LIP-STUDENTS-25-23

Search for rare phenomena with ATLAS Open Data

Maria Barros[1,a], Joana Feio[1,b] and Miguel Saganha[1,c]

[1] *LIP - Minho, Universidade do Minho, Braga, Portugal*

Project supervisor: Nuno Castro

December 2, 2025

Abstract. We present an analysis of the Higgs boson decay channel $H \to ZZ^* \to 4\ell$ using the 13 TeV ATLAS Open Data. Events with exactly four leptons were selected and filtered with different criteria. The two Z bosons were reconstructed using three complementary pairing strategies: (i) invariant mass proximity to the nominal Z mass, (ii) angular separation in the azimuthal angle ϕ, and (iii) transverse missing momentum. The reconstructed four-lepton invariant mass distributions exhibit good agreement between experimental data and Monte Carlo simulations, with a clear enhancement near 125 GeV, consistent with the Higgs boson signal. This study demonstrates how publicly available ATLAS data can be used to explore Higgs boson reconstruction techniques and provides a pedagogical framework for testing different selection and pairing methods in the 4ℓ final state.

Keywords: LHC, Higgs, Z Boson reconstruction

1 Introduction

1.1 The Standard Model (SM)

The Standard Model (SM) of particle physics is the theoretical framework that describes all the fundamental particles and three of the four interactions between them (electromagnetic, weak and strong). The particles are divided in two separate groups:

- Fermions: particles that have half-integer spin and are the building blocks of matter. They are separated in two types of particles, depending on how they interact with other particles. While quarks interact mostly through strong and electromagnetic forces, being mediated by gluons, the leptons interact through electromagnetic and weak forces (primarily the weak force).
- Bosons: particles with integer spin that mediate the fundamental forces. The photon is the mediator of the electromagnetic force and it is massless and chargeless; the W and Z bosons mediate the weak force, with Z boson being neutral and W boson existing in positively and negatively charged forms; the gluons are carriers of the strong force , binding quarks with hadrons and carrying the colour charge. The Higgs Boson has spin 0 and imparts masses to other particles giving every elementary massive particle its rest mass.

1.2 The Large Hadron Collider (LHC)

The LHC is a particle accelerator located at CERN. It was built to test the Standard Model of particle physics, look for its limitations, and find new particles and phenomena to gain a better understanding of the fundamental nature of matter and energy.

As the biggest and most powerful particle accelerator in the world, the LHC speeds up two beams of particles, like protons, in opposite directions to nearly the speed of light within its 27-kilometer-long ring. When these beams collide, detectors like ATLAS measure the results of the high-speed interaction. [1]

1.3 The ATLAS Detector

The ATLAS detector, a general-purpose detector at the LHC, is a large instrument measuring approximately 46 meters in length and 25 meters in diameter. [2] Its design features multiple layers to collect extensive data on a wide variety of particles created in the LHC's high-energy collisions. This data includes information on their trajectories, momentum, energy, and other identifying characteristics.

The main layers of the detector are:

- **Inner Detector**: closest to the collision point, tracking the paths of charged particles;
- **Calorimeters**: surrounding the Inner Detector, measuring the energy of particles. The electromagnetic calorimeter measures electrons and photons, while the hadronic calorimeter measures hadrons;
- **Muon Spectrometer**: using large superconducting magnets and precision chambers to track muons;
- **Magnet Systems**: bending the paths of charged particles, allowing the measurement of their momentum.

ATLAS uses a right-handed coordinate system for particle tracking, where the x-axis points toward the LHC center and the z-axis is aligned with the beamline. The detector is designed symmetrically around the beam axis, specifically in the azimuthal angle ϕ.

The origin of detected electrons can be identified by whether they are isolated or not. **Isolated electrons** are not surrounded by other particles and typically come from the decay of Z or W bosons, or τ leptons. In contrast, **non-isolated electrons** are often produced by the decay of hadrons. Isolation is determined by analyzing the particle's pseudorapidity (η), defined in 1.5, and its azimuthal angle ϕ.

[a] e-mail: mariainesferreirabarros@gmail.com
[b] e-mail: joanafeio2005@gmail.com
[c] e-mail: miguelsaganha15@gmail.com

In proton-proton collisions, the momentum along the beamline is unknown because the interacting partons (quarks and gluons) carry an unknown fraction of the proton's momentum. However, the total momentum perpendicular to the beamline is zero before the collision. Therefore, the **transverse momentum** (p_T), the momentum component perpendicular to the beamline, is a key variable in the analysis of these collisions.

