

# CMS Draft Analysis Note

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

2017/04/07  
Head Id: 398031  
Archive Id: 398126  
Archive Date: 2017/04/06  
Archive Tag: trunk

## Search for lepton flavor violating decays of the Higgs boson in 2016 data

Daniel Troendle<sup>1</sup>, Colin Jessop, Silvia Taroni, Nabarun Dev, Fanbo Meng<sup>2</sup>, Sridhara Dasu,  
Cecile Caillol, Aaron Levine<sup>3</sup>, and María Cepeda<sup>4</sup>

<sup>1</sup> University of Hamburg

<sup>2</sup> University of Notre Dame

<sup>3</sup> University of Wisconsin-Madison

<sup>4</sup> CERN

### Abstract

A search for lepton flavor violating decays of the 125 GeV Higgs boson in the  $e\tau$  and  $\mu\tau$  decay modes is presented. The search is based on a dataset of  $36 \text{ fb}^{-1}$  of proton-proton collisions collected by the CMS detector in 2016, at a center-of-mass energy of 13 TeV. The tau leptons are reconstructed in the leptonic and hadronic decay modes. The analysis is expected to exclude the branching fractions of the Higgs boson greater than  $\mathcal{B}(H \rightarrow \mu\tau) < 0.26\%$  and  $\mathcal{B}(H \rightarrow e\mu) < 0.32\%$  at the 95% confidence level.

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

PDFAuthor: Cecile Caillol, Maria Cepeda, Nabarun Dev, Colin Jessop, Fanbo Meng, Silvia Taroni  
PDFTitle: Search for lepton flavor violating decays of the Higgs boson in 2016 data  
PDFSubject: CMS  
PDFKeywords: CMS, physics, Higgs, Taus

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



## 1 Contents

1	1	Introduction . . . . .	3
2	1.1	Recap of Run 1 LFV H Analysis technique . . . . .	4
3	1.2	Overview of Run 2 Analysis . . . . .	5
4	2	Data and simulation samples . . . . .	6
5	3	Event reconstruction . . . . .	8
6	4	Background estimation . . . . .	9
7	4.1	$Z \rightarrow \tau\tau$ . . . . .	9
8	4.2	$Z \rightarrow \ell\ell$ . . . . .	9
9	4.3	$t\bar{t} + \text{jets}$ . . . . .	10
10	4.4	Misidentified Leptons . . . . .	10
11	4.5	Other backgrounds . . . . .	14
12	5	$H \rightarrow \mu\tau_h$ channel . . . . .	15
13	5.1	Loose selection . . . . .	15
14	5.2	Cut-based analysis . . . . .	15
15	5.3	Multivariate analysis . . . . .	15
16	5.4	Validation of the background methods . . . . .	18
17	6	$H \rightarrow e\tau_h$ channel . . . . .	20
18	6.1	Loose selection . . . . .	20
19	6.2	Cut-based analysis . . . . .	23
20	6.3	Multivariate analysis . . . . .	23
21	6.4	Validation of background methods . . . . .	24
22	7	$H \rightarrow \mu\tau_e$ channel . . . . .	29
23	7.1	loose selection . . . . .	29
24	7.2	Cut-based analysis . . . . .	29
25	7.3	Multivariate analysis . . . . .	29
26	7.4	Validation of the background methods . . . . .	30
27	8	$H \rightarrow e\tau_\mu$ channel . . . . .	37
28	8.1	Loose selection . . . . .	37
29	8.2	Cut-based analysis . . . . .	40
30	8.3	Multivariate analysis . . . . .	40
31	8.4	Background validation . . . . .	47
32	9	Systematic uncertainties . . . . .	48
33	9.1	Luminosity . . . . .	48
34	9.2	$\tau_h$ identification . . . . .	48
35	9.3	Muon triggering, identification and isolation . . . . .	49
36	9.4	Electron triggering, identification and isolation . . . . .	49
37	9.5	$e/\mu \rightarrow \tau_h$ fake rates . . . . .	49
38	9.6	Normalization of MC samples . . . . .	49
39	9.7	Reducible background from the fake rate method . . . . .	49
40	9.8	Theoretical uncertainties on the signal . . . . .	49
41	9.9	B-tagging efficiency . . . . .	50
42	9.10	$\tau_h$ energy scale . . . . .	50

---

44	9.11	$e/\mu \rightarrow \tau_h$ energy scale . . . . .	50
45	9.12	Unclustered energy scale . . . . .	51
46	9.13	Electron energy scale . . . . .	51
47	9.14	Muon energy scale . . . . .	51
48	9.15	Jet energy scale . . . . .	51
49	9.16	Pile-up uncertainties . . . . .	51
50	9.17	Bin-by-bin uncertainties . . . . .	52
51	9.18	Propagation of the uncertainties to the BDT output . . . . .	52
52	10	Results . . . . .	53
53	10.1	$H \rightarrow \mu\tau$ results . . . . .	53
54	10.2	$H \rightarrow e\tau$ results . . . . .	53
55	10.3	Summary and comparison with past results . . . . .	60
56	10.4	Limits on the Yukawa couplings . . . . .	60
57	11	Summary . . . . .	62
58	A	Theoretical Uncertainties . . . . .	67
59	A.1	PDF uncertainty . . . . .	67
60	A.2	Scale uncertainty . . . . .	68
61	B	BDT method A . . . . .	68
62	B.1	$H \rightarrow \mu\tau_h$ . . . . .	68
63	B.2	$H \rightarrow e\tau_h$ . . . . .	69
64	B.3	$H \rightarrow \mu\tau_e$ . . . . .	70
65	B.4	$H \rightarrow e\tau_\mu$ . . . . .	70
66	C	BDT output distributions in control regions . . . . .	74
67	D	Tuning of the cut based analysis for the $H \rightarrow \mu\tau_h$ . . . . .	82

---

## 68 1 Introduction

69 A particle with a mass of 125 GeV and compatible with the standard model (SM) Higgs boson  
 70 (H) was discovered in 2012 by the ATLAS and CMS experiments at the CERN Large Hadron  
 71 Collider (LHC) [1–3]. The subsequent studies of its properties and decays have all been consis-  
 72 tent with SM expectations. However, the possibility of beyond standard model (BSM) decays  
 73 of the H boson remains of considerable interest. A recent analysis of the combined results of  
 74 both CMS and ATLAS constrained  $\mathcal{B}(H \rightarrow BSM) < 0.34$  at the 95% confidence level [4]. This  
 75 allows for the possibility of unexpected or forbidden decays such as the lepton flavour violat-  
 76 ing (LFV) decays  $H \rightarrow e\mu$ ,  $H \rightarrow e\tau$ ,  $H \rightarrow \mu\tau$ . Although forbidden in the SM [5], they can occur  
 77 naturally in models with more than one Higgs boson doublet [6]. They also arise in supersym-  
 78 metric models [7–9], composite Higgs models [10, 11], models with flavour symmetries [12],  
 79 Randall-Sundrum models [13–15], and others [16–23].

80 Searches for LFV Higgs boson decays have been performed at the LHC with data collected in  
 81 Run-1 and in Run-2. Most of the authors of the present note were also responsible for the earlier  
 82 CMS searches. Our first CMS search was for  $H \rightarrow \mu\tau$ . It is documented in analysis note AN-13-  
 83 244 and paper HIG-14-005. We observed an excess of data with respect to the SM background-  
 84 only hypothesis at  $m_H = 125$  GeV with a significance of 2.4 standard deviations in  $19.7 \text{ fb}^{-1}$   
 85 of pp data collected at a center-of-mass energy of 8 TeV in 2012 [24]. This corresponds to a  
 86 constraint on the branching fraction  $\mathcal{B}(H \rightarrow \mu\tau)$  less than 1.51% at 95% confidence level (CL),  
 87 while the best fit branching fraction is  $\mathcal{B}(H \rightarrow \mu\tau) = 0.84 \pm 0.39\%$ . The ATLAS Collaboration  
 88 also searched for  $H \rightarrow \mu\tau$  in the Run-1 dataset using a different techniques. They found a less  
 89 significant deviation from the background expectation of  $1.3\sigma$  significance and set an upper  
 90 limit of  $\mathcal{B}(H \rightarrow \mu\tau) < 1.85\%$  at 95% CL with a best fit branching fraction of  $\mathcal{B}(H \rightarrow \mu\tau) =$   
 91  $(0.77 \pm 0.62)\%$  [25].

92 We have also searched for  $H \rightarrow e\tau$  and  $H \rightarrow e\mu$  in the same Run 1 CMS data set. These searches  
 93 are documented in Analysis notes AN-14-232 and AN-14-231 respectively and paper HIG-14-  
 94 040. No evidence was found for lepton flavour violating decays in either final state. Upper  
 95 limits on the branching fractions,  $\mathcal{B}(H \rightarrow e\tau) < 0.69\%$  and  $\mathcal{B}(H \rightarrow e\mu) < 0.035\%$ , were set at  
 96 the 95% confidence level. The ATLAS collaboration recently reported searches for  $H \rightarrow e\tau$  and  
 97  $H \rightarrow \mu\tau$ , finding no significant excess of events over the background expectation. The searches  
 98 in channels with leptonic tau decays are sensitive only to a difference between  $\mathcal{B}(H \rightarrow e\tau)$  and  
 99  $\mathcal{B}(H \rightarrow \mu\tau)$ . These are combined with the searches in channels with hadronic tau decays to set  
 100 limits of  $\mathcal{B}(H \rightarrow e\tau) < 1.04\%$ ,  $\mathcal{B}(H \rightarrow \mu\tau) < 1.43\%$  at 95% CL [26].

101 The limits from all the direct searches described above are used to constrain the corresponding  
 102 flavour violating Yukawa couplings:  $Y_{e\mu}$ ,  $Y_{e\tau}$  and  $Y_{\mu\tau}$ . There are also indirect constraints com-  
 103plementary to the direct searches. The presence of LFV Higgs boson couplings allows  $\mu \rightarrow e$ ,  
 104  $\tau \rightarrow \mu$ , and  $\tau \rightarrow e$  to proceed via a virtual Higgs boson [27, 28]. The experimental lim-  
 105 its on these decays have been translated into constraints on  $\mathcal{B}(H \rightarrow e\mu)$ ,  $\mathcal{B}(H \rightarrow \mu\tau)$  and  
 106  $\mathcal{B}(H \rightarrow e\tau)$  [5, 29]. The null result for  $\mu \rightarrow e\gamma$  [30] strongly constrains  $\mathcal{B}(H \rightarrow e\tau_\mu) < \mathcal{O}(10^{-8})$ .  
 107 However, the constraint  $\mathcal{B}(H \rightarrow e\tau) < \mathcal{O}(10\%)$  is much less stringent. This comes from  
 108 searches for rare  $\tau$  decays [31] such as  $\tau \rightarrow e\gamma$ , and the measurement of the electron mag-  
 109 netic moment. Exclusion limits on the electric dipole moment of the electron [32] also provide  
 110 complementary constraints.

111 This note describes the Run 2 CMS searches for  $H \rightarrow e\tau$  and  $H \rightarrow \mu\tau$ . The Run 1  $H \rightarrow e\mu$  search  
 112 was four orders of magnitude less sensitive than the corresponding indirect limits, whereas  
 113 the  $H \rightarrow e\tau$  and  $H \rightarrow \mu\tau$  searches were an order of magnitude improvement over the indirect  
 114 limits. The  $H \rightarrow e\mu$  search is therefore lower priority and will be updated later. A brief recap

of the The Run 1 analysis is described below followed by a summary of the differences in the present Run 2 analysis.

### 1.1 Recap of Run 1 LFV H Analysis technique

The  $H \rightarrow e\tau$  and  $H \rightarrow \mu\tau$  searches were cut based analysis. Each search was performed in two decay channels according to whether the tau lepton decays electronically ( $\tau_\mu/\tau_e$ ) or hadronically ( $\tau_h$ ). The decay channels,  $H \rightarrow e\tau_e$ , and  $H \rightarrow \mu\tau_\mu$  are not considered due to the large background contribution from  $Z \rightarrow ee$  and  $Z \rightarrow \mu\mu$  decays respectively.

The main observable used to discriminate between the signal and the background is the collinear mass,  $M_{\text{col}}$ , which provides an estimate of  $M_H$  using the observed decay products. It is constructed using the collinear approximation based on the observation that, since  $M_H \gg M_\tau$ , the  $\tau$  decay products are highly Lorentz boosted in the direction of the  $\tau$  [33]. The neutrino momenta can be approximated to have the same direction as the other visible decay products of the  $\tau$  ( $\tau^{\text{vis}}$ ) and the component of the  $\vec{E}_T^{\text{miss}}$  in the direction of the visible  $\tau$  decay products is used to estimate the transverse component of the neutrino momentum ( $p_T^{\nu, \text{est}}$ ). The collinear mass can then be derived from the visible mass of the  $\tau$ - $l$  system ( $M_{\text{vis}}$ ) as  $M_{\text{col}} = M_{\text{vis}} / \sqrt{x_\tau^{\text{vis}}}$ , where  $x_\tau^{\text{vis}}$  is the fraction of energy carried by the visible decay products of the  $\tau$  ( $x_\tau^{\text{vis}} = p_T^{\tau^{\text{vis}}} / (p_T^{\tau^{\text{vis}}} + p_T^{\nu, \text{est}})$ ).

The primary backgrounds are from  $Z \rightarrow \tau\tau$  decays and mis-identified leptons in  $W + \text{jets}$ . The  $Z \rightarrow \tau\tau$  decays were estimated with a  $Z \rightarrow \mu\mu$  data sample with the muons replaced with simulated tau lepton decays. The misidentified lepton background was estimated from data by defining a sample with the same selection as the signal sample, but inverting the isolation requirements on one of the leptons, to enrich the contribution from  $W + \text{jets}$  and QCD multijets. The probability for PF objects to be misidentified as leptons is measured using an independent collision data set and this probability was applied to the background enriched sample to compute the misidentified lepton background in the signal sample. Smaller backgrounds such as  $t\bar{t}$  were estimated with simulated samples.

A loose pre-selection was first made to give a high statistics sample to validate the background estimation. This selection required two oppositely charged isolated fiducial leptons with loose  $p_T$  requirements. The events are then divided into categories within each channel according to the number of jets in the event. Jets are required to pass identification criteria, have  $p_T > 30 \text{ GeV}$ , and lie in the region  $|\eta| < 4.7$ . The 0-jet and 1-jet categories contain events primarily produced by gluon fusion (GF). The 2-jet category is defined to enrich the contribution from events produced via the vector boson fusion (VBF) process. The background estimation is validated in all the kinematic variables that are subsequently used to suppress the background. The analysis is subsequently performed blinded by using a fixed selection and checking the agreement between relevant observed and simulated distributions outside the sensitive region  $100 \text{ GeV} < M_{\text{col}} < 150 \text{ GeV}$ .

Next, a set of kinematic variables is defined, and the event selection criteria are set to maximise the significance  $S/\sqrt{S+B}$ , where  $S$  and  $B$  are the expected signal and background event yields in the mass window  $100 \text{ GeV} < M_{\text{col}} < 150 \text{ GeV}$ . The variables used are: the lepton transverse momenta  $p_T^\ell$  with  $\ell = e, \mu, \tau_h$ ; azimuthal angles between the leptons  $\Delta\phi_{\vec{p}_T^{\ell_1}-\vec{p}_T^{\ell_2}}$ ; azimuthal angle between the lepton and the  $E_T^{\text{miss}}$  vector  $\Delta\phi_{\vec{p}_T^\ell-\vec{E}_T^{\text{miss}}}$ ; the transverse mass  $M_T^\ell = \sqrt{2p_T^\ell E_T^{\text{miss}}(1 - \cos \Delta\phi_{\vec{p}_T^\ell-\vec{E}_T^{\text{miss}}})}$ .

After the full selection, a binned likelihood is used to fit the distributions of  $M_{\text{col}}$  for the signal

159 and the background contributions. The modified-frequentist  $CL_s$  method [34, 35] is used to set  
160 upper bounds on the signal strength  $\mu$ , or determine a signal significance.

## 161 1.2 Overview of Run 2 Analysis

162 The Run 2 analysis is a refinement of the Run-1 analysis technique with the following differ-  
163 ences

- 164 • The 2-jets category is now divided into a GG and VBF category. There are now four  
165 categories for each of the two channels instead of three
- 166 • Additional kinematic variables that employ the angles between the two leptons are  
167 used.
- 168 • The selection criteria are set by determining the fitted expected limit and significance  
169 of each selection including all systematic errors
- 170 • A multivariate combination of the kinematic variables using a boosted decision tree  
171 (BDT) has been employed. It provides a significantly better sensitivity than a re-  
172 optimized cut based analysis and is the baseline analysis.
- 173 • The cut based analysis has been retained as cross-check analysis. It has been re-  
174 optimized as described above but with an additional constraint. It was found that  
175 the tight lepton  $p_T$  cuts used in the Run 1 analysis caused some of the backgrounds  
176 to peak under the signal which was criticized as an undesirable feature. Hence these  
177 criteria are relaxed. The subsequent loss of sensitivity is partially offset by the new  
178 kinematic variables and the re-optimization. However, additional sensitivity in the  
179 cut based analysis could be gained by restoring the tight lepton  $p_T$  cuts at the cost of  
180 peaking backgrounds.
- 181 • The  $Z \rightarrow \tau\tau$  background is estimated with simulation alone rather than the embed-  
182 ded samples.
- 183 • The method to estimate of the misidentified lepton backgrounds in the  $\tau_e$  and  $\tau_\mu$   
184 channels is different. We now use a semi-data driven method. In the  $\tau_h$  channels we  
185 use the same fully data driven methods as in Run1. Both methods are compared in  
186 the leptonic decay channels and the semi-data driven technique is chosen because  
187 for greater statistical precision.

## 188 2 Data and simulation samples

189 The data studied in this analysis corresponds to  $36 \text{ fb}^{-1}$  collected with single electron, single  
 190 muon, or muon+electron triggers, in 2016. The analysis will be topped-up with the full inte-  
 191 grated luminosity collected in 2016 by the CMS detector before approval, and re-reco datasets  
 192 will then be used. The datasets used are given in Tab. 1.

Dataset	Run range	Luminosity
/MuonEG/Run2016B-PromptReco-v2/MINIAOD	272007–275376	5.892 /fb
/MuonEG/Run2016C-PromptReco-v2/MINIAOD	275657–276283	2.646 /fb
/MuonEG/Run2016D-PromptReco-v2/MINIAOD	276315–276811	4.353 /fb
/MuonEG/Run2016E-PromptReco-v2/MINIAOD	276831–277420	4.117 /fb
/MuonEG/Run2016F-PromptReco-v2/MINIAOD	277932–278808	3.174 /fb
/MuonEG/Run2016G-PromptReco-v2/MINIAOD	278820–280385	7.540 /fb
/MuonEG/Run2016H-PromptReco-v2/MINIAOD	281613–284044	8.606 /fb
/SingleMuon/Run2016B-PromptReco-v2/MINIAOD	272007–275376	5.892 /fb
/SingleMuon/Run2016C-PromptReco-v2/MINIAOD	275657–276283	2.646 /fb
/SingleMuon/Run2016D-PromptReco-v2/MINIAOD	276315–276811	4.353 /fb
/SingleMuon/Run2016E-PromptReco-v2/MINIAOD	276831–277420	4.117 /fb
/SingleMuon/Run2016F-PromptReco-v2/MINIAOD	277932–278808	3.174 /fb
/SingleMuon/Run2016G-PromptReco-v2/MINIAOD	278820–280385	7.540 /fb
/SingleMuon/Run2016H-PromptReco-v2/MINIAOD	281613–284044	8.606 /fb
/SingleElectron/Run2016B-PromptReco-v2/MINIAOD	272007–275376	5.892 /fb
/SingleElectron/Run2016C-PromptReco-v2/MINIAOD	275657–276283	2.646 /fb
/SingleElectron/Run2016D-PromptReco-v2/MINIAOD	276315–276811	4.353 /fb
/SingleElectron/Run2016E-PromptReco-v2/MINIAOD	276831–277420	4.117 /fb
/SingleElectron/Run2016F-PromptReco-v2/MINIAOD	277932–278808	3.174 /fb
/SingleElectron/Run2016G-PromptReco-v2/MINIAOD	278820–280385	7.540 /fb
/SingleElectron/Run2016H-PromptReco-v2/MINIAOD	281613–284044	8.606 /fb

Table 1: Datasets studied in the analysis.

193 The Monte Carlo (MC) simulations used in the analysis are given in Tab. 2. Z+jets and W+jets  
 194 are simulated at leading order (LO) with Madgraph, interfaced with Pythia8. Inclusive samples  
 195 are merged with exclusive samples corresponding to different generated jet multiplicities. The  
 196 stitching depends on the cross section of the various processes, on the number of generated  
 197 events for each sample, and on the number of generated jets. A k-factor of 1.16 is used for the  
 198 Z+jets samples, while 1.21 is used for W+jets samples. The  $t\bar{t}$  and single top quark samples, as  
 199 well as the SM Higgs boson productions, are generated with Powheg.

200 Signal events are generated with Powheg, and the corresponding samples are listed in Tab. 3.

201 All simulations are passed through Geant4 to model the interactions with the detector, and are  
 202 reconstructed with the same CMSSW version as used for the data.

203 Due to the high luminosities attained during data-taking, many events have multiple proton-  
 204 proton interactions per bunch crossing (pileup). All simulated samples are reweighted to match  
 205 the pileup distribution observed in data. An event weight is applied based on the number of  
 206 simulated pileup events and the instantaneous luminosity per bunch-crossing, averaged over  
 207 the run period.

MC simulations	Cross section
/DYJetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	4954.0 pb (LO)
/DY1JetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	1012.5 pb (LO)
/DY2JetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	332.8 pb (LO)
/DY3JetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	101.8 pb (LO)
/DY4JetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	54.8 pb (LO)
/DYJetsToLL_M-10to50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	18610 pb (LO)
/TT_TuneCUETP8M1_13TeV-powheg-pythia8	831.76 pb
/WJetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	50380 pb (LO)
/W1JetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	9644.5 pb (LO)
/W2JetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	3144.5 pb (LO)
/W3JetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	954.8 pb (LO)
/W4JetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/	485.6 pb (LO)
/WZTo1L3Nu_13TeV_amcatnloFXFX_madspin_pythia8	3.05 pb
/WZTo1L1Nu2Q_13TeV_amcatnloFXFX_madspin_pythia8	10.71 pb
/WZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8	5.595 pb
/ST_tW_antitop_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1	35.85 pb
/ST_tW_top_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1	35.85 pb
/ST_t-channel_antitop_4f_leptonDecays_13TeV-powheg-pythia8_TuneCUETP8M1	80.95 pb
/ST_t-channel_top_4f_leptonDecays_13TeV-powheg-pythia8_TuneCUETP8M1	136.02 pb
/WWTo1L1Nu2Q_13TeV_amcatnloFXFX_madspin_pythia8	1.212 pb
/ZZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8	3.22 pb
/VVTo2L2Nu_13TeV_amcatnloFXFX_madspin_pythia8	11.95 pb
/WGToLNuG_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8	489.0 pb
/WGstarToLNuMuMu_012Jets_13TeV-madgraph	2.793 pb
/WGstarToLNuEE_012Jets_13TeV-madgraph	3.526 pb
/GluGluHToTauTau_M125_13TeV_powheg_pythia8	44.14 pb X 0.0627
/VBFHToTauTau_M125_13TeV_powheg_pythia8	3.782 pb X 0.0698

Table 2: Monte Carlo samples used in the search, together with their respective cross sections. The k-factors for the W+jets and Z+jets samples are 1.21 and 1.16, respectively. The NNLO+NNLL QCD cross sections for the Higgs boson samples in gluon-gluon fusion production mode are considered in this version of the analysis, and will be replaced by N3LO cross sections before the approval.

MC simulations	Cross section
/GluGlu_LFV_HToMuTau_M125_13TeV_powheg_pythia8	44.14 pb
/VBF_LFV_HToMuTau_M125_13TeV_powheg_pythia8	3.782 pb
/GluGlu_LFV_HToETau_M125_13TeV_powheg_pythia8	44.14 pb
/VBF_LFV_HToETau_M125_13TeV_powheg_pythia8	3.782 pb

Table 3: Simulated signal samples used in the analysis.

### 208 3 Event reconstruction

209 Individual particles are reconstructed with a particle-flow (PF) algorithm [36, 37], which relies  
 210 on the information from the various CMS subdetectors. More complicated objects, such as jets,  
 211 hadronically-decaying taus, or transverse missing energy, are reconstructed from a combina-  
 212 tion of these individual particles.

213 Hadronically decaying taus,  $\tau_h$ , are reconstructed with the Hadrons-Plus-Strips (HPS) algo-  
 214 rithm [38, 39]. The HPS algorithm combines tracks and energy deposits in strips of the ECAL  
 215 to build taus in the 1-prong, 1-prong +  $\pi^0$ s, and 3-prong decay modes. Identification and  
 216 isolation conditions are enforced by requiring the  $\tau_h$  candidates to pass some working point  
 217 of an MVA discriminator that takes as input various isolation variables and variables related  
 218 to  $\tau$  lifetime information. The tight working point is used as it brings the best sensitivity to  
 219 the analysis. Additionally, electrons and muons faking  $\tau_h$  candidates are rejected by using an  
 220 MVA-based or a cut-based discriminator, respectively. The exact working points depend on the  
 221 final state studied and on the background composition. The energy scale of genuine hadronic  
 222 taus is corrected based on the reconstructed decay mode, as recommended by the TAU POG:  
 223 -1.8% for 1-prong, +1.0% for 1prong+ $\pi^0$ s, and +0.4% for 3-prong. The energy scale of electrons  
 224 faking hadronic taus is corrected by +10% in the 1prong+ $\pi^0$ s decay mode, and is not corrected  
 225 for other decay modes. The correction is obtained by taking the correction that corresponds to  
 226 an optimum in the negative loglikelihood function curves obtained by fitting templates with  
 227 various corrections in the region  $70 < m_{vis} < 110$  GeV. The energy scale of muons faking  
 228 hadronic taus is corrected by +1.5% in the 1prong+ $\pi^0$ s decay mode, and -0.2% in the aprong  
 229 decay mode.

230 The isolation of muons (electrons) is measured relative to their transverse momentum  $p_T^\ell$ , by  
 231 summing over the transverse momenta of PF particles in a cone with  $\Delta R < 0.4$  (0.3) around the  
 232 lepton:

$$R_{\text{Iso}}^\ell = \left( \sum p_T^{\text{charged}} + \max \left[ 0, \sum p_T^{\text{neutral}} + \sum p_T^\gamma - p_T^{\text{PU}}(\ell) \right] \right) / p_T^\ell. \quad (1)$$

233 Muons are required to pass the medium identification tuned for the ICHEP dataset ("Medium2016").  
 234 Electrons should pass one of the working points of the non triggering MVA identification.

235 The  $\sum p_T^{\text{charged}}$  is the scalar sum of the transverse momenta of charged hadrons originating from  
 236 the primary vertex. The  $\sum p_T^{\text{neutral}}$  and  $\sum p_T^\gamma$  are the scalar sums of the transverse momenta for  
 237 neutral hadrons and photons, respectively. The neutral contribution to isolation from pileup,  
 238  $p_T^{\text{PU}}(\ell)$ , is estimated using a *jet area* method [40] for the electron. For the muon, the  $p_T^{\text{PU}}(\ell)$  con-  
 239 tribution is estimated using the sum of transverse momenta of charged hadrons not originating  
 240 from the primary vertex scaled by a factor of 0.5.

241 Jets are reconstructed by clustering all PF candidates using the anti- $k_T$  algorithm with distance  
 242 parameter  $\Delta R = 0.4$  [41]. Charged hadrons that are not originating from the primary vertex are  
 243 subtracted from the PF candidates (hadron subtraction). Jets are only considered if they pass  
 244 the loose working point of the PF jet ID discriminator provided by the JETMET POG [42] and  
 245 if they have a corrected transverse momentum above 30 GeV and  $|\eta| < 4.7$ . Additionally, they  
 246 have to be separated from every lepton selected in the analysis by  $\Delta R > 0.5$ .

247 The missing transverse energy is obtained from all PF constituents reconstructed in each event.  
 248 Type-1 corrections are applied.

## 249 4 Background estimation

250 The signal is a prompt isolated lepton,  $e$  or  $\mu$ , along with an oppositely charged isolated lepton  
 251 of different flavour ( $\tau_\mu$ ,  $\tau_e$  or  $\tau_h$ ) where  $\tau_\mu$ ,  $\tau_e$  or  $\tau_h$  denote a tau lepton decaying to a muon, elec-  
 252 tron or hadrons. The neutrino in the tau lepton decay leads to the presence of significant miss-  
 253 ing energy. The largest backgrounds to this signature are from  $Z \rightarrow \tau\tau$  decays and from  $W+jets$   
 254 and QCD multijet production. In the latter, PF objects (predominantly jets), are misidentified  
 255 as leptons. These backgrounds are estimated with data driven methods. Smaller backgrounds  
 256 are estimated with simulation. These include  $t\bar{t}$  pairs, SM Higgs boson production ( $H \rightarrow \tau\tau$ ),  
 257  $WW$ ,  $WZ$ ,  $ZZ + jets$ ,  $W\gamma^{(*)} + jets$  processes, and single top quark production. The background  
 258 estimation techniques are described in detail below. To validate the background estimation we  
 259 define a loose selection for each channel (described in following sections) and compare the dis-  
 260 tributions of the kinematic variables with data. We also validate with control regions than are  
 261 enhanced with the dominant backgrounds.

### 262 4.1 $Z \rightarrow \tau\tau$

263 The  $Z \rightarrow \tau\tau$  background is currently estimated from simulation. The  $m_{\ell\ell}$  distribution in the  
 264  $Z + jets$  samples is found to be different in data and simulation. To correct for this a reweighting  
 265 is applied to the samples to the generator-level  $p_T$  and generator-level  $m_{\ell\ell}$  distributions  
 266 in the LO Madgraph samples. The generator-level  $p_T$  is found by summing all generator par-  
 267 ticles that satisfy the `(fromHardProcessFinalState && (isMuon || isElectron ||`  
 268 `isNeutrino)) || isDirectHardProcessTauDecayProduct` requirements. The visible  
 269 generator-level  $p_T$  is found the same way, but the `isNeutrino` requirement is removed. The  
 270 weights were originally produced for the MSSM  $\Phi \rightarrow \tau\tau$  analysis [43], and more details can be  
 271 found in [4].

### 272 4.2 $Z \rightarrow \ell\ell$

273 The  $Z \rightarrow \ell\ell$  background is also estimated from simulation. Corrections for  $e \rightarrow \tau_h$  and  $\mu \rightarrow \tau_h$   
 274 fake rates are applied when the reconstructed  $\tau_h$  candidate is matched to an electron or a muon,  
 275 respectively, at the generator level. They are measured by the TAU POG in Drell-Yan events.  
 276 The correction factors depend on the pseudorapidity of the lepton and are given in Tab. 4.

Table 4: Efficiency scale factors for the discriminators used to reject prompt leptons in the tau identification process.

Tau discriminator	$\eta$ range	Scale factor
againstElectronMVAVLoose	$ \eta  < 1.460$	1.292
againstElectronMVATight	$ \eta  < 1.460$	1.505
againstElectronMVAVLoose	$ \eta  > 1.558$	1.536
againstElectronMVATight	$ \eta  > 1.558$	1.994
againstMuonLoose3	$ \eta  < 1.2$	$1.14 \pm 0.03$
againstMuonTight3	$ \eta  < 1.2$	$1.28 \pm 0.06$
againstMuonLoose3	$1.2 <  \eta  < 1.7$	$1.2 \pm 0.2$
againstMuonTight3	$1.2 <  \eta  < 1.7$	$2.6 \pm 2.6$
againstMuonLoose3	$1.7 <  \eta  < 2.3$	$1.3 \pm 0.3$
againstMuonTight3	$1.7 <  \eta  < 2.3$	$2.1 \pm 0.9$

<sup>277</sup> **4.3  $t\bar{t}$ +jets**

<sup>278</sup> The  $t\bar{t}$ +jets background is particularly important in the  $e\mu$  final state. It is estimated with sim-  
<sup>279</sup> ulation. A correction based on the generated  $p_T$  of the top and antitop quarks is applied to  
<sup>280</sup> match the  $p_T$  distribution observed in data, as described in Section ???. The weights have been  
<sup>281</sup> measured with Run-2 data [44].

<sup>282</sup> **4.4 Misidentified Leptons**

<sup>283</sup> Jets misidentified as leptons are a source of background. They arise from two sources  $W + jets$   
<sup>284</sup> and QCD multijet events. In  $W + jets$  events one lepton candidate is a real lepton from the  $W$   
<sup>285</sup> decay and the other is a jet misidentified as a lepton. In QCD multijet events both lepton can-  
<sup>286</sup> didates are misidentified jets. In Run 1 analysis we used a fully data driven technique, referred  
<sup>287</sup> to as the “fake rate” method to estimate the misidentified background. This technique is fully  
<sup>288</sup> data driven and estimates the misidentified background from the two sources combined. In  
<sup>289</sup> each of the four channels for this analysis ( $\mu\tau_h, e\tau_h, \mu\tau_e, e\tau_\mu$ ) we have estimated the misidentified  
<sup>290</sup> lepton background using the fake rate method. In the  $\mu\tau_e$  and  $e\tau_\mu$  channels we have also esti-  
<sup>291</sup> mated this background using a semi-data driven technique used previously in the SM  $H \rightarrow \tau\tau$   
<sup>292</sup> analysis in which the misidentified  $W + jets$  background is estimated using simulation and the  
<sup>293</sup> QCD background using data. We have checked that the two techniques are consistent with the  
<sup>294</sup> loose selection sample. However we use the semi-data driven technique with the full selection  
<sup>295</sup> in these channels as the fully data driven technique suffers from lack of statistics.

<sup>296</sup> **4.4.1 Fully data driven technique**

<sup>297</sup> The misidentified lepton background is estimated from collision data by defining a sample  
<sup>298</sup> with the same selection as the signal sample, but loosing the isolation requirements on one of  
<sup>299</sup> the leptons, to enrich the contribution from  $W + jets$  and QCD multijets. The probability for PF  
<sup>300</sup> objects to be misidentified as leptons is measured using an independent collision data set, de-  
<sup>301</sup> fined below, and this probability is applied to the background enriched sample to compute the  
<sup>302</sup> misidentified lepton background in the signal sample. The technique is shown schematically in  
<sup>303</sup> Table 5 in which four regions are defined including the signal (I) and background (III) enriched  
<sup>304</sup> regions and two control Regions (II & IV), defined with the same selections as Regions I & III  
<sup>305</sup> respectively, except with leptons of the same charge.

