

Optimization of BDT binning

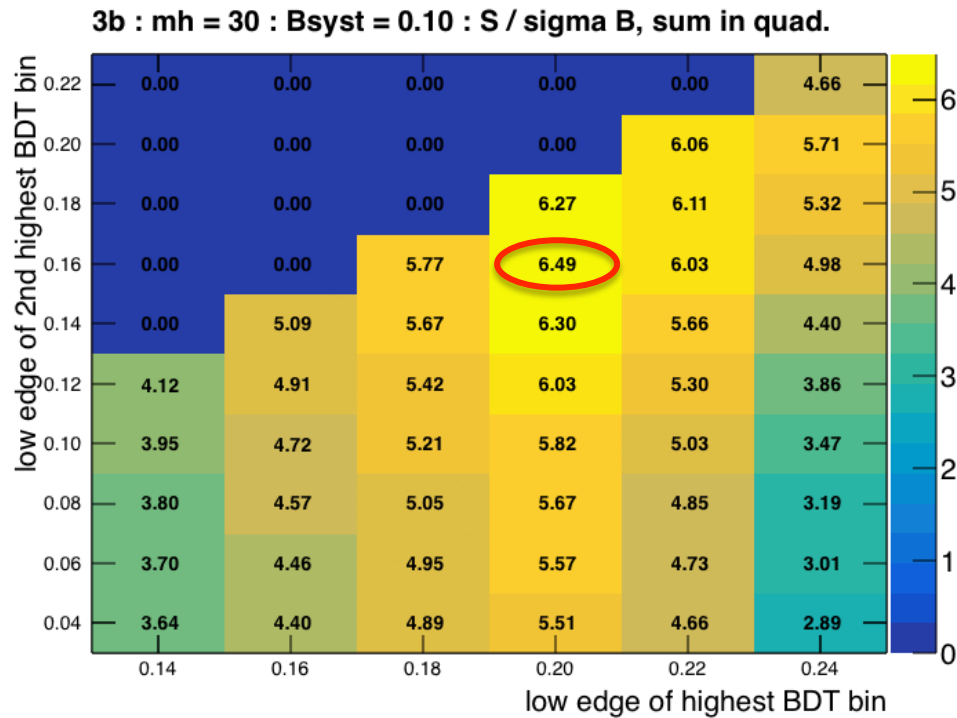
Owen

Oct. 25, 2018

Optimization study description

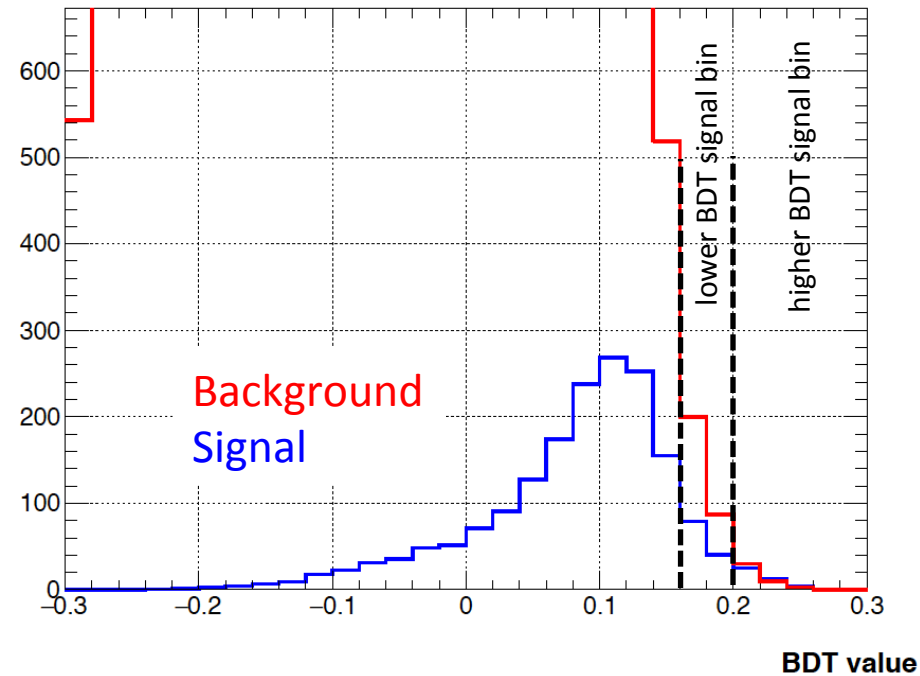
- Using S / σ_B as metric for finding optimal binning of BDT.
 - similar to S / \sqrt{B} or $S / \sqrt{S+B}$ except using σ_B including a fractional "systematic" uncertainty that's in addition to statistical uncertainty.
 - $\sigma_B = \sqrt{B + (B \cdot \text{syst})^2}$, where syst is a fractional uncertainty. That is, syst = 0.10 for a 10% BG uncertainty that's in addition to the statistical uncertainty from the expected number of BG events.
 - S / σ_B calculated in each BDT bin. To combine them, add the values in quadrature: $S / \sigma_B(\text{total}) = \sqrt{S / \sigma_B(\text{bin1})^2 + S / \sigma_B(\text{bin2})^2}$
 - Should be close to the significance of the signal if it exists at the predicted cross section (and BF's set to 1).
- Optimizing the highest two BDT bins for signal sensitivity.
- Assuming that 3 to 5 lower background-dominated BDT bins will also be included so that the BG scale factors can be fit, but won't contribute to signal sensitivity.
- Note: BG histograms that I have now only have $t\bar{t}$ and triboson. Missing QCD and other significant BGs.

