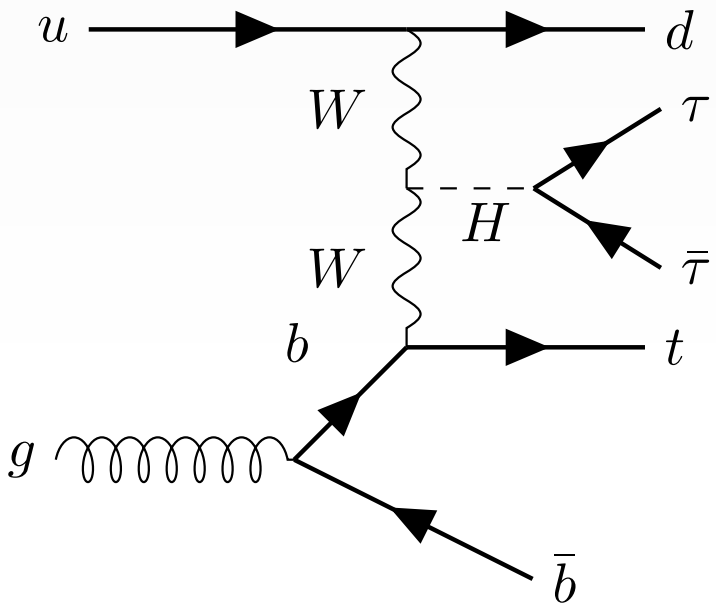
tH($\tau\tau$) Review and plan

Christian Kirfel
on behalf of the tH tau channels team

10th November 2021

UNIVERSITÄT BONN

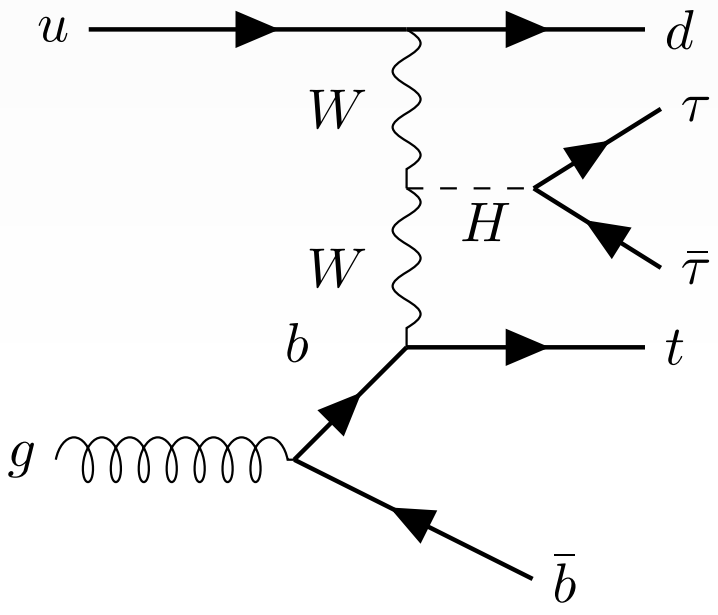
Selection lepditau



- n-jets: 2-6 (b-jets: **1**)
- b-jet WP: 70 DL1r
- nLeptons & nTaus: **1e**/μ **2**τ_{had}OS
- $E_{\text{T,miss}}$: no cut (to 800 GeV)

- jets:
  - $p_T > 25$ GeV
  - $|\eta| < 4.5$
  - EMPFlow
- electrons:
  - $p_T > 20$ GeV trigger matched 27 GeV
  - $|\eta| < 2.5$ not in 1.37 - 1.52
  - WP: Tight ; isolation: PLIVTight
- muons:
  - $p_T > 20$ GeV trigger matched 27 GeV
  - $|\eta| < 2.5$
  - WP: Tight ; isolation: PLIVTight
- taus:
  - $p_T > 20$ GeV trigger matched 27 GeV
  - $|\eta| < 2.5$ not in 1.37 - 1.52
  - WP: RNNMedium
  - ASG recommended OLR (τ_{had} remove jets)

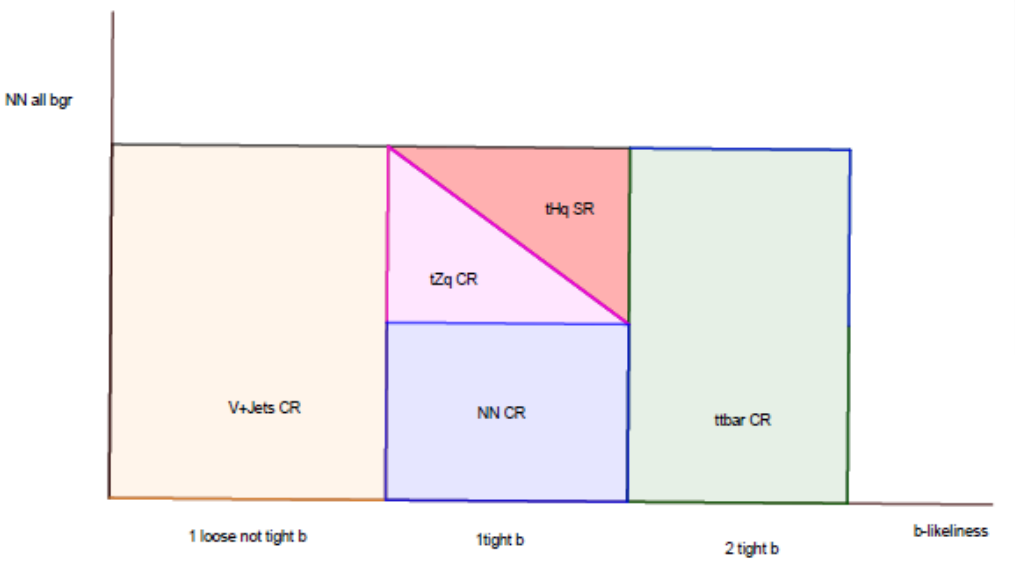
Selection dileptau



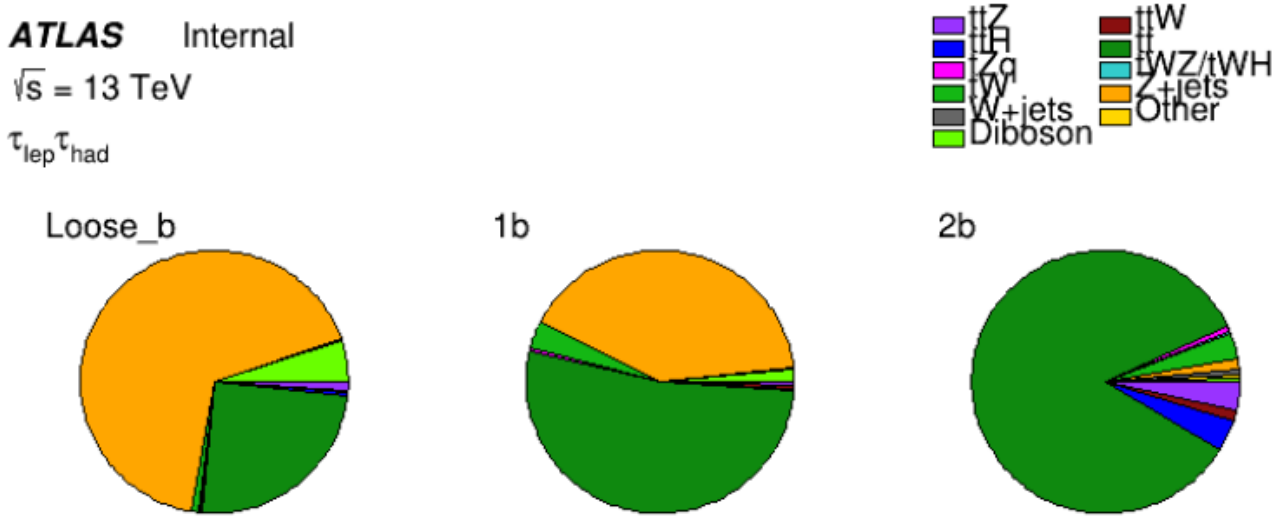
- n-jets: 2-6 (b-jets: **1**)
- b-jet WP: 70 DL1r
- nLeptons & nTaus:
  2e/μ **1**τ_{had}(1 OS light lepton)
- $E_{\text{T,miss}}$: no cut (to 800 GeV)

- jets:
  - $p_T > 25$ GeV
  - $|\eta| < 4.5$
  - EMPFlow
- electrons:
  - $p_T > 20$ GeV trigger matched 27 GeV
  - $|\eta| < 2.5$ not in 1.37 - 1.52
  - WP: Tight ;
    isolation: PLIVTight
- muons:
  - $p_T > 20$ GeV trigger matched 27 GeV
  - $|\eta| < 2.5$
  - WP: Tight ; isolation: PLIVTight
- taus:
  - $p_T > 20$ GeV trigger matched 27 GeV
  - $|\eta| < 2.5$ not in 1.37 - 1.52
  - WP: RNNMedium
  - ASG recommended OLR (τ_{had} remove jets)

