CMS Draft Analysis Note

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

2019/11/15
Archive Hash: fe73570-D
Archive Date: 2019/11/14

Search for VBF Higgs bosons decaying to invisible particles at 13 TeV with 2017 and 2018 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 $\sqrt{s} = 13\ TeV$ in 2017 and 2018, corresponding to integrated luminosities of $41.3 fb^{-1}$ and $59.7 fb^{-1}$, respectively.

Contents

1 Introduction 2
2 Samples 2
2.1 Data Samples 2
2.2 Background Samples 2
2.3 Signal simulation samples 4
3 Trigger 4
3.1 Efficiency measurement 5
4 Physics objects 9
4.1 Jets 10
4.2 Missing transverse momentum and recoil 11
4.3 Leptons 13
4.4 Photons 14
5 Reweighting of simulated events 16
5.1 Trigger efficiency reweighting 16
5.2 Pileup reweighting 16
5.3 Lepton and photon identification/reconstruction efficiency reweighting . 19
5.4 Higher-order reweighting 25
6 Event selection 30
6.1 Signal region selection 30
6.2 Single muon control region selection 33
6.3 Single electron control region selection 38
6.4 Double muon control region selection 43
6.5 Double electron control region selection 48
6.6 Photon control region 53
7 Background estimation 57
7.1 Signal extraction strategy 57
7.2 Transfer factors 57
7.3 Systematic uncertainties 60
7.4 Control sample validation 62
7.5 QCD multijet estimation 62
8 Results 65
A Comparison of 1D and 2D QCD k-factors 66
B Trigger efficiencies in double muon CR 72

1 Introduction

This analysis note describes a search for events in which the Higgs boson is produced by vector boson fusion (VBF), decaying into invisible particles in the final state. The signature of such a final state will be two seperated jets and an imbalance in $\vec{p}_{\mathrm{T}}$ due to the undetected particles. Two seperated 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 and 2018 at $\sqrt{s} = 13\,\mathrm{TeV}$, corresponding to an integrated luminosity of 41.3 fb^{-1} and 59.7 fb^{-1}, respectively.

The analysis strategy is similar to that of 2016 VBF analysis, with several improvements. One of the major improvements in this analysis is the addition of the photon control region as the fifth control region. The addition of this new control region provides additional constraint and sensitivity on the final fit. In addition to this, the theory scale factors are re-derived in this analysis for VBF topology.

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 fb^{-1}. The MC simulation samples for the background processes have been produced in the `Fall17` and `Autumn18` 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 measurments on physics objects used in the analysis (e.g. trigger turn-ons).

Table 1: List of datasets used to select events in the signal and control regions. Datasets for both years correspond to the `Nano1June2019` campaign, otherwise known as `v5`. For the 2017 data, the `31Mar2018` reconstruction is used for all periods. For 2018 data, the `17Sep2018` reconstruction is used for runs A to C, and the `22Jan2019` 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

2.2 Background Samples

Simulation datasets for the background processes are listed in Table 2 and 3. 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 seperation and invariant mass, along with $p_{\mathrm{T}}^{\mathrm{miss}}$. These processes are as follows:

$\mathbf{Z}(\nu\nu) + \mathbf{jets}$: 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 $H_{\mathrm{T}}$for the QCD case, and in one inclusive sample for the EWK case.

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 $H_{\mathrm{T}}$for the QCD case, and in once inclusive sample for the EWK case.

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 $W \to \ell\nu$, 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 $H_{\mathrm{T}}$for the QCD case, and in once inclusive sample for the EWK case.

Top: Top-quark decays (both $\mathrm{t\bar{t}}$ and single top) also contribute background events to this analysis. In these processes, the W boson produced in a top-quark decay further decays leptonically, which produces genuine $p_{\mathrm{T}}^{\mathrm{miss}}$ in the event. Next-to-leading order (NLO) $\mathrm{t\bar{t}}$ simulation samples have been produced with the aMC@NLO generator with two additional partons in the matrix element. Single-top events have been generated with the Powheg generator at NLO in QCD with one additional matrix element parton.

Dibosons: Decays of diboson pairs (WW, WZ, ZZ) also constitute background processes. Typically, one of the bosons decays leptonically ($W \to \ell\nu$,$Z \to \nu\nu$) while the other boson decays hadronically, thus producing jets and $p_{\mathrm{T}}^{\mathrm{miss}}$ 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 $p_{\mathrm{T}}^{\mathrm{miss}}$. 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 [1] 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 [2] (FxFx [3]) prescription to match jets from matrix element calculations and parton shower description for LO (NLO) samples. The NNPDF 3.1 NNLO [4] parton distribution functions (PDFs) are used in all generated samples. The propagation of all final-state particles through the CMS detector is simulated with the GEANT 4 software [5]. 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).

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 2: Simulated datasets for the modelling of single electroweak boson backgrounds. All datasets are accessed at the NanoAOD data tier from the `v5` campaign, also known as `1Jun19`.

2.3 Signal simulation samples

Simulated signal samples for invisible decays of the Higgs boson are generated using the Powheg generator at NLO in QCD. The samples are normalized to the standard model cross section for Higgs production in the given mode [6]. They are listed in Tab. 4.

3 Trigger

Events for the signal region are collected using a set of dedicated triggers designed to select events with large $p_{\mathrm{T}}^{\mathrm{miss}}$ and large $H_{\mathrm{T}}^{\mathrm{miss}}$ based on the online particle flow (PF) algorithm. In these dedicated trigger algorithms, identified PF muons are removed from the event before the $p_{\mathrm{T}}^{\mathrm{miss}}$ and the $H_{\mathrm{T}}^{\mathrm{miss}}$ 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 [7]. 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

Data set name	Cross section (pb)	Order in QCD
EWKWMinus2Jets_WToLNu_M-50_TuneCP5_13TeV-madgraph-pythia8	20.35	LO
EWKWPlus2Jets_WToLNu_M-50_TuneCP5_13TeV-madgraph-pythia8	25.81	LO
EWKZ2Jets_ZToLL_M-50_TuneCP5_13TeV-madgraph-pythia8	4.321	LO
EWKZ2Jets_ZToNuNu_TuneCP5_13TeV-madgraph-pythia8	10.04	LO
AJJ_EWK_TuneCP5_13TeV_amcatnlo-pythia8	6.096	NLO
GJets_Mjj-500_SM_5f_TuneCP5_EWK_13TeV-madgraph-pythia8	4.937	LO
GJets_SM_5f_TuneCP5_EWK_13TeV-madgraph-pythia8	32.91	LO
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 3: Simulated datasets for the modelling of EWK V production and other processes. All datasets are accessed at the NanoAOD data tier from the `v5` campaign, also known as `1Jun19`.

Data set name	Cross section (pb)	Order in QCD
GluGlu_HToInvisible_M125_TuneCP5_13TeV_powheg_pythia8	48.58	NLO
VBF_HToInvisible_M125_13TeV_TuneCP5_powheg_pythia8	3.782	NLO

Table 4: Simulated signal datasets for the modelling of invisible Higgs boson decays. All datasets are accessed at the NanoAOD data tier from the `v5` campaign, also known as `1Jun19`.

shown in Table 5.

3.1 Efficiency measurement

3.1.1 $p_{\mathrm{T}}^{\mathrm{miss}} + H_{\mathrm{T}}^{\mathrm{miss}}$ triggers

The performance of the $p_{\mathrm{T}}^{\mathrm{miss}} + H_{\mathrm{T}}^{\mathrm{miss}}$ triggers is measured using single muon events. The events are selected from the SingleMuon using the `HLT_IsoMu27` (`HLT_IsoMu24`) 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. 6.2):

1. Veto on additional leptons, photons, b jets, τ_{had} candidates.
2. $\Delta\phi(jet, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}) > 0.5$ for the four leading jets with $p_{\mathrm{T}} > 30$ GeV.

Table 5: 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	`L1_ETMHF70`	MET
	HLT_PFMETNoMu120_PFMHTNoMu120_IDTight_PFHT60	`L1_ETMHF80_HTT60er`	MET
	HLT_Ele35_WPTight_Gsf	`L1_SingleEG24`	SingleElectron

	HLT_Photon200	`L1_SingleEG30`
		`L1_SingleJet170`
		`L1_SingleTau100er2p1`
			SinglePhoton

2018	HLT_PFMETNoMu120_PFMHTNoMu120_IDTight	`L1_ETMHF100` `L1_ETM150`	MET
	HLT_PFMETNoMu120_PFMHTNoMu120_IDTight_PFHT60	`L1_ETMHF90_HTT60er`	MET

	HLT_Ele32_WPTight_Gsf	`L1_SingleIsoEG24er2p1`
		`L1_SingleEG26er2p5`
		`L1_SingleEG60`
			EGamma

	HLT_Photon200	`L1_SingleEG34er2p5`
		`L1_SingleJet160er2p5`
		`L1_SingleJet180`
		`L1_SingleTau120er2p1`
		`L1_SingleEG60`
			EGamma

3. (Calo $p_{\mathrm{T}}^{\mathrm{miss}}$- PF $p_{\mathrm{T}}^{\mathrm{miss}}$) / recoil ¡ 0.5
4. $M_T(\ell, p_{\mathrm{T}}^{\mathrm{miss}}) < 160$ GeV.
5. Leading AK4 jet with $p_{\mathrm{T}} > 80\ GeV$, passing the tight jet ID.
6. Subleading AK4 jet with $p_{\mathrm{T}} > 40\ GeV$.
7. $\Delta\eta_{jj} > 1.0$
8. $\Delta\phi_{jj} < 1.5$

To understand the dependence of the efficiencies on the jet kinematics, we measured the efficiency in three different categories:

- Two central VBF jets: Leading two jets both satisfying $|\eta| \leq 2.4$
- Two forward VBF jets: Leading two jets both satisfying $|\eta| > 2.4$
- One central and one forward VBF jet: One of the leading jets is central with $|\eta| \leq 2.4$ and the other jet in the pair is forward with $|\eta| > 2.4$

In these three categories, efficiencies are measured as a function of M_{jj} and recoil, using both data and MC samples for 2017 and 2018, seperately. Figs. 59, 60, 1 and 2 show the results.

Fig. 1 shows the efficiencies as a function of recoil in data and MC for the three categories, for 2017 samples, whereas Fig. 2 shows the efficiencies as a function of recoil in data and MC for the three categories, for 2018 samples.

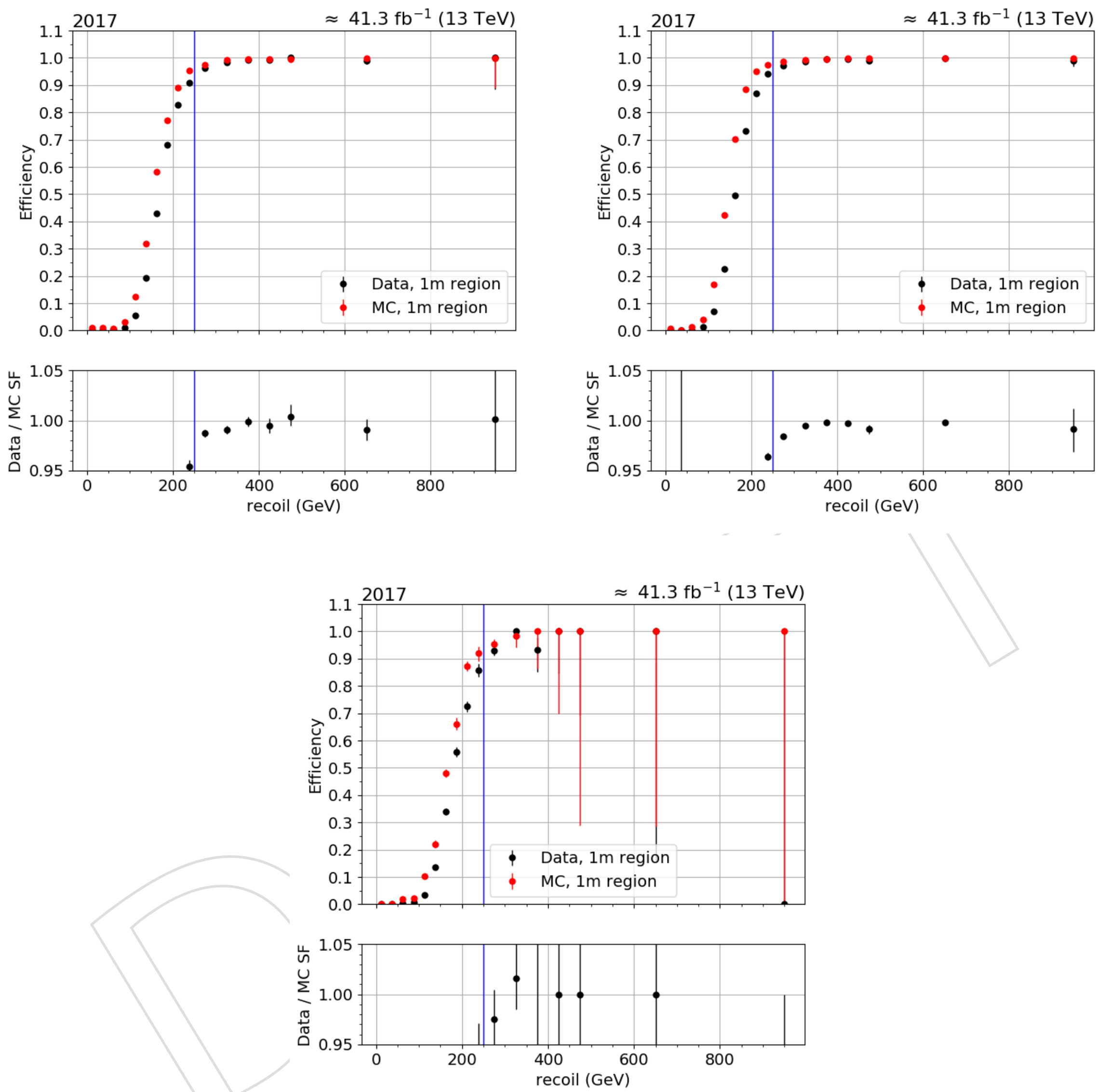

Figure 1: MET trigger efficiency as a function of recoil in three categories: One forward jet and one central jet, two central jets and two forward jets. These results are obtained from 2017 data and MC samples with the selection of single muon events.

3.1.2 Photon trigger

The photon trigger efficiency is measured using events from the JetHT dataset collected with the `HLT_PFHT1050` trigger, which was fully unprescaled in 2017 and 2018 [1]. Events are selected in the same way as for the photon analysis control region (cf. sec. 6.6), except for the photon p_T, recoil and trigger requirements. The trigger efficiency ϵ is then determined as:

[1] The other, prescaled `HLT_PFHTXXX` paths yield lower statistical precision.

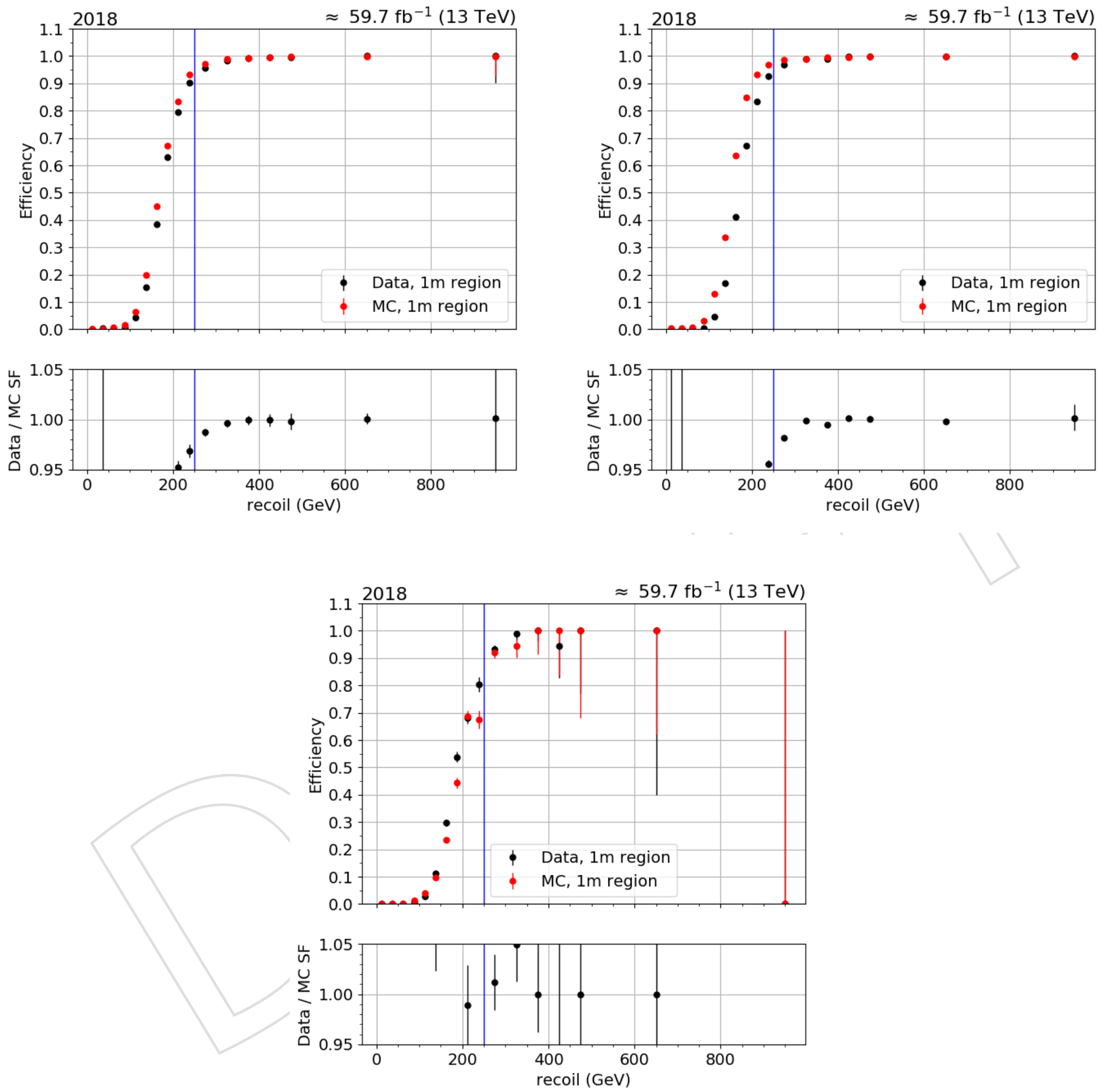

Figure 2: MET trigger efficiency as a function of recoil in three categories: One forward jet and one central jet, two central jets and two forward jets. These results are obtained from 2018 data and MC samples with the selection of single muon events.

$$\epsilon(\texttt{HLT\_Photon200}) = \frac{\text{Offline selection \&\& }\texttt{HLT\_PFHT1050}\text{ \&\& }\texttt{HLT\_Photon200}}{\text{Offline selection \&\& }\texttt{HLT\_PFHT1050}}$$

The resulting efficiency in data and `GJets` H_T-binned simulation is shown in Fig. 3. 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 $215\tilde{G}eV$, where is 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.

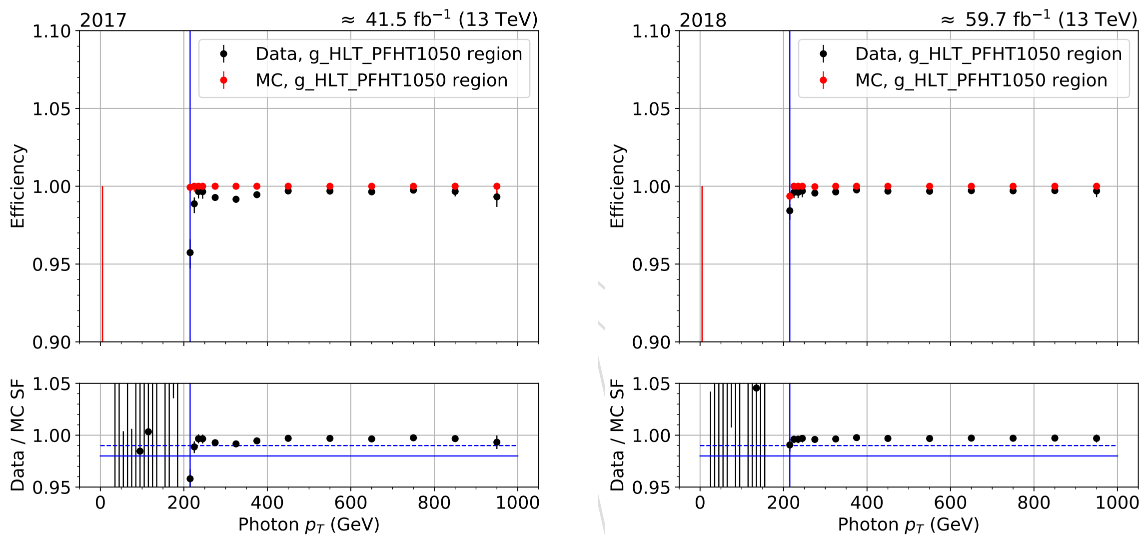

Figure 3: Efficiency of the `HLT_Photon200` trigger in data (black) and H_T-binned `GJets` simulation (red) for 2017 (left) and 2018 (right) as a function of photon p_T. The bottom panel shows the MC-to-data efficiency scale factor. The blue vertical line indicates a photon p_T of 215 GeV, which is the requirement used in the analysis selection.

3.1.3 Electron trigger

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

4 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.

4.1 Jets

In this analysis, jets are reconstructed by clustering PF candidates using the infrared and collinear safe anti-k_T algorithm [8]. 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 [8, 9] 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, γ+jet, and leptonic Z+jet events [10]. The `Fall17_17Nov2017_V32` and `Autumn18_V8` 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 $50 GeV$ 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 η distribution in Fig. 4.

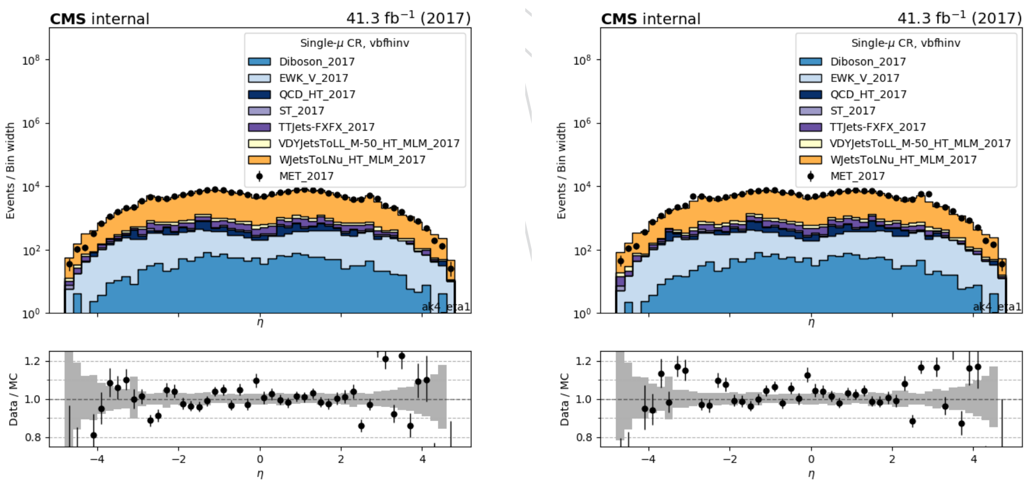

Figure 4: Subleading AK4 jet η distribution in single muon control region with pileup ID requirement (left) and without pileup ID requirement (right).

In addition to the pileup ID requirement, JER corrections are applied to all AK4 jets. Application of this correction on all jets is observed to improve the modeling of the jets, especially near the horn regions near $|\eta| = 2.9$. The effect of JER corrections on the subleading jet eta distribution is shown in Fig. 5.

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

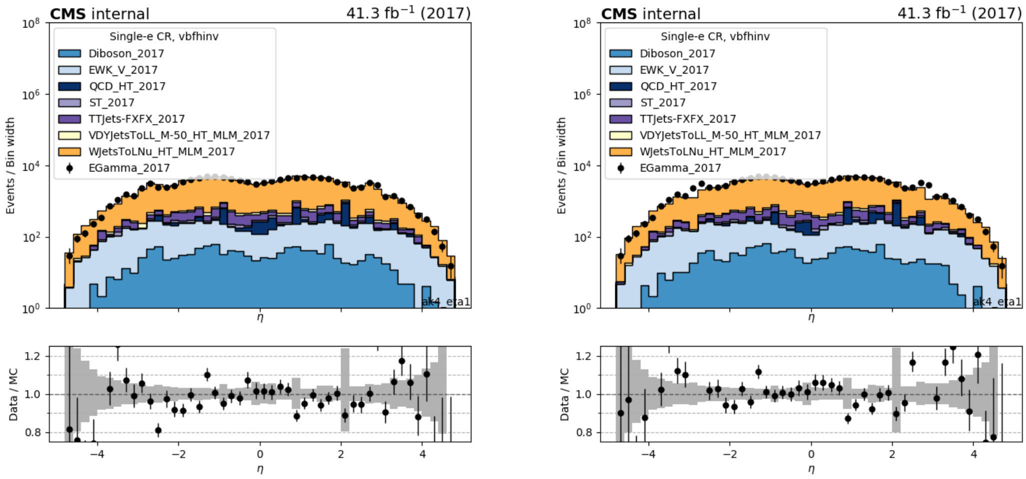

Figure 5: Subleading AK4 jet η distribution in single electron control region with JER corrections applied (left) and without JER corrections applied (right).

4.1.1 b-tagged jets

In this analysis b jets with $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.4$ are identified using the "DeepCSV" algorithm [11], 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.

4.2 Missing transverse momentum and recoil

The vector $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ 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, $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ corresponds to the transverse momentum that is carried by undetected particles such as neutrinos. Practically, $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ is computed as the negative of the vectorial sum of transverse momenta of all PF candidates and is therefore also referred to as PF $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$. The magnitude of the $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ is referred to as $p_{\mathrm{T}}^{\mathrm{miss}}$.

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 $p_{\mathrm{T}}^{\mathrm{miss}}$. The bias on the $p_{\mathrm{T}}^{\mathrm{miss}}$ measurement is reduced by propagating the effect of the jet energy corrections introduced in section 4.1 according to

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

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

This "type-I" correction for $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ uses jet energy scale corrections for all corrected jets with $p_{\mathrm{T}} > 15\,\mathrm{GeV}$ 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, $p_{\mathrm{T}}^{\mathrm{miss}}$ 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 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

$$\vec{U} = \vec{p}_{\mathrm{T}}^{\mathrm{miss}} + \sum_{i \,\in\, \text{leptons, photons}} \vec{p}_{T}^{\,i} \tag{2}$$

where the sum takes into account the leptons and photons used to define the respective control region. The uncertainty of $p_{\mathrm{T}}^{\mathrm{miss}}$ has a strong dependence on the event topology. Therefore, the uncertainty on $p_{\mathrm{T}}^{\mathrm{miss}}$ 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 $p_{\mathrm{T}}^{\mathrm{miss}}$ uncertainty comes from the variations of the jet energy scale correction and the magnitude of the uncertainty is estimated to be 4% for the $\mathrm{Z}(\nu\nu) +$ jets events. This uncertainty is not the most recent one and it will be recalculated.

Anomalous high-$p_{\mathrm{T}}^{\mathrm{miss}}$ 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 supress anomalous high $p_{\mathrm{T}}^{\mathrm{miss}}$ events [12]. The recommended filters are listed in Tab. 6 and are applied in the analysis.

Table 6: The $p_{\mathrm{T}}^{\mathrm{miss}}$ filters recommended by the JME POG [12]. 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 miminize the contribution of anomalous high-$p_{\mathrm{T}}^{\mathrm{miss}}$ events (specifically due to spurious charged hadrons) in this analysis, a quantity based on the relative ratio of calorimetry based $p_{\mathrm{T}}^{\mathrm{miss}}$ and PF based $p_{\mathrm{T}}^{\mathrm{miss}}$ is employed. Events satisfying $|E_{\mathrm{T\,calo}}^{\mathrm{miss}} - E_{\mathrm{T\,PF}}^{\mathrm{miss}}|/U < 0.5$ are selected in this analysis.

4.3 Leptons

4.3.1 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 [13]. 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 [14]. 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. [13] and subtracted from the isolation sum. In this note, 'veto' [15] 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' [15] 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 7.

Table 7: 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)
$\lvert\Delta\eta_{in}\rvert$	< 0.00255 (< 0.00501)	< 0.00463 (< 0.00814)
$\lvert\Delta\phi_{in}\rvert$	< 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 (ρ correction)	$< 0.0287 + 0.506/pT$ $(< 0.0445 + 0.963/p_{\mathrm{T}})$	$< 0.198 + 0.506/p_{\mathrm{T}}$ $(< 0.203 + 0.963/p_{\mathrm{T}})$
1/E - 1/p	< 0.159 (< 0.0197)	< 0.209 (< 0.132)
$\lvert d_{xy}(\mathrm{vtx})\rvert$	< 0.050 (< 0.100)	< 0.050 (< 0.100)
$\lvert d_{z}(\mathrm{vtx})\rvert$	< 0.100 (< 0.200)	< 0.100 (< 0.200)
Expected Inner Missing Hits	$<= 1$ $(<= 1)$	$<= 2$ $(<= 3)$
Pass conversion veto	Yes (Yes)	Yes (Yes)

4.3.2 Muons

Muons within the geometrical acceptance of $|\eta| < 2.4$ are reconstructed by combining information from the silicon tracker and the muon system [16]. 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" [17] muons with $p_{\mathrm{T}} > 10$ GeV are selected with an average efficiency of 98% and are used as a condition to reject events,whereas "tight" [18] muons with $p_{\mathrm{T}} > 20$ 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 χ^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

4.3.3 Taus

Hadronically decaying τ leptons are required to pass identification criteria using the hadron-plus-strips algorithm [19]. 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 τ candidate. Candidate τ jets are reuired to pass both the "DecayModeNewDMs" and "DecayMode" identifiers.

In addition, τ 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. [19]. The "VLoose_IsolationMVArun2v1DBnewDMwLT" isolation working point [20] is employed in this analysis for tau candidates with p_{T} larger than 18 GeV within $|\eta| < 2.3$.

4.4 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 [21]. 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.

Two candidate definitions are employed. "Loose" photons are used to reject events with unwanted photons. These photons are required to pass the EGamma POG 'loose' identification criteria [22], have a transverse momentum of at least 15 GeV and be within $|\eta| < 2.5$. The exact identification criteria are summarized in Table 8. A "tight" photon definition is used for the photon in the dedicated photon control region. These photons are required to be in the barrel ($|\eta| < 1.479$), and have $p_{\mathrm{T}} > 215$ GeV. The POG medium ID is employed, which is summarized in Table 9.

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 8: Loose photon identification criteria.

Variable	Selection Barrel
Full 5x5 $\sigma_{i\eta i\eta}$	< 0.01015
H/E	< 0.02197
charged hadron isolation	< 1.141
neutral hadron isolation	$< 1.189 + 0.01512 \times p_T + 2.259 \times 10^{-5} \times {p_T}^2$
photon isolation	$< 2.08 + 0.004017 \times p_T$
Conversion safe electron veto	Yes

Table 9: Tight photon identification criteria. The criteria are only given for the barrel region since endcap photons are not taken into account.

