

# CMS Draft Analysis Note

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

2019/11/09

Archive Hash: 729478e-D

Archive Date: 2019/11/09

## Search for VBF Higgs bosons decaying to invisible particles at 13 TeV with 2017 data

Alp Akpinar, Zeynep Demiragli, Andreas Albert, and Siqi Yuan  
Boston University (US)

### Abstract

This note describes the search for invisible decays of Higgs boson produced with vector boson fusion (VBF). The search is performed using a shape-based analysis, using the data collected by CMS at  $\sqrt{s} = 13 \text{ TeV}$  in 2017 and 2018, corresponding to integrated luminosities of  $41.3 \text{ fb}^{-1}$  and  $59.7 \text{ fb}^{-1}$ , respectively.

This box is only visible in draft mode. Please make sure the values below make sense.

PDFAuthor: Alp Akpinar  
PDFTitle: Run-II VBFHinv BU  
PDFSubject: CMS  
PDFKeywords: CMS, physics, your topics

Please also verify that the abstract does not use any user defined symbols



**Contents**

1	1	Introduction . . . . .	2
2	2	Samples . . . . .	2
3	2.1	Data Samples . . . . .	2
4	2.2	Background Samples . . . . .	3
5	2.3	Signal simulation samples . . . . .	5
6	3	Physics objects . . . . .	6
7	3.1	Jets . . . . .	6
8	3.2	Missing transverse momentum and recoil . . . . .	8
9	3.3	Leptons . . . . .	9
10	3.4	Photons . . . . .	12
11	4	Reweighting of simulated events . . . . .	12
12	4.1	Trigger efficiency reweighting . . . . .	12
13	4.2	Pileup reweighting . . . . .	14
14	4.3	Lepton and photon identification/reconstruction efficiency reweighting .	15
15	4.4	Higher-order reweighting . . . . .	16
16	4.5	Generator-level boson construction . . . . .	16
17	5	Event selection . . . . .	20
18	5.1	Signal region selection . . . . .	20
19	5.2	Single muon control region selection . . . . .	29
20	5.3	Single electron control region selection . . . . .	34
21	5.4	Double muon control region selection . . . . .	39
22	5.5	Double electron control region selection . . . . .	44
23	5.6	Photon control region . . . . .	49
24	6	Signal extraction strategy . . . . .	53
25	6.1	Multijet background estimation . . . . .	53
26	6.2	EW background estimation and fitting procedure . . . . .	55
27	6.3	Systematic Uncertainties . . . . .	61
28	6.4	Control sample validation . . . . .	65
29	7	Results . . . . .	71
30			

## 1 Introduction

Several astrophysical observations [?] provide compelling evidence for the existence of dark matter (DM), a type of matter not accounted for in the standard model (SM). To date only gravitational interactions of DM have been observed and it remains unknown if DM has a particle origin and could interact with ordinary matter via SM processes. However, many theoretical models have been proposed in which DM and SM particles interact with sufficient strength that DM may be directly produced with observable rates in high energy collisions at the CERN LHC. While the DM particles would remain undetected, they may recoil with large transverse momentum ( $p_T$ ) against other detectable particles resulting in an overall visible  $p_T$  imbalance in a collision event. This type of event topology is rarely produced in SM processes and therefore enables a highly sensitive search for DM. Similar event topologies are predicted by other extensions of the SM, such as the Arkani-Hamed, Dimopoulos, and Dvali (ADD) model [?] of large extra spatial dimensions (EDs).

This analysis note describes a search for Higgs boson produced by vector boson fusion (VBF), decaying into invisible particles. The signature of such a final state will be two separated jets and an imbalance in  $\vec{p}_T$  due to the undetected particles. Two separated jets are the result of the hadronization of two final state quarks emerging from the VBF process.

This analysis makes use of data collected with the CMS detector in proton-proton (pp) collisions in 2017 at  $\sqrt{s} = 13$  TeV, corresponding to an integrated luminosity of  $41.5 \text{ fb}^{-1}$ .

This paper describes a search for new physics resulting in final states with one or more energetic jets and an imbalance in  $p_T$  due to undetected particles. The jets are the result of the fragmentation and hadronization of quarks or gluons, which may be produced directly in the hard scattering process as initial-state radiation or as the decay products of a vector boson  $V$  ( $W$  or  $Z$ ). These final states are commonly referred to as “monojet” and “mono- $V$ ”. Several searches have been performed at the LHC using the monojet and mono- $V$  channels [?]. This analysis makes use of a data sample of proton-proton (pp) collisions at  $\sqrt{s} = 13$  TeV collected with the CMS detector at the LHC, corresponding to an integrated luminosity of  $101.3 \text{ fb}^{-1}$ . This sample is approximately three times larger than the one used in Ref. [?]. The analysis strategy is similar to that of previous CMS searches, and simultaneously employs event categories to target both the monojet and mono- $V$  final states. In an improvement compared to previous searches, in this paper revised theoretical predictions and uncertainties for  $\gamma + \text{jets}$ ,  $Z + \text{jets}$ , and  $W + \text{jets}$  processes based on recommendations of Ref. [?] are used.

## 2 Samples

The analysis described in this note is performed using data collected in 2017 by CMS at 13 TeV and corresponds to an integrated luminosity of  $41.5 \text{ fb}^{-1}$ . The MC simulation samples for the background processes have been produced in the and campaigns. Further details are given in the following sub-sections.

### 2.1 Data Samples

The datasets listed in Tab. 1 are used to select events in the signal and the control regions, and to perform measurements on physics objects used in the analysis (e.g. trigger turn-ons).

Events for the signal region are collected using a set of dedicated triggers designed to select events with large  $p_T^{\text{miss}}$  and large  $H_T^{\text{miss}}$  based on the online particle flow (PF) algorithm. In these dedicated trigger algorithms, identified PF muons are removed from the event before the

Table 1: List of datasets used to select events in the signal and control regions. Datasets for both years correspond to the campaign, otherwise known as . For the 2017 data, the reconstruction is used for all periods. For 2018 data, the reconstruction is used for runs A to C, and the reconstruction is used for run D.

Year	Dataset name	Events selected for
both	/MET/Run201*/NANOAOD	Signal , single muon, double muon control regions
2017	/SingleElectron/Run2017*/NANOAOD	Single electron, double electron control regions
	/SinglePhoton/Run2017*/NANOAOD	Single photon control region
2018	/EGamma/Run2018*/NANOAOD	Single electron, double electron control, single photon control regions

$p_T^{\text{miss}}$  and the  $H_T^{\text{miss}}$  objects are calculated. With this definition, the signal trigger paths can also be used to select single and double muon events for the W and Z control regions, respectively.

Electron events for the W and Z regions are selected using a single electron trigger. To ensure the trigger efficiency also for high- $p_T$  electrons, the single electron trigger is used in combination with a single photon trigger [? ]. The same photon trigger is used to select events for the photon control region.

The full list of triggers used, along with the L1 seeds and the associated primary datasets are shown in Table 2.

## 2.2 Background Samples

Simulation datasets for the background processes are listed in Table 3 and 4. There are several Standard Model processes that pose as backgrounds to the VBF  $H_{inv}$  signal, experimental signature of the final state being two jets with large rapidity separation and invariant mass, along with  $p_T^{\text{miss}}$ . These processes are as follows:

**Z( $\nu\nu$ ) + jets** : This process yields the largest irreducible background in the analysis, consisting of a Z boson and 2 or more jets coming from either QCD or EWK vertices. Simulated samples for this background have been produced at leading order (LO) in QCD using the aMC@NLO generator in several bins of  $H_T$ .

**W + jets** : This process is the second largest source of background in this analysis, consisting of a W boson and 2 or more jets coming from either QCD or EWK vertices. The contribution of this background can be reduced by rejecting events with charged lepton candidates (electron/muon/tau). However, this process becomes irreducible in the case where the charged leptons are outside of the detector acceptance. Simulation samples for this background have been generated at LO in QCD using the aMC@NLO generator in several bins  $H_T$ .

**Z( $\ell\ell$ ) + jets** : This process mimics signal-like events in the case where the leptons coming from the Z boson decay are not reconstructed. As in the case of  $W \rightarrow \ell\nu$ , the contribution of this background is reduced by rejecting events with charged leptons. Simulation samples for this process have been generated at LO in QCD using the aMC@NLO generator in bins of  $H_T$ .

**Top**: Top-quark decays (both  $t\bar{t}$  and single top) also contribute background events to this analysis. In these processes, the W boson produced in a top-quark decay further decays

Table 2: HLT paths and the associated L1 seeds used in the analysis for the 2017 and 2018 datasets. The HLT paths ending in “\_HT60” are backup triggers introduced to mitigate noise rate problems in 2017. Their inclusion is not strictly necessary for 2018, but is done for consistency.

Year	HLT path	L1 seed	Primary dataset
2017	HLT_PFMETNoMu120_PFMHTNoMu120_IDTight		MET
	HLT_PFMETNoMu120_PFMHTNoMu120_IDTight_PFHT60		MET
	HLT_Ele35_WPTight_Gsf		SingleElectron
	HLT_Photon200		SinglePhoton
2018	HLT_PFMETNoMu120_PFMHTNoMu120_IDTight		MET
	HLT_PFMETNoMu120_PFMHTNoMu120_IDTight_PFHT60		
	HLT_Ele32_WPTight_Gsf		EGamma
	HLT_Photon200		EGamma

leptonically, which produces genuine  $p_T^{\text{miss}}$  in the event. Next-to-leading order (NLO) tt simulation samples have been produced with the aMC@NLO generator. Single-top events have been generated with the Powheg generator at NLO.

**Dibosons:** Decays of diboson pairs (WW, WZ, ZZ) also constitute background processes. Typically, one of the bosons decays leptonically ( $W \rightarrow \ell\nu, Z \rightarrow \nu\nu$ ) while the other boson decays hadronically, thus producing jets and  $p_T^{\text{miss}}$  in the final state. Simulated samples for WW, WZ and ZZ production have been generated at LO with Pythia 8.

**QCD Multijet:** QCD multijet events typically do not have large genuine  $p_T^{\text{miss}}$ . However, given the large cross section with which these events are produced, even a small fraction of events with jet mismeasurement results in a non-zero contribution of this process as background in the analysis. Simulated QCD samples have been generated at LO in QCD using the MadGraph generator in several bins of  $H_T$ .

The MC samples produced using MADGRAPH5\_aMC@NLO, and POWHEG generators are interfaced with PYTHIA using the CP5 tune [?] for the fragmentation, hadronization, and underlying event description. In the case of the MADGRAPH5\_aMC@NLO samples, jets from the matrix element calculations are matched to the parton shower description following the MLM [?] (FxFx [?]) prescription to match jets from matrix element calculations and parton shower description for LO (NLO) samples. The NNPDF 3.1 NNLO [?] parton distribution functions (PDFs) are used in all generated samples. The propagation of all final-state particles through

Dataset name	Cross section (pb)	Order in QCD
WJetsToLNu_HT-70To100_TuneCP5_13TeV-madgraphMLM-pythia8	1296	LO
WJetsToLNu_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8	1392	LO
WJetsToLNu_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8	410.2	LO
WJetsToLNu_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8	57.95	LO
WJetsToLNu_HT-600To800_TuneCP5_13TeV-madgraphMLM-pythia8	12.98	LO
WJetsToLNu_HT-800To1200_TuneCP5_13TeV-madgraphMLM-pythia8	5.39	LO
WJetsToLNu_HT-1200To2500_TuneCP5_13TeV-madgraphMLM-pythia8	1.08	LO
WJetsToLNu_HT-2500ToInf_TuneCP5_13TeV-madgraphMLM-pythia8	0.008098	LO
ZJetsToNuNu_HT-100To200_13TeV-madgraph	305.3	LO
ZJetsToNuNu_HT-200To400_13TeV-madgraph	91.86	LO
ZJetsToNuNu_HT-400To600_13TeV-madgraph	13.13	LO
ZJetsToNuNu_HT-600To800_13TeV-madgraph	3.242	LO
ZJetsToNuNu_HT-800To1200_13TeV-madgraph	1.501	LO
ZJetsToNuNu_HT-1200To2500_13TeV-madgraph	0.3431	LO
ZJetsToNuNu_HT-2500ToInf_13TeV-madgraph	0.005146	LO
DYJetsToLL_M-50_HT-70to100_TuneCP5_13TeV-madgraphMLM-pythia8	146.9	LO
DYJetsToLL_M-50_HT-100to200_TuneCP5_13TeV-madgraphMLM-pythia8	160.9	LO
DYJetsToLL_M-50_HT-200to400_TuneCP5_13TeV-madgraphMLM-pythia8	48.68	LO
DYJetsToLL_M-50_HT-400to600_TuneCP5_13TeV-madgraphMLM-pythia8	6.998	LO
DYJetsToLL_M-50_HT-600to800_TuneCP5_13TeV-madgraphMLM-pythia8	1.745	LO
DYJetsToLL_M-50_HT-800to1200_TuneCP5_13TeV-madgraphMLM-pythia8	0.8077	LO
DYJetsToLL_M-50_HT-1200to2500_TuneCP5_13TeV-madgraphMLM-pythia8	0.1923	LO
DYJetsToLL_M-50_HT-2500toInf_TuneCP5_13TeV-madgraphMLM-pythia8	0.003477	LO
GJets_HT-40To100_TuneCP5_13TeV-madgraphMLM-pythia8	18640	LO
GJets_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8	8641	LO
GJets_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8	2196	LO
GJets_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8	258.4	LO
GJets_HT-600ToInf_TuneCP5_13TeV-madgraphMLM-pythia8	85.23	LO

Table 3: Simulated datasets for the modelling of single electroweak boson backgrounds. All datasets are accessed at the NanoAOD data tier from the campaign, also known as .

the CMS detector is simulated with the GEANT 4 software [? ]. The simulated events include the effects of pileup, with the multiplicity of reconstructed primary vertices matching that in data. The average number of pileup interactions per proton bunch crossing is found to be 32 for both the 2017 and 2018 data samples used in this analysis (assuming a total inelastic proton-proton cross-section).

## 2.3 Signal simulation samples

In this section we should add description of the models, feynman diagrams, important model parameters

### DMSimp Models

#### t-channel Model

#### ADD Model

#### Leptoquark Model

#### Higgs Invisible Model



Dataset name	Cross section (pb)	Order in QCD
TTJets_TuneCP5_13TeV-amcatnloFXFX-pythia8	831.76	NLO
ST.t-channel_top_4f_inclusiveDecays_TuneCP5_13TeV-powhegV2-madspin-pythia8	137.458	NLO
ST.t-channel_antitop_4f_inclusiveDecays_TuneCP5_13TeV-powhegV2-madspin-pythia8	83.0066	NLO
ST.tW_top_5f_inclusiveDecays_TuneCP5_*_13TeV-powheg-pythia8	35.85	NLO
ST.tW_antitop_5f_inclusiveDecays_TuneCP5_*_13TeV-powheg-pythia8	35.85	NLO
WW_TuneCP5_13TeV-pythia8	75.91	LO
WZ_TuneCP5_13TeV-pythia8	27.56	LO
ZZ_TuneCP5_13TeV-pythia8	12.14	LO
QCD_HT1000to1500_TuneCP5_13TeV-madgraph-pythia8	1095	LO
QCD_HT1000to1500_TuneCP5_13TeV-madgraph-pythia8	1095	LO
QCD_HT100to200_TuneCP5_13TeV-madgraph-pythia8	2.369e+07	LO
QCD_HT1500to2000_TuneCP5_13TeV-madgraph-pythia8	99.27	LO
QCD_HT2000toInf_TuneCP5_13TeV-madgraph-pythia8	20.25	LO
QCD_HT200to300_TuneCP5_13TeV-madgraph-pythia8	1.554e+06	LO
QCD_HT300to500_TuneCP5_13TeV-madgraph-pythia8	324300	LO
QCD_HT300to500_TuneCP5_13TeV-madgraph-pythia8	324300	LO
QCD_HT500to700_TuneCP5_13TeV-madgraph-pythia8	29990	LO
QCD_HT50to100_TuneCP5_13TeV-madgraphMLM-pythia8	1.85e+08	LO
QCD_HT700to1000_TuneCP5_13TeV-madgraph-pythia8	6374	LO
QCD_HT700to1000_TuneCP5_13TeV-madgraph-pythia8	6374	LO

Table 4: Simulated datasets for the modelling of subleading background processes. All datasets are accessed at the NanoAOD data tier from the campaign, also known as .

### 3 Physics objects

All physics objects are used to identify signal-like events, to suppress backgrounds and to define control regions for background estimation. For the object definitions, we mostly follow the CMS POG endorsed recommendations. The physics objects and the selection requirements are described below.

#### 3.1 Jets

In this analysis, jets are reconstructed by clustering PF candidates using the infrared and collinear safe anti- $k_T$  algorithm [? ]. Jets are clustered with a distance parameter of 0.4 and are referred to as AK4 jets. The reconstructed vertex with the largest value of summed physics-object  $p_T^2$  is taken to be the primary pp interaction vertex. The physics objects are those returned by a jet finding algorithm [? ? ] applied to all charged PF candidates associated with the vertex, plus the corresponding associated missing transverse momentum.

Jet momentum is determined as the vector sum of all particle momenta in the jet, and is found from simulation to be within 5 to 10% of the true momentum over the full  $p_T$  spectrum and detector acceptance. An offset correction is applied to jet energies to take into account the contribution from additional proton-proton interactions within the same or nearby bunch crossings (pileup). Jet energy corrections are derived from simulation, and are confirmed with *in situ* measurements of the energy balance in dijet, multijet,  $\gamma$ +jet, and leptonic Z+jet events [? ]. The and versions of the jet energy corrections are used for the 2017 and 2018 data sets, respectively.

The AK4 jets used in this analysis are required to pass loose jet identification criteria. In addition, all the jets with  $p_T$  smaller than  $50\text{GeV}$  must pass the medium pileup ID criteria. This additional constraint on all AK4 jets is found to improve the modeling of jet distributions, especially in the horn regions near  $|\eta| = 2.9$ . The effect of this requirement is demonstrated in



the sub-leading jet  $\eta$  distribution in Fig. 1.

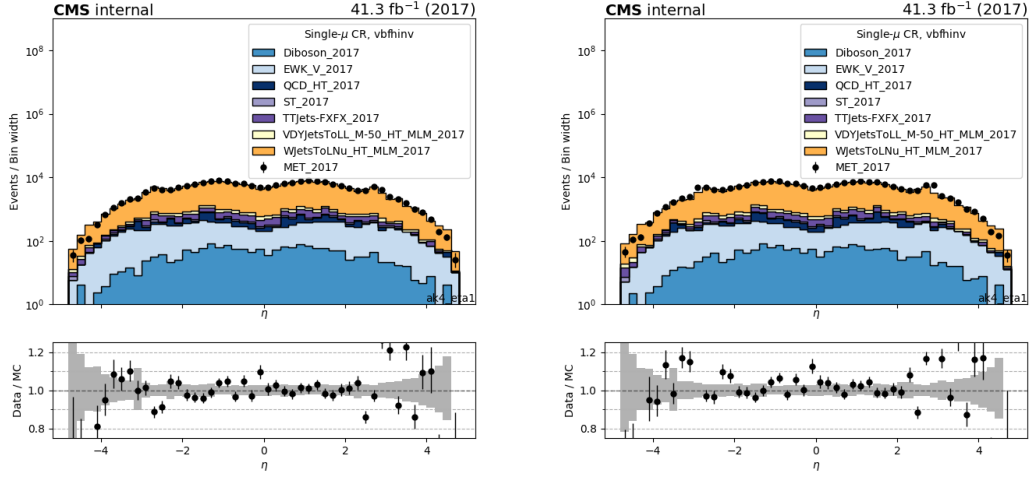


Figure 1: Subleading AK4 jet  $\eta$  distribution with pileup ID requirement (left) and without pileup ID requirement (right) in the single muon control region.

Lastly, to suppress the contributions due to non-collision backgrounds, the following requirements are applied on the leading AK4 jet:

- Charged hadron energy fraction  $> 0.1$
- Neutral hadron energy fraction  $< 0.8$

The effect of these cleaning requirements are demonstrated in the jet  $\phi$  distribution of the signal region events in Fig. 2.



Figure 2: Leading AK4 jet  $\phi$  distribution before and after applying the jet cleaning requirements in the signal region.

### 3.1.1 b-tagged jets

**Update the b-tagging working point when the new tagging working points are defined** In this analysis b jets with  $p_T > 20$  GeV and  $|\eta| < 2.4$  are identified using the “DeepCSV” algo-

rithm [? ], adopting a working point (medium) corresponding to correctly identifying a b-jet with a probability of 80%, and misidentifying a light-flavor jet with a probability of 10%. This working point corresponds to the value of DeepCSV tagger to be greater than 0.4941. Events with identified b jets are rejected to reduce the contamination from top quark processes.

### 3.1.2 V-tagged jets

Update the V-tagging working point when the new tagging working points are defined

## 3.2 Missing transverse momentum and recoil

The vector  $\vec{p}_T^{\text{miss}}$  is defined as the imbalance in the transverse momentum of all particles that interact with the detectors. Due to momentum conservation in the plane transverse to the beam axis,  $\vec{p}_T^{\text{miss}}$  corresponds to the transverse momentum that is carried by undetected particles such as neutrinos. Practically,  $\vec{p}_T^{\text{miss}}$  is computed as the negative of the vectorial sum of transverse momenta of all PF candidates and is therefore also referred to as PF  $\vec{p}_T^{\text{miss}}$ . The magnitude of the  $\vec{p}_T^{\text{miss}}$  is referred to as  $p_T^{\text{miss}}$ .

Minimum energy thresholds in the calorimeters, inefficiencies in the tracker, nonlinearity of the response of the calorimeter for hadronic particles can lead to an over- or underestimation of  $p_T^{\text{miss}}$ . The bias on the  $p_T^{\text{miss}}$  measurement is reduced by propagating the effect of the jet energy corrections introduced in section 3.1 according to

$$\vec{p}_T^{\text{miss}}(\text{corr}) = \vec{p}_T^{\text{miss}} - \sum_{\text{jets}} (\vec{p}_{T,\text{jet}}(\text{corr}) - \vec{p}_{T,\text{jet}}), \quad (1)$$

where the “corr” refers to the scale energy corrected measurements of the related objects.

This “type-I” correction for  $\vec{p}_T^{\text{miss}}$  uses jet energy scale corrections for all corrected jets with  $p_T > 15 \text{ GeV}$  that have less than 0.9 of their energy deposited in the ECAL. Furthermore, if a muon is found in a jet, its 4-momentum is subtracted from the 4-momentum of the jet when performing the correction and is added back to a corrected object.

Since signal events in this analysis contain only jets and no other reconstructed candidates,  $p_T^{\text{miss}}$  is equivalent to the total hadronic momentum in the event. For the leading backgrounds, this also corresponds to the transverse momentum of the W or Z boson. To mimic this behavior in the control regions of this analysis, the transverse momentum of the hadronic recoil  $\vec{U}$ , defined as the vectorial sum of the transverse momenta of all particles except the vector boson (or its decay products), is used. The variable is computed as

$$\vec{U} = \vec{p}_T^{\text{miss}} + \sum_{i \in \text{leptons, photons}} \vec{p}_T^i \quad (2)$$

where the sum takes into account the leptons and photons used to define the respective control region. The uncertainty of  $p_T^{\text{miss}}$  has a strong dependence on the event topology. Therefore, the uncertainty on  $p_T^{\text{miss}}$  is often factorized into its components of jets, leptons and unclustered energy. Each sub-component is then varied within its scale and resolution uncertainty. In this analysis, the largest contribution on the final  $p_T^{\text{miss}}$  uncertainty comes from the variations of the jet energy scale correction and the magnitude of the uncertainty is estimated to be 4% for the  $Z(\nu\nu) + \text{jets}$  events. **Up to date?**

Anomalous high- $p_T^{\text{miss}}$  events can appear due to various phenomena. In the ECAL, spurious deposits may appear due to particles striking sensors in the ECAL photodetectors, or from real

showers with non-collision origins such as those caused by beam halo particles. ECAL dead cells can cause real energy to be missed, again leading to a spurious imbalance. In the HCAL, spurious energy can arise due to noise in the hybrid photodiode and readout box electronics, as well as direct particle interactions with the light guides and photomultiplier tubes of the forward calorimeter. A number of filters has been developed by the POG/DPG groups to identify and suppress anomalous high  $p_T^{\text{miss}}$  events [? ]. The recommended filters are listed in Tab. 5 and are applied in the analysis.

Table 5: The  $p_T^{\text{miss}}$  filters recommended by the JME POG [? ]. The recommendations apply to both 2017 and 2018. Except for the bad super cluster filter (“ee badSC”), all filters are applied both in data and simulation.

Filter	Name in NanoAOD	Applied in data (MC)
primary vertex filter	Flag_goodVertices	✓(✓)
beam halo filter	Flag_globalSuperTightHalo2016Filter	✓(✓)
HBHE noise filter	Flag_HBHENoiseFilter	✓(✓)
HBHEiso noise filter	Flag_HBHENoiseIsoFilter	✓(✓)
ECAL TP filter	Flag_EcalDeadCellTriggerPrimitiveFilter	✓(✓)
Bad PF Muon Filter	Flag_BadPFMuonFilter	✓(✓)
ee badSC noise filter	Flag_eeBadScFilter	✓(×)
ECAL bad calibration filter update	Flag_ecalBadCalibFilterV2	✓(✓)

To further minimize the contribution of anomalous high- $p_T^{\text{miss}}$  events (specifically due to spurious charged hadrons) in this analysis, a quantity based on the relative ratio of calorimetry based  $p_T^{\text{miss}}$  and PF based  $p_T^{\text{miss}}$  is employed. Examples of the distribution in single muon and single electron control regions can be found in Fig. 3, which are produced after requiring the whole set of the standard  $p_T^{\text{miss}}$  filters provided by the POG. Events satisfying  $|E_{T \text{ calo}}^{\text{miss}} - E_{T \text{ PF}}^{\text{miss}}|/U < 0.5$  are selected in this analysis.

### 3.3 Leptons

#### 3.3.1 Electrons

Electrons within the geometrical acceptance of  $|\eta| < 2.5$  are reconstructed by associating tracks reconstructed in the silicon detector with clusters of energy in the ECAL [? ]. Well-identified electron candidates are required to satisfy additional identification criteria based on the shower shape of the energy deposit in the ECAL and the consistency of the electron track with the primary vertex [? ]. Electron candidates that are identified as coming from photon conversions in the detector material are removed. An isolation variable is calculated based on the sum of the energies of the PF candidates within a cone of  $\Delta R < 0.3$  around the electron. The mean energy deposit in the isolation cone of the electron coming from pileup is estimated following the method described in Ref. [? ] and subtracted from the isolation sum. In this note, ‘veto’ [? ] electrons with a minimum  $p_T$  of 10 GeV are selected with an average efficiency of 95% and their presence is used as a condition to reject events, whereas ‘tight’ [? ] electrons with a minimum  $p_T$  of 40 GeV and an average efficiency of 70% are used to select the events in the control regions. Full selection criteria are shown in Table 6.

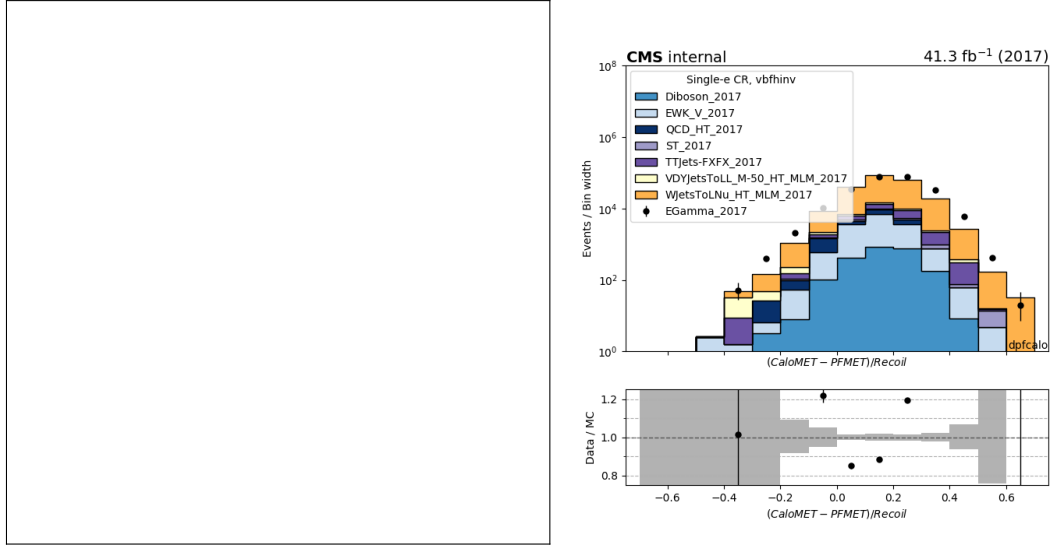


Figure 3: The difference between calorimeter-based and PF MET relative to the hadronic recoil in data and simulated events passing the analysis selection for the single muon (left) and single electron (right) control regions. Events with values of  $< 0.5$  are considered in the analysis.

### 3.3.2 Muons

Muons within the geometrical acceptance of  $|\eta| < 2.4$  are reconstructed by combining information from the silicon tracker and the muon system [?]. The muons are required to pass set of quality criteria based on the number of spatial points measured in the tracker and in the muon system, the fit quality of the muon track and its consistency with the primary vertex of the event. Similar to electron case, the isolation requirements for muons are also based on the sum of the energies of the PF candidates, but a different cone size of a  $\Delta R < 0.4$  is used. The muon isolation variable is corrected for pileup effects by subtracting half of the sum of the transverse momenta of charged particles that are inside the isolation cone and not associated with the primary vertex. In this note, “loose” [?] muons with  $p_T > 10$  GeV are selected with an average efficiency of 98% and are used as a condition to reject events, whereas “tight” [?] muons with  $p_T > 20$  GeV are selected with an average efficiency of 95% and are used to select events in the control samples. A full list of tight identification criteria is given here:

- Muon reconstructed as a global muon
- Muon reconstructed as a particle flow muon
- Normalized  $\chi^2$  of the global track less than 10
- At least one muon chamber hit included in the global track fit
- Muon segments in at least two muon stations
- Transverse impact parameter w.r.t. the primary vertex less than 2 mm.
- Longitudinal impact parameter w.r.t. the primary vertex less than 5 mm.
- At least one pixel hit
- Hits on at least 5 tracker layers
- $\Delta\beta$  relative isolation less than 0.15

Table 6: Tight and veto electron identification criteria.

