Highly boosted Higgs bosons decaying to W pairs

Daniel Ocampo-Henao[1]

Supervisor: Cristina Mantilla Suarez
CERN Summer Student Programme 2021

We present the search for the standard model Higgs boson with $p_T > 200$ GeV decaying to W vector boson pairs and reconstructed as a single, large-radius jet. The search uses a data sample of proton-proton collisions at $\sqrt{s} = 13$ TeV corresponding to an integrated luminosity of 41.5 fb^{-1}. We optimize the pre-selection of the signal regions in two decay channels and compare the data and simulation agreement.

1 Introduction

One of the major achievements of the LHC has been the observation of the Higgs boson (H) in 2012 by the ATLAS and CMS experiments [1, 5]. Since then, there has been a great effort to measure its properties with very good results in this regard. Nowadays, the Higgs boson have been observed in almost all of its main production and decay modes. Nevertheless, the properties of the Higgs boson at very high transverse momenta p_T remain largely unconstrained.

In this project, we are interested in studying the properties of the Higgs boson when it is produced at very high p_T in the decay channel H(WW), which has the second highest branching ratio as shown in Fig. 1.

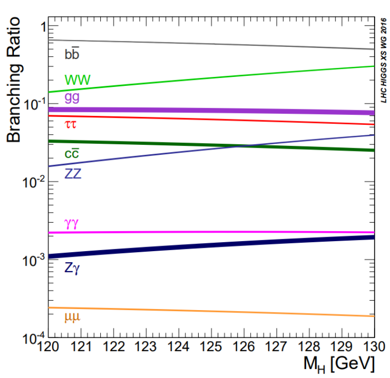

Figure 1 : Higgs boson branching ratios and their uncertainties for the mass range around 125 GeV [7]

The W vector boson can decay into a lepton and neutrino or two quarks, which means that in the H(WW) process we can have three possible final states: $l\nu_l l\nu_l$ (leptonic channel), $qqqq$ (hadronic channel) and $l\nu_l qq$ (semileptonic channel).

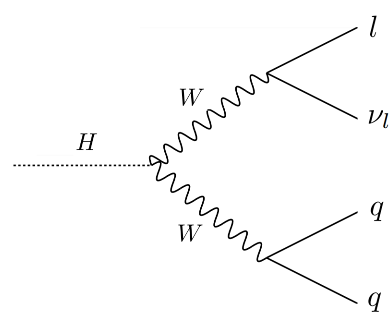

Figure 2 : H(WW) semileptonic channel

At very high p_T the decay products of the Higgs boson get very close together and collimated, so one can reconstruct the Higgs with a single jet with distance parameter $R = 0.8$.

2 CMS detector

The Large Hadron Collider (LHC) accelerates protons to nearly the velocity of light and then collides them at four locations around its ring. The Compact Muon Solenoid (or CMS) detector lies at one of these four collision points.

[1]daniel.ocampoh@udea.edu.co

The CMS detector consists of an inner tracking detector made of individual electronic sensors arranged in concentric layers immersed in a magnetic field of 3.8 T, which reconstructs the curved track left by a charged particle that passes through it. Externally to the tracker, there is an electromagnetic calorimeter (ECAL) which measures the energy of electrons and photons, and a hadronic calorimeter (HCAL) which measures the energy of the hadronic particles. Finally, the most external layer consists of muon chambers, again immersed in a large magnetic field, which reveal the track of muons produced in the interaction point.