4.4.1 Photon purity studies

Photons are reconstructed from ECAL clusters, and can discriminated from other sources of ECAL deposits due to the properties of the cluster, as well as their lack of other associated signatures such as tracks or HCAL deposits. This discrimination is not perfect, however, and in some cases, non-photon objects will incorrectly be identified as photons ("fakes"). The leading source of fake photons is QCD production of multijet events, where a jet is misidentified. This process is relevant mainly because of its large cross-section, which yields non-negligible contributions to the photon selection even if the per-jet probability of misreconstruction is small.

To estimate the fake contribution to the photon control region, a purity measurement is performed. The photon purity is defined as the fraction of reconstructed photons that is actually due to an isolated photon from the hard scattering event, as opposed to a fake. The purity is obtained from a template fit to the distribution of the $\sigma_{i\eta i\eta}$ variable in data, which $\sigma_{i\eta i\eta}$ variable represents the width of the ECAL shower in the η direction. Due to the different shower behavior of photons and hadrons, the $\sigma_{i\eta i\eta}$ distribution shows characteristically different behavior between these two classes of reconstructed photons: Real photons show a large peak with a cut-off around $\sigma_{i\eta i\eta} \approx 0.01$, while fake photons will have a smaller peak (stemming from actual photons inside jets), and an additional non-peaking bulk region at larger $\sigma_{i\eta i\eta}$ (stemming from hadrons interacting in the ECAL). The inputs to the fit are defined as follows:

- Photons in data are selected by applying the same identification criteria as for the tight selection defined above, with the exception of the $\sigma_{i\eta i\eta}$ requirement. By removing this requirement, the full $\sigma_{i\eta i\eta}$ distribution can be observed.
- A real photon template is obtained from GJETS simulation. The same identification criteria are applied as in data.
- A fake photon template is obtained from data. In this case, the identification criteria are modified by requiring that at least one of the isolation criteria is not passed. Therefore, the photons in this template do not overlap with the the 'data' template, and represent reconstructed photons inside jets.

The templates are derived in separate bins of the photon p_{T} and the measurement is performed separately for 2017 and 2018. Events are selected using the following criteria:

- At least one jet with $p_{\mathrm{T}} > 100$ and $|\eta| < 2.4$, which is not overlapping with a photon of interest.
- The event must pass the `HLT_Photon200` trigger, as well as the MET filters.
- The event must have $p_{\mathrm{T}}^{\mathrm{miss}} < 60\,\mathrm{GeV}$.

The fits are shown in Figs. 6 and 7, and the resulting purity as a function of photon p_{T} is shown in Fig. 8. While the fits are performed in the full range of $\sigma_{i\eta i\eta} < 0.15$, the purity is evaluated only taking into account the contributions with $\sigma_{i\eta i\eta} < 0.01015$, which is the requirement posed in the identification criteria used in the analysis. The resulting purity ranges between 2 and 4%, with a decreasing trend with photon p_{T}.

5 Reweighting of simulated events

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

5.1 Trigger efficiency reweighting

The efficiency is calculated as a function of the recoil and M_{jj}. 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% (c.f. sec. 3.1).

5.2 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 section of 69.2±4.6% mb, following the recommendations in in Ref. [23]. The true pileup distributions in data and simulation are shown in Fig. 9. The distribution of the number of reconstructed vertices for $W \rightarrow \mu\nu$ events before and after PU reweighting is shown in Fig. 10. 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 ρ is shown in Fig. 10, 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.

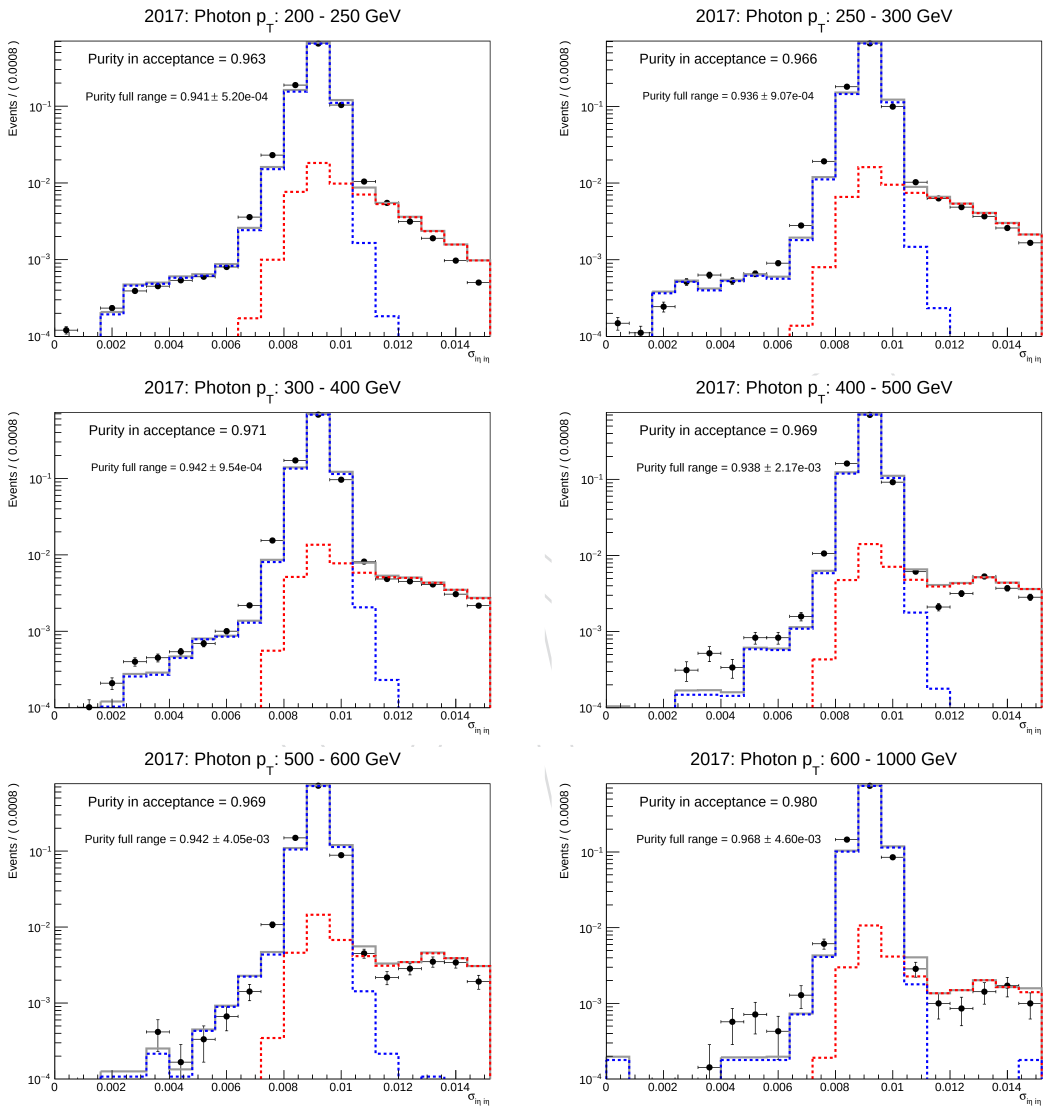

Figure 6: Template fits used to determine the photon purity in the 2017 dataset. The fits are shown in bins of photon p_T, increasing from left to right and top to bottom.

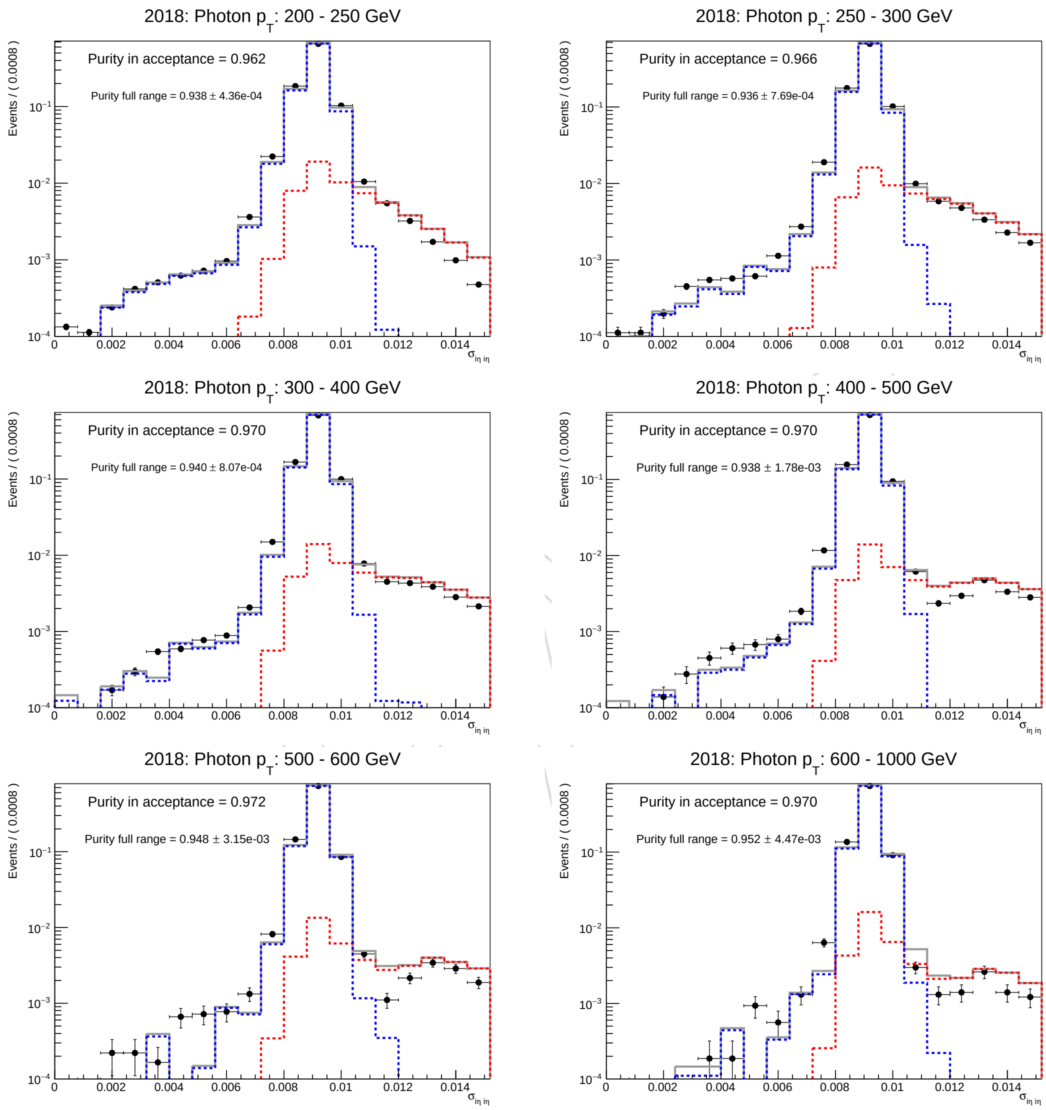

Figure 7: Template fits used to determine the photon purity in the 2018 dataset. The fits are shown in bins of photon p_{T}, increasing from left to right and top to bottom.

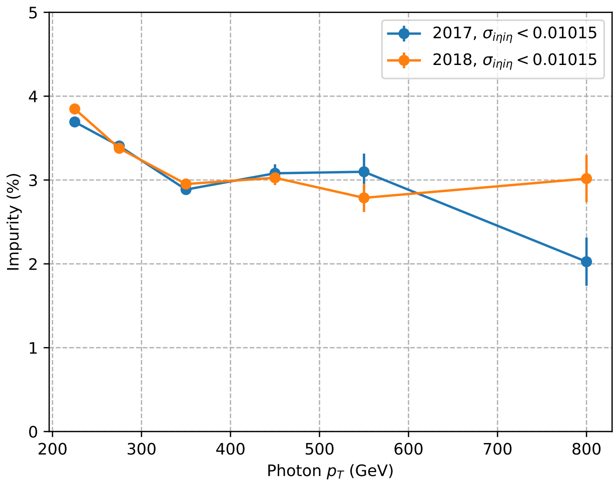

Figure 8: Photon purity as a function of photon p_{T}. This figure shows the aggregated results obtained from the fits shown in Figs. 6 and 7.

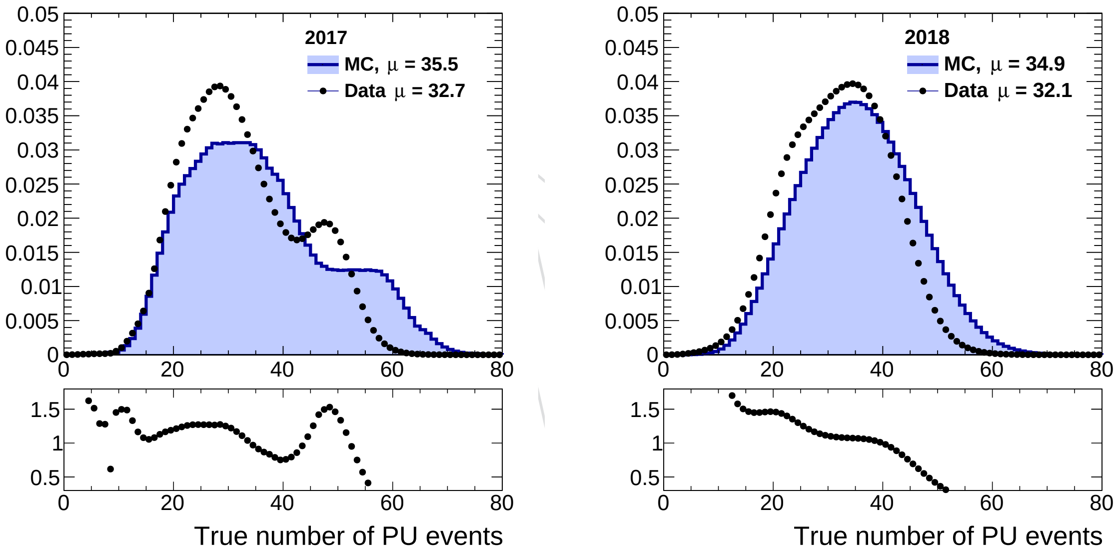

Figure 9: Distribution of the true number of PU events in data and simulation for 2017 (left) and 2018 (right). The distributions for data are extracted assuming a minimum bias cross section of 69.2 mb.

5.3 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 simulationn. 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 η in both data and simulation. These scale factors are provided by the relevant POGs.

The reconstruction scale factors for electrons are shown in Fig. 12. The corresponding identification scale factors for veto and tight electrons are shown in Fig. 13, and include the effect of the isolation efficiency.

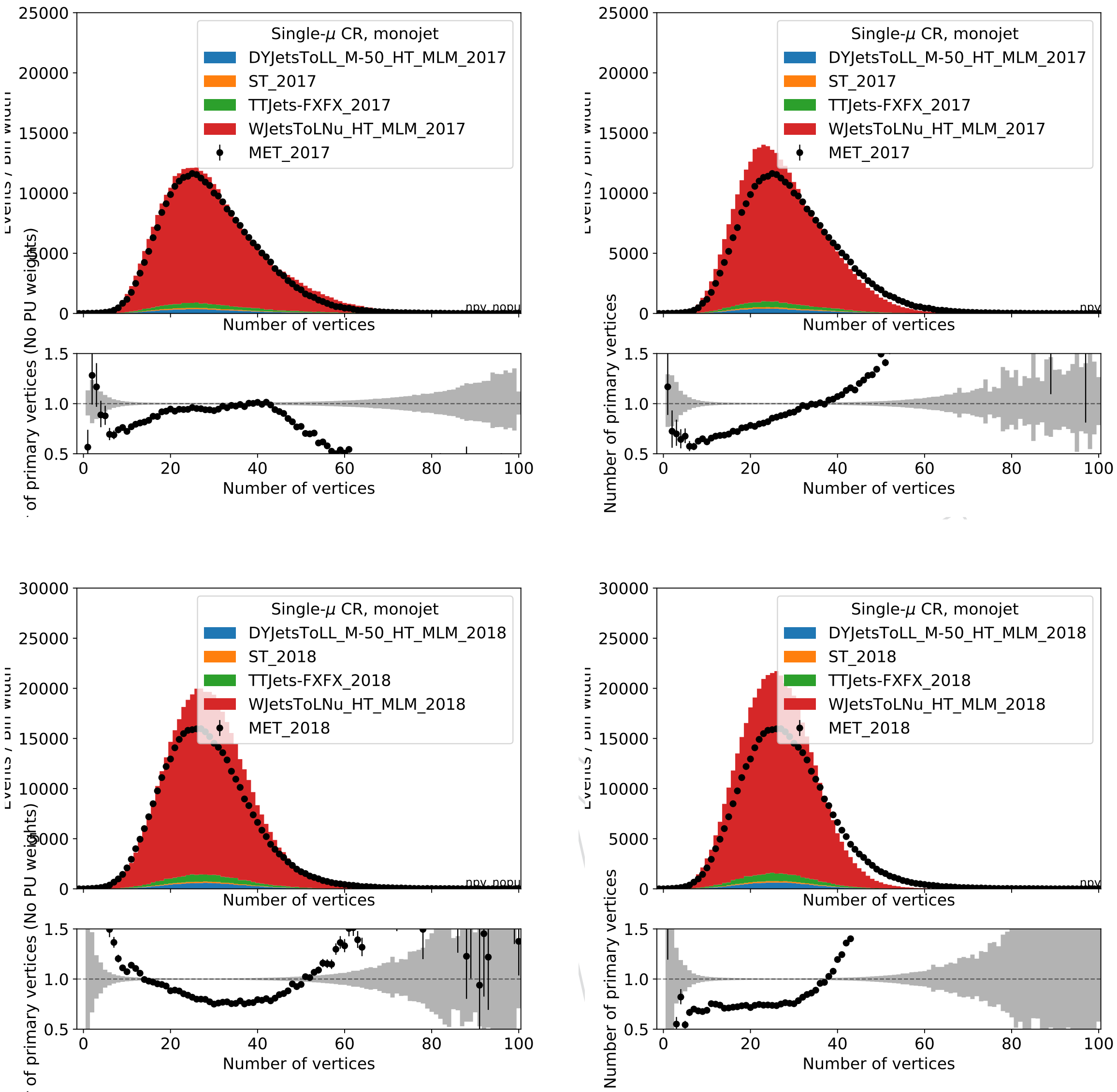

Figure 10: Distribution of the number of vertices in W $\rightarrow \mu\nu$ events in data and simulation before pileup re-weighting (left) and after pileup reweighting (right). The Monte Carlo is normalized to the luminosity of 41.53 and 59.7 fb^{-1}, respectively for 2017 and 2018.

The identification scale factors for muons are shown in Fig. 14. Here, isolation scale factors are applied separately and are shown in Fig. **??**. The corresponding corrections for muons are deemed negligible [24].

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

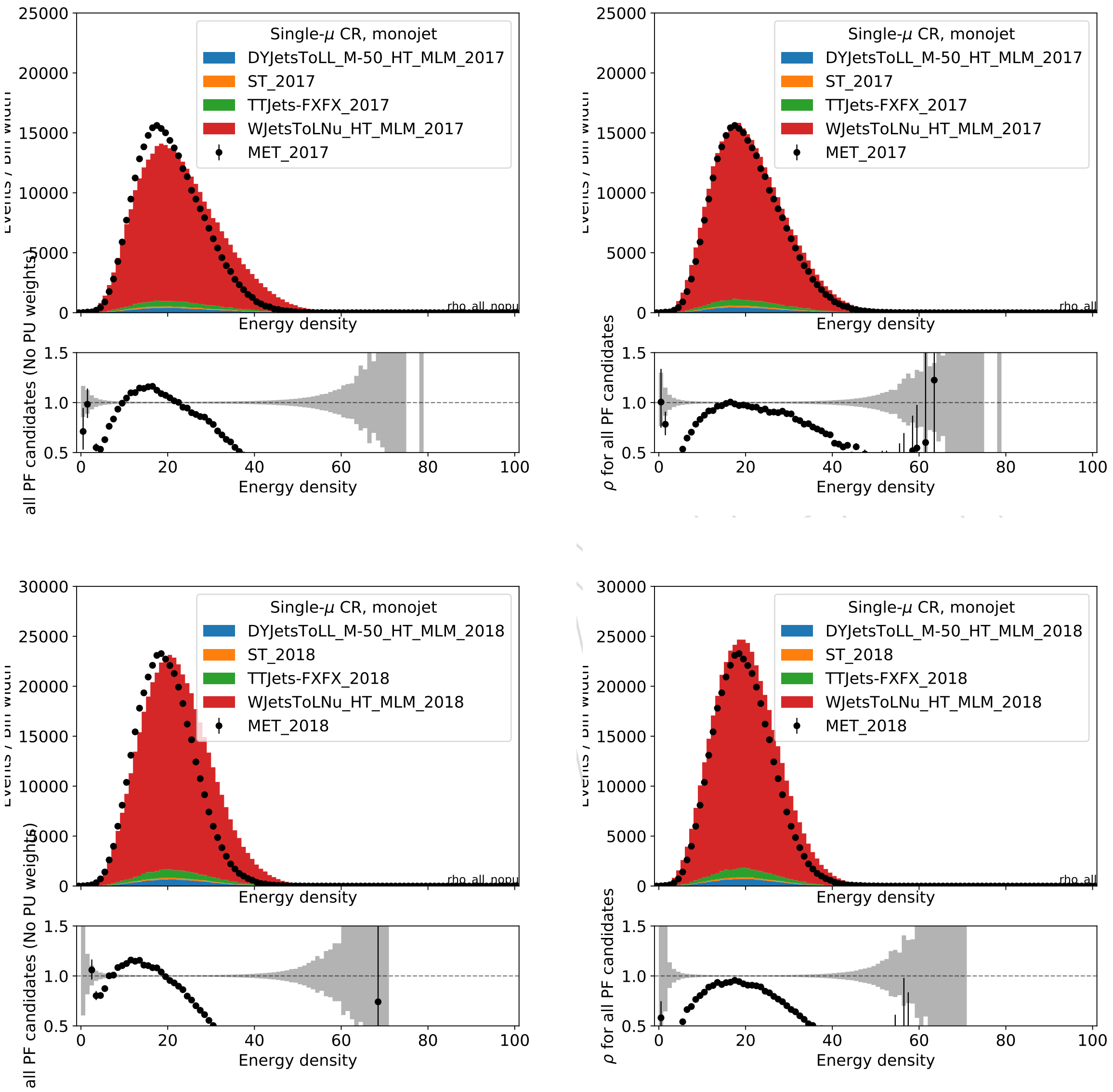

Figure 11: Distribution of the event energy density ρ in $W \rightarrow \mu\nu$ events in data and simulation before pileup re-weighting (left) and after pileup reweighting (right). The Monte Carlo is normalized to the luminosity of 41.53 and 59.7 fb^{-1}, respectively for 2017 and 2018.

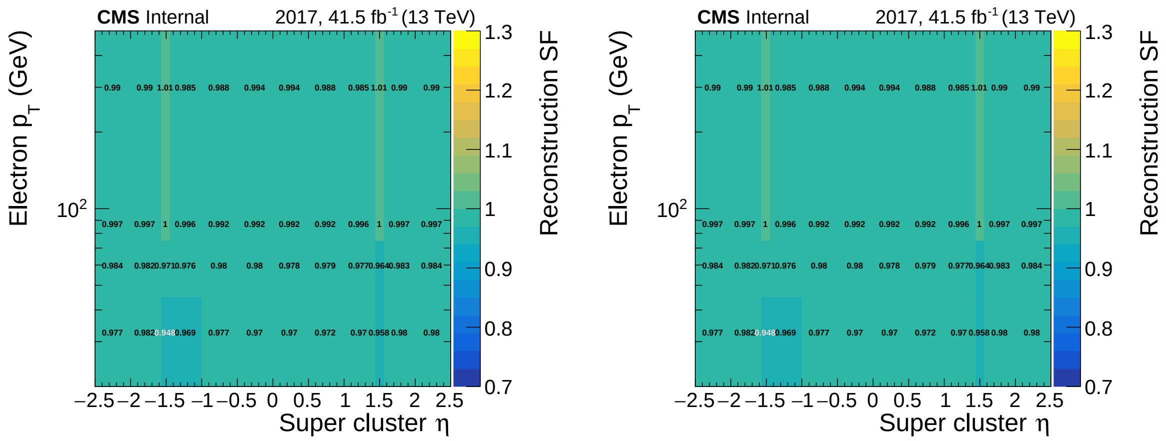

Figure 12: Scale factors for the reconstruction efficiency of electrons starting from a super cluster for 2017 (left) and 2018(right)

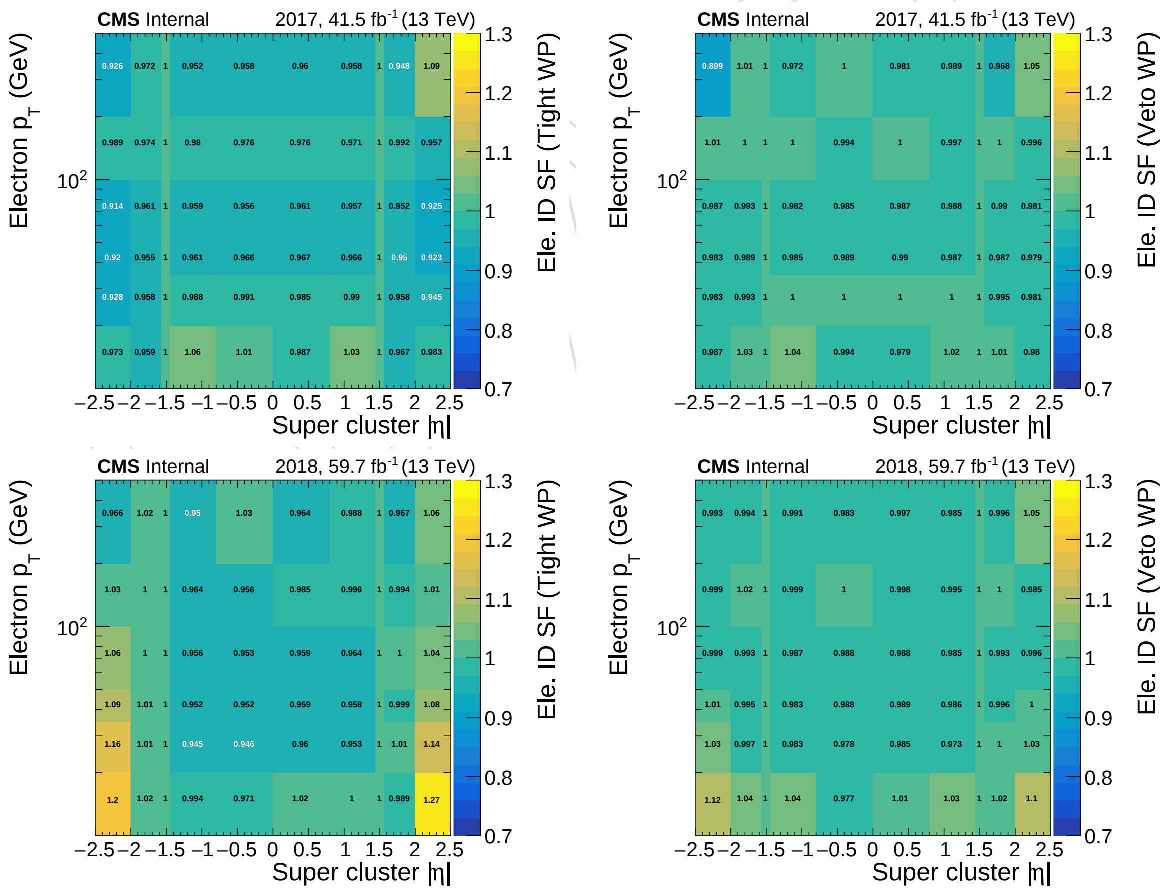

Figure 13: Scale factors for tight (left) and veto (right) electrons are shown for 2017 (top) and 2018 (bottom). The scale factors are provided in bins of electron p_T and η.

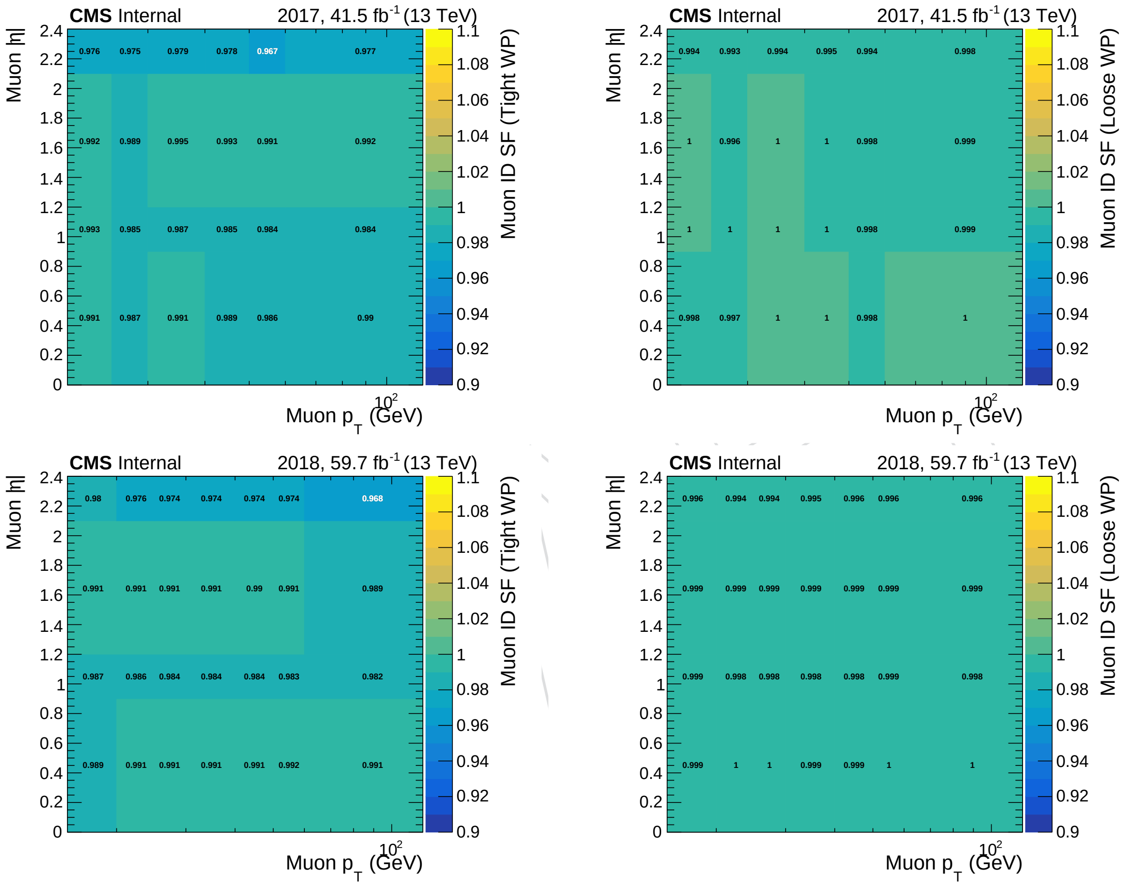

Figure 14: Scale factors for tight (left) and veto (right) muon identification are shown for 2017 (top) and 2018 (bottom). The scale factors are provided in bins of electron p_T and η.

