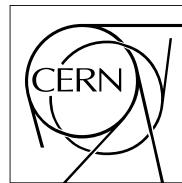


The Compact Muon Solenoid Experiment
Analysis Note

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



20 January 2017 (v5, 04 March 2017)

Search for ttH in multilepton final states with the full 2016 dataset

Daniel Bloch, Cristina Botta, Nicolas Chanon, Xavier Coubez, Kevin Lannon, Anne-Catherine Le Bihan, Binghuan Li, Jing Li, Charlie Mueller, Marco Peruzzi, Giovanni Petrucciani, Francesco Romeo, Sergio Sanchez Cruz, Kirill Skovpen, Geoffrey Smith, Benjamin Stieger, Thomas Strebler, Pietro Vischia, Huaqiao Zhang

Abstract

A search for the standard model Higgs boson produced in association with a top quark pair is presented, using the full pp collision dataset collected by the CMS experiment in 2016. We target final states where the Higgs boson decays to either a W, Z or tau pair by selecting events with two isolated same-sign leptons, or more than three leptons, and b-jets.

1 Contents

2	1	Introduction	2
3	2	Data and MC Samples	4
4	2.1	Data Samples	4
5	2.2	Signal Samples	4
6	2.3	Background Samples	4
7	2.4	Triggers	6
8	3	Event reconstruction and object identification	12
9	3.1	Jets and B-tagging	12
10	3.2	Missing Energy	12
11	3.3	Lepton Identification	13
12	3.4	Validation of lepton identification variables	16
13	3.5	Taus	22
14	4	Event selection	25
15	4.1	Two lepton same-sign category	25
16	4.2	Three and four lepton categories	25
17	5	Signal extraction	31
18	5.1	2lss event BDT	31
19	5.2	3l event BDT	37
20	5.3	2D BDT binning	40
21	5.4	Event subcategories	43
22	6	Signal modeling	45
23	6.1	Correction factors and experimental uncertainties	45
24	6.2	Theoretical uncertainties	45
25	7	Background predictions	47
26	7.1	Irreducible backgrounds	47
27	7.2	Di-bosons backgrounds	49
28	7.3	Charge misassignment background	52
29	7.4	Fake lepton background	55
30	8	Results	65
31	A	Control region plots	68
32	A.1	Lepton MVA sideband region	68
33	A.2	Jet multiplicity sideband region	71
34	A.3	$WZ \rightarrow 3\ell$	73
35	A.4	$t\bar{t}Z \rightarrow 3\ell$	74
36	B	Matrix Element Method	76
37	B.1	The algorithm	76
38	B.2	Treatment of jets and permutations	77
39	B.3	MEM discriminant	80
40	B.4	MEM distributions after event selection	83
41	C	The H_{jj} tagger	91
42	D	Signal Extraction using the k -means algorithm	93
43	E	Training of the kinematic BDT discriminators	95

44 1 Introduction

45 The LHC Run I data have been exploited to measure all the accessible properties of the newly-
 46 discovered Higgs boson [1, 2]. ATLAS and CMS have combined their effort in order to reach
 47 an already very precise measurement of the boson mass, 125.09 ± 0.21 (stat.) ± 0.11 (syst.)
 48 GeV [3]. This precise mass result has created an opportunity to test the predictions of the stan-
 49 dard model by measuring the other properties of the Higgs boson. Measurements of the Higgs
 50 boson production and decay rates and constraints on its couplings have been performed by
 51 both experiments [4, 5], and, in general, agreement with the SM predictions given the current
 52 uncertainties (10-30 %) have been found. It is of great interest to use the 13 TeV LHC data to
 53 further constrain these measurements, as any deviation from expectation could be a sign of
 54 new physics.

55

56 Among these measurements, it is of particular interest to measure the coupling of the Higgs
 57 boson to the top quark ($t\bar{t}H$) because the top quark could play a special role in the context of
 58 electroweak symmetry breaking due to its large mass. The Higgs boson does not decay to top
 59 quarks. The $t\bar{t}H$ interaction vertex, however, is present in a rare production mechanism where
 60 the Higgs boson is produced in association with a top quark-antiquark pair as shown in Fig. 1.
 61 At LHC energies the largest contribution to the standard model Higgs boson production is
 62 a gluon-gluon induced loop dominated by virtual top exchange. The comparison of a direct
 63 measurement of the $t\bar{t}H$ coupling with the one inferred from the cross section measurement
 64 can put limits on the contribution of new physics to the gluon-gluon loop.

65

66 The $t\bar{t}H$ process has been used by both ATLAS and CMS experiments to directly measure the
 67 top-Higgs coupling at tree level with the 20 fb^{-1} of 8 TeV collisions of the LHC Run I. Via
 68 this process, both experiments reached a 30% accuracy on the top Yukawa coupling, a great
 69 achievement given that the production cross section (130 fb at 8 TeV at next-to-leading order
 70 (NLO) [6]) was two orders of magnitude lower with respect to the dominant Higgs produc-
 71 tion mode (gluon-gluon fusion). In order to achieve this result several decay channels of the
 72 Higgs boson have been considered by both experiments, and three main searches have been
 73 designed. The first channel searches for $t\bar{t}H$ in events where the Higgs boson decays to $b\bar{b}$; the
 74 best fit value for the combined signal strength obtained by the CMS experiment is $0.7^{+1.9}_{-1.9}$ (95%
 75 CL) [7]. The second channel searches for $t\bar{t}H$ in events where the Higgs boson decays to $\gamma\gamma$;
 76 the best fit value for the combined signal strength obtained by the CMS experiment is $2.7^{+2.6}_{-1.8}$
 77 (95% CL)) [7].

78

79 We designed the third search to probe $t\bar{t}H$ events where the Higgs boson decays into ZZ^* ,
 80 WW^* , or $\tau\tau$, with at least one Z , W or τ decaying leptonically. Despite the small branching
 81 ratio, the presence of one or two additional leptons from the top quark pair decays leads to the
 82 following clean experimental signatures:

- 83 • two same-sign leptons (electrons or muons) plus b-tagged jets;
- 84 • three leptons plus b-tagged jets;
- 85 • four leptons plus b-tagged jets.

86 Examples of Feynman diagrams for $t\bar{t}H$, followed by the decays of the top quark and the Higgs
 87 boson that lead to the signatures described above are shown in Fig. 1. With this search we ob-
 88 tained the most precise measurement of the $t\bar{t}H$ signal strength: $3.7^{+1.9}_{-1.9}$ (95%CL)) [8].

89

90 The combined best-fit signal strength obtained assuming a Higgs boson mass of 125 GeV was
 91 $\mu = 2.9^{+1.1}_{-0.9}$. This result corresponds to a 3.5 standard deviation excess over the background-
 92 only ($\mu = 0$) hypothesis, and represents a 2.1 standard deviation upward fluctuation on the SM
 93 $t\bar{t}H$ ($\mu = 1$) expectation. Although the combined observed signal strength is consistent with
 94 SM expectations, with a roughly 2 standard deviation upward fluctuation, it is interesting to
 95 point out that the excess was mainly driven by the multilepton analysis, and in particular by
 96 the same-sign di-muon subsample [8].

97

98 With respect to 8 TeV, the 13 TeV $t\bar{t}H$ cross section increased by a factor of 4 with the higher center
 99 of mass energy, while the cross sections of the main backgrounds $t\bar{t}W$, $t\bar{t}Z$, $t\bar{t}+jets$ increased
 100 by roughly a factor of 3. We thus expect to increase our sensitivity during Run II, compared to
 101 Run I.

102 The first multilepton search at 13 TeV analyzed 2.3 fb^{-1} of the 2015 dataset. It measured the
 103 expected 95% confidence level upper limit on the Higgs boson production cross section for a
 104 Higgs boson mass of $125 \text{ GeV}/c^2$ to be 2.6 times the standard model expectation, compared
 105 to the observed limit of 3.3. The signal strength μ , relative to the expectation for the standard
 106 model Higgs boson, was measured to be $0.6^{+1.4}_{-1.1}$ [9].

107

108 The 2016 data has been preliminarily analysed for the ICHEP conference considering 12.9 fb^{-1}
 109 [10]. The results have been combined with the 2015 dataset and yield a $t\bar{t}H$ signal strength
 110 of $2.0^{+0.8}_{-0.7}$ times the standard model prediction. They are used to set a 95% confidence level
 111 upper limit on the signal production cross section of 3.4 times the standard model expectation,
 112 compared to an expected upper limit of $1.3^{+0.6}_{-0.4}$ in the absence of a signal.

113 In this note we perform the $t\bar{t}H$ multilepton search with the full 2016 data, corresponding to
 114 36.9 fb^{-1} , collected by the CMS experiment at $\sqrt{s} = 13 \text{ TeV}$. The general strategy remains simi-
 115 lar to the previous searches. Multivariate analysis techniques are used to identify objects with
 116 high purity and to distinguish background from signal events. The amount of signal is fit to
 117 the multivariate discriminant output distribution in all the final states simultaneously.

118

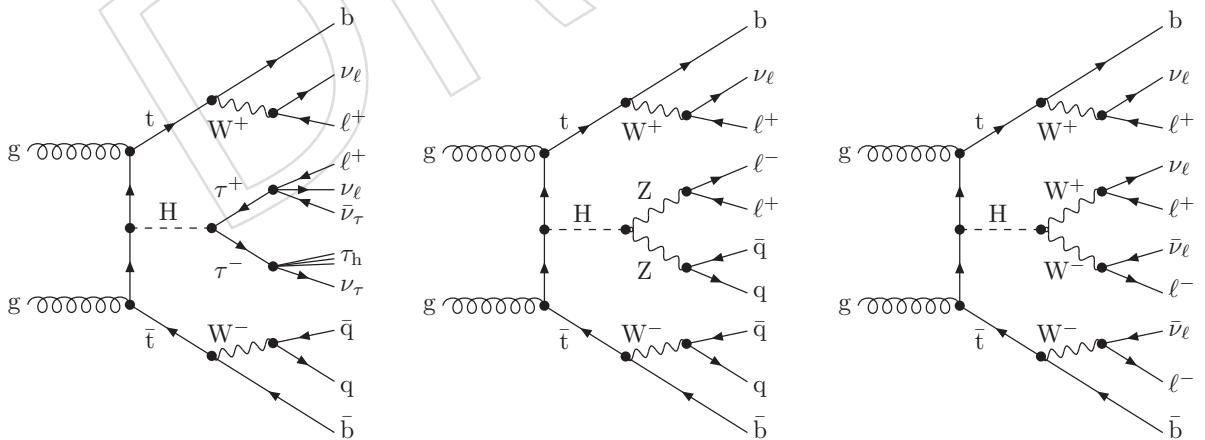


Figure 1: Examples of leading order Feynman diagrams for $t\bar{t}H$ production at pp colliders, with the Higgs boson decaying to $\tau\tau$, ZZ^* , and WW^* (from left to right). The first, second, and third diagrams are examples of the two same-sign lepton signature, the three lepton signature, and the four lepton signature, respectively.

119 2 Data and MC Samples

120 In the following we list as reference the set of samples of the RunIISummer16MiniAODv2
 121 (CMSSW 80X) campaign used for the studies and results presented in this analysis.

122 2.1 Data Samples

123 The data used in this analysis has been collected with the CMS detector in 2016. We use data
 124 collected only during periods when the CMS magnet was on, corresponding to a total integrated
 125 luminosity of 36.9 fb^{-1} [11]. We use the 23Sep2016 (Run B to G) and PromptReco (Run
 126 H) versions of the datasets.

127 2.2 Signal Samples

128 We use simulated Monte-Carlo events to model both the ttH signal and the various standard
 129 model backgrounds. Details on the used sample are in Table 1.

Sample	Xsec (pb)
/ttHToNonbb_M125_TuneCUETP8M2_ttHtranche3_13TeV-powheg-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM	0.5085*(1-0.577)

Table 1: Signal samples used in this analysis (CMSSW 80X).

130 2.3 Background Samples

131 In this section, we list the Monte-Carlo samples used to model our backgrounds. A variety of
 132 methods are used to produce these samples: for our dominant backgrounds (ttbar, ttW, ttZ), we
 133 use the same aMC@NLO→PYTHIA chain used for the ttH signal and the NLO cross sections.
 134 Other minor backgrounds may use other generators, such as PYTHIA or POWHEG matched
 135 to PYTHIA. See table 2 for more details.

136

137 The nominal $t\bar{t}Z$ MC sample is generated with the requirement $m_{\ell^+\ell^-} > 10 \text{ GeV}$, because of
 138 this, we also use LO Madgraph to generate the remaining part of the phase space ($1 < m_{\ell^+\ell^-} <$
 139 10 GeV). This private sample was produced both in the context of the RunIISpring16MiniAODv2
 140 and RunIISpring16MiniAODv1 (CMSSW 80X) campaigns. Another sample not yet available in
 141 both campaigns is the $t\bar{t}WW$. Nevertheless in the RunI analysis we proved that in all the ana-
 142 lyzed final states, the contribution of the process is found to be at least an order of magnitude
 143 smaller than $t\bar{t}W$ and $t\bar{t}Z$.

Sample	Xsec (pb)
/TTWJetsToLNu_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8/RunIISummer16MiniAOdV2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6_ext1-v3/MINIAODSIM	0.2043
/TTWJetsToLNu_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8/RunIISummer16MiniAOdV2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6_ext2-v1/MINIAODSIM	0.2043
/TTZToLLNuNu_M-10_TuneCUETP8M1_13TeV-amcatnlo-pythia8/RunIISummer16MiniAOdV2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6_ext1-v1/MINIAODSIM	0.2529
/TTZToLL_M-1to10_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAOdV2-80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM	0.0493
/WGTolNuG_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8/RunIISummer16MiniAOdV2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6_ext1-v1/MINIAODSIM	585.8
/ZGto2LG_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8/RunIISummer16MiniAOdV2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6_ext1-v1/MINIAODSIM	131.3
/TGejets_TuneCUETP8M1_13TeV-amcatnlo_madspin_pythia8/RunIISummer16MiniAOdV2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM	2.967
/TGejets_TuneCUETP8M1_13TeV-amcatnlo_madspin_pythia8/RunIISummer16MiniAOdV2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6_ext1-v1/MINIAODSIM	2.967
/TTGjets_TuneCUETP8M1_13TeV-amcatnloFX-madspin-pythia8/RunIISummer16MiniAOdV2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM	3.697
Rares /WpWpjj_EWK-QCD_TuneCUETP8M1_13TeV-madgraph-pythia8/RunIISummer16MiniAOdV2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6_ext1-v1/MINIAODSIM	0.03711
Rares /WWTo2L2Nu_DoubleScattering_13TeV-amcatnlo-pythia8/RunIISummer16MiniAOdV2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM	0.1729
Rares /WWTo2L2Nu_TuneCUETP8M1_13TeV-amcatnlo-pythia8/RunIISummer16MiniAOdV2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM	0.2086
Rares /WZZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8/RunIISummer16MiniAOdV2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM	0.1651
Rares /ZZZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8/RunIISummer16MiniAOdV2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM	0.05565
Rares /Zq_ll_4f_13TeV-amcatnlo-pythia8/RunIISummer16MiniAOdV2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6_ext1-v1/MINIAODSIM	0.01398
Rares /ZTTT_TuneCUETP8M1_13TeV-amcatnlo-pythia8/RunIISummer16MiniAOdV2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM	0.0758
	0.099103
/T1Jets_SingleLeptFromTBar_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAOdV2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM	182.18
/T1Jets_SingleLeptFromTBar_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAOdV2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6_ext1-v1/MINIAODSIM	182.18
/T1Jets_SingleLeptFromT_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAOdV2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM	182.18
/T1Jets_SingleLeptFromT_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAOdV2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6_ext1-v1/MINIAODSIM	182.18
/T1Jets_DiLept_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAOdV2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM	87.3
/T1Jets_DiLept_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAOdV2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6_ext1-v1/MINIAODSIM	87.3
/ST_tW_top_5f_inclusiveDecays_13TeV_powheg_pythia8_TuneCUETP8M1/RunIISummer16MiniAOdV2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6_ext1-v1/MINIAODSIM	35.6
/ST_tW_antitop_5f_inclusiveDecays_13TeV_powheg_pythia8_TuneCUETP8M1/RunIISummer16MiniAOdV2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6_ext1-v1/MINIAODSIM	35.6
/ST_1-channel_top_4f_inclusiveDecays_13TeV_powhegV2-madspin-pythia8_TuneCUETP8M1/RunIISummer16MiniAOdV2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM	136.02
/ST_1-channel_antitop_4f_inclusiveDecays_13TeV_powhegV2-madspin-pythia8_TuneCUETP8M1/RunIISummer16MiniAOdV2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM	80.95
/ST_1-channel_4f_leptonDecays_13TeV-amcatnlo_pythia8_TuneCUETP8M1/RunIISummer16MiniAOdV2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM	3.68
/DYjetsToLL_M-10to50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAOdV2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM	18610
/DYjetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAOdV2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6_ext1-v2/MINIAODSIM	6025.2
/WZTo3LNu_TuneCUETP8M1_13TeV-powheg_pythia8/RunIISummer16MiniAOdV2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM	61526.7
/WWTo2LNu_13TeV_powheg/RunIISummer16MiniAOdV2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM	4.42965
/ZZTo4L_13TeV_powheg_pythia8/RunIISummer16MiniAOdV2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM	10.481
	1.256

Table 2: List of background samples used in this analysis (CMSSW 80X). In the first section of the table are listed the samples of the processes for which we use the simulation to extract the final yields and shapes, in the second section the samples of the processes we will estimate from data. The MC simulation is used to design the data driven methods and derive the associated systematics.

144 2.4 Triggers

145 In this analysis, we consider online-reconstructed events triggered on one, two or three leptons.
 146 The inclusion of single-lepton triggers boosts acceptance by including events where the p_T of
 147 the subleading lepton falls below the threshold of the double-lepton triggers. In addition, by
 148 including double-lepton triggers in the ≥ 3 lepton category, as well as single-lepton triggers in
 149 all categories, we increase efficiency by considering the logical “or” of the trigger decisions of
 150 all the individual triggers in a given category. Table 3 shows the lowest-threshold unprescaled
 151 triggers present in the High-Level Trigger (HLT) menus for both Monte-Carlo and data in 2016
 152 (will be updated if subsequent data have different unprescaled triggers).

153

154 Figures 2, 3, 4, and 5 show a comparison of the trigger efficiency between data and Monte-
 155 Carlo in each of the analysis categories. In general, we find that the trigger efficiencies in the
 156 data agree well with simulation. Measuring the efficiency in simulated events is straightfor-
 157 ward because there is no trigger bias with simulated events. To measure the efficiency in data
 158 we follow the procedure described here [12], which was also the same procedure used in the
 159 Run I, 2015 Run II, and 2016 Run II (ICHEP) multilepton analyses. We first select a set of events
 160 that were recorded on a trigger that is uncorrelated with the lepton triggers. We use events
 161 recorded on a MET trigger as a unbiased sample. We then look for candidate events with ex-
 162 actly two good leptons (and separately, events with exactly three good leptons). We measure
 163 the efficiency for the candidate events to pass the logical “or” of triggers being considered in a
 164 given event category (i.e., the triggers listed by category in table 3).

