Group: 20

Board Member: Matthew Wing

Project Title: Rediscovering the Higgs boson at the CMS experiment

Word Count: 2867

Identifier	Role
{AA}	Minute-taker
{BB}	Data analysis
{CC}	Literature review
{DD}	Data analysis
{EE}	Group lead

Abstract

The discovery of the Higgs boson represented a significant milestone in particle physics. In this project, we analyse simulated and real proton-proton collision data accessed through the CMS Open Data repository. Focusing on reconstructing the Higgs boson through its prominent decay channels, namely $H \rightarrow \gamma\gamma$ (diphoton) and $H \rightarrow ZZ \rightarrow 4l$ (four-lepton), were reconstructed. For diphoton decay, simulated data from 2011, 2012 and 2015 was processed to plot and analyse the diphoton invariant mass distributions. After applying Gaussian fits, the peak positions for the Higgs mass were found to be(125.09 ± 0.15)GeV (2011), (124.43 ± 0.16) GeV (2012), and (124.71 ± 0.09) GeV (2015), respectively. Similarly, data from 2011 and 2012 were used to reconstruct the Higgs boson mass for the four-lepton decay channel. The reconstructed masses for the Higgs boson are found to be (124.95 ± 0.12) GeV (4μ, 2011), (123.76 ± 0.16)GeV ($4e$, 2011), (124.58 ± 0.18)($2e2\mu$, 2011) GeV, (125.75 ± 0.60) GeV (4μ, 2012), (125.29 ± 0.28) GeV ($4e$, 2012), and (126.74 ± 0.71) GeV ($2e2\mu$, 2012), respectively. Real double-electron collision data from 2011 were analysed using a Crystal Ball fitting function, yielding a reconstructed mass of (120 ± 0.66) GeV, which deviates from the expected Higgs mass primarily due to dataset selection biases. By comparing our findings to the CMS publications, we evaluate the reliability of publicly available CMS data for independent research and validate the results obtained on the rediscovery of the Higgs boson. The Higgs boson signals were successfully identified and reconstructed using simulated diphoton and four-lepton mass data from the CMS experiment at CERN. A Gaussian fit was applied to the invariant mass distributions of these channels to derive the results. Overall, these results clearly agree with the expected Higgs boson mass of approximately 125 GeV, with slight variations possibly arising due to differences in integrated luminosity, cross-section uncertainties, and filtering criteria applied to the datasets.

Introduction

Following the discovery of the Higgs boson in 2012, the data from the ATLAS and CMS experiments at CERN was made publicly available in order to "make scientific research more reproducible, accessible, and collaborative." [15] This project aims to reproduce the results from the CMS experiment using the CERN Open Data platform and evaluate its reliability and ease of use. Allowing public access to level 3 data, which is the data used in most studies [15], means that the information can be used for educational purposes, or scientists in other fields will be able to create their analyses, for example.

The CMS (Compact Muon Solenoid) is a general-purpose detector at the LHC in Geneva, collecting data to study a wide range of unanswered questions in physics – from the Standard Model to the search for extra dimensions [16]. The CMS is structured in a way that identifies different particles in different regions of the detector. For example, the paths of charged particles are bent by powerful magnets and tracked in the inner tracker; electrons and photos deposit their energy in the electromagnetic calorimeter (ECAL); hadrons in the hadronic calorimeter (HCAL); and muons are measured in the muon chambers [17].

The Higgs mechanism and boson are key parts of the Standard Model, which is a theory within quantum mechanics that classifies the elementary particles and the associated fundamental forces (strong, weak, and electromagnetic). The Lagrangian density (referred to simply as the 'Lagrangian') of the Standard Model defines all of these fundamental interactions, describing the motion and energy of fields and their interactions. The Lagrangian consists of kinetic and potential energy terms with the kinetic term equal to $\frac{1}{2}(\partial_{\mu}\phi)(\partial^{\mu}\phi)$ and $V(\phi) = \mu^2\phi^2 + \lambda\phi^4$ where μ2 and l determine the behaviour of the field. When μ2 < 0, the Higgs field acquires a non-zero vacuum expectation value (VEV), given by $v = \sqrt{-\frac{\mu^2}{\lambda}}$. This non-zero VEV breaks the Lagrangian's symmetry causing particles to gain mass. The process is known as spontaneous symmetry breaking. [6]

Fermions interact with the Higgs field via Yukawa couplings, giving rise to their mass through the relation $m_f = y_f v$, where y_f is the coupling constant and v is the VEV, approximately 246 GeV. Hence, this explains why heavier particles, e.g. the top quark, have stronger couplings to the Higgs boson. [6]

The Standard Model predictions align with experimental results that state that the Higgs boson itself is a scalar particle, with no intrinsic spin and remains rotationally invariant ($J^P = 0^+$). It can, in theory, decay into any particle within the Standard Model, but because of the coupling to the mass of the particles involved, the branching ratio is much greater for heavier particles. For a Higgs boson of mass 125 GeV, the decay mode $H \rightarrow b\bar{b}$ has the highest branching ratio of 57.8%. However, when attempting to find the boson, this was not the decay channel that was chiefly considered – CMS investigated "channels with distinctive final-state topologies" such as $H \rightarrow \gamma\gamma$ and $H \rightarrow l^+l^-l'^+l'^-$ which meant that they were more clearly distinguishable from the background. [6]

