A search for a Standard Model Higgs boson decaying to two τ leptons in the lepton-hadron final state

Kieran Matthew Bristow 309234

July 21, 2014

Supervisor: Dr. Trevor Vickey

A Dissertation Presented to

The Faculty of Science

University of the Witwatersrand

Johannesburg

In Fulfillment

of the Requirements of the Degree

Master of Science

Abstract

This thesis presents multiple studies contributing to the research conducted at the ATLAS experiment at the CERN facility in Switzerland. The areas of contribution include the ATLAS physics validation developments and the Standard Model Higgs boson decaying to two tau leptons ($H \rightarrow \tau^+\tau^-$) search effort. A section outlining the use and maintenance of the ATLAS TauValidation package details contributions made towards the ATLAS physics validation program. Studies investigating the comparison of Pythia and Herwig Monte Carlo generators facilitated in a move from Herwig to Pythia based Monte Carlo samples in the $H \rightarrow \tau^+\tau^-$ search. A trigger treatment investigation exposes biases introduced when using the embedding trigger treatment within the $H \rightarrow \tau^+\tau^-$ in the lepton-hadron final state analysis.

Acknowledgements

A number of thank yous need to be made to those who have greatly aided me throughout my thesis work. Firstly a huge thank you to Dr Trevor Vickey, my supervisor, for without his help my CERN visit along with the amazing introduction to the world of experimental particle physics would never have been possible. The constant advice and guidance throughout the period of my studies has been invaluable.

Another huge thank you goes out to Dr. Katharine Leney who was my biggest source of hands-on help through out the entire period of my studies. Her experience and willingness to help out at almost any time was at many times a life-saver.

Finally, a thank you goes to my brother, Timothy Bristow. You helped keep me motivated throughout this endeavour and the constant pieces of advice you gave were always greatly appreciated.

Contents

1 Introduction

The Standard Model of particle physics is our most accurate theory to date for describing the universe's elementary particles and how they interact with each other. It has seen huge success at predicting and describing many experimental results and has therefore been an important area of research for particle physicists. For a long time however, there was an important missing link that was essential to the theory but had never been observed. This was the particle essential for the mechanisms behind electroweak symmetry breaking and giving elementary particles mass. This mechanism is known as the BEH (Brout–Englert–Higgs) mechanism or just the Higgs mechanism [1, 2, 3] and the particle is the Higgs boson [1, 4].

For a long time many different experiments around the world were searching for the Higgs boson. Finally, on the 4th of July 2012 an announcement was made at CERN that a new boson had been observed by both the ATLAS and CMS experiments. The boson closely matched the expected properties for a Standard Model Higgs boson at $\approx$ 125 GeV[1] [5, 6]. However, the book is not yet closed. There is still work to be done in confirming if it is indeed the Higgs boson predicted by the Standard Model and not some other theory. Part of the continued work involves studying the properties of this boson. There are many predicted decay modes of the Higgs for all models but even for the highly likely candidate model for the observed boson, the Standard Model, only the $\gamma\gamma$, ZZ and WW final states of a Higgs decay were observed to date. The other final states such as tau tau ($\tau\tau$) [7, 8] and bb [9, 10] do have results that suggest the existence of a Higgs in these channels but more research is required to make this claim with certainty. In addition to property measurements for the newly observed particle, the search to observe these other predicted decay modes of the Higgs still continues.

Of particular interest to this thesis is the unobserved channel for the Standard Model (SM) Higgs decaying to two oppositely charged τ leptons. A branching ratio of 6.3%[11] for a Higgs with m_H = 125 GeV means it is one of the leading decay modes that could help measure the coupling of the Higgs to fermions. The $H \to \tau^+\tau^-$ channel consists of three sub-channels depending on whether the taus decay leptonically or hadronically:

[1] GeV stands for Giga electron Volt where 1 eV = $1.60217657 \times 10^{-10}$ joules. GeV actually represents GeV/c^2 but using natural units, c is set to 1.

- lepton-lepton ($H \to \tau_{lep}\tau_{lep}$)
- lepton-hadron ($H \to \tau_{lep}\tau_{had}$)
- hadron-hadron ($H \to \tau_{had}\tau_{had}$).

These final states describe how each of the two τ leptons decay after their production where 'lepton' describes a tau decaying to an electron(e) or a muon (μ) and 'hadron' describes a tau that decays to hadrons (τ_h) where hadrons are particles comprised of quarks. Currently in this search an excess of 4.1 standard deviations above the expected background has been observed for a Standard Model Higgs of $m_H \approx 125$ GeV at the ATLAS experiment [12]. A standard deviation of 5 or higher is required in the world of particle physics to claim discovery of the particle. An observed excess of 4.1 standard deviations means there is strong evidence for the existence a Higgs boson with this decay mode but more data is required to conclude the search.

This thesis presents research done as part of the ATLAS search for $H \to \tau_{lep}\tau_{had}$. The analysis uses data collected from proton-proton collisions by the ATLAS experiment at the LHC (Large Hadron Collider) and corresponding to integrated luminosities of 4.6 fb^{-1} and 13.0 fb^{-1} at center of mass energies of $\sqrt{s} = 7$ TeV and 8 TeV respectively. Here the center of mass energy is the sum of the energy of the two colliding beams. At the LHC the $H \to \tau_{lep}\tau_{had}$ channel which contributes to 46% of the entire branching ratio for $H \to \tau^+\tau^-$ is being studied across various production processes. These production processes include gluon-gluon fusion $gg \to H$, $W^\pm$ or Z vector-boson fusion $qq \to qqH$(denoted VBF) and Higgs-strahlung $qq \to VH$ (denoted VH) in association with a hadronically decaying vector boson ($V = W^\pm/Z$) [13]. The feynman diagrams depicting these processes can be seen in Figure 1. For $m_H = 125$ GeV the gluon-gluon fusion production process has the largest production cross-section for SM Higgs boson production, of 15.3 pb and 19.5 pb at $\sqrt{s} = 7$ TeV and 8 TeV respectively [11]. VBF has the second largest production cross-section of 1.22 pb (7 TeV) and 1.57 pb (8 TeV) and finally VH has cross-sections 0.5729 pb (7 TeV) and 0.7046 pb (8 TeV) for WH and 0.3158 pb (7 TeV) and 0.4153 pb (8 TeV) for ZH [11]. The standard signal topology consists of an electron or muon, a low track multiplicity jet resulting from the τ_h, and missing transverse energy resulting from the three neutrinos which accompany the τ decays. Here jets are cones of hadrons that form when high energy quarks and gluons undergo a process called hadronization. Each of the production processes described above result in additional jets in the final state which give distinct experimental signatures. By requiring additional jets with high transverse momentum (p_T) in the event the boost of the Higgs boson in the transverse plane can be exploited. These boosted topologies frequently result in the decay products of the τ having larger p_T which helps separate signal from background processes and helps in the measurement of the Higgs resonance signal [14].

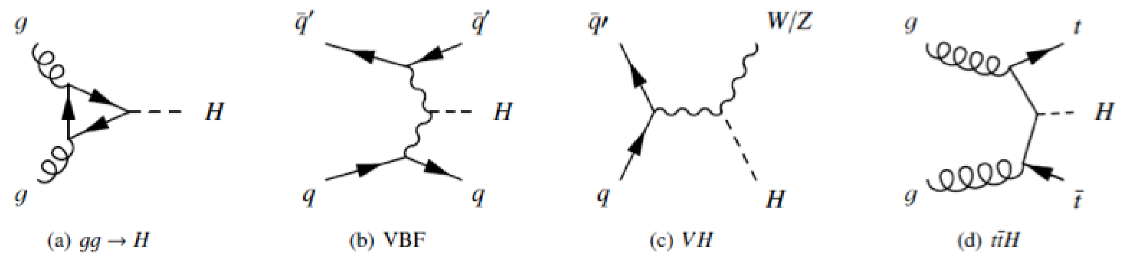


Figure 1: Leading order Feynman diagrams showing the different production mechanisms of the Standard Model Higgs boson at the ATLAS experiment [15].

In the remainder of the thesis a discussion is presented in Chapter 2 pertaining to the design of LHC with a large focus on the ATLAS detector. The theory relevant to a search for a Standard Model Higgs boson is discussed in Chapter 3. Chapter 4 explains an overview of Monte Carlo production at ATLAS. A look at required service work performed for the ATLAS Collaboration's Tau Validation Group is shown in Chapter 5. Contributions made to the $H \to \tau_{lep}\tau_{had}$ data analysis and search are shown in Chapters 6 and 7.

2 The Detector

2.1 The Large Hadron Collider

The LHC is a particle collider and along with a number of pre-accelerator systems is situated in a set of tunnels beneath the Swiss-French border which previously housed the Large Electron Positron collider. The LHC first became operational in 2008. Whilst the LHC has a design center-of-mass energy of $\sqrt{s} = 14$ TeV it has not yet operated at these energies. In 2012 the LHC ran at its highest operational energies to date of 7 TeV and 8 TeV [16] following which it entered a shutdown to receive the upgrades necessary to allow it to operate at the higher design center-of-mass energy [17]. The collider is expected to resume operation in 2015.

The LHC is mainly used to perform collisions between two high energy proton beams but can also facilitate the high energy collisions of two heavy ion beams. After passing through a number of pre-acceleration facilities the particles injected into the main ring are further accelerated using a system of boosting structures along the ring to reach a speed of 99.999997% of the speed of light. This main ring of the LHC sits in a 27 km long tunnel system and it is made up of super-conducting magnets that are maintained at a temperature of -271° C. These magnets produce the 8 T magnetic field that is used to guide the high energy beams of charged particles around the path of the collider's main ring. The two particle beams travel in two separate high vacuum tubes and once they acquire the necessary energy these beams are brought to collide at the centers of the four main detectors [18, 19].

The LHC hosts seven detectors at different positions around the ring. These detectors and the experiments associated with them are called ATLAS, CMS, ALICE, LHCb, TOTEM, LHCf and MoEDAL. The experiments have unique designs with varying sizes. ATLAS and CMS are both general purpose detectors which investigate a wide range of physics. All seven experiments were designed with specific purposes from finding evidence to support many current theories extending to the search for new physics.

Besides the LHC ring there also exists a set of pre-accelerators. The overview of the ring layouts can be seen in Figure 2. The operation of the LHC starts with particles that are initially produced by a particular source depending on the type of run. The particles are initially accelerated by a linear accelerator following which they are injected into the Booster and CPS (CERN Proton Synchrotron). These accelerate the particle beam until it reaches the injection momentum of 26 GeV necessary to enter the SPS (Super Proton Synchrotron) ring. The beam is accelerated further in the SPS ring and is then split in two with the two beams now injected into the main LHC ring at 450 GeV travelling in opposite directions. In this final ring each of the beams are accelerated to 4 TeV after which they are brought to collide with each other. The collision points are situated at different detectors on the main ring. [19]

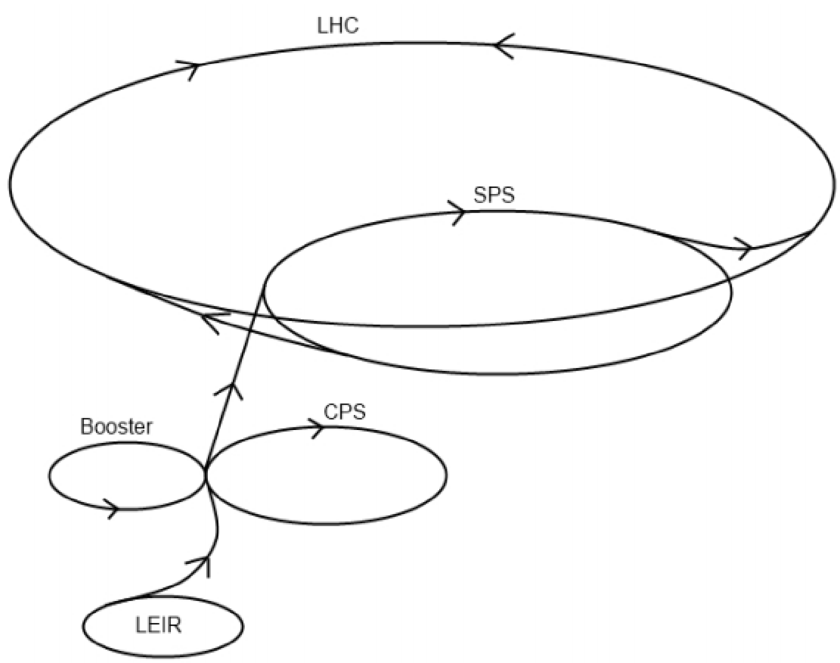


Figure 2: Overview of the collider layout at the CERN research facility.

2.2 The ATLAS Detector

The ATLAS detector is a general purpose detector with a barrel and two endcaps design. Some of the areas of research investigated by ATLAS include the search for dark matter, looking for Higgs bosons in multiple models, looking for supersymmetric particles, looking for extra dimensions and since the discovery of a Higgs boson, this of course extends to Higgs property measurements. Collision data is collected by the ATLAS detector using an arrangement of several sub-detection systems. These detection systems sit in a set of magnetic fields that bend the paths of charged particles which allows for the detector to identify the charge of the particles along with performing energy and momentum measurements. This complex and large (46m long, 50m in diameter, 7000 tonnes in weight) device has played an important role in leading research and results produced at CERN by teams from all over the world [20].

2.2.1 Design

The ATLAS detector utilises an 'onion' layout system with a barrel and two endcaps design. More specifically this 'onion' layout refers to how the detector is built as multiple layers of detection systems in the form of concentric cylinders and endcaps build surrounding a collision point on the beam path. The six sub detection systems fall into a simple hierarchy of sections which have been named the inner tracking detector, the calorimeters and the muon chambers. Each of these sections, whilst consisting of a number of specialised detection systems themselves, fulfill specific roles in identifying particles produced in collisions.

The magnetic field that these sub-detection systems sit in is generated by two systems of magnets. The Inner Detector is housed by a thin superconducting solenoid which produces a near uniform 2 T magnetic field allowing for highly precise measurements to be made. Further out, surrounding the calorimeters and sitting within the Muon Spectrometers is a large magnetic system which produces a non-uniform magnetic field with bending power ranging from 1 to 7.5 T·m. This barrel and two end-caps magnetic system is what lead the design of the rest of the ATLAS detector [21, 22, 23, 24]. Figure 3 provides an illustration of the design of the detector.

2.2.2 Co-ordinate System

In order to understand and properly record the result of a given particle collision, it is important that a rigid and efficient co-ordinate system is constructed in the ATLAS detector. The co-ordinate system is chosen in such a way to be centred around the interaction point of the collisions. The z-axis is defined in the direction of the beams path, the positive x-axis is defined pointing towards the centre of the LHC ring and the y-axis positive direction points upwards. The azimuthal angle (ϕ) and polar angle (θ) are then defined in the normal spherical coordinate fashion. The polar angle θ is used to define a more useful value known as the pseudorapidity

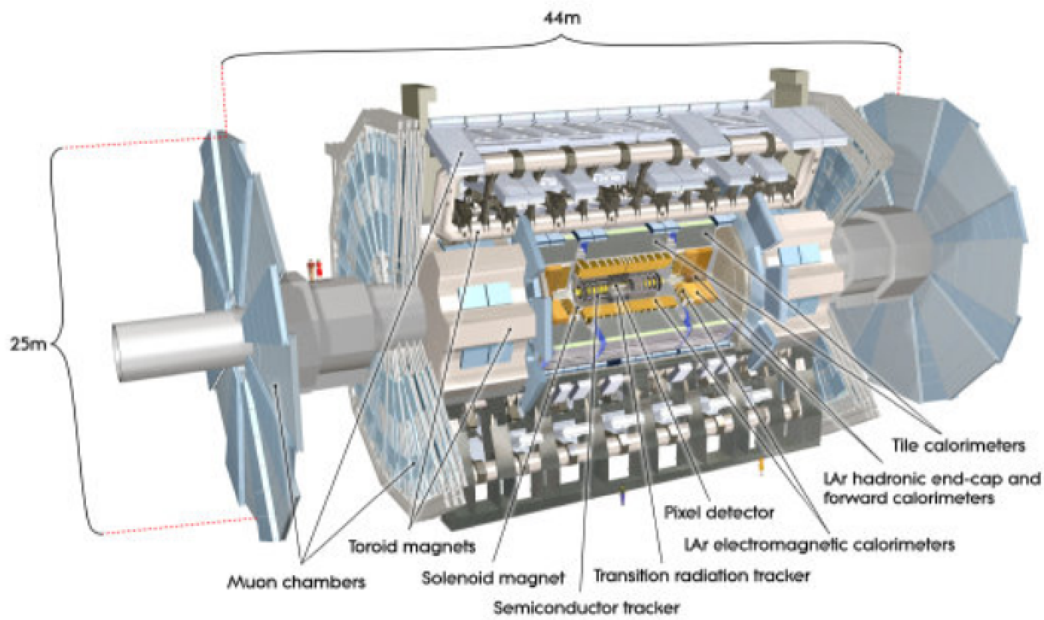


Figure 3: Design of the ATLAS detector [25].

which is denoted η,

$$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right). \tag{2.1}$$

The pseudorapidty and the azimuthal angle are then used to define the angular separation between tracks denoted as ΔR

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}. \tag{2.2}$$

Finally it is important to note that the transverse energy E_T, transverse momentum p_T, and missing transverse energy E_T^{miss} are all defined in the x-y plane.

2.2.3 Trigger System

The rate at which data is produced by the LHC and in ATLAS is so high that it is unrealistic to consider recording all collision events that occur. There are also a lot of QCD interactions that occur that are not considered important to record. To cope with this problem the ATLAS detector uses a specialised trigger system to filter which events should be kept and which should be discarded. The decision to keep or discard an event has to be made extremely quickly in order to be effective at reducing the initial bunch crossing rate. When running at the design luminosity of $10^{34}\text{cm}^{-2}\text{s}^{-1}$ the bunch crossing rate will be 40 MHz and the trigger system has been designed to reduce this to an output rate of ≈ 200 Hz.

The trigger system operates using three levels. L1 is the first level and is a hardware based system and uses information from the calorimeter and muon sub-detectors. L2 is the second level and along with the Event Filter (EF), which is the third level, is a software based system. These two systems use information from all sub-detectors and are together known as the High Level Trigger (HLT). For each event hundreds of conditions are checked and if one or more is passed the event will be kept. These conditions are based on identifying potentially interesting physics objects such as electrons, muons, taus, jets, b-jets, photons or missing energy [26].

2.2.4 Inner Detector

The Inner Detector, as the name suggests, sits in the innermost layer of the ATLAS detector. The layout of the Inner Detector has high resolution detectors in the layers closest to the interaction point with continuous tracking elements further out. All these elements are contained in the inner solenoid which produces a magnetic field surrounding the Inner Detector. The Inner Detector consists of three sub-detector systems; a set of high resolution semiconductor pixel trackers (Pixel), silicon microstrip trackers (SCT) and transition radiation trackers (TRT). The Inner Detector and its sub-systems are arranged in a barrel and two endcaps layout and provide coverage for $|\eta| < 2.5$. The fine granularity detector systems allow for the high precision tracking of charged particles which provides a detailed measurement of particle types and their momentum [21]. A more detailed discussion on the individual sub-detection systems follows. An illustration of the Inner Detector system can be seen in Figure 4.

2.2.4.1 Pixel detector

Pixel detectors in hadronic accelerators are high precision tracking detectors and play the vital role of providing good vertex reconstruction. The ATLAS Pixel detector is the innermost detector component. The design of the Pixel detector consists of three coaxial barrel pixel layers and two endcap disk sections each consisting of three disks. The barrel layers are a cylindrical arrangement of long structures each housing multiple pixel modules. The inner barrel sits closest to the beam pipe with a radius of $r = 50.5$ mm, the middle barrel at $r = 88.5$ mm and the outer barrel sits

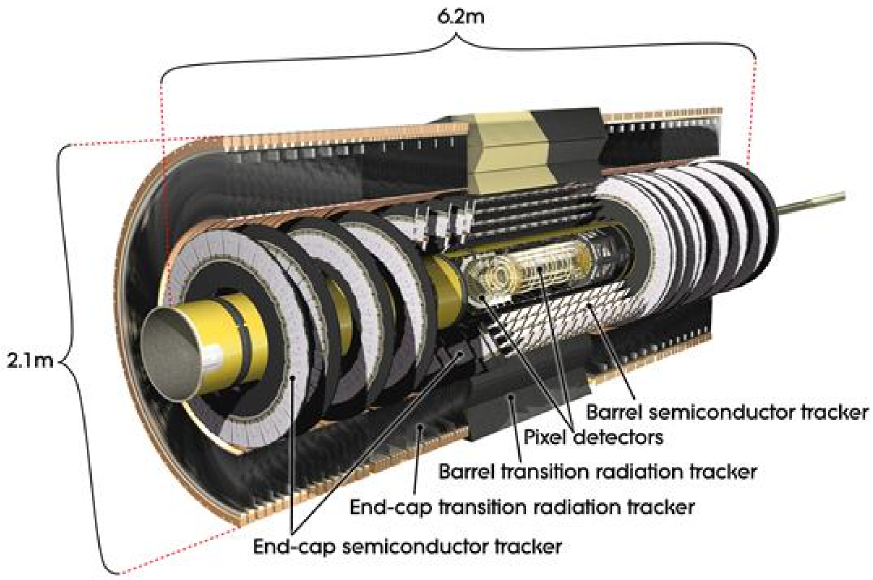


Figure 4: Design of the ATLAS Inner Detector system [25].

at $r = 122.5$ mm. The pixel detector is made up of a total of 1744 pixel modules. The basic design of these pixel modules consists of a silicon pixel sensor, bump-bonded to sixteen radiation-hard front-end readout chips which are then connected to a hybrid circuit holding the control chip and supporting interfaces to all necessary cable connections. Each silicon pixel module covers an active area of 16.4 mm $\times$ 60.8 mm and is divided into 47,268 pixels. With a total of 1744 pixel modules this gives $\sim 8 \times 10^7$ pixels which provides the very high granularity necessary to perform accurate vertex reconstruction [27, 28].

The silicon sensors consist of a silicon strip sandwiched between two charge conductive plates which maintain an electric field between them. High energy particles passing through this silicon layer ionize multiple electron hole pairs along their path. The electric field pulls the holes and electrons in opposite directions which are then absorbed by doped areas in the silicon. Charge build up in these doped areas can then be detected and read out. The three layers of pixels all produce readings which when combined together can be used to determine the paths of particles and form tracks. These tracks can then be used to reconstruct vertices.

2.2.4.2 Semi-Conductor Tracker Detector

The SCT detector is another high precision tracking detector within the Inner Detector system. The SCT system sits between the Pixel and TRT detectors. The system consists of a barrel section comprised of four coaxial layers and two end-cap disk sectors of nine disks each. The innermost layer of the SCT sits at a radial distance of $r = 299$ mm from the beam axis with outer reaching $r = 560$ mm and is 5.6 m in length. The SCT design follows that of the pixel detector but instead of pixel modules uses silicon-strip sensors. The 2112 modules on the barrels have constant pitch of 80 μm and are arranged approximately parallel to the magnetic field and beam axis. The 1976 modules on the end-caps are arranged radially and have variable $57 - 94$ μm pitch.

The silicon strip modules are made of two pairs of silicon microstrip sensors which are glued back to back. To make two-dimensional track reconstruction possible the strips are glued together at a 40 mrad stereo angle. This back-to-back design along with the four barrel layers typically provides eight measurements, resulting in four space-points for particles that pass through the SCT. The sensors on each side of the strip are daisy-chained to each other to make 768 strips of ≈ 12 cm in length. The strips are read out using radiation-hard front-end readout chips and are mounted onto a hybrid circuit similar to the pixel detectors hybrid circuits. Each chip reads out 128 channels and in total gives the detector $\approx 6.3 \times 10^6$ channels. Although the SCT detector has a coarser granularity than the pixel detector, its position further from the interaction point means it still does an excellent job at reconstructing charged particle interactions [28, 29].

2.2.4.3 Transition Radiation Tracker

The TRT is the outermost sub-detector system in the Inner Detector. This detector uses thin-wall drift tubes (straws) arranged in the barrel and two end-cap design. The straws are 4 mm in diameter and have lengths of 39 cm in the end-cap regions and 144 cm in the barrel regions. The barrel region has the straws aligned along the direction of the beam axis in three coaxial layers and the end-cap regions are comprised of 80 wheel-like arrangements of straws. The TRT detectors overall structure is 2.1 m in diameter and 6.2 m long and contains a total of 298,304 straws which are readout by 350,848 channels. The overall layout of these straws means that particles with transverse momentum $p_T > 0.5$ GeV and pseudorapidity $|\eta| < 2.0$ will cross on average 30 straws providing 30 two-dimensional space-points for the particle tracks. Between these straws are woven polypropylene fibres in the barrel region and polypropylene foils in the endcaps. These foils and fibres cause high velocity particles to produce transition radiation x-ray photons which can then be measured. The main objective of the TRT detector is providing accurate charged particle track measurement but the measurement of transition radiation also allows for improved particle identification. Particularly the measurement of transition radiation allows for electrons and hadrons to be distinguished from one another [28, 30, 31].

The straws themselves are constructed using two strips of film which are wound together in spirals to create a tube. The film is coated on one side with a conductive layer and on the other side with an insulating layer. A gold plated tungsten wire is placed in the middle of the tube which is filled with a gas mixture of 70% Xe, 27% CO_2 and 3% O_2. A strong electric field can then be maintained in the tube by creating a high potential difference between the wire and the tube itself. A charged particle that passes through the tube will ionize the gas along its path. The electric field maintained causes the positive ions to move towards the tube wall and the negative ions to move to the central wire. Upon reaching either the wall or the wire the ions produce a signal which can then be measured [31, 32].

Due to the polypropylene fibres and foils that fill the space between the straws particles that pass through the detector transition between interfaces of materials with different dielectric constants. When a charged particle passes through such an interface a transition radiation x-ray is created. These x-rays pass through the straws and produce signals which can help to distinguish between the different types of particles. The particle loses energy proportional to its Lorentz factor leading to the transition radiation measurement allowing the momentum of the original particle to be determined, and helping to identify what type of particle it is [32].

2.2.5 Calorimeters

Calorimetry is a particle detection technique that is used in accelerator based particle physics primarily to detect and measure energies of electrons, photons and hadrons. Calorimeters are built of materials that absorb the energy of these particles as they pass through and provide a means for that absorbed energy to be measured. Calorimeters use either an electromagnetic or hadronic design whilst the absorber material can be sampling or homogeneous. The ATLAS experiment has both an electromagnetic and a hadronic calorimeter which each use sampling absorber materials [33].

Sampling calorimeters are divided into layers of ‘active’ and ‘passive’ materials whereas a single material fulfilling the purpose of both these layer types is used in a homogeneous calorimeter. Since ATLAS uses sampling calorimeters, they will be the focus of this discussion. When a particle enters a sampling calorimeter it interacts with the ‘passive’ material which is a dense absorber material. The strong or electromagnetic interactions with this dense material cause the particle to decay into a shower of multiple particles of lower energies. This shower of particles then proceed to interact with the next layer which is an ‘active’ sampling material. The particles deposit energy into this material and it will produce a detectable signal. The particles in the shower will then continue to the next absorber layer, produce a new shower of lower energy particles and continue to the next sampling layer. This continues until they are completely absorbed or they pass through the detector. It is important that the calorimeters are designed to be thick enough to be able to completely absorb these showers. The advantage of using these alternating materials over a single material is that the two materials can be chosen to do their specific job with very high efficiency. The disadvantage however is that some energy gets deposited in the absorber layer. The calorimeter can then only sample a fraction of the particles energy from which the total energy is then estimated [33, 34]. The layout of the ATLAS calorimeter systems can be seen in Figure 5 which shows both the Liquid Argon electromagnetic Calorimeter (LAr EM Calorimeter) and the hadronic Tile Calorimeter.

2.2.5.1 Argon Calorimeter

The LAr EM Calorimeter design consists of a barrel region and two endcaps. The barrel region is comprised of two half barrels sharing a single cryostat collectively called the electromagnetic barrel (EMB) calorimeter. This barrel region provides calorimetry coverage for the region of $|\eta| < 1.475$. The two endcaps are housed in separate cryostats and cover the range $1.375 < |\eta| < 3.2$. The endcap cryostats contain a number of disk-like detector systems including the Electromagnetic Endcap Calorimeter (EMEC), the hadronic endcap colorimeter (HEC) and the Forward Calorimeter (FCal). There is also a LAr presampler which corrects for energy that the particles lose before entering the calorimeters. All systems in the LAr calorimeter use liquid argon as a sampling material [24, 35].

EMEC is a system of two disks with one in each endcap. Each of the EMEC

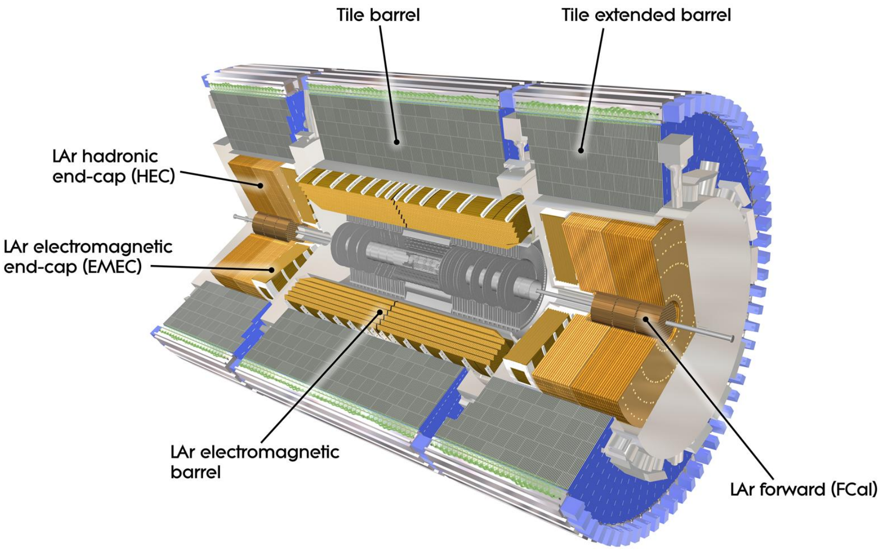


Figure 5: The ATLAS Calorimeter system [25].

disks is comprised of two separate co-axial disks. The inner disk covers the region of $2.5 < |\eta| < 3.2$ with the outer disk covering $1.375 < |\eta| < 2.5$. Each endcap also contains a HEC with each HEC consisting of two parallel-plate calorimeters. These HECs assist the ATLAS hadronic calorimeter, that will be discussed later, by covering the range of $1.5 < |\eta| < 3.2$ and utilise copper as an absorber material. Lastly, the FCal is a system of three sequential cylindrical modules in each endcap covering the range of $3.1 < |\eta| < 4.9$. The module closest to the interaction point uses copper as an absorber material being designed for EM measurements, whilst the other two use tungsten as an absorber and are designed for hadronic measurements [35].

Both EMEC and EMB use accordion-shaped copper-kapton electrodes which are placed between lead absorber plates. The absorbers and electrodes together are used to maintain an electric field created between them. These electrodes and absorber plates are held in position by honeycomb spacers and the entire system is immersed in liquid argon. Particle showers caused by high energy particles interacting with the absorber plates ionize the liquid argon causing many electron hole pairs to be created. The electric field between the plates and electrodes cause these electron hole pairs to drift in opposite directions. Upon reaching the electrodes the electrons create a pulse that can then be measured. These EM systems will typically contain electrons and photons whilst hadrons will pass through to the hadronic calorimeter [35]. The operation of the HEC and FCal systems works on the same basis as discussed here.

2.2.5.2 Tile Calorimeter

The hadronic tile calorimeter surrounds the LAr calorimeter. The tile calorimeter is designed in the same barrel and two endcaps fashion used for the LAr calorimeter. The barrel region consists of a long cylinder split into two separate read out channels with the endcaps, called extended barrels, consisting of a single shorter cylindrical structure on each end of the main barrel. The tile calorimeter uses plastic scintillating plates as a sampling material with iron plates used as the absorbers. The particle showers produced through incident particles interacting with the absorbers interact with the scintillating plates and the emitted photons are collected. The intensity of the light produced by these plates is proportional to the energy of the particles that passes through them. The light can be transmitted using special fibers and photomultiplier tubes allowing for the signal to be recorded, providing measurements for the energy of particles passing through the calorimeter. The calorimeter is used to make measurements of hadrons, jets and taus along with generating input required by the ATLAS trigger system. The entire design provides hadronic calorimetry coverage for the range of $|\eta| < 1.7$ [36].

2.2.6 Muon Spectrometer

Muons have a mean lifetime of $\approx 2.2 \times 10^{-6}$ s [37] but the beam collision results in muons with such high momentum that they can be considered stable within the frame of the detector. Typically muons are the only detectable particles that will pass through the calorimeters. The Muon Spectrometer is the outer most detection layer. The Muon Spectrometer operates using superconducting air-core magnet systems to bend the paths of muon tracks and performs measurements using separate trigger and high-precision tracking chambers. With the same barrel and two endcaps design used throughout ATLAS, the different trigger and high-precision tracking chambers are split between the barrel and endcaps with each sector having its own magnet system. The barrel magnetic system provides 1.5 to 5.5 T·m of bending power whilst the endcap magnetic systems provide 1 to 7.5 T·m of bending power. An overview of the muon spectrometer design can be seen in Figure 6 and Figure 7

[23, 24].

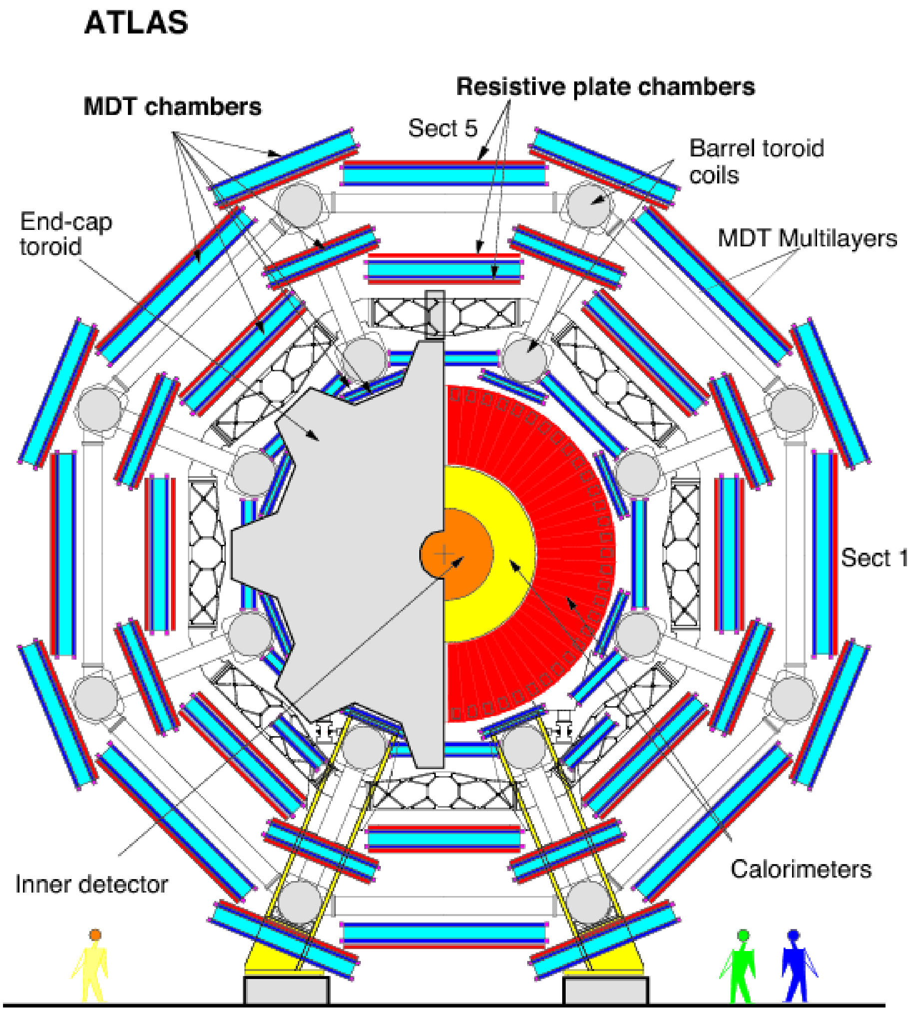


Figure 6: ATLAS Muon Spectrometer in the x-y projection [38].

In the barrel region ($|\eta| < 1.4$) the magnetic field is generated using a large toroid magnet. The endcaps each have smaller magnets placed in them which cover a range of $1.6 < |\eta| < 2.7$. The magnetic bending between the barrel and endcap regions ($1.4 < |\eta| < 1.6$) is contributed to by both magnet systems. This arrangement produces a magnetic field which is approximately orthogonal to all muons travelling through the detector [24].

2.2.6.1 Precision Tracking

The precision-tracking system in the muon spectrometer consists of Monitored Drift Tubes (MDT) and Cathode-Strip Chambers (CSC). These precision detectors are

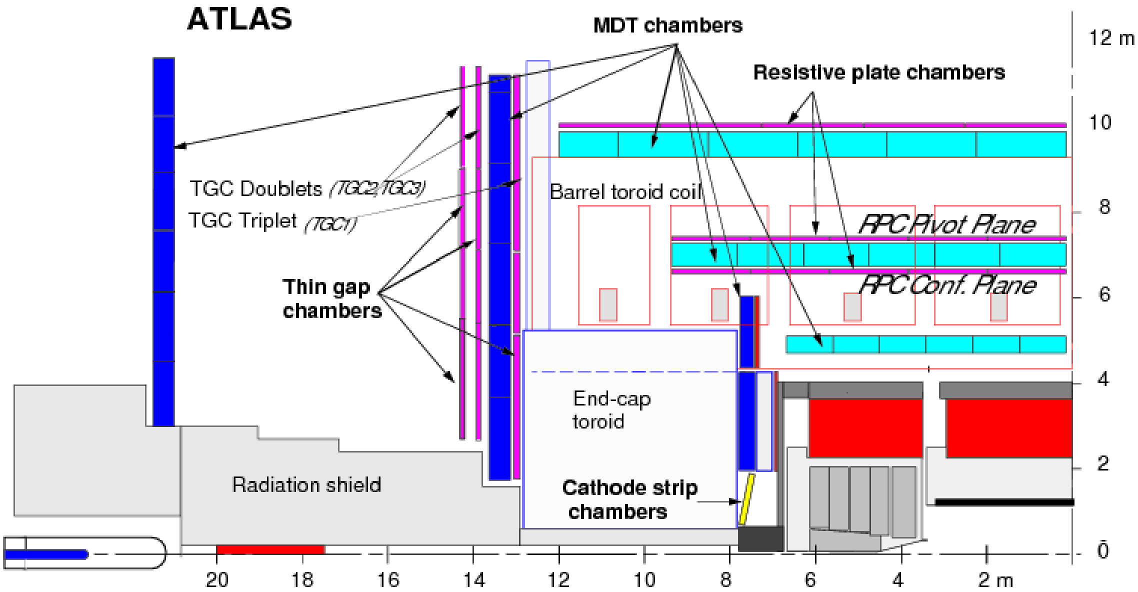


Figure 7: ATLAS Muon Spectrometer in the z-y projection [38].

arranged in three coaxial layers in the barrel region and three disk sectors (or layers) in each endcap region as seen in Figure 7. Since these disk and barrel sectors work in the same way the term *station* will be used to refer to them in general, that is, a station is one of the three layers in the spectrometer. Each station is structured as multiple layers of monitored drift tubes or cathode-strip detectors. The cathode-strip chambers cover the region of $2 < |\eta| < 2.7$ with the monitored drift tube chambers covering the region of $|\eta| < 2$ (see Figure 6). These systems allow for the detector to identify muons with momenta greater than 3 GeV and take accurate measurements of their p_T in the range $p_T < 1$ TeV [23, 24].

The drift tubes used in the MDTs are filled with pressurised gas with a wire which acts as an electrode running down the center of the tube. The diameter of this pipe is 29.970 mm and is filled with a gas mixture of Ar(97%) and CO_2(7%) maintained at a pressure of 3 bar. The wire electrode is made of tungsten-rhenium and an electric field is maintained between this wire and the tube. Muons that pass through the tube ionize the gas and the electrons that are produced drift towards the central wire and create a measurable pulse upon reaching it [24].

The CSCs consist of multiple anode wires within two planar cathodes made of strips which enclose a volume that is filled with pressurised gas. The wires are arranged in the radial direction. One of the cathodes has its strips running perpendicular to the wires and the other has the strips running parallel to the wires. The enclosed chamber volume is filled with Ar (30%) CO_2 (50%) and CF_4(20%) [39]. When a muon passes through it ionizes the gas, creating electron-hole pairs which then travel towards the strips on either side. These electrons create a measurable charge on the individual strips on either cathode. The strips are arranged perpendicularly to each other providing a two-dimensional reading of the location of the particles interaction as seen in Figure 8. Layers of these cathode strips exist in each chamber which allows for the track to be reconstructed by interpolating the measurements on neighbouring layers [23, 24].

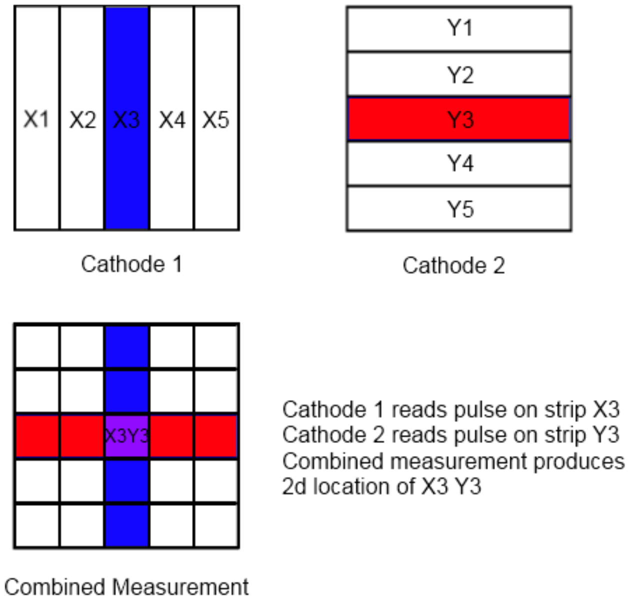


Figure 8: Depiction of how perpendicular strip arrangement produces two-dimensional space point.

2.2.6.2 Trigger

The muon spectrometers trigger system employs both resistive plate chambers (RPC) and thin gap chambers (TGC). These trigger chambers are designed to provide fast measurements of muon tracks allowing the L1 trigger system to identify their multiplicity and energy range. The system covers the range of $|\eta| < 2.4$ and due to varying requirements with changes in η the barrel and endcap regions needs to use different detector technology. These requirements change due to varying average muon momenta, the non-uniform magnetic field and differing radiation levels. In the barrel range of $|\eta| < 1.05$, RPCs are used. In the endcap regions covering the remaining range of $1.05 < |\eta| < 2.4$ TGCs are used. These RPCs and TGCs are placed at selected stations in the muon spectrometer [23, 24].

The RPCs are gaseous detectors comprised of two parallel electrode-plates which are resistive plates made of phenolic-melaminic plastic laminate. These plates are kept 2 mm apart and an electric field is maintained between them. The space between the plates is filled with a gas mixture of $C_2H_2F_4$ (94.7%), iso–C_4H_{10} (5%) and SF_6 (0.3%). The gas is ionized when a charged particle passes through the detector and the electron hole pairs that are created drift towards the plates creating a measurable pulse upon reaching them [23, 24].

The TGCs are designed and work in the same way that the CSCs work. They differ only in the distance between the anodes and cathodes and in that they use a different gas mixture of CO_2 (55%) and n-pentane (45%) [39]. In addition to being used for the trigger system these chambers also provide a measurement of a second azimuthal co-ordinate to complement the MDTs. Each middle endcap station of the muon spectrometer has seven layers of TGC and each inner most endcap station has two layers of TGCs [23, 24].

2.3 Grid Computing

The data produced at the LHC needs to be accessible to scientists all around the world who use it in their analyses. The LHC Computing Grid (LCG) was formed with the goal of building and maintaining a data storage and analysis infrastructure to accommodate this need. The model chosen to fulfil these requirements was a globally distributed computing Grid using a four-tiered approach [40].

Tier-0 is located at CERN and stores all the raw data that emerges from the data acquisition systems associated with the different experiments at the LHC. This data is used in a first pass reconstruction done at Tier-0, and this reconstructed data is also stored at Tier-0. Copies of the raw data and the first pass reconstructed data are also distributed to different Tier-1 sites.

Tier-1 sites have the prime responsibility of managing permanent data storage for raw, simulated and processed data. These sites are also required to provide computational facilities for reprocessing data and analysis that require access to large amounts of data.

Tier-2 sites are responsible for storing Monte Carlo event simulation and for providing computation facilities for scientists around the world to perform computational analysis. The Monte Carlo data created here is sent to Tier-1 for permanent storage and these Tier-2 sites will then pull data from the Tier-1 sites as necessary. There are over 100 Tier-2 sites.

Tier-3 sites are located around the world at many universities. These are generally used by smaller experimental groups for analysis only and are provided with access to the data facilities in the same way as Tier-2 sites are.

In practice, when computation needs to be run on large amounts of data it is not practical to download the datasets locally due to their huge sizes. Instead, CERN computing facilities provide tools that allow jobs to be created and sent to different computing sites on the LHC Computing Grid where the required data is located. The computational jobs are then run remotely and the results alone can be downloaded by the analyser.

3 Theory

3.1 Elementary Particles

Modern particle physics describes a set of fundamental, or elementary, particles. These particles have no sub-structure and through interaction with each other, form all observed matter and anti-matter in our universe. The elementary particles include a set of fermions, that is quarks, leptons and their anti-partners, and a set of bosons which include the gauge bosons and the newly discovered Higgs boson. All the fundamental particles in the formulation of the Standard Model are discussed briefly before looking at their roles in the Standard Model. This introduction to the fundamental particles in the Standard Model is summarised from [41, 42, 43]. An introductory discussion on the Standard Model Higgs boson was presented in the Introduction section of this thesis. The Higgs boson will be further discussed in the Standard Model section of this thesis.

3.1.1 Leptons

Leptons are part of the fundamental fermion particles all of which carry spin $\frac{1}{2}$. They exist in three generations of particle doublets. For each lepton, there exists an anti-particle partner, as predicted by the Dirac equation. The anti-lepton and lepton have opposite charges and upon meeting, they will annihilate.

Each generation of lepton doublet consists of a charged particle and a neutral particle. The three charged leptons are the electron, the muon and the tau, each of which carries the same charge of $\pm e$. For each of these charged leptons the corresponding neutral particles are their associated neutrinos which are the electron neutrino, the muon neutrino and the tau neutrino. The electron is the only stable charged lepton whilst all three neutrinos are stable. A summary of the lepton generations can be seen in Table 1.

Leptons only experience the force of gravity, weak interactions and, if they carry a charge, electromagnetic interactions. All interactions with leptons must obey charge conservation. Each generation of leptons can be assigned a lepton number, with a given generation's anti-particles having the negated lepton number. This lepton number has been observed to be conserved in lepton interactions.

Leptons			
Generation	I	II	III
Name	Electron (e)	Muon (μ)	Tau (τ)
	Electron Neutrino (ν_e)	Muon Neutrino (ν_μ)	Tau Neutrino (ν_τ)

Table 1: Table of Leptons in their three generations.

3.1.2 Quarks

Quarks form the other half of the fundamental fermion particles. Quarks also exist in three generations of doublets. Each quark has a corresponding anti-quark which upon meeting both will annihilate. Each type of quark has a specific quark flavour. Each flavour of quark carries a different fractional charge. Quarks are known to experience the strong force in addition to the force of gravity, the electromagnetic force and the weak force experienced by leptons. The three generations of quarks can be seen in Table 2.

Quarks cannot exist freely and can only be seen in bound states of multiple quarks called hadrons. Hadrons can either be baryons or mesons where a baryon is made up of three quarks and a meson is a quark and an anti-quark. A baryon quark configuration of three up quarks (uuu) in the same quantum state was discovered to be the necessary configuration for the Δ^{++} baryon. It appeared as though there was a hidden quantum number associated with quarks, for then this would otherwise violate the Pauli exclusion principle. The new quantum number associated with quarks was called colour. Each quark posses one of three colours arbitrarily named red(R), green(G) and blue(B). The Δ^{++} baryon could then be explained by having each of these three up quarks having one of the three possible quark colours with the configuration written as $\mathrm{u}_R\mathrm{u}_G\mathrm{u}_B$. Further, the phenomenon known as colour confinement requires that hadrons are white or colourless. The simplest configurations

Quarks			
Generation	I	II	III
Name	Up(u)	Charm(c)	Top(t)
	Down(d)	Strange(s)	Bottom(b)

Table 2: Table of Quarks in their three generations.

that fulfil this requirement are three quarks of different colour making up baryons, and a quark with an anti-quark having the corresponding anti-colour to the quark making up mesons. The three anti-quark colours are denoted $\bar{\text{R}}$, $\bar{\text{G}}$ and $\bar{\text{B}}$.

3.1.3 Gauge Bosons

Leptons and quarks are the fundamental constituents of matter in our universe. However, there exists more fundamental particles that mediate the interactions between the leptons and quarks. These particles are all bosons and are the transmitters of the different forces that are present in these particle interactions. These particles are the photon, the W^+ and W^- bosons, the Z^0 boson, eight gluons and possibly the graviton. A description of these particles and which forces they mediate can be seen in Table 3. Whilst the graviton is predicted to act on all particles, a complete theory of quantum gravity has not yet been produced. Modern elementary particle physics assumes that gravity is too weak to play an important role in the interaction between the fundamental particles. The photon is a massless, neutral particle of spin 1. The eight gluons are massless, have spin 1 and are neutral in charge, but all carry different colour charges as will be seen later. The W^+ and W^- bosons are spin 1, massive ($M_W \approx 80.4$ GeV [37]) and charged bosons. Finally the Z^0 particle is a spin 1, massive ($M_Z \approx 91.2$ GeV [37]) and neutral particle. Further details about these gauge bosons will be discussed in the Standard Model section of this thesis.

3.2 Global Symmetry Breaking

The concept of symmetry breaking is important in understanding how the Standard Model works. Symmetry breaking in particle physics is what leads to the BEH mechanism giving mass to certain particles and also shows how the Higgs boson is

Gauge Boson	Force Mediated	Relative Strength
Gluon(g)	Strong	10
Photon(γ)	Electromagnetic	10^{-2}
$\text{W}^\pm$ and Z^0	Weak	10^{-13}
Graviton	Gravity	10^{-42}

Table 3: Description of the different Gauge bosons.

exposed in the Standard Model. First a discussion on global symmetry breaking is presented which will then be extended to a discussion on local symmetry breaking.

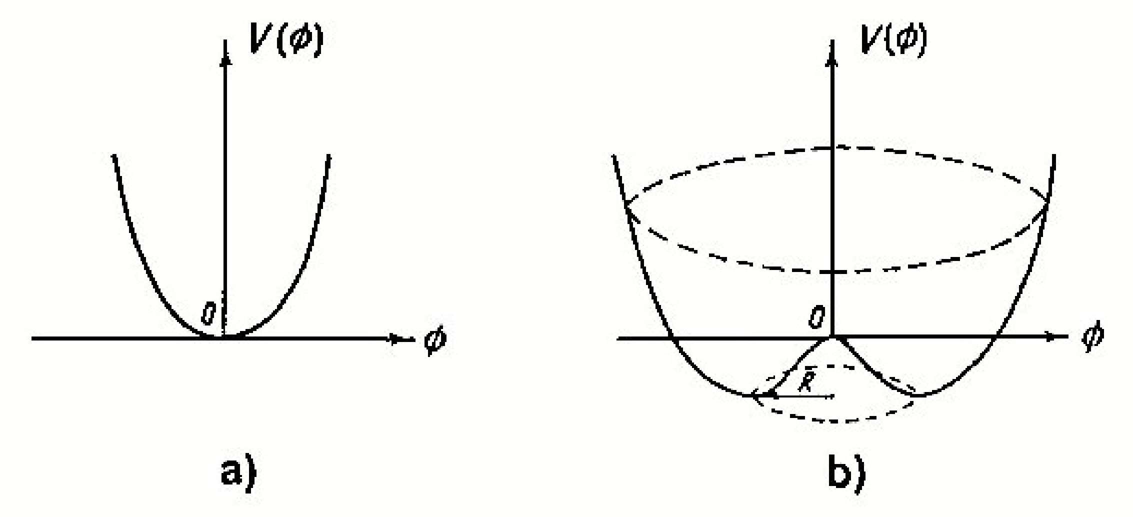


Figure 9: a) Plot of the potential in equation (3.2). b) Plot of the Mexican hat potential in equation (3.3).

Start by considering the Lagrangian for a complex scalar boson ϕ

$$\mathcal{L} = -\frac{1}{2}\partial^{\mu}\phi^{\dagger}\partial_{\mu}\phi - V\left(\phi^{\dagger}, \phi\right). \tag{3.1}$$

Initially take the potential term to be of the form

$$V\left(\phi^{\dagger}, \phi\right) = \frac{1}{2}m^2\phi^{\dagger}\phi. \tag{3.2}$$

This potential term allows the Lagrangian to have a $U(1)$ symmetry. A plot of V against $|\phi|$ can be made which is seen in Figure 9 a). Here, V_{min}, the minimum potential or vacuum expectation value (VEV), is situated at $|\phi|^2 = 0$. This indicates that the vacuum for this potential is at $\phi = 0$ and corresponds to no particles being present. Change the potential to be of the form

$$V\left(\phi^{\dagger}, \phi\right) = \frac{1}{2}\lambda m^2\left(\phi^{\dagger}\phi - \Phi^2\right)^2. \tag{3.3}$$

Here λ and Φ are real constants and the previous $U(1)$ symmetry is maintained. A plot of this new potential has a "Mexican hat" form as seen in Figure 9 plot b). V_{min} is now at $|\phi| = \Phi$ and is shown by the dashed circle of radius R in the figure. Since it is possible to choose an infinite number of points on this circle it shows that there are an infinite number of vacuums for this potential. This vacuum can be parametrized by $e^{i\alpha}$ meaning that there will be a vacuum for every value of α. So a vacuum needs to be chosen manually by choosing a value for α. Choose α such that the ϕ is real and is equal to Φ. This process is called gauge fixing the symmetry. It has broken the $U(1)$ symmetry that was in the Lagrangian. Expand around the constant vacuum giving the result

$$\phi = \Phi + \alpha + i\beta. \tag{3.4}$$

Here α and β are new real scalar fields. Using this new information the Lagrangian can now be written out and fully expanded to

$$\begin{aligned} \mathcal{L} = & -\frac{1}{2}\partial^\mu \alpha \partial_\mu \alpha - \frac{1}{2} 4\lambda m^2 \Phi^2 \alpha^2 - \frac{1}{2}\partial^\mu \beta \partial_\mu \beta \qquad . \\ & -\frac{1}{2}\lambda m^2 \left(4\Phi\alpha^3 + 4\Phi\alpha\beta^2 + \alpha^4 + \alpha^2\beta^2 + \beta^4\right) \end{aligned} \tag{3.5}$$

It can be seen that now there exists a massive real scalar field α with a mass of $\sqrt{4\lambda m^2 \Phi^2 \alpha^2}$ and massless real scalar field β. This Lagrangian no longer shows a $U(1)$ symmetry. The symmetry has been broken and it is for this reason that the process followed above is called symmetry breaking. As seen here, global symmetry breaking always leads to to an extra massless field being created and these particles have been named Goldstone Bosons. Local symmetry breaking is the next step which works very similarly to this process and is what leads to the BEH mechanism in the Standard Model.

3.3 Local Symmetry Breaking

Local symmetry breaking uses the same process that global symmetry breaking follows. Start by considering the Lagrangian for a complex scalar ϕ

$$\mathcal{L} = -\frac{1}{2}\left[\left(\partial^\mu - iqA^\mu\right)\phi^\dagger\right]\left[\left(\partial_\mu + iqA_\mu\right)\phi\right] - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} - V\left(\phi^\dagger, \phi\right). \tag{3.6}$$

This Lagrangian has been made to be symmetric under a local $U(1)$ gauge using a process known as symmetry gauging. This results in the term $-\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$ being added which is a kinetic term with $F^{\mu\nu}$ defined as

$$F^{\mu\nu} = \frac{i}{q}\left[D^\mu, D^\nu\right] = \partial^\mu A^\nu - \partial^\nu A^\mu. \tag{3.7}$$

Choose the potential V to be of the same form as previously seen in (3.3). The vacuum still has the same degeneracy at $|\phi| = \Phi$ but the local symmetry now means that α has a local dependency and is written $\alpha(x)$. This means $\alpha(x)$ can be chosen so that both the vacuum is real and ϕ is real. Expanding around the vacuum then gives

$$\phi = \Phi + h. \tag{3.8}$$

Here h is a real scalar field which represents the fluctuations around the vacuum. Substituting back into (3.6) and expanding gives the result

$$\mathcal{L} = -\frac{1}{2}\partial^{\mu} h \partial_{\mu} h - \frac{1}{2} 4\lambda m^2 \Phi^2 h^2 - \frac{1}{4} F^{\mu\nu} F_{\mu\nu} - \frac{1}{2} q^2 \Phi^2 A^2 + \mathcal{L}_{int}\,. \tag{3.9}$$

Now $\mathcal{L}_{int}$ is all the interaction terms which are not important in writing out for this discussion. The interesting result is that the theory started out with a complex scalar field ϕ and a vector field A^{μ} but there now exists a field A^{μ} with a mass of $q\Phi$ and a real scalar h with a mass of $\sqrt{4\lambda m^2 \Phi^2}$. Local symmetry breaking has removed a Goldstone boson resulting in a massive gauge boson A^{μ} [44]. This process that has given mass to the theory is called the BEH Mechanism and the field h is the Higgs boson. The particles in the theory have gained mass through interacting with the field associated with the Higgs boson. This mechanism plays a large role in the formulation of the Standard Model which will be discussed now.

3.4 The Standard Model

For decades now the Standard Model of particle physics has been at the forefront of the particle physics world. This theory predicting the universe's fundamental particles and how they interact with each other has seen huge success in its predictions, pre-dating experimental observations by many years. The model is the best model to date at describing the interactions of the fundamental particles of our universe. The model's success has lead to a huge interest in further studying it and trying to investigate all aspects of it experimentally. Whilst this thesis focuses on the experimental aspect of studying the Standard Model, understanding and discussing the Standard Model is still important. There are a number of ways to formulate the Standard Model theory and the specific methodology chosen here follows the structure presented in the *Symmetry and the Standard Model: Mathematics and Particle Physics* [45] by Matthew Robinson.

The Standard Model is based on a symmetry group written $SU(3)_C \times SU(2)_L \times U(1)_Y$. $SU(3)_C$ is the special unitary colour (C subscript) group of order three, $SU(2)$ is the special unitary group of order two representing weak isospin which acts only on left-handed (L subscript) states and finally, $U(1)_Y$ is the unitary group of order one where Y stands for hypercharge which are used to distinguish this group from the $U(1)$ group for electromagnetism [46].

Table 4 shows how the leptons and quarks are presented in the Standard Model symmetry groups. Only the first generations of the quarks and leptons are shown here since, as is discussed later, this is all that is needed in the formulation that will be used. The third column in the table is a triplet showing the representation of the corresponding set of particles in each of the three factors making up the Standard Models symmetry group. The first entry is the $SU(3)_C$ representation, the second the $SU(2)_L$ representation and the third is the $U(1)_Y$ representation. These representations play important roles in the derivations that follow. Also note that right handed leptons and quarks have the same representation as their left handed anti-partners.

3.4.1 Electroweak Theory

The Standard Model Lagrangian is contributed to by multiple sectors in the theory. The Lagrangian will be built piece by piece and the journey starts with the electroweak theory behind the Standard Model. The mathematics describing the electroweak sector arise from the $SU(2)_L \times U(1)_Y$ part of the main symmetry group. Using this sub-symmetry group a renormalisable gauge theory with a short range for the short ranged weak interactions and a long range $U(1)_{EM}$ for electromagnetism will be created.

