CMS Draft Analysis Note

The content of this note is intended for CMS internal use and distribution only

Search for supersymmetry using boosted Higgses and missing transverse momentum in proton-proton collisions at 13 TeV

O. Long[1], J. Bradmiller-Feld[2], J. Richman[2], M. Franco Sevilla[2], A. Ovcharova[2], W. T. Ford[3], F. O. Jensen[3], J. Hirschauer[4], A. Whitbeck[4], C. Vernieri[4], K. A. Ulmer[5], and R. Patel[5]

[1] University of California, Riverside
[2] University of California, Santa Barbara
[3] University of Colorado Boulder
[4] Fermi National Accelerator Laboratory
[5] Texas A&M University

Abstract

A search is presented here for gluinos decaying into Higgs bosons with large missing transverse momentum in proton-proton collisions collected by the CMS experiment at the LHC at $\sqrt{s} = 13$ TeV. We target Higgs bosons produced with large momentum, where the hadronization products of a pair of b-quarks can be reconstructed in single jet. The analysis makes use of an algorithm to identify jets reconstructed from overlapping $b\bar{b}$. The background from multijet and $t\bar{t}$ events is significanctly reduced with requirements on the heavy flavour tagging of the jet and its mass. The data sample corresponds to an integrated luminosity of 35.9 fb^{-1} and was collected in 2016 and 2017.

1 Introduction

This note presents a search for Supersymmetry (SUSY) in the event topology where the final state particles have considerably larger momentum compared to their mass. In particlular the standard model (SM) Higgs bosons decays to b-quarks, allow for explicitly tagging the heavy flavor jets coming from the boson decay. For the lower momentum Higgs decays, the angular seperation between b-quarks will allow them to be resolved in seperate jets (cite manuel), which can then be used to reconstruct the Higgs mass. For significantly large momentum than the Higgs mass ($p_{\mathrm{T}} > 300 GeV$), the b-quarks can merge into a single jet. For these boosted Higgs bosons, a jet with a large cone (R=0.8) can be tagged based on its sub-structure and its jet mass. The composition of the jet is analyzed based on sub-structure variables, the associated tracks, and the vertex information to gauge how likely the jet contains two b-hadrons. The narrow mass window of the Higgs boson allows the jet mass to characterize a jet as coming from the boson decay. This analysis identifies events with two Higgs-tagged and single Higgs-tagged events allowing for SUSY decays with HH and ZH production. The SUSY scenarios here assume R-parity is conserved, so the event topology also has large missing energy from the lightest SUSY particle escaping the detector. The final analysis categories are divided into regions with increasing missing energy to capture the falling spectrum of the background and maximizing seperation between the signal and the background.

2 Event samples

2.1 Standard model MC samples

Monte Carlo (MC) samples reconstructed with CMSSW release 8_0_X (Summer16) are used for all processes. The SM samples are listed in Tables 1,4, 2, and 3. The cross sections listed correspond to next-to-next-to-leading-order (NNLO) calculations unless otherwise noted. All samples use the PU25bx25 pileup scenario, which simulates a pileup distribution with an average of 25 interactions per bunch crossing and a 25 ns interval between bunches. The Monte Carlo samples are used to validate the data-driven background estimation methods.

Table 1: SM $t\bar{t}$ MC samples used in the analysis. The cross sections are calculated to NNLO.

Dataset	σ (pb)	$\int \mathcal{L}\, dt$ (fb^{-1})
TTJets_SingleLeptFromT_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	182.72	283.90
TTJets_SingleLeptFromTbar_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	182.72	326.48
TTJets_DiLept_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	88.34	346.25
TTJets_HT-600to800_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	2.734	5231.81
TTJets_HT-800to1200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	1.121	9416.61
TTJets_HT-1200to2500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	0.198	14819.34
TTJets_HT-2500toInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	0.002	221088.29

Table 2: SM QCD MC samples used in the analysis. All cross sections are calculated to LO.

| Dataset	σ (pb)	$\int \mathcal{L}\, dt$ (fb^{-1})
QCD_HT200to300_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	1735000	0.03
QCD_HT300to500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	366800	0.16
QCD_HT500to700_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	29370	1.95
QCD_HT700to1000_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	6524	6.68
QCD_HT1000to1500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	1064	12.62
QCD_HT1500to2000_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	121.5	32.63
QCD_HT2000toInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	25.42	239.30

27 3 Signal models
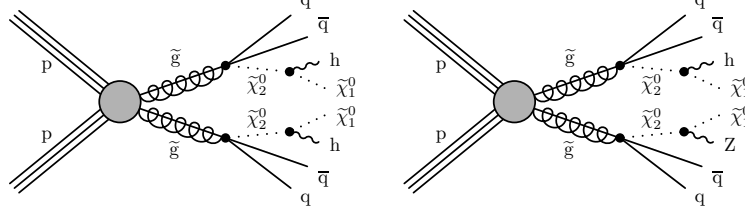
Figure 1: Signal diagrams for the boosted Higgs search via gluino strong production. We consider 100% branching fraction to the Higgs boson and also a mixed state where there is a 50% branching fraction for Higgs and Z-bosons.

28 Fig. 1 shows the signal diagrams for the interpretations considered in this analysis. The mass
29 splitting between $\tilde{g}$ and $\tilde{\chi}_2^0$ is fixed to 50 GeV which results in the quarks to be quite soft from the
30 gluino decay and the mass of the $\tilde{\chi}_1^0$ is fixed to 1 GeV so that the Higgs p_T is directly proportional
31 to $m_{\tilde{g}}$. The signal regions for this analysis are designed to be sensitive for purely Higgs in the
32 final state with $H \rightarrow b\bar{b}$ or a mixed state of Higgs and Z in the final state with a relative
33 branching fraction. Fig. 2 shows the Higgs p_T for various $m_{\tilde{g}}$ model points. For high gluino
34 mass, the b-quarks from $H \rightarrow b\bar{b}$ can be contained in a large jet cone with $\Delta R < 0.8$. Events are
generated with the Full Simulation using the reconstruction in CMSSW 8_0_X.
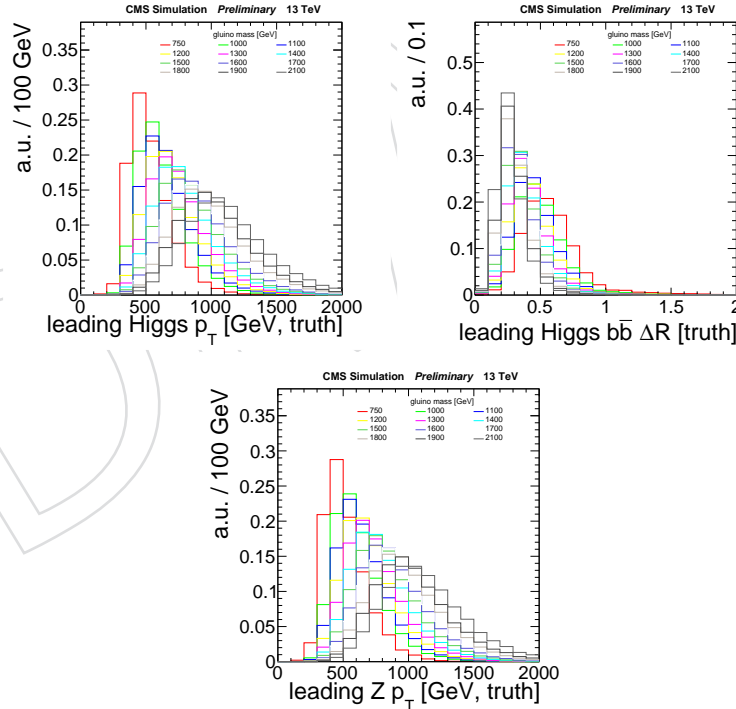

Figure 2: The generator-level p_T of the leading Higgs boson for a range of $m_{\tilde{g}}$. For $m_{\tilde{g}} > 1000 GeV$, the Higgs p_T is well above 300 GeV. In this Higgs p_T range the angle between daughter b-quarks is small so that they are merged into a single jet cone. The Z-boson has a similar boost for the same range of $m_{\tilde{g}}$.

35

Figure 3: Single muon trigger efficiencies measured relative to the inclusive E_T^{miss} triggers in events with in events with exactly 1 reconstructed muon.
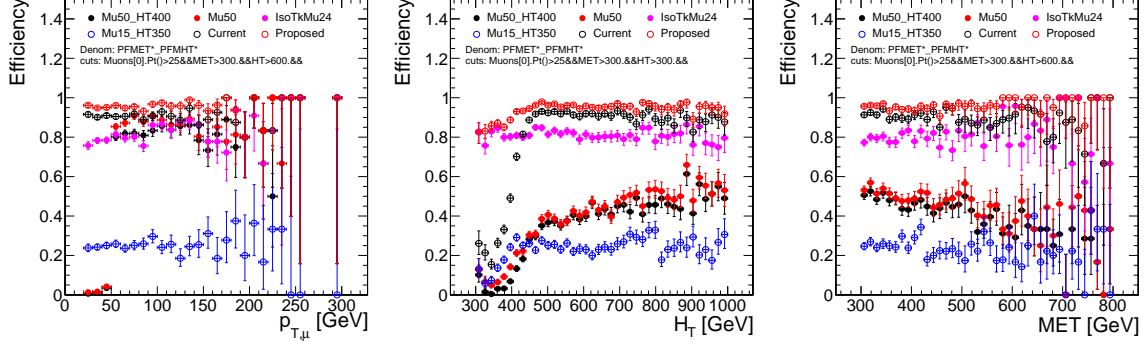
4 Triggers

4.1 Signal region

Events in the search region are collected by the logical OR of the following triggers:

- HLT_PFMET100_PFMHT100_IDTight, HLT_PFMETNoMu100_PFMHTNoMu100_IDTight
- HLT_PFMET110_PFMHT110_IDTight, HLT_PFMETNoMu110_PFMHTNoMu110_IDTight
- HLT_PFMET120_PFMHT120_IDTight, HLT_PFMETNoMu120_PFMHTNoMu120_IDTight

4.2 Single-μ region

Events in the single-μ validation region are triggered with the logical OR of several muon and muon-HT triggers:

- HLT_IsoTkMu24, HLT_Mu50
- HLT_Mu15_IsoVVVL_PFHT350_PFMET50, HLT_Mu15_IsoVVVL_PFHTXXX
- HLT_Mu50_IsoVVVL_PFHT400,

where XXX can refer to PFHT thresholds of 350, 400, or 600.

To measure the efficiency of these triggers in data, the inclusive E_T^{miss} triggers that are used for the signal region, Sec. 4.1, are used as a set of reference triggers. Figure 3 shows the efficiency of the single muon triggers versus $p_{T\,\mu}$, H_T, and E_T^{miss}. These efficiencies are found to be consistent with those derived in [], which are based on a single muon reference trigger; we apply those efficiencies corresponding to the single muon reference trigger to the MC for all studies based on the single muon validation region.

4.3 Single-e region

Events in the single-e validation region are triggered with a logical OR of the following triggers:

- HLT_Ele15_IsoVVVL_PFHTXXX
- HLT_Ele45_WPLoose_Gsf, HLT_Ele50_IsoVVVL_PFHT400
- HLT_Ele27_eta2p1_WPLoose_Gsf, HLT_Ele25_eta2p1_WPTight_Gsf
- HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ

The efficiency for these triggers is measured using the inclusive E_T^{miss} triggers discussed in Section 4.1. Figure 4 shows the efficiency versus $p_{T,e}$, H_T, and E_T^{miss}. Table 5 shows the efficiency

Table 3: SM $Z \rightarrow \nu\nu$+jets MC samples used in the analysis. The cross sections are calculated to NNLO.

Dataset	σ (pb)	$\int \mathcal{L}\,\mathrm{d}t$ (fb^{-1})
ZJetsToNuNu_HT-100To200_13TeV-madgraph	344.8	54.13
ZJetsToNuNu_HT-200To400_13TeV-madgraph	95.53	208.46
ZJetsToNuNu_HT-400To600_13TeV-madgraph	13.20	77.30
ZJetsToNuNu_HT-600To800_13TeV-madgraph	3.148	1795.26
ZJetsToNuNu_HT-800To1200_13TeV-madgraph	1.451	1486.09
ZJetsToNuNu_HT-1200To2500_13TeV-madgraph	0.355	1029.81
ZJetsToNuNu_HT-2500ToInf_13TeV-madgraph	0.0085	47498.87

Table 4: SM $W \rightarrow \ell\nu$+jets MC samples used in the analysis. The cross sections are calculated to NNLO.

Dataset	σ (pb)	$\int \mathcal{L}\,\mathrm{d}t$ (fb^{-1})
WJetsToLNu_HT-100To200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	1627.45	18.16
WJetsToLNu_HT-200To400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	435.24	45.88
WJetsToLNu_HT-400To600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	59.18	123.64
WJetsToLNu_HT-600To800_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	14.58	221.32
WJetsToLNu_HT-800To1200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	6.66	1123.13
WJetsToLNu_HT-1200To2500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	1.608	153.44
WJetsToLNu_HT-2500ToInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	0.039	6497.28

Figure 4: Single electron trigger efficiencies measured relative to the inclusive E_T^{miss} triggers in events with in events with exactly 1 reconstructed electron.
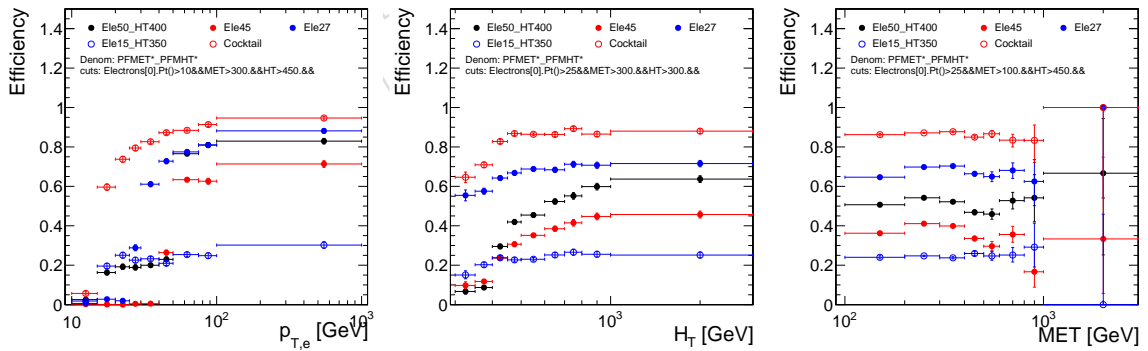

Table 5: Single electron cocktail trigger efficiency parameterization.

e trig. ϵ [%]	25< p_T <30	30< p_T <40	40< p_T <50	50< p_T <75	75< p_T <100	p_T >100
300< H_T <450	0.794	0.826	0.872	0.884	0.913	0.947
H_T >450	0.572	0.775	0.858	0.861	0.932	1.