Variable	Selection Tight Barrel (Endcaps)	Selection Veto Barrel (Endcap)
Full 5x5 $\sigma_{i\eta i\eta}$	$< 0.0104$ ( $< 0.0353$ )	$< 0.0126$ ( $< 0.0457$ )
$ \Delta\eta_{in} $	$< 0.00255$ ( $< 0.00501$ )	$< 0.00463$ ( $< 0.00814$ )
$ \Delta\phi_{in} $	$< 0.022$ ( $< 0.0236$ )	$< 0.148$ ( $< 0.19$ )
H/E	$< 0.026 + 1.15/E_{SC} + 0.0324\rho/E_{SC}$ ( $< 0.0188 + 2.06/E_{SC} + 0.183 * \rho/E_{SC}$ )	$< 0.05 + 1.16/E_{SC} + 0.0324\rho/E_{SC}$ ( $< 0.05 + 2.54/E_{SC} + 0.183\rho/E_{SC}$ )
Relative isolation ( $\rho$ correction)	$< 0.0287 + 0.506/p_T$ ( $< 0.0445 + 0.963/p_T$ )	$< 0.198 + 0.506/p_T$ ( $< 0.203 + 0.963/p_T$ )
1/E - 1/p	$< 0.159$ ( $< 0.0197$ )	$< 0.209$ ( $< 0.132$ )
$ d_{xy}(vtx) $	$< 0.050$ ( $< 0.100$ )	$< 0.050$ ( $< 0.100$ )
$ d_z(vtx) $	$< 0.100$ ( $< 0.200$ )	$< 0.100$ ( $< 0.200$ )
Expected Inner Missing Hits	$\leq 1$ ( $\leq 1$ )	$\leq 2$ ( $\leq 3$ )
Pass conversion veto	Yes (Yes)	Yes (Yes)

### 3.3.3 Taus

Hadronically decaying  $\tau$  leptons are required to pass identification criteria using the hadron-plus-strips algorithm [? ]. The algorithm identifies a jet as an hadronically decaying tau lepton candidate if a subset of the particles assigned to the jet is consistent with the decay products of a  $\tau$  candidate. Candidate  $\tau$  jets are required to pass both the “DecayModeNewDMs” and “DecayMode” identifiers.

In addition,  $\tau$  candidates are required to be isolated from other activity in the event. The isolation requirement is computed by summing the  $p_T$  of the charged PF candidates and PF photon candidates within an isolation cone of  $\Delta R = 0.5(0.3)$ , around the tau candidate direction. The charged and photon candidates associated with the tau candidate are removed from this sum and further described in Ref. [? ]. The “VLoose.IsolationMVArun2v1DBnewDMwLT” isolation working point [? ] is employed in this analysis for tau candidates with  $p_T$  larger than 18 GeV within  $|\eta| < 2.3$ .

### 3.4 Photons

Photon candidates are reconstructed from energy deposits in the ECAL using algorithms that constrain the clusters to the size and shape expected from a photon [? ]. The identification of the candidates is based on shower-shape and isolation variables. For isolated photons, scalar sums of the  $p_T$  of PF candidates within a cone of  $\Delta R < 0.3$  around the photon candidate are required to be below the bounds defined. Only the PF candidates that do not overlap with the EM shower of the candidate photon are included in the isolation sums. The photon candidates used in this analysis are required to have a minimum transverse momentum of 15 GeV and to be within  $|\eta| < 2.5$  passing the ‘loose’ [? ] identification criteria in. The full identification criteria is also given in Table 7.

Variable	Selection
	Barrel (Endcap)
Full 5x5 $\sigma_{i\eta i\eta}$	$< 0.0106 (< 0.0272)$
H/E	$< 0.04596 (< 0.0590)$
charged hadron isolation	$< 1.694 (< 2.089)$
neutral hadron isolation	$< 24.032(19.722) + 0.01512(0.0117) \times p_T + 2.259(2.3) \times 10^{-5} \times p_T^2$
photon isolation	$< 2.876(4.162) + 0.004017(0.0037) \times p_T$
Conversion safe electron veto	Yes (Yes)

Table 7: Loose photon identification criteria.

#### 3.4.1 Photon purity studies

To be added.

## 4 Reweighting of simulated events

Simulated signal and background samples are corrected for various effects through reweighting procedures outlined in this section.

### 4.1 Trigger efficiency reweighting

#### 4.1.1 $p_T^{\text{miss}} + H_T^{\text{miss}}$ triggers

The performance of the  $p_T^{\text{miss}} + H_T^{\text{miss}}$  triggers is measured using single muon events. The events are selected from the SingleMuon using the () trigger for 2017 (2018), and the offline muon is required to be well-identified and have  $p_T$  larger than 40 GeV. The same selection is required as for the single-muon control region used in the final fit (cf. sec. 5.2):

1. Veto on additional leptons, photons, b jets,  $\tau_{had}$  candidates.
2.  $\Delta\phi(jet, \vec{p}_T^{\text{miss}}) > 0.5$  for the four leading jets with  $p_T > 30$  GeV.
3.  $(\text{Calo } p_T^{\text{miss}} - \text{PF } p_T^{\text{miss}}) / \text{recoil} \geq 0.5$
4.  $M_T(\ell, p_T^{\text{miss}}) > 160$  GeV.
5. Central AK4 jet with  $p_T > 100\text{GeV}$ , passing the tight jet ID.

The efficiency is calculated as a function of the hadronic recoil  $p_T$  and is shown in Fig. 4. The trigger is found to be more than 95% efficient for events with a recoil larger than 250 GeV, and more than 99% efficient for events with a recoil larger than 375 GeV. The MC-to-data scale factor is found to be within 1% of unity everywhere except for the lowest recoil bin at 250 GeV, where it is within 2%.

To investigate the stability of the measurement in single muon events, the same method is used to extract the efficiency from samples of double muon and single electron events. Again, an identical selection to the analysis control regions is used (cf. sec. 5.4 and 5.3), with the exception of requiring the leading muon  $p_T$  to be larger than 40 GeV and omitting the  $p_T^{\text{miss}}$  cut in the electron region. The  $H_T$ -binned and simulation samples are used. The resulting data-to-MC efficiency scale factors for all regions are shown in Fig. 5. For 2017, a clear trend is present: The scale factor in dimuon events is larger in absolute terms than the one from the single muon region, and the scale factor from single electron events is the smallest. The difference between all regions is within 1% relative to the single muon region. For 2018, no clear trend is observed above a recoil of 250 GeV: The scale factors from all three regions agree within the available statistical precision. Finally, the scale factors obtained from the single muon region are used to reweight the simulation. The difference to the other regions is taken into account by assigning an overall 1% uncertainty.



Figure 4: MET trigger turn-on curve measured in single muon events as a function of hadronic recoil  $p_T$  in the 2017 and 2018 datasets and  $H_T$ -binned simulation samples. The vertical blue line indicates a recoil value of 250 GeV, which is the requirement used in the analysis selection. The bottom panel shows the MC-to-data scale factor, with the blue horizontal lines indicating deviations of 1% and 2% from unity, respectively.

#### 4.1.2 Photon trigger

The photon trigger efficiency is measured using events from the JetHT dataset collected with the trigger, which was fully unscaled in 2017 and 2018<sup>1</sup>. Events are selected in the same way as for the photon analysis control region (cf. sec. 5.6), except for the photon  $p_T$ , recoil and trigger requirements. The trigger efficiency  $\epsilon$  is then determined as:

<sup>1</sup>The other, prescaled paths yield lower statistical precision.



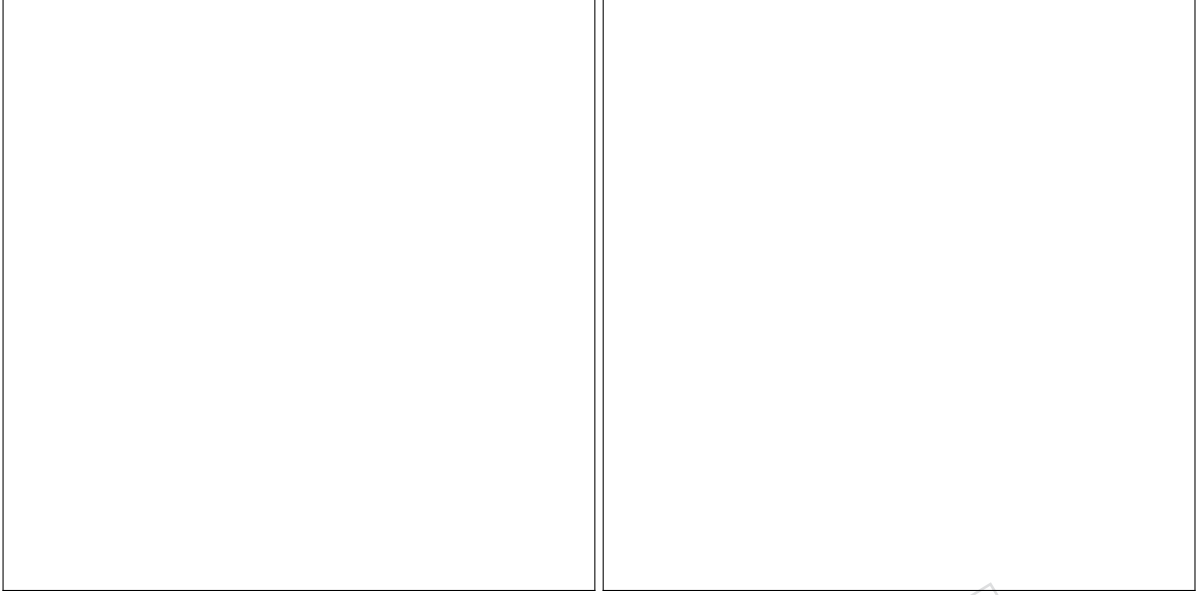


Figure 5: MC-to-data efficiency scale factors measured single muon events ('1m'), dimuon events ('2m') and single electron events ('1e'). For the single-lepton regions, the  $H_T$ -binned are used, and the  $H_T$ -binned samples are used for the dimuon region. The bottom panel shows the ratio of the scale factors obtained from each individual region to the scale factor obtained from the single muon region. The vertical blue line indicates a recoil value of 250 GeV, which is the requirement used in the analysis selection, while the horizontal lines indicate deviations of  $\pm 1\%$  from unity.

$$\epsilon() = \frac{\text{Offline selection \&\& \&\&}}{\text{Offline selection \&\&}}$$

The resulting efficiency in data and  $H_T$ -binned simulation is shown in Fig. 6. The trigger efficiency in data is larger than 95% for a photon  $p_T$  of larger than 215 GeV, and larger than 99% for photon  $p_T$  larger than 400 GeV. Between 250 and 400 GeV, there is a slight inefficiency amounting to approximately 1% at the most, with a larger amplitude in 2017 than in 2018. In both years, the turn-on behavior is almost immediate in simulated events, resulting in an MC-to-data scale factor almost entirely driven by the efficiency in data. The scale factor is within 1% of unity for all bins except the lowest 2017 bin at 215 GeV, where it deviates from unity by about 4%.

Based on these results, the offline  $p_T$  cut for the photon in the photon control region is chosen to be 215 GeV.

#### 4.1.3 Electron trigger

We also need to add information about the single electron trigger efficiencies

## 4.2 Pileup reweighting

The pileup (PU) conditions in the simulated samples are not identical to the ones observed measured in data, and a reweighting is applied to remove the difference. The reweighting is performed by matching the true pileup distribution of each simulated sample with the pileup distribution in data, obtained through the pileupCalc tool assuming a minimum bias cross

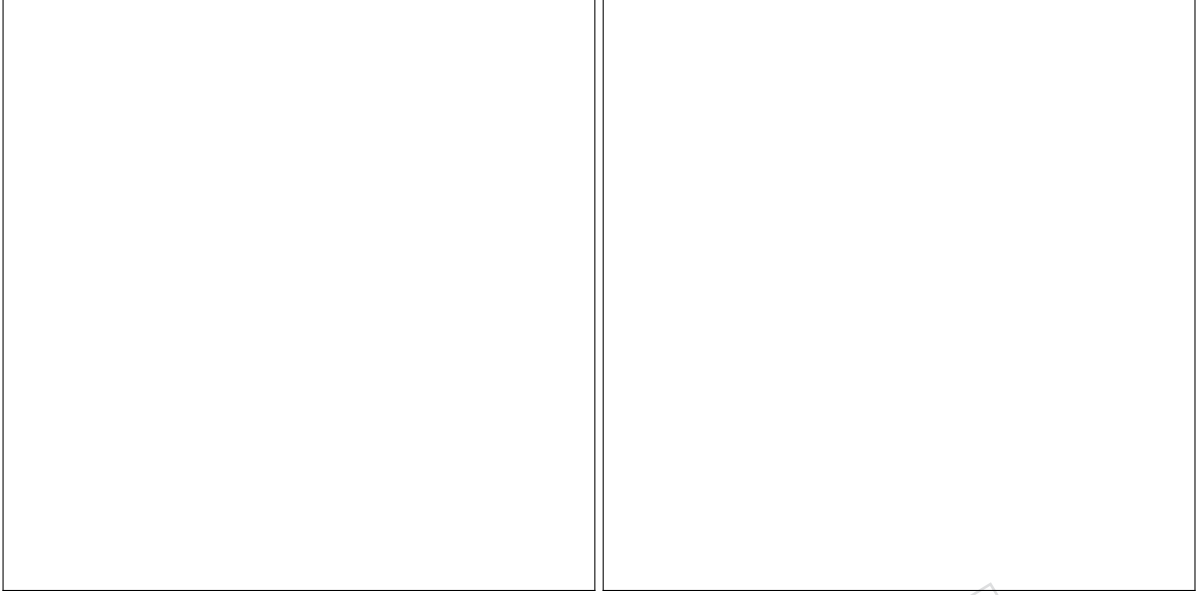


Figure 6: Efficiency of the trigger in data (black) and  $H_T$ -binned simulation (red) for 2017 (left) and 2018 (right) as a function of photon  $p_T$ . The bottom panel shows the MC-to-data efficiency scale factor. The blue vertical line indicates a photon  $p_T$  of 215 GeV, which is the requirement used in the analysis selection.

section of  $69.2 \pm 4.6\%$  mb, following the recommendations in in Ref. [? ]. The true pileup distributions in data and simulation are shown in Fig. 7. The distribution of the number of reconstructed vertices for  $W \rightarrow \mu\nu$  events before and after PU reweighting is shown in Fig. 8. In this variable, the PU reweighting method leads to a worse overall agreement between data and simulation. To check this behavior, the distribution of the event energy density  $\rho$  is shown in Fig. 8, again before and after PU reweighting. Here, the agreement before PU reweighting is worse than in the primary vertex distribution and the PU reweighting clearly improves the agreement.

### 4.3 Lepton and photon identification/reconstruction efficiency reweighting

Data-to-simulation scale factors are applied to events in the control regions to account for differences in the reconstruction, identification and isolation of leptons between data and simulation. These data-to-MC scale factors are derived from the efficiencies that are measured for the electron and muon selections in bins of  $p_T$  and  $\eta$  in both data and simulation. These scale factors are provided by the relevant POGs.

The reconstruction scale factors for electrons are shown in Fig. 10. The corresponding identification scale factors for veto and tight electrons are shown in Fig. 11, and include the effect of the isolation efficiency.

The identification scale factors for muons are shown in Fig. 12. Here, isolation scale factors are applied separately and are shown in Fig. ???. The corresponding corrections for muons are deemed negligible [? ].

The scale factors for id and isolation for tight muons are shown in Fig. ??.

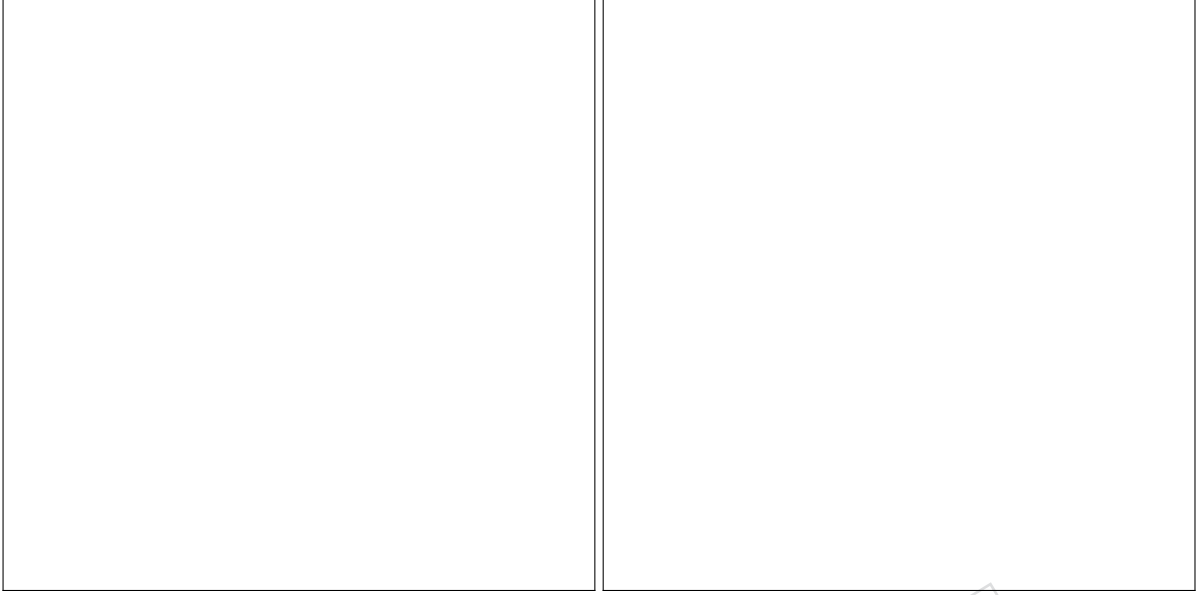


Figure 7: Distribution of the true number of PU events in data and simulation for 2017 (left) and 2018 (right). The distributions for data are extracted assuming a minimum bias cross section of 69.2 mb.

#### 4.4 Higher-order reweighting

This analysis uses the ratios of the recoil distributions in signal and control regions to constrain the final background estimate in a partially data driven way. As signal and control regions both have large statistical power, precise predictions of these ratios are necessary. To achieve this goal, the LO simulation samples for the samples W, DY and photon backgrounds are reweighted using higher-order corrections separately corresponding to NLO QCD, NLO EW and NNLO QCD terms.

#### 4.5 Generator-level boson construction

All theory-based corrections of the W, DY and photon backgrounds are parametrized as a function of the generator-level  $p_T$  of the respective boson  $p_{T,V}$ . For each simulated event, this quantity is calculated as follows. For DY and W samples, generator-level dilepton candidates are built from:

1. “dressed” final-state electrons and muons. Lepton dressing means to collect all photons radiated off the lepton within a cone of  $\Delta R < 0.1$  and adding their four-momenta back to the lepton four-momentum. This procedure is meant to undo the effect of final state photon radiation, which would otherwise distort the value of the reconstructed boson four-momentum. This effect is especially relevant as electrons and muons follow different radiation patterns. Lepton dressing is performed in central NanoAOD production following the procedure used in the RIVET software.
2.  $\tau$  leptons with generator status 2. As  $\tau$  leptons are unstable, they are not present as final state particles (status 1) in the generator record. The  $\tau$  lepton before its decay has status 2.
3. neutrinos with generator status 1.

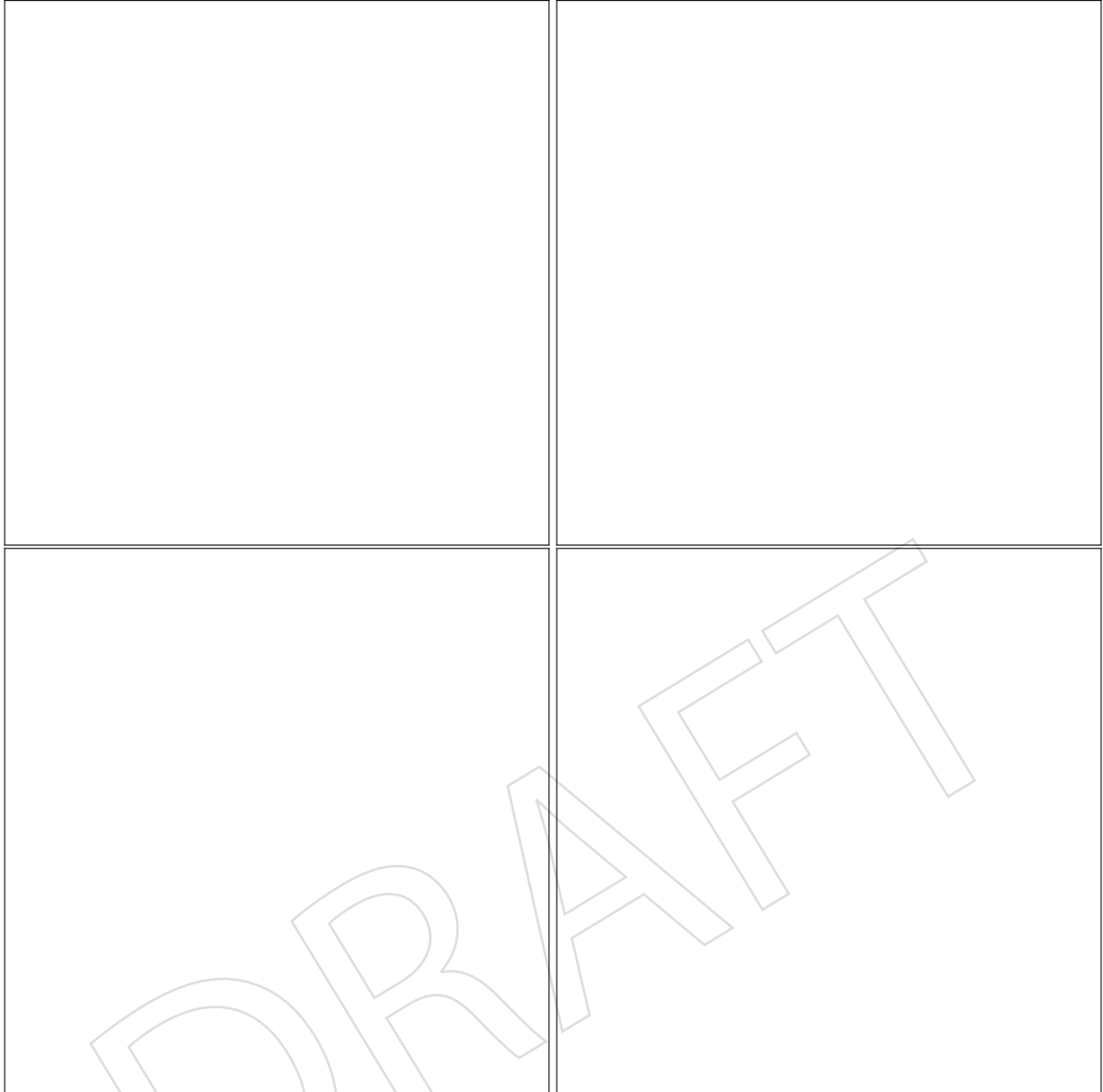


Figure 8: Distribution of the number of vertices in  $W \rightarrow \mu\nu$  events in data and simulation before pileup re-weighting (left) and after pileup reweighting (right). The Monte Carlo is normalized to the luminosity of 41.53 and 59.7  $\text{fb}^{-1}$ , respectively for 2017 and 2018.

375 The dilepton candidates are checked for flavour consistency with the desired boson candidate.  
 376 If multiple candidates are found in an event, the one with the highest invariant mass is used.  
 377 For photon events, the generator photon with highest  $p_T$  and status 1 is used.

#### 378 4.5.1 EW NLO

379 Scale factors corresponding to NLO EW corrections are obtained from Ref. [?] and applied as  
 380 a function of the generator-level boson  $p_T$ . The scale factors are shown in Fig. 15.

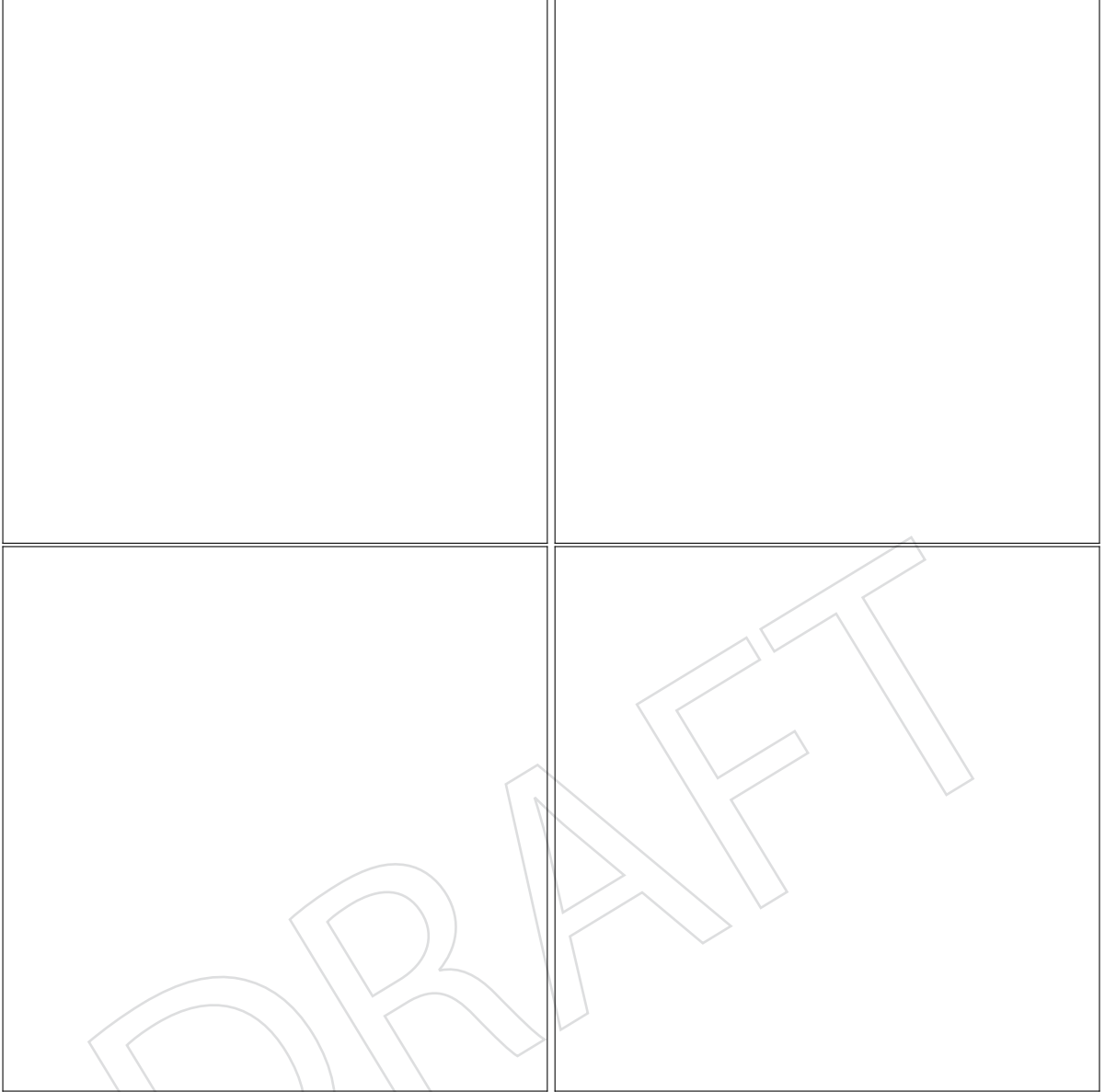


Figure 9: Distribution of the event energy density  $\rho$  in  $W \rightarrow \mu\nu$  events in data and simulation before pileup re-weighting (left) and after pileup reweighting (right). The Monte Carlo is normalized to the luminosity of 41.53 and 59.7  $\text{fb}^{-1}$ , respectively for 2017 and 2018.

#### 4.5.2 QCD NLO

Scale factors corresponding to NLO QCD corrections for W and Z production are obtained from central CMS samples produced in the “Fall17” campaign, in which large LO and NLO samples are available. Both samples are generated using . The LO samples are identical to the ones used in in the analysis, and are generated with up to four partons in the matrix element. The NLO samples are generated with up to two additional partons in the matrix element calculation. Further jet multiplicities are handled by the parton shower, which in both cases is performed using with tune .

The scale factors are derived by obtaining the generator-level boson  $p_T$  distribution in both samples, normalizing the distributions to their respective cross sections, and then dividing

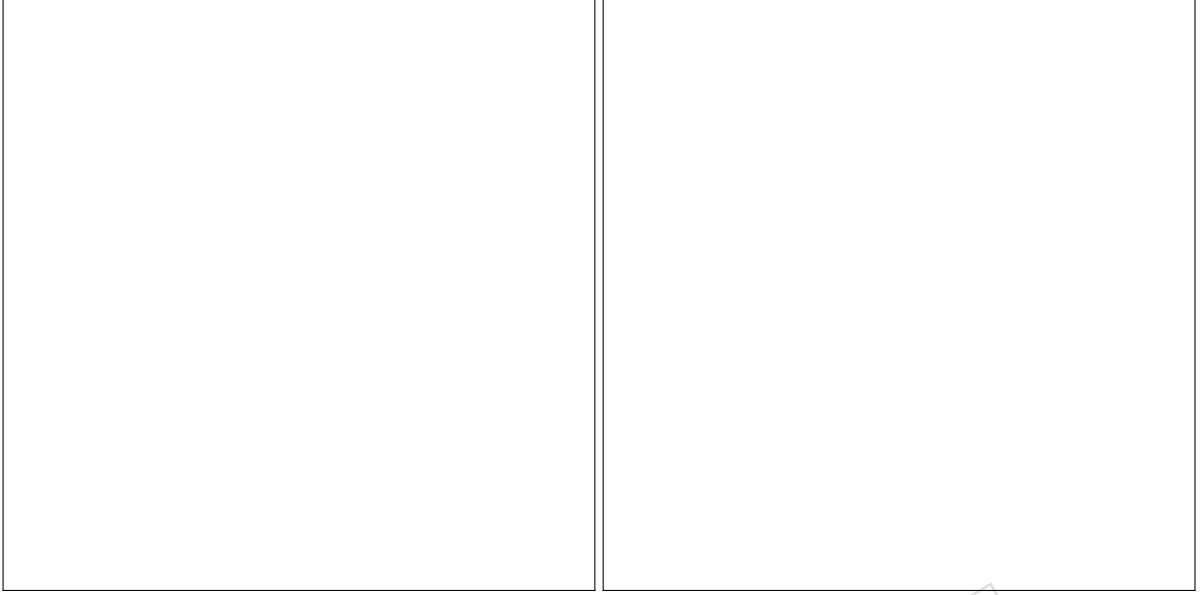


Figure 10: Scale factors for the reconstruction efficiency of electrons starting from a super cluster for 2017 (left) and 2018(right)

them as  $SF = NLO / LO$ . Identical selection criteria are applied to both samples based on the generator-level boson and generator-level AK4 jets, which are clustered using all visible generator particles with status 1. The requirements are:

1. The leading jet must have  $p_T > 100$  GeV and  $|\eta| < 2.4$ .
2. The difference in the azimuthal angle ( $\Delta\phi$ ) between the boson and the four leading jets in the event is required to be larger than 0.5. Only jets with  $|\eta| < 2.4$  and  $p_T > 30$  GeV are considered.

Compared to an inclusive derivation of the SF, the inclusion of the selection criteria leads to an approximately constant increase in the SF of about 2%.

The resulting SFs are shown in Fig. 14 for DY and W production. To protect the outcome of the reweighting procedure from binning effects, the binned scale factor is interpolated using a falling exponential function:

$$SF = a \times \exp(-b \times p_T) + c, \quad (3)$$

where  $p_T$  is the boson transverse momentum and  $a$ ,  $b$  and  $c$  are determined by a fit to the scale factor histogram. The resulting values of the fit parameters, as well as the resulting interpolated shape are also shown in Fig. 14. A comparison of the agreement of data and simulation between the reweighted LO samples and the NLO samples is shown in Appendix ??, where good agreement between the two methods is observed.

