

IMPERIAL COLLEGE LONDON

DEPARTMENT OF PHYSICS

---

# The Search for Axions in $B^0$ meson decays at LHCb

---

*Author:*

01726085

*Supervisor:*

Dr. Michael McCann

*Assessor:*

Dr. Mitesh Patel

*Word Count:*

9869

*Project Code:*

HEPH-McCann2

Submitted in partial fulfillment of the requirements for the MSci degree in Physics  
of Imperial College London

April 2024

## Abstract

In this analysis, signatures of ALPs were searched in the decay of  $B^0 \rightarrow K^{*0}a \rightarrow K^+\pi^-\gamma\gamma$ .  $B^0 \rightarrow K^{*0}\gamma\gamma$ . Signal events simulating  $B^0 \rightarrow K^{*0}\gamma\gamma$  were generated using software which was then adjusted for the L0 trigger efficiencies. After applying selection cuts, two XGBoost models were trained. One isolated the  $B^0$  decays from the LHCb data whilst the other removed the  $K1$  meson peak. The decay channel of  $B^0 \rightarrow K^{*0}\eta$  was used as a validation channel to ensure that the selection process was accurate. This was done by comparing the measured and theoretical  $\eta$  yield. The measured value was  $376 \pm 27$  whilst the theoretical value was  $392 \pm 49$ . Then, sWeighting was used to conduct a fit for  $B^0 \rightarrow K^{*0}\gamma\gamma$  and calculate the branching fraction  $BF(K^{*0} \rightarrow \gamma\gamma)$  which was determined to be  $8.52 \pm 1.16 \times 10^{-6}$ . This branching fraction was used to calculate the axion-boson coupling constant  $g_{aW}$  for axion masses from  $0 - 4388$  MeV.

---

## Acknowledgments

I want to thank my supervisor, Dr. Michael McCann, for his incredible support throughout this project. Having worked on this project alone, he has helped and guided me immensely, and this work would not have been possible without his mentorship. I would also like to thank Dr. Mitesh Patel for his advice regarding the project during the viva. A final thanks to my peers at Imperial College, who constantly encouraged and supported me during this journey.

---

## Layperson's Summary

If you were to take this paper and zoom in further, you would find fibers. Go even further, and you will find molecules, then atoms, till you end at neutrons, protons, and electrons. That's the end of the story, right? No. As it turns out, electrons are fundamental particles, but protons and neutrons can be split even further into smaller, more fundamental particles called quarks. Each of these quarks carries a charge, and the sum of the quark charges is the charge of the neutron or proton. It is postulated that these quarks move around freely inside the neutrons. Therefore, we would expect the neutrons to have a non-uniform charge distribution. This is called the neutron electric dipole moment.

Surprisingly, when the neutron electric dipole moment was measured, it was found that the charge distribution in a neutron was almost uniform despite the quarks moving randomly inside the neutron. This is known as the 'strong CP problem'. To resolve this issue, a new particle called the 'axion' was postulated, which would provide a mechanism by which the charge distribution in the nucleus would remain uniform. Although there are other proposed solutions to the strong CP problem, the axion remains the most popular one because of its simplicity and the ability of the axion to solve the strong CP problem without altering already known properties of particles. The objective of this research is to detect the axion particle.

The most obvious way to search for these particles is to open up these neutrons and protons. But how? ... by using the Large Hadron Collider. This is a collider that accelerates protons to near the speed of light and smashes them together at very high energies. The quarks are ripped from protons at these energies and start creating new and exotic particles. One of these particles is the  $B^0$  meson (mesons are 'two-quark' structures). When the  $B^0$  meson decays, an axion is produced. The axion then further decays to two photons. The analysis in this research focuses on looking for two photon signatures originating from a  $B^0$  meson. However, the axion cannot decay by itself. It requires another particle called the 'W boson'. This research measures the strength of the axion interaction with the W boson, known as the axion-boson coupling constant  $g_{aW}$ . Furthermore, this research provides new limits on the axion-boson coupling constant as a function of the axion mass in a region where others have not searched before. No evidence of the axion was found for these axion masses with the axion-boson coupling constant specified. Further research must be conducted to search different masses and coupling constants.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Background Theory</b>	<b>3</b>
2.1	QCD axions as a solution to the strong CP problem . . . . .	3
2.2	Proposed Search for Axions . . . . .	4
<b>3</b>	<b>The LHCb detector</b>	<b>6</b>
3.1	Truth cuts . . . . .	6
3.2	LHCb detector . . . . .	7
3.3	L0 and HLT triggers . . . . .	8
3.4	Calculating trigger efficiency . . . . .	9
3.4.1	Determining binning scheme . . . . .	10
3.4.2	Re-weighting Signal Events . . . . .	13
3.4.3	Total efficiency error . . . . .	13
<b>4</b>	<b>Selection</b>	<b>16</b>
4.1	Selection cuts . . . . .	16
4.2	XGBoost binary classification . . . . .	17
4.2.1	Feature selection . . . . .	17
4.2.2	Hyperparameter tuning . . . . .	21
<b>5</b>	<b>Validation and fitting</b>	<b>24</b>
5.1	$B^0 \rightarrow K^{*0} \eta$ fit . . . . .	24
5.1.1	K1 XGBoost model . . . . .	25
5.1.2	Measuring $\eta$ yield . . . . .	29
5.1.3	Calculating the efficiency factor . . . . .	30
<b>6</b>	<b>Axion-boson coupling</b>	<b>31</b>
6.1	sWeighting for $N_\eta$ and $N_{\gamma\gamma}$ . . . . .	31
6.2	Efficiency ratio . . . . .	34
6.3	Determining $g_{aW}$ . . . . .	35
<b>7</b>	<b>Conclusion</b>	<b>37</b>
<b>8</b>	<b>Appendix</b>	<b>38</b>
8.1	Truth Cuts . . . . .	38
8.2	Transverse Momentum and Pseudorapidity . . . . .	39

8.3	RooFit and Gradient Descent . . . . .	40
8.4	Crystal Ball Function . . . . .	41
8.5	Decision Trees . . . . .	42
8.5.1	XGBoost Algorithm . . . . .	42
8.5.2	Hyperparameters . . . . .	42
8.5.3	ROC Curves and AUC score . . . . .	43
8.5.4	K-Fold Cross Validation . . . . .	43
8.5.5	Youden's J statistic . . . . .	44
8.6	sWeights . . . . .	45
8.7	Bootstrapping . . . . .	46



# Chapter 1

## Introduction

The Standard Model (SM) is one of the most successful empirical theories of Particle Physics. The SM has accurately predicted certain experimental properties, such as the muon magnetic moment to 1 part per billion [1] and the existence of the Higgs boson [2].

Despite all its triumphs, the SM is known to be incomplete. For instance, it does not incorporate gravity into the model at all [3], nor does it explain the matter-antimatter asymmetry observed in the Universe [4]. Moreover, recent estimates suggest that Dark Matter (DM) constitutes roughly 27% of all matter [5], but the SM does not incorporate DM into the theory either. As a result, many Beyond the Standard Model (BSM) theories have been proposed to resolve these issues.

One of the limitations of the SM is the inability to explain the strong Charge-Parity (CP) [6] problem. The Quantum Chromodynamics (QCD) Lagrangian includes a CP-violating term. However, the strong force does not show CP-violation in nature. This was determined by measuring the neutron electric dipole moment (NEDM), which sets the CP-violating term to be smaller than  $10^{-27}$  [7]. There is no satisfactory theoretical justification as to why the CP-violating term is strongly suppressed for the strong force. To resolve this issue, a new particle called the axion was proposed by Peccei and Quinn [8], which negates the CP-violating term in the QCD Lagrangian.

The hypothetical axion particle is weakly interacting with matter, making it hard to detect. There have been numerous experiments that have tried to detect axions employing a range of techniques from cosmological axion searches to laser resonant cavities [9, 10, 11]. Due to their weak coupling with matter, axions have a longer lifetime. This property of axions has made them a strong candidate as a DM particle [12]. Furthermore, axions provide an insight into stellar structure as the weak coupling of axions allows them to carry energy over a period of time [13].

In this thesis, a search was conducted for axion-like particles (referred to as axions henceforth for brevity), a more general class of particles whose mass and coupling constant are independent of each other and do not necessarily couple to the gluon. The analysis involved examining  $B^0$  meson decays through  $B^0 \rightarrow K^{*0} a$  and then

a further decay to  $a \rightarrow \gamma\gamma$ .  $B^0$  meson decays are particularly interesting because of their sensitivity to New Physics and the high accuracy of experimental data available on  $B^0$  meson decays [14]. The  $B^0$  meson decay branching fractions are well documented, so any deviation would potentially indicate the presence of axions or, more likely, provide a better constraint to the values of the axion-boson coupling parameter  $g_{aW}$  further.

The advantage of searching this neutral  $B^0$  meson decay channel is due to the production of  $K^{*0}$  mesons which further decays via  $K^{*0} \rightarrow K^+\pi^-$ . The charged Kaons and Pions then deflect in the magnetic fields at LHCb, leading to more accurate vertex reconstruction than neutral hadrons. Moreover, the axion decay via  $a \rightarrow \gamma\gamma$  requires no minimum axion mass for axion production as photons can be massless, whereas investigating axion decay via hadronic modes requires a minimum mass for axion production.

The remainder of this thesis is divided into six further chapters. Chapter 2 motivates the requirement for axions and provides current constraints on  $m_a - g_{aW}$  parameter space. Chapter 3 describes the LHC experiment and the LHCb detector. The chapter also presents the simulation data for decay modes such as  $B^0 \rightarrow K^{*0} \eta$ , which was not generated in the analysis but already provided. Finally, the chapter discusses the need to account for the efficiency of the LHCb detector and the analysis strategy used to estimate the efficiency. Chapter 4 presents the selection cuts applied to the LHCb data and explains the analysis strategy to train and test machine learning models to identify  $B^0 \rightarrow K^{*0} \gamma\gamma$  decays. Chapter 5 outlines the need to train an additional machine learning model to remove contributions from partly reconstructed backgrounds. The chapter also explains the fits conducted using a combination of probability distribution functions to measure the yield of  $\eta$  mesons and compare to the theoretical yield of  $\eta$  mesons in the validation channel of  $B^0 \rightarrow K^{*0} \eta$  to ensure that the selection process is accurate. Chapter 6 presents the final fit for the  $B^0 \rightarrow K^{*0} \gamma\gamma$  channel and uses the branching fraction of  $BF(K^{*0} \rightarrow \gamma\gamma)$  to determine the axion-boson coupling constant  $g_{aW}$  for a range of axion masses. Finally, the chapter discusses the determination of errors using bootstrapping. Chapter 7 then concludes the work.

# Chapter 2

## Background Theory

The need for axions became apparent through the strong CP problem. The QCD Lagrangian is as described as follows

$$\theta \frac{g^2}{32\pi^2} G_{\mu\nu}^a \tilde{G}^{a\mu\nu} \in \mathcal{L}_{QCD}. \quad (2.1)$$

Here, the  $G_{\mu\nu}$  represents the gluon fields, and  $\tilde{G}_{\mu\nu}$  is defined as the dual vector where  $G_{\mu\nu} = \epsilon_{\mu\nu\alpha\beta} \tilde{G}^{\alpha\beta}$ . The second term in this equation is CP-odd; therefore, the strong force does not obey CP symmetry mathematically. However, the  $\theta$  term corresponds to an observable quantity that can be measured via the NEDM. The current most accurate measurements constrain the NEDM to  $d_n < 2.9 \times 10^{-26} \text{ e cm}$  [15] which then restricts  $\theta \leq 10^{-9}$ .

### 2.1 QCD axions as a solution to the strong CP problem

The simplest solution to resolve the strong CP problem would be to introduce a particle with coupling constant  $f_a$  and a field  $a(\mathbf{r}, t)$  such that it couples to the CP violating term in the QCD Lagrangian in equation 2.1. Therefore, the modified QCD Lagrangian becomes [16],

$$\left( \theta + \frac{a}{f_a} \right) \frac{g^2}{32\pi^2} G_{\mu\nu}^a \tilde{G}^{a\mu\nu} \in \mathcal{L}_{QCD}. \quad (2.2)$$

The QCD axion requires that the minimum energy of the particle should be achieved, implying  $\theta + \frac{a}{f_a} = 0$ . This means the CP-violating term disappears completely from the QCD Lagrangian.

Although QCD axions provide an elegant solution to the strong CP problem, there are problems associated with the theory. For example, there is no exact prediction for either the QCD axion mass or the QCD axion coupling to SM particles. Therefore, extensive searches need to be conducted to constrain these coupling parameters. Furthermore, the QCD axion has compatibility issues with other prominent BSM

theories such as quantum gravity, string theory, and others [17, 18].

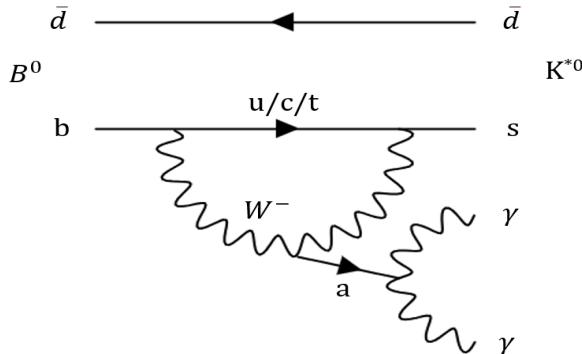
## 2.2 Proposed Search for Axions

In this thesis, axion production was investigated through  $B^0$  meson decays due to their sensitivity to new Physics.

The Lagrangian for the axion-boson interactions for this decay is given by [19]

$$\mathcal{L} = -\frac{1}{4} g_{aW} a(\mathbf{r}, t) W_{\mu\nu}^b \tilde{W}^{b\mu\nu}, \quad (2.3)$$

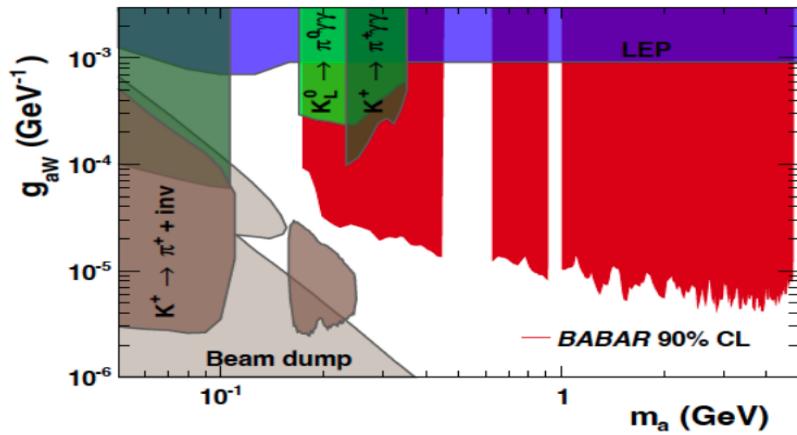
where  $a(\mathbf{r}, t)$  still represents the axion field,  $W_{\mu\nu}^b$  represents the boson field and its dual vector  $\tilde{W}^{b\mu\nu}$ . The  $g_{aW}$  denotes the axion-boson coupling constant. The axion then further decays into two photons. For  $m_a \leq m_W$ , the branching fraction is almost 100% for this decay.



**Figure 2.1:** This figure shows the Feynman diagrams for the decay channel  $B^0 \rightarrow K^{*0} \gamma\gamma$ . The  $K^{*0}$  meson then further decays to  $K^{*0} \rightarrow K^+ \pi^-$  [20].

The axion can also be searched for in  $\bar{B}^0$  meson decays which has a similar Feynman diagram to figure 2.1 but the process is mediated by a  $W^+$  boson instead. For the analysis, the distinction between  $B^0$  and  $\bar{B}^0$  mesons was not required because both decays will produce an axion.

There have been recent experiments conducted at BaBar, which have constrained the current estimates for the  $m_a - g_{aW}$  parameter space. The BaBar detector searched for  $B^\pm \rightarrow K^\pm a$  decays followed by  $a \rightarrow \gamma\gamma$  in the mass range  $0.175 \text{ GeV} \leq m_a \leq 4.78 \text{ GeV}$ . However, the mass intervals  $0.45 - 0.63 \text{ GeV}$  and  $0.91 - 1.01 \text{ GeV}$  are removed due to backgrounds from  $\eta$  and  $\eta'$  mesons respectively. Figure 2.2 shows the constraints on the parameter space.



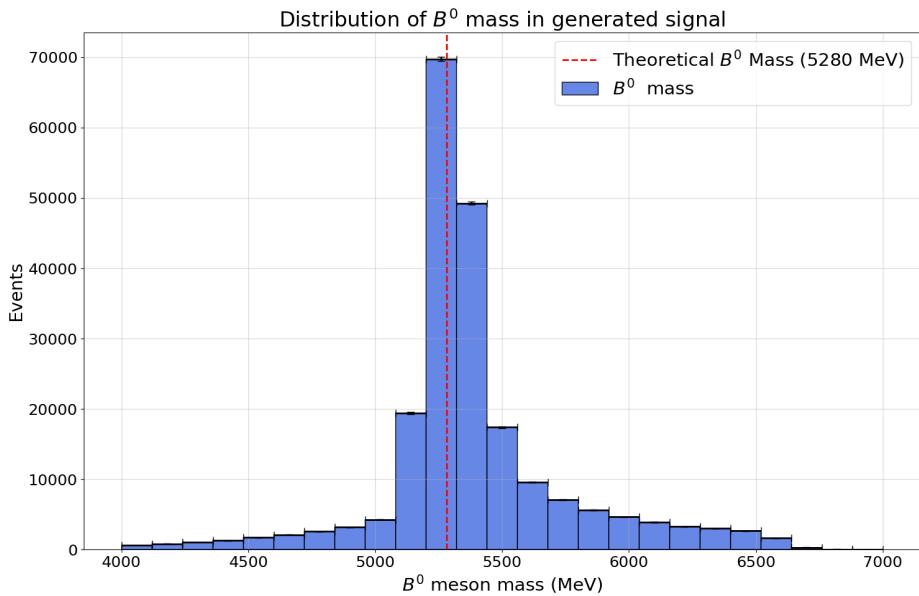
**Figure 2.2:** This figure shows the  $m_a - g_{aW}$  parameter space with the red sections highlighting the recent results from BaBar at 90% confidence interval. The notches in the data represent the removed data due to the  $\eta$  and  $\eta'$  meson backgrounds [21].

