$HH \to bb\tau\tau$ Analysis with Run2+Run3

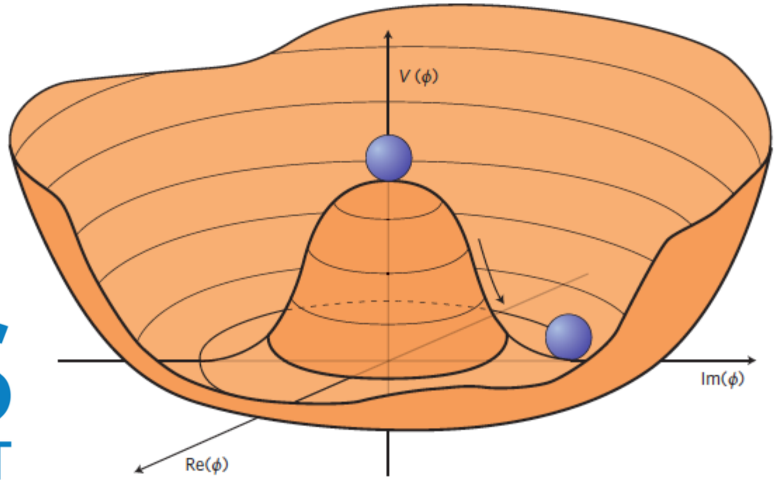
with the ATLAS Detector

Bhupesh Dixit

Supervised by:

Carl Gwilliam, Jordy Degens, Katharina Behr

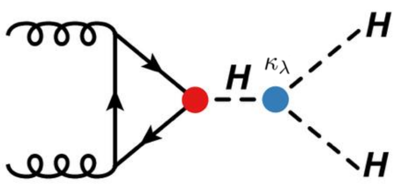
April 08, 2025



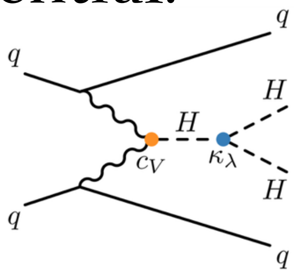
$HH \rightarrow b\bar{b}\tau^+\tau^-$ Analysis

- Di-Higgs allows measuring Higgs self-coupling which will be the ultimate test of EWSB and verify the Mexican hat shape of the Higgs potential.

- Higgs production modes: ggF + VBF



Gluon-Gluon Fusion (ggF)
~87% at 13TeV
Source: PDG2016



Vector Boson Fusion (VBF)
~7% at 13TeV
Source: PDG2016

- Higgs decay channels:

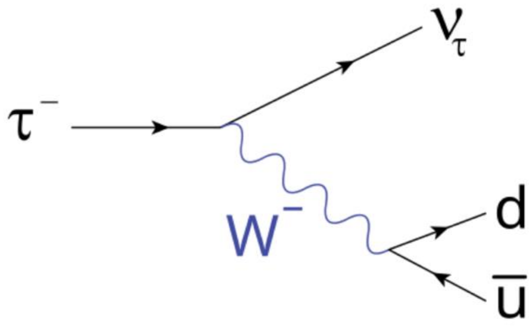
	bb	WW	ττ	ZZ	γγ
bb	33%				
WW	25%	4.6%			
ττ	7.4%	2.5%	0.39%		
ZZ	3.1%	1.2%	0.34%	0.076%	
γγ	0.26%	0.10%	0.029%	0.013%	0.0005%

Source: Bill Balunas : HH at ATLAS

- $HH \rightarrow b\bar{b}b\bar{b}$
  - High branching ratio
  - High QCD background
- $HH \rightarrow b\bar{b}\gamma\gamma$
  - Low background
  - Low branching ratio
- $HH \rightarrow b\bar{b}\tau^+\tau^-$
  - Significant branching ratio
  - Relatively low background
  - Highest expected sensitivity to di−Higgs (as seen in Run2) coupling for $\kappa_\lambda > 1$

$HH \rightarrow b\bar{b}\tau^+\tau^-$ Analysis

The major channels in the $HH \rightarrow bb\tau\tau$ analysis are:



Hadhad
2 b-jets
+2 Hadronic Tau
(Opposite Sign)

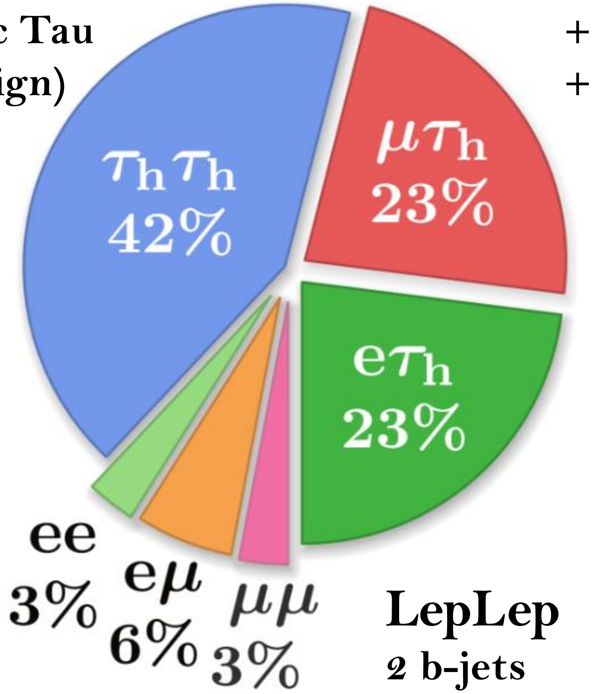
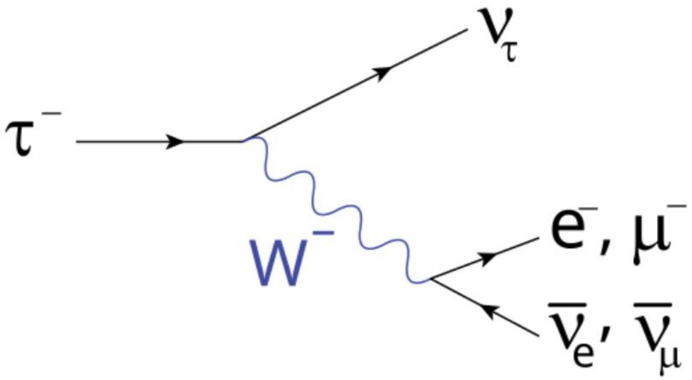


