

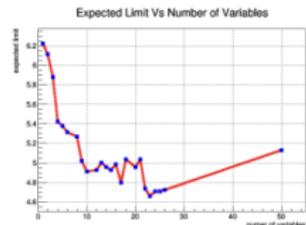
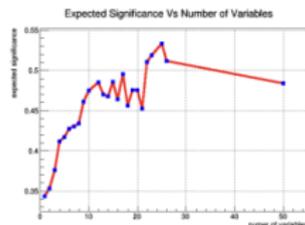
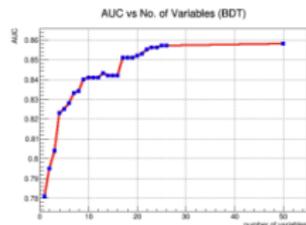
Status of 4Top analysis

Huiling Hua¹, Fabio lemmi¹, Ducan Leggat², Hongbo Liao¹, Hideki Okawa², Yu Zhang²

1,IHEP, 2,Fudan University

April 20, 2022

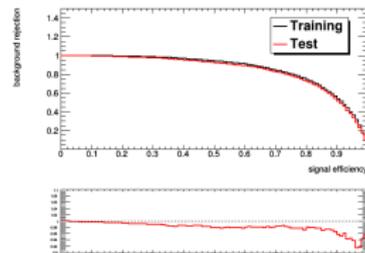
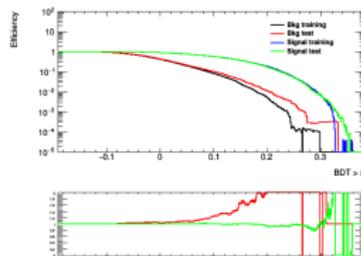
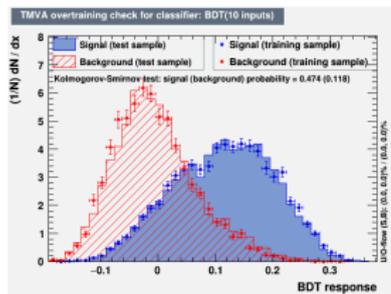
Review Huiling's significance curve



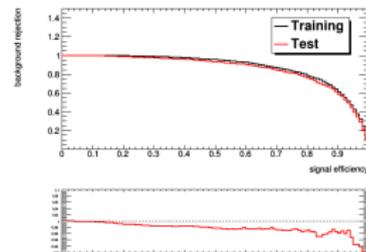
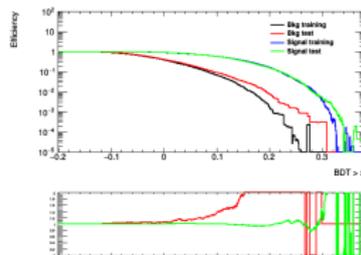
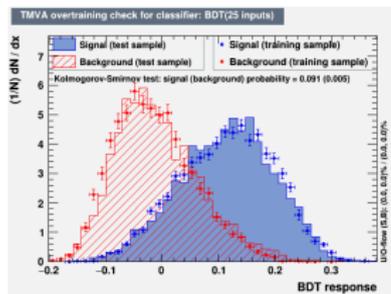
We need to understand:

- Why the significance jumps from 11 to 20 variables
- Whether the peak at 25 variables is reliable
- How we choose the optimal binning
- Statistics
 - Signal : 95703
 - Background : 37267

Overtraining check — 10 variables



Overtraining check — 25 variables



Cut-and-Count significance

- The histograms are 11 bins from -0.8 to 0.8 and almost all the background and signal are in bin 5 to 8 (-0.21 to 0.36).
- Calculate the bin-by-bin significance and error since the "Combine" tool does not provide error of significance.

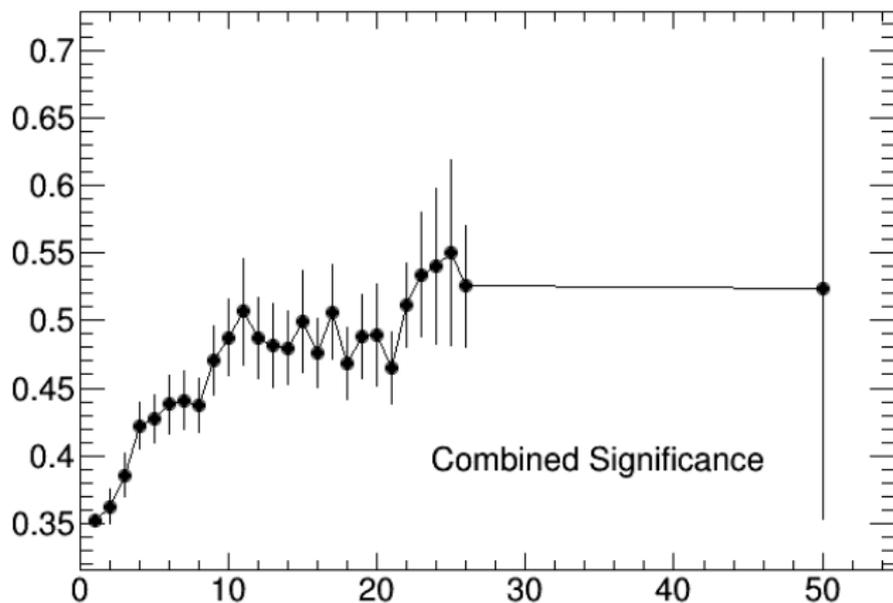
$$\sigma = \sqrt{2((s + b)\log(1 + s/b) - s)}$$

