Background Estimations for High Energy Higgs Boson Measurements with the CMS Experiment

Keston Smith

Department of Physics, University of Alabama at Birmingham-SIST Intern

August 13, 2021

Abstract

In the events following proton-proton collisions, we look for Higgs produced through vector boson fusion (VBF) at high p_T ranges. Usually gluon-gluon fusion is responsible for 90% of the Higgs cross sections, but this isn't the case at high p_T. In our analysis we strive estimate the $t\bar{t}$ processes where there are Higgs signal which will in turn allow us to become more certain a Higgs is observed.

Introduction

Recent CMS experiments at the LHC tells us that gluon fusion is responsible for about 90% of Higgs cross sections at 125 GeV. In Figure 1 we see that percentage decreases as we look at Higgs cross sections in high transverse momentum ranges, and we see other production contributions become more significant. The production mode of interest for my project is vector boson fusion. VBF is a process in proton collisions where two quarks or anti quarks scatter after the exchange of gauge bosons; the emitted bosons then fuse to produce a Higgs boson. Due to the instability of the Higgs, it immediatly decays into other particles within the constraints of conservation laws. For the likes of my project, the bottom quark decay channel is considered. The Standard Model predicts that that the Higgs will decay to a pair of b-quarks about 58% of the time. The Higgs decay hadronically into b-quark jets which registers in the CMS detector as a one large radius jet; these b-jets posses a distinct two prong substructure. This enables a more efficient method of distiguishing Higgs events from the pool of others.

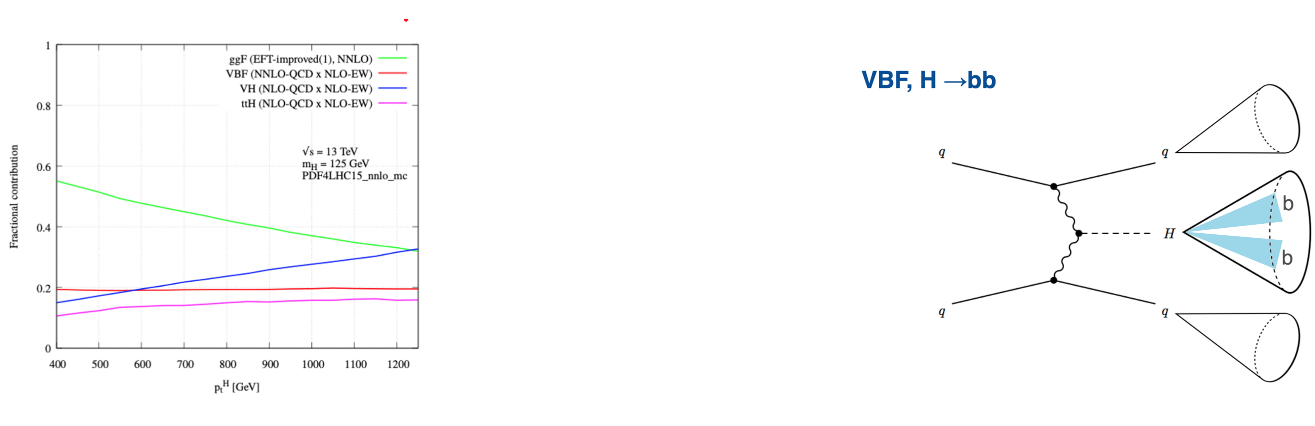

Figure 1: From left to right, a plot of the fractional contribution of various Higgs production modes as a function of p_T and an example of a Higgs production through VBF and its decay into bottom quarks. It is important to note the decrease in the gluon fusion contribution as p_T increase.

A control region made up of mostly $t\bar{t}$ processes is used to correct the $t\bar{t}$ process in the signal prediction. Particularly, top quark processes in this control region involves semi-leptonic productions of a muon with $p_T > 55$ GeV. We use Monte Carlo (MC) event generators and VBF Higgs jet criterion to model a signal region. The control region is selected to be the same as the signal with the exeption of the presence of the lepton. It is then used to derive a scale factor which is then used to estimate the $t\bar{t}$ events in the signal prediction. To understand this process in the signal region to then provide accurate Higgs neasurements is the goal of my work.

