

# $HH \rightarrow b\bar{b}\tau^+\tau^-$ ATLAS analysis overview: experimental status and plans

Alessandra Betti  
on behalf of the ATLAS collaboration

Double Higgs Production at Colliders Workshop  
Fermilab, 6 September 2018



# Hunting the Higgs boson pair

- SM HH production cross-section very small: 33.41 fb at 13 TeV
- To study a rare process like HH production need to look at the H decay channels with high BRs:

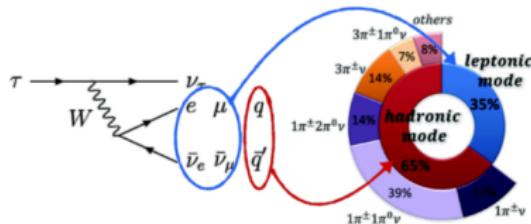
## HH-decay BRs:

- $b\bar{b}b\bar{b}$ : 33%
- $b\bar{b}WW$ : 25%
- $b\bar{b}\tau\tau$ : 7.4%
- $b\bar{b}\tau\tau$  has the third highest accessible BR
- it is relatively clean compared to other channels with higher BR but also larger background and more difficult reconstruction  
→ very promising channel!

	bb	WW	$\tau\tau$	ZZ	W
bb	33%				
WW	25%	4.6%			
$\tau\tau$	7.4%	2.5%	0.39%		
ZZ	3.1%	1.2%	0.34%	0.076%	
$\gamma\gamma$	0.26%	0.10%	0.029%	0.013%	0.0005%

ATLAS Run2  $b\bar{b}\tau\tau$  analysis is performed in two channels depending on the  $\tau$  lepton decay modes:

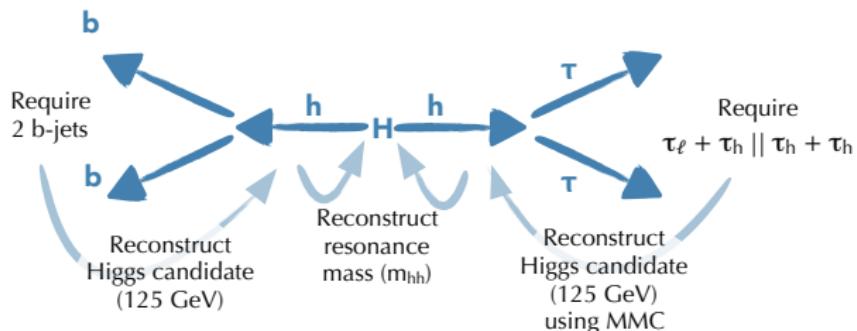
Leptonic+Hadronic (LepHad) channel (45.6%) and Hadronic+Hadronic (HadHad) channel (42%)



# Di-Higgs production in the $bb\tau\tau$ final state: what are we looking for?

Look for production of 2 Higgs bosons decaying one in  $bb$  and one in  $\tau\tau$

- 2 b-jets  
 $m_{BB} = m_h = 125 \text{ GeV}$
- 2 taus  
 $m_{\tau\tau} = m_h = 125 \text{ GeV}$



- Resonant case: reconstruct the invariant mass of the resonance from the 4-momenta of the 2 Higgs candidates  
 $m_X = m_{hh} = ?$   
→ constrain the  $bb$  and  $\tau\tau$  invariant mass to the Higgs mass (scaling the 4-vectors) in the calculation of  $m_{hh}$

## Signal hypothesis

**Resonant analysis:** resonances with mass from 260 GeV to 1 TeV

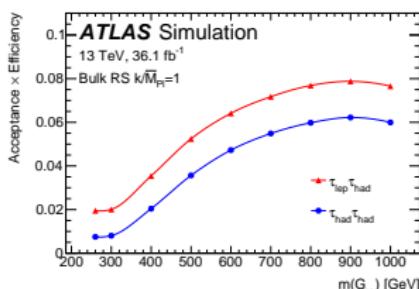
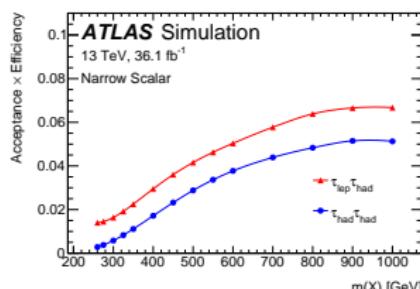
- Two Higgs Doublet Model (2HDM) for spin 0
- Randall-Sundrum graviton (RSG) model for spin 2

## Non-resonant analysis:

- SM non-resonant Higgs pair production
- Non-resonant Higgs pair production with varied Higgs self coupling  $\lambda_{hhh}$

## Analysis description: event selection

- **Data:**  $\simeq 36.1 \text{ fb}^{-1}$  recorded by ATLAS in 2015/2016 with LHC pp collisions at 13 TeV
- **Triggers:**
  - HadHad channel: di-tau and single tau triggers (DTT and STT)
  - LepHad channel: single lepton ( $e/\mu$ ) and lepton-tau triggers (SLT and LTT) (lowest unprescaled triggers)
    - detailed information in the "Trigger strategies for bbtautau" talk from Agni Bethani
- **Preselection region:** 2 signal taus (or 1 light lepton ( $e/\mu$ ) and 1 signal tau) with opposite sign of the charge (OS) and at least 2 jets
  - $m_{MMC} > 60 \text{ GeV}$ ,  $\text{Jet0 } p_T > 45 \text{ GeV}$ ,  $\text{Jet1 } p_T > 20 \text{ GeV}$
- **Signal region:** preselection + 2b-tagged jets
- **Control and validation regions:** Use 0b-tags, 1b-tag and same sign (SS) regions as validation regions for checking and understanding the background modeling
- tau-ID working point: "medium" (55% efficiency for 1-prong and 40% efficiency for 3-prongs)
- b-tagging working point: 70% efficiency

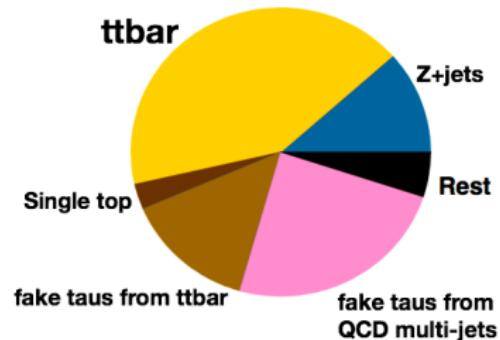


Acceptance x Efficiency for SM non-resonant:

- HadHad: 2.9%
- LepHad: 4.2%

# Backgrounds in the HadHad channel SR

- $t\bar{t}$ : 42% (with 2 real taus)
- fake taus from multijet: 25%
- Z+jets: 15%
- fake taus from  $t\bar{t}$ : 15%
- Rest (single top, W+jets, Diboson): 5%



- multijet tau fakes estimated with an ABCD data-driven fake factor (FF) method
- $t\bar{t}$  tau fakes from MC corrected with a fake-rate (FR) method from data
- Other backgrounds estimated from MC

# Multijet tau fakes estimation - HadHad channel

Background of jets faking taus from multijets estimated with ABCD data-driven FF method  
Define 4 regions depending on the tau leptons ID and charge:

A OS ID- $\tau$	B SS ID- $\tau$
C OS AntiID- $\tau$	D SS AntiID- $\tau$

Assuming:

$$\frac{N_A}{N_C} = \frac{N_B}{N_D}$$

Compute the **Fake Factors (FF)** as:

$$FF = \frac{N_B}{N_D}$$

as a function of leading and subleading tau  $p_T$ ,  
separate for 1p1p, 1p3p, 3p1p and 3p3p

