FYS5555 Project 2

Search for the Higgs boson decaying to 4 leptons in ATLAS open data

William Hirst

Spring 2022

University of Oslo

In this rapport I will be studying the higgs decay in a proton proton collision. Specifically I will be studying the higgs decay into 4-leptons. I will be recreating results published by Atlas in both 2012 and 2020, displaying the invariant mass, transverse momentum and energy of the 4-leptons. In addition we will be attempting to search for the effect of the higgs in the histograms we create.

1 Higgs boson decay

In this report I will be studying the decay of the higgs boson using the ATLAS open data, Monte-Carlo simulations and the root framework. More specifically we will be studying the higgs decay in the proton-proton collision at the LHC. At the LHC, the higgs can decay through the following processes; $b\bar{b}$, w^+w^-, gg, $\tau\bar{\tau}$, $c\bar{c}$, $ZZ^{(*)}$, $\gamma\gamma$, $Z+\gamma$ and others. The process most commonly detected, is the higgs into two photons ($H \to \gamma\gamma$). As far as higgs production there are also many channels, like; $gg \to H$, $qq \to qqH$, $q\bar{q} \to WH$ and others. The most common way is the gluon fusion through a heavy quark loop.

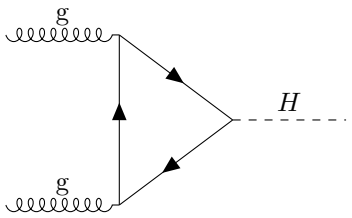


Figure 1: Feynman diagram of higgs production through a gluon fusion and a heavy quark loop.

From this large range of decay there are potentially many higgs bosons to be found from the data recorded at the LHC. But, given the challenge of separating the higgs from all the other data, certain processes must be chosen. Given that $H \to ZZ^*$ is one of the most sensitive channels, this will be the main focus of this rapport. In a proton-proton collision the decay of a higgs into two Z-bosons will quickly follow by a decay into 4-leptons ($H \to ZZ^* \to l^+l^-l^+l^-$). These four leptons will be two pairs of either e^+e^- or $\mu^+\mu^-$ or a combination (SFOS), and thereby discarding neutrinos and τ from the analysis.

All results will be compared to results from the articles written by the the ATLAS Collaboration from 2012 [2] and with the article written in 2020 [4]. All calculations in this rapport from both simulations and data were collected with a luminosity of 1064 fb^{-1} and $\sqrt{s} = 13TeV$. Therefore most comparisons will be done with the latter article given that the article worked with a similar luminosity.

2 Contribution Feynman diagrams

In the analysis I have chosen out a set of channels which are allowed to contribute. The channels are split in to two categories; signal and background. For the signal events in this analysis correspond to the ones containing the higgs boson. In figure (2) I have drawn the Feynman diagram corresponding to the signal event. In the diagram we see a higgs boson decaying to a pair of Z-bosons, where one is on-shell and the other is off.

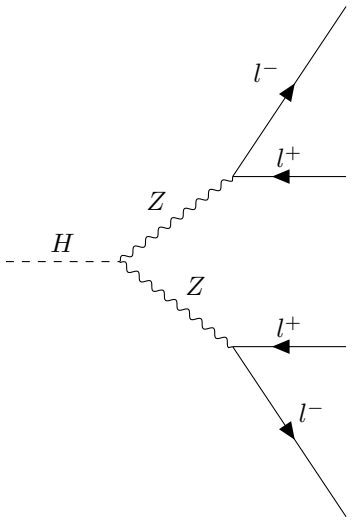


Figure 2: The Feynman diagram for the signal channel of H

For the background events I have chosen out three channels. The first and largest background is the diboson channel. The Feynman diagram is drawn in figure (3). In the diagram we see a quark and and an anti-quark forming a pair of Z-boson. Either one or both of the Z bosons can be on- or off-shell. As in the signal case both Z-bosons quickly decay into pairs of leptons.