The four-momentum $p^u = (E, p_x, p_y, p_z)$ of the reconstructed events was derived through the equations:

$$p_x = p_T \cos(\phi) \qquad (1)$$

$$p_y = p_T \sin(\phi) \quad (2)$$

$$p_z = p_T \sinh(\eta) \quad (3)$$

$$E = \sqrt{p_x^2 + p_y^2 + p_z^2 + m^2} \quad (4)$$

Where p_T represents the particle's transverse momentum, η is pseudorapidity, ϕ is azimuthal angle, and m is mass. The Higgs particle is identified by examining the invariant mass, shown in equation 2, of the particles involved in a proton-proton collision at the LHC.

$$M^2 = (\sum E_i)^2 - |\sum p_i|^2 \quad (5)$$

where E_i is the energy of each particle and $p_i = (p_x, p_y, p_z)$, the three-momentum of each particle. This invariant mass will correspond to $m_H = \sqrt{2\lambda}v$, so by looking in the range predicted by the Standard Model, $110 \leq m_H \leq 150\ GeV$, the mass of the Higgs particle could be found. Both the CMS and ATLAS experiments reported consistent results with CMS measuring $125.3\ \pm 0.6\ GeV$ and ATLAS measuring $126\ \pm 0.6\ GeV$. [6]

In this study, we will use the diphoton and four lepton decay channels to reconstruct invariant mass spectra and determine the mass of the Higgs boson, which is identified through a characteristic peak in the invariant mass spectrum. We seek to demonstrate the feasibility of independent Higgs research using publicly available data and provide valuable insights into the effectiveness of the CMS Open Data platform for future research in particle physics.

Methodology

VM {AA, CC, EE}

Following the guidelines [5] provided by CMS, we investigated different methods of processing the datasets. A method we experimented with was using CMS-specific CERN Virtual Machine (VM), which should provide a pre-configured environment that is compatible with the ROOT framework. However, this processing method did not work in the group project due to several possible reasons. Different system configurations may have led to inconsistency in the VM, causing difficulties in running the datasets. Also, the VM lacked the required CMS shell icon on its desktop, forcing reliance on an X terminal emulator incompatible with the CMSSW software. Attempts to manually mount CVMFS required administrative privileges were unsuccessful due to password restrictions.

AWS {AA, CC, EE}

One method we used to analyse the data from the CMS experiment was Amazon Web Service (AWS). This method, while time-consuming, proved to be effective for data processing. In order to apply this service, an S3 bucket was used to store the required data files downloaded from the CERN Open Data repository, and a virtual server was set up using an EC2 instance. The instance was configured with appropriate security settings to only allow certain traffic from allowed devices and sites. Once this was completed, Ubuntu, a Linux based operating system, was installed along with other required packages, such as ROOT and python. Ubuntu was chosen for its ease of use, stability, and ability to manage packages. To begin the analysis, the chosen data file was copied from the bucket to the instance where it was unpacked and processed using the ROOT framework. The plots could then be generated using python. The issue with this method is that the total time is far greater than with Docker as everything is done manually, so in the end that method came out on top.

We analysed simulated and real proton-proton collision datasets from the Open Data repository to validate the CMS findings and rediscover the Higgs boson.

Docker {BB, DD}

To ensure a consistent and isolated analysis environment, Docker was first installed. Following CERN's official guidelines [5], the python-vnc container image (gitlab-registry.cern.ch/cms-cloud/python-vnc) was pulled from the CERN Github registry. Necessary libraries, including uproot, awkward, fsspec, XRooTD, were installed via the terminal. After launching the container, the Jupyter Notebook server was set up as the workspace, enabling interactive code development, documentation, and visualisation. Port 8888 was exposed for browser access. Then, via XRooTD function, the dataset from Open Data could be accessed remotely.

Due to variations in naming conventions across different decay channels and dataset periods, slight adjustments in search mechanisms were employed.

