

Electron Identification using PICNN

Department of Physics and Astronomy
The University of Sheffield, UK

September 13, 2024

Anirudh Malik

Abstract

Electron identification in proton-proton collisions at the ATLAS detector is the main subject of this thesis. The tag-and-probe methodology is used to measure the efficiency of electron identification, and the sPlot statistical method is used to extract background information from shower shape variables for estimation. Discrepancies in electron identification are investigated by comparing data with Monte Carlo simulations (AF3 and Geant4). By analyzing important characteristics related to the form of the shower, which differentiates electrons from background, and by applying optimal transport, the alignment between the data and simulations is improved.

1 Introduction

Investigating the fundamental units of matter and their interplay is the domain of particle physicists. Weak, strong, and electrical forces are the three pillars upon which the standard model of particle physics rests. For now, the usual model only doesn't explain gravity.

One of the main goals of the ATLAS (Section:[1.3]) at the LHC is to correctly identify and reconstruct particles that are made when particles collide. Elementarily, electrons are very important, especially in actions involving the Z boson (Section:[1.2.2]), which breaks down into an electron-positron pair (Section:[1.2.3]). Finding and putting together the right electrons is very important if we want to get accurate measurements of how particles interact, check the Standard Model (Section:[1.2]), and search for new physics.

Finding electrons is a big part of this thesis. There are a lot of complicated methods we use, such as the tag-and-probe method (Section:[3.2]), the sPlot statistical method (Section:[3.7.1]), and AF3 (Section:[4.1.1]) and Geant4 (Section:[4.1.2]) Monte Carlo models to compare our data. These techniques are very important for studying how electrons behave in high-energy collisions because they keep out signals that aren't needed and only use signals from real electrons in the analysis.

Things going on in the background, like photon conversions and jets being confused for electrons, make it harder to find and recover electrons (Section: 2]). Electron detection uses a variety of detector data, including shower shapes (Section:[3.6]), isolation cuts (Section:[2.2.1]), and track-

cluster matches (Section:[2.1]), to deal with these issues. This thesis looks at the differences between data (Rel 22, 2018 data) and Monte Carlo models (AF3 and Geant4). To make electron identification methods (Section:[3]) work better and check them, these changes are important.

This study also looks at optimal transport theory (Section:[5]) as a new way to make virtual and real data more similar. This method cuts down on differences between distributions, which usually improves the accuracy of electron identification. We want to help people learn more about how particles interact with each other, especially when the Z boson and electrons are present. This will be done by making these detecting methods better.

1.1 Outline of the Thesis

- Work has been undertaken to refine the electron identification process within the ATLAS detector utilising the tag-and-probe method (Section:[3.2]). This approach is crucial for validating the performance and accuracy of electron identification algorithms by identifying electron positron pairs originating from Z boson decays and separating them from false electron positron pairs.
- The thesis compares Z mass data with an emphasis on signal-to-background separation. Methods like the polynomial fitting (Section:[3.5.2]), which is good at simulating background changes, and the Breit-Wigner fit (Section:[3.5.1]), well known for accurately fitting resonance shapes, are utilised to fit the background and compared to see which performs better using the chi-squared test (Section:[4.2]). The purpose of this comparison is to find out how these fitting models affected the derived Z mass signal's accuracy and reliability.
- Another focus of this study is on the features of the shower shape (Section:[3.6]) that distinguish electron-induced showers from other particle showers using the criteria (Section:[3.1.2]). This thesis delves into many methods for optimising shower form analysis, such as alternative Monte Carlo simulation frameworks, sophisticated fitting algorithms for refining shape extraction (Section:[3.7]), and the loose catch method (Section:[3.1.2]) for less strict data selection. In order to learn how well these models represent electromagnetic showers and how they help with precise electron identification, we compare them.
- In addition, the thesis explores how optimal transport theory (Section:[5]) can be used to make the match between Monte Carlo models and real data better. By moving mass around in the best way possible, the goal of optimal transport is to close the gap between groups as much as possible. The goal of this thesis is to improve the precision of MC-to-data comparisons by using this theory. This will lead to more accurate models and fewer differences. This method might make electron recognition much more accurate, which would lead to more accurate models of particle interactions and better agreement between what was simulated and what was observed.

1.2 Standard Model

The Standard Model [1], which is the main theory of modern particle physics (except for gravity), explains all known fundamental particles and three of the four basic interactions. These are weak, strong, and electromagnetic. It has correctly predicted and shown the existence of quarks, gluons, vector bosons, the tau-neutrino, and the top-quark [2], among other particle properties and behaviours, since it was made fifty years ago. The most recent major revision was in 2002

when neutrino oscillations were explained; in 2013, the T2K experiment confirmed this. The confirmation of the Higgs boson [3] in 2013 was another significant breakthrough.

1.2.1 Ingredients of the Standard Model

These are the fundamental particles [4] & [5] that comprise our universe according to the Standard Model:

- **Quarks:** Participate in all three interactions; they possess colour, weak, and electromagnetic charges.
- **Leptons:** Except strong interactions, they engage in other type of interactions.
- **Gauge Bosons:** Force carriers for the interactions.
- **Gluons:** Strong force carrier particles belonging to SU(3) group.
- **Z and W^\pm bosons:** carriers of weak interactions SU(2) group.
- **Photons:** Carriers of electromagnetic force belonging to U(1) group.
- **Higgs Boson:** Particle, which is known as the reason of mass for anything in the universe. [6]

Parameter	Description	Value
m_u	Up quark mass	2.3 MeV
m_d	Down quark mass	4.8 MeV
m_s	Strange quark mass	95 MeV
m_c	Charm quark mass	1.275 GeV
m_b	Bottom quark mass	4.18 GeV
m_t	Top quark mass	173.5 GeV
m_e	Electron mass	511 keV
m_μ	Muon mass	105.7 MeV
m_τ	Tau mass	1.78 GeV
m_Z	Z boson mass	91.18 GeV
m_W	W boson mass	80.35 GeV
m_H	Higgs boson mass	126 GeV

Table 1: The list of the parameters of the Standard Model, assuming the masses of neutrinos to be zeros (this is a valid approximation since the mass of neutrinos is very small, $m_\nu < 2$ eV). [7]

The weak and electromagnetic forces come together in the $SU(2) \times U(1)$ gauge group. This is called the electroweak interaction. The Higgs field breaks the symmetry by itself, making the heavy and 0 bosons and the light photon with no mass. Quantum Chromodynamics (QCD) [8–10] says that the $SU(3)$ symmetry group controls the strong interaction. Gluons are the force carriers that hold quarks together.

Even though the Standard Model works, it doesn't explain a lot of things. For example, what dark matter is, how neutrinos work, and how forces at very high energies come together. Particle physics study is being driven by these open questions, both now and in the future.

1.2.2 Z Boson and Electron Interactions

The role of the Z boson in neutral weak interactions is a key part of the Standard Model that is important to this argument. The decay of Z bosons, especially into pairs of electrons and positrons $Z \rightarrow e^+ + e^-$, is an exact way to study how particles interact and test what the Standard Model says should happen.

To accurately measure the properties of the Z boson, the electron identification method needs to be able to find and rebuild electrons, which are the main products of decay. So, knowing about the Z boson and how it interacts with other particles is very important to this thesis because it lets us measure how well electron detection works and how much background noise is blocked.

1.2.3 Drell-Yan process

In high-energy proton-proton collisions, like those at the LHC, the Drell-Yan process is what makes electron-containing lepton pairs. When the quark and antiquark of a proton hit each other, they make a simulated photon or Z boson. This boson will then break down into two leptons.

This thesis is mostly about figuring out the electrons that are made when Z decays. The Drell-Yan process is also a good way to learn more about the Z boson's features. tag-and-probe and shower shape analysis are two methods that will be used to boost the accuracy of electron recognition. Experiment data will also be compared with Monte Carlo simulations like AF3 and Geant4.

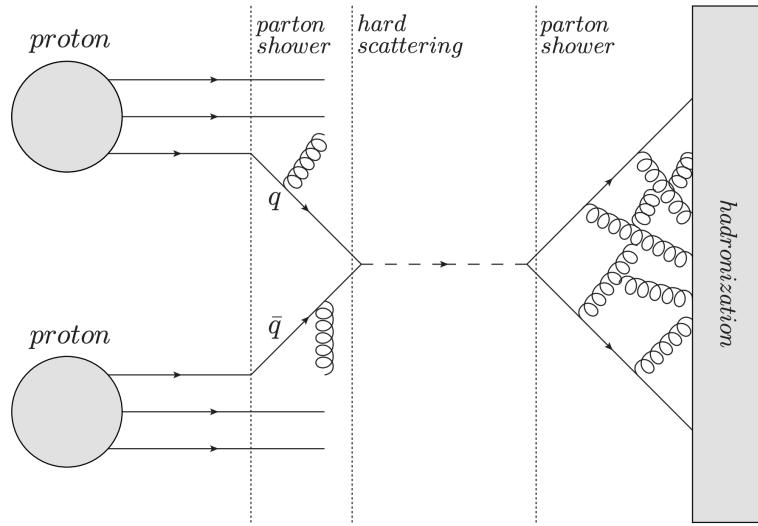


Figure 1: There are three distinct stages to inelastic two-proton scattering: initial radiation, hard scattering (which produces a gauge boson), and final state radiation. When the gauge boson decays into quarks, the quarks undergo hadronization and become color-neutral hadrons.

1.3 ATLAS

The ATLAS detector [11] is a significant general-purpose detector at the Large Hadron Collider (LHC), with a specific focus on studying proton-proton and heavy ion collisions. The fifteen-year development, construction, installation, and operation of ATLAS was the joint effort of many scientists, engineers, and technicians. The ATLAS collaboration uses the detector to support

a wide variety of experiments, which include B-meson studies, supersymmetry investigations, Z and W boson decays, Higgs boson studies, and more. A high degree of complexity and accuracy in the various detector subsystems is required for the wide range of currently underway research.

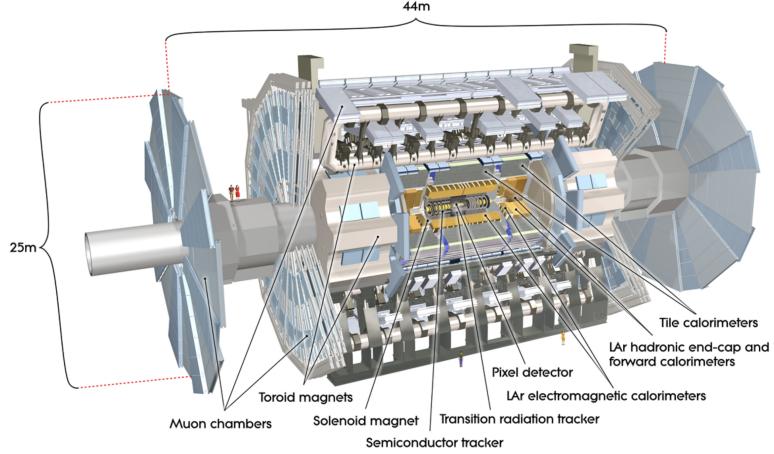


Figure 2: ATLAS detector, with all the major sub-systems and the human-sized scale figures. [12]

1.3.1 Physics Requirements and An Overview of the Detector

Located at the interaction point, the cylindrical ATLAS detector is positioned along the LHC beamline. The geometry of the detector is well-suited to its coordinate system [11, 12], which uses the interaction point as its origin, a z-axis that follows the beam pipe, and a transverse x-y plane. Axis positives indicate a direction towards the LHC ring centre, and y-axis positives indicate an upward direction. To create a right-handed coordinate system, the z-axis is selected. Around the beam axis is where the azimuthal angle, ϕ , is measured, and from the z-axis, the polar angle, θ , is found. The pseudo-rapidity η is related to the polar angle θ (the angle from the z-axis) by $\eta = -\ln \tan \frac{\theta}{2}$.

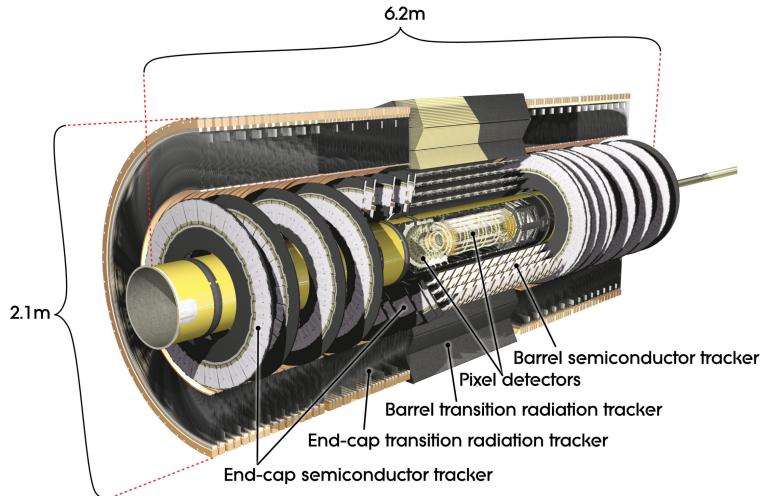


Figure 3: SCT, TRT and pixel detectors are shown. [11]

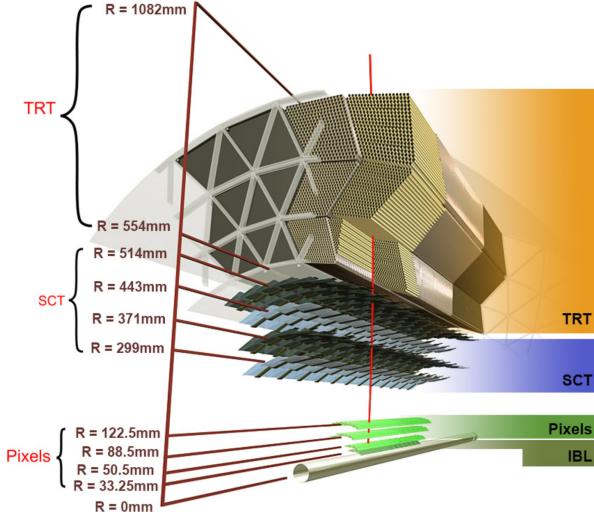


Figure 4: The ATLAS inner detector [13]. Shown are the main components, pixel tracker – SCT – TRT, as well as the distances from the beam.

1.3.2 Inner Detector (ID)

The Inner Detector (ID) is responsible for reconstructing the trajectories of charged particles produced in collisions. It spans a total length of approximately 7 meters and a radius of 1.15 meters, covering a pseudorapidity range of $|\eta| < 2.7$. The ID is crucial for precise momentum measurements, accurate primary vertex reconstruction, and for separating pile-up (PU) interactions, especially in high-luminosity environments. The spatial resolution of the ID for the transverse impact parameter d_0 , which is the shortest distance between the particle track and the z-axis, is $10 \mu\text{m}$. The detector is shown in Figure 4.