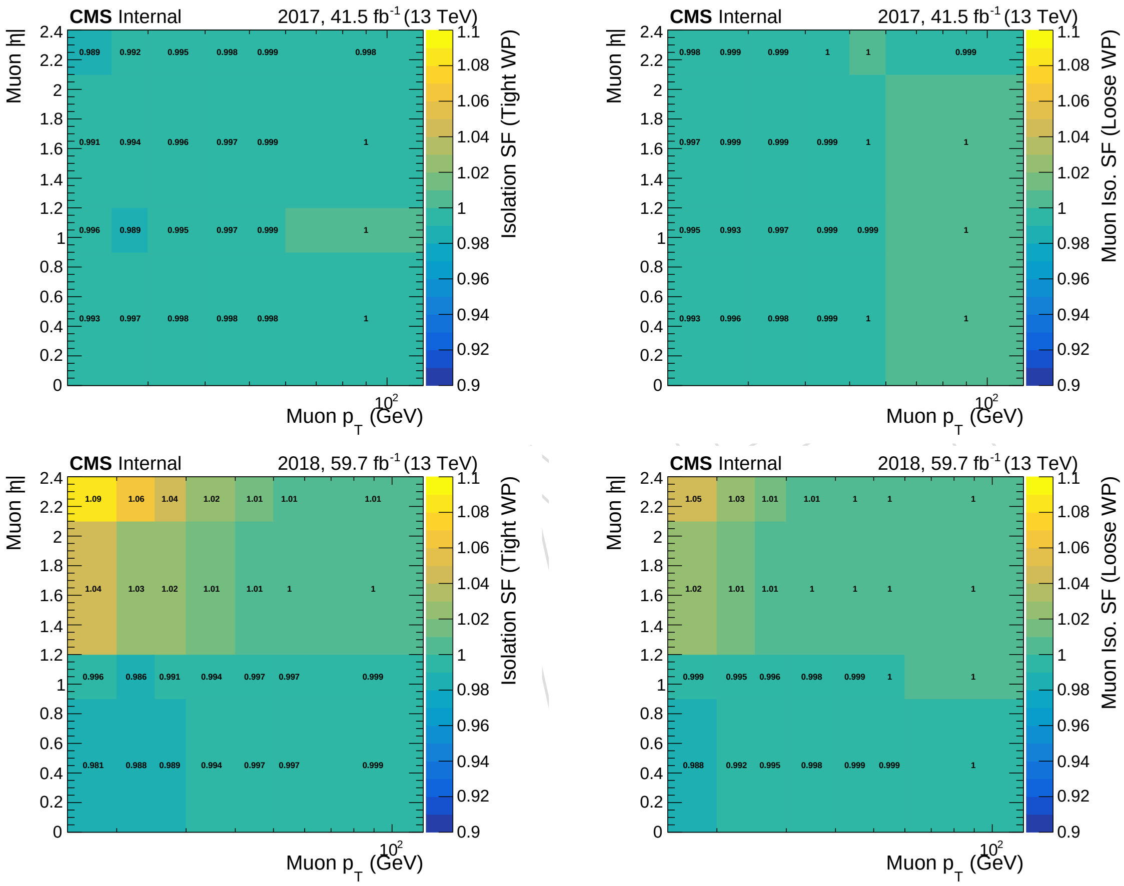

Figure 15: Scale factors for tight (left) and veto (right) muon isolation are shown for 2017 (top) and 2018 (bottom). The scale factors are provided in bins of electron p_{T} and η.

5.4 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. The individual corrections are described in more detail in this section. A concise overview of which corrections are applied to which processes is given in Tab. 10.

Table 10: Summary of higher-order corrections applied to simulated samples. For each boson production process, separate samples and corrections are available for the EWK and QCD production modes. "MC order" reflects the perturbative order used in the generation of the simulation sample, while the further columns represent corrections applied on a per-event level in the analysis process.

Boson	production mode	MC order	NLO QCD	NNLO QCD	NLO EWK
Z	QCD	LO	✓	✓	✓
	EWK	LO	✓	–	–
W	QCD	LO	✓	✓	✓
	EWK	LO	✓	–	–
γ	QCD	LO	✓	✓	✓
	EWK	LO	–	–	–

5.4.1 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 $p_{\mathrm{T},V}$. 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. τ leptons with generator status 2. As τ leptons are unstable, they are not present as final state particles (status 1) in the generator record. The τ lepton before its decay has status 2.

3. neutrinos with generator status 1.

The dilepton candidates are checked for flavour consistency with the desired boson candidate. If multiple candidates are found in an event, the one with the highest invariant mass is used.

For photon events, the generator photon with highest p_{T} and status 1 is used.

5.4.2 QCD NLO corrections to QCD V processes

Scale factors corresponding to NLO QCD corrections for W and Z production are obtained from central CMS. For the DY and W processes, samples from "Fall17" campaign, while "Summer16" samples are used for the γ+jets process. In both cases, all samples are generated using `MadGraph5_amc@NLO`. The LO samples are binned in HT and are equivalent 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 `Pythia8` with tune `CP5`.

The scale factors are derived by obtaining the distribution of interest at the generator-level in both samples, normalizing the distributions to their respective cross sections, and then dividing 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. At least two generator-level jets, with the leading (trailing) p_T of at least 80 GeV (40 GeV) and $|\eta| < 4.7$.
2. The two leading jets must be in opposite hemispheres of the detector, $\eta_1 \times \eta_2 < 0$.
3. 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 $p_T > 30$ GeV are considered.

Compared to an inclusive derivation of the SF, the inclusion of the selection criteria leads to an increase in the value of the SF of about $10 - 12\%$.

The scale factors are derived either as a one-dimensional function of the generator-level boson $p_{T,V}$ ("1D")or two-dimensionally in $p_{T,V}$ and M_{jj}("2D").

The 1D SFs are shown in Fig. 16. To protect the outcome of the reweighting procedure from binning effects, the binned scale factor is interpolated using a falling exponential function:

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

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. 16.

The 2D versions of the SFs are only derived for the W and DY processes, for which the available simulation samples are bigger and thus allow for finer binning. The scale factors are shown in Fig. 17.

5.4.3 QCD NNLO corrections to QCD V processes

NNLO corrections are obtained from the fixed-order calculations results in [25]. They are parametrized as a function of the generator-level boson p_T. The correction is shown in Fig. 18.

5.4.4 EW NLO corrections to QCD V processes

Scale factors corresponding to NLO EW corrections are obtained from Ref. [25] and applied as a function of the generator-level boson p_T. The scale factors are shown in Fig. 19.

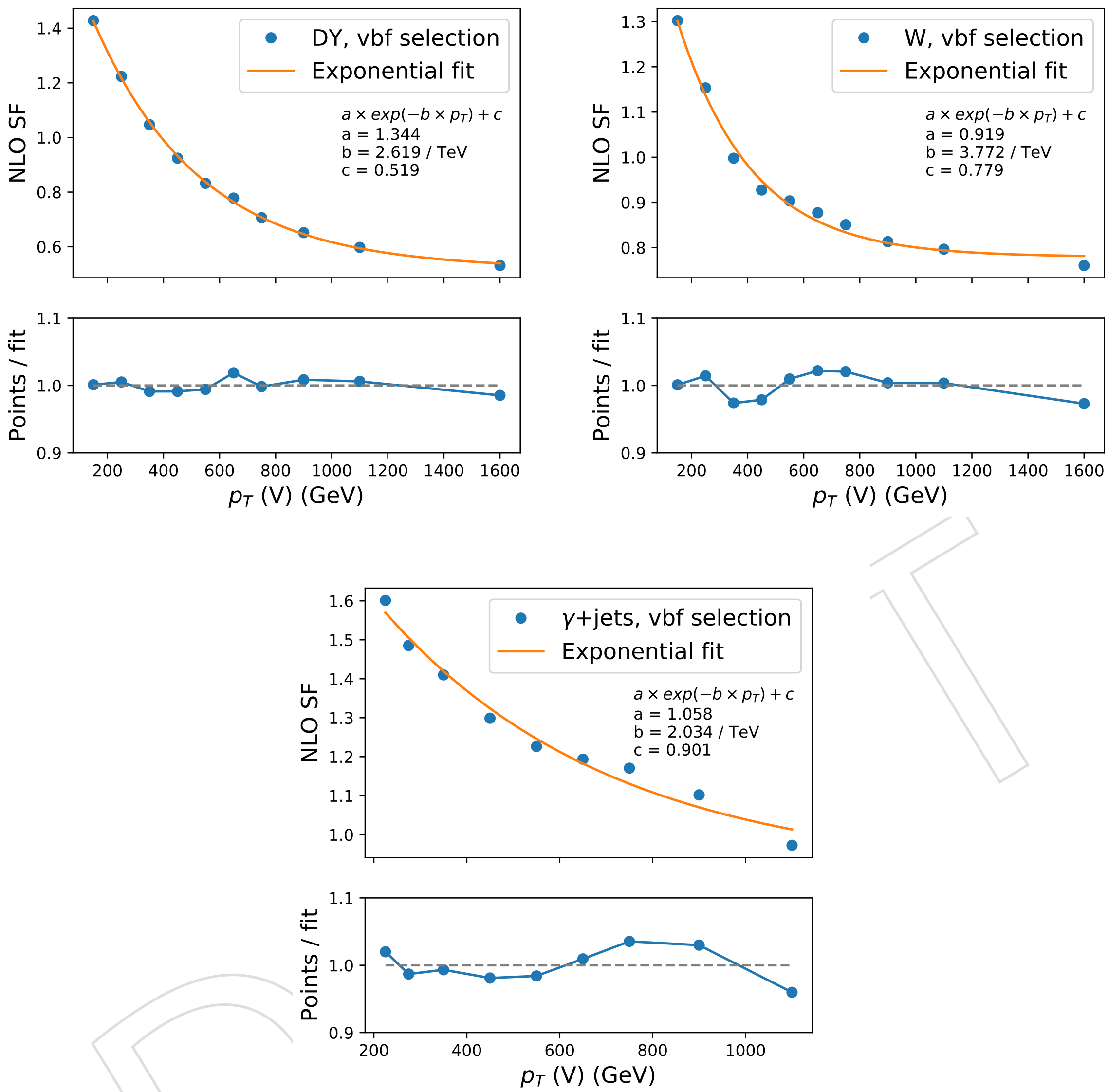

Figure 16: QCD NLO scale factors for the DY (top left), W (top right) and γ+jets processes. The k factors are derived within the generator-level VBF selection described in the text. In the top panel of each plot, the blue markers show the NLO SF derived from the simulated samples. The orange line shows 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.

5.4.5 QCD NLO corrections to EWK V processes

The QCD NLO corrections to EWK W and Z production have been calculated in Ref. [26] using the VBF@NLO program. They are parametrized in $p_{\mathrm{T},V}$ and $M_{jj}$and are shown in Fig. 20.

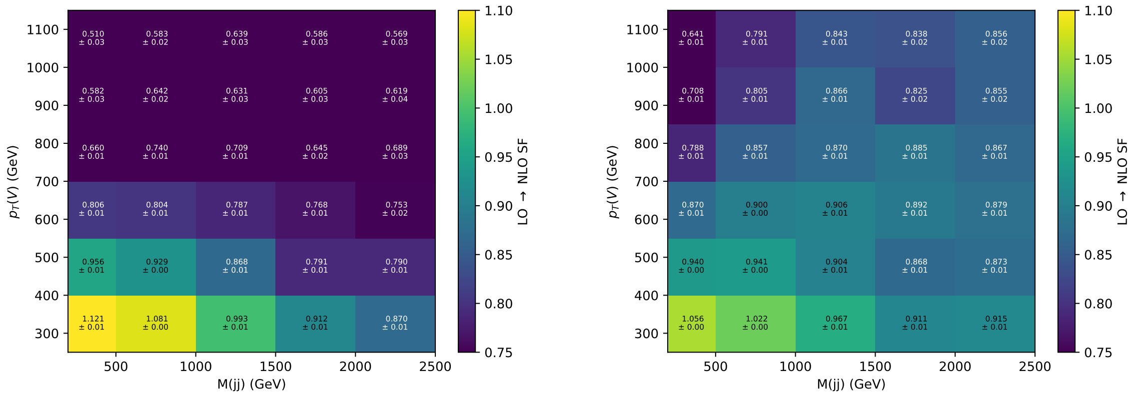

Figure 17: Same as Fig. 16, but now binned in two dimensions of the generator-level boson p_{T} and M_{jj}. The k factors are derived within the generator-level VBF selection described in the text. The uncertainties quoted in each bin are the statistical uncertainties due to the finite size of simulated samples.

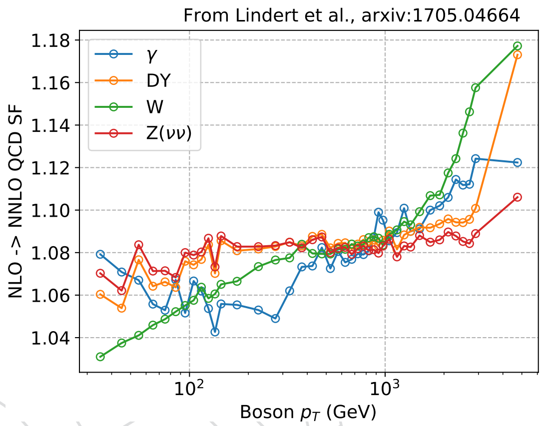

Figure 18: QCD NNLO scale factors for DY, W and photon production as a function of $p_{\mathrm{T},V}$.

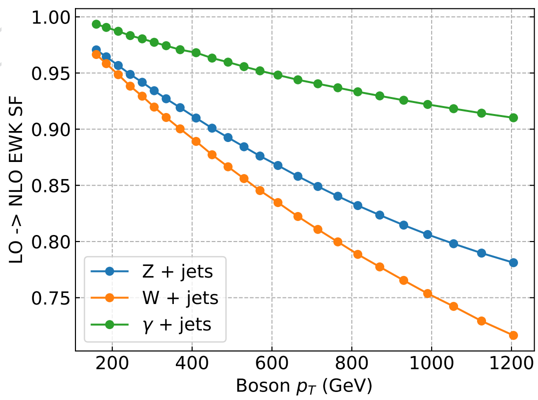

Figure 19: EW NLO scale factors for DY, W and photon production as a function of $p_{\mathrm{T},V}$.

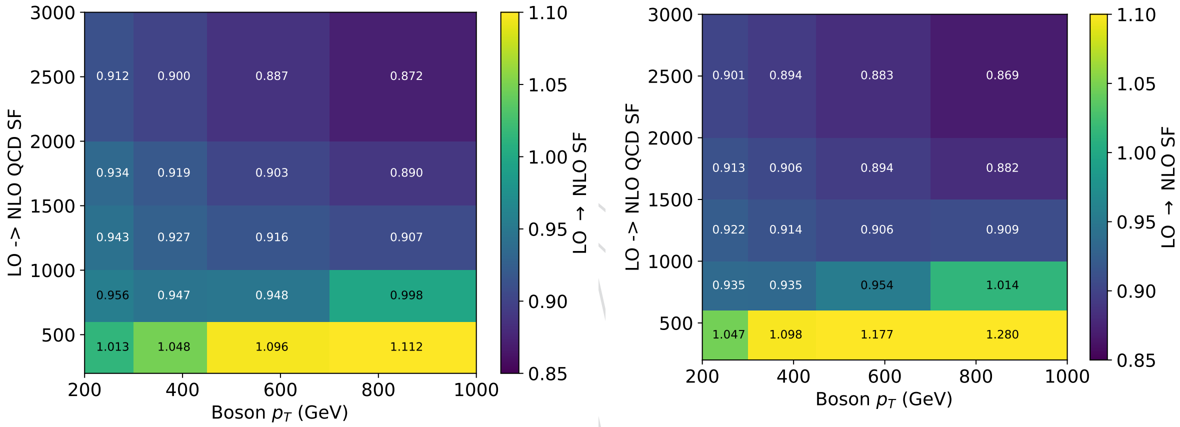

Figure 20: QCD NLO scale factors for EWK DY, W production of $p_{\mathrm{T},V}$ and M_{jj}.

6 Event selection

6.1 Signal region selection

Signal region events are selected using triggers with thresholds of 120 GeV on both $p_{\mathrm{T,trig}}^{\mathrm{miss}}$ and $H_{\mathrm{T,trig}}^{\mathrm{miss}}$. The $p_{\mathrm{T,trig}}^{\mathrm{miss}}$ corresponds to the magnitude of the vector $\vec{p}_{\mathrm{T}}$ sum of all the PF candidates reconstructed at the trigger level, while the $H_{\mathrm{T,trig}}^{\mathrm{miss}}$ is computed as the magnitude of the vector $\vec{p}_{\mathrm{T}}$ sum of jets with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ 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 prediction, muon candidates are not included in the $p_{\mathrm{T,trig}}^{\mathrm{miss}}$ nor $H_{\mathrm{T,trig}}^{\mathrm{miss}}$ computation. The trigger efficiency is measured to be 96% for events passing the analysis selection for $p_{\mathrm{T}}^{\mathrm{miss}} > 250\,\mathrm{GeV}$ and becomes more than 99% efficient for events with $p_{\mathrm{T}}^{\mathrm{miss}} > 350\,\mathrm{GeV}$.

Candidate events are required to have $p_{\mathrm{T}}^{\mathrm{miss}} > 250\,\mathrm{GeV}$. The leading AK4 jet in the signal event is required to have $p_{\mathrm{T}} > 80\,\mathrm{GeV}$ and $|\eta| < 4.7$, and the subleading AK4 jet is required to have $p_{\mathrm{T}} > 40\,\mathrm{GeV}$ and $|\eta| < 4.7$. In addition, if the leading jet is within the tracker range, $|\eta| < 2.5$, it is required to have at least 10% of its energy coming from charged particles and less than 80 % of its energy attributed to neutral hadrons, as discussed in section 4. This selection helps to remove events originating from beam-induced backgrounds. In addition, the analysis employs various event filters to reduce events with large misreconstructed $p_{\mathrm{T}}^{\mathrm{miss}}$ [27] originating from noncollision backgrounds.

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 seperation and small azimuthal seperation. Therefore, this analysis employs several requirements on M_{jj}, $\Delta\eta_{jj}$ and $\Delta\phi_{jj}$, which can be found in Table 11.

The main background processes in this search are the $\mathrm{Z}(\nu\nu)+$jets and $\mathrm{W}(\ell\nu)+$jets processes. The $\mathrm{Z}(\nu\nu)+$jets process is an irreducible background and constitutes the largest background in the search. In contrast, the background from $\mathrm{W}(\ell\nu)+$jets is suppressed by imposing a veto on events containing one or more loose muons or electrons with $p_{\mathrm{T}} > 10\,\mathrm{GeV}$, or hadronically decaying τ leptons with $p_{\mathrm{T}} > 18\,\mathrm{GeV}$. Events that contain a loose, isolated photon with $p_{\mathrm{T}} > 15\,\mathrm{GeV}$ and $|\eta| < 2.5$ are also rejected. This helps to suppress electroweak (EW) backgrounds in which a photon is radiated from the initial state. To reduce the contamination from top quark backgrounds, events are rejected if they contain a b tagged jet with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ and $|\eta| < 2.4$. These jets are identified using the DeepCSV algorithm [11, 28], adopting the “medium” 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 $E_{\mathrm{T}}^{\mathrm{miss}}$ arising from mismeasurements of the jet momenta is suppressed by requiring the minimum azimuthal angle between the $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ 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 11.

Figs. 21, 22 show the distribution of the $E_{\mathrm{T}}^{\mathrm{miss}}$, the number of jets, p_{T} and η distribution of the leading AK4 jet for events in the VBF signal category respectively for 2017 and 2018 datasets.

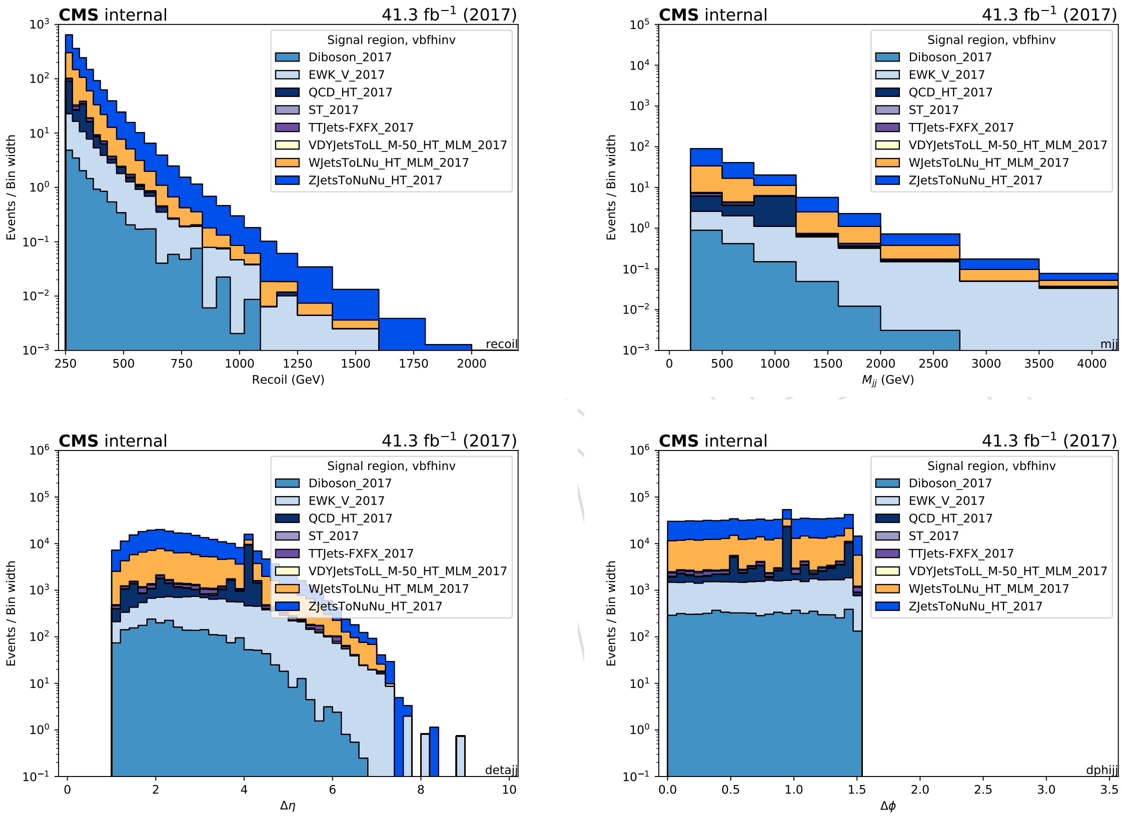

Figure 21: Recoil distribution, M_{jj} distribution, $\Delta\eta_{jj}$ and $\Delta\phi_{jj}$ distribution in the VBF signal region, using 2017 dataset.

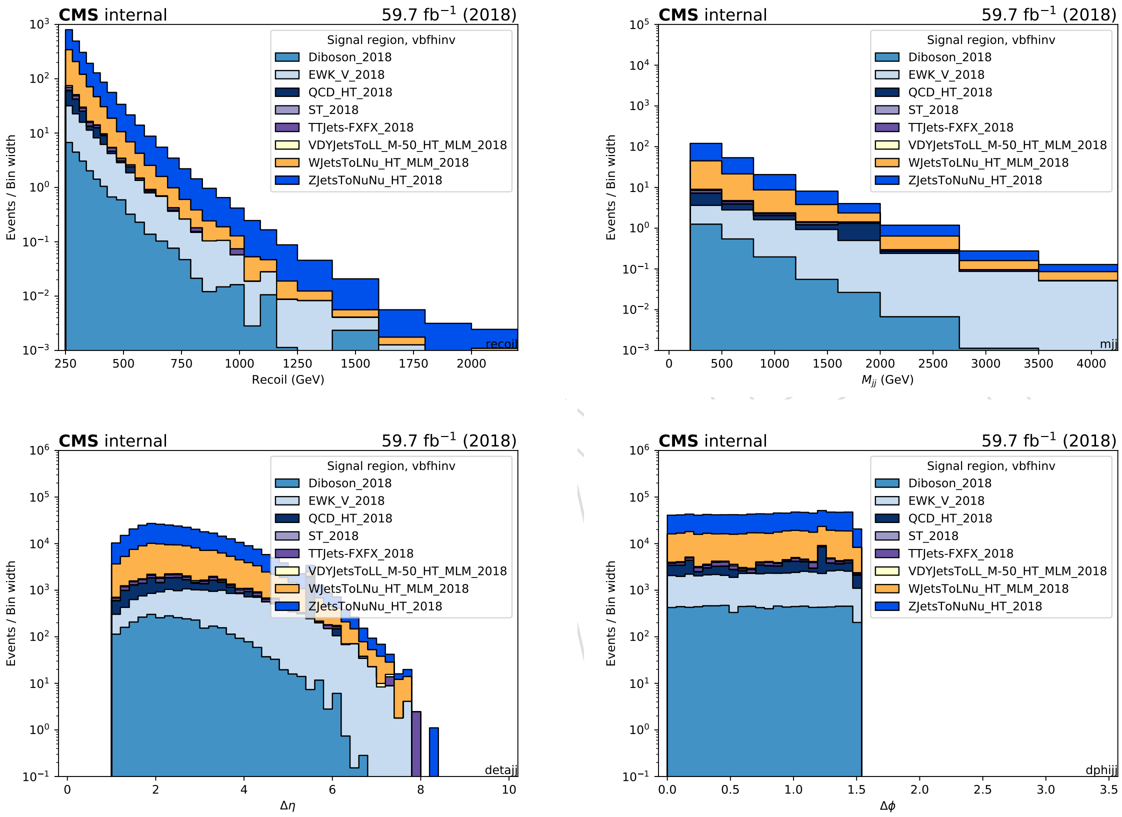

Figure 22: Recoil distribution, M_{jj} distribution, $\Delta\eta_{jj}$ and $\Delta\phi_{jj}$ distribution in the VBF signal region, using 2018 dataset.

Table 11: Summary of the common selection requirements

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

6.2 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 $p_{\mathrm{T}}^{\mathrm{miss}}$ requirement is replacement an identical requirement on the hadronic recoil, which is defined as the sum of $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ and the muon $\vec{p}_{\mathrm{T}}$, and thus corresponds to the distribution of the W p_{T}. In the single-muon control sample, exactly one tightly identified, isolated muon with $p_{\mathrm{T}} > 20$ GeV is required. No additional loose muons or electrons with $p_{\mathrm{T}} > 10$ GeV are allowed. In addition, the transverse mass of the muon-$\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ system is required to be smaller than than 160 GeV. The transverse mass (M_{T}) is computed as $M_{\mathrm{T}} = \sqrt{2E_{\mathrm{T}}^{\mathrm{miss}} p_{\mathrm{T}}^{\mu}(1-\cos\Delta\phi)}$, where p_{T}^{μ} is the p_{T} of the muon, and $\Delta\phi$ is the angle between $\vec{p}_{\mathrm{T}}^{\,\mu}$ and $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$.

Figs. 23 and 25 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. 24 and 26 show the distributions of the leading muon p_{T} and η, as well as the muon-$p_{\mathrm{T}}^{\mathrm{miss}}$ transverse mass, again for 2017 and 2018, respectively.

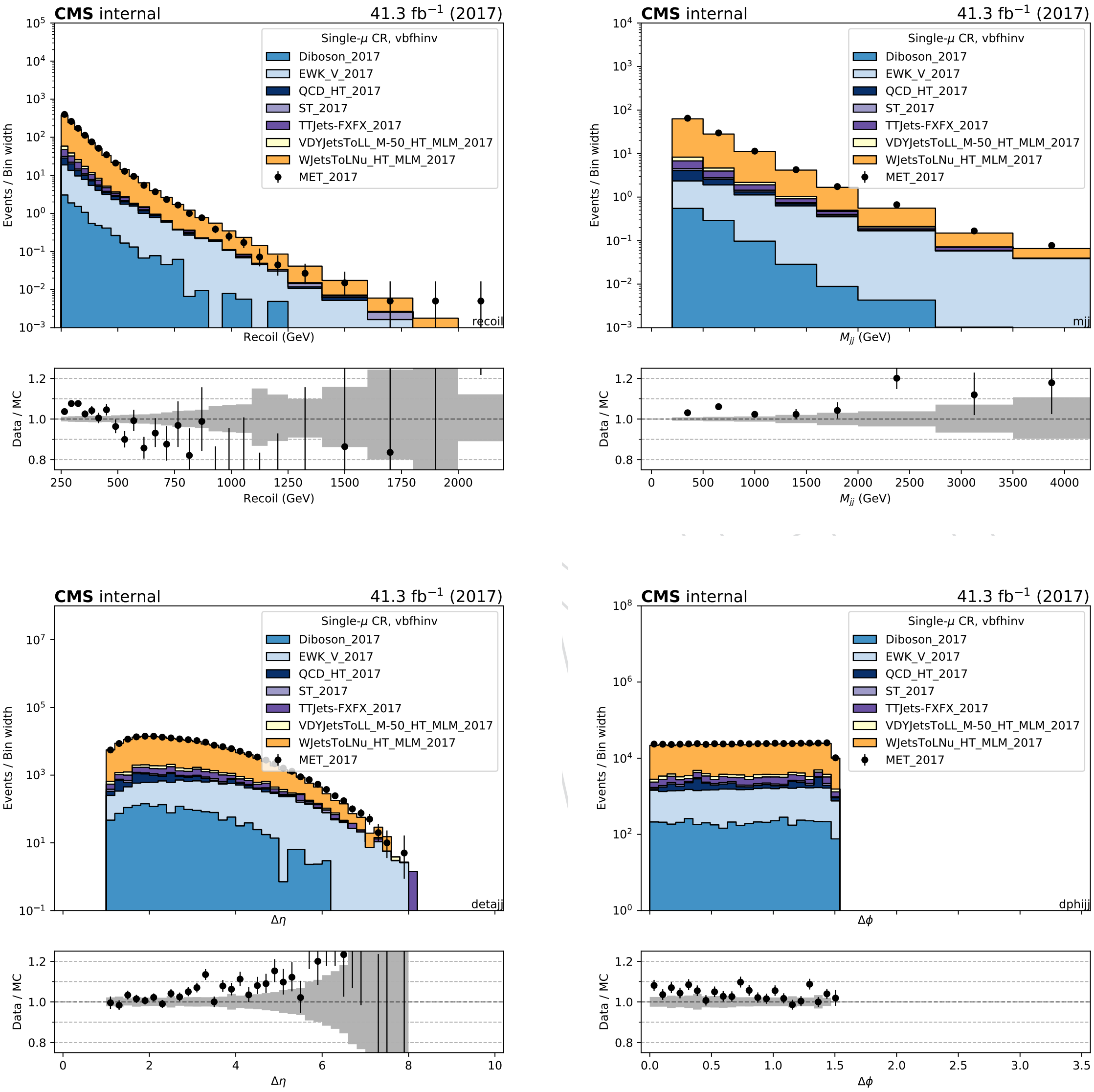

Figure 23: Comparison between 2017 data and Monte Carlo simulation in the single muon control sample for the recoil distribution, the M_{jj} distribution, $\Delta\eta_{jj}$ distribution and $\Delta\phi_{jj}$ distribution for the two leading AK4 jets with the VBF selection.

