Simplified Template Cross Section measurements of the $V(H \to b\bar{b})$ process with the ATLAS detector at $\sqrt{s} = 13$ TeV

Ryan Justin Atkin, Sahal Yacoob

Department of Physics, University of Cape Town, Private Bag X3, Rondebosch 7701, Cape Town, South Africa

E-mail: rjatkin93@gmail.com

Abstract. Presented are the studies of the production of the Standard Model Higgs boson in association with a W or a Z boson, where the Higgs boson decays to $b\bar{b}$ and the W/Z bosons decay leptonically. The data used is the full Run-2 ATLAS dataset, corresponding to 139 fb^{-1} of integrated luminosity, which was collected in proton-proton collisions at a centre of mass energy of $\sqrt{s} = 13$ TeV. At this energy, the $H \to b\bar{b}$ decay has a branching fraction of $\sim 58\%$, so this study allows the probing of the dominant Higgs boson decay mode. This production mode also provides the best sensitivity to the WH and ZH interactions and allows the study of the Higgs boson at high transverse momentum, both of which are important for the interpretation of the Higgs boson measurements in Effective Field Theories. The focus of the latest round of this analysis is to measure the cross-sections using the Simplified Template Cross Section method. Here, the cross sections are measured as a function of the W/Z boson transverse momentum in different fiducial volumes based on kinematic selections. In particular, the jet p_T selections are studied to see if any improvements can be achieved by tightening the selections.

1. Introduction

The dominant decay of the Higgs boson is to two bottom quarks ($H \to b\bar{b}$), with a branching fraction of $\sim 58\%$ at $\sqrt{s} = 13$ TeV [1]. However, when searching for these events with the dominant Higgs boson production mode of gluon-gluon fusion, the multi-jet backgrounds are too large. Therefore, if we instead look at the production of the Higgs boson in association with a vector boson V (W/Z boson), we can use the leptonic decays of the vector bosons to help reduce those backgrounds [2]. A Feynman diagram of this process is given in Figure 1. This production and decay mode provides the best sensitivity of the Higgs coupling to the W/Z bosons and allows the study of the Higgs boson at high transverse momentum. This sensitivity and high momentum are important for the interpretation of the Higgs boson measurements in Effective Field Theories (EFTs) [2]. Since b-hadrons are the only down-type hadrons that can be effectively identified, this decay mode also allows the study of the Yukawa coupling of the Higgs boson to the down-type quarks.

This paper will discuss the work going into improving the Simplified Template Cross Section (STXS) measurements in the $V(H \to b\bar{b})$ analysis by tightening the jet transverse momentum selections. The previous analysis [2] found no discrepancies in the cross sections between the Standard Model and the data. The ATLAS detector is described in Section 2 and the event

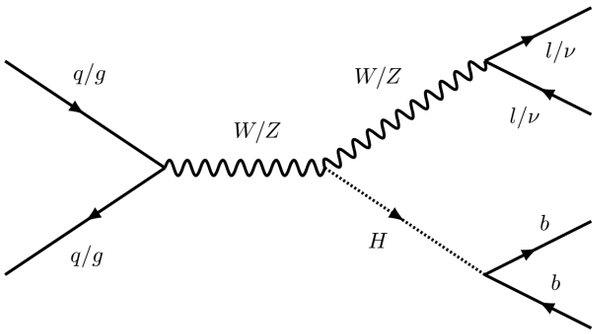

Figure 1. Feynman diagram for the $V(H \to b\bar{b})$ process. The W boson can only be produced by the quark initiated process due to charge conservation, while the Z boson can be produced via quark or gluon annihilations. The W/Z boson emits a Higgs boson which subsequently decays into a bottom anti-bottom quark pair. The Z boson then decays leptonically into two leptons or two neutrinos, while the W boson decays leptonically into a lepton and neutrino.

selections in Section 3. The STXS are explained in Section 4, with the new jet selections and results in Sections 5 and 6 respectively.

2. The ATLAS detector

The ATLAS detector[3], located at the Large Hadron Collider[4] at CERN, is an all-purpose particle detector with almost 4π coverage around the interaction point. The detector uses cylindrical coordinates (r, ϕ) in the transverse plane, with ϕ being the azimuthal angle around the z-axis. The pseudorapidity is defined as $\eta = -\ln(\tan\frac{\theta}{2})$, where θ is the polar angle and the distance in (η, ϕ) coordinates is defined as $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$. A convenient variable used is the transverse momentum $p_T = |p| \sin(\theta)$.

