



CERN Summer Student Report 2017

CP Violation for decay $D^0 \rightarrow \phi \gamma$

Athina Papachristodoulou

Supervised by: Chris Burr

Abstract

In order to make a CP violation measurement re-stripping of the 2012 data set is required due to lack of a reference mode. This study modeled the mass distribution of the D^0 meson when decaying via $D^0 \rightarrow \phi \gamma$. From the analysis, 1040 ± 46 signal events were measured from the 2012 LHCb Run I data set using an integrated luminosity of 2 fb^{-1} and recorded at a centre-of-mass energy of 8 TeV.

Contents

1	Introduction	1
2	LHCb Detector	2
2.1	Particle Identification	2
2.2	Calorimeter	3
3	Experimental method	3
3.1	Candidate Selection	3
3.2	Linear Selection	3
3.3	Helicity Angle Calculation	5
4	Multivariate selection	5
5	Results	10
5.1	2D FIT	10
5.1.1	D^0 invariant Mass distribution	10
5.1.2	Helicity Angle	12
5.2	2012 Data Fitting	13
5.3	Helicity Plots without fitting	16
6	Conclusions and future plans	18
	References	19

1 Introduction

CP symmetry refers to a combination of charge conjugation, where a particle replaced by it's antiparticle and parity transformations, which reverses the spatial coordinates of a system. In case that CP symmetry exist, the process of a particle will have the same interaction rate as its antiparticle equivalent. In the other case, the CP symmetry is violated and then we have the CP Violation, which we will discuss in this project.

Nowadays, flavor-changing radiative decays of the charmed meson system $D \rightarrow V\gamma$, where V is a vector meson is of particular interest. According to the Standard Model, these decays have a short-distance contribution ,which is negligible with branching fraction less than 10^{-8} and the long-distance contribution which has a branching fraction in the range 10^{-5} [1] .Short-range contributions transpire via a rare radiative loop-level transition which is expected to violate CP symmetry and a dominate non-CP violating transition. In our case, we study the decay of D^0 meson ,which is disintegrating to ϕ meson and a photon according to the following decay:

$$D^0 \rightarrow \phi\gamma$$

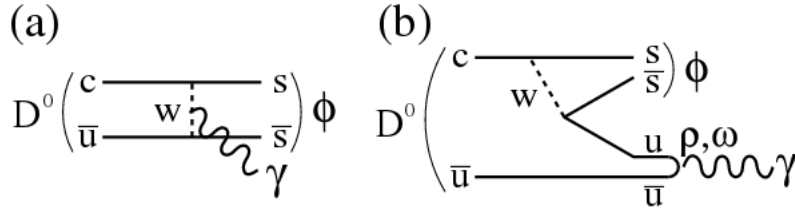


Figure 1: (a) Short distance process diagram, (b) long distance process diagram for the decay $D^0 \rightarrow \phi\gamma$

In this decay, the CP violation occurs directly and we can calculate directly the asymmetry according to the equation [2]:

$$A_{raw} = \frac{N(D^0 \rightarrow \phi\gamma) - N(\bar{D}^0 \rightarrow \phi\gamma)}{N(D^0 \rightarrow \phi\gamma) + N(\bar{D}^0 \rightarrow \phi\gamma)} \quad (1)$$

where N is the number of events identified from each decay. In order to identify if the meson was D^0 or $\bar{D}^0$, it needs to have originated from a D^{*+} meson, which decays via $D^{*+} \rightarrow D^0\pi^+$. The charge of the pion in this decay, can then be used to identify the nature of the D^0 meson.

2 LHCb Detector

The LHCb Detector is a single-arm forward spectrometer and one of the experiments at the Large Hadron Collider (LHC), which is primarily used for precision measurements of CP violation and b or c decays. LHCb consists of Vertex Locator(VELO) which is at the point closest to where the beams collide and particles containing b and anti-b quarks are produced, two Ring Imaging Cherenkov (RICH) detectors which are for particle identifications, working by measuring the emission of Cherenkov radiation. In addition, the detector contains a Trigger Tracker (TT) prior to the dipole magnet and three others Tracker stations (T1,T2,T3) and a calorimeter system. [3]

2.1 Particle Identification

Particle identification (PID) is very useful for identify the types of particles which are detected on LHCb. PID uses a combination of information given by the trackers (TT, Inner and Outer) and by the PID system for differentiating particles. It consists of two RICH detectors, a calorimeter system and a muon system which are described above. Using the velocity obtained from the RICH detectors and the momentum which is can be determined from the tracking system, we can determine the mass of a charged particle.

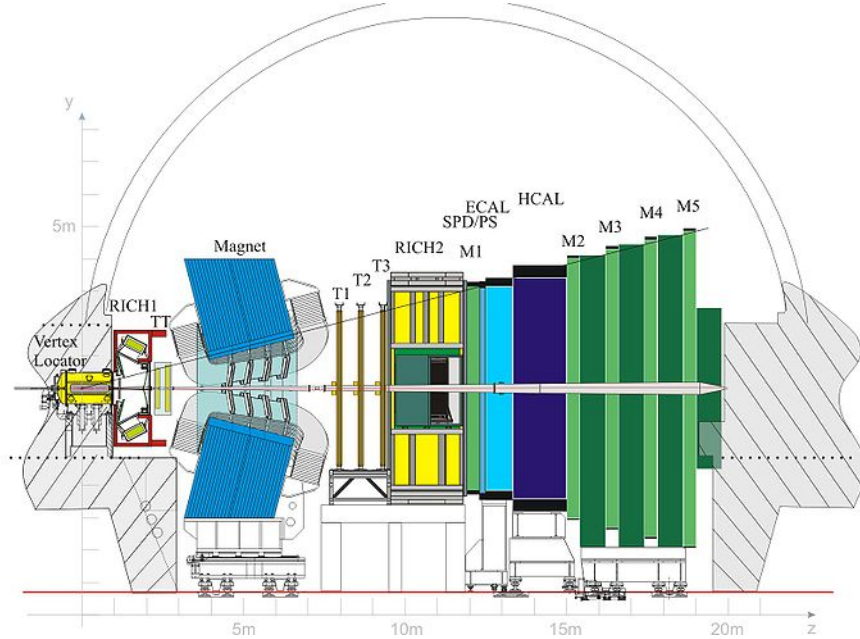


Figure 2: Layout of LHCb Detector

2.2 Calorimeter

The calorimeter system of LHCb is comprised of four components: a scintillator pad detector (SPD), a Pre-Shower Detector (PS), an electromagnetic calorimeter (ECAL), which measures the energy of lighter particles (like electrons and photons) and a Hadronic calorimeter (HCAL) which measures the energy of protons, neutrons and other particles with quarks. The main purpose of the calorimeter system is to differentiate particles by stopping them as are through the detector, measuring the amount of energy lost and determine the transverse energy used later for high level triggers.