1.4 The Higgs Boson

The Standard Model of particle physics proposes the Higgs mechanism, a process that gives mass to all elementary particles and predicts the existence of the Higgs boson. This particle is an excitation of the Higgs field, and a particle's mass is determined by how strongly it interacts with this field.

At the LHC, the Higgs boson can be produced by colliding two protons at extremely high energy. To detect the Higgs boson, physicists analyze the particles into which it decays. According to the Standard Model, it can decay into pairs of fermions or bosons. The ways the Higgs boson decays, and how often, depend significantly on its mass.

Detecting these decay products is challenging because the cross-section for producing a Higgs boson is much smaller than that of most other Standard Model processes. This makes it a very rare event, requiring a large amount of data from many collisions to draw meaningful conclusions about the Higgs boson.

The Higgs boson was identified in 2012 through an analysis similar to the one explored in this work. It is important to highlight that the Higgs boson has an extremely short lifetime, on the order of 10^{-22}s [3], which makes direct detection impossible. Instead, it is reconstructed from its decay products. As stated in [4], the Higgs boson can decay into various combinations of particles, with probabilities depending on its mass. While several decay channels are possible, one of them stands out: the so-called **golden channel**, because it "presents the clearest and cleanest signature" [5]. This process,

$$H \to ZZ^* \to \ell\ell\ell\ell,$$

results in a final state containing four leptons (electrons, e, or muons,mu).

1.5 Key Variables

Here, we present a list of the main variables that can be measured using the detector, which are integral to this analysis:

- θ: polar angle of a particle's trajectory relative to the counterclockwise proton beam;
- $\eta = -\ln(\tan(\theta/2))$: pseudorapidity;
- ϕ: angle of the object's trajectory in the plane perpendicular to the direction of the proton beams;
- $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$: separation between reconstructed objects within the detector;
- p_T: transverse momentum;
- p_T^{miss}: missing transverse momentum, the magnitude of the negative vector sum of the transverse momenta of all reconstructed objects;
- E_T^{miss}: hadronic missing transverse momentum, the magnitude of the negative vector sum of the transverse momenta of all reconstructed jets (where $p_T > 20$ GeV and $|\eta| < 5.2$);
- m: invariant mass;
- N: number of objects (for instance, N_{aj} denotes the number of additional jets).

Note that some of these variables, such as θ, η, ϕ and p_T, are specific for each of the 4 leptons, while the others are global.

2 Data

2.1 Data access

This analysis utilizes both collision data and simulated Monte Carlo (MC) samples from the ATLAS Open Data initiative, which provides curated datasets from proton–proton collisions at $\sqrt{s} = 13$ TeV. While the full research-grade datasets typically exceed the terabyte scale with minimal preprocessing, this work employs specially prepared educational skims — optimized subsets with reduced event counts and pre-applied event selection criteria.

The `exactly4lep` educational skim was selected for targeted analysis of the $H \to ZZ^* \to 4\ell$ process, as it contains pre-selected events with four identified leptons.

Data retrieval was performed using the ATLAS Open Magic Python package [6], which provides programmatic access to metadata and remote storage locations (URLs) for ATLAS Open Data resources.

2.2 Data structure

The datasets provide essential kinematic variables for lepton reconstruction: transverse momentum (p_T), pseudorapidity (η), azimuthal angle (ϕ), and energy (E). Together, these four-vector components enable the kinematic reconstruction of particle trajectories and the calculation of invariant masses.

In addition, the datasets include several pieces of supplementary information:

- Trigger system data recording single-lepton trigger decisions
- Particle identification flags distinguishing electrons and muons, including quality working points
- Electric charge information enabling opposite-sign lepton pair identification

2.3 Monte Carlo

Monte Carlo simulations are also included in the datasets and play a crucial role in this analysis by providing theoretical predictions for both signal and background processes. The generated events undergo the same reconstruction algorithms applied to collision data, ensuring a consistent framework for comparison.

The signal sample corresponds to Higgs boson production in the $H \to ZZ^* \to 4\ell$ decay channel at the known mass of 125 GeV. The background simulations include major Standard Model processes such as ZZ production, top-quark pairs ($t\bar{t}$), and WZ events. Each background contribution is carefully modeled to ensure accurate representation in the final analysis.

A key feature of the MC samples is the inclusion of event weights. These weights correct the simulated events to match the expected yields in real data, accounting for factors such as cross-sections, detector efficiency, and pileup effects.

3 General Analysis

For each graph we have a representation of:

- Real ATLAS experimental data values, represented as 'Data';
- Values from Monte Carlo (MC) simulations that mimic the ATLAS experiments, including:
  - Background data (*SM processes:* $Z, t\bar{t}, t\bar{t}+V, VVV$; ZZ^*).
  - Signal data.

