Background Estimation For High Energy Higgs Boson Measurements with the CMS Experiment

Keston Smith

Supervisor: Jennet Dickinson, Nhan Tran - Fermilab

SIST/GEM Final Presentation

30 July 2021

Outline

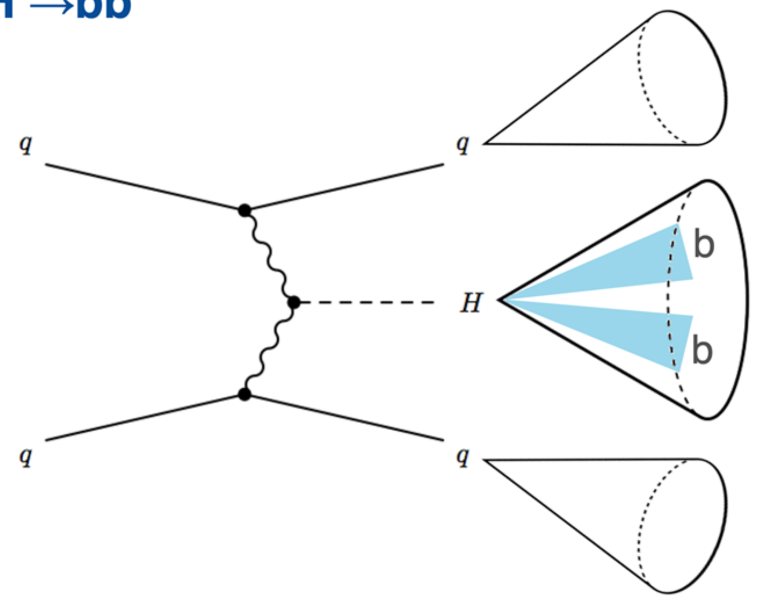
- Introduction/background
- Objective and Analysis
- Results

Large Hadron Collider (LHC)

Introduction

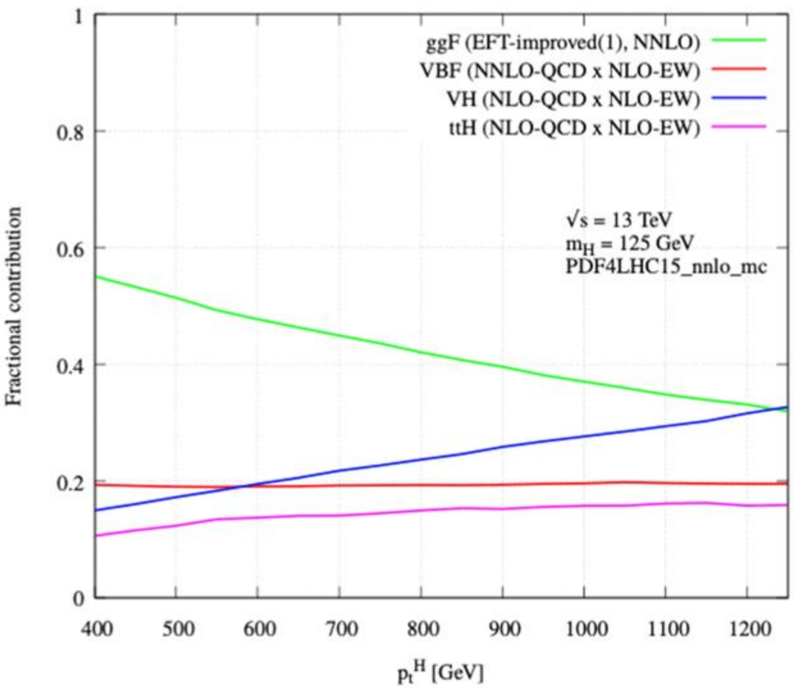
- There are **several ways a Higgs can be produced** in proton-proton collisions.
  - e.g. ggF, VBF, VH etc.
- We study Higgs bosons through vector boson fusion with high transverse momentum.
- We also look at its decay into bottom quarks.

VBF, H →bb



An example of a Higgs produced through vector boson fusion (VBF) and decaying into b-quark jets.

Why High Transverse Momentum?



Fraction contributions to Higgs production as a function of p_T. Gluon-gluon contributes less as p_T increase.