165

166 We use scale factors to correct for small differences in the trigger efficiency between data and
 167 Monte-Carlo. We compare the efficiency measured in data to the efficiency in simulation, and
 168 derive correction factors which are the ratio of these two efficiencies. Overall, we see good
 169 agreement between the efficiency in data and the efficiency in simulation, so we apply a flat
 170 scale factor close to unity in each category. These scale factors are summarized in table 4.



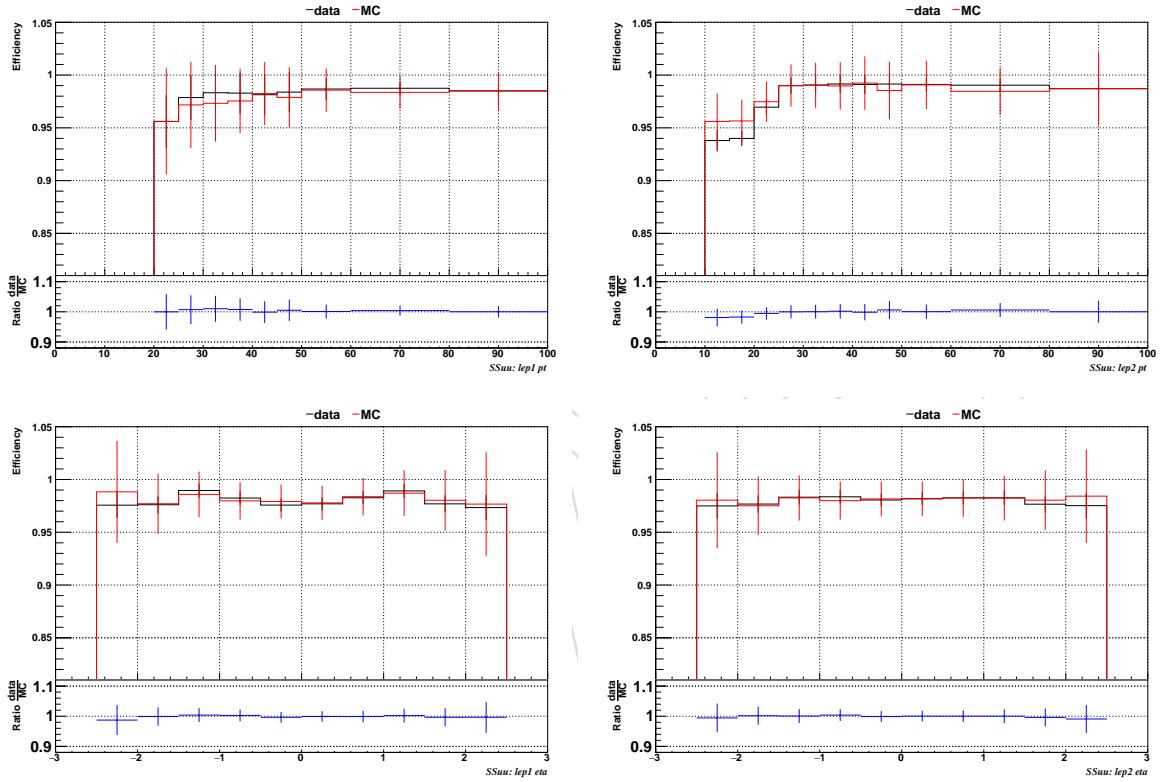


Figure 2: Comparison of the trigger efficiency in the same-sign dimuon category before corrections, shown as a function of the p_T and η of the leading lepton (left) and the sub-leading lepton (right).

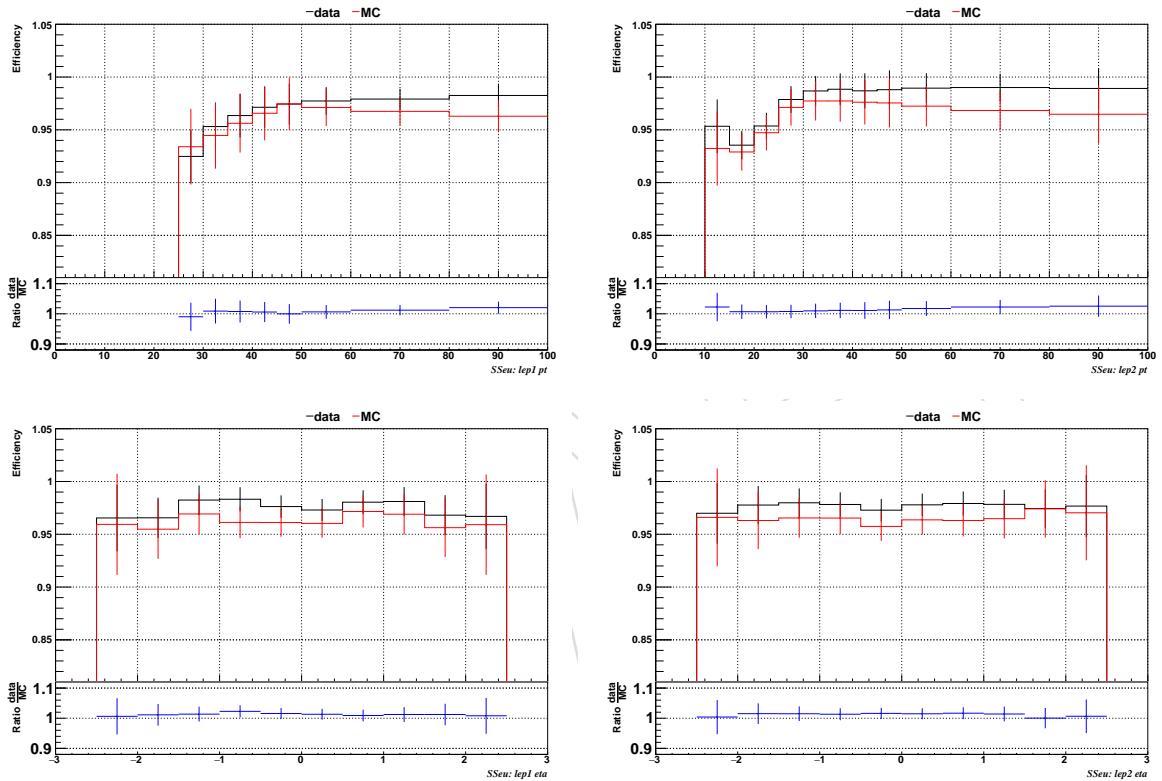


Figure 3: Comparison of the trigger efficiency in the same-sign muon+electron category before corrections, shown as a function of the p_T and η of the leading lepton (left) and the sub-leading lepton (right).

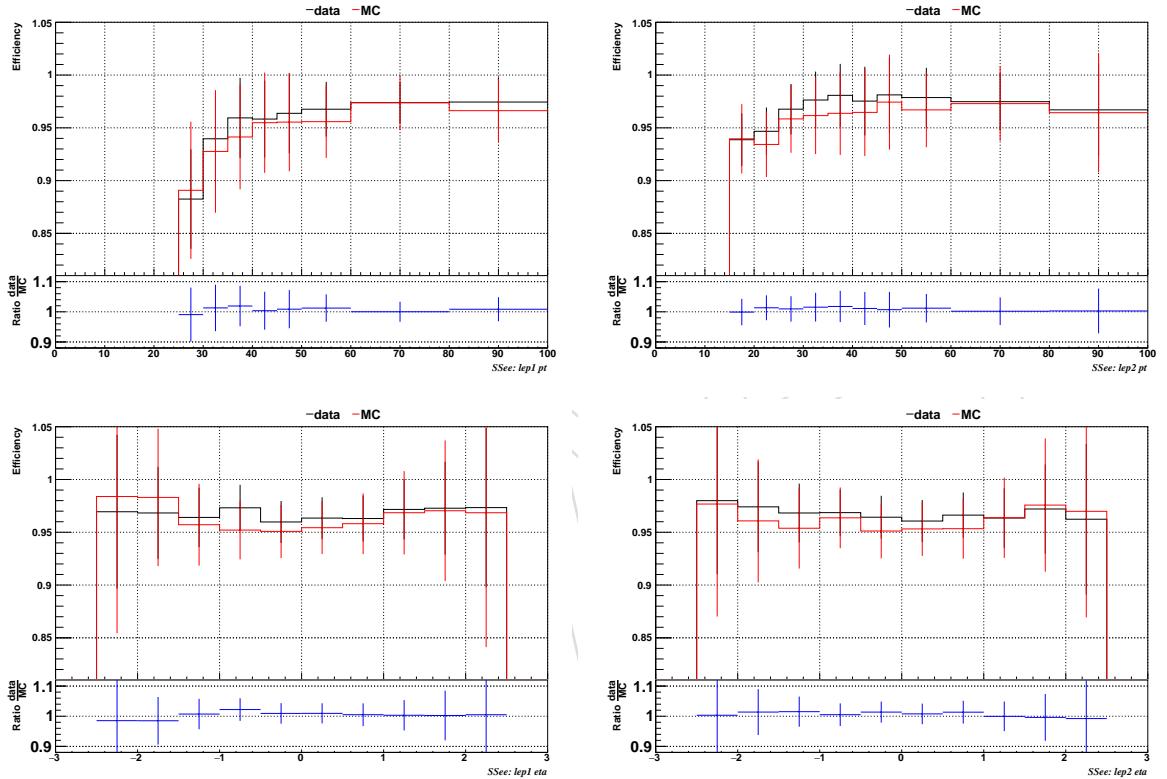


Figure 4: Comparison of the trigger efficiency in the same-sign dielectron category before corrections, shown as a function of the p_T and η of the leading lepton (left) and the sub-leading lepton (right).

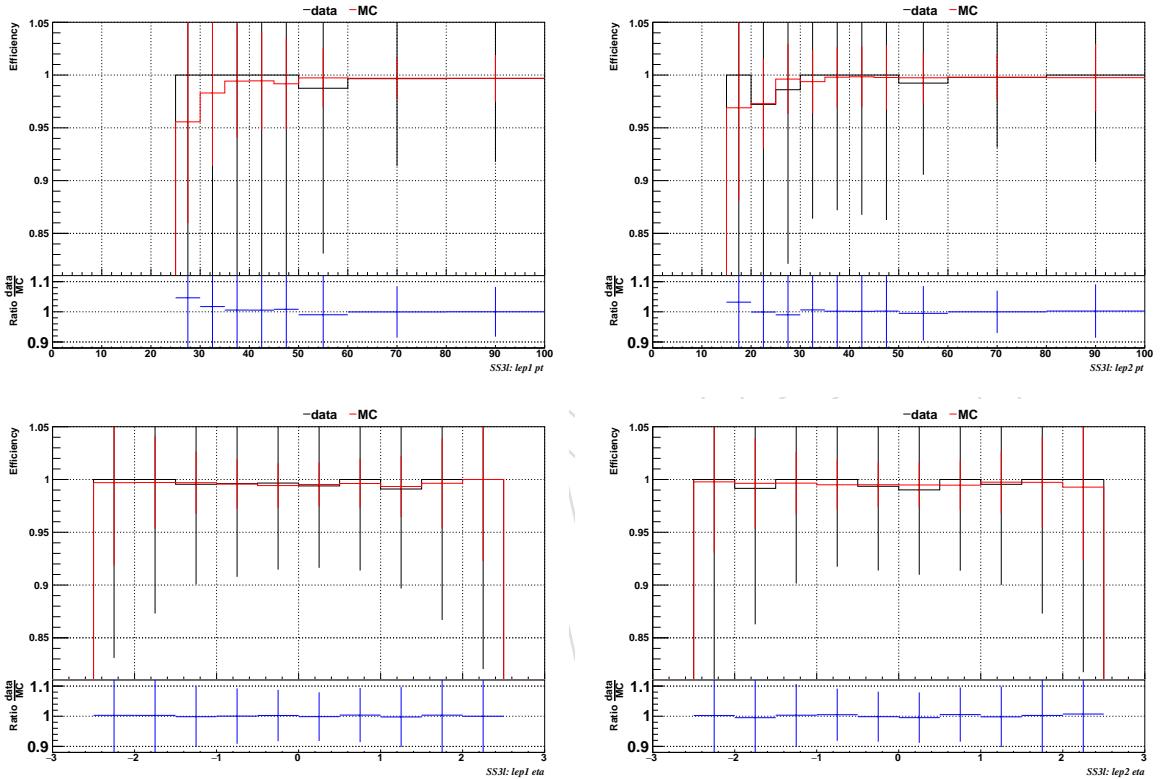


Figure 5: Comparison of the trigger efficiency in the ≥ 3 -lepton category before corrections, shown as a function of the p_T and η of the leading lepton (left) and the sub-leading lepton (right).

Same-sign dilepton (==2 muons)
HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_v*
HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ_v*
HLT_IsoMu22_v*
HLT_IsoTkMu22_v*
HLT_IsoMu22_eta2p1_v*
HLT_IsoTkMu22_eta2p1_v*
HLT_IsoMu24_v*
HLT_IsoTkMu24_v*
Same-sign dilepton (==2 electrons)
HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v*
HLT_Ele27_WPtight_Gsf_v*
HLT_Ele25_eta2p1_WPtight_Gsf_v*
HLT_Ele27_eta2p1_WPloose_Gsf_v*
Same-sign dilepton (==1 muon, ==1 electron)
HLT_Mu23_TrkIsoVVL_Ele8_CaloIdL_TrackIdL_IsoVL_v*
HLT_Mu23_TrkIsoVVL_Ele8_CaloIdL_TrackIdL_IsoVL_DZ_v*
HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_v*
HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v*
HLT_IsoMu22_v*
HLT_IsoTkMu22_v*
HLT_IsoMu22_eta2p1_v*
HLT_IsoTkMu22_eta2p1_v*
HLT_IsoMu24_v*
HLT_IsoTkMu24_v*
HLT_Ele27_WPtight_Gsf_v*
HLT_Ele25_eta2p1_WPtight_Gsf_v*
HLT_Ele27_eta2p1_WPloose_Gsf_v*
Three lepton and Four lepton
HLT_DiMu9_Ele9_CaloIdL_TrackIdL_v*
HLT_Mu8_DiEle12_CaloIdL_TrackIdL_v*
HLT_TripleMu_12_10_5_v*
HLT_Ele16_Ele12_Ele8_CaloIdL_TrackIdL_v*
HLT_Mu23_TrkIsoVVL_Ele8_CaloIdL_TrackIdL_IsoVL_v*
HLT_Mu23_TrkIsoVVL_Ele8_CaloIdL_TrackIdL_IsoVL_DZ_v*
HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_v*
HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v*
HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v*
HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_v*
HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ_v*
HLT_IsoMu22_v*
HLT_IsoTkMu22_v*
HLT_IsoMu22_eta2p1_v*
HLT_IsoTkMu22_eta2p1_v*
HLT_IsoMu24_v*
HLT_IsoTkMu24_v*
HLT_Ele27_WPtight_Gsf_v*
HLT_Ele25_eta2p1_WPtight_Gsf_v*
HLT_Ele27_eta2p1_WPloose_Gsf_v*

Table 3: Table of high-level triggers that we consider in the analysis.

Category	Scale Factor
2e	1.01 ± 0.02
e+mu	1.01 ± 0.01
2mu	1.00 ± 0.01
3 and 4l	1.00 ± 0.03

Table 4: Trigger efficiency scale factors and associated uncertainties, shown here rounded to the nearest percent.

171 3 Event reconstruction and object identification

172 A complete reconstruction of the individual particles from each collision event is obtained
 173 via the particle-flow (PF) algorithm. The technique uses the information from all CMS sub-
 174 detectors to identify and reconstruct individual particles in the collision event [13, 14]. The
 175 particles are classified into mutually exclusive categories: charged hadrons, neutral hadrons,
 176 photons, muons, and electrons.

177 3.1 Jets and B-tagging

178 Jets are reconstructed by clustering PF candidates using the anti- k_T algorithm with distance
 179 parameter $\Delta R = 0.4$ as implemented in the FASTJET package [15, 16]. The charged hadrons not
 180 coming from the primary vertices are subtracted from the PF candidates considered in the clus-
 181 tering. The primary vertex is chosen as the vertex with the highest sum of p_T^2 of its constituent
 182 tracks. The prescribed jet energy corrections are applied as a function of the jet E_T and η [17]. In
 183 addition, a multivariate discriminator is applied to distinguish between jets coming from the
 184 primary vertex and jets coming from pile-up vertices. The discrimination is based on the dif-
 185 ferences in the jet shapes, in the relative multiplicity of charged and neutral components, and
 186 in the different fraction of transverse momentum which is carried by the hardest components.
 187 Within the tracker acceptance the jet tracks are also required to be compatible with the primary
 188 vertex. Jets are only considered if they have a transverse energy above 25 GeV and $|\eta| < 2.4$.
 189 In addition, they have to be separated from any lepton candidates passing the Fakeable Object
 190 selection, described below, by requiring $\Delta R = \sqrt{(\eta^\ell - \eta^{jet})^2 + (\phi^\ell - \phi^{jet})^2} > 0.4$.

191 The CSVv2 b-tagging algorithm [18] is used to identify jets that are likely to originate from the
 192 hadronization of bottom quarks. This algorithm combines both secondary vertex information
 193 and track impact parameter information together in a likelihood discriminant. The discrimi-
 194 nant output value ranges from zero to one. It distinguishes between b -jets and jets originating
 195 from light quarks, gluons and charm quarks. The efficiency to tag b -jets and the rate of misiden-
 196 tification of non- b jets depend on the operating point chosen. Both the efficiency and the fake
 197 rate are parameterised as a function of the transverse momentum and pseudorapidity of the
 198 jets. These performance measurements are obtained directly from data in samples that can be
 199 enriched in b jets, such as $t\bar{t}$ and multijet events where a muon can be found inside the one
 200 of jets. Two working points for the CSVv2 output discriminant are used in the analysis. The
 201 *loose* one ($\text{CSVv2} > 0.5426$) has approximately 85% efficiency to tag jets with b quarks and a
 202 10% chance to tag jets with only light quarks or gluons. The *medium* working point ($\text{CSVv2} >$
 203 0.8484) has approximately 70% efficiency for tagging jets with b quarks and 1.5% efficiency to
 204 tag jets with only light quarks or gluons [18].

206 3.2 Missing Energy

The missing transverse energy vector is calculated offline as the negative of the vector sum
 of transverse momenta of all PF candidates identified in the event. The magnitude of this
 vector is referred to as E_T^{miss} . To recover from the performance degradation of the missing
 transverse energy due to pile-up interactions, we also consider the H_T^{miss} variable, computed in
 the same way as the E_T^{miss} , but using only the selected jets and leptons (the lepton selection will
 be described in the following paragraphs). The H_T^{miss} variable has worse resolution than E_T^{miss}
 but is more robust as it does not rely on the soft part of the event. In this analysis the event
 selection makes use of a linear discriminator based on the two variables, $E_T^{\text{miss}} \text{LD}$, exploiting
 the fact that E_T^{miss} and H_T^{miss} are less correlated in events with instrumental missing energy with

respect to events with real missing energy. The $E_T^{\text{miss}} \text{LD}$ is defined as

$$E_T^{\text{miss}} \text{LD} = E_T^{\text{miss}} * 0.00397 + H_T^{\text{miss}} * 0.00265 \quad (1)$$

207 and the working point used is $E_T^{\text{miss}} \text{LD} > 0.2$, as in HIG-15-008.