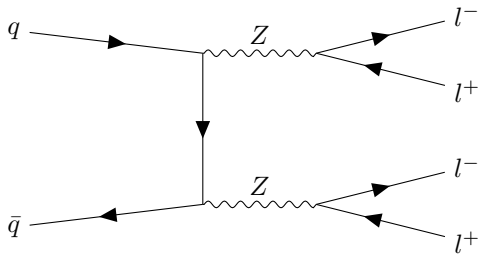


Figure 3: The Feynman diagram for the background diboson background channel $ZZ^{(*)}$.

The remaining channels I have grouped into the categories "others", similarly to the 2020 article [2]. These channels are $t\bar{t}$ and Z-jets. Both these channels can introduce one or more fake leptons into the quadruple. Below I have drawn two examples of this. The first diagram is drawn in figure (4). In the diagram we find a quark-anti quark pair created from a gluon pair(alternatively it could be a gluon decaying into a $t\bar{t}$-pair). The quark pair quickly annihilates into a W-boson and a quark/anti quark. The W-boson will then decay into a lepton with some missing energy. The remaining jets will then act as a pair of fake leptons, which "tricks" the MC calculations during the reconstruction.

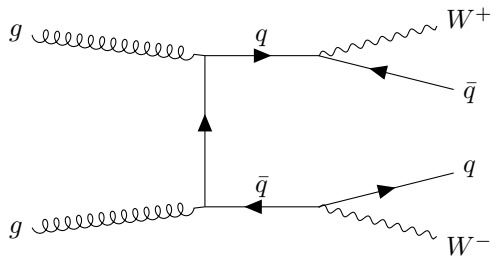


Figure 4: The Feynman diagram for the background channel-$t\bar{t}$.

The second channel in the "others" category is Z+jets-channel. In this channel a pair of gluons annihilate to form a quark, anti quark-pair connected by a third quark. The third quark then emits a Z-boson which annihilates to form a lepton pair. Again the first quark, anti quark-pair acts as a pair of fake leptons.

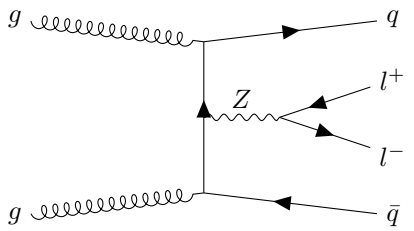


Figure 5: The Feynman diagram for the background channel Z + jets.

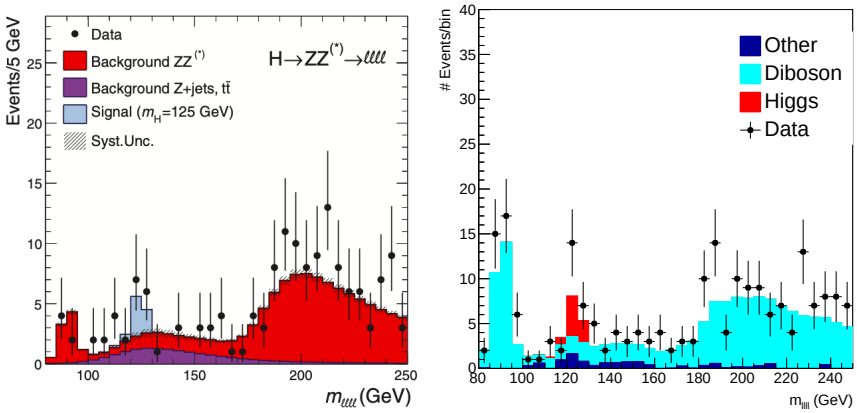
3 Cuts

When running calculations for this rapport there were several cuts needed in order to complete the analysis. Cuts are simply thresholds and limits put on the different variables to decide which events shall be included in the analyses and which shall be ignored, either for the purpose of specific analysis at hand or for realism. In this way we hope to best replicate results from the 2012 and 2020 Atlas articles, [2] and [4] and to preform an as realistic analysis as possible.