Pixel Detector

The Pixel Detector is the innermost layer of the ID and consists of four layers in the barrel and three disks in each end-cap. Since Run-2, the Insertable B-Layer (IBL) has been added as the innermost layer, significantly enhancing the resolution of the detector. The typical pixel size for the original layers is $50 \mu\text{m} \times 400 \mu\text{m}$, while the IBL features smaller pixels at $50 \mu\text{m} \times 250 \mu\text{m}$ [11]. The pixel detector provides intrinsic accuracies of $10 \mu\text{m}$ ($R - \psi$) and $115 \mu\text{m}$ (z) in the barrel, and $10 \mu\text{m}$ ($R - \psi$) and $115 \mu\text{m}$ (R) in the disks. [14, 15]

The IBL adds an additional 12 million pixels, increasing the total to 80 million pixels for precise vertex identification and improved tracking efficiency, especially for particles with low transverse momentum.

Silicon Microstrip Detector (SCT)

The Silicon Microstrip Detector (SCT) consists of four extra cylindrical tracking layers in the barrel and nine discs in each end-cap. The addition of 6.3 million readout channels enhances the tracking precision even more. The SCT layer comprises two sets of silicon micro-strips, enabling the measurement of both $R - \psi$ coordinates. The strips within the barrel are orientated parallel to the axis of the beam, whilst the strips in the end-cap are organised in a radial manner.

Strip Size and Resolution:

Each strip is 6.4 cm long with a strip width of $80 \mu\text{m}$, providing intrinsic accuracies of $17 \mu\text{m}$

$(R - \psi)$ and $580 \mu\text{m}$ (z) in the barrel and $17 \mu\text{m}$ ($R - \psi$) and $580 \mu\text{m}$ (R) in the disks. [11]

Transition Radiation Tracker (TRT)

The Transition Radiation Tracker (TRT) incorporates straw-tube detectors to introduce an extra tracking layer, hence greatly augmenting the total tracking capabilities of the ID. The TRT has 351,000 readout channels, with each straw tube having a diameter of 4 mm and a length of 140 cm. The TRT encompasses a pseudorapidity range of $|\eta| < 2.0$, offering 36 hits per track with a spatial resolution of $130 \mu\text{m}$ in the $R - \psi$ plane.

The TRT also plays a key role in electron identification through the detection of transition radiation, emitted when electrons pass through the straw tubes filled with a xenon-based gas mixture. However, some parts of the detector use an argon-based gas mixture, which reduces electron identification performance due to gas leaks in certain areas. [16]

The Pixel Detector, SCT, and TRT are the three tracking subsystems that are inside a 2 T solenoidal magnetic field. This lets us measure momentum accurately over a large range, from 0.5 GeV to 150 GeV.

1.3.3 Electromagnetic Calorimeter (ECAL)

Electromagnetic Calorimeter (ECAL) is responsible for measuring the energy of electrons and photons. It plays a major role in the reconstruction of electrons. It helps differentiating electrons from other particles like hadrons and muons. The ECAL also captures the electromagnetic showers produced when high-energy electrons and photons interact with the calorimeter material. [11]

Structure of the ECAL

ECAL in ATLAS is a liquid argon (LAr) sampling calorimeter, composed of back and forth layers of liquid argon and lead absorbers. It covers a pseudorapidity range of $|\eta| < 3.2$ and consists of two main regions: the barrel calorimeter and the endcap calorimeters. The barrel calorimeter covers the central pseudorapidity range ($|\eta| < 1.475$), while the two endcap calorimeters extend the coverage to higher values of η . [17]

- **Barrel Calorimeter:**

- Located between $|\eta| = 0$ and $|\eta| = 1.475$.
- Consists of two half-barrels placed symmetrically around the beam axis.
- Thickness varies across different layers, providing optimal energy resolution for electron and photon interactions.

- **Endcap Calorimeters:**

- Two endcaps cover the region $1.375 < |\eta| < 3.2$, extending the coverage to more forward areas of the detector.
- Endcaps are also made of liquid argon and lead but thinner, to deal with the higher particle flux in the forward region.

Significance of ECAL in Electron Identification

ECAL in electron identification provides information on the energy and shower shape of the particle. Electrons create narrow electromagnetic showers in the ECAL. Hadrons produce wider

showers due to their interactions with both the ECAL and the hadronic calorimeter (HCAL). By comparing the shower shapes and energy distributions from ECAL and HCAL, electrons can be distinguished from hadrons and photons.

The lateral and longitudinal shower shapes are used as critical variables for electron identification. These variables allow the identification algorithm to efficiently separate electrons from the background.

Calibration and Corrections

The purpose of calibrating the ECAL is to guarantee precise energy measurements. Multiple corrections are implemented, including:

Energy loss corrections:

Compensating for loss of energy that occurs before to the particle's arrival in the calorimeter, specifically within the inner detector and solenoid.

Pile-up corrections:

Compensating for the extra energy deposits resulting from the overlapping of proton-proton collisions (pile-up) that happen at high luminosities.

Cluster energy corrections:

Adjusting the energy of the reconstructed clusters to compensate for energy that escapes outside the cluster borders and energy that is deposited in inactive areas of the detector.

These calibrations are necessary to ensure accurate measurements of electron energy, which are vital for high-precision analysis in physics, such as investigations on Z bosons and Higgs bosons [18, 19].

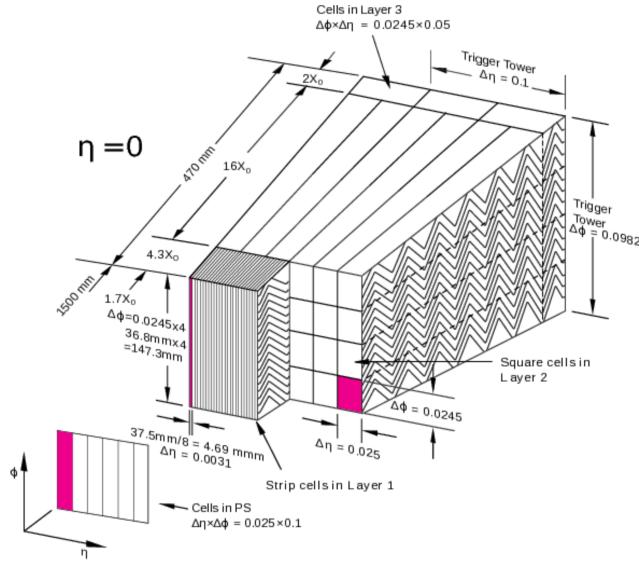


Figure 5: Schematic of the electromagnetic calorimeter, illustrating the accordion-shaped electrodes. The calorimeter is divided into three longitudinal layers, with the granularity of the cells in η and ϕ for each layer, as well as the dimensions of the trigger towers, indicated [20].

2 Reconstruction of Electrons

Reconstruction of Electrons is an important process that involves interpreting detector signals and heat signatures to identify electron with high efficiency and great accuracy. One of the major work of reconstruction is to study the decay process of the Z boson. This section discusses the techniques and methodologies employed in the reconstruction of electrons, with a focus on using information from the electromagnetic calorimeter and the inner detector.

Electron which are produced in the pp collision, such as from the Z boson decays interacts with the detector and this interaction results in the characteristic signatures and can be reconstructed.

The reconstruction process utilizes the following key components of the ATLAS detector:

- **Electromagnetic Calorimeter (EM Calorimeter):** In here measurement of energy of electrons and photons takes place. When electron enters the EM Calorimeter it produces an electromagnetic shower composed of secondary particles and photons. These shower interacts with calorimeter material. These energies get deposited in the specific cells, forming clusters of energy deposits that can be analysed.
- **Inner Detector (ID):** Inner detector tracks the charged particles which they move through the magnetic field and measures their momentum, path and charge.

These both are essential to reconstruct the electron.

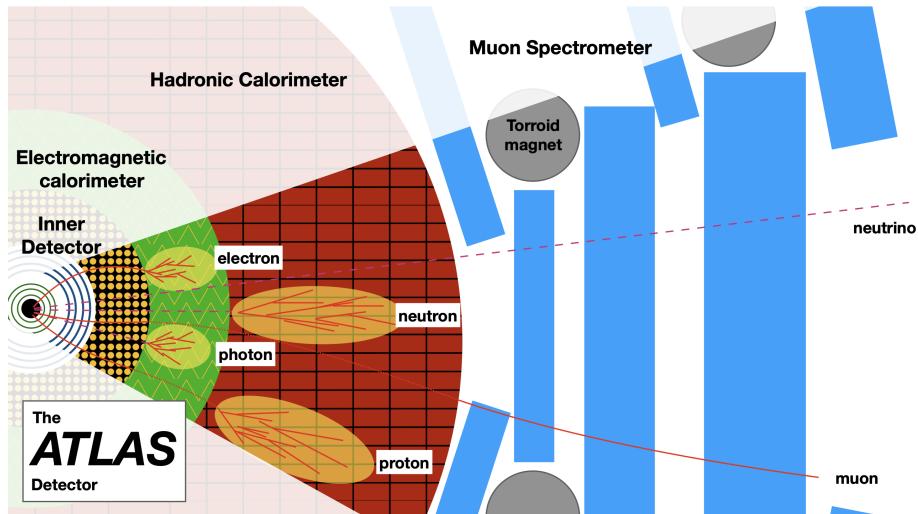


Figure 6: Particle identification in ATLAS. Here we can see the fundamental parts of the ATLAS detector and how they work to detect particles. A green circle represents the pixel detector and a blue circle the silicon strip detector; a black region with yellow dots represents the TRT; and these three colours together make up the inner detector. Green represents the electromagnetic calorimeter, whereas red represents the hadronic calorimeter. The blue boxes stand for the muon chambers. The red lines represent the particle tracks, while the yellow ellipses indicate the calorimeter showers. Purple dashed lines denote invisible particles.

2.1 Reconstruction Algorithm

Topo-cluster reconstruction algorithms [21–26] are implemented at ATLAS to recreate the electrons from the signals detected in EM and ID. To distinguish between electron and other particles, like hadrons and photons, a cut-based method is used.

- **Cluster Formation:** Clusters of energy deposits are formed in the EM Calorimeter, this is the initial step in electron reconstruction. The clustering algorithm uses a sliding window technique to scan for regions with significant energy deposits. When a cluster's transverse energy exceeds a threshold (typically around 2.5 GeV), it is flagged as a potential electron candidate. The cluster's size is usually fixed at 3×5 cells in the barrel region and 5×5 cells in the endcap region, reflecting the different geometries and particle densities in these detector areas. [27]
- **Track Reconstruction:** Along with constructing cluster, ID helps to reconstruct track of the electron. This involves fitting the signals in the tracking detectors to identify the particle's path through the magnetic field. When the Global χ^2 Track Fitter and the Gaussian Sum Filter method are used together, they correctly reconstruct tracks while taking into account any energy loss that might happen because of bremsstrahlung, which is when high-energy electrons send out photons and lose energy. [28]
- **Track-Cluster Matching:** When clusters and tracks are identified, the next step is to match them with each other. A successful match indicates a candidate electron. In order to complete the matching procedure, the reconstructed track is extrapolated to the position of the calorimeter, and then the location of the calorimeter is compared to the cluster. An asymmetric matching window is possible because the matching requirements take into account the bending of the electron's trajectory that occurs as a result of the magnetic field. For instance, the bending side must have a value that is less than 0.1, while the other side must have a value that is less than 0.05.

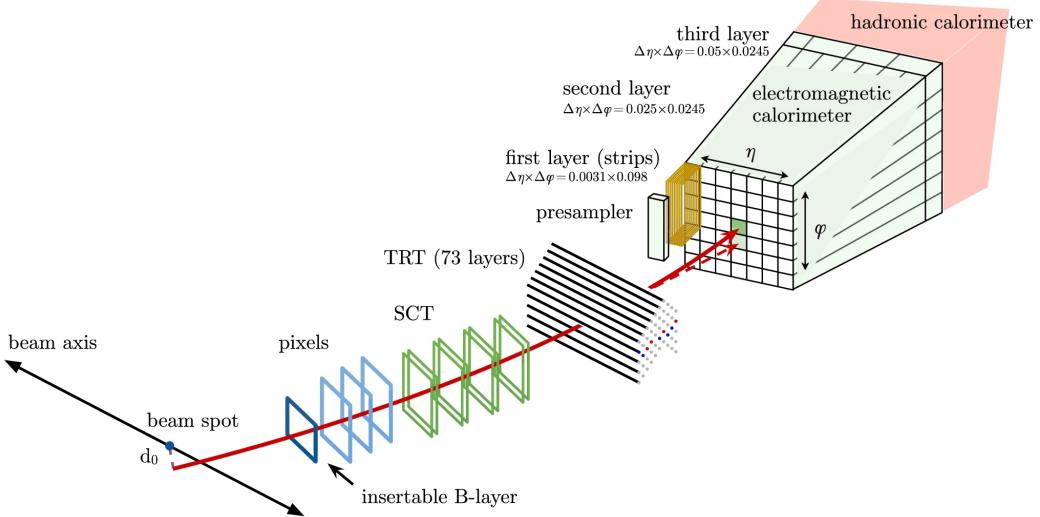


Figure 7: Clustering process around seed cells in the electromagnetic calorimeter.

- **Energy Calibration:** Energy loss before reaching the EM calorimeter, energy leakage outside the cluster boundaries, and energy deposited in regions beyond the EM calorimeter are taken into account after the Track and Cluster are matched.

2.2 Background Suppression

Electrons reconstructing from the detector's signals, from processes like Z boson decay may contain background. Background events can include jets or hadrons misidentified as electrons

from one way or another. So, background suppression is required and employed to improve the purity of the reconstructed electrons. These techniques include various identification criteria, energy deposit comparisons, and isolation cuts.

2.2.1 Isolation Cuts

Isolation cuts [29] are one way to get rid of background noise. Electrons made when Z bosons decay don't usually have many friends, so they are pretty much alone in the detector. On the other hand, background electrons typically have other particles surrounding them. Isolation cuts find out what amount of extra energy remains in a cone around the electron's trajectory. This energy of the cone is the sum of all transverse energy E_T within a cone radius ΔR around the electron's track. Mathematically,

$$E_T^{\text{iso}} = \sum_{\Delta R < 0.4} E_T,$$

where, $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$, the angular distance, defines the size of the cone around the electron. The parameter $\Delta\eta$ is the difference in pseudorapidity, and $\Delta\phi$ is the difference in azimuthal angle.

Electrons are considered isolated if the transverse energy within the cone, $E_T^{\text{iso}} < 5\%$ of the electron's own transverse momentum, p_T : $E_T^{\text{iso}} < 5\% \times p_T$.

3 Electron Identification

Researchers can accurately investigate the properties of electrons by using electron identification to separate them from other particles. Electron identification with great efficiency and purity is accomplished using various techniques.

