Analysing Z-boson Events With the ATLAS Experiment
Third Year Lab Report

Tommy C. Smith[1]
(Experiment performed in collaboration with Yii Hanks Tan)
[1] *Department of Physics and Astronomy, University of Manchester*
(Dated: April 6, 2025)

The main goal of this investigation was to measure the cross-section for Z-boson production during proton-proton collision. The cross-section was determined to be $\sigma = 2.007 \pm 0.001\,(\text{stat.})\,^{+0.001}_{-0.002}\,(\text{syst.}) \pm 0.034\,(\text{lumi.})$, and $\sigma = 1.970 \pm 0.001\,(\text{stat.})\,^{+0.001}_{-0.004}\,(\text{syst.}) \pm 0.033\,(\text{lumi.})$ for $Z \to e^+e^-$ and $Z \to \mu^+\mu^-$ decay channels respectively. Results were obtained using Monte Carlo (MC) simulations and optimized selection criteria. Disagreement between decay channels suggest inaccuracies in underlying assumptions, providing insights for improving future analyses. This report outlines the analytical methodology and discusses the results and their implications for understanding physical processes.

1. INTRODUCTION

1.1. *ATLAS and the Z-boson*

This investigation utilizes open data provided by the ATLAS experiment for proton-proton collisions at a centre-of-mass energy of 13TeV. ATLAS (A Toroidal LHC Apparatus) is the largest of the four major experiments operating at CERN's large hadron collider (LHC). It is a general-purpose detector designed to explore a wide range of phenomena, including precision measurements of the Standard Model. Monte Carlo (MC) simulations were used in comparison with ATLAS data, forming a cornerstone of this analysis.

Our focus is on the Z-boson. Discovered in 1983 at CERN by the UA1 and UA2 experiments[1], the Z-boson is a neutral mediator of the electroweak interaction. Its discovery confirmed the theoretical framework of electroweak unification proposed by Glashow, Salam, and Weinberg, who were later awarded the Nobel Prize for this work. [2]

The Z-boson is produced in proton-proton collisions when a quark and antiquark annihilate. While protons do not contain valence antiquarks, antiquarks are present in the form of sea quarks which originate from gluons splitting into virtual quark-antiquark pairs. Once produced, the Z-boson decays almost immediately, predominantly into a lepton-antilepton pair, such as e^+e^- or $\mu^+\mu^-$, providing a clean and identifiable signature for analysis.

Studying the Z-boson remains essential, and precise measurements of its production cross-section and decay channels help refine our understanding of electroweak interactions and can reveal potential deviations indicative of new physics.

1.2. *Cross-secton*

The cross-section, σ, represents the effective target area of a process and is directly related to the probability of its occurrence. It was broadly calculated via the following formula:

$$\sigma = \frac{N_{\text{selected}} - N_{\text{background}}}{\epsilon \cdot \int \mathcal{L}\, dt}, \tag{1}$$

where:

- N_{selected} is the number of observed events after selection cuts.
- $N_{\text{background}}$ is the number of background events.
- ϵ is the selection efficiency.
- $\int \mathcal{L}\, dt = 10.064 \pm 1.7\%\,\text{fb}^{-1}$ is the integrated luminosity of the dataset.[3]

This investigation focuses on optimizing selection cuts to maximize the signal-to-background ratio, mitigate the effects of mismodeling, reduce systematic uncertainties, and improve the overall accuracy of the analysis. For brevity, only the figures for the $Z \to e^+e^-$ results are presented in this report. The $Z \to \mu^+\mu^-$ results have been analyzed similarly and are discussed where relevant.

2. INVARIANT MASS ANALYSIS

The invariant mass represents the rest mass of the parent particle (e.g the Z-boson). It is Lorentz invariant, and can be calculated in any reference frame by measuring the momenta of detected lepton pairs. This method is invaluable for identifying particles via their decay products. The invariant mass was calculated from lepton decay pairs via the equation:

$$m_{\ell\ell} = \sqrt{2p_{T1}p_{T2}\left(\cosh(\eta_1 - \eta_2) - \cos(\phi_1 - \phi_2)\right)} \tag{2}$$

where:

- $m_{\ell\ell}$ is the invariant mass of the lepton pair
- p_T is lepton transverse momentum
- η is pseudorapidity
- ϕ is azimuthal angle
- the subscripts 1,2 refer to the individual leptons in the decay pair

This utilises the Lorentz invariant quantities η and ϕ defined in the transverse plane, which are measured by ATLAS. By plotting an invariant mass distribution for our detected leptons, we can identify the processes responsible for their production.