For $H \rightarrow ZZ^{*} \rightarrow 4l$ decay channel, the simulated datasets '***Simulated dataset SMHiggsToZZTo4L_M-125_7TeV powheg15-JHUgenV3-pythia6***' (299683 events, 54.2GiB) [10] and '***Simulated dataset SMHiggsToZZTo4L_M-125_8TeV-powheg15-JHUgenV3-pythia6 in AODSIM format for 2012 collision data***' (299973 events, 98.9GiB) [11] were chosen. The ROOT files were accessed via the uproot library {BB}.
Within the 'Event' tree, the following branches were selected for muons:
'recoMuons_muons__RECO.obj.pt_' (p_T),
'recoMuons_muons__RECO.obj.eta_' (η),
'recoMuons_muons__RECO.obj.phi_' (ϕ),
'recoMuons_muons__RECO.obj.mass_' (m).
Similarly, the corresponding branches for electrons were selected:
'recoGsfElectrons_gsfElectrons__RECO.obj.pt_' (p_T),
'recoGsfElectrons_gsfElectrons__RECO.obj.eta_' (η),
'recoGsfElectrons_gsfElectrons__RECO.obj.phi_' (ϕ),
'recoGsfElectrons_gsfElectrons__RECO.obj.mass_' (m).
For $H \rightarrow ZZ^{*} \rightarrow 4\mu$ and $H \rightarrow ZZ^{*} \rightarrow 4e$ decay cannels, the events with 4 muons and 4 electrons were selected respectively, and for $H \rightarrow ZZ^{*} \rightarrow 2e2\mu$ decay, the events with both 2 muons and 2 electrons were selected. Then, the invariant masses M_{4e}, $M_{4\mu}$ and M_{2e2u} were calculated using equations (1) to (5).

For the $H \rightarrow \gamma\gamma$ decay channel, the relevant simulated diphoton datasets were then obtained from the CERN Open Data Portal {DD}. Specifically, datasets corresponding to the gluon-gluon fusion Higgs production process (GluGluToHToGG) were selected for 2011, 2012 and 2015. The 2011 simulated dataset ***'GluGluToHToGG_M-125_7TeV-powheg-pythia6'*** (AODSIM format, containing 99988 events, 19.9 GiB), the 2012 simulated dataset

‘***GluGluHToGG_M-125_8TeV-pythia6’*** (AODSIM, containing 99912 events, 27.5 GiB) and the 2015 simulated dataset ‘***GluGluHToGG_M125_13TeV_amcatnloFXFX_pythia8’*** (MINIAODSIM, containing 812652 events, 18.1 GiB) contain reconstructed proton-proton collision events stored as ROOT files. Each event records photon candidate properties, including p_T, η, φ, and m (assumed to be zero for photons). The data is stored under the branch names ***‘recoPhotons_photons__RECO.obj.pt_/eta_/phi_/mass_’*** for the 2011 and 2012 datasets and ***‘patPhotons_slimmedPhotons__PAT.obj.m_state.p4Polar_.fCoordinates.f Pt/fEta/fPhi/fM’*** for the 2015 dataset. Events with fewer than two photons were excluded to ensure valid diphoton mass calculations. The invariant mass of each diphoton pair was then computed using the same process utilising equations (1) through (5). A complete diphoton invariant mass distribution was produced for each dataset by iterating over all events.

For 2011 Double Electron collision data, the dataset ‘***DoubleElectron primary dataset in AOD format from RunA of 2011’*** (49693737 events, 6.1TiB) was chosen {BB, CC, DD}. The same electron branches as those used in the $H \rightarrow \mathrm{ZZ}^* \rightarrow 4l$ simulated dataset were chosen. Then, electrons satisfying $p_T > 7\ GeV$ and $|\eta| < 2.5$ were then selected. The invariant mass, M_{4l}, was calculated through Equations (1) to (5).

Results

Simulated data

Histograms were plotted to provide a clear view of the invariant mass distributions. These histograms aggregate the data into uniform mass intervals, highlighting the overall shapes and facilitating the identification of potential peaks, particularly around 125 GeV. The x-axis represents the invariant mass in GeV, while the y-axis shows the event count for each mass interval. Vertical error bars were calculated using the Poisson distribution, based on the square root of the event count in each bin, reflecting the statistical uncertainty. Horizontal error bars were derived from the standard deviation of the Gaussian fit applied to the data around the Higgs mass region. For the datasets, Gaussian fits were applied to zoomed-in regions, around 100–150 GeV for $H \rightarrow \gamma\gamma$ decay and 80-150 GeV for $H \rightarrow \mathrm{ZZ}^* \rightarrow 4l$ decay, where the Higgs boson signal is expected.

Figures 1, 2, and 3 depict the distributions of four-lepton invariant masses for 2011 simulated data. For comparison, the background from the Standard Model ZZ^*process is represented by shaded areas. The red lines of the figures depict distinct peaks around 125GeV, consistent with the simulated mass value. The peaks values obtained from the Gaussian fit for $H \rightarrow \mathrm{ZZ} \rightarrow 4\mu$, $H \rightarrow \mathrm{ZZ} \rightarrow 4e$, and $H \rightarrow \mathrm{ZZ} \rightarrow 2e2\mu$ channels are (124.95 ± 0.12) GeV, (123.76 ± 0.16) GeV, and (124.58 ± 0.18) GeV, respectively. Figures 4, 5, and 6 depict the distributions of four-lepton invariant masses for 2012 simulated data, and shaded areas also represent the backgrounds. The peak values obtained from the fit for $H \rightarrow ZZ \rightarrow 4\mu$, $H \rightarrow ZZ \rightarrow 4e$, and $H \rightarrow ZZ \rightarrow 2e2\mu$ channels are (125.75 ± 0.60) GeV, (125.29 ± 0.28) GeV, and (126.74 ± 0.71) GeV, respectively. Comparing the histograms for two years, the number of events for all three channels was higher in 2012 than in 2011. This is due to the fact that the 8TeV dataset collected in 2012 was at a higher integrated luminosity of 20 fb^{-1}, whereas the 7TeV dataset collected in 2011 was at an integrated luminosity of 5fb^{-1}.