To begin with I started with the default DileptonAnalysis notebook. Therefore the first task was to alter from a dilepton to a 4-lepton analysis. To do so I simply cut all processes with more or less than 4 leptons. Secondly I had to ensure that all dilepton pairs were either a e^+e^- or $\mu^+\mu^-$. This was done by checking that the sum of the lepton-type was one of the possible combinations $[4 \cdot 11,\ 4 \cdot 13,\ 2 \cdot 11 + 2 \cdot 13]$ ($11 = e^{\pm}$ and $13 = \mu^{\pm}$).
Now that we have ensured a SFOS 4-lepton event, we can focus on the kinematic cuts needed to replicate realistic results, i.e finding the "good leptons". All kinematic cuts used in the analysis are found in table 1 and 3 in [4] as well as page 20.
To begin with we choose out only the events where the momentum is high enough. These thresholds were chosen to be 25GeV, 15GeV, 10GeV and 7 GeV for the particles with the first, second, third and forth largest transverse momentum respectively. Next we add a cut to the track isolation variables, P_t^{cone} and E_t^{cone}. Here we use the loose criteria described in table 3 in [4], 0.15. But, after further examination I found that this alone did not give the right results. Instead I needed a larger threshold for the particles with the second, third and fourth highest transverse momentum. The ideal threshold was found to be 0.30. An explanation for this is found in the article [3] (page 35). In the article published by Atlas it is found that the threshold shrinks for particles with higher transverse momentum.
Additionally we need to cut for a realistic track. The cut is done through the variable, η or pseudo rapidity. All electrons are therefore required to have an $\eta < 2.47$ and all muons must have an $\eta < 2.5$. A final cut is done to restrict the distance of closest approach to the primary vertex, z_0 to a realistic range. We do this by demanding that $z_0 \sin(\theta) < 0.5mm$, where θ is the polar angle.
In addition to the cuts above I also needed to choose which channels should be allowed in the analysis. This was done by studying the background text file in the input folder in the jupyter notebook. The channels chosen for this rapport were the channels described in [4] (page. 36).

4 Results

In this section we will study the results using the channels and cuts described in the sections above. To begin with we created a histogram of the invariant mass of the 4 leptons for the range of $m_{llll} \in [80, 250]$GeV, to compare with the histogram in Atlas article from 2012 [2].

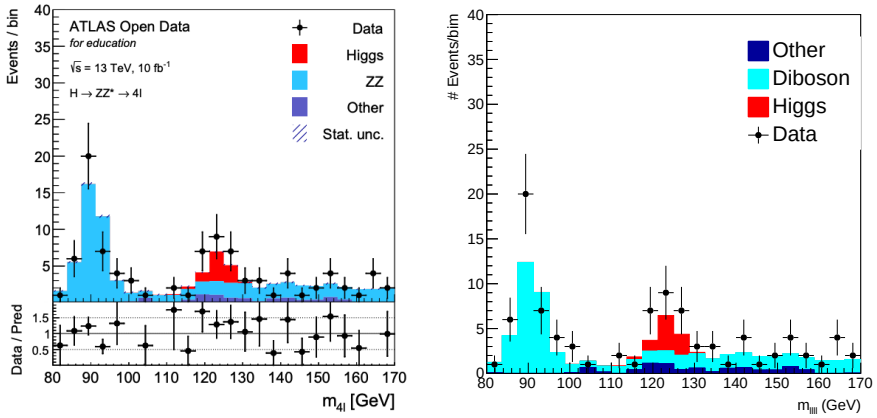


(a) Atlas open data review

(b) My computations

Figure 6: Histograms of the invariant mass distribution for the 4-lepton state in the interval $m_{llll} \in [80, 250]$GeV. Others correspond to $t\bar{t}$ and Z-jets.