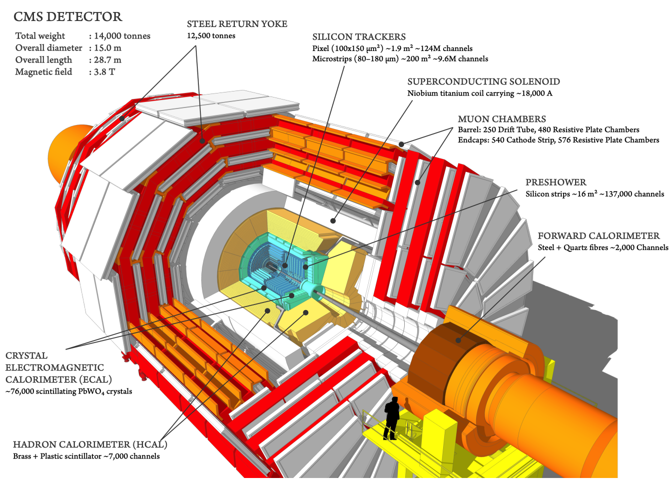

Figure 3 : Diagram of the CMS detector [9]

Events of interest are selected using a two level trigger system [2]. The first level, uses information from the calorimeters and muon detectors to select events at a rate of around 100 kHz within a time interval of less than 4 µs. The second level, known as the high-level trigger, consists of a event reconstruction software optimized for fast processing, and reduces the event rate to around 1 kHz before data storage [2].

A more detailed description of the CMS detector, together with a definition of the coordinate system used and the relevant kinematic variables, can be found in Ref. [2].

3 Data and simulated samples

3.1 Data

In this analysis we use events from datasets with a single muon or a single electron. These datasets, from the 2017 run, correspond to an integrated luminosity of 41.5 fb^{-1}. Table 1 shows the datasets used in the analysis.

Dataset
/SingleElectron/Run2017B-02Apr2020-v1/NANOAOD
/SingleElectron/Run2017C-02Apr2020-v1/NANOAOD
/SingleElectron/Run2017D-02Apr2020-v1/NANOAOD
/SingleElectron/Run2017E-02Apr2020-v1/NANOAOD
/SingleElectron/Run2017F-02Apr2020-v1/NANOAOD
/SingleMuon/Run2017B-02Apr2020-v1/NANOAOD
/SingleMuon/Run2017C-02Apr2020-v1/NANOAOD
/SingleMuon/Run2017D-02Apr2020-v1/NANOAOD
/SingleMuon/Run2017E-02Apr2020-v1/NANOAOD
/SingleMuon/Run2017f-02Apr2020-v1/NANOAOD

Table 1: Datasets used for the analysis

We use a logical "OR" of the following triggers

- Electron triggers:
  - Ele50_CaloIdVT_GsfTrkIdT_PFJet165
  - Ele115_CaloIdVT_GsfTrkIdT
  - Ele15_IsoVVVL_PFHT600
- Muon triggers:
  - Mu50
  - Mu15_IsoVVVL_PFHT600

3.2 Background simulation

We consider the following background processes: multijet QCD, W($\rightarrow l\nu$) + Jets, $Z(ll)$ + Jets, single t and dileptonic, semileptonic and hadronic $t\bar{t}$. All the background MC samples used in the analysis are shown in Table 2.

Dataset	σ (pb)
QCD_HT300to500_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8	347700
QCD_HT500to700_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8	32100
QCD_HT700to1000_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8	6831
QCD_HT1000to1500_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8	1207
QCD_HT1500to2000_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8	119.9
QCD_HT2000toInf_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8	25.24
WJetsToLNu_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8	359.7 * 1.21
WJetsToLNu_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8	48.91 * 1.21
WJetsToLNu_HT-600To800_TuneCP5_13TeV-madgraphMLM-pythia8	12.05 * 1.21
WJetsToLNu_HT-800To1200_TuneCP5_13TeV-madgraphMLM-pythia8	5.501 * 1.21
WJetsToLNu_HT-1200To2500_TuneCP5_13TeV-madgraphMLM-pythia8	1.329 * 1.21
TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8	369.487
DYJetsToLL$_{Pt}t-50To100_{M}atchEWPDG20_{T}uneCP5_{1}3TeV-amcatnloFXFX-pythia8$	398.8
DYJetsToLL_Pt-100To250_MatchEWPDG20_TuneCP5_13TeV-amcatnloFXFX-pythia8	93.61
DYJetsToLL_Pt-250To400_MatchEWPDG20_TuneCP5_13TeV-amcatnloFXFX-pythia8	3.67
DYJetsToLL_Pt-400To650_MatchEWPDG20_TuneCP5_13TeV-amcatnloFXFX-pythia8	0.5
DYJetsToLL_Pt-650ToInf_MatchEWPDG20_TuneCP5_13TeV-amcatnloFXFX-pythia8	0.04704
ST_s-channel_4f_leptonDecays_TuneCP5_13TeV-amcatnlo-pythia8	3.549
ST_t-channel_antitop_5f_InclusiveDecays_TuneCP5_13TeV-powheg-pythia8	71.74
ST_t-channel_eleDecays_TuneCP5_13TeV-comphep-pythia8	24.78
ST_t-channel_muDecays_TuneCP5_13TeV-comphep-pythia8	24.73
ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8	34.97
ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8	34.91