Table 5: Definition of the samples used to estimate the misidentified lepton ( $\ell$ ) background. They are defined by the charge of the two leptons and by the isolation requirements on each. The definition of not-isolated differs in each channel.

Region I	Region II
$\ell_1^\pm$ (isolated)	$\ell_1^\pm$ (isolated)
$\ell_2^\mp$ (isolated)	$\ell_2^\pm$ (isolated)
Region III	Region IV
$\ell_1^\pm$ (isolated)	$\ell_1^\pm$ (isolated)
$\ell_2^\mp$ (not-isolated)	$\ell_2^\pm$ (not-isolated)

<sup>306</sup> The rates with which jets are misidentified as electrons ( $f_e$ ), muons ( $f_\mu$ ), or  $\tau_h$  ( $f_{\mu_h}$ ), are esti-  
<sup>307</sup> mated in  $Z + jets$  events in data. This event sample is obtained by requiring a good  $Z$  boson  
<sup>308</sup> candidate plus one jet. The  $Z$  boson candidate is formed from two muons with  $p_T > 25$  GeV,  
<sup>309</sup>  $|\eta| < 2.4$ , passing the medium ID requirement, and with relative isolation less than 0.15. In  
<sup>310</sup> the case of jet  $\rightarrow e$  fake rate, the muon isolation is loosened to 0.25 to increase the statistics.  
<sup>311</sup> The muons are required to have an opposite sign charge and an invariant mass between 70 and

312 110 GeV. The fake rates are measured in both data and in  $Z + \text{jets}$  simulation. The contribution  
 313 from diboson events, where the jet candidate corresponds to a real lepton, is subtracted using  
 314 simulation.

The fake rates are then defined as

$$f_i = \frac{N_i(\text{regionI})}{N_i(\text{regionIII})}$$

315 where  $N_i$  is the number of events passing the requirements for the regions as defined below  
 316 and  $i = e, \mu, \tau_h$ .

### 317 Jets $\rightarrow \tau_h$ Estimation

318 In region III, the definition of not-isolated for  $\tau_h$  candidates is:

- 319 • Reconstructed as  $\tau_h$ , with decay mode finding;
- 320 • Very loose MVA  $\tau_h$  isolation.
- 321 • NOT pass the tight requirement for region I

322 In region I (signal region),  $\tau_h$  candidates are required to pass the tight MVA isolation. The  $\tau_h$   
 323 fake rates obtained from these requirements are given in Fig. 1. Fake rates show a  $p_T$  depen-  
 324 dence that depends on DM and  $\eta$ . The fakerates are applied as a function of  $p_T$  in different  
 325 decay modes and  $\eta$  region ( $|\eta| > 1.5$  or  $|\eta| < 1.5$  ).

### 326 Jets $\rightarrow e$ Estimation

327 In region III the definition of not-isolated for the electron candidates is:

- 328 • Identification WP80;
- 329 • Relative isolation less than 0.5. This isolation condition comes from the fact that the  
 330 electron isolation condition at trigger level for  $H \rightarrow e\tau_h$ .
- 331 • NOT pass the tight requirement for region I

322 In region I (signal region) the electrons are required to pass the WP80 identification and have a  
 323 relative isolation less than 0.1. The fake rate is given in Fig. 2 and shows a dependence on  $p_T$   
 324 and it can be considered as flat in  $|\eta|$ .

### 335 Jets $\rightarrow \mu$ Estimation

336 In region III the definition of not-isolated for the muon candidates is:

- 337 • Medium muon ID;
- 338 • Relative isolation less than 0.25.
- 339 • NOT pass the tight requirement for region I.

340 In the signal region, muon isolation is required to be less than 0.15. The fake rates for muons  
 341 are given in Fig. 3. No dependence on  $|\eta|$  is seen. The dependence of  $p_T$  is taken into account and  
 342 applying the fake rate in  $p_T$  bins.

The number of misidentified background events in the signal region  $N_i(\text{regionI})$  is then given by

$$N_i(\text{misidentified}) = \frac{f_i}{1 - f_i} N_i(\text{regionIII})$$

Double-counted events with two fake leptons are subtracted. For example events with a fake  
 muon (electron) and a fake  $\tau_h$  are subtracted in the  $H \rightarrow \mu\tau_h$  and  $H \rightarrow e\tau_h$  channels. This

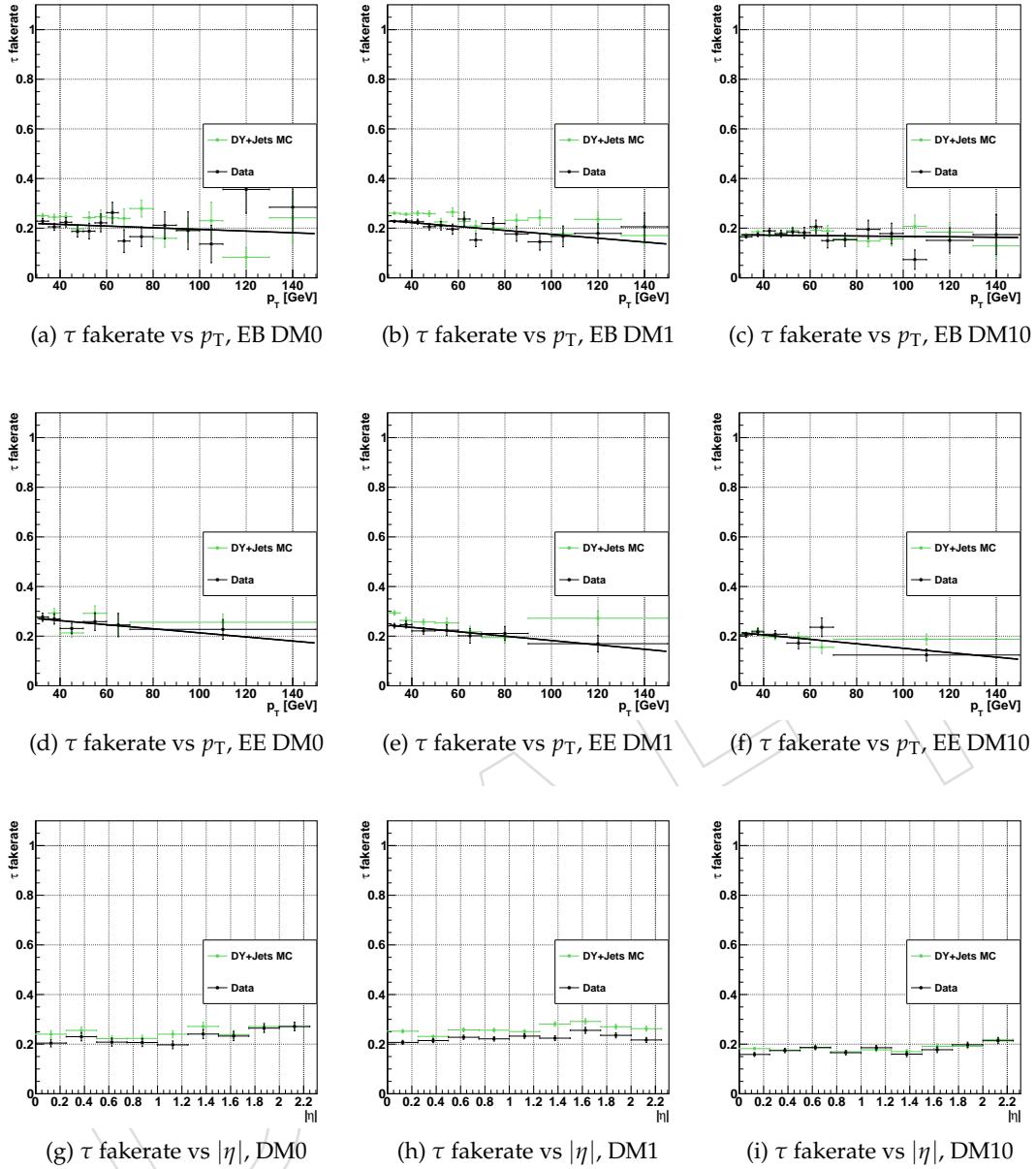


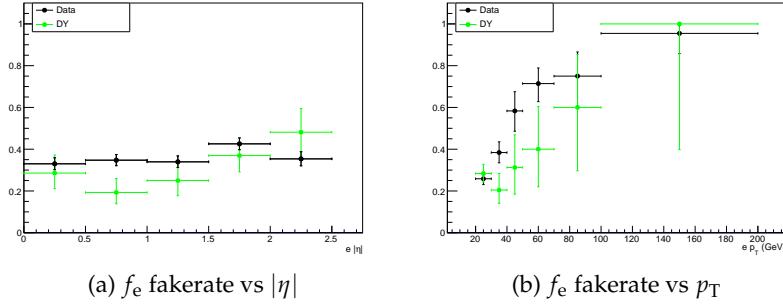
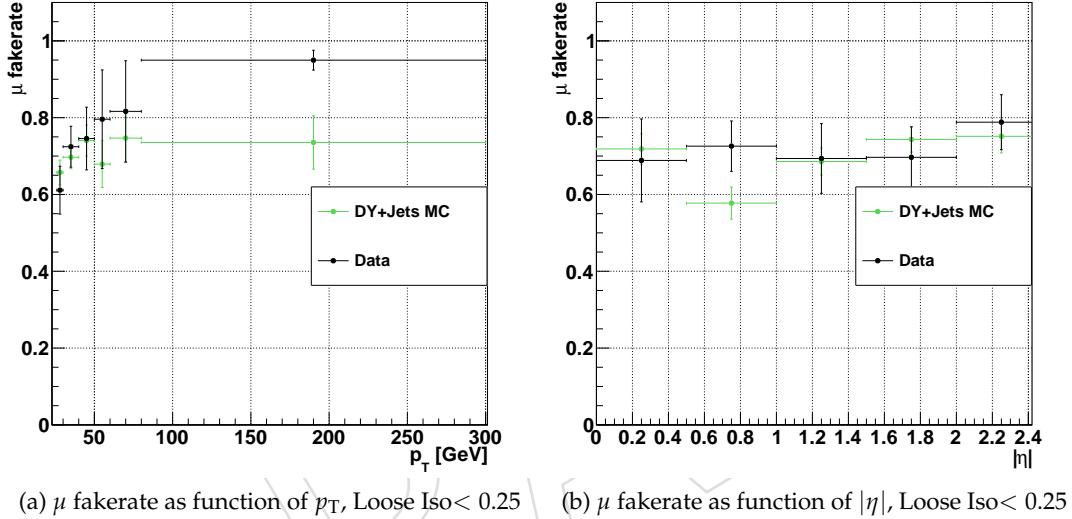
Figure 1:  $\tau$  fake rates as a function of  $p_T$  for different decay modes in EB and EE and  $|\eta|$  for different decay modes.

contribution is estimated as

$$N_j(\text{misidentified}) = N_j \frac{f_\tau \cdot f_j}{(1 - f_\tau) \cdot (1 - f_j) N_j} (\text{regionIII})$$

343 where  $j = e$  or  $\mu$ .

- 344 The same strategy is applied to estimate the mis-identified background for the  $H \rightarrow \mu\tau_e$  and  
 345  $H \rightarrow e\tau_\mu$  channels using the electron and muon fakerate. The contribution of Drell-Yan,  $t\bar{t}$ ,  
 346 and diboson events is estimated by applying the fake rate method to the corresponding MC  
 347 samples, and subtracting these contribution to the misidentified background.
- 348 The background estimation is validated in like sign events by following the same procedure

Figure 2: Electron fake rates as function of  $|\eta|$  and  $p_T$ .Figure 3: Muon fake rates as a function of  $p_T$  and  $|\eta|$ .

349 but with regions II and IV instead of I and III in the above. The background estimation is  
 350 also be validated in a W-enriched control region. The region is obtained by selecting opposite  
 351 sign events with  $m_T(\mu/e, \text{MET}) > 60$  GeV and  $m_T(\tau_h, \text{MET}) > 80$  GeV. Selecting opposite  
 352 sign events better reproduces the jet composition in the signal region. The validations for each  
 353 channel are given in sections 5.4 and 6.4.

#### 354 4.4.2 Semi Data Driven Technique

355 The W + jets contribution to the misidentified lepton background is estimated with simulation  
 356 for  $H \rightarrow e\tau_\mu$  and  $H \rightarrow \mu\tau_e$  decay channels. Simulated samples with different generated jet  
 357 multiplicities are combined to improve the statistical precision. The QCD multijet contribution  
 358 is estimated with like sign events that pass the signal region requirement. The expected yield  
 359 from non-QCD processes is subtracted using simulation. The resulting sample is then rescaled  
 360 by 2.3, to account for like-sign/opposite-sign differences, according to the studies performed  
 361 by the  $H \rightarrow \tau\tau$  working group. The studies can be found in AN-16-374 and PAS HIG-16-037  
 362 (MSSM  $H \rightarrow \tau\tau$ ) and AN-16-355 and PAS HIG-16-043 (SM  $H \rightarrow \tau\tau$ ).

<sup>363</sup> **4.5 Other backgrounds**

<sup>364</sup> The diboson and single top quark backgrounds, as well as the SM Higgs boson decays to tau  
<sup>365</sup> pairs, are estimated purely from MC simulations.

DRAFT

## 366 5 $H \rightarrow \mu\tau_h$ channel

### 367 5.1 Loose selection

368 In the  $H \rightarrow \mu\tau_h$  channel events are required to pass either of the isolated single muon  
 369 triggers ; HLT\_IsoMu24 or HLT\_IsoTkMu24. It is then required to have an isolated muon and an  
 370 isolated hadronic tau of opposite charge within the fiducial region and separated by  $\Delta R > 0.5$ .  
 371 Events with additional electrons, muons, or  $\tau_h$  candidates or least one b-jet are rejected vetoed.  
 372 The muon is required to have  $p_T > 25$  GeV,  $|\eta| < 2.4$ , and to pass the ICHEP Medium working  
 373 point selection criteria from the Muon POG. Hadronic taus are required to have  $p_T$  greater  
 374 than 30 GeV, and  $\eta < 2.3$ . They should pass the old decay mode finding, the tight WP of the  
 375 MVA isolation, the very loose MVA rejection against electrons, and the cut-based tight rejection  
 376 against muons.

377 The events are then divided into categories within each sample according to the number of  
 378 jets in the event. This is designed to enhance the contribution of different Higgs production  
 379 mechanisms. The four categories are:

380 **0-jet:** The events do not have any jet with  $p_T > 30$  GeV,  $|\eta| < 4.7$ , and pass the loose PF  
 381 ID. This category enhances the gluon-gluon fusion contribution.

382 **1-jet:** The events have one jet with  $p_T > 30$  GeV and  $|\eta| < 4.7$ , and pass the loose PF  
 383 ID. This category enhances the gluon-gluon fusion production with initial state ra-  
 384 diation (ISR). Some VBF events where one jet has escaped detection also enter this  
 385 category.

386 **2-jets GG:** The events have two jets with  $p_T > 30$  GeV and  $|\eta| < 4.7$ , and pass the loose  
 387 PF ID, with an invariant mass,  $M_{jj} < 550$  GeV. The dominant signal contribution  
 388 comes from gluon-gluon fusion Higgs boson production.

389 **2 jets VBF:** The events have two jets with  $p_T > 30$  GeV and  $|\eta| < 4.7$ , and pass the loose  
 390 PF ID, with an invariant mass,  $M_{jj} > 550$  GeV. The dominant signal contribution  
 391 comes from VBF Higgs production.

392 The threshold on  $M_{jj}$  has been optimized to give the best expected limits.

### 393 5.2 Cut-based analysis

394 After the loose selection a set of kinematic selection criteria are also required. The thresholds  
 395 are optimized to give the most stringent expected limits based on an Asimov dataset. The  
 396  $p_T$  of the  $\mu$  and  $\tau_h$  are constrained to be very loose to avoid making the backgrounds peak  
 397 under the signal in the  $M_{col}$  distribution. Signal events are characterized by a lower trans-  
 398 verse mass  $m_T(MET, \tau_h)$ , so we require events in 0-jet, 1-jet, 2-jet GG (VBF) categories to have  
 399  $m_T(MET, \tau_h) < 105(85)$  GeV. Requirements on  $m_T(MET, \mu)$ ,  $\Delta\phi(p_T^\mu, MET)$ , and  $\Delta\phi(p_T^\tau, p_T^\mu)$ ,  
 400 were investigated but were not found to give any improvement to the expected limits based  
 401 on an Asimov dataset. Note that the lepton  $p_T$  is a powerful background discriminant so that  
 402 tightening this requirement would enhance sensitivity but at the cost of peaking backgrounds.

### 403 5.3 Multivariate analysis

404 A Boosted decision tree (BDT) is exploited to provide greater sensitivity. It is trained after  
 405 the loose selection but combining all categories. The signal is a mixture of gg and VBF signal  
 406 events, weighted according to their respective production cross sections. The background is a  
 407 data sample of misidentified lepton events. This is the dominant background in this channel.  
 408 The leptons are required to be like sign and be anti-isolated in order to be orthogonal to the

409 signal region, yet have the same kinematic properties. Two options for the BDT analysis were  
 410 studied in previous iterations of this analysis, called Method A and Method B. It was found that  
 411 Method B gives significantly improved sensitivity to either the cut based analysis or Method A,  
 412 maintaining good signal background shape discrimination. As a result, only the BDT Method  
 413 B, relabelled simply "BDT" has been continued for the final results with  $36\text{fb}^{-1}$ . Details of the  
 414 differences between BDT Method A and B can be found in Appendix B.

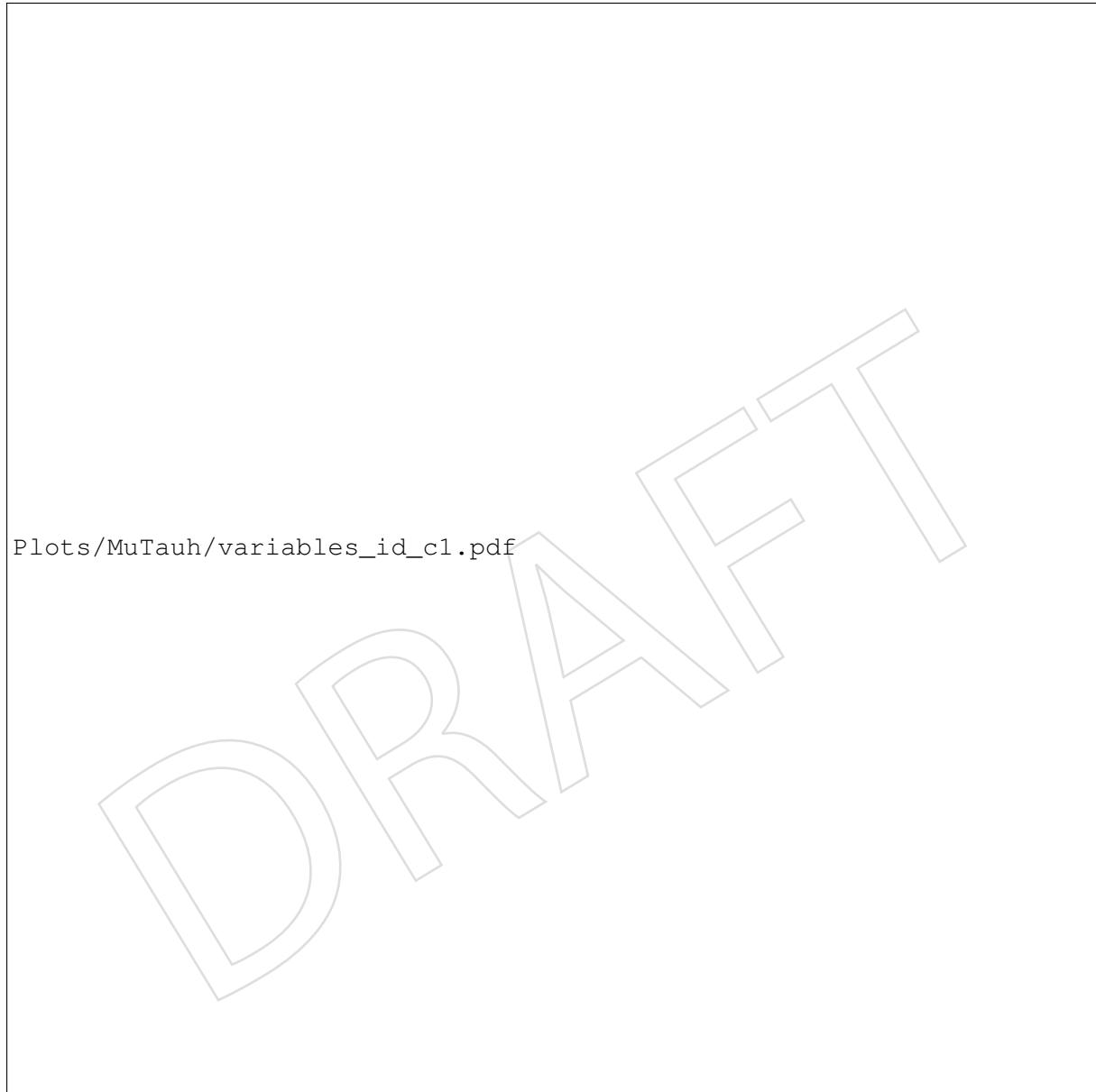


Figure 4:  $H \rightarrow \mu\tau_h$ : BDT input variable distributions. Signal distributions are drawn in blue and the background distributions in red.

415 The input variables to the BDT are:

- 416 • Transverse mass between the  $\tau_h$  and  $E_T^{\text{miss}}$   $M_T(\tau_h, E_T^{\text{miss}})$ .  
 417 • Missing transverse energy,  $E_T^{\text{miss}}$ .  
 418 • Pseudorapidity difference between the muon and the  $\tau_h$  candidate,  $\Delta\eta(e, \tau_h)$ .

- Azimuthal angle between the muon and the  $\tau_h$ ,  $\Delta\phi(\mu, \tau_h)$ .
- Azimuthal angle between the  $\tau_h$  and the  $E_T^{\text{miss}}$ ,  $\Delta\phi(\tau_h, E_T^{\text{miss}})$ .
- Collinear mass,  $M_{\text{col}}$ .
- Muon  $p_T$ .
- $\tau_h p_T$ .

No threshold requirement is made on the BDT output: the expected limits are estimated by fitting the BDT output distributions in the four jet categories. Fig. 5 shows the correlations between the input variables.

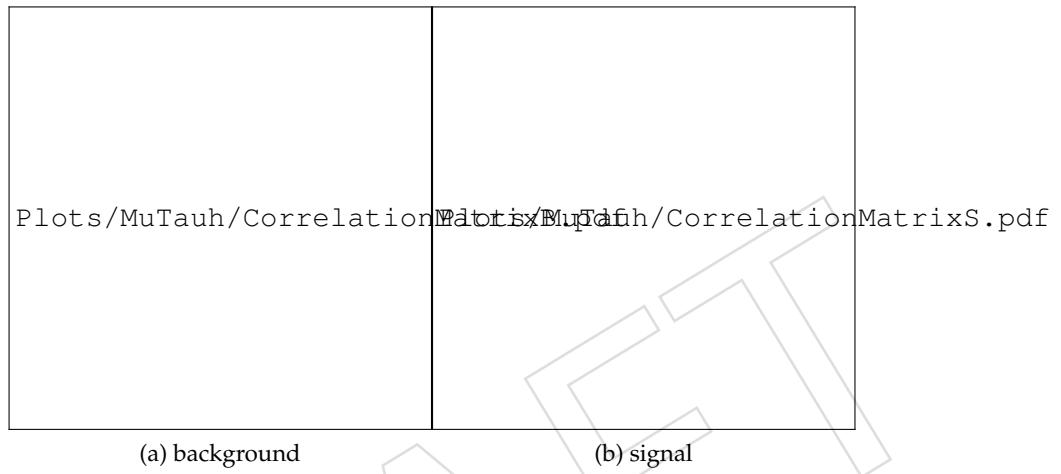


Figure 5:  $H \rightarrow \mu\tau_h$ : BDT input variable correlations for background (a) and signal (b).



Figure 6:  $H \rightarrow \mu\tau_h$ : BDT response.

The BDT response distributions for the signal and background are shown in Fig. 6. The data to simulation distributions of the BDT outputs are shown in Section 10, Fig. 40. There is good agreement between observed and predicted backgrounds in the unblinded regions.

#### 430 5.4 Validation of the background methods

431 The misidentified lepton backgrounds are estimated with the full data driven method de-  
 432 scribed in Sec 4. The  $\tau_h$  and/or  $\mu$ candidate isolation requirements are relaxed to enhance  
 433 statistics. This background is validated with a like-sign control region. Figure 7 shows the  
 434 data compared to the background estimation in the like sign control region. This sample is  
 435 dominated by the misidentified lepton background component and the agreement is good. A  
 436 30% uncertainty is assigned to the misidentified lepton background estimation. This is shown  
 437 in the figure. The uncertainty accounts for uncertainties in the fits of the fake rate functions,  
 438 and for differences of background composition or jet flavour. . It also covers the possible data-  
 background difference observed in the control region. As the jet composition of the like-sign

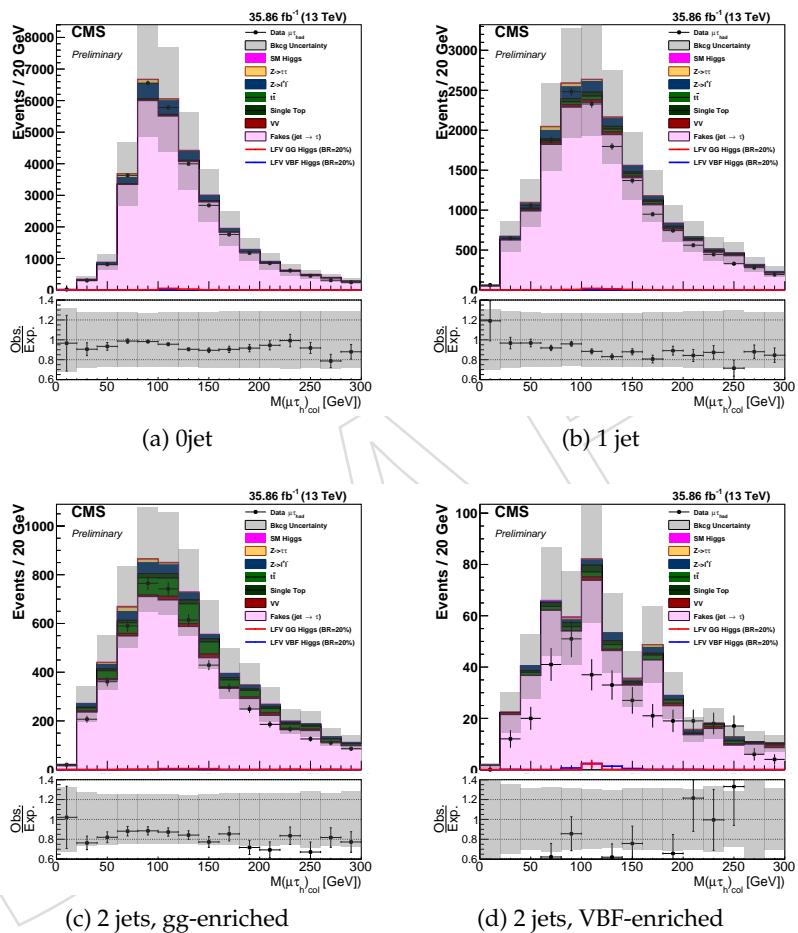


Figure 7:  $H \rightarrow \mu\tau_h:M_{col}$  in the four categories of the  $\mu\tau_h$  analysis, after the loose selection for like-sign pairs. The plots are before fitting, and both statistical and systematics uncertainties are shown.

439 control region is different with respect to the signal region because of the like-charge require-  
 440 ment, the background estimate is cross-checked using another  $W +$  jets enriched control region  
 441 obtained selecting events having the transverse mass between the muon and the missing trans-  
 442 verse energy to be greater than 60 GeV and the transverse mass between the  $\tau_h$  and the trans-  
 443 verse missing energy greater than 80 GeV. Figure 8 shows the data - background comparison  
 444 for this control region.

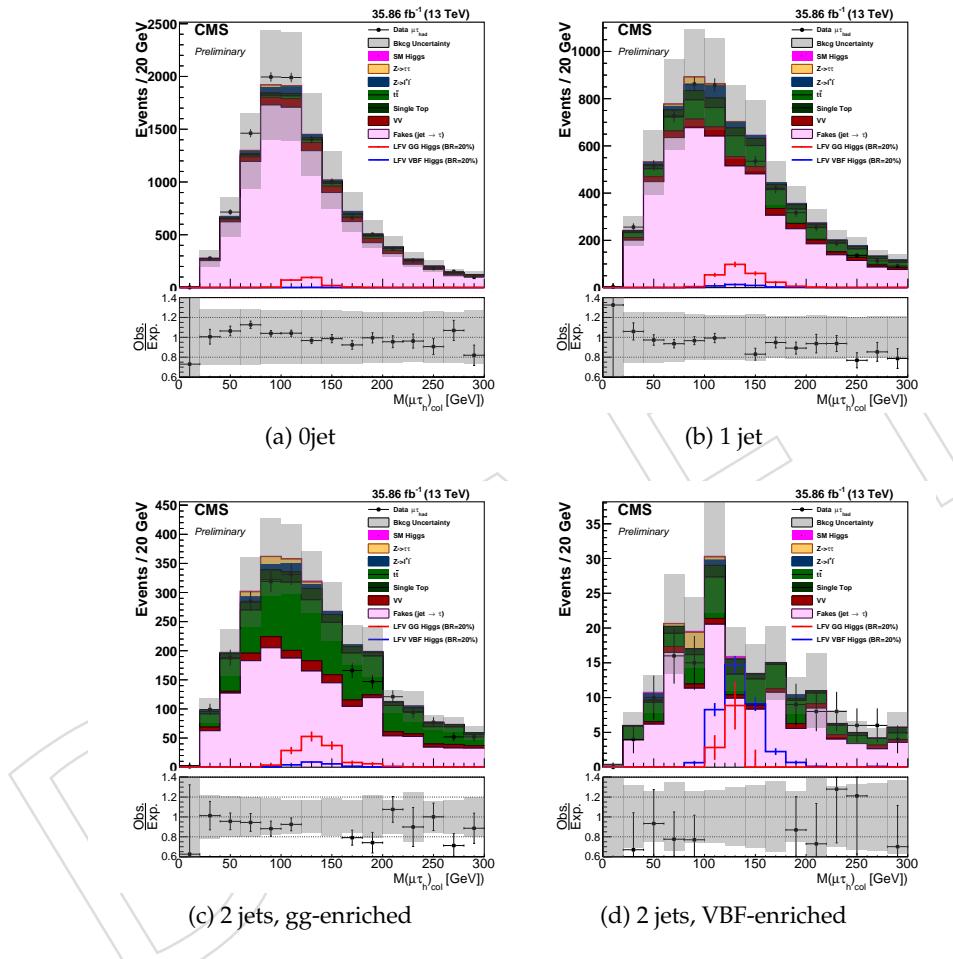


Figure 8:  $H \rightarrow \mu\tau_h$ : Collinear mass in the four categories of the  $\mu\tau_h$  analysis, after the loose selection for the W-enriched control region. The plots are prefit, and both statistical and systematic uncertainties are shown.

## 446 6 $H \rightarrow e\tau_h$ channel

### 447 6.1 Loose selection

448 In the  $H \rightarrow e\tau_h$  channel events are required to pass the single electron trigger HLT\_Ele25WPTight\_eta2p1.  
 449 The electron candidates are then required to have  $p_T > 26$  GeV and  $|\eta| < 2.1$ . They are addi-  
 450 tionally required to pass the MVA non triggering isolation with 80% efficiency for real electrons,  
 451 and to have a relative isolation less than 0.1. Hadronic taus are required to have  $p_T$  greater than  
 452 30 GeV, and  $|\eta| < 2.3$ . They should pass the tight MVA isolation, the tight rejection against elec-  
 453 trons, and the loose rejection against muons. In each event, the electron and the  $\tau_h$  candidate  
 454 are required to have opposite sign charge, and to be separated by  $\Delta R > 0.5$ . Events with addi-  
 455 tional electrons, muons, or  $\tau_h$  candidates are rejected. No requirement is made on the number  
 456 of b-tagged jets as the  $t\bar{t}$  contribution is small. Prefit control plots of several variables after the  
 457 selection just described, are shown in Fig. 9, 10, and 11.

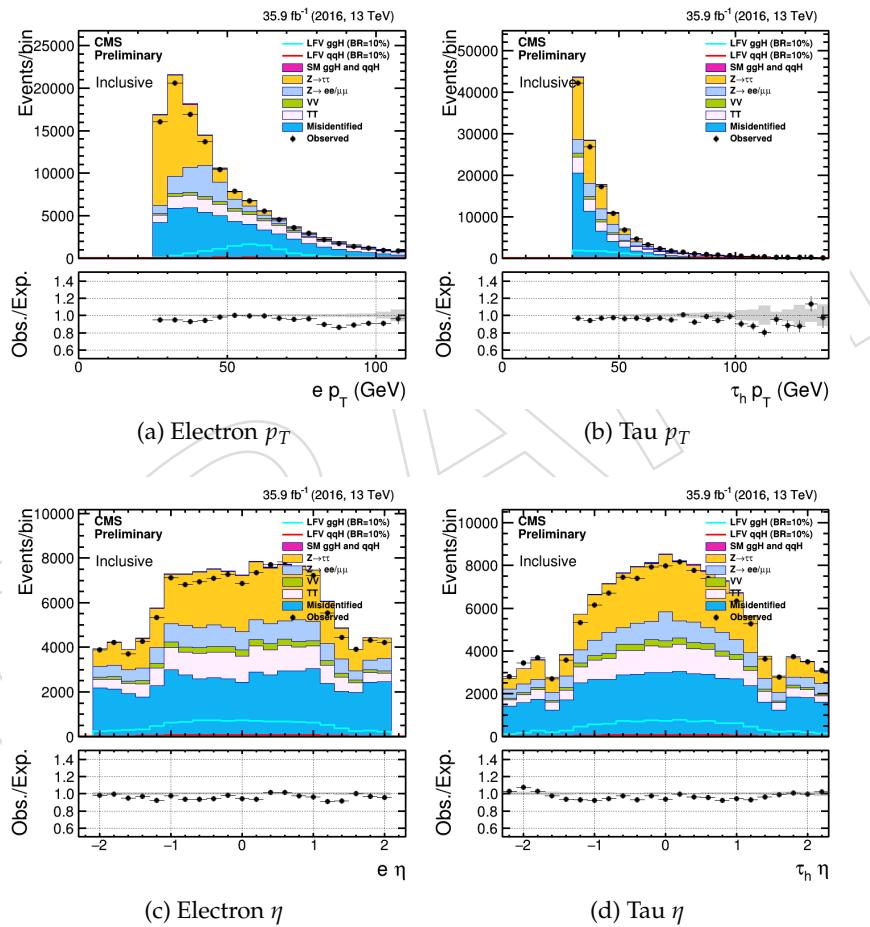


Figure 9: Control distributions of the tau and electron  $p_T$  and  $\eta$  after the loose selection, in the  $e\tau_h$  channel. Only the statistical uncertainty is shown. Plots are prefit.