Example, 3b

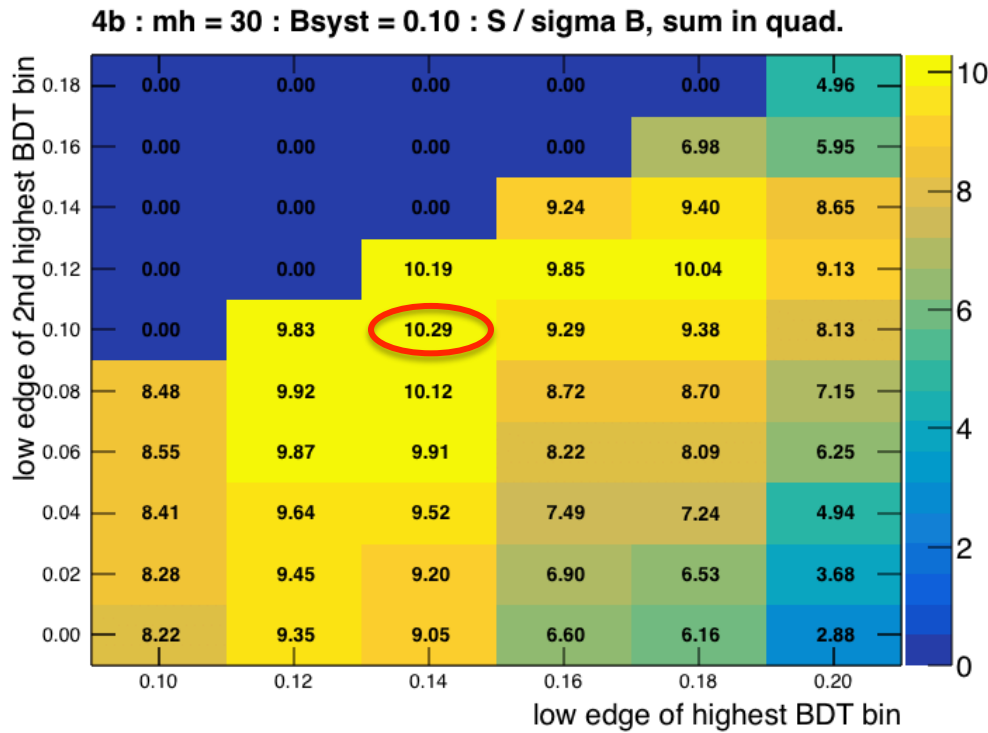


For a mass of 30 GeV, the binning for the 3b BDT that gives the highest S / sigmaB is shown, assuming a BG "systematic" uncertainty of 10%.