The background data from the MC simulations aim to reproduce events that yield final states similar to those expected from Higgs boson production. In particular:

Background ZZ^* — "ZZ diboson production in pp interactions at the LHC provides a test of the electroweak sector of the SM. The SM ZZ production can proceed via a SM Higgs boson propagator, although this contribution is suppressed in the region where both Z bosons are produced on-shell. Hence, non-resonant ZZ diboson production constitutes an important background for searches of the SM Higgs boson decaying to ZZ^*, as discussed in Section 3.9. The analysis presented here focuses on implementing selection criteria for ZZ diboson events where both Z bosons decay to leptons." [7, sec. 3.8]

Background $Z, t\bar{t}, t\bar{t}+V, VVV$ — "Other background contributions arise from $t\bar{t}$ and Z+jets production (studied in Section 3.3 and Section 3.4, respectively), where charged lepton candidates originate either from hadron decays involving b- or c-quark content, or from misidentification of jets." [7, p. 20]

For each histogram, the data could be filtered with custom user-defined criteria, with a pre-made option, or left unfiltered. These pre-made filters correspond to requiring leptons with transverse momenta of $p_T > 25, 15, 10, 7$ GeV, respectively, as specified in the 13 TeV ATLAS Open Data release [7, p. 20]. For most of the graphs, we applied these filters to reduce background events more effectively, since the Higgs signal is comparatively rare.

3.1 Basic Variables (Single lepton)

This section reflects on three histograms describing the distribution of the transverse momentum (see Fig. 1), the pseudorapidity, and the multiplicity of leptons per event available in Appendix 7.1, with the preset p_T filter applied.

For the first histogram (Fig. 1), due to the applied cut, there are no events with leptons having p_T lower than 7 GeV. As a result of the filtering and selection, the number of events for the *Background* $Z, t\bar{t}, t\bar{t}+V, VVV$ is significantly reduced, while the distributions of *Background* ZZ^* and the signal remain very similar. Therefore, these graphs are not particularly helpful for identifying the Higgs boson. The real data, as shown, follows these latter distributions, suggesting that it primarily consists of such events.

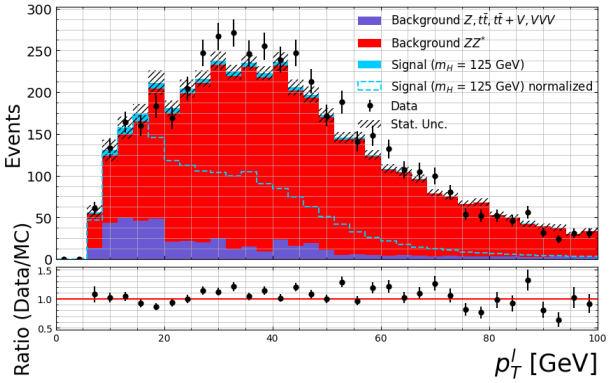

Figure 1: Transverse momentum (p_T) distribution for all individual leptons in all events from both experimental data and Monte Carlo simulations.

3.2 4-Lepton System Reconstruction

With the kinematic variables of the individual leptons, we reconstruct the properties of the parent 4-lepton system for each event. By calculating the invariant mass and transverse momentum of the system, we can identify resonant structures that may correspond to intermediate particles such as the Higgs boson.

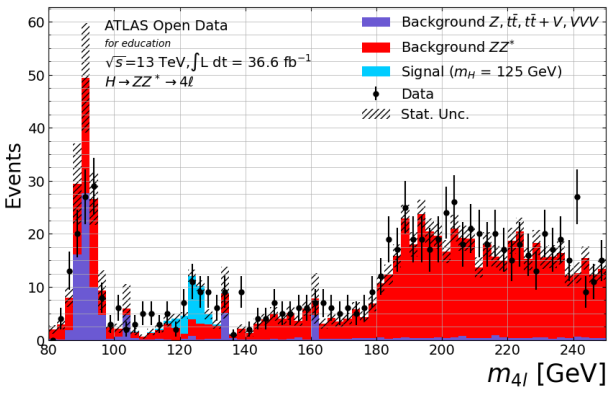

Figure 2: 4-lepton invariant mass distribution for each event from both experimental data and Monte Carlo simulations.

This histogram (Fig. 2) is very relevant for determining the Higgs boson signal, as it represents the calculated invariant mass of the four-lepton system in each event.

We observe an overall agreement between the data and the MC background/signal. Around 125 GeV, the expected Higgs peak is visible. Additionally, a peak around 90 GeV (the Z boson peak) is observed, which arises mostly from background.