Image Source: Tikz.net

LepHad
2 b-jets
+1 Hadronic Tau
+1 e/mu
(Opposite Sign)



Studied under HH->bbll Analysis(link)

LepLep
2 b-jets
+ e+mu
or e+e
or mu+mu
(Opposite Sign)

$HH \to b\bar{b}\tau^{+}\tau^{-}$ Analysis

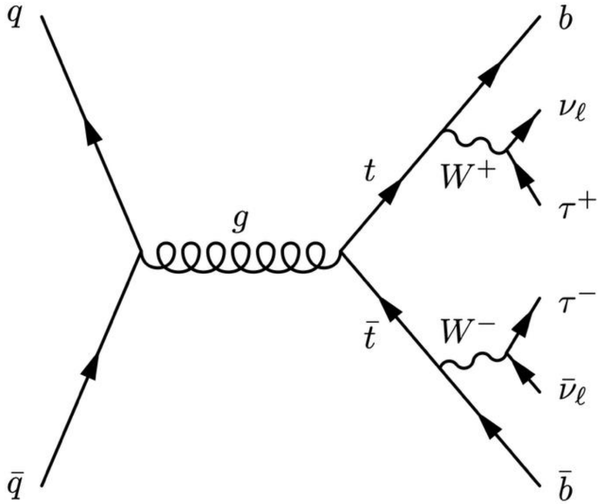
$HH \to bb\tau\tau$ analysis workflow:

Using GNNs Signal Region is split:
- VBF Like
- ggF Like

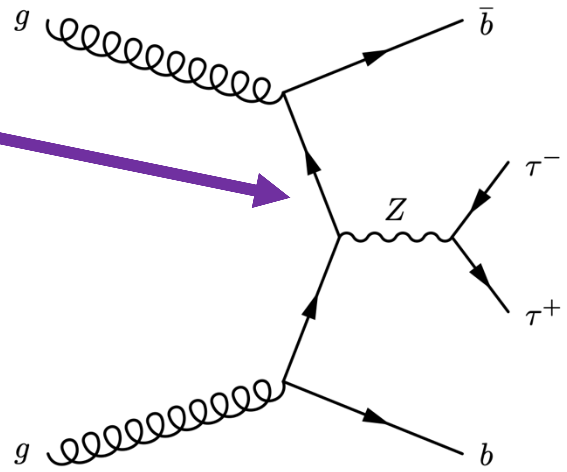
ggF is further split into:
- ggF Low (mHH < 350 GeV)
- ggF High (mHH > 350 GeV)

Workflow:
- Selected Events
- Categorized Events
- MVA Analysis
- Bkg. Estimation & Control Regions
- Systematic Uncertainties
- Statistical Analysis
- Summary