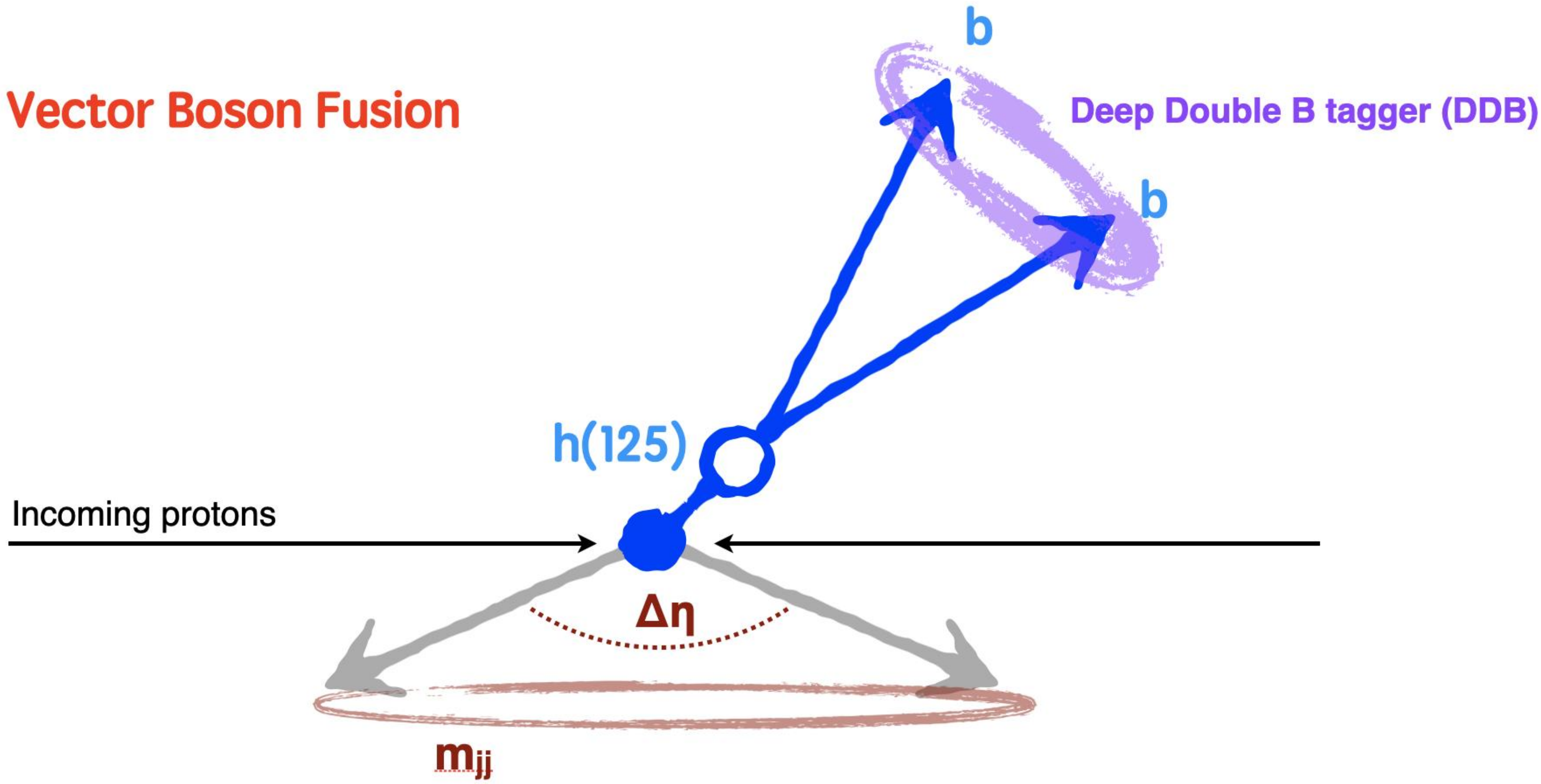
- At high p_T ranges, we observe that the ggF production becomes **less prominent.**
- VBF and VH production modes become of more importance.
- High p_T categories provide more sensitivity to interaction beyond the standard model.

Objective

- The focus of my project is to define a **control region** to estimate the top quark (ttbar) processes in the **signal region**. With the help of MC simulations, we can derive ttbar estimates in a control region to then compute a scale factor to renormalize the signal ttbar prediction.

- We then validate our prediction, by observing how much the simulation and the data agree.

