Backgrounds in $H \to WW^{(*)} \to \ell\nu\ell\nu$ with ATLAS

Tomo Lazovich[1,a] On behalf of the ATLAS Collaboration

[1]*Harvard University, Department of Physics, 17 Oxford St., Cambridge, MA 02138, USA*

Abstract. We present techniques used to estimate the backgrounds in the search for the Standard Model Higgs boson in the $H \to WW^{(*)} \to \ell\nu\ell\nu$ decay channel with the ATLAS experiment at the LHC. The dataset corresponds to 13 fb^{-1} of integrated luminosity taken at a center of mass energy of 8 TeV. Only the final states with an electron, muon, and zero or one jet are presented here.

1 Introduction

In July 2012 the ATLAS [1] and CMS [3] experiments at the LHC announced the discovery of a new particle consistent with the long-sought Higgs boson [4, 5]. The results presented here constitute an update of the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis with a dataset of 13 fb^{-1} taken at a center of mass energy of 8 TeV [2]. In particular, we summarize the methods of background estimation for this search channel, which focuses on the low mass Higgs signal region.

The $WW^{(*)} \to \ell\nu\ell\nu$ decay channel of the Higgs boson has a final state defined by two leptons and missing transverse energy (E_T^{miss}) from neutrinos which escape detection. The analysis presented here considers only the final states with one electron, one muon, and zero or one jets with transverse momentum (p_T) greater than 25 GeV. The leptons are required to be isolated and the leading (subleading) lepton must have $p_T > 25$ (15) GeV. Additionally, the event must have relative missing transverse energy ($E_{T,\mathrm{rel}}^{\mathrm{miss}}$) greater than 25 GeV, where $E_{T,\mathrm{rel}}^{\mathrm{miss}} = E_T^{\mathrm{miss}} \sin(\min(\Delta\phi, \frac{\pi}{2}))$ and $\Delta\phi$ is the azimuthal angle between the E_T^{miss} and the nearest reconstructed lepton or jet. This definition helps to reject events where a mismeasurement of one of the reconstructed objects is a major source of the E_T^{miss}. After these pre-selection cuts, the signal region is divided into zero and one jet bins and additional topological cuts (specific to each bin) are applied to discriminate the Higgs signal from background contributions.

Many processes in the Standard Model (SM) produce final states similar to that in $H \to WW^{(*)} \to \ell\nu\ell\nu$. The largest background contribution is the irreducible SM WW background. The next largest background contributions come from $t\bar{t}$ and single top production. These backgrounds are primarily relevant for larger jet multiplicity bins but are also present in the zero jet bin. Another important background for the analysis is the W+jets background. Here a single W boson is produced in association with one or more jets, and one of the jets fakes a final state
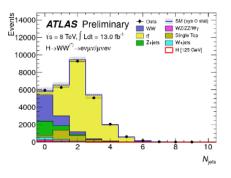

Figure 1. The jet multiplicity distribution after signal preselection cuts are applied [2].

lepton. Other backgrounds which will not be discussed in great detail include the Z+jets and diboson (WZ, ZZ, $W\gamma$) backgrounds. Figure 1 shows the jet multiplicity distribution before the selection separates the events into jet multiplicity bins. After the pre-selection, the WW background is the dominant background in the zero jet bin while the top background dominates the higher jet multiplicity bins.
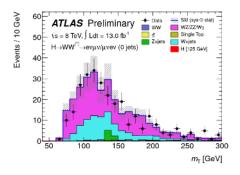
2 W+jets and Other Minor Backgrounds

The W+jets background arises from SM W boson production in association with jets where one jet produces an object reconstructed as a lepton. This can be a real lepton produced by heavy quark decay or a product of the jet fragmentation that is incorrectly reconstructed as an electron. The W+jets background contribution is estimated by a data-driven method called the "fake factor" method. First, a control region is defined in data by requiring one lepton with the same identification and isolation criteria as the signal leptons. The second lepton is required to be anti-identified, satisfying loosened isolation criteria and failing at least one identification requirement. These events are then required to pass the full signal selection.