Table 2. List of background MC samples with their corresponding cross sections

3.3 Signal simulation

The MC sample we use as signal is a semileptonic gluon-fusion production mode of Higgs:

GluGluHToWWToLNuQQ_M125_TuneCP5_PSweight_13TeV-powheg2-jhugen727-pythia8

with a cross section $\sigma = 0.9913$ pb.

4 Object reconstruction

The object reconstruction is based on a particle-flow (PF) algorithm [10], which aims to reconstruct and identify each individual particle with an optimized combination of information from the various elements of the CMS detector. The algorithm identifies each reconstructed particle as an electron, a muon, a photon, or a charged or neutral hadron. The missing transverse momentum vector is defined as the negative vector sum of the transverse momenta of all the particles identified in the event, and its magnitude is referred to as p_T^{miss}, representing the energy carried away by the neutrino(s).

4.1 Jet

Jets used in the analysis are reconstructed by clustering particle flow candidates in the event using the anti-k_T algorithm [4] with a distance parameter of $R = 0.8$ (AK8 jet). Next, we select the jet closest to the missing energy in the events. These jets are required to have $p_T > 200$ GeV and be within a pseudorapidity region $|\eta| < 2.5$.

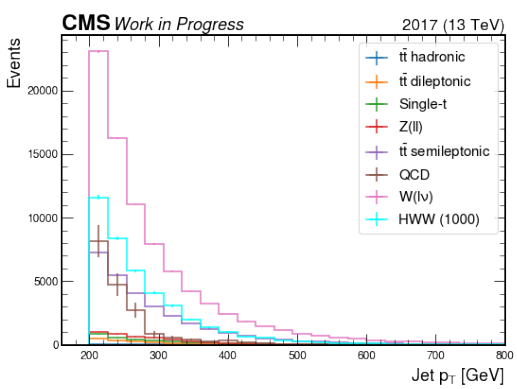

Figure 4 : AK8 Jet p_T distributions for signal and background samples normalized to their cross sections, before any selection.

4.1.1 Jet mass

We also apply the soft drop algorithm [8] that removes soft contributions from the jet in order to mitigate the initial state radiation and pileup effects. The soft drop jet mass m_{SD} is used in the analysis requiring that $m_{sd} > 15$ GeV. Specific corrections to the soft drop mass are applied to account for a small shift in the generated vector boson mass [3].

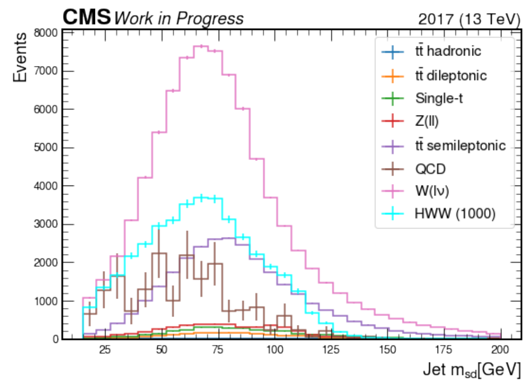

Figure 5 : Corrected AK8 Jet soft drop mass distributions for signal and background samples normalized to their cross sections.