458 Events are divided into four categories:

459 **0-jet:** The events do not have any jet with  $p_T > 30$  GeV,  $|\eta| < 4.7$ , and pass the loose PF  
 460 ID. This category enhances the gluon-gluon fusion contribution.

461 **1-jet:** The events have one jet with  $p_T > 30$  GeV and  $|\eta| < 4.7$ , and pass the loose PF  
 462 ID. This category enhances the gluon-gluon fusion production with initial state ra-

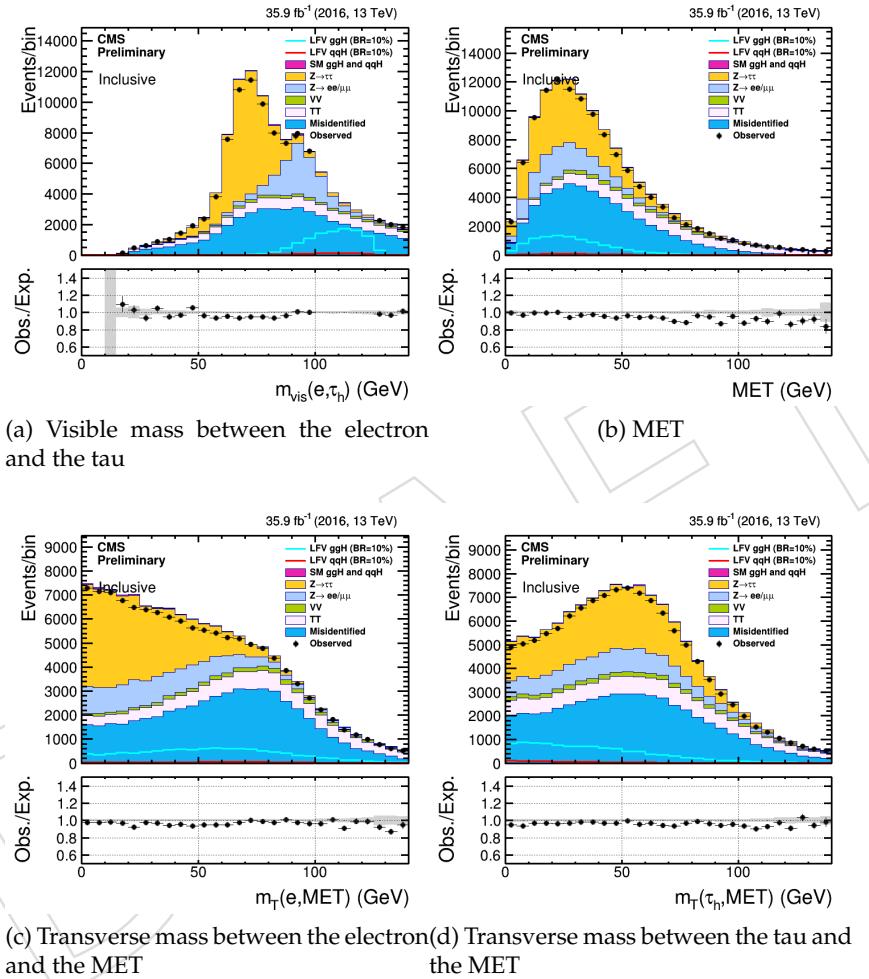


Figure 10: Control distributions of the tau and electron  $p_T$  and  $\eta$  after the loose selection, in the  $e\tau_h$  channel. Only the statistical uncertainty is shown. Plots are prefit.

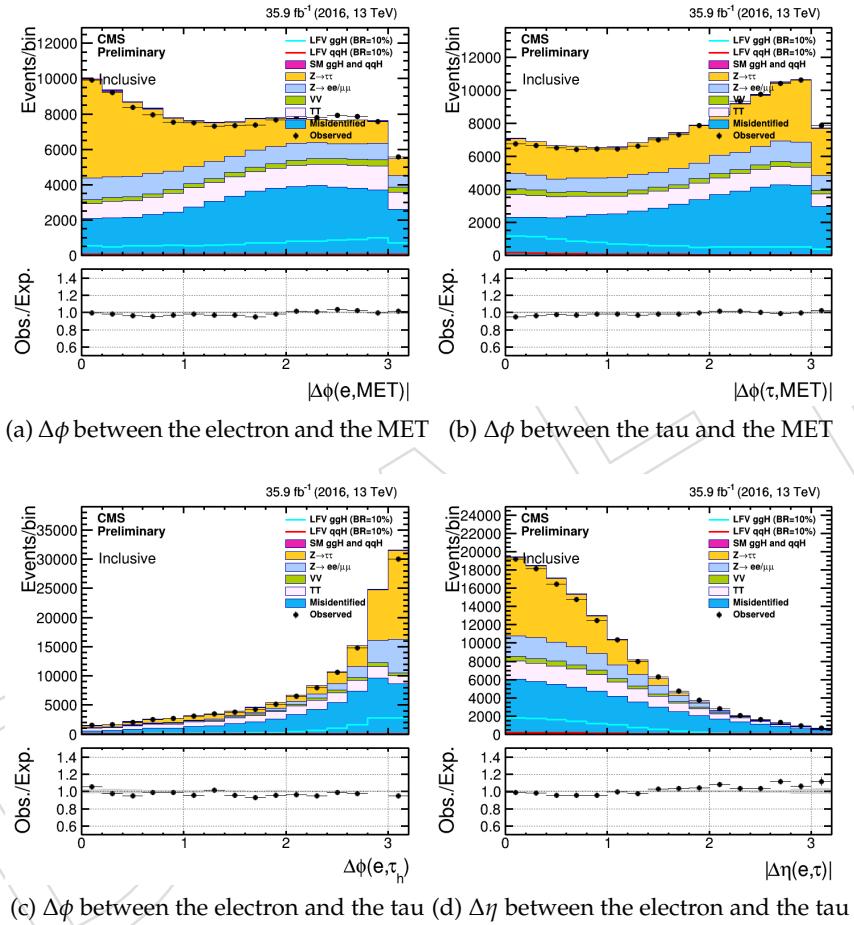


Figure 11: Control distributions of the angles between the electron, tau, and MET, in the  $e\tau_h$  channel. Only the statistical uncertainty is shown. Plots are prefit.

diation (ISR). Some VBF events where one jet has escaped detection also enter this category.

**2-jets GG:** The events have two jets with  $p_T > 30$  GeV and  $|\eta| < 4.7$ , and pass the loose PF ID, with an invariant mass,  $M_{jj} < 500$  GeV. The dominant signal contribution comes from gluon-gluon fusion Higgs boson production.

**2 jets VBF:** The events have two jets with  $p_T > 30$  GeV and  $|\eta| < 4.7$ , and pass the loose PF ID, with an invariant mass,  $M_{jj} > 500$  GeV. The dominant signal contribution comes from VBF Higgs production.

The threshold on  $M_{jj}$  has been optimized to give the best expected limits.

## 6.2 Cut-based analysis

After the loose selection a set of kinematic selection criteria are also required. The thresholds are optimized to give the most stringent expected limits based on an Asimov dataset. The  $p_T$  of the  $\tau_h$  and  $\tau_h$  are constrained to be very loose to avoid making the backgrounds peak under the signal in the  $M_{col}$  distribution. In the signal region, events that have a transverse mass  $m_T(MET, \tau_h) > 60$  GeV are rejected. This cut rejects almost 50% of the misidentified lepton background, while retaining around 90% of the signal events. The  $t\bar{t}$  background, which contributes largely in events with two jets, is also reduced by a factor two after this selection criterion is applied. Although the missing transverse energy is expected to be approximately collinear to the  $\tau_h$  candidate, selecting events with a small  $\Delta\phi$  between the  $E_T^{\text{miss}}$  and the  $\tau_h$  does not improve the expected limits. Similarly, requirements on  $\Delta\phi(e, \tau_h)$  or  $\Delta\eta(e, \tau_h)$  do not offer any improvement.

## 6.3 Multivariate analysis

A Boosted decision tree (BDT) is exploited to provide greater sensitivity. It is trained after the loose selection but combining all categories. The signal is a mixture of gg and VBF signal events, weighted according to their respective production cross sections. The background is a data sample of misidentified lepton events. This is the dominant background in this channel. The leptons are required to be like sign and be anti-isolated in order to be orthogonal to the signal region, yet have the same kinematic properties. Two options for the BDT analysis were studied in previous iterations of this analysis, called Method A and Method B. It was found that Method B gives significantly improved sensitivity to either the cut based analysis or Method A, maintaining good signal background shape discrimination. As a result, only the BDT Method B, relabelled simply "BDT" has been continued for the final results with  $36\text{fb}^{-1}$ . Details of the differences between BDT Method A and B can be found in Appendix B.

The input variables to the BDT are:

- Transverse mass between the  $\tau_h$  and the  $E_T^{\text{miss}}$ ,  $M_T(\tau_h, E_T^{\text{miss}})$ .
- Missing transverse energy,  $E_T^{\text{miss}}$ .
- Pseudorapidity difference between the electron and the  $\tau_h$  candidate,  $\Delta\eta(e, \tau_h)$ .
- Azimuthal angle between the electron and the  $\tau_h$ ,  $\Delta\phi(e, \tau_h)$ .
- Azimuthal angle between the  $\tau_h$  and the  $E_T^{\text{miss}}$ ,  $\Delta\phi(\tau_h, E_T^{\text{miss}})$ .
- Collinear mass,  $M_{col}$ .
- Electron  $p_T$ .
- $p_T$  of the  $\tau_h$ .

The distributions of these variables for the signal and the background are given in Fig. 12. Note

505 that the variables are identical to the  $\mu\text{-}\tau_h$  channel except with  $\mu \rightarrow e$ .

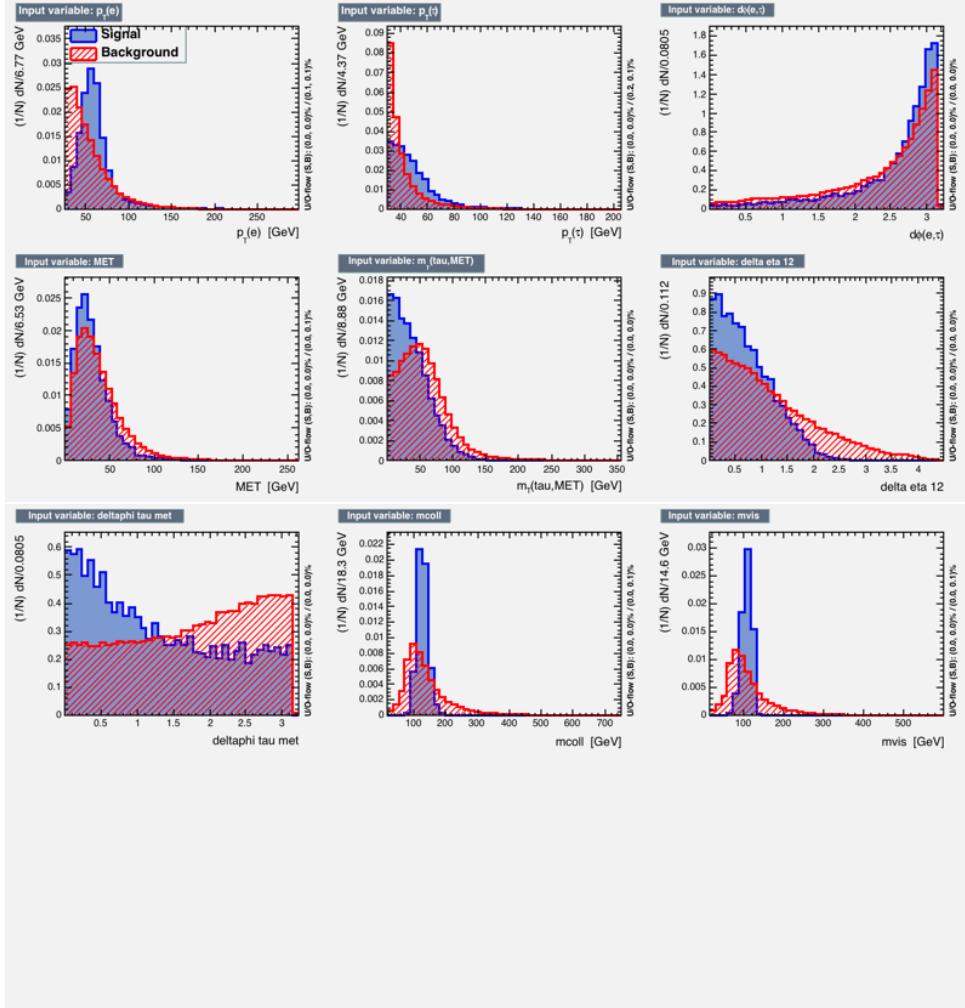
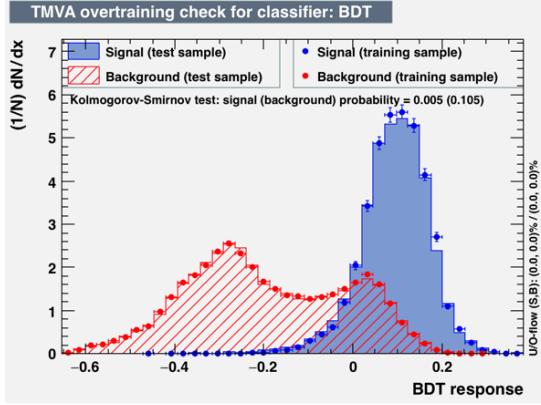
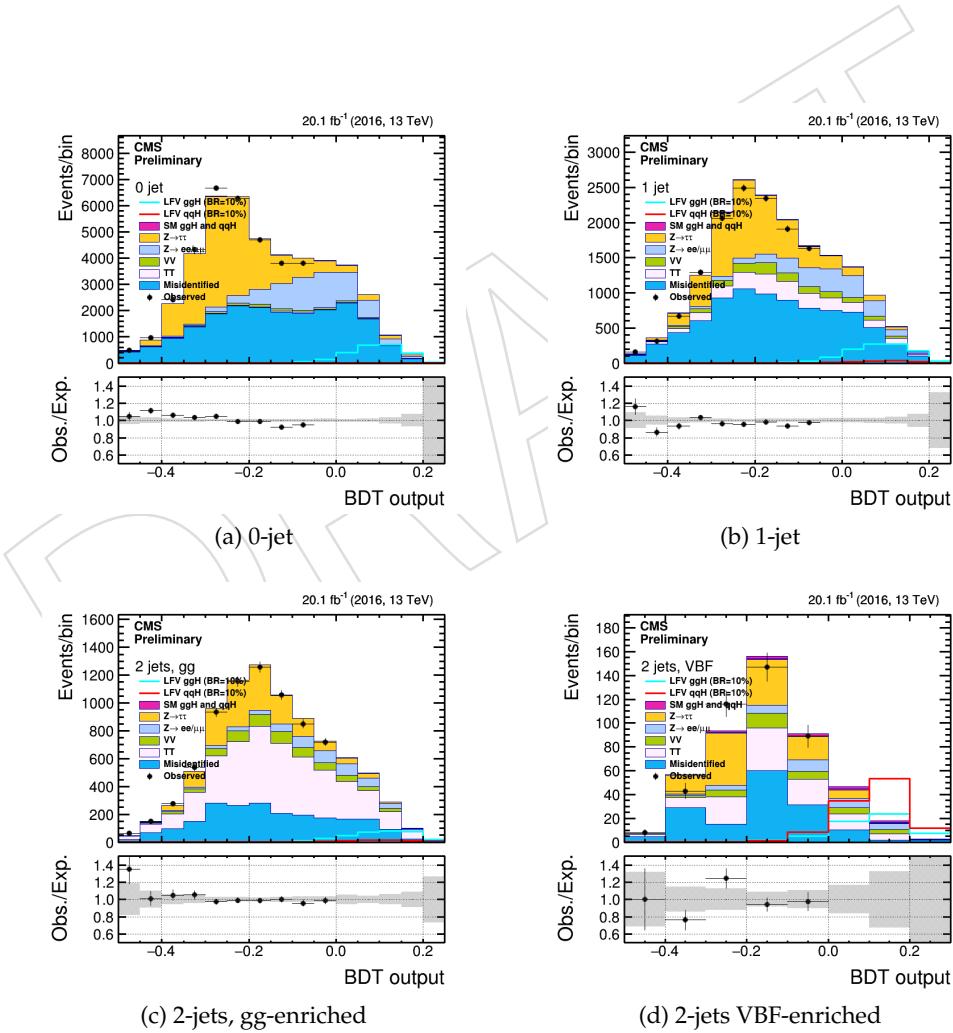


Figure 12:  $H \rightarrow e\tau_h$ : Normalized distributions of the variables that enter the BDT for the  $e\tau_h$  analysis. The signal (blue) is composed of a weighted mixture of gg and VBF  $H \rightarrow e\tau$  events, whereas the background (red) is made of background events selected in data with inverted lepton isolation and same-sign requirement. The events used in the BDT pass the same  $p_T$  and  $\eta$  requirements as in the signal region.

506 The BDT output distributions for the signal and  $t\bar{t}$  background are shown in Fig. 13. The figures  
 507 also show the results of Kolmogorov-Smirnov tests to assess the compatibility of the distributions  
 508 on the training sample and on a test sample composed of different events. The data/MC  
 509 distributions are shown in Fig. 14. Good agreement is seen between observed data and predicted  
 510 backgrounds in the unblinded regions with low signal significance. No threshold require-  
 511 ment is made on the BDT output. Instead the expected limits are estimated by fitting the  
 512 BDT output distributions in the four jet categories.

#### 513 6.4 Validation of background methods

514 The misidentified lepton backgrounds are estimated with the full data driven method de-  
 515 scribed in Sec 4. This background is validated with a like-sign control region. Figs. 15 and  
 516 16 show the data compared to the background estimation in the like sign control region. This

Figure 13:  $H \rightarrow e\tau_h$ : BDT output distributions.Figure 14:  $H \rightarrow e\tau_h$ : BDT output in the four categories. The plots are partially blinded, and only the statistical uncertainty is shown in the ratio plots.

sample is dominated by the misidentified lepton background component and the agreement is generally good. The reducible background purity is lower in the 2 jets categories, where the  $t\bar{t}$  background contributes more.

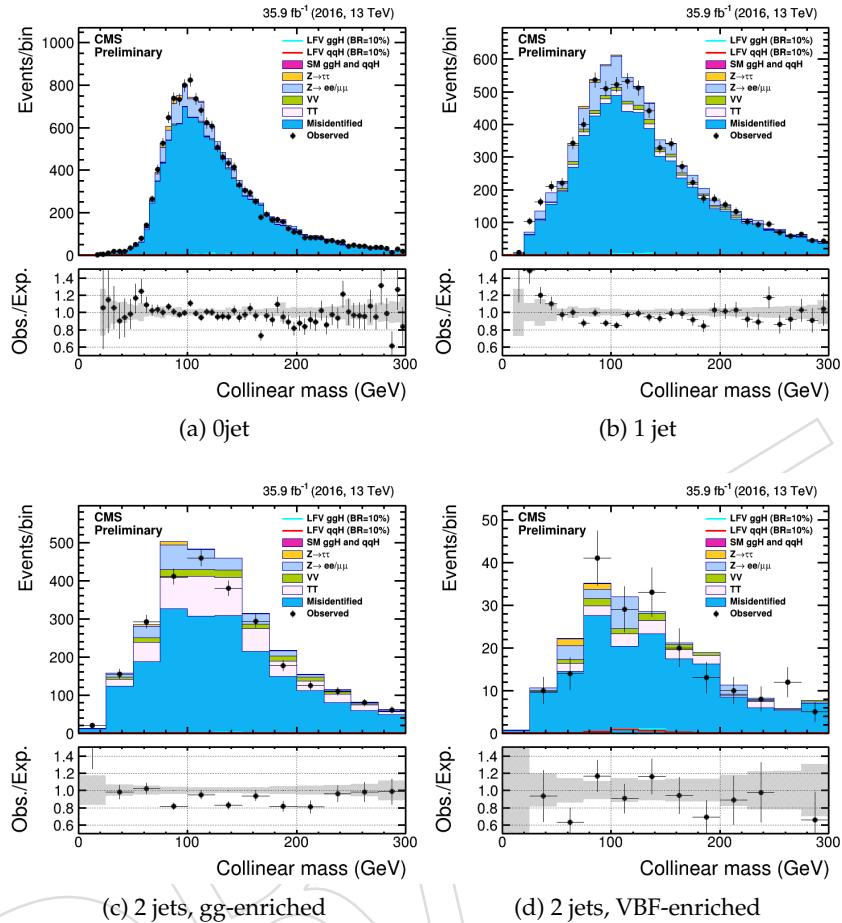


Figure 15:  $H \rightarrow e\tau_h$ : The  $M_{col}$  distribution in the four categories of the analysis, after the loose selection. The electron and the  $\tau_h$  candidate are required to have the same charge. The plots are prefit, and only the statistical uncertainty is shown.

519  
 520 The control regions described previously have a different jet composition than the signal region  
 521 because of the inversion of the charge requirement, and different behavior could be expected  
 522 for jets with different flavors. A signal-free control region with jet composition closer to the sig-  
 523 nal region can be obtained by requiring the electron to have a different charge, and by requiring  
 524 the transverse mass between the electron and the missing transverse energy to be greater than  
 525 60 GeV and the transverse mass between the  $\tau_h$  and the transverse missing energy greater than  
 526 80 GeV. This region is enriched in W+jets events. As shown in Fig. 17, data and predicted  
 527 backgrounds are also in reasonable agreement in all the categories of the analysis. Only the  
 528 statistical uncertainty is shown

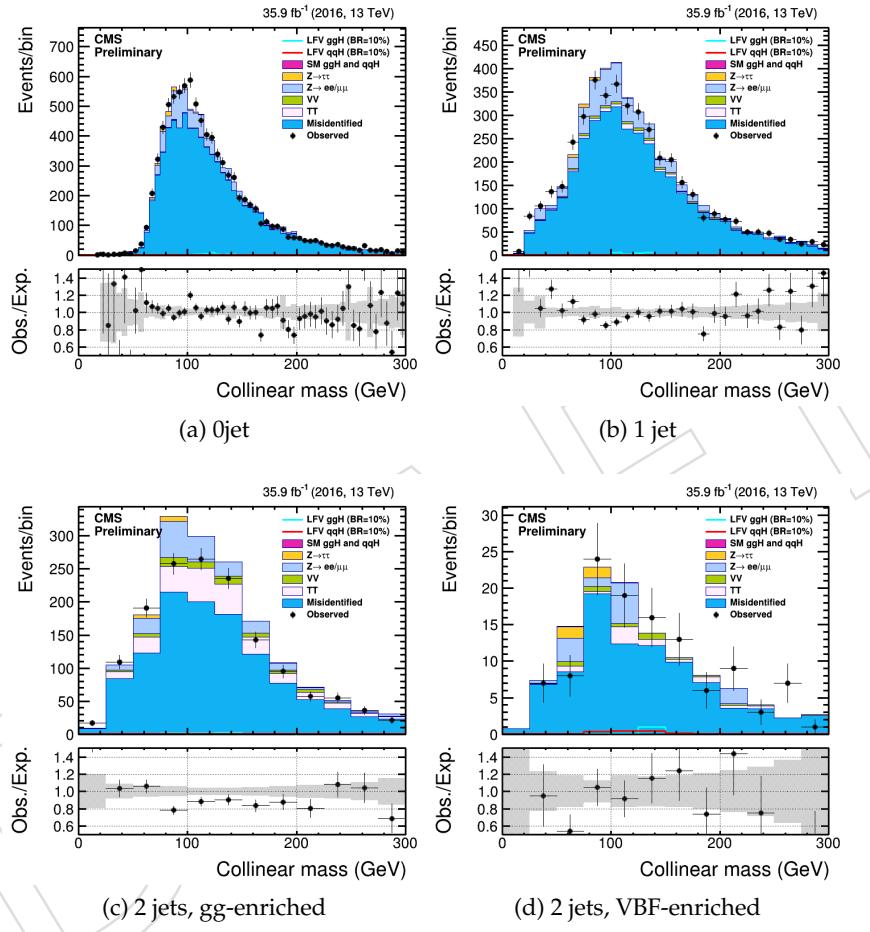


Figure 16:  $H \rightarrow e\tau_h$ : Collinear mass in the four categories of the analysis, after the optimized selection. The electron and the  $\tau_h$  candidate are required to have the same charge. The plots are prefit, and only the statistical uncertainty is shown.

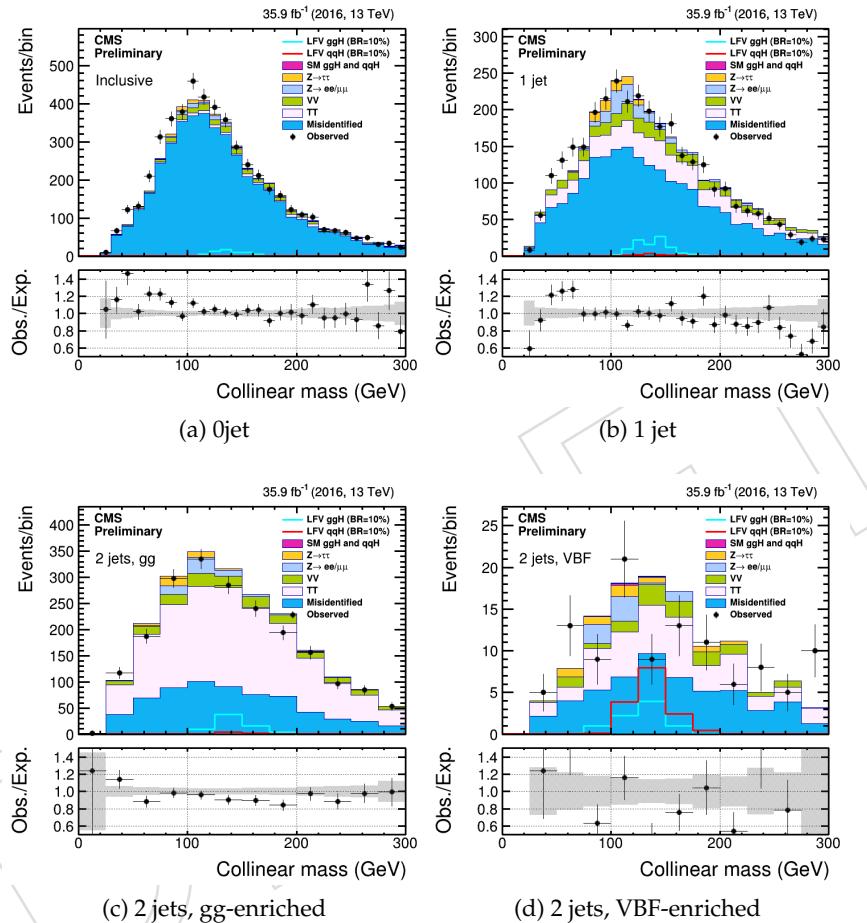


Figure 17:  $H \rightarrow e\tau_h$ : Collinear mass in the four categories of the analysis. The electron and the  $\tau_h$  candidate are required to have opposite sign charge. The transverse mass between the electron and the missing transverse energy is greater than 60 GeV and the transverse mass between the  $\tau_h$  and the transverse missing energy greater than 80 GeV to enrich the samples in  $W + jets$  events. The plots are prefit, and only the statistical uncertainty is shown.

## 529 7 $H \rightarrow \mu\tau_e$ channel

### 530 7.1 loose selection

531 In the  $H \rightarrow \mu\tau_e$  channel events are required to pass either of the isolated single muon  
 532 triggers; HLT\_IsoMu24 or HLT\_IsoTkMu24. It is then required to have an isolated muon and  
 533 an isolated electron of opposite charge within the fiducial region and separated by  $\Delta R > 0.3$ .  
 534 Events with additional electrons, muons, or  $\tau_h$  candidates are rejected. Events with one or more  
 535 b-jets are also rejected to reduce the  $t\bar{t}$  background contribution. The muon is then required to  
 536 have  $p_T > 26$  GeV,  $|\eta| < 2.4$ , and to be isolated from other objects in the event having a relative  
 537 isolation  $R_{\text{iso}}^{\ell} < 0.15$  in a cone of  $R=0.4$  around the direction of the muon and also to pass the PF  
 538 Medium/ICHEP Medium working point selection criteria from the Muon POG. The electron  
 539 is required to have  $p_T$  greater than 10 GeV, and  $|\eta| < 1.4442$  or  $1.566 < |\eta| < 2.3$ . They  
 540 are additionally required to pass the MVA non triggering isolation with 80% efficiency for real  
 541 electrons, and have a relative isolation less than 0.1 in the cone  $R=0.3$  around the direction of  
 542 the electron. The events are then divided into categories within each sample according to the  
 543 number of jets in the event. This is designed to enhance the contribution of different Higgs  
 544 production mechanisms. The four categories are:

545 **0-jet:** The events do not have any jet with  $p_T > 30$  GeV,  $|\eta| < 4.7$ , and pass the loose PF  
 546 ID. This category enhances the gluon-gluon fusion contribution.

547 **1-jet:** The events have one jet with  $p_T > 30$  GeV and  $|\eta| < 4.7$ , and pass the loose PF  
 548 ID. This category enhances the gluon-gluon fusion production with initial state ra-  
 549 diation (ISR). Some VBF events where one jet has escaped detection also enter this  
 550 category.

551 **2-jets GG:** The events have two jets with  $p_T > 30$  GeV and  $|\eta| < 4.7$ , and pass the loose  
 552 PF ID, with an invariant mass,  $M_{jj} < 550$  GeV. The dominant signal contribution  
 553 comes from gluon-gluon fusion Higgs boson production.

554 **2 jets VBF:** The events have two jets with  $p_T > 30$  GeV and  $|\eta| < 4.7$ , and pass the loose  
 555 PF ID, with an invariant mass,  $M_{jj} \geq 550$  GeV. The dominant signal contribution  
 556 comes from VBF Higgs production.

557 The threshold on  $M_{jj}$  has been optimized to give the best expected limits.

### 558 7.2 Cut-based analysis

559 After the loose selection a set of stricter kinematic selection criteria are also required. These  
 560 thresholds are optimized to give the most stringent expected limits based on an Asimov dataset.  
 561 The  $p_T$  of the  $\mu$  and  $e$  are kept loose to avoid making the backgrounds peak under the signal  
 562 in the  $M_{\text{col}}$  distribution. The muon  $p_T$  is required to be greater than 30 GeV for the 0-jet cate-  
 563 gory and 26 GeV for the other categories. Signal events are characterized by a higher transverse  
 564 mass between the muon and the  $E_T^{\text{miss}}$ : In the 0-jet, 1-jet , 2-jets GG and 2-jets VBF categories the  
 565 transverse mass  $M_T(\text{MET}, \mu)$  is required to be greater than 60,40,15 and 15 GeV respectively.  
 566 The neutrinos from the tau lepton decay are approximately collinear to the electron direction;  
 567 so an additional requirement is made on the azimuthal angle between the electron and the  
 568  $E_T^{\text{miss}}$ ;  $\Delta\phi(p_T^e, \text{MET}) < 0.7, 0.7, 0.5, 0.3$  for the 0-jet, 1-jet, 2-jets GG and 2 jets VBF categories  
 569 respectively. In the 0-jet and 1-jet category it is further required that  $\Delta\phi(p_T^e, p_T^{\mu\mu}) > 2.5, 1.0$ .

### 570 7.3 Multivariate analysis

571 A Boosted decision trees (BDT) is exploited to provide greater sensitivity. It is trained after  
 572 the loose selection but combining all categories. The signal is a mixture of gg and VBF signal

events, weighted according to their respective production cross sections. The background is a mixed sample of  $t\bar{t}$  and DY events also weighted by their cross-sections. The  $t\bar{t}$  background is the dominant background in this channel for the 2 jet category and very significant in 1 jet. It has many kinematic characteristics in common with other backgrounds such as diboson and single top. The DY background is the dominant background in 0-jet and 1-jet category. Two options for the BDT analysis were studied in previous iterations of this analysis, called Method A and Method B. It was found that Method B gives significantly improved sensitivity to either the cut based analysis or Method A, maintaining good signal background shape discrimination. As a result, only the BDT Method B, simply relabelled "BDT" has been continued for the final results with  $36\text{fb}^{-1}$ . Details of the differences between BDT Method A and B can be found in Appendix B.

The input variables of this BDT, shown in Fig. 18, are the following:

- Transverse mass between the  $\mu$  and  $E_T^{\text{miss}}$ ,  $M_T(\mu, E_T^{\text{miss}})$ .
- Transverse mass between the  $e$  and  $E_T^{\text{miss}}$ ,  $M_T(e, E_T^{\text{miss}})$ .
- Azimuthal angle between the  $e$  and  $\mu$ ,  $\Delta\phi(e, \mu)$ .
- Azimuthal angle between the  $e$  and  $E_T^{\text{miss}}$ ,  $\Delta\phi(e, E_T^{\text{miss}})$ .
- Azimuthal angle between the  $\mu$  and  $E_T^{\text{miss}}$ ,  $\Delta\phi(\mu, E_T^{\text{miss}})$ .
- Collinear mass,  $M_{\text{col}}$ .
- Muon  $p_T$ .
- Electron  $p_T$ .

The distribution of these variables for signal and background is shown in figure 18

No threshold requirement is made on the BDT output. Instead the expected limits are estimated by fitting the BDT output distributions in the four jet categories. Fig. 19 shows the correlations between the input variables for the BDT with lepton  $p_T$  as input.

The data to simulation distributions of the input variables to the BDT are shown in Fig. 20 and Fig. 21. There is good agreement between observed and predicted backgrounds in the unblinded regions.

