A Contribution to ATLAS Open Data Collaboration at CERN

Candidate Number: 149596
Supervisor: Lily Asquith
Word Count: 5738
Date: May 15th 2019

University of Sussex

Abstract

The ATLAS Open Data project is a collaborative effort to make high energy proton-proton collision data generated at the Large Hadron Collider accessible to all. The collaboration has previously released an 8 TeV dataset with a corresponding analyses framework. The collaboration aims to release a higher energy, more recent 13 TeV data set with corresponding analyses. The purpose of this project is to contribute to the teaching tool. This was done by aiding in updating the framework built for the previous data set and performing analyses of the Z boson, $W^{\pm}$ boson, ZZ diboson, $t\bar{t}$ and boosted jet to validate them on the 13 TeV Open Data set. It was discovered that the inclusive Monte Carlo simulations generated with Pythia and Powheg did not model the boosted jet and $W^{\pm}$ analysis well. This was resolved for the boosted jet analysis by re-performing the analysis on slices of Monte Carlo generated with Sherpa.

Contents

List of Tables

List of Figures

Preface

Preface

- **Chapter 1:** All work has been written by myself and is a review of information sourced from text books and other researchers. All sources used are cited.
- **Chapter 2:** All information on detector has been sourced from papers and written in my own words. Again all sources cited.
- **Chapter 3:** Written in my own words, sources cited where used.
- **Chapter 4:** The code used to perform analyses on the 13 TeV Open Data set discussed in this chapter was adapted from code written to perform analysis on the previous 8 TeV data set. Much of the core functionality and logic of the code was left unchanged with many minor revisions to correct errors encountered when running over a new data set. For particularly challenging errors, help and advice was received from PhD student Meirin Evans. The files used to setup analyses, such as the configuration files and event selection requirements were created me. Figures used have been cited correctly, the flow chart to represent the codes logic was created by me. The real and Monte Carlo Open Data set came from the ATLAS experiment at the LHC.
- **Chapter 5:** All graphs generated from various analyses were created myself and with the use of CERN's supercluster lxplus server.
- **Chapter 6:** All content was composed by me.
- **Chapter 7:** All content was composed by me.

Chapter 1

Introduction

1.1 ATLAS Open Data

1.1.1 Motivation for this Dissertation

This project aims to contribute to the development of a universal ATLAS Open Data [1] educational tool, accessible to all. The purpose of the first 3 chapters is to present the basic concepts and necessary terminology in a simple format that aids understanding the rationale behind analyses developed and performed on the 13 TeV ATLAS Open Data set. Many publications in the field of experimental particle physics utilise complex language that is inaccessible and discouraging for beginners and non-experts. This dissertation has been written with accessibility in mind and complements the educational nature of the tools developed; where possible explanations have been simplified.

1.1.2 The Educational Tool

When confronted with the odd realities of particle physics and the Standard Model (SM) [2]; textbooks and traditional teaching tend to alienate and disengage students, leaving them to ponder why the material presented to them is relevant to actual particle physics experiments. This is where the ATLAS Open Data project shines and demonstrates itself as a powerful and unique teaching tool; requiring students to not only ask why, but search for answers. Whether needing to isolate specific events, understand what a peak on a graph means or any other thought provoking mystery; it teases inquisitive students to learn the underlying physics of processes to help resolve conundrums. The ATLAS Open Data project gives public access to cutting edge data being generated by high energy collisions at the Large Hadron Collider (LHC) [3] and offers a totally unique teaching opportunity by allowing students to gain hands on data-analysis experience. Easy-to-use tutorials will be provided to demonstrate and teach students and educators how to successfully run pre-made analysis on the 2016 data set and to understand the visualisations created. Results are generated quickly and students are encouraged to delve deeper with suggestions on what to do next, an experience designed to be pain free, yet, very rewarding! Nothing is comparable to seeing a concept taught in a classroom context actually being useful and returning visible results from analysis you have performed. It grounds obscure ideas and subject matter into reality and is essential to show students the importance of these concepts. With no apparent light at the end of the tunnel, no application; students are left with a sense of ambiguity and question whether or not they would like to continue studying the field at all. The ATLAS Open Data project provides this application that is so crucial in swaying and intriguing young physicists to continue their studies and become brilliant particle physicists.

1.2 Physics for Particle Accelerators

1.2.1 Events

At such high energies, collisions can create particles more massive than the initial colliding particles as energy and mass are equivalent. [4] Each collision, also known as an event, produces lots of new particles; many of which are heavy so decay very quickly. These decay products are lighter and have longer lifetimes allowing them to travel far enough to interact with the detector. This is why particle accelerators are used to study the elementary constituents of matter, the decay products of heavy particles can be used to infer their properties. Theoretically predicted particles with large masses and tiny lifetimes, such as the Higgs or Z bosons, can be (and have been) studied and validated through this inference.

1.2.2 Background vs Signal

Inside the ATLAS detector [5] is a very chaotic place while protons are colliding, with over a billion [6] particle interactions per second. Desirable events, known as signal are hidden amongst undesirable events, known as background. The data collection process at ATLAS has many stages dedicated to removing and filtering unwanted background processes. This is done to maximise the ratio of signal to background events, known as the data purity which is defined as:

$$\text{Purity} = \frac{\text{Number of signal events in sample}}{\text{Total number of events in sample}} \tag{1.1}$$

1.2.3 The Quark Parton Model

Traditionally, protons are thought to be formed of three quarks; two up and a down. Yet, this is far from a complete picture of the busy internal sea of gluons and virtual quarks inside the proton. The term parton refers to the charged point-like individual constituents of the proton, now identified as quarks and gluons; the name parton originates from a time before their discovery. In high energy collisions it is the internal partons of the protons that collide. To correctly predict the creation of new particles it is important to determine a distribution function of partons. A good approximation is to view the proton as a stream of quarks and gluons, each carrying a fraction of the proton's overall momentum.

1.2.4 Particle Creation

In order to understand why new particles are created in proton collisions, a concept known as colour charge from the overarching topic of quantum chromo-dynamics (QCD) [7] must be introduced. Electric charge is positive and negative, or charge and anti-charge. In contrast, colour charge has three types; red, blue and green, each with a corresponding anti-colour. All composite particles are colour neutral. Baryons consist of three quarks each having a different colour, the mixture combination produces a colour-neutral particle, much like combining red, green and blue light to produce white light. Mesons consist of a quark and anti-quark, the anti-colour of the antiquark neutralises the meson's colour, this is similar to positive and negative charge cancelling one another out. Quarks interact with one another via the strong force by the exchange of gluons. Gluons are the mediators of the strong nuclear force and are vector bosons, meaning they have spin $1\hbar$. Gluon exchange in the strong interaction is analogous to the exchange of photons in the electromagnetic interaction. However, unlike photons, gluons can also interact with other gluons. As a result, there is a constant force between colour charges as their separation is increased, requiring ever more amounts of energy to increase the distance of separation. [4] In high energy interactions where quarks are knocked out of protons, the force between the quark and the rest of the partons acts like a rubber band that is being continuously stretched; in this analogy the rubber band is the colour force. When the band is stretched enough it can break into segments and the energy stored within each segment is converted into matter and antimatter in a process known as hadronisation, a phenomenon responsible for the confinement of quarks. [8] This leads to a shotgun-like blast of paired quarks and antiquarks all travelling in the same direction and is known as a jet.

1.2.5 Jets

Jets are narrow cones of strongly interacting particles that are the dominant final state objects of high energy proton collisions inside the ATLAS detector. [9]

1.2.6 Centre-of-mass energy

The LHC consists of two beams of equally accelerated protons moving in opposite directions. Bringing two particles to collision allows much higher centre-of-mass energies to be achieved than accelerating one particle into another at rest. The centre of mass energy (E_{cm} or $\sqrt{s}$) of a particle collision is essentially the available energy for physics experiments. With greater E_{cm}, heavier particles can be created and observed in the lab. [10] The 4-momentum $\mathbf{p} = (E, \vec{p})$ of a particle is an expression of its energy and momentum. The square of $\mathbf{p}^2$ is equal to:

$$\mathbf{p}^2 = E^2 - \vec{p}^2 = m^2, \tag{1.2}$$

where m is the particle's invariant mass. In the collision of two particles, the total centre of mass energy is:

$$E_{cm}^2 = (\mathbf{p}_1 + \mathbf{p}_2)^2 = (E_1 + E_2)^2 - (\vec{p}_1 + \vec{p}_2)^2. \tag{1.3}$$

The beams of protons in the LHC have equal momenta in opposite directions, hence $|\vec{p}_1| = |\vec{p}_2|$. The centre of mass energy becomes:

$$E_{cm} = (E_1 + E_2)^2, \tag{1.4}$$

where E_1 and E_2 are the energy of a proton in...