4.2 AK4 b-jets

Since the top quark can decay via the weak interaction producing a W vector boson and a b-quark, we can use a jet flavor tagging variable to reduce the $t\bar{t}$ semileptonic contamination. The DeepFlavor algorithm [6] is used to identify AK4 jets that come from b-quarks. Figure 6 shows the distributions of the maximum DeepCSV discriminator value (btagDeepFlavB) for AK4 jets in the opposite hemisphere of the AK8 jet.

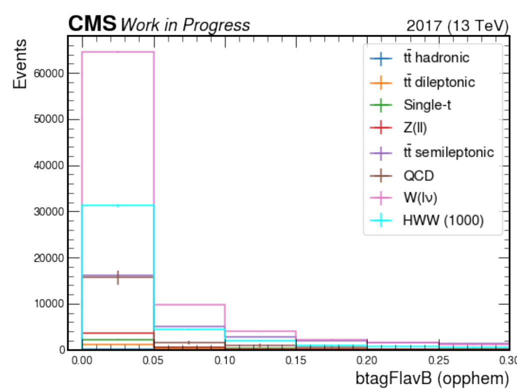

Figure 6 : Distribution of the maximum DeepCSV discriminator value (btagDeepFlavB) for AK4 jets in the opposite hemisphere of the AK8 jet for signal and background samples normalized to their cross sections.

4.3 Lepton selection

Leptons are required to pass the following identification and kinematic criteria:

Muons:

- $p_T > 25$ GeV.
- $|\eta| < 2.4$.
- Loose ID

Electrons:

- $p_T > 25$ GeV.
- $|\eta| < 2.5$.
- mvaFall17V2noIso_WP80

Tau veto:

- $p_T > 20$ GeV.
- $|\eta| < 2.3$.
- idAntiEle ≥ 8
- idAntiMu ≥ 1

4.4 Missing transverse energy

Events in data can end up with large spurious p_T^{miss} due to detector noise and beam backgrounds. In order to remove these events, the following p_T^{miss} filters have been implemented in the analysis:

- goodVertices
- globalSuperTightHalo2016Filter
- HBHENoiseFilter
- HBHENoiseIsoFilter
- EcalDeadCellTriggerPrimitiveFilter
- BadPFMuonFilter

We require that $p_T^{miss} > 20$ GeV in the analysis.

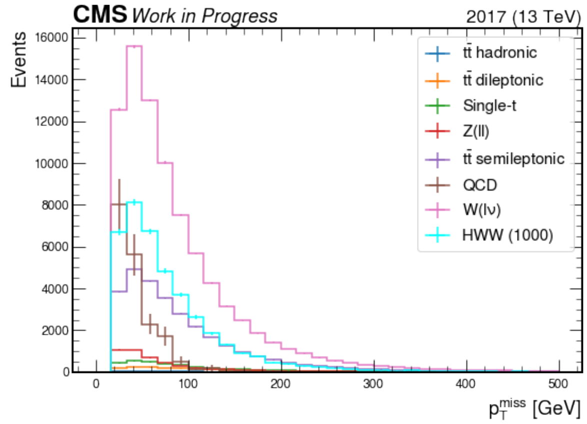

Figure 7 : p_T^{miss} distributions for signal and background samples normalized to their cross sections.

In the semi-leptonic decay channel of the W vector boson, we can reduce the background contamination using the transverse mass m_T, defined by

$$m_T = \sqrt{2\, p_T^{\text{lep}} p_T^{miss} \left[1 - \cos \Delta\phi(\vec{p}_T^{\text{lep}}, \vec{p}_T^{miss})\right]} \quad (1)$$

where $\Delta\phi$ is the difference in the azimuthal angle between $\vec{p}_T^{\text{lep}}$ and $\vec{p}_T^{miss}$. We require that $m_T < 80$ GeV.