Start by considering a complex scalar doublet field ϕ with the representation $(2, -\frac{1}{2})$ in $SU(2)_L \times U(1)_Y$ and look at finding its co-variant derivative. A gauge field A^μ can be written as the scalar co-efficients of each generator of a symmetry group as

$$A^\mu = A_a^\mu T_\mu^a, \tag{3.10}$$

where A_a^μ is a scalar function and T_μ^a are the symmetry group generators. $SU(2)_L \times U(1)_Y$ has two symmetry groups so both sets of generators need to be taken into

Particles	Symbol	Representation $(SU(3)_C, SU(2)_L, U(1)_Y)$
Left-handed leptons	$L_i = \begin{pmatrix} v_i \\ e_i \end{pmatrix}_L$	$(1, 2, -\frac{1}{2})$
Left-handed anti-leptons	$\bar{e}_{iL}$	$(1, 1, 1)$
Left-handed quarks	$Q_i = \begin{pmatrix} u_i \\ d_i \end{pmatrix}_L$	$(3, 2, \frac{1}{6})$
Left-handed anti-up quark	$\bar{u}_{iL}$	$(\bar{3}, 1, -\frac{2}{3})$
Left-handed anti-down Quark	$\bar{d}_{iL}$	$(\bar{3}, 1, \frac{1}{3})$

Table 4: The $(SU(3)_C, SU(2)_L, U(1)_Y)$ representations of quarks and leptons.

account. The $SU(2)$ generators are defined as

$$T^a = \frac{1}{2}\sigma^a, \tag{3.11}$$

where σ represents the Pauli matrices. Denote the gauge fields for this group as W_μ^a. For $U(1)_Y$ the generators are defined as

$$Y = C \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{3.12}$$

$$= -\frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{3.13}$$

Here, $C = -\frac{1}{2}$ is the hypercharge for the complex scalar field ϕ. The gauge fields for $U(1)_Y$ are denoted as B_μ. Using these structures the covariant derivative of ϕ is then

$$\left(D_\mu \phi\right)_i = \partial_\mu \phi_i - i\left[g_2 W_\mu^a T^a + g_1 B_\mu Y\right]_{ij} \phi_j, \tag{3.14}$$

where g_1 and g_2 are coupling constants. Using equation (3.11), equation (3.13) and the definition of the Pauli matrices the following can be written

$$g_2 W_\mu^a T^a + g_1 B_\mu Y = \frac{1}{2} \begin{pmatrix} g_2 W_\mu^3 - g_1 B_\mu & g_2\left(W_\mu^1 - iW_\mu^2\right) \\ g_2\left(W_\mu^1 + iW_\mu^2\right) & -g_2 W_\mu^3 - g_1 B_\mu \end{pmatrix} . \tag{3.15}$$

Now the matrix form of the covariant derivative becomes

$$\left(D_\mu \phi\right)_i = \begin{pmatrix} \partial_\mu \phi_1 + \frac{i}{2}\left(g_2 W_\mu^3 - g_1 B_\mu\right)\phi_1 + \frac{ig_2}{2}\left(W_\mu^1 - iW_\mu^2\right)\phi_2 \\ \partial_\mu \phi_2 - \frac{i}{2}\left(g_2 W_\mu^3 + g_1 B_\mu\right)\phi_2 + \frac{ig_2}{2}\left(W_\mu^1 + iW_\mu^2\right)\phi_1 \end{pmatrix} . \tag{3.16}$$

This covariant derivative can be used to write out the locally gauge invariant part of the full Lagrangian. In general this is

$$\mathcal{L}_\phi = D_\mu \phi_i^\dagger D^\mu \phi_i - V\left(\phi^\dagger, \phi\right). \tag{3.17}$$

$D_\mu \phi_i^\dagger D^\mu \phi_i$ is the kinetic term where V is a potential term. Pick the potential term to be of the form

$$V\left(\phi^\dagger, \phi\right) = \frac{1}{4}\lambda\left(\phi^\dagger \phi - \frac{1}{2}v^2\right)^2. \tag{3.18}$$

λ and v are both constants with $\lambda > 0$. The vacuum for (3.18) is at $|\phi| = \frac{v}{\sqrt{2}}$. Using a global $SU(2)$ transformation the vacuum expectation can be shifted to be completely on the first component of ϕ. The field can the be made real by using a global $U(1)$ transformation. The result is

$$\langle 0\,|\phi|\,0\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} v \\ 0 \end{pmatrix}. \tag{3.19}$$

Expansion around this new vacuum will then give

$$\phi(x) = \frac{1}{\sqrt{2}}\begin{pmatrix} v + h(x) \\ 0 \end{pmatrix}. \tag{3.20}$$

This followed the process of symmetry breaking that was discussed in Section 3.3. As expected, symmetry has been broken and only a local $U(1)$ symmetry remains unbroken by this vacuum expectation value. Putting equation (3.20) back into the covariant derivative (3.16) it becomes

$$(D^\mu \phi)_i = \begin{pmatrix} \partial^\mu \frac{h}{\sqrt{2}} + \frac{i}{2}\left(g_2 W^{3\mu} - g_1 B^\mu\right)\left(v + \frac{h}{\sqrt{2}}\right) \\ \frac{ig_2}{2}\left(W^{1\mu} + iW^{2\mu}\right)\left(v + \frac{h}{\sqrt{2}}\right) \end{pmatrix}. \tag{3.21}$$

This can be used to write out the kinetic part of the Lagrangian which will be investigated in the Higgs coupling section. Instead here it is useful to investigate the Lagrangrian at the VEV. This produces a mass matrix from which the gauge field masses can be produced. The Lagrangian when using the VEV (3.19) can be worked into the form

$$\mathcal{L}_{\langle\phi\rangle} = -\frac{1}{8}\begin{pmatrix} v & 0 \end{pmatrix}\begin{pmatrix} g_2 W_\mu^3 - g_1 B_\mu & g_2\left(W_\mu^1 - iW_\mu^2\right) \\ g_2\left(W_\mu^1 + iW_\mu^2\right) & -g_2 W_\mu^3 - g_1 B_\mu \end{pmatrix}^2 \begin{pmatrix} v \\ 0 \end{pmatrix}. \tag{3.22}$$

Four new gauge bosons are now defined using the current four gauge fields that are in use. The new gauge fields are specially defined as the following linear combinations of the previous gauge fields

$$W_\mu^+ = \frac{1}{\sqrt{2}}\left(W_\mu^1 - iW_\mu^2\right) \tag{3.23}$$

$$W_\mu^- = \frac{1}{\sqrt{2}}\left(W_\mu^1 + iW_\mu^2\right) \tag{3.24}$$

$$Z_\mu = c_w W_\mu^3 - s_w B_\mu \tag{3.25}$$

$$A_\mu = s_w W_\mu^3 + c_w B_\mu. \tag{3.26}$$

Here the following definitions have been used

$$c_w = \frac{g_1}{\sqrt{g_1^2 + g_2^2}} \tag{3.27}$$

$$s_w = \frac{g_2}{\sqrt{g_1^2 + g_2^2}}. \tag{3.28}$$

To bring these new gauge bosons into the Lagrangian for the VEV they need to be shifted so that the previous gauge fields are the subjects of the equations. This can be done simply using simultaneous equations which result in

$$W_\mu^1 = \frac{1}{\sqrt{2}}\left(W_\mu^+ + W_\mu^-\right) \tag{3.29}$$

$$W_\mu^2 = \frac{i}{\sqrt{2}}\left(W_\mu^+ - W_\mu^-\right) \tag{3.30}$$

$$W_\mu^3 = c_w Z_\mu + s_w A_\mu \tag{3.31}$$

$$B_\mu = -s_w Z_\mu + c_w A_\mu. \tag{3.32}$$

These can be substituted into (3.17) which then results in the following which allows the masses of these new gauge fields to be defined as

$$\begin{aligned} \mathcal{L}_{\langle\phi\rangle} &= -\frac{1}{8}g_2^2 \begin{pmatrix} v & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{c_w}Z_\mu & \sqrt{2}W_\mu^+ \\ \sqrt{2}W_\mu^- & X \end{pmatrix}^2 \begin{pmatrix} v \\ 0 \end{pmatrix} \\ &= -M_W^2 W^{+\mu}W^{-\mu} - \frac{1}{2}M_Z^2 Z^\mu Z_\mu. \end{aligned} \tag{3.33}$$

X is placed in the matrix since this term is always multiplied by zero and is therefore unimportant in listing. Now (3.33) has directly shown the masses for the gauge fields as

$$M_W = \frac{g_2 v}{2} \tag{3.34}$$

$$M_Z = \frac{g_2 v}{2c_w}. \tag{3.35}$$

This has shown that through symmetry breaking the $W_\mu^\pm$ and Z_μ field have gained mass. A_μ has however remained a massless field. The $W_\mu^\pm$ and Z_μ fields are the massive gauge bosons that transmit the weak force and A_μ is the massless photon which transmits the electromagnetic force. What has been demonstrated here is that at very high energies (the electroweak scale is $\sim$ 100 GeV [47]) above energies where symmetry breaking occurs there exist four photon like mass gauge fields $\left(W_\mu^1, W_\mu^2, W_\mu^2, B_\mu\right)$ and a complex Higgs scalar field ϕ. The electromagnetic and weak forces are unified at these energies and the theory is called the electroweak theory. As energies are lowered and symmetry breaking occurs, the theory changes drastically with a new set of gauge fields $\left(W_\mu^+, W_\mu^-, Z_\mu^0, A_\mu\right)$ coming into the picture and the electroweak force breaking into two separate forces: the weak force and the electromagnetic force.

3.4.2 Gauge bosons and the Higgs field

To see how the gauge bosons in the theory interact with the Higgs field the full Lagrangian needs to be written out. First consider again the potential (3.18) with the field in (3.20) and perform the substitution and expansion to get

$$V\left(\phi^\dagger, \phi\right) = \frac{1}{4}\lambda v^2 h^2 + \frac{1}{4}\lambda v h^3 + \frac{1}{16}\lambda h^4. \tag{3.36}$$

The kinetic term is just of the standard form of $-\frac{1}{2}\partial_\mu h \partial^\mu h$. The kinetic terms for the original four gauge fields $\left(W_\mu^1, W_\mu^2, W_\mu^2, B_\mu\right)$ can be shown to be written as

$$\mathcal{L}_{kin} = -\frac{1}{4}F_a^{\mu\nu}F_{\mu\nu}^a - \frac{1}{4}B^{\mu\nu}B_{\mu\nu} \tag{3.37}$$

where the following field tensor definitions [48] exist

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu, \tag{3.38}$$

$$F_{\mu\nu}^1 = \partial_\mu W_\nu^1 - \partial_\nu W_\mu^1 + g_2\left(W_\mu^2 W_\nu^3 - W_\nu^2 W_\mu^3\right), \tag{3.39}$$

$$F_{\mu\nu}^2 = \partial_\mu W_\nu^2 - \partial_\nu W_\mu^2 + g_2\left(W_\mu^3 W_\nu^1 - W_\nu^3 W_\mu^1\right), \tag{3.40}$$

$$F_{\mu\nu}^3 = \partial_\mu W_\nu^3 - \partial_\nu W_\mu^3 + g_2\left(W_\mu^1 W_\nu^2 - W_\nu^1 W_\mu^2\right). \tag{3.41}$$

It should also be noted that the photon A_μ and the mediator Z_μ are actually orthogonal and normalised linear combinations of B_μ and W^3_μ. These combinations can be written

$$A_\mu = \cos\theta_W B_\mu + \sin\theta_W W^3_\mu \tag{3.42}$$

$$Z_\mu = -\sin\theta_W B_\mu + \cos\theta_W B_\mu + \cos\theta_W W^3_\mu \tag{3.43}$$

which is used to define the mixing angle θ_W [49]. This information along with the transformations from the original gauge fields to the post symmetry breaking gauge fields in (3.29) - (3.32) can be used to build the Lagrangian that is under investigation. The end result after substitution, expansion and simplification is

$$\begin{aligned}
\mathcal{L}_{eff} = \; & -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} - \frac{1}{4}Z^{\mu\nu}Z_{\mu\nu} \qquad . \\
& - D^{\dagger\mu}W^{-\nu}D_\mu W^+_\nu + D^{\dagger\mu}W^{-\nu}D_\nu W^+_\mu \\
& - ie\left(F^{\mu\nu} + \cot\theta_W Z^{\mu\nu}\right)W^+_\mu W^-_\nu \\
& - \frac{1}{2}\left(\frac{e^2}{\sin^2\theta_W}\right)\left(W^{+\mu}W^-_\mu W^{+\nu}W^-_\nu - W^{+\mu}W^+_\mu W^{-\nu}W^-_\nu\right) \\
& - \left(M^2_W W^{+\mu}W^-_\mu + \frac{1}{2}M^2_Z Z^\mu Z_\mu\right)\left(1 + \frac{h}{v}\right)^2 \\
& - \frac{1}{2}\partial^\mu h \partial_\mu h - \frac{1}{2}m^2_h h^2 - \frac{1}{2}\frac{m^2_h}{v}h^3 - \frac{1}{8}\frac{m^2_h}{v^2}h^4.
\end{aligned} \tag{3.44}$$

Here the following definitions have been used

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu, \tag{3.45}$$

$$Z_{\mu\nu} = \partial_\mu Z_\nu - \partial_\nu Z_\mu, \tag{3.46}$$

$$D_\mu = \partial_\mu - ie\left(A_\mu + \cot\theta_W Z_\mu\right). \tag{3.47}$$

The Lagrangian in (3.44) shows the interactions between the gauge bosons and the Higgs boson. The results show that A_μ does not couple to the Higgs but the $W^\pm$ and Z^0 gauge bosons do. These couplings to the Higgs reveal production mechanisms of the Higgs such as VBF and VH as discussed earlier [50].

3.4.3 Leptons in the Standard Model

This section discusses the role of leptons in the Standard Model. The different generations of leptons all act very similarly. This similarity allows for only a single generation's theory to be formulated after which the theory can be generalised to the remaining two generations. Here the first generation of leptons will be discussed.

The lepton sector still falls under the $SU(2)_L \times U(1)_Y$ symmetry of the Standard Model. Recall Table 4 and note that in this section the $SU(2)_L$ doublet L and the singlet $\bar{e}$. For $SU(2)_L \times U(1)_Y$, L has the representation $\left(2, -\frac{1}{2}\right)$ and $\bar{e}$ has the representation $(1, 1)$. The co-variant derivatives can be written as

$$\left(D_\mu L\right)_i = \partial_\mu L_i - i \left[g_2 W_\mu^a \left(T^a\right)_{ij} L_j + g_1 B_\mu Y_L L_i\right] \tag{3.48}$$

$$D_\mu \bar{e} = \partial_\mu \bar{e} - i g_1 B_\mu Y_{\bar{e}} \bar{e} \ . \tag{3.49}$$

The two unseen generators are defined as

$$Y_L = -\frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{3.50}$$

$$Y_{\bar{e}} = (1) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} . \tag{3.51}$$

The kinetic part of the Lagrangian can then be written out as

$$\mathcal{L}_{kin} = i L^{\dagger i} \bar{\sigma}^\mu \left(D_\mu L\right)_i + i \bar{e}^\dagger \bar{\sigma}^\mu D_\mu \bar{e}. \tag{3.52}$$

The Lagrangian here has no mass terms for L and $\bar{e}$. Since electrons have mass it is obvious that mass terms need to be included in the Lagrangian. Directly adding in mass terms always leads to the symmetry breaking and the loss of Lorentz invariance which is not acceptable. A different approach must be used and by instead introducing a Yukawa term this problem can be solved. The Yukawa term used is

$$\mathcal{L}_{\text{Yuk}} = -y \epsilon^{ij} \phi_i \left(L_j \bar{e}\right) + h.c \tag{3.53}$$

which is an interaction term. Here y is the Yukawa coupling constant, ϵ^{ij} is the totally antisymmetric tensor and $h.c$ is the hermitian conjugate of the first term. Now the symmetry can be broken using the same symmetry breaking process used in the electroweak theory section of this thesis. Choose $\phi_2 = 0$ and $\phi_1 = \frac{1}{\sqrt{2}} \left(v + h\left(x\right)\right)$.

Substituting into (3.53) produces the result

$$
\begin{aligned}
\mathcal{L}_{\text{Yuk}} &= -y\left(\phi_1 L_2 - \phi_2 L_1\right)\bar{e} + h.c && (3.54)\\
&= -\frac{1}{\sqrt{2}}y\left(v+h\right)L_2\bar{e} + h.c && (3.55)\\
&= -\frac{1}{\sqrt{2}}y\left(v+h\right)e\bar{e} - \frac{1}{\sqrt{2}}y\left(v+h\right)\bar{e}^{\dagger}e^{\dagger} && (3.56)\\
&= -\frac{1}{\sqrt{2}}yv\bar{\varepsilon}\varepsilon - \frac{1}{\sqrt{2}}yh\bar{\varepsilon}\varepsilon. && (3.57)
\end{aligned}
$$

Here the following definitions exist

$$
\varepsilon = \begin{pmatrix} e \\ i\sigma^2\bar{e}^{\dagger} \end{pmatrix} \qquad (3.58)
$$

$$
\bar{\varepsilon} = \varepsilon^{\dagger}\gamma^0. \qquad (3.59)
$$

This results in the electron and the anti-electron now having mass terms $\left(\frac{yv}{\sqrt{2}}\right)$. To see how the leptons interact with the gauge bosons of $SU(2)_L \times U(1)_Y$ the covariant derivative in (3.48) needs to be expanded. The bracketed term in (3.48) can be separated into two important parts. The first part can be written out and simplified, using the fact that the generators are the Pauli matrices, as

$$
\begin{aligned}
g_2\left(W_\mu^1 T^1 + W_\mu^2 T^2\right) &= \frac{g_2}{2}\begin{pmatrix} 0 & W_\mu^1 - iW_\mu^2 \\ W_\mu^1 + iW_\mu^2 & 0 \end{pmatrix} && (3.60)\\
&= \frac{g_2}{\sqrt{2}}\begin{pmatrix} 0 & W_\mu^+ \\ W_\mu^- & 0 \end{pmatrix}.
\end{aligned}
$$

The second part can be simplified in the same way and the result is

$$
\begin{aligned}
g_2 W_\mu^3 T^3 + g_1 B_\mu Y &= \frac{e}{s_w}\left(s_w A_\mu + c_w Z_\mu\right)T^3 + \frac{e}{c_w}\left(c_w A_\mu - s_w Z_\mu\right)Y && (3.61)\\
&= e\left(A_\mu + \cot\theta_w Z_\mu\right)T^3 + e\left(A_\mu - \tan\theta_w Z_\mu\right)Y\\
&= e\left(T^3 + Y\right)A_\mu + e\left(\cot\theta_w T^3 - \tan\theta_w Y\right)Z_\mu\\
&= eQA_\mu + e\left(\cot\theta_w T^3 - \tan\theta_w Y\right)Z_\mu.
\end{aligned}
$$

Here $Q = T^3 + Y$ is the generator of electrical charge. These covariant derivatives can then be used to write out the full Lagrangian. Writing out the full Lagrangian is not important here and all that needs to be noted is that these covariant derivatives will show how the leptons interact with the gauge bosons. The result is that the electron neutrino, the electron and the anti-electron all interact with the new gauge bosons $\left(W_\mu^+, W_\mu^-, Z_\mu^0, A_\mu\right)$. In these interactions the Z_μ^0 and A_μ gauge bosons interact neutrally with the leptons and the lepton charge remains unchanged. The W_μ^+ gauge boson acts as the $SU(2)$ raising operator and raises the charge of leptons it interacts with. The W_μ^- gauge boson acts as the $SU(2)$ lowering operator and lowers the charge of leptons it interacts with. As stated earlier, only the first generation of leptons was discussed, but the process here can be repeated for the second and third generation of leptons giving a similar result for each of the three generations.

3.4.4 The Quark Sector

Just as was done with the lepton sector, the formulation of the theory for only one generation of the quarks needs to be investigated. The result can easily be extended using the same process to include all generations of quarks. First recall the quark fields from Table 4 which are Q, $\bar{u}$ and $\bar{d}$. The representations for these fields in $SU(3)_C \times SU(2)_L \times U(1)_Y$ are $\left(3, 2, \frac{1}{6}\right)$, $\left(\bar{3}, 1, -\frac{2}{3}\right)$ and $\left(\bar{3}, 1, -\frac{2}{3}\right)$ for Q, $\bar{u}$ and $\bar{d}$ respectively. The covariant derivatives can be written out as

$$(D_\mu Q)_{\alpha i} = \partial_\mu Q_{\alpha i} - ig_3 A_\mu^a \left(T_3^a\right)_\alpha^\beta Q_{\beta i} - ig_2 W_\mu^a \left(T_2^a\right)_i^j Q_{\beta i} - ig_1 \left(\frac{1}{6}\right) B_\mu Q_{\alpha i} \quad (3.62)$$

$$(D_\mu \bar{u})^\alpha = \partial_\mu \bar{u}^\alpha - ig_3 A_\mu^a \left(T_{\bar{3}}^a\right)_\beta^\alpha \bar{u}^\beta - ig_1 \left(-\frac{2}{3}\right) B_\mu \bar{u}^\alpha \quad (3.63)$$

$$\left(D_\mu \bar{d}\right)^\alpha = \partial_\mu \bar{d}^\alpha - ig_3 A_\mu^a \left(T_{\bar{3}}^a\right)_\beta^\alpha \bar{d}^\beta - ig_1 \left(\frac{1}{3}\right) B_\mu \bar{d}^\alpha. \quad (3.64)$$

Here the index i is an $SU(2)_L$ index and α is an index for $SU(3)_C$. The values α can take are either 1, 2 or 3 which represent the quark colours of red, green and blue. A_μ^a is the gluon field and is made distinguishable from the photon by the upper index. Just as was done for the lepton sector a kinetic and potential term can be written out for the Lagrangian but once again there are no mass terms present. Since quarks are known to be massive this cannot be left uncorrected, but adding in the mass terms directly presents the same problems as before for the lepton sector. The necessary approach requires the use of a Yukawa term. The Yukawa term for the quark sector is as follows

$$\mathcal{L}_{\text{Yuk}} = -y' \epsilon^{ij} \phi_i Q_{\alpha i} \bar{d}^\alpha - y'' \phi^{\dagger i} Q_{\alpha i} \bar{u}^\alpha + h.c. \quad (3.65)$$

Taking the Yukawa term through the process of symmetry breaking gives

$$\mathcal{L}_{\text{Yuk}} = -\frac{1}{\sqrt{2}} y' (v+h) \left(d_\alpha \bar{d}^\alpha + \bar{d}_\alpha^\dagger d^{\dagger\alpha} \right) - \frac{1}{\sqrt{2}} y'' (v+h)(v+h) \left(u_\alpha \bar{u}^\alpha + \bar{u}_\alpha^\dagger u^{\dagger\alpha} \right) \tag{3.66}$$

$$= -\frac{1}{2} y' (v+h) \bar{\mathcal{D}}^\alpha \mathcal{D}_\alpha - \frac{1}{\sqrt{2}} y'' (v+h) \bar{\mathcal{U}}^\alpha \mathcal{U}_\alpha \tag{3.67}$$

with

$$\mathcal{D}_\alpha = \begin{pmatrix} d_\alpha \\ \bar{i}\sigma^2 d_\alpha^* \end{pmatrix} \tag{3.68}$$

$$\mathcal{U}_\alpha = \begin{pmatrix} u_\alpha \\ \bar{i}\sigma^2 u_\alpha^* \end{pmatrix}. \tag{3.69}$$

Expanding the Yukawa term in (3.67) will produce mass terms for the quarks. The up quark gets a mass term $\frac{y''v}{\sqrt{2}}$ and the down quark gets a mass term of $\frac{y'v}{\sqrt{2}}$. The covariant derivatives can also be expanded which will show how the quarks interact with gauge bosons. The same parts of the covariant derivatives need to be looked at as in the lepton sector. These can be written as

$$g_2 \left(W_\mu^1 T^1 + W_\mu^2 T^2 \right) = \frac{g_2}{\sqrt{2}} \begin{pmatrix} 0 & W_\mu^+ \\ W_\mu^- & 0 \end{pmatrix}. \tag{3.70}$$

The second part of the covariant derivative has the form

$$g_2 W_\mu^3 T^3 + g_1 B_\mu Y = eQ)A_\mu + \frac{e}{s_w c_w} \left(T^3 - s_w^2 Q \right) Z_\mu. \quad . \tag{3.71}$$

These give a result where the quark doublet Q behaves in the same way that the lepton doublet L behaves when interacting with the gauge bosons $W^\pm$, A_μ and Z_μ. The quarks also interact with another set of gauge fields which are known as gluons. There are eight gluons associated with eight generators of $SU(3)$ which are called the Gell-Mann matrices [51]. Gluons each carry a colour and an anti colour and they can be written out as

$$g_\beta^\alpha = \begin{pmatrix} r\bar{r} & r\bar{g} & r\bar{b} \\ g\bar{r} & g\bar{g} & g\bar{b} \\ b\bar{r} & b\bar{g} & b\bar{b} \end{pmatrix}. \tag{3.72}$$

The upper index on g_β^α denotes the column and is called the anti-colour index. The lower index denotes the row and is called the colour index. Gluons will only interact with quarks that have the corresponding colour index to the anti-colour index and vice versa for anti-quarks. As as example, a gluon with an anti-red anti-colour will only interact with a red coloured quark.

These particle interactions described by the Standard Model provide a theoretical foundation for the studies presented in this thesis. The principles explained here are used as the basis for developing many algorithms used to simulate and reconstruct particle interactions at ATLAS.

4 ATLAS Monte Carlo

Monte Carlo simulations (MC) are important to the offline analysis at ATLAS in trying to predict and understand actual data. It is used in background estimation, signal estimation and helps test an analysis before applying it to collision data since it is better known what to expect. MC plays an important part in the analysis presented in this thesis and to properly understand the analysis it is important to understand how MC is produced at ATLAS. MC production is broken down into a few steps which are discussed in turn. In general the MC simulation chain is split into three steps. These are generation of events and the immediate decays, simulation of the detector and physics interactions and digitization of the energy deposited in the detectors. Whilst these are covered in more detail in later sections, prior knowledge of these steps is necessary since they are mentioned in the following ATLAS offline software section.

4.1 ATLAS offline software

The offline software used and developed by ATLAS is contained within a framework called Athena. Athena uses Python as a scripting language to configure algorithms and objects designed in C++. This allows for Athena to be compiled once and then new configurations can be applied without recompiling through the interpreted Python language. These configurations are done using what are called ‘job option’ files. These are Python scripts that can define things such as what algorithms should be run and how they should be configured. The C++ backend also ensures the speed necessary for many complex algorithms to run in a timely manner. Athena was based on the Gaudi framework [52, 53] originally designed for LHCb. The infrastructure uses the CLHEP common libraries [54] which include utility classes designed specifically for use in high-energy physics. Each Athena release contains all the ATLAS simulation software but will differ in certain areas of functionality. The event generation, simulation and digitization steps in MC production are all run using the Athena framework.

4.2 Event Generation

The event generators model the particles involved in specific processes and the immediate decays in these processes. Event generation is run within the Athena framework and the results are fed into the simulation step that follows. The event generators are written and maintained by development teams external to ATLAS but Athena has a number of interface packages which allow it to drive the use of these generators. The Athena event manager generates run numbers and event numbers which are passed to the event generator which then produces an event. ATLAS uses multiple event generators and it is therefore important that all their output be mapped to a common format so that any algorithms are independent of which generator was used. The format used is the HepMC event record format [55]. The HepMC format stores particles that are considered stable within the detector (muons are included as they live long enough to pass through the detector) to use as input to the simulation step. Other particles considered as unstable in the detector are used for physics studies and diagnostics. Which particles are produced and kept depends on the generator and process being modelled. The particles are stored in a connected tree which mimics the structure of a classical decay chain. This entire connected tree is stored as the truth information in the MC datasets. Truth information gives the analyser the ability to investigate exactly what happened in the process which can be useful for checking things such as the accuracy of reconstruction algorithms. HepMC is not a standard implemented among all generators however and so Athena has to handle conversion of all different data formats into the HepMC format.

The list of generators used by ATLAS includes Pythia [56], PythiaB [57], Pythia 8 [58], Herwig [59], Herwig++ [60], Sherpa [61], Hijing [62], Alpgen [63], MC@NLO [64], PowHeg [65], AcerMC [66], Tauola [67], Photos [68], EvtGen [69] and ISAJET [70]. These generators fall into two categories which are general purpose generators and specialised generators which will be discussed along with the generators that are relevant to this thesis.

4.3 General Purpose Generators

The general purpose generators are used to produce complete events which can be simulated from a number of different initial states (e.g the proton-proton initial state which is relevant to this thesis). The general purpose generators include Pythia, PythiaB, Pythia8, Herwig, Herwig++, Sherpa and Hijing. Those relevant to this thesis include Pythia and Herwig.

4.3.1 Pythia

Pythia was developed in FORTRAN and has been extensively used and validated at multiple collider experiments. The event generation done by Pythia starts with a hard process calculated to lowest order in QCD. Additional QCD and QED radiation is added in a particle shower approximation. This shower approximation is most accurate for radiation emitted at small angles. A model for hard and soft scattering processes in a single event is used to then simulate underlying activity in minimum bias events. Here hard scattering processes are clearly perturbative processes involving high p_T where soft processes are non-perturbative processes and low p_T [71]. Minimum bias events are events that are accepted by the trigger with no energy requirement imposed. Underlying events include all particles from a single particle collision excluding the hard process.

Pythia has two models for QCD radiation, but ATLAS uses the showering model mentioned previously which is expected to agree more with theory on QCD showers [56]. The shower model produces higher jet activity. At the end of a simulated shower Pythia then uses a phenomenological model to group quarks and gluons into hadrons which is known as hadronization. Pythia contains many built in processes that can be modelled this way and can have new ones added by modifying the code. It is also often used in conjunction with other specialised generators when the other generator is expected to produce a better description of certain processes or final states.

Pythia has recently been re-implemented in C++. ATLAS is currently shifting to the use of this C++ version which is called Pythia 8.

4.3.2 Herwig

Herwig is another flexible generator focusing particularly on the detailed simulation of QCD parton showers. It uses a similar showering model to that of Pythia from the basic way the interaction is modelled to the hadronization process. In particular ATLAS uses Herwig with the Jimmy [72] implementation of the underlying event. Herwig can also be used with other specialised generators to improve the physical description of the final states.

Herwig is implemented in FORTRAN but has recently been rewritten in C++. ATLAS is currently shifting to the use of the C++ version which is called Herwig++.

4.4 Specialised Generators

The specialised generators do not produce complete events to be passed to the simulation step. They are used together with the general purpose generators to produce these full events. These generators are used when they can produce better physical descriptions of the final states for a process than the general purpose generators can. Most of the specialised generators follow the Les Houches Accords [73, 74] and are known as Les Houches generators. This means that they run alone and produce a file containing four-vectors describing different particles in the format agreed upon in the Les Houches Accords. These files can then be read in by Athena and prepared for use in the general purpose generator to fully model the process. The list of specialised generators used by ATLAS include ISAJET, Photos, Tauola, EvtGen, Alpgen, MC@NLO, PowHeg and AcerMC. The ones that are relevant to discuss are Alpgen, MC@NLO and PowHeg since they are used extensively in this thesis.

4.4.1 Alpgen

Alpgen is used to produce final states with several well separated hadronic jets. Herwig or Pythia are used to perform the hadronization and produce the final state QCD radiation when Alpgen is used. Alpgen implements a fixed order QCD matrix element method to simulate processes and is used for processes where this method is expected to give better results than the shower models used by Herwig and Pythia. Particularly, Alpgen performs well for final states containing a $W^{\pm}$ or a Z and many jets.

4.4.2 MC@NLO

MC@NLO uses fundamental processes evaluated at next-to-leading order in QCD. Once the process is modelled by MC@NLO the output is passed to Herwig to complete the full event. MC@NLO includes one loop corrections with a side effect that the events appear with negative and positive weights which needs to be taken into account. It is better at modelling the transverse momentum distribution of top quarks than Herwig and Pythia. For this reason MC@NLO is mostly used to generate top events and in the production of $W^{\pm}$ and Z events.

4.4.3 PowHeg

PowHeg works in the same way that MC@NLO does, using next-to-leading order in QCD evaluations [65]. PowHeg can be used with both Herwig and Pythia and is also able to produce events with only positive weights which is easier to handle for Athena. It is used to model the same type of processes that MC@NLO does.

4.5 Simulation

The simulation step in the production of MC involves simulating the detector and physical interactions that occur within the detector. In order to do this it is necessary to build a virtual representation of the ATLAS detector before actually running the simulation. The ATLAS detector simulation and the core simulation are discussed.

4.5.1 ATLAS Detector Simulation

The ATLAS detector is a highly complex piece of machinery, so modelling it accurately requires a detailed description. This virtual description is done using the GeoModel toolkit [75]. The GeoModel toolkit provides functionality that allows for descriptions of complex detector systems to be created with minimal memory usage. GeoModel allows for basic volumes to be created, rotated and positioned. Volumes can be placed within one another and physical and logical properties can be defined. It is also possible to model different materials for the volumes.

The actual simulation is done using GEANT4 [76] and the GeoModel description needs to be translated into the GEANT4 format before the simulation can be run. Since GeoModel contains geometry features very similar to GEANT4 this makes it a simple procedure which is another reason the GeoModel toolkit is used. When creating the ATLAS geometry all volume overlaps and touching surfaces require special handling. Overlaps of more than 1 picometer and touching surfaces mean particle tracks can get stuck between the two surfaces. GEANT4 then does not know which volume the tracks belong to and the event is discarded. This is handled by creating small gaps between all the volumes at a cost of an extra step to move each particle through these gap regions.

The simulation, digitization and reconstruction steps all need to be able to use the same geometry. Layouts of the detector need to be maintained in such a way that they are not specific to any of these steps. These layouts are stored in a geometry database which can be accessed by each step. For each layout all parts of the detector are modelled separately which allows for job option configurations in Athena to select which parts of the detector it wants to be active in the simulations. This allows for different geometries of the detector to be tested easily. There are also alternate non-standard layouts of the detector that exist which can be used in the MC generation. Such layouts include models of future upgrades to the detector or layouts that were used to do cosmic ray data-taking before all sub-detector systems were complete.

4.5.2 Core Simulation

The GEANT4 toolkit models physics and infrastructure for particle transportation through materials in the detector. The GEANT4 accepts the ATLAS geometry from the GeoModel description which can then be specially configured. For example, range event selection criteria can be imposed or the magnetic field can be turned on or off before the simulation. These additional options are particularly useful for testing and debugging purposes. The simulation first goes through an initialisation procedure where a number of steps need to be performed to prepare the simulation environment such as loading external modules, locking in job options, configuring the event generation and loading along with configuring the detector geometry. Once initialised the simulation is run using a simulation event loop and the output is stored in HepMC format and fed to the reconstruction and digitization steps. The simulation process also adds extra information to the truth record created during the event generation.

4.6 Digitisation and Reconstruction

The digitisation step converts any hits produced in the core simulation into detector responses. These responses are called digits and are typically produced when a voltage or a current on a readout channel rises above a certain level over a given time. Each sub-detector has specific modelling software that describes charge collection in the sub-detector system. This software can be individually configured to drive the modelling in different ways as necessary. The tuning of the software systems are constantly updated to deal with changes in the detector so that the modelling is accurate to current run conditions. For example, if certain channels in ATLAS are not working for certain runs the digitisation software needs to take this into account too.

Up until this point only single hard scattering processes are generated and simulated. However, in practice within the detector, multiple proton-proton collisions occur per bunch crossing. Along with these interactions there are other interactions that can occur at the same time including cavern background and beam halo effects. These interactions that were treated individually before are now overlaid with each other before the total detector response is calculated. The multiple responses can interfere with each other and this effect which is seen in practice and simulated here is known as pileup.

Once the hits have been turned into digits that represent the detector responses this data is then used in the reconstruction process. Reconstruction uses the detector responses as input and runs a number of algorithms that attempt to identify the particles along with their properties that caused the given responses. The output is stored using ROOT file formats which are the native file format for the ROOT data analysis framework [77]. ROOT is the primary data analysis framework used by the ATLAS collaboration in their multiple analyses and has also been used extensively throughout this thesis.

4.7 Fast Simulations with ATLFast

The complicated detector geometry and physics descriptions used in GEANT4 for ATLAS cause the simulation step of MC generation to be a slow process. So much so that it would be unrealistic to produce the number of events to give good enough statistics for many physics studies. The design of the ATLFast simulation software allows it to perform these simulations at a very fast rate but at the cost of a less accurate description of the results. ATLFast comes in two versions which are ATLFast-I and ATLFast-II.

ATLFast-I is the lesser accurate simulation method. The detail detector simulation is no longer used and is instead replaced with parameterisations of the subdetectors and different reconstruction effects. The simulation time is reduced by a factor of ∼1000 using ATLFast-I.

ATLFast-II works similarly to ATLFast-I except that it can use the full GEANT4 simulation for specific sub-detectors if desired. ATLFast-II by default uses fast tracking simulation for the muon system and inner detector and uses fast calorimeter simulation for the calorimeters. Using this default configuration ATLFast-II sees a speed increase of a factor of ∼100 over the full simulation [78]. The output includes all properties for the reconstructed objects which allows for the standard ATLAS reconstruction software to process the sample.

4.8 Naming Conventions

It is important for the analyser using a dataset to know some details of how it was produced. The naming conventions used for MC datasets allow for easy identification of what configurations were used to produce it. The name of a dataset can contain the generators used and what process was modelled, a number that identifies the dataset called the dataset ID, the format of the dataset, the event generation configuration (evgen), the simulation configuration (sim) and the reconstruction configuration (rec).

For the following discussion, consider the dataset

valid1.106052.PythiaZtautau.recon.AOD.e1950_s1487_r3478.

The generator name and process that are modelled are a single string containing a brief description of what the dataset contains. In the example, this is "PythiaZtautau" which can be easily interpreted as the sample contains data for a process where a Z boson decays to two tau leptons which was modelled by the Pythia event generator. This particular dataset is given the dataset ID 106052. This dataset ID is always the same for the PythiaZtautau dataset regardless of the evgen, sim and rec configurations. The format of this dataset is AOD (Analysis Object Data). AOD is one of the formats used by ATLAS which describes how the data should be stored in root files. Other formats include ESD (Event Summary Data) and Ntuple. The evgen, sim and rec configurations are combined into one string separated but underscores. The evgen configuration is prefixed by "e", sim by "s" and rec by "r". In this example the full string is "e1950_s1487_r3478". These configurations are not specific to the chosen generator but can describe configurations for many generators. After production, samples with different sim or rec configurations can be merged together which leads to there being multiple sim tags or multiple rec tags. A merging of two different rec configurations, for example, would be "e1950_s1487_r3478_r3567". Finally in the example there is also the tag "valid1" denoting the sample is for validation purposes, and the tag "recon" denoting that it the dataset contains reconstructed data. Not all MC datasets will use this format but the general discussion here allows for a good understanding of the datasets that are used within this thesis.

Finally it is important to note that samples produced with ATLFast can have a different configuration tag. The configuration tag still contains the evgen configuration prefixed by an "e" but all following tags are prefixed by an "a". An example tag could be "e1950_a1145_a1189".

5 Tau Validation

This section goes into detail about what Tau Validation is and how the validation process is performed. A discussion on contributions made whilst in charge of the Tau Validation package is presented before finalising the section by looking at the future of Tau Validation at ATLAS.

5.1 What is Tau Validation?

The process behind how MC simulations are produced at ATLAS constantly changes. Before any new releases of this MC production software can be put into production they need to be checked for integrity. This integrity checking is performed by the Physics Validation groups at ATLAS. The Physics Validation group is sub-divided to share the load of this validation with each group focussing on validating certain areas of MC integrity. The Tau Validation group is one such sub-group and it focusses on validating hadronic τ lepton reconstruction in the MC samples. The changes that occur in the MC production typically include updates to the actual generator software, improvements in the configuration of simulation and reconstruction phases of MC generation, development changes in the Athena analysis framework or changes that allow for testing the upgrade geometries planned for the future of ATLAS. Samples with these new changes are compared to baseline samples to check that they show reasonable results given what has been previously validated against theory and experiment.

5.2 Physics Validation

Physics Validation groups have to check the integrity of new releases before they can be put into use by ATLAS collaborators. The procedure for the validation of these new releases follows a specific step-by-step process. First the request for a new cache, or configuration, is made with all details on production being specified. A specifically chosen set of samples is produced using the new release and configuration that is to be tested. These samples contain different processes produced with multiple physics objects and generators. Once these samples are ready, information about the validation tasks is circulated to the validation groups which, for each task, includes in particular the “evgen”, “rec” and “sim” configuration (collectively called a configuration from now on) along with the required release of Athena for the configuration. For each task this information is provided for a test configuration and one or more reference configurations.

For each task additional information on what is being tested and what deviations from the references are expected is provided. An example of a validation task description can be seen in Figure 10. The parenthesis next to the test and reference “rec” tags in the figure describe the required Athena release for the configuration. Links to the sample production pages are also given which provide a way to find the full “evgen_sim_rec” configuration tags for the reference and test samples. Each physics validation group runs an analysis on the test and reference samples producing histograms of variables which are important to the group for their extended analysis. For example the Tau Validation group is responsible for validating variables related to hadronic τ reconstruction. The histograms for the test and reference samples can be compared once produced which allows for any discrepancies to be spotted in the new test samples.

Task 1:
Validation of Truth Tracking with 17.2.11.10
This is an update of the validation done on 18/10/2013: a problem was found with pileup options set during the production of the last test sample (r4941), a new test was produced (r4965).
This setup runs the truth tracking on the pileup and the standard setup on the signal event. Differences expected in the number of tracks/vertices + additional effect likely in pileup, energy resolutions, efficiencies etc

Test : r4965 (17.2.11.10)
Reference: r4764 (17.2.11.1)

Figure 10: An example of a Physics Validation task description.

The tasks are labelled based on how well the test matches the reference sample excluding any expected differences. Major differences mean the task is labelled red, minor differences lead to the task being labelled yellow, and negligible differences mean the task is labelled green. The results of the validation are then presented and discussed in a group meeting. Upon peer agreement these results and their colour flags are recorded in a bookkeeping database. The developers can then reference these results and make the necessary corrections to the releases or configurations.

Evgen Tags	Sample Type	Sample Name
e850, e1127	$Z \rightarrow \tau\tau$	106052.PythiaZtautau
	$t\bar{t}$	105200.T1_McAtNlo_Jimmy
	$Z \rightarrow ee$	106046.PythiaZee_no_filter
	QCD	105015.J6_pythia_jetjet
e1574, e1900, e1934	$Z \rightarrow \tau\tau$	147808.PowhegPythia8_AU2CT10_Ztautau
	$t\bar{t}$	105200.McAtNloJimmy_CT10_ttbar_LeptonFilter
	$Z \rightarrow ee$	147806.PowhegPythia8_AU2CT10_Zee

Table 5: A summary of the samples used in Tau Validation.

5.3 Tau Validation Samples

As mentioned previously, a set of samples are produced to perform the validation for each validation task. Each validation group uses a select few of these samples to perform their validation tests. The specific set of samples produced are dependant on the given "evgen" tag that is being used for the configuration. The "evgen" tags with the associated samples used by the Tau Validation group are summarised in Table 5. For "evgen" tags "e850" and "e1127" Tau Validation uses four Monte Carlo samples which include a $Z \rightarrow \tau\tau$, $t\bar{t}$, $Z \rightarrow ee$ and QCD sample. When moving from "e1127" to "e1574" and beyond, a new set of samples were chosen that needed to be investigated by the validation groups. Specifically for the Tau Validation group for "e1574" and beyond only three samples need to be validated which include a $Z \rightarrow \tau\tau$, $t\bar{t}$ and $Z \rightarrow ee$ sample.

Within the Tau Validation package, $Z \rightarrow ee$ and QCD samples are considered as background samples whilst the $Z \rightarrow \tau\tau$ and $t\bar{t}$ considered as signal samples. These may be referred to as such in further discussions.

5.4 Tau Identification and Reconstruction

Tau Validation is coordinated by the Tau Working Group at ATLAS. The Tau Working Group is responsible for developing and maintaining algorithms used in identifying and reconstructing hadronically decaying tau leptons for the whole of ATLAS. Tau Validation needs to check that variables important to these algorithms are correct, along with making sure that the result of the algorithms are acceptable for different generator and simulation configurations.

As a result of their high mass, tau leptons are the only leptons that can decay into hadrons. As discussed in the introduction to this thesis this means that taus can be classified as having either a leptonic or hadronic decay. Taus have a proper decay length of 87 μm. In the ATLAS detector this means that tau leptons decay too soon to be directly detected and instead need to be identified by reconstructing them from their decay products. Leptonic decay products of taus cannot be distinguished from prompt muons and electrons so are not considered in tau identification. In the hadronic decay mode of the tau, only the hadrons are detected and used in the reconstruction and identification process. This means that only the visible part of the tau decay is reconstructed ($\tau_{\text{had-vis}}$). These hadrons appear in the detector reconstructed as a jet of particles and for this reason are hard to distinguish from jets initiated by quarks and gluons. The shower shape of the hadronic tau decay products is narrow due to the taus being Lorentz boosted and has an odd number of tracks [79]. Occasionally three prong taus can appear to have two prongs due to the third not being identified. These characteristics are useful in aiding identification of taus.

Each identified jet object is considered as a possible $\tau_{\text{had-vis}}$ decay. Information from the calorimeter clusters associated with the $\tau_{\text{had-vis}}$ candidate is used in the calculation of kinematic quantities. Tracks fulfilling a set of selection criteria are matched with these calorimeter clusters. Identification variables can then be calculated using the information from the tracks and the calorimeters. The identification variables can then be used with multivariate techniques (boosted decision tree (BDT) and projective likelihood (LLH) methods are available) or selection criteria based discriminants to reject the fake $\tau_{\text{had-vis}}$ candidates [80]. For electron discrimination a BDT discriminant is used whilst for muon discrimination a selection criteria based approach is currently used. Variables used in these algorithms and the algorithm results need to be validated by the Tau Validation group.

5.5 Pre-update Tau Validation

This section looks at how the Tau Validation package worked before necessary changes and updates were implemented. The overview given here only discusses the broad view of the variables that the tau validation investigates with the reason for this being that ≈800 variables are investigated. Further, a look at the way the variables were initially validated is discussed.

Hadronic taus can be classified in terms of the number of associated tracks. These categories of classification include "one-prong" for taus associated with a single track, "three-prong" for taus associated with three tracks and "multi-prong" which includes taus with more than one associated track. Objects that pass tau identification can be compared to Monte Carlo truth information to check if the object was correctly identified. Incorrectly identified objects are called tau fakes. The number of tau fakes and number of correctly identified taus are used to determine tau identification efficiencies. These can in turn can be used to determine scale factors used to account for the inefficiencies when analysing MC samples. This means that one-prong, three-prong and multi-prong taus can be further categorised as either true or fake taus. For each of these categories certain variables can then be investigated (e.g one-prong fake tau η distributions). In particular the set of core variables that are investigated include tau η, ϕ, transverse energy, number of tracks, LLH scores, BDT scores[2] and charge. There are a number of variables that are currently or have been previously used in the tau identification algorithm, in electron vetos and other related sub-detector information. The number of these extra variables is too long to list in this thesis and are also not of huge importance as will be discussed later.

Efficiencies are another aspect of the algorithms that are validated. It is important to verify that the algorithms perform with the same rate of success across multiple samples. The efficiencies investigated are categorised by different identification algorithms which include the BDT and LLH techniques. Different variable efficiencies can be investigated within these categories (e.g three-prong ϕ distribution for taus identified using the BDT with a loose signature). Generally efficiencies for η, ϕ, transverse energy and number of identified taus are always investigated. Other variables are also investigated but once again the number is too large to list here. All efficiencies and distributions were investigated for low p_T taus ($p_T < 20$ GeV) and for all other taus ($p_T > 20$ GeV).

The Tau Validation package provided rough automatic grading functionality to help narrow down which plots need to be investigated further. In the pre-updated version of the tau validation package two separate methods are used for grading

[2]BDT and LLH scores refer to a score produced by these two algorithms describing how confidently a physics object is identified.

efficiency distributions and general distribution plots. For general distributions the reference and test histograms were iterated through bin-by-bin and the maximum ratio between a test bin and the corresponding reference bin was found. Regions had been defined and depending which region the ratio fell into the histogram was graded red, yellow or green. Here red means the reference and test histograms are very different, yellow means they still need to be manually checked for validity and green means they should be acceptable. Generally if there are red distributions the source of the problem permeates through all similar variables to varying extents leading to most yellow distributions being related to those in the red category. This means that usually only the red distributions need to be investigated and simply looking at the names of the yellow distributions easily shows related problems without needing to investigate the plots. If there are no red distributions the yellow distributions will be investigated more thoroughly. Since for each task there are thousands of plots, methods used to filter what is likely to be problematic are useful. For efficiency distributions a similar methodology was devised but rather uses the average difference across all bins. Neither of the general distributions or efficiency plots were normalised. Overall, inaccurate automatic grading of these histograms existed. This is investigated further later in this section.

5.6 Tau Validation Updates

A number of changes were proposed to update the Tau Validation software. The suggested changes were split between those related to the physics behind the validations and changes to improve the implementation of the code base used for Tau Validation. This section discusses the various updates made and contributions made to the Tau Validation group.

5.6.1 Automatic Validation

In the initial state the Tau Validation package was in, it was decided that automation of the validation process was necessary. To understand this statement a discussion on what shall be called "manual" validation is required. Reasons for the automation can then be discussed along with the goals it would aim to fulfil and finally the implementation of the automation system. It is also useful to note that the software described here was designed to run on the CERN LXPLUS computing cluster which consists of machines accessible to CERN users through SSH for interactive work. The machines run SLC5 (Scientific Linux CERN 5) and SLC6 (Scientific Linux CERN 6).

Tau Validation is broken into two parts which have been arbitrarily and simply named “Part 1” and “Part 2”. Each physics validation typically includes a number of individual validation tasks. Each task describes a test configuration that needs to be checked against one or more reference configurations. For each configuration a number of samples needs to be validated. In Part 1 of the validation all the samples for all configurations are configured to be run over by the Tau Validation analysis package and the jobs are then submitted to the CERN computing GRID. The tau analysis package plots the necessary variables that need validation and produces output datasets which are used as input to Part 2 of the validation. Before the actual validation occurs, Part 2 initially requires a script named the AMI (ATLAS Metadata Interface) script to be run that pulls information from a specific database about all the samples used. The information provided by this database includes details such as information about the event generator, simulation and reconstruction configurations. This database is known as the AMI database. Next, for each reference sample in the task, Part 2 of the validation produces plots with the histograms from the references and the corresponding test sample histograms overlain. The automatic grading methods are run over the histograms being compared and the results are stored. An HTML report is automatically produced using the AMI information, the histograms and the results from the automatic grading. The report is uploaded to a web area where the results can then be viewed, interpreted and reported on. Examples of the web-page design can be seen in Figures 11, 12 and 12.

Manual validation requires most of this process to be configured by hand for each sample one by one. For Part 1 this includes producing GRID submission shell commands for each sample, for one configuration at a time, for each task. Each configuration is run separately since between submitting the jobs generally a new version of Athena needs to be setup and changes to the job option files may be required. Once all of the jobs are submitted to the CERN computing GRID and have successfully run the output is downloaded and Part 2 can then be executed. The AMI script is run one by one for each configuration. A script is then configured for each test-reference pair which will generate a script for each sample related to the configuration. Executing these scripts will run all the necessary steps to go from the datasets with individual histograms to generating the validation HTML report. The results can then be uploaded to a web location and reviewed.

Author 1 and 3 Seeded Tau (Pt 20GeV-10TeV)

Type of Tau Candidates	Number of Variables Classified as:			
	Green	Yellow	Red	All
Distributions ID	248	57	0	305
Distributions MATCHED	166	45	0	211
Distributions NEWCORE	25	3	2	30
Distributions Total	737	87	2	826
Efficiencies MATCHED	21	0	0	21
Efficiencies Total	183	0	0	183

Figure 11: An example of an individual validation task results summary page.

The Total Table

Title	Chi2	Histogram
Identification Variables: Single track (Fake) Impact Parameter Z0 Sin Theta Significance Lead Track	0.305749548922	calo_ID_1t_N_13
Identification Variables: Single track (Fake) Impact Parameter Significance Lead Track	0.422698581691	calo_ID_1t_N_12
Identification Variables: Single track (Fake) dR Max	0.771115732986	calo_ID_1t_N_11
Identification Variables: Single track (Fake) Is Tau bit	0.101120032918	calo_ID_1t_N_10
Electron Veto Variables: All (Truth Matched) Ratio of High to Low Threshold TRT Hits	0.0	calo_EleVeto_Al_M_05
Electron Veto Variables: All (Truth Matched) Hadronic Radius	12.3359425499	calo_EleVeto_Al_M_04
Electron Veto Variables: All (Truth Matched) Electron BDT Score	1.24996597359	calo_EleVeto_Al_M_06
Core Variables: Multi track (Fake) JetBDTSig Trans Score	0.171152496122	calo_Core_mt_N_13
Core Variables: Multi track (Fake) JetBDTSig Score	0.420229704444	calo_Core_mt_N_12
Core Variables: Multi track (Fake) Safe LLH Score	0.0	calo_Core_mt_N_11
Core Variables: Multi track (Fake) LLH Score	0.0	calo_Core_mt_N_10
Core Variables: Multi track (Fake) JetBDT Electron Score	0.193897089726	calo_Core_mt_N_15
Core Variables: Multi track (Fake) JetBDTBkg Trans Score	0.0	calo_Core_mt_N_14
Identification Variables: All centrality Fraction	5.97190072804	calo_ID_Al_Al_04
Identification Variables: All EM Fraction	0.198474403699	calo_ID_Al_Al_05

Figure 12: An example of an individual validation task results total distributions details page.

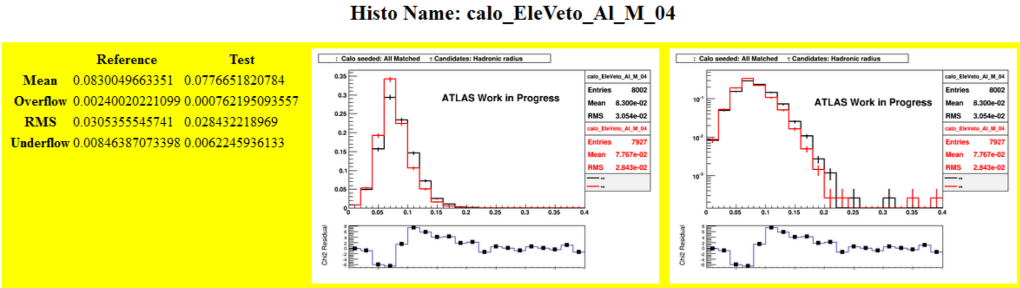


Figure 13: An example of an individual variable in a validation task results page.

For newcomers to Tau Validation this large number of steps is difficult to remember accurately, and the room for human error is large due to all the manual manipulation. The above process also requires a lot of repetitive small steps to be executed which is ideal for automation provided the use of external tools such as Athena and the DQ2 [81] dataset manager allow for it. Investigation into this showed that these tools could successfully be automated but problems can arise with Python bindings clashing when setting up Athena and DQ2 in the same environment. A solution to this problem is presented later in this section. Testing Bash scripting to automate the validation process was conducted but technical limitations and the growing scope of the automation lead to the current implementation using a combination of smaller Bash scripts being driven by a larger Python system. A more detailed explanation is discussed later.

The implemented automation for this system needed to be able to meet a few criteria to be considered a successful investment for Tau Validation. Criteria included reducing the time it takes to do a validation, making the validation process easier to learn so that newcomers to the group could produce useful results quicker, be highly configurable to cater for all the different environment configurations available and have a maintainable code base that would be possible to extend should the need arise. The following list of goals were set:

- A single configuration file should be used for Part 1 of the validation
- Job option file manipulation should be automated
- All dataset look ups using DQ2 must be automated
- All Athena release setups should be automated

- Part 1 must produce the necessary setup files for Part 2 to run without manual configuration

- Part 2 should use these setup files from Part 1 to run all of the Part 2 validation with a single script

- Uploading to the web area should be script controlled

- The code base should be easy to change and be well documented/commented

- The system should work on both SLC5 and SLC6 machines.

The list of goals were successfully implemented using a system created using a combination of a Python object oriented system and procedurally generated Bash scripts. The single configuration file allows for the job submission of the validation tasks to be simply described but has enough options to fine control the process. The Python framework reads in the configuration file creating data objects that describe the individual tasks and the individual configuration tags for the tests and references. These individual task objects are used to automatically format and create the Part 2 setup files and hold information needed for generating Bash scripts that handle the Athena installation and job submission. As stated earlier it is known that DQ2 installations can clash with Athena installations in the same environment. An option was added that finds and writes the dataset names needed for the tasks to disk. A new shell environment can then be started and these datasets are read from disk instead of using DQ2 which allows Athena to be installed and run without DQ2 to get the input dataset names. The Athena installation and job submission is handled using automatically generated Bash scripts that are called by the Python framework using the Python subprocess module [82]. The Part 2 setup files are read in when all the submitted jobs are successfully finished. The setup files include information that allow for the output datasets from these jobs to be automatically downloaded and formatted, the AMI database information to be acquired for the necessary tags and for the actual validation to be run. Running the validation produces the HTML report of the validation. The user would previously need to use information from this report to manually produce a summary in the form of a presentation given to the Physics Validation group. For user convenience additional functionality was added that produces a skeleton Latex Beamer [83] presentation containing the validation results. The user can use the framework to upload all results to the web area used by the Physics Validation groups to allowing them to be easily viewed in a web browser.

5.6.2 χ^2 Test

Another change that needed to be investigated was to look at implementing a more accurate automatic plot grading method. A more accurate method would improve how accurately and easily the user can perform the validation and warranted investigation. There are a number of approaches that can be used to prove that two distributions are different or consistent with one another such as the χ^2 test [84] and Kolmogorov-Smirnov test [85] which are the most widely accepted methods. The χ^2 test is used for testing for differences between binned data where the Kolmogorov-Smirnov test is used for testing continuous distributions. Binned data is used in Tau Validation and the χ^2 method was chosen. The ROOT analysis package provides a χ^2 test function for histograms but investigation showed that it would not work for all cases of this validation due to certain requirements imposed on input histograms such as not allowing empty bins. This lead to the other methods described earlier for testing being chosen initially. An investigation into manually implementing a version of the χ^2 test was investigated to overcome this problem.

Comparing whether or not two data distributions are similar can be restated in a more mathematical fashion as asking whether, to a certain level of significance, the null hypothesis that two data sets are results of the same population distribution function can be disproven [86]. If the null hypothesis is disproven a conclusion can be drawn that the two distributions are not of the same distribution function. Failing to disprove the null hypothesis can only draw the conclusion that it is likely or possible that the two distributions are from the same distribution function. That is, we can describe significance levels or thresholds that given a χ^2 value calculated from two distributions we can say whether the distributions show unacceptable differences. In Tau Validation this would allow problematic variables to be investigated. The theory of the χ^2 test is discussed in this section.

Consider two binned distributions where one distribution needs to be checked for consistency against the other. Call the distribution being checked for consistency the test distribution and the other distribution will be called the reference distribution. Let T_i be the number of entries in the ith bin of the test and R_i be the number of entries in the ith bin of the reference. The χ^2 distribution is then

$$\chi^2 = \sum_i \frac{(T_i - R_i)^2}{T_i + R_i} \tag{5.1}$$

The value that is used for the validation is however not just the χ^2 distribution here but its quotient with the number of degrees of freedom (χ^2/N_{df} where N_{df} is the number of degrees of freedom). The total N_{df} is initially the total number of bins in the histogram. The histograms are both normalised which adds a constraint and removes one degree of freedom. When summing over T_i and R_i, if $T_i = R_i = 0$, this bin is ignored and a degree of freedom is removed.

The χ^2 test helps identify problematic distributions in the test samples. A final decision as to whether the distributions are acceptable is however made by the user. Tau Validation produces ~1000 plots for each sample in a test-reference pair. Each test-reference pair contains three or four samples that need validation, each task contains at least one test-reference pair and in a given validation there can be multiple tasks. A lower limit of ~3000 plots are produced for any given validation. The χ^2 test is there to help the user filter through these thousands of plots. The χ^2 check was calibrated by running a number of tests and comparing what χ^2/N_{df} values were generally associated with acceptable and unacceptable distribution. Thresholds chosen for the χ^2 test are summarised in Table 6. As before, green indicates that there are negligible differences, yellow indicates that the distribution could be inconsistent and red requires further investigation. The red and yellow graded plots are still investigated where as the green plots are assumed as consistent. The χ^2 test is only used as a rough indication of which distributions must be investigated at this stage for Tau Validation so a strict investigation is not warranted provided the green plots can be trusted as consistent. The limits chosen fulfil the necessary requirement.

The implementation was done in Python using the χ^2 test C++ implementation

χ^2/N_{df} Range	Grading
$\chi^2/N_{df} < 10$	green
$10 <= \chi^2/N_{df} < 100$	yellow
$\chi^2/N_{df} >= 100$	red

Table 6: The χ^2/N_{df} ranges used in automatic distribution grading.

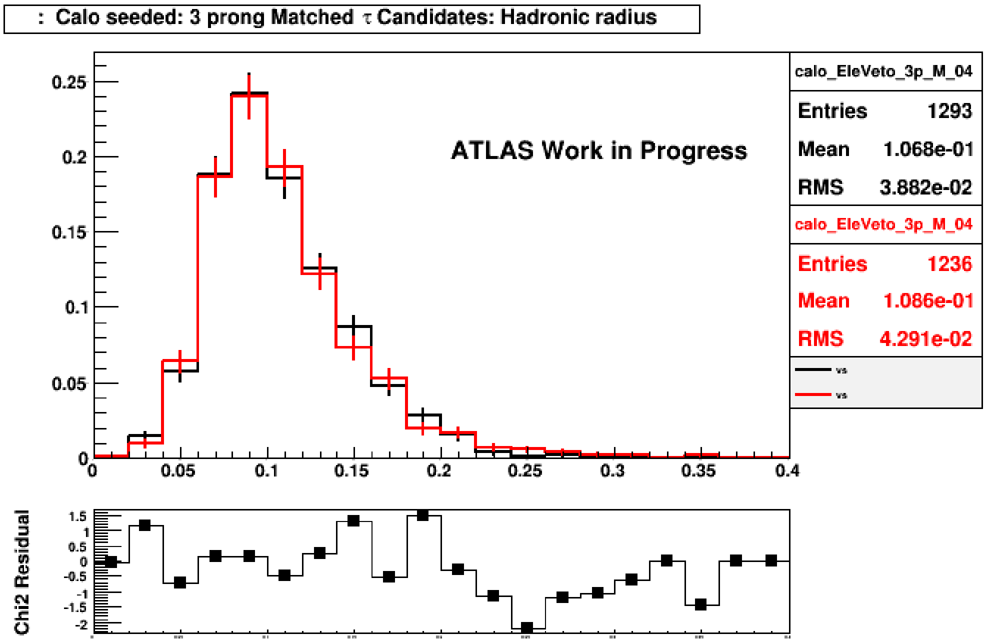
in *Numerical Recipes: The Art of Scientific Computing* [86] by William H. Press. Results of the χ^2 test can be discussed by looking at Figures 14, 15 and 16 which show distributions that were marked green, yellow and red respectively. It is not of great importance here to understand the plots as just the result of the χ^2 test needs to be discussed but the names of the distributions are given in the figure descriptions. It is also useful to note that the χ^2 residuals shown in this plot are calculated using the ROOT χ^2 functionality. A χ^2 residual here refers to the χ^2 value between the two distributions for a given bin. Figure 15 shows a case where ROOT fails to produce the χ^2 Residuals. More on this is discussed in Section 5.7.