MVACutFlowTribMVA.root : signal mass 30

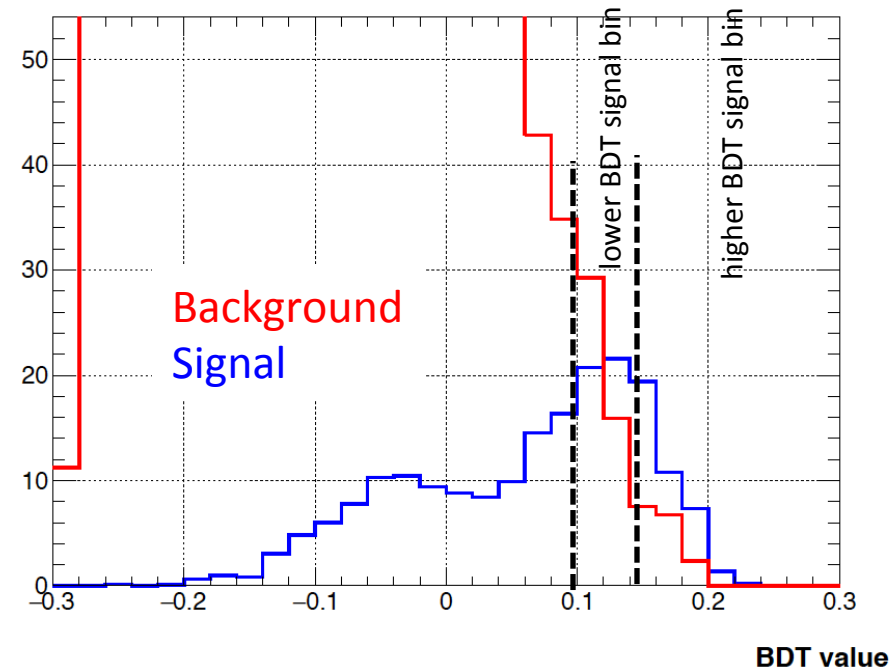


Example, 4b

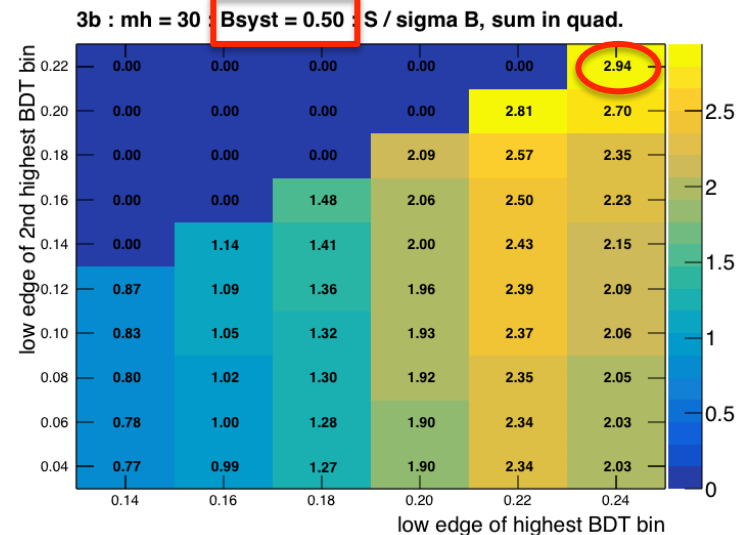
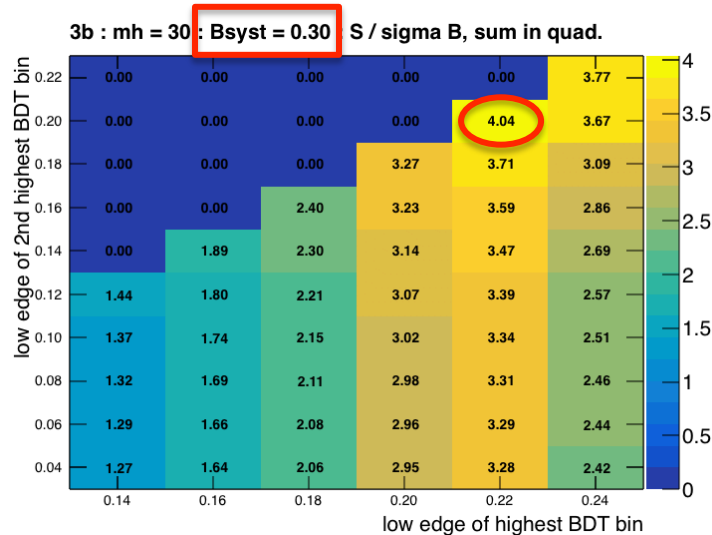
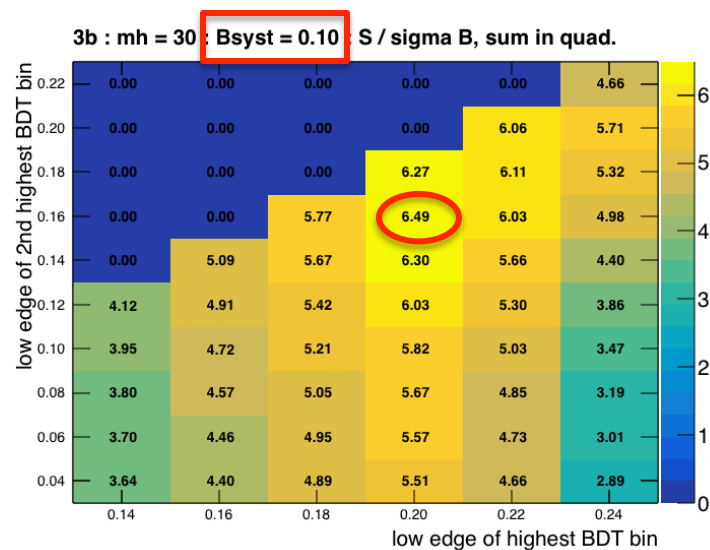