Figure 5: Single photon trigger efficiencies measured relative to the single photon triggers in events with in events with exactly 1 reconstructed photon with $p_T > 100$ GeV.
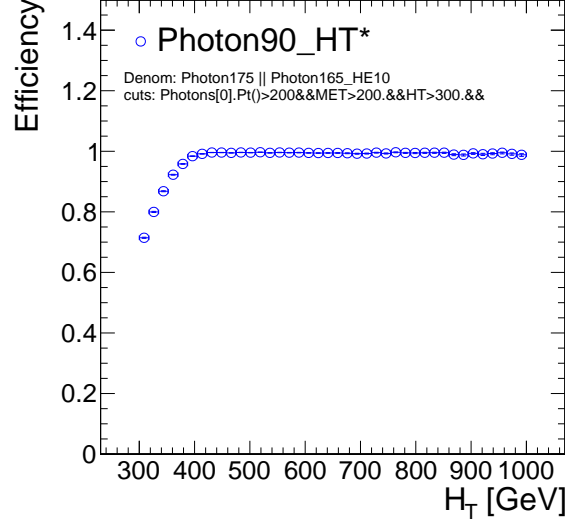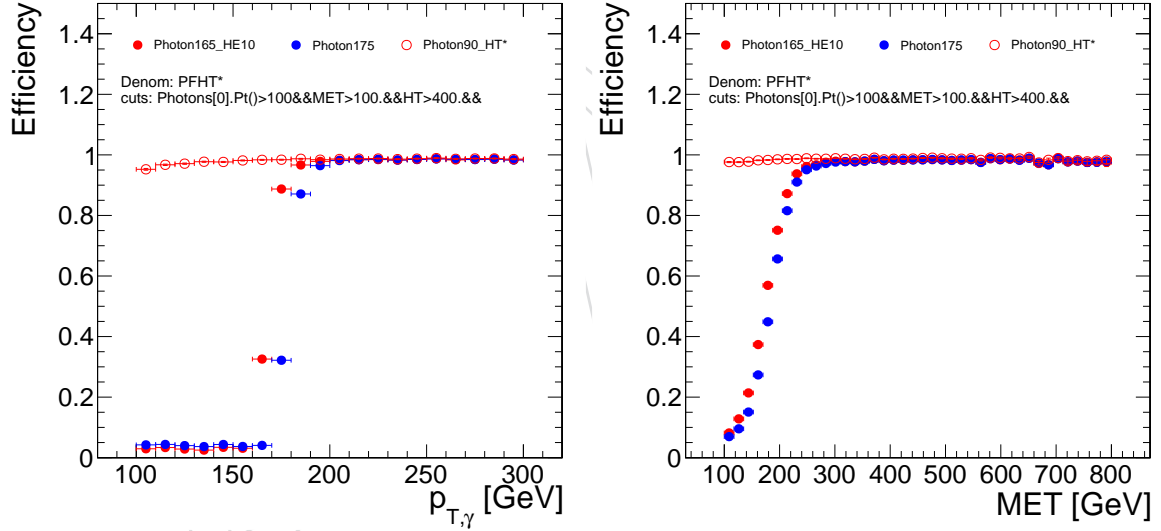
Figure 6: Single photon trigger efficiencies measured relative to the inclusive H_T triggers in events with in events with exactly 1 reconstructed photon with $p_T > 100$ GeV.

parameterization that is applied to the MC for all studies related to the single-e validation region.

4.4   Photon region

Events in the photon validation region are triggered with the `HLT_Photon90_CaloIdL_PFHTXXX`, where `XXX` refers the PFHT thresholds of 500 or 600. For these triggers, the photon's energy is part of the PFHT energy sum, while for the photon validation region, the photon's energy is removed from the event before computing jet-level and event-level variables, such as H_T or E_T^{miss}. As a result the turn-on for events with 100 GeV photons us reduced.

To measure the H_T dependence of the photon trigger efficiency, an OR of the `HLT_Photon165_HE10` and `HLT_Photon175` triggers are used as reference. Figure 5 shows the efficiency versus H_T.

Figure 7: Low-$\Delta\phi$ trigger efficiencies measured relative to the inclusive H_T triggers in events with in events passing the low-$\Delta\phi$ selection.
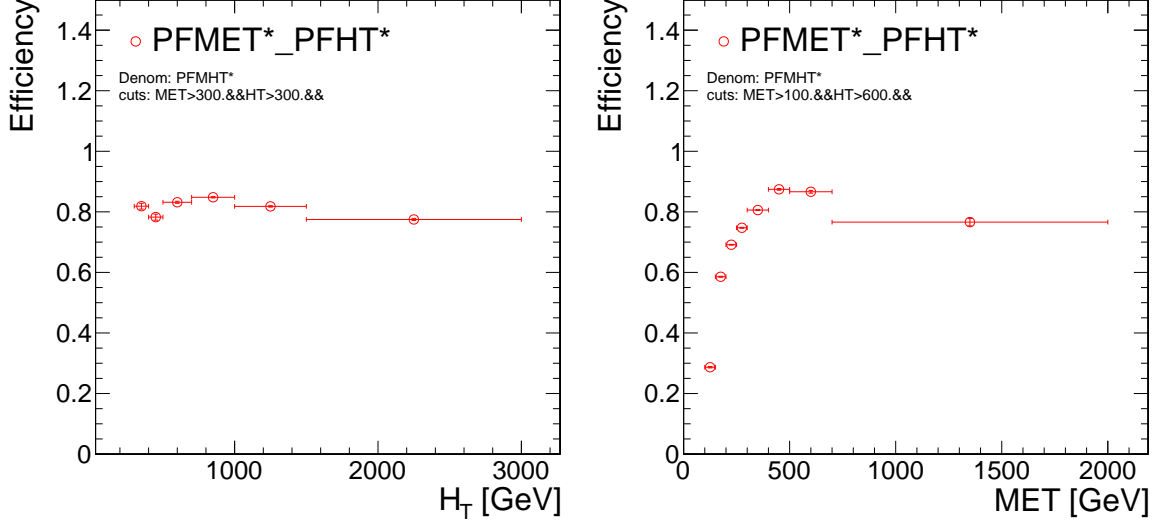

Table 6: Low-$\Delta\phi$ trigger efficiencies.

low $\Delta\phi$ ϵ [%]	$100 < E_\mathrm{T}^{\mathrm{miss}} < 200$	$200 < E_\mathrm{T}^{\mathrm{miss}} < 300$	$300 < E_\mathrm{T}^{\mathrm{miss}} < 400$
	0.500	0.712	0.806
	$400 < E_\mathrm{T}^{\mathrm{miss}} < 500$	$500 < E_\mathrm{T}^{\mathrm{miss}} < 700$	$E_\mathrm{T}^{\mathrm{miss}} > 700$
	0.874	0.866	0.766

The turn-on is found to plateau at nearly 100% above $H_\mathrm{T} > 400$ GeV.

To measure the p_T and $E_\mathrm{T}^{\mathrm{miss}}$ dependence of the photon trigger efficiency, a suite of `HLT_PFHTXXX` reference triggers are used. The PFHT tresholds vary from 300 to 900, each with varying degree of prescales. Figure 5 shows the efficiency, as measured relative to the inclusive HT triggers versus $p_{T,\gamma}$ and $E_\mathrm{T}^{\mathrm{miss}}$ for events with $H_\mathrm{T} > 400$ GeV.

4.5 Low-$\Delta\phi$ region

Events in the low-$\Delta\phi$ validation region are triggered using the same suite of MET-MHT triggers as used for the signal region. However, because of the different detector response for fake-$E_\mathrm{T}^{\mathrm{miss}}$ events, the average efficiency in the low-$\Delta\phi$ control region, which is dominated by fake-$E_\mathrm{T}^{\mathrm{miss}}$, is lower than that of the signal region, which is dominated by genuine-$E_\mathrm{T}^{\mathrm{miss}}$, i.e. $t\bar{t}$ and W+jets.

To measure the efficiency of the low-$\Delta\phi$ triggers, the same set of PFHT reference triggers described in Section **??** are used. Figure 7 shows the measured efficiency versus H_T and $E_\mathrm{T}^{\mathrm{miss}}$. Table 6 shows the efficiency parameterization that is applied to MC events for all studies related to the low-$\Delta\phi$ validation region.

5 Event selection

The search regions for the analysis require large $E_\mathrm{T}^{\mathrm{miss}}$, large H_T, and no leptons. The selection makes use of the same kinematic variables as more inclusive SUSY analyses [1]. In addition, selection is applied to the large jet cones to ensure a boosted topology and the compositions of jets to have two b-quarks. Jets used in this analysis are reconstructed from particle-flow

92 (PF) candidates using the anti-k_T algorithm [2] with the size parameter 0.8. The PF algorithm
93 is used to individually identify and reconstruct all particles produced in the collision (PF can-
94 didates); namely charged hadrons, photons, neutral hadrons, muons, and electrons [3]. The
95 large jet cones are reclustered from their original jet constituents, and the clustering sequence
96 is modified to remove soft and wide-angle particles or groups of particles. This "pruned jet" is
97 used to compute the mass after removing the soft radiation to provide a narrower Higgs mass
98 window [4]. The analysis also uses anti-k_T jets with the size parameter 0.4 to compute the H_T
99 and $\Delta\phi(j_i)$ variables.

Cut	Total Bkg.	$Z \to \nu\nu$	$W \to \ell\nu$	$t\bar{t}$	QCD	T5HH1200	T5HH1600
Pre-Selection for Skim	119215	76939.1	30973.7	6840.16	709.29	1391.31	150.09
$HT > 600\,\text{GeV}\,E_{\text{T}}^{\text{miss}} > 300\,\text{GeV}$	26924.7	16376.7	6632.57	2296.66	468.36	1128.75	122.76
2 AK8 Jets	18128.7	10950.9	4685.7	1313.06	421.84	962.08	107.27
AK8 Jet $p_{\text{T}} > 300\,\text{GeV}$	6588.63	4143.35	1715.54	296.18	199.70	738.0	91.57
Jet Mass in [50, 250] GeV	847.72	475.11	205.90	89.84	21.39	446.22	55.64
AK8 Jets Double-b	14.42	3.70	1.59	5.49	0.36	109.07	14.02
Jet Mass in [85, 135] GeV	1.71	0.51	0.26	0.59	0.055	70.47	9.64

Table 7: Cut Flow table on MC for each of the main background samples and the sum all the background components. The table also includes the signal for the 2-Higgs final state.

The following requirements define the baseline selection:

- $H_T > 600\,\text{GeV}$, where $H_T = \sum_{\text{jets}} p_T$

- $E_T^{\text{miss}} > 300\,\text{GeV}$ where $E_T^{\text{miss}} = \left| \sum_{\text{pfpart}} \vec{p}_T \right|$.

- Angular cut: The majority of QCD multijet events in our high-E_T^{miss} search region have jets with undermeasured momenta and thus a spurious momentum imbalance. A signature of such an event is a jet closely aligned in direction with the E_T^{miss} vector. To suppress this background, we reject all events in which the two highest-p_T jets lie within 0.5 radians of the E_T^{miss} vector in the azimuthal coordinate. The third and fourth highest-p_T jets are rejected if they lie within 0.3 radians.

- Muon veto:

  Muon candidates are selected using the POG-recommended "Medium Muon" selection [5] with the additional requirements:

$$d_{xy}(\mu, \text{PV}) < 0.2 \text{ mm}$$
$$d_z(\mu, \text{PV}) < 0.5 \text{ mm} \tag{1}$$

  Muon candidates are required to have $p_T > 10\,\text{GeV}$ and $|\eta| < 2.4$. To distinguish between prompt muons and muons from b-hadron decays, muons are required to satisfy an isolation requirement, $I_{\text{mini}} < 0.2$, where I_{mini} is the mini-isolation variable described in Ref. [6]. Any event with a muon satisfying all of the above criteria is vetoed.

- Electron veto:

  Electron candidates are selected using the POG-recommended "Cut Based VETO" selection [7]. Electron candidates are required to have $p_T > 10\,\text{GeV}$ and $|\eta| < 2.5$. Electron candidates are required to satisfy an isolation requirement of $I_{\text{mini}} < 0.1$. Any event with an electron satisfying all of the above criteria is vetoed.

- Isolated track vetoes: Following the event selection described above, including the muon and electron event vetoes, the dominant source of background for many of the search regions is $t\bar{t}$, single-top, and W+jets events with one $W \to \ell\nu$ decay. In about half these background events, the W boson decays to a τ lepton and the τ lepton decays hadronically, while in the other half, an electron or muon is not identified or does not satisfy the criteria for an isolated electron or muon candidate given above. To suppress these backgrounds, we reject events with one or more isolated charged track.