208 3.3 Lepton Identification

209 In this analysis, *background leptons* are defined as leptons coming from b-hadron decays, the
 210 misidentification of light jets, and photon conversions. We define *signal leptons* as the isolated
 211 leptons coming from W, Z, and τ decays.

212 The reconstruction and identification of electron and muon candidates is described first, fol-
 213 lowed by the advanced identification criteria used to retain the highest possible efficiency for
 214 signal leptons while rejecting background leptons.

215 3.3.1 Muons reconstruction and identification

216 Muon candidates are reconstructed combining the information from both the silicon tracker
 217 and the muon spectrometer in a global fit [19]. An identification selection is performed using
 218 the quality of the geometrical matching between the tracker and the muon system measure-
 219 ments.

220 Two working points are considered for the muon identification. The loose working point,
 221 "POG Loose ID" described in [20], and a tighter working point given by the list of require-
 222 ments on the muon segment-compatibility variable, known as the "POG Medium Id", defined
 223 in [21]. The usage of each working point will be described in Table 6. Only muons within the
 224 muon system acceptance $|\eta| < 2.4$ and minimum p_T cuts of 5 GeV are considered.

225

226 3.3.2 Electron reconstruction and identification

227 Electrons are reconstructed using tracking and electromagnetic calorimeter information by
 228 combining ECAL superclusters and gaussian sum filter (GSF) tracks. We require electrons to
 229 have $|\eta| < 2.5$ to ensure that they are within the tracking volume and a minimum p_T of 7 GeV.
 230 The electron identification is performed using a multivariate discriminant built with shower-

231 shape variables ($\sigma_{i\eta i\eta}$, $\sigma_{i\phi i\phi}$, the cluster circularity, widths along η and ϕ , R_9 , H/E , $E_{\text{inES}}/E_{\text{raw}}$),
 232 track-cluster matching variables ($E_{\text{tot}}/p_{\text{in}}$, $E_{\text{Ele}}/p_{\text{out}}$, $\Delta\eta_{\text{in}}$, $\Delta\eta_{\text{out}}$, $\Delta\phi_{\text{in}}$, $1/E - 1/p$) and track
 233 quality variables (χ^2 of the KF and GSF tracks, the number of hits used by the KF/GSF filters,
 234 fbrem). A complete description of the multivariate discriminant (MVA ID) and training
 235 used can be found in [22]. A loose selection based on eta-dependent cuts on this discriminant
 236 is used to preselect our electron candidates, the full shape of the discriminant is used in the
 237 lepton multivariate selection to separate signal leptons from background leptons. Additional
 238 identification criteria are applied for electrons with p_T greater than 30 GeV to mimic the identi-
 239 fication applied at trigger level in order to ensure consistency between the *measurement* region
 240 and *application* region of the fake-rate (as it will be described in dedicated sections).

241 All the selection criteria will be described in Table 7.

242 3.3.3 Lepton vertexing

243 With the goal of rejecting pile-up or mis-reconstructed tracks, and more importantly to reject
 244 background leptons from b-hadron decays, the following impact parameter variables are also
 245 considered: impact parameter in the transverse plane d_0 , impact parameter along the z axis d_z ,
 246 and the impact parameter significance in the detector space SIP_{3D} .

- 247 Loose cuts are applied on this variables to achieve the first goal, while the full shape of the same
 248 variables is used in a multivariate approach to reach the best separation between the signal and
 249 the background leptons.
 250 The details of the selections are provided in Table 6 and Table 7.

251 **3.3.4 Lepton isolation**

The charged leptons produced in decays of heavy particles, such as W and Z bosons, are typically spatially isolated from the hadronic activity in the event, while the leptons produced in the decays of hadrons or misidentified leptons are usually embedded in jets. This distinction becomes less evident moving to highly boosted systems where decay products tend to overlap. Therefore, given the higher collision energy, instead of using the standard PF Isolation where all the neutral, charged hadrons and photons are considered in a cone of $\Delta R = \sqrt{(\eta^\ell - \eta^i)^2 + (\phi^\ell - \phi^i)^2} < 0.3$ around the leptons, a new isolation variable is constructed: the mini isolation I_{mini} .

Requiring I_{mini} below a given threshold ensures that the lepton is locally isolated, even in boosted topologies. The impact of pileup is mitigated using the so-called effective area correction:

$$I_{\text{mini}} = \frac{\sum_R p_T(h^\pm) - \max(0, \sum_R p_T(h^0) + p_T(\gamma) - \rho \mathcal{A} \left(\frac{R}{0.3}\right)^2)}{p_T(\ell)}. \quad (2)$$

where ρ is the pileup energy density, while $\sum_R p_T(h^\pm)$, $\sum_R p_T(h^0)$ and $\sum_R p_T(\gamma)$ refers to the sum of the transverse momentum of the charged hadrons, neutral hadrons and photons, respectively, within a cone R , dependent of the lepton p_T :

$$R = \frac{10}{\min(\max(p_T(\ell), 50), 200)} \quad (3)$$

The effective areas \mathcal{A} used are listed in Table 5. A very loose cut on this variable is applied to

$ \eta $ range	$\mathcal{A}(e)$ neutral/charged	$\mathcal{A}(\mu)$ neutral/charged
0.0 – 0.8	0.1607 / 0.0188	0.1322 / 0.0191
0.8 – 1.3	0.1579 / 0.0188	0.1137 / 0.0170
1.3 – 2.0	0.1120 / 0.0135	0.0883 / 0.0146
2.0 – 2.2	0.1228 / 0.0135	0.0865 / 0.0111
2.2 – 2.5	0.2156 / 0.0105	0.1214 / 0.0091

Table 5: Effective areas, for muons and electrons

- 252 pre-select the muon and electron candidates, while the full shape is used in the multivariate
 253 discriminator for the signal lepton selection. Again, details of the selections are provided in
 254 Table 6 and Table 7.

256 **3.3.5 Jet-related variables**

In this analysis the most important source of misidentified leptons comes from the decay of b-hadrons (from $t\bar{t}$ +jets, DY+jets, and W+jets events). We therefore want to use in addition to the vertexing and isolation variables described above additional handles to target the rejection of this particular type of background leptons. These additional variables are related to the jet reconstructed in the event closest to the lepton. In particular we use the PF jets reconstructed around the leptons, requiring $\Delta R = \sqrt{(\eta^\ell - \eta^{\text{jet}})^2 + (\phi^\ell - \phi^{\text{jet}})^2} < 0.5$; charged hadrons from pile-up primary vertices are not removed prior to the jet clustering. The four

considered variables are the ratio between the p_T of the lepton and the p_T of the jet, the CSV b-tagging discriminator value of the jet, the number of charged tracks of the jet, and the p_T^{rel} variable:

$$p_T^{\text{rel}} = \frac{(\vec{p}(\text{jet}) - \vec{p}(\ell)) \cdot \vec{p}(\ell)}{||\vec{p}(\text{jet}) - \vec{p}(\ell)||}. \quad (4)$$

257 In order to avoid an over-correction on prompt leptons, the application of the jet energy cor-
 258 rectio is only applied on the hadronic part of the jet, using the following formula $\text{jet} = \ell + (\text{jet-PU-}\ell) * \text{JEC} - \text{PU}$, where ℓ is the lepton, PU the pileup energy clustered into the jet,
 259 and JEC the jet energy scale correction to be applied to any jet.
 260

261 3.3.6 Lepton MVA discriminator

262 In order to profit from all these handles together, we first preselect our leptons candidates with
 263 the *Loose* selection that will be described in the following, and we then use a multivariate dis-
 264 criminator based on boosted decision tree (BDT) techniques to distinguish signal leptons (from
 265 W , Z , or τ decays) from background leptons (mostly from b-hadron decays). We refer to it
 266 as the *lepton MVA discriminator* throughout this document. The multivariate discriminator is
 267 trained using simulated signal *Loose* leptons from the $t\bar{t}H$ MC sample and fake leptons from
 268 the $t\bar{t}$ +jets MC sample, separately for muons and electrons. The training used in this analysis
 269 is unchanged with respect to the 2015 analysis, detailed in [23]. It uses as input variables the
 270 vertexing, isolation and jet-related variables described so far, the p_T and η of the lepton and two
 271 additional variables that contribute to make it robust also in the rejection of leptons from light
 272 jets mis-identification: the electron MVA ID discriminator and the muon segment-compatibility
 273 variables.
 274

275 3.3.7 Additional requirements

276 In the dilepton final state additional requirements on the quality of the charge assignment are
 277 applied to suppress opposite-sign events in which the charge of one of the leptons is mismea-
 278 sured. For the electrons we require consistency between the independent measurements of the
 279 charge from the ECAL supercluster and the tracker, while for the muons we require the track
 280 transverse momentum to be well measured ($\Delta p_T / p_T < 0.2$). We will refer to these cuts as *tight-charge*
 281
 282 Moreover in order to suppress as much as possible background electrons from photon conver-
 283 sions we reject electrons with missing hits in the innermost layer or associated with a success-
 284 fully reconstructed conversion vertex [24].

285 3.3.8 *Loose, Fakeable Object, Tight* definitions

286 Three different selections are used both for the electron and the muon objects identification: the
 287 *Loose*, the *Fakeable Object*, the *Tight* selection. In the description of the analysis strategy it will
 288 be explained for which purposes the different criteria are used.
 289 For reasons that will explained in the data-driven background prediction session, for the *Fake-
 290 able Object* selections the lepton p_T is intended to $0.90 * p_T(\text{jet})$ with the jet being the one asso-
 291 ciated to the lepton as defined for the jet-related variables computation.
 292 In Table 6 and Table 7 all the criteria on the variables previously described are listed.

293 3.4 Validation of lepton identification variables

294 We validate the modelling of the lepton identification variables in simulation by looking at
 295 three control regions: one enriched in prompt leptons from dileptonic $t\bar{t}$, one enriched in non-
 296 prompt leptons from semi-leptonic $t\bar{t}$, and one enriched in $Z+jets$ events. The first control
 297 region is obtained selecting opposite-sign dilepton events with at least two jets and at least
 298 one medium b-tagged jet or two loose ones; events with more than two leptons are vetoed.
 299 The second control region is obtained selecting same-sign dilepton events with exactly three or
 300 four jets, and exactly one medium b-tagged jet, in order to suppress the contributions from $t\bar{t}V$
 301 and $t\bar{t}H$, and similarly events with more than two leptons are vetoed. The third control region
 302 is obtained selecting same-flavor and opposite-sign pairs of leptons with higher than 25 and
 303 15 GeV plus a third candidate with p_T lower than 50 GeV, for which the transverse mass with
 304 the MET is lower than 55 GeV and the MET < 60 GeV. In all control regions, the trailing lepton
 305 is required only to pass the loose selection, not the lepMVA requirement, so that its properties
 306 can be studied in an unbiased way. A data to simulation comparison is done for the lepton
 307 MVA discriminant and some of the more important inputs: the mini-isolation, SIP_{3D} , p_T^{ratio} , p_T^{rel}
 308 and the b-tagging discriminator of the associated jet (Fig. 6 and 7). In all cases, the simulation
 309 is normalized to data, scaling all contributions by the same factor. The contribution from $t\bar{t}$ is
 310 split according to the origin of the lepton in the simulation: prompt, non-prompt from B hadron
 311 decays ($b \rightarrow \ell_{\text{np}}$), or non-prompt from other origins ($j \rightarrow \ell_{\text{np}}$). A good agreement between data
 312 and simulations is observed overall, while minor discrepancies may be related to the fact that
 313 no corrections are applied to the simulation.

Table 6: Requirements on each of the three muon selections. In the cases where the cut values change between the selections, those values are listed in the table. Otherwise, whether the cut is applied is indicated. A few extra requirements are applied for fakeable objects that fail the lepton MVA requirement, to better control the extrapolation in fragmentation and flavour composition: $p_T^{\text{ratio}} > 0.5$ (for all), jet CSV < 0.3 , segmentCompatibility > 0.3 for muons, electron ID MVA $> 0.0(0.7)$ for electrons in the barrel (endcaps).

Cut	Loose	Fakeable Object	Tight
$ \eta < 2.4$	✓	✓	✓
p_T	> 5	> 15	> 15
$ d_{xy} < 0.05$ (cm)	✓	✓	✓
$ d_z < 0.1$ (cm)	✓	✓	✓
$SIP_{3D} < 8$	✓	✓	✓
$I_{\text{mini}} < 0.4$	✓	✓	✓
is Loose Muon	✓	✓	✓
jet CSV	–	< 0.8484	< 0.8484
is Medium Muon	–	–	✓
tight-charge	–	–	✓
lepMVA > 0.90	–	–	✓

Table 7: Requirements on each of the three electron selections. In the cases where the cut values change between the selections, those values are listed in the table. Otherwise, whether the cut is applied is indicated. In some cases, the cut values change for different η ranges. These ranges are $0 < |\eta| < 0.8$, $0.8 < |\eta| < 1.479$, and $1.479 < |\eta| < 2.5$ and the respective cut values are given in the form (value₁, value₂, value₃). Cuts marked with \dagger are applied only to objects failing the tight selection.

Cut	Loose	Fakeable Object	Tight
$ \eta < 2.5$	✓	✓	✓
p_T	> 7	> 15	> 15 2lss(3l)
$ d_{xy} < 0.05$ (cm)	✓	✓	✓
$ d_z < 0.1$ (cm)	✓	✓	✓
$SIP_{3D} < 8$	✓	✓	✓
$I_{\text{mini}} < 0.4$	✓	✓	✓
MVA ID $> (0.0, 0.0, 0.7)$	✓	✓	✓
$\sigma_{i\eta i\eta} < (0.011, 0.011, 0.030)$	—	✓	✓
$H/E < (0.10, 0.10, 0.07)$	—	✓	✓
$\Delta\eta_{in} < (0.01, 0.01, 0.008)$	—	✓	✓
$\Delta\phi_{in} < (0.04, 0.04, 0.07)$	—	✓	✓
$-0.05 < 1/E - 1/p < (0.010, 0.010, 0.005)$	—	✓	✓
p_T^{ratio}	—	$> 0.5\dagger / -$	—
jet CSV	—	$< 0.3\dagger / < 0.8484$	< 0.8484
tight-charge conversion rejection	—	—	✓
Number of missing hits	< 2	$== 0$	$== 0$
lepMVA > 0.90	—	—	✓

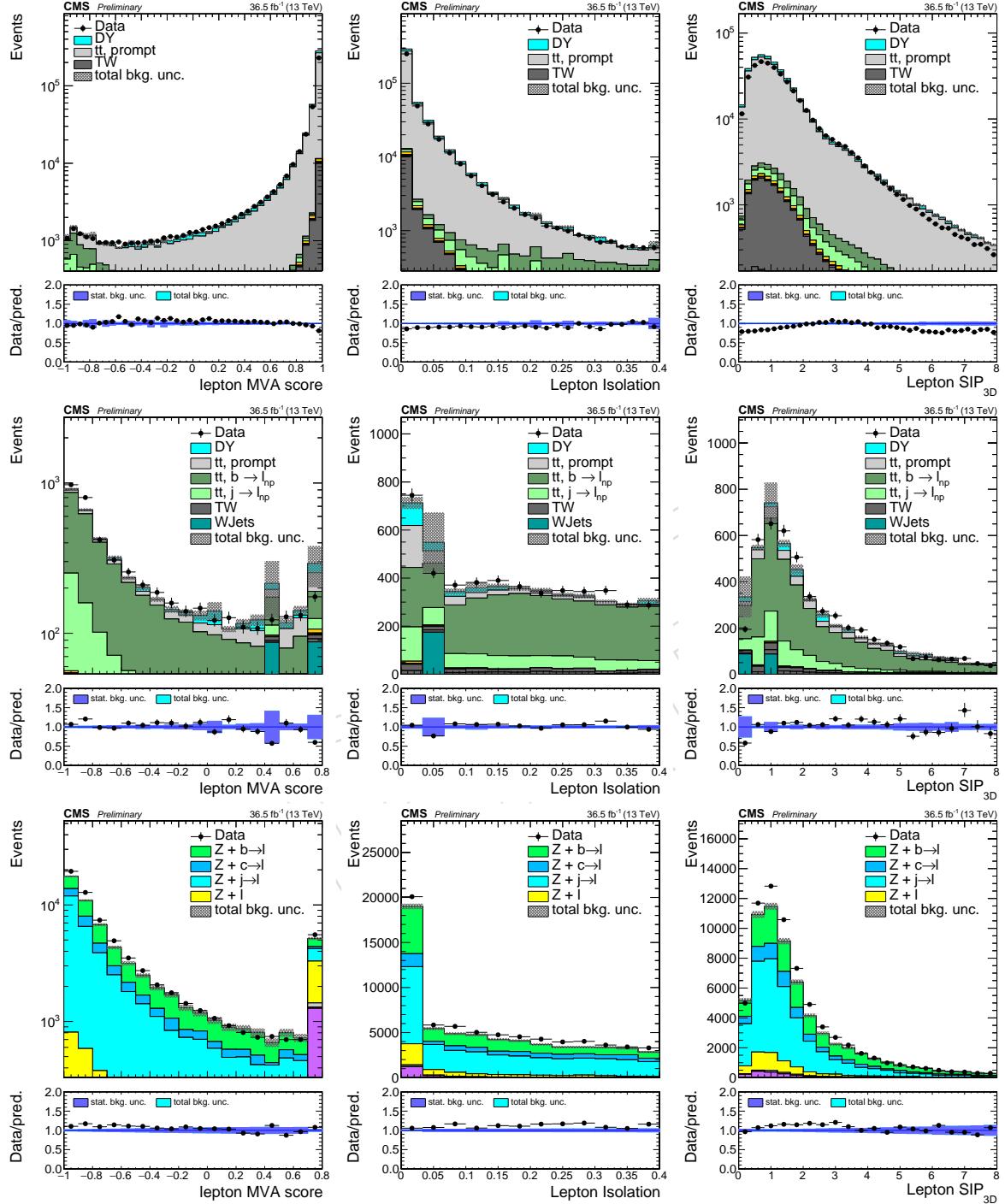


Figure 6: Comparison of the distributions for the lepton MVA (left), mini-isolation (center), and SIP_{3D} (right) between data and simulations in control regions enriched in prompt leptons (top), non-prompt leptons (center), and $Z+jets$ events (bottom), as described in the text. The uncertainty shown on the simulation is only statistical.

