Boosted $t\bar{t}H$, H$\rightarrow$bb, with ATLAS in LHC Run 2

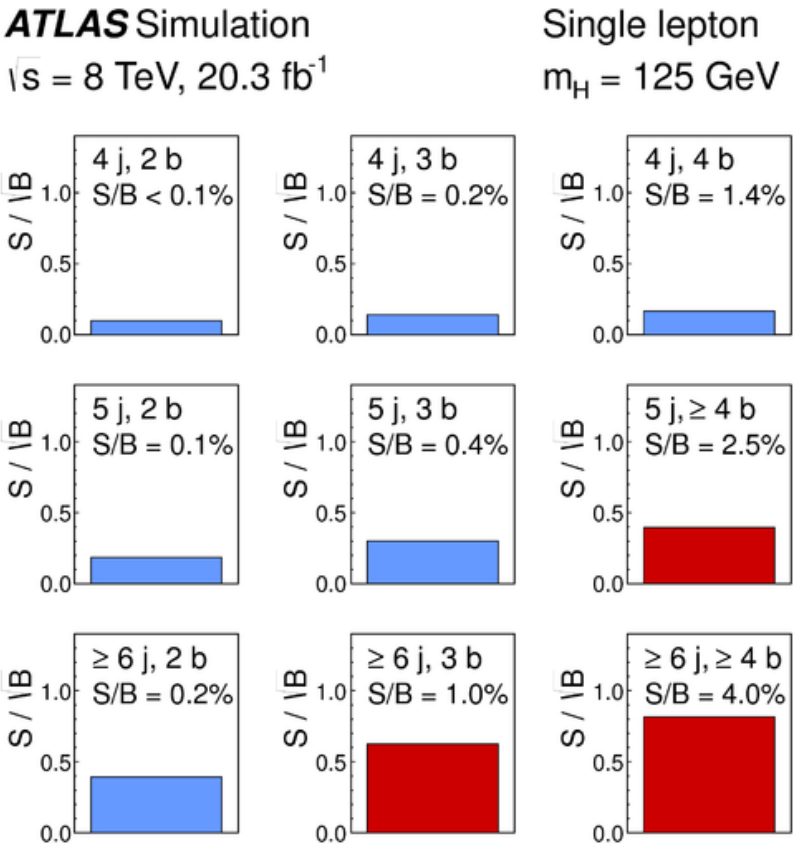
Sam Crawley

University of Glasgow | Experimental Particle Physics

22nd March 2016

Recap resolved semi-leptonic $t\bar{t}H$

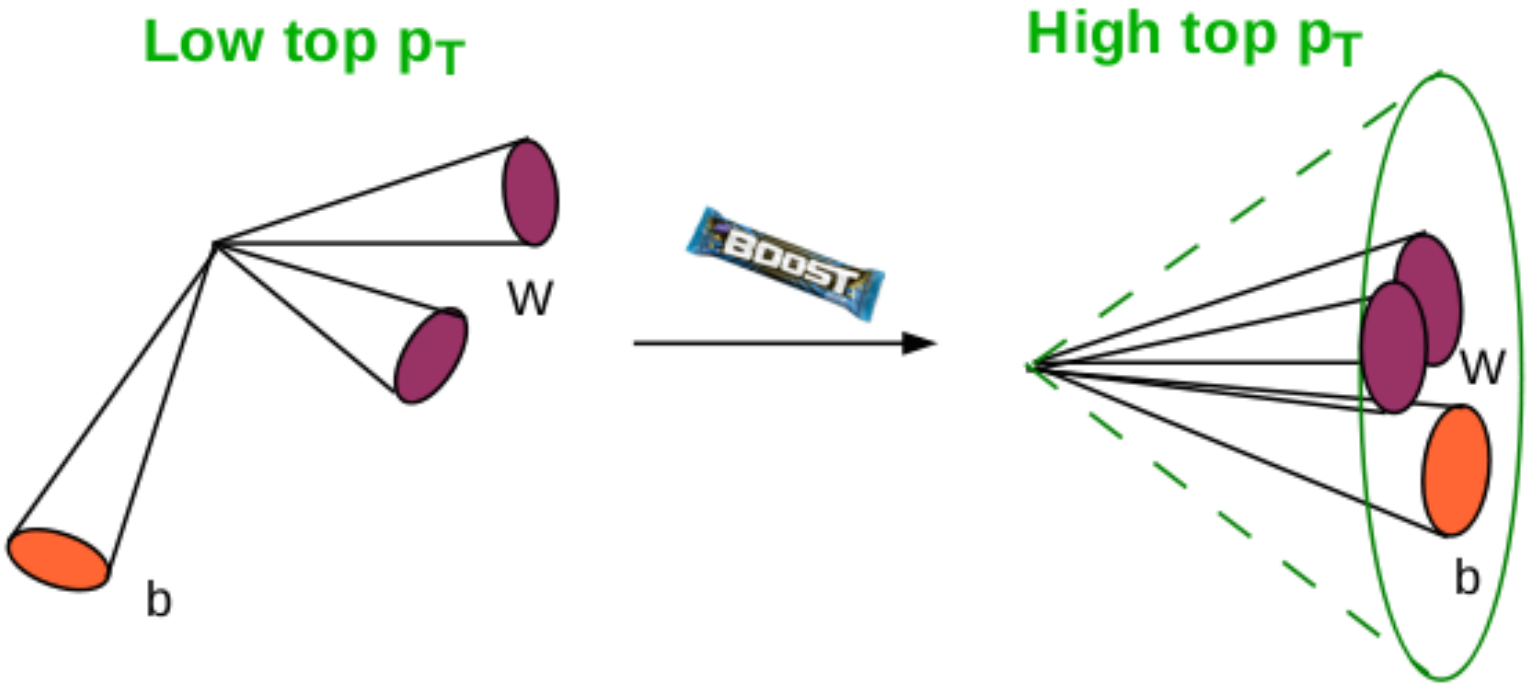
- Split events into regions based on jet and b-jet multiplicities.
- Use regions with low signal to constrain the backgrounds and associated uncertainties.
- Train and apply an MVA (more on this later) on signal-rich regions. (Neural network in run 1)