1.2.7 Luminosity

An important indication of the performance of the LHC is instantaneous luminosity; it is the number of particle interactions per second. The integrated luminosity is the number of interaction over a period of time, calculated from the integral of luminosity ($\mathcal{L}$) over the run time of the collider (T) [10]:

$$\mathcal{L}_{int} = \int_0^T \mathcal{L}(t')\, dt' \tag{1.5}$$

and is measured in inverse femtobarns fb^{-1}. 1 fb^{-1} equates to 100 trillion (10^{12}) proton-proton collisions. [11]

1.2.8 Cross Section

Cross-section, represented by σ is defined as:

$$\sigma = \frac{N}{\mathcal{L}_{int}}\epsilon, \tag{1.6}$$

where N is the number of events, $\mathcal{L}_{int}$ is the integrated luminosity and ϵ is a factor accounting for the efficiencies of the detector and process selection. [4] It is proportional to the probability for a process to occur during any given particle-particle collision. Processes with larger cross sections occur more often than processes with smaller cross sections. Cross sections are used as they enable the comparison of data between particle physics experiments with different measurement techniques.

1.2.9 Missing Transverse Momentum

Transverse momentum (p_T) is an extremely useful variable to monitor in analyses; it is the component of momentum perpendicular to the beamline. Most standard model particles leave tracks inside the detector, yet some such as SM neutrinos and particles predicted beyond the standard model do not interact with the matter in the detector. [12] Since they are not directly measurable, when searching for these particles we must infer their existence from apparent violations of the law of conservation of momentum. The total transverse momentum before the collision is zero, any 'missing' p_T after the collision can be accredited to these non-interacting, invisible particles.

1.2.10 Pseudorapidity

The standard coordinate of a particle relative to the beamline is pseudorapidity (η). It is defined as:

$$\eta \equiv -\ln\left[\tan\left(\frac{\theta}{2}\right)\right] \tag{1.7}$$

where θ is the angle of the particle's $\vec{p}$ with the beam axis. [13]

1.2.11 Pile-up

When the two beams are crossed at the centre of the ATLAS detector, interesting particles are created when protons collide in head-on, hard scattering [14], interactions. These interactions are relatively rare and in order to increase their probability the beams of protons in the LHC are bunched rather than continuous. [15] When bunches cross at the centre of the detector multiple proton collisions often occur, the number of collisions per bunch-crossing is known as pile-up. Increasing pile-up is a double edged sword, on one hand a higher number of interesting events occur, on the other the difficulty of isolating signal events also increases.

1.3 Physics for Analyses

1.3.1 The Z^0 and $W^{\pm}$ Bosons

The neutral Z and charged $W^{\pm}$ bosons are all mediators of the weak force, analogous to the photon being the mediator of the electromagnetic force.

Z bosons allow neutrinos to interact with other particles, [16] neutrinos are charge-neutral and so cannot interact electromagnetically via photon exchange. Photons and Z bosons are closely related, however, their masses drastically change the range of their corresponding

mediated force. Electric charges can feel each other at very large distances due to the massless nature of the photon, enabling it to travel an infinite distance. On the other hand, the Z boson has a very large mass (91.1876 ± 0.0021 GeV [17]) and therefore an extremely short lifetime (half life of 3×10^{25} s [17]), as a consequence it can only travel a short distance.
Generation percentages and decay Z bosons are produced in two main processes; Drell-Yan [18] accounts for roughly 65% of Z bosons generated in collisions and quark gluon scattering which accounts for 35%. In the Drell-Yan process a quark from one proton and an oppositely-charged anti-quark from another annihilate, producing a photon or a Z boson as the net electric charge is zero. In the quark gluon scattering process a quark and gluon interact, resulting in the quark emitting a Z boson. Due to its high mass, the Z boson decays almost immediately after production. The charges of the decay products of the Z boson must sum to zero due to its neutral-charge; all particle interactions must conserve charge. Z bosons therefore must decay into a particle-antiparticle pair. Lepton-antilepton pairs account for roughly 10% of Z boson decays. [19] Each of the three flavours of lepton pairs; electron-positron, tau-antitau and muon-antimuon have equal probability. Neutrino-antineutrino pairs account for approximately 20% of decays [19], again the three flavours of neutrino are equally probable. Neutrinos do not interact with the detector so must be inferred from missing transverse momentum (see section 1.2.9). The remaining 70% of Z boson decay events produce a quark-antiquark pair, which are seen as jets.

Z^0 Decay Product	Symbolic	Percentage
Lepton-Antilepton pair	$Z^0 \to \ell^- \ell^+$	10%
Neutrino-Antineutrino pair	$Z^0 \to \nu\bar{\nu}$	20%
Lepton-Antiquark pair	$Z^0 \to q\bar{q}$	70%

The $W^\pm$ bosons are responsible for transforming protons into neutrons and neutrons into protons, allowing the process of nuclear fusion. Since $W^\pm$ are charged particles, to conserve charge their decay products must have an overall charge of ±1. Production of a quark anti-quark pair occurs two thirds of the time and creation of a lepton and neutrino of the same flavour occurs the remaining third. [19]

$W^\pm$ Decay Product	Symbolic	Percentage
Lepton-Antineutrino pair	$W^- \to \ell^- \bar{\nu}$	
Antilepton-Neutrino pair	$W^+ \to \ell^+ \nu$	33.3%
Quark-Antiquark pair	$W^\pm \to q\bar{q}$	66.6%

1.3.2 Diboson

The production of diboson pairs in the LHC allows the study of the electromagnetic and weak sector of the SM at the highest available energies. Dibosons are formed in SM processes such as the production of ZZ in the decay of the Higgs boson [20]. However, many theories beyond the standard model predict new heavy particles that decay into dibosons and so an understanding of these events is critical in making future discoveries. The ZZ dibosons leptonic decay events are considered in the ZZ analysis.

Diboson	Decay
ZZ diboson pair	$Z \rightarrow \ell^- \ell^+ \ell^- \ell^+$

1.3.3 $t\bar{t}$

The top quark has the largest mass of all the quarks in the standard model and thus a top anti-top quark pair is highly energetic. These high energy particles push the limits of the theoretical models used to build Monte Carlo (MC) data sets (see section 3.1). Hence, top pair data is an ideal process to test and validate the theoretical models used. In addition, top-antitop pair production is a key background in many Higgs boson analyses. [21] The two top quarks almost always decay into two W bosons and two bottom quarks. [17] The W bosons decay as described in section 1.3.1 and the bottom quarks undergo hadronisation and are seen as jets in the detector. A specialised algorithm identifies jets originating from b quarks and b-tags them. [22]

$t\bar{t}$ Decay Product	Symbolic	Percentage
Two W bosons and bottom quark pair	$t\bar{t} \rightarrow W^+W^- \, b\bar{b}$	100%

1.3.4 Boosted Jets

Objects described as boosted are particles with high velocities relative to the lab frame. When colliding particles at energies as high as those seen at the LHC, new particles are often created well above their invariant masses and any remaining creation energy can highly boost them. [D] If a boosted particle decays into a quark-antiquark pair their individual jets will overlap. With sufficient boost, the two jets become indistinguishable from one another and the decay 'fuses' into a single high energy, boosted, 'fat jet'. The invariant mass and internal structure of these highly energetic fat jets provide additional clues to the process responsible for their creation in the collision event. The immense resolution of the ATLAS calorimeters allow for these properties to be investigated. [23]

Chapter 2

The ATLAS Detector

2.1 Detector Structure

In order to generate useful event data, the ATLAS detector must have a high enough resolution to separate approximately 1000 particles emerging from the collision point every 25 ns in both space and time. [13] It is necessary for different types of particle to be distinguishable from one another and their energies and momenta measurable. The ATLAS detector consists of two main types of detector components; trackers and calorimeters (see figure 2.1. Some components are immersed in powerful magnetic fields, up to 2 T, [6] to bend the trajectories of charged particles; the curvature of these particles can be used to measure their momenta. [24] The inner detector is the first element that decay products of proton-proton collisions pass through. It is immersed in a 2T magnetic field and provides precise tracking, momentum and charge measurement of electrically-charged particles in the pseudorapidity range $|\eta| < 2.5$. Calorimeters are designed to measure the energy and position of particles by total absorption. Two types of calorimeter can be used to measure almost all of the SM particles, excluding neutrinos and muons. [4] Photons and leptons are detected with electromagnetic calorimeters, abbreviated to ECal. Two ECals are utilised in the detector; a higher resolution one covering the pseudorapidity range $|\eta| < 3.2$ and a lower resolution one covering $|\eta| < 4.9$. Hadrons are detected with the hadronic calorimeter, HCal, which covers the entire pseudorapidity range $|\eta| < 4.9$. To make measurements of muons, several types of muon spectrometers are located at the outermost section of the ATLAS detector covering ranges $|\eta| < 2.4$ and $|\eta| < 2.7$. [25] There is often a trade-off between more precise higher resolution measurements and the range covered by the measuring device.

2.2 Triggers

With over 40 million proton collisions per second [6] inside the ATLAS detector, it is impossible to record and save everything. A filter, called a trigger, must be used to cut out unwanted events that make up the vast majority of the total. The conditional requirements set are chosen very carefully to remove as many unwanted events as possible without removing interesting the interesting ones. Candidate particles must meet conditions such as a minimum energy or have a particular property in an allowed range. Ultimately, triggers are designed to make data analysis simpler and increase the purity of recorded data. However, with too high or too narrow ranged a threshold, desirable events will also be lost. Within the competitive environment at the LHC with research groups rushing to be credited with the initial discovery of new particles and each having access to the same energy of proton collisions, choice of trigger conditions are extremely important in gaining an edge to find new particles.

