respectively. As a comparison, the events surviving the cut track fit reduced $\chi^2 \leq 2$ were also evaluated. For each cut, the number of signal events was computed by subtracting the number of events in the invariant mass[4] sidebands $[m_{B^+} - 120$ MeV, $m_{B^+} - 80$ MeV$]$ and $[m_{B^+} + 80$ MeV, $m_{B^+} + 120$ MeV$]$ from the number of events in the signal region $[m_{B^+} - 40$ MeV, $m_{B^+} + 40$ MeV$]$, where $m_{B^+} = 5281.7$ MeV was determined from a fit to the data. The signal efficiencies of the different cuts, i. e. the ratio of signal events after a cut to the number of signal events before the cut, are given in Table 1. The invariant mass distributions of the events before and after the cut $GP \leq 0.3$ on all tracks are shown in Fig. 8.

The measured signal efficiencies for the GP cuts are high. Even cutting on the ghost

[4]Throughout this note, natural units are used.



(a) μ^- tracks in the $B^+ \to J/\psi\, K^+$ channel (b) h^- tracks in the $B \to hh$ channel

Figure 7: GP distributions of two different channels from 2012 LHCb data. The $B^+ \to J/\psi\, K^+$ channel (left) features a clean resonance without significant background (note the logarithmic scale), while in the $B \to hh$ channel (right) the background from ghosts is important.

Cut	Signal efficiency
$GP \leq 0.3$ on the μ^- track	99.78% ± 0.01%
$GP \leq 0.3$ on the μ^+ track	99.79% ± 0.01%
$GP \leq 0.3$ on the K^+ track	99.55% ± 0.02%
$GP \leq 0.3$ on all tracks	99.11% ± 0.02%
track fit reduced $\chi^2 \leq 2$ on all tracks	92.90% ± 0.06%

Table 1: Signal efficiencies of different ghost-reducing cuts at the $B^+ \to J/\psi\, K^+$ resonance peak in 2012 LHCb data as described in the text. The uncertainties are purely statistical and have been calculated as described in [8]. The cuts on the ghost probability were found to be very efficient.

probability of all three tracks only leads to a signal loss of 0.89%. Especially the cuts on the muon tracks are very efficient (99.8% per track), while cutting on the K^+ track leads to a slightly higher loss of signal events (signal efficiency 99.6%). The cut on the reduced χ^2 of the tracks fits gives a significantly worse signal efficiency of only 92.9%.

