

# Search for the Standard Model Higgs boson produced in association with top quarks and decaying into $b\bar{b}$ in pp collisions with ATLAS

Nihal BRAHIMI

On behalf of the ATLAS collaboration

Top LHC France, Grenoble



April 25, 2019

Search for the standard model Higgs boson produced in association  
with top quarks and decaying into a  $b\bar{b}$  pair in  $pp$  collisions  
at  $\sqrt{s} = 13$  TeV with the ATLAS detector

M. Aaboud *et al.*\*  
(ATLAS Collaboration)

(Received 27 December 2017; published 30 April 2018)

A search for the standard model Higgs boson produced in association with a top-quark pair,  $t\bar{t}H$ , is presented. The analysis uses  $36.1 \text{ fb}^{-1}$  of  $pp$  collision data at  $\sqrt{s} = 13$  TeV collected with the ATLAS detector at the Large Hadron Collider in 2015 and 2016. The search targets the  $H \rightarrow b\bar{b}$  decay mode. The selected events contain either one or two electrons or muons from the top-quark decays, and are then categorized according to the number of jets and how likely these are to contain  $b$ -hadrons. Multivariate techniques are used to discriminate between signal and background events, the latter being dominated by  $t\bar{t} + \text{jets}$  production. For a Higgs boson mass of 125 GeV, the ratio of the measured  $t\bar{t}H$  signal cross-section to the standard model expectation is found to be  $\mu = 0.84^{+0.64}_{-0.61}$ . A value of  $\mu$  greater than 2.0 is excluded at 95% confidence level (C.L.) while the expected upper limit is  $\mu < 1.2$  in the absence of a  $t\bar{t}H$  signal.

DOI: [10.1103/PhysRevD.97.072016](https://doi.org/10.1103/PhysRevD.97.072016)

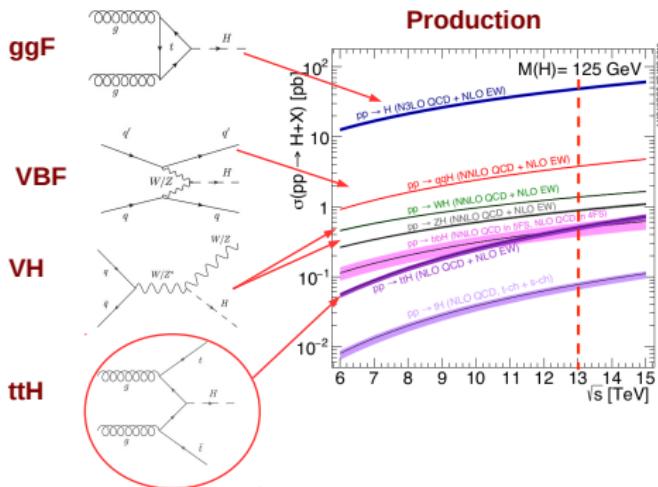
DOI: [Phys. Rev. D 97, 072016 \(2018\)](https://doi.org/10.1103/PhysRevD.97.072016)

This talk summarizes:

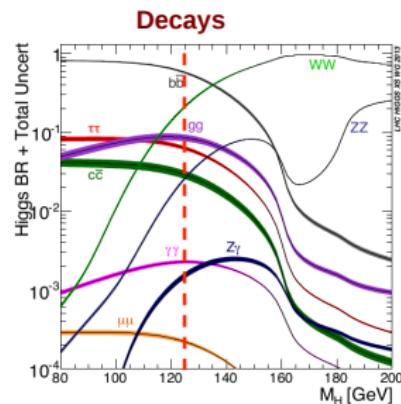
⇒ the  $t\bar{t}H(b\bar{b})$  search results obtained with  $36.1 \text{ fb}^{-1}$  of  $pp$  collision data @13 TeV

# Higgs boson @the LHC

arXiv:1610.07922



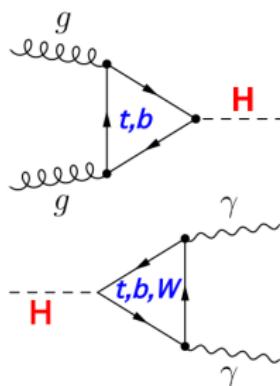
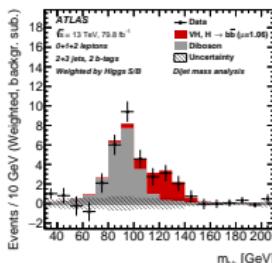
arXiv:1307.1347



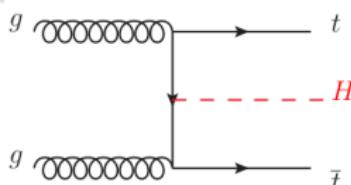
- 4 main production modes of the SM Higgs @LHC.  
 $\Rightarrow t\bar{t}H \sim 1\%$  of the total Higgs boson production cross section
- Several decay modes accessible:  $H \rightarrow bb$  largest BR @125 GeV.

# Higgs boson measurements and Top Yukawa coupling

Phys. Lett. B 786 (2018) 59

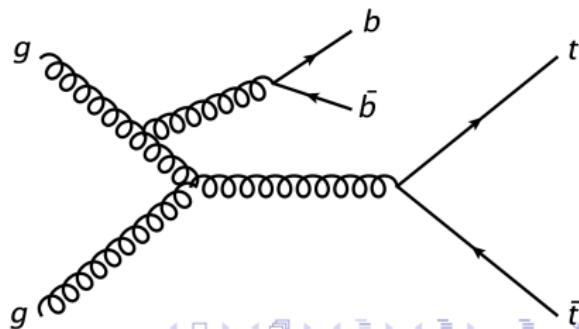
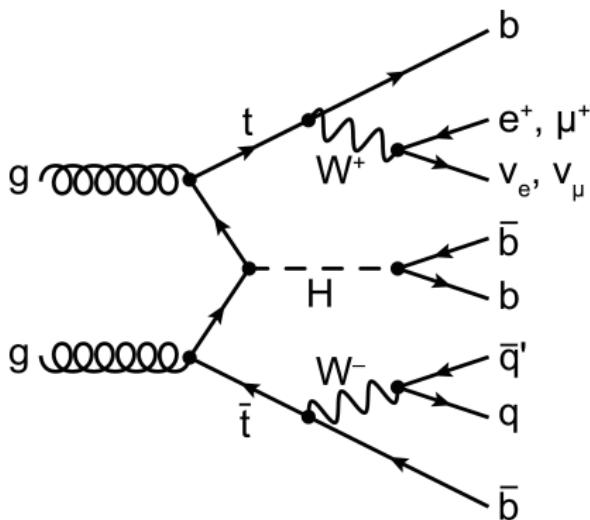


- Higgs boson discovered in 2012 by ATLAS and CMS.
- Higgs decays to bosons observed:  
     $\Rightarrow \gamma\gamma, WW^*, ZZ^*$
- Higgs couplings to 3<sup>rd</sup> generation fermions established:  
     $\Rightarrow H \rightarrow \tau\tau, H \rightarrow bb, t\bar{t}H$  (See Nazlim's talk)
- Indirect constraints on top Yukawa coupling prior to  $t\bar{t}H$  observation:  
     $\Rightarrow ggF$  and  $H \rightarrow \gamma\gamma$ , Run 1 results ([arXiv:1606.02266](https://arxiv.org/abs/1606.02266))
- Top quark heaviest in the SM  
     $\Rightarrow$  strongest Yukawa coupling ( $y_t \sim 1$ ).
- Plays a major role in the stability of the SM at high energies.
- $t\bar{t}H$  production: best way to have direct access to  $y_t$ .