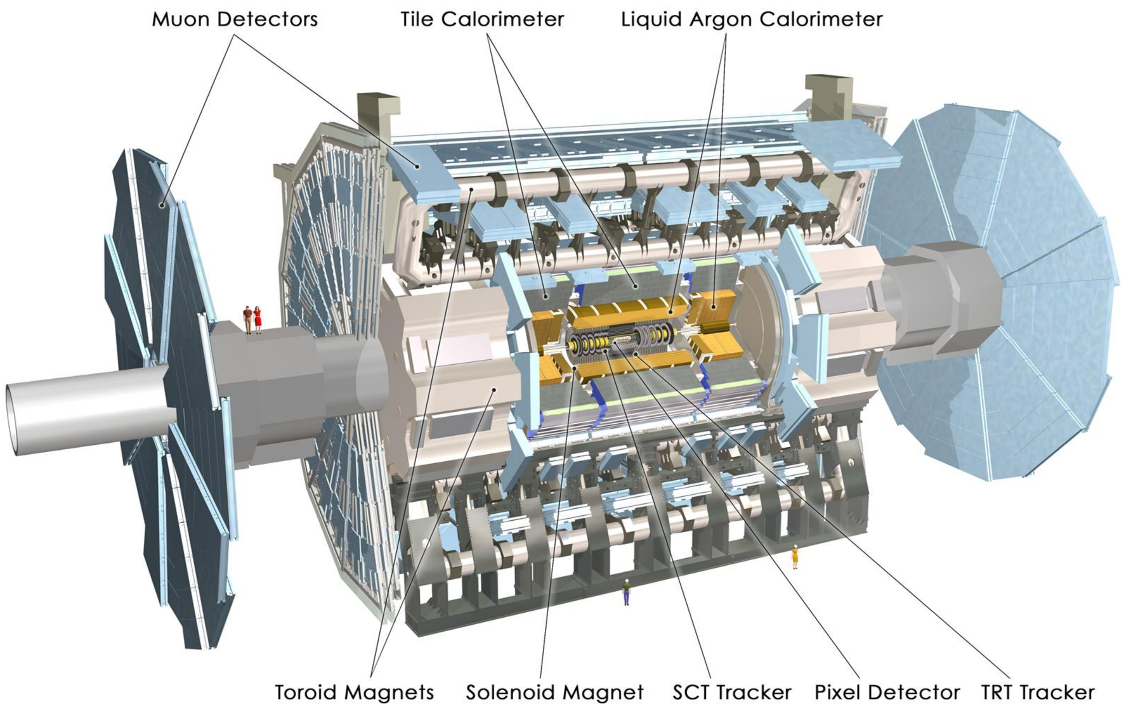

Figure 2.1: A labelled computer generated image of a cross section of the ATLAS detector. [26]

Chapter 3

Monte Carlo Simulations

3.1 Introduction to Monte Carlo

Monte Carlo simulations (MC) have applications in many fields; at their most fundamental level they predict a range of possible outcomes and the probability that each will occur. MC can be used to simulate the collisions of protons occurring inside a virtual ATLAS detector, creating theoretical data in the same format as actual data. Since the events of the two data sets are reconstructed from raw detector output, the same analyses can be applied to both data sets. This allows the validation of theoretical models and predictions with direct comparison of sets. MC generation can be split into two main parts; the generation of the collision events and the simulation of particle interactions with the detector. Each of these complex models is built on theoretical research, empirical relationships and experimental inputs. MC assumes systems are described, and hence can be modelled, by probability density functions; particle interactions are dictated by quantum mechanics and therefore fundamentally probabilistic. It is important to note that the theoretical data created is not the truth of what actually occurs when real protons collide, as it is unknown exactly how nature works. However, the models used have held up against scrutiny of the scientific community and are the current best guess. In addition, not all physics processes are simulated and MC cannot account for all detector effects.

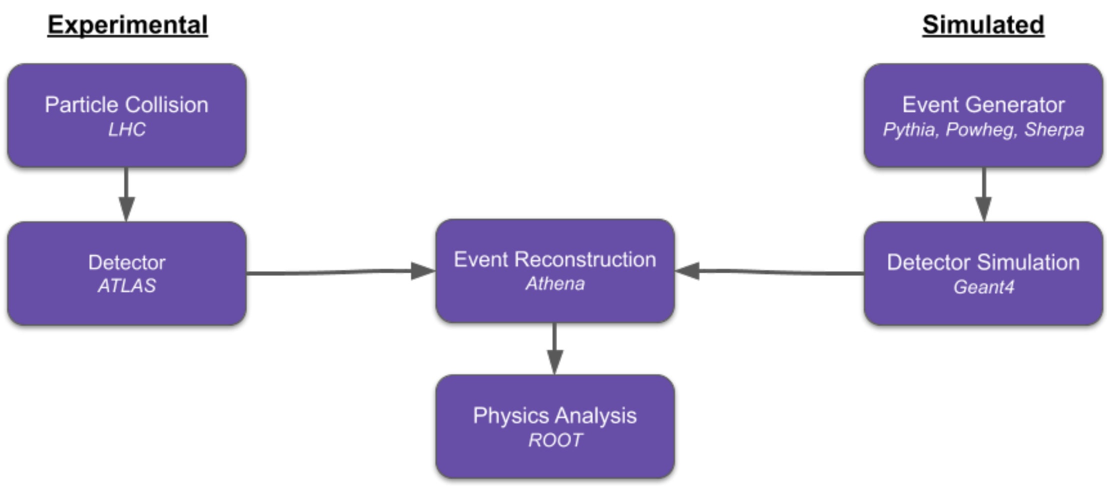

Figure 3.1: Flow chart overview of how MC and real data is processed.

3.2 Event Generator

This stage generates a set of final state particles created from simulated collisions; complete 4-vector information is known about every final state particle, along with all parent-daughter relationships. Simulation of events involves many steps including random number generation, parton showering and hadronisation. With several ways to perform each, uncertainties are introduced into theoretical modelling. Different models are implemented by various MC programs. MC programs used at this stage are; Pythia 8 [27], Powheg [28] and Sherpa [29]. Often, physical processes are simulated with multiple programs and the difference in predictions are assigned as systematic uncertainty. It is important to briefly discuss how each program generates event data; Pythia and Powheg generate events to next-to-leading order (NLO) while Sherpa generates them to next-to-next-to-leading order (NNLO). Pythia and Powheg generate 'inclusive' MC that contains many events at lower transverse momentum (p_T), yet, very few events at higher p_T. Sherpa generates 'slices' of MC split into bins of transverse momentum and jet flavour (see figure 3.2). For the analyses performed in chapter 4 both sets of MC sets were available.

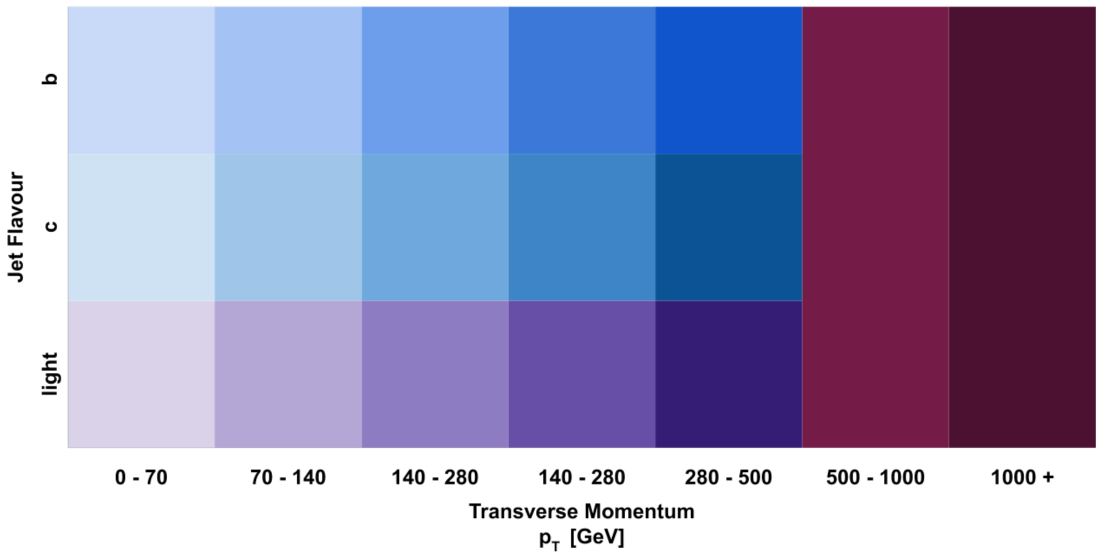

Figure 3.2: Diagram exhibiting breakdown of the slices of 'sliced' MC, each MC slice contains events generated with Sherpa. Slices separated into bins of transverse momentum p_t and the three flavours of identified jet; b, c and light.