$$\Delta\sigma = \sqrt{(A\delta s)^2 + (B\delta b)^2}$$

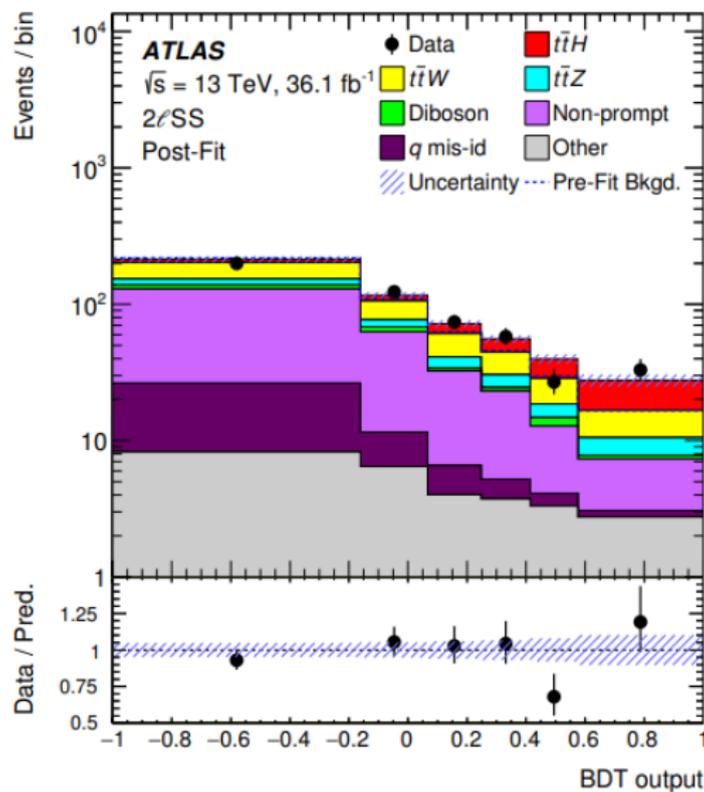
$$A = \partial\sigma/\partial s = \log(1 + s/b)/\sigma$$

$$B = \partial\sigma/\partial b = (\log(1 + s/b) - s/b)/\sigma$$

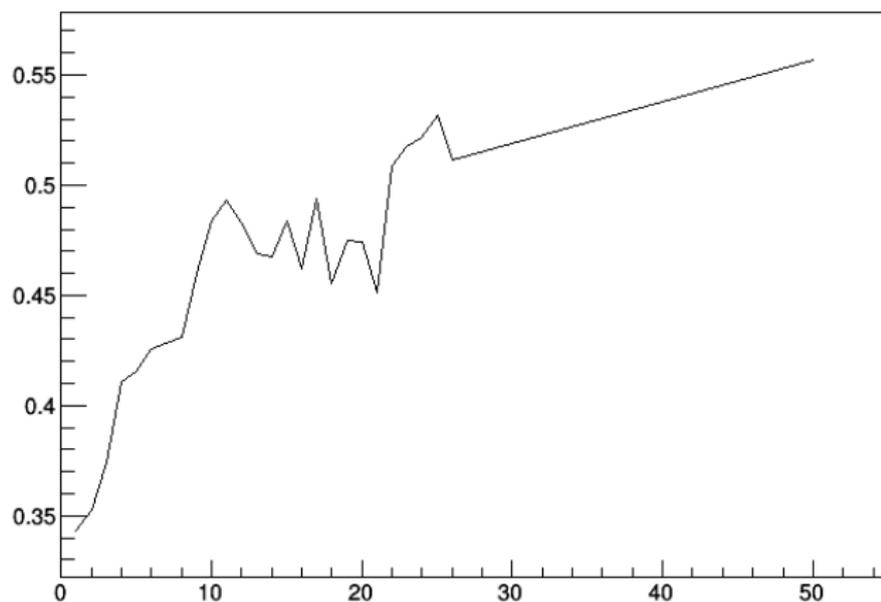
Cut-and-Count significance



Binning of ATLAS ttH analysis



Significance vs number of variables with 6 bins



Next step

- Find optimal binning of each set of variables by scanning the bin edge.
- Check the significance vs number of variables to determine the number of variables.