Figure 14 shows two distributions that are marked as green by the χ^2 test. Here distribution (a) has a χ^2/N_{df} value of 9.06 and (b) a value of 8.43. These values are both close to the boundary between yellow and green and show similarity. The χ^2 test will often be less accurate for lower statistic samples however as large differences will generally be covered by error bars. Here there are no major differences and for the Tau Validation would be marked green even with the small errors. Here we did not expect exactly the same distributions since there are changes in the configurations.

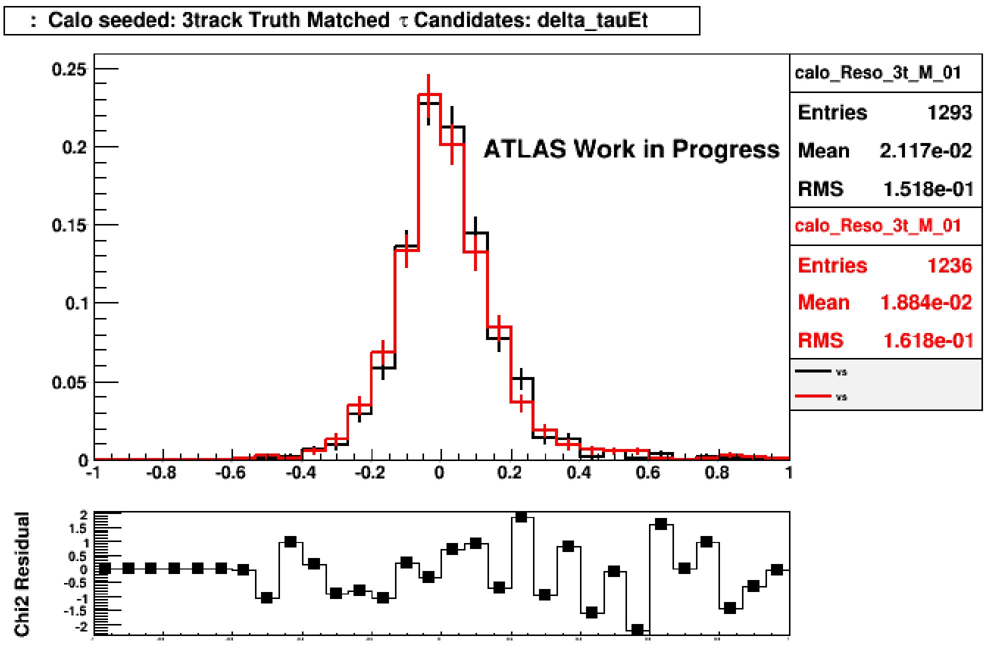
Figure 15 shows two distributions that are marked as yellow by the χ^2 test. Here we can see two different cases that fall under this yellow grading. Figure 15 (a) would be considered acceptable even though being marked as yellow whereas (b) would be reported as a problem to the production team. Here (a) has a χ^2/N_{df} value of 12.1 where as (b) has a χ^2/N_{df} value of 75.9. Typically the range 10 to 100 can include both acceptable and unacceptable distributions which is why all in this range should be highlighted for further investigation.

Figure 16 shows two distributions that are marked as red by the χ^2 test. Both the distributions would need to be reported to the production teams as problems and the χ^2 test identifies this. Here (a) has a χ^2/N_{df} value of 390.5 and (b) has a value of 184.5.

In summary, the χ^2 test in ATLAS Tau Validation is not used to make the final decision about whether two distributions are consistent with one another, but rather to help the user narrow down the number of plots they need to manually investigate for problems. The χ^2/N_{df} ranges which were chosen through manual tweaking and calibration were chosen such that only those distribution comparisons which are highly consistent according to the χ^2 test would be marked as green. The yellow category range was chosen such that it covers the range where both inconsistent and consistent comparisons are present. Red plots are identified as distributions that are extremely likely to be inconsistent. The categorisation of these plots helps the user perform validations faster and with better accuracy than when using the previous automatic grading techniques.

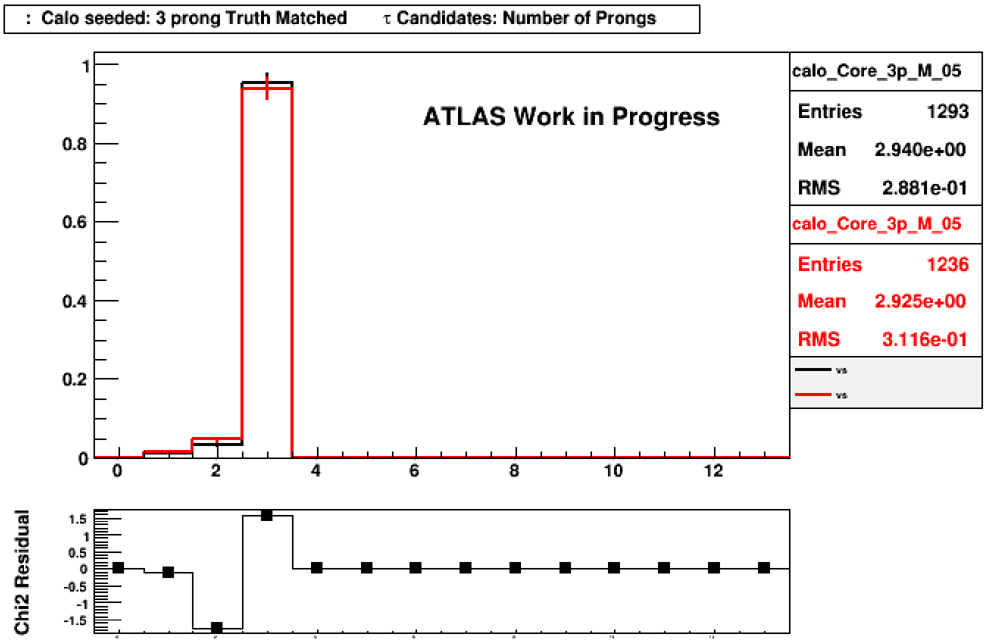


(a) Three track truth matched tau candidate hadronic radius.

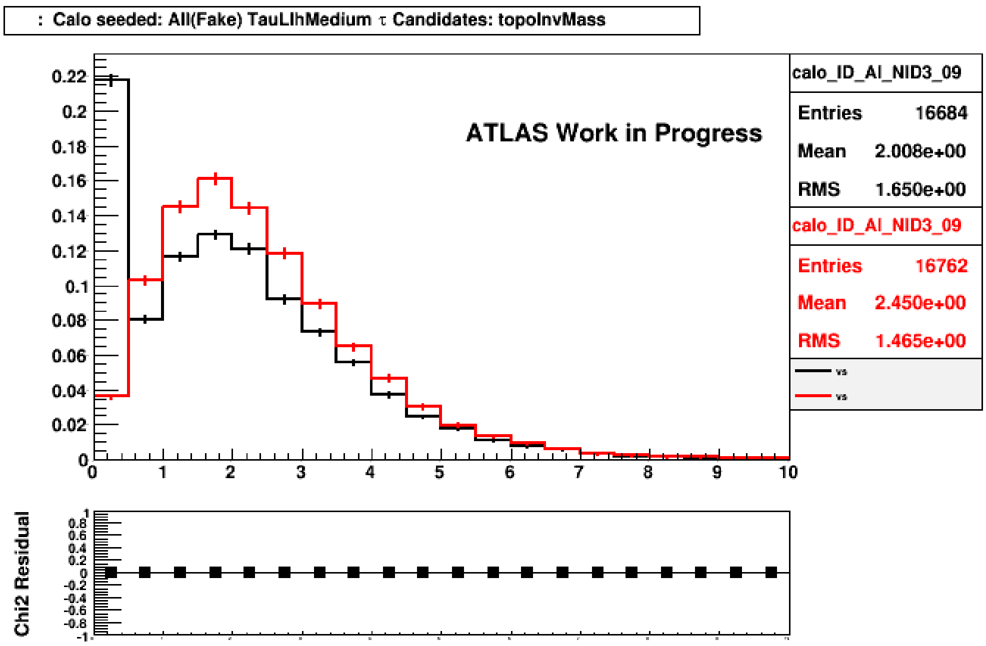


(b) Three track truth matched tau candidate delta E_T.

Figure 14: Examples of plots marked green using the χ^2 test.

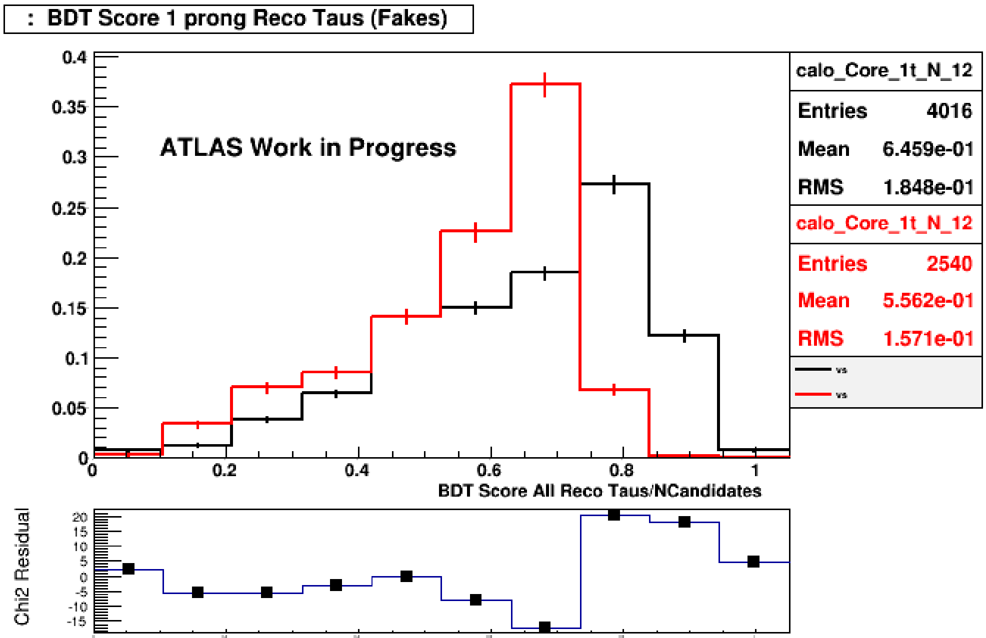


(a) Three prong truth matched tau candidates number of truth prongs.

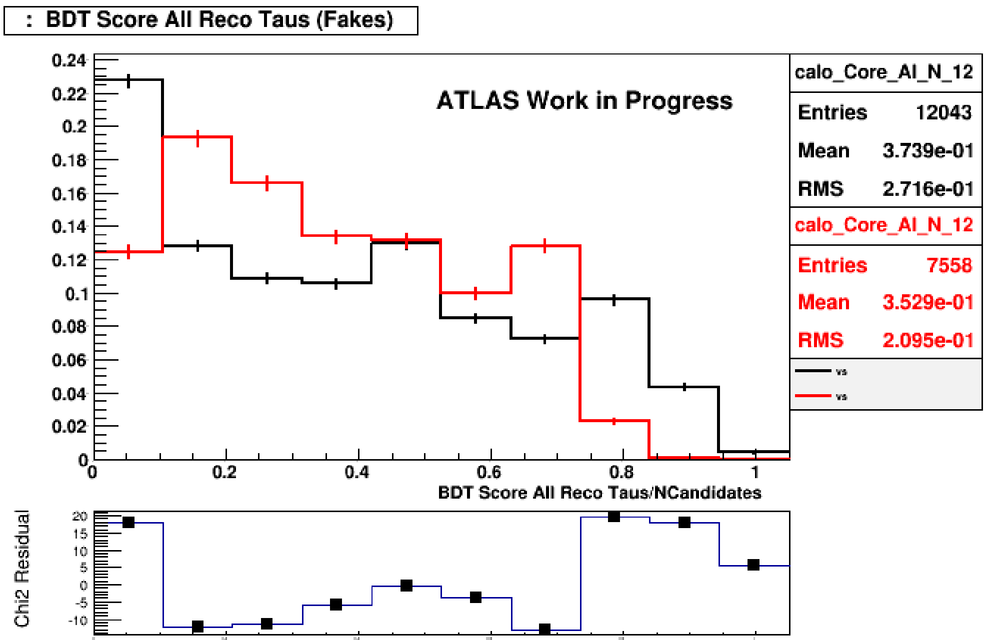


(b) All fake Tau candidates identified using the LLH Medium ID topo Invariant Mass.

Figure 15: Examples of plots marked yellow using the χ^2 test.



(a) One prong fake tau candidates BDT score.



(b) All fake tau candidates BDT score.

Figure 16: Examples of plots marked red using the χ^2 test.

5.6.3 New Core Variables

The previous section discusses the χ^2 testing method that helps users identify variables and distributions that are problematic in new releases. The test is not 100% accurate and it is possible that some smaller problems could go unseen. As discussed before, thousands of plots are produced for each validation. With this number of distributions being investigated it is not surprising that certain distributions would be considered less important than others. A list of core variables were identified that need to be manually investigated for every validation regardless of the automatic grading. Most of these core variables are those used in the identification and reconstruction algorithms of the taus [80]. These differ from other distributions in the thousands that are plotted which investigate only various results of the reconstruction and identification algorithms. Since these core variables are used as inputs to the identification algorithms then it is likely that the output will be consistent if these inputs are consistent.

A list of 28 new core variables were chosen that needed to be validated. These variables include common variables and discriminating variables. The variables implemented are listed and explained here.

- Charge: The charge of the tau candidate calculated as the sum of the charge of tracks associated with the tau candidate.
- E_T: The transverse component of the energy of the tau candidate.
- E: The energy of the tau candidate.
- η: The pseudo-rapidity of the tau candidate.
- ϕ: The azimuthal angle of the tau candidate.
- m: The mass of the tau candidate.
- p_T: The transverse component of the momentum of the tau candidate.
- N_{π^0}: The number of reconstructed neutral pions.
- Number of wide tracks: This variable stores the number of wide tracks that are associated with the tau candidate. Wide tracks are tracks that are contained within the isolation annulus of $0.2 < \Delta R \leq 0.4$.

- Track width: The angular separation ΔR between the track and the tau candidate.

- Lead Track $\frac{\eta}{p_T}$: The quotient of the tau candidate's leading track pseudorapidity and transverse momentum.

- $S_{lead\,track}$: The impact significance of the leading track associated with the tau candidate defined as

$$S_{lead\,track} = \frac{d_0}{\delta d_0}. \tag{5.2}$$

  Here d_0 is the distance of the closest approach of the track to the tau vertex in the transverse plane and δd_0 is the estimated uncertainty in d_0.

- $S_{lead\,loose\,track}$: similar to $S_{lead\,track}$ except that the tracks are identified using looser requirements.

- Lead track $z_0 \sin\theta$: The transverse longitudinal distance of the closest approach the leading track made to the reconstructed primary vertex.

- N^{iso}_{track}: The number of tracks in the isolation annulus of $0.2 < \Delta R \leq 0.4$.

- m_{tracks}: The invariant mass of the track system calculated using both core and isolation tracks.

- S^{flight}_T: The transverse flight path significance. This is the decay-length significance of the secondary vertex for multi-track tau candidates calculated in the transverse plane. It is defined as

$$S^{flight}_T = \frac{L^{flight}_T}{\delta L^{flight}_T}. \tag{5.3}$$

  Here L^{flight}_T is the reconstructed signed decay length with δL^{flight}_T being the estimated uncertainty in L^{flight}_T.

- R_{track}: The average track distance or track radius which is defined as the p_T-weighted track width. It is calculated using

$$R_{track} = \frac{\sum_i^{\Delta R_i \leq 0.4} p_{T,i} \Delta R_i}{\sum_i^{\Delta R_i \leq 0.4} p_{T,i}}. \tag{5.4}$$

  The sum here has i run over all core and isolation tracks with $\Delta R_i \leq 0.4$.

- R_{EM}: The electromagnetic radius of the shower associated with the tau candidate. It is defined as

$$R_{\mathrm{EM}} = \frac{\sum_{i\in\{\mathrm{EM}\,0\,-\,2\}}^{\Delta R_i<0.2} E_{T,i}^{\mathrm{EM}} \Delta R_i}{\sum_{i\in\{\mathrm{EM}\,0\text{-}2\}}^{\Delta R_i<0.2} E_{T,i}^{\mathrm{EM}}}. \tag{5.5}$$

The sum runs over all cells in Layer 1 and 2 of the EM calorimeter within $\Delta R_i \leq 0.2$. $E_{T,i}^{\mathrm{EM}}$ is the transverse energy deposited in Cell i.

- R_{had}: The hadronic radius of the shower associated with the tau candidate. It is defined as

$$R_{\mathrm{had}} = \frac{\sum_{i\in\{\mathrm{Had}\}}^{\Delta R_i<0.2} E_{T,i}^{\mathrm{EM}} \Delta R_i}{\sum_{i\in\{\mathrm{Had}\}}^{\Delta R_i<0.2} E_{T,i}^{\mathrm{EM}}}. \tag{5.6}$$

The sum runs over all cells in the hadronic calorimeter and Layer 3 of the EM calorimeter within $\Delta R_i \leq 0.2$. $E_{T,i}^{\mathrm{EM}}$ is the transverse energy deposited in Cell i and is calibrated at the EM energy scale.

- StripWidth2: The energy weighted width in the calorimeter strips associated with the hadronic tau.

- E_T^{iso}: The cluster isolation energy of the tau candidate. It is defined as the transverse energy of the clusters in the isolation annulus and is calculated using

$$E_T^{\mathrm{iso}} = \sum_i^{0.2<\Delta R_i<0.4} E_{T,i}. \tag{5.7}$$

- E_T^{iso}: The corrected cluster isolation energy. It is calculated using

$$E_{T,corr}^{\mathrm{iso}} = E_T^{\mathrm{iso}} - \delta E_T^{\mathrm{iso}} = \sum_i^{0.2<\Delta R_i<0.4} E_{T,i} - \delta E_T^{\mathrm{iso}}. \tag{5.8}$$

Here $\delta E_T^{\mathrm{iso}}$ is a pileup correction term and is defined as

$$\delta E_T^{\mathrm{iso}} = (1 - f_{JVF}) \times \sum p_{T,trk}. \tag{5.9}$$

f_{JVF} is the jet-vertex fraction which is the fraction of the p_T from tracks associated to the tau candidates primary vertex to the p_T from all associated tracks.

- f_{core} : The core energy fraction. This is the fraction of transverse energy in the region $\Delta R_i < 0.1$ for the tau candidate. It is defined as

$$f_{\text{core}} = \frac{\sum_{i\in\{all\}}^{\Delta R_i<0.x_1} E_{T,i}}{\sum_{j\in\{all\}}^{\Delta R_j<0.x_2} E_{T,j}}. \tag{5.10}$$

The sum for i runs over all cells in the calorimeters within $\Delta R < 0.1$ with the sum for j running cells in the range $\Delta R < 0.2$. $E_{T,i}^{\text{EM}}$ ($E_{T,j}^{\text{EM}}$) is the transverse energy deposited in Cell $i(j)$ and is calibrated at the EM energy scale.

- f_{iso}: The ring isolation of the tau candidate. This is defined as

$$f_{\text{iso}} = \frac{\sum_{i\in\{\text{EM0-2}\}}^{0.1<\Delta R_i<0.2} E_{T,i}}{\sum_{j\in\{\text{EM0-2}\}}^{\Delta R_j<0.2} E_{T,j}}. \tag{5.11}$$

The sum for i runs over all cells in the calorimeters within the annulus $0.1 < \Delta R < 0.2$ with the sum for j running cells in the range $\Delta R < 0.2$. $E_{T,i}^{\text{EM}}$ ($E_{T,j}^{\text{EM}}$) is the transverse energy deposited in Cell $i(j)$ and is calibrated at the EM energy scale.

- $f_{2(3)\text{ lead clusters}}$: The ratio of the two (three) leading cluster energies to the total energy of all clusters associated with the tau candidate.

- n_{strip}: The number of calorimeter strips with energy deposited in them associated with the hadronic tau.

5.6.4 New tau ID

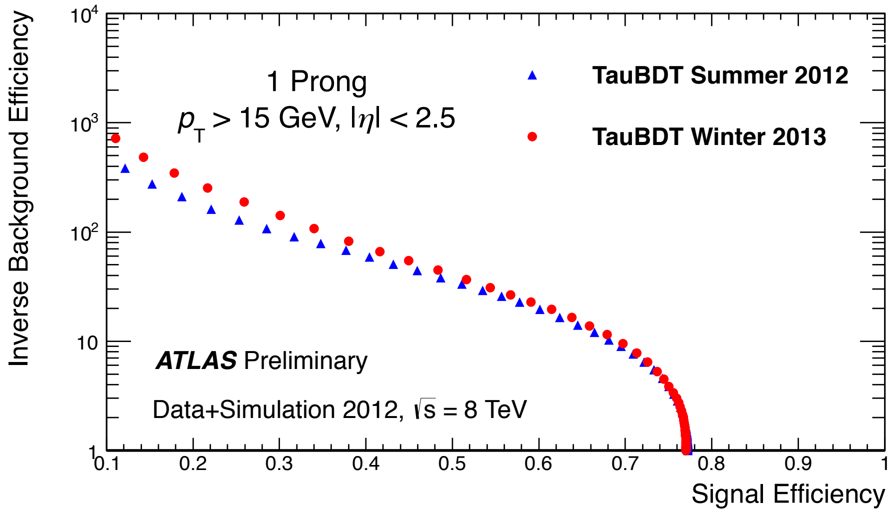
In 2013 changes were made to the Tau identification (tau ID) algorithms. Compared to the old tau ID (Summer 2012 tau ID) this new tau ID (Winter 2013 tau ID) uses a more pileup robust set of variables for the BDT training by including π^0 reconstruction information [80]. For the training of these algorithms $Z \to \tau\tau$, $W \to \tau\nu$ and $Z' \to \tau\tau$ MC samples were used. Number of vertex (N_{vtx}) weighting of background to MC is performed followed by p_T-weighting of MC to the N_{vtx} weighted background[80]. Training is performed using 2/3 of these entries with testing being performed on the remaining 1/3. Only truth matched tau candidates are used that pass the following requirements:

- $p_T > 15$ GeV

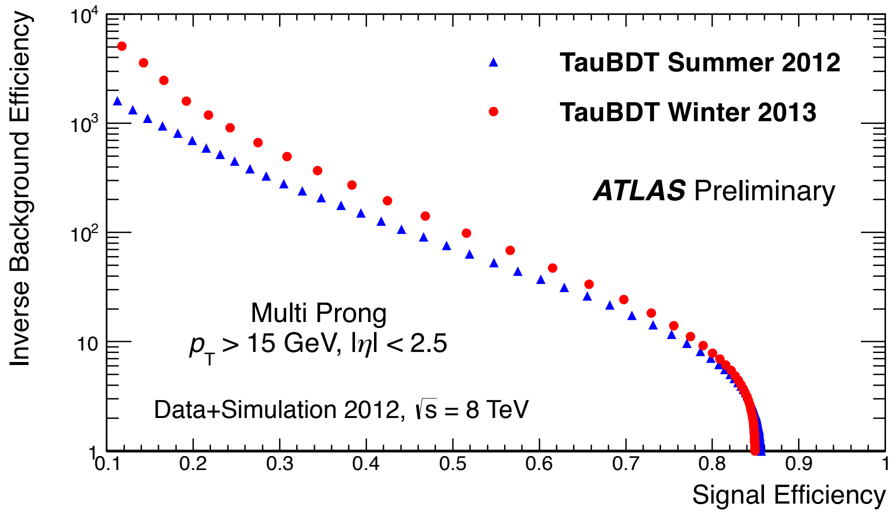
- $\eta < 2.5$
- $N_{tracks} = 1||3$
- $|p_T^{true}| > 10$ GeV
- $|\eta^{true}| < 2.5$
- $N_{tracks}^{true} = 1||3$.

Separate BDTs were developed for single-prong and multi-prong identification. These new BDTs were not only designed to be more pileup robust but performance increase was another obvious goal. The new BDT saw a candidate rejection increase at medium efficiency of about 15% and 100% for single-prong and multi-prong identification respectively [80]. A comparison of the signal vs background efficiencies for the Summer 2012 BDT and Winter 2013 can be seen in Figure 17.

The new tau ID as of the end of 2013 does not yet appear in the AOD datasets that are used for Tau Validation. For this reason it needs to be manually calculated when performing these validations. Different versions of tau reconstruction software exist for different Athena releases. The automatic validation was updated to manage the installation of the correct versions of the tau reconstruction software and changes were made to update to the new tau ID used. The Tau Validation system is not built to see how well the new tau ID works but rather test that the current implementation of the tau ID maintained by the Tau Working Group runs consistently across different releases. After implementing the new tau ID the only significant changes expected are in the BDT scores for tau candidates. To test this a validation was run where the reference used the old tau ID and the test was run using the new tau ID. An MC12 (2012 Monte Carlo) sample was chosen which had standard pileup and geometry configurations. The validation showed significant changes in the tau BDT scores with no detectable changes seen in any other variables. Figure 18 shows examples of differences seen in the validation.

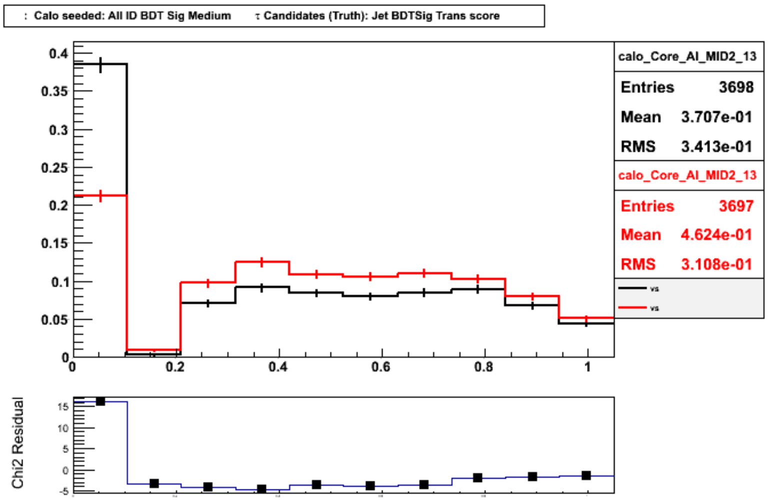


(a) One-prong efficiencies.

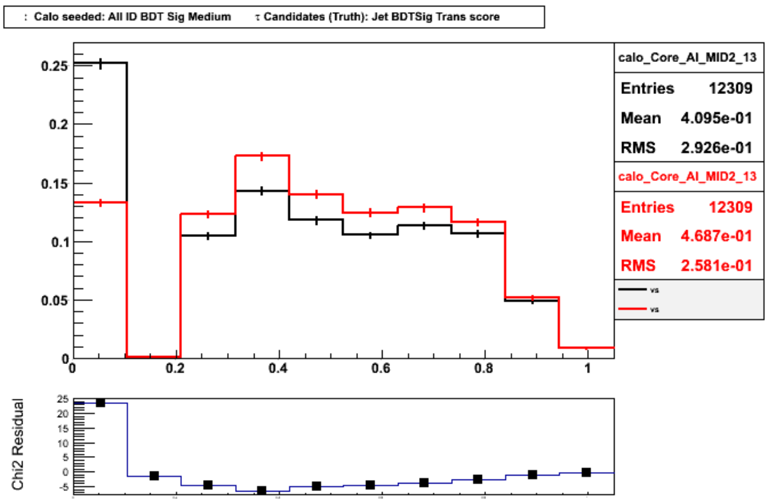


(b) Multi-prong efficiencies.

Figure 17: Comparison of Summer 2012 and Winter 2013 tau ID efficiencies [80].



(a) BDT Medium scores for all truth tau candidates for the $Z \rightarrow ee$ sample.



(b) BDT Medium scores for all truth tau candidates for the TTBar sample.

Figure 18: Comparison of old tau ID to new tau ID in Tau Validation. Here the red line is the new tau ID and the black is the old tau ID.

5.6.5 General Improvements

A few other changes were made to the tau validation package which warrant mentions. These changes are general clean up of the tau validation package to make it more user friendly. The improvements are explained here and why they are of use.

For each validation it has been mentioned that an HTML report is generated which is then uploaded to a web area and the results are then viewed as a web page. The results are collected and presented in table form with the index page being a large information table with a result table linking to all the plots sorted by grading. This was shown earlier in Figures 11, 12 and 13. The website functionality was improved by removing irrelevant information, adding the ability to hide the bulky information table, and making table sortable which allows users to easily find plots by name or sort them by χ^2 value. The overall website user experience is improved with these changes.

Previously results of the validation were split into two categories. These two categories were results for tau candidates with $p_T < 20$ GeV (low p_T) and results for tau candidates with $p_T \geq 20$ GeV. It was decided that the low p_T comparisons were no longer necessary since no analyses use tau candidates in this p_T range. The tau validation package was edited to stop producing these results since they take up unnecessary memory and time to produce.

These changes mentioned so far required modifying the automatic HTML reporting code. In order to simplify future maintenance this code was rewritten providing developers better control when expanding the website. A simple Python HTML generation system was produced with a focus on being easy to understand and modify should the need arise. The automatic HTML report generator was rewritten using this system with full commenting provided to make sure it was understandable.

Finally with ~1000 histograms being produced per sample, many of these had poor binning and x-axis ranges assigned to them. To make plots more readable and understandable many of these histograms needed the values to be fine tuned over a series of validation tests.

5.7 The Future of Tau Validation

The Tau Validation is constantly undergoing changes and improvements. Some plans already exist for the future of the Tau Validation package which are listed here.

- The $Z \to \tau\tau$ sample is being switched to the new 147818.Pythia8 because it was discovered that the current sample has a τ lifetime error.
- Names of π^0 variables are changing along with new ones being introduced in

future Athena releases and the Tau Validation package needs to adapt to these changes.

- The automation system along with current validation framework cannot easily handle comparisons of the same test and reference with different Athena installations or job option settings. This is important to validate that new Athena installations work correctly and to allow changes in the Tau Validation package to be tested (as was needed for testing the new tau ID).

- Introducing a high p_T sample is being considered to test that the reconstruction algorithms work across different high p_T data sets.

The first two feature changes in this list are currently being developed and tested with the expected date of completion being early 2014. The automation system update is a suggestion but not a necessity since with some manual work and renaming tricks the validation system can still produce these results. This will be updated should time be available. The high p_T sample is still in discussion as to if or when it would be necessary.

6 Pythia vs Herwig validation studies

It was decided that the $H \to \tau\tau$ group wanted to move from Alpgen Herwig (AH) MC samples to Alpgen Pythia (AP) samples. In both cases here Alpgen models the initial process whilst Herwig and Pythia perform the hadronization step. The reason for wanting to move was because the newer AP samples were produced with more statistics than the older AH samples. Before switching to the new samples it was necessary to investigate whether the two generators produced similar results. A validation study comparing the AP samples against the AH samples is presented in this section.

6.1 Validation Requirements

The two sets of MC samples both use Alpgen to model the hard processes but use different MC generators to perform the hadronization of the event generation. For this reason if any differences exist they are expected in jet-related variables and these are what need to be validated when comparing Herwig against Pythia. The samples that are of interest to this study are $Z \to ee$, $Z \to \mu\mu$, $Z \to \tau\tau$, $W \to e\nu$, $W \to \mu\nu$ and $W \to \tau\nu$. For all of these samples the Jimmy implementation of Herwig is used to produce them; the word Jimmy is used to refer to the Herwig-related samples. The variables that need to be validated were investigated at two stages. The first was chosen to be before all standard event selection criteria are applied and only truth information for jets with $p_T > 30$ GeV was investigated. The second stage was done at what is known as the pre-selection level where both truth and reconstructed variables were investigated. The details of what event selection criteria make up the pre-selection are discussed in the following section.

6.2 Pre-Selection

Pre-selection is a stage in the full $H \to \tau\tau$ analysis which is just before the events are categorised into different production modes such as VBF or $gg \to H$. The event selection criteria that need to be applied to reach the pre-selection level in this analysis are detailed here.

6.2.1 Object Selection

The selection criteria described here are used to define objects used in the analysis including electrons, muons, hadronic taus and jets.

6.2.1.1 Electrons

The electron candidates are reconstructed using clusters in the electromagnetic calorimeter matched to a track in the inner detector. Electrons require $p_T > 15$ GeV, $|\eta| < 2.47$ and must be identified with a loose ID. Electrons in the 'crack region' $1.37 < |\eta| < 1.52$ (where the EM calorimeter barrel meets the end-cap) are rejected. Electron quality criteria are also applied which means ignoring electrons identified using poor quality data. The electrons have the following isolation requirements

- $I(E_T, 0.2) < 0.06$ – the total other E_T within a cone of $\Delta R < 0.2$ of the electron must be $< 6\%$ of the electron's E_T.
- $I(p_T, 0.4) < 0.06$ – all other tracks with $p_T > 1$ GeV combined p_T within a cone of $\Delta R < 0.4$ of the electron must be $< 6\%$ of the electron's p_T.

6.2.1.2 Muons

Muons are reconstructed using the STACO [87] algorithm with a tight ID[3] and are required to be combined muons. They require $p_T > 10$ GeV and $|\eta| < 2.5$. Quality criteria are also applied in the same way as for the electrons. Muons have the same isolation requirements as the electrons.

6.2.1.3 Jets

Jets are identified and reconstructed using the anti$-k_t$ algorithm[88] with a distance parameter $R = 0.4$. The local hadronic calibration scheme[87] is used to calibrate energy depositions from hadrons for both reconstructed jets and E_T^{miss} in the detector. Jets are required to be in the range of $|\eta| < 4.5$ and jets in the range $|\eta| < 2.4$ are required to have a jet-vertex fraction greater than 0.5.

[3] Note that tight, medium and loose IDs refer to the number of requirements imposed when identifying a particle. The tight ID has the strictest requirements leading to low identification rate but a high background rejection rate. The loose ID has the most lenient requirements leading to a high identification rate but a low background rejection rate.

6.2.1.4 Hadronic Taus

Hadronic taus are reconstructed using clusters in the electromagnetic and hadronic calorimeters [23]. Taus are seeded by jets with $p_T > 20$ GeV, an anti$-k_t$ distance parameter $R = 0.4$ and are required to be in the region $|\eta| < 2.47$. The BDT tau identification method [79] is used to identify taus and requires the tau to pass a medium ID. A standard muon veto and medium electron veto are applied based on a BDT algorithm to both one-prong and three-prong taus. Tau candidates are also required to possess a charge of ± 1 and 1 or 3 associated tracks in a cone of radius of $\Delta R < 0.2$.

6.2.1.5 Overlap Removal

Object overlap removal is done to make sure that no two objects that geometrically overlap with each other within a cone of $\Delta R < 0.2$ are kept for analysis. Only one of the objects that overlap within this cone will be kept for analysis. The selection is performed in the order of muons, electrons, hadronic taus and then jets. For overlap removal against hadronic taus only, the muon p_T requirement is dropped to 4 GeV. The electron ID requirement is also relaxed to loose for the overlap removal step.

6.2.2 General Selection Requirements

There are a number of general selection requirements that are applied to events which are detailed in this section below.

- Events require at least one primary vertex, with at least four associated tracks.
- For both $\tau_e \tau_{had}$ and $\tau_\mu \tau_{had}$ final states, the event requires that either a single lepton trigger (SLT) or combined trigger (CT) are fired.
- pT thresholds are required on the τ_{had} candidate depending on which trigger is fired. These thresholds are detailed in Table 7.
- Exactly one lepton and one τ_{had} candidate passing the p_T thresholds are required and must be of opposite charge.

Trigger	p_T Thresholds in GeV
Single Electron	$p_T^e > 26$ $p_T^{\tau_{had}} > 20$
Single Muon	$p_T^\mu > 26$ $p_T^{\tau_{had}} > 20$
Combined $e + \tau_{had}$	$20 < p_T^e < 26$ $p_T^{\tau_{had}} > 25$
Combined $e + \tau_{had}$	$17 < p_T^e < 26$ $p_T^{\tau_{had}} > 25$

Table 7: Table of trigger related lepton p_T selection criteria.

6.3 Variables to Validate

There are different variables that were investigated and validated for the truth information and for the reconstructed information. This section lists and describes those variables. Note that for all variables being validated no event weighting was applied.

6.3.1 Truth Jet Information

The list that follows describes truth jet information that was validated across the different samples.

- Jet multiplicity - The number of jets per event.
- Jet η - The pseudorapidity of the jets in the event. The jet η was plotted in four different categories. First all jets η is plotted, then it is separated into plots for the leading, sub-leading and sub-sub-leading jets η distributions.
- Jet p_T - The transverse momentum of the jets in the event. The p_T is also investigated for all jets, the leading jets, sub-leading jets and sub-sub-leading jets separately.

6.3.2 Reconstructed Jet Information

The list that follows describes the reconstructed jet information that was validated across the different samples.

- Jet Multiplicity - The number of jets per event.
- Jet Mass - The reconstructed jet mass for all jets.
- Jet η - The pseudorapidity of the jets in the event. The reconstructed jet η is also investigated for all jets, the leading jets, sub leading jets and sub sub leading jets separately.
- Jet p_T - The transverse momentum of the jets in the event. The reconstructed p_T is also investigated for all jets and the leading, sub leading and sub sub leading jets separately.

6.3.3 Other reconstructed variables

The list that follows describes three values that are affected by changes to the jets in the events which means they also need to be investigated.

- p_T^{Tot} - This variable is defined as the sum of the leading and sub-leading jet's p_T. In the case where there is only a leading jet p_T^{Tot} is just the leading jet's p_T. In the case where there are no jets in the event, $p_T^{Tot} = 0$.
- m_{MMC} - This is the reconstructed invariant $\tau\tau$ mass which is reconstructed using what is known as the Missing Mass Calculator (MMC)[89]. The MMC method fully reconstructs the event topology to produce this invariant (the same in all frames of reference) mass. This requires six to eight unknowns to be solved for. These unknowns are the x-, y- and x-components of the momentum for the neutrinos related to the two τ leptons in the event ($p_{miss_{1,2}}$), and the invariant mass of the neutrinos related to leptonic τ decays ($m_{miss_{1,2}}$). These unknowns are contained within four equations:

$$E_{T_x}^{miss} = p_{miss_1} \sin\theta_{miss_1} \cos\phi_{miss_1} + p_{miss_2} \sin\theta_{miss_2} \cos\phi_{miss_2}, \tag{6.1}$$

$$E_{T_y}^{miss} = p_{miss_1} \sin\theta_{miss_1} \sin\phi_{miss_1} + p_{miss_2} \sin\theta_{miss_2} \sin\phi_{miss_2}, \tag{6.2}$$

$$M_{\tau_1}^2 = m_{miss_1}^2 + m_{vis_1}^2 + 2\sqrt{p_{vis_1}^2 + m_{vis_1}^2}\sqrt{p_{miss_1}^2 + m_{miss_1}^2} - 2p_{vis_1} p_{miss_1} \cos\Delta\theta_{vm_1}, \tag{6.3}$$

$$M_{\tau_2}^2 = m_{miss_2}^2 + m_{vis_2}^2 + 2\sqrt{p_{vis_2}^2 + m_{vis_2}^2}\sqrt{p_{miss_2}^2 + m_{miss_2}^2} - 2p_{vis_2} p_{miss_2} \cos\Delta\theta_{vm_2}. \tag{6.4}$$

Here $E_{T_x}^{miss}$ and $E_{T_y}^{miss}$ are the x- and y-components of E_T^{miss}, $p_{vis_{1,2}}$, $m_{vis_{1,2}}$, $\theta_{vis_{1,2}}$, $\phi_{vis_{1,2}}$ are the momenta, invariant masses, polar and azimuthal angles of the visible τ decay products, and $M_\tau = 1.777$ GeV [37]. $\Delta\theta_{vm_{1,2}}$ are the angles between the vectors p_{vis} and p_{miss} for the two τ leptons. The remaining variables are the unknowns described previously. For a τ that decays hadronically m_{miss} is then zero which reduces the number of unknowns. This means that in the $H \rightarrow \tau_l\tau_{had}$ final state seven unknowns need to be solved for. The system of equations is underconstrained and the MMC has been designed to overcome this problem. This is done by performing a scan of the variables related to the visible decay products associated with the τ lepton and the full τ lepton four-vector. This produces a probability associated with each scan point which is used to weight the point. The mass of the $\tau\tau$ system is then estimated by taking the scan point with the largest weighting. The probability weighting is calculated using

$$L = -\log\left(P\left(\Delta R_1, p_{\tau_1}\right) \times P\left(\Delta R_2, p_{\tau_2}\right)\right). \tag{6.5}$$

Here P represents probability functions for different τ decay topologies.

- E_T^{miss} - This is the missing transverse energy in the event. In the $H \rightarrow \tau_l\tau_{had}$ analysis missing energy is always present due to the presence of undetectable τ neutrinos. The reconstruction of E_T^{miss} is done using calorimeter cells which are calibrated according to the reconstructed physics objects they are associated with [90]. The cells are associated with parent objects in the order of electrons, photons, hadronic τ leptons, jets and then muons. The cells not associated with these physics objects are also considered and the contribution is named $E_T^{miss,CellOut}$. The E_T^{miss} calorimeter term is then calculated using

$$E_{T\,x,y}^{miss,calo} = \sum_o E_{T\,x,y}^{miss,o}. \tag{6.6}$$

Here $o = e,\ \gamma,\ \tau$, jets, soft jets, calo-μ and CellOut and each term is calculated using

$$E_{T\,x}^{miss,o} = -\sum_{i=1}^{o-\text{term cell}} E_i \sin\theta_i \cos\phi_i, \tag{6.7}$$

$$E_{T\,y}^{miss,o} = -\sum_{i=1}^{o-\text{term cell}} E_i \sin\theta_i \cos\phi_i. \tag{6.8}$$

E_i is the energy, θ_i is the polar energy and ϕ_i is the azimuthal angle. The summation is done only for the specified physics objects in the range $|\eta| < 4.5$ and then the vector sum p_T of the muons is added to the result.

6.4 Results of the Validation

This section shows and discusses the results obtained from the validation for the different samples. Not all results are shown here as some plots are meaningless, due to low statistics, to be relevant or no statistics at all, particularly in the case of sub-sub-leading jet values. Important or interesting plots are shown here whilst all results with acceptable statistics to be interesting are shown in the Appendix of this thesis.

6.4.1 Initial MC Truth Information

First the results of the initial MC truth information are discussed here. As mentioned before, all results here are for truth jets with $p_T > 30$ GeV. For truth information it is expected that results will be very similar. The observed truth variables show good correlation between the AP samples and the AH samples with only small differences. In general it can be seen that Pythia produces slightly fewer low p_T jets, as seen in Figure 19, and slightly more central jets (η close to 0), as seen in Figure 20, than Jimmy does across all samples but the differences are not large so can be accepted. These initial plots show that at the event generation level there are no large differences and that we would not expect to see any huge differences later on after event selection criteria are applied.

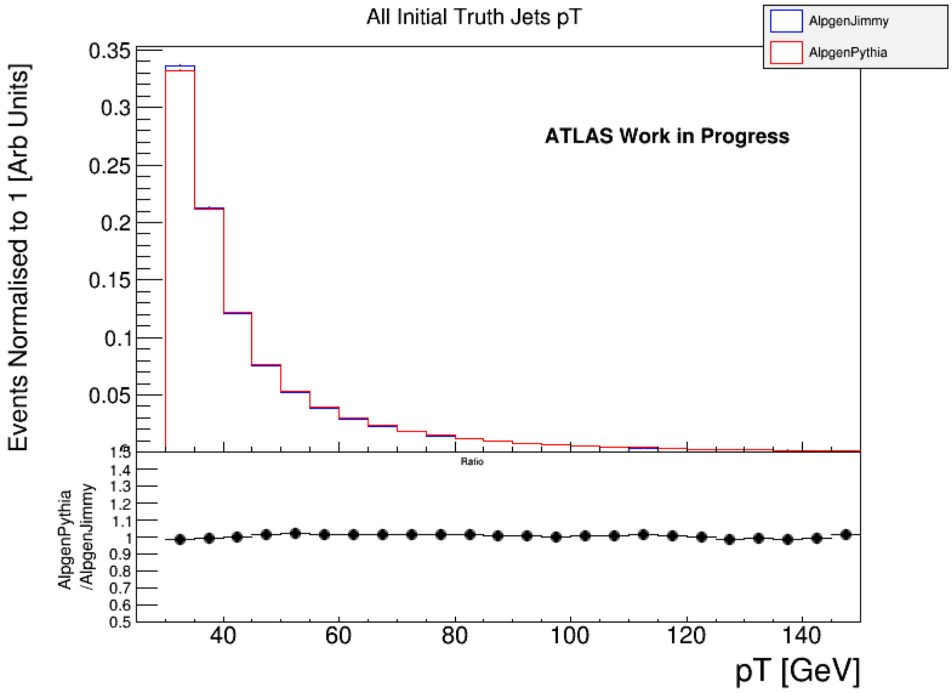


Figure 19: Initial $W \to \tau\nu$ truth jet p_T distribution.

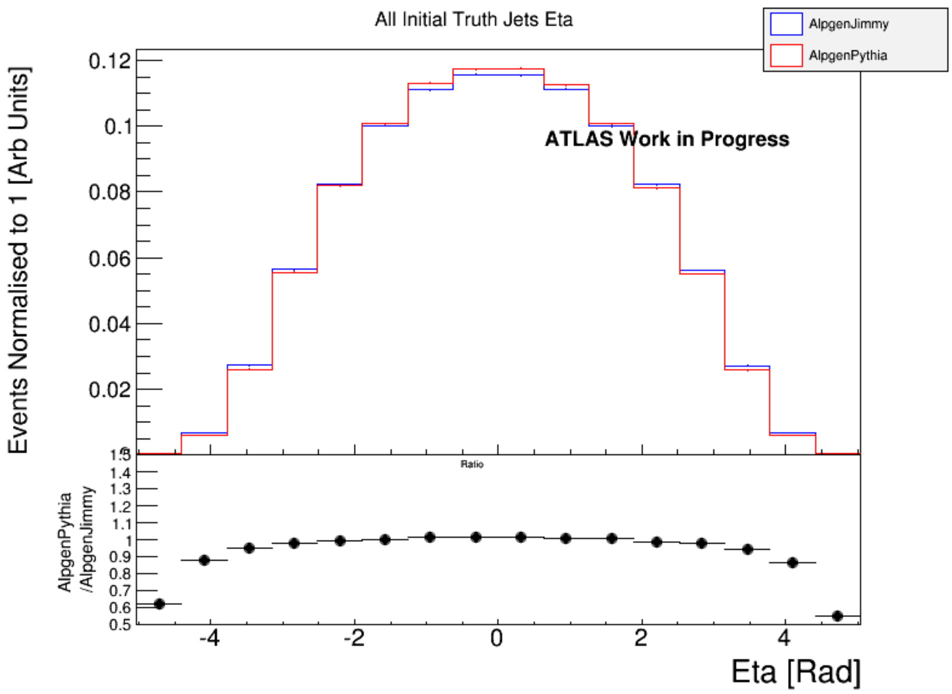


Figure 20: Initial $W \to \tau\nu$ truth jet η distribution.

6.4.2 Pre-selection Level Information

This section is used to discuss the results produced after all the pre-selection event selection criteria have been applied. The truth information plots at this stage shows us how jet related variables have changed from the initial stage of the analysis to the pre-selection stage on the truth level. This allows us to see that the truth information is still consistent across both samples at this stage and allows a comparison between the truth level information and reconstructed information to be drawn. The reconstructed variables allow a comparison to be made across the two sample sets to verify that the reconstruction algorithms used to reconstruct jets work similarly across the two samples. Statistics are expected to be much lower and differences may be present.

6.4.2.1 Pre-selection Truth Information

Looking at the truth information at pre-selection level large changes can be seen from the initial plots. In particular, the truth jet multiplicity plots have changed and in general most events at pre-selection either have no truth jets or just one truth jet. The number of events with sub-leading and sub-sub-leading jets at this level is negligible or none across all sample bases. This leads to statistics for these variables being so low that either the plots are empty or uninformative. Another consequence is that the leading truth jet plots closely mimic the plots for all truth jets so can be ignored too. Changes in the η distributions are expected since requirements on the η range are imposed on the jets for pre-selection.

Investigating the actual results of the validation at this stage it can be seen that both Pythia and Jimmy produce similar results for the truth jet multiplicity plots. Relative differences on the number of events with single jets in them could be considered non negligible as seen in Figure 21 but were considered to be within acceptable range. The truth p_T plots show differences, however all major discrepancies are covered by the uncertainties shown in the plots to a degree that makes the results acceptable, as seen in Figure 22. The η plots show similar results with differences observed in the plots, but they are within uncertainties to a degree considered acceptable, as seen in Figure 23.

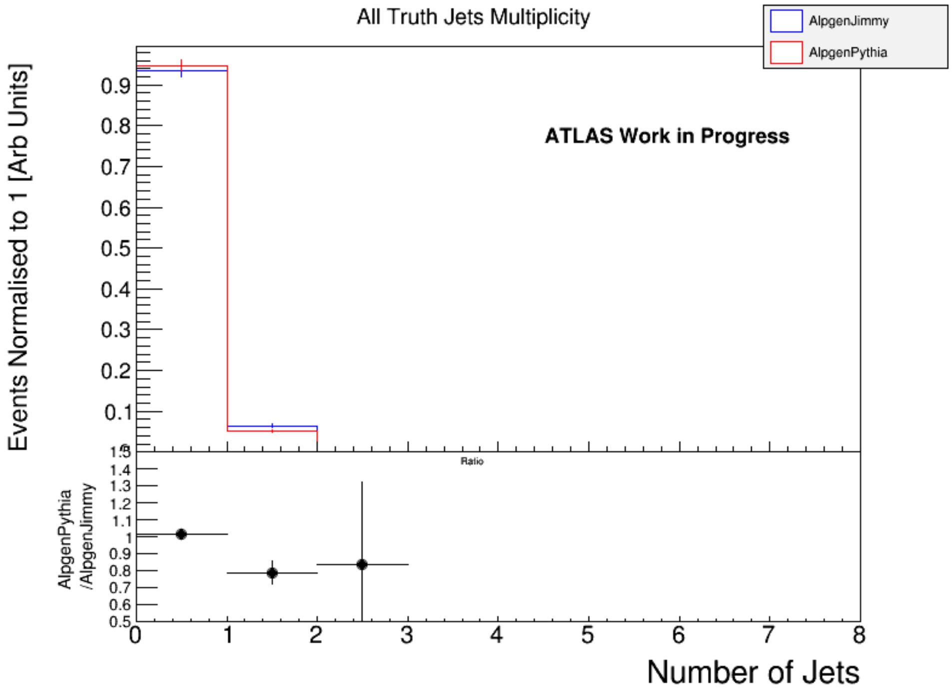


Figure 21: Pre-selection $Z \to ee$ truth jet multiplicity distribution.

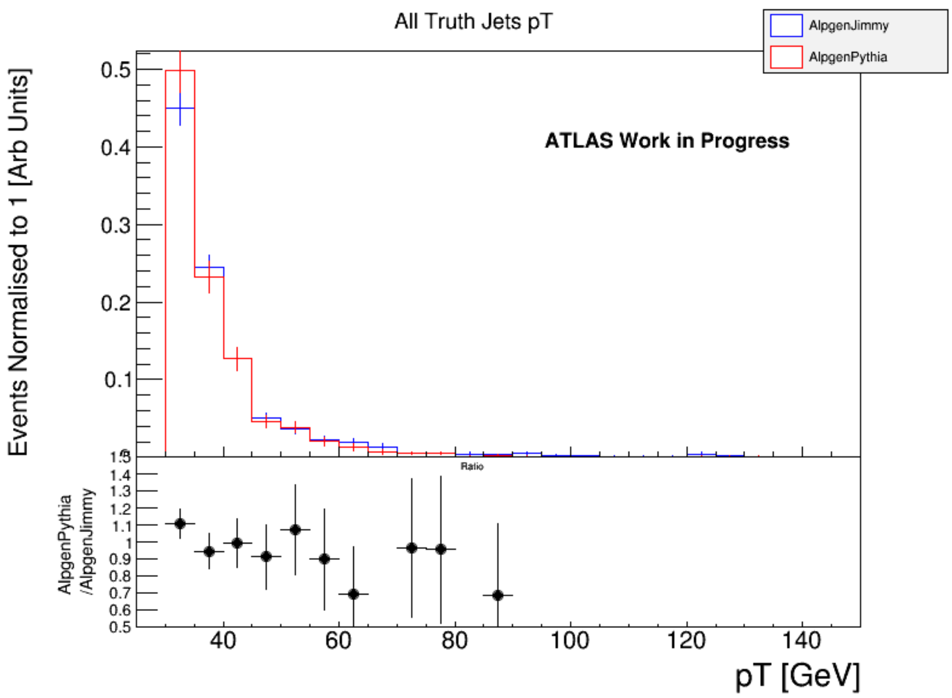


Figure 22: Pre-selection $W \to e\nu$ truth jet p_T distribution.

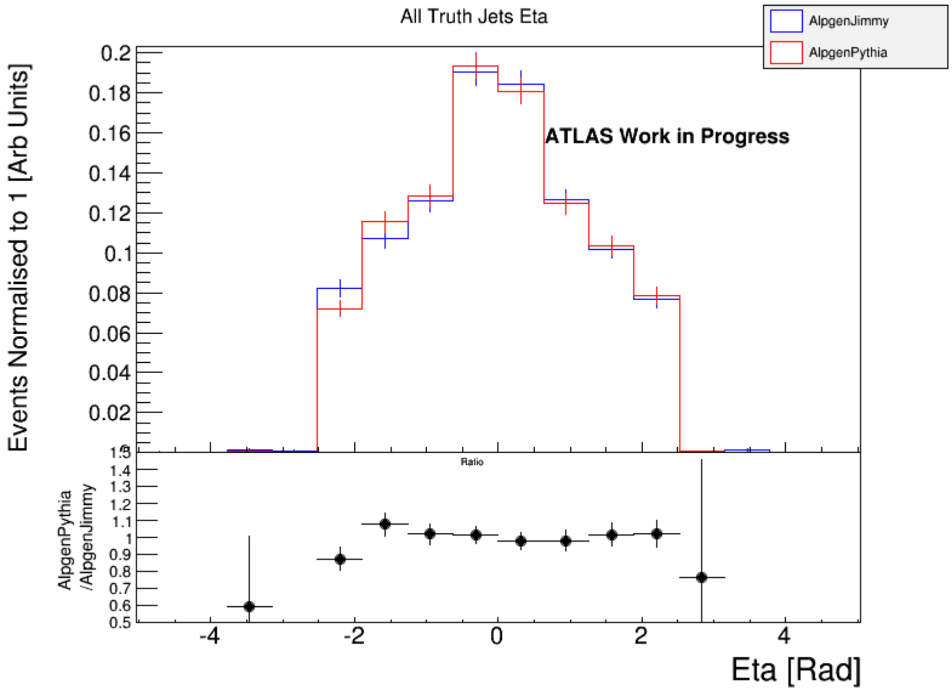


Figure 23: Pre-selection $Z \to \tau\tau$ truth jet η distribution.

6.4.2.2 Pre-selection Reconstructed Information

All the reconstructed variables at the pre-selection level can be seen in the Appendix but the overview of the results is discussed here. Firstly a discussion on the variables that are not present in the truth level plots is made with a discussion and comparison of the results corresponding to the truth level plots to follow.

All reconstructed variables show good correlation between the samples. Reconstructed jet energies and masses show good correlation with error bars covering all discrepancies between the two samples as can be seen in the examples in Figure 24 and Figure 25. The reconstructed jet ϕ shows lower agreement but taking error bars into account takes the differences down to acceptable levels as seen in Figure 26. The more important variables that need to be investigated were the p_T^{Tot}, m_{MMC} and E_T^{miss} plots. The p_T^{Tot} results show good correlation between the two samples with all values being within uncertainties. The m_{MMC} distributions are reasonable with some small differences noticeable in the $W \to e\nu$ and $Z \to ee$ plots that are not covered by the error bars as seen in Figure 27. The differences were still considered within acceptable ranges and the m_{MMC} plots pass the validation requirements. The E_T^{miss} plots allow for a similar conclusion to be drawn. Small differences are observed that are not covered by the error bars, as seen in Figure 28, but these differences are considered small enough making the E_T^{miss} results acceptable.

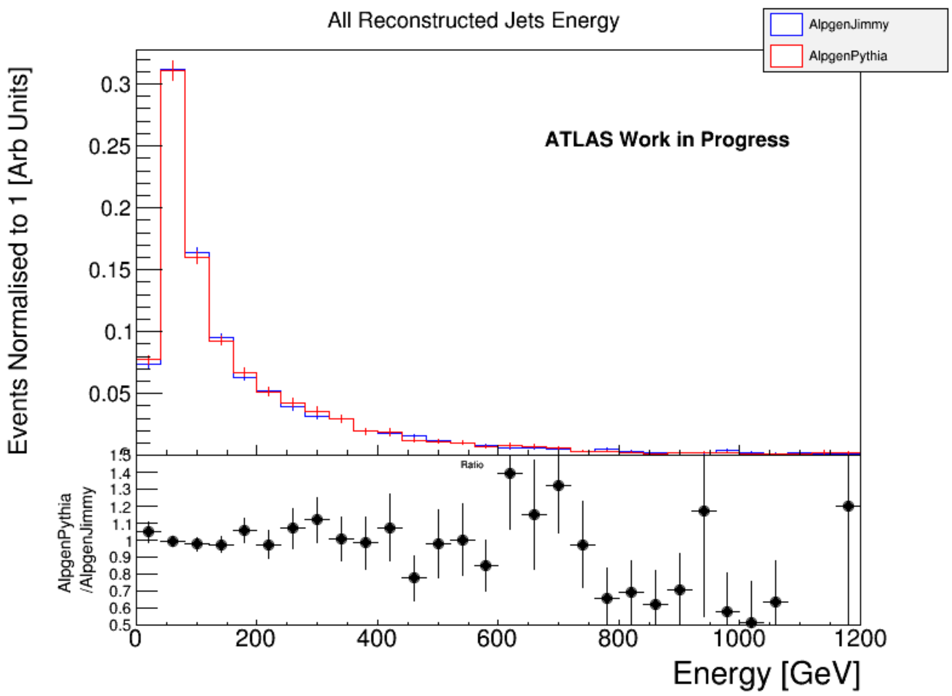


Figure 24: Pre-selection $W \rightarrow \mu\nu$ reconstructed jet energy distribution.

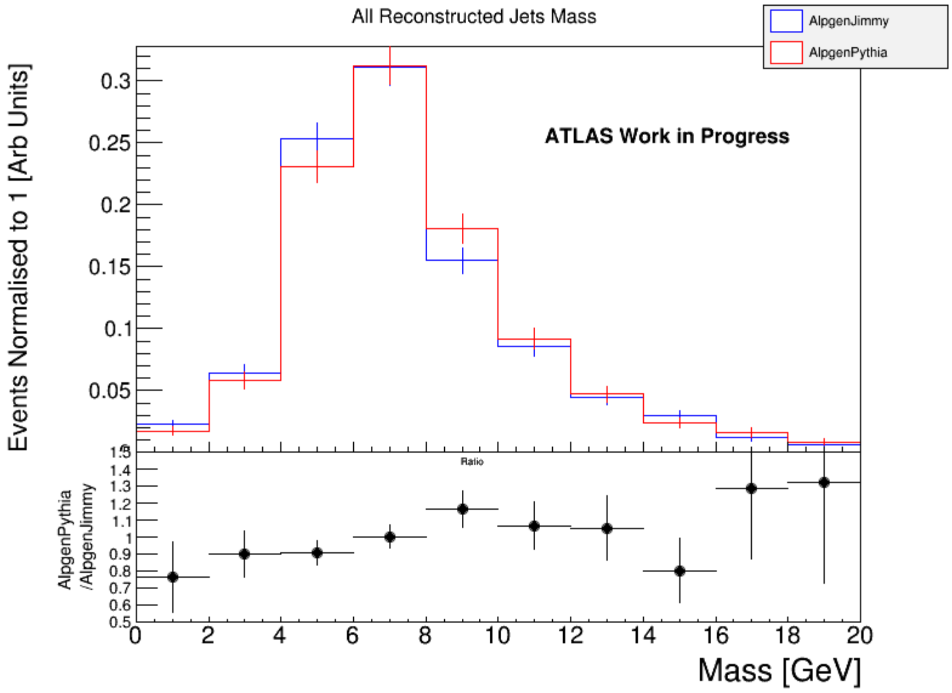


Figure 25: Pre-selection $Z \to \mu\mu$ reconstructed jet mass distribution.