# $t\bar{t}H(b\bar{b})$ analysis

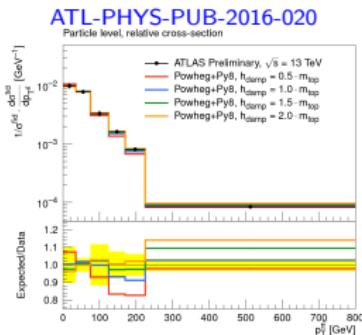
- Profits from the large  $H \rightarrow b\bar{b}$  branching fraction (58%)  
⇒ Contribution from other decay modes is negligible.
- Complex final state:  
⇒ 4 b-jets, 1 or 2 leptons ( $e/\mu$ ), jets from hadronic W or radiation (additional)
- Channel dominated by the overwhelming  $t\bar{t} + \text{jets}$  background.



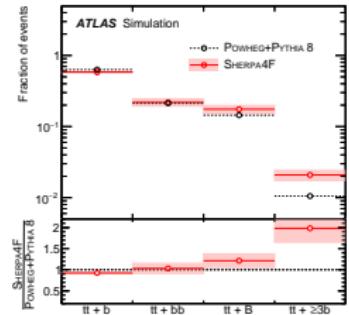
# $t\bar{t} + \text{jets}$ background in the $t\bar{t}H(b\bar{b})$ channel

- Most dominant background ( $> 90\%$ ).
- Split into 3 categories based on the flavor of additional jets:  $t\bar{t}+ \geq 1b$  (irreducible),  $t\bar{t}+ \geq 1c$  and  $t\bar{t} + \text{light}$
- Extensive modeling studies to choose and tune the MC generators based on the 7/8/13 TeV data.
  - ⇒ Powheg+Pythia8 is chosen as nominal.
  - ⇒ Normalized to NNLO+NNLL cross section.

- $t\bar{t}+ \geq 1b$  further divided into 4 sub-components.
- Relative contribution from each subcategory reweighted to match a 4F NLO  $ttbb$  prediction with massive b-quarks and  $g \rightarrow b\bar{b}$  from ME (Sherpa+OpenLoops)

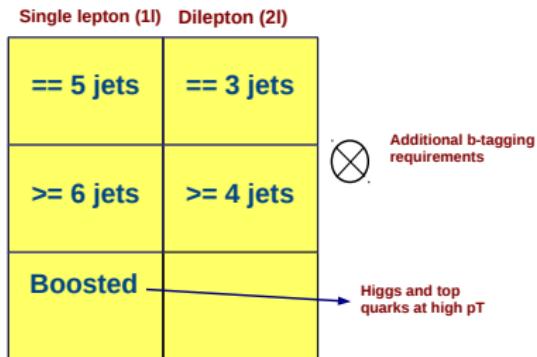


Phys. Rev. D 97, 072016 (2018)



# Event selection and categorization

- Two channels defined requiring 1 or 2 isolated leptons:
  - Single lepton (*resolved*):  $\geq 5$  jets,  $\geq 2$  b-tag jets @60% WP or  $\geq 3$  b-tag jets @77% WP
  - Dilepton:  $\geq 3$  jets and  $\geq 2$  b-tag jets @77 % WP.



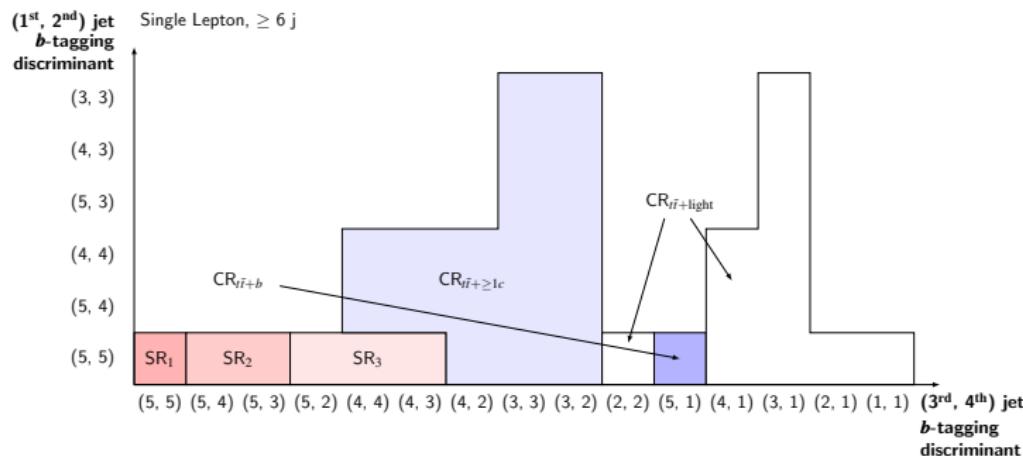
- 4 b-jet identification working points:
  - b-tagging efficiency [85%-60%]
  - c-jet rejection [3-35]
  - light rejection [30-1500]

- Discrete b-tagging discriminant defined as:

	Not tagged	loose	medium	tight	very tight
b-tagging efficiency	-	85%	77%	70%	60%
Discriminant value	1	2	3	4	5