The BDT output distributions for the signal and  $t\bar{t}$  background are shown in Fig. 22. The figures also show the results of Kolmogorov-Smirnov tests to assess the compatibility of the distributions on the training sample and on a test sample composed of different events.

The data to simulation distributions of the BDT outputs are shown in Section 10, Fig. 40. There is good agreement between observed and predicted backgrounds in the unblinded regions.

## 7.4 Validation of the background methods

The misidentified lepton background in this channel is much less important than in the hadronic final states. Two different methods were used to estimate this background. These are the fully data driven and semi-data driven methods described in section 4.

Further data to simulation distributions for different control regions targetting the key backgrounds of the analysis can be found in Appendix C.

611

Figs. 23 and 24 compare the full and semi data driven background estimation. The two methods are in agreement but the semi-data driven method gives greater statistical precision. As the reducible background is quite small in the  $\mu e$  final state, the expected limits measured with

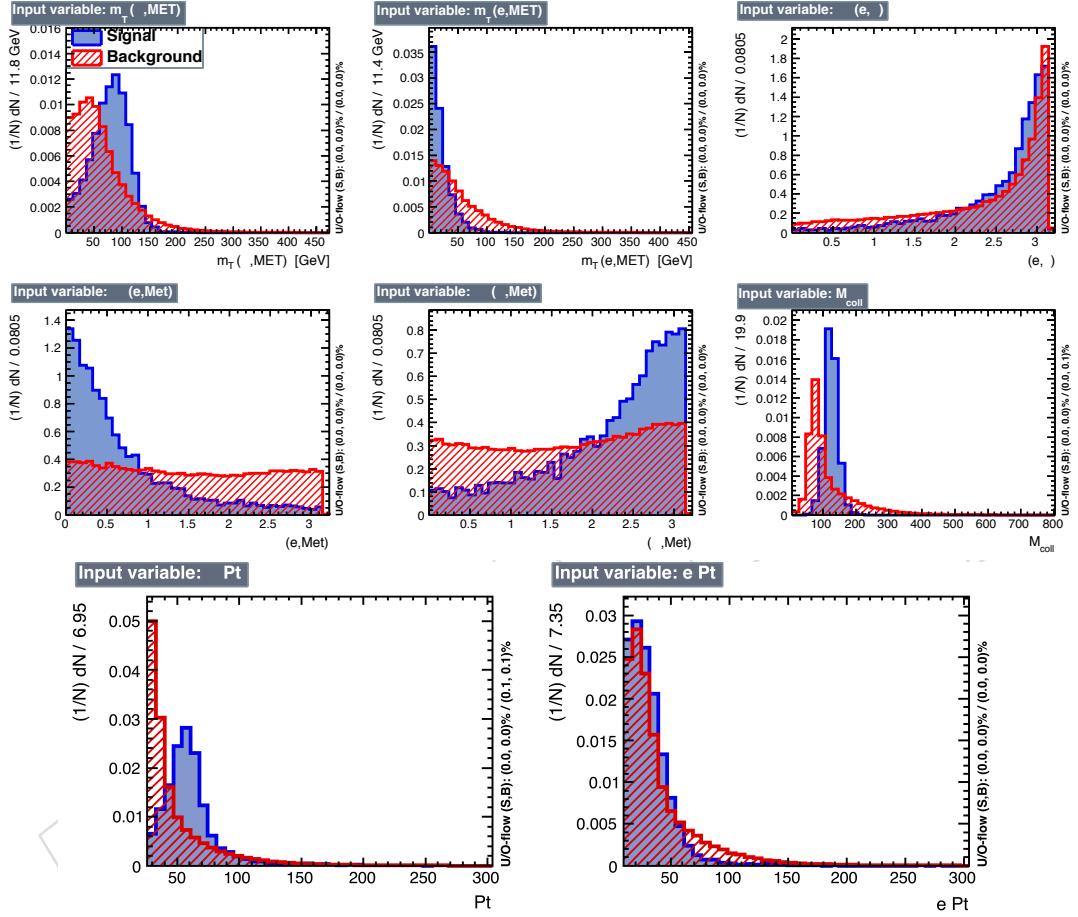


Figure 18:  $H \rightarrow \mu\tau_e$ : Normalized distributions of the variables that enter the second BDT for the  $\mu\tau_e$  analysis with collinear mass and lepton  $p_T$ s as input variables. The signal (blue) is composed of a weighted mixture of gg and VBF  $H \rightarrow \mu\tau$  events, whereas the background (red) is made of  $t\bar{t}$  and DY events. The events used in the BDT pass the same loose selection cuts as in the signal region.

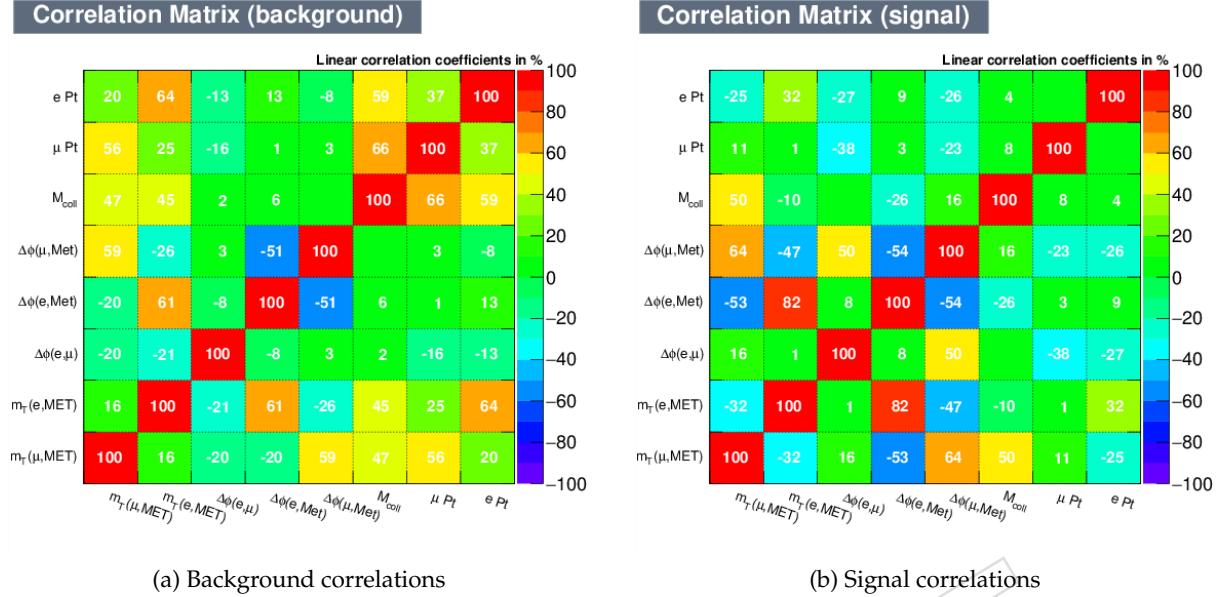


Figure 19:  $H \rightarrow \mu\tau_e$ : Correlations between input variables for signal (right) and background (left) for BDT with the collinear mass and the lepton  $p_T$  as inputs, in the  $\mu e$  final state.

615 both reducible background estimates are comparable between each other. The semi-data driven  
 616 technique is used in the analysis.

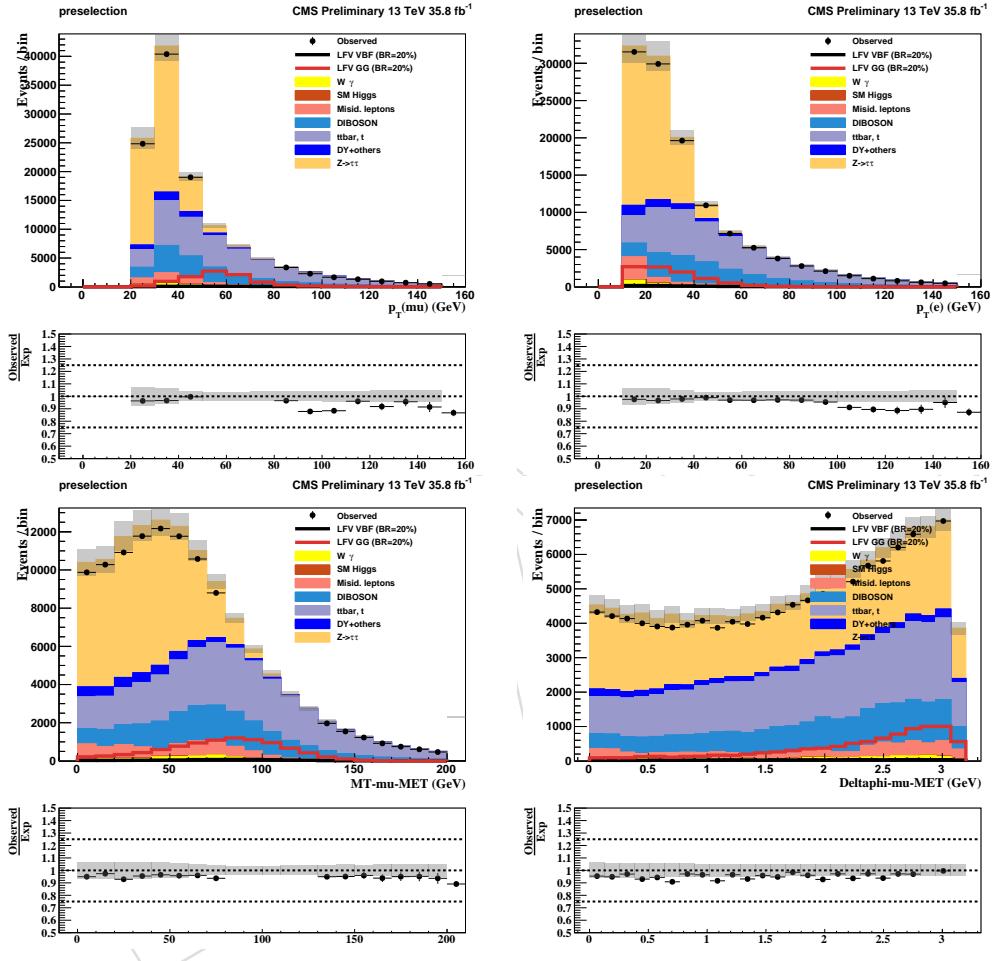
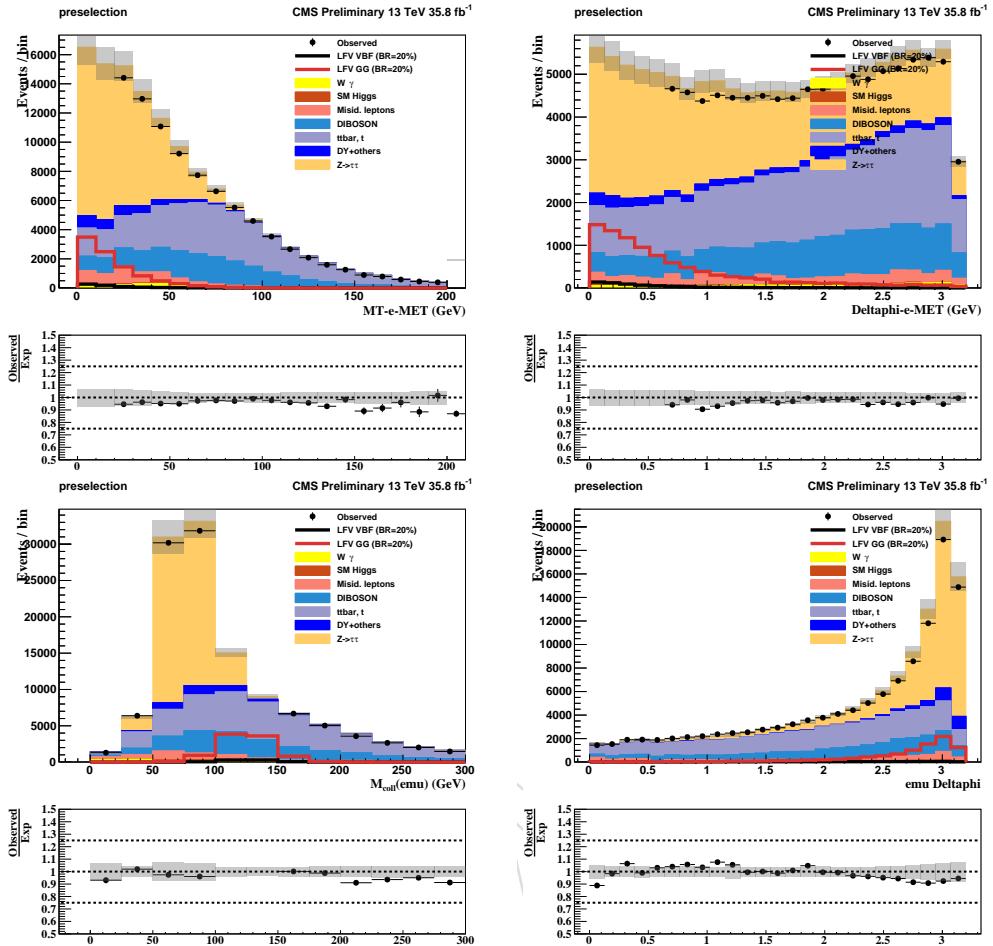
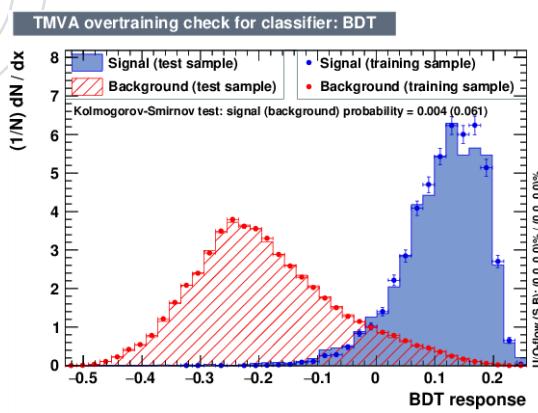


Figure 20:  $H \rightarrow \mu\tau_e$ : Control distributions for the different variables that enter the BDT

Figure 21:  $H \rightarrow \mu\tau_e$ : Control distributions for the different variables that enter the BDTFigure 22:  $H \rightarrow \mu\tau_e$ : BDT output distributions

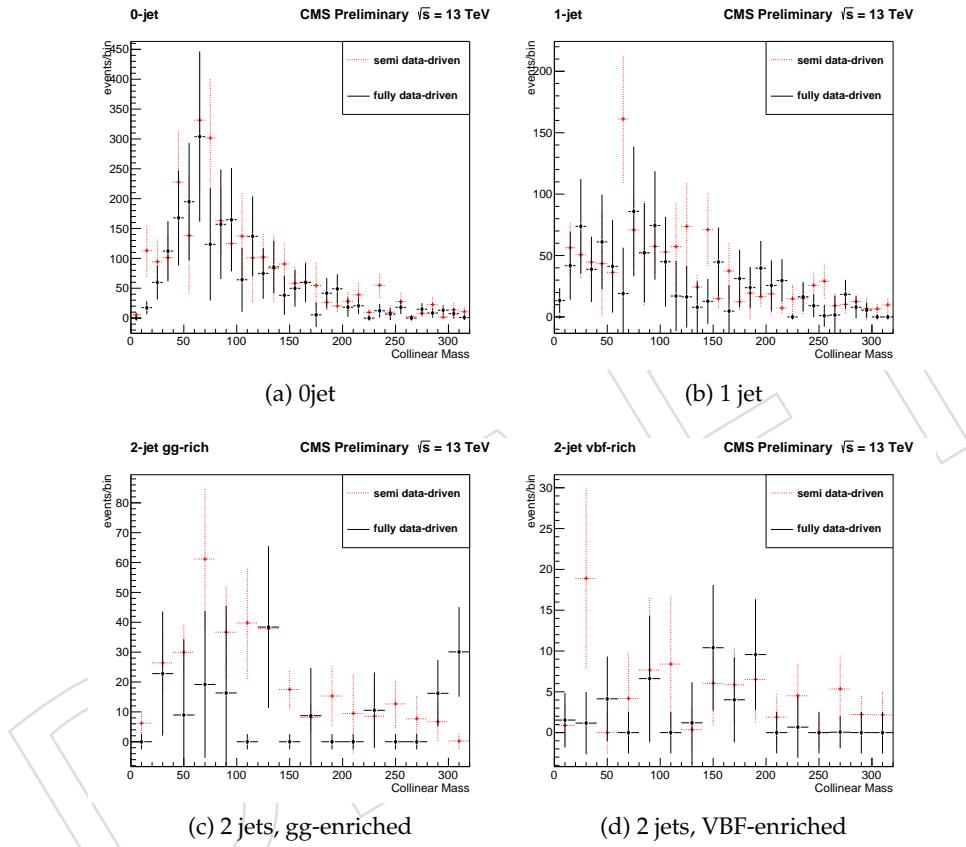


Figure 23:  $H \rightarrow \mu\tau_e$ : Comparison of the reducible background estimates with the semi data-driven (W from MC samples, QCD from SS data) and the fully data-driven (W and QCD from fake rate method) methods after loose selection, in the  $\mu e$  final state.

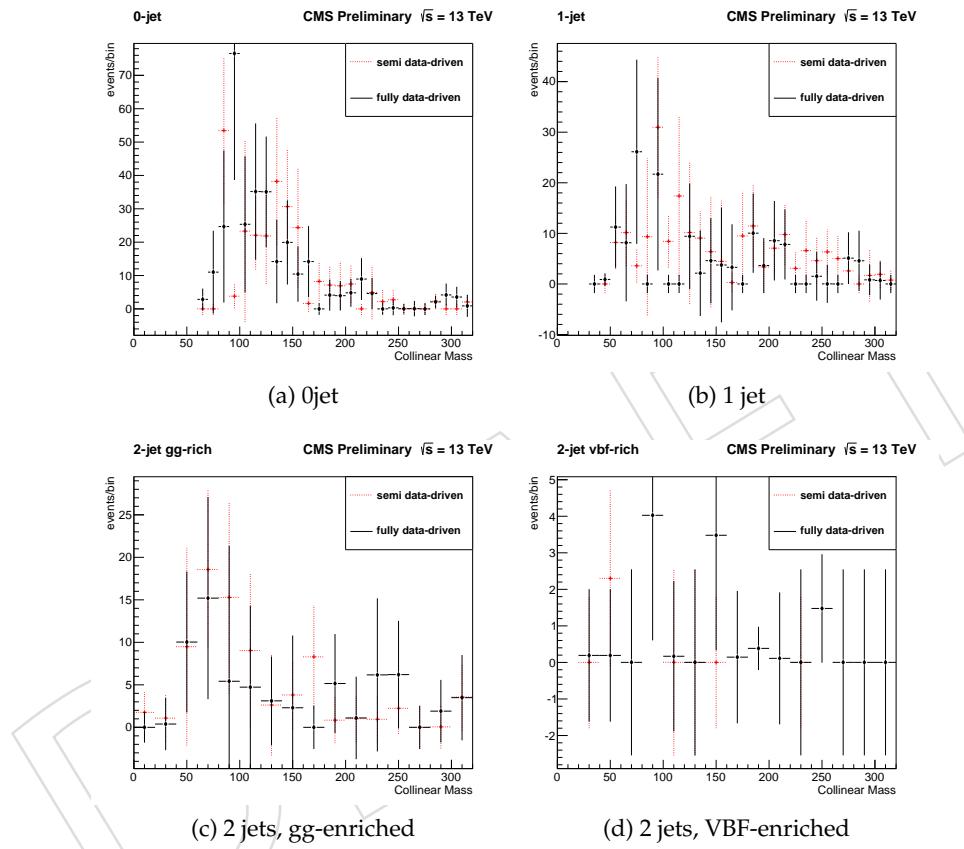


Figure 24:  $H \rightarrow \mu\tau_e$ : Comparison of the reducible background estimates with the semi data-driven (W from MC samples, QCD from SS data) and the fully data-driven (W and QCD from fake rate method) methods after final selection, in the  $\mu e$  final state.

## 617 8 $H \rightarrow e\tau_\mu$ channel

### 618 8.1 Loose selection

619 In the  $H \rightarrow e\tau_\mu$  channel events are required to pass the HLT\_Mu8\_TrkIsoVVL\_Ele23\_CaloIdL\_TrackIdL\_IsoVL  
 620 or HLT\_Mu23\_TrkIsoVVL\_Ele12\_CaloIdL\_TrackIdL\_IsoVL triggers. The electron candidates are  
 621 then required to have  $p_T > 26$  GeV, and  $|\eta| < 2.1$ . They are additionally required to pass the  
 622 MVA non triggering isolation with 80% efficiency for genuine electrons, and to have a relative  
 623 isolation less than 0.15. Muons are required to have  $p_T$  greater than 10 GeV and  $|\eta| < 2.4$ . They  
 624 should pass the medium ICHEP identification, and have a relative isolation less than 0.10. The  
 625 electron and the muon candidates are required to have an opposite sign charge, and to be sep-  
 626 arated by  $\Delta R > 0.5$ . Events with additional electrons, muons, or  $\tau_h$  candidates are rejected. No  
 627 requirement is made on the number of b-tagged jets. Prefit control plots of several variables  
 628 after the selection just described, are shown in Fig. 25, 26, and 27.

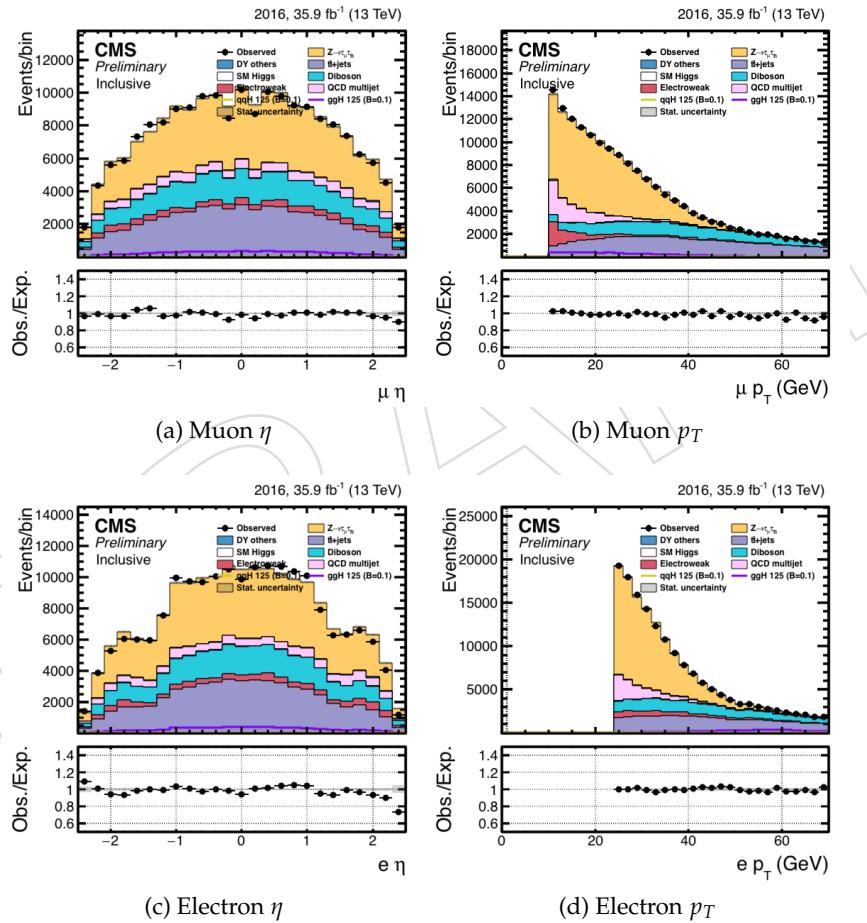


Figure 25: Control distributions of the muon and electron  $p_T$  and  $\eta$  after the loose selection, in the  $e\tau_\mu$  channel. Only the statistical uncertainty is shown. Plots are prefit.

629 The events are then divided into categories within each sample according to the number of  
 630 jets in the event. This is designed to enhance the contribution of different Higgs production  
 631 mechanisms. The four categories are:

632 **0-jet:** The events do not have any jet with  $p_T > 30$  GeV,  $|\eta| < 4.7$ , and pass the loose PF  
 633 ID. This category enhances the gluon-gluon fusion contribution.

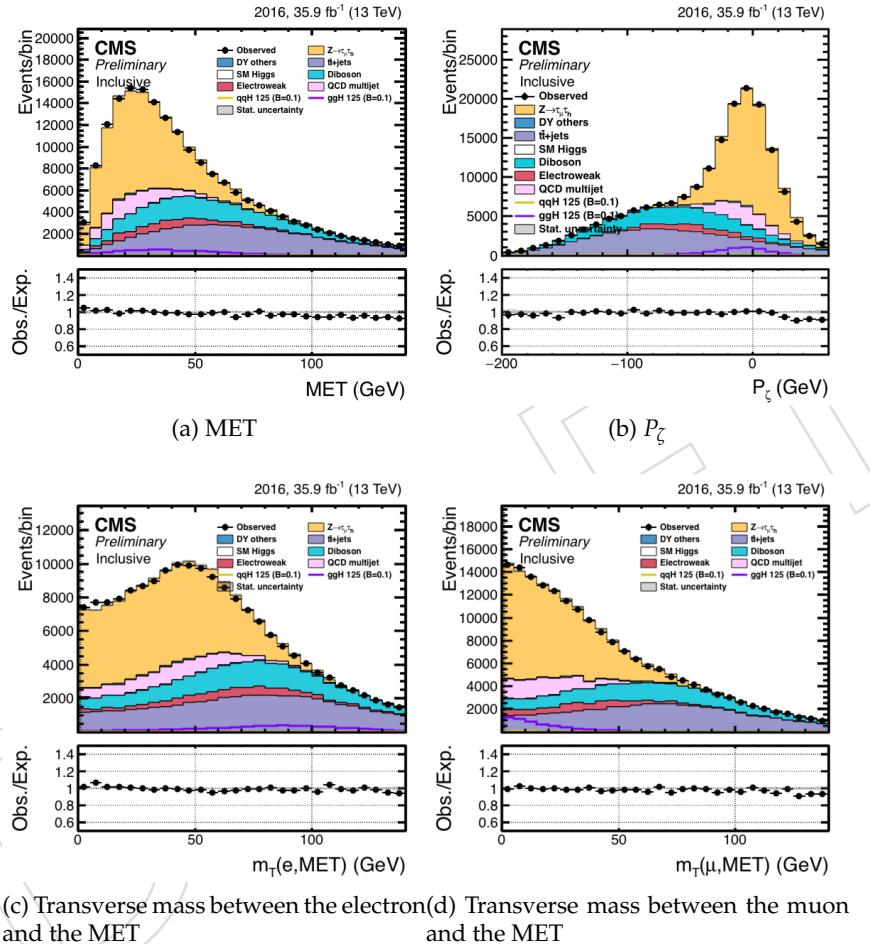


Figure 26: Control distributions of the MET,  $P_\zeta$ , transverse mass between the muon and the MET, and the transverse mass between the electron and the MET. Only the statistical uncertainty is shown. Plots are prefit.

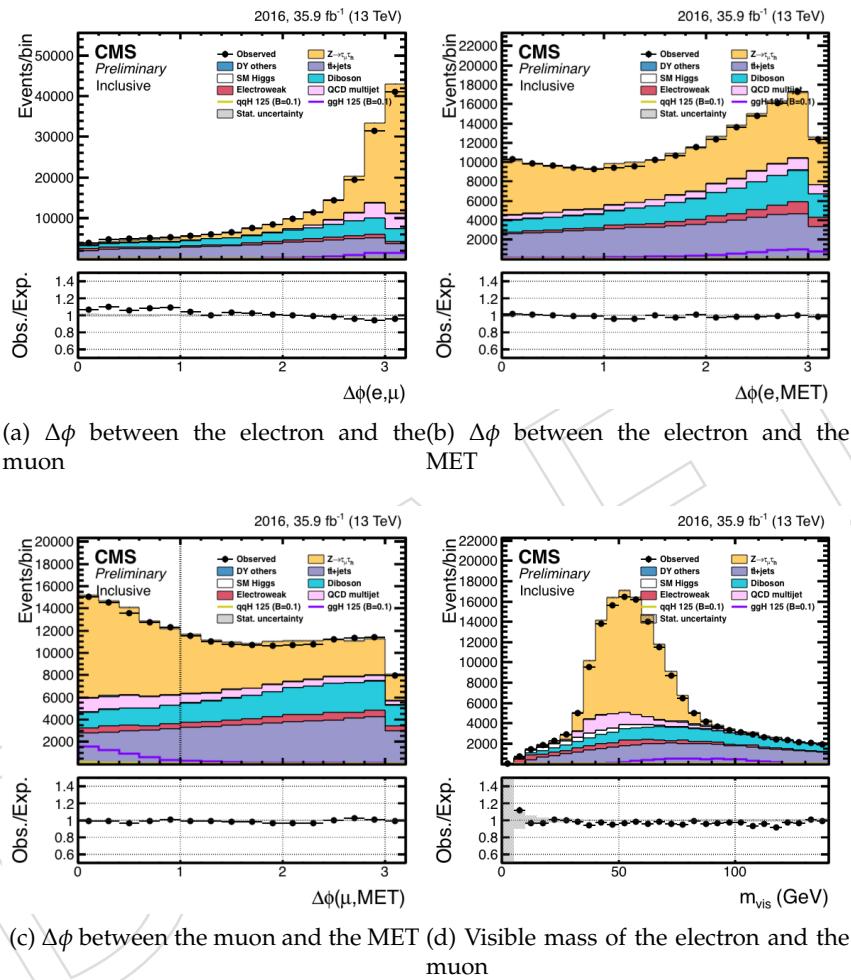


Figure 27: Control distributions of the azimuthal angle between the electron and muon, between the electron and the MET, and between the muon and the MET. Only the statistical uncertainty is shown. Plots are prefit.

634 **1-jet:** The events have one jet with  $p_T > 30$  GeV and  $|\eta| < 4.7$ , and pass the loose PF  
 635 ID. This category enhances the gluon-gluon fusion production with initial state ra-  
 636 diation (ISR). Some VBF events where one jet has escaped detection also enter this  
 637 category.

638 **2-jets GG:** The events have two jets with  $p_T > 30$  GeV and  $|\eta| < 4.7$ , and pass the loose  
 639 PF ID, with an invariant mass,  $M_{jj} < 500$  GeV. The dominant signal contribution  
 640 comes from gluon-gluon fusion Higgs boson production.

641 **2 jets VBF:** The events have two jets with  $p_T > 30$  GeV and  $|\eta| < 4.7$ , and pass the loose  
 642 PF ID, with an invariant mass,  $M_{jj} > 500$  GeV. The dominant signal contribution  
 643 comes from VBF Higgs production.

644 The threshold on  $M_{jj}$  has been optimized to give the best expected limits.

## 645 8.2 Cut-based analysis

646 After the loose selection a set of kinematic selection criteria are also required. The thresholds  
 647 are optimized to give the most stringent expected limits based on an Asimov dataset. The  $p_T$   
 648 of the  $\mu$  and  $e$  are kept very loose to avoid making the backgrounds peak under the signal in the  
 649  $M_{col}$  distribution.

650 The neutrinos from the tau lepton decay are approximately to be collinear to the muon direction  
 651 so an additional requirement is made on the azimuthal angle between the electron and the  
 652  $E_T^{\text{miss}}$ ;  $\Delta\phi(p_T^e, \text{MET}) < 1.0$ . The limits improve by about 15% when applying this criterion on  
 653 top of the loose selection described above. The  $\Delta\phi(\mu, \text{MET})$  distributions in the four categories  
 654 of the analysis, after the loose selection, are shown in Fig. 28. The large  $t\bar{t}$  background can be  
 655 reduced by requiring  $P_\zeta - 60$  GeV.  $P_\zeta - 60$  is defined in the previous section. This selection  
 656 criterion further improves the expected limits by about 10%. More than half of the  $t\bar{t}$  and  
 657 diboson backgrounds are rejected, whereas the signal efficiency is close to 95%, as shown in  
 658 Fig. 29. Finally, the transverse mass between the electron and the missing transverse energy  
 659  $M_T(\text{MET}, \mu)$  is required to be greater than 60 GeV. This removes about 70% of the Drell-Yan  
 660 events, and 30% of other processes, including signal. Distributions of this variable after the  
 661 loose selection are shown in Fig. 30. This requirement improves the limits by approximately  
 662 15%.

## 663 8.3 Multivariate analysis

664 A Boosted decision trees (BDT) is exploited to provide greater sensitivity. It is trained after the  
 665 loose selection but combining all categories. The signal is a mixture of ggH and VBF signal  
 666 events, weighted according to their respective production cross sections. The background is  
 667 a sample of  $t\bar{t}$  events. The  $t\bar{t}$  background is the dominant background in this channel for the  
 668 2 jet category and very significant in 1 jet. It has many kinematic characteristics in common  
 669 with other backgrounds such as diboson and single top. Two options for the BDT analysis  
 670 were studied in previous iterations of this analysis, called Method A and Method B. It was  
 671 found that Method B gives significantly improved sensitivity to either the cut based analysis  
 672 or Method A, maintaining good signal background shape discrimination. As a result, only the  
 673 BDT Method B, relabelled simply "BDT" has been continued for the final results with  $36\text{fb}^{-1}$ .  
 674 Details of the differences between BDT Method A and B can be found in Appendix B.