  The requirements for the definition of an isolated track differ slightly depending on whether the track is identified as leptonic or hadronic by the PF algorithm. For leptonic tracks, we require:

  - $p_T > 5\,\text{GeV}$,
  - $I_{\text{tk}} < 0.2$,

  where I_{tk} is the scalar p_T sum of other charged tracks within $\Delta R \equiv \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} < 0.3$ of the primary track, divided by the p_T value of the primary track. For hadronic tracks, we apply slightly tighter requirements to reduce hadronic (non-τ) signal loss:

  - $p_T > 10\,\text{GeV}$,
  - $I_{\text{tk}} < 0.1$.

  To retain more signal and thus to improve the signal-to-background event discrimi-

nation, isolated tracks are considered only if they satisfy

$$m_T(\text{tk}, E_T^{\text{miss}}) = \sqrt{2p_T^{\text{tk}} E_T^{\text{miss}}(1 - \cos \Delta\phi)} < 100 \text{ GeV}, \quad (2)$$

where p_T^{tk} is the transverse momentum of the track and $\Delta\phi$ is the azimuthal separation between the track and $\vec{p}_T^{\text{miss}}$.

To reduce the influence of tracks from extraneous pp interactions (pileup), isolated tracks are considered only if their nearest distance of approach along the beam axis to a reconstructed vertex is smaller for the primary event vertex than for any other vertex.

- Event cleaning:

  We reject events with a jet that satisfies $p_T > 30$ GeV and $|\eta| < 5$ if the jet fails the loose jet ID criteria given above.

  We apply event filters designed by various POGs to reject events with spurious E_T^{miss} signals. The current list includes:

    - globalTightHalo2016Filter
    - HBHENoiseFilter
    - HBHEIsoNoiseFilter
    - eeBadScFilter
    - EcalDeadCellTriggerPrimitiveFilter
    - BadChargedHadronFilter
    - BadMuonFilter
    - Good vertex filter (requiring at least one reconstructed vertex satisfying !isFake && $N_{\text{dof}} > 4$ && $|z| < 24$ && $\rho < 2$)

 To protect against particle flow failures, events are rejected if PFMET/CaloMET > 5.

  We apply the following "muon jet filter" to reject events with misreconstructed muons:

    - Veto events if any jet in the event has $p_T > 200$ GeV, muon energy fraction > 0.5, and $\Delta\phi(\text{jet}, E_T^{\text{miss}}) > \pi - 0.4$.

- To ensure events with a boosted topology, we select events based on high p_T AK8 jets with the following criteria:

    - Require the event to have at least two AK8 jets with $p_T > 300 GeV$
    - The pruned mass of the two highest p_T jets are required to be between 50 and 250 GeV

These cuts constitute the baseline phase space for the analysis in which all of the control regions and signal regions are defined. Events are categorized in using jet propoerties of the leading and sub-leading, ordered by p_T, AK8 jet: the double-b discriminant, $\mathcal{D}_{bb}$, described in more detail in Section 5.1, and the pruned-jet mass. Events are considered double tagged if both the two highest p_T jets satisfy $\mathcal{D}_{bb} > 0.3$. Events are considered single tagged if either one of the two highest p_T jets satisfy $\mathcal{D}_{bb} > 0.3$. All other events are considered anti-tagged. For each of these three regions, a signal mass window (SR) and sideband (SB) are defined and used as control regions. Events are considered to be within the signal mass window if the pruned mass of the two highest p_T jets must be within the mass window [85, 135] GeV for all tagged regions, else they are in the sideband.

The selection described above has the effect of selecting events with one or more candidate

Figure 8: Top: Jet p_T distribution after all baseline selection except the jet mass and p_T requirements. Bottom: Jet mass distributions after all baseline selection except the jet mass requirements.
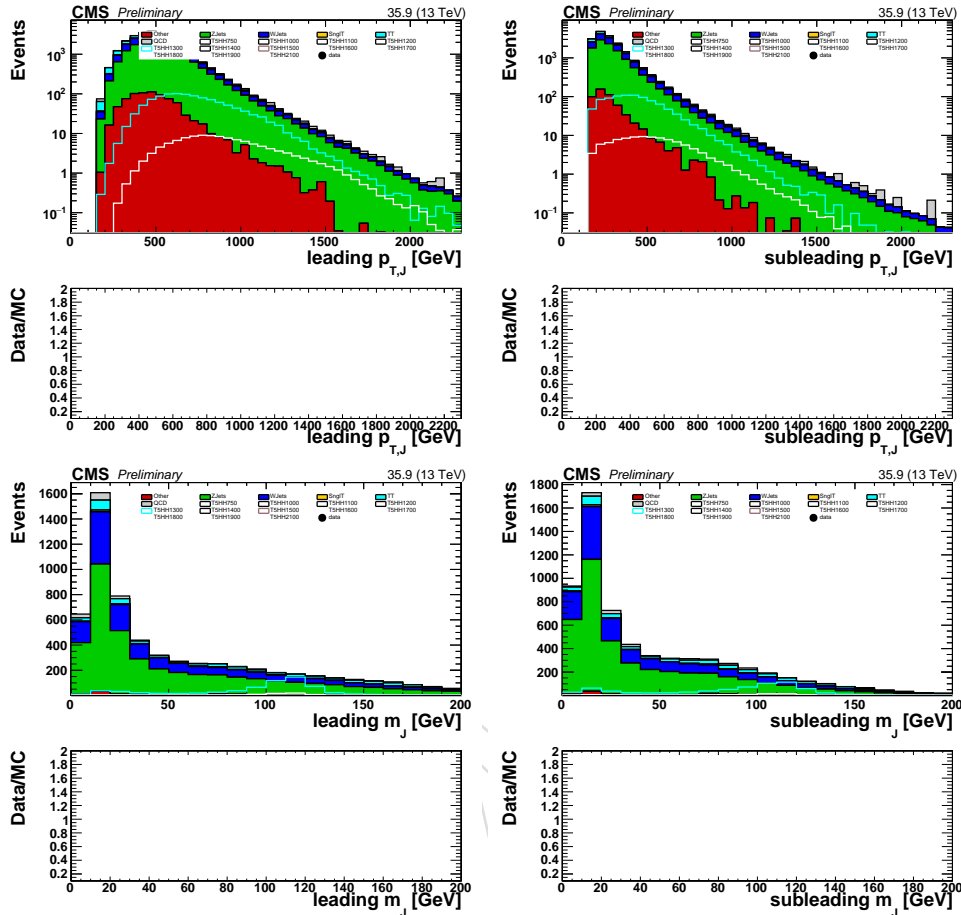
boosted objects. The dominant background for high E_T^{miss}, all hadronic events typically involved either fake MET, from QCD multijet events, or leptonic decays of weak vector bosons, which produce neutrinos. In either case, hadronically decaying boosted objects are not copiously producd in these process, which the exception of semileptonic $t\bar{t}$. Figure 8 shows the expected distributions of jet p_T for SM backgrounds p_T and mass requirements and the pruned-jet mass distributions after before jet mass requirements. Figure 9 shows the expected MET distribution for signal and background processes after baseline selection, single-tagged signal reguirements, and double-tagged signal requiremnts. As can be seen in Figure 9 and Table **??**, the signal region requirements are very effective at reducing the SM background yields to $\mathcal{O}(1)$ event with a signal efficiency of $\mathcal{O}(10\%)$.

5.1 **Double-b Tagger**

A novel approach has been studied to identify large cone jets composed of two b-quarks [8]. A dedicated multivariate (MVA) tagging algorithm is implemented to combine the information from the secondary vertices, the tracks, and the τ-axes of the jet substructure and optimize the discimination between jets from two b-hadron decays and single-parton jets. To generalize the algorithm so that it does does not depend on the mass of the b-parents or the boost, the input variables are chosen so that they do not depend strongly on jet mass or jet p_T. All tracks with $p_T > 1\,\text{GeV}$ are associated to jets in a cone $\Delta R < 0.8$ around the jet axis, which is defined by

Figure 9: $E_\mathrm{T}^\mathrm{miss}$ distribution after all baseline requirements.
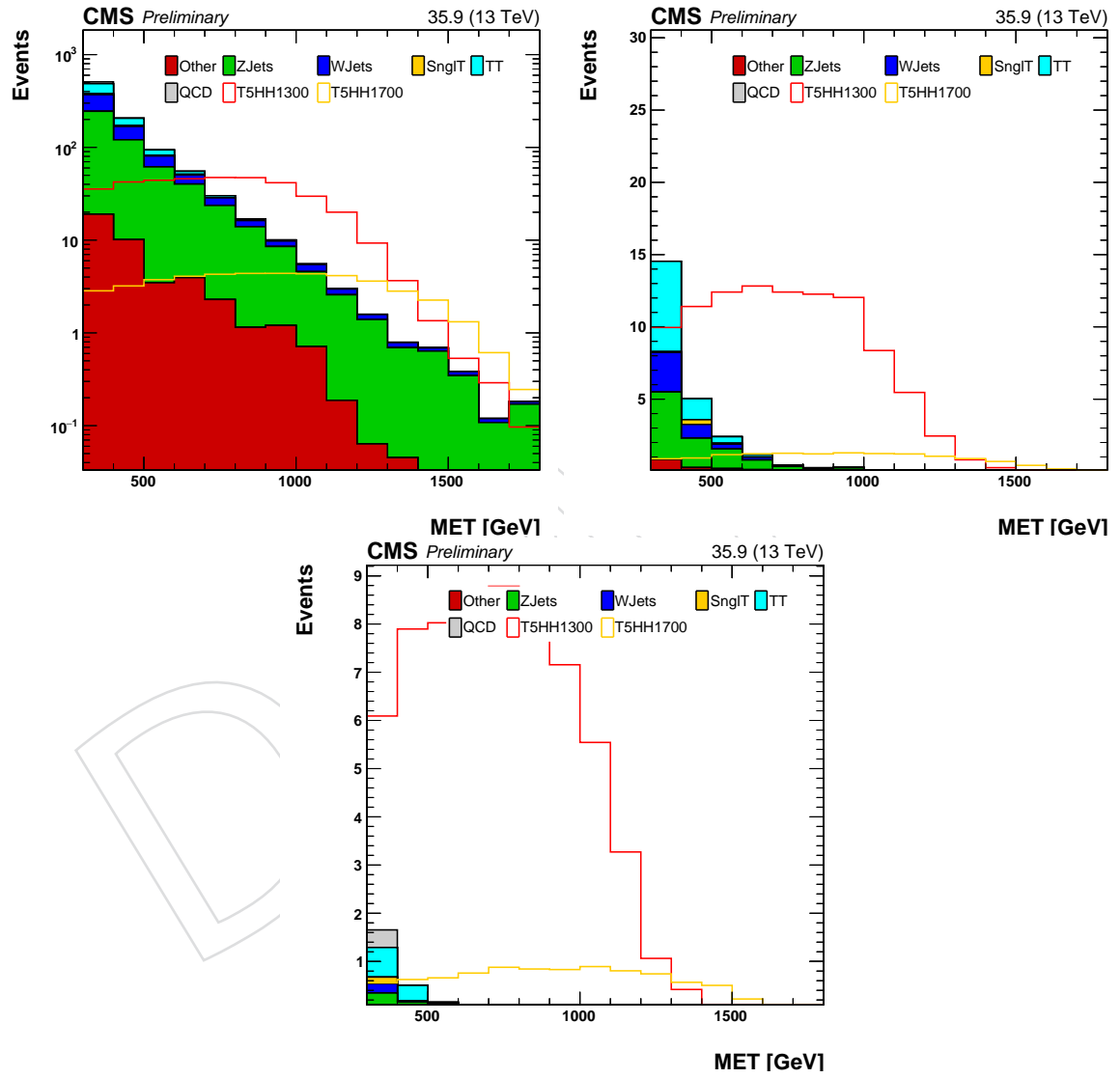
the primary vertex and the direction of the jet momentum. The tracks are then associated to the closest τ-axes. The distance between the sub-structure axis and the track is the distance of closest approach and it is required to be within 700 μm and the point of closest approach on the track is required to be 5cm from the primary vertex. The impact parameter significance (w.r.t to the primary vertex) is used to discriminate tracks from the b-decay from prompt tracks. Several input variables make use of the secondary vertices wich are reconstructed using the Inclusive Vertex Finder, which identifies seocndary vertices independantly of the jet clustering. The secondary vertices are then associated to the closest sub-structure axis. For each τ-axis, track momenta of the constituents tracks from all the secondary vertices associated to a given τ-axis are added to compute the SV mass and transverse momementum.

The input variables to the double-b tagger MVA discriminate are in [8].
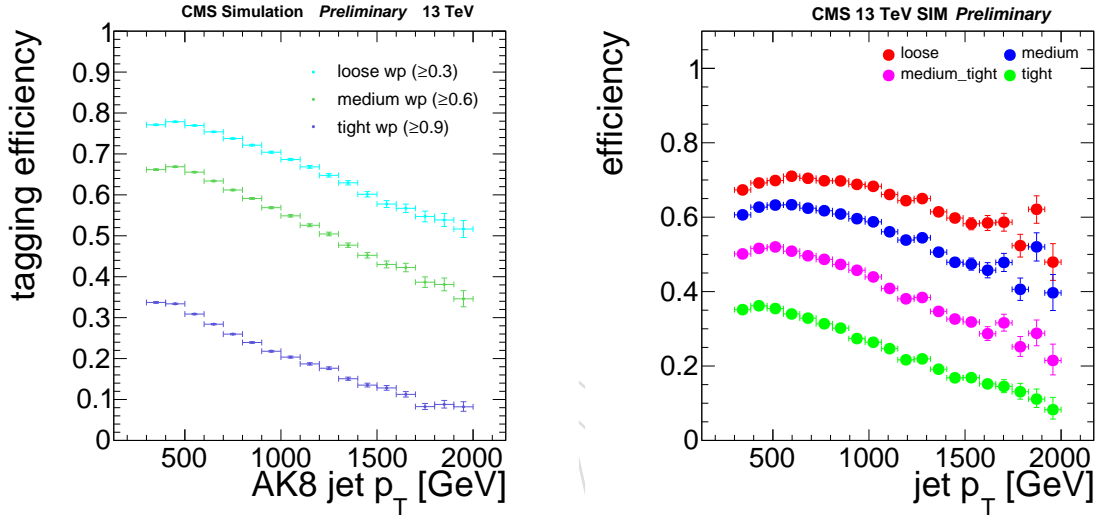

Figure 10: Efficiency of a jet from $H \rightarrow b\bar{b}$ to pass a Double-b discriminator cut at a given working point(left) and the efficiency for the jet to pass both the Double-b discriminator cut and have pruned jet mass within [85,135] GeV (right). We use the loosest available working point corresponding to a cut value of 0.3.