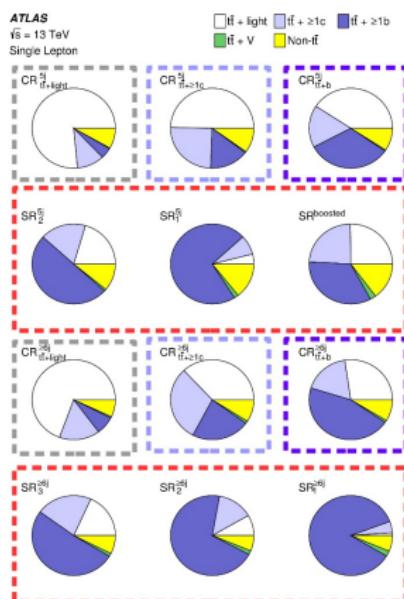
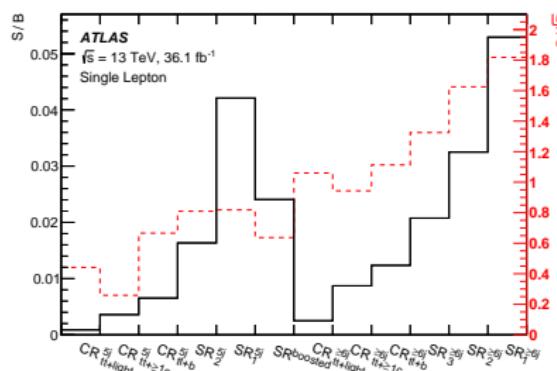
# Signal /Control regions

- **Signal** and **Control** regions are defined **varying the b-tagging discriminant requirements**.
- **Boosted** events are not further classified.
- Resolved events are categorized further into:
  - ⇒ **SRs** enhanced in signal and dominated by ttbb background.
  - ⇒ **CRs** rich in ttb,  $t\bar{t} + \geq 1c$  and  $t\bar{t} + \text{light}$  by loosening the b-tagging requirements.



# Background composition in Signal/Control regions

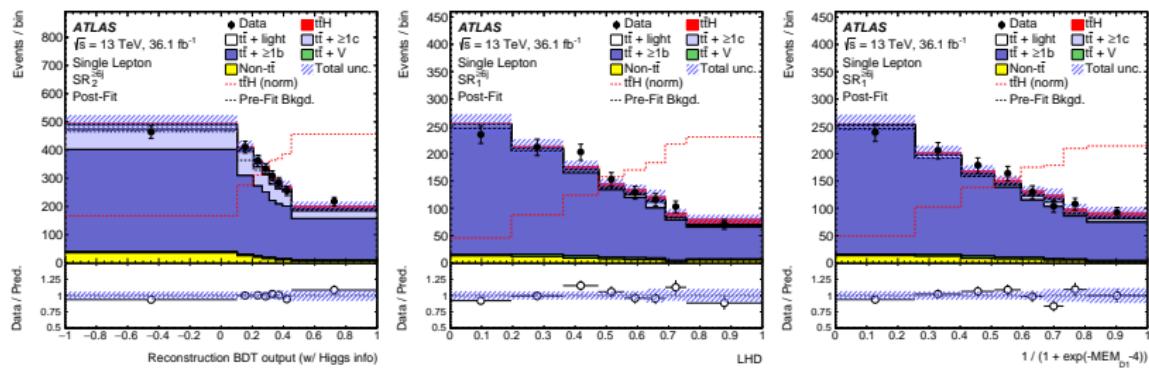
- 6 (3) **SRs** in Single lepton (Dilepton) channel.
- 6 (4) **CRs** in Single lepton (Dilepton) channel.
- Signal purity < 6%:  
⇒ MVAs to separate signal and ttbb.



Phys. Rev. D 97, 072016 (2018)

# MVA techniques in $t\bar{t}H(b\bar{b})$ -I

- The use of MVAs in  $t\bar{t}H(b\bar{b})$  intervenes at **2 stages** (more details in Ziyu's talk):
- Intermediate stage:**
  - Reconstruction BDT: used to select the **best combination of jets** to reconstruct the  $t\bar{t}H(b\bar{b})$  system.
  - Likelihood Discriminant (LHD): combines signal/background hypotheses probabilities based on 1D discriminating variables (invariant mass..etc).
  - Matrix Element Method (MEM): uses matrix element calculation at parton level to compute signal/background probabilities.



Phys. Rev. D 97, 072016 (2018)

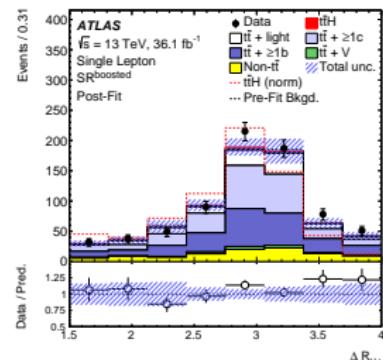
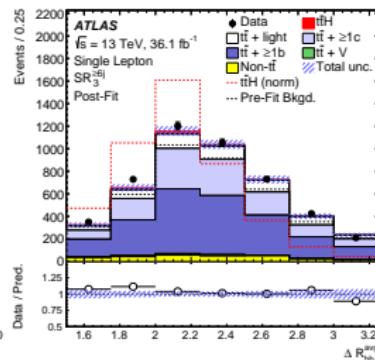
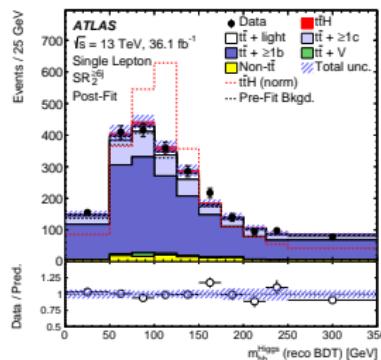
# MVA techniques in $t\bar{t}H(b\bar{b})$ -II

- **Classification BDT: final discriminant**

- ⇒ Trained to separate signal from background ( $t\bar{t}$ )
- ⇒ Only variables with good MC modeling are used as input.

- **Input variables:**

- ⇒ General kinematic variables
- ⇒ Discrete b-tagging discriminant
- ⇒ Reconstruction BDT, LHD and MEM (where available).
- ⇒ In boosted SR, large R-jets kinematics and their jet constituents are used.

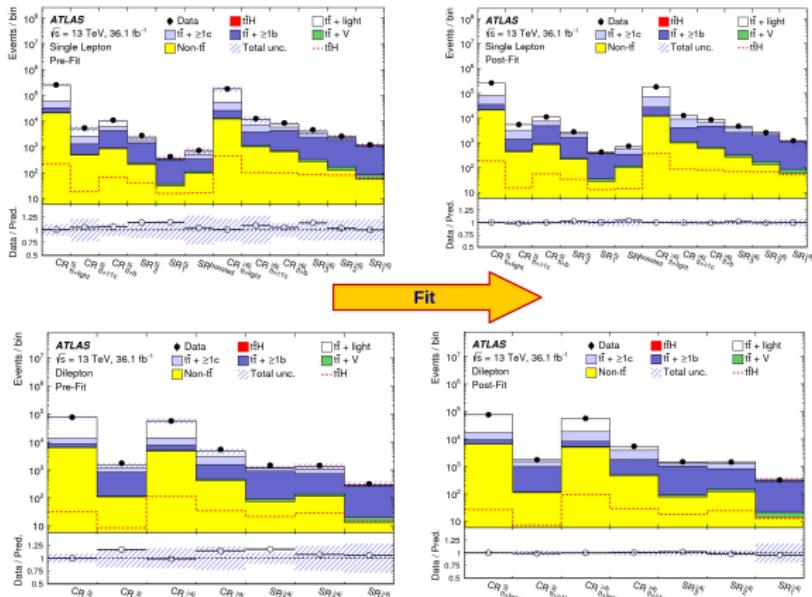


Phys. Rev. D 97, 072016 (2018)

# Fitting procedure

- Simultaneous profile likelihood fit to data in all SRs and CRs.
- Fit improves the Data/MC agreement and reduces the uncertainties.