The detector is layered, with the inner layer being the Inner Detector (ID) used to track charged particles and calculate their momenta, and covers the pseudorapidity range $|\eta| < 2.5$. This is surrounded by a 2 T superconducting solenoid which bends the charged particles to allow for momenta and charge measurements. Following this are the electromagnetic and hadronic calorimeters used to measure the energy of the particles, covering the pseudorapidity range $|\eta| < 4.9$. The final layer is the Muon Spectrometer (MS) and superconducting toroidal magnets. The MS performs the same function as the ID but solely for muons, and the tracking covers the pseudorapidity range $|\eta| < 2.7$.

3. Event selections

The final objects for the $V(H \to b\bar{b})$ process contains leptons (e/μ) and two b-jets. For the Z boson production, there is either the 0-lepton decay $(Z \to \nu\nu)$ or the two lepton decay $(Z \to ll)$. The W boson only has a one lepton decay $(W^{\pm} \to l^{\pm}\nu)$. This analysis is split into three channels based on these leptonic decays of the vector bosons. For this paper, the focus will be on the 0-lepton channel. Since neutrinos are undetected in the ATLAS detector, a lepton veto is applied to the events to ensure we have no reconstructed leptons in the final state. The events are then split by the p_T of the vector boson (p_T^V), since the sensitivity is higher at larger p_T^V. In the 0-lepton channel, the missing transverse momentum (E_T^{miss}) is used for p_T^V as this accounts for the neutrino p_T. The resultant p_T^V regions are $150 < E_T^{miss} < 250$, $250 < E_T^{miss} < 400$ and $E_T^{miss} > 400$, with the split at 400 being new for this round of the analysis.

Events are then further split by the number of jets in the event, which are reconstructed using the anti-k_T algorithm[5] with a radius parameter of $R = 0.4$. For the 0-lepton channel, there

Table 1. Some of the main selections for the analysis in the zero lepton channel.

Selection	selection
Leptons	0 leptons
p_T^V regions	$150 < p_T^V < 250$
	$250 < p_T^V < 400$
	$p_T^V > 400$
Jet p_T	> 20 GeV for central jets ($\|\eta\| < 2.5$)
	> 30 GeV for forward jets ($2.5 < \|\eta\| < 4.5$)
b-jets	Exactly 2
Leading b-tagged jet p_T	> 45 GeV
Jet categories	Exactly 2 / Exactly 3 / Exactly 4

are the 2-jet, 3-jet and 4-jet regions, with 4-jet being new for this round of the analysis. The jets need to pass a p_T selection of 20 GeV if the jet is central ($|\eta| < 2.5$) or 30 GeV if forward ($2.5 < |\eta| < 4.5$). A neural-network based tagging algorithm (DL1r [6]) is applied to the central jets to identify the b-jets, and tuned to an average b-tagging efficiency of 70%. Exactly two b-tagged jets are required, with the addition of the highest p_T b-tagged jet having a p_T larger than 45 GeV. A summary of these selections is given in Table 1.

4. Simplified Template Cross Section

Usually when cross section measurements are performed, the differential cross section is calculated using the entire phase space possible within the detector. However, for this analysis a Simplified Template Cross Section (STXS)[7] measurement is performed which calculates the cross section in smaller predefined kinematic regions. This reduces the theoretical uncertainties directly folded into the measurement, and helps isolate certain BSM effects in each region. A diagram for the STXS regions of the $V(\rightarrow\text{leptons})H$ process is shown in Figure 2. These are the regions that could possibly be studied and are defined by the theorists. They are split by vector boson production, p_T^V and the number of additional jets above the two tagged jets. Due to experimental and statistical constraints, not all of these regions are used and some have to be merged. For example, in the previous round of the analysis [2], the five resultant regions were; WH with $150 < p_T^V < 250$ and $p_T^V > 250$, and ZH with $75 < p_T^V < 150$, $150 < p_T^V < 250$ and $p_T^V > 250$. In this round of the analysis, we are looking at splitting both the WH and ZH, $150 < p_T^V < 250$ regions into zero and greater than zero additional jets regions, for which the split can be seen in Figure 2, resulting in seven regions.

Important for this paper is that when the STXS theoretical calculations are performed, the particle-level b-jets are selected to have $p_T > 20$ GeV, while all other particle-level jets have $p_T > 30$ GeV. This is different to the reconstruction level selections which require all the central jets to have $p_T > 20$, and only the forward jets have $p_T > 30$ GeV.