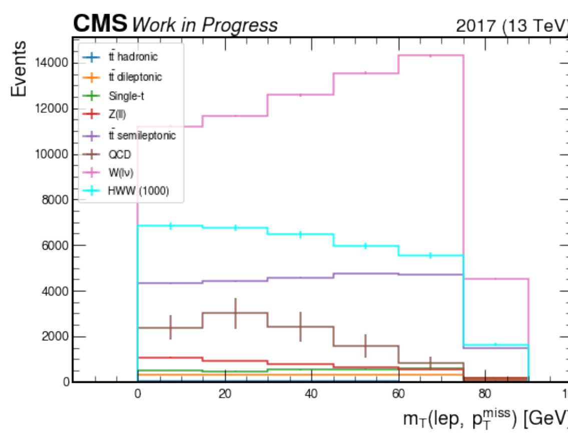

Figure 8 : m_T distributions for signal and background samples normalized to their cross sections.

5 Signal regions

Additional to the baseline event selection defined in section 4, we define the following signal regions with two categories: muon and electron regions:

Muon	**Electron**	**Description**
triggermu	triggerel	Trigger
met_filters	met_filters	MET filters
lep_in_fj	lep_in_fj	ΔR(lep, AK8 Jet) < 0.8
fjmsd	fjmsd	AK8 Jet $m_{sd} > 15$ GeV
onemuon	oneelectron	One single lepton
mu_iso		$p_T < 55$ GeV & RelIso < 0.25
		$p_T \geq 55$ GeV & miniIso < 0.1
	el_iso	$p_T < 120$ GeV & RelIso < 0.25
		$p_T \geq 120$ GeV & miniIso < 0.1
btag_med	btag_med	btagFlavB (opphem) < 0.3
mt_lep_met	mt_lep_met	$m_T < 80$ GeV

Table 3. Signal regions

where the btagFlavB (opphem) cut corresponds to the medium working point for the 2017 data.

The lepton isolation is computed as the sum of the transverse energy of the particles inside a cone of radius R around the lepton direction divided by the lepton transverse momentum. The Relative Isolation (RelIso) uses a cone size of $R = 0.4$, while the Mini Isolation (miniIso) uses a cone size of

$$R = \begin{cases} 0.2, & p_T^{\text{lep}} \leq 50 \text{ GeV} \\ \frac{10 \text{ GeV}}{p_T^{\text{lep}}}, & p_T^{\text{lep}} \in (50 \text{ GeV}, 200 \text{ GeV}) \\ 0.05, & p_T^{\text{lep}} \geq 200 \text{ GeV} \end{cases} \quad (2)$$

where p_T^{lep} is the magnitude of the transverse momentum vector of the lepton candidate.

For both regions we have added pileup corrections and lepton scale factors to the weights. We have also implemented luminosity and trigger corrections. Fig. 9 and Fig. 10 show the cut-flow diagram for the two defined regions, where we can see the efficiency of each of the requirement described above for data and simulated samples

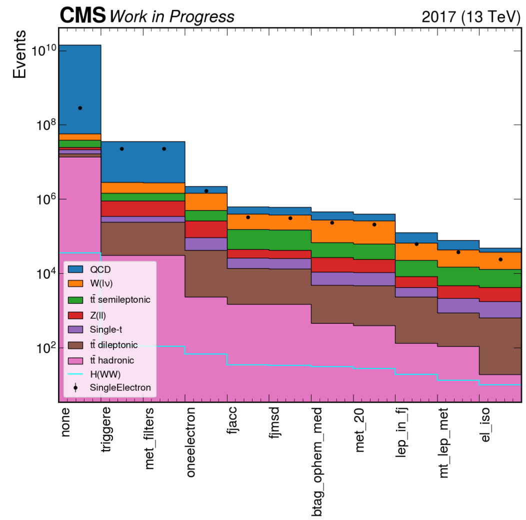

Figure 9 : Cut-flow in the electron control region for 2017 category showing yield sorted sample contributions.

