All hadronic $t\bar{t}H$ ($H \rightarrow b\bar{b}$) with the ATLAS detector

Daniele Madaffari

October 29, 2015

PESBLADe meeting

CENTRE DE PHYSIQUE DES PARTICULES DE MARSEILLE
CPPM

ATLAS

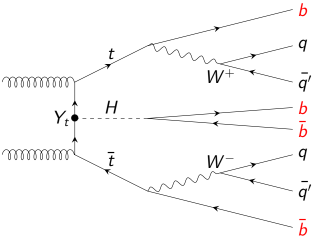
Outline

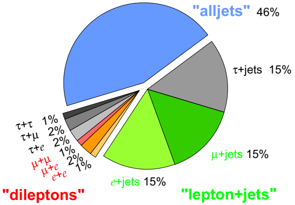
- Introduction to fully hadronic $t\bar{t}H$ ($H \rightarrow b\bar{b}$) analysis
  - Motivations
  - Analysis preselection
- Modellizzation of event selections from per-jet properties
  - The effect of applying selections to a sample is reproduced by the application of event weights function of jet properties
    - Tag Rate function method in MC :TRF_{MC}
    - Tag Rate function method for multijet background : TRF_{MJ}
    - Trigger selection
- Final discriminant: Boosted Decision Tree
- Systematic uncertainties considered in the analysis
  - Details on TRF_{MJ} method systematics
- Results of standalone fully hadronic $t\bar{t}H$ and ATLAS $t\bar{t}H$ combination
  - $t\bar{t}H$ cross section limit and best-fit
  - Higgs couplings
- Conclusions

Introduction

- Fully hadronic $t\bar{t}H$ ($H \to b\bar{b}$) analysis:
  - $t\bar{t}H$: Direct acces to Yukawa coupling of Higgs boson to top-quark (Y_t)
  - $H \to b\bar{b}$: the largest barching ratio of SM Higgs (56%)
  - Full hadronic $t\bar{t}$ BR = 46%
  - First measurement at the LHC
- Multijet final state: $\sim$ 8 jets, $\sim$ 4 b-jets:
  - Multi-jet trigger:
    - At least 5 jets with $E_T > 55$ GeV

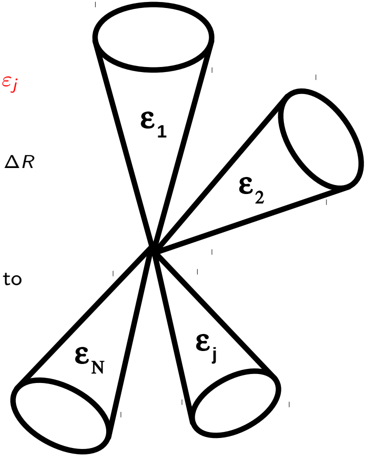
  Offline requirements:
  - At least 5 jets with $p_T > 55$ GeV
  - Other jets with $p_T > 25$ GeV
  - b-tagging: MV1 with 60% efficiency WP
  - Lepton veto
- Main background: Multijet production (MJ)
  - Data driven description through dedicated technique: TRF_{MJ}
- Other backgrounds:
  - $t\bar{t}$ + jets, $t\bar{t}V$, single top
  - Using TRF_{MC} method to enhance statistics of MC samples





Event description from per-jet properties

Selections on event quantities can be described as function of properties of the jets in the event



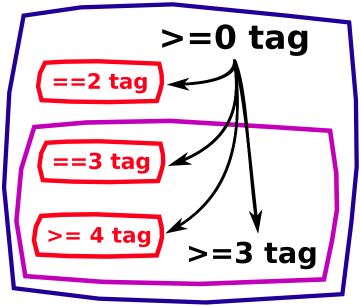
- To each jet it is possible to associate an **efficiency** ε_j depending on
  - Jet p_T, η, flavour (in MC only)
  - Relation of the jet with other jets in the event, like ΔR
- Event weights W are evaluated as function of the per-jet efficiencies: $W = f(\varepsilon_1, \ldots, \varepsilon_N)$
- The effect of applying a selection to a sample is reproduced by the application of the event weights to each event in the sample

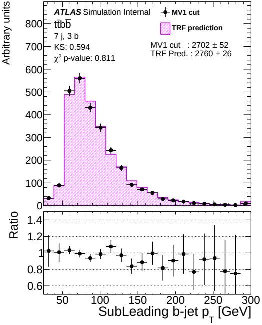
Benefit:

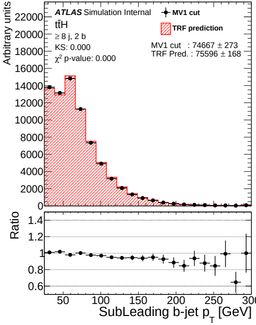
- Avoid loss of statistics

Tag Rate Function method for MC: TRF_{MC}

- **b-tagging selection is described by the application of event weight $W = f(\varepsilon_1, \ldots, \varepsilon_N)$**
  - $\varepsilon(p_T, \eta, \text{flavour})$ = probability of the jet to be b-tagged
  - W = probability to have n_b number of b-tagged jets in the events
- Using full MC data set without any b-tagging requirement
  - **Avoid loss of statistics** when selecting events with high b-tag multiplicities
- TRF_{MC} method predicts **normalization and shapes** of variables
  - TRF_{MC} method allows to select a configuration of jets to consider b-tagged based on the probability of the configuration itself



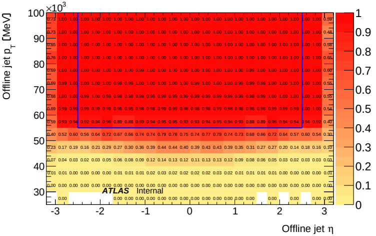




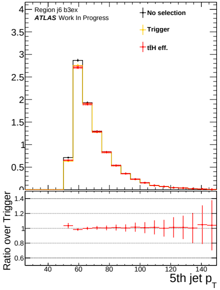
Trigger efficiency and SF

- **Trigger selection emulated by the application of event weight $W = f(\varepsilon_1, \ldots, \varepsilon_N)$**
- Using full MC data set without trigger requirement
  - $\varepsilon(p_T, \eta)$ = probability to fire a trigger chain
  - W = probability to have at least 5 trigger chains
    - ◇ Trigger efficiency estimation validated in MC
- Max signal acceptance reached by requiring as low offline p_T cut as possible
  - Trigger **plateau**: 5^{th} leading jet offline $\mathbf{p_T > 65}$ **GeV**
  - Analysis **selection**: 5^{th} leading jet offline $\mathbf{p_T > 55}$ **GeV**
- Working below plateau requires estimating data/MC trigger Scale Factors (SF)
  - SF is evaluated comparing per-jet trigger efficiencies in data and PYTHIA8 di-jet MC
  - Sample dependance, derived in MC, is assigned as systematic uncertainty

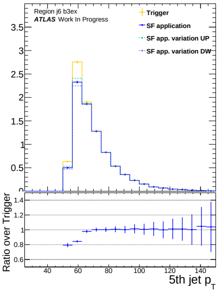
Trigger efficiency evaluated in $t\bar{t}H$



Validation in $t\bar{t}H$

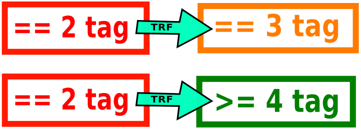


SF application $t\bar{t}H$

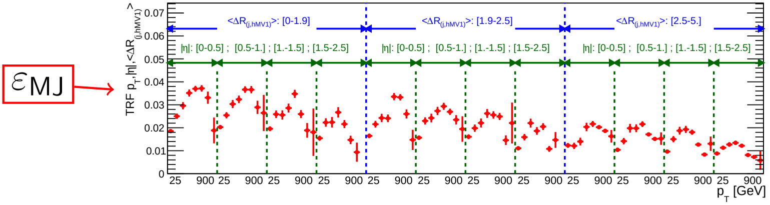


MJ background estimation: TRF_{MJ} method

- TRF_{MJ} method is a **data-driven** method
- ***b*-tagging selection modelled by application of event weight** $W = f(\varepsilon_{\mathbf{MJ,1}}, \ldots, \varepsilon_{\mathbf{MJ,N}})$
  - $\varepsilon_{MJ}(p_T, \eta, \langle \Delta R_{(j,\mathrm{hMV1})} \rangle)$ = probability to be *b*-tagged
    - ◇ ε_{MJ} is evaluated in data in a dedicated MJ dominated sample
    - ◇ True flavour of the jet is unknown in data, $\langle \Delta R_{(j,\mathrm{hMV1})} \rangle$ sensitive to heavy-flavour production
    - ◇ $\langle \Delta R_{(j,\mathrm{hMV1})} \rangle$: Average of the distances of the jet from the two jets with the highest MV1 weight
- TRF_{MJ} method is applied in regions where the amount of MJ background is known
  - Regions with exactly 2 *b*-tagged jets
  - MJ (2b) = DATA (2b) - $\sum$ $MC_{\mathrm{background}}$ (2b)



