BDTs in the Search for $t\bar{t}H$ Production with Higgs Decays to $b\bar{b}$ at ATLAS

ATLAS-CONF-2016-080

An observation of Higgs boson production in association with a top quark pair ($t\bar{t}H$) provides a direct measurement of the top quark Yukawa coupling and probes the Standard Model. The ATLAS search for ttH->bb relies on Boosted Decision Trees for Higgs reconstruction and signal-background discrimination. Higgs boson decays to two bottom quarks, and top pair decays with one or two leptons are considered.
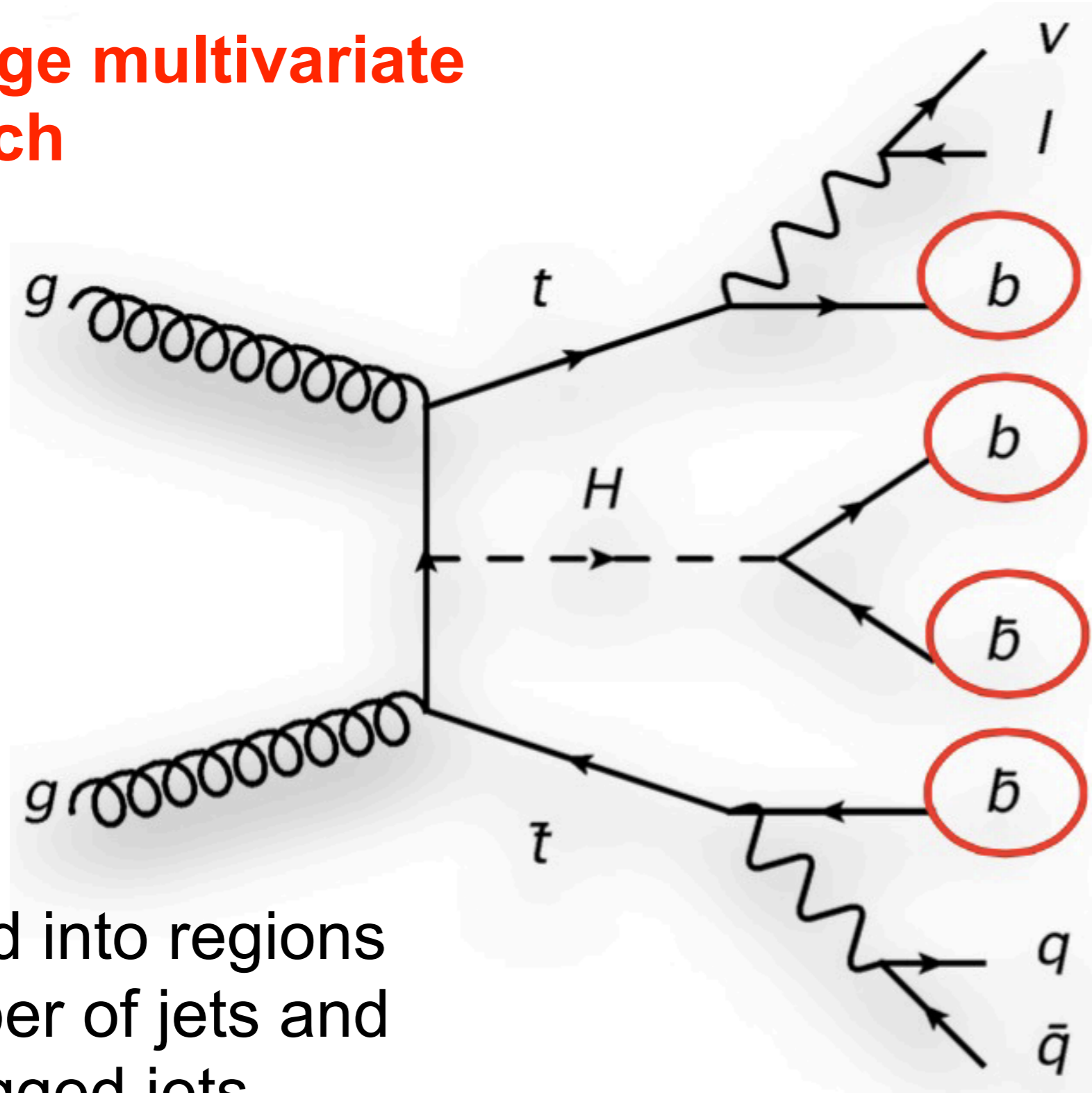
$t\bar{t}H$, H->$b\bar{b}$