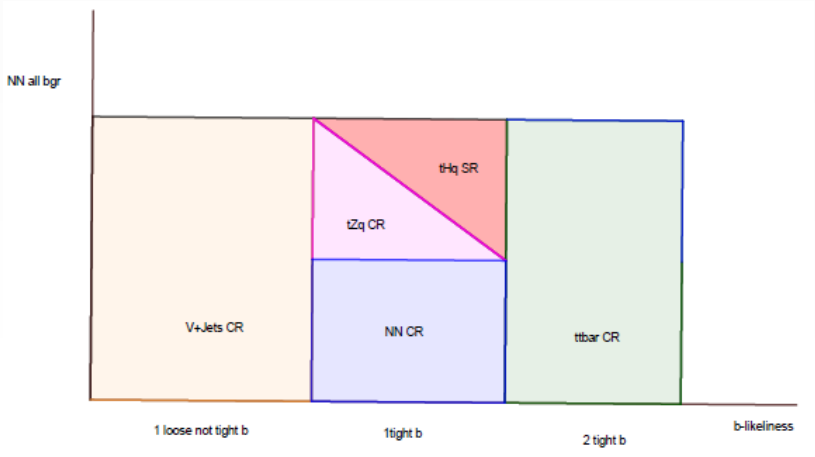
Region definitions dileptau



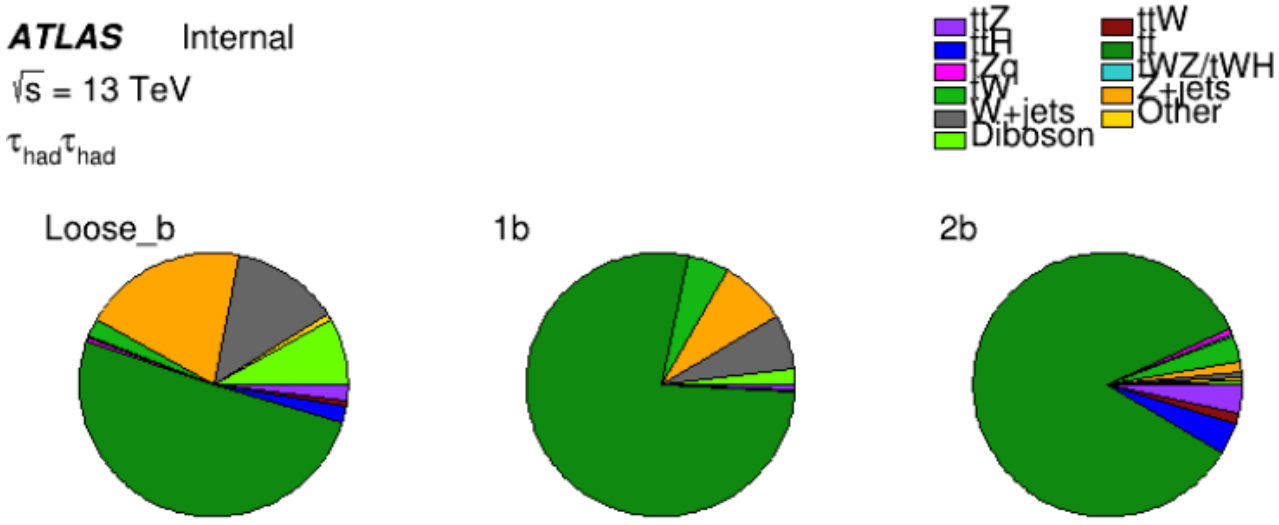
1. Cut on number ob b-jets
2. Cut on MVA score
3. Separate tZq



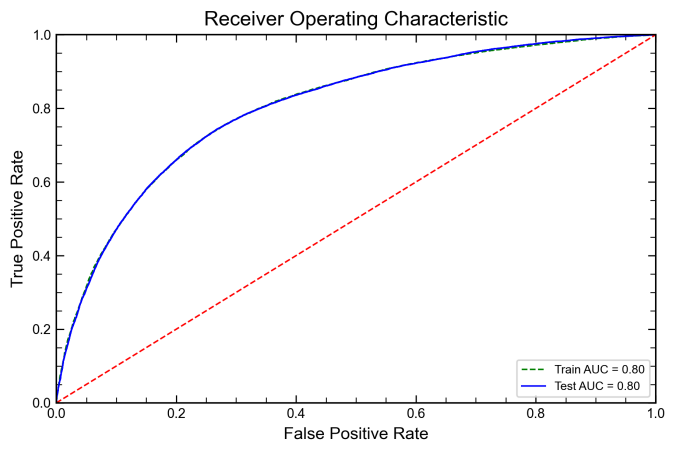
Region definitions lepditau

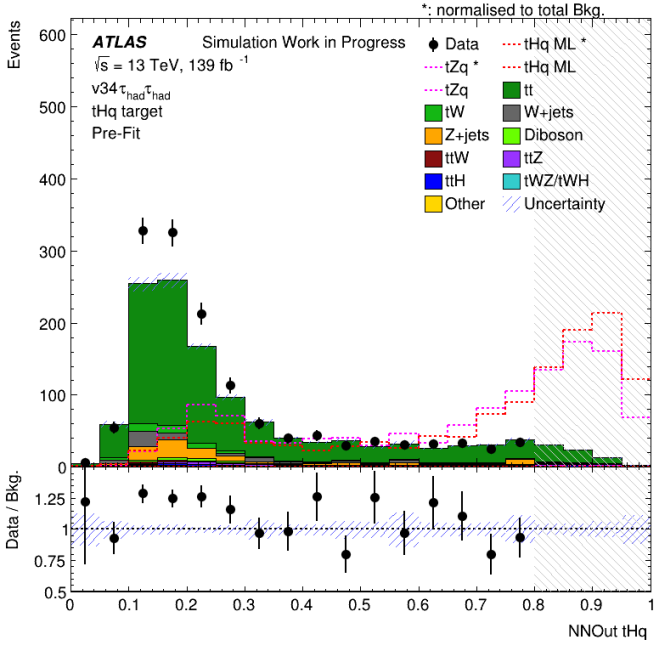


1. Cut on number ob b-jets
2. Cut on MVA score
3. Separate tZq



Lepditau Neural network





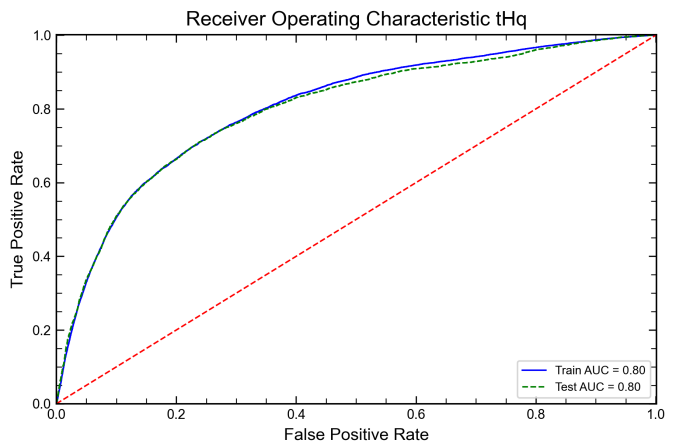
Setup

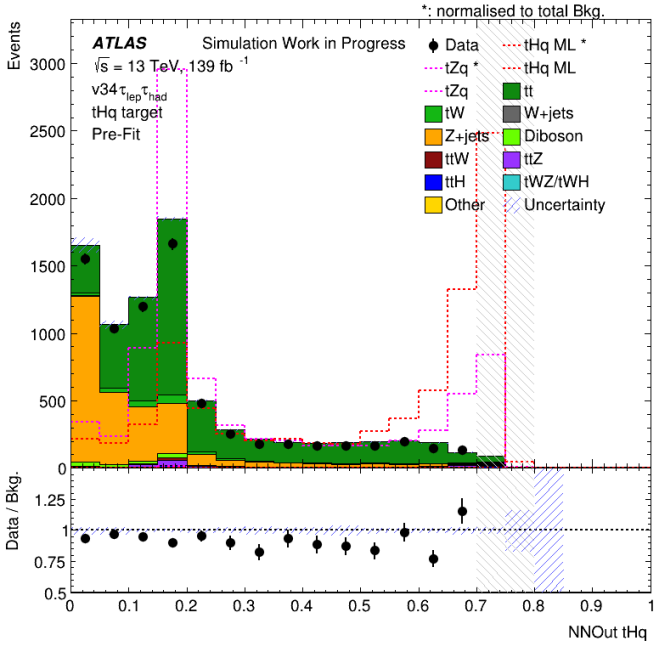
- Optimisation: Evolutionary + grid search
- Model: Categorical (currently binary for v34)
- Variables: final state kinematics
- Currently performance problems for v34
- Train on absolute, predict on full weights

Plans

- Validate categorical setup in lepditau
- Rerun the optimisation for v34
- Rank and test variables