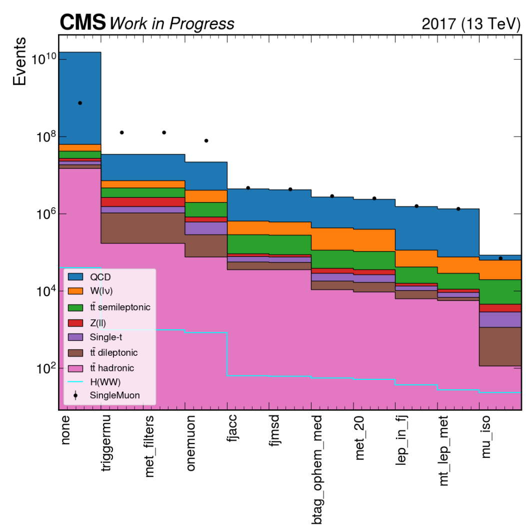

Figure 10 : Cut-flow in the muon control region for 2017 category showing yield sorted sample contributions.

In Fig. 11 and Fig. 12 the Jet p_T, Jet m_{sd}, btagFlavB (opphem) and missing energy p_T^{miss} distributions are shown for data and simulated samples events after all the selection criteria described in Table 3.

In all of these distributions the W($\rightarrow l\nu$) + Jets process dominates, although the multijet QCD and semileptonic $t\bar{t}$ processes also contribute significantly. In both distributions and cut-flow diagrams, we see a disagreement between data and background, particularly in the electron region. This results may be improved by updating the corrections to UL corrections which are slightly more accurate. b-tagging scale factors could be also added to improve further the agreement on both regions.

6 Summary

We presented the optimization of the event selection of the search for the standard model Higgs boson with $p_T > 200$ GeV decaying to W vector boson pairs and reconstructed as a single, large-radius jet. The search uses a data sample of proton-proton collisions at $\sqrt{s} = 13$ TeV corresponding to an integrated luminosity of 41.5 fb^{-1}.

We define two signal regions (muon and electron regions) in which we analyze the p_T and m_{sd} of the candidate jet, as well as the btagFlavB (opphem) and missing energy p_T^{miss}. We find a slight disagreement between data and background, particularly in the electron region.

This is a work in progress in which some corrections need to be added in order to improve further the results presented.

References

[1] Georges Aad et al. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Physics Letters B* 716.1 (2012), pp. 1–29.

[2] Roman Adolphi et al. "The CMS experiment at the CERN LHC". In: *Jinst* 803 (2008), S08004.

[3] T. K. Aarrestad et al. "Identification and calibration of boosted hadronic W/Z bosons at 13 TeV". In: *Technical report* AN-2016/215 (2016).

[4] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. "The anti-kt jet clustering algorithm". In: *Journal of High Energy Physics* 2008.04 (2008), p. 063.

[5] CMS collaboration et al. "A new boson with a mass of 125 GeV observed with the CMS experiment at the Large Hadron Collider". In: *Science* 338.6114 (2012), pp. 1569–1575.

[6] CMS collaboration et al. "Identification of b quark jets at the CMS Experiment in the LHC Run 2". In: (2016).

[7] Daniel de Florian et al. "Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector". In: *arXiv. org* (2016).

[8] Andrew J Larkoski et al. "Soft drop". In: *Journal of High Energy Physics* 2014.5 (2014), pp. 1–46.

[9] Tai Sakuma and Thomas McCauley. "Detector and event visualization with SketchUp at the CMS experiment". In: *Journal of Physics: Conference Series*. Vol. 513. 2. IOP Publishing. 2014, p. 022032.

[10] Albert M Sirunyan et al. "Particle-flow reconstruction and global event description with the CMS detector". In: *Journal of Instrumentation* (2017), p. 86.