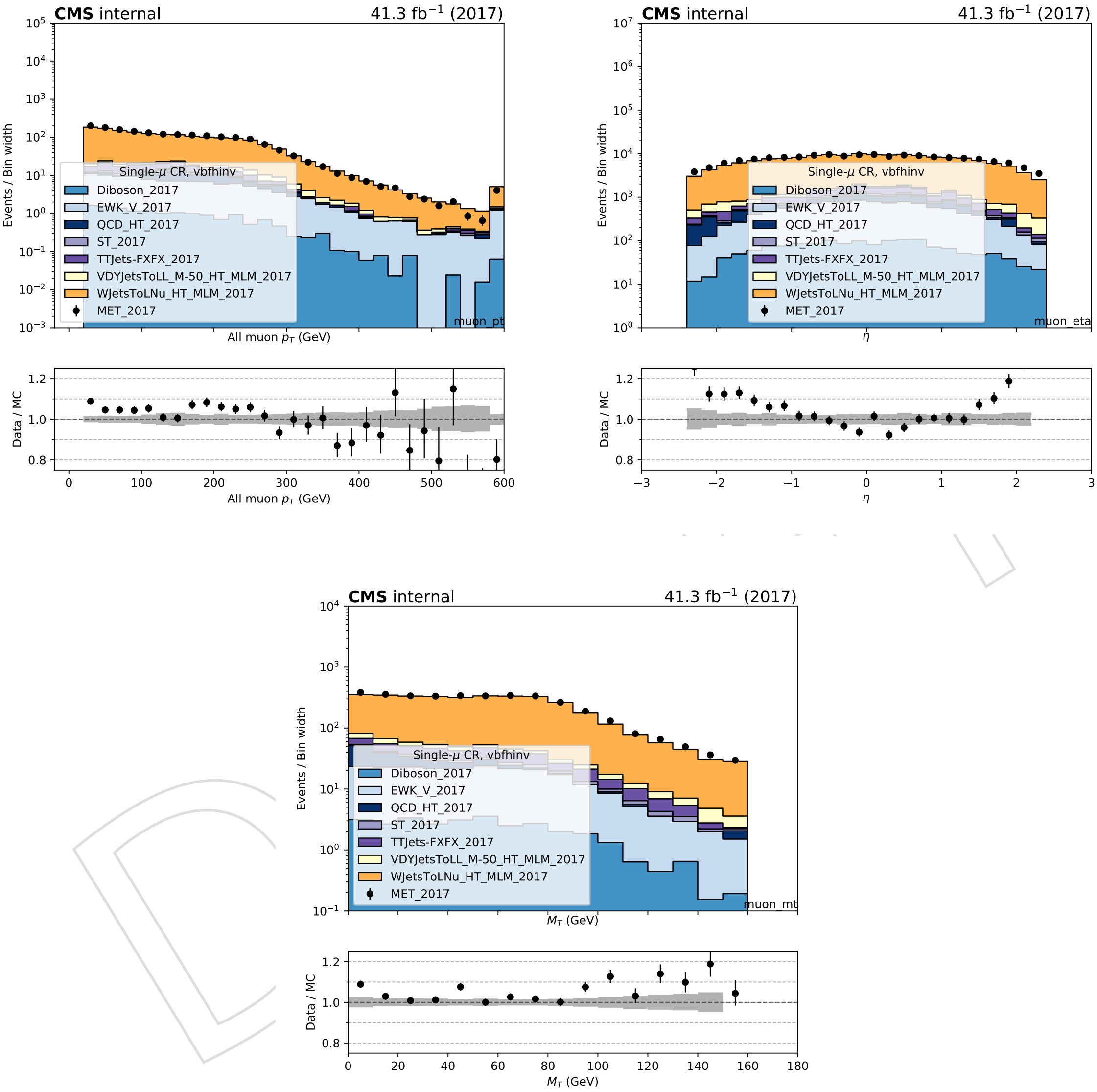

Figure 24: Comparison between 2017 data and Monte Carlo simulation in the single muon control sample for the p_{T} and η of the leading muon and the transverse mass distribution with the VBF selection.

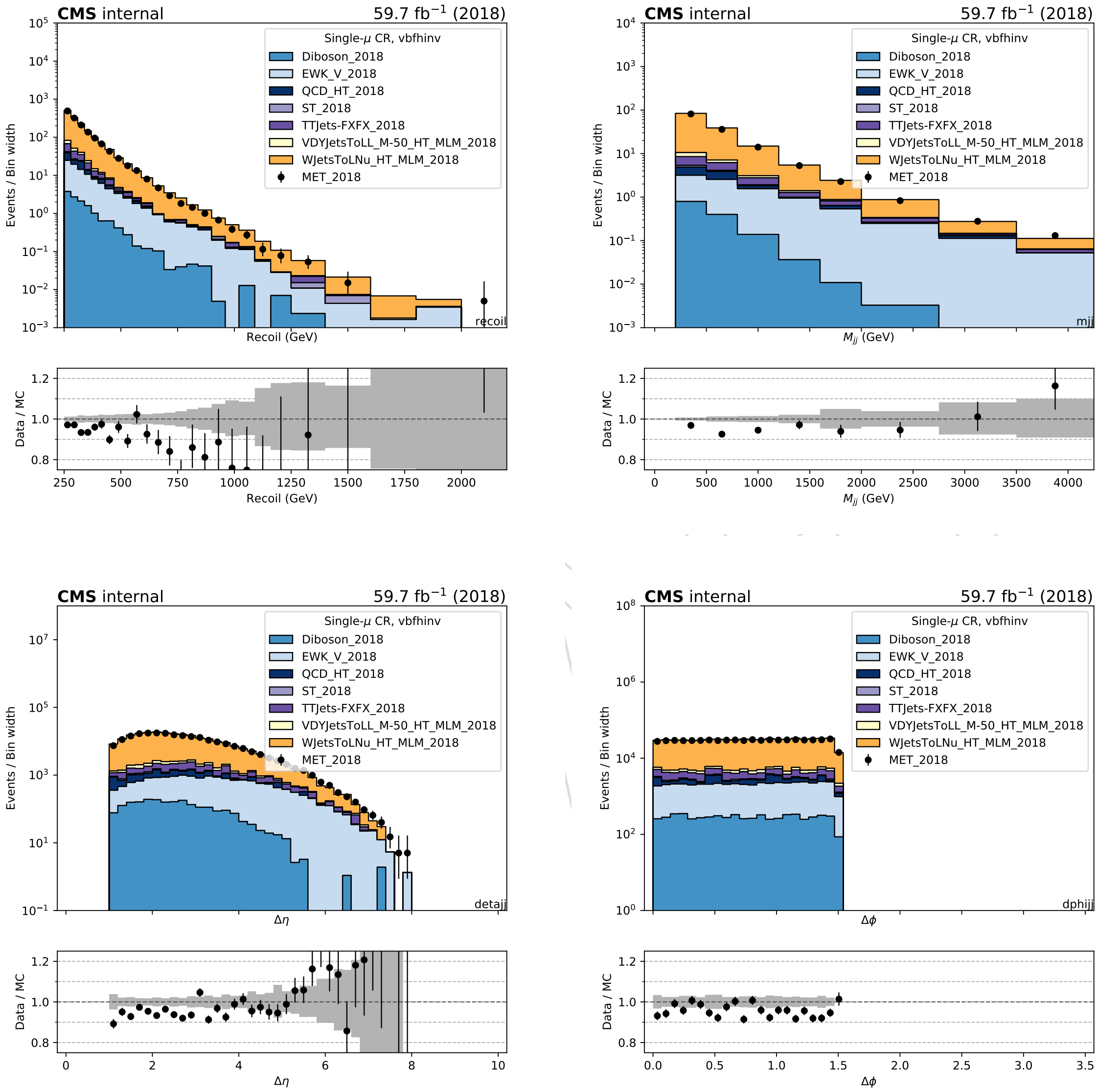

Figure 25: Comparison between 2018 data and Monte Carlo simulation in the single muon control sample for the recoil distribution, the M_{jj} distribution, $\Delta\eta_{jj}$ distribution and $\Delta\phi_{jj}$ distribution for the two leading AK4 jets with the VBF selection.

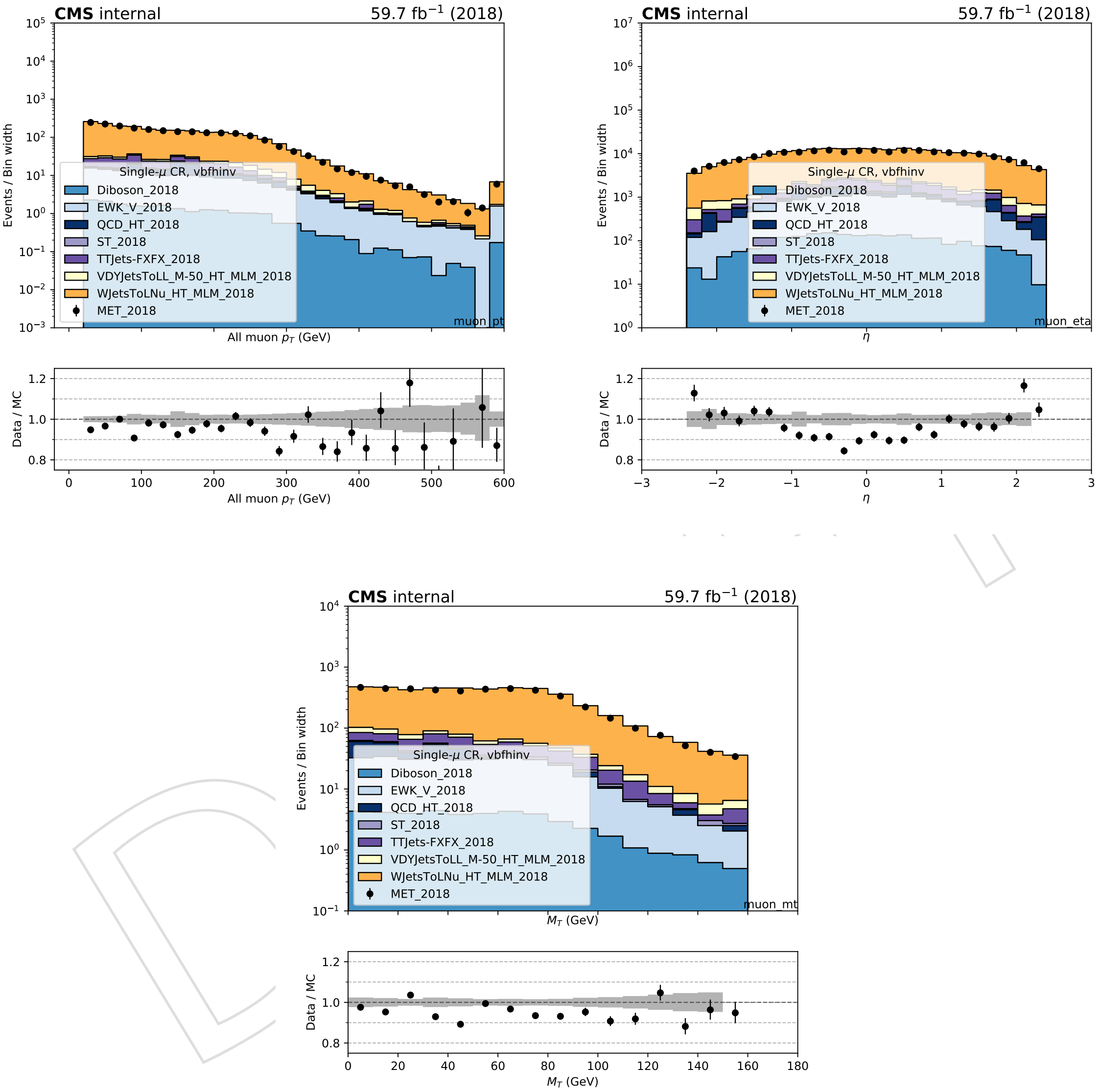

Figure 26: Comparison between 2018 data and Monte Carlo simulation in the single muon control sample for the p_{T} and η of the leading muon and the transverse mass distribution with the VBF selection.

6.3 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 $p_{\mathrm{T}}^{\mathrm{miss}}$ requirement is replaced with an identical requirement on the hadronic recoil, which is defined as the sum of $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ and the electron $\vec{p}_{\mathrm{T}}$, 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 $p_{\mathrm{T}} > 40$ GeV. In addition, the contamination from QCD multijet events in this control sample is suppressed by requiring $E_{\mathrm{T}}^{\mathrm{miss}} > 50$ GeV and $M_{\mathrm{T}} < 160\,\mathrm{GeV}$.

Figs. 27 and 29 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. 28 and 30 show the distributions of the leading electron p_{T} and η, as well as the electron-$p_{\mathrm{T}}^{\mathrm{miss}}$ transverse mass, again for 2017 and 2018, respectively.

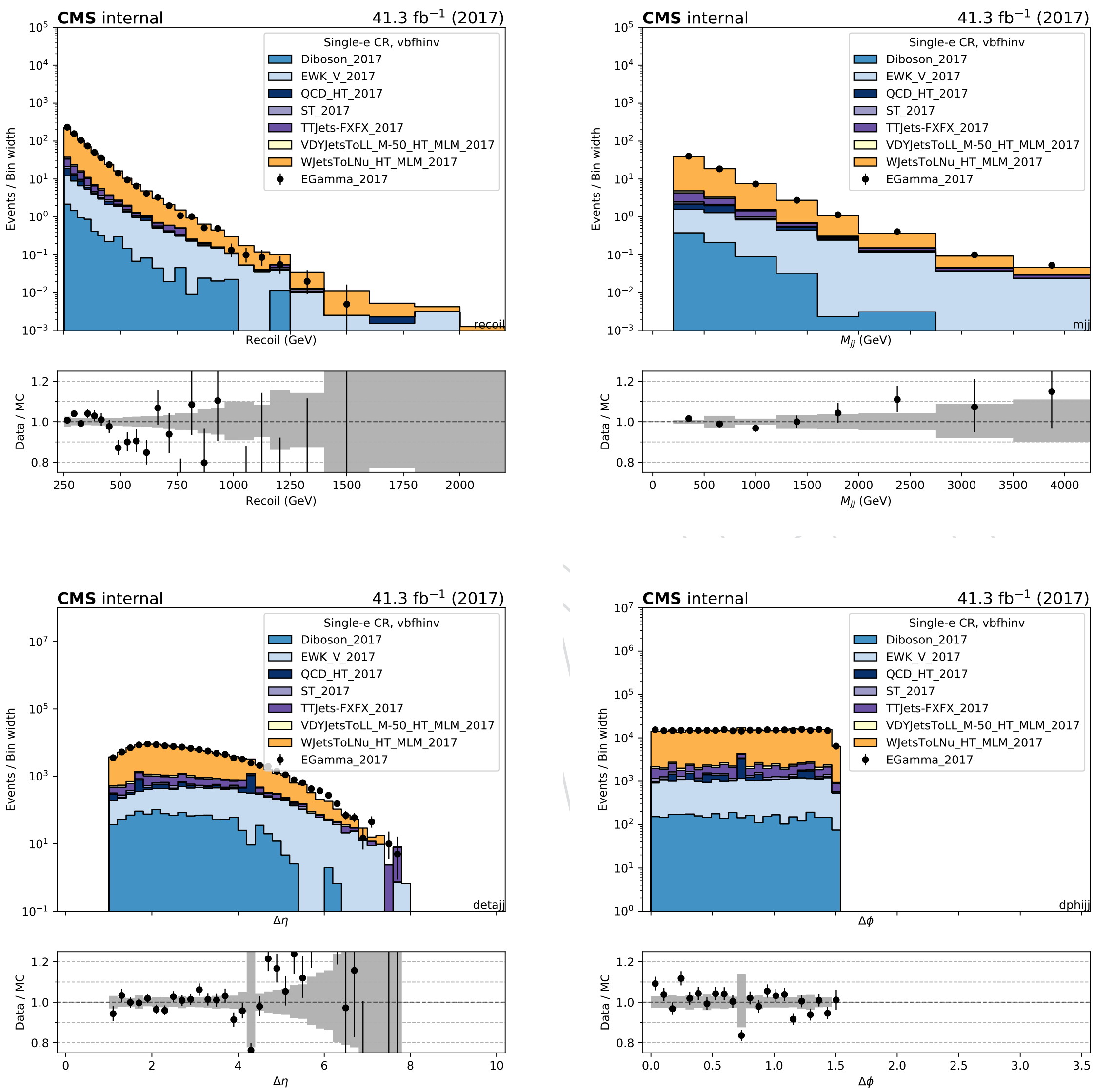

Figure 27: Comparison between 2017 data and Monte Carlo simulation in the single electron control sample for the recoil distribution, the M_{jj} distribution, $\Delta\eta_{jj}$ distribution and $\Delta\phi_{jj}$ distribution for the two leading AK4 jets with the VBF selection.

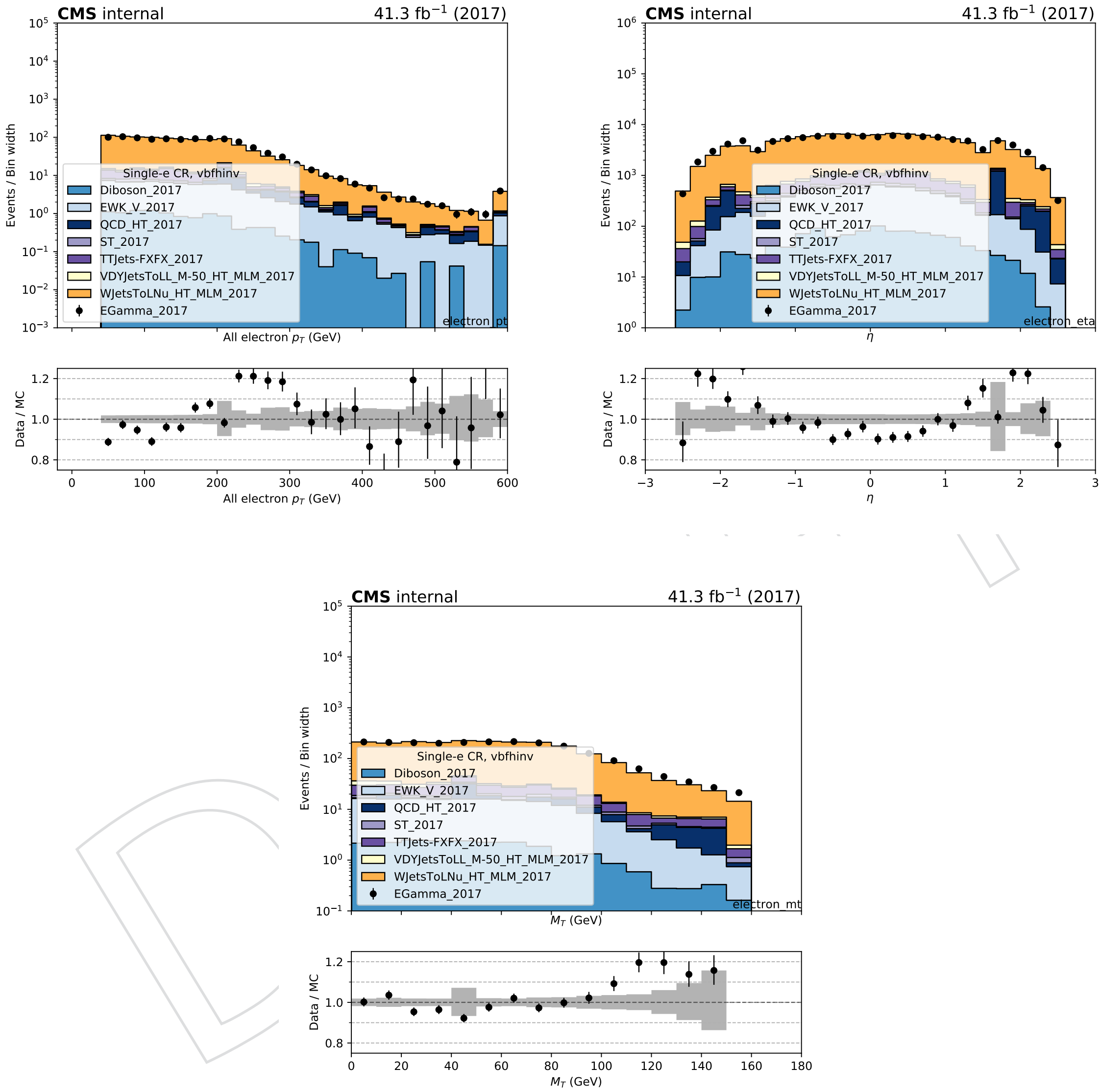

Figure 28: Comparison between 2017 data and Monte Carlo simulation in the single electron control sample for the p_{T} and η of the leading electron and the transverse mass distribution with the VBF selection.

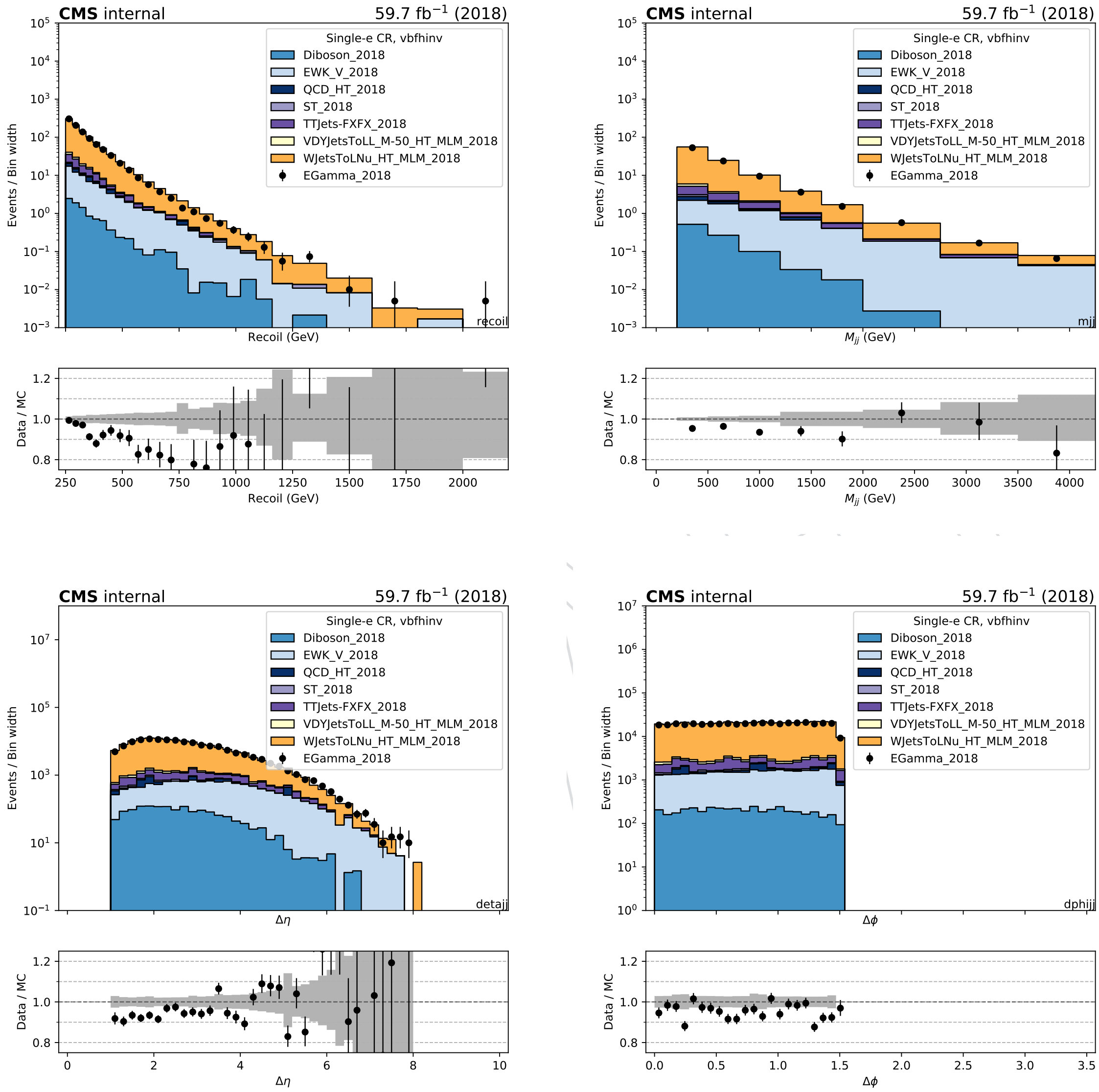

Figure 29: Comparison between 2018 data and Monte Carlo simulation in the single electron control sample for the recoil distribution, the M_{jj} distribution, $\Delta\eta_{jj}$ distribution and $\Delta\phi_{jj}$ distribution for the two leading AK4 jets with the VBF selection.

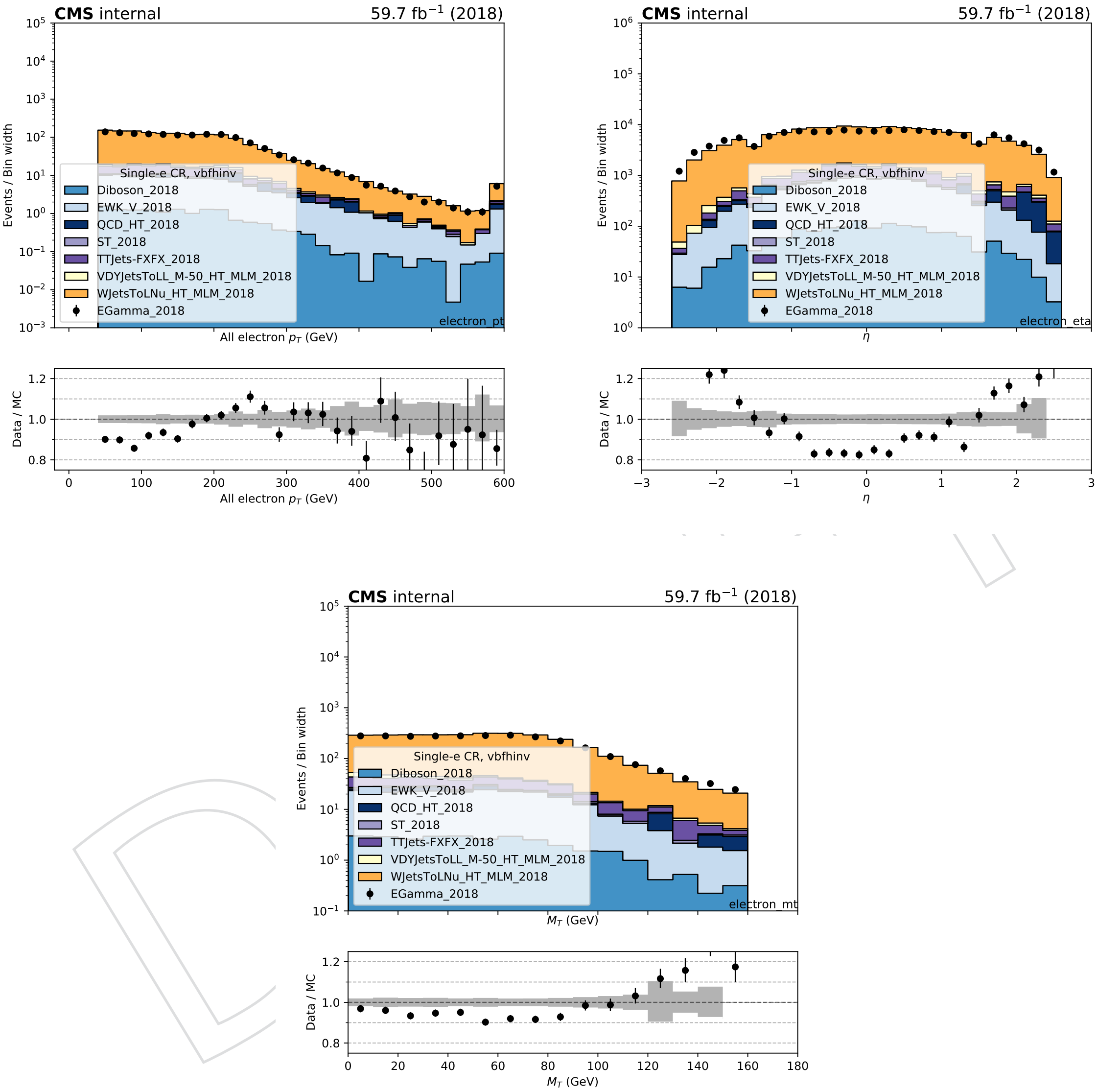

Figure 30: Comparison between 2018 data and Monte Carlo simulation in the single electron control sample for the p_{T} and η of the leading electron and the dilepton mass distribution with the VBF selection.

6.4 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 $p_{\mathrm{T}} > 10$ GeV. The SR $p_{\mathrm{T}}^{\mathrm{miss}}$ requirement is replacement an identical requirement on the hadronic recoil, which is defined as the sum of $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ and the muon $\vec{p}_{\mathrm{T}}$, and thus corresponds to the distribution of the Z p_{T} smeared with the $p_{\mathrm{T}}^{\mathrm{miss}}$ resolution.

Figs. 31 and 33 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. 32 and 34 show the distributions of the leading muon p_{T} and η, as well as the dimuon mass and p_{T}, again for 2017 and 2018, respectively.

