Matthew Heath
University of Edinburgh

Background Modelling in the ttH(H→γγ) Channel at ATLAS

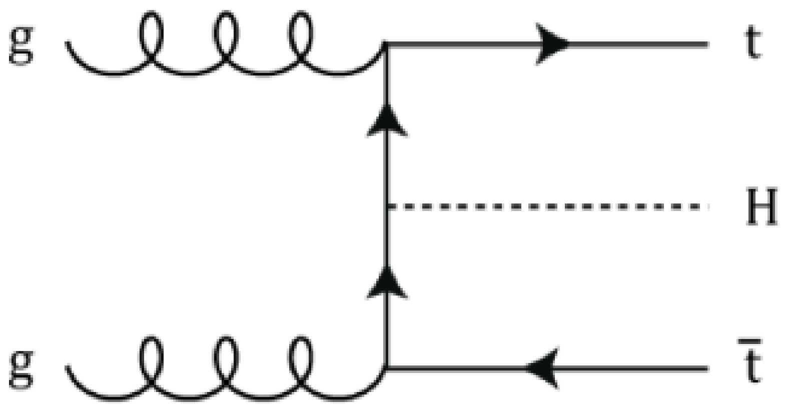
IoP Joint APP and HEPP Conference 2019
Imperial College London

Overview

- ttH(γγ) Motivation
- Analysis Process
- Background Modelling
- Monte Carlo Statistics
- Results and Conclusions

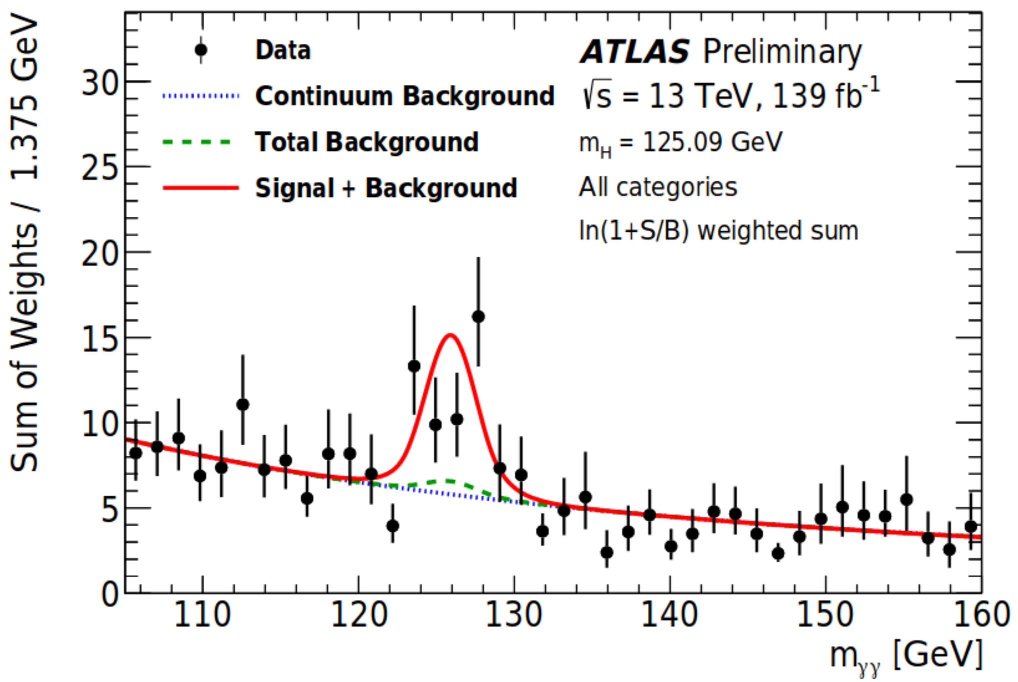
ttH Production

- ttH cross-section for a 125GeV Higgs at $\sqrt{s} = 13TeV$ is $\sigma_{ttH} = 0.507pb$. (~1% of that from ggH).
- Gives a tree-level measurement of the top Yukawa coupling, y_t.
- y_t measurement is sensitive to BSM physics.
- Analyses are split up by the Higgs decay.