Signal & Control Region

As mentioned before, the signal and control samples were modeled using a collection MC event genterators. A separate simulation program was used to replicate the CMS detector. The signal model have specific ”cuts” in place to select VBF and b-quark events associated with a Higgs and to mitigate the large QCD background that correlates to high transverse momentum processes. A VBF Higgs jet candidate must have a p_T >450 GeV, $|\eta|$ >2.5, $\Delta\eta$ >3.5, m_{jj} >1000 GeV. The variable $|\eta|$ is the angle of the jet with respect to the direction of the beam and it is measured in radians. $\Delta\eta$ is the angle of separation between the two smaller jets, and m_{jj} is the jet invariant mass. The deep double b tagger (DDBT) algorithm is based on a neural network and is used to select jets with b-quark decay vertices. A score from 0 to 1 is assigned to each jet candidate. I used two versions of this algorithm for my analysis; the score is split up into passing and failing regions; a jet must have a DDB score >0.89 for version one and a score >0.64 for version two. In the control region $|\eta| < 2.1$ and the leading large radius jets must have a $p_T > 400$ GeV. With the strict VBF selections, the statistics in the control region are limited which leads to large statistical uncertainties in the $t\bar{t}$ background estimate. In order to increase the statistics in the $t\bar{t}$ control region, we make successively looser VBF cuts. Then we must verify that the scale factor stays consistent, within uncertainties, in order to use the $t\bar{t}$ scale factor prediction from the control region with looser cuts.

VBF Cuts	Data ($t\bar{t}$)	Control($t\bar{t}$)	Scale Factor ($\frac{N_D}{N_C}$)
m_{jj}(0,4000) $\Delta\eta$(0,7)	2584.03	3733	1.44
m_{jj}(0,4000) $\Delta\eta$(0.5,7)	399.34	679	1.70
m_{jj}(350,4000) $\Delta\eta$(1,7)	216.76	382	1.76
m_{jj}(500,4000) $\Delta\eta$(1.5,7)	46.3	72	1.55
m_{jj}(1000,4000) $\Delta\eta$(2,7)	44.27	72	1.63
m_{jj}(1000,4000) $\Delta\eta$(2.5,7)	40.96	68	1.66
m_{jj}(1000,4000) $\Delta\eta$(3,7)	36.74	54	1.46
m_{jj}(1000,4000) $\Delta\eta$(3.5,7)	22.38	36	1.60

Table 1: A table of the number of events for the $t\bar{t}$ process for the data and control region for successive m_{jj} and $\Delta\eta$ cuts. The last column are the scale factors used to make the signal predictions.

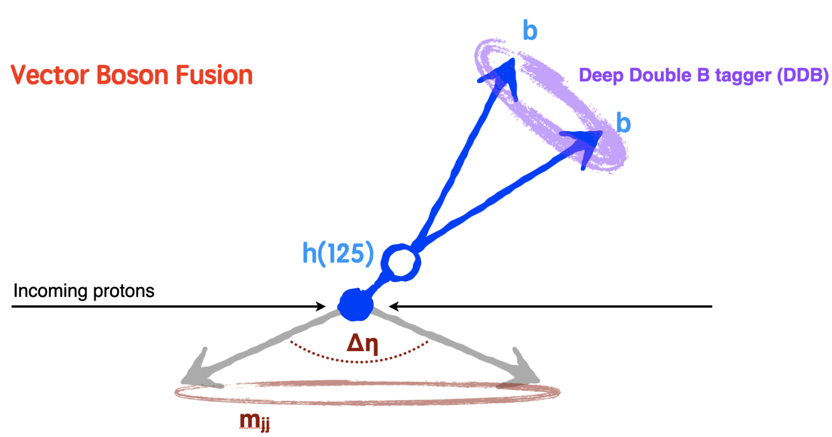

Figure 2: A more deatiled diagram of jet kinematics and variables mentioned in the previous paragraph. It shows the production of a Higgs through VBF and its decay to a bottom quark pair.

The event counts in the control region for the DDB version two become servely low. This resulted in large uncertainties in the scale factor and signal prediction post scale factor. A synopsis of the statistical anaysis done will be included in the followng section.