In figure (10) we observe three peaks; one around 90GeV, one around 125 GeV and one around 200 GeV. This is the case for both histograms. The first histogram is from the case of one on-shell and one off-shell Z-boson decaying through the diboson channel. The off-shell Z-boson has a relatively small mass, and the peak therefore resonates around the mass of the Z-boson. The second peak shows a clear contribution from the higgs channel, which is expected due to the peak laying in the area of resonance for this channel.
Lastly there is a wide peak around the mass of two Z-bosons. This comes from the diboson-channel and is the case when both bosons are on-shell. This peak is far wider than the other two peaks. This could be explained by the fact that in the first two peaks, only one would be on-shell so there would not be great variation around the peaks. In the last peak, both bosons are on-shell and can be larger than the mass of resonance, leading to greater variation. Regardless this shows that most of the 4-lepton events go through the diboson-channels, and also that there is more often two on-shell bosons than one.
As far as the accuracy of trying to recreate the figure from 2012 article (6a), there are some deviations. But, this is most likely a consequence of the difference in luminosity. We see that the the 1 on-shell and the higgs peak are the ones that are mostly effected. For accuracy between MC and data the general trend seems to fit well. The largest deviation is around the mass of the higgs, but we see that the contribution from the higgs-channel greatly improves the accuracy.

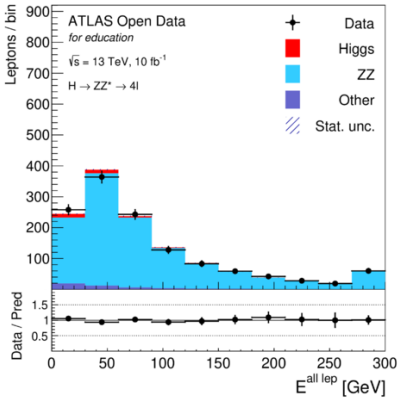


(a) Atlas open data review

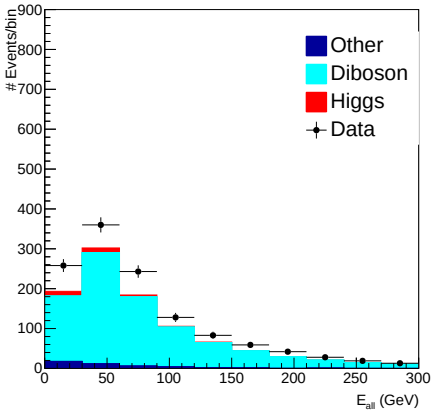
(b) My computations

Figure 7: Histograms of the invariant mass distribution for the 4-lepton state in the interval $m_{llll} \in [80, 170]$GeV. Others correspond to $t\bar{t}$ and Z-jets.

Next I changed the range of the mass to better compare to the histograms in the 2020 article. Again we want to draw the invariant mass distribution, but this time in the range of $m_{llll} \in [80, 170]$GeV. Using this range we see a far better resemblance to the ATLAS figure. The only real deviation is around the Z-mass,
Given the range of the histogram, the two on-shell peak is no longer included. Though we can better see that the higgs peak is isolated to a small interval around its mass. This is because we have defined the higgs contributions as channels where both dilepton pairs come directly from the higgs. Another interesting note is the symmetry around the peak. From the histogram it seems that there is no significant tail on either side.
In addition we can see how the higgs contribution perfectly fills the otherwise empty gap between the MC and the data. This is one of the thing that lead to the discovery of the higgs and its mass.