Phys. Rev. D 97, 072016 (2018)

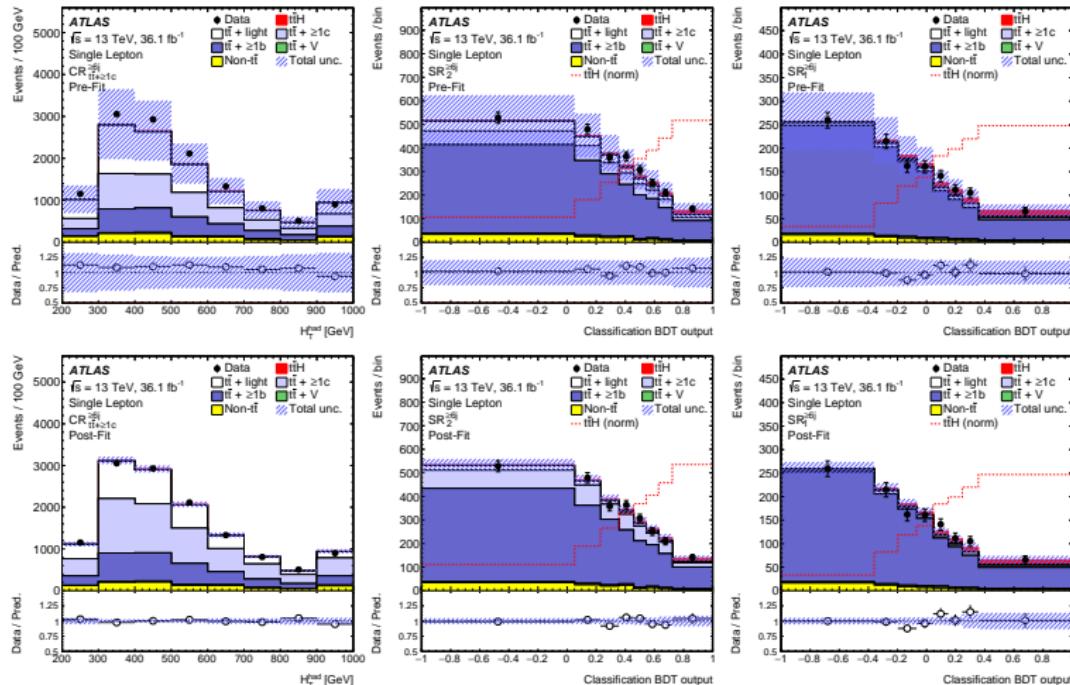


11

21

# Pre-/Post-fit distributions: Single lepton channel

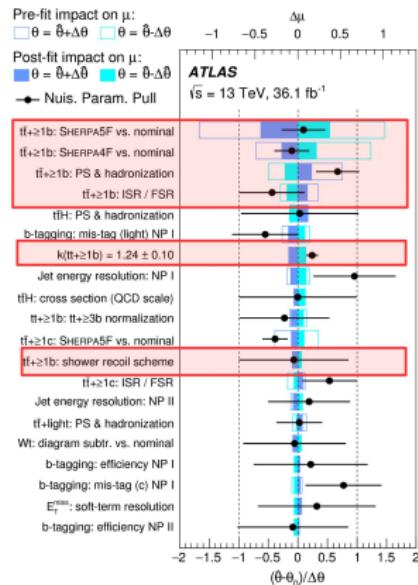
Phys. Rev. D 97, 072016 (2018)



No significant data/MC mis-modeling observed after the fit.

# Fit behavior

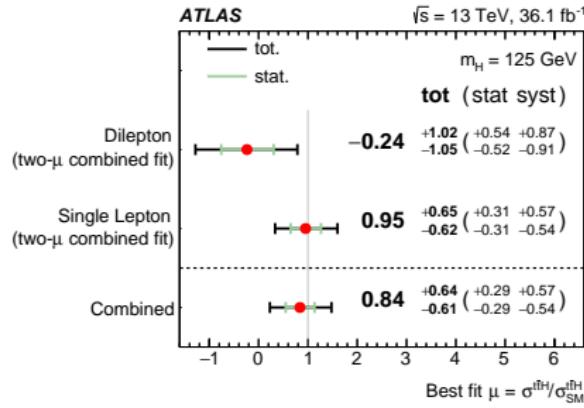
- Systematics with largest impact:
  - ⇒ related to  $t\bar{t}+ \geq 1b$  modeling
  - ⇒ Main challenge of the analysis.
- Norm. factors for  $t\bar{t}+ \geq 1b$  /  $t\bar{t}+ \geq 1c$  :
  - ⇒  $t\bar{t}+ \geq 1b$  :  $1.24 \pm 0.1$
  - ⇒  $t\bar{t}+ \geq 1c$  :  $1.63 \pm 0.23$
- Few systematics pulled by the fit to correct Data/MC mis-modeling
  - ⇒ Thorough checks performed to validate the fit model.



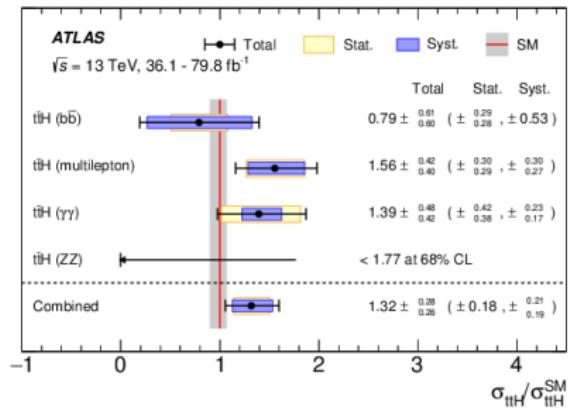
Phys. Rev. D 97, 072016 (2018)