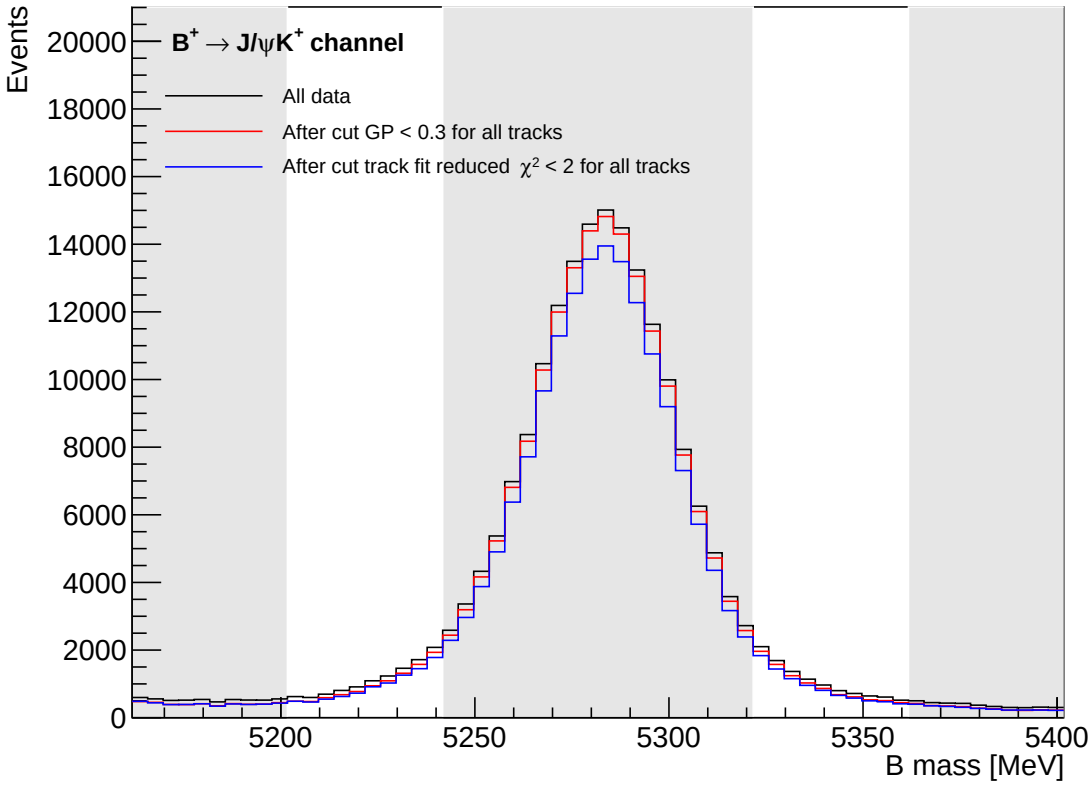
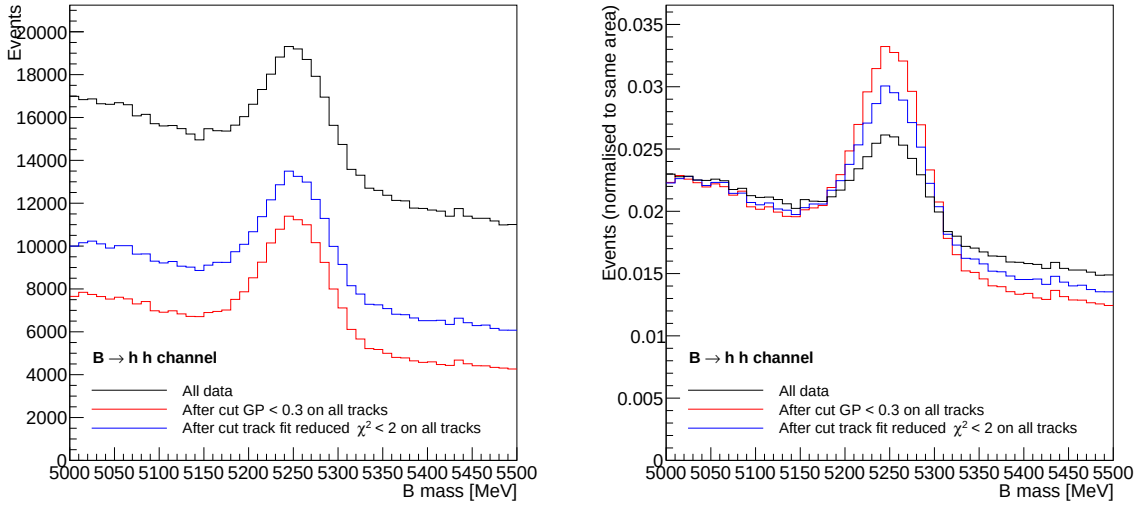


Figure 8: Events from 2012 LHCb data in the $B^+ \to J/\psi K^+$ channel, before (in black) and after (in red) the cut ghost probability ≤ 0.3 on all three tracks. Signal region and sidebands that are used to determine the number of signal events are shaded in grey. By simply subtracting the sideband events from the signal region events, the signal efficiency of the GP cut was determined to be 0.9911 ± 0.0002.

4.2 Ghost reduction in the $B \to hh$ channel

In 2012 LHCb data, the $B \to hh$ channel from the $B_s^0 \to \mu^+\mu^-$ analysis contains a large background from ghosts. The GP distribution of the μ^- track is shown in Fig. 7b. The B resonance in this channel was used to analyse the background rejection of ghost probability cuts. For this purpose, the data before any ghost-reduction cuts was compared to the events remaining after the cut $GP \leq 0.3$ on both tracks as well as to the events remaining after requiring track fit reduced χ^2 values of less than 2.

The invariant mass distribution of the events in the $B \to hh$ channel before and after these cuts are shown in Fig. 9. Both the GP and the track fit reduced χ^2 cut improve the signal / background ratio dramatically. However, the cut on the ghost probability classifier clearly gives the better purity.



(a) No normalisation

(b) Histograms normalised to same area

Figure 9: Events from 2012 LHCb data in the $B \to hh$ channel. The black line shows the distribution before any ghost reduction cuts, the coloured lines display the events passing cuts on the ghost probability (red), the old tuning of the ghost proability (blue) and the track fit χ^2.Both cuts improve the signal / background ratio drastically. The cut on the ghost probability classifier leads to the best purity.