Analysis Strategy Higgs Bottom quark decay Signal



Analysis Strategy Higgs Bottom quark decay Signal

- VBF like events include one large radius jets and two smaller jets.
  - Each jet candidate is required to have p_T >450 GeV, |eta| < 2.5.
  - DDB score >0.89 for deep double b tagger v1; DDB score >0.64 for deep double b tagger v2.
  - No leptons, no b-jets in the opposing hemisphere.
  - deta>3.5, m_{jj}>1000 GeV

We use Higgs jet criterion to define the signal region.

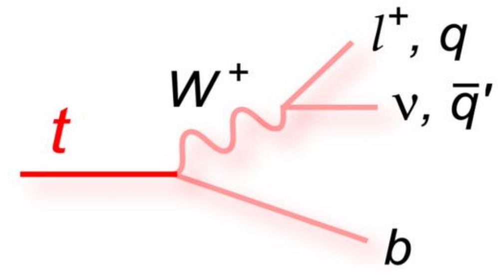
Muon Control Region

To extract a Higgs signal, we must first **understand the associated background**. Of this background we focus on the **ttbar process**.

- This segment of the analysis has the same requirements as the signal region, except we allow a “loose” muon.
- The scale factor is computed by taking the ratio between the CMS data and the simulation events.
- We then use the scale factor to renormalize the ttbar prediction in the signal region.

n_SRpred = n_SRMC * SF

SF = n_CRdata / n_CRMC



Feynman diagram of top quark decay

Statistics Report

- Using the coffea module, I was able to generate signal region event counts for various deta and m_{jj} cuts.
- The uncertainty in the scale factor is derived from the CMS data; this uncertainty then propagates to the signal region ttbar prediction.