2.1. *Results*

Figure 1 reveals a clear peak at $\sim 90MeV/c^2$ consistent with the Z-boson rest mass of $M_Z = 91.1876 \pm 0.0021\,\text{GeV}/c^2$[4]. This represents our signal event. It appears to be Poisson distributed with a width characterised by its standard deviation. Background noise due to the various other processes that can create lepton pairs is observed particularly for low invariant mass values. In the $0 - 20MeV/c^2$ range we can see a distinct peak. Focusing our distribution to this range we were able to discern two peaks which we identified as the J/ψ and Υ mesons.

In the invariant mass distribution generated with the MC data however, we saw that the events fell to zero below an invariant mass of $\sim 60GeV/c^2$. This led us to introduce a selection cut which excluded any events with an invariant mass below $60GeV/c^2$. This was

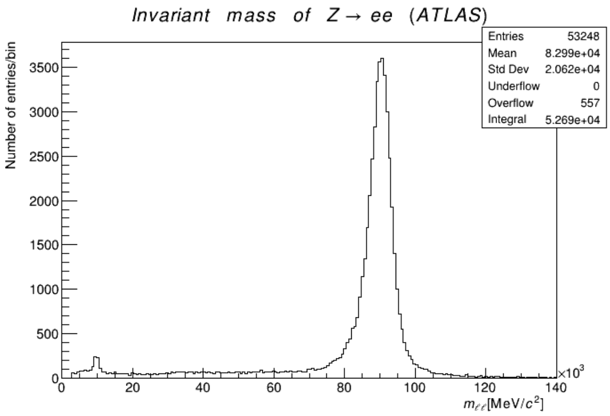

FIG. 1. Invariant mass distribution for electron positron pairs detected by ATLAS. Selection cuts were used to isolate decay pairs of the same flavour.

done to avoid an extra contribution in low invariant mass values which is unsimulated by the MC dataset and to exclude a portion of the background events such as the J/ψ and Υ mesons which may skew our results.

3. p_T^{cone} AND E_T^{cone} ANALYSIS

Another method for reducing background contamination is by considering the variables p_T^{cone} and E_T^{cone}. Whilst leptons produced by W and Z-boson decay are generally isolated, decaying directly into lepton-antilepton pairs, those produced in background processes such as b-quark decay are often embedded within particle jets due to hadronization.

To distinguish isolated signal leptons from those embedded in jets, we can define a cone of constant $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$ [5] centred around the track of the detected lepton. We define the variables p_T^{cone} and E_T^{cone} as the sum of transverse momenta of all tracks within these cones. $\Delta R = 0.3$ is used for p_T^{cone} and $\Delta R = 0.2$ for E_T^{cone} [3]. These two variables essentially quantify the same thing, however p_T^{cone} sums the transverse momenta of charged particles from tracks in the inner detector, while E_T^{cone} includes the transverse energy deposited by both charged and neutral particles in the calorimeter. We would expect isolated leptons to have low p_T^{cone} and E_T^{cone} values and those in jets to have high values.

Varying selection cuts on these variables, which reject events above a threshold, allows us to optimize the signal-to-background ratio and reveals potential biases in our results, making it a powerful tool for quantifying systematic uncertainty. To evaluate the effectiveness of these cuts, we calculated the significance, $\mathcal{S} = \frac{N_{\text{signal}}}{\sqrt{N_{\text{background}}}}$, where N_{signal} and $N_{\text{background}}$ represent the number of signal, and background MC simulated events. The significance quantifies how well the signal stands out from the background, accounting for statistical fluctuations in the background, modeled as Poisson noise through the term $\sqrt{N_{\text{background}}}$. Producing p_T^{cone} and E_T^{cone} distributions allows us to visually compare ATLAS data with signal and background simulations, and provides context regarding signal-to-background ratios.

3.1. *Results*

The analysis was performed iteratively. First, significance and cross-section curves were generated by varying p_T^{cone} cuts without restricting E_T^{cone}, revealing optimal cuts where significance peaked, at $3\,\text{GeV}/c$ for both $Z \to e^+e^-$ and $Z \to \mu^+\mu^-$. Then, fixing p_T^{cone} at this value, E_T^{cone} was varied, yielding optimal cuts of $5\,\text{GeV}/c$ and $4\,\text{GeV}/c$ respectively. Variations in E_T^{cone} caused smaller cross-section shifts ($\sim 1\%$) compared to p_T^{cone} ($\sim 4.65\%$), so systematic error analysis focused on p_T^{cone}. The blue results in Figure 2 show the final iteration, varying p_T^{cone} again with E_T^{cone} now fixed at its optimal value.