3 Experimental method

3.1 Candidate Selection

The data samples used in this data analysis are recorded by the LHCb experiment during LHC Run I in 2012. We used Monte Carlo simulations corresponding to the 2012 LHCb running conditions, with a centre-of-mass energy of 8TeV.

3.2 Linear Selection

In this part, we have applied more strictly cuts listed in the Table 1 to stripped MC samples of the signal and the merged pion background corresponding to the luminosity and centre-of-mass energy at the LHCb 2012 data set.

Delta Mass Cut

The delta mass Δ is defined as the difference between the invariant mass of the D^{*+} meson and the D^0 meson. The delta mass cut is useful to reduce the combinatorial background because by this cut $|\Delta - 145.4 \text{ MeV}| < 2 \text{ MeV}$ where 145.4 is the mass of the pion it ensures that the pion is of low momentum as expected.

Variable	Condition
$K^\pm$ PIDK	> 10
D^0 direction angle	< 0.02 rad
K^+ and π_s ghost probability	< 0.3
γ CL	> 0.25
$\pi^0 \rightarrow \gamma\gamma$	$= 0$
$ M_\phi - 1020 \text{ MeV} $	$< 10 \text{ MeV}$
$ \Delta - 145.4 \text{ MeV} $	$< 2 \text{ MeV}$

Table 1: A table of the acceptance selections that we have applied to MC samples

Triggers
L0Global_TOS
L0HadronDecision_TOS
L0PhotonDecision_TOS
Hlt1TrackPhotonDecision_TOS
Hlt1TrackAllL0Decision_TOS
Hlt2CharmHadD02HHXDst_hhXDecision_TOS

Table 2: A table of the triggers that we used in this analysis

According to the Table 2 above, we applied triggers using the LHCb trigger system which is consist of two levels: level zero (L0) and the high level triggers (HLT). The level zero(L0) is a trigger on events with large transverse energy using the calorimeter. The high level triggers that we applied, are divided into two levels, the HLT1 which performs a partial event reconstruction using track segments from the VELO and the HLT2 where are fully reconstructed all tracks with a large transverse momentum. Triggers also categorise each event selected into TOS (Trigger On Signal) and TIS (Trigger Independent of Signal). If the objects are associated with the signal, then the specific event is classified as TOS and if is independent of the signal decay, is classified as TIS. By applying these triggers, helped us to have a more strictly cut particularly for the background events.

3.3 Helicity Angle Calculation

One of the most important calculations which help us to differentiate the signal ($D^0 \rightarrow \phi\gamma$) and the merged pion background ($D^0 \rightarrow \phi\pi^0$) is the difference in helicity angle, defined as the cosine of θ , the angle between the D^0 meson and one of the two kaons (K^+ or K^-) as illustrated in Figure 3. To calculate the helicity angle, we defined a helicity calculation

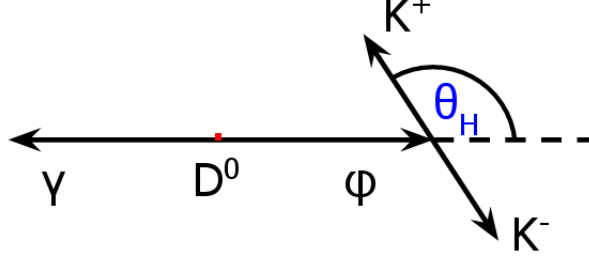


Figure 3: Helicity Angle θ between D^0 and one of the two kaons

function, which uses only the mass and momentum components of D^0 , K^+ and K^- to calculate the momentum components and the energy of ϕ , π^0 and γ according to the following decays:

$$\phi \rightarrow K^+ + K^- \quad (2)$$

$$\text{for the signal } D^0 \rightarrow \phi + \gamma \quad (3)$$

$$\text{for the merged pion background } D^0 \rightarrow \phi + \pi^0 \quad (4)$$

because of the conservation of energy and angular momentum. We calculated the cosine of helicity angle θ in the ϕ rest frame.

4 Multivariate selection

In this section, we will describe the multivariate selection (particularly the BDT method) which is used to discriminate between the signal and background events. As the behaviour of Monte Carlo data is already known, we used Boosting Decision Trees on real-data by training our Monte Carlo data in order to confirm that we have the right results. Firstly, we stripped MC by applying the selections as described in the Table 1 and then we applied

Variable	Condition
γ CL	0.25
$\pi^0 \rightarrow \gamma\gamma$ veto	= 0

Table 3: A table of pre-selection cuts used on $D^0 \rightarrow \phi\gamma$ and $D^0 \rightarrow \phi\pi^0$ MC samples

a pre-selection as described in Table 3. After the pre-selection cuts, the remaining samples of signal events and merged pion events are divided by two sub-samples, the training sample and the testing sample (both of them for the MVA). For the training and testing samples we used 12 input variables out of which 6 are calorimeter variables and 6 Pre-Shower variables. For each variable, we made a distribution for the Signal and Background MC samples as we can see in Figures 4, 5.

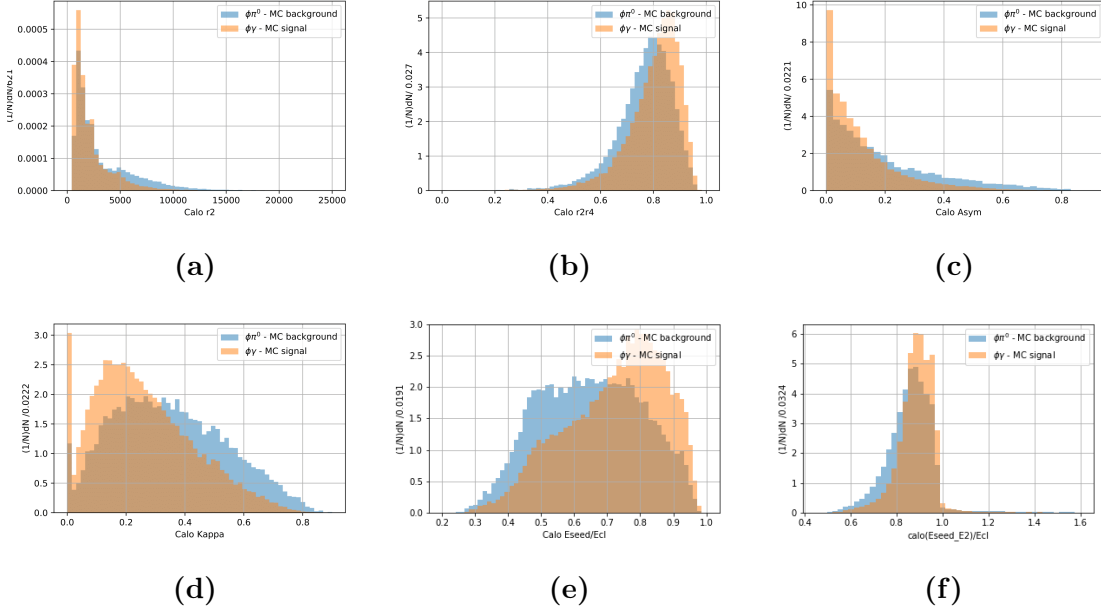


Figure 4: Distributions of the Electromagnetic Calorimeter (ECAL) cluster hits for r2, r2r4, asym, κ , E_{seed}/E_{cl} and $(E_{seed} + E2)/E_{cl}$ for Signal MC samples (in blue) and merged pion background MC (in orange) samples.