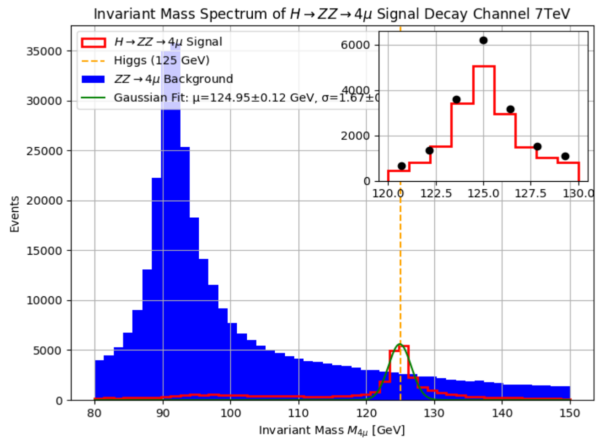


Figure 1

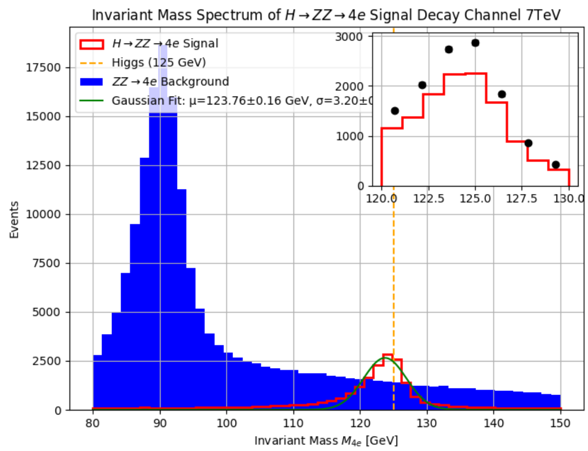


Figure 2

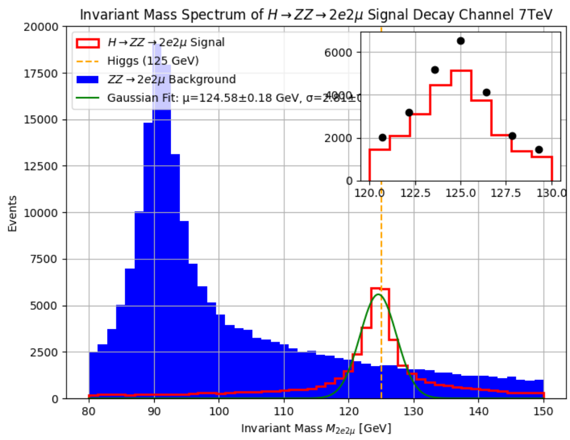


Figure 3

Figures.1, 2, and 3. The invariant masses, $M_{4\mu}, M_{4e}, M_{2e2\mu}$, distribution. The red lines depict the signals extracted from the 2011 Higgs Bosons decays; the filled histogram represents the corresponding backgrounds from 2011 SM ZZ^* process.