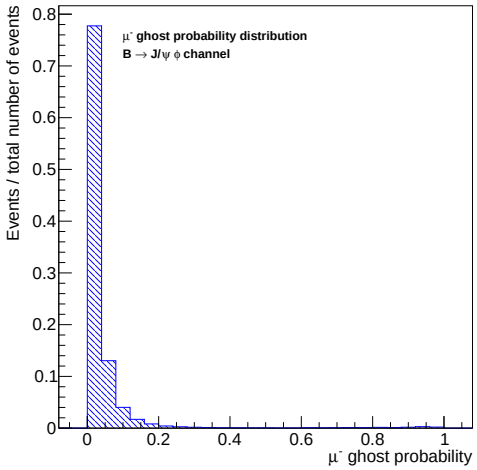
Summarising the studies above, the cut $GP \leq 0.3$ on each track leads to a higher signal efficiency in the $B^+ \to J/\psi K^+$ channel, but also to a better background suppression and therefore a better signal-to-background ratio in the $B \to hh$ channel than the cut on the reduced χ^2 of the track fit. In a similar way, GP cuts were compared to cuts on the old ghost probability variable, and again found to give a better performance.

4.3 Signal efficiencies in the $B_s^0 \to J/\psi\,\phi$ channel

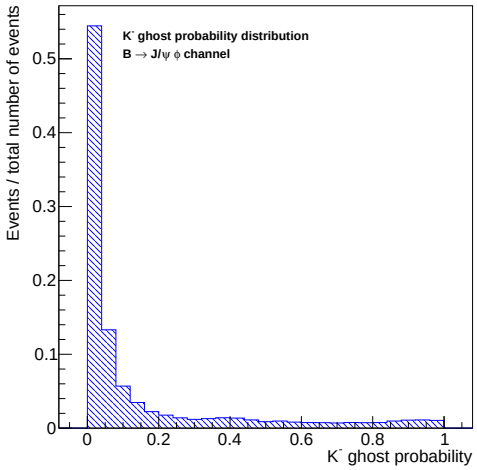
The signal efficiency of GP cuts has also been analysed in the $B_s^0 \to J/\psi\,\phi$ channel, where the J/ψ decays to two muons and the ϕ to a K^+K^- pair. This channel is also part of the $B_s^0 \to \mu^+\mu^-$ analysis. Due to the small difference in mass between two kaons and a ϕ, the opening angle between K^+ and K^- is typically very small, so this decay mode was used to evaluate whether small openining angles can lead to problems with the ghost probability classifier.

The GP distributions of the μ^- tracks and the K^- tracks in 2012 LHCb data are shown in Fig. 10. Cleanup cuts including kaon PID cuts have been applied. While the muon tracks are very clean, a significant number of kaon candidates have a large GP value.

Again, the effect of different GP cuts has been evaluated. The efficiencies were measured



(a) GP distribution of the μ^- track



(b) GP distribution of the K^- track

Figure 10: GP distributions of different tracks in the $B \to J/\psi\,\phi$ channel from 2012 LHCb data.

Cut	Signal efficiency	Background reduction
$GP \leq 0.3$ on the K^+ track	98.96% ± 0.17%	14.3% ± 1.9%
$GP \leq 0.3$ on the K^- track	98.74% ± 0.18%	17.8% ± 2.0%
$GP \leq 0.5$ on the K^+ track	99.53% ± 0.12%	10.2% ± 1.6%
$GP \leq 0.5$ on the K^- track	99.54% ± 0.13%	13.2% ± 1.8%
$GP \leq 0.3$ on the μ^+ track	99.87% ± 0.07%	2.9% ± 1.0%
$GP \leq 0.3$ on the μ^- track	99.81% ± 0.07%	2.7% ± 0.9%
$GP \leq 0.3$ on all four tracks	97.52% ± 0.24%	25.3% ± 2.2%
$GP \leq 0.3$ (0.5) on the μ (K) tracks	98.83% ± 0.19%	20.9% ± 2.1%

Table 2: Signal efficiencies and background reduction rates of ghost probability cuts at the $B \to J/\psi\,\phi$ resonance in 2012 LHCb data. The GP cuts are highly efficient, the small opening angle between the two kaons does not lead to problems. Cuts on the tracks with different charge give the same behaviour.

at the B resonance, where the signal and background contribution were determined by fitting a double Gaussian + exponential function to the data. The fit was simultaneously performed for the data before the ghost-reduction cuts and the events passing each cut. The resulting signal and background efficiencies are given in Table 2.