4 Z Boson Pair Reconstruction

In this analysis, the dataset was utilized to reconstruct the intermediate ZZ^* bosons originating from the decay of the Higgs boson, specifically focusing on the $H \rightarrow ZZ^* \rightarrow 4\ell$ decay channel, where the final state consists of four leptons. Having already selected events containing four leptons, categorized into two opposite-sign same-flavor (OSSF) pairs (which could consist of two electron pairs, two muon pairs, or one of each), the next step involved reconstructing the two Z bosons from these leptons using the available kinematic features.

The reconstruction was performed using three distinct methodologies.

4.1 Mass Criterion

The first method involved enumerating all six possible pairwise combinations of the four leptons that satisfy the OSSF condition. For each pair, the invariant mass was calculated. The pair whose invariant mass was closest to the nominal Z boson mass of 91.188 GeV [10] was identified as the first Z candidate. The remaining two leptons naturally formed the second Z candidate. This approach leverages the well-known mass of the Z boson as a key discriminant to identify the most probable pairing consistent with the decay hypothesis.

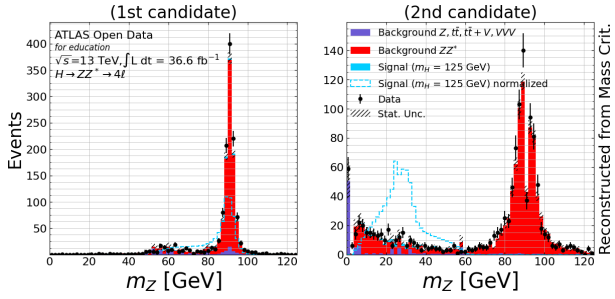

Figure 3: Calculated mass distribution for the real Z-boson candidate (left) as well as the virtual one (right), using the mass criterion.

4.2 Angular Separation Criterion

The second method focused on the angular separation between leptons, specifically utilizing the azimuthal angle ϕ measured in the detector coordinate system. Again, all six possible lepton pairings were considered. Instead of invariant mass, the difference in ϕ between leptons in each pair was computed. The pair exhibiting a $|\ \Delta\phi\ |$ value closest to π was selected as the first Z candidate, under the assumption that leptons produced from the decay of the on-shell Z boson have higher transverse momentum and therefore tend to be emitted in opposite directions ($\vec{p}_T^{\,\ell^-} + \vec{p}_T^{\,\ell^+} \approx 0$), resulting in the largest angular separation. The complementary pair of leptons formed the second Z candidate.

This method is motivated by the decay kinematics of the Z boson, where the decay products are expected to be correlated in direction, providing an alternative criterion for pairing leptons.

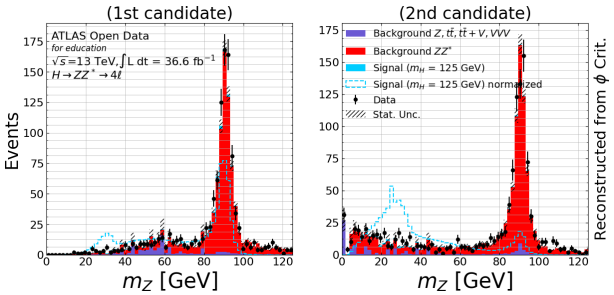

Figure 4: Calculated mass distribution for the real Z-boson candidate (left) as well as the virtual one (right), using the ϕ criterion.

4.3 Transverse Momentum Criterion

The final reconstruction method exploited the transverse momentum balance between leptons in each possible OSSF pair. For every event containing four selected leptons, all six unique lepton pair combinations were considered. For each pair, the scalar sum of the individual lepton transverse momenta was computed. Only OSSF pairs were retained, consistent with the $Z \rightarrow \ell^-\ell^+$ decay signature.

From the ensemble of valid lepton pairs, the pair exhibiting the largest scalar sum of lepton transverse momenta was designated as the primary candidate, assumed to originate from the on-shell Z boson (m_Z = 91.188 GeV). The remaining leptons were assigned to the secondary candidate, interpreted as the virtual Z^* boson decay products. This strategy is motivated by the principle of mass-energy conservation: the on-shell Z boson, with its larger mass, produces leptons with higher average transverse momentum compared to those from the lighter virtual state.

The p_T sum criterion thus provides a complementary kinematic discriminant to invariant mass and angular separation, offering enhanced performance in regions where mass resolution is limited or angular correlations are less distinctive.