- Higgs in association with top quarks is a rare production mode: only 1% of total Higgs cross section
- Largest Higgs branching ratio is to b quarks, 58%
- Select $t\bar{t}$ decays with one or two leptons to reduce backgrounds

Challenging analysis:
- High **combinatorics from 4 b-jets** final state
- **Large backgrounds** from $t\bar{t}$ + b-jets

➡ **two stage multivariate approach**

Analysis divided into regions based on number of jets and number of b-tagged jets.

Boosted Decision Trees

A Boosted Decision Tree (BDT) is a multivariate algorithm used for **particle and event classification**.
- Shallow learner: robust against over-training
- Decision tree ensemble built by incrementally targeting previously mis-classified training instances (Adaptive Boosting)
- BDTs are trained on signal and background Monte Carlo simulated data

Reconstruction BDT

Matches observed **b-jets** to final state quarks from either **Higgs or top** decays.
- reconstruction of the Higgs boson mass from large combinatorics of b-jets
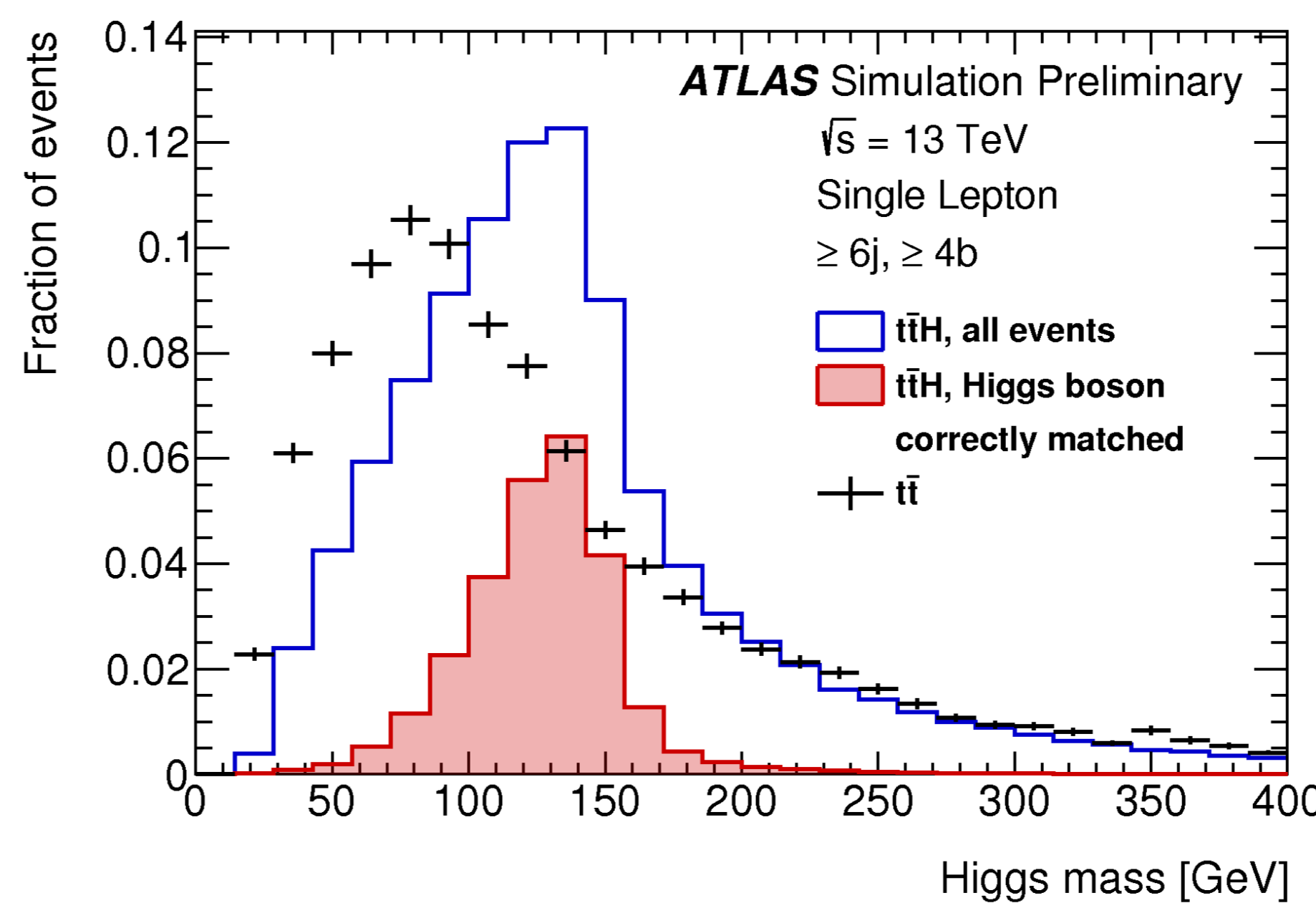