Dileptau Neural network





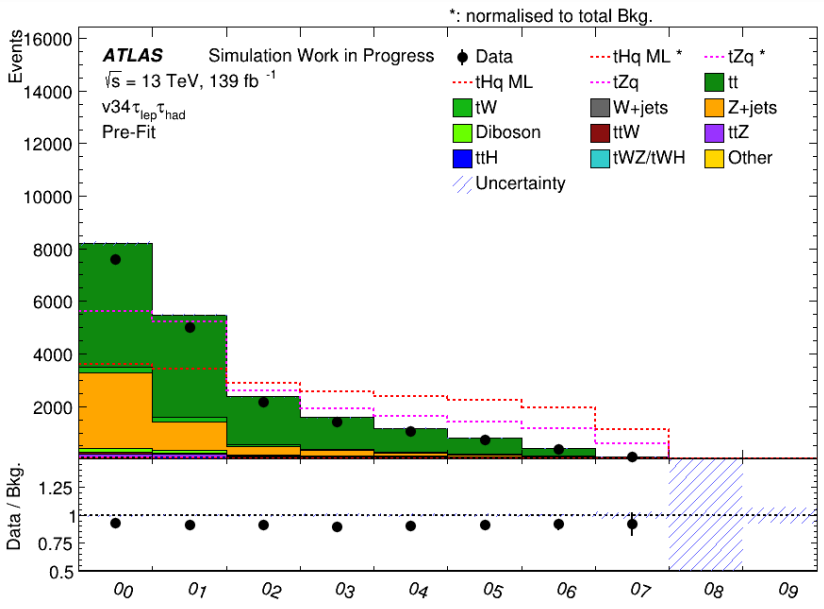
Setup

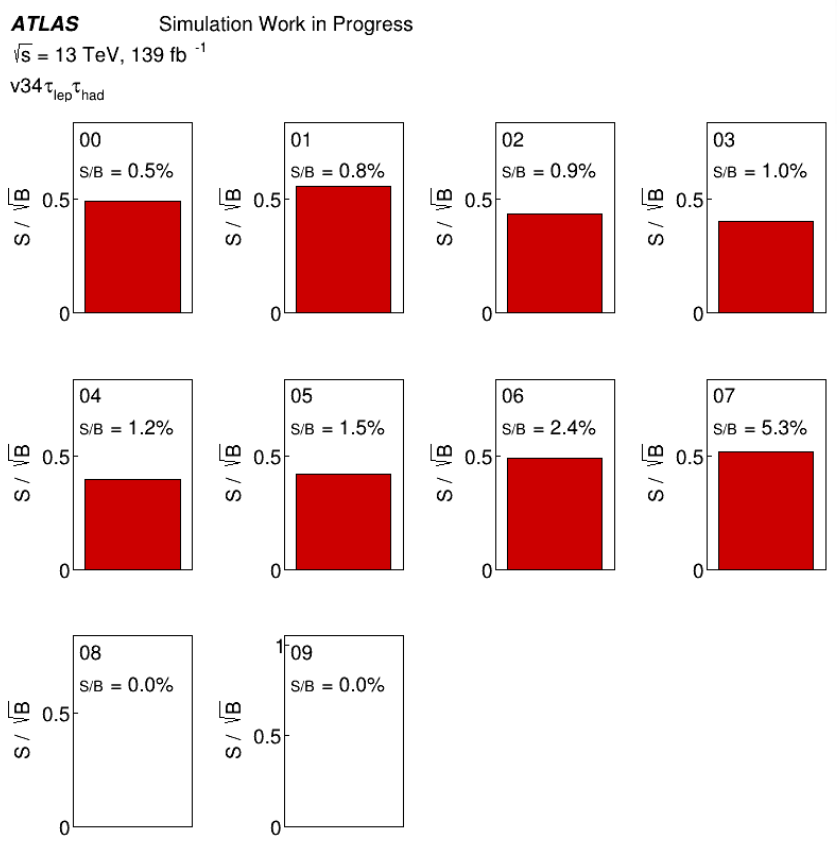
- Optimisation: Evolutionary + grid search
- Model: Categorical (Treating tZq separately)
- Variables: final state kinematics
- Train on absolute, predict on full weights

Plans

- Rerun the optimisation for v34
- Rank and test variables

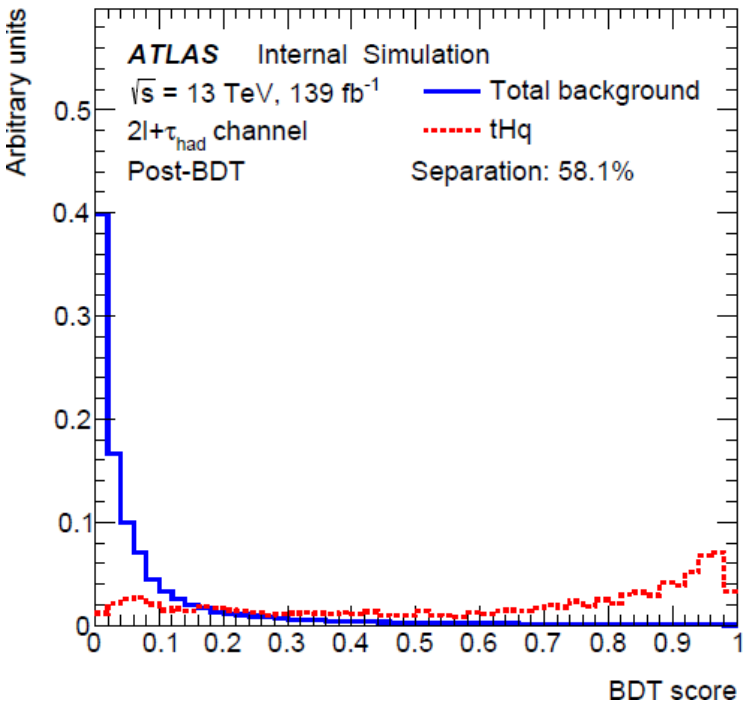
Dileptau S/B


*: normalised to total Bkg.
ATLAS Simulation Work in Progress
√s = 13 TeV, 139 fb -1
v34τlepτhad
Pre-Fit
Data
tHq ML
tW
Diboson
ttH
Uncertainty
tHq ML *
tZq
W+jets
ttW
tWZ/tWH
tZq *
tt
Z+jets
ttZ
Other
Events
Data / Bkg.
00 01 02 03 04 05 06 07 08 09


ATLAS Simulation Work in Progress
√s = 13 TeV, 139 fb -1
v34τlepτhad
00 S/B = 0.5%
01 S/B = 0.8%
02 S/B = 0.9%
03 S/B = 1.0%
04 S/B = 1.2%
05 S/B = 1.5%
06 S/B = 2.4%
07 S/B = 5.3%
08 S/B = 0.0%
09 S/B = 0.0%
S / √B

Dileptau BDT

- Trained using v34
- PreBDT: $n(b\text{-jets}) \geq 1$
- Hyperparameters optimised using Genetic Algorithm (see Backup slides)
- These metrics and distribution include negative weights



max_depth	4
objective	binary:logistic'
learining_rate	0.1
n_estimators	1500
min_child_weight	3.39609×10^{-6}
tree_method	gpu_hist
n_jobs	-1
scale_pos_weight	534.3914

Hyperparameters of the BDT after first GA optimisation.

K-Fold	roc_auc	log_loss
0	0.727393864	0.4087744
1	0.731473309	0.4086942
2	0.728292148	0.4099594
3	0.732439999	0.4123000
4	0.732329792	0.4113235

Metrics for BDT. Using the set of 76 features and the first results form the GA.
AUC = 0.7304 ± 0.0009

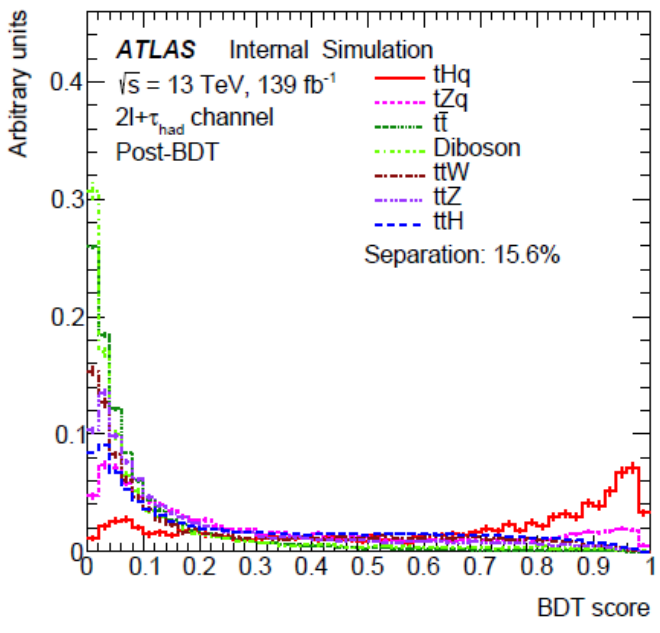
Dileptau BDT

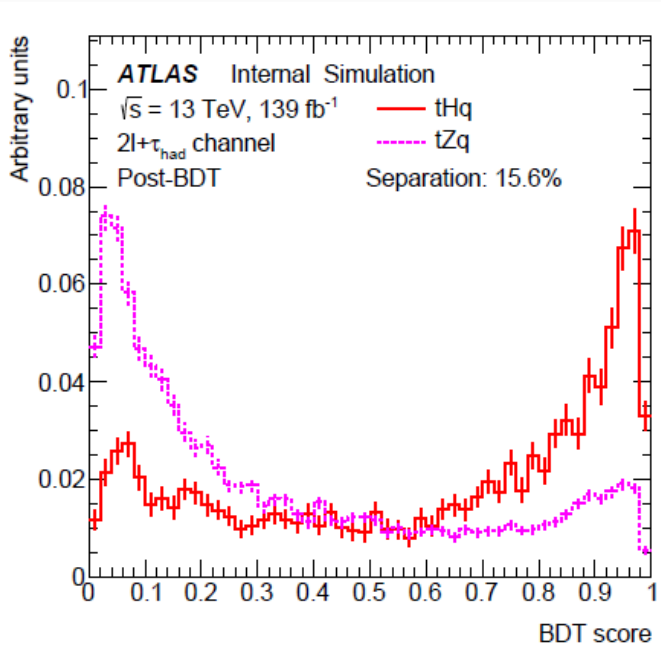
Region	tHq	Total	S/B (%)	Significance
n(b-jets) ≥ 1	2.5	12565.9	0.020	0.022
BDT ≥ 0.10	2.2	2760.9	0.080	0.042
BDT ≥ 0.20	2.0	1412.5	0.142	0.053
BDT ≥ 0.40	1.7	563.1	0.303	0.072
BDT ≥ 0.50	1.6	380.7	0.422	0.082
BDT ≥ 0.60	1.5	245.9	0.614	0.096
BDT ≥ 0.65	1.4	196.1	0.719	0.100
BDT ≥ 0.70	1.3	152.3	0.861	0.106
BDT ≥ 0.75	1.2	117.3	1.034	0.111
BDT ≥ 0.80	1.0	82.9	1.221	0.110
BDT ≥ 0.85	0.9	51.8	1.768	0.126
BDT ≥ 0.87	0.8	40.5	2.015	0.127
BDT ≥ 0.90	0.7	26.6	2.703	0.137
BDT ≥ 0.92	0.6	16.8	3.704	0.148
BDT ≥ 0.95	0.3	6.7	4.688	0.118
BDT ≥ 0.97	0.2	2.1	10.526	0.143

Yields, S/B and significance depending on the BDT cut

Dileptau BDT

- BDT not using categorical approach
- tZq separation is still quite good





Fake estimation dileptau

Analysis goal: Estimate Fake τ abundance in lep-had SR and correct MC to data ratio

1-Bin Method:

- likelihood fit of one component
- jets faking taus
- normalization from data

Quark/Gluon Method:

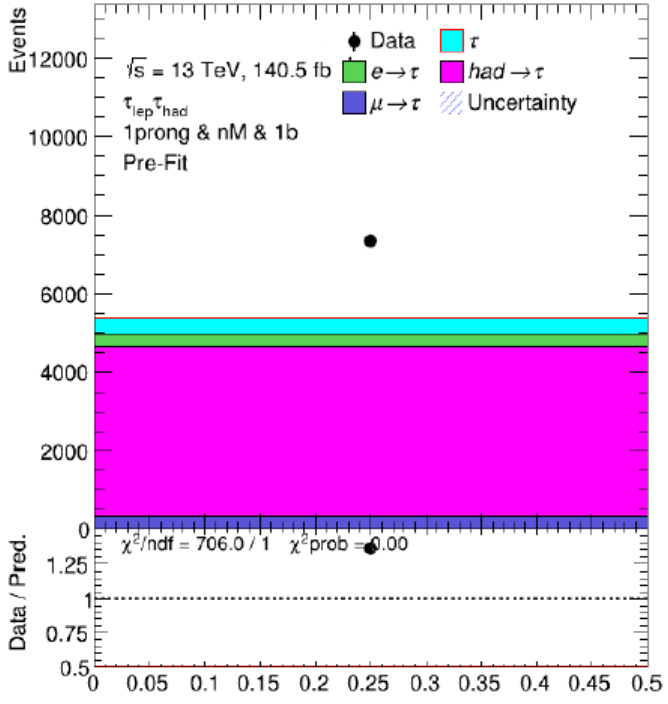
- likelihood fit of two components
- quark and gluon jets faking taus
- normalization from data and shape from MC

- normalization of templates for truth τ, electron faking τ and μ faking τ are fixed
- fixed sources treated as nuisances in the likelihood fit
- determine the dependence on p_T, $|\eta|$ and prongs

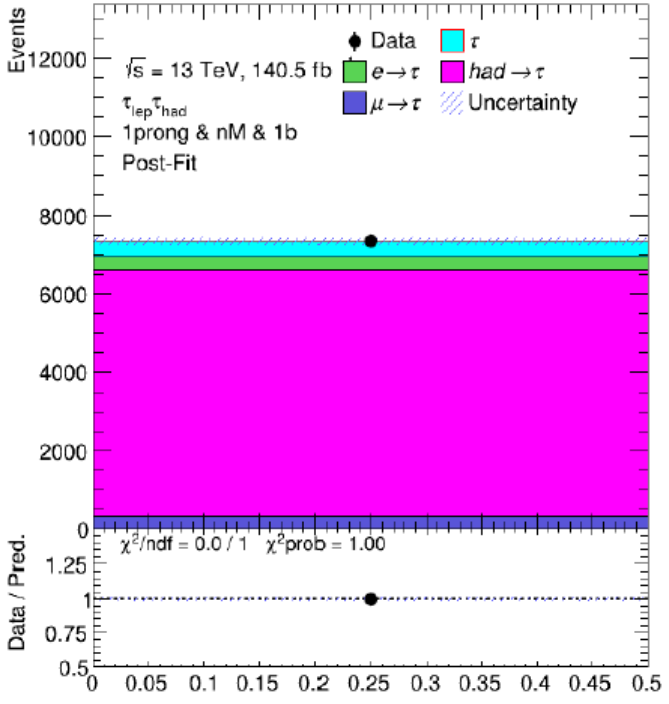
One bin method dileptau Control Region

- fit only jet faking tau SF → analogous to solving linear eq. for α_{had}:

$$N_{\text{data}} = N_{\text{el}} + N_{\mu} + N_{\tau} + \alpha_{\text{had}} N_{\text{had}}$$