Eur. Phys. J. C (2015) 75:349

What does "Boosted" mean?

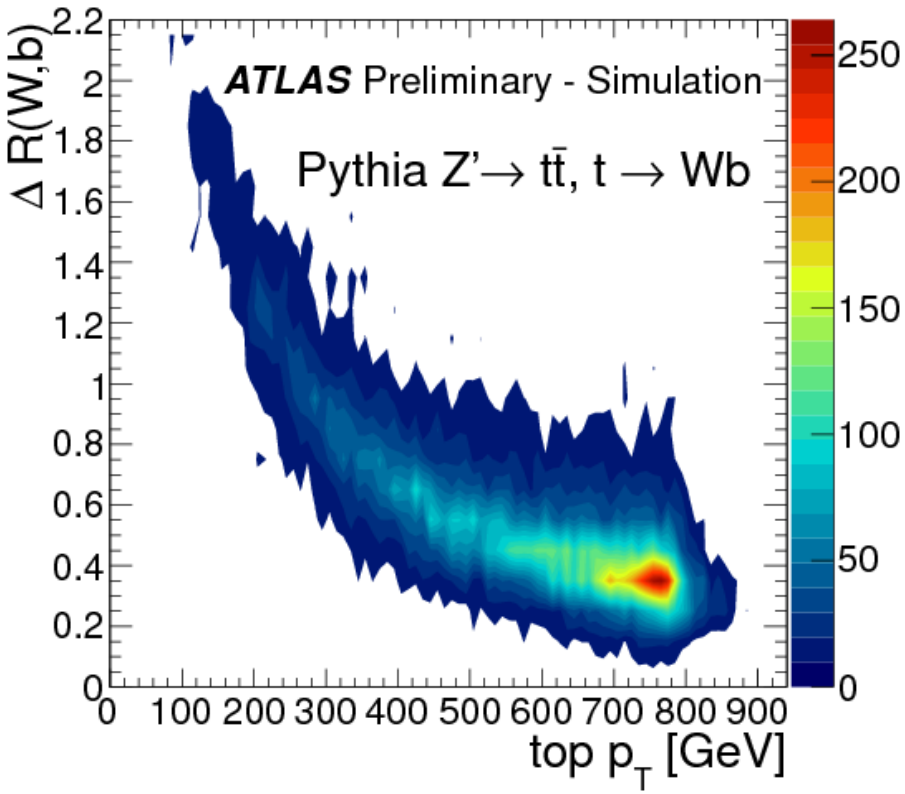
- When a decaying particle has high p_T, its products will take longer to separate and be collimated when hitting the detector.
- Smaller jets will merge into one jet with a larger R parameter.



