Big Data: ATLAS Proving the Existence of the Higgs Boson

The Tiffin Girls' School

Ayda, Soniya & Phoebe

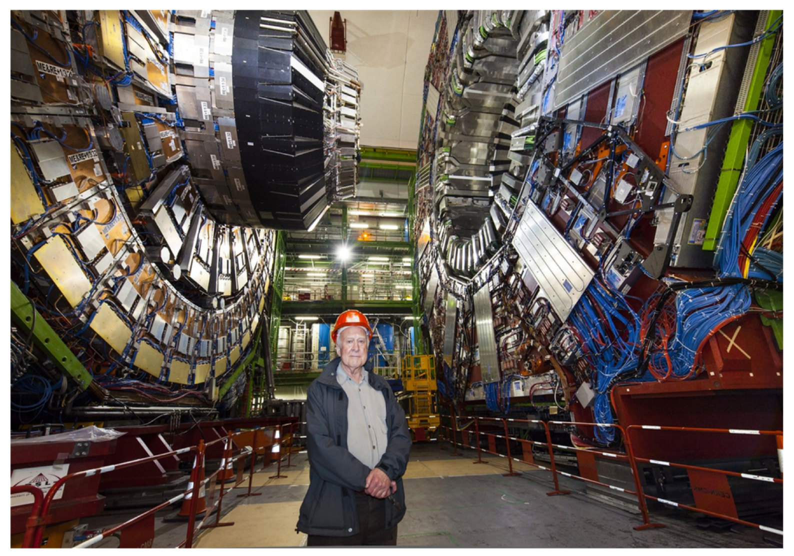
Summary

The Higgs field plays a critical role in the Standard Model of Particle Physics: all elementary particles interact with the field, through the Higgs mechanism, and acquire mass – enabling them to form stable matter and give rise to the complex structures one can observe in the universe. The Higgs boson is the mediator the Higgs field, and is therefore responsible for the existence of mass in the universe.

The overall aim of our project was to reprove the existence of the Higgs boson, so that we may experience the satisfaction of taking part in our own research. To do this, we had smaller aims: develop Python coding techniques, gain a deeper understanding of particle physics, and of course the Higgs boson.

Background information

After first being proposed in the 1960s by theoreticians Robert Brout, François Englert and Peter Higgs, the existence of the Higgs boson was finally confirmed in 2012. Data collected from the detectors, namely ATLAS and CMS, in experiments which had been carried out at the Large Hadron Collider in CERN to prove the Higgs boson's existence displayed a deviation from expectations: the invariant mass distribution of the two photons produced via events concerning the diphoton channel showed a slight bump near 126GeV, corresponding to a global significance of 2.2 standard deviations above the Standard Model (SM) background, consistent with the Higgs boson's hypothesised mass[1].

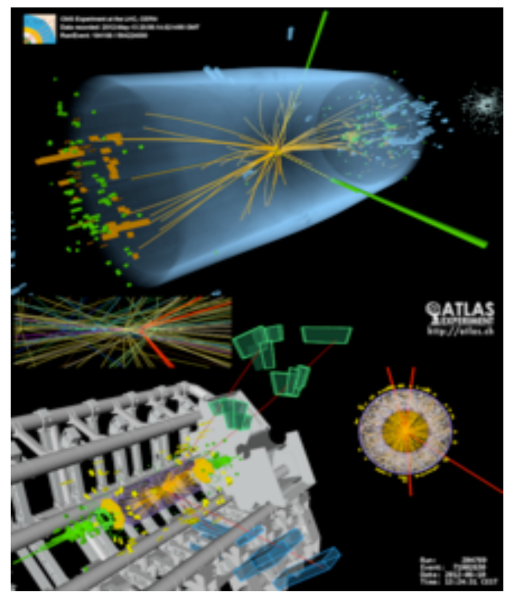


Peter Higgs, standing in front of the CMS detector at the Large Hadron Collider at CERN.[5]