Fig. 10 shows the signal efficiency for an $H \rightarrow b\bar{b}$ jet to pass the Double-b tag cut. The efficiency is above 80% for only the Double-b discriminator and it drops by about 15% when applying the pruned mass cut. At high jet p_T there is an inefficiency in the Double-b tag as tracks from the b-quarks become more collimated. The QCD background consists of some mistagged light flavor jets and jets with true b-hadrons which can come from gluon-splitting. The $t\bar{t}$ background has both true heavy flavor jets and true E_T^{miss}, but a good number of the single b-hadron jets can be rejected with the Double-b tag requirement. Fig. 11 shows the discrimination between between signal $H \rightarrow b\bar{b}$ and background jets based on the true number of b-hadrons.

The two signal models described in Sec 2 are targetted by defining a single Higgs tag and a two Higgs tag. The single Higgs region requires both jets to be within the mass window of $[85, 135]$ GeV but only one of them to be Double-b tagged. This allows to capture the Z decays to light flavor quarks in the T5HZ SMS model. For the two Higgs tag region, both jets are in the mass window $[85, 135]$ GeV and both are double-b tagged. These signal regions are designed to supress as much of the background as possible while preserving efficiency for the T5ZH and T5HH signal models. Figure 12 shows the signal efficiency for each signal point vs. the
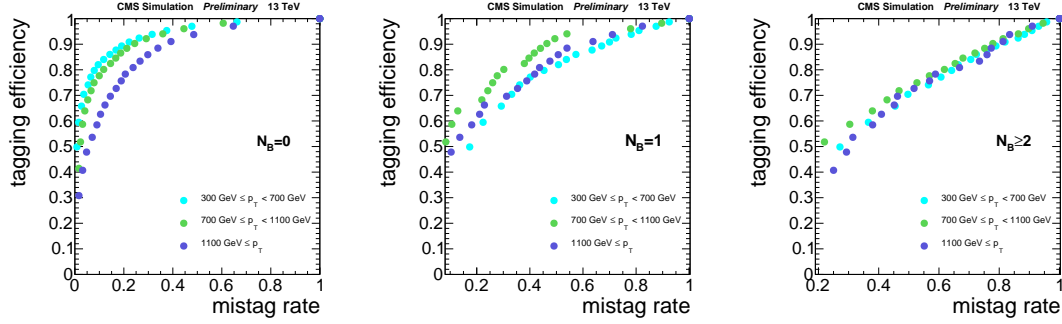
224   Double-b discriminator.

Figure 11: The double-b tagger tag vs mis-tag efficiency is shown for all backgrounds based on the true b-hadron content of the jets.

225   The efficiency of the double-b tagger is measured in a data sample consisting of high p_T jets
226   enriched in $b\bar{b}$ from gluon-splitting. To select a boosted topology similar to the signal, the AK8
227   jet p_T is required to be above 300 GeV and the AK8 pruned mass is required to be above 50 GeV.
228   The jet must also be matched to at least two muons with p_T ¿7 GeV and within $\Delta R < 0.4$ of the
229   sub-jet axis so that the jet can be double-muon tagged". The comparison between data and MC
230   is compared to give a data/MC scale factor which is fairly close to unity. The mis-tag rate is
231   evaluated by comparing data and MC for top-quark jets faking H jets in $t\bar{t}$ production. The
232   studies are based on single lepton $t\bar{t}$ events and the event selection requires one isolated muon
233   with p_T ¿50GeV and an AK4 jet in the same hemisphere.
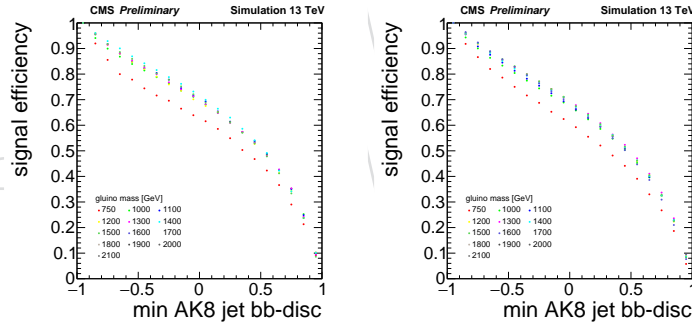

Figure 12: The signal efficiency for the two signal models considered in this analysis: (left) Higgs and Z boson final state (right) Two Higgs bosons in the final state

234   6   **Background Estimation Method**

235   The background is estimated with a simple ABCD method, using the background dominated
236   region where the lead and sub-lead AK8 jet are below the Double-b discrimantor cut (Anti-
237   tagged events) and a sideband in pruned AK8 jet mass.

238   The "Two Higgs-Tagged" signal region is defined as:

239   • Lead and sub-lead AK8 Jet above Double-b discrimantor cut of 0.3

240   • Lead Jet pruned mass within [85, 135] GeV and sub-lead jet pruned mass within [85,
241   135] GeV

Also included in the signal region is the "Single Higgs-Tagged" region which has only one Higgs boson with $H \rightarrow b\bar{b}$:

- Lead AK8 Jet above Double-b discrimantor cut of 0.3 or sub-lead AK8 Jet above Double-b discrimantor cut of 0.3 and the other jet is below 0.3

- Both Lead and sub-lead AK8 jets have pruned mass within $[85, 135]$ GeV

The mass sideband region is the larger window with:

- Lead Jet pruned mass with $[50, 250]$ GeV excluding $[85, 135]$ GeV or sub-lead pruned mass with $[50, 250]$ GeV excluding $[85, 135]$ GeV

The mass sideband is then divided into a Double-b tagged and Double-b anti-tagged regions:

- The Double-b tagged region has both lead and sub-lead AK8 Jet above Double-b discrimantor cut of 0.3 for the Two Higgs Tag region

- The Double-b tagged region has either lead or sub-lead AK8 Jet above Double-b discrimantor cut of 0.3 for the sideband corresponding to the Single Higgs Tag region

- The Double-b anti-tagged region has both lead and sub-lead AK8 Jet below Double-b discrimantor the cut of 0.3

The full set of Signal and background control regions are then:

- A1: Signal Region where the lead jet and sub-lead jet is in the Higgs mass window $[85, 135]$ GeV and one of the jets has Double-b discrimantor cut > 0.3

- A2: Signal Region where the lead jet and sub-lead jet are in the Higgs mass window $[85, 135]$ GeV and both jets have Double-b discrimantor cut > 0.3

- B: Anti-tagged region where both jets have Double-b discrimantor cut < 0.3 and one of the jets is within $[85, 135]$ GeV

- C1: 1 Higgs sideband region where either the lead or sub-lead jet has Double-b discrimantor cut > 0.3 and either jet has pruned mass within $[50, 250]$ GeV excluding $[85, 135]$ GeV

- C2: 2 Higgs sideband region where both jets have Double-b discrimantor cut > 0.3 and either lead or sub-lead jet has pruned mass within $[50, 250]$ GeV excluding $[85, 135]$ GeV

- D: Sideband region where both jets have Double-b discrimantor cut < 0.3 and either the lead or sub-lead jet has pruned mass within Higgs Mass window $[50, 250]$ GeV excluding $[85, 135]$ GeV

The method assumes the Double-b tagged to Anti-Double-b tag ratio is the same in the Higgs mass window and the sideband. The background prediction in the signal region is then given by $A = (C/D) B$. The signal and background in each of the regions determined simultaneously in the likelihood fit, where signal contamination in the control regions is included in the fit. The major source of uncertainty on the background are the available statistics in each of the background control regions which are modeled as Poisson observables in the likelihood.

6.1 Background Systematics

The systematic uncertainty on the background estimimation relies on the independence of the Double-b tagged to Anti-Double-b tag ratio and the pruned jet mass. The closure of the method can be quantified with the double-ratio κ. This double-ratio will be close to unity if the following conditions are met:

1. The Double-b tagged to Anti-Double-b tag ratio is independent of the jet mass for each background independently.

2. The relative abundance of each background component does not affect the independence of the Double-b tagged to Anti-Double-b tag ratio and the jet mass.

In the next sections, we define the dedicated data control regions that allow us to study the closure of the background estimation method for each background process, which can then be propogated to the overall κ and its uncertainty. The systematic uncertainty on the overall κ is obtained as follows: sure of the method can be quantified with the double-ratio κ. This double-ratio will be close to unity if the following conditions are met:

1. For each data control control region, we derive a set of data/MC scale factors for each of the ABCD regions.

2. The scale factors are appied to the MC to gauge the non-closure and compute the κ and the scale factor uncertainty is propagated to the kappa uncertainty.

3. Since no $E_{\mathrm{T}}^{\mathrm{miss}}$ dependence is observed in the simulation or the observed data, the scale factors are derived intetgrating over $E_{\mathrm{T}}^{\mathrm{miss}}$.

Figure 13 and Figure 14 show the closure of the background method in MC for the 2-Higgs and 1-Higgs regions respectively for each of the main background processes. In general, the prediction is compatible with the events in the signal region, but the test is limited by the available MC statistics. The precision in the looser 1-Higgs signal region in Figure 13 is more precise due to the higher MC statistics compared to Figure 14. The bottom panel shows how consistent κ in MC is with 1.0, which indicates how the jet mass and double-b tag to anti-tag ratio is independent. The uncertainty on this ratio should take into account the statistical fluctuations in the numerator and the denominator. We compute the uncertainty on the ratio by throwing gaussian toys for the prediction and the observation seperately based on the central values and statisitical uncertainty. The κ is computed per toy and the uncertainty is taken based on the standard deviation of the distribution. For each background process, the value of 1.0 is within the kappa uncertainty. One exception is the W+jets background in the 1-Higgs tag region which over-predicts the background.

Figure 15 shows the closure of the total background in the signal regions based on the nominal compostion predicted in the MC. This tests if the relative abundance of a particular background process influences the overall closure. For the 1-Higgs and 2-Higgs signal regions the kappa ratio is consistent with 1.0.

The nominal MC may not reflect the relationship between the jet mass and the double-b tag to anti-tag ratio in data. The background composition in data control regions may also be different between data and MC. In order to assess and correct for these differences, scale factors are derived in validation regions that are enriched in each background process. The validation regions also boost the available MC statistics to test the closure. In Sections 6.2, 6.3, and 6.4, each of the MC used for a background process will be corrected based on the data in validation regions. The scale factors and the closure observed in the validation regions will be used to gauge the overall κ and uncertainty in the data. The final closure uncertainty assigned for the overall kappa will be obtained from :

₃₂₇ • The level of non-closure due to the relative abundance of each background process.
₃₂₈ This uncertainty is measured after correcting the MC based on the validation region
₃₂₉ scale factors.

₃₃₀ • The closure uncertainty for each background source obtained in the dedicated val-
₃₃₁ idation region. This uncertainty is propagated to the corrected MC process in the
₃₃₂ signal regions.

₃₃₃ • Since no E_T^{miss} dependence is observed, the scale factors are computed by integrating
₃₃₄ over the E_T^{miss} search bins to improve the statistical uncertainty. This uncertainty is
₃₃₅ propagated to the corrected MC process in the signal regions.