It was found that all the analysed cuts give high signal efficiencies. As expected from the GP distributions, the cuts on the muon track GP lead to a negligible signal loss of around 0.1% per track, but do not give a large background reduction, either. Cutting on

Cut	Signal efficiency	Background reduction
$GP \leq 0.3$ on the μ^- track	99.9% ± 0.1%	6.8% ± 0.6%
$GP \leq 0.3$ on the μ^+ track	99.8% ± 0.1%	9.7% ± 0.7%
$GP \leq 0.3$ on both tracks	99.7% ± 0.2%	15.3% ± 0.8%

Table 3: Signal efficiencies and background reduction rates of ghost probability cuts at the $J/\psi \to \mu^+\mu^-$ resonance in 2012 LHCb data. The high signal efficiencies agree for the cuts on the μ^- and μ^+ tracks, but the background reduction is different.

the GP of the kaon tracks at 0.3 yields a significant background reduction of around 15% per track, while the signal efficiencies are around 99% per track. Comparing this number to the signal efficiency of the cut on the GP of the K^+ candidate in the $B^+ \to J/\psi$ channel of 99.6% (see Table 1), it seems that the small opening angle does lead to a slight decrease in signal efficiency. However, the performance of the ghost probability classifier is still very good, and the small opening angle of the K tracks does not seem to lead to significant problems. It should also be noted that there is no difference in the signal efficiency of GP cuts on differently charged tracks.

4.4 Signal efficiencies for prompt $J/\psi \to \mu^+\mu^-$

The signal efficiencies of GP cuts were also measured using a prompt $J/\psi \to \mu^+\mu^-$ selection in 2012 LHCb data. The analysed data covers only one magnet polarity (magnet up). The selection criteria are described in [9]. Again, the data before any ghost probability cuts were compared to the remaining events after requiring $GP \leq 0.3$ for the μ^- track, the μ^+ track, or both tracks, respectively. In each case, the number of signal and background events were determined with a fit to a double crystal ball + polynomial function, see [9] for details. The resulting efficiencies and background reduction rates of the different GP cuts are given in Table 3.

Again, the signal losses for the GP cuts are very small. The signal efficiencies for the GP cuts on the μ^- track and on the μ^+ track agree within the uncertainties. However, the background is reduced more strongly by the μ^+ track cut. This might be explained by detector asymmetries and the fact that the analysed data only contains one magnet polarity.

4.5 Asymmetry studies in D^0 decays

In order to check for possible charge asymmetry effects, the effect of ghost probability cuts in the channels $B^+ \to \bar{D}^0\mu^+\nu_\mu$ and $B^- \to D^0\mu^-\bar{\nu}_\mu$ where the D mesons decay to $\pi^+\pi^-$, $\pi^\pm K^\mp$, or $K^\pm K^\mp$ pairs has been analysed. Events from 2012 LHCb data representing an integrated luminosity of approximately 500 pb^{-1} and passing the selection criteria described in [10] were used. The signal efficiency of the cut $GP \leq 0.3$ on all three tracks

Channel	Signal efficiency	Background reduction
$B^+ \to \bar{D}^0(\pi^+\pi^-)\mu^+\nu_\mu$	99.82% ± 0.17%	11.5% ± 0.2%
$B^- \to D^0(\pi^+\pi^-)\mu^-\bar{\nu}_\mu$	99.56% ± 0.15%	11.7% ± 0.3%
$B^+ \to \bar{D}^0(K^+\pi^-)\mu^+\nu_\mu$	99.54% ± 0.02%	17.6% ± 0.3%
$B^- \to D^0(K^-\pi^+)\mu^-\bar{\nu}_\mu$	99.55% ± 0.02%	18.5% ± 0.3%
$B^+ \to \bar{D}^0(K^+K^-)\mu^+\nu_\mu$	99.46% ± 0.08%	14.8% ± 0.1%
$B^- \to D^0(K^+K^-)\mu^-\bar{\nu}_\mu$	99.42% ± 0.08%	15.1% ± 0.1%