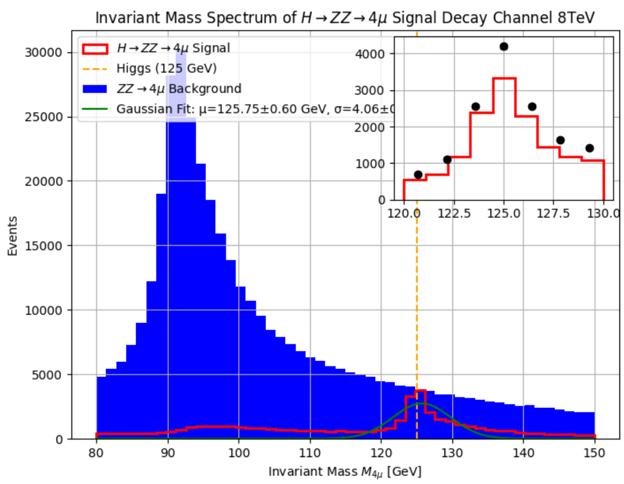


Figure 5

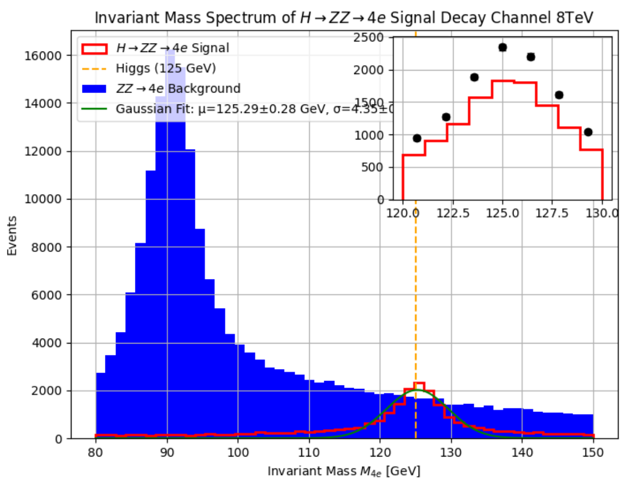


Figure 4

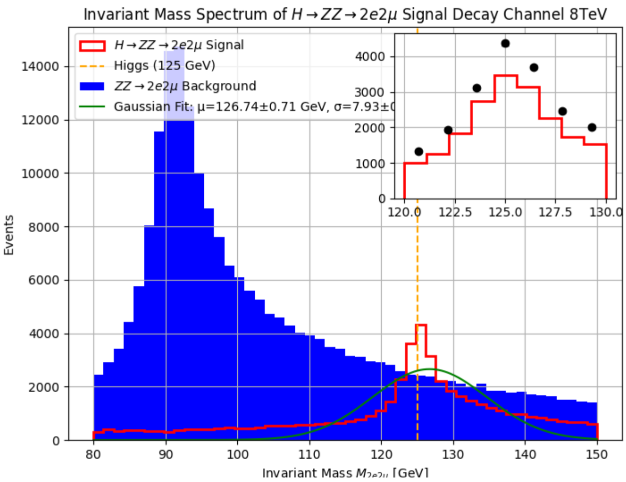


Figure 6

Figures.4, 5, and 6. The invariant masses, $M_{4\mu}, M_{4e}, M_{2e2\mu}$, distribution. The red lines depict the signals extracted from the 2012 Higgs Bosons decays; the filled histogram represents the corresponding backgrounds from 2012 SM ZZ^* process

For the 2011, 2012, and 2015 $H \to \gamma\gamma$ datasets, Gaussian fits were performed on the diphoton invariant mass spectrum within the 100-150 GeV range, where the Higgs boson signal is expected. The peak positions were (125.09 ± 0.15) GeV for 2011, (124.43 ± 0.16 GeV) for 2012, and (124.71 ± 0.09) GeV for 2015. These values align well with the expected Higgs boson mass of approximately 125 GeV, confirming the presence of a peak in the diphoton mass spectrum across all three years. The slight variations observed can be attributed to differences in condition filters. This analysis excluded events with fewer than two photons, but no restrictions were placed on other photon properties, such as transverse momentum. Introducing a minimum transverse momentum requirement could refine the selection, reducing background noise and yielding results closer to the expected 125 GeV.