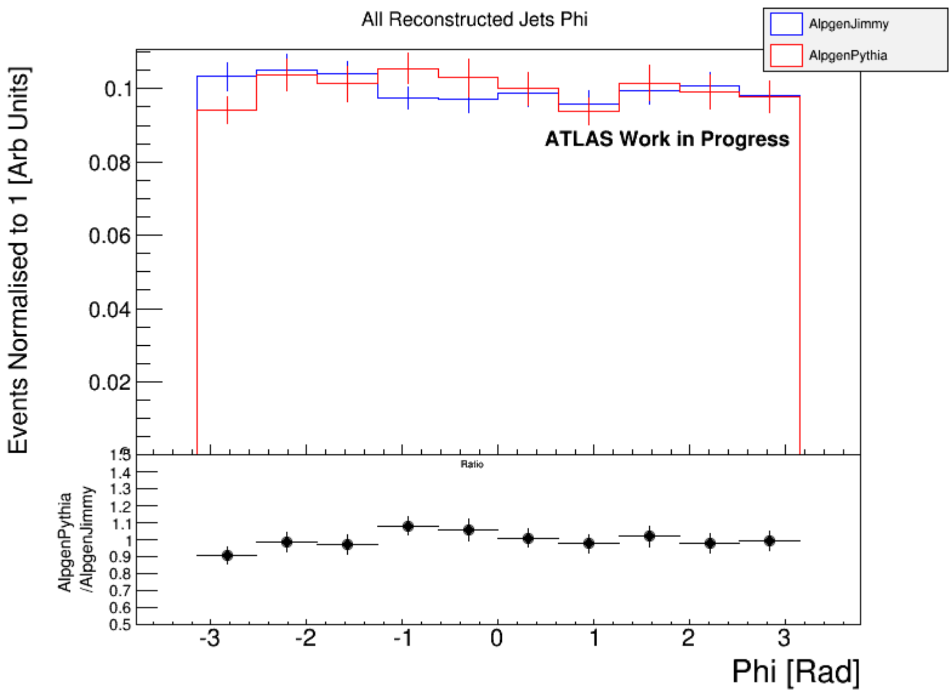
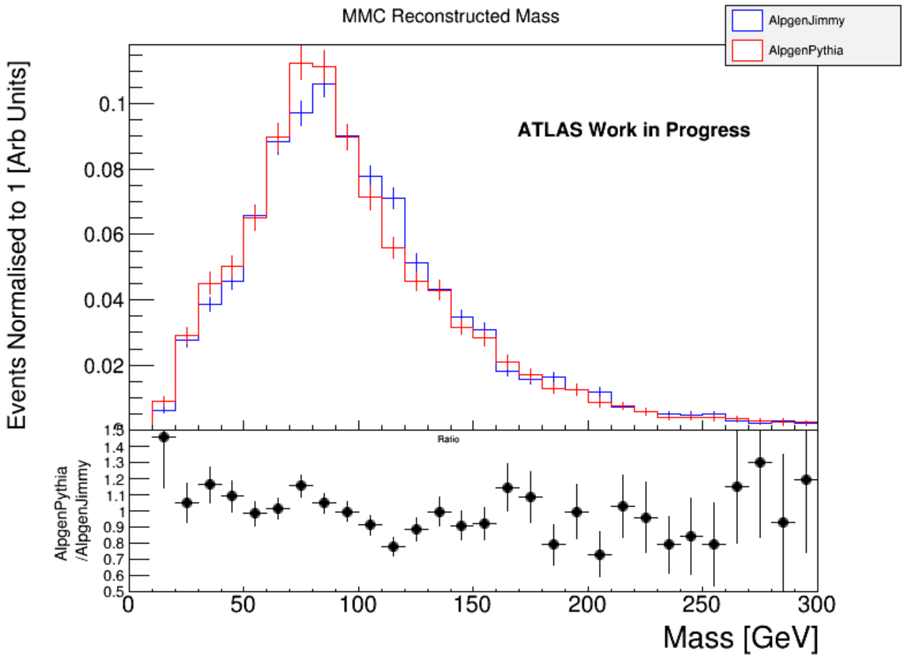
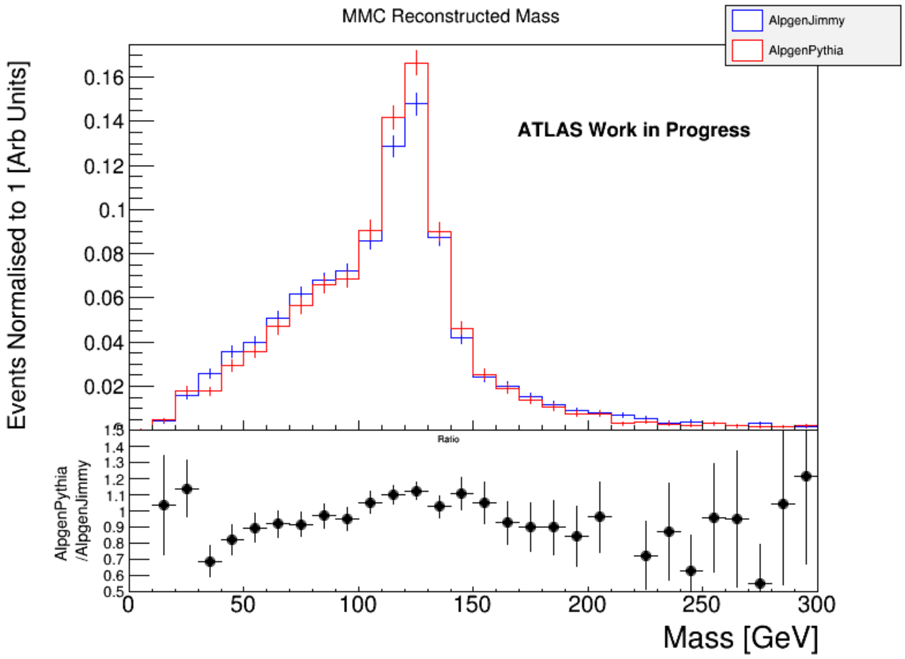


Figure 26: Pre-selection $Z \to ee$ reconstructed jet ϕ distribution.

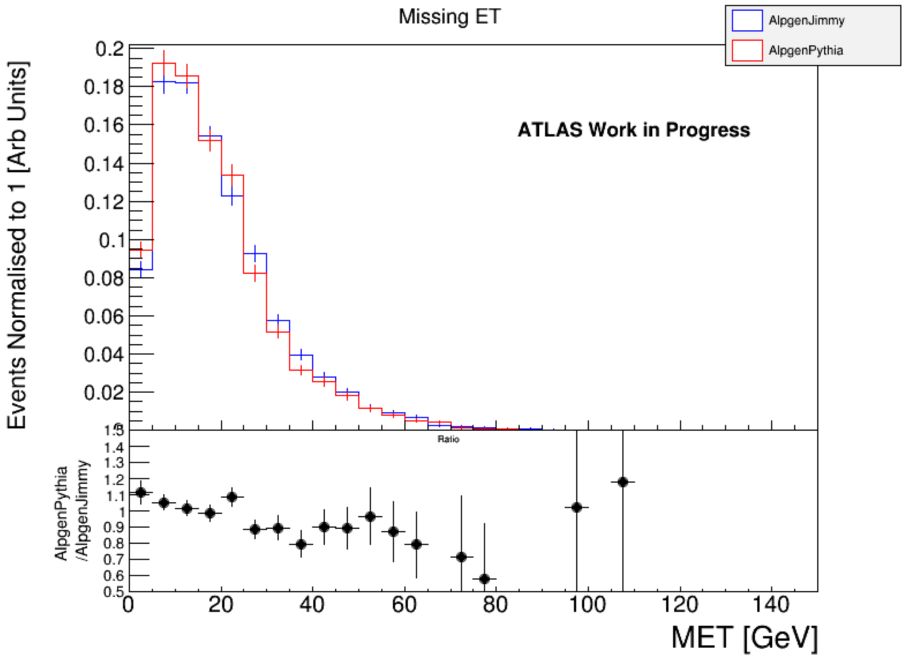


(a) $W \to e\nu$ m_{MMC} distribution.



(b) $Z \to ee$ m_{MMC} distribution.

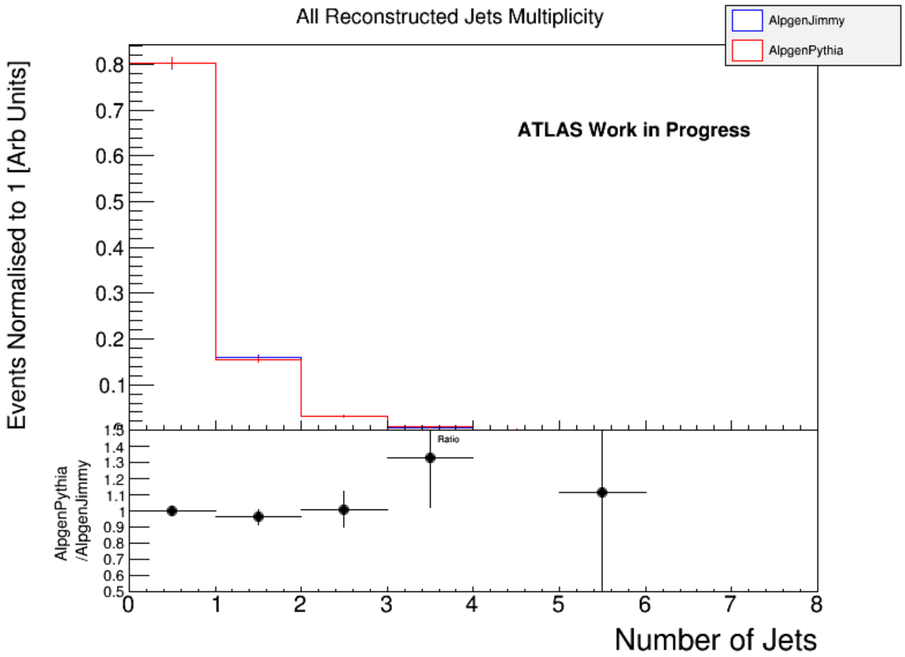
Figure 27: Pre-selection level m_{MMC} distributions.



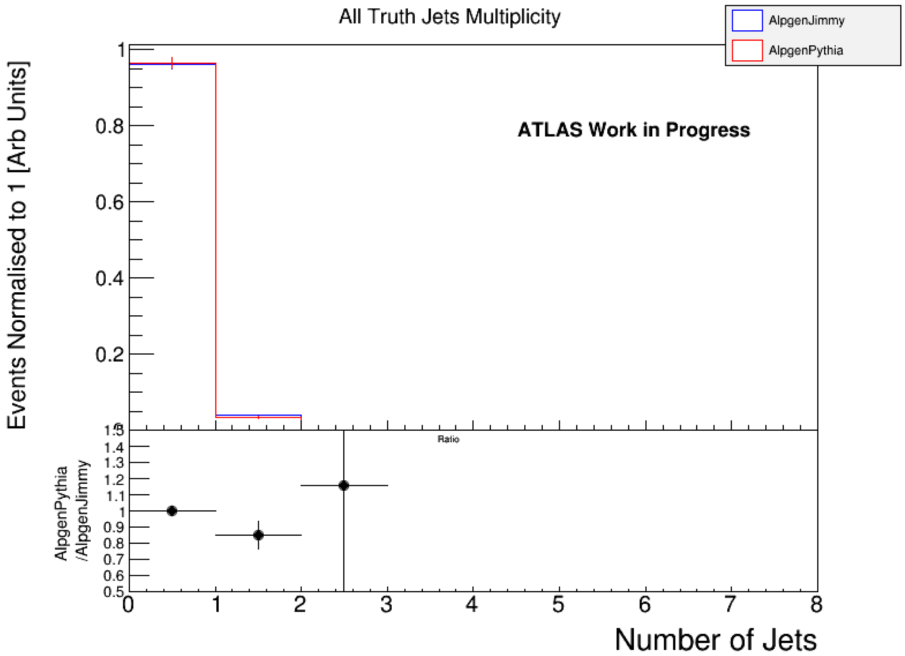
(a) $Z \rightarrow ee$ E_T^{miss} distribution.

Figure 28: Pre-selection level E_T^{miss} distributions.

Moving to the plots of reconstructed variables that are also produced using MC truth information. First, looking at the jet multiplicity most plots show good correlation but some large differences can be seen in the $W \rightarrow e\nu$ and the $Z \rightarrow ee$ samples but the differences are within uncertainties. The p_T and η plots all are acceptable with some minor discrepancies within uncertainties. The plots at this level are all considered acceptable with the minor differences being considered acceptable in exchange for the benefit of the higher statistics sample. There are however a few other questions that need answering when looking at these plots. Firstly when comparing the truth level jet multiplicity at pre-selection with the reconstructed jet multiplicity it is easy to see they are not consistent with each other as seen in Figure 29. Another question was raised about the unexpected peaks in the reconstructed $Z \rightarrow ee$ η plots in the range of approximately $2.5 < |\eta| < 3$, as seen in Figure 30, and needed to be investigated.



(a) Pre-selection $Z \to \mu\mu$ reconstructed jet multiplicity distribution.



(b) $Z \to \mu\mu$ truth jet multiplicity distribution.

Figure 29: Pre-selection level jet truth and reconstructed jet multiplicity distributions.

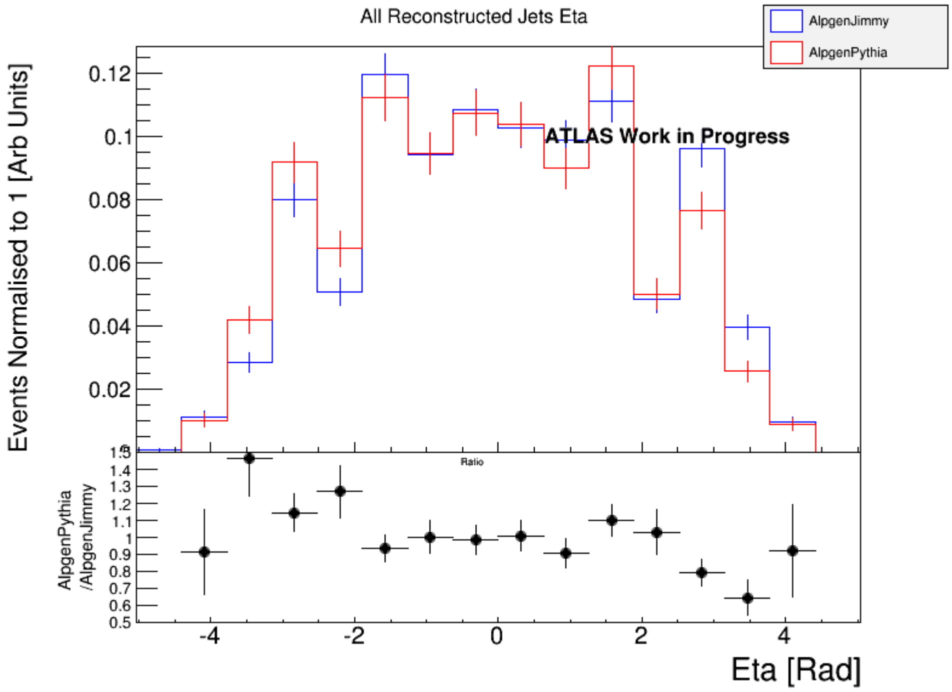


Figure 30: Pre-selection $Z \to ee$ reconstructed jet η distribution.

After some investigation the reason behind the η peaks was discovered as being a result of the overlap removal recommendations. Before overlap removal occurs electrons in the "crack region" of the electromagnetic calorimeter are excluded. This means that jets do not undergo overlap removal with electrons in this region which leads to electrons in this region being incorrectly identified as jets. In electron heavy processes, such as seen in the $Z \to ee$ sample, this results in a large number of incorrectly identified jets in this region. This was not considered a problem however since the $Z \to ee$, which is a negligible background in this analysis, is the only sample that is noticeably affected by the uncovered problem.

The difference in the jet multiplicity plots was suggested to be a possible result of pileup jets in the events. To investigate this a strategy of performing truth matching on the jets at the pre-selection level was proposed. The truth matching process involves taking all reconstructed jets at pre-selection level and comparing them against all the initial truth jets. A reconstructed jet is considered to be matched to a truth jet if $\Delta R < 0.2$. The result of the truth matching process can be seen in the truth matched jet multiplicity plots in Figure 31. This plot shows similar results to those of the truth level plots at pre-selection with few to no events with more than a single jet being present at pre-selection. The conclusion was drawn that the differences in the jet multiplicity plots between reconstructed jets and truth jets are due to the presence of pileup jets. All the truth matched jet multiplicity plots can be found in the Appendix of this thesis.

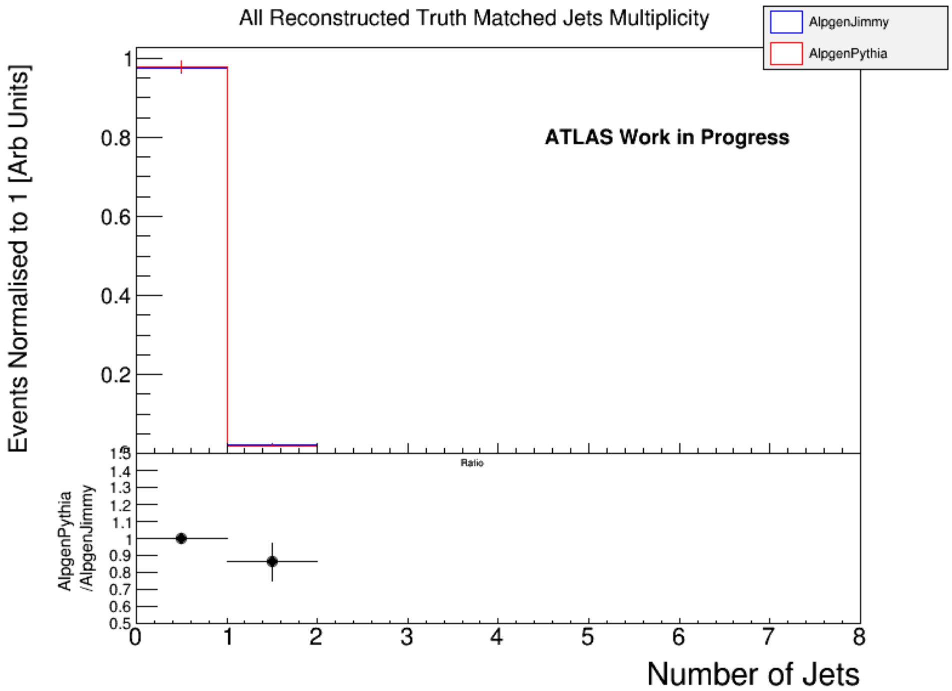


Figure 31: Pre-selection $Z \to \mu\mu$ truth matched jet multiplicity distribution.

As a final decision the $H \to \tau\tau$ analysis groups moved to using the Pythia samples. Whilst differences could be seen in the results they were generally covered by error bars or the differences were considered negligible and the need for a sample with better statistics could permit these small differences in the results.

7 Monte Carlo versus Embedding trigger treatment

This section presents a study investigating differences between MC and Embedding trigger treatment on $Z \rightarrow \tau\tau$ MC simulated data. A discussion is presented on how Embedding data is created to highlight how it differs from MC simulated data before the details of the study and the results are shown.

7.1 τ-Embedding data

An irreducible background process in the $H \rightarrow \tau\tau$ analysis which is difficult to model is $Z \rightarrow \tau\tau$. Firstly, $Z \rightarrow \tau\tau$ process is badly modelled by MC simulations. Further, identification of τ decays is a difficult task which leads to the result that high-statistic $Z \rightarrow \tau\tau$ samples cannot be directly estimated using data-driven techniques either. However, $Z \rightarrow \mu\mu$ decays are similar to $Z \rightarrow \tau\tau$ decays and can be used as a basis for modelling $Z \rightarrow \tau\tau$ events. These $Z \rightarrow \mu\mu$ decays result in two isolated, high-energy muons with opposite charge and are a near signal-free signature in the $H \rightarrow \tau\tau$ analysis. Using embedding techniques these muons are replaced with simulated τ decays. The τ-lepton kinematic properties are taken from the replaced muons and the decays are simulated using MC techniques. This approach is justified due to the well-understood decays of Z-bosons and τ leptons. Whilst these embedding samples are primarily used for analysis purposes, they are also used for comparison to try improve the $Z \rightarrow \tau\tau$ MC modelling accuracy.

7.2 Trigger treatments

Since MC and embedding are not the same modelling process, different trigger treatments exist for these two modelling techniques. This study investigates discrepancies seen between using the two treatments on a single MC sample. Both treatments weight events based on which one of four types of trigger combinations fire, but do so differently. The types of trigger combinations are:

- Muon Single Lepton Trigger (Muon SLT)
- Muon Combined Trigger (Muon CT)

- Electron SLT
- Electron CT

The same trigger selection criteria as explained in Chapter 6 are used for both embedding and MC but the events are weighted differently between the two treatments.

7.3 Monte Carlo versus Embedding

Whilst embedding samples model $Z \to \tau\tau$ process with far greater accuracy than MC simulations, the embedding samples have no simulation of the ATLAS trigger system. Selection criteria and trigger efficiencies are applied to events in an effort to estimate the ATLAS trigger. This trigger treatment may however bias the shape of kinematic variables in the analysis. The $H \to \tau\tau$ analysis group is looking into making corrections to cater for these biases. The purpose of the study presented here is to investigate what biases exist in the $H \to \tau_{lep}\tau_{had}$ analysis. These biases can be investigated by applying both the embedding and MC trigger treatment to the $Z \to \tau\tau$ MC sample and investigating discrepancies in different kinematic distributions. The following kinematic distributions were investigated:

- m_{MMC} - The reconstructed invariant $\tau\tau$ mass reconstructed using MMC technique.
- $|\Delta\phi|$ - The separation in ϕ between the τ_{lep} and τ_{had}.
- $\Delta\phi_{\text{Sum}}$ - This is defined as

$$\Delta\phi_{\text{Sum}} = \left|\Delta\phi_{\text{MET}}^{lep}\right| + \left|\Delta\phi_{\text{MET}}^{had}\right|, \tag{7.1}$$

  where $\Delta\phi_{\text{MET}}^{lep(had)}$ is $\Delta\phi$ calculated between the $\tau_{lep(had)}$ and the MET vectors.
- m_{vis} - The visible mass of the $\tau\tau$ decay.
- p_T^{lep} - The p_T of the τ_{lep}.

These distributions were investigated at pre-selection level in the $H \to \tau_{lep}\tau_{had}$ analysis just as was used in Chapter 6. The distributions across both trigger treatments are normalised before making comparisons since the biases in the distribution shapes are the focus of this study.

7.4 Results

The results presented here show the discrepancies found in the specified distributions using the different trigger treatments. The distributions can be seen in Figures 32-36. As expected, all distributions show discrepancies and reveal that the embedding trigger treatment does bias the shapes of these kinematic distributions. The results shown here will be used to help make corrections to the embedding trigger treatment through producing scaling factors to be applied to events that pass this trigger treatment.

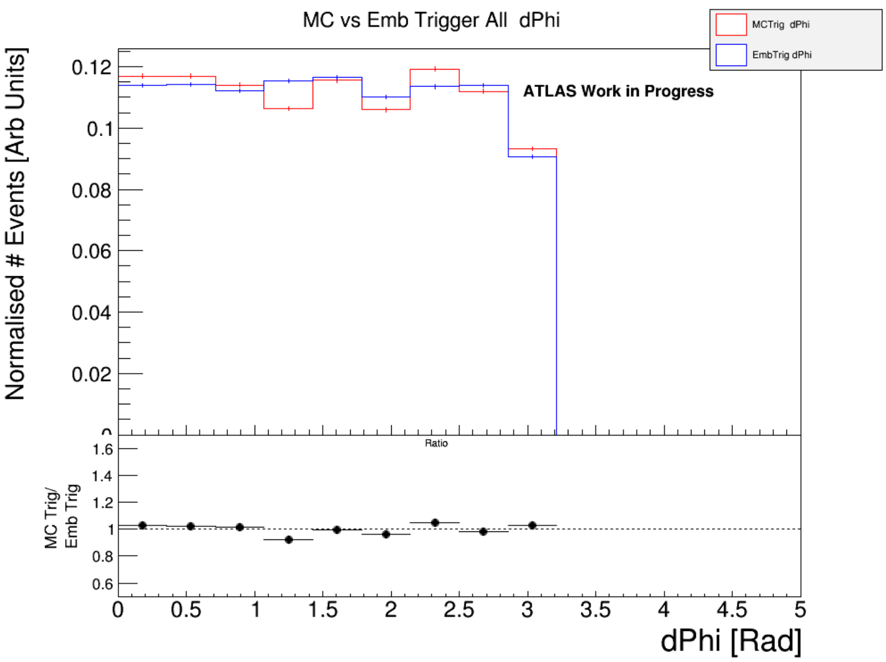


Figure 32: MC versus embedding trigger treatment $\Delta\phi$ distribution for all events passing pre-selection.

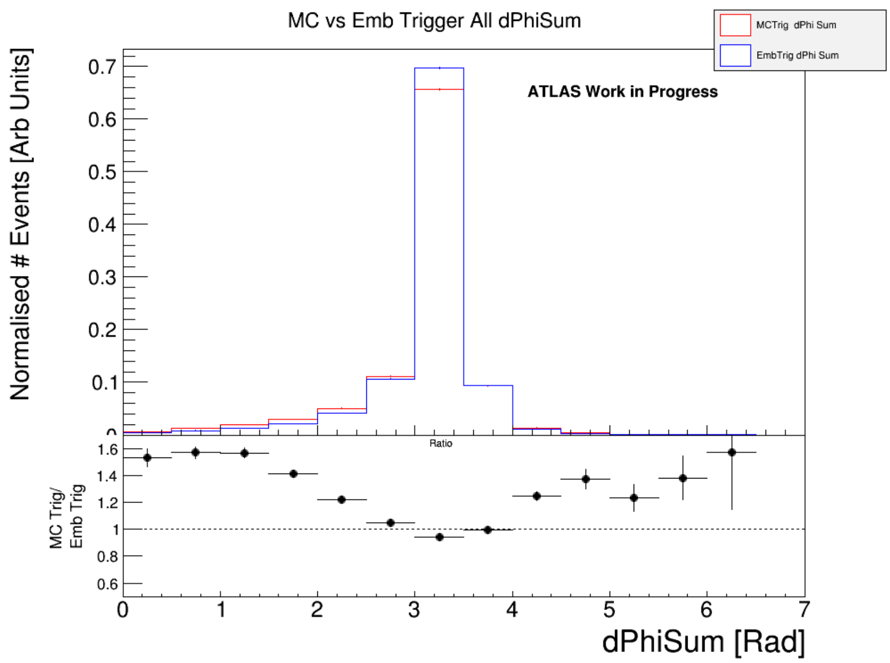


Figure 33: MC versus embedding trigger treatment $\Delta\phi_{\text{Sum}}$ distribution for all events passing pre-selection.

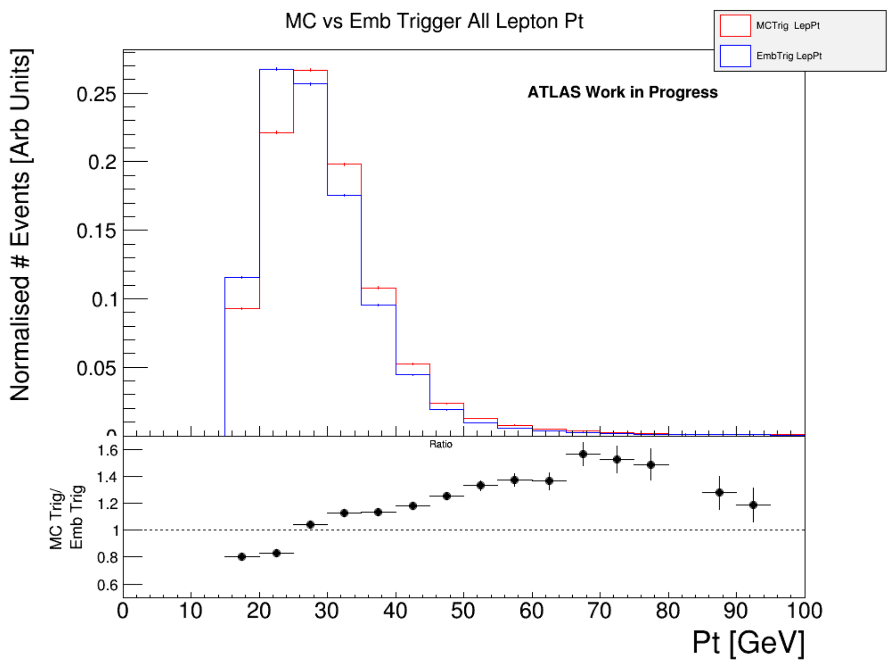


Figure 34: MC versus embedding trigger treatment p_T^{lep} distribution for all events passing pre-selection.

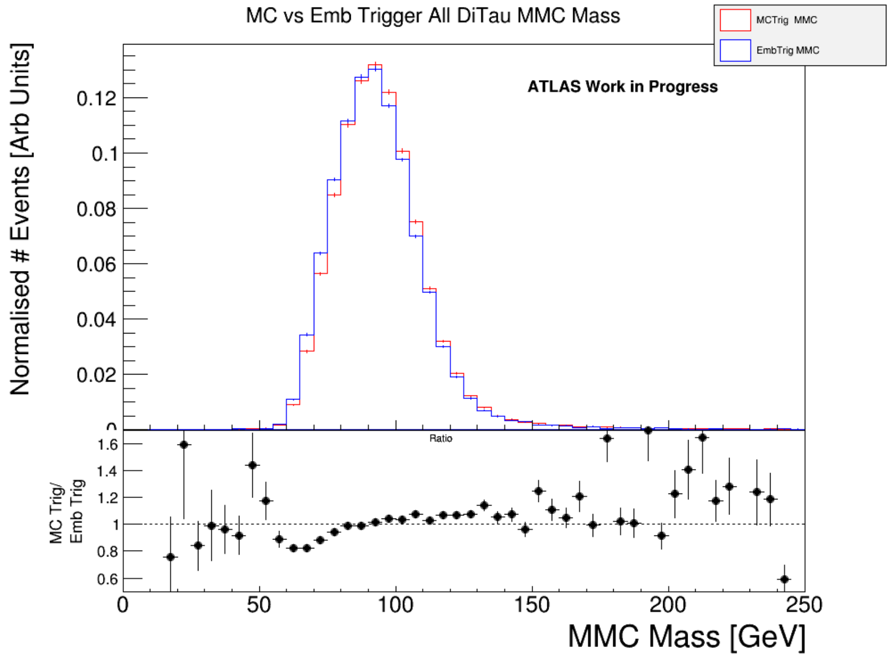


Figure 35: MC versus embedding trigger treatment m_{MMC} distribution for all events passing pre-selection.

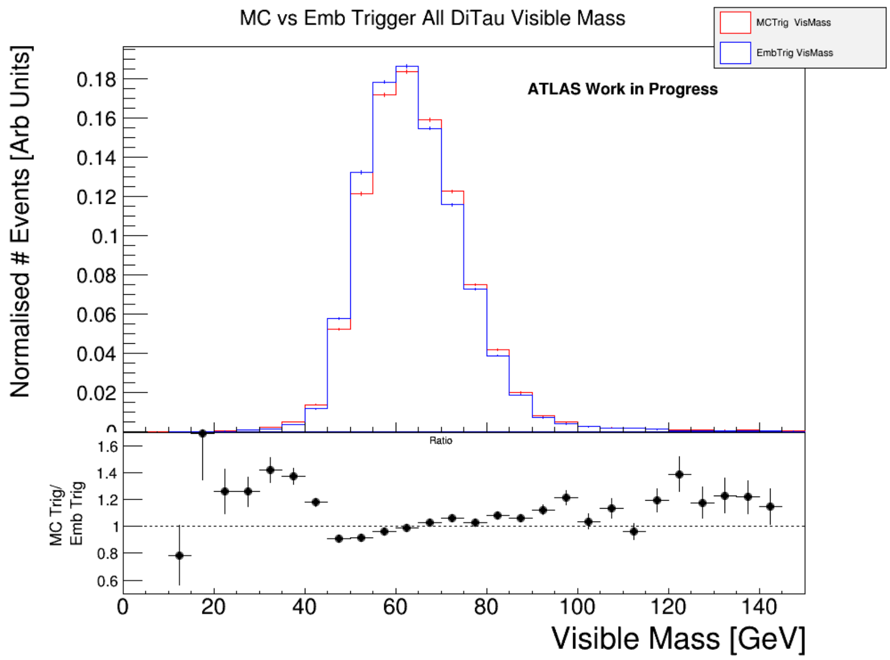


Figure 36: MC versus embedding trigger treatment m_{vis} distribution for all events passing pre-selection.

8 Summary

8.1 Summary of thesis work

This thesis is presented in order to satisfy requirements for a Master degree from the University of the Witwatersrand. The thesis focuses on research done using the ATLAS experiment at CERN. The research contributed to studies managed by the Standard Model $H \rightarrow \tau_{lep}\tau_{had}$ group and the Tau Working Group at ATLAS.

A summary of the LHC is provided along with a detailed description of the ATLAS experiment. This provides the experimental background knowledge necessary to comprehend how and which data are collected from the detector. A theory section outlines the basics of modern particle physics and gives a description of the Standard Model exposing the theoretical reasoning for the Standard Model Higgs boson search. The presentation of how ATLAS Monte Carlo simulations are run concludes the necessary background knowledge required to understand the results of the studies in the thesis.

The workings of the ATLAS Tau Validation package are presented along with a number of goals for improving the package. A solution to automating the validation process in order to reduce the time spent to process the validation tasks is explained. A χ^2 test is implemented which improves the ability of a user to spot problematic variables in the datasets. Several new variables used as input for tau reconstruction algorithms are introduced to the validation. The new and more relevant tau identification algorithms were implemented in the validation package improving the validation's scope. Minor changes were made to the overall structure of the software package so that future maintenance and extensions are possible. The current future of the Tau Validation package is outlined.

Validation studies are shown which facilitated a switch from the Alpgen Jimmy MC samples to the Alpgen Pythia MC samples. The $H \rightarrow \tau_{lep}\tau_{had}$ analysis was implemented up to the pre-selection level. Jet related variables are validated at the truth level before any selection criteria are imposed and again at the truth and reconstructed level at the pre-selection stage of the analysis. Discrepancies are seen across the samples but the values fall within uncertainties. Large differences between truth level and reconstruction level plots at the pre-selection stage of the analysis are investigated using a truth matching approach. The truth matching results show that the discrepancies can be attributed to the pileup effects and validation of the samples is accepted.

The investigation into the MC versus embedding trigger treatment showed that the embedding treatment is not completely accurate at simulating the ATLAS trigger. Biases were shown to exist in all of the kinematic distributions investigated just as was predicted. The differences in shape introduced by this bias will be used to improve embedding efficiencies used in the various $H \rightarrow \tau\tau$ study groups.

8.2 Overview of the ATLAS SM $H \rightarrow \tau_{lep}\tau_{had}$ search

A general overview of the current SM $H \rightarrow \tau_{lep}\tau_{had}$ group analysis is presented here along with the most recent results giving insight into the greater goal that this thesis contributed towards. The analysis strategy used by the group starts with preselection of physics objects. This preselection process was fully described in Chapter 6. Following preselection the events are put through more selection criteria designed to categorize depending on which SM Higgs production mechanism they originate from. Two categories are defined to separate these events:

- VBF: this category is designed to select events with two jets and a large η separation. It will mainly select VBF events but a small number of $gg \rightarrow H$ and VH events are selected too.

- Boosted: this category selects those events that fail VBF selection with $p_T^H >$ 100 GeV. Mainly $gg \rightarrow H$ events are selected in this category but a small number of VBF and VH events are selected too.

However, after categorization, this $H \rightarrow \tau_{lep}\tau_{had}$ signal is not free from background. A BDT multivariate technique is applied to help separate the signal from the background. These background processes include W+Jets, $Z \rightarrow \tau\tau$, $Z \rightarrow ee$, $Z \rightarrow \mu\mu$, $t\bar{t}$ and single-top quark production backgrounds. For each category a separate BDT is specifically optimized to try perform this separation of background and signal. The VBF BDT is trained with only a VBF signal sample where the Boosted BDT is trained with $gg \rightarrow H$, VBF and VH samples. This separated training exploits the event kinematic differences between these two categories improve the accuracy with which these BDTs can categorize the events. The BDT's attempt to identify which events are most signal-like by distributing the events into bins of varying signal purity levels. This purity distribution undergoes a transformation such that all events are given a purity rating between -1 and 1 where 1 represents most signal-like events and -1 represents least signal-like events. These distributions help identify signal-free control regions in data. These control regions are then used to normalise Monte Carlo backgrounds to data.

It is now possible to extract the signal from the data. This involves finding the variable μ, which is known as the signal strength parameter. μ is defined as the ratio of the measured cross section times branching ratio normalised to the Standard Model cross section times the branching ratio for $H \to \tau^+\tau^-$. In this analysis a fitting technique is used to estimate the value of μ. A value of $\mu = 1$ represents the presence of a Higgs boson whilst a value of $\mu = 0$ represents no Higgs boson.

Using the analysis strategy described above encouraging results have been produced in the ATLAS search for a SM Higgs boson decaying to two tau leptons in the lepton-hadron final state. The search contributed to the overall ATLAS search for SM Higgs boson of mass 125 GeV decaying to two tau leptons. The most recent search was performed using the full ATLAS 2012 dataset, which had an integrated luminosity of $\mathcal{L} = 20.3$ fb^{-1} of proton-proton collisions at a center-of-mass energy of 8 TeV. The combined results released in November of 2013 show a signal strength value of $\mu = 1.4^{+0.5}_{-0.4}$ has been measured. This can been seen in Figure 37 where a best fit of $\mu = 1.4$ is found for a Standard Model Higgs boson with $m_H = 125$ GeV. This value of $\mu = 1.4$ corresponds to a measured signal with a significance of 4.1 standard deviations which constitutes strong evidence for the Higgs boson decaying to fermions [12]. When the LHC resumes operation in 2015 [17], new data will be produced and the search will continue and the signal strength parameter will continue to be refined.

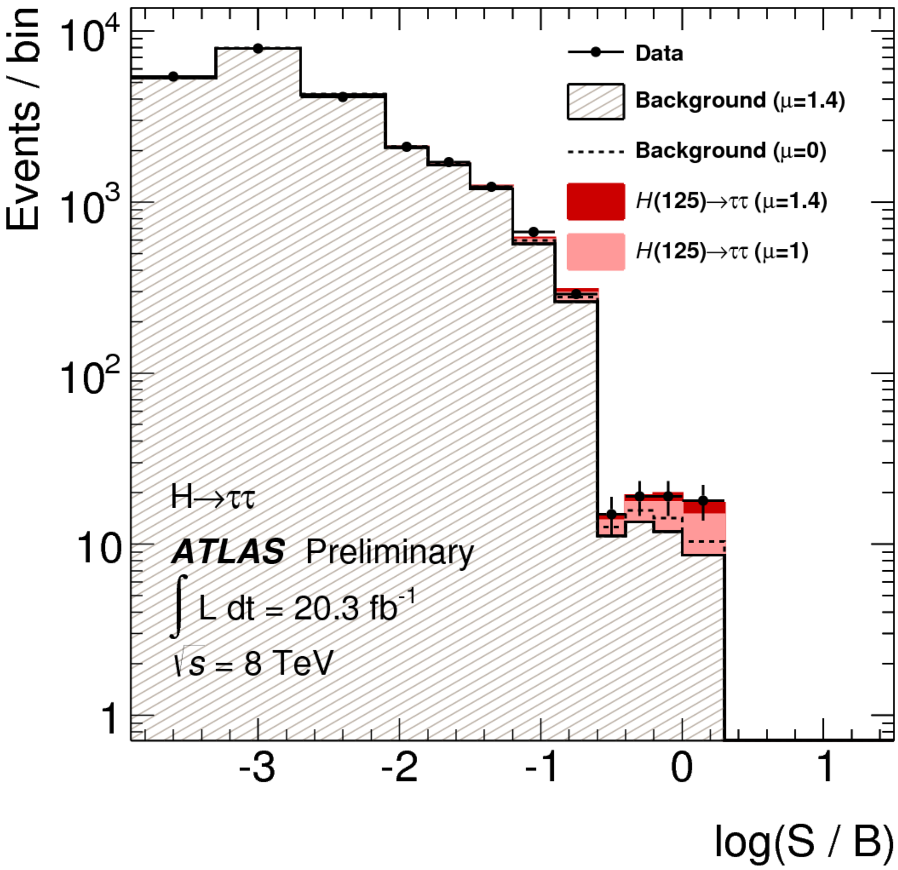


Figure 37: The figure shows a plot of the event yields as a function of $\log{(S/B)}$. Here S is signal yield and B is background yield. It can be seen that a best fit for signal strength gives a value of $\mu = 1.4$ for a Standard Model Higgs with $m_H = 125$ GeV [7].

References

[1] P. W. Higgs, "Broken symmetries, massless particles and gauge fields," *Phys. Rev. Lett. 13*, p. 508, 1964.

[2] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons," *Phys. Rev. Lett. 13*, p. 321, 1964.

[3] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global Conservation Laws and Massless Particles," *Phys. Rev. Lett. 13*, p. 585, 1964.

[4] P. W. Higgs, "Broken symmetries and the Masses of Gauge Bosons," *Phys. Lett. 12*, p. 132, 1964.

[5] ATLAS Collaboration, "Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc," *Phys.Lett. B716 (2012) 1-29*, no. CERN-PH-EP-2012-218, 2012. arXiv:1207.7214v2 [hep-ex].

[6] CMS Collaboration, "Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc," *Phys. Lett. B 716 (2012) 30*, no. CMS-HIG-12-028; CERN-PH-EP-2012-220, 2013. arXiv:1207.7235v2 [hep-ex].

[7] ATLAS Collaboration, "Evidence for Higgs Boson Decays to the $\tau^+\tau^-$ Final State with the ATLAS Detector," no. ATLAS-CONF-2013-108, 2014.

[8] CMS Collaboration, "Evidence for the 125 gev higgs boson decaying to a pair of tau leptons," no. CMS-HIG-13-004, CERN-PH-EP-2014-001, 2014. arXiv:1401.5041v1 [hep-ex].

[9] ATLAS Collaboration, "Search for the bb decay of the Standard Model Higgs boson in associated $(W/Z)\,H$ production with the ATLAS detector," no. ATLAS-CONF-2013-079, 2014.

[10] CMS Collaboration, "Search for the standard model higgs boson produced in association with a w or a z boson and decaying to bottom quarks," *Phys. Rev. D 89, 012003 (2014)*, no. CMS-HIG-13-012, CERN-PH-EP-2013-188, 2014. arXiv:1310.3687v2 [hep-ex].

[11] L. H. C. S. W. Group, "Handbook of lhc higgs cross sections: 1. inclusive observables," no. CERN-2011-002, 2011. arXiv:1101.0593v3 [hep-ph].

[12] ATLAS collaboration, "Evidence for Higgs Boson Decays to the $\tau^+\tau^-$ Final State with the ATLAS Detector," no. ATLAS-CONF-2013-108, ATLAS-COM-CONF-2013-095, 2013.

[13] J. Baglio, "Higgs production at the lhc," *JHEP 1103:055,2011*, no. CERN-PH-TH/2010-289; LPT Orsay 10-89, 2012. arXiv:1012.0530v5 [hep-ph].

[14] ATLAS Collaboration, "Search for the standard model higgs boson in the h to tau+ tau- decay mode in sqrt(s) = 7 tev pp collisions with atlas," *JHEP (2012) 070*, no. CERN-PH-EP-2012-140, 2012. arXiv:1206.5971v1 [hep-ex].

[15] Luke Weston, "The Higgs Boson Explained," 2012. , Retrieved February 26, 2014, from `http://youngausskeptics.com/2012/07/the-higgs-boson-explained/`.

[16] L. H. C. S. W. Group, "Handbook of lhc higgs cross sections: 3. higgs properties," no. CERN-2013-004, 2013. arXiv:1307.1347v2 [hep-ph].

[17] C. Duc, "Long Shutdown 1: Exciting times ahead," 2013. , Retrieved January 7, 2014, from `http://home.web.cern.ch/about/updates/2013/02/long-shutdown-1-exciting-times-ahead`.

[18] CERN, "Why the LHC," 2008. , Retrieved January 7, 2014, from `http://public.web.cern.ch/public/en/LHC/WhyHC-en.html`.

[19] M. Kramer and F. Soler, *The Large Hadron Collider Phenomenolgy.* 2004. ISBN: 0750309865.

[20] CERN, "ATLAS," n.d. , Retrieved December 2, 2013, from `http://home.web.cern.ch/about/experiments/atlas`.

[21] CERN, "ATLAS: Detector and physics performance technical design report. Volume 1," no. CERN-LHCC-99-14, ATLAS-TDR-14, 1999.

[22] CERN, "ATLAS: Detector and physics performance technical design report. Volume 2," no. CERN-LHCC-99-15, ATLAS-TDR-15, 1999.

[23] ATLAS Collaboration, "Expected performance of the atlas experiment - detector, trigger and physics," 2009. arXiv:0901.0512v4 [hep-ex].

[24] ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider," *Journal of Instrumentation*, vol. 3, no. 08, 2008.

[25] "ATLAS Photos," n.d. , Retrieved Novemeber 13, 2013, from `http://www.atlas.ch/photos/`.

[26] ATLAS Collaboration, "Performance of the atlas trigger system in 2010," *Eur.Phys.J.C 72 (2012) 1849*, no. CERN-PH-EP-2011-078, 2011. arXiv:1110.1530v2 [hep-ex].

[27] F. Huegging, "The atlas pixel detector," *IEEE Trans.Nucl.Sci. 53 (2006) 1732-1736*, no. BONN-IP-2004-12, 2005. arXiv:physics/0412138v2 [physics.ins-det].

[28] ATLAS Collaboration, "The atlas inner detector commissioning and calibration," *Eur.Phys.J.C70:787-821,2010*, no. CERN-PH-EP-2010-043, 2010. arXiv:1004.5293v2 [physics.ins-det].

[29] V. A. Mitsou, "Atlas silicon microstrip tracker: operation and performance," *Astroparticle, Particle, Space Physics and Detectors for Physics Applications, World Scientific, pp. 636-640 (2012)*, no. ATL-INDET-PROC-2011-020, 2011. arXiv:1110.1983v1 [physics.ins-det].

[30] S. Y. Smirnov, "Trtviewer: the atlas trt detector monitoring and diagnostics tool," no. CERN report no: ATL-INDET-PROC-2011-044, 2012. arXiv:1208.0502v1 [physics.ins-det].

[31] E. Hines, "Performance of particle identification with the atlas transition radiation tracker," 2011. arXiv:1109.5925v1 [physics.ins-det].

[32] M. Mandl, "Transition-Radiation Tracker," 2000. , Retrieved November 18, 2013, from `http://www.martin-mandl.com/thesis/thesis-8.html`.

[33] C. W. Fabjan and F. Gianotti, "Calorimetry for particle physics," *Rev. Mod. Phys.*, vol. 75, pp. 1243–1286, 2003.

[34] J. Proudfoot, "Hadron Calorimetry at the LHC," no. SSI-2006-L012, 2006.

[35] N. Nikiforou, "Performance of the atlas liquid argon calorimeter after three years of lhc operation and plans for a future upgrade," 2013. arXiv:1306.6756v2 [physics.ins-det].

[36] B. S.-M. Peralva, "Calibration and performance of the atlas tile calorimeter," 2013. arXiv:1305.0550v1 [physics.ins-det].

[37] J. Beringer et al. (Particle Data Group), "Particle Data Group summary tables," *Phys. Rev. D86, 010001 (2012) and 2013 partial update for the 2014 edition.*, 2013.

[38] G. Aad et al, "Commissioning of the ATLAS Muon Spectrometer with Cosmic Rays," *Eur.Phys.J.*, vol. C70, pp. 875–916, 2010.

[39] Y. Assran, "Transport properties of operational gas mixtures used at lhc," 2011. arXiv:1110.6761v1 [physics.ins-det].

[40] C. Eck et al. , *LHC computing Grid: Technical Design Report. Version 1.06 (20 Jun 2005).* Technical Design Report LCG, CERN, 2005.

[41] G. Kane, *Modern Elementary Particle Physics.* 1993. ISBN: 0201624605.

[42] D. Griffiths, *Introduction to Elementary Particles.* 1987. ISBN: 0471603864.

[43] W. N. Cottingham and D. A. Greenwood, *An introduction to the Standard Model of Particle Physics.* 2007. ISBN: 0521852498.

[44] S. Dawson, "Introduction to electroweak symmetry breaking," no. BNL-HET-99/1, 1999. arXiv:hep-ph/9901280v1.

[45] M. Robinson, *Symmetry and the Standard Model: Mathematics and Particle Physics.* 2011. ISBN: 1441982663.

[46] J. F. Donoghue, E. Golowich, and B. R. Holstein, *Dynamics of the Standard Model.* 1994. ISBN: 0521476526.

[47] S. Roth, *Precision Electroweak Physics at Electron-Positron Colliders.* 2007. ISBN: 3540351647.

[48] W. Hollik, "Quantum field theory and the standard model," *CERN Yellow Report CERN-2010-002, pp.1-44*, 2010. arXiv:1012.3883v1 [hep-ph].

[49] G. Altarelli, "Collider physics within the standard model: a primer," no. RM3-TH/13-1, CERN-PH-TH/2013-020, 2013. arXiv:1303.2842v2 [hep-ph].

[50] M. M. M. Gomez-Bock, M. Mondragn and M. Spira, "Concepts of electroweak symmetry breaking and higgs physics," no. CERN-PH-TH/2007-262, DESY 07-214, LAPTH-CONF-1223/2007, LPT-ORSAY 07-128, PITHA 07/20, PSI-PR-07-11, 2007. arXiv:0712.2419v1 [hep-ph].

[51] P. Langacker, *The Standard Model and Beyond.* 2011. ISBN: 1420079069.

[52] M. Cattaneo et al., "Status of the GAUDI event-processing frame-work," *Computing in High Energy and Nuclear Physics 2001 Conference*, 2001.

[53] G. Barrand et al., "GAUDI-a software architecture and framework for building LHCb data processing applications," *Computing in High Energy and Nuclear Physics 2000 Conference*, 2000.

[54] L. Lonnblad, "CLHEP: a project for designing a C++ class library for high-energy physics," *Comput. Phys. Commun. 84*, 1994.

[55] M. Dobbs, "The HepMC C++ Monte Carlo event record for high energy physics," *Comput. Phys. Commun. 134*, 2001.

[56] T. Sjostrand, S. Mrenna, and P. Skands, "PYTHIA 6.4 physics and manual.," *High Energy Phys. 05*, 2006.

[57] C. Anatopoulos et al., "Physics analysis tools for beauty physics in ATLAS," *Phys. Conf. 119*, 2008.

[58] T. Sjostrand, S. Mrenna, and P. Skands, "A brief introduction to PYTHIA 8.1," *Comput. Phys. Commun. 178*, vol. 2008.

[59] C. Corcella et al., "HERWIG 6: an event generator for hadron emission reactions with interfering gluons (including super symmetric processes)," *High Energy Phys. 01*, 2001.

[60] M. Bahr et al., "Herwig++ physics and manual," *Eur. Phys. J. C 58*, 2008.

[61] T. Gleisberg, "Sherpa 1.alpha, a proof-of-concept version," *High Energy Phys. 02*, 2004.

[62] M. Gyulassy and X. N. Wang, "HIJING 1.0: a Monte Carlo program for parton and particle production in high-energy hadronic and nuvlear collisions," *Comput. Phys. Commun. 83*, 1994.

[63] M. Mangano et al., "ALPGEN, a generator for hard multiparton processes in hadronic collision," *High Energy Phys. 07*, 2003.

[64] S. Frixione, P. Nason, and B. Webber, "Matching NLO QCD and parton showers in heavy flavour production," *High Energy Phys. 08*, 2003.

[65] C. Oleari, "The powheg-box," *Nucl.Phys.Proc.Suppl.205-206:36-41,2010*, 2010. arXiv:1007.3893v1 [hep-ph].

[66] B. Kersevan and E. Rcihter-Was, "The Monte Carlo event generator AcerMC version 2.0 with interfaces to PYTHIA 6.2 and HERWIG 6.5," 2004.

[67] S. Jadach, J. Kuhn, and Z. Was, "TAUOLA: a Library of Monte Carlo programs to simulate decays of polarized tau leptons," *Comput. Phys. Commun. 64*, 1990.

[68] E. Barberio, B. van Eijk, and Z. Was, "PHOTOS: a universal Monte Carlo for QED radiative corrections in decays," *Comput. Phys. Commun. 66*, 1991.

[69] D. Lange, "The EvtGen particle decay simulation package," *Nucl. Instrum. Methods A 462*, 2001.

[70] F. Paige, S. Protopopescu, H. Baer, and X. Tata, "ISAJET 7.69: a Monte Carlo event generator for p p, anti-p pm and e+ e- reactions," p. 2003.

[71] M. Leyton, "Minimum bias and underlying event measurements with atlas," no. ATL-PHYS-PROC-2012-033, 2012. arXiv:1202.2090v1 [hep-ex].

[72] J. Butterworth, J. Foreshaw, and M. Seymour, "N," *Z. Phys. C 72*, pp. 637–646, 1996.

[73] E. Boos et al., "Generaic user process interface for event generators," *Proceedings of Les Houches 2001: Physics at TeV colliders*, 2001.

[74] J. Alwall et al., "A standard format for Les Houches event files," *Comput. Phys. Commun. 176*, 2007.

[75] J. Boudreau and V. Tsulaia, "The GeoModel toolkit for detector description," *Computing in High Energy and Nuclear Physics 2004*, 2004.

[76] M. Asai and R. Mount, "Geant4 – Current and Future : A Snowmass 2013 White Paper," 2013. arXiv:1308.1994v1 [hep-ex].

[77] The ROOT Team, "About ROOT," 1995. , Retrieved December 15, 2013, from `http://root.cern.ch/drupal/content/about`.

[78] L. Wolfgang, "Fast Simularion for ATLAS: Atlfast-II and ISF," *J. Phys.: Conf Ser. 396*, 2012.

[79] "Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons," Tech. Rep. ATLAS-CONF-2011-077, CERN, 2011.

[80] "Identification of the Hadronic Decays of Tau Leptons in 2012 Data with the ATLAS Detector," Tech. Rep. ATLAS-CONF-2013-064, CERN, 2013.

[81] M. Branco, D. Cameron, B. Gaidioz, V. Garonne, B. Koblitz, M. Lassnig, R. Rocha, P. Salgado, and T. Wenaus, "Managing ATLAS data on a petabyte-scale with DQ2," *Journal of Physics: Conference Series*, vol. 119, no. 6, 2008.

[82] "Python Subprocess," n.d. , Retrieved December 15, 2013, from `http://docs.python.org/2/library/subprocess.html`.

[83] "Latex Beamer class," n.d. , Retrieved March 22, 2014, from `https://bitbucket.org/rivanvx/beamer/wiki/Home`.

[84] Phil McLean, "The Chi-Square Test," 2000. , Retrieved January 18, 2014, from `http://www.ndsu.edu/pubweb/~mcclean/plsc431/mendel/mendel4.htm`.

[85] Encyclopedia of Mathematics, "Kolmogorov–Smirnov test," 2012. , Retrieved January 18, 2014, from `http://www.encyclopediaofmath.org/index.php?title=Kolmogorov%E2%80%93Smirnov_test&oldid=22659`.

[86] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *NUMERICAL RECIPES: The Art of Scientific Computing.* third ed., 2007. ISBN: 0521880688.

[87] ATLAS Collaboration, "Local Hadronic Calibration," Tech. Rep. ATL-LARG-PUB-2009-001-2. ATL-COM-LARG-2008-006. ATL-LARG-PUB-2009-001, CERN, Geneva, 2008.

[88] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-kt jet clustering algorithm," *JHEP 04*, 2008.

[89] P. A.Elagin and A.Pranko, "A new mass reconstruction technique for resonances decaying to di-tau," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, Volume 654, Issue 1, 21 October 2011, Pages 481-489*, 2011. arXiv:1012.4686v2 [hep-ex].

[90] G. Aad et al., "Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 7$ TeV with ATLAS," *The European Physical Journal C - Particles and Fields 72*, 2012.

A Appendix

This appendix provides all extra necessary plots as described in the Pythia vs Herwig validation studies section of this thesis. Only the plots related to this study with acceptable statistics are shown here. Particularly sub-leading and sub-sub-leading truth jet related distributions show very low statistics and are not included.

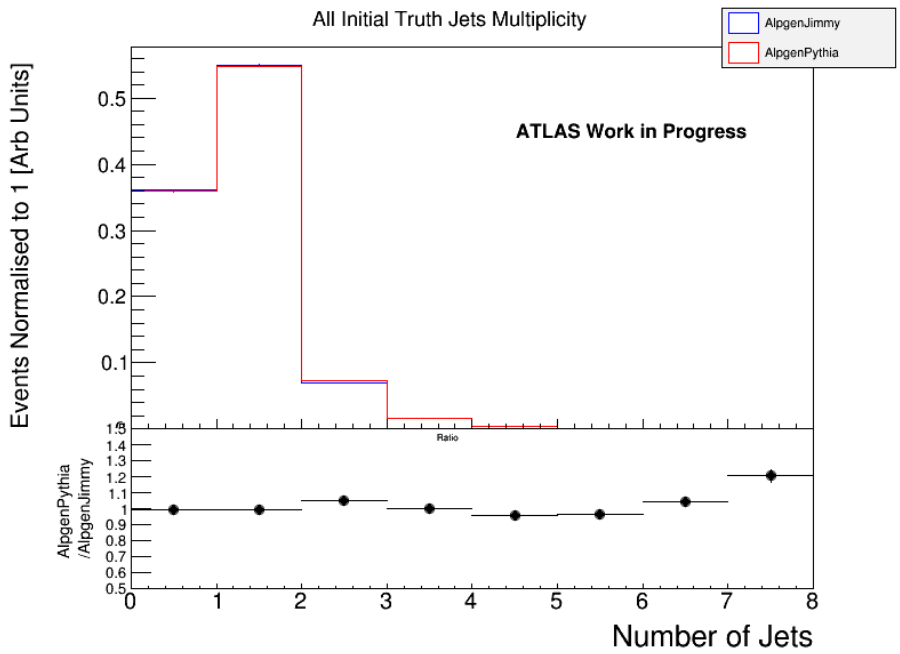


Figure 38: $W \rightarrow e\nu$ truth jet multiplicity.

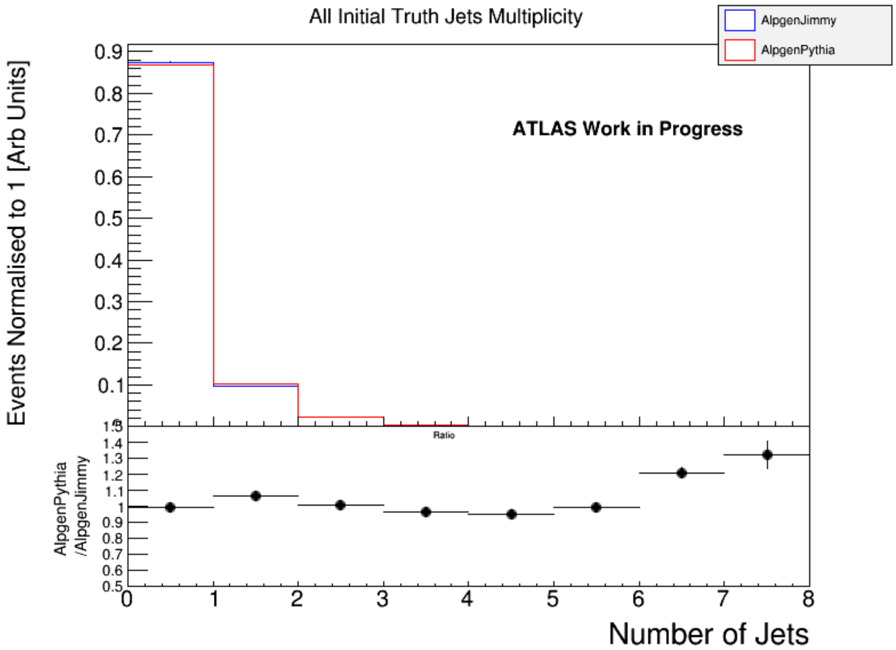


Figure 39: $W \rightarrow \mu\nu$ truth jet multiplicity.

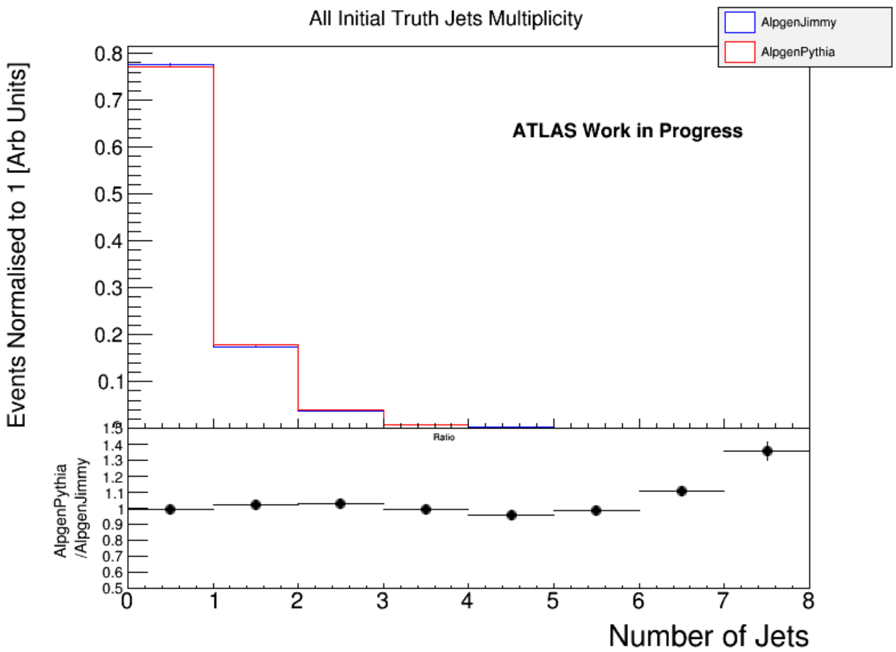


Figure 40: $W \rightarrow \tau\nu$ truth jet multiplicity.

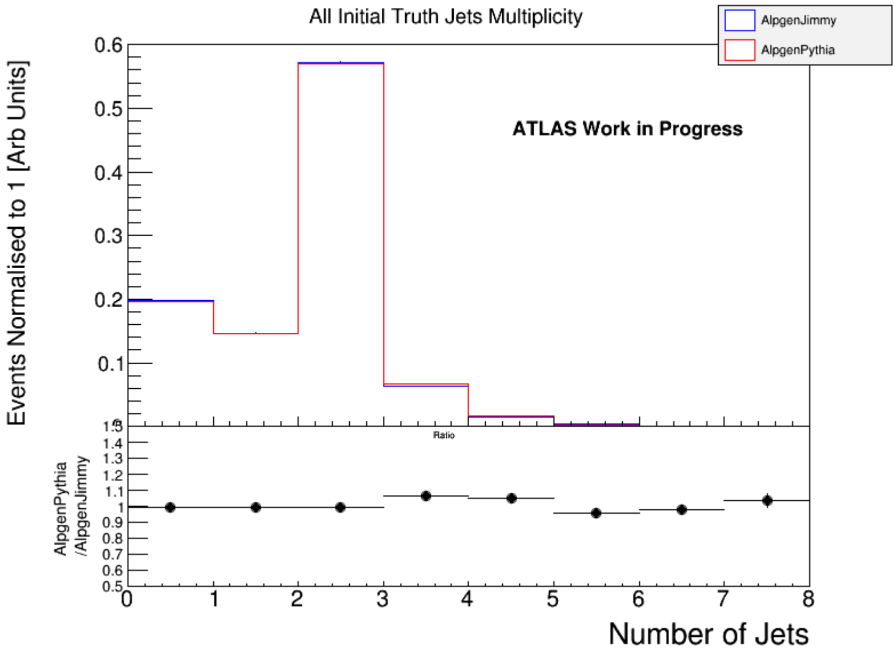


Figure 41: $Z \to ee$ truth jet multiplicity.