And applying them to region C we obtain the  
template for the SR A:

$$N_A = \frac{N_B}{N_D} \times N_C = FF \times N_C$$

- A = SR
- ID= 2 signal taus
- Anti-ID= At least 1 anti-tau
- Anti-tau= tau candidate failing the signal tau requirements
- OS/SS= tau candidates with opposite/same charge sign

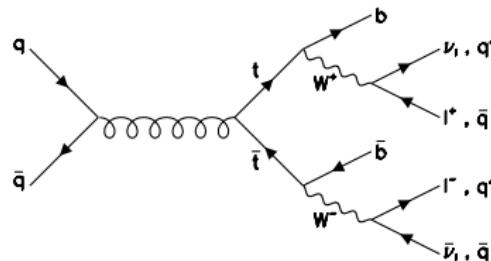
Contribution from all other backgrounds is  
subtracted

## $t\bar{t}$ tau fakes estimation - HadHad channel

Rate of jets faking taus (FR) in  $t\bar{t}$  events computed from data in LepHad  $t\bar{t}$  CR  
(LepHad preselection + 2btags +  $M_T^W > 80$  GeV)

- contribution from true  $t\bar{t}$  (with a real tau) and other backgrounds is subtracted
- defined as:

$$FR = \frac{N_{passID}}{N_{total}}$$



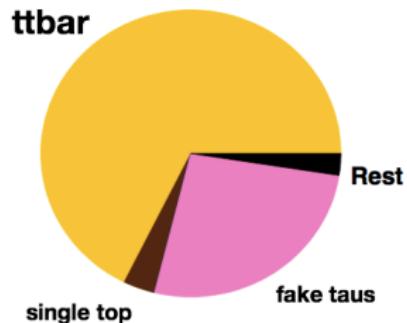
as a function of tau  $\eta$ , separate for 1p and 3p

in the HadHad SR  $t\bar{t}$  events from MC are split in true (T) and fake (F) by truth-matching the reconstructed tau to the truth lepton:

- TT: true component  $\rightarrow \simeq 26\%$  of our  $t\bar{t}$  is fake (at least one fake tau from a jet)
- TF, FT, FF: fake
- FRs computed in the LepHad CR applied to fake tau  $t\bar{t}$  events from MC that pass the HadHad channel selection to correct the MC prediction in the HadHad SR

# Backgrounds in the LepHad channel SR

- $t\bar{t}$ : 62% (with a real tau)
- fake taus from  $t\bar{t}$ , W+jets and QCD: 31% (mostly from  $t\bar{t}$ )
- Rest (Z+jets, single top, W+jets, diboson): 7%



- Background with jets faking taus estimated with a combined fake factor method from data (for  $t\bar{t}$ , multi-jets, W+jets)
- Other backgrounds estimated from MC

## tau fakes estimation - LepHad channel

Backgrounds with jets faking hadronic taus (from ttbar, multijets and W+jets) estimated with a combined fake factor method from data:

- FFs computed separately for each process ( $t\bar{t}$ , W+jets, multi-jets) in a dedicated CR
- contribution from true taus and other backgrounds is subtracted
- defined as:

$$FF = \frac{N_{tau}}{N_{anti-tau}}$$

binned as a function of tau  $p_T$ , separate for 1p and 3p

then a combined fake factor is derived:

$$FF(comb) = FF(QCD) \times rQCD + FF(W/t\bar{t}) \times (1 - rQCD)$$

where rQCD is defined as the fraction of fakes from multi-jets in the anti-ID tau region:

$$rQCD = \frac{N(multijets,data)}{N(data) - N(true\ tau,MC)}$$
 and  $N(multijets,data) = data - (true\ tau\ MC + fake\ tau\ MC)$

Combined FFs applied to anti-tau events from data (SR selection apart from tauID) to estimate the fake tau background in the SR

## Systematic uncertainties on the fake taus background estimation

Data-driven methods for the estimation of backgrounds with jets faking taus come with systematic uncertainties from several different sources:

- **Statistics:** vary FFs and FRs up and down by their statistical uncertainties
- **true MC contamination in CRs:** vary the subtracted true  $t\bar{t}$  MC by propagating the cross-section, CP and modeling shape uncertainties and vary other small subtracted MC backgrounds up and down by 50%
- **extrapolation from CR to SR:** calculate FFs with alternative definition of the CR and take the difference
- **fake composition variation between anti-ID and ID regions:** closure tests in VRs

# Analysis Strategy: Boosted Decision Tree (BDT)

The analysis makes use of BDTs to separate signal and background

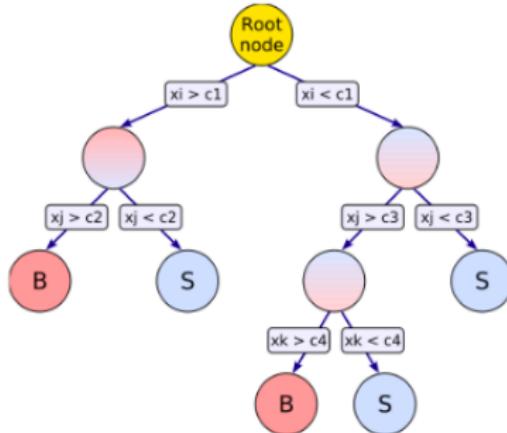
- Train each signal hypothesis against sum of main backgrounds

## Resonant analysis:

- One training for each mass point (masses from 260 GeV to 1 TeV) for 2HDM and Graviton model to have the best sensitivity at each mass
- In each training three neighbouring mass points are combined to ensure sensitivity between mass points

## Non-resonant analysis:

- Training on the SM non-resonant di-Higgs signal
- Training on the non-resonant signal with  $\lambda_{HHH}/\lambda_{SM} = 20$  for the  $\lambda$  variation limits