- Special mathematical treatment to estimate MJ in non-overlapping regions
  - W linked to the probability to have $n_b = 3$ or ≥ 4 number of *b*-tagged jets



TRF_{MJ} validation in data and MC

Closure test is performed applying TRF_{MJ} method in data and MC

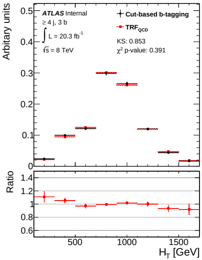
- **Data:** TRF_{MJ} extraction sample
  - Normalizations agree within 5%
  - Good shapes description

Data	3j,3b	≥4j, 3b	≥4j, ≥4b
TRF_{MJ}	632 ± 4	7952 ± 25	452 ± 2
Direct *b*-tag	641	7585	425

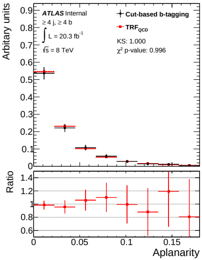
- **MC:** PYTHIA8 di-jet
  - Normalizations agree within 6%
  - Good shapes description
    - Plots are made using sub-sample with more statistics

Di-jet MC	≥4j, 3b	≥4j, ≥4b
TRF_{MJ}	15.5 ± 0.1	0.89 ± 0.01
Direct *b*-tag	14.6 ± 0.5	0.9 ± 0.1

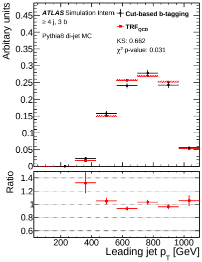
≥4j, 3b



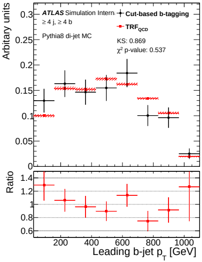
≥4j, ≥4b



≥4j, 3b



≥4j, ≥4b



Boosted Decison Tree

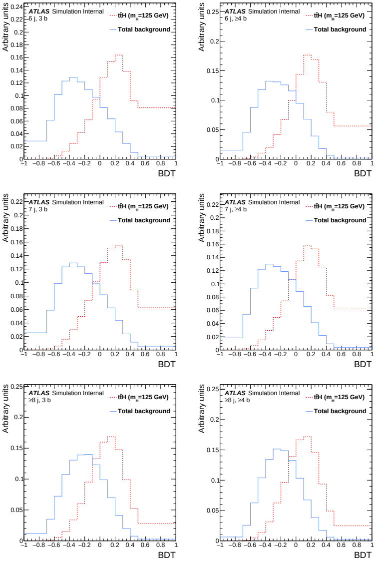
- Boosted Decision Trees (BDT) are trained **one for each fit region**

  Signal : $t\bar{t}H$, inclusive in top and Higgs decays

  Background : Multijet + all MC backgrounds

  Input variables selection:

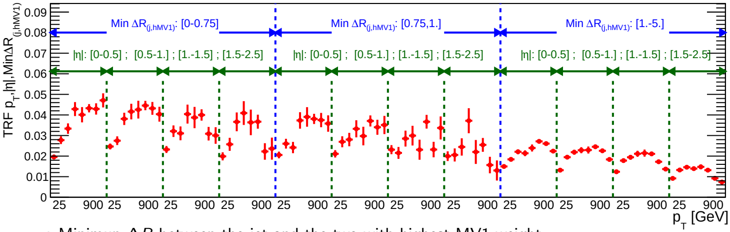
- Start with a pool of interesting variables (~ 35)
- Rank the best variables
  - Iteratively add one variable in the BDT training and select the one giving the best improovement in the discrimination
  - Stop when the addition of more variables does not improove the performance anymore
    - i.e. Reach a plateau in the BDT performance -
  - Roughly 11 variables per region



Systematics on TRF_{MJ}: description of ε_{MJ} I

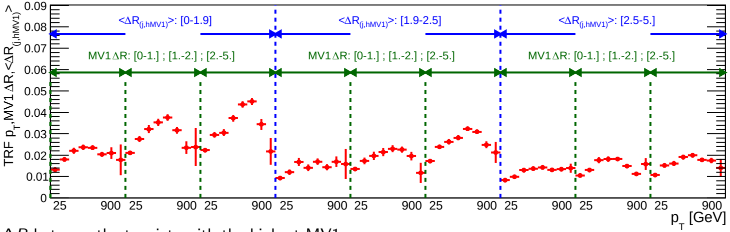
Different sets of variables have been used to parametrize ε_{MJ}