Statistical Analysis

The uncertainties in the simluated counts is assumed to be negligible, only the uncertainties in the data is the focus here. The data follows a poisson distribution where the mean is the number of events for particular VBF cuts.

$$P(n;\lambda) = \frac{e^{-\lambda}\lambda^{n}}{n!} \tag{1}$$

The uncertainties can then be computed by taking the square root of the mean.

$$\sigma = \sqrt{\lambda} \tag{2}$$

The computation of the uncertainties in the data, scale factor, and signal prediction are given by,

$$\sigma_{SF} = \frac{\sigma_D}{N_C} \tag{3}$$

$$\sigma_D = \sqrt{N_D} \tag{4}$$

$$\sigma_S = \sigma_{SF} \cdot N_s \tag{5}$$

Using the matplotlib plotting module in python, I've created plots of the signal $t\bar{t}$ prediction counts as a function of m_{jj} and $\Delta\eta$ cuts. The first set is for the passing DDBT version one and the second is for version two passing score. These plots serve as a good visual aid to understanding how the uncertainties grow as VBF cuts become more strict.

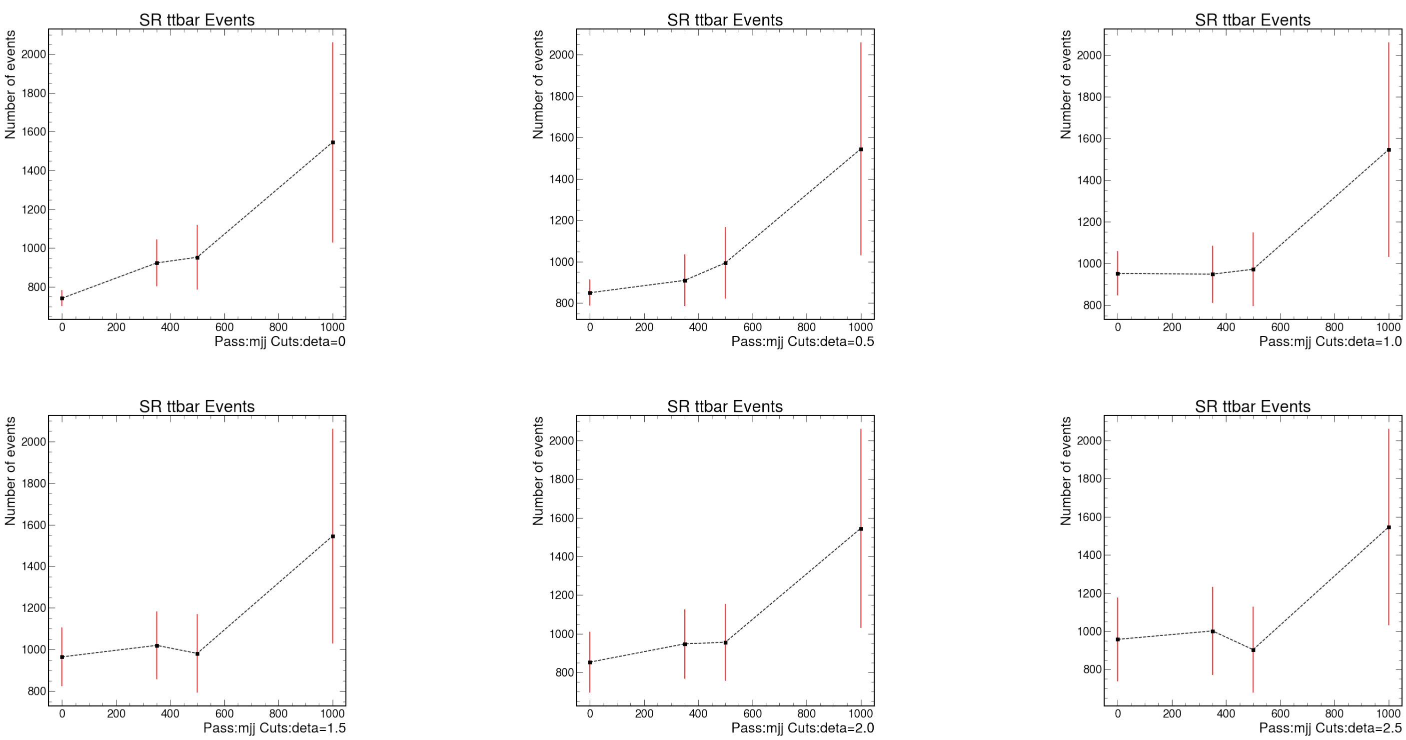

Figure 3: Predicted number events for the $t\bar{t}$ process in the signal region after the scale factor adjustments for DDBT v1 passing score. The x-axis is the range on m_{jj} cuts and each plot is for different fixed $\Delta\eta$ cuts. The uncertainties become larger as the tigher cuts are introduced.