(a) Magnet polarity up

Channel	Signal efficiency	Background reduction
$B^+ \to \bar{D}^0(\pi^+\pi^-)\mu^+\nu_\mu$	99.62% ± 0.18%	11.2% ± 0.2%
$B^- \to D^0(\pi^+\pi^-)\mu^-\bar{\nu}_\mu$	99.89% ± 0.22%	11.7% ± 0.3%
$B^+ \to \bar{D}^0(K^+\pi^-)\mu^+\nu_\mu$	99.58% ± 0.02%	17.3% ± 0.3%
$B^- \to D^0(K^-\pi^+)\mu^-\bar{\nu}_\mu$	99.50% ± 0.02%	17.5% ± 0.3%
$B^+ \to \bar{D}^0(K^+K^-)\mu^+\nu_\mu$	99.38% ± 0.09%	15.6% ± 0.2%
$B^- \to D^0(K^+K^-)\mu^-\bar{\nu}_\mu$	99.41% ± 0.09%	15.3% ± 0.2%

(b) Magnet polarity down

Channel	Signal efficiency	Background reduction
$B^+ \to \bar{D}^0(\pi^+\pi^-)\mu^+\nu_\mu$	99.73% ± 0.12%	11.4% ± 0.1%
$B^- \to D^0(\pi^+\pi^-)\mu^-\bar{\nu}_\mu$	99.71% ± 0.12%	11.7% ± 0.3%
$B^+ \to \bar{D}^0(K^+\pi^-)\mu^+\nu_\mu$	99.56% ± 0.01%	17.4% ± 0.2%
$B^- \to D^0(K^-\pi^+)\mu^-\bar{\nu}_\mu$	99.53% ± 0.01%	18.1% ± 0.2%
$B^+ \to \bar{D}^0(K^+K^-)\mu^+\nu_\mu$	99.42% ± 0.03%	15.5% ± 0.1%
$B^- \to D^0(K^+K^-)\mu^-\bar{\nu}_\mu$	99.41% ± 0.04%	15.5% ± 0.3%

(c) Combined magnet polarities

Table 4: Signal efficiencies and background reduction rates of the cut $GP \leq 0.3$ on all tracks on different $B \to D^0(hh')\mu\nu_\mu$ channels in 2012 LHCb data for different magnet polarities. The results for the D^0 and $\bar{D}^0$ channels and for the different magnet polarities agree well, no charge asymmetry effects have been found.

Quantity	No. variables	Input to classifier for				
		L	V	U	D	T
Total track fit χ^2, dof	2	✓	✓	✓	✓	✓
Track fit χ^2, dof contributions from VELO	2	✓	✓	✓		
Track fit χ^2, dof contributions from T stations	2	✓			✓	✓
Track fit χ^2 contribution from match	1	✓				
Observed hits assigned to track in VELO	1	✓	✓	✓		
Observed hits assigned to track in TT	1	✓		✓	✓	
Observed hits assigned to track in IT, OT	2	✓			✓	✓
Excess over expected hits in VELO	1	✓	✓	✓		
Excess over expected hits in TT	1	✓	✓	✓	✓	
Excess over expected hits in IT, OT	2	✓	✓	✓	✓	✓
OT outliers + bad drift times	1	✓			✓	✓
Total hit multiplicities in VELO, TT, IT, OT	4	✓	✓	✓	✓	✓
Pseudorapidity	1	✓	✓	✓	✓	✓
Transverse momentum	1	✓	✓	✓	✓	✓

Table 5: Input variables used for the ghost probability classifiers for different track types (L: long tracks, V: VELO tracks, U: upstream tracks, D: downstream tracks, T: Ttracks).

has been evaluated in each of these channels by simultaneously fitting a double Gaussian + exponential function to the events before and after the ghost probability cut. The results are shown in Table 4.

Again the ghost probability cuts yield high signal efficiencies of more than 99%. No significant difference between the $B^+ \to \bar{D}^0\mu^+\nu_\mu$ and the $B^- \to D^0\mu^-\bar{\nu}_\mu$ channels was found. Thus, there is no evidence for any charge asymmetry introduced by cuts on the ghost probability classifier.

5 Extension to other track types

In addition to the GP variable for long tracks presented above, ghost probability classifiers for VELO, upstream, downstream tracks and Ttracks have been constructed in the same way. The neural networks use the same input variables as the classifier for long tracks as long as they are available for a given track type. The full list of input variables for all classifiers is given in Table 5. The classifiers for other track types are also based on MLP neural networks implemented in TMVA and have been trained on the same inclusive J/ψ sample of good tracks and ghosts tracks. The numbers of good tracks and ghosts and the ghost rates for the different track types in this sample are given in Table 6.