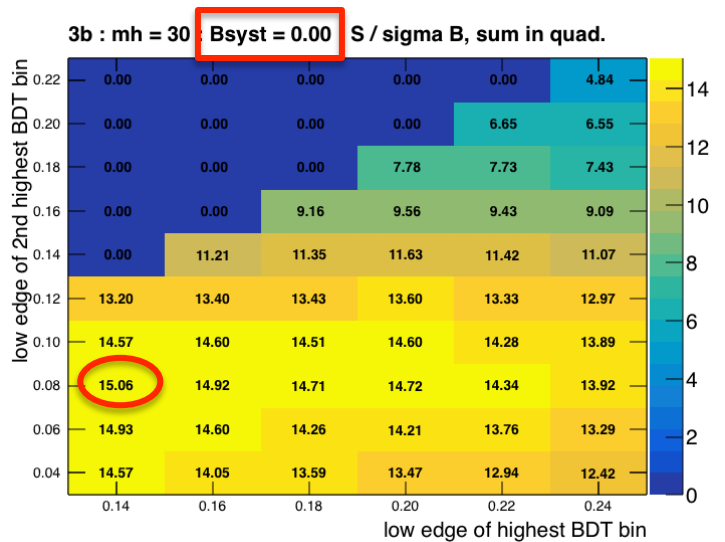


For a mass of 30 GeV, the binning for the 4b BDT that gives the highest S / sigmaB is shown, assuming a BG "systematic" uncertainty of 10%.