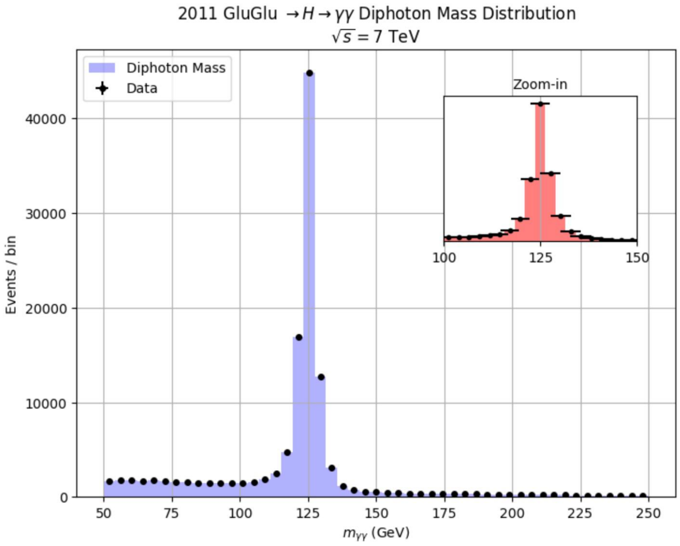


Figure 7

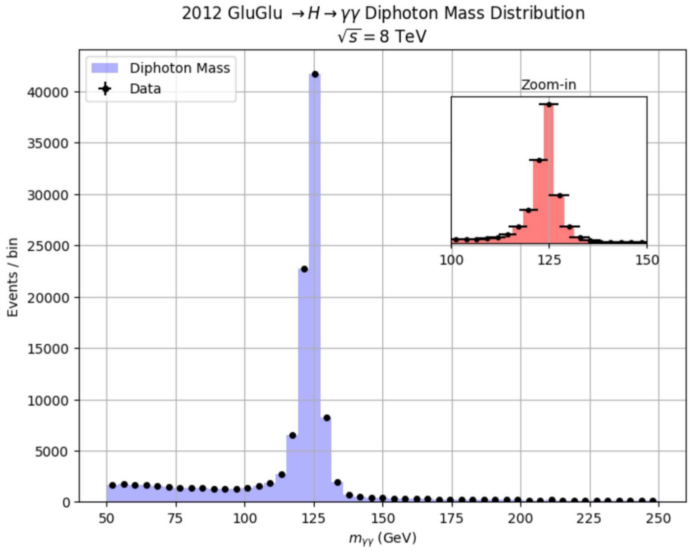


Figure 8

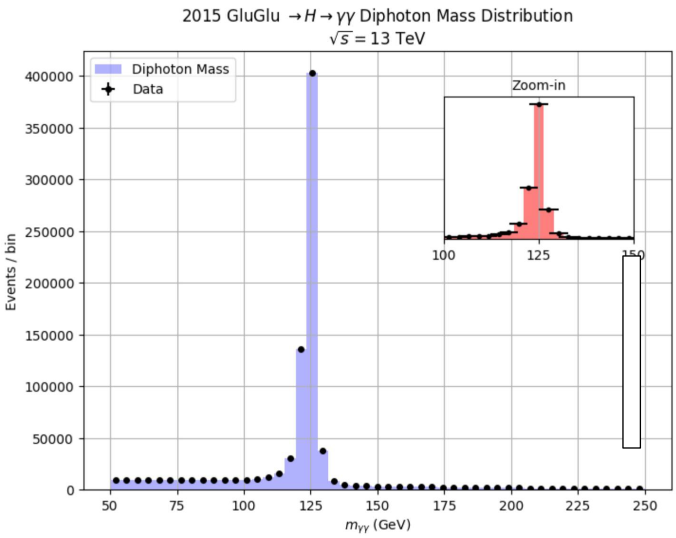
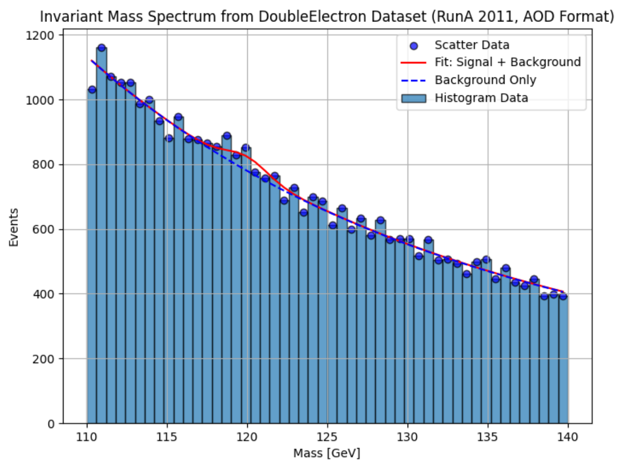


Figure 9

Figures.7, 8, and 9. Diphoton invariant mass distribution from 2011, 2012, 2015 dataset presented as histograms

Real data

In the analysis of 2011 Double-Electron collision data, the histogram for the invariant masses of the final states were drawn from 110 GeV to 140 GeV with an equal bin width of 0.6 GeV, as shown in Figure.10. The blue curve represents the background, which is fitted by power functions. The red line is the combination of background and signal, where signal is fitted by the Crystal Ball function. The red curve depicts a decreasing trend in the number of events as the invariant mass increases. Around 120 GeV, there is a bump where a slight excess of background is observed, indicating a potential candidate for the Higgs boson. Figure.11 shows the invariant mass spectrum after background subtraction, highlighting the signal contribution more clearly. From the fit, the reconstructed mass of the Higgs boson is (120 ± 0.66) GeV, which deviates from the theoretical value by $7.60\ \sigma$.



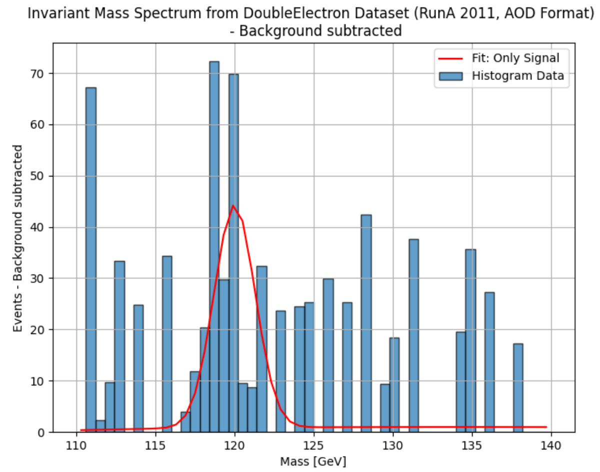


Figure.10. The shaded histogram represents invariant mass distribution of 2011 Double-Electron collision dataset. The blue line is the fitting background, and the red lines is the combination of signal and background.

Figure.11. The same dataset with Fig.10. where the background is now subtracted. The red line depicts the fitting signals.

Conclusion

For the 2011 Double Electron collision data, the final result deviates by $7.60\,\sigma$ from the expected mass of the Higgs boson. The primary reason for this discrepancy is the limitation in the dataset selection. The dataset contains events with more than two high-energy electrons, meaning that the final states are predominantly from the $H \to ZZ^* \to 4e$ decay channel. The absence of 4μ and $2e2\mu$ channels significantly reduces the overall event statistics and introduces bias in the final state.