3.3 Detector Simulation

The previous stage's simulated events assume protons collide in empty space without interaction with the detector or other apparatus. The MC program Geant4 [30] uses the 4-vector output from the event generator to simulate the interactions of particles with all the major components and materials of the detector. This includes; simulations of particle ionisation in the tracker, energy deposits in the calorimeters, curvature of charged particles in magnetic fields and muons in the muon chamber. In addition, particle interactions with the support structure, cooling pipes, the 3000 km [6] of wiring and known imperfections in detector components must also be accounted for. The final output is simulated raw data in the same format as the real output of the detector.

3.4 Event Reconstruction

The electronic signals from the detector are reconstructed into tracks and calorimeter deposits, which are then reconstructed into leptons, jets, photons and missing transverse momentum. The spatial positions of proton-proton collisions are known as vertices. Particles and jets created in these collisions interact with the detector along their trajectories, the points of interaction are called 'hits'. The tracks of particles can be reconstructed from the hits inside the detector with algorithms designed to connect them. Each proton bunch-crossing has multiple collisions and the locations of all vertices are simultaneously identified using the tracks as input to a three-dimensional imaging software. [31] However, this software occasionally incorrectly reconstructs vertices due the finite time resolution of the detector; hits from particles generated in previous and following bunch crossing can be mistaken for the current bunch crossing. [32] False tracks may be reconstructed that lead to inaccurate or non-existent vertices. The same algorithms are used to reconstruct both real and MC data. Each type of particle has a unique signature that allows them to be distinguished.

Particle	Signature
Photon	Energy deposit in the EM calorimeter without an inner detector track.
Electron	Energy deposit in the EM calorimeter with an inner detector track.
Tau, Jet	Energy deposit in the Hadronic calorimeter with many inner dectector tracks.
Neutral Hadron	Energy deposit in the Hadronic calorimeter without an inner detector track
Charged Hadron	Energy deposit in the Hadronic calorimeter with an inner detector track.
Muon	Inner detector and muon spectrometer track.
Neutrino	Missing transverse momentum

Chapter 4

Methodology

4.1 Analysis Techniques

All of the analyses were performed with Python in conjunction with ROOT. [33] ROOT is an object-oriented framework aimed at solving data-analysis challenges encountered in high energy particle physics. Essentially, ROOT performs the computationally intensive tasks and provides much of the general functionality used in all analyses. Python is used as an interface for ROOT, calling upon ROOT functions. Python is also used to define the specific selection conditions and useful functions in a more readable format. The object-oriented nature of the code allows re-usability throughout; leading to the creation of a pipeline structure that reads in a set of data and MC, performs a defined analysis and returns results. This is critical in creating an educational tool where the user interface must assume no knowledge of the the inner workings of the code. Figure 4.1 shows the overall logic of the code used to perform analyses.

4.2 Luminosity Weighting of Data and MC

To compare theoretical and real data, the luminosity of both sets must match. However, the luminosity of LHC is dependent on many factors and data collection runs typically have different values. Since generating MC events is computationally expensive and time consuming, it is logical to create it independently of luminosity. This allows the MC luminosity weightings to be changed after their creation in order to match real data. This re-weighting was performed in the plotting stage of analyses and was a common fix for disagreements between MC and real data. All analyses, except boosted jet, were performed on the ATLAS Open Data set 'STMD4'. The luminosity of this set was $10fb^{-1}$. Boosted analysis was performed on the ATLAS Open Data set '1fatjet1lep'; the luminosity of this set was $401.3fb^{-1}$.

4.3 Event Selection

Removing background signal and maximising the purity of the data is crucial to obtain high quality analyses. Event selection can be broken up into two key phases; general selection and specific selection. The general selection cuts eliminate events well known to be background in all analyses, unlike specific selection cuts that require careful consideration to isolate particles of interest in the remaining events. It is important to note that the 13 TeV Open Data set has been made with a pre-selection determined by the 'Good run list'. [34]

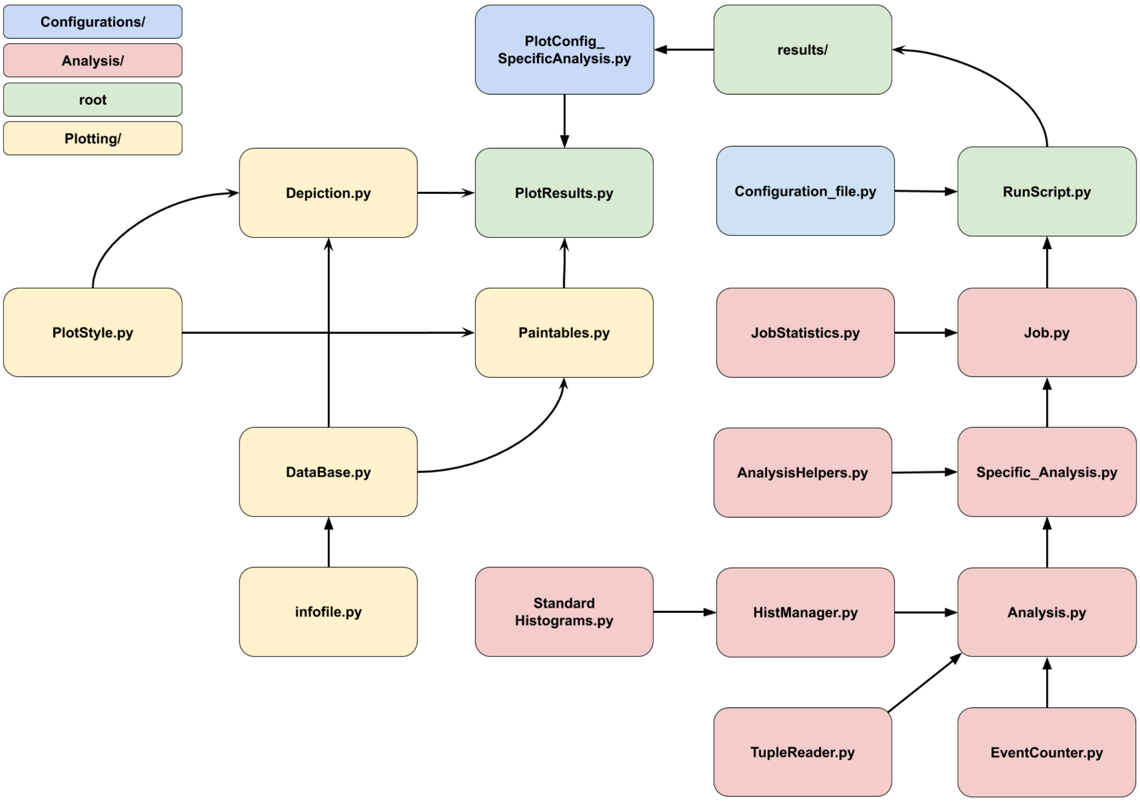

Figure 4.1: A diagram showing the basic overall logic of code used. Arrows indicate the file being calling into another.

General Selection	Requirements
Must have triggered	an electron or a muon or tau or di tau or photon.
Primary vertex cuts	Number of tracks $N_{\text{tracks}} > 2$

4.3.1 $\mathbf{Z^0}$ and $\mathbf{W^\pm}$

The measured masses of the Z^0 and $W^\pm$ bosons are roughly 100 times greater than the mass of the proton. [17] As consequence, these bosons are often nearly at rest when they are formed from colliding protons, as most of the available energy is converted into mass. If the bosons are nearly at rest in the lab frame and they are expected to produced leptons and neutrinos with have high transverse momentum.
The Z boson decay considered in this analysis is into two leptons 'back-to-back' in its rest frame. The specific selection requirements for this Z analysis searches for exactly two oppositely-charged leptons of the same flavour with high transverse momentum. Electron and muon particle-antiparticle pairs are easily identifiable, however, the tau leptons are not due to their very short half lives ($(2.903 \pm 0.005) \times 10^{13}$ s [17]) and large range of subsequent decay products. The Z boson has a cross section of $\sigma_Z^{tot} = 1.981 \pm 0.087$ [35]. This is a relatively high value for cross section therefore Z boson events are very common, see Figure 4.3 for comparison to other processes. a simple way to increase the purity of the data is to exclude the tau when searching for Z decaying to leptons-antilepton pairs.
The $W^\pm$ boson decay considered in this analysis is to a lepton-neutrino pair almost instantly (half life of 3×10^{25} s [17]) after the boson's production. The neutrino does not interact with the detector, so its existence must be inferred from missing transverse momentum. The specific selection for $W^\pm$ analysis performed here searches for strictly one lepton with corresponding missing transverse momentum. As done so for the Z, to further

increase purity, tau flavour lepton decays are excluded as they decay before reaching the detector.

Specific Selection	Requirement
Z Analysis	2 oppositely charged leptons of the same flavour. Transverse momentum of each lepton: $p_T > 25$ GeV.
W Analysis	Exactly 1 lepton. Transverse momentum of lepton: $p_T > 25$ GeV. Missing transverse momentum: $p_{T,miss} > 30$ GeV.