Identification of electrons is based on a likelihood method that brings together data from several detector subsystems. The goal is to reduce background signals from hadrons or photon conversions while keeping real electrons with high efficiency. In order to identify a particle as an electron or background, this procedure uses shower shape characteristics from the electromagnetic calorimeter, track quality from the inner detector, and track cloud matching.

In this chapter, we also cover several fitting models, the tag-and-probe method, and how to improve electron identification accuracy with shower shape variables.

3.1 Likelihood Method for Electron Identification

The likelihood method [30] uses a number of observables to figure out how likely it is that a particle is either a real electron (signal) or a background particle. These are physical quantities that can be seen because of the detector's response. For example, shower shapes and track properties in the calorimeter are examples of energy deposition patterns.

3.1.1 Likelihood and Discriminant Calculation

Monte Carlo simulations are used to figure out the probability density functions (PDFs) for signal and background for each observable x , like shower shape, track-cluster matching, or track

quality. The probability of witnessing a specific value of x for an electron (signal) or background particle (such as a hadron or photon conversion) is represented by these PDFs $P_S(x)$ and $P_B(x)$. When we multiply the PDFs for n separate observables, we get the likelihood $L_S(x)$ for the signal and the likelihood $L_B(x)$ for the background:

$$L_S(x) = \prod_{i=1}^n P_S(x_i)$$

$$L_B(x) = \prod_{i=1}^n P_B(x_i)$$

where x_i represents the value of the $i - th$ observable for a particular electron candidate.

To distinguish signal from background, we calculate the discriminant:

$$d_L = \frac{L_S}{L_S + L_B}$$

The discriminant d_L takes values between 0 (background-like) and 1 (signal-like). However, the sharp peaks at 0 and 1 make it difficult to apply straightforward thresholds, so we apply the inverse sigmoid function to improve resolution:

$$d'_L = -\ln \left(\frac{1}{d_L} - 1 \right)$$

Each electron candidate is put into one of two groups based on its d'_L value: signal or background. By putting different limits on d'_L , the Loose, Medium, and Tight operating points can be found.

It is possible to set these operating points by giving the modified likelihood discriminant d'_L certain limits. When the value of d'_L is greater than the threshold, the electron candidates are called signals. When it is lower than the threshold, they are called background.

3.1.2 Loose, Medium, and Tight Criteria:

The major difference between loose, medium, and tight criteria [31] is in the levels of background rejection in comparison to signal efficiency while doing the electron identification. The loose criteria focus on identification efficiency and allow more background contamination. The medium and tight criteria are stricter on shower shape variables, track-cluster matching, and hadronic leakage. When these cuts are applied, signal electrons are identified with high purity, isolating signals from background processes.

Loose criteria focus on broader shower shapes and simpler track-matching conditions, making it more inclusive. On the other hand, tight criteria demand a narrower energy spread in the calorimeter and stricter requirements on track reconstruction and matching, leading to higher purity in electron identification.

3.2 Tag-and-Probe Method

The tag-and-probe method [32] is widely used in high-energy physics to evaluate and calibrate particle identification algorithms. It provides a straightforward way to measure the efficiency and performance of electron identification.

3.2.1 Electron Pairing and Event Selection

The initial step is to select events likely to contain electron pairs. These events are often the result of the decay of particles like the Z boson, which decays into an electron and a positron. Once such events are selected, electron candidates, known as "probe" electrons, are identified within these events.

3.2.2 Probe Electron Identification

Probe electrons are analysed to evaluate how well the identification algorithms perform. Key physical properties such as energy (E), pseudorapidity (η), and transverse momentum (p_T) are recorded. These properties are crucial for distinguishing electrons from other particles and understanding their behaviour in detector interactions.

3.2.3 Tag Electron Matching

To verify the probe electron's identification, each probe is paired with a "tag" electron. Tag electrons are identified with high confidence using stringent selection criteria. The assumption is that, in processes like Z boson decays, the electron and positron will exhibit correlated characteristics, which aids in accurate identification.

3.2.4 Invariant Mass Calculation

The invariant mass of the tag-probe pair is calculated to determine the parent particle from which the electron pair originated. For example, in Z boson decays, the invariant mass should be close to the known Z boson mass of approximately 91 GeV:

$$m = \sqrt{(E_{\text{probe}} + E_{\text{tag}})^2 - (\mathbf{p}_{\text{probe}} + \mathbf{p}_{\text{tag}})^2}$$

Firstly, the tag and probe is run on nominal data and is saved and then quality standards, such as the loose criteria is applied to probe electrons to ensure accurate identification. These standards help differentiate between true electron events and background noise. Nominal and loose differences are studied.

For each valid tag-and-probe pair, essential characteristics such as momentum, energy, and computed invariant mass are stored. The data is organized for efficient analysis, allowing future comparisons of different identification algorithms' performance and the development of new techniques.

3.3 Efficiency and Performance Evaluation

To evaluate the performance of the electron reconstruction and identification process, several metrics are used, including efficiency, purity, and fake rate.

3.3.1 Efficiency Calculation

The efficiency is defined as the fraction of probe electrons that are successfully reconstructed and identified:

$$\text{Efficiency} = \frac{N_{\text{reconstructed}}}{N_{\text{total}}}$$

where $N_{\text{reconstructed}}$ is the number of probe electrons that are reconstructed and pass the identification criteria and N_{total} is the total number of probe electrons selected for the analysis. Efficiencies are often measured in bins of transverse momentum and pseudorapidity to ensure uniform performance across different regions of the detector.

Using the RooFit package [33], likelihood fits are performed, and the sPlot technique is subsequently applied to extract signal and background contributions.

3.4 Purity and Fake Rate

Ratio of the electrons identified *correctly* to the electrons reconstructed by the algorithm is known as the **purity rate**. Higher purity rate means the identified electrons are more than the background.

$$\text{Purity} = \frac{N_{\text{true}}}{N_{\text{identified}}}$$

Where:

- N_{true} is the number of electrons *correctly* identified as genuine signal electrons.
- $N_{\text{identified}}$ is the total number of electrons identified by the algorithm, including both true electrons and fakes.

Ratio of the electrons identified *incorrectly* to the electrons reconstructed by the algorithm is known as the **fake rate**, higher fake rate means they are background more than the identified electrons.

$$\text{Fake Rate} = \frac{N_{\text{fake}}}{N_{\text{identified}}}$$

Where:

- N_{fake} is the number of electrons *incorrectly* identified as genuine signal electrons.
- $N_{\text{identified}}$ is the total number of electrons identified by the algorithm, including both true electrons and fakes.

ATLAS uses techniques like isolation cuts, background suppression, and performance indicators like efficiency, purity, and fake rate to make sure that the reconstructed electrons are as clean as they can be for analysis. This is an important part for studies that involve Z boson decays because it helps with precisely identifying electron pairs that are needed for accurate measurements of particle properties and interactions.

3.5 Z mass fitting

3.5.1 Breit-Wigner Fit

The Breit-Wigner function is commonly used to model the resonance of the Z boson:

$$f(m) = \frac{1}{(m^2 - M^2)^2 + M^2 \Gamma^2}$$

where M is the mass of the resonance, and Γ is the width. This model is effective in describing the peak structure of invariant mass distributions.

3.5.2 Polynomial Fit Model

Polynomial models provide flexibility in fitting the background shape. A third-degree polynomial, for example, can be written as:

$$f(m) = p_0 + p_1 m + p_2 m^2 + p_3 m^3$$

3.6 Shower Shape

Shower shapes are distributions of the energy deposited in the ECAL by the particles like electrons that goes through it. When electron enters the Calorimeter it produces a shower of secondary particles and photons. These secondary particles deposit their energy in the calorimeter cells, forming a characteristic pattern or “shape” that can be analysed.

Electrons make a shower that is small and focused. Hadrons and photons, make showers that are wider and less focused. Shower variables can tell the difference between electron-induced showers and other particles by looking at the shape of the energy deposits.

3.6.1 Shower Shape Variables

Shower shape variables are critical for differentiating electrons from other particles, such as hadrons. These include:

Name	Description	Type
R_{had}	Ratio of E_T in the first layer of the hadronic calorimeter to E_T of the EM cluster (used in the ranges $ \eta < 0.8$ or $ \eta > 1.37$)	Hadronic leakage
R_η	Ratio of the energy deposited in the middle layer of the EM calorimeter in 3×7 cells and the energy deposited in 7×7 cells centered at the electron cluster position	Lateral shower shape
$w_{\eta 2}$	Energy-weighted lateral width of the shower, $\sqrt{(\sum E_i \eta_i^2) / (\sum E_i) - ((\sum E_i \eta_i) / (\sum E_i))^2}$, where E_i is the energy and η_i is the pseudorapidity of cell i and the sum is calculated within a window of 3×7 cells	
E_{ratio}	Ratio of the energy difference between the largest and second largest energy deposits in the cluster over the sum of these energies	Profile of shower in strip layer of EM calorimeter
w_{stot}	Energy-weighted shower width, $\sqrt{(\sum E_i (i - i_{\max})^2) / (\sum E_i)}$, where i runs over all strips in a window of $\Delta\eta \times \Delta\phi \approx 0.0625 \times 0.2$, corresponding typically to 20 strips in η , and i_{\max} is the index of the highest-energy strip	

Table 2: Summary of selected shower shape variables and their application in electron identification. in the range $|\eta| < 2.47$. [34]

3.7 Methods to extract Shower Shapes

3.7.1 SPlot Technique

The SPlot technique [35] is a statistical method used to extract signal and background distributions from data. It provides weights for each event based on likelihood of being a signal or background, enabling the separation of signal and background components. These weights are used to enhance the purity of signals in distributions.

sPlot technique is applied after using tag-and-probe, in which an initial sample of signal and background events is selected. sPlot method refines this selection by assigning each event a signal or background weight, based on a multivariate likelihood fit of the data. This helps in isolating the pure signal component from the background contamination.

3.7.2 Principle of sPlot

The key idea behind the sPlot technique is to perform a fit on a variable that discriminates between signal and background, such as the invariant mass of a tag-and-probe pair. Both the signal and background probability density functions (PDFs) are estimated using a likelihood fit. After the fit is done, the likelihood of each event being in the signal or background class is indicated by its weight.

For an event i , the weight assigned is given by:

$$w_i = \frac{S_i}{S_i + B_i}$$

where S_i and B_i are the signal and background probabilities for the i th event.

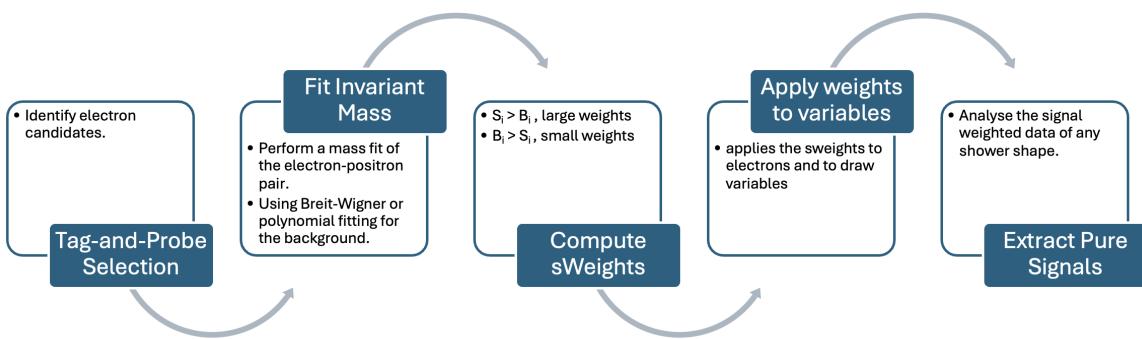


Figure 8: Identification algorithm

Fitting models are essential for analyzing the invariant mass distribution of tag-and-probe pairs and for distinguishing signal electrons from background. The choice of fitting model affects the precision of electron identification.

4 Results

In this section, we compare the experimental data with Monte Carlo (MC) simulations using two different frameworks: AF3 (ATLFAST3) and Geant4. These simulations are essential for understanding the performance of electron identification techniques, particularly in relation to shower

shape variables. We also calculate the chi-squared values to quantify the level of agreement between data and simulations.

This analysis is based on Rel 22, 2018 data.

4.1 Simulations: AF3 and Geant4

4.1.1 AF3 (ATLFAST3)

ATLFAST3 [36] is a fast simulation framework used in the ATLAS experiment. It reduces the computational time required for simulating particle collisions by employing parameterized models for certain detector responses, especially in the calorimeters. While AF3 provides a good approximation of detector behaviour, it may not reproduce fine details, particularly in the shower shape variables critical for electron identification. These approximations lead to faster simulations but with some loss in accuracy.

AF3 is especially useful for high-statistics studies where speed is more critical than precise modelling. However, when detailed shower shapes are required, as in electron identification, AF3 can show discrepancies when compared with experimental data, especially in the calorimeter.

4.1.2 Geant4

Geant4 [37], on the other hand, provides a full simulation of the ATLAS detector. It tracks every particle interaction using detailed physics models, making it highly accurate but computationally expensive. Geant4 simulates the entire detector system, including precise modelling of the electromagnetic showers produced by electrons. For this reason, Geant4 is often considered the gold standard for comparisons with experimental data.

Due to its accuracy, Geant4 is better suited for studies that require detailed shower shape analysis. However, its high computational cost limits the number of events that can be simulated, making it less feasible for very large datasets.

4.2 Chi-squared Calculation

To quantify the agreement between data and simulations, we calculated the Chi-squared (χ^2) value for each comparison. The χ^2 is calculated using the formula:

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{\sigma_i^2}$$

where O_i is the observed value from the data, E_i is the expected value from the simulation, and σ_i is the uncertainty in the measurement. The χ^2 values were normalized to account for the number of degrees of freedom. [38]

4.3 Shower Shape Variables

We analyzed key shower shape variables for the selected η and p_T bins, including:

- R_{had} - Hadronic leakage.
- $w_{\eta 2}$ - Lateral shower width.

- E_{ratio} - Energy ratio in the EM calorimeter.

The overall study was done on these η & p_T bins:

variable	bin edges
$E_T[\text{GeV}]$	15, 25, 30, 35, 40, 45, 50, 60, 90, 150
η	-2.47, -2.01, -1.52, -1.37, -0.8, -0.1, 0.0, 0.1, 0.8, 1.37, 1.52, 2.01, 2.47

4.4 Comparison of Fit Models for Background Estimation

In this analysis, two different fitting models were applied to estimate the background distribution: the Breit-Wigner (BW) fit and the polynomial (poly) fit. The primary goal of applying these fits was to compare how well each model approximates the background shape and subsequently assess their effectiveness by comparing chi-squared χ^2 values. Both the MC models and data were used in the zmass fitting without any cuts. This was done across three scenarios:

MC vs. BW Fit The first comparison evaluates the χ^2 value between the Monte Carlo (MC) simulated data and the BW fit. The BW function, which is typically used for resonant processes, was applied here as a candidate for modeling the background around the Z mass peak.