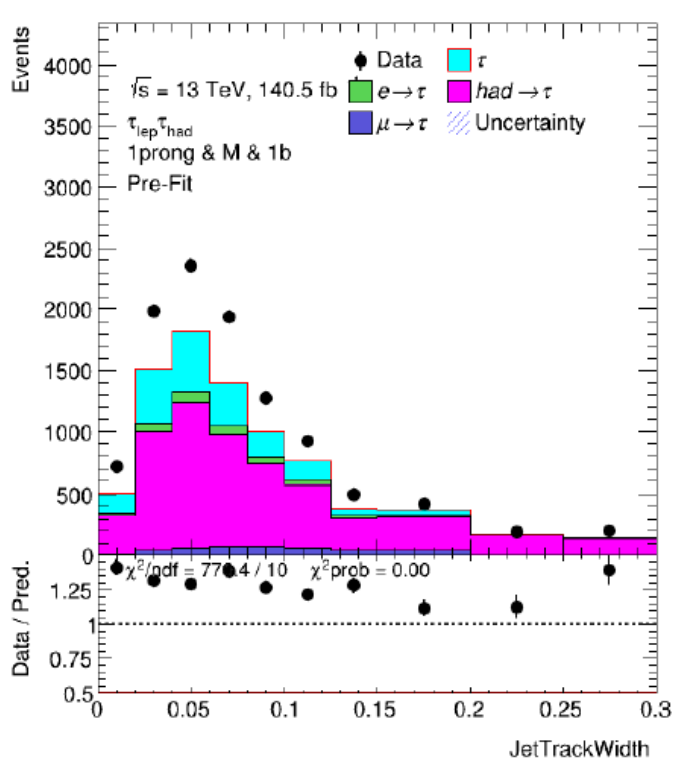


Pre-Fit: for $p_T \in [20, 30)$ GeV, 1-prong and τ on loose

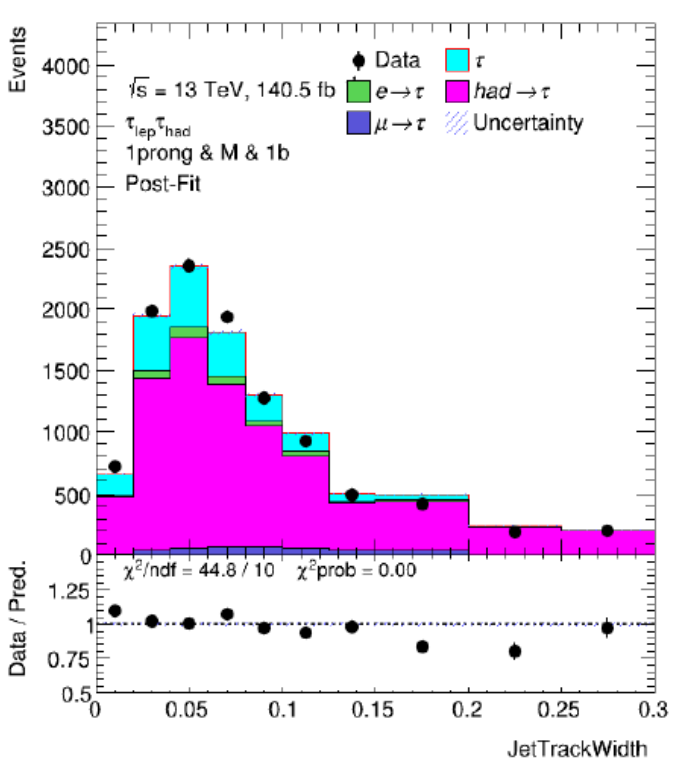


Post-Fit: for $p_T \in [20, 30)$ GeV, 1-prong and τ on loose

One bin method dileptau Signal Region

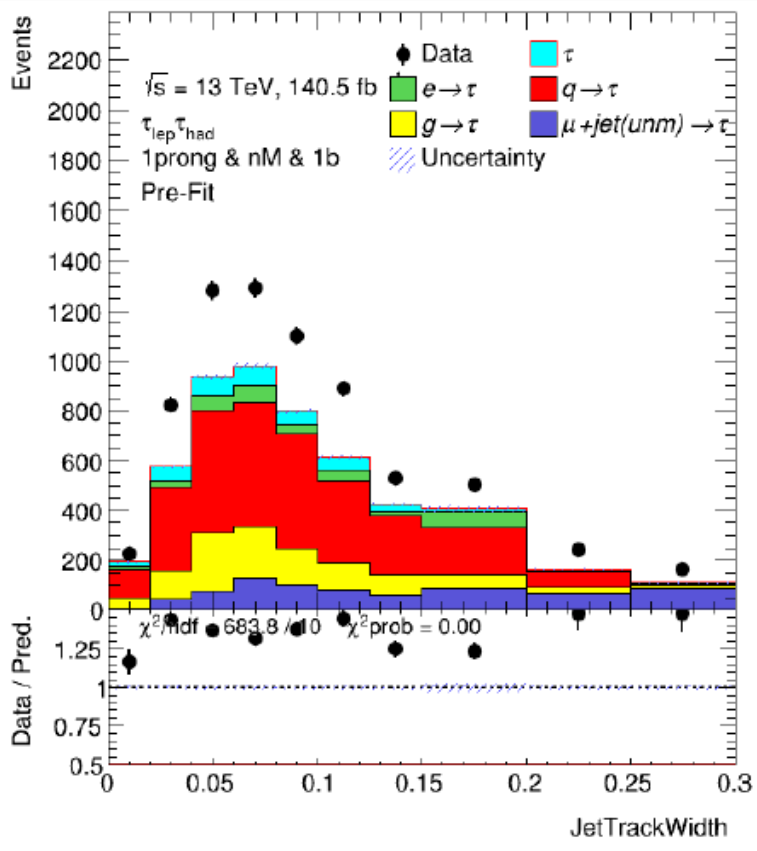


Pre-Fit: for $p_T \in [20, 30)$ GeV, 1-prong and τ on medium

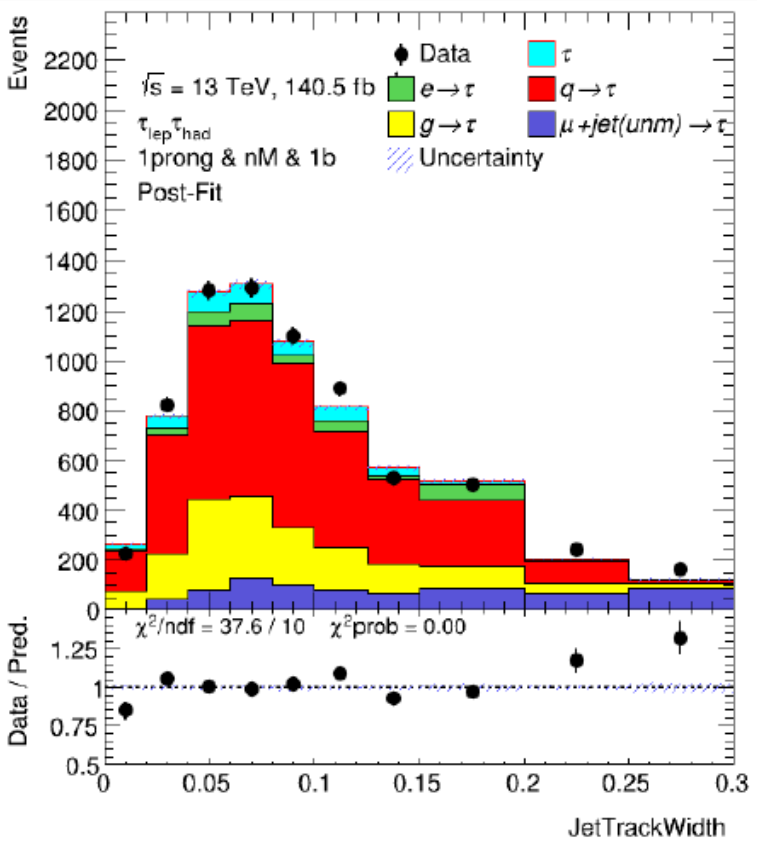


Post-Fit: for $p_T \in [20, 30)$ GeV, 1-prong and τ on medium

Quark/gluon method dileptau Control Region

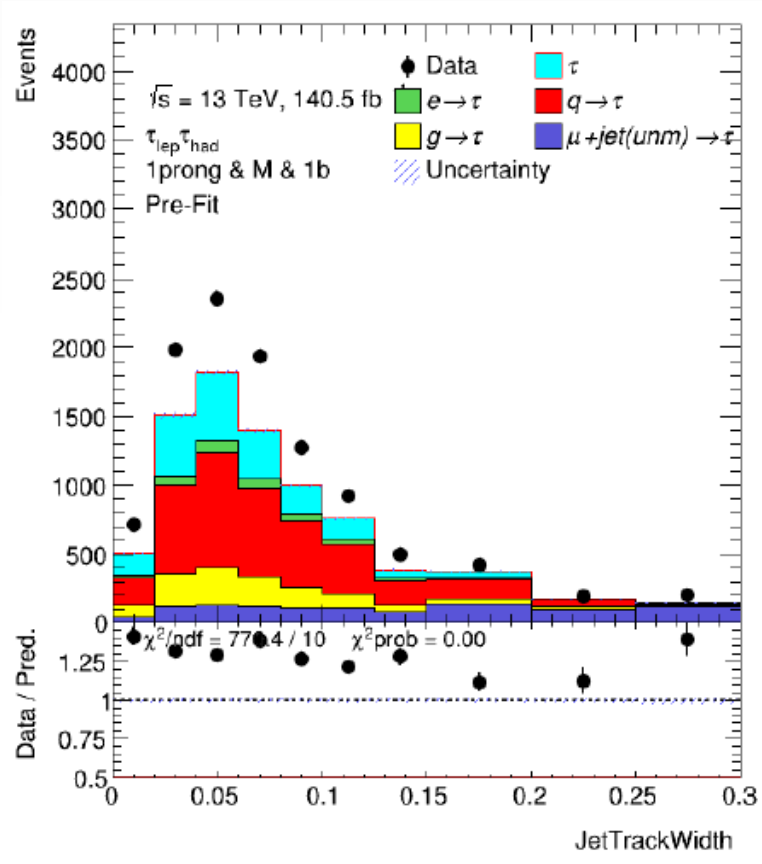


Pre-Fit: for $p_T \in [20, 30)$ GeV, 1-prong and τ on loose

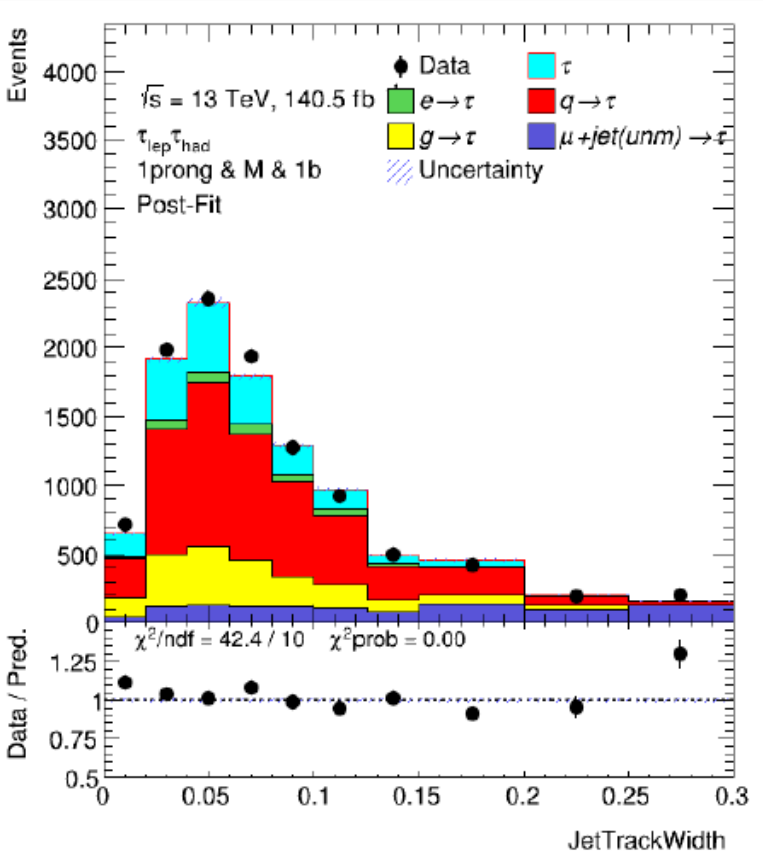


Post-Fit: for $p_T \in [20, 30)$ GeV, 1-prong and τ on loose

Quark/gluon method dileptau Signal Region



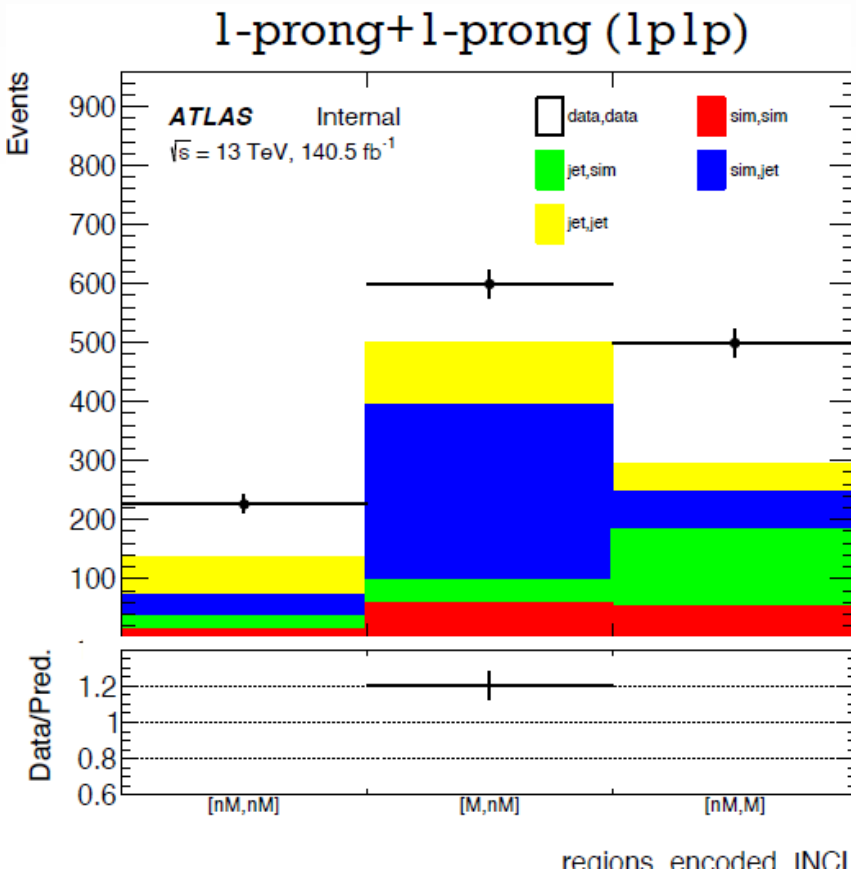
Pre-Fit: for $p_T \in [20, 30)$ GeV, 1-prong and τ on medium