$$\sqrt{\sigma_{ttH}} \propto y_t = \sqrt{2}\frac{m_t}{v} \approx 1$$

ttH(γγ)



ATLAS-CONF-2019-004

- H→γγ has a tiny BR of 0.227%. Though this channel produces very high signal purity.
- High resolution of the ATLAS EM calorimeter results in high photon reconstruction and isolation efficiency.
- Allows for a background fit in a one-parameter function in the $m_{\gamma\gamma}$ spectrum.
- High resolution gives a narrow signal peak in the parametrised background continuum.

ttH(γγ) Event Selection

- Require two tight, isolated photons with (sub)leading photon possessing p_T > (25) 35GeV in the [105, 160] GeV $m_{\gamma\gamma}$ range. $p_T/m_{\gamma\gamma}$ > (0.25)0.35 for the (sub)leading photon.
- ≥ 1 jet with p_T > 25 GeV, and containing a b-jet tagged with 77% efficiency.
- Use BDTs to help discriminate ttH events from ggH and multi-jet backgrounds.

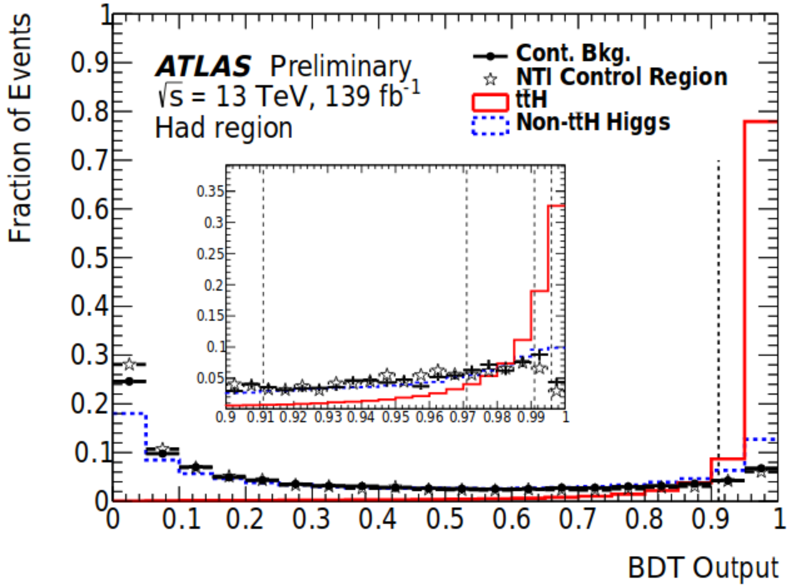
BDT Training Variables

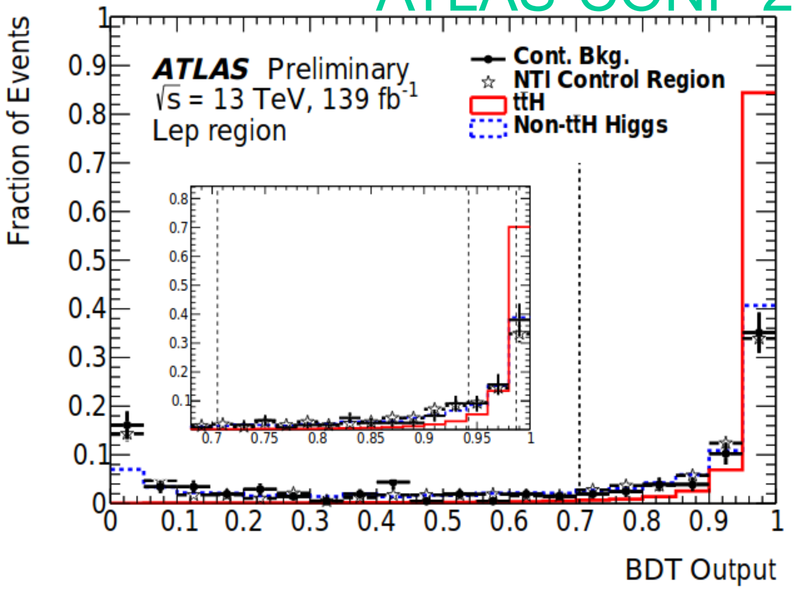
- Hadronic (Had) category for fully hadronic decays of the tops.
- BDT trained on:
- p_T, η, φ, E, b-tag of up to 6 leading jets in p_T.
- Magnitude and φ of MET.
- $p_T/m_{\gamma\gamma}$, η, φ of each photon.