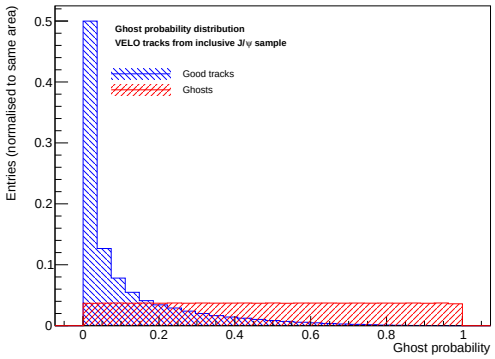
The distribution of ghost probability output values for different track types is shown in Fig. 11. Again, the classifier output is constructed in such a way that the GP distribution of ghost tracks of the training sample is flat in the range [0,1]. The performance of GP

Track type	Good tracks	Ghost tracks	Ghost rate
Long	1 055 379	386 938	26.8%
VELO	599 875	132 270	18.1%
Upstream	75 686	16 874	18.2%
Downstream	140 555	175 163	55.5%
Ttrack	1 013 729	138 497	12.0%

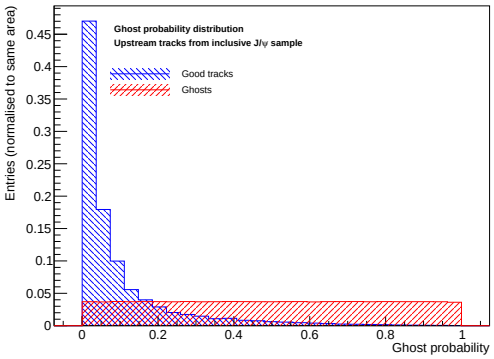
Table 6: Number of good tracks and ghosts in the training samples used for the ghost probability classifiers for the different track types.

cuts on different track types is given by the ROC curve in Fig. 12. On downstream tracks from the inclusive J/ψ sample, the cut $GP \leq 0.3$ removes 70% of the ghost tracks and 14.5% of the good tracks. The cut $GP \leq 0.5$ gives a ghost rejection rate of 50% and a signal loss of 1.7% per track. This performance of the ghost probability is significantly better than that of the other classifiers analysed.

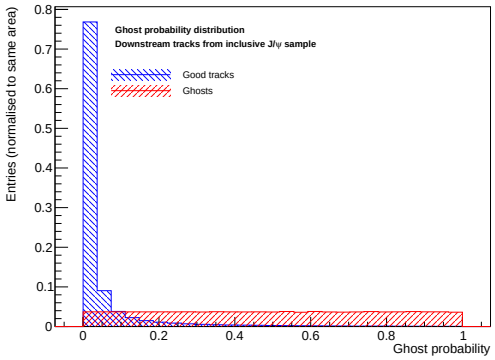
In the current reconstruction version `Reco14`, the ghost probability classifier has only been implemented for long tracks. The classifiers for other track types will be added in the next version.



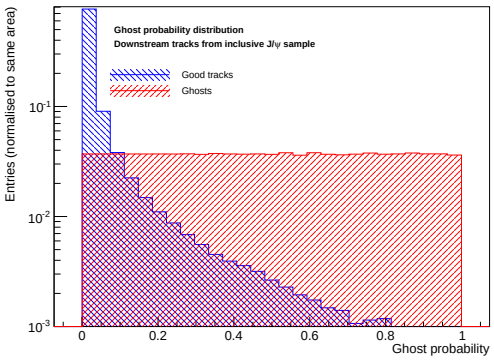
(a) VELO tracks



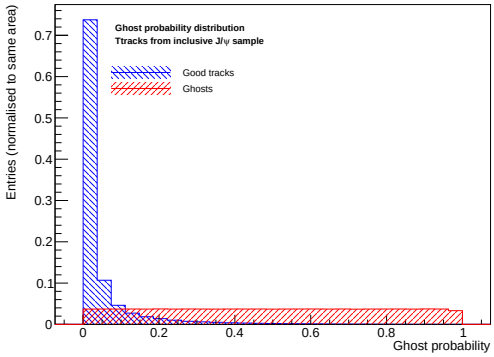
(b) Upstream tracks



(c) Downstream tracks

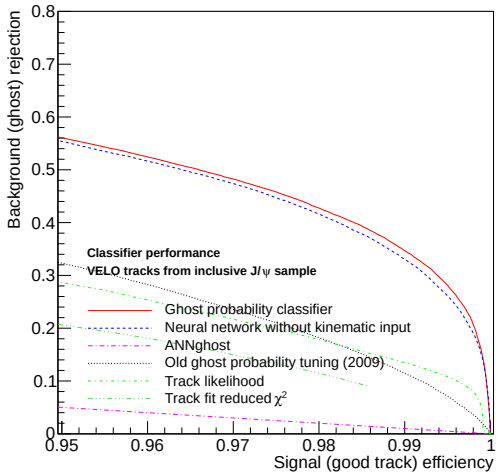


(d) Downstream tracks (logarithmic scale)