Run 2 Higher Energy

- Run 2 increases $\sqrt{s}$ to 13/14 TeV.
- More energy available, more likely to have a boosted Top and/or Higgs.
- Detector can't resolve individual jets from Top and/or Higgs if sufficiently boosted.

JHEP09 (2013) 076



- Boosted signal events may not fall into the traditional resolved signal regions.

Advantages of the Boosted regime

- $t\bar{t}H(bb)$ lepton + jets channel suffers from a high combinatorial background.
  - Especially difficult to separate $t\bar{t}H$ from $t\bar{t}$ background.
- Working at higher energies:
  - Allows for the tagging of boosted hadronic Tops and Higgs using jet substructure variables.
  - Reduces the combinatorial background.
  - Provides access to variables involving large-R jets that can be discriminating against background processes.
- We can use all this to better reconstruct the event and separate signal and background.

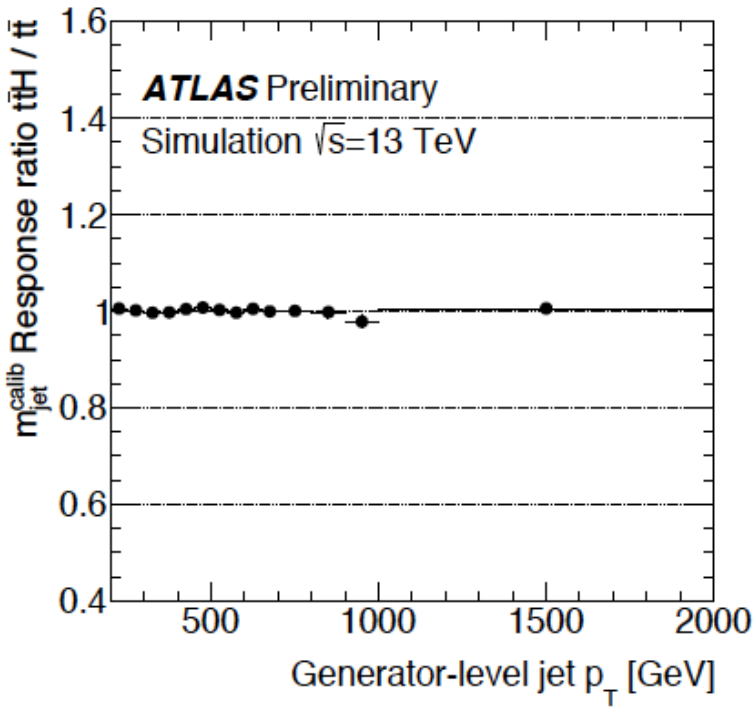
- Aim to analyse the boosted events separately to gain an overall improvement in sensitivity - need to develop orthogonal selections

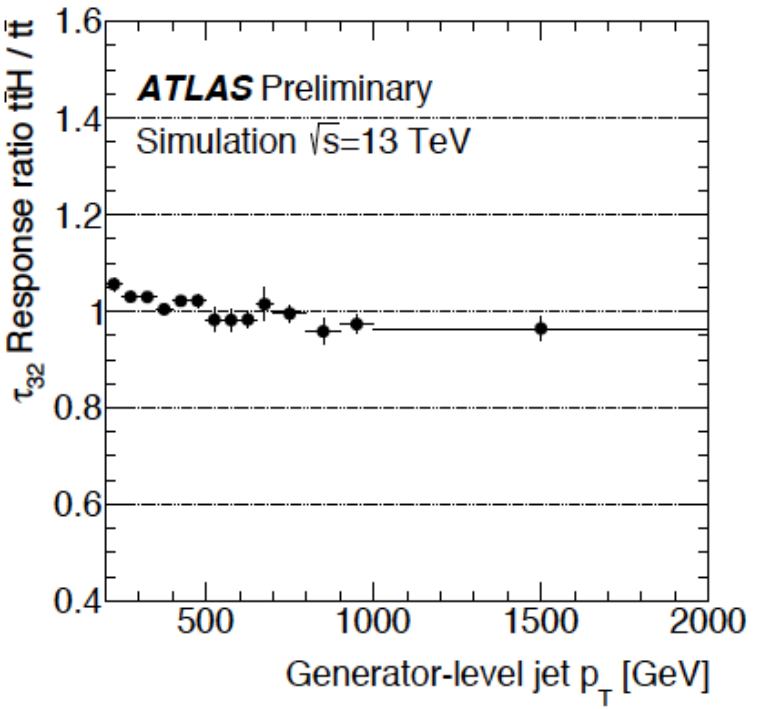
Boosted Signal Region

- Preliminary event selection requires at least 1 large jet with R = 1.0 and $p_T > 250$ GeV, at least 1 of which is top-tagged.
- Removed small jets that were within $\Delta R = 1.5$ of the large top-tagged jets.
- Signal Region defined as requiring $\geqslant 3$ small-R jets of which at least 2 are b-tagged.
- B-tagging at 77% WP and Top-tagging at 80% WP) efficiencies.