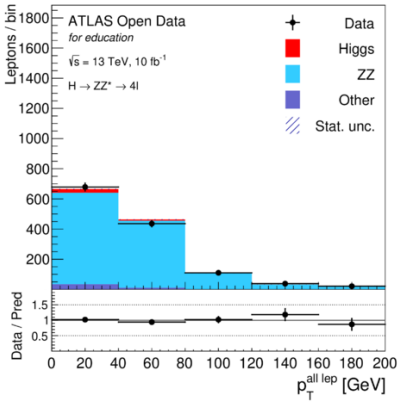
(a) Atlas open data review



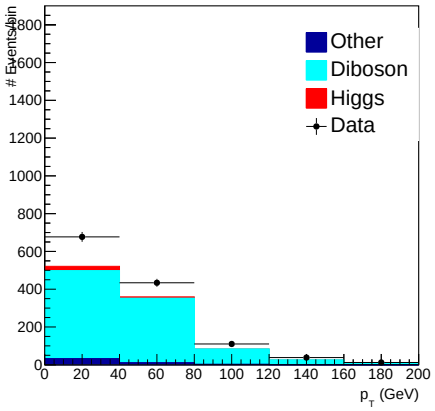
(b) My computations

Figure 8: Histograms of the distribution of energy for all four leptons. Others correspond to $t\bar{t}$ and Z-jets.

Then I created a histogram of the transverse momentum and energy of all the leptons. Figure (8) shows the energy of all of the leptons within the range of [0-300] GeV, where as figure (9) shows the transverse momentum. We can observe that the higgs yet again fills a gap in both the energy and transverse momentum histogram. Most of the higgs contribution is found in the lower bins, [0-100]Gev and [0-40]Gev for energy and momentum respectively.



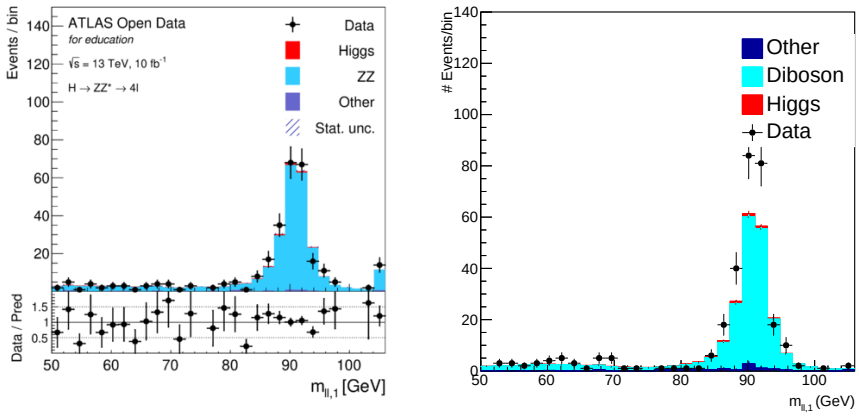
(a) Atlas open data review



(b) My computations

Figure 9: Histograms of the distribution of the transverse momentum for all four leptons. Others correspond to $t\bar{t}$ and Z-jets.

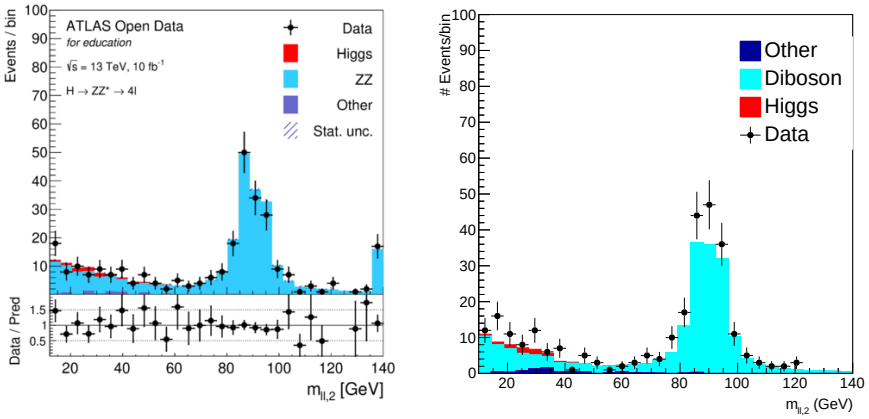
Finally two histograms were made to display the distribution of the dilepton mass of each pair. To separate the leptons in pairs there were two scenarios I needed to handle; one where all 4 leptons were same flavor and one where the two pairs were different flavours. When the pairs were different flavour I simply divided the pairs by which leptons shared the same flavour. In the case of equal flavour I first separated the leptons by charge, then found the pair with mass closest to the Z-mass.
In figure (10) we see a clear peak around the Z-mass, and an otherwise flat structure. In figure (11) we also see a peak around the Z-mass, but also with a rising trend towards the lower masses, peaking around 10-30GeV. The two largest peaks in figures (10) and (11) correspond to the two on-shell peak in figure (6). The largest peak in figure (10) also has a large contribution of higgs. This is explained by the smaller peak in figure (11), where we see that higgs also contributes. A combination of the two peaks make up the higgs peak found in (7). A final observation is that in both dilepton pairs, we see that the highest peaks are around the Z-mass, once again confirming the systems preference towards a two on-shell boson collision.