₃₃₆ Section **??** summarizes the sources of background uncertainty and how they are correlated in
the background fit.
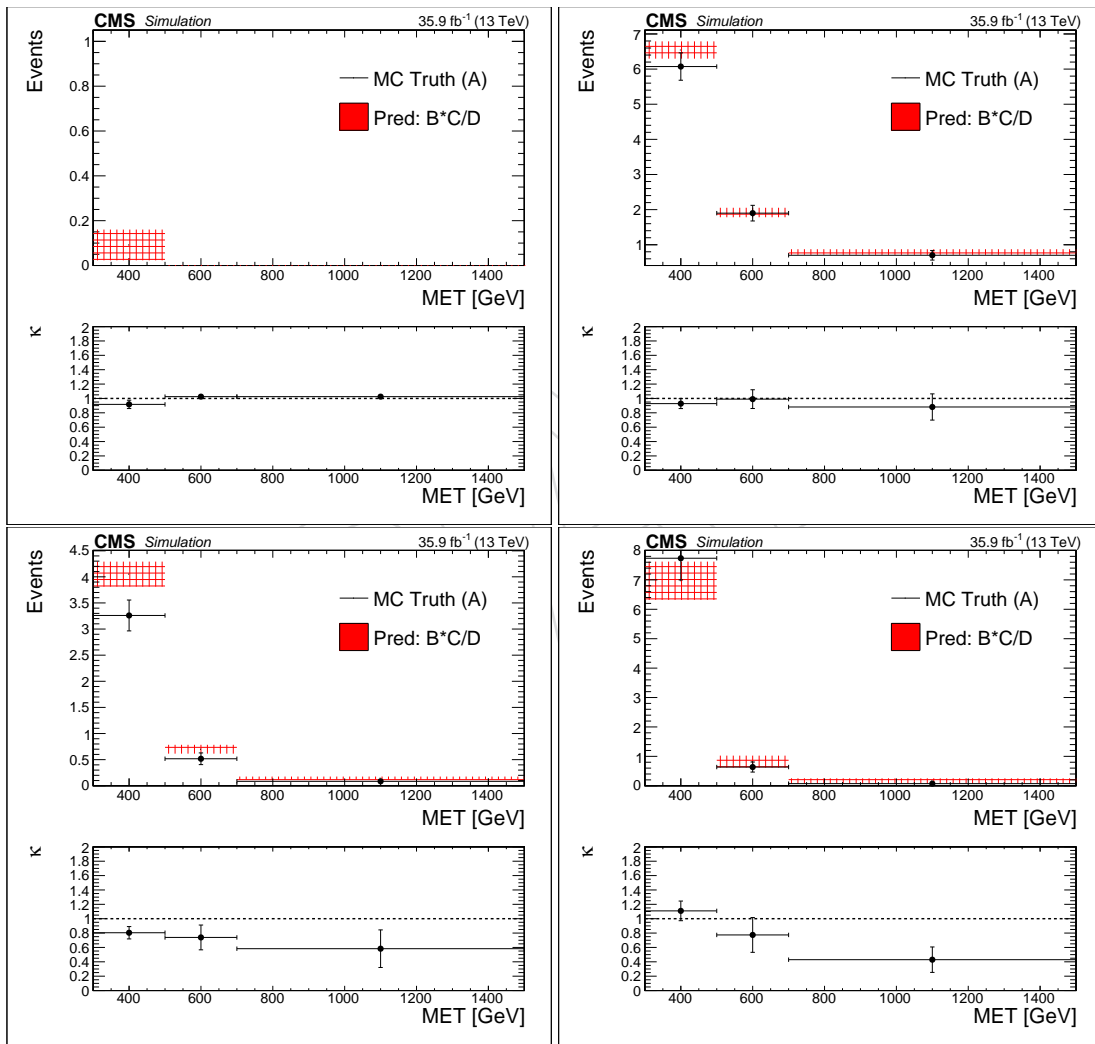
Figure 13: The kappa for each background process for the 1 Higgs tag region: Top (left) QCD (right) $Z \to \nu\nu$ Bottom (left) W+Jets (right) $t\bar{t}$. The uncertainties shown for the prediction and the observation in region A is the MC statistical unc. The uncertainty on κ is computed by throwing gaussian toys for the predict ion and the MC truth and computing the standard deviation of the distribution.
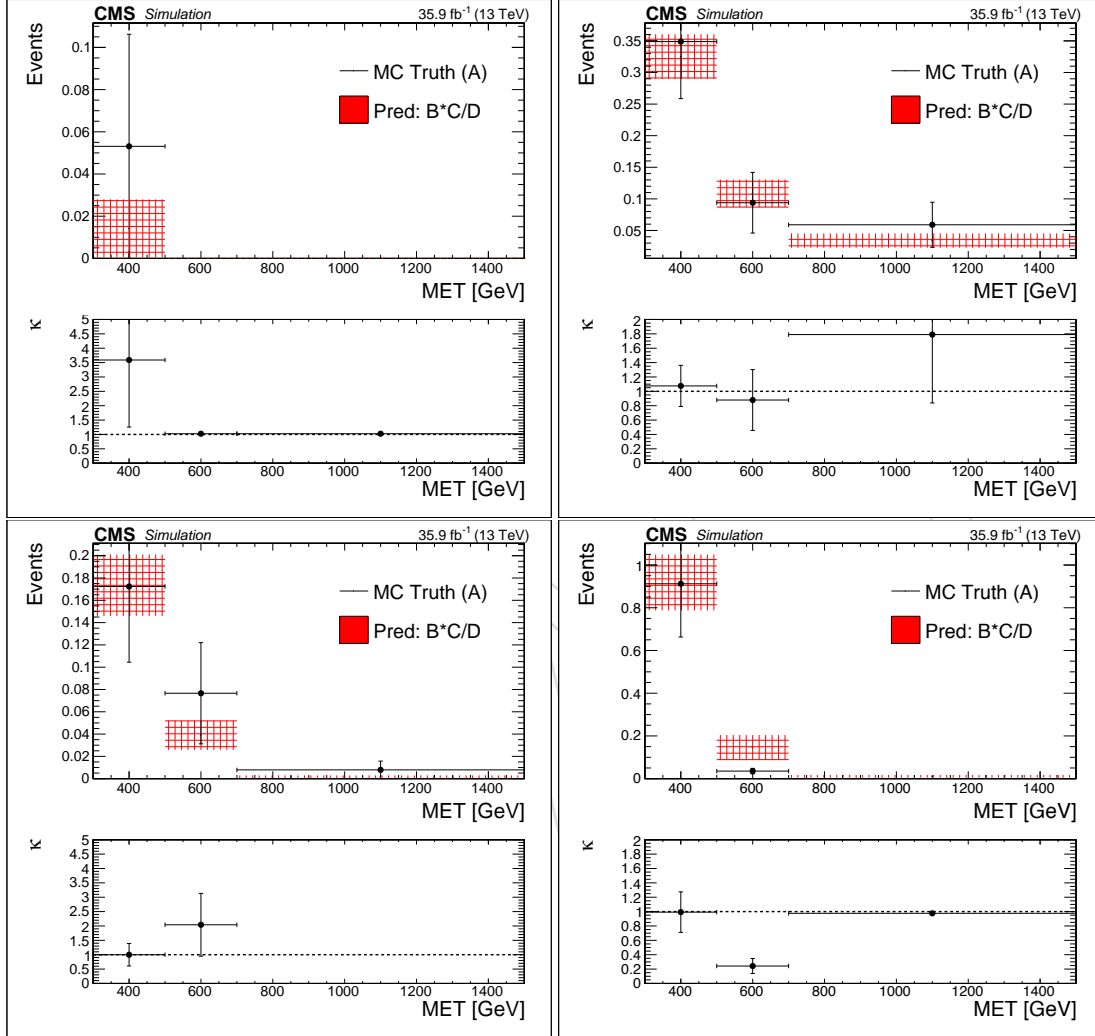
₃₃₇

Figure 14: The kappa for each background process for the 2 Higgs tag region: Top (left) QCD (right) $Z \rightarrow \nu\nu$ Bottom (left) W+Jets (right) $t\bar{t}$. The uncertainties shown for the prediction and the observation in region A is the MC statistical unc. The uncertainty on κ is computed by throwing gaussian toys for the prediction and the MC truth and computing the standard deviation of the distribution.
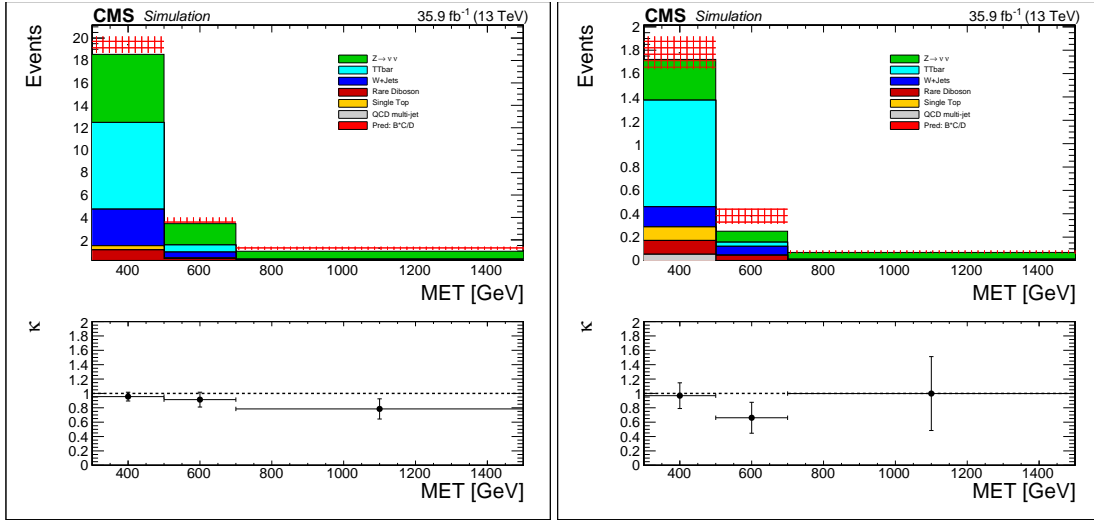
Figure 15: The closure for the total background from the SM monte-carlo is shown for (left) the 1-Higgs signal region and (right) the 2-Higgs signal region.

6.2 Low $\Delta\phi$ QCD Validation Region

We define a QCD enriched validation region by inverting the $\Delta\phi$ cut. This cut is designed to reject QCD multi-jet events with E_T^{miss} resulting from under-measured jets. Inverting the cut gives a rich sample of QCD events in data and MC but with some amount of contamination from the other backgrounds. As seen in Figure 14 and Figure 13 the QCD background component in search regions is very small, but can only be assessed in the lowest MET bin. The validation region can be used to test the background method for QCD across a range of MET. Figure 16 shows the closure for data events in the QCD enriched validation region and shows that the κ ratios are consistent with 1 within the available data statistics.

The low $\Delta\phi$ validation region is used to derive data/MC scale factors that are applied to the MC in the signal region. The scale factors are derived in each of the ABCD regions to correct the MC for any mismodeling of variables like the double-b discriminator and the pruned jet mass. Figure 17 shows that the pruned mass has no major difference in shape between data and MC. This implies we expect the scale factors in the signal window and the sideband to be similar so that they divide out in the ABCD prediction. Figure 18 shows the data comparison to MC for the double-b discriminator. For this variable there is a shape difference between data and MC that will affect the ABCD prediction. The MC underpredicts the events that pass the double-b discriminator cut and over-predicts those that would fail which will be reflected in the scale-factors.

Figure 19 shows the scale factors for each of the ABCD regions in the low $\Delta\phi$ validation region. Each ratio panel shows scale-factor for each E_T^{miss} bin, and it shows that the scale-factor does not depend on MET. Computing the scale factor integrated over MET reduces the statistical uncertainty. The number of events that do not pass the double-b discriminator, the anti-tag events, tend to be lower in data than in the MC. Figure 18 also shows that the data in the double-b discriminator falls more slowly than in the MC. The sub-lead double-b discriminator shows a bump at high double-b values corrsponding to the fraction of gluon splitting to $b\bar{b}$. The data to MC comparison suggests mis-modeling of the light flavor and heavy flavor fractions in the MC.
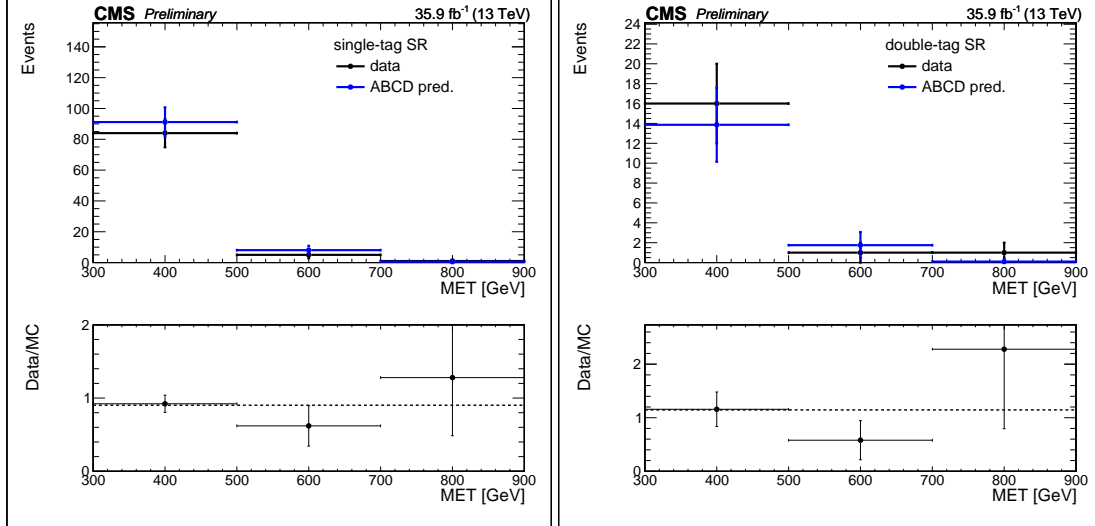
Figure 16: The closure for the low $\Delta\phi$ data is shown for (left) the 1-Higgs signal region and (right) the 2-Higgs signal region.
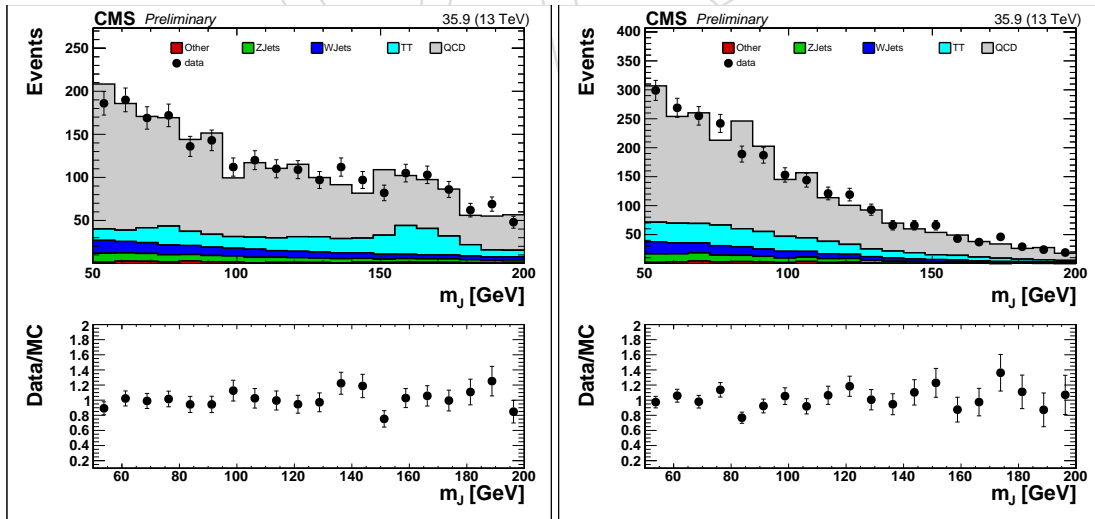