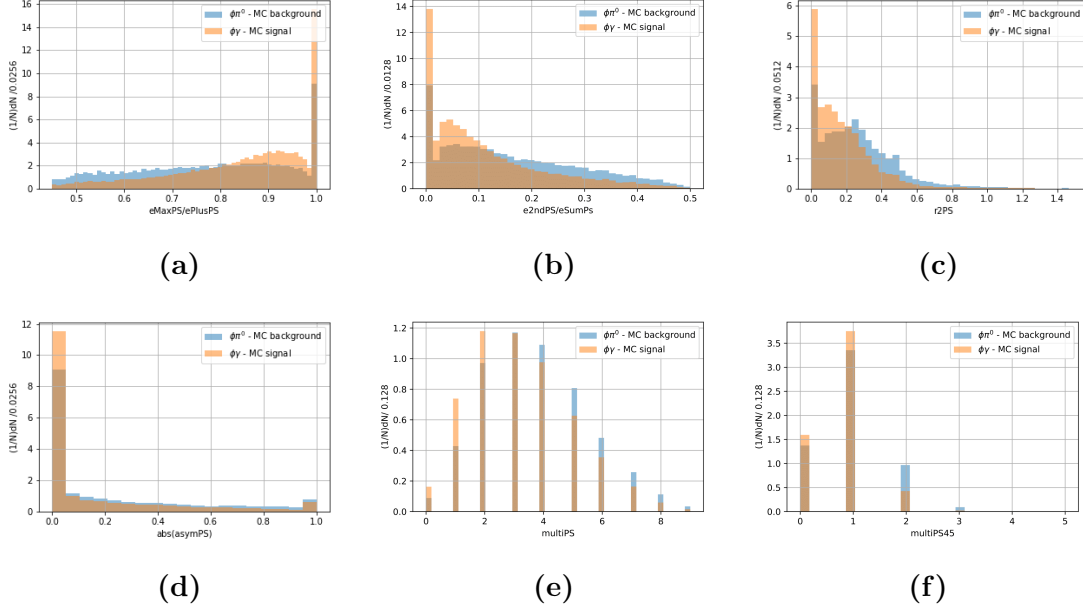


Figure 5: Distributions of the following Pre-Shower (PS) variables : E_{max}/E_{sum} , $E2/E_{sum}$, $r2$ PS, absolute asym, multi and multi45 PS for signal MC (in blue) and merged pion background MC samples (in orange).

Our next step is to find the best set of parameters for the BDT technique, which are not learnt from the training data. In order to find the best set of hyper-parameters for which the area under the ROC curve is maximized, we used Cross Validation splitting our development in training and testing sub-samples, performing a grid search over the parameter space. Specifically, in our study the hyper-parameters are `n_estimators`, `learning rate` and `max_depth` and the best set of them was shown in Table 4 .

Hyper Parameters	
<code>n_estimators</code>	250
<code>learning rate</code>	0.085
<code>max_depth</code>	5

Table 4: Best set of Hyper Parameters for the BDT method

The distribution for the output of BDT (Boosting Decision Trees) method can be found in Figure 6. The best set of hyper-parameters are shown in Table 4, we had the maximum area ($= 0.74$) in ROC plot according to the Figure 7 and 8.

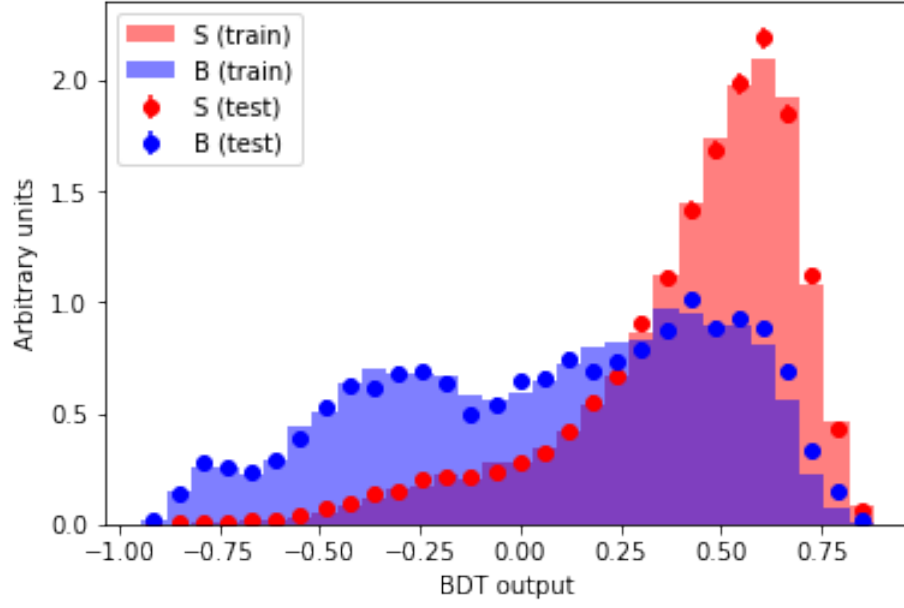


Figure 6: Output of the BDT for Signal(in red) and Background (in blue) MC training and testing samples