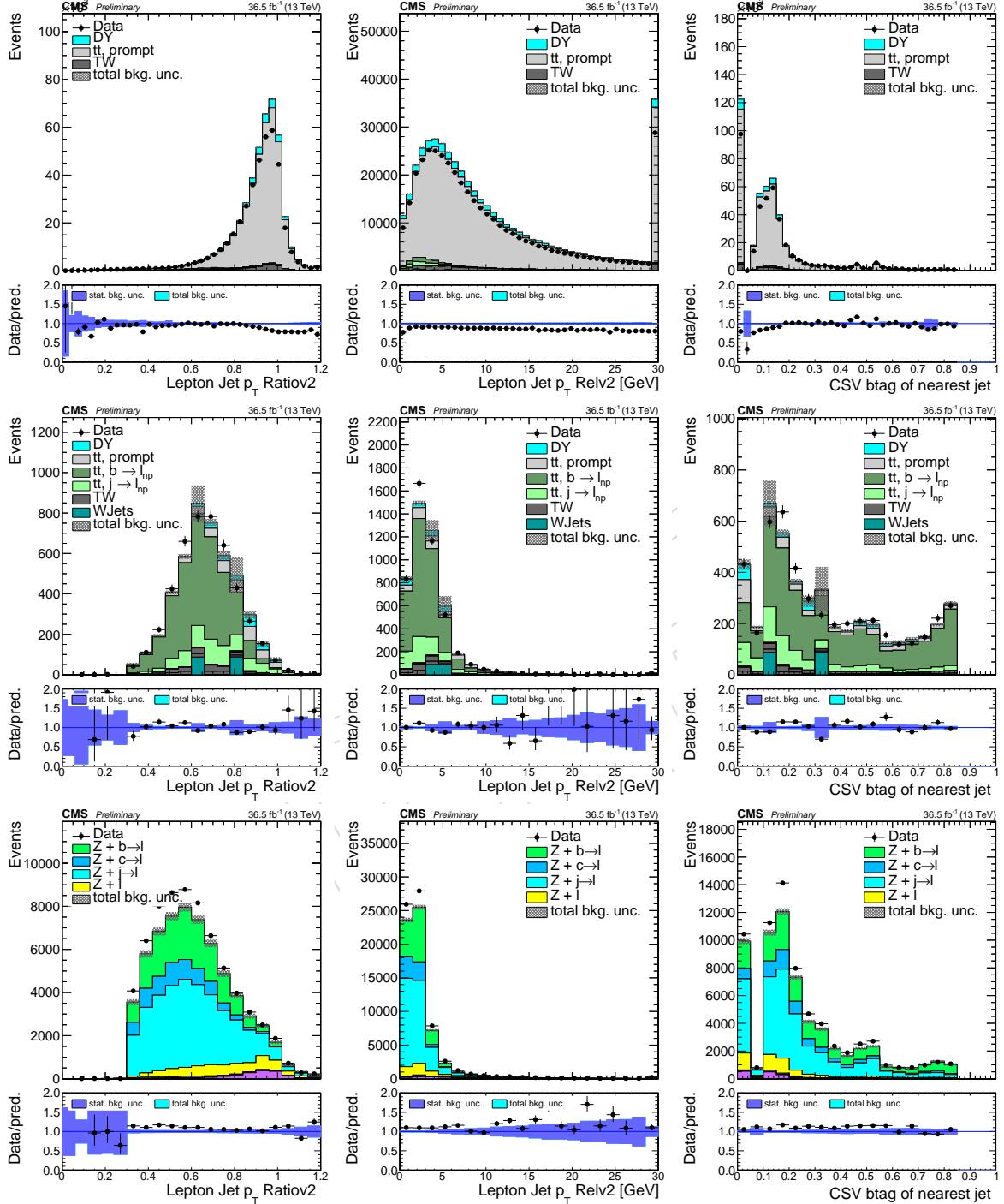


Figure 7: Comparison of the distributions for the lepton p_T^{ratio} (left), p_T^{rel} (center) and the b-tagging discriminator of the associated jet (right), between data and simulations in control regions enriched in prompt leptons (top), non-prompt leptons (center), and Z +jets events (bottom), as described in the text. The uncertainty shown on the simulation is only statistical.

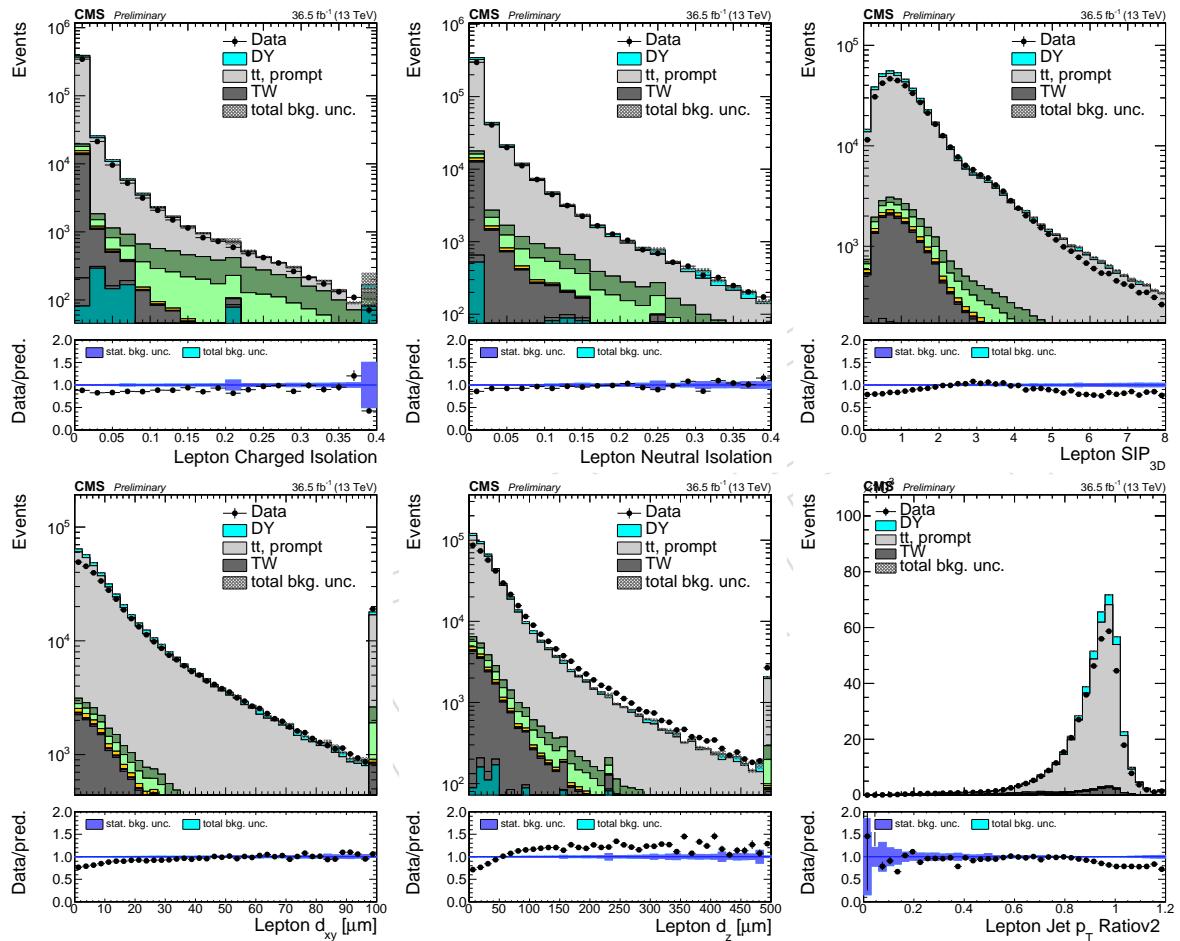


Figure 8: Summary of the input variables to the lepton MVA in a $t\bar{t}$ enriched sample, in data and simulation. The uncertainty shown on the simulation is only statistical.

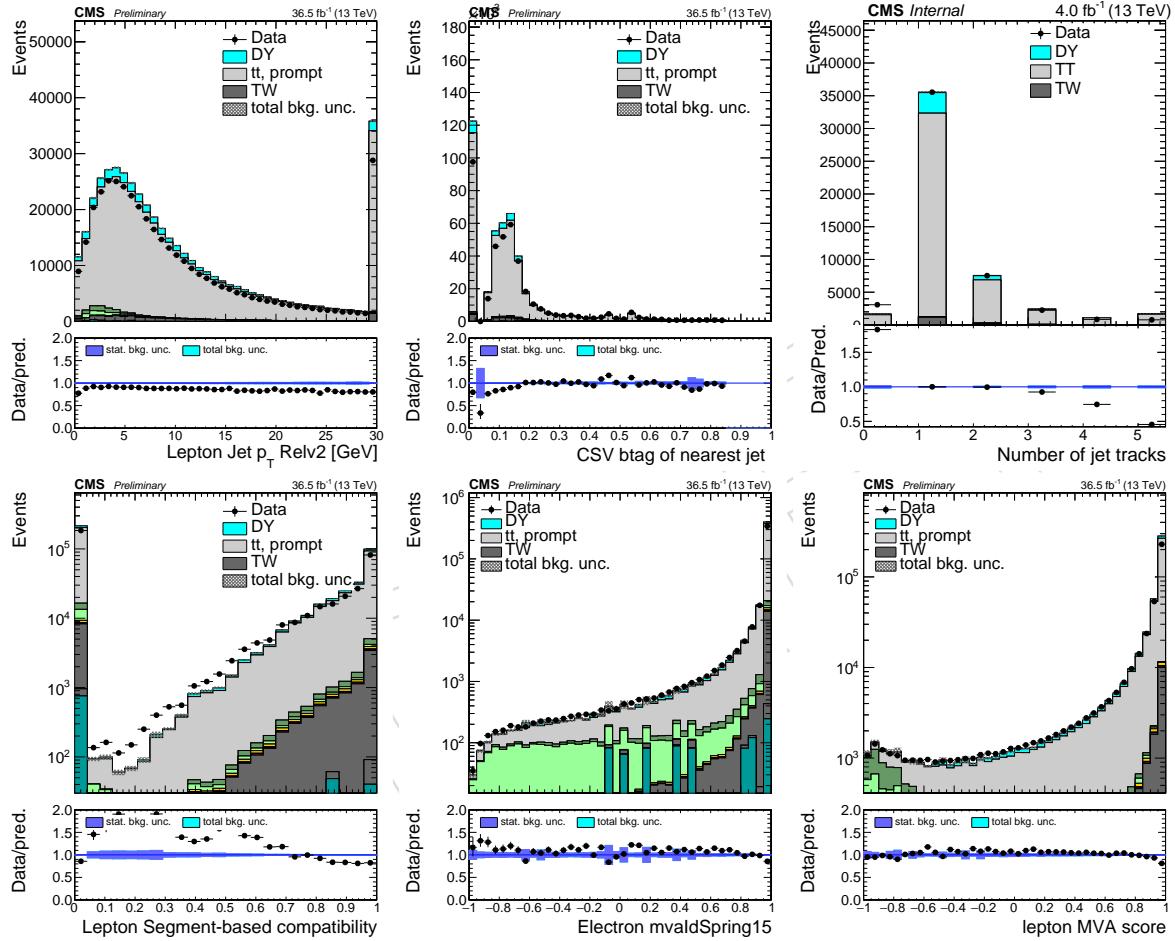


Figure 9: Summary of the input variables (continued) to the lepton MVA and output discriminator in a $t\bar{t}$ enriched sample, in data and simulation. The uncertainty shown on the simulation is only statistical.

314 **3.4.1 Loose selection efficiency**

The reconstruction and loose identification efficiency are computed both for muons and electrons using the Tag and Probe technique with $Z \rightarrow \ell^+ \ell^-$ events, in data and in simulation separately. The efficiency scale factor is therefore defined as:

$$\rho(p_T, \eta) = \frac{\varepsilon_{\text{data}}(p_T, \eta)}{\varepsilon_{\text{MC}}(p_T, \eta)}, \quad (5)$$

where $\varepsilon_i(p_T, \eta)$ is the efficiency measured for a given lepton in the process i (data or simulation). The scale factor is used afterward to correct the weight of the simulated event. The full simulation correction from the lepton side is thus given by the product of all scale factors :

$$\rho = \prod_{j \in \text{leptons}} \rho(p_T(j), \eta_j) \quad (6)$$

315 We apply scale factors for the loose electron selection efficiency that have been derived in the
 316 context of the SUSY lepton SF working group. The efficiency of tight selection cuts has been
 317 measured with the same procedure, with respect to the loose selection. Systematics on these
 318 scale-factors related to the method have been evaluated and are of the order of 2% for both
 319 lepton flavors, in all the considered kinematic range.

320

321 **3.4.2 Tight vs Loose selection efficiency**

322 The efficiencies of applying the tight selection as defined in Tables 6 and 7, on the loose leptons
 323 are determined using a tag and probe method on a sample of DY-enriched events. Numerator
 324 cuts for the same-sign dilepton efficiencies include the tight-charge requirement.

325 Number of passed and failed probes are determined from a fit to the invariant mass of the
 326 dilepton system. The resulting efficiencies are shown in Figures 10, 11, 12, and 13.

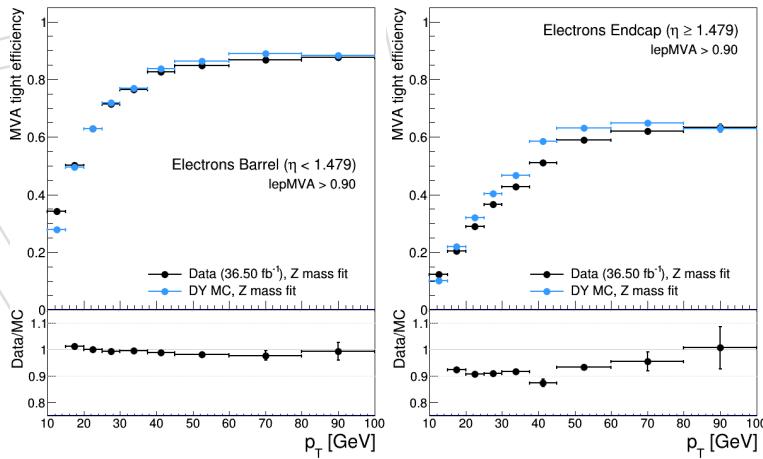


Figure 10: Tight vs loose selection efficiencies for electrons, for the same-sign dilepton lepton definition (i.e. including the tight-charge requirement).

327 We use these (η, p_T) dependent scale factors to correct the simulation.

328 **3.5 Taus**

329 Hadronically decaying taus (τ_h) are reconstructed using the hadron-plus-strips algorithm [25].
 330 τ_h candidates are required to pass the “decay mode finding” discriminator, either being recon-

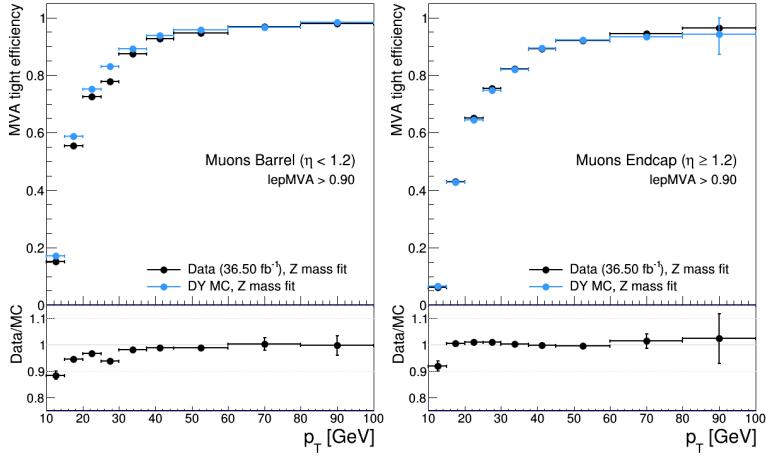


Figure 11: Tight vs loose selection efficiencies for muons, for the same-sign dilepton lepton definition (i.e. including the tight-charge requirement).

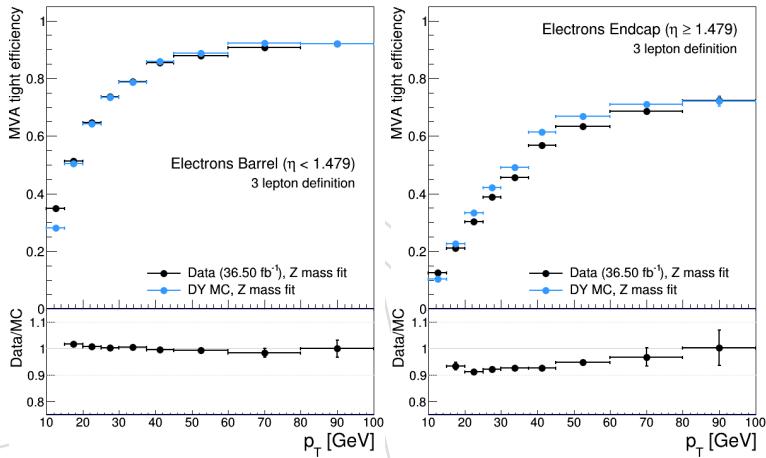


Figure 12: Tight vs loose selection efficiencies for electrons, for the three lepton channel (i.e. not including the tight-charge requirement).

- 331 struncted in 1- or 3-prong decay modes with or without additional π^0 's. In addition, they have
 332 to fulfill $p_T > 20$ GeV and $|\eta| < 2.3$, following Tau POG recommendations.
- 333 The tau identification criteria applied are based on a tau discriminator, using an MVA specifi-
 334 cally trained with $t\bar{t}$ and $t\bar{t}H$ events with an isolation cone of $\Delta R = 0.3$ [26], which increases the
 335 efficiency of the tau isolation in $t\bar{t}H$ with respect to the default discriminators using an isolation
 336 cone of $\Delta R = 0.5$. The medium working point is used for the tau selection ("byMediumIsola-
 337 tionMVArun2v1DBdR03oldDMwLT").
- 338 Reconstructed τ_h candidates are removed if they overlap within $\Delta R = 0.4$ with *loose* electrons
 339 or muons. No dedicated discriminators against background from prompt electrons and muons
 340 are applied since the contribution from background events with additional prompt electrons
 341 and muons passing the τ_h selection criteria but not the muon and electron pre-selection require-
 342 ments is negligible.
- 343 In order to ensure orthogonality of our selection from the phase space covered by the $2lss+\tau_h$,
 344 $3l+\tau_h$ channels, covered in HIG-17-003, we veto the presence in the event of any τ_h passing the
 345 above selection.

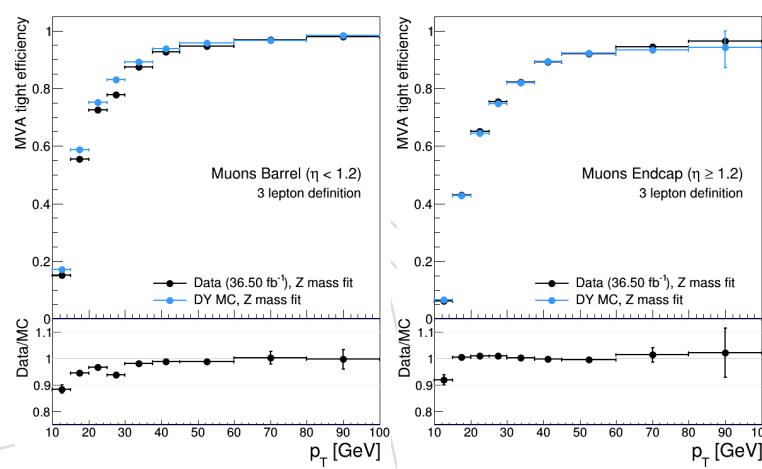


Figure 13: Tight vs loose selection efficiencies for muons, for the three lepton channel (i.e. not including the tight-charge requirement).