Boosted Top Tagger

ATL-PHYS-PUB-2015-053

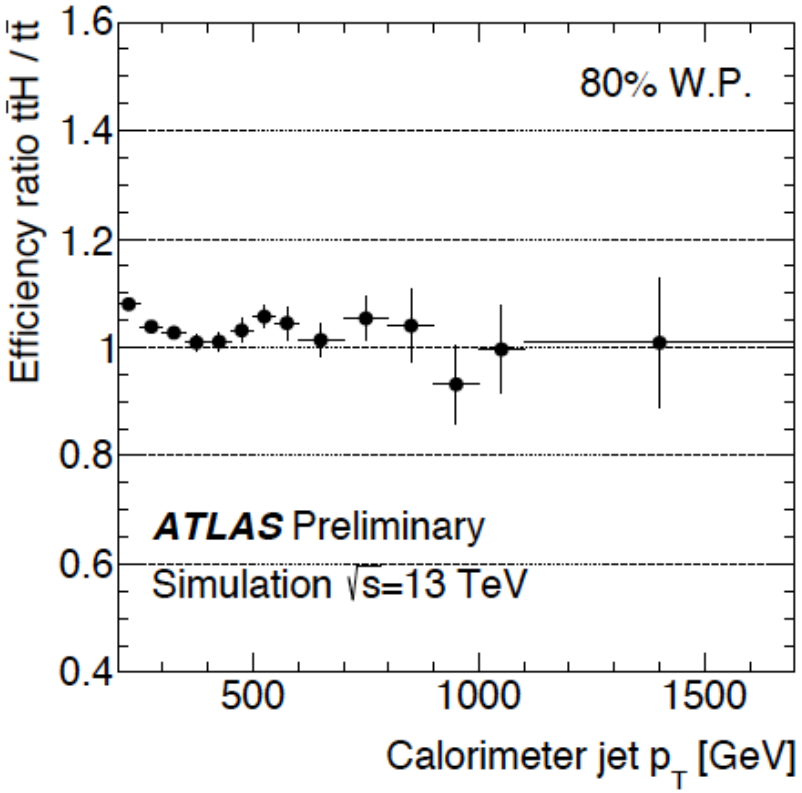




- Attempt to identify large R jet as the hadronic top using the algorithm described here., designed for boosted hadronic tops in Standard Model $t\bar{t}$.
- Algorithm uses the calibrated jet mass, m_{jet}^{calib} and the N-subjetiness ratio, τ_{32}.
- Testing was done in $t\bar{t}H$ due to its busy final state, to see

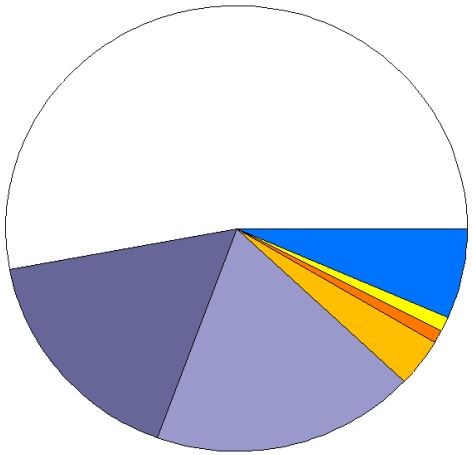
Boosted Top Tagger

ATL-PHYS-PUB-2015-053



- The efficiency is higher for $t\bar{t}H$ at low p_T and is consistent with $t\bar{t}$ at higher p_T.
- More visible in top-tagged jets that are geometrically close the the Higgs.
- Shows the efficiency is not diminshed by contamination of the large-R by other jets in a busy environment.

Signal Region Yields at 3.2 fb^{-1}.