The advantage of searching for axions in the decay chain of  $B^0 \rightarrow K^{*0} \rightarrow K^+ \pi^- \gamma\gamma$  is that the axion mass ranging from 0 - 4388 MeV can be searched. The upper limit is determined from the  $B^0$  rest mass minus the  $K^{*0}$  rest mass.

# Chapter 3

## The LHCb detector

To investigate the decay of  $B^0$  mesons, signal events were first generated using software such as EVTGEN and GEANT4 to simulate the  $B^0 \rightarrow K^{*0}\gamma\gamma$  decay. The plot of the  $B^0$  meson mass distribution from these simulated software decays is shown in figure 3.1.



**Figure 3.1:** This figure shows the distribution of the  $B^0$  meson mass in the generated signal events with the red line indicating the theoretical  $B^0$  mass of 5280 MeV.

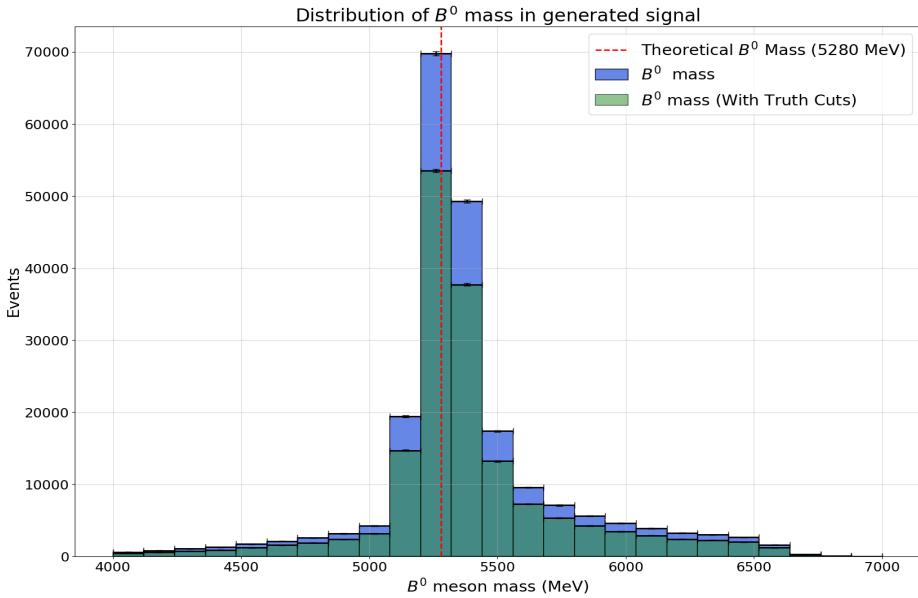
The decays simulated originated from a  $B^0$  meson, hence, the peak around the  $B^0$  mass of 5280 MeV. The width came from the measurement and reconstruction errors.

### 3.1 Truth cuts

These softwares simulated a plethora of decay chains for the  $B^0$  meson, however, only the decay  $B^0 \rightarrow K^{*0}a \rightarrow K^+\pi^-\gamma\gamma$  was of interest. As certain particles must

exist in our decay channel, such as the  $B^0$  and  $K^{*0}$ , therefore, Particle Data Group Monte-Carlo (PDG MC) codes were applied to the signal events which isolated the decay chain of interest [22]. A full list of truth cuts can be found in the appendix 8.1.

Figure 3.2 shows the  $B^0$  mass again with the truth cuts applied:



**Figure 3.2:** This figure shows the  $B^0$  mass distribution with the truth cuts applied in green. These  $B^0$  events corresponded to the decay chain of interest.

## 3.2 LHCb detector

The Large Hadron Collider (LHC) at the European Council for Nuclear Research (CERN) is one of the world's most powerful tools for Particle Physics research. The collider is designed to collide proton-proton beam at a center-of-mass energy of 14 TeV, which provides an unprecedented insight into probes of BSM Physics. There are 1232 superconducting dipole magnets, each 15 meters in length, which bend the beam, and 392 quadropole magnets, about 5 - 7 meters long, which focus the proton beams for collision. The proton beams are made to collide at four points inside the 26.7 km diameter tunnel. These four collision points correspond to four different particle detectors: ATLAS, CMS, ALICE, LHCb [23].

The Large Hadron Collider beauty (LHCb) is made for primarily observing b quark decays, which mainly focuses on answering questions such as CP violation and matter-antimatter symmetry. The LHCb can search for decay modes that are forbidden in the Standard Model. However, these rare decay modes predicted by different types of BSM theories are so suppressed that lots of statistics are required to prove extensions to the SM and not be attributed to statistical fluctuations due to a lack of understanding of SM parameters. The LHCb produces approximately  $10^{12}$

$b\bar{b}$  events in one year. The LHC proton-proton collisions provide an abundance of quarks that decay quickly; therefore, the LHCb contains movable tracking detectors that can be moved close to the proton-proton beam. This is different from detectors like ATLAS or CMS, whereby the whole collision point is enclosed by a detector [24].

The LHCb detector has a unique forward cone design because most of the b quarks produced from the proton-proton collision are produced in a forward or backward cone [25]. The range of angles covered by the LHCb detector goes from 10 mrad to 250 mrad, as shown in figure 3.3.



**Figure 3.3:** This figure shows the layout of the LHCb detector. Unlike the ATLAS or CMS detectors, the LHCb detector contains many sub-detectors totaling about 20 meters in length. The vertex locator on the left is the closest to the point of collision, enabling accurate vertex reconstruction. The Ring-Imaging-Cherenkov-Detectors (RICH) allow the identification of charged hadrons. The trackers T1, T2, and T3 track the path of the particles. The Electromagnetic Calorimeter (ECal) measures the energy of the electrons and photons whilst the Hadronic Calorimeter (HCal) measures the energy of the hadrons. The magnets are used to bend charged particles and thus calculate their momenta. The green sections on the right represent the muon systems that track and detect muons.

The generated signal events that were previously created represented the  $B^0 \rightarrow K^{*0}\gamma\gamma$  decay channel very well, but the softwares used for simulating the signal events did not account for the efficiency of the detectors (the number of generated signal events that survive the trigger cuts) at the LHCb. Therefore, the efficiency of the simulated signal events was determined manually later in the thesis.

### 3.3 L0 and HLT triggers

The LHCb can track and detect b and c quarks efficiently because any hadrons containing b or c quarks have the following properties: large lifetimes, high masses ( $\geq 1\text{GeV}$ ), and usually contain muons in the final decay state. These properties enable the LHCb detector to accurately determine b or c quark containing hadrons

from all the harsh hadronic backgrounds generated from proton-proton collisions. However, there are still far too many events that are collected from the LHCb detector to be analysed. Therefore, the LHCb contains a trigger system that makes a split-second decision on whether an event is of interest, thus reducing the data read-out from the LHCb. The trigger system is split into two different parts: a hardware trigger (Level 0 trigger) and a software trigger (High-Level Trigger).

The L0 trigger has to be hardware-based for computational speed because the L0 trigger has to decide whether an event is interesting enough or not in under  $4 \mu\text{s}$ . To achieve this, the L0 trigger mainly uses the calorimeter and muon systems (as shown in figure 3.3) because these are the only pieces of information available at such a high rate. The L0 trigger has three components: L0 electron trigger, L0 photon trigger, and L0 hadron trigger meant to detect the respective particles. The L0 trigger selects muon events with significant transverse momenta (refer to appendix 8.2) satisfying the condition  $p_T > 1.76 \text{ GeV}$  whilst the electron and photon triggers require transverse energy deposits of  $E_T > 3 \text{ GeV}$ . The L0 hadron trigger requires a transverse energy deposit of  $E_T > 3.68 \text{ GeV}$  [26].

The HLT trigger has two components, HLT1 and HLT2. HLT1 partially reconstructs the tracks whilst HLT2 fully reconstructs the tracks that pass the HLT1 trigger. As the HLT triggers are usually 90% or more efficient, therefore, in this analysis the effect of the HLT triggers on the particles were ignored.

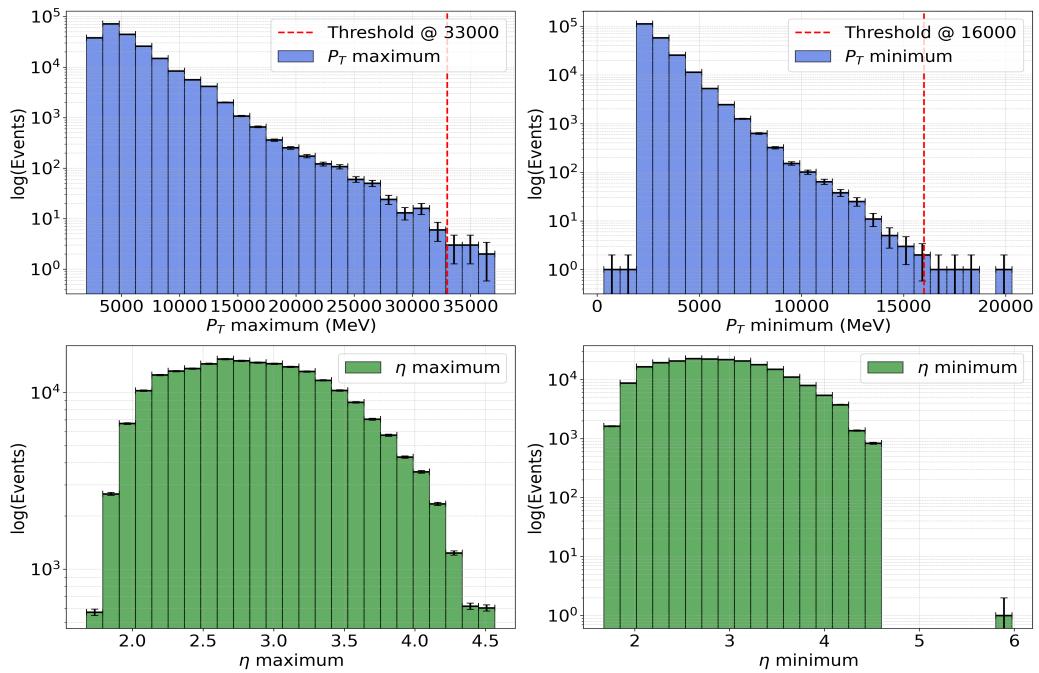
The L0 electron and photon triggers are approximately 70% efficient whilst the L0 hadron trigger is roughly 40% efficient but can fall as low as 20% for certain decay modes. Therefore, the L0 hadron trigger was not used in this analysis.

## 3.4 Calculating trigger efficiency

To investigate the L0 electron and photon trigger efficiency for the signal events, the data was split into four variables maximum  $P_T$ , minimum  $P_T$ , maximum  $\eta$  and minimum  $\eta$ , where  $P_T$  represents the transverse momentum and  $\eta$  here, represents pseudorapidity (refer to appendix 8.2). The L0 triggers depend mainly on the variables  $P_T$  and  $\eta$ , so the signal events were split into these variables. Each signal event produces two photons; the photon with the higher transverse momentum is assigned maximum  $P_T$ , and the other photon is assigned minimum  $P_T$ . The maximum  $\eta$  is the pseudorapidity of the maximum  $P_T$  events and vice versa for minimum  $\eta$ .

A calibration data set was also used, mimicking photons coming from  $\eta \rightarrow \gamma\gamma$ . The photons produced from the decay sometimes generated an electron-positron pair. Hence, there was a need to consider the hits on both the L0 photon and L0 electron triggers.

The fiducial cuts in figure 3.4 removed some stand-alone events in the high  $P_T$  and  $\eta$  regions. To determine the efficiency of the generated signal events, they were grouped into bins of  $P_{T\text{s}}$  and  $\eta\text{s}$ .



**Figure 3.4:** This figure shows the distribution of the four variables mentioned for the generated signal events. There are two red lines for both the  $P_T$  variables, which are the fiducial cuts that were later applied to the signal data. These fiducial cuts were introduced to overlap signal and calibration data.

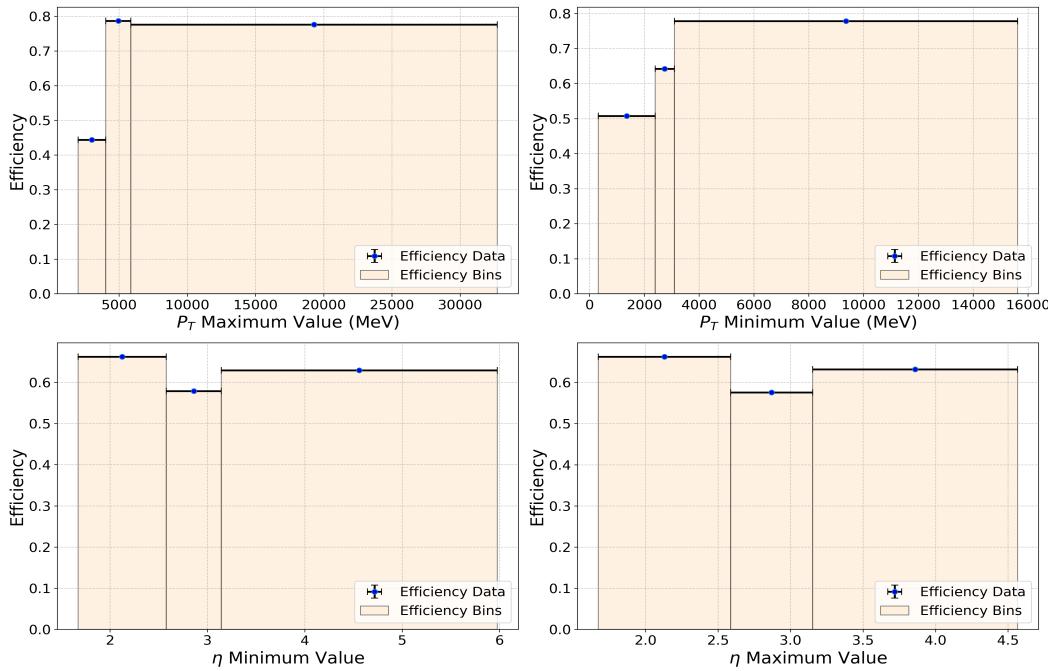
### 3.4.1 Determining binning scheme

Examples of three and five bins for each of the variables are shown in figures 3.5 and 3.6.

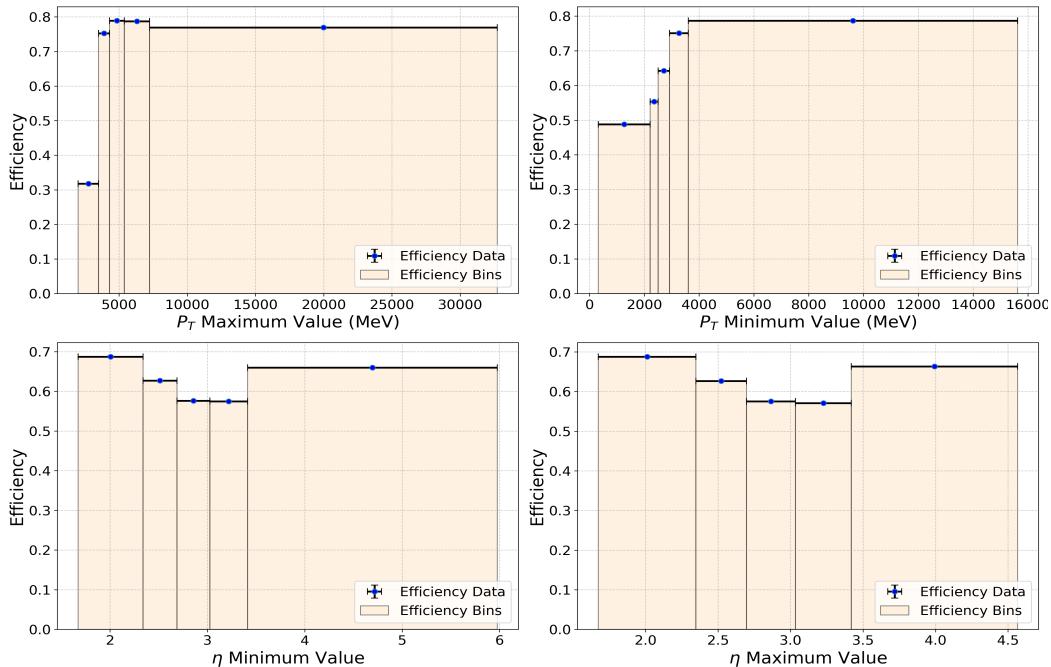
For high  $P_T$  events, the efficiency had a large value because those events almost always passed the L0 trigger cuts, and similarly, for low  $P_T$  events, the efficiency had a lower value shown by the figures 3.5 and 3.6. The strategy, therefore, was to choose enough bins to capture the variation of efficiency between the two extremes but also to ensure not to choose too many bins such that in 4D, there were not enough events per bin (because of the four variables as discussed already).

Once a binning scheme had been chosen, the binning scheme was checked in 2D, shown in figure 3.7 to ensure that the efficiency did not vary rapidly between bins.