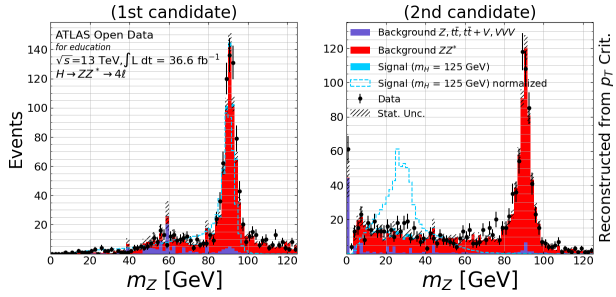

Figure 5: Calculated mass distribution for the real Z-boson candidate (left) as well as the virtual one (right), using the p_T criterion.

All of these reconstruction strategies aim to optimize the correct identification of the ZZ^* system, thereby improving the fidelity of the Higgs boson decay channel analysis. Comparing the performance of these methods provides insight into the underlying event topology and detector resolution effects, and may inform future refinements in event selection and reconstruction algorithms.

Distributions of additional ZZ^* boson characteristics are provided in Appendix 7.2 to facilitate direct comparison between the three reconstruction methods.

5 Jets Analysis

In hadronic collisions it is common to have jets in the final result. We obtained the distribution of the number of jets per event (see Appendix 7.3). This typology should not, in first order, produce any jets. As the final state of this decay are four leptons, that are not subject to strong nuclear force neither interact via this force, they cannot fragment and hadronize into sprays of particles (jets). So the jets detected in our events do not come from the Higgs decay, but from or additional hadronic activity from the proton-proton collision.

To investigate the impact of jets on the kinematics of the particles in our analysis, we compared the transverse momentum distributions for events with and without reconstructed jets at three distinct reconstruction levels:

- 4-lepton System (potential Higgs candidate),
- Z Boson Subsystem (reconstructed from lepton pairs),
- Individual Leptons.

5.1 4-Lepton System

When the Higgs boson is produced without additional jets, transverse momentum conservation implies that its p_T should be low. Since given its high mass it is produced almost at rest and no significant additional objects are present to balance it, the Higgs candidate should remain close to zero p_T. The higher values observed in the histograms (Fig. 6 and Fig. 7) arise from the fact that, in practice, some jets are not reconstructed and additional particles, such as photons, contribute to the final state, forcing the Higgs candidate to compensate with a nonzero p_T.

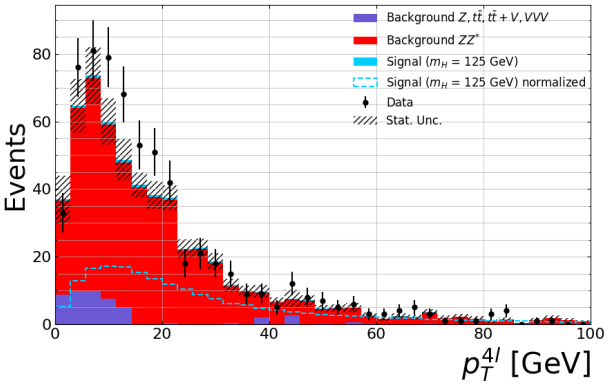

Figure 6: Transverse momentum (p_T) distribution of the 4-lepton system in events with zero jets.

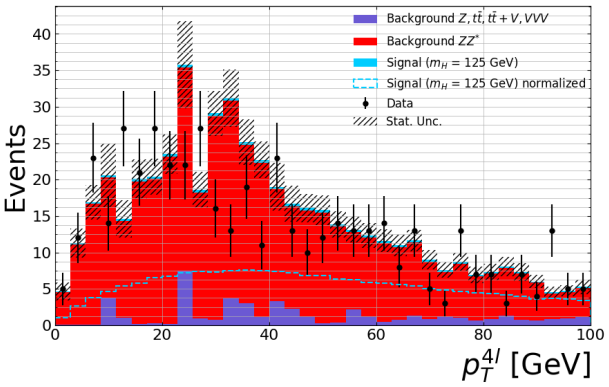

Figure 7: Transverse momentum (p_T) distribution of the 4-lepton system in events with nonzero jets.

When the Higgs boson is produced in association with jets, the jets carry transverse momentum that must be balanced by the Higgs candidate in order to conserve momentum in the transverse plane. This typically results in a higher Higgs p_T correlated with jet activity, as confirmed in the histogram.

5.2 Z Boson Subsystem

As expected, the transverse momentum of the reconstructed ZZ^* system is lower in events without additional jets (see Fig.8 and Fig.9). However, it is not as low as in the 4-lepton system case, since the reconstruction of ZZ^* involves multiple objects that interfere with the final result, allowing more possibilities to compensate the extra transverse momentum and cancel its vector sum.