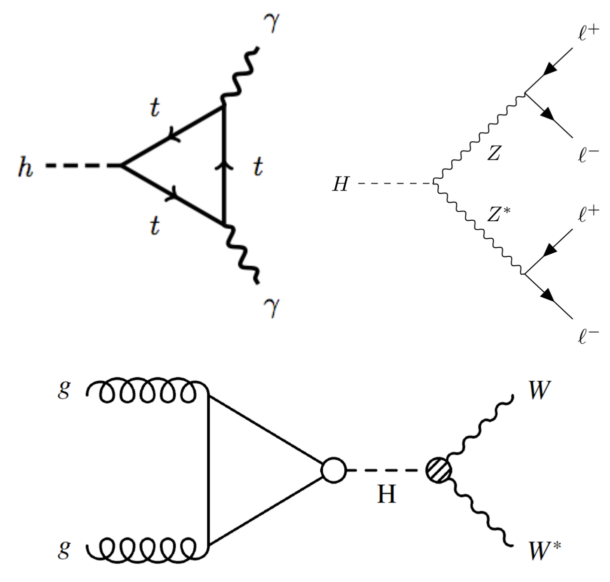
Research aims

Our project was aimed at proving the existence of the Higgs Boson through two of the most sensitive Higgs Boson decay channels. First, we wanted to explore the $H \rightarrow \gamma\gamma$ decay channel (2.8σ local significance[1]), and look for evidence of the Higgs boson in the graphs of diphoton invariant mass, expecting a mass of 125 GeV. Next, we investigated the $H \rightarrow WW$ decay channel (1.4σ local significance[1]), and this required a different method involving non-resonant research. Finally, we looked at the $H \rightarrow ZZ$ channel (2.1σ local significance[1])

To complete this research, we used the ATLAS Open Data for 13TeV events2, and explored using different data files to get the more reliable and accurate results.



"Candidate Higgs boson events from collisions between protons in the LHC. The top event in the CMS experiment shows a decay into two photons (dashed yellow lines and green towers). The lower event in the ATLAS experiment shows a decay into four muons (red tracks)"[3]



Top left: Higgs to gamma gamma channel Feynman diagram
Top right: Higgs to Z-bosons channel Feynman diagram
Bottom: Higgs to W-bosons channel Feynman diagram[4]

Experimental Method

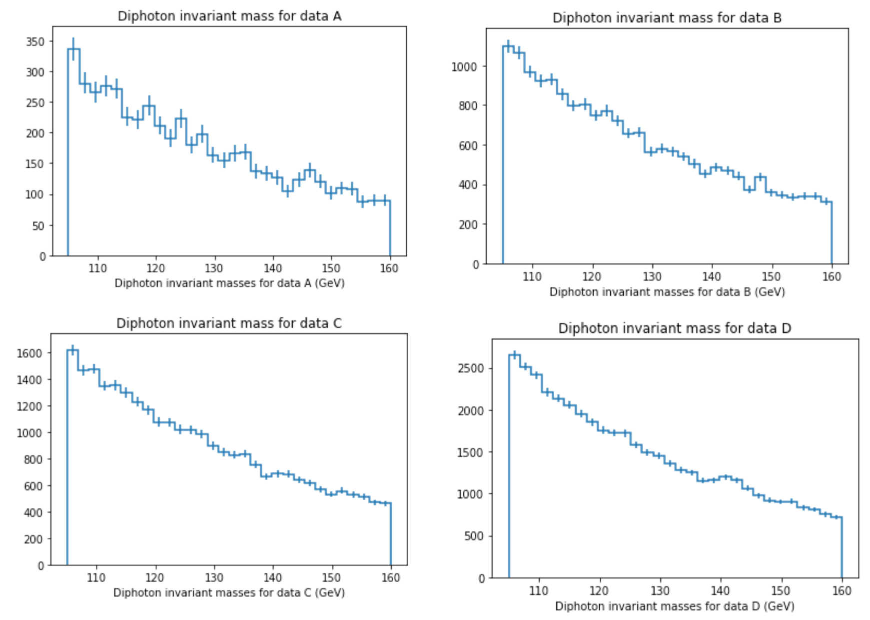
We utilised the 13TeV ATLAS Open Datasets, which provided us with raw data produced from the real proton-proton collisions detected by ATLAS, in addition to simulation samples. These event data files formed the 'root' files for our analysis using a selection of Python libraries suited to our research: the uproot library in order to process the large datasets and carry out statistical analyses on them using the numpy library for numerical computing and the hist library within the larger Matplotlib library for data visualisation which allowed us to plot histograms from our analyses.