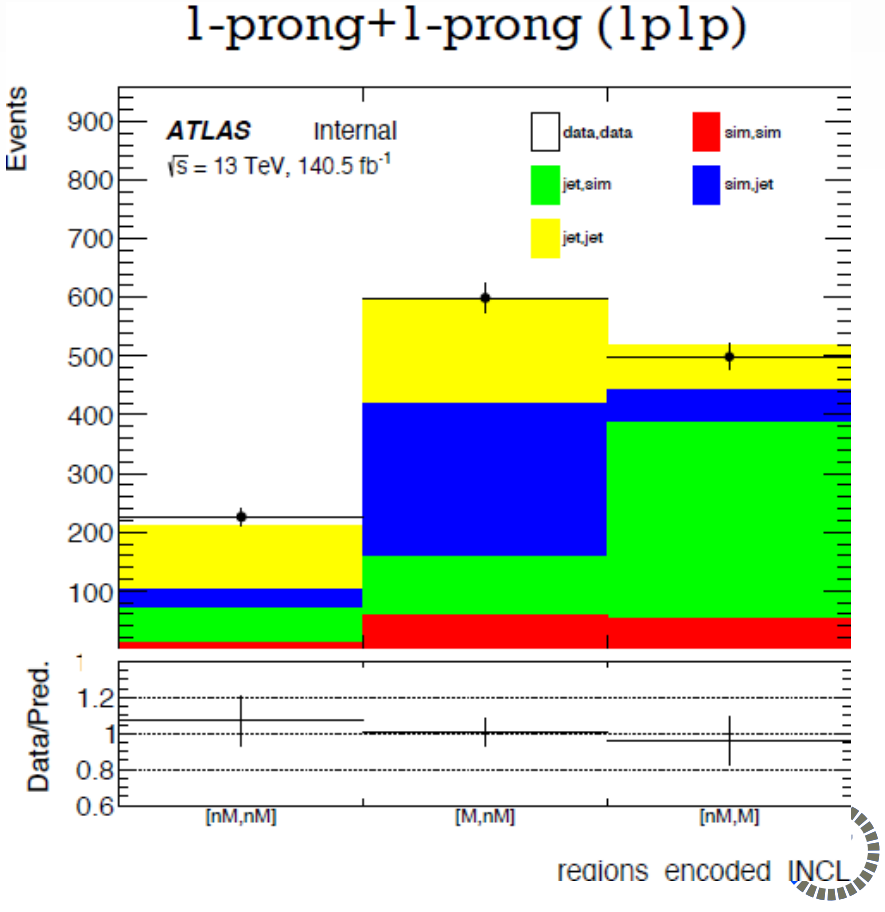


Post-Fit: for $p_T \in [20, 30)$ GeV, 1-prong and τ on medium

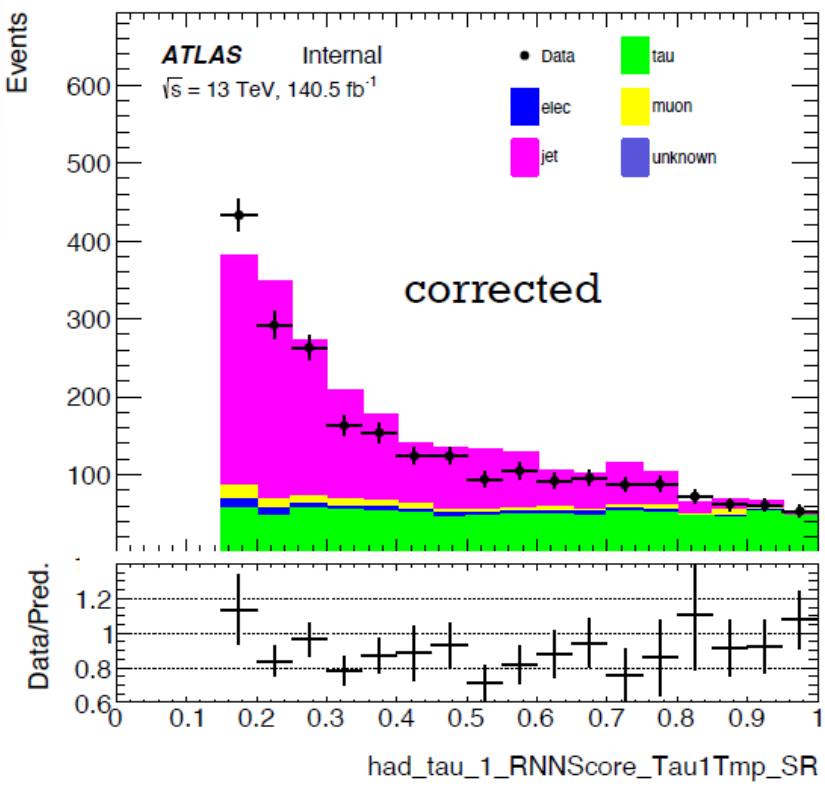
Fake Estimation Lepditau

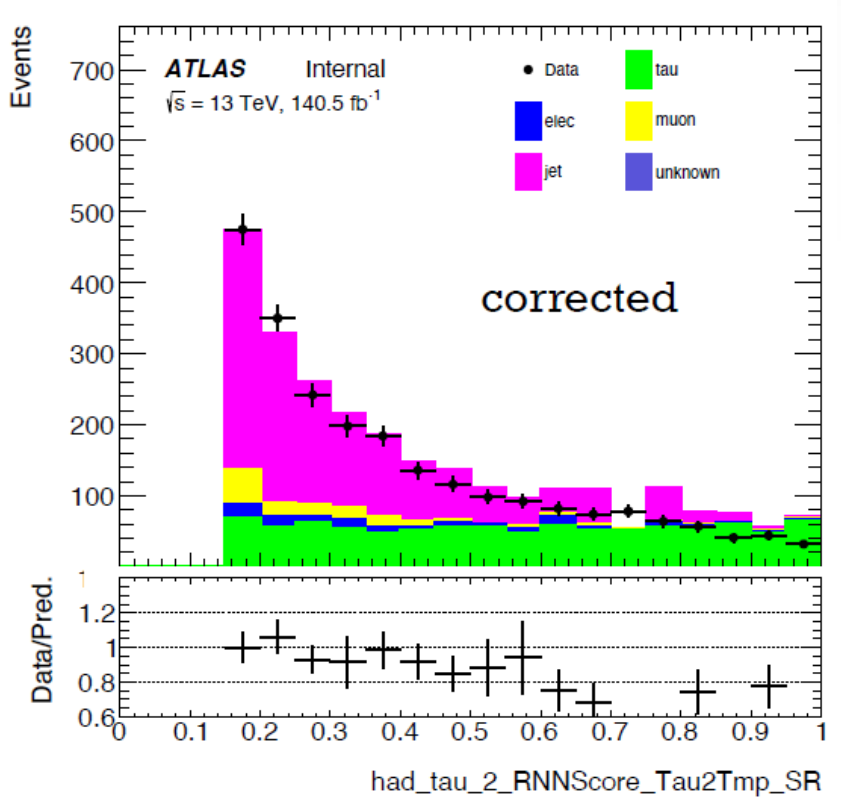
- Using template fit method



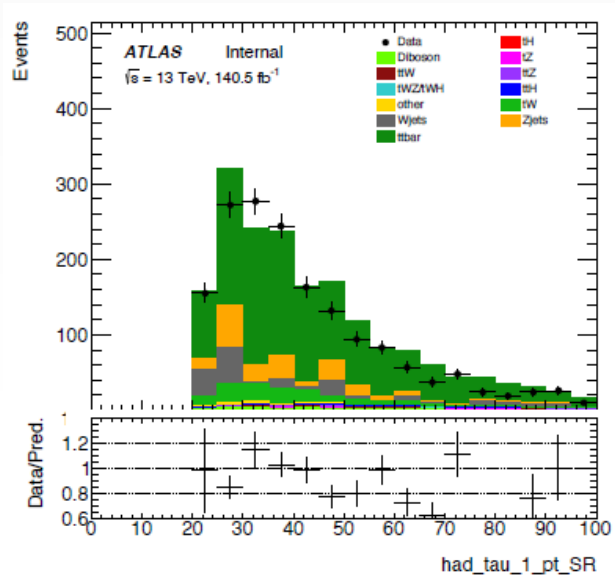


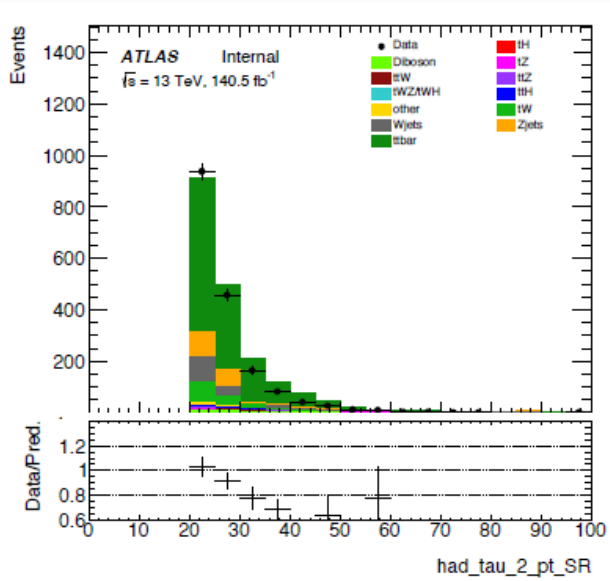
Lepditau fake estimation fit results

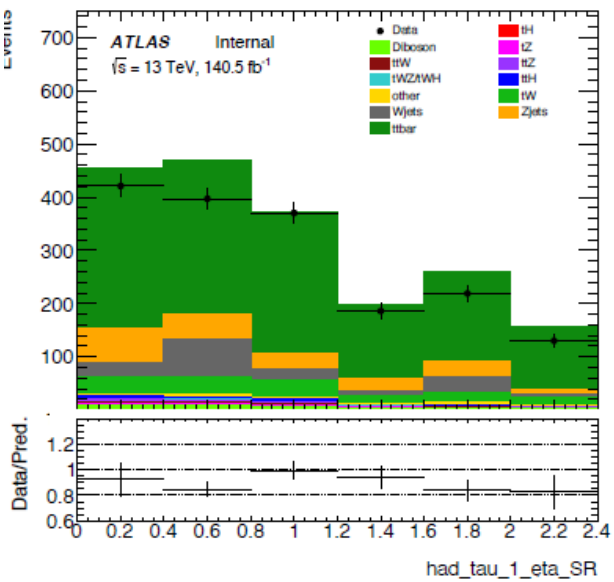


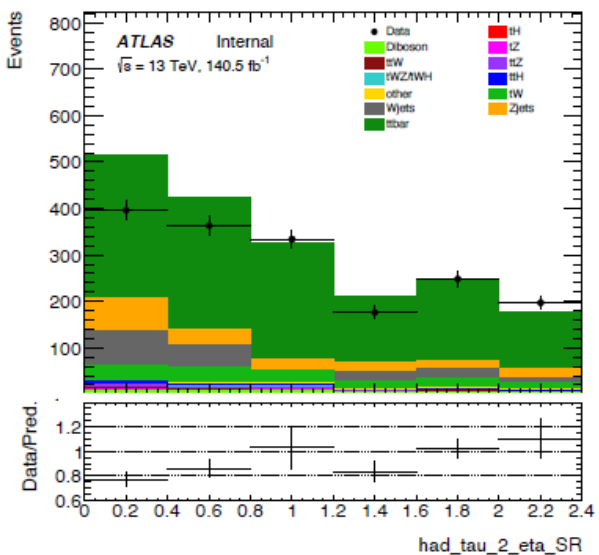


Lepditau fake estimation fit results





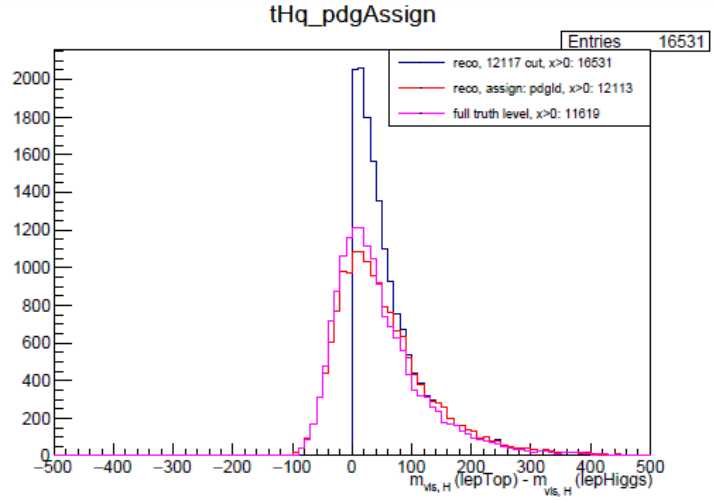


