1 Move to ℓb classifier

1.1 Motivation

Move due to poor performance on event level BDT (was using 50% TRF and 50% testing/training). ST, minMax, HT and a few other variables performed better than the BDT → most likely due to the fact that the BDT score variable had half the MC than just taking the full MC dataset as normal.

1.2 Different Iterations/Setups

Common setup:

- use $t\bar{t}$ Bakoven sample (Ntuple-level selection of 1 lepton)
- Signal: ℓb from the same top
- Signal: ℓb **NOT** from the same top

1.2.1 Matching ℓ AND b to truth

Match ℓ and b to their truth particles, by use of smallest ΔR. Small ΔR is indication of a good match to truth.

- Signal: ℓ and b with the smallest ΔR (below some threshold → 0.3) and come from the same top
- Background: ℓ and b with the smallest ΔR (below some threshold → 0.3) and DON'T come from the same top

Note: threshold chosen of 0.3 is VERY loose

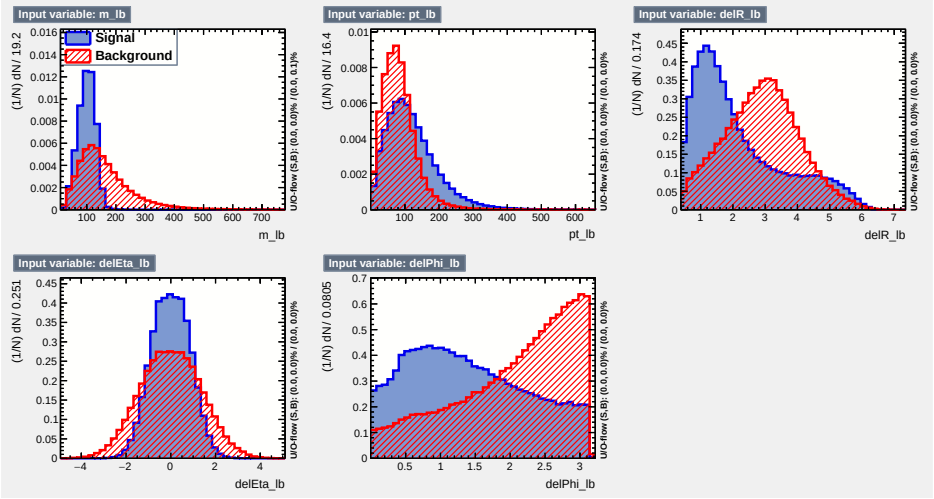
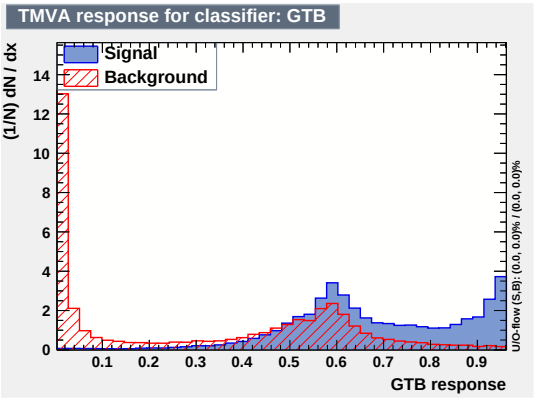


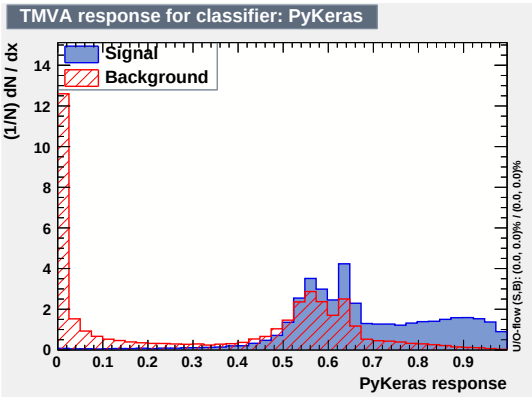
Figure 1: Input variables

- Num events[1] training ∼ 240 000 (sig and back)
- Num events testing ∼ 60 000 (sig and back)

