STUDY OF LEPTON FLAVOUR UNIVERSALITY VIOLATIONS IN THE $B^0 \to K^{*0}ee$ DECAY

Using machine learning algorithms to differentiate between signal and background events

July 17, 2022

Author:Mathias Ajami

University of Copenhagen
Niels Bohr Institute
Supervisor: Troels Petersen

Abstract

This thesis is part of a larger project which aims to analyse data from the ATLAS experiment at the Large Hadron Collider, in order to calculate $R(K^{*0})$ for the $B^0 \to K^{*0}ll$ decay, in order to find violations of the Lepton Universality predicted by the Standard Model. The work in the thesis, which deals with the case were ll is an electron-positron pair, focuses on separating signal events (the ones corresponding to the decay in question) from background events (all the other interactions happening in the ATLAS detector), obtained from both Monte Carlo (MC) simulations and experimental results from the ATLAS detector, in order to help obtain the proper values for the calculation of $R(K^{*0})$. This was done by using the data to train a Gradient Boosted Decision Tree, a type of machine learning algorithm, which was then applied to the Monte Carlo and ATLAS data in order to test its ability to accept signal events and reject background ones. The resulting selection was able to effectively separate signal from background, albeit with some fluctuations in the efficiency of background rejection.

The B^0 and ee mass distributions for the signal events were also fitted to gaussian and crystal-ball curves in order to better understand their expected shape, which can also help separate them from the background. We were able to successfully fit them to a curve formed by the sum of a gaussian and a crystal-ball function. Finally, the different sources of systematic uncertainty for the results were considered.

Hopefully, the work done here will complement work already done in the LHCb experiment [6][54][55][56], in order to provide solid evidence for violations of Lepton Universality, and the presence of Physics beyond the Standard Model.

Contents

Acknowledgements

I would like to thank my supervisor, Troels Petersen, for his patience working with me for over a year. I also owe my thanks to everyone in the RKStarRun2 ATLAS group, but especially Tomáš Jakoubek and Ann-Kathrin Perrevoort.

I finally owe a big thanks to everyone who helped and supported me in all kinds of ways while I was working on this thesis, including my family, my close friends, and also Vivi Nawroz and Gunjan Nagda from Academic Support.

Introduction

The Standard Model (SM) is one of the most important theories in Physics, as it explains the fundamental particles that make up the matter and carry (most of) the forces present in our Universe. The existence of all the funndamental particles predicted by it has been experimentally confirmed since 2012. That being said, the theory fails to explain some important aspects of our Universe, such as the nature of dark matter, or how gravity works (gravity is the only fundamental force not included in the Standard Model). Physicists have therefore tried to study the so-called "New Physics" (NP), which is Physics that goes beyond the Standard Model, tackling and explaining effects and findings that cannot be explained by the SM. One of the most important ways of experimentally studying Particle Physics theories, including the SM and NP, is through particle colliders, of which the Large Hadron Collider (LHC) is the largest. Some of the main detectors/experiments at the LHC include ATLAS ("A Toroidal LHC ApparatuS"), CMS ("Compact Muon Solenoid") and the LHCb ("Large Hadron Collider beauty") experiment.

This thesis is part of a larger project within ATLAS which aims to find evidence of NP in the $B^0 \to K^{*0}ll$ decay, where "l" refers to a lepton (for example, and electron or a muon). A property of the SM called Lepton Universality (LU) predicts that the decay rates for $B^0 \to K^{*0}ll$ should be the same when l is a muon or an electron, and the ratio between them, called $R_{K^{*0}}$, is predicted by the SM to be 1, and deviations from this value can indicate the existence of NP processes or particles in the decay. LHCb has already obtained values of $R_{K^{*0}}$ quite smaller than 1 [54], and if an ATLAS team manages to itself obtain deviations in the same direction, the results will be significant, since these are two independent experiments obtaining similar results.

The idea is then to analyse measurements from the ATLAS detector to study the relevant decays. But since most of the interactions happening in ATLAS are not relevant to us, we must find ways of separating these background interactions from the signal ones. We are thus developing machine learning (ML) algorithms that can perform this separation, and finding other ways of analysing and understanding the different interactions. The work in this thesis involves using a ML algorithm called a Gradient Boosted Decision Tree (GBDT), and also analysing and fitting the curves of mass distributions of signal interactions, in order to be able to better identify and separate them. The focus will be on the case where the leptons in the decay are electrons (the muon case is being studied by other people within the larger project I'm a part of), and we will use both real data from ATLAS and simulations of ATLAS decays and their measurements, in order to develop the relevant algorithms.

The first 3 chapters of the thesis introduce concepts and information relevant to the results and how they were obtained. The first chapter explains the SM, its particles and limitations , the need for NP, and the basics of B-Physics (field of Physics that studies B hadrons, which contain at least one bottom quark), which includes the $B^0 \to K^{*0}ll$ decay.

Chapter 2 presents the LHC and the ATLAS detector, including all its components, and also its trigger systems and the problems involved with the identification, isolation and reconstruction of electrons in the detector. Chapter 3 describes ML and GBDT, and explains the process of developing an ML algorithm.

Chapter 4 contains the actual results of our analysis. It showcases the data (both real and simulated) chosen for the algorithm, the result of the pre-selection imposed on this data, the training and results of the GBDT, and the fitting of the distribution of the B^{*0} and di-electron pair masses. Finally, a study of possible sources of systematic uncertainty in our results is carried out, which does not include measurements of our own for these uncertainties, but references some measurementss from other experiments, alongside estimates on how ATLAS results might compare to these.

Chapter 1

Theory

1.1 The Standard Model of Particle Physics

The Standard Model (SM) is arguably the most important model in Particle Physics, with numerous success stories since its introduction in the early 1970s [19]. These include the prediction of the existence of the W and Z particles (experimentally observed in 1983 at CERN [31]), the bottom and top quarks (observed at Fermilab in 1977 [23] and 1995 [51] [53], respectively), the tau neutrino (observed at Fermilab in the year 2000 [24]) and the gluons (observed in 1978-1979 at DESY [43] [45]). Finally, the Higgs boson was found in 2012 at CERN by both the ATLAS [47] and CMS [52] experiments, and with this discovery, all the building blocks of the SM have been experimentally found [37].

a Elementary particles: fermions and bosons

The basic components of the Standard Model are found in the image below:

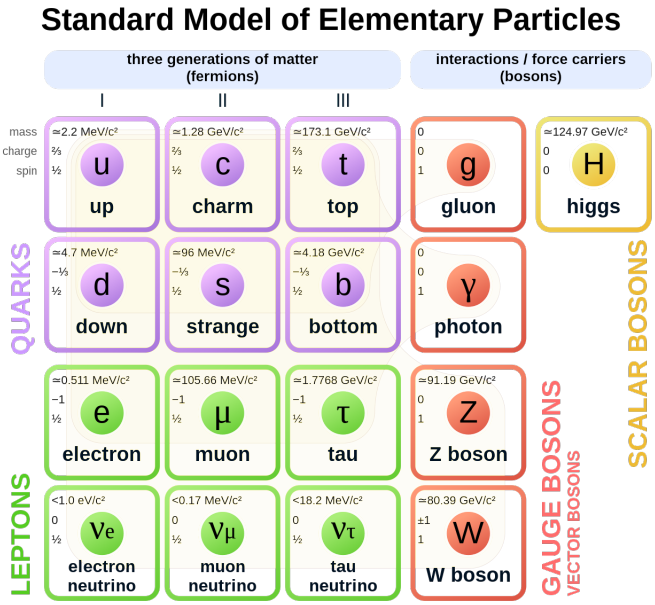


Figure 1.1: Elementary Particles of the Standard Model [36]

As we see in figure 1.1, the elementary particles can be divided in two major types: fermions and bosons.

. Fermions are the building blocks of matter. They have half-integer spin (specifically 1/2 for all the ones we know of) and follow the Pauli Exclusion Principle (meaning that multiple identical fermions cannot be in the same quantum state [41, p. 264]), which leads to them following Fermi-Dirac statistics (hence the name) [39, p. 205]. The fermions can be divided into quarks, which interact via the strong force, and leptons, which do not. Additionally, they can be divided into three generations. The particles in the first generation are less massive and more stable than ones in generation II and III, and due to this most of the matter we encounter in our daily life is made up of fermions from generation I [19](atoms, the basic building blocks of everyday matter, are made up of electrons, neutrons and protons, and these last two are made up of trios of up and down quarks). While the figure only showcases 12 types of fermions (the term "flavour" is generally used for these types, so we say there are 6 flavours of both leptons and quarks), there are actually 24, since every particle has a corresponding antiparticle with the same mass and opposite charge (e.g. the positron, which is the positively charged antiparticle of the electron).
Bosons have integer spin, are not beholden to the Pauli Exclusion principle (so multiple identical particles can occupy the same state) and consequently follow Bose-Einstein statistics [39, p. 205] (hence, once again, the name). The gauge bosons have spin 1 (making them vector bosons) and are force-carriers for three of the fundamental forces: electromagnetism (carried by the photon), the strong nuclear force (carried by the gluon) and the weak nuclear force (carried by the $W^{\pm}$ and Z bosons)[19]. The Higgs boson has spin 0 (making it a scalar boson), and is the visible quantum manifestation of the Higgs field, which interacts with all elementary particles to give them mass through the Brout-Englert-Higgs mechanism. [17]

b **Hadrons: baryons and mesons**

The particles mentioned until now were the elementary particles of the Standard Model (i.e. the most fundamental ones, not made up of any smaller components-at least not according to the SM), but these can combine to form more complex composite particles. Notably, quarks combine into bound states called hadrons, which can be divided into two main types, : baryons and mesons. The quarks are connected through the strong nuclear force (mediated by the gluon).[38]

The quantum field theory describing how quarks interact through the strong force is called quantum chromodynamics (QCD). According to this theory, each quark can be assigned a property called "colour" (note that this term is used for the sake of analogy, and isn't physically connected to what we normally understand as being the colour of an object, which depends on the frequency of the light reflected or emitted by it). Quarks can then be "red", "green" or "blue", and antiquarks are "antired", "antigreen" or "antiblue". Colour confinement tells us that free particles united by the strong force (baryons) must be "white" or "colourless", meaning that they must either be composed of three red green and blue quarks (baryons) or a quark and anti-quark pair with complementing colours (mesons) (n.b. technically baryons and mesons can contain more particles than indicated here, needing only to have an odd or even amount of quarks/antiquarks, respectively, but most hadrons will only have the 2 or 3 quarks). Due to the rules of spin addition, baryons are fermions, while mesons are bosons [38].

The two most well known hadrons are the proton (quark composition: uud) and the neutron (quark composition: udd), both baryons that make up the atomic nucleus.

Mesons aren't something most come across in their daily life, but some of them are very relevant to this project, and will be discussed later on in section 1.2. [38]

Hadrons have antiparticles, just like their elementary constituents. The quark composition of an antihadron is the corresponding composition of its hadron but with the quarks replaced with their respective antiquarks, and vice-versa. For example, the antiproton's quark composition is $\overline{u}\overline{u}\overline{d}$ (n.b. in this context, a line over the symbol of a quark indicates its repective antiquark, so e.g. $\overline{u}$ is the up antiquark). It should be noted that some mesons, such as J/ψ (quark composition: $c\overline{c}$) can be their own antiparticle.

c Limitations of the Standard Model and New Physics

Despite the successes of the SM, it is not perfect, and has issues that indicate it is not the final theory for Particle Physics. There are quite a few unexplained phenomena that the SM does not account for. These include the fact that we find more matter than antimatter in our Universe (when the SM predicted they should have been generated in equal amounts), the existence and unexplained nature of dark matter and energy, the fact that neutrinos appear to have mass, among others. [19]

One of the most prominent features missing from the SM is a Quantum Field Theory for gravity. While we currently have quantum theories for the other three fundamental forces (electromagnetism and the weak and strong nuclear forces), such as Quantum Electrodynamics (QED) for electromagnetism and QCD for the strong force, and have found the corresponding gauge bosons for these interactions, our best theory with experimental evidence for the gravitational force is Einstein's General Theory of Relativity, which is not quantum in nature. Particle Physics was developped through the combination of Special Relativity and Quantum Mechanics, but the fusion of General Relativity and Quantum Mechanics has still not been completely achieved. Still, theoretical predictions have been made, and we even have a hypothetical gauge boson for gravity, the graviton, which we expect to have spin 2 (tensor boson). [39, p. 403-407] But no experimental evidence for such a particle has been found yet.

Because of these issues, physicists have worked on new theories that go beyond the Standard Model and try to make up for its shortcomings. These theories and hypotheses are collectively referred to as New Physics (NP), or "Physics Beyond the Standard Model" (BSM) and have attracted a lot of projects and experiments (including the one my work is tied to).

1.2 B-Physics

a B mesons

B mesons are a small family of mesons containing a bottom antiquark. The other particle must either be a strange (B_s^0), charm (B_c^+), up (B^+) or down($B^0/B_d/B_d^0$) quark (n.b. the other particle being a top quark is thought impossible, and if its a bottom quark the meson is called bottomonium and is not considered a B meson). There are of course also antiparticles for each of these with a bottom quark and corresponding antiquarks for each type.

b The $B^0 \to K^{*0}ll$ decay

The B^0 meson has a mean life of only 1.5 ps (1 ps=10^{-12} s)[38], meaning that it must quickly decay to other particles. It has over 500 channels of decay, but the one relevant to this thesis is the $B^0 \to K^{*0}ll$ decay (n.b. K^{*0} here refers specifically to the $K^*(892)^0$ variation of the kaon, and we usee ll to indicate l^+l^-, i.e. an lepton-antilepton pair, particularly muons and electrons). The K^{*0} will go on to decay into a $K^+\pi^-$ pair. As always, there is a corresponding antiparticle decay $\overline{B^0} \to \overline{K^{*0}}ll$, where $\overline{K^{*0}} \to K^-\pi^+$.

If this decay happens directly , it is referred to as a non-resonant decay, but it can also happen through intermediary particles, which call a resonant decay. It is of specific interest to us cases where the intermediate particles go on to decay into the leptons. One example is the J/ψ resonance: $B^0 \to K^{*0}J/\psi$ where $J/\psi \to ll$. Other possible resonances are through the D^0, $\psi(2s)$, η, ω, and ϕ mesons, among others.

The table below lists some of the mesons most relevant to this thesis:

Meson	**Quark composition**	**Mass (MeV/c²)**	**Antiparticle**	**Mean lifetime (ps)**
B^0	$d\bar{b}$	5280	$\overline{B}^0(b\bar{d})$	1.519
B^+	$u\bar{b}$	5279	$B^-(b\bar{u})$	1.638
K^{*0}	$d\bar{s}$	892	$\overline{K}^{*0}(s\bar{d})$	Negligible
K^+	$u\bar{s}$	494	$K^-(s\bar{u})$	$1.238 \cdot 10^4$
π^+	$u\bar{d}$	140	$\pi^-(d\bar{u})$	$2.603 \cdot 10^4$
π^0	$\approx \frac{u\bar{u}-d\bar{d}}{2}$	135	Itself	Negligible
η	$\approx \frac{u\bar{u}+d\bar{d}-2s\bar{s}}{\sqrt{2}}$	548	Itself	Negligible
J/ψ	$c\bar{c}$	3097	Itself	Negligible
$\psi(2S)$	$c\bar{c}$	3686	Itself	Negligible
ϕ	$s\bar{s}$(almost purely)	1019	Itself	Negligible

Table 1: List of relevant mesons and their basic properties.[38]

c Some useful concepts: Feynman Diagrams and Branching Ratios

Feynman Diagrams are some of the most important tools in Particle Physics. On a surface level they may just look like basic representations of interactions between particles, but they actually represent mathematical terms that can be summed together to predict results. A certain interaction, with given inputs and outcomes, can be represented by different diagrams, and by summing all the diagrams together (there are generally infinite diagrams in theory, but we can limit ourselves to the leading order contributions, as is customary in all of Physics) we can obtain the complex amplitude $\mathcal{A}$ for this interaction. The square of the absolute value of this amplitude corresponds to the probability of the interaction occurring, and this in turn can be used to calculate various quantities such as cross-sections or, more relevant to us, decay rates (Γ).

Branching ratios ($\mathcal{B}$), also known as branching fractions, indicate what fraction of the decays of a certain particle correspond to a specific decay. For example the branching ratio of the $B^0 \to J/\psi K^{*0}$ decay is around $1.27 \cdot 10^{-3}$, meaning that approximately 0.127

% of decays of the B^0 mesons will be into $J/\psi K^{*0}$[38]. The branching ratio can also be defined as the ratio of the decay rate of a specific decay mode to the total decay rate of the particle (Γ_i/Γ), which is mathematically equivalent to the previous definition.

d Flavour Changing Neutral Currents, Lepton Flavour Universality and New Physics

The $B^0 \to K^{*0}ll$ decay is an example of a flavour-changing neutral current (FCNC), which is a process where a lepton changes its flavour but not its charge. In this particular case, the bottom antiquark of the neutral B meson turns into the strange antiquark of the neutral Kaon, and both antiparticles have the charge $+\frac{1}{3}$. According to the SM, the Feynman diagrams of FCNCs must contain a loop, i.e. they cannot be tree diagrams. But in NP we can consider the possibility of tree diagrams for FCNCs, involving hypothetical new particles like leptoquarks or unknown gauge bosons. The figure below exemplifies leading-order contributions from both SM and NP diagrams:

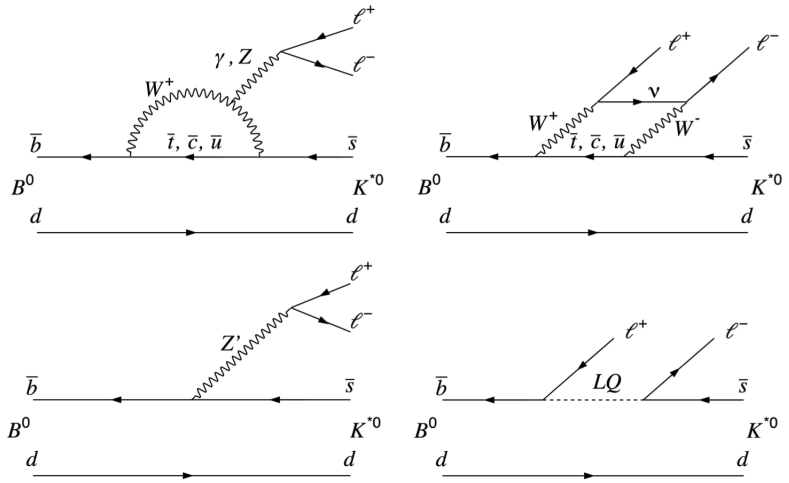


Figure 1.2: Feynman Diagrams for the $B^0 \to K^{*0}ll$ decay. The top ones are allowed by the SM and contain loops with electroweak couplings (i.e. couplings involving electromagnetic and/or weak nuclear forces and gauge bosons), more especifically a penguin loop for the left and a box loop for the right. The bottom ones are tree-level diagrams mediated by hypothetical NP particles, like a new gauge boson we refer to as Z' (left) or a leptoquark we refer to as LQ (right) [54]

We can therefore separate the contributions to the total amplitude of this decay into 2: the contributions from processes compliant with the SM, whose sum we call $\mathcal{A}_{SM}$, and the ones only allowed in Physics beyond the SM, which add up to $\mathcal{A}_{BSM}$. Thus, according to the SM, the probability of the decay will be given by $|\mathcal{A}_{SM}|^2$, while in NP it expands to $|\mathcal{A}_{SM} + \mathcal{A}_{BSM}|^2$.

Changes in which Feynman diagrams contribute to the amplitude and probability of the decay can lead to observable differences in the way the decay occurs, and we can try to measure these. One method is to do an angular analysis of the decay, defining and measuring various angular parameters and comparing these to the SM predictions, to see if there are significant differences. This method is not relevant to this thesis.

Another method involves a property from the SM called Lepton Universality (LU). According to LU, the interactions between electroweak gauge bosons ($W^{\pm}$,Z^0 and γ) and

leptons must be independent of the flavour of the latter. Thus the upper diagrams in figure 1.2 must contribute equally to the amplitude regardless of whether l is an electron or a muon (we concentrate on these 2 lepton flavours because the heavy tau lepton is harder to measure and calculate with, and neutrinos are very hard to detect at all). Since the amplitudes will directly affect decay rates (Γ), as mentioned, in the SM we expect the following ratio to be equal to 1:

$$R_H = \frac{\int \frac{d\Gamma(B^0 \to H\mu^+\mu^-)}{dq^2} dq^2}{\int \frac{d\Gamma(B^0 \to He^+e^-)}{dq^2} dq^2} \tag{1.1}$$

Here H can represent any hadron with an s quark, particularly K^{*0} [54]. We integrate the decay rate over every possible value of the square of the dilepton invariant mass, from $q^2_{min} = 4m_l^2$ to $q^2_{max} = (m_{B^0} - m_H)^2$, assuming we're working in the rest frame of the B meson. Moving onto the specific $H = K^{*0}$ case, this formula can also be represented in the following form (keep again in mind that branching rations can be represented as ratios between decay rates):

$$R_{K^{*0}} = \frac{\mathcal{B}(B^0 \to K^{*0}\mu^+\mu^-)}{\mathcal{B}(B^0 \to K^{*0}e^+e^-)} \tag{1.2}$$

However, we will calibrate both of these non-resonant channels by dividing their branching ratio by the branching ratio of the corresponding J/ψ resonant channel. By doing so, we can remove many systematic uncertainties that the resonant and non-resonant channels have in common:

$$R_{K^{*0}} = \frac{\mathcal{B}(B^0 \to K^{*0}\mu^+\mu^-)}{\mathcal{B}(B^0 \to K^{*0}e^+e^-)} \cdot \frac{\mathcal{B}(B^0 \to K^{*0}J/\psi(\to e^+e^-))}{\mathcal{B}(B^0 \to K^{*0}J/\psi(\to \mu^+\mu^-))} \tag{1.3}$$

In order for the two versions of the ratio be equal, we must have $r_{J/\psi} = \frac{\mathcal{B}(B^0 \to K^{*0}J/\psi(\to \mu^+\mu^-))}{\mathcal{B}(B^0 \to K^{*0}J/\psi(\to e^+e^-))} = 1$. This ratio was measured in [54] to be $1.043 \pm 0.006(stat.) \pm 0.045(syst.)$, which is in good agreement with unity once we consider the statistical and systematic uncertainties.