675 The input variables to the BDT are:

- 676 • Transverse mass between the muon and the  $E_T^{\text{miss}}$ ,  $m_T(\mu, E_T^{\text{miss}})$ ;
- 677 • Azimuthal angle between the electron and the muon,  $\Delta\phi(e, \mu)$ ;

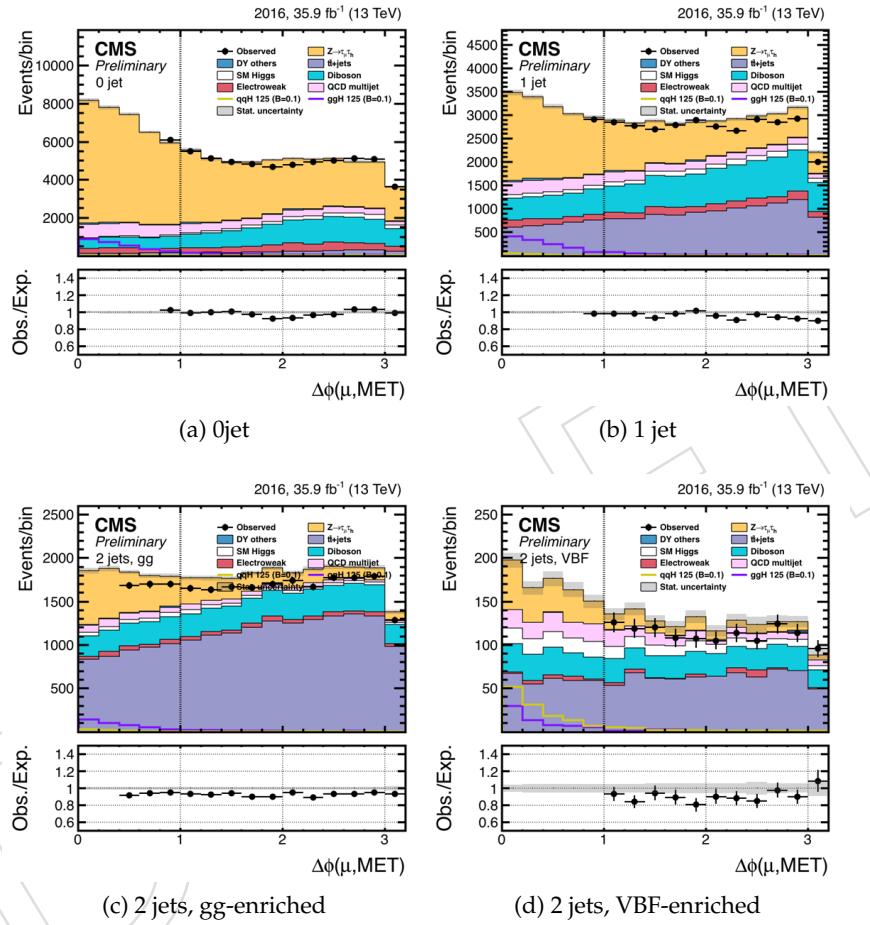


Figure 28:  $H \rightarrow e\tau_\mu$ :  $\Delta\phi(\mu, \text{MET})$  distributions in the four categories of the analysis. The optimized threshold for the cut is indicated by the vertical dashed line. The plots are prefit, and partially blinded according to the signal significance in each bin. Only the statistical uncertainty is shown.

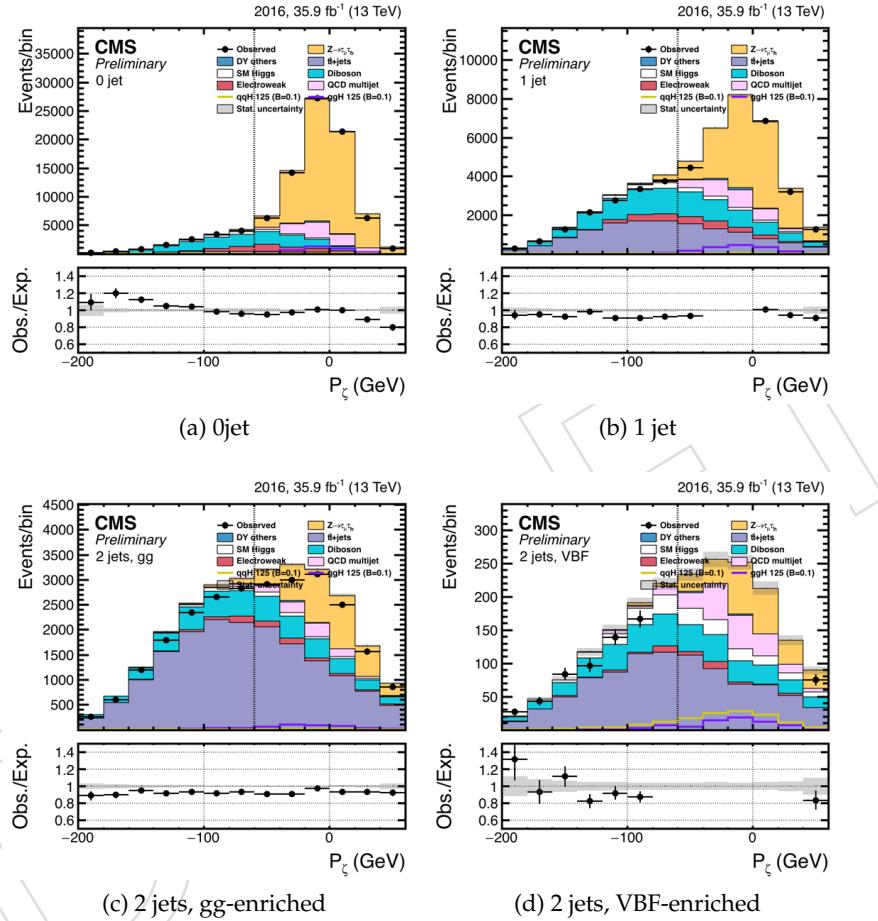


Figure 29:  $H \rightarrow e\tau_\mu$ :  $P_\zeta$  distributions in the four categories of the analysis. The optimized threshold for the cut is indicated by the vertical dashed line. The plots are prefit, and partially blinded according to the signal significance in each bin. Only the statistical uncertainty is shown.

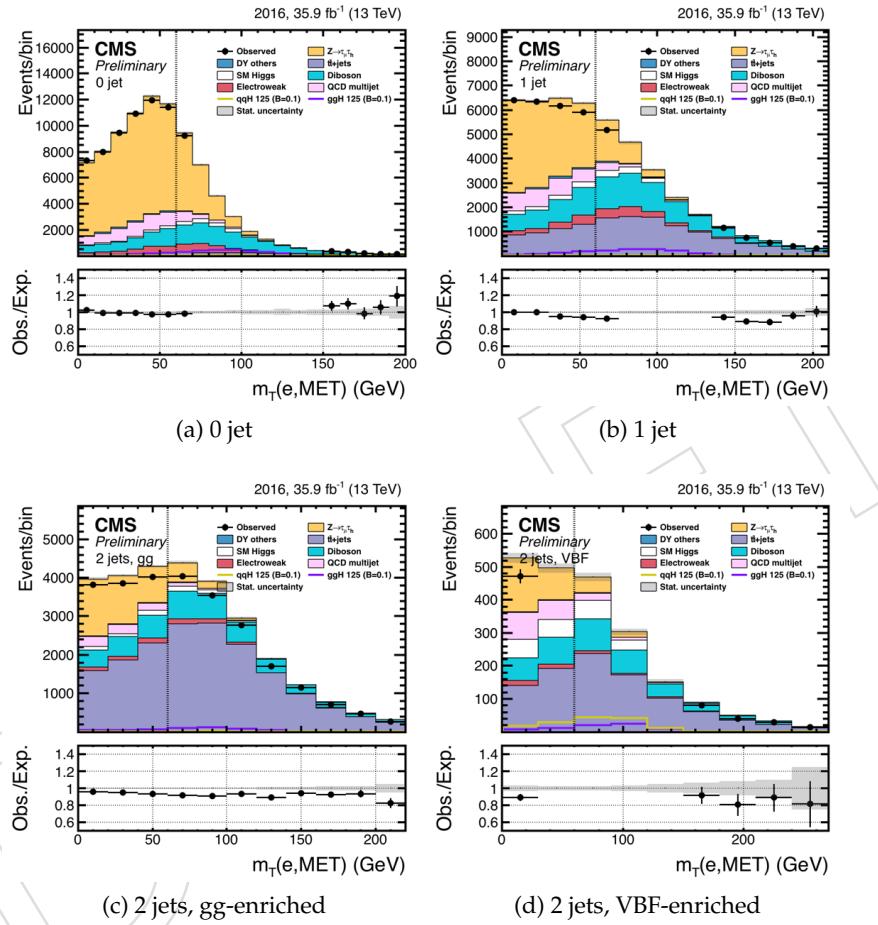


Figure 30:  $H \rightarrow e\tau_\mu$ :  $m_T(e, MET)$  distributions in the four categories of the analysis. The optimized threshold for the cut is indicated by the vertical dashed line. The plots are prefit, and partially blinded according to the signal significance in each bin. Only the statistical uncertainty is shown.

- Azimutal angle between the electron and the missing transverse energy,  $\Delta\phi(e, E_T^{\text{miss}})$ ;
- Azimutal angle between the muon and the missing transverse energy,  $\Delta\phi(\mu, E_T^{\text{miss}})$ ;
- Collinear mass,  $M_{\text{col}}$ .
- Visible mass, .
- Muon  $p_T$ .
- Electron  $p_T$ .

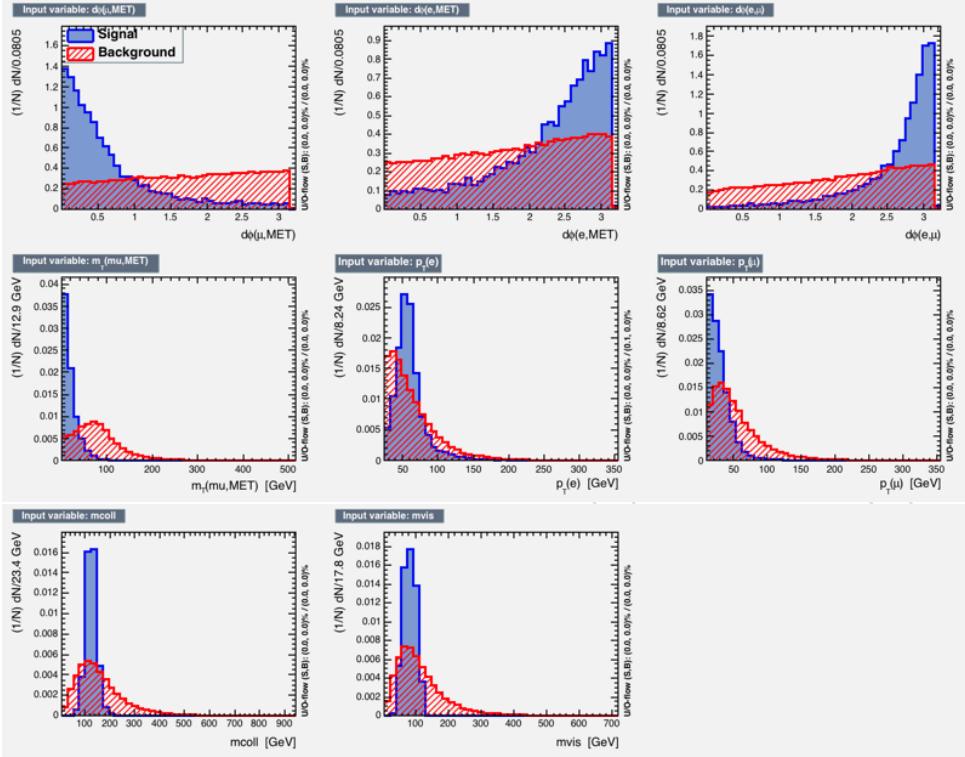
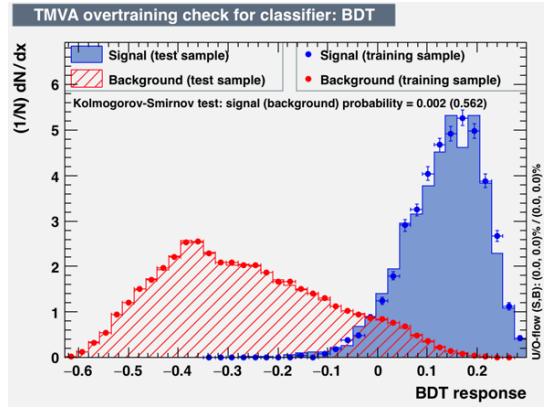
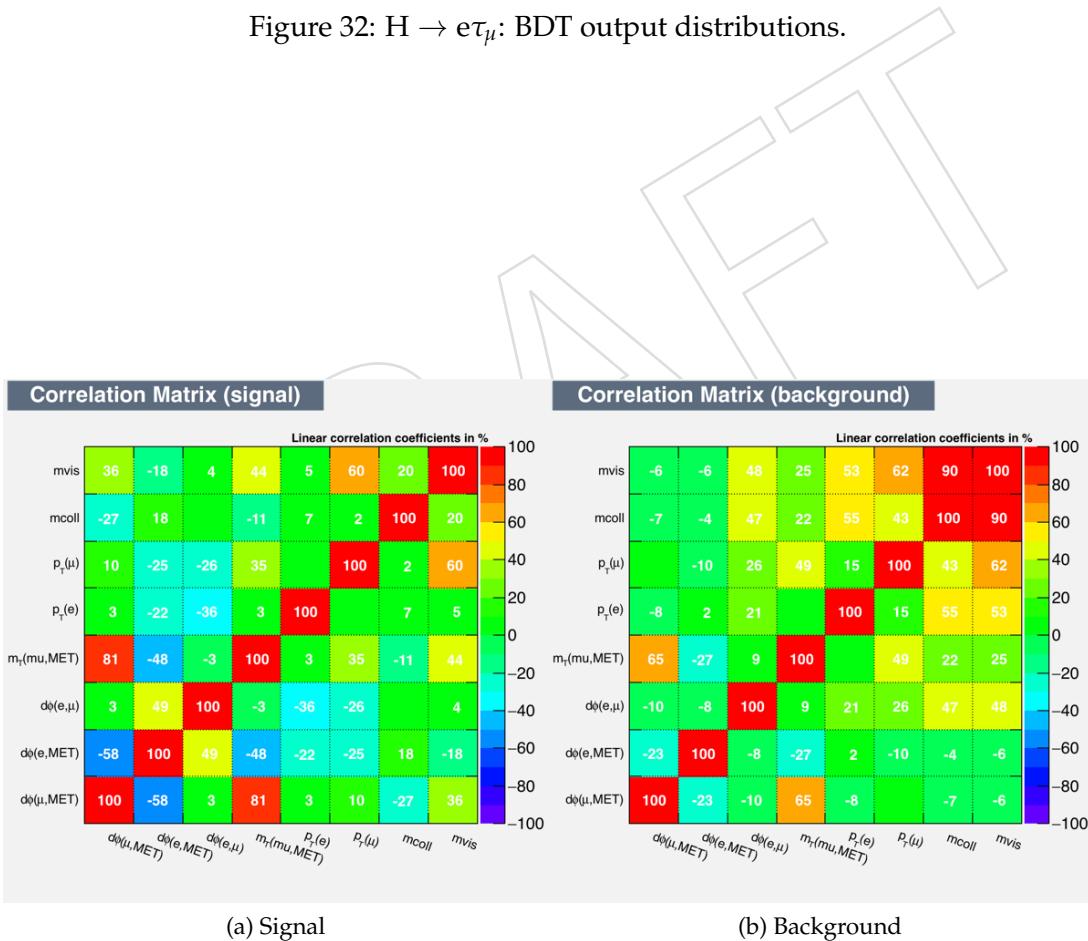


Figure 31:  $H \rightarrow e\tau_\mu$ : Normalized distributions of the variables that enter the BDT for the  $e\tau_\mu$  analysis. The signal (blue) is composed of a weighted mixture of ggH and VBF  $H \rightarrow e\tau$  events, whereas the background (red) is made of  $t\bar{t}$  events. The events used in the BDT pass the same  $p_T$ ,  $\eta$ , and isolation criteria as in the signal region.

The distributions of these variables for the signal and the reducible background are given in Fig. 31. No threshold requirement is made on the BDT output. The expected limits are estimated by fitting the BDT output distributions in the four jet categories. The BDT output distributions for the signal and  $t\bar{t}$  background are shown in Fig. 32. The figures also show the results of Kolmogorov-Smirnov tests to assess the compatibility of the distributions on the training sample and on a test sample composed of different events.

The correlation between the variables used in the BDT is shown in Fig. 33 for signal and background events.

The data vs simulation distributions of both BDT outputs are shown in Fig. 34. Good agreement is seen between observed data and predicted backgrounds in the unblinded regions with low signal significance.

Figure 32:  $H \rightarrow e\tau_\mu$ : BDT output distributions.

(a) Signal

(b) Background

Figure 33:  $H \rightarrow e\tau_\mu$ : Correlation between the variables used in the BDT, for signal (left) and background (right) events.

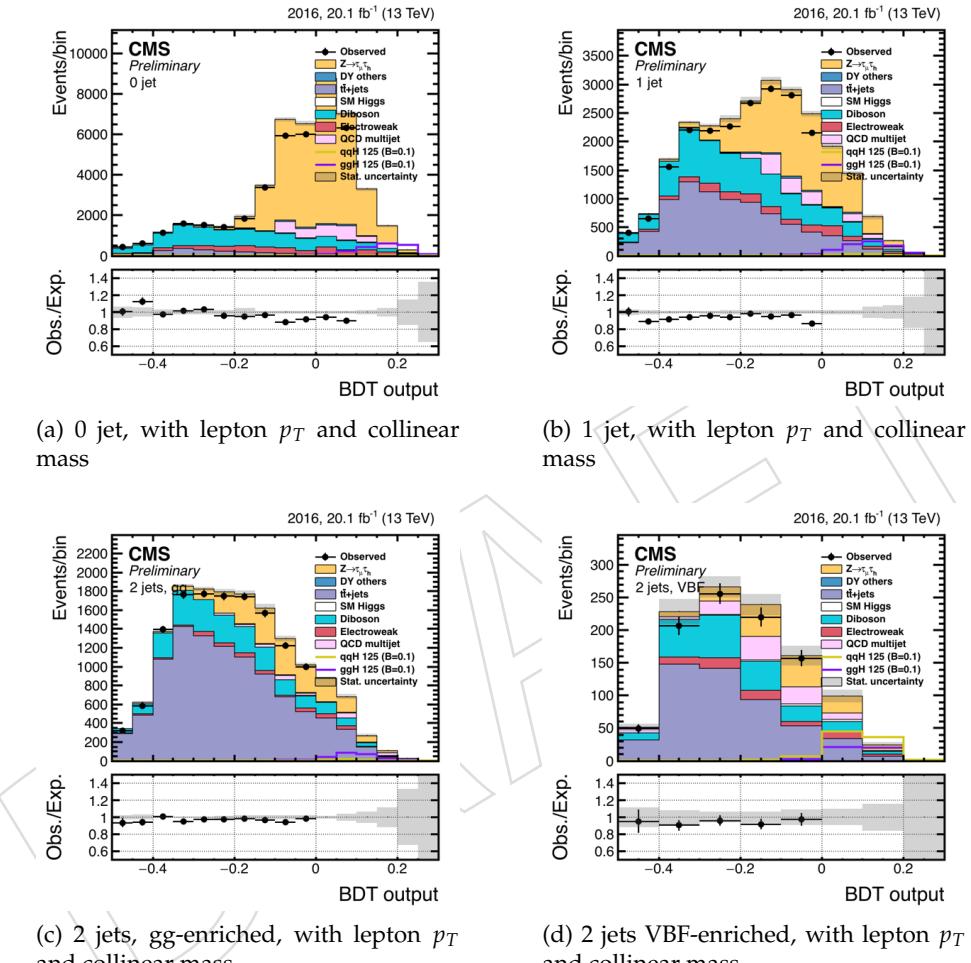


Figure 34:  $H \rightarrow e\tau_\mu$ : BDT output in the four categories for the training with the collinear mass and the lepton  $p_T$ , in the  $e\mu$  final state. The plots are partially blinded, and only the statistical uncertainty is shown in the ratio plots.

## 695 8.4 Background validation

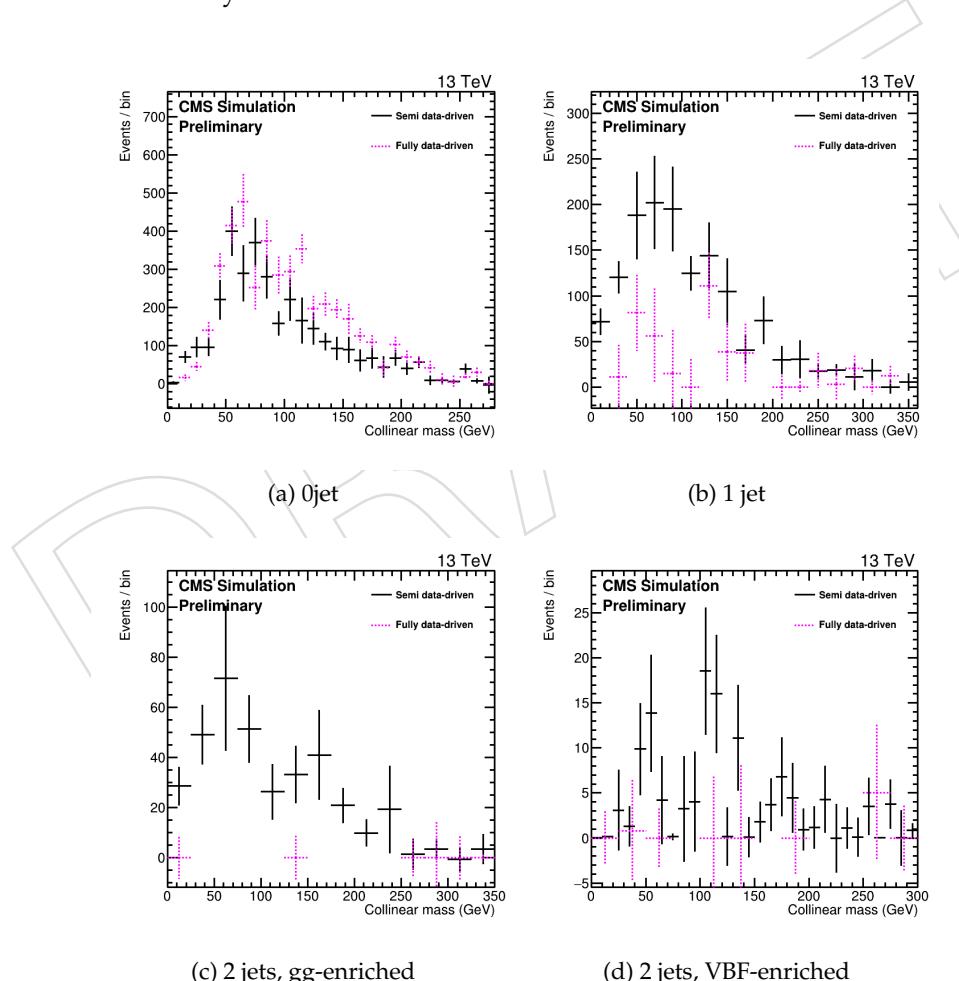
696 The misidentified lepton background in this channel is much less important than in the hadronic  
 697 final states. Two different methods are used to estimate this background. These are the fully  
 698 data driven and semi-data driven methods described in section ??.

### 700 Fully Data Driven

701 This background estimation is validated with a like-sign control region. Figure 35 shows the  
 702 data - background estimation comparison for the  $M_{col}$  distribution in the like sign control re-  
 703 gion. The agreement is relatively good.

### 704 Semi Data Driven

705 Figs. 35 compare the full and semi data driven background estimation. The two methods are in  
 706 agreement but the semi-data driven method gives greater statistical precision. As the reducible  
 707 background is quite small in the  $\mu e$  final state, the expected limits measured with both re-  
 708ducible background estimates are comparable between each other. The semi-data driven tech-  
 709 nique is used in the analysis.



708 Figure 35:  $H \rightarrow e\tau_\mu$ : Comparison of the reducible background estimates with the semi data-  
 709 driven (W from MC samples, QCD from SS data) and the fully data-driven (W and QCD from  
 fake rate method) methods, in the  $e\mu$  final state.

## 709 9 Systematic uncertainties

710 The systematic uncertainties are summarized in Table 6.

Table 6: Systematic uncertainties in the expected event yield. All uncertainties are treated as correlated between the categories, except those which have two values indicated. In this case the first value is correlated as above, while the second value (following the  $\oplus$  symbol) represents an uncorrelated uncertainty for each individual category. The total uncertainty in a given category is the sum in quadrature of the two values.

Systematic uncertainty	$H \rightarrow \mu\tau_e$	$H \rightarrow \mu\tau_h$	$H \rightarrow e\tau_\mu$	$H \rightarrow e\tau_h$
Muon trigger/ID/isolation	2%	2%	2%	-
Electron trigger/ID/isolation	2%	-	2%	2%
Hadronic $\tau$ efficiency	-	10%	-	10%
$\mu \rightarrow \tau_h$ fake rate	-	25%	-	-
$e \rightarrow \tau_h$ fake rate	-	-	-	12%
b-tagging veto	0.2-6.5%	0.2-6.5%	0.2-6.5%	-
Z+jets background	10% $\oplus$ 5%	10% $\oplus$ 5%	10% $\oplus$ 5%	10% $\oplus$ 5%
Misidentified $\tau_h$ background	-	30% $\oplus$ 10%	-	30% $\oplus$ 10%
W+jets estimation	10%	-	10%	-
QCD multijet estimation	30%	-	30%	-
Diboson background	5%	5%	5%	5%
t $\bar{t}$ background	6%	6%	6%	6%
W + $\gamma$ background	10% $\oplus$ 5%	-	10% $\oplus$ 5%	-
Single top production background	5%	5%	5%	5%
Jet energy scale	3-20%	3-20%	3-20%	3-20%
Hadronic $\tau$ energy scale	-	3%	-	3%
Electron energy scale	$\pm\sigma$	-	$\pm\sigma$	$\pm\sigma$
Muon energy scale	$\pm\sigma$	$\pm\sigma$	-	$\pm\sigma$
Unclustered energy scale	$\pm\sigma$	$\pm\sigma$	$\pm\sigma$	$\pm\sigma$
Theory uncertainty	10%	10%	10%	10%
Bin-by-bin	Shape	Shape	Shape	Shape
Luminosity	6.2%	6.2%	6.2%	6.2%

711 Systematic uncertainties affect the normalization or the shape of the distribution of the various  
 712 processes, and come from experimental or theoretical sources. Other uncertainties also affect  
 713 the distributions, and are detailed in the second part of this section.

714 

### 9.1 Luminosity

715 The uncertainty on the integrated luminosity collected by the CMS detector in 2016 amounts to  
 716 2.6%. It affects all processes with normalization taken directly from simulation.

717 

### 9.2 $\tau_h$ identification

718 The uncertainty in the data/MC  $\tau_h$  identification scale factor amounts to 5%. The  $\tau_h$  identifi-  
 719 cation uncertainty is applied to all events where the reconstructed  $\tau_h$  is matched to a  $\tau_h$  at  
 720 generated level.

### 721 **9.3 Muon triggering, identification and isolation**

722 The uncertainty in the muon selections (trigger, identification, and isolation efficiencies) is es-  
 723 timated using a tag and probe measurement in  $Z \rightarrow \mu\mu$  data [45]. It amounts to 2% per muon.

### 724 **9.4 Electron triggering, identification and isolation**

725 The uncertainty in the electron selections (trigger, identification, and isolation efficiencies) is es-  
 726 timated using a tag and probe measurement in  $Z \rightarrow ee$  data [46]. It amounts to 2% per  
 727 electron.

### 728 **$e/\mu \rightarrow \tau_h$ fake rates**

729 The uncertainty associated to the scale factors related to the rates with which electrons are  
 730 misidentified as  $\tau_h$  candidates depends on the pseudorapidity and ranges from 7 to 17%. An  
 731 inclusive uncertainty of 12% is associated to events where the reconstructed  $\tau_h$  is matched to  
 732 an electron at generated level.

733 The uncertainty associated to the scale factors related to the rates with which muons are misiden-  
 734 tified as  $\tau_h$  candidates depends on the pseudorapidity. An inclusive uncertainty of 25% is asso-  
 735 ciated to events where the reconstructed  $\tau_h$  is matched to a muon at generated level.

### 736 **9.6 Normalization of MC samples**

737 The uncertainty on the Drell-Yan normalization amounts to 10%. The data/MC agreement in  
 738 a  $Z \rightarrow \mu\mu$  control region has been checked with different  $p_T$  thresholds for the muons, and is  
 739 found to be compatible with 1.0 within the uncertainty. The uncertainty in the  $t\bar{t}$  normalization  
 740 amounts to 6% and comes from a recent CMS measurement. The uncertainty related to the  
 741 diboson, single top, and  $W\gamma$  backgrounds, is, respectively, 5, 5, and 10%.

### 742 **9.7 Reducible background from the fake rate method**

743 The uncertainty on the estimation of the reducible background with the fake rate method is  
 744 taken from the agreement in background-enriched control regions obtained by inverting the  
 745 charge requirement on the leptons, or by selecting events with a large transverse mass between  
 746 the leptons and the missing transverse energy. The final uncertainty amounts to 30% and is  
 747 uncorrelated between all channels, but correlated between categories of a same channel. An  
 748 additional 10% is added to partially uncorrelate it between categories in the same channel. The  
 749 uncertainty is quite conservative and gets constrained in the tails of the collinear mass distri-  
 750 butions, where the misidentified background dominates. No shape uncertainty is considered  
 751 for the jet  $\rightarrow e$  and jet  $\rightarrow \mu$  fakerate as it would be negligible with respect the bin by bin uncer-  
 752 tainty. The shape uncertainties are instead considered for the jet  $\rightarrow \tau$  fakerate: the parameters  
 753 of the fitted function are varied inside their error and the new fakerate values are applied to  
 754 obtain alternative collinear mass and BDT distributions.

### 755 **9.8 Theoretical uncertainties on the signal**

756 A 10% of uncertainties on the Higgs production cross sections has been added to take into  
 757 account the changes in acceptance when the factorization and the renormalization scale are  
 758 shifted. This 10% has been checked to be a good estimation using the 100 variation on the  
 759 PDF contained in NNPDF PDF set and adding the uncertainties due to the  $\alpha_s$  variations and  
 760 the uncertainties reported by the LHC Higgs cross section working group. The quoted 10%

	0 jet	1 jet	2 jets gg-enriched	2 jets VBF-enriched
Diboson	0.2%	1.4%	2.3%	2.2%
$t\bar{t}$	2.2%	4.8%	6.5%	5.1%

Table 7: Normalization uncertainties on the diboson and  $t\bar{t}$  backgrounds in the different categories of the  $e\mu$  final state, due to the veto of events with b-tagged jets. These uncertainties are uncorrelated. Uncertainties on other processes are negligible.

761 include also the scale variation of the renormalization and factorization scale. Details on the  
 762 estimation are described in Appendix A.

### 763 9.9 B-tagging efficiency

764 The data/MC scale factors for the b-tagging efficiency are varied within their uncertainties, and  
 765 the effect is propagated to the various processes. Modifying these scale factors does not affect  
 766 the collinear mass distributions, and this is considered as a normalization uncertainty. The  
 767 uncertainties on the b-tagging efficiency for real b jets and for light-flavor jets are considered as  
 768 uncorrelated. The uncertainties considered in the final states where a b jet veto is applied are  
 769 given in Tab. 7.

### 770 9.10 $\tau_h$ energy scale

771 The uncertainty on the  $\tau_h$  energy scale is 1.2%, as recommended by the TAU POG. It is uncor-  
 772 related based on the reconstructed decay mode. The uncertainty is propagated to the collinear  
 773 mass and BDT distributions. It also has an impact on the global acceptance of processes with  
 774 real  $\tau_h$ . The nominal collinear mass distribution together with the distributions corresponding  
 775 to  $\pm 1\sigma$  variation of the  $\tau_h$  energy scale, are shown in Fig. 36 for the signal in two categories  
 776 with different jet multiplicity.

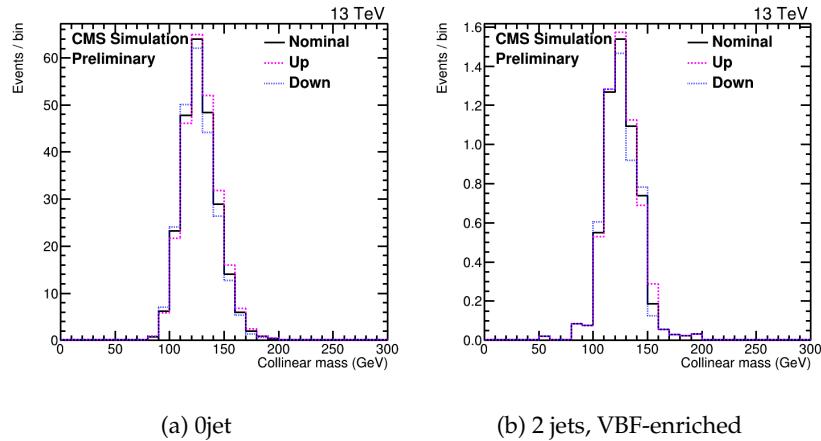


Figure 36: Nominal and  $\pm 1\sigma$  distributions of the collinear mass for the signal for the tau energy scale systematic uncertainty, in the 0-jet (left) and 2 jets VBF enriched (right) categories, for the  $e\tau_h$  final state.

### 777 9.11 $e/\mu \rightarrow \tau_h$ energy scale

778 The uncertainty in the energy scale of electrons and muons faking taus is, respectively, 3 and  
 779 1.5%.

## 9.12 Unclustered energy scale

The uncertainty is propagated to the collinear mass distributions. The nominal collinear mass distribution together with the distributions corresponding to  $\pm 1\sigma$  variation of the unclustered energy scale, are shown in Fig. 37 for the signal in two categories with different jet multiplicity.

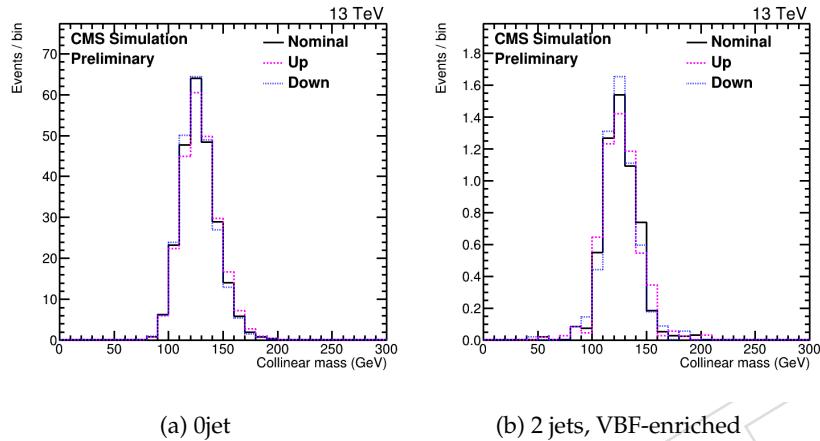


Figure 37: Nominal and  $\pm 1\sigma$  distributions of the collinear mass for the signal for the unclustered energy scale systematic uncertainty, in the 0-jet (left) and 2 jets VBF enriched (right) categories, for the  $e\tau_h$  final state.