Figure 17: Data and MC comparison at low $\Delta\phi$ for the pruned mass of (left) the lead jet (right) and the sub-lead jet.
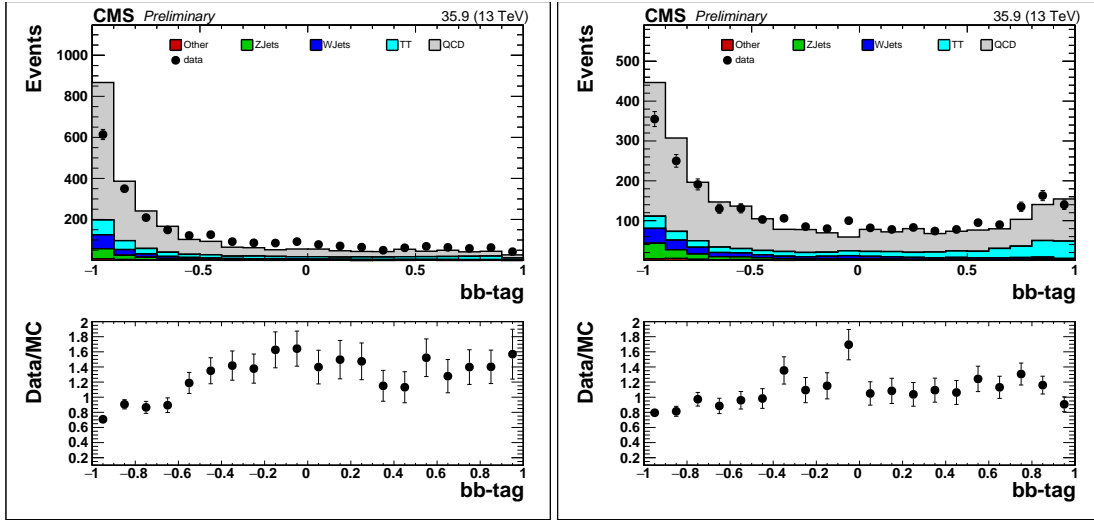
Figure 18: Data and MC comparison at low $\Delta\phi$ for the double-b discriminator of (left) the lead jet (right) and the sub-lead jet.

6.3 Single Lepton Validation Region

The $t\bar{t}$ and W backgrounds result from events where the lepton is lost in the event, and the E_T^{miss} originates from a neutrino from the W decay. To test the closure for these background components, we define a single lepton control region that represents a similar kinematic phase space to the signal region. The lepton in the event is not considered in the analysis variables. The presence of the lepton allows us to trigger on events with lower E_T^{miss} and increase the statistics in determining the scale-factors. The single electron and single muon regions can be used seperately to cross-check any mis-modeling in independant validation regions.

The closure in the single lepton data shows that the prediction is compatiable within the statistical unceratinty with the observed data. There is some deviation seen in the lowest E_T^{miss} bin for 2-Higgs tag region.

Figure 20 shows the lead and sub-lead jet mass for the single muon sample. The lead jet pruned mass shows a peak at the W-mass and also near the top mass values, and the sub-lead jet pruned mass has a broader peak near the W-mass. Both plots show similar shapes between data and MC. Figure 21 the same features with the same level of correspondence. These plots show the same level of agreement as observed in the QCD validation region.

Figure 22 shows the double-b discriminator comparison in data and MC. The main mis-modeling seems to occur at the very negative values of the double-b discriminator where the data is 80% to 60% of the MC prediction. Similar trends are seen with the single electron sample in Figure 23. This difference accounts for some but not all of the data/MC scale-factors observed in Figure 24 tend to be larger than the shape differnce in the double-b discriminator. The scale-factors account for differences in the modeling of AK8 jet p_T which is observed to have a different shape in data and in MC at high p_T.

6.4 Photon Validation Region

Direct prompt photons, which make up about 85% of the photon control sample, are most useful for predicting the $Z \to \nu\bar{\nu}$ background because their production process neatly maps to Z boson production processes. Photon events are selected based on whether they are in the barrel $|\eta| < 1.4442$ or endcap $1.4442 < |\eta| < 2.5$. The requirements for barrel (endcap) are:

Figure 19: E_T^{miss} distribution in each of the signal and control regions of the low-$\Delta\phi$ validation region. The hashed red distribution denote the prediction from simulation; the solid points denote the observed yields in data.
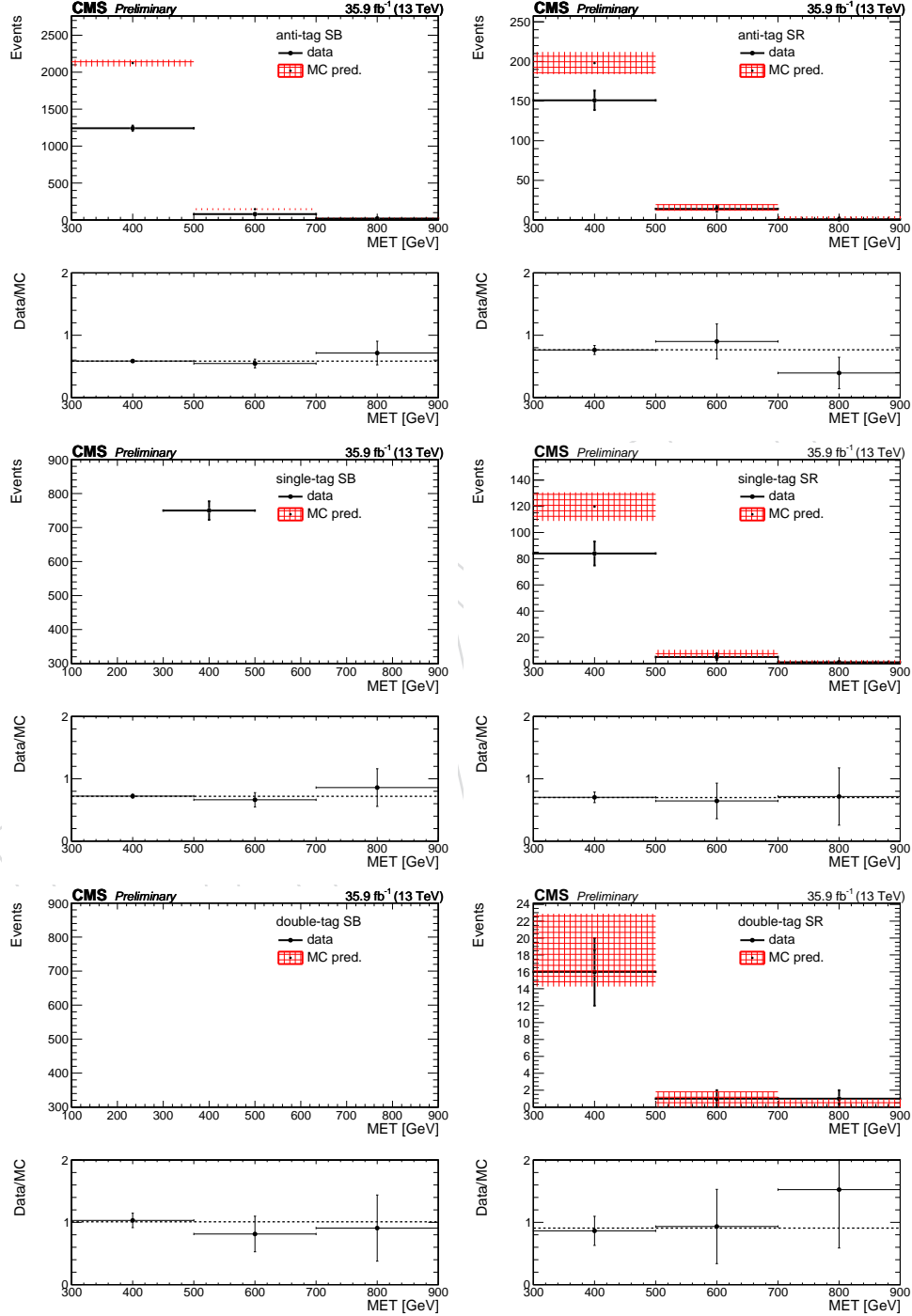
Figure 20: Distribution of pruned-jet mass of the leading (left) and sub-leading (right) AK8 jet for events in the muon validation region after baseline selections have been applied.
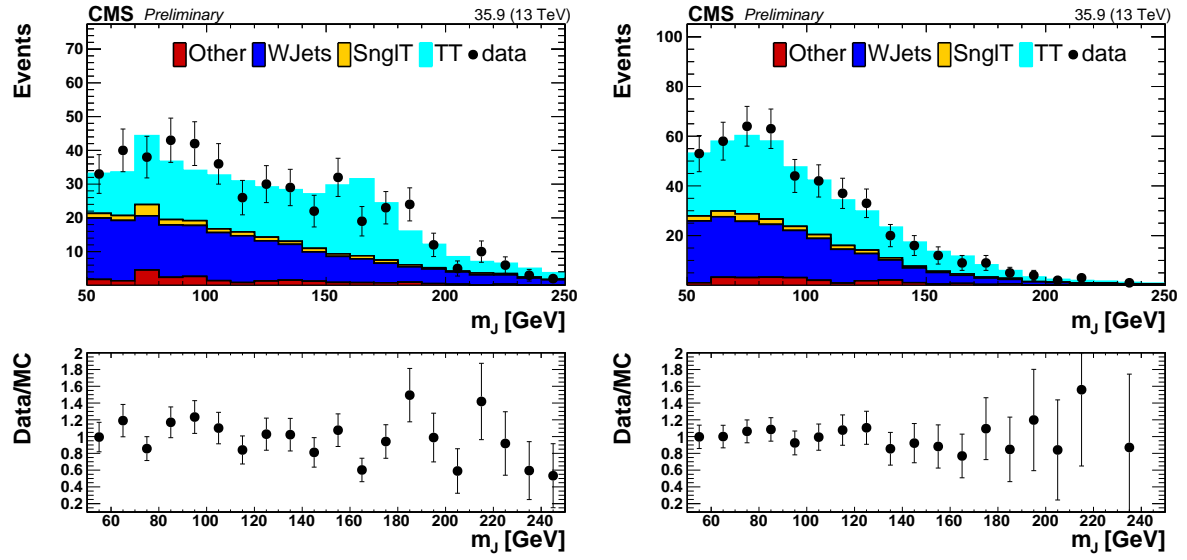

Figure 21: Distribution of pruned-jet mass of the leading (left) and sub-leading (right) AK8 jet for events in the electron validation region after baseline selections have been applied.
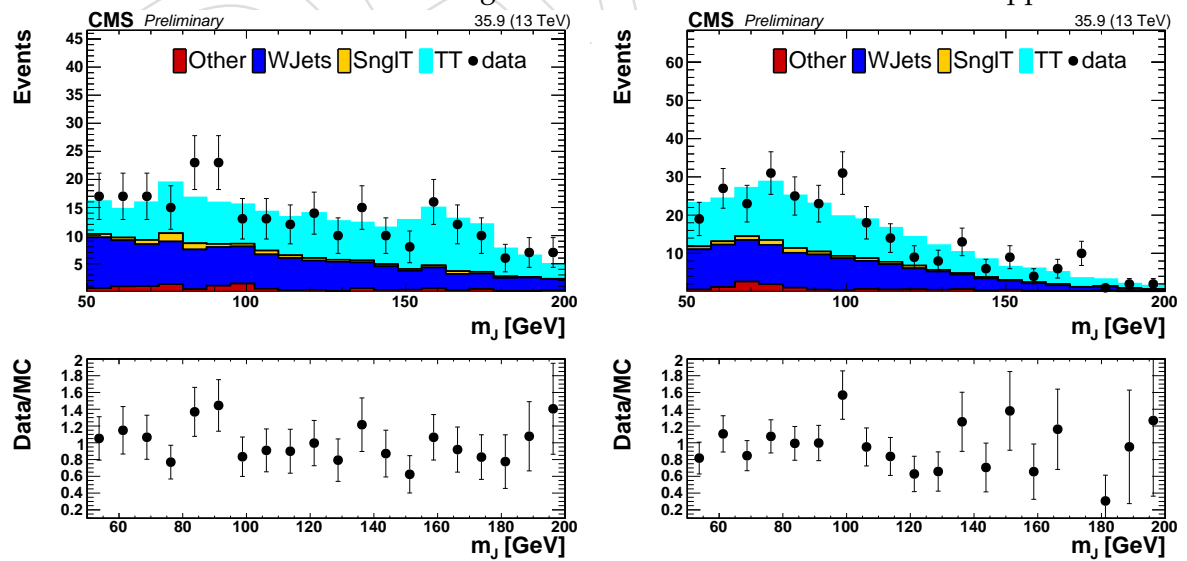
Figure 22: Distribution of double-b discriminant of the leading (left) and sub-leading (right) AK8 jet for events in the muon validation region after baseline selections have been applied.
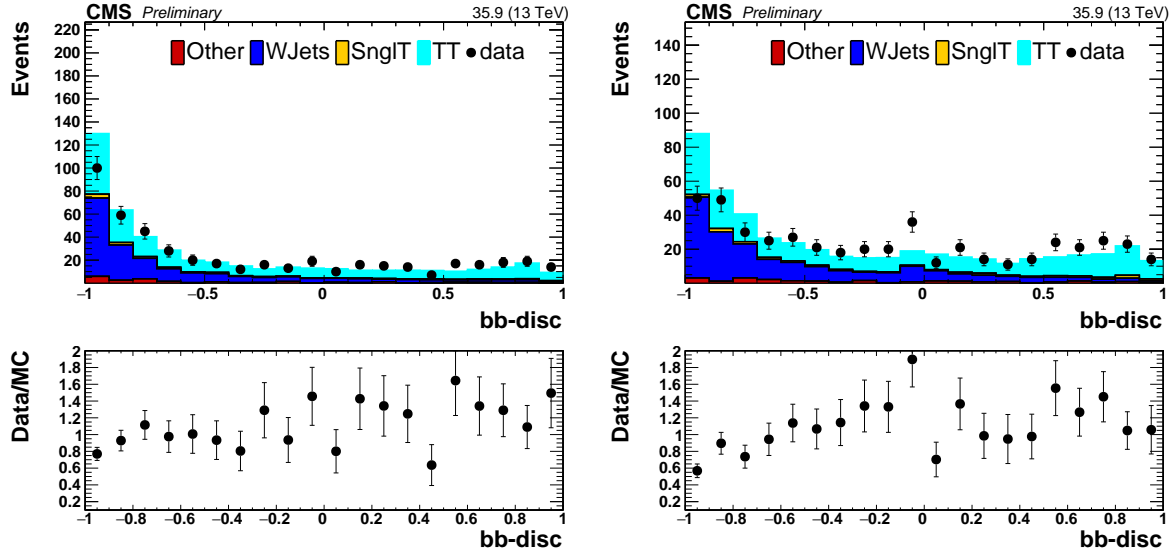