Same signature as signal



Same signature as signal

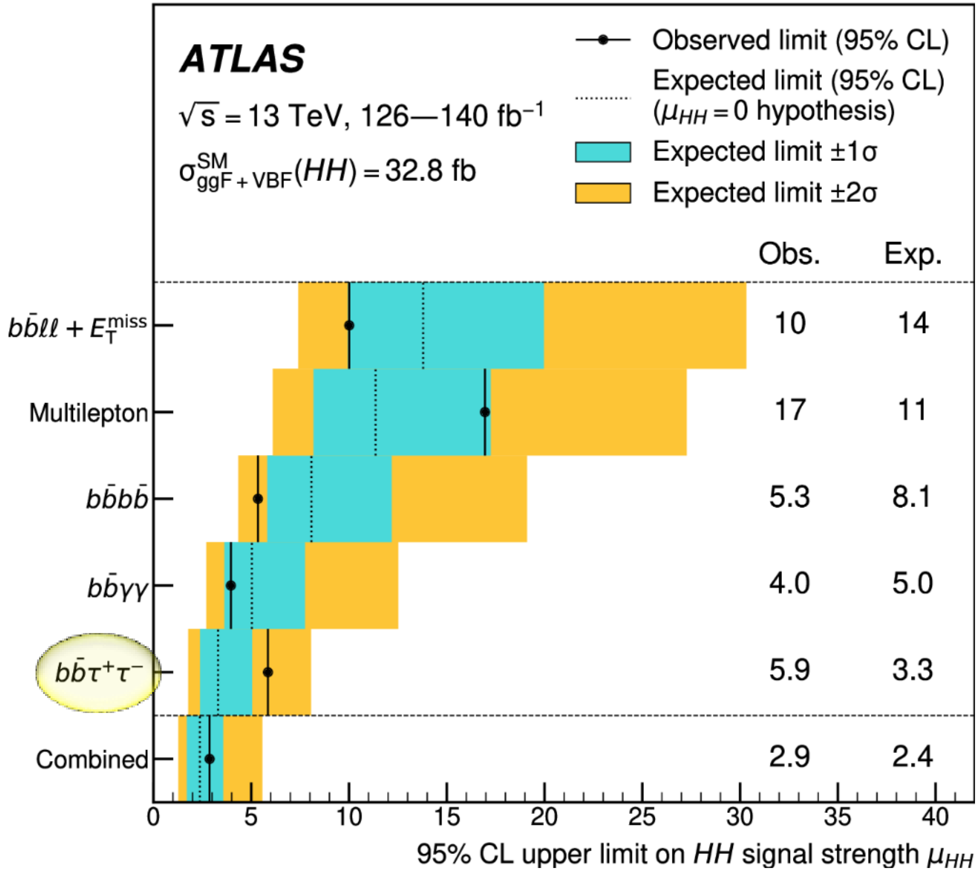


Recap of Run2

Upper Limit on Signal Strength

$$\mu_{HH} = (\sigma_{\text{ggF}} + \sigma_{\text{VBF}})/(\sigma^{\text{SM}}_{\text{ggF}} + \sigma^{\text{SM}}_{\text{VBF}})$$

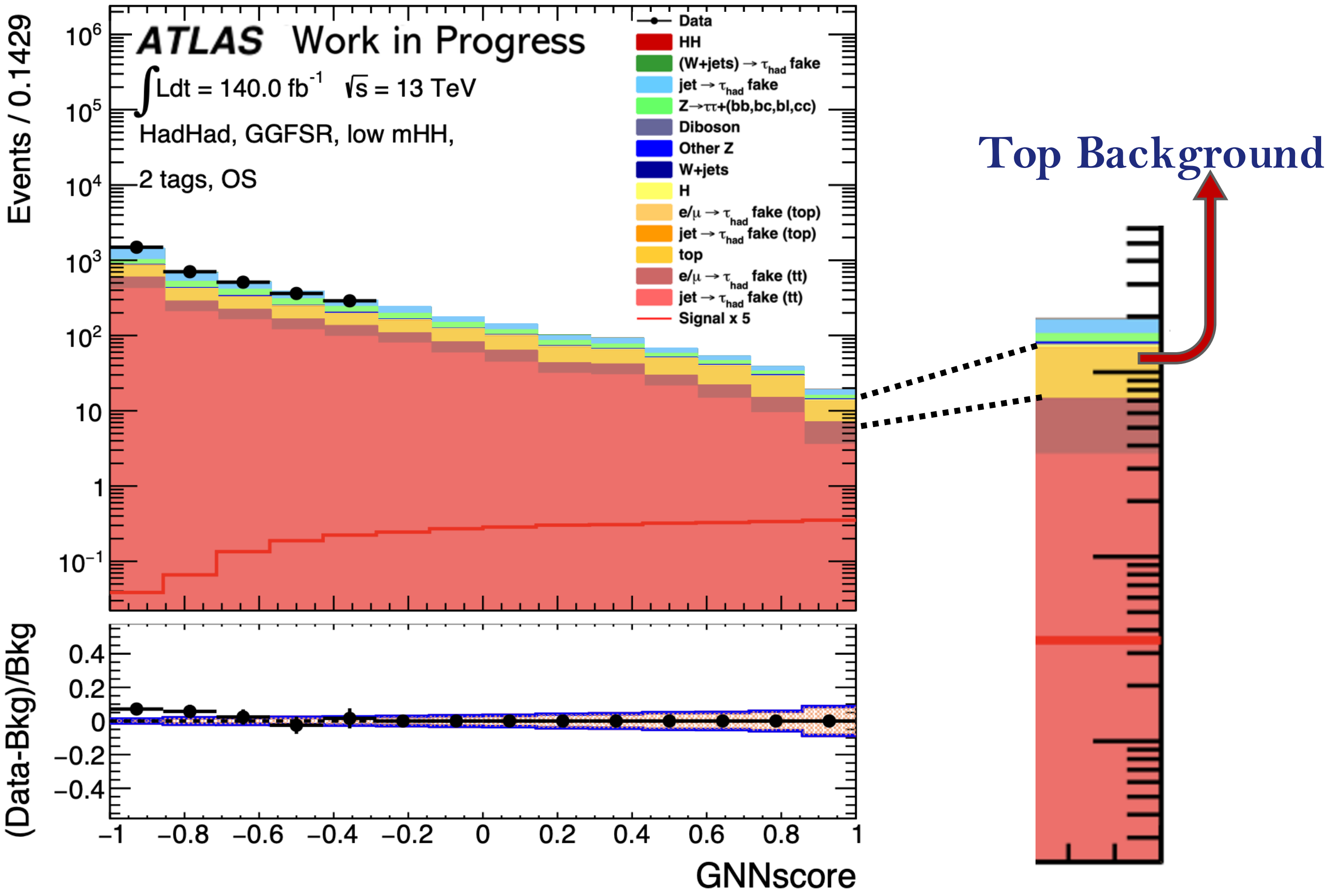
Source: HH->bbtautau Run2 Legacy



$HH \rightarrow b\bar{b}\tau^{+}\tau^{-}$ was the most sensitive channel in Run2

An Example of GNN Output

ggF Low Signal Region



GNN Score for ggF Low Mass in the Signal Region. Note the Top and Z backgrounds in the highest order bin.