346 4 Event selection

347 The event selection aims at rejecting events that do not match the decay signatures targeted
 348 by this analysis. We require that at least two leptons passing the tight selection are present in
 349 the event. Moreover, events where a pair of loose leptons with an invariant mass smaller than
 350 12 GeV is found are rejected, as they are not modeled by the simulation.

351 For all events passing the selection, we require at least two jets with transverse momentum
 352 greater than 25 GeV be reconstructed in the $|\eta| < 2.4$ region. We also require that both jets
 353 satisfy the loose working point of the CSV b-tag algorithm, or that at least one of them satisfies
 354 the medium working point, as a top quark pair decaying into b-jets is present in all signal
 355 events.

356 4.1 Two lepton same-sign category

357 In events where no additional tight lepton with a transverse momentum greater than 10 GeV is
 358 present, we require that the two tight leptons have the same charge and transverse momenta
 359 greater than 25 GeV and 15 GeV respectively. These events constitute the two lepton same-sign
 360 ($2lss$) category of the analysis.

361 In addition to the requirements described above, we discard $2lss$ events that contain less than
 362 four jets with transverse momentum greater than 25 GeV and $|\eta| < 2.4$ in the final state.

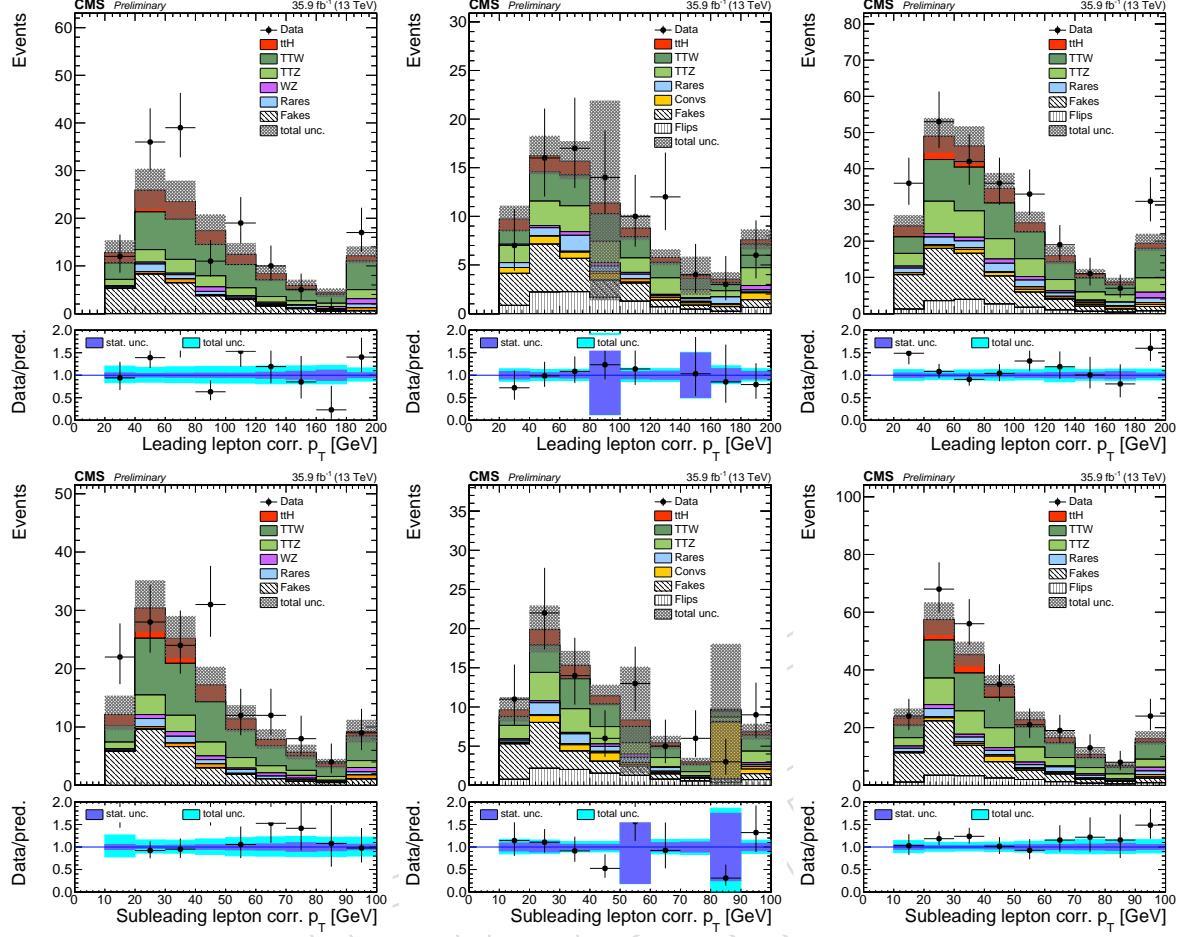
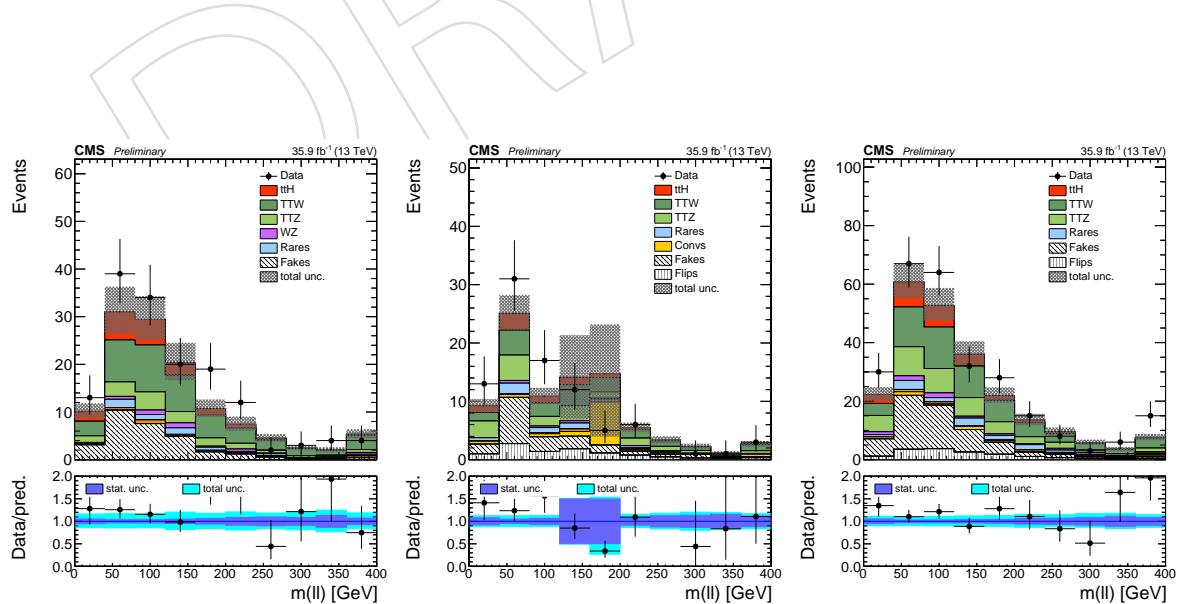
363 The event is also rejected if the two selected leptons do not pass the requirements aimed at re-
 364 jecting leptons from conversions and those on the quality of the charge measurement described
 365 in Section 3. The background from electrons from Z decays, where the charge of one electron
 366 is mismeasured, is further reduced by vetoing events where the di-electron invariant mass dif-
 367 fers by less than 10 GeV from the Z mass. For the same reason, we also require that the $E_T^{\text{miss}} LD$
 368 variable is larger than 0.2 in di-electron events.

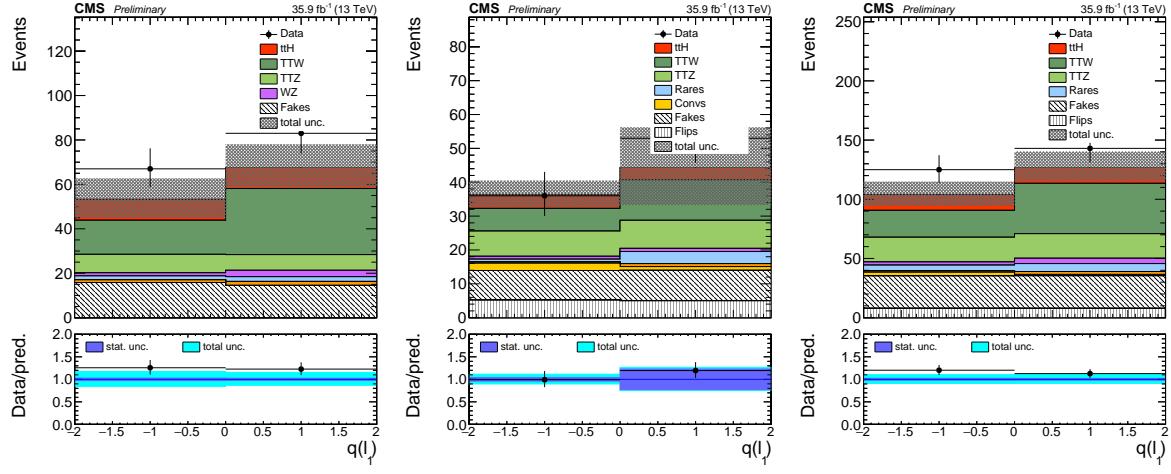
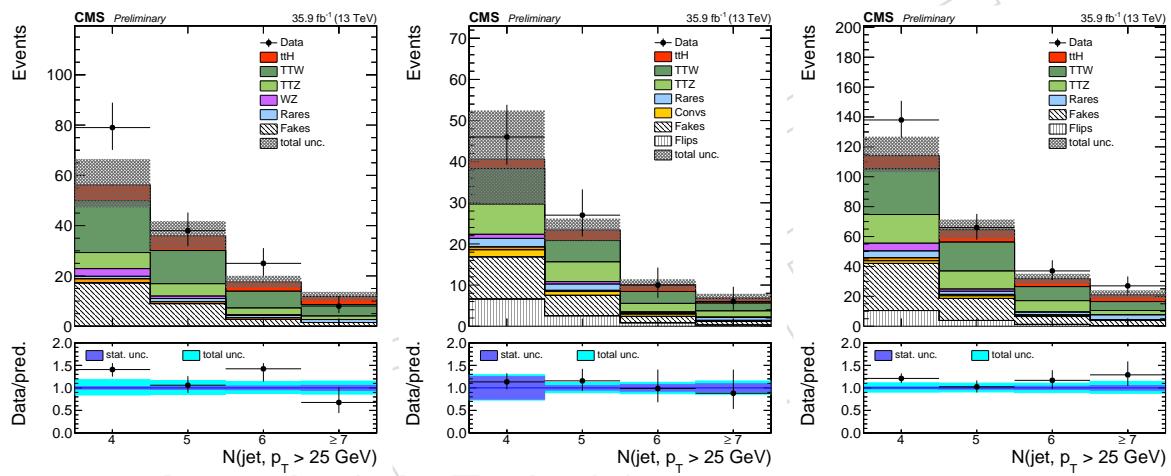
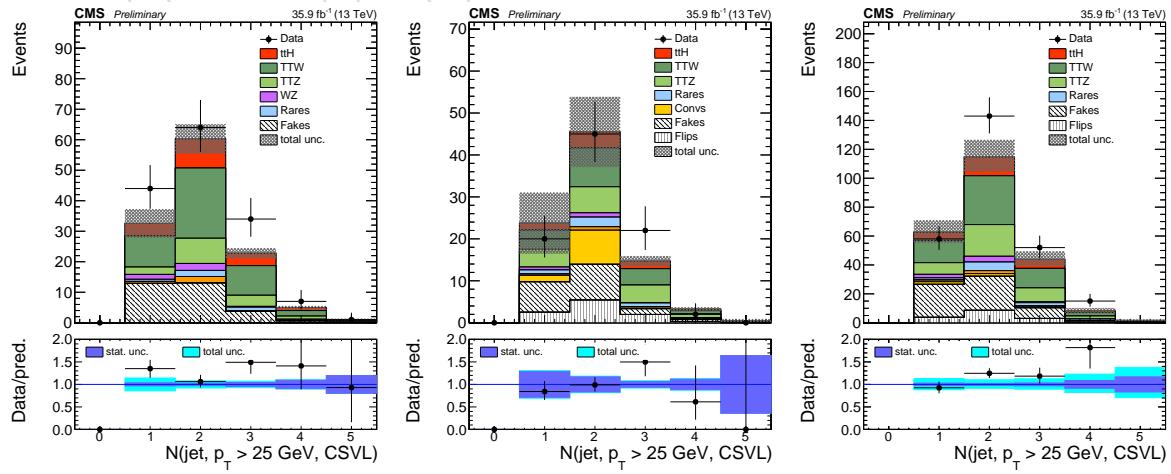
369 4.2 Three and four lepton categories

370 The three lepton ($3l$) category consists of events that contain exactly three tight leptons with a
 371 transverse momenta greater than 25, 15 and 15 GeV respectively. In order to reject backgrounds
 372 from processes with Z bosons in the final state, we require that no pair of same-flavor opposite-
 373 sign loose leptons has an invariant mass closer than 10 GeV to the mass of the Z boson. We then
 374 add an $E_T^{\text{miss}} LD > 0.2$ requirement. The $E_T^{\text{miss}} LD$ threshold is tighter (0.3) if the event has a pair
 375 of leptons with the same flavor and opposite sign. For events with large jet multiplicity (≥ 4
 376 jets), where the contamination from the Z background is smaller, no requirement on $E_T^{\text{miss}} LD$ is
 377 applied. The event is also rejected if the three selected leptons do not pass the conversion veto
 378 requirements, or if the sum of their charges is not equal to +1 or -1. We further veto events with
 379 two OSSF pairs and their 4 lepton mass lower than 140 GeV, where the leptons pass the loose
 380 identification.

381 The four lepton ($4l$) category is defined exactly as the ($3l$) category, except that it requires the
 382 presence of at least four tight leptons with a transverse momenta greater than 25, 15, 15, and 10
 383 GeV respectively.

384 Figures 14-22 and 23-25 show the main event observables (lepton and jet multiplicities and
 385 spectra, energy sums) for events passing the $2l$ and $3l$. Figure 26 shows the number of events
 386 in the $4l$ category.

Figure 14: Lepton transverse momentum spectra in the 2ℓ ($\mu\mu$, ee , $e\mu$) selections.Figure 15: Di-lepton invariant mass spectra in the 2ℓ ($\mu\mu$, ee , $e\mu$) selections.

Figure 16: Sum of lepton charges in the 2ℓ ($\mu\mu$, ee , $e\mu$) selections.Figure 17: Jet multiplicity in the 2ℓ ($\mu\mu$, ee , $e\mu$) selections.Figure 18: Multiplicity of jets passing the loose working point of the CSV tagger in the 2ℓ ($\mu\mu$, ee , $e\mu$) selections.

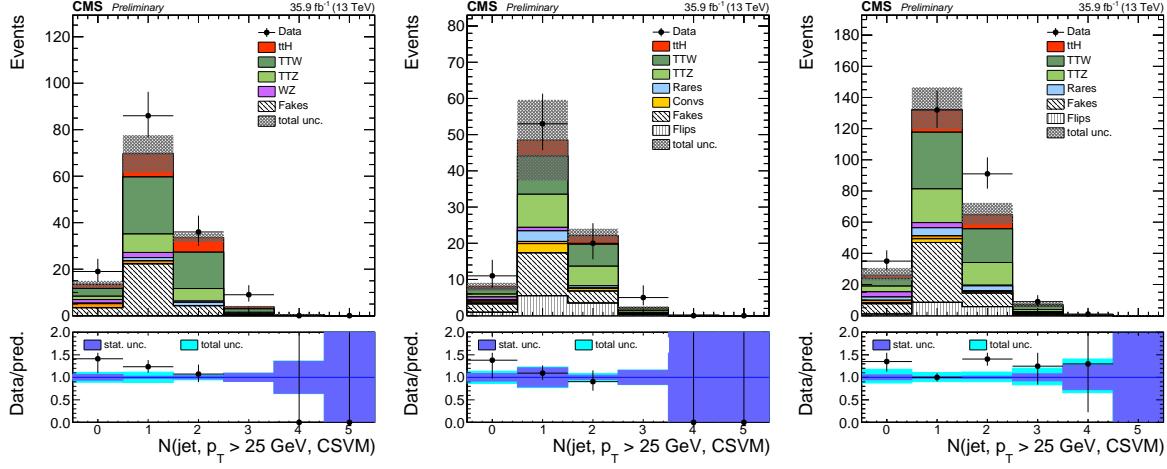


Figure 19: Multiplicity of jets passing the medium working point of the CSV tagger in the 2ℓ ($\mu\mu$, ee , $e\mu$) selections.

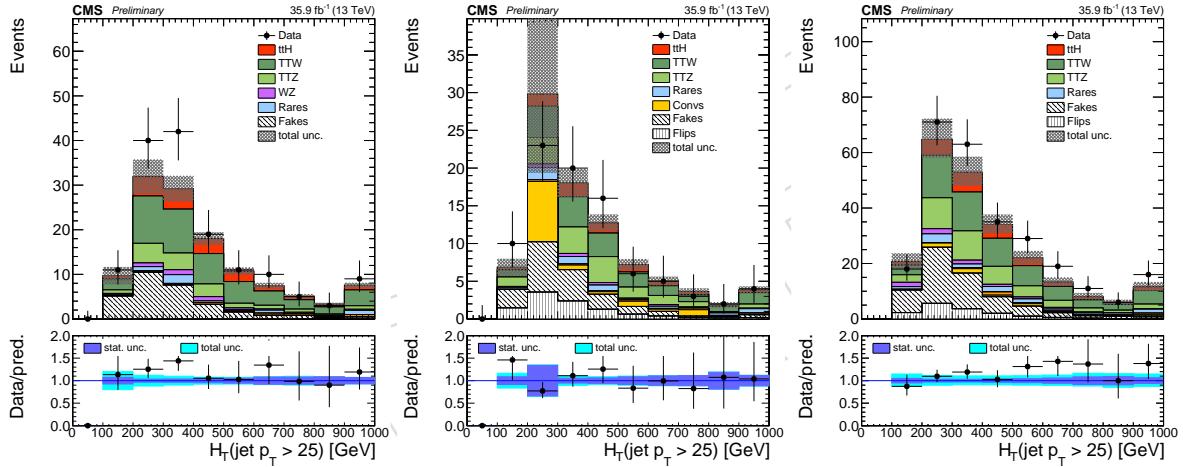


Figure 20: H_T spectra in the 2ℓ ($\mu\mu$, ee , $e\mu$) selections.