5. Optimising jet selections

The focus of the previous round of this analysis was to maximise the overall VH significance. In this round, the focus is on improving the STXS measurements. As part of this improvement, we are looking at changing the jet p_T selections at reconstruction level, as shown in Table 1, to match more closely with the theoretical selections as described in Section 4. The two possible scenarios are to increase only the non-tagged central jet p_T cuts to 30 GeV, or to increase all central jet p_T cuts to 30 GeV. The main motivation for the second option over the first is that it would simplify the implementation, and so this scenario will be the focus of this paper. The hope is that increasing the jet p_T cuts will improve the correlation between the particle-level regions

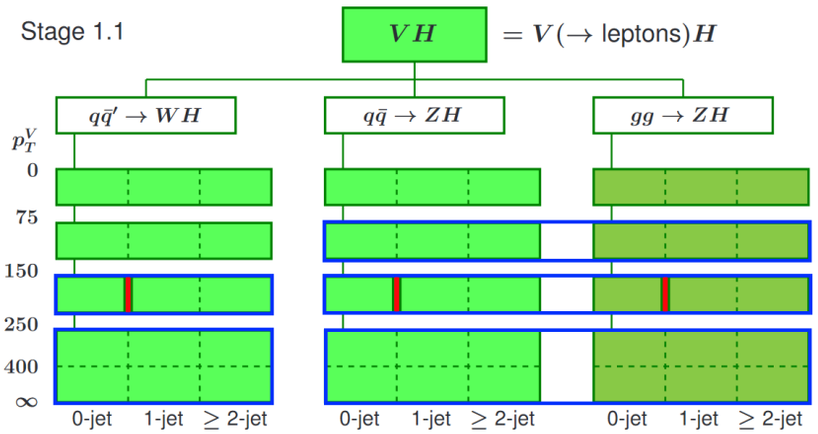

Figure 2. Diagram of the regions used in the STXS scheme for the $V(\rightarrow\text{leptons})H$ production [7]. The dashed lines show the theoretically possible regions that can be studied. However, due to limited statistics and the ability of the detector, some of the regions experimentally can't be used or need to be merged. This leaves with us with five regions, defined by the blue boxes; WH with $150 < p_T^V < 250$ and $p_T^V > 250$, and ZH with $75 < p_T^V < 150$, $150 < p_T^V < 250$ and $p_T^V > 250$. The red lines show the additional splits studied for this paper, giving seven regions.

(hereby referred to as STXS truth regions) and the reconstruction regions, i.e., that an event that falls in the truth region of no additional jets (only the two tagged jets) and $150 < p_T^V < 250$ also falls in the reconstruction region of two jets (only the two tagged jets) and $150 < p_T^V < 250$. The improved correlations should then increase the significance in the STXS fit and reduce the uncertainties on the cross section measurements.

6. Results

In order to study the correlation between the regions, or bins, we can look at matrices like that shown in Figure 3. In this plot, the y-axis has the nine reconstruction bins (three n-jet regions each with three p_T^V regions), while the x-axis has the STXS truth bins. The entries are the relative fraction of each reconstruction bin in each of the truth bins, in percentage, so the rows add up to 100%. Ideally the main diagonal would all be 100%, but due to uncertainties in the reconstruction, many events lie off this diagonal. As can be seen for the nominal case in Figure 3, with no change in the jet p_T cuts, the correlations improve for increasing p_T^V and decrease for increasing number of jets. The worst directly correlated bin (4-jet and $150 < p_T^V < 250$) only has 16.7% correlation.

For the case where all the central jet p_T cuts have been increased to 30 GeV, the matrix is shown in Figure 4. Clearly increasing the cut has improved the correlations overall. All the 2-jet bins have decreased slightly by about 3.75%, although they did start off a lot higher than the other bins. It's nice to see that the worse the correlation in the bin was, the larger the improvement. So for the worst bin discussed earlier, there is an improvement in the correlation by 118%, with the other 4-jet bins improving by 80-100% and the 3-jet bins improving by 32-45%.