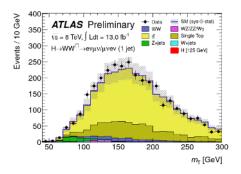[a]e-mail: tlazovich@physics.harvard.edu

A fake factor, the ratio of the number of lepton candidates passing all identification requirements and signal selections to the number that are anti-identified, is derived in an inclusive data dijet sample. This factor is used to scale the number of events in the control region to the signal region. The total relative uncertainty on the estimate is 50%, dominated by the systematic uncertainty on the fake factor.

Figure 2 shows the transverse mass (m_T) in a same sign validation region (where two same sign rather than two opposite sign leptons are required). This region is composed largely of W+jets and $WZ/ZZ/W\gamma$ backgrounds and is used to validate the modeling of kinematic variables for these samples. This region shows that the m_T is well modeled (within statistics) for these samples.

Figure 2. m_T in the zero jet same sign validation region [2].

Here we briefly mention other minor backgrounds which will not be discussed in further detail. First, the Z+jets background comes from a case where the Z decays to two leptons and there is fake missing transverse energy in the event due to the calorimeter resolution. This background is normalized to data in a control region requiring $m_{\ell\ell} < 80$ GeV and $\Delta\phi_{\ell\ell} > 2.9$. Finally, the normalizations for the remaining backgrounds ($WZ/ZZ/W\gamma$) are taken from Monte Carlo simulation.

Figure 3. m_T distribution in the top one jet control region. The rate predicted by Monte Carlo simulation has not yet been normalized to the data. [2]

3 $t\bar{t}$ and Single Top Backgrounds

The $t\bar{t}$ and single top backgrounds are normalized together in control regions (CR) separated by jet multiplicity. The

Background	Stat. (%)	Theory (%)	Expt. (%)	Crosstalk (%)	Total (%)
WW, H + 0-jet	3.3	7.2	1.5	6.2	13
WW, H + 1-jet	9	8	12	34	54
top, H + 1-jet	2	8	29	1	37

Table 1. Total uncertainties on backgrounds normalized using simple NF scaling in CRs [2].

one jet bin is normalized in a CR defined with the same pre-selection as the signal region (SR) and at least one b-tagged jet. For the zero jet bin, there are two CRs used for the background estimate. First, a CR with only the SR pre-selection is used to estimate the fraction of top events passing a jet veto. A second, b-tagged CR is then used to estimate the probability of having no other jets reconstructed in the event and is used as a correction to the fraction estimate from the first CR.

Figure 3 shows the m_T in the one jet CR before any normalization factors are applied. The normalization factors (NF), or ratio between the data and Monte Carlo predictions, derived via these methods are 1.04 ± 0.05 (stat.) for the zero jet channel and 1.03 ± 0.02 (stat.) for the one jet channel.
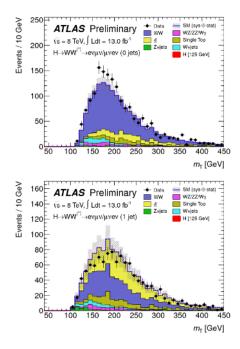
4 Standard Model WW Background

The SM WW background is estimated in a CR which uses the SR pre-selection cuts (two leptons, missing transverse energy) and is separated into jet bins. While the SR requires $m_{\ell\ell} < 50$ GeV, the WW CR requires $m_{\ell\ell} > 80$ GeV. This is the largest background in the signal region. The WW modeling in simulation is done with a tune of Powheg for event generation and Pythia 8 for parton showering.