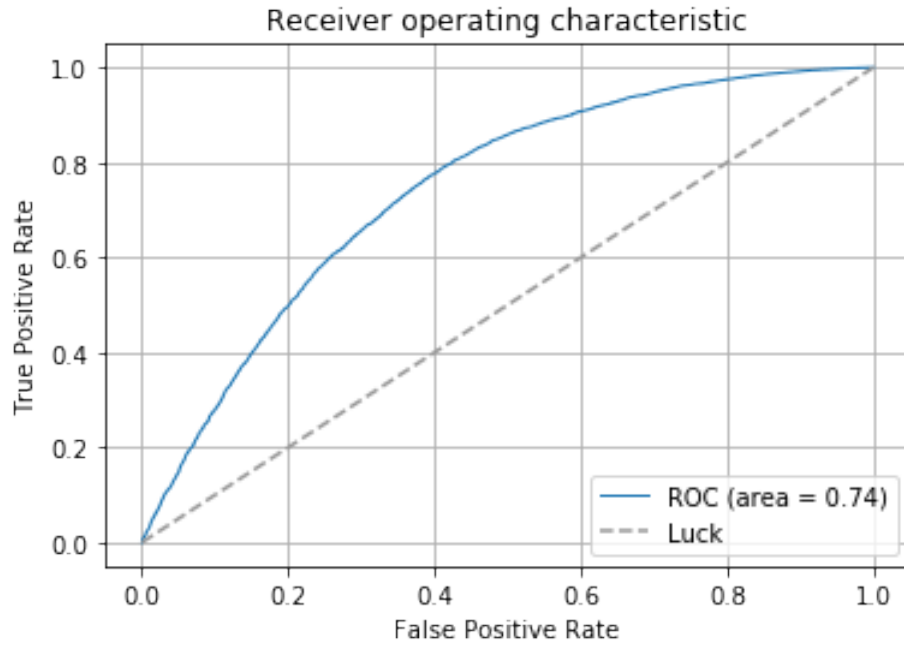


Figure 7: ROC (Receiver operating characteristic) plot with maximum area =0.74 by using the best set of Hyper-Parameters.

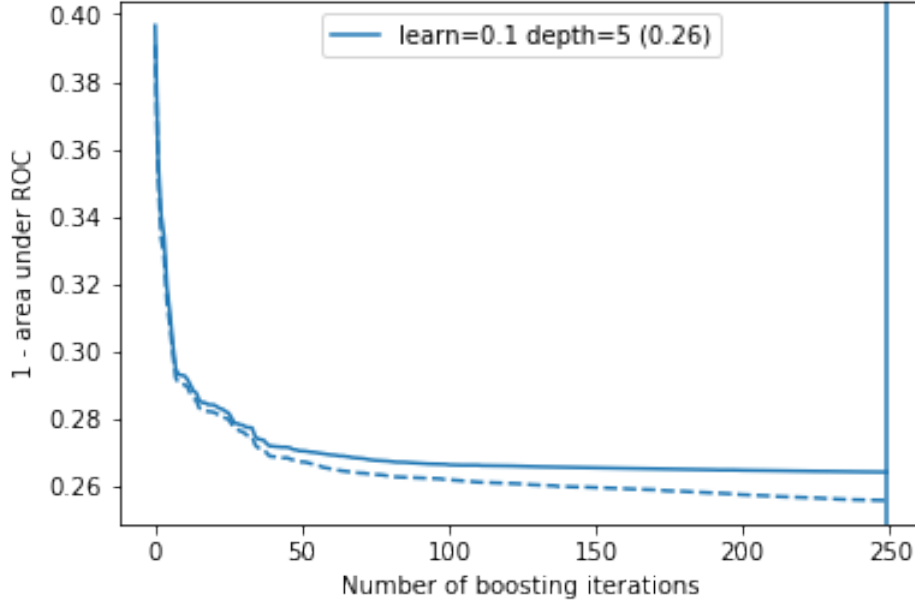


Figure 8: (a) Plot of the area under the ROC after each boosting iteration for each of the classifiers. The solid line shows the performance on the testing sample and the dashed line shows the performance on training sample. The vertical line mark the boosting iteration (= 250) at which our classifier achieves its best performance ($n_estimators=250$) as measured on the testing set.

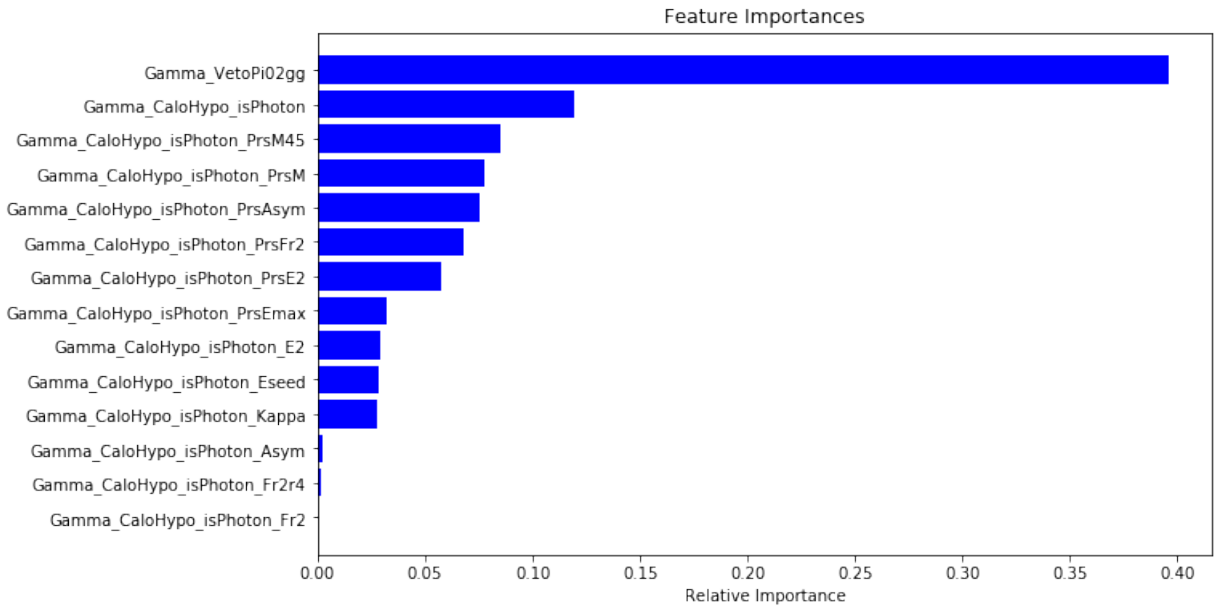


Figure 9: Plot of feature importances, in which the most important feature is `Gamma_VetoPi02gg`.

5 Results

5.1 2D FIT

Raw signal yield estimation is extracted from a 2-dimensional fit to the D^0 invariant mass and helicity angle calculations. The distribution for the invariant mass of D^0 meson and the helicity angle is separated into 3 components; the signal mode, the merged pion background mode and the combinatorial background mode. The combinatorial background fit was extracted from the real data in contrast with the invariant mass distributions of the signal and background mode, which were predicted using MC simulations of $D^0 \rightarrow \phi \gamma$ and $D^0 \rightarrow \phi \pi^0$ events respectively.