ATLAS Work in Progress

$\sqrt{s}$ = 13 TeV --boosted--

- $t\bar{t}$+light
- $t\bar{t}$+$b\bar{b}$
- $t\bar{t}$+$c\bar{c}$
- W+jets
- Z+jets
- Dibosons
- Single top

Event Yields

	Yield	+/-
$t\bar{t}$+light	1120	8.94
$t\bar{t} + b\bar{b}$	349	4.87
$t\bar{t} + c\bar{c}$	393	5.17
W+jets	71.9	4.56
Z+jets	19.2	1.30
Diboson	21.8	1.02
Single Top	139	1.80
$t\bar{t} + V$	20.9	0.18
Sum bkg.	2130	5.28
Data	2310	
ttH exp.	16.0	0.19

Resolved minus Boosted at 3.2 fb^{-1}.

Need to investigate the overlap as we will veto events from the resolved analysis that fall into the boosted signal region to avoid double counting.

ATLAS WORK IN PROGRESS

Existing Resolved

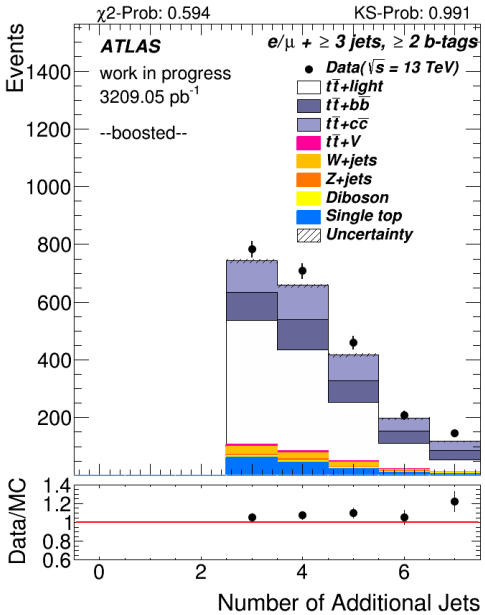
	4j,2b	5j,2b	≥6j,2b	4j,3b	5j,3b	≥6j,3b	4j,≥4b	5j,≥4b	≥6j,≥4b	
ttH	18.4	27.0	56.6	8.3	15.5	38.9	1.4	4.7	18.3	
all bkg	60319.4	35590	23989.8	6217.3	5029.3	4787.6	128.8	264.7	571.6	
S/B	0.0003	0.0008	0.0024	0.0013	0.0031	0.0081	0.0112	0.0177	0.0320	
S/√B	0.08	0.14	0.37	0.11	0.22	0.56	0.13	0.28	0.77	

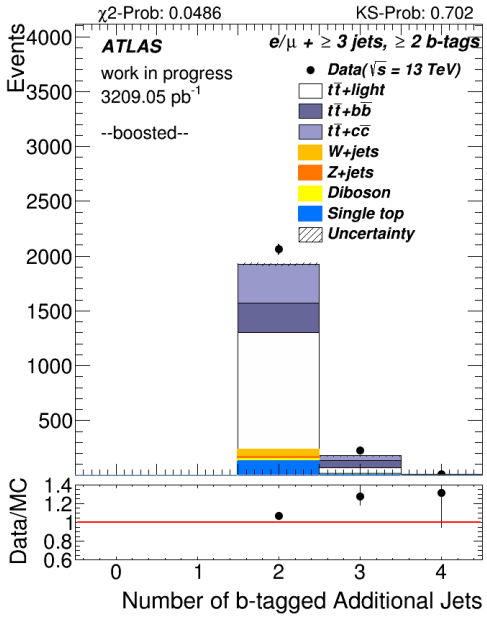
Resolved-minus-boosted **+ boosted**

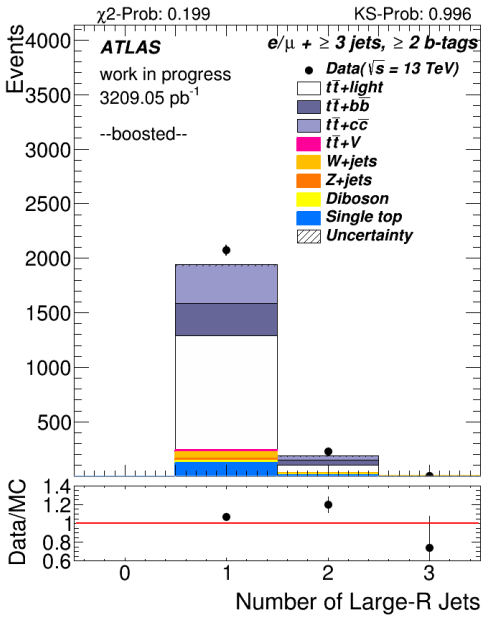
	4j,2b	5j,2b	≥6j,2b	4j,3b	5j,3b	≥6j,3b	4j,≥4b	5j,≥4b	≥6j,≥4b	"3211"
ttH	18.4	26.2	52.9	8.2	14.6	32.9	1.4	4.3	14.4	16.0
all bkg	60254.5	35204.6	23140.3	6163.7	4858.4	4304.8	127.5	251.5	478.6	2116.2
S/B	0.0003	0.0007	0.0023	0.0013	0.0030	0.0076	0.0110	0.0169	0.0300	0.0076
S/√B	0.07	0.14	0.35	0.10	0.21	0.50	0.12	0.27	0.66	0.35