Several improvements can be made for other channels in the project. Firstly, employing different fitting methods and ranges could provide valuable insights. Comparing the results obtained from alternative fitting methods and ranges, such as the Crystal Ball, fitting with a narrower range would help expand the scope of the investigation and offer a more comprehensive analysis. Machine learning techniques could be explored to increase the event classification efficiency of data analysis, allowing for the examination of more entries, the inclusion of additional datasets, and improve signal-to-background discrimination. For instance, a Boosted Decision Tree (BDT) [12] could be applied and trained to identify Higgs boson related events by leveraging results and patterns derived from simulated data. Also, unsupervised anomaly detection algorithms like k-nearest neighbours (k-NN) [13] and Isolation Forest algorithms [14] could be implemented to achieve a higher signal-to-noise ratio.

The methodological insights and analysis pipelines developed in this study serve as valuable references for future analyses involving other challenging decay channels, such as $H \to b\bar{b}$ particularly for data extraction and analysis. Having tested the reliability and usability of the CMS Open data sets, some improvements to the site can be made to enhance the user experience. Clearer labelling and consistent naming of files can be implemented to easier search for a particular set of collision data. Pre-configured machine learning tools can be embedded directly for data processing as well, to aid independent researchers with insufficient computational power. Despite that, the high quality and accessibility of the CMS Open Data repository makes it well-suited for educational purposes and can be applied to other research areas.

References

[1] Chatrchyan, S., et al. (2012). Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. Physics Letters B, 716(1), 30–61. https://doi.org/10.1016/j.physletb.2012.08.021

[2] CERN Open Data. (n.d.). Retrieved from https://opendata.cern.ch/.

[3] Jomhari, N. Z., Geiser, A., & Bin Anuar, A. A. (2017). Higgs-to-four-lepton analysis example using 2011-2012 data. CERN Open Data Portal. DOI:10.7483/OPENDATA.CMS.JKB8.RR42. https://opendata.cern.ch/record/5500

[4] Tripathee, A., et al. (2017). Phys. Rev. D 96, 074003. DOI: https://doi.org/10.1103/PhysRevD.96.074003

[5] CMS Guide to Research Use of CMS Open Data. (n.d.). Retrieved from https://opendata.cern.ch/docs/cms-guide-for-research

[6] Thomson M. Modern Particle Physics. Cambridge: Cambridge University Press; 2013. p.460–497

[7] CMS Collaboration, "Simulated dataset GluGluToHToGG_M-125_7TeV-powheg-pythia6 in AODSIM format for 2011 collision data," CERN Open Data Portal, 2018. [Online]. Available: https://opendata.cern.ch/record/1681. [Accessed: Mar. 6, 2025].

[8] CMS Collaboration, "Simulated dataset GluGluHToGG_M-125_8TeV-pythia6 in AODSIM format for 2012 collision data," CERN Open Data Portal, 2019. [Online]. Available: https://opendata.cern.ch/record/7782. [Accessed: Mar. 6, 2025].

[9] CMS Collaboration, "Simulated dataset GluGluHToGG_M125_13TeV_amcatnloFXFX_pythia8 in MINIAODSIM format for 2015 collision data," CERN Open Data Portal, 2021. [Online]. Available: https://opendata.cern.ch/record/16815. [Accessed: Mar. 6, 2025].

[10] CMS collaboration. "/SMHiggsToZZTo4L_M-125_7TeV-powheg15-JHUgenV3-pythia6/Summer11LegDR-PU_S13_START53_LV6-v1/AODSIM". [online] Cern.ch. Available at: https://opendata.cern.ch/record/1507 [Accessed 10 Mar. 2025].

[11] CMS collaboration (2025). "/SMHiggsToZZTo4L_M-125_8TeV-powheg15-JHUgenV3-pythia6/Summer12_DR53X-PU_S10_START53_V19-v1/AODSIM". [online] Cern.ch. Available at: https://opendata.cern.ch/record/9356 [Accessed 10 Mar. 2025].

[12] T. Chen and T. He, "Higgs Boson Discovery with Boosted Trees," Neural Information Processing Systems, vol. 42, pp. 69–80, Jan. 2014.

[13] D. Diaz et al., "Improving Di-Higgs Sensitivity at Future Colliders in Hadronic Final States with Machine Learning," ArXiv (Cornell University), Jan. 2022, doi: https://doi.org/10.48550/arxiv.2203.07353.

[14] L. Jiang, Y.-C. Guo, and J.-C. Yang, "Detecting anomalous quartic gauge couplings using the isolation forest machine learning algorithm," Physical review. D/Physical review. D., vol. 104, no. 3, Aug. 2021, doi: https://doi.org/10.1103/physrevd.104.035021.