# t̄H(bb) results

Phys. Rev. D 97, 072016 (2018)



arXiv:1806.00425



- Results consistent with the SM:  
 ⇒ significance w.r.t bkg only hypothesis:  $1.4$  ( $1.6$ )  $\sigma$  obs. (exp.).
- Sensitivity dominated by the “1l” channel,  $1l/2l$  channels compatibility: 19%.
- **t̄H observation achieved when combining with other Higgs decay modes channels:**  
 ⇒ obs.(exp.) significance of  $5.8$  ( $4.9$ )  $\sigma$ :  $WW^*$ ,  $\tau^+\tau^-$ ,  $(\gamma\gamma, ZZ^*$  based on  $80\text{fb}^{-1}$ ) )

# Towards full Run 2 analysis

A lot of work ongoing for the next round of the  $t\bar{t}H(b\bar{b})$  analysis adding the full Run 2 data

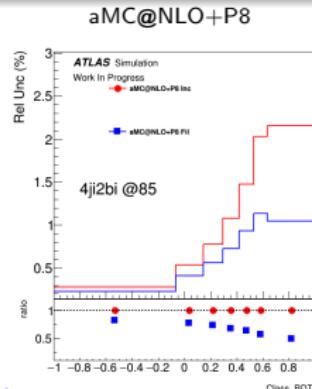
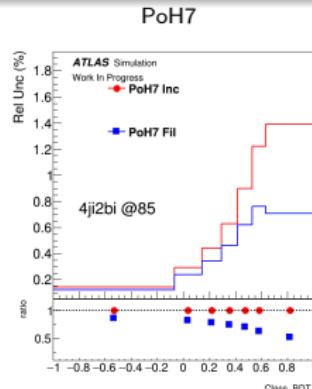
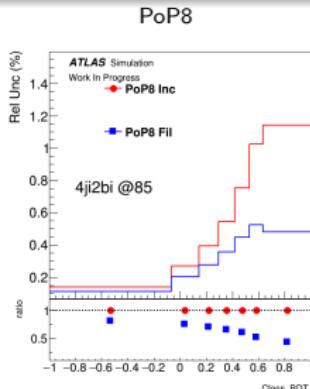
- My work focuses on tackling the two leading systematics:
  - ⇒ **Limited MC statistics** ⇒ increase  $t\bar{t} + \text{HF}$  statistics.
  - ⇒  **$t\bar{t}$  modeling systematics** ⇒ alternative model based on data

Phys. Rev. D 97, 072016 (2018)

Uncertainty source	$\Delta\mu$
$t\bar{t} + \geq 1b$ modeling	+0.46 -0.46
Background-model stat. unc.	+0.29 -0.31
$b$ -tagging efficiency and mis-tag rates	+0.16 -0.16
Jet energy scale and resolution	+0.14 -0.14
$t\bar{t}H$ modeling	+0.22 -0.05
$t\bar{t} + \geq 1c$ modeling	+0.09 -0.11
JVT, pileup modeling	+0.03 -0.05
Other background modeling	+0.08 -0.08
$t\bar{t} + \text{light}$ modeling	+0.06 -0.03
Luminosity	+0.03 -0.02
Light lepton ( $e, \mu$ ) id., isolation, trigger	+0.03 -0.04
Total systematic uncertainty	+0.57 -0.54
$t\bar{t} + \geq 1b$ normalization	+0.09 -0.10
$t\bar{t} + \geq 1c$ normalization	+0.02 -0.03
Intrinsic statistical uncertainty	+0.21 -0.20
Total statistical uncertainty	+0.29 -0.29
Total uncertainty	+0.64 -0.61

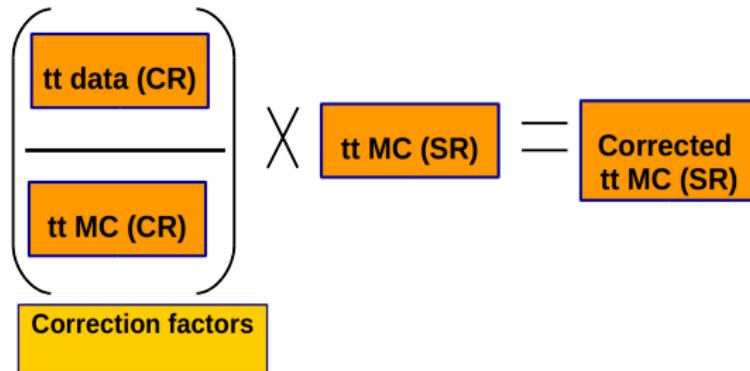
# $t\bar{t} + HF$ filters optimization

- Loose filters were developed in the previous round of the  $t\bar{t}H(b\bar{b})$  analysis
  - ⇒ Cutting on hadrons kinematics at truth level.
  - ⇒ Only b-filtered nominal and alternative  $t\bar{t}$  samples were produced.
- However, more stats needed for  $t\bar{t} + HF$  in the signal-rich regions to reduce uncertainties.
- Multiple new filters tested tightening kinematic and multiplicity requirements on hadrons:
  - ⇒ Optimized for higher stats without compromising the phase space coverage.
- The new filters were used to produce b(c)-filtered  $t\bar{t}$  samples for the ongoing analysis.
  - ⇒ For instance,  $\times 4$  more effective stat in the most sensitive BDT bin is achieved.
  - ⇒ Main concern: negative weights reducing the effective stats.