So for events with more than two jets, increasing the jet p_T improves the STXS correlations, but it slightly decreases in the 2-jet regions. Unfortunately, the 2-jet regions dominant in statistics and correlations. It is therefore possible that the benefits of the increased cuts will only be seen when the final STXS bins used for the fit are split into the zero and greater than zero additional jets regions, as explained in Section 4. Without the split it is possible that the 2-jet regions

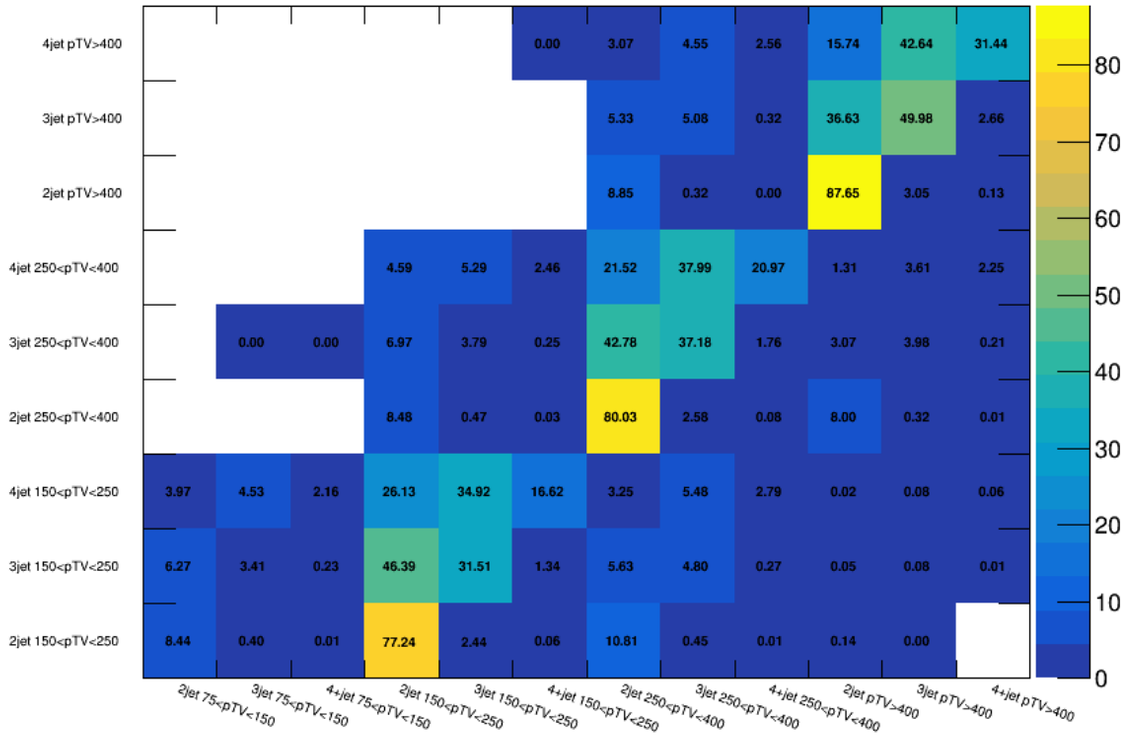

Figure 3. The correlation matrix comparing the nine reconstruction bins (y-axis) to the STXS truth bins (x-axis) for the nominal case. The bin values are the relative fractions in percent of each reconstruction bin in each truth bin, so the rows add up to 100%. The direct correlations between the reconstruction and particle-level bins lie on the diagonal from the bottom bin fourth from the left up to the top right bin. The reconstruction bins are defined by the number of jets in the event, while the truth bins are defined by the number of additional jets above the two b-tagged jets in the event.

dominate and the improvements in the higher jet regions is lost. As part of the next steps in this study, the full maximum likelihood fit will be performed in the STXS bins both with the increased cuts and with the extra split in the different number of jets.

7. Conclusion

The production of the Standard Model Higgs boson in association with a W or a Z boson, where the Higgs boson decays to $b\bar{b}$ and the W/Z bosons decay leptonically, is an important process due to the W/Z boson couplings to the Higgs boson and the decay of the Higgs boson in it's dominant decay channel to two bottom quarks. This process is being studied in the framework of the Simplified Template Cross Section (STXS) and possible improvements over the previous round are being explored. One of the ways for this is to increase the jet p_T cuts at reconstruction level to match more closely to the cuts applied at the STXS truth level.