The deviation of $R_{K^{*0}}$ from unity is therefore a good indicator of violation of LU, and the presence of NP processes.

Now, it can be asked why we decided to use the J/ψ resonance instead of some of the others already mentioned. Well, beside knowing that $r_{J/\psi}$ has been measured to be close to unity, which would minimize further systematic uncertainties brought forth by the calibration, there is also the fact that the J/ψ resonance has the highest branching ratio of them all, and also its decay to leptons also has higher branching ratio than the others, and these higher branching rations reduce statistical error.

e Current status of measurements

Various experiments for detecting NP and LU violations in B meson decays have already been carried out at CERN. Angular analyses of the $B^0 \to K^{*0}\mu^+\mu^-$ decay have been done by LHCb [6] and ATLAS [4] back in 2016 and 2018, respectively. Meanwhile, in 2017 LHCb calculated $R_{K^{*0}}$ to be $0.66^{+0.11}_{-0.07}(stat) \pm 0.03(syst)$, for $0.045 < q^2 < 1.1 \quad GeV^2/c^4$, and $0.69^{+0.11}_{-0.07}(stat) \pm 0.05(syst)$, for $1.1 < q^2 < 6 \quad GeV^2/c^4$, which corresponds to a difference of 2.1-2.3 and 2.4-2.5 standard deviations from the expected SM value, respectively [54].

LU violations have also been studied in other B meson decays. In March 2021, for the $B^+ \to K^+ll$ decay, the LHCb experiment calculated R_{K^+} to be $0.846^{+0.042}_{-0.039}(stat)^{+0.013}_{-0.012}(syst)$, for $1.1 < q^2 < 6 \quad GeV^2/c^4$, which corresponds to a difference of 3.1 standard deviations from the expected SM value[55]. In October of the same year, for the $B^0 \to K^0_S ll$ and $B^+ \to K^{*+}ll$ decays, LHCb calculated $R_{K^0_S} = 0.66^{+0.20}_{-0.14}(stat)^{+0.02}_{-0.04}(syst)$, for $1.1 < q^2 < 6 \quad GeV^2/c^4$ and $R_{K^{*+}_S} = 0.70^{+0.18}_{-0.13}(stat)^{+0.03}_{-0.04}(syst)$, for $0.6 < q^2 < 6 \quad GeV^2/c^4$, corresponding to a separation of 1.5 and 1.4 standard deviations from the SM prediction, respectively[56].

Lepton Universality tests have been done in other facilities outside of the LHC and CERN, such as BaBar at SLAC, in the USA, and Belle at Kek, in Japan. Tests have also been done with decays of other particles than B mesons, such as e.g. Z and W bosons, pions, kaons, D mesons, and even decays of the τ lepton into neutrinos, electrons and muons. [16]

Chapter 2

The ATLAS experiment

2.1 The LHC

The Large Hadron Collider (LHC), is a particle accelerator located between the border of France and Switzerland, near Geneva, developed by the European Organization for Nuclear Research (CERN). It began operating on the 10th of September 2008, but due to technical issues had to stop 9 days later. It eventually was repaired and particle beams began circulation in November 2009, leading to the first collisions of protons. [44]

a Structure and functioning of the CERN Accelerator Complex and the LHC

The LHC is part of a larger system called the CERN accelerator complex, whose structure is visualized in the image below:

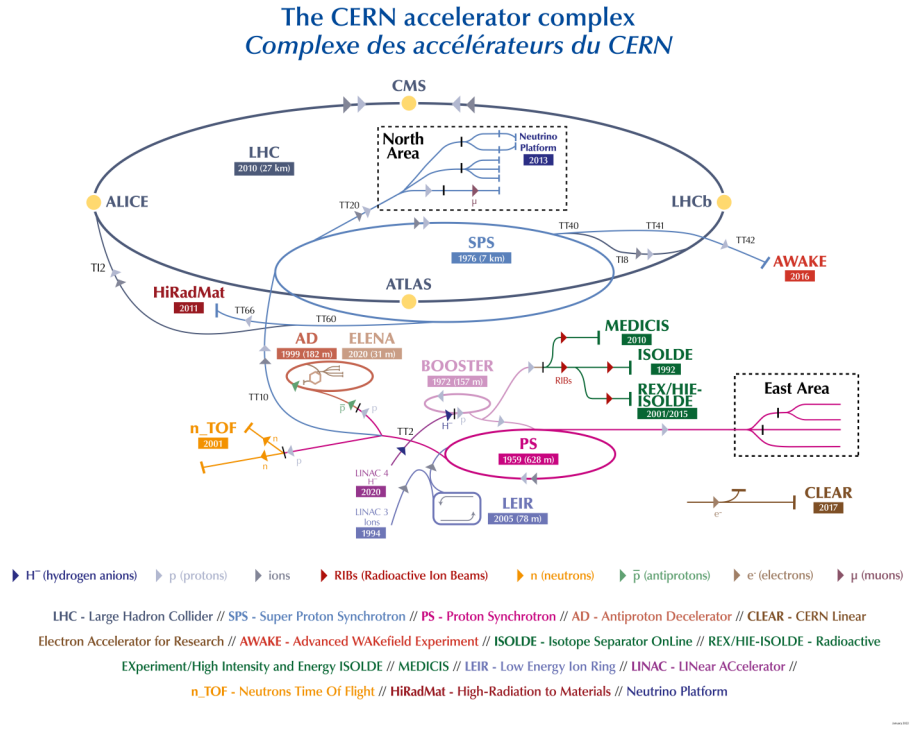


Figure 2.1: The CERN Accelerator Complex [34]

This complex accelerates the hadrons using various different accelerators, which gradually give them more and more energy. Protons (technically hydrogen anyons, protons with two electrons), for example, are accelerated in the Linear Accelerator 4 (Linac4), the PS Booster (upon entering this the electrons are removed, leaving only protons), the Proton Synchroton (PS) and the Super Proton Synchroton (SPS), after which they are finally inserted into the LHC in two beam pipes, and both beams are accelerated to speeds close to the speed of light, and acquire a maximum energy of 6.5 TeV (this value was increased to 6.8 TeV in Run 3 of the LHC, leading to a total energy of 13.6 TeV). The complex can also accelerate lead ions instead of protons, using the Linac3 and the LEIR(Low Energy Ion Ring), before sending the ions into the PS. [18]

Shaped like a ring, the LHC has a circumference of 27 kilometers. It uses superconducting electromagnets (kept at a temperature of -271.3^{o}C, through the use of e.g. liquid helium. Their superconducting state allows us to neutralize resistivity and loss of energy through heat) to guide the two beams of high-energy hadrons around the structure in opposite directions, until their eventual collision. Each proton beam is divided into 2808 bunches which each contain $11 \cdot 10^{11}$ protons at start, giving us a peak design luminosity (number of particles per second per unit of area) of $10^{34} \cdot cm^{-2} \cdot s^{-1}$. An ultrahigh vacuum is kept inside the two beam pipes (the pressure inside is of the order of 10^{-10} to 10^{-11} mbar) the hadrons travel in. The particles are accelerated here using 16 radiofrequency (RF) cavities, who operate in a superconducting state and use electric pulses (created by klystrons, tubes with electrons inside) to increase the particles' energy from 450 GeV to their final 6.5 TeV, a process which takes around 20 minutes. At this energy, the particles complete 11 245 revolutions per second. The LHC uses almost 9600 magnets to guide the beams, including 1232 beam-bending dipole magnets and 392 beam-focusing magnets.

There are 4 main LHC experiments, located at 4 possible collision points for the hadrons: ATLAS ("A Toroidal LHC ApparatuS"), CMS ("Compact Muon Solenoid"), ALICE ("A Large Ion Collider Experiment") and the LHCb ("Large Hadron Collider beauty") experiment. There are also 3 smaller experiments installed close to one of the main ones: the LHCf ("Large Hadron Collider forward") experiment, the TOTEM ("TOTal Elastic and diffractive cross section Measurement") experiment, and MoEDAL ("Monopole and Exotics Detector at the LHC"). These experiments can contain multiple detectors, for the purpose of detecting and analyzing the particles resulting from the collisions. [18]

2.2 The ATLAS detector

This thesis was written as part of the ATLAS collaboration, so we will be focusing on this experiment/detector.

Located 100 m underground, ATLAS is a 44 meters long and 25 meters wide cylindrical detector weighing 7000 tonnes. These dimensions make it the detector at the LHC with the largest volume. More than a billion collisions occurs in here every second, though only a millionth of these are further studied. [46] [10]

Below we see a labelled diagram of the detector and its components.

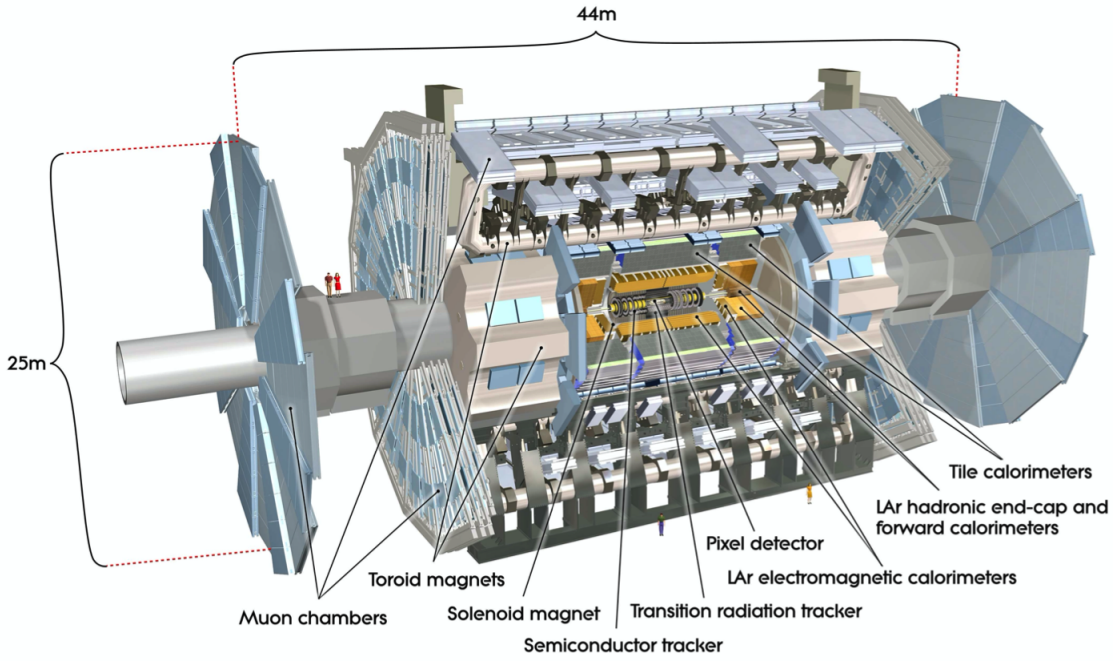


Figure 2.2: The ATLAS detector [46]

a Coordinate system

Both ATLAS and CMS employ a coordinate system with an origin at the collision/interaction point of the hadrons. The z axis has the direction of the beam, the x-axis points at the center of the collider, and the y-axis points upward. The x-y plane is of course transverse to the beam, and this is where the transverse momentum p_T and the radial distance r of the different particles are defined. We can define the directions of the particles using the polar angle θ, measured from the positive z axis, and the azimutal angle ϕ, measured in the transverse plane from the positive z-axis. This allows us to define the transverse energy $E_T = E \cdot \sin\theta$, where E is the total energy of the particles.

From the polar angle, we define the pseudorapidity as $\eta = -ln\,[tan(\theta/2)]$, whose absolute value can vary between 0(if the particle is moving perpendicular to the beam) and infinity (for particles moving along the beam) (see fig. 2.3). The advantage of using this quantity over θ is the fact that differences in pseudorapidity are Lorentz invariant under boosts along the beam axis, which is useful when dealing with beams of particles moving very close to the speed of light. Angular distances between particles are generally defined in the (η,ϕ) plane as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ (this quantity will of course also be Lorentz invariant under boosts, as it depends on $\Delta\eta$, not η).

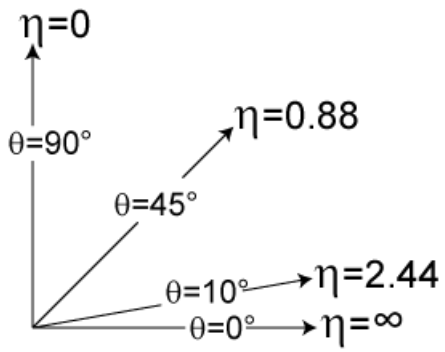


Figure 2.3: Values of the pseudorapidity for different values of the polar angle [27]

b The different ATLAS detector systems

ATLAS contains multiple detector systems that contribute to the detection and measurement of the properties of particles resulting from the hadron collisions. They are generally cylindrically shaped and enclose each other like layers in an onion [46].

The inner detector

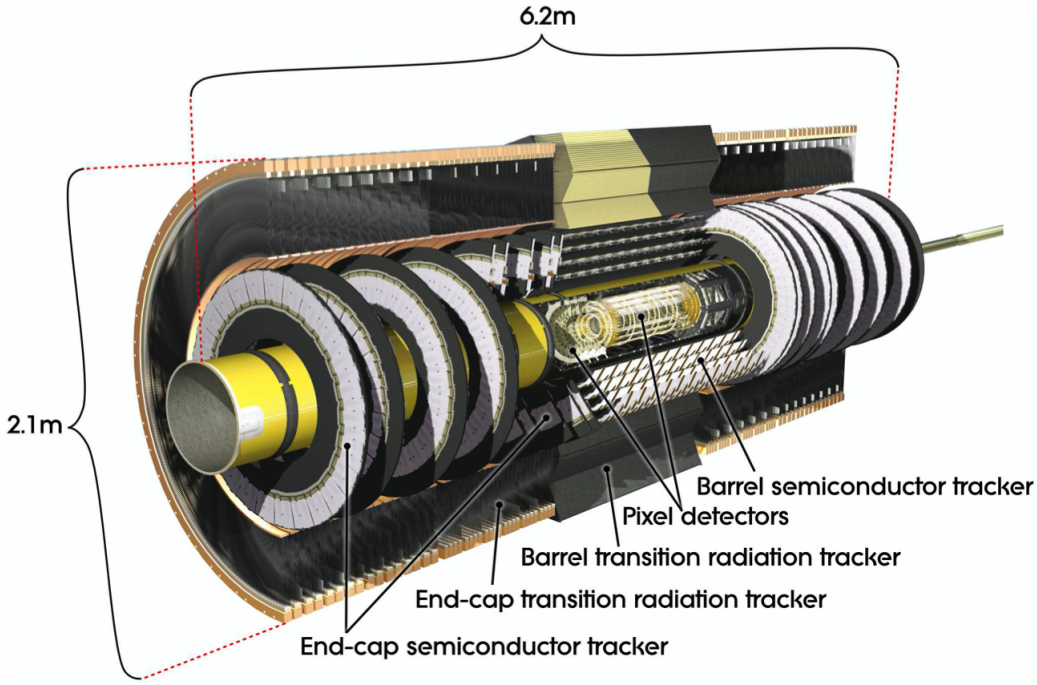


Figure 2.4: The ATLAS inner detector [46]

The ATLAS Inner Detector (ID) is the first detector system the resulting particles of the p-p collisions encounter, and its main function is tracking these particles and measuring their momentum, direction and charge. As seen in fig. 2.4, it contains three types of detectors: Pixel detectors, Semiconductor Trackers (SCT) and Transition Radiation Trackers (TRT).

The pixel detectors contain over 92 million pixels in total, which are divided between the 4 barrel layers around the beam (note: the innermost layer, the Insertable B-layer, was inserted in 2014 and is made up of 14 staves), containing 1736 sensor modules, and 3 disks at each end-cap, containing 288 modules. When a particle passes through the pixel

detectors, they trigger signals that can be used to measure the particles momentum and determine its origin. The pixel detectors are located in the radial interval of 33-150 mm.

Around the pixel detector is the SCT, which reconstructs the tracks of the charged particles using silicon sensors, with 4088 modules containing over 6 million "micro-strips" of silicon sensors. These are contained in 4 barrel layers and 9 discs at each end-cap. The SCT is contained in the 299–560 mm radial interval.

Around the SCT is the TRT, which is made up of 300000 "straws", thin-walled drift tubes with a diameter of 4 mm, which contain a goldplated tungsten wire inside of them, and which are filled with a mixture of gases. The wire is kept at high voltage, thus creating an electric field around t. When passing by, the particles ionise the gasses, and the free electrons move to the wires under the influence of their fields, thus creating signals that can be used to reconstruct their tracks. Fibres (in the barrel) and foils (in the endcaps) of a polymer material fill out the space between the straws, and when a relativistic particle traverses the boundary between the different materials it emits transiition radiation, which can be used to identify it (this effect depends only on $\gamma = \frac{E}{m}$ is therefore strongest for lower mass particles like the electrons and weaker for e.g. hadrons, if the particles have similar energy. This allows for particle discrimination by imposing probability thresholds) [57]. 50000 of the straws are located in the barrel region, and the rest at the end-caps. The TRT is contained in the 563–1066 mm radial interval.

Around the ID is a solenoid electromagnet, which creates a magnetic field with the direction of the field and an intensity of 2T, in which the previously mentioned sensors are immersed. Liquid argon cryostats are used for cooling the ID, in order to minimize thermal noise.[46] [11] Figure 2.5 is a diagram showcasing a cross-section of the entire ID in the r-z plane:

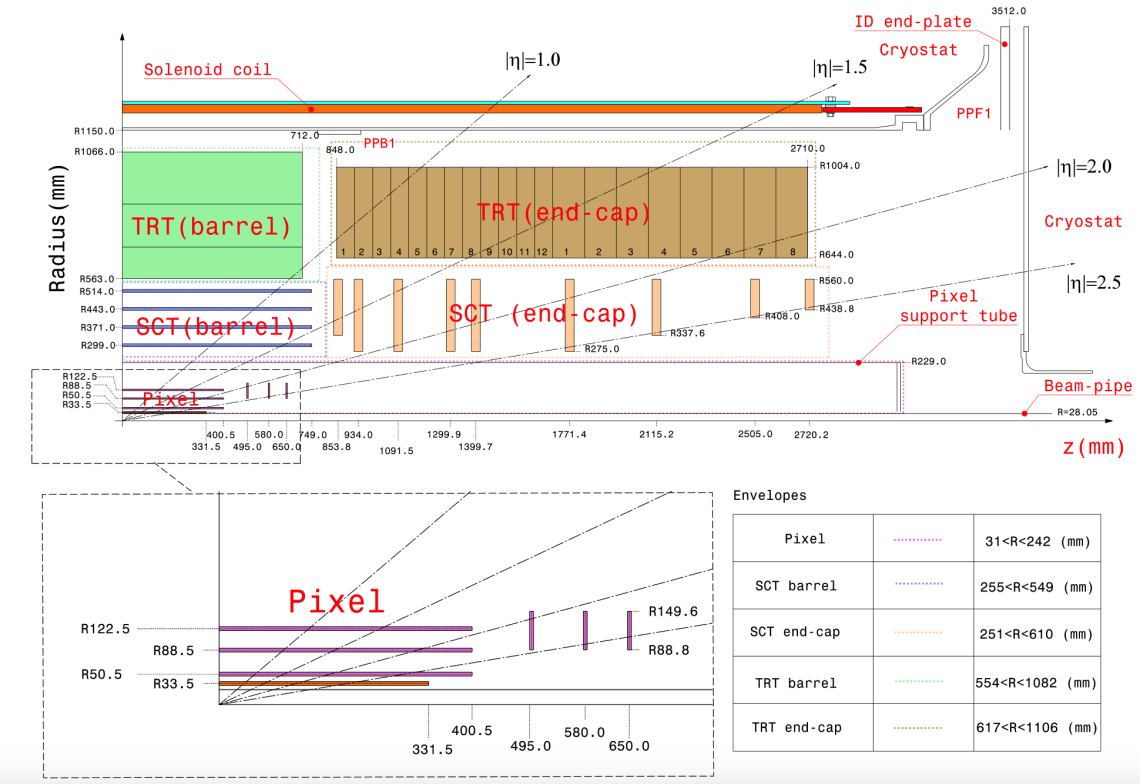


Figure 2.5: Diagram of the ID in the r-z plane, including a magnified view of the pixel detector. The new Insertable B-layer is shown in orange [49]

The calorimeters

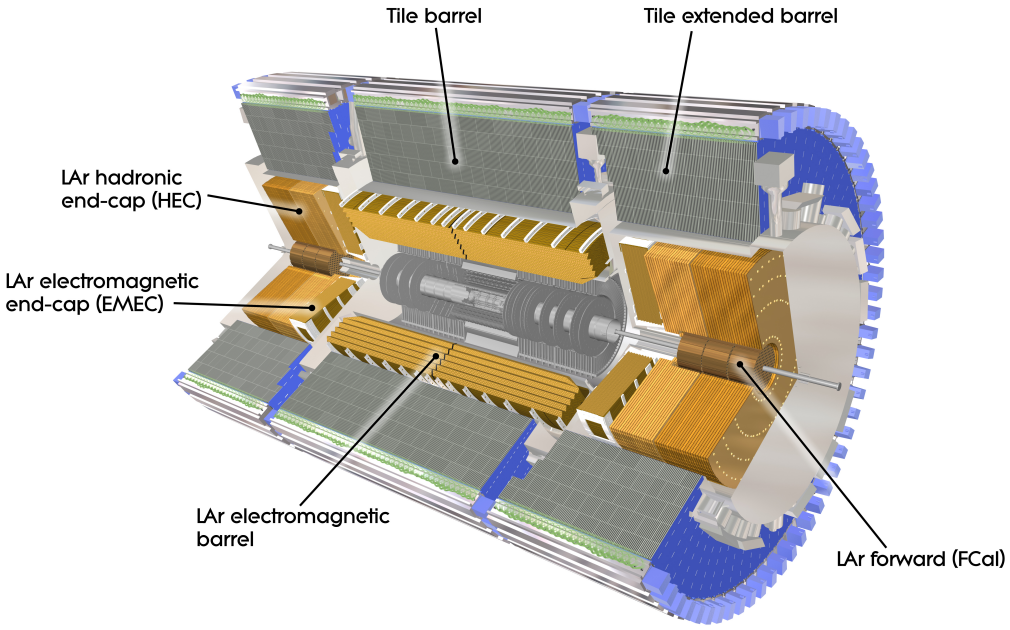


Figure 2.6: The ATLAS calorimetry system [46]

After interacting with the inner detector, the particles encounter the calorimeters. These will stop most of the particles (notable expceptions being muons and neutrinos) and absorb their energy, which can then be measured. (At ATLAS, the calorimeters can also be divided into two parts, based on their structure and composition: the Liquid Argon (LAr) Calorimeter and the Tile Hadronic Calorimeter). For electrons and photons, electromagnetic (EM) calorimeters measure their energy through their interactions with matter, while hadrons' energy is measured by hadronic calorimeters through their interaction with the nuclei of atoms.

