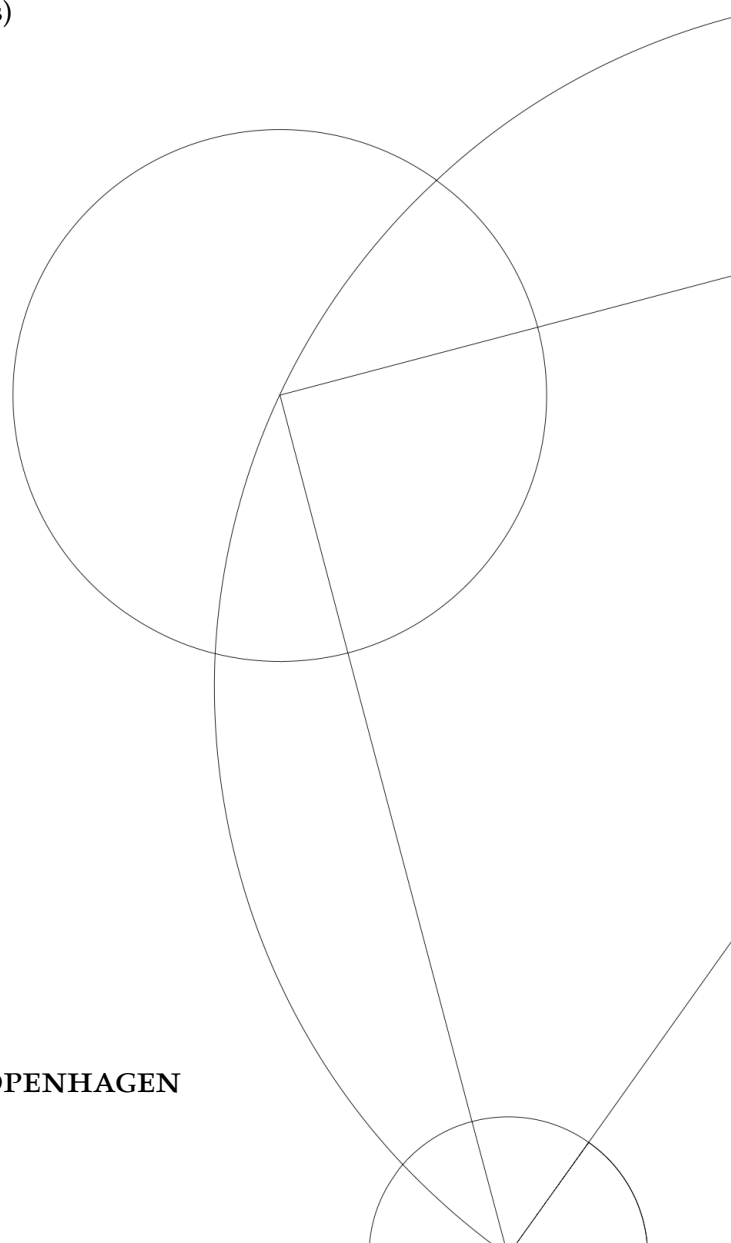
Improving Z and Higgs selections in ATLAS using Machine Learning

Sara Elisabeth Dahl Pinholt

Master's Thesis (Cand.scient in Computational Physics)

May 14, 2021

Supervised by
Troels Petersen



University of Copenhagen

MSc Thesis,
Niels Bohr Institute, University of Copenhagen

Thesis submitted 14th of May 2021 for the completion of the degree
Master of Science (MSc) in Computational Physics

LaTeXtemplate: A Tufte-Style Book
The color map used in the thesis is inspired by color palettes of Wes Anderson films.

First printing, May 2021

Abstract

This thesis presents a novel approach to search for the $H \to Z\gamma$ decay. Previously, only cut-based methods has been explored [1]. Here, a machine learning-based approach is introduced, using tree-based methods in the Higgs boson search.

The analysis presented is based on the decay products of the $H \to Z\gamma$ decay; muons, electrons—their combination into a Z boson—and photons. The $Z \to ll\gamma$ decay is utilized as a control channel, as it shares decay channels with the Higgs boson, allowing for testing in data from the ATLAS detector.

All Z and Higgs models are evaluated against the selection in [1] on Monte Carlo data (MC) and real data. The $Z \to \mu\mu\gamma$ model increases the signal with $40 - 106\%$ ($17 - 30\%$) in MC (data) while the $Z \to ee\gamma$ model shows increase in signal of $12 - 44\%$ ($6 - 48\%$) in MC (data) for same amount of background as the ATLAS selection. The Higgs trained models are evaluated in MC only, where an increase of $44 - 53\%$ ($29 - 32\%$) is seen for the $H \to Z(\to ll)\gamma$ decay for muons (electrons). As a large decrease is seen for the MC-trained $Z \to \mu\mu\gamma$ model when applied to data, expected increase for the $H \to Z(\to \mu\mu)\gamma$ decay is 16% while for the $H \to Z(\to ee)\gamma$ model, 22% more signal is expected compared with the selection in [1].

In summary, the results of this thesis strongly suggests that adoption of a tree-based approach would drastically improve the signal selection of the $H \to Z\gamma$ decay compared to previous cut-based methods.

Disclaimer

It is important to note, that this thesis succeeds several Master's Theses on the same subject, all with different foci. Previously, emphasis was put on electrons and photons whereas my work primarily investigates muons. Furthermore, while previous work relied mainly on MC data, I have implemented the evaluation in data for both electrons, muons and photons. Since changes were necessary to extend the analysis to data, I have trained new models for electrons and photons, based on previous work on MC data.

The thesis concludes my Masters degree in Computational Physics. Prior to this project, I had very little experience with Particle Physics. It has been an interesting topic to learn, and it has served as a great subject to apply computational methods.

Acknowledgments

First and foremost, I would like to thank my supervisor, Troels Petersen—for his enthusiasm and brilliant ideas on what to pursue next. Secondly, I would like to thank Christian Michelsen and Malte Algren for great office companionship (the times where we were not trapped at home!) and thanks to new and previous group members such as Mads Storr-Hansen, Rasmus Ørsøe, Helle Leerberg, and Daniel Nielsen.

Contents

Introduction

In 1964, two independent papers, proposing a theory to explain mass generation, where published [2, 3]. The papers gave name to the Brout-Englert-Higgs mechanism (later shortened to the Higgs mechanism). The mechanism required a spin-less boson to exist and the search of the *Higgs boson* thus began in 1975 [4], even though no one really knew exactly what to look for.

In the late 1990's and early 2000's, searches ruled out masses above 200 GeV [5] and below 114 GeV [6]. On September 30th 2008, the Large Hadron Collider (LHC) was commissioned. The accelerator was CERN's flagship accelerator, built to—among other things—prove whether the Higgs boson existed. In 2012, two LHC experiments, A Toroidal LHC ApparatuS (ATLAS) [7] and the Compact Muon Solenoid (CMS) [8], announced results consistent with the Higgs particle at $m_H = 125\,\text{GeV}$.

The discovery-channels in 2012 were $H \to ZZ^{(*)} \to 4l$, $H \to \gamma\gamma$ and $H \to WW^* \to e\nu\mu\nu$. Current searches [1] include the $H \to Z(\to ll)\gamma$ decay, which is explored in this thesis. The well-investigated $Z \to ll\gamma$ decay is used as a control channel, where models are tested in data. Current and previous searches of the $H \to Z(\to ll)\gamma$ decay uses a cut-based method [1, 9–12] to separate signal from background. This thesis will present a machine learning-based approach. The procedure presented is split in several parts where each part is optimized and evaluated against current methods in ATLAS.

Like the cut-based ATLAS methods, the ML based approach focuses on the different decay products separately. First, models for single leptons (e, μ) and photons (γ), are created. The models helps us identify and select well-isolated particles. These models are referred to as **Pid** and **Iso** models. Secondly, the Z boson decay into two leptons, $Z \to \mu\mu$ and $Z \to ee$, is studied and models are created focusing on the combination of the leptons, the **Zmm** and **Zee** (**Zll**) models. These takes the Pid and Iso models for the muons and electrons as input, together with track variables. Lastly, models are created for the $Z \to ll\gamma$ and $H \to Z(\to ll)\gamma$ decays, **Zllg** and **Hllg**. Here, the Zll models are combined with the Pid and Iso models for the photon, creating a model for the triplet of particles. The results is evaluated against [1] aiming to get a clearer signal at the same level of background. [1] focuses on the $H \to Z(\to ll)\gamma$ decay, but the cuts

introduced are applied to Z decays in thesis as well.

The aim of the thesis is to create models independent of mass and kinematics. This is achieved by creating "bottlenecks" in the models, i.e. not putting all variables into one single model, but rather creating several smaller ones serving as input to each other. This prevents the final model to focus on one specific variable, but instead forces the model to see the event as a whole.

The models are mainly trained on Monte Carlo generated data, but data-trained models are created for muons and electron in Pid, Iso and Zll. The Pid and Iso models are evaluated on the test set, i.e. the same type of data as it was trained upon and the Z models (both Zll and Zllg) are evaluated on both data and MC. The $H \rightarrow Z(\rightarrow ll)\gamma$ models are evaluated in MC only, but from the performance of the Zllg models in data, performance in Higgs data will be estimated.

1 The Standard Model of Particle Physics

The following sections dives into the current theory of Particle Physics, the Standard Model. It describes the basic building blocks of matter and interactions. The chapter will give a short introduction to the particles and forces in the theory before elaborating more on the particles focused on in the thesis.

1.1 The Standard Model

The Standard Model (SM) is a framework describing our current understanding of Particle Physics. The SM relies on two field theories to describe three of four fundamental forces. The field theory called Quantum Chromo Dynamics (QCD) describes the strong interaction, while Electroweak Interaction (EWK) describes the electromagnetic and weak forces. The fourth fundamental force—gravity—is not included in the present formulation of SM. The model has been developed in stages driven by both theorists and experimentalists, where the latest addition to the theory was the confirmation of the Higgs particle [7, 8].

An overview of the elementary particles described by the Standard Model can be seen in Fig. 1.1.

The SM groups the particles into two part: The fermions and the bosons. A brief introduction to the groups will be given.

1.1.1 Fermions

Fermions are particles with spin $\frac{1}{2}$ and all have a corresponding anti-particle. There are 12 types of elementary fermions in the Standard

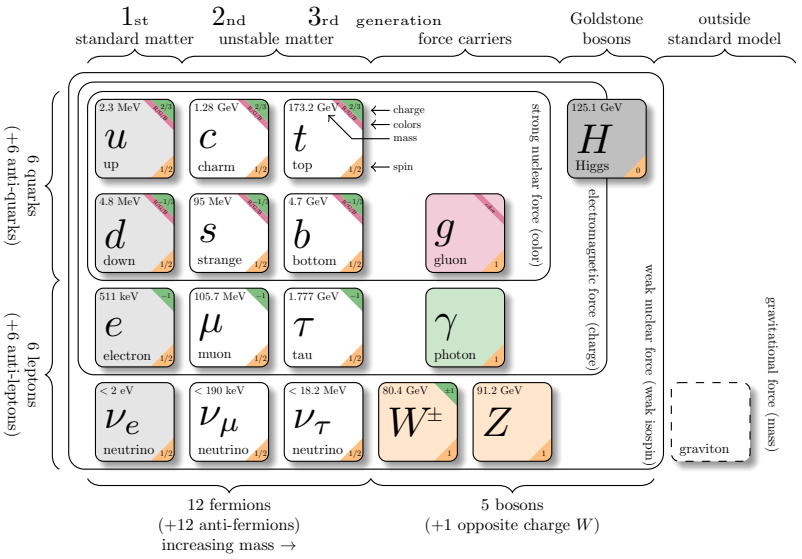


Figure 1.1: The Standard Model (SM) of Particle Physics [13]. The particles of the SM is grouped into two; the fermions and the bosons. The fermions are further split into quarks and leptons which are the building blocks of all matter. The bosons are force carrying particles mediating the weak, stong and electromagnetic forces.

Model: Six quarks and six leptons. They are the basic building blocks of all matter and have all been experimentally observed. They are separated into two groups, depending on whether they interact via the strong interaction or not.

1.1.1.1 *Quarks*

The six quarks are up, down, charm, strange, top, and bottom (u, d, c, s, t and b). We group them in *generations*, written as doublets:

$$\begin{pmatrix} u \\ d \end{pmatrix}, \begin{pmatrix} s \\ c \end{pmatrix}, \begin{pmatrix} b \\ t \end{pmatrix} \tag{1.1}$$

Quarks interact via the strong, electromagnetic, and weak forces. The quarks also carry color change, meaning that they are strongly bound to each other. The color change allows for the formation of color-neutral composites. This can both be as quark-antiquark pair (mesons, e.g. $\pi^+ = u\bar{d}$, $\kappa^0 = s\bar{d}$) or three quarks (baryons, e.g. proton (uud), neutron (udd)). This means that the quarks are the fundamentals constituents of mesons and baryons, the two types of particles also known under the joint name hadrons. Tab. 1.1 shows a list of the quarks along with a few of their properties.

Particle	Anti-particle	Spin	Charge
u	$\bar{u}$	1/2	+2/3
d	$\bar{d}$	1/2	−1/3
c	$\bar{c}$	1/2	+2/3
s	$\bar{s}$	1/2	−1/3
t	$\bar{t}$	1/2	+2/3
b	$\bar{b}$	1/2	−1/3

Table 1.1: List of the quarks, their antiparticle, spin and charge.

1.1.1.2 *Leptons*

The six leptons are also grouped into three generations:

$$\begin{pmatrix} \nu_e \\ e^- \end{pmatrix}, \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}, \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix} \tag{1.2}$$

These are the three negatively charged leptons; the electron, muon, and tau along with their neutral lepton; electron neutrino, muon-neutrino, and tau-neutrino, respectively. The leptons has—similar to the quarks—a corresponding anti-particle with opposite charge

Particle	Anti-particle	Spin	Charge
e^-	e^+	1/2	−1
ν_e	$\bar{\nu_e}$	1/2	0
μ^-	μ^+	1/2	−1
ν_μ	$\bar{\nu_\mu}$	1/2	0
τ^-	τ^+	1/2	−1
ν_τ	$\bar{\nu_\tau}$	1/2	0

Table 1.2: List of the leptons, their antiparticle, spin, and charge.

(Tab. 1.2). The electron, muon, and tau interact via the electromagnetic and weak forces. We will focus on the electron and the muon in this thesis, as taons are not measured in the ATLAS experiment.

1.1.2 *Bosons*

In order to mediate interactions and understand the forces between the fermions, we need the bosons. Bosons form the second of the two fundamental classes of particles and are different from the fermions as they all have integer spin. There are four types of bosons with spin 1: The photon, gluon, Z boson, and W bosons. The newest particle in the SM is the Higgs boson, which has spin 0.

The bosons are all force-carriers, meaning that they each mediate a type of interaction in the Standard Model. The photon, which is massless, mediates the electromagnetic interaction between electrically charged particles. The Z and W bosons are responsible for the weak interaction between quarks and leptons. The gluon mediates the strong interactions between color-charged particles (quarks and the gluon itself). Finally, the Higgs bosons interacts with all particles with mass, which are all particles, except the photon and the gluon. Tab. 1.3 summarizes this class of particles in the Standard Model. All particles and their interactions are summarized in Figure 1.2.

Particle	Anti-particle	Spin	Charge	Interaction
γ	self	1	0	Electro
W^-	W^+	1	-1	Weak
Z^0	self	1	0	Weak
g	self	1	0	Strong
H^0	self	0	0	Mass

Table 1.3: List of the bosons, their antiparticle, spin, charge, and the force mediated by the particle. "Self" means that the particle is its own anti-particle.

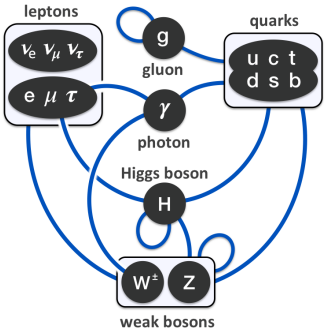


Figure 1.2: The interactions between particles in the Standard Model.

In this thesis, the focus will be on the Z and Higgs bosons. Below, a deeper introduction to the two particles is given.

1.2 *The Z boson*

The Z boson was first observed in the UA1 and UA2 experiments at CERN in 1983 [14]. The neutral vector is named after having zero electric charge opposed to its co-mediator of the weak force, the $W^{\pm}$ boson. Mathematically, the weak fields are the W_1, W_2, W_3 and the electromagnetic field, B. They all give rise to the gauge bosons, mediating the electroweak interaction, a combination of the weak and electromagnetic fields. The W_1 and W_2 fields combine to yield the $W^{\pm}$ bosons,

$$W^{\pm} = \frac{1}{\sqrt{2}} \left(W_1 \mp i W_2 \right). \tag{1.3}$$

The $W^{\pm}$ boson is left-handed, meaning that it interacts only with left-handed particles. Spontaneous symmetry breaking rotates the W_3 and B fields into the Z and γ bosons as follows,

$$\begin{pmatrix} \gamma^* \\ Z \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} B \\ W_3 \end{pmatrix}, \tag{1.4}$$

where θ_W is the weak mixing angle [15]. The Z boson is thus sometimes written as Z/γ^*, where the γ boson is the mediator of electromagnetic force and Z, the weak force. The combination of

the B and W fields means that the Z boson is neither right- or left-handed, but it is a combination of the two which affects the Z to favor some decays over others. The lifetime of the Z boson is very short ($\sim 10^{-23}$ s), and it is therefore only its decay products that are detected in the ATLAS detector.

The Z boson decay into a fermion and its anti-particle and the most likely decay for the Z boson is $Z \to q\bar{q}$ accounting for 69.2% of the decays of the Z boson (see Fig. 1.3). However, the quarks are difficult to separate from jets in the detector and the production of Z bosons at the LHC is dominated by the less likely Drell-Yan mechanism [17], where a quark from each of the incoming protons annihilate into a pair of leptons; muons or electrons (see the Feynmann diagram in Fig. 1.4).

This thesis will investigate the Z boson decays $Z \to ll$ and $Z \to ll\gamma$ where in the latter, the hard process causes emission of a photon from one of the leptons. The Z boson is valuable to investigate, as it a part of many analyses and searches for new physics phenomena [18]. In our case, the Z boson is especially interesting as the Higgs and Z boson share decay products, allowing us to investigate the impacts on a known high statistic resonance before testing the method on the Higgs boson.

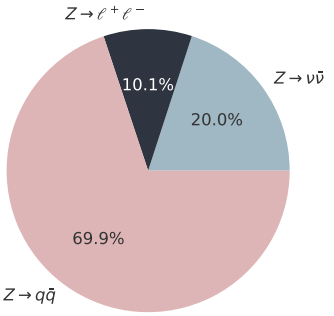


Figure 1.3: The decay rates of the Z boson. Figure made with data from [16]. The Z boson has equal probability of decaying into the three types of leptons, making it a lepton universality. However, only electrons and muons are considered in ATLAS.

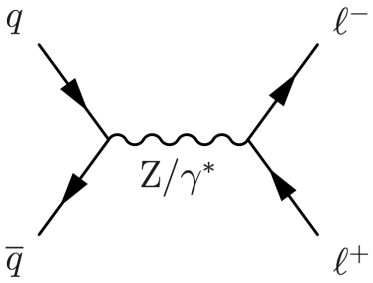


Figure 1.4: The Drell-Yan production of the Z boson decaying into a dilepton final state

1.3 The Higgs boson

In 1964, the Brout–Englert–Higgs mechanism was proposed as a theory to explain mass generation [2, 3, 19–22]. The mechanism required a spinless boson to exist. The boson was given the name *the Higgs boson*. In 2012, the ATLAS [7] and the CMS [8] experiments, at LHC in CERN, announced results consistent with the Higgs particle. In 2013, Peter Higgs and François Englert was awarded the Nobel Prize in Physics for their theory.

Their theory introduces the Higgs field which interact with all massive particles and thereby allowing them to have masses. This field is mediated by the Higgs particle discovered in 2012, filling the last missing piece in the Standard Model theory.

The Higgs particle couples with any particle having mass, and the coupling is proportional to the mass of the particle. The stronger the coupling, the more likely the Higgs decay into that particle is. Thus we are more likely to observe some decays than others. Fig. 1.5 shows the possible decays for the Higgs particle and the branching ratio[1] as a function of mass.

From this figure we see, given $m_H = 125\,\text{GeV}$, that we are most likely to observe the $H \to b\bar{b}$ decay. However, the discovery-channels from 2012 are the $H \to ZZ^{(*)} \to 4l$, $H \to \gamma\gamma$, and $H \to WW^* \to e\nu\mu\nu$ decays with a much lower branching ratio. The rare Higgs $b\bar{b}$ decay is difficult to separate from the abundant production of $b\bar{b}$ and other jets in the detector [24].

[1] See Sect. 2.2.2

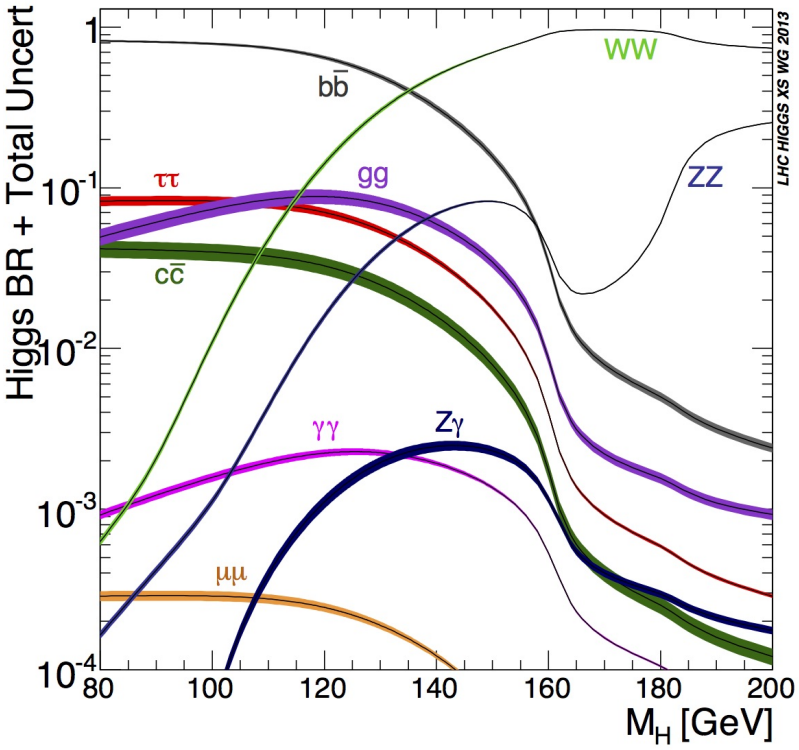


Figure 1.5: Possible decays for the Higgs boson as a funciton of mass and branching ratio [23]. For $m_H = 125\,\text{GeV}$, the most likely decay is the $H \to b\bar{b}$.

In this thesis, the $H \to Z(\to ll)\gamma$ decay will be investigated along the $Z \to ll\gamma$ boson decay. Several studies have been made by CMS and ATLAS on the $H \to Z(\to ll)\gamma$ decay [1, 9–12].

The predicted branching ratio of the $H \to Z\gamma$ decay is $(1.54 \pm 0.09) \cdot 10^{-3}$ at $m_{\text{H}} = 125.9\,\text{GeV}$ [1] which is comparable to that of $H \to \gamma\gamma$ of $2.27 \cdot 10^{-3}$ [25] (see Fig. 1.5). However, combining the rates of the $H \to Z\gamma$ and the $Z \to ll$, the $H \to Z(\to ll)\gamma$ is rarer than all discovery channels.

As mentioned in Sect. 1.2, the Z boson is well-investigated it is thus valuable to study the shared decay modes, as the analysis of the Z boson can provide knowledge of background events and signatures useful for the Higgs search.

In [1], an improved analysis compared to [10], is presented, including optimized lepton and photon identification criteria. The dominant background is the production of a Z boson together with a random photon. The events are classified into six categories designed to maximize the sensitivity to the presence of the SM Higgs boson. The categories each have different expected signal-to-background ratios and mass resolutions. A fit to the reconstructed $Z\gamma$ invariant mass distribution in all categories is performed to extract the overall signal. Fig. 1.6 shows the invariant mass of the $Z\gamma$ events satisfying the $H \to Z\gamma$ selection in Data in [1]. The observed data are consistent with the expected background with a p-value of 1.3% while the expected p-value in the presence of a SM Higgs boson is 12.3%. Thus,

increase in signal is very much striven for.

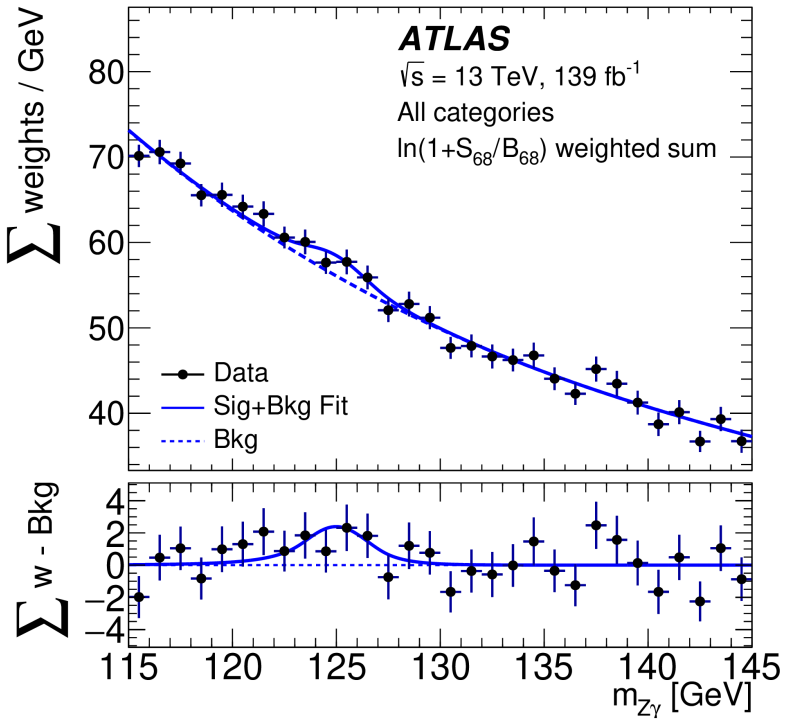


Figure 1.6: Weighted $Z\gamma$ invariant mass by the signal-to-background ratio. Figure from [1]

2 The ATLAS Experiment

ATLAS is one of the four major experiments at the Large Hadron Collider (LHC). This chapter will describe LHC and the associated experiments, before diving into the ATLAS detector itself. Unless otherwise stated, this chapter is based on the ATLAS detector paper [26].

2.1 The Large Hadron Collider

The LHC is the World's largest particle accelerator. It is located in Geneva, Switzerland, at the Consiel Européen pour la Rescherche Nucléaire (CERN). The accelerator is a 27-kilometer ring of superconducting magnets, where protons and heavy ions are brought to collision in one of the four different detectors associated with the

accelerator: Large Hadron Collider beauty (LHCb), A Large Ion Collider Experiment (ALICE), Compact Muon Solenoid (CMS), and A Toroidal LHC ApparatuS (ATLAS). The four detectors take on different jobs - LHCb is specialized in b-physics and ALICE is, as the acronym implies, designed for heavy ion physics. CMS and ATLAS are general-purpose detectors built to probe the Standard Model and it was these two experiments which discovered the Higgs Boson in 2012. The two differs in technical solutions and magnet-system design [27].

The LHC is the last element in the CERN accelerator complex—a succession of machines to increase the energy of particles. Each machine boosts the energy of the particle beam before it is passed on to the next machine in the sequence. Fig. 2.1 shows the accelerator complex and the placement of the four different detectors associated with the LHC.

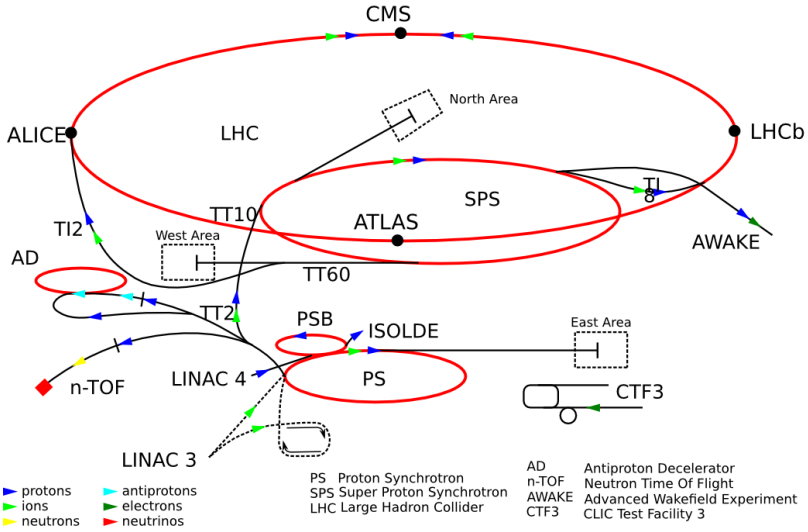


Figure 2.1: The CERN accelerator complex. The LHC is the last element in the succession of machines built to increase energy of particles. For detectors are located on the 27-kilometer ring. Picture from [28].

2.2 *Detector requirements*

Before we move on to the description of the ATLAS detector, I will first go through some necessary general information.

2.2.1 *Coordinate system*

ATLAS uses a right-handed coordinate system with cylindrical coordinates to describe the detector geometries. The collision point of the particle is defined as the origin while the beam direction defines the z-axis. The x-axis points in the direction of the centre of the LHC and the y-axis points towards the sky (see Fig. 2.2). Several angles are defined as well: The angle φ is measured in the xy-plane from the x-axis while θ is measured from the positive z-axis in the yz-plane.

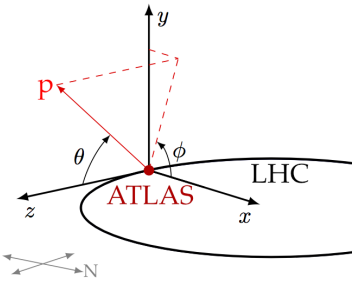


Figure 2.2: The coordinate system of the ATLAS detector. Modified but originally from [29]

Fig. 2.2 similarly shows a vector p. This is the momentum vector

[30] which is defined from the angles θ and φ, $p = (|p|, \theta, \varphi)$. The transverse momentum, p_T, is the component of p that is in the xy-plane transverse to the beam pipe, $p_T = p \sin\theta$. The transverse energy E_T is defined similarly.

The pseudorapidity, η, is defined from the angle θ, and is, unlike θ, invariant under boosts along the beam axis,

$$\eta = -\ln(\tan(\theta/2)).$$

For a particle perpendicular to the beam axis, the pseudorapidity η would be zero and for a particle in the beam axis, the pseudorapidity η would be infinite. The relationship between θ and η can be seen in Fig. 2.3. η is symmetric around the perpendicular axis.

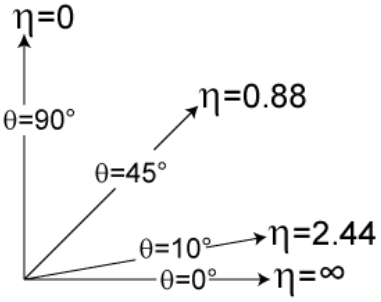


Figure 2.3: The relationship between θ and η. Picture from [31]

2.2.2 *Cross-section, luminosity and pileup*

The *cross-section* σ is a measure of the probability of producing a specific particle during collisions. Some particles are produced more frequently—and thus have a higher cross-section—than the ones being produced more rarely. When a particle has a small cross-section, a larger amount of interactions is needed to observe these rare particles and their decay. The cross-section is measured in barns ($b = 10^{-28} m^2$). The branching ratio of a particle is the fraction of events measured to decay in a certain way. The branching ratios sum to one for each particle.

To measure how many interactions happens, we use the concept of *luminosity*. Both the instantaneous and integrated luminosity exist.

Instantaneous luminosity is a measure of the interaction rate calculated from the cross-section σ and the number of events per second,

$$\mathcal{L} = \frac{1}{\sigma}\frac{dN}{dt}.$$

$\mathcal{L}$ is given in units of interactions per square centimeters per second, $cm^2 s^{-1}$. The integrated luminosity is, as the name implies, the luminosity collected over time in units of inverse barns, b^{-1},

$$\mathcal{L}_{int} = \int \mathcal{L} dt.$$

Thus $\mathcal{L}_{int}$ is the total amount of data recorded. Fig. 2.4 shows a plot of the total integrated luminosity as a function of time during Run2[1]. Given the cross-section and luminosity, the number of produced events from a specific process can be calculated by $\mathcal{L}_{int} \cdot \sigma$. The higher the number, the more occurrences of the process.

Pileup is the accumulation of luminosity. As we have many proton-proton collisions for each bunch crossing, the accumulated measure is used. More specifically, it is the integrated luminosity as a function of mean number of interactions, $\langle\mu\rangle$. Fig. 2.5 shows the distribution of pileup for the years 2015-2018 along with the mean number of μ for each year. The estimated pileup for Run3 is $\langle\mu\rangle \approx 60$.

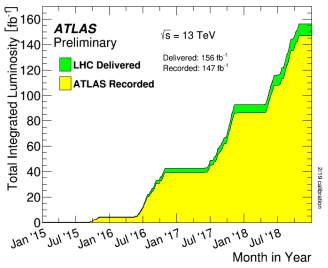


Figure 2.4: The total integrated luminosity of Run2 as a function of time. Picture from [32].

[1] ATLAS divides data-taking periods into "Runs". Run2 was the period 2015-2018, while Run3 is planed to run 2021-2023.

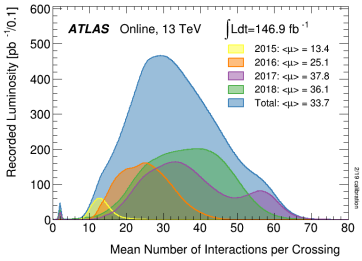


Figure 2.5: The luminosity-weighted distribution of the mean number of interactions per crossing. The mean value of μ is shown for each year. Picture from [33].

2.3 The ATLAS Detector

The ATLAS Detector is a general-purpose detector. This means that it investigates the fundamentals of matter and basic forces shaping the universe [34].

Beams of particles are sent to collide in the detector, forming new particles. The ATLAS Detector consists of several layers of detection systems, recording paths, momentum, and energy of the particles allowing for identification [27].

The detector is a 46 meter long and 25 meter wide forward-backward symmetric cylinder. It consists of the Inner Detector (ID), Calorimeters, the Muon Spectrometer (MS), and the Magnet System.

2.3.1 Inner Detector

The ID is the first part of the ATLAS detector the particle meets, starting only few centimeters from the beam axis and extending to a radius of 1.2 meters. Approximately 1000 particles emerge every 25 ns making it very important that the ID has high resolution. The main purpose of the Inner Detector is to track charged particles, revealing information about charge and momentum. This is achieved in the Pixel Detector, Semiconductor Tracker (SCT), and Transition Radiation Tracker (TRT) constituting the Inner Detector together with the 2T magnetic field surrounding the ID. The magnetic field bends the tracks of the charged particles, the direction and curvature of which disclose the charge and momentum.

All three sub-detectors in the ID is seperated into a barrel and end-cap part (see Fig. 2.6). The Pixel Detector and SCT covers the range $\eta < |2.5|$ while the TRT covers $\eta < |2|$. It is designed for particles above the transverse momentum threshold of $p_T > 0.5$ GeV.

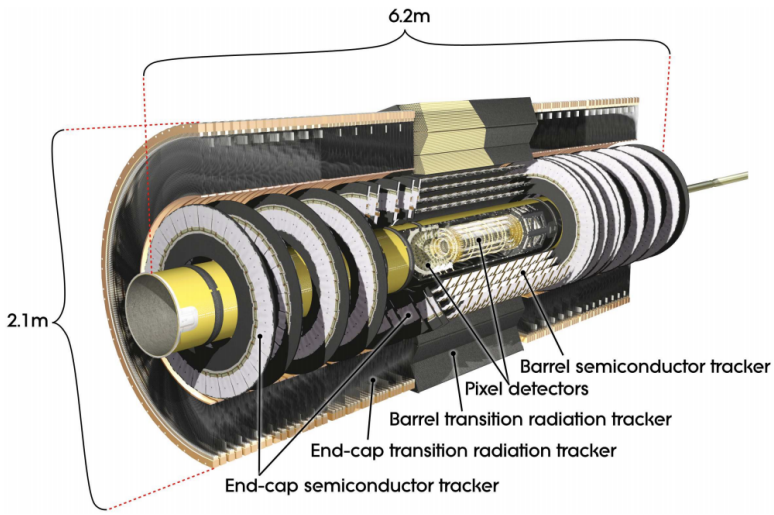


Figure 2.6: Cut-away view of the Inner Detector showing the Pixel, SCT, and TRT detectors. Picture from [26].

2.3.1.1 *Pixel Detector*

The pixel detector is the part of the detector closest to the beam pipe and is thus designed both to provide high-precision measurements with its 80 million read-out channels [35] and to withstand significant exposure to radiation. It is built out of semiconductors, consisting of three pixel disks in each end-cap region and three barrel disks designed to provide at minimum three points of the track from the collision. It has a resolution of 10 µm $(R-\phi)$ and 115 µm (z) for the barrels, and 10 µm $(R-\phi)$ and 115 µm (R) for the disks.

2.3.1.2 *The Semiconductor Tracker (SCT)*

As the name implies, the SCT also utilizes semiconductors, but in strips rather than pixels. The strips are glued back-to-back yielding two-side modules with good resolution in both dimensions. The SCT has four barrel layers and nine layers in each end-cap [35]. It has a resolution of 17 µm $(R-\phi)$ and 580 µm (z) for the barrels, and 117 µm $(R-\phi)$ and 580 µm (R) for the disks. The SCT, like the Pixel Detector, provides points of the track. The two parts combined provides about 10 coordinates with very high precision [36].

2.3.1.3 *The Transition Radiation Tracker (TRT)*

This part of the detector is a gaseous detector providing almost continuous tracking [36]. It is a combination of a straw tracker and a transition radiation detector. Long straw tubes with 4mm diameter with a gold-plated tungsten wire in the center are filled with gas [35]. A high bias voltage is applied causing a charged particle passing the tube to ionize the gas. This will be detected as a current creating a pattern allowing the path of the particle to be determined. Particles with a large γ factor ($\gamma > 10,000$, speed near the speed of light) emit transition radiation due to the tubes being filled mostly with xenon. This enlarges the signal improving the identification of these particles which are especially electrons.