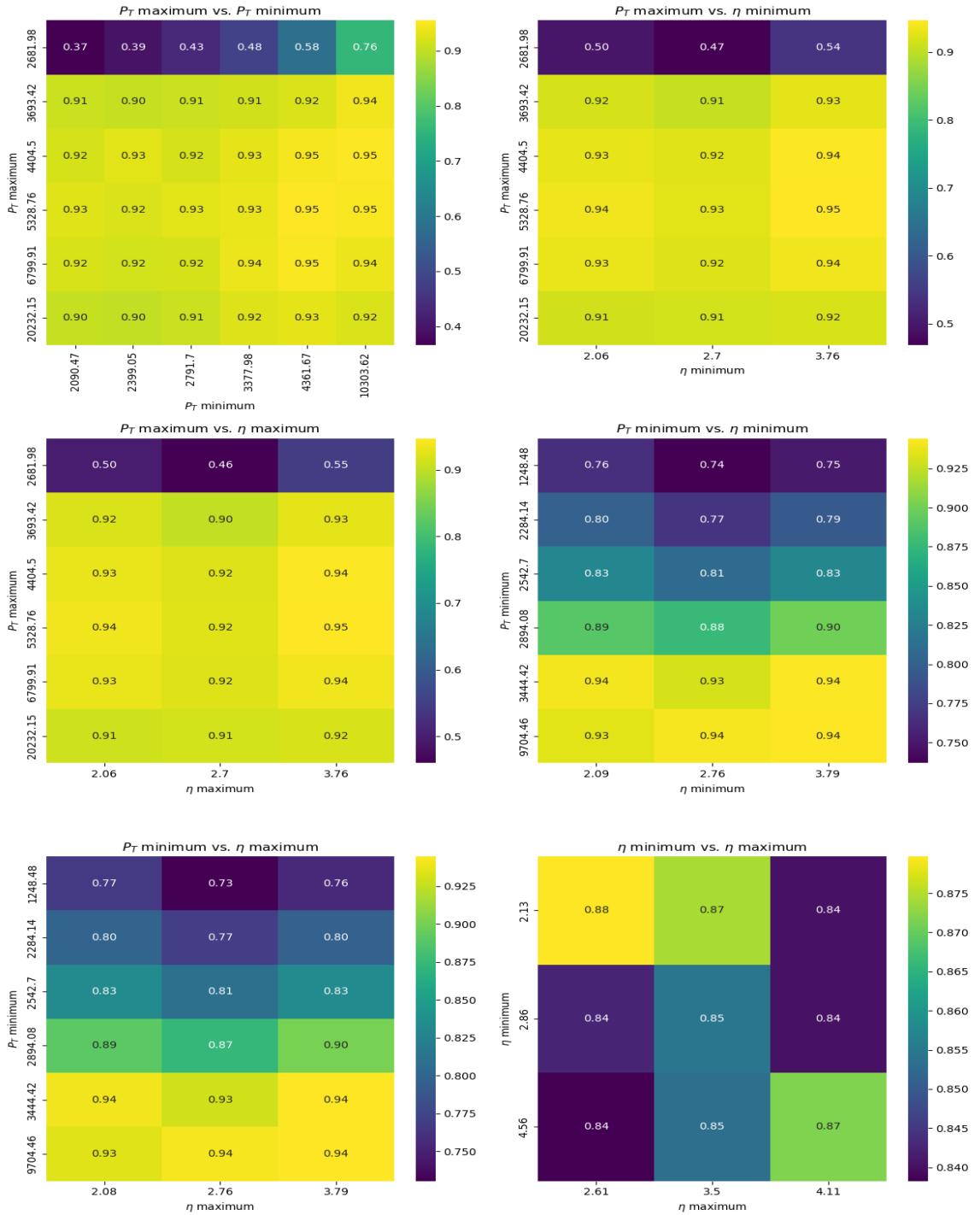
Ideally, a check should have been carried out in 3D and 4D to ensure that the efficiencies do not vary rapidly between bins. However, this was not done in the interest of computational time. The final binning scheme chosen for this analysis was 6 bins in  $P_T$  and 3 bins in  $\eta$ .



**Figure 3.5:** This figure shows the variation of the efficiency with respect to the four different variables in bins of three. The errors on the efficiency are Binomial rather than Poissonian, as the latter would result in unphysical error ranges for the efficiency (discussed in detail in [27]).



**Figure 3.6:** This figure shows the four variables in bins of five.



**Figure 3.7:** This figure shows the variation of all of the four variables with each other in 2D. The efficiency was roughly estimated in each bin as the number of events that passed the LO trigger cuts divided by the total number of events in each bin. This heat map aimed to ensure that there were no sudden changes in efficiency between the bins.

### 3.4.2 Re-weighting Signal Events

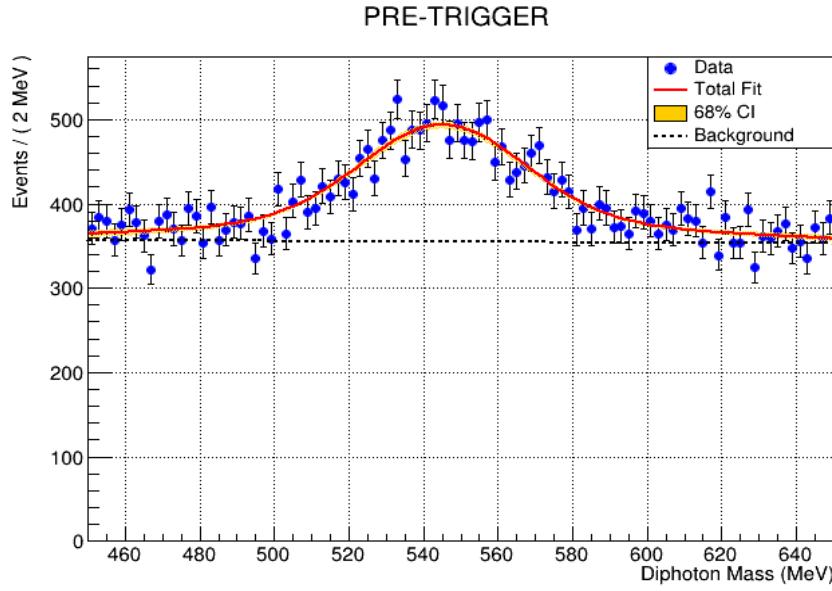
In the previous cases, the efficiency was estimated simply as the number of events that pass a cut divided by the total number of events. However, this does not take into account some of the background photon events that do not originate from our decay. Therefore, a fit needs to be done only to extract photon events corresponding to our decay chain. The RooFit software was used to fit all data in this analysis (refer to appendix 8.3).

The fit was conducted over the diphoton mass (the invariant mass of the two photons). A double crystal ball (refer to appendix 8.4) function was used to model the asymmetric signal distribution with two power law tails on either end whilst a linear function was used to model the background. Alternatively, Legendre, Chebyshev, or higher-order polynomials could have been used to model the data. However, the additional parameters required by those functions would have increased the computational complexity of the fit and would have led to convergence errors due to the larger number of floating parameters that needed to be fit. The photon peak shape was first fixed using the diphoton mass spectrum from the signal events. Then, the photon peak and the linear background were fitted to the calibration data diphoton mass. The shape of the photon peak was fixed for two reasons: firstly, the generated signal events will model the physics of the diphoton signal. Secondly, by fixing the photon peak shape, the number of parameters in the final fit was reduced, thereby reducing the fit complexity and decreasing the computational cost of the RooFit algorithm. 4D bins in the variables  $P_T$  and  $\eta$  were made, corresponding to 324 bins in total (6 bins for the  $P_T$  variables and 3 bins for the  $\eta$  variables). A sample fit for a bin corresponding to the kinematic region  $1993.55 < P_T \text{ maximum} \leq 3496.28$ ,  $323.26 < P_T \text{ minimum} \leq 2118.33$ ,  $1.68 < \eta \text{ maximum} \leq 2.96$ ,  $1.69 < \eta \text{ minimum} \leq 2.54$  bin is shown in figure 3.8.

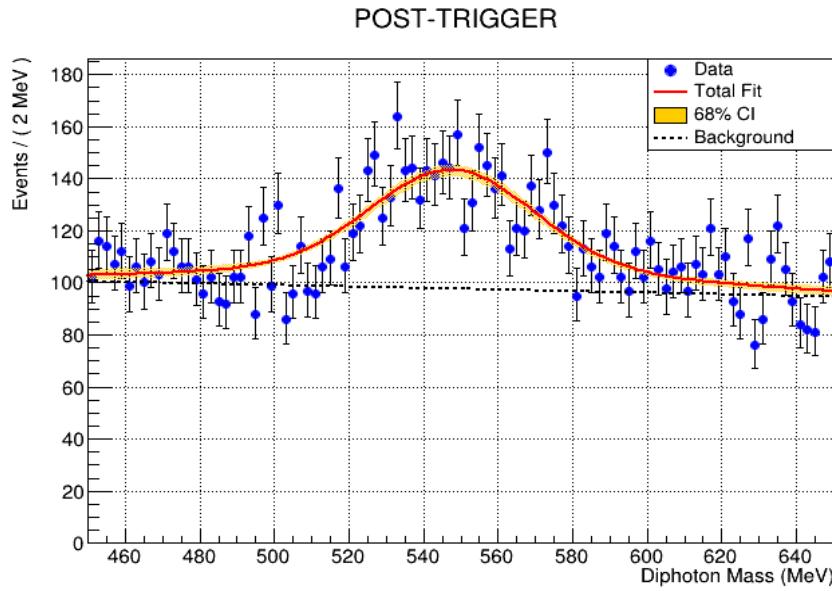
To ensure that the RooFit algorithm found the global minimum in figures 3.8 and 3.9, the starting parameter values were varied by about 10% of the starting value. The algorithm converged to the same point, indicating that it had converged to the global minimum. The efficiency of this particular bin was calculated by taking the ratio of signal events from figures 3.8 and 3.9.

### 3.4.3 Total efficiency error

There were three sources of errors for the efficiency calculated in each bin. There was a statistical error when calculating the number of events that pass the trigger cuts, which were calculated assuming binomial errors as discussed in section 3.4. There was also a fitting error from the RooFit algorithm, and this error is indicated as the 68% confidence interval in figures 3.8 and 3.9.



**Figure 3.8:** This figure shows the fit in one of the bins out of the 324 bins. The red line is the total fit, a combination of two double crystal balls and a linear polynomial background. The black line shows the linear background projected from the fit. The orange area around the total fit indicates the error of the fit at a 68% confidence interval. The small area of the confidence interval indicated a smaller error, showing that the fit quality was good.



**Figure 3.9:** This figure shows the same fit for the bin but with the trigger cuts for the LO electron and photon trigger applied. The number of events has been reduced compared to figure 3.8 as the triggers have not detected certain events.

As the number of bins chosen can vary, there was a systematic error associated with changing binning schemes. To determine the systematic error, the total efficiency from all the bins was calculated first using,

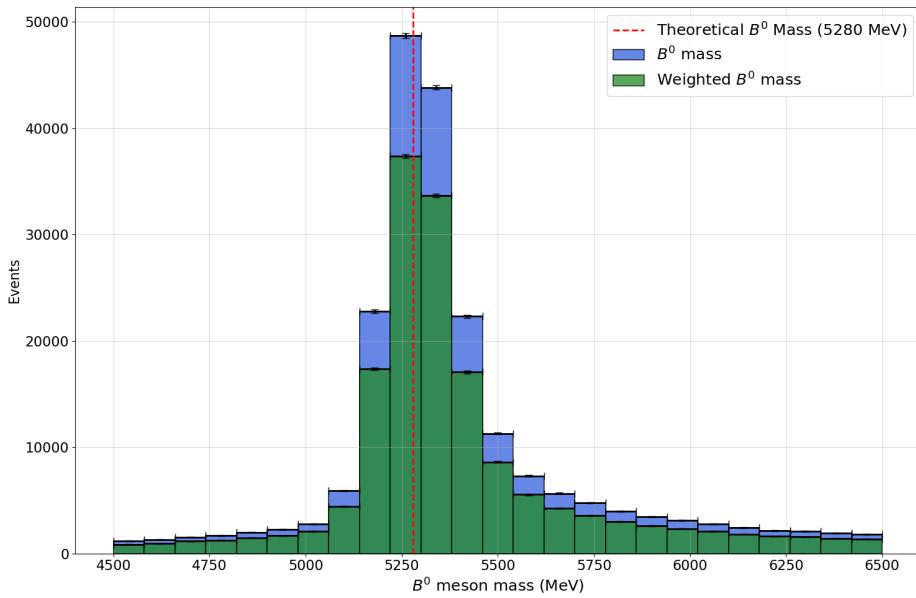
$$\epsilon_T = \frac{\sum_N \epsilon_i}{N}. \quad (3.1)$$

Here,  $\epsilon_T$  represents the total efficiency, and  $\epsilon_i$  represents the efficiency in each bin, whilst  $N$  represents the number of events. Once the total efficiency was determined, the binning scheme was changed from the 6 bins in  $P_T$  and 3 bins in  $\eta$  to 8 bins in  $P_T$  and 4 bins in  $\eta$ . Then, the total efficiency was calculated again, and the difference in total efficiency yielded the systematic error. In this case  $\delta\epsilon_{sys} = 0.012$ . Therefore, the total error in each bin was given by,

$$\delta\epsilon_i^2 = \delta\epsilon_{sys}^2 + \delta\epsilon_{fit}^2 + \delta\epsilon_{stat}^2. \quad (3.2)$$

The errors were added in quadrature because the three errors are independent of each other. In the end, each region of parameter space defined by  $P_T$  and  $\eta$  had an efficiency and uncertainty associated, which was used to re-weight the generated signal events to account for the L0 trigger.

Figure 3.10 demonstrated how certain generated signal events were discarded mainly because those events did not possess enough transverse momenta or possess the correct range of pseudorapidities to register a hit on the L0 triggers.



**Figure 3.10:** This figure shows the same  $B^0$  mass from the generated signal events, but the green distribution now represents the  $B^0$  mass accounting for the L0 trigger efficiency. In this case, the total efficiency was roughly 75%.

# Chapter 4

## Selection

The data presented in this analysis was a smaller subset of the 2018 LHCb data corresponding to a luminosity of  $2\text{fb}^{-1}$ . Three main types of backgrounds were removed from the data to isolate the  $B_0$  meson signal. The three types of backgrounds were:

**Combinatorial Background:** This background was formed from any combination of particles that replicated the final state particles of the decay chain. An exponential distribution usually models the combinatorial background.

**Partly Reconstructed Background:** Partly Reconstructed Background refers to the case when certain particles are misidentified at the LHCb, causing them to mimic the decay products. In this case, the source of the most important partly reconstructed background was the  $K^{*0}$  meson, which the LHCb misidentified as the  $K1$  meson.

**Peaking Background:** This background refers to peaks from other processes that mimic  $B^0$  signal peaks. In this decay chain, the main peaking backgrounds will be from  $B^0 \rightarrow K^{*0}\eta$ ,  $B^0 \rightarrow K^{*0}\pi^0$  and  $B^0 \rightarrow K^{*0}\eta'$ . For this particular decay mode, if the axion were to lie near the mass ranges of the  $\eta$ ,  $\eta'$  and  $\pi^0$ ; it would not be possible to find the axion using this analysis because the production of peaking backgrounds was dominated by QCD effects.

### 4.1 Selection cuts

To begin filtering out the data, cuts were applied to the background events. A particle identification (PID) cut was applied for the kaon and pion because a kaon and pion must be produced in the decay. The PID cuts used the RICH detectors (figure 3.3) to identify the charged hadrons emitted from a collision. In this analysis, PID cuts of  $P(K^+) > 0.1$  and  $P(\pi^-) > 0.1$  were chosen.  $P(K^+) > 0.1$  represents the probability that a detected particle was a Kaon and vice-versa for  $P(\pi^-) > 0.1$ . The probability cuts were chosen randomly. Hence, the values of the chosen cuts were low, meaning that the hadrons discarded had a less than 10% chance of being a Kaon or Pion. This loose cut ensured that events that may have been part of the  $B^0 \rightarrow K^{*0}\gamma\gamma$  decay were not discarded.

Additionally, as a  $K^{*0}$  meson was produced, an additional cut was applied to the  $K^{*0}$  meson mass. Therefore, a cut was imposed on the data such that the  $K^{*0}$  mass lay in the range 792 – 992 MeV as the theoretical  $K^{*0}$  mass is 892 MeV. The 100 MeV mass range was applied to account for measurement and detection errors.

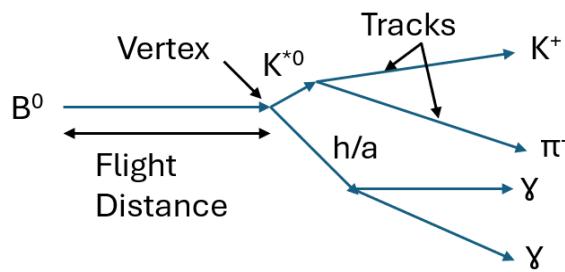
These initial cuts reduced the size of the dataset. Afterward, an XGBoost (refer to appendix 8.5.1) machine learning model was applied to isolate the  $B^0$  signal events from the dataset further.

## 4.2 XGBoost binary classification

To train the XGBoost model, the weighted signal events previously generated were used to mimic the  $B^0$  signal in the LHCb data. The background data was obtained by applying a mass cut  $m_{B^0} > 5800$  MeV (henceforth referred to as the upper mass sideband) on the data. This was done because the decay chain originates from a  $B^0$  meson. Therefore, in the data, it was expected that the  $B^0$  events would lie around the theoretical mass of the  $B^0$  at 5280 MeV. Hence, it was assumed that there would be few to almost no  $B^0$  events above the 5800 MeV mass, even accounting for measurement errors and detector effects.

### 4.2.1 Feature selection

The XGBoost model was supplied with features (refer to appendix 8.5.1) to differentiate between signal and background. These features were selected primarily from the decay topology and kinematics as shown in figure 4.1. Then, some features were added to improve the performance of the model.



**Figure 4.1:** This figure shows the topology of the decay chain. The  $h/a$  indicates the production of a hadron or an axion and their subsequent decay to two photons.

The vertex indicated in figure 4.1 was the primary vertex from which the  $B^0$  decays. The secondary vertex was the point from which the  $K^{*0}$  decays. However, as the  $K^{*0}$  lifetime is small, the distance between the primary and secondary vertex was

assumed to be negligible for practical purposes.