- Variables used: p_T, $|\eta|$, Min$\Delta R_{(j,\text{hMV1})}$



Min $\Delta R_{(j,\text{hMV1})}$: Minimun ΔR between the jet and the two with highest MV1 weight

- Variables used: p_T, $\langle \Delta R_{(j,\text{hMV1})} \rangle$, MV1 ΔR

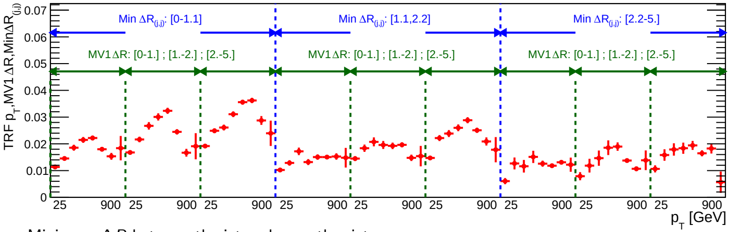


MV1 ΔR: ΔR between the two jets with the highest MV1

Systematics on TRF_{MJ}: description of ε_{MJ} II

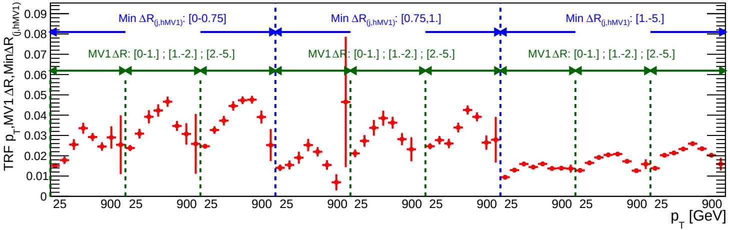
Different sets of variables have been used to parametrize ε_{MJ}

- Variables used: p_T, Min $\Delta R_{(j,j)}$, MV1 ΔR



Min $\Delta R_{(j,j)}$: Minimum ΔR between the jet and any other jet

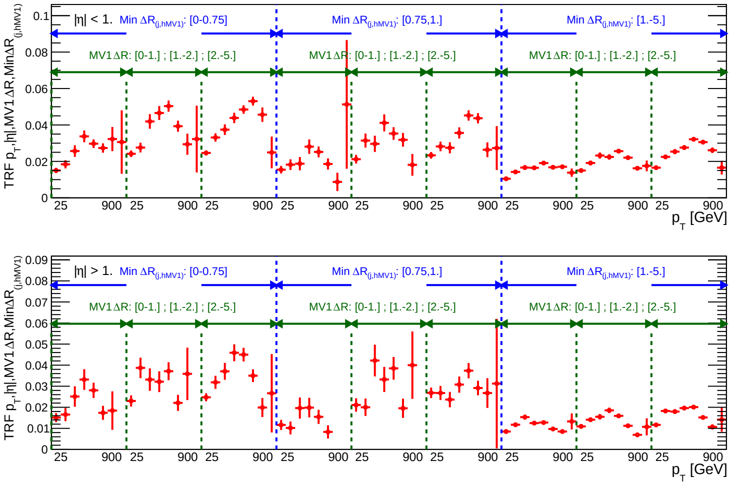
- Variables used: p_T, Min $\Delta R_{(j,\text{hMV1})}$, MV1 ΔR



Systematics on TRF_{MJ}: description of ε_{MJ} III

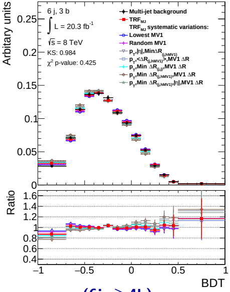
Different sets of variables have been used to parametrize ε_{MJ}

- Variables used: p_T, $|\eta|$, Min $\Delta R_{(j,\mathrm{hMV1})}$, MV1 ΔR

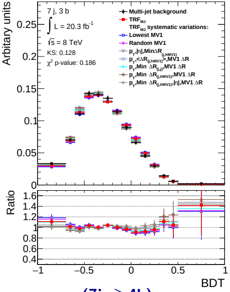


Systematics on TRF_{MJ}: description of ε_{MJ} IV

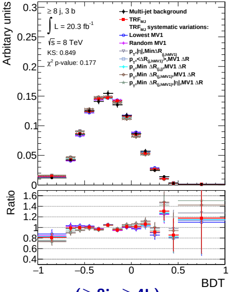
(6j, 3b)