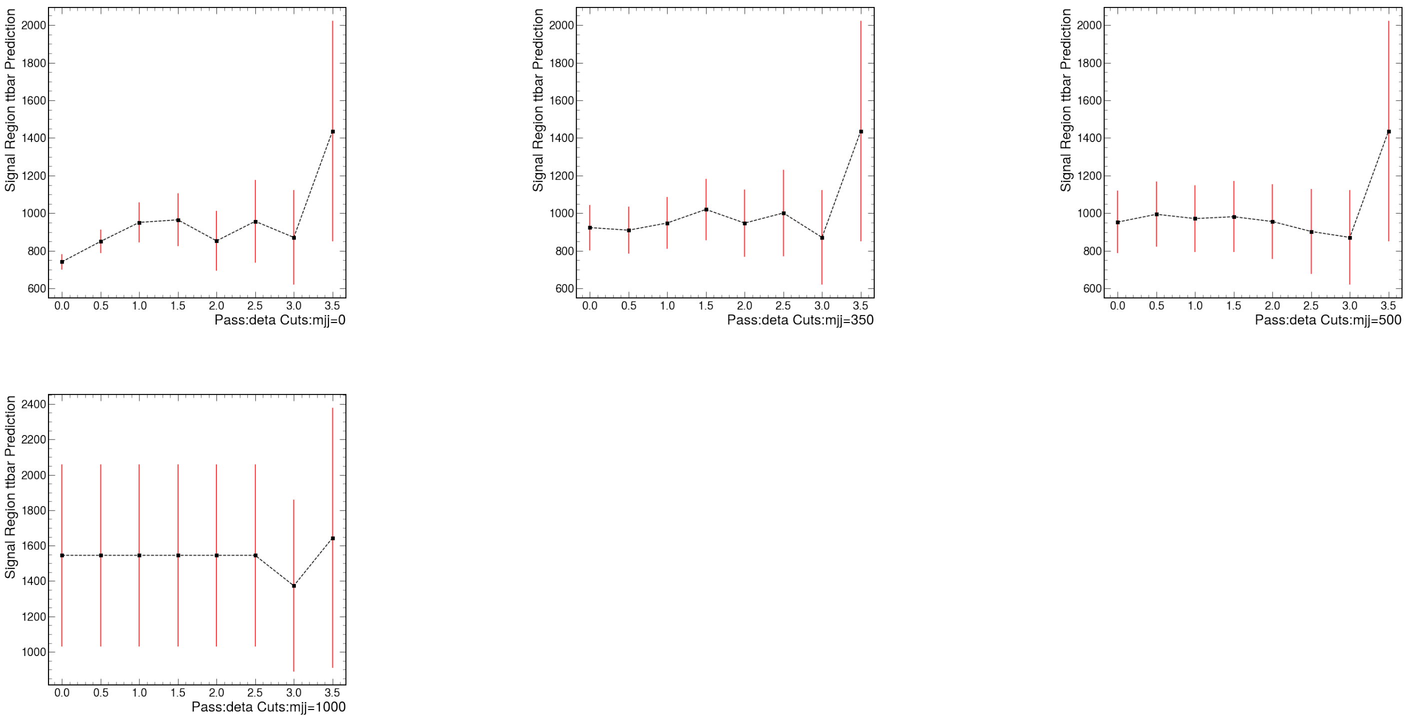

Figure 4: Signal region prediction for the range of $\Delta\eta$ cuts and fixed m_{jj} cuts for DDBT v1 passing score.