TopCR: Why do we need it?

Run2 Flashback…

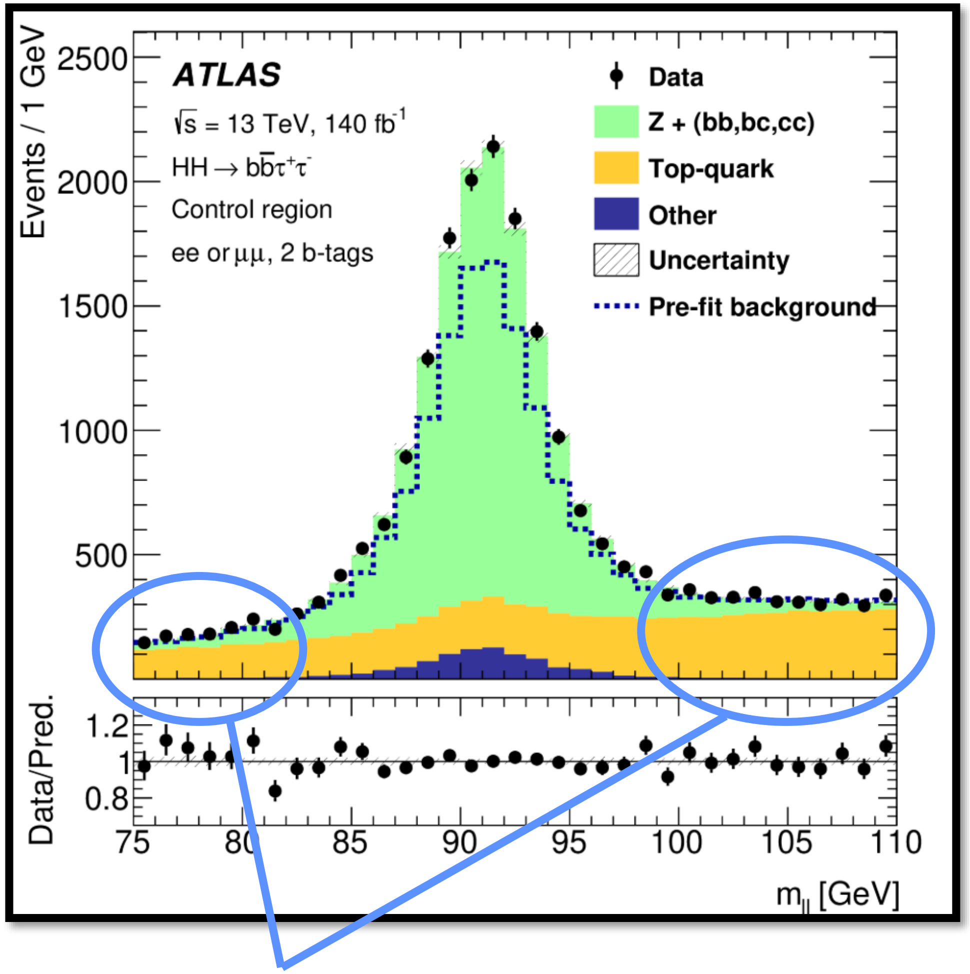
Source: HH->bbtautau Run2 Legacy

- **Assumption**: Kinematics of top events are the same for tau and e/mu decays of the W.
- Extract Z and ttbar normalization from CR
  - Cut: 2 e/mu of opposite charge + 2b-jets
  - Dilepton invariant mass ($m\ell\ell$) within the range of 75 GeV < $m\ell\ell$ < 110 GeV

So…what's the problem here?

  - **Large Z background**

- Hence, dedicated TopCR is needed.
- With selection to remove Z-background:
  - Exactly 1 electron and 1 muon of opposite charge
  - Exactly two b-tagged jets
- I led the TopCR modelling and systematic unc.

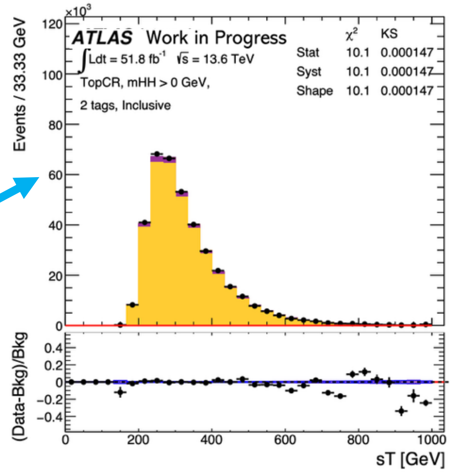


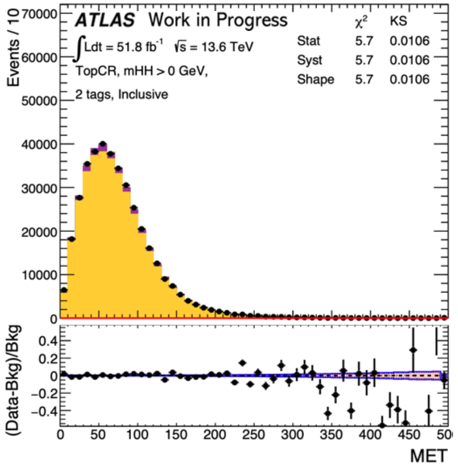
Regions from ZCR used to constrain Top background in Run2 Legacy

New Analysis with TopCR: How is it useful?