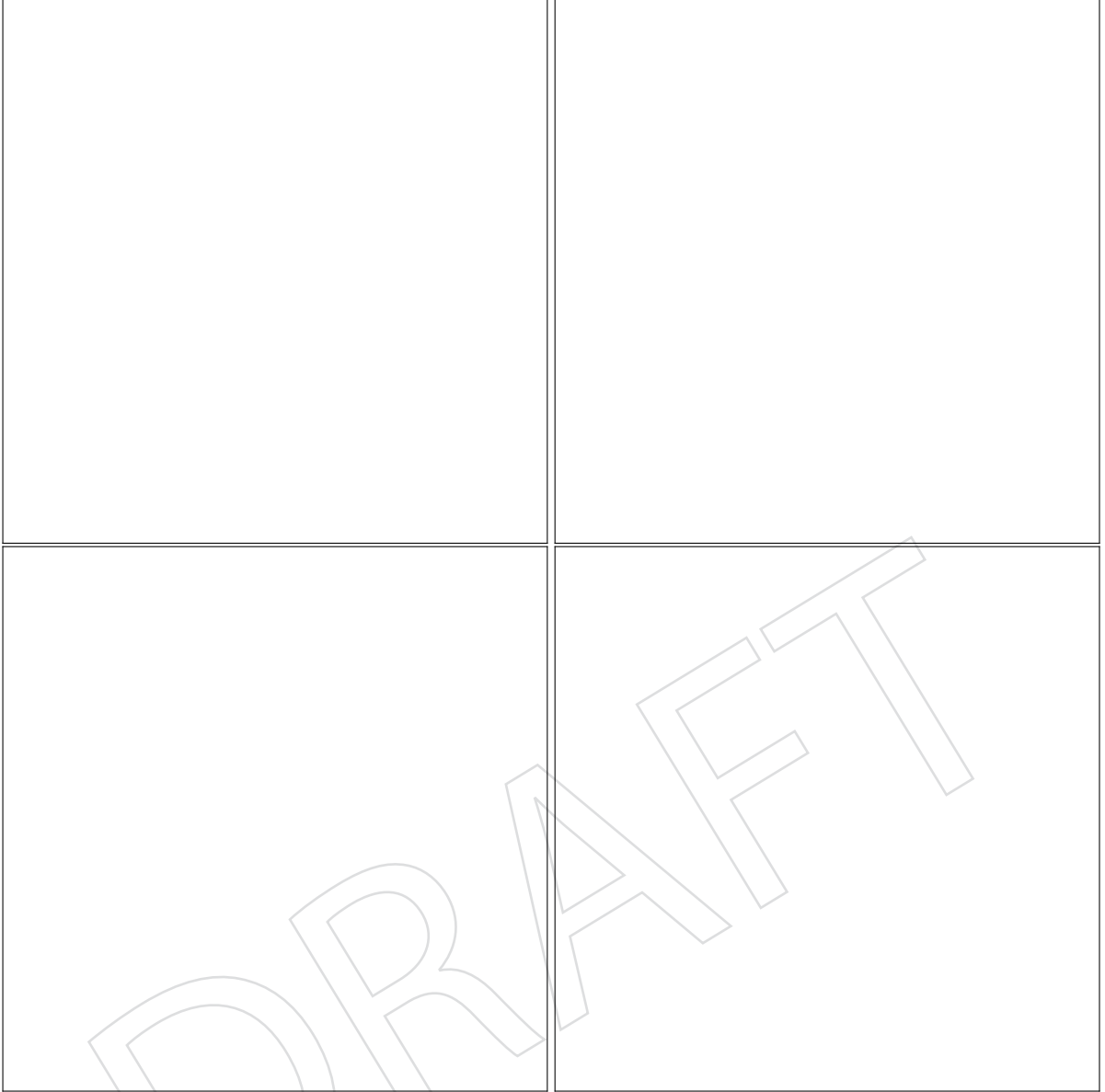


Figure 11: Scale factors for tight (left) and veto (right) electrons are shown for 2017 (top) and 2018 (bottom). The scale factors are provided in bins of electron  $p_T$  and  $\eta$ .

### 4.5.3 QCD NNLO

## 5 Event selection

### 5.1 Signal region selection

Signal region events are selected using triggers with thresholds of 120 GeV on both  $p_{T,\text{trig}}^{\text{miss}}$  and  $H_{T,\text{trig}}^{\text{miss}}$ . The  $p_{T,\text{trig}}^{\text{miss}}$  corresponds to the magnitude of the vector  $\vec{p}_T$  sum of all the PF candidates reconstructed at the trigger level, while the  $H_{T,\text{trig}}^{\text{miss}}$  is computed as the magnitude of the vector  $\vec{p}_T$  sum of jets with  $p_T > 20 \text{ GeV}$  and  $|\eta| < 5.0$  reconstructed at the trigger level. The energy fraction attributed to neutral hadrons in these jets is required to be smaller than 0.9. This requirement suppresses anomalous events with jets originating from detector noise. To be able to use the same triggers for selecting events in the muon control samples used for background



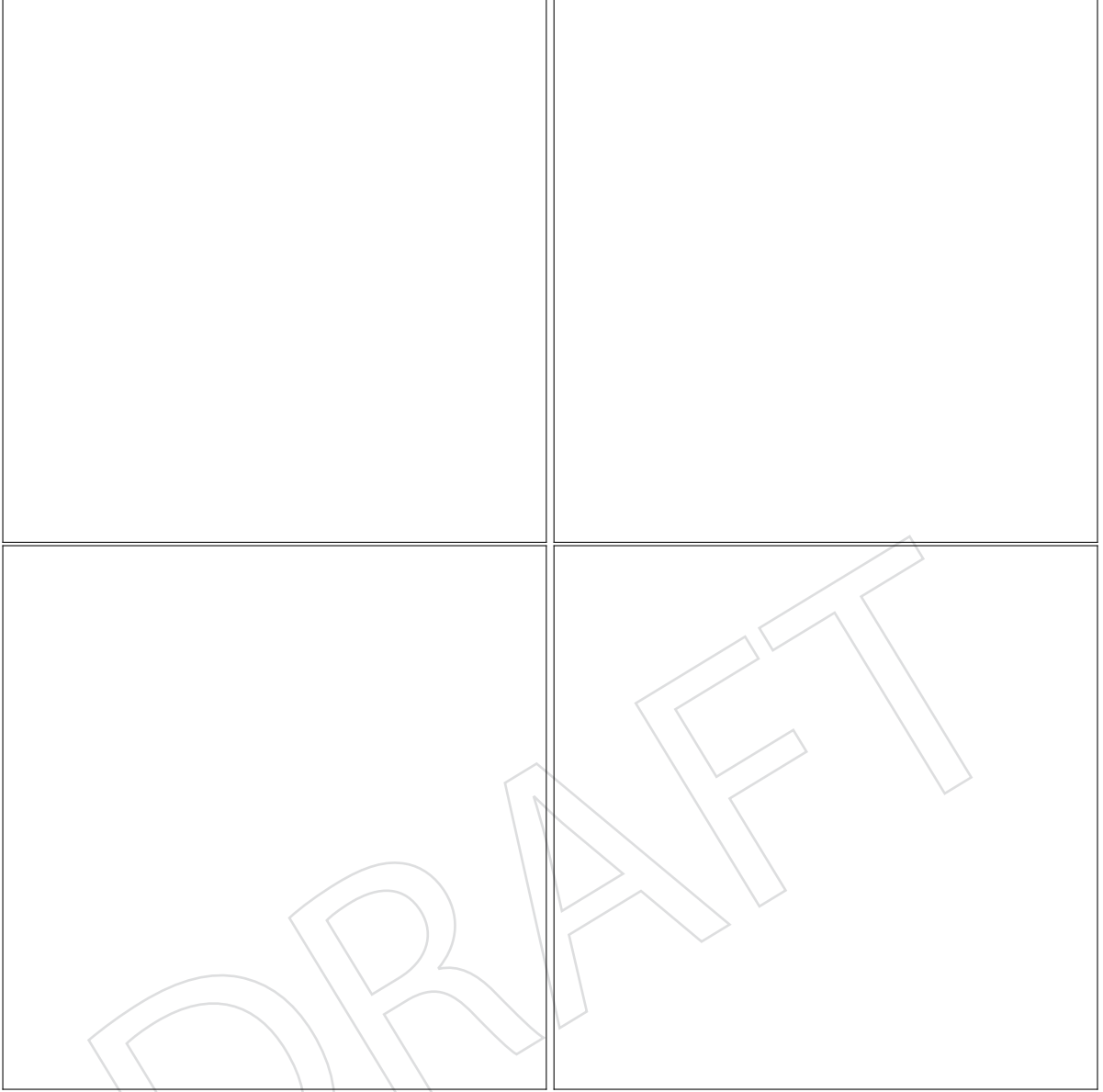


Figure 12: Scale factors for tight (left) and veto (right) muon identification are shown for 2017 (top) and 2018 (bottom). The scale factors are provided in bins of electron  $p_T$  and  $\eta$ .

415 prediction, muon candidates are not included in the  $p_{T,\text{trig}}^{\text{miss}}$  nor  $H_{T,\text{trig}}^{\text{miss}}$  computation. The trigger  
 416 efficiency is measured to be 96% for events passing the analysis selection for  $p_T^{\text{miss}} > 250$  GeV  
 417 and becomes more than 99% efficient for events with  $p_T^{\text{miss}} > 350$  GeV.

418 Candidate events are required to have  $p_T^{\text{miss}} > 250$  GeV. The leading AK4 jet in the signal event  
 419 is required to have  $p_T > 80$  GeV and  $|\eta| < 4.7$ , and the subleading AK4 jet is required to have  
 420  $p_T > 40$  GeV and  $|\eta| < 4.7$ . In addition, if the leading jet is within the tracker range,  $|\eta| < 2.5$ ,  
 421 it is required to have at least 10% of its energy coming from charged particles and less than 80  
 422 % of its energy attributed to neutral hadrons, as discussed in section 3. This selection helps to  
 423 remove events originating from beam-induced backgrounds. In addition, the analysis employs  
 424 various event filters to reduce events with large misreconstructed  $p_T^{\text{miss}}$  [?] originating from  
 425 noncollision backgrounds.

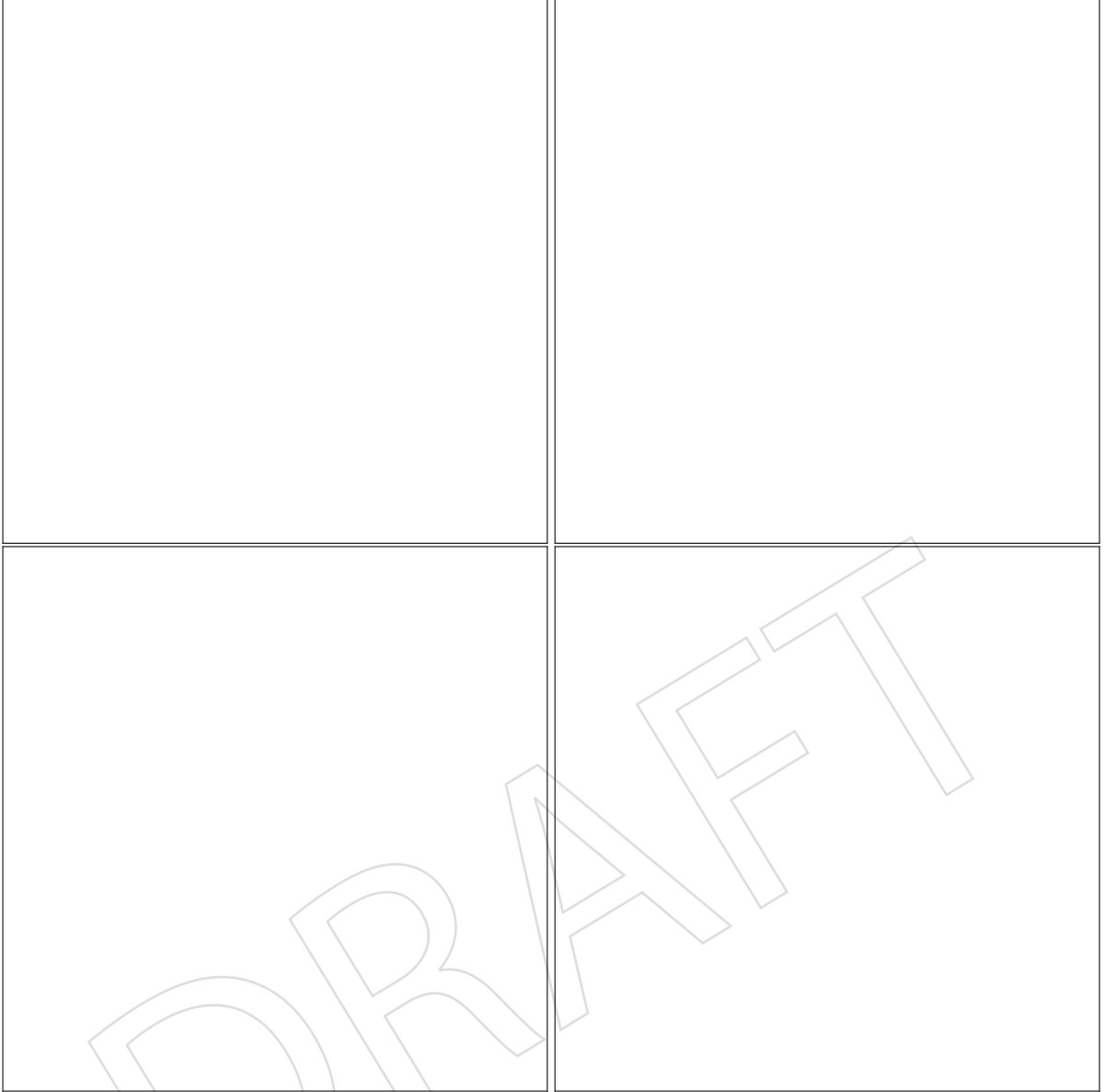


Figure 13: Scale factors for tight (left) and veto (right) muon isolation are shown for 2017 (top) and 2018 (bottom). The scale factors are provided in bins of electron  $p_T$  and  $\eta$ .

For the VBF signal events, two leading jets in opposite hemispheres are expected, with large dijet mass. Furthermore, these jets are expected to have large rapidity separation and small azimuthal separation. Therefore, this analysis employs several requirements on  $M_{jj}$ ,  $\Delta\eta_{jj}$  and  $\Delta\phi_{jj}$ , which can be found in Table 1.

426 The main background processes in this search are the  $Z(\nu\nu) + \text{jets}$  and  $W(\ell\nu) + \text{jets}$  processes.  
 427 The  $Z(\nu\nu) + \text{jets}$  process is an irreducible background and constitutes the largest background  
 428 in the search. In contrast, the background from  $W(\ell\nu) + \text{jets}$  is suppressed by imposing a veto  
 429 on events containing one or more loose muons or electrons with  $p_T > 10$  GeV, or hadronically  
 430 decaying  $\tau$  leptons with  $p_T > 18$  GeV. Events that contain a loose, isolated photon with  $p_T >$   
 431  $15$  GeV and  $|\eta| < 2.5$  are also rejected. This helps to suppress electroweak (EW) backgrounds  
 432 in which a photon is radiated from the initial state. To reduce the contamination from top  
 433 quark backgrounds, events are rejected if they contain a b tagged jet with  $p_T > 20$  GeV and  
 434  $|\eta| < 2.4$ . These jets are identified using the DeepCSV algorithm [??], adopting the “medium”

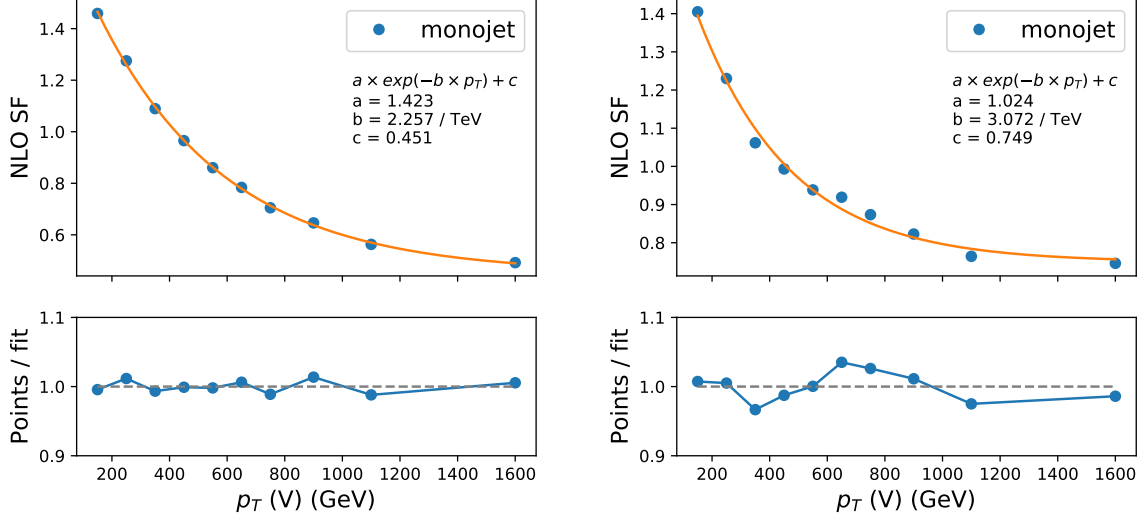


Figure 14: QCD NLO scale factors for the DY (left) and W processes (right) after applying a monojet-like selection on generator level. In the top panel, the blue markers show the NLO SF derived from the simulated samples. The orange line shows a fit function used to interpolate the SF. The functional form and resulting parameters are given in the figure. In the bottom panel, the blue markers show the ratio of the histogram to the fit result in each bin.

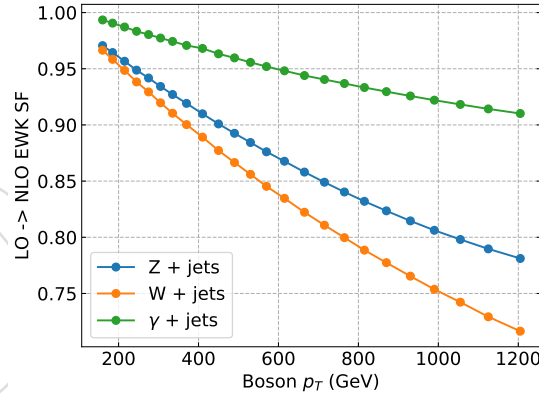


Figure 15: EW NLO scale factors for DY, W and photon production as a function of  $p_{T,V}$ .

working point, which corresponds to correctly identifying a jet originating from a bottom quark with a probability of 80% and misidentifying a jet originating from a charm quark (light-flavor jet) with a probability of 12 (2)%. Lastly, QCD multijet background with  $E_T^{\text{miss}}$  arising from mismeasurements of the jet momenta is suppressed by requiring the minimum azimuthal angle between the  $\vec{p}_T^{\text{miss}}$  direction and each of the first four leading jets with  $p_T$  greater than 30 GeV and  $|\eta| < 2.4$  to be larger than 0.5 radians.

The selection requirements for this analysis are summarized in Table 8.

The  $N$ -subjettiness variable  $\tau_N$  [?] is also employed to further isolate jets arising from hadronic decays of W or Z bosons. This observable measures the distribution of jet constituents relative

to candidate subjet axes in order to quantify how well the jet can be divided into  $N$  subjets. Therefore, the ratio of the ‘2-subjettiness’ to the ‘1-subjettiness’ ( $\tau_2/\tau_1$ ) has excellent capability for distinguishing jets originating from boosted vector bosons from jets originating from light quarks and gluons. The pruned jet mass and  $N$ -subjettiness requirements, whose use is referred to as V tagging, result in a 70% efficiency for tagging jets originating from V bosons and a 5% probability of misidentifying a jet as a V jet. Events that do not qualify for the mono-V category are assigned to the monojet category.

Table 8: Summary of the common selection requirements

Variable	Selection	Target background
Muon (electron) veto	$p_T > 10 \text{ GeV},  \eta  < 2.4(2.5)$	$Z(\ell\ell) + \text{jets}, W(\ell\nu) + \text{jets}$
$\tau$ lepton veto	$p_T > 18 \text{ GeV},  \eta  < 2.3$	$Z(\ell\ell) + \text{jets}, W(\ell\nu) + \text{jets}$
Photon veto	$p_T > 15 \text{ GeV},  \eta  < 2.5$	$\gamma + \text{jets}$
Bottom jet veto	DeepCSV medium $< 0.4941/0.4184$ (2017 / 2018) for all jets with $p_T > 20 \text{ GeV},  \eta  < 2.4$	Top quark
$p_T^{\text{miss}}$	$> 250 \text{ GeV}$	QCD, top quark, $Z(\ell\ell) + \text{jets}$
$\Delta\phi(\vec{p}_T^{\text{jet}}, \vec{p}_T^{\text{miss}})$	$> 0.5 \text{ radians}$	QCD
Leading AK4 jet $p_T$ and $\eta$	$> 80 \text{ GeV}$ and $ \eta  < 4.7$	All
Subleading AK4 jet $p_T$ and $\eta$	$> 40 \text{ GeV}$ and $ \eta  < 4.7$	All
$M_{jj}$	$> 200 \text{ GeV}$	$\Delta\eta_{jj}$
$< 1.5$ height		$> 1.0 \Delta\phi_{jj}$

Fig. 16, 17, 18, 19, shows the distribution of the  $E_T^{\text{miss}}$ , the number of jets,  $p_T$  and  $\eta$  distribution of the leading AK4 jet for events in the monojet and mono-V signal categories respectively for 2017 and 2018 datasets.

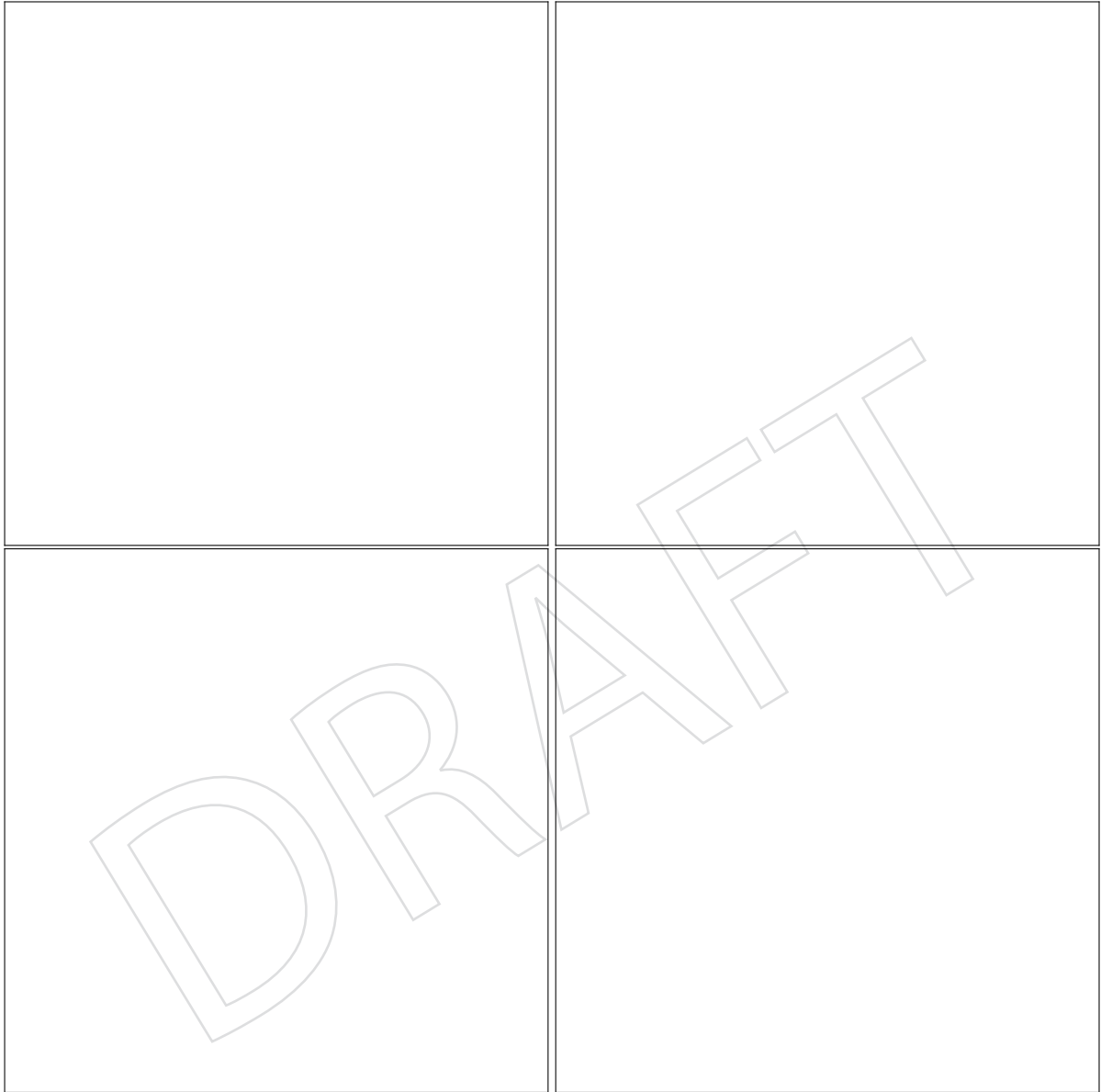


Figure 16: Comparison between data and monte carlo simulation in the monojet signal region for the recoil distribution, the AK4 jet multiplicity distribution,  $p_T$  and  $\eta$  distribution of the leading AK4 jet in 2017 dataset.

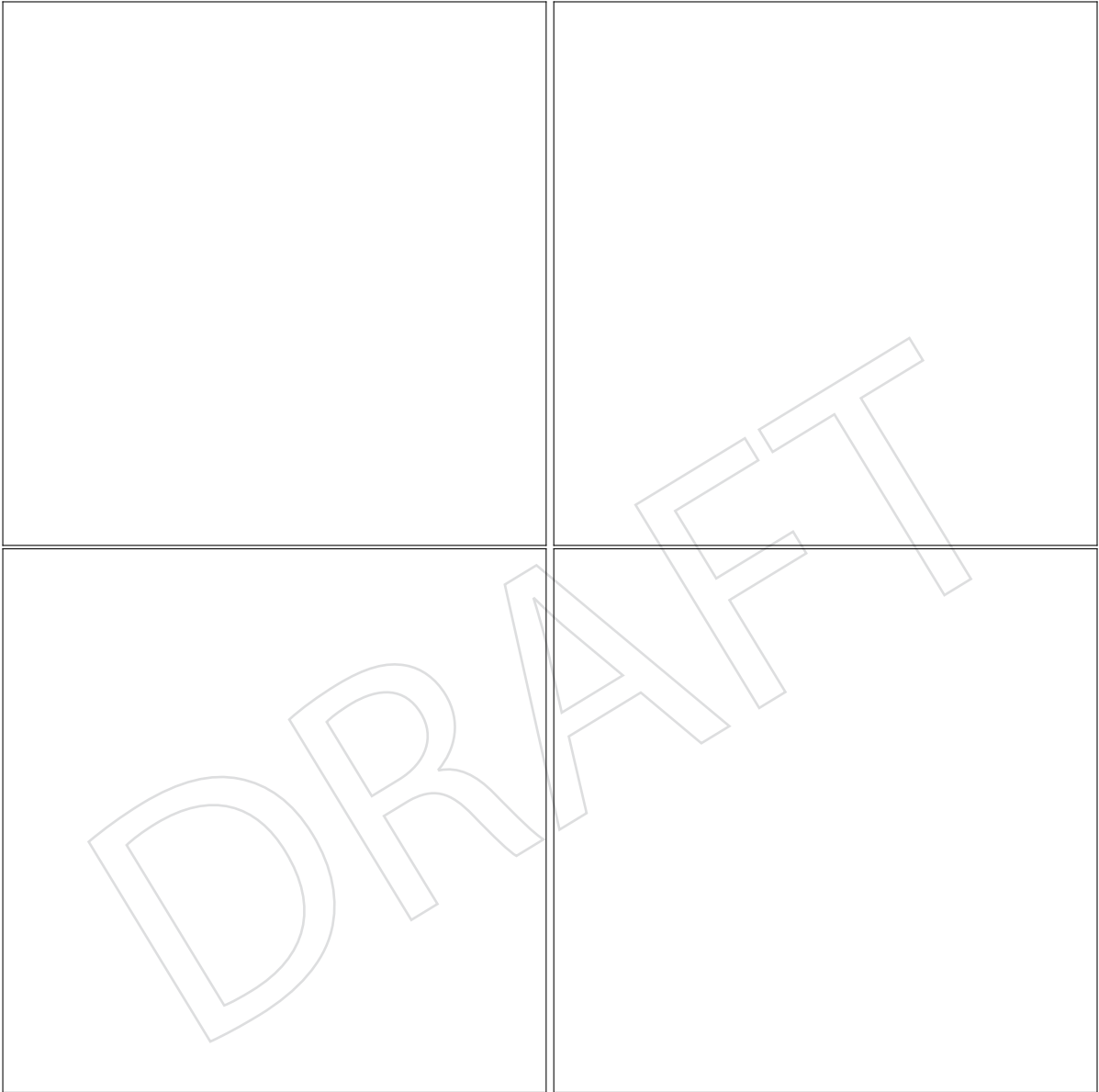


Figure 17: Comparison between data and monte carlo simulation in the monojet signal region for the recoil distribution, the AK4 jet multiplicity distribution,  $p_T$  and  $\eta$  distribution of the leading AK4 jet in 2017 dataset.

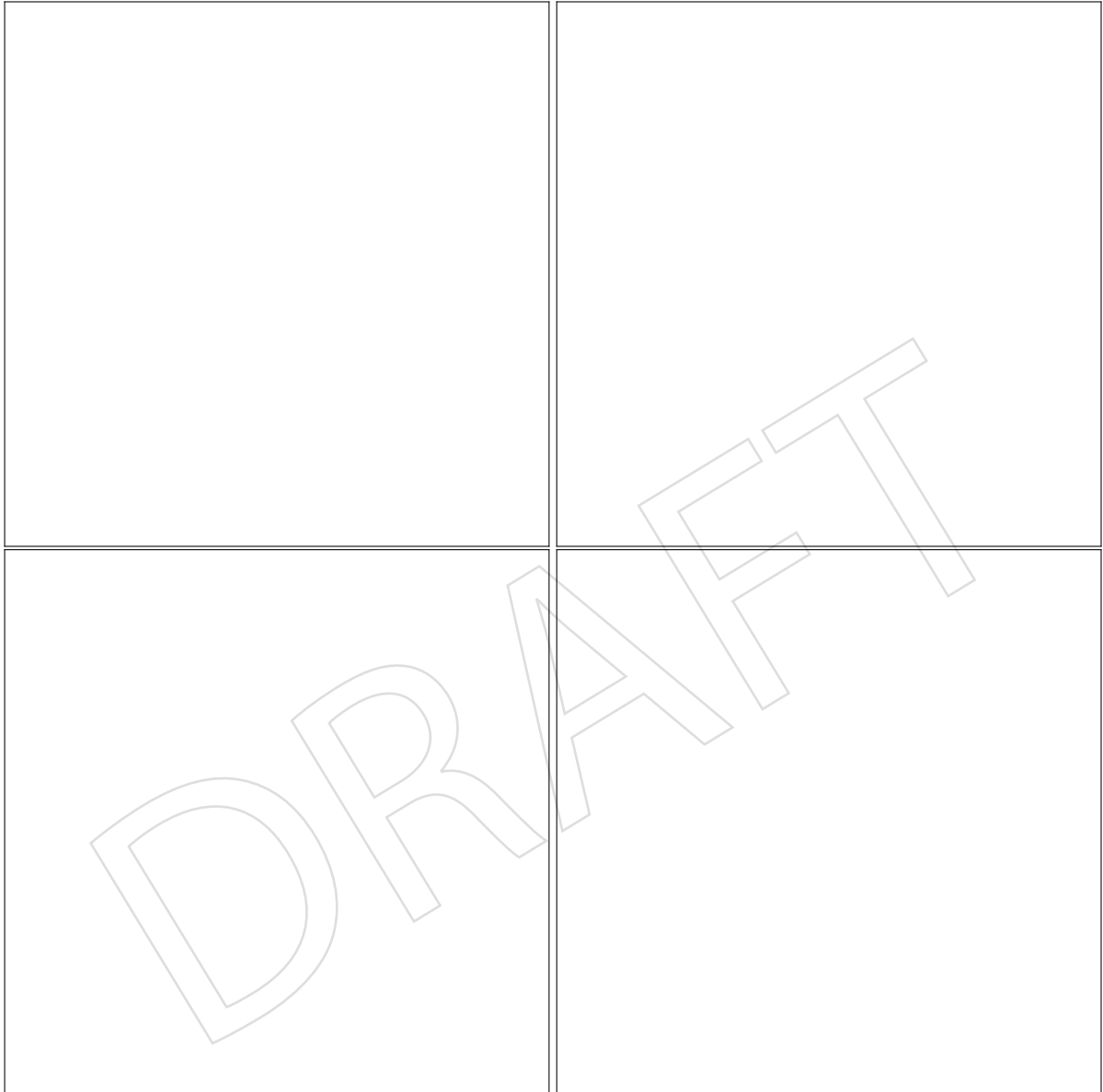


Figure 18: Comparison between data and monte carlo simulation in the monojet signal region for the recoil distribution, the AK4 jet multiplicity distribution,  $p_T$  and  $\eta$  distribution of the leading AK4 jet in 2017 dataset.