Surrounding the ID are the LAr calorimeters, which are made up of metal layers with liquid Argon (kept at -184^{o}C) between them. Through absorption by the metal, particles turn into a shower of particles (see fig. 2.7) which ionise the Argon, leading to the creation of electrical currents which can be used to calculate the energy of the original particles. The electromagnetic barrel part has a length of 6.4 m, while the endcaps are disks with a radius of 2.09 m, and can be either electromagnetic or hadronic. Imbedded in the endcaps is the forward calorimeter, which is composed of three modules (first one is electromagnetic, the other two hadronic) with a radius of 0.455 m.

Surrounding the LAr Calorimeters is the Tile Calorimeter, whose purpose is to absorb and measure the remaining energy of the hadrons, since these are not fully absorbed by the LAr Hadronic Calorimeters. They contain steel layers, which turn incident particles into particle showers that interact with the ca. 420000 plastic scintillator tiles, causing these to emit photons which generate electrical currents that can once again be used to calculate the particle's energy. The Tile Calorimeter is composed 3 barrels with 64 wedges each: a center barrel with 5.6 m long wedges, and two extended barrels with 2.6 m long wedges. [46] [8]

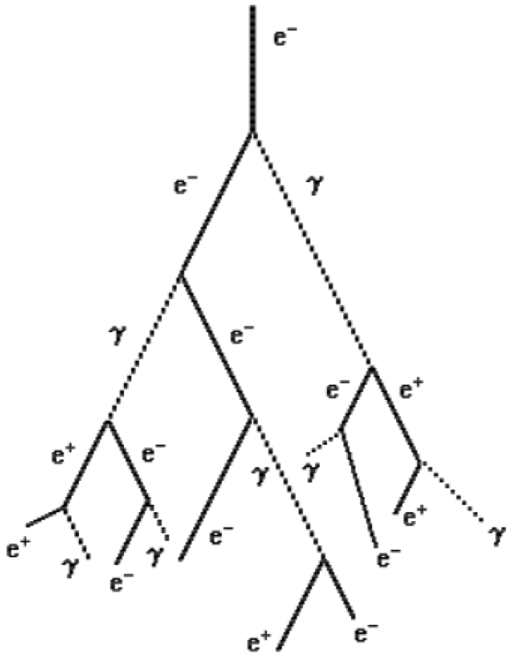


Figure 2.7: Particle shower created by an electron. When a high energy electron (or positron) interacts with the material in the calorimeter, it is decelerated and loses energy by emitting photons (bremsstrahlung). These photons can then turn into an electron-positron pair, and the process continues, until all the original electron's energy has been deposited in the calorimeter [42]

The muon spectrometer

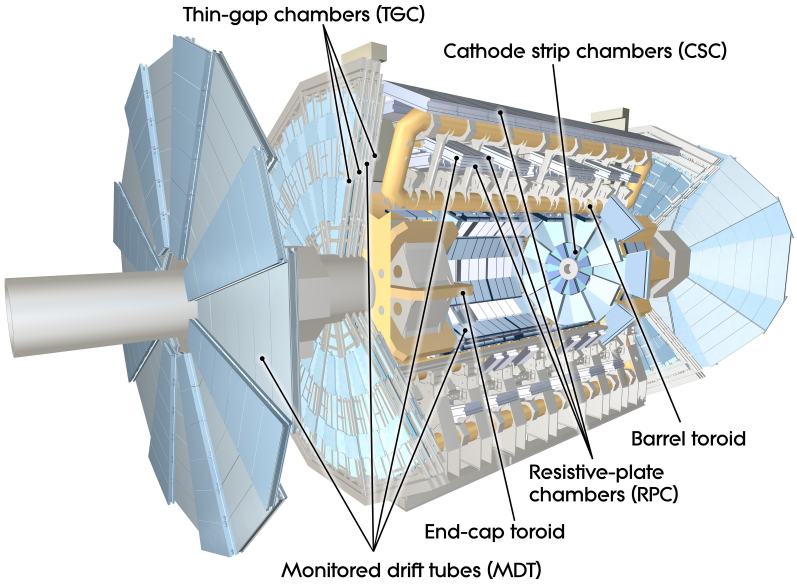


Figure 2.8: The ATLAS muon spectrometer [46]

The muons leaving the calorimeter are identified by the muon spectrometer, which also measures their momentum, based on the bending of their tracks under the influence of a magnetic field created by a toroid. The muon system contains 4000 chambers and employs 4 different technologies. The triggering and measurement of the 2nd coordinate (direction perpendicular to the plane where the bending occurs) is done by Resistive

Plate Chambers at the barrel region and Thin Gap Chambers at the endcap. Precision measurements of the curves of tracks is done by the Monitored Drift Tubes, while the Cathode Strip Chambers carry out precision measurements at the innermost plane of the spectrometer (see fig. 2.8). [46] [9]

c The trigger system

As mentioned, ATLAS experiences more than a billion collisions every single second. Analysing all of these would be extremely taxing, even for computers (the data obtained from all these collision corresponds to more than 1 PB every second). The purpose of the Trigger and Data Acquisition (TDAQ) system is selecting the fraction of these events which might actually be interesting to study and analyse. The structure of the TDAQ is visualized in the diagram below, which will be (partially) explained afterwards:

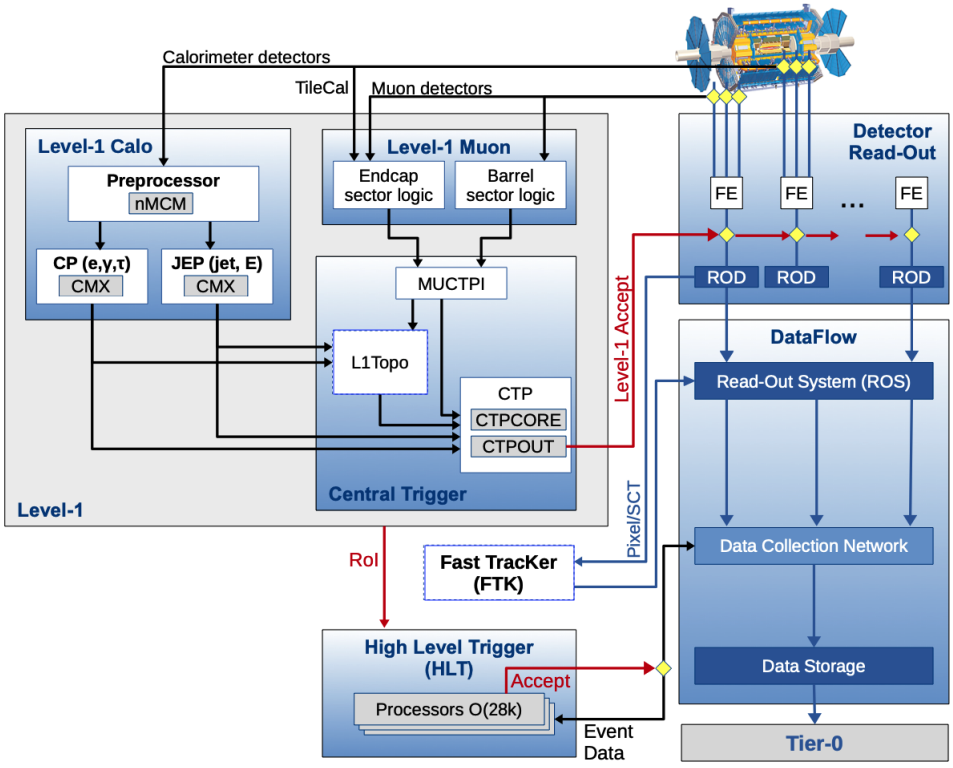


Figure 2.9: The ATLAS TDAQ system during Run 2 of the LHC. The image was slightly edited for this thesis, its source being [48]

There are two levels for the trigger system. The first level (L1) is hardware based, and decides whether to accept an event based on information obtained from the calorimeters and muon spectrometer. It takes less than 2.5 μs after an event's occurrence for it to be rejected or accepted by L1. When an event is accepted, the information from the detector's front-end (FE) electronics related to the event is transferred to its read-out drivers (ROD) and then buffered in the Read-Out System (ROS). Up to 100000 events can be accepted at L1 and transferred to the second level every second.

The second level, known as the High-Level trigger (HLT) is software based. With its processing power of about 40000 CPU cores, it analyses the events in 200 μs. Its triggering algorithms are based on Region of Interest (RoI) information received from L1.

Around 1000 events per second are approved by the HLT, and these are subsequentaly transferred to storage and then to the CERN TIER-0 facility where they can be analysed and reconstructed offline later. [48] [12]

d Electron particle identification

A big issue for the analysis is the identification, isolation and reconstruction of electrons (n.b. electron can mean either electron or positron in this context). Electrons are detected through their tracks in the ID and the energy they deposit in the calorimeters. However, given the low energy of electrons in this analysis, the signal is not very striking and the background is enormous, and these methods bring with them the risk of misidentifying the particles, and sometimes it can be hard to separate the electrons resulting from the decays we are studying from e.g. misidentified hadrons, or even electrons from different reactions we are not interested in.

When interacting with the material in the detector, an electron might lose energy (causing it to slow down) through the emission of radiation known as bremstrahlung ("braking radiation"), in the form of photons. These photons might give origin to an electron-positron pair, and these particles may in turn emit more radiation, continuing the process. Thus, a single electron resulting from the decay can turn into an entire electromagnetic cluster of photons and electrons. Reconstruction of the original electron is done by matching tracks left by the electrons/positrons in the ID and clusters of energy left by the particles in the calorimeters. Error might happen here if tracks are missed or wrongfully added to an electron's reconstruction.

Identification/selection of prompt electrons is done using a likelihood function:

$$L_{S(B)}(x) = \prod_{i=1}^{n} P_{S(B),i}(x_i) \tag{2.1}$$

The n inputs of the likelihood function, saved in the vector x, are values obtained from the calorimeters and tracking systems. The likelihood can be calculated for signal (S) or background (B) electrons, using their respective probability density functions for each variable, P_i. From the likelihood, we can calculate a discriminant for each candidate:

$$d'_L = -\tau^{-1} ln(d_L^{-2} - 1), \quad with \qquad d_L = \frac{L_S}{L_S + L_B} \tag{2.2}$$

Here τ is set at 15. The reason we use the transformed discriminant d'_L instead of d_L is because the later has sharp peaks at 0 and 1, making it inconvenient to deal with, both numerically and in subsequent analysis. Four values of the discriminant are used as operating points, called VeryLoose, Loose, Medium and Tight, corresponding to tresholds for acceptance of a candidate. For all of these some amount of background and signal will always be accepted and rejected, and the tighter the operating point, the higher the rejection is for both.

Differentiating the signal promp electrons from background candidates can be done by analyzing the activity in the vicinity of a candidate, e.g. by defining a cone with a certain angular distance ΔR value around the candidate, and summing transverse energy (from the calorimeters) or transverse momenta (of tracks in the ID) inside this cone. Thus, results from the calorimeter and ID are used to quantify the electron isolation. Once again, requirements on the isolation of the particles can be varied, with different efficiencies for signal and background rejection.

Lastly, the identification of the electron charges is done by analysing the curvature of the reconstructed track in the ID. Sources of misidentification include matching incorrect tracks to the candidate (remember that multiple tracks can be produced by the same electron, and there might be other electrons/photons close-by that can cause issues), and difficulties in determining the curvature of the track of the primary electron, e.g. due to bad reconstruction, emission of photons or tracks becoming very straight at high energies or large values of η, which makes it harder to measure their curvature. [13] [14]

Chapter 3

Introduction to Machine Learning and LightGBM

Machine learning (ML) models, as implied by their name, allow computers to learn by themselves how to solve a problem. They do this by first trying out a solution, analyzing their errors, and adjusting the models in order to minimize these errors and improve. Two of the main tasks in machine learning are regression, where the model outputs real numbers (or sequences) based on the data points it uses as input, and classification, where the output is an integer that indicates a class (e.g. 1 for signal and 0 for background), based again on the inputs.[25, ch. 1.5] This project is a classification problem, since we need to determine whether each event is a background or signal event.

3.1 Mathematical definitions

Since we're trying to develop algorithms for computers, it is useful to mathematically describe what we're doing, and define terms such as "errors".

We have the data, which will be our input for the model. In our specific case, we have for each event a collection of n attributes $x_1, ...x_n$, which are something like the mean lifetime of the B meson or the transverse momentum of the electrons, and also a label y_T (T for "TRUE") which can be 0 (for background) or 1 (for signal).

The model will assign to each event a label $\hat{y}$ based on the inputs, i.e. $\hat{y} = f(x_1...x_n)$, where f is a function representing the model. $\hat{y}$ is an estimate of y_T, although it can take values between 0 and 1.

A loss function is a way of mathematically defining the "error" of the model. For regression problems it's generally the root mean squared of the difference between $\hat{y}$ and y_T, but for classification problems with the options "0" and "1"we can use the binary cross entropy loss function (logloss), shown below:

$$Loss = -\frac{1}{N}\sum_i y_{T,i} \cdot log(\hat{y}_i) + (1 - y_{T,i}) \cdot log(1 - \hat{y}_i) \tag{3.1}$$

Where N is the total number of events. [20]

3.2 Gradient boosted decision trees and LightGBM

One of the simplest machine learning models are decision trees (DT). They divide the events on the data into different nodes depending on the values of the inputs. If these nodes don't divide further they are considered "leaves" and all the events in it are assigned a classification. The maximum number of decisions an event can go through before being assigned a leaf is called the depth of the tree. Below we see a hypothetical decision tree with three leaves and a depth of two:

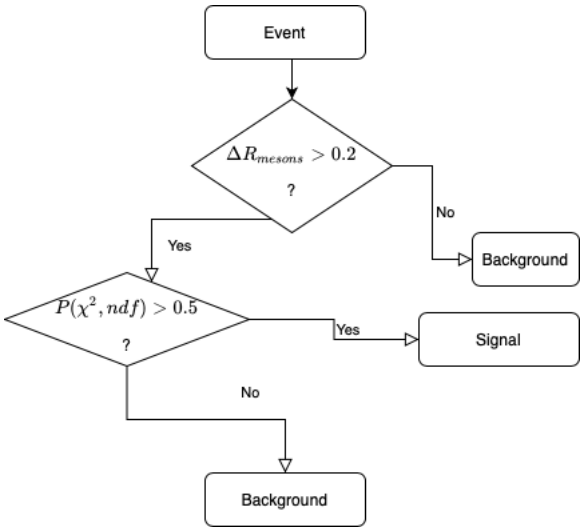


Figure 3.1: Basic example of a decision tree, created by me at `https://app.diagrams.net`

Decision trees by themselves are quite simple and considered weak learners, but can be improved upon through the use of gradient boosting. While some events might be clearly within the signal or background groups based on their parameters, others are less clear and might end up mislabelled. Gradient boosting allows us to develop more trees where we give more weight to the mislabelled events, allowing us to focus more on them and their proper separation. Thus we obtain gradient boosted decision tress (GBDT). By creating more trees with this boosting, the idea is to reduce the loss function (the gradient in GBDT of the loss function as it decreases). A parameter called the learning rate (α) determines how fast the learning happens and how quickly the loss function decreases. A good machine learning algorithm must have a learning rate that is neither too slow, which would make it inefficient, or too fast, which could cause problems like missing the minimum of the loss. For each event, the GBDTs vote on how to classify it through their output.[7]

A good framework for creating GBDTs is LightGBM[30]. Its trees grow in a leaf-wise manner, meaning the algorithm develops the tree by picking the leaf whose splitting will maximize the loss difference: [33]

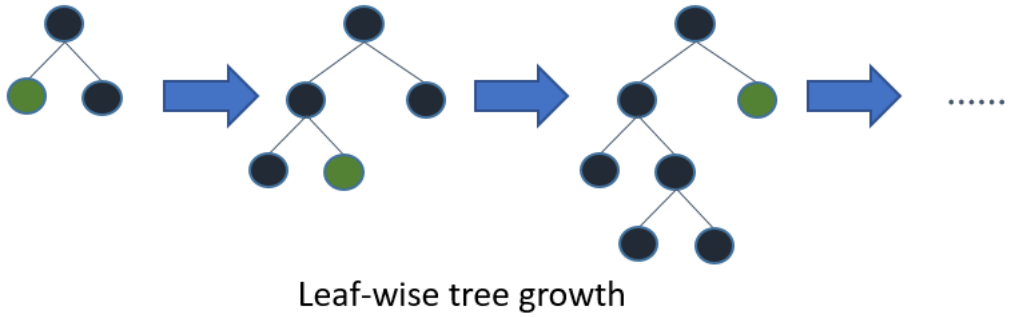


Figure 3.2: Showcasing of the development of decision trees in LightGBM [32]

3.3 Training, testing and validation sets. Overfitting and early stopping

After creating the algorithm using our data, it is good to verify how accurate it is. So we apply the model to the data and see how well the results match. However, it doesn't make sense to apply the algorithm to the same data that we used to train the model, since they will obviously be very correlated and biased. And therefore we divide our data into three sets: training, validation and testing. The training data, which was 60% of the total in this project, is the data the algorithm learns from to develop the trees.

As mentioned before, the model can keep creating more trees with the intention of reducing the loss function. And therefore we can constantly decrease the loss of the validation set towards 0. However, if we go too far we can start overfitting, where we create a model that tries to follow too closely the particular set we used for training, including statistical fluctuations, while not being applicable to the rest of the data. The purpose of the validation set (we use 20 % of the data for this) is to stop the process before we enter that zone. So we apply the model created both to the training and the validation data, and while the training loss will keep on decreasing (although it does plateau, because it won't go below zero), the loss of the validation will not only eventually stop decreasing as we increase the number of trees, but will even start increasing as we go on. So we stop at the point where the validation set's reaches its lowest value. This method is called early stopping.

After we've determined the ideal number of trees for our model, we can then test it on the testing set, which makes up the remaining 20 % of the data. [26] [40]

Chapter 4

Analysis

4.1 Analysis overview

a The decays: signal and background. Monte Carlo and data

As mentioned in chapter 1, the project aims to study the $B^0 \to K^{*0}ll$ decay, with the objective of calculating R_{K*0}. My analysis focuses specifically on the electron channel,i.e. the $B^0 \to K^{*0}ee$ decay (once again, electron here can denote either an electron or positron, and ee means e^+e^-).

A big problem with the analysis is the large amount of background decays happening at the same time. Since we can't observe the decays directly, but only make measurements in the detector of the properties of products resulting from these decays, it is important to find ways to differentiate the resulting particles from decays we are studying (signal decays), from similar resulting particles of other decays (background decays).

Monte Carlo (MC) simulations were used to simulate both the signal and background decays we wanted to consider. Through these simulations, we can not only simulate the events themselves, but also the measured properties of the final particles resulting from the event. When trying to reconstruct an event, there might be multiple "candidates" for the final particles, which for the decay we're studying should be two electrons and a $K^+\pi^-$ pair (note that since the detectors can't completely identify the resulting hadrons, there are many particles that could be mistaken for a K^+ or a π^-), and an advantage of simulation is that the signal files include a flag indicating which of the candidates is the correct one for reconstructing the events.