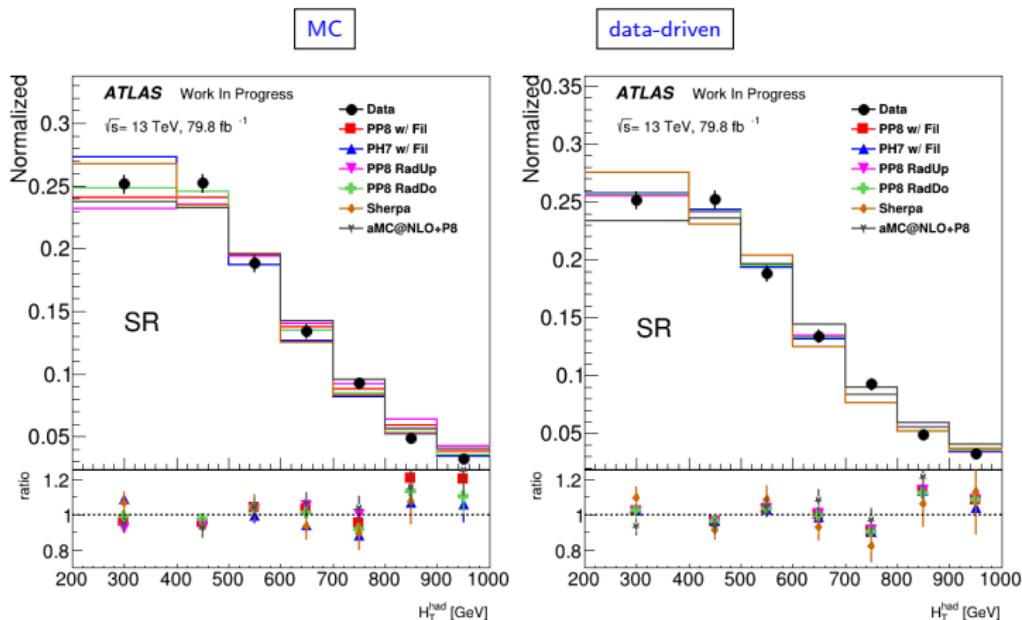


# Alternative Data-driven $t\bar{t}$ model: motivation and idea

- Traditional MC-based  $t\bar{t}$  model does not perfectly describe data especially for  $t\bar{t}+ \geq 1b$ .  
    ⇒ Large theoretical uncertainties even with the available NLO ttbb 4F predictions.
- Developing a more data-driven approach would improve the sensitivity of the analysis.
- The idea: correct MC to data in a CR and extrapolate correction to  $t\bar{t}+ \geq 1b$  rich region.
- CR is chosen to be less enhanced in  $t\bar{t}+ \geq 1b$  based on b-tagging requirements.
- Still reliant on some MC input:  
    ⇒ Instead of modeling systematics on  $t\bar{t} + \text{jets}$  MC in SR → systematics on extrapolation from CR to SR.



# Comparing predictions: MC Vs data-driven



- Shape differences reduced after Data/MC correction.
- Data/MC agreement improved  $\Rightarrow$  Better modeling.
- Still limited by stats  $\Rightarrow$  more checks needed to validate the method.

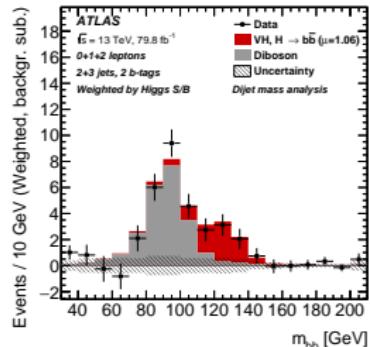
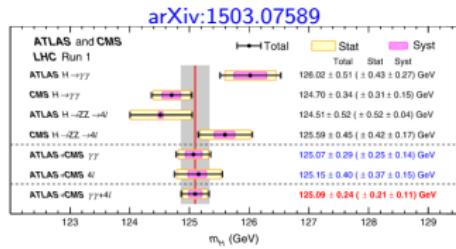
# Summary

- Results of the  $t\bar{t}H(b\bar{b})$  search performed using  $36.1 \text{ fb}^{-1}$  of pp collision data presented.
- $t\bar{t}H(b\bar{b})$  is a **very challenging** analysis:
  - ⇒ Overwhelming and poorly modeled  $t\bar{t} + \geq 1b$  background.
  - ⇒ Heavy use of MVAs to separate signal from background
- Results obtained are **consistent with the SM expectations**:  $1.4$  ( $1.6$ )  $\sigma$  obs.(exp.)
- **An observation of the  $t\bar{t}H$**  production mode is achieved when combining with other H decay modes:  $5.8$  ( $4.9$ )  $\sigma$  obs (exp.).
- A lot of work ongoing to tackle the leading systematics limiting the sensitivity of the analysis:
  - ⇒ New tighter filters were developed to increase the  $t\bar{t} + \text{HF}$  stats in the SRs.
  - ⇒ An alternative approach to estimate  $t\bar{t}$  based on data is being explored.

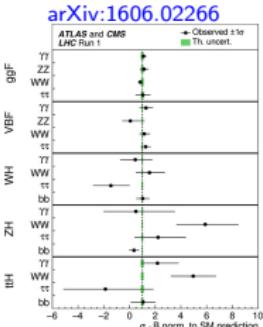
# Backups

# Higgs boson measurements catalogue

Phys. Lett. B 786 (2018) 59

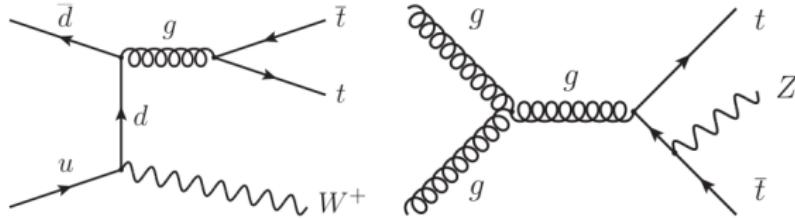


- Higgs boson discovered in 2012 by ATLAS and CMS.
- $m_H \approx 125 \text{ GeV}$ , spin and parity ( $0^+$ )
- Higgs boson observed in:
  - ⇒ bosonic decay modes:  $\gamma\gamma$ ,  $WW^*$ ,  $ZZ^*$
  - ⇒ fermionic decay modes:  $\text{Higgs} \rightarrow \tau\tau$ ,  $\text{H} \rightarrow b\bar{b}$
- All measurements compatible with the SM.



# Other backgrounds in the $t\bar{t}H(b\bar{b})$ channel

- Other backgrounds also considered in the analysis but **negligible w.r.t  $t\bar{t} + \text{jets}$**  :
  - ⇒  $t\bar{t} + V$  ( $V=W,Z$ )
  - ⇒ Single top ( $Wt$ ,  $s(t)$ -channel)
  - ⇒  $V+\text{jets}$
  - ⇒  $tHW$ ,  $tHqb$
  - ⇒  $t\bar{t}t\bar{t}$ ,  $t\bar{t}WW$  , ...
  - ⇒ fake leptons.



# Event selection

- Two channels defined depending on the number of leptons ( $e/\mu$ ): Single lepton (1l), Dilepton (2l).
- **Dilepton:** 2 opposite-sign leptons with  $p_T > 27$  GeV (leading) and  $p_T > 15(10)$  GeV (subleading) in ee ( $e\mu/\mu\mu$ ) channels.
  - ⇒ In ee/ $\mu\mu$ :  $m_{ll} > 15$  GeV, outside the  $m_Z$  window and events with  $\geq 1 \tau_{had}$  are vetoed.
  - ⇒ Require  $\geq 3$  jets and  $\geq 2$  b-tagged jets @77 % WP.
- **Single lepton:** 1 l with  $p_T > 27$  GeV (veto events with  $\geq 2 \tau_{had}$ )
  - ⇒ **High  $p_T$  category (boosted):**
    - \* Require  $\geq 1$  Boosted Higgs and  $\geq 1$  boosted top candidates (large R-jets reclustered from  $R = 0.4$  jets) and  $\geq 1$  additional b-tagged jet @85 % WP.
  - ⇒ if not boosted → **resolved event:**
    - \* Require  $\geq 5$  jets and ( $\geq 2$  b-tagged jets @60 % WP or  $\geq 3$  b-tagged jets @77% WP).