By increasing the p_T cuts of all the central jets from 20 GeV to 30 GeV, an overall improvement in the correlation between the reconstruction and STXS truth bins is observed. In particular, the regions with 3 jets improve by 32-45% while the 4-jet regions improve by 80-118%. There is a very slight decrease in the 2-jet region of about 3.75%, but these regions started off with high correlations. Future work would be to perform the full maximum likelihood fit in the STXS bins

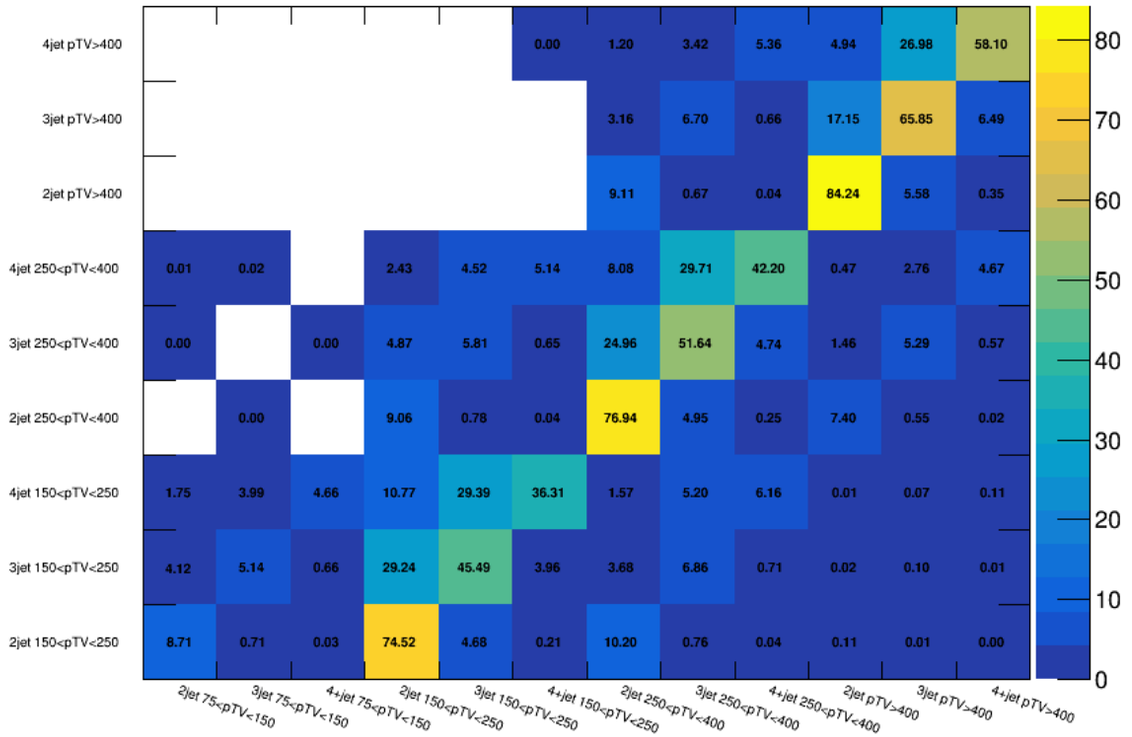

Figure 4. The correlation matrix comparing the nine reconstruction bins (y-axis) to the STXS truth bins (x-axis) for the increased jet p_T cut scenario. The bin values are the relative fractions in percent of each reconstruction bin in each truth bin, so the rows add up to 100%. The direct correlations between the reconstruction and truth bins lie on the diagonal from the bottom bin fourth from the left up to the top right bin. The reconstruction bins are defined by the number of jets in the event, while the truth bins are defined by the number of additional jets above the two b-tagged jets in the event.

to determine if the increased jet p_T cuts improve the significances as well, and if increasing the cuts would still be beneficial if the fit is not split by the number of jets in the events.

References

[1] D de Florian et al. Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector. 2016.
[2] ATLAS Collaboration. Measurements of WH and ZH production in the $H \to b\bar{b}$ decay channel in pp collisions at 13 TeV with the ATLAS detector. *Eur. Phys. J. C*, 81(2):178, 2021.
[3] ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *Journal of Instrumentation*, **3**(08):S08003, 2008.
[4] L Evans and P Bryant. LHC Machine. *Journal of Instrumentation*, **3**(08):S08001, 2008.
[5] M Cacciari, G P Salam, and G Soyez. The anti-kt Jet Clustering Algorithm. *JHEP*, 0804(063), 2008.
[6] ATLAS Collaboration. ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV. *Eur. Phys. J. C*, 79(11):970, 2019.
[7] N Berger et. al. Simplified template cross sections - stage 1.1, 2019. arXiv:1906.02754 [hep-ph].