**Strategy:** Use the BDT score distribution as final discriminant variable for the limit setting fit

# BDT input variables

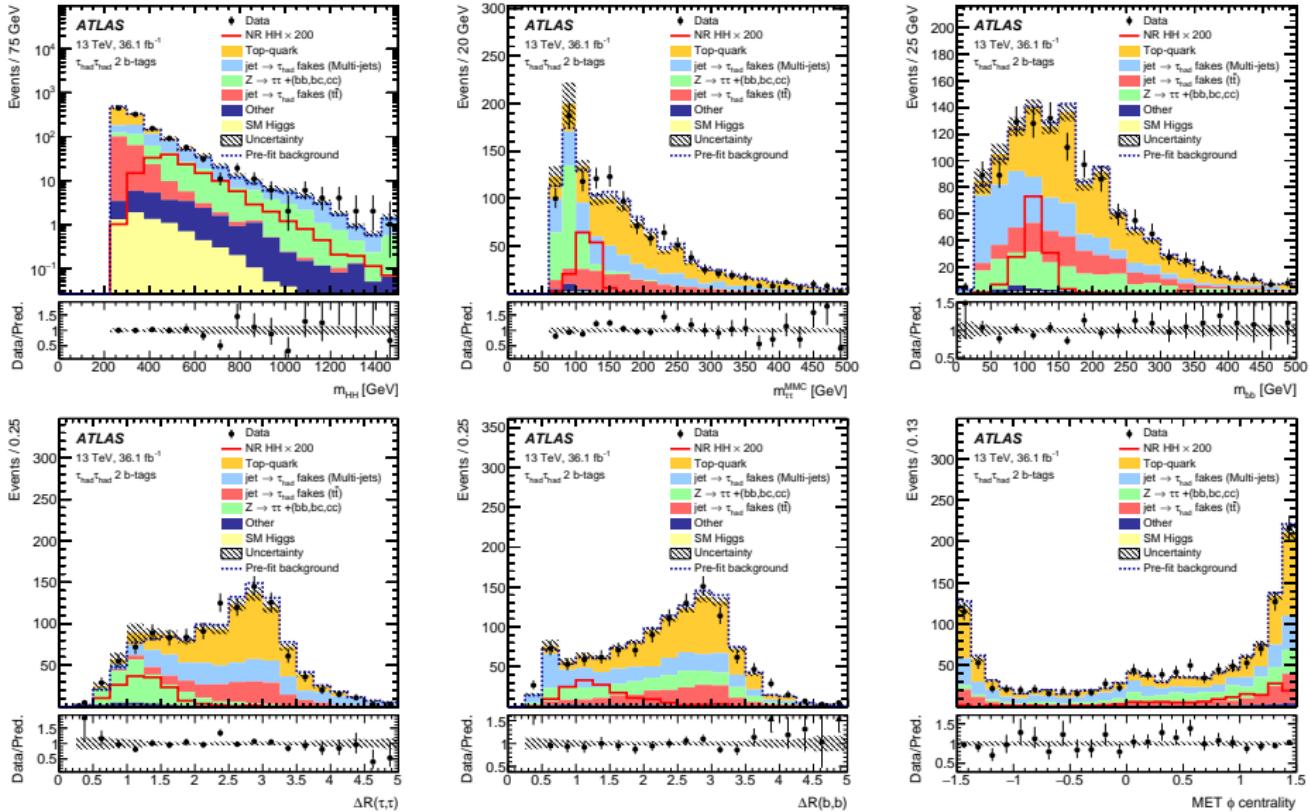
- BDT input variables chosen to exploit the characteristics of signal events, where one can reconstruct the two Higgs boson candidates in the  $b\bar{b}\tau^+\tau^-$  final state
- List of variables optimized for each channel and event category, for resonant and non-resonant signals

Variable	$\tau_{\text{lep}}\tau_{\text{had}}$ channel (SLT resonant)	$\tau_{\text{lep}}\tau_{\text{had}}$ channel (SLT non-resonant & LTT)	$\tau_{\text{had}}\tau_{\text{had}}$ channel
$m_{HH}$	✓	✓	✓
$m_{\tau\tau}^{\text{MMC}}$	✓	✓	✓
$m_{bb}$	✓	✓	✓
$\Delta R(\tau, \tau)$	✓	✓	✓
$\Delta R(b, b)$	✓	✓	✓
$E_T^{\text{miss}}$	✓		
$E_T^{\text{miss}}$ $\phi$ centrality	✓		✓
$m_T^W$	✓	✓	
$\Delta\phi(H, H)$	✓		
$\Delta p_T(\text{lep}, \tau_{\text{had-vis}})$	✓		
Sub-leading $b$ -jet $p_T$	✓		

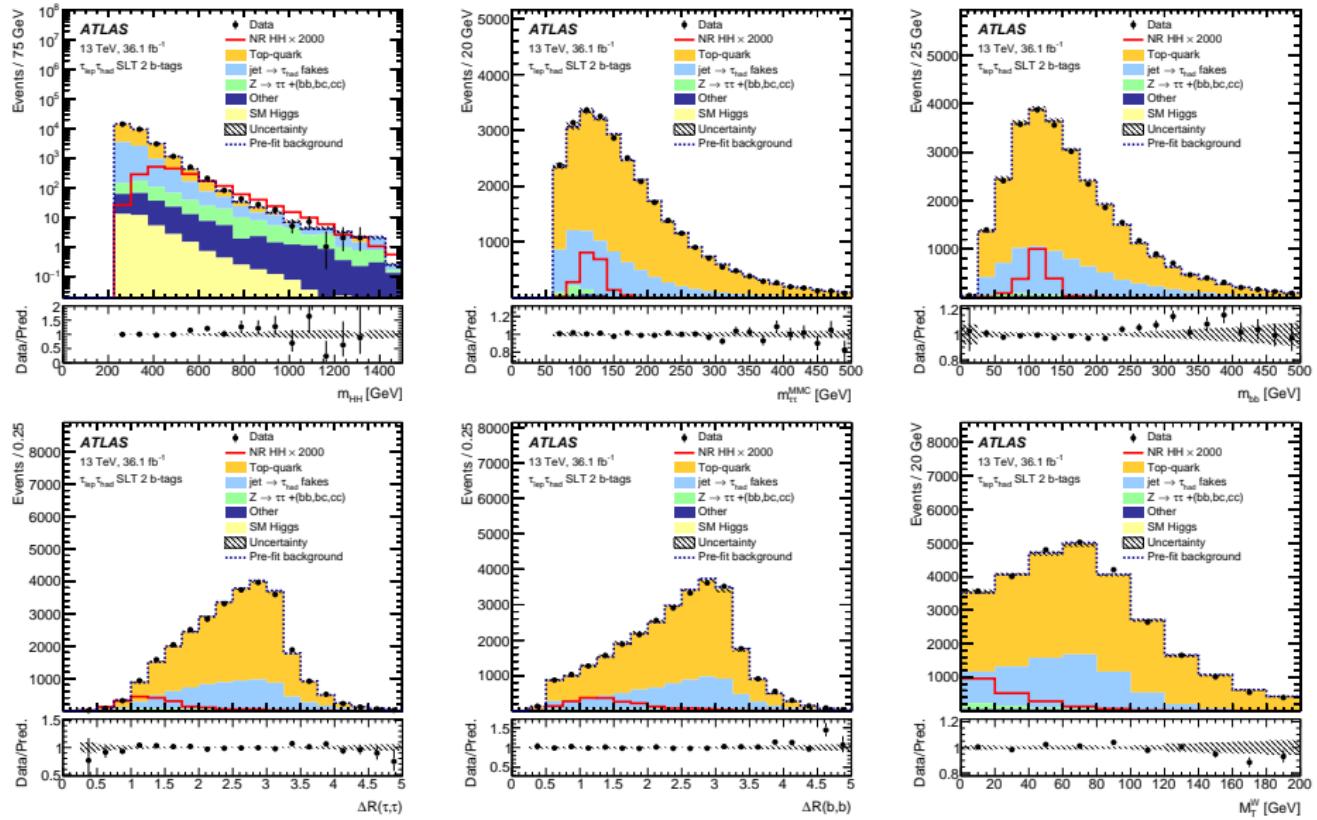
$E_T^{\text{miss}}$   $\Phi$  centrality: relative angular position of the  $E_T^{\text{miss}}$  relative to the taus in the transverse plane

$$(A+B)/(\sqrt{A^2 + B^2}), \text{ where } A = \sin(\phi_{E_T^{\text{miss}}} - \phi_{\tau_2})/\sin(\phi_{\tau_1} - \phi_{\tau_2}), B = \sin(\phi_{\tau_1} - \phi_{E_T^{\text{miss}}})/\sin(\phi_{\tau_1} - \phi_{\tau_2})$$