(a) Atlas open data review

(b) My computations

Figure 10: Histograms of the distribution mass for the first lepton pair, ie. the pair with the mass closest to m_Z. Others correspond to $t\bar{t}$ and Z-jets.



(a) Atlas open data review

(b) My computations

Figure 11: Histograms of the distribution mass for the second lepton pair, ie. the pair with the mass second closest to m_Z. Others correspond to $t\bar{t}$ and Z-jets.

Lastly we can note that there are small deviation from the Atlas figures and my computations. The deviations are around the largest possible mass in both figures. We can also see a clear difference in the shape of the peak in figure (11). Given the accuracy of the other figures, I believe this could be because of how I have chosen to divide the pairs. It seems that where Atlas have found a sufficiently large number of events in the high mass range, I have placed these in the Z-mass peak area. I suspect that the error could emerge from the case of equal flavour on all leptons. Perhaps Atlas used a better way of separating the leptons into pairs, taking advantage of some attribute of each lepton that I am not aware of.

5 Scaling the crossection

In the results presented in the section above I observed a systematic lack of event between the Atlas figures and the histograms I created. After all cuts were applied to the calculations I noticed a deviation between the diboson channel contribution in my results and those produced in the ATLAS article.
In figure (7b) I observed that there are fewer events around the mass of the Z-boson, where the diboson channel dominates. After reading several articles I finally discovered the article Cranmer et ai. [1][1]. Through reading the article I discovered that the data and the Monte Carlo calculations might differ in that the MC data does not include LO non-resonant channel $gg \to ZZ$. To correct for this and additional factor of 1.3 is added to the LO diboson channels

[1]Later I also discovered another example where such a scaling is done. The link shows a code preforming similar calculations, with the same data. Link is found here.

cross-section, as mention in the article (page.9). The Feynman diagram of the non-resonant contribution is found in figure (12).

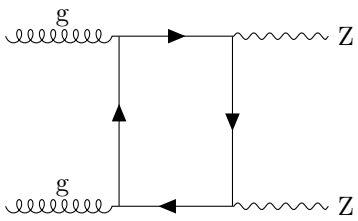
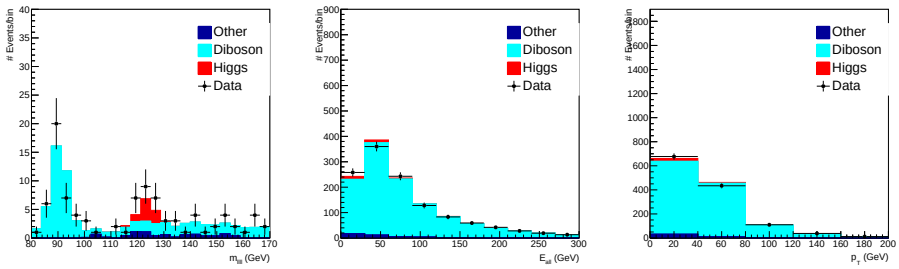
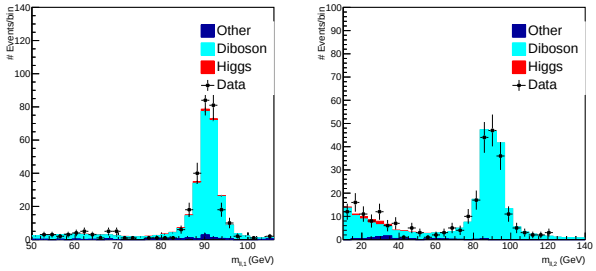


Figure 12: Feynman diagram for the non-resonant channel of $gg \to ZZ$.