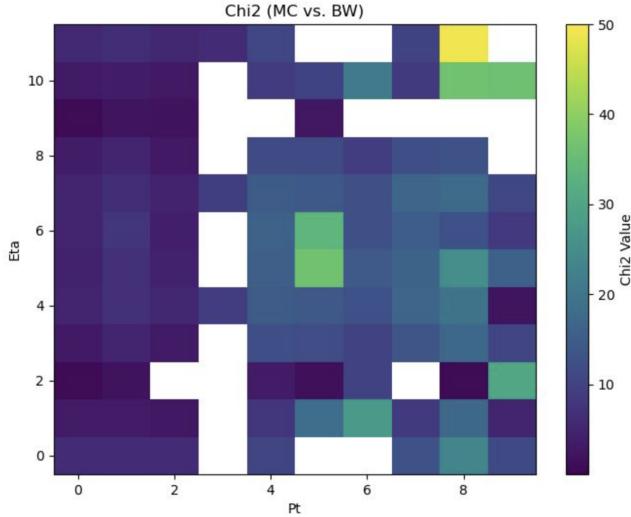


Figure 9: Background χ^2 for MC vs. BW fit. The blank area shows that the fit failed for those η and p_T bins.

MC vs. Poly Fit The second comparison assesses the χ^2 value between the MC simulated data and the polynomial fit. The polynomial fit, with a third-degree function, offers flexibility in capturing smooth variations in the background and was tested for its ability to match the MC distribution.

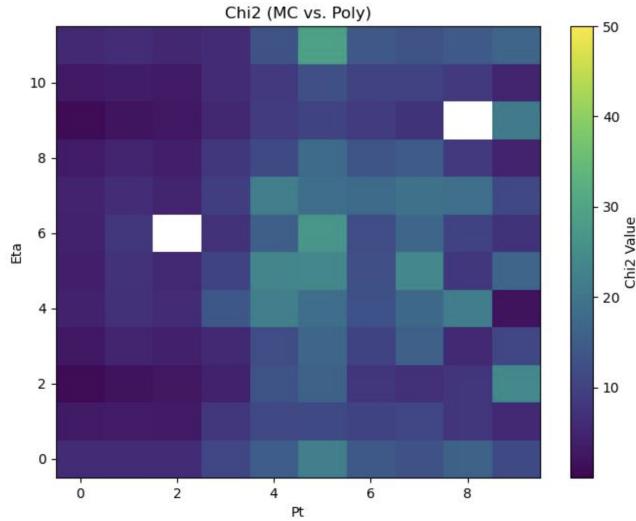


Figure 10: Background χ^2 for MC vs. Poly fit. The blank area shows that the fit failed for those η and p_T bins.

BW vs. Poly Fit Lastly, the χ^2 value was calculated between the BW fit and the polynomial fit themselves. This comparison allows us to directly observe the differences in how both functions model the background.

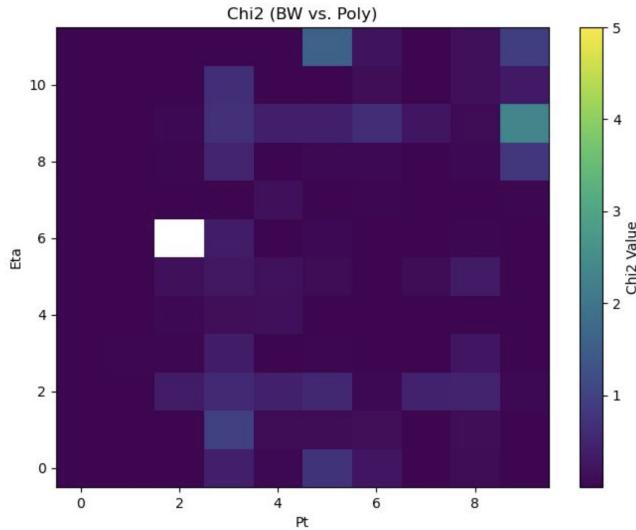


Figure 11: Background χ^2 for BW vs. Poly fit. The blank area shows that the fit failed for those η and p_T bins.

To visualize these comparisons, I will present the BW Z mass fit and poly Z mass fit. This will provide a clearer insight into how closely each fit follows the observed data and simulated background.

We focus on specific bins in pseudorapidity (η) and transverse momentum (p_T):

- Two η bins: one central bin at $\eta = 0.8$ and one forward bin at $\eta = 2.01$.
- Three p_T bins: two low p_T bin at 15 & 25 GeV and one high p_T bin at 45 GeV.

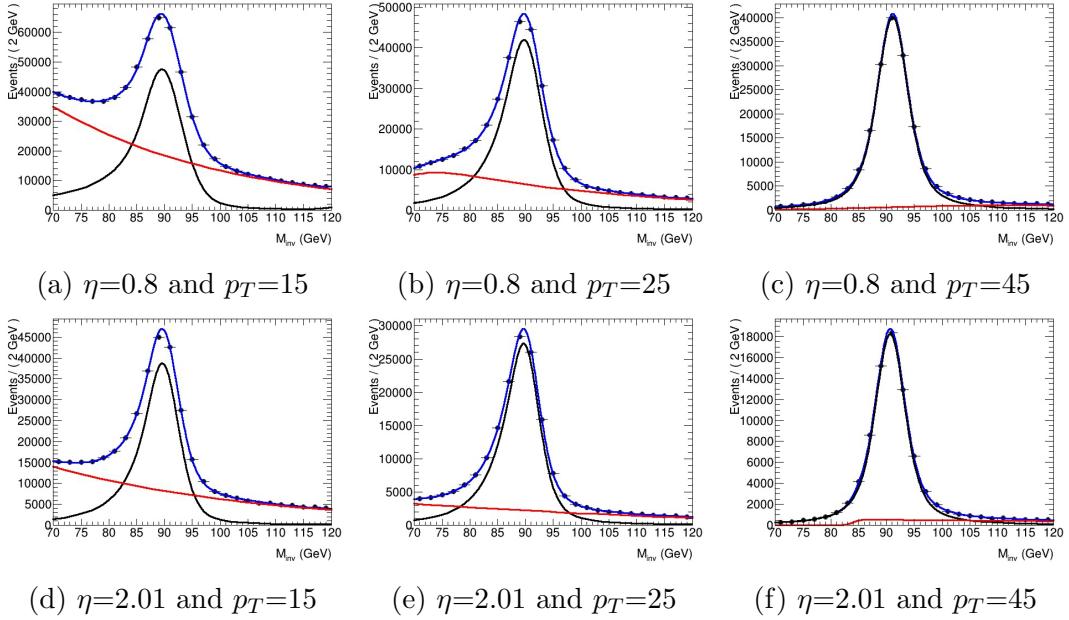


Figure 12: Background fit using BW Fitting

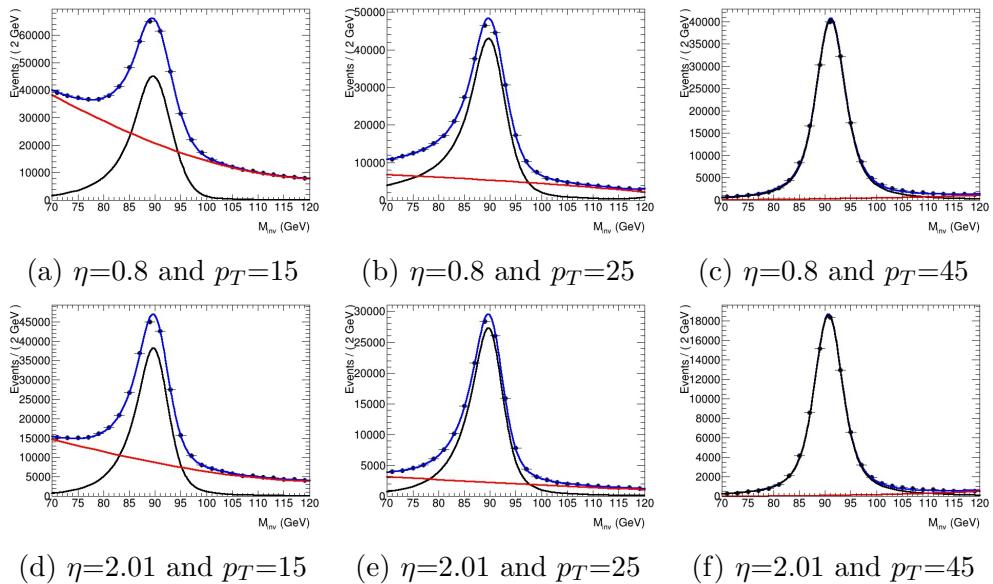


Figure 13: Background fit using Polynomial Fitting

These comparisons ultimately help determine that the polynomial model fits the background more accurately as Breit-Wigner failed at a lot of bins, guiding the choice of background estimation in the broader context of the Z boson mass study.

4.5 Shower Shape Comparisons of Fitting Models

Shower Shape Comparisons for BW vs Poly for AF3/Full Simulation with Nominal Cut/Loose Cut This study helps us know with what combination of fitting models, simulations and cut methods better shower shapes are attained. Figure 14, 15, 16, 17, 18 & 19 represent them and contain the χ^2 heatmaps for all the bins and shower shapes for w_η^2 for aforementioned η and p_T bins.

The low χ^2 value of all the variables for each cut method represents that it doesn't matter which fitting method is used to determine weights in the splot technique to attain shower shapes. The shower shapes remain the same for both, ie. polynomial and Breit-Wigner fittings.

4.6 Shower Shape Comparison of Loose Cut with Nominal

Shower Shape Comparisons for Loose Cut vs Nominal for AF3/Full Simulation with BW/Poly fit. As per the χ^2 values of Loose and Nominal for BW and Polyfit are shown in Figure 20, they match each other, meaning the shower shapes are independent of the fit models. But comparing the AF3 and Fullsim, AF3 has lower χ^2 values in comparison of the Fullsim.

Shower shapes and the ratios of the corresponding shower shapes are shown below each of them for the shower shape variable w_η^2 in the figures 21, 22, 23, 24, 25 & 26