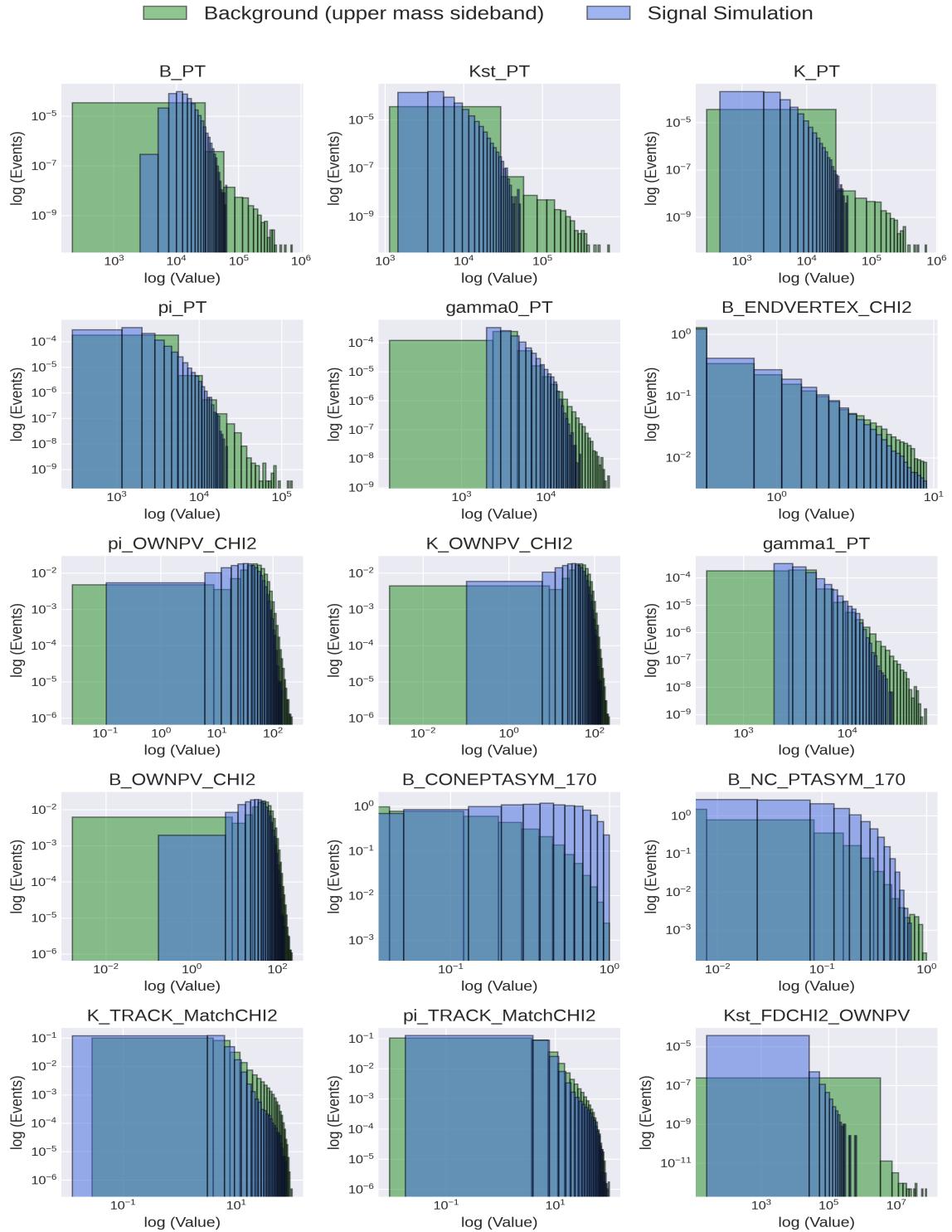
A total of 15 features (explained in detail in table 4.1) were selected to isolate the  $B^0 \rightarrow K^{*0}\gamma\gamma$  events in the LHCb data. The distribution of the weighted  $B^0 \rightarrow K^{*0}\gamma\gamma$  events compared to the upper mass sideband background is shown in figure 4.2 for all 15 features.

To determine if a feature improved the performance of a model or not, the model was trained and tested on the LHCb data, and the area-under-curve (AUC) score (refer to appendix 8.5.3) of the XGBoost model was evaluated. The feature was removed, and then the model was re-trained and tested again. The AUC score was re-evaluated and compared to the previous AUC score. If the AUC score was higher compared to the previous score, then it was deemed that the feature improved the performance of the XGBoost model.

The background from the LHCb data (upper mass sideband) was split into 10-folds (refer to appendix 8.5.4), and the weighted signal events were also divided into 10-folds. To train the XGBoost models, nine folds were combined and used for training, and one was used for testing. This corresponded to a 90% training and 10% testing split for the XGBoost model. Training lets the XGBoost model learn the difference between signal and background events based on the 15 features identified previously. Testing allows the XGBoost model to apply the learned knowledge to separate known signal and background events so that the performance of the model can be evaluated. The testing process prevents the XGBoost model from overfitting or underfitting (refer to appendix 8.5.4). This process was repeated so that 10 XGBoost models were trained. The models were trained in such a way that no two models had the same training and testing dataset.

Variable Name	Reason
B_PT	Transverse momentum of $B^0$ . This was included because the momentum would differ for the decay chain of interest and other background processes.
Kst_PT	Transverse momentum of $K^{*0}$ . Same reason as above.
K_PT	Transverse momentum of $K^+$ . Same reason as above.
pi_PT	Transverse momentum of $\pi^-$ . Same reason as above.
gamma0_PT	Transverse momentum of one $\gamma$ . Same reason as above.
gamma1_PT	Transverse momentum of the other $\gamma$ . Same reason as above.
B_ENDVERTEX_CHI2	$\chi^2$ values of the $B^0$ vertex. This was added because the decay chain must originate from a $B^0$ meson. The vertex fit's quality can indicate the event reconstruction's reliability.
pi_OWNPV_CHI2	$\chi^2$ values of the $\pi^-$ primary vertex. This variable was included because the kaon and pion must originate from a single point
K_OWNPV_CHI2	$\chi^2$ values of the $K^+$ . same reason as above.
pi_Track_MatchCHI2	$\chi^2$ values of the Pion track. This variable tracked the flight of the Pion and was included because the Kaon and Pion tracks would have deflected in different directions due to the opposite charges.
K_Track_MatchCHI2	$\chi^2$ values of the Kaon track. Same reason as above.
Kst_FDCHI2_OWNPV	$\chi^2$ values of the flight distance of the $K^{*0}$ meson. The $K^{*0}$ meson was expected to decay quickly to the $K^+$ and $\pi^-$ mesons. Therefore, the flight distance would have been different compared to other background decays.
B_CONEPTASYM_170	This represents the difference in the transverse momenta of the particles in a cone around a $B^0$ meson. This variable was added to improve the performance of the model.
B_NC_PTASYM_170	This represents the difference in the transverse momenta of the neutral particles in a cone around a $B^0$ meson. This variable was added to improve the performance of the model.

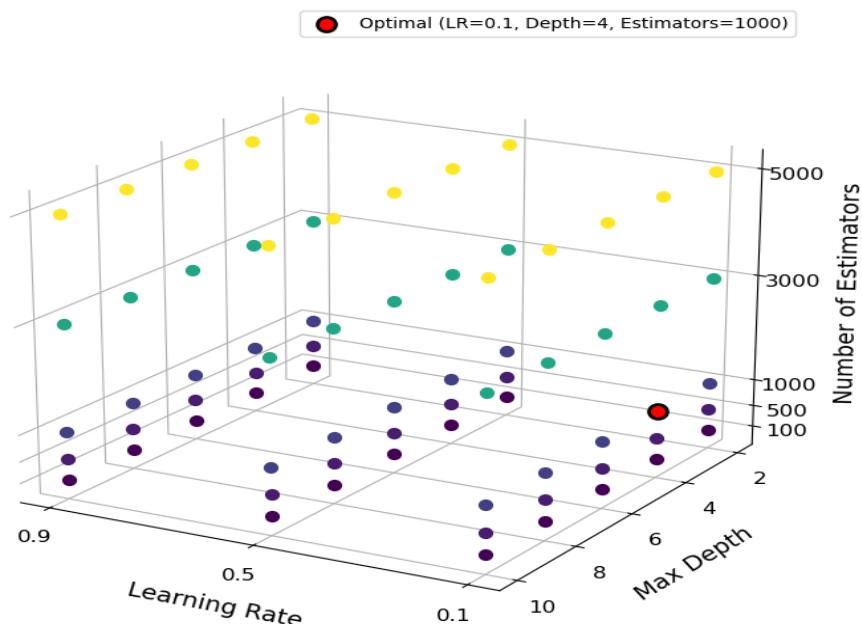
**Table 4.1:** This table lists all the features used for the XGBoost model. For some features, the  $\chi^2$  value was used because the differences in the features mattered and not the absolute values of the differences. The latter two variables were introduced to improve the AUC score of the model



**Figure 4.2:** This figure shows the different distributions for the features chosen for the generated signal events and the background data events. For clarity, the error bars have been omitted.

### 4.2.2 Hyperparameter tuning

To optimise the performance of the XGBoost model, the hyperparameters (refer to appendix 8.5.1): maximum depth, learning rate, and the number of estimators needed to be fine-tuned to improve the AUC score of the model as much as possible. To do this, a coarse 3D grid search was conducted in all three hyperparameters to identify the region where the optimal hyperparameters were located. The hyperparameters were varied for each of the 10 XGBoost models, and the model was tested on the corresponding fold. For example, model 0 was tested on fold 0, and model 1 was tested on fold 1, and so forth. The hyperparameter value that gave the best AUC score was the optimal point. The reason for performing a coarse 3D grid search was to reduce the computational time associated with training and testing XGBoost models. The analysis strategy is highlighted in figure 4.3.



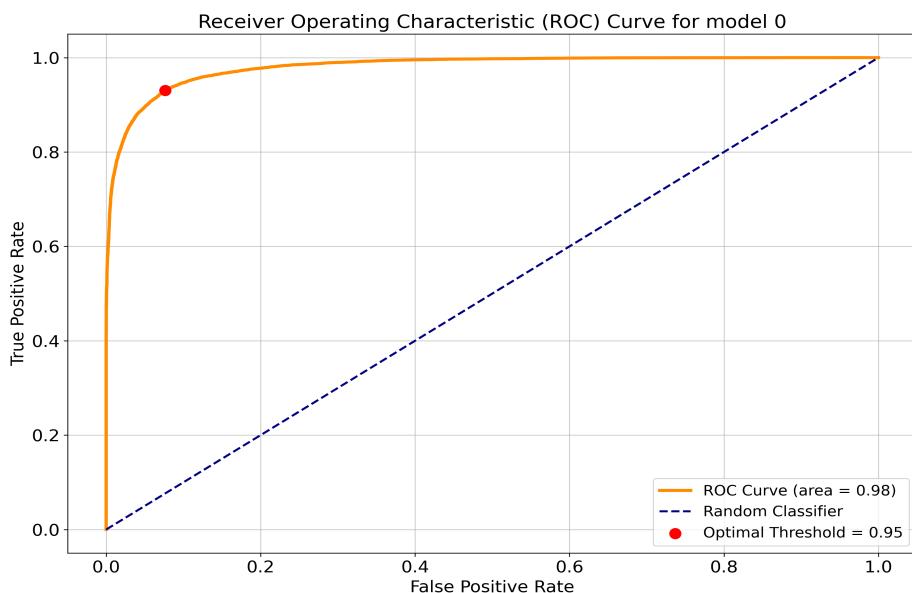
**Figure 4.3:** This figure shows the 3D grid search used and the corresponding values used in the search. The red point indicates the values where the optimal hyperparameters were found.

For all of the 10 XGBoost models, the optimal region learning rate was 0.1, the optimal max depth was 4, and the optimal number of estimators was 1000. Once the region of optimal hyperparameters was roughly identified, a finer search was conducted around these hyperparameter values to determine the optimal hyperparameters that gave the highest AUC score for all the XGBoost models. The finer search involved searching around the values of the learning rate 0.01 - 0.20 in increments of 0.01. The max depth was searched from 2 to 6 in increments of 1, and the number of estimators searched was from 800 to 1200 in increments of 100. The final optimal hyperparameters for all 10 of the XGBoost models were 0.02 for the

learning rate, 3 for the maximum depth, and 1000 for the number of estimators.

Finally, all 10 of the XGBoost models were applied to the LHCb data so each event had a probability assigned to it. The probability corresponded to the likelihood of the event belonging to  $B^0 \rightarrow K^{*0}\gamma\gamma$  decay or background. For events with a  $B^0$  mass  $m_{B^0} > 5800$  MeV, the models were applied on the corresponding fold, as explained above. For mass,  $m_{B^0} \leq 5800$  MeV, all 10 XGBoost models were applied to each event, and the probability was determined depending on which model gave the highest probability for that event.

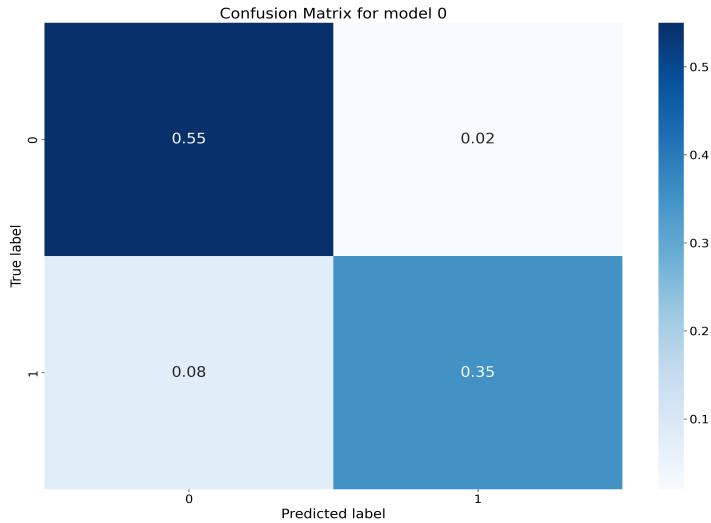
The optimal threshold is the probability cut-off value such that any events with a higher probability than the cut-off value were deemed  $B^0 \rightarrow K^{*0}\gamma\gamma$  events, and the events with a lower probability were deemed background. To determine the optimal threshold, Youden's J statistic (refer to appendix 8.5.5) was used, which yielded an optimal threshold of 0.95 for all 10 of these XGBoost models.



**Figure 4.4:** This figure shows the ROC curve for a particular model. The AUC score is 0.98, which indicates a high level of accuracy. The red point on the ROC curve shows the optimal threshold point where the true positive rate is maximum and the false positive rate is minimum.

To further evaluate the performance of the XGBoost model, the confusion matrix shown in figure 4.5 was used. The confusion matrix is a plot of four variables: true positive, false positive, true negative, and false negative. The true or false labels come from whether the event was taken from the  $B^0 \rightarrow K^{*0}\gamma\gamma$  (signal) decay or background in the testing dataset. The positive or negative labels originated from the XGBoost model where events deemed as signal corresponded to positive and events deemed as background corresponded to negative. The true positive and false negatives, therefore, identify the events correctly classified by the XGBoost model as

signal or background, respectively. However, the events incorrectly classified by the model fall into two categories: false positive and true negative. True negatives are signal events that the model incorrectly classified as background, whilst false positives are events that are background but incorrectly classified as signal.



**Figure 4.5:** This figure shows the confusion matrix for the model. The top left indicates the true negative events and the bottom right indicates the number of true positive events. Most of the events were classified correctly, and as there was more background than signal, therefore, the model correctly identified most of the background.

The bottom left corner in figure 4.5, was the false negatives and this represented original signal events which the model mistakenly classed as background. Whilst this was not ideal, it was manageable because the false positives (top right) were smaller, meaning the signal was being modeled properly and was not contaminated with background events. In the case that the false positives were higher, the signal would have many background events, which would have been a problem when conducting further analysis with the decay chain  $B^0 \rightarrow K^{*0}\gamma\gamma$ .

# Chapter 5

## Validation and fitting

Once the selection process was completed, it needed to be tested to ensure the selection process worked as expected. As mentioned in section 4, there are three sources of peaking background. One such example was the decay chain of  $B^0 \rightarrow K^{*0}\eta \rightarrow K^+\pi^-\gamma\gamma$ . In these regions, the production of  $\pi^0$ ,  $\eta$  and  $\eta'$  dominated via QCD effects. Therefore, if the axion had a mass near the peaking background, it would not have been possible to detect it using this analysis method. Therefore, these peaking backgrounds can be used to validate whether the selection process worked as intended, as the physics behind the peaking background decays is already known. The peaking background from the  $\eta$  meson was used as it had the most data available compared to the other peaking backgrounds.