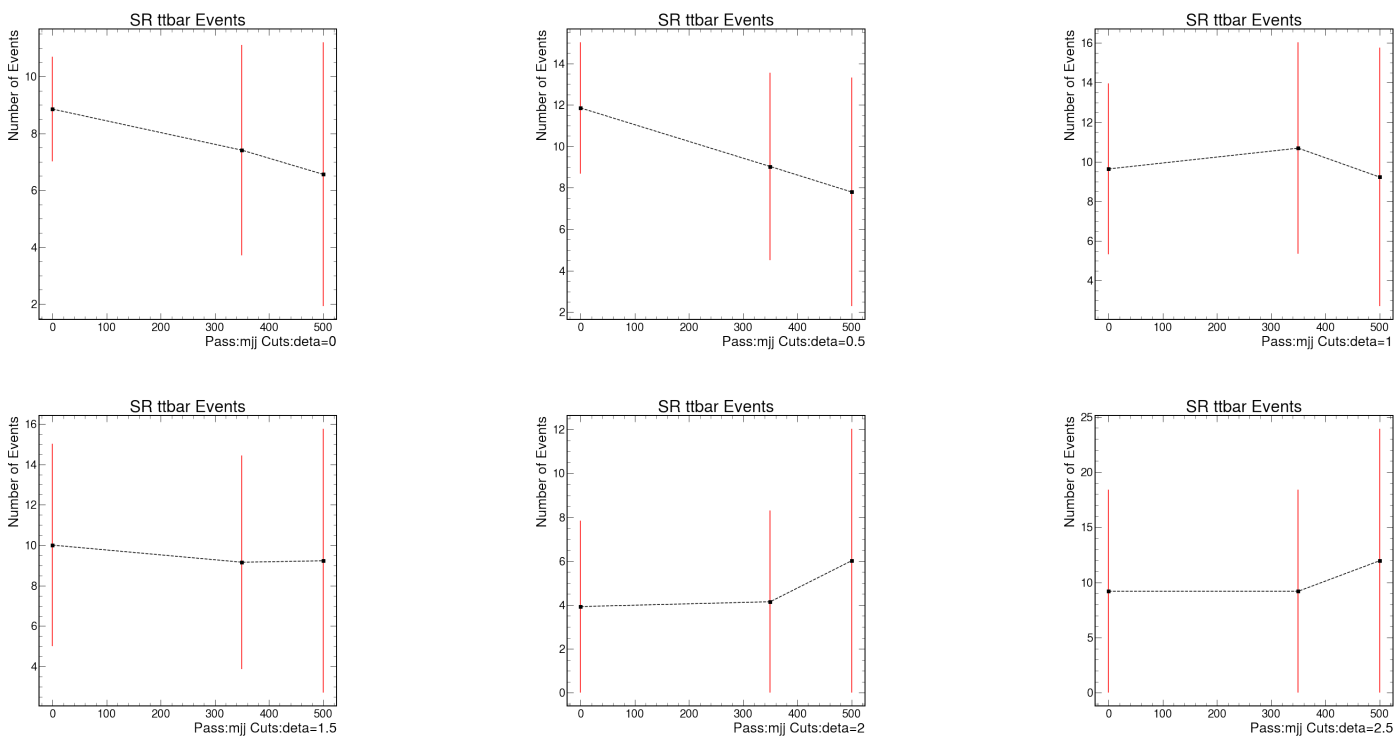

Figure 5: Signal predictions for DDBT version two for the range of m_{jj} cuts and fixed $\Delta\eta$ cuts.

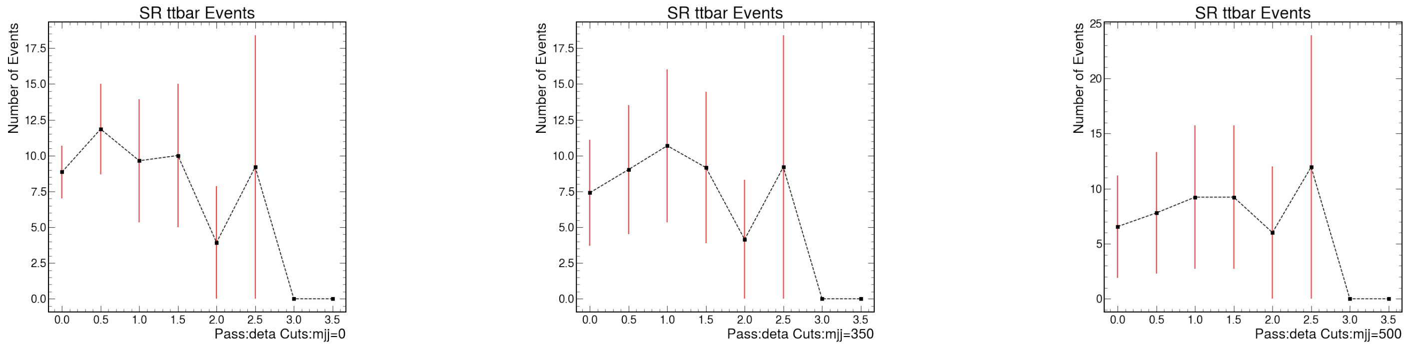

Figure 6: Signal predictions for DDBT version two for the range of $\Delta\eta$ cuts and fixed m_{jj} cuts.