- Leptonic (Lep) region for semi-leptonicly enriched top decays (≥ 1 lepton).
- BDT trained on:
- p_T, η, φ, E of the up to 4 (2) leading jets (leptons) in p_T.
- Magnitude and φ of MET.
- $p_T/m_{\gamma\gamma}$, η, φ of each photon.

Had and Lep BDT Results

ATLAS-CONF-2019-004



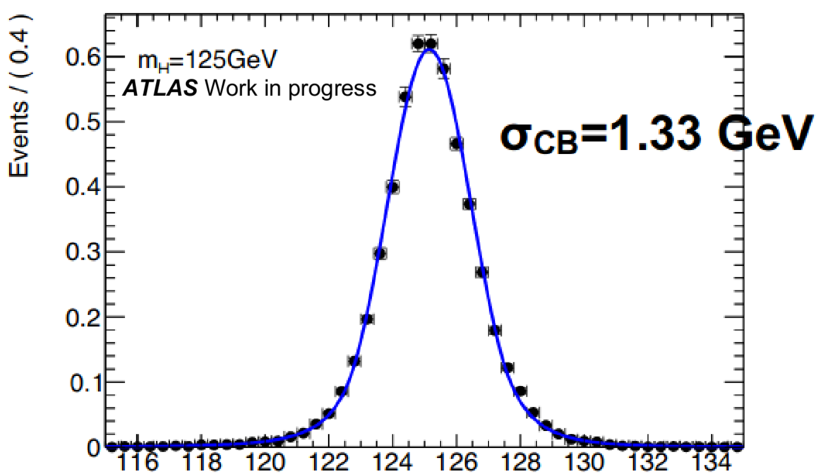


- Events with low BDT response are removed.
- About 85% (97%) of the ttH signal events are selected and about 89% (43%) of the non-resonant background events are rejected in the Had (Lep) region.
- Remaining events are categorised into four (three) bins in the Had (Lep) region depending on BDT response. These bins are chosen to optimise expected sensitivity to ttH signal.

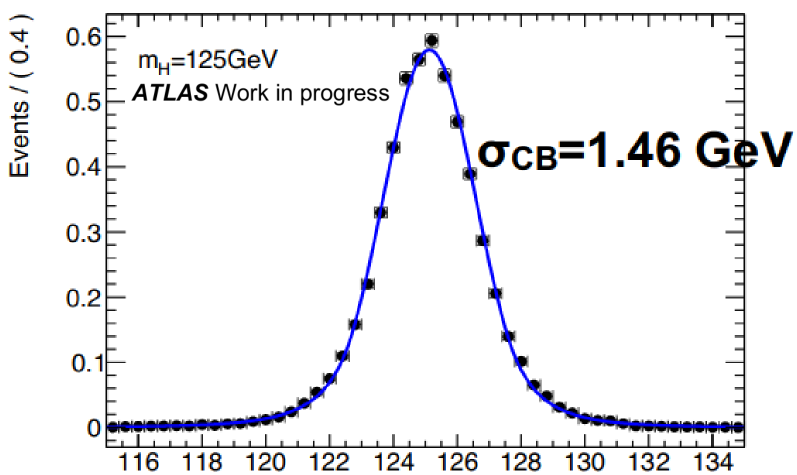
ttH(γγ) Signal and Background Modelling

- Signal is modelled as a Double Sided Crystal Ball function.
- The DSCB parameters for each Had and Lep BDT bin are determined using fits to the Monte Carlo Sample.