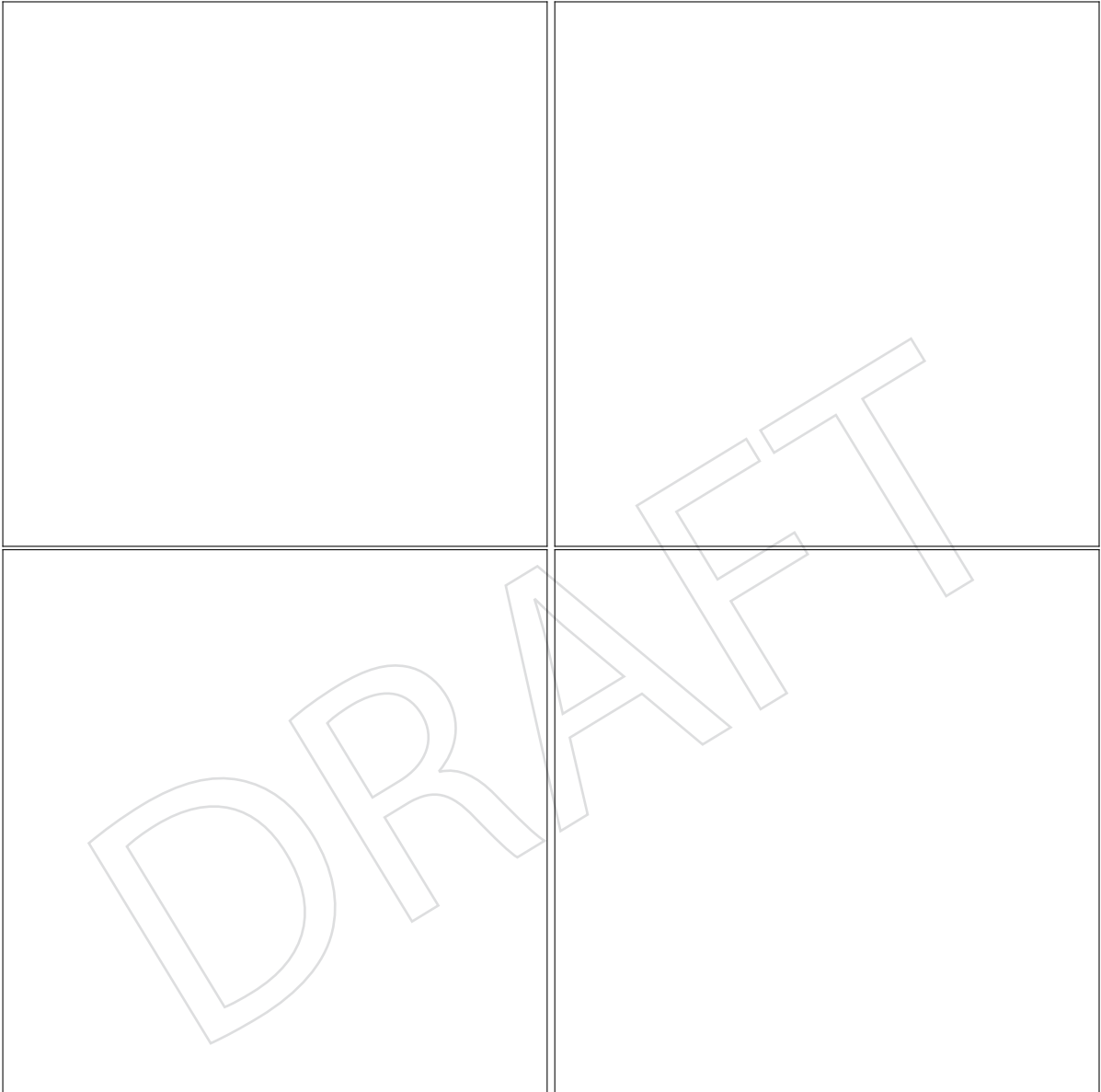


Figure 19: Comparison between data and monte carlo simulation in the monojet signal region for the recoil distribution, the AK4 jet multiplicity distribution,  $p_T$  and  $\eta$  distribution of the leading AK4 jet in 2017 dataset.

## 5.2 Single muon control region selection

Single-muon control sample events are selected using full signal region criteria of VBF selection with the exception of the muon veto. The  $p_T^{\text{miss}}$  requirement is replaced by an identical requirement on the hadronic recoil, which is defined as the sum of  $\vec{p}_T^{\text{miss}}$  and the muon  $\vec{p}_T$ , and thus corresponds to the distribution of the  $W$   $p_T$ . In the single-muon control sample, exactly one tightly identified, isolated muon with  $p_T > 20$  GeV is required. No additional loose muons or electrons with  $p_T > 10$  GeV are allowed. In addition, the transverse mass of the muon- $\vec{p}_T^{\text{miss}}$  system is required to be smaller than 160 GeV. The transverse mass ( $M_T$ ) is computed as  $M_T = \sqrt{2E_T^{\text{miss}}p_T^\mu(1 - \cos\Delta\phi)}$ , where  $p_T^\mu$  is the  $p_T$  of the muon, and  $\Delta\phi$  is the angle between  $\vec{p}_T^\mu$  and  $\vec{p}_T^{\text{miss}}$ .

Figs. 20 and 22 show the distributions of the recoil,  $M_{jj}$ ,  $\Delta\eta_{jj}$  and  $\Delta\phi_{jj}$  of the two leading AK4 jets for events in the single-muon control sample for the VBF category in 2017 and 2018 datasets, respectively. Figs. 21 and 23 show the distributions of the leading muon  $p_T$  and  $\eta$ , as well as the muon- $p_T^{\text{miss}}$  transverse mass, again for 2017 and 2018, respectively.

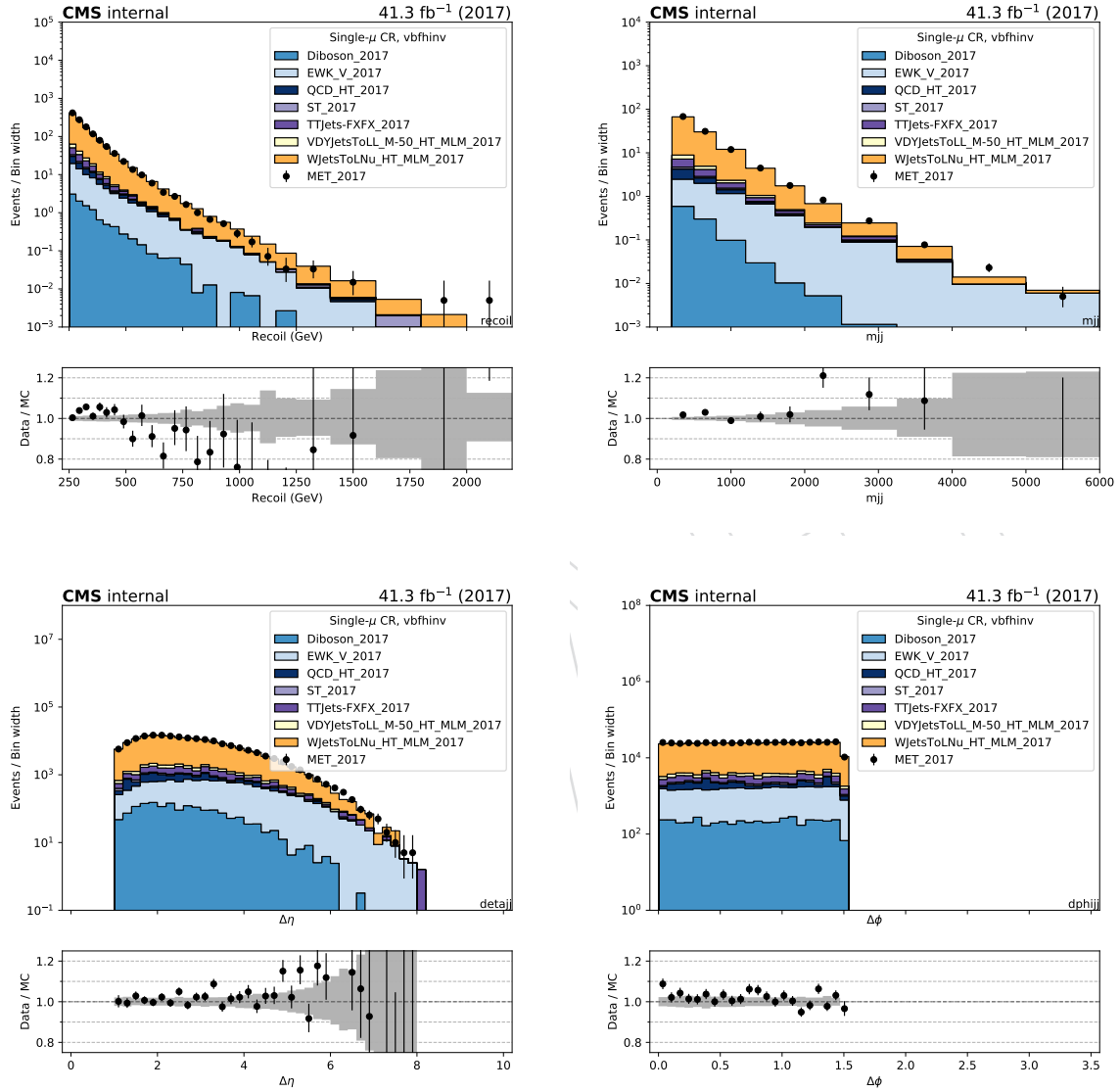


Figure 20: Comparison between 2017 data and Monte Carlo simulation in the single muon control sample for the recoil distribution, the  $M_{jj}$  distribution,  $\Delta\eta_{jj}$  distribution and  $\Delta\phi_{jj}$  distribution for the two leading AK4 jets with the VBF selection.

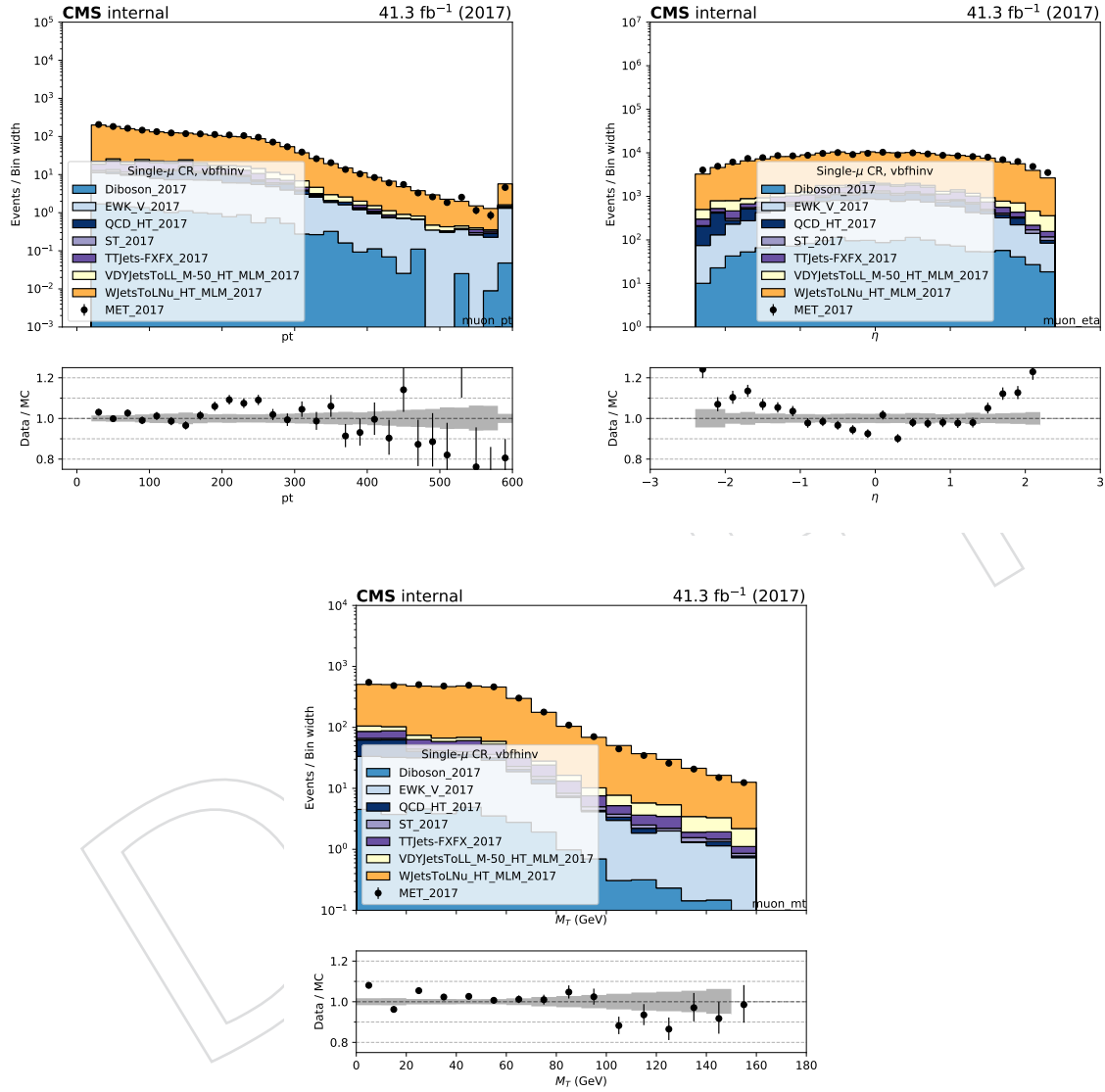


Figure 21: Comparison between 2017 data and Monte Carlo simulation in the single muon control sample for the  $p_T$  and  $\eta$  of the leading muon and the transverse mass distribution with the VBF selection.

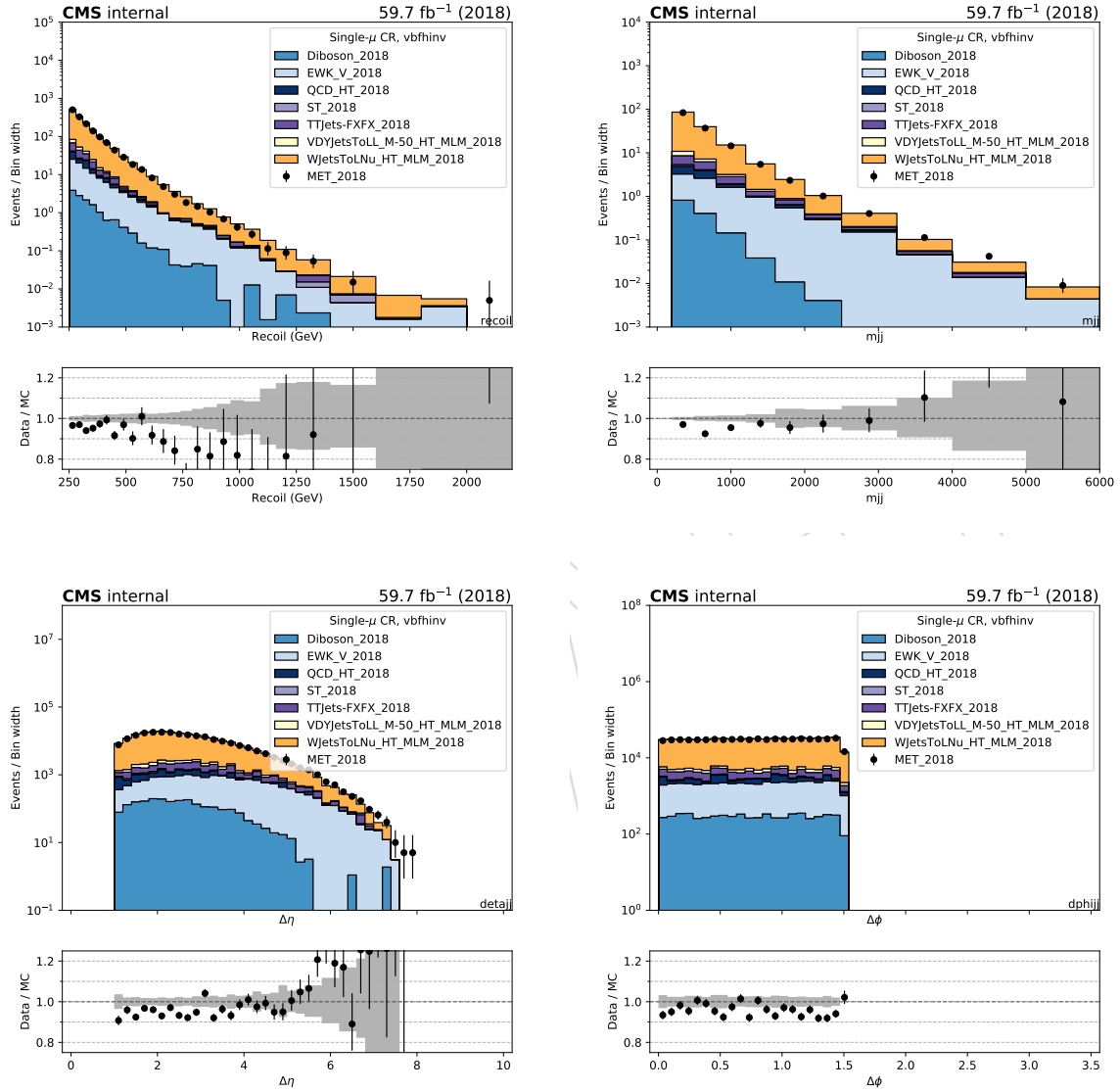


Figure 22: Comparison between 2018 data and Monte Carlo simulation in the single muon control sample for the recoil distribution, the  $M_{jj}$  distribution,  $\Delta\eta_{jj}$  distribution and  $\Delta\phi_{jj}$  distribution for the two leading AK4 jets with the VBF selection.

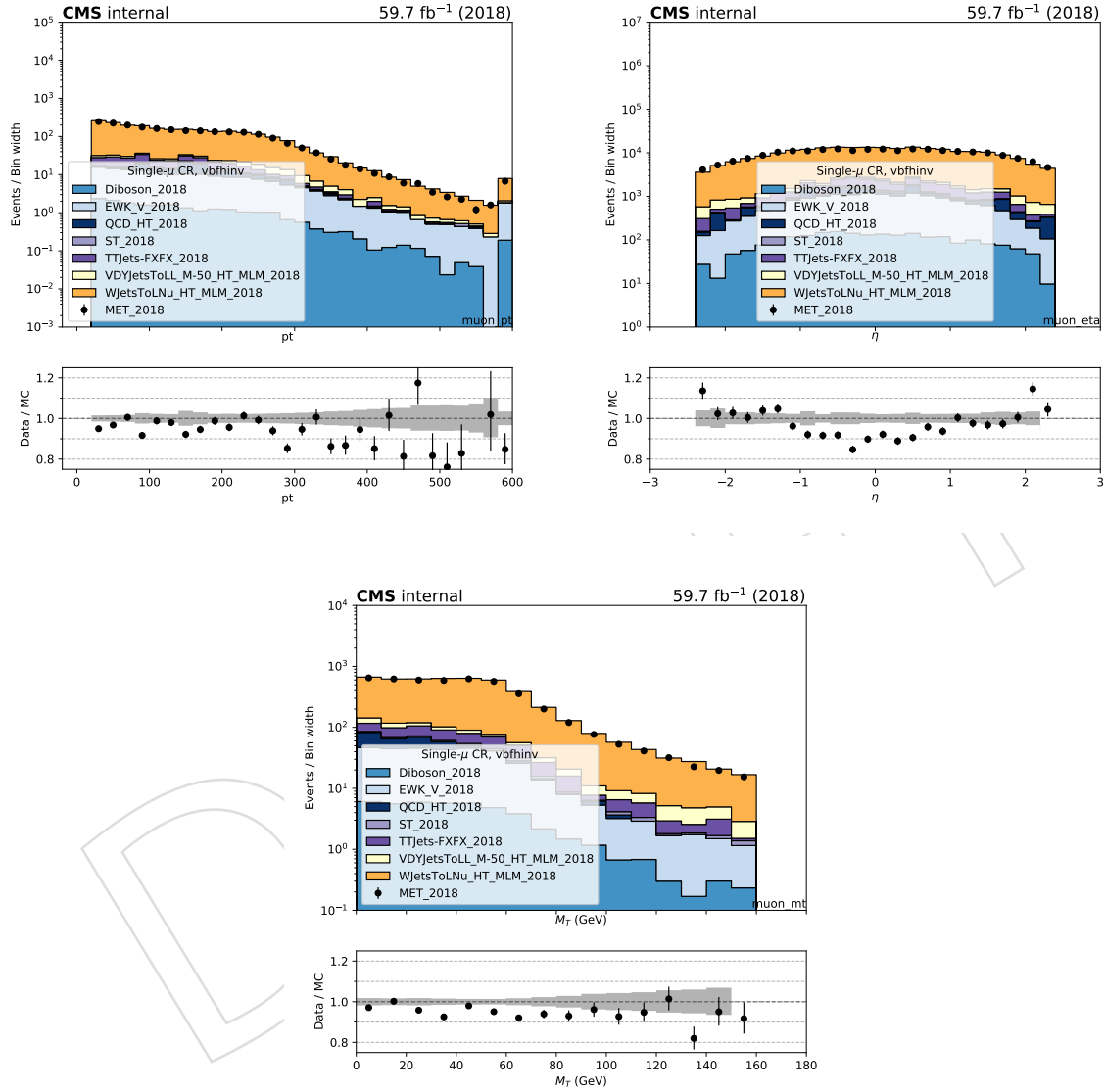


Figure 23: Comparison between 2018 data and Monte Carlo simulation in the single muon control sample for the  $p_T$  and  $\eta$  of the leading muon and the transverse mass distribution with the VBF selection.

### 5.3 Single electron control region selection

Events for the single-electron control sample are collected with the single-electron and photon triggers described in Sec. 2. The  $p_T^{\text{miss}}$  requirement is replaced with an identical requirement on the hadronic recoil, which is defined as the sum of  $\vec{p}_T^{\text{miss}}$  and the electron  $\vec{p}_T$ , and thus corresponds to the distribution of the W  $p_T$ . The events in the single-electron control sample are required to contain exactly one tightly identified and isolated electron with  $p_T > 40$  GeV. In addition, the contamination from QCD multijet events in this control sample is suppressed by requiring  $E_T^{\text{miss}} > 50$  GeV and  $M_T < 160$  GeV.

Figs. 24 and 26 show the distributions of the recoil,  $M_{jj}$ ,  $\Delta\eta_{jj}$  and  $\Delta\phi_{jj}$  of the two leading AK4 jets for events in the single-electron control sample for the VBF category in 2017 and 2018 datasets, respectively. Figs. 25 and 27 show the distributions of the leading electron  $p_T$  and  $\eta$ , as well as the electron- $p_T^{\text{miss}}$  transverse mass, again for 2017 and 2018, respectively.



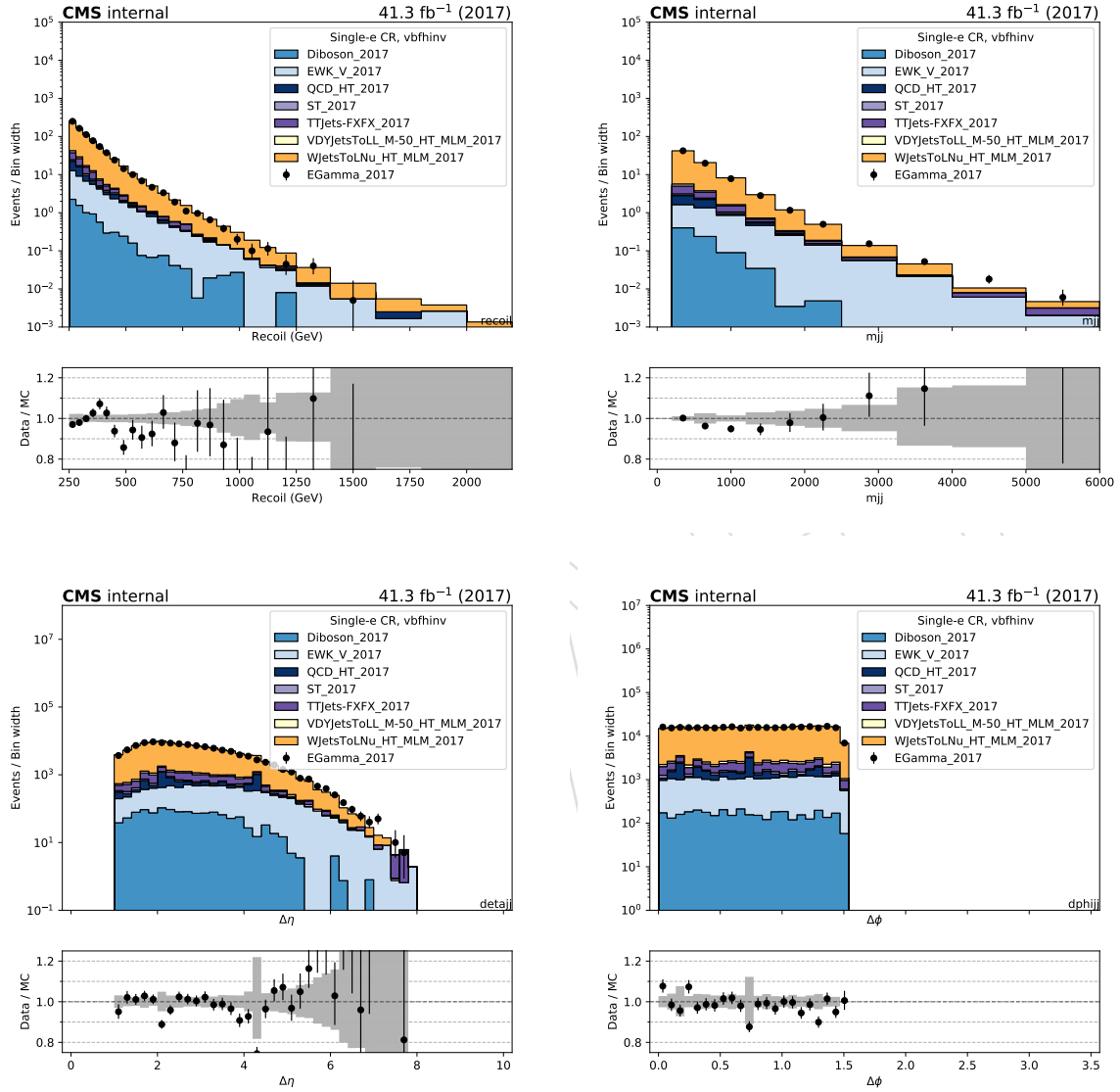


Figure 24: Comparison between 2017 data and Monte Carlo simulation in the single electron control sample for the recoil distribution, the  $M_{jj}$  distribution,  $\Delta\eta_{jj}$  distribution and  $\Delta\phi_{jj}$  distribution for the two leading AK4 jets with the VBF selection.

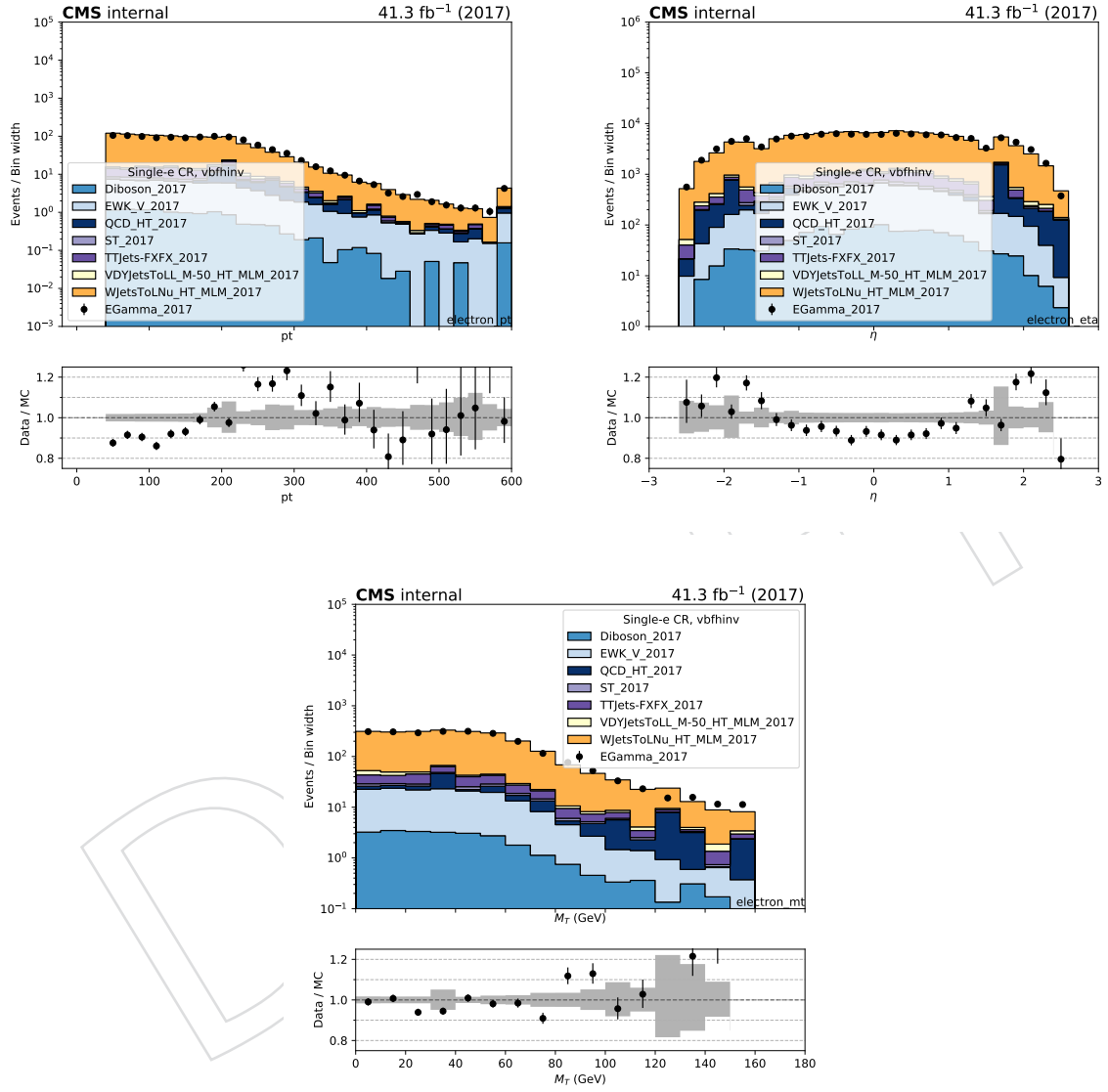


Figure 25: Comparison between 2017 data and Monte Carlo simulation in the single electron control sample for the  $p_T$  and  $\eta$  of the leading electron and the transverse mass distribution with the VBF selection.