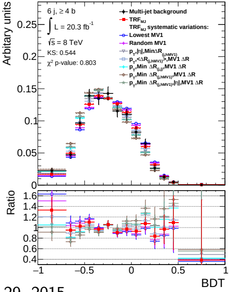
(7j, 3b)



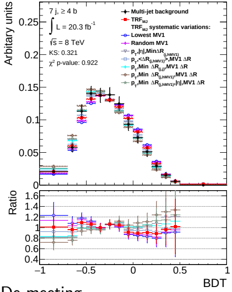
(≥8j, 3b)



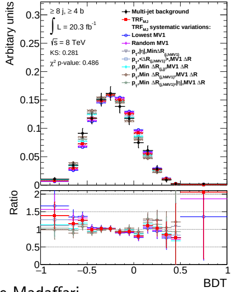
(6j, ≥4b)



(7j, ≥4b)



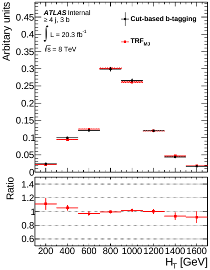
(≥8j, ≥4b)

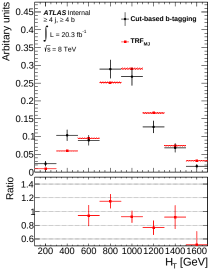


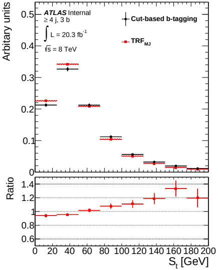
Systematics on TRF_{MJ}: residual mismodeling

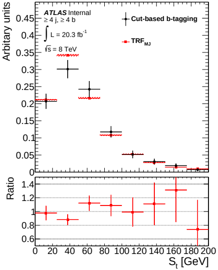
- Mismodeling is obesrved for H_T and S_t variables in the TRF_{MJ}**extraction region**
- A reweight, evaluated in the same region, is applied to compensate for this effect

- H_T reweight:
  - Mismodelling is observed in ≥ 4 b-tag regions
  - Reweight is applied to ≥ 4 b-tag regions only
- S_t reweight:
  - Mismodelling is observed in all regions





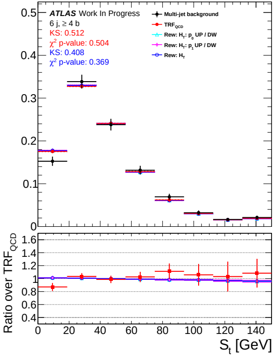




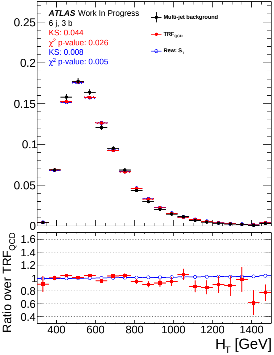
Systematics on TRF_{MJ}: independence of reweight

- The two reweights are independent
  - Reweght w.r.t H_T has no effect on S_t and vice-versa