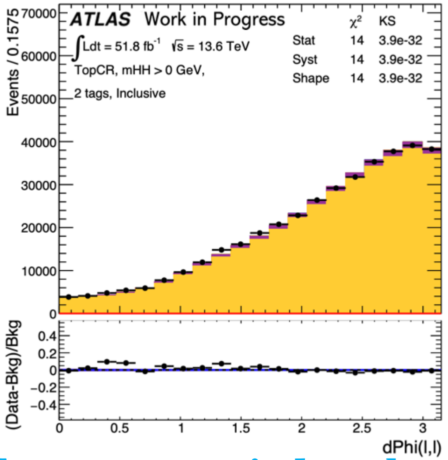
Background to Data Corrections &Verifying the MVA input variable modelling in MC

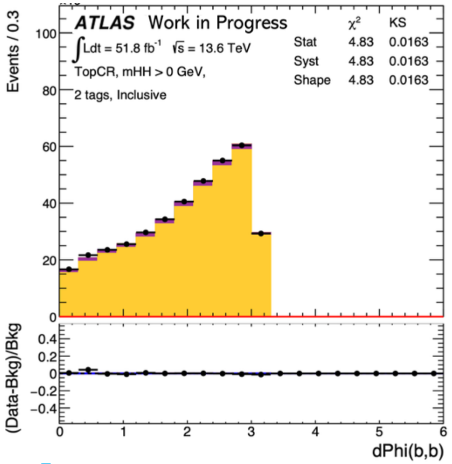
Energy-related Variables

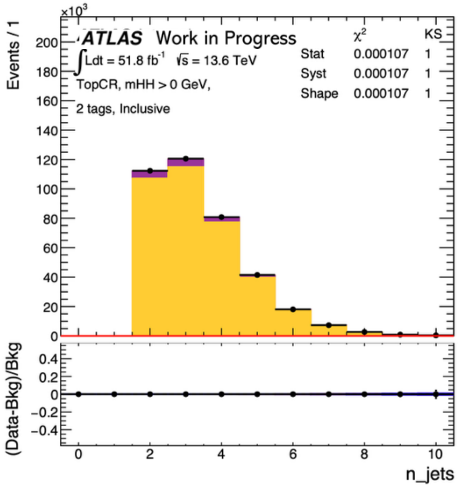


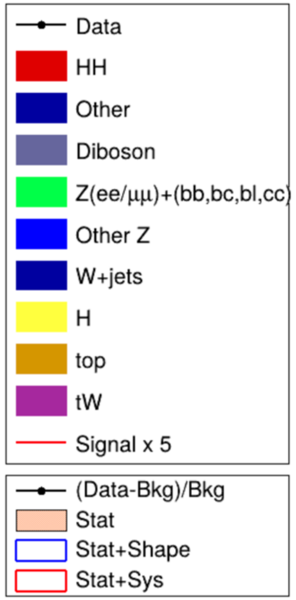


Angular Variables









Object Multiplicity Variable

- These plots have been re-weighted to data.
- The weights (SF) are calculated bin-wise using:

$$r_i = (N_{i,\text{data}} - N_{i,\text{non-}t\bar{t}})/N_{i,t\bar{t}}$$

- The MC plots use Powheg+Pythia at NLO accuracy.
- Both Data & MC use Run3 2022 at 51.8/fb Int. Lumi.
- SF = 1.00 ± 0.0025 (Statistical uncert. Only)

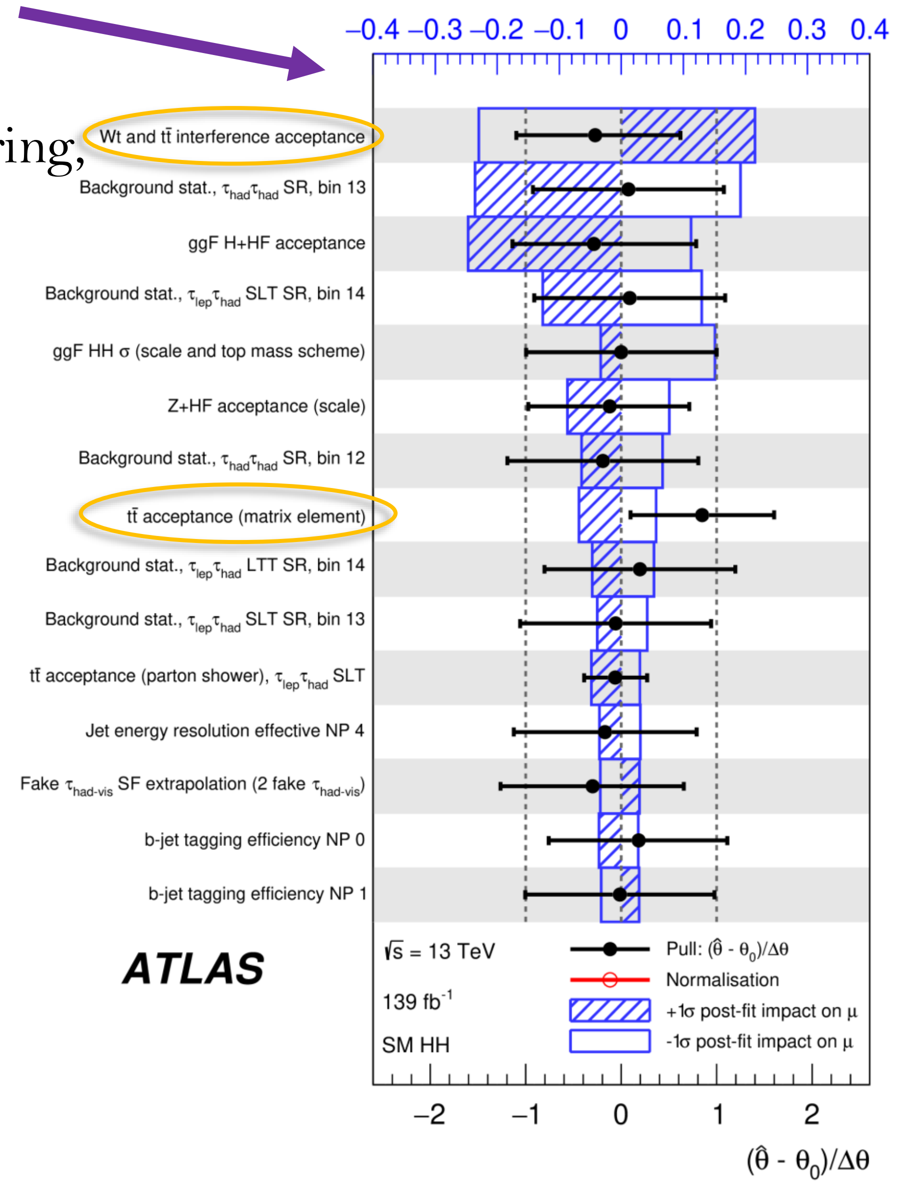
ttbar Systematics in HadHad Signal Region