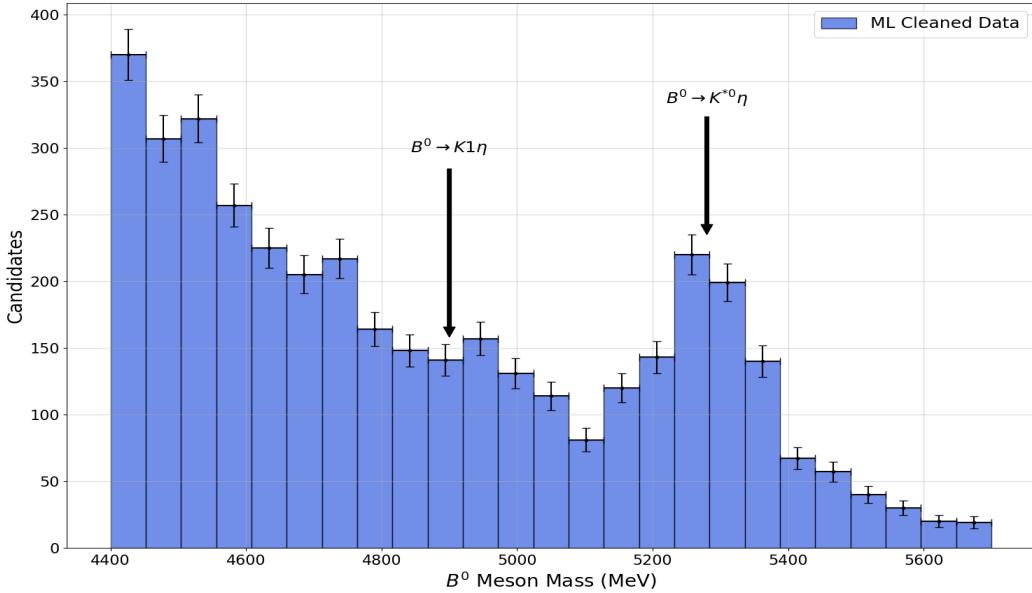
### 5.1 $B^0 \rightarrow K^{*0}\eta$ fit

The validation was done by measuring the number of  $\eta$  mesons observed in the LHCb data and comparing that value to the theoretical number of  $\eta$  mesons expected to be found from the branching fraction of this decay.

The  $\eta$  meson has a mass of 550 MeV, and it decays to photons. Therefore, the  $\eta$  region was isolated further by applying the constraint of the diphoton mass between 450 - 650 MeV to the LHCb data. A mass window of 100 MeV was included to account for measurement errors and detector effects.

The selection filtered LHCb data shown in figure 5.1 shows the  $B^0$  mass between the range of 4500 to 6500 MeV and in the diphoton mass range of 450 - 650 MeV. The large peak around 5280 MeV was caused via the  $B^0 \rightarrow K^{*0}\eta$  decay channel. The smaller peak around 4900 MeV was caused by the partly reconstructed background from the  $B^0 \rightarrow K1\eta$  decay channel. To calculate the yield of  $\eta$  mesons to validate the selection process, a fit was required to extract the  $\eta$  mesons from the peak around 5280 MeV. However, before performing a fit, the contribution from the K1 peak needed to be reduced. This was required because the contribution of the combinatorial and the K1 partly reconstructed backgrounds to the total background

in figure 5.1 was unknown. Once the contribution from the  $K1$  peak was reduced, the assumption could be made that the background was dominated by combinatorial background, leading to a more robust fit quality.

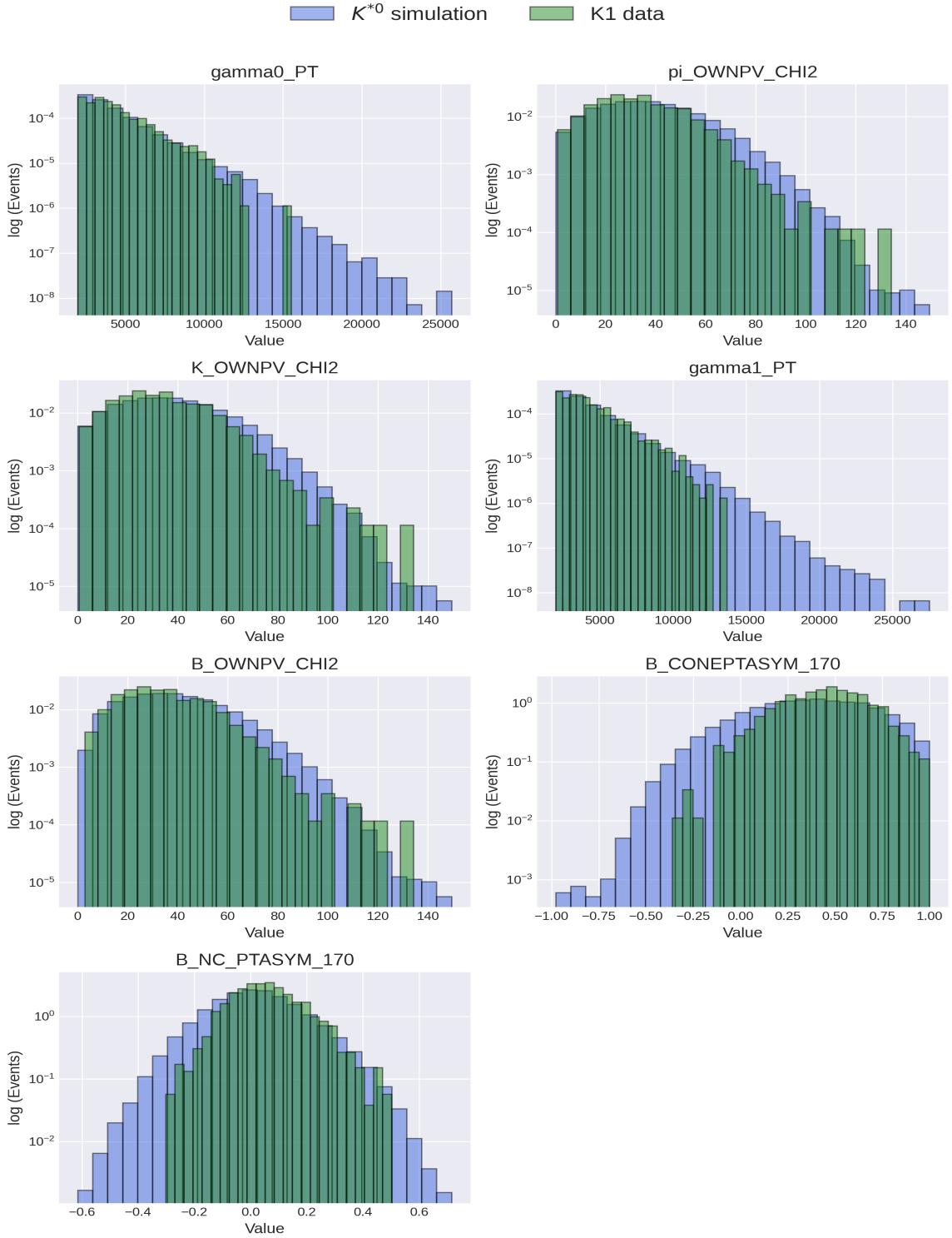


**Figure 5.1:** This figure shows the  $B^0$  mass distribution in the  $\eta$  region. This was after the selection cuts and the machine learning cut-off of 0.95 had been applied. There are three features of interest in this plot: firstly, the peak around 5280 MeV indicates that the selection process had been able to isolate the  $B^0$  decays of interest. Secondly, the combinatorial background can be seen roughly representing an exponential. Thirdly, there is another peak at approximately 4900 MeV, which represents the partly reconstructed background peak from the  $K1$  meson.

### 5.1.1 K1 XGBoost model

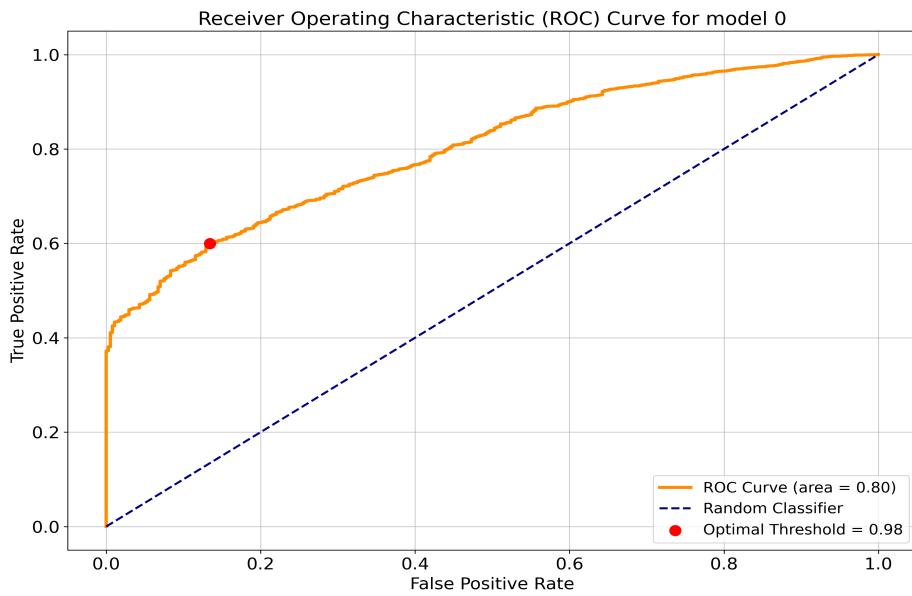
To remove the partly reconstructed background peak, another XGBoost model was trained that specifically distinguished between  $K^{*0}$  and  $K1$  meson. The  $K1$  meson events were taken from a range of  $4500 < m_{B^0} < 4900$  from figure 5.1. Care was taken to ensure that the  $K1$  meson events were not selected too close to the  $B^0$  signal peak to prevent any  $B^0$  events from being incorporated into the  $K1$  meson events. It was also ensured that  $K1$  meson events were not taken too far from the 4900 MeV peak, as outside of the peak region, there was no certainty if the events selected were from the partly reconstructed  $K1$  background or the combinatorial background.

The  $K1$  data distribution data was compared to the  $K^{*0}$  data to determine the features. There were seven features used in total. The features chosen were the same ones used in table 4.1, but some of those features were omitted because those features had no difference in the  $K1$  and  $K^{*0}$  data distributions. Figure 5.2 shows the feature comparison histograms.



**Figure 5.2:** This figure shows the comparison of the distributions between the  $K^{*0}$  events in the generated signal and the K1 events obtained from the LHCb data.

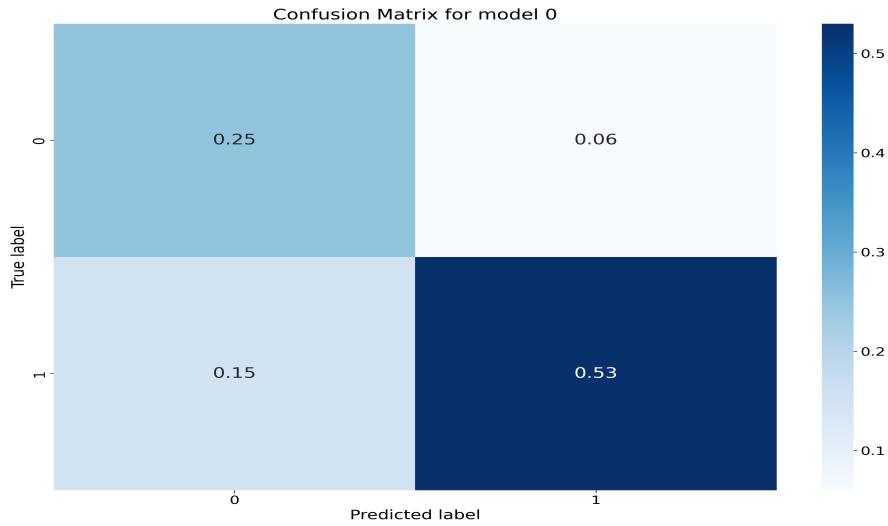
Again, the background K1 data from 4900 MeV peak and the generated signal events containing  $K^{*0}$  mesons were split into 5 folds corresponding to an 80% - 20% training-testing split and 5 different XGBoost models were trained with each one being trained and tested on different folds to avoid overfitting. This time 5 folds were used instead of 10 folds because in the 4500 - 4900 MeV region from figure 5.1, there were few K1 meson events. Splitting these events further into folds and then into training and testing data would have created insufficient events in certain folds for the XGBoost model to make predictions accurately. A similar procedure was followed to tune the hyperparameters using the same grid search method in figure 4.3 to train and obtain the optimal models. For all 5 models, the best learning rate was 0.1, a maximum depth of 4, and the number of estimators was 1000. Again, a fine search was used in this region where the learning rate was varied from 0.01 to 0.2 in increments of 0.01, the maximum depth was varied from 2 to 6 in increments of 1, and the number of estimators was varied from 800 to 1200 in increments of 100. The optimal hyperparameters for all 5 models were determined to be 0.02 for the learning rate, 3 for the maximum depth, and 1000 for the number of estimators. Finally, Yourden's J statistic was used to determine the optimal thresholds for all 5 models from figure 5.3



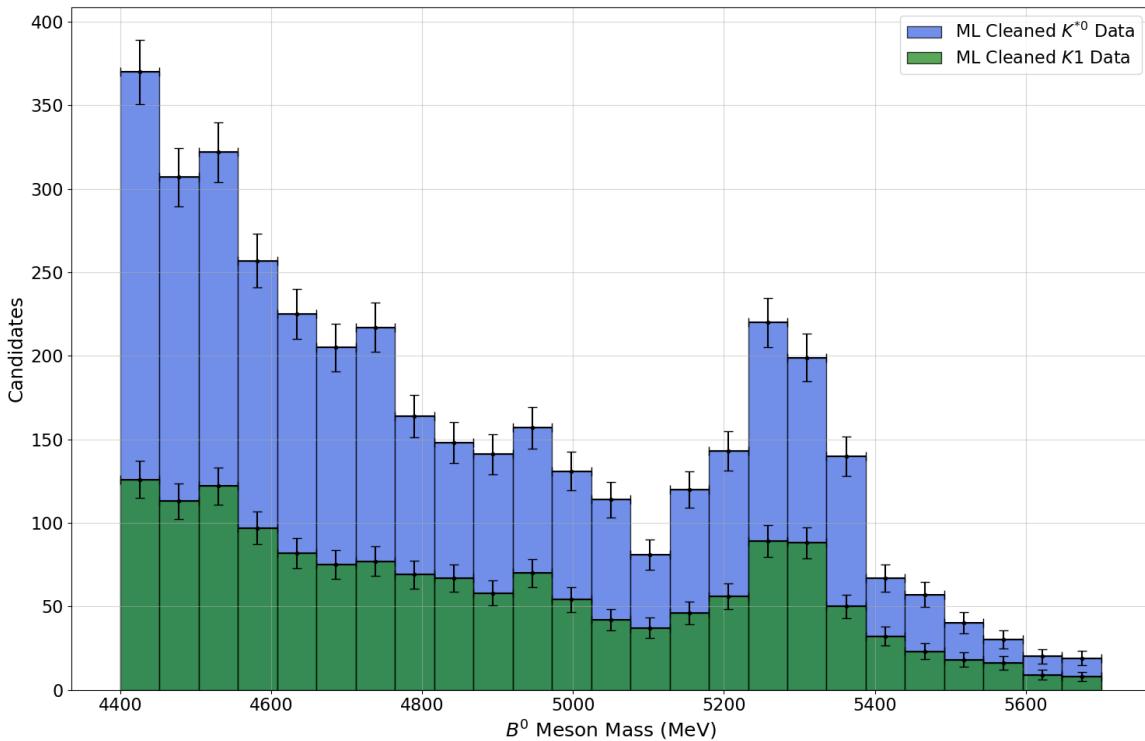
**Figure 5.3:** This figure shows the ROC curve for the XGBoost model. The optimal threshold in this case was 0.98. The AUC score was lower due to the similar nature of the  $K^{*0}$  and K1 mesons. The higher optimal threshold is required due to a lower separating power of the XGBoost model.

The confusion matrix shown in figure 5.4 was also obtained alongside the ROC curve to evaluate the performance of the models.

The final data is shown in figure 5.5 after applying the five XGBoost models for the partly reconstructed background and the ten XGBoost models for the  $B^0$  meson-background separation.



**Figure 5.4:** This figure shows the confusion matrix for the model. There were more errors in the models, as indicated by the higher proportion of false positives and true negatives. However, the rate of false positives was once again much lower, meaning the  $B^0 \rightarrow K^{*0}\eta$  peak at 5280 MeV in figure 5.1 would not get contaminated with background events.



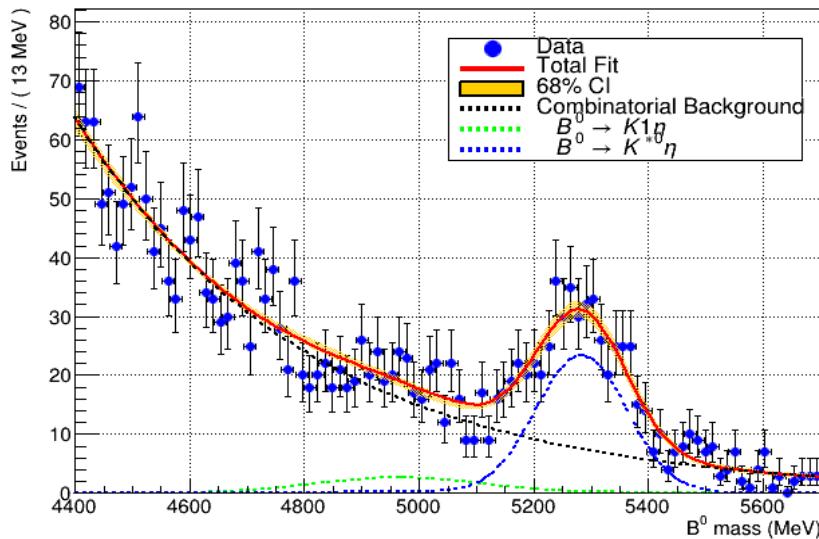
**Figure 5.5:** This figure shows the LHCb data with the  $K1$  decays indicated and the  $K^{*0}$  meson decays indicated. The partly reconstructed peak was almost removed after the second round of machine learning.

Although quite a few  $B^0 \rightarrow K^{*0}\gamma\gamma$  events were removed from the  $B^0$  peak, as shown in figure 5.5, this was still acceptable because it reduced the number of partly reconstructed backgrounds to a minimal level, meaning most of the background left was combinatorial. This allowed for better fits to the data later on.

### 5.1.2 Measuring $\eta$ yield

The peak at 5280 MeV in figure 5.5 represented the decay  $B^0 \rightarrow K^{*0}\eta$ . Therefore, counting the number of  $B^0 \rightarrow K^{*0}\eta$  events in the  $B^0$  peak would give the measured yield of  $\eta$  mesons, which was then compared to the theoretical yield of  $\eta$  mesons.