## 9.13 Electron energy scale

It is computed event by event applying the energy smearing and regression from the EGamma POG. Both resolution and scale systematics are the result of the sum in quadrature of three components: electron selection, pseudorapidity, and R9 categorization. The four-momentum of the electron is re-computed adding or subtracting the systematics and the uncertainties propagated to the collinear mass and BDT distributions. The resolution systematics resulted to be negligible and thus not considered in the fit.

## 9.14 Muon energy scale

The uncertainty is propagated to the collinear mass and BDT distributions. It amounts to 0.2%, as prescribed by the Muon POG.

## 9.15 Jet energy scale

The shape and normalization uncertainties related to jet energy scale are computed for each background by calculating the differences in yield and shape when the jet energy scale is altered by  $\pm 1\sigma$ , which results in a shift of 5% to 10% in jet  $p_T$ . The nominal collinear mass distribution together with the distributions corresponding to  $\pm 1\sigma$  variation of the jet energy scale, are shown in Fig. 38 for the signal in two categories with different jet multiplicity.

## 9.16 Pile-up uncertainties

Shape uncertainties related to the pile up have been considered varying by 5% the minimum bias cross section in the computation of the pileup events in data, following the official prescription by CMS. The new values are then used to compute the weights to apply to the MC

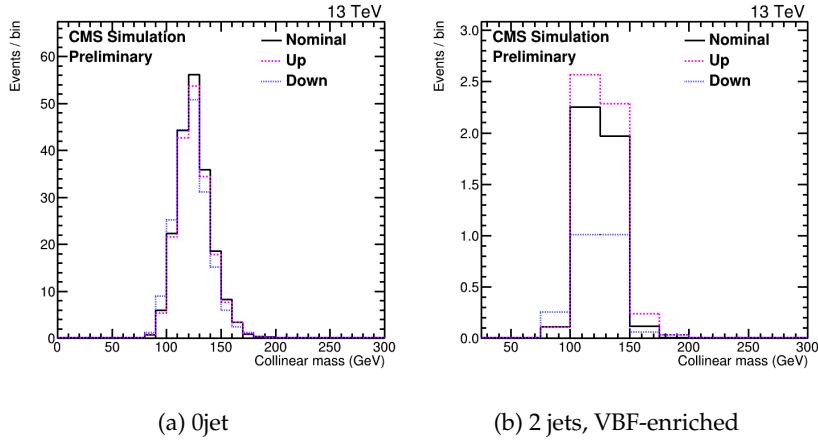


Figure 38: Nominal and  $\pm 1\sigma$  distributions of the collinear mass for the signal for the jet energy scale systematic uncertainty, in the 0-jet (left) and 2 jets VBF enriched (right) categories, for the  $e\tau_h$  final state.

804 samples and these are applied, event by event, to produced alternative collinear mass and BDT  
 805 distributions used as shape uncertainties in the fit.

### 806 9.17 Bin-by-bin uncertainties

807 Bin-by-bin uncertainties are considered to take into account the statistical uncertainties in every  
 808 bin of the distributions of every process. They are fully uncorrelated between bins, processes,  
 809 and categories.

### 810 9.18 Propagation of the uncertainties to the BDT output

811 The propagation of the uncertainties to the BDT output is done by propagating the various  
 812 uncertainties to the input variables, and recomputing the BDT output. It has been checked that  
 813 the response of the BDT varies linearly with the input variables (namely a 2-sigma shift of the  
 814 tau energy scale would give a 2-sigma shift of the BDT output).

## 815 10 Results

816 After applying the selection criteria, a maximum likelihood fit is performed in a final discrimination  
 817 variable to derive the expected limits or significance. Each systematic uncertainty is  
 818 used as a nuisance parameter in the fit. Two different fits are considered, with differences in  
 819 the discrimination variable employed and changes to the selection:

820 **Cut Based Selection** The  $M_{col}$  distribution is fit, after applying cut-based selection criteria.

821 **BDT Selection** The BDT discriminator is fit, after the loose selection. This BDT training in-  
 822 cludes the  $M_{col}$  variable, and the lepton  $p_T$ .

823 The fits are performed per channel and category, and then combined to set 95% CL upper  
 824 limits on the branching fraction of LFV H decay in the  $\mu\tau$  and  $e\tau$  channels,  $\mathcal{B}(H \rightarrow \mu\tau)$  and  
 825  $\mathcal{B}(H \rightarrow e\tau)$  respectively. A profiled likelihood method is used to derive all results assuming  
 826 the asymptotic approximation [47], and to set upper bounds on the branching fraction the CL<sub>s</sub>  
 827 method [34, 35] is used.

### 828 10.1 $H \rightarrow \mu\tau$ results

829 The distributions of the collinear mass  $M_{col}$  after the cut-based selection, compared to the signal  
 830 and background estimated contributions, are shown in Figures 39. Similarly, Figure 40 shows  
 831 the BDT discriminator distributions used in the BDT selection method, per jet category. The  
 832 median expected 95% CL upper limits on  $\mathcal{B}(H \rightarrow \mu\tau)$ , for a Higgs boson mass of 125 GeV, are  
 833 given for each category in Tables 8 (cut-based selection), and 9 (BDT selection). The limits are  
 834 also summarized graphically in Figure 41.

Table 8: The observed and expected upper limits and the best-fit branching fractions for different  $n$ -jet categories for the  $H \rightarrow \mu\tau$  process, after the cut-based selection.

Cut-Based  $H \rightarrow \mu\tau$  results

Expected 95% CL limits on BR					
	0-jet (%)	1-jet (%)	2-jets (%)	VBF (%)	Combined (%)
$\mu\tau_h$	< 1.13	< 1.19	< 2.15	< 1.59	< 0.68
$\mu\tau_e$	< 1.00	< 1.60	< 3.42	< 2.31	< 0.76
$\mu\tau$			< 0.51		

Expected significance for $BR(H \rightarrow \mu\tau) = 1\%$					
	0-jet (%)	1-jet (%)	2-jets (%)	VBF (%)	Combined (%)
$\mu\tau_h$	1.74	1.63	0.89	1.21	2.83
$\mu\tau_e$	1.96	1.32	0.62	0.71	2.57
$\mu\tau$			3.83		

### 835 10.2 $H \rightarrow e\tau$ results

836 The distributions of the collinear mass  $M_{col}$  after the cut-based selection, compared to the signal  
 837 and background estimated contributions, are shown in Figures 42. Similarly, Figure 43 shows  
 838 the BDT discriminator distributions used in the BDT selection method, per jet category. The  
 839 median expected 95% CL upper limits on  $\mathcal{B}(H \rightarrow \mu\tau)$ , for a Higgs boson mass of 125 GeV, are  
 840 given for each category in Tables 10 (cut-based selection), and 11 (BDT selection). The limits  
 841 are also summarized graphically in Figures 44.

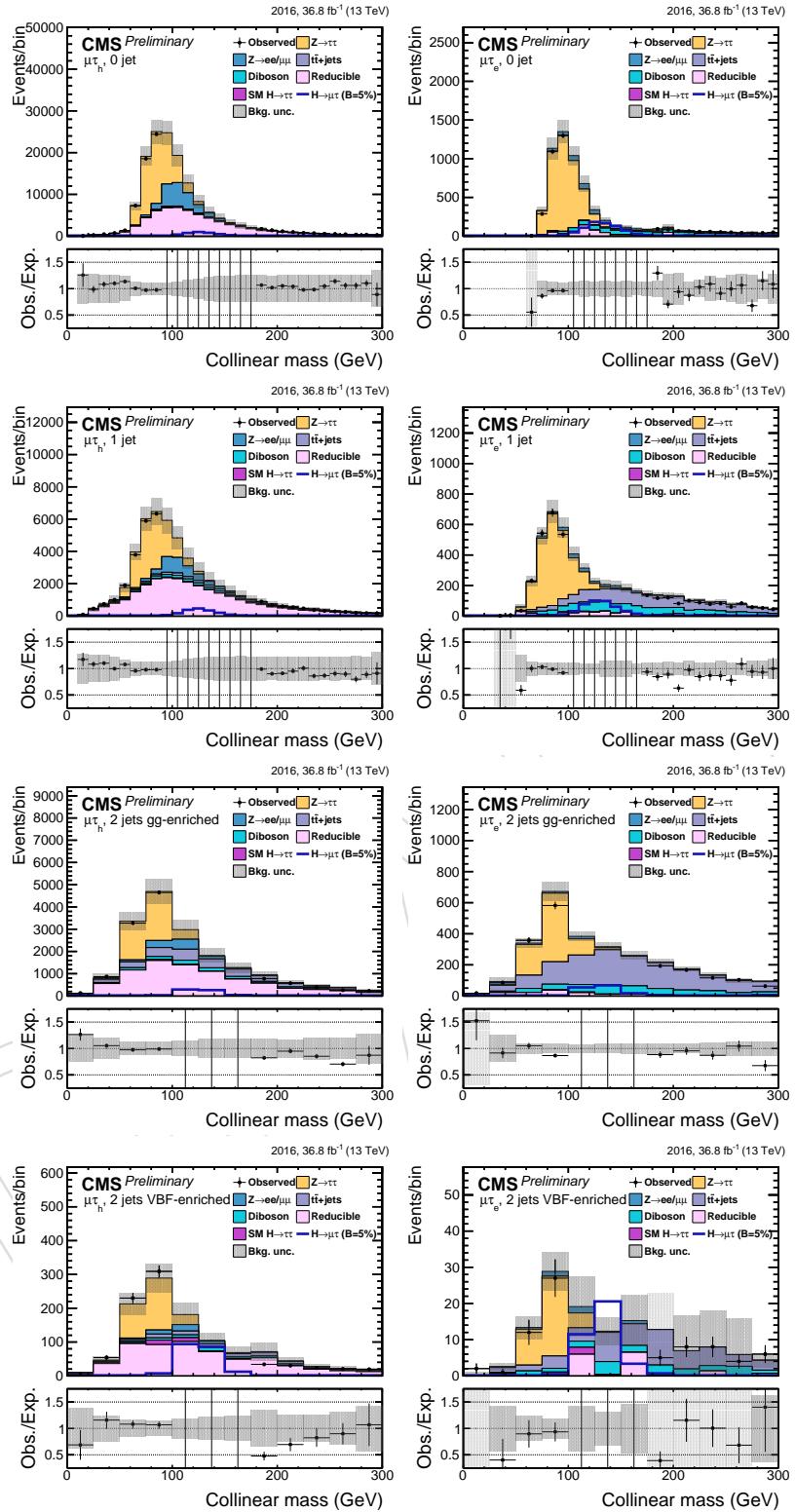


Figure 39: Distribution of the collinear mass  $M_{\text{col}}$  for the  $H \rightarrow \mu\tau$  process in the signal region after the Cut-Based selection, in the different channels and categories compared to the signal and background estimation. The background is normalized to the best-fit values from the signal plus background fit while the signal is normalized to  $\mathcal{B}(H \rightarrow \mu\tau) = 5\%$ . The lower panel in each plot shows the fractional difference between the observed data and the fitted background. The left column of plots correspond to the  $H \rightarrow \mu\tau_h$  categories, from 0-jets (first row) to VBF (fourth row). The right one to their  $H \rightarrow \mu\tau_e$  counterparts.

Table 9: The observed and expected upper limits and the best-fit branching fractions for different  $n$ -jet categories for the  $H \rightarrow \mu\tau$  process, after a BDT selection and a fit on the BDT discriminator (BDT Method).

BDT  $H \rightarrow \mu\tau$  results

Expected 95% CL limits on BR					
	0-jet (%)	1-jet (%)	2-jets (%)	VBF (%)	Combined (%)
$\mu\tau_h$	< 0.42	< 0.58	< 0.96	< 0.62	< 0.29
$\mu\tau_e$	< 0.79	< 1.09	< 2.37	< 1.71	< 0.56
$\mu\tau$			< 0.26		

Expected significance for $BR(H \rightarrow \mu\tau) = 1\%$					
	0-jet (%)	1-jet (%)	2-jets (%)	VBF (%)	Combined (%)
$\mu\tau_h$	4.85	3.31	2.08	2.97	6.79
$\mu\tau_e$	2.20	1.51	0.85	0.95	3.01
$\mu\tau$			7.42		

Table 10: The observed and expected upper limits and the best-fit branching fractions for different  $n$ -jet categories for the  $H \rightarrow e\tau$  process, after a cut-based selection.

Cut Based  $H \rightarrow e\tau$  results

Expected 95% CL limits on BR					
	0-jet (%)	1-jet (%)	2-jets (%)	VBF (%)	Combined (%)
$e\tau_h$	< 1.50	< 1.71	< 3.39	< 1.55	< 0.86
$e\tau_\mu$	< 0.97	< 1.20	< 3.61	< 2.66	< 0.72
$e\tau$			< 0.55		

Expected significance for $BR(H \rightarrow e\tau) = 1\%$					
	0-jet (%)	1-jet (%)	2-jets (%)	VBF (%)	Combined (%)
$e\tau_h$	1.62	1.21	0.64	1.15	2.46
$e\tau_\mu$	1.84	1.64	0.57	0.79	2.56
$e\tau$			3.58		

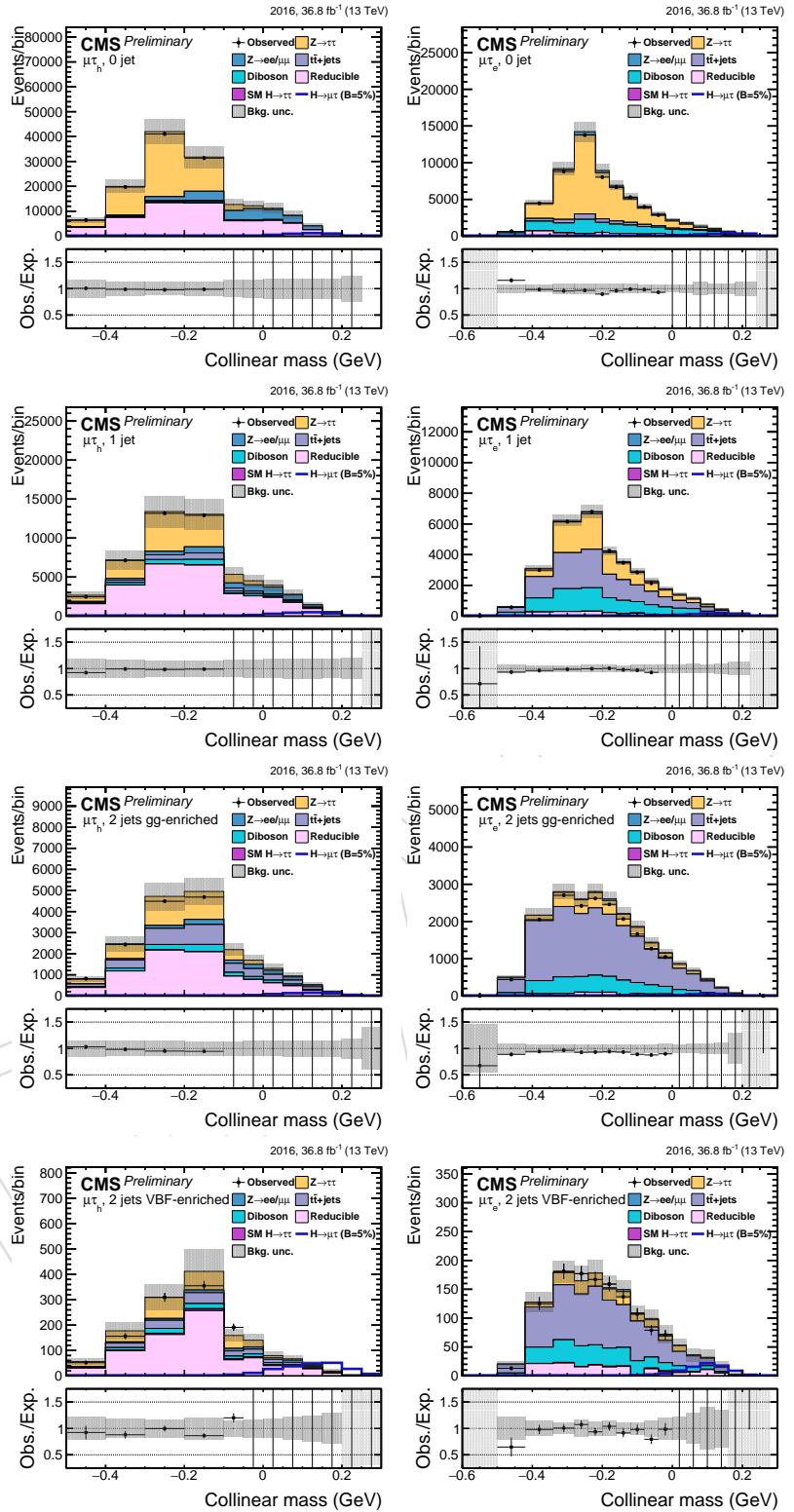


Figure 40: Distribution of the collinear mass  $M_{\text{col}}$  for the  $H \rightarrow \mu\tau$  process in the signal region in the BDT analysis, in the different channels and categories compared to the signal and background estimation. The background is normalized to the best-fit values from the signal plus background fit while the signal is normalized to  $\mathcal{B}(H \rightarrow \mu\tau) = 5\%$ . The lower panel in each plot shows the fractional difference between the observed data and the fitted background. The left column of plots correspond to the  $H \rightarrow \mu\tau_h$  categories, from 0-jets (first row) to VBF (fourth row). The right one to their  $H \rightarrow \mu\tau_e$  counterparts.

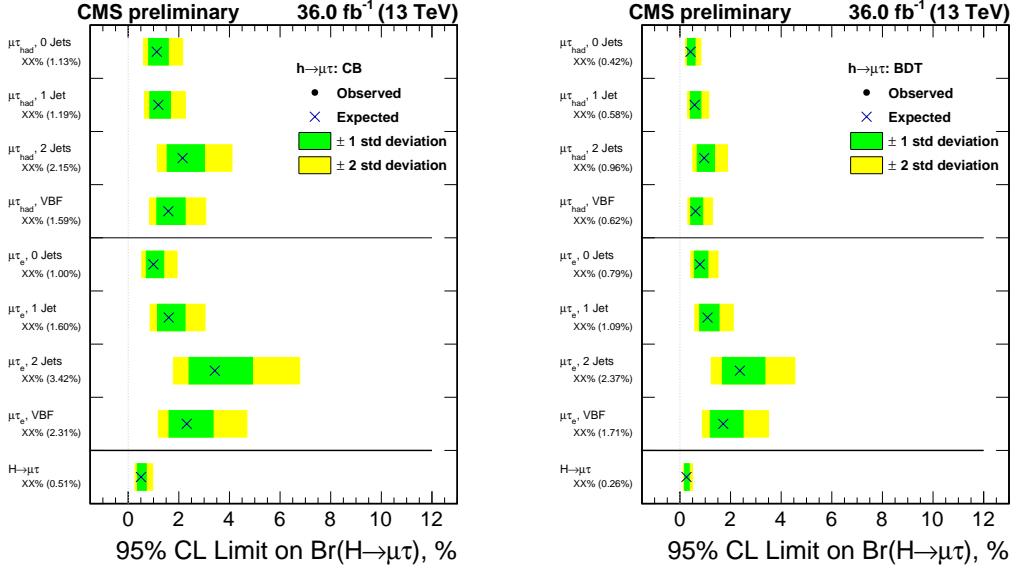


Figure 41: Observed and expected 95% CL upper limits on the  $\mathcal{B}(H \rightarrow \mu\tau)$  for each individual category and combined. Left: Cut-Based results. Right: BDT results.

Table 11: The observed and expected upper limits and the best-fit branching fractions for different  $n$ -jet categories for the  $H \rightarrow e\tau$  process, after the BDT analysis (fit on BDT discriminator).

#### BDT $H \rightarrow e\tau$ results

Expected 95% CL limits on BR					
	0-jet (%)	1-jet (%)	2-jets (%)	VBF (%)	Combined (%)
$e\tau_h$	< 0.56	< 0.98	< 1.62	< 0.69	< 0.38
$e\tau_\mu$	< 0.83	< 1.54	< 2.27	< 1.77	< 0.66
$e\tau$			< 0.32		

Expected significance for $\mathcal{B}(H \rightarrow e\tau)=1\%$					
	0-jet (%)	1-jet (%)	2-jets (%)	VBF (%)	Combined (%)
$e\tau_h$	2.69	1.83	1.30	2.84	4.44
$e\tau_\mu$	2.02	1.35	0.86	1.09	2.69
$e\tau$			5.44		

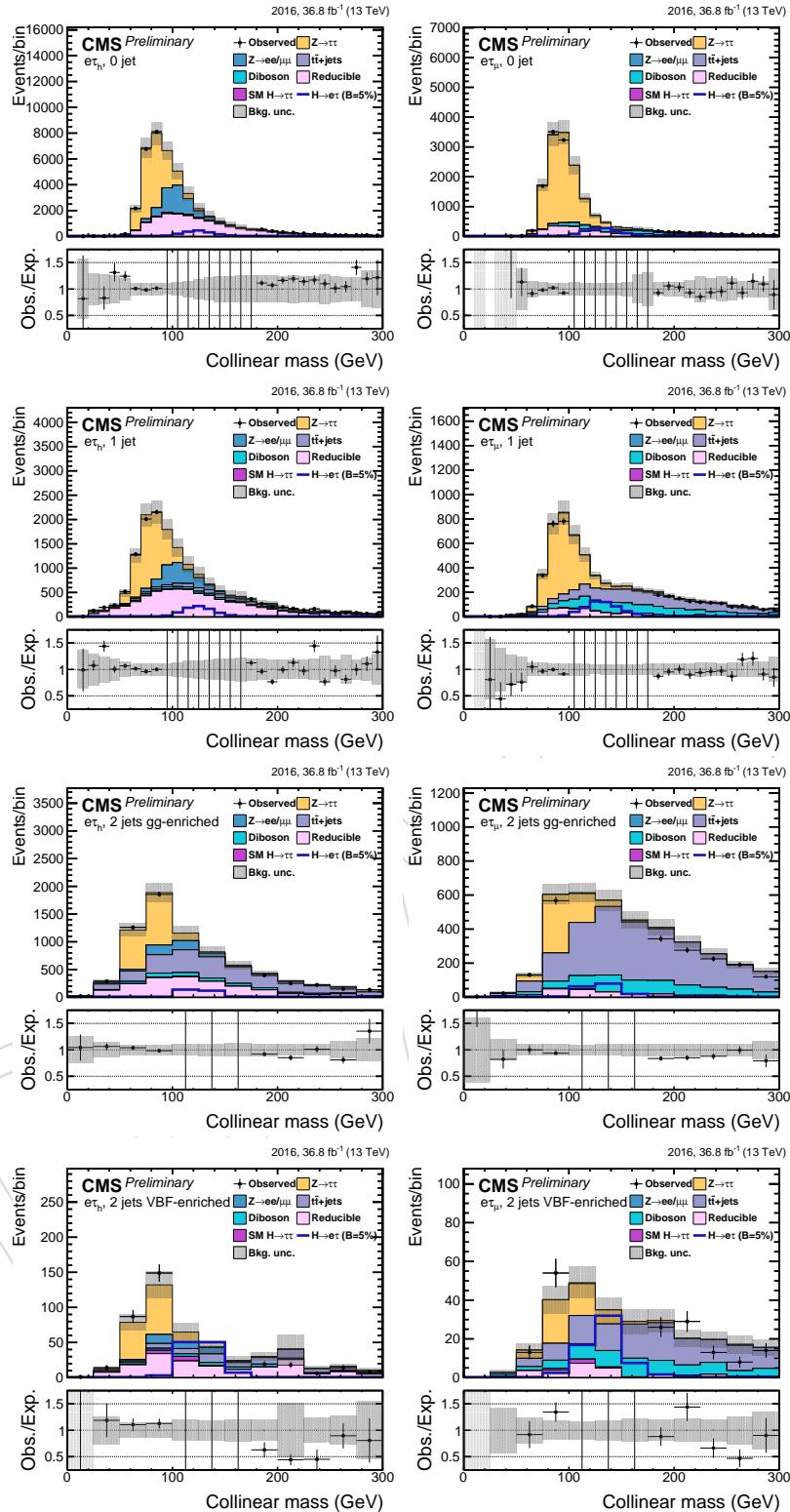


Figure 42: Distribution of the collinear mass  $M_{\text{col}}$  for the  $H \rightarrow e\tau$  process in the signal region after the Cut-Based selection, in the different channels and categories compared to the signal and background estimation. The background is normalized to the best-fit values from the signal plus background fit while the signal is normalized to  $\mathcal{B}(H \rightarrow e\tau) = 5\%$ . The lower panel in each plot shows the fractional difference between the observed data and the fitted background. The left column of plots correspond to the  $H \rightarrow e\tau_h$  categories, from 0-jets (first row) to VBF (fourth row). The right one to their  $H \rightarrow e\tau_\mu$  counterparts.

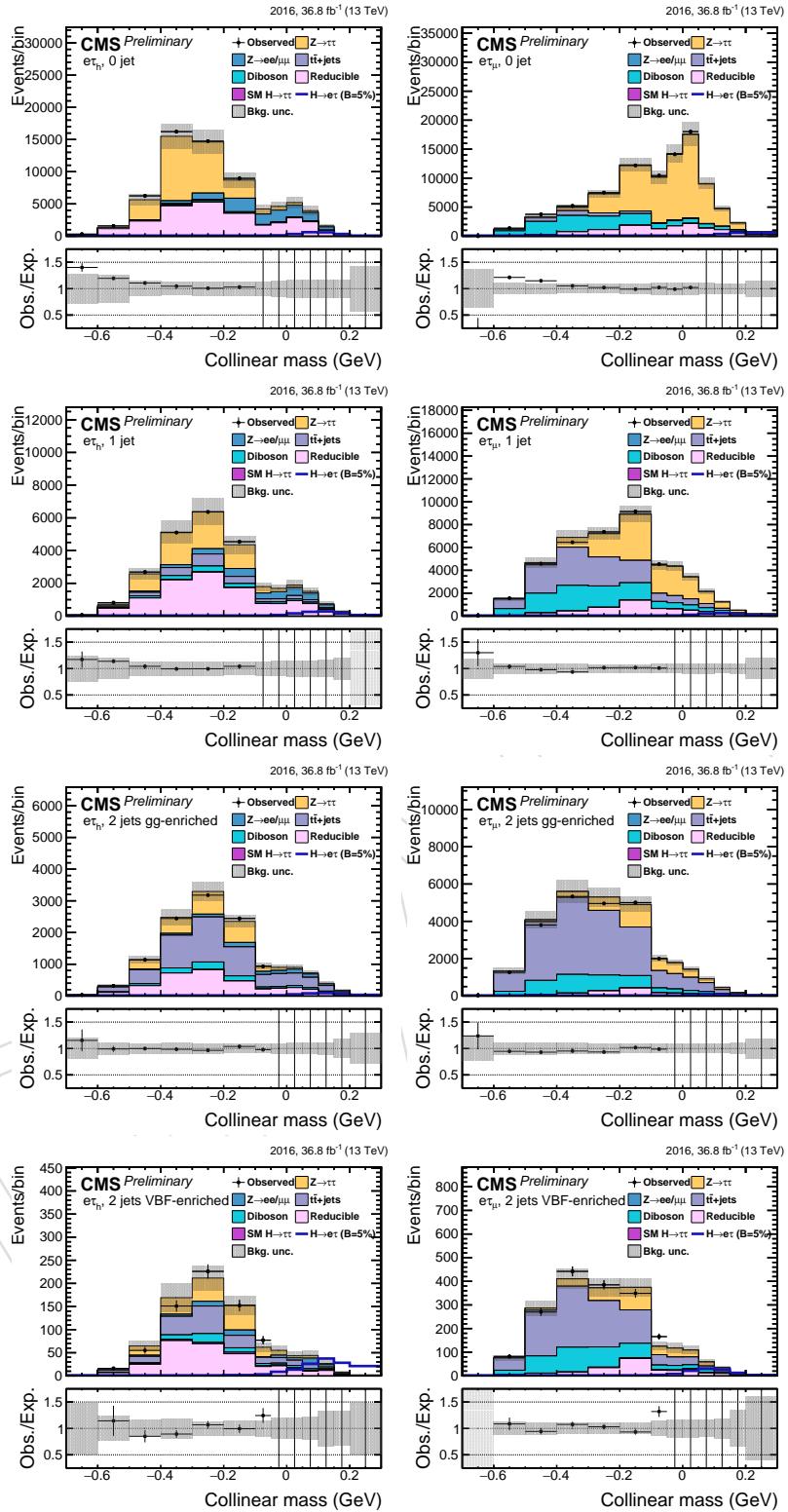


Figure 43: Distribution of the BDT output for the  $H \rightarrow e\tau$  process in the signal region in the BDT analysis, in the different channels and categories compared to the signal and background estimation. The background is normalized to the best-fit values from the signal plus background fit while the signal is normalized to  $\mathcal{B}(H \rightarrow e\tau) = 5\%$ . The lower panel in each plot shows the fractional difference between the observed data and the fitted background. The left column of plots correspond to the  $H \rightarrow e\tau_h$  categories, from 0-jets (first row) to VBF (fourth row). The right one to their  $H \rightarrow e\tau_\mu$  counterparts.

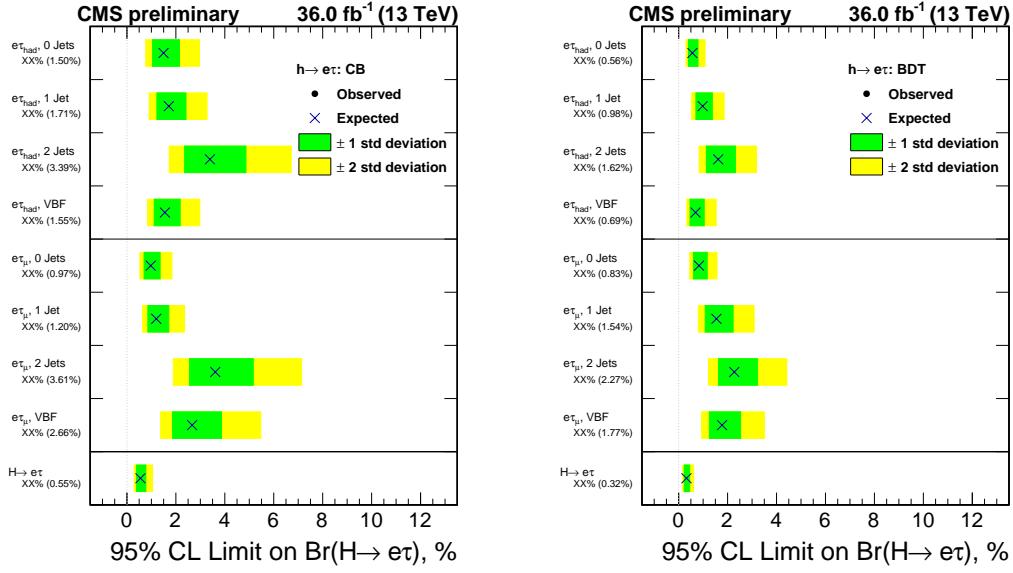


Figure 44: Observed and expected 95% CL upper limits on the  $\mathcal{B}(H \rightarrow \mu\tau)$  for each individual category and combined. Left: Cut-Based results. Right: BDT results.

### 10.3 Summary and comparison with past results

The expected limits obtained with the analyzed  $36 \text{ fb}^{-1}$  dataset at 13 TeV are significantly tighter than the 8 TeV [24, 48] and the early 13 TeV 2015 results [49].

Table 12 presents a comparison of the new expected limits on the 95% CL upper limits. The BDT is much more sensitive, with limits reduced by about a factor two, and it is chosen as the default method to extract the results. The cut-based analysis is still performed as a cross-check and its results will be quoted in the paper.

Table 12: The observed and expected upper limits and the best-fit branching fractions for the  $H \rightarrow \tau\tau$  and the for the  $H \rightarrow e\tau$  processes, comparing the different selections and signal extraction methods described (Cut-Based, BDT).

Expected 95% CL limits on BR		
	Cut Based (%)	BDT (%)
$H \rightarrow \mu\tau$	< 0.51	< 0.26
$H \rightarrow e\tau$	< 0.55	< 0.32
Expected significance for $\text{BR}(H \rightarrow \ell\tau)=1\%$		
	Cut Based	BDT
$H \rightarrow \mu\tau$	3.83	7.42
$H \rightarrow e\tau$	3.58	5.44

### 10.4 Limits on the Yukawa couplings

The constraints on  $\mathcal{B}(H \rightarrow \mu\tau)$  and  $\mathcal{B}(H \rightarrow e\tau)$  can be interpreted in terms of LFV Yukawa couplings [5]. The LFV decays  $e\tau$  and  $\mu\tau$  arise at tree level from the assumed flavour violating Yukawa interactions,  $Y_{\ell^\alpha \ell^\beta}$  where  $\ell^\alpha, \ell^\beta$  denote the leptons,  $\ell^\alpha, \ell^\beta = e, \mu, \tau$  and  $\ell^\alpha \neq \ell^\beta$ . The decay width  $\Gamma(H \rightarrow \ell^\alpha \ell^\beta)$  in terms of the Yukawa couplings is given by:

$$\Gamma(H \rightarrow \ell^\alpha \ell^\beta) = \frac{m_H}{8\pi} (|Y_{\ell^\beta \ell^\alpha}|^2 + |Y_{\ell^\alpha \ell^\beta}|^2),$$

and the branching fraction by:

$$B(H \rightarrow \ell^\alpha \ell^\beta) = \frac{\Gamma(H \rightarrow \ell^\alpha \ell^\beta)}{\Gamma(H \rightarrow \ell^\alpha \ell^\beta) + \Gamma_{SM}}.$$

- 850 The SM H decay width is assumed to be  $\Gamma_{SM} = 4.1 \text{ MeV}$  [50] for  $M_H = 125 \text{ GeV}$ .  
 851 The 95% CL constraint on the Yukawa couplings derived the expression for the branching frac-  
 852 tion above is:

- Cut-Based result:

$$\sqrt{|Y_{\mu\tau}|^2 + |Y_{\tau\mu}|^2} < 2.047 \times 10^{-3}, \sqrt{|Y_{e\tau}|^2 + |Y_{\tau e}|^2} < 2.141 \times 10^{-3}$$

- BDT result:

$$\sqrt{|Y_{\mu\tau}|^2 + |Y_{\tau\mu}|^2} < 1.462 \times 10^{-3}, \sqrt{|Y_{e\tau}|^2 + |Y_{\tau e}|^2} < 1.625 \times 10^{-3}$$

DRAFT

**853 11 Summary**

854 This note presents the search for LFV decays of the Higgs boson in the  $\mu\tau$  and  $e\tau$  final states,  
855 with the 2016 data collected by the CMS detector. This version shows the analysis with  $36 \text{ fb}^{-1}$ .  
856 The results are extracted by a fit to the output of BDT trained to discriminate the signal from  
857 backgrounds. The results are cross-checked with a cut-based analysis. The expected limits on  
858 the branching fraction of the Higgs boson to  $\mu\tau$  and to  $e\tau$  are expected to be less than 0.26 and  
859 0.32%, respectively, at 95% confidence level.