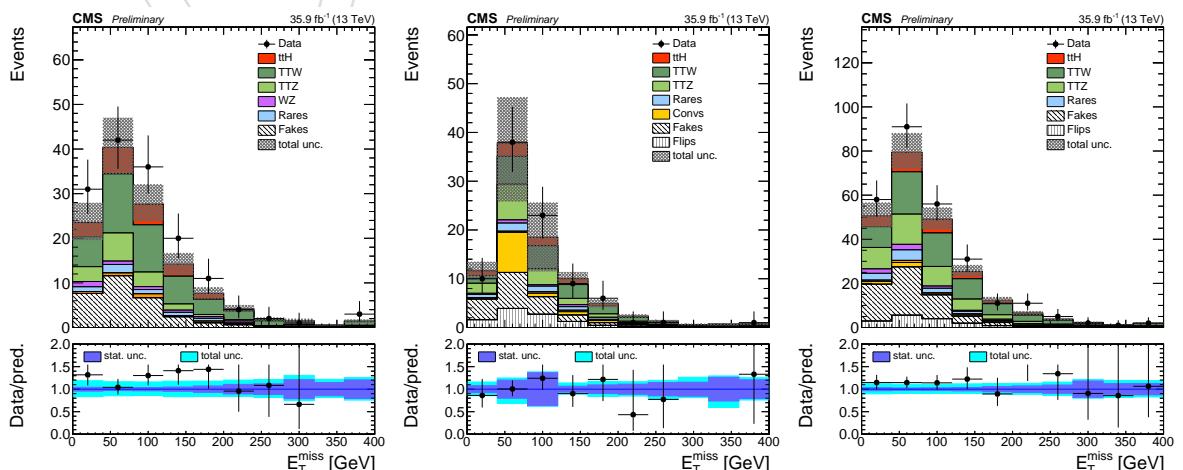
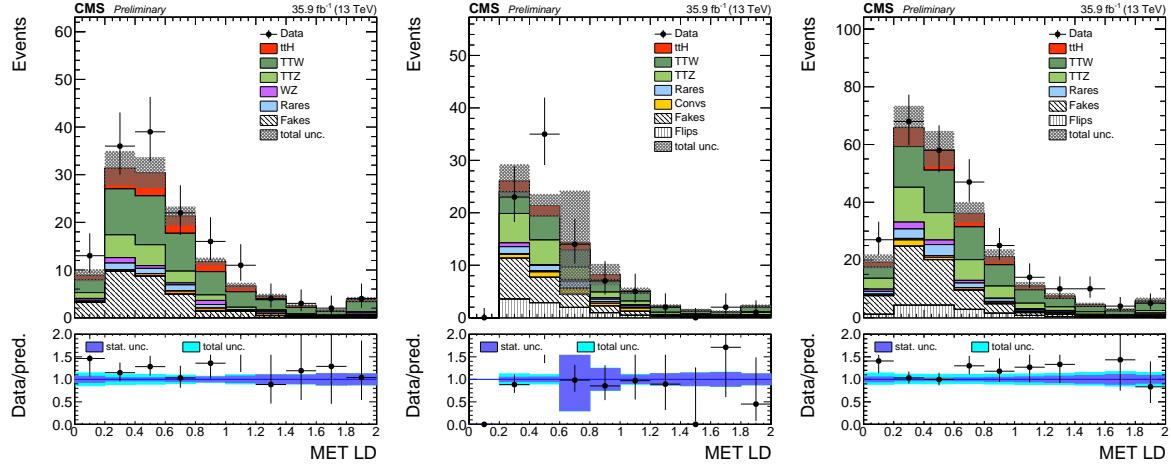
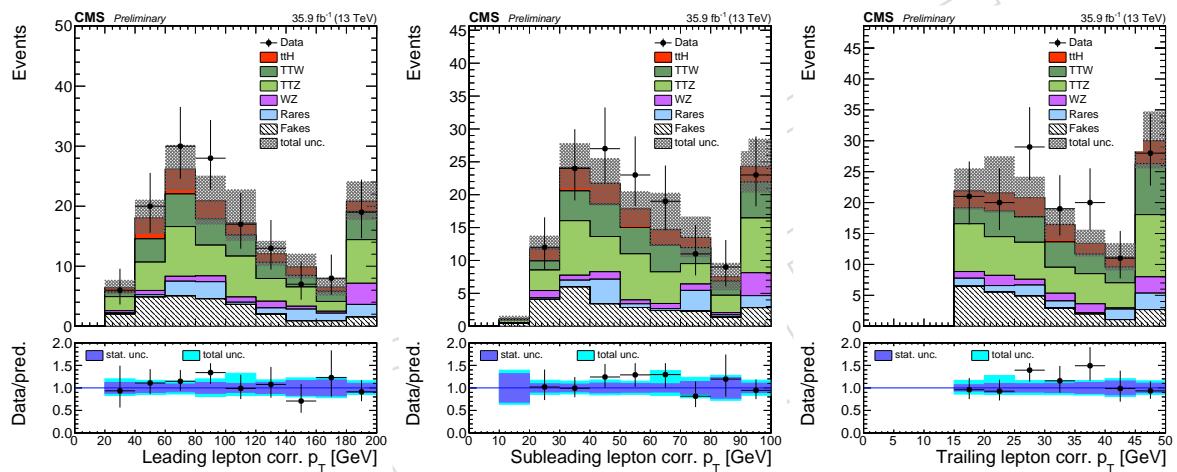
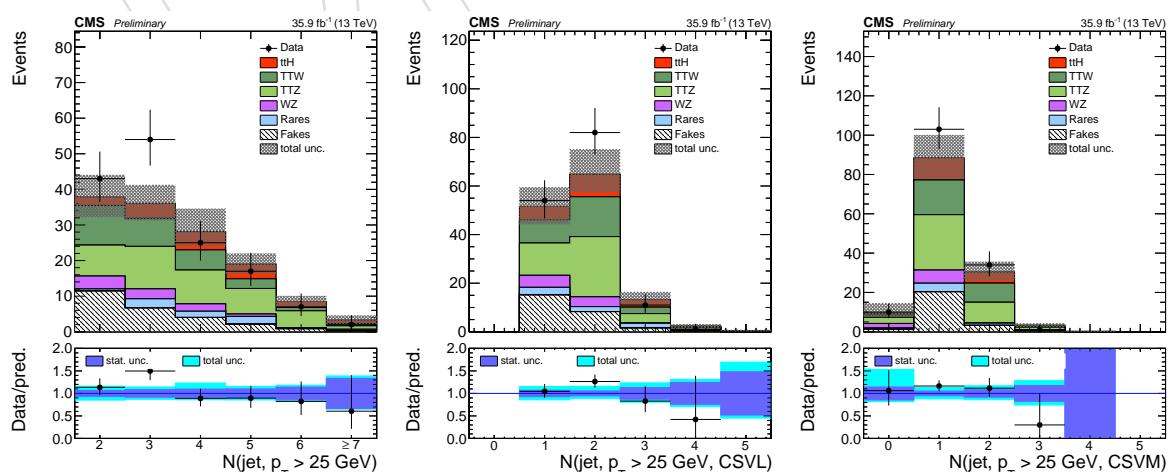


Figure 21: E_T^{miss} spectra in the 2ℓ ($\mu\mu$, ee , $e\mu$) selections.

Figure 22: E_T^{miss} LD spectra in the 2ℓ ($\mu\mu$, ee , $e\mu$) selections.Figure 23: Lepton transverse momentum spectra in the 3ℓ selection.Figure 24: Jet multiplicities (all jets, jets passing the loose working point of the CSV tagger, jets passing the medium working point of the CSV tagger) in the the 3ℓ selection.

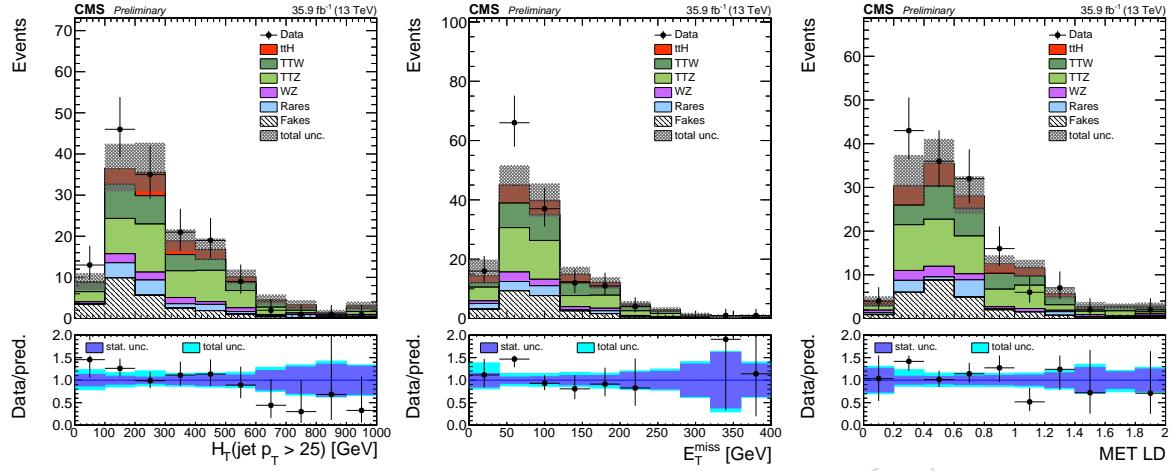


Figure 25: H_T , E_T^{miss} and $E_T^{\text{miss}} \text{LD}$ distributions in the 3ℓ selection.

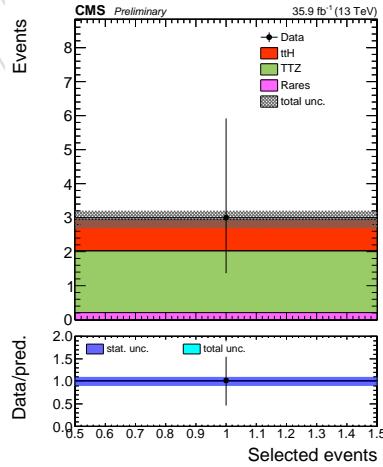


Figure 26: Number of events selected in the 4ℓ category.

387 5 Signal extraction

388 Despite the event selection requirements previously described, the post-selection yields are
 389 still dominated by the backgrounds, and have insufficient statistics to determine the presence
 390 of the $t\bar{t}H$ signal, making further discrimination necessary. The approach adopted in this search
 391 is to split the selected events into several mutually exclusive categories with different signal to
 392 background ratios. In each of these categories the signal is extracted from the distribution of a
 393 suitable discriminating variable.

394 In order to exploit the topological characteristics and specificities of the $t\bar{t}H$ signal with respect
 395 to the most dominant backgrounds, the output of the boosted decision tree (BDT), trained using
 396 a selection of kinematic variables, is used as the discriminating variable for the signal extrac-
 397 tion. In this analysis, the samples used for BDT training and evaluation are the ones reported
 398 in Sec. 2.

399 Both final states with two same sign leptons (2lss) and at least three leptons ($\geq 3l$) have dom-
 400 inant backgrounds originating from the $t\bar{t}$ and $t\bar{t}V$ ($V=W/Z$) processes. In order to have an
 401 efficient discrimination against both of these processes, a two-dimensional (2D) BDT approach
 402 is introduced. For each of the 2lss and $\geq 3l$ final states, the BDT is separately trained against
 403 the $t\bar{t}$ and $t\bar{t}V$, selecting a set of kinematic variables that provide the largest separation in each
 404 training. The BDT outputs of the training against these two processes are used to construct the
 405 2D space, effectively a scatter plot of the two discriminators. The consequent 2D distribution
 406 is then partitioned to rectangular sectors and $t\bar{t}H$ signal and background contributions of each
 407 sector are summed and folded to a one-dimensional histogram. With a convenient partitioning
 408 of the 2D space, the resulting difference of the signal and background shapes is enhanced with
 409 respect to the one-dimensional case, for example against the $t\bar{t}$, and that is provided by the
 410 training of the additional BDT, against the $t\bar{t}V$ process.

411 In the following subsections we detail the definitions of the BDTs used as discriminating vari-
 412 able for the signal extraction in the 2lss and $\geq 3l$ categories. We then describe the criterion to
 413 decide the binning of the 2D BDT. Finally, we detail the precise event subcategorization.

414 5.1 2lss event BDT

415 The training is performed using a relaxed event selection that requires at least two preselected
 416 same sign leptons with leading and trailing lepton transverse momentum larger than 25 and
 417 15 GeV, respectively, plus at least four jets in the event of which either two loose selected b-jets
 418 or one medium b-tagged jet.

419 The training against the $t\bar{t}$ background is performed considering the following input variables:

- 420 • maximum absolute pseudorapidity of the two leading leptons
- 421 • multiplicity of hadronic jets
- 422 • minimum distance between the leading lepton and closest jet
- 423 • minimum distance of the trailing lepton and closest jet
- 424 • transverse mass of the leading lepton and missing transverse energy
- 425 • hadronic top reconstruction (score of the discriminator for best permutation)

426 The training against the $t\bar{t}V$ background is performed considering the following input vari-
 427 ables:

- 428 • maximum absolute pseudorapidity of the two leading leptons

- multiplicity of hadronic jets
- minimum distance between the leading lepton and closest jet
- minimum distance of the trailing lepton and closest jet
- transverse mass of the leading lepton and missing transverse energy
- leading lepton transverse momentum
- trailing lepton transverse momentum
- H_j tagger score after hadronic top jets triplet removal (best permutation)

The sets of observables listed above include new variables that have not been used before, such as the score of the hadronic top reconstruction and the H_j tagger, which are explained in more detail below. The aforementioned hadronic top jet triplet, when using the H_j tagger in the training against the $t\bar{t}V$ background, indicates a triplet of jets with the highest likelihood to originate from hadronic top decay in 2lss events. This jet triplet is found through the event reconstruction BDT explained below, whose score is used as an input in the training against $t\bar{t}$. The H_j tagger, described more in detail in the following, is an algorithm designed to identify jets originating from one of the W of the Higgs in 2lss events. The Higgs jets search is performed against all the jets that enter the 2lss region, except those forming a triplet compatible with an hadronic top decay (from which the expression hadronic top jets triplet removal).

5.1.1 Hadronic top reconstruction

The reconstruction of the hadronic top decay is performed with a BDT. The objective is to correctly match each selected jet and lepton to a final state particle in a $t\bar{t}H$ event, then use the BDT response and other variables from the reconstruction to discriminate against events without a hadronic top.

Event reconstruction targets the 2lss category, specifically where the Higgs decays to W bosons. In the 2lss category, this means that one lepton originates from the top system, and the other from the Higgs. For the jets, one of the W bosons from the Higgs decays hadronically, one of the top quarks decays hadronically, producing a total of two b-jets, and four light-flavor jets from the hadronic W decays.

Training

The event reconstruction BDT is trained using the $t\bar{t}H$ monte-carlo powheg signal sample described previously.

The signal is correctly matched $t\bar{t}H$ events, which pass the 2lss selection. Because the 2lss event selection requires at least four jets, the vast majority of signal events used for training are only partially reconstructed, since a full reconstruction necessitates six matched jets. Because so few events can be fully reconstructed, we must consider partial reconstructions for events that have fewer than six matched jets. The strategy for this is to use 'null' jets whose four-vectors are set to zero to substitute missing jets in the event. Finally we require the signal events to have two correctly matched selected leptons, and at least four correctly matched selected jets.

The background consists of all jet and lepton permutations of incorrectly matched $t\bar{t}$ events. For the background, the null jets are added according to the jet multiplicity. For events with seven or fewer selected jets, three null jets are added, for events with eight selected jets, two null jets are added, and one null jet is added for events with greater than eight selected jets. To reduce the computation time and improve performance, several cuts are applied at each permutation to remove unlikely reconstructions. These cuts include applying the b-tag requirement on the

472 two jets being considered as b-jets (1 b-tight, 2 b-loose) described earlier, requiring that no
 473 reconstructed W have a mass greater than 120 GeV, requiring the leptonic top mass be less
 474 than 180 GeV, and requiring the hadronic top mass be less than 220 GeV. Additionally, we
 475 ignore permutations arising from swapping two light flavor jets from the same W boson, as the
 476 reconstruction is identical.

477 The BDT uses eight input variables, consisting of the CSV of the b-jets from the top system,
 478 the transverse momentum of the reconstructed hadronic top, the mass of the reconstructed
 479 hadronic top, the mass of the W originating from the hadronic top, the transverse momentum
 480 ratio of lepton from the Higgs with the lepton from the top, and the solid angles between
 481 the lepton from the top and each b-jet from the top system, and between the lepton from the
 482 Higgs. This approach focuses on the hadronic top decay, as the other aspects of the event are
 483 more difficult to reconstruct due to the missing energy from the neutrinos.

484 Evaluation

485 The event reconstruction BDT is evaluated by iterating over all possible lepton and jet permu-
 486 tations, and selecting the highest scoring permutation as the reconstruction for each event. For
 487 the evaluation and usage, the null jet prescription and permutation cuts used are identical to the
 488 background training. The reconstruction is designed identify events that have a hadronic top
 489 present and thus offers some discrimination against the semi-leptonic $t\bar{t}$ background, shown
 490 below in Figure 27.

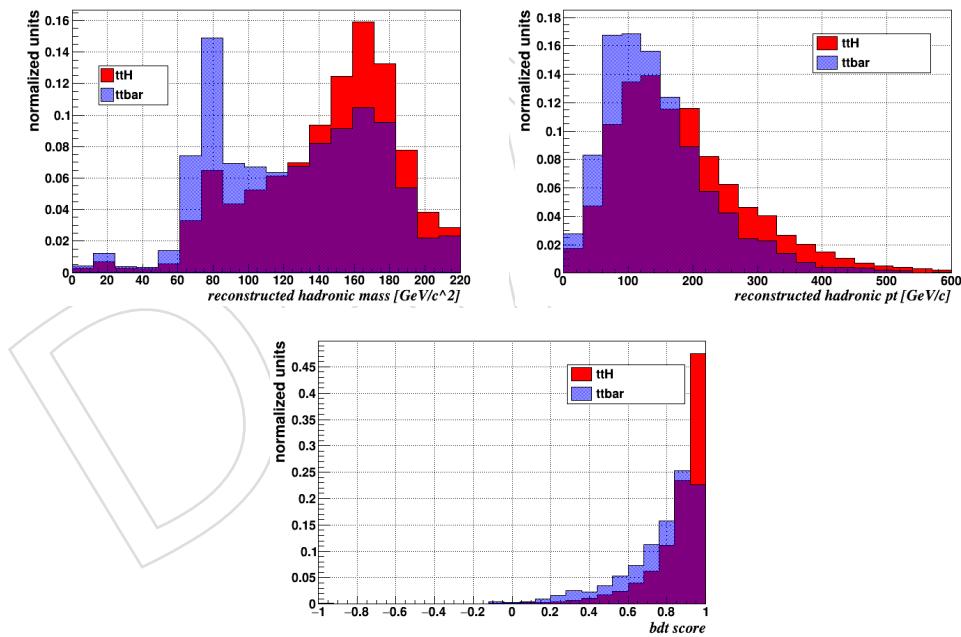


Figure 27: Best BDT score and associated quantities from hadronic top reconstruction.