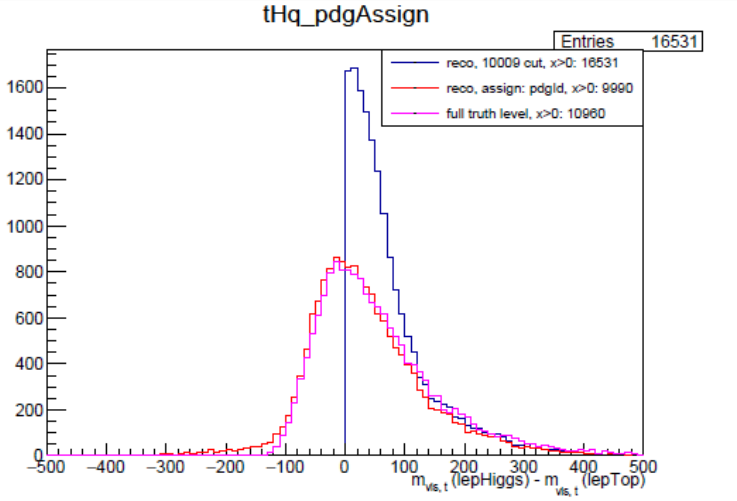


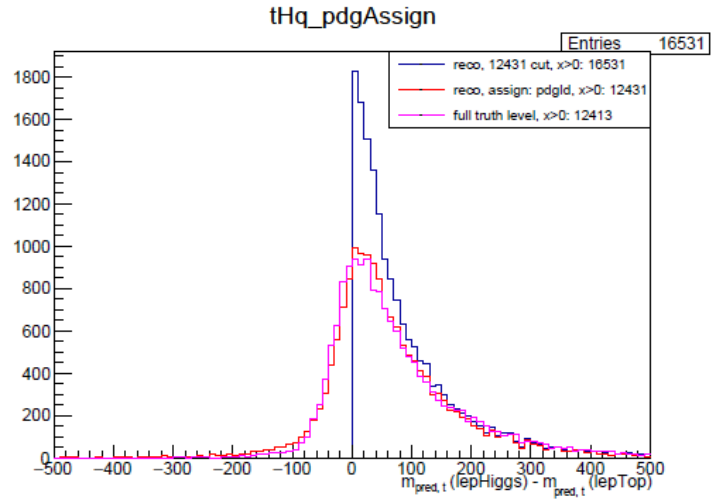
Lepton Assigment Method

- Establish method for lepton assignment
- Tested promising variables for assignment
- $m_{pred,t}(lep(Higgs)) - m_{pred,t}(lep(top)) > 0$
- Generality allows to expand to SS events

Lepton Assigment Results





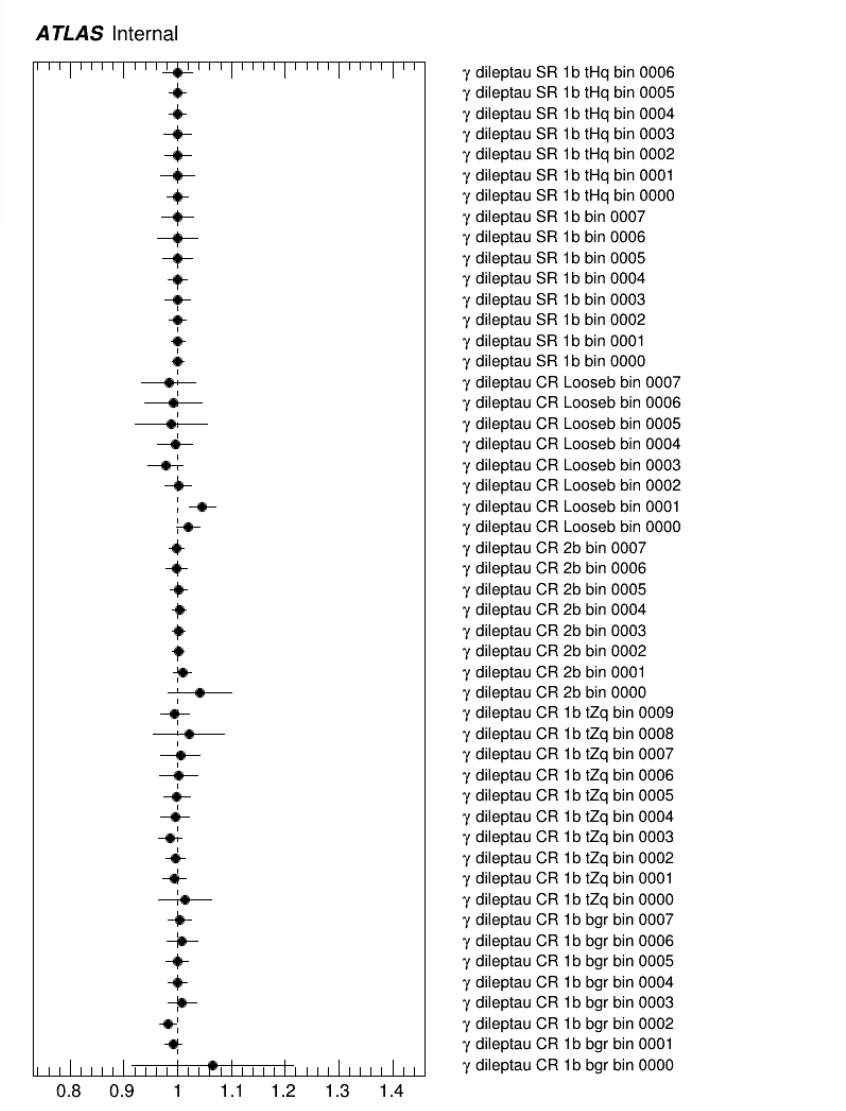


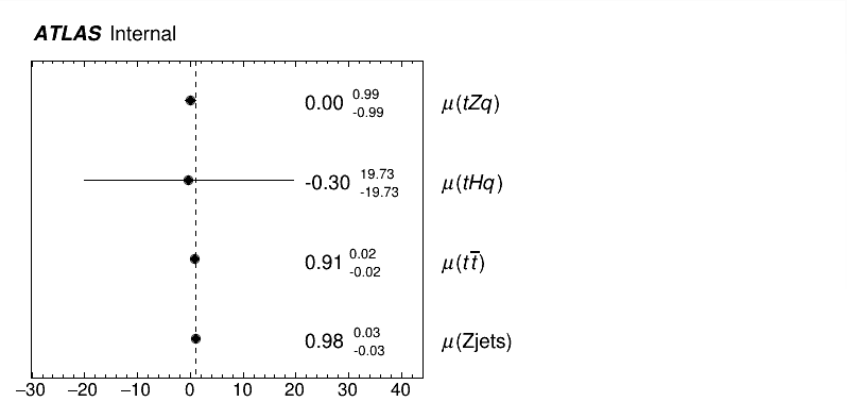
OSSF+OSDF+SSDF="pdgAssign"

assignment correct:

- $m_{\mathrm{pred},t}$: 75.2 %
- $m_{\mathrm{vis},H}$: 73.3 %
- $m_{\mathrm{vis},t}$: 60.5 %