H_T reweight

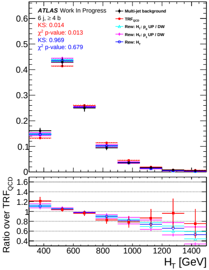


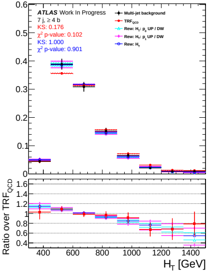
S_t reweight

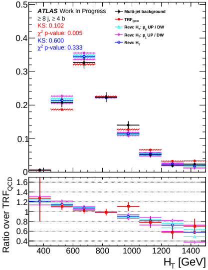


Systematics on TRF_{MJ}: effect of reweight

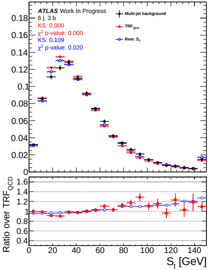
H_T **reweight** $\Rightarrow$

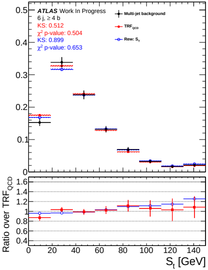


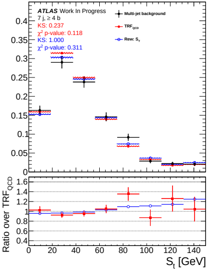




S_t **reweight** $\Rightarrow$

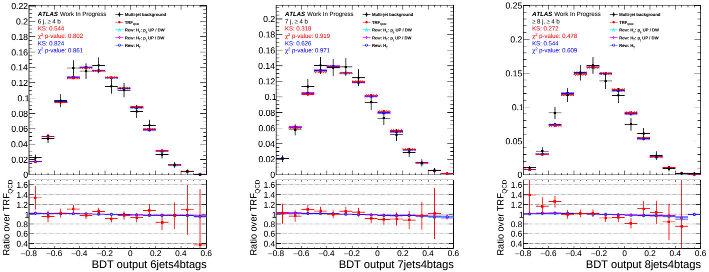




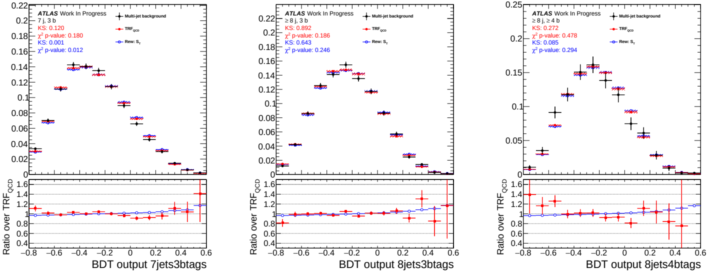


Systematics on TRF_{MJ}: effect of reweight – BDT

H_T reweight ⇒



S_t reweight ⇒



Systematics uncertainties

- MJ background estimation:

  Shape:
  - 5 components for ε_{MJ} description
  - 2 components for b-tagged jet selection
  - 2 components for H_T and S_t residual mismodeling

  Normalization:
  - 6 SF, MJ normalization free floating in each region

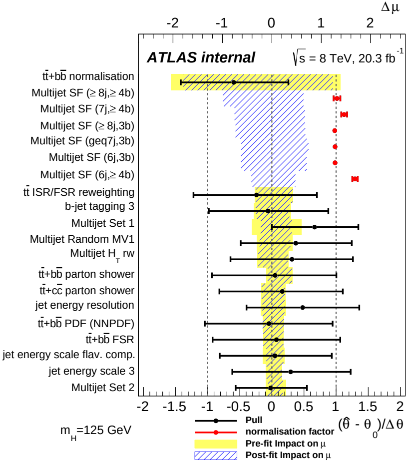
- Jet Energy scale:
  - Split in 22 uncorrelated components

- b-tagging:
  - $b/c/$light-tagging split into 6/6/12 uncorrelated components

- $t\bar{t}$ +jets modelling
  - Shape and normalization uncertainties derived from variation of renormalization scale and PDF

- $t\bar{t}$ +HF normalizations
  - 50% on $t\bar{t}+b\bar{b}$, 50% on $t\bar{t}+c\bar{c}$

- top p_T reweighting
  - Scale variation, shower model and PDF for $t\bar{t}+b\bar{b}$ reweighting
  - 9 leading systematic uncertainties from differential $t\bar{t}$ cross-section measurement



Pre- and post-fit yields

Pre-fit yields:

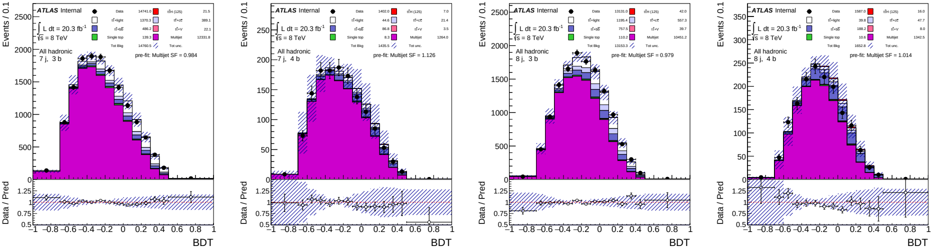
	6j,3b	6j,≥4b	7j,3b	7j,≥4b	8j,3b	8j,≥4b
Multijet	16400 ± 130	1100 ± 33	12500 ± 12	1100 ± 33	10600 ± 100	1300 ± 36
single top	170 ± 63	6.0 ± 3.7	140 ± 55	8.3 ± 4.6	110 ± 50	11 ± 5.9
$t\bar{t}+V$	14 ± 6.3	1.8 ± 1.5	22 ± 9.0	3.5 ± 2.3	40 ± 15	8.0 ± 4.2
$t\bar{t}+b\bar{b}$	330 ± 180	44 ± 26	490 ± 270	87 ± 51	760 ± 450	190 ± 110
$t\bar{t}+c\bar{c}$	280 ± 170	17 ± 12	390 ± 240	21± 15	560 ± 350	48 ± 33
$t\bar{t}+light$	1500 ± 400	48 ± 18	1370 ± 400	45 ± 18	1200 ± 500	40 ± 23
$t\bar{t}H$ (125)	13 ± 4.5	3.3 ± 2.1	21 ± 6.2	7.0 ± 3.2	42 ± 11	16 ± 6.1
Total bkg.	18700 ± 500	1200 ± 50	14960 ± 580	1300 ± 65	13380 ± 77	1650 ± 130
Data	18508	1545	14741	1402	13131	1587