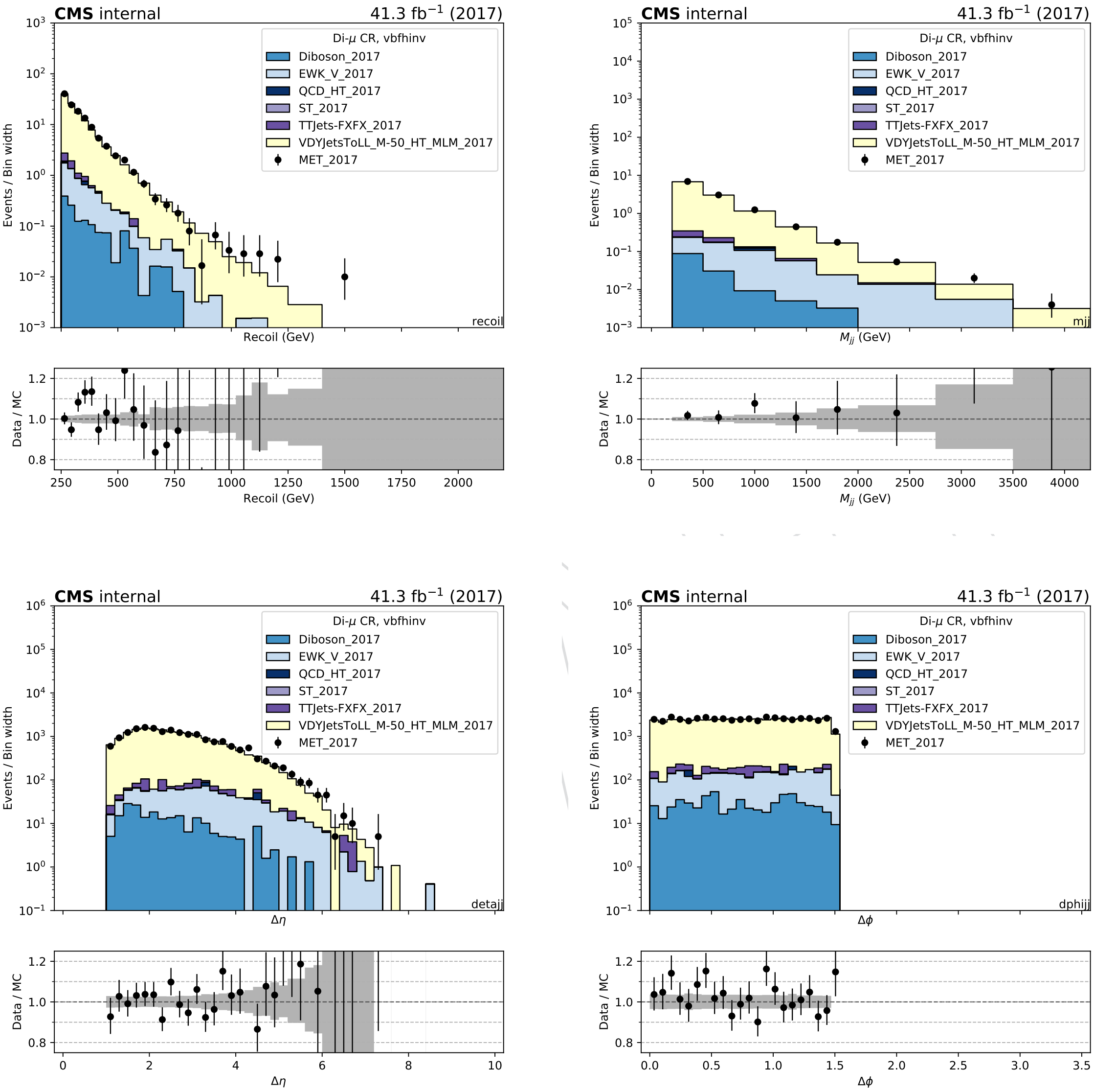

Figure 31: Comparison between 2017 data and Monte Carlo simulation in the double muon control sample for the recoil distribution, the M_{jj} distribution, $\Delta\eta_{jj}$ distribution and $\Delta\phi_{jj}$ distribution for the two leading AK4 jets with the VBF selection.

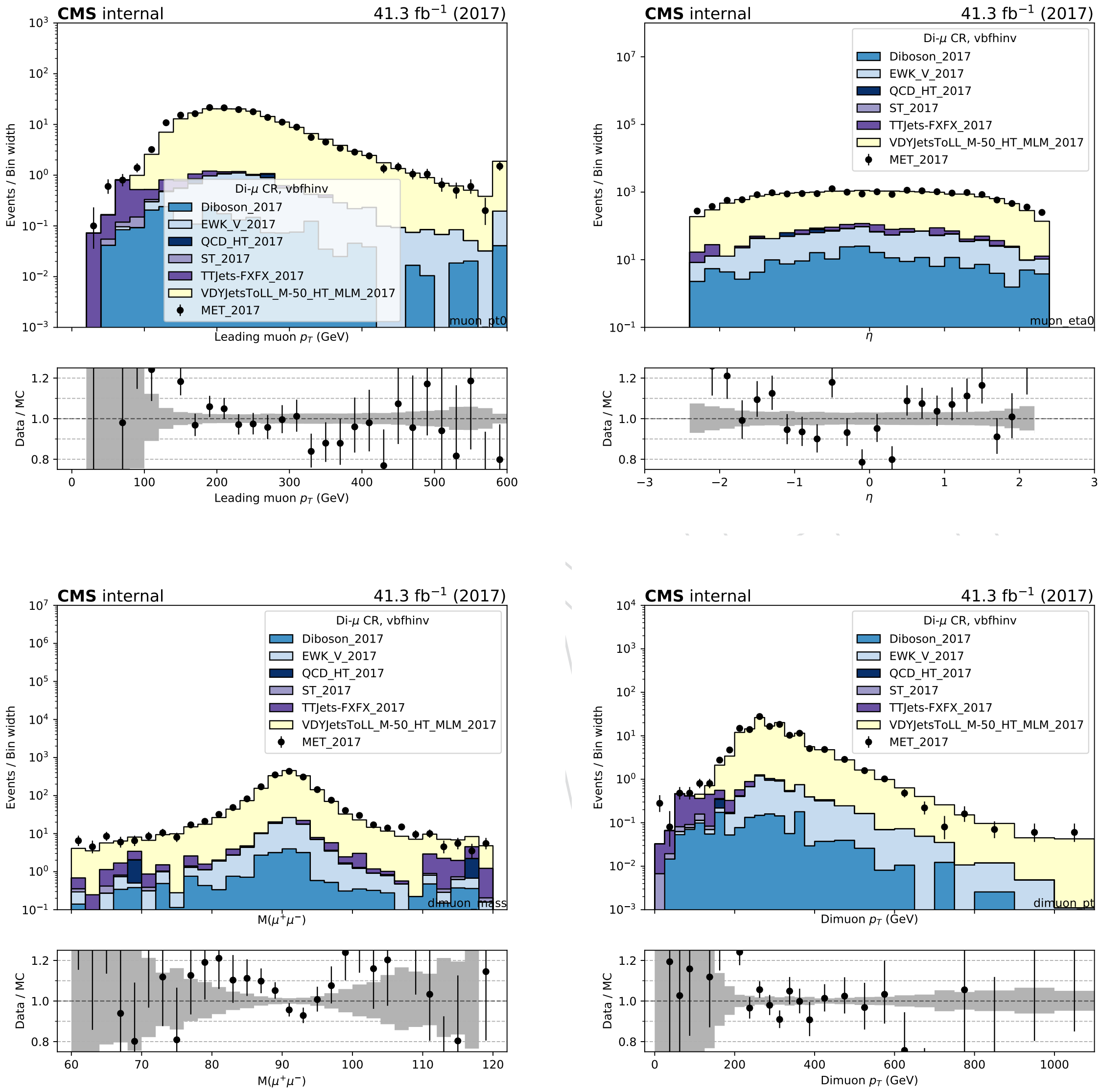

Figure 32: Comparison between 2017 data and Monte Carlo simulation in the double muon control sample for the p_{T} and η of the leading muon and the transverse mass and p_{T} of the dimuon candidate with the VBF selection.

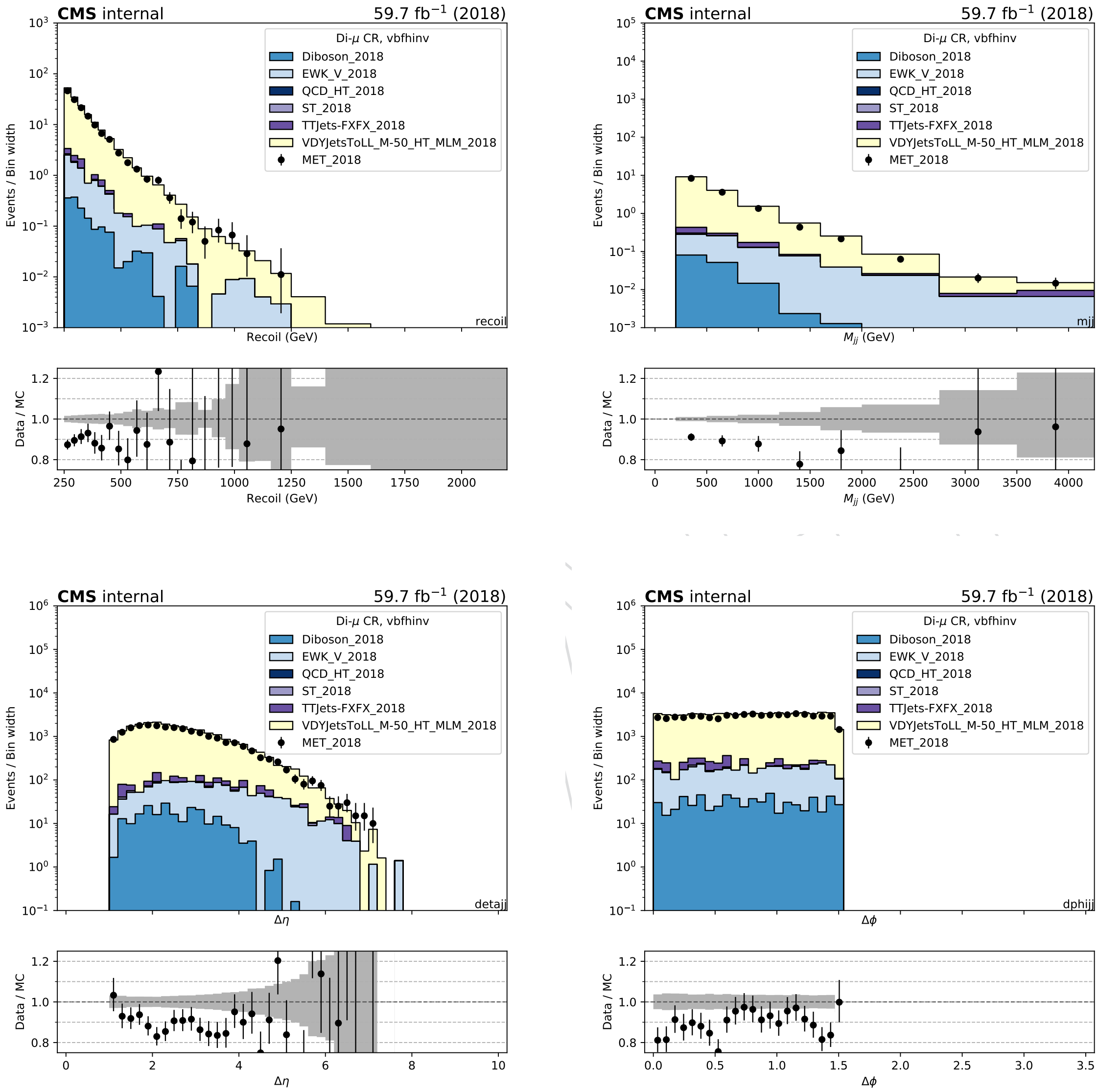

Figure 33: Comparison between 2018 data and Monte Carlo simulation in the double muon control sample for the recoil distribution, the M_{jj} distribution, $\Delta\eta_{jj}$ distribution and $\Delta\phi_{jj}$ distribution for the two leading AK4 jets with the VBF selection.

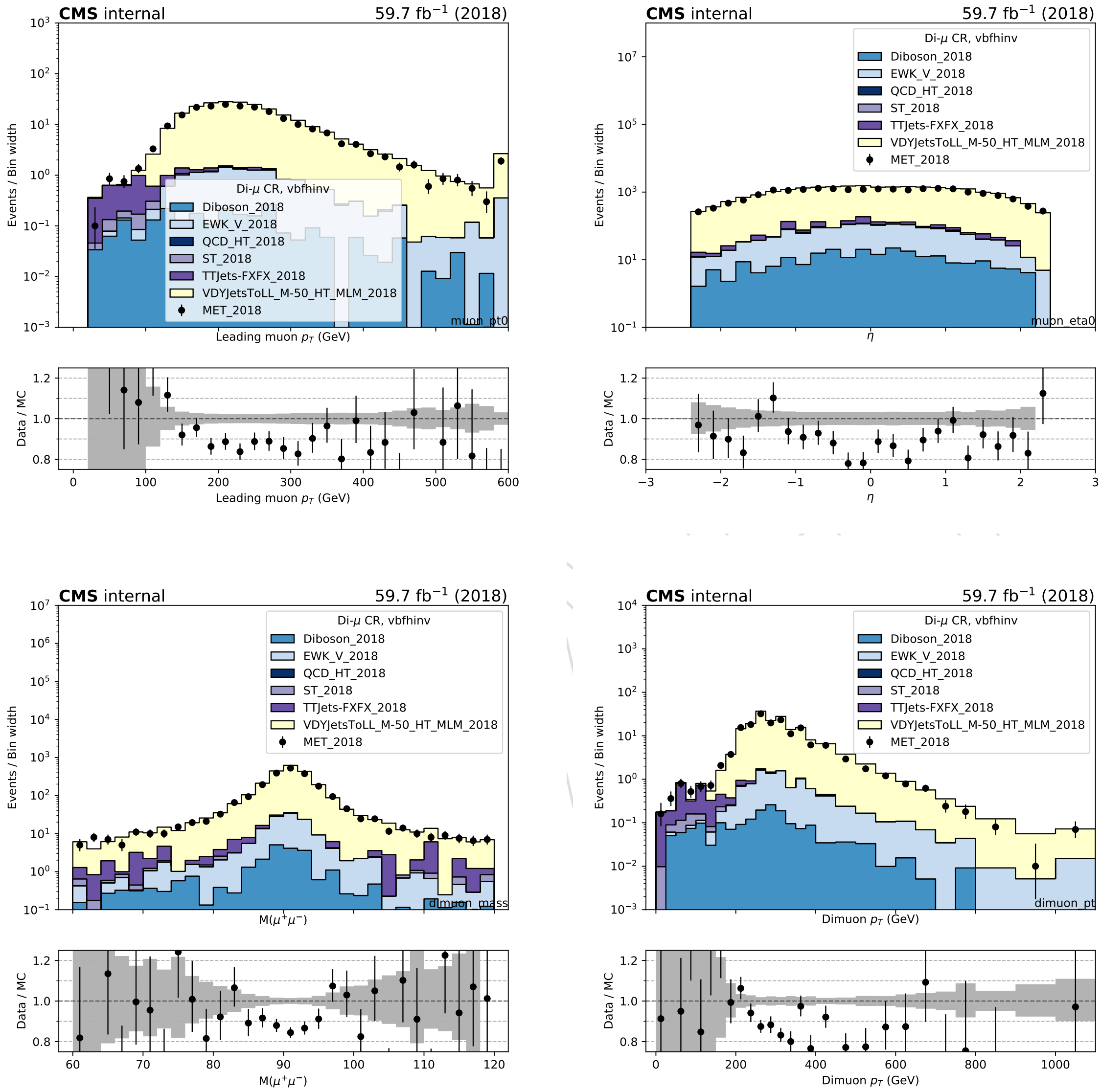

Figure 34: Comparison between 2018 data and monte carlo simulation in the double muon control sample for the p_{T} and η of the leading muon and the transverse mass and p_{T} of the dimuon candidate with the VBF selection.

6.5 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 $p_{\mathrm{T}}^{\mathrm{miss}}$ requirement is replacement an identical requirement on the hadronic recoil, which is defined as the sum of $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ and the muon $\vec{p}_{\mathrm{T}}$, and thus corresponds to the distribution of the Z p_{T} smeared with the $p_{\mathrm{T}}^{\mathrm{miss}}$ 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. 35 and 37 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. 36 and 38 show the distributions of the leading electron p_{T} and η, as well as the dielectron mass and p_{T}, again for 2017 and 2018, respectively.

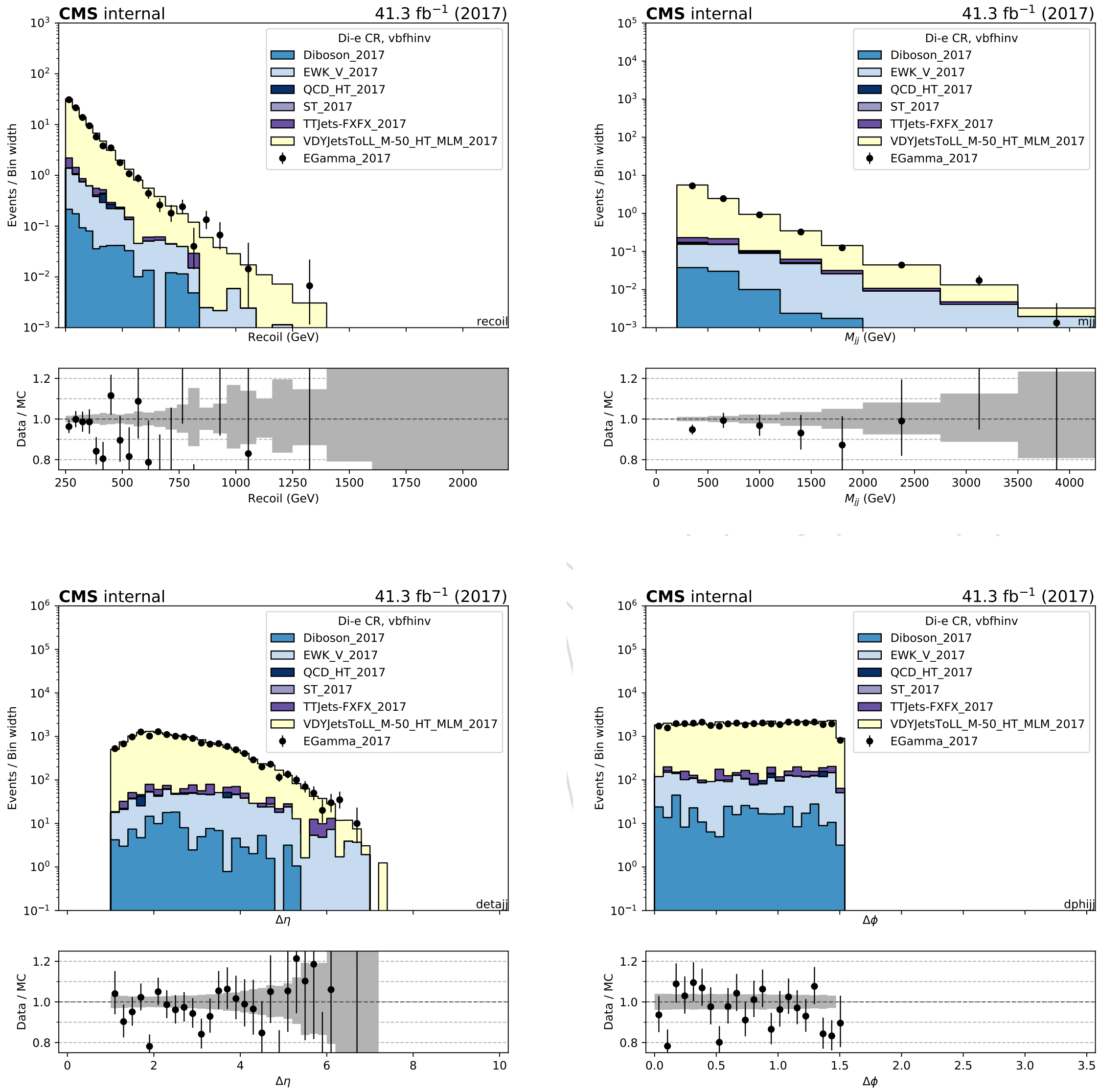

Figure 35: Comparison between 2017 data and Monte Carlo simulation in the double electron control sample for the recoil distribution, the M_{jj} distribution, $\Delta\eta_{jj}$ distribution and $\Delta\phi_{jj}$ distribution for the two leading AK4 jets with the VBF selection.

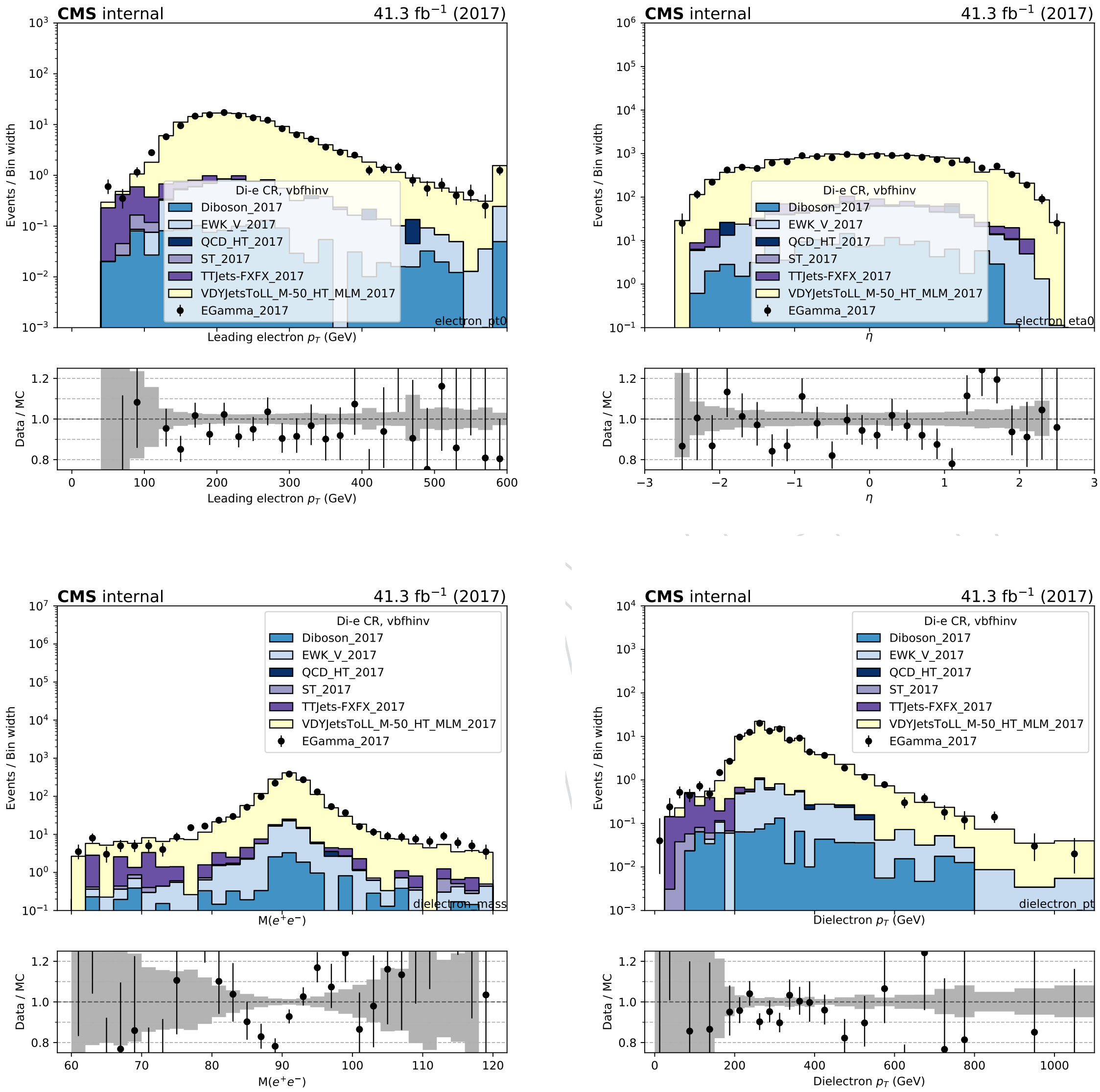

Figure 36: Comparison between 2017 data and Monte Carlo simulation in the double electron control sample for the p_{T} and η of the leading electron and the transverse mass and p_{T} of the dielectron candidate with the VBF selection.

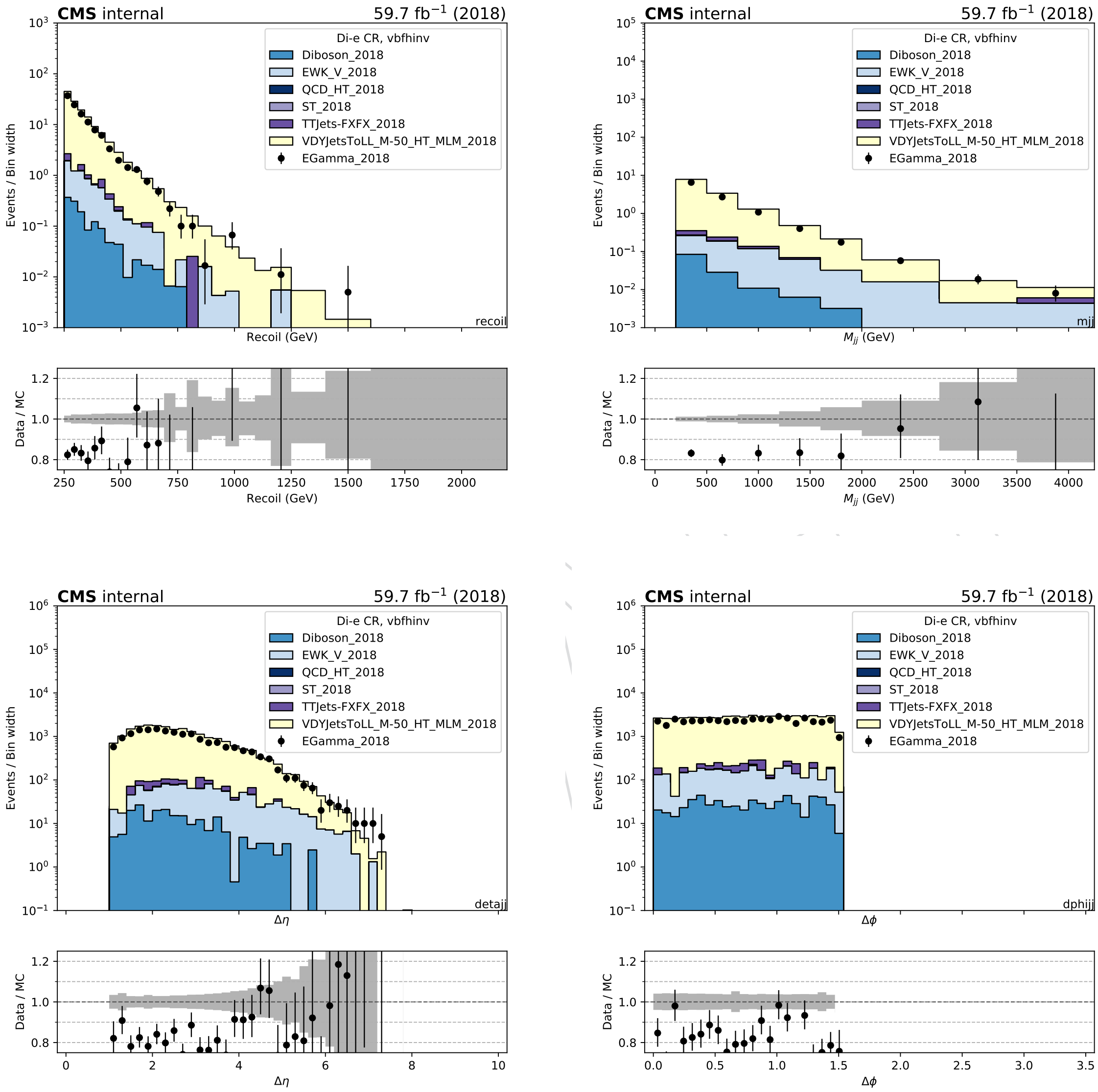

Figure 37: Comparison between 2018 data and Monte Carlo simulation in the double electron control sample for the recoil distribution, the M_{jj} distribution, $\Delta\eta_{jj}$ distribution and $\Delta\phi_{jj}$ distribution for the two leading AK4 jets with the VBF selection.

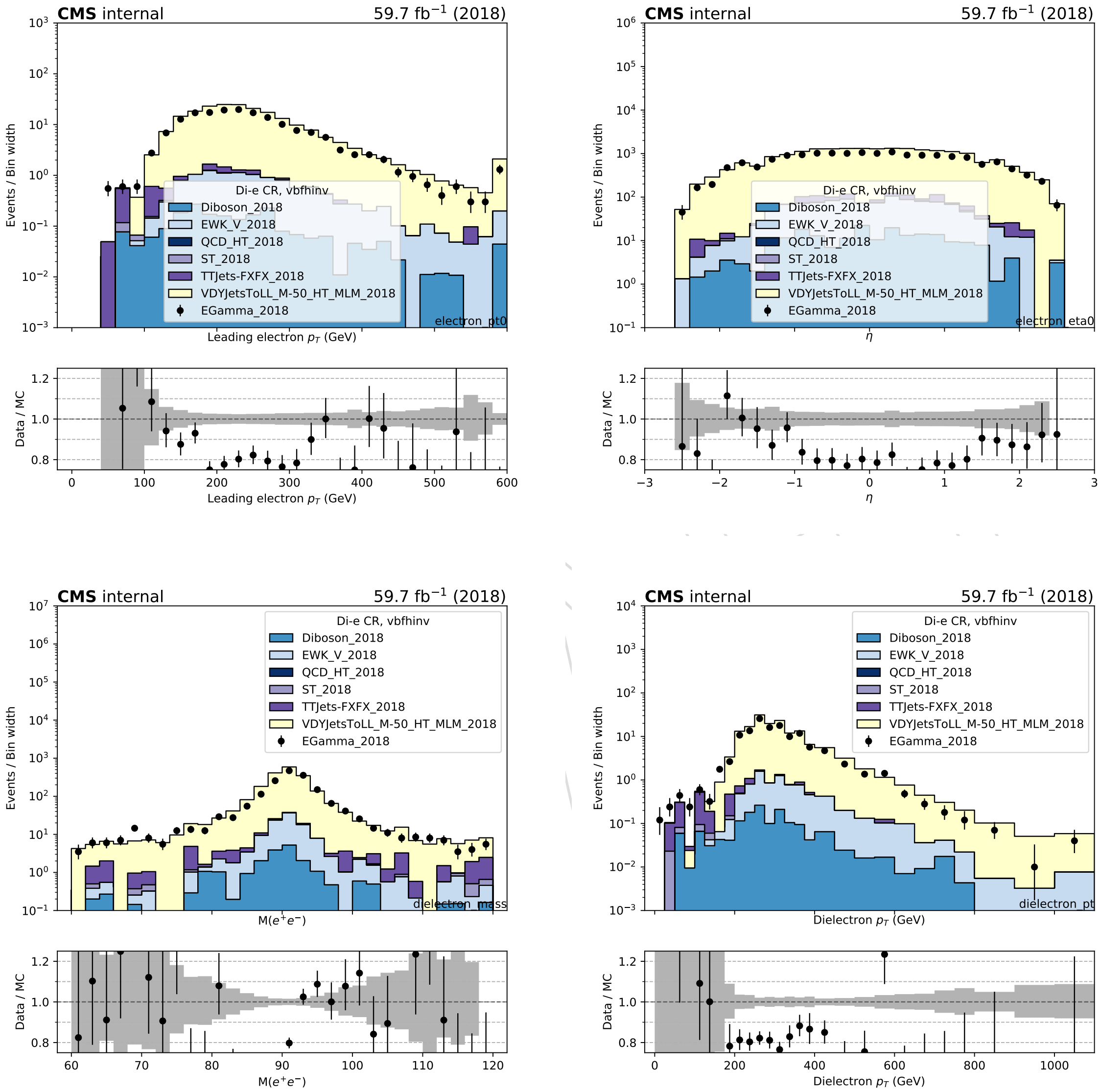

Figure 38: Comparison between 2018 data and Monte Carlo simulation in the double electron control sample for the p_{T} and η of the leading electron and the transverse mass and p_{T} of the dielectron candidate with the VBF selection.