Aside from the results of the MC simulation, we also used real data from measurements of the resulting particles from pp collisions, taken in 2018 during Run-2 of the LHC. The data comes from different periods, and the center-of-mass energy of the collisions is 13 TeV. An overwhelming amount of the results obtained from the data will be from background events, especially for masses significantly different from the B^0 mass of 5280 MeV, so we can use the candidates from these data events as background as well.

b Selection of event candidates: pre-selection and GBDT

Before plotting the results, all the candidates were subjected to a pre-selection process whose aim was to remove any candidates that are too different from the expected values for results of the signal decays, or real results of a decay in general. For this pre-selection we consider things like angular distance (ΔR) between electrons, $|\eta|$ and p_T of electrons

and tracks, quality of tracks and the masses of the electron pair, the hadron pair and the reconstructed B meson. Note that this pre-selection was already chosen by the ATLAS group I'm working with.

The events that pass the pre-selection will then be used to train a Gradient Boosted Decision Tree (GBDT), specifically one obtained from the Light Gradient Boosting Machine (LightGBM) framework. This algorithm is quite good at separating background from signal, and is much faster than the Graph Neural Networks (GNN) used by other members of the ATLAS group, as the GBDT can be trained with more than 300000 signal and background events in less than 10 seconds.

Since we are studying and graphing the distribution of the invariant B mass and di-electron mass of the different events, and we will also observe the effect of the GBDT selection on these distributions, we want the features fed to the ML algorithm to be independent from both of these masses, so that the output is also independent. Otherwise, we risk the algorithm becoming familiar with the shape of these distributions and trying to replicate it in its selection.

c Fitting

We studied the mass distributions of the signal events further through fitting them to certain functions.

The distribution of the invariant mass of the reconstructed B meson for each signal event was fitted to a sum of a gaussian and a crystal ball function. This was done for different regions of the value of η of each electron, corresponding to different parts of the detector that were reached by these.

4.2 The data used for analysis

The data was extracted from measurements carried out in the ATLAS detector in 2018 during Run 2 of the LHC. This data started being collected in June of 2018, leading to an integrated luminosity in the ATLAS detector of approximately 30 $fb^{-}1$ until the end of the year, as can be seen in figure 4.1 (keep in mind again that data gathering only started in June). Of course, most of this data will not come from the decay we're studying, but other processes (and on the other hand, most of the B meson decays will not be detected). The center-of-mass energy in these interactions ($\sqrt{s}$) was 13 TeV.

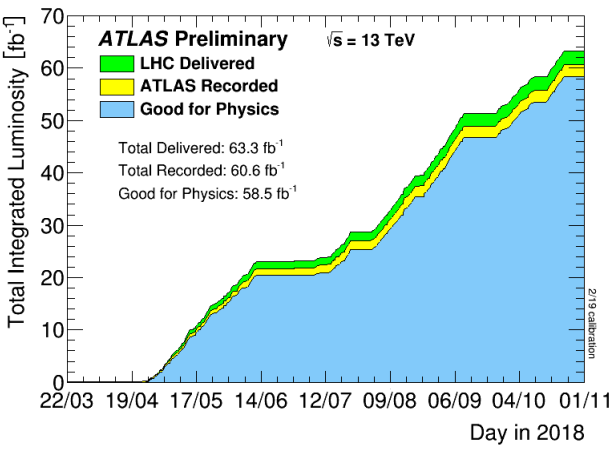


Figure 4.1: Graph of the total cumulative luminosity in the ATLAS detector throughout 2018, with distinctions between delivered luminosity (green), recorded luminosity (yellow) and luminosity of events with good data quality (blue). [50]

The advantage of using data from towards the end of Run-2 of the LHC is the availability of a more developped trigger system, specifically the addition of the Level-1 Topological Processor (L1Topo), which is one of the components appearing in figure 2.9. Due to the increase in luminosity, and consequently trigger rates in Run-2, it was necessary to filter more events, but simply raising the existing thresholds would remove too many of the useful events. L1Topo allows the hardware system to make new topological selections using treshold for new angular and kinematic quantities of individual objects in the events. B-physics project like this one benefit a lot from the new triggering options afforded by the L1Topo, since it can make selections based on e.g. the angular distance (ΔR) between the resulting leptons[5].An example of the difference in trigger rates with or without topological selections can be seen below, for muons:

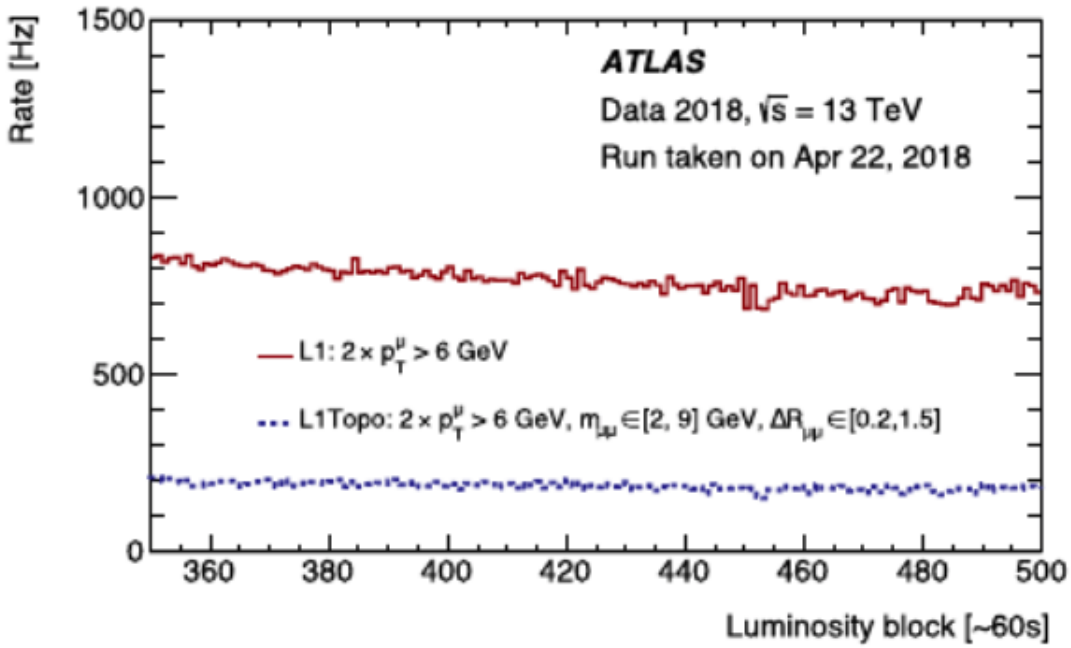


Figure 4.2: Comparison between dimuon trigger rates for the non-topological (red) and topological (blue) L1 triggers. Both types make a selection on the transverse momentum of the two muons, both the topological trigger also considers their angular distance and invariant mass. This leads us to a reduction of 4 in the trigger rate, with only a loss of 20 % in signal efficiency [5]

a Simulation

We simulated the following signal and background events in MC:

Decay type	Decay formula	DSID	$\mathcal{B}$
Signal	$B^0 \to K^{*0}ee$	300590	1.03×10^{-6}
	$\overline{B^0} \to \overline{K^{*0}}ee$	300591	
	$B^0 \to K^{*0}J/\psi(ee)$	300592	1.27×10^{-3}
	$\overline{B^0} \to \overline{K^{*0}}J/\psi(ee)$	300593	
Background	$B^+ \to \pi^+ J/\psi(ee)$	300718	3.92×10^{-5}
	$B^- \to \pi^- J/\psi(ee)$	300719	
	$B^+ \to K^+\pi^0(ee\gamma)$	300722	1.29×10^{-5}
	$B^- \to K^-\pi^0(ee\gamma)$	300723	
	$B^+ \to \pi^+\pi^0(ee\gamma)$	300724	5.5×10^{-6}
	$B^- \to \pi^-\pi^0(ee\gamma)$	300725	
	$B^+ \to \pi^+\eta(ee\gamma)$	300726	4.02×10^{-6}
	$B^- \to \pi^-\eta(ee\gamma)$	300727	
	$B^+ \to K^+\eta(ee\gamma)$	300730	2.4×10^{-6}
	$B^- \to K^-\eta(ee\gamma)$	300731	
	$B^0 \to K^+\pi^- J/\psi(ee)$	300734	1.15×10^{-3}
	$\overline{B^0} \to K^-\pi^+ J/\psi(ee)$	300735	
	$B^0 \to K^+\pi^-\psi(2S)(ee)$	300738	5.8×10^{-4}
	$\overline{B^0} \to K^-\pi^+\psi(2S)(ee)$	300739	
	$B^0 \to K^+\pi^-\pi^0(ee\gamma)$	300742	3.78×10^{-5}
	$\overline{B^0} \to K^-\pi^+\pi^0(ee\gamma)$	300743	
	$B^0 \to K^{*0}\eta(ee\gamma)$	300744	1.59×10^{-5}
	$\overline{B^0} \to \overline{K^{*0}}\eta(ee\gamma)$	300745	
	$B^0 \to K^{*0}\pi^0(ee\gamma)$	300748	3.3×10^{-6}
	$\overline{B^0} \to \overline{K^{*0}}\pi^0(ee\gamma)$	300749	

Table 1: Decays simulated in Monte Carlo. DSID stands for "Dataset Identification", and they are internal numbers assigned to the simulated interaction's datasets. The branching ratios were obtained from [38].

On top of simulating the decays, the detection of the resulting particles and consequent measured values were also simulated. Many times, there were different "candidates" among the results for reconstructing the original particles, and they are all registered alongside their results in the simulation files.
The signal decays include the direct (non-resonant) decay and the decay through the J/ψ resonance, which is the control decay, as mentioned in chapter 1 (the antiparticle versions are also included, of course). Although there were available simulation results for the $\psi(2S)$ resonance, they didn't include information of which candidate was the true candidate, so they weren't used.
The chosen background decays are ones whose final decay results can be mistaken for results of the signal decays. These final results will include a pair of electrons and $K^\pm$ and/or $\pi^\pm$ particles that can be mistaken for the $K\pi$ pair resulting from the K^{*0} decay. It should be noted that these are only some of the possible decays that might serve as background for the decay, and they were chosen based on their frequency and branching

ratios. Ideally, more could have been chosen, and the group I have worked with have already acquired simulations of further background progress, for example occuring from decays of the Λ_b or B_s^0 meson.

Below we see the distribution of the presupposed invariant B mass, invariant mass of the 2 electrons, and invariant mass of the pressuposed $K\pi$ pair resulting from the decay of the K^{*0} meson. These distributions were created using ROOT [1] version 6.24/04, through which the resulting .root files containing information from the simulations were accessed. For each signal file, 23000 events were picked, and for each the candidate flagged as "true" was chosen for the distribution. For each background file, 3500 events were chosen, and for each a random candidate was chosen, since they're all false. If an event contained no candidates or,for signal events, contained either zero or multiple true candidates, it was skipped. After all these eliminations, we ended up with information from 61758 signal candidates and 65147 background candidates, and these numbers are large and similar enough to be usable for the LGBM algorithm.

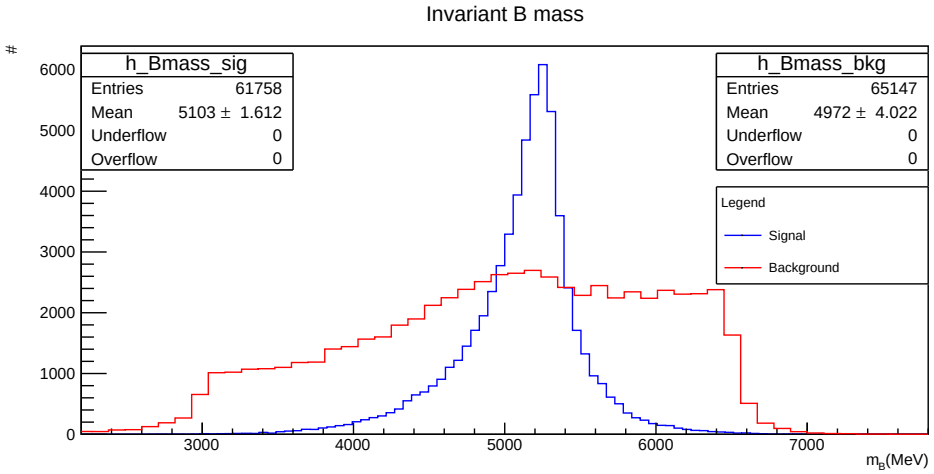


Figure 4.3: Histogram of invariant B mass for the signal (blue) and background (red) candidates

As expected the signal has a peak around the actual value of the mass of B (5280 MeV)[38]. The background is more of a plateau.

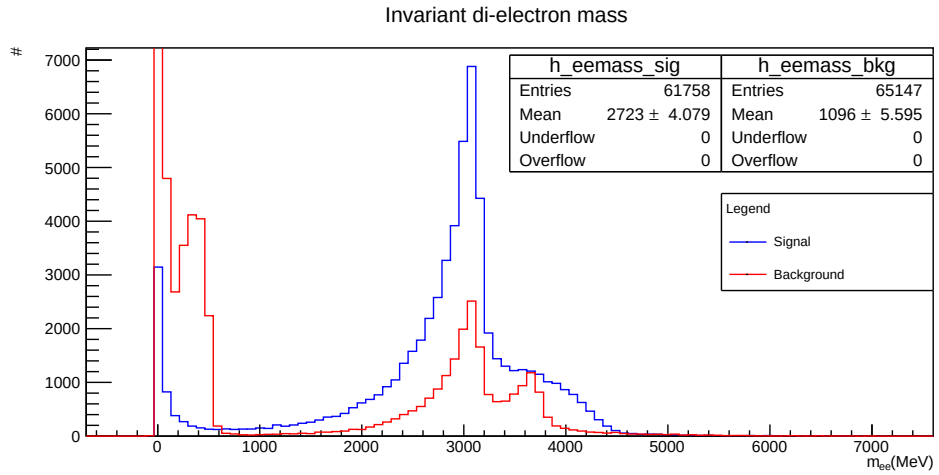


Figure 4.4: Histogram of invariant di-electron mass for the signal (blue) and background (red) candidates

There are peaks around the J/Ψ mass (3097 MeV)[38], as expected. Background also has a peak around the $\Psi(2S)$ mass (3686 MeV [38]), and signal would probably also had had a peak there if the decays $B^0 \rightarrow K^0\psi(2S)(ee)$ and the corresponding anti-particle version had been included in the signal. The peak around 0, especially for signal, can maybe be explained by the way the measurement of the energy and momentum of the electrons was made (consider also that they have very low rest mass).

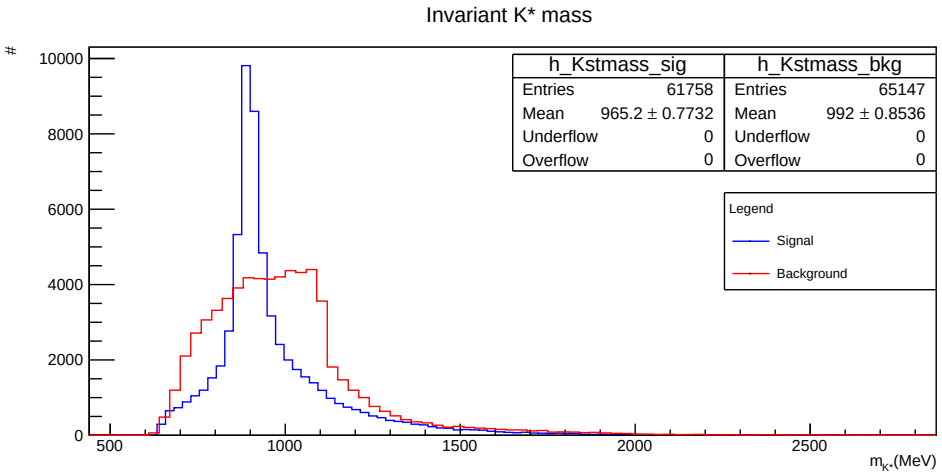


Figure 4.5: Histogram of invariant K^{*0} mass for the signal (blue) and background (red) candidates

The signal is peaking more or less around the K^{*0} mass (895 MeV) [38].

It should be noted that a more realistic depiction of the background would include events from each decay based on the frequency of each decay, instead of uniformly picking from every decay file. This is a problem that can be addressed in future work.

4.3 Pre-selection

The pre-selection, whose goal is to remove events that are too far off the expected values for the signal decays (or even any real processes at all) was decided by the group I was a part of before I even joined, so it uses the same conditions for approval of them, which will be listed below (note that "track" means track and we hypothesise these are the tracks of the K or π particles):

- $\Delta R(ee) > 0.1$. The system of the two electrons coming from the decay should have most of its invariant mass around the invariant mass of the J/ψ. Meanwhile, many background electrons will probably come from the common $\gamma \to ee$ decays happening in the detector, and since the photon has invariant mass of 0, these decays will have a bigger tendency to create electrons going in similar directions,which give smaller values for ΔR, and therefore this condition can help remove many of these background candidates.

- $p_T(e) > 5\ GeV$ and $p_T(trk) > 0.5\ GeV$. In general, having bigger transverse momentum is a good indication that particles came from an interaction, since background particles tend to have most of their momentum in the beam direction.

- $|\eta(e)| < 2.5$ and $|\eta(trk)| < 2.5$. Particles moving in directions wit η above 2.5 will not be detected by any of the tracking detectors (SCT and TRT) in the ID (see fig. 2.5).

- $m(ee) < 7\ GeV$. If the invariant mass of the di-electron system is much bigger than the invariant mass of the J/psi, then we discard the candidate as background. If it's much smaller than that mass, it will probably be discarded anyway by the trigger system or the other conditions.

- $[m(B^0)\ in\ (3000, 6500)\ MeV$ and $m(K^{*0})\ in\ (690, 1110)\ MeV]$ or $[m(\overline{B^0})\ in\ (3000, 6500)\ MeV$ and $m(\overline{K^{*0}})\ in\ (690, 1110)\ MeV]$. We reconstruct the hypothesized K^{*0} mass from the two meson tracks we assume to be a $K^{\pm}\pi^{\mp}$ pair, and we can reconstruct the hypothesized B^0 meson from these and the two electrons. Of course, since we're also considering the $\overline{B^0} \to \overline{K^{*0}}ee$ decay as signal, we can also reconstruct the original mesons assuming they are $\overline{K^{*0}}$ and $\overline{B^0}$. In either case, since particles have the same mass as their antiparticle, we expect the reconstructed mass of $B^0/\overline{B^0}$ and $K^{*0}/\overline{K^{*0}}$ not to be too far from the theoretical values of their rest mass.

- The charges of the two leptons must be opposite, since we expect an electron-positron pair). The same goes for the two mesons, which we assume to be a $K^{\pm}\pi^{\mp}$ pair.

- trk_qual=loose,loose electron. The quality of a track is related to how close it follows its expected helix shape. For both the meson and electron tracks we only expect it to be loosely close to this shape.

These pre-selection conditions significantly reduced the number of background candidates that were accepted. Due to this, we chose to accept all candidates for each event in the background files.

Meanwhile, we also chose to include information from data files, which as mentioned in section 4.2 correspond to measurements taken in 2018, starting in June. Specifically,

we used data from periods K, L, M, O and Q. 2 data.root files were used for each period, giving us a total of 10 data files. The advantage of using these files is the fact that they contain "real information", and are therefor expected to be more realistic. Of course, an overwhelming amount of the events in the data files are background, since our signal decay is only a small fraction of all the interactions happening after the pp collision. We therefore will be considering data as background when training our GBDT.

An important thing to keep in mind when using data is that we want our analysis to be blinded. This means that we want to develop all our algorithms of analysis before we apply them to the real data obtained from the LHC and calculate $R(K^{*0})$, and we in fact don't even want to see the real data. This prevents us from being influenced by the actual data when developping our algorithms and methods, since seeing and working on the actual data before could influence us to fine tune our procedures in order to obtain specific results. So we decided to avoid using and studying the data of events close to the B^0 mass, and we therefore introduced an additional pre-selection condition for the data files:

- $[m(B^0) < 4000\ MeV$ and $m(\overline{B^0}) < 4000\ MeV]$ or $[m(B^0) > 6250\ MeV$ and $m(\overline{B^0}) > 6250\ MeV]$

Requiring that the data events have masses far from the B^0 mass also had the advantage of increasing even further the probability of them being background.

In the end, we used 15000 events from each MC signal file, where the candidates marked "true" were used in each event. From each MC background file we used around 900 candidates (as mentioned we used all candidates from the events, but if we reached 900 and there were still candidates in the current event, we also picked those. This was done because the candidates in each event are sorted according to p_T measurements, and suddenly stopping the candidate collection could introduce bias). And finally, we used 13500 events from each data file, and the candidate was chosen randomly in each.
After collecting the candidates, we ended up with 60000 MC signal events/candidates, 18118 MC background candidates, and 135000 data events/candidates. Although there is a larger amount of background (MC background + data) candidates than signal candidates for the training of the ML algorithm, they are still within the same order of magnitude, so it's no problem. It is also a better reflection of reality, where background interactions vastly outnumber signal ones.

We made graphs comparing the distribution of the reconstructed B_d and $\overline{B_d}$ mass distributions to make sure they were similar enough to justify only looking at the B_d distributions.These graphs were made for signal MC, background MC and data.