In the DDBT version two plots, the uncertainties are notably larger. This is so due the number of events in the control region becoming even lower. Since this quantity is used to obtain the scale factor, the associated uncertainties become large which promotes even larger uncertainties in the signal $t\bar{t}$ prediction.

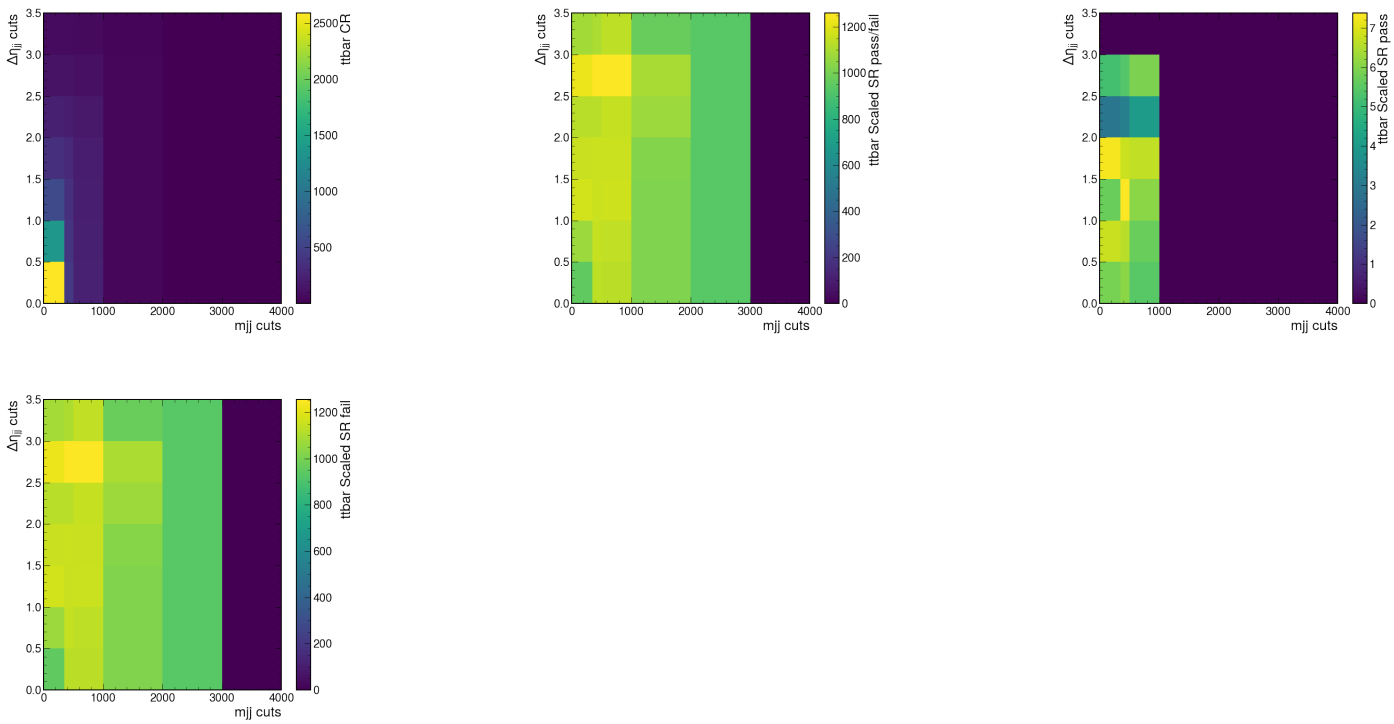

Figure 7: 2d histograms of control and signal regions. These plots show how the vbf cuts and DDBT scores affect the number of events in each region. The first plot is the control region and the other three are signal regions for different DDBT scores.

Results

Searching for Higgs in high p_T ranges come with the issue of large number of QCD multijet events, this makes sense energentically. For this reason, the VBF cuts plays an important role in filtering through the events that has nothing to do with Higgs processes. For the control sample, VBF cuts and DDBTv2 are the reason behind the low counts which contributes to the ever growing uncertainties in the signal predictions. As one could imagine the effects of large uncertainties to a measurement. The suggestive solution to this problem was to remove all VBF cuts in the control region and allow events with less than the minimum required number of jets to increase the counts. This allowed for better validation plots to be produced and overall better statistics. To cover potential biases in the signal prediction, a large contrived uncertainty is given to the scale factor.