$$\Delta N_D = \sqrt{N_D}$$

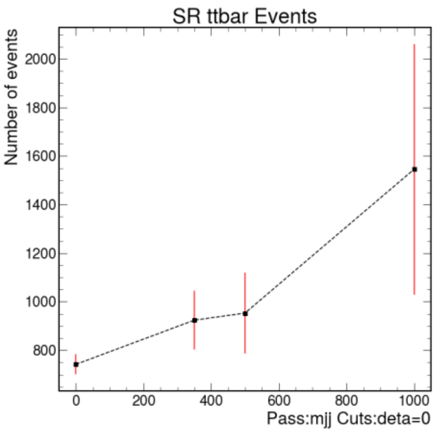
$$\Delta SF = \Delta N_D / N_C$$

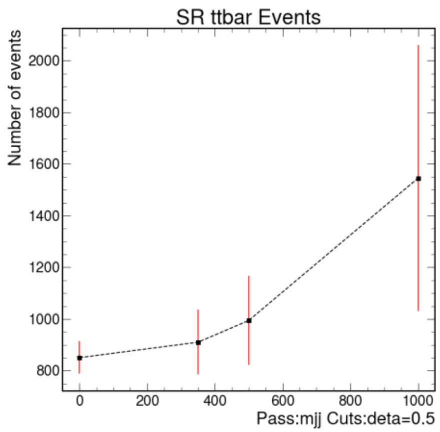
$$\Delta N_S = \Delta SF * N_S$$

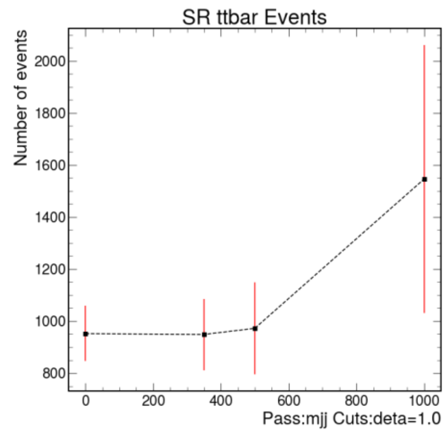
Ddbv1 passing score

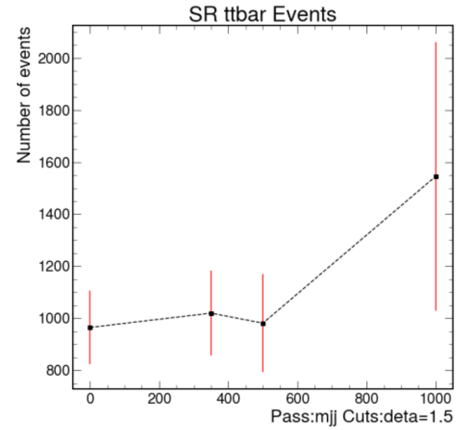
Statistics Report

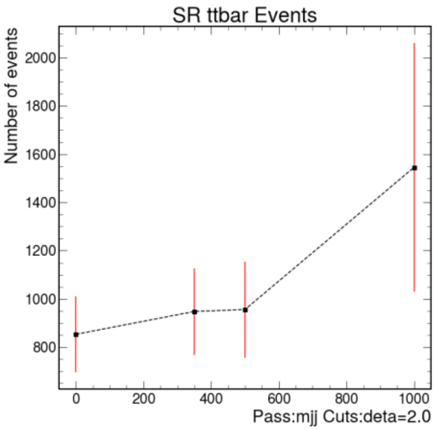
DDBv1 passing score-Fixed deta cuts

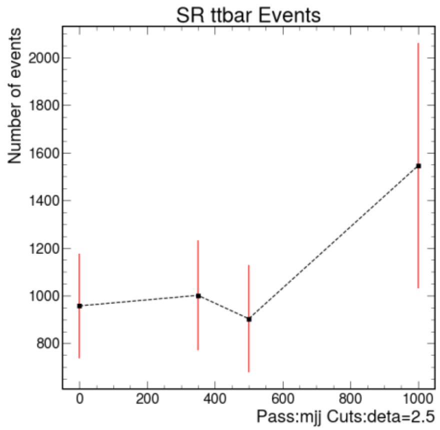


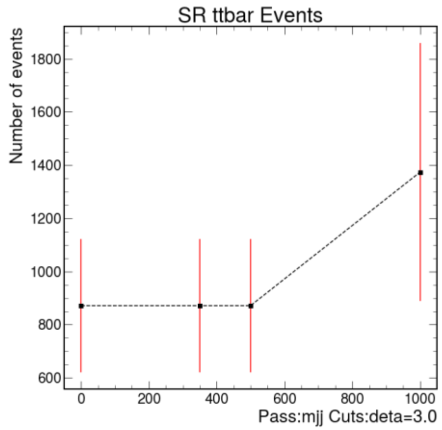