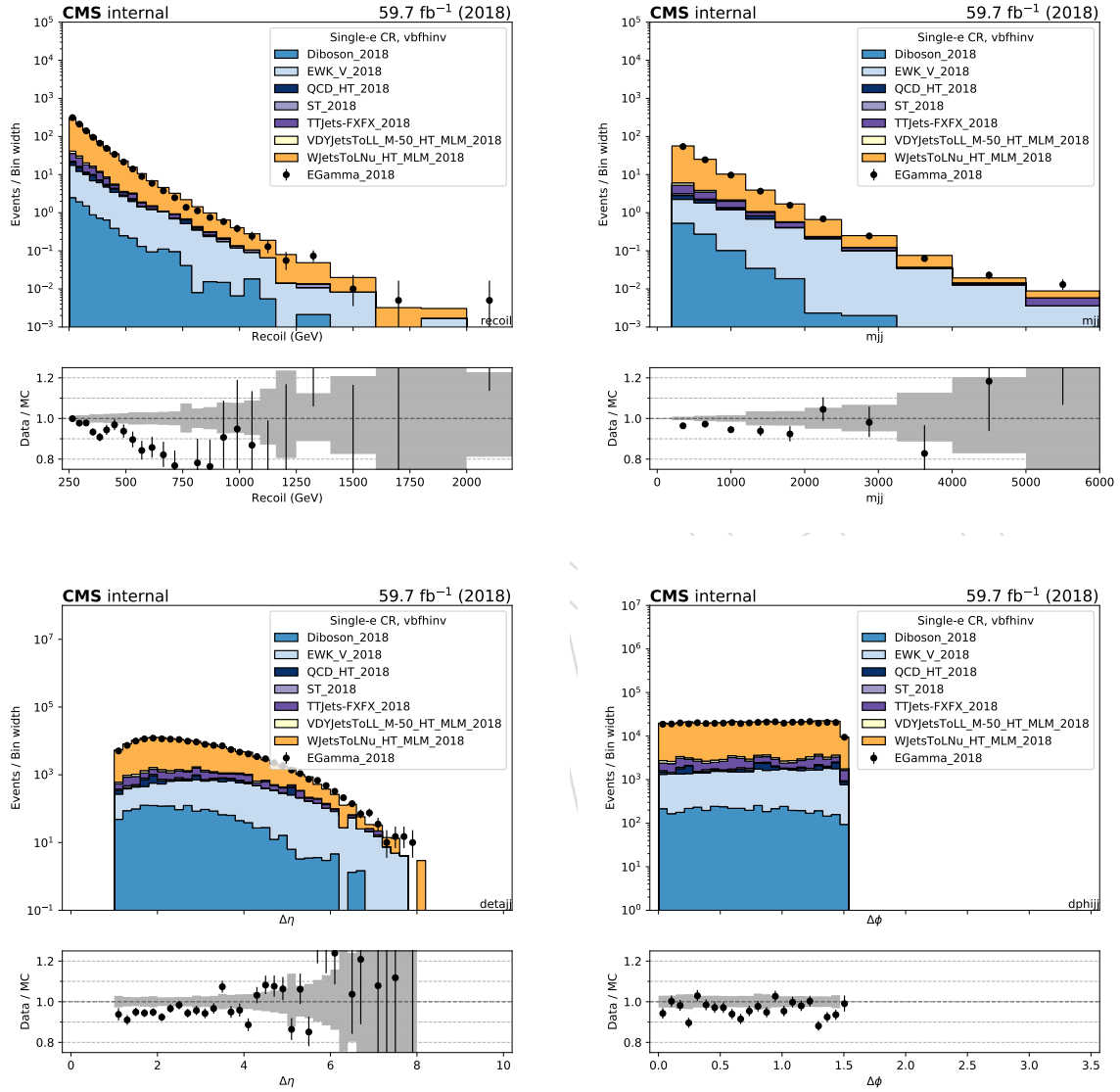


Figure 26: Comparison between 2018 data and Monte Carlo simulation in the single electron control sample for the recoil distribution, the  $M_{jj}$  distribution,  $\Delta\eta_{jj}$  distribution and  $\Delta\phi_{jj}$  distribution for the two leading AK4 jets with the VBF selection.

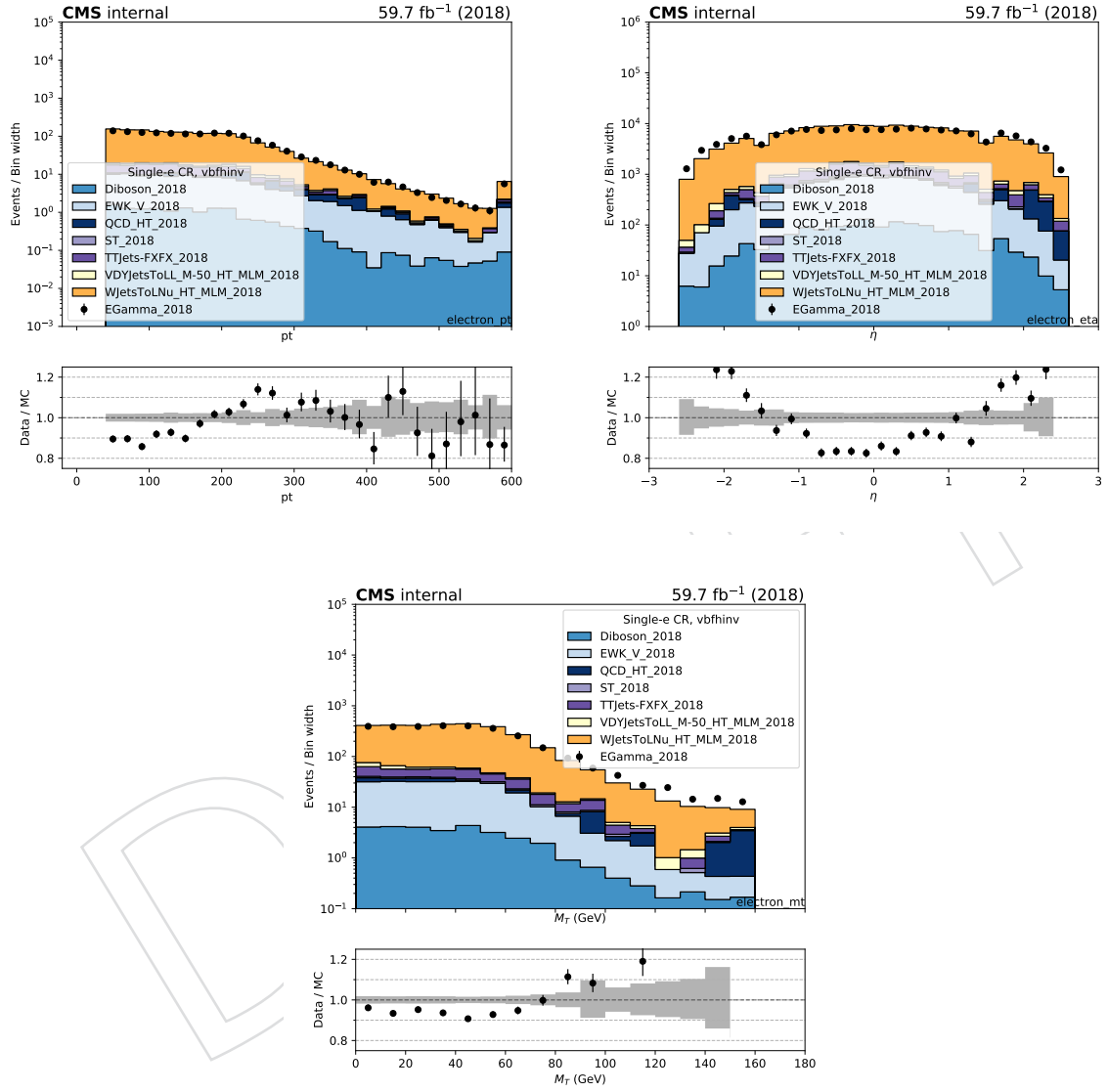


Figure 27: Comparison between 2018 data and Monte Carlo simulation in the single electron control sample for the  $p_T$  and  $\eta$  of the leading electron and the dilepton mass distribution with the VBF selection.

## 5.4 Double muon control region selection

Double-muon control sample events are selected using full signal region criteria of VBF category with the exception of the muon veto. In the double-muon control sample, events are selected requiring leading (subleading) muon  $p_T$  greater than 20 (10) GeV and an invariant mass in the range 60 to 120 GeV, compatible with a Z boson decay. At least one of the two muons is required to pass the tight candidate definition. Events are rejected if there is an additional loose muon or electron with  $p_T > 10$  GeV. The SR  $p_T^{\text{miss}}$  requirement is replacement an identical requirement on the hadronic recoil, which is defined as the sum of  $\vec{p}_T^{\text{miss}}$  and the muon  $\vec{p}_T$ , and thus corresponds to the distribution of the Z  $p_T$  smeared with the  $p_T^{\text{miss}}$  resolution.

Figs. 28 and 30 shows the distributions of the recoil,  $M_{jj}$ ,  $\Delta\eta_{jj}$  and  $\Delta\phi_{jj}$  of the two leading AK4 jets for events in the double-muon control sample for the VBF category in 2017 and 2018 datasets, respectively. Figs. 29 and 31 show the distributions of the leading muon  $p_T$  and  $\eta$ , as well as the dimuon mass and  $p_T$ , again for 2017 and 2018, respectively.

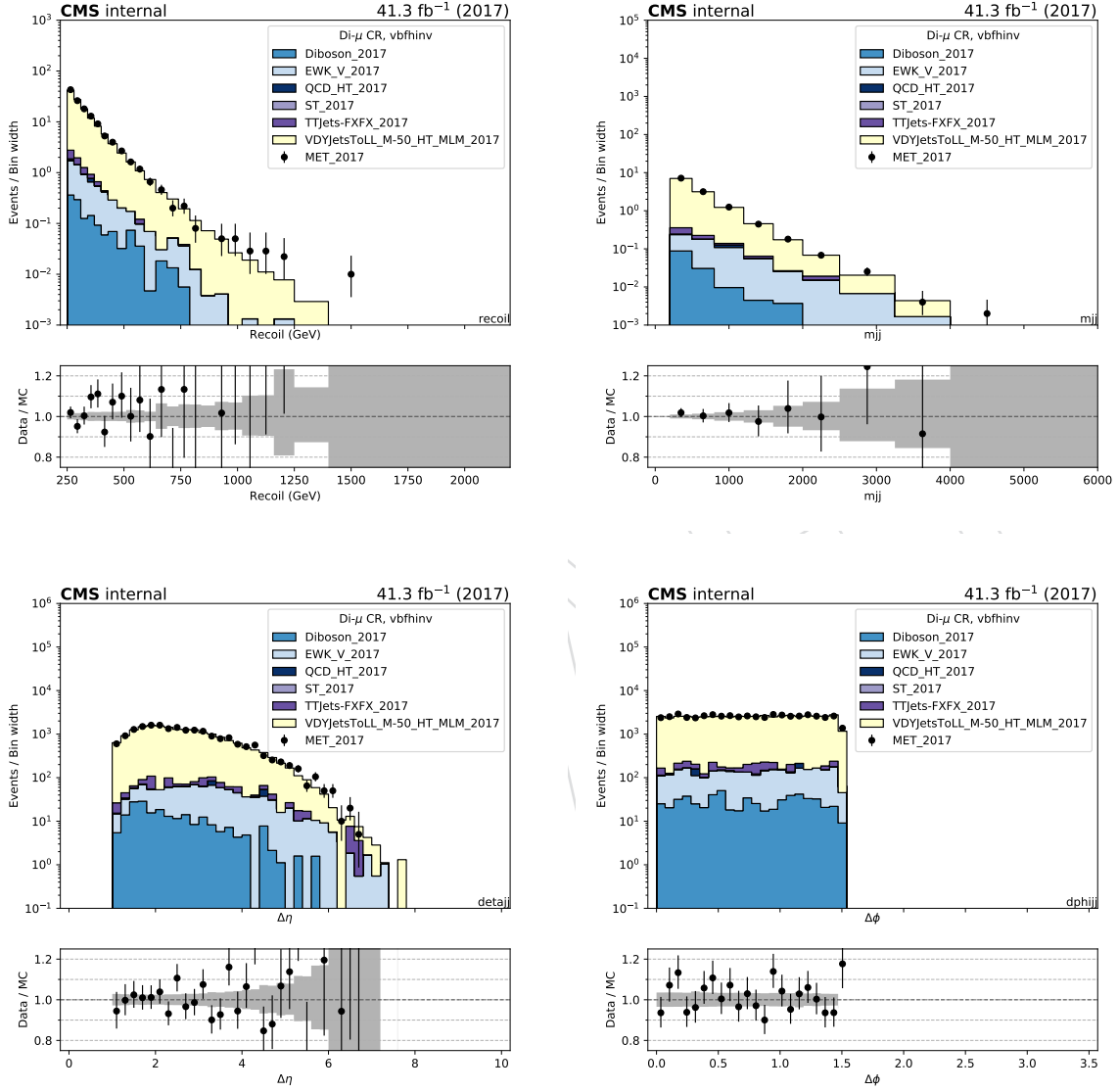


Figure 28: Comparison between 2017 data and Monte Carlo simulation in the double muon control sample for the recoil distribution, the  $M_{jj}$  distribution,  $\Delta\eta_{jj}$  distribution and  $\Delta\phi_{jj}$  distribution for the two leading AK4 jets with the VBF selection.

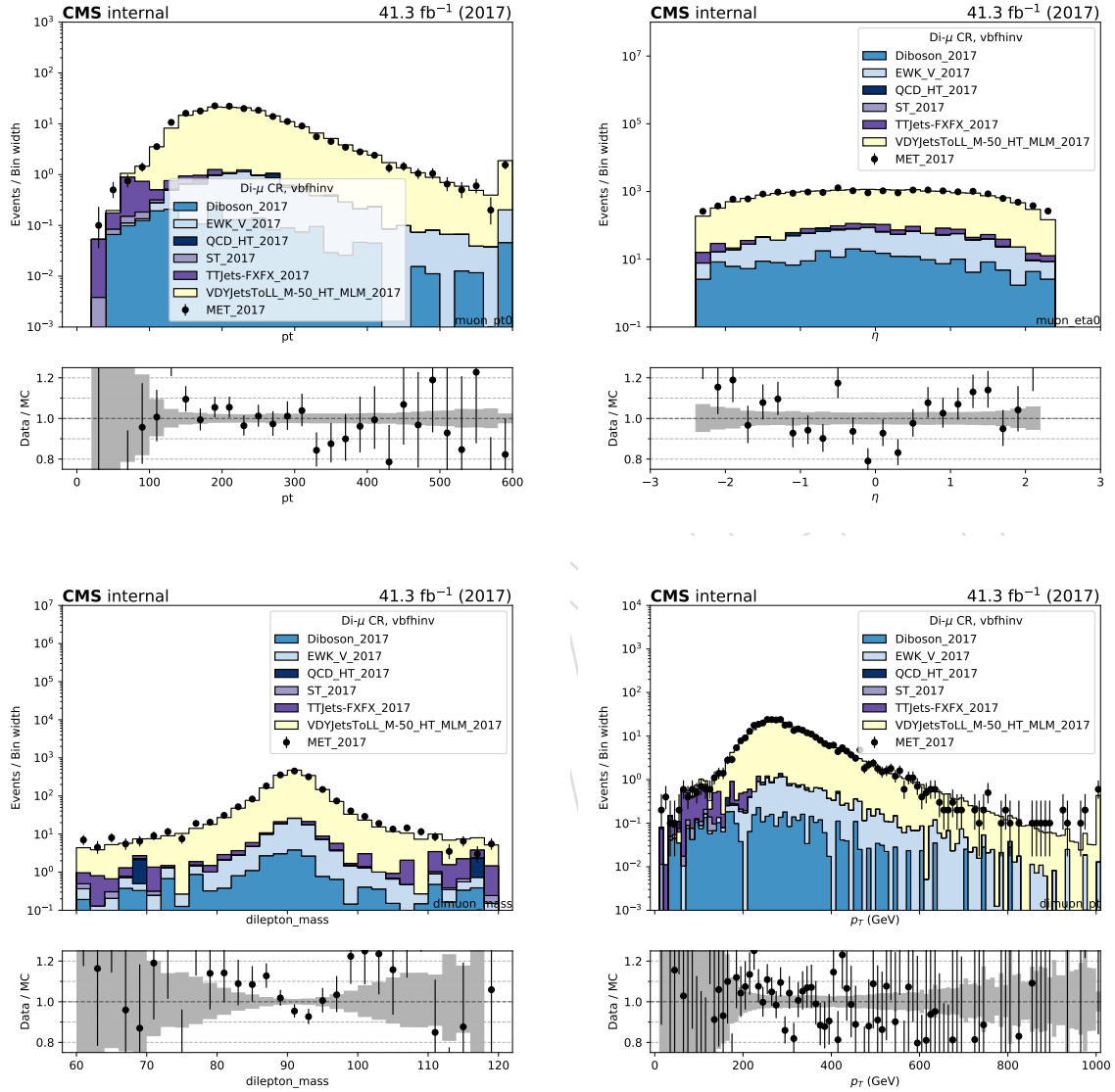


Figure 29: Comparison between 2017 data and Monte Carlo simulation in the double muon control sample for the  $p_T$  and  $\eta$  of the leading muon and the transverse mass and  $p_T$  of the dimuon candidate with the VBF selection.

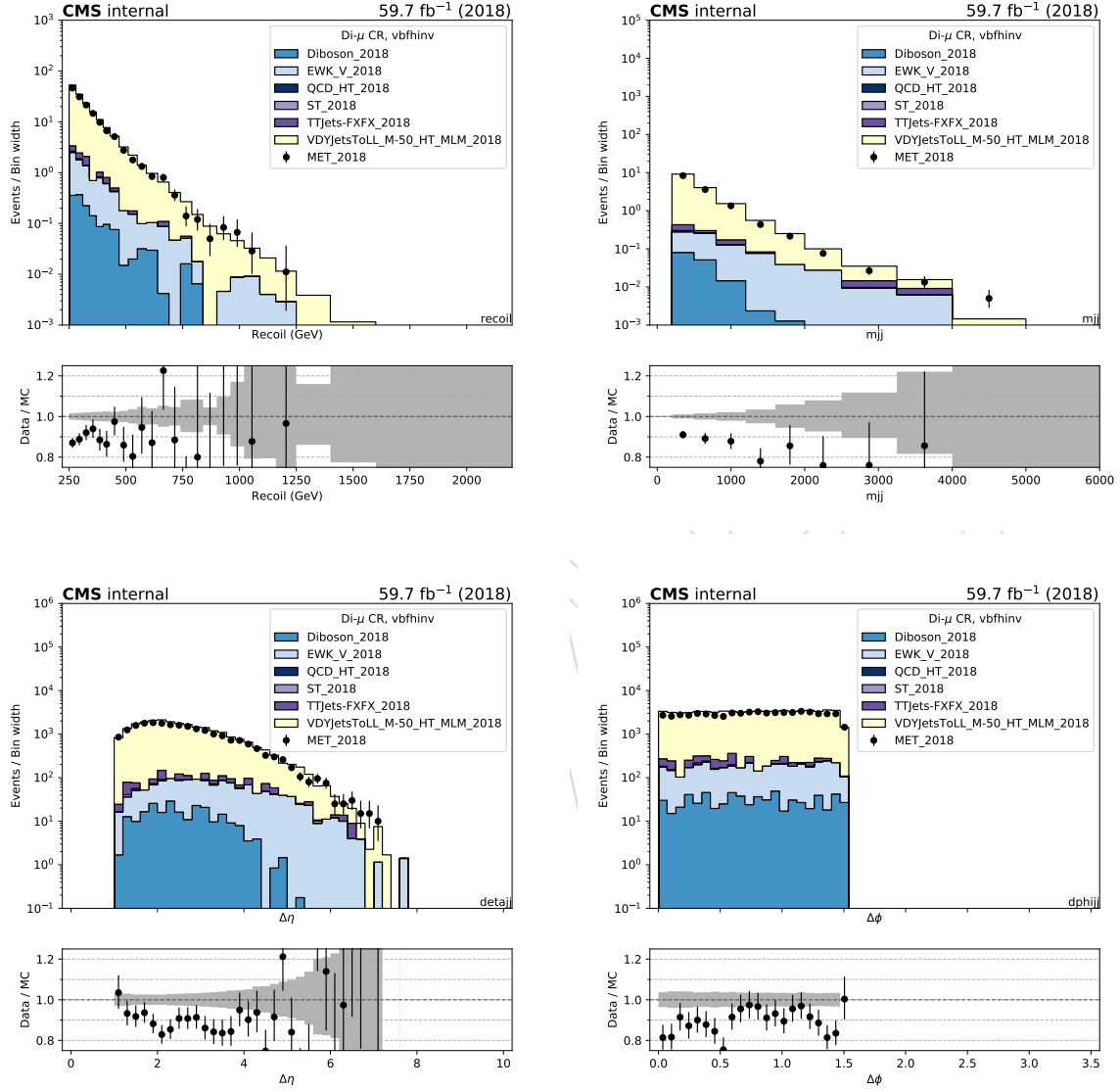


Figure 30: Comparison between 2018 data and Monte Carlo simulation in the double muon control sample for the recoil distribution, the  $M_{jj}$  distribution,  $\Delta\eta_{jj}$  distribution and  $\Delta\phi_{jj}$  distribution for the two leading AK4 jets with the VBF selection.



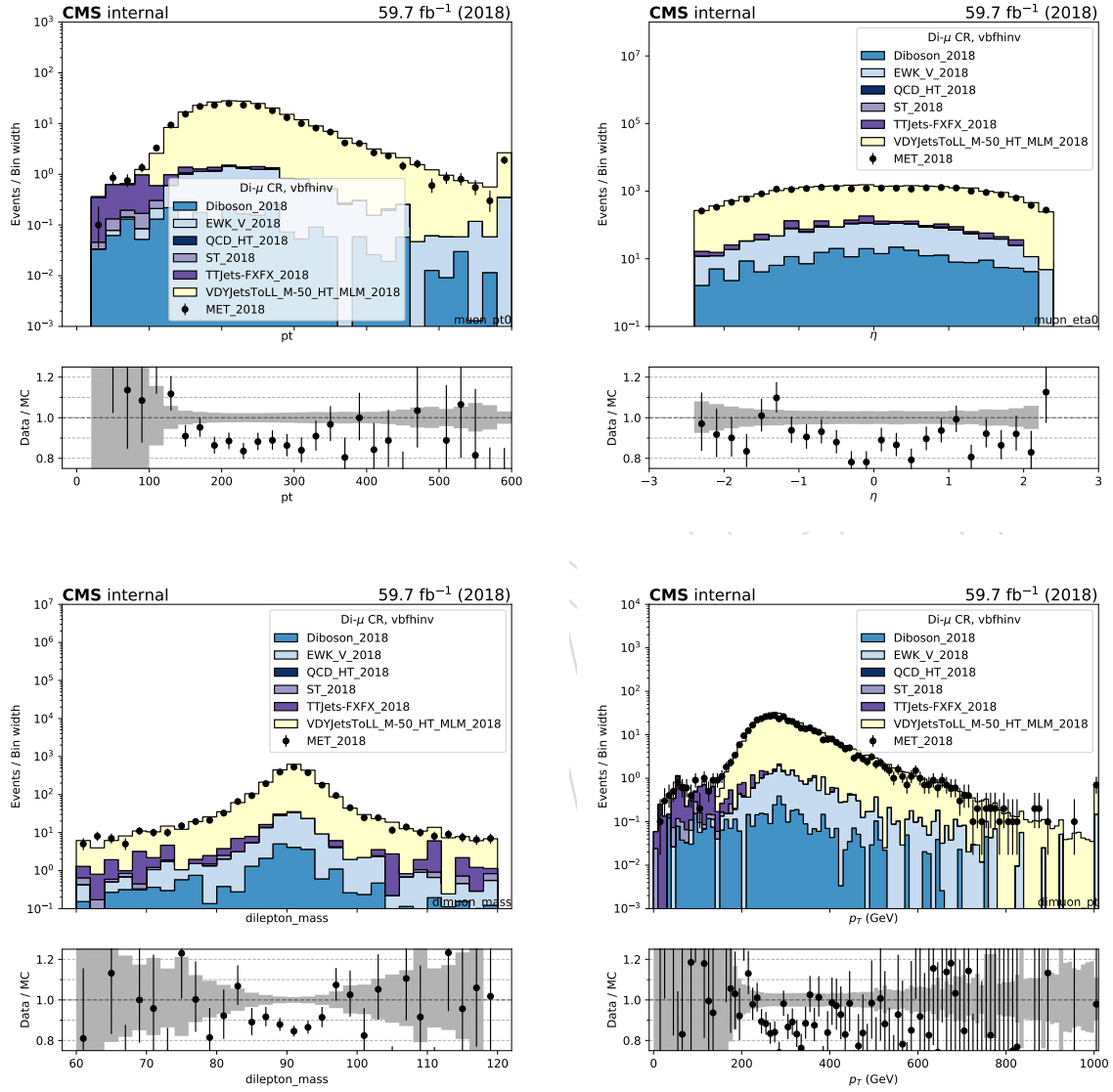


Figure 31: Comparison between 2018 data and monte carlo simulation in the double muon control sample for the  $p_T$  and  $\eta$  of the leading muon and the transverse mass and  $p_T$  of the dimuon candidate with the VBF selection.

## 5.5 Double electron control region selection

Events for the double-electron control sample are collected with the single-electron and photon triggers described in Sec. 2. In the offline analysis, events in the dielectron control sample are required to contain exactly two oppositely charged electrons with leading (trailing) electron  $p_T$  greater than 40 (10) GeV, with at least one of the two passing the tight candidate definition. The SR  $p_T^{\text{miss}}$  requirement is replacement an identical requirement on the hadronic recoil, which is defined as the sum of  $\vec{p}_T^{\text{miss}}$  and the muon  $\vec{p}_T$ , and thus corresponds to the distribution of the Z  $p_T$  smeared with the  $p_T^{\text{miss}}$  resolution. Similar to the dimuon control sample case, the invariant mass of the dielectron system is required to be between 60 and 120 GeV to be consistent with a Z boson decay.

Figs. 32 and 34 shows the distributions of the recoil,  $M_{jj}$ ,  $\Delta\eta_{jj}$  and  $\Delta\phi_{jj}$  for the two leading AK4 jets for events in the double-electron control sample for the VBF category in 2017 and 2018 datasets, respectively. Figs. ?? and ?? show the distributions of the leading electron  $p_T$  and  $\eta$ , as well as the dielectron mass and  $p_T$ , again for 2017 and 2018, respectively.

DRAFT

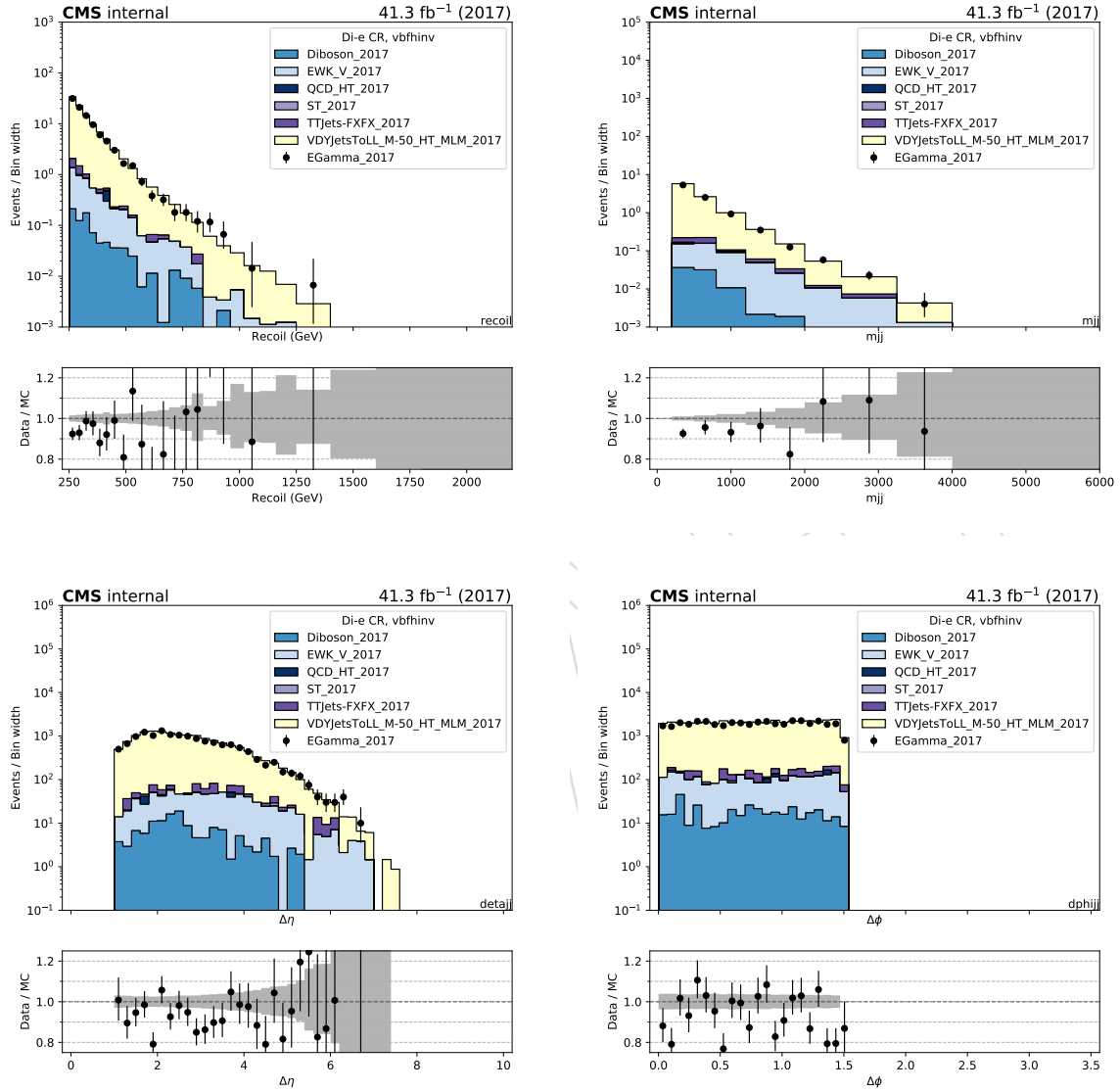


Figure 32: Comparison between 2017 data and Monte Carlo simulation in the double electron control sample for the recoil distribution, the  $m_{jj}$  distribution,  $\Delta\eta_{jj}$  distribution and  $\Delta\phi_{jj}$  distribution for the two leading AK4 jets with the VBF selection.