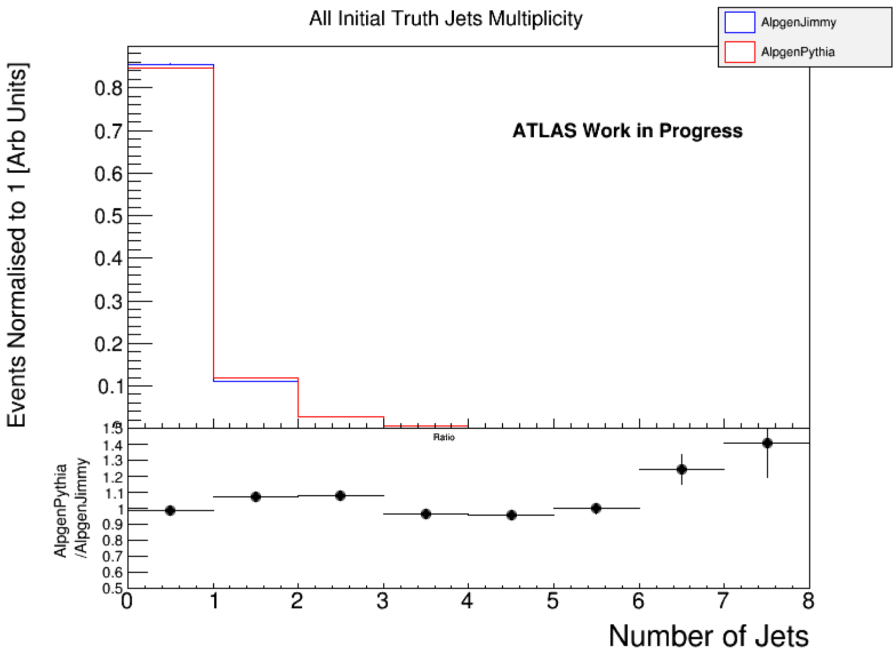


Figure 42: $Z \to \mu\mu$ truth jet multiplicity.

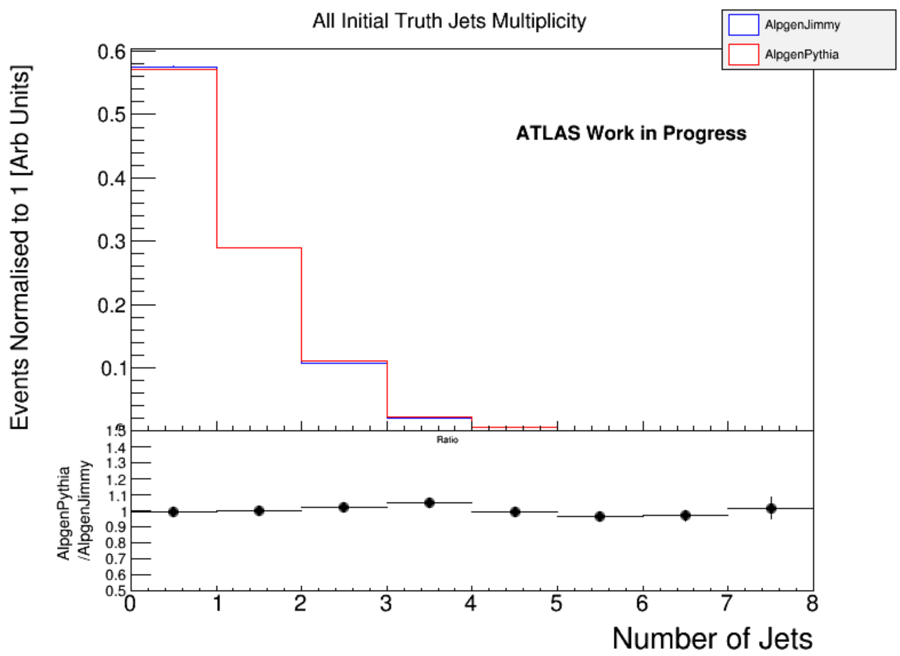


Figure 43: $Z \to \tau\tau$ truth jet multiplicity.

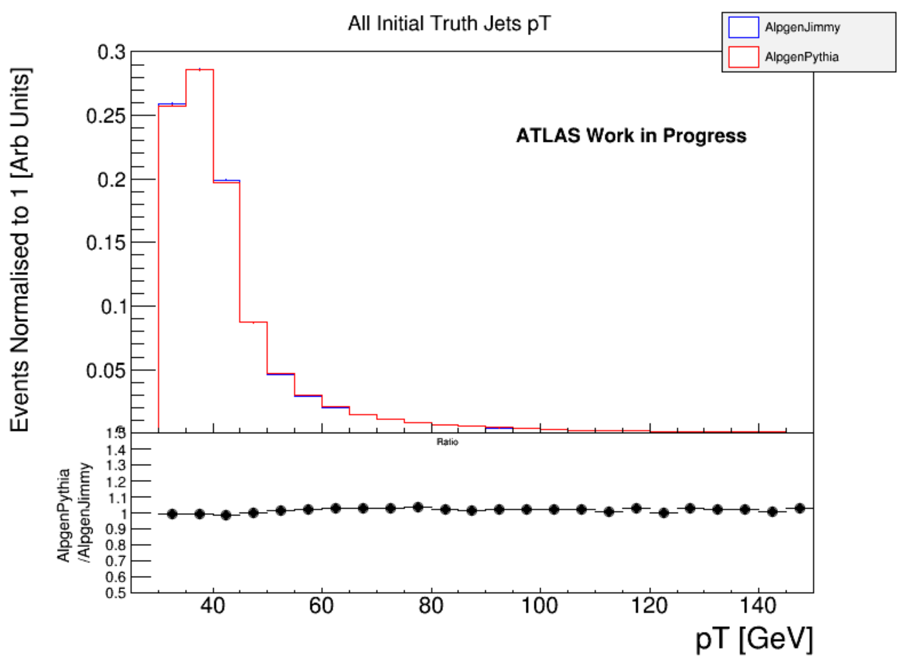


Figure 44: $W \to e\nu$ truth jet p_T.

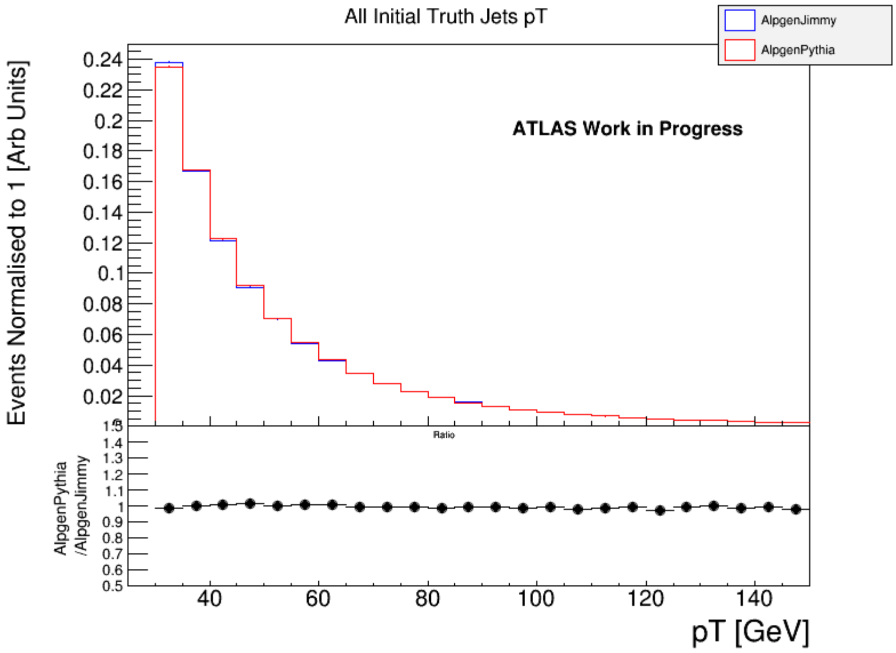


Figure 45: $W \rightarrow \mu\nu$ truth jet p_T.

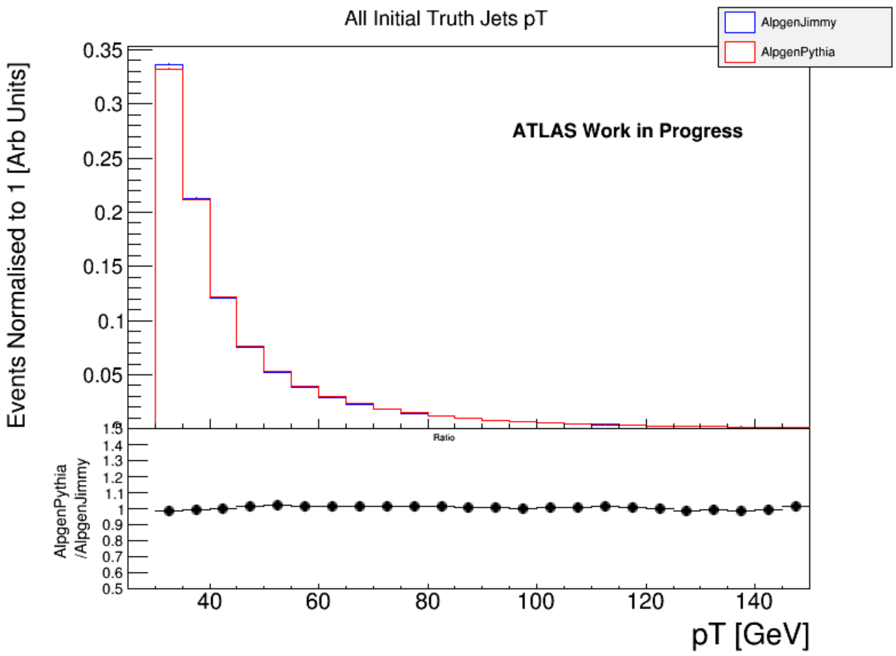


Figure 46: $W \rightarrow \tau\nu$ truth jet p_T.

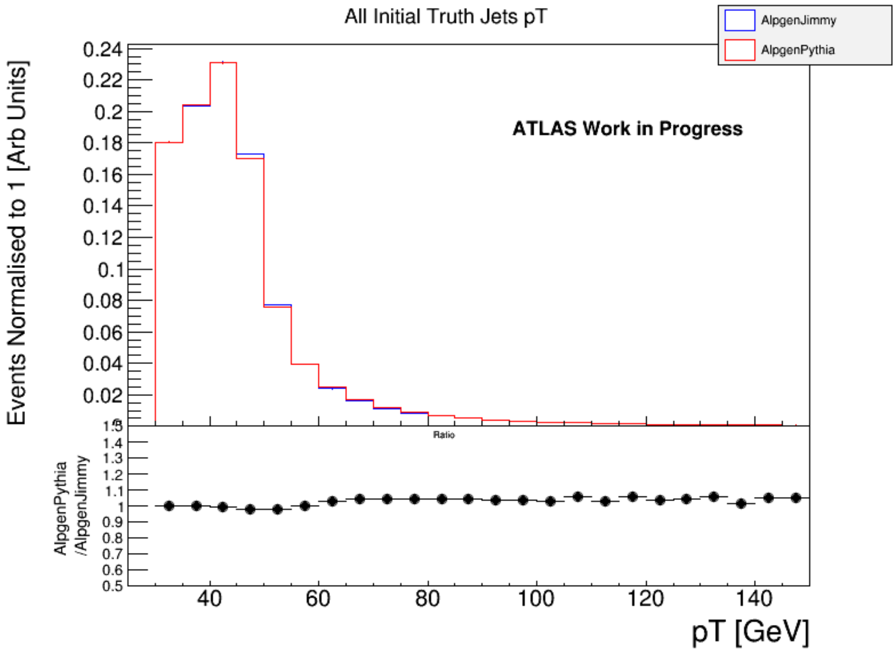


Figure 47: $Z \to ee$ truth jet p_T.

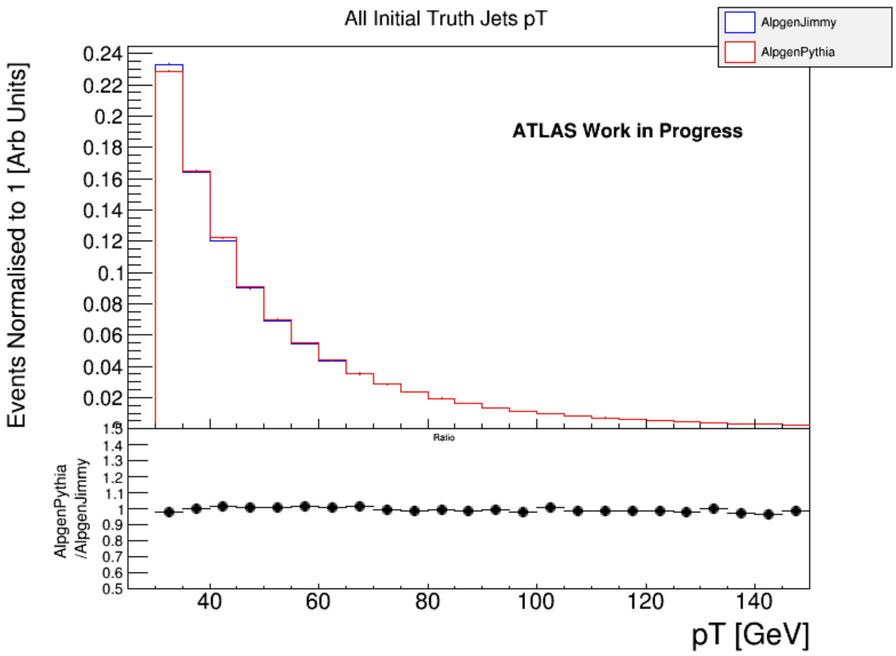


Figure 48: $Z \to \mu\mu$ truth jet p_T.

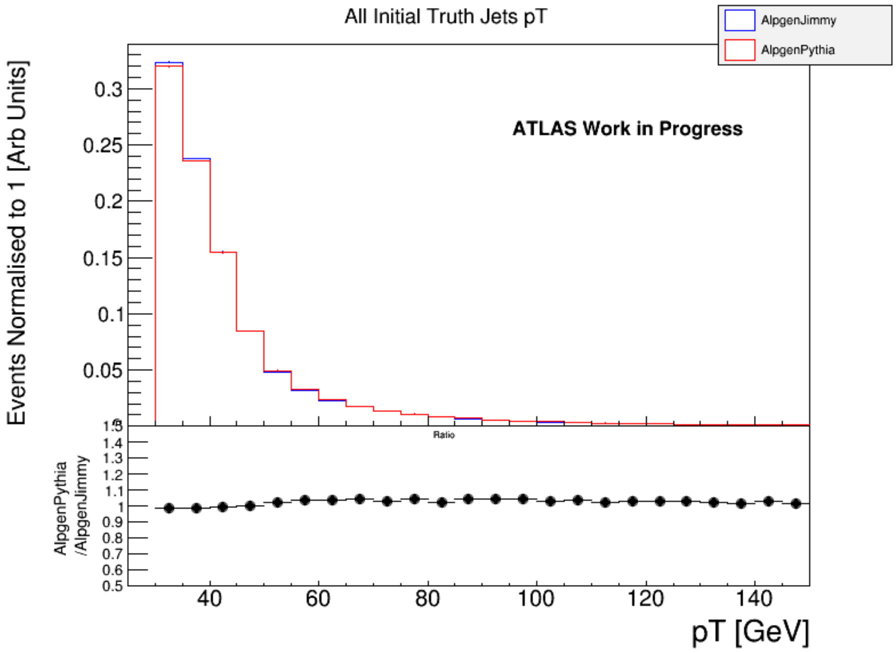


Figure 49: $Z \to \tau\tau$ truth jet p_T.

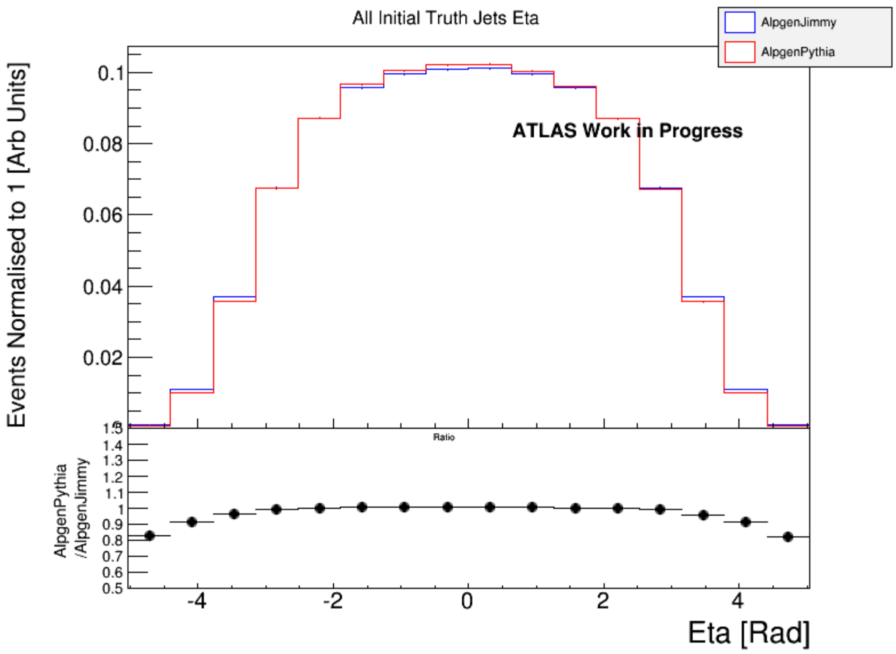


Figure 50: $W \to e\nu$ truth jet η.

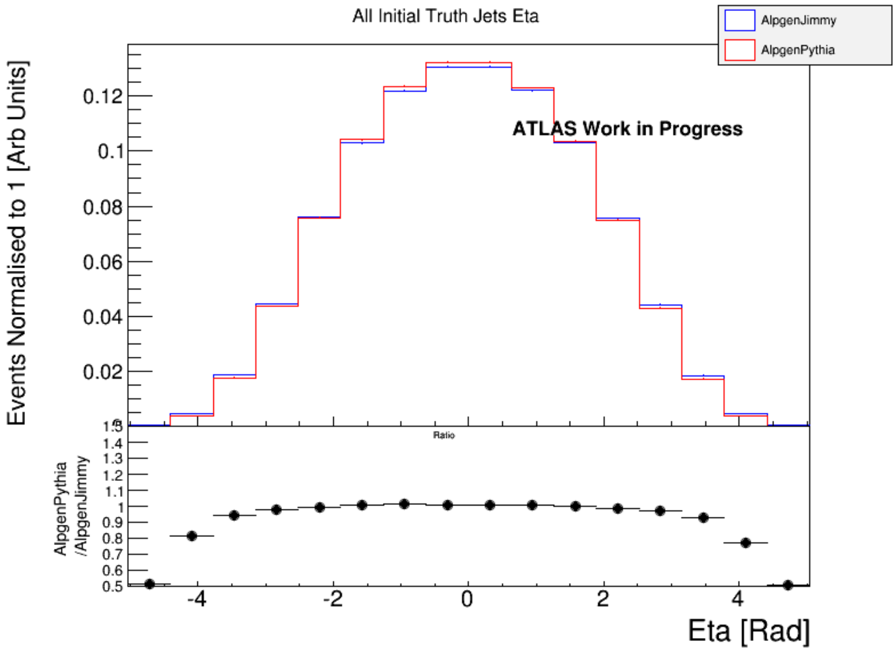


Figure 51: $W \to \mu\nu$ truth jet η.

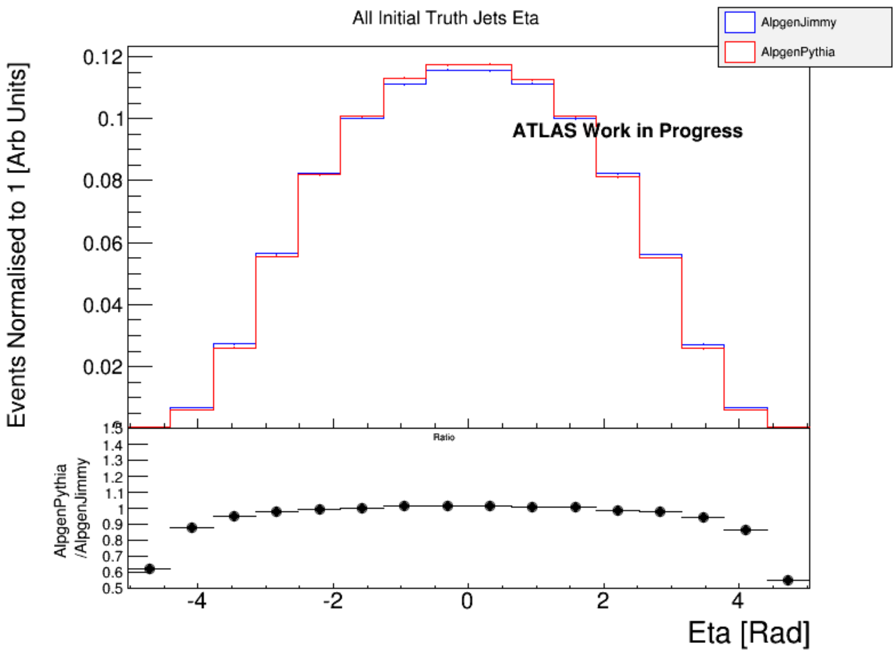


Figure 52: $W \to \tau\nu$ truth jet η.

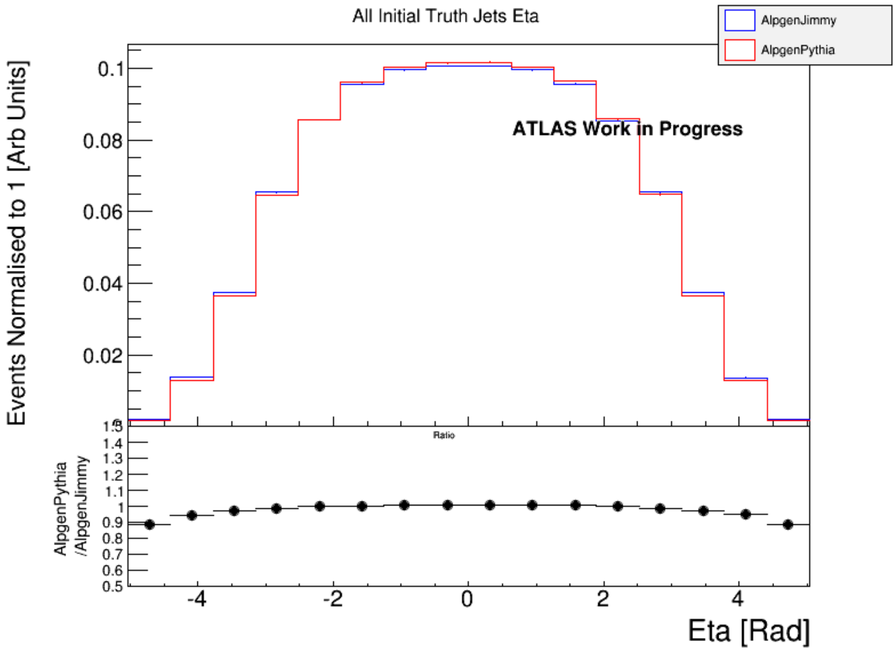


Figure 53: $Z \rightarrow ee$ truth jet η.

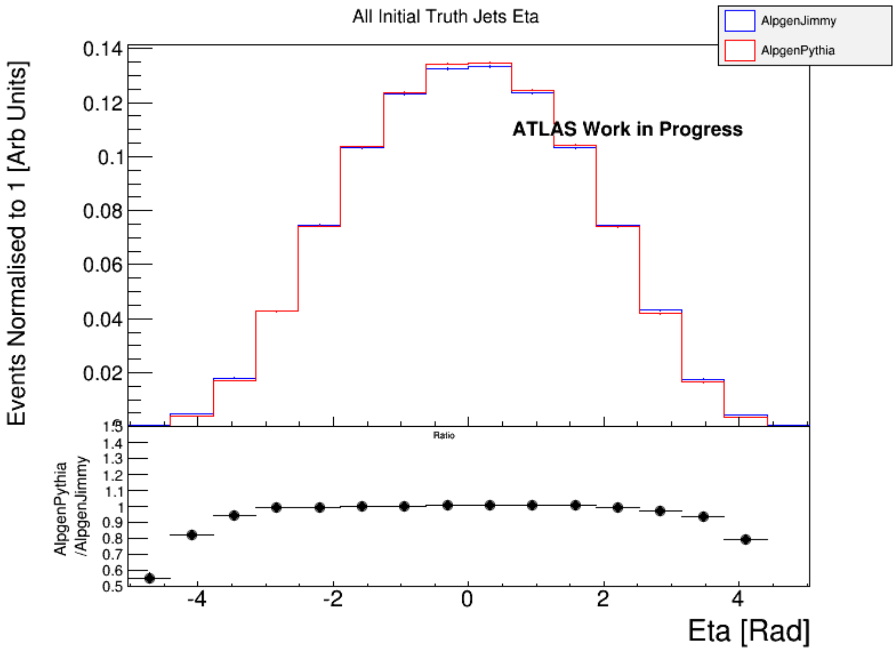


Figure 54: $Z \rightarrow \mu\mu$ truth jet η.

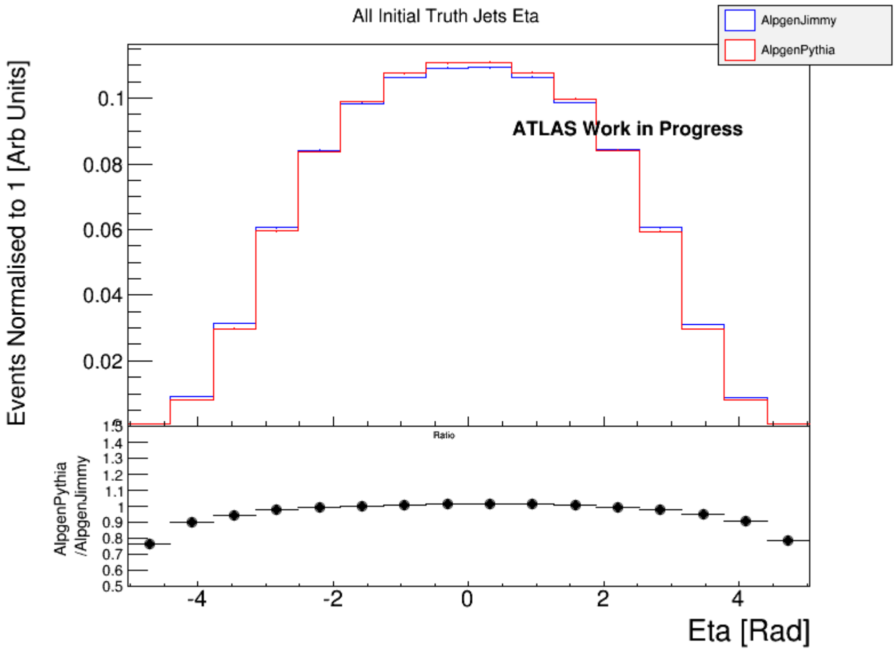


Figure 55: $Z \to \tau\tau$ truth jet η.

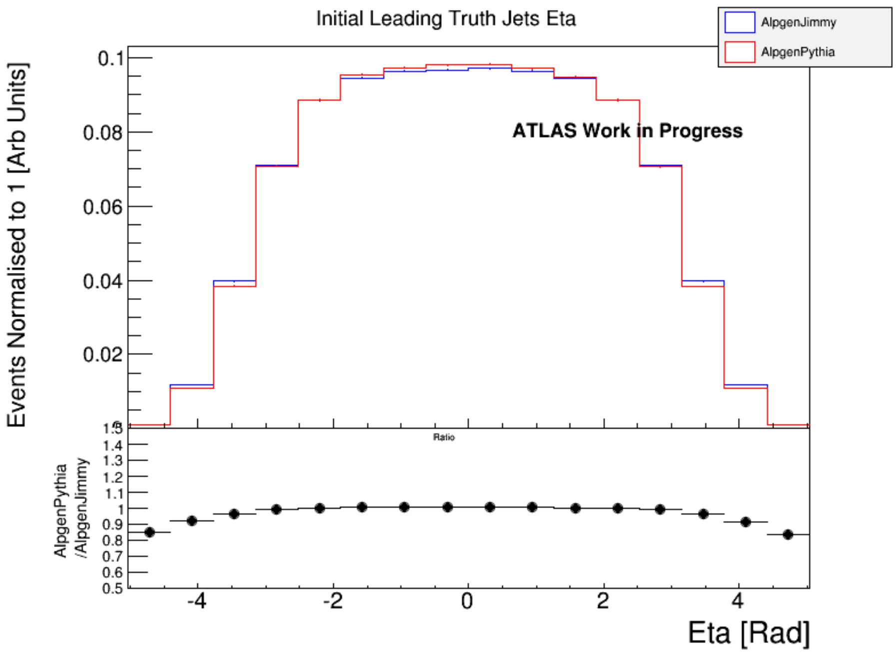


Figure 56: $W \to e\nu$ truth leading jet η.

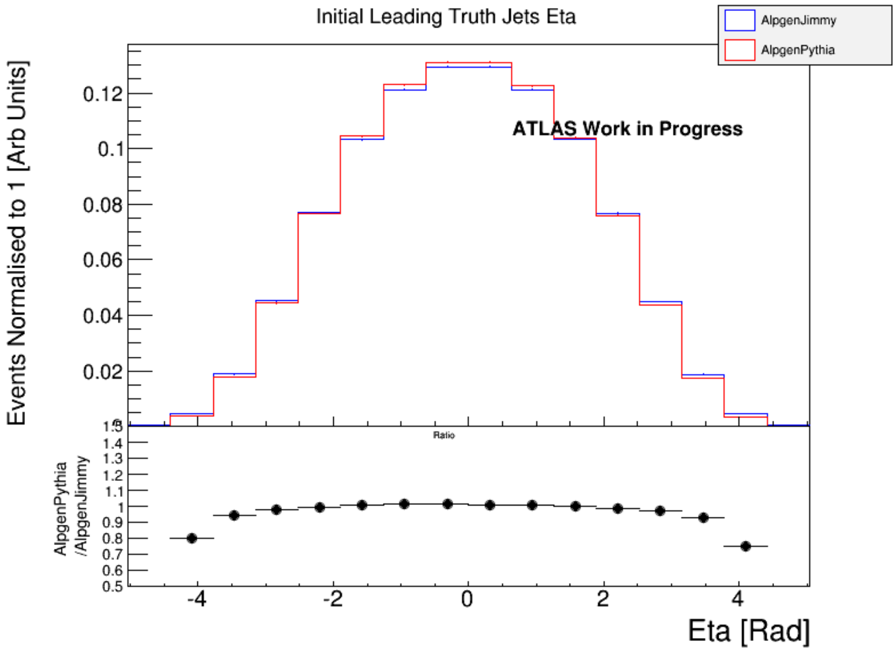


Figure 57: $W \to \mu\nu$ truth leading jet η.

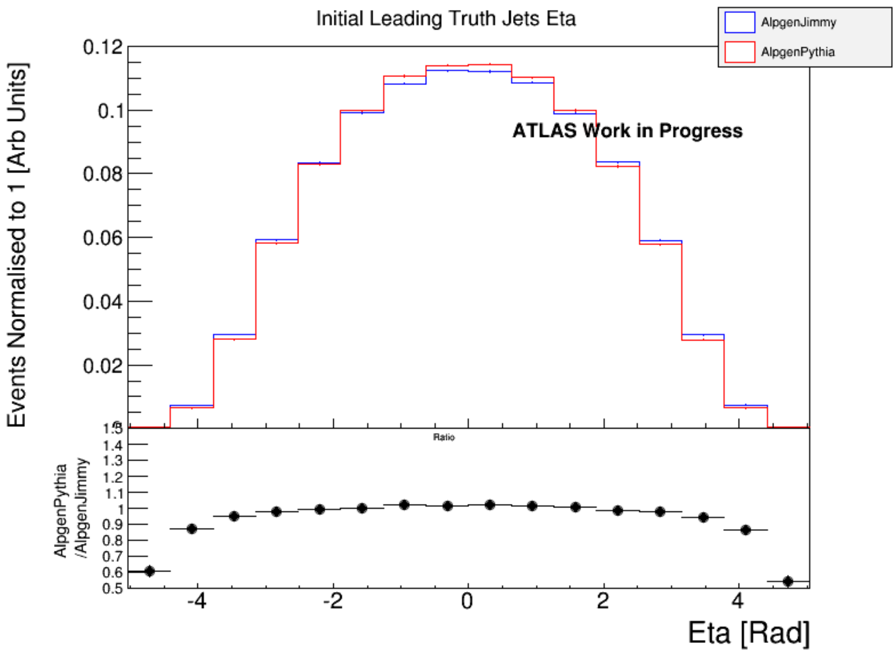


Figure 58: $W \to \tau\nu$ truth leading jet η.

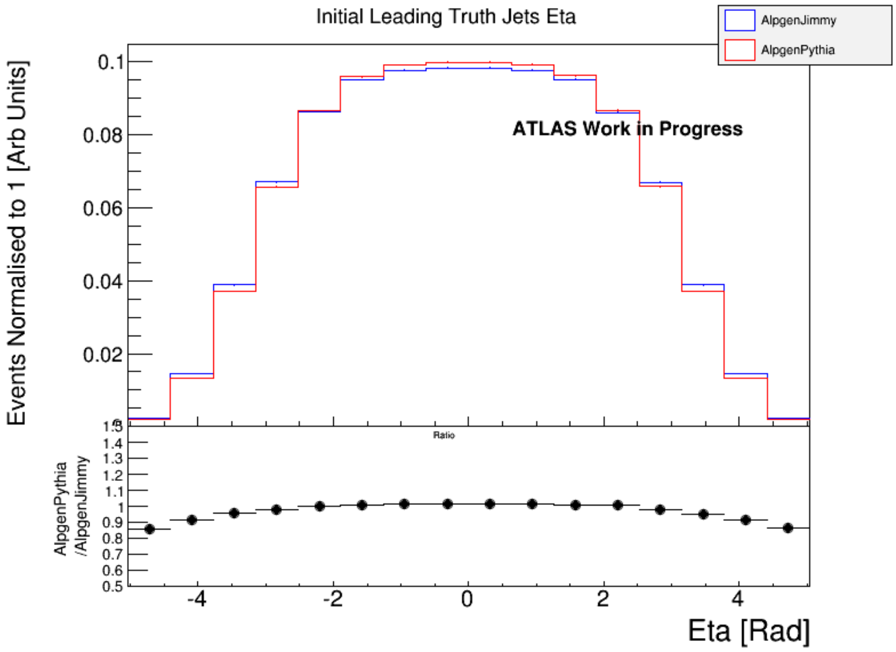


Figure 59: $Z \to ee$ truth leading jet η.

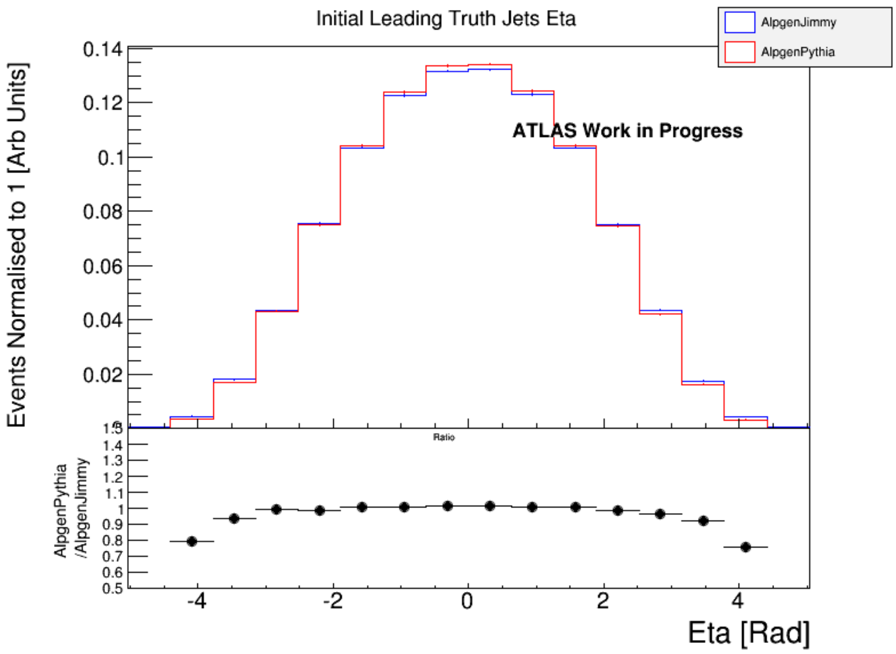


Figure 60: $Z \to \mu\mu$ truth leading jet η.

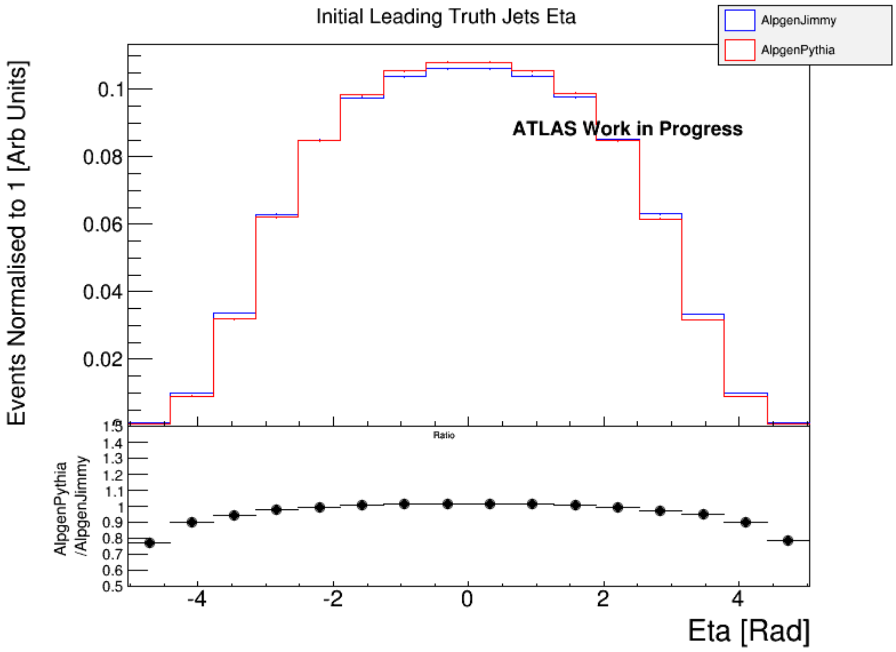


Figure 61: $Z \to \tau\tau$ truth leading jet η.

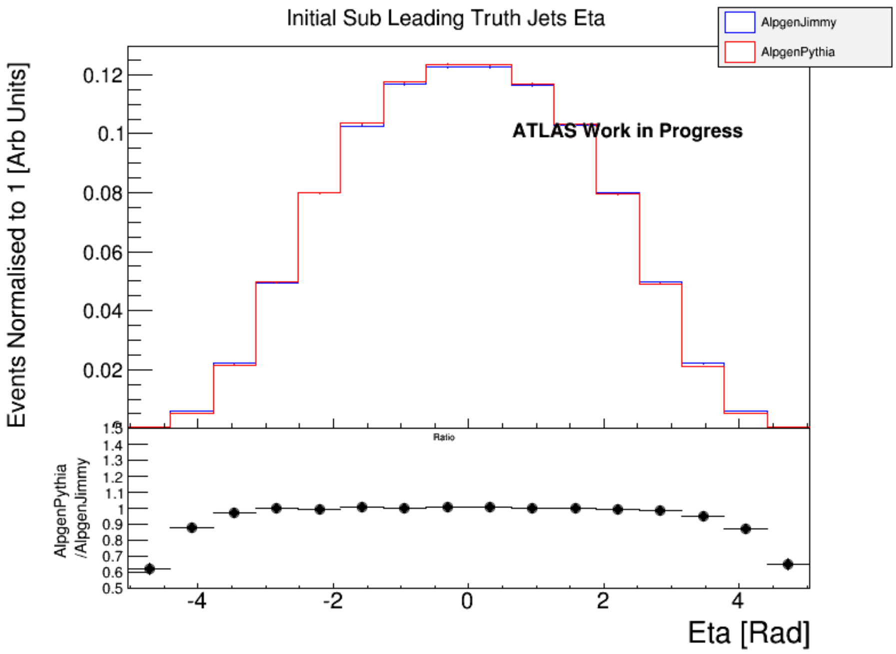


Figure 62: $W \to e\nu$ truth sub-leading jet η.

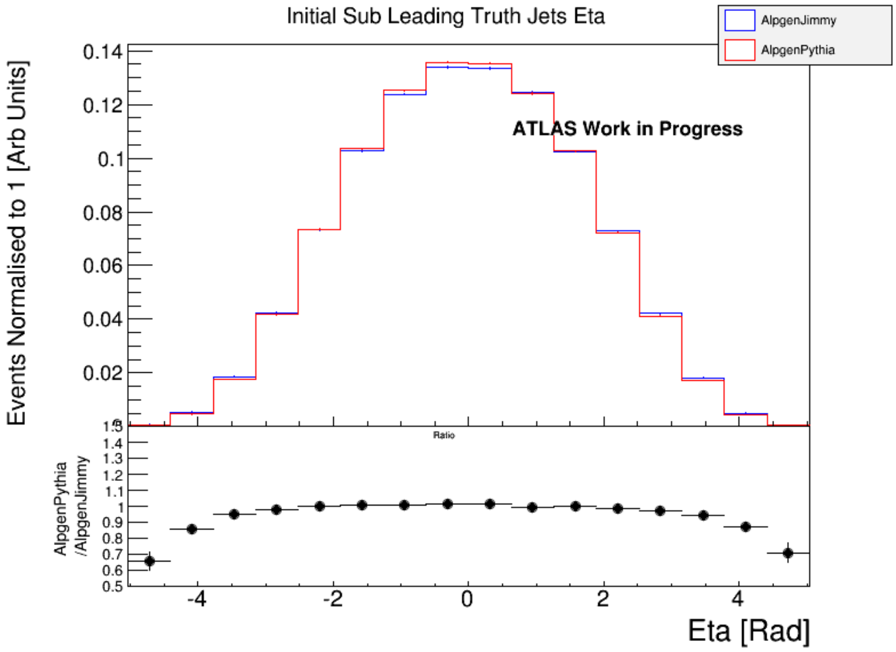


Figure 63: $W \to \mu\nu$ truth sub-leading jet η.

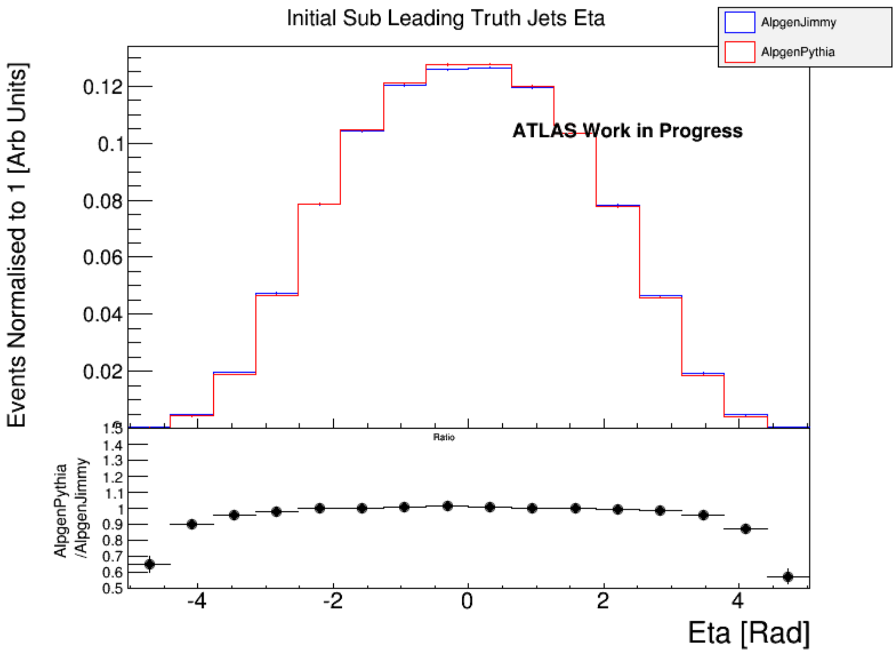


Figure 64: $W \to \tau\nu$ truth sub-leading jet η.

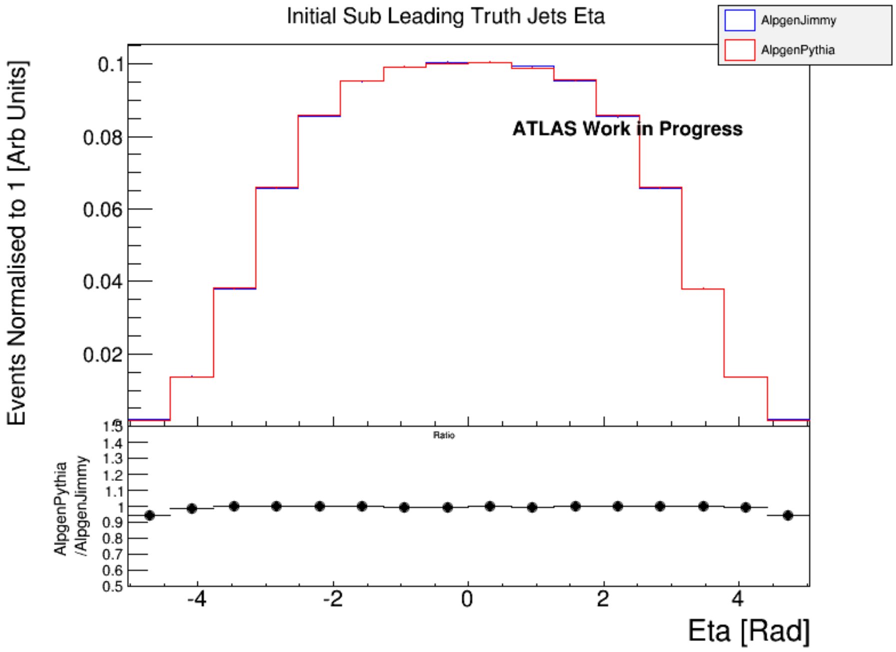


Figure 65: $Z \to ee$ truth sub-leading jet η.

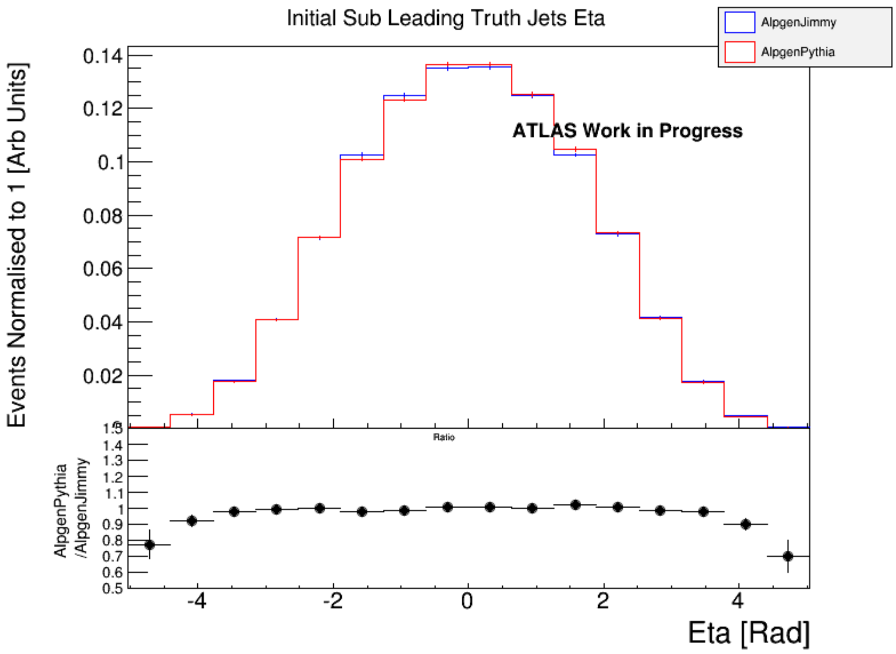


Figure 66: $Z \to \mu\mu$ truth sub-leading jet η.

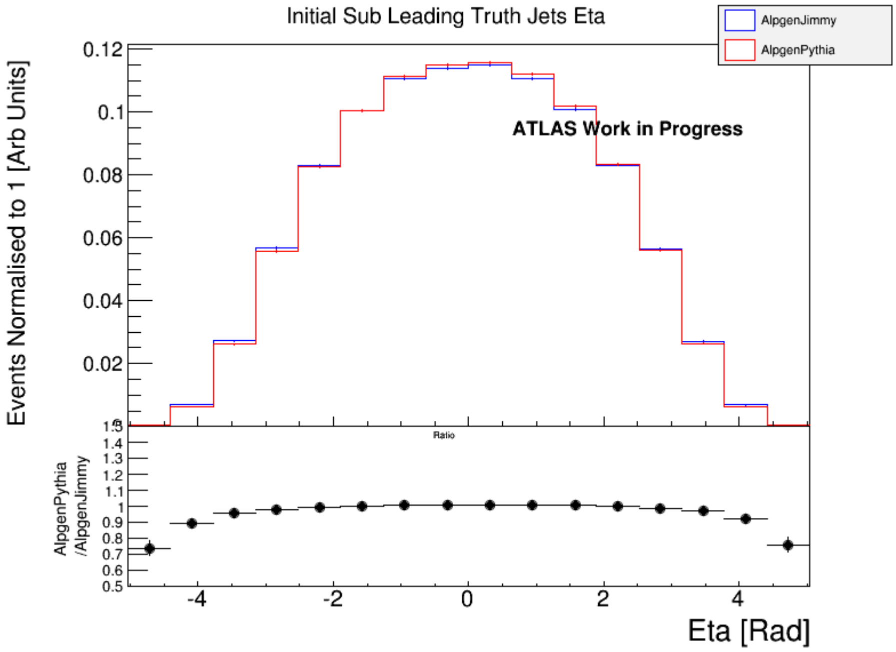


Figure 67: $Z \to \tau\tau$ truth sub-leading jet η.

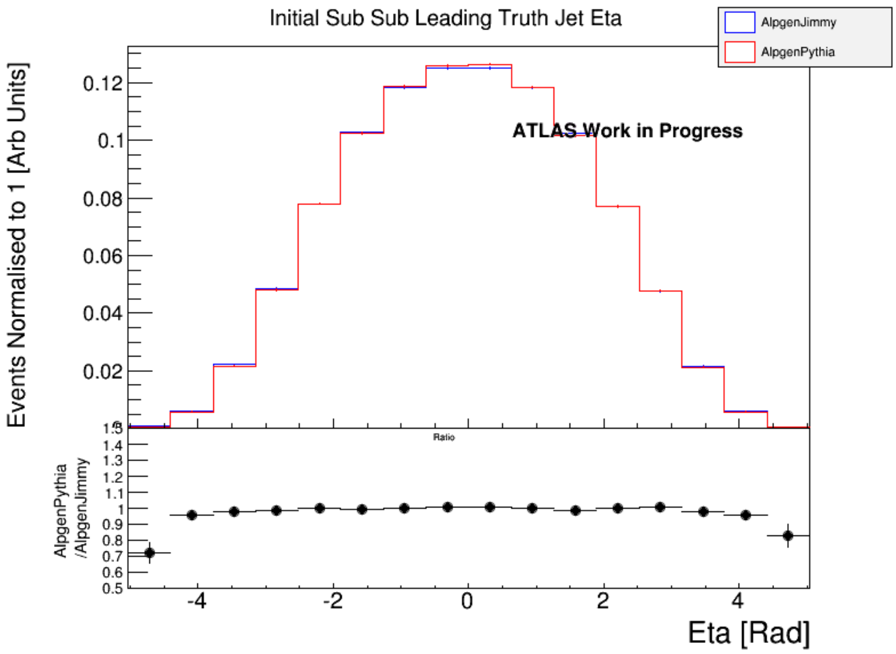


Figure 68: $W \to e\nu$ truth sub-sub-leading jet η.

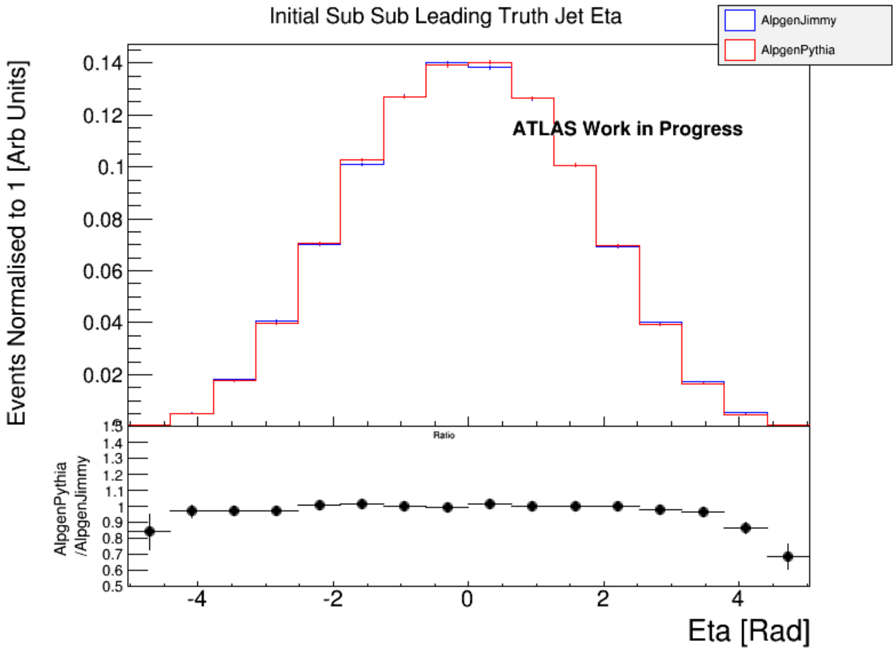


Figure 69: $W \to \mu\nu$ truth sub-sub-leading jet η.

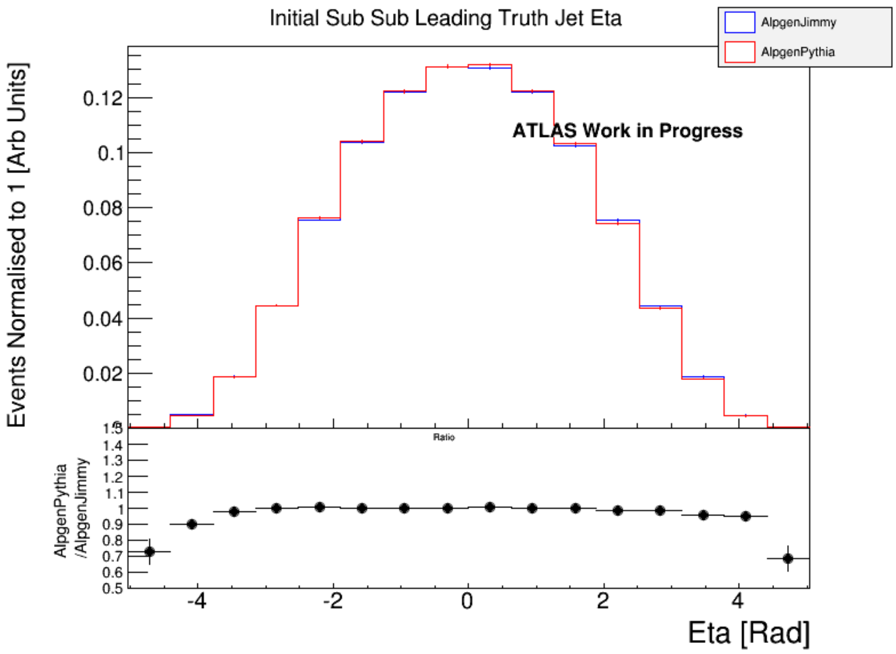


Figure 70: $W \to \tau\nu$ truth sub-sub-leading jet η.

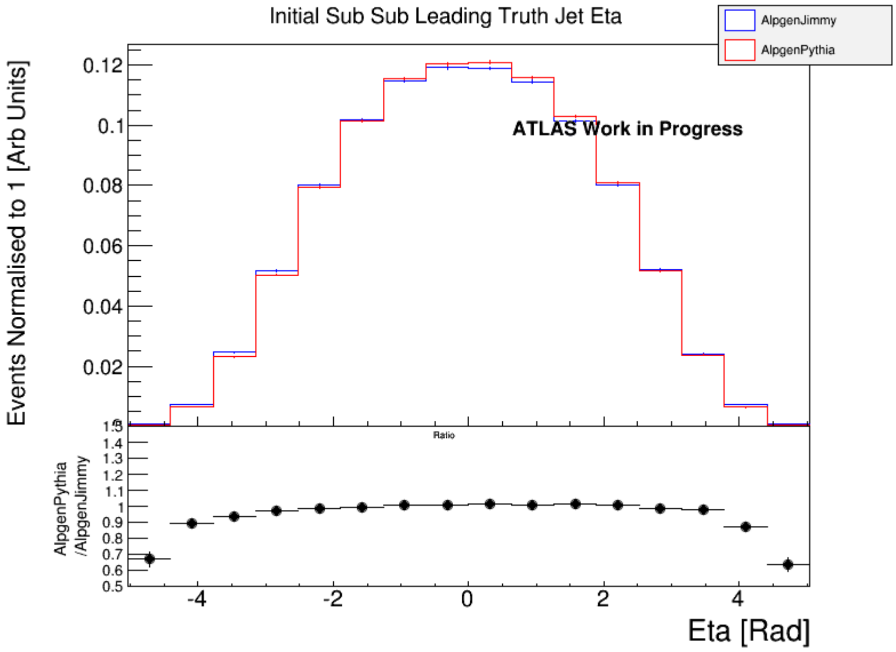


Figure 71: $Z \to ee$ truth sub-sub-leading jet η .

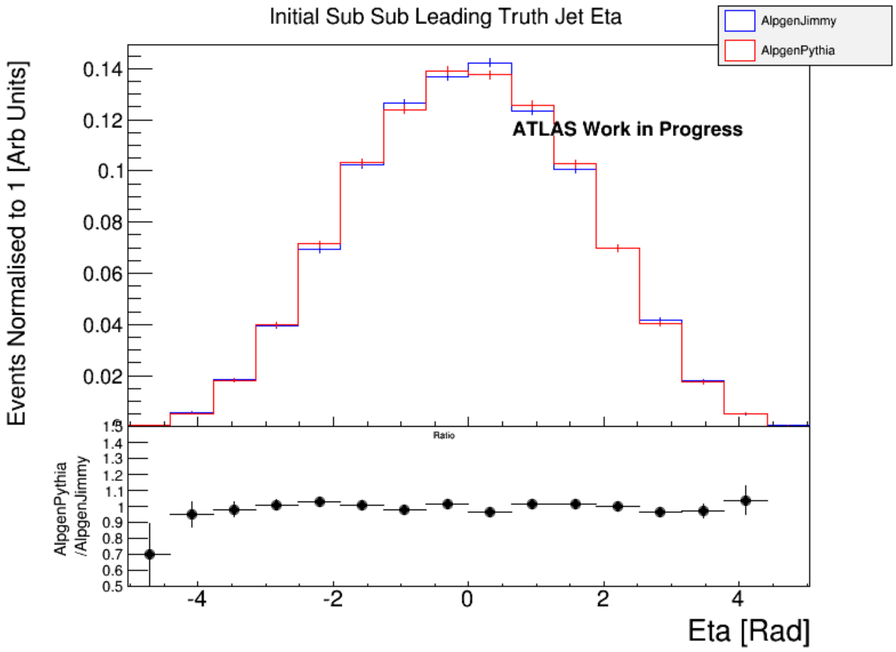


Figure 72: $Z \to \mu\mu$ truth sub-sub-leading jet η.

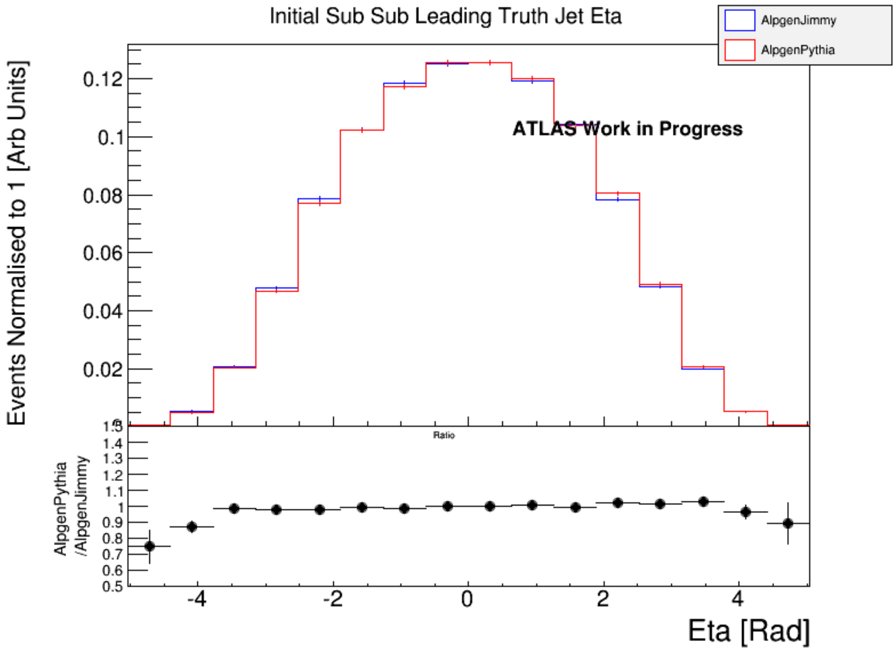


Figure 73: $Z \to \tau\tau$ truth sub-sub-leading jet η.