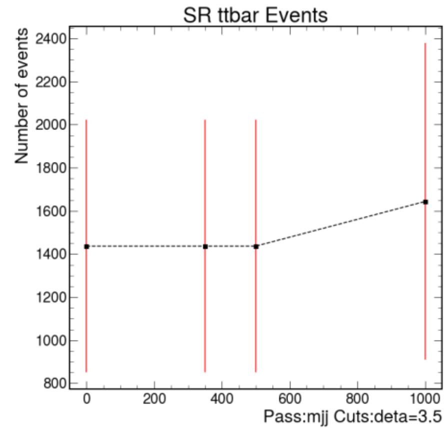






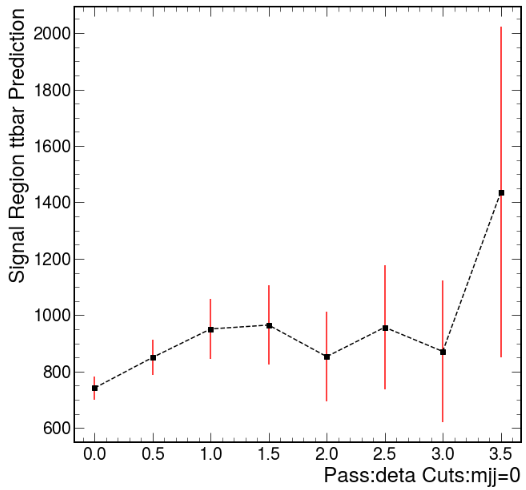


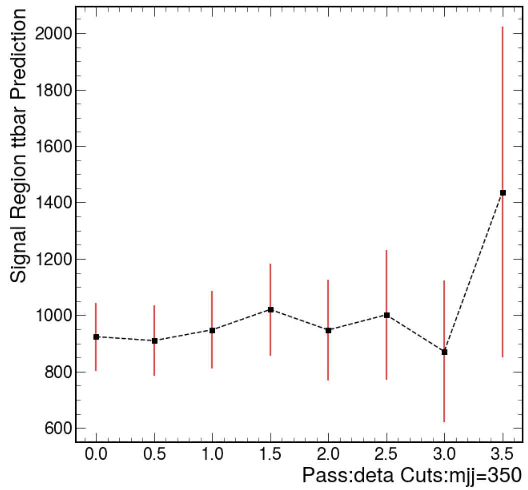


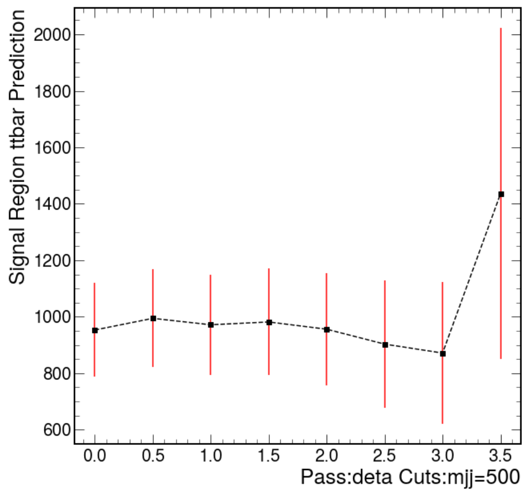


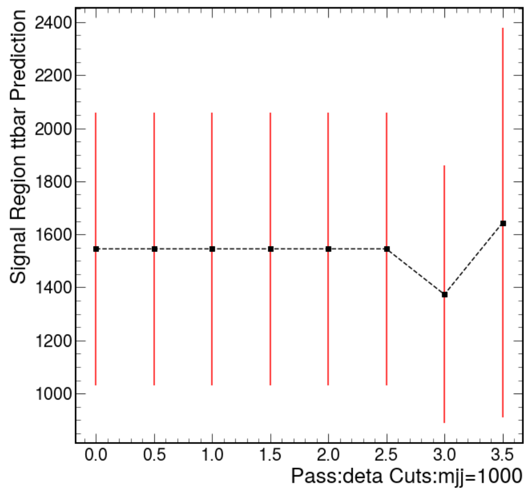
Statistics Report

Passing score for fixed m_{jj} cuts



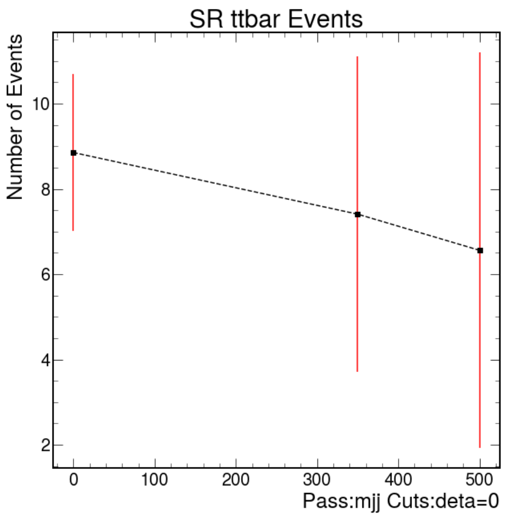


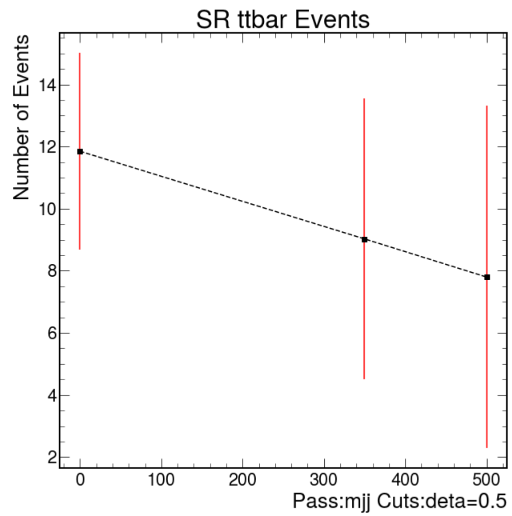


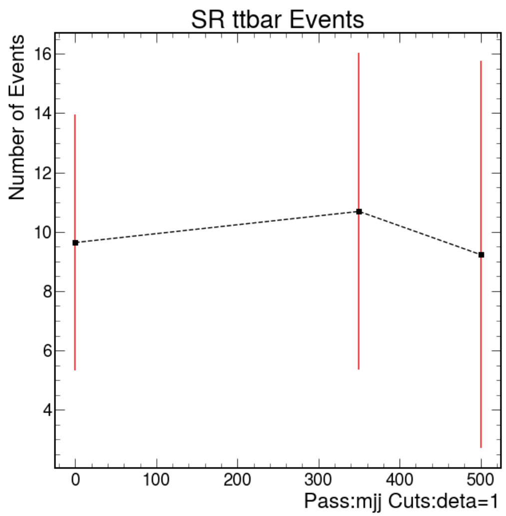


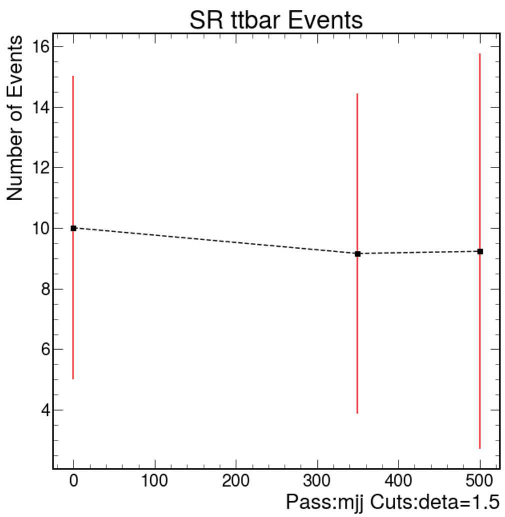
Statistics Report

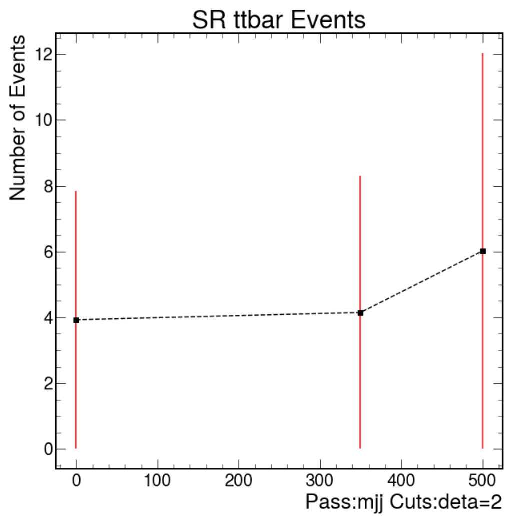