<u>H to gamma gamma channel</u>

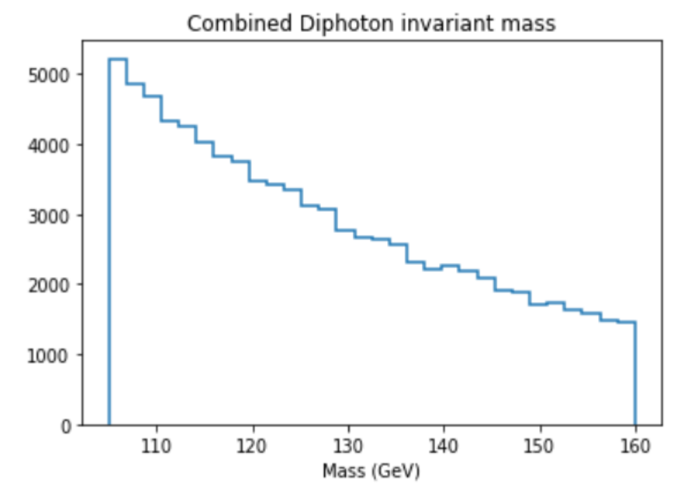
In order to prove the existence of the Higgs Boson through the Higgs to gamma diphoton channel, we needed to plot the combined mass of the two photons produced in the selected events chosen after making suitable cuts to ensure that only events involving a two-photon system produced by this decay channel were looked at. The process, including the cuts involved, was:

1. Loop through each event in the TTree named mini from the ROOT file providing the event data
2. In each event, search for good quality photons (they must pass the diphoton trigger, must pass "Tight" requirements, have pT > 25GeV, be in the 'central' region of ATLAS with |eta| < 2.37, not fall in the 'transition region' between ATLAS's inner detector barrell and ECal endcap i.e. 1.37 <= |eta| <= 1.52
3. If the two photons are well-isolated, extract their 4 momentum from the pT, η, ϕ and energy Ttree branches, and store in a TLorentz vector (which packages together the energy and momentum of a particle into a 4-vector). A pre-written function allowing you to extract the mass from a TLorentz vector was provided inside the Jupyter Notebook.
4. Add the TLorentz vectors of the two vectors of the two photons together.
5. Calculate the invariant mass of the two-photon system
6. Check each photon makes up a minimum fraction of the diphoton system invariant mass
7. Fill the histogram with the invariant mass of the two-photon system

We repeated this process for all the datasets provided by the documentation, plotting separate histograms for each dataset:

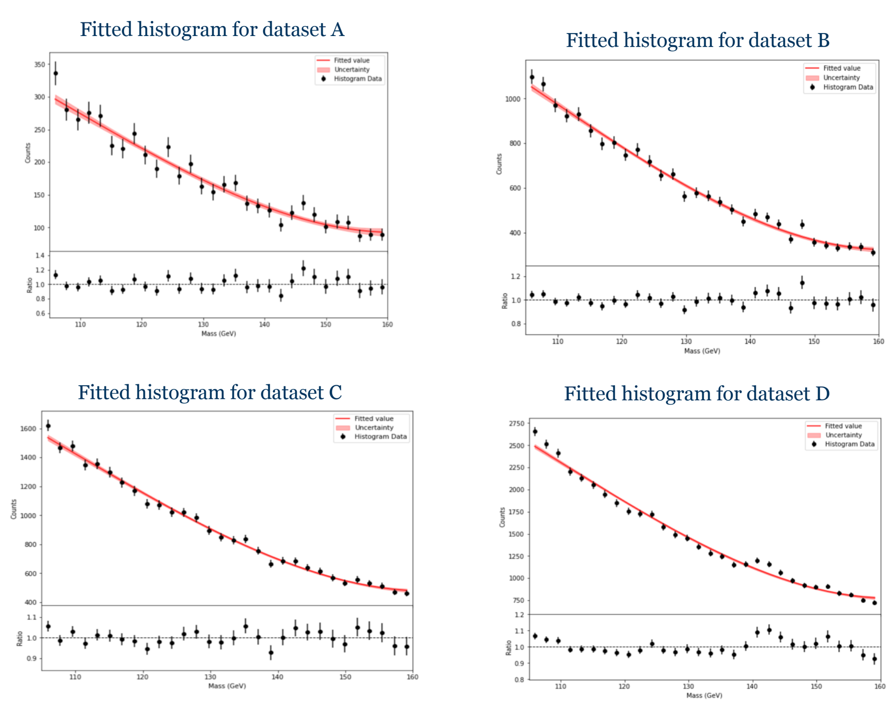


The datasets had varying numbers of events, with dataset D having the highest number (3,600,000 events), making the process take longer for this dataset. Due to this, when merging the histograms for the different datasets in order to make the Higgs boson mass' bump more visible, we were unable to include all of dataset D's data as we were limited by our computers' processing power.



Merged histogram for the diphoton invariant masses obtained from all events in A, B, C and 2,900,000 events of D (ran out of memory).

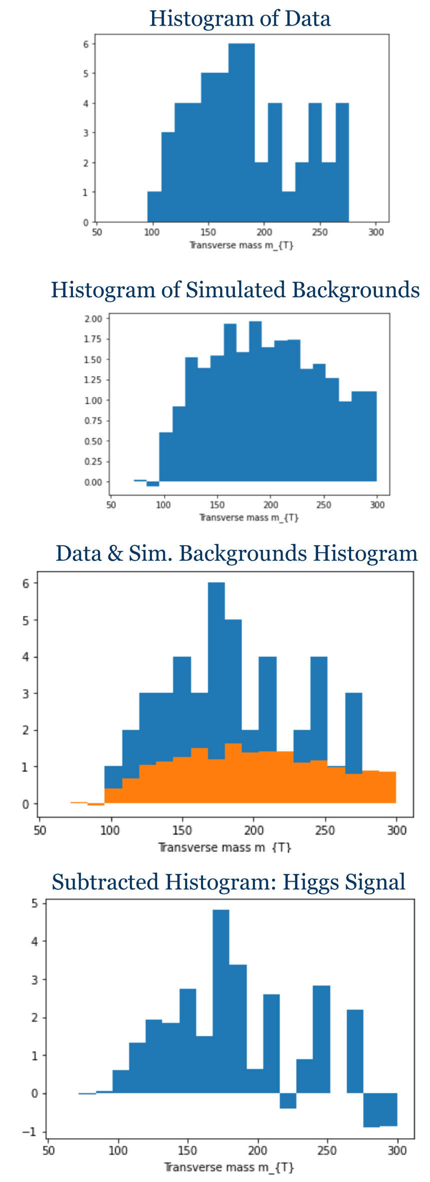
To make the Higgs bump clearer to see, we produced a prediction of what the background would look like using a cubic function to fit the graph and plotted the data against this to make it stand out more. Dataset D