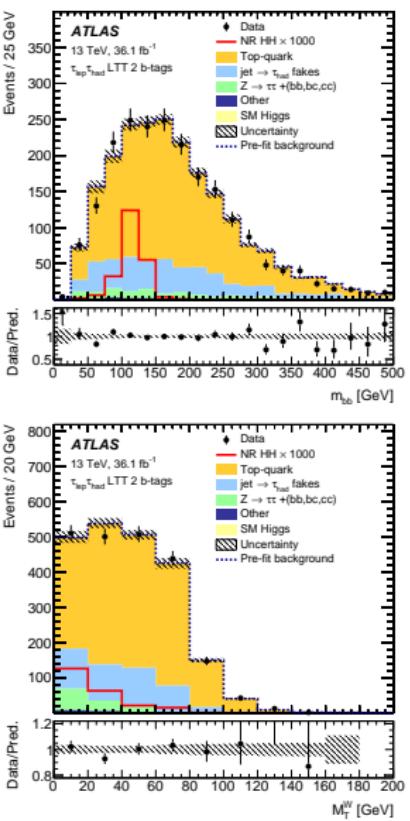
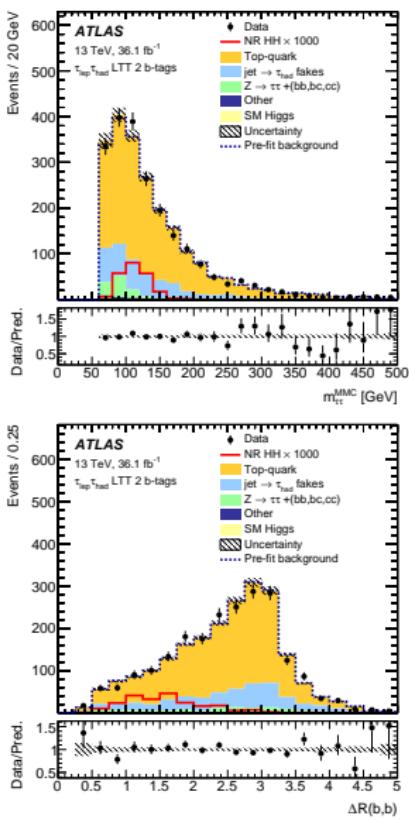
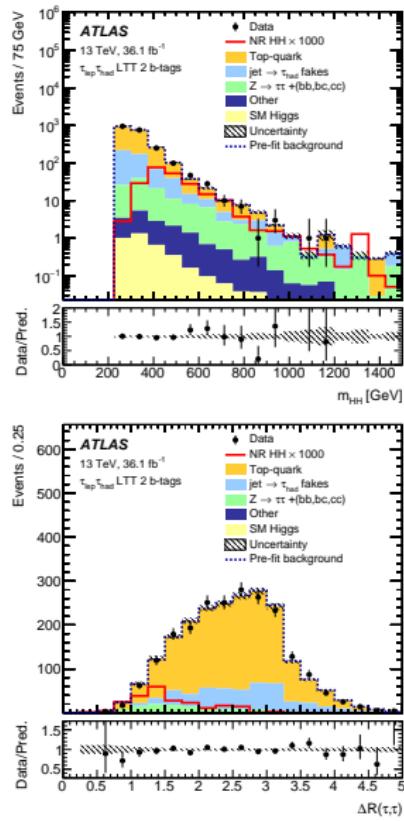
## BDT input variables - HadHad channel



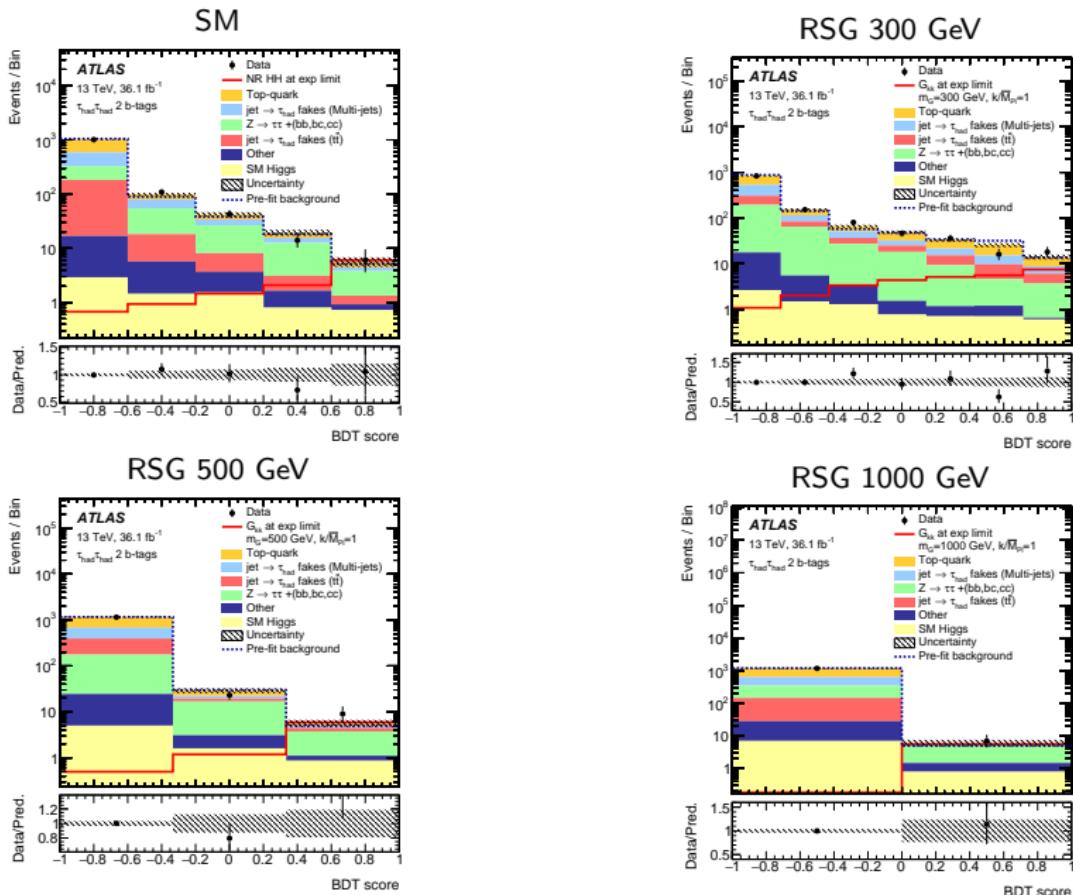
## BDT input variables - Lephad channel (SLT) (non-resonant)



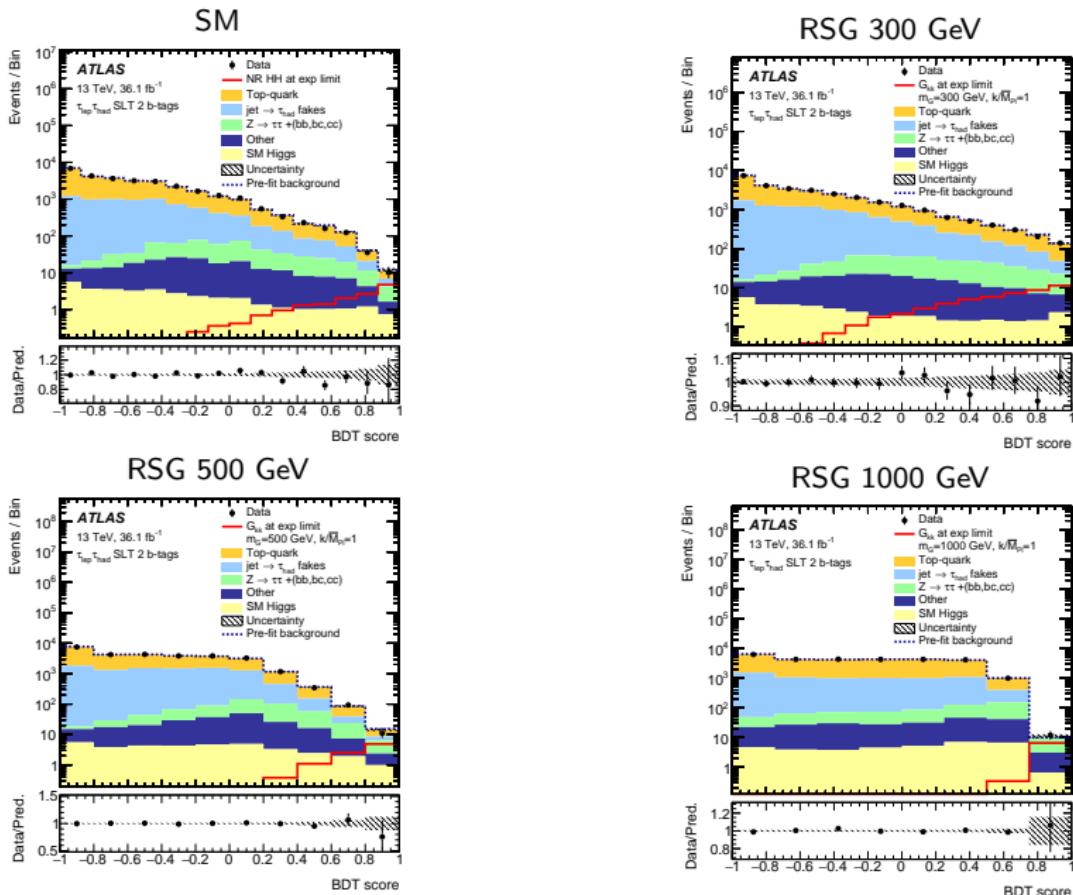
## BDT input variables - Lephad channel (LTT)