Train on up to 22 kinematic variables depending on region.
- must be well modelled against data.

Normalised reconstructed mass of Higgs candidate from ttH MC and fraction of events correctly matched, compared to $t\bar{t}$ background.

Matching within ΔR<0.3 is up to 42% efficient.

Results

Likelihood fit of MC to data in 18 regions, where high S/B signal regions are high score bins of BDT output:
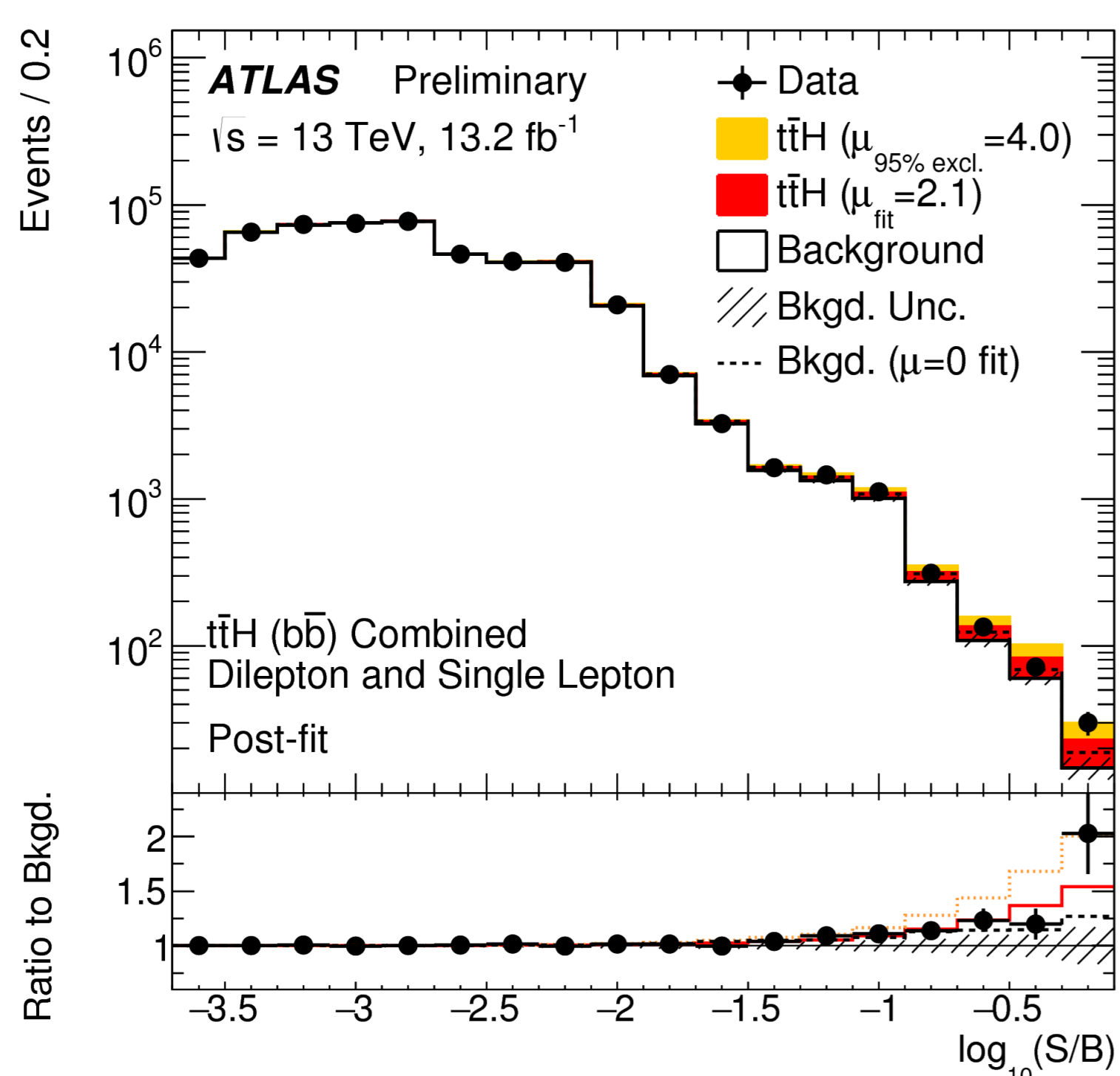
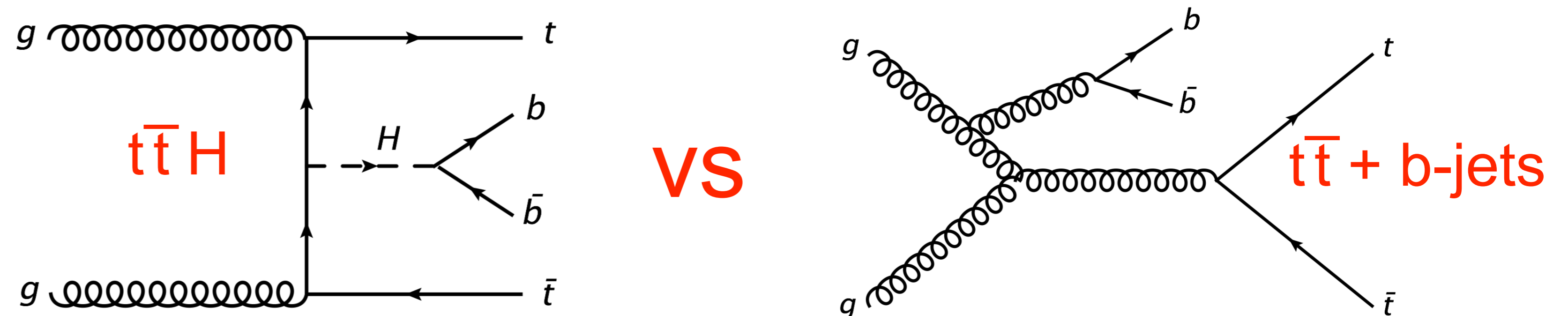$t\bar{t}H$ signal larger than 4.0 times the SM prediction excluded at 95% C.L.