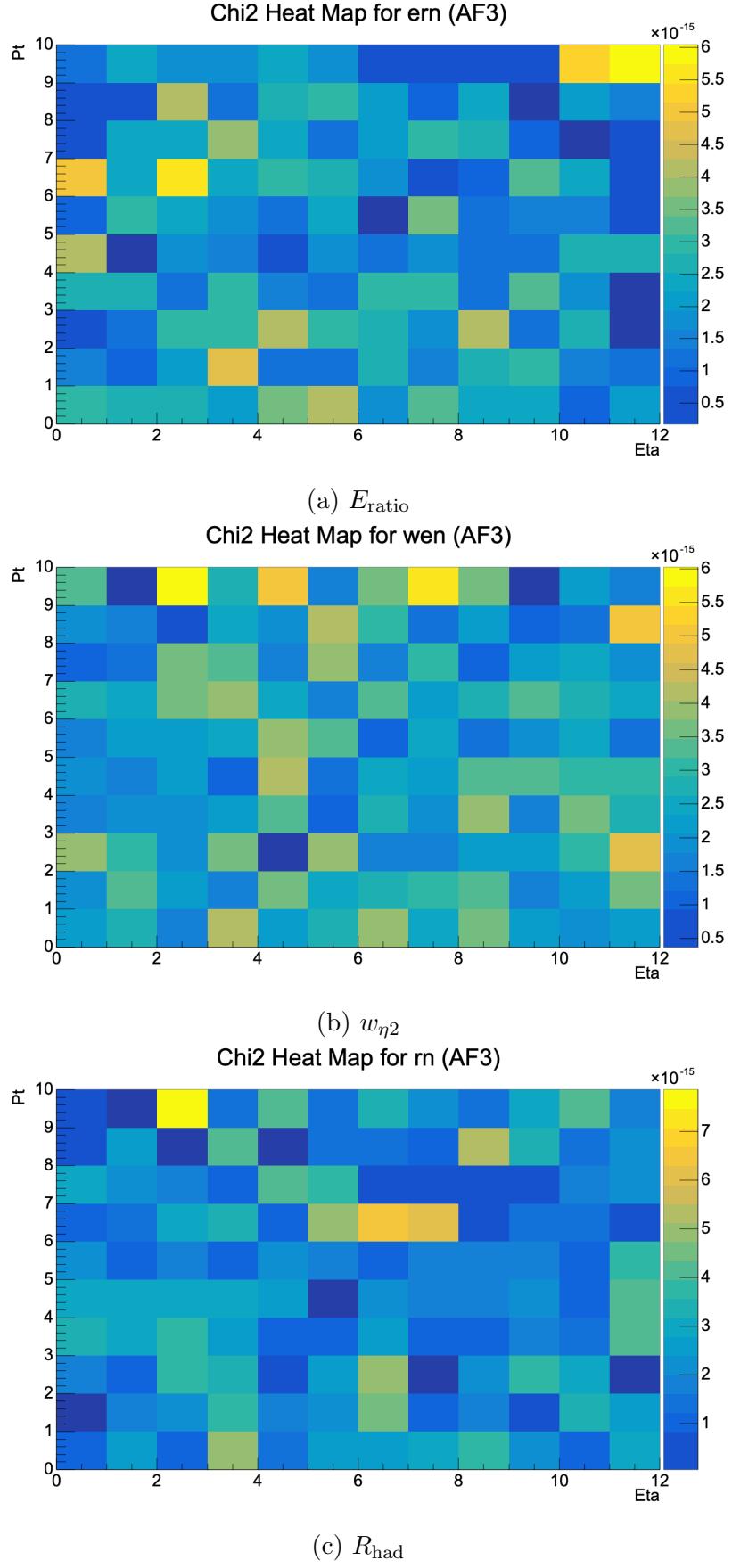


Figure 14: Shower Shape Comparisons for BW vs Poly for AF3 with Nominal Cut

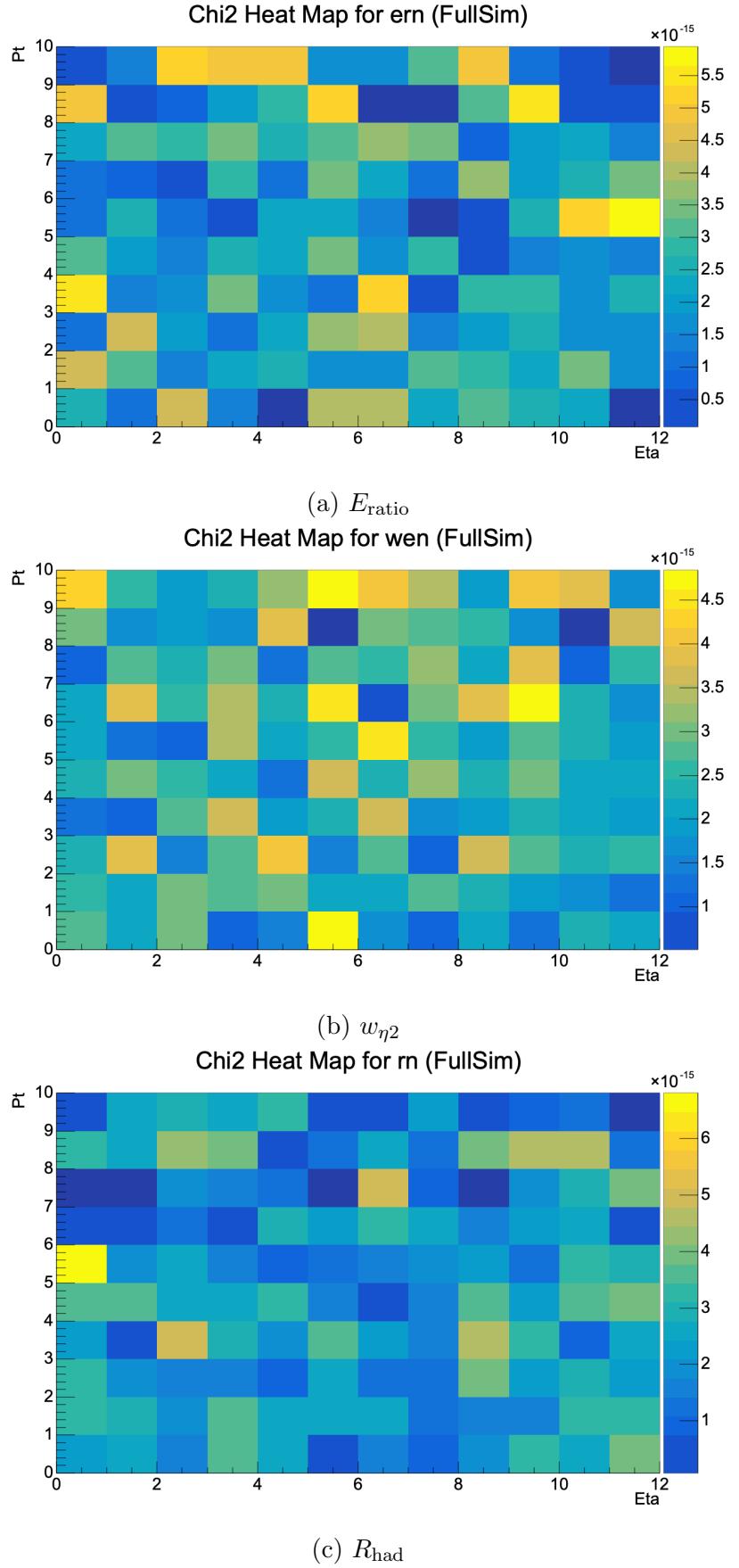


Figure 15: χ^2 of Shower Shape Comparisons for BW vs Poly for Full Simulation with Nominal Cut

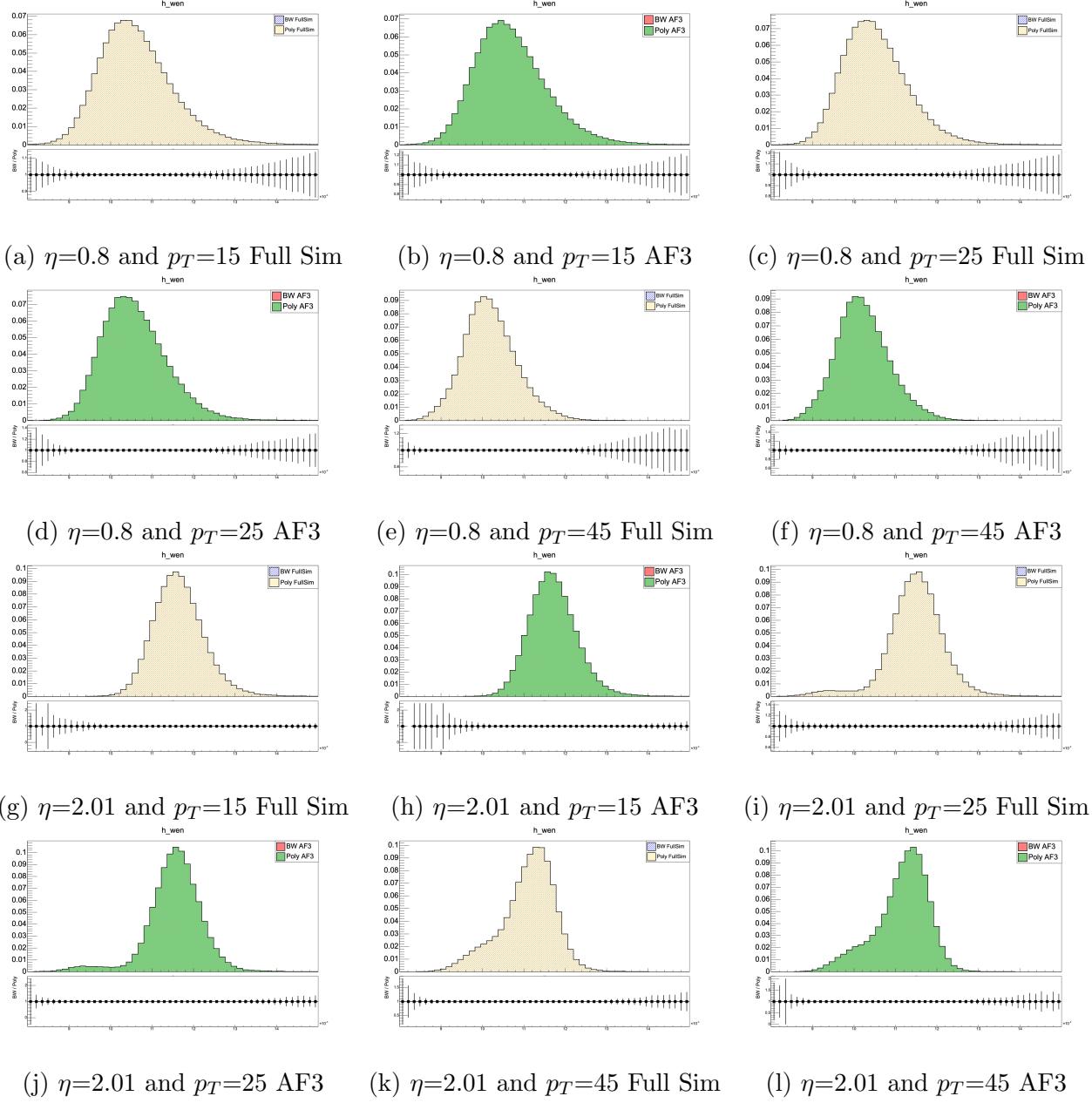


Figure 16: Shower Shape Comparisons for BW vs Poly for AF3 and Full Simulation with Nominal Cut for specific η and p_T for the variable $w_{\eta 2}$.

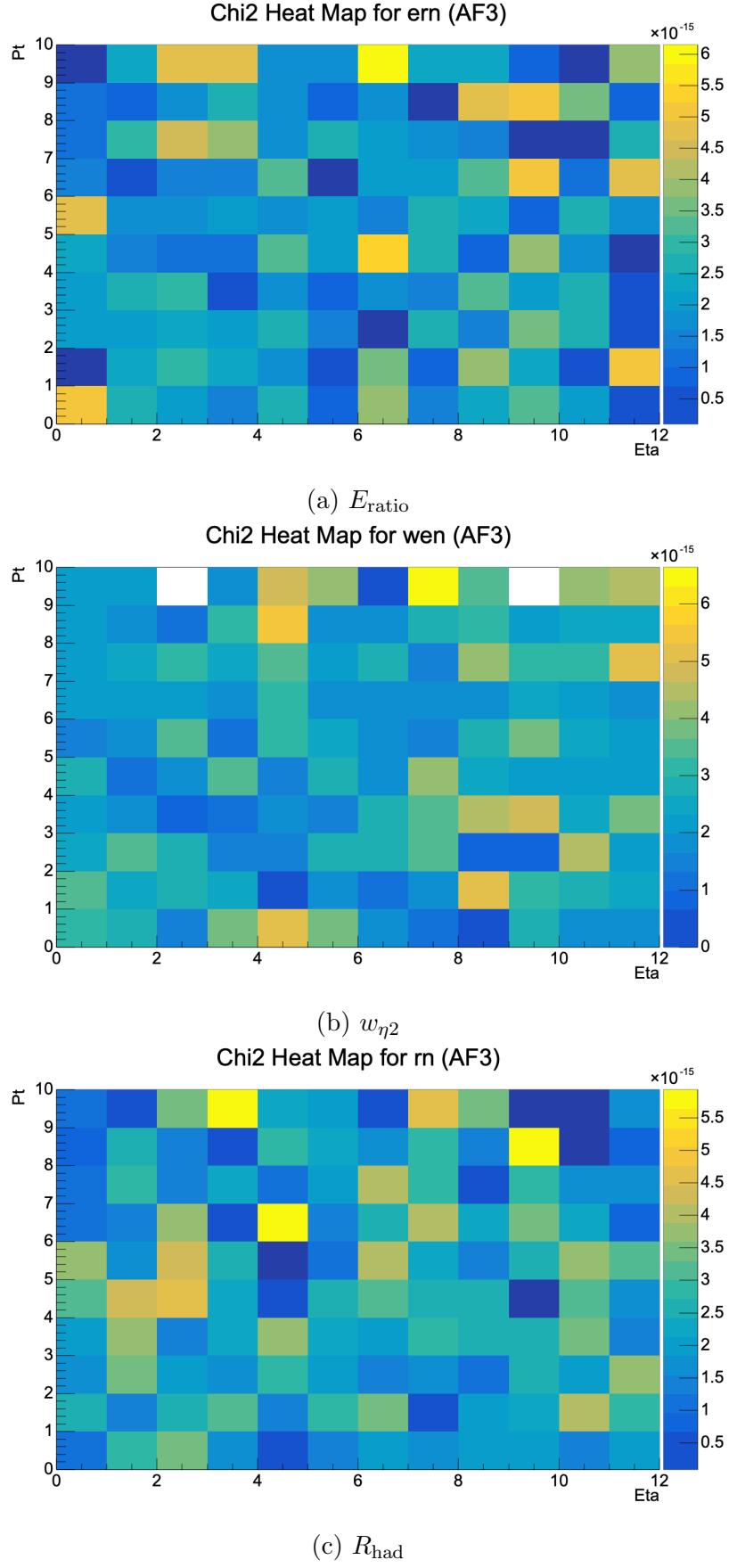


Figure 17: Shower Shape Comparisons for BW vs Poly for AF3 with Loose Cut

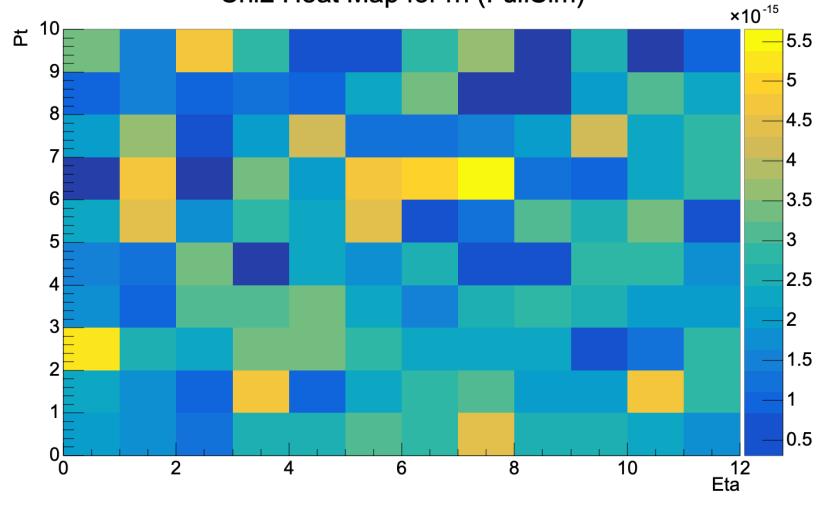
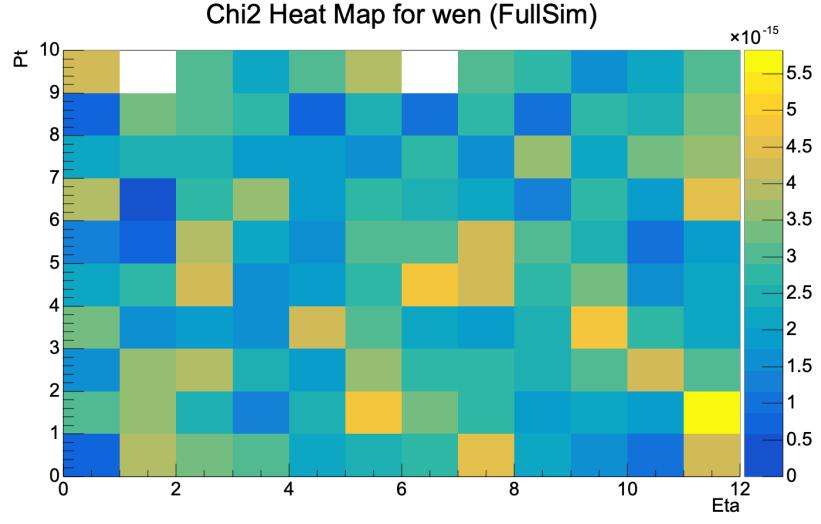
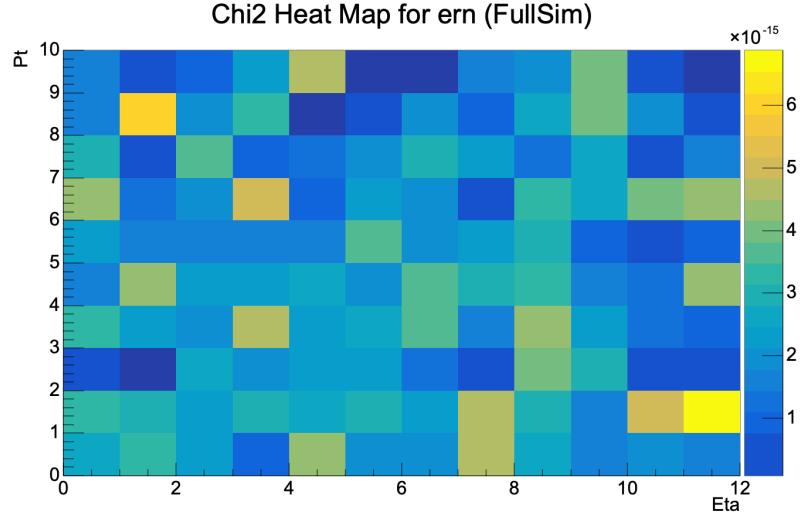