[15] CERN Press Release, CERN announces new open data policy in support of open science, 12/2020 [Accessed: 03/2025], Available from: https://library-guides.ucl.ac.uk/referencing-plagiarism/vancouver#s-lg-box-wrapper-18252586

[16] CERN. CMS [online]. CERN: Geneva; [Accessed 03/2025]. Available from: https://home.cern/science/experiments/cms

[17] CERN, CMS. Detector [online]. CERN: Geneva; [Accessed 03/2025]. Available from: https://cms.cern/detector

Appendices

Agenda and Minutes

Meeting	Members present	Agenda items	Summary	Actions	Board member comments
14/01/25	All except [DD]	Role Allocation - Chair, communication with Board member: [EE] - Minutes: [AA] Look into papers for literature review - CMS - ATLAS - Other experiments - Textbook Planning and organisation - Set up computational environment - Subgroups to work on different parts (Dividing literature review and computational analysis)	Roles have been allocated to group members, and tasks have been identified and listed.	Start reading Have tasks allocated	It is good to have the workload distributed amongst the members.
21/01/25	All	Create Timeline Start understanding the theory Allocation: [AA], [EE] investigate Docker environment [BB], [CC], [DD] look at the literature review.	A project timeline was created, with a clear allocation of tasks.	Have the allocated tasks in item 3 completed.	
30/01/25	All	Docker issues - Problems running the data - Try Virtual Machine Distribute analysis Start with a small fraction and understand what is happening - Move on to larger data sets after understanding Improve efficiency - Run on all our laptops - High-power computing system Lit review - Coming along - Look at the original Higgs paper		Readings - Ref 3, 4 - Extra reading [AA] & [CC]: Virtual Machine [EE] & [BB]: Selection cuts and plotting, [DD]continuing with Docker	Recommended reading: - Textbooks: Modern Particle Physics, Mark Thomson (2013), CMS detector, LHC physics
04/02/25	All	Issues running data with VM and Docker [BB] had success with graph acquired on the mass spectrum Look into other platforms for running data - AWS - Google platforms	Other computational services were considered and experimented with.		Suggested use of SWAN
11/02/25	All	AWS - How to pay for		[AA]: Set up AWS/Look into Google & Microsoft platforms	

		- Reimbursement - Start with the free version Google & Microsoft - Tutorial for Azure and CERN open data - SWAN Check the consistency of results - Comparison to published data [BB] & [DD]: mass spectrum, slightly different graphs - Data filtering [EE]: Docker is working but only producing half a plot		References [EE], [BB], [DD] - Continue with Docker, fix issues - Virtual Machine, [CC] need to fix issues with missing CMSSW environment	
20/02/25	All	Progress with AWS Issues with Docker and VM - Asking for a password - Issues with container [BB] got results from simulation		2011-12 & 15-16 Photon – photon and 4 leptons simulation data done -[AA] look at W boson simulated data Start planning report(s) - Make a start on the group introduction [EE]	-Missing CMSSW environment of VM due to missing CMS shell icon on the desktop
27/02/25	All	Need more Financial Proposal Limited data space on personal laptops (include in the write-up) Create a timeline for write-up Individual Report - Give background, but mainly focus on the work YOU have done and how you have personally helped the group achieve the goals Proofread individual reports (future)		All data from ref 3 analysed Sim data for the 2b quark - Challenging - Pair is produced frequently (high background) Finance request - [EE] completed writing	
04/03/25	All	[BB] & [AA] working on the first link - [BB] working on page 1 - [AA] from page 10 [DD] completed 4 lepton sim data for 2011, 12, 15 [DD] now working on double photon		Stop data analysis on 14/3	
12/03/25	All	[BB] has produced some good results from real data - Using branches (applied filters to generate graphs) - ~121 GeV (+-2 GeV) *is the deviation large? (applied too many restrictions?) - small deviations (can be refined by applying different restrictions/filters?) - Will be using the same data			Write more on the sections and worry about editing the document later

		with different parameters to refine the results [CC] has worked on the report - Looked at equations (background) [DD] has produced some results from real data as well - No obvious peak - No fitted graph - Will apply filters and restrictions on the graphs - *[DD] and [BB] are working on the same data, different parts -[CC] is working on the other set of real data			

Risk Assessment

Event	Hazard	Risk	Controls	Likelihood	Impact
Participation in group meetings.	Group members break into arguments during discussion.	May harbour animosity and acrimony among members.	Agree to be professional and polite during meetings and discussions.	Unlikely.	Heated arguments may cause delays in completing goals within the project timeline.
Participation in group meetings.	Group members/members not showing up to meetings.	May harbour animosity and acrimony among members.	Set timetabled sessions for the group to discuss the process, allocating the next meeting during meetings.	Unlikely	Not all members showing up may lead to uneven workload distribution and delays in the completion of tasks.
Allocation of tasks amongst group members.	Uneven workload amongst group members.	Large work burden on members/members.	Discuss the allocation of tasks in meetings and have all members agree on the uptake of tasks.	Unlikely	Uneven workload distribution may cause arguments between members and delay task completion.
Conducting computational analysis.	Datasets are too large for analysis.	Laptops/Computers are unable to run analyses of the data sets.	Agree to use external computational services if there is a need to.	Likely	Unable to analyze the CMS data sets.