5.1.1 D^0 invariant Mass distribution

Due to energy losses from interactions within the detector medium a Crystal Ball [4] function, given by equation:

$$f(x; \alpha, n, \bar{x}, \sigma) = \begin{cases} \exp(-\frac{(x-\bar{x})^2}{2\sigma^2}), & \text{for } \frac{x-\bar{x}}{\sigma} > -\alpha \\ A \cdot (B - \frac{(x-\bar{x})}{\sigma})^{-n}, & \text{for } \frac{x-\bar{x}}{\sigma} \leq -\alpha \end{cases} \quad (5)$$

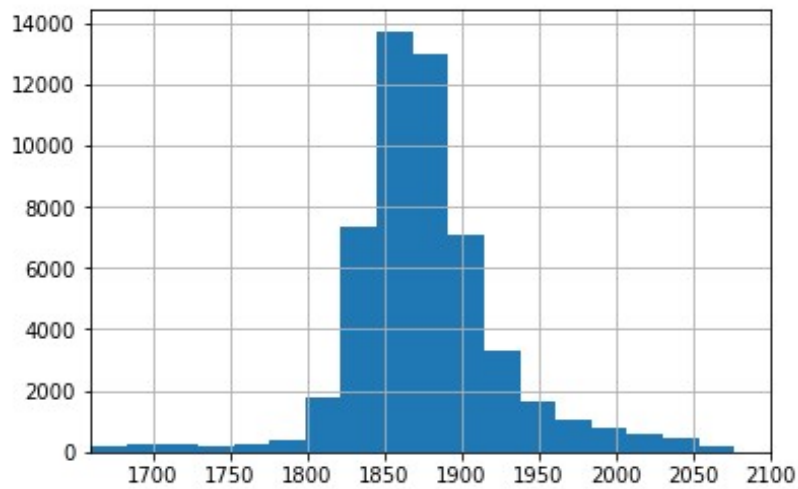
$$\text{where: } A = (\frac{n}{|\alpha|})^2 \cdot \exp(-\frac{|\alpha|^2}{2}) \quad (6)$$

$$B = \frac{n}{|\alpha|} - |\alpha| \quad (7)$$

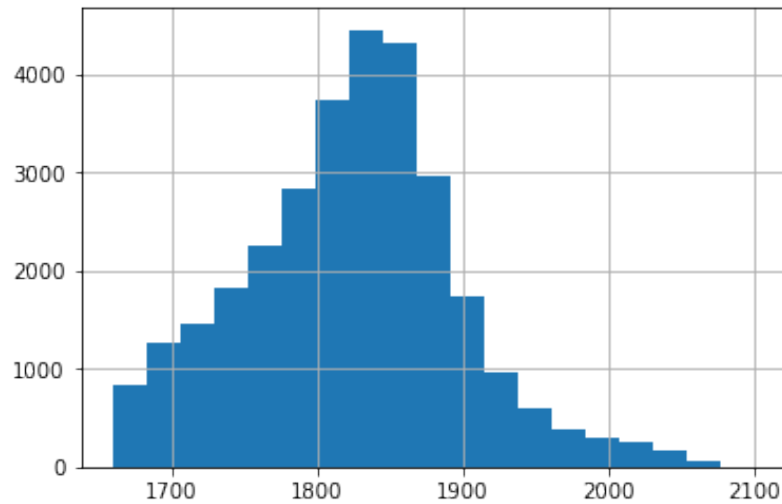
N is a normalization factor and α , n, x and σ are parameters which are fitted with data, is expected to provide us an accurate description. Furthermore, we applied 3 more cuts according to the Table 5 in order to obtain the invariant mass distribution of D^0 meson for the signal mode and merged pion background mode. We have tried 3 different options of BDT cut in order to check the behaviour of training list in 3 different selections of BDT value and we observed that the shape of the invariant mass distributions was the same for the 3 of them.

Variable	Condition
$Dstr_BKGCAT$	<11
$ delta_mass - 145.2 \text{ MeV} $	$<2 \text{ MeV}$
bdt_value	$< -0.25 \text{ or } 0 \text{ or } 0.25$

Table 5: Applying extra cuts in order to have a better results of training list and checked the differences in invariant mass distributions of signal mode and merged pion background mode



(a) A histogram to show the Invariant mass distribution for the Signal mode where x-axis is the Invariant Mass of D^0 (MeV/c^2) and y-axis the number of counts



(b) A histogram to show the Invariant mass distribution for the merged pion background mode where x-axis is the Invariant Mass of D^0 (MeV/c^2) and y-axis the number of counts

According to the plots shown in Figure 10, we obtained more counts for background mode instead of signal mode. In Figure 10b a tail was observed on the right hand side of mass peak between 1940-2060 MeV/ c^2 which is in line with the signal mode and well-described by Crystal Ball function.

5.1.2 Helicity Angle

The distributions of helicity angle in the signal and merged pion background are very different because of the intrinsic parities [5] of the associate particles. According to the Figure 11, the helicity angle for the signal and the merged pion background are follow $\sin^2\theta$ and $\cos^2\theta$ distribution respectively. The theoretical distributions are in line with those which we expected from the Monte Carlo simulation.

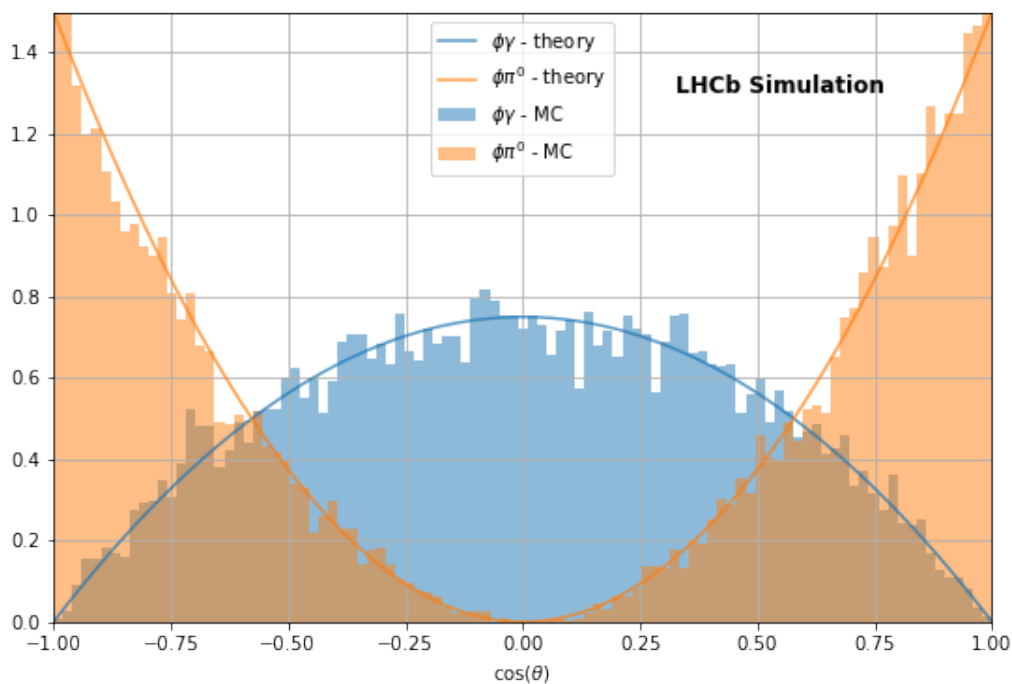


Figure 11: Plot of the Helicity Angle θ for the signal (in blue) and the merged pion background (in orange) for theoretical distribution and MC simulation (without cuts) .