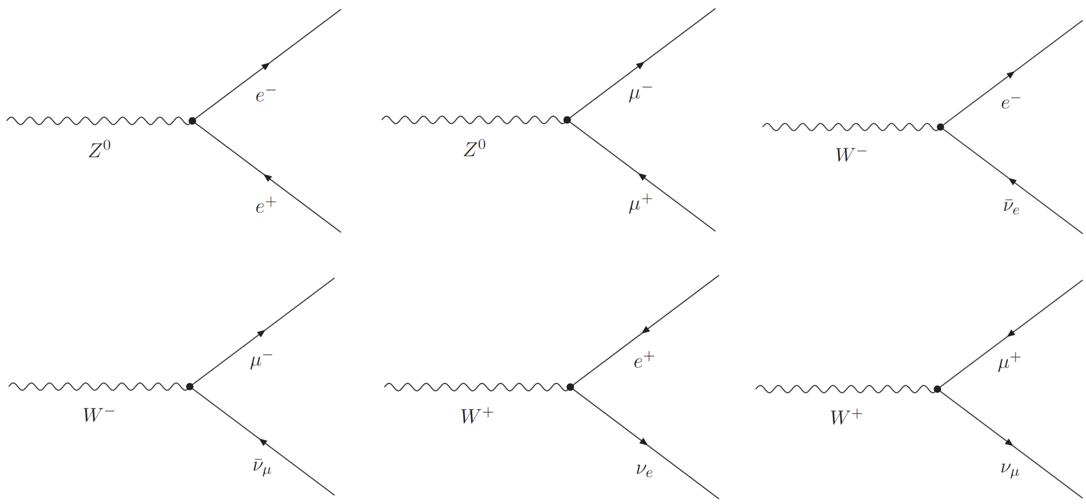

Figure 4.2: Feynman diagrams showing the signature decays of Z^0 and $W^\pm$ bosons selected in each corresponding analysis. [36]

4.3.2 ZZ

In ZZ analysis events with leptonic decay of both Z bosons are selected. The specific selection requirements are for two pairs of leptons, each pair having the same flavour and with at least one lepton having a relatively high transverse momentum.

Specific Selection	Requirement
ZZ Analysis	2 pairs of oppositely-charged leptons. Each pair of leptons must have the same flavour. Transverse momentum of at least one lepton: $p_T > 25$ GeV.

4.3.3 $t\bar{t}$

The $t\bar{t}$ analysis performed only selects semi-leptonic decay events. Semi-leptonic refers to exactly one of the $W^\pm$ bosons produced decaying into a lepton and a neutrino pair. In addition two jets from the other $W^\pm$ decay and two b-tagged jets from the bottom quarks are required.

Specific Selection	Requirement
$t\bar{t}$ Analysis	Exactly 1 lepton with $p_T > 25$ GeV. Missing transverse momentum: $p_{T,miss} > 30$ GeV. At least 4 jets with $p_T > 25$ GeV and $\|\eta\| > 2.5$ At least 2 of these jets must be b-tagged

4.3.4 Boosted Jet Analysis

In boosted jet analysis events with at least one lepton and one 'fat' jet are selected. These jet are highly boosted relative to the lab frame and therefore expected to have a very high energy deposit in the hadron calorimeter.

Specific Selection	Requirement
Boosted Analysis	At least one lepton with $p_T > 25$ GeV At least one fat jet with $p_T > 250$ GeV

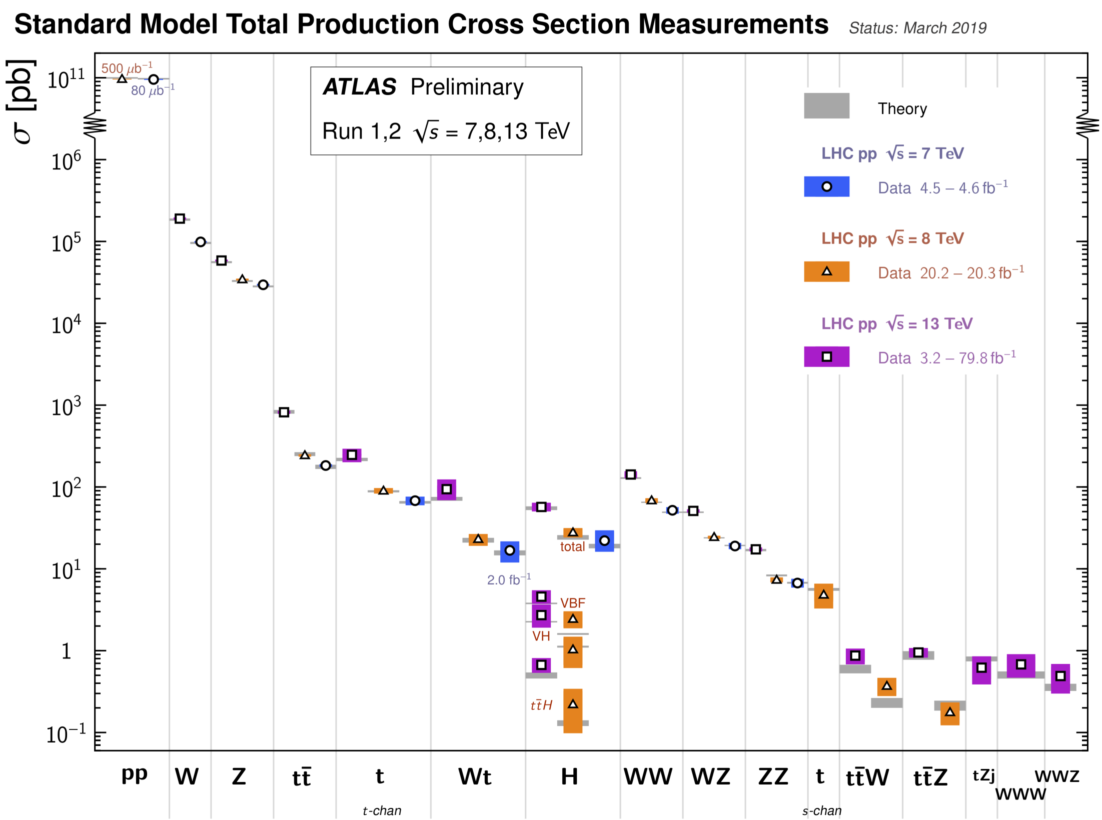

Figure 4.3: Graph of the comparison of SM total production cross section measurements of various processes against their theoretical values. [37]

Chapter 5

Results

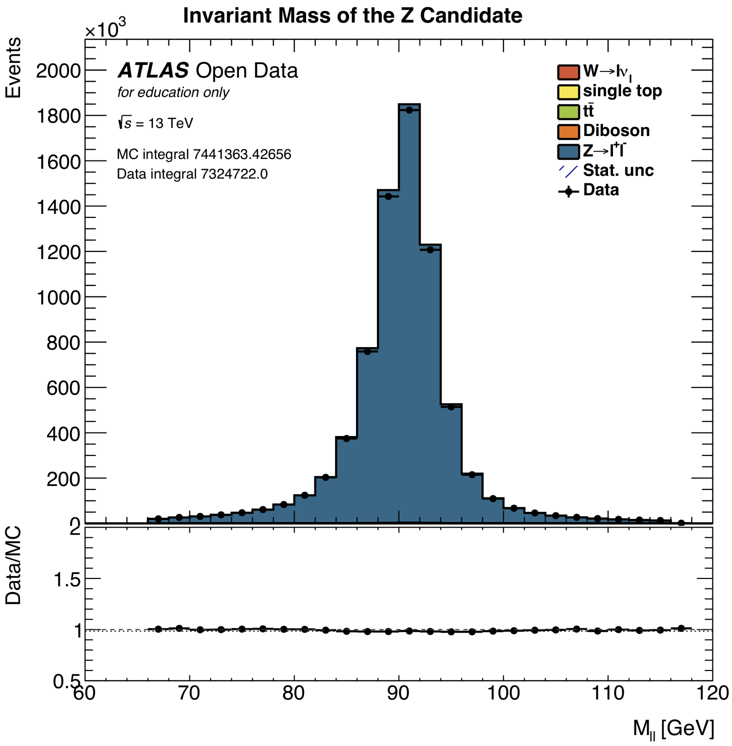

Figure 5.1: Histogram of number of events against invariant mass of the Z candidate. The histogram shows that the vast majority of data events passing the Z analysis selection requirements are formed of $\mathrm{Z}^0 \rightarrow \mathrm{e}^-\mathrm{e}^+$ and $\mathrm{Z}^0 \rightarrow \mu^-\mu^+$ processes.