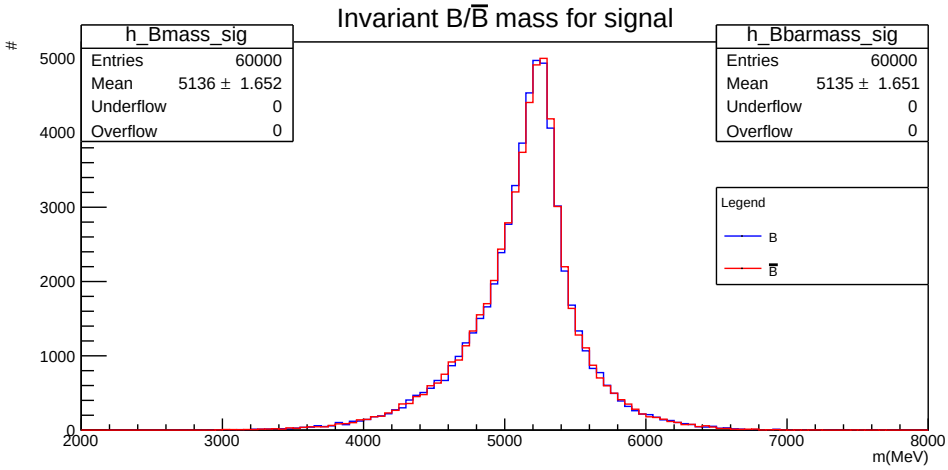


Figure 4.6: Histogram of invariant B^0 and $\overline{B^0}$ mass for the signal candidates (MC)

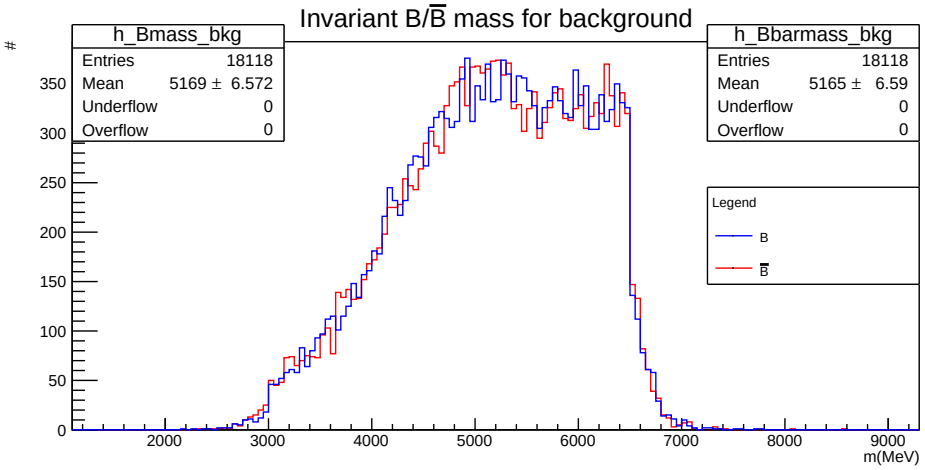


Figure 4.7: Histogram of invariant B_d and $\overline{B_d}$ mass for the background candidates (MC)

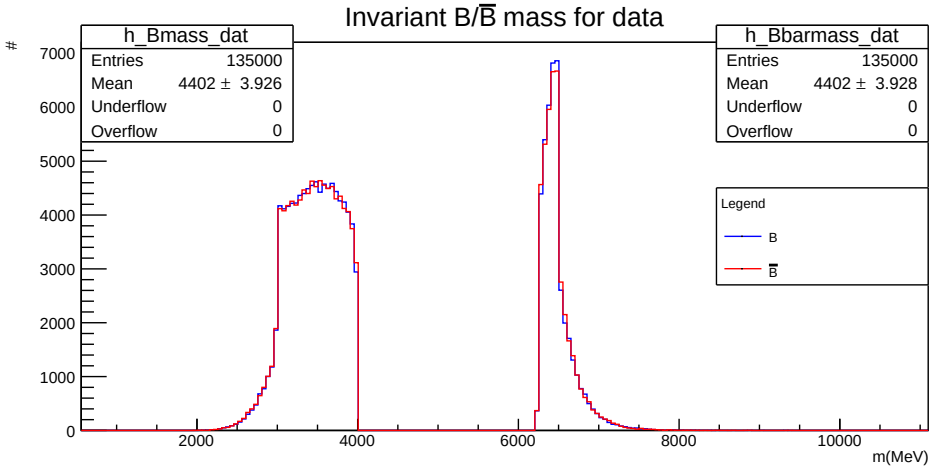


Figure 4.8: Histogram of invariant B_d and $\overline{B_d}$ mass for the data candidates

As we can see, the two mass distributions are quite similar for all types of candidates. There appears to be some variation in the background MC and part of the data, but it doesn't significantly affect the mean, and could maybe be explained by the fact that the creation of all the particles come from the collision of two protons, which are matter, not

antimatter, and this might affect a bit the expected symmetry.

4.4 Training and performance of the GBDT

Now that we had our data for the ML algorithm ready, it was time to apply the ML algorithm to it. We will be using a GBDT through LightGBM.[33]. We ran the algorithms of this package using Python [2] version 3.9.5

Since we wanted to use the GBDT to apply a selection on the distribution of both the reconstructed invariant B^0 mass and the di-electron mass, the selection algorithm, and therefore the features used for its training, had to be independent from both of these. We wanted enough features to allow the algorithm to more easily distinguish between signal and background. Below is a list of the features used:

- m(K^{*0}) and m($\overline{K^{*0}}$).
- η and p_T of both electrons and both mesons.
- Angular distance (ΔR) between the mesons. We can't use the angular distance between the electrons because when combined with their η and p_T it correlates with their mass.
- p(χ^2,$ndof$). In the reconstruction of the initial event that gave us the measured results the program fits models to these measured results.χ^2 here gives a measure of the difference between the measured values and the ones predicted by the fitted model, by summing the squares of these differences divided by the expected values. $ndof$ is the number of degrees of freedom of the mode, which is the difference between the number of experimental points and the number of parameters of the model. χ^2 is expected to follow a certain distribution for a given $ndof$, and p(χ^2,$ndof$) is a p-value that gives us the probability that an observed value of χ^2 would by chance happen to be bigger than the one we actually measured.
- $\tau(B^0)$. This is the calculated mean lifetime of the reconstructed B meson. This can be calculated in different ways, the one we used was calculated by keeping the mass constant and minimizing the transverse distance a_0 to the primary vertex (the point where the pp collision happened).
- Isolation cone values of both electrons, with width $\Delta R = 40$. These values correspond to the sum of the p_T of all the particles within the isolation cone of the electrons.

The values for these features were taken from the MC and data candidates/events previously described. The (MC) signal candidates were given a label of "1", denoting that they are our signal events, while the MC background and the data were given the value 0, indicating they are background. The events/candidates were then shuffled and split into training (60%), validation (20%) and testing (20%) sets. Aside from the features, a few other parameters had to be established for the GBDT:

The first 2 parameters are explained in chapter 3. The last one is not the actual number of estimators used, since that was determined using the early stopping method previously described, as can be seen in the figure below:

Parameter	Value
Learning rate	0.09
Tree depth	Limitless
Maximum number of estimators (i.e. trees)	2000

Table 2: Parameters of the decision trees

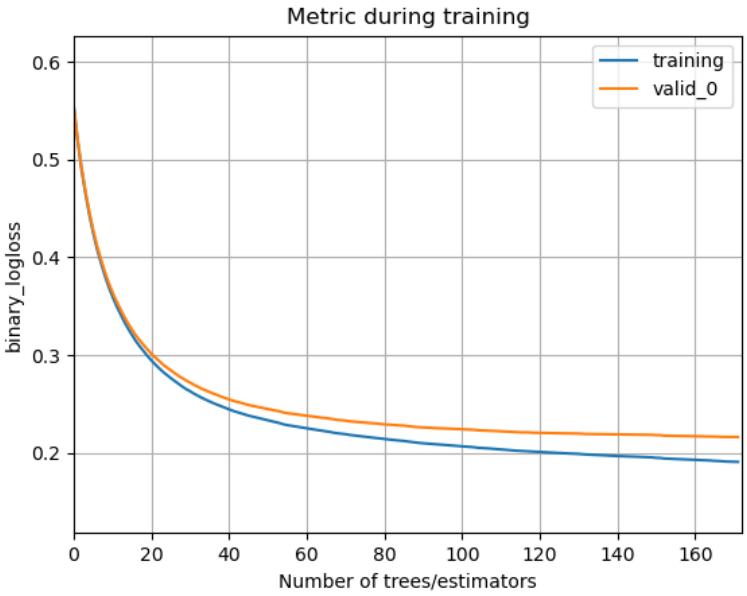


Figure 4.9: Logloss function for training and validation

170 estimators were used based on this early stopping method.
After training the algorithm with the established number of trees, it was applied on the test sample. Each candidate was given a value between 0 and 1 for the predicted probability of being signal.

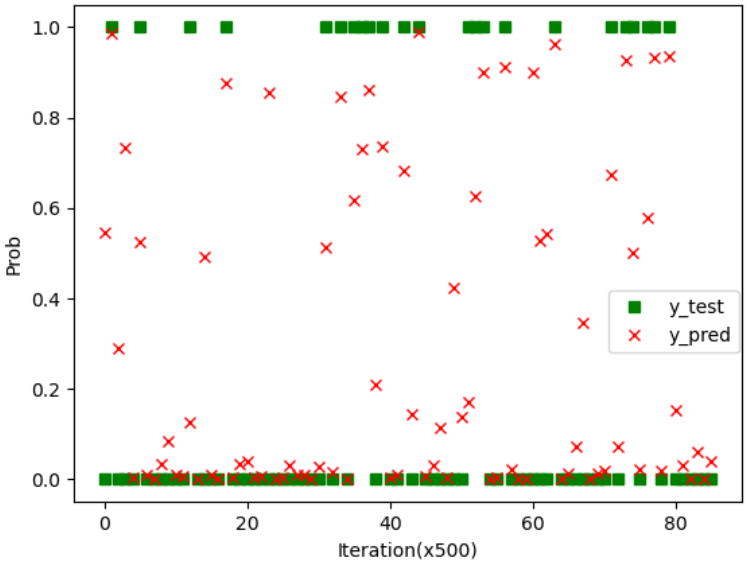


Figure 4.10: Predicted probability of samples belonging to signal, compared to their actual classification (1 means signal, 0 background). Calculated for every 500th testing candidate, for the sake of comprehension

Looking at fig. 4.10, we can see that generally the predicted probabillities tend to stay closer to the actual labels, and there don't appear to be any non-uniformities or strange effects in the distribution of these predictions.

Below we have the distribution of the predicted probability of being signal for the testing signal and background samples.

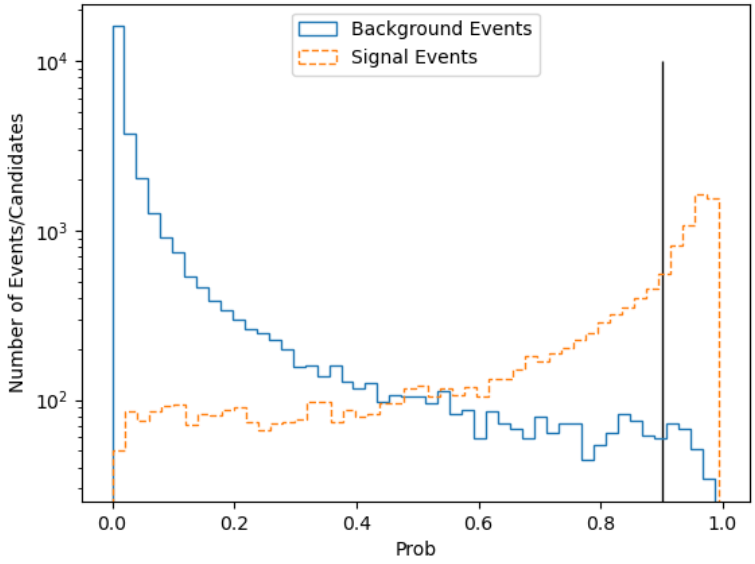


Figure 4.11: Distribution of the predicted probability of being signal for the testing signal and background samples. The vertical line indicates our selectiion cut.

By looking at the figure above, it is clear that it is of course impossible to set a minimum requirement for the value of the assigned probability that completely distinguishes background from signal. However, the peaks of background and signal are still quite distinguishable, and by carefully choosing the minimum assigned probability we can try to find the balance between rejecting as much background and accepting as uch signal as possible, a balance which is graphed in the ROC (receiving operator characteristic) [21] curve below:

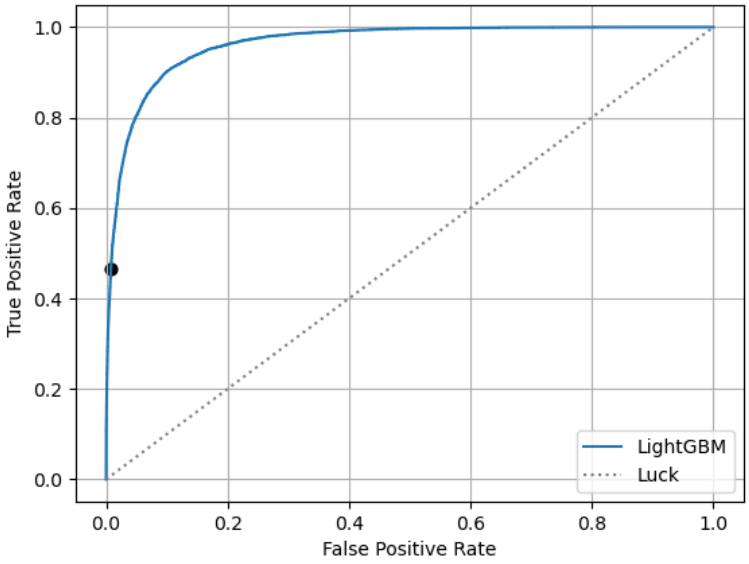


Figure 4.12: ROC curve, which shows the relation between the fraction of accepted signal candidates (True Positive Rate) and the fraction of accepted background candidates (False Positive Rate). The dot indicates our selection cut (0.9)

The curve is quite far from the diagonal, which would correspond to randomly choosing the assignment of signal and background (called luck in the figure). This once again shows that background and signal can be well separated.

Due to the larger amount of background compared to signal, it is best to choose a tight cut, and our selection therefore only accepted candidates with an assigned probability above 0.9. This selection cut is shown in the 4.11 and 4.12 figures.

Using the algorithm we developed, we made a selection on the MC signal, MC background and data invariant mass distributions of the B^0 and di-electron systems by applying the algorithm to their candidates and requiring that they pass the aforementioned cut at 0.9. The results were presented in histograms. We also graphed out the efficiency (ϵ) graph, where for each histogram the efficiency was calculated by dividing the number of selected candidates/events by the total number of candidates/events in each bin (this was only calculated in bins where the second number wasn't 0 of course). The error for the efficiency graphs was calculated by considering no error for the total number of candidates/events, and considering that each event/candidate in each bin follows a binomial distribution with $p = \epsilon$ of being selected, and has therefore a variance $\epsilon(1-\epsilon)$, so the error for all the events/candidates in each bin is given by:

$$s_\epsilon = \sqrt{\frac{\epsilon(1-\epsilon)}{N_t}} \tag{4.1}$$

Where N_t is the total number of candidates/events in each bin. The error for the value of the invariant mass (in the x-axis) was given by the bin width.

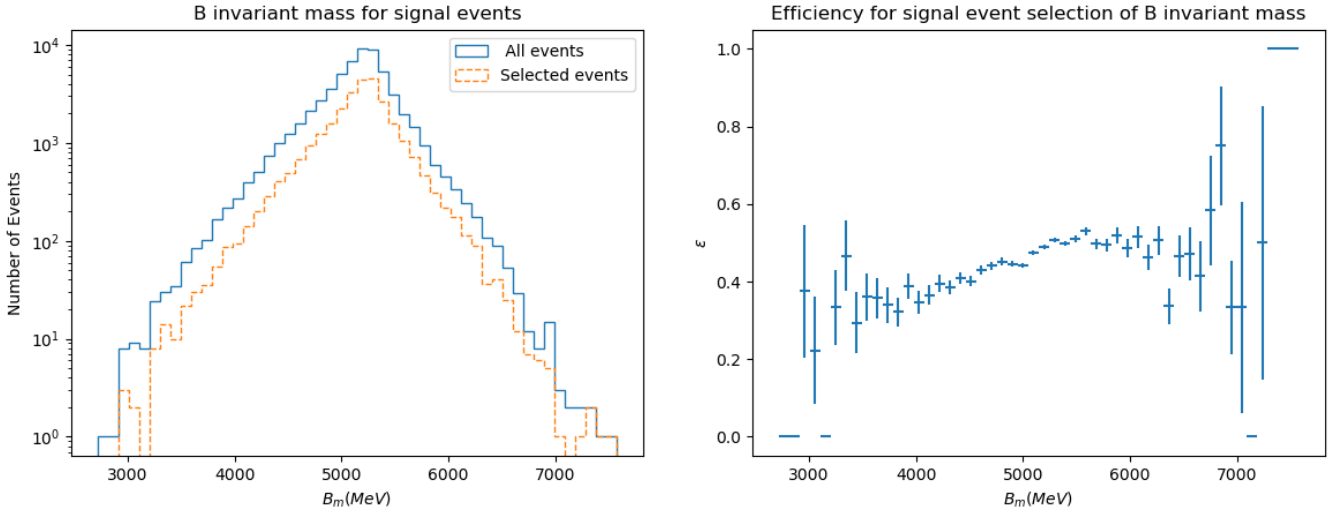


Figure 4.13: B^0 invariant mass histogram for MC signal before and after selection (left), and its respective efficiency graph (right)

Although the selection removes around half of the signal events due to our previously mentioned tight requirements (keep in mind that the y-scale in the histogram is on logarithmic scale), this is fine as long as the original shape of the distribution is reasonably kept, which means that we expect the efficiency graph to remain constant. A visual analysis of the histogram and the efficiency graph shows us that results are more inconsistent (and less precise) towards the sides, but in the range around the B^0 mass (where we see the peak) the results are reasonably what we expected, although the value of the efficiency does slowly increase with the measured mass, but its not so pronounced that it's a major problem.

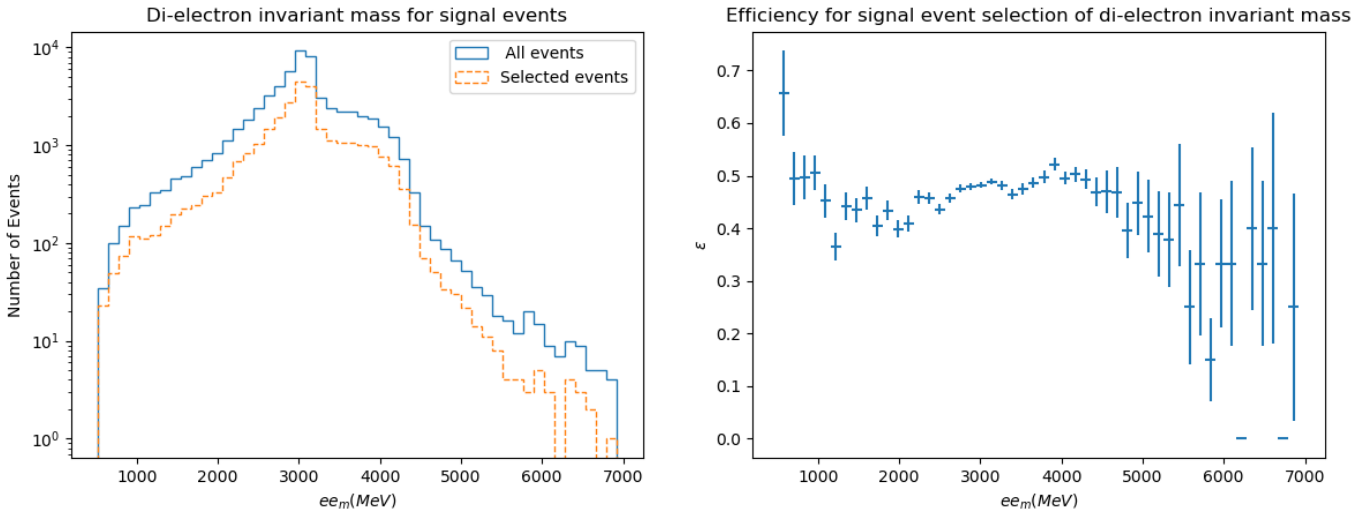


Figure 4.14: Di-electron invariant mass histogram for MC signal before and after selection (left), and its respective efficiency graph (right)

For the signal di-electron mass distribution, we see a similar trend: despite irregular fluctuations at the limits, around the J/ψ mass the efficiency remains more or less constant, with a bit of oscillation and a general tendency to increase.