DDBv2 passing score- Fixed deta cuts

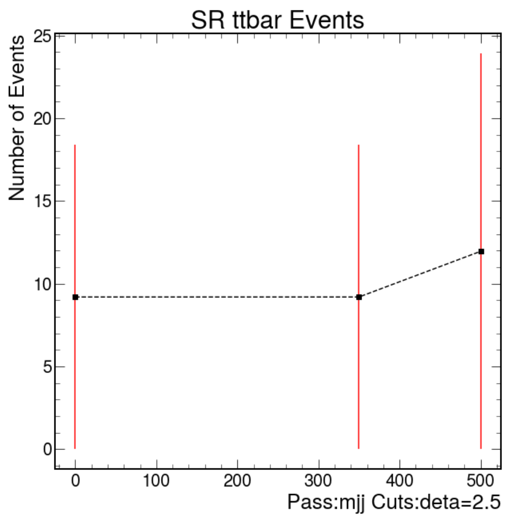






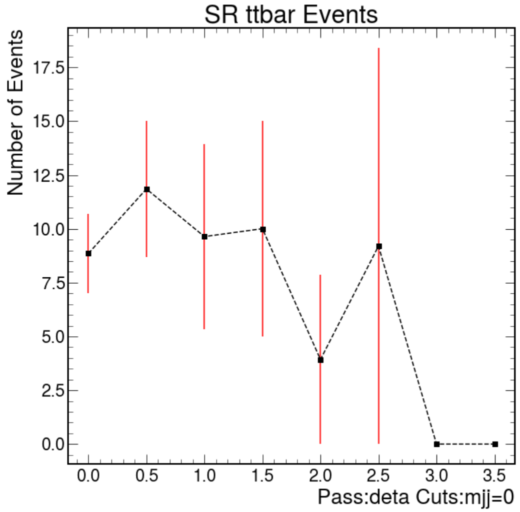


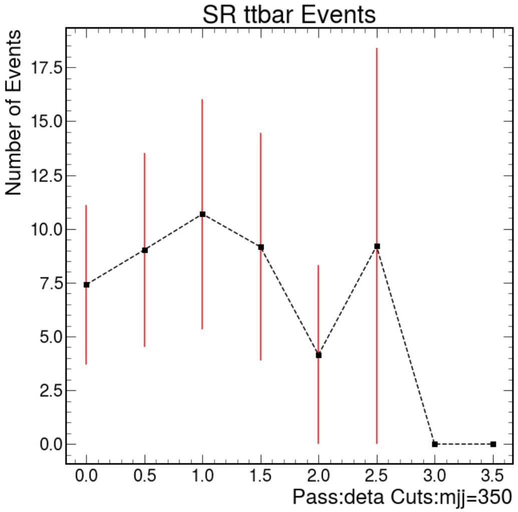


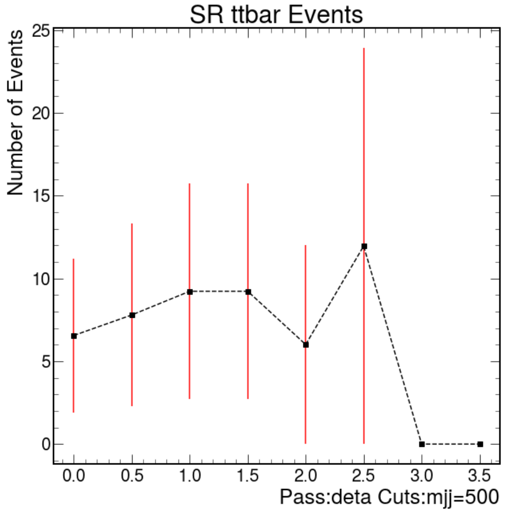


Statistics Report

Passing score for fixed m_{jj} cuts

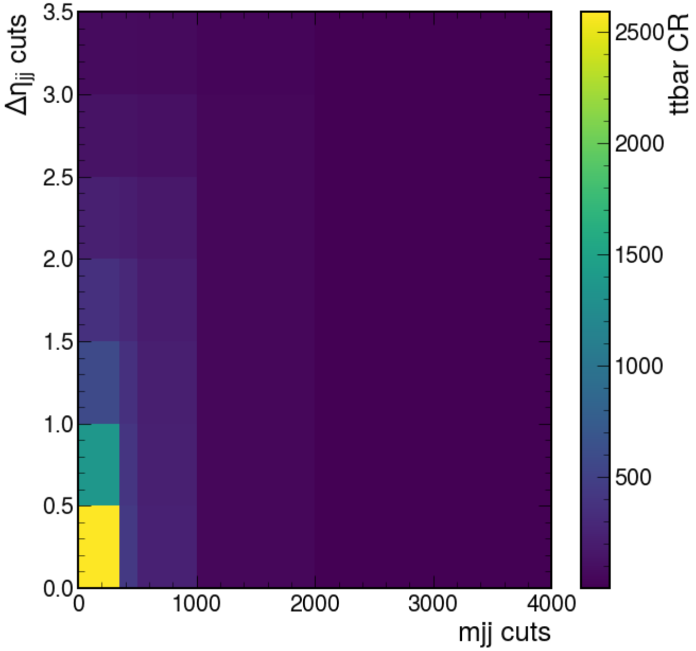






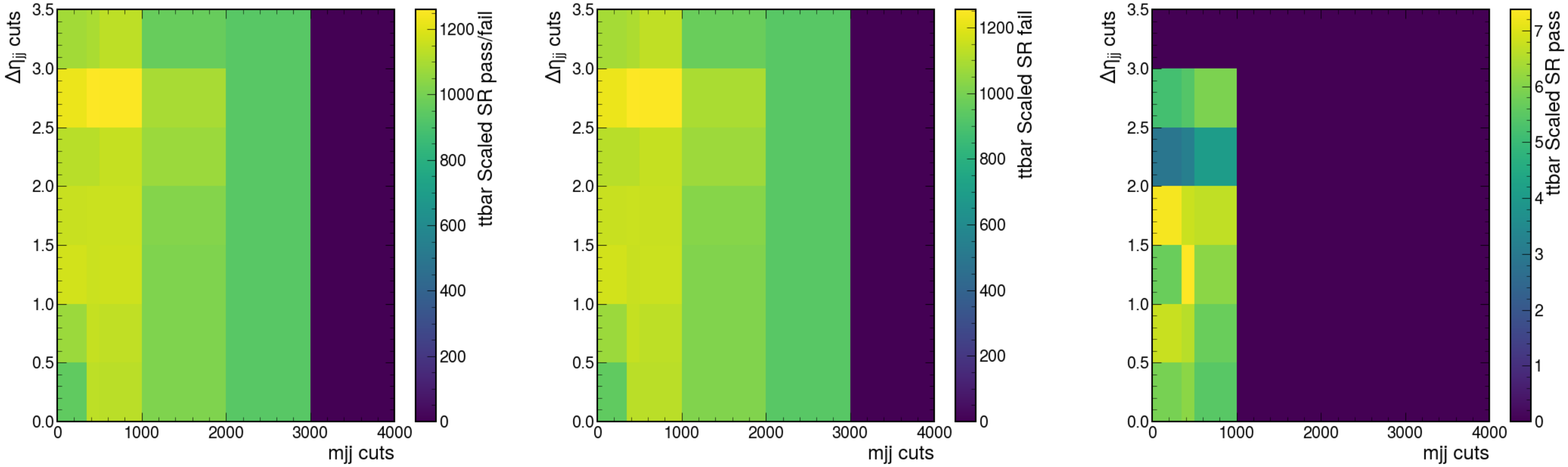
Statistics Report

- The event count becomes critically low in the passing regions for ddbv2.