Figure 4 shows the m_T in the WW zero and one jet CR, before the application of any WW NF. In both the zero and one jet (but particularly in the one jet) WW CRs, there is a non-negligible contrbution from the top backgrounds. Therefore, the top backgrounds are first normalized using the procedures described in Section 3 before all of the non-WW backgrounds are subtracted from the event yields in the CR to derive the final normalization. The ratio of the data (with non WW background subtracted) to the WW simulation prediction is 1.13 ± 0.04 (stat.) in the zero jet CR and 0.84 ± 0.08 (stat.) in the one jet CR. The NF differ between the zero and one jet channels because these CR are correcting for the over-prediction of the jet multiplicity distribution by the current Powheg+Pythia 8 tune used by ATLAS.

5 Background Predictions and Uncertainties

Table 1 shows the total uncertainties on the background normalization for the backgrounds which use the simple data to simulation scaling in the CR for normalization. Theoretical uncertainties on the estimates include differences due to the choice of generator and parton shower/underlying event as well as other contributions.

Figure 4. The m_T distribution in the WW zero (one) jet CR is shown in the top (bottom) plot. The WW rate predicted by Monte Carlo simulation has not yet been normalized to the data. The top backgrounds have been normalized according to the procedure described in the text [2].

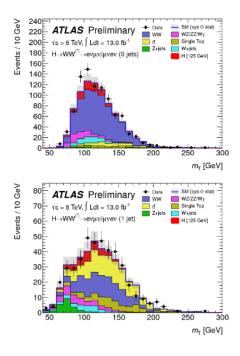| Background	0 jet NF	1 jet NF
Top	1.04 ± 0.05	1.03 ± 0.02
WW	1.13 ± 0.04	0.84 ± 0.08
Z+jets	0.87 ± 0.03	0.85 ± 0.03

Table 2. Normalization factors (NF) for all backgrounds whose normalizations are taken from data [2].

The experimental uncertainties are dominated the jet energy scale and resolution and, in the one jet bin, the *b* tagging efficiency. The "Crosstalk" column refers to uncertainties on other backgrounds which must be subtracted from the CR before the normalization of the desired background can be computed. Notice in particular that the WW one jet normalization has a large contribution from crosstalk due to the fact that a top background contribution must be subtracted from the CR before the normalization is computed.

Table 2 shows the NF derived for all of the backgrounds whose normalizations are taken from data. In the case of everything except the top zero jet background, this factor is simply the ratio of the number of Monte Carlo events to data events in the appropriate CR for that background. We can see that most of the backgrounds do not require very large corrections to their normalization (none more than 16%).

Figure 5 shows the m_T distribution after the signal selection cuts have been applied for the zero and one jet bin. It can be seen here that the WW background is dominant in the zero jet bin, while both WW and top are dominant

in the one jet bin. In zero jet, there is a total of 774 ± 9 (stat.) expected background events, and 555 ± 5 (stat.) of those are SM WW events. In the one jet, out of 386 ± 5 (stat.) total expected background events, 118 ± 2 (stat.) are SM WW events while 134 ± 5 (stat.) are $t\bar{t}$ events. The difference between the background expectation and the 917 (433) events observed in data in the zero (one) jet bin are due to the presence of the Higgs-like signal.

Figure 5. m_T distribution in the zero (one) jet bin after all signal selection cuts in the top (bottom) plot [2]. The expectation for a 125 GeV Higgs signal is shown in red.

6 Conclusion

A wide array of estimation methods can be employed to understand the complicated background processes that factor into a search for a $H \to WW^{(*)} \to \ell\nu\ell\nu$ signal. Simple data to simulation scaling in control regions is used for backgrounds such as SM WW (or top in the one jet bin) where the variable shapes are well modeled but their normalizations may be incorrect. More complicated data-driven methods, such as the W+jets fake factor method, can also be used when the backgrounds are not well modeled by simulation alone.

References

[1] ATLAS Collaboration, JINST **3** S08003 (2008)

[2] ATLAS Collaboration, ATLAS-CONF-2012-158, https://cds.cern.ch/record/1493601

[3] CMS Collaboration, JINST **3** S08004 (2008)

[4] ATLAS Collaboration, Phys. Lett. B, **716**, 1-29 (2012)

[5] CMS Collaboration, Phys. Lett. B, **716**, 30-61 (2012)