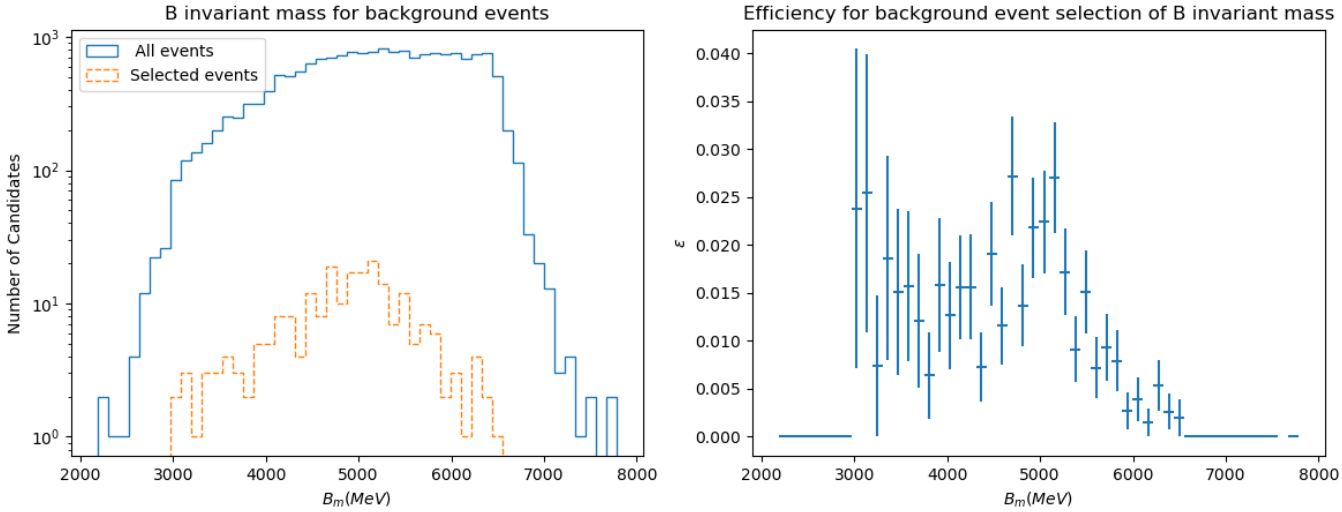


Figure 4.15: B^0 invariant mass histogram for MC background before and after selection (left), and its respective efficiency graph (right)

For the MC background candidates, we can see in the above figures that they were significantly cut (once again, remember that the y-axis in the histogram is logarithmic), so that less than 4% remain in each bin. This is good, since ideally we want as much background to be removed as possible. However, it can be observed in the figures that the distribution of the selected candidates shows a peak around the B^0 mass, even if this peak wasn't present in the original distribution, which is a problem, since it suggests the selection process appears to have some correlation with the B^0 mass distribution after all, and is therefore trying to replicate the peak. Ideally, we would want the only contribution to the peak after the ML selection to be due to signal.

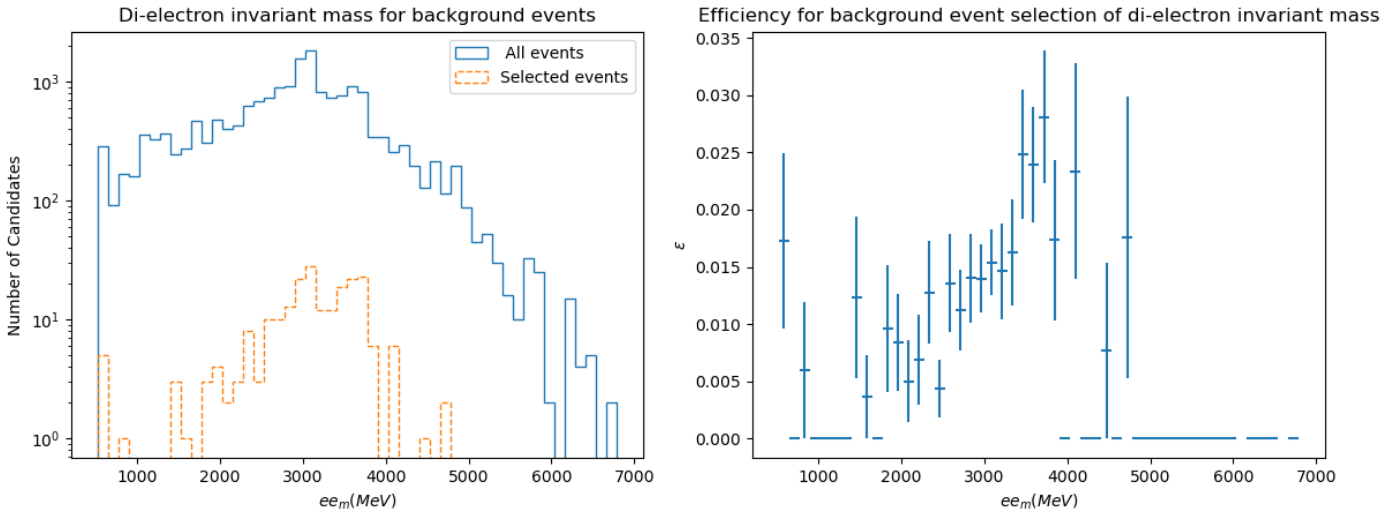


Figure 4.16: Di-electron invariant mass histogram for MC background before and after selection (left), and its respective efficiency graph (right)

Here we can notice in the histogram the selected events emulating the slight peak of the total events around the J/ψ mass. But the efficiency graph appears to have a peak closer to the $\psi(2S)$ mass, although the uncertainties around that point are quite big, so this result might not have much significance.

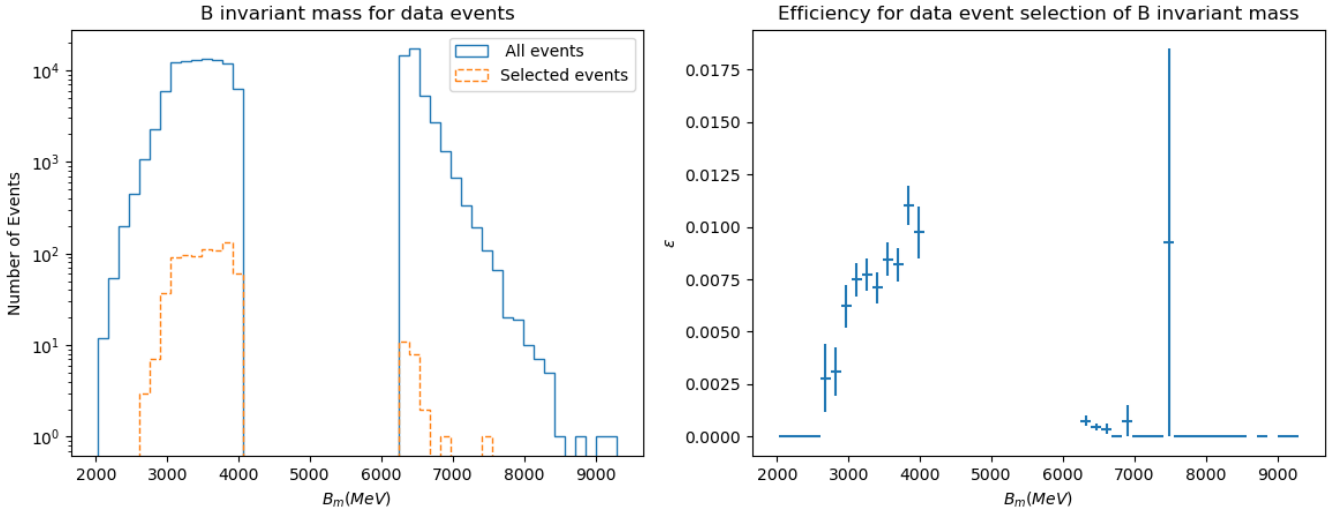


Figure 4.17: B^0 invariant mass histogram for data before and after selection (left), and its respective efficiency graph (right)

The data events were also greatly reduced by the selection, which should be expected, as a lot of them are considered background, and they were used as background for the training of the algorithm. The efficiency appears to be growing towards a peak for the lower values of mass, and the selected distribution is sort of following the distribution of all events, but by design we can't know what is happening around the B^0 mass.

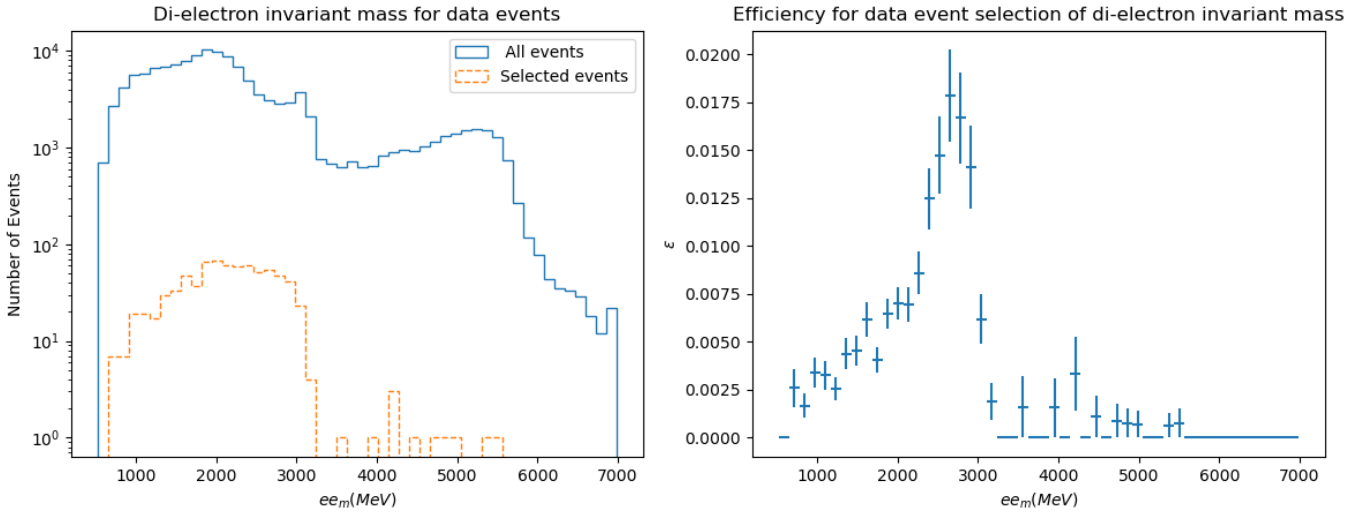


Figure 4.18: Di-electron invariant mass histogram for data before and after selection (left), and its respective efficiency graph (right)

For lower values of mass the selected distribution seems to emulate more or less the total distribution. The efficiency showcases a peak at a value slightlybelow $3000MeV$, and therefore below the J/ψ mass. This peak is bizarre, but it should be noted that, although the exclusion of events with B^0 masses close to the theoretical value is less obvious here than in the B^0 mass distribution, it is still affecting results and we're not sure what conclusions we can then take from them.

4.5 Fitting the signal ditribution

We created a new histogram with the B^0 mass distribution from the MC signal files, this time using 20000 events from each file, leading to a total of 80000 events. The actual histogram of the distribution, as well as all other histograms mentioned here, has been moved to the appendix, while we only keep its counts represented in a scatter plot (in the y-axis of the figures below the (/50) indicates that the number of events represented at each point is counted in bins of width 50 MeV).

In order to gain a better understanding of the expected shape of the signal curve, we used RooFit [3] v. 3.60 to try and to fit it to some common distributions in particle physics (Gaussian and Crystal-Ball), in the 3500-6500 MeV range, in order to cut out the longer tails of the original histogram that might make fitting less successful. The result is visible below:

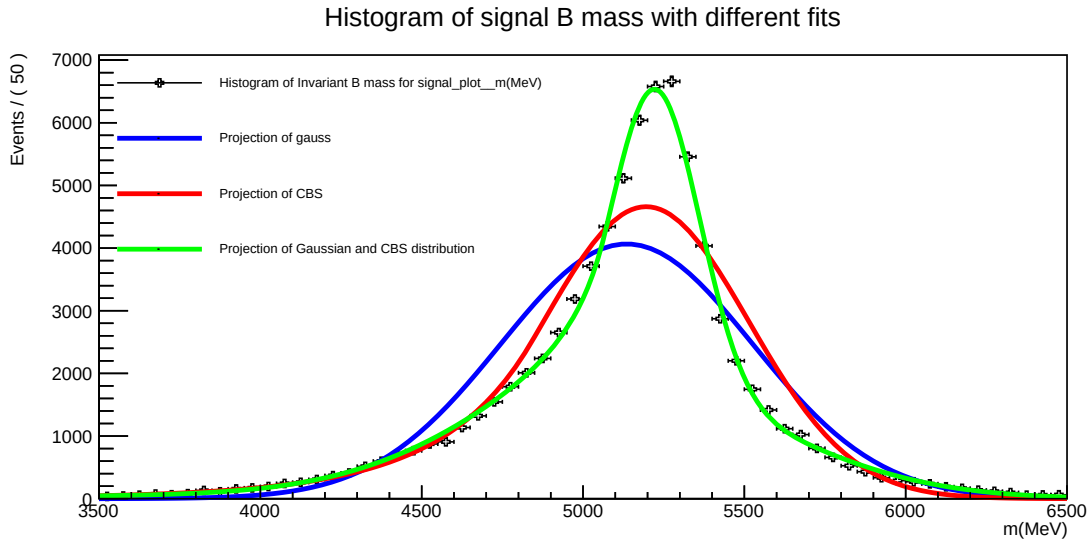


Figure 4.19: The fits of a gaussian curve, a crystall-ball shaped curve and the sum of both to a histogram of the invariant B^0 mass distribution for signal

As can be observed, neither the gaussian or the crystal-ball shaped (CBS) curves were able to properly fit the data points near the peak, but a sum of both was. This indicates that the best curve for fitting the signal distributions might be a mix of these two types.

We decided to separate the events based on the η values of the two electrons resulting from the decay(note that, for the sake of convenience, we use here the word "electron" and the letter "e" to refer both to the electron and the positron resulting from the decay). This is useful because, due to the cylindrical symmetry of the detector, η alone determines what sensors and materials the electron nteracts with. We the η value range into 3 different zones: "low" values corresponding to $\eta < 0.8$, "medium" values corresponding to $0.8 < \eta < 1.52$, and "high" values corresponding to $1.52 < \eta < 2.47$. Electrons going in directions with η above 2.47 will not interact with the tracking system, and $\eta = 1.52$ is where a "crack" region which started at $\eta = 1.37$ ends, where results will have less quality. Meanwhile $\eta = 0.8$ marks the beginning of a zone of higher material budget as can be seen in figure 4.20:

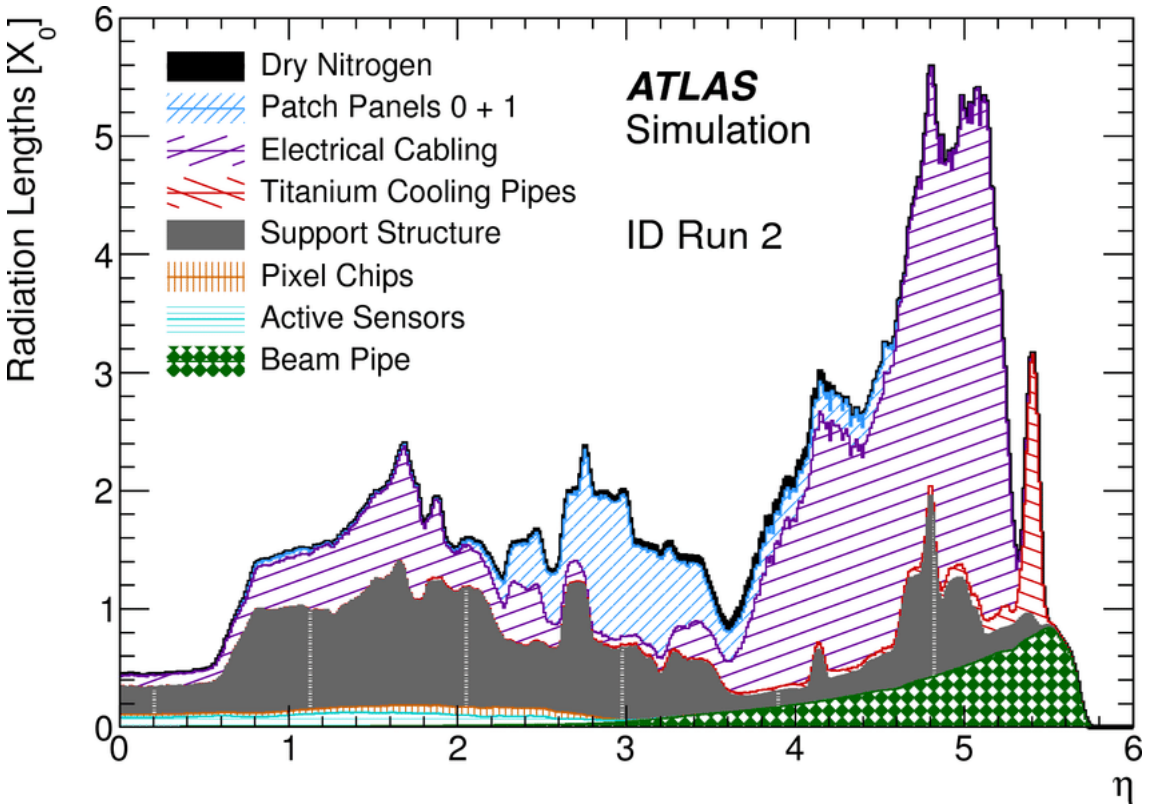


Figure 4.20: Material budget in the Inner Detector during run 2 [35]

We also separated the resonant from the non-resonant events, to see if their distribution varied significantly in the different zones. This also has the advantage of identifying and isolating zones with better and worse distributions, thus preventing the power of the better distribution results to be "diluted" by the worse ones.The results can be seen in the figures below (keep in mind that the labels e_0 and e_1 are just used to emphasize that we are studying two distinct electrons, but they don't refer to a specific one. This means that e.g. if any of the electrons is in the "low" range and the other is in the "medium" one, the event will appear in figure 4.24)

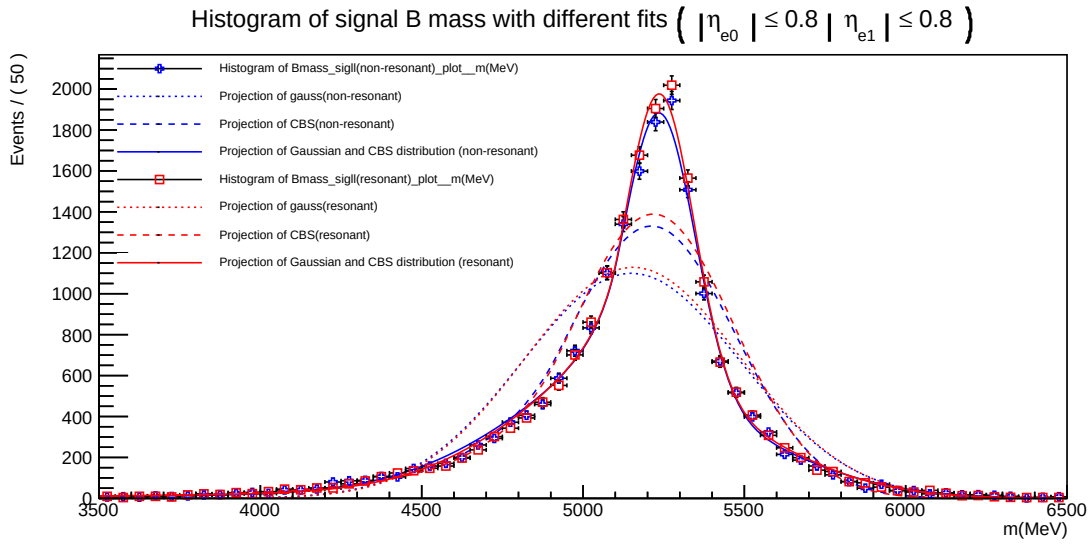


Figure 4.21: Invariant B^0 mass distribution for low values of η for both electrons, with the fits of the different curves to both the resonant and non-resonant histograms

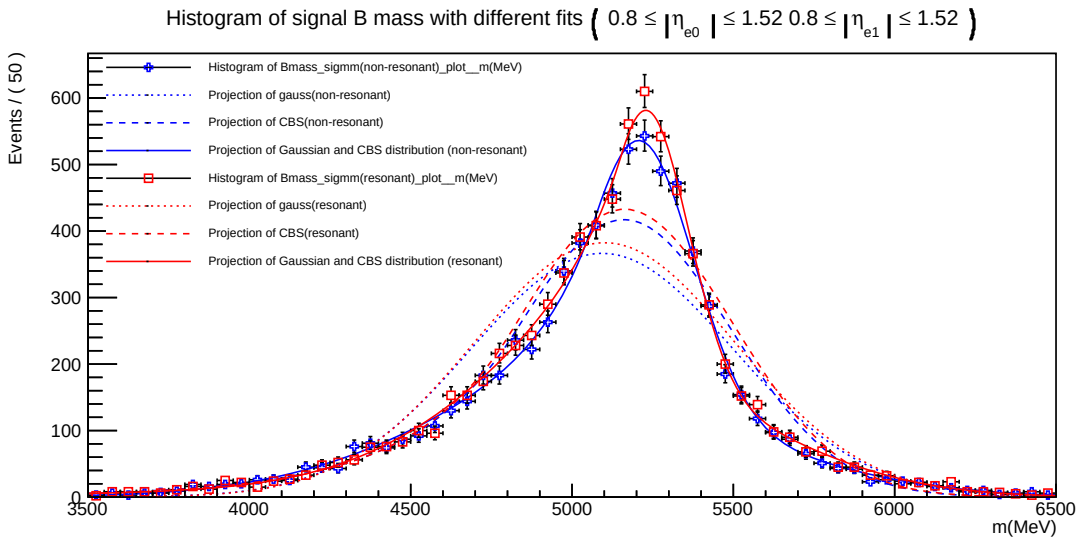


Figure 4.22: Invariant B^0 mass distribution for medium values of η for both electrons, with the fits of the different curves to both the resonant and non-resonant histograms