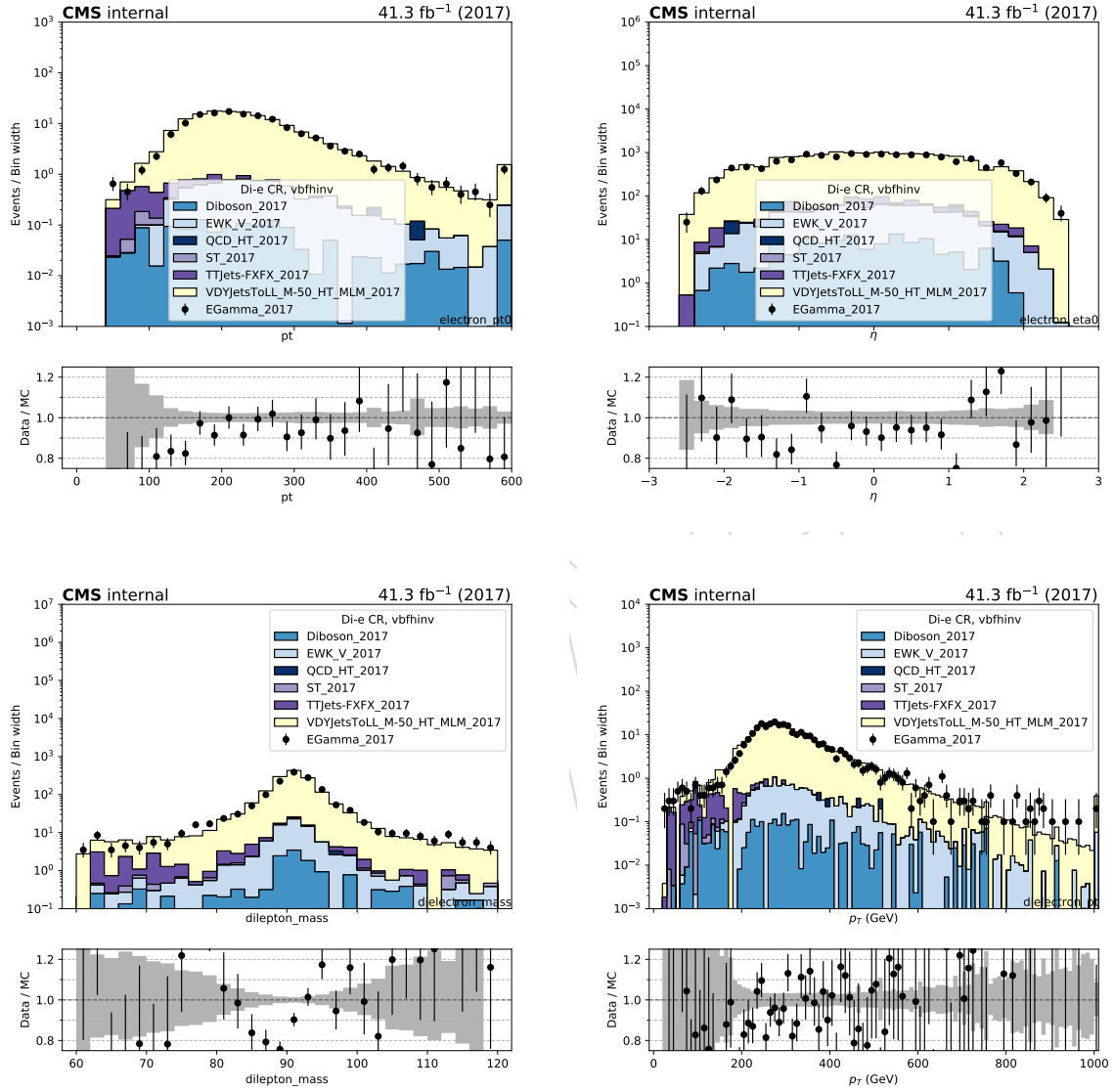


Figure 33: Comparison between 2017 data and Monte Carlo simulation in the double electron control sample for the  $p_T$  and  $\eta$  of the leading electron and the transverse mass and  $p_T$  of the dilepton candidate with the VBF selection.

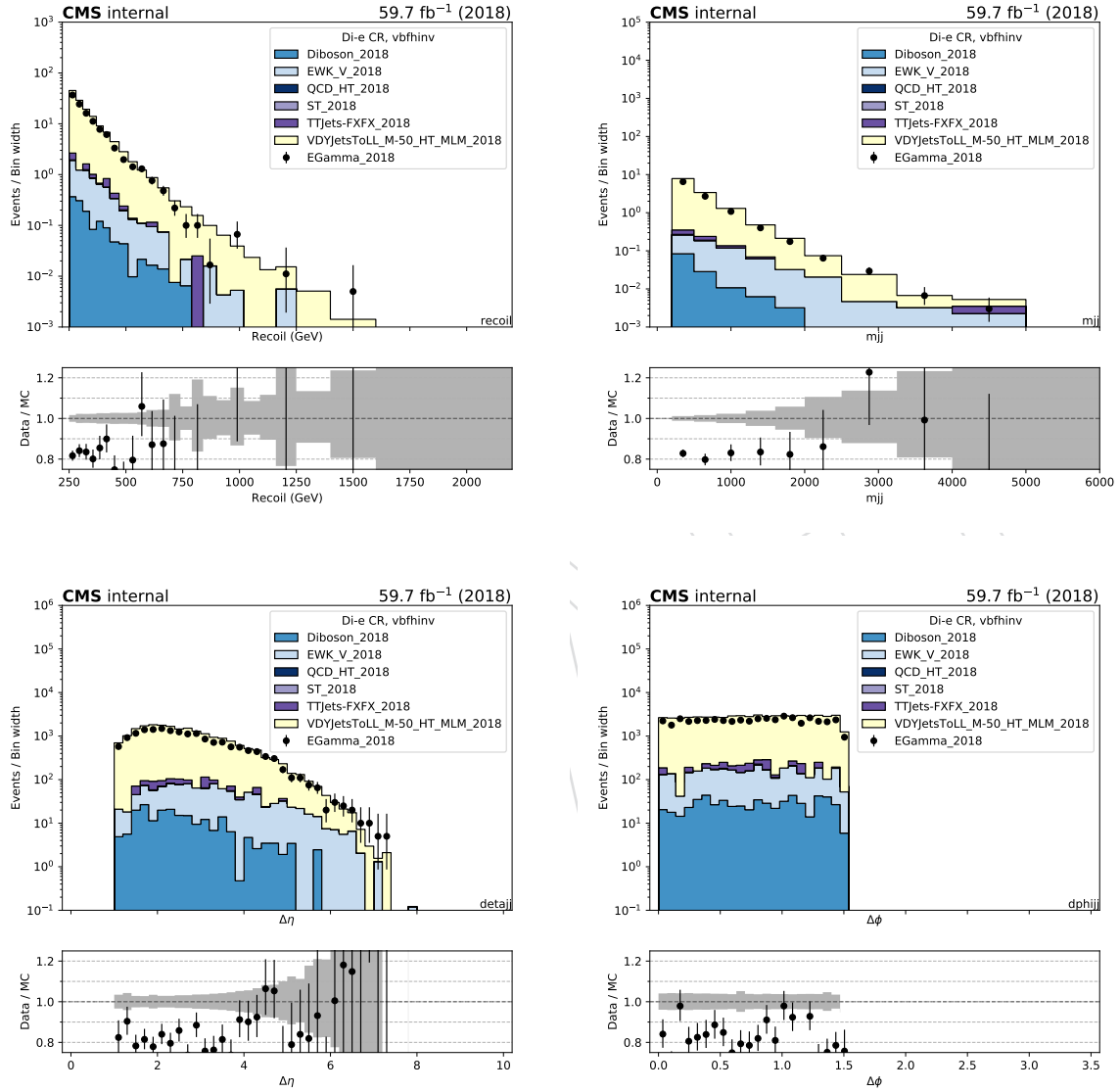


Figure 34: Comparison between 2018 data and Monte Carlo simulation in the double electron control sample for the recoil distribution, the  $M_{jj}$  distribution,  $\Delta\eta_{jj}$  distribution and  $\Delta\phi_{jj}$  distribution for the two leading AK4 jets with the VBF selection.

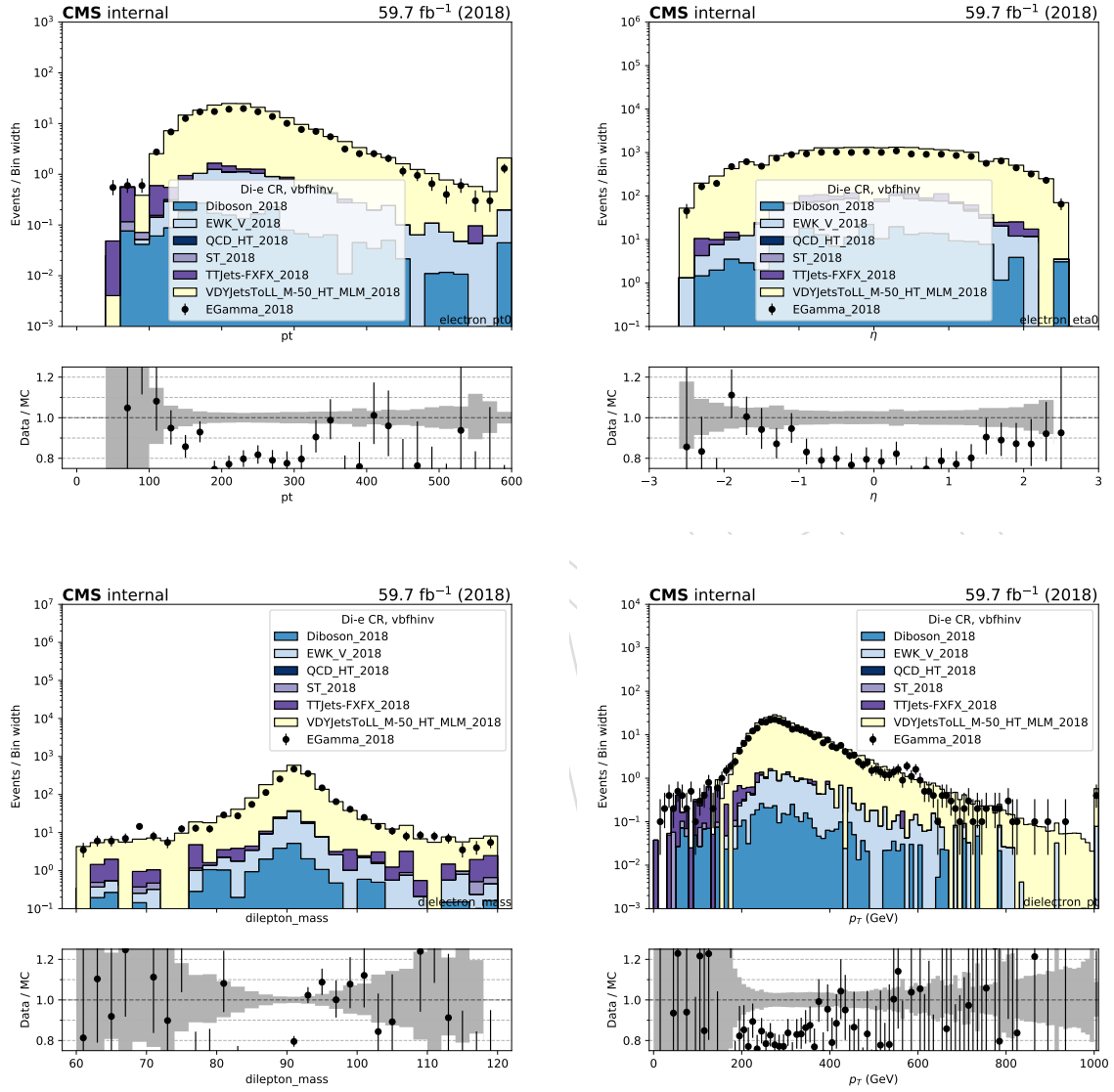


Figure 35: Comparison between 2018 data and Monte Carlo simulation in the double electron control sample for the  $p_T$  and  $\eta$  of the leading electron and the transverse mass and  $p_T$  of the dilepton candidate with the VBF selection.

## 5.6 Photon control region

The  $\gamma + \text{jets}$  control sample is selected using events with one high- $p_T$  photon collected using single-photon triggers with  $p_T$  thresholds of 165 or 175 GeV, depending on the data taking conditions. The photon is required to have  $p_T > 175$  GeV and to pass tight identification and isolation criteria, to ensure a high trigger efficiency of 98%.

Figs. 36 and 36 show the distributions of the recoil,  $M_{jj}$ ,  $\Delta\eta_{jj}$  and  $\Delta\phi_{jj}$  distribution of the two leading AK4 jets for events in the photon control sample for the VBF category in the 2017 and 2018 datasets, respectively. Similarly, Figs. 37 and 39 show the distributions of the photon  $p_T$  and  $\eta$ .

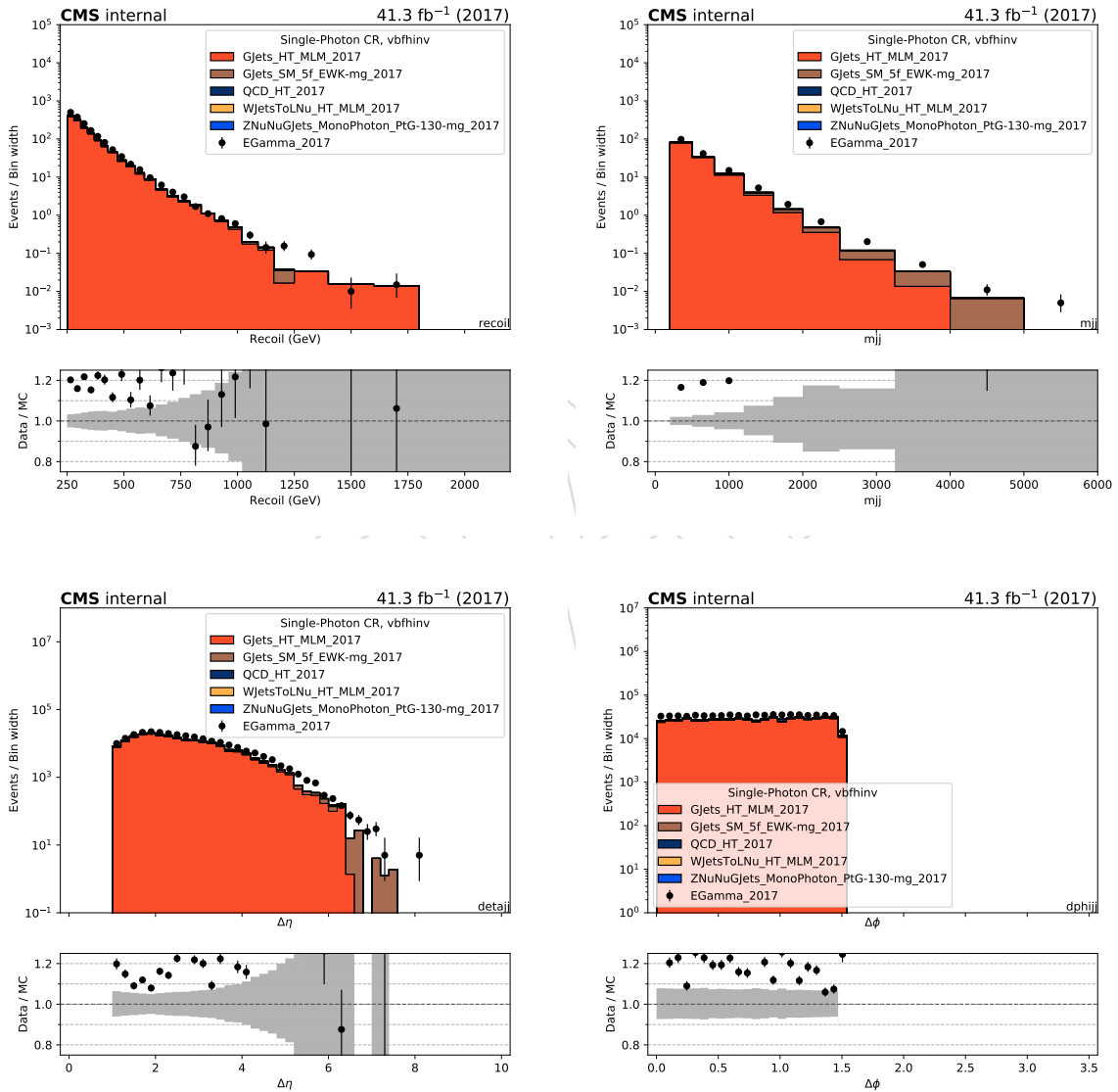


Figure 36: Comparison between 2017 data and Monte Carlo simulation in the photon control sample for the recoil distribution, the  $M_{jj}$  distribution,  $\Delta\eta_{jj}$  distribution and  $\Delta\phi_{jj}$  distribution for the two leading AK4 jets with the VBF selection.

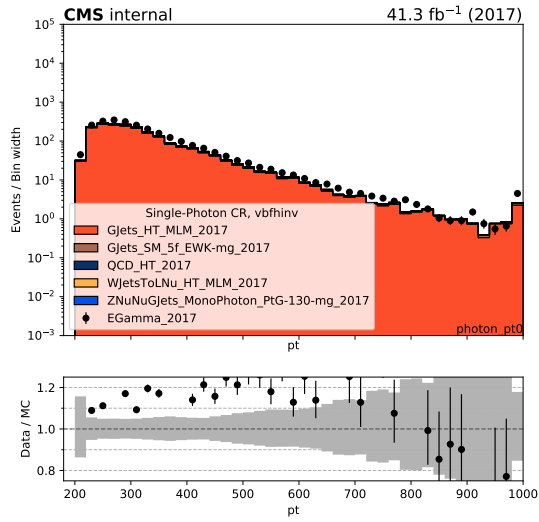


Figure 37: Comparison between 2017 data and Monte Carlo simulation in the photon control sample for the  $p_T$  and  $\eta$  of the leading photon with the VBF selection.



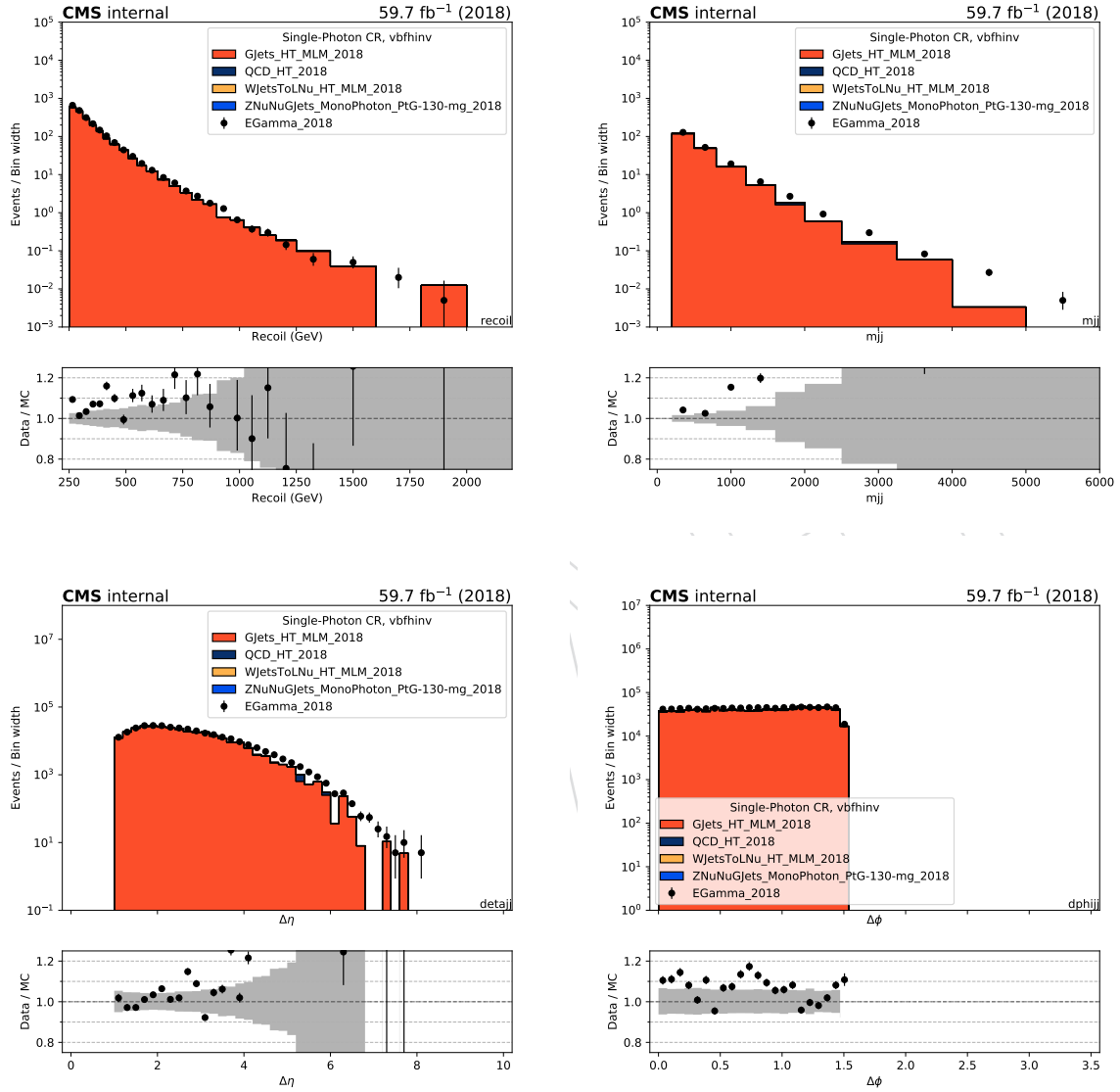


Figure 38: Comparison between 2018 data and Monte Carlo simulation in the photon control sample for the recoil distribution, the  $M_{jj}$  distribution,  $\Delta\eta_{jj}$  distribution and  $\Delta\phi_{jj}$  distribution for the two leading AK4 jets with the VBF selection.

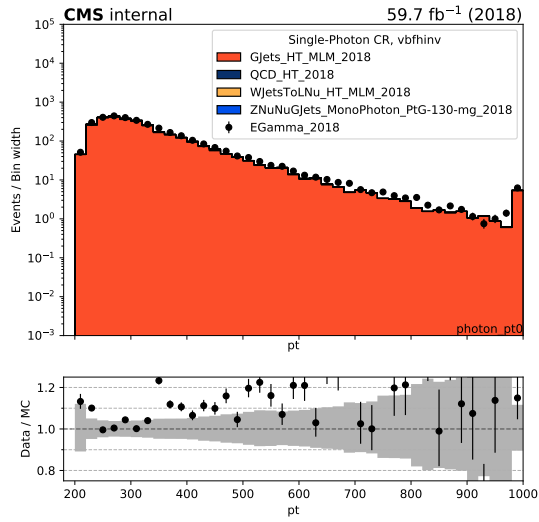


Figure 39: Comparison between 2018 data and Monte Carlo simulation in the photon control sample for the  $p_T$  and  $\eta$  of the leading photon with the VBF selection.

## 6 Signal extraction strategy

The largest background contributions, from  $Z(\nu\nu) + \text{jets}$  and  $W(\ell\nu) + \text{jets}$  processes, are estimated using data from five mutually exclusive control samples selected from dimuon, dielectron, single-muon, single-electron, and  $\gamma + \text{jets}$  final states as explained below. The hadronic recoil  $p_T$  is used as a proxy for  $p_T^{\text{miss}}$  in these control samples, and is defined by excluding identified leptons or photons from the  $p_T^{\text{miss}}$  calculation.

The remaining backgrounds that contribute to the total event yield in the signal region are much smaller than those from  $W(\ell\nu) + \text{jets}$  processes. These smaller backgrounds include QCD multijet events which are measured from data using a  $\Delta\phi$  extrapolation method and will be discussed in 6.1, and top quark and diboson processes, which are obtained directly from simulation.

### 6.1 Multijet background estimation

QCD multijet production yields transversely well-balanced events. However, due to jet energy mismeasurements, uninstrumented or non-functional detector regions, or neutrinos from semileptonic decays of heavy-flavour mesons; the balance might no longer be preserved yielding a final state with  $E_T^{\text{miss}}$ . Although such effects are rare, due to large QCD production cross section such QCD multijet events can show up in final states with large  $E_T^{\text{miss}}$ . In this analysis, event selection is constructed to suppress contributions from such events to below percent level (0.02%) of the total background. On the other hand, this background is known to be badly reproduced in simulation, therefore, we employed a data driven prediction for this process to be used in the final simultaneous fit.

For this measurement the ABCD method is employed using the  $\min \Delta\phi(\text{jet}, E_T^{\text{miss}})$  and  $E_T^{\text{miss}}$  variables.  $\min \Delta\phi(\text{jet}, E_T^{\text{miss}})$  variable gives a strong handle on QCD multijet events since the jet momentum mismeasurements yielding large  $E_T^{\text{miss}}$  will also naturally yield low separation in  $\phi$  between these objects.

The distribution of  $\min \Delta\phi(\text{jet}, E_T^{\text{miss}})$  without imposing any selection on this variable after signal region event selection is shown in Fig. 40.

Large amounts of QCD multijet events are found at low  $\min \Delta\phi(\text{jet}, E_T^{\text{miss}})$  region. As a result, the selection for the A, B, C and D regions are determined to be as follows:

- Region A:  $E_T^{\text{miss}} \leq 250$  and  $\min \Delta\phi(\text{jet}, E_T^{\text{miss}}) \leq 0.5$
- Region B:  $E_T^{\text{miss}} > 250$  and  $\min \Delta\phi(\text{jet}, E_T^{\text{miss}}) \leq 0.5$
- Region C:  $E_T^{\text{miss}} \leq 250$  and  $\min \Delta\phi(\text{jet}, E_T^{\text{miss}}) > 0.5$
- Region D:  $E_T^{\text{miss}} > 250$  and  $\min \Delta\phi(\text{jet}, E_T^{\text{miss}}) > 0.5$

The  $E_T^{\text{miss}}$  distribution split in regions for the QCD multijet simulation is shown in Fig. 41. Each contribution from each of the regions specified above is shown with a different color. The two dimensional distribution of the  $\min \Delta\phi(\text{jet}, E_T^{\text{miss}})$  vs  $E_T^{\text{miss}}$  for QCD multijet simulation is also shown.

The ratio between the regions of low and high  $\min \Delta\phi(\text{jet}, E_T^{\text{miss}})$  then computed as:

$$r = \frac{\min \Delta\phi(\text{jet}, E_T^{\text{miss}}) > 0.5}{\min \Delta\phi(\text{jet}, E_T^{\text{miss}}) < 0.5}$$

This ratio computed using the QCD multijet simulation is shown in Fig. 6.1 and is fitted by a double exponential function. It has been shown in earlier studies [?] that the shape difference

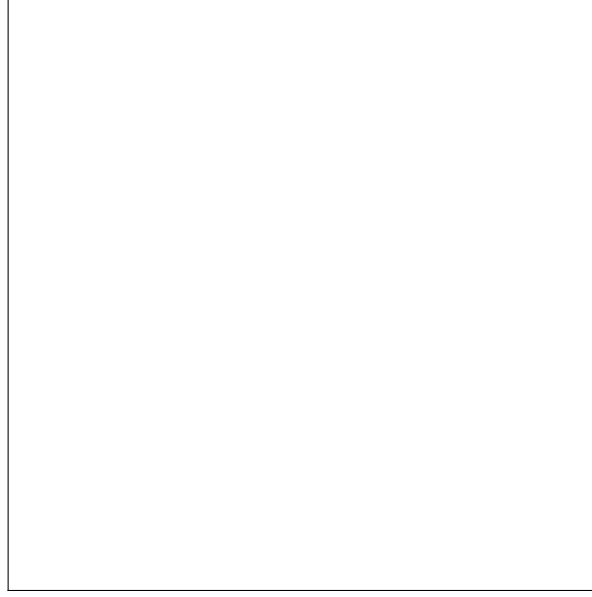


Figure 40: The distribution of  $\min \Delta\phi(\text{jet}, E_T^{\text{miss}})$  without imposing any selection on this variable

in this ratio at different  $E_T^{\text{miss}}$  regions arises from threshold effects, pileup at lower  $E_T^{\text{miss}}$  values, where as at higher values, gaussian core of the jet energy resolution dictates the shape and various over and under measurements of the jet energies could morph the exponential shape. This is the reason why this ratio is fitted by a double exponential.

The data-driven estimation starts by taking the events from region B defined by  $E_T^{\text{miss}} > 250$  and  $\min \Delta\phi(\text{jet}, E_T^{\text{miss}}) \leq 0.5$  selections after imposing the full event selection in MET dataset. All other backgrounds corresponding to V+jets processes are subtracted from these selected events. The remaining events are scaled by the double exponential function fitted to the ratio as shown in Fig. 42. The systematic uncertainty sources are also studied. The largest source of uncertainty in the measurement is due to the variations of jet energy scale, resolution and the fit. To estimate the effect of this variation in the measurment, the jet energy scale and resolution is varied up and down by  $1\sigma$  and the above procedure is repeated. The corresponding variations in the ratio are treated as the systematic uncertainty. The total uncertainty is found to vary from 20% to 150% as a function of  $E_T^{\text{miss}}$  and is shown in Fig. 42 as well. The largest deviation in each bin is taken as the uncertainty.

The final estimation including the uncertainties and the comparison of the final estimation to those estimations coming from the simulated events is shown in Fig. 6.1.

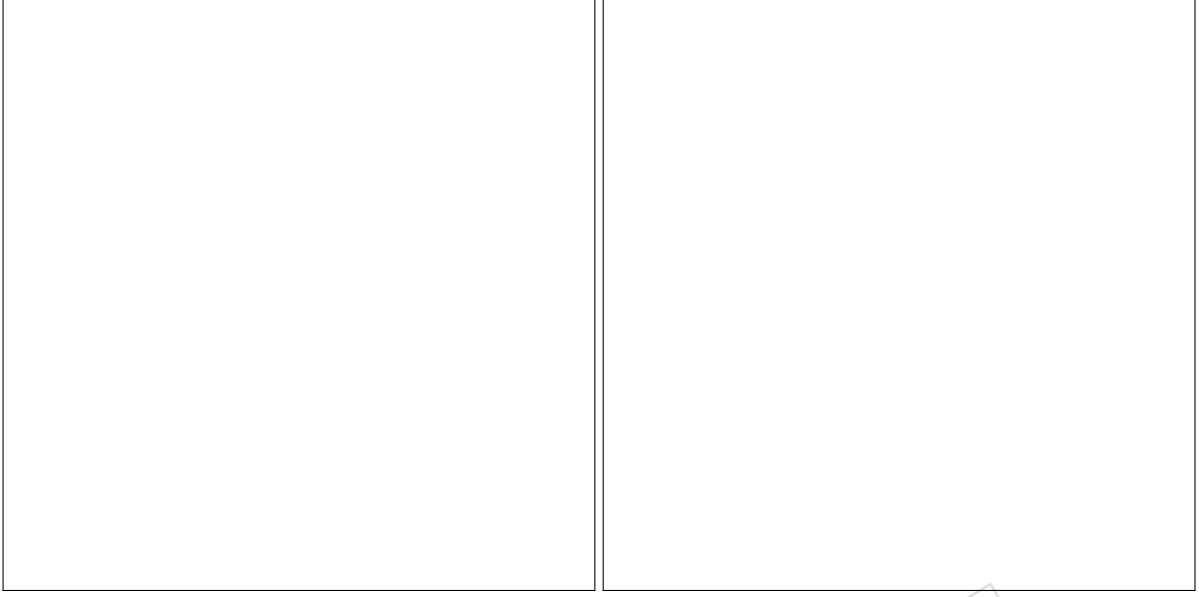


Figure 41: The  $E_T^{\text{miss}}$  distribution split in regions for the QCD multijet simulation. Each contribution from each of the regions specified above is shown with a different color.