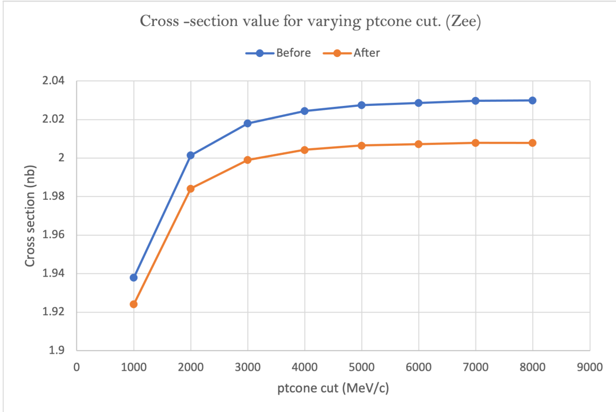

FIG. 2. Shows the variation in the cross-section value for e^+e^- channels as the p_T^{cone} selection cut is varied in steps of $1\,\text{GeV}/c^2$. The blue and orange curves represent values before and after considering unsimulated background estimates.

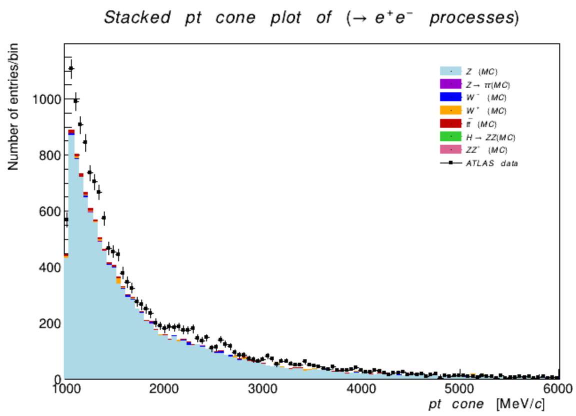

FIG. 3. The p_T^{cone} distribution for e^+e^- events in the range $1\text{–}6\,\text{GeV}/c$. Events with $p_T^{\text{cone}} < 1\,\text{GeV}/c$ are accumulated in the zero bin, which is excluded for clarity. ATLAS data is overlaid with a stacked MC histogram of signal and background events.

The results in Figure 2 show that the cross-section value is stable at high p_T^{cone} cuts but decreases as the cuts are tightened, indicating systematic error. This is likely due to two factors: background events in the ATLAS data not simulated in the MC dataset and signal mismodeling in the MC simulation.

This was supported by the p_T^{cone} and E_T^{cone} distributions. Figure 3 shows a discrepancy between ATLAS and MC datasets. At low values, the ATLAS experiment recorded more events than simulated by MC. This deficiency in the MC dataset could be down to either signal mismodelling, unsimulated background or both. The E_T^{cone} distribution exhibited an uneven bell shape, steep on the negative side with a peak at zero. Negative values arise from detector noise corrections, where very low E_T^{cone} values are adjusted below zero. The MC data oversimulates events at zero E_T^{cone} and undersimulates those in the negative range. This suggests signal mismodelling may be prominent, as unsimulated background cannot explain the lower ATLAS peak or the negative-value discrepancy, which likely correspond to signal events due to their low E_T^{cone}.

4. UNSIMULATED BACKGROUND

One possible source of our systematic error was background events detected by ATLAS but not simulated in MC. To address this, we estimated unsimulated background by analyzing events in which leptons of different flavors ($e\mu$) were detected. Assuming random flavor production in such events, $e\mu$ pairs should occur twice as often as ee or $\mu\mu$, leading to the following estimate:

$$N_{unsim} = \frac{N_{e\mu}^{\text{ATLAS}} - N_{e\mu}^{\text{MC background}}}{2} \tag{3}$$

where $N_{e\mu}^{\text{ATLAS}}$ isolates different flavour events in the ATLAS data, and $N_{e\mu}^{\text{MC background}}$ are the events that pass the same cuts in the simulated background events.

By incorporating this estimation into the $N_{\text{background}}$ term in Equation 2 and the significance calculation, we regenerated the cross-section and significance curves for p_T^{cone} variation, assessing the impact of this adjustment on the nominal cross-section value and associated systematic uncertainty.