The gas tubes are leaking and, due to cost reasons, some are being refilled with argon. The Ar gas performs similar to Xe for track reconstruction, but does not absorb as many transition radiation photons, and does thus not identify electrons as well. The TRT only obtains information with the resolution 130 µm $(R-\phi)$ direction as the tubes are aligned parallel to the beam.

2.3.2 *The Calorimeters*

The Calorimeters are located outside the magnet, surrounding the Inner Detector. It measures the energy of particles by absorbing it, and is built to contain the energy showers and limit the punch-through to the muon system. It is seperated into the Electromagnetic Calorimeter (ECAL) ($|\eta| < 3.2$), Hadronic Calorimeter (HCAL) ($|\eta| < 3.2$), and the Forward Calorimeter (FCAL) ($3.1 < |\eta| < 4.9$). All three calorimeters are sampling calorimeters, meaning that they have both an active and

passive layer. In the active layer, the energy is deposited and in the passive layer, the particles create showers of particles. See Fig. 2.7 for an overview of the calorimeters.

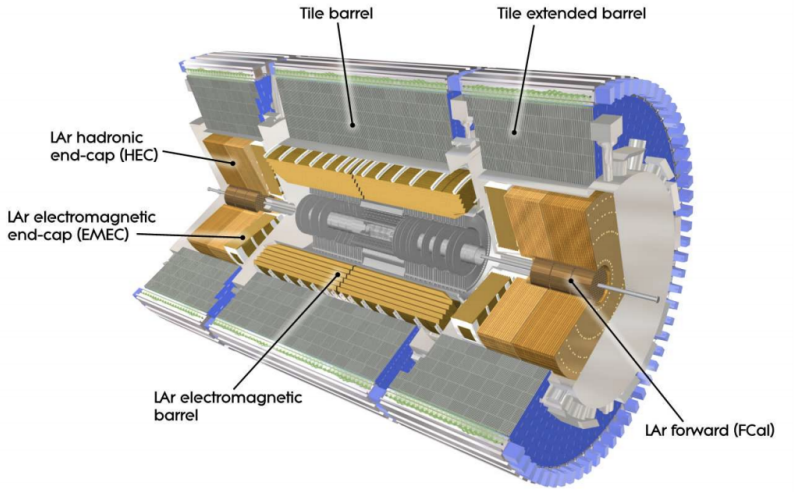


Figure 2.7: Cut-away view of the Calorimeters showing the ECAL, HCAL, and FCAL. Picture from [26].

2.3.2.1 *The Electromagnetic Calorimeter (ECAL)*

The EM calorimeter consists of a barrel part ($|\eta| < 1.475$) and two end-cap parts ($1.375 < |\eta| < 3.2$). It is a lead-LAr detector, meaning that the active material is liquid argon (LAr), and the passive absorbing material is lead. An overview of the calorimeter can be seen in Fig. 2.8. The ECAL absorbs energy from particles interacting electromagnetically, such as electrons and photons [37]. When the electron interacts with the calorimeter, a cascade of secondary particles (photons) will be produced with lesser energy. Each of these interact in the same way, producing a shower of particles. The electrons radiate off a photon via bremsstrahlung with the length scale of a radiation length (X_0). The total thickness in the ECAL is $> 22X_0$ in the barrel and $> 24X_0$ in the end-caps.

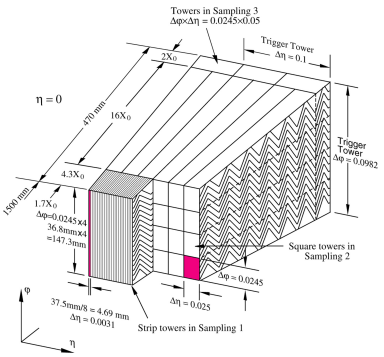


Figure 2.8: Overview of the Electromagnetic Calorimeter. Picture taken from [38].

2.3.2.2 *The Hadronic Calorimeter (HCAL)*

The HCAL consists of three parts, a barrel, an extended barrel, and end-caps. The barrels are built out of tile scintillators and steel as passive material. The barrels cover $|\eta| < 1.7$. The gaps between the barrel and extended barrel are called the "crack" containing some steel-scintillator modules recovering some of the lost energy of the region. The end-caps has a total of four layers with LAr as active material and copper as passive absorbing material covering the range $2.5 < |\eta| < 3.2$. Hadrons interact with the protons and neutrons in the material when passing the HCAL. Here, secondary particles are created interacting further with the material causing a hadronic shower. For both the HCAL and the ECAL, the intensity of the shower

is measured and converted into an electric current (signal).

2.3.2.3 *The Forward Calorimeter (FCAL)*

The FCAL covers the region closest to the beam pipe with $3.1 < |\eta| < 4.9$. It has three layers. The passive material is copper in the first layer and tungsten in the second and third layers. The active material is LAr in all layers. The first layer is mostly measuring electromagnetic interactions while the second and third layers are measuring hadronic interactions.

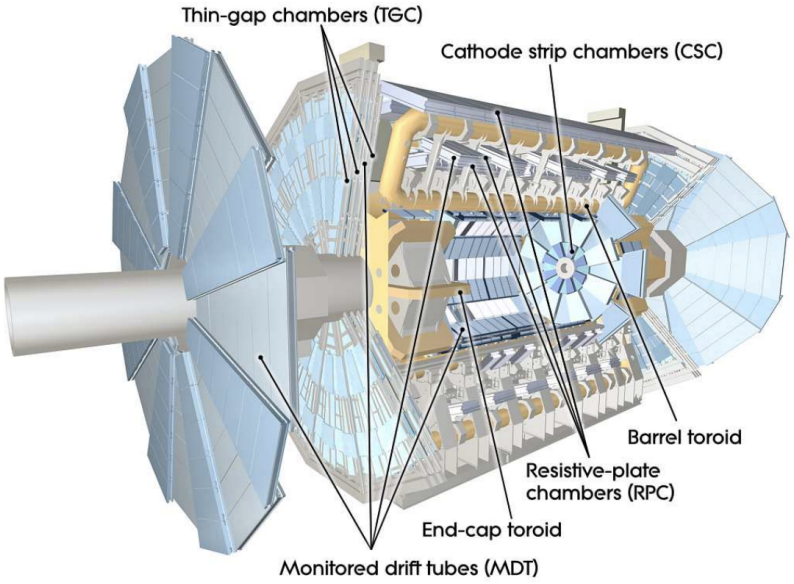


Figure 2.9: Overview of the Muon Spectrometer. Picture taken from [26].

2.3.3 *The Muon System*

As muons pass through both the ID and the Calorimeters, a special setup is needed for the particle. The Muon Spectrometer is based on deflection of muons tracks in toroid magnets, instrumented with trigger and high-precision tracking chambers. Fig. 2.9 provides an overview of the large system while Tab. 2.1 lists the main parameters. The MS extends over the radius from 4.25 m to 11 m. The size is required to measure the momentum of muons accurately. For the range $|\eta| < 1.4$ the track is bend by the large barrel toroid magnet while for $1.4 < |\eta| < 1.6$ the magnetic deflection is provided by the barrel and end-cap fields in combination.

Monitored drift tubes (MDT)	
Coverage	$\|\eta\| < 2.0$
No. of chambers	1150
No. of channels	354000
Function	Precision tracking
Cathode strip chambers (CSC)	
Coverage	$2.0 < \|\eta\| < 2.7$
No. of chambers	32
No. of channels	31000
Function	Precision tracking
Resistive plate chambers (RPC)	
Coverage	$\|\eta\| < 1.05$
No. of chambers	606
No. of channels	373000
Function	Triggering, second coordinate
Thin gap chambers (TGC)	
Coverage	$1.05 < \|\eta\| < 2.7$ (2.4 for triggering)
No. of chambers	3588
No. of channels	318000
Function	Triggering, second coordinate

Table 2.1: Main parameters of the MS. Numbers for the MDT and RPC refer to the final configuration of the detector in 2009. Table from [26].

2.3.3.1 *The toroid magnets*

Three large air-core toroid generates the magnetic field for the Muon Spectrometer. There is one barrel toroid and two end-cap toroids. The barrel magnet provides bending in the range $0 < |\eta| < 1.4$ and the end-cap magnets in $1.6 < |\eta| < 2.7$. The magnets overlap in the region $1.4 < |\eta| < 1.6$ giving a slightly lower bending power.

2.3.3.2 *The muon chambers*

The precision measurements of the track coordinates is provided by the Monitored Drift Tubes (MDT's) combining high measurement accuracy and simplicity of construction. They cover $|\eta| < 2.7$ (except for the innermost end-cap region, where $|\eta| < 2.0$). In the forward region ($2 < |\eta| < 2.7$), the Cathode-Strip Chambers (CSC) are used.

The trigger system covers $|\eta| < 2.4$ consisting of Resistive Plate Chambers (RPC's) in the barrel and Thin Gap Chambers (TGC's) in the end-cap regions. They serve the purpose of providing bunch-crossing identification, provide p_T thresholds, and measure muon coordinates. There exists two different trigger systems as the radiation in the forward region is high and should withstand the exposure.

2.3.4 *Overview of particle interactions*

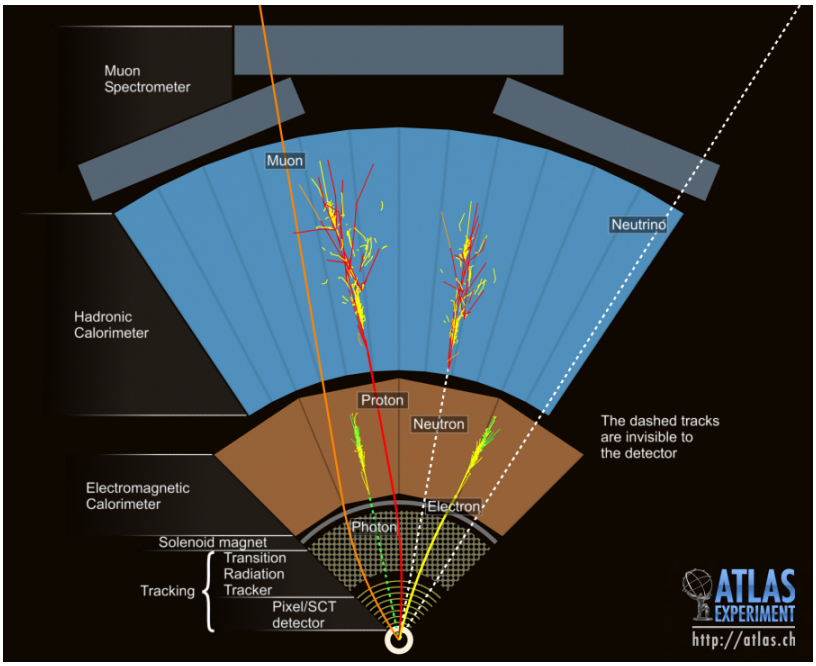


Figure 2.10: Computer generated image of an event cross section in the ATLAS detector. Picture from [39].

Fig. 2.10 shows a computer generated image of an event cross section in the ATLAS detector. The image shows the interaction of final-state particles in the detector layers discussed in this chapter.

Electrons and photons: Deposit their energy in the ECAL. The electron will leave a track in the ID while the photon leaves no track here (unless it is converted).

Hadrons (protons and neutrons): Deposit their energy in the ECAL and the HCAL. A charged particle (e.g. a proton), will leave a track in the ID and interact with the ECAL before leaving a shower in the HCAL. The neutral neutron will not leave a track in the ID but interact with the ECAL and HCAL. Note that the figure does not show the neutron interacting with the ECAL, this is a mistake.

Muons: Are electrically charged and thus leaves a bent track in the ID and $1 - 3\,\text{GeV}$ energy depositions in the ECAL and HCAL. They are detected in the MS by tracking, but not stopped.

Neutrinos: Passes the detector undetected. The existence of neutrinos is inferred by missing transverse energy.

2.3.5 *Trigger System*

The purpose of the trigger system is to reconstruct events in real-time to identify the most interesting events to retain. There where two trigger levels in Run2 combined with a final event selection.

L1: The level-1 trigger is implemented in hardware at the detector site. It searches for high transverse-momentum muons, electron, photons, jets, and τ-leptons decaying into hadrons. It uses information from calorimeters and muon spectrometer and reduces the rate of events in the read-out to 100 kHz. The L1 trigger also defines Regions-of-Interest (RoI's) within the detector with interesting features. This is passed to the high level trigger.

L2/HLT: The level-2 trigger, also known as the high level trigger, is implemented in software. It utilizes the RoI's from L1 and all available detector data from the regions. The L2 reduces the trigger rate to approx. 3.5 kHz averaged over all events.

Event filter: The final stage of the event selection reduces the event rate[2] to approx. 200 Hz which can be recorded for offline analysis.

[2] i.e. saved events

The events saved for offline analysis are available for members of the ATLAS experiment. It is this data that will be used for investigation of relevant decays for the thesis. The trigger information is saved and used later to recognize "good" particles.

2.3.6 *ATLAS Likelihood and Working Points*

Prompt electrons in the central region of the detector are selected using a likelihood-based (LH) identification. It is based on simulated samples with prompt electrons as signal, and a combination of jets as background [40]. For different binned variables, products of the signal and background likelihoods are calculated,

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

where n is the number of variables and x is the vector of variables. P_i is the PDF for the i'th variable, and $L_{S(B)}$ is the final signal (background) LH value. A discriminant is formed upon where the LH is based,

$$d = \frac{L_S}{L_S + L_B}, \tag{2.2}$$

which is logit transformed into d' without the sharp peaks at zero and one,

$$d' = -\tau^{-1} \ln (d^{-1} - 1), \tag{2.3}$$

with $\tau = 15$. The PDFs and discriminant are formed in bins of η and E_T. These can be seen in Tab. 2.2.

Bin boundaries in $\lvert\eta\rvert$									
0.0	0.6	0.8	1.15	1.37	1.52	1.81	2.01	2.37	2.47

Bin boundaries in E_T [GeV]												
4.5	7	10	15	20	25	30	35	40	45	80	150	∞

Table 2.2: Boundaries of the bins in $\lvert\eta\rvert$ and E_T of the electron candidate.

ATLAS has four fixed operating points for the likelihood, referred to as *VeryLoose*, *Loose*, *Medium*, and *Tight*. They have different requirements on tracking criteria. E.g. the Loose, Medium, and Tight require the same number of hits in the pixel detector, but further requirements exists for Medium and Tight in terms of hits in the innermost pixel layer [40].

Previously, the electron selection in ATLAS was cut-based, but the LH has advantages such that if a prompt (signal) electron fails the cut-based identification on only one parameter, it might survive the LH selection as the parameters are considered as a whole [40].

The muon and photon selections are still cut-based, having criteria of e.g. p_T, hits, and η. These selections are thus not based on likelihoods, but are referred to as Working Points (WP). The photon has the two operating points *Loose* and *Tight* while the muon selection has the WPs *Low-p_T*, *Loose*, *Medium*, *Tight*, and *High-p_T*. In [41], a multivariate *Low-p_T* WP based on a boosted decision tree is introduced, showing promising results.

2.3.7 *MC production*

In order to study the vast data in ATLAS, a very detailed simulation has been implemented, allowing the study of the detector response for the wide range of physics phenomena. Generally, the simulation is generated in three steps: The generation of the event and decays, the simulation of the detector and physics interactions, and the digitization of the energy deposited; creating signals in the detector electronics.

In the event generation, sets of particles are produced and brought to collision [42]. Several generators exists, each responsible for different types of event productions. Each event contains particles from a single interaction, with a vertex located in the geometric origin. No detector geometry is considered at this step.

After the events are generated, each particle is propagated through the ATLAS detector. The energy deposited is recorded and stored. Furthermore, the *truth* information (i.e. the particle id, true energy etc.) is also stored for each event, which can be used to quantify the success of the reconstruction [42].

The last step, the digitization, takes the stored information about energy deposits and generate the detector signal. To reproduce particular runs, dead channels and noise rates are read from a database and added to the simulation. The first level trigger, implemented in hardware on the real detector, is also simulated [42].

All of these steps produces the Monte Carlo simulated data which forms the foundation of this thesis. The samples allows for model production and analysis which will be discussed in the next chapters.

3 Machine Learning

The goal of this thesis is to use data from the ATLAS detector to improve the search of the Higgs particle. The large-scale MC production in ATLAS facilitate ideal conditions to train machine learning (ML) models, as processes with known heritage can be investigated. In this chapter, it will be described how one learns from data, and the specific machine learning-based method used in the thesis will be covered.

3.1 Learning methods

Generally, learning from data allows one to utilize historical data to improve decisions tomorrow. A set of observations are used to uncover an underlying process. The main type of learning is called *supervised learning*, where the training process can be seen as being supervised by a teacher presenting the learner with training data. Since we are in a teaching situation, the training data has examples of correct output. The training data and the corresponding output is given to the learner, in order to grasp the underlying function[1].

Other types of learning are *unsupervised learning* and *reinforcement learning*. In the former, no correct output is provided and the data is

[1] See the next section for a more formal definition

instead characterized based on inherent properties while the latter utilizes scoring of the output to force the model to learn.

In this thesis, I have used the method of supervised learning, because the correct output is known. The decay processes for the Z and Higgs bosons are well-described. The training of the models are based on Monte Carlo simulated data allowing us supervise the algorithm. The following sections will focus on the paradigm of supervised learning.

3.2 The Learning Problem

To understand the learning problem, a hands-on example will be used—a bank. They get a lot of loan applications. Can the process of approving costumers be automated? By using data available from previous loans and the profit earned, together with information on the costumers; salary, years in residence, outstanding loans etc., a supervised learning task is formed. Historical data serves as a good formula for credit approval. Below, the learning example is formalized [43].

input x: Customer information (salary, previous loans etc.)

input space $\mathcal{X}$: Set of all possible inputs $\mathbf{x}$

output space $\mathcal{Y}$: Set of all possible outputs (can be binary (yes/no, 0/1) or multi-class (1,2,3..., red, green, blue...). In this example, it would be to approve or not approve the loan application. Is often referred to as *labels*

data set $\mathcal{D}$: Set of input-output examples $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)$, where $y_n = f(\mathbf{x}_n)$ for $n = 1, \ldots, N$. In the credit approval example, $\mathbf{x}$ is information about previous customers and y is the correct decision of approval in hindsight.

hypothesis set $\mathcal{H}$: A set of candidate formulas to describe the input-output relation. Could be a set of linear functions.

unknown target function $f : \mathcal{X} \to \mathcal{Y}$: Ideal formula for credit approval

approximated target function $g : \mathcal{X} \to \mathcal{Y}$: The target function best approximating the unknown function f. This is chosen from $\mathcal{H}$, e.g. the function with the best fit to data, $\hat{y} = g(x_N)$.

When a new customer applies for a loan, the decision will be based on g and how good a decision that is depends on how well g replicates f [43]. g is more commonly referred to as a "model".

3.2.1 Training a model

A data set, used to train the model and find the best fitting approximation to the unknown target function, is usually divided into two or three sets. The first, a *training set*, is used to train the model and fit the parameters to create a model best fitting the training set. The

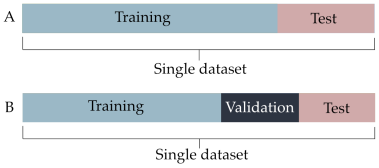


Figure 3.1: Examples of ways to partition a dataset. Method A only has a training and a test set, while method B also has a validation set. Figure inspired by [44].

performance of the model is evaluated on the *test set*. This is a part of the data not known by the model. If the model is trained properly, it is able to predict the output y_N of the test set. The third set, the *validation set* is not always used, but it can be utilized to evaluate the fit of a model, tune parameters unbiased, and to implement early stopping to ensure that a model is not overfitted[2]. See Fig. 3.1 for an example of the partition.

[2] See Sect. 3.2.3: If the model is fitted "too well" on the training data, it will not be able to generalize and is overfitted.

It is important that the datasets all follow the same probability distribution, i.e. represents the same data. If the training set only contains people accepted for loan and the test set only contains people rejected for a loan, the model will not be able to predict the data in the test set correctly.

3.2.2 *Error measure*

It is not expected that a trained model g will predict the correct output every time, as it is an approximation to the target model f, $f(x) \neq g(x)$. We define an *error measure*[3] to evaluate the model. The choice of an error measure may affect the learning outcome, so its not unimportant which measure is chosen.

[3] is also often referred to as a *loss*, *cost* or *objective*

We will only work with binary classification in this thesis, where only two outputs are possible; 0/1, true/false, signal/background. Models used for binary classification often outputs a score between 0 and 1, where the higher the score, the more signal-like the data-point is.

An error measure used for such classification is the Receiver Characteristic Operator (ROC) curve corresponding and Area Under Curve (AUC) score. These measures takes a one-dimensional variable—the output score of the model—and computes the *true positive rate* (TPR) and *false positive rate* (FPR) at various threshold settings.

$$TPR = \frac{TP}{TP + FN} \qquad FPR = \frac{FP}{TN + FP}, \tag{3.1}$$

where TP are the true positives, FP are the false positives, TN are the true negatives and FN are the false negatives. This means that TP are the samples which the model (at the specific threshold) predicted to be "positive" or "signal" correctly, while FP are the ones that are incorrectly predicted as signal. These can also be visualized in a confusion matrix (see Tab. 3.1), where the correct predictions are those in the diagonal.

True / Predicted	**P**	**N**
P	TP	FP
N	FN	TN

Table 3.1: Confusion matrix

The ROC curve is a plot of the TPR against the FPR, and thus the ability for the model to separate signal from background at various threshold settings. One way to visualize this, is to look at distributions of true/signal values and false/background values (see Fig. 3.2). In the top figure, a cut at $x = 0.45$ is shown. Here, $TP = 0.5$ and $FP = 0.16$ (of the total true and false, respectively), corresponding to the dashed lines in the bottom figure. Creating cuts at all thresholds in the top plot yields the bottom curve, having an area under the curve of 0.76. If the histograms where completely separable, the AUC

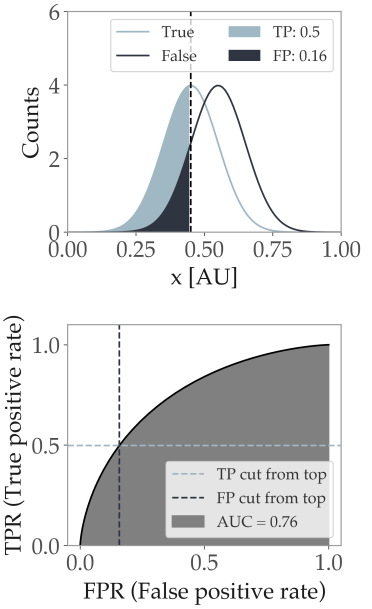


Figure 3.2: Top: two histograms showing true and false values. At $x = 0.45$ there are 0.5 true positives and 0.16 false positives (of the total true and false, respectively). Bottom: Area under curve (AUC) including the cut of $x = 0.45$.

would be 1 and if they where indistinguishable the AUC would be 0.5. In this thesis, the ROC curve and AUC will be used for the Pid and Iso models.

3.2.3 *Overfitting and early stopping*

The in-sample error, E_{in} [43], is based on data points used for training of the model. It measures training performance, adjusting the model towards improvement to create the model best fitting the training data. The out-of-sample error, E_{out} [43], measures how well the training has generalized on data not seen before. We are interested in training the model well and minimize E_{in}, but not so well that it does not generalize. The optimal will be to minimize E_{out}.

In practice, this is achieved using the validation set. While training the model using the training set, the model is simultaneously evaluated on the validation set. If evaluation of the validation set stops improving, the optimal training is achieved and is stopped. This is referred to as early stopping. See Fig. 3.3. If the evaluation stops before a minimum in the validation error is reached, the model is underfitted, meaning that it does not capture the underlying structure of the data. The final evaluation will be on the test set, as both the training and validation set is used during training and the error on these will be biased.

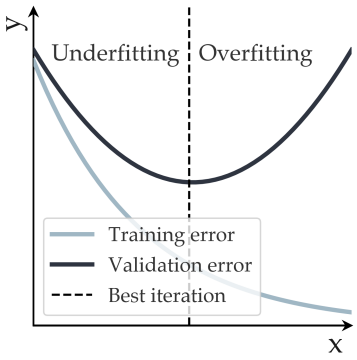


Figure 3.3: The error during training for the training and validation set. If the training stops before the minimum validation error is achieved, the model is underfitted.

3.2.4 *Hyper-parameters*

A hyper-parameter is a parameter used to control the learning process of a machine learning model. They can be optimized to optimally solve the learning task. This is done by trying various sets of parameters for the algorithm. The values yielding the optimal model minimizing the loss is chosen.

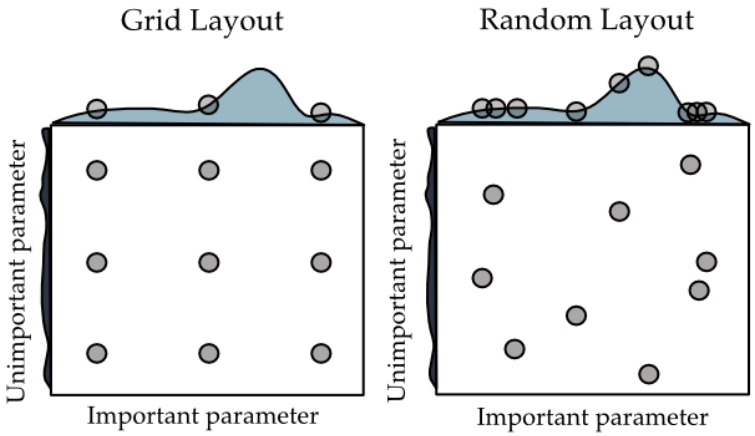


Figure 3.4: Grid and random search in nine trials. The method of random search randomly searches parameters in a given range and can often outmatch the search in a grid. Picture inspired by [45].

For the search of the optimal hyper-parameters a random search can be used. Here, parameters to be tested are chosen randomly

within a given range. Opposed to grid search, where specified combinations of parameters are searched (see Fig. 3.4), the number of parameter searches can be controlled explicitly. Random search is thus more efficient than grid search in many cases [45].

3.3 Models

Many types of machine learning models exist. Examples of supervised learning algorithms are regression, K-nearest neighbors, and neural networks. In this thesis, the tree based method (Gradient) Boosted Decision Trees (BDT's) with the framework LightGBM [46] is used. It is a simple and fast algorithm. Different algorithms has been explored by previous students in the group [47], where BDT's showed similar performance as neural networks. Furthermore, BDT's are already implemented in the ATLAS experiment framework allowing for easier implementation of new BDT models. In the following, I will therefore only review different types of tree based methods.

3.3.1 Decision trees

Decision trees are used for classification and regression [48]. The goal is to create a model, which uses simple decision rules inferred from data features to predict the value of a target. Fig. 3.5 shows a simple decision tree used to separate two species of the Iris flower (the "setosa" and "versicolor" species) based on sepal and petal length. Here, four cuts were required to create a total separation of the two species.

The optimal split for the decision trees is calculated using the *Gini impurity* [49], which is the probability that a randomly selected object will be misclassified. If allowed, the model will continue the splitting until a pure node[4] is reached. In practice, the training can be controlled by setting the maximum depth of the tree to avoid over-training. Still, decision trees have a tendency to over-fit as this parameter can be difficult to set, and a small variation in the data may even lead to a completely different prediction [48].

[4] All events in the node have the same label

A way to reduce the downsides of the decision trees and keep the benefits of simplicity is to use ensemble methods. Some ensemble methods based on decision trees are Random Forest, Bagging, Adaptive Booster, and Gradient Boosting. I will discuss two of them below.

3.3.2 Random Forest

The Random Forest learning method create a multitude of decision trees, and the predicted output is based on a majority vote from the different trees. The name "random" is chosen, because the trees are build in randomly selected sub-spaces of the feature space to reduce over-fitting [50].

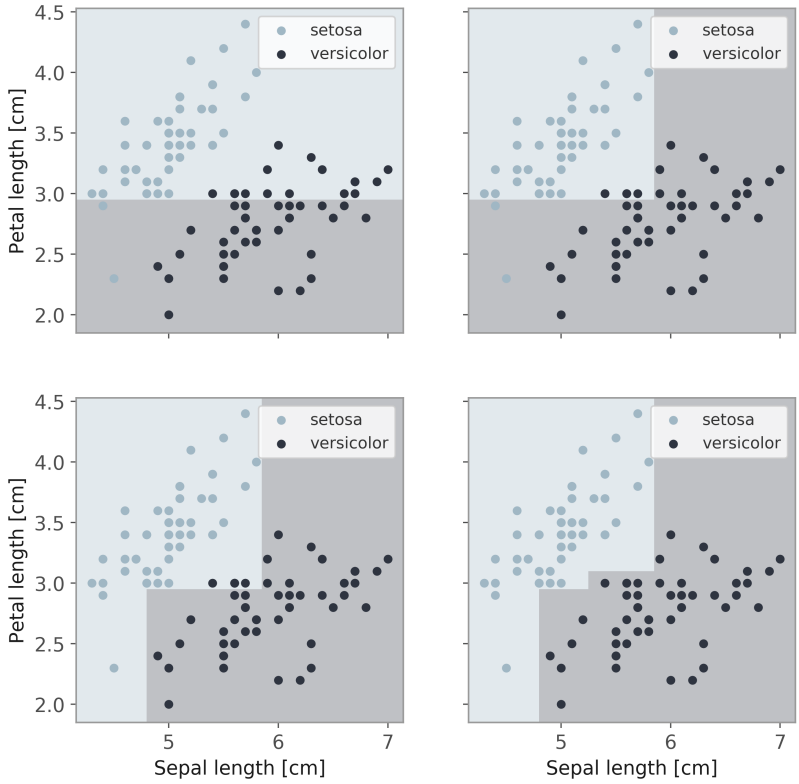


Figure 3.5: Classification of two different species of the Iris flower using a decision tree. Four cuts are combined to separate the two species in sepal length and petal length.

3.3.3 *Gradient Boosting*

The gradient boosting method combines *weak learners* to create a strong model. The weak learners are simple decision trees, and these are combined stage-wise to compensate for the mistakes of the previous learners [51]. When working with *gradient* boosting, the mistakes are identified by gradients in the loss function.

3.3.3.1 *LightGBM*

In previous comparisons of different gradient boosting algorithms for ATLAS [47], the framework LightGBM was found optimal due to superior speed. For smaller datasets, LGBM also outperformed its competitor in terms of accuracy. The LightGBM is a gradient boosting framework using histogram-based algorithms, which speeds up training and reduces memory usage [46]. Most decision tree learners grow trees depth-wise, and reduces the loss by adding levels to the tree. LightGBM grows leaf-wise, so only a single leaf is expanded instead of all leafs on that level. The leaf with the highest gain is split. Thus, the main parameter controlling the complexity of the model is the maximum number of leaves. When optimizing the hyper-parameters for models, the parameters `num_leaves`, `max_depth`, and `min_data_in_leaf` are tuned. The first and second both controlling the depth of the tree and the latter controlling if a split can be made for that leaf [52].

3.3.3.2 *Hyper-parameter search for the LightGBM model*

When performing the search of the hyper-parameters for the models, cross validation will be utilized with the package `lightgbm.cv()`.

Here, the dataset is split into K parts, e.g. 5 as in Fig. 3.6. In the first round, the first part of the dataset is used as the validation set and the rest is used for training. The model is then re-trained, but this time with the second part of the dataset as validation. The training is completed K times, and the performance is measured as a mean of the performance on the K validation sets.

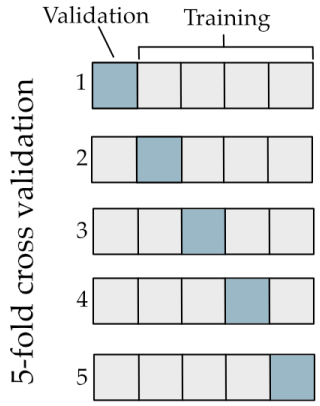


Figure 3.6: Splitting of the dataset in 5-fold cross validation

When we use this method for hyper-parameter search, 20 random configurations are drawn and a 5-fold cross validation is performed with a number of boosting rounds of 500 and early stopping rounds of 100. The configuration of parameters yielding the best performance (measured using AUC) is chosen as the parameters for the training.

For all models, the number of leaves, maximum depth, and minimum data in leaf is searched in ranges `num_leaves` $= [20, 40]$, `max_depth` $= [-20, 20]$ (negative being unlimited) and `min_data_in_leaf` $= [10; 100]$. `num_leaves < 2^(max_depth)` is required [52].

3.3.3.3 *Reweighing for the LightGBM model*

Before training a model, the data is *reweighed* to ensure that distributions for certain variables are statistically identical in signal and background. We do so, such that the model is trained *independent* on these variables and the model cannot infer something about the signal and background distribution. E.g. the models for the di-lepton pair in the $Z \to ll$ decay is also used as input for the $Z \to ll\gamma$ model, where the di-lepton pair has a lower invariant mass, $m_{ll}^{Z\to ll\gamma} < m_{ll}^{Z\to ll}$, as both $m_Z = 91.2\,\text{GeV}$. Thus it is important that the model is trained without knowledge of the mass and is unable to infer the information from other variables. The reweighing is performed with respect to 3-4 different variables that we wish to train models independent on—including the invariant mass—and a single weight is assigned to each event; it is not variable dependent.

The reweighing is performed using the module `GBReweighter` from the package `hep_ml` [53]. It is a gradient boosted reweigher based on ensembles of regression trees. The parameter `n_estimators` sets the number of trees used [54]. The number of estimators will determine how well the reweighed data fits the target, and a high number might over-fit the data. If so, the algorithm will have difficulties matching the distribution for the validation data. Thus, the reweighing is run with different number of estimators, and the best is chosen. For the Pid and Iso models, it is the distributions of η, $\langle\mu\rangle$ and E_T that will be reweighed to match between signal and background. For the Zll(g) and Hllg models, also the distribution of invariant mass will be reweighed. An example of reweighing of distributions for the ePid model can be seen in Fig. 3.7.

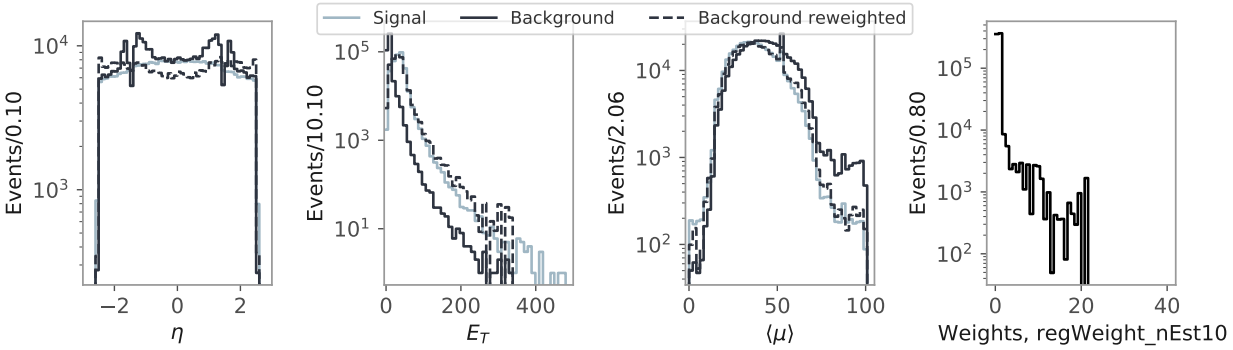


Figure 3.7: An example of reweighted data. The plot shows the signal and background data for the ePid model reweighted using 10 estimators. The reweighted background is plotted in dashed lines. The right-most plot shows a distribution of the weights.

4 Data processing and selection

A lot of data is produced in the ATLAS detector. This chapter will describe the processing and selection of the original data before it is used in the Machine Learning models.

4.1 Derivations

All events produced in ATLAS (both in Monte Carlo and Data[1]) are saved in xAOD files. From the xAOD files several DxAODs (derived xAODs) are produced. As the name implies, not all content of the xAODs are saved when producing the DxAODs. They are produced to reduce the size of the data such that the analysis can be run smoother and faster. The three main operations in the derivation are [55]:

Skimming: Removing non-relevant events
Thinning: Removing non-relevant objects from events
Slimming: Removing non-relevant information from objects

The DxAODs are produced by the working groups in ATLAS[2] saving the events and variables relevant for the analysis of the group. Each group has different derivations where the relevant ones for this thesis can be seen in Tab. 4.1.

During the work in this thesis, it was found necessary to get more variables for certain type of decay processes than was available in the derivations, as we wished to have a MC dataset for the decay $Z \to \mu\mu$ including variables for photons. The MUON1 derivation only saves very little information about the background photons and the EGAM4 production is only run on the $Z \to \mu\mu\gamma$ decay. The solution was to run the DxAOD software for EGAM4 (using the skimming, thinning, and slimming defined for the derivation) on the $Z \to \mu\mu$ AOD file. This allowed for getting more variables for the background photons in the events. Similarly, a EGAM8 $Z \to ee$ Data file was produced as this was not available.

[1] Monte Carlo generated data and Data taken in the ATLAS detector will be distinguished in the following by capitalizing Data.

[2] e.g. E/gamma, Muon, Tau, Higgs, Standard Model etc.

Name	Selection
EGAM1	$Z \to ee$
EGAM3	$Z \to ee\gamma$
EGAM4	$Z \to \mu\mu\gamma$
EGAM8	$Z \to ee$ with at least one electron from the forward region
MUON1	One CB muon w/ $p_T >$ 24 GeV and $\lvert\eta\rvert < 2.5$, second muon w/ $p_T >$ 4 GeV and dimuon invariant mass > 70 GeV.
HIGG1D2	At least one photon and two opposite-sign electrons or muons. Invariant mass of the three above 40 GeV. Further cuts in E_T, p_T and $\lvert\eta\rvert$ for the particles.

Table 4.1: Different derivations relevant for the thesis. EGAM derivations are produced for the specific decays while MUON and HIGG derivations are based on skimming events [56–58].

4.2 Ntuple production

The DxAODs are accessible to all members of ATLAS via Rucio[3] and can be downloaded to a local server. Before using the (still quite big) .root files, they are processed to even smaller files by creating ntuples. These are produced using the same principles of skimming, thinning, and slimming as the production of DxAODs with the difference that this is done locally. The ntuple production of this group has been developed by Helle Leerberg, a previous student of the group [59], and has been adapted for my own needs during the work of the thesis[4].

[3] The data management system used in ATLAS, see https://rucio.cern.ch/

[4] The latest version of the ntuple production of the group is available at [60]

The only skimming process used by the ntuple production is overlap removal [61], it focuses mainly on slimming and thinning. The thinning is dependent on the type of file, e.g. if you are working with a $Z \to \mu\mu$ file, only variables for muons are kept. Similarly, only the relevant variables are saved for each type of particle, for example the E_T is saved for the electron (if the file contains electrons), while p_T is saved for the muon.