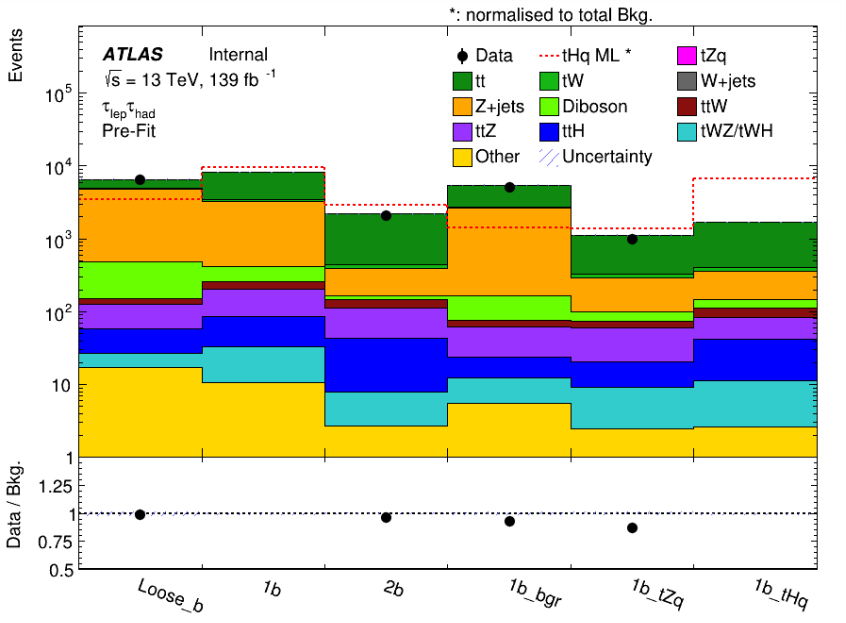
First fit results dileptau

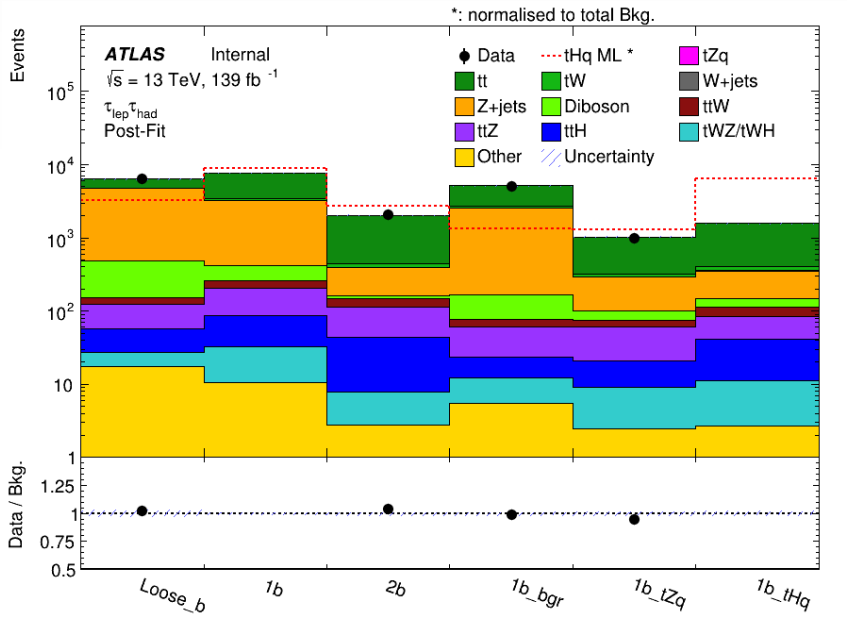




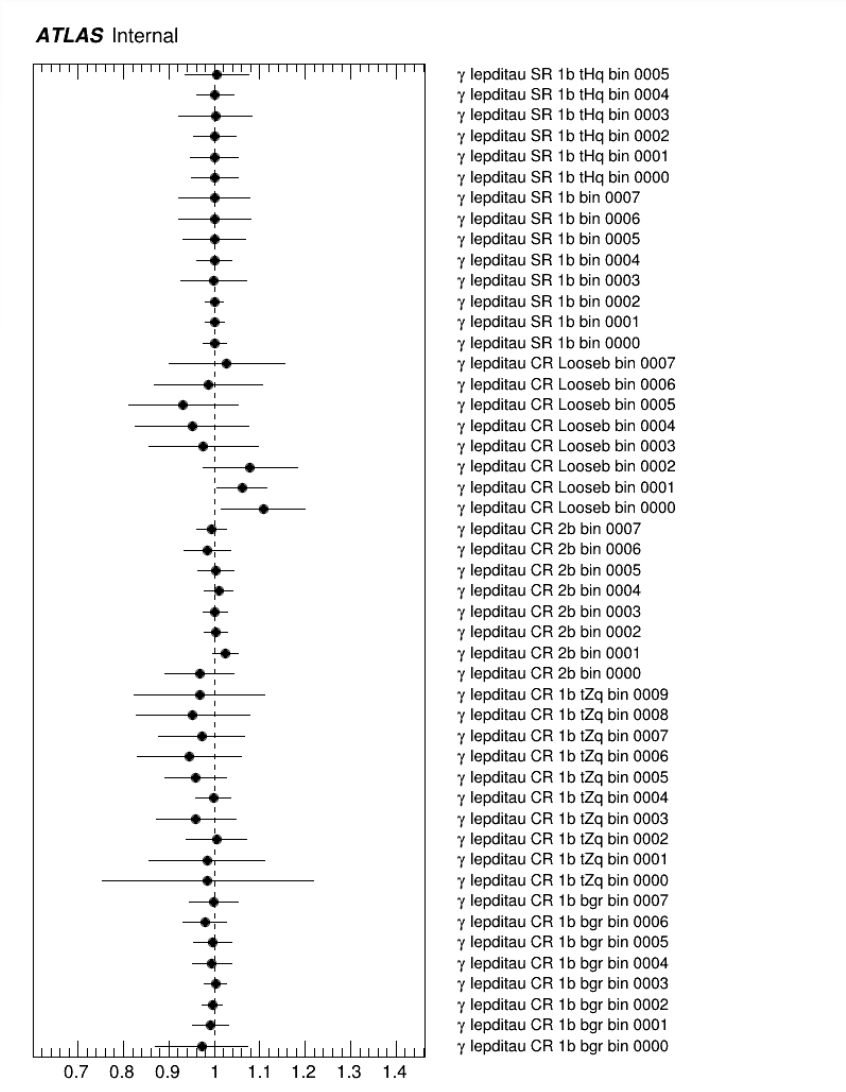
- Technically fitting is possible
- Much room for improvement left
- Separation for V+jets and $t\bar{t}$ already good
- Higher sensitivity expected from SS event inclusion

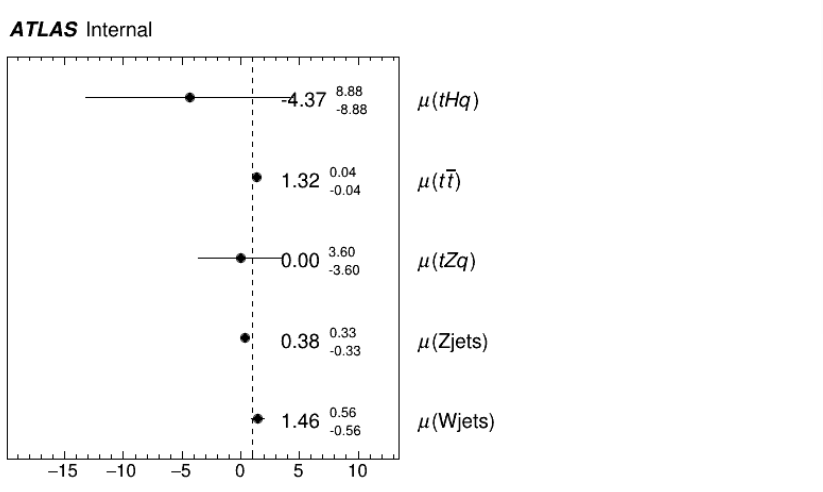
First fit results dileptau





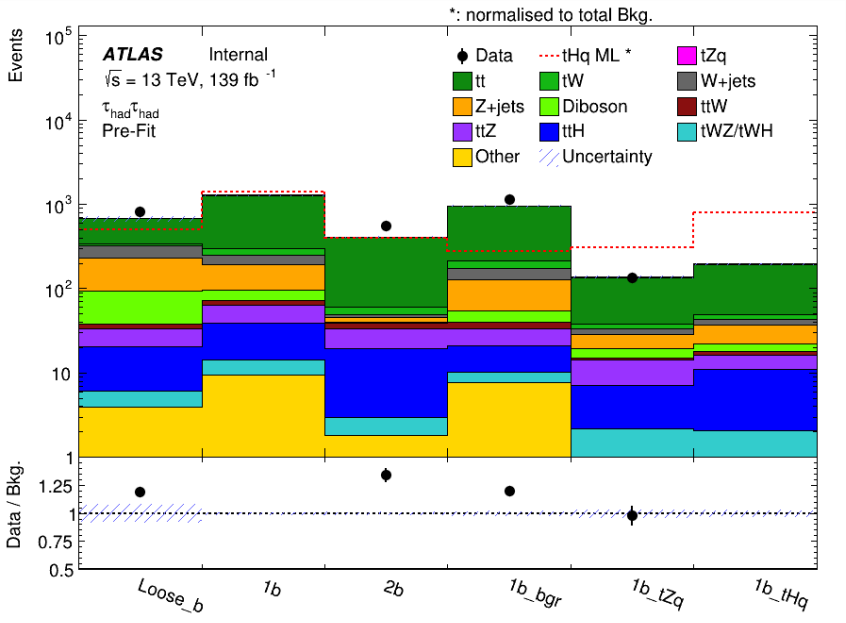
First fit results lepditau

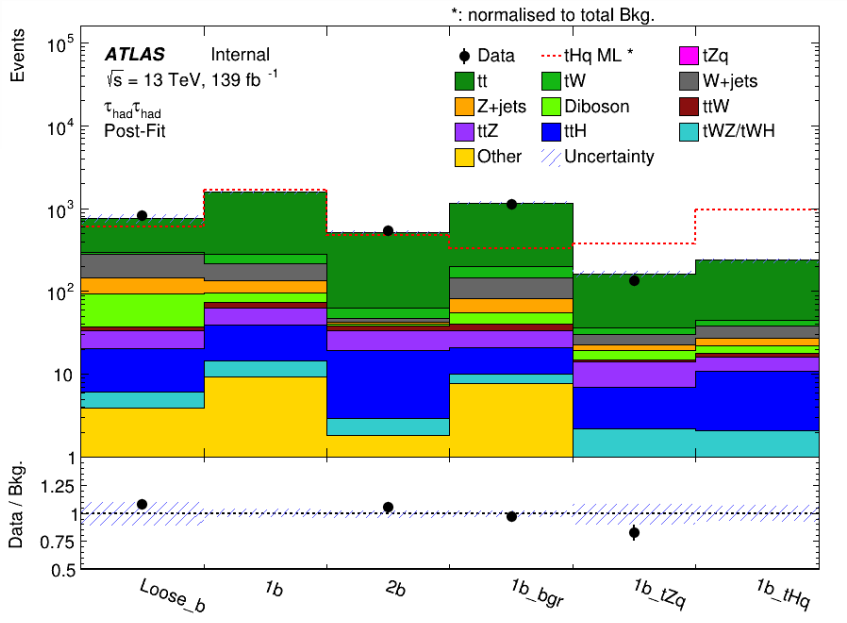




- Technically fitting is possible
- Much room for improvement left
- Separation for V+jets and $t\bar{t}$ already good
- Improvement in fake tau scale factors to be expected

First fit results lepditau





Status of the Int Note

- https://gitlab.cern.ch/atlas-physics-office/HIGG/ANA-HIGG-2020-02/ANA-HIGG-2020-02-INT1
- Majority is finished
- For all the missing parts authors have been assigned and have accepted
- A first version is planned for next week

More information

- Checks on v34
- Lepditau Fakes
- Dileptau Fakes
- Categorical MVA
- Dileptau BDT, feature importance
- Lepton assignment