RooFit was used again to do an unbinned fit on the data in figure 5.5 to calculate the number of  $B^0 \rightarrow K^{*0}\eta$  events. The combinatorial background was modeled via an exponential function because it was expected to be Poissonian in nature. The  $\eta$  and  $K1$  peaks were both modeled using double crystal ball functions. Once again, the  $\eta$  peak shape and the  $K1$  shape were fixed from the generated signal events to preserve the Physics considerations and also reduce the number of floating parameters, which in turn allowed the RooFit program to converge faster with less computational power and avoid any instabilities in the fit due to too many floating parameters. The total fit is shown in figure 5.6.



**Figure 5.6:** This figure shows the total fit, which was a combination of all three components. The orange area shows the 68% confidence interval associated with the fit. The blue peak represents the  $\eta$  peak. The green peak represents the  $K1$  peak.

In figure 5.6, the  $K1$  peak was smaller than the  $\eta$  peak. This proved that the XGBoost separated enough of the partly reconstructed background to make a small contribution. Thus, the remaining background was assumed to be combinatorial, making the

fit quality better. The measured  $\eta$  yield from the peak in figure 5.6 was  $376 \pm 27$ .

The theoretical yield of  $\eta$  needed to be calculated in multiple steps. Firstly, the number of  $B^0$  mesons produced was computed from  $N_{B^0} = \sigma \mathcal{L}$  where  $\sigma$  was the cross-section of the  $B^0$  at LHCb and  $\mathcal{L}$  represents the Luminosity [28]. Using  $86.6 \pm 9 \mu\text{b}$  and  $2 \pm 0.3 \text{fb}^{-1}$  for the cross-section and the Luminosity respectively, the number of  $\eta$  was

$$N_\eta = \epsilon \text{BF}(B^0 \rightarrow \eta) \times N_{B^0}, \quad (5.1)$$

where  $\epsilon$  was some efficiency factor to account for the  $\eta$  events lost during the selection process. The branching fraction  $\text{BF}(B^0 \rightarrow \eta)$  is a combination of three other branching fractions in the following manner,

$$\text{BF}(B^0 \rightarrow \eta\gamma\gamma) = \text{BF}(B^0 \rightarrow K^{*0}) \times \text{BF}(K^{*0} \rightarrow K^+\pi^-) \times \text{BF}(\eta \rightarrow \gamma\gamma). \quad (5.2)$$

The branching fractions for  $\text{BF}(B^0 \rightarrow K^{*0})$  and  $\text{BF}(\eta \rightarrow \gamma\gamma)$  were obtained from PDG data which were  $1.59 \pm 0.10 \times 10^{-5}$  and  $0.3936 \pm 0.0018$  respectively. The branching fraction  $\text{BF}(K^{*0} \rightarrow K^+\pi^-)$  was  $2/3$  due to isospin.

### 5.1.3 Calculating the efficiency factor

The efficiency factor was required because the selection cuts, such as the PID cuts, the machine learning cuts, and the trigger cuts, removed some  $B^0$  signal events from the LHCb data. Therefore, the  $\epsilon$  factor accounts for the number of  $B^0$  events that were removed. The efficiency factor was estimated for each cut by applying the cuts to the generated signal events to estimate how many  $B^0$  events were removed from the LHCb data.

The efficiency factor is a combination of four different efficiencies,

$$\epsilon = \epsilon_{\text{strip+reco}} \times \epsilon_{\text{trigger|strip+reco}} \times \epsilon_{\text{PID|trigger}} \times \epsilon_{\text{ML|PID}}. \quad (5.3)$$

When the signal events were generated in figure 3.1, particles were produced in all directions. However, only the particles entering the LHCb detector from figure 3.3 were retained, the other events were discarded, and a loose cut was applied to the generated signal events. The  $\epsilon_{\text{strip+reco}}$  efficiency accounts for this effect. The other efficiencies represent part of the selection process that was applied.

The total efficiency  $\epsilon$  was calculated to be  $\epsilon = 0.00031 \pm 0.00003$ . Therefore, the final measured theoretical value of  $\eta$  yield was  $392 \pm 49$  (equation 5.1), which showed that there was good agreement between the theoretical and measured yield of  $\eta$ , which implied that the selection process was accurate.

Once the selection procedure was validated and checked to be accurate, the selection procedure was applied to the whole range of the LHCb data to determine the branching fraction  $\text{BF}(B^0 \rightarrow K^{*0}\gamma\gamma)$  which was then used to determine the axion-boson coupling constant  $g_{aW}$ .

# Chapter 6

## Axion-boson coupling

To calculate the axion-boson coupling constant, the branching fraction  $BF(B^0 \rightarrow K^{*0}\gamma\gamma)$  was calculated. For axion masses  $m_a$  less than the mass of the W boson, the axion decays to two photons and occurs with a branching fraction of almost 100%. Therefore, the branching fraction  $BF(B \rightarrow K^{*0}\gamma\gamma)$  was used to determine the decay width  $\Gamma(B^0 \rightarrow K^{*0}a)$ . From this decay width, the axion-boson coupling parameter was theoretically determined [29].

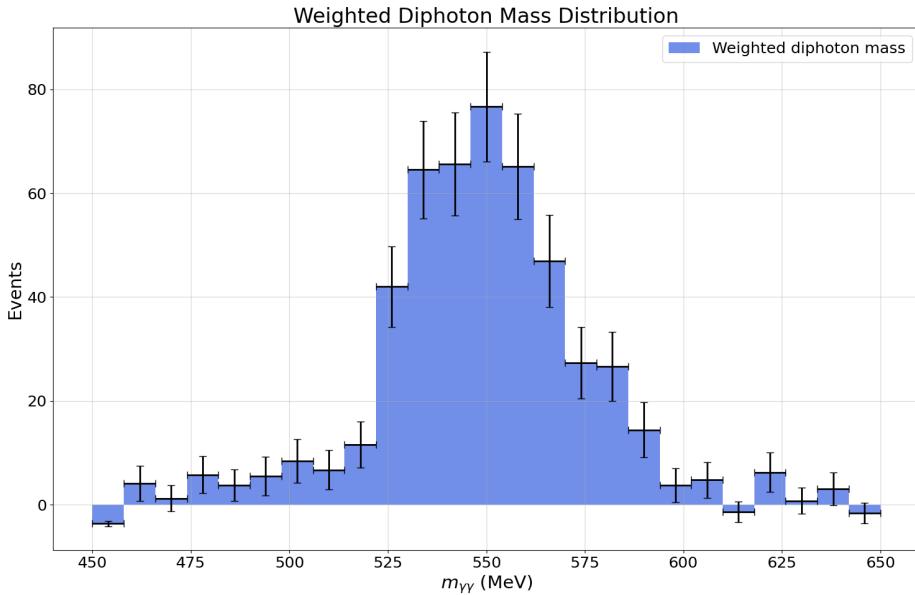
To determine the branching fraction  $BF(K^{*0} \rightarrow \gamma\gamma)$ , the branching fraction for the  $BF(B^0 \rightarrow K^{*0}\eta)$  can be used to remove the dependencies on  $\sigma$  and  $\mathcal{L}$  which have a large source of uncertainty. Therefore, replacing  $\sigma$  and  $\mathcal{L}$  and re-arranging yields

$$BF(B^0 \rightarrow K^{*0}\gamma\gamma) = \frac{\epsilon_\eta}{\epsilon_{\gamma\gamma}} \frac{N_{\gamma\gamma}}{N_\eta} BF(B^0 \rightarrow K^{*0}\eta). \quad (6.1)$$

The branching fraction of  $BF(B^0 \rightarrow K^{*0}\eta)$  was already documented from the PDG data. To calculate the number of  $\eta$  and  $\gamma\gamma$  events, sWeighting (refer to appendix 8.6) was used to separate the  $B^0 \rightarrow K^{*0}\eta$  distribution from the background distribution. This was a more accurate method of counting rather than counting the number of  $B^0 \rightarrow K^{*0}\eta$  events directly in figure 5.6 because sWeighting removes any potential overlap between  $B^0 \rightarrow K^{*0}\eta$  events and background distributions. Therefore, the number of  $\eta$  events  $N_\eta$  was re-calculated using sWeighting.

### 6.1 sWeighting for $N_\eta$ and $N_{\gamma\gamma}$

The number of  $\eta$  events  $N_\eta$  was re-calculated by summing the sWeights for all the events. The sWeights were generated from figure 5.6, and the sWeighted plot is shown in figure 6.1. This had the advantage of removing the background distribution from the  $B^0 \rightarrow K^{*0}\eta$  events, allowing for a more accurate calculation for  $\eta$  events. The measured number of  $\eta$  events was  $N_\eta = 356 \pm 36$ . This value was slightly lower than the previous value obtained from figure 5.6 as the background distribution overlap with the  $B^0 \rightarrow K^{*0}\eta$  events was removed.

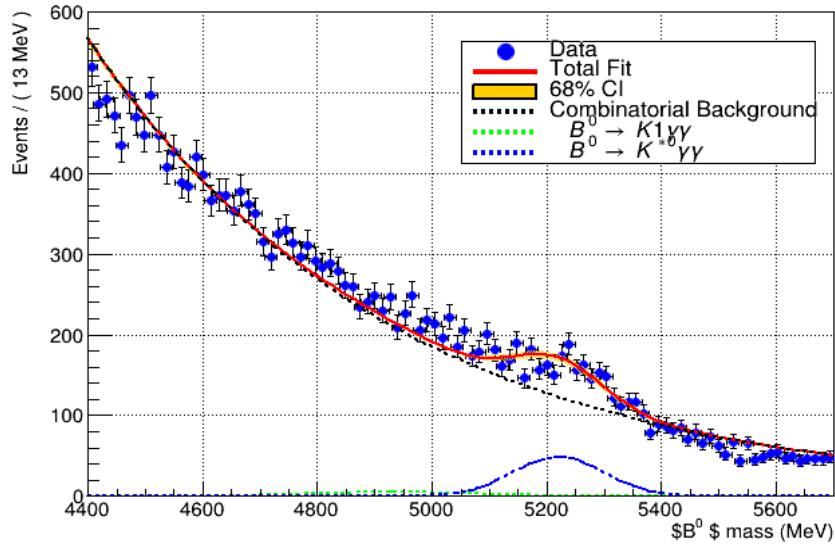


**Figure 6.1:** This figure shows the sWeighted plot for the diphoton events. The peak came from  $B^0 \rightarrow K^{*0}\eta$  events whilst the negatively weighted bins indicated strong backgrounds in those regions.

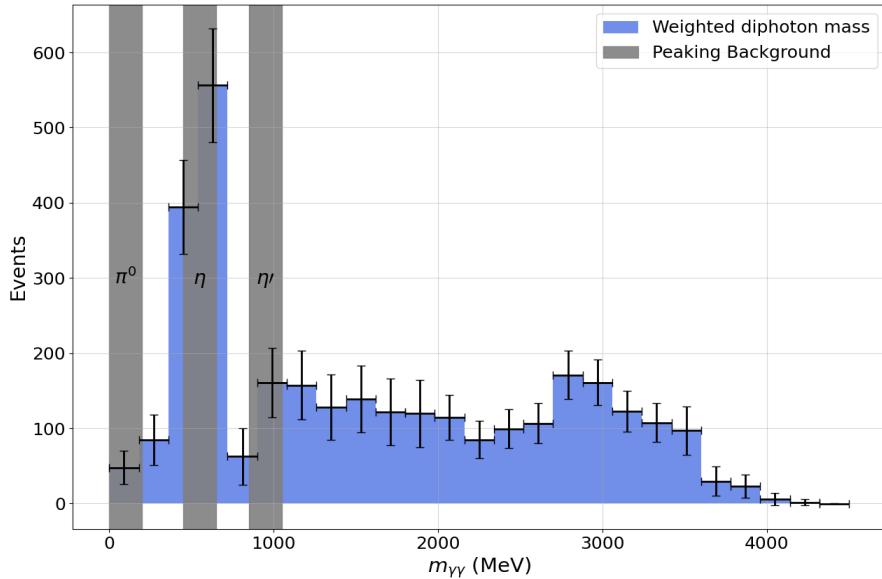
Figure 6.1 peaks at 550 MeV, which was expected due to the  $\eta$  mesons. The peak around 550 MeV provided a sanity check that the sWeighting procedure was working as intended and did not change the shape of the distribution. To calculate  $N_{\gamma\gamma}$  in equation 6.1, the fit in figure 5.6 had to be extended to the whole decay channel of  $B^0 \rightarrow K^{*0}\gamma\gamma$  so that sWeighting could be applied to calculate  $N_{\gamma\gamma}$ .

Once again, to calculate the  $\gamma\gamma$  events, another set of sWeights was generated from figure 6.2. The resulting sWeighted plot is shown in figure 6.3. In the grey regions of figure 6.3, the peaking backgrounds from the  $\pi^0$ ,  $\eta$  and  $\eta'$  mesons are shown. The  $N_{\gamma\gamma}$  value was determined to be  $751 \pm 83$ .

In addition, systematic errors for both the XGBoost models were calculated by comparing the sWeighted  $B^0 \rightarrow K^{*0}\gamma\gamma$  events in figure 6.3 to the original weighted  $B^0 \rightarrow K^{*0}\gamma\gamma$  events generated in figure 3.10. The GBReweighting (similar to XGBoost) model was applied to both distributions. It was determined that the statistical error was much larger than the systematic error:  $\delta\epsilon_{\text{stat}}^{\text{ML}} \gg \delta\epsilon_{\text{sys}}^{\text{ML}}$ , hence, the systematic error was not considered further.



**Figure 6.2:** This figure shows the same fit as figure 5.6 but to the whole data. The blue peak represents the  $B^0 \rightarrow K^{*0} \gamma\gamma$  events.



**Figure 6.3:** This figure shows the sWeighted diphoton mass plot for the whole fit from figure 6.2. The grey regions indicated the regions removed due to strong peaking backgrounds from  $\pi^0$ ,  $\eta$  and  $\eta'$  mesons.

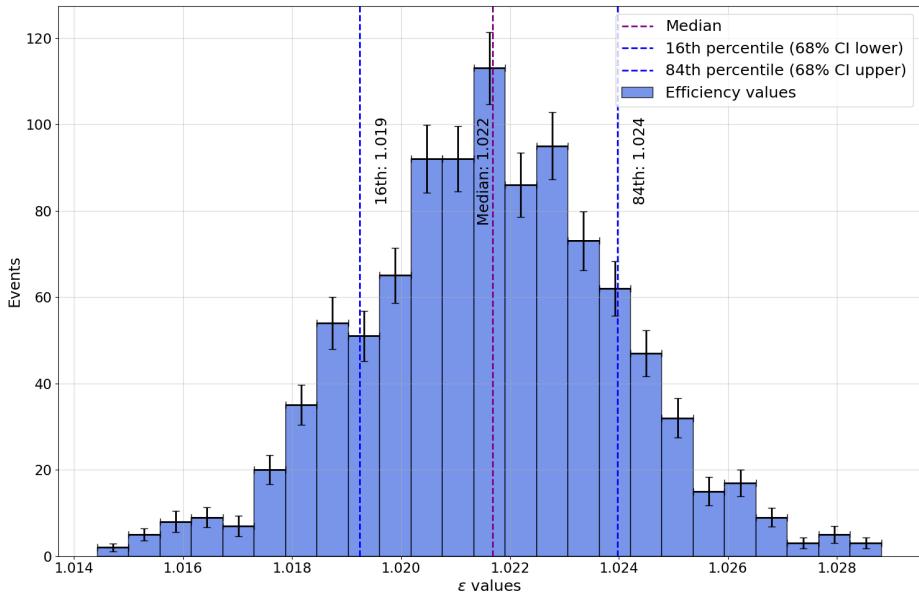
## 6.2 Efficiency ratio

Once the  $N_\eta$  and  $N_{\gamma\gamma}$  events were calculated, the efficiency  $\frac{\epsilon_\eta}{\epsilon_{\gamma\gamma}}$  (in equation 6.1) of the decay needed to be re-evaluated to account for the removal of  $B^0 \rightarrow K^{*0} \gamma\gamma$  events by the selection process. The ratio of efficiencies was calculated under the assumption that systematic uncertainties present in both  $\epsilon_\eta$  and  $\epsilon_{\gamma\gamma}$  would cancel, minimising the error on the efficiency. The ratio was evaluated for all parts of the selection process as shown by equation 6.2. The ratio of each section was evaluated in the following manner:

$$\frac{\epsilon_\eta}{\epsilon_{\gamma\gamma}} = \frac{\epsilon_{\text{strip+reco}}^\eta}{\epsilon_{\text{strip+reco}}^{\gamma\gamma}} \times \frac{\epsilon_{\text{trigger|strip+reco}}^\eta}{\epsilon_{\text{trigger|strip+reco}}^{\gamma\gamma}} \times \frac{\epsilon_{\text{PID|trigger}}^\eta}{\epsilon_{\text{PID|trigger}}^{\gamma\gamma}} \times \frac{\epsilon_{\text{ML|PID}}^\eta}{\epsilon_{\text{ML|PID}}^{\gamma\gamma}}. \quad (6.2)$$

Bootstrapping (refer to appendix 8.7) was used on each part of the efficiency ratios to calculate the value and error. The trigger efficiency ratio calculated via bootstrapping is shown in figure 6.4.

For most of the efficiency ratios, the main source of error was statistical. However, for the trigger efficiency, the error of the 4D bins was taken into consideration. This was done with a Gaussian function with the mean at the bin's efficiency value and the Gaussian width defined by the error of the efficiency in that bin according to equation 3.2.



**Figure 6.4:** This figure shows the bootstrapping of the ratio  $\epsilon_\eta/\epsilon_{\gamma\gamma}$  calculated for the trigger efficiency whereby a Gaussian was placed at the weights, and the widths were defined by the error of the weights. 1000 samples were generated, and the median and the 68% confidence intervals are indicated in the figure.