Furthermore, according to the plot there is a deviation between the theory and MC distributions. The deviation is most dramatic for the both signal and background in the edges region (for $\cos(\theta)$ between ± 0.60 and ± 0.85) .

5.2 2012 Data Fitting

We used a fit function to described the components for the 2012 LCHb data set. The number of events associated with each component, was equal in both the D^0 mass distribution and the helicity angle. This analysis found :

$$N(D^0 \rightarrow \phi \gamma) = 1272 \pm 46$$

$$N(D^0 \rightarrow \phi \pi^0) = 4960 \pm 74$$

The model was used to fit a 2-dimensional fit in order to confirm the lack of correlation between the mass of the D^0 meson and the helicity angle .

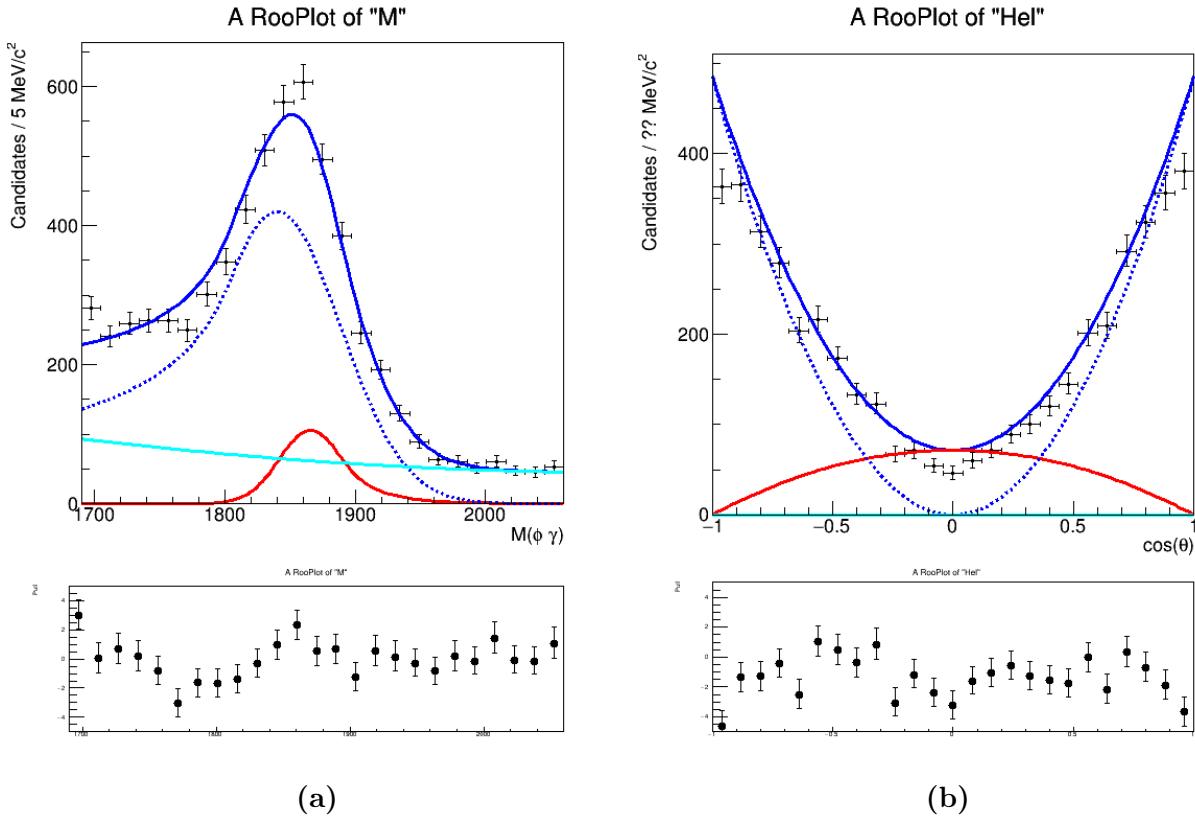


Figure 12: (a)Plot of Invariant mass of D^0 meson from 2012 real data set. We fitted the data using the second order of Chebyshev polynomial. The signal, background and combinatorial background are plotted in red, dashed blue and cyan respectively. (b) Plot of Helicity Angle from 2012 real data set. We fitted the data using the second order of Chebyshev polynomial. The signal, background and combinatorial background are plotted in red, dashed blue and cyan respectively.

The distribution of the helicity angle showed in Figure 12b, is separated into 3 components;

the signal mode (in red), the background mode (in dashed blue) and the combinatorial background (in cyan). According to the Figure 12b, for negative helicity angle the majority of data points lie above the model, while for positive helicity angle the data points lie below the model except of some of them which lie consistently above. The fact that we had some data points above the model and others below, leads us to an asymmetry. This kind of asymmetry may arise from an additional contribution that has not been accounted for. In order to fitted the D^0 meson invariant mass and the helicity angle combinatorial background (2012 real data set) we used a second order Chebyshev polynomial of the first kind given by equation :

$$f(x) = T_3(2x^2-1) + T_2x + T_1$$

where T_1 , T_2 and T_3 are parameters. The fit of combinatorial background in both distributions is shown in Figures 12b, 12a in cyan colour. As depicted in the helicity angle plot (Figure 12b), the combinatorial background seems to be flat, so we had re-plotted it by replacing the parameters T_1 with a unit and T_2 , T_3 of the Chebyshev polynomial with zero. Comparing the helicity angle distributions in Figures 12b and 13b with and without T_1 , T_2 , T_3 parameters respectively, both of them have the same behaviour and the combinatorial background is flat. As the helicity distribution of combinatorial background was flat in both cases, we have as a result that there were not events and we re-plotted the helicity histograms for the 2012 real dataset, without using the fitting function (Section 3.1.1).