# BDT distributions - HadHad channel

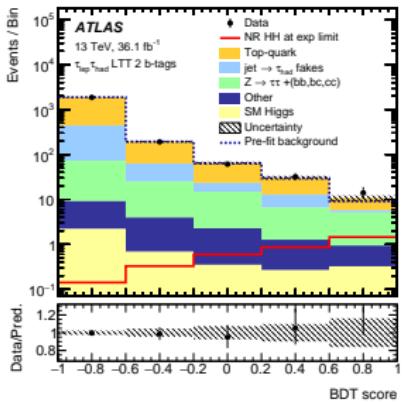


# BDT distributions - LepHad channel (SLT)

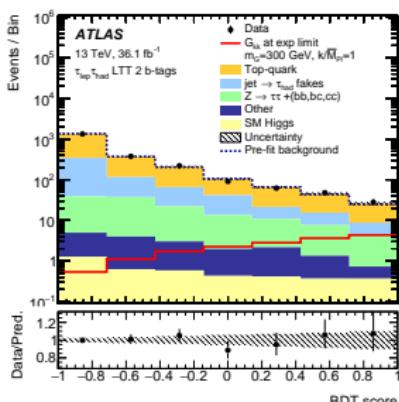


# BDT distributions - LepHad channel (LTT)

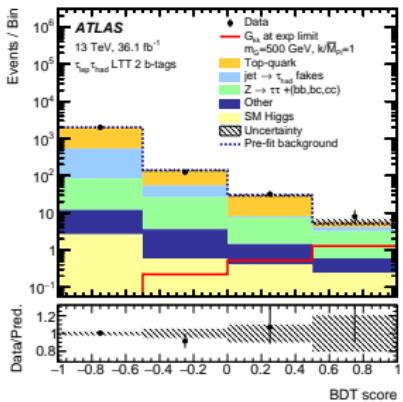
SM



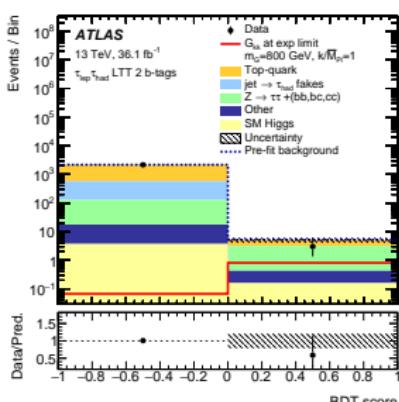
RSG 300 GeV



RSG 500 GeV



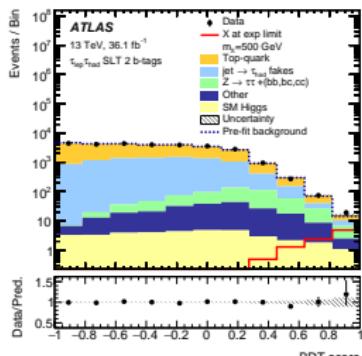
RSG 800 GeV



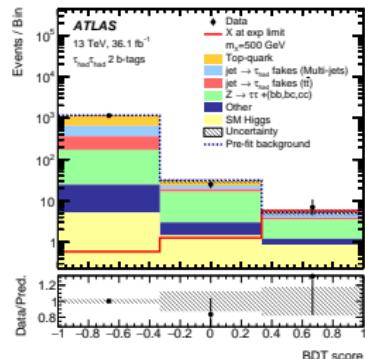
# Fit setup

- Fit of the full BDT distribution
- Simultaneous fit of 3 BDT distributions (HadHad, LepHad SLT and LepHad LTT) for each signal hypothesis
- Included in the fit also a 1 bin CR for the normalization of the Z+HF background
- 2 freely floating normalizations:  
Z+HF constrained by the dedicated CR and  $t\bar{t}$  constrained by the low BDT region of the LepHad SLT SR:
  - Z+HF =  $1.34 \pm 0.16$
  - $t\bar{t}$  =  $1.06 \pm 0.13$

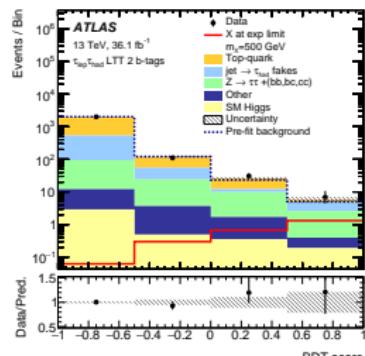
## 2HDM 500 GeV - LepHad SLT



## 2HDM 500 GeV - HadHad



## 2HDM 500 GeV - LepHad LTT

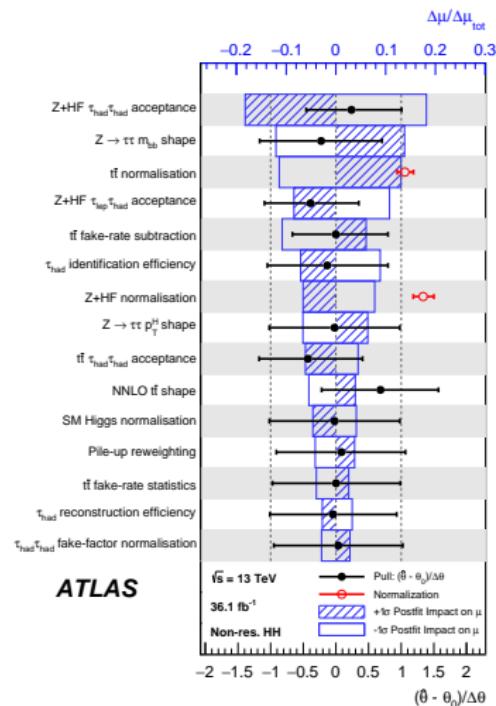


## Systematic uncertainties

- Analysis dominated by statistical uncertainties on data
  - Percentage uncertainties on the simulated non-resonant signal strength assuming a cross-section times branching fraction equal to the 95% CL expected limit of 14.8 times the SM expectation:

Source	Uncertainty (%)
Total	$\pm 54$
Data statistics	$\pm 44$
Simulation statistics	$\pm 16$
Experimental Uncertainties	
Luminosity	$\pm 2.4$
Pileup reweighting	$\pm 1.7$
$\tau_{\text{had}}$	$\pm 16$
Fake- $\tau$ estimation	$\pm 8.4$
$b$ -tagging	$\pm 8.3$
Jets and $E_T^{\text{miss}}$	$\pm 3.3$
Electron and muon	$\pm 0.5$
Theoretical and Modeling Uncertainties	
Top	$\pm 17$
Signal	$\pm 9.3$
$Z \rightarrow \tau\tau$	$\pm 6.8$
SM Higgs	$\pm 2.9$
Other backgrounds	$\pm 0.3$

- Fractional impact of systematic uncertainties for the fitted non-resonant HH signal-strength parameter  $\mu$



## Results: non-resonant

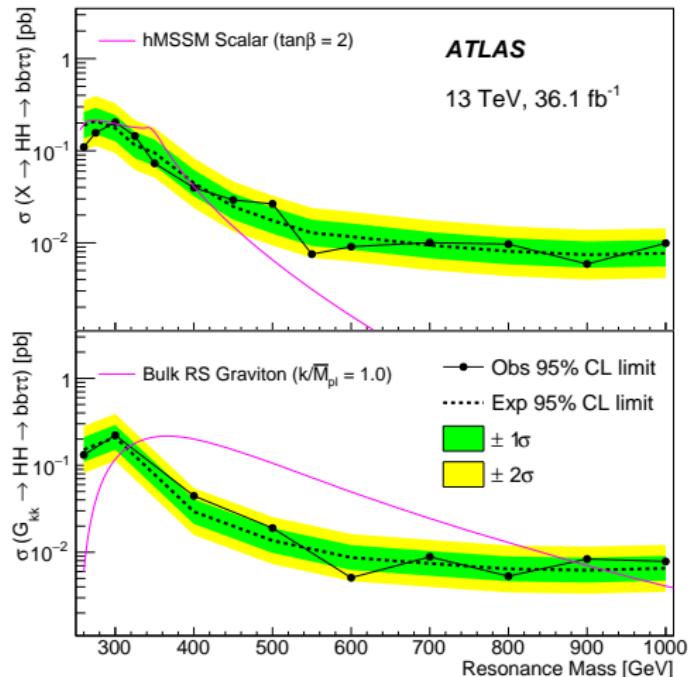
Upper limit on the SM di-Higgs production cross-section

		Observed	$-1\sigma$	Expected	$+1\sigma$
$\tau_{\text{lep}} \tau_{\text{had}}$	$\sigma(HH \rightarrow bb\tau\tau) [\text{fb}]$	57	49.9	69	96
	$\sigma/\sigma_{\text{SM}}$	23.5	20.5	28.4	39.5
$\tau_{\text{had}} \tau_{\text{had}}$	$\sigma(HH \rightarrow bb\tau\tau) [\text{fb}]$	40.0	30.6	42.4	59
	$\sigma/\sigma_{\text{SM}}$	16.4	12.5	17.4	24.2
Combination	$\sigma(HH \rightarrow bb\tau\tau) [\text{fb}]$	30.9	26.0	36.1	50
	$\sigma/\sigma_{\text{SM}}$	12.7	10.7	14.8	20.6

Observed (expected) an upper limit on the di-Higgs production cross-section of  $12.7(14.8) \times \sigma_{\text{SM}}$   
→ current best upper limit on this process!

## Results: resonant

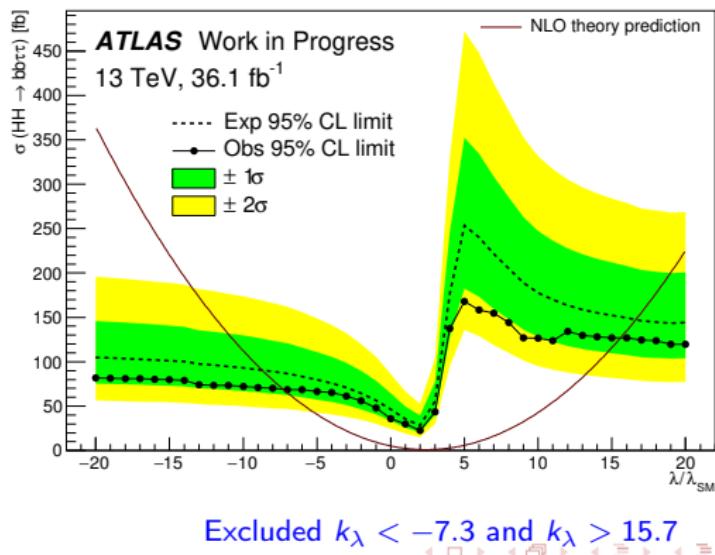
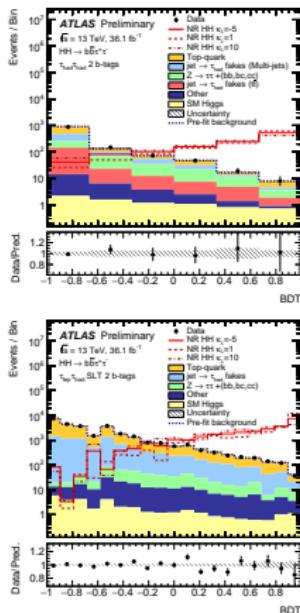
Upper limits on  $\sigma_X \times BR_{X \rightarrow hh \rightarrow bb\tau\tau}$  as a function of the mass for 2HDM and RSG



- Excluded resonances  $X$  in the mass range  $305 \text{ GeV} < m_X < 402 \text{ GeV}$  in the simplified hMSSM minimal supersymmetric model for  $\tan\beta = 2$
- Excluded bulk Randall–Sundrum gravitons  $G_{KK}$  in the mass range  $325 \text{ GeV} < m_{G_{KK}} < 885 \text{ GeV}$  for  $k/M_{Pl} = 1$

# Results: non-resonant with varied Higgs self coupling $\lambda$

- Non-resonant HH production with variations of the Higgs self coupling  $\lambda_{HHH}$   
→ detailed information in the "ATLAS HH combination results" talk from Petar Bokan
- BDT trained on the  $\lambda_{HHH}/\lambda_{SM} = 20$  signal used for all values  
(gives better sensitivity compared to the SM BDT training over the full  $\lambda$  scan range, compatible with a  $\lambda$  specific BDT training)
- Fit the full BDT distributions to set **upper limits on the HH production cross-section as a function of the coupling**



# Summary and plans for end of Run2

## Summary:

- ATLAS di-Higgs to  $bb\tau\tau$  analysis now public using 2015/2016 data ( $\simeq 36 \text{ fb}^{-1}$ ): arXiv:1808.00336v1, submitted to Phys. Rev. Lett.
- Use of different data-driven methods for estimation of backgrounds with jets faking taus
- Use of BDTs for separating signal and background
- Fit of the BDT distribution to get upper limits on:
  - the cross-section for non-resonant SM di-Higgs production
  - the cross-section for non-resonant di-Higgs production as a function of the Higgs self coupling
  - the cross-section for the production of a heavy Higgs or a graviton decaying into two Higgs bosons in the mass range 260 GeV - 1 TeV

## Plans for end of Run2:

- Working on improvements for a new analysis with the full Run2 dataset
- Focusing on obtaining the best possible result for the non-resonant SM di-Higgs production:
  - Testing several multivariate methods and analysis strategies to:
    - improve overall performance
    - reduce the number of trainings/generated signal samples
    - reduce impact of systematics
    - constrain better individual bkg
  - Testing including additional SRs like 1btag region or loosening the 2btags region
  - Including new categories and channels like VBF and LepLep

## Back-up slides

## Signal samples:

- SM non-resonant: MadGraph5\_aMC@NLO at next-to-leading order (NLO) + Herwig++
- Resonant 2HDM and RSG: MadGraph5\_aMC@NLO at leading order (LO) + Pythia8