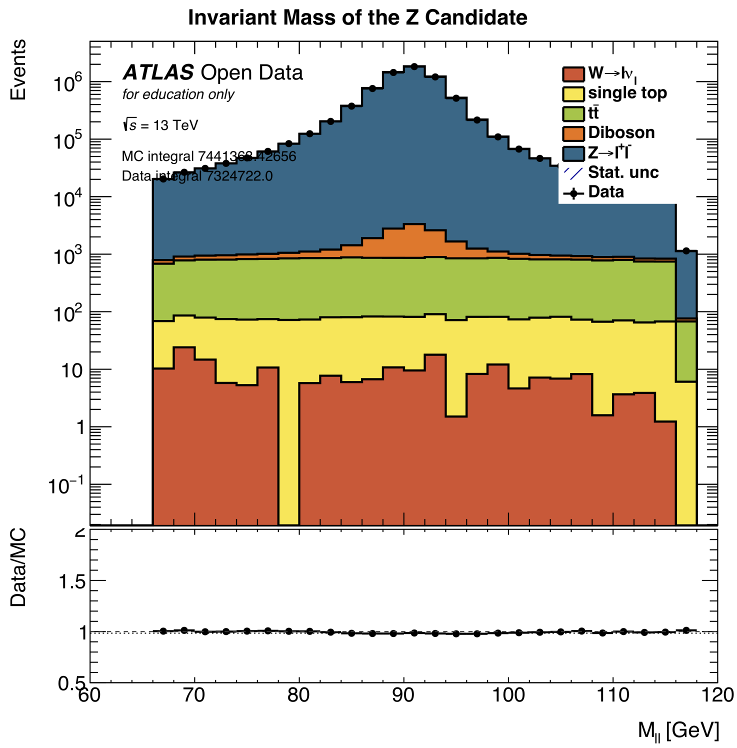

Figure 5.2: Logged scale Y axis version of figure 5.1 to highlight other contributing processes. In this case the processes $ZZ \rightarrow \ell^-\ell^+\ell^-\ell^+$, $t\bar{t} \rightarrow W^+W^-b\bar{b}$, $t \rightarrow W\ b$ and $W^{\pm} \rightarrow \ell\nu$.

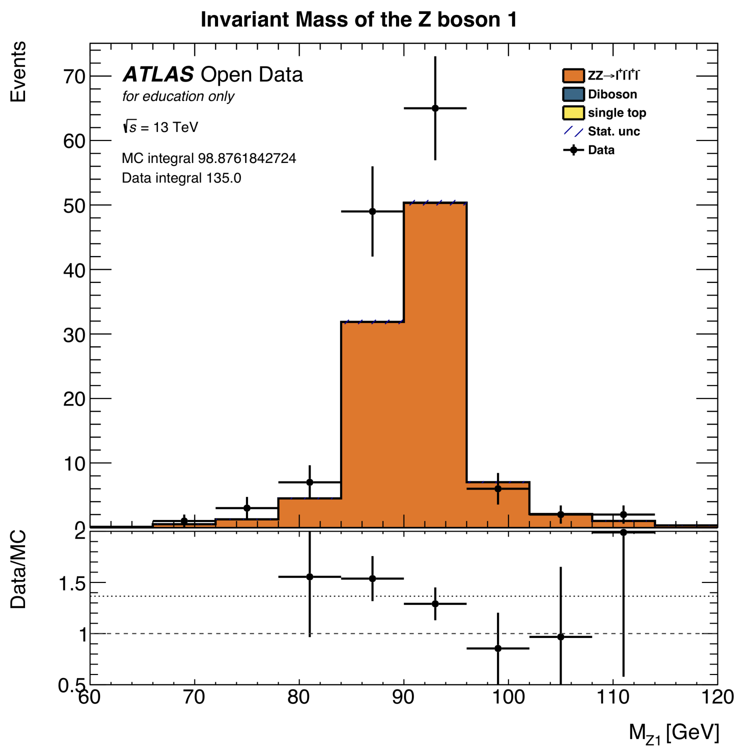

Figure 5.3: Histogram showing the number of events against the invariant mass of Z boson candidate number 1. The histogram shows the majority of data events passing the ZZ analysis selection requirements are $ZZ \rightarrow \ell^-\ell^+\ell^-\ell^+$.

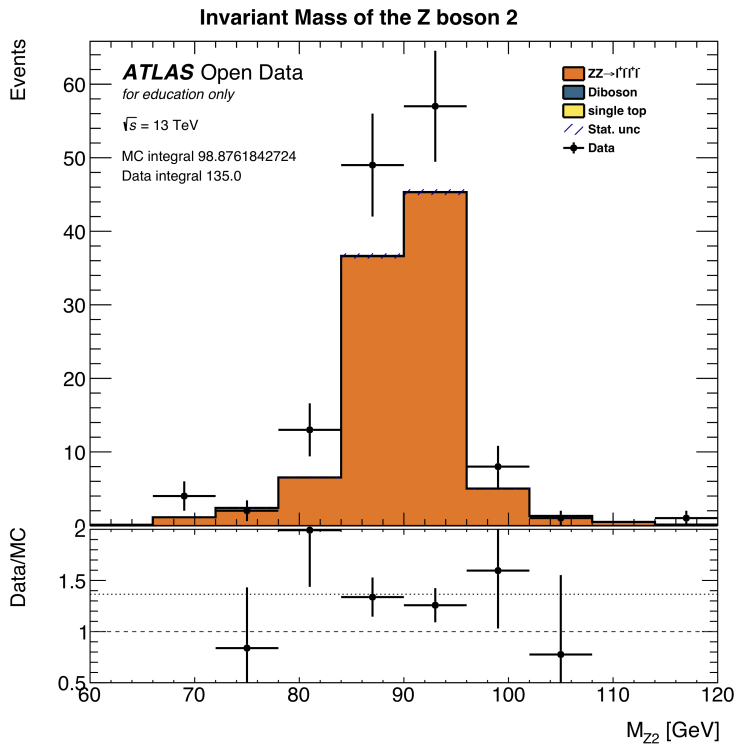

Figure 5.4: Histogram showing the number of events against the invariant mass of Z boson candidate number 2. The histogram shows the majority of data events passing the ZZ analysis selection requirements are ZZ $\rightarrow \ell^-\ell^+\ell^-\ell^+$.

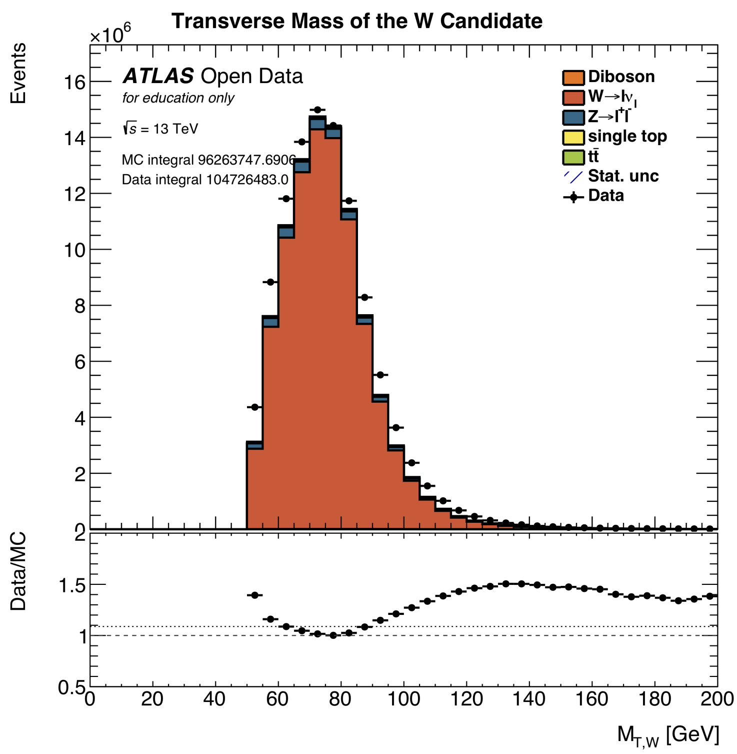

Figure 5.5: Histogram showing the number of event against the transverse mass of the W candidate. The majority of data events passing the W analysis are $W^{\pm} \rightarrow \ell\nu$ with background processes $Z^0 \rightarrow \ell^+\ell^-$, $t\bar{t} \rightarrow W^+W^- b\bar{b}$ and various diboson.

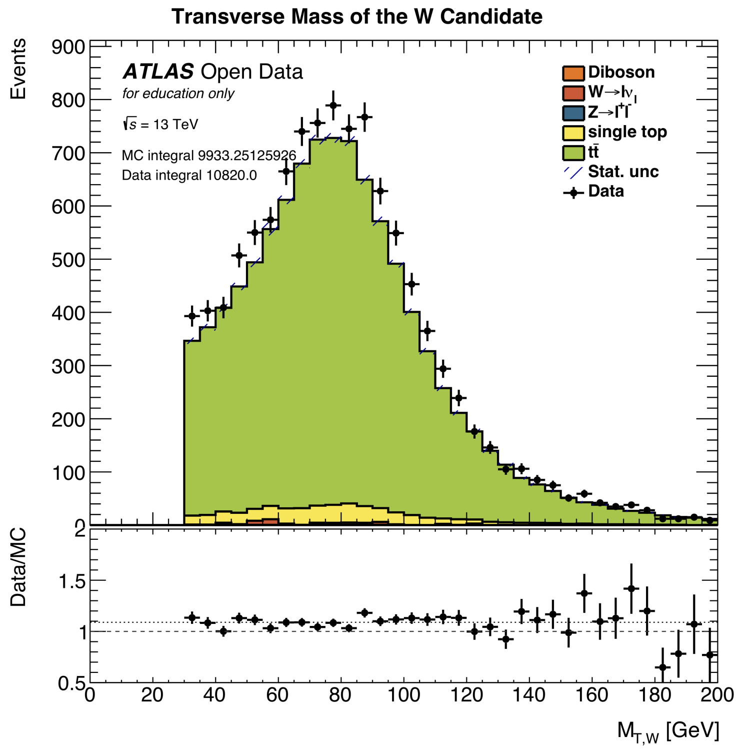

Figure 5.6: Histogram showing the number of events against transverse mass of the W candidate. The histogram shows the majority of data event passing the $t\bar{t}$ analysis are $t\bar{t} \rightarrow W^+W^- b\bar{b}$ process with contribution from $t \rightarrow W\ b$. Background includes $Z^0 \rightarrow \ell^+\ell^-$ and diboson processes.