The ratio of trigger efficiencies was determined to be

$$\frac{\epsilon_{trigger|strip+reco}^\eta}{\epsilon_{trigger|strip+reco}^{\gamma\gamma}} = 1.022 \pm 0.003.$$

Similarly, calculating the other efficiency ratios and finally taking the product yielded

$$\frac{\epsilon_\eta}{\epsilon_{\gamma\gamma}} = 1.025 \pm 0.006$$

for the total efficiency. Therefore, the final branching fraction was calculated to be

$$BF(B^0 \rightarrow K^{*0}\gamma\gamma) = 8.52 \pm 1.16 \times 10^{-6}.$$

The branching fraction  $BF(\eta \rightarrow \gamma\gamma) = 1.59 \pm 0.10 \times 10^{-5}$  was greater than the calculated branching fraction for  $BF(B^0 \rightarrow K^{*0}\gamma\gamma)$  which made sense because the axion decay was expected to be suppressed.

### 6.3 Determining $g_{aW}$

Using the decay width  $\Gamma(B^0 \rightarrow K^{*0}a)$ , the axion-boson coupling parameter was theoretically determined from [29]:

$$\Gamma(B^0 \rightarrow K^{*0}a) = \frac{M_{B^0}^3}{64\pi} |g_{abs}|^2 A_0^2(M_a^2) \lambda_{K^{*0}a}^{3/2}. \quad (6.3)$$

Here, the decay width  $\Gamma(B^0 \rightarrow K^{*0}a)$  was calculated as the ratio of branching fractions as:  $\frac{BF(B \rightarrow K^{*0}\gamma\gamma)}{BF(B^0 \rightarrow \text{all decays})}$ . The lifetime of the  $B^0$  was used to determine  $BF(B^0 \rightarrow \text{all decays})$ .  $M_B$  is the mass of the  $B^0$  meson.  $\lambda_{K^{*0}a}$  is defined as:

$$\lambda_{K^{*0}a} = \left(1 - \frac{(M_a + M_{K^{*0}})^2}{M_{K^{*0}}}\right) \left(1 - \frac{(M_a - M_{B^0})^2}{M_{B^0}}\right).$$

In the above equation,  $M_a$  is the mass of the axion,  $M_{K^{*0}}$  is the mass of the  $K^{*0}$  meson.  $A_0(M_a^2)$  is the hadronic form factor [30] defined as:

$$A_0(M_a^2) = \frac{r_1}{1 - M_a^2/M_{B^0}^2} + \frac{r_2}{1 - M_a^2/M_{\text{fit}}^2}.$$

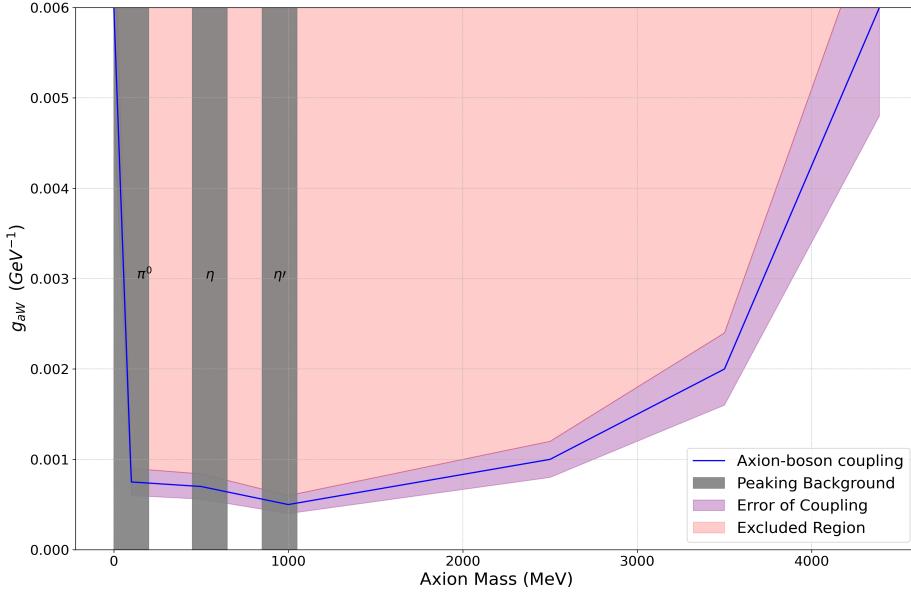
The constants  $r_1$ ,  $r_2$ ,  $M_{\text{fit}}$  were already calculated in [30]. The values used were  $r_1 = 1.363$ ,  $r_2 = -0.990$ ,  $M_{\text{fit}} = 36.78$ .  $g_{abs}$  is defined as:

$$g_{abs} = -\frac{3\sqrt{2}G_F M_W^2 g_{aW}}{16\pi^2} \sum_{\alpha \in c,t} V_{\alpha i} V_{\alpha j}^* f\left(\frac{M_\alpha^2}{M_W^2}\right). \quad (6.4)$$

In the equation above,  $G_F$  is the Fermi constant,  $M_W$  is the mass of the W-boson, and  $g_{aW}$  is the axion-boson coupling constant.  $V_{\alpha i}$ ,  $V_{\alpha j}^*$  are the CKM matrix elements where the  $\alpha$  sum is taken over the charm and top quarks.  $M_\alpha$  is the mass of the charm or the top quark. The function  $f(x)$  is defined as:

$$f(x) = \frac{x(1 + x(\ln(x) - 1))}{(1 - x)^2}.$$

Using the calculated value of the decay width  $\Gamma(B^0 \rightarrow K^{*0}a)$ , the final axion-boson coupling constant  $g_{aW}$  was plotted for a range of axion masses and is shown in figure 6.5. The figure shows the  $m_a - g_{aW}$  parameter space for axion masses from



**Figure 6.5:** This figure shows the coupling of the axion to the  $W$  boson ( $g_{aW}$ ) varying as a function of the axion mass. The grey region indicates the peaking backgrounds where the axion could not be searched for. The red region is the region excluded by this search.

0 - 4388 MeV. The results obtained were similar to the BaBar results. However, by using a larger dataset, the statistical error can be reduced, enabling a more accurate constraint on the  $m_a - g_{aW}$  parameter space.

# Chapter 7

## Conclusion

In summary, the  $B^0 \rightarrow K^{*0}\gamma\gamma$  decay channel was used to measure the axion-boson  $g_{aW}$  coupling constant. Initially, signal events were generated using simulation software, which was adjusted to account for the L0 electron and photon trigger efficiencies at the LHCb. Selection cuts were applied, and 2 XGBoost models were trained, one to isolate  $B^0 \rightarrow K^{*0}\gamma\gamma$  decays from the LHCb data and another one to remove the  $K1$  meson peaking background contribution. The  $B^0 \rightarrow K^{*0}\eta$  decay channel was used to verify the selection process was correct by measuring the theoretical yield of  $\eta$  compared to the measured yield from the LHCb data. Then, sWeighting was used to separate the  $B^0 \rightarrow K^{*0}\gamma\gamma$  and background distributions and to calculate the branching fraction  $BF(K^{*0} \rightarrow \gamma\gamma)$ . Afterwards, this branching fraction was used to calculate an axion-boson  $g_{aW}$  coupling constant. This provided a limit to the coupling for certain mass regions and provided exclusion zones where the axion could not exist. This provided a branching fraction for  $BF(B^0 \rightarrow K^{*0}\gamma\gamma)$  which had not been measured and also searched the  $m_a - g_{aW}$  parameter space in a region not previously explored.

Although, theoretically, the range of masses searched ranged from 0 – 4388 MeV for the axion corresponding to the rest mass of the  $B^0$  minus the  $K^{*0}$  meson, the lower mass ranges cannot be searched practically because at low masses of around  $\approx 1$  MeV, the L0 electron or photon triggers won't register a hit. Furthermore, the detector resolution will not be accurate enough to identify two photons as the angle between them will be small due to the lack of energy. In addition, if the decay length of the axion were longer than 10m (the length until the photon and electron detectors), then the axion would not be detected at all by the LHCb. To improve the analysis, the HLT trigger efficiencies could have been considered as well. Additionally, when the  $B^0$  meson decays at the LHCb, the detector sometimes detects two events for the same  $B^0$  decay. These are called multiple candidate events, and removing them would improve the analysis. Lastly, The  $K1$  meson XGBoost model could be improved to prevent the loss of  $B^0$  events from the LHCb data when the model is applied. The  $K1$  events for the model to train on were obtained from the data around 4500 – 4900 MeV. In this range, there would have been other backgrounds, such as combinatorial backgrounds, which meant the training was not done purely on  $K1$  mesons, which reduced the accuracy of the model.

# Chapter 8

## Appendix

### 8.1 Truth Cuts

The truth cuts for each particle were obtained from the PDG MC codes [31]. The table below shows the list of codes used:

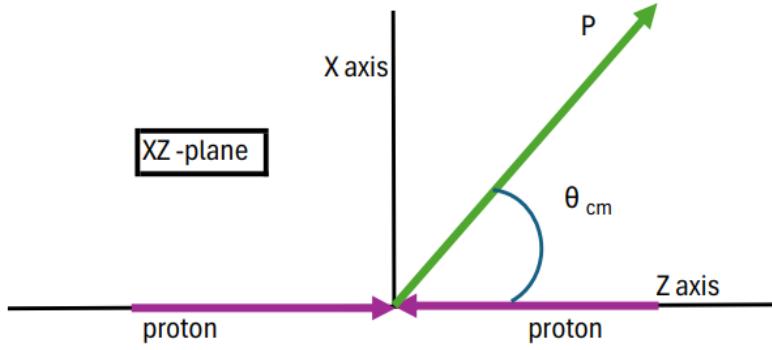
Particle	Truth Cut	Reason
$B^0$	511	Identify particle
$K^{*0}$	313	Identify particle
$K^+$	321	Identify particle
$\pi^-$	211	Identify particle
$\gamma$	22	Identify particle

**Table 8.1:** This table shows the full list of PDG MC codes applied as truth cuts to the signal events.

The last truth cut applied was the background cut called ‘B\_BKGCAT’ in the signal events. This was set to 30 to obtain the correct decay topology.

## 8.2 Transverse Momentum and Pseudorapidity

Due to the cylindrical symmetry at the LHCb, particle momenta can be split into two orthogonal components: transverse momenta and longitudinal momenta.



**Figure 8.1:** This figure shows the convention by which the momentum components are split. The origin is at the collision point indicated by the collision of protons. The z-axis is the beamline of the LHCb detector (the axis through which the proton-proton beams from the LHC pass through). The diagram is drawn in the center-of-mass frame [32].

The momentum  $\mathbf{P}$  can be split into a z-component:  $P_L$  and a momentum component in the x-y plane:  $P_T$ .

In Particle Physics, it is customary to use pseudorapidity, which is defined as:

$$\eta = -\ln \left( \tan \frac{\theta}{2} \right) \quad (8.1)$$

Using pseudorapidity ( $\eta$ ) rather than the polar angle ( $\theta$ ) is better because  $\eta$  is a Lorentz invariant in the massless limit.

This can be rewritten in terms of transverse momentum and total momentum as:

$$\eta = \operatorname{arctanh} \left( \frac{P_L}{|\mathbf{P}|} \right) \quad (8.2)$$

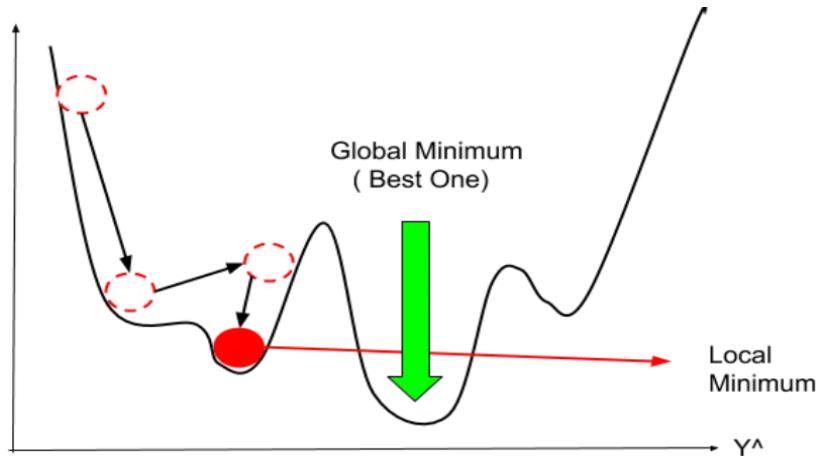
## 8.3 RooFit and Gradient Descent

RooFit is a software program used for fitting in high energy physics. RooFit has the advantage of being able to do unbinned fits, which means every data point's true value is used, increasing the statistical power of the fit. Furthermore, unbinned fits do not have artificial fluctuations introduced into the data due to sub-optimal binning schemes. However, unbinned fits do have the trade-off of being computationally expensive and usually lead to more complicated models. RooFit also allows multiple PDFs to be combined [33]. These multiple PDF combinations are especially useful for modeling signal and background. The PDFs are combined as follows:

$$S(x) = c_0 \cdot F_0(x) + c_1 \cdot F_1(x) + \cdots + (1 - \sum_{i=0}^{n-1} c_i) F_i(x).$$

Here, in the equation,  $c_i$  represents the coefficient of each individual PDF function  $F_i(x)$ . The sum of PDFs is also a PDF so the final function  $S(x)$  is also normalised.

RooFit calculates the optimal parameters for a given fit using maximum likelihood estimation (MLE). The MLE algorithm firstly defines a likelihood function  $L(w|y)$ , which is the probability of obtaining the parameters  $w$  given the data  $y$  [34]. The RooFit algorithm uses the negative log likelihood to iterate through different parameter values till the minimum is found.



**Figure 8.2:** This figure shows the concept behind the gradient descent algorithm used to determine the minimum of the negative log likelihood. [35].

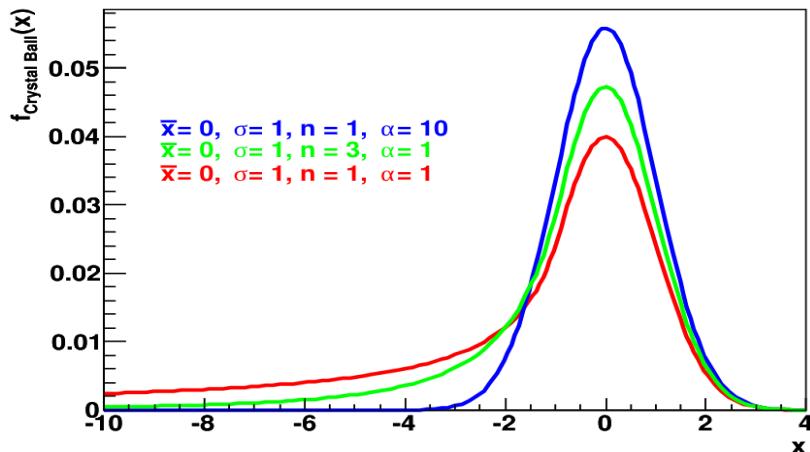
When using the gradient descent method, care must be taken to ensure that the algorithm converges to the global minimum and not some local minimum.

## 8.4 Crystal Ball Function

A crystal ball function is commonly used as a fitting function in High Energy Physics. It is essentially a Gaussian core with a power law tail that is useful for modeling distributions. The crystal ball function also has the advantage of being able to fit to lossy processes. The formula for a crystal ball function is,

$$f(x) = \begin{cases} e^{-\frac{1}{2}\left(\frac{x-\bar{x}}{\sigma}\right)^2} & \frac{x-\bar{x}}{\sigma} > \alpha \\ \left(\frac{n}{|\alpha|}\right)^n e^{-\frac{\alpha^2}{2}} \left(\frac{n}{|\alpha|} - |\alpha| - \frac{x-\bar{x}}{\sigma}\right)^{-n} & \frac{x-\bar{x}}{\sigma} < \alpha. \end{cases} \quad (8.3)$$

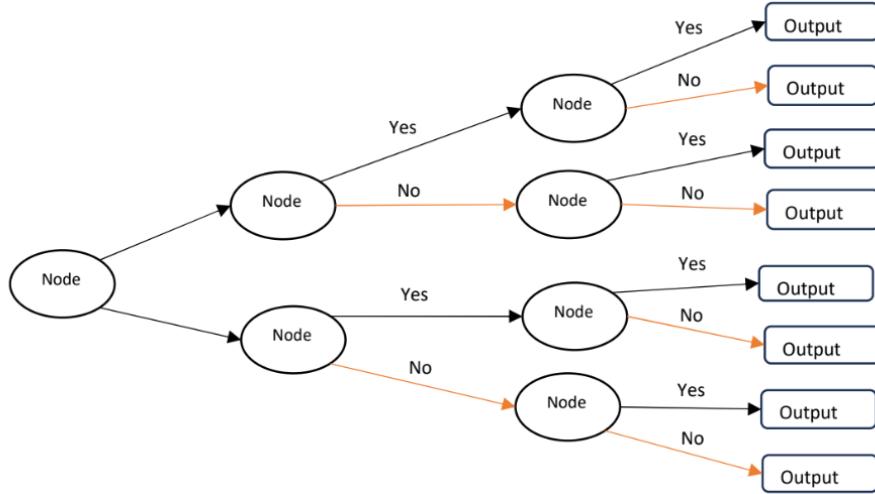
In the above equation,  $x$  is some continuous variable,  $\bar{x}$  is the mean of the Gaussian,  $\sigma$  is the standard deviation of the Gaussian,  $n$  is the index of the power law tail, and  $\alpha$  is the Z-score at which the Gaussian and the power law tail combine shown in figure 8.3. The crystal ball function is preferred over the Gaussian function because the crystal ball function can deal with asymmetric tails whilst having a better fit quality to data than the Gaussian, allowing for a better signal-to-background ratio. However, the crystal ball function has the drawback that due to the increased number of parameters, the computational time is increased, and the greater number of parameters compared to a Gaussian leads to convergence errors [36].