Post-fit yields:

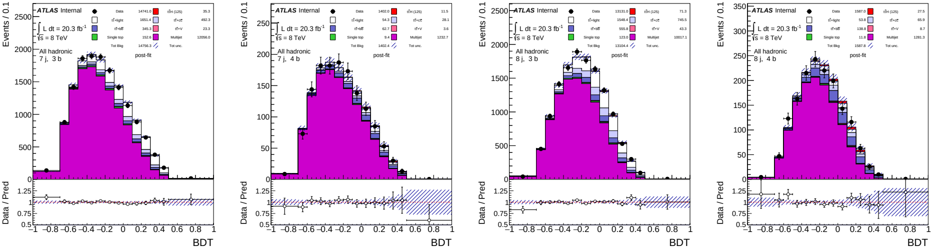
	6j,3b	6j,≥4b	7j,3b	7j,≥4b	8j,3b	8j,≥4b
Multijet	16000 ± 320	1400 ± 66	12000 ± 350	1230 ± 78	10000 ± 490	1300 ± 100
single top	180 ± 59	6.7 ± 3.6	153 ± 12	9.4 ± 4.4	120 ± 47	12 ± 5.7
$t\bar{t}+V$	15 ± 6.2	1.9 ± 1.5	23 ± 8.9	3.6 ± 2.1	43 ± 15	8.7 ± 4.2
$t\bar{t}+b\bar{b}$	230 ± 120	31 ± 17	340 ± 190	63 ± 34	560 ± 320	140 ± 75
$t\bar{t}+c\bar{c}$	350 ± 170	22 ± 11	490 ± 240	28 ± 15	740 ± 360	66 ± 32
$t\bar{t}+light$	1750 ± 270	55 ± 13	1650 ± 340	54 ± 19	1500 ± 450	54 ± 21
$t\bar{t}H$ (125)	21 ± 6.1	5.5 ± 2.7	35 ± 8.6	11 ± 4.4	71 ± 15	27 ± 8.4
Total bkg.	18500 ± 310	1540 ± 61	14700 ± 300	1400 ± 69	13100 ± 340	1590 ± 72
Data	18508	1545	14741	1402	13131	1587

Pre- / post-fit comparisons

Pre-fit:

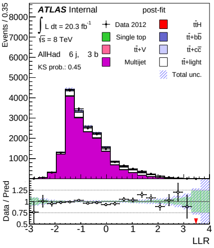


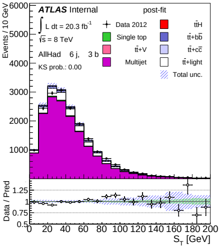
Post-fit:



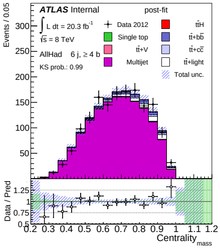
Post-fit variables

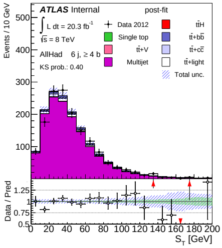
(6j, 3b)



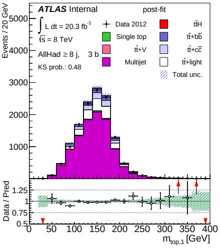


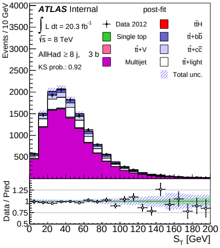
(6j, ≥4b)



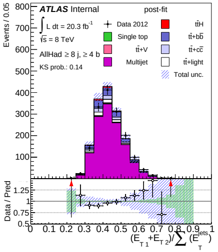


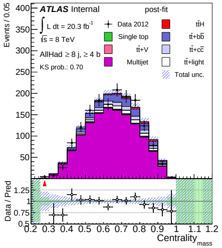
(≥8j, 3b)



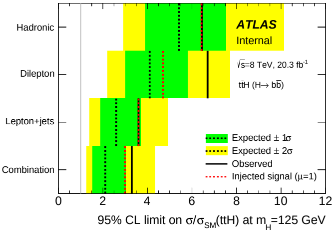


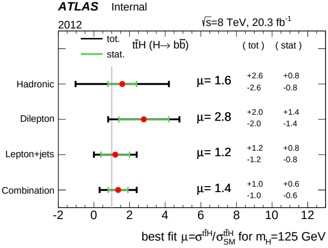
(≥8j, ≥4b)