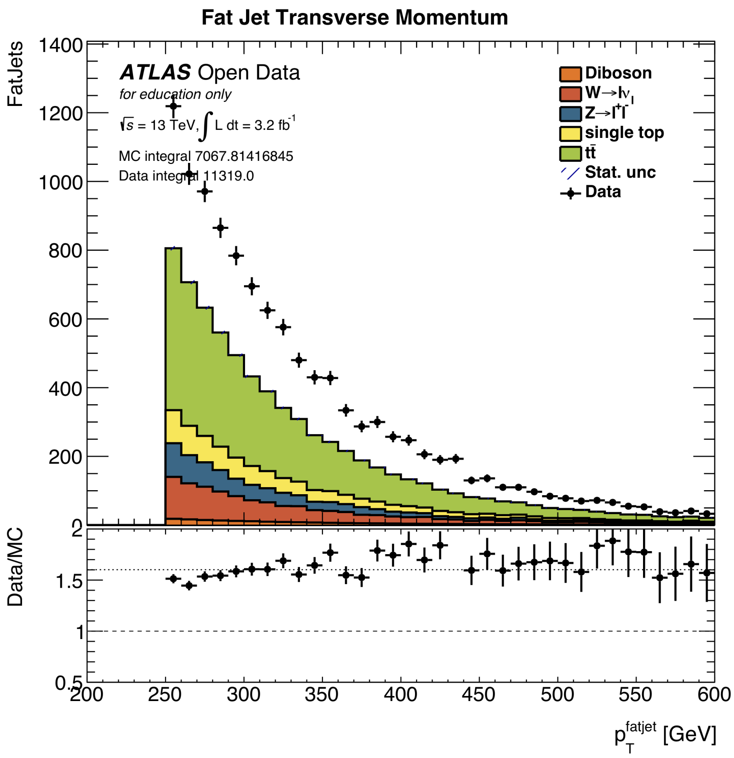

Figure 5.7: Histogram showing the number of events against the transverse momentum of fat jet. The inclusive event data passing fat analysis selection requirements are formed of $t\bar{t} \rightarrow W^+W^- b\bar{b}$, $W^\pm \rightarrow \ell\nu$, $t \rightarrow W\ b$ and $Z^0 \rightarrow \ell^+\ell^-$. Also include some background formed of diboson processes.

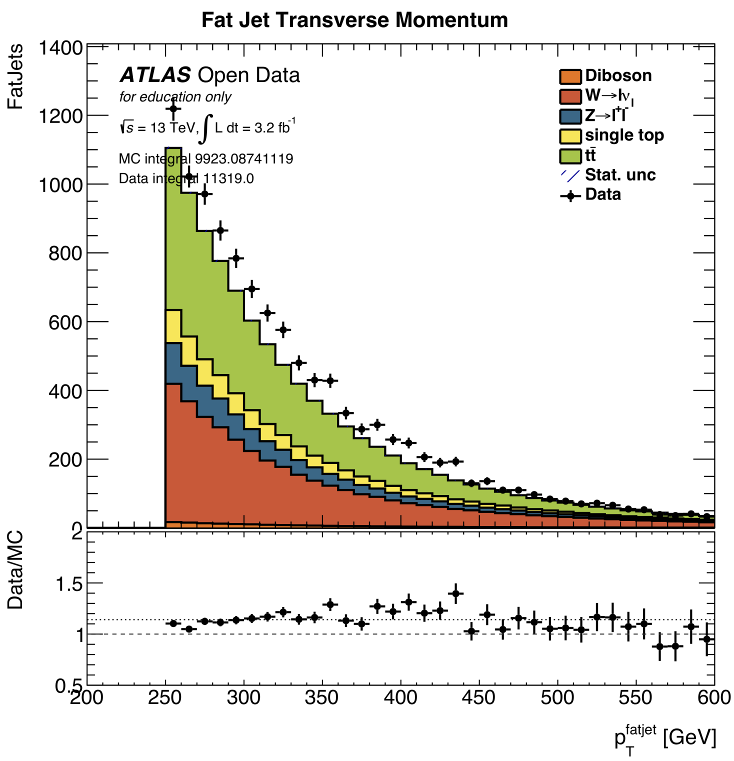

Figure 5.8: Histogram similar to figure 5.7, fat analysis performed on sliced MC (see section 3.2)

Chapter 6

Discussion

6.1 Results Discussion

The measured mass of the Z boson is 91.1876 ± 0.0021 GeV. [17] The histogram in Figure 5.1 was obtained from the Z analysis performed on the 13 TeV Open Data set. The distribution of invariant mass can be seen to form a sharp peak roughly around the Z measured mass. In addition, the flat line representing the ratio of real data to MC suggests a strong agreement between the sets. This indicates that the physical theory used to build the MC data set accurately describes the reality of real Z decays. The invariant Z mass with a logarithmic scale in Figure 5.2 highlights other processes with small contributions. The second largest contribution after the decay of a single Z into a lepton pair, $Z \rightarrow \ell^+\ell^-$, is the decay of ZZ into two lepton pairs, $ZZ \rightarrow \ell^+\ell^-\ell^+\ell^-$. Other smaller contributing processes, such as W bosons decaying leptonically, are seen as minor background. The minimal background suggests the analysis performed almost entirely isolated the desired Z boson decay events and therefore demonstrates itself as high purity.

Results for the ZZ diboson analysis performed on the 13 TeV open data set, presented in Figure 5.3 and in Figure 5.4 show the invariant masses of both Z bosons created in selected events. From observation it can be seen that the MC does not account for all of the ZZ decays seen in real data, however, the shape of the MC histograms matches that of the real data. The full ATLAS ZZ analysis [38] includes contributions to ZZ production initiated by gluon interaction. Whereas the ATLAS Open Data analysis only accounts for this proton-proton interaction contribution. This may be responsible for the discrepancy between real and MC data sets. The total number of ZZ events is much smaller than the number of single Z events. This is to be expected as diboson ZZ production in collisions is much rarer than Z production due to it's higher mass. Although there is not total agreement between the MC and data, a relatively sharp peak in the invariant mass distribution is seen around the accepted mass of the Z boson for both candidate particles. This strongly indicates the ZZ analysis is successfully isolating desirable events.

Figure 5.6 was obtained from performing $t\bar{t}$ analysis, it shows the transverse mass distribution of the W candidate. This can be seen to peak around the mass of the W boson. The ratio of real data to MC is relatively flat so may be correct after the systematic uncertainy on the MC is considered.

Results of the fat jet analysis performed on the 13 TeV Open Data set are presented in 5.7 and 5.8, both show the distribution of number of events against fat jet transverse momentum, but, are made with different sets of MC. Figure 5.7 was made by running analysis on inclusive MC data; inclusive containing events generated to NLO. Figure 5.8 was made by running analysis on sliced MC data; sliced containing events generated to

NNLO. It is important to note that not all sliced MC processes were available at the time of plotting and as a result there is a slight disagreement between MC and real data sets in Figure 5.8. With direct comparison it is shown that the sliced MC is a much better description of the data than inclusive MC. A possible reason for this is that NNLO, implemented in the sliced MC, provides a more accurate representation of the physical processes occurring in these high energy jets. This conclusion is important as it demonstrates that inclusive MC data is not suitable for the boosted jet analysis.

Figure 5.5 illustrates the transverse mass of the W candidate and was obtained by performing the W analysis on the 13 TeV Open Data set. The accepted mass of the $W^{\pm}$ is 80.379 ± 0.012 GeV [17]. From inspection of Figure 5.5 the mass distribution peaks around the accepted mass of the W boson. The MC integral value is: 96263748 and the data integral value is: 104726483. The difference in these values and observation of the histogram shows there is less MC than data being plotted. Initially this was thought to be incorrect weightings of the MC data, however, after much cross referencing of weights it does not appear to be the cause of disagreement. The ratio of real data to MC suggests higher p_T are not modelled correctly. A possible reason is that inclusive MC is not a perfect descriptor of the physical processes occurring in the real data.

6.2 Further Progression

The W analysis took a very long time to run over the entire data set. With more time, the analysis would be run on the sliced MC to test if this is the cause of discrepancy between MC and real data sets. At the time of plotting, some of the MC slices for the '1fatjet1lep' (see section 4.2) MC data set were not available. These slices should be added to the boosted jet analysis to ensure this is responsible for the remaining disagreement between real and MC data sets. Additional analyses such as Higgs decay to ZZ, Higgs decay to WZ and boosted Z would be performed next.