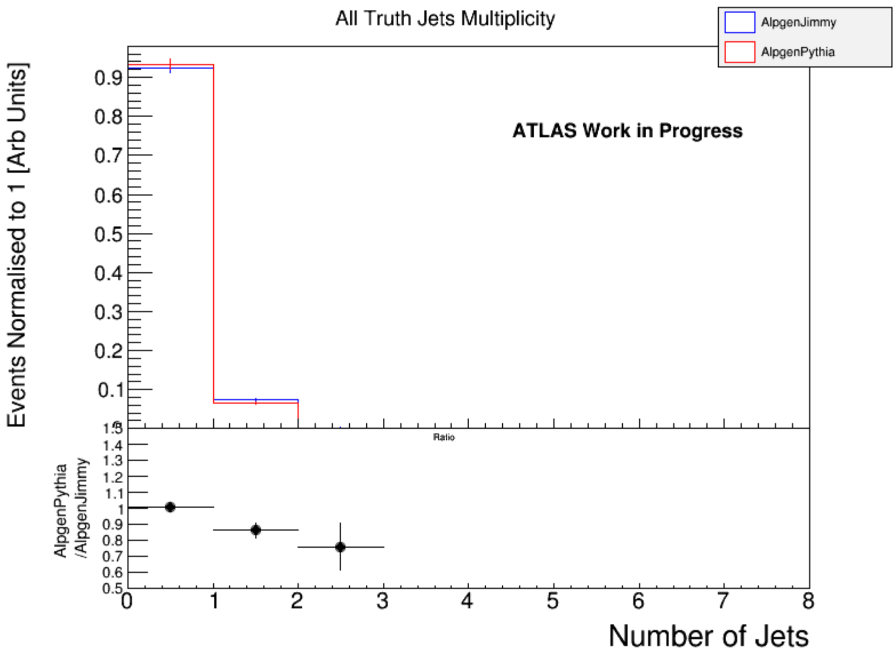


Figure 74: $W \to e\nu$ truth jet multiplicity.

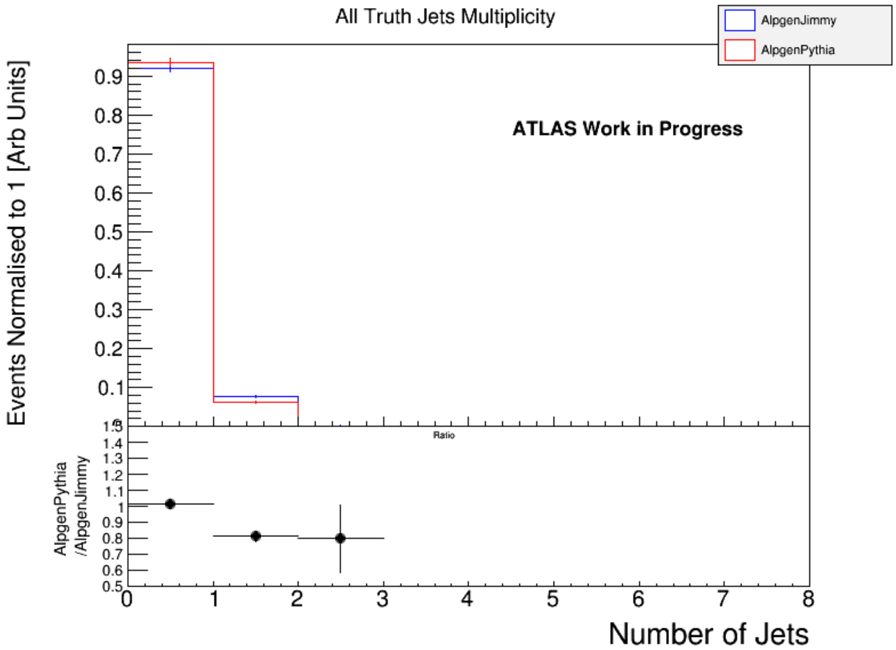


Figure 75: $W \to \mu\nu$ truth jet multiplicity.

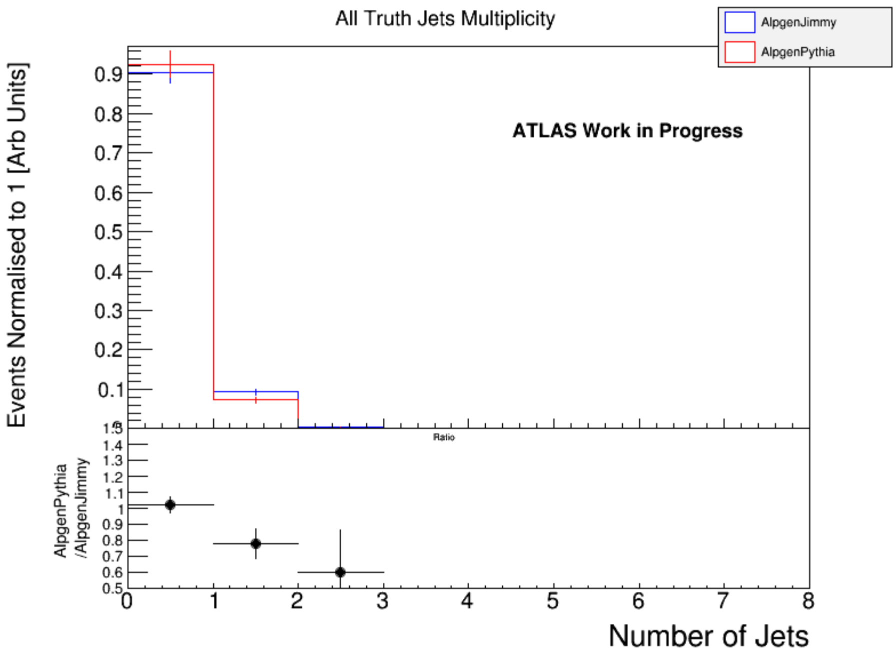


Figure 76: $W \to \tau\nu$ truth jet multiplicity.

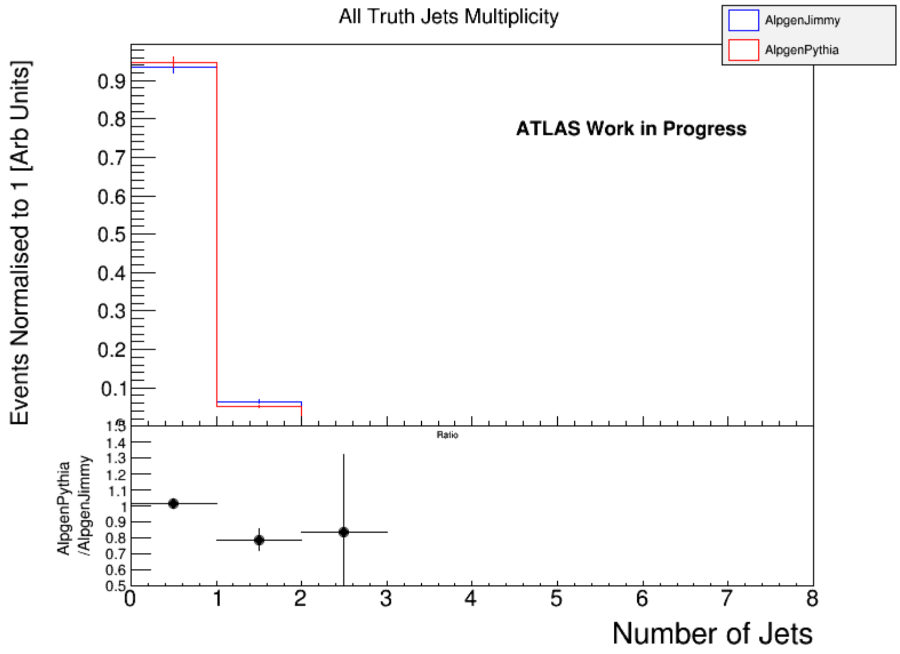


Figure 77: $Z \to ee$ truth jet multiplicity.

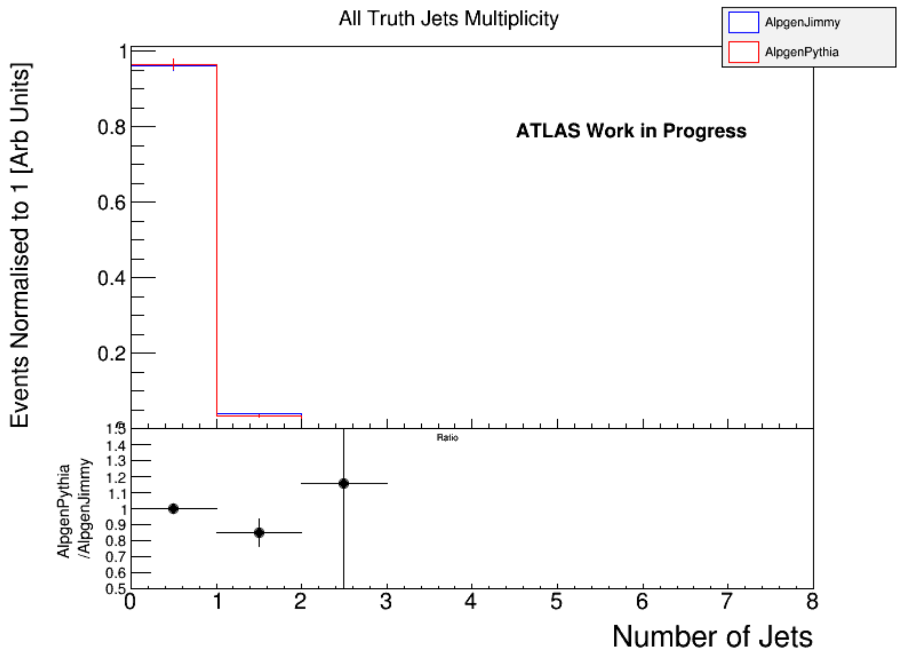


Figure 78: $Z \to \mu\mu$ truth jet multiplicity.

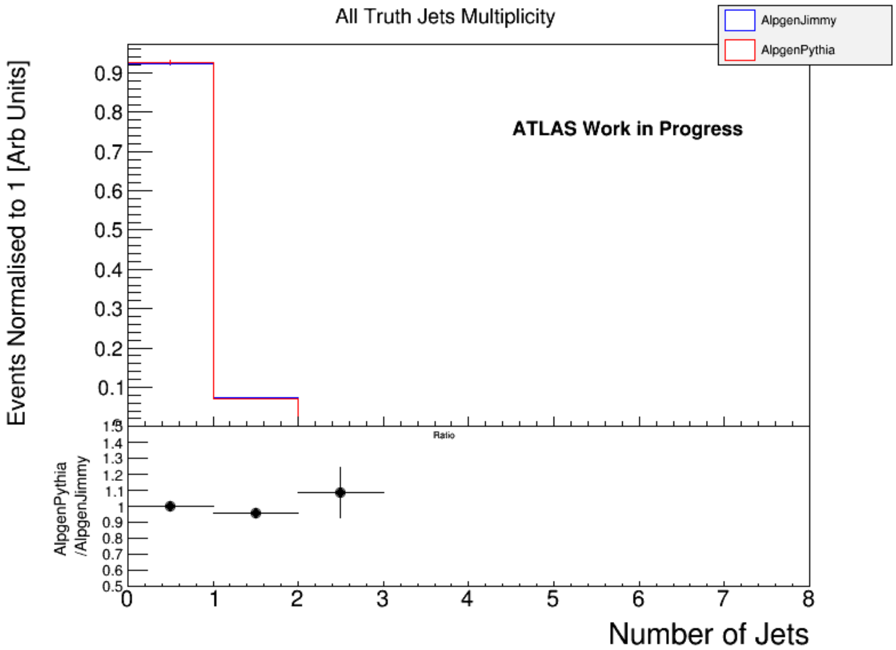


Figure 79: $Z \to \tau\tau$ truth jet multiplicity.

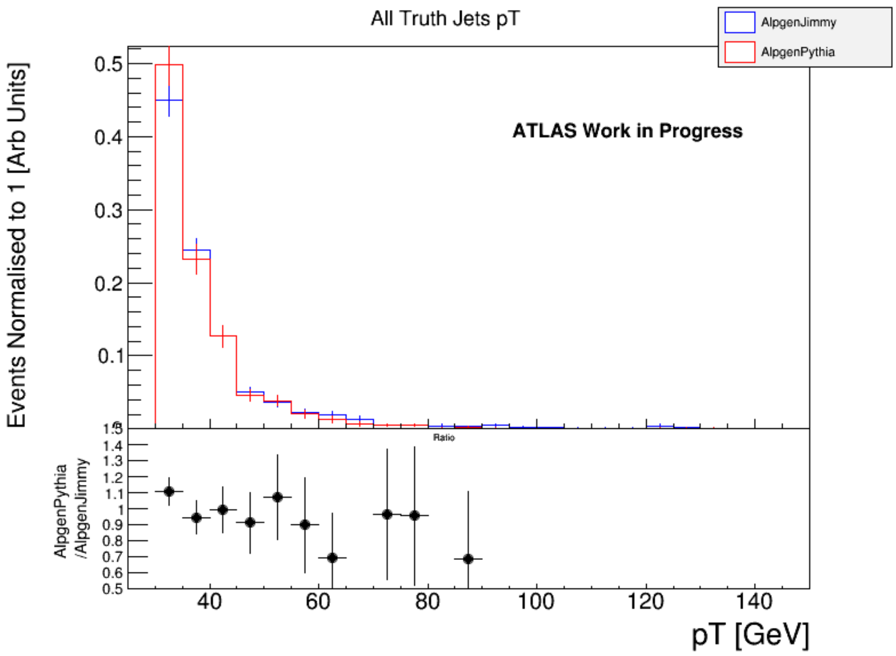


Figure 80: $W \to e\nu$ truth jet p_T.

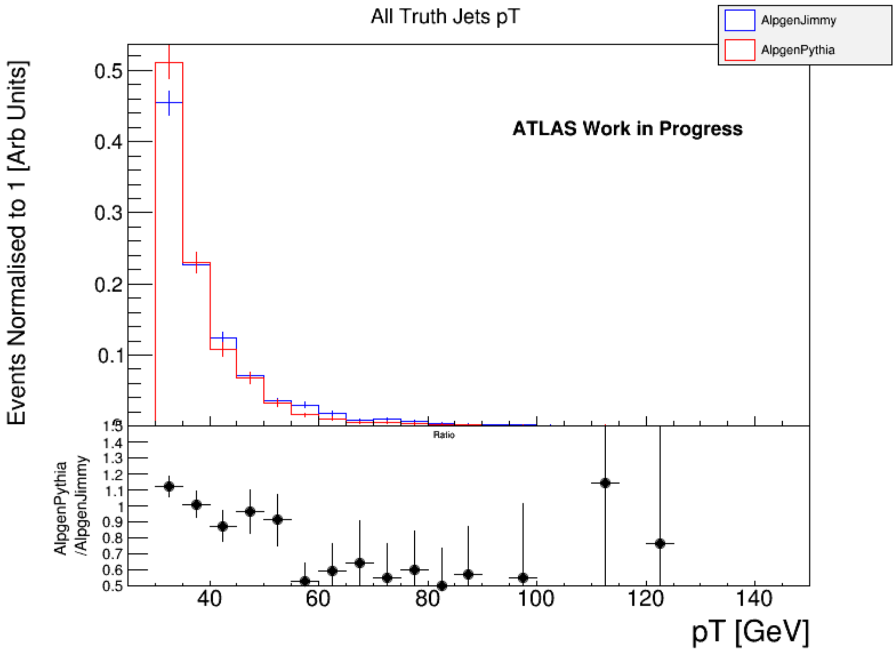


Figure 81: $W \to \mu\nu$ truth jet p_T.

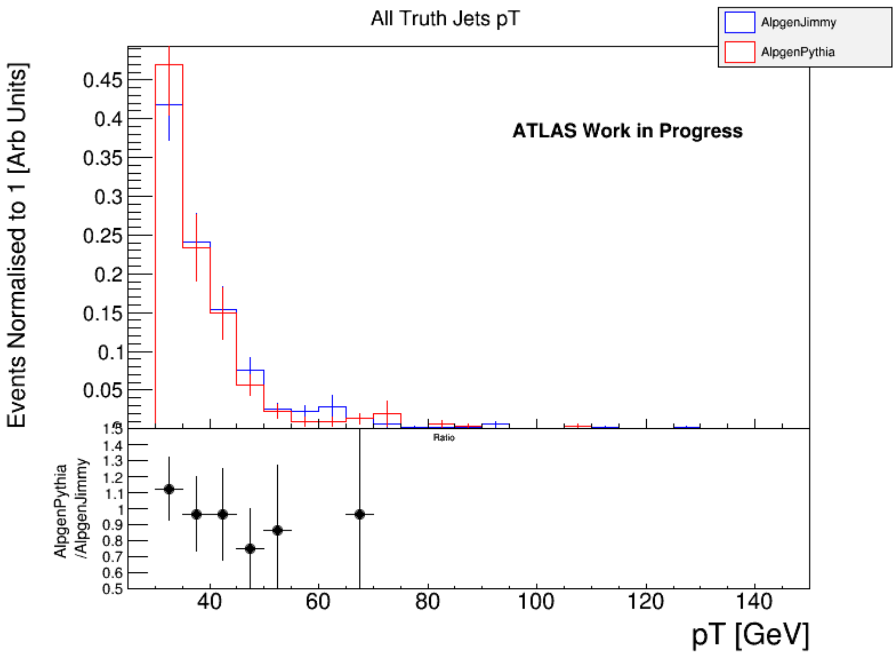


Figure 82: $W \to \tau\nu$ truth jet p_T.

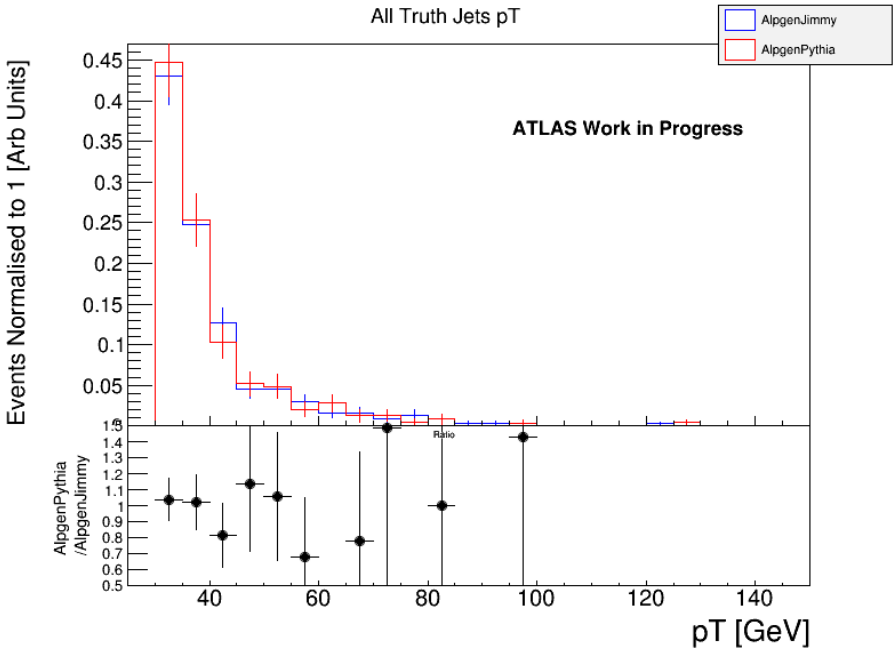


Figure 83: $Z \to ee$ truth jet p_T.

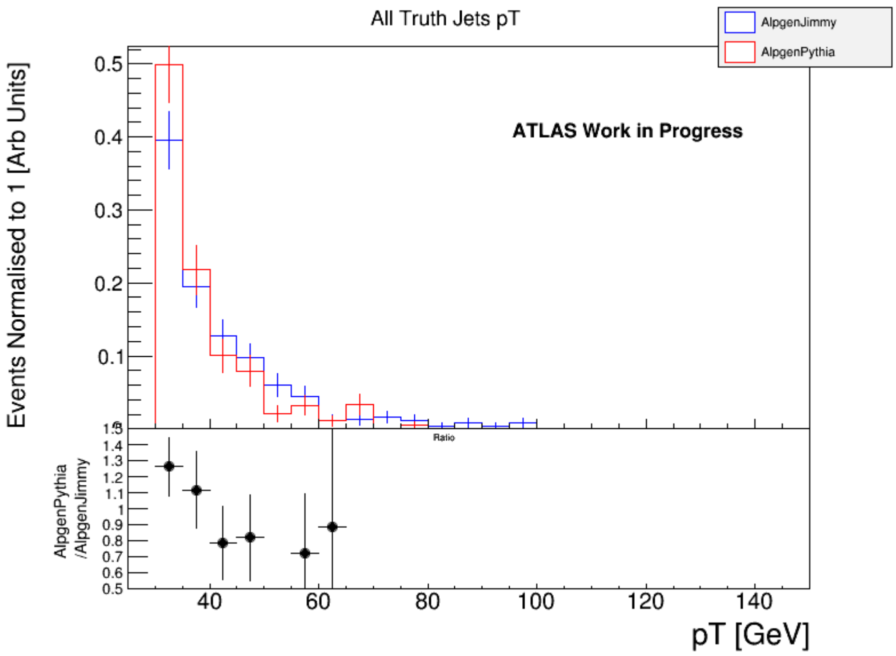


Figure 84: $Z \to \mu\mu$ truth jet p_T.

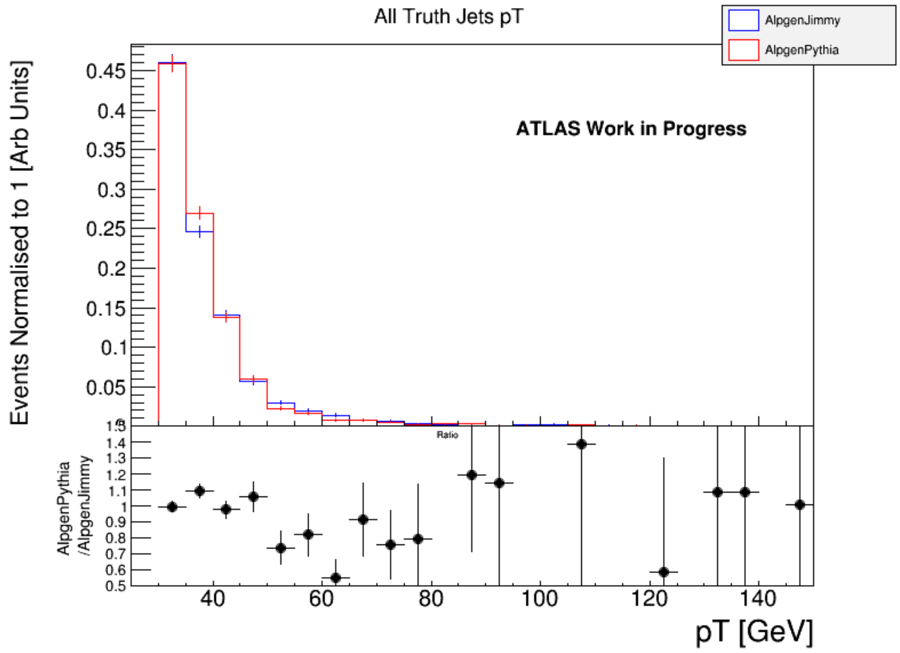


Figure 85: $Z \rightarrow \tau\tau$ truth jet p_T.

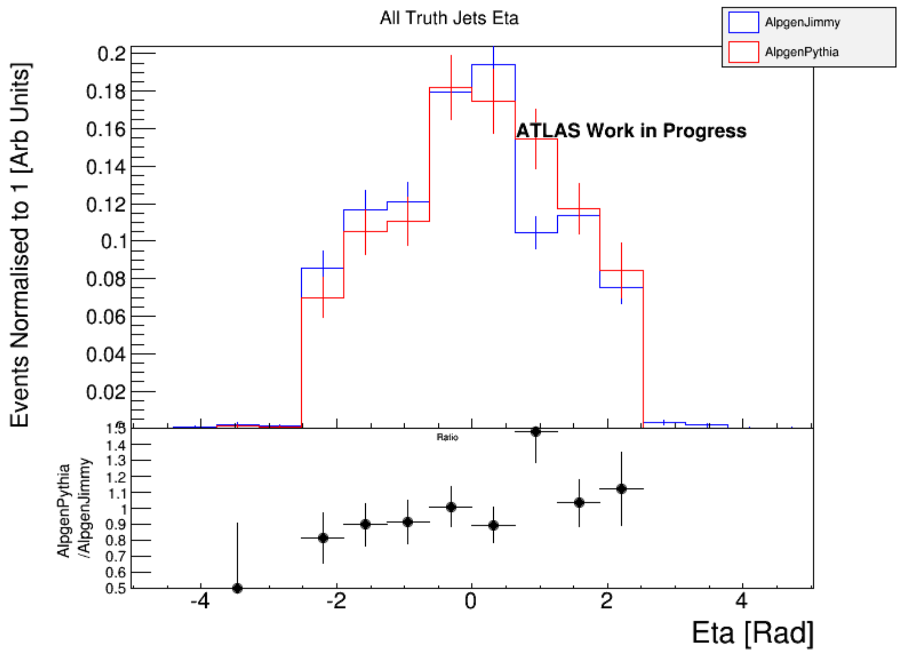


Figure 86: $W \rightarrow e\nu$ truth jet η.

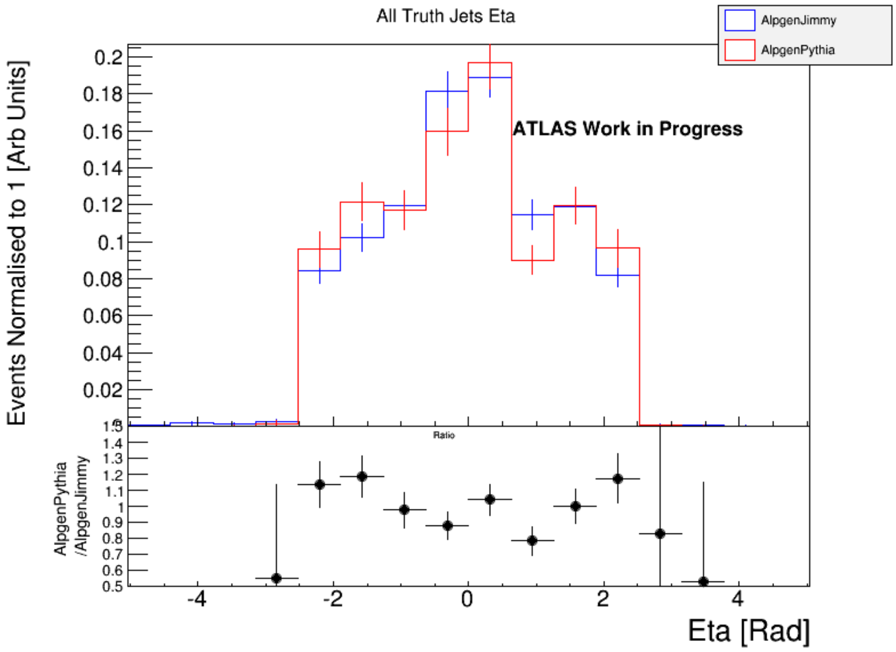


Figure 87: $W \rightarrow \mu\nu$ truth jet η.

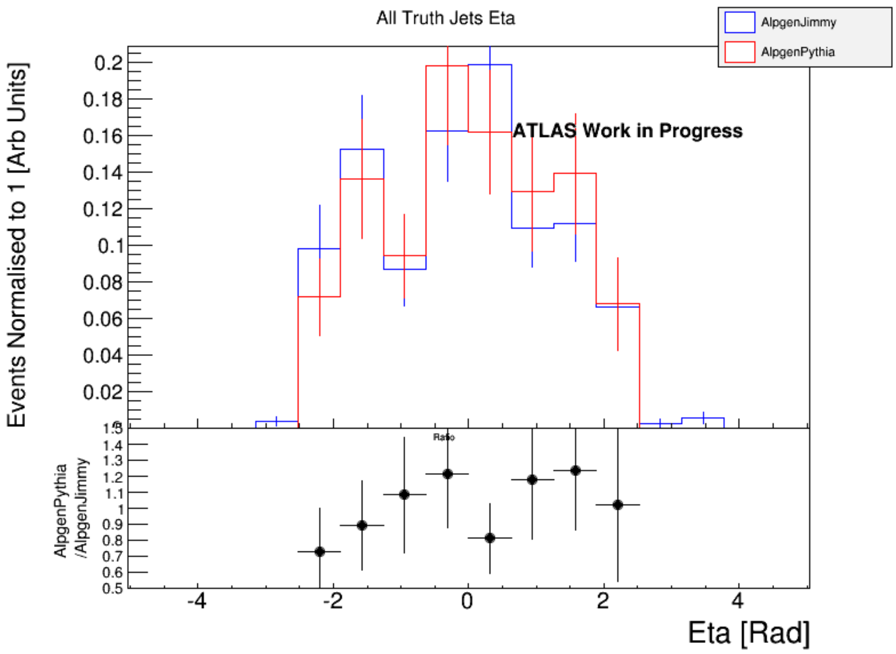


Figure 88: $W \rightarrow \tau\nu$ truth jet η.

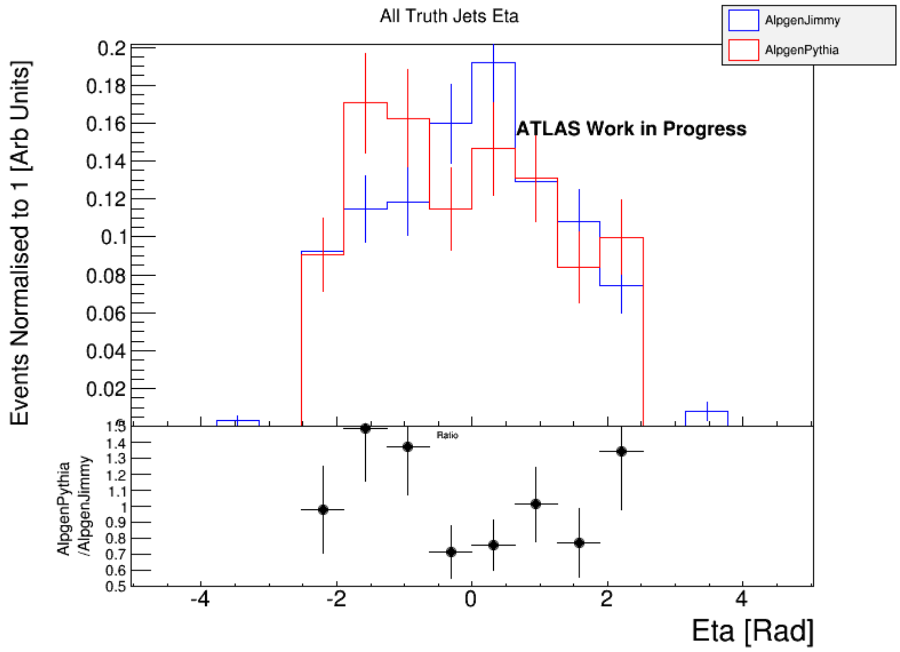


Figure 89: $Z \to ee$ truth jet η.

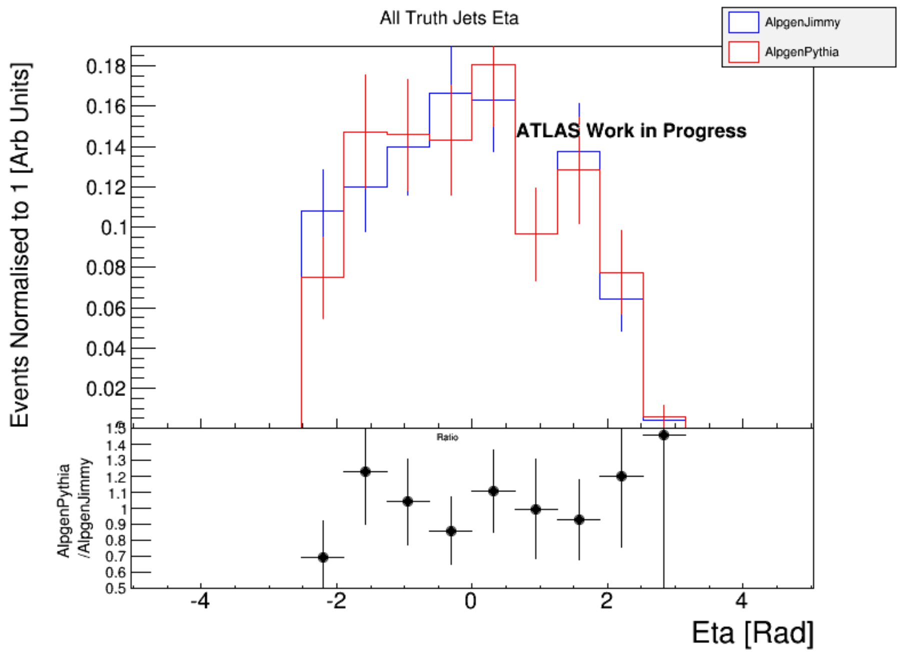


Figure 90: $Z \to \mu\mu$ truth jet η.

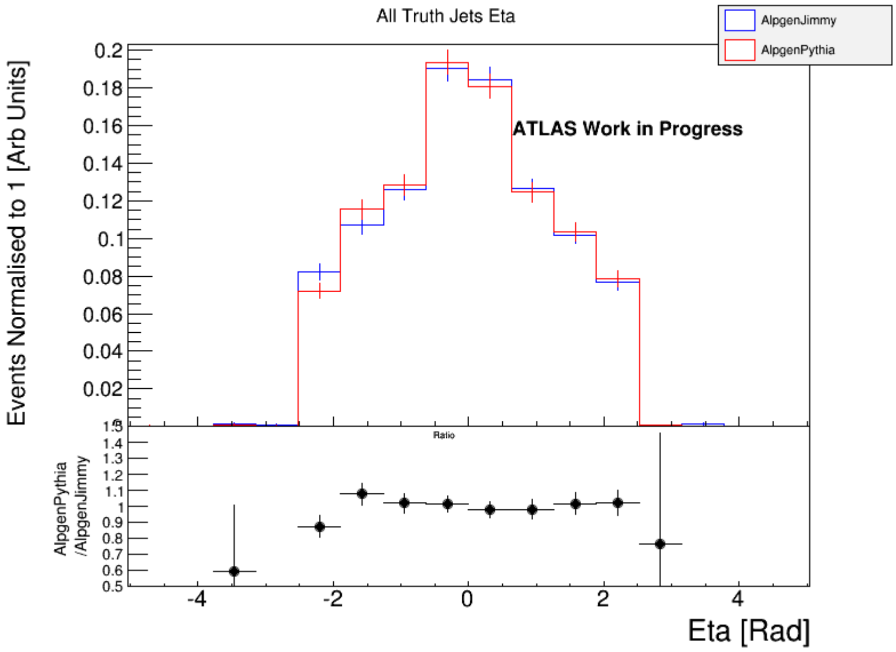


Figure 91: $Z \to \tau\tau$ truth jet η.

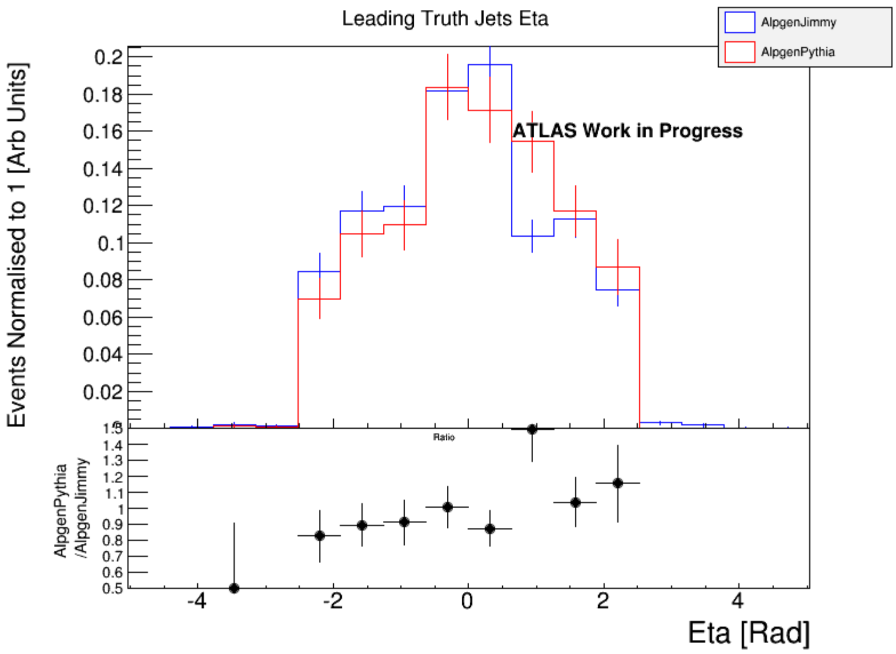


Figure 92: $W \to e\nu$ truth leading jet η.

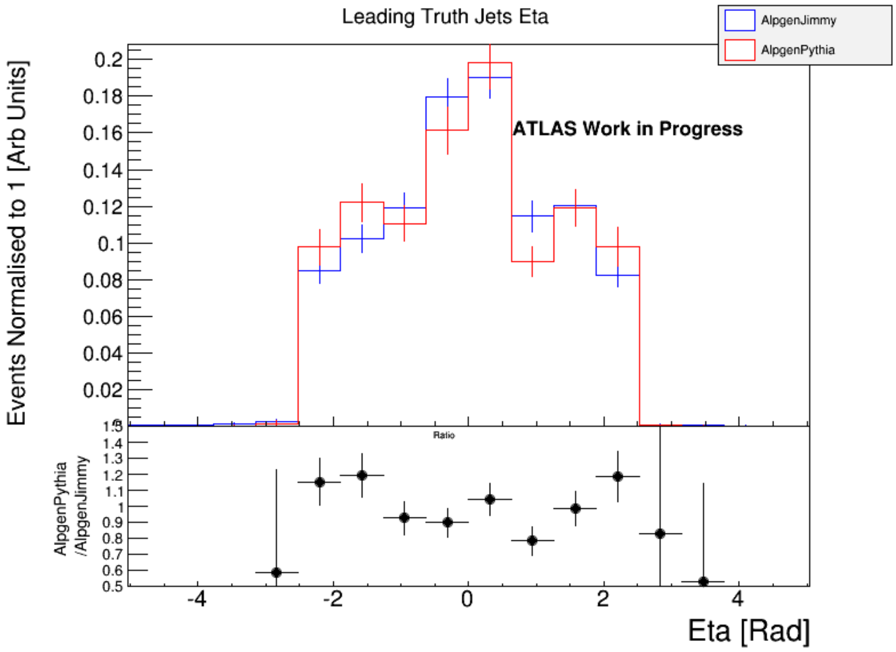


Figure 93: $W \to \mu\nu$ truth leading jet η.

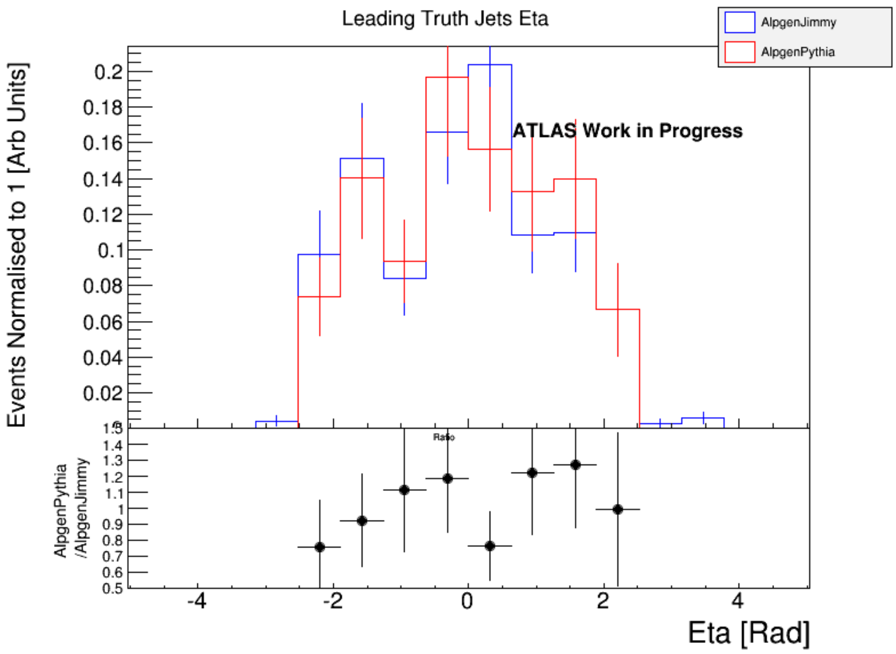


Figure 94: $W \to \tau\nu$ truth leading jet η.

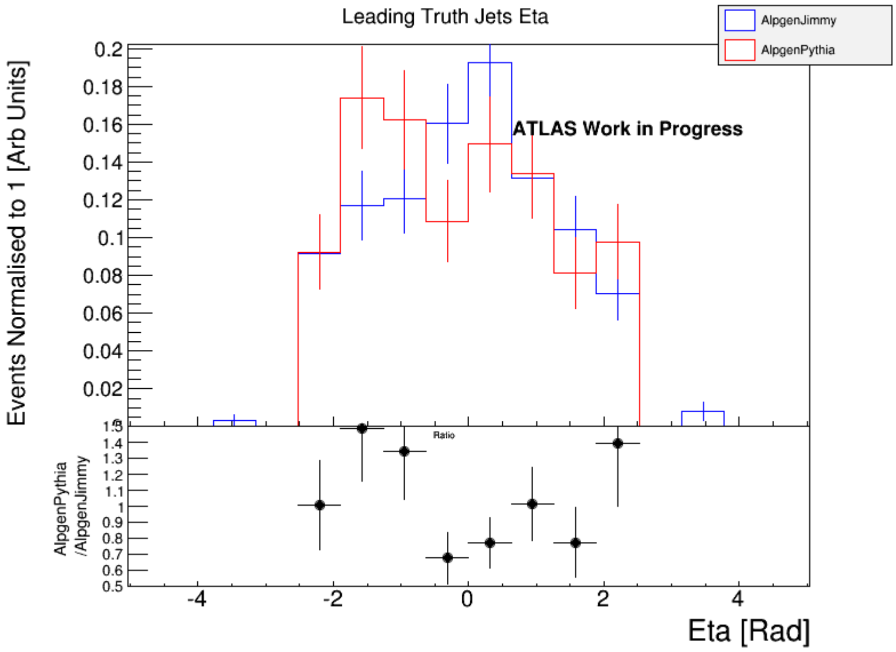


Figure 95: $Z \to ee$ truth leading jet η.

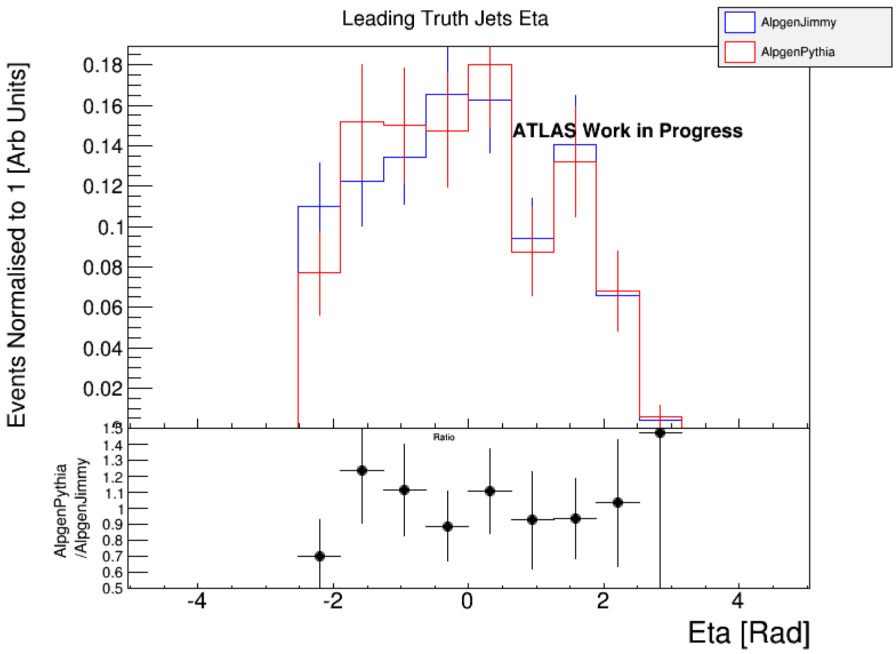


Figure 96: $Z \to \mu\mu$ truth leading jet η.

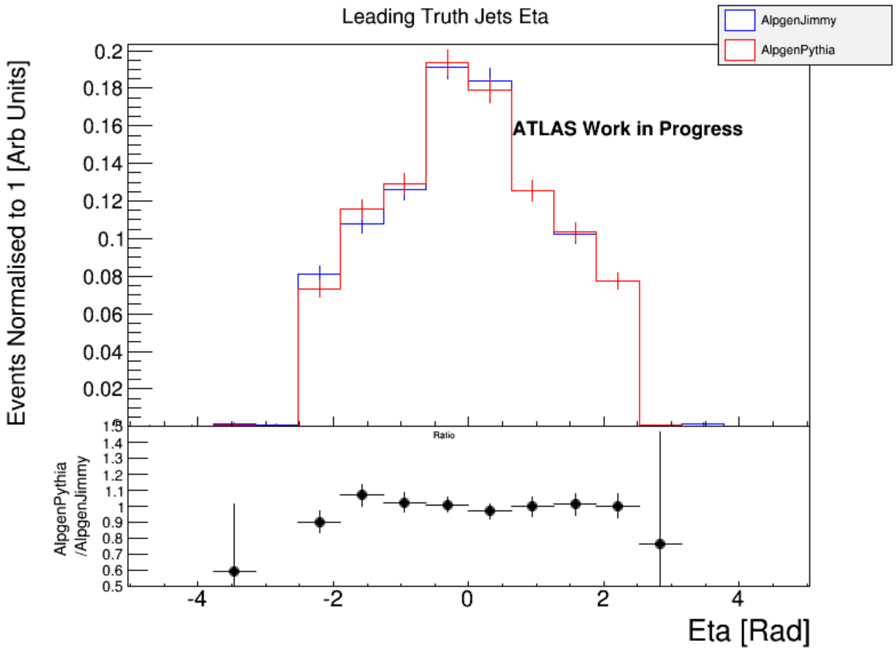


Figure 97: $Z \to \tau\tau$ truth leading jet η.

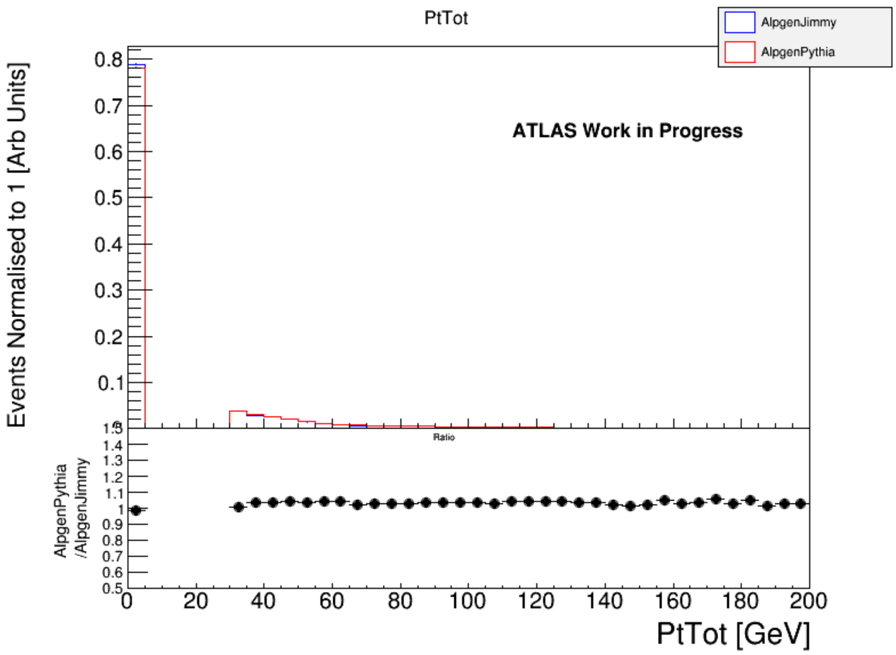


Figure 98: $W \to e\nu$ p_T^{Tot} distribution.

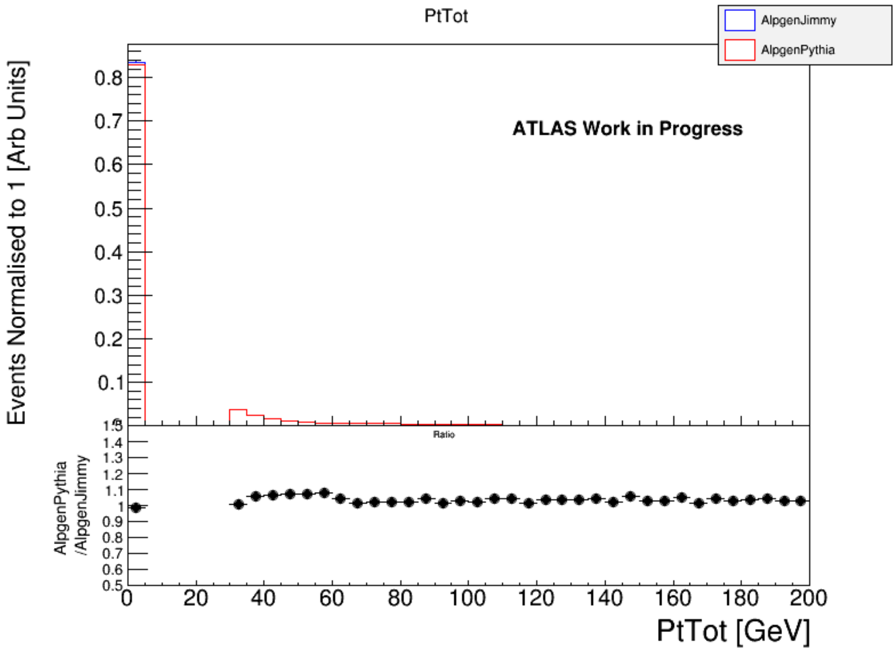


Figure 99: $W \to \mu\nu$ p_T^{Tot} distribution.

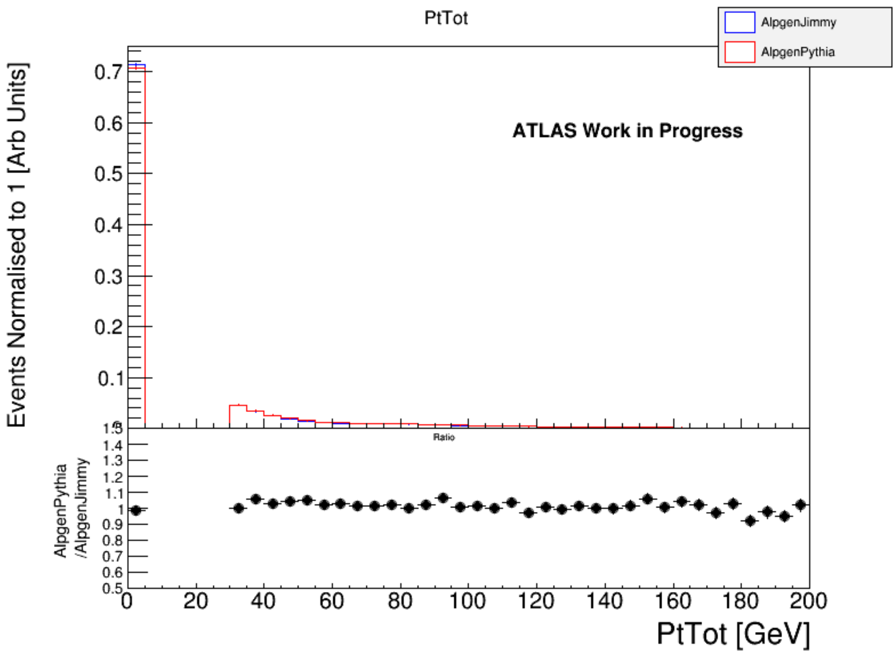


Figure 100: $W \to \tau\nu$ p_T^{Tot} distribution.

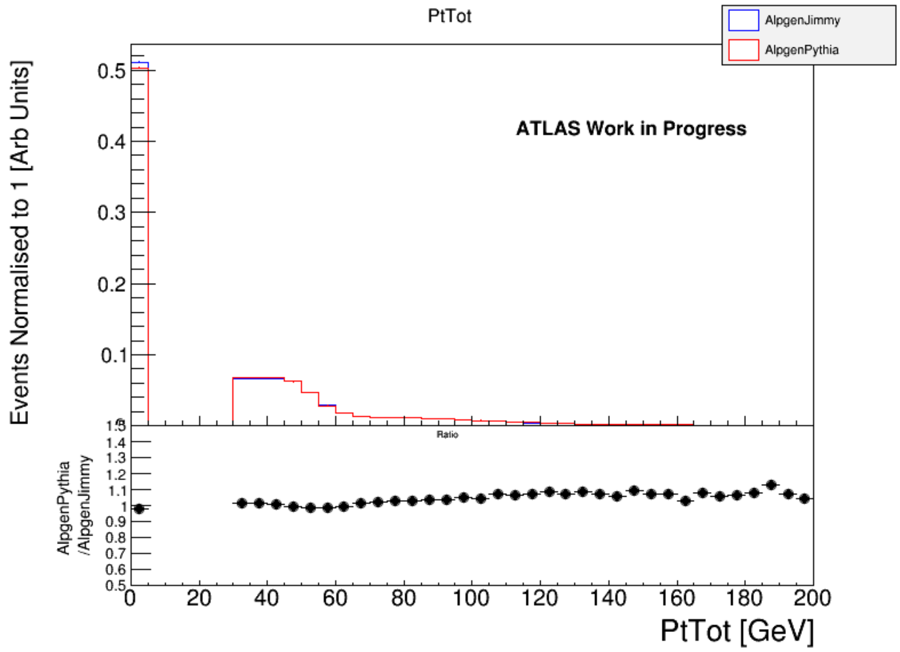


Figure 101: $Z \to ee\ p_T^{Tot}$ distribution.

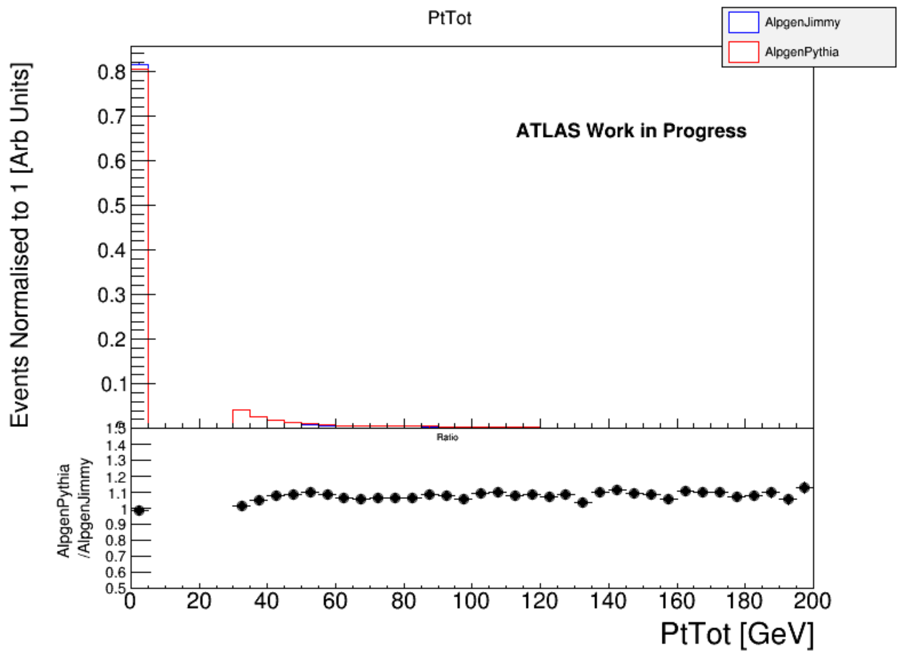


Figure 102: $Z \to \mu\mu\ p_T^{Tot}$ distribution.

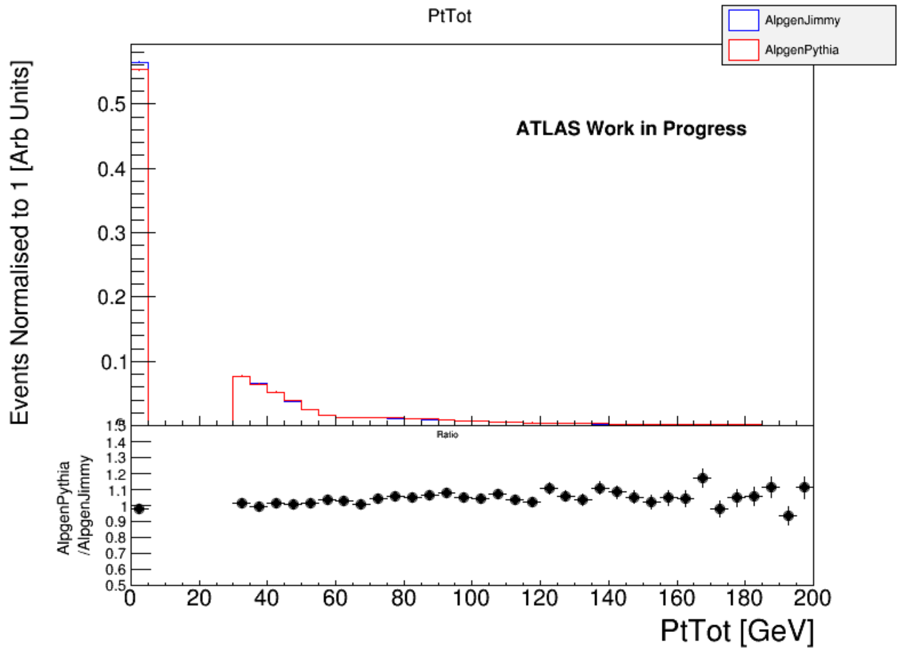


Figure 103: $Z \to \tau\tau$ p_T^{Tot} distribution.

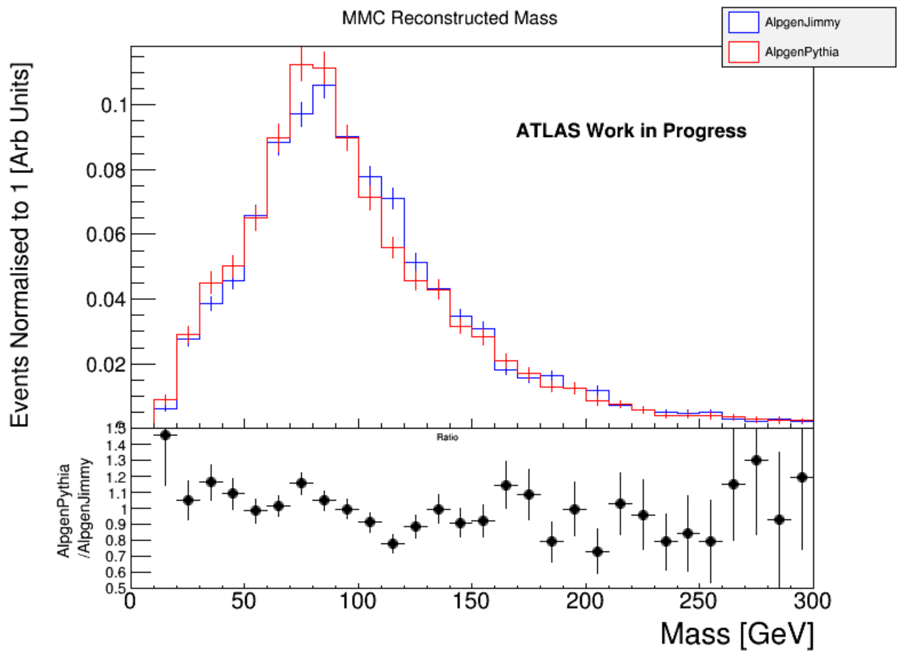


Figure 104: $W \to e\nu$ m_{MMC} distribution.

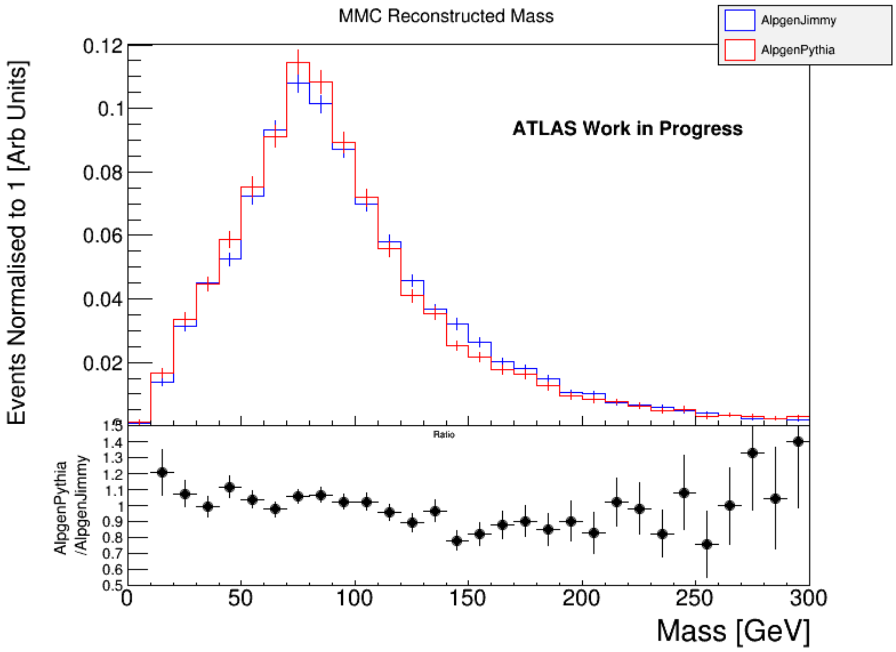


Figure 105: $W \to \mu\nu$ m_{MMC} distribution.