6.6 Photon control region

The $\gamma+$jetscontrol 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. 39 and 39 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. 40 and 42 show the distributions of the photon p_T and η.

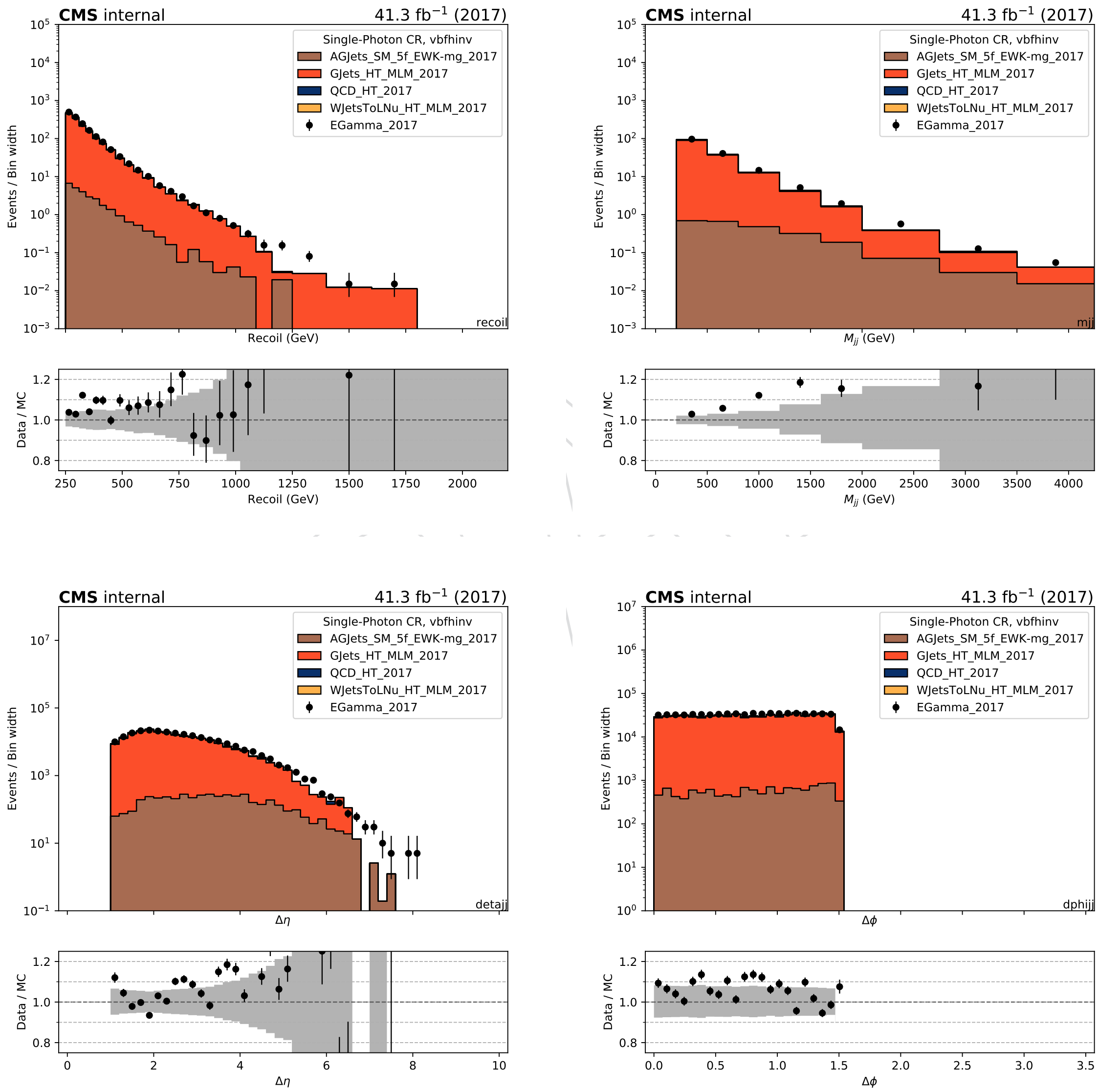

Figure 39: Comparison between 2017 data and Monte Carlo simulation in the photon control sample for the recoil distribution, the M_{jj} distribution, $\Delta\eta_{jj}$ distribution and $\Delta\phi_{jj}$ distribution for the two leading AK4 jets with the VBF selection.

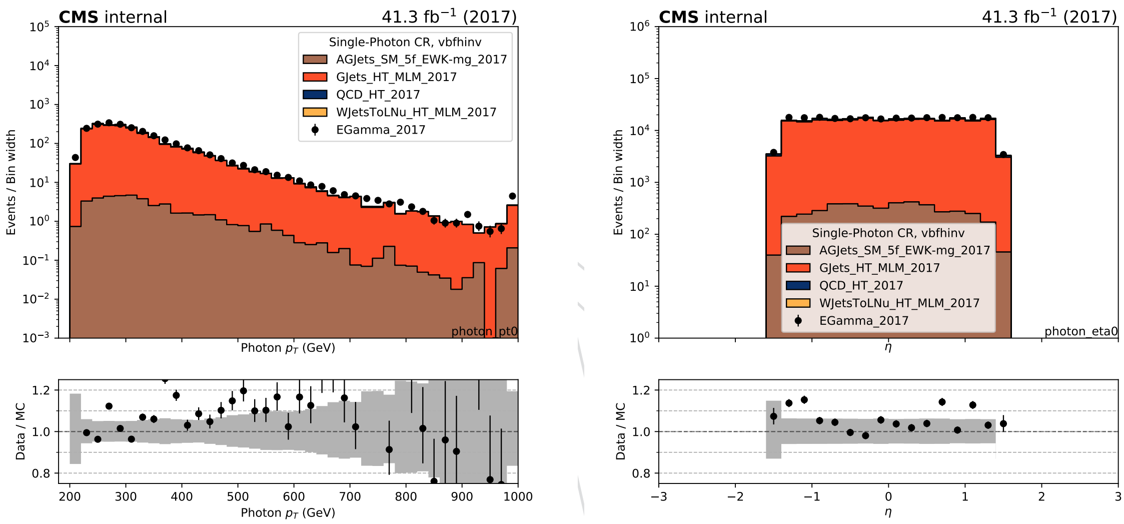

Figure 40: Comparison between 2017 data and Monte Carlo simulation in the photon control sample for the p_{T} and η of the leading photon with the VBF selection.

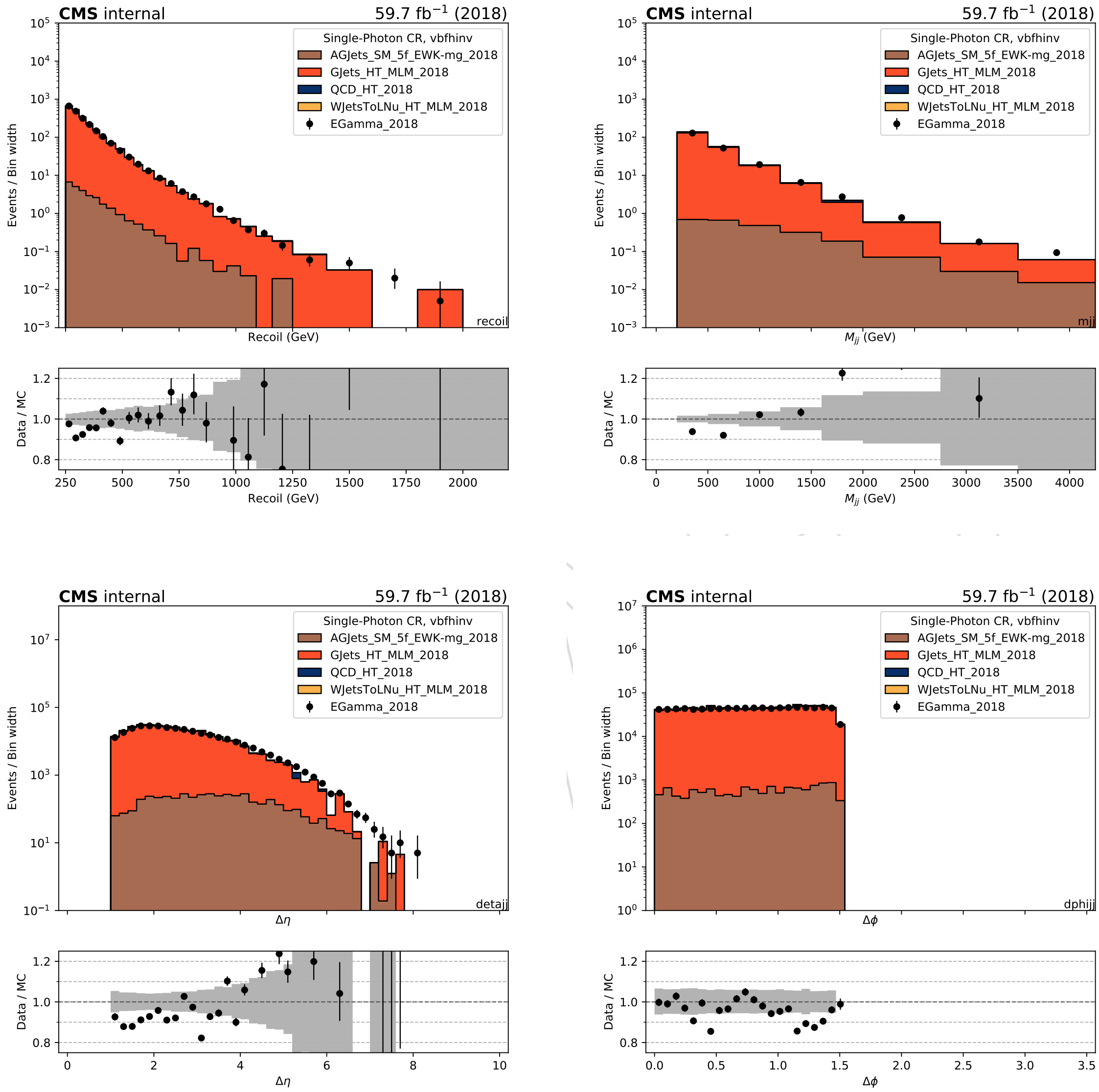

Figure 41: Comparison between 2018 data and Monte Carlo simulation in the photon control sample for the recoil distribution, the M_{jj} distribution, $\Delta\eta_{jj}$ distribution and $\Delta\phi_{jj}$ distribution for the two leading AK4 jets with the VBF selection.

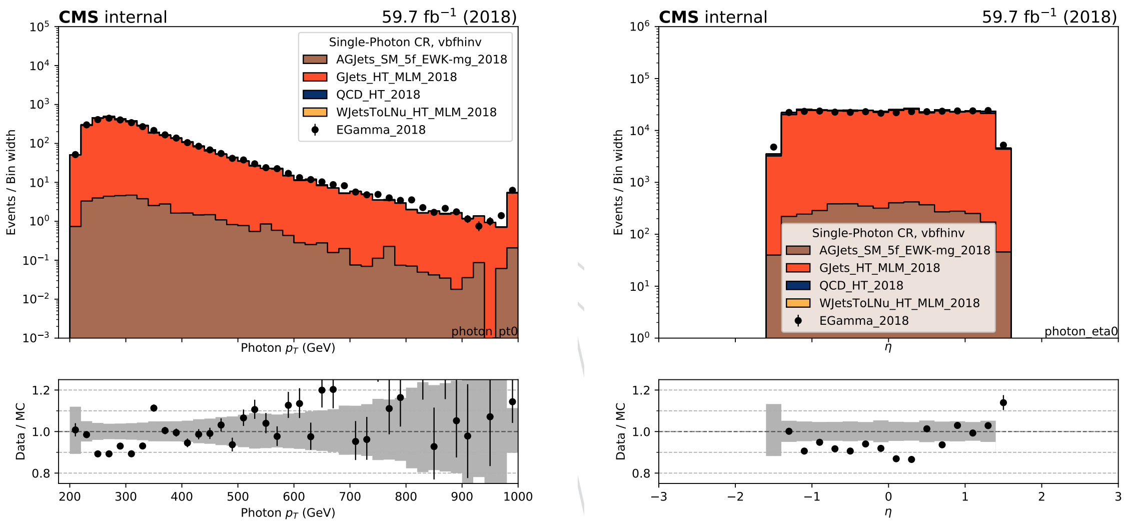

Figure 42: Comparison between 2018 data and Monte Carlo simulation in the photon control sample for the p_{T} and η of the leading photon with the VBF selection.

7 Background estimation

The largest background contributions from $Z(\nu\nu) + $ jets and $W(\ell\nu) + $ jets processes are estimated using data from four mutually exclusive control samples selected from dimuon, dielectron, single-muon, and single-electron states, as explained below.

The remaining backgrounds that contribute to the total event yield in the signal region are much smaller than those from $Z(\nu\nu) + $ jets and $W(\ell\nu) + $ jets processes. These backgrounds include QCD multijet events which are measured from data using a $\Delta\phi$ extrapolation method [29, 30], and top-quark and diboson processes, which are obtained directly from simulation and are explained in the previous sections.

7.1 Signal extraction strategy

A binned likelihood fit to the data (constructed as a product of poisson probabilities) as presented here:

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\mu}^{Z\rightarrow\nu\nu}, \mu, \boldsymbol{\theta}) = &\prod_i \mathrm{Pois}\left(d_i | B_i(\boldsymbol{\theta}) + (1 + f_i(\boldsymbol{\theta})_{\mathrm{QCD}})\boldsymbol{\mu}_i^{Z\rightarrow\nu\nu} + R_i^{Z\rightarrow\circ\circ}(1 + f_i(\boldsymbol{\theta})_{\mathrm{EW}})\boldsymbol{\mu}_i^{Z\rightarrow\nu\nu} + \mu S_i(\boldsymbol{\theta})\right) \times \\
&\prod_i \mathrm{Pois}\left(d_i^Z | B_i^Z(\boldsymbol{\theta}) + \frac{\boldsymbol{\mu}_i^{Z\rightarrow\nu\nu}}{R_i^Z(\boldsymbol{\theta})_{\mathrm{QCD}}} + \frac{\boldsymbol{\mu}_i^{Z\rightarrow\nu\nu}}{R_i^Z(\boldsymbol{\theta})_{\mathrm{EW}}}\right) \\
&\prod_i \mathrm{Pois}\left(d_i^W | B_i^W(\boldsymbol{\theta}) + \frac{f_i(\boldsymbol{\theta})_{\mathrm{QCD}}\,\boldsymbol{\mu}_i^{Z\rightarrow\nu\nu}}{R_i^W(\boldsymbol{\theta})_{\mathrm{QCD}}} + \frac{f_i(\boldsymbol{\theta})_{\mathrm{EW}}\,\boldsymbol{\mu}_i^{Z\rightarrow\nu\nu}}{R_i^W(\boldsymbol{\theta})_{\mathrm{EW}}}\right) \times
\end{aligned} \tag{4}$$

The fit is performed simultaneously in the four different control samples and in the signal region, for events selected in the vbf category, to estimate the $Z(\nu\nu) + $ jets and $W(\ell\nu) + $ jets rate in each m_{jj} bin. In this likelihood, the expected number of $Z(\nu\nu) + $ jets events in each bin of m_{jj} are the free parameters of the fit. The $Z(\nu\nu) + $ jets and $W(\ell\nu) + $ jets rates are estimated seperately for the QCD and EW components seperately in each m_{jj} bin. However the fit is constrained using the $R_i^{Z\rightarrow\circ\circ}$ which demonstrates the ratio between the QCD and EW components of the $Z(\nu\nu) + $ jets background. This ratio does not have any additional uncertainty. The systematic uncertainties ($\boldsymbol{\theta}$) enter the likelihood as additive perturbations to the transfer factors $R_i^{/Z/W}$, and are modeled as Gaussians. The parameter $\boldsymbol{\mu}_i^{Z\rightarrow\nu\nu}$ represents the yield of the $Z(\nu\nu) + $ jets background in the i dijet mass bin into the signal region, and is left freely floating in the fit. The function $f_i(\boldsymbol{\theta})$ is the transfer factor between the $Z(\nu\nu) + $ jets and $W(\ell\nu) + $ 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 μ being the signal strength parameter also left floating in the fit.

7.2 Transfer factors

Transfer factors, derived from simulation, are used to link the yields of the $Z(\ell\ell) + $ jets and $W(\ell\nu) + $ jets processes in the control regions with the $Z(\nu\nu) + $ jets and $W(\ell\nu) + $ jets background estimates in the signal region. These transfer factors are defined as the ratio of expected yields of the target process in the signal region and the process being measured in the control sample. As an example:

$$R_i^Z = \frac{N_{i,MC}^{Z\rightarrow\mu^+\mu^-}}{N_{i,MC}^{Z\rightarrow\nu\nu}} \tag{5}$$

where N_i is the number of events in bin i of the dijet mass distribution, R_i is the transfer factor between the dimuon control region and $Z \rightarrow \nu\nu$ background. Other transfer factors are constructed in similar manner.

To estimate the $W(\ell\nu) + \text{jets}$ background in the signal region, the transfer factors between the $W(\mu\nu) + \text{jets}$ and $W(e\nu) + \text{jets}$ event yields in the single-lepton control samples and the $W(\ell\nu) + \text{jets}$ background estimates in the signal region are constructed. These transfer factors take into account the impact of lepton acceptances and efficiencies, lepton veto efficiencies, and the difference in the trigger efficiencies in the case of the single-electron control sample. These transfer factors are shown in Figure 43. The dotted red (blue) line shows the ratio of the processes where the V-boson is produced via EW (QCD) production where as the solid lines show the ratio of the processes where the V-boson is produced both via EW and QCD production.

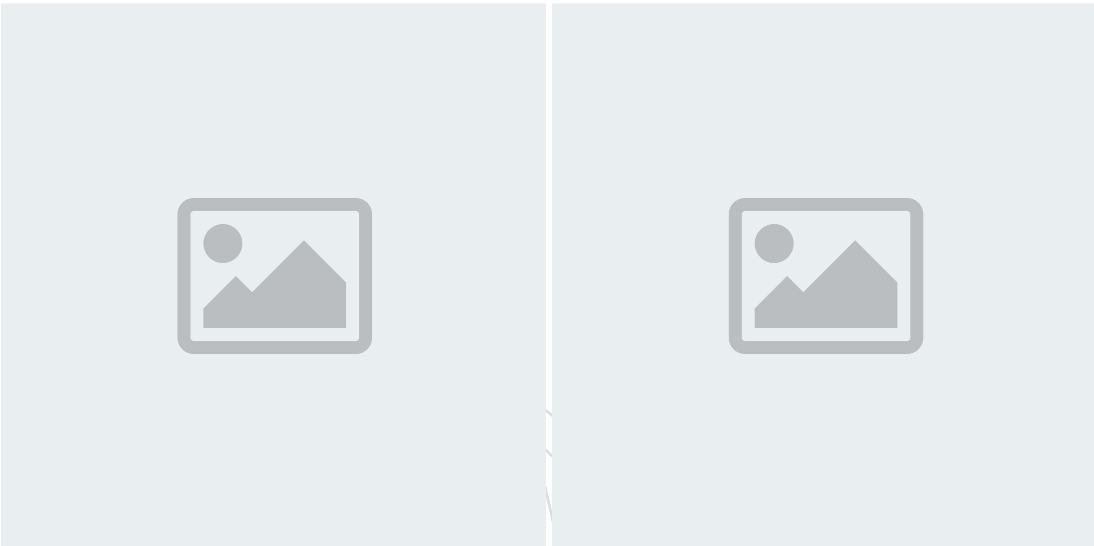

Figure 43: Transfer factors for the $W \rightarrow \ell\nu$background as a function of the dijet mass using the single muon and single electron control regions for the vbf final state. The grey band shows the statistical and systematic uncertainties on the ratios.

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 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 $Z \rightarrow \nu\nu$ background from the dilepton control samples is limited by the statistical uncertainty in the dilepton control samples due to the large branching fraction difference of the Z boson decays to muons and electrons compared to that to neutrinos. These transfer factors are shown in Figure 44. The dotted red (blue) line shows the ratio of the processes where the V-boson is produced via EW (QCD) production where as the solid lines show the ratio of the processes where the V-boson is produced both via EW and QCD production.

Finally, a transfer factor is also defined to connect the $Z(\nu\nu) + \text{jets}$ and $W(\ell\nu) + \text{jets}$ background yields in the signal region to further benefit from larger statistical power that $W(\ell\nu) + \text{jets}$ background has making it possible to experimentally constrain $Z(\nu\nu) + \text{jets}$ production at high dijet masses. These transfer factors are shown in Figure 45. The dotted red (blue) line shows the ratio of the processes where the V-boson is produced via EW (QCD) production where as the solid lines show the ratio of the processes where the V-boson is produced both via EW and QCD production.

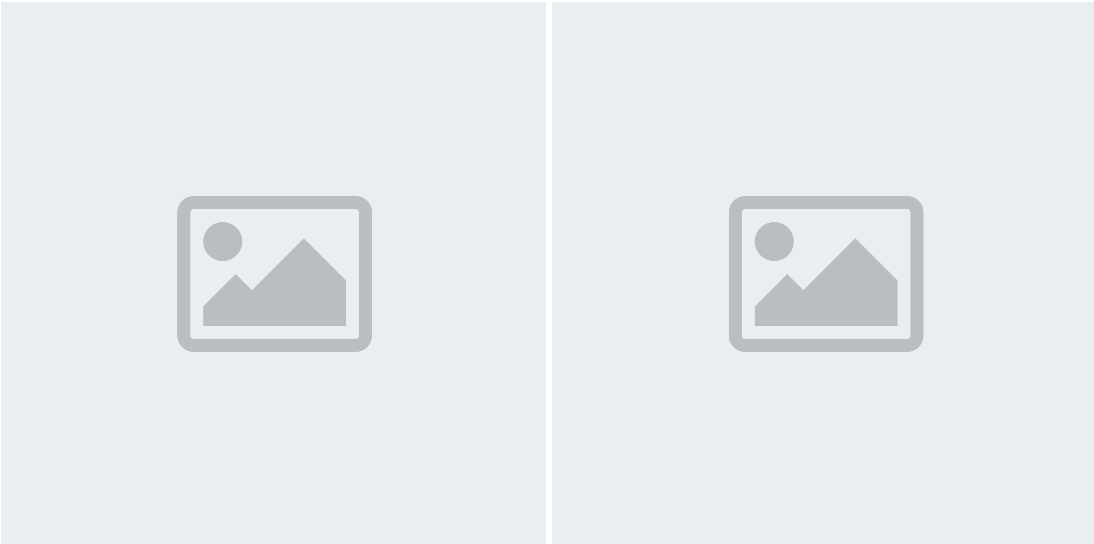

Figure 44: Transfer factors for the $Z(\nu\nu) +$ jetsbackground as a function of the dijet mass using the dimuon, and dielectron control regions in vbf final state. The grey band shows the statistical and systematic uncertainties on the ratios.

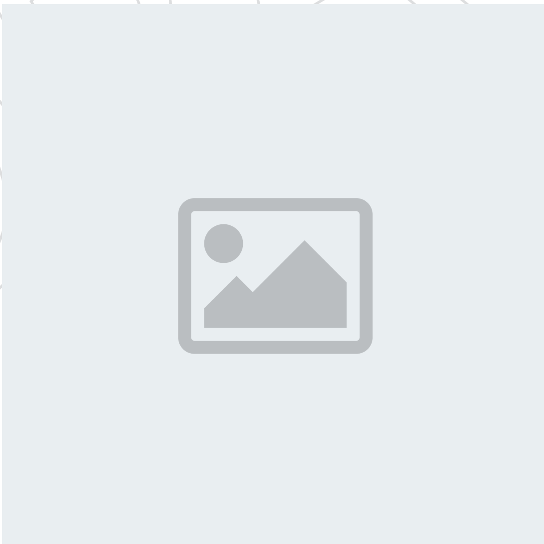

Figure 45: Transfer factors for the to estimate the $Z(\nu\nu)$+jets background from W+jets in the signal region as a function of the dijet mass for the vbf final state. The grey band shows the statistical and systematic uncertainties on the ratios.

7.3 Systematic uncertainties

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

7.3.1 Theoretical uncertainties

Theoretical uncertainties in V+jets processes include effects from QCD and EW higher-order corrections along with PDF modeling uncertainty. 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. This uncertainty is treated to be uncorrelated across the Z + jets, W + jets processes, but correlated across the bins of the dijet mass distribution.

The PDF uncertainty has been estimated using the standard deviations of weights provided in the NNPDF3.0 parton distribution set and using the RMS of each bin of the distribution after varying the full spectra by these weights. This uncertainty is treated to be correlated across the Z + jets, W + jets processes, and the bins of the dijet mass distribution.

The uncertainty due the EW corrections is assumed to be the full correction itself as a conservative approach. The uncertainty is treated correlated across the processes and across the bins of the dijet mass spectra.

These uncertainties are applied to the QCD and EW V+jets processes, but are assumed to be uncorrelated. That is, the QCD W + jets component has a factorization scale uncertainty that is the same size as the factorization scale uncertainty on the EW W + jets component, but the two are treated as separate nuisances in the signal extraction fit.

The summary of the aforementioned theoretical uncertainties including their magnitude per process and on the ratio are shown in Figure 46

7.3.2 Experimental uncertainties

Check numbers for 2017/18 Experimental uncertainties including the reconstruction efficiency (1% per muon or electron), and selection efficiencies of leptons (1% per muon and 2% per electron), and hadronically decaying τ leptons (5%) are incorporated to the analysis. These reconstruction and selection efficiencies further translate into an uncertainty in the lepton veto efficiency of 3%. The lepton veto uncertainties 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 both the flavour composition of the W + jets process the acceptance of the lepton. The overall magnitude of the lepton-veto uncertainty is found to be around 3 (4)% and is found to be dominated by the τ-veto uncertainty.

The uncertainty in the modeling of $E_{\mathrm{T}}^{\mathrm{miss}}$ in simulation [31] is estimated to be 1-2% on the ratios and is dominated by the uncertainty on the jet energy scale. The jet energy scale uncertainties found to be not-cancelling in the ratio due to the differences in the jet kinematics between control and signal regions. The resulting jet energy scale uncertainties in the ratios can be seen in Figure 47

Lastly, uncertainties in the efficiency of the electron (2%), and $p_{\mathrm{T}}^{\mathrm{miss}}$ (2%) triggers, are included

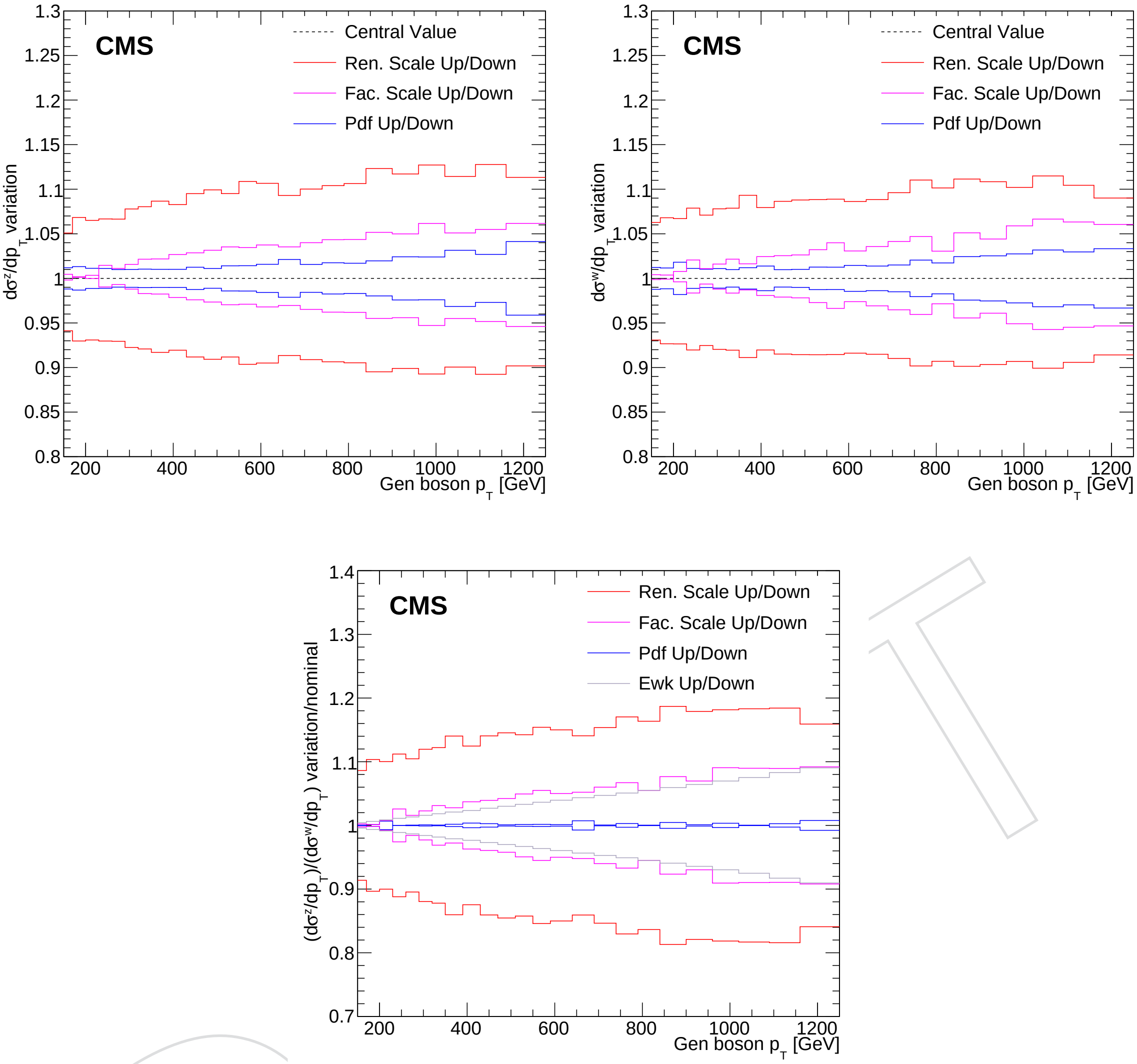

Figure 46: Theoretical uncertainties due to QCD and EW higher-order corrections and the PDF variation is shown for individual processes and for the ratio.

and are fully correlated across all the bins of dijet mass distribution.

The remaining uncertainties are for the monte carlo based backgrounds. A systematic uncertainty of 10% for the top quark background due to the modeling of the top quark p_{T} distribution in simulation. The uncertainty in the efficiency of the b jet veto is estimated to be 3% (1)% for the top quark (diboson) background. In addition, systematic uncertainties of 10 and 20% are included in the normalizations of the top quark [32] and diboson backgrounds [33, 34], 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 between 50–150% due to the variations of the jet response and the statistical uncertainty of the extrapolation factors.