- Background is modelled using a dedicated data control region for Had, and simulated background events for the Lep region.
- The functional form is determined as either Exponential or Power-Law by the “spurious signal” method.



ttHhad BDT1



ttHlep BDT1

Spurious Signal

- Fit is performed on background simulation in the $m_{\gamma\gamma}$ range [105, 160] GeV (same as analysis).
- Values of $m_{\gamma\gamma}$ are then scanned in steps of 1 GeV between 121 GeV to 129 GeV (~$m_H \pm 2\sigma$), parametrising the background model to an analytical function.
- The spurious signal is the resulting fit bias and is evaluated as

$$N_{sp} = \max_{121 < m_H < 129\,GeV} \left| N_s(m_H) \right|$$

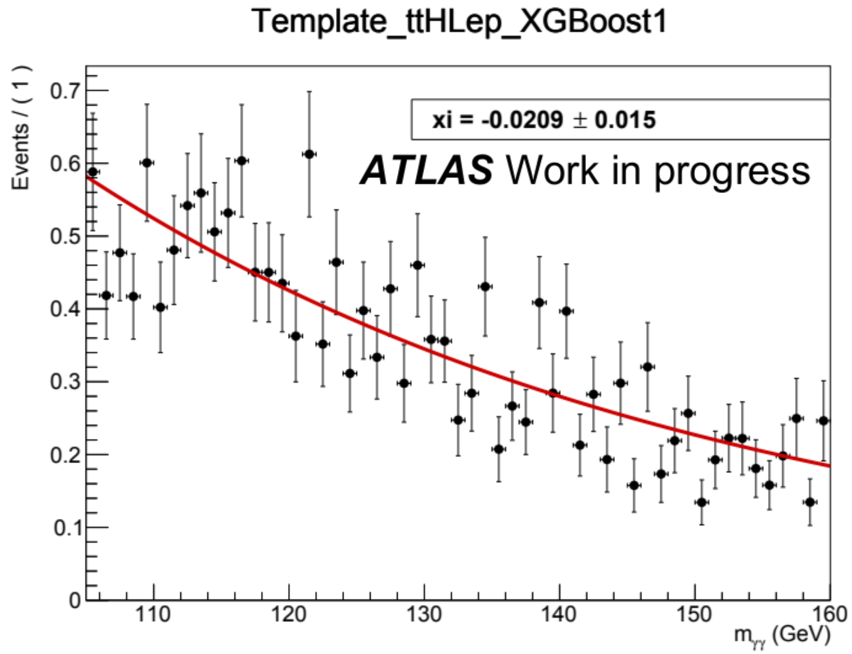
- In order to accept a functional form as the background template is that the spurious signal must satisfy at least one of:
  - $N_{sp} < 10\%\ N_{s,\,exp}$
  - $N_{sp} < 20\%\ \sigma_{bkg}$
- N_{sp} is then used as the systematic uncertainty on the background modelling if a function passes this spurious signal test.

Spurious Signal Problem

- When calculating the spurious signal, MC generation is relied on to provide a background-only shape in the signal region.
- As luminosity of data increases, the statistical uncertainty on the background MC samples in comparison to the expected number of signal events increases.
- The background functional form becomes more and more dependent on the statistical uncertainty of the MC.
- N_{sp} increases as discrepancies between the fitted background and the true form increase.
- Systematic error on the background model could become dominant in the measurement.