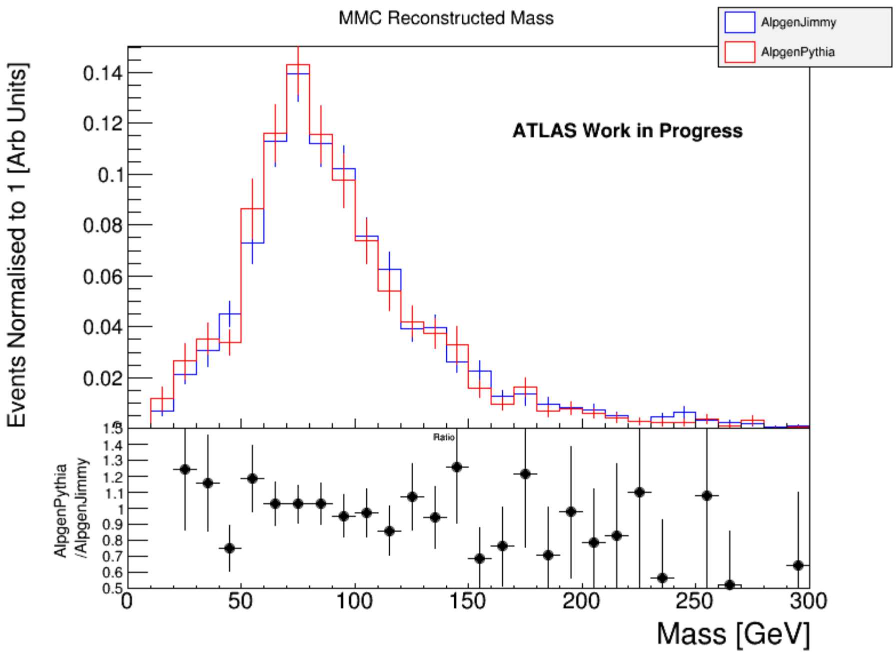


Figure 106: $W \to \tau\nu$ m_{MMC} distribution.

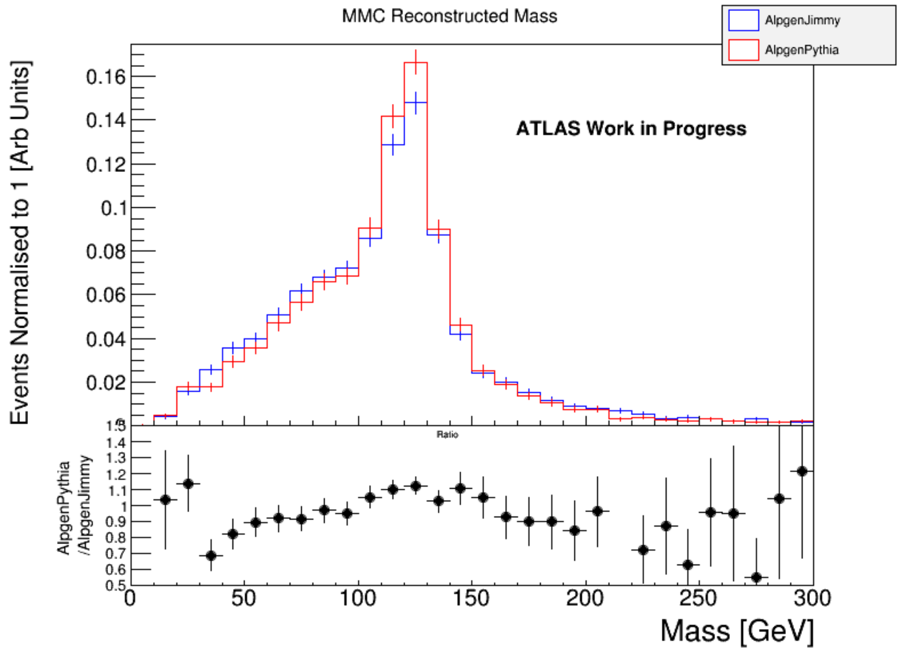


Figure 107: $Z \to ee\ m_{MMC}$ distribution.

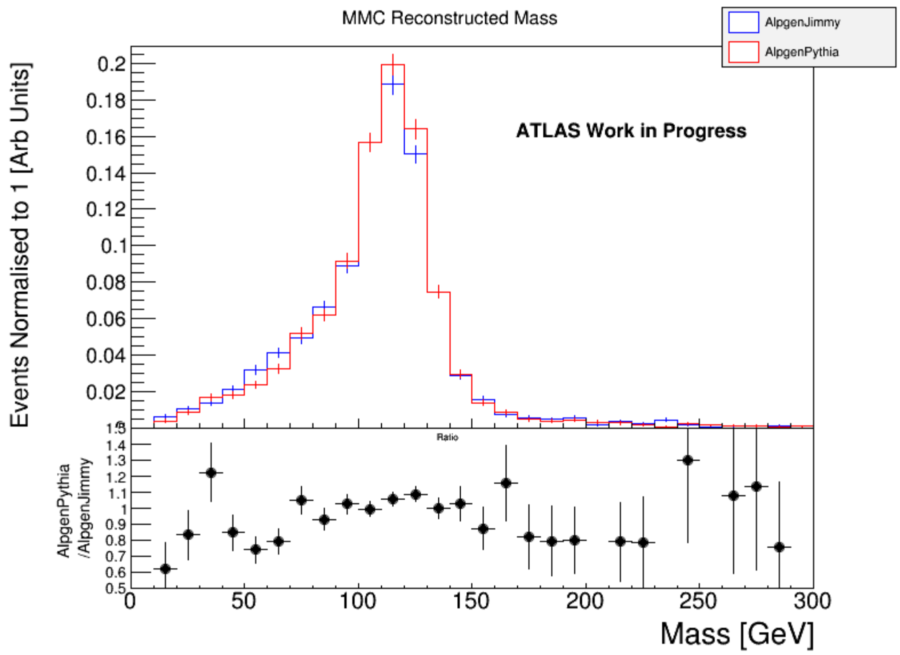


Figure 108: $Z \to \mu\mu\ m_{MMC}$ distribution.

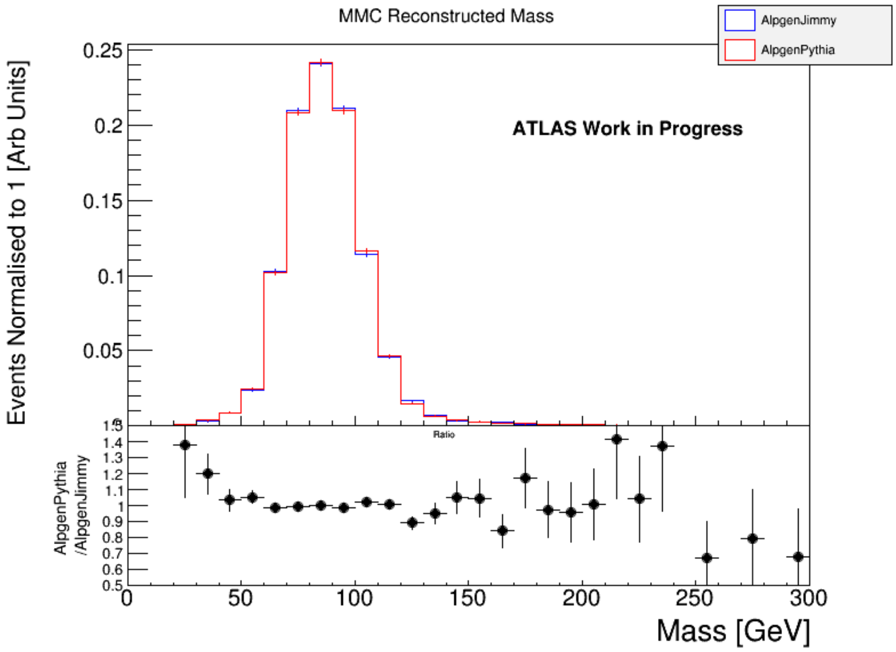


Figure 109: $Z \to \tau\tau$ m_{MMC} distribution.

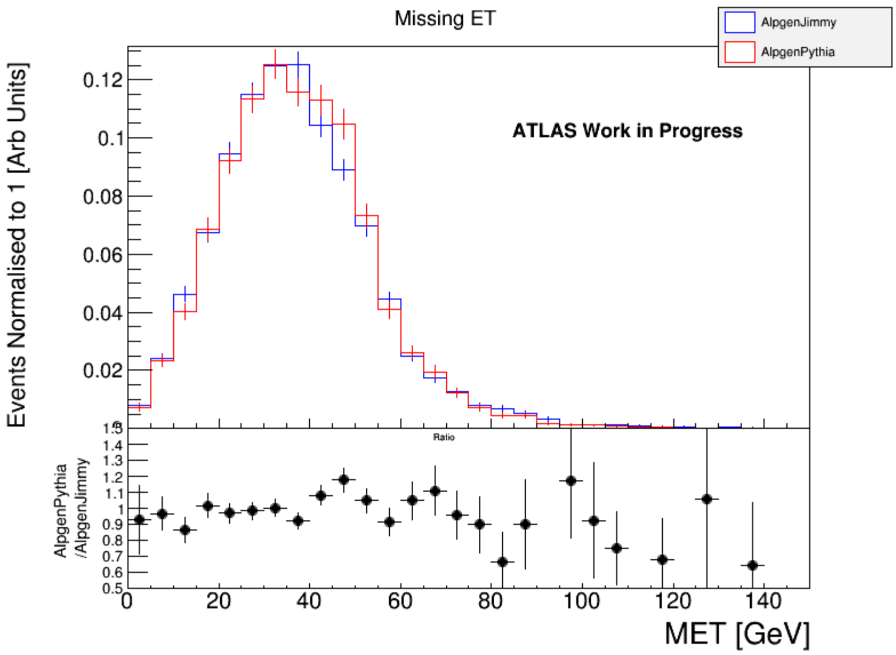


Figure 110: $W \to e\nu$ E_T^{miss} distribution.

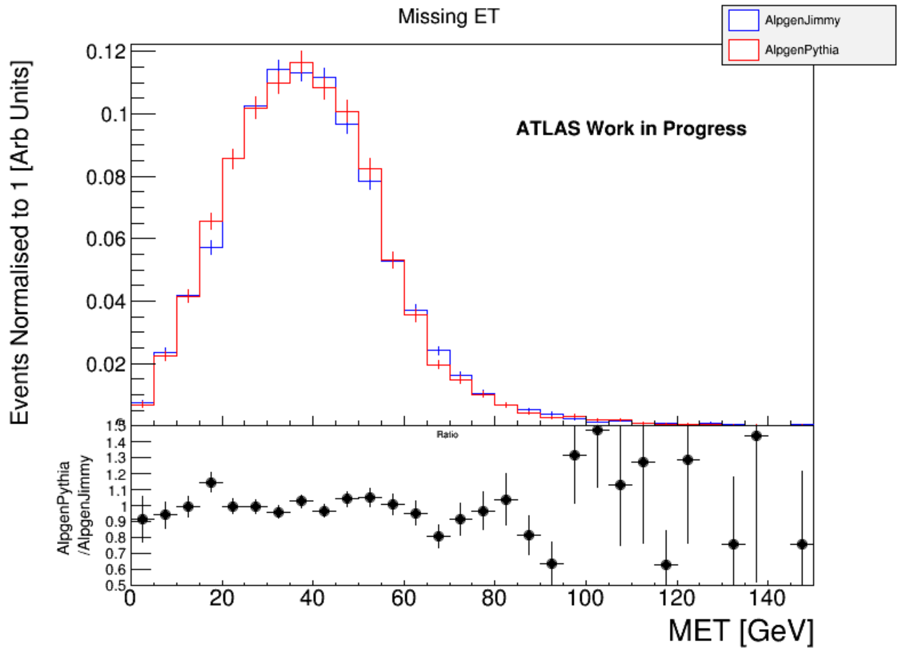


Figure 111: $W \to \mu\nu$ E_T^{miss} distribution.

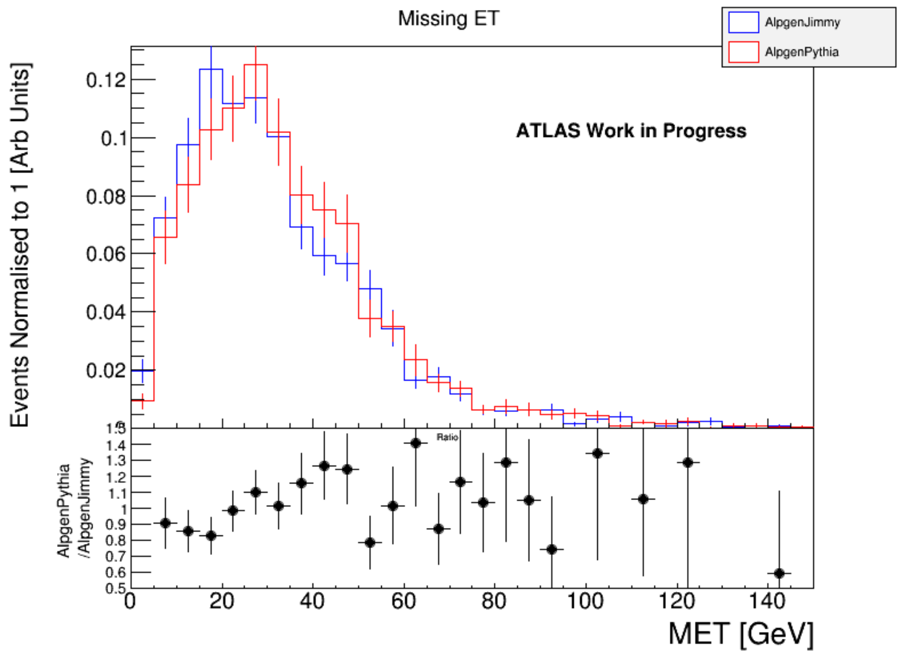


Figure 112: $W \to \tau\nu$ E_T^{miss} distribution.

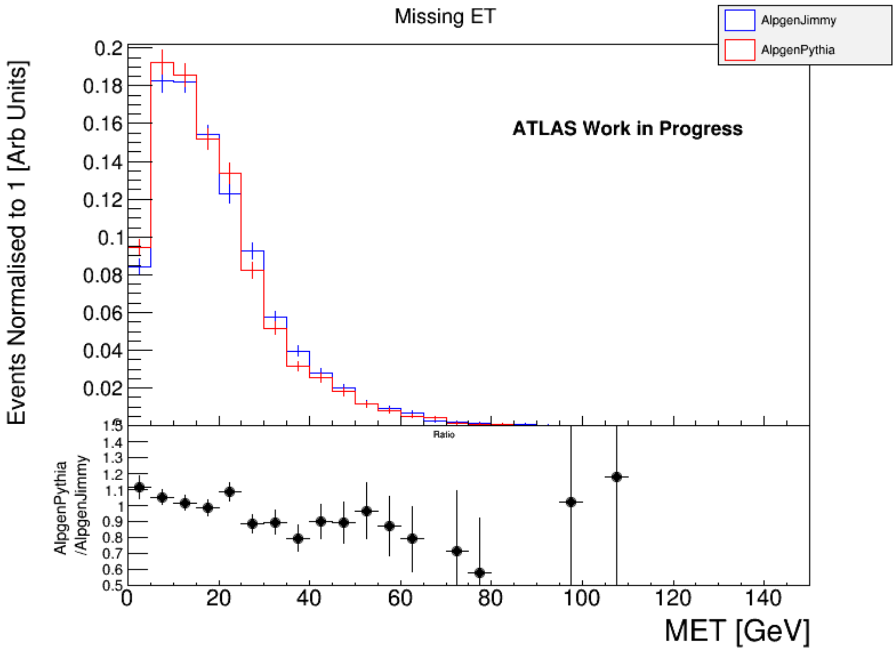


Figure 113: $Z \to ee$ E_T^{miss} distribution.

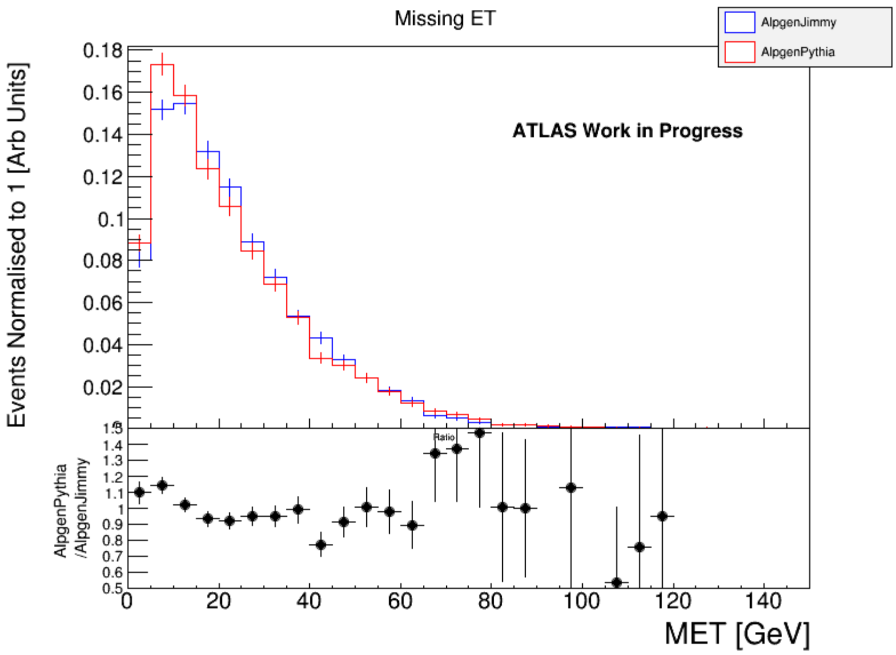


Figure 114: $Z \to \mu\mu$ E_T^{miss} distribution.

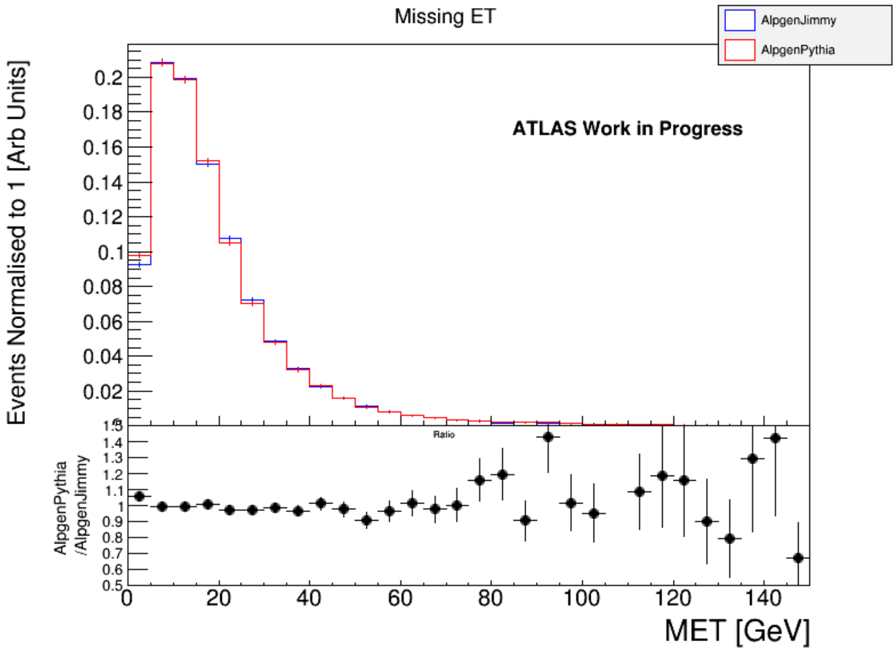


Figure 115: $Z \to \tau\tau$ E_T^{miss} distribution.

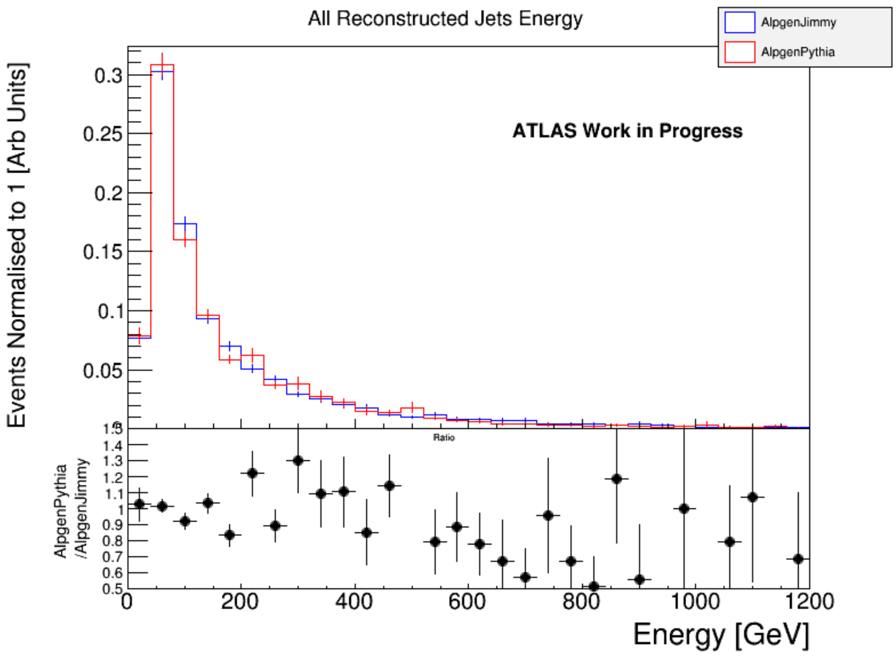


Figure 116: $W \to e\nu$ reconstructed jet energies.

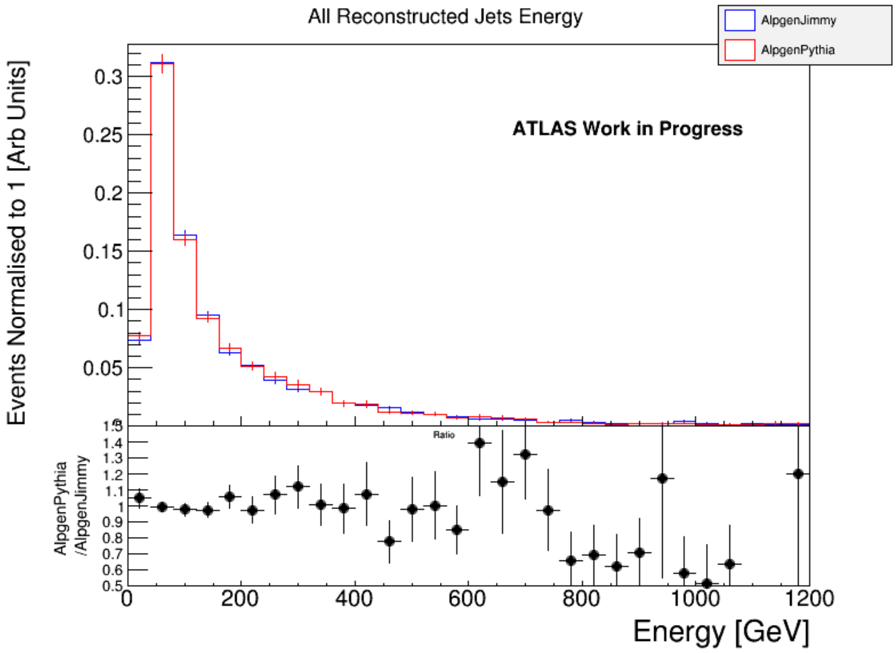


Figure 117: $W \to \mu\nu$ reconstructed jet energies.

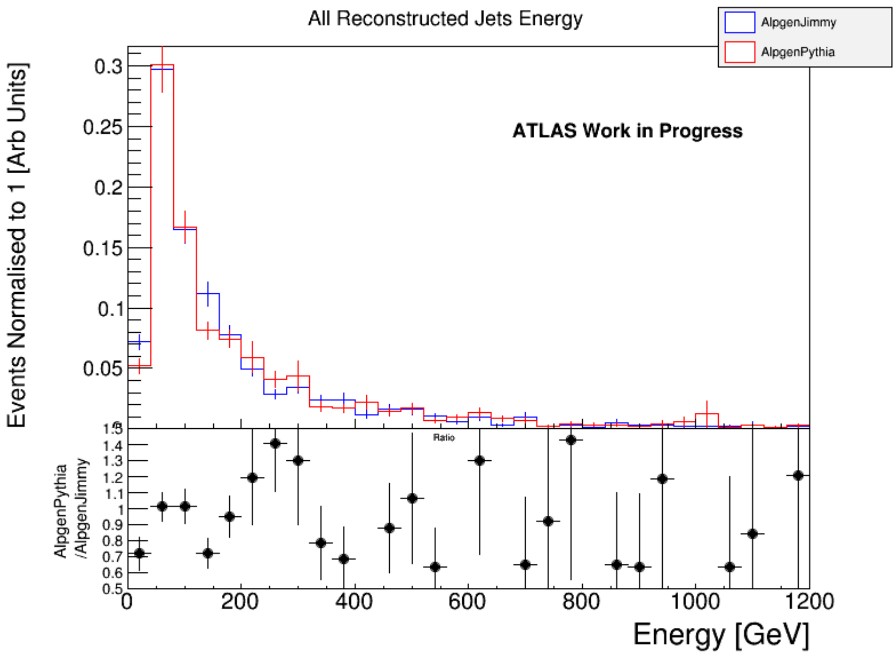


Figure 118: $W \to \tau\nu$ reconstructed jet energies.

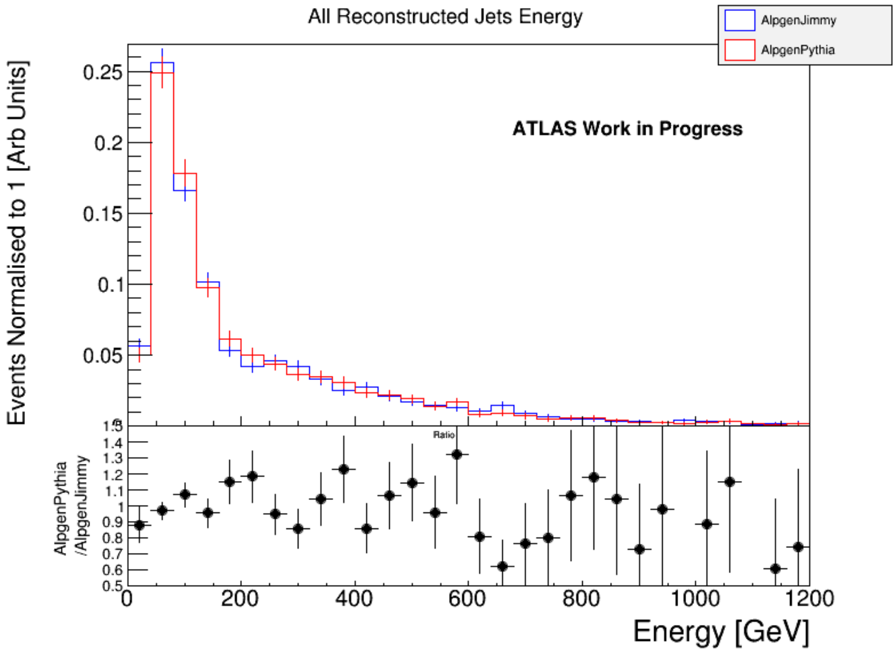


Figure 119: $Z \rightarrow ee$ reconstructed jet energies.

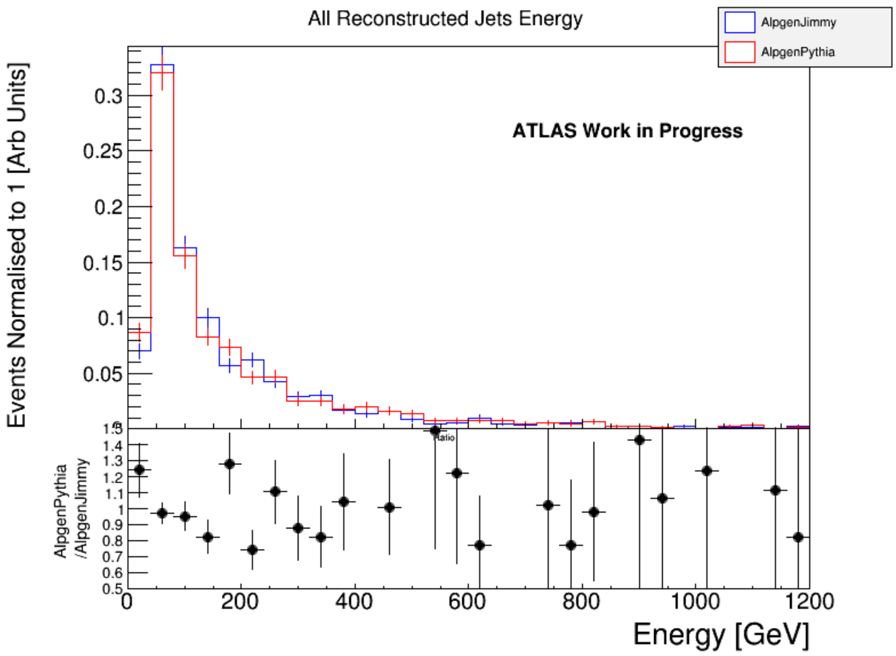


Figure 120: $Z \rightarrow \mu\mu$ reconstructed jet energies.

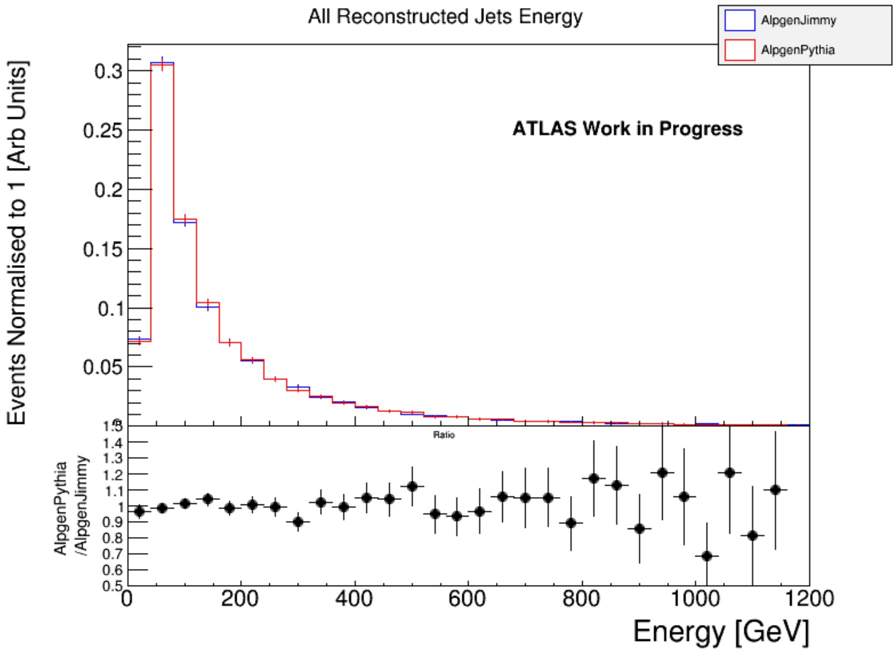


Figure 121: $Z \to \tau\tau$ reconstructed jet energies.

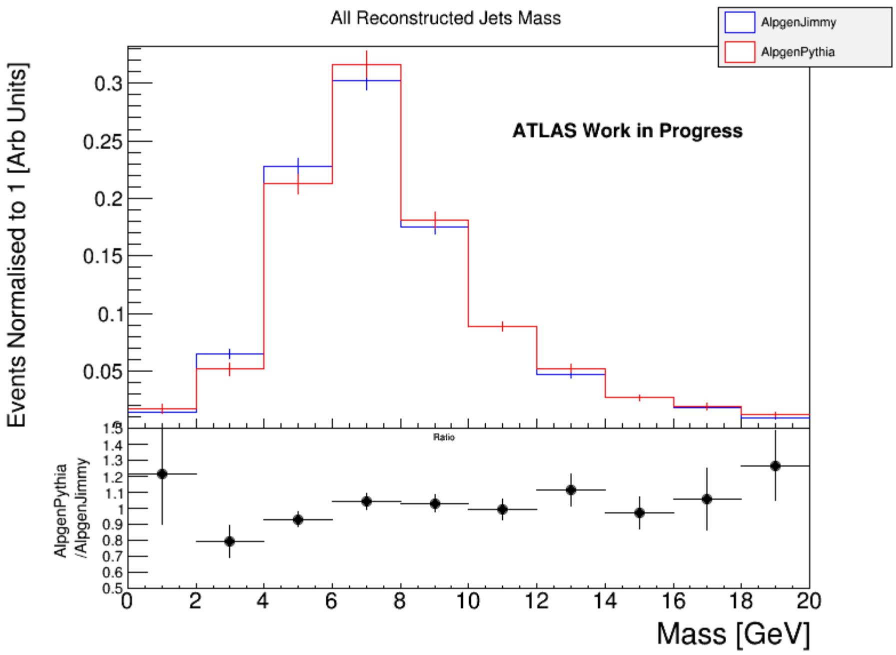


Figure 122: $W \to e\nu$ reconstructed jet mass.

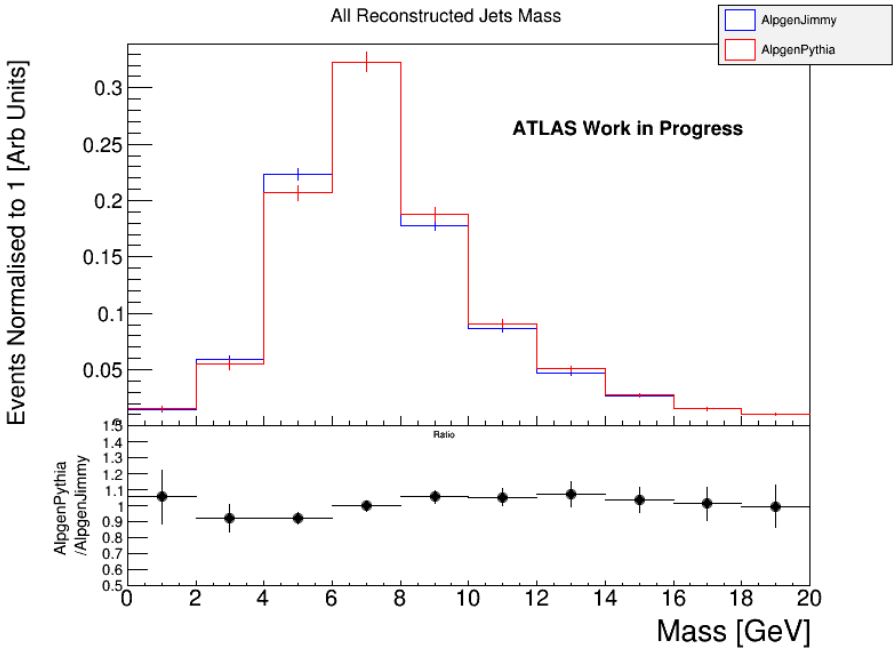


Figure 123: $W \to \mu\nu$ reconstructed jet mass.

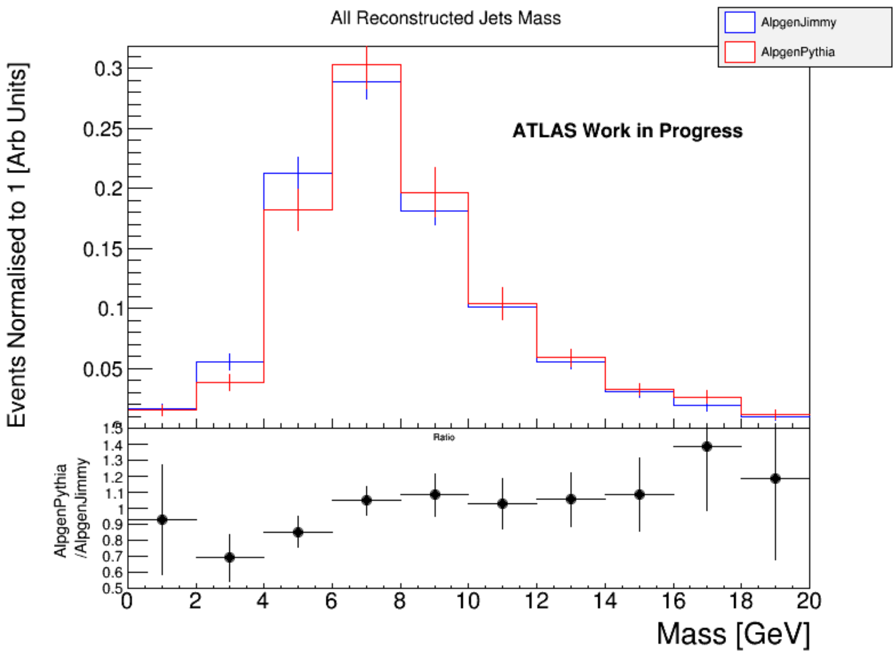


Figure 124: $W \to \tau\nu$ reconstructed jet mass.

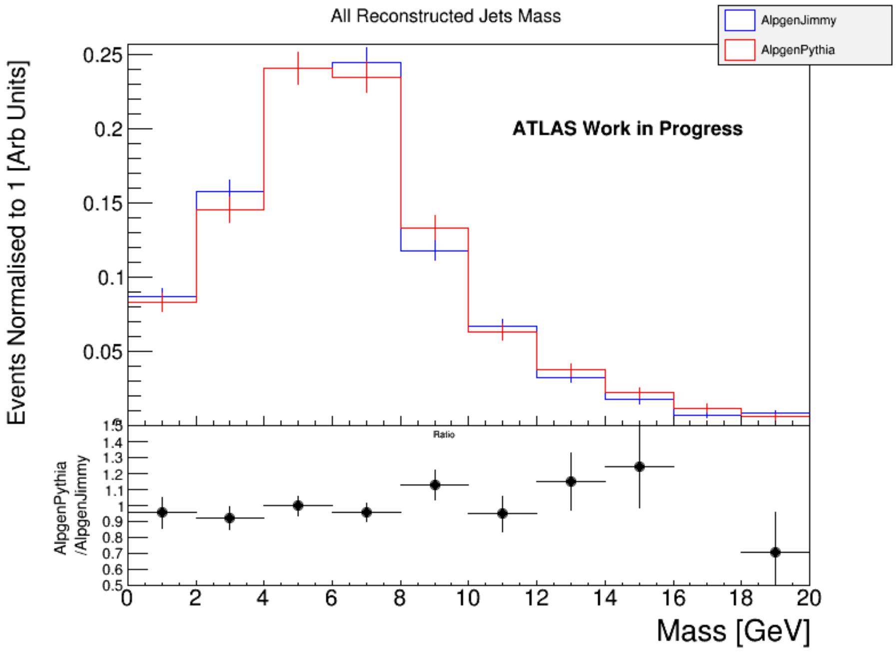


Figure 125: $Z \to ee$ reconstructed jet mass.

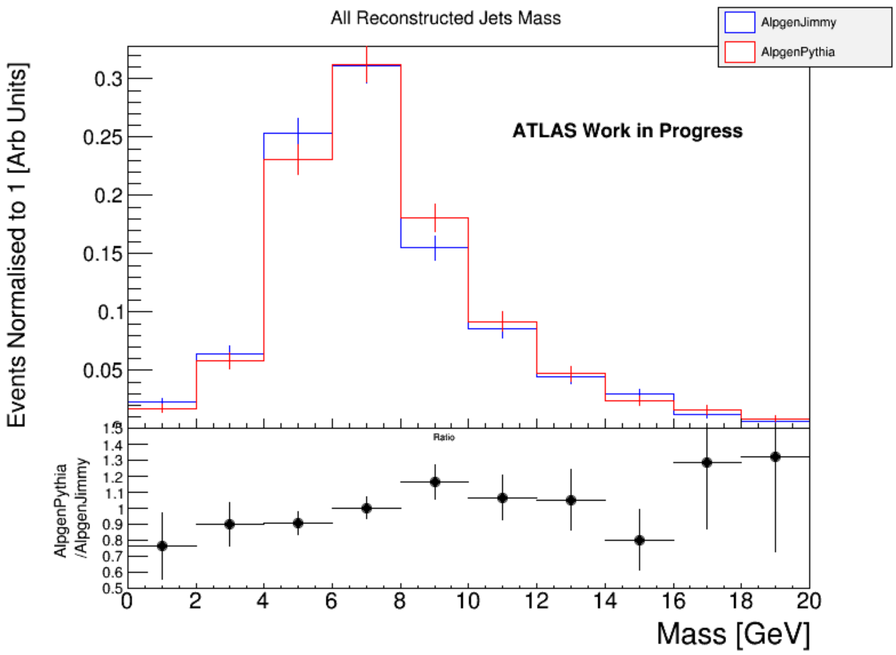


Figure 126: $Z \to \mu\mu$ reconstructed jet mass.

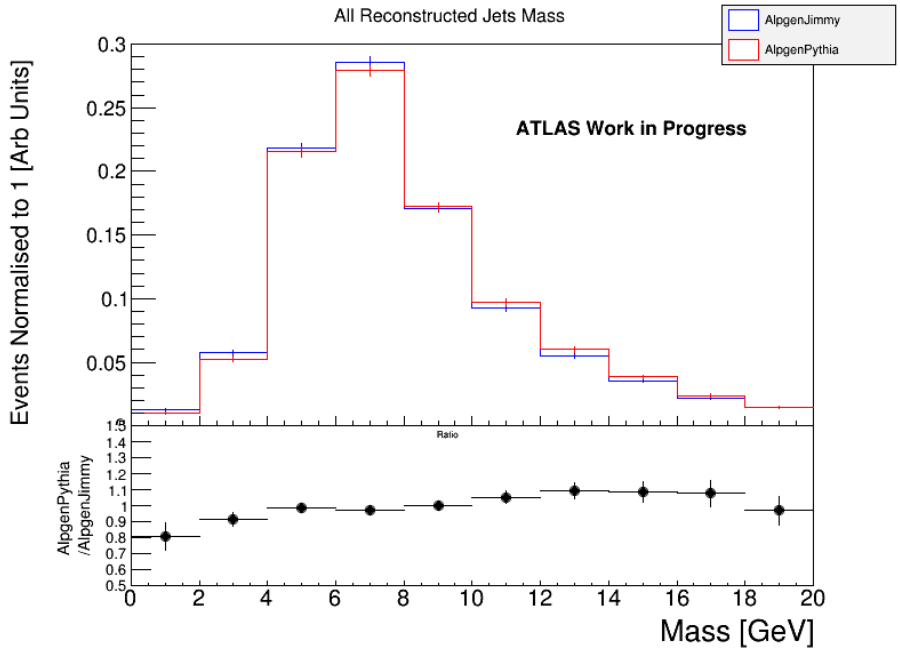


Figure 127: $Z \to \tau\tau$ reconstructed jet mass.

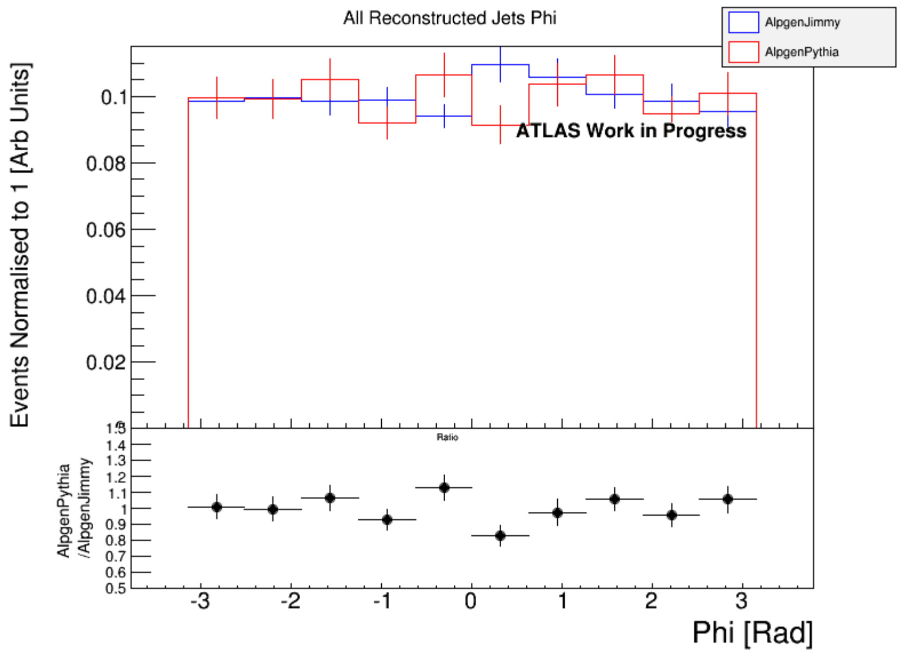


Figure 128: $W \to e\nu$ reconstructed jet ϕ.

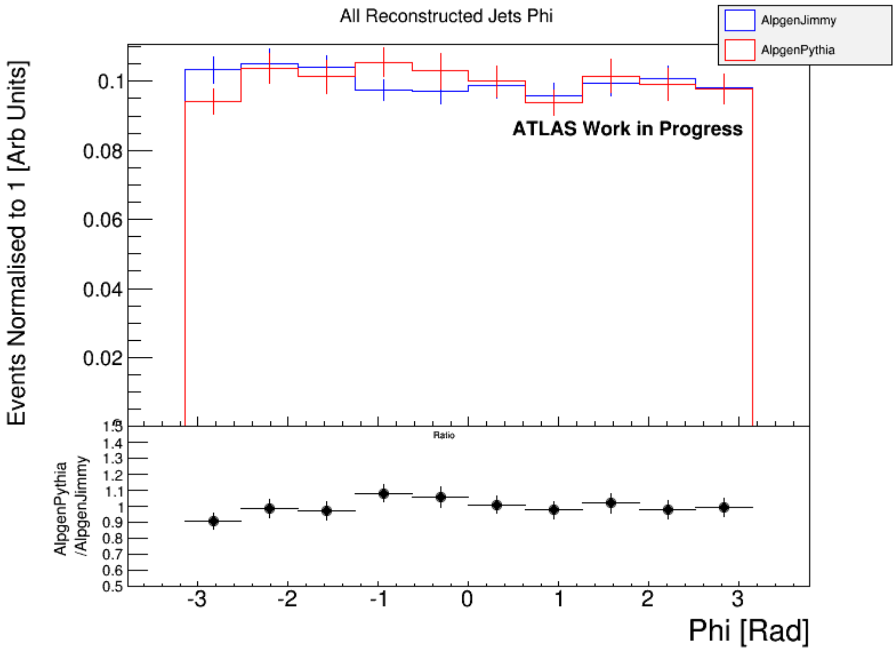


Figure 129: $W \rightarrow \mu\nu$ reconstructed jet ϕ.

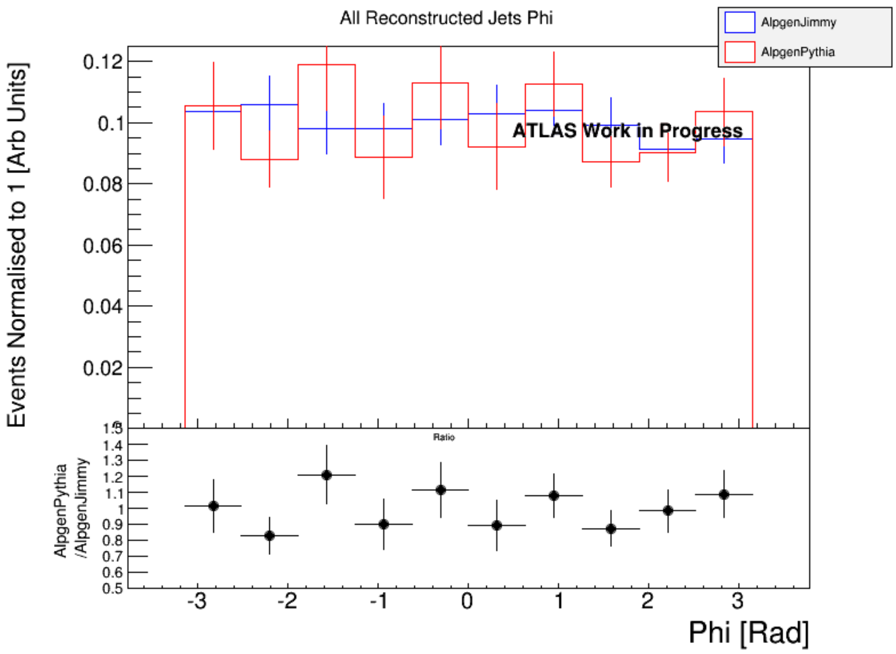


Figure 130: $W \rightarrow \tau\nu$ reconstructed jet ϕ.

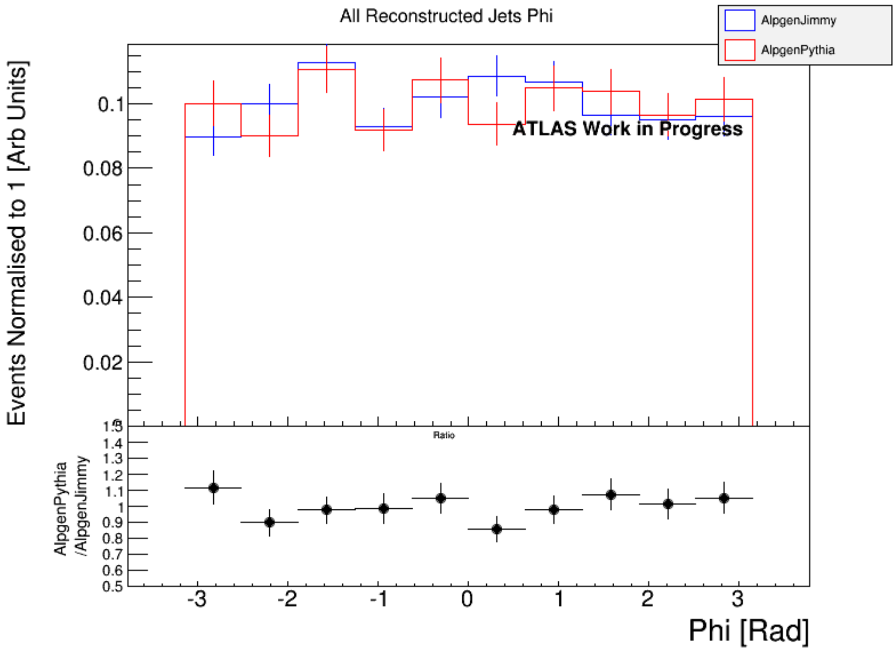


Figure 131: $Z \to ee$ reconstructed jet ϕ.

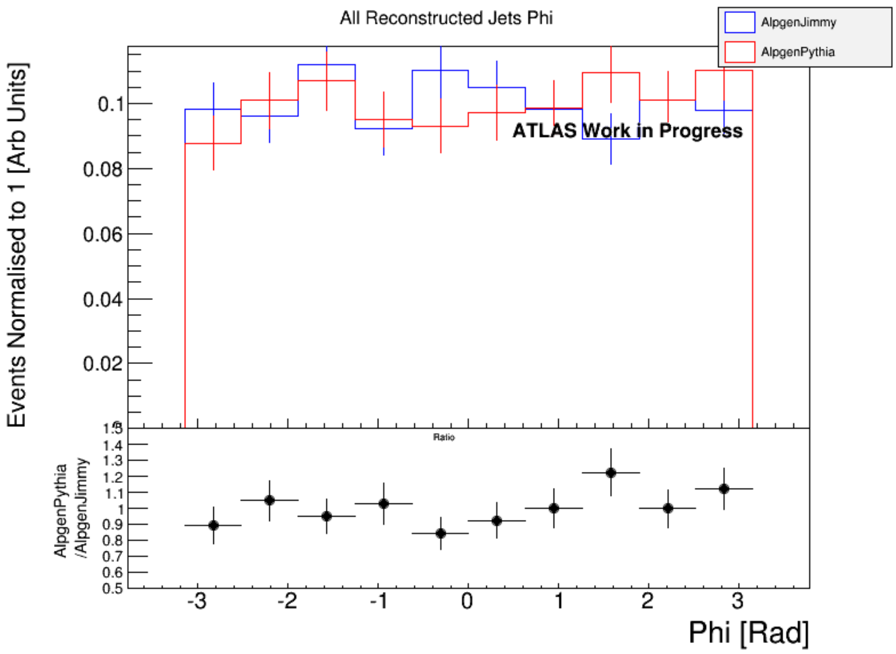


Figure 132: $Z \to \mu\mu$ reconstructed jet ϕ.

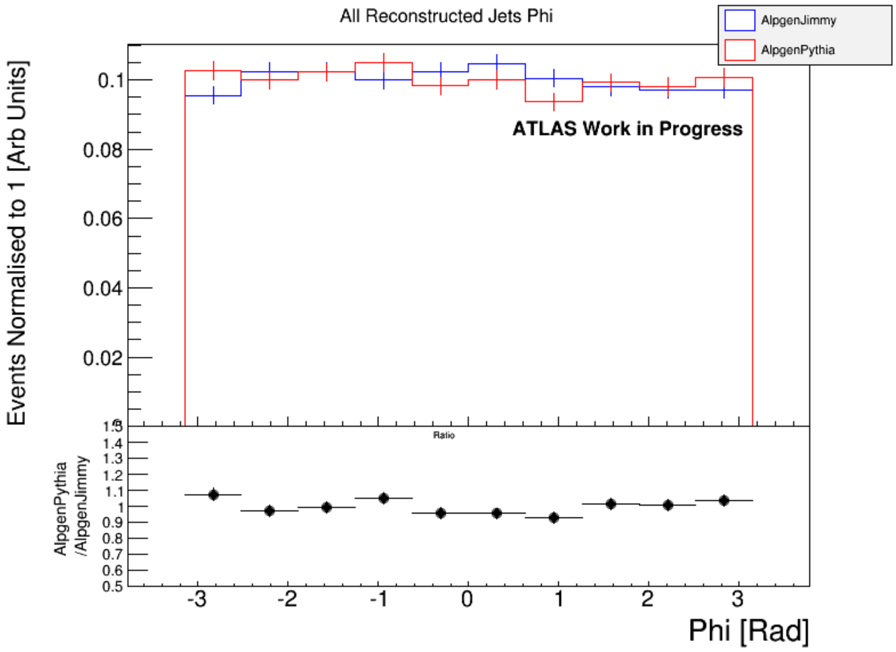


Figure 133: $Z \to \tau\tau$ reconstructed jet ϕ.

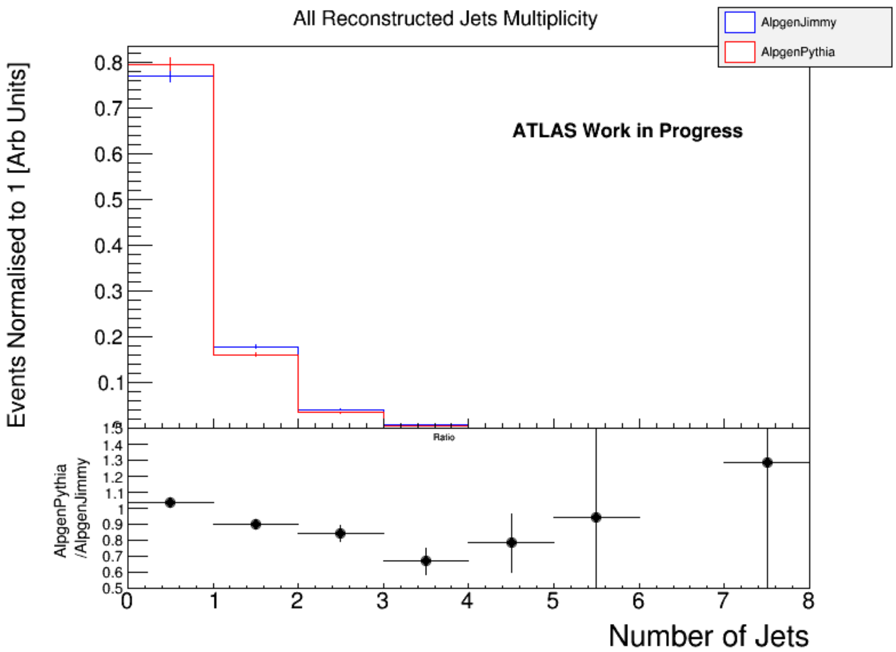


Figure 134: $W \to e\nu$ reconstructed jet multiplicity.

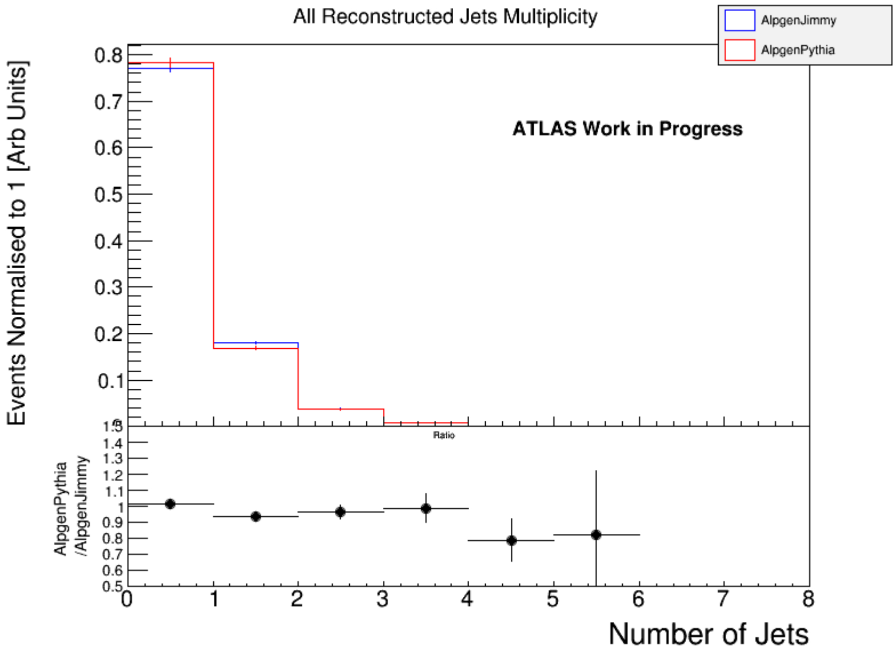


Figure 135: $W \to \mu\nu$ reconstructed jet multiplicity.

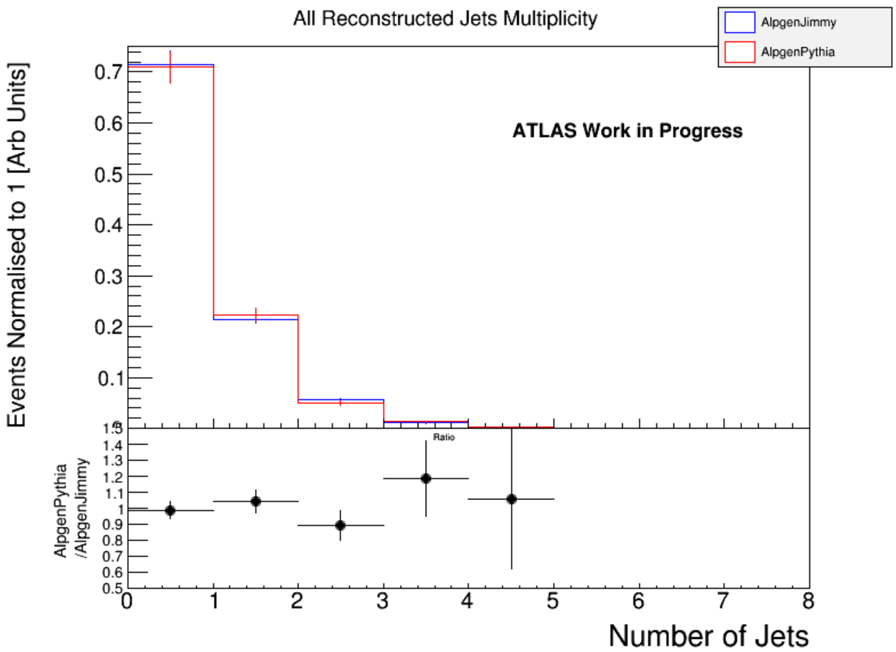


Figure 136: $W \to \tau\nu$ reconstructed jet multiplicity.

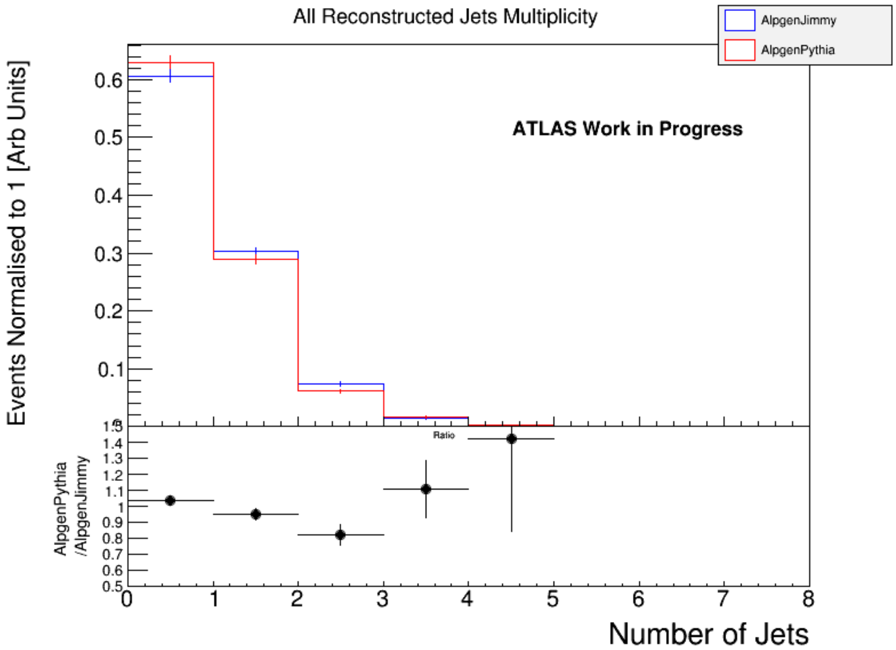


Figure 137: $Z \to ee$ reconstructed jet multiplicity.