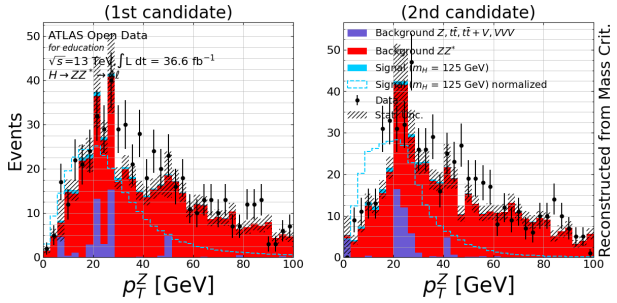

Figure 8: Transverse momentum (p_T) distribution of the reconstructed Z-boson candidates (real and virtual) in events with zero (top) jets.

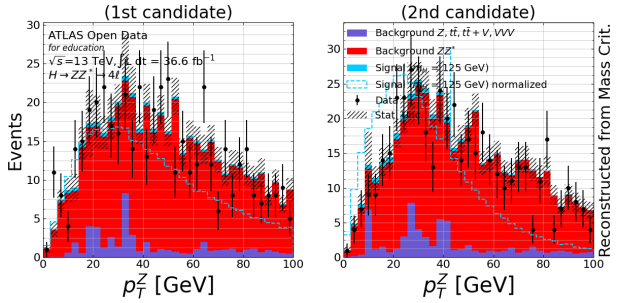

Figure 9: Transverse momentum (p_T) distribution of the reconstructed Z-boson candidates (real and virtual) in events with nonzero (bottom) jets.

Although the difference is less pronounced, we observe the same effect as in the 4-lepton system: events with additional jets exhibit higher transverse momentum to balance the jet activity.

5.3 Individual Leptons

For individual leptons in events without additional jets (see Fig. 10), the transverse momentum distribution is not as close to zero as in the 4-lepton system. This is expected, since the decay produces several objects, which means there are more opportunities for transverse momentum compensation among the decay products.

In contrast, when comparing events with and without jets (Fig. 10 and Fig. 11), the distributions are very similar, indicating that the presence of jets has a much smaller effect at the single-lepton level. This explains why the jet-related differences are more clearly visible in the 4-lepton system (Sec. 5.1) than at the level of individual leptons.

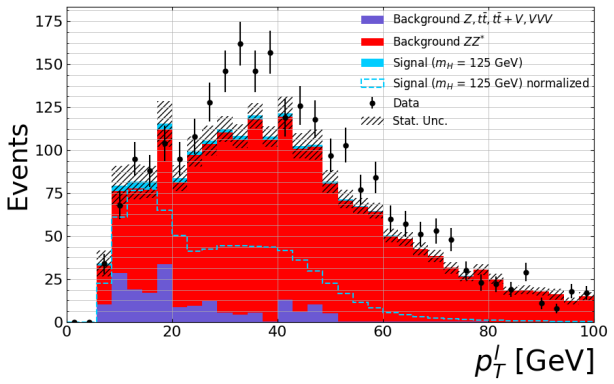

Figure 10: Transverse momentum (p_T) distribution of individual leptons in events with zero jets.

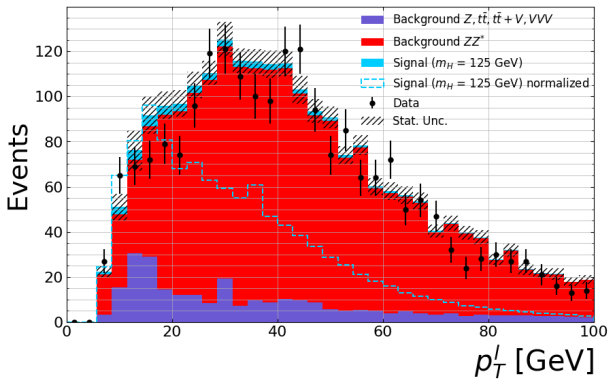

Figure 11: Transverse momentum (p_T) distribution of individual leptons in events with nonzero jets.

6 Results and Conclusions

In this study, we carried out an analysis of ATLAS Open Data for a particular typology, extending beyond the standard Higgs boson rediscovery to perform a detailed reconstruction of Z boson candidates using multiple methodological approaches. By implementing three distinct reconstruction criteria, we identified varying levels of success, yielding valuable insights into the challenges of particle reconstruction in proton–proton collisions.