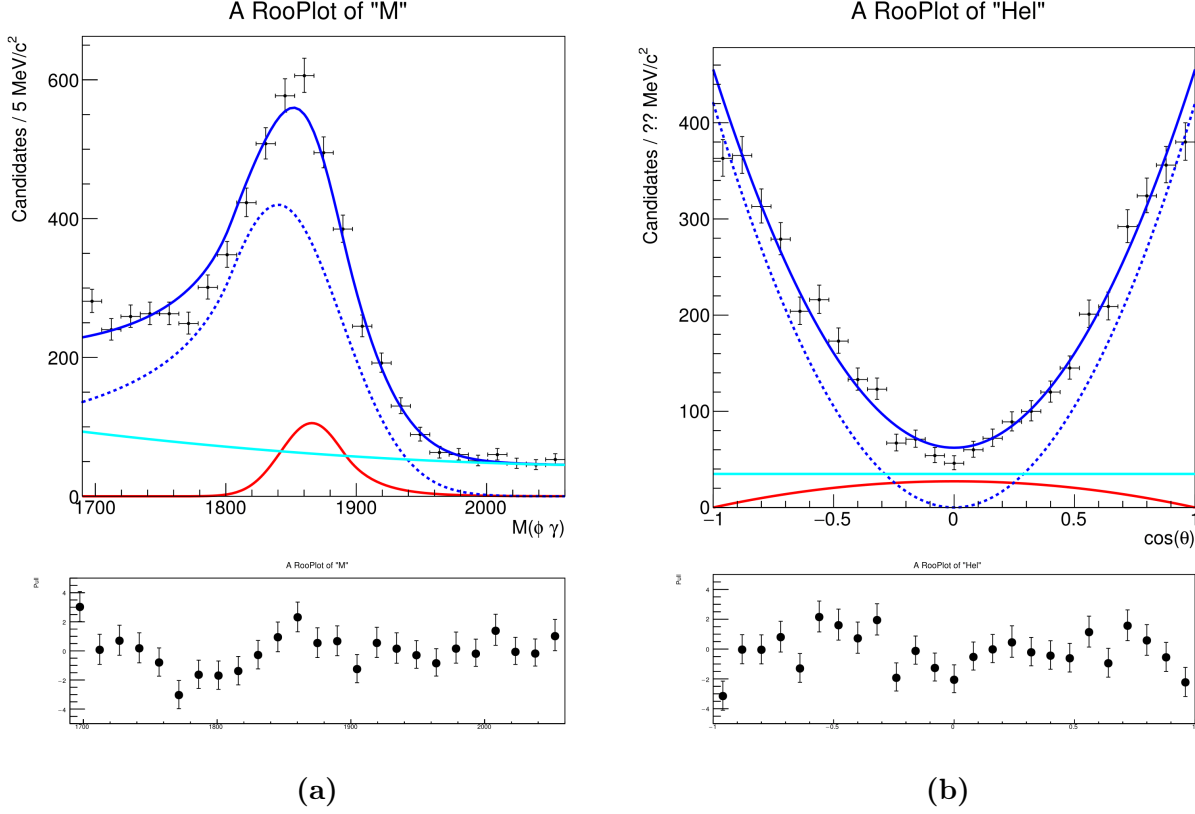


Figure 13: (a)Plot of Invariant mass of D^0 meson from 2012 real data set. We replaced T_1 parameter with unit and T_2 , T_3 parameters from fitting function with zero. The signal, background and combinatorial background are plotted in red, dashed blue and cyan respectively. (b) Plot of Helicity Angle from 2012 real data set. We replaced T_1 parameter with unit and T_2 , T_3 parameters from fitting function with zero. The signal, background and combinatorial background are plotted in red, dashed blue and cyan respectively.

The distribution of the invariant mass of the D^0 meson showed in Figure 12a, is separated into 3 components; the signal mode (in red), the background mode (in dashed blue) and the combinatorial background (in cyan). The distributions of signal mode and the merged pion background mode were predicted using MC simulations of 2012 data and the distribution of combinatorial background was predicted using the real data. According to the Figure 12a we can see a prominent tail on the right hand side of the mass peak between 1960 and 2040 MeV/c^2 . This tail existed because an over correction by the detector, when we taking into account energy losses of the photon [5]. According to the mass fitting plot using the second order of Chebyshev polynomial (with parameters T_1 , T_2 , T_3) in Figure 12a, for the invariant mass of D^0 between 1740 $\sim$ 1840 MeV/c^2 the data points lie below the model while all

others lie above the model. The mass distribution had the same shape in both cases, with and without the parameters of Chebyshev polynomial as we can see in Figures 12a, 13a .

5.3 Helicity Plots without fitting

In this part, we ignored the fitting part and we replaced the condition for the M_ϕ (which is shown in Table 1) with 2 different cuts for the mass of ϕ meson according to the Table 6.

Additional cut	Selections
$M_\phi > 1030$ MeV	full selection
» »	without PIDK cut
» »	without NGhost cut
» »	without PIDK and NGhost cuts
$M_\phi < 1010$ MeV	full selection
» »	without PIDK cut
» »	without NGhost cut
» »	without PIDK and NGhost cuts

Table 6: Extra conditions for the invariant mass of ϕ meson in order to check for any improvements

We made the helicity angle histograms for 2012 real data set applied 4 different selection combinations for each M_ϕ cut, which are shown in Figures 14, 15 for $M_\phi > 1030$ MeV and $M_\phi < 1010$ MeV respectively in order to check the shape of the combinatorial background helicity distribution in each case.

According to the histograms above, we observed that all of them look flat (except of the lower side of the $M_\phi < 1010$ MeV) so we can describe the shape of the combinatorial background by replacing the parameters of the Chebyshev polynomial T2,T3 with zero and T1 with unit. In addition between these 4 selections, PIDK cut is the most useful for removing the background in line with the Figures 14b, 15b.