  - This due to the MC and data events running low.



2d histogram plot that demonstrates how the ttbar MC events of the control region decrease as the cuts become stricter. The ddbv2 tagger also causes a notable decrease in count.

Statistics Report

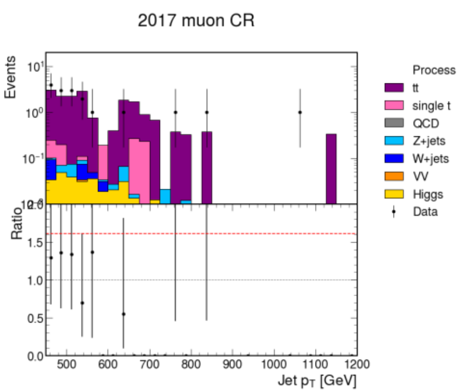


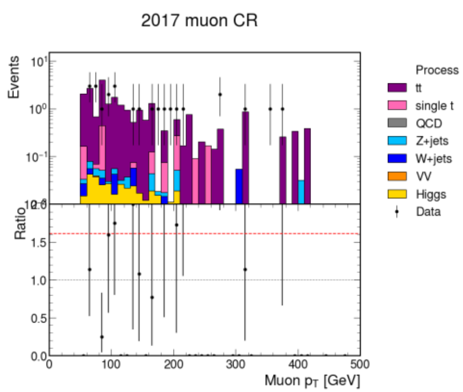
These 2d histograms shows how the ddbv2 tagger decreases the event count significantly in the passing region.

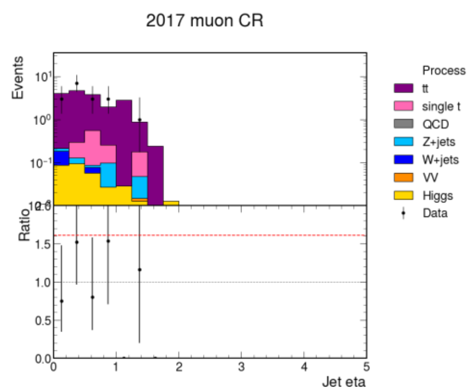
To increase the counts, we removed the m_{jj} and deta cuts, and allowed events with less than 2 jets. We issue the large uncertainty on the scale factor to account for any potential biases.

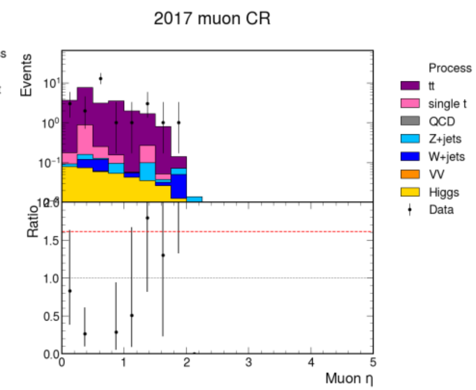
Validation

Within statistical uncertainty, we have pretty good simulation to data agreement for the control region.









The value for the scale factor in the passing region is indicated by the red horizontal line in the ratio plot.