$t\bar{t}H$ $(H \to b\bar{b})$ combination



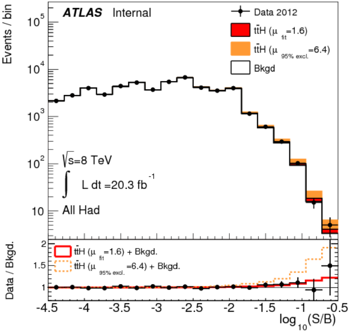


Fully hadronic $t\bar{t}H$ $(H \to b\bar{b})$:

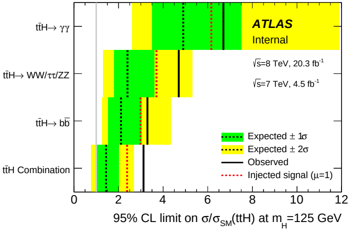
- 95% CL observed (expected) upper limit on $t\bar{t}H$ cross section 6.4 (5.4) $\times$ SM cross section
- Best fit value $\mu = \frac{\sigma_{t\bar{t}H}}{\sigma^{SM}_{t\bar{t}H}} = 1.6 \pm 2.6$

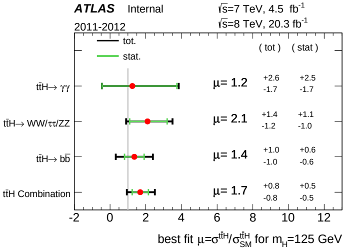
$t\bar{t}H$ $(H \to b\bar{b})$ combination:

- Best fit value $\mu = 1.4 \pm 1$
- Improvement with the addition of fully hadronic channel
  - Upper limit 3.4 (2.2) $\to$ **3.3 (2.1)** $\times$ SM cross section



Run 1 $t\bar{t}H$ combination



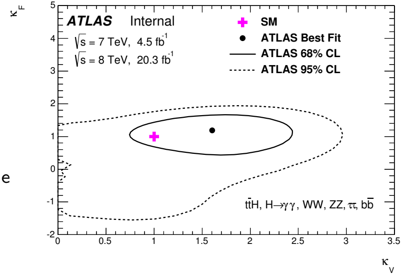


Signal strength and limit:

- 95% CL observed (expected) upper limit on $t\bar{t}H$ cross section 3.1 (1.4) $\times$ SM cross section
- Best fit value $\mu = 1.7 \pm 0.8$

Higgs couplings:

- Best-fit of couplings modifiers κ_V and κ_F is compatible with SM prediction within 1 σ



Conclusions

- First fully hadronic $t\bar{t}H$ ($H \to b\bar{b}$) analysis ever performed
- Description of the tools used in the analysis
  - $\mathrm{TRF_{MC}}$ method for emulation of b-tagging selection in MC
  - Evaluation of trigger efficiency and SF
  - $\mathrm{TRF_{MJ}}$ data-driven method to model MJ background
    - Events with exactly 2 b-tags are used to describe events with exactly 3 and $\geq$4 b-tagged jets
- Description of the main systematic uncertainties
  - $\mathrm{TRF_{MJ}}$ method shape systematics: 5 parametrization of $\varepsilon_{\mathrm{MJ}}$ + 2 reweighting
    - MJ normalization free floating in the fit
  - Uncertainty on $t\bar{t} + b\bar{b}$ cross section is the leading uncertainty of the analysis
- Results of the standalone analysis
  - Best fit signal strength value $\mu = 1.6 \pm 2.6$
  - 95% CL upper limit observed (expected) 6.4 (5.4) $\times$ SM cross section
- Results of the combination with all $t\bar{t}H$ ATLAS channels
  - Best fit signal strength value $\mu = 1.7 \pm 0.8$
  - 95% CL upper limit observed (expected) 3.1 (1.4) $\times$ SM cross section
  - Best-fit of couplings modifiers κ_V and κ_F is compatible with SM within 1 σ

Back-up

Analysis strategy

nj \ nb	2	3	≥4
6	**multijet (MJ)** background *extraction region.* **MJ** defined here as the difference between data and the MC based top-quark background	TRF_{MJ} → TRF_{MJ} →	Fit region
7		TRF_{MJ} → TRF_{MJ} →	Fit region
≥8		TRF_{MJ} → TRF_{MJ} →	Fit region