Figure 23: Distribution of double-b discriminant of the leading (left) and sub-leading (right) AK8 jet for events in the electron validation region after baseline selections have been applied.
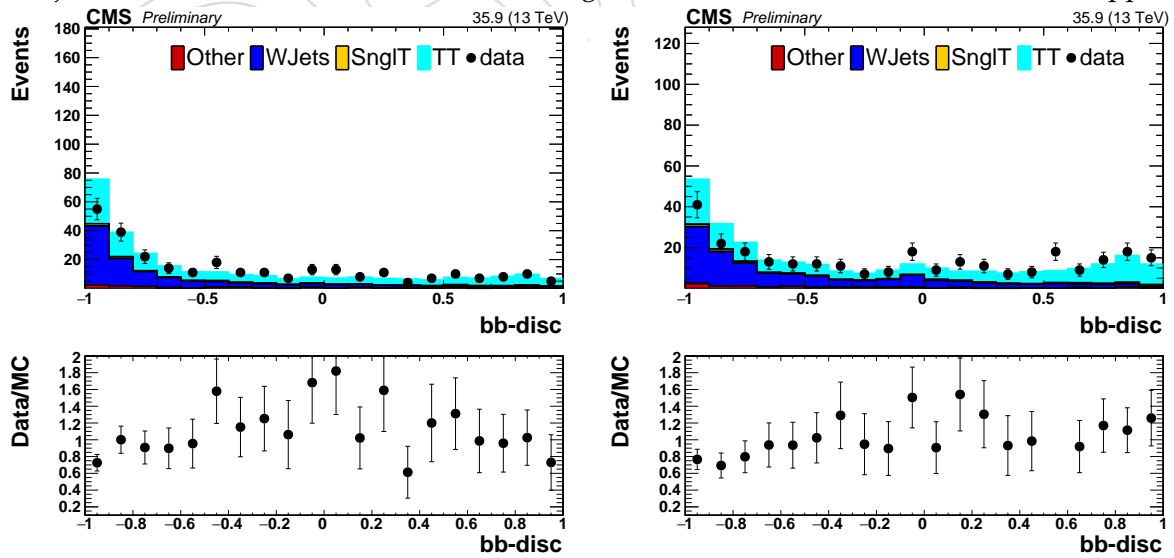
Figure 24: E_T^{miss} distribution in each of the signal and control regions of the e/μ validation region. The hashed red distribution denote the prediction from simulation; the solid points denote the observed yields in data.
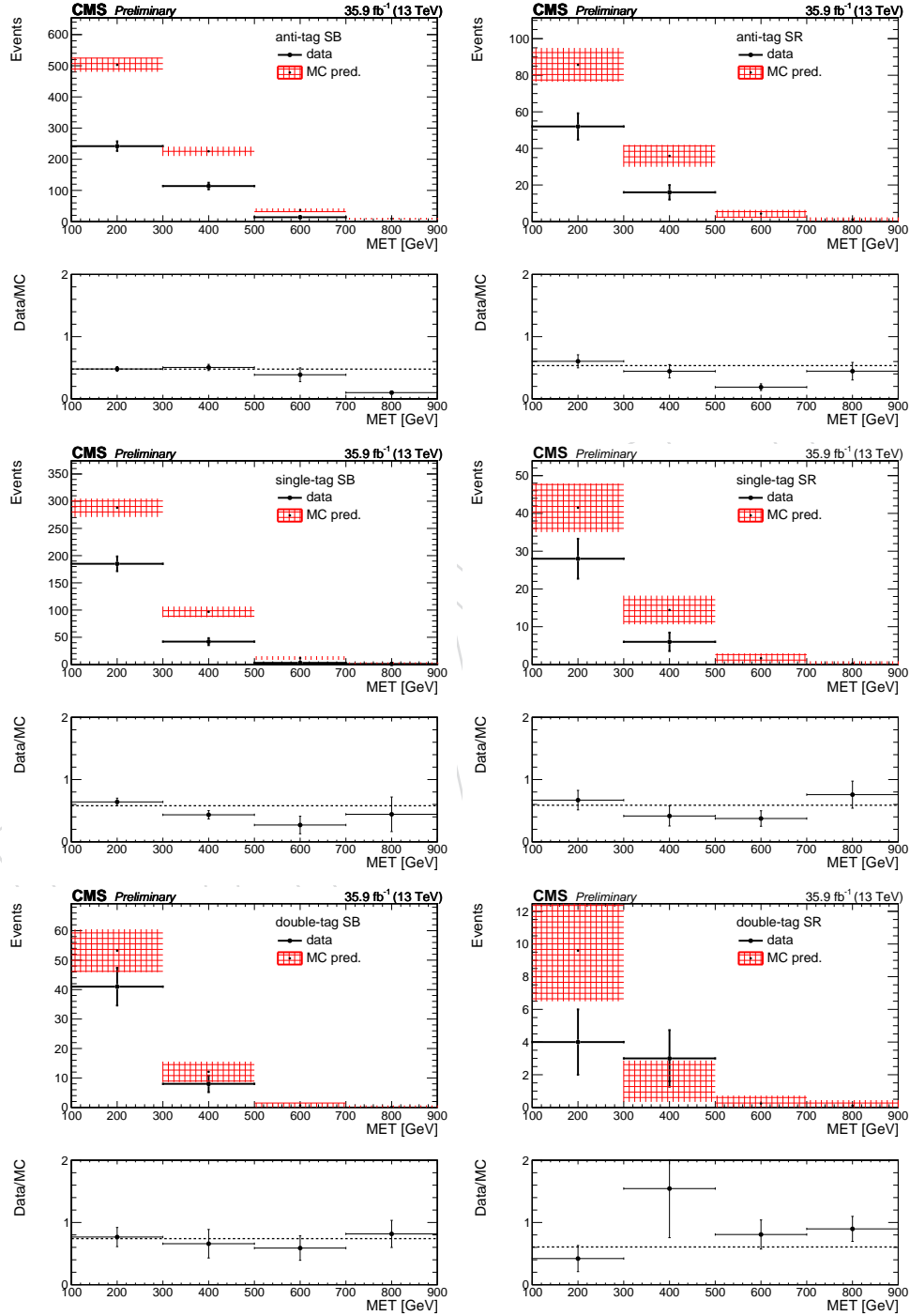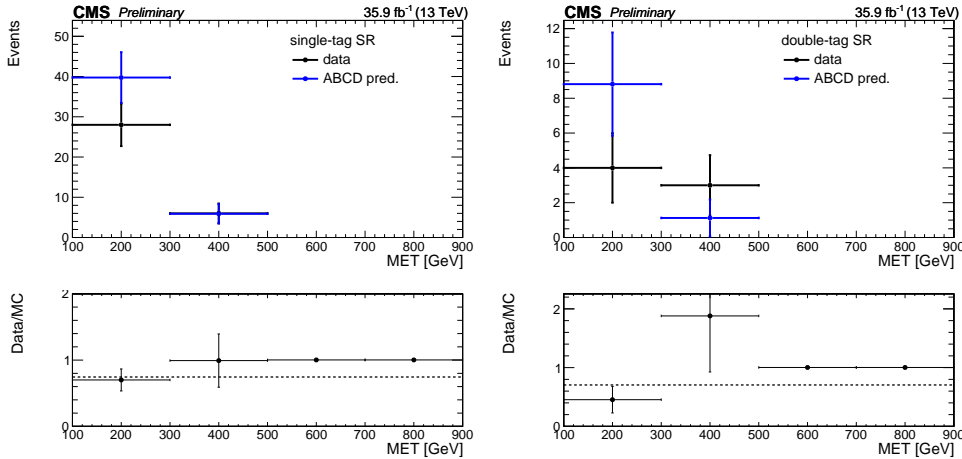
Figure 25: Closure observed in the E_T^{miss} search bins in the e/μ validation region data.(left) 1-Higgs Tag (right) 2-Higgs tag

396 • $H/E < 0.28$ (0.093)

397 • $\sigma_{i\eta i\eta} < 0.0107$ (0.0272)

398 • Charged hadron isolation < 2.67 (1.79), Neutral Hadron isolation $< 7.23e^{0.0028p_T+0.5408}$
399    $(8.89 + 0.01725p_T)$, and electromagnetic particle isolation $< 2.11 + 0.0014p_T$ $(3.09 +$
400    $0.0091p_T)$ The isolation requirements are computed in a $\Delta R < 0.3$

The photon kinematics in gamma jets events are used to test any differences between MC and data for Z-production. Kinematic differences between photons and Z-production are assessed as a seperate double ratio SF $\frac{R_{Z/\gamma}^{MC}}{R_{Z/\gamma}^{Data}}$ as shown in Eq. 3. The Z background will have an additional uncertainty that covers any deviation of this ratio from 1.0.

$$SF_\gamma = \frac{N_\gamma^{Data}}{N_\gamma^{MC}}, SF_Z = \frac{N_Z^{MC}}{N_Z^{Data}}, R_{Z/\gamma}^{MC/Data} = \frac{N_Z^{MC/Data}}{N_\gamma^{MC/Data}}, N_Z^{MC/Data} = R_{Z/\gamma}^{MC/Data} N_\gamma^{MC/Data}, MC_{SFCorr}^Z = N_Z^{MC} SF_Z = N$$

(3)

401 Figure 26 shows reasonable agreement in the mass shape between data and MC. Figure 27
402 shows the same shape difference as observed in the low $\Delta\phi$ validation region which is due
403 to the same mis-modeling of the light-flavor and heavy flavor content in MC. The closure in
404 Figure 28 is compatible with kappa of 1.0 within the statistical uncertaony for the 1-Higgs tag,
405 while for the 2-Higgs tag there is only data available in the loosest E_T^{miss} bin where the kappa is
406 also compatible with 1.0. The photon scale factors are shown in Figure 29.

407 **6.5   Summary of Systematic Uncertainties**

408 Taken into account the information from all the validation regions, we have obtained a set of
409 scale-factors that can be applied to the signal region to test the closure in MC that mimics the
410 data. We have also checked the closure uncertainty in each data validation region, but these
411 tests are statistically limited at large E_T^{miss} values. Table **??** and Table **??** show the κ factors
412 for the 1-Higgs and 2-Higgs regions respectively across the 3 E_T^{miss} search bins. In general, κ is
413 statistically compatible with 1 for the loosest E_T^{miss} region in each validation region. The 2-Higgs
414 tables shows more statistical fluctuations in the kappa factors due to limited data statistics.
415 Figure **??** shows the closure integrating over E_T^{miss} for the three validation regions.
416 The scale factors derived from each of the validation regions are shown in Table **??**.