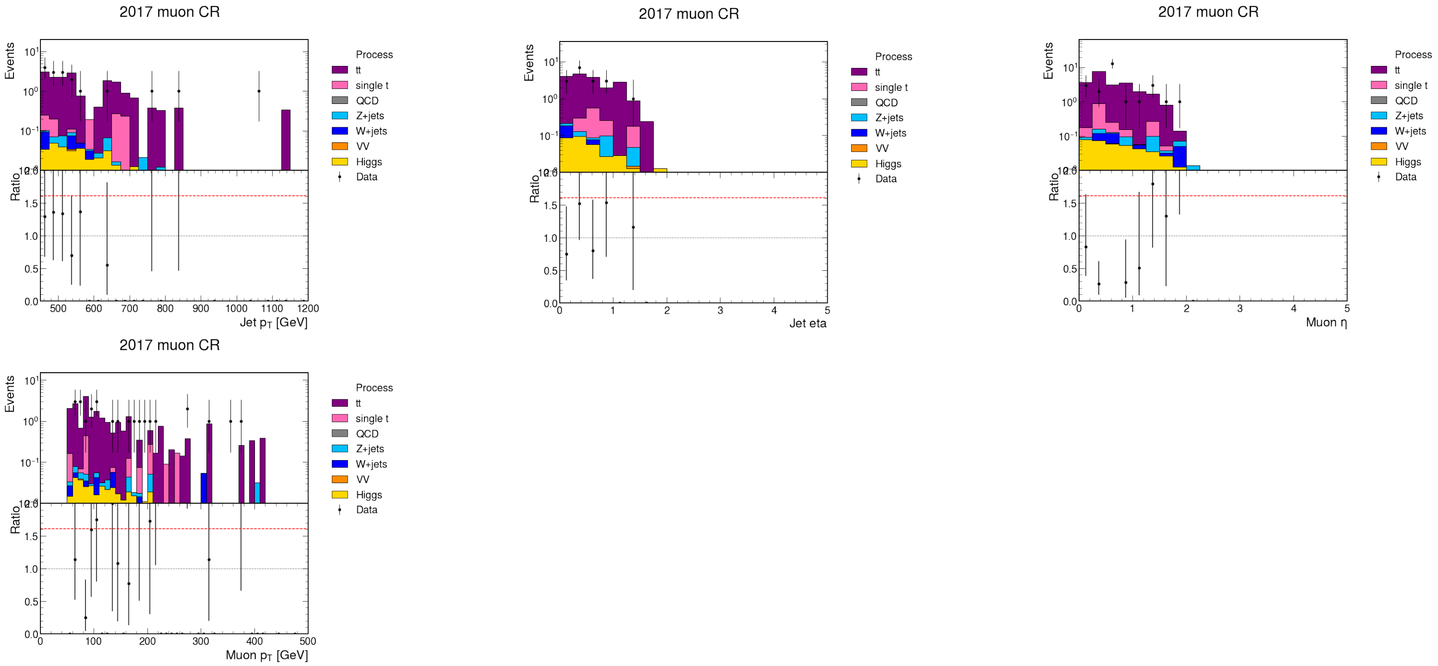

Figure 8: Simulation to data validation plots; the value of the of the scale factor with full cuts is 1.60 which is indicated by the red horizontal line in the ratio plots. There is good agreement within the statistical uncertainties about this value.

It is important that the $t\bar{t}$ estimate is as accurate possible in the signal prediction such that the correct amount of Higgs is measured. Accurate Higgs measurements corresponds to discovering more about its nature which will allow one to learn more about its interations beyond Standard Model predictions.

References

[1] CMS Collaboration, "Inclusive search for highly boosted Higgs bosons decaying to bottom quark antiquark pairs in proton proton collisions at $\sqrt{s} = 13$ TeV", *Phys. Lett. B*120 (2018) 071802

[2] CMS Collaboration, "Observation of Higgs boson decay to bottom quarks", *Phys. Rev. Lett.* 121 (2018) 121801

[3] CMS Collaboration, "Inclusive search for a highly boosted Higgs boson decaying to a bottom quark-antiquark pair",*Phys. Rev. Lett.* 120 (2018) 071802,