## Background samples:

- $t\bar{t}$  and single-top: Powheg + Pythia6
- Z+jets, W+jets, Diboson, Drell-Yan: Sherpa 2.2.1
- quark-induced ZH: Pythia8
- gluon-induced ZH: Powheg + Pythia8
- ttH: MadGraph5\_aMC@NLO + Pythia 8

## Object selection

**Electrons:** Tight ID

SLT:  $\text{pt} > 25 \text{ or } 27 \text{ GeV}$

LTT:  $\text{pt} > 18 \text{ GeV} \text{ & } \text{pt} < 25(27) \text{ GeV}$

**Jets:** AntiKt4,

$\text{pt} > 45, 20 \text{ GeV}$

in LTT: at least one jet with  $\text{pt} > 80 \text{ GeV}$  due to  
L1 requirement

**Muons:** Medium ID

SLT:  $\text{pt} > 25 \text{ or } 27 \text{ GeV}$

LTT:  $\text{pt} > 15 \text{ GeV} \text{ & } \text{pt} < 25(27) \text{ GeV}$

**Hadronic Taus:** Medium ID

LepHad: SLT:  $\text{pt} > 20 \text{ GeV}$ , LTT:  $\text{pt} > 30 \text{ GeV}$

HadHad: STT:  $\text{pt} > 180, 140, 100/20 \text{ GeV}$

DTT:  $\text{pt} > 40, 30 \text{ GeV}$

di-tau mass (MMC)  $> 60 \text{ GeV}$

opposite charge lepton/ $\tau$  - $\tau$  pairs (OS)

at least 2 (b-tagged) jets

lephad channel: veto additional leptons and taus

hadhad channel: veto any leptons and additional taus

## Control regions

**ttbar CR** for calculation of HadHad channel ttbar Fake Rates:

- LepHad channel preselection
- 2btags
- $M_T^W > 80 \text{ GeV}$

**ttbar CR** for calculation of LepHad channel ttbar Fake Factors:

- LepHad channel preselection
- 2btags
- $M_T^W > 40 \text{ GeV}$

**W+jets CR** for calculation of LepHad channel ttbar Fake Factors:

- LepHad channel preselection
- 0btags
- $M_T^W > 40 \text{ GeV}$

**Multijet CR** for calculation of LepHad channel ttbar Fake Factors:

- LepHad channel preselection but inverse lepton isolation
- 1btag

Z( $\mu\mu$ )+HF CR included in the fit to constrain the Z+HF normalization:

- single muon trigger
- =2  $\mu$  with  $p_T > 27$  GeV
- $\geq 2$  b-jets with  $p_T > 45, 20$  GeV
- $81 < m_{\mu\mu} < 101$  GeV
- Higgs veto  $80 < m_{bb} < 140$  GeV

# Systematic uncertainties on the fake taus background estimation - HadHad channel

Data-driven methods for the estimation of backgrounds with jets faking taus come with systematic uncertainties from several different sources:

- **Statistics:** vary FFs and FRs up and down by their statistical uncertainties
- **true MC contamination in CRs:**
  - multijet FFs: vary small subtracted MC background by 50%
  - $t\bar{t}$  FRs: vary the subtracted true  $t\bar{t}$  MC by propagating the cross-section, CP and modeling shape uncertainties

## multijet FFs

- **fake composition variation between anti-ID and ID regions:** closure test in 1b-tag OS region and parameterize difference as a function of tau  $p_T$
- **extrapolation from 1 b-tag to 2 b-tags:** vary the normalization transfer factor up and down by the statistical uncertainty
- **extrapolation from SS to OS:** compare SS/OS FFs in a QCD enriched region

## $t\bar{t}$ FRs

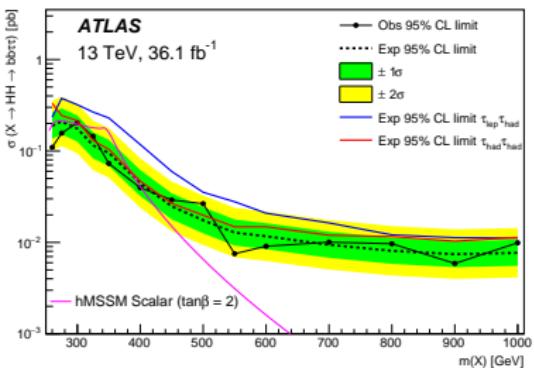
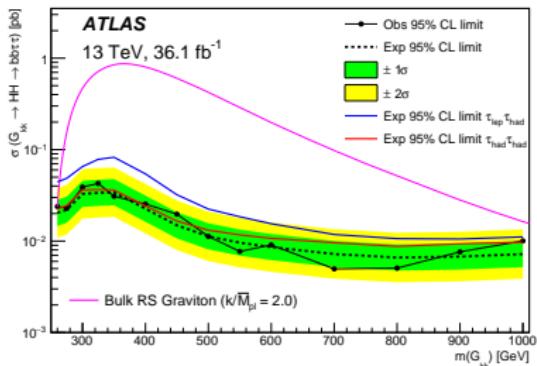
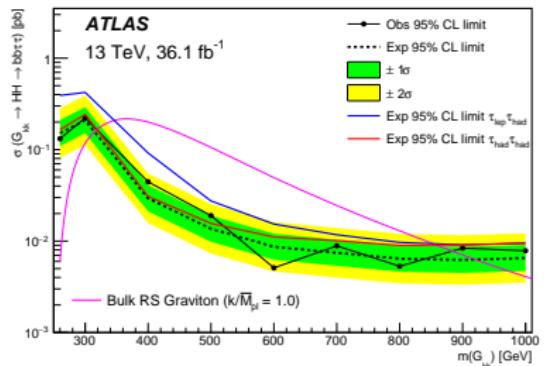
- **extrapolation to low  $M_T^W$  from high  $M_T^W$  CR:** reduce  $M_T^W$  in the definition of the CR and take the difference
- **trigger selection systematic:**  $p_T$  dependent variation to account for difference in  $p_T$  distribution from combining STT+DTT

# Systematic uncertainties on the fake taus background estimation - LepHad channel

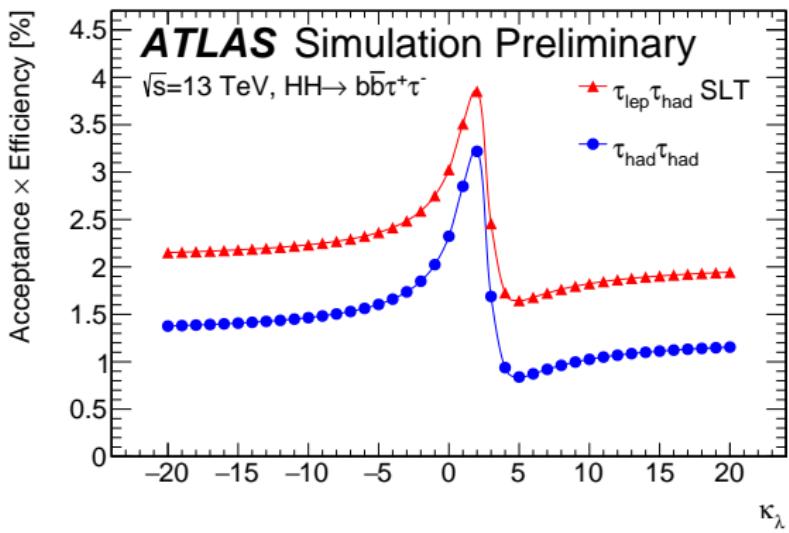
Data-driven methods for the estimation of backgrounds with jets faking taus come with systematic uncertainties from several different sources:

- **Statistics:** vary FFs up and down by their statistical uncertainties
- **true MC contamination in CRs:** vary the subtracted true  $t\bar{t}$  MC by propagating the cross-section, CP and modeling shape uncertainties and vary other small subtracted MC backgrounds by 50%
- **fake composition variation between anti-ID and ID regions:** closure test in 2b-tags SS region and parameterize difference as a function of tau  $m_{HH}$
- **extrapolation of  $t\bar{t}$  and W+jets FFs from CR to SR:** calculate MC-based FFs in SR and CR and take the difference
- **extrapolation of multijet FFs from CR to SR:** calculate FFs in a CR defined using SS (not inverse lepton isolation) and take the difference

# Results: resonant

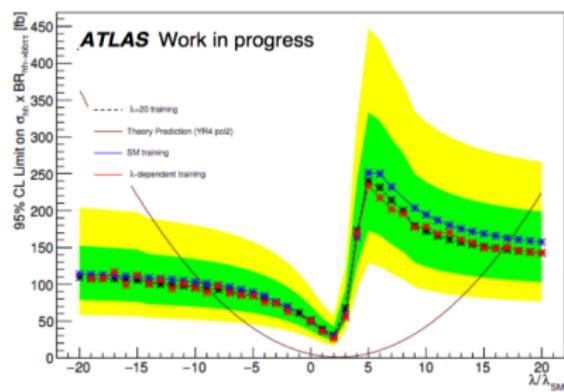


# Acceptance $\times$ Efficiency vs $k_\lambda$

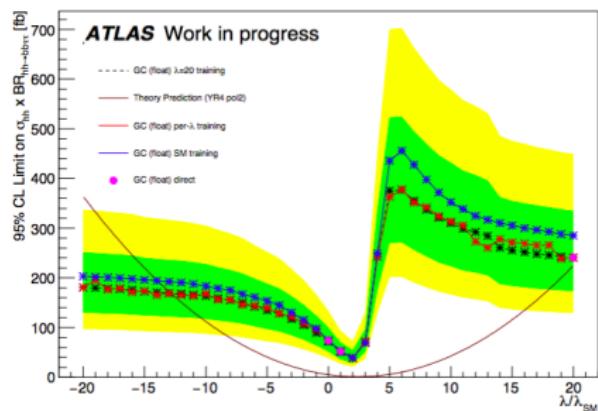


# BDT choice for $k_\lambda$ variation limits

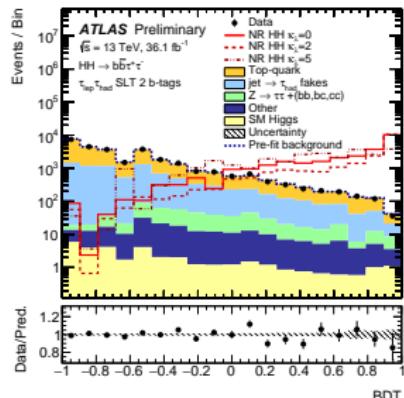
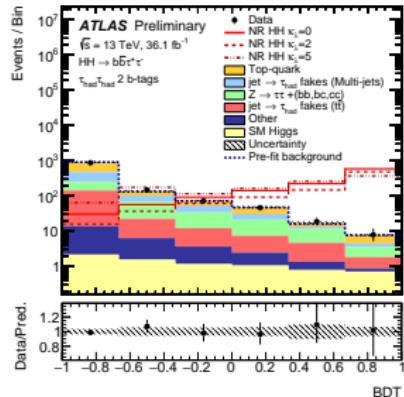
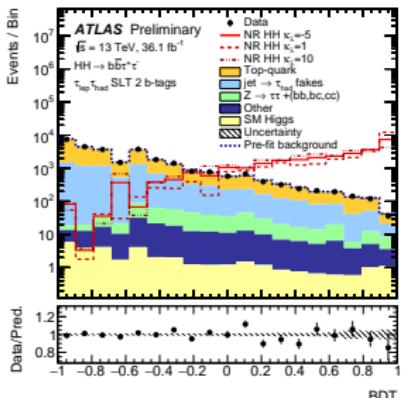
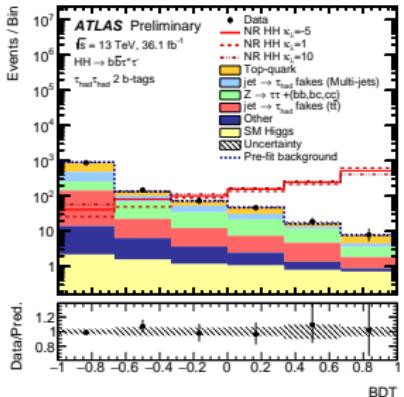
HadHad



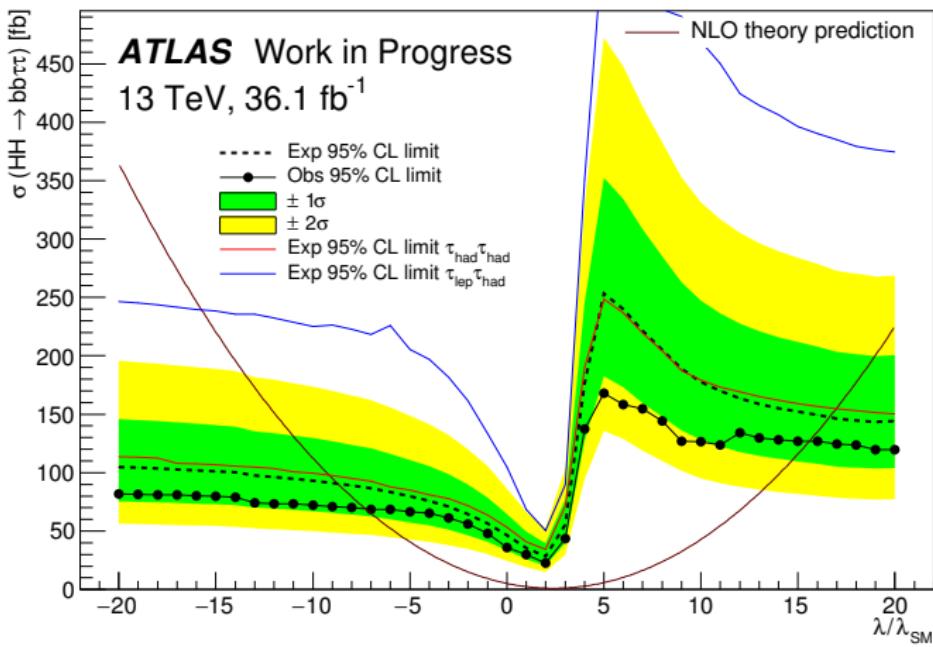
LepHad



# BDT distributions - $k_\lambda$ variations

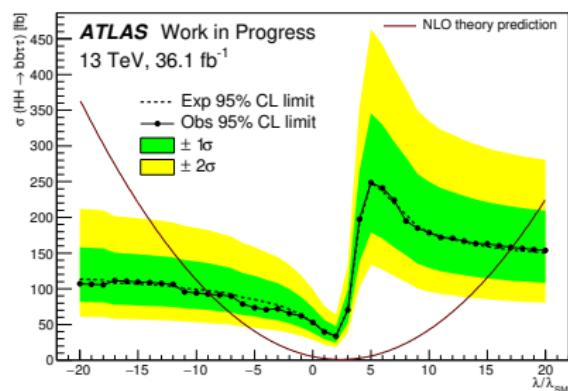


# $k_\lambda$ variation limits



# $k_\lambda$ variation limits

HadHad



LepHad