MVACutFlowQuabMVA.root : signal mass 30



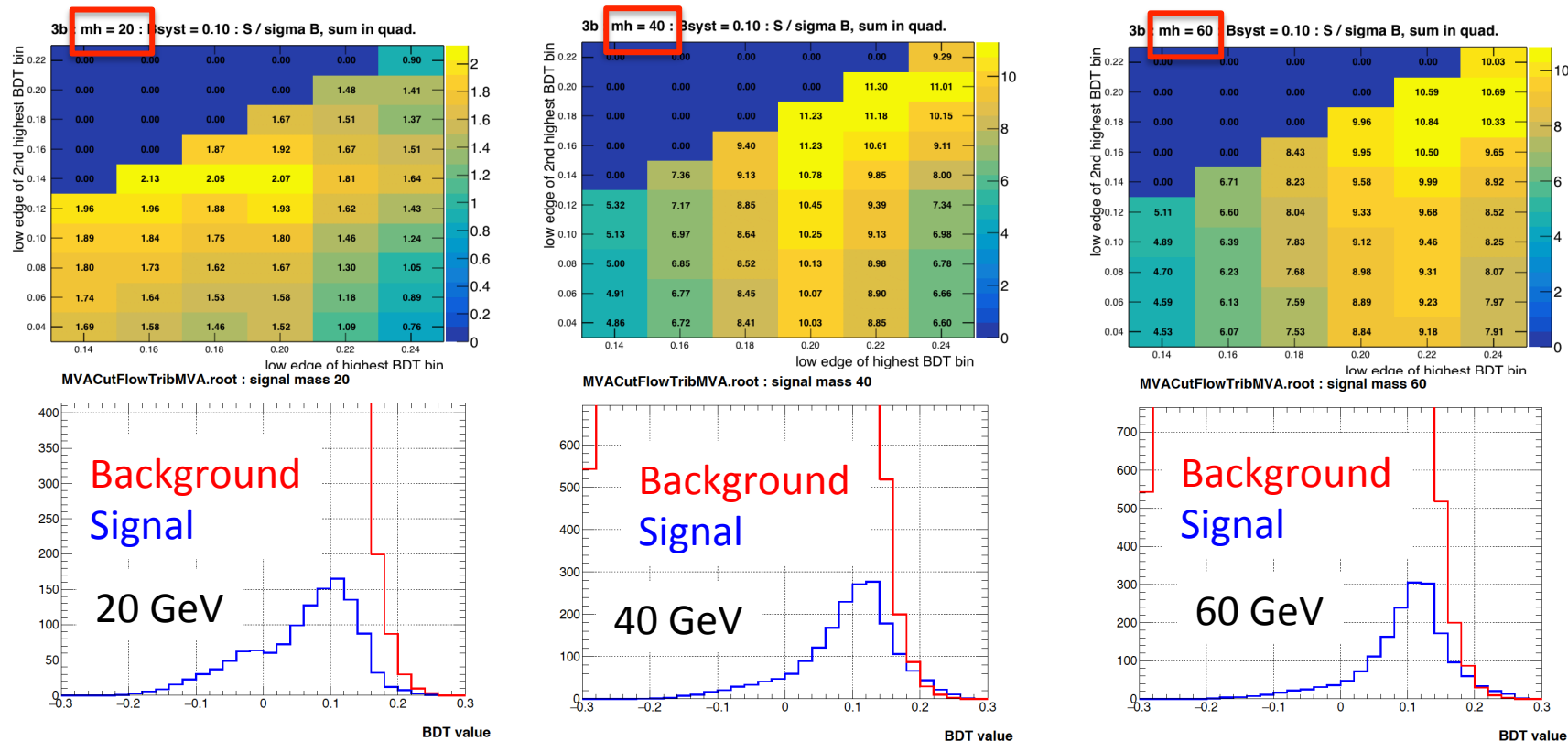
Dependence on assumed BG systematic



Tighter bins are more optimal when BG uncertainty is large.

Dependence on a mass

To have good sensitivity for low a masses, should have looser bins.



Some conclusions

- Simple optimization study like this can give guidance on how to choose BDT bins.
- Will need to redo it with all important backgrounds included.
- Results are fairly sensitive to the BG uncertainty.
- Some dependence on the signal mass.
- General conclusions
 - Good to have a tight bin where the signal fraction is at least 50%.
 - To cover different signal masses and BG uncertainties, could have 3 signal bins, plus around 4 or 5 background-dominated bins?
- Once we have it narrowed down, can test a handful of options with the full combine setup.