<u>H to WW Channel</u>

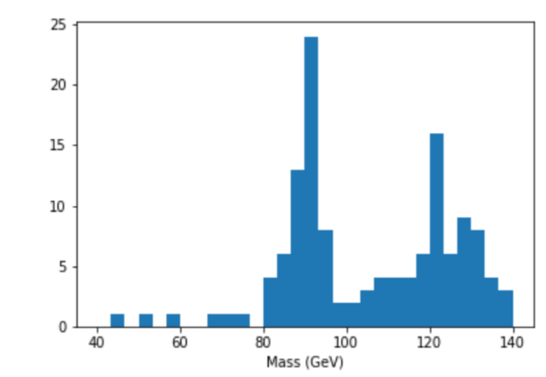
To find a signal for the Higgs boson decaying to 2 W-bosons, we used the non-resonant search technique to histogram the transverse mass of the decay products of the 2 W-bosons. To perform this analysis, we selected leptons with opposite charges and high transverse momenta, which are consistent with the decay of the Higgs boson to two W bosons. Then, we applied various selection criteria to ensure that the selected events are consistent with the signal hypothesis and reject events that may have arisen from background processes. Next, we used the Monte Carlo simulations to model the background contributions. These events were processed through the same selection criteria as the data and were plotted on histograms (shown to the right).

We drew the data and simulated backgrounds on the same canvas. Lastly, we subtracted the backgrounds from the data to obtain our Higgs signal.



<u>H to ZZ Channel</u>

To investigate the $H \rightarrow ZZ$ channel we used our knowledge from Book 4 (when we calculated the Z boson's invariant mass by reconstructing two leptons). We modified our code so that it would check for 4 SFOS (same-flavour opposite-sign) leptons in each event and from that we were able to reconstruct two Z bosons using two pairs of lead and trail vectors. Then we added the two Z boson vectors in order to reconstruct the Higgs boson and plotted its histogram to obtain the signal (including backgrounds). Below is the combined histogram for all the files.



4 Lepton Invariant Mass

Conclusion

After viewing all of our histograms, it was clear that there was a pronounced bump at 125 GeV in all of them, the expected invariant mass we would expect from the Higgs boson. Since we conducted multiple investigations through different decay channels, it solidified our evidence for observing the Higgs boson. We ensured that we worked well as a team and divided up sections to focus our analysis on, and met regularly to share our findings, and create a plan of our next steps. It also meant we could work through coding challenges together, as we were able to share similar experiences and collaborate to find a solution. We also watched some lectures and videos to enable us to understand the physics prior to diving deep into particles without knowing what was really happening, so that we could get the most out of the project.

Citations

1. ATLAS Collaboration. "Combined search for the Standard Model Higgs boson using up to 4.9 fb^{-1} of pp collision data at $\sqrt{s}$ = 7TeV with the ATLAS detector at the LHC" Physics Letters B, vol. 710, 2012, pp. 49-66, doi: 10.1016/j.physletb.2012.02.044
2. Opendata.atlas.cern. (2020). ROOT files & collections - ATLAS Open Data 13 TeV Documentation. [online] Available at: http://opendata.atlas.cern/release/2020/documentation/datasets/files.html (Accessed 24 Apr. 2023).
3. ATLAS and CMS, C. (2013) Candidate Higgs Boson events from collisions between protons in the LHC, CERN Document Server. Available at: https://cds.cern.ch/record/1630222 (Accessed: April 27, 2023).
4. CERN accelerating science (no date) ATLAS Experiment at CERN. Available at: http://atlas.cern/ (Accessed: April 27, 2023).
5. Schirber, M. (2013) Nobel prize - Why particles have mass, Physics. American Physical Society. Available at: https://physics.aps.org/articles/v6/111 (Accessed: April 25, 2023).

L008

IRIS The Institute for Research in Schools

University of Oxford

Science & Technology Facilities Council Rutherford Appleton Laboratory

The Tiffin Girls' School Sapere Aude