## 6.2 EW background estimation and fitting procedure

A binned likelihood fit to the data as presented in Ref. [?] is performed simultaneously in the five different control samples and in the signal region, for events selected in both the monojet and mono-V categories, to estimate the  $Z(\nu\nu) + \text{jets}$  and  $W(\ell\nu) + \text{jets}$  rate in each  $p_T^{\text{miss}}$  bin. The full likelihood can be seen below:

$$\begin{aligned}
 \mathcal{L}_c(\mu^{Z \rightarrow \nu\nu}, \mu, \theta) = & \prod_i \text{Poisson} \left( d_i^\gamma | B_i^\gamma(\theta) + \frac{\mu_i^{Z \rightarrow \nu\nu}}{R_i^\gamma(\theta)} \right) \\
 & \times \prod_i \text{Poisson} \left( d_i^Z | B_i^Z(\theta) + \frac{\mu_i^{Z \rightarrow \nu\nu}}{R_i^Z(\theta)} \right) \\
 & \times \prod_i \text{Poisson} \left( d_i^W | B_i^W(\theta) + \frac{f_i(\theta) \mu_i^{Z \rightarrow \nu\nu}}{R_i^W(\theta)} \right) \\
 & \times \prod_i \text{Poisson} \left( d_i | B_i(\theta) + (1 + f_i(\theta)) \mu_i^{Z \rightarrow \nu\nu} + \mu S_i(\theta) \right)
 \end{aligned} \tag{4}$$

In the above likelihood,  $d_i^{\gamma/Z/W}$  are the observed number of events in each bin of the photon, dimuon/dielectron and single-muon/single-electron control regions, and  $B_i^{\gamma/Z/W}$  is the background in the respective control regions. The systematic uncertainties ( $\theta$ ) enter the likelihood as additive perturbations to the transfer factors  $R_i^{\gamma/Z/W}$ , and are modeled as Gaussians. The parameter  $\mu^{Z \rightarrow \nu\nu}$  represents the yield of the  $Z \rightarrow \nu\nu$  background in the signal region, and is left freely floating in the fit. The function  $f_i(\theta)$  is the transfer factor between the  $Z \rightarrow \nu\nu$  and  $W$ +jets backgrounds in the signal region and represents a constraint between these backgrounds. The likelihood also includes the signal region with  $B_i$  representing all the backgrounds,  $S$  representing the nominal signal prediction, and  $\mu$  being the signal strength parameter also left floating in the fit.

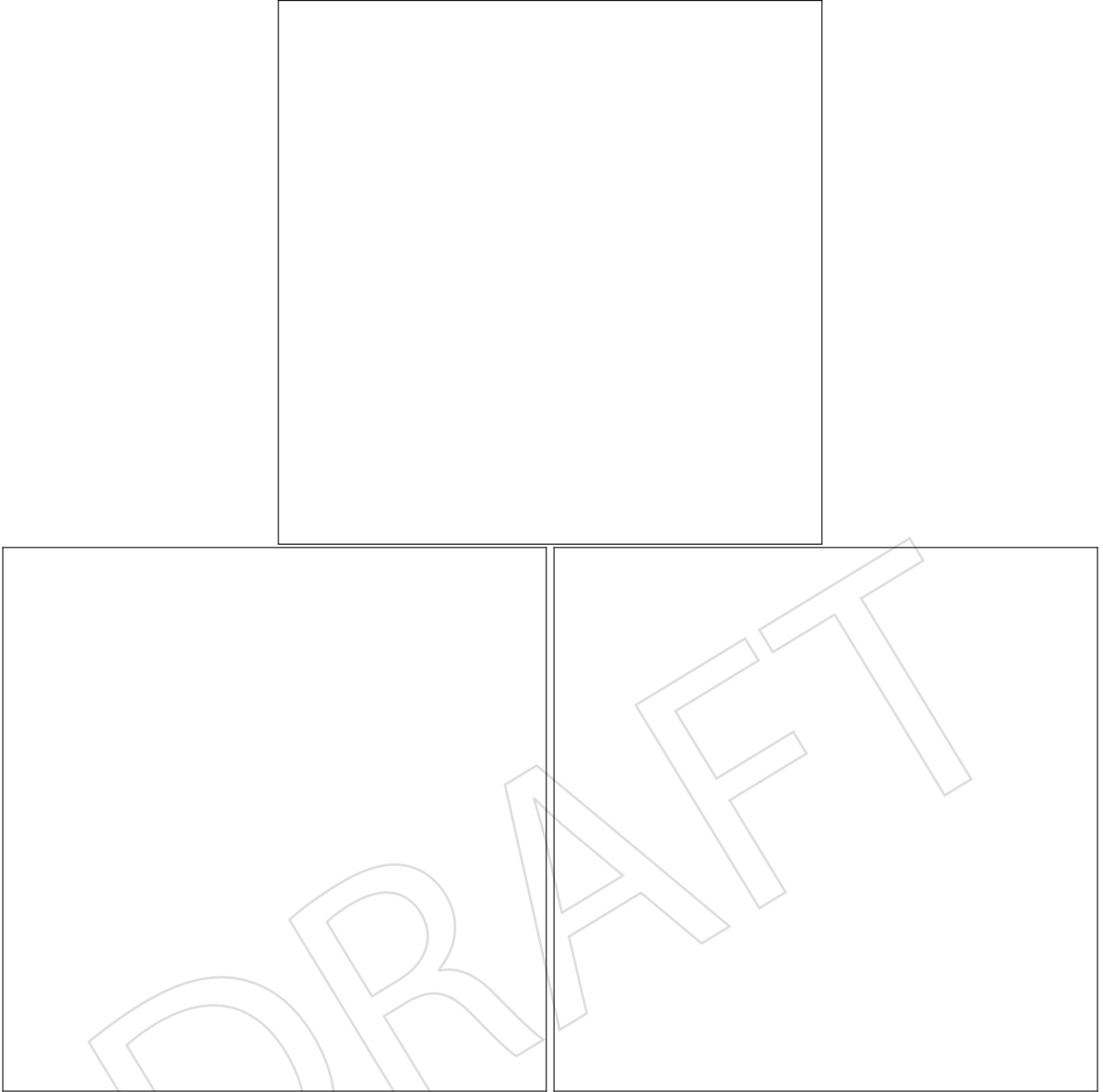


Figure 42: The distribution and the fit on the ratio is shown on the top. The distributions and the fits for the up and down variations are shown on the bottom left and right, respectively.

583 In this likelihood, the expected numbers of  $Z(\nu\nu) + \text{jets}$  events in each bin of  $p_T^{\text{miss}}$  are the free  
 584 parameters of the fit. Transfer factors, derived from simulation, are used to link the yields of the  
 585  $Z(\ell\ell) + \text{jets}$ ,  $W(\ell\nu) + \text{jets}$  and  $\gamma + \text{jets}$  processes in the control regions with the  $Z(\nu\nu) + \text{jets}$  and  
 586  $W(\ell\nu) + \text{jets}$  background estimates in the signal region. These transfer factors are defined as  
 587 the ratio of expected yields of the target process in the signal region and the process being  
 588 measured in the control sample.

589 To estimate the  $W(\ell\nu) + \text{jets}$  background in the signal region, the transfer factors between the  
 590  $W(\mu\nu) + \text{jets}$  and  $W(e\nu) + \text{jets}$  event yields in the single-lepton control samples and the esti-  
 591 mates of the  $W(\ell\nu) + \text{jets}$  background in the signal region are constructed. These transfer fac-  
 592 tors take into account the impact of lepton acceptances and efficiencies, lepton veto efficiencies,  
 593 and the difference in the trigger efficiencies in the case of the single-electron control sample.

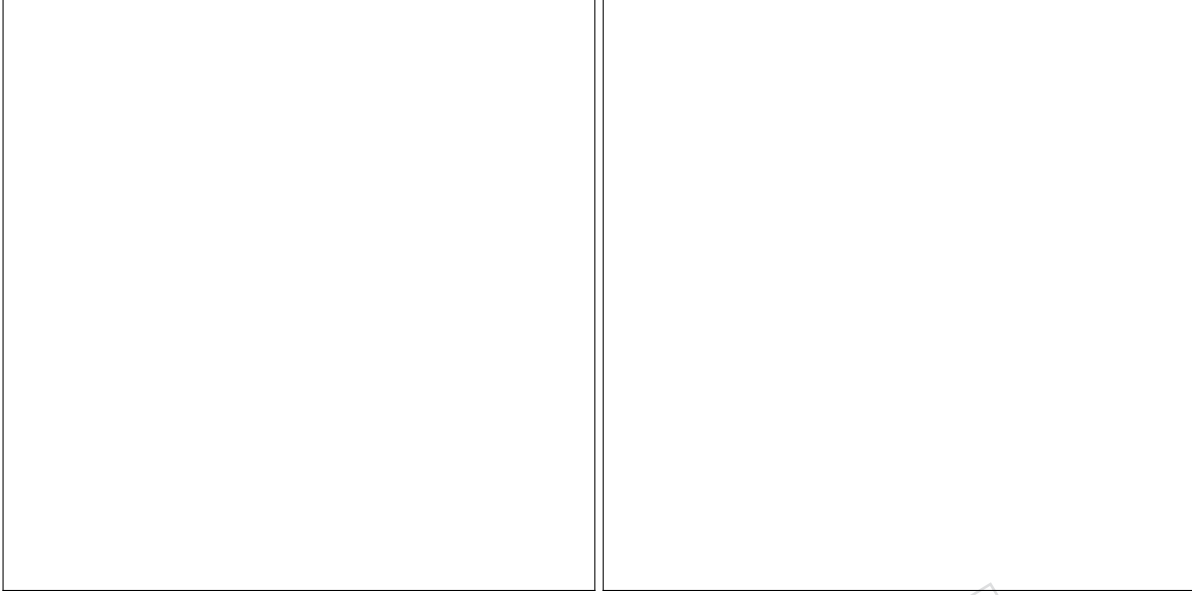


Figure 43: The data driven estimation of multijet events using the fit is compared to the yields obtained by simulation is shown on the left.

The distribution on the right row show the central estimation with the associated uncertainties coming from the systematic fits. The ratio pad shows the % uncertainty per bin.

The  $Z \rightarrow \nu\nu$  background prediction in the signal region is connected to the yields of  $Z \rightarrow \mu^+\mu^-$  and  $Z \rightarrow e^+e^-$  events in the dilepton control samples. The associated transfer factors account for the differences in the branching ratio of Z bosons to charged leptons relative to neutrinos and the impact of lepton acceptance and selection efficiencies. In the case of dielectron events, the transfer factor also takes into account the difference in the trigger efficiencies. The resulting constraint on the  $Z(\nu\nu) + \text{jetsprocess}$  from the dilepton control samples is limited by the statistical uncertainty in the dilepton control samples because of the large difference in branching fractions between Z boson decays to neutrinos and Z boson decays to muons and electrons.

The  $\gamma + \text{jetscontrol}$  sample is also used to predict the  $Z(\nu\nu) + \text{jetsprocess}$  in the signal region through a transfer factor, which accounts for the difference in the cross sections of the  $\gamma + \text{jets}$  and  $Z(\nu\nu) + \text{jetsprocesses}$ , the effect of acceptance and efficiency of identifying photons along with the difference in the efficiencies of the photon and  $p_T^{\text{miss}}$  triggers. The addition of the  $\gamma + \text{jetscontrol}$  sample mitigates the impact of the limited statistical power of the dilepton constraint, because of the larger production cross section of  $\gamma + \text{jetsprocess}$  compared to that of  $Z(\nu\nu) + \text{jetsprocess}$ .

Finally, a transfer factor is also defined to connect the  $Z(\nu\nu) + \text{jets}$  and  $W(\ell\nu) + \text{jetsbackground}$  yields in the signal region, to further benefit from the larger statistical power that the  $W(\ell\nu) + \text{jetsbackground}$  provides, making it possible to experimentally constrain  $Z(\nu\nu) + \text{jetsproduction}$  at high  $p_T^{\text{miss}}$ .

These transfer factors rely on an accurate prediction of the ratio of Z + jets, W + jets, and  $\gamma + \text{jets}$  cross sections. Therefore, LO simulations for these processes are corrected using boson  $p_T$ -dependent NLO QCD K-factors derived using MADGRAPH5\_aMC@NLO. They are also corrected using  $p_T$ -dependent higher-order EW corrections extracted from theoretical calculations [? ? ? ? ? ? ]. The higher-order corrections are found to improve the data-to-

simulation agreement for both the absolute prediction of the individual  $Z + \text{jets}$ ,  $W + \text{jets}$ , and  $\gamma + \text{jets}$  processes, and their respective ratios. All transfer factors are shown in Fig 44 45 46.

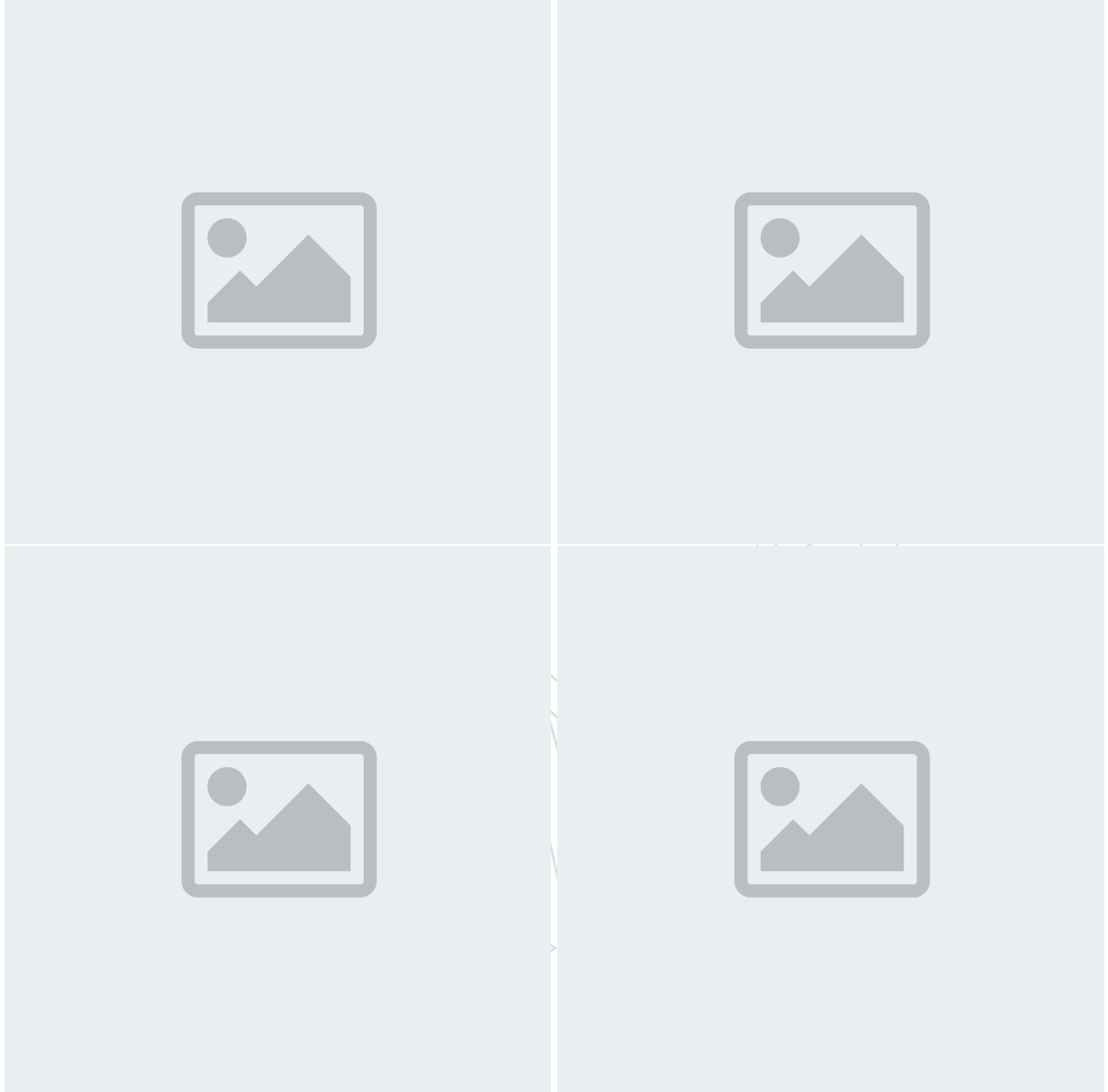


Figure 44: Transfer factors for the  $Z \rightarrow \nu\nu$  background as a function of the recoil using the dimuon, dielectron control regions in monojet (left) and mono-V (right) final state. The orange band shows the theoretical uncertainties on the ratios.

The remaining backgrounds that contribute to the total event yield in the signal region are much smaller than those from  $Z(\nu\nu) + \text{jets}$  and  $W(\ell\nu) + \text{jets}$  processes. These smaller backgrounds include QCD multijet events which are measured from data using a  $\Delta\phi$  extrapolation method [? ?], and top quark and diboson processes, which are obtained directly from simulation.



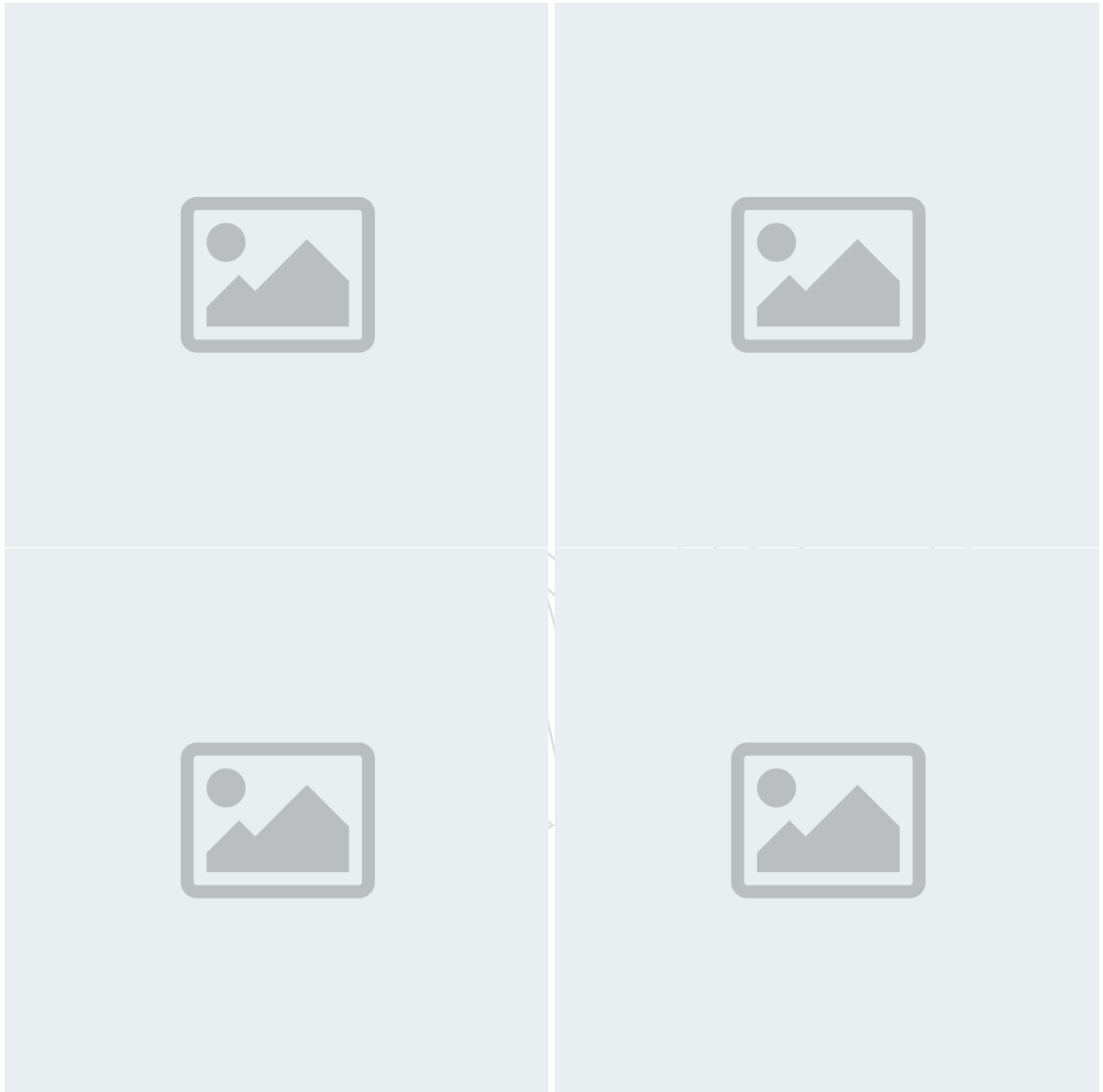


Figure 45: Transfer factors for the to estimate the  $Z \rightarrow \nu\nu$  background from photon control region and  $W$ +jets in the signal region as a function of the recoil for the monojet (left) and mono- $V$  (right) final state.

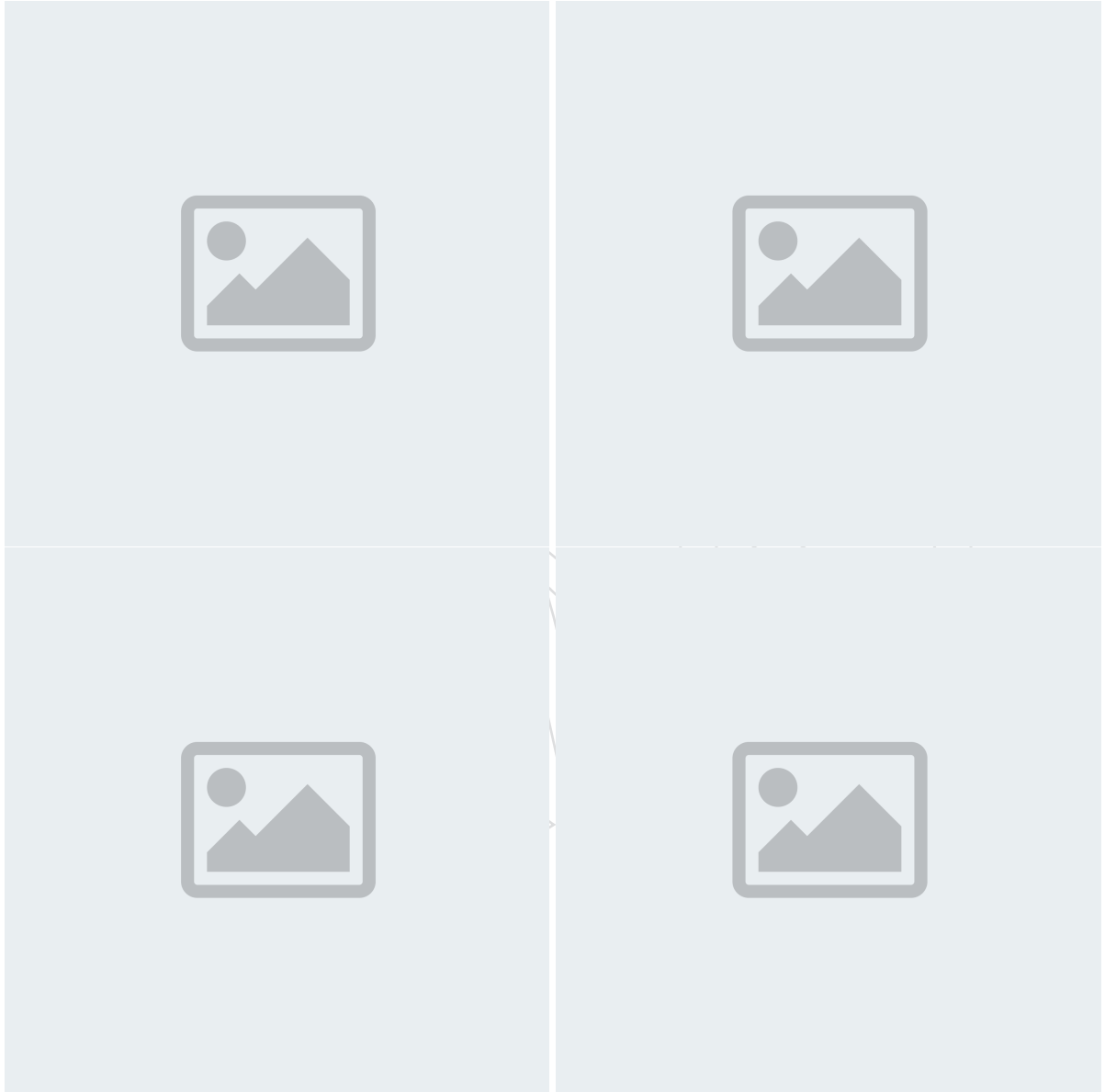


Figure 46: Transfer factors for the  $W(\ell\nu) + \text{jets}$  background as a function of the recoil using the singlemuon, singleelectron control regions in monojet (left) and mono-V (right) final state. The orange band shows the theoretical uncertainties on the ratios.

## 6.3 Systematic Uncertainties

Systematic uncertainty discussion is divided into sections separating out the uncertainties in the transfer factors from the ones in the MC based backgrounds.

### 6.3.1 Uncertainties in the transfer factors

Systematic uncertainties in the transfer factors are modeled as constrained nuisance parameters and include both experimental and theoretical uncertainties in the  $\gamma + \text{jets}$  to  $Z + \text{jets}$  and  $W + \text{jets}$  to  $Z + \text{jets}$  differential cross section ratios.

Theoretical uncertainties in V-jets and  $\gamma + \text{jets}$  processes include effects from QCD and EW higher-order corrections along with PDF modeling uncertainty. To estimate the theoretical uncertainty in the V-jets and  $\gamma + \text{jets}$  ratios due to QCD and EW higher-order effects as well as their correlations across the processes and  $p_T$  bins, the recommendations of Ref. [?] are employed, as detailed in the following explanation.

Three separate sources of uncertainty associated with QCD higher order corrections are used. One of the uncertainties considered comes from the variations around the central renormalization and factorization scale choice. It is evaluated by taking the differences in the NLO cross section as a function of boson  $p_T$  after changing the renormalization and factorization scales by a factor of two and a factor of one-half with respect to the default value. These constant scale variations mainly affect the overall normalization of the boson  $p_T$  distributions and therefore underestimate the shape uncertainties that play an important role in the extrapolation of low- $p_T$  measurements to high- $p_T$ . A second, conservative shape uncertainty derived from altered boson  $p_T$  spectra is used to supplement the scale uncertainties and account for the  $p_T$  dependence of the uncertainties. The modeling of the correlations between the processes assumes a close similarity of QCD effects between all V-jets and  $\gamma + \text{jets}$  processes. However, the QCD effects in  $\gamma + \text{jets}$  production could differ compared to the case of  $Z + \text{jets}$  and  $W + \text{jets}$  productions. In order to account for this variation, a third uncertainty is computed based on the difference of the known QCD K-factors of the  $W + \text{jets}$  and  $\gamma + \text{jets}$  processes with respect to  $Z + \text{jets}$  production. All QCD uncertainties are correlated across the  $Z + \text{jets}$ ,  $W + \text{jets}$ ,  $\gamma + \text{jets}$  processes, and also correlated across the bins of the hadronic recoil  $p_T$ .

For the V-jets and  $\gamma + \text{jets}$  processes, nNLO EW corrections are applied, which correspond to full NLO EW corrections [?] supplemented by two-loop Sudakov EW logarithms [?]. We also considered three separate sources of uncertainty arising from the following: pure EW higher-order corrections failing to cover the effects of unknown Sudakov logarithms in the perturbative expansion beyond NNLO, missing NNLO effects that are not included in the nNLO EW calculations, and the difference between the next-to-leading logarithmic (NLL) Sudakov approximation at two-loop and simple exponentiation of the full NLO EW correction. The variations due to the effect of unknown Sudakov logs are correlated across the  $Z + \text{jets}$ ,  $W + \text{jets}$ , and  $\gamma + \text{jets}$  processes and are also correlated across the bins of hadronic recoil  $p_T$ . On the other hand, the other two sources of EW uncertainties are treated as uncorrelated across the V-jet and  $\gamma + \text{jets}$  processes, and an independent nuisance parameter is used for each process.

A recommendation that includes a factorized approach to partially include mixed QCD-EW corrections is outlined in Ref. [?]. An additional uncertainty is introduced to account for the difference between the corrections done in the multiplicative and the additive approaches, to account for the non-factorized mixed EW-QCD effects.

The summary of the aforementioned theoretical uncertainties including their magnitude and correlation is outlined in Table 9.

Table 9: Theoretical uncertainties considered in the V-jets and  $\gamma + \text{jets}$  processes, and their ratios. The correlation between each process and between the  $p_T$  bins are described.

Uncertainty source	Process (magnitude)	Correlation
Fact. & renorm. scales (QCD)	$Z \rightarrow \nu\nu/W \rightarrow \ell\nu$ (0.1 – 0.5%) $Z \rightarrow \nu\nu/\gamma + \text{jets}$ (0.2 – 0.5%)	Correlated between processes; and in $p_T$
$p_T$ shape dependence (QCD)	$Z \rightarrow \nu\nu/W \rightarrow \ell\nu$ (0.4 – 0.1%) $Z \rightarrow \nu\nu/\gamma + \text{jets}$ (0.1 – 0.2%)	Correlated between processes; and in $p_T$
Process dependence (QCD)	$Z \rightarrow \nu\nu/W \rightarrow \ell\nu$ (0.4 – 1.5%) $Z \rightarrow \nu\nu/\gamma + \text{jets}$ (1.5 – 3.0%)	Correlated between processes; and in $p_T$
Effects of unknown Sudakov logs (EW)	$Z \rightarrow \nu\nu/W \rightarrow \ell\nu$ (0 – 0.5%) $Z \rightarrow \nu\nu/\gamma + \text{jets}$ (0.1 – 1.5%)	Correlated between processes; and in $p_T$
Missing NNLO effects (EW)	$Z \rightarrow \nu\nu$ (0.2 – 3.0%) $W \rightarrow \ell\nu$ (0.4 – 4.5%) $\gamma + \text{jets}$ (0.1 – 1.0%)	Uncorrelated between processes; correlated in $p_T$
Effects of NLL Sudakov approx. (EW)	$Z \rightarrow \nu\nu$ (0.2 – 4.0%) $W \rightarrow \ell\nu$ (0 – 1.0%) $\gamma + \text{jets}$ (0.1 – 3.0%)	Uncorrelated between processes; correlated in $p_T$
Unfactorized mixed QCD-EW corrections	$Z \rightarrow \nu\nu/W \rightarrow \ell\nu$ (0.15 – 0.3%) $Z \rightarrow \nu\nu/\gamma + \text{jets}$ (<0.1%)	Correlated between processes; and in $p_T$
PDF	$Z \rightarrow \nu\nu/W \rightarrow \ell\nu$ (0 – 0.3%) $Z \rightarrow \nu\nu/\gamma + \text{jets}$ (0 – 0.6%)	Correlated between processes; and in $p_T$

Experimental uncertainties including the reconstruction efficiency (1% per muon or electron) and the selection efficiencies of leptons (1% per muon and 2% per electron), photons (2%), and hadronically decaying  $\tau$  leptons (5%), are also incorporated. Uncertainties in the purity of photons in the  $\gamma + \text{jets}$  control sample (2%), and in the efficiency of the electron (2%), photon (2%), and  $p_T^{\text{miss}}$  (1–4%) triggers, are included and are fully correlated across all the bins of hadronic recoil  $p_T$  and  $p_T^{\text{miss}}$ . The uncertainty in the efficiency of the b jet veto is estimated to be 6% (2)% for the contribution of the top quark (diboson) background.