In the production, truth-matching is run (for MC files), where particles, e.g. muon candidates, are matched with the truth events to get the *truth variables*. This is the simulated origin, p_T, η etc. for the particle.

The ntuple production is made specifically for the derivations seen in Tab. 4.1 as the different groups have different variables available in their DxAOD, and the production should thus be adapted if one considers a new type of DxAOD file.

Lastly, the ntuple dataset is converted into hdf5 files [62] before the data files for each specific analysis is created.

4.3 Datasets

All decay processes used for training can be seen in Tab. 4.2, along with the derivation used. The right-most column list the models where the specific decay is included in training.

Some datasets are used to train a series of models. When ntuple datasets are transformed into hdf5 files, they are split into 10 equal-size files. It is ensured that, if the models trained on the same decay will later be evaluated together, the files used for training are different. An example is the $Z \to \mu\mu$ decay, where the MUON1 Zmm file is used in training of the mPid, mIso, and Zmm models. Here, file 0-3 are used for Pid and Iso, while 4-9 are used for the Zmm model (both training and test), as we do not want to risk the test sample of the Zmm model to include events previously seen by the Pid and Iso models in training. However, the $Z \to ee\gamma$ data is used in the Zeeg and Heeg models will never "meet", meaning that the same data can be used for training and predicting without any trouble.

Process	**Derivation**	**Model**
$Z \to \mu\mu$	MUON1	mPid, mIso, Zmm
$Z \to \mu\mu$ (incl. γ in events)	EGAM4 (created from MUON1 AOD)	Zmmg
$Z \to \mu\mu\gamma$	EGAM4	mPid, mIso, pPid, pIso, Zmm, Zmmg, Hmmg
$W^+ \to \mu\nu$	MUON1	mPid, mIso, Zmm
$W^- \to \mu\nu$	MUON1	mPid, mIso, Zmm
$Z \to ee$	EGAM1	ePid, eIso, Zee
$Z \to ee$ (fwd)	EGAM8	ePid, eIso, Zee
$Z \to ee$ (incl. γ in events)	EGAM3	ePid, eIso, Zeeg
$Z \to ee\gamma$	EGAM3	ePid, eIso, pPid, pIso, Zee, Zeeg, Heeg
$W^+ \to e\nu$	EGAM1	ePid, eIso, Zee
$W^- \to e\nu$	EGAM1	ePid, eIso, Zee
$H \to ll\gamma$	HIGG1D2	Hmmg, Heeg

Table 4.2: The different processes and derivations used. A full list of the containers can be seen in Tab. A.1

5 Data Analysis

In this chapter I will present the analysis-part of the thesis. I will first give a broad overview of the work done and put it into context, before going into detail with the models and considerations behind. Models are trained to identify single particles—muons, electrons, and photons—and their combinations into Z and Higgs bosons.

5.1 Overview

This thesis presents a new approach to the $H \to Z\gamma$ search which will be evaluated against that of [1]. In that ATLAS selection, a cut-based approach is taken: The decay products are to satisfy the medium (loose) working point (WP) criteria, be within $|\eta| < 2.7$ (2.47, with $1.37 < |\eta| < 1.52$ excluded), and have $p_{\mathrm{T}} > 10\,\mathrm{GeV}$ ($E_{\mathrm{T}} > 10\,\mathrm{GeV}$) for muons (electrons). Photons are to satisfy same requirements as the electrons, except the WP criteria, where photons should meet the Tight WP identification criteria, and $|\eta|$, where $|\eta| < 2.37$. Furthermore, muons and electrons are ensured to be from the primary vertex with the requirements of $|\Delta z_0 \cdot \sin\theta| < 0.5\,\mathrm{mm}$. To suppress heavy-flavour decays, the requirement $|d_0|/\sigma_{d_0} < 3\,(5)$ for muons (electrons), should be met. See Tab. 5.1 for an overview of the cuts.

ATLAS selection

	Cut	Type	Description
Muons (electrons)	$Q_1 \cdot Q_2 < 0$	Kinematic	Leptons should have opposite charge
	$p_T\,(E_T) > 10\,\mathrm{GeV}$	Kinematic	The transverse momentum (energy) p_T (E_T) should be above 10 GeV
	$\lvert\eta\rvert < 2.7$ (2.47, with $1.37 < \lvert\eta\rvert < 1.52$ excluded)	Kinematic	The pseudorapidity should be lower than 2.7 (2.47, the crack and forward calorimenter are excluded)
	WP medium (LH loose)	Identification	Both leptons are identified with the medium (loose) working points
	$\lvert d_0\rvert/\sigma_{d_0} < 3\ (5)$	Vertex	The d_0 significance of both leptons should be less than 3 (5)
	$\lvert\Delta z_0 \cdot \sin\theta\rvert < 0.5\,\mathrm{mm}$	Vertex	Leptons are required to be associated with the primary vertex.
	track isolation	Isolation	Cut requiring 99% efficiency over the lepton p_T range. That means removing all pairs containing lepton with the 1% highest ptvarcone20.
Photons	$p_T > 10\,\mathrm{GeV}$	Kinematic	The transverse momentum p_T should be above 10 GeV
	$\lvert\eta\rvert < 2.37$, with $1.37 < \lvert\eta\rvert < 1.52$ excluded	Kinematic	Photons from forward and crack regions are excluded
	WP Tight	Identification	Photons are identified with the tight working point

Table 5.1: Selection from [1] used in this thesis to evaluate against ATLAS.

In this chapter, the new approach—an ML based one—will be aimed to improve the selection compared to the cut-based selection from ATLAS. The ATLAS selection focuses on the different decay products of the $H \to Z\gamma$ decay. The cuts of the selection are either based on kinematic variables, identification of the particle, or isolation of the particle. The ML based method presented in this chapter will take a similar approach. To replace the different cuts in Tab. 5.1, several models will be trained, and the output from some models will serve as input to the next. Below, I will introduce the different type of models and their purpose:

Pid: Particle identification models. Trained for muons, electrons,

and photons. Answers whether the particle we are looking at, is indeed the particle we are looking for. Sometimes an electron candidate can actually be a photon in-disguise, and the Pid models are trained to identify these. A separate Pid model for forward electrons (i.e. electrons in the forward calorimeter with $\eta > 2.47$) is trained. These types of electrons are usually not included in ATLAS selections. Different variables exist for the forward electrons, as the FCAL include other variables compared to the ECAL, and thus a separate model is necessary. The Pid models have similar tasks as the ATLAS Likelihoods/Working points for muons, electrons, and photons.

Iso: Isolation models. Trained for muons, electrons, and photons. Answers if the particle is isolated from nearby tracks. A true identified muon (by the Pid model) can originate from hadron decays in flight, pileup interactions, or cosmic rays [41] instead of the hard-scattering proton-proton interaction. These background muons, however, are not isolated in the detector. The Iso models takes different **kinematic** variables as input.

Zll: Di-lepton models. Combines the Pid and Iso models for muons and electrons with track and **vertex** variables. Trained for the $Z \to \mu\mu$ and $Z \to ee$ decays. The output of this model is compared with the cuts for muons and electrons in Tab. 5.1 (i.e. the top part of the table).

Zllg/Hllg: Models for the full $Z \to ll\gamma$ and $H \to Z(\to ll)\gamma$ decays. Combines the di-lepton models with Pid and Iso models for the photon. The output of this model is compared with all cuts in Tab. 5.1.

The Pid and Iso models will be referred to *mPid/mIso*, *ePid/eIso*, and *pPid/pIso* to distinguish the models for each particle—m, e, and p representing the three particles; muons, electrons, and photons. The models for the $Z \to \mu\mu$ and $Z \to ee$ decays will be referred to under the joint name *Zll*, while *Zmm* and *Zee* will refer to the model for the specific decay. Same applies to *Zllg* and *Hllg* models.

Fig. 5.1 represents an overview of the muon models and their combination into succeeding models. The mPid and mIso models are combined with track and vertex variables into the Zmm model. This Zmm model is then combined with the photon models and other variables to the Zmmg and Hmmg models. The figure shows the combination for muon models, but the same format applies to the electron models. The other variables, serving as input to the models, will be elaborated upon later in the chapter including the identification and kinematic variables being input to the Pid and Iso models.

The models are mainly trained using Monte Carlo generated data (see Sect. 2.3.7). Here, the truth variables (e.g. origin) are used to identify signal and background. These models are referred to as MC

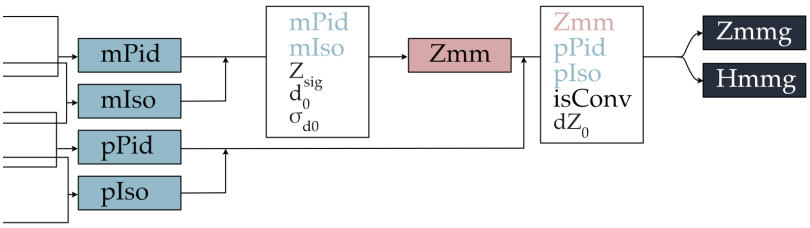


Figure 5.1: Overview of the models and their combinations. Shown for muon models. Same concept applies for the electron models. Pid and Iso models are trained and combined with track variables into a Zmm model. The Zmm model and photon variables are combined into a Zmmg/Hmmg model.

Truth models. Some models are trained in real Data where truth variables do not exist. To give labels (and distinguish signal and background) a Tag & Probe method is used, which will be explained in later sections. Tag & Probe-based models are also trained in MC data. These models are referenced to as MC T&P and Data T&P.

Fig. 5.2 shows the timeline for muon models and which of the models are trained on the types of datasets. The lepton Pid and Iso models, as well as the di-lepton model, are trained in MC and Data T&P. They will each take the corresponding Pid and Iso models as input to the di-lepton model. The Pid and Iso models for the photon, as well as the Zllg and Hllg models, are trained in MC Truth only. A total of 25 models are trained. In this chapter, it will be aimed to introduce all types of models and their performance. However, many of the details will be put in appendix.

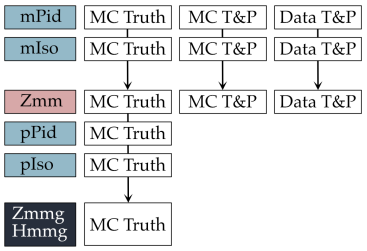


Figure 5.2: Type of models trained. MC Truth, MC T&P, and Data T&P models are trained for muons (and electrons) and Zll models. For the pPid, pIso, and Zllg/Hllg, only MC Truth models are trained.

When training in Monte Carlo data, it is important to consider the dominant background and the signal-to-background ratios for the true Data the final model is predicted upon. We wish to train on data resembling the true distribution of Data. It has been aimed to do so, but it has not been optimized fully. The Data-trained models are introduced in order to have a better relationship between the training and testing data. However, since we still need labels on the training data, in order to supervise the model, it is not certain that this model will be the most successful one, as we might not assign the labels perfectly.

When evaluating the models, it is important to consider the labeling and distributions. The Pid and Iso models will be evaluated using the ROC curve of the test sets. Here, a perfect ROC score does not necessarily mean that the model generalizes well to Data, but instead that it generalizes well to the test data with the same distribution in variables (and labels) as the training data.

The Zmm and Zee models are evaluated in both MC and Data, and compared to the selection in [1]. When evaluated in MC, they will be evaluated against their own test set. This means that the MC Truth model will be evaluated against the MC Truth labels while the Data T&P model is evaluated based on the T&P labels[1]. When the Zmm and Zee models are evaluated on Data, all types of models will be shown together on the same Data along with the ATLAS selection. The Zllg models will similarly be evaluated in MC and Data, while

[1] Thus, the Data T&P is actually not evaluated in MC, but in Data with labels

the Hllg models are evaluated on MC data only.

5.2 *Lepton and photon models*

In this section, single particle models will be described. In the Higgs decay $H \to Z(\to ll)\gamma$, single leptons and photons are decay products, and will be detected in the ATLAS detector. Thus, it is key that we have models identifying these particles and answering questions such as "Is it a muon?" or "Is the muon isolated?". Models to answer those two questions are described in this chapter. Pid models will answer the particle identification question, while Iso models tell us whether the particle is isolated. A separate Pid model for electrons in the forward calorimeter will be trained, which will be referred to as the eFwd model.

The procedure for data selection, training, and evaluation of Pid and Iso models will be described in the following.

5.2.1 *Datasets*

This section will describe the datasets for all types of training data.

All models will be trained single particles, i.e. one electron, one muon, or one photon. Tab. 5.2 shows the data files used for training the Pid and Iso models for electrons and muons, while Tab. 5.3 shows the photon files.

The datasets are separated into a training, validation, and test sample. As described in Chap. 3, the training and validation set is used in training. The results in this chapter will be shown on the test set as we want the unbiased estimate of the model output.

Signal files	
Electrons	Muons
$Z \to ee$	$Z \to \mu\mu$
$Z \to ee\gamma$	$Z \to \mu\mu\gamma$
Background files	
Electrons	Muons
$W \to e\nu$	$W \to \mu\nu$

Table 5.2: The decay processes used for the electron and muon data selection described in this section.

Datasets for photons
$Z \to ee\gamma$
$Z \to \mu\mu\gamma$

Table 5.3: The decay processes used for the photon selection described in this section.

5.2.1.1 *MC Truth dataset for muons and electrons*

Monte Carlo (MC) generated data has the advantage that one knows the true identity of a particle. This information is stored in *truth variables*, such as `truthType` [63], `truthPdgId` [2] [64], or `truthOrigin` [63].

The datasets based on the truth variables in the MC data will be slightly different for the Pid and Isolation models. In the Pid case, the signal and background will be distinguished using the truth variable truthPdgId. In the muon case, the truthPdgId is 13 for a negatively charged muon and -13 for a positively charged muon. For the election, this number is 11 [64]. Thus, signal will be muons (electrons) with $|\text{pdgId}| = 13$ (11) and background $|\text{pdgId}| \neq 13$ (11). For the eFwd model, the signal and background electrons are chosen the same way, however of course only considering electrons from the forward calorimeter.

For the Isolation models, we look at the truthType, as this allows for distinction of isolated and non-isolated particles. Particles with truthType $= 6$ (2) are isolated muons (electrons), i.e. signal, and those with truthType $= 7$ (3) are non-isolated muons (electrons), i.e. background.

[2] pdgId is an abbreviation for Particle Data Group ID

For both the Pid and Iso models, we will use the datasets listed in Tab. 5.2 and will not distinguish whether the particle came from a signal or background *file*. That means that muons not passing the signal cut, coming from the $Z \to \mu\mu$ file, will be background along with the events from the $W \to \mu\nu$ file.

5.2.1.2 *Tag & Probe datasets*

The method of Tag & Probe has the advantage that it can be used in both MC and Data. Pairs of particles are created using this method, where the "tag" particle passes tight selections; it should be a trigger particle, have a transverse energy above 26 GeV, be isolated ($p_T^{varcone} < 0.15 \cdot E_T$), and meet the Tight WP. This is paired with a "probe" particle from the same event passing very loose selection cuts, only checking that $E_T > 10$ GeV (see Fig. 5.3).

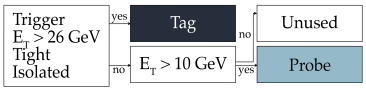


Figure 5.3: Tag & Probe selection. The tag particle is found from tight selection cuts, while only a very loose cut is made on the probe particle.

From this, we choose signal and background events based on the charge and the invariant mass between the particles (see Fig. 5.4). If the particles have opposite sign and are within 5 GeV of the Z boson mass, it is a signal event. Background is selected either from events where the leptons have same charge and invariant mass between 10 and 40 GeV from the Z boson mass, or opposite-sign events with invariant mass between 20 and 40 GeV from the Z boson mass.

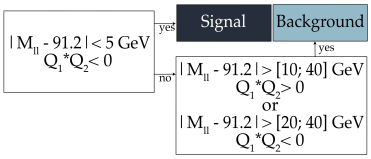


Figure 5.4: Signal and background selection for Tag & Probe selection. Only the probe particle is used as input in the training.

Furthermore, for the MC, we only choose signal events from the signal files in Tab. 5.2 and background events only from the background files. The background will be dominated by signal events if events from signal files are also considered as background. In Data, all events are considered for both signal and background.

When implementing the models, data was lacking and the files containing the background events were small. To get a proper signal/background ratio in terms of number of events, events are copied and shuffled from the existing background to augment the data. When training the model, only the probe particle is used, the tag particle is discarded.

5.2.1.3 *MC Truth dataset for photons*

The signal and background photons will be found slightly different than for the muons and electrons. When training the photon models, both Pid and Isolation, we will use isolated photons as signal. These have the truthType 14. For the isolation model, the background will be photons with truthType 16 (background photons). The pPid models are trained with all other types available as background. All photons with $E_T < 4.5$ GeV are cut away.

5.2.1.4 *A note on T&P dataset for photons*

A T&P model can also be trained for the photons, however, due to time constraints, this was not considered in this thesis. This could be done using signal events from the $Z \to ll\gamma$ decays, ensuring $E_T < 10$ GeV for the photon and invariant mass within 5 GeV of the Z boson mass. A lot of different files that could serve as background

files in the photon case exist. Considering the data we have available, random photon candidates from $Z \to ll$ files would be the preferred source of background.

Tab. 5.4 shows the number of events in the datasets for each model trained. From these numbers, the datasets are split into training, validation, and test set using the split 60-20-20.

Pid and Iso data selection

Model	MC Truth		MC T&P		Data T&P	
	Signal	Background	Signal	Background	Signal	Background
mPid	498,220	476,379	53,292	150,250	18,204	12,198
mIso	489,168	489,375	53,292	150,250	18,204	12,198
ePid	551,256	644,507	56,852	32,535	19,228	75,460
eFwd	30,967	402,927				
eIso	551,256	371,853	56,852	32,535	19,228	75,460
pPid	121,126	524,555				
pIso	121,126	109,482				

Table 5.4: Number of events for the Pid and Iso data sets. Rows show the model and columns show the signal and background events for each type of model trained. Note that the same datasets are used in Pid and Iso for T&P trained models.

5.2.2 *Reweighing*

The reweighing is made separately for each type of dataset, but is done in the variables $\langle\mu\rangle$, E_T or p_T, and η for all datasets. Fig. 5.5 shows the MC Truth training dataset for the ePid model is shown reweighed using 100 estimators. We see that for $\langle\mu\rangle$ and η, the background distribution resembles the signal well, but for E_T, it has some trouble fitting the signal distribution. It is clear that the signal has higher values of E_T than the background, it can thus not be reweighed to match completely. This is also not striven for, as the data then might be overfitted.

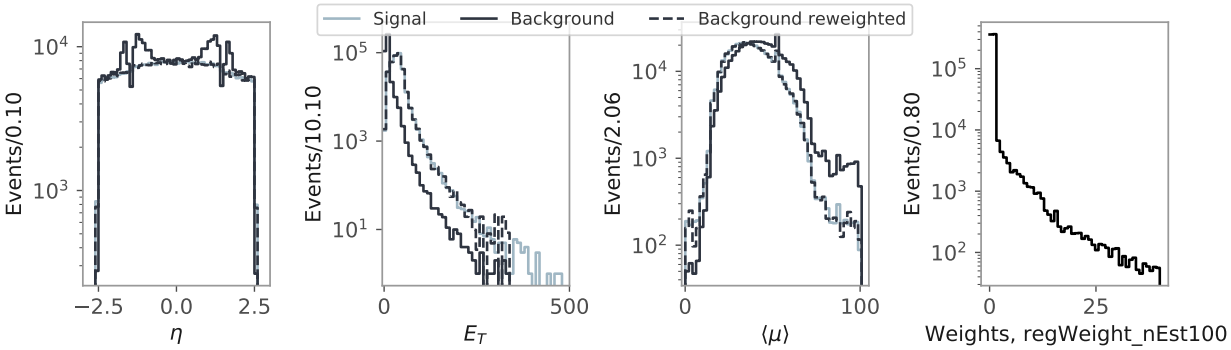


Figure 5.5: Signal and background distributions of $\langle\mu\rangle$, E_T and η for the MC truth dataset for the electron PID model. The background is reweighed with 100 estimators to match the signal distribution.

5.2.3 *Model training*

In the next parts, I will describe how the models are trained, their input parameters, and the performance of the models.

All models will be trained using the framework LightGBM. To ensure that the models are trained with the best parameters and that they are not over-fitted, hyper-parameter optimization and early stopping is implemented in the training as described in Sects. 3.2.3 and 3.2.4.

I will only show plots for the MC Truth muon model as all Pid and Iso model has same general implementation. However, variables for the other models will be shown. Furthermore, comments will be made and conclusions will be drawn on all Pid and Iso models, including the T&P trained ones. The full description for all models can be found in App. A.2.

5.2.3.1 *Particle Identification: Pid*

As mentioned in the introduction to Sect. 5.2, the Pid model helps answering the question of whether the particle we are looking at, is indeed the particle we are looking for. Input features to these models are tracking and calorimeter variables, which are also used in the ATLAS Likelihood (see Sect. 2.3.6).

The mPid models are trained using the parameters in Tab. 5.5. A description of the variables can also be found here. The features for the electron and photon models can be found in Tab. 5.6 and for a detailed description, see Tabs. A.2 and A.4.

ML mPid input features

`muo_numberOfPrecisionLayers`	The number of hits in the precision layers
`muo_numberOfPrecisionHoleLayers`	Number of expected hits in the precision layer not seen
`muo_quality`	Type of muon (Combined, Segment tagged, StandAlone etc.)
`muo_MuonSpectrometerPt`	The p_T deposited in the muon spectrometer
`muo_scatteringCurvatureSignificance`	Used to search for a kink along the track, computed as the change of the normalised integral of the scattering angle significances in the track bending plane.
`muo_scatteringNeighbourSignificance`	Used to search for a kink along the track, defined as the largest value of scattering angle significance over the entire track
`muo_momentumBalanceSignificance`	Quantifies the difference between the energy loss in the calorimeters and the corresponding direct calorimeter measurement.
`muo_EnergyLoss`	Energy deposited in calorimeter
`muo_energyLossType`	Type of energy deposited in calorimeter

Table 5.5: Input features for the mPid models.

For the electrons, a separate model will be trained for the events in the forward calorimeter. The forward electrons does not include as many variables as the prompt electrons, as the detector is different, and thus the same model cannot be used. Usually, the events in the forward detector are excluded—as in the ATLAS selection we will

be comparing our models against—but they are included here. The model will be referred to as the eFwd model.

All Pid models will be compared to the ATLAS Likelihood working points Loose, Medium, and Tight for muons and electrons, and Loose and Tight for the photons.

ML ePid input features	ML eFwd input features	ML pPid input features
ele_Rhad	ele_centerLambdaCluster	correctedScaledAverageMu
ele_Rhad1	ele_lateralCluster	pho_eta
ele_f3	ele_fracMaxCluster	pho_et
ele_weta2	ele_topoetconecoreConeEnergyCorrection	pho_Rhad
ele_Rphi	ele_longitudinalCluster	pho_Rhad1
ele_Reta	NvtxReco	pho_weta2
ele_Eratio		pho_Rphi
ele_f1		pho_Reta
ele_d0		pho_Eratio
ele_d0Sig		pho_f1
ele_dPOverP		pho_wtots1
ele_TRTPID		pho_weta1
ele_deltaEta1		pho_fracs1
ele_deltaPhiRescaled2		pho_ConversionType
		pho_ConversionRadius
		pho_VertexConvEtOverPt
		pho_VertexConvPtRatio

Table 5.6: Features for the ePid, eFwd, and pPid models. The list including description of the variables can be seen in Tabs. A.2 and A.4.

SHapley Additive exPlanations (SHAP) is a game theoretic approach to explain the output of a Machine Learning model [65]. We will use this method to quantify the importance of the features used in the model. The features are ranked in terms of largest contribution or importance to the final model. The SHAP values for the mPid MC Truth model is seen in Fig. 5.6. The top ranking variable is `muo_numberOfPrecisionLayers`, which contains information on the quality of the muon track. The ranking is the same for the MC T&P model, while they are slightly different for the Data trained T&P model, however still with the `muo_numberOfPrecisionLayers` as the most important feature (see Fig. A.1).

For the ePid models (Figs. A.7 and A.10), the rankings differ to some extent for the three trained models. The feature `ele_Reta` scores high for all models, and also `ele_Eratio` and `ele_deltaEta1` are in the top. The Reta and Eratio variables both relates to the energy deposit in the calorimeter, while deltaEta1 is a tracking variable. For the eFwd model (Fig. A.7), it is also a calorimeter variable, `ele_fracMaxCluster`, that has the highest SHAP value.

For the pPid model (Fig. A.19), the variables `pho_Reta`, `pho_Eratio`, and `pho_deltaE` that has the biggest influence of the model output.

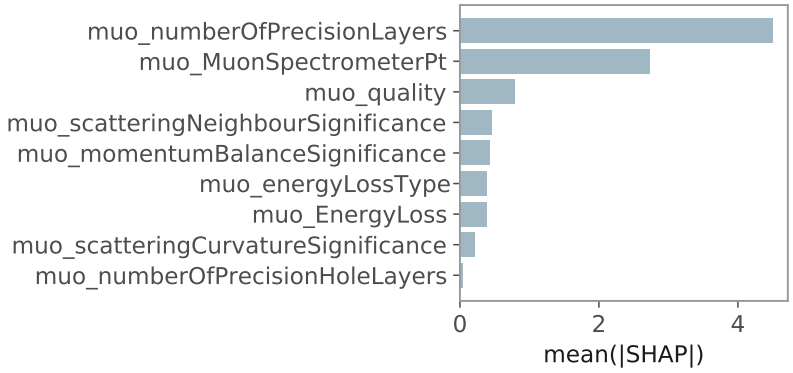


Figure 5.6: SHAP values for the mPid model. The variable with the highest influence on the model is `muo_numberOfPrecisionLayers`

These are all similar to the top features for the ePid models just described.

Each of the different models is tested on the corresponding test set by predicting the labels for the events. When predicting using the LGBM model, each input event is given a output score between 0 and 1. The higher the value, the more signal-like the event is. The (logit transformed) output for the final mPid model can be seen in Fig. 5.7. The logit transformation [66] maps the value in the range $(0, 1)$ to $(-\infty, \infty)$. In Fig. 5.7 we can see that the signal and background are separated nicely, which tells us that the model performs well. Another way to see this, is to plot the ROC curve. Fig. 5.8 shows the ROC curve for the mPid MC Truth test set. We see that the AUC for the LGBM model is 0.998, meaning that it has almost perfect performance and the output distribution of signal and background are nearly completely seperable. The figure also shows the ATLAS working-points (WP) Loose, Medium, and Tight. Plots of the LGBM scores and ROC curves for the rest of the muon models along with the electron and photon models can be found in App. A.2.

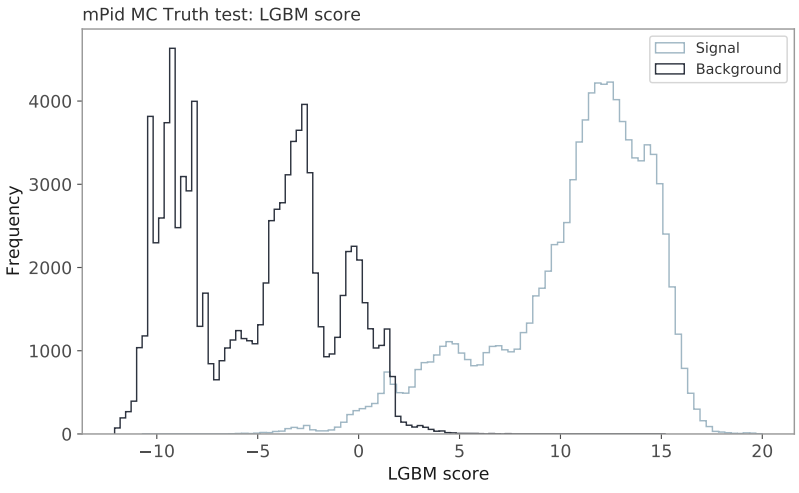


Figure 5.7: Logit transformed LGBM score for the final mPid MC Truth model. There are some evident features in the prediction scores, but the signal and background are separated nicely.

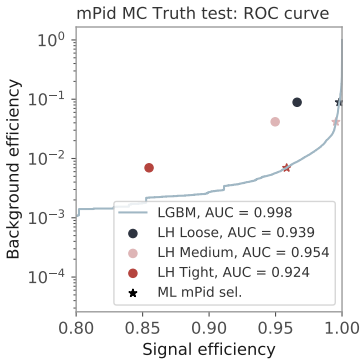


Figure 5.8: The ROC curve for the mPid MC Truth test set. The ATLAS Loose, Medium and Tight WP are shown on the figure as well as the ML selection that matches their background efficiency.

Tab. 5.7 shows the TPR (signal efficiency) for the ATLAS selection and the corresponding TPR for the LGBM Pid models at similar background efficiency. The change for each model is shown, indicating the improvement or deterioration of the ML models compared to the ATLAS WPs. The TPRs are similarly shown in Fig. 5.8 for both the ATLAS selection and the LGBM, the latter as *ML mPid sel.*. Since the outputs of the WPs are binary, compared to the continuous output of the LGBM model as seen in Fig. 5.7, they are depicted as single points rather than a curve.

The Data T&P models has efficiency much worse than the ATLAS prediction, but for the electron and muon models, the MC trained model improves the selection from $2-28\%$, where the largest improvement is for the ePid MC T&P model at the Tight cut, with 27.6%. The signal efficiency of the eFwd model drops significantly compared to the ATLAS WP. However, ATLAS does not include the forward electrons, so simply doing so, is likely to yield an increase in the final comparison.

For the photon model, the ML model has an improvement of 9% for the Loose WP while a slight worsening is seen for the Tight WP. The corresponding ROC curve can be found in Fig. A.20.

Pid models performance

Model	**Loose WP**			**Medium WP**			**Tight WP**		
	ATLAS	ML	Change	ATLAS	ML	Change	ATLAS	ML	Change
mPid MC Truth	0.966	0.997	3.1%	0.950	0.995	4.5%	0.855	0.958	10.8%
mPid MC T&P	0.982	0.999	1.7%	0.958	0.998	4%	0.882	0.997	11.5%
mPid Data T&P	0.940	0.741	−26.9%	0.919	0.701	−31.0%	0.860	0.680	−26.5%
ePid MC Truth	0.837	0.944	11.4%	0.766	0.888	13.7%	0.681	0.801	15.0%
ePid MC T&P	0.885	0.982	9.8%	0.814	0.976	16.5%	0.708	0.977	27.6%
ePid Data T&P	0.799	0.41	−94.8%	0.716	0.337	−112.4%	0.614	0.274	−124.2%
eFwd	0.797	0.46	−73.3%	0.624	0.254	−145.6%	0.429	0.164	−161.1%
pPid	0.739	0.813	9.3%				0.526	0.522	−0.7%

Table 5.7: Performance of the Pid models for muons, electrons, and photons. Three (two) working points are shown for muons and electrons (photons). These are the Loose, Medium, and Tight (Loose, Tight). The signal efficiency is shown for the WPs, where the ML efficiency is measured at the same background efficiency as the ATLAS WP.

5.2.3.2 *Isolation: Iso*

The Iso models give us information about the isolation of the particle. Variables used here are energy and energy deposits in the detector. The training features for the mIso model can be seen in Tab. 5.8, while features for the eIso and pIso models can be seen in Tab. 5.9 (Tabs. A.3 and A.5 for full description). All models share similar features—the pileup, number of vertices and different energy measures for the particle.

The performances of the mIso models are compared to the isolation WP from [41], $p_T^{varcone30} < 0.06\, p_T$, and the eIso models to $p_T^{varcone20} < 0.15\, E_T$ from [67]. No direct comparison can be made for the pIso model, but for visualization purposes, the muon WP will be depicted on the photon isolation ROC curve.

ML mIso input features	
`correctedScaledAverageMu`	Pileup
`NvtxReco`	Number of reconstructed vertecies in the event
`muo_etcone20`	A variable of the sum of E_T tracks in the 0.20 cone around the object
`muo_pt`	The transverse momentum p_T of the muon
`muo_ptcone20`	A variable of the sum of p_T tracks in the 0.20 cone around the object
`muo_etconecoreConeEnergyCorrection`	The correction of the etcone variable

Table 5.8: Input features for the mIso models.

ML eIso input features	**ML pIso input features**
`correctedScaledAverageMu`	`correctedScaledAverageMu`
`NvtxReco`	`NvtxReco`
`ele_et`	`pho_et`
`ele_ptvarcone20`	`pho_ptvarcone20`
`ele_topoetcone20`	`pho_topoetcone20`
`ele_topoetcone40`	`pho_topoetcone40`

Table 5.9: Features for the eIso and pIso models. The list including description of the variables can be seen in Tabs. A.3 and A.5.

As for the Pid models, the SHAP values are investigated (see Fig. 5.9). Here, `muo_pt` has the largest importance in the mPid model training. Generally for all isolation models, the transverse momentum/transverse energy (p_T/E_T) variables are important together with the `topoetcone20` variable for electrons and photons. The topoetcone variable is the sum of the E_T tracks in a cone around the object, making it ideal as an isolation variable. The SHAP listings for the other isolation models can be found in App. A.2.

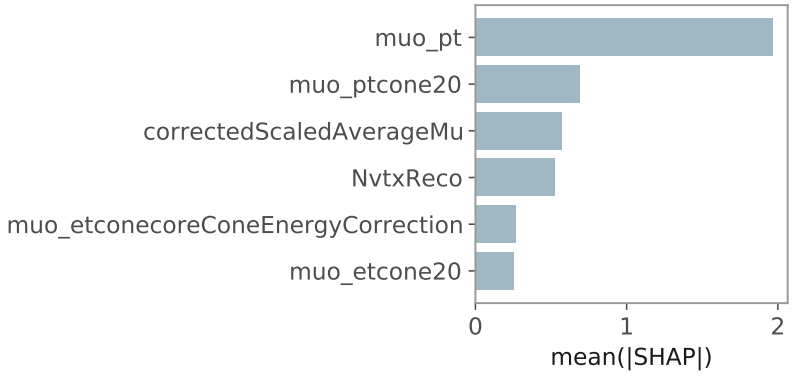


Figure 5.9: SHAP values for the mIso model. The variable with the highest influence on the model is `muo_pt`.

Fig. 5.11 shows the logit transformed LGBM score for the MC Truth trained mIso model, while the ROC curve for the test set can be seen in Fig. 5.10. The mIso model gets an AUC of 0.939, and, compared to the ATLAS WP plotted in Fig. 5.10, $p_T^{varcone30} < 0.06\, p_T$, yielding an AUC of 0.694, the mIso model performs well. For the T&P trained models, we get an AUC of 0.958 for the MC and 0.703 for the Data trained model.

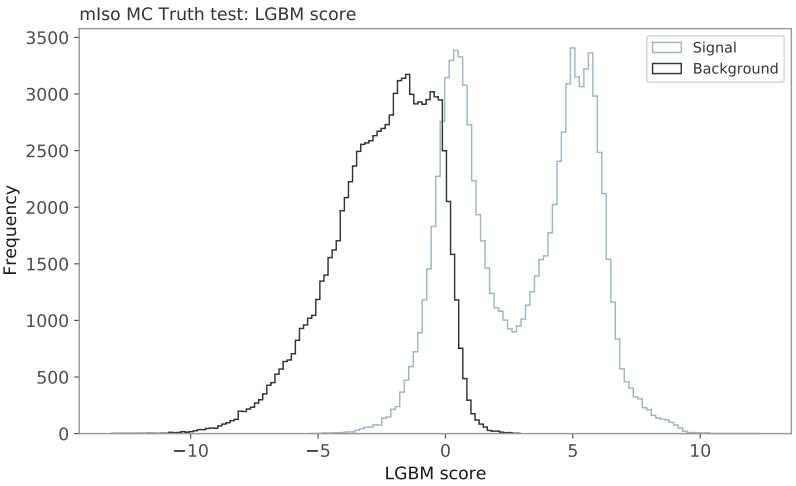


Figure 5.11: Logit transformed LGBM score for the final mIso MC Truth model. As can also be seen on the ROC curve in Fig. 5.10, the signal and background are less seperable than in the mPid model in Fig. 5.7.

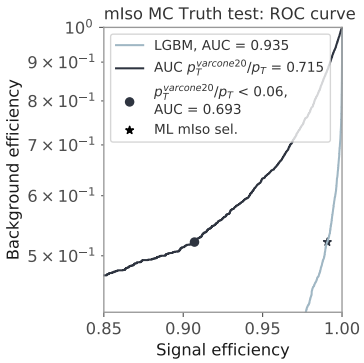


Figure 5.10: The ROC curve for the mIso MC Truth test set. The ATLAS WP $p_T^{varcone30} < 0.06\, p_T$ is shown on the figure.

The eIso MC models has AUCs of 0.986 and 0.991, while the Data trained eIso model has an AUC of 0.885, lower than its MC trained siblings, just as the mIso and Pid cases. Here—different from the Pid models—the Data-trained models performs better than ATLAS. The photon isolation has an AUC of 0.893.

The signal efficiencies and their improvement can be seen in Tab. 5.10, where it is evident that all model improves compared to the ATLAS WPs, varying between $6-12\%$. However, as can be seen on the slightly lower AUCs compared to those of the Pid models, isolation is a difficult task, and definitely a place where further

improvements can be made.

Note that the photon is evaluated against a muon WP and should thus not be seen as a direct comparison to the ATLAS photon isolation. See App. A.2 for all ROC curves and LGBM scores for the isolation models.

Iso models performance

Model	Isolation WP			
	ATLAS	ML	Change	Note
mIso MC Truth	0.907	0.991	8.5%	
mIso MC T&P	0.941	1.0	5.9%	
mIso Data T&P	0.899	0.975	7.9%	
eIso MC Truth	0.885	0.986	10.3%	
eIso MC T&P	0.876	0.998	12.3%	
eIso Data T&P	0.811	0.919	11.7%	
pIso	0.82	0.93	11.8%	Compared with muon WP

Table 5.10: Performance of the Iso models for muons, electrons, and photons. One isolation working points are shown. The signal efficiency is shown for the WPs, where the ML efficiency is measured at the same background efficiency as the ATLAS WP. Note that the photon is evaluated against a muon WP.

5.2.4 *Conclusions on lepton and photon models*

In the last section, the training and performance of the lepton and photon models has been described. Two types of models have been trained for all particles; a particle identification model and an isolation model. For the muon and electron, we have used three different datasets for training.