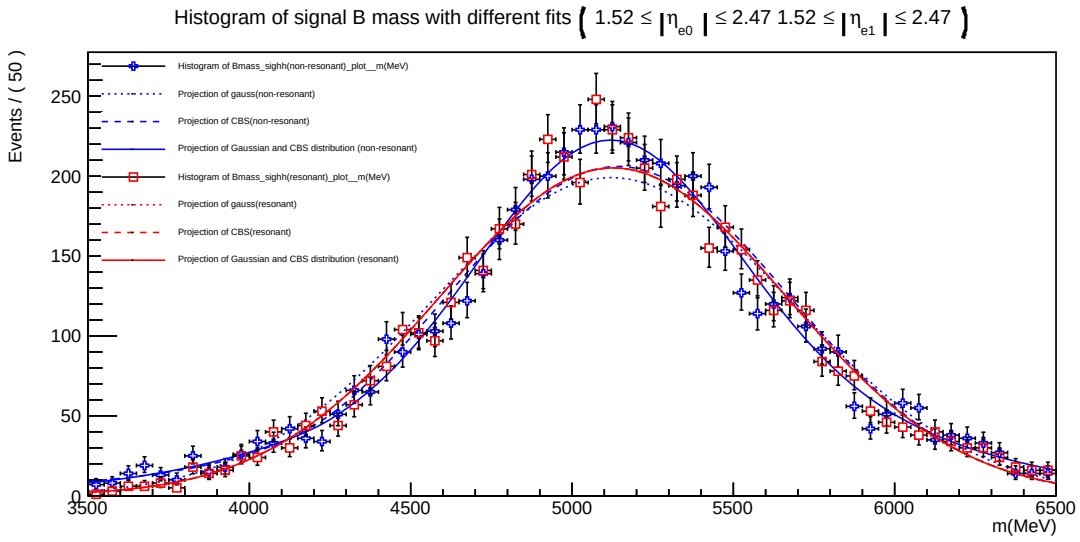


Figure 4.23: Invariant B^0 mass distribution for high values of η for both electrons, with the fits of the different curves to both the resonant and non-resonant histograms

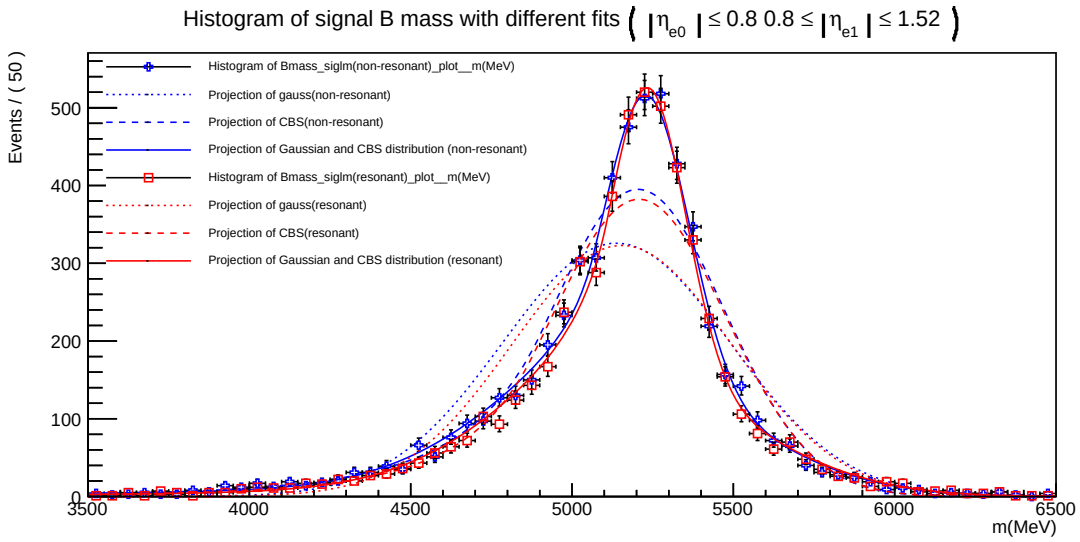


Figure 4.24: Invariant B^0 mass distribution when one electron has a low value of η and the other has a medium one, with the fits of the different curves to both the resonant and non-resonant histograms

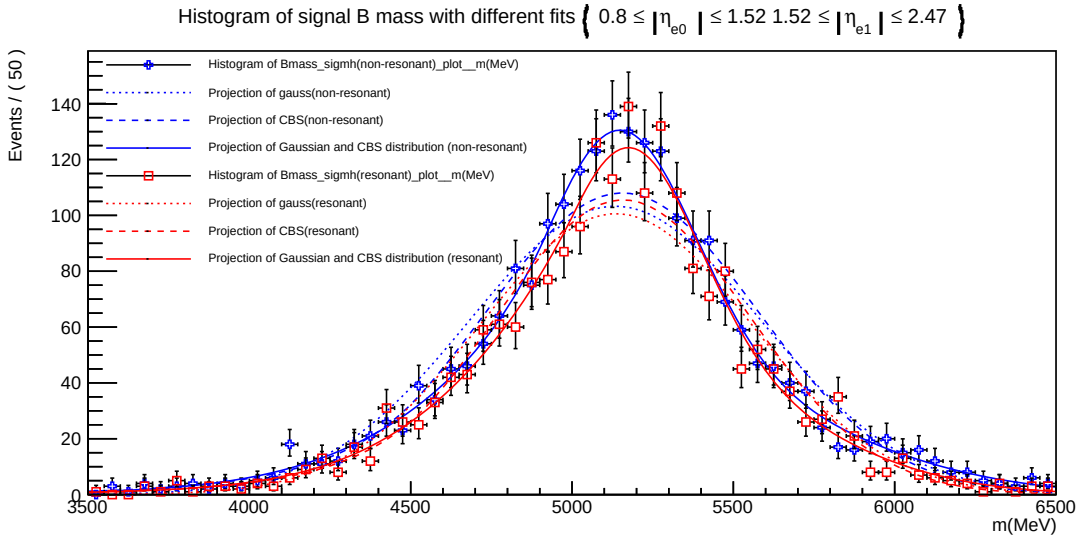


Figure 4.25: Invariant B^0 mass distribution when one electron has a medium value of η and the other has a high one, with the fits of the different curves to both the resonant and non-resonant histograms

The case were one of the electrons stayed in the "low" range and the other in the "high" was not included because only 2 non-resonant events were in those graphs, and they were most likely just a statistical anomaly, indicating that there's almost no event where the difference between the η of the two electrons is that high. This is because such a large $\Delta\eta$ would lead to a large ΔR between the electrons, which would require p_T of one of the electrons to be so low that they would not pass the trigger requirement.

Observing the figures above it is clear that lower values of η for the electron correspond to narrower (and therefore more pure) peaks, and also fitted curves closer to the data points. This might be related to the fact that electrons with higher η will travel through a larger part of the detector, and more material (see fig. 4.20), and therefore subjected to worse tracking and measuring.

The tables below showcase how the 80000 events are shared between the different regions for resonant and non-resonant cases:

e_0 \ e_1	l	m	h
l	18719	5779	2
m	-	7526	2339
h	-	-	5620

Table 3: Number of events in the different η regions of the electrons for non-resonant signal

e_0 \ e_1	l	m	h
l	18993	5517	0
m	-	7793	2129
h	-	-	5549

Table 4: Number of events in the different η regions of the electrons for resonant signal

It is clear that more events are registered for lower values of pseudorapidity for the electrons, which can also contribute to the smaller peaks and better fits mentioned above. A possible reason for part of this imbalance is the fact that particles with lower η have more likelihood to pass the p_T trigger and pre-selection requirements. It is also of note how the distribution of events between the zones is similar for both non-resonant and resonant cases, which indicates that the resonant case is indeed a good choice for calibration in the calculation of the $R_{K^{*0}}$ ratio.

We also decided it might be useful to separate the signal di-electron invariant mass distribution into the resonant and non-resonant parts (the histograms are once again in the appendix)

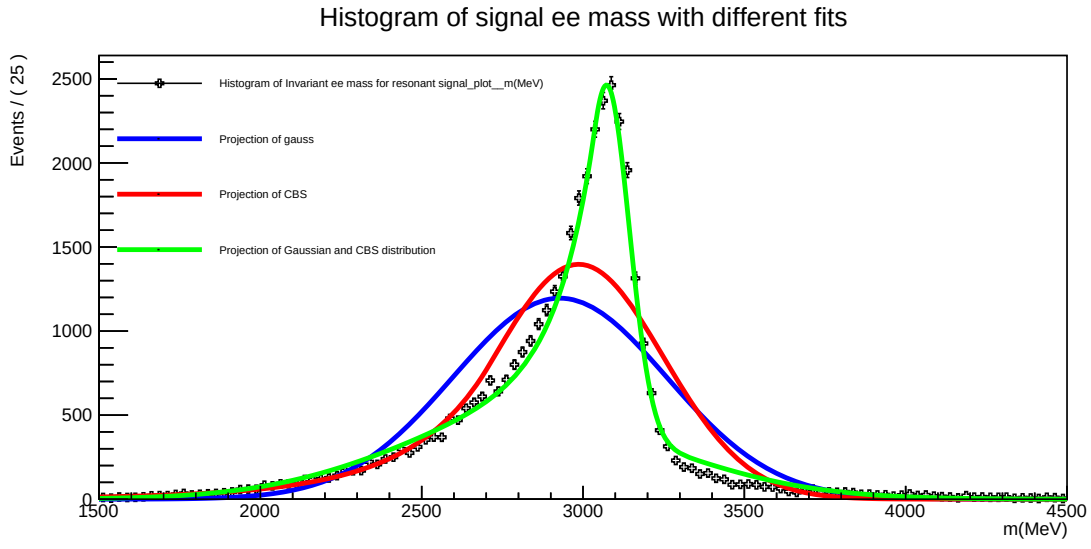


Figure 4.26: The fits of a gaussian curve, a crystall-ball shaped curve and the sum of both to a histogram of the invariant *ee* mass distribution for resonant signal

Once again, the sum of the gaussian and the Crystal-Ball function makes for the best fit at the peak. However, at the peak slopes (around 2700-2900 MeV and especially 3200-3700 MeV) we can observe that the fit deviates a bit from the datapoints. The long tail towards lower values of mass is also an issue, as it will be background for the non-resonant case.

4.6 Systematic uncertainties

Although we have not been able to properly calculate the values for our systematic uncertainties, we have considered the possible sources for it, looked into values obtained by LHCb, and tried to estimate how our values would compare to those. The LHCb values, corresponding to $\Delta R_{K^{*0}}/R_{K^{*0}}$, were obtained from table 4 in [54], where we chose to use the values corrresponding to the hardware electron trigger (L0E) in the [0.045,1.1] GeV/c^2 (low-q^2) and [1.1,6.5] GeV/c^2 (central-q^2) ranges of the di-electron invariant mass. So here are the main sources of systematic uncertainty:

- **Electron trigger and reconstruction efficiency**. Many electrons can be missed by the trigger system, or be lost in the reconstruction process, which will increase the uncertainty of our final results. LHCb calculates the trigger efficiency uncertainty to be 0.1 % for low-q^2 and 0.2 % for central-q^2. We expect our uncertainty to be significantly higher since LHCb is designed to study B-Physics, and has better systems for that, since we don't have huge amounts of $J/\psi \to ee$ samples.

- **Fit model**. Our fit of the signal with the sum of a Crystall-Ball and function distribution can carry uncertainties from the parametrization. LHCb calculates their uncertainty for background and signal (they used the sum of two CB functions for electrons and the sum of a CB and a gaussian function for muons) mass fits to be 1.4 % for low-q^2 and 2.0% for medium-q^2, and we expect similar values.

- **Kinematic reweighing**. The process of reweighing for the kinematic variables can introduce additional uncertainty. The LHCb paper doesn't mention reweighing, but for kinematic selection they estimate an uncertainty of 2.1 % for both low and central q^2. We expect similar values, since this is mostly a theoretical systematic uncertainty.

- **Machine learning input**. Uncertainties from the data inserted into the ML algorithm will contribute to uncertainty of the results. LHCbs paper doesn't involve machine learning, so they don't have uncertainty here. We could study this uncertainty by observing data and (MC) simulated distributions of the variables, and altering the MC distributions (through e.g. shifting or introduction of noise, or "jittering") to observe how that affects the signal selection. Doing this to different degrees and analysing the distributions of the effect on the signal selection can help us obtain the uncertainty.

- **Tracking efficiency**. Like the trigger efficiency, the tracking of the particles in the detector is imperfect and can introduce uncertainty. LHCb didn't include this uncertainty either. In ATLAS we expect this to be at the level of a few percent, as can be verified in e.g. [15].

- **Pile-up reweighing**. Pileup include all the background interactions we are not interested in. LHCb include an uncertainty due to residual background only for central-q^2, which was calculated to be 5 %. We expect similar values (or even lower), since e.g. the systematic uncertainties for pile-up effects of R calculated in [15] were only 1 %.

- **Acceptance**. Acceptance is related to the range of kinematic values a particle must have to be detectable. Since some particles will be excluded based on their

properties, this introduces uncertainty. LHCb didn't go into this, but it is again mostly theoretical, and related to kinematic reweighing.

- **Selection efficiency (data and background)**. As observed, some data and background were still accepted by the algorithm which introduces uncertainty. This will at least partially be eliminated by the ratio. LHCb didn't consider this.

- **Best candidate selection**. For each event, the choice of the best candidate to be used can be tricky, especially when dealing with real data, as a random choice is not the best option. Using ML, I simply chose the candidate with highest BDT score, and this might be different between data and MC, but a deeper study of this has not yet been realized. LHCb does not go into this.

One major advantage of using the double ratio when calculating $R_{K^{*0}}$ is the cancellation of multiple other sources of uncertainty. Considering the very roughly estimated values, one would expect an overall systematic on $R_{K^{*0}}$ of around 10-12 %. The results are summarized below:

Sources of systematic uncertainty	**LHCb (low-q^2)**	**LHCb (central-q^2)**	ATLAS
Electron trigger and reconstr. efficiency	0.1%	0.2 %	++
Fit model (signal and background)	1.4%	2 %	0
Kinematic reweighing	2.1%	2.1 %	0
Machine learning input	-	-	-
Tracking efficiency	-	-	$\approx$ 0.1,1%
Pile-up reweighing	-	5 %	0, low
Acceptance	-	-	-
Selection efficiency (data and bkg.)	-	-	-
Best candidate selection	-	-	-

Table 5: Possible sources of systematic uncertainty, with relevant LHCb results, and estimates of whether our values (ATLAS) should be similar (0), higher (+). much higher (++), or lower(low)

4.7 Outline of Combined Analysis

We will now describe how the work undertaken in the previous sections can be useful towards achieving the final objective, that is, the measurement of R(K^{*0}).

The selection through the GBDT has a simple objective: create an algorithm for cleaning out the real data it will be applied to. In order to get accurate values for the branching ratios of the resonant and non-resonant decays, we need to be able to remove as much of the background as possible, without (completely) removing the real signal decay, a task made harder by the fact that there is overwhelmingly more background interactions than signal ones in real data. Of course, as discussed, this "cleaning" up of the data can never be perfect, but we need to optimize it as much as possible. That is the role of the GBDT.

The fitting of the signal B^0 mass curve is important for understanding how we expect the peak corresponding to the interaction to look in the data, to once again better separate it from the background distribution, which can also be fitted. This way we can know the amount of remaining background and overall signal (resonant and non-resonant). The fitting of the resonant signal ee mass curve and representation of both the resonant and non-resonant distributions is important to separating these two forms of the signal decay, whose branching ratios both need to be considered separately to calculate R(K^{*0}) through the double ratio.

People from the ATLAS group I'm working with have also estimated that, from the approximate integrated luminosity of 30 fb^{-1} from the 2018 data, 28 true signal events and 581 true background events can be extracted from it (non-resonant). This gives a total amount of 609 events, and applying the Poisson approximation (were we assume the statistical error to be given by the square root of the total amount of events), the statistical uncertainty for the values is around 25, which is around 90 % of the total amount of signal events (in actuality, upon fitting, the standard deviation was found to be around 16, corresponding to around 57 % of the total amount of signal events. The Poisson approximation assumes indistinct events, which might not be the case)[29]. We expect an integrated luminosity of around 200 fb^{-1} for the 2022-2024 period of Run-3 of the LHC [22], and this would give us a combined luminosity of around 230 fb^{-1} by the end of 2024. Scaling up, we would then estimate 215 true signal events and 4454 true background events in all this data, to a total of 4669 events. Using the Poisson approximation, we obtain a statistical uncertainty of around 68, which is around 32 % of the total amount of signal events (and assuming the ratio between the Poisson error and the fit error remains the same, we can estimate the fit uncertainty to be around 38, corresponding to around 38 % of the total amount of signal events). This indicates that our precision for the estimates of $R(K^{*0})$ will increase with time, and we might be able to finally get a significant deviation from unity larger than 5 σ, especially when combining our result with the results from $LHCb$, which are independent from ours.

Chapter 5

Conclusion and outlook

The main conclusion to be drawn from this thesis is that the use of Gradient Boosted Decision Trees is a viable way of analysing measurements from the ATLAS experiment in order to isolate the signal events from the background ones, with signal here corresponding to the resonant ($B^0 \rightarrow K^{*0} J/\psi(ee)$) and non-resonant ($B^0 \rightarrow K^{*0} ee$) decays. This can be seen for example in the assigned probability distribution of figure 4.11 and its corresponding ROC curve in figure 4.12. We were able to apply a strict threshold (shown in the previous figures) where it was required that the probability assigned by the Machine Learning algorithm be above 0.9 for an event to be selected, and still only remove around half of the signal events, while the background and data events/candidates (both of which were treated as background for the traning of the algorithm) were signifcantly cut down. The results can be seen in the histograms and efficiency graphs of figures 4.13-4.18. Although there are some observed discrepancies, the algorithm was able to perform its task quite well.

The results from the fitting of the signal events to different curves can also be quite useful, as a good understanding of the expected shape of these signal distributions of the invariant B^{*0} and ee masses is also important to gain a better understanding of what these signal curves are expected to look like, making it easier to separate signal from background, and resonant decays from non-resonant ones. Finally, the consideration of the different sources of systematic uncertainty is something very important, as the quality of our final results will depend both on the statistical and systematic uncertainty.

As previously mentioned, the work on this thesis is part of a bigger project within ATLAS that aims to calculate $R(K^{*0})$. We believe that the work done here can be used and improved upon to achieve this final goal, as the identification of relevant signal events within the multiple interactions in the detector is fundamental when it comes to calculating this ratio precisely. Within this ATLAS group there are other people dealing with the case of the decay into muons ($(B^0 \rightarrow K^{*0} \mu\mu)$), and calculating the relevant branching fractions for it. Members of this group have also used different machine learning algorithms to carry out similar analysis to mine, specifically by using Graph Neural Networks (GNN) instead of GBDT, and while the former is better able to separate signal from background than the latter, it is also much slower (the training of the GBDT algorithm only takes a few seconds, even when processing large amounts of data), indicating that both methods have strengths and weaknesses, and it is therefore advantageous to use different algorithms. Finally, it should be noted that even within the Niels Bohr Institute there are already other students working on Bachelor and Master Thesis that are using and improving upon the work presented in this thesis.

People in the ATLAS group have already created Monte Carlo simulation files for additional decays that weren't considered here, but can be found listed in [28]. Ideally, one should also take events/candidates from each file depending on the corresponding decays frequency and branching ratio, to make the approach more realistic. It is also important to try fitting background mass distributions, and non-resonant ee distrubutions, to get an even better understanding of the differences between the different types of events. And finally, we can try to study in more detail the systematic uncertainty sources and make more accurate estimates of their value.

Hopefully, with the continuous improvement of the methods of analysis of both the electron and muon versions of the decays, and the future influx of new data from ATLAS due in Run 3 of the LHC, we will eventually be able to evaluate $R(K^{*0})$, which when combining with the LHCb findings might open up a whole new understanding of Particle Physics beyond the Standard Model, i.e. New Physics.

Bibliography

[1] `https://root.cern`. Last accessed 27 June 2022.

[2] Python. `https://www.python.org`. Last accessed 4 July 2022.

[3] Roofit. `https://root.cern/manual/roofit/`. Last accessed 7 July 2022.

[4] Aaboud, M., et al. Angular analysis of $B_d^0 \to K^*\mu^+\mu^-$ decays in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *JHEP 10* (2018), 047.

[5] Aad, G., Abbott, B., Abbott, D., Abed Abud, A., Abeling, K., Abhayasinghe, D., Abidi, S., AbouZeid, O., Abraham, N., Abramowicz, H., Abreu, H., Abulaiti, Y., Hoffman, A., Acharya, B., Achkar, B., Adam, L., Bourdarios, C., Adamczyk, L., Adamek, L., and Zwalinski, L. Performance of the atlas level-1 topological trigger in run 2. *The European Physical Journal C 82* (01 2022), 7.

[6] Aaij, R., et al. Angular analysis of the $B^0 \to K^{*0}\mu^+\mu^-$ decay using 3 fb^{-1} of integrated luminosity. *JHEP 02* (2016), 104.