These reconstruction methods provided insight into the underlying physical processes, particularly the interplay between real and virtual Z boson production in $H \to ZZ^* \to 4\ell$ events. Our analysis demonstrated that the different reconstruction strategies capture complementary aspects of the decay kinematics, with each method offering distinct advantages for specific analysis objectives.

References

[1] ATLAS Open Data, https://opendata.atlas.cern/docs/documentation/introduction/introductory_page, (Accessed: 08-2025)

[2] Introduction to the ATLAS experiment, https://opendata.atlas.cern/docs/documentation/introduction/introduction_ATLAS

[3] CERN, The Higgs Boson, https://home.cern/science/physics/higgs-boson/how

[4] Standard Model of Particle Physics and Beyond, https://opendata.atlas.cern/docs/documentation/introduction/SM_and_beyond

[5] ATLAS Experiment, *New ATLAS precision measurements of the Higgs Boson in the 'golden channel'*,https://atlas.cern/updates/briefing/higgs-golden-channel

[6] ATLAS Open Data, ATLAS Open Magic, https://opendata.atlas.cern/docs/data/atlasopenmagic

[7] The ATLAS Collaboration, *Review of the 13 TeV ATLAS Open Data release*, https://cds.cern.ch/record/2707171/files/ANA-OTRC-2019-01-PUB-updated.pdf

[8] ATLAS Open Data, 13 TeV 2025 Data - Beta, https://opendata.atlas.cern/docs/data/for_education/13TeV25_details

[9] ATLAS Open Data, https://opendata.atlas.cern/docs/documentation/physic_objects/jets, (Accessed: 08-2025)

[10] Particle Data Group, *See the related review(s): Z Boson*, https://pdg.lbl.gov/2024/listings/rpp2024-list-z-boson.pdf

7 Appendix

7.1 Single lepton variables

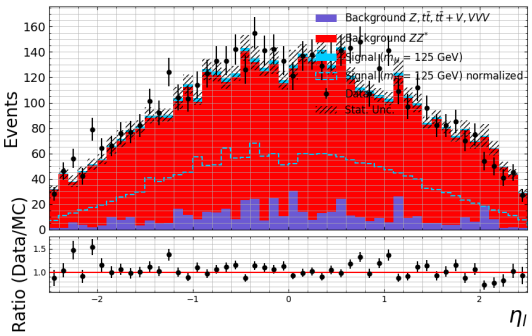

Figure 12: Pseudorapidity (η) distribution of individual leptons.

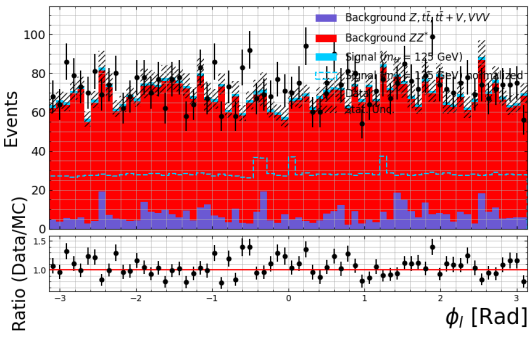

Figure 13: Phi angle (ϕ) distribution of individual leptons.

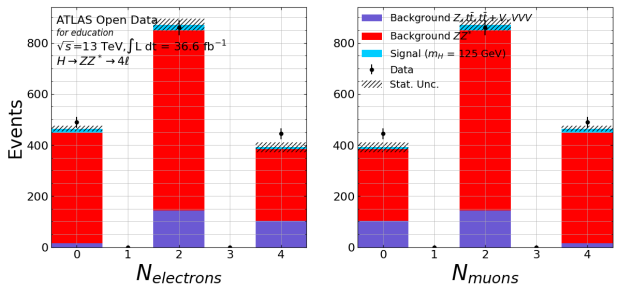

Figure 14: Number and flavor (electron or muon) distribution of individual leptons.

7.2 Z Boson reconstruction variables

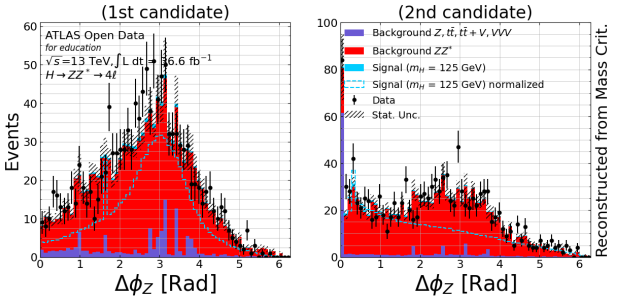

Figure 15: Calculated phi angle differ ence for the real Z-Boson (left) as well as the virtual one (right), using the mass criterion.