The Pid models have been compared to ATLAS Likelihood WPs in Tab. 5.7, where most muon and electron models have yielded similar or better results. Exceptions are the forward model for the electron and the Data trained T&P models. For the forward model, ATLAS does not even include the electrons in this part of the detector, and as mentioned, simply including the particles will be likely to yield an increase in performance. For the Data trained models, the low statistics might play a role in the bad performance. The pPid model shows an improvement compared to the Loose WP and a slight worsening for the Tight WP.

Isolation models have all been compared to isolation working points in Tab. 5.10, and we have seen an increase in performance for all models. The isolation models is generally a place where improvements could be made compared to the ATLAS selection.

Common to all ML trained models is that the model output is continuous opposite to ATLAS, where all WPs are binary. For the next part of the analysis, the continuous output is important, as we will use the models as input features for new models. This gives the

models the ability to find the best cut in the continuous feature rather than accepting the "cut" already made when working with a binary variable.

For all models, the performance is measured on the test set, where the labels are created in the same manner as the training set. However, we might find that the distribution of the test data does not match the final Data distribution that we are trying to predict upon. Thus we will still keep all types of models and test all in the next part of the data analysis.

5.3 Boson models

In this section the training and evaluation of the boson models will be presented. As described in Sect. 5.1, these are the **Zll** (di-lepton models) and **Zllg/Hllg** models. The decays and the model names are listed in Tab. 5.11.

Boson decays	Model names
$Z \to ee$	Zee (Zll)
$Z \to ee\gamma$	Zeeg (Zllg)
$H \to Z(\to ee)\gamma$	Heeg (Hllg)
$Z \to \mu\mu$	Zmm (Zll)
$Z \to \mu\mu\gamma$	Zmmg (Zllg)
$H \to Z(\to \mu\mu)\gamma$	Hmmg (Hllg)

Table 5.11: The boson decays investigated in Sect. 5.3

The Zmm and Zee models will be trained on three different types of datasets, as for the lepton models, the MC Truth, MC T&P, and Data T&P. The Zllg and Hllg models will be trained on MC Truth only.

The models for the $Z \to ll\gamma$ decays serves as control models for the $H \to Z(\to ll)\gamma$ decays, since the Z-decays are discovered and well-investigated—opposed to the H-decay. The Z and Higgs decays will be compared in MC and the H Data behavior will be extrapolated from the Z Data performance.

The performance of the ML models will be measured with respect to the ATLAS selection from [1] as described in Sect. 5.1 and summarized in Tab. 5.12. The ATLAS and ML selections will be shown in the figures and the improvement achieved by the ML models will can be seen in the figure legend as well as tables.

ATLAS selection	
Muons (electrons)	$Q_1 \cdot Q_2 < 0$ $p_T > 10\,\text{GeV}$ $\lvert\eta\rvert < 2.7$ (2.47, with $1.37 < \lvert\eta\rvert < 1.52$ excluded) LH medium (LH loose) $\lvert d_0\rvert/\sigma_{d_0} < 3$ (5) $\lvert\Delta z_0 \cdot \sin\theta\rvert < 0.5\,\text{mm}$ track isolation
Photons	$p_T > 10\,\text{GeV}$ $\lvert\eta\rvert < 2.37$, with $1.37 < \lvert\eta\rvert < 1.52$ excluded LH Tight

Table 5.12: Selection from [1]. See Tab. 5.1 for an elaboration of the cuts.

The section will be structured as follows: First, the data production and selection will be introduced. This also includes the evaluation methods used throughout the section. Then, the models will be presented: First, the Zll models are introduced and their performance in MC and Data is presented and discussed. The presentation of the Zllg and Higgs models will follow. The behavior of the Zllg models in Monte Carlo and Data from the ATLAS detector will be reviewed, before the Higgs models are evaluated in MC and their expected performance in Data is discussed.

5.3.1 Datasets

We are using datasets with decay products from the Z and Higgs boson. This means that we are creating datasets with multiple particles and their variables. Event variables are calculated, such as the energy, η, and invariant mass of the combination of products—thus the origin particle.

As mentioned, three types of datasets will be created for training the Zee and Zmm model. These are—again—the MC Truth, MC T&P, and Data T&P. The procedure has some similarities with the lepton model datasets. The difference here is, that we need to keep both the tag and probe particle, and thus we cannot perform as many checks on the tag as we could when this was discarded, since this will bias our data selection.

First, the procedure for production of datasets for the Zee and Zmm model will be described and afterwards we will do the same for Zllg and Hllg.

5.3.1.1 *MC Truth datasets for Zee and Zmm*

When creating the MC Truth datasets, we again rely on the truth variables. But first, we create a pair using the tag and probe-method. We loop through all particles, and if the particle activated the trigger, it is labeled as *tag* and if not, it is a *probe*, see Fig. 5.12.

The tag and probe candidates are combined in pairs and their truth variables are checked. The pair is labeled as signal if both particles have truthOrigin 13 (Z boson), if their truthPgdId match the correct (± 13 for muons and ± 11 for electrons) and if they have opposite charge. They are labeled as background if they match one of the following conditions:

Bkg0Lep: Both are muons (electrons), but not from Z,
Bkg1Lep: One is a muon (electron) or
Bkg2Lep: No muons (electrons).

Particles where one lepton origin from the Z boson will be discarded. The largest contribution to the background is Bkg1Lep, where one of the tag and probe candidates are the correct lepton. This will most likely be the tag lepton, as it has triggered.

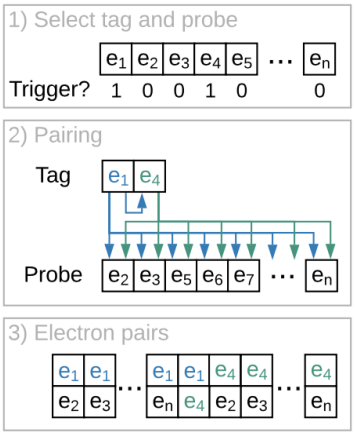


Figure 5.12: Illustration of the T&P pair selections from [59]. 1) All electrons in the event are seperated into tag and probe using trigger information. 2) The tag electrons are combined with the probe electrons. 3) The electron pairs are outputted in array.

5.3.1.2 *Tag & Probe datasets for Zee and Zmm*

For the tag & probe datasets, we find tag and probe pairs by checking that the tag particle has triggered, and we then demand the probe to have $p_T\,(E_T) > 10\,\text{GeV}$ for muons (electrons).

In MC, the signal is found from signal files (as for lepton models, see Tab. 5.2) where we check the invariant mass is in the range $|m_Z - m_{ll}| < 5$ and that the particles have opposite sign.

For the background, we consider the background files in Tab. 5.2 and accept same-sign candidates in $10 < |m_Z - m_{ll}| < 40$ and opposite-sign candidates with $20 < |m_Z - m_{ll}| < 40$. This also means that we only have data in these ranges; $m_Z \pm 5\,\text{GeV}$ and $m_Z \pm (10 - 40)\,\text{GeV}$.

For the Data T&P data set, we find signal and background considering all events in the data file and enforcing same cuts as for the MC.

As for the lepton models, the data is augmented by adding more background by copying shuffled events. This means that the difference in from the T&P datasets for single leptons, is, that we do not check if the tag particle is tight, isolated, and have $p_T\,(E_T) > 26\,\text{GeV}$ and that both the tag and probe particles are kept.

Tab. 5.13 shows the number of events for all different datasets for the Zll models. These are used for training, validation, and test with the split 60-20-20.

ML Zll data selection

Decay	MC Truth		MC T&P		Data T&P	
	Signal	Background	Signal	Background	Signal	Background
$Z \to \mu\mu$	809,151	1,800,617	73,376	315,875	19,228	75,460
$Z \to ee$	126,480	306,778	82,983	748,875	19,988	31,327

Table 5.13: Number of events for the Zll data sets. The rows shows the two different decays. The columns are split into the MC Truth, MC T&P, and Data T&P datasets, and numbers of signal and background events for the different datasets are shown.

5.3.1.3 *MC datasets for Zllg and Hllg*

The data selection for Zllg and Hllg is based on the truth variables and the tag and probe method. Since we are considering three decay particles, there will be one tag and two probes (TP&P; tag, probe, and probe). As for the Zll models, all leptons triggering will be considered as tags while the rest are considered as probes. All photons with $E_T > 4.5\,\text{GeV}$ will be labeled as probes.

To label the events, all three particles in the event are considered. For signal leptons, their origin should be 13 (14) for Z (Higgs), truthPdgId ± 13 (± 11) for muons (electrons), and have opposite charge. The signal photon should have origin 3 (14) for Z (Higgs) and truthPdgId 22. There will be the following different combinations for the signal and background:

Signal: All particles meets the signal requirements
Bkg0: The leptons meets the signal requirements, but the photon does not
Bkg1: The leptons does not meet the signal requirements, but the photon does
Bkg2: None of the particles meets the signal requirements

The main background will be from a $Z \to ll$ decay with a random photon for both the Zllg and Hllg data files, Bkg0. A cut will be performed for the Zllg datasets in the invariant mass—we will require $40 < m_{ll\gamma} < 83\,\text{GeV}$ as in [68]. For both the $Z \to ll\gamma$ and $H \to Z(\to ll)\gamma$ decays, a cut on the photon energy, $p_T^\gamma > 10\,\text{GeV}$, will also be performed.

Tab. 5.14 shows the selection for the Zllg and Hllg datasets including cuts. Here we see that for the Zllg models, the cut in the invariant mass reduces the signal to around 75% while reducing the background to $26 - 28\%$. The photon cuts further reduces the amount signal events to around 60% and the background to $3 - 4\%$ of the original size. For the Higgs datasets, the photon cuts only reduces the signal amount to 92% while the background is heavily reduced to $13 - 14\%$. From the data, the training, validation, and test sets are created, again using the split 60-20-20.

5.3.1.4 *Datasets for testing in Data*

We wish to test both the Zmm, Zee, and Zllg models on Data and need a test dataset. We find tags and probes using the trigger particles

ML Zllg and Hllg data selection

Decay	N before cuts		N after cuts		Cuts applied			
					$40 < M_{ll}[\text{GeV}] < 83$		$p_T^\gamma > 10\,\text{GeV}$	
	Signal	Background	Signal	Background	Signal	Background	Signal	Background
$Z \to \mu\mu\gamma$	237,725	4,289,018	142,404	168,049	75.1%	38.2%	59.9%	3.9%
$Z \to ee\gamma$	113,156	5,860,330	67,922	174,223	75.3%	26.4%	60%	2.9%
$H \to Z_{(\to\mu\mu)}\gamma$	84,101	2,171,535	77,736	290,949			92%	13.3%
$H \to Z_{(\to ee)}\gamma$	82,572	6,728,870	75,927	972,480			92%	14.4%

Table 5.14: Number of events for the Zllg and Hllg datasets. Each row corresponds to a different decay. The first two columns show the absolute number of events in the data sets before and after the cuts are applied. The cuts and their percentage of the data left after cutting, are shown in the last two columns of the table. The cut in the invariant mass is not relevant for the Higgs datasets. A table with absolute numbers on the cuts can be seen in Tab. A.8.

as for the MC datasets. Only photons with E_T above 10 GeV will be considered. The lepton and photon candidates are combined into triplets without further checks.

For the Zllg data, the cut on the $Z \to ll$ mass will be made as for the MC data.

Tab. 5.15 shows the number of events for each Data file before and after the different cuts. No cuts are performed on the Zmm and Zee datasets, so the table simply shows the number of events.

Zll(g) Data selection

Decay	N before cuts	N after cuts	Cuts applied [GeV]	
			$40 < M_{ll} < 83$	$p_T^\gamma > 10$
$Z \to \mu\mu$		145,815		
$Z \to ee$		1,099,400		
$Z \to \mu\mu\gamma$	12,791,783	424,880	29.5%	3.3%
$Z \to ee\gamma$	21,176,293	493,441	24.3%	2.3%

Table 5.15: Number of events for the Zll(g) Data data sets. Each row shows the two different decays. The first two columns show the absolute number of events in the data sets before and after the cuts are applied. The cuts and their percentage of the data left after cutting, are shown in the last two columns of the table. No cuts are made on the $Z \to \mu\mu$ and $Z \to ee$ decays, so the table simply shows the total number of events in the Data file.

5.3.2 *Reweighing*

As for the lepton and photon models, the data will be reweighed before training. Background events are given weights to ensure that the signal and background distribution matches for the variables $\langle\mu\rangle$, p_T, η, and $m_{Z/H}$. Fig. 5.13 shows the reweighed data for the Zee MC Truth model using 40 estimators.

5.3.3 *Evaluation methods*

We will evaluate all models on both MC and Data. In Monte Carlo, truth variables are available allowing for easy evaluation of the models. However these are not available in Data, so alternative methods are needed. When evaluating the MC or Data T&P methods, we will still use the labeling of the events as an error measure. It will be considered how many background-labeled events are passed in the selection (false positives). This evaluation will be made alongside the following described below, which can also be applied to Data where no labels exist. We do so, to check if the method of evaluation is consistent with our labeling.

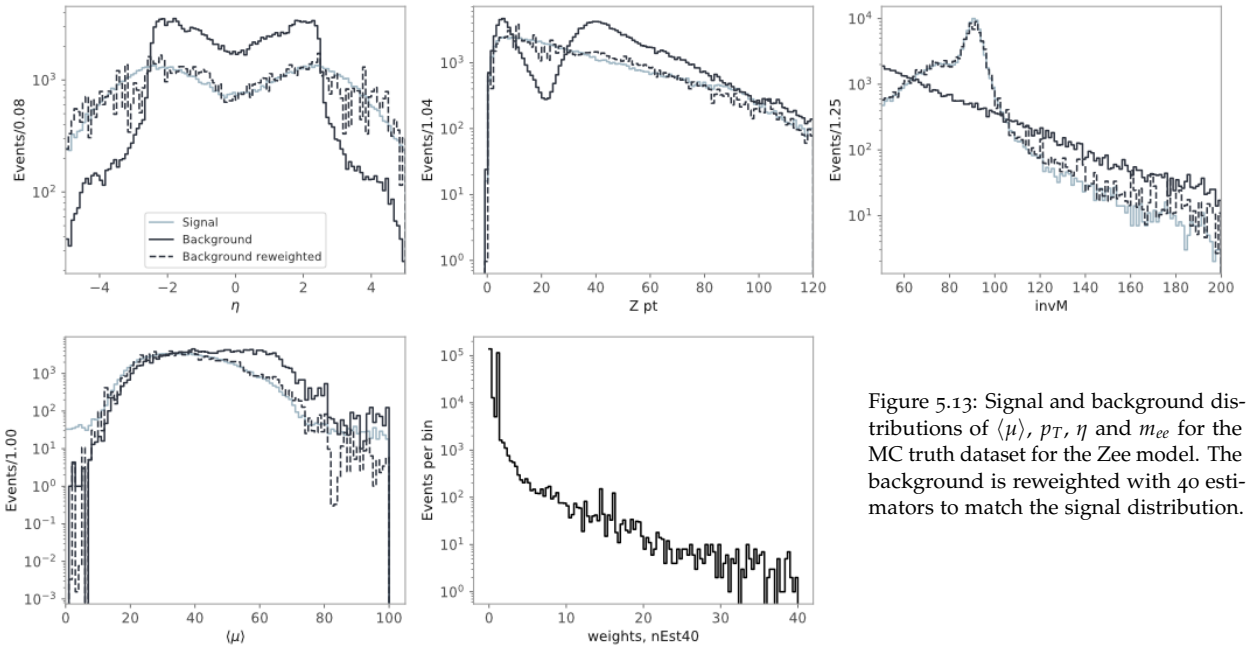


Figure 5.13: Signal and background distributions of $\langle\mu\rangle$, p_T, η and m_{ee} for the MC truth dataset for the Zee model. The background is reweighted with 40 estimators to match the signal distribution.

One method of evaluation is by looking at the same-sign data. We know that the leptons decaying from either the Z or the Higgs boson will be of opposite sign, to make up for the zero electric charge of the bosons. We will consider the amount of same-sign data in the tails of the signal peak in invariant mass (see Fig. 5.14).

For the Z and H decays, the signal will be calculated as the number of opposite-sign leptons in the signal peak, with $70 < m_{ll(\gamma)} < 110$. And the background, the number of same-sign leptons for $15 < |m_{ll(\gamma)} - m_Z| > 40$. For the Higgs decays, the number of same-sign leptons will be counted in the the tails $(m_H - 25) < m_{ll\gamma} < (m_H - 15)$ and $(m_H + 40) > m_{ll\gamma} > (m_H + 15)$, that is, in the range $[100; 115]$ and $[140; 165]$. The signal will be counted as the opposite-sign events in $115 < m_{ll\gamma} < 135$. I will refer to this evaluation method by the *same-sign* method.

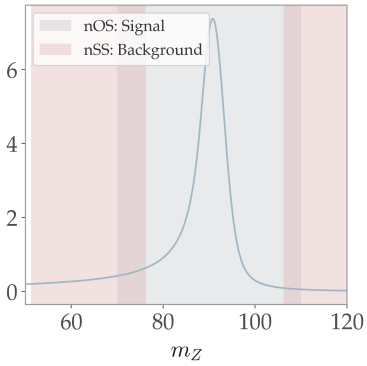


Figure 5.14: The signal and background calculation for the nSS method.

Another way to evaluate the performance of the model is by fitting the peak and estimating the number of signal and background events from the fit. The framework *RooFit* developed by ROOT, is used. For Z, the signal will be fitted with a Breit-Wigner function convoluted with a Crystal-Ball (BWxCB). For the Higgs decay, the signal is fitted with a Crystal-Ball only. The background will be fitted with an exponential decay.

An example of a fit using the simple background model can be seen in Fig. 5.15. This is the $Z \to \mu\mu$ decay. The fitting variables estimate the fraction of background to be $f_{bkg} = 0.5284$, yielding $N_{bkg} = 317,400 \cdot 0.5284 = 167,714$ and $N_{sig} = 317,400 \cdot (1 - 0.5284) = 149,686$. The example here shows the full dataset with no cuts in the background, thus the high rate of background. I will refer

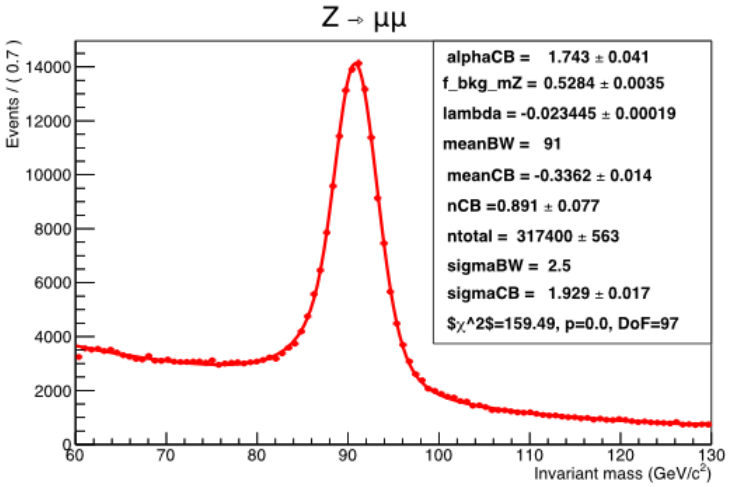


Figure 5.15: A fit of the $Z \to \mu\mu$ decay using RooFit. Background is fitted with a exponential decay and signal is fitted with a Breit-Wigner function convoluted with a Crystal-Ball. Fitting estimates fraction of background f_{bkg} to be 0.5284 and the total number of events 316,400, yielding $N_{bkg} = 317,400 \cdot 0.5284 = 167,714$

to this evaluation method by the f_{bkg} method.

The third performance measure will be to count the number of events in the high-mass tails. This simply counts the number of events above 110 GeV (150 GeV) for the Z (H) decays. The background is counted in the high-mass tail rather than the low mass tail, as the Z peak will have (signal) events in the low mass tail. These exist due to the Z/γ^* relationship and are avoided in the high mass tail. This method will only be used for the Zllg and Hllg models and will be referred to as the *bkg-count* method.

We will use these evaluation models for both the MC and Data for the boson models to be able to compare the performance. In both cases, we will compare with the ATLAS cut. Thus, we will aim for the same number of same-sign leptons, bkg events in the high-mass tail, or estimated number of background events by the fit as ATLAS gets in their selection based on the same type of evaluation. I.e. the events passed in the ATLAS selection are fitted, estimating the amount of background in the selection. Then, the ML selected events are fitted and it is aimed to get the same amount of background as the ATLAS selection.

The LGBM models give each event a score, which indicates the probability of that event being either signal or background. To know at which threshold the model will have the same amount of background compared to the ATLAS selection, a series of cuts will be made in the score. The cut giving the result closest to the desired is chosen and the search is terminated. This is not necessarily the same cut for the methods of evaluation. See Fig. 5.16 for an example of a LGBM score distribution and a series of cuts. The cut giving the desired output (marked with a solid line) is chosen.

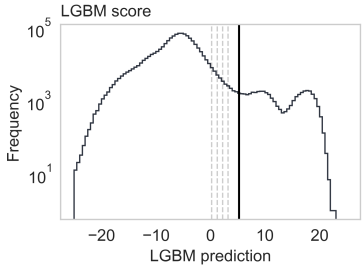


Figure 5.16: Example of a series of cuts in the LGBM score

5.3.4 *Model training and testing*

Now I will move on to describing how the models are trained and their performance is in both MC and Data. I will start with the Zmm and Zee models. Since the approach is similar for both models, these will be described together. Afterwards we will move on to the Zllg and Hllg models.

5.3.4.1 *Zmm and Zee training*

The training features for the Zmm and Zee models can be seen in Tab. 5.16. All features with "N", exist for both leptons, and is thus referring to the first or second lepton in the event. Both models are trained with similar variables except for the `ele2_isFwd` feature, existing for the Zee model only. If the probe electron is from the forward calorimeter, the ePid score for the probe electron will be replaced with the eFwd score. Tag electrons cannot be from the forward calorimeter as no triggering is available here.

ML Zll input features

`NvtxReco`	`NvtxReco`	The number of reconstructed verticies
`Z_sig`	`Z_sig`	$Z_{sig} = (l_{1,z0} - l_{2,z0}) \cdot \left(\sqrt{(l_{1,\sigma_{z0}})^2 + (l_{2,\sigma_{z0}})^2}\right)^{-1}$, where l is the relevant lepton
`muoN_PID_score`	`eleN_PID_score`	The mPid and ePid score for both leptons in the event. If the probe electron is from the forward calorimeter, the Pid score for the probe electron will replaced with `ele_Fwd_score`
`muoN_ISO_score`	`eleN_ISO_score`	The mIso and eIso score for both leptons in the event
`muoN_priTrack_d0`	`eleN_d0`	d_0 for the lepton
`muoN_priTrack_d0Sig`	`eleN_d0Sig`	σ_{d_0} for the lepton
	`ele2_isFwd`	Is the probe electron from the forward calorimeter? Not relevant for the muons

Table 5.16: Input features for the Zmm and Zee models.

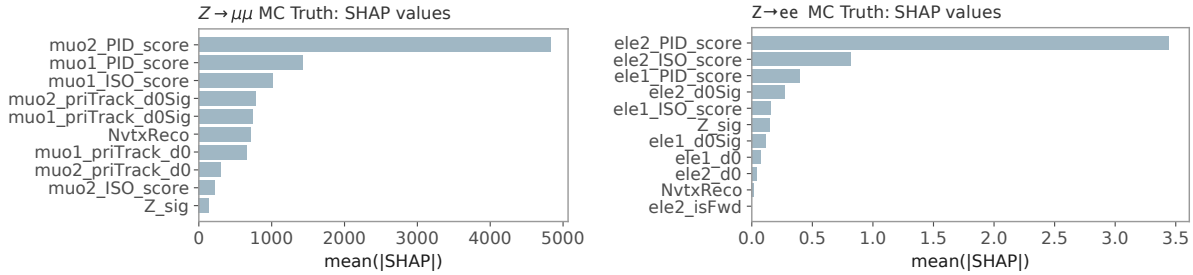


Figure 5.17: SHAP values for the Zmm (left) and Zee (right) models.

Fig. 5.17 shows the SHAP values for the Zmm and Zee models. For both models, the Pid score for the second lepton has the highest influence on the model prediction. When creating the dataset, the tag lepton gets the number "1" and the probe, "2". We are almost certain that lepton 1 is a good lepton from a Z, since it has triggered, and thus if the second lepton has a good Pid score, they are most likely

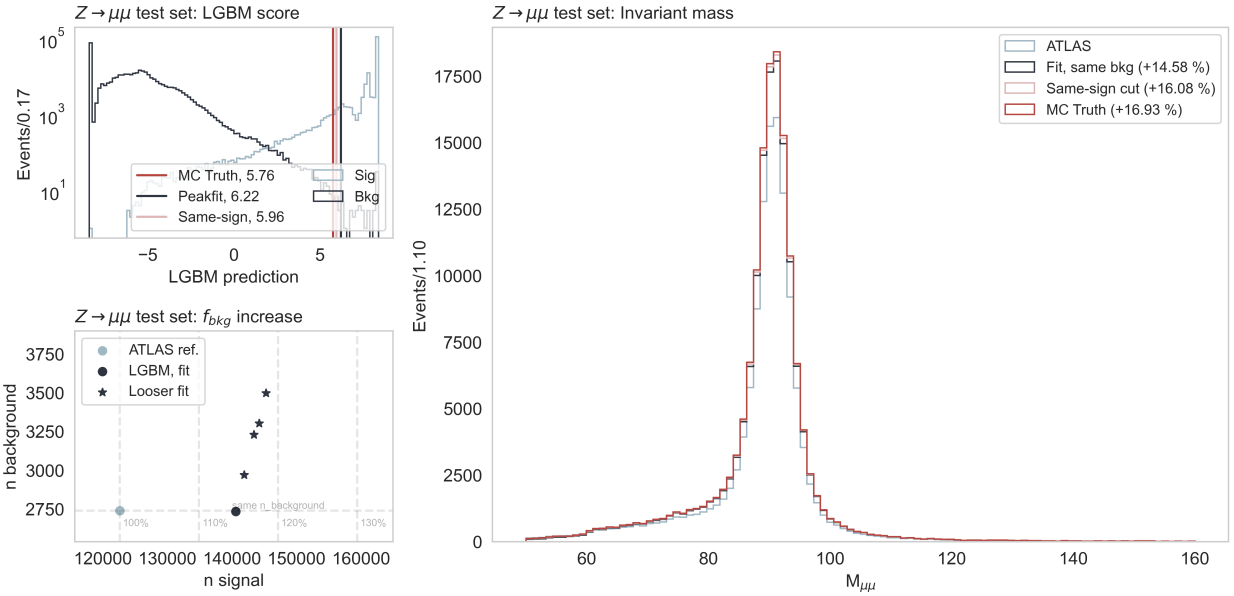


Figure 5.18: Top left: Plot of Zmm test set LGBM score and the corresponding cuts. Right: Plot of the invariant mass of the $Z \to \mu\mu$ decay. Bottom left: Plot of the signal vs. background for the f_{bkg} method. Also shown with looser cuts. The plot of the invariant mass is shown on log scale in Fig. A.24.

both originating from a Z boson, indicating that the Pid for the first particle is not as important as the Pid for the second particle. It could be interesting to see how the model performance would change if the leptons were shuffled.

5.3.4.2 *Zmm and Zee testing in MC*

The trained models are evaluated on the MC test set and on Data. We will use the same-sign and f_{bkg} evaluation methods on the MC data as well as the truth, referring to the labeling of the data.

Fig. 5.18 shows the LGBM score and the invariant mass for the Zmm model. On the figure of the LGBM score, we see the cuts of the two different evaluation methods. The same number of same-sign events yield an increase in the signal peak of 16.1%. The fit has a slightly stricter cut, yielding increase of +14.6% compared to ATLAS. The figure also shows the MC Truth cut. This is the true increase in signal based on the MC labeling of the data, where we get +16.9%, similar to the same-sign cut.

The cuts are the ones that achieve the same background as ATLAS. In Fig. 5.18 bottom left, the f_{bkg} cut is loosened to see how much signal we can get if we don't restrict our selves on the number of background events. The same plot can be found for the n_{SS} method in Fig. A.23. For both methods, the number of background events rise a lot when loosening the cut. Some further signal is achieved.

In Fig. 5.19, the same plots are plotted for the Zee model. The two methods yields increase in signal of 22.5% and 23.4% for the fit and same-sign methods, which is more than the Zmm model. The increase for MC Truth is close to both methods method with +23.2%.

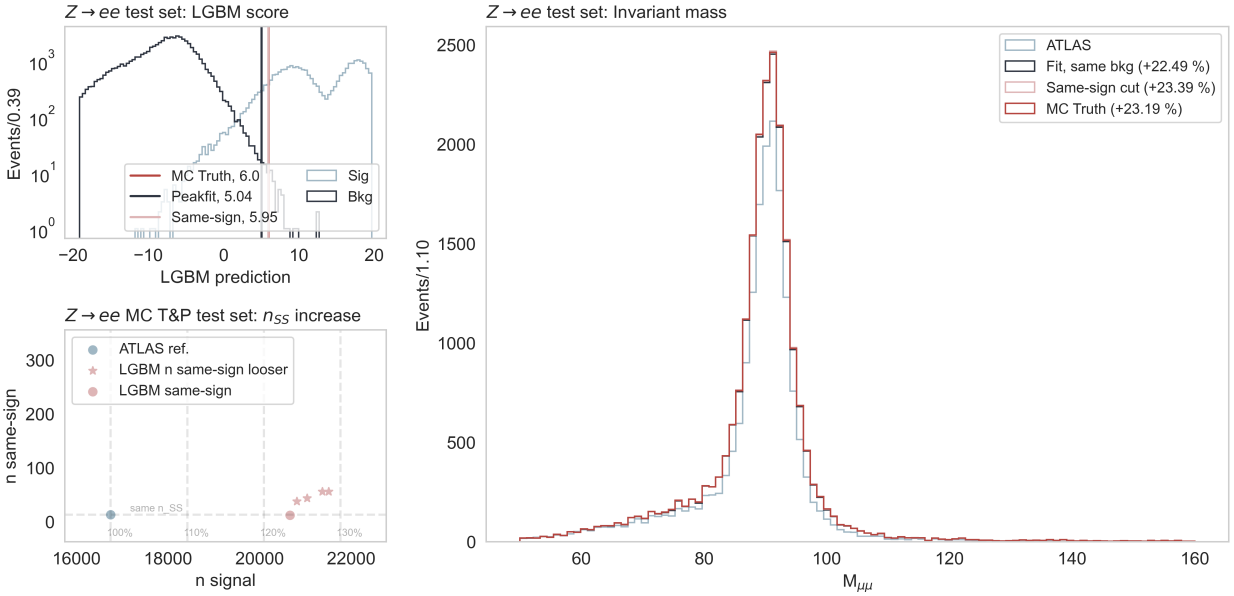


Figure 5.19: Top left: Plot of Zee test set LGBM score and the corresponding cuts. Right: Plot of the invariant mass of the $Z \to ee$ decay. Bottom left: Plot of the signal vs. background for the f_{bkg} method. Also shown with looser cuts. The plot of the invariant mass is shown on log scale in Fig. A.24.

On the LGBM score plot in Fig. 5.19 top left, we see that especially the same-sign cut is well into the signal-part of the LGBM score, and on Fig. 5.19 bottom left and Fig. A.23, we can see that more signal can be achieved without adding much more background.

The corresponding plots for the MC T&P and Data T&P models can be found in Apps. A.3.2 and A.3.3. The Zmm T&P trained models improves the selection in the test set with $10 - 11\%$ for the Data trained and 19% for the MC trained models. For both models, the most important parameter according to the SHAP values is the mPid score for the second muon, as was also seen in Fig. 5.17 for the MC truth trained model. For Zee T&P trained models, the increase for the Data trained T&P model is ranging from $21 - 27\%$ and $24 - 30\%$ for the MC trained model. The SHAP values related to their training is, as the muon models, also similar to the MC Truth trained Zee model.

Tab. 5.17 shows an overview of the performance for the labeling of the data while the performance for the n_{SS} and f_{bkg} evaluations can be seen in Tabs. A.6 and A.7.

The Zll models are later being used for the $Z \to ll\gamma$ decays. The di-lepton pair in this decay will have an invariant mass below m_Z, and it thus relevant to look at how the performance changes at lower invariant mass. Figs. 5.20 and 5.21 shows the efficiency of the Zll MC Truth models as a function of the invariant mass. The shaded area marks m_Z, where both the LGBM and ATLAS models have the highest signal efficiency.

For the Zmm model, the efficiency of the LGBM models is high for $M_{\mu\mu} > 70$ but for low mass, the efficiency drops.

For the Zee model, the efficiency of all models has a slight drop in

Zll models performance, truth labeling					**MC**
Model	**N ATLAS**		**N LGBM**		**Change**
Zmm	Sig	Bkg	Sig	Bkg	
MC Truth	119,328	23	139,525	24	16.93%
MC T&P	12,232	2	14,599	3	19.35%
Data T&P	3,183	438	3,521	439	10.62%
Zee					
MC Truth	16,745	16	20,750	15	23.19%
MC T&P	12,434	495	16,255	497	30.73%
Data T&P	2,841	182	3,441	183	21.12%

Table 5.17: Performance of the Zll models. The number of signal events are measured in the range [70;110] GeV while the number of background events are measured for [50;150] GeV.

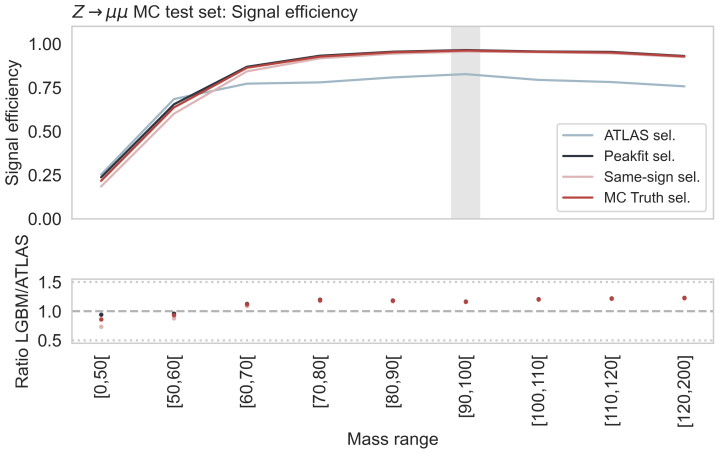


Figure 5.20: Efficiency of the Zmm model evaluated as a function of the invariant mass. The shaded area marks m_Z.

the range [100;110], where the ATLAS selection has the biggest drop. This is an interesting feature, that has also been seen by previous students in the group [59]. The most relevant range to consider in both the Zee and Zmm plots, however, is below the Z mass, as this where the invariant mass of the di-lepton pair in the Zllg models will be.

For Zee, The LGBM model has a higher efficiency for all ranges, however they are comparable for [50, 60] GeV. For Zmm, the efficiency of the LGBM model is just below the ATLAS efficiency for [0, 60] GeV. The drop for low mass in both Zmm and Zee, might be due to limited amount of training data in low mass. Data from the $Z \to ll\gamma$ decay is included in training, to get low-mass signal, however, the drop in performance could indicate that the statistics are still too low for low-mass. The invariant mass of the di-lepton pair in the $Z \to ll\gamma$ decay is assumed to mainly be in the range [60, 80] GeV which is thus the most important mass range to consider. It is important to note, that the efficiencies of the Zmm and Zee models are measured at the *specific* cut in LGBM score that gives the same amount of background as the ATLAS cut. Loosening the cuts is likely to yield different distributions.

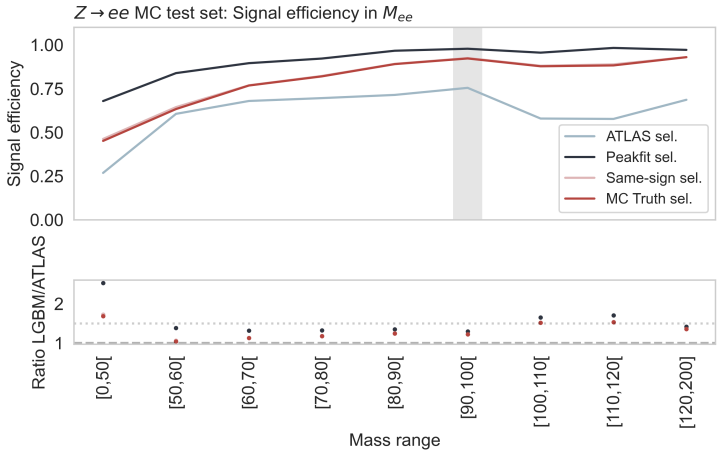


Figure 5.21: Efficiency of the Zee model evaluated as a function of the invariant mass. The shaded area marks m_Z, where it is evident that both the LGBM and ATLAS selections have the highest efficiency.

Fig. 5.23 shows the efficiency of the Zee model as a function of $|\eta|$ to give us information on where the increase of the model is gained. The ATLAS selection has a drop in the crack (they do not include events in $1.37 < |\eta| < 1.52$ at all) and goes to zero where the forward detector starts ($|\eta| > 2.47$). The efficiency for the LGBM model drops in the forward part of the detector, but some signal events are caught. Thus, one can see, that the increase is gained both from crack and forward, but some also from the fiducial part of the detector. The same distribution for Zmm is shown in Fig. 5.22, where no special features are seen across the ranges of $|\eta|$.

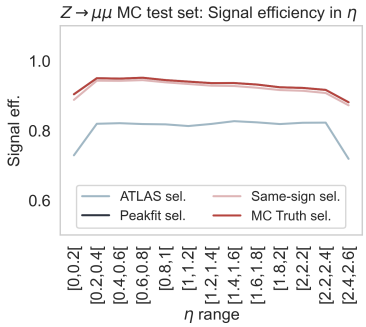


Figure 5.22: Efficiency of the Zmm model evaluated as a function of $|\eta|$. The efficiency is stable for the full range of $|\eta|$, only a small drop for high and low η, most significant for the ATLAS selection.