Observed signal strength **μ = 2.1 ± 1.0** . Disfavours background-only hypothesis by 2 std. deviations

Sensitivity limited by $t\bar{t}$ + b-jets modelling uncertainty.
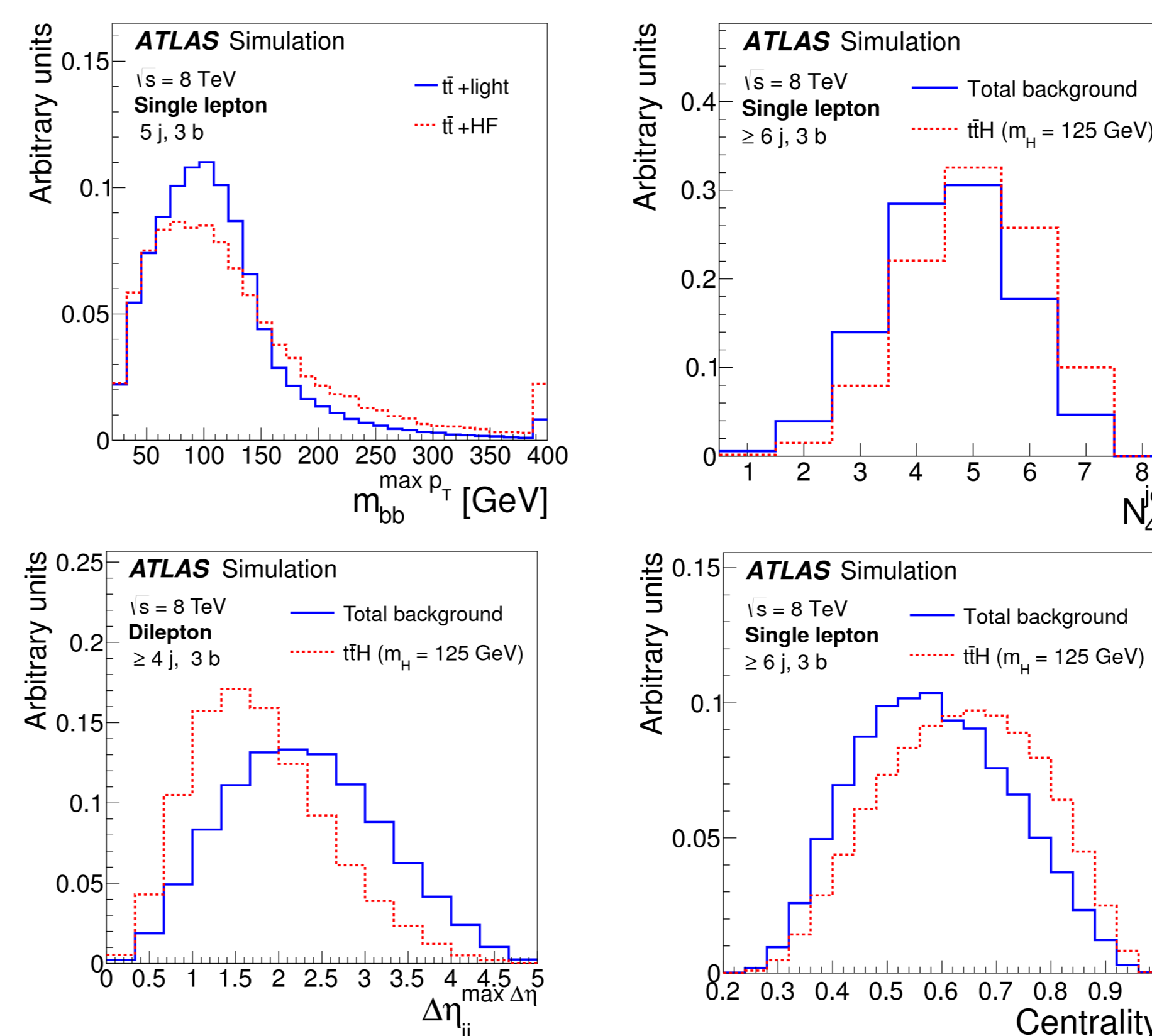
Stay tuned for full 2016 results !