# Boosted channel

- Single lepton: 1l with  $p_T > 27$  GeV (veto events with  $\geq 2 \tau_{had}$ )

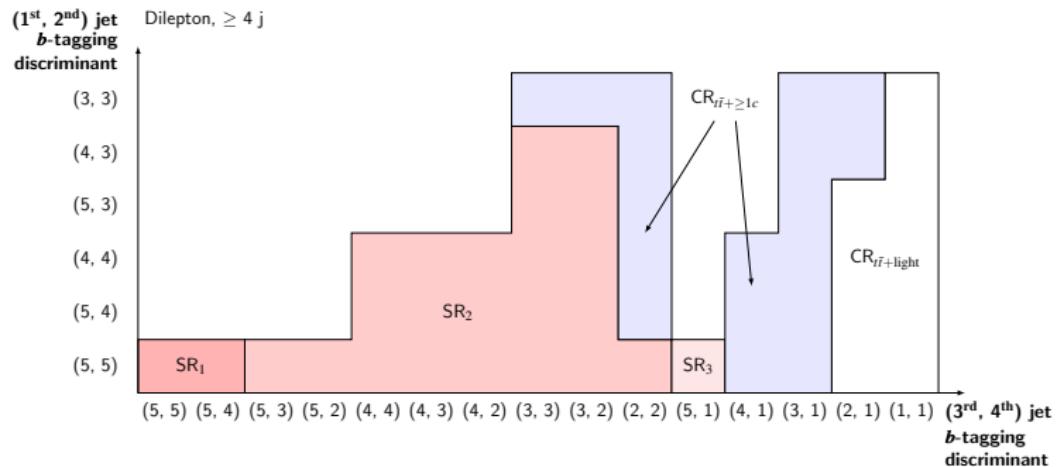
⇒ **High  $p_T$  category (boosted):**

- \* Require  $\geq 1$  Boosted Higgs and  $\geq 1$  boosted top candidates (large R-jets reclustered from  $R = 0.4$  jets) and  $\geq 1$  additional loose b-tagged jet.
- \* Boosted Higgs candidate:  $p_T > 200$  GeV,  $\geq 2$  loose b-tagged jets.
- \* Boosted top candidate:  $p_T > 250$  GeV, exactly one loose b-tagged jet and  $\geq 1$  additional non-tagged jet.

⇒ if not boosted → **resolved event:**

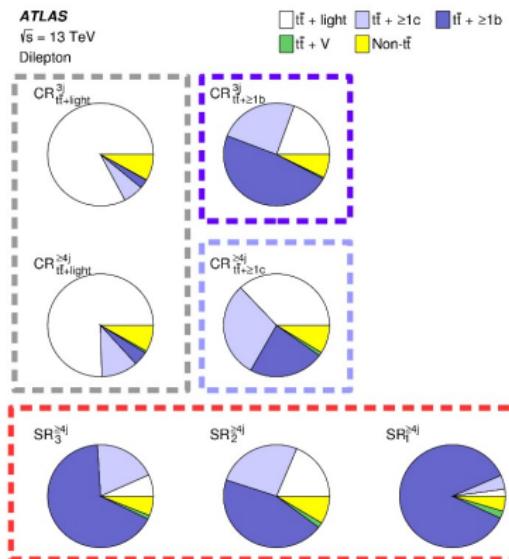
- \* Require  $\geq 5$  jets and  $\geq 2$  very tight or  $\geq 3$  medium b-tagged jets.

# Signal /Control regions: Dilepton channel



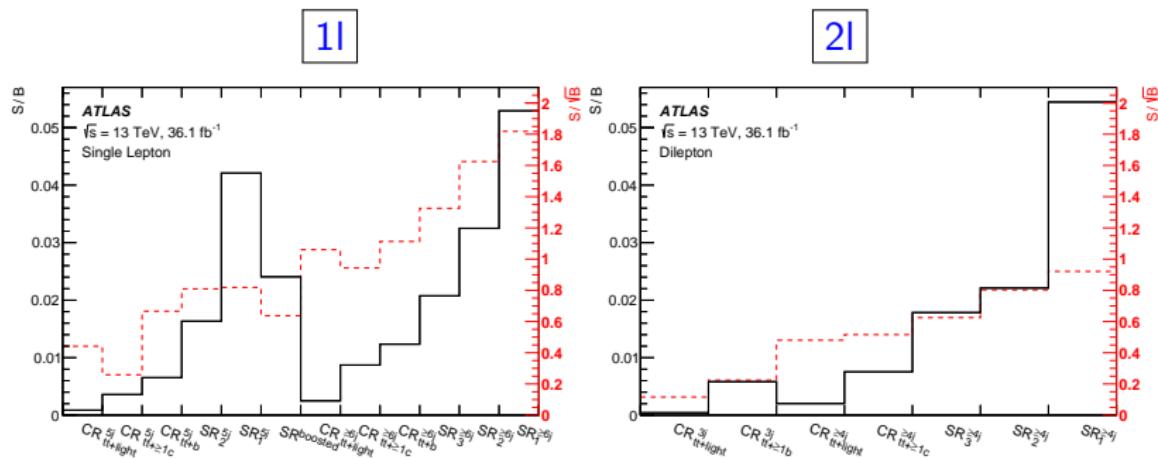
Phys. Rev. D 97, 072016 (2018)

# Background composition in Signal/Control regions: 2)



Phys. Rev. D 97, 072016 (2018)