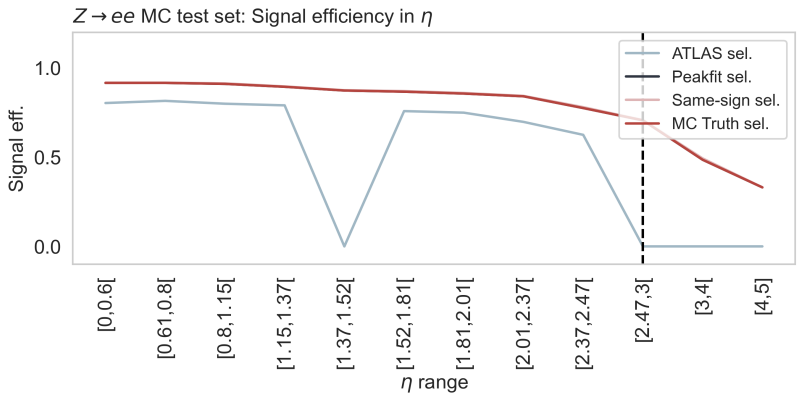


Figure 5.23: Efficiency of the Zee model evaluated as a function of $|\eta|$. The ATLAS efficiency is zero in the crack between HCAL and ECAL and after the beginning of the forward detector, marked by the dashed line. Efficiencies for all MC selections are on top of each other.

5.3.4.3 *Zmm and Zee testing in Data*

We'll now consider the performance of the Zll models in Data. Fig. 5.24 shows the invariant mass of the Zmm model inlcuding increases achieved by the ML models. Fig. 5.25 shows the same for the Zee model. The f_{bkg} evaluation and n_{SS} are separated into two plots showing all three models together for clarity. For examples of the fits, see Fig. A.25, where the fits of the ATLAS and Data T&P selections for the Zmm model are shown.

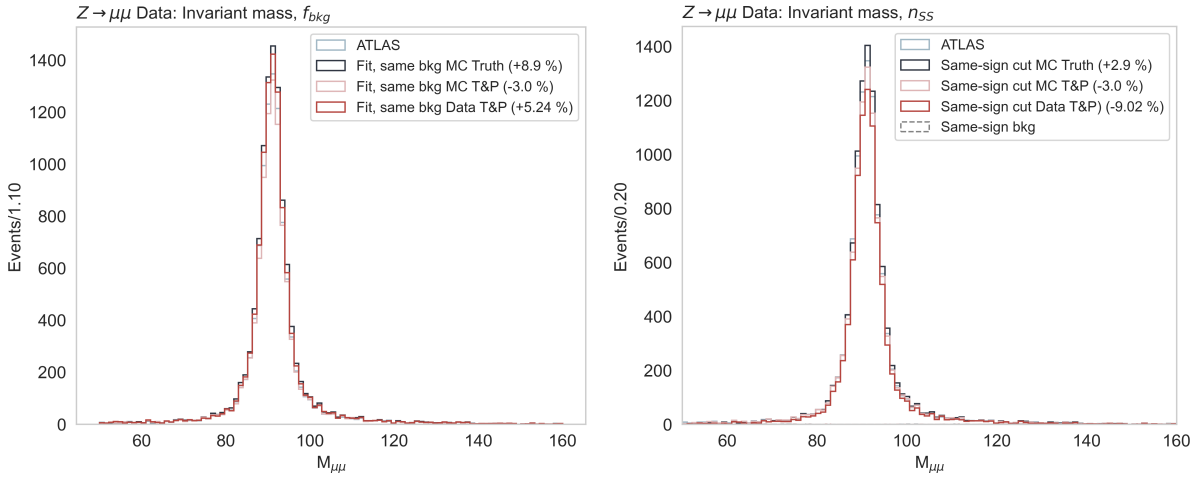


Figure 5.24: Plot of the invariant mass for the $Z \to \mu\mu$ decay in Data. The left plot shows the MC Truth, MC T&P, and Data T&P performance for the f_{bkg} evaluation. The right plot shows the n_{SS} evaluation. Here, the very few same-sign events are also plotted on the figure. They are better visualized in Fig. A.26, which is the same plot in logarithmic scale.

For both decays, the MC Truth model is the best performing one. It yields an increase in signal at the same amount of background as ATLAS for the $Z \to \mu\mu$ decay of 8.9% and 2.9% for the fit and same-sign cut respectively and 20.1% and 11.9% for the $Z \to ee$ decay.

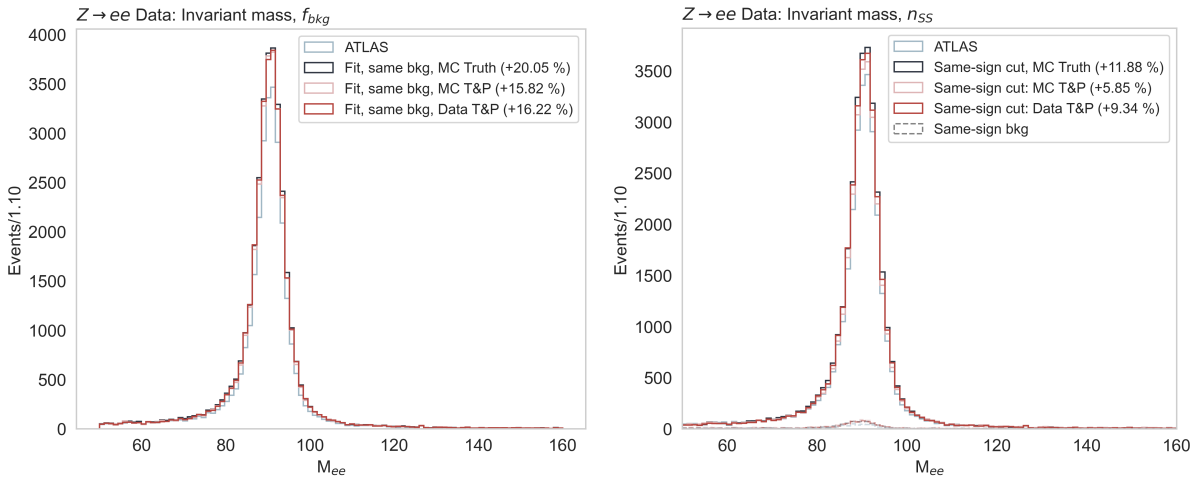


Figure 5.25: Plot of the invariant mass for the $Z \to ee$ decay in Data. The left plot shows the MC Truth, MC T&P, and Data T&P performance for the f_{bkg} evaluation. The right plot shows the same models but for the n_{SS} evaluation. Here, the same-sign events are also plotted on the figure. The same plots in logarithmic scale can be found in Fig. A.27.

The improvement for the Zmm model is lower in Data compared with the MC Truth test set. For the same-sign method, this is especially low. ATLAS has 7 same-sign events and 9,423 opposite-sign events, making the selection extremely effective. The two T&P methods, especially the MC trained one, shows very bad performance in Data.

Fig. 5.26 shows the LGBM distribution for the MC T&P trained model in MC and Data (the MC T&P-trained model predicted on Data). This figure shows that the Data distribution does not match the MC distribution at all, indicating that the training data is not a good approximation to the Data we are predicting upon. The same-sign cut in Zmm M&P includes 16 same-sign events compared to ATLAS' 7, but no further cuts can be made for the MC T&P model, as 16

same-sign events has a LGBM score of 1.

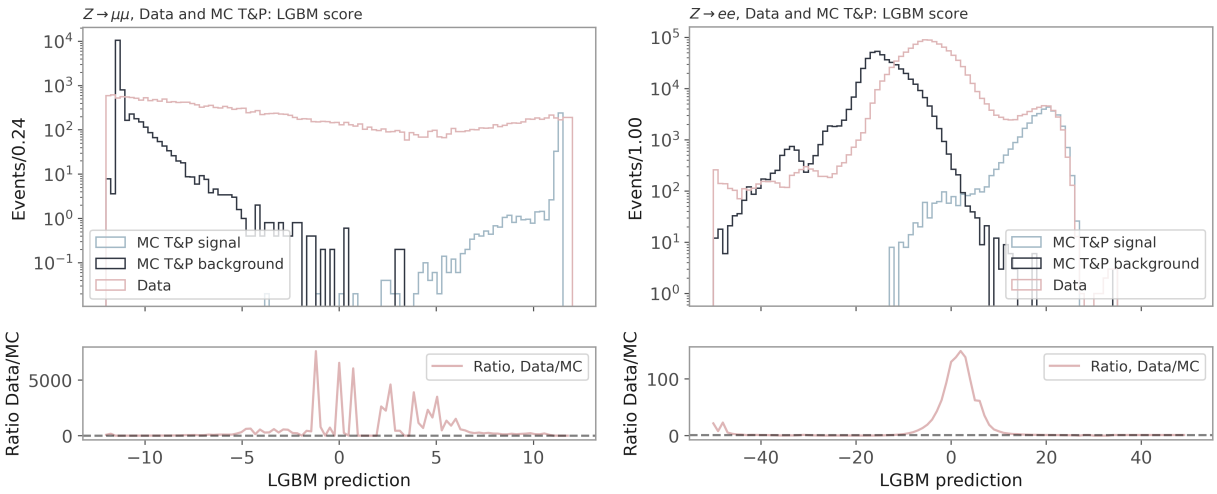


Figure 5.26: The LGBM score of the MC T&P trained models plotted in the MC test set and the Data. The MC data is weighted in signal and background separately, in order to compare the distribution to Data. Left shows the Zmm model while right shows the Zee model.

The Zee improvement in Data for the same-sign cuts is lower than the improvement in MC while comparable for f_{bkg} cuts in Data. Fig. A.37 shows that more signal can be gained from loosening the cuts, where doubling the background yields +25% for both same-sign and f_{bkg} methods for the MC Truth model. The MC T&P selection for Zee has the lowest improvement. Looking at Fig. 5.26, the LGBM score matches the predicted Data better than was seen for the Zmm MC T&P model, but it seems that the background distribution in MC gets a score of 10 lower than the Data. This indicates that the signal and background is not as well-separated in Data as we saw in MC, accounting for the lower improvement.

Tab. 5.18 shows the number of signal and background events for each model and method in Data with corresponding increase in signal compared to the ATLAS selection. For the Zee model, data

Zll models performance								**Data**
Eval	**ATLAS**		**MC Truth**		**MC T&P**		**Data T&P**	
Zmm	Sig	Bkg	Sig	Bkg	Sig	Bkg	Sig	Bkg
nSS	9,423	7	9,696 (+3%)	8	9,140 (−3%)	16	8,573 (−9%)	8
f_{bkg}	9,423	392	10,262 (+9%)	398	9,140 (−3%)	393	9,917 (+5%)	400
Zee								
nSS	27,320	173	30,566 (+12%)	174	28919 (+6%)	177	29872 (+9%)	178
f_{bkg}	27,282	1,110	32,108 (+18%)	1,116	31,237 (+15%)	1,113	31,581 (+16%)	1,108

Table 5.18: Performance of the Zll models in Data. For the same-sign method, the number of signal events are measured in the range [70; 110] GeV while the number of background events are measured for [50; 150] GeV. For the fitting, the number of signal and background events are estimated by the fit in the range [60; 130] GeV.

from the forward container has been included in both the training data and prediction Data. It should be noted that only a very small data sample was available in MC, with only 6000 signal events of the 126,000 total signal events in the Zee decay. In Data, no forwards events were available, and the dataset was thus acquired by applying the EGAM8 derivation to the Data AOD, as described in Chap. 4. Still,

the forward events are only 4200 out of the total 1,099,000 events in the Data sample. Thus, it is important to remember that an even stronger increase could be achieved, since we saw in Fig. 5.23 that the Zee model has around 50% signal efficiency in this area of the detector where ATLAS has zero.

Fig. 5.27 shows the LGBM distribution for MC and Data for the MC Truth trained Zmm and Zee models, showing best performance in Data. To compare distributions, densities for signal and background in MC has been weighted to match the Data distribution. The ratio of the LGBM score on the different datasets are plotted in the bottom of the figure.

These plots shows very similar distributions of the LGBM scores. The Zee MC Truth data seem to have a little more background than the Data. However, both distributions match well and the MC Truth training data seem to approximate the Data we are predicting upon well. Distributions of the LGBM score for the Zll Data T&P trained models can be seen in Fig. A.38.

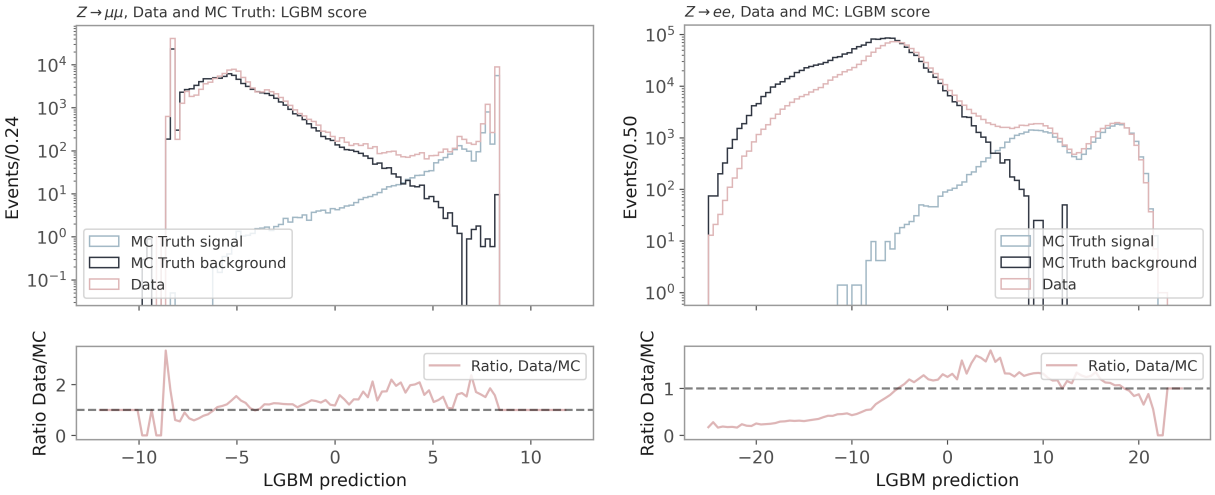


Figure 5.27: The LGBM score of the MC Truth trained models plotted in the MC test set and the Data. The MC data is weighted in signal and background separately, in order to compare the distribution to Data. Left shows the Zmm model while right shows the Zee model.

5.3.4.4 *Zllg and Hllg training*

Zllg and Hllg models all aim to identify triplets of particles, two leptons and one photon, decaying from either the Z or Higgs bosons. Thus, the training features for the Zmmg/Hmmg and Zeeg/Heeg models (see Tab. 5.19) are identical, only diverting at the point that the feature `Z_score` is taken from the Zmm model for Zmmg/Hmmg and the Zee model for Zeeg/Heeg.

The models for the Z boson and Higgs boson are trained separately—for the Z boson on a datasets with Z's and for the Higgs on datasets with H's. The dominant background for both decays is the $Z \to ll$ decay combined with a random photon. For the Zllg model, a cut for the invariant mass of the two leptons is made, $40 < m_{ll} < 83\,\text{GeV}$ as in [68], to reduce the amount of background around m_Z. This cut will

ML Zllg and Hllg input features

`dZ0`	$dZ_0 = (\gamma_{z0} - ll_{z0}) \cdot \left(\sqrt{(\gamma_{\sigma_{z0}})^2 + (ll_{\sigma_{z0}})^2}\right)^{-1}$, where ll_{z0} is the weighted average of the z_0 of the two leptons, $l_1/(l_{1,\sigma_{z0}})^2 + l_2/(l_{2,\sigma_{z0}})^2$ and $ll_{\sigma_{z0}} = \left(1/(l_{1,\sigma_{z0}})^2 + 1/(l_{2,\sigma_{z0}})^2\right)^{-1}$
`Z_score`	The Zee score for the Zeeg/Heeg model and the Zmm score for Zmmg/Hmmg
`pho_isConv`	Is the photon converted?
`pho_PID_score`	The pPid score
`pho_ISO_score`	The pIso score

Table 5.19: Input features for the Zllg and Hllg models.

not be performed for the Hllg models, as this will also cut away the signal, where $m_{ll} = 91.2\,\text{GeV}$.

Fig. 5.28 shows the SHAP values all models. The models all agree on their top contributor to the output score; `Z_score`. The pPid and pIso takes the second and third spot for all models, however the Hmmg model ranks the pIso higher than the pPid. The feature `dZ0` should give information about whether the particles emerge from the same vertex, but does not seem to have a big importance for the models. If all decay particles receive high Pid and Iso scores, they are unlikely to be from different vertices, which might make the `dZ0` feature redundant. The uncertainty for the photon $\gamma_{\sigma_{z0}}$ is also high, making the variable very uncertain.

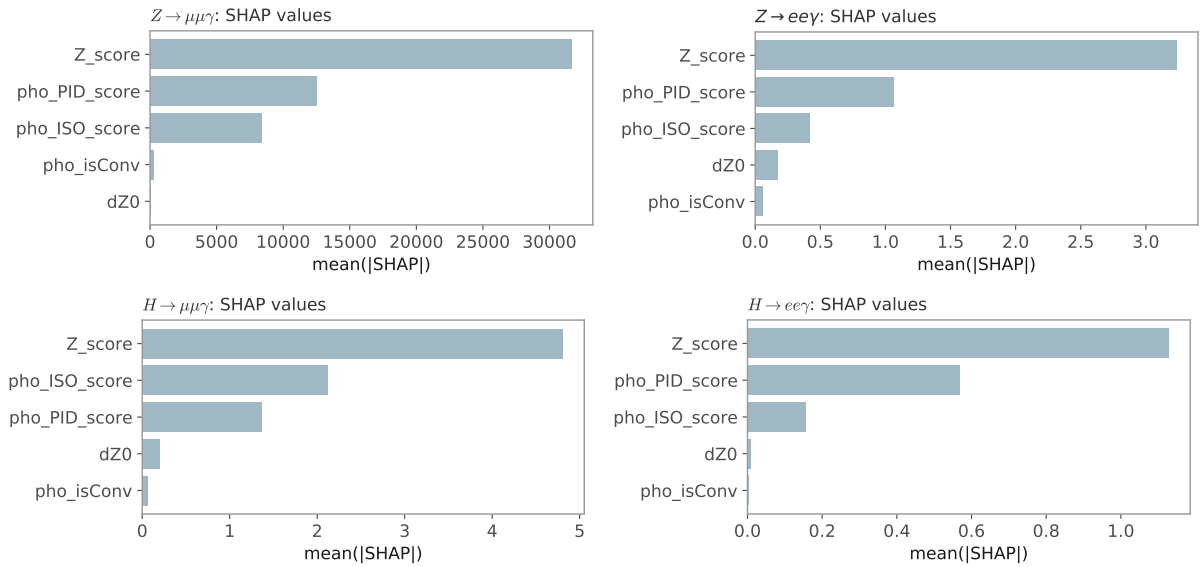


Figure 5.28: SHAP values for the Zllg (top) and Hllg (bottom) models

5.3.4.5 *Zllg testing in MC*

Fig. 5.29 shows the model performance of the Zllg models in the MC test set. The MC truth labeling shows an increase for both models, with +74.3% for the Zmmg model and +43.8% for the Zeeg model.

Same-sign evaluation for the Zeeg model is about 15% lower than the Truth evaluation while the bkg-count mehod lowers the increase

to 60% of that of the Truth, with only 12.1% increase, significantly lower than the MC Truth increase on the same dataset. The fitting shows an increase of 64.1%.

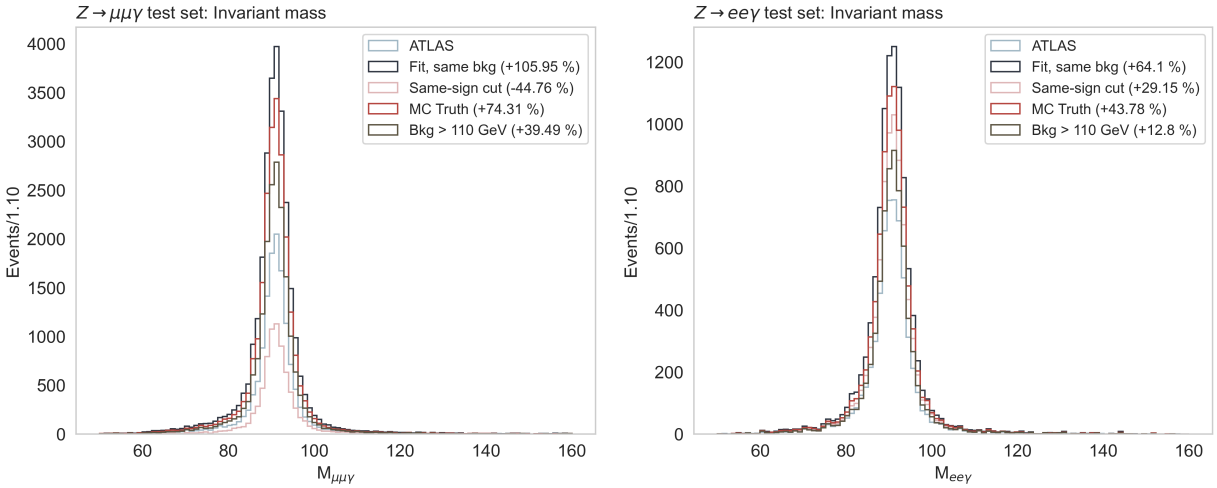


Figure 5.29: Plot of the invariant mass for the $Z \to ll\gamma$ decays in MC test. The left plot shows the $Z \to \mu\mu\gamma$ decay and the right plot shows the invariant mass for the $Z \to ee\gamma$ decay. Same plots can be seen in Fig. A.39 in logarithmic scale.

For the Zmm model, the same-sign evaluation is again low. Here the ATLAS selection only has 1 same-sign event compared to 13,276 opposite-sign signal events. The evaluation method yields a decrease of 44.8%. However, the fit and bkg-count methods has a high increase, with 105% for the fit and 39% for the bkg-count respectively.

Zllg models performance					**MC**
Eval	**N ATLAS**		**N LGBM**		**Change**
Zmmg	Sig	Bkg	Sig	Bkg	
f_{bkg}	13,498	107	27,951	105	106%
n_{SS}	13,276	1	7,334	0	−45%
bkg-count	13,276	304	18,519	309	39%
MC Truth	13,174	514	23,274	515	74%
Zeeg					
f_{bkg}	6,123	88	10,045	88	64%
n_{SS}	6,030	7	7,788	8	29%
bkg-count	6,030	122	6,802	124	13%
MC Truth	5,744	380	8,365	381	44%

Table 5.20: Performance of the Zllg models in MC. The number of signal events are measured in the range $[70;110]$ GeV while the number of background events are measured for $[50;150]$ GeV.

An overview of the different evaluation methods and their performance can be seen in Tab. 5.20. The evaluation methods have a high variance in the estimation of the signal increase. It could seem that one method overestimates while other underestimates the gained increase of the model.

As was seen in Figs. 5.20 and 5.21, the Zee and Zmm models were most efficient for $M_{ll} > 70$. This measure was based on the getting the same background efficiency as ATLAS in the di-lepton models. In Figs. 5.30 and 5.31, the invariant masses of the di-lepton pair in the

$Z \to ee\gamma$ and $Z \to \mu\mu\gamma$ decays are shown along with the MC Truth selection from the Zeeg and Zmmg models. For the Zmmg model, we seem to have an increase in signal for the full range of the mass, differently than what we saw in the efficiency plot in Fig. 5.20.

However, in the Zmmg model, both the Zmm model and the photon models are combined. The Zmmg model can put more or less weight on the inputs—it might be that background is removed mainly by cutting in the pPid score—thus we might reach the same desired background efficiency at a different cut in the di-lepton score than for in the Zmm model. If we use the cut found for the $Z \to \mu\mu$ decay in the Zmmg model, the story is different. Then the increase is gained from the range $[65; 83]$ GeV, as expected. It indicates that the Zmmg model can be less strict compared to the Zmm model when cutting in the di-lepton score, and that the main background cut away by the Zmmg model, seems to be based on stricter cuts in the photon models. For the Zeeg model, it seems that it is the same cut in the Z score for both Zee and Zeeg models where increase in signal is mainly for m_{ll} in $[60; 83]$ GeV.

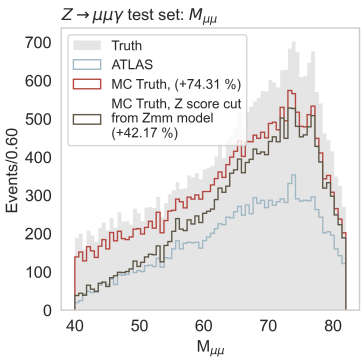


Figure 5.30: Plot of the $M_{\mu\mu}$ for the Zmmg test set.

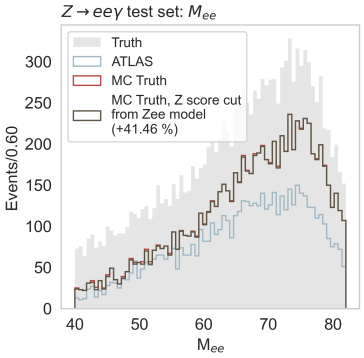


Figure 5.31: Plot of the M_{ee} for the Zeeg test set. It is evident that the increase is for $M_{ee} > 70$ GeV. The shaded area shows all events labeled as signal, indicating that the most signal is in the region $M_{ee} > 70$ GeV.

5.3.4.6 *Zllg testing in Data*

The invariant mass plot for the two models on Data is shown in Fig. 5.32 while Tab. 5.21 summarizes the number of signal and background events and the performance.

For the Zmmg decay, an improvement of $+30.9\%$ is found for the fitting method. Again, a decrease is seen for same-sign evaluation, here of -39.4%. The bkg-count method, increases the selection with 17.6%. Thus a significant drop compared with selections in the corresponding MC data.

Zllg models performance					**Data**
Eval	**N ATLAS**		**N LGBM**		**Change**
Zmmg	Sig	Bkg	Sig	Bkg	
f_{bkg}	30,572	3350	40,628	3518	31%
n_{SS}	32,148	21	19,483	22	-39%
bkg-count	32,148	2752	37,800	2761	18%
Zeeg					
f_{bkg}	7,736	696	9,438	697	21%
n_{SS}	7,964	58	11,809	58	48%
bkg-count	7,964	621	8,479	631	6%

Table 5.21: Performance of the Zllg models in Data. The number of signal events are measured in the range $[70; 110]$ GeV while the number of background events are measured for $[50; 150]$ GeV.

The Zeeg model shows increase between 6% and 48% depending on the method of evaluation. Again, the bkg-count method shows the least improvement, while the same-sign the highest improvement, with around 10% higher increase compared with the same method in MC.

A plot of the LGBM scores can be seen in Fig. A.40, showing similar distributions for both Zmmg and Zeeg models in MC and Data.

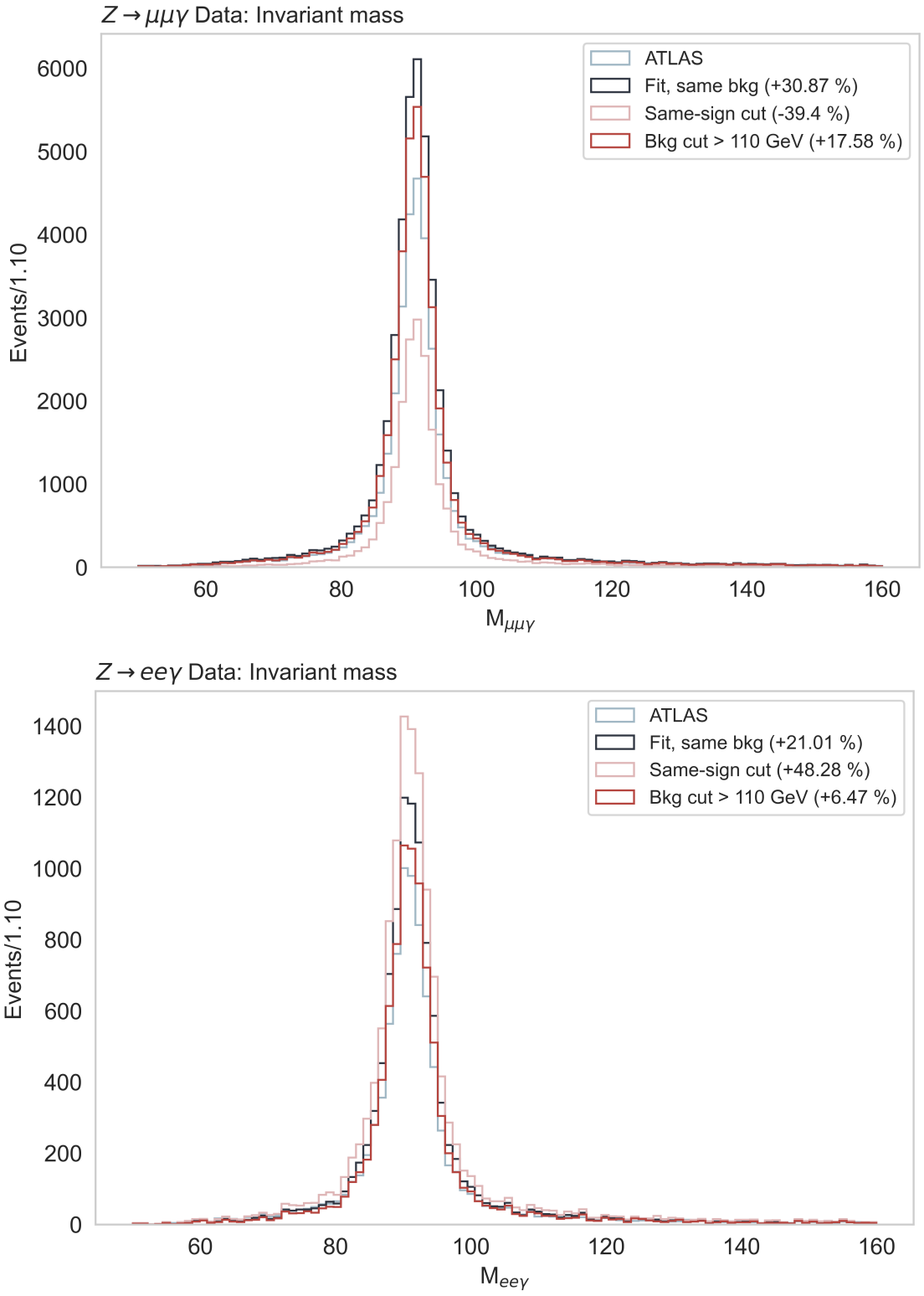


Figure 5.32: Plot of the invariant mass for the $Z \to ll\gamma$ decays in Data. The top plot shows the $Z \to \mu\mu\gamma$ decay and the bottom plot shows the invariant mass for the $Z \to ee\gamma$ decay and corresponding increases in Data. They are shown in linear scale to better see the large increase in signal. See Fig. A.41 for same plots on logarithmic scale.

5.3.4.7 *Hllg testing in MC*

Fig. 5.33 shows the invariant mass of the Higgs MC test set along with the different evaluation methods. Generally, it is noticeable that the evaluation methods here all yield results in similar ranges (i.e. a smaller variation between evaluation methods is seen) for both models opposed to what was seen for the Zllg models.

The MC Truth prediction yields an increase of 52.6% for the Hmmg model and 28.8% for the Heeg model. For the $H \to Z(\to ee)\gamma$ decay, it is again, similar to the Zeeg model, the result same-sign method that is closest to the MC truth increase, here with 29%. For the $H \to Z(\to \mu\mu)\gamma$ decay, the same-sign cut again yields a decrease, while the two other methods are lower than the MC truth, bkg-count with the closest improvement of 48.8%.

All models for muon decays has shown a decrease in signal when using the same-sign evaluation method. It could indicate that the ATLAS selection for muons is optimized with respect to the number of same-sign events in some way. The charge of the electron is more uncertain than the charge of the muon. This is due to charge misidentification mainly caused by small-angle bremsstrahlung in the calorimeter leading to incorrect determination of the track curvature [68]. This could explain why the selection for muons might be optimized against the charge and the selection for the electrons is not.

See Tab. 5.22 for an overview of the improvements.

Hllg models performance					**MC**
Eval	**N ATLAS**		**N LGBM**		**Change**
Hmmg	Sig	Bkg	Sig	Bkg	
f_{bkg}	8,959	587	12,582	588	44%
n_{SS}	8,652	3	8,236	3	−5%
bkg-count	8,652	177	12,873	178	49%
MC Truth	8,429	611	13,033	612	53%
Heeg					
f_{bkg}	7,676	562	9,897	552	29%
n_{SS}	7,678	15	9,958	14	30%
bkg-count	7,678	144	10,167	145	32%
MC Truth	7,363	706	9,623	707	29%

Table 5.22: Performance of the Hllg models. The number of signal events are measured in the range [115; 135] GeV while the number of background events are measured for [100; 170] GeV.

5.3.4.8 *Hllg performance in Data*

We cannot test our Higgs models in Data, but from the performance of the Zllg models, we will try to extrapolate the Higgs Data behavior.

Considering the Zmmg model, the MC Truth yields an increase of 74.3%. The mean increase of the bkg-count and f_{bkg} methods is 73%. In Data, the mean performance of the same methods is 24.5%,

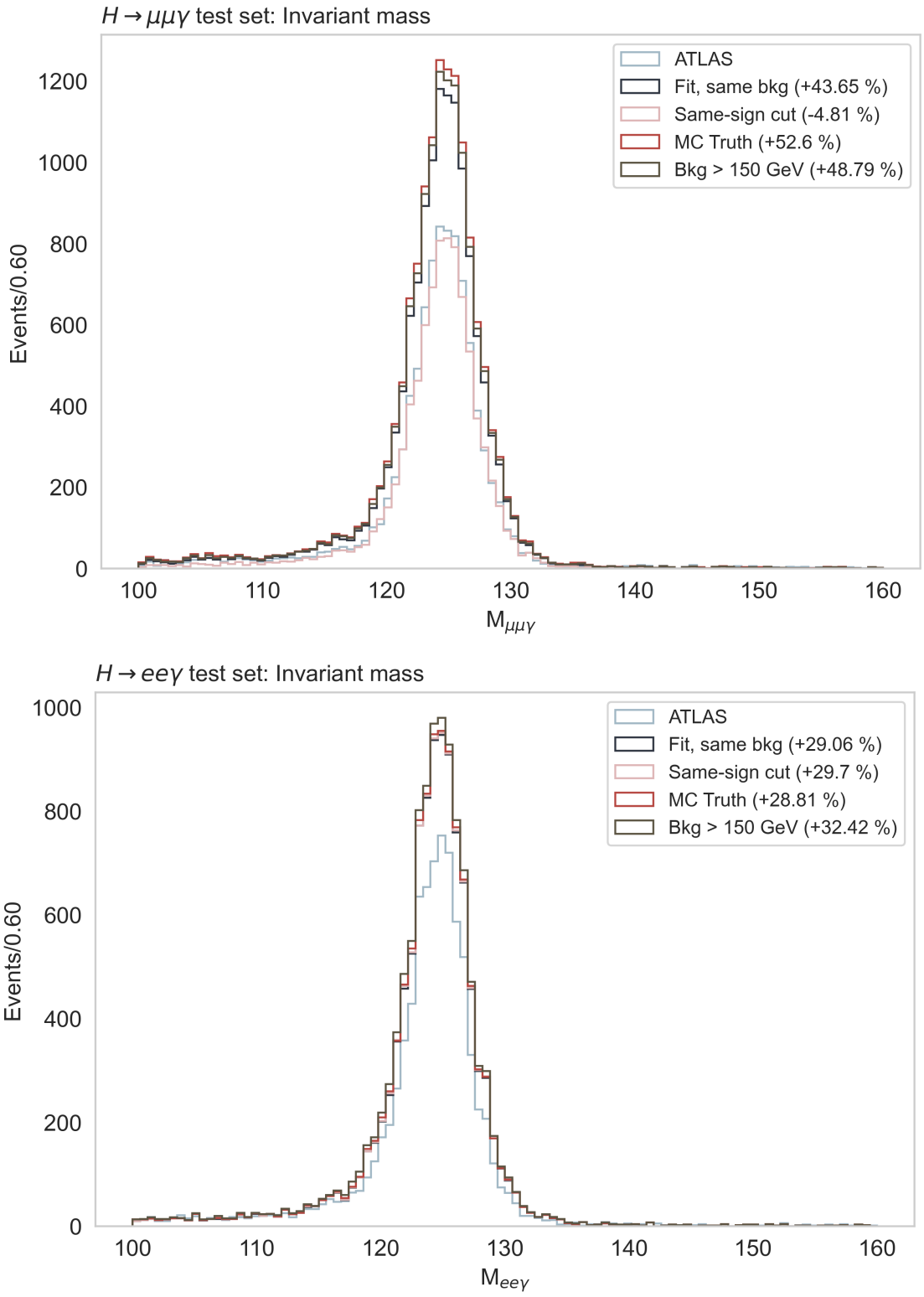


Figure 5.33: Plot of the invariant mass for the $H \to Z(\to ll)\gamma$ decays in MC test. The top plot shows the $H \to Z(\to \mu\mu)\gamma$ decay and the bottom plot shows the invariant mass for the $H \to Z(\to ee)\gamma$ decay. They are shown in linear scale to better see the large increase in signal. See Fig. A.42 for same plots on logarithmic scale.

corresponding to a decrease of 66% compared to the MC performance. In $H \to Z(\to \mu\mu)\gamma$, we see a mean performance of 46.5%. Assuming similar decrease going from MC to Data, it can be expected to see an increase of 16% in $H \to Z(\to \mu\mu)\gamma$ Data. See Tab. 5.23 for an overview of the numbers.

Extrapolated H $\to$ Z($\to\mu\mu$)γ model performance				**Data**
	Z $\to\mu\mu\gamma$		**H $\to$ Z($\to\mu\mu$)γ**	
Eval	**MC**	**Data**	**MC**	**Data**
f_{bkg}	106%	31%	44%	
bkg-count	40%	18%	49%	
Mean	73%	24.5% (-66%)	46.5%	16% (est.)

Table 5.23: Extrapolated performance of the Hmmg model in Data. The table shows the performance of the evaluation methods (excluding the same-sign method) in MC and Data for the Zmmg model and the mean performance of the models where a decrease of 66% is seen going from MC to Data. Assuming similar behavior for the Hmmg model, the expected increase in Data is $+16\%$.

Extrapolated H $\to$ Z($\to$ ee)γ model performance				**Data**
	Z $\to$ eeγ		**H $\to$ Z($\to$ ee)γ**	
Eval	**MC**	**Data**	**MC**	**Data**
f_{bkg}	64%	21%	29%	
n_{SS}	29%	48%	30%	
Mean	47%	34.5% (-27%)	29.5%	22% (est.)

Table 5.24: Extrapolated performance of the Heeg model in Data. The table shows the performance of the evaluation methods in MC and Data for the Zeeg model and the mean performance of the models. The bkg-count evaluation method is excluded, as a significant worsening in performance is seen compared with the MC Truth evaluation in $Z \to ee\gamma$. We see a decrease in the Zeeg model going from MC to Data of -27%. Assuming similar behavior for the Heeg model, the expected increase in Data is $+22\%$.