DRAFT

## 860 References

- [1] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys. Lett. B* **716** (2012) 1, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.
- [2] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Phys. Lett. B* **716** (2012) 30, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.
- [3] CMS Collaboration, "Observation of a new boson with mass near 125 GeV in pp collisions at  $\sqrt{s} = 7$  and 8 TeV", *JHEP* **06** (2013) 081, doi:10.1007/JHEP06(2013)081, arXiv:1303.4571.
- [4] ATLAS, CMS Collaboration, "Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at  $\sqrt{s} = 7$  and 8 TeV", *JHEP* **08** (2016) 045, doi:10.1007/JHEP08(2016)045, arXiv:1606.02266.
- [5] R. Harnik, J. Kopp, and J. Zupan, "Flavor violating Higgs decays", *JHEP* **03** (2013) 26, doi:10.1007/JHEP03(2013)026.
- [6] J. D. Bjorken and S. Weinberg, "Mechanism for Nonconservation of Muon Number", *Phys. Rev. Lett.* **38** (1977) 622, doi:10.1103/PhysRevLett.38.622.
- [7] J. L. Diaz-Cruz and J. Toscano, "Lepton flavor violating decays of Higgs bosons beyond the standard model", *Phys. Rev. D* **62** (2000) 116005, doi:10.1103/PhysRevD.62.116005, arXiv:hep-ph/9910233.
- [8] T. Han and D. Marfatia, " $h \rightarrow \mu\tau$  at Hadron Colliders", *Phys. Rev. Lett.* **86** (2001) 1442, doi:10.1103/PhysRevLett.86.1442, arXiv:hep-ph/0008141.
- [9] A. Arhrib, Y. Cheng, and O. C. Kong, "Comprehensive analysis on lepton flavor violating Higgs boson to  $\mu\bar{\tau} + \tau\bar{\mu}$  decay in supersymmetry without R parity", *Phys. Rev. D* **87** (2013) 015025, doi:10.1103/PhysRevD.87.015025, arXiv:1210.8241.
- [10] K. Agashe and R. Contino, "Composite Higgs-mediated flavor-changing neutral current", *Phys. Rev. D* **80** (2009) 075016, doi:10.1103/PhysRevD.80.075016, arXiv:0906.1542.
- [11] A. Azatov, M. Toharia, and L. Zhu, "Higgs mediated flavor changing neutral currents in warped extra dimensions", *Phys. Rev. D* **80** (2009) 035016, doi:10.1103/PhysRevD.80.035016, arXiv:0906.1990.
- [12] H. Ishimori et al., "Non-Abelian Discrete Symmetries in Particle Physics", *Prog. Theor. Phys. Suppl.* **183** (2010) 1, doi:10.1143/PTPS.183.1, arXiv:1003.3552.
- [13] G. Perez and L. Randall, "Natural neutrino masses and mixings from warped geometry", *JHEP* **01** (2009) 077, doi:10.1088/1126-6708/2009/01/077, arXiv:0805.4652.
- [14] S. Casagrande et al., "Flavor physics in the Randall-Sundrum model I. Theoretical setup and electroweak precision tests", *JHEP* **10** (2008) 094, doi:10.1088/1126-6708/2008/10/094, arXiv:0807.4937.

- [15] A. J. Buras, B. Duling, and S. Gori, "The impact of Kaluza-Klein fermions on Standard Model fermion couplings in a RS model with custodial protection", *JHEP* **09** (2009) 076, doi:10.1088/1126-6708/2009/09/076, arXiv:0905.2318.
- [16] M. Blanke et al., " $\Delta F = 2$  observables and fine-tuning in a warped extra dimension with custodial protection", *JHEP* **03** (2009) 001, doi:10.1088/1126-6708/2009/03/001, arXiv:0809.1073.
- [17] G. F. Giudice and O. Lebedev, "Higgs-dependent Yukawa couplings", *Phys. Lett. B* **665** (2008) 79, doi:10.1016/j.physletb.2008.05.062, arXiv:0804.1753.
- [18] J. Aguilar-Saavedra, "A minimal set of top-Higgs anomalous couplings", *Nucl. Phys. B* **821** (2009) 215, doi:10.1016/j.nuclphysb.2009.06.022, arXiv:0904.2387.
- [19] M. E. Albrecht et al., "Electroweak and flavour structure of a warped extra dimension with custodial protection", *JHEP* **09** (2009) 064, doi:10.1088/1126-6708/2009/09/064, arXiv:0903.2415.
- [20] A. Goudelis, O. Lebedev, and J. H. Park, "Higgs-induced lepton flavor violation", *Phys. Lett. B* **707** (2012) 369, doi:10.1016/j.physletb.2011.12.059, arXiv:1111.1715.
- [21] D. McKeen, M. Pospelov, and A. Ritz, "Modified Higgs branching ratios versus CP and lepton flavor violation", *Phys. Rev. D* **86** (2012) 113004, doi:10.1103/PhysRevD.86.113004, arXiv:1208.4597.
- [22] A. Pilaftsis, "Lepton flavour nonconservation in  $H^0$  decays", *Phys. Lett. B* **285** (1992) 68, doi:10.1016/0370-2693(92)91301-O.
- [23] J. G. Körner, A. Pilaftsis, and K. Schilcher, "Leptonic CP asymmetries in flavor-changing  $H^0$  decays", *Phys. Rev. D* **47** (1993) 1080, doi:10.1103/PhysRevD.47.1080.
- [24] CMS Collaboration, "Search for lepton-flavour-violating decays of the Higgs boson", *Phys. Lett. B* **749** (2015) 337, doi:10.1016/j.physletb.2015.07.053, arXiv:1502.07400.
- [25] ATLAS Collaboration, "Search for lepton-flavour-violating  $H \rightarrow \mu\tau$  decays of the Higgs boson with the ATLAS detector", (2015). arXiv:1508.03372. Submitted to JHEP.
- [26] ATLAS Collaboration, "Search for lepton-flavour-violating decays of the Higgs and Z bosons with the ATLAS detector", arXiv:1604.07730.
- [27] B. McWilliams and L.-F. Li, "Virtual effects of Higgs particles", *Nucl. Phys. B* **179** (1981) 62, doi:10.1016/0550-3213(81)90249-2.
- [28] O. U. Shanker, "Flavour violation, scalar particles and leptoquarks", *Nucl. Phys. B* **206** (1982) 253, doi:10.1016/0550-3213(82)90534-X.
- [29] G. Blankenburg, J. Ellis, and G. Isidori, "Flavour-changing decays of a 125 GeV Higgs-like particle", *Phys. Lett. B* **712** (2012) 386, doi:10.1016/j.physletb.2012.05.007, arXiv:1202.5704.
- [30] K. Olive et al., "Review of Particle Physics", *Chin. Phys. C* **38** (2014) 090001, doi:10.1088/1674-1137/38/9/090001.

- [31] A. Celis, V. Cirigliano, and E. Passemar, “Lepton flavor violation in the Higgs sector and the role of hadronic tau-lepton decays”, *Phys. Rev. D* **89** (2014) 013008, doi:10.1103/PhysRevD.89.013008, arXiv:1309.3564.
- [32] S. M. Barr and A. Zee, “Electric dipole moment of the electron and of the neutron”, *Phys. Rev. Lett.* **65** (1990) 21, doi:10.1103/PhysRevLett.65.21.
- [33] R. K. Ellis, I. Hinchliffe, M. Soldate, and J. van der Bij, “Higgs Decay to  $\tau^+\tau^-$ : A possible signature of intermediate mass Higgs bosons at high energy hadron colliders”, *Nucl. Phys. B* **297** (1988) 221, doi:10.1016/0550-3213(88)90019-3.
- [34] T. Junk, “Confidence level computation for combining searches with small statistics”, *Nucl. Instrum. Meth. A* **434** (1999) 435, doi:10.1016/S0168-9002(99)00498-2, arXiv:hep-ex/9902006.
- [35] A. L. Read, “Presentation of search results: the  $CL_s$  technique”, *J. Phys. G* **28** (2002) 2693, doi:10.1088/0954-3899/28/10/313.
- [36] CMS Collaboration, “Particle–Flow Event Reconstruction in CMS and Performance for Jets, Taus, and  $E_T^{\text{miss}}$ ”, *CMS Physics Analysis Summary CMS-PAS-PFT-09-001* (2009).
- [37] CMS Collaboration, “Commissioning of the particle–flow event reconstruction with leptons from  $J/\psi$  and  $W$  decays at 7 TeV”, *CMS Physics Analysis Summary CMS-PAS-PFT-10-003* (2010).
- [38] CMS Collaboration, “Reconstruction and identification of lepton decays to hadrons and at CMS”, *JINST* **11** (2016) P01019, doi:10.1088/1748-0221/11/01/P01019, arXiv:1510.07488.
- [39] CMS Collaboration, “Performance of reconstruction and identification of tau leptons in their decays to hadrons and tau neutrino in LHC Run-2”.
- [40] M. Cacciari and G. P. Salam, “Pileup subtraction using jet areas”, *Phys. Lett. B* **659** (2008) 119–126, doi:10.1016/j.physletb.2007.09.077, arXiv:0707.1378.
- [41] 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.
- [42] JetMET POG, “Jet Identification for the 13 TeV data Run2016”. <https://twiki.cern.ch/twiki/bin/view/CMS/JetID13TeVRun2016>.
- [43] CMS Collaboration, “MSSM Higgs to taunu search at 13 TeV using the 2015 data sample”, *CMS-PAS-HIG-16-006* (2016).
- [44] Top PAG. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/TopPtReweighting>.
- [45] CMS Collaboration, “Performance of CMS muon reconstruction in pp collision events at  $\sqrt{s} = 7 \text{ TeV}$ ”, (2012). arXiv:1206.4071. Submitted to *J. Inst.*
- [46] CMS Collaboration, “Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at  $\sqrt{s} = 8 \text{ TeV}$ ”, *JINST* **10** (2015) P06005, doi:10.1088/1748-0221/10/06/P06005, arXiv:1502.02701.
- [47] ATLAS and CMS Collaborations, LHC Higgs Combination Group, “Procedure for the LHC Higgs boson search combination in Summer 2011”, Technical Report ATL-PHYS-PUB 2011-11, CMS NOTE 2011/005, CERN, 2011.

- 978 [48] CMS Collaboration, “Search for lepton flavour violating decays of the Higgs boson to  $e\tau$   
979 and  $e\mu$  in protonproton collisions at  $\sqrt{s}=8$  TeV”, *Physics Letters B* **763** (2016) 472,  
980 doi:<http://dx.doi.org/10.1016/j.physletb.2016.09.062>.
- 981 [49] CMS Collaboration Collaboration, “Search for Lepton Flavour Violating Decays of the  
982 Higgs Boson in the mu-tau final state at 13 TeV”, Technical Report CMS-PAS-HIG-16-005,  
983 CERN, Geneva, 2016.
- 984 [50] A. Denner et al., “Standard model Higgs-boson branching ratios with uncertainties”,  
985 *Eur. Phys. J. C* **71** (2011) 1753, doi:[10.1140/epjc/s10052-011-1753-8](https://doi.org/10.1140/epjc/s10052-011-1753-8),  
986 arXiv:[1107.5909](https://arxiv.org/abs/1107.5909).

DRAFT

## 987 A Theoretical Uncertainties

988 Evaluation of the theoretical uncertainties is one of the important aspect of the Higgs analysis  
 989 either for the limit calculation or the signal extraction. The theoretical uncertainty arises  
 990 from our limited knowledge on the Higgs cross-section and/or on the signal acceptance. The  
 991 typical sources that can affect the former are the Parton Distribution Function (PDF) and the  
 992 scale choice such as the renormalization and the factorization scale, while in the latter case it  
 993 is the parton shower modeling. There is also an uncertainty arising from the choice of the  $p_T$   
 994 matching threshold by which the matrix element and the parton shower are treated separately.  
 995 The theoretical uncertainties are directly translated into the uncertainty on the yields which is  
 996 assumed to be 100% correlated among all event categories.

Note that all the uncertainties shown in the following paragraphs are on the signal acceptance which is

$$A = \frac{\text{# of events passing the category selection}}{\text{#of inclusive events}} \quad (2)$$

997 The lepton flavour violation samples  $H \rightarrow \mu\tau$  has been used to estimate these uncertainties.  
 998 The estimation has been developed for the analysis of the 2015 dataset and it is based on MC  
 999 samples simulated and analysed with CMSSW\_7\_6\_3. A set of cut similar to the analysis selec-  
 1000 tion is applied on reco quantites and it has been cross check using generation level information  
 1001 of the selected reconstructed candidate (MC truth matching). A global muon is required to have  
 1002  $p_T > 25$  GeV for  $H \rightarrow \mu\tau_e$  or  $30$  GeV for  $H \rightarrow \mu\tau_h$ ,  $|\eta| < 2.1$ , to pass the tight ID from the Muon  
 1003 POG, to come from the primary vertex  $|d_{xy}| < 0.2$ , and to be isolated ( $RelIso < 0.15$ ). The  $\tau$   
 1004 has to be identified using the DecayModeFinding discriminator, againstElectronMediumMVA6  
 1005 and againstMuonTight3; it is required having  $p_T > 30$  GeV and  $|\eta| < 2.3$ . The electron from  $\tau_e$   
 1006 have  $p_T > 10$  GeV,  $|\eta| < 2.3$ ,  $|d_{xy}| < 0.2$ ,  $RelIso < 0.1$ , and LooseID (WP90).

### 1007 A.1 PDF uncertainty

1008 The PDF uncertainty is determined by calculating the acceptance using the full NLO calculation  
 1009 following the PDF4LHC[?] recommendations for Run2.  
 1010 It has been estimated using the 100 variations of NNPDF30 PDF set and looking the effects on  
 1011 acceptance. As not all the acceptance distribution are gaussian, the width that contains the 68%  
 1012 of the values has been considered. Table 13 reports the uncertainties due to the PDF variations  
 1013 together with the uncertainties due to  $\alpha_s$  variation, LHC Higgs cross section WG uncertainties  
 and the total PDF uncertainties obtained summing the three contributions.

	gluon gluon fusion			vector boson fusion		
	0-jet	1-jet	2-jets	0-jet	1-jet	2-jets
NNPDF30	-0.4/0.4%	-0.3/0.3%	-0.2/0.1%	-0.3/0.4%	-0.3/0.4 %	-0.5/0.3%
$\alpha_s$	0.2%	0.1%	0.01%	0.1%	0.2 %	0.1%
LHC H x-sec WG	3.1%	3.1%	3.1%	2.1%	2.1 %	2.1%
total PDF uncertainty	-3.5/3.5%	-3.4/3.4%	-3.3/3.2%	-2.4/2.5%	-2.5/2.5 %	-2.6/2.4%

Table 13: Uncertainties due to the PDF variation

1015 **A.2 Scale uncertainty**

The scale uncertainty is determined by calculating the half of the difference between the maximum and minimum acceptance divided by acceptance obtained with the default scale.

$$\text{scale uncertainty} = \frac{A_{\max} - A_{\min}}{2} \cdot \frac{1}{A_{\mu_R=\mu_F=\text{default}}} \quad (3)$$

1016 The obtained scale uncertainty are listed in table 14.

These estimation are consistent with the 10% values quoted as systematics in the analyses

	gluon gluon fusion			vector boson fusion		
	0-jet	1-jet	2-jets	0-jet	1-jet	2-jets
scale variation	1.3 %	0.6 %	2.4%	0.7%	0.4 %	0.6%
LHC H x-sec WG	-8.1/7.6 %	-8.1/7.6 %	-8.1/7.6%	-0.3/0.4%	-0.3/0.4%	-0.3/0.4%
total scale variation	-9.4/8.9%	-8.7/8.2%	-10.5/10.0%	-1.0/1.1%	-0.7/0.8%	-0.9/1.0%

Table 14: Scale uncertainties

1017

1018 described in this note.

1019 **B BDT method A**

1020 Two different BDT methods have been exploited at different stages of the analysis, called BDT  
1021 method A and BDT method B. Method B resulted in the better sensitivity and has been chosen  
1022 as strategy for the result.

1023 The main difference between the two methods is the input variable set and the variable used  
1024 for the fit. In method A the input variable sets do not include the  $p_T$  of the product of the Higgs  
1025 decay and the collinear or visible mass. The  $p_T$  rule has been chosen to avoid the background  
1026 peaking in the signal region. Not using the mass as input variable lets it available to be fitted.  
1027 The BDT output is used as discriminating variable and the distribution of the collinear mass  
1028 of the events passing the cut on the BDT output is fitted to extract the limit. The sensitivity of  
1029 method A has been found to be comparable with the cut based analysis.

1030 **B.1  $H \rightarrow \mu\tau_h$**

1031 The input variables to the BDT are:

- 1032 • Transverse mass between the  $\tau_h$  and  $E_T^{\text{miss}}$   $M_T(\tau_h, E_T^{\text{miss}})$ .
- 1033 • Missing transverse energy,  $E_T^{\text{miss}}$ .
- 1034 • Pseudorapidity difference between the muon and the  $\tau_h$  candidate,  $\Delta\eta(e, \tau_h)$ .
- 1035 • Azimuthal angle between the muon and the  $\tau_h$ ,  $\Delta\phi(\mu, \tau_h)$ .
- 1036 • Azimuthal angle between the  $\tau_h$  and the  $E_T^{\text{miss}}$ ,  $\Delta\phi(\tau_h, E_T^{\text{miss}})$ .
- 1037 • Lepton asymmetry;

$$A = \frac{p_T^\mu - p_T^{\tau_h}}{p_T^\mu + p_T^{\tau_h}}. \quad (4)$$

1037 The distribution of signal and background events for each variable is shown in Fig. 45. A  
1038 threshold requirement is made on the the BDT output ( $> -0.01$ ) and the limit is estimated by  
1039 a fit to the  $M_{\text{col}}$  distribution. The lepton  $p_T$  is not included to avoid biasing the shape of the  
1040 background.

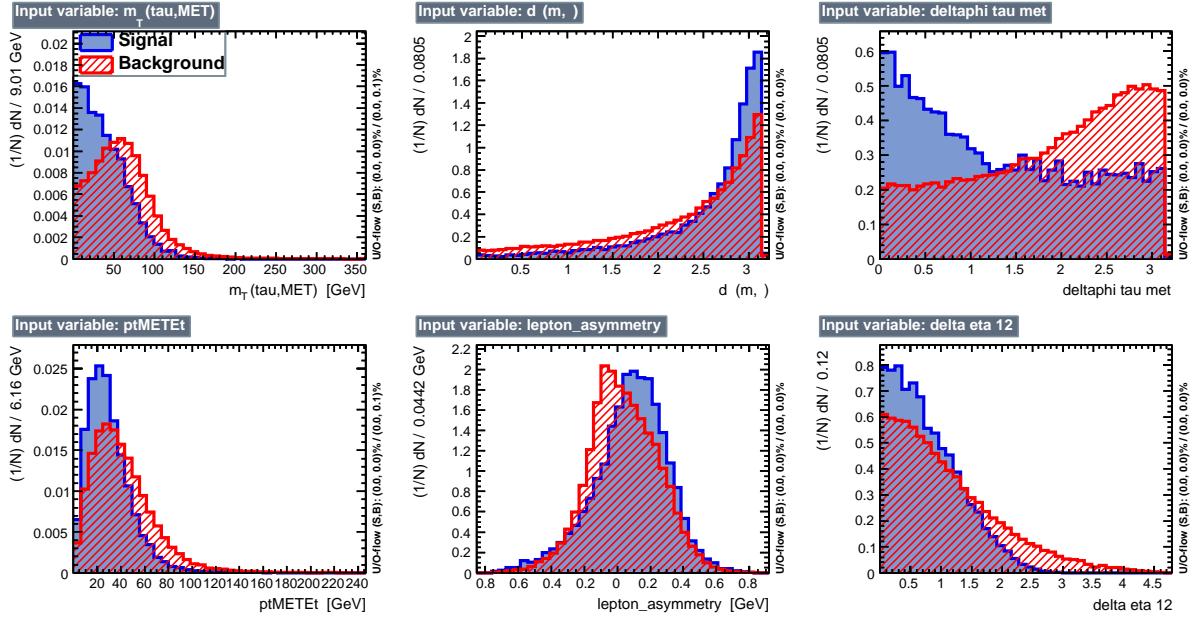


Figure 45:  $H \rightarrow \mu\tau_h$ : BDT input variable distributions. Signal distributions are drawn in blue and the background distributions in red.

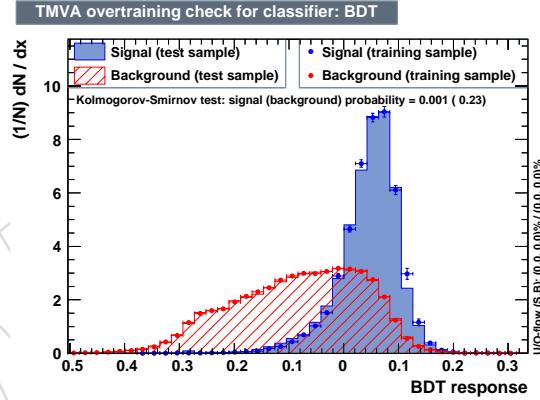


Figure 46:  $H \rightarrow \mu\tau_h$ : BDT output distribution

## B.2 $H \rightarrow e\tau_h$

The variables included in the BDT are:

- Transverse mass between the  $\tau_h$  and the  $E_T^{\text{miss}}$ ,  $M_T(\tau_h, E_T^{\text{miss}})$ .
- Missing transverse energy,  $E_T^{\text{miss}}$ .
- Pseudorapidity difference between the electron and the  $\tau_h$  candidate,  $\Delta\eta(e, \tau_h)$ .
- Azimuthal angle between the electron and the  $\tau_h$ ,  $\Delta\phi(e, \tau_h)$ .
- Azimuthal angle between the  $\tau_h$  and the  $E_T^{\text{miss}}$ ,  $\Delta\phi(\tau_h, E_T^{\text{miss}})$ .

The distribution of signal and background events for each variable is shown in Fig. 12. A threshold requirement is made on the the BDT output ( $> xxx$ ) and the limit is estimated by a fit to the  $M_{\text{col}}$  distribution. The lepton  $p_T$  is not included to avoid biasing the shape of the background. The BDT output distributions for the signal and  $t\bar{t}$  background are shown in

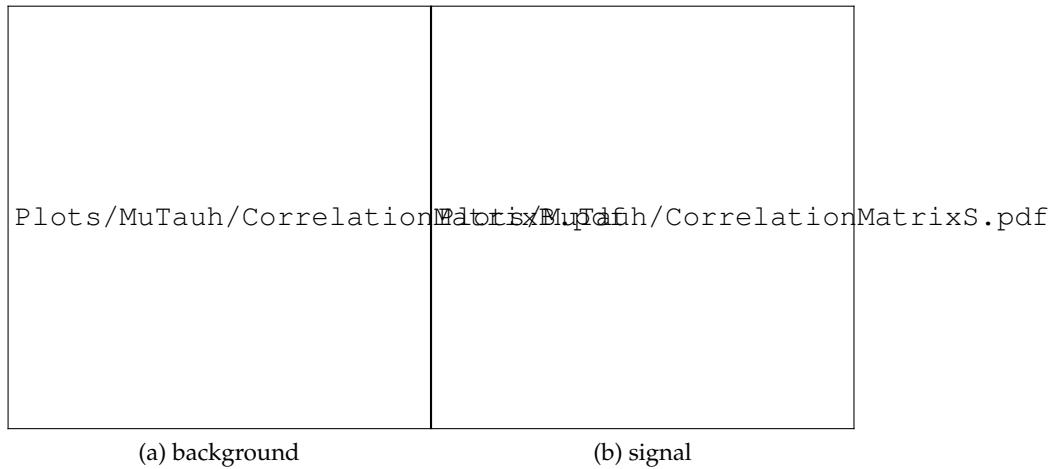


Figure 47:  $H \rightarrow \mu\tau_h$ : BDT input variable correlations for background (a) and signal (b).

Fig. 48. The data/MC distributions are shown in Fig. 49.

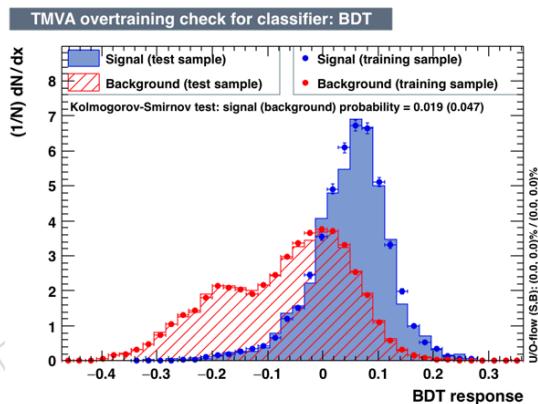


Figure 48:  $H \rightarrow e\tau_h$ : BDT output distributions

### B.3 $H \rightarrow \mu\tau_e$

The input variables of this BDT, shown in Fig. 18, are the following:

- Transverse mass between the  $\mu$  and  $E_T^{\text{miss}}, M_T(\mu, E_T^{\text{miss}})$ .
  - Transverse mass between the  $e$  and  $E_T^{\text{miss}}, M_T(e, E_T^{\text{miss}})$ .
  - Azimuthal angle between the  $e$  and  $\mu, \Delta\phi(e, \mu)$ .
  - Azimuthal angle between the  $e$  and  $E_T^{\text{miss}}, \Delta\phi(e, E_T^{\text{miss}})$ .
  - Azimuthal angle between the  $\mu$  and  $E_T^{\text{miss}}, \Delta\phi(\mu, E_T^{\text{miss}})$ .

#### B.4 $H \rightarrow e\tau_\mu$

The variables included in the BDT are:

- Transverse mass between the muon and the  $E_T^{\text{miss}}$ ,  $m_T(\mu, E_T^{\text{miss}})$ ;
  - Azimuthal angle between the electron and the muon,  $\Delta\phi(e, \mu)$ ;
  - Azimuthal angle between the electron and the missing transverse energy,  $\Delta\phi(e, E_T^{\text{miss}})$ ;

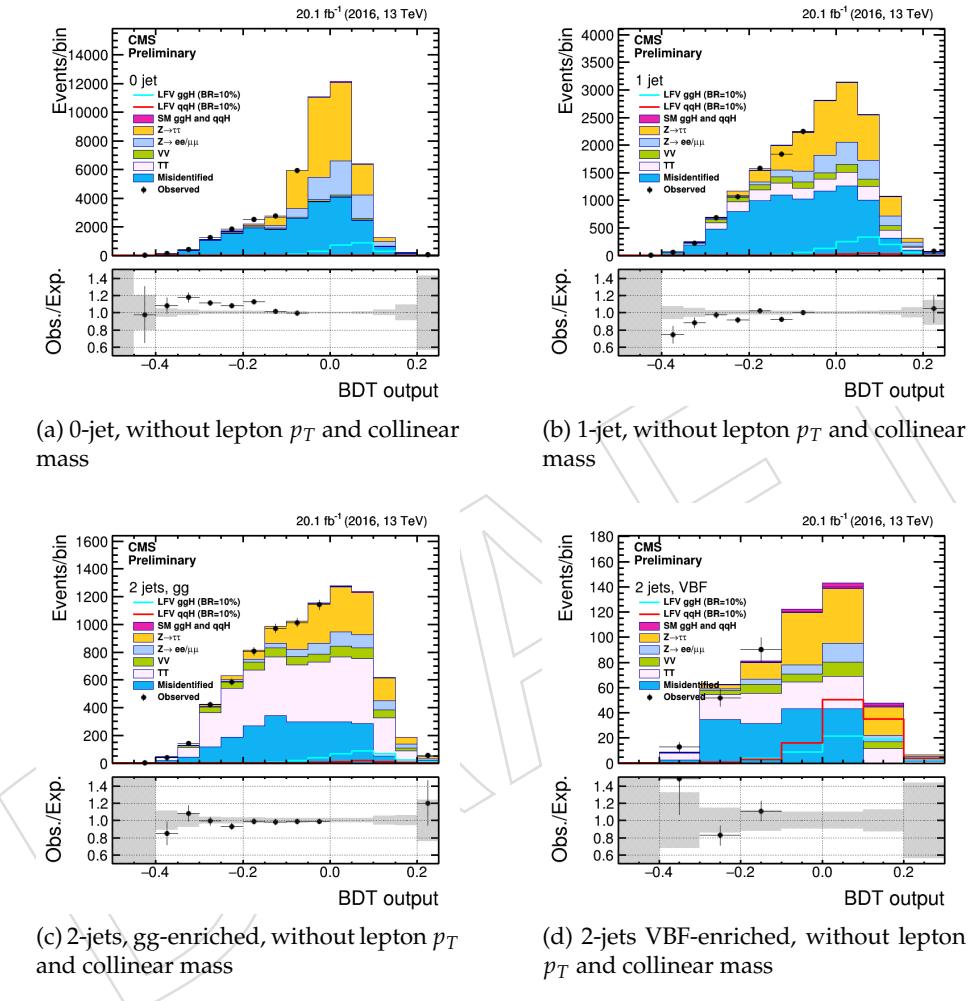


Figure 49:  $H \rightarrow e\tau_\mu$ : BDT output in the four categories for the training without the collinear mass and the lepton  $p_T$ . The plots are partially blinded, and only the statistical uncertainty is shown in the ratio plots.

- 1065 • Azimuthal angle between the muon and the missing transverse energy,  $\Delta\phi(\mu, E_T^{\text{miss}})$ ;

1066 The distributions of these variables for the signal and the reducible background are given in  
 1067 Fig. 50. A threshold requirement is made on the the BDT output given and the limit is estimated  
 1068 by a fit to the  $M_{\text{col}}$  distribution. The lepton  $p_T$  is not included to avoid biasing the shape of the  
 1069 background.

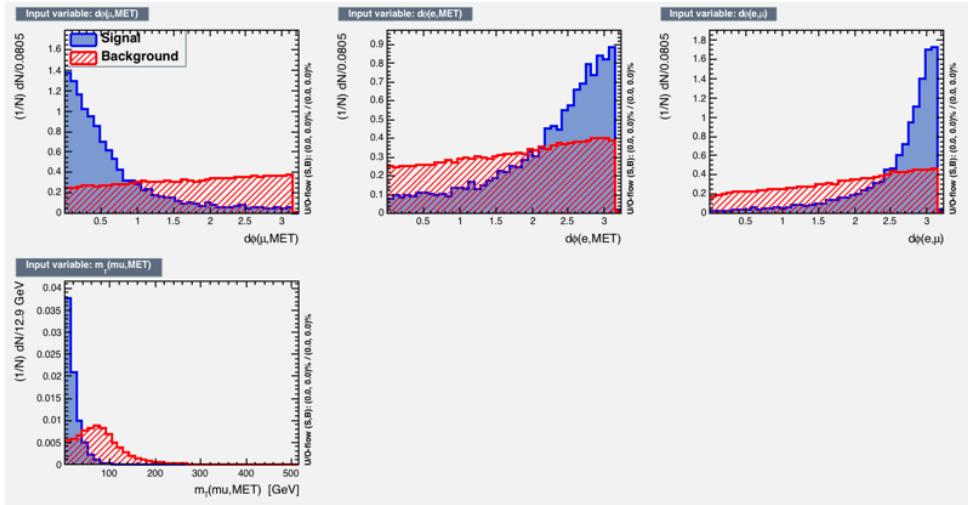


Figure 50:  $H \rightarrow e\tau_\mu$ : Normalized distributions of the variables that enter the BDT for the  $e\tau_\mu$  analysis. The signal (blue) is composed of a weighted mixture of ggH and VBF  $H \rightarrow e\tau$  events, whereas the background (red) is made of  $t\bar{t}$  events. The events used in the BDT pass the same  $p_T$ ,  $\eta$ , and isolation criteria as in the signal region.

1070 The expected limits are estimated by fitting the BDT output distributions in the four jet categories. The BDT output distributions for the signal and  $t\bar{t}$  background are shown in Fig. 51.  
 1071 The figures also show the results of Kolmogorov-Smirnov tests to assess the compatibility of  
 1072 the distributions on the training sample and on a test sample composed of different events.

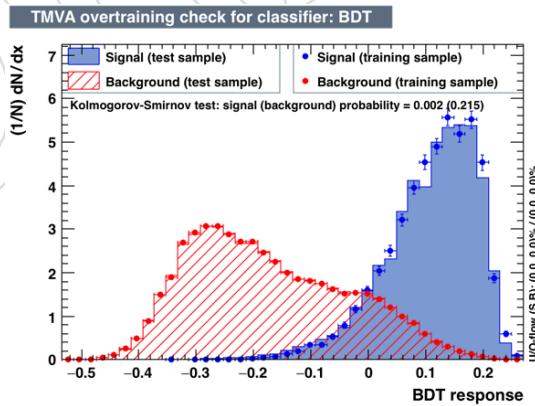


Figure 51:  $H \rightarrow e\tau_\mu$ : BDT output distributions.

1073  
 1074 The data vs simulation distributions of both BDT outputs are shown in Fig. 52. Good agreement  
 1075 is seen between observed data and predicted backgrounds in the unblinded regions with low  
 1076 signal significance.