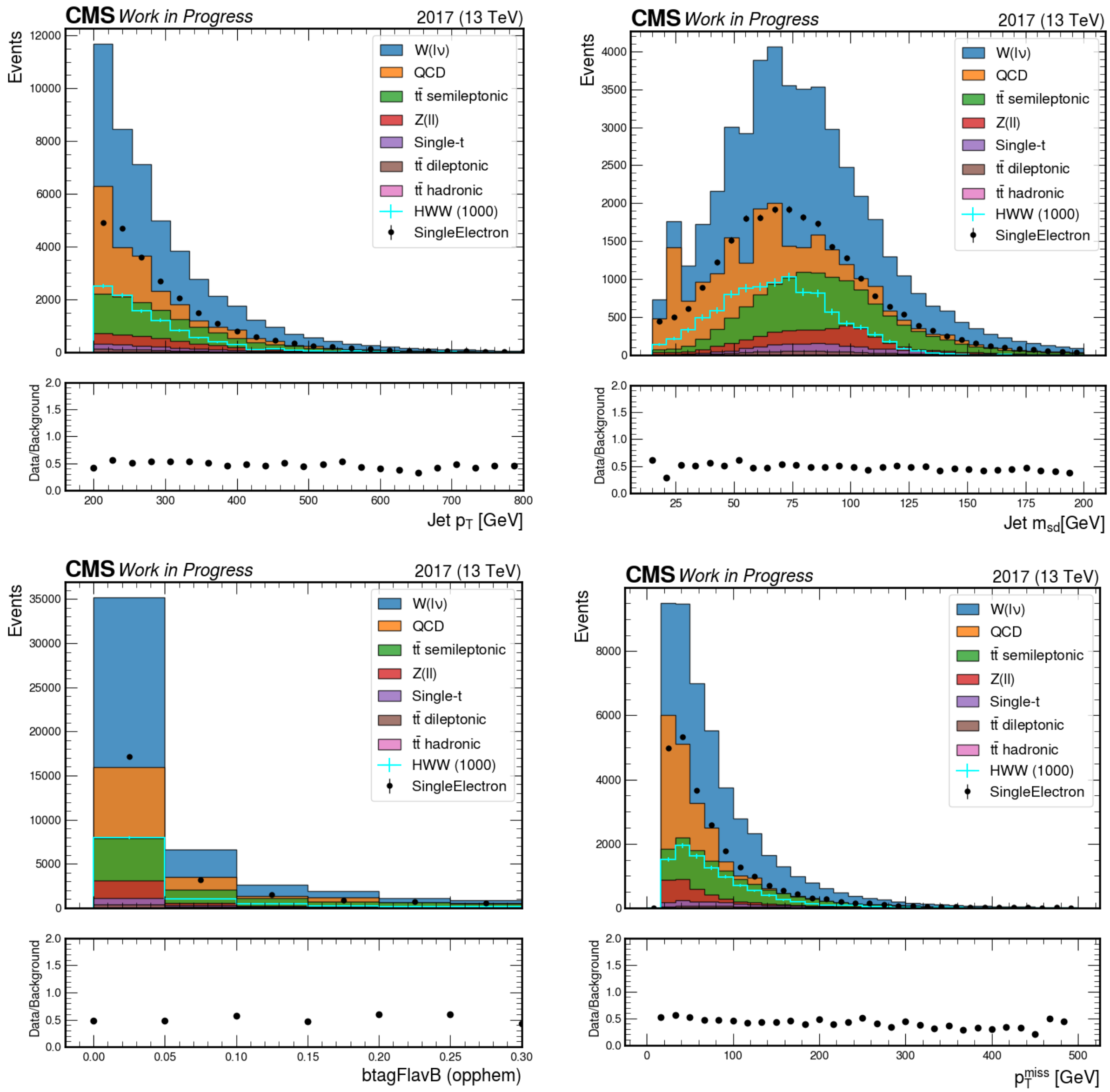

Figure 11 : Jet p_T, Jet m_{sd}, btagFlavB (opphem) and missing energy p_T^{miss} distributions in the electron region.