⋆ Check non-linear correlation coefficient graphs (In TMVAGui)

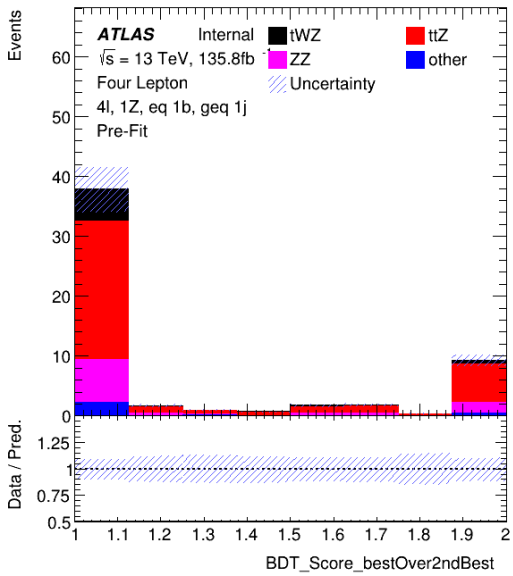
[0]Date: Early August 2020
[1]James says this is sufficient and any more would just slow down training

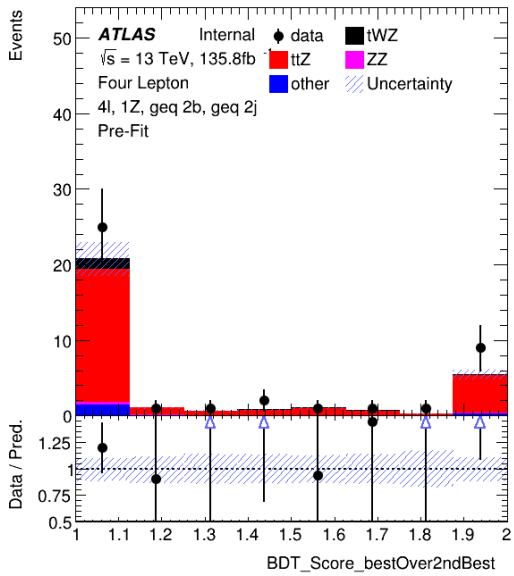




⋆ Weird bump?

Results:





Limit → 2.429 (All ASIMOV), 2.488 (Data in CR)

Previously (Full Run 2 $m(\ell b)_{minMax}$) → 2.474 (All ASIMOV), 2.535 (Data in CR)

** Figured out that the '-1' default value is being used in the fit (underflow), as can see in first bin **

1.2.2 Matching ℓb system to truth

2 Delft Ntuples

2.1 Yield Discrepancy

Supposed difference between Delft and Constantia which has the largest effect → using a tighter lepton isolation parameter

This causes much less stats/yields in Delft (compared to Constantia) → also seen in Cameron's analysis. I compared 2018 Constantia and Delft 2018 (ttZ region), 1 data event in Constantia and 0 in Delft. This was concerning, so looked at INT note to see stats/yields Full Run 2 Constantia had. INT note had around 40 data events (would expect 2018 Constantia to have more than 1, probably around 16 (since luminosity of mc16e is a bit less than half of Full Run 2 luminosity)). The only differences between INT note plot and 2018 Constantia control plot are that

we are using Full Run 2 and the 77% btag WP (referring to INT note plots).

Recall:

- Tight jet: 77 WP or less
- Loose jet: 85 WP

- Delft 2018: 0 ttZ data events ($\geq$ 1 tight, $\geq$ 1 loose, loose + tight = 2)
- Constantia 2018: 1 ttZ data event ($\geq$ 1 tight, $\geq$ 1 loose, loose + tight = 2)
- Constantia in INT Note (Full Run 2): $\sim ttZ$ 40 data events ($\geq$ 1 tight, $\geq$ 0 loose, loose + tight = 2)

Changed Constantia 2018 ttZ region to $\geq$ 1 tight, $\geq$ 0 loose, loose + tight = 2 (essentially what is in INT note (these plots are before this bjet WP was implemented)). This gave us the expected total number of events, $\approx$ 17.