The same train of thought can be used for the $H \to Z(\to ee)\gamma$ decay. Here, the bkg-count evaluation method is excluded, as it performs significantly worse than the MC Truth method in Zeeg. The two other, same-sign and f_{bkg} methods, has a mean increase in performance of 47%, similar to the 43.8% increase seen for the MC Truth method. In Data, their mean increase drops 27%, to 34.5%. For the $H \to Z(\to ee)\gamma$ decay, we see a mean increase of 29.5% in MC and a 27% decrease will yield a final increase in Data of 22%. See Tab. 5.24.

It is peculiar that the mean increase for both Higgs models in MC corresponds to the equivalent MC Truth increase. It seems that one method overestimates the performance while the second underestimates the performance. In the muon case, the fit yields a higher increase for the Z model than the bkg-count method. In Higgs, they get comparable results, however the bkg-count method actually yields the highest increase.

For the electrons, it is the fit that has the high increase for Z in MC while the n_{SS} has the high increase in Data. In MC, ATLAS has 7 same-sign events compared to 6,000 opposite-sign (signal) events. In Data, the ATLAS selection gets 58 same-sign events and 8,000 opposite-sign events, that is 8 times the background at 1.3 times the signal. It is not evident why there is such a big difference, but it is likely to be the reason for the high increase of the same-sign method in Data compared to the same method in MC.

5.3.5 *Conclusions on boson models*

The training and performance on boson models has been described in the last section. Models for the $Z \to ll$, $Z \to ll\gamma$, and $H \to Z(\to ll)\gamma$ decays have been trained based on the lepton and photon models from Sect. 5.2. The di-lepton models have been trained for three different types of datasets, where the best performing models were

the ones trained in MC using truth variables as labeling.

The model performance has been measured against the ATLAS selection of [1]. Models for the Z boson has been evaluated in both Monte Carlo and Data, while models for the Higgs boson has been evaluated in MC only. In MC, improvements (based on the labeling of data) ranging between $10 - 74\%$ has been seen. Generally a small drop in performance happens when applying the models to Data, where improvements are in the range $6 - 48\%$. The difference in estimated increase between the methods of evaluation is quite high. It seems that some methods overestimates while other underestimates the gained increase of the model.

From the performance in $Z \rightarrow ll\gamma$ in MC and Data, the Data performance in $H \rightarrow Z(\rightarrow ll)\gamma$ is extrapolated. Improvements for $H \rightarrow Z(\rightarrow ee)\gamma$ around 20% is expected and for $H \rightarrow Z(\rightarrow \mu\mu)\gamma$ expected improvements are around 15%.

The di-lepton models shows the least improvement in Data. However, it is shown that more signal can be gained, especially for the electron model. This is both by including the crack and forward regions, but also in fiducial parts of the detector. By loosening the cut on the model, and allowing more background in the selection, an even stronger improvement is seen. This can definitely be of relevance for several physics searches in the ATLAS detector, as the Z boson is part of final states in many ATLAS analyses, both measurements and searches.

Generally, the ML models presented in the section has shown large improvements in Z and Higgs selections with very few exceptions.

6 Conclusions and outlook

This chapter will draw the final conclusion and suggest further work that can be done.

6.1 Conclusions

In this thesis, the $H \to Z(\to ll)\gamma$ decay has been investigated. Machine Learning models has been created to optimize the search for the decay, and the work has been evaluated against the cut-based selection of [1].

Particle identification and isolation models for single leptons and photons, e, μ and γ, has been trained. The models for muons and electrons were trained using three different datasets—two based on Monte Carlo data and one based on Data—while photon models were trained on one MC dataset only.

The Pid models has been evaluated against ATLAS working points Loose, Medium, and Tight. The Data trained models showed bad results for particle identification, with decrease in signal up to 115%. The MC trained muon models showed increase in signal of $4 - 4.5\%$ compared to the Medium WP and 11% compared to the Tight WP while electron models increased signal selection of $14 - 17\%$ and $15 - 28\%$ for Medium and Tight WP respectively. The photon Pid model showed increase of 9% for the Loose WP while a decrease in signal of -0.7% was seen for the Tight WP. A separate Pid model for the electrons in the forward calorimeter has also been trained. Evaluated against the ATLAS WPs, a decrease in signal was seen. However, forward electrons are discarded in the ATLAS selection, so simply including electrons from this part of the detector was likely to yield an increase in performance.

The isolation models has been evaluated against ATLAS isolation WPs, all showing increase in signal with $6 - 8\%$ for the muon and $10 - 12\%$ for the electron models respectively. Photon models does not have an ATLAS comparison, but was evaluated against the muon WP for visualization purposes. The AUC of the photon model was

found to be 0.893.

The Pid and Iso models for muons and electrons were used as input variables for a di-lepton model trained on the $Z \to ee$ and $Z \to \mu\mu$ datasets. The di-lepton models were again trained on three different datasets, just as the single muon and electron models. The models has been evaluated in MC and Data. The MC performance showed increase in signal compared with the selection from [1] of $20 - 30\%$ for the electrons and $10 - 19\%$ for the muons. The increase for electrons was gained from the fiducial part of the detector, but also from the crack and forward calorimeter, which are excluded in the ATLAS selection. For the muons, the increase was gained from the fiducial part of the detector. In Data, the performance of especially the MC T&P trained models dropped, and a plot of the LGBM score in MC and Data revealed that the distributions did not match. For the other models, the increase was $9 - 18\%$ for the electrons and $5 - 9\%$ for the muons, excluding same-sign evaluation for muons.

Lastly, the di-lepton models served as input, together with the photon Pid and Iso models, to the $Z \to ll\gamma$ and $H \to Z(\to ll)\gamma$ models. The $Z \to \mu\mu\gamma$ and $Z \to ee\gamma$ models were trained on MC Truth data only and evaluated in both MC and Data. The $H \to Z(\to \mu\mu)\gamma$ and $H \to Z(\to ee)\gamma$ models were trained and evaluated in MC.

The $Z \to \mu\mu\gamma$ model showed increase in signal in MC of $40 - 106\%$ dropping to $17 - 30\%$ when applied to Data, excluding same-sign evaluations in both cases. The $Z \to ee\gamma$ model showed MC increase of $12 - 44\%$ and Data increase of $6 - 48\%$. The Higgs model for muon decays, $H \to Z(\to \mu\mu)\gamma$, showed increase of $44 - 53\%$ in MC while for the $H \to Z(\to ee)\gamma$ model, an increase of $29 - 32\%$ was seen. Considering the application of the $Z \to ll\gamma$ MC trained models in Data, where the largest drop in performance was seen for the muon model, expected increase of the muon and electron models is at 15% and 20% respectively.

The ML approach for the $H \to Z\gamma$ decay presented in the thesis has shown results consistently improving the ATLAS selection of [1]. Machine learning-based methods have shown beneficial in particle identification tasks, and the results strongly suggests that adoption of ML-based methods would improve signal selections compared to current cut-based methods in ATLAS.

6.2 *Outlook and further work*

The analysis in [1] showed results consistent with both signal and background. The Data collection in Run3 will increase the luminosity, which hopefully increases the possibility of observing the decay. However, a larger amount of signal is still relevant to aim for.

Improving the selection presented in this thesis could be achieved

with better performing photon models. The pPid model trained in Sect. 5.2 showed a decrease in signal of -0.7% compared to the Tight WP, which is used in the ATLAS selection from [1]. The photon models in this analysis has been trained for unconverted and converted photons together, but splitting them up and training distinct models for the two types could be considered. Furthermore, the training data for the photon models was data from the $Z \to ll\gamma$ decays. Here, the photon energy is lower than for the Higgs decays, and thus it is relevant to include high-energy photons in the training data as well.

Furthermore, more electrons from the forward calorimeter should be included in the $Z \to ee$ dataset. Unfortunately, these were unavailable, but could in principle be a source to further improvements for the Zee model.

The MC data used in the thesis has been aimed to correspond to real Data in terms of type of background and also, in some extent, signal-background ratios. However this could certainly be optimized which could give less performance difference between models in MC and Data.

The Z boson models serves as a great candidates for scale-factor calculation. By fitting, one can estimate the efficiency of the Z selection in MC and Data. In principle, three selections—a loose, medium, and tight—could be made with corresponding scale factors. This will give the possibility to correct simulation to match Data. The scale factors are often calculated in different regimes, e.g. in high and low p_T, which is also possible here.

Ongoing studies in the group include better energy resolution, narrowing the signal peak of a detected particle. If successful, more background (and thus also more signal) can be included while still having a decent signal-to-background ratio. Models for this analysis is trained in Data and MC together, which could also be applied to the analysis presented in this thesis. This ensures a better transaction from the trained model to the data it is applied to.

Current searches in ATLAS also include the $H \to \gamma^*\gamma$ decay, where the γ^* decays into two low-mass leptons ($m_{ll} < 30\,\text{GeV}$). The di-lepton model described in this analysis could in principle be applied to the decay together with the photon Pid and Iso models. However, the di-lepton models showed decreasing efficiency for low-mass, but with training samples from the Drell-Yan process or a $\Upsilon \to l^+l^-$ decay sample, a better signal selection for low-mass could be achieved. Furthermore, the muon lepton models could also be utilized on the decay $H \to \mu\mu$.

A Appendix

A.1 Datasets

Process	Container	Model
$Z \to \mu\mu$	mc16_13TeV.361107.PowhegPythia8EvtGen_AZNLOCTEQ6L1_Zmumu.deriv.DAOD_MUON1.e3601_e5984_s3126_r10724_r10726_p3629	mPid, mIso, Zmm
$Z \to \mu\mu$ (incl. γ in events)	mc16_13TeV.361107.PowhegPythia8EvtGen_AZNLOCTEQ6L1_Zmumu.merge.DAOD_EGAM4.e3601_e5984_s3126_r10724_r10726	Zmmg
$Z \to \mu\mu\gamma$	mc16_13TeV.301536.Sherpa_CT10_mumugammaPt10_35.deriv.DAOD_EGAM4.e3952_s3126_r10201_r10210_p3956	mPid, mIso, Zmm, Zmmg
$W^+ \to \mu\nu$	mc16_13TeV.361101.PowhegPythia8EvtGen_AZNLOCTEQ6L1_Wplusmunu.deriv.DAOD_MUON1.e3601_e5984_s3126_s3136_r10724_r10726_p4145	mPid, mIso, Zmm
$W^- \to \mu\nu$	mc16_13TeV.361104.PowhegPythia8EvtGen_AZNLOCTEQ6L1_Wminusmunu.deriv.DAOD_MUON1.e3601_e5984_s3126_s3136_r10724_r10726_p4145	mPid, mIso, Zmm
$Z \to ee$	mc16_13TeV.361106.PowhegPythia8EvtGen_AZNLOCTEQ6L1_Zee.deriv.DAOD_EGAM1.e3601_e5984_s3126_r10724_r10726_p4323	ePid, eIso, Zee
$Z \to ee$ (fwd)	mc16_13TeV.361106.PowhegPythia8EvtGen_AZNLOCTEQ6L1_Zee.deriv.DAOD_EGAM8.e3601_s3126_r10201_r10210_p3648	ePid, eIso, Zee
$Z \to ee$ (incl. γ in events)	mc16_13TeV.361106.PowhegPythia8EvtGen_AZNLOCTEQ6L1_Zee.deriv.DAOD_EGAM3.e3601_e5984_s3126_r10724_r10726_p4250	ePid, eIso, Zeeg
$Z \to ee\gamma$	mc16_13TeV.301535.Sherpa_CT10_eegammaPt10_35.deriv.DAOD_EGAM3.e3952_s3126_r10201_r10210_p3956	ePid, eIso, Zee, Zeeg
$W^+ \to e\nu$	mc16_13TeV.361100.PowhegPythia8EvtGen_AZNLOCTEQ6L1_Wplusenu.deriv.DAOD_EGAM1.e3601_e5984_s3126_r10201_r10210_p3916	ePid, eIso, Zee
$W^- \to e\nu$	mc16_13TeV.361103.PowhegPythia8EvtGen_AZNLOCTEQ6L1_Wminusenu.deriv.DAOD_EGAM1.e3601_e5984_s3126_r10201_r10210_p3916	ePid, eIso, Zee
$H \to ll\gamma$	mc16_13TeV.345833.PowhegPythia8EvtGen_NNPDF30_AZNLOCTEQ6L1_VBFH125_Zllgam.deriv.DAOD_HIGG1D2.e7111_s3126_r10724_p4062	Hmmg, Heeg
$H \to ll\gamma$	mc16_13TeV.344303.PowhegPythia8EvtGen_CT10_AZNLOCTEQ6L1_MPIOFF_ggH125_Zllgam.deriv.DAOD_HIGG1D2.e5145_s3126_r10724_p4062	Hmmg, Heeg

Table A.1: Used datasets

A.2 Lepton and photon models

A.2.1 Particle Identification: Muons

Here, the plots for the muon Pid models trained in MC T&P and Data T&P are shown. The procedure of the training and prediction is described in Sect. 5.2.3.1.

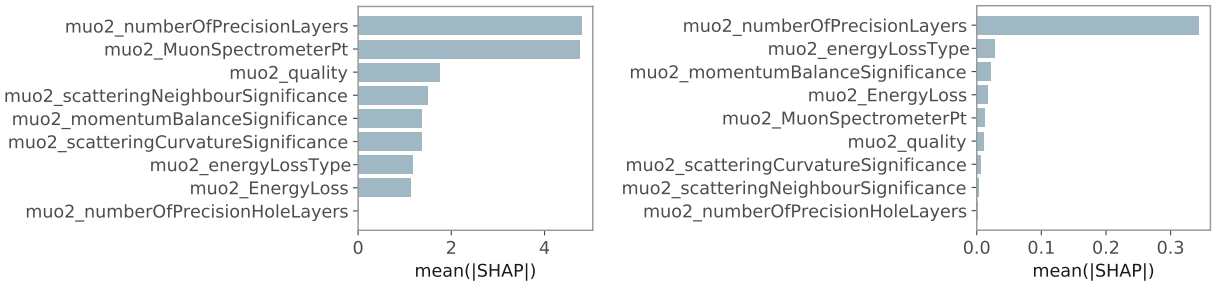


Figure A.1: SHAP values for the MC and Data T&P mPid model. The variable with the highest influence on both models is `muo_numberOfPrecisionLayers`.

The LGBM score plots in Fig. A.2 shows complete seperation of signal and background for the MC trained model. The ROC curve for the same model, shown in Fig. A.3 also shows that the model has an AUC of 1, i.e. perfect performance. The Data trained T&P model shows performance worse than the ATLAS selection. Both features could be due to the datasets being sparse—for the MC model the total seperation could indicate that only specific regions are covered, making the decision easy for the model. For the Data trained model, we have low statistics in the training data, meaning that the model cannot generalize well.

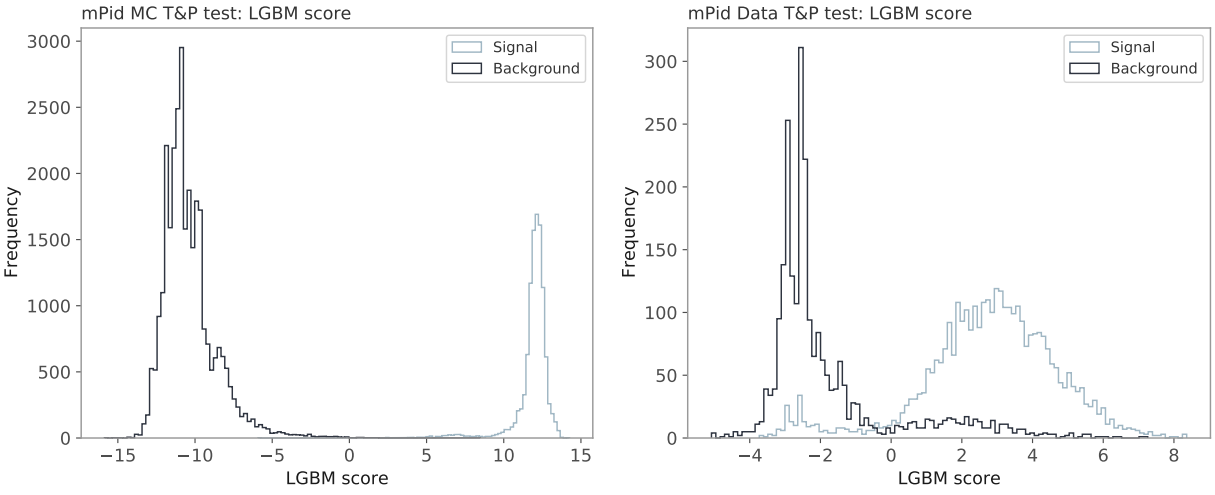


Figure A.2: LGBM score for the T&P mPid models.

A.2.2 *Isolation: Muons*

Here, the plots for the muon Iso models trained in MC T&P and Data T&P are shown. The procedure of the training and prediction is described in Sect. 5.2.3.2.

The LGBM scores (Fig. A.5) show some overlap between signal and background, which is also evident in the ROC curve in Fig. A.6, where the AUC is 0.958 and 0.703 for the MC and Data model respectively.

A.2.3 *Particle Identification: Electrons*

For the electron, the ePid and eFwd for the MC Truth will first be shown, and then the T&P trained ePid models will follow.

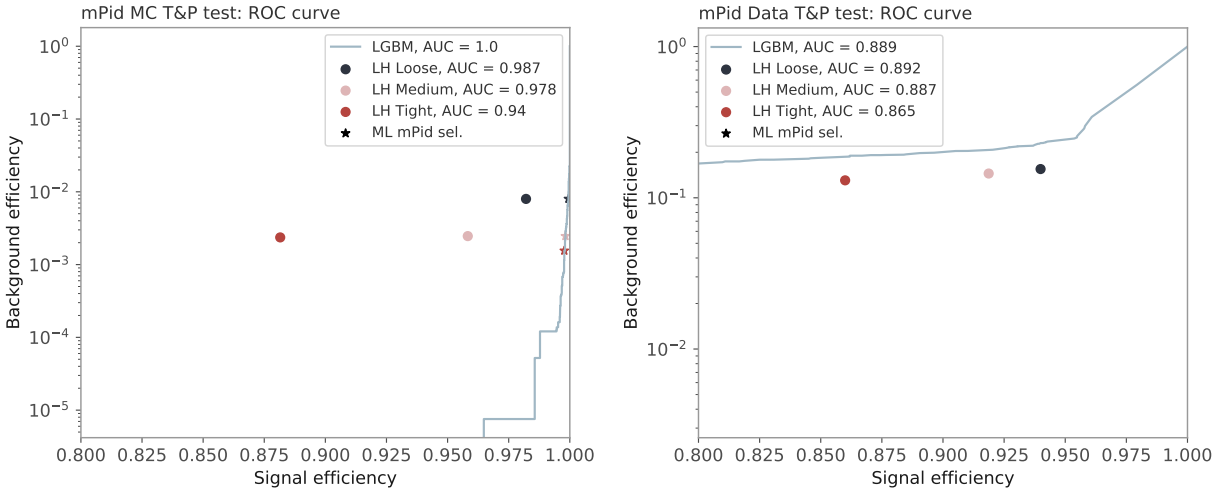


Figure A.3: ROC curve for the T&P mPid models

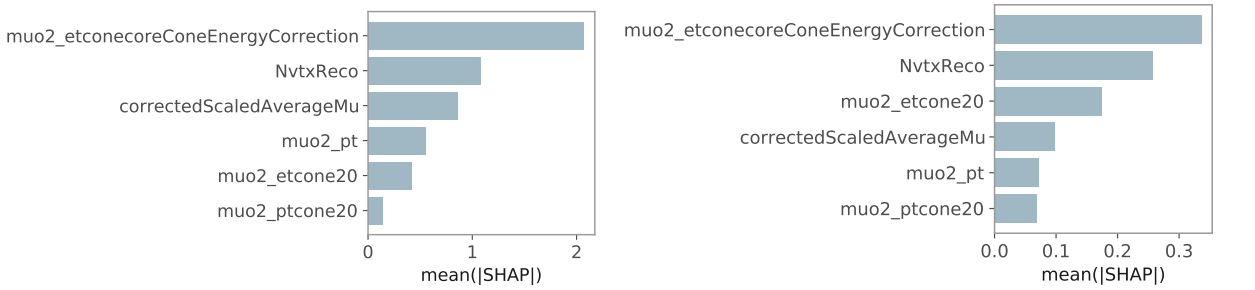


Figure A.4: SHAP values for the MC T&P mIso model. The variable with the highest influence on the model is `muo_etconecoreConeEnergyCorrection`.

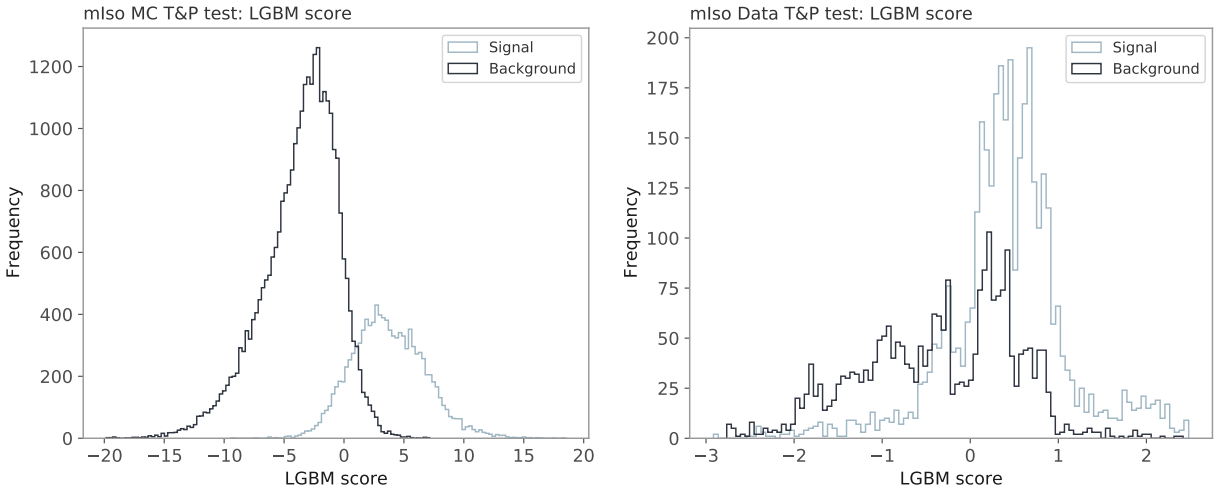


Figure A.5: LGBM score for the T&P mIso models. The signal and background has some overlap in their distributions.

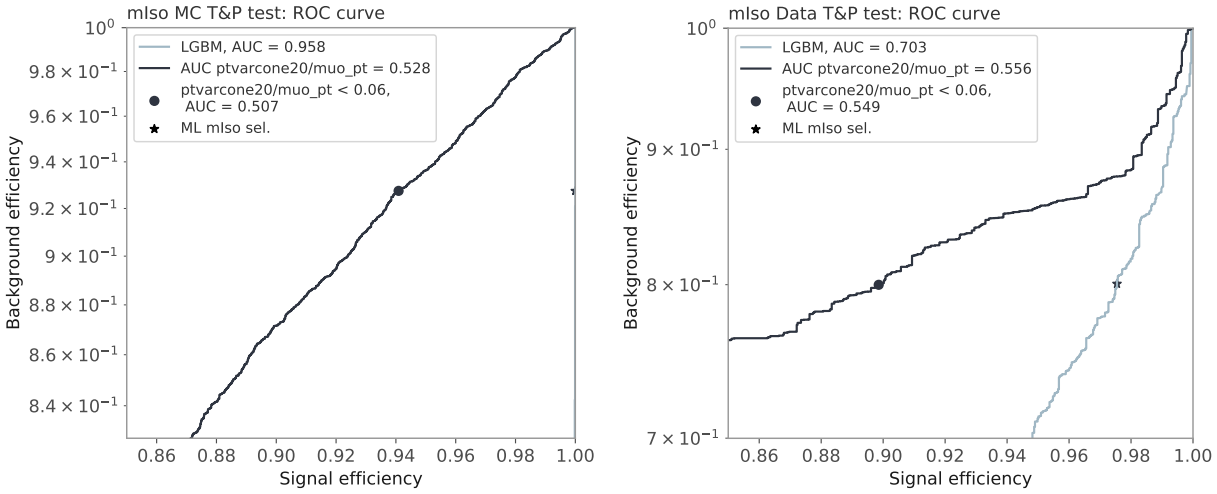


Figure A.6: ROC curve for the T&P mIso models. The MC trained model has an AUC of 0.958 while the Data trained model has AUC = 0.703.

The parameters for the ePid and eFwd models can be seen in Tab. A.2. The ePid input features are chosen to be the same as the likelihood identification used by ATLAS and will thus also be compared to the ATLAS WPs. The eFwd input features are chosen as the top 8 in [69], where the forward region has been well-researched. Unfortunately, it was not possible to use the model trained in [69], as one variable was missing in the available data.

SHAP values for the ePid and eFwd MC Truth trained models can be seen in Fig. A.7.

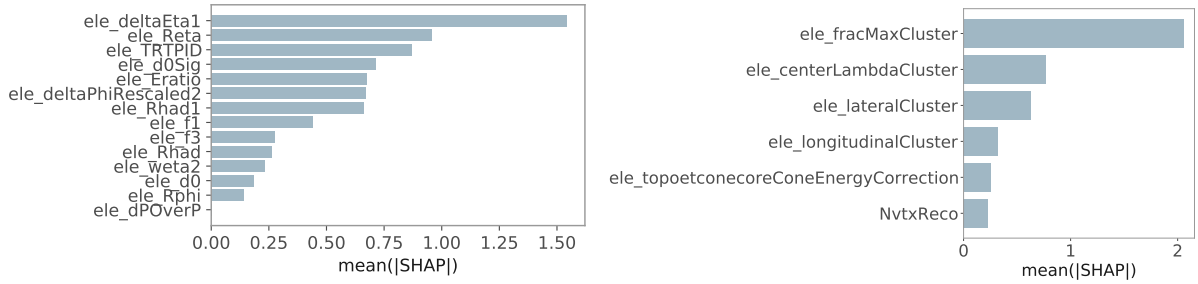


Figure A.7: SHAP values for the ePid (left) and eFwd (right) models. The variable with the highest influence on the model is `ele_deltaEta1` for the ePid model and `ele_fracMaxCluster` for the eFwd model.

Fig. A.8 shows the logit transformed LGBM scores for the ePid and eFwd models. The ePid score is well-seperated and also yields an AUC of $AUC = 0.992$ (see Fig. A.9). Here, we also see clear improvement compared to ATLAS WPs Loose, Medium and Tight. For the eFwd model, the signal and background are not as clearly seperated which is also evident on the bottom ROC curve in Fig. A.9, where the ATLAS WPs actually yields a higher signal efficiency than the LGBM model. However, ATLAS does not use forward electrons in the analysis, so simply including that part of the detector will most likely yield improvements in the output.

For the T&P trained models, we get an AUC of 0.998 for the MC and 0.893 for the Data trained ePid model, see Fig. A.12. The LGBM scores (Fig. A.11) also shows a clear seperation between signal and

ML ePid input features	
`ele_Rhad`	Ratio of E_T in the hadronic calorimeter to E_T of the EM cluster
`ele_Rhad1`	Ratio of E_T in the first layer of the hadronic calorimeter to E_T of the EM cluster
`ele_f3`	Ratio of the energy in the back layer to the total energy in the EM accordion calorimeter
`ele_weta2`	Lateral shower width
`ele_Rphi`	Ratio of the energy in 3x3 cells over the energy in 3x7 cells centered at the electron cluster position
`ele_Reta`	Ratio of the energy in 3x7 cells over the energy in 7x7 cells centered at the electron cluster position
`ele_Eratio`	Ratio of the energy difference between the largest and second largest energy deposits in the cluster over the sum of these energies
`ele_f1`	Ratio of the energy in the strip layer to the total energy in the EM accordion calorimeter
`ele_d0`	Transverse impact parameter with respect to the beam-line
`ele_d0Sig`	Significance of transverse impact parameter defined as the ratio of d_0 and its uncertainty
`ele_dPOverP`	Momentum lost by the track between the perigee and the last measurement point divided by the original momentum
`ele_TRTPID`	Likelihood probability based on transition radiation in the TRT
`ele_deltaEta1`	$\Delta\eta$ between the cluster position in the strip layer and the extrapolated track
`ele_deltaPhiRescaled2`	Defined as $\Delta\phi^2$, but the track momentum is rescaled to the cluster energy before extrapolating the track from the perigee to the middle layer of the calorimeter
ML eFwd input features	
`ele_centerLambdaCluster`	Distance of the shower barycentre from the calorimeter front face measured along the shower axis.
`ele_lateralCluster`	Lateral width of the shower cluster.
`ele_fracMaxCluster`	Fraction of cluster energy in the most energetic cell.
`ele_topoetconecoreConeEnergyCorrection`	The energy of the candidate electron for subtracting and thereby correcting the isolation topoetcone energy.
`ele_longitudinalCluster`	Lateral width, thereby depth, of the shower cluster.
`NvtxReco`	The number of reconstructed vertices in the event.

Table A.2: Input features for the ePid and eFwd models. Description of ePid from [67]. The eFwd feature description is from [69].

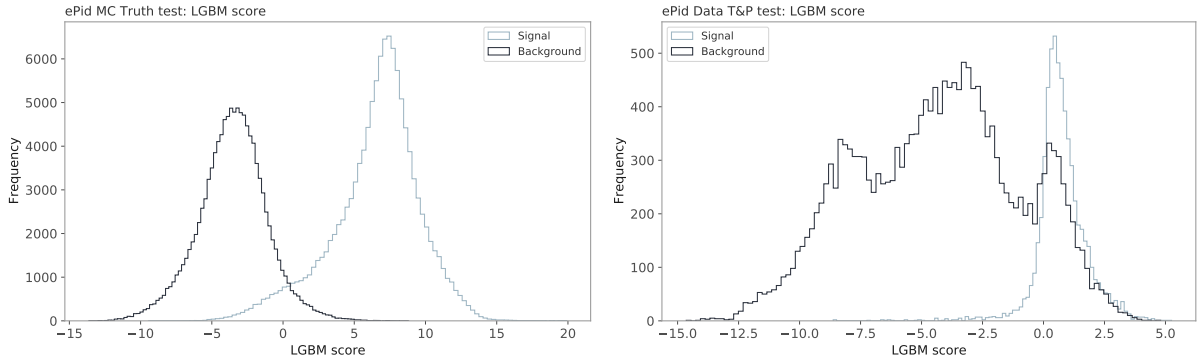


Figure A.8: Logit transformed LGBM score for the final ePid and eFwd MC Truth models.

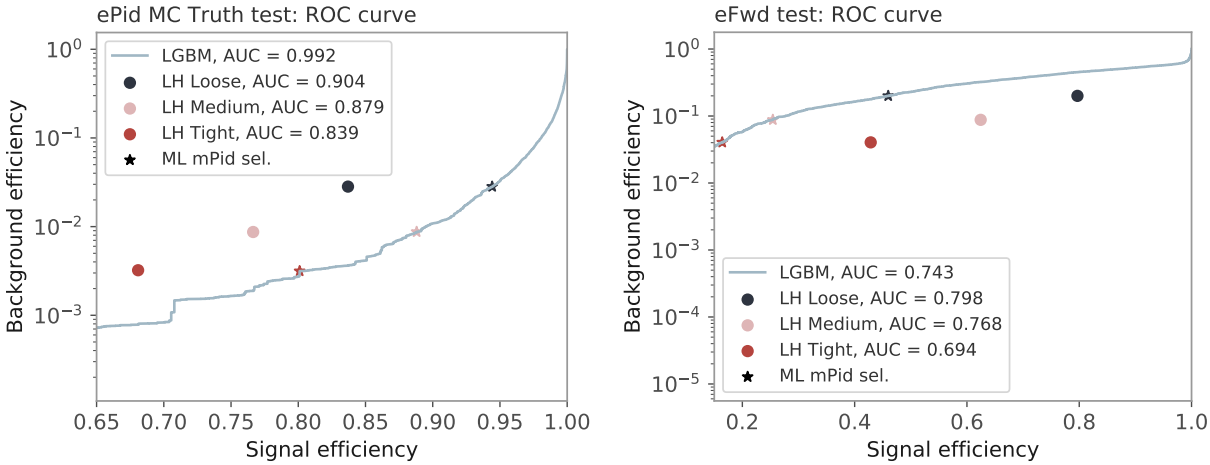


Figure A.9: The ROC curve for the ePid and eFwd MC Truth test sets. The ATLAS Loose, Medium and Tight WP are shown on the figure as well as the ML selection that matches their background efficiency.

background. However, the performance for the Data trained model is worse than the ATLAS selection, as was also seen in the muon model.

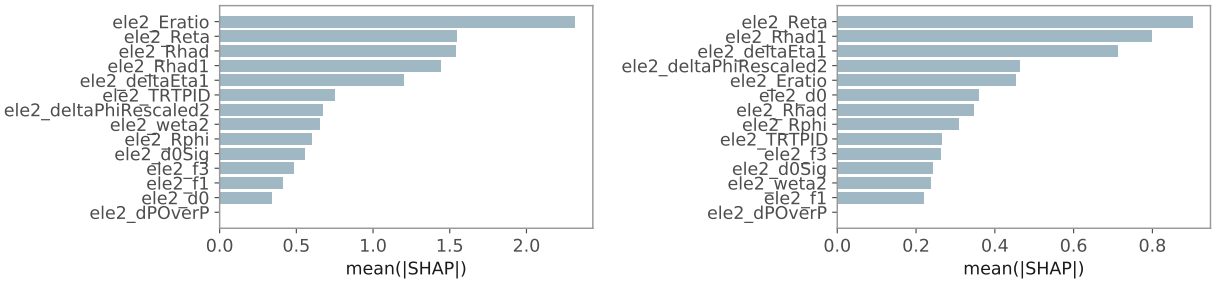


Figure A.10: SHAP values for the T&P ePid model. The variable with the highest influence on the model is `muo_numberOfPrecisionLayers`.

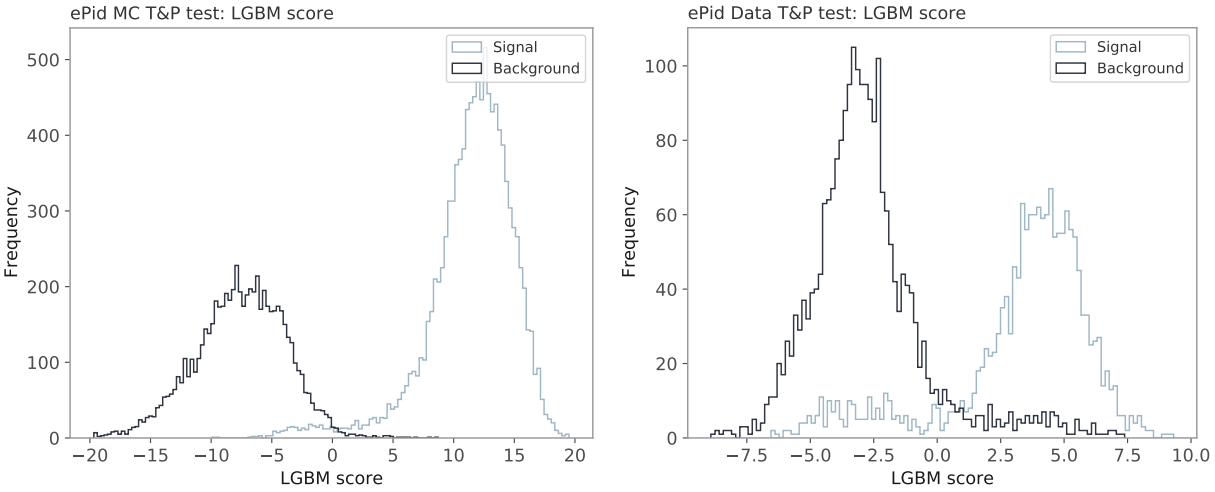


Figure A.11: LGBM score for the T&P ePid models.

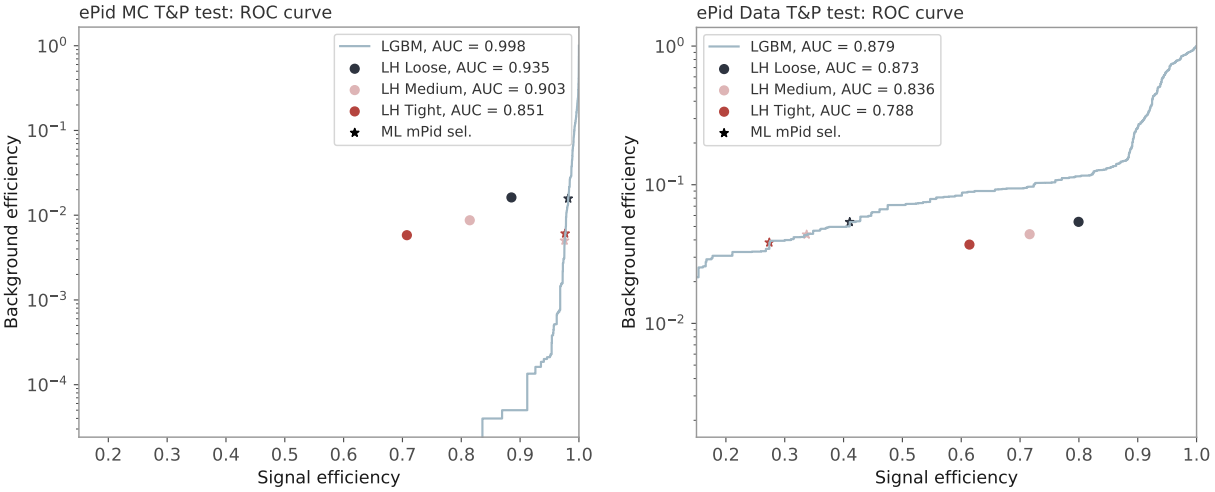


Figure A.12: ROC curve for the T&P ePid models

A.2.4 *Isolation: Electrons*

For the eIso models, the training features can be seen in Tab. A.3. The performance will be compared to the isolation WP from [67], $p_T^{varcone20} < 0.15\,E_T$.

Table A.3: Input features for the eIso models [70].