491 **5.1.2 The H_j tagger**

In this section we describe a discriminator aiming at identifying jets that originates from a Higgs decaying to two Ws. In particular, we target the 2lss category in which the ttH signal decays, in the highest fraction of events, according to the following chain:

$$(t)(t)(H) \rightarrow (bW)(bW)(WW^*) \rightarrow (bjj)(b\ell\nu_\ell)(\ell\nu_\ell jj)$$

492 We therefore expect 2 b quark jets, 4 jets, 2 same-sign leptons, and missing energy in the final
 493 state, altough, in order to increase the signal acceptance, the analysis requires the presence of
 494 at least 4 jets overall. This means that the Higgs jets do not necessarily enter the signal region.

495 In order to deal with both the complicated jet combinatoric and the possibility that not all the
 496 jets originating from Higgs are selected, one discriminator is developed: the Higgs-jet (H_j)
 497 tagger. The H_j tagger is an object discriminator that exploits jet identification and kinematic
 498 properties in order to assess the likelihood of a jet of originating in the decay $H \rightarrow WW^* \rightarrow$
 499 $\ell\nu_\ell jj$.

500 The discriminator is developed considering the BDT multivariate technique. We rely on the
 501 powheg ttH sample and ttV sample to define the signal and the background in the training.
 502 For the H_j tagger the signal is represented by reconstructed jets that are matched at gen-level
 503 to jets of the process $H \rightarrow WW^* \rightarrow \ell\nu_\ell jj$, while the background is given by the reconstructed
 504 jets in ttV events. For the training, we consider the phase space of events that enter the 2lss
 505 category, with 0 τ_h .

506 In the following subsection we list the variables used for the H_j taggers and their expected BDT
 507 distributions.

508 **H_j variables and performances**

509 The variables used for the H_j tagger are:

- 510 • minimum dR of the jet and one of the lepton
- 511 • maximum dR of the jet and one of the lepton
- 512 • jet pT
- 513 • jet b-tagging discriminator
- 514 • jet quark-gluon discriminator

515 The performances of the H_j tagger are illustrated in Fig. 28. The ROC curve (Fig. 28, right)
 516 highlights the improvement in performance of the Moriond 2017 BDT (red line) versus the
 517 ICHEP 2016 BDT (black line).

518 Figures 29 and 30 show a comparison of the simulated signal (ttH) and background (t \bar{t} or t $\bar{t}V$)
 519 processes for each of the input variables to the BDT discriminator. Figure 31 shows the separa-
 520 ration power of the BDT discriminators.

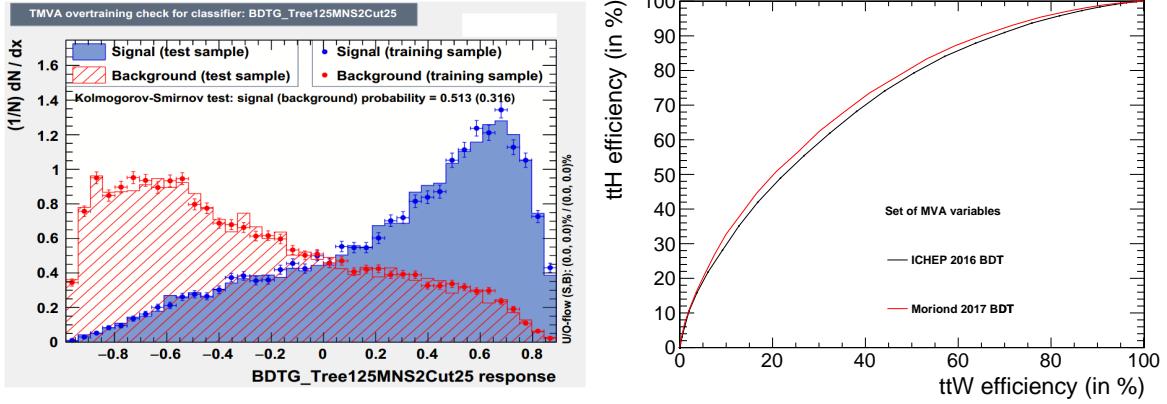


Figure 28: The H_j distribution (left) and ROC (right). The ROC curve highlights the improvement in performance of the Moriond 2017 BDT (red line) versus the ICHEP 2016 BDT (black line). Signal and background composition are described in the text.

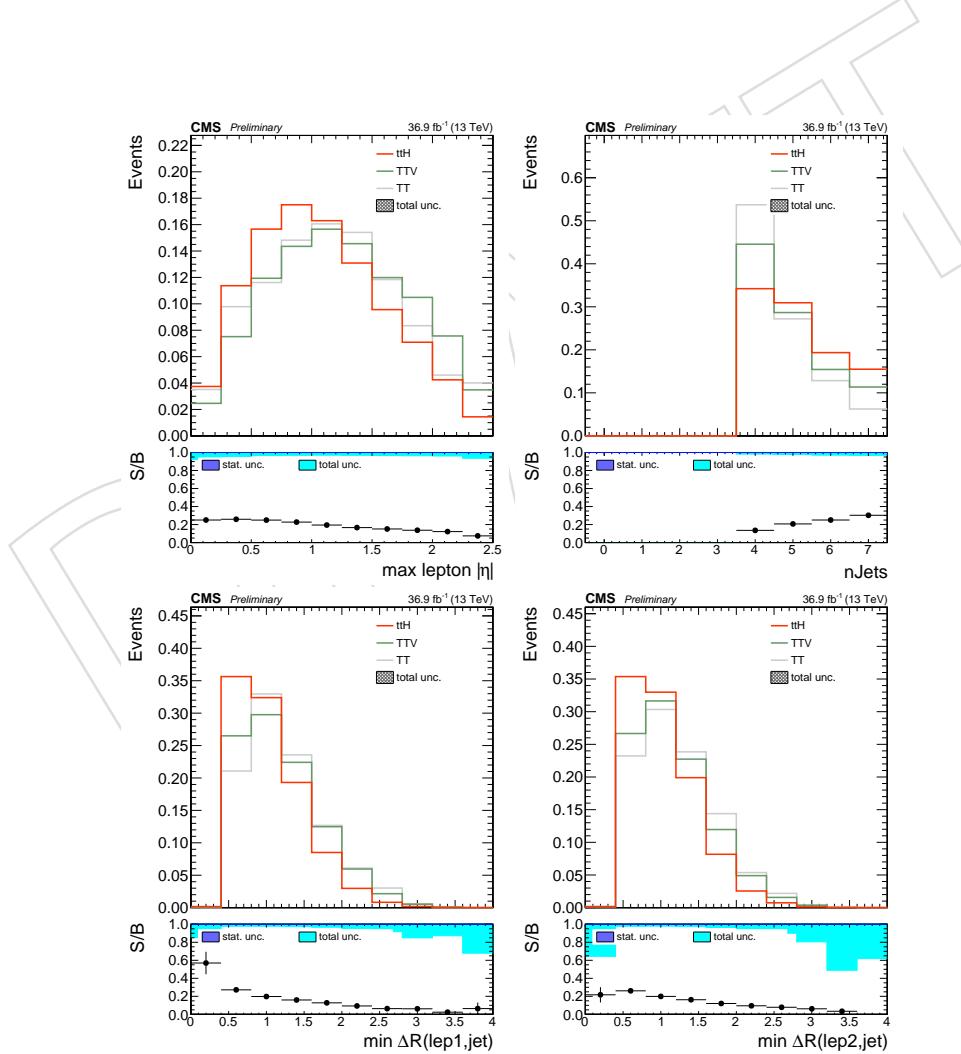


Figure 29: The separation power of the variables used for BDT trainings, in the two same sign leptons channel.

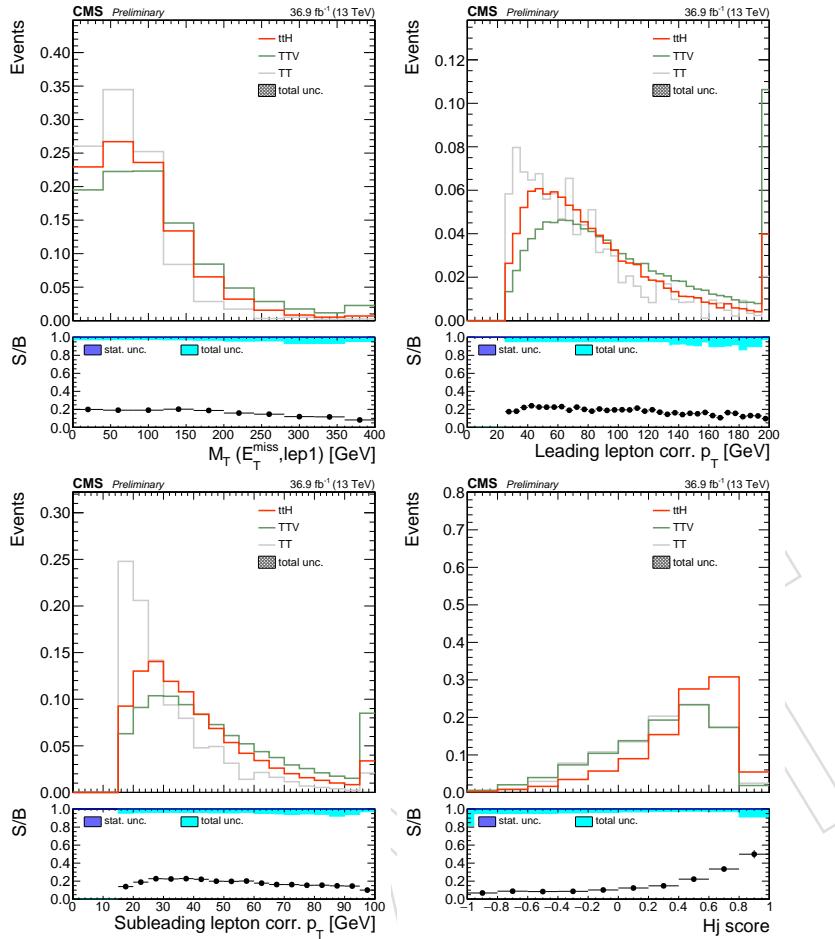


Figure 30: The separation power of the variables used for BDT trainings, in the two same sign leptons channel.

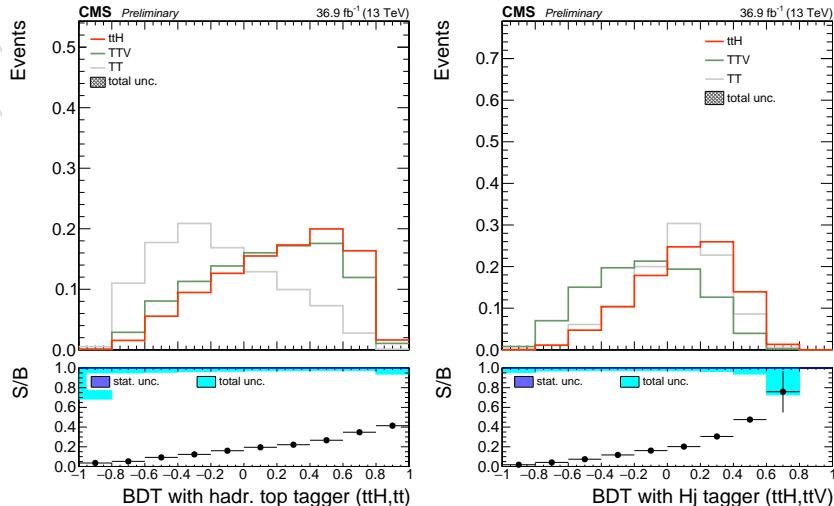


Figure 31: The separation power of the BDT output against the $t\bar{t}$ background (left) and $t\bar{t}V$ (right) background, in the two same sign leptons channel.

521 **5.2 3l event BDT**522 The three lepton category has the following variables for the training against the $t\bar{t}$ background:

- 523
- maximum absolute pseudorapidity of the two leading leptons
 - multiplicity of hadronic jets
 - minimum distance between the leading lepton and closest jet
 - minimum distance of the trailing lepton and closest jet.
 - transverse mass of the leading lepton and missing transverse energy

528 For the training against the $t\bar{t}V$ background, the input variables are:

- 529
- maximum absolute pseudorapidity of the two leading leptons
 - transverse mass of the leading lepton and missing transverse energy
 - multiplicity of hadronic jets
 - minimum distance between the leading lepton and closest jet
 - minimum distance of the trailing lepton and closest jet
 - leading lepton transverse momentum and the third lepton transverse momentum.

535 Furthermore, the performance of the Matrix Element Method (MEM, as described in Appendix B,
536 on page 76) in the three lepton category warrants its inclusion as input to the training against
537 the $t\bar{t}V$ process. For ICHEP2016, the MEM was included using the log of weights for the $t\bar{t}H$,
538 $t\bar{t}W$ and $t\bar{t}Z$ hypotheses. For this iteration of the analysis, it was found that given the $t\bar{t}V$ BDT
539 parameters used, it is more efficient to include solely the likelihood of background vs signal +
540 backgrounds (one variable instead of three). This improves the performance by a few percent.541 In addition, it was found that the $t\bar{t}$ BDT could also be improved by including the MEM among
542 the input variables, using the MEM weights of $t\bar{t}H$, $t\bar{t}$, and the weights obtained from kinematic
543 reconstruction of fully and semi leptonic $t\bar{t}H$ hypothesis are included. It improves the perfor-
544 mance of the $t\bar{t}$ BDT by about 5%. *In the current version of the analysis, this is not yet implemented*
545 *because of CPU time constraints (MEM not yet run on the $t\bar{t}$ sample).*546 Because the currently available Monte-Carlo statistics prevent us from training with the full
547 event selection, a relaxed selection has been applied instead for training. This relaxed selec-
548 tion requires at least three preselected leptons where neither lepton pair has an invariant mass
549 within 10 GeV of the mass of the Z boson, the leading, trailing and sub-trailing lepton trans-
550 verse momentum larger than 25, 15 and 15 GeV, respectively, the MET LD requirement applied
551 and at least two loose selected b-jets in the event. In Figures 32 and 33 a comparison of the
552 simulated signal ($t\bar{t}H$) and background ($t\bar{t}$ or $t\bar{t}V$) processes for each of the input variables to
553 the BDT discriminator is presented. Further details on the BDT training are available in Ap-
554 pendix E.

555 Figure 34 shows the separation power of the BDT discriminators.

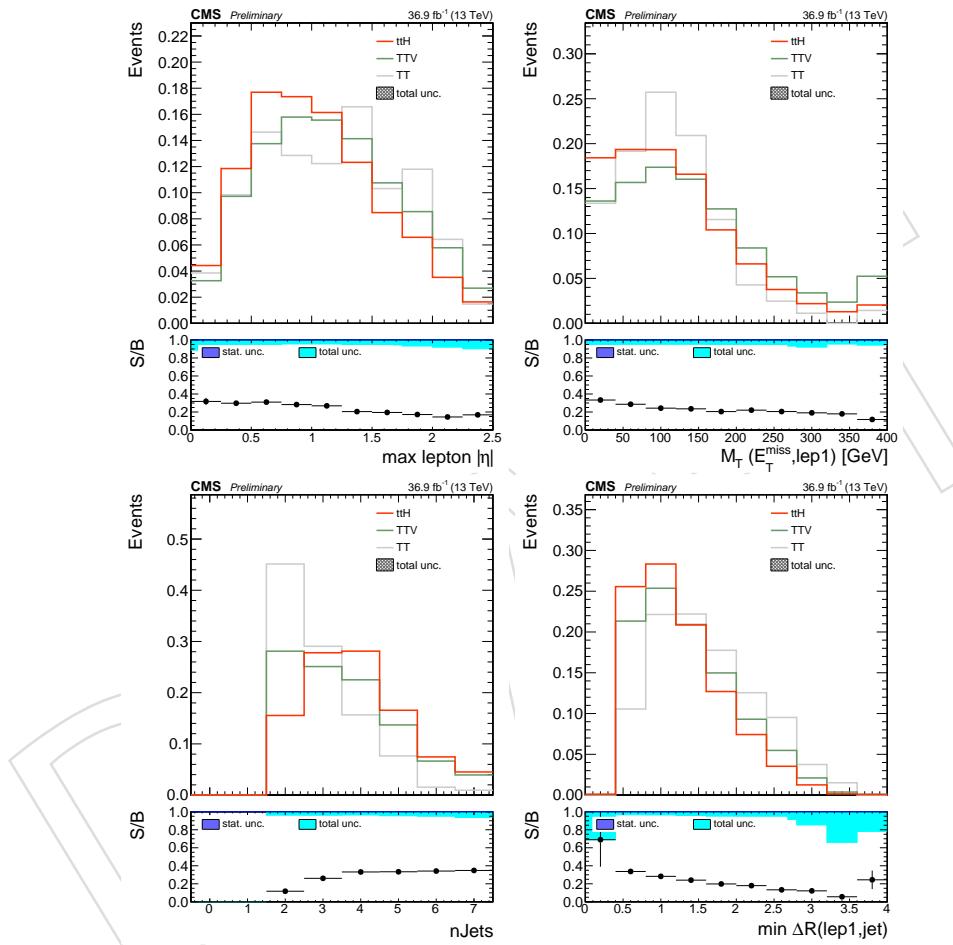


Figure 32: The separation power of the variables used for BDT trainings, in the three lepton channel.