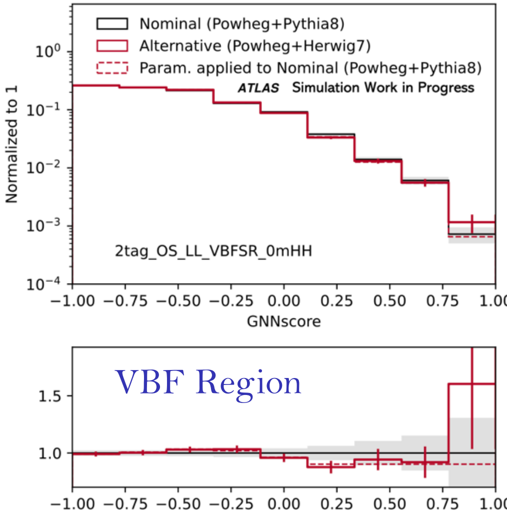
ttbar + tW uncertainties rank high in terms of their impact!

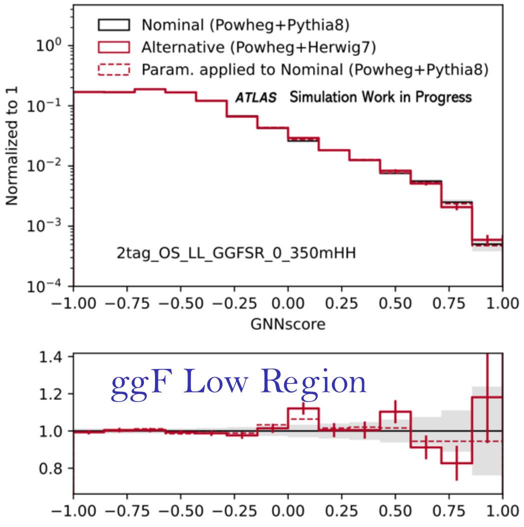
Source: HH->bbtautau Run2

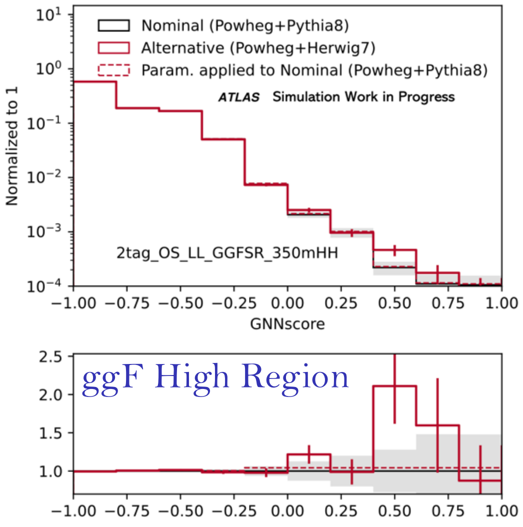
- There are two types of unc. implemented:
  - Shape unc. : From variations in physics processes e.g. showering, scattering etc.
  - Extrapolation unc. : From the discrepancy b/w Data & MC e.g. extrapolation from ZCR to Signal Region
- Following uncertainties are to be considered:

1. Parton Showering Uncertainty: Estimated using two samples
   - Powheg+Pythia8
   - Powheg+Herwig7



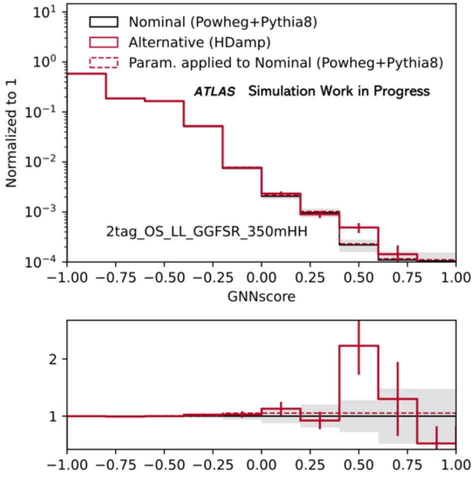






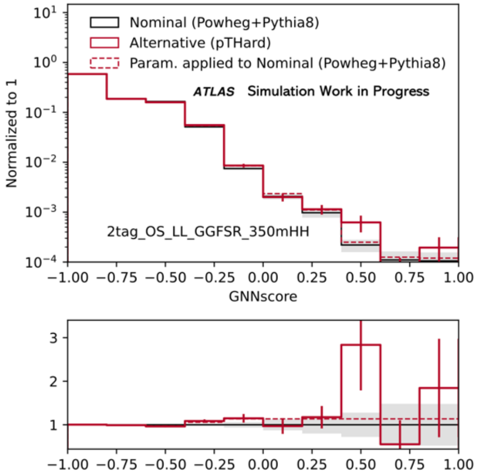
ttbar Systematics in HadHad Signal Region

2. Matrix Element –Parton Showering Matching Uncertainty: By changing the h_damp parameter which regulates 1st pT emission in MC



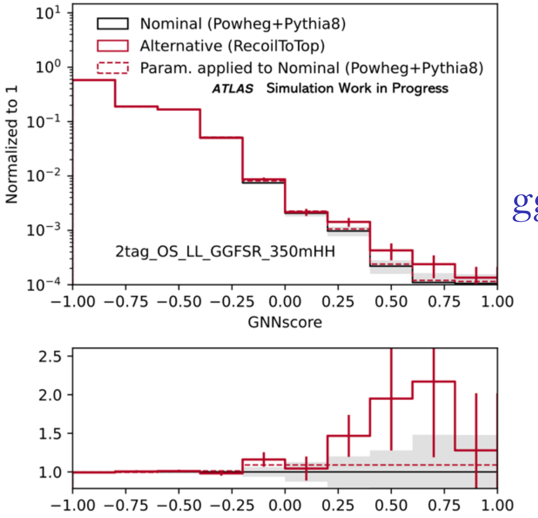
ggF High Region

3. pTHard: Unc. from the order of emissions which are related to the pT of the emissions i.e. hardenss



ggF High Region

4. ttbar Recoil to Top: Unc. from the recoil passed to the top.



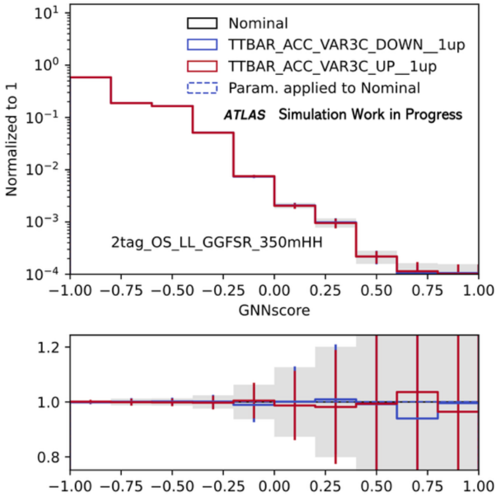
ggF High Region

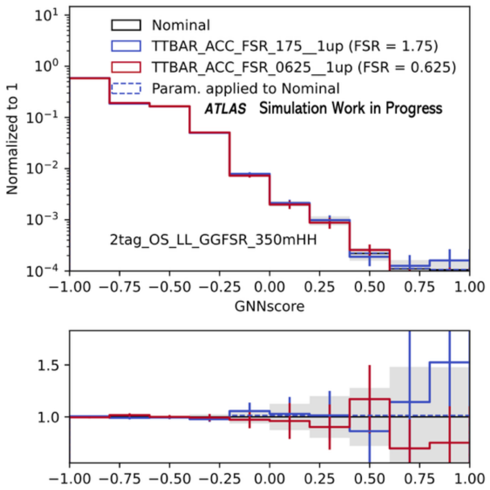
- These uncertainties are derived by comparing two samples produced using different generators.
- Only the significant ones are used in final fit
  - e.g. ttbar-recoil unc. in ggF High Region
- There are other unc. that are obtained by varying the generator weights.

ttbar Systematics in HadHad Signal Region

5. <u>ISR and FSR uncertainties</u>: Uncertainties from the Initial and the Final State radiation.