I could have improved the efficiency of performing analyses by utilising the 'screen' function in the remote linux server terminal earlier in my project. Often, analyses that would require upwards of 2 hours would fail to run due to my computer being disconnected from the remote server. The screen command allowed analyses to continue to run even when disconnected. Another performance improvement would have been to make use of the parallel computing option earlier in the project, allowing multiple stages of analyses to run at the same time.

Chapter 7

Conclusion

This project has made a contribution to the development of the ATLAS Open Data educational tool aiming to release a 13 TeV dataset. Z boson, $W^{\pm}$ boson, ZZ, $t\bar{t}$ and boosted jet analyses have been conducted and on this data set. The MC provided by the Open Data project has no systematic uncertainties assigned to them, as a result only visual comparison can be made. It has been shown that inclusive Monte Carlo data is unsuitable to model the boosted jet analysis and possibly the W analysis. $t\bar{t}$ and ZZ boson analyses have been validated and with minor adjustments is ready to be used in the educational tool. The Z boson analysis is ready to be implemented now.

Acknowledgements

I would like to take this opportunity to express my immense gratitude to PhD student Meirin Evans; for all the time, support and help he has given to me throughout the course of this dissertation. He is an exemplary teacher and his genuine passion for education is inspiring.

Bibliography

[1] Open Data Collaboration. *The Open Data Project*. URL: http://opendata.atlas.cern/.

[2] CERN. *The Standard Model*. URL: https://home.cern/science/physics/standard-model.

[3] CERN. *Large Hadron Collider*. URL: https://home.cern/science/accelerators/large-hadron-collider.

[4] B. R. Martin and G. Shaw. *Particle Physics. Third Edition*. Wiley, 2008.

[5] ATLAS Collaboration. *The ATLAS experiment*. URL: https://atlas.cern/.

[6] ATLAS Collaboration. *ATLAS fact sheet*. URL: https://cds.cern.ch/record/1457044/files/.

[7] Frank Wilczek. "QCD made simple". In: *Phys. Today* 53N8 (2000), pp. 22–28. DOI: 10.1063/1.1310117.

[8] Kurt Riesselmann. *Computing the quarks*. 2005. URL: https://www.symmetrymagazine.org/article/november-2005/computing-the-quarks.

[9] *Jet mass reconstruction with the ATLAS Detector in early Run 2 data*. Tech. rep. ATLAS-CONF-2016-035. Geneva: CERN, July 2016. URL: https://cds.cern.ch/record/2200211.

[10] Werner Herr and B Muratori. "Concept of luminosity". In: (2006). URL: https://cds.cern.ch/record/941318.

[11] CERN writing guidelines. *Inverse Femtobarn*. URL: https://writing-guidelines.web.cern.ch/entries/inverse-femtobarn.

[12] Andrée Robichaud-Véronneau and. "Searches for Supersymmetry and Exotics phenomena with the ATLAS detector". In: *Journal of Physics: Conference Series* 455 (Aug. 2013), p. 012012. DOI: 10.1088/1742-6596/455/1/012012. URL: https://doi.org/10.1088%5C%2F1742-6596%5C%2F455%5C%2F1%5C%2F012012.

[13] The ATLAS Collaboration. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08003–S08003. DOI: 10.1088/1748-0221/3/08/s08003. URL: https://doi.org/10.1088%5C%2F1748-0221%5C%2F3%5C%2F08%5C%2Fs08003.

[14] John W Harris. "Introduction to Hard Scattering Processes and Recent Results from Hard Probes at RHIC and LHC". In: *Journal of Physics: Conference Series* 630 (July 2015), p. 012052. DOI: 10.1088/1742-6596/630/1/012052. URL: https://doi.org/10.1088%5C%2F1742-6596%5C%2F630%5C%2F1%5C%2F012052.

[15] M. zur Nedden. "The LHC Run 2 ATLAS trigger system: design, performance and plans". In: *Journal of Instrumentation* 12.03 (Mar. 2017), pp. C03024–C03024. DOI: 10.1088/1748-0221/12/03/c03024. URL: https://doi.org/10.1088%2F1748-0221%2F12%2F03%2Fc03024.

[16] V.A. Novikov et al. “Theory of Z boson decays”. In: *Reports on Progress in Physics* 62 (June 1999). DOI: 10.1088/0034-4885/62/9/201.

[17] Particle Data Group. “Review of Particle Physics”. In: *Phys. Rev. D* 98 (3 Aug. 2018), p. 030001. DOI: 10.1103/PhysRevD.98.030001. URL: https://link.aps.org/doi/10.1103/PhysRevD.98.030001.

[18] I R Kenyon. “The Drell-Yan process”. In: *Reports on Progress in Physics* 45.11 (Nov. 1982), pp. 1261–1315. DOI: 10.1088/0034-4885/45/11/002. URL: https://doi.org/10.1088%2F0034-4885%2F45%2F11%2F002.

[19] ATLAS Open Data Collaboration. *ATLAS open data documentation*. URL: http://opendata.atlas.cern/books/current/get-started/_book/.

[20] Morad Aaboud et al. “Measurement of inclusive and differential cross sections in the $H \to ZZ^* \to 4\ell$ decay channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: *JHEP* 10 (2017), p. 132. DOI: 10.1007/JHEP10(2017)132. arXiv: 1708.02810 [hep-ex].

[21] The ATLAS collaboration. “Measurement of gluon fusion and vector boson fusion Higgs boson production cross-sections in the $H \to WW^* \to e\nu\mu\nu$ decay channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: (2018).

[22] Chen Gang and Liu Lianshou. “A Monte Carlo study of the quark- and gluon-jets of different hardness in three-jet events identified by b-tag method”. In: *Journal of Physics G: Nuclear and Particle Physics* 30.10 (Aug. 2004), pp. 1399–1406. DOI: 10.1088/0954-3899/30/10/007. URL: https://doi.org/10.1088%2F0954-3899%2F30%2F10%2F007.

[23] *Measurement of Jet Mass and Substructure for Inclusive Jets in s = 7 TeV pp Collisions with the ATLAS Experiment*. Tech. rep. ATLAS-CONF-2011-073. Geneva: CERN, May 2011. URL: http://cds.cern.ch/record/1352454.

[24] ATLAS. *Detector Technology*. URL: https://atlas.cern/discover/detector.

[25] Morad Aaboud et al. “Performance of the ATLAS Trigger System in 2015”. In: *Eur. Phys. J.* C77.5 (2017), p. 317. DOI: 10.1140/epjc/s10052-017-4852-3. arXiv: 1611.09661 [hep-ex].

[26] ATLAS Collaboration. *Labelled image of the ATLAS detector*. 2008. URL: http://cdsweb.cern.ch/record/1095924.

[27] PYTHIA 8. *PYTHIA 8 MC event generator*. URL: http://home.thep.lu.se/~torbjorn/pythia81php/Welcome.php.

[28] POWHEG. *POWHEG MC event generator*. URL: http://powhegbox.mib.infn.it/.

[29] Sherpa. *Sherpa MC event generator*. URL: https://sherpa.hepforge.org/doc/SHERPA-MC-2.1.1.html.

[30] CERN. *Geant4 Detector simulation MC Program*. URL: https://geant4.web.cern.ch/.

[31] G Borissov et al. “ATLAS strategy for primary vertex reconstruction during Run-2 of the LHC”. In: *Journal of Physics: Conference Series* 664.7 (Dec. 2015), p. 072041. DOI: 10.1088/1742-6596/664/7/072041. URL: https://doi.org/10.1088%5C%2F1742-6596%5C%2F664%5C%2F7%5C%2F072041.

[32] The CMS Collaboration. “Description and performance of track and primary-vertex reconstruction with the CMS tracker”. In: *Journal of Instrumentation* 9.10 (Oct. 2014), P10009–P10009. DOI: 10.1088/1748-0221/9/10/p10009. URL: https://doi.org/10.1088%5C%2F1748-0221%5C%2F9%5C%2F10%5C%2Fp10009.

[33] CERN. *ROOT data analysis framework*. URL: https://root.cern.ch/.

[34] T Golling et al. "The ATLAS data quality defect database system". In: *European Physical Journal C* 72 (Oct. 2011). DOI: 10.1140/epjc/s10052-012-1960-y.

[35] "Measurement of W and Z-boson production cross sections in pp collisions at s=13 TeV with the ATLAS detector". In: *Physics Letters B* 759 (2016), pp. 601–621. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2016.06.023. URL: http://www.sciencedirect.com/science/article/pii/S0370269316302763.

[36] ATLAS Collaboration. *Feynman diagrams of W decays*. 2015. URL: https://atlas.physicsmasterclasses.org/wpath_files/img/Feynman/.

[37] ATLAS Open Data Collaboration. *ATLAS Collaboration*. URL: https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2019-010/fig_01.pdf.

[38] Georges Aad et al. "Measurement of the ZZ Production Cross Section in pp Collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector". In: *Phys. Rev. Lett.* 116.10 (2016), p. 101801. DOI: 10.1103/PhysRevLett.116.101801. arXiv: 1512.05314 [hep-ex].