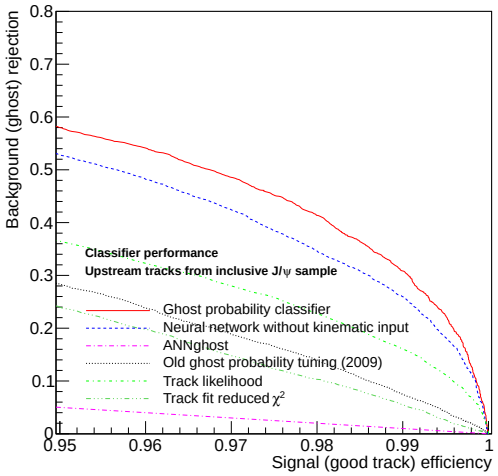


(e) Ttracks

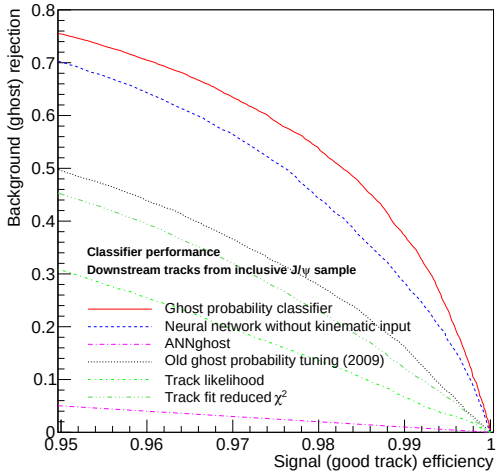
Figure 11: *GP* distributions for different track types on the training sample.



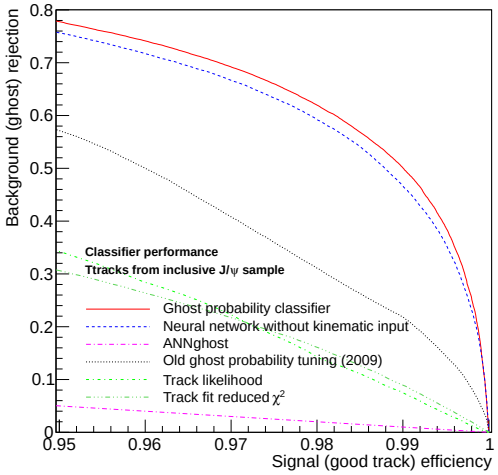
(a) VELO tracks



(b) Upstream tracks



(c) Downstream tracks



(d) Ttracks

Figure 12: ROC curves for ghost probability cuts on different track types.

6 Conclusions

In this note we presented a new multivariate classifier, the ghost probability (GP). It is based on a neural network with an input of 22 tracking variables and quantifies the likelihood of a track being misreconstructed ("ghost"). It was trained on an inclusive J/ψ Monte-Carlo sample based on 2012 reconstruction conditions.

The signal efficiencies and ghost reduction rates of different ghost probability cuts have been evaluated on Monte-Carlo samples. Cutting on this classifier as compared to other quantities gives a higher ghost reduction for the same signal efficiency. The cut $GP \leq 0.3$ removes 70% of the ghost tracks on the inclusive J/ψ sample, while losing only 0.3% of the good tracks. It does not seem to introduce significant bias or asymmetry effects. Various checks on 2012 LHCb data also showed very high signal efficiencies and sizable background reduction rates. Most importantly, no charge asymmetry effects were found.

The superb performance of this classifier together with the almost negligible loss in signal efficiency were compelling arguments for many analyses (e. g. $B_s^0 \to \mu^+\mu^-$, $\tau \to 3\mu$, $B^0 \to K^*\mu^+\mu^-$, $B \to DX$ and many more) to immediately implement this classifier in their stripping selections. This quick implementation allowed to achieve the computing requirements of the `Stripping20` campaign.