Figure 18: Shower Shape Comparisons for BW vs Poly for Full Simulation with Loose Cut

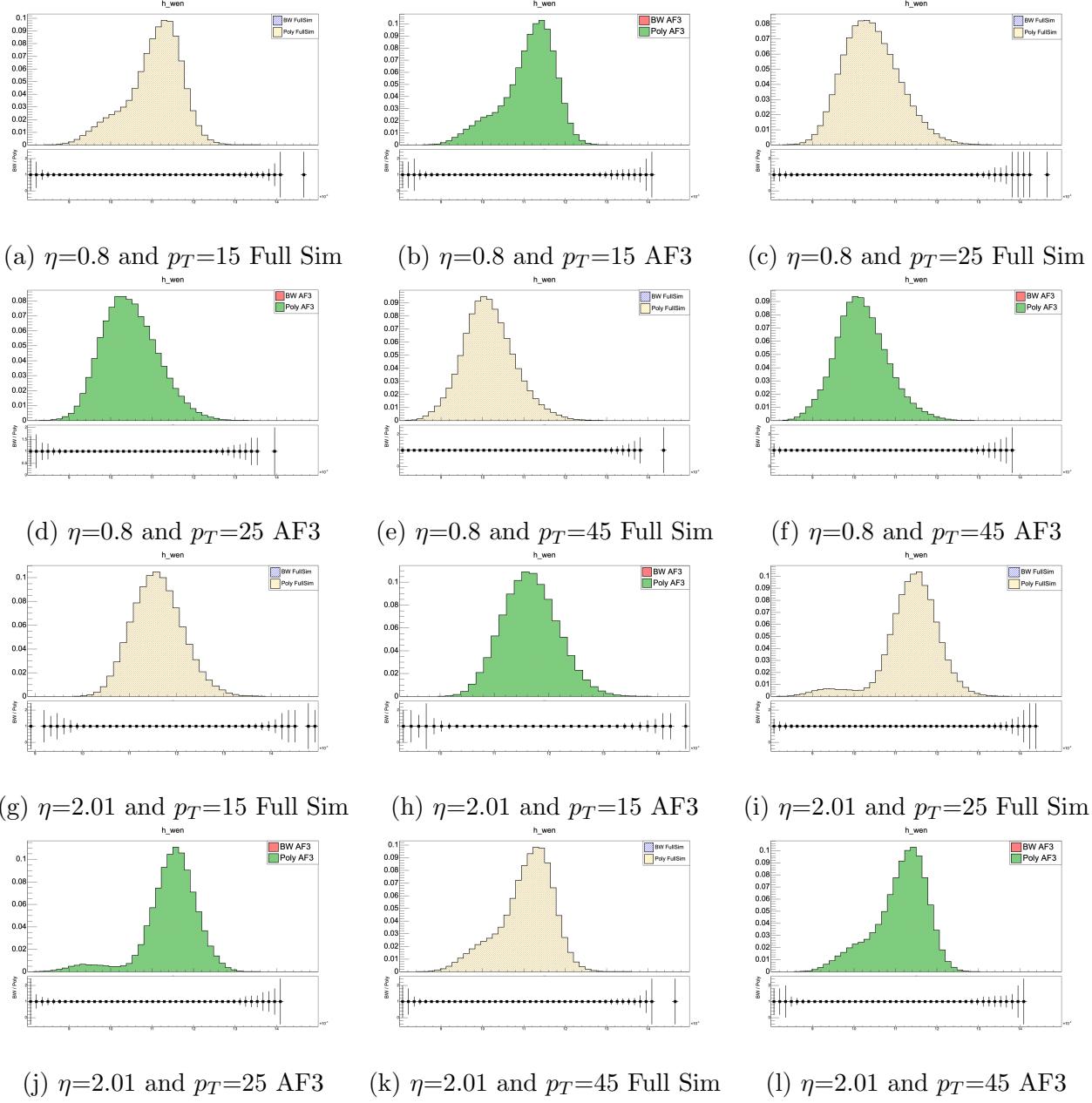


Figure 19: Shower Shape Comparisons for BW vs Poly for AF3 and Full Simulation with loose Cut for specific η and p_T for the variable $w_{\eta 2}$.

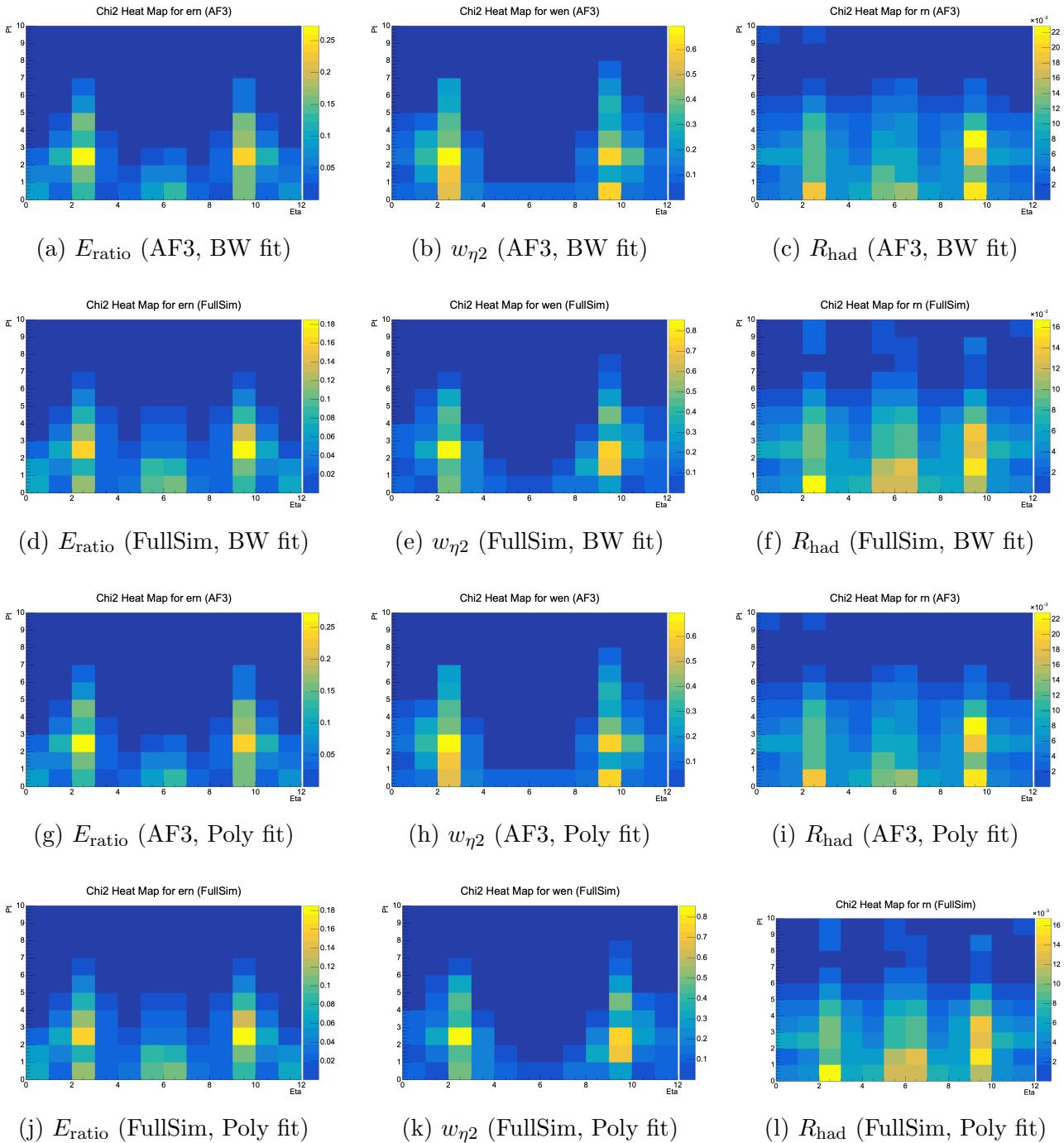


Figure 20: Shower Shape Comparisons for Loose Cut with Nominal for AF3 and Full Simulation with BW and Poly fits.

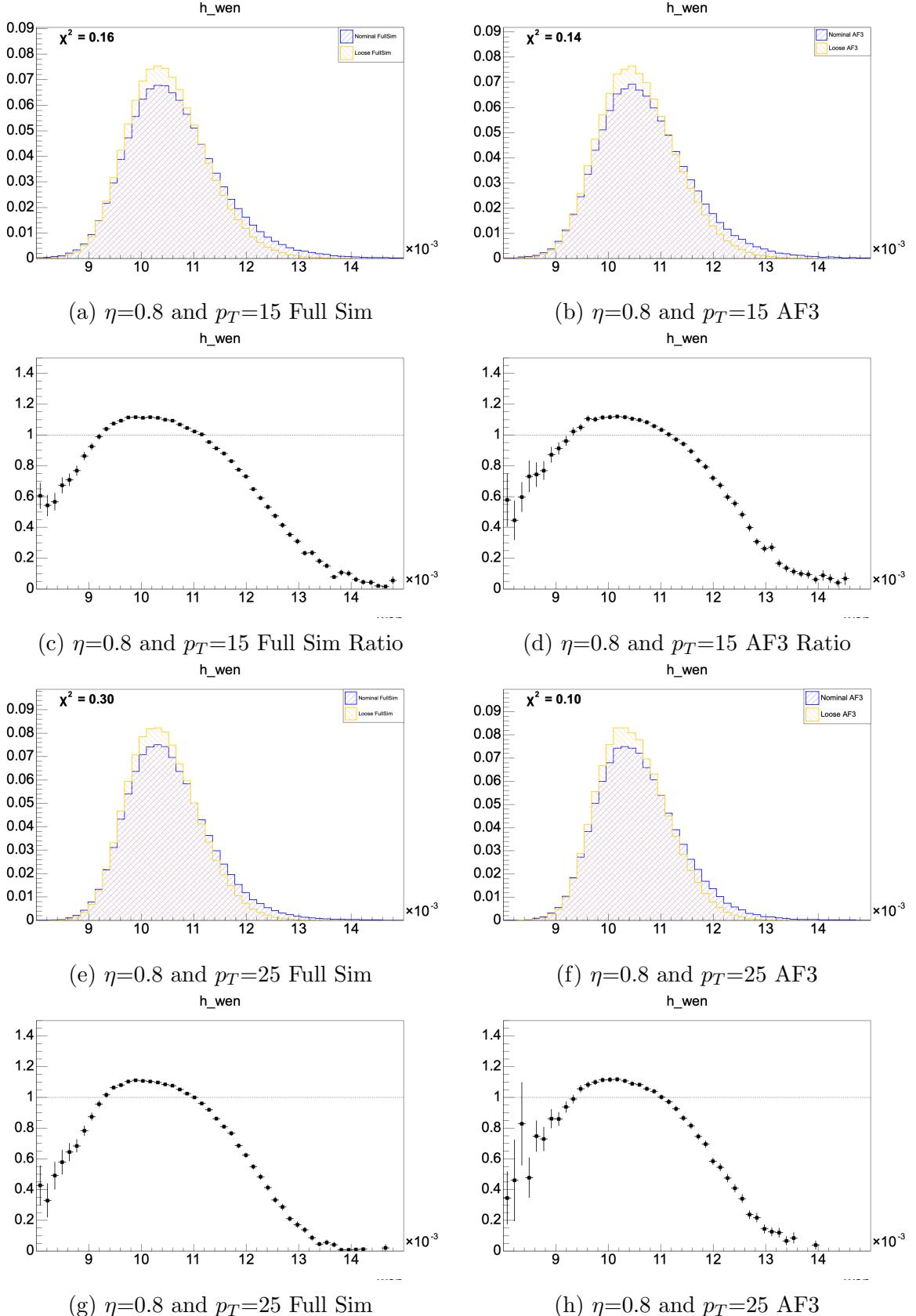


Figure 21: Shower Shape Comparisons for Nominal vs Loose Cut with bw for specific η and p_T for the variable $w_{\eta 2}$ and the Ratio plot of Loose / Nominal.

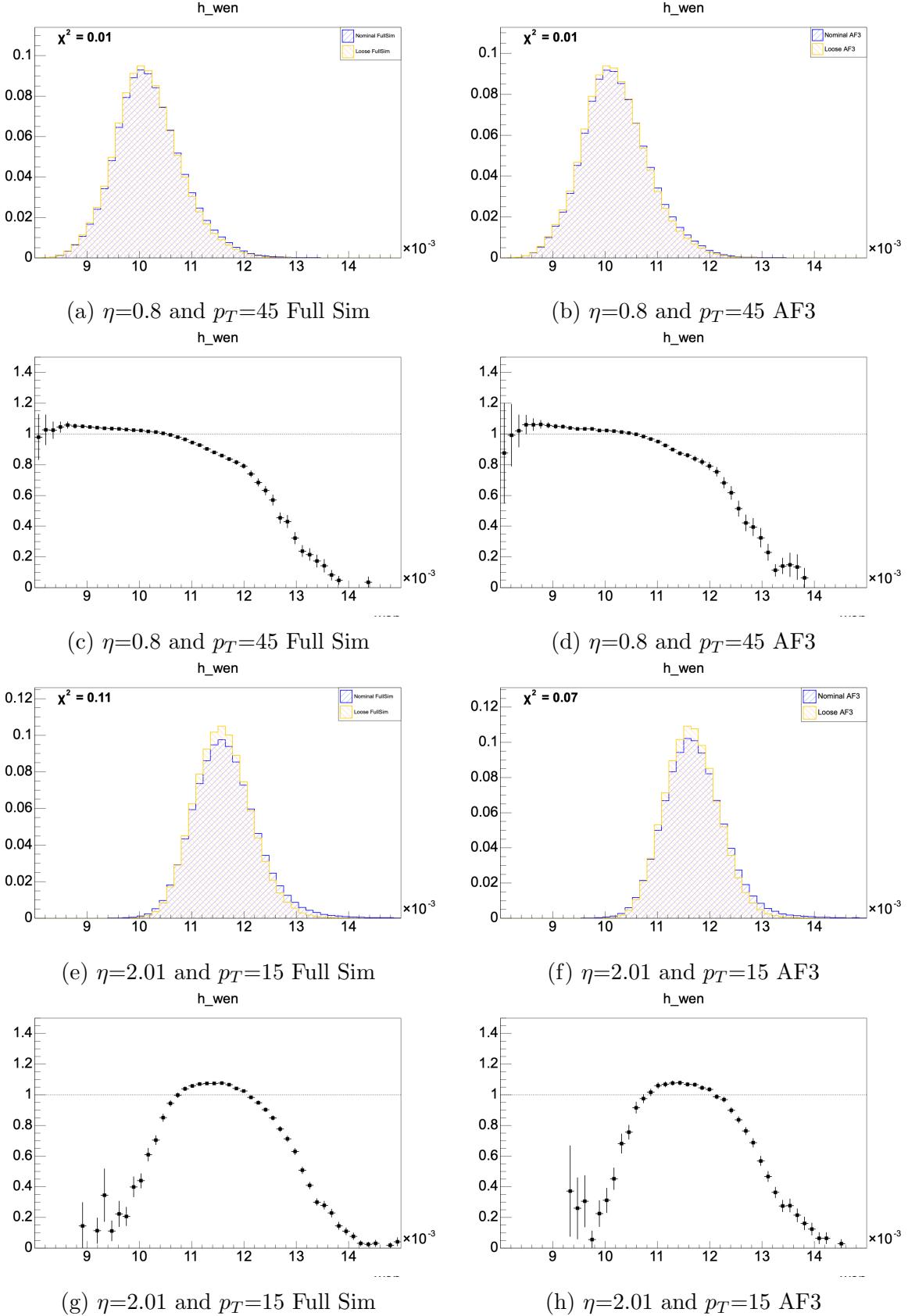


Figure 22: Shower Shape Comparisons for Nominal vs Loose Cut with bw for specific η and p_T for the variable $w_{\eta 2}$ and the Ratio plot is Loose / Nominal.