In addition, the lepton veto uncertainties are also included in the transfer factors and are estimated through propagating the overall uncertainty on the tagging scale factor (loose-muon ID, veto-electron ID and loose MVA-tau ID) into the vetoed selection based on the flavour composition of the  $W + \text{jets}$  process. The resulting uncertainties are shown in Fig. 47(b). These uncertainties are only applicable to the leptons in the signal region that are in acceptance and not rejected by the lepton veto. The fraction of each flavor of the leptons both in and out of acceptance are shown in Fig. 48. The overall magnitude of the lepton-veto uncertainty is found to be around 1 (2)% and is found to be dominated by the  $\tau$ -veto uncertainty.

The uncertainty in the efficiency of the V tagging requirements is estimated to be 9% in the mono-V category. The uncertainty in the modeling of  $p_T^{\text{miss}}$  in simulation [?] is estimated to be 4% and is dominated by the uncertainty in the jet energy scale.

The full list of uncertainties on the transfer factors are summarized in Table 10. It should be noted that uncertainties that are common both in the denominator and the numerator of these transfer factors, such as jet energy scale, jet energy resolution, luminosity do not contribute to the total uncertainty as they cancel out in the ratio.

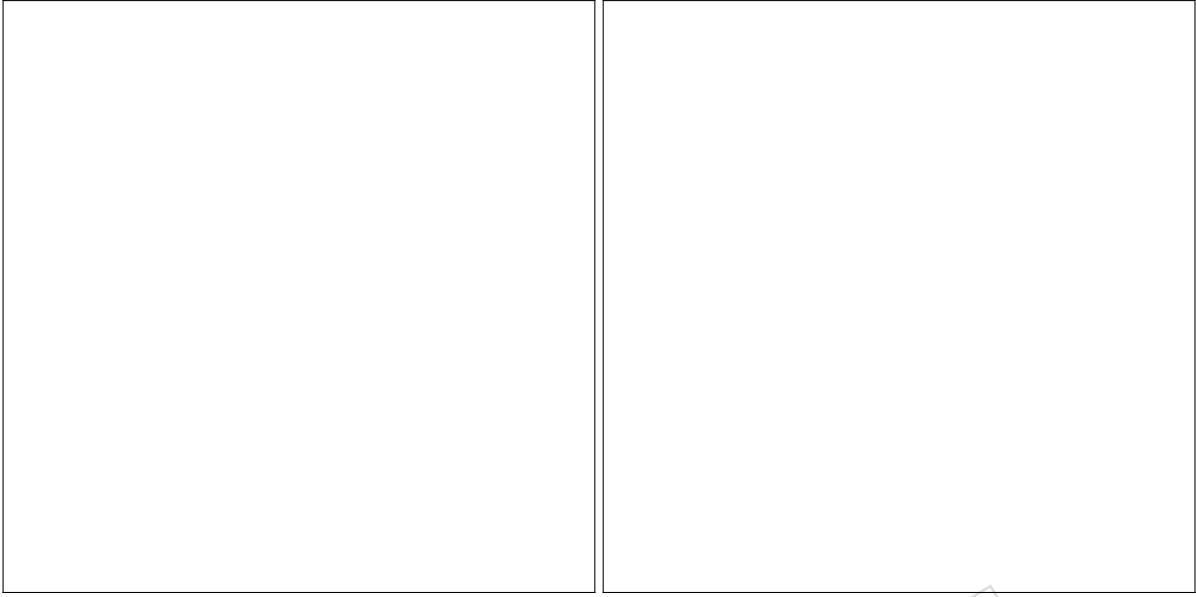


Figure 47: PDF uncertainty and veto uncertainty

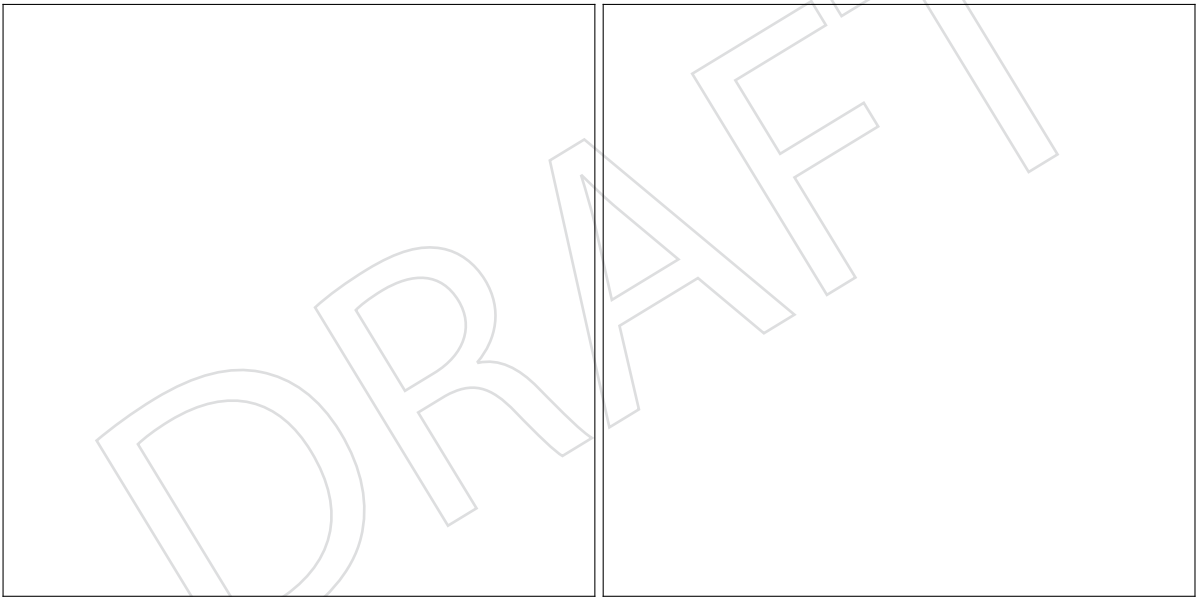


Figure 48: The hadronic  $\tau$  lepton contribution is shown with dotted (solid) blue lines for in (out) of acceptance leptons, where as electron contribution is shown with green, and muon with orange lines for the Wjets background in the signal region after the lepton vetos.

### 6.3.2 Uncertainties assigned to the MC based processes

Uncertainties assigned to the simulation based processes include the uncertainty in the efficiency of the b-jet veto and is estimated to be 3(1)% for the top quark (diboson) background. A systematic uncertainty of 10% is included for the top quark background normalization due to the modeling of the top quark  $p_T$  distribution in simulation. Systematic uncertainties of 10% and 20% are included in the normalizations of the top quark [?] and diboson backgrounds [? ?], respectively, to account for the uncertainties in their cross sections in the relevant kinematic phase space. Lastly, the uncertainty in the QCD multijet background estimate is found to be

Table 10: Experimental uncertainties affecting transfer factors used in the analysis to estimate  $W \rightarrow \ell\nu$  background in the signal region.

Source	Process	Uncertainty
Electron trigger	$W_{\text{SR}}/W_{e\nu}$	1%
$E_{\text{T}}^{\text{miss}}$ trigger	$W_{\text{SR}}/W_{e(\mu)\nu}$	2% (shape)
Muon-reco efficiency	$W_{\text{SR}}/W_{\mu\nu}$	1%
Muon-ID efficiency	$W_{\text{SR}}/W_{\mu\nu}$	1%
Electron-reco efficiency	$W_{\text{SR}}/W_{e\nu}$	1%
Electron-ID efficiency	$W_{\text{SR}}/W_{e\nu}$	1.5%
Muon veto	$W_{\text{SR}}/W_{e(\mu)\nu}$	< 1% (shape)
Electron veto	$W_{\text{SR}}/W_{e(\mu)\nu}$	< 1% (shape)
Tau veto	$W_{\text{SR}}/W_{e(\mu)\nu}$	2% (shape)
PDF	$W_{\text{SR}}/W_{e(\mu)\nu}$	2% (shape)

701 between 50–150% due to the variations of the jet response and the statistical uncertainty of the  
 702 extrapolation factors. All the above mentioned uncertainites are summarized 11.

Table 11: Uncertainties assigned to the simulation based processes.

Source	Process	Uncertainty
Luminosity	All processes	2.3%
Electron trigger	All processes in single-electron CR	1%
$E_{\text{T}}^{\text{miss}}$ trigger	All processes in signal region and single-muon CR	2%
Jet/ $E_{\text{T}}^{\text{miss}}$ energy scale	All processes	4%
Pileup	All processes	2% (shape)
Muon-reco efficiency	All processes in single-muon CR	1%
Muon-ID efficiency	All processes in single-electron CR	1%
Electron-reco efficiency	All processes in single-electron CR	1%
Electron-ID efficiency	All processes in single-electron CR	1.5%
b-jet veto	Top in SR and all CRs	3%
	All remaining in SR and all CRs	1%
Top $p_{\text{T}}$ reweight	Top	10%
Top norm	Top	10%
VV norm	VV	15%
$Z(\ell\ell) + \text{jets}$ norm	$Z(\ell\ell) + \text{jets}$ (SR)	20%
	in SR	50-100% (shape)
Fake muons	in $W_{\mu\nu}$	50%
Fake electrons	in $W_{e\nu}$	50%

## 6.4 Control sample validation

An important cross-check of the application of  $p_T$ -dependent NLO QCD and EW corrections is represented by the agreement between data and simulation in the ratio of  $Z + \text{jetsevents}$  to both  $\gamma + \text{jetsevents}$  and  $W + \text{jetsevents}$  in the control samples, as a function of hadronic recoil  $p_T$ .

Figure ?? shows the ratio between  $Z(\ell\ell) + \text{jets}$  and  $\gamma + \text{jets}$  (left),  $Z(\ell\ell) + \text{jets}$  and  $W(\ell\nu) + \text{jets}$  (middle), and the one between  $W(\ell\nu) + \text{jets}/\gamma + \text{jets}$  processes (right) as a function of the recoil for events selected in the monojet category. While we do not explicitly use a  $W(\ell\nu) + \text{jets}/\gamma + \text{jets}$  constraint in the analysis, the two cross sections are connected through the  $Z + \text{jets}/\gamma + \text{jets}$  and  $Z + \text{jets}/W + \text{jets}$  constraints. Therefore, it is instructive to examine the data-MC comparison of the  $W(\ell\nu) + \text{jets}/\gamma + \text{jets}$  ratio. Good agreement is observed between data and simulation after the application of the NLO corrections as shown in Fig. 49 and Fig. 50 for 2017 and 2018 data respectively.

Figures 51–53 show the results of the combined fit in all control samples and the signal region. Data in the control samples are compared to the pre-fit predictions from simulation and the post-fit estimates obtained after performing the fit. The control samples with larger yields dominate the fit results. A normalization difference of 7% is observed in the pre-fit distributions for the mono-V category in the single-lepton and dilepton control regions. The sources of the differences are identified to be the modeling of the pruned mass variable and the large theoretical uncertainties in the diboson and top quark backgrounds, which are the leading backgrounds in these regions. The normalization difference is found to be fully mitigated by the fitting procedure.

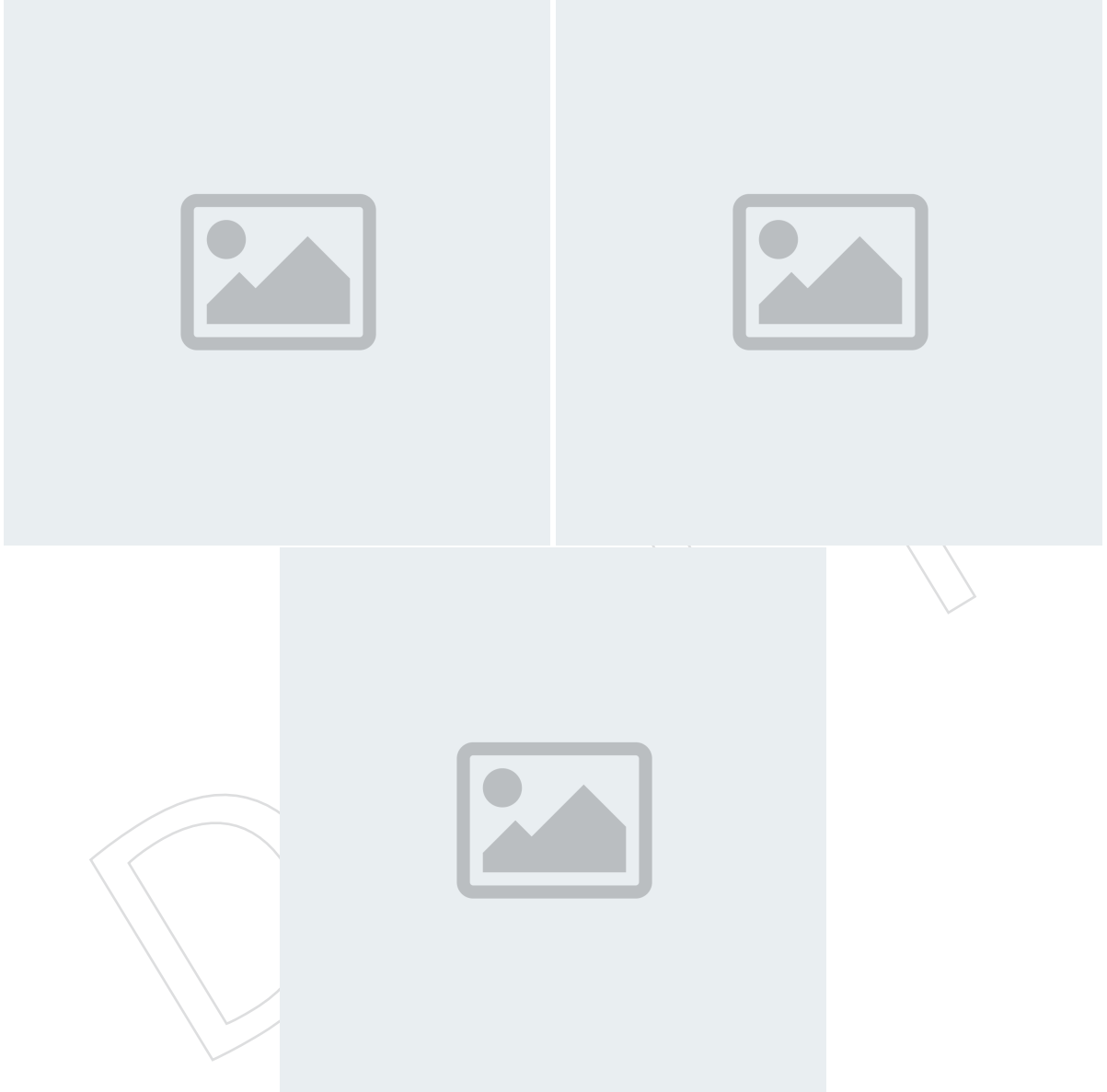


Figure 49: Comparison between 2017 data and MC simulation for the  $Z(\ell\ell)/\gamma + \text{jets}$ ,  $Z(\ell\ell)/W(\ell\nu)$ , and  $W(\ell\nu)/\gamma + \text{jets}$  ratios as a function of the hadronic recoil in the monojet category. In the lower panels, ratios of data with the pre-fit background prediction are shown. The gray bands include both the pre-fit systematic uncertainties and the statistical uncertainty in the simulation.



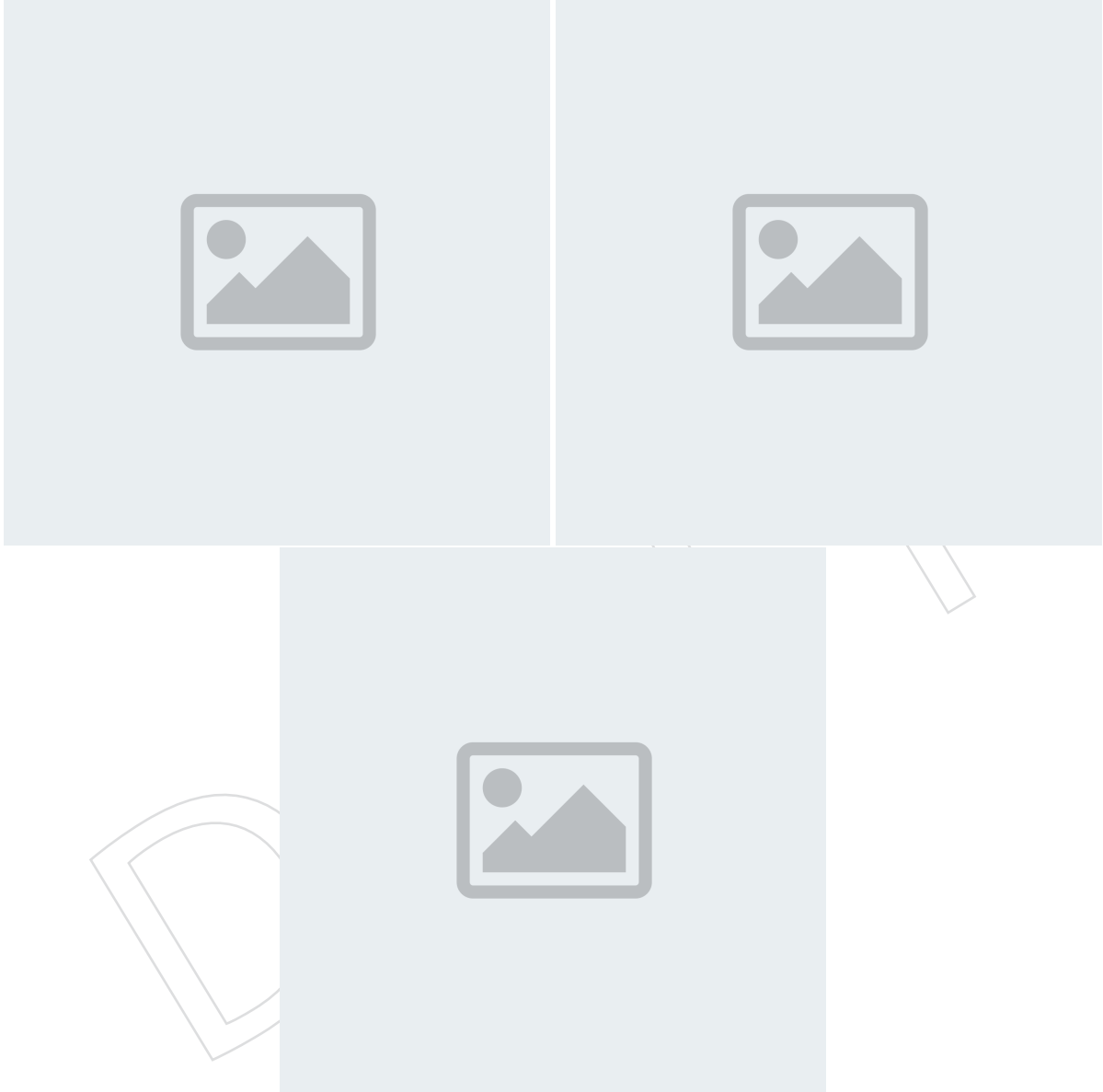


Figure 50: Comparison between 2018 data and MC simulation for the  $Z(\ell\ell)/\gamma + \text{jets}$ ,  $Z(\ell\ell)/W(\ell\nu)$ , and  $W(\ell\nu)/\gamma + \text{jets}$  ratios as a function of the hadronic recoil in the monojet category. In the lower panels, ratios of data with the pre-fit background prediction are shown. The gray bands include both the pre-fit systematic uncertainties and the statistical uncertainty in the simulation.

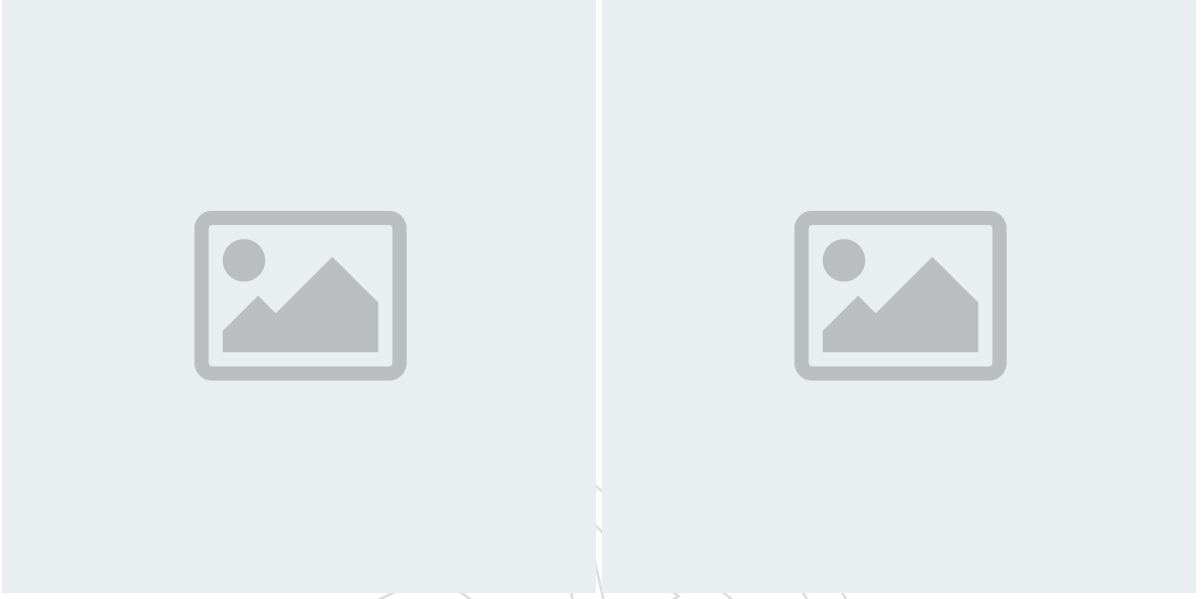


Figure 51: Comparison between data and MC simulation in the  $\gamma + \text{jets}$  control sample before and after performing the simultaneous fit across all the control samples and the signal region assuming the absence of any signal. The left plot shows the monojet category and the right plot shows the mono-V category. The hadronic recoil  $p_T$  in  $\gamma + \text{jet}$  events is used as a proxy for  $p_T^{\text{miss}}$  in the signal region. The last bin includes all events with hadronic recoil  $p_T$  larger than 1250 (750) GeV in the monojet (mono-V) category. In the lower panels, ratios of data with the pre-fit background prediction (red open points) and post-fit background prediction (blue full points) are shown for both the monojet and mono-V categories. The gray band in the lower panel indicates the post-fit uncertainty after combining all the systematic uncertainties. Finally, the distribution of the pulls, defined as the difference between data and the post-fit background prediction relative to the quadrature sum of the post-fit uncertainty in the prediction and statistical uncertainty in data, is shown in the lowest panel.

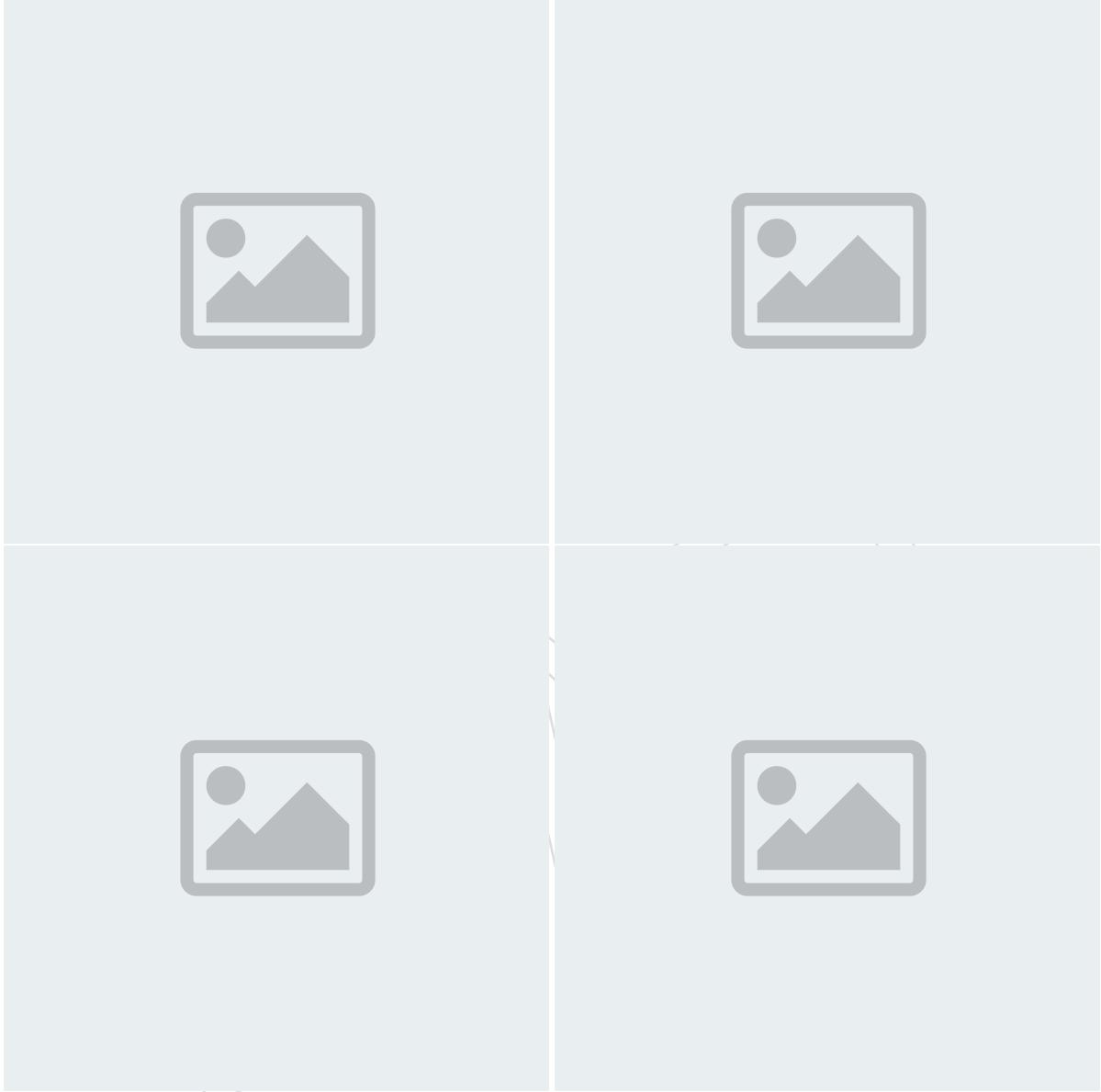


Figure 52: Comparison between data and MC simulation in the dimuon (upper row) and dielectron (lower row) control samples before and after performing the simultaneous fit across all the control samples and the signal region assuming the absence of any signal. Plots correspond to the monojet (left) and mono-V (right) categories, respectively, in the dilepton control sample. The hadronic recoil  $p_T$  in dilepton events is used as a proxy for  $p_T^{\text{miss}}$  in the signal region. The other backgrounds include top quark, diboson, and  $W + \text{jets}$  processes. The description of the lower panels is the same as in Fig. 51.

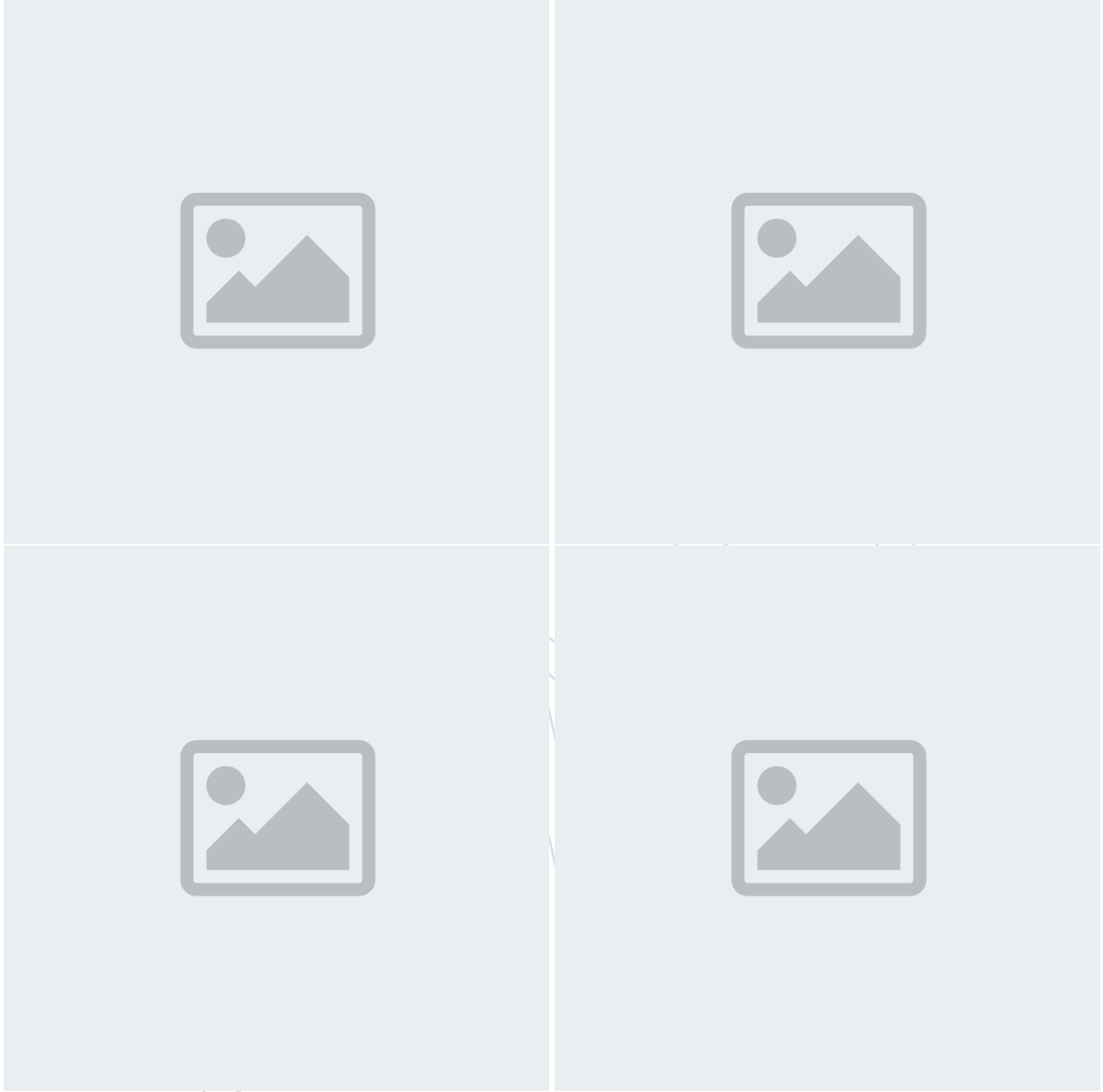


Figure 53: Comparison between data and MC simulation in the single-muon (upper row) and single-electron (lower row) control samples before and after performing the simultaneous fit across all the control samples and the signal region assuming the absence of any signal. Plots correspond to the monojet (left) and mono-V (right) categories, respectively, in the single-lepton control samples. The hadronic recoil  $p_T$  in single-lepton events is used as a proxy for  $p_T^{\text{miss}}$  in the signal region. The other backgrounds include top quark, diboson, and QCD multijet processes. The description of the lower panels is the same as in Fig. 51.

---

## 724 **7 Results**

DRAFT

725 **Acknowledgments**

726 **References**

DRAFT