**Figure 8.3:** This figure shows an example of the crystal ball function with the parameters discussed above indicated in the figure. The power law tail highlights the ability of the crystal ball function to model asymmetric distributions whilst adapting to distributions that may not be ideally Gaussian.

## 8.5 Decision Trees

Decision Trees are a type of machine-learning algorithm that uses a tree-like structure to mimic a human decision process.



**Figure 8.4:** This figure shows the example of a decision tree. The circles are the nodes where decisions are made. At the end, the square boxes display the probability associated with that path. The depth refers to the number of nodes from the start till the end where the probabilities are displayed [37].

### 8.5.1 XGBoost Algorithm

XGBoost (eXtreme Gradient Boost) uses multiple decision trees (figure 8.4), whereby each new tree attempts to reduce the error of the previous one. As a result, a more accurate prediction is obtained from the model. XGBoost also has the advantage of being very computationally efficient, and it can work on smaller datasets while maintaining a high accuracy level in its predictions. The trees are combined in the following manner,

$$T(x) = \sum_{i=0}^{N_{trees}} \alpha T_i(x), \quad (8.4)$$

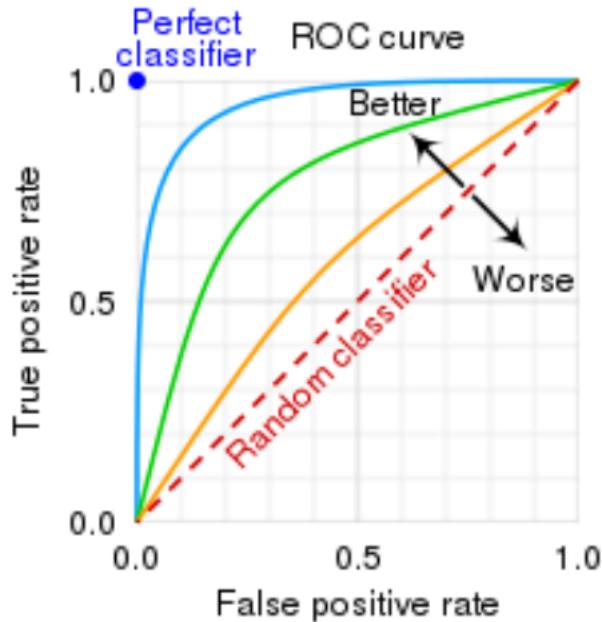
where the parameter  $\alpha$  is referred to as the learning rate and  $N_{trees}$  is the number of trees used or more commonly referred to as the number of estimators [38].

### 8.5.2 Hyperparameters

Each machine learning model comes with hyperparameters, which are parameters that can be fine-tuned to improve the model's performance. XGBoost has relatively fewer hyperparameters than other machine learning models, making it simple and effective. XGBoost has three hyperparameters that can be tuned: depth, learning rate, and the number of estimators [39].

### 8.5.3 ROC Curves and AUC score

The two most important metrics for evaluating a model's performance in terms of how well it can classify signal and background are the Receiver Operating Characteristics (ROC) curve and the Area-Under-Curve (AUC) score. An example of an arbitrary ROC curve is shown below:

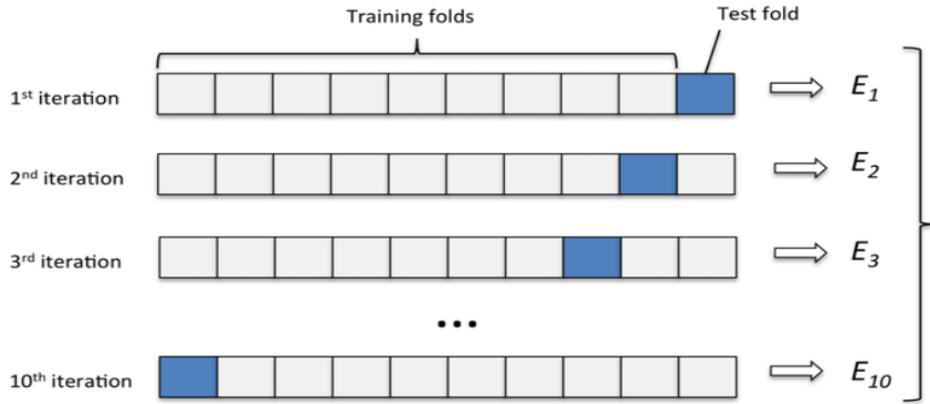


**Figure 8.5:** This figure shows an example of a ROC curve. The y-axis is the true positive rate, which represents the number of signal events correctly classified by the model. On the x-axis is the false positive rate, which represents the number of events that were background but have been incorrectly classed as signal by the machine learning model. The dotted line represents the ROC curve's appearance if an event was randomly classed as signal or background. Therefore, the ideal curve would be higher than the random curve [40].

The best models will aim to maximise the number of true positive events whilst minimising the number of false positive events. The AUC score is the area under the ROC curve and summarises the information on the ROC curve. The AUC score ranges from 0 to 1. A score of 0.5 represents the random classifier. An AUC score of 1 would represent the perfect classifier [41].

### 8.5.4 K-Fold Cross Validation

K-Fold cross-validation is a commonly used technique in machine learning where the data is split evenly into K different datasets (or folds). One set of folds is reserved to evaluate the performance of the model whilst the other K - 1 folds are used to train the model.



**Figure 8.6:** This figure shows the logic behind K-folding. Once the data has been split evenly, a total of  $K$  different machine learning models are trained, all of which have different training and testing folds [42].

The advantage of using K-folds is that overfitting is mitigated via this process. Overfitting refers to when a particular machine learning algorithm learns the training data well but fails to perform to a good standard on unseen data (test data) [43].

### 8.5.5 Youden's J statistic

To determine the optimal cut-off point for the model probabilities, such as the maximum number of true positive events passing whilst minimising the number of false positives. The optimal threshold point is usually the elbow point (turning point) on a ROC curve and serves as the point for determining whether an event will be classed as signal or background. The optimal threshold can be calculated in various ways, but the method used in this analysis was Yourden's J statistic. This is defined as:

$$J = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}} + \frac{\text{true negatives}}{\text{true negatives} + \text{false positives}}. \quad (8.5)$$

Yourden's J statistic can be evaluated for all the points on a ROC curve, and the point which maximises the statistic is the optimal threshold [44].

## 8.6 sWeights

sWeighting is a special type of weighting procedure that separates signals and backgrounds from a distribution. sWeighting assigns a weight to each event in a dataset, which represents the probability of an event belonging to signal or background. A dataset can be represented as [45]:

$$f(x) = N_0 d_0(x) + N_1 d_1(x). \quad (8.6)$$

Here,  $x$  represents the variable being fit to.  $N_0$  and  $N_1$  represent the yields, while  $d_0(x)$  and  $d_1(x)$  represent the distribution. sWeights are obtained from the fit parameters calculated in a fitting algorithm, and then the covariance matrix is used to assign the sWeights.

## 8.7 Bootstrapping

Bootstrapping is a method used to produce confidence intervals from an unknown distribution of data. There are many methods for bootstrapping, but the one used in this analysis was percentile bootstrapping [46]. In this method,  $n$  samples are generated for a parameter  $\theta$  of interest from a particular distribution. The confidence intervals are constructed according to the percentile  $100 \times \frac{\alpha}{2}$  and  $100 \times (1 - \frac{\alpha}{2})$  where  $\alpha$  is the desired confidence interval. Therefore, the confidence interval becomes:

$$[\hat{\theta}_{lower}, \hat{\theta}_{upper}] = [\hat{\theta}_j^*, \hat{\theta}_k^*] \quad (8.7)$$

Here, the  $\hat{\theta}_j^*$  and  $\hat{\theta}_k^*$  variables represent the lower and upper bounds calculated from percentiles.

# Bibliography

- [1] B. Odom, D. Hanneke, B. d'Urso, and G. Gabrielse, "New measurement of the electron magnetic moment using a one-electron quantum cyclotron," *Physical review letters*, vol. 97, no. 3, p. 030801, 2006. pages 1
- [2] S. D. Bass, A. De Roeck, and M. Kado, "The higgs boson implications and prospects for future discoveries," *Nature Reviews Physics*, vol. 3, no. 9, pp. 608–624, 2021. pages 1
- [3] V. A. Kostelecký, "Gravity, lorentz violation, and the standard model," *Physical Review D*, vol. 69, no. 10, p. 105009, 2004. pages 1
- [4] F. Stecker, "The matter-antimatter asymmetry of the universe," *arXiv preprint hep-ph/0207323*, 2002. pages 1
- [5] V. Rubin, "Dark matter in the universe," *Scientific American*, vol. 1, pp. 106–110, 1998. pages 1
- [6] J. E. Kim and G. Carosi, "Axions and the strong c p problem," *Reviews of Modern Physics*, vol. 82, no. 1, p. 557, 2010. pages 1
- [7] M. Chabab, "Theta term in qcd sum rules at finite temperature and the neutron electric dipole moment," in *AIP Conference Proceedings*, vol. 660, pp. 170–176, American Institute of Physics, 2003. pages 1
- [8] R. D. Peccei, T. T. Wu, and T. Yanagida, "A viable axion model," *Physics Letters B*, vol. 172, no. 3-4, pp. 435–440, 1986. pages 1
- [9] F. Björkeroth, E. J. Chun, and S. F. King, "Flavourful axion phenomenology," *Journal of High Energy Physics*, vol. 2018, no. 8, pp. 1–33, 2018. pages 1
- [10] V. Anastassopoulos, S. Aune, K. Barth, A. Belov, G. Cantatore, J. Carmona, J. Castel, S. Cetin, F. Christensen, J. Collar, *et al.*, "New cast limit on the axion-photon interaction," *arXiv preprint arXiv:1705.02290*, 2017. pages 1
- [11] Y. Krita, T. Hasada, M. Hashida, Y. Hirahara, K. Homma, S. Inoue, F. Ishibashi, Y. Nakamiya, L. Neagu, A. Nobuhiro, *et al.*, "Search for sub-ev axion-like particles in a stimulated resonant photon-photon collider with two laser beams based on a novel method to discriminate pressure-independent components," *Journal of High Energy Physics*, vol. 2022, no. 10, pp. 1–20, 2022. pages 1

- [12] L. D. Duffy and K. Van Bibber, “Axions as dark matter particles,” *New Journal of Physics*, vol. 11, no. 10, p. 105008, 2009. pages 1
- [13] G. G. Raffelt and D. S. Dearborn, “Bounds on hadronic axions from stellar evolution,” *Physical Review D*, vol. 36, no. 8, p. 2211, 1987. pages 1
- [14] M. Gavela, A. Le Yaouanc, L. Oliver, O. Pene, J.-C. Raynal, M. Jarfi, and O. Lazrak, “How to isolate penguin diagrams through exclusive b meson decays,” *Physics Letters B*, vol. 154, no. 5-6, pp. 425–430, 1985. pages 2
- [15] C. Baker, D. Doyle, P. Geltenbort, K. Green, M. Van der Grinten, P. Harris, P. Iaydjiev, S. Ivanov, D. May, J. Pendlebury, *et al.*, “Improved experimental limit on the electric dipole moment of the neutron,” *Physical review letters*, vol. 97, no. 13, p. 131801, 2006. pages 3
- [16] R. D. Peccei and H. R. Quinn, “Constraints imposed by cp conservation in the presence of pseudoparticles,” *Physical Review D*, vol. 16, no. 6, p. 1791, 1977. pages 3
- [17] J. Alvey and M. Escudero, “The axion quality problem: global symmetry breaking and wormholes,” *Journal of High Energy Physics*, vol. 2021, no. 1, pp. 1–28, 2021. pages 4
- [18] K. Hamaguchi, Y. Kanazawa, and N. Nagata, “Axion quality problem alleviated by nonminimal coupling to gravity,” *Physical Review D*, vol. 105, no. 7, p. 076008, 2022. pages 4
- [19] J. Lees, V. Poireau, V. Tisserand, E. Grauges, A. Palano, G. Eigen, D. Brown, Y. G. Kolomensky, M. Fritsch, H. Koch, *et al.*, “Search for an axionlike particle in b meson decays,” *Physical review letters*, vol. 128, no. 13, p. 131802, 2022. pages 4
- [20] S. Chakraborty, M. Kraus, V. Loladze, T. Okui, and K. Tobioka, “Heavy qcd axion in  $b \rightarrow s$  transition: Enhanced limits and projections,” *Physical Review D*, vol. 104, no. 5, p. 055036, 2021. pages 4
- [21] S. Middleton, “Recent results from babar: Dark matter, axion-like particles and heavy neutral leptons, a contribution to the 2023 electroweak session of the 57th rencontres de moriond,” *arXiv preprint arXiv:2305.00808*, 2023. pages 5
- [22] T. Sjöstrand, “Monte carlo tools,” *arXiv preprint arXiv:0911.5286*, 2009. pages 7
- [23] L. Evans and P. Bryant, “Lhc machine,” *Journal of instrumentation*, vol. 3, no. 08, p. S08001, 2008. pages 7
- [24] A. A. Alves Jr, L. Andrade Filho, A. Barbosa, I. Bediaga, G. Cernicchiaro, G. Guerrer, H. Lima Jr, A. Machado, J. Magnin, F. Marujo, *et al.*, “The lhcb detector at the lhc,” *Journal of instrumentation*, vol. 3, no. 08, p. S08005, 2008. pages 8

- [25] I. Belyaev, G. Carboni, N. Harnew, C. Matteuzzi, and F. Teubert, “The history of lhcb,” *The European Physical Journal H*, vol. 46, no. 1, p. 3, 2021. pages 8
- [26] T. Head, “The lhcb trigger system,” *Journal of Instrumentation*, vol. 9, no. 09, p. C09015, 2014. pages 9
- [27] M. Paterno, “Calculating efficiencies and their uncertainties,” 2004. pages 11
- [28] F. Abe, M. Albrow, S. Amendolia, D. Amidei, J. Antos, C. Anway-Wiese, G. Apollinari, H. Areti, M. Atac, P. Auchincloss, *et al.*, “Measurement of the b meson differential cross section  $d\sigma/dp_t$  in  $p\bar{p}$  collisions at  $s = 1.8$  tev,” *Physical review letters*, vol. 75, no. 8, p. 1451, 1995. pages 30
- [29] E. Izaguirre, T. Lin, and B. Shuve, “A new flavor of searches for axion-like particles,” *arXiv preprint arXiv:1611.09355*, 2016. pages 31, 35
- [30] P. Ball and R. Zwicky, “B d,  $s \rightarrow \rho, \omega, k^*, \phi$  decay form factors from light-cone sum rules reexamined,” *Physical Review D*, vol. 71, no. 1, p. 014029, 2005. pages 35
- [31] L. Basso, A. Belyaev, D. Chowdhury, M. Hirsch, S. Khalil, S. Moretti, B. O’Leary, W. Porod, and F. Staub, “Proposal for generalised supersymmetry les houches accord for see-saw models and pdg numbering scheme,” *Computer Physics Communications*, vol. 184, no. 3, pp. 698–719, 2013. pages 38
- [32] S. Chatterjee and P. Božek, “Pseudorapidity profile of transverse momentum fluctuations in heavy ion collisions,” *Physical Review C*, vol. 96, no. 1, p. 014906, 2017. pages 39
- [33] W. Verkerke, “Roofit,” in *EPJ Web of Conferences*, vol. 4, p. 02005, EDP Sciences, 2010. pages 40
- [34] J.-X. Pan, K.-T. Fang, J.-X. Pan, and K.-T. Fang, “Maximum likelihood estimation,” *Growth curve models and statistical diagnostics*, pp. 77–158, 2002. pages 40
- [35] S. Hageboeck, “What the new roofit can do for your analysis,” *arXiv preprint arXiv:2012.02746*, 2020. pages 40
- [36] S. Das, “A simple alternative to the crystal ball function,” *arXiv preprint arXiv:1603.08591*, 2016. pages 41
- [37] F. P. García Márquez, I. Segovia Ramírez, and A. Pliego Marugán, “Decision making using logical decision tree and binary decision diagrams: a real case study of wind turbine manufacturing,” *Energies*, vol. 12, no. 9, p. 1753, 2019. pages 42

- [38] B. P. Roe, H.-J. Yang, J. Zhu, Y. Liu, I. Stancu, and G. McGregor, “Boosted decision trees as an alternative to artificial neural networks for particle identification,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 543, no. 2-3, pp. 577–584, 2005. pages 42
- [39] P. Probst, A.-L. Boulesteix, and B. Bischl, “Tunability: Importance of hyperparameters of machine learning algorithms,” *Journal of Machine Learning Research*, vol. 20, no. 53, pp. 1–32, 2019. pages 42
- [40] J. Fan, S. Upadhye, and A. Worster, “Understanding receiver operating characteristic (roc) curves,” *Canadian Journal of Emergency Medicine*, vol. 8, no. 1, pp. 19–20, 2006. pages 43
- [41] S. Rosset, “Model selection via the auc,” in *Proceedings of the twenty-first international conference on Machine learning*, p. 89, 2004. pages 43
- [42] Y. Jung, “Multiple predicting k-fold cross-validation for model selection,” *Journal of Nonparametric Statistics*, vol. 30, no. 1, pp. 197–215, 2018. pages 44
- [43] D. Anguita, L. Ghelardoni, A. Ghio, L. Oneto, S. Ridella, *et al.*, “The’k’in k-fold cross validation.,” in *ESANN*, vol. 102, pp. 441–446, 2012. pages 44
- [44] N. J. Perkins and E. F. Schisterman, “The youden index and the optimal cut-point corrected for measurement error,” *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, vol. 47, no. 4, pp. 428–441, 2005. pages 44
- [45] M. Pivk and F. Le Diberder, “splot: a statistical tool to unfold data distributions. nucl. istrum. methods a 555, 356–369 (2005),” *arXiv preprint physics/0402083*. pages 45
- [46] T. J. Diciccio and J. P. Romano, “A review of bootstrap confidence intervals,” *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 50, no. 3, pp. 338–354, 1988. pages 46