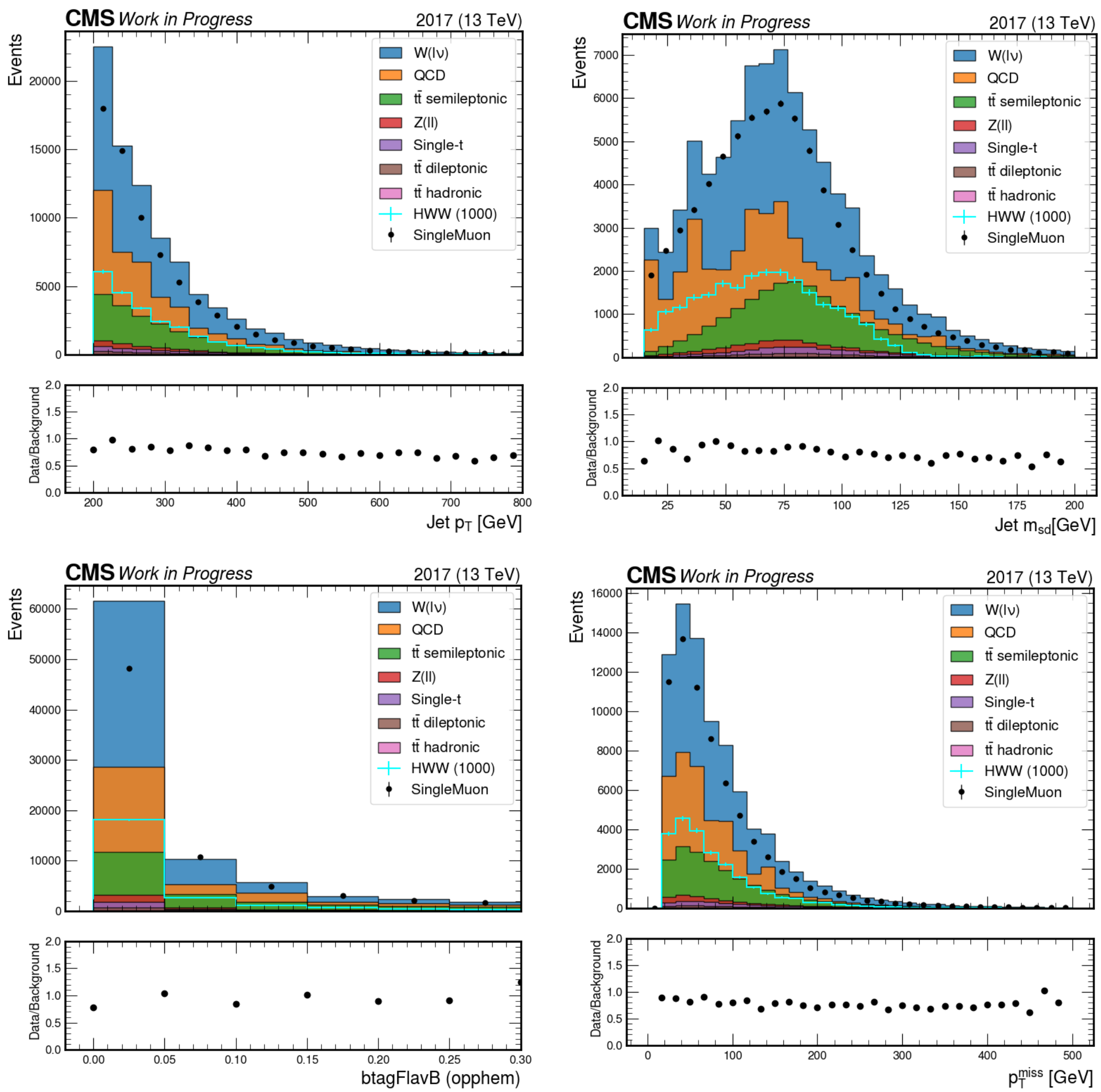

Figure 12 : Jet p_T, Jet m_{sd}, btagFlavB (opphem) and missing energy p_T^{miss} distributions in the the muon region.