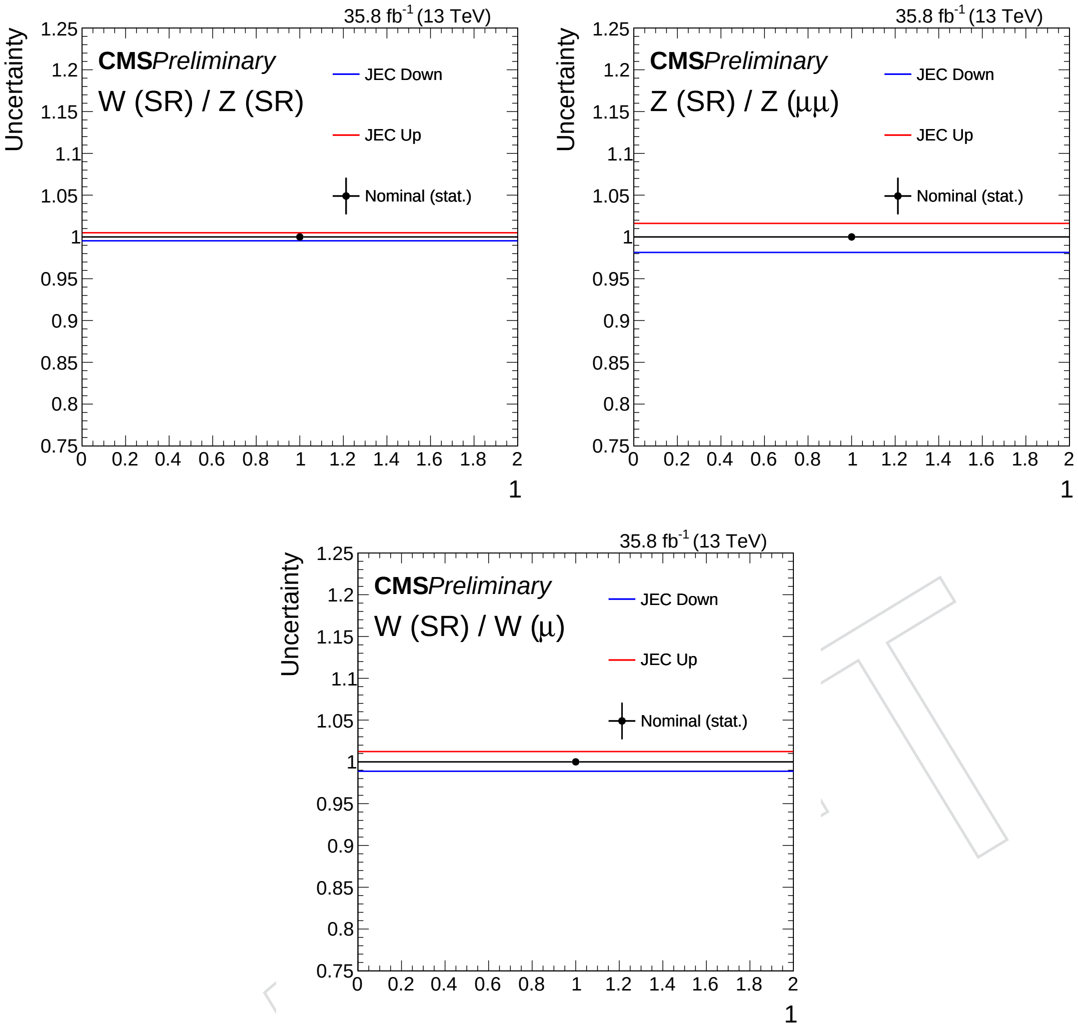

Figure 47: Redo for for 2017/18 The resulting jet energy scale uncertainties in the ratios of $W_{SR}/Z_{\nu\nu}$, $Z_{\nu\nu}/Z_{\mu\mu}$ and $W_{SR}/W_{\mu\nu}$, respectively.

7.4 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 + jets events and W + jets events in the control samples as a function of m_{jj}.

Figure 48 shows the ratio between $Z(\mu\mu)$ + jets and $W(\mu\nu)$ + jets (left), and the one between $Z(ee)$ + jetsand $W(e\nu)$ + jets processes (right) as a function of the dijet mass distribution for events selected both in multi-bin analysis (top) and single-bin analysis (bottom). Good agreement is observed between data and simulation after the application of the NLO corrections.

7.5 QCD multijet estimation

Update for 2017/18

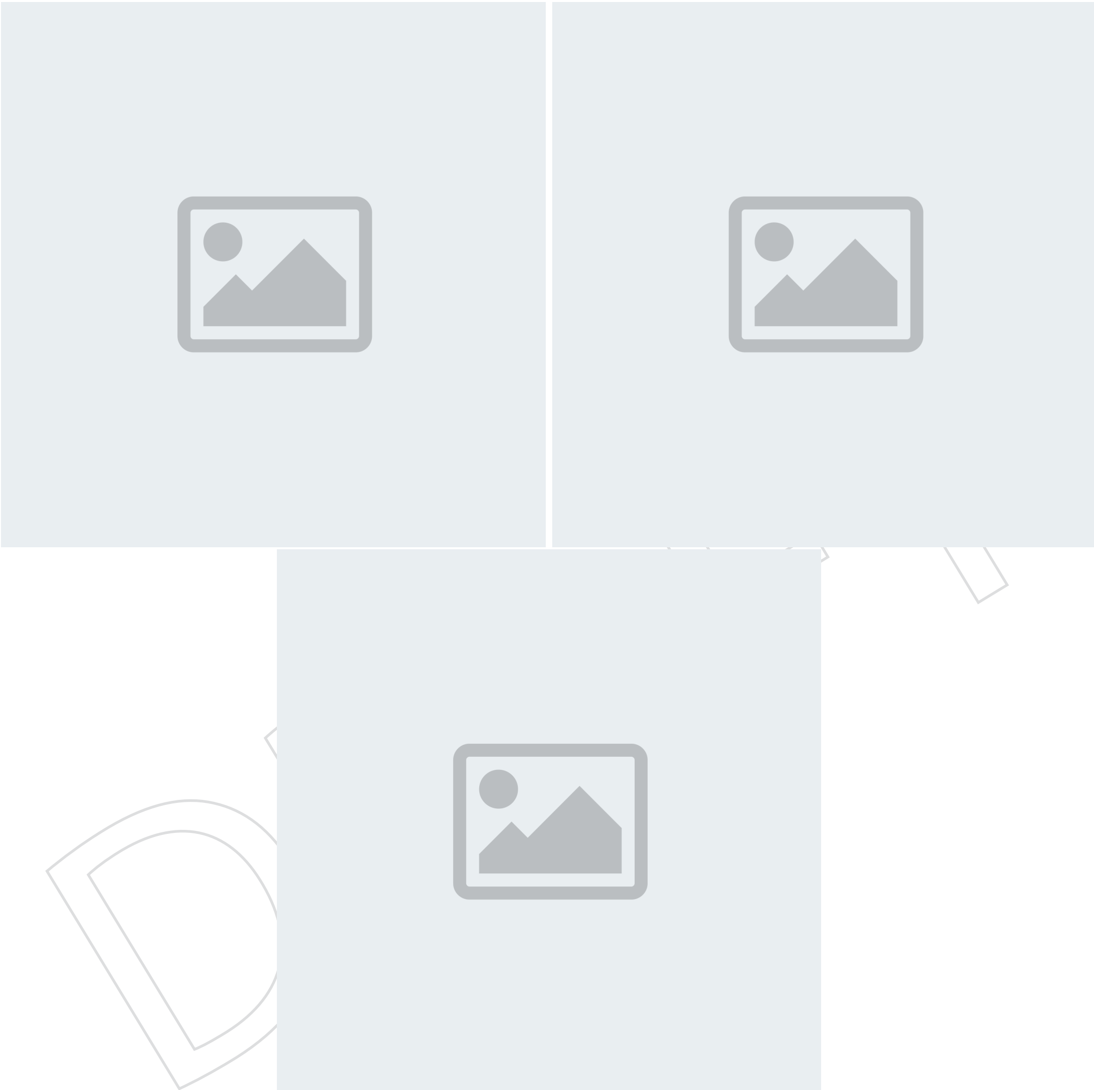

Figure 48: Comparison between data and MC simulation for, $\mathrm{Z}(\mu\mu)+\mathrm{jets}$ and $\mathrm{W}(\mu\nu)+\mathrm{jets}$ (top left), $\mathrm{Z}(ee)+\mathrm{jets}$and $\mathrm{W}(e\nu)+\mathrm{jets}$ processes (top right) ratio as a function of m_{jj}. The combined ratio is shown on bottom. The gray bands include both the pre-fit systematic uncertainties and the statistical uncertainty in the simulation.

Source	Process	Uncertainty
Electron trigger	$W_{SR}/W_{e\nu}$, $Z_{\nu\nu}/Z_{ee}$	1%
E_{T}^{miss} trigger	W_{SR}/W_{CR}, $Z_{\nu\nu}/Z_{CR}$, Z/W,	2%
Muon-reco efficiency	$W_{SR}/W_{\mu\nu}$, $Z_{\nu\nu}/Z_{\mu\mu}$	1% (per leg)
Muon-ID efficiency	$W_{SR}/W_{\mu\nu}$, $Z_{\nu\nu}/Z_{\mu\mu}$	1% (per leg)
Muon-Iso efficiency	$W_{SR}/W_{\mu\nu}$, $Z_{\nu\nu}/Z_{\mu\mu}$	0.5% (per leg)
Electron-reco efficiency	$W_{SR}/W_{e\nu}$, $Z_{\nu\nu}/Z_{ee}$	1% (per leg)
Electron-IDiso efficiency	$W_{SR}/W_{e\nu}$, $Z_{\nu\nu}/Z_{ee}$	1.5% (per leg)
Lepton veto	$W_{l,SR}$	3% (QCD), 5.0% (EW)
Jet energy scale	W_{CR}/W_{SR}	2%
Jet energy scale	Z_{CR}/Z_{SR}	1%

Table 12: Check numbers for 2017/18 Summary of experimental uncertainties affecting used in the analysis.

8 Results

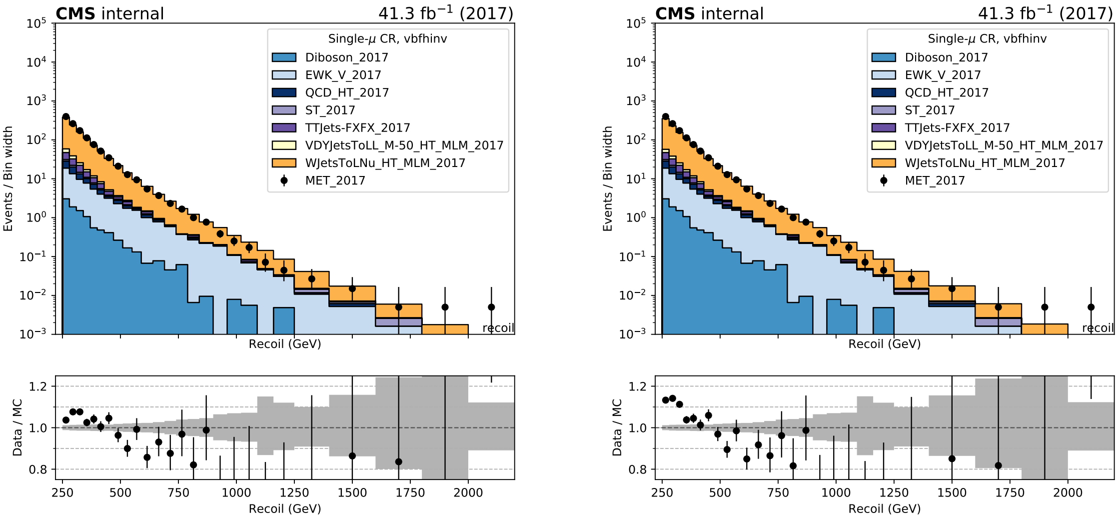

Figure 49: The recoil distribution in the single muon CR for the case in which 1D k-factors are used (left), compared to the case in which 2D k-factors are used (right), using 2017 samples.

A Comparison of 1D and 2D QCD k-factors

In this appendix, several distributions such as recoil, M_{jj} and $\Delta\eta_{jj}$ are compared for two cases in the single muon control region:

- Plots with 1D k-factors applied: Function of gen-boson pt
- Plots with 2D k-factors applied: Function of gen-boson pt and M_{jj}

Both k-factors are derived from samples in which the VBF selections are applied at the gen level. 1D k-factor is a function of generator level vector boson p_{T}, whereas 2D k-factor is a function of both generator level boson p_{T} and M_{jj}. The distributions are shown in the following pages.

Fig. 49 shows the comparison of the recoil distribution with 2017 samples, while Fig. 50 shows the same comparison with 2018 samples. Fig. 51 shows the comparison of the M_{jj} distribution with 2017 samples, while Fig. 52 shows the same comparison with 2018 samples. Fig. 53 shows the comparison of the $\Delta\eta_{jj}$ distribution with 2017 samples, while Fig. 54 shows the same comparison with 2018 samples. Fig. 55 shows the comparison of the muon p_{T} distribution with 2017 samples, while Fig. 56 shows the same comparison with 2018 samples. Fig. 57 shows the comparison of the muon η distribution with 2017 samples, while Fig. 58 shows the same comparison with 2018 samples.

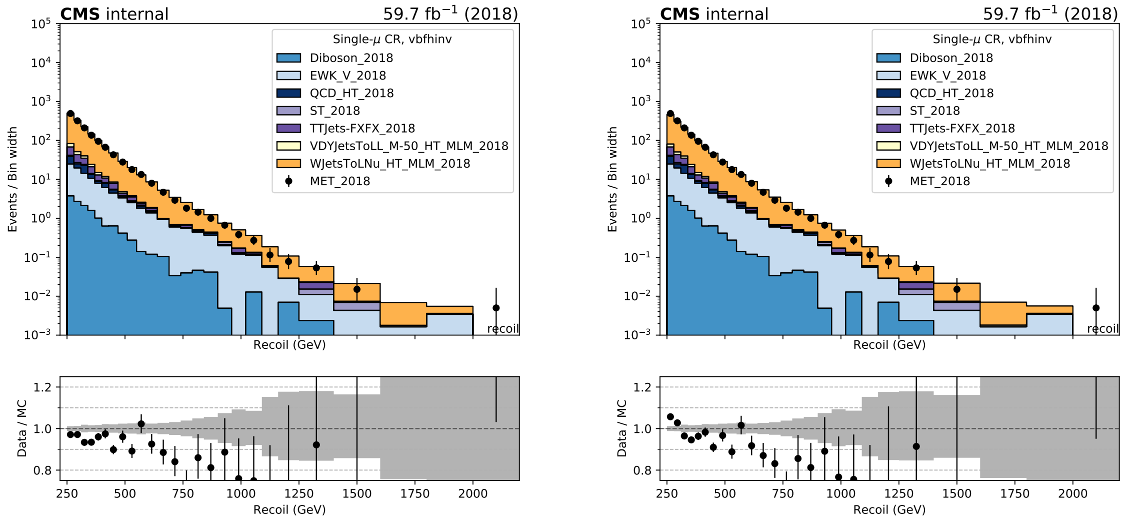

Figure 50: The recoil distribution in the single muon CR for the case in which 1D k-factors are used (left), compared to the case in which 2D k-factors are used (right), using 2018 samples.

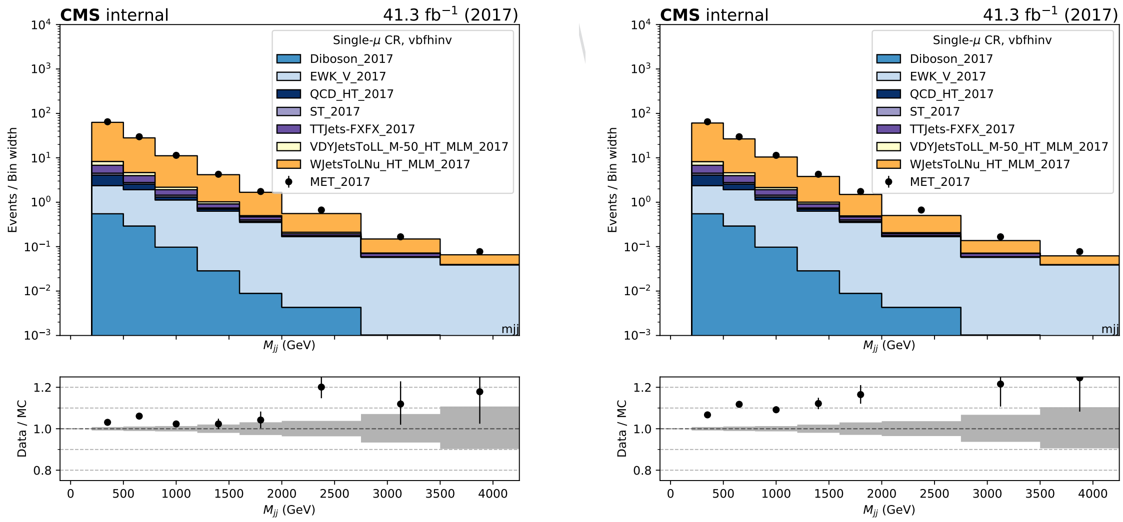

Figure 51: The M_{jj} distribution in the single muon CR for the case in which 1D k-factors are used (left), compared to the case in which 2D k-factors are used (right), using 2017 samples.

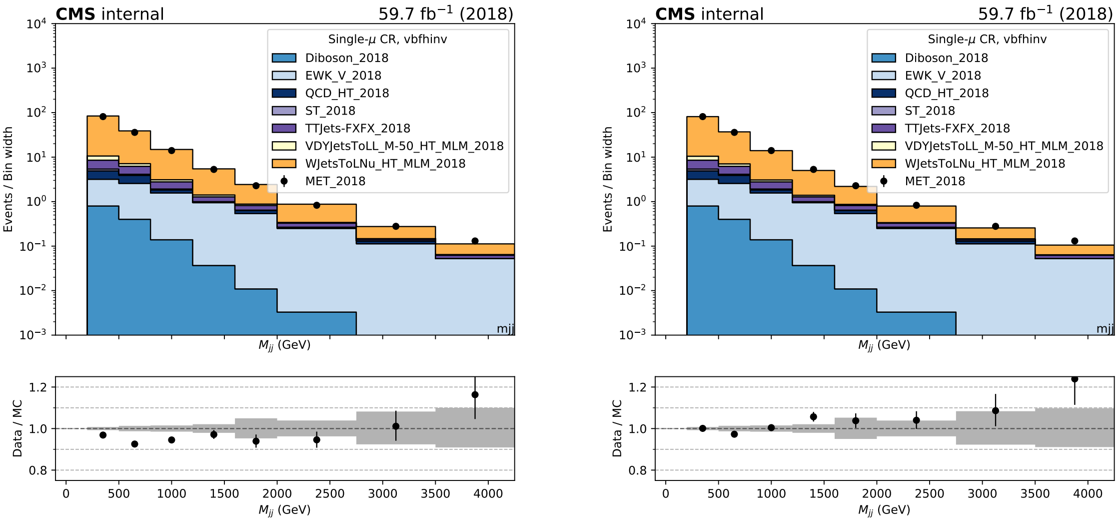

Figure 52: The M_{jj} distribution in the single muon CR for the case in which 1D k-factors are used (left), compared to the case in which 2D k-factors are used (right), using 2018 samples.

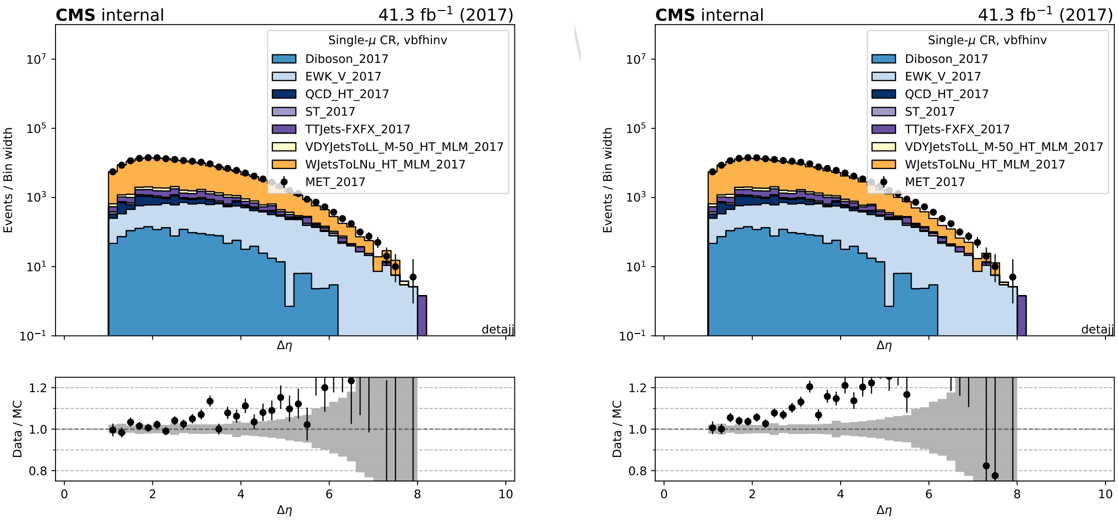

Figure 53: The $\Delta\eta_{jj}$ distribution in the single muon CR for the case in which 1D k-factors are used (left), compared to the case in which 2D k-factors are used (right), using 2017 samples.

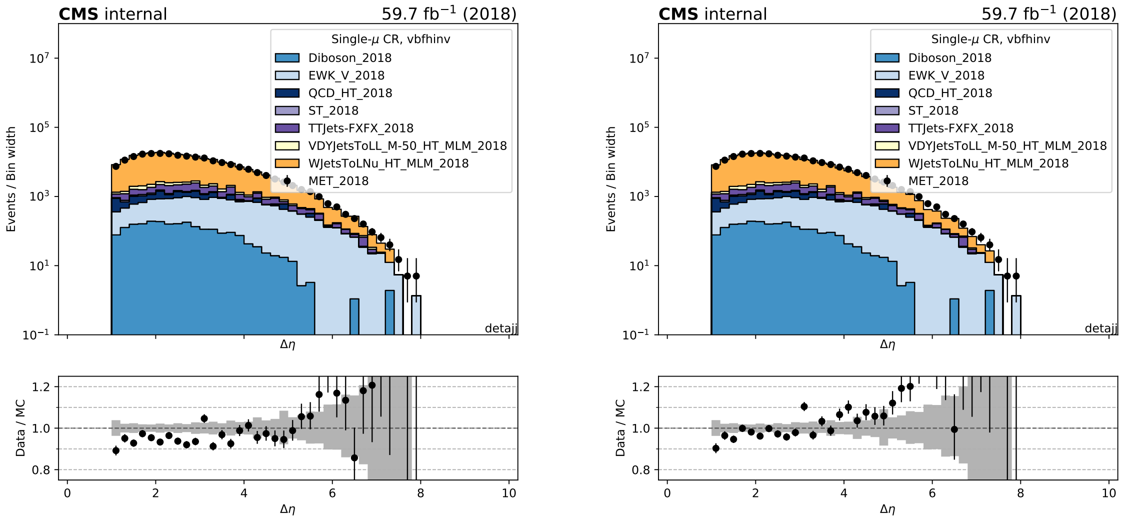

Figure 54: The $\Delta\eta_{jj}$ distribution in the single muon CR for the case in which 1D k-factors are used (left), compared to the case in which 2D k-factors are used (right), using 2018 samples.

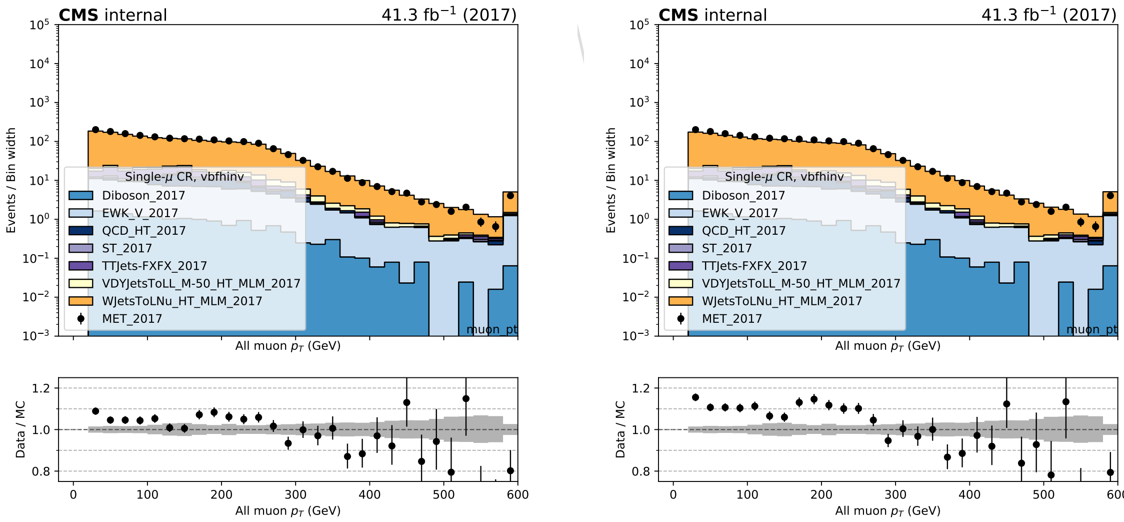

Figure 55: The muon p_{T} distribution in the single muon CR for the case in which 1D k-factors are used (left), compared to the case in which 2D k-factors are used (right), using 2017 samples.

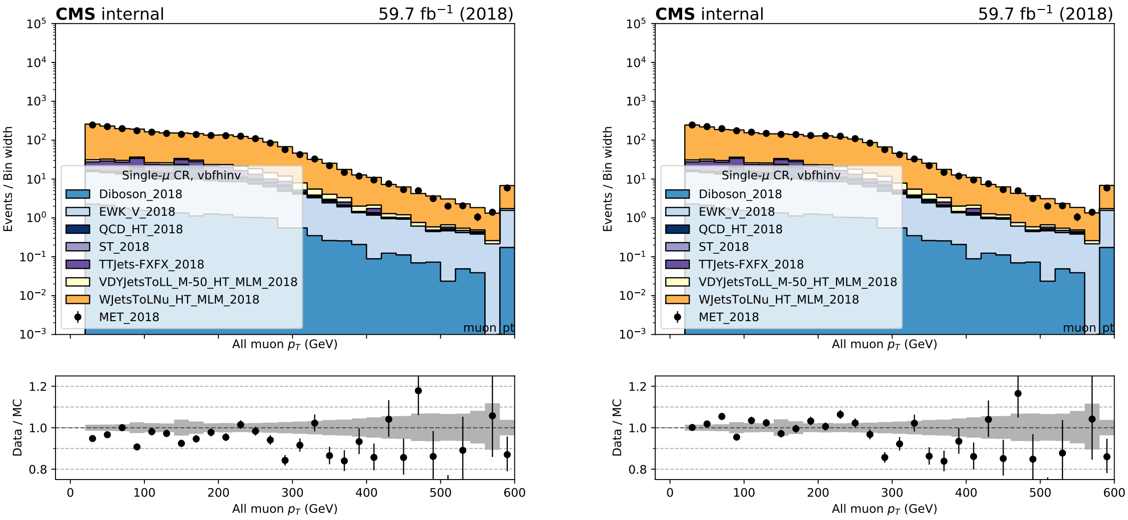

Figure 56: The muon p_{T} distribution in the single muon CR for the case in which 1D k-factors are used (left), compared to the case in which 2D k-factors are used (right), using 2018 samples.

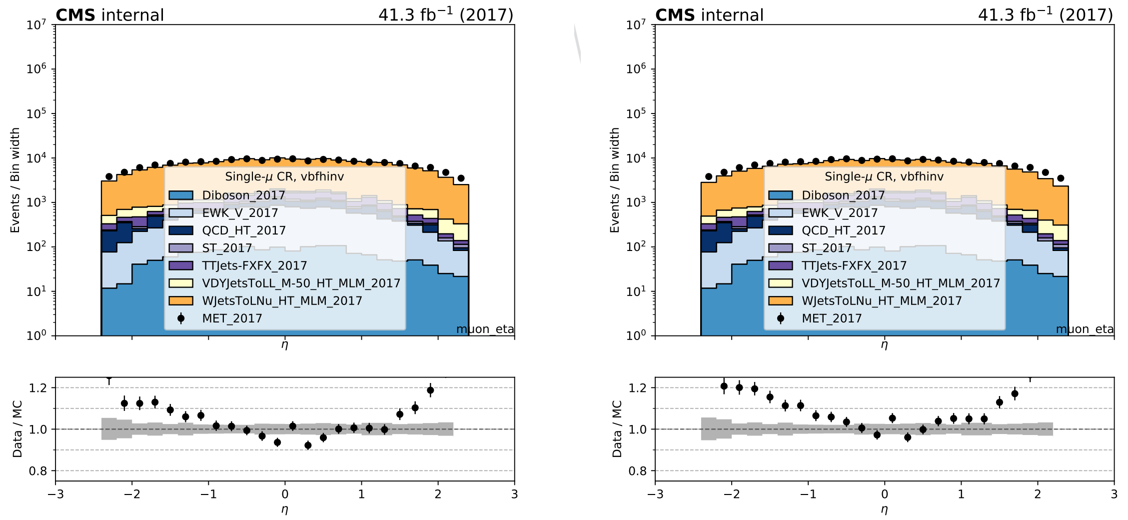

Figure 57: The muon η distribution in the single muon CR for the case in which 1D k-factors are used (left), compared to the case in which 2D k-factors are used (right), using 2017 samples.

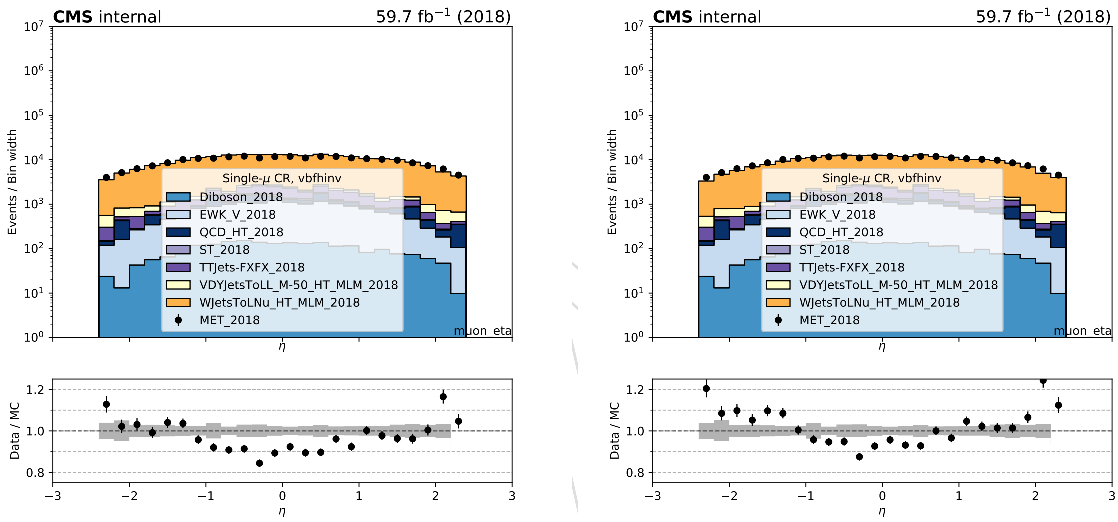

Figure 58: The muon η distribution in the single muon CR for the case in which 1D k-factors are used (left), compared to the case in which 2D k-factors are used (right), using 2018 samples.