4.1. Results

The results on cross-section variation are displayed in orange in Figure 2. The cross-section values decreased relative to results before unsimulated background considerations, but still displayed the same shape in low p_T^{cone} cuts indicative of systematic error. Furthermore, the significance peak value shifted from a cut of $3\,\text{GeV}/c$ to $2\,\text{GeV}/c$ for both $Z \to e^+e^-$ and $Z \to \mu^+\mu^-$ channels. This was a slightly disturbing result as our nominal cross-section value, determined by this optimal cut was shifted into the region with clear instability due to systematic error. This had severe impact on our systematic uncertainty (now $\sim 3\%$) which was generated by varying our cuts by $1\,\text{GeV}/c$ either side of our central value.

As a result, we were forced to question the validity of significance as a measure of selection cut effectiveness. Significance measures the ratio of signal events to statistical fluctuations in the number of background events. Since our systematic uncertainty at this point was $\sim 75\times$ our statictical uncertainty, it is clear that this is not the most powerful method for determining our selection cuts. We know that the higher we cut, the more background we are including and thus a sensible cut is perhaps the lowest value at which the cross-section value appears stable. Using Figure 2 and the $Z \to \mu^+\mu^-$ equivalent, these were determined as 5 and $4\,\text{GeV}/c$ respectively. Before we could confidently shift our selection cuts however, it was important to validate our suspicions that signal mismodelling is the source of systematic error in the low p_T^{cone} cut region.

5. INVESTIGATING SIGNAL MISMODELLING

Using selection cuts we were able to isolate events that (for $Z \to e^+e^-$) were kept when cuts were placed at $5\,\text{GeV}/c$ and rejected with cuts at $2\,\text{GeV}/c$. These represent the events responsible for the difference in cross-section value when the cuts were placed at these two values. By plotting the invariant mass of these events for both ATLAS and MC data, we should gain some insight on the nature of the systematic error.

5.1. Results

The resulting distribution exhibited a peak at the Z-boson mass in both datasets, with a defecit in the MC data most pronounced at the peak. This indicates fewer signal events in MC than in ATLAS, supporting the conclusion that systematic error arises from signal mismodeling and justifying a looser p_T^{cone} cut.

With this we generated our final cross-section value: $\sigma = 2.007 \pm 0.001\,(\text{stat.})\,^{+0.001}_{-0.002}\,(\text{syst.}) \pm 0.034\,(\text{lumi.})$, $\sigma = 1.970 \pm 0.001\,(\text{stat.})\,^{+0.001}_{-0.004}\,(\text{syst.}) \pm 0.033\,(\text{lumi.})$ for decays into electron and muon pairs respectively. The discrepancy between the two results highlights inaccuracies in some of the assumptions made during the analysis, indicating potential areas for refinement. Specifically, when estimating the unsimulated background, the assumption that lepton flavors are produced randomly proved to be inaccurate, as differences in selection efficiency for muons and electrons were observed throughout the analysis. Further studies could involve varying the invariant mass or transverse momentum cuts to optimize the selection criteria and improve the robustness of the method.

6. CONCLUSION

This investigation aimed to measure the cross-section of Z-boson decays into lepton pairs using ATLAS open data and Monte Carlo simulations, while optimizing selection criteria to reduce background contamination and systematic uncertainties. By analysing invariant mass distributions, p_T^{cone} and E_T^{cone} variables, we identified mismodelling of signal events and deficiencies in unsimulated background estimates as key contributors to systematic errors. These findings highlight the importance of refining simulation techniques and selection methodologies to improve the accuracy of future cross-section measurements.

[1] L. D. Lella and C. Rubbia, Reviews of Modern Physics **57**, 699 (2015).

[2] T. N. Foundation, The nobel prize in physics 1979, `https://www.nobelprize.org/prizes/physics/1979/summary/` (1979), accessed: 2024-8-10.

[3] U. o. M. Dept. of Physics & Astronomy, *Third Year Laboratory: Analysing Events with Z, W and Higgs Bosons with the ATLAS Experiment*, version 3.3 ed. (2024).

[4] ALEPH, DELPHI, L3, OPAL Collaborations and LEP Electroweak Working Group , (2005), arXiv:hep-ex/0509008.

[5] L. University, Studies of muon isolation and fsr photons, `https://indico.cern.ch/event/162332/contributions/1415964/attachments/190074/266809/skinnari_muonIso_nov16.pdf` (2011), accessed: 2024-9-10.