ISR

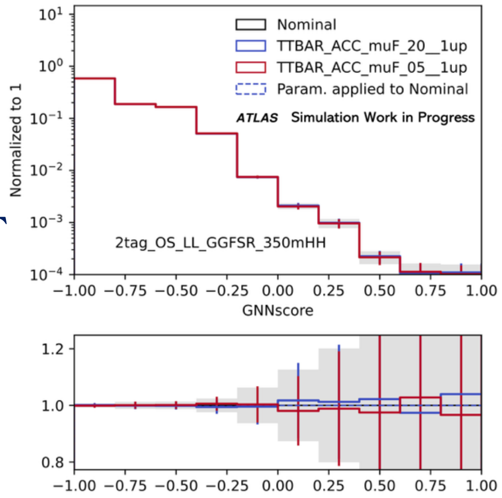


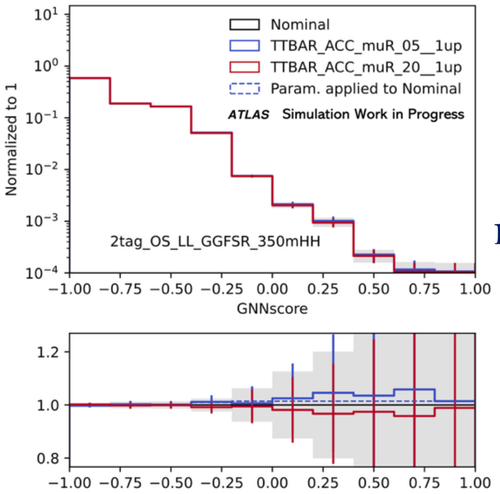


FSR

6. <u>Factorisation and Renormaliasation unc.</u>

muF





muR

ggF High Region

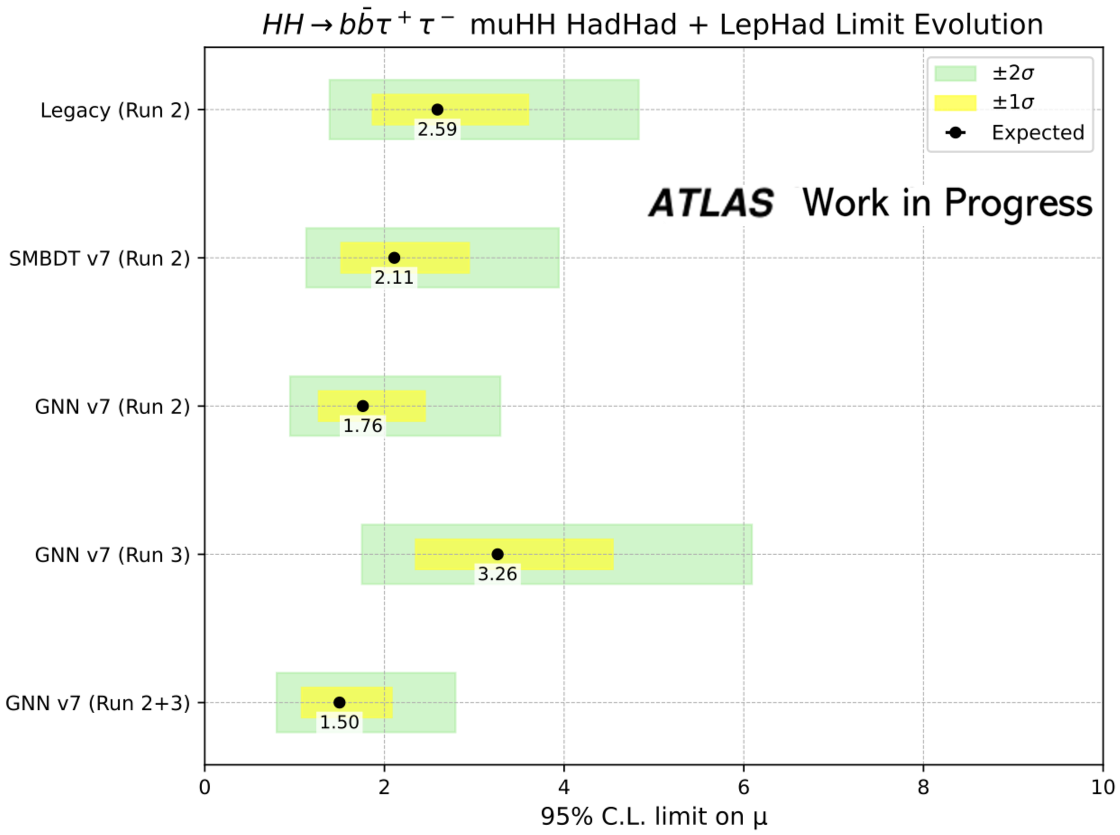
- To account for the the top uncertainties in the TopCR derived from the ZCR in the final fit, extrapolation uncertainties are required.
- These basically account for the differences in the MC in the CR to the SR.
- <u>The ttbar unc. for the HadHad signal region will go in the final fit.</u>

e.g. for ggF High Region

HDAMP	up: -4.3%, down: 4.3%
PS	up: -10.0%, down: 10.0%
PTHARD	up: -4.9%, down: 4.9%
RECOIL TO TOP	up: 2.2%, down: -2.2%

Summary

- A new Run-2+3 analysis with ATLAS detector is underway with many improvements already implemented at various levels (object definition, event selection, background estimate, MVA etc)
- The next steps for me:
  - Add the unc. to the final fit.
- These contributions in the ttbar modelling will help in constraining the ttbar background in the signal region.
- Significant improvement expected based on the preliminary results.



Upper limit on signal strength for GNNs using pre-fit Asimov datasets vs previous results.

Thank You!

Backup

Why Split ggF Region

- ggF region is split into low mHH (i.e. mHH < 350 GeV) and high mHH regions (i.e. mHH >= 350 GeV)
- This allows for better resolution of SM Higgs and BSM Higgs.

Selection Cuts (ToP_CR)

- Selections:
  - pass_TopEMuCR
  - Two b-jets: pcbt_GN2v01 >= 3 i.e. >85% WP
  - e_mu Opposite Sign

- Triggers:
  - pass_TopEMuCR = !ele1 && !mu1 && ele0->pt() > 40GeV && mu0->pt() > 40GeV

- Particle selection:
  - TWO_BJETS = (bjets0_pt > 45GeV && bjets1_pt > 45GeV)
  - mLL > 60 GeV
  - Lepton1_pt & Lepton2_pt > 40GeV
  - Hbbm = (bbtt_H_bb_m_{systematic.suffix}<40) || (bbtt_H_bb_m_{systematic.suffix}>210) (removed)