B Trigger efficiencies in double muon CR

In this section, the efficiencies of $p_{\mathrm{T}}^{\mathrm{miss}} + H_{\mathrm{T}}^{\mathrm{miss}}$ triggers measured in double muon control region are presented. In addition, efficiencies of these triggers as a function of mjj both in single muon and double muon control region are also presented.

Fig. 59 shows the efficiencies as a function of M_{jj} in data and MC for the three categories in 2017, whereas Fig. 60 shows the efficiencies as a function of M_{jj} in data and MC for the three categories in 2018.

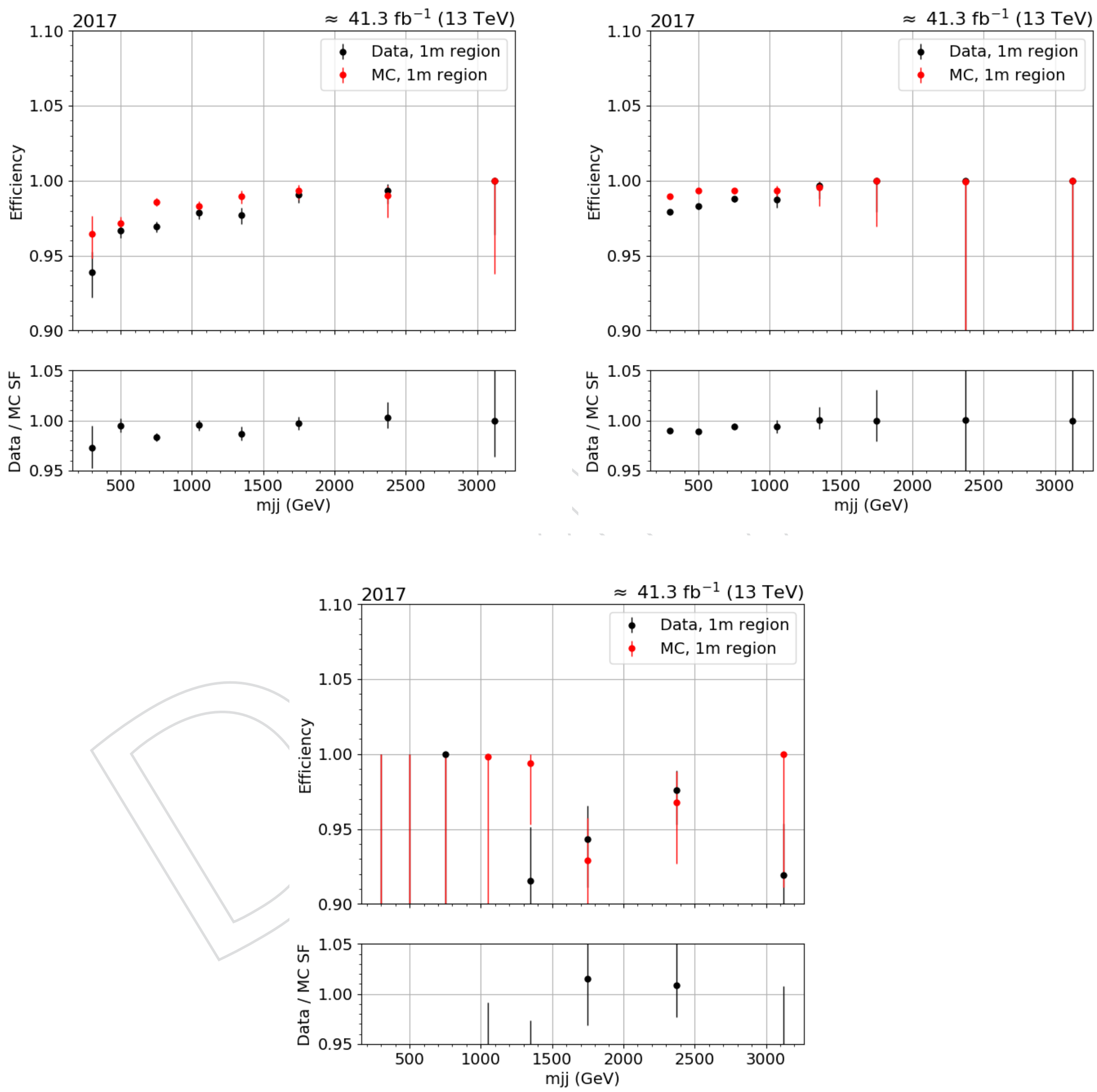

Figure 59: MET trigger efficiency as a function of mjj in three categories: One forward jet and one central jet, two central jets and two forward jets. These results are obtained from 2017 data and MC samples with the selection of single muon events.

Figs. 61 and **??** show the efficiencies obtained in double muon CR as a function of recoil in data and MC for the three jet categories in 2017 and 2018, respectively. Figs. 63 and **??** show the efficiencies obtained in double muon CR as a function of M_{jj} in data and MC for the three jet

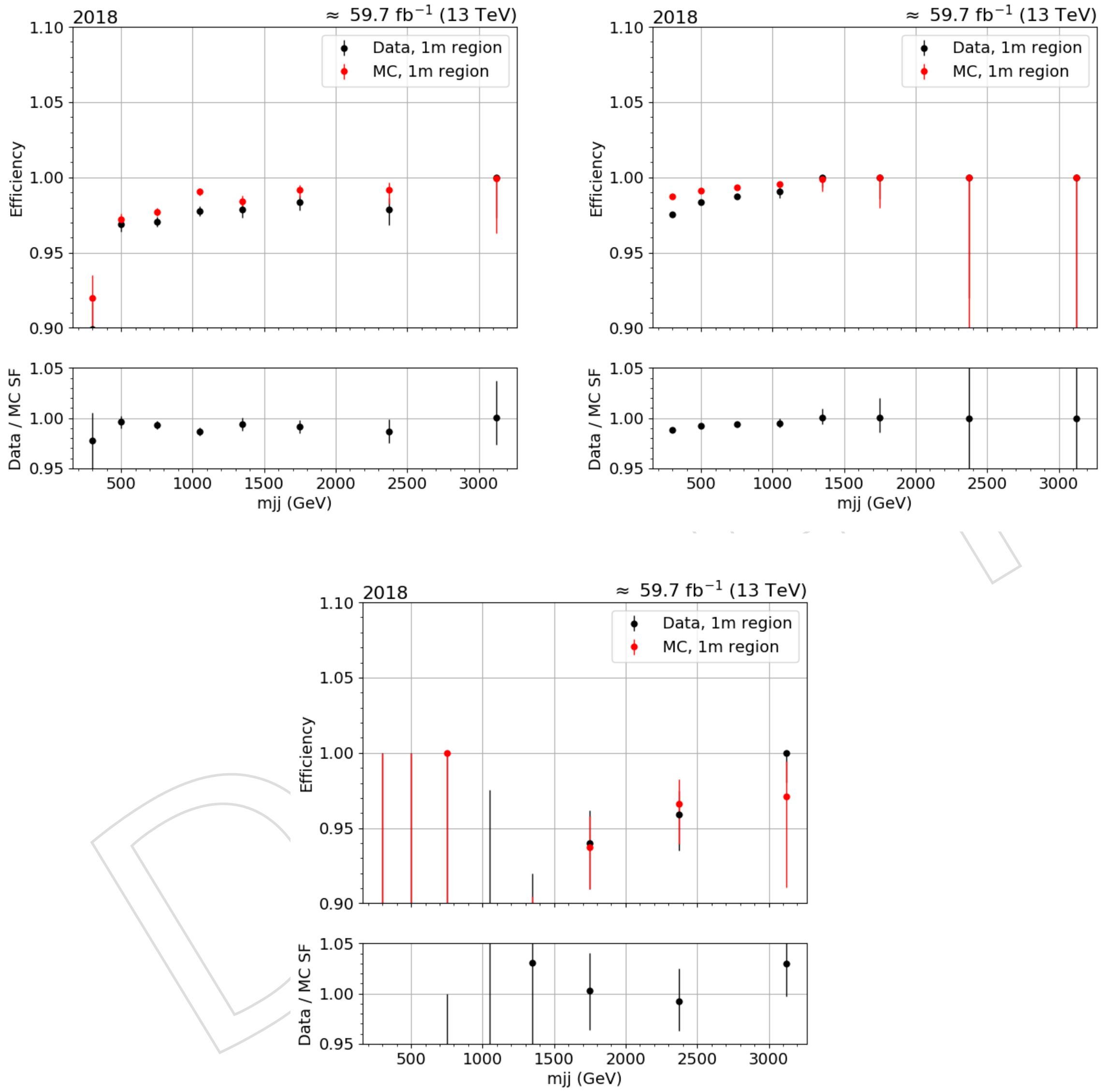

Figure 60: MET trigger efficiency as a function of mjj in three categories: One forward jet and one central jet, two central jets and two forward jets. These results are obtained from 2018 data and MC samples with the selection of single muon events.

categories in 2017 and 2018, respectively.

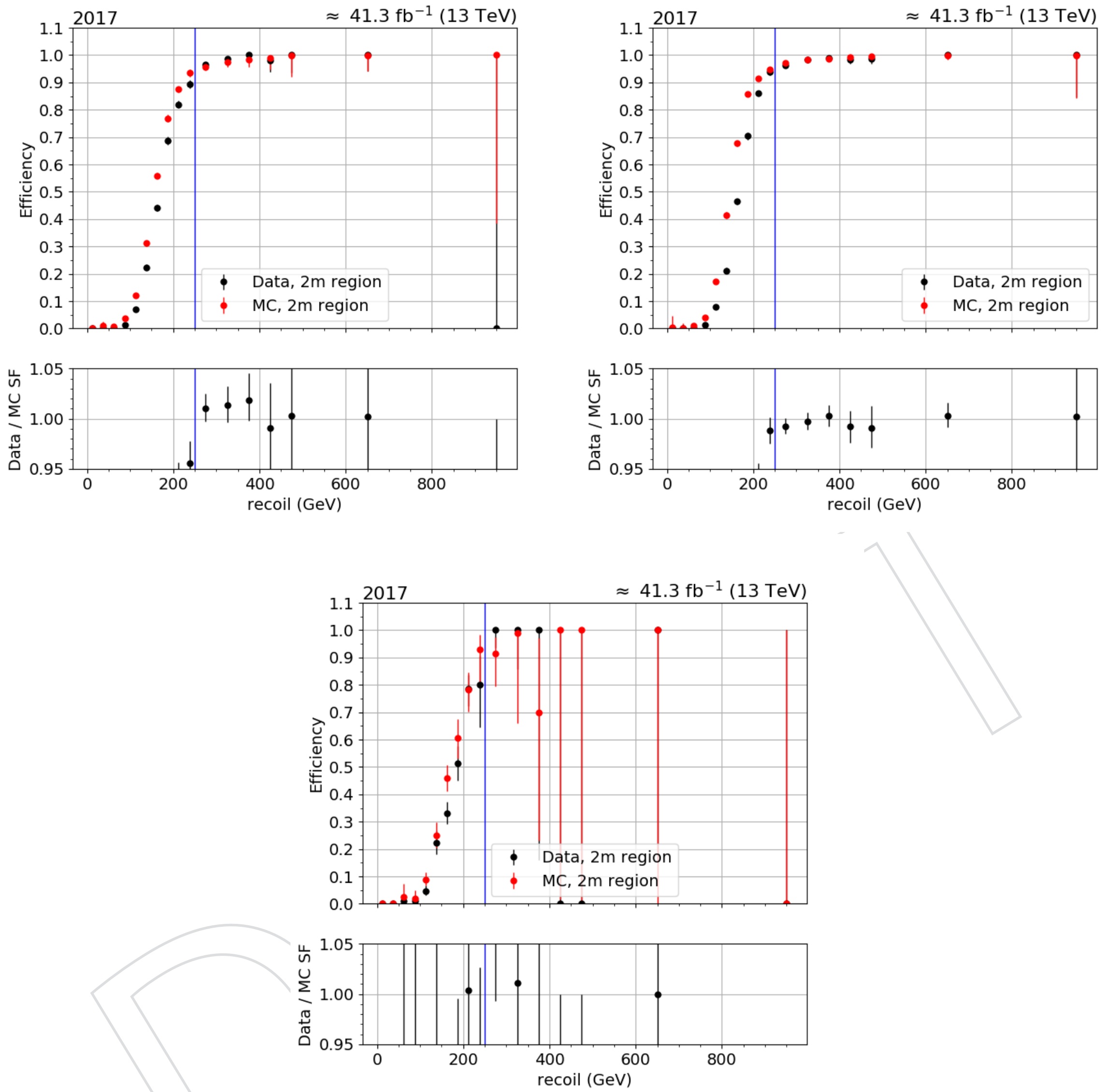

Figure 61: MET trigger efficiency as a function of recoil in three categories: One forward jet and one central jet, two central jets and two forward jets. These results are obtained from 2017 data and MC samples with the selection of double muon events.

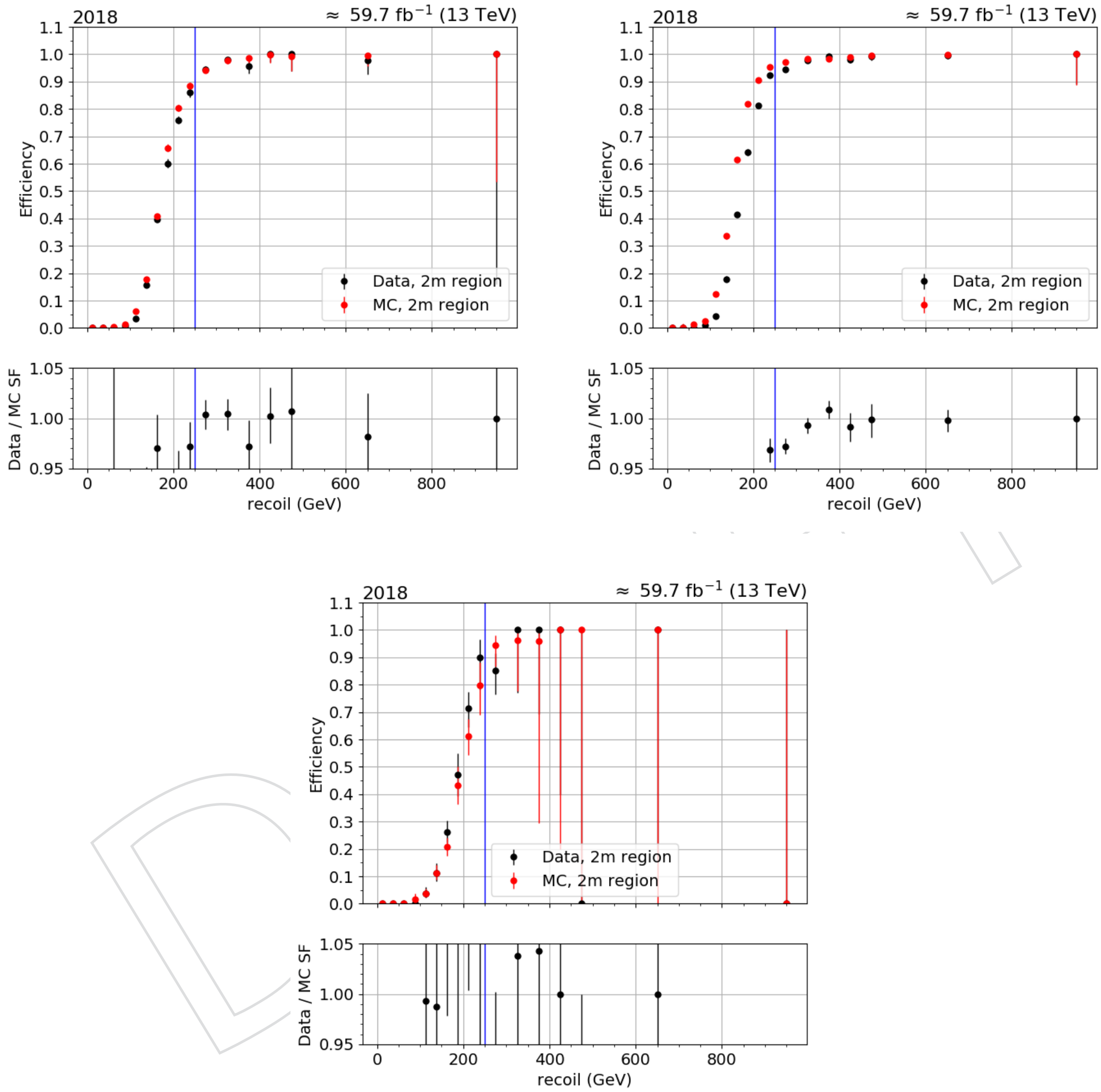

Figure 62: MET trigger efficiency as a function of recoil in three categories: One forward jet and one central jet, two central jets and two forward jets. These results are obtained from 2018 data and MC samples with the selection of double muon events.

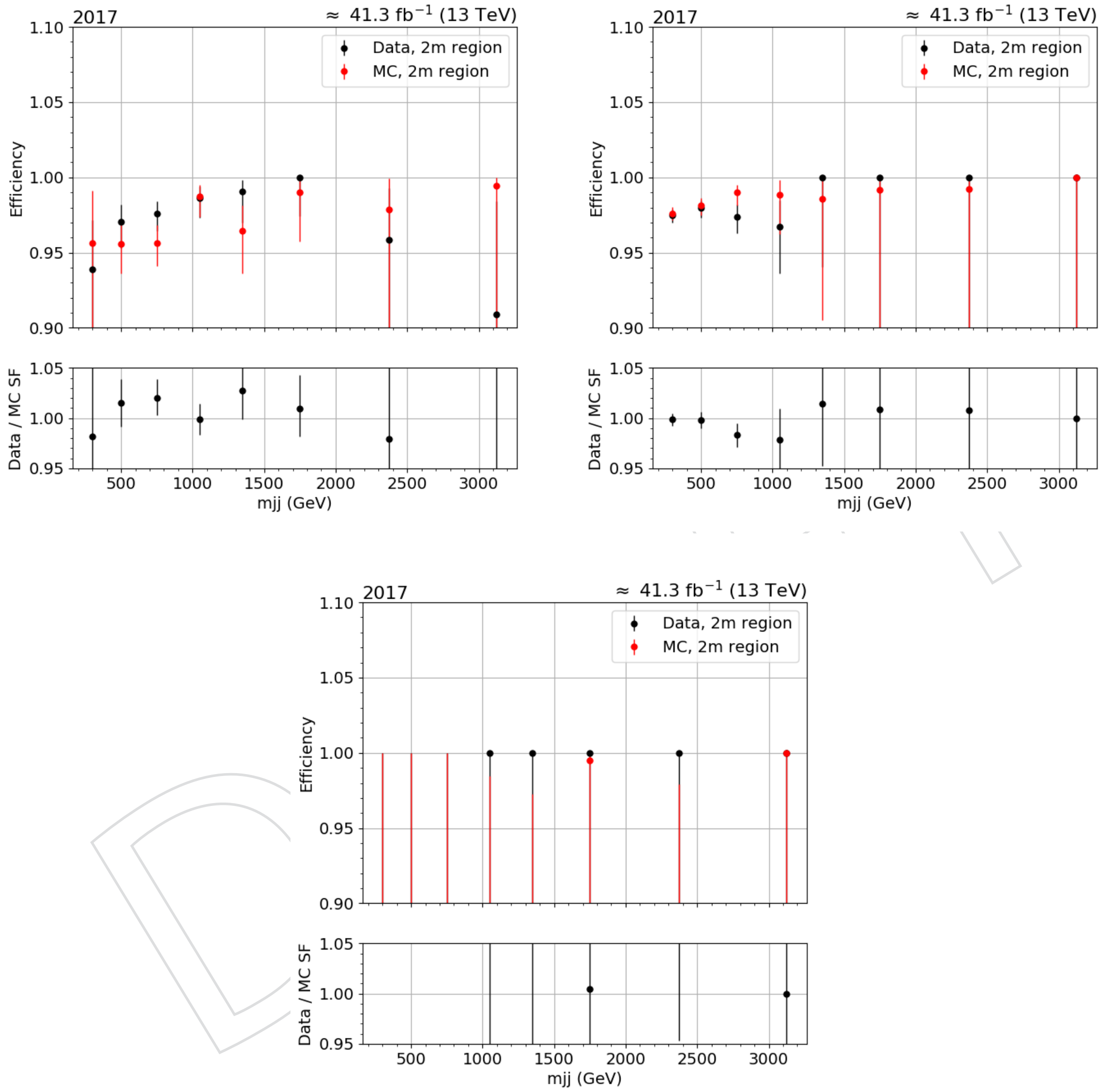

Figure 63: MET trigger efficiency as a function of mjj in three categories: One forward jet and one central jet, two central jets and two forward jets. These results are obtained from 2017 data and MC samples with the selection of double muon events.

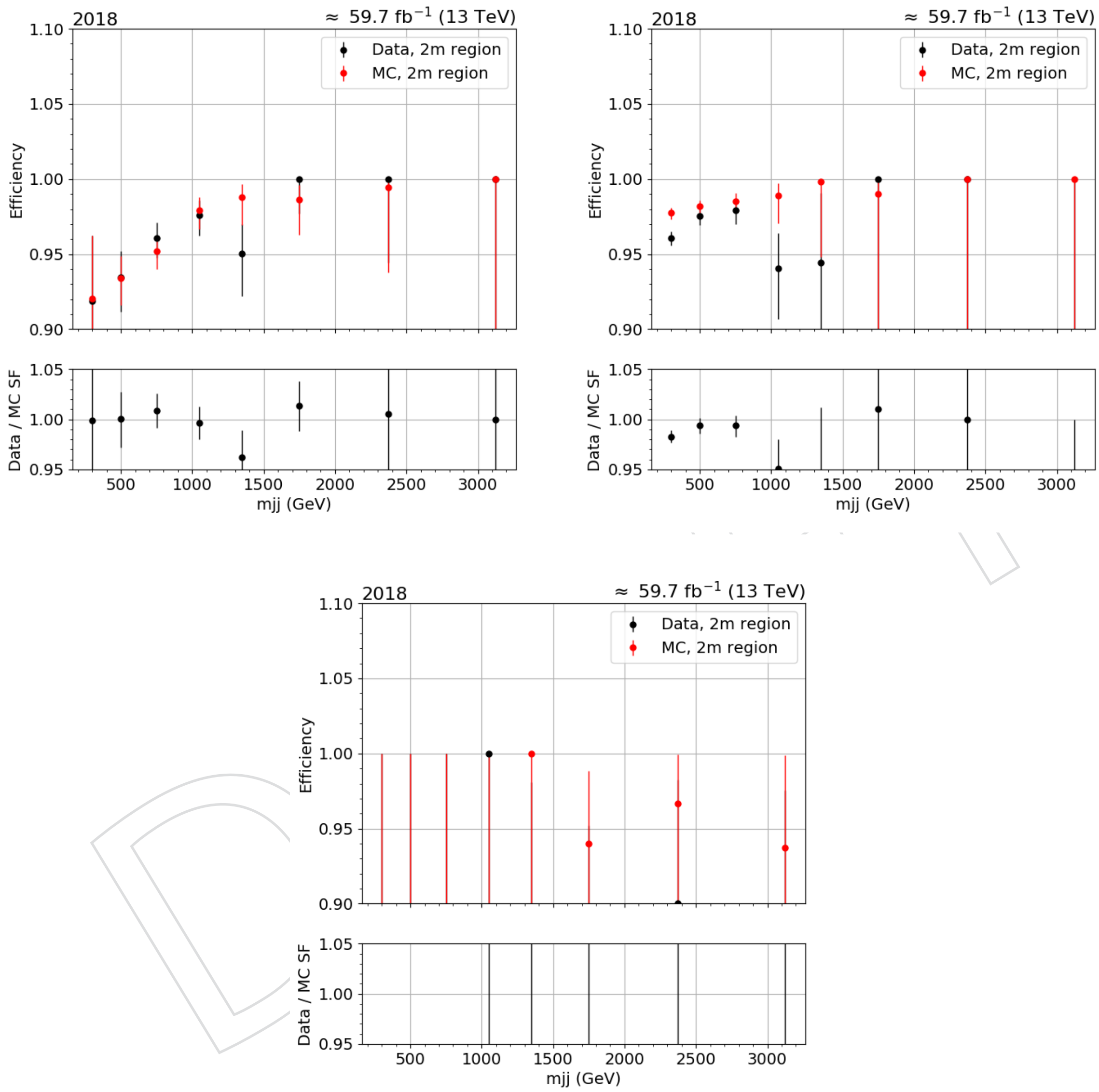

Figure 64: MET trigger efficiency as a function of mjj in three categories: One forward jet and one central jet, two central jets and two forward jets. These results are obtained from 2018 data and MC samples with the selection of double muon events.

Acknowledgments

References

[1] CMS Collaboration, "Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements", `arXiv:1903.12179`.

[2] M. L. Mangano, M. Moretti, F. Piccinini, and M. Treccani, "Matching matrix elements and shower evolution for top-quark production in hadronic collisions", *JHEP* **01** (2007) 013, `doi:10.1088/1126-6708/2007/01/013`, `arXiv:hep-ph/0611129`.

[3] R. Frederix and S. Frixione, "Merging meets matching in MC@NLO", *JHEP* **12** (2012) 061, `doi:10.1007/JHEP12(2012)061`, `arXiv:1209.6215`.

[4] NNPDF Collaboration, "Parton distributions from high-precision collider data", *Eur. Phys. J.* **C77** (2017), no. 10, 663, `doi:10.1140/epjc/s10052-017-5199-5`, `arXiv:1706.00428`.

[5] GEANT4 Collaboration, "GEANT4: A simulation toolkit", *Nucl. Instrum. Meth. A* **506** (2003) 250, `doi:10.1016/S0168-9002(03)01368-8`.

[6] LHC Higgs Cross Section Working Group Collaboration, "Handbook of LHC Higgs cross sections: 4. deciphering the nature of the Higgs sector", `doi:10.2172/1345634,10.23731/CYRM-2017-002`, `arXiv:1610.07922`.

[7] CMS EGamma POG, "EgHLTRunIISummary". `https://twiki.cern.ch/twiki/bin/view/CMS/EgHLTRunIISummary`.

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

[9] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet user manual", *Eur. Phys. J. C* **72** (2012) 1896, `doi:10.1140/epjc/s10052-012-1896-2`, `arXiv:1111.6097`.

[10] CMS Collaboration, "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV", *JINST* **12** (2017) P02014, `doi:10.1088/1748-0221/12/02/P02014`, `arXiv:1607.03663`.

[11] CMS Collaboration, "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV", *JINST* **13** (2018), no. 05, P05011, `doi:10.1088/1748-0221/13/05/P05011`, `arXiv:1712.07158`.

[12] CMS JME POG. `https://twiki.cern.ch/twiki/bin/view/CMS/MissingETOptionalFiltersRun2`.

[13] CMS Collaboration, "Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8\,\mathrm{TeV}$", *JINST* **10** (2015) P06005, `doi:10.1088/1748-0221/10/06/P06005`, `arXiv:1502.02701`.

[14] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *JINST* **9** (2014) P10009, `doi:10.1088/1748-0221/9/10/P10009`, `arXiv:1405.6569`.

[15] CMS EGamma POG. `https://twiki.cern.ch/twiki/bin/viewauth/CMS/CutBasedElectronIdentificationRun2#Spring15_selection_25ns`.

[16] CMS Collaboration Collaboration, "Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7\,\mathrm{TeV}$", *JINST* **7** (2012) P10002, doi:10.1088/1748-0221/7/10/P10002, arXiv:1206.4071.

[17] CMS Muon POG. https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideMuonIdRun2#Loose_Muon.

[18] CMS Muon POG. https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideMuonIdRun2#Tight_Muon.

[19] CMS Collaboration, "Reconstruction and identification of τ lepton decays to hadrons and ν_τ at CMS", *JINST* **11** (2016) P01019, doi:10.1088/1748-0221/11/01/P01019, arXiv:1510.07488.

[20] tau-POG. https://twiki.cern.ch/twiki/bin/view/CMS/TauIDRecommendation13TeV#Measurement_in_Z_tautau_events.

[21] CMS Collaboration, "Performance of photon reconstruction and identification with the CMS detector in proton-proton collisions at $\sqrt{s} = 8\,\mathrm{TeV}$", *JINST* **10** (2015) P08010, doi:10.1088/1748-0221/10/08/P08010, arXiv:1502.02702.

[22] CMS EGamma POG. https://twiki.cern.ch/twiki/bin/viewauth/CMS/CutBasedPhotonIdentificationRun2#SPRING15_selections_bunch_crossi.

[23] CMS Lumi POG. https://twiki.cern.ch/twiki/bin/view/CMS/PileupJSONFileforData#Pileup_JSON_Files_For_Run_II.

[24] CMS Muon POG. https://twiki.cern.ch/twiki/bin/view/CMS/MuonReferenceSelectionAndCalibrationsRun2.

[25] J. M. Lindert et al., "Precise predictions for V+jets dark matter backgrounds", (2017). arXiv:1705.04664.

[26] "Search for vbf higgs boson decays to invisible particles with 2016 dataset", 2017. CMS Analysis Note : AN-17-267. http://cms.cern.ch/iCMS/jsp/openfile.jsp?tp=draft&files=AN2017_267_v4.pdf.

[27] CMS Collaboration, "Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector", *Submitted to: JINST* (2019) arXiv:1903.06078.

[28] P. Lenzi, C. A. Palmer, J. Thomas-Wilsker et al., "Performance of flavour tagging algorithms at 13 tev with 2018 data", CMS Note 2018/323, 2019.

[29] CMS Collaboration, "Search for New Physics with Jets and Missing Transverse Momentum in pp collisions at $\sqrt{s} = 7$ TeV", *JHEP* **08** (2011) 155, doi:10.1007/JHEP08(2011)155, arXiv:1106.4503.

[30] CMS Collaboration, "Search for dark matter produced with an energetic jet or a hadronically decaying W or Z boson at $\sqrt{s} = 13$ TeV", *JHEP* **07** (2017) 014, doi:10.1007/JHEP07(2017)014, arXiv:1703.01651.

[31] CMS Collaboration, "Performance of the CMS missing transverse momentum reconstruction in pp data at $\sqrt{s}$ = 8 TeV", *JINST* **10** (2015) P02006, doi:10.1088/1748-0221/10/02/P02006, arXiv:1411.0511.

[32] CMS Collaboration, "Measurement of the top quark pair production cross section in proton-proton collisions at $\sqrt{s} = 13$ TeV", *Phys. Rev. Lett.* **116** (2016) 052002, doi:10.1103/PhysRevLett.116.052002, arXiv:1510.05302.

[33] CMS Collaboration, "Measurement of the ZZ production cross section and Z $\to \ell^+\ell^-\ell'^+\ell'^-$ branching fraction in pp collisions at $\sqrt{s}$=13 TeV", *Phys. Lett. B* **763** (2016) 280–303, doi:10.1016/j.physletb.2016.10.054, arXiv:1607.08834.

[34] CMS Collaboration, "Measurement of the WZ production cross section in pp collisions at $\sqrt{s}$=13 TeV", *Phys. Lett. B* **766** (2017) 268–290, doi:10.1016/j.physletb.2017.01.011, arXiv:1607.06943.