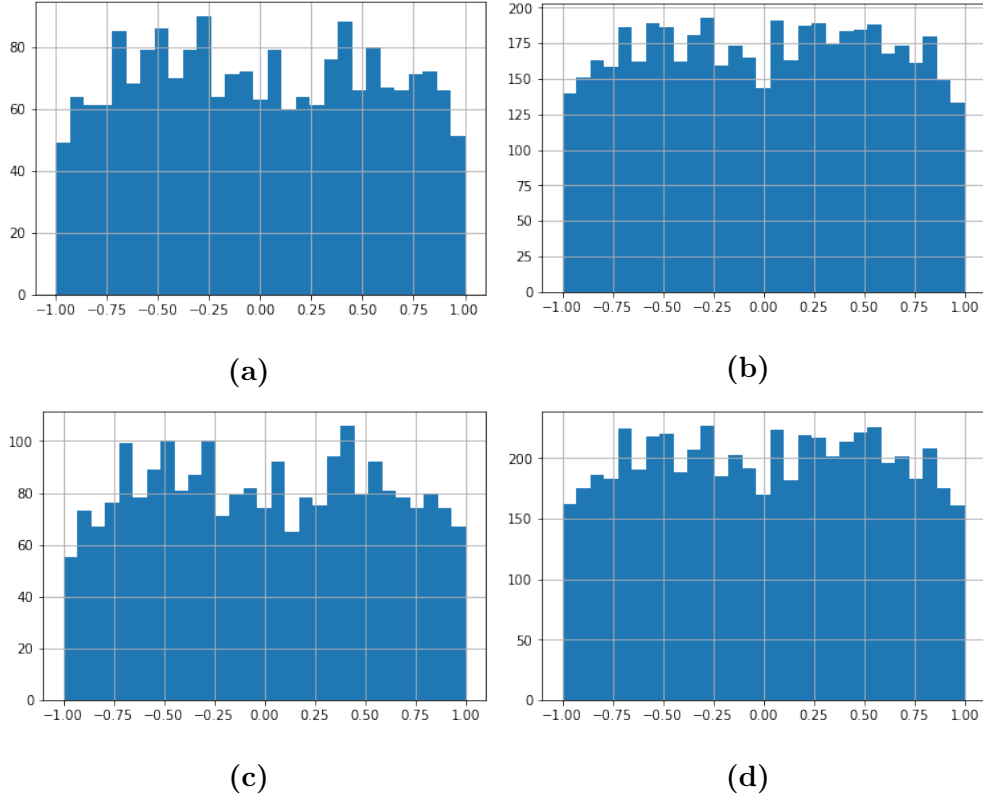


Figure 14: Helicity histograms of real data for $M_\phi < 1030$ MeV where: (a) all selections, (b) without PIDK cut, (c) without NGhost cut and (d) without PIDK and NGhost cuts

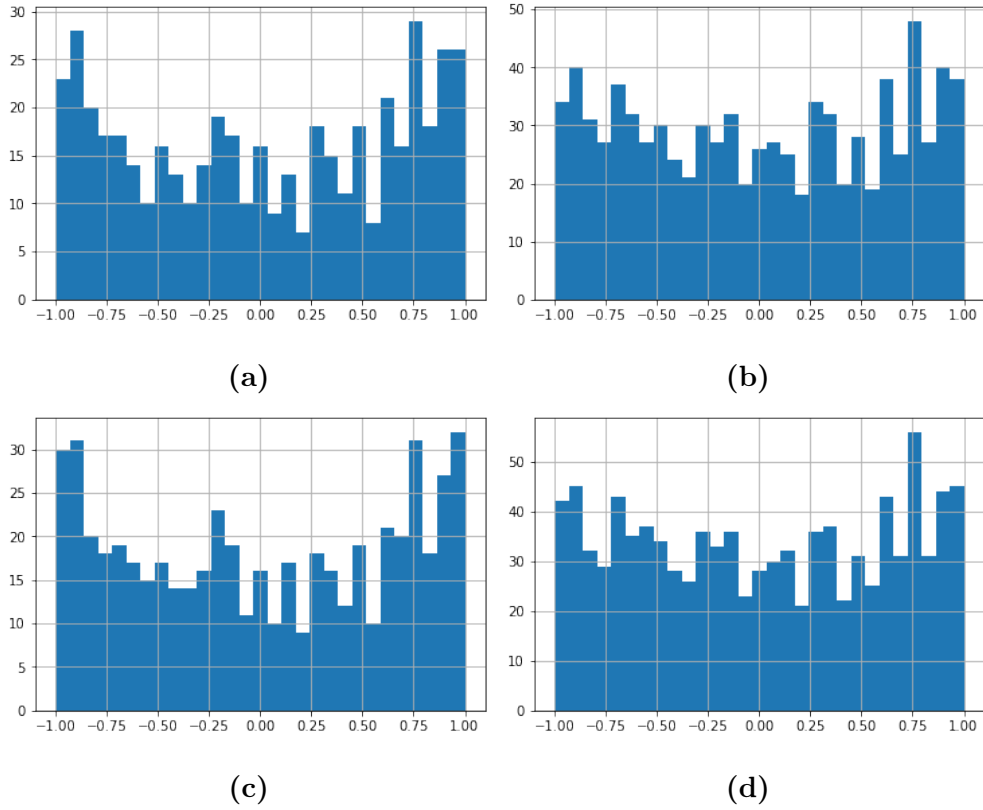


Figure 15: Helicity histograms of real data for $M_\phi > 1010$ MeV where: (a) all selections, (b) without PIDK cut, (c) without NGhost cut and (d) without PIDK and NGhost cuts

6 Conclusions and future plans

The yields of signal, merged pion background and combinatorial background modes are shown in Table 7. According to this table, in this study the background is almost 4 times bigger than the signal. Comparing our results with previous results [5], we observed that in both cases the background is still bigger than the signal. In particular, from the 2012 data set this analysis yielded 1270 ± 46 signal events compared to 847 ± 53 obtained previously. In addition, we concluded that the combinatorial background can be described without the Chebyshev polynomial, as its helicity distributions seems to be flat.

Parameter	Value	Error
$N(D^0 \rightarrow \phi\pi^0)$	4961	± 74
$N(D^0 \rightarrow \phi\gamma)$	1272	± 46
$N(\text{Combinatorial Background})$	392	± 0.50

Table 7: Table of yields for merged pion background, signal and combinatorial background respectively.

Future Plans

We need a further investigation for a function which will describe the shape of the combinatorial background more accurately. Furthermore, as mentioned in Section 5.1.2, the fact that there is a deviation between the theory and MC simulation on the edges of the distribution of the Helicity angle for the signal and the merged pion background, needs a deeper understanding in order to ensure the results of the fit are accurate.

Acknowledgements

I would like to express my thankfulness to my supervisor Chris Burr for his help and his guidelines during the whole project. He was always available for any question and dedicated a lot of time to discuss with me the next steps of the project. In addition, I

would like to thank my second supervisor Dr.Alex Pearce who introduced me to the project and the LHCb software and Professor Chris Parkes for his useful advice.

References

- [1] T. B. Collaboration, “Measurements of the branching fractions of the radiative charm decays $D^0 \rightarrow \bar{K}^{*0} \gamma$ and $D^0 \rightarrow \phi \gamma$,” *Physical Review Letters*, vol. 78, 2008.
- [2] T. B. Collaboration, “Observation of $D^0 \rightarrow \rho^0 \gamma$ and search for cp violation in radiative charm decays,” *Physical Review Letters*, vol. 118, 2017.
- [3] <http://lhcb-public.web.cern.ch/lhcb-public/en/Detector/Detector-en.html>, 2008.
- [4] T. Skwarnicki, *A study of the radiative CASCADE transitions between the Upsilon-Prime and Upsilon resonances*. PhD thesis, Cracow, INP, 1986.
- [5] L. Crow, “A feasibility study for the measurement of the cp asymmetry in the $D^0 \rightarrow \phi \gamma$ decay at the LHCb,” Master’s thesis, University of Manchester, May 2017.