[7] Analythics Vidhya. Tree Based Algorithms: A Complete Tutorial from Scratch (In R & Python). `https://www.analyticsvidhya.com/blog/2016/04/tree-based-algorithms-complete-tutorial-scratch-in-python/#one`. Last accessed 17/07/2022.

[8] ATLAS. Calorimeter. `https://atlas.cern/Discover/Detector/Calorimeter`. Last accessed 17/07/2022.

[9] ATLAS. Muon Spectrometer. `https://atlas.cern/Discover/Detector/Muon-Spectrometer`. Last accessed 17/07/2022.

[10] ATLAS. The ATLAS Detector. `https://atlas.cern/Discover/Detector`. Last accessed 17/07/2022.

[11] ATLAS. The Inner Detector. `https://atlas.cern/Discover/Detector/Inner-Detector`. Last accessed 17/07/2022.

[12] ATLAS. Trigger and Data Acquisition. `https://atlas.cern/Discover/Detector/Trigger-DAQ`. Last accessed 17/07/2022.

[13] ATLAS Collaboration. Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data. *Journal of Instrumentation 14*, 12 (dec 2019), P12006–P12006.

[14] ATLAS Collaboration. Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at $$\sqrt{s} = 13$$ $$\text {TeV}$$. *The European Physical Journal C 79*, 8 (aug 2019).

[15] ATLAS Collaboration. Study of $b_c^+ \to j/\psi d_s^+$ and $b_c^+ \to j/\psi d_s^{*+}$ decays in pp collisions at $\sqrt{s} = 13$ tev with the atlas detector, 2022.

[16] Bifani, S., Descotes-Genon, S., Romero Vidal, A., and Schune, M.-H. Review of Lepton Universality tests in B decays. *J. Phys. G 46*, 2 (2019), 023001.

[17] CERN. The higgs boson. `https://home.web.cern.ch/science/physics/higgs-boson`. Last accessed 28 February 2022.

[18] CERN. LHC faq-the guide. `https://press.cern/sites/default/files/2018-07/CERN-Brochure-2017-002-Eng_0.pdf`. Last accessed 16/07/2022.

[19] CERN. The standard model. `https://home.web.cern.ch/science/physics/standard-model`. Last accessed 28 February 2022.

[20] Dembla, G. Intuition behind Log-loss score. `https://towardsdatascience.com/intuition-behind-log-loss-score-4e0c9979680a`. Last accessed 17/07/2022.

[21] Developers, G. Classification: ROC Curve and AUC. `https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc`. Last accessed 17/07/2022.

[22] Fartoukh, S., Kostoglou, S., Solfaroli Camillocci, M., Arduini, G., Bartosik, H., Bracco, C., Brodzinski, K., Bruce, R., Buffat, X., Calviani, M., Cerutti, F., Efthymiopoulos, I., Goddard, B., Iadarola, G., Karastathis, N., Lechner, A., Metral, E., Mounet, N., Nuiry, F.-X., Papadopoulou, P. S., Papaphilippou, Y., Petersen, B., Persson, T. H. B., Redaelli, S., Rumolo, G., Salvant, B., Sterbini, G., Timko, H., Tomas Garcia, R., and Wenninger, J. LHC Configuration and Operational Scenario for Run 3. Tech. rep., CERN, Geneva, Nov 2021.

[23] Fermilab. Discoveries at fermilab - discovery of the bottom quark. `https://www.fnal.gov/pub/inquiring/physics/discoveries/bottom_quark_pr.html`, 8 1977. Last accessed 22 February 2022.

[24] Fermilab. Physicists find first direct evidence for tau neutrino at fermilab. `https://news.fnal.gov/2000/07/physicists-find-first-direct-evidence-tau-neutrino-fermilab/`, 7 2000. Last accessed 22 February 2022.

[25] Goodfellow, I., Bengio, Y., and Courville, A. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[26] Goswami, D. S. Introduction to Early Stopping: an effective tool to regularize neural nets. `https://towardsdatascience.com/early-stopping-a-cool-strategy-to-regularize-neural-networks-bfdeca6d722e`. Last accessed 17/07/2022.

[27] JabberWok (Wikimedia Commons). A graphic showing the relationship between angle and pseudorapidity. https://commons.wikimedia.org/wiki/File:Pseudorapidity2.png, 5 2007. Last accessed 6 April 2022.

[28] Jakoubek, T. R(K^*) analysis in Run-2. https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/R_KstarRun2?sortcol=0;table=5;up=0#MC_samples. Last accessed 14 July 2022. Only authorized people can access this page.

[29] Jakoubek, T. R(K*) measurement. https://indico.cern.ch/event/1149016/contributions/4821613/attachments/2465932/4228728/jakoubek_rkstar_atlas_week_20220621.pdf. Last accessed 17/07/2022. Internal source, only available for members of ATLAS.

[30] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.-Y. Lightgbm: A highly efficient gradient boosting decision tree. In *Proceedings of the 31st International Conference on Neural Information Processing Systems* (Red Hook, NY, USA, 2017), NIPS'17, Curran Associates Inc., p. 3149–3157.

[31] Lella, L. D., and Rubbia, C. The discovery of the w and z particles. *Advanced Series on Directions in High Energy Physics— 60 Years of CERN Experiments and Discoveries* (9 2015), 137–163.

[32] LightGBM. Lightgbm-features. https://lightgbm.readthedocs.io/en/latest/Features.html. Last accessed 15 March 2022.

[33] LightGBM. Welcome to lightgbm's documentation! https://lightgbm.readthedocs.io/en/latest/index.html. Last accessed 4 July 2022.

[34] Lopienska, E. The cern accelerator complex. https://cds.cern.ch/images/CERN-GRAPHICS-2022-001-1, 2 2022. Last accessed 9 March 2022.

[35] Meng, L. Atlas itk pixel detector overview, 05 2021.

[36] MissMJ, Cush (Wikimedia Commons). Standard model of elementary particles. https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg, 9 2019. Last accessed 22 February 2022.

[37] Rao, A. The higgs boson: Revealing nature's secrets. https://home.cern/news/series/lhc-physics-ten/higgs-boson-revealing-natures-secrets, 7 2020. Last accessed 22 February 2022.

[38] R.L. Workman *et al.* [Particle Data Group]. Review of Particle Physics. https://pdg.lbl.gov. to be published (2022).

[39] Schwartz, M. D. *Quantum field theory and the standard model*. Cambridge University Press, Cambridge, 2014.

[40] Shah, T. About Train, Validation and Test Sets in Machine Learning. https://towardsdatascience.com/train-validation-and-test-sets-72cb40cba9e7. Last accessed 17/07/2022.

[41] Shankar, R. *Principles of quantum mechanics*, 2nd ed. ed. Springer, New York, 1994.

[42] Stachel, J. Detectors in Nuclear and Particle Physics. https://www.physi.uni-heidelberg.de/~fschney/detektoren/detector8.pdf. Last accessed 16/07/2022.

[43] Stella, B. R., and Meyer, H.-J. Υ (9.46 gev) and the gluon discovery (a critical recollection of pluto results). *The European Physical Journal H 36* (9 2011), 203–243.

[44] Symmetry-dimensions of particle Physics. LHC circulates first beams of 2009. https://www.symmetrymagazine.org/breaking/2009/11/20/lhc-circulates-first-beams-of-2009. Last accessed 16/07/2022.

[45] Söding, P. On the discovery of the gluon. *The European Physical Journal H 35* (7 2010), 3–28.

[46] The ATLAS Collaboration. The ATLAS experiment at the CERN large hadron collider. *Journal of Instrumentation 3*, 08 (aug 2008), S08003–S08003.

[47] The ATLAS Collaboration. Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc. *Physics Letters B 716* (9 2012), 1–29.

[48] The ATLAS Collaboration. Performance of the ATLAS trigger system in 2015. *The European Physical Journal C 77*, 5 (may 2017).

[49] The ATLAS Collaboration. Study of the material of the ATLAS inner detector for run 2 of the LHC. *Journal of Instrumentation 12*, 12 (dec 2017), P12009–P12009.

[50] The ATLAS Collaboration. Public atlas luminosity results for run-2 of the lhc. https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2, 2018. Last accessed 9 June 2022.

[51] The CDF Collaboration. Observation of top quark production in $\overline{p}$p collisions with the collider detector at fermilab. *Physical Review Letters 74* (4 1995), 2626–2631.

[52] The CMS Collaboration. Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc. *Physics Letters B 716* (9 2012), 30–61.

[53] The DØ collaboration. Observation of the top quark. *Physical Review Letters 74* (4 1995), 2632–2637.

[54] The LHCb collaboration. Test of lepton universality with $b^0ßk^{*0}l^+l^-decays$. *Journal of High Energy Physics 2017* (8 2017), 55.

[55] The LHCb Collaboration. Test of lepton universality in beauty-quark decays, 2021.

[56] The LHCb Collaboration. Tests of lepton universality using $b^0 \to k_s^0\ell^+\ell^-$ and $b^+ \to k^{*+}\ell^+\ell^-$ decays, 2021.

[57] Vogel, A. ATLAS Transition Radiation Tracker (TRT): Straw Tube Gaseous Detectors at High Rates.

Appendix A

Histograms of B^0 signal mass distributions from section 4.5

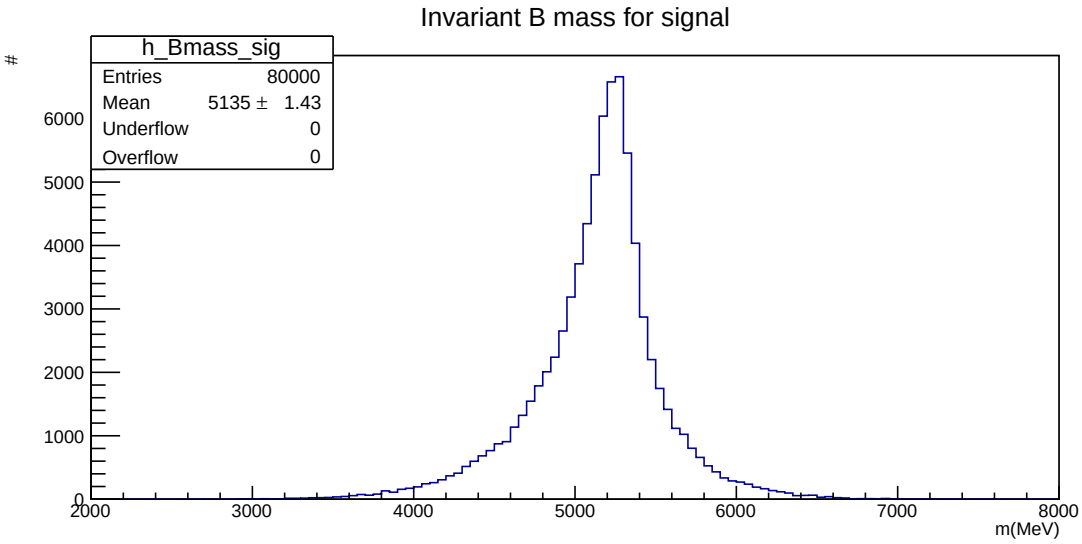


Figure A.1: Histogram of the invariant B mass distribution for signal

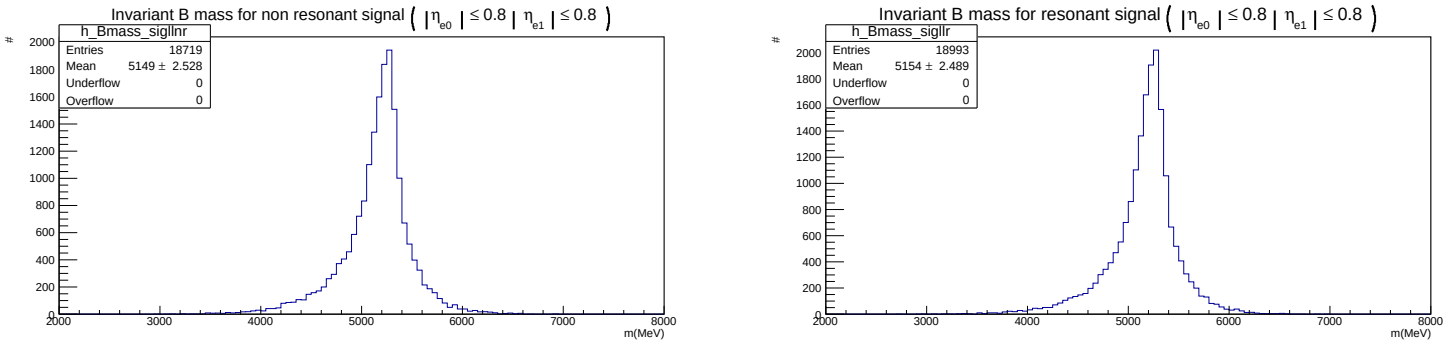
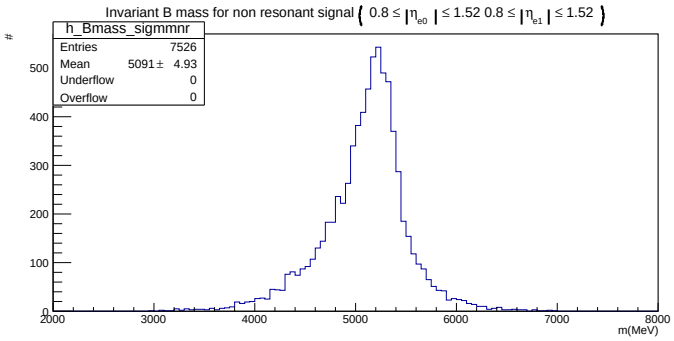


Figure A.2: Invariant B mass distribution for low values of η for both electrons. We have the histogram for non-resonant signal (left) and resonant signal (right)



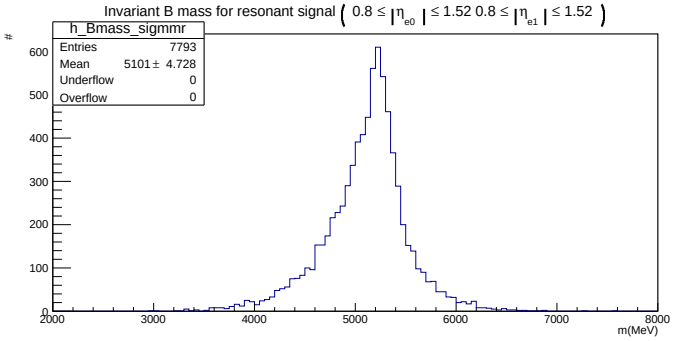
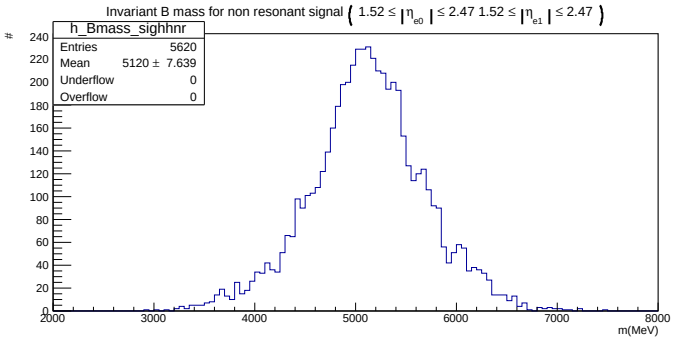


Figure A.3: Invariant B mass distribution for medium values of η for both electrons. We have the histogram for non-resonant signal (left) and resonant signal (right)



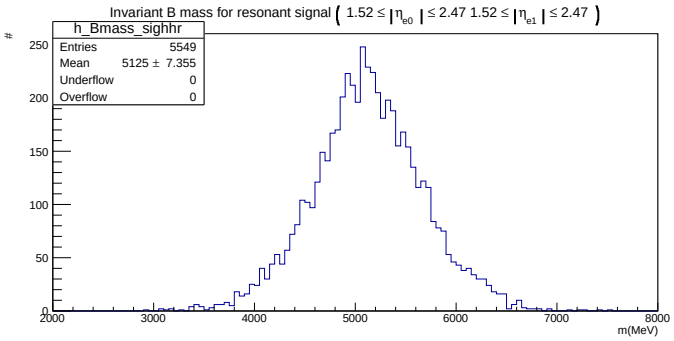
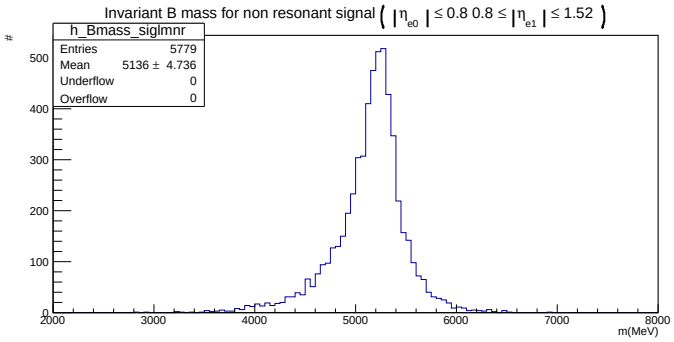


Figure A.4: Invariant B mass distribution for high values of η for both electrons. We have the histogram for non-resonant signal (left) and resonant signal (right)



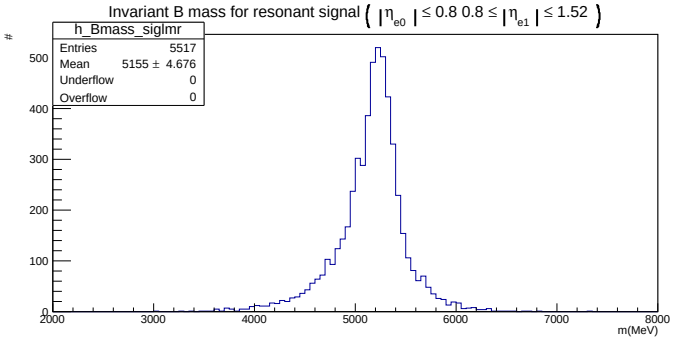
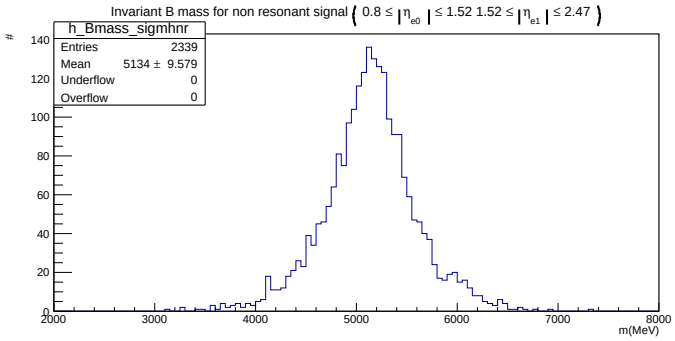


Figure A.5: Invariant B mass distribution when one electron has a low value of η and the other has a medium one. We have the histogram for non-resonant signal (left) and resonant signal (right)



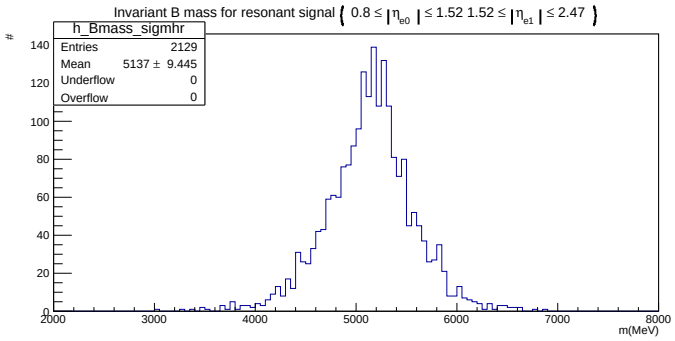
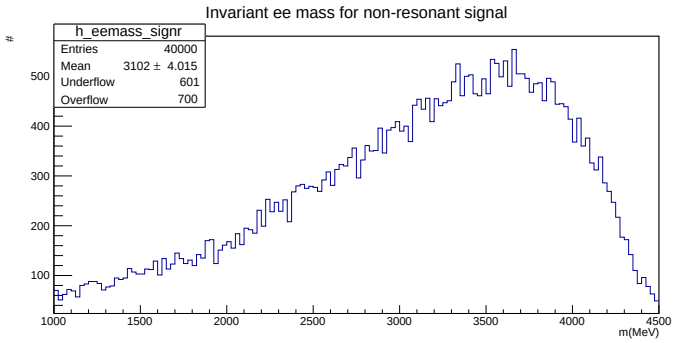


Figure A.6: Invariant B mass distribution when one electron has a medium value of η and the other has a high one. We have the histogram for non-resonant signal (left) and resonant signal (right)



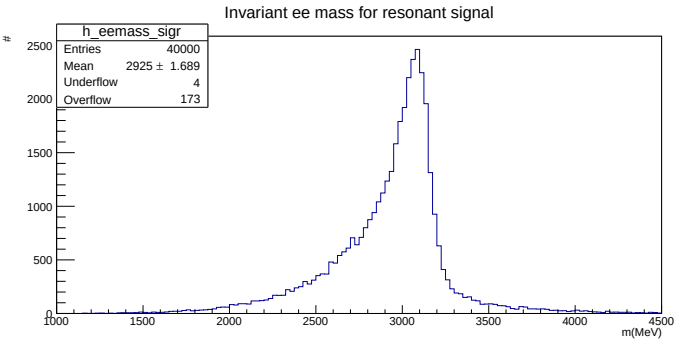


Figure A.7: Invariant ee mass distributions. We have the histogram for non-resonant signal (left) and resonant signal (right)