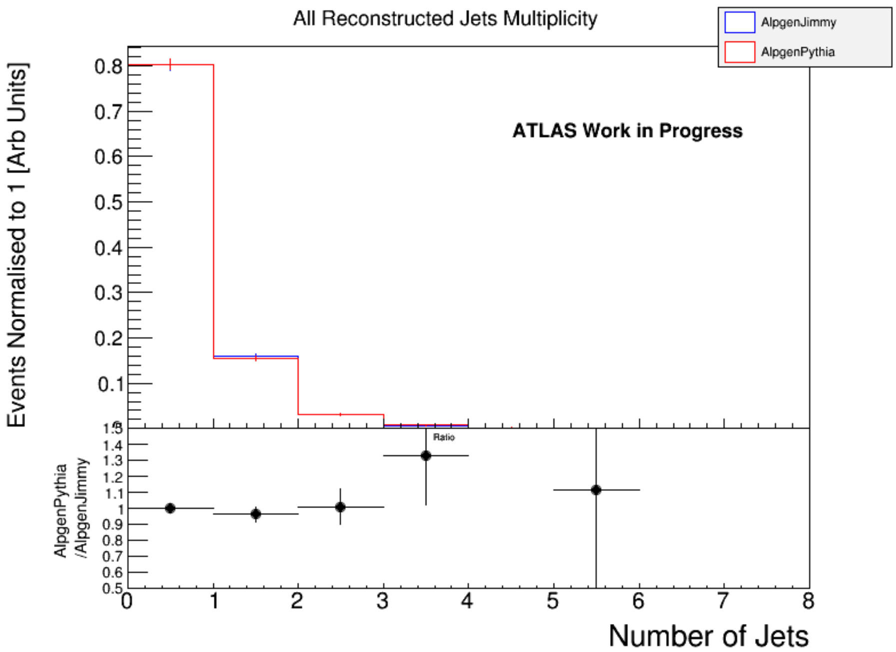


Figure 138: $Z \to \mu\mu$ reconstructed jet multiplicity.

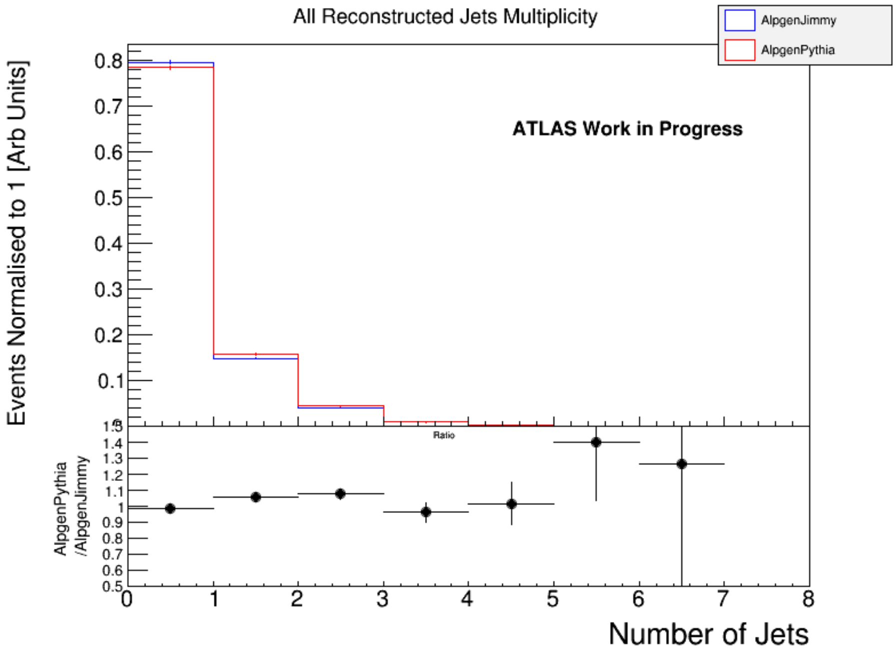


Figure 139: $Z \to \tau\tau$ reconstructed jet multiplicity.

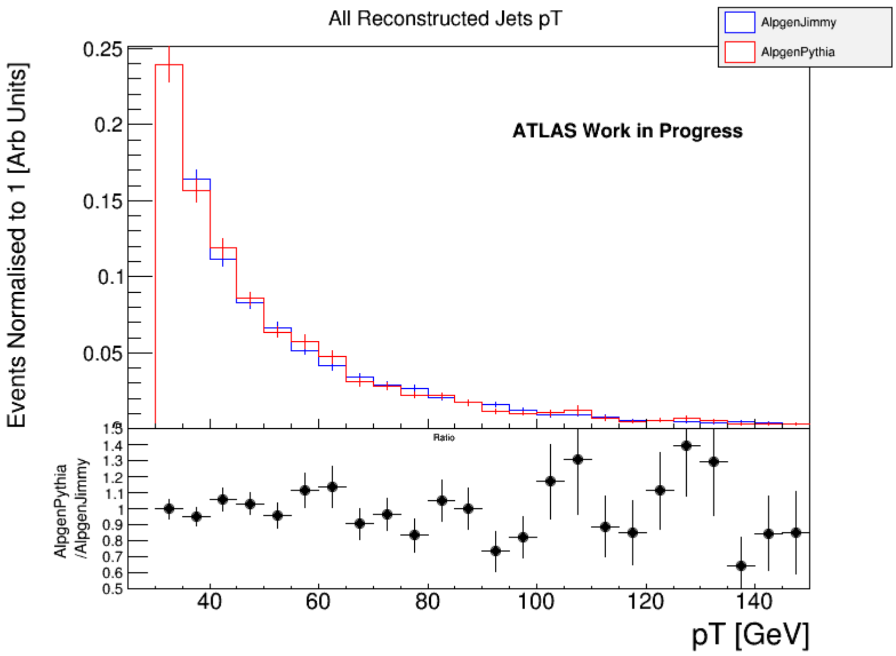


Figure 140: $W \to e\nu$ reconstructed jet p_T.

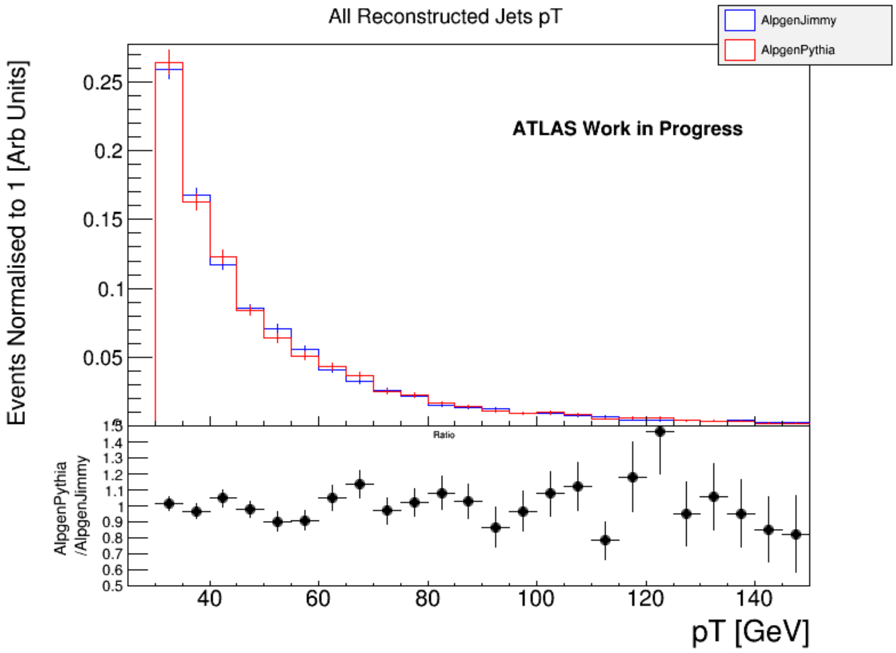


Figure 141: $W \to \mu\nu$ reconstructed jet p_T.

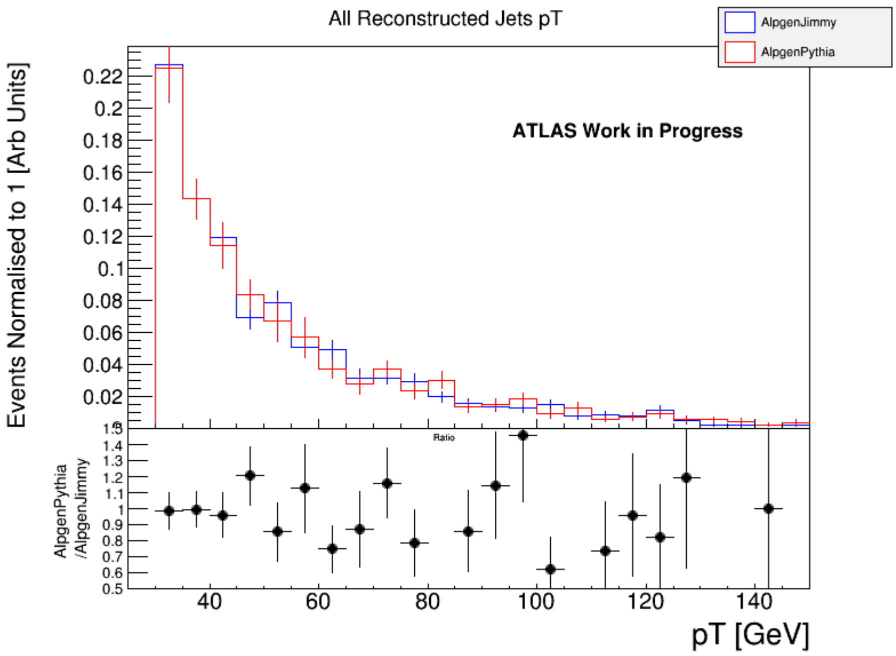


Figure 142: $W \to \tau\nu$ reconstructed jet p_T.

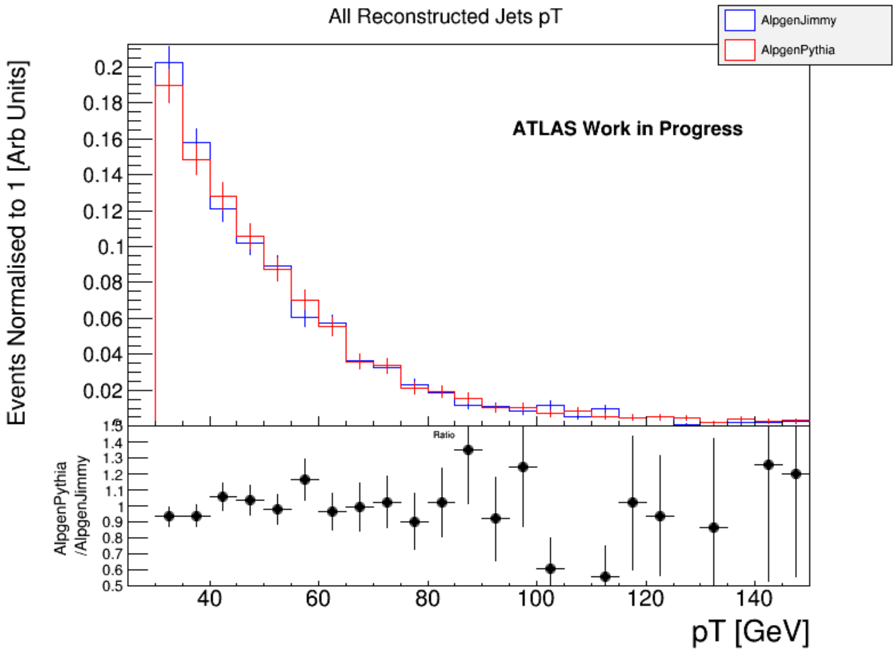


Figure 143: $Z \to ee$ reconstructed jet p_T.

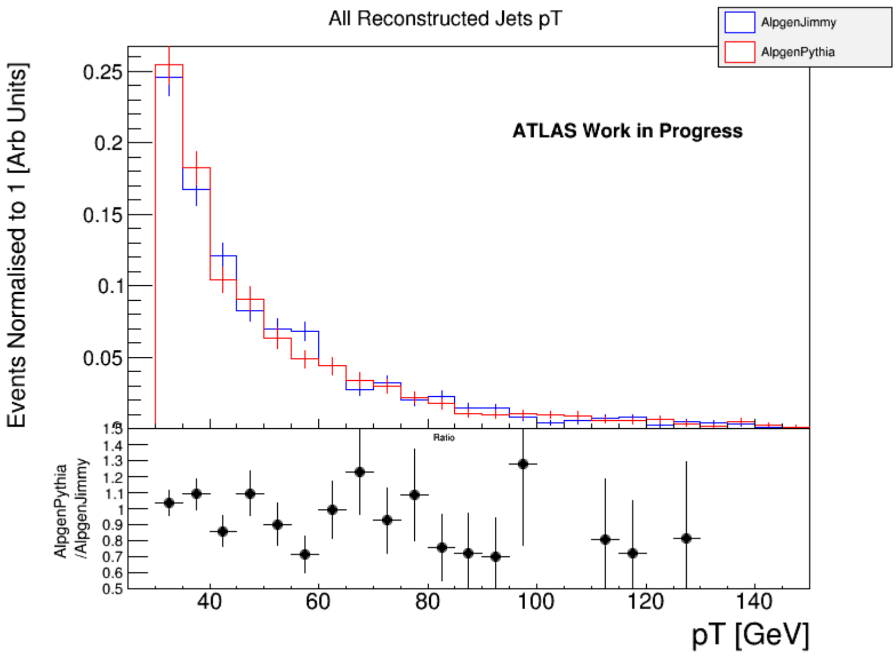


Figure 144: $Z \to \mu\mu$ reconstructed jet p_T.

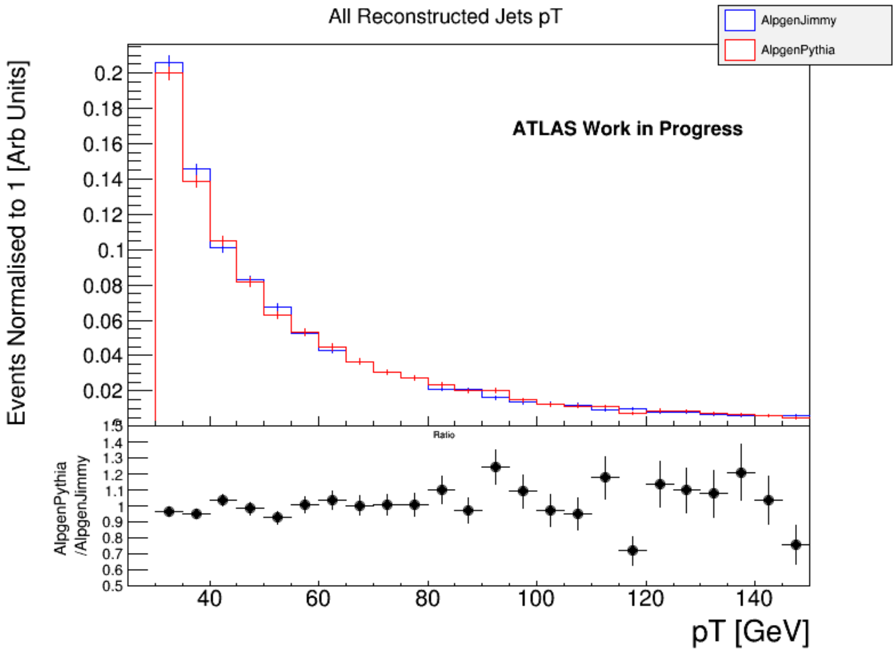


Figure 145: $Z \to \tau\tau$ reconstructed jet p_T.

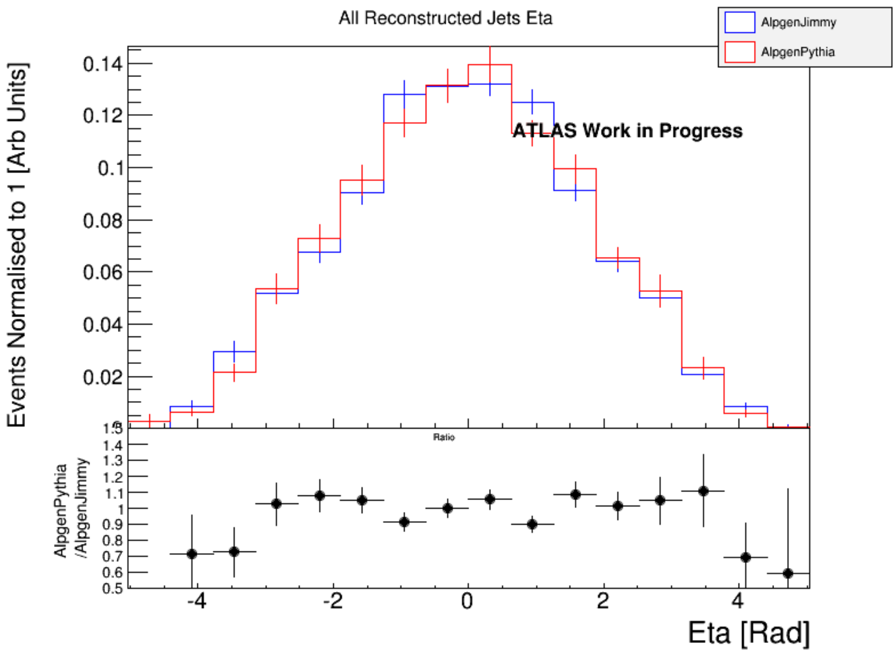


Figure 146: $W \to e\nu$ reconstructed jet η.

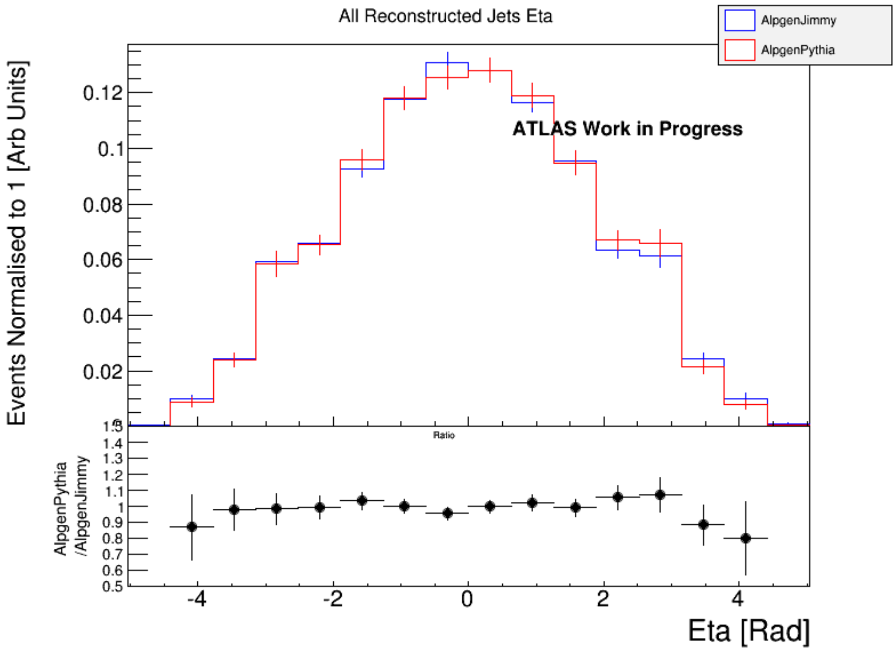


Figure 147: $W \to \mu\nu$ reconstructed jet η.

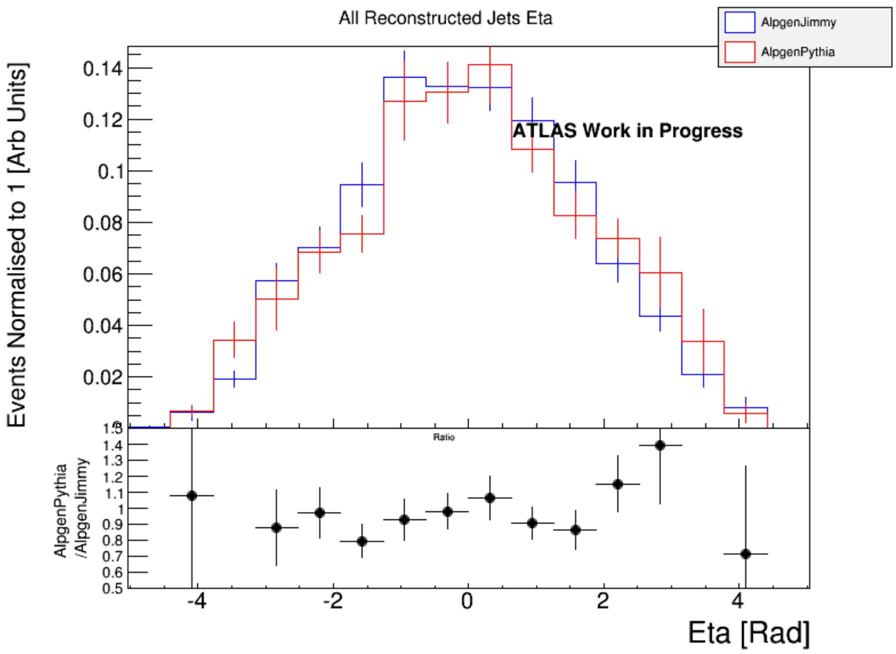


Figure 148: $W \to \tau\nu$ reconstructed jet η.

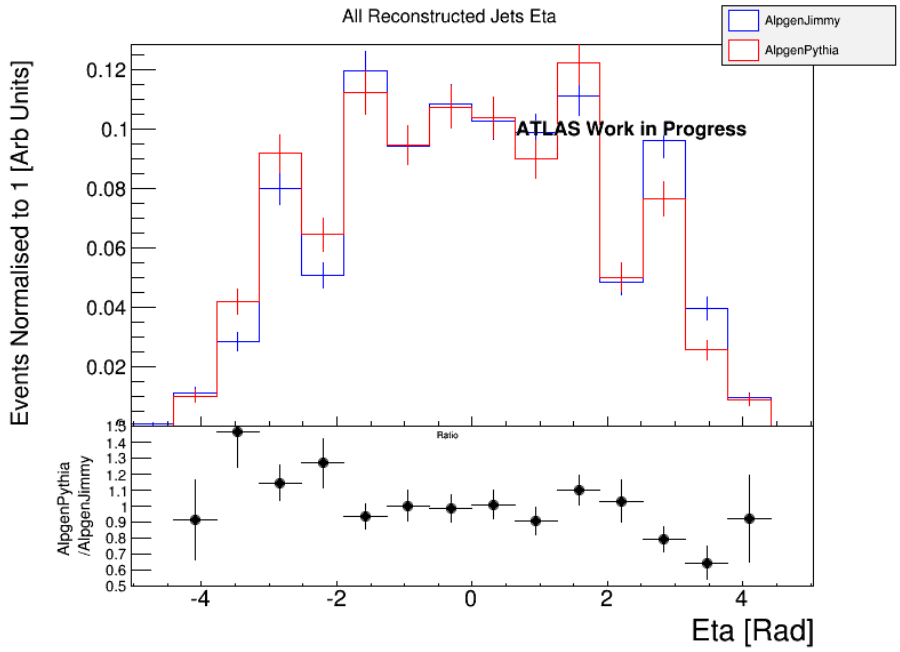


Figure 149: $Z \to ee$ reconstructed jet η.

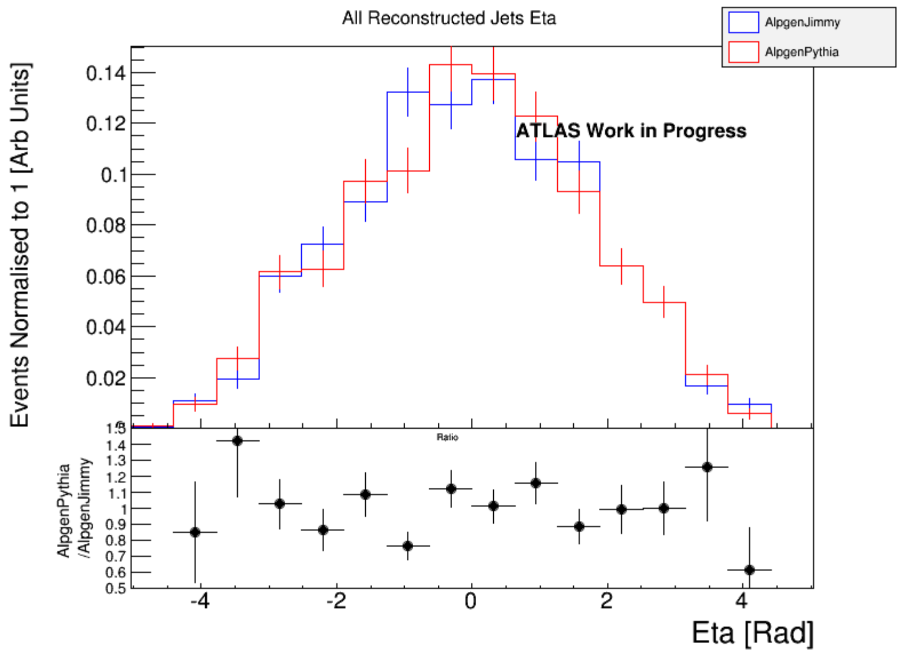


Figure 150: $Z \to \mu\mu$ reconstructed jet η.

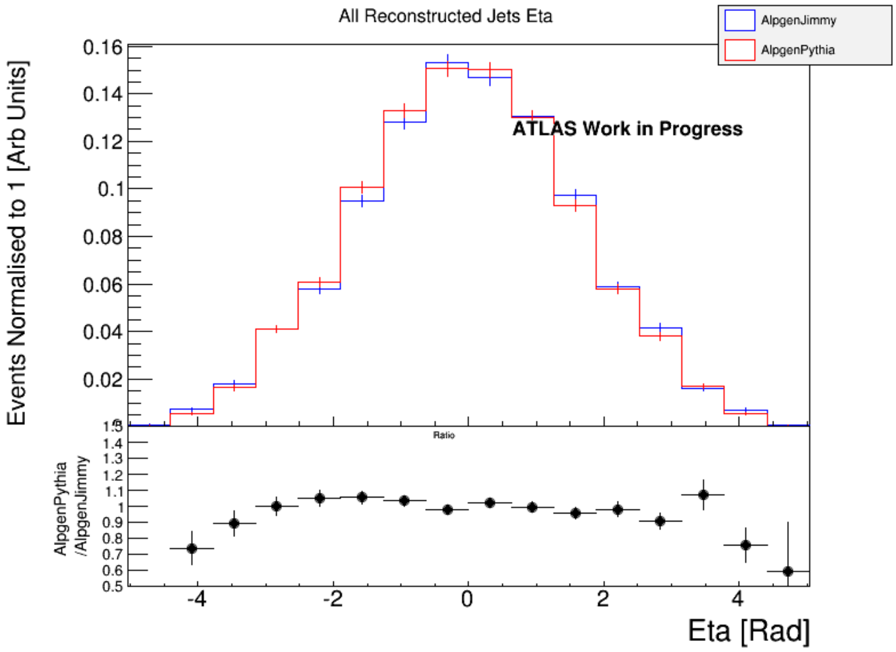


Figure 151: $Z \to \tau\tau$ reconstructed jet η.

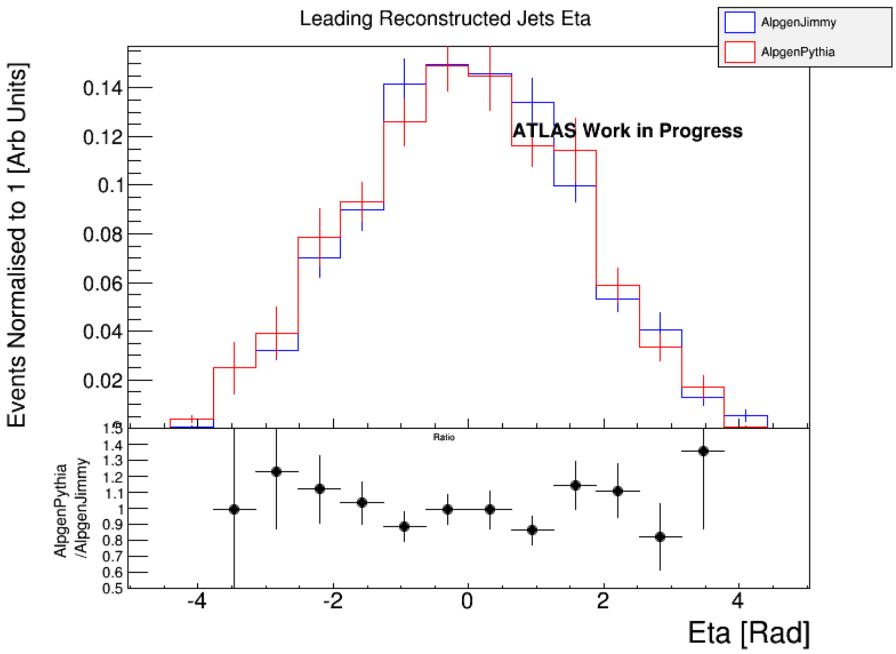


Figure 152: $W \to e\nu$ reconstructed leading jet η.

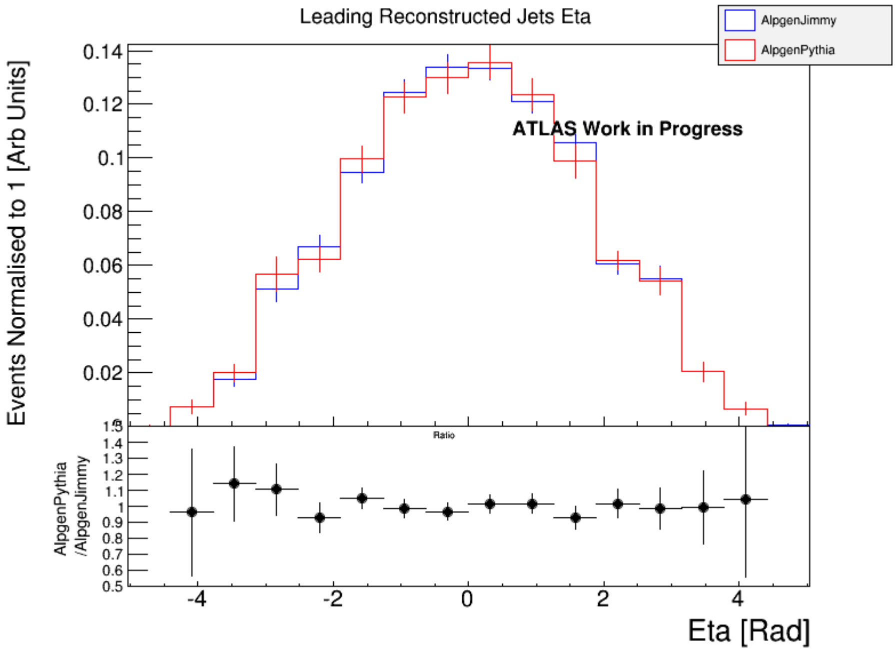


Figure 153: $W \to \mu\nu$ reconstructed leading jet η.

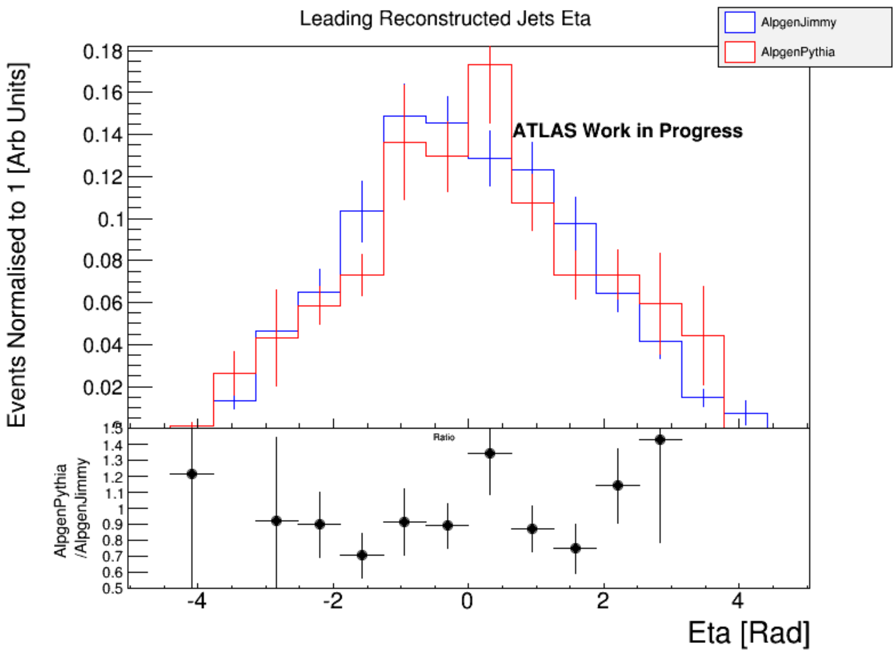


Figure 154: $W \to \tau\nu$ reconstructed leading jet η.

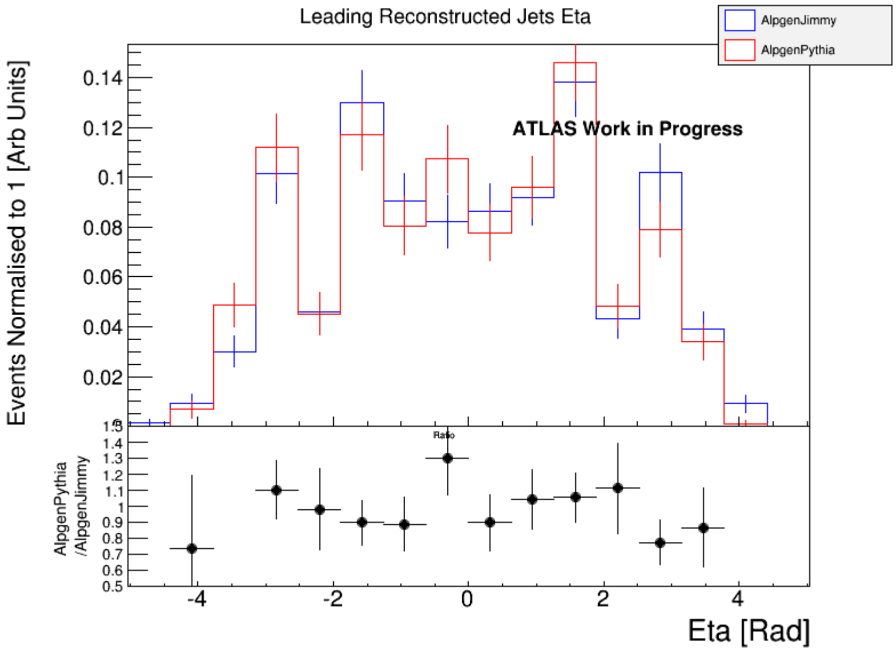


Figure 155: $Z \to ee$ reconstructed leading jet η.

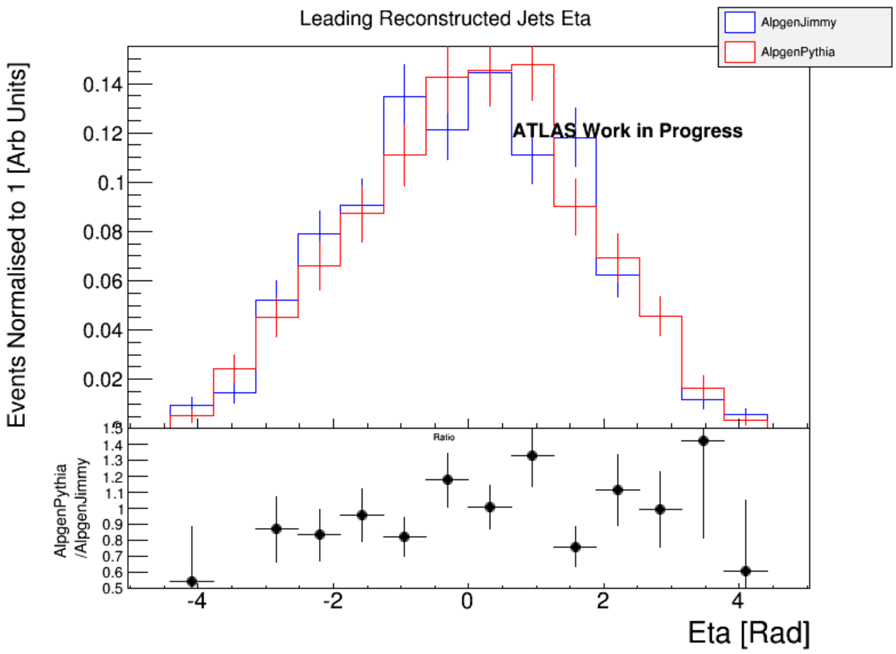


Figure 156: $Z \to \mu\mu$ reconstructed leading jet η.

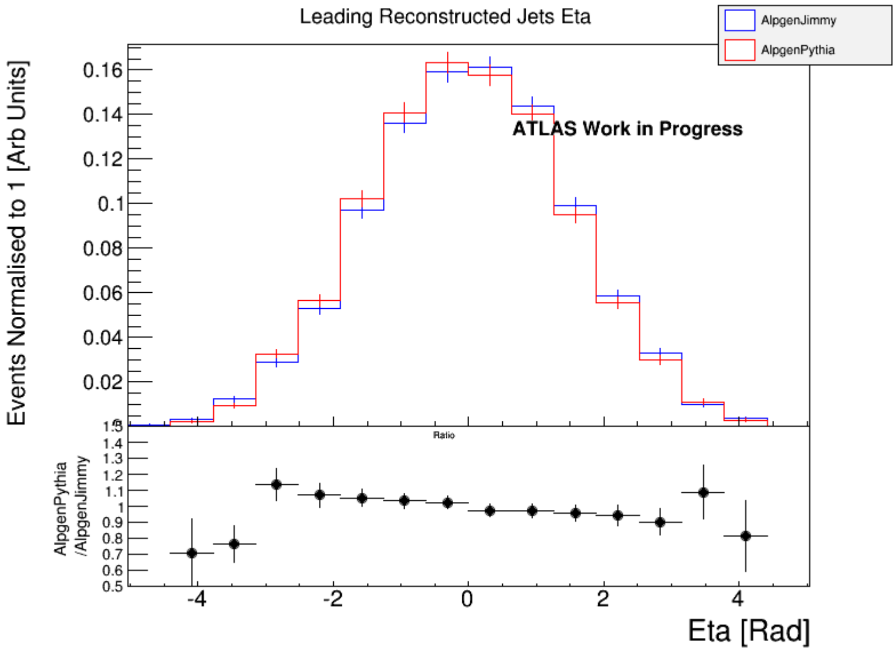


Figure 157: $Z \to \tau\tau$ reconstructed leading jet η.

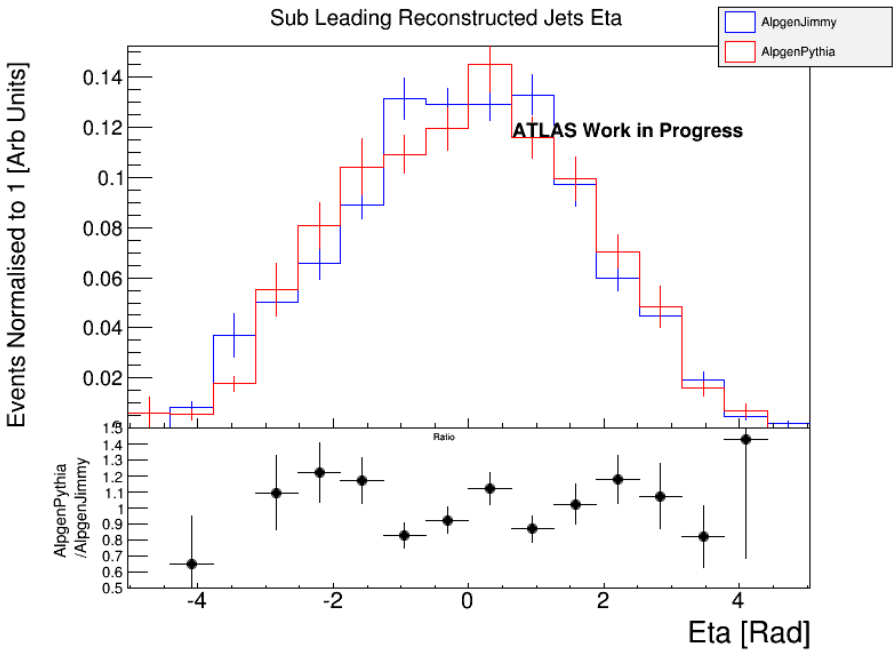


Figure 158: $W \to e\nu$ reconstructed sub-leading jet η.

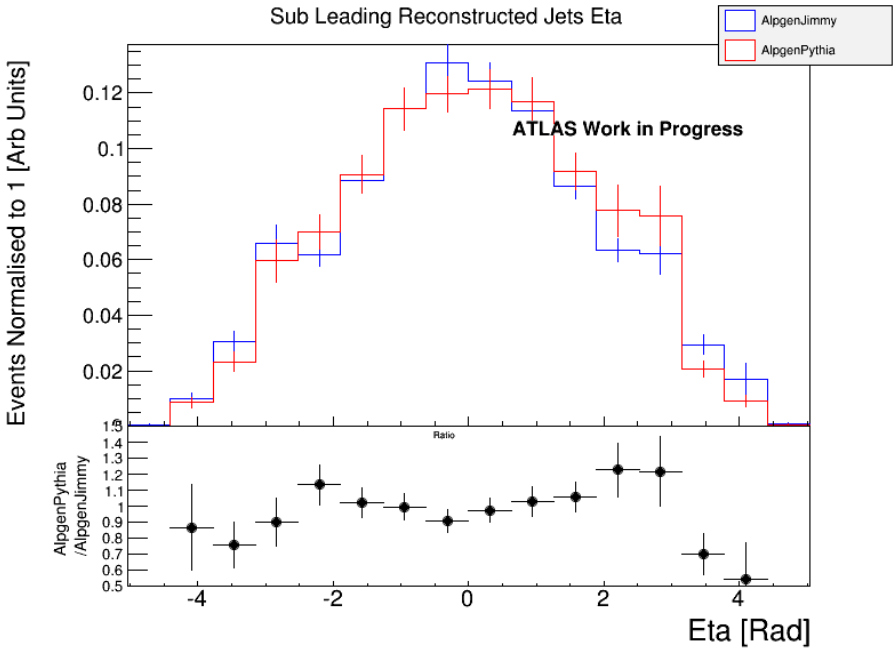


Figure 159: $W \to \mu\nu$ reconstructed sub-leading jet η.

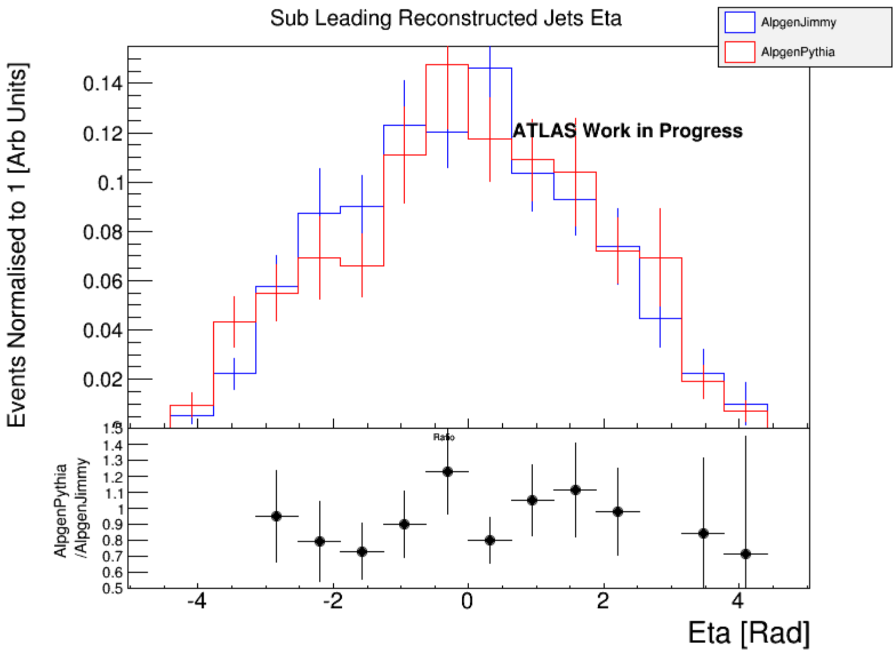


Figure 160: $W \to \tau\nu$ reconstructed sub-leading jet η.

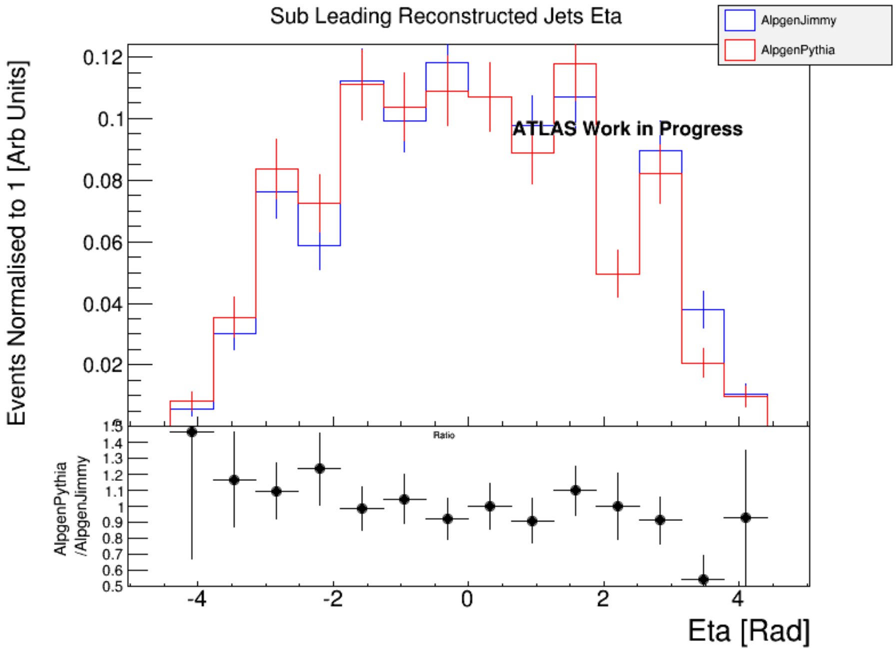


Figure 161: $Z \to ee$ reconstructed sub-leading jet η.

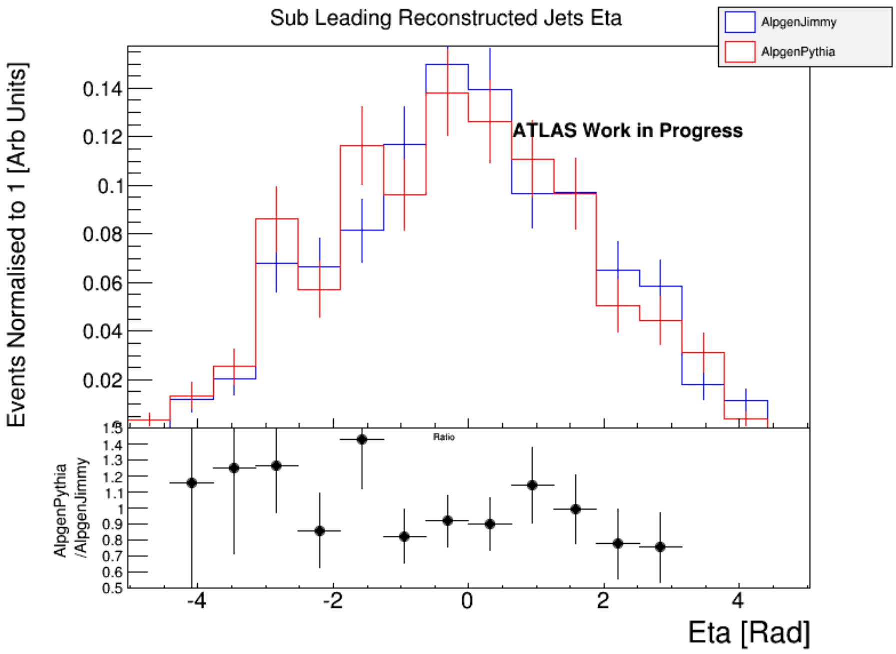


Figure 162: $Z \to \mu\mu$ reconstructed sub-leading jet η.

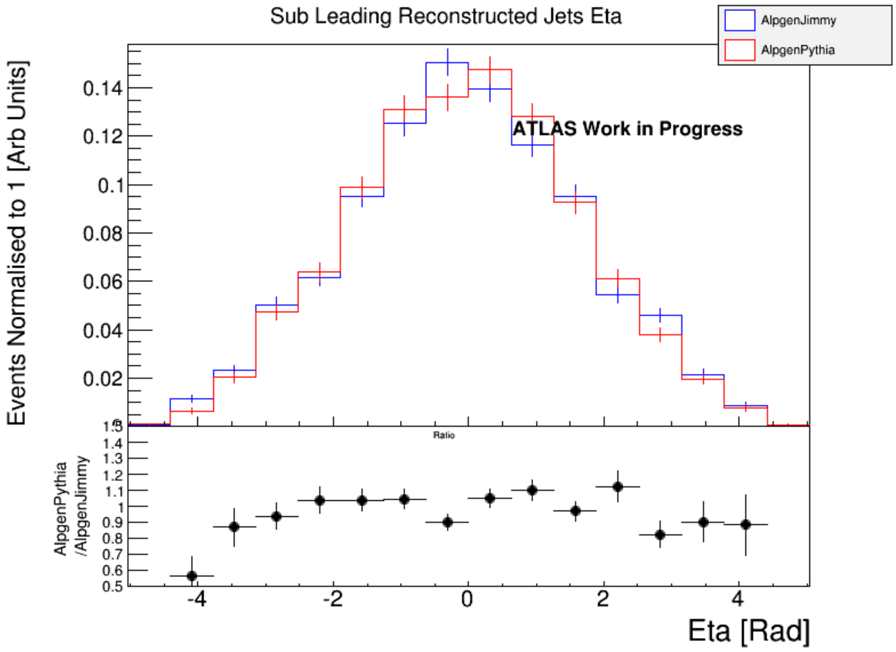


Figure 163: $Z \to \tau\tau$ reconstructed sub-leading jet η.

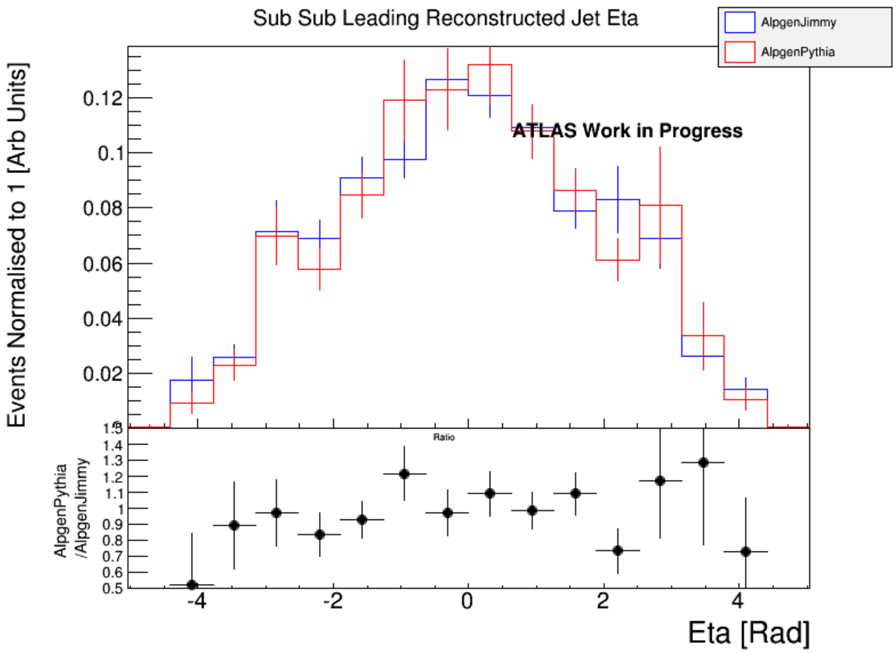


Figure 164: $W \to e\nu$ reconstructed sub-sub-leading jet η.

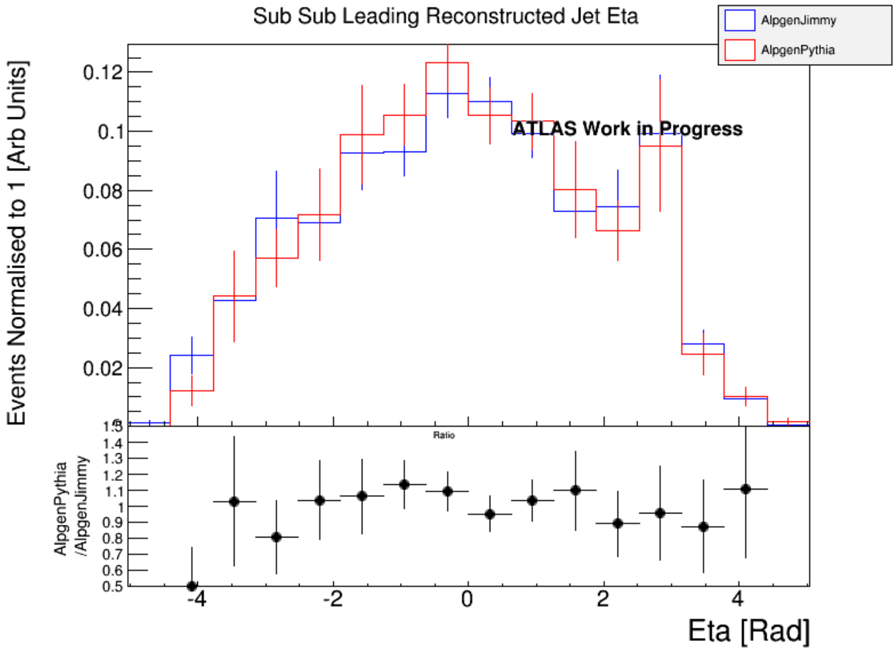


Figure 165: $W \to \mu\nu$ reconstructed sub-sub-leading jet η.

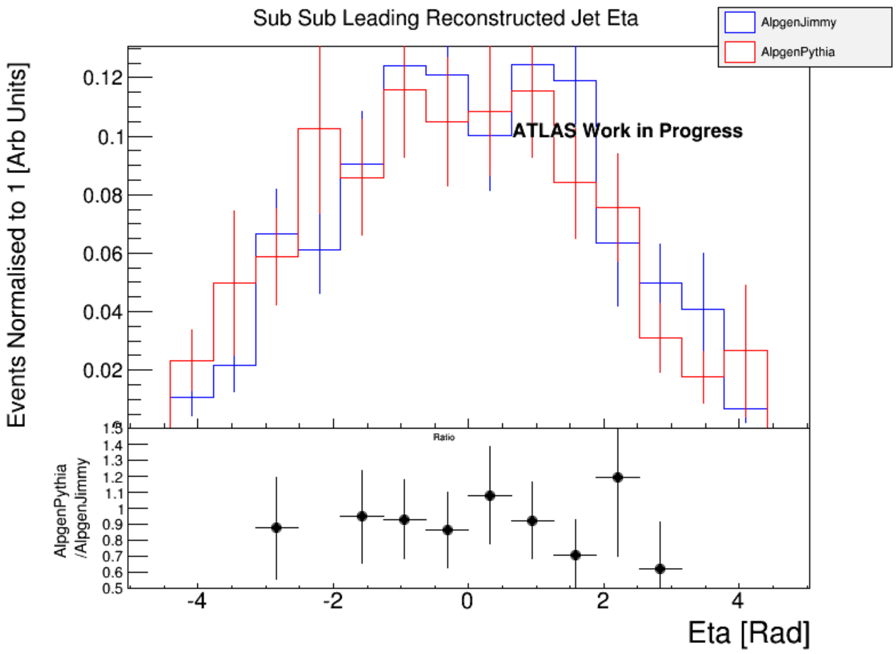


Figure 166: $W \to \tau\nu$ reconstructed sub-sub-leading jet η.

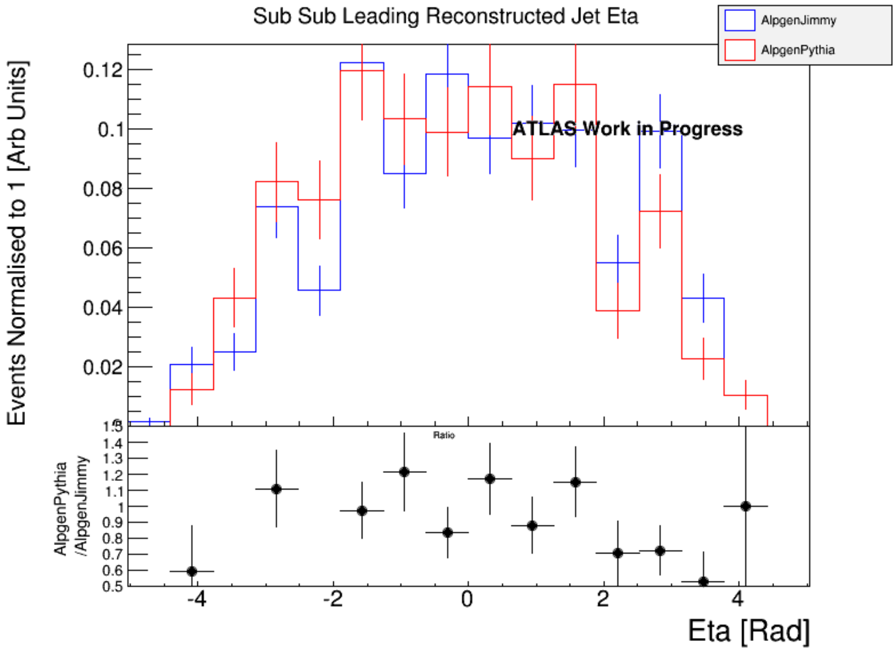


Figure 167: $Z \to ee$ reconstructed sub-sub-leading jet η.

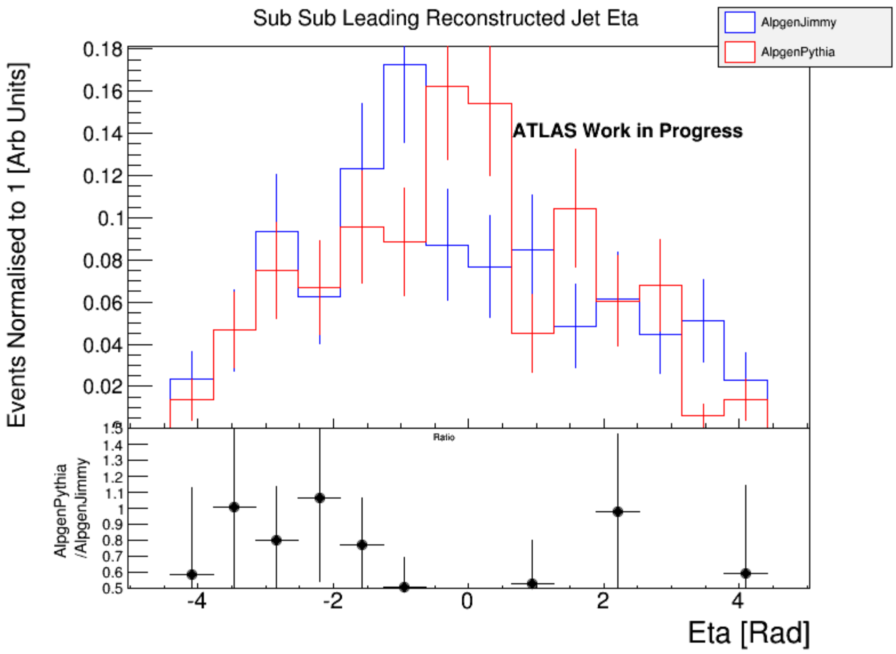


Figure 168: $Z \to \mu\mu$ reconstructed sub-sub-leading jet η.

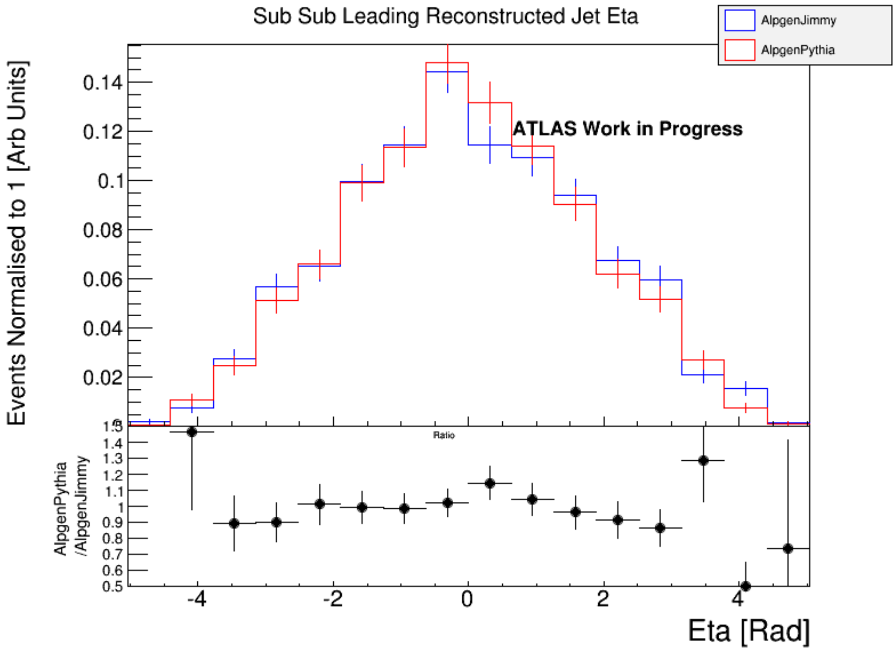


Figure 169: $Z \to \tau\tau$ reconstructed sub-sub-leading jet η.