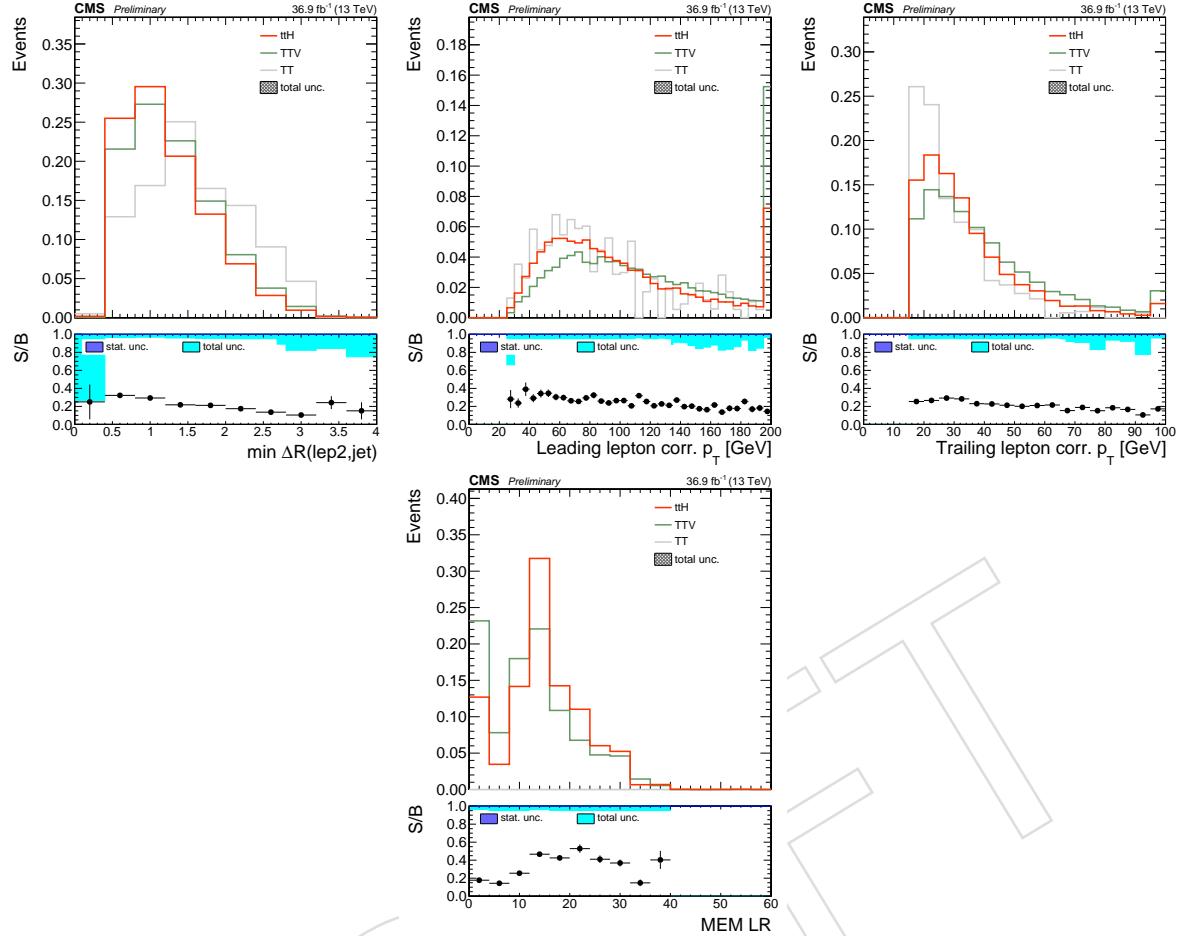


Figure 33: The separation power of the variables used for BDT trainings, in the three lepton channel.

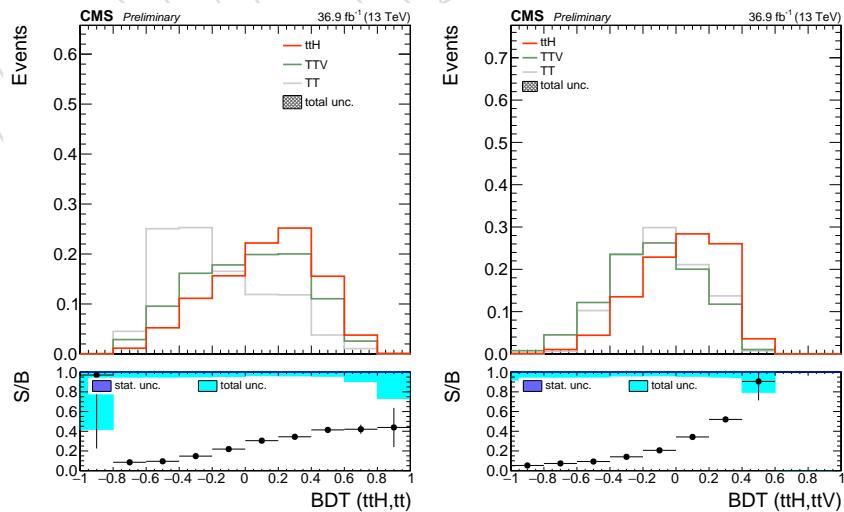


Figure 34: Distribution of the discriminator against the $t\bar{t}$ bar (left) and $t\bar{t}V$ (right) backgrounds, in the three lepton channel, with reducible background prediction from MC (top) and from data (bottom).

556 **5.3 2D BDT binning**

557 The two-dimensional plane spanned by the output of the two BDTs is populated by signal and
 558 background events in a non-uniform way: the two-dimensional distribution of the discrimina-
 559 tors for the ttH signal (top), and for the ttbar (bottom left) and ttV (bottom right) backgrounds,
 560 are shown in Figures 35 and 36 for the same-sign dileptonic and the multileptonic inclusive
 561 signal regions, respectively.

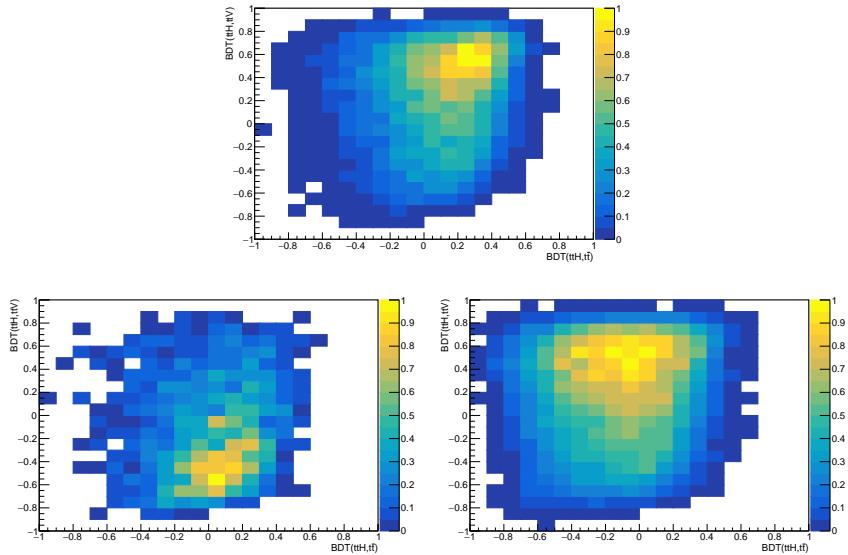


Figure 35: Two-dimensional distribution of the discriminators for the $t\bar{t}H$ signal (top), and for the $t\bar{t}$ (bottom left) and $t\bar{t}V$ (bottom right) backgrounds, in the two same sign leptons channel, estimated from MC.

562 The two-dimensional plane is then partitioned in several regions, depending on their different
 563 signal or background composition, with the objective of maximizing the analysis sensitivity
 564 with the available luminosity. To do so, a method based on the likelihood ratio of signal and
 565 background is used.

566 Formally, this corresponds to building a multivariate classifier to reduce the dimensionality of
 567 the problem from two to one.

568 The plane is populated with events taken from simulated samples that are not used in the main
 569 analysis. Furthermore, for each simulated sample ($t\bar{t}H$ signal, $t\bar{t}$ and $t\bar{t}V$ backgrounds) half of
 570 the events are used for training, and the other half are used to evaluate the performance of the
 571 classifier.

572 *As the MEM method is vastly computer intensive, it has not been possible to compute the matrix element
 573 for the $t\bar{t}$ sample in the three leptons final state. For this reason, $t\bar{t}V$ events are used to assess the
 574 eventual correlation between the $BDT(t\bar{t}H, t\bar{t}V)$ outputs obtained when training with or without the
 575 MEM discriminator among the inputs. Figure 37 shows the positive correlation found between the
 576 $BDT(t\bar{t}H, t\bar{t}V)$ output in $t\bar{t}V$ events. A linear dependence is assumed, and under this assumption a
 577 function relating the non-MEM-trained BDT output and the MEM-trained BDT output is found via a
 578 χ^2 fit to the $t\bar{t}V$ simulated events. The parameters of the fit are determined with 2–10% accuracy, and
 579 the resulting estimator, $\hat{BDT}_{mem}(t\bar{t}H, t\bar{t}V) = 0.04804 + 0.6902 \times BDT(t\bar{t}H, t\bar{t}V)_{nomem}$, is used in
 580 place of the output of the $BDT(t\bar{t}H, t\bar{t}V)$ trained without MEM discriminator as an input. This relies
 581 on the assumption that the introduction of the MEM discriminator as an input to the $BDT(t\bar{t}H, t\bar{t}V)$
 582 training has the same effect in $t\bar{t}$ events as in $t\bar{t}V$ events. As soon as it will be possible to compute MEM*

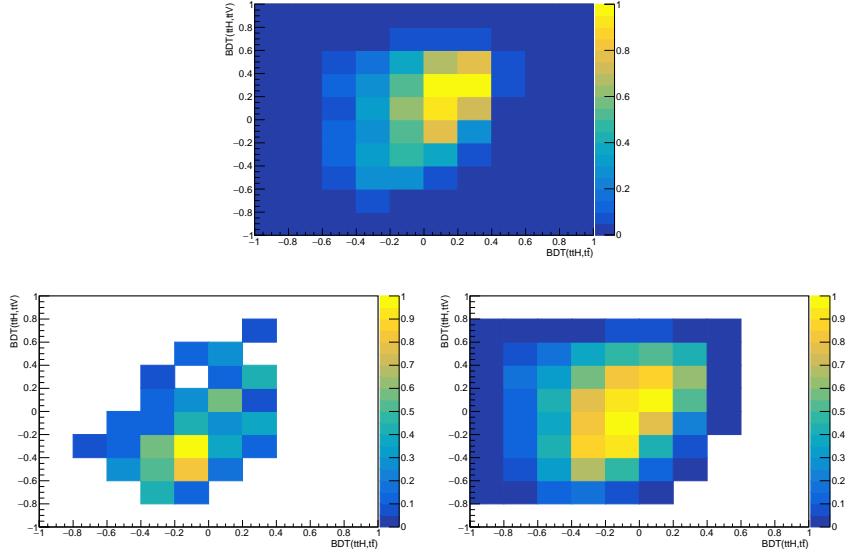


Figure 36: Two-dimensional distribution of the discriminators for the $t\bar{t}H$ signal (top), and for the $t\bar{t}$ (bottom left) and $t\bar{t}V$ (bottom right) backgrounds, in the three lepton channel, estimated from MC.

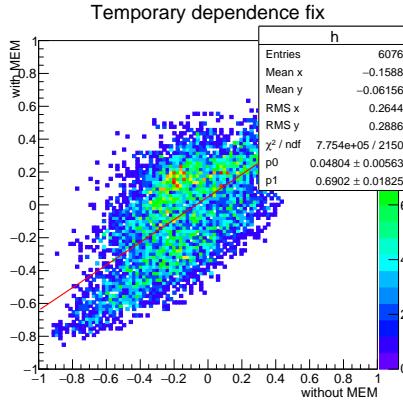


Figure 37: The discriminator of the $BDT(t\bar{t}H, t\bar{t}V)$ classifier for simulated $t\bar{t}V$ events, when the training is performed with or without the MEM discriminator among the input variables. A χ^2 fit is performed, assuming linear dependence.

583 for $t\bar{t}$ events, the conversion function will be removed.

584 The plane, populated with training events, is binned finely, depending on the available statis-
 585 tics: the binning is finer (20×20) for the two same sign leptons final state, and coarser (10×10)
 586 for the three leptons final state. For each bin, the likelihood ratio between signal and back-
 587 ground is computed.

588 In order to obtain an optimal one-dimensional classifier starting from the two-dimensional
 589 likelihood ratio distribution, an ordering criterion for the bins is devised. First the likelihood
 590 distribution for background events is obtained by assigning to each background event the like-
 591 lihood ratio corresponding to the bin the event belongs to. The fine granularity of the binning
 592 ensures that this assignment is a good approximation of the true underlying likelihood func-
 593 tion. From the distribution of the likelihood ratio for background events, the corresponding
 594 cumulative distribution is computed. A choice for the final number of bins is performed, by
 595 taking into account the available statistics in the training samples, and the quantiles of the

596 cumulative likelihood ratio distribution for background distribution are used to divide it into
 597 the target number of final bins. The cumulative likelihood ratio distribution for background
 598 events, and its partition in quantiles, are shown in Fig. 38 for the two same sign leptons (left)
 599 and three leptons (right) final states respectively.

600 The map between the initial and the final binning is shown in the left part of Fig. 39 and 40 for
 601 the two same sign leptons and three leptons final states respectively. Such construction guar-
 602 antees that the final bins are populated by background events in an uniform way. The resulting
 603 one-dimensional distributions, obtained using the testing events, are shown in the right part
 604 of Fig. 39 and 40 for the two same sign leptons and three leptons final states respectively. The
 605 population of the background the testing distribution is not perfectly uniform in the testing
 606 sample because of statistical fluctuation of the testing and training samples.

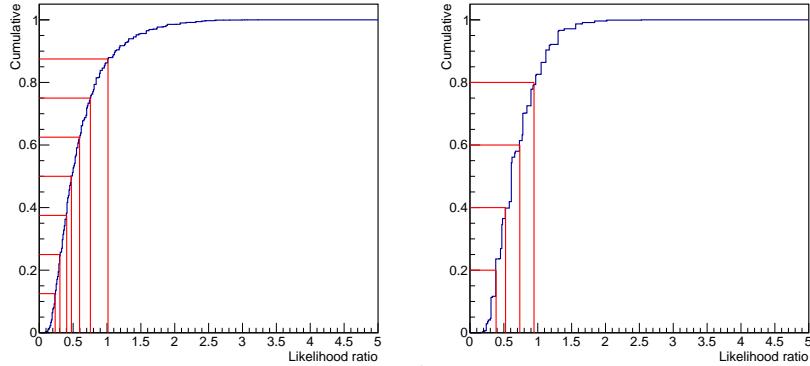


Figure 38: Cumulative likelihood ratio distribution for background events for the two same sign leptons final state (left) and the three leptons final state (right), estimated from MC.

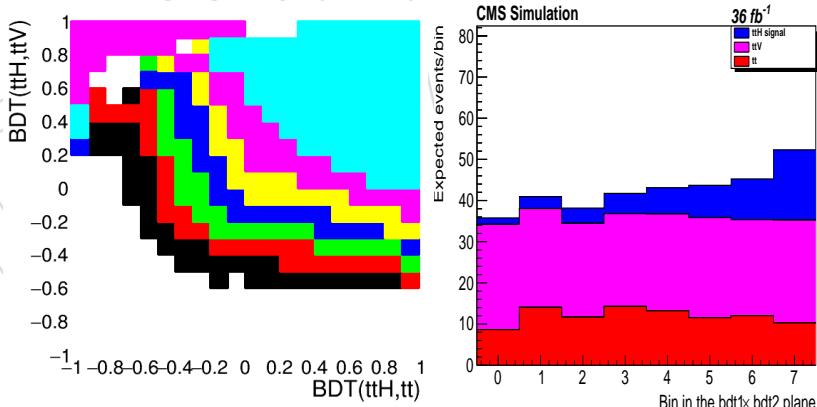


Figure 39: The location of the different bins in the two-dimensional plane (left), as well as the number of expected events in each bin (right) in the two same sign leptons channel, estimated from MC.

607 To check that the choice number of final bins in each final state is indeed suitable, an alternative
 608 method that does not require any assumption on the final number of bins is used, based on the
 609 k -means algorithm [27, 28] is used, yielding similar results and thus confirming the soundness
 610 of the choice of number of final bins. Such method is detailed in Appendix D.

611 Figure 41 shows the event yield as a function of the bins defined above.

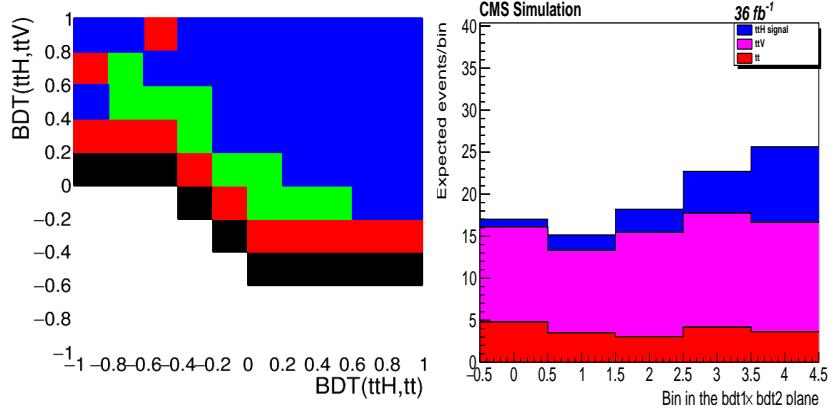


Figure 40: The location of the different bins in the two-dimensional plane (left), as well as the number of expected events in each bin (right) in the three leptons channel, estimated from MC.

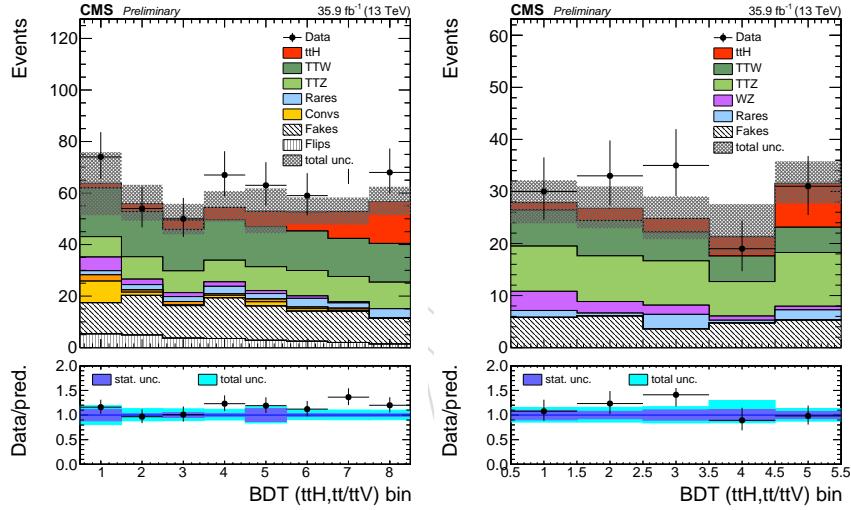


Figure 41: Binned distributions of the pair of discriminators in 2lss (left) and 3l (right) channels.

5.4 Event subcategories

Events are further split into lepton flavours: two electrons, two muons and electron and muon.

These three categories, except the two electrons, are further divided according to the presence

(or absence) of two medium tagged b-jets, the b-tight (b-loose) categories.

The events with at least three leptons are only separated into the b-tight and b-loose categories.

Finally, to exploit the charge asymmetry present in several backgrounds (ttW, WZ, single top

and W+jets), but not present in ttH, events in each of the categories described above are further

categorized by the positive or negative sum of the lepton charges. The summary of all event

categories is summarized in Figure 42.

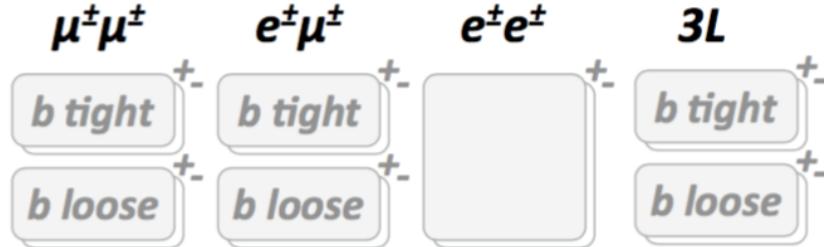


Figure 42: Diagram of all event categories in the analysis. Categories are based on lepton multiplicity and flavor, b-jet composition, and the sign of the sum of the lepton charges.