In figure (13) I display all 5 of the histograms with the additional factor of 1.3. By comparing the results to the Atlas histograms in the figures above one observes a clear improvement on all histograms. Most of the histograms are practically identical to the Atlas figures. Why such are difference between the data and the MC simulation exist is unknown. There is also reason to believe that one can not perfectly add a channel to an analysis by uniformly multiplying a factor on a pre-existing channel. Because of this the factor was not added to the main results, but given further research one might find sufficient reasoning to why such a factor can be added.



(a) m_{llll} (b) E (c) P_t



(d) $m_{ll,1}$ (e) $m_{ll,2}$

Figure 13: Histograms of the mass of the 4-leptons (m_{llll}), the energy (E), the transverese momentum (P_t) and the mass of the first and second dilepton pair($m_{ll,1}$ and $m_{ll,2}$) with tha crossectionfactor of 1.3.

6 The analysis

From all histograms produced for the sake of this analysis we see that the higgs contribution improves the comparison between data and MC. In figure (14) I removed the contribution from the higgs in the MC calculations and we can see a clear deviation around the higgs mass. The benefit of the higgs contribution was also mentioned in the dilepton mass histograms. There we found the need for the higgs contribution around the Z-mass for the first pair and around 20-30Gev area.
A final test of the analysis is to calculate the significance of the signal using the equation of

$$Z = \sqrt{2\left[(s+b)ln\left(1+\frac{s}{b}\right)-s\right]}. \quad (1)$$

Using the control region [120-125] and the data points measured I find the rough estimate of the significance equalling 4.7. This was calculated using the bin sizes and limits of the figure (6), where I found ca. 14 signal events and 5 background. Though this is slightly bellow the threshold of discovery, 5 it is a result well worth the time and effort of further investigation.

I found that I was able to partly reproduce the figures from the articles of 2012 and almost perfectly reproduce the figures from the 2020 article. An even better comparison was found, although this required to multiply a uniform factor to the diboson-channel. The only slight deviation from the 2020 article was in the dibson mass histograms, where I suspect an error was made in the separation into dilepton pairs.

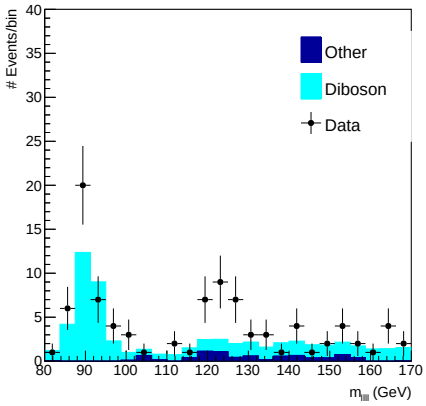


Figure 14: The invariant mass of the 4 leptons without higgs channel.

References

[1] Kyle Cranmer, Bruce Mellado, William Quayle, Sau Lan Wu. "Application of K Factors in the H $\rightarrow$ ZZ $\rightarrow$ 4l Analysis at the LHC." In: (2003). DOI: https://arxiv.org/pdf/hep-ph/0307242.pdf.

[2] The ATLAS Collaboration. "A Particle Consistent with the Higgs Boson Observed with the ATLAS Detector at the Large Hadron Collider." In: (2020). DOI: https://www.science.org/doi/full/10.1126/science.1232005.

[3] The ATLAS Collaboration. "Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton–proton collision data." In: (2019). DOI: https://arxiv.org/pdf/1908.00005.pdf.

[4] The ATLAS Collaboration. "Review of the 13 TeV ATLAS Open Data release." In: (2020). DOI: http://cds.cern.ch/record/2707171/files/ANA-OTRC-2019-01-PUB-updated.pdf?version=2.