ML eIso input features	
`correctedScaledAverageMu`	Pileup
`NvtxReco`	Number of reconstructed vertecies in the event
`ele_et`	the energy E_T of the electron
`ele_ptvarcone20`	A variable of the sum of pT tracks in the 0.20 cone around the object
`ele_topoetcone20`	The sum of E_T tracks in the topo-clusters the 0.20 cone around the object
`ele_topoetcone40`	The sum of E_T tracks in the topo-clusters the 0.40 cone around the object

The SHAP values for the MC Truth trained eIso model can be seen in Fig. A.13. We see that `ele_et` has the largest importance in the model training corresponding to the top variable in the mIso model, `muo_pt`.

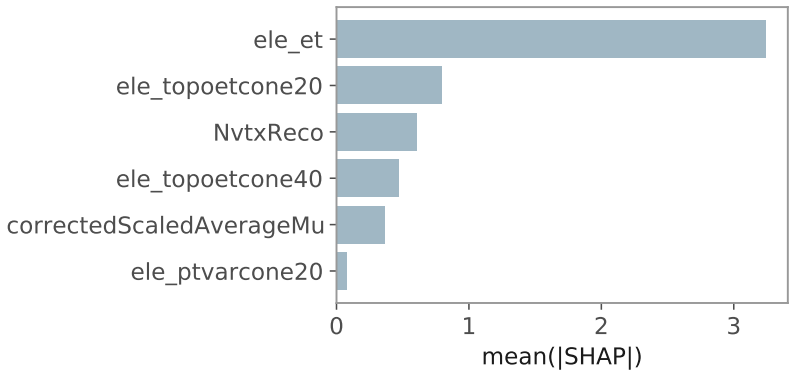


Figure A.13: SHAP values for the eIso model. The variable with the highest influence on the model is `ele_et`.

Fig. A.15 shows the logit transformed LGBM score for the MC Truth trained eIso model, the ROC curve for the test set can be seen in Fig. A.14. The AUC for the eIso model is 0.899 which is a big improvement compared to the ATLAS isolation WP $p_T^{varcone20} < 0.15\,E_T$, yielding an AUC of 0.722. For the T&P trained models, we get an AUC of 0.988 for the MC and 0.896 for the Data trained model, all with increase in performance compared to the ATLAS WP.

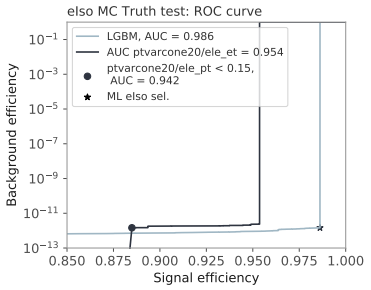


Figure A.14: The ROC curve for the eIso MC Truth test set. The ATLAS WP $p_T^{varcone30} < 0.15\,E_T$ is shown on the figure.

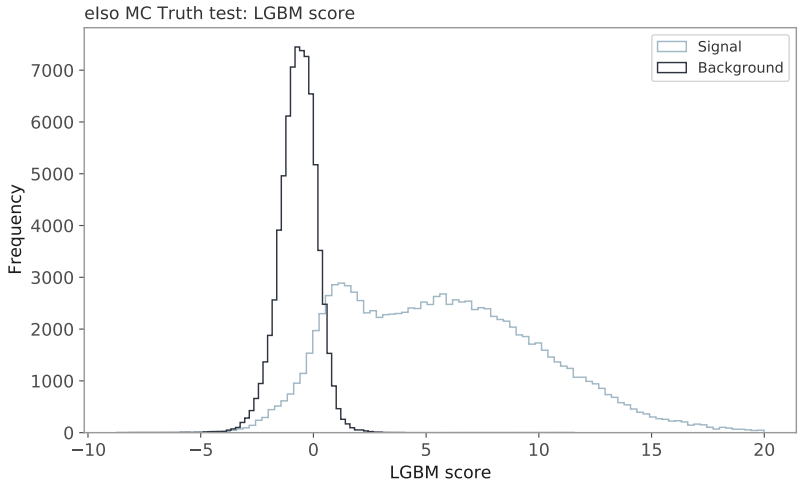
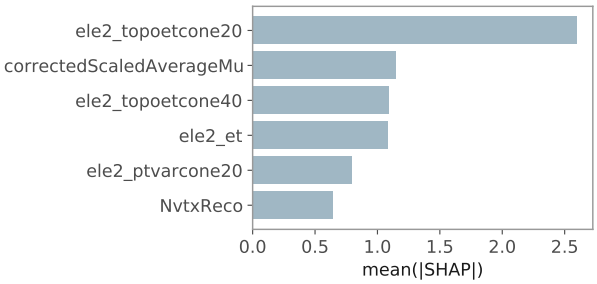


Figure A.15: Logit transformed LGBM score for the final eIso MC Truth model



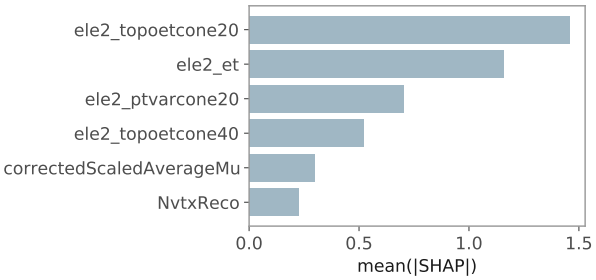
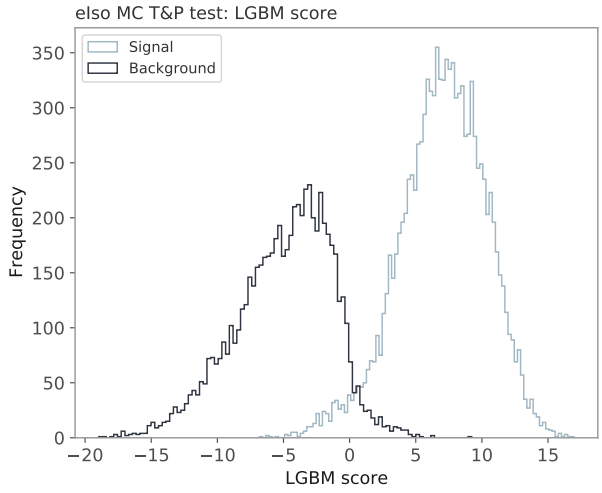


Figure A.16: SHAP values for the MC T&P eIso model. The variable with the highest influence on the model is `muo_numberOfPrecisionLayers`.



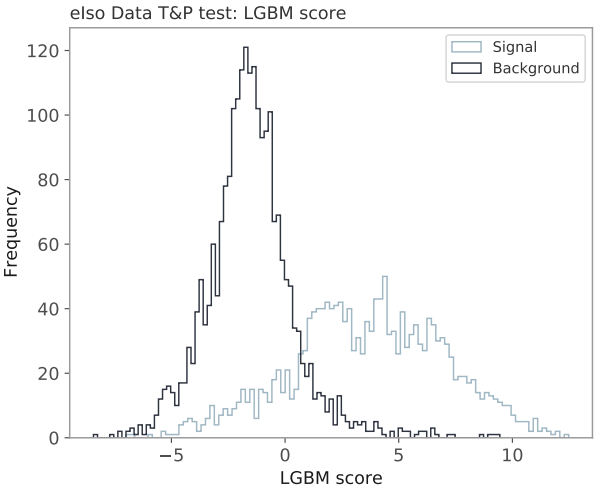


Figure A.17: LGBM score for the MC T&P eIso model.

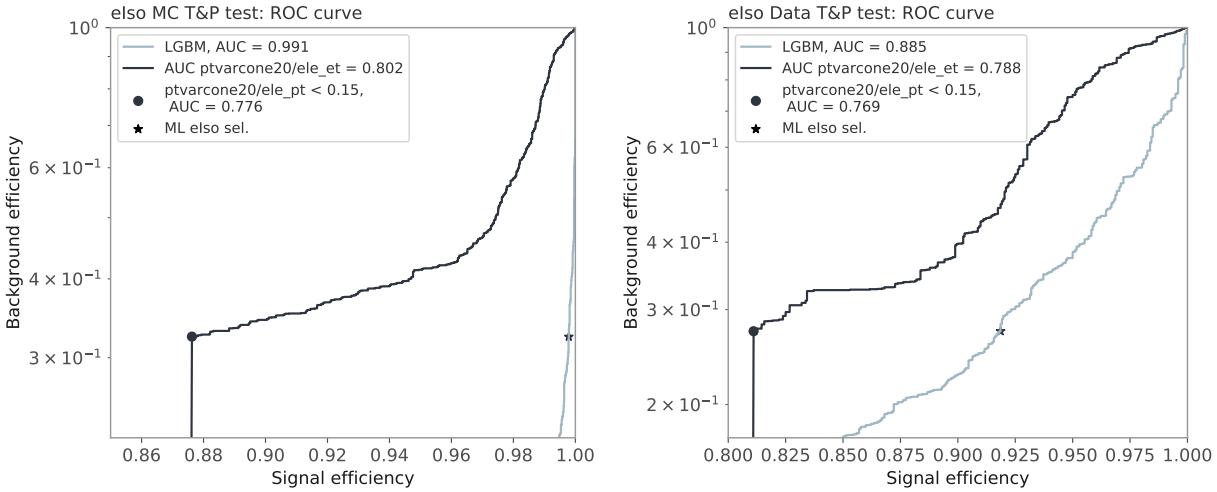


Figure A.18: ROC curve for the MC T&P eIso model

A.2.5 Particle Identification: Photons

The photon models, both the Pid and Iso, are only trained on the MC Truth dataset. The variables for the pPid model, that resembles those used in the ATLAS Likelihood model, can be seen in Tab. A.4. In this case, we will compare the result with the ATLAS Likelihood Loose and Tight working points, as these are the ones available for photons.

The SHAP values for the MC Truth trained pPid model can be seen in Fig. A.19. We see that `pho_Reta` has the largest importance in the model training.

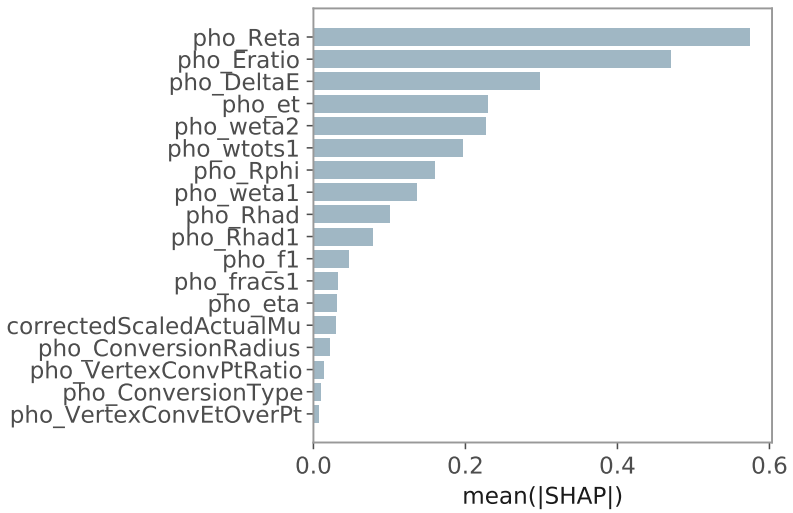


Figure A.19: SHAP values for the pPid model. The variable with the highest influence on the model is `pho_Reta`.

Left plot in Fig. A.20 shows the logit transformed LGBM score for the MC Truth trained pPid model, the ROC curve for the test set can be seen in right plot in Fig. A.20. The AUC for the pPid model is 0.87 which is a big improvement compared to the ATLAS likelihood WP Loose yielding an AUC of 0.741. The Tight working point has an AUC of 0.71, but we can see in Fig. A.20 that we get similar signal and background efficiencies.

ML pPid input features

`correctedScaledAverageMu`	Pileup
`pho_eta`	The pseudorapidity
`pho_et`	The energy
`pho_Rhad`	Ratio of E_T in the hadronic calorimeter to E_T of the EM cluster
`pho_Rhad1`	Ratio of E_T in the first layer of the hadronic calorimeter to E_T of the EM cluster
`pho_weta2`	Lateral shower width in the EM middle layer
`pho_Rphi`	Ratio of the energy in 3x3 cells over the energy in 3x7 cells centered at the electron cluster position
`pho_Reta`	Ratio of the energy in 3x7 cells over the energy in 7x7 cells centered at the electron cluster position
`pho_Eratio`	Ratio of the energy difference between the largest and second largest energy deposits in the cluster over the sum of these energies
`pho_f1`	Ratio of the energy in the strip layer to the total energy in the EM accordion calorimeter
`pho_wtots1`	Lateral shower width in the EM strip layer
`pho_DeltaE`	Difference between the energy of the strip associated with the second maximum in the strip layer, and the energy reconstructed in the strip with the minimal value found between the first and second maxima
`pho_weta1`	
`pho_fracs1`	
`pho_ConversionType`	The type of conversion of the photon. 0 is unconverted
`pho_ConversionRadius`	The radius of conversion
`pho_VertexConvEtOverPt`	
`pho_VertexConvPtRatio`	

Table A.4: Input features for the pPid model [70].

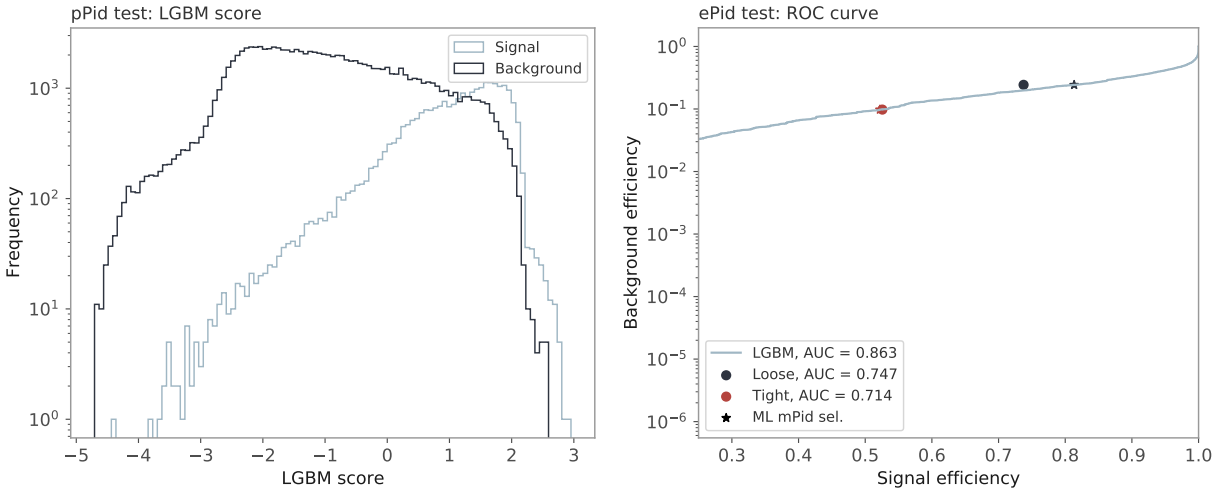


Figure A.20: (left) Logit transformed LGBM score for the final eIso MC Truth model. (right) The ROC curve for the pPid MC Truth test set. The ATLAS WPs Loose and Tight are shown on the figure.

A.2.6 Isolation: Photons

Input features for the photon isolation model can be found in Tab. A.5. They are similar to the features from the mIso and eIso models.

ML pIso input features	
`correctedScaledAverageMu`	Pileup
`NvtxReco`	Number of reconstructed verticies
`pho_et`	The energy E_T of the photon
`pho_topoetcone20`	The sum of E_T tracks in the topo-clusters the 0.20 cone around the object
`pho_topoetcone40`	The sum of E_T tracks in the topo-clusters the 0.40 cone around the object
`pho_ptvarcone20`	A variable of the sum of pT tracks in the 0.20 cone around the object

Table A.5: Input features for the pIso model.

The SHAP values for the MC Truth trained mIso model can be seen in Fig. A.21. We see that `pho_topoetcone20` has the largest importance in the model training.

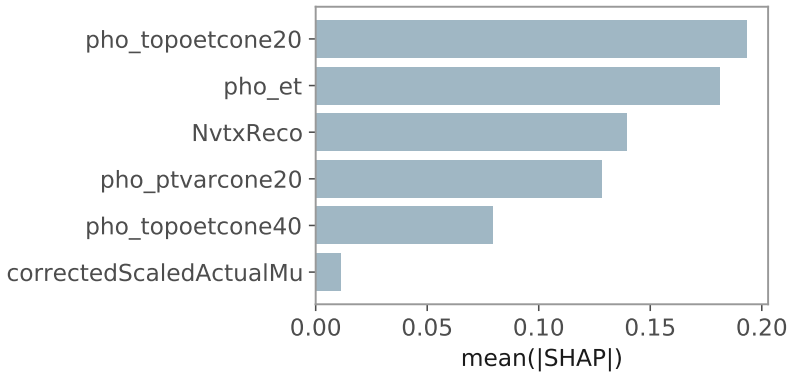


Figure A.21: SHAP values for the pIso model. The variable with the highest influence on the model is `pho_topoetcone20`.

Fig. A.22 (left) shows the logit transformed LGBM score for the MC Truth trained pIso model. The signal and background are not so easily seperable which is also evident in the ROC curve in the right plot in Fig. A.22. The pIso model gets an AUC of 0.893. On the figure, this performance is compared to the ATLAS WP for muons from [41], $p_T^{varcone30} < 0.06\, p_T$ for visualisation purposes. This WP yields an AUC of 0.73.

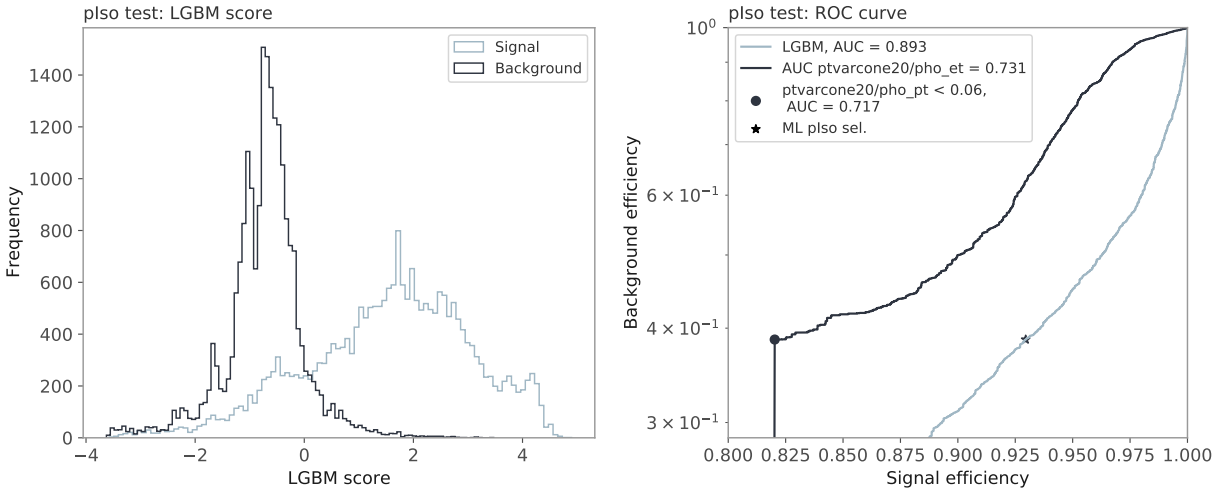


Figure A.22: (left) Logit transformed LGBM score for the final pIso MC Truth model. As can also be seen on the ROC curve in the right plot, the signal and background are less seperable than in the pPid model in Fig. 5.7. (right) The ROC curve for the pIso MC Truth test set. The ATLAS WP $p_T^{varcone30} < 0.06\, p_T$ is shown on the figure. This WP is from [41], and is only shown for visualisation purposes.

A.3 Zll models

A.3.1 Zll Truth

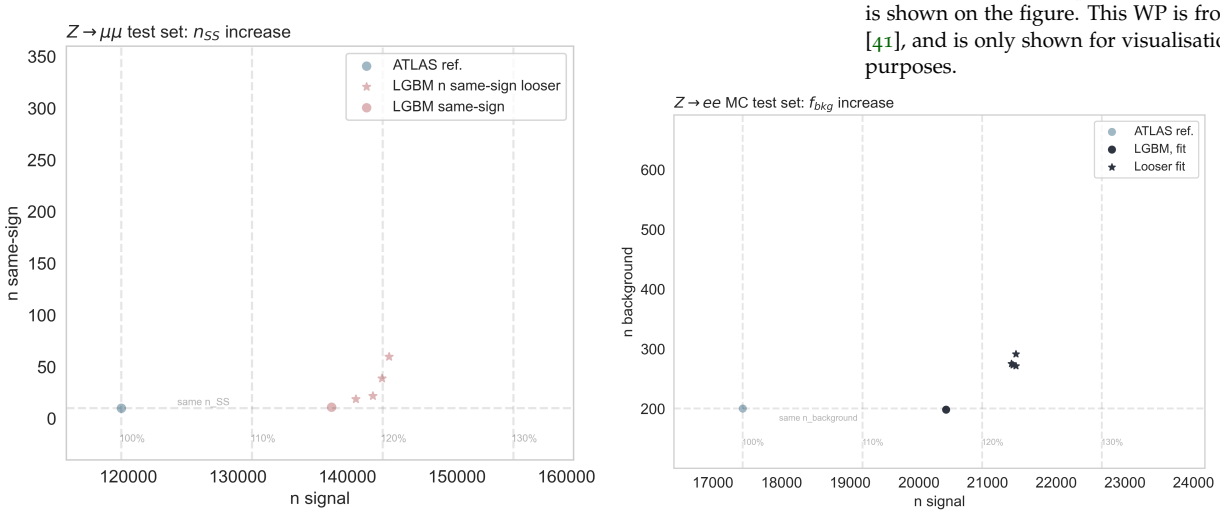


Figure A.23: Number of background events for the $Z \to \mu\mu$ and $Z \to ee$ decays. Left plot shows the number of same-sign events for the $Z \to \mu\mu$ model while the right plot shows the number of background events for the fit for the $Z \to ee$ decay.

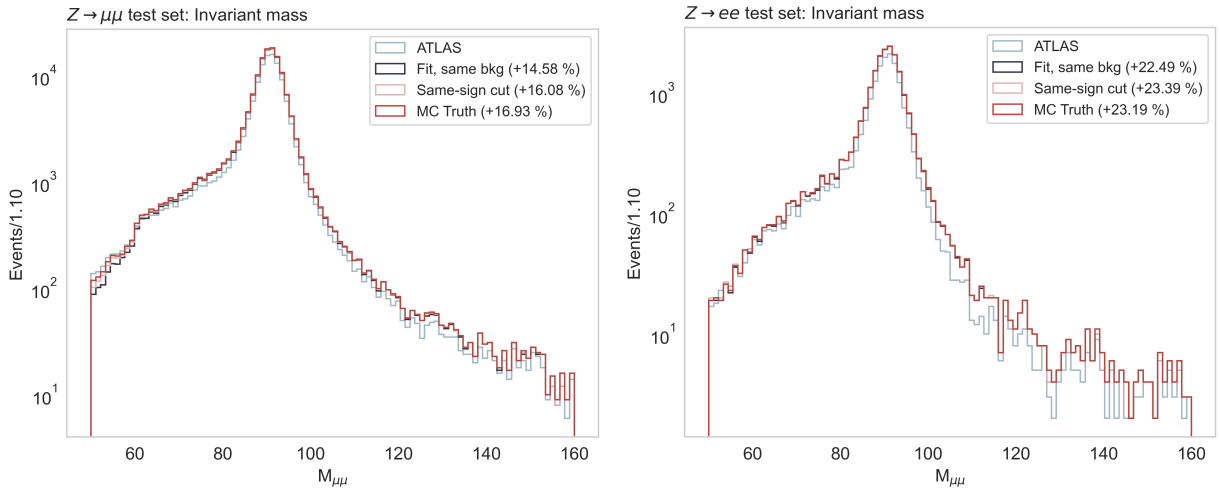


Figure A.24: Invariant mass of the $Z \to \mu\mu$ and $Z \to ee$ decays (log scale)

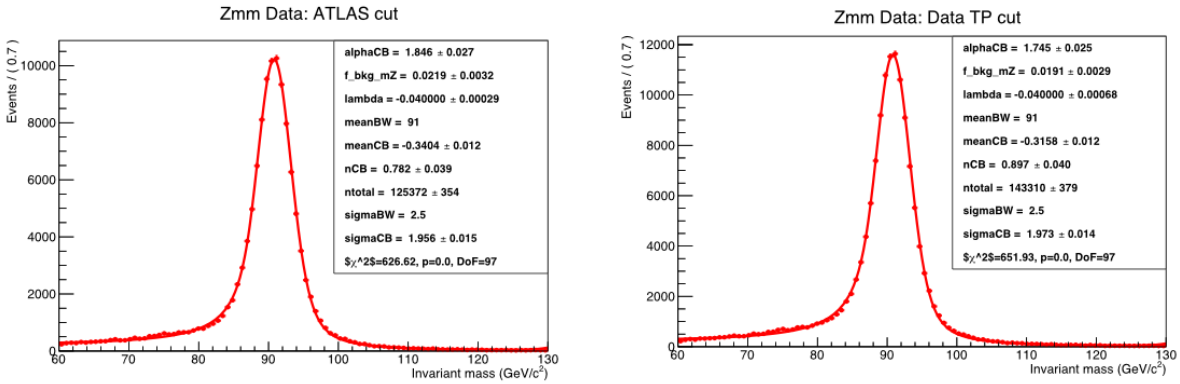


Figure A.25: Examples of fits for the Zmm model. The fit of the ATLAS selection and the Data T&P selection is shown.

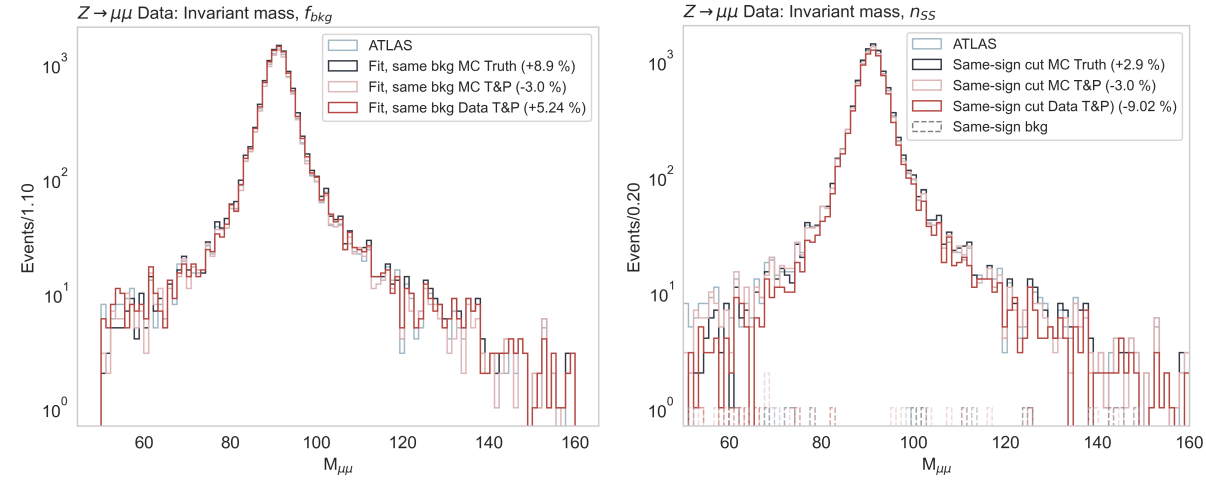


Figure A.26: Plot of the invariant mass for the $Z \to \mu\mu$ decay in Data. The left plot shows the MC Truth, MC T&P and Data T&P performance for the f_{bkg} evaluation. The right plot shows the same models but for the n_{SS} evaluation. Here, the very few same-sign events are also plotted on the figure. Shown in logarithmic scale.

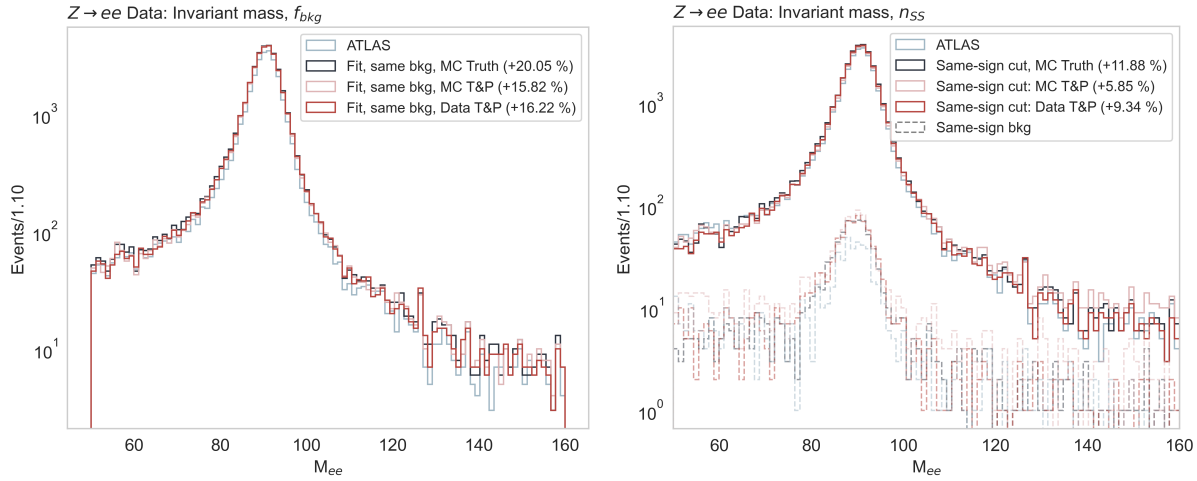


Figure A.27: Plot of the invariant mass for the $Z \to ee$ decay in Data. The left plot shows the MC Truth, MC T&P and Data T&P performance for the f_{bkg} evaluation. The right plot shows the same models but for the n_{SS} evaluation. Here, the same-sign events are also plotted on the figure. Shown in logarithmic scale.

A.3.2 Zmm T&P

A.3.2.1 Training

The T&P models are trained using the same variables as listed in Tab. 5.16. They each take their "own" mPid and mIso score as input, i.e. the MC T&P Zmm model takes the mPid and mIso MC T&P trained models as input.

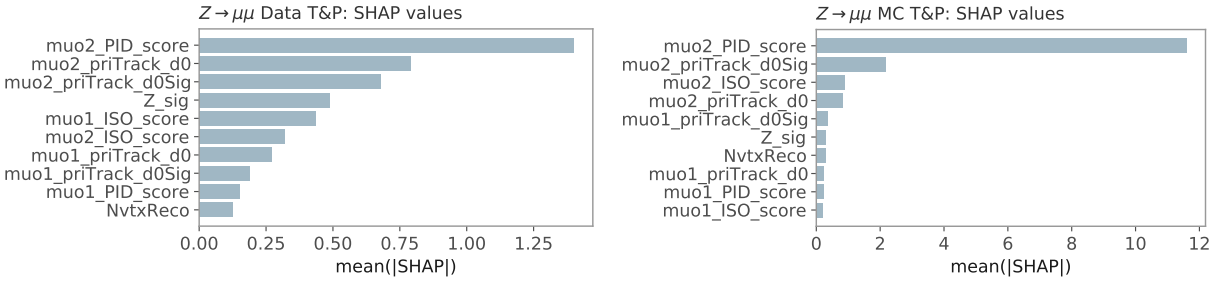


Figure A.28: The SHAP values for the T&P trained Zmm models

The SHAP values for the models are shown in Fig. A.28. For both models, the feature with the highest importance is the mPid score for the second muon. This is as in the MC Truth trained model. However, we see that it is almost the only important variable for the MC trained model, while the rest of the features has some importance for Data trained T&P model.

A.3.2.2 Testing

Fig. A.29 shows the invariant mass for the Data and MC T&P models. As mentioned, the cuts defining the signal and background, removes all data not in the ranges $m_Z \pm 5\,\text{GeV}$ and $m_Z \pm (10-40)\,\text{GeV}$, responsible for the empty tails in the figure.

We see an increase in signal of $10-11\%$ for the Data T&P model and around 19% for the MC T&P model for all evaulation methods.

Fig. A.30 shows the increase in signal and background when loos-

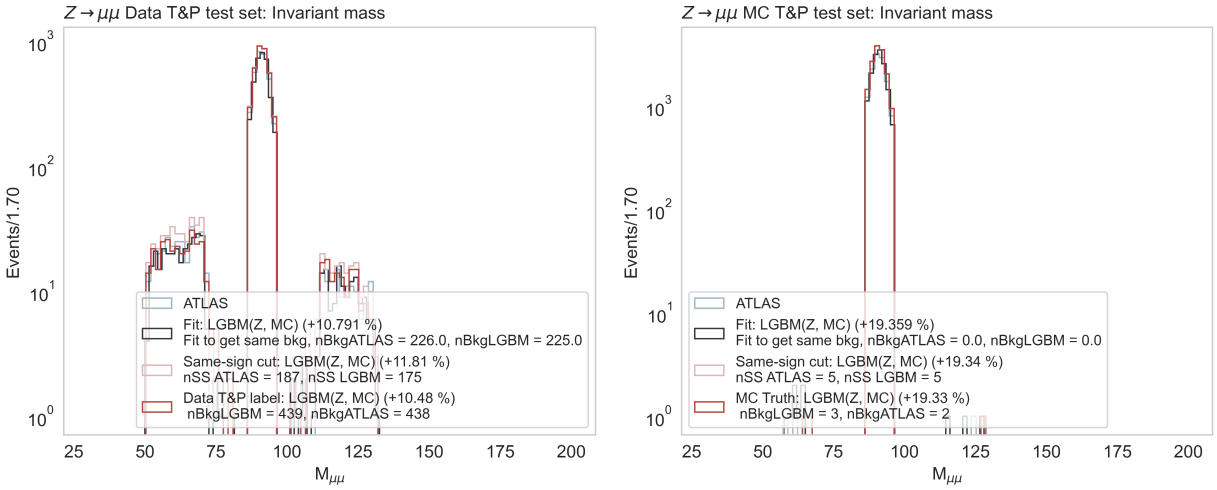


Figure A.29: Plot of the invariant mass for the Zmm Data and MC T&P models.

ening the cuts for the Data T&P model along with the cut in the LGBM score. We see that the fitting method is unable to include more signal, but that loosening the same-sign cut, we can achieve around +20% more signal with the double amount of background.

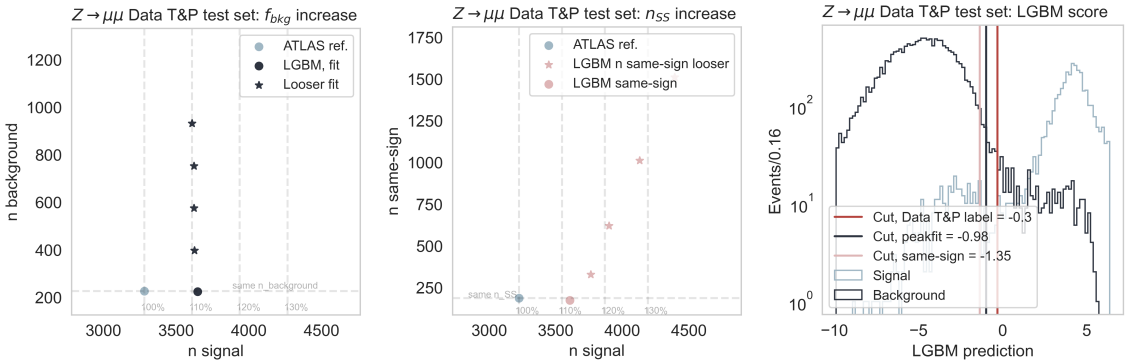


Figure A.30: (left) Signal and background plots for the Zmm Data T&P model. (right) LGBM score for the Zmm Data T&P model.

Fig. A.31 shows the increase in signal and background when loosening the cuts for the MC T&P model along with the cut in the LGBM score. Here, it is evident that there is not much more signal to gain, but the LGBM score also shows that the signal and background are completely seperated.

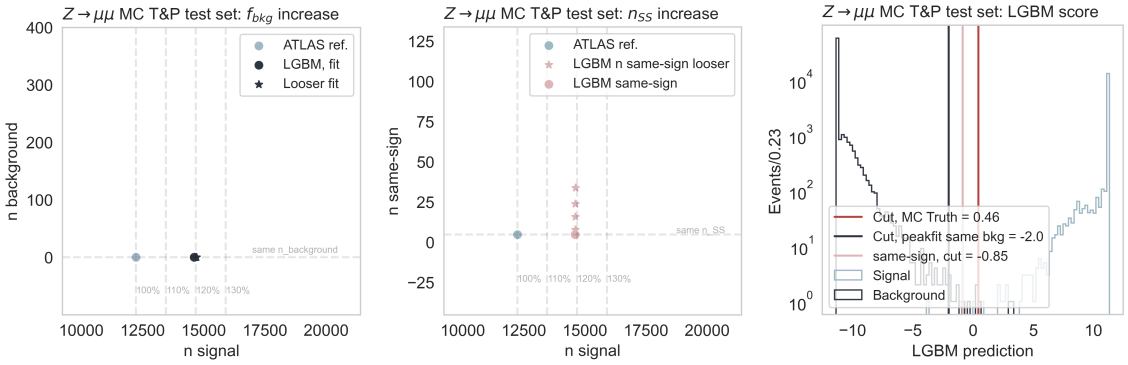


Figure A.31: (left) Signal and background plots for the Zmm MC T&P model. (right) LGBM score for the Zmm MC T&P model.

A.3.3 Zee T&P

A.3.3.1 Training

The T&P models are again trained using the same variables as listed in Tab. 5.16 with their corresponding ePid and eIso model.

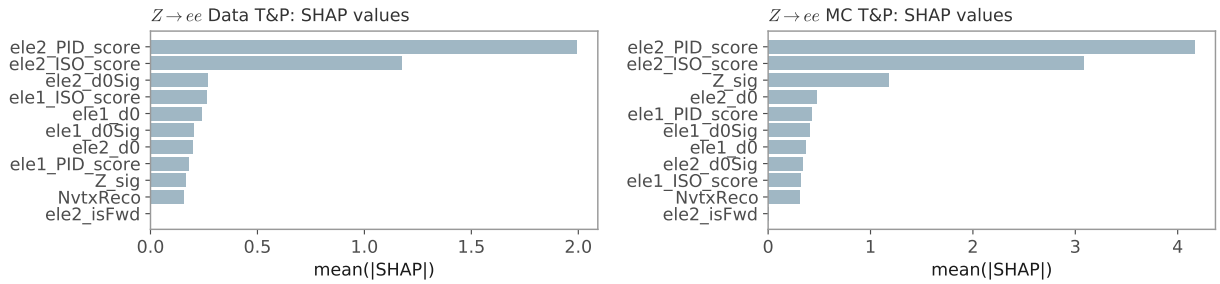


Figure A.32: The SHAP values for the T&P trained Zee models

The SHAP values for the models are shown in Fig. A.32. Again, for both models, the feature with the highest importance is the ePid score for the second election. This is as in the MC Truth trained model. Here, opposed to the muon models, also the eIso score for the second electron has high importance for both models.