Creating Pseudo-Datasets

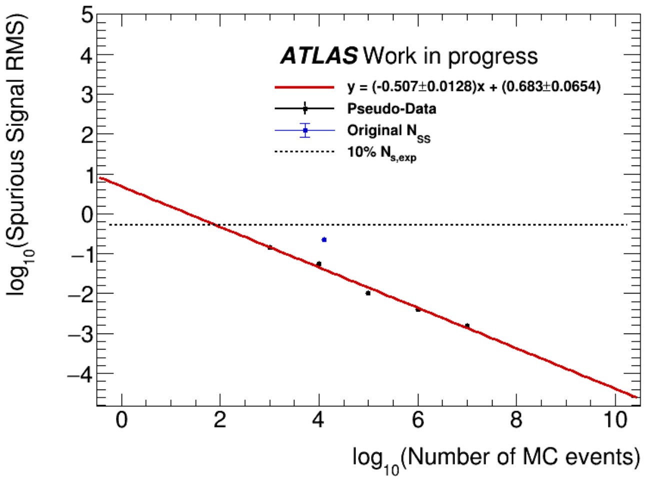
- To quantify how the statistics of a Monte Carlo sample affect the spurious signal, need to produce a number of different templates.
- To create pseudo-datasets of the background models with differing statistics, spurious signal software was run on the background category templates to obtain their most relevant functional forms.
- Using ROOFIT, the original templates are fitted with these most relevant functional forms which are used as a PDF to randomly generate events in the $m_{\gamma\gamma}$ spectrum.
- Using this method 100 pseudo-datasets containing 1k, 10k, 100k, 1M, and 10M events were created for each category and renormalised to the number of events in the original templates.



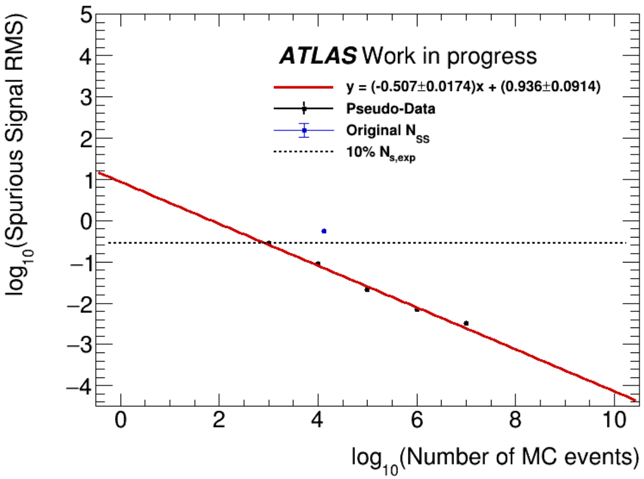
MC Statistics Study

- The spurious signal software was then run over each pseudo-dataset, testing for different background PDF forms.
- The maximum number of spurious signal events for the selected best fit is then recorded.
- The RMS of the 100 templates can then be plotted against the number of events used to create the sample to observe any trends.

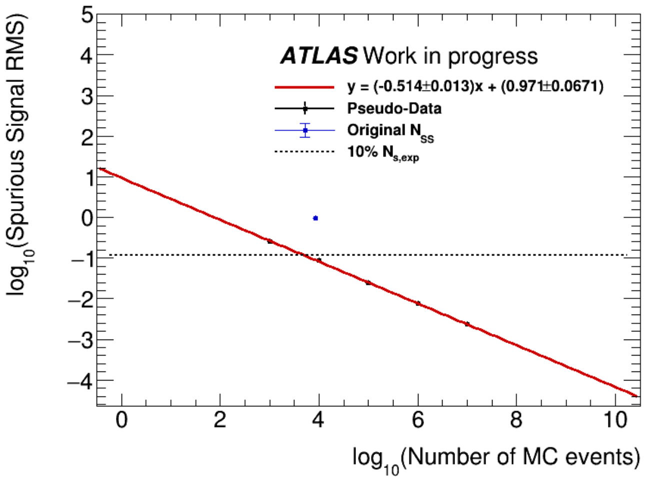
Results



- Lep, BDT Bin 1



- Lep, BDT Bin 2



- Lep, BDT Bin 3

It should be noted that the blue point on each plot, the Original N_{ss}, shows the spurious signal result from the background templates used to create the pseudo-datasets. They are not aligned with the pseudo-datasets due to the latter's averaging over 100 datasets, being less prone to fluctuation.

Conclusions

- Uncertainties on the background model are becoming a dominant systematic in the future of this analysis as data luminosity increases.
- These uncertainties are already more pressing in other $H \rightarrow \gamma\gamma$ categories.
- This is mainly due to the statistical fluctuations of the Monte Carlo forming the background templates.
- Study to quantify how these statistics affect the spurious signal method was performed.
- This allows for the estimation of the sample size required for a given systematic uncertainty.