# S/B and $S/\sqrt{B}$ for 1l/2l channels

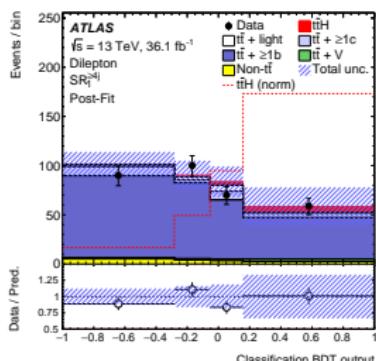
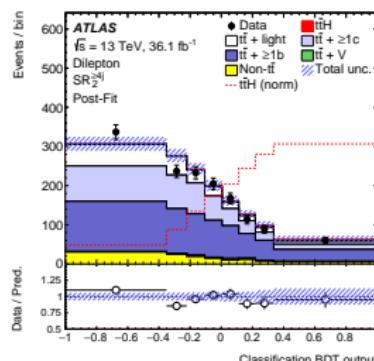
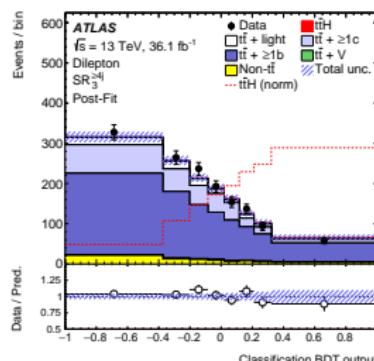
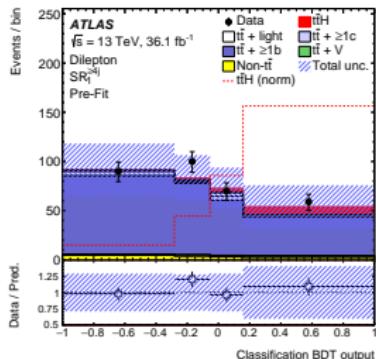
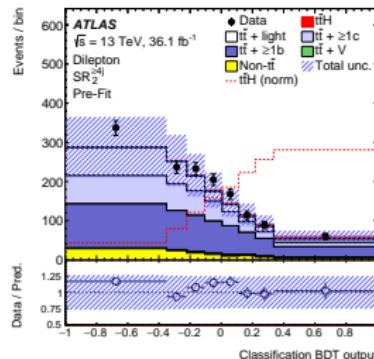
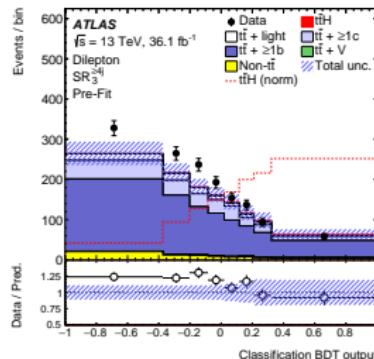


Phys. Rev. D 97, 072016 (2018)

- Signal purity < 6%  $\Rightarrow$  need MVAs to further separate the signal from ttbb.

# Pre/Post fit SR distributions: Dilepton channel

Phys. Rev. D 97, 072016 (2018)



# t $\bar{t}$ H(bb) uncertainties breakdown

Phys. Rev. D 97, 072016 (2018)

- Analysis dominated by systematics.
- Most relevant uncertainties:
  - ⇒ t $\bar{t}$ + ≥ 1b background modeling.
  - ⇒ Limited MC statistics.
  - ⇒ b-tagging uncertainties.
  - ⇒ Jet uncertainties.

Uncertainty source	$\Delta\mu$
t $t$ + >1b modeling	+0.46 -0.46
Background-model stat. unc.	+0.29 -0.31
b-tagging efficiency and mis-tag rates	+0.16 -0.16
Jet energy scale and resolution	+0.14 -0.14
t $\bar{t}H$ modeling	+0.22 -0.05
t $\bar{t}$ + ≥1c modeling	+0.09 -0.11
JVT, pileup modeling	+0.03 -0.05
Other background modeling	+0.08 -0.08
t $\bar{t}$ + light modeling	+0.06 -0.03
Luminosity	+0.03 -0.02
Light lepton ( $e, \mu$ ) id., isolation, trigger	+0.03 -0.04
Total systematic uncertainty	+0.57 -0.54
t $\bar{t}$ + ≥1b normalization	+0.09 -0.10
t $\bar{t}$ + ≥1c normalization	+0.02 -0.03
Intrinsic statistical uncertainty	+0.21 -0.20
Total statistical uncertainty	+0.29 -0.29
Total uncertainty	+0.64 -0.61

# Systematics uncertainties assigned to $t\bar{t} + \text{jets}$ modeling

Systematic source	Description	$t\bar{t}$ categories
$t\bar{t}$ cross-section	Up or down by 6%	All, correlated
$k(t\bar{t} + \geq 1c)$	Free-floating $t\bar{t} + \geq 1c$ normalization	$t\bar{t} + \geq 1c$
$k(t\bar{t} + \geq 1b)$	Free-floating $t\bar{t} + \geq 1b$ normalization	$t\bar{t} + \geq 1b$
SHERPA5F vs. nominal	Related to the choice of NLO event generator	All, uncorrelated
PS & hadronization	POWHEG+HERWIG 7 vs. POWHEG+PYTHIA 8	All, uncorrelated
ISR / FSR	Variations of $\mu_R$ , $\mu_F$ , $h_{\text{damp}}$ and A14 Var3c parameters	All, uncorrelated
$t\bar{t} + \geq 1c$ ME vs. inclusive	MG5_aMC@NLO+HERWIG++: ME prediction (3F) vs. incl. (5F)	$t\bar{t} + \geq 1c$
$t\bar{t} + \geq 1b$ SHERPA4F vs. nominal	Comparison of $t\bar{t} + b\bar{b}$ NLO (4F) vs. POWHEG+PYTHIA 8 (5F)	$t\bar{t} + \geq 1b$
$t\bar{t} + \geq 1b$ renorm. scale	Up or down by a factor of two	$t\bar{t} + \geq 1b$
$t\bar{t} + \geq 1b$ resumm. scale	Vary $\mu_Q$ from $H_T/2$ to $\mu_{\text{CMMPS}}$	$t\bar{t} + \geq 1b$
$t\bar{t} + \geq 1b$ global scales	Set $\mu_Q$ , $\mu_R$ , and $\mu_F$ to $\mu_{\text{CMMPS}}$	$t\bar{t} + \geq 1b$
$t\bar{t} + \geq 1b$ shower recoil scheme	Alternative model scheme	$t\bar{t} + \geq 1b$
$t\bar{t} + \geq 1b$ PDF (MSTW)	MSTW vs. CT10	$t\bar{t} + \geq 1b$
$t\bar{t} + \geq 1b$ PDF (NNPDF)	NNPDF vs. CT10	$t\bar{t} + \geq 1b$
$t\bar{t} + \geq 1b$ UE	Alternative set of tuned parameters for the underlying event	$t\bar{t} + \geq 1b$
$t\bar{t} + \geq 1b$ MPI	Up or down by 50%	$t\bar{t} + \geq 1b$
$t\bar{t} + \geq 3b$ normalization	Up or down by 50%	$t\bar{t} + \geq 1b$

# Analysis strategy

- Control the poorly modeled  $t\bar{t} + \text{jets}$  ( $t\bar{t}+ \geq 1b$ ) background.
- Exploit b-tagging to define regions enhanced in  $t\bar{t}+ \geq 1b$ ,  $t\bar{t}+ \geq 1c$  and  $t\bar{t} + \text{light}$  contributions.
- Use MVA techniques to further distinguish the signal from background in regions where the signal contribution is relatively large.
- Fit all channels and regions simultaneously, to the observed data and extract the  $t\bar{t}H$  cross section.