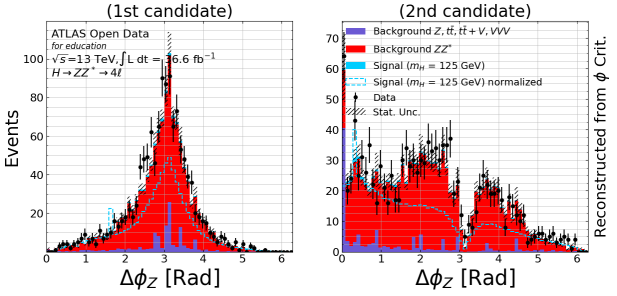

Figure 16: Calculated phi angle difference for the real Z-Boson (left) as well as the virtual one (right), using the angular separation criterion.

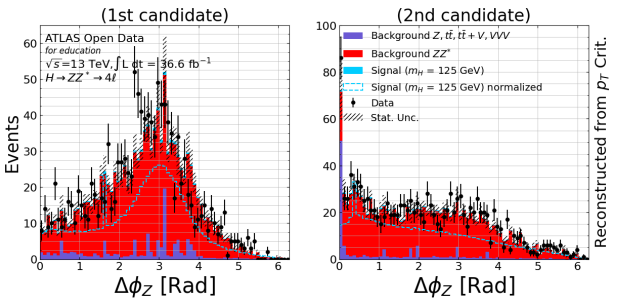

Figure 17: Calculated phi angle difference for the real Z-Boson (left) as well as the virtual one (right), using the transverse momentum criterion.

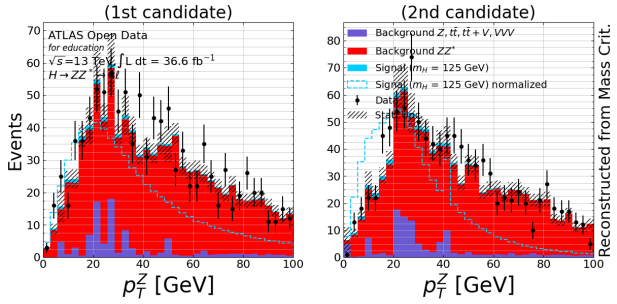

Figure 18: Calculated transverse momentum sum for the real Z-Boson (left) as well as the virtual one (right), using the mass criterion.

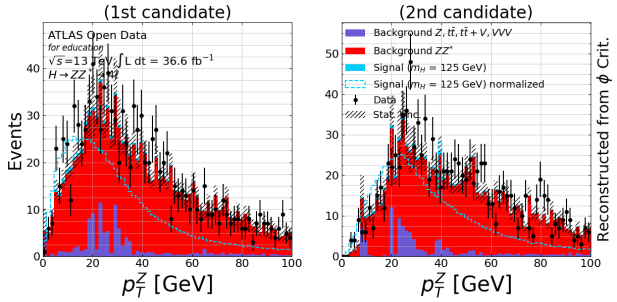

Figure 19: Calculated transverse momentum sum for the real Z-Boson (left) as well as the virtual one (right), using the angular separation criterion.

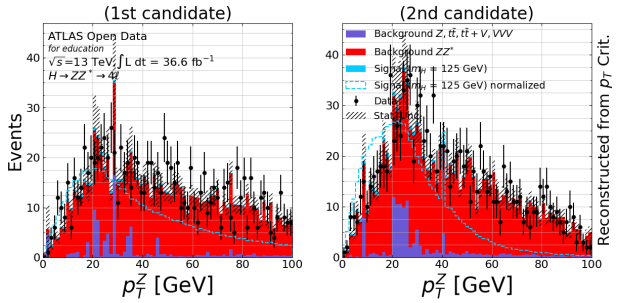

Figure 20: Calculated transverse momentum sum for the real Z-Boson (left) as well as the virtual one (right), using the transverse momentum criterion.

7.3 Distribution of the number of jets

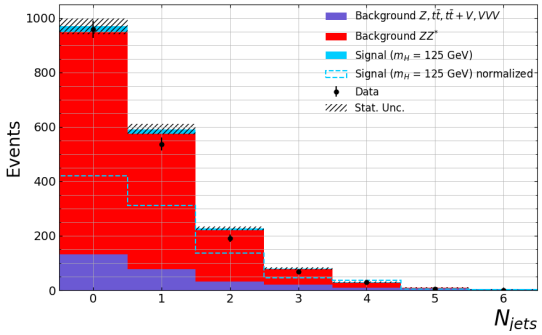

Figure 21: Number of jets distribution per event.