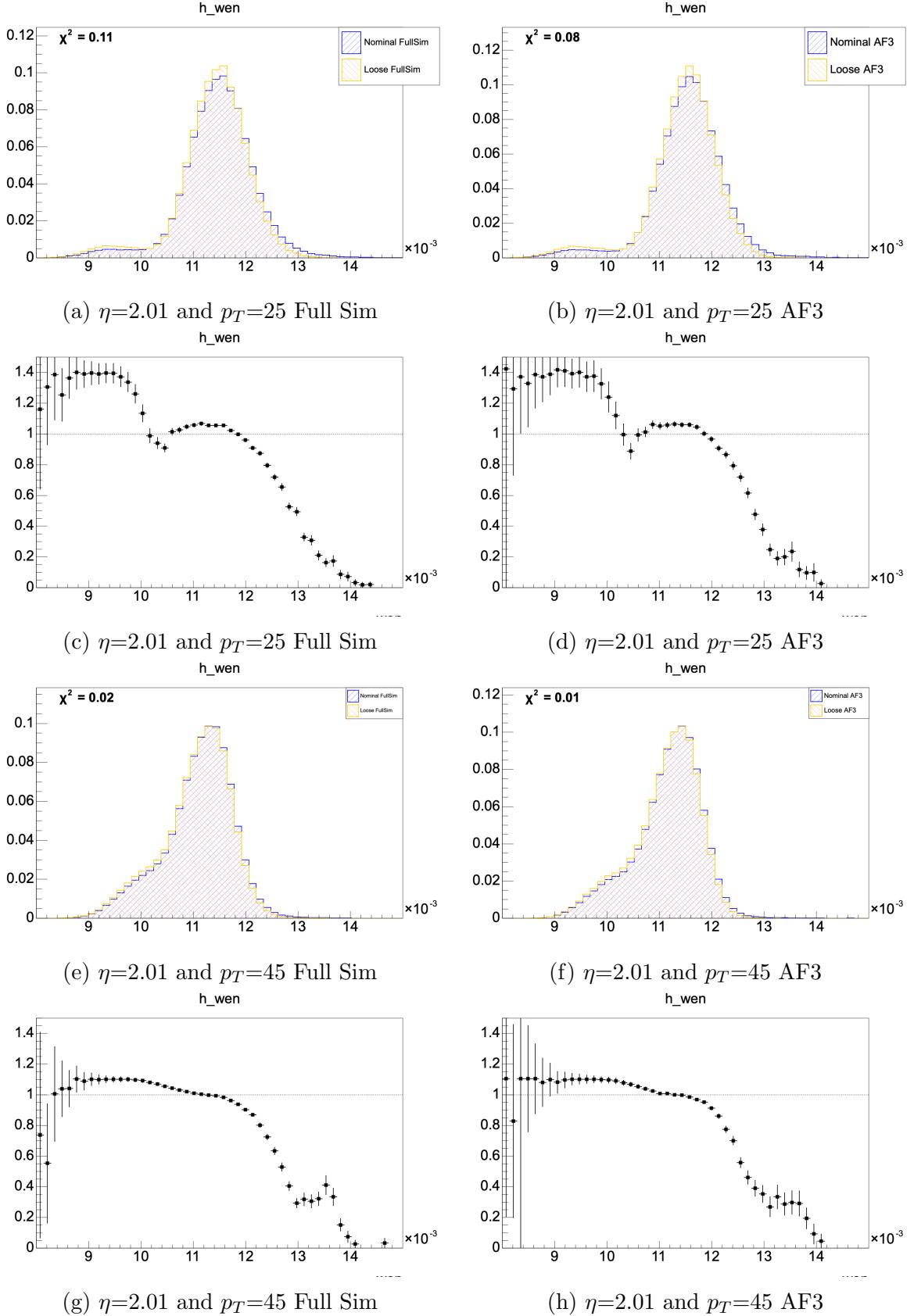


Figure 23: Shower Shape Comparisons for Nominal vs Loose Cut with bw for specific η and p_T for the variable $w_{\eta 2}$ the Ratio plot is Loose / Nominal.

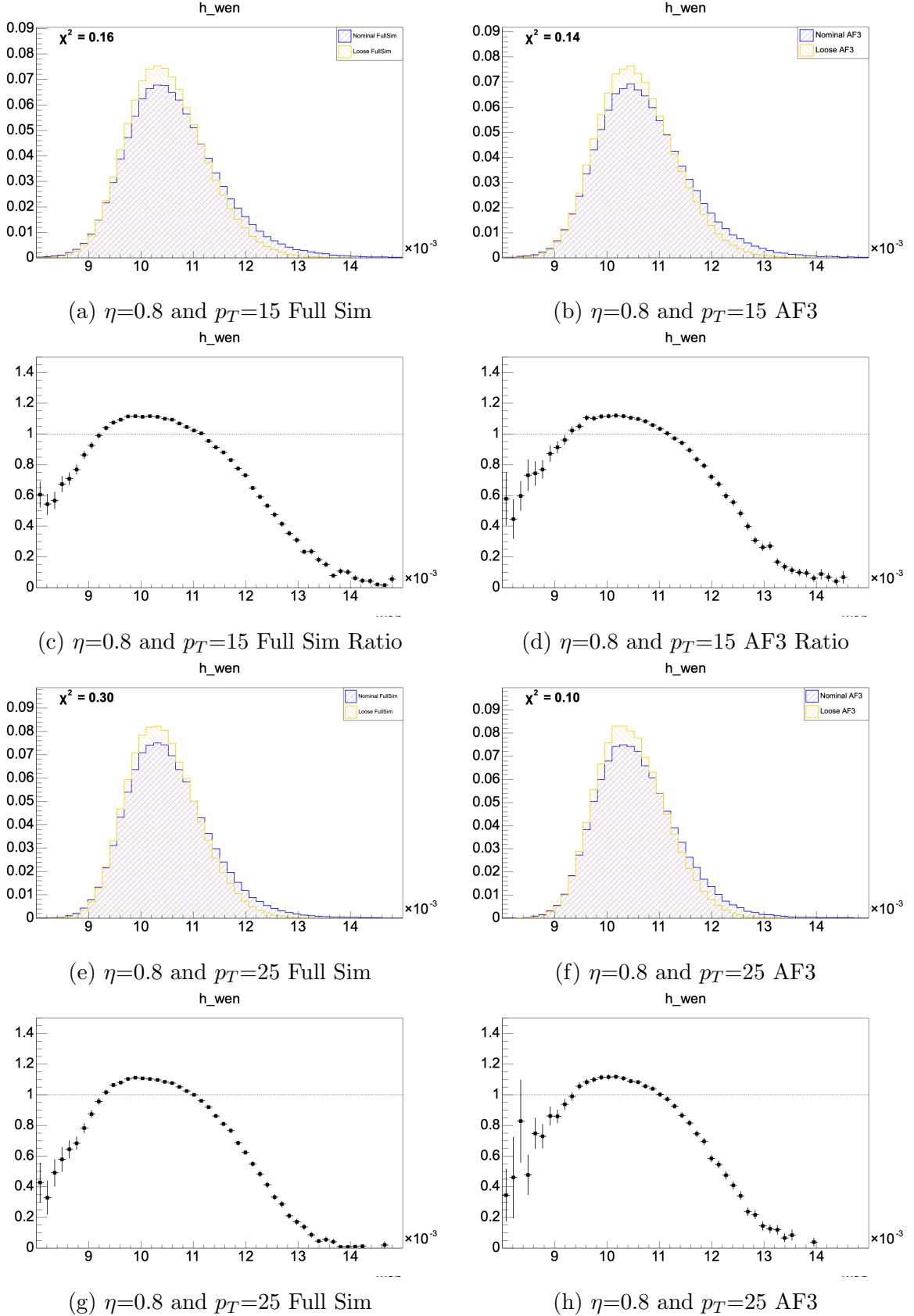


Figure 24: Shower Shape Comparisons for Nominal vs Loose Cut with Poly for specific η and p_T for the variable $w_{\eta 2}$ the Ratio plot is Loose / Nominal.

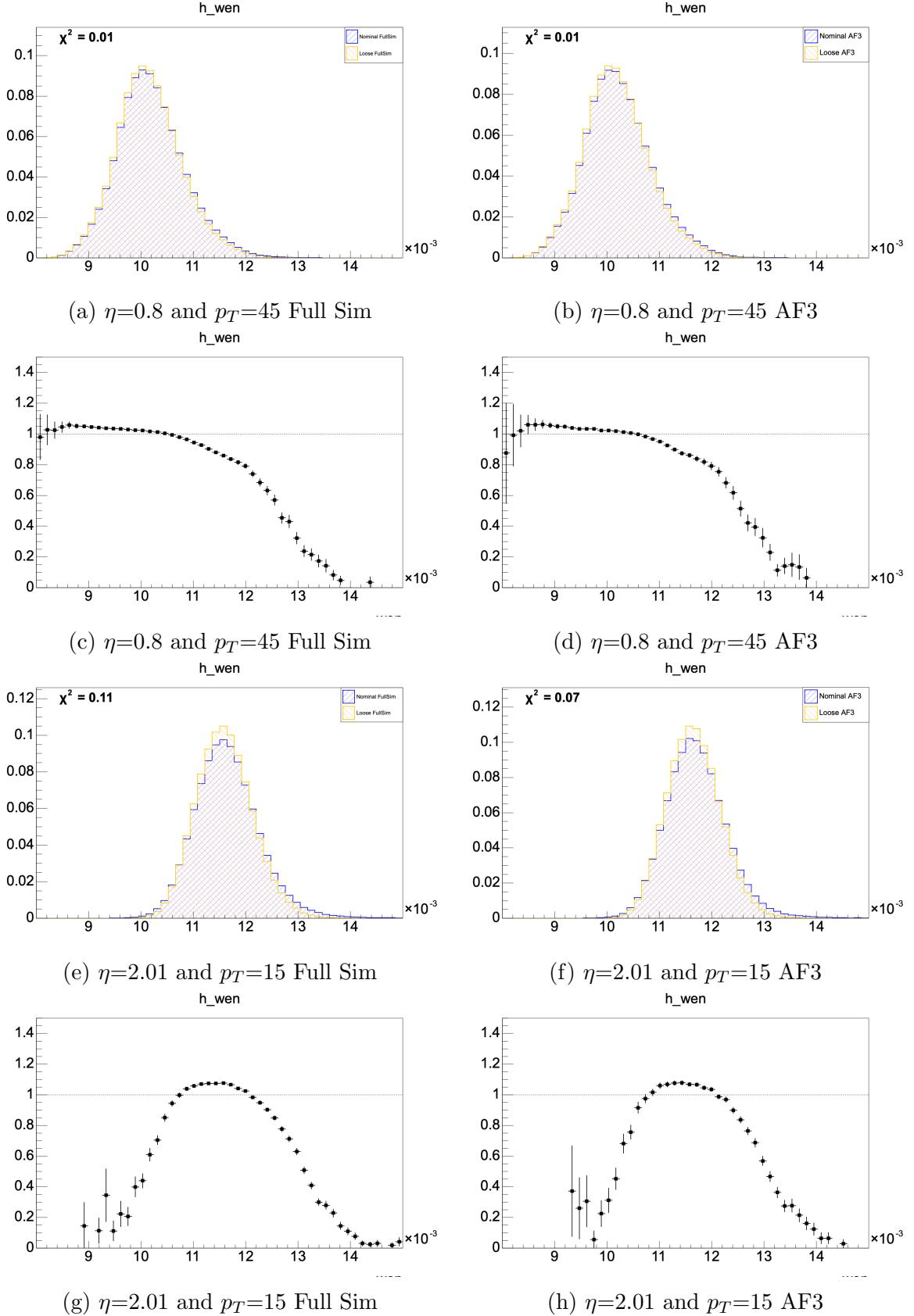


Figure 25: Shower Shape Comparisons for Nominal vs Loose Cut with Poly for specific η and p_T for the variable $w_{\eta 2}$ the Ratio plot is Loose / Nominal.