A.3.3.2 Testing

Fig. A.33 shows the invariant mass for the Data and MC T&P models. The Data trained model shows an increase of $7-17\%$, meaning that it is highly dependent on the evaluation method. The MC trained model has improvements in the range $24-30\%$.

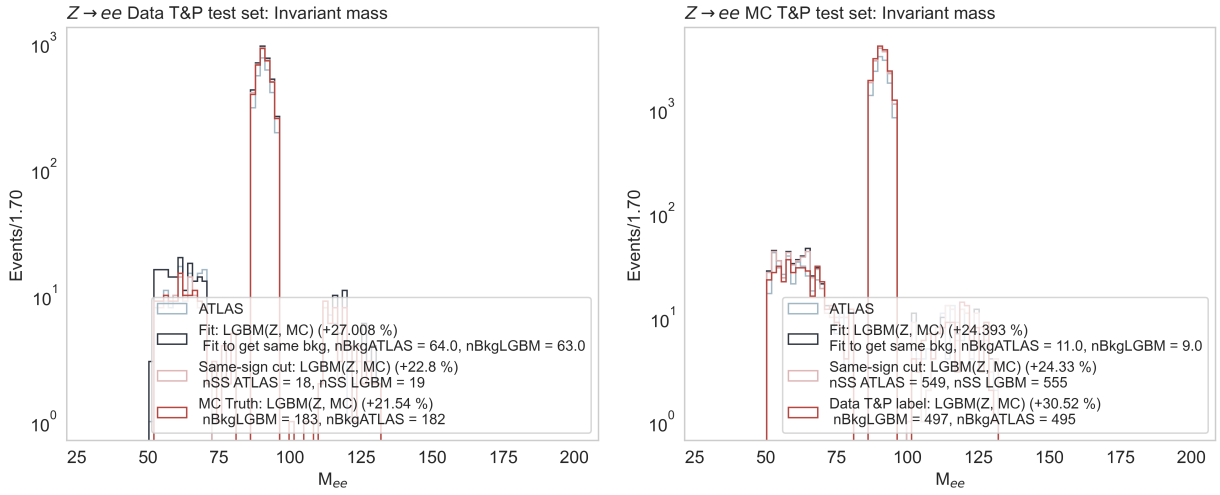


Figure A.33: Plot of the invariant mass for the Zee Data and MC T&P models.

Fig. A.34 shows the increase in signal and background when loosening the cuts for the Data T&P model along with the cut in the LGBM score. Some signal can be achieved by loosening the cuts, but especially for the same-sign method. Here we see an increase of around $+20\%$ with double background.

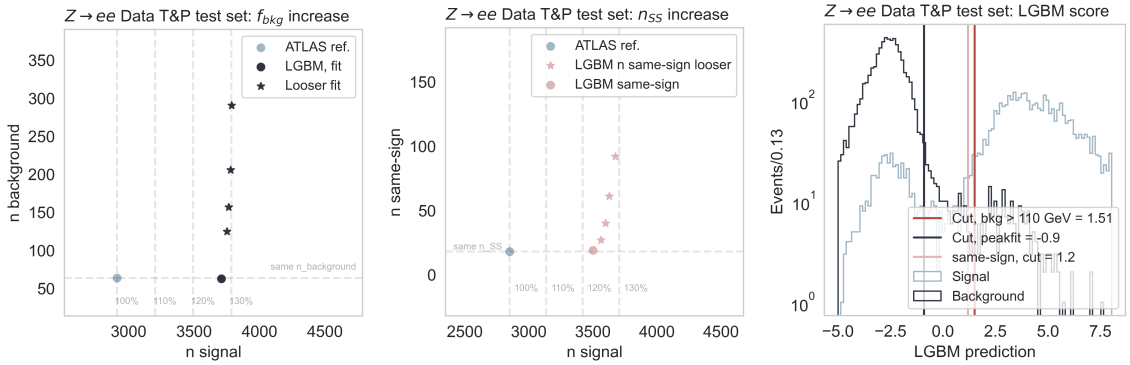


Figure A.34: (left) Signal and background plots for the Zee Data T&P model. (right) LGBM score for the Zee Data T&P model.

Fig. A.34 shows the increase in signal and background when loosening the cuts for the MC T&P model along with the cut in the LGBM score. Here, it is evident that there is not much more signal to gain. The plot of the LGBM score also shows that the selection is already well into the background part of the score.

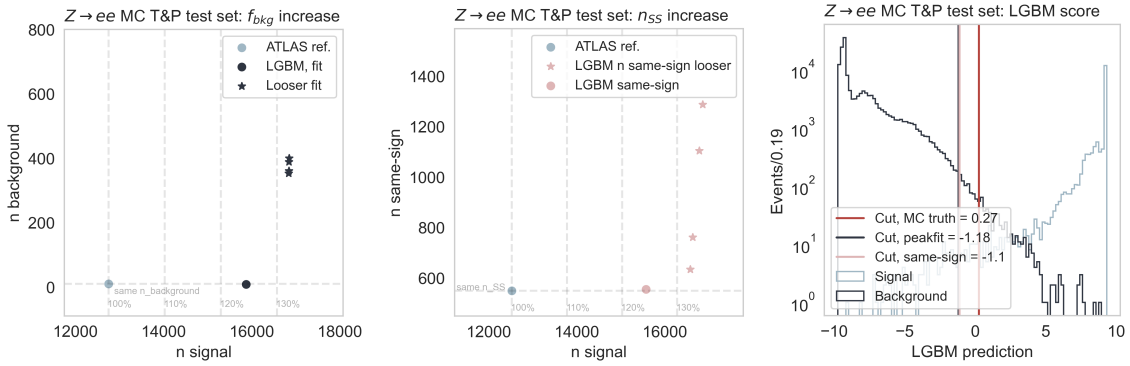


Figure A.35: (left) Signal and background plots for the Zee MC T&P model. (right) LGBM score for the Zee MC T&P model.

Zll models performance, f_{bkg}					**MC**
Model	**N ATLAS**		**N LGBM**		**Change**
Zmm	Sig	Bkg	Sig	Bkg	
MC Truth	122,632	2740	140,575	2735	14.58%
MC T&P	12,233	0	14,603	0	19.37%
Data T&P	3,281	226	3,649	225	10.79%
Zee					
MC Truth	17,435	200	20,396	198	22.49%
MC T&P	12,750	10	15,859	10	24.19%
Data T&P	2,912	64	3710	63	27.01%

Table A.6: Performance of the Zll models for f_{bkg} evaluation. The number of signal events are measured in the range $[70;110]$ GeV while the number of background events are measured for $[50;150]$ GeV.

A.3.4 Predicting in Data

Fig. A.36 shows the relation between signal and background events in Data. When the cuts are loosened, mainly background is let in. This was also seen for MC data in the $Z \to \mu\mu$ channel. In Fig. A.37, the same plot for the $Z \to ee$ decay is shown, indicating that more signal events can be found by loosening the cuts. Fig. A.38 shows

Zll models performance, n_{SS}					**MC**
Model	**N ATLAS**		**N LGBM**		**Change**
Zmm	Sig	Bkg	Sig	Bkg	
MC Truth	119,341	10	138,530	11	16.08%
MC T&P	12,232	5	14,600	5	19.36%
Data T&P	3,225	187	3,606	175	11.81%
Zee					
MC Truth	16,745	13	20,661	12	23.39%
MC T&P	12,524	549	15,551	552	24.17%
Data T&P	2,851	18	3,501	19	22.8%

Table A.7: Performance of the Zll models for n_{SS} evaluation. The number of signal events are measured in the range $[70;110]$ GeV while the number of background events are measured for $[50;150]$ GeV.

the LGBM score for the Data and the Data T&P trained models. The distribution for the Zmm model has a sharper peak in signal for the Data T&P data than the Data, otherwise, the ratio plot shows similar distributions.

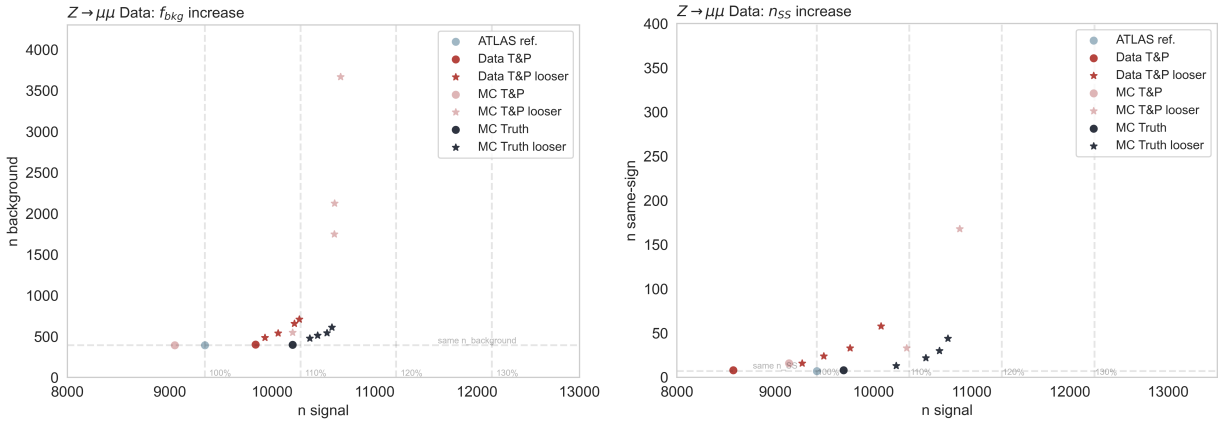


Figure A.36: The number of background events for the Zmm model.

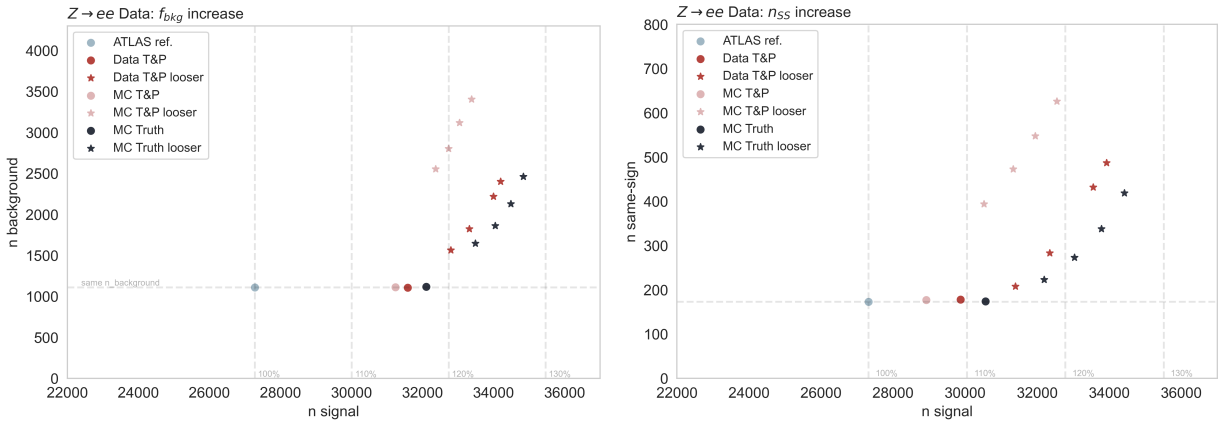


Figure A.37: The number of background events for the Zee model.

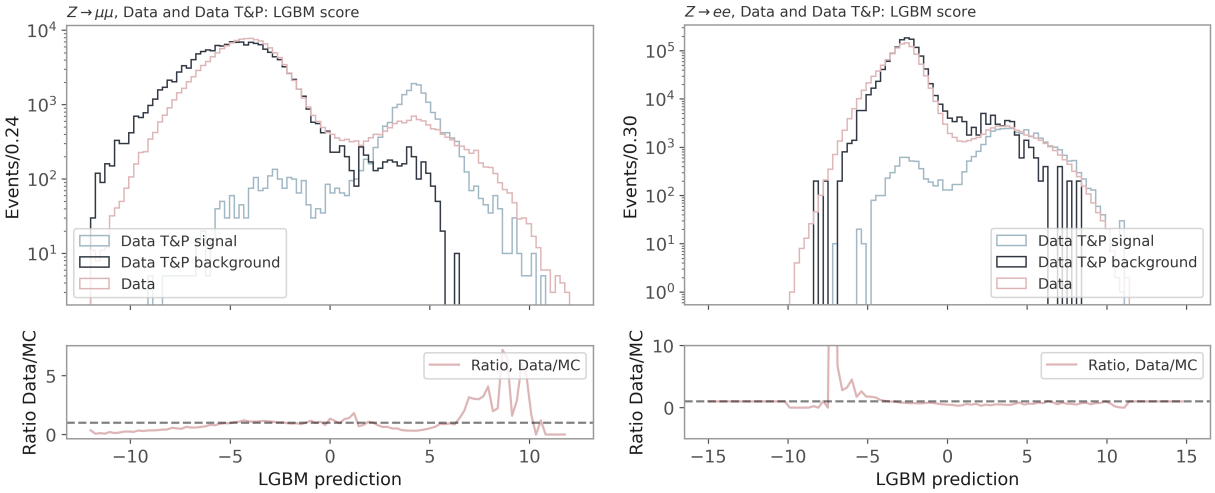


Figure A.38: LGBM score of the Data T&P trained models for Zee and Zmm

A.4 *Zllg and Hllg*

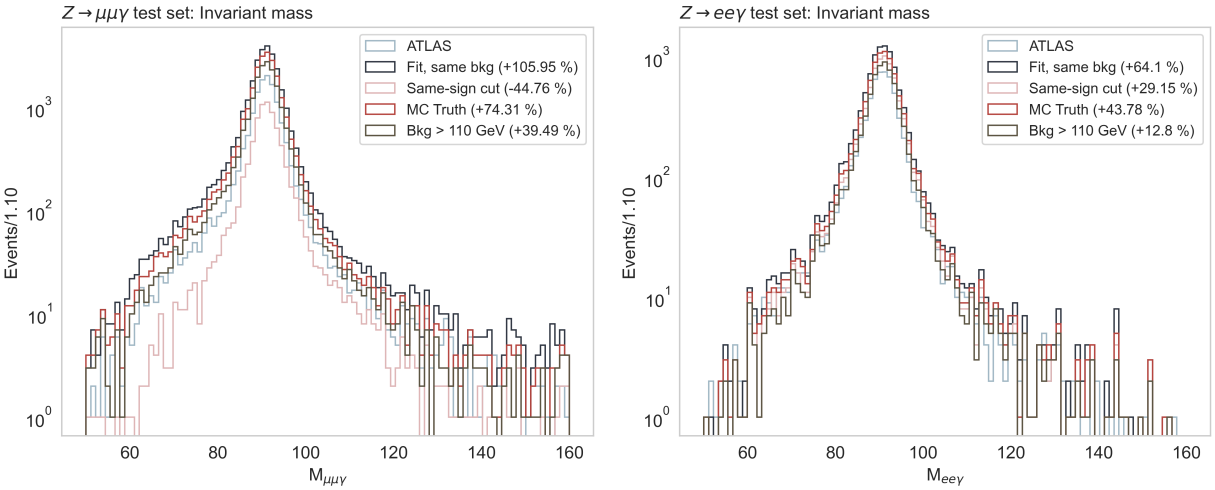


Figure A.39: Plot of the invariant mass for the $Z \to ll\gamma$ decays in MC test. The left plot shows the $Z \to \mu\mu\gamma$ decay and the right plot shows the invariant mass for the $Z \to ee\gamma$ decay

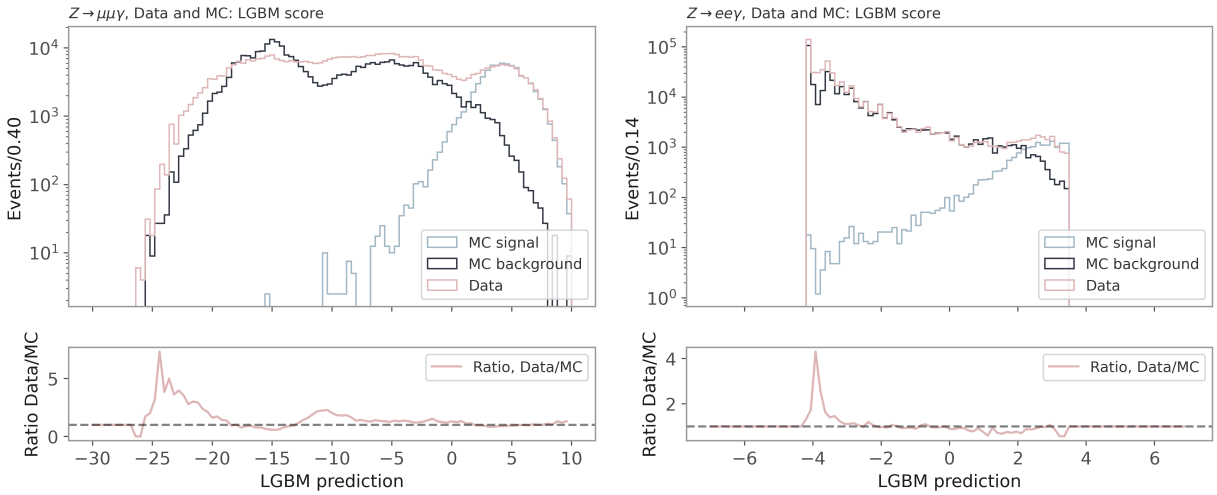


Figure A.40: LGBM score of the Zmmg and Zeeg models for MC and Data

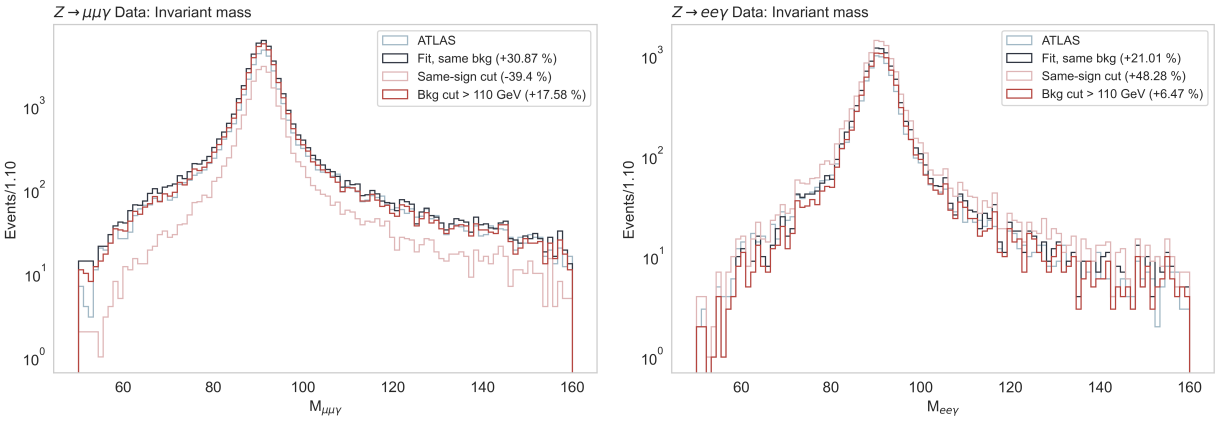


Figure A.41: Plot of the invariant mass for the $Z \to ll\gamma$ decays in Data on logarithmic scale. The top plot shows the $Z \to \mu\mu\gamma$ decay and the bottom plot shows the invariant mass for the $Z \to ee\gamma$ decay and corresponding increases in Data

ML Zllg and Hllg data selection

$Z \to \mu\mu\gamma$	Signal		Background	
	Absolute	Relative	Absolute	Relative
All events	237,725	100%	4,289,018	100%
$83 > M_{ee}[\text{GeV}] > 40$	178,619	75.1%	1,639,869	38.2%
$p_T^\gamma > 10\,\text{GeV}$	142,404	79.7%	168,049	10.2%

Final data sets	Training		Validation		Test	
	data points	percentage	data points	percentage	data points	percentage
Signal	90,856	45.73%	22,924	46.15%	28,624	46.1%
Background	107,834	54.27%	26,748	53.85%	33,467	53.9%

$Z \to ee\gamma$	Signal		Background	
	Absolute	Relative	Absolute	Relative
All events	113156	100%	5860330	100%
$83 > M_{ee}[\text{GeV}] > 40$	85212	75.3%	1547834	26.4%
$p_T^\gamma > 10\,\text{GeV}$	67922	79.7%	174223	11.2%

Final data sets	Training		Validation		Test	
	data points	percentage	data points	percentage	data points	percentage
Signal	43587	28.13%	10854	28.02%	13481	27.84%
Background	111386	71.87%	27889	71.98%	34948	72.16%

$H \to Z(\to \mu\mu)\gamma$	Signal		Background	
	Absolute	Relative	Absolute	Relative
All events	84101	100%	2171535	100%
$p_T^\gamma > 10\,\text{GeV}$	77362	92%	290949	13.3%

Final data sets	Training		Validation		Test	
	data points	percentage	data points	percentage	data points	percentage
Signal	49489	21%	12498	21.21%	15375	20.87%
Background	186229	79%	46432	78.79%	58288	79.13%

$H \to Z(\to ee)\gamma$	Signal		Background	
	Absolute	Relative	Absolute	Relative
All events	82572	100%	6728870	100%
$p_T^\gamma > 10\,\text{GeV}$	75927	92%	972480	14.4%

Final data sets	Training		Validation		Test	
	data points	percentage	data points	percentage	data points	percentage
Signal	48657	7.25%	12118	7.22%	15152	7.23%
Background	622323	92.75%	155627	92.78%	194530	92.77%

Table A.8: Data set selection for Hllg

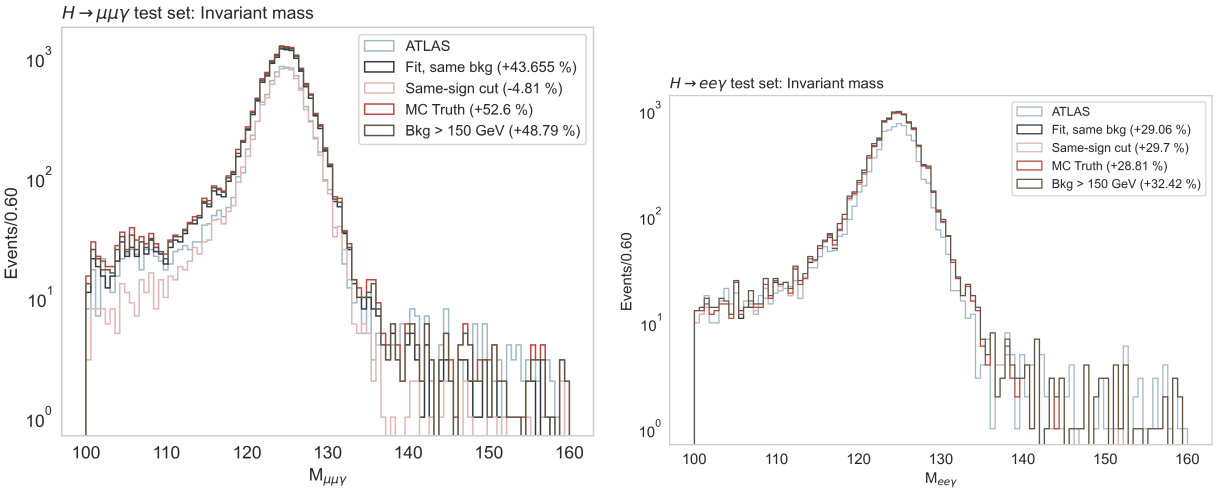


Figure A.42: Plot of the invariant mass for the $H \to Z(\to ll)\gamma$ decays in MC test on logarithmic scale. The top plot shows the $H \to Z(\to \mu\mu)\gamma$ decay and the bottom plot shows the invariant mass for the $H \to Z(\to ee)\gamma$ decay.

Bibliography

[1] ATLAS Collaboration. A search for the $Z\gamma$ decay mode of the Higgs boson in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector. *Physics Letters B*, 809:135754, Oct. 2020. ISSN 03702693. doi: 10.1016/j.physletb.2020.135754. URL http://arxiv.org/abs/2005.05382. arXiv: 2005.05382.

[2] P. W. Higgs. Broken Symmetries and the Masses of Gauge Bosons. *Physical Review Letters*, 13(16):508–509, Oct. 1964. doi: 10.1103/PhysRevLett.13.508. URL https://link.aps.org/doi/10.1103/PhysRevLett.13.508. Publisher: American Physical Society.

[3] F. Englert and R. Brout. Broken Symmetry and the Mass of Gauge Vector Mesons. *Physical Review Letters*, 13(9):321–323, Aug. 1964. doi: 10.1103/PhysRevLett.13.321. URL https://link.aps.org/doi/10.1103/PhysRevLett.13.321. Publisher: American Physical Society.

[4] Peter Higgs. My life as a Boson. URL https://inspirehep.net/files/d2b9efd520f4288b3ba6e47bf1a77132.

[5] T. C. Collaboration. Observation of Top Quark Production in Pbar-P Collisions. *Physical Review Letters*, 74(14):2626–2631, Apr. 1995. ISSN 0031-9007, 1079-7114. doi: 10.1103/PhysRevLett.74.2626. URL http://arxiv.org/abs/hep-ex/9503002. arXiv: hep-ex/9503002.

[6] G. Abbiendi. Search for the Standard Model Higgs Boson at LEP. *Physics Letters B*, 565:61–75, July 2003. ISSN 03702693. doi: 10.1016/S0370-2693(03)00614-2. URL http://arxiv.org/abs/hep-ex/0306033. arXiv: hep-ex/0306033.

[7] 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(1):1–29, Sept. 2012. ISSN 03702693. doi: 10.1016/j.physletb.2012.08.020. URL http://arxiv.org/abs/1207.7214. arXiv: 1207.7214.

[8] 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(1):30–61, Sept. 2012. ISSN 03702693. doi: 10.1016/j.physletb.2012.08.021. URL http://arxiv.org/abs/1207.7235. arXiv: 1207.7235.

[9] ATLAS Collaboration. Search for Higgs boson decays to a photon and a Z boson in pp collisions at sqrt(s)=7 and 8 TeV with the ATLAS detector. *Physics Letters B*, 732:8–27, May 2014. ISSN 03702693. doi: 10.1016/j.physletb.2014.03.015. URL http://arxiv.org/abs/1402.3051. arXiv: 1402.3051.

[10] ATLAS Collaboration. Searches for the $Z\gamma$ decay mode of the Higgs boson and for new high-mass resonances in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Journal of High Energy Physics*, 2017(10):112, Oct. 2017. ISSN 1029-8479. doi: 10.1007/JHEP10(2017)112. URL http://arxiv.org/abs/1708.00212. arXiv: 1708.00212.

[11] CMS Collaboration. Search for a Higgs boson decaying into a Z and a photon in pp collisions at sqrt(s) = 7 and 8 TeV. *Physics Letters B*, 726(4-5):587–609, Nov. 2013. ISSN 03702693. doi: 10.1016/j.physletb.2013.09.057. URL http://arxiv.org/abs/1307.5515. arXiv: 1307.5515.

[12] CMS Collaboration. Search for the decay of a Higgs boson in the $\ell\ell\gamma$ channel in proton-proton collisions at $\sqrt{s} =$ 13 TeV. *Journal of High Energy Physics*, 2018(11):152, Nov. 2018. ISSN 1029-8479. doi: 10.1007/JHEP11(2018)152. URL http://arxiv.org/abs/1806.05996. arXiv: 1806.05996.

[13] Carsten Burgard. Standard model of physics. URL https://texample.net/tikz/examples/model-physics/.

[14] UA1 Collaboration. Experimental observation of lepton pairs of invariant mass around 95 GeV/c2 at the CERN SPS collider. *Physics Letters B*, 126(5):398–410, July 1983. ISSN 0370-2693. doi: 10.1016/0370-2693(83)90188-0. URL https://www.sciencedirect.com/science/article/pii/0370269383901880.

[15] Wikipedia. Weinberg angle. URL https://en.wikipedia.org/wiki/Weinberg_angle.

[16] C. Patrignani et al. (Particle Data Group). Gauge and Higgs Boson Summary Table. URL https://iopscience.iop.org/1674-1137/40/10/100001/media/rpp2016_0029-0031.pdf.

[17] Wikipedia. Drell–Yan process. URL https://en.wikipedia.org/wiki/Drell%E2%80%93Yan_process.

[18] The ATLAS Collaboration. ATLAS simulation of boson plus jets processes in Run 2. Technical report. URL https://cds.cern.ch/record/2261937.

[19] T. W. B. Kibble. Symmetry Breaking in Non-Abelian Gauge Theories. *Physical Review*, 155(5):1554–1561, Mar. 1967. doi: 10.1103/PhysRev.155.1554. URL https://link.aps.org/doi/10.1103/PhysRev.155.1554. Publisher: American Physical Society.

[20] P. W. Higgs. Spontaneous Symmetry Breakdown without Massless Bosons. *Physical Review*, 145(4):1156–1163, May 1966. doi: 10.1103/PhysRev.145.1156. URL https://link.aps.org/doi/10.1103/PhysRev.145.1156. Publisher: American Physical Society.

[21] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. Global Conservation Laws and Massless Particles. *Physical Review Letters*, 13(20):585–587, Nov. 1964. doi: 10.1103/PhysRevLett.13.585. URL https://link.aps.org/doi/10.1103/PhysRevLett.13.585. Publisher: American Physical Society.

[22] P. W. Higgs. Broken symmetries, massless particles and gauge fields. *Physics Letters*, 12(2):132–133, Sept. 1964. ISSN 0031-9163. doi: 10.1016/0031-9163(64)91136-9. URL https://www.sciencedirect.com/science/article/pii/0031916364911369.

[23] cheatham1. The Higgs Boson. URL https://cheatham1.gitbooks.io/get-started/content/the-higgs-boson.html.

[24] ATLAS International Masterclasses. The Higgs boson. URL https://atlas.physicsmasterclasses.org/en/zpath_hboson.htm.

[25] M. Carena et al. Status of Higgs Boson Physics. URL https://pdg.lbl.gov/2016/reviews/rpp2016-rev-higgs-boson.pdf.

[26] The ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *Journal of Instrumentation*, 3(08):S08003–S08003, Aug. 2008. ISSN 1748-0221. doi: 10.1088/1748-0221/3/08/S08003. URL https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08003.

[27] CERN. The ATLAS Experiment. URL https://home.cern/science/experiments/atlas.

[28] Wikipedia. Large Hadron Collider. URL https://en.wikipedia.org/wiki/Large_Hadron_Collider.

[29] iwn. CMS Wiki Pages. URL https://wiki.physik.uzh.ch/cms/latex:example_spherical_coordinates.

[30] X. C. Vidal and R. C. Manzano. Momentum: Taking a closer look at LHC, Visited 13th of January 2021. URL https://www.lhc-closer.es/taking_a_closer_look_at_lhc/0.momentum.

[31] JabberWok. As polar angle approaches zero, pseudorapidity tends towards infinity, Visited 13th of January 2021. URL https://en.wikipedia.org/wiki/Pseudorapidity#/media/File:Pseudorapidity2.png.

[32] E. Torrence. Luminosity Public Results Run2, Visited 13th of January 2021, . URL https://atlas.web.cern.ch/Atlas/GROUPS/DATAPREPARATION/PublicPlots/2018/DataSummary/figs/intlumivstimeRun2.png.

[33] E. Torrence. Luminosity Public Results Run2, Visited 13th of January 2021, . URL https://atlas.web.cern.ch/Atlas/GROUPS/DATAPREPARATION/PublicPlots/2018/DataSummary/figs/mu_2015_2018.png.

[34] University of Cambridge. ATLAS: a general purpose detector for the LHC at CERN. URL https://www.phy.cam.ac.uk/research/research-groups/hep/atlas.

[35] ATLAS Experiment. The Inner Detector, Visited 14th of January 2021. URL https://atlas.cern/discover/detector/inner-detector.

[36] P. Hansen. Particle detectors and accelerators, Lecture notes, third edition, University of Copenhagen, 2016. URL https://www.nbi.dk/~phansen/lecturenote.pdf.

[37] CERN. How a detector works. URL https://home.cern/science/experiments/how-detector-works.

[38] F. Hubaut. Performance of the atlas electromagnetic calorimeter under beam tests. URL https://www.sciencedirect.com/science/article/pii/S0168900203027116.

[39] J. Pequenao. Event cross section in a computer generated image of the atlas detector. URL https://cds.cern.ch/record/1096081.

[40] ATLAS Collaboration. Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at $\sqrt{s}$ = 13 TeV. *The European Physical Journal C*, 79(8):639, Aug. 2019. ISSN 1434-6044, 1434-6052. doi: 10.1140/epjc/s10052-019-7140-6. URL http://arxiv.org/abs/1902.04655. arXiv: 1902.04655.

[41] ATLAS Collaboration. Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at $\sqrt{s}=13$ TeV. *arXiv:2012.00578 [hep-ex]*, Dec. 2020. URL http://arxiv.org/abs/2012.00578. arXiv: 2012.00578.

[42] The ATLAS Collaboration. The ATLAS Simulation Infrastructure. *The European Physical Journal C*, 70(3):823–874, Dec. 2010. ISSN 1434-6044, 1434-6052. doi: 10.1140/epjc/s10052-010-1429-9. URL http://arxiv.org/abs/1005.4568. arXiv: 1005.4568.

[43] M. M.-I. Y. S. Abu-Mostafa and H.-T. Lin. *Learning From Data - a short course*. AMLbook.com, 2012. ISBN 10:1-60049-006-9.

[44] Km121220. Examples of ways to partition a dataset. URL https://en.wikipedia.org/wiki/Training,_validation,_and_test_sets#/media/File:ML_dataset_training_validation_test_sets.png.

[45] J. Bergstra and Y. Bengio. Random Search for Hyper-Parameter Optimization. *The Journal of Machine Learning Research*, Feb. 2012. doi: 10.5555/2188385.2188395. URL https://dl.acm.org/doi/pdf/10.5555/2188385.2188395.

[46] Microsoft Corporation. Revision 42d1633a. LightGBM Documentation. URL https://lightgbm.readthedocs.io/en/latest/index.html.

[47] L. F. Ehrke. Machine Learning and Electrons. Master's thesis, Niels Bohr Institute, University of Copenhagen, 9 2019.

[48] scikit-learn developers. Decision Trees, 2020. URL https://scikit-learn.org/stable/modules/tree.html.

[49] Wikipedia. Decision tree learning: Gini Impurity. URL https://en.wikipedia.org/wiki/Decision_tree_learning#Gini_impurity.

[50] T. K. Ho. Random decision forests. URL https://ieeexplore.ieee.org/document/598994/.

[51] Cheng Li. A Gentle Introduction to Gradient Boosting. URL http://www.chengli.io/tutorials/gradient_boosting.pdf.

[52] Microsoft Corporation. Revision 42d1633a. LightGBM Documentation: Parameters Tuning. URL https://lightgbm.readthedocs.io/en/latest/Parameters-Tuning.html.

[53] Yandex; Alex Rogozhnikov and contributors. hep ml documentation, . URL https://arogozhnikov.github.io/hep_ml/index.html.

[54] Yandex; Alex Rogozhnikov and contributors. hep ml documentation: Reweighing algorithms, . URL https://arogozhnikov.github.io/hep_ml/reweight.html.

[55] ATLAS TWiki. Derivation Framework. URL https://twiki.cern.ch/twiki/bin/view/AtlasProtected/DerivationFramework#What_is_the_derivation_framework.

[56] Giovanni Marchiori. egamma AOD derivations: Derivations defined for egamma. URL https://twiki.cern.ch/twiki/bin/view/AtlasProtected/EGammaxAODDerivations#Derivations_defined_for_egamma.

[57] Dongliang Zhang. MCPDxAOD: Definitions. URL https://twiki.cern.ch/twiki/bin/view/AtlasProtected/MCPDxAOD#Definitions.

[58] Sandrine Laplace. HGamma DxAOD samples: Information about release 20.7. URL https://twiki.cern.ch/twiki/bin/view/AtlasProtected/HggDerivationSamples?fbclid=IwAR3KU_uAB2areIKw0fBXiTC6YnXgB5fgOHZT8HdZAZBqHcM7OcJACyNa1eo#Information_about_release_20_7.

[59] H. K. Leerberg. Higgs Decay to Zgamma — A Study of Efficiency. Master's thesis, Niels Bohr Institute, University of Copenhagen, 8 2020. URL https://www.nbi.dk/~petersen/HelleKogsboellLeerberg_MasterThesisFinal_17aug2020_AsHandedIn.pdf.

[60] Disciples of Troels. Ntuple production. URL https://gitlab.cern.ch/disciples-of-troels/ntupleproduction.

[61] Particle Wiki. Overlap removal (ATLAS). URL https://particle.wiki/wiki/Overlap_removal_(ATLAS).

[62] Andrew Collette and contributors. HDF5 for Python. URL https://docs.h5py.org/en/stable/.

[63] CERN. MC Truth Classifiers Definitions, . URL https://gitlab.cern.ch/atlas/athena/blob/21.2/PhysicsAnalysis/MCTruthClassifier/MCTruthClassifier/MCTruthClassifierDefs.h.

[64] CERN. PdgId, . URL https://twiki.cern.ch/twiki/bin/view/Main/PdgId.

[65] Scott Lundberg. shap, 2021. URL https://github.com/slundberg/shap.

[66] Wikipedia. Logit, 2021. URL https://en.wikipedia.org/wiki/Logit.

[67] The ATLAS Collaboration. Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data, June 2016. URL https://cds.cern.ch/record/2157687. Institution: ATLAS-COM-CONF-2016-028 Number: ATLAS-CONF-2016-024.

[68] 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): P12006–P12006, Dec. 2019. ISSN 1748-0221. doi: 10.1088/1748-0221/14/12/P12006. URL http://arxiv.org/abs/1908.00005. arXiv: 1908.00005.

[69] L. N. Wentzel. Using Machine Learning on electron identification and energy regression in the forward region in ATLAS. Master's thesis, Niels Bohr Institute, University of Copenhagen, 1 2020.

[70] Dongliang Zhang and Matthew Klein. Run2IsolationHarmonisation, 2019. URL https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/Run2IsolationHarmonisation.

List of Figures

List of Tables