417

Figure 26: Distribution of pruned-jet mass of the leading (left) and sub-leading (right) AK8 jet for events in the photon validation region after baseline selections have been applied.
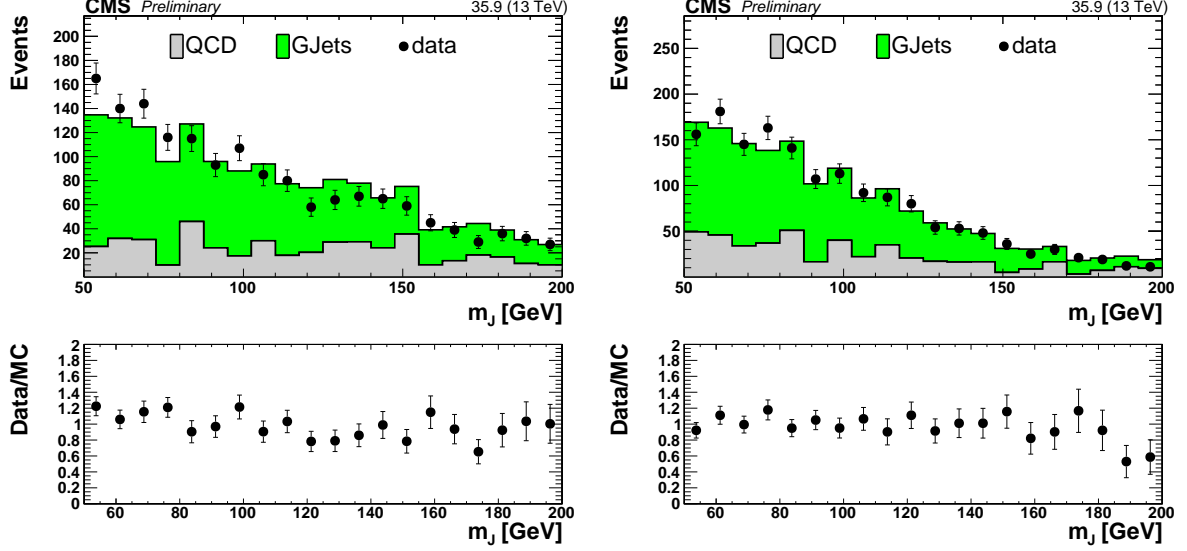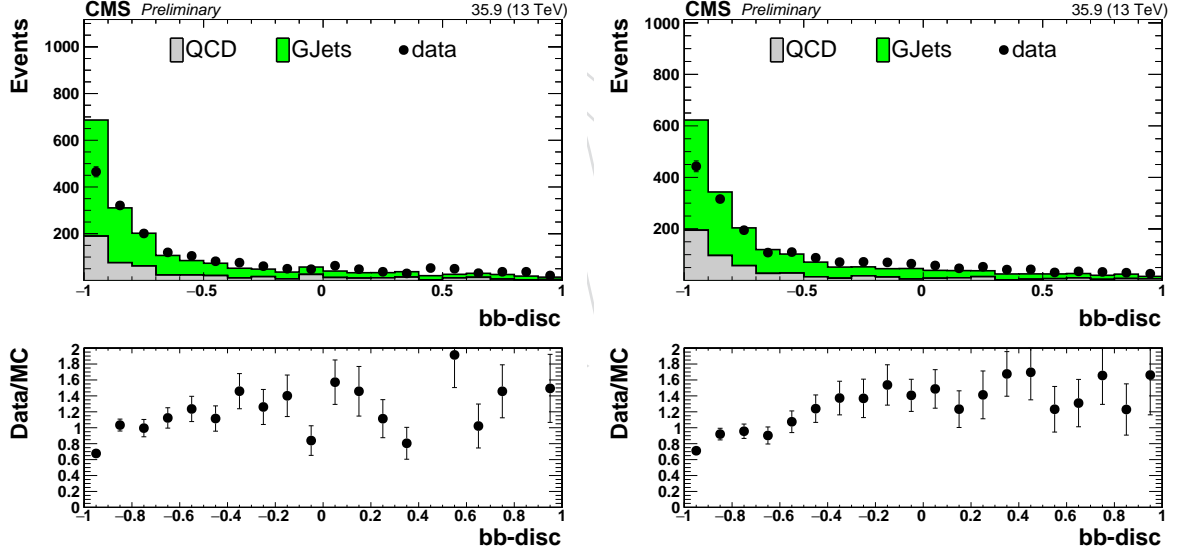

Figure 27: Distribution of double-b discriminant of the leading (left) and sub-leading (right) AK8 jet for events in the photon validation region after baseline selections have been applied.

Low $\Delta\phi$			
$E_\mathrm{T}^\mathrm{miss}[100,300]$	$E_\mathrm{T}^\mathrm{miss}[300,500]$	$E_\mathrm{T}^\mathrm{miss}[500,700]$	$E_\mathrm{T}^\mathrm{miss}>700$
	0.92 ± 0.12	0.62 ± 0.28	1.72 ± 1.14
Single Lepton			
$E_\mathrm{T}^\mathrm{miss}[100,300]$	$E_\mathrm{T}^\mathrm{miss}[300,500]$	$E_\mathrm{T}^\mathrm{miss}[500,700]$	$E_\mathrm{T}^\mathrm{miss}>700$
0.70 ± 0.16	$1.02 + 0.43$		
Photon			
$E_\mathrm{T}^\mathrm{miss}[100,300]$	$E_\mathrm{T}^\mathrm{miss}[300,500]$	$E_\mathrm{T}^\mathrm{miss}[500,700]$	$E_\mathrm{T}^\mathrm{miss}>700$
0.84 ± 0.18	0.81 ± 0.24	1.95 ± 0.96	1.39 ± 0.83

Table 8: κ closure ratios found in each validation region for the 1-Higgs tag search bins.

Figure 28: Closure observed in the $E_{\mathrm{T}}^{\mathrm{miss}}$ search bins in the γ validation region data (left) 1-Higgs region and (right) 2-Higgs region
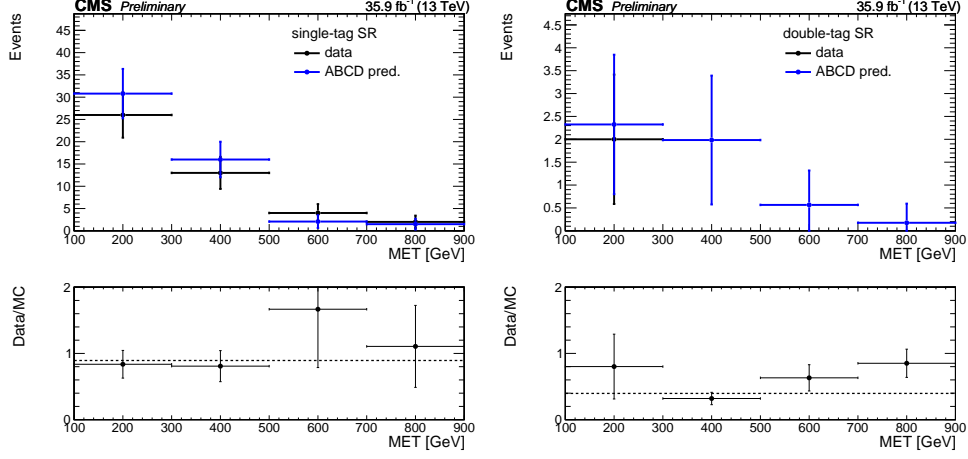
Low $\Delta\phi$			
$E_{\mathrm{T}}^{\mathrm{miss}}[100,300]$	$E_{\mathrm{T}}^{\mathrm{miss}}[300,500]$	$E_{\mathrm{T}}^{\mathrm{miss}}[500,700]$	$E_{\mathrm{T}}^{\mathrm{miss}} > 700$
	$1.15 + 0.30$	0.68 ± 0.43	2.71 ± 1.73
Single Lepton			
$E_{\mathrm{T}}^{\mathrm{miss}}[100,300]$	$E_{\mathrm{T}}^{\mathrm{miss}}[300,500]$	$E_{\mathrm{T}}^{\mathrm{miss}}[500,700]$	$E_{\mathrm{T}}^{\mathrm{miss}} > 700$
0.46 ± 0.21	$2.73 + 1.46$		
Photon			
$E_{\mathrm{T}}^{\mathrm{miss}}[100,300]$	$E_{\mathrm{T}}^{\mathrm{miss}}[300,500]$	$E_{\mathrm{T}}^{\mathrm{miss}}[500,700]$	$E_{\mathrm{T}}^{\mathrm{miss}} > 700$
0.92 ± 0.52	$0.34 + 0.06$	0.64 ± 0.10	0.85 ± 0.1

Table 9: κ closure ratios found in each validation region for the 2-Higgs tag search bins.

Low $\Delta\phi$					
A_{SF}^{1H}	A_{SF}^{2H}	B_{SF}	C_{SF}^1	C_{SF}^2	D_{SF}
Single Lepton					
A_{SF}^{1H}	A_{SF}^{2H}	B_{SF}	C_{SF}^1	C_{SF}^2	D_{SF}
0.58 ± 0.12	0.6 ± 0.25	0.53 ± 0.08	0.58 ± 0.049	0.74 ± 0.14	0.48 ± 0.031
Photon					
A_{SF}^{1H}	A_{SF}^{2H}	B_{SF}	C_{SF}^1	C_{SF}^2	D_{SF}
0.79 ± 0.15	0.44 ± 0.26	0.49 ± 0.046	0.91 ± 0.06	1.7 ± 0.41	0.6 ± 0.02

Table 10: Summary of the Validation region scale-factors derived integrated over the $E_{\mathrm{T}}^{\mathrm{miss}}$ search bins. They are computed for each of the regions defined at the beginning of the Section.

Figure 29: $E_\mathrm{T}^\mathrm{miss}$ distribution in each of the signal and control regions of the γ validation region. The hashed red distribution denote the prediction from simulation; the solid points denote the observed yields in data.
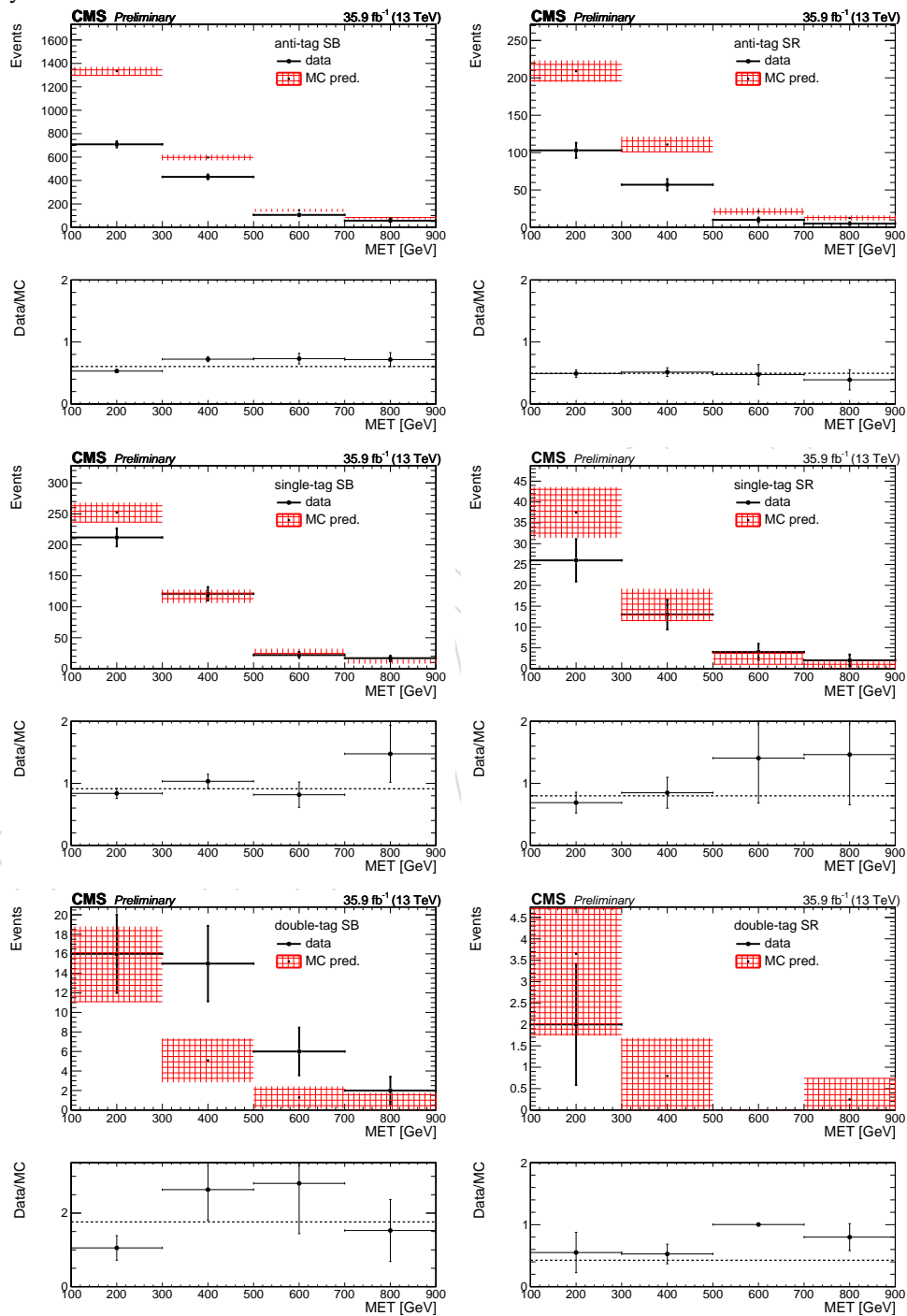
7 Results

References

[1] CMS Collaboration, "Search for supersymmetry in multijet events with missing transverse momentum in proton-proton collisions at 13 TeV", `arXiv:1704.07781`.

[2] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm", *JHEP* **04** (2008) 063, `doi:10.1088/1126-6708/2008/04/063`, `arXiv:0802.1189`.

[3] CMS Collaboration, "Particle–Flow Event Reconstruction in CMS and Performance for Jets, Taus, and E_T^{miss}", CMS Physics Analysis Summary CMS-PAS-PFT-09-001, CERN, 2009.

[4] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, "Recombination Algorithms and Jet Substructure: Pruning as a Tool for Heavy Particle Searches", *Phys. Rev.* **D81** (2010) 094023, `doi:10.1103/PhysRevD.81.094023`, `arXiv:0912.0033`.

[5] CMS Muon POG, "Baseline muon selections for 2012 data (CMSSW 52X and above)", (2012). `https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideMuonId#The2012Data`.

[6] Rohan Bhandari, Adam Dishaw, Ryan Heller, Ana Ovcharova, Jeffrey Richman, Manuel Franco Sevilla, David Stuart, Chris West, Jae Hyeok Yoo, "Search for supersymmetry in pp collisions at $\sqrt{s} = 13$ TeV in the single-lepton final state using the sum of masses of large radius jets", *CMS Analysis Note* **AN-15-139** (2015).

[7] CMS E/gamma POG, "Cut Based Electron ID for Run 2", (2014). `https://twiki.cern.ch/twiki/bin/viewauth/CMS/CutBasedElectronIdentificationRun2`.

[8] CMS Collaboration Collaboration, "Identification of double-b quark jets in boosted event topologies", Technical Report CMS-PAS-BTV-15-002, CERN, Geneva, 2016.