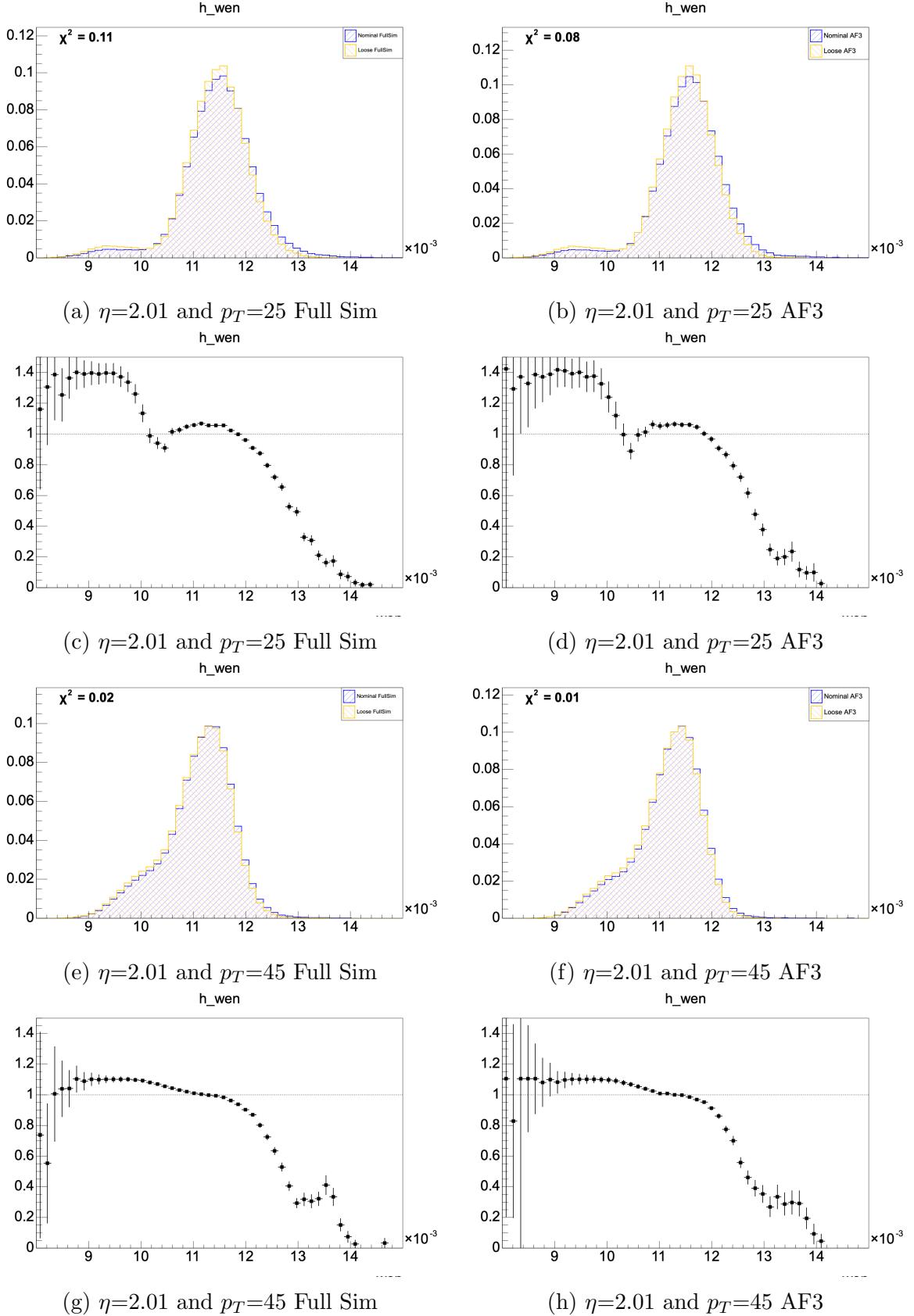


Figure 26: Shower Shape Comparisons for Nominal vs Loose Cut with Poly for specific η and p_T for the variable $w_{\eta 2}$ the Ratio plot is Loose / Nominal.

5 Optimal Transport

Optimal transport theory [39] provides a mathematical structure that can change one probability distribution into another by minimising a cost function that has already been set. To match the distributions from Monte Carlo (MC) simulations, like AF3 and Geant4, with the real data from the detector in high-energy physics studies, like those done at the ATLAS detector, this framework can be used. We want to make models more accurate by using optimal transport to improve correlations between simulations and observations.

Monte Carlo simulations are critical for modeling particle interactions, but they often exhibit differences when compared with real data. These discrepancies, particularly in shower shape variables, can reduce the reliability of the simulations in electron identification. The goal of optimal transport is to align the distributions of the simulated data with real data, improving the accuracy of the modeled variables such as shower shapes.

5.1 Mathematical Model

Theoretically optimal transport tries to find a transformation T that maps the simulated distribution P_{MC} to the real data distribution P_{data} , minimizing a predefined cost function. The problem is formulated as follows:

$$T = \arg \min_{\text{maps}} \sum c(x, T(x)),$$

where x represents the points in the simulated distribution, $T(x)$ is the corresponding point in the real distribution, and $c(x, T(x))$ represents the cost of moving x to $T(x)$. The squared Euclidean distance is often used as the cost function because it shows how big of a shift is needed to line up the distributions.

5.2 Attempted Implementation with PICNN

In my analysis, I attempted to apply the *Partial Input Convex Neural Networks (PICNN)* algorithm to perform optimal transport on shower shape variables from Monte Carlo simulations. PICNN is particularly useful because it models convex potential functions, which are used to compute the optimal transport map between distributions. The gradient $\nabla\phi(x)$ can be found using the convex potential function $\phi(x)$. This gradient is used to shift the Monte Carlo distributions so that they fit the data better.

It was planned to use this algorithm on shower shape distributions from the AF3 and Geant4 models and match them up with real ATLAS detector data. To get the PICNN model to agree with real data and simulations better, it would have been trained on shower shape factors like R_{had} , $w_{\eta 2}$, and E_{ratio} .

5.3 Challenges and Partial Results

While I began the implementation process of optimal transport using PICNN, time constraints prevented the completion of this task. This meant that the final comparison between Monte Carlo models and real data using optimal transport could not be made. Even though the execution wasn't finished, this first work set the stage for further investigation. Using optimal

transport is still a good way to improve the accuracy of electron recognition by lowering the differences between real and simulated data. As this method is improved, it might be possible to make shower shape modelling and electron detection much more accurate.

Even though the execution wasn't finished, this first work set the stage for further investigation. Using optimal transport is still a good way to improve the accuracy of electron recognition by lowering the differences between real and simulated data. As this method is improved, it might be possible to make big steps forward in modelling shower shapes and finding electrons.

6 Conclusion

In this dissertation, we have explored various methodologies for improving electron identification in proton-proton collisions using the ATLAS detector. Electron identification is crucial for high-precision studies, particularly those involving Z boson decays. We were able to measure how well electrons could be identified and get rid of background contamination in shower shape variables by using the tag-and-probe method and the sPlot statistical technique. We investigated potential solutions to the gaps between experimental and simulated results by comparing data with Monte Carlo simulations (AF3 and Geant4).

Model fitting for background estimation comprised an important section of the study. We compared the Breit-Wigner (BW) fit and the polynomial (poly) fit to assess their effectiveness in background modelling. Our chi-squared test showed that the polynomial fit was a better representation of the background shape across a number of pseudorapidity and transverse momentum bins. This was particularly evident in the shower shape comparisons, where the polynomial fit showed superior performance over the Breit-Wigner fit, especially when dealing with the AF3 simulation.

In addition, we performed shower shape comparisons between different fitting models and simulations, examining key variables such as R_{had} , w_{η^2} , E_{ratio} . The chi-squared test showed that the BW and polynomial fits gave the same results for these variables. This suggests that the choice of fit model does not really matter when modelling the shape of a shower. When we looked at AF3 and Full Simulation (Geant4) side by side, we saw that AF3 has lower chi-squared values, which means it fits the data better in some areas, but Geant4 is still the more accurate model overall.

While significant progress was made in modeling and fitting electron identification, one of the more innovative approaches we attempted was the application of optimal transport theory. We did not have enough time to fully implement the Partial Input Convex Neural Networks (PICNN) approach, but this method looks like it could be very useful for aligning Monte Carlo simulations with real data by reducing differences in shower shape distributions. Future work in this area could lead to further improvements in simulation accuracy and electron identification.

In conclusion, this dissertation demonstrates that, while conducting investigations in particle physics, a combination of sophisticated statistical approaches and meticulous model and simulation framework selection can lead to more precise electron identification. This research not only improves our knowledge of how electrons behave in high-energy collisions, but it also lays the groundwork for future work to improve particle identification methods and reduce the discrepancy between theoretical predictions and experimental results. The area of particle physics can acquire the precision necessary for the identification of new physics outside the Standard Model if it keeps investigating novel approaches like optimum transport.

References

- [1] Pich A 2007 The standard model of electroweak interactions (*Preprint* 0705.4264) URL <https://arxiv.org/abs/0705.4264>
- [2] Nakamura K and Group) P D 2010 *Journal of Physics G: Nuclear and Particle Physics* **37** 075021 URL <https://dx.doi.org/10.1088/0954-3899/37/7A/075021>
- [3] Dittmaier S and Schumacher M 2013 *Progress in Particle and Nuclear Physics* **70** 1–54 ISSN 0146-6410 URL <http://dx.doi.org/10.1016/j.ppnp.2013.02.001>
- [4] Peskin M E and Schroeder D V 1995 *Westview Press*
- [5] Halzen F and Martin A D 1984 *Wiley*
- [6] Glashow S L 1961 *Nuclear Physics* **22** 579–588 ISSN 0029-5582 URL <https://www.sciencedirect.com/science/article/pii/0029558261904692>
- [7] et al J B (Particle Data Group) 2012 *Phys. Rev. D* **86**(1) 010001 URL <https://link.aps.org/doi/10.1103/PhysRevD.86.010001>
- [8] Gross D J and Wilczek F 1973 *Phys. Rev. D* **8**(10) 3633–3652 URL <https://link.aps.org/doi/10.1103/PhysRevD.8.3633>
- [9] Fritzsch H, Gell-Mann M and Leutwyler H 1973 *Physics Letters B* **47** 365–368 ISSN 0370-2693 URL <https://www.sciencedirect.com/science/article/pii/0370269373906254>
- [10] David Politzer H 1974 *Physics Reports* **14** 129–180 ISSN 0370-1573 URL <https://www.sciencedirect.com/science/article/pii/0370157374900143>
- [11] Collaboration T A and et al G A 2008 *Journal of Instrumentation* **3** S08003 URL <https://dx.doi.org/10.1088/1748-0221/3/08/S08003>
- [12] Collaboration T A and et al G A 2010 *The European Physical Journal C* **70** 823–874 ISSN 1434-6052 URL <http://dx.doi.org/10.1140/epjc/s10052-010-1429-9>
- [13] Potamianos K 2016 The upgraded pixel detector and the commissioning of the inner detector tracking of the atlas experiment for run-2 at the large hadron collider (*Preprint* 1608.07850) URL <https://arxiv.org/abs/1608.07850>
- [14] Potamianos K 2016 The upgraded pixel detector and the commissioning of the inner detector tracking of the atlas experiment for run-2 at the large hadron collider (*Preprint* 1608.07850) URL <https://arxiv.org/abs/1608.07850>
- [15] ATLAS Collaboration M A e a 2017 *The European Physical Journal C* **77** ISSN 1434-6052 URL <http://dx.doi.org/10.1140/epjc/s10052-017-4887-5>
- [16] Mindur B 2017 *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **845** 257–261 ISSN 0168-9002 proceedings of the Vienna Conference on Instrumentation 2016 URL <https://www.sciencedirect.com/science/article/pii/S0168900216301905>
- [17] 1994 *ATLAS: technical proposal for a general-purpose pp experiment at the Large Hadron Collider at CERN* LHC technical proposal (Geneva: CERN) URL <https://cds.cern.ch/record/290968>

- [18] Collaboration T A 2009 Expected performance of the atlas experiment - detector, trigger and physics (*Preprint* 0901.0512) URL <https://arxiv.org/abs/0901.0512>
- [19] 1994 *ATLAS: Technical proposal for a general-purpose pp experiment at the Large Hadron Collider at CERN LHC* technical proposal (Geneva: CERN) URL <https://cds.cern.ch/record/290968>
- [20] ATLAS Collaboration M A e a 2016 *The European Physical Journal C* **76** ISSN 1434-6052 URL <http://dx.doi.org/10.1140/epjc/s10052-016-4507-9>
- [21] Collaboration A 2017 *The European Physical Journal C* **77** ISSN 1434-6052 URL <http://dx.doi.org/10.1140/epjc/s10052-017-5004-5>
- [22] Capeans M, Darbo G, Einsweiller K, Elsing M, Flick T, Garcia-Sciveres M, Gemme C, Pernegger H, Rohne O and Vuillermet R (ATLAS) 2010 ATLAS Insertable B-Layer Technical Design Report Tech. rep. URL <https://cds.cern.ch/record/1291633>
- [23] Sjöstrand T, Mrenna S and Skands P 2008 *Computer Physics Communications* **178** 852–867 ISSN 0010-4655 URL <http://dx.doi.org/10.1016/j.cpc.2008.01.036>
- [24] Alioli S, Nason P, Oleari C and Re E 2008 *Journal of High Energy Physics* **2008** 060 URL <https://dx.doi.org/10.1088/1126-6708/2008/07/060>
- [25] Davidson N, Przedzinski T and Was Z 2015 Photos interface in c++; technical and physics documentation (*Preprint* 1011.0937) URL <https://arxiv.org/abs/1011.0937>
- [26] Ball R D, Bertone V, Carrazza S, Deans C S, Del Debbio L, Forte S, Guffanti A, Hartland N P, Latorre J I, Rojo J and Ubiali M 2013 *Nuclear Physics B* **867** 244–289 ISSN 0550-3213 URL <https://www.sciencedirect.com/science/article/pii/S0550321312005500>
- [27] Lampl W, Laplace S, Lelas D, Loch P, Ma H, Menke S, Rajagopalan S, Rousseau D, Snyder S and Unal G 2008 Calorimeter Clustering Algorithms: Description and Performance Tech. rep. CERN Geneva all figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-LARG-PUB-2008-002> URL <https://cds.cern.ch/record/1099735>
- [28] Cornelissen T, Elsing M, Fleischmann S, Liebig W, Moyse E and Salzburger A 2007 Concepts, Design and Implementation of the ATLAS New Tracking (NEWT) Tech. rep. CERN Geneva all figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-SOFT-PUB-2007-007> URL <https://cds.cern.ch/record/1020106>
- [29] Hance M, Olivito D and Williams H 2011 Performance Studies for e/gamma Calorimeter Isolation Tech. rep. CERN Geneva URL <https://cds.cern.ch/record/1379530>
- [30] Eadie W, Drijard D, James F, Roos M and Sadoulet B 2013 *Journal of the American Statistical Association*
- [31] 2011 Expected electron performance in the ATLAS experiment Tech. rep. CERN Geneva all figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2011-006> URL <https://cds.cern.ch/record/1345327>
- [32] Collaboration T A 2011 Performance of the atlas trigger system in 2010 (*Preprint* 1110.1530) URL <https://arxiv.org/abs/1110.1530>

- [33] Verkerke W and Kirkby D 2003 The root toolkit for data modeling (*Preprint physics/0306116*) URL <https://arxiv.org/abs/physics/0306116>
- [34] The ATLAS Collaboration Aad G A B e a 2012 *Eur. Phys. J. C* **72**, 1909
- [35] Pivk M and Le Diberder F 2005 *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **555** 356–369 ISSN 0168-9002 URL <http://dx.doi.org/10.1016/j.nima.2005.08.106>
- [36] AATLAS Collaboration A c 2010 *The European Physical Journal C* **70** 823–874 ISSN 1434-6052 URL <http://dx.doi.org/10.1140/epjc/s10052-010-1429-9>
- [37] GEANT4 Collaboration S A e a 2003 *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506** 250–303 ISSN 0168-9002 URL <https://www.sciencedirect.com/science/article/pii/S0168900203013688>
- [38] Lyons L and Moneta L 2016 URL <https://cds.cern.ch/record/2235733>
- [39] Pollard C and Windischhofer P 2022 *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **1027** 166119 ISSN 0168-9002 URL <http://dx.doi.org/10.1016/j.nima.2021.166119>