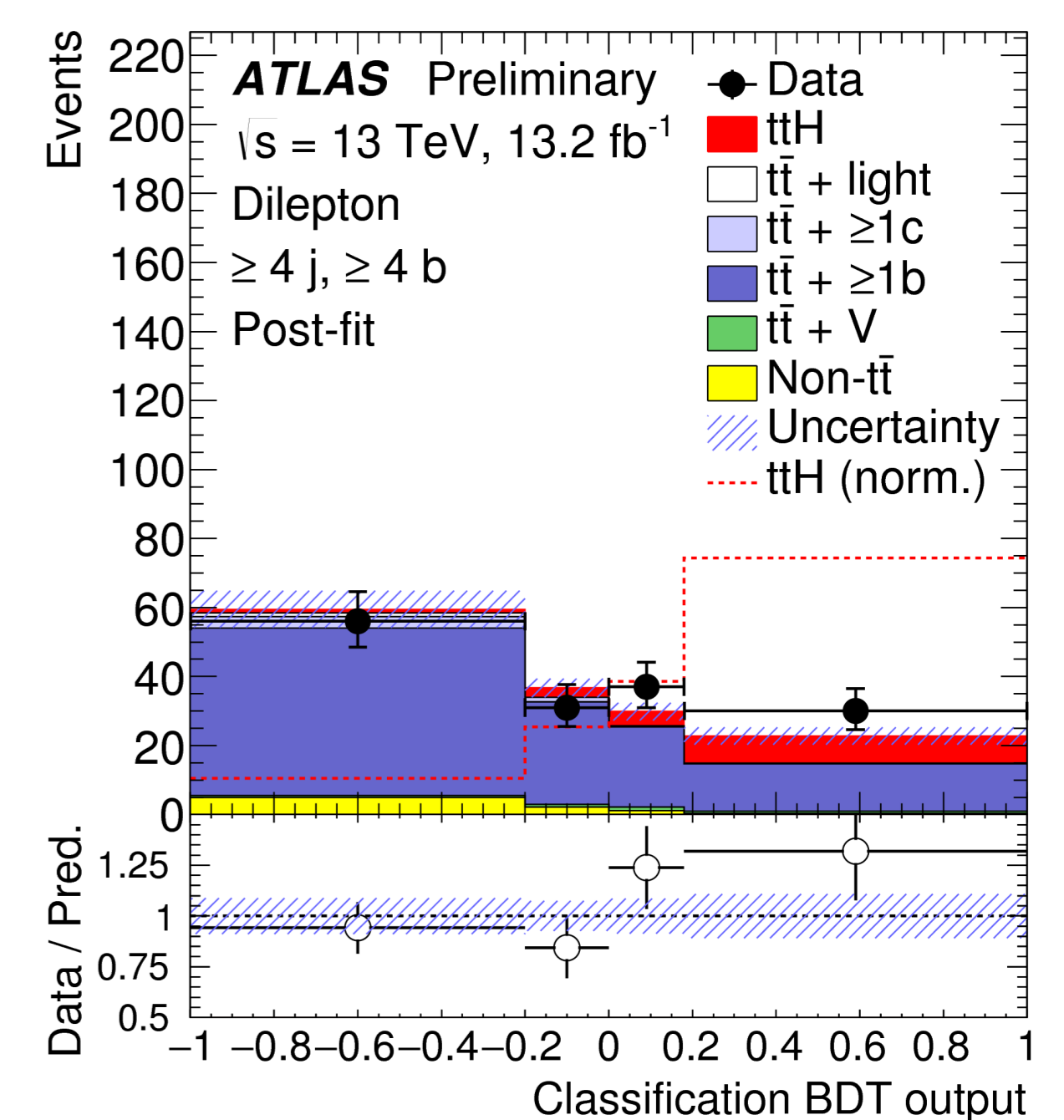## Classification BDT

$t\bar{t}H$ VS $t\bar{t}$ + b-jets
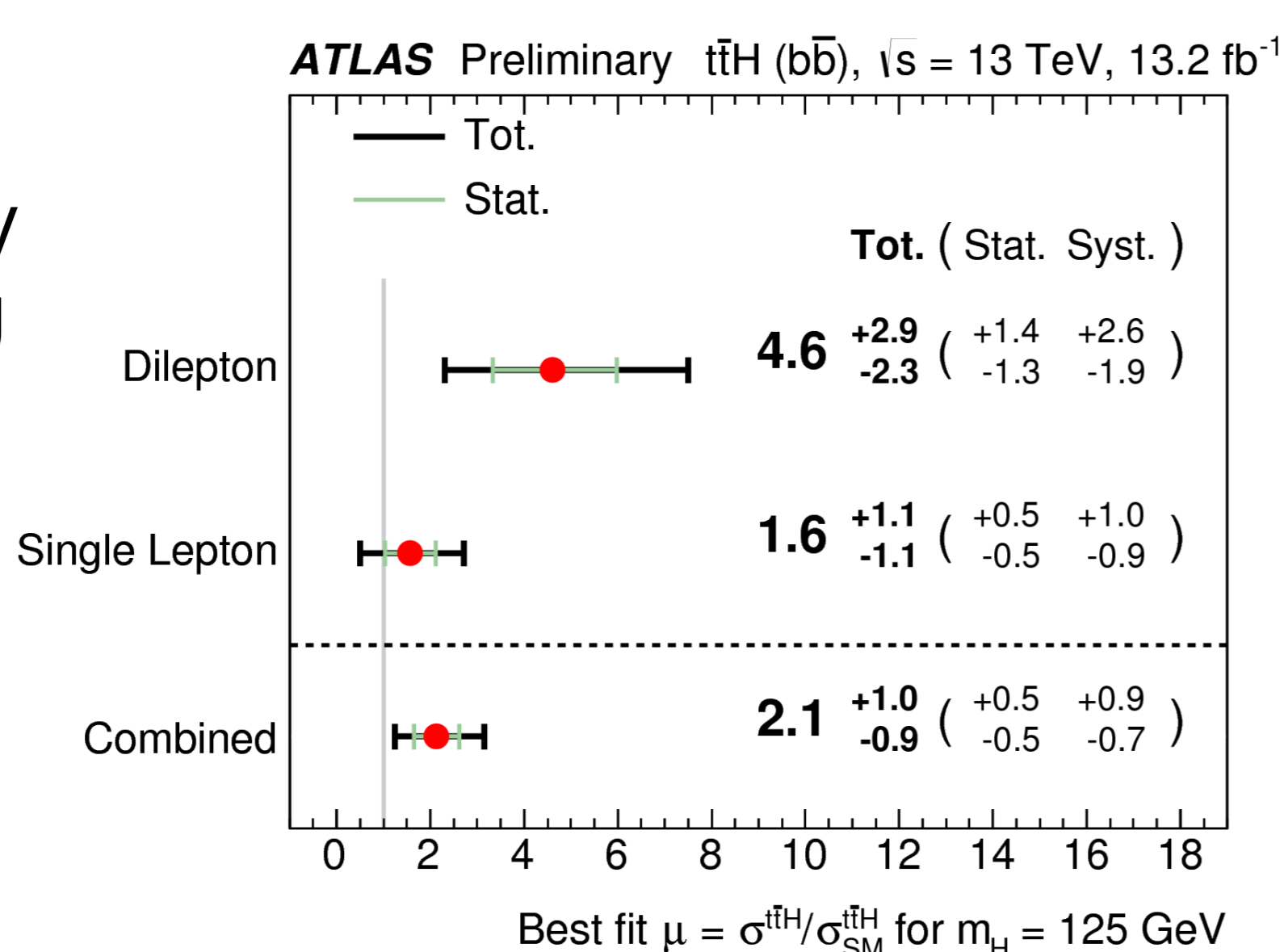
Variables using and not using event reconstruction information are inputs to the Classification BDT, trained to classify events as more signal or background-like.

Train on up to 14 variables, examples are:

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

Individual small kinematic differences combined in the BDT provide better separation of signal from the background.

BDT scores used in fit

Paul Glaysher (DESY), for the ATLAS Collaboration