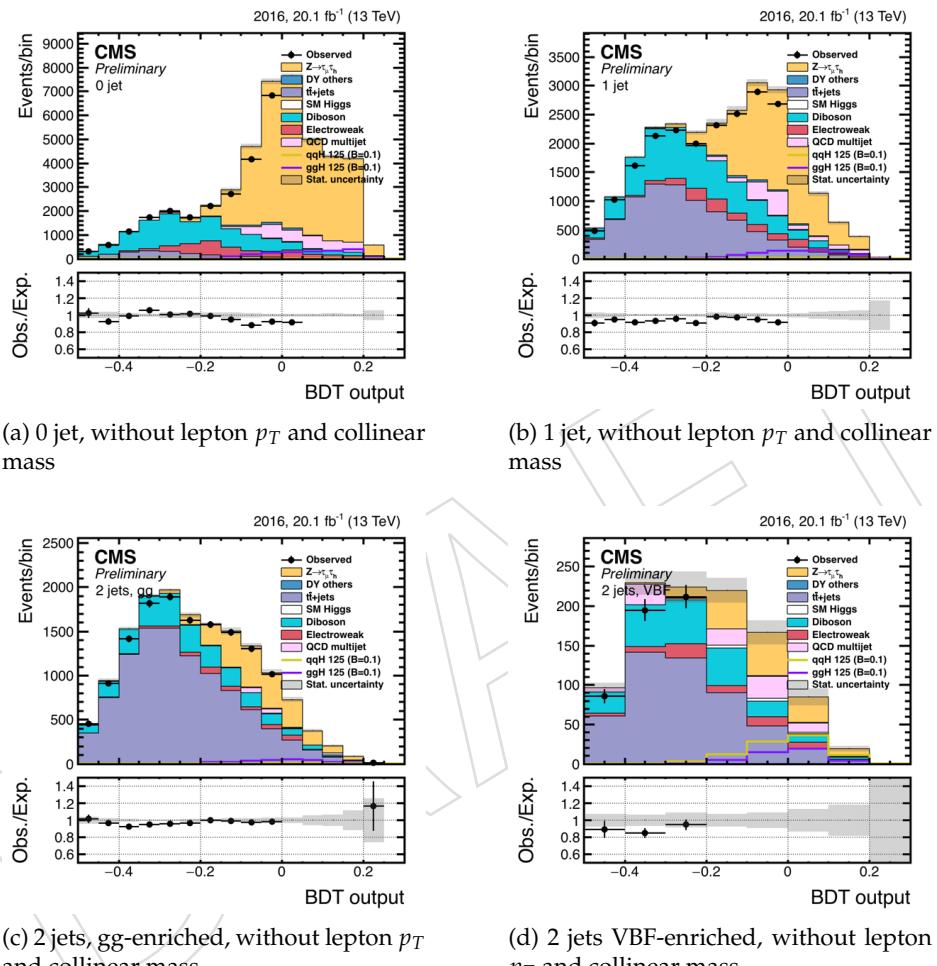


Figure 52:  $H \rightarrow e\tau_\mu$ : BDT output in the four categories for the training without the collinear mass and the lepton  $p_T$ , in the  $e\mu$  final state. The plots are partially blinded, and only the statistical uncertainty is shown in the ratio plots.

## 1077 C BDT output distributions in control regions

1078 Some control regions enriched in specific backgrounds are designed to verify the data/MC  
 1079 agreement of the BDT output distributions for various processes independently. These control re-  
 1080 gions are not orthogonal to the signal region, even if they are signal-poor, and cannot therefore  
 1081 be used as control regions in the fit to extract the analysis results. They are however useful as a  
 1082 check that the BDT output shapes in the signal region are reliable within uncertainties.

1083 In the  $e\tau_h$  and  $\mu\tau_h$  final states, a region enriched in  $Z \rightarrow \tau\tau$  is obtained by selecting events  
 1084 with an invariant mass between the muon (or electron) and the  $\tau_h$  candidate between 40 and  
 1085 80 GeV, with a transverse mass between the muon (or electron) and the MET less than 40 GeV  
 1086 to reduce the  $W+jets$  background, and with  $P_T^{\gamma}$  more than -25 GeV to reject the  $t\bar{t}$  and the QCD  
 1087 backgrounds. The BDT output distributions in this region are shown in Fig. 53 and 54, for the  
 1088  $e\tau_h$  and  $\mu\tau_h$  final states, respectively. Good data/MC agreement is seen.

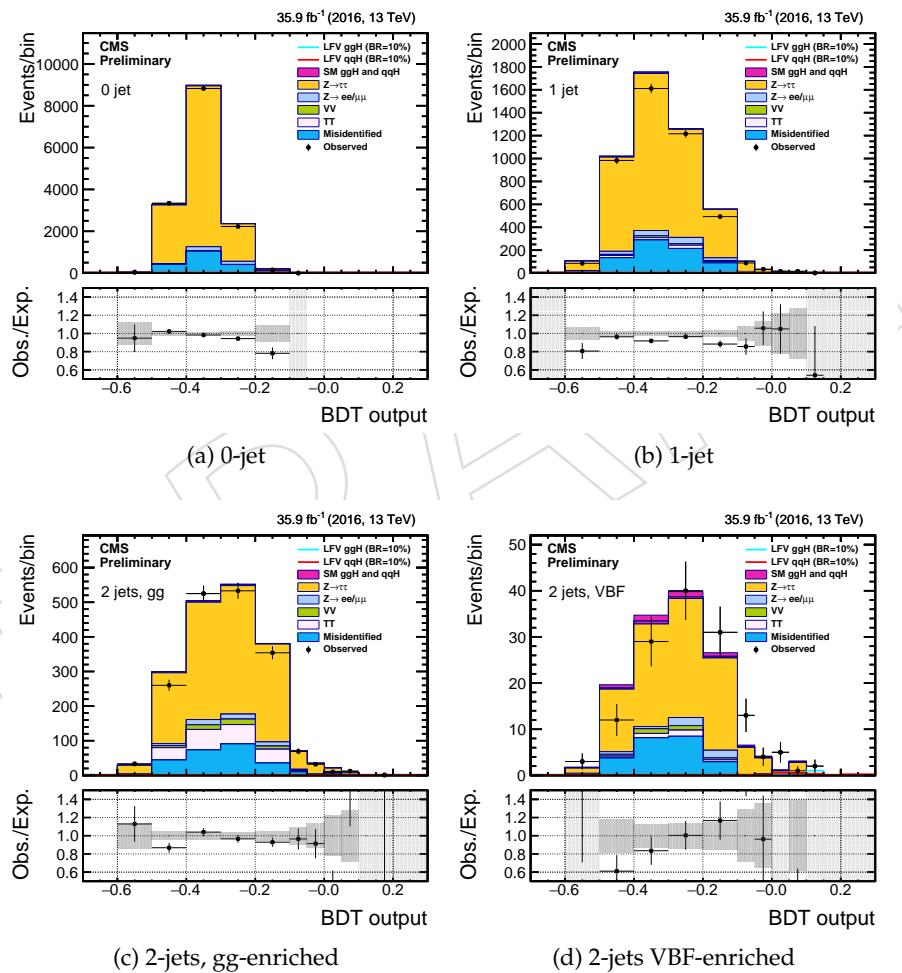


Figure 53: BDT output in  $Z \rightarrow \tau\tau$ -enriched regions in the four categories of the analysis for the  $e\tau_h$  final state. Only the statistical uncertainty is shown. The exact selection is detailed in the text.

1089 In the  $\mu\tau_e$  ( $e\tau_{\mu\mu}$ ) final state, the selection to obtain a  $Z \rightarrow \tau\tau$ -enriched control region is slightly  
 1090 different to account for different kinematics as the light lepton from tau decay is typically softer  
 1091 than a reconstructed hadronic tau  $\tau_h$ . In practice, events are selected with an invariant mass

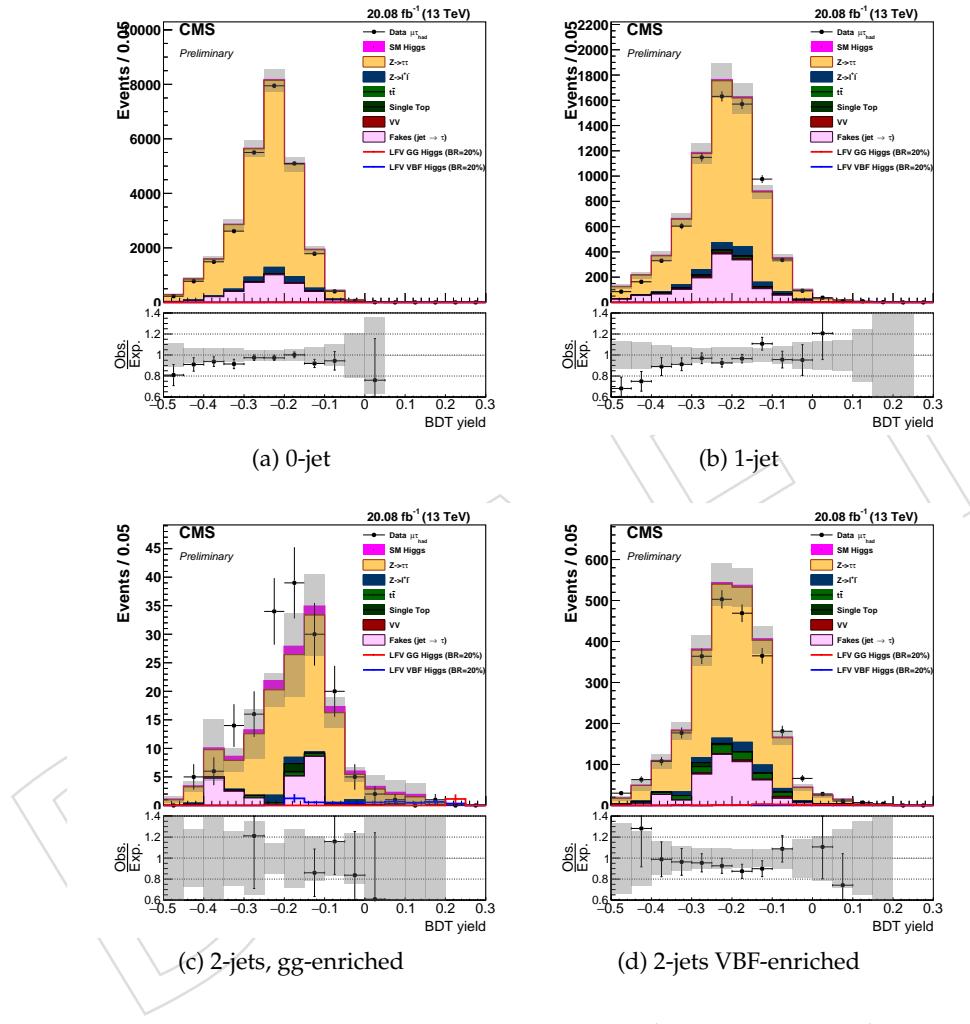


Figure 54: BDT output in  $Z \rightarrow \tau\tau$ -enriched regions in the four categories of the analysis for the  $\mu\tau_h$  final state. Only the statistical uncertainty is shown. The exact selection is detailed in the text.

1092 between the electron and the muon between 30 and 70 GeV, with a transverse mass between  
 1093 the muon and the MET less than 60 GeV, and with a muon transverse momentum less than 40  
 1094 GeV. The BDT output distributions in this region are shown in Fig. 55 for the  $\mu\tau_e$  final state.

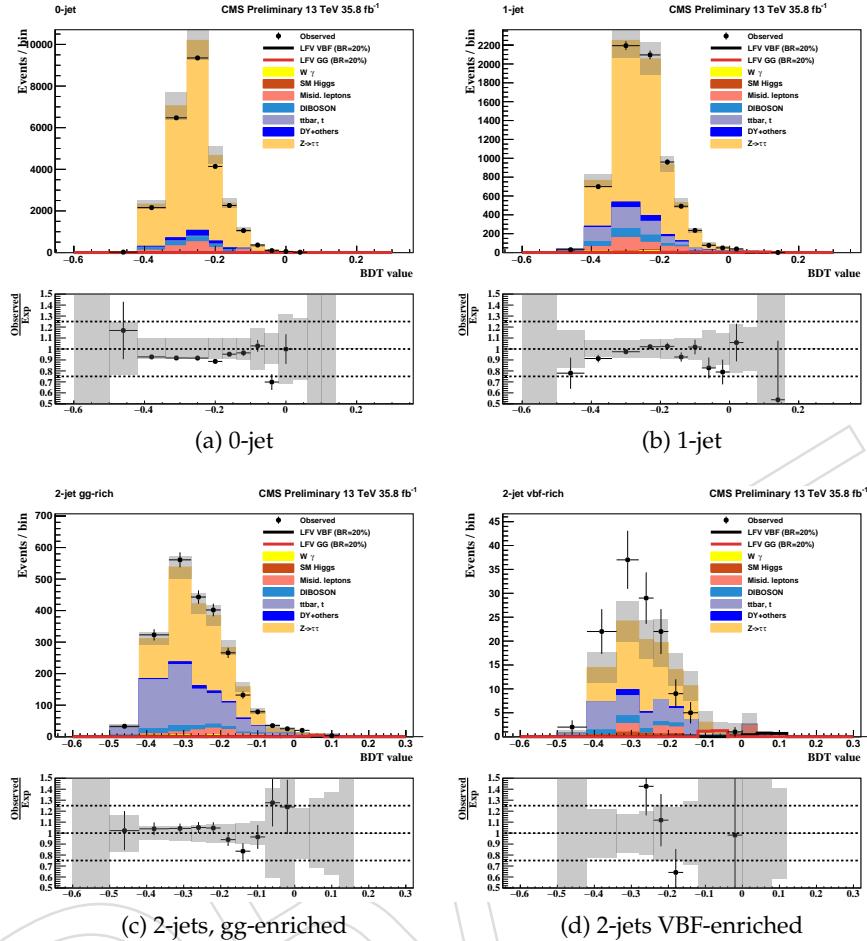


Figure 55: BDT output in  $Z \rightarrow \tau\tau$ -enriched regions in the four categories of the analysis for the  $\mu\tau_e$  final state. The exact selection is detailed in the text.

1095 In the  $e\tau_h$  ( $\mu\tau_h$ ) final state, a region enriched in  $Z \rightarrow ee$  ( $Z \rightarrow \mu\mu$ ) is obtained by selecting events  
 1096 with  $|m_{vis}(e/\mu, \tau_h) - m_Z| < 5$  GeV, MET less tan 25 GeV, and with the transverse mass between  
 1097 the electron (muon) and MET less than 40 GeV. The BDT output distributions in this region are  
 1098 shown in Fig 56 and 57 for the  $e\tau_h$  and  $\mu\tau_h$  final states, respectively.

1099 In the  $e\tau_h$  ( $\mu\tau_h$ ) final state, a region enriched in  $t\bar{t}$  events is obtained by selecting events with  
 1100  $m_{vis}(e/\mu, \tau_h) > 60$  GeV to reject the  $Z \rightarrow \tau\tau$  background, and with  $P_\zeta$  less than -50 GeV. No b-  
 1101 tagging requirement is made to stay close to the selection in the signal region. The BDT output  
 1102 distributions in this region are shown in Fig 58 for the  $e\tau_h$  final state.

1103 Two different  $t\bar{t}$  control regions were obtained for  $\mu\tau_e$  ( $e\tau_\mu$ ) final states.

1104 First, a region enriched in  $t\bar{t}$  events is obtained by selecting events with the transverse mass  
 1105 between the electron (muon) and MET greater than 50 GeV, and by removing the btag veto.  
 1106 The BDT output distributions for this region are shown in Fig. 59  
 1107 for the  $\mu\tau_e$  final state.

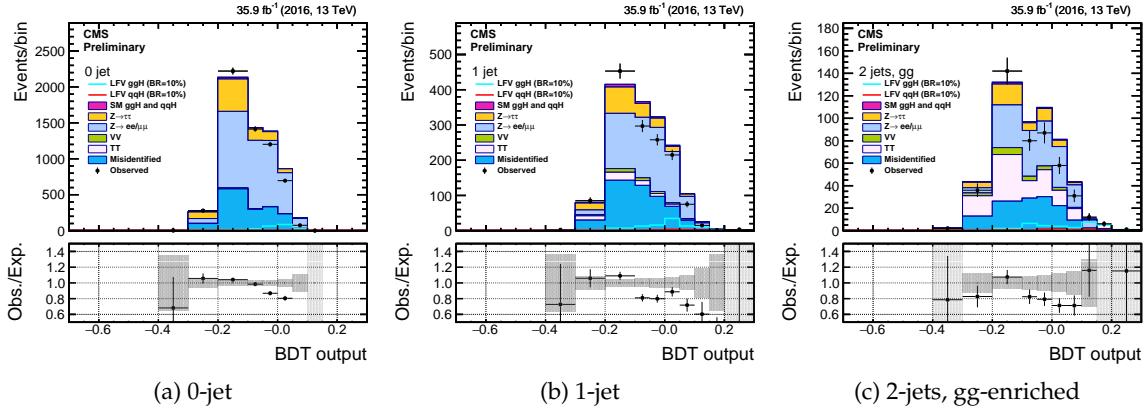


Figure 56: BDT output in  $Z \rightarrow ee$ -enriched regions in three of the four categories of the analysis for the  $e\tau_h$  final state. The 2 jets VBF category is too limited by statistics and is therefore not shown. Only the statistical uncertainty is shown. The exact selection is detailed in the text.

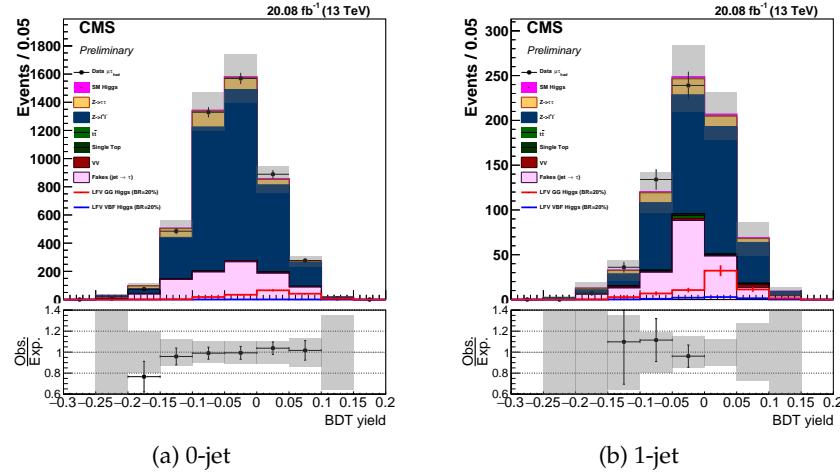


Figure 57: BDT output in  $Z \rightarrow \mu\mu$ -enriched regions in the four categories of the analysis for the  $\mu\tau_h$  final state. The exact selection is detailed in the text.

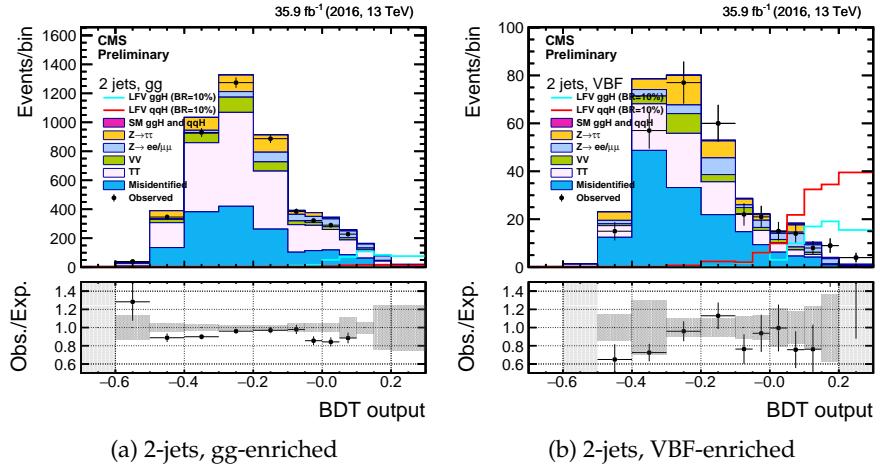


Figure 58: BDT output in  $t\bar{t}$ -enriched regions in categories with two jets for the  $e\tau_h$  final state. Only the statistical uncertainty is shown. The exact selection is detailed in the text.

- 1108 the second  $t\bar{t}$  control region has been obtained in the  $\mu\tau_\mu$  ( $e\tau_e$ ) final states by requiring that a
- 1109 b-tagged jet be present in the event.
- 1110 In the  $e\tau_h$  ( $\mu\tau_h$ ) final state, a region enriched in W+jets events is obtained by selecting events
- 1111 with a transverse mass between the electron (muon) and MET greater than 90 GeV. As in the
- 1112 signal region the leptons are required to have an opposite sign charge. The BDT output distri-
- 1113 butions in this region are shown in Fig. 61 and 62, for the  $e\tau_h$  and  $\mu\tau_h$  final states, respectively.
- 1114 In the  $\mu\tau_e$  final state, a region enriched in QCD events is obtained by selecting events with
- 1115 leptons passing inverted isolation conditions:  $0.15 < \text{iso}(\mu) < 0.25$  or  $0.1 < \text{iso}(e) < 1.0$ . The
- 1116 BDT output distributions in this region are shown in Fig. 63 for the  $\mu\tau_e$  final state.

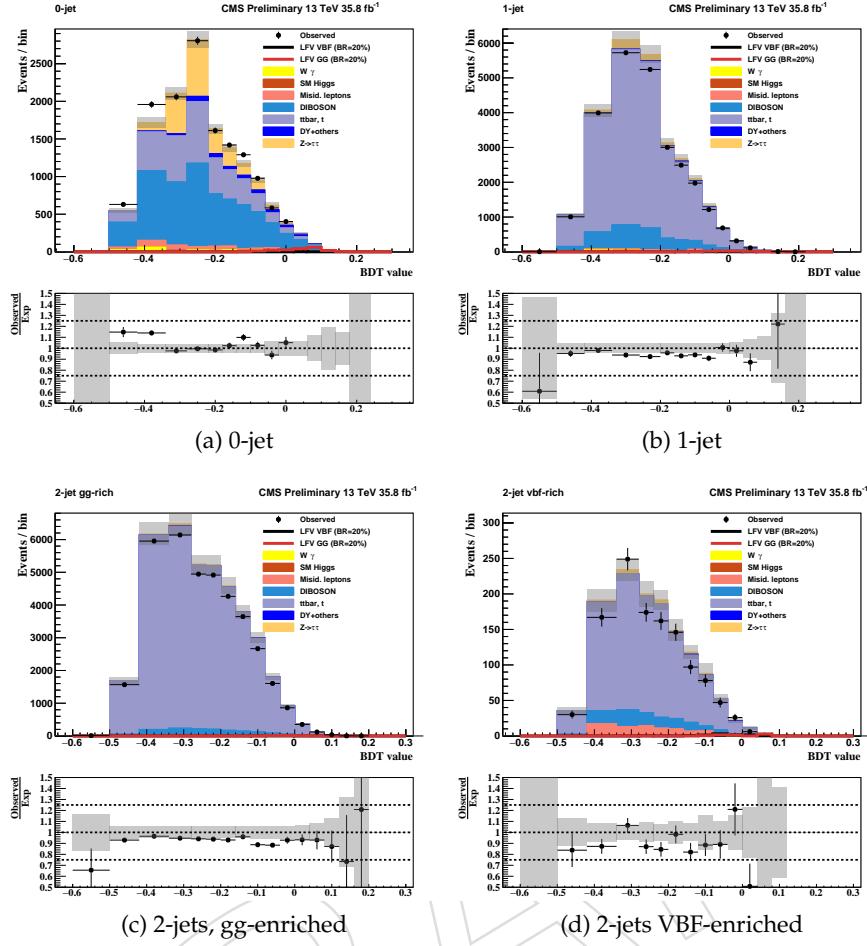


Figure 59: BDT output in  $t\bar{t}$ -enriched regions in the four categories of the analysis for the  $\mu\tau_e$  final state by removing the bTag veto and placing a greater than 50 GeV cut on transverse mass of electron and MET.

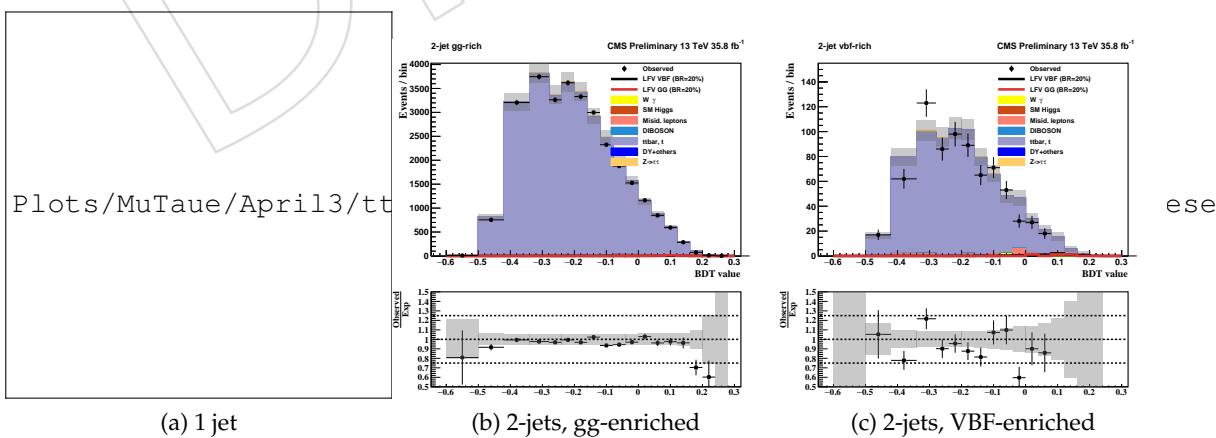


Figure 60: BDT output in  $t\bar{t}$ -enriched regions in categories with one jet in the  $\mu\tau_e$  final state. The exact selection is detailed in the text.

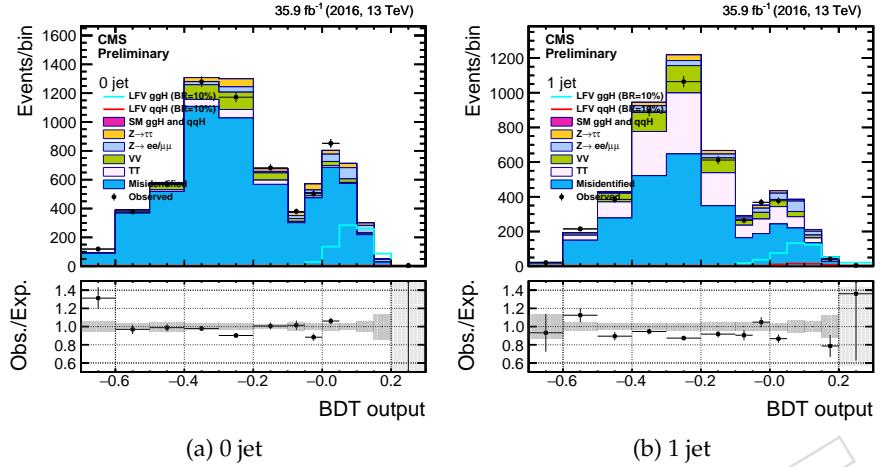


Figure 61: BDT output in W+jets-enriched regions in categories with less than two jets for the  $e\tau_h$  final state. Only the statistical uncertainty is shown. The exact selection is detailed in the text.

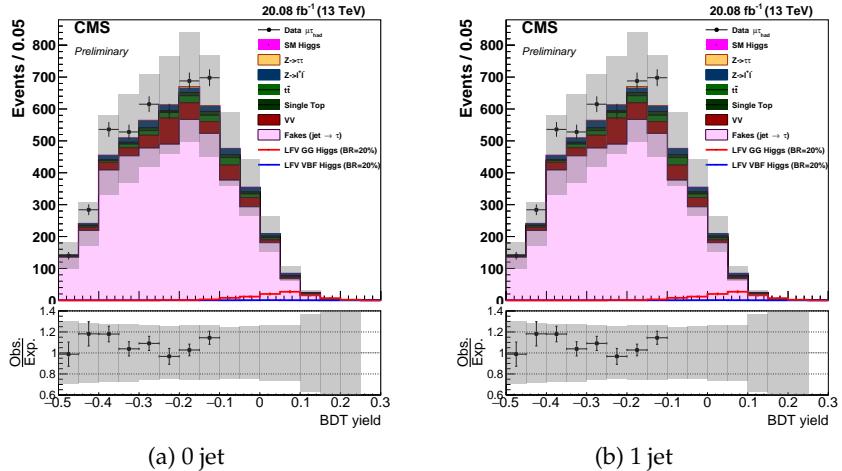


Figure 62: BDT output in W+jets-enriched regions in categories with less than two jets for the  $\mu\tau_h$  final state. Only the statistical uncertainty is shown. The exact selection is detailed in the text.

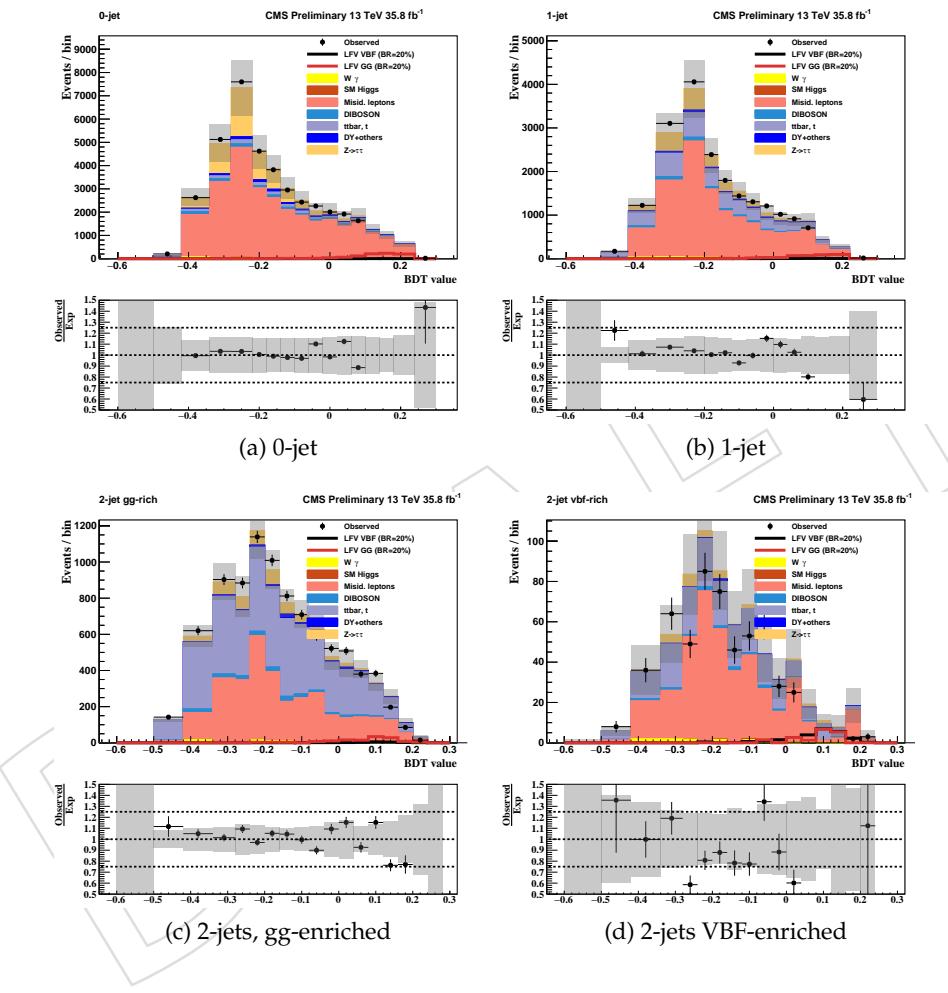
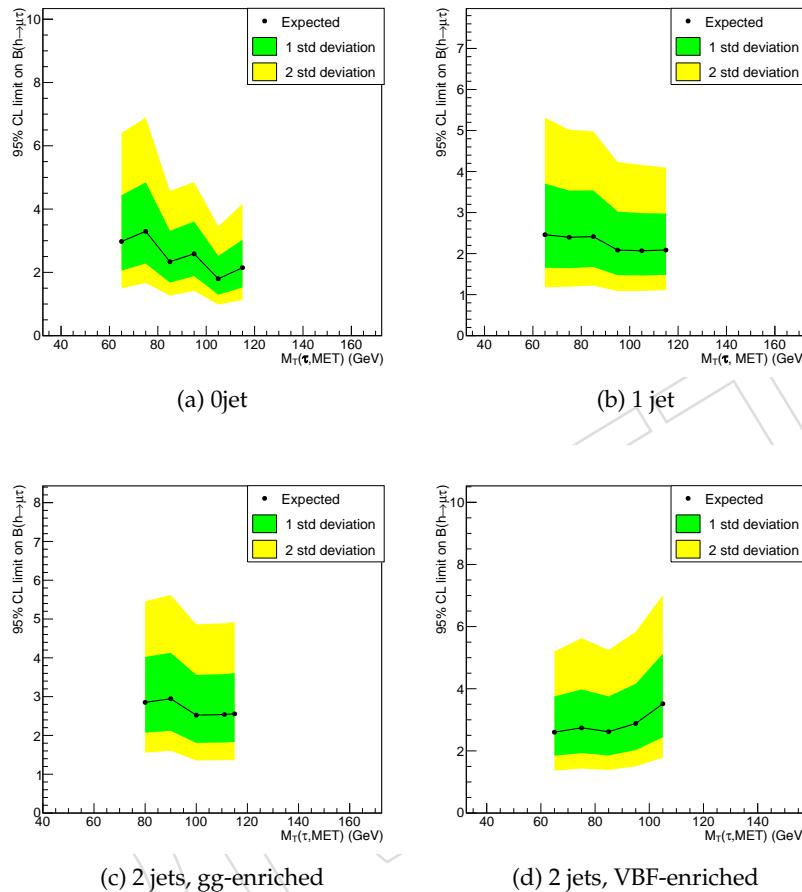


Figure 63: BDT output in QCD-enriched regions in the four categories of the analysis for the  $\mu\tau_e$  final state. The exact selection is detailed in the text.

## 1117 D Tuning of the cut based analysis for the $H \rightarrow \mu\tau_h$

1118 For the cut based analysis described in Sec. 5.2 the thresholds have been optimized to give the  
 1119 most stringent expected limits based on an Asimov dataset. Examples of limits obtained from  
 1120 different threshold are shown in Fig.64 for the  $M_T(\tau, MET)$  for the different categories. Fig. 65  
 shows the optimization for the di-jet invariant mass for the two 2-jet categories.



1121 Figure 64: Expected limits based on an Asimov dataset as a function of  $M_T(\tau, MET)$  for the different categories.

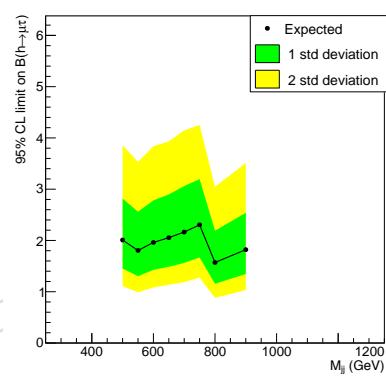


Figure 65: Expected limits based on an Asimov dataset as a function of  $M_{jj}$  for the 2 jet categories.