Acknowledgements

We would like to thank Jeroen Van Tilburg and Chris Jones for their help and useful discussions.

References

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

[2] J. Albrecht, *Neural Net Based Track Quality*, presentation at the tracking workshop, 28.January 2010.

[3] C. Jones, *http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/davinci /releases/v32r2p1/doxygen/d6/dfc/class_rich_1_1_tabulated_function1_d.html*, .

[4] M. Needham, *Identification of Ghost Tracks using a Likelihood Method*, Tech. Rep. LHCb-2008-026. CERN-LHCb-2008-026. LPHE-2008-004, CERN, Geneva, May, 2008.

[5] C. Jones, *Charged PID*, presentation at the software week, 24. April 2012.

[6] M. Needham, *Classification of Ghost Tracks*, Tech. Rep. LHCb-2007-128. CERN-LHCb-2007-128, CERN, Geneva, Oct, 2007.

[7] C. Adrover *et al.*, *Search for the rare decays $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ with 1.02 fb^{-1}*, CERN-LHCb-ANA-2011-102 (2012).

[8] T. Ullrich and Z. Xu, *Treatment of Errors in Efficiency Calculations*, `arXiv:physics/0701199`.

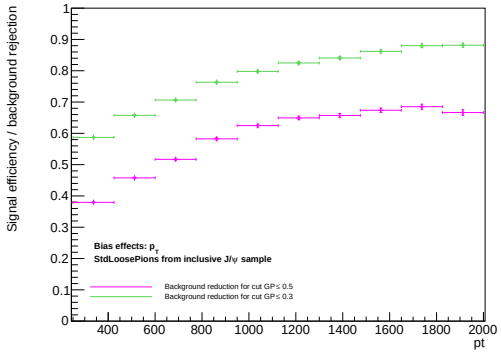
[9] J. Albrecht *et al.*, *Updated measurement of the J/ψ production cross-section in LHCb*, CERN-LHCb-ANA-2011-102 (2010).

[10] L. Darmé *et al.*, *Search for direkt CP violation in $D^0 \to K^-K^+, \pi^-\pi^+$ using semileptonic B decays*, CERN-LHCb-ANA-2012-012 (2012).

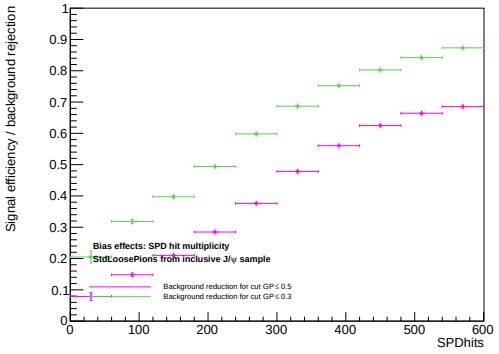
Appendix

A Additional plots

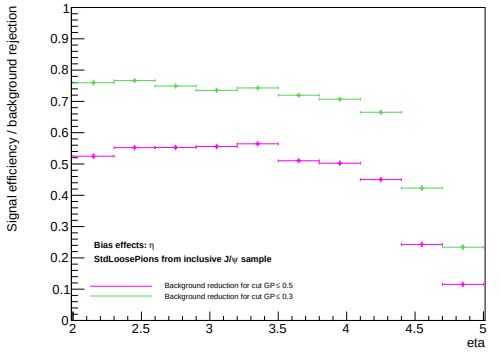
In Fig. 13, signal efficiencies and background rejection rates of GP cuts are shown as functions of different variables. Again, StdLoosePions from an inclusive J/ψ MC sample under 2012 conditions have been used.



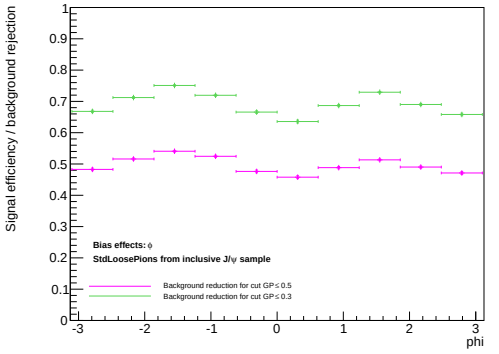
(a) As a function of p_T



(b) As a function of the total hit multiplicity in the SPD



(c) As a function of η



(d) As a function of φ

Figure 13: Signal efficiency and background reduction as a function of different variables for the cuts $GP \leq 0.3$ and $GP \leq 0.5$.