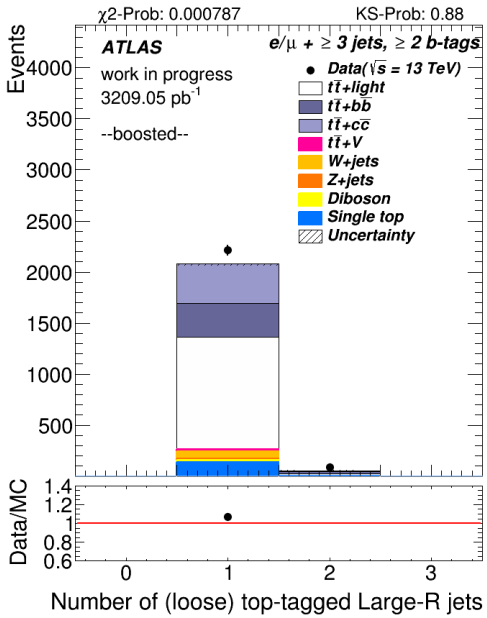
Boosted events spread out, low impact on resolved SRs.

Signal Region Kinematics (1)

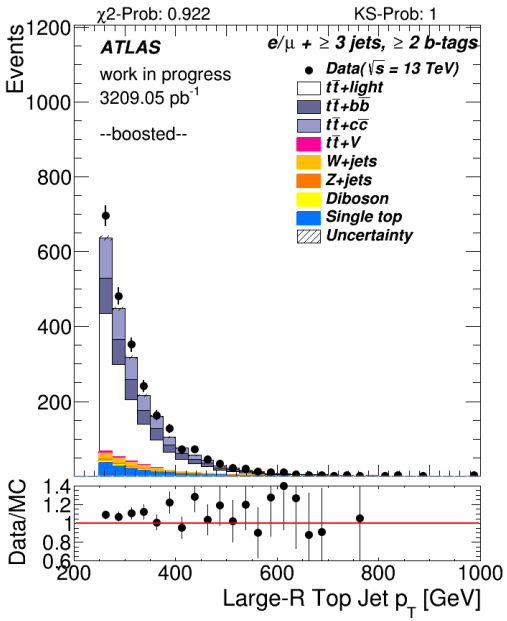


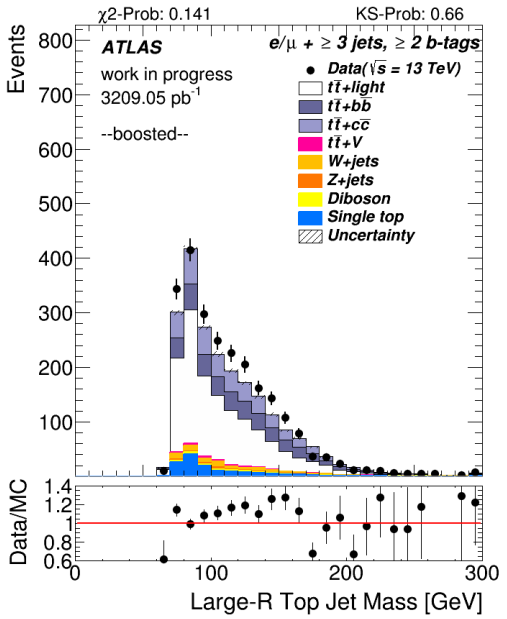


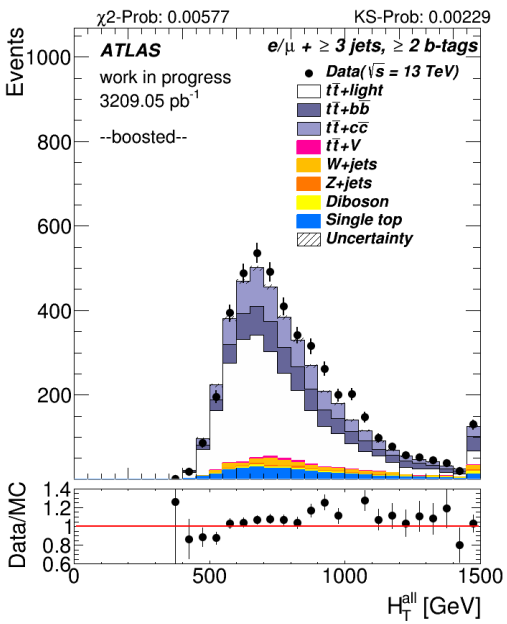


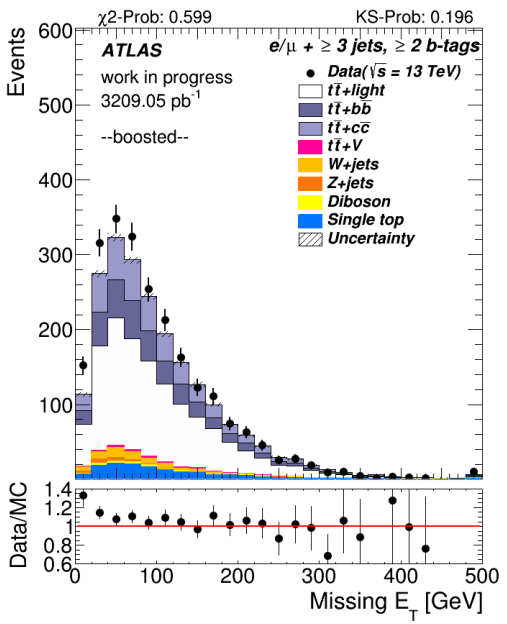


Signal Region Kinematics (2)



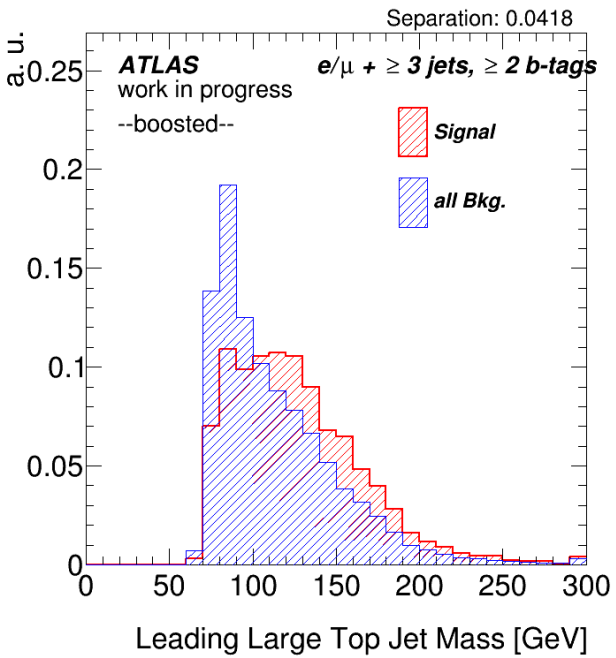


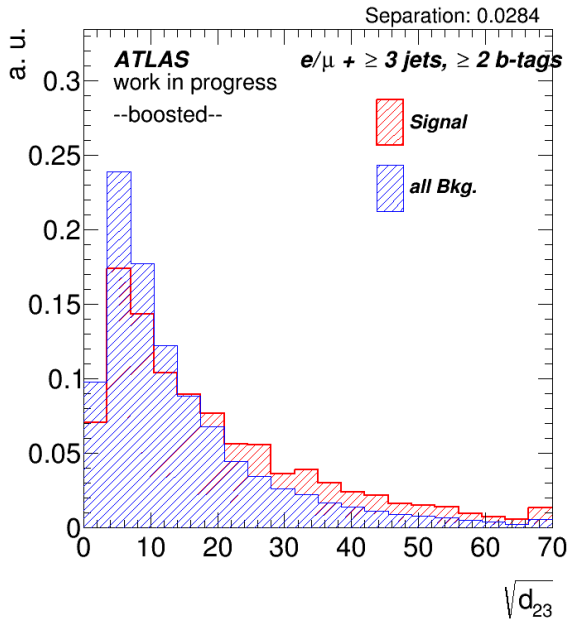


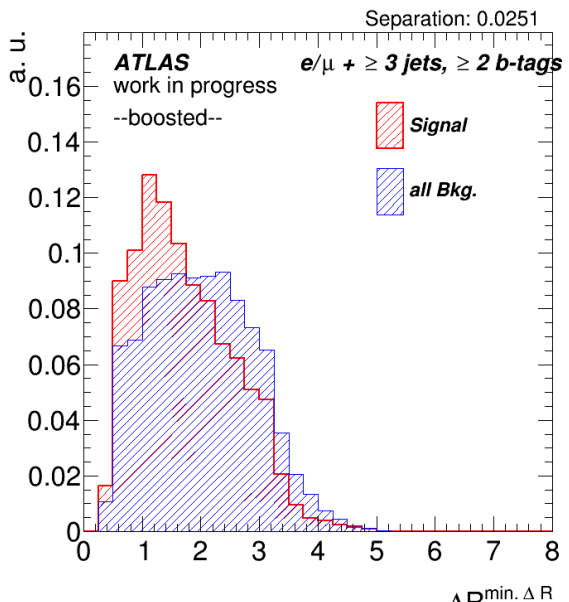


Discriminating Variables

- Some promising variables that help separate $t\bar{t}H$ from its backgrounds.
- As with the resolved analysis, none are good enough to cut on alone.

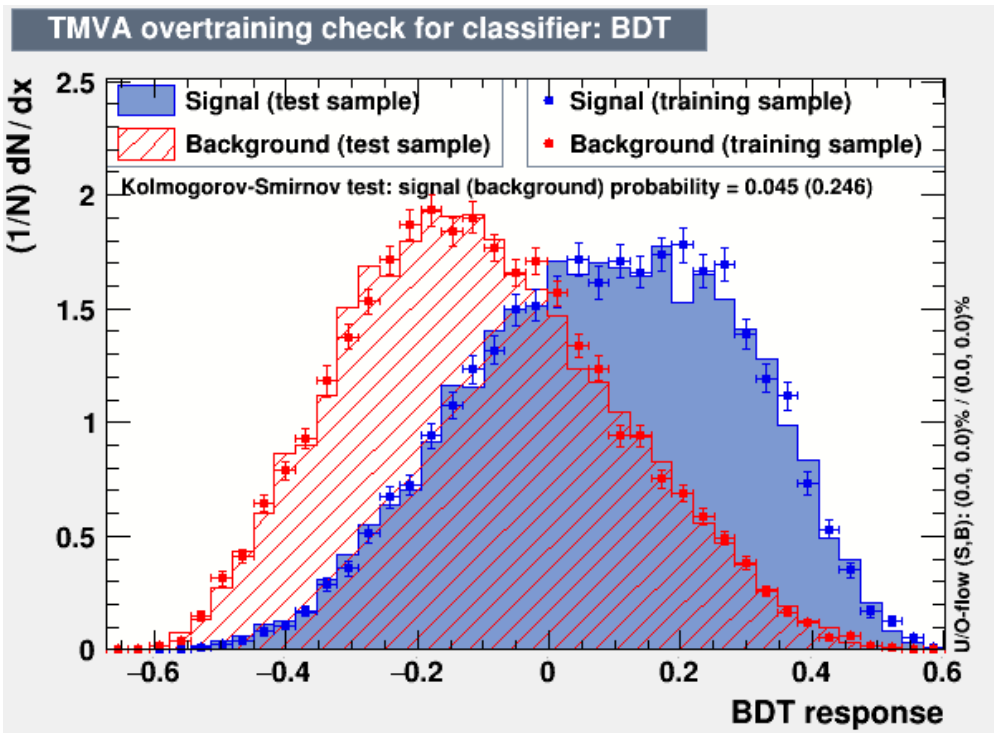






TMVA Boosted Decision Tree

- Multiple discriminating variables are used to train a BDT on the signal region.
- Currently going through iterations to improve the separation of the BDT output and then fit to gain an improved combined limit.



Conclusions

- More energy = more boosted hadronically decaying particles.
- Opportunity to take advantage of the boosted events in $t\bar{t}H$.
- Separate selection and analysis needed for these events.
- Using discriminating variables to train an MVA on the boosted region and then fit. Iterate with different variables and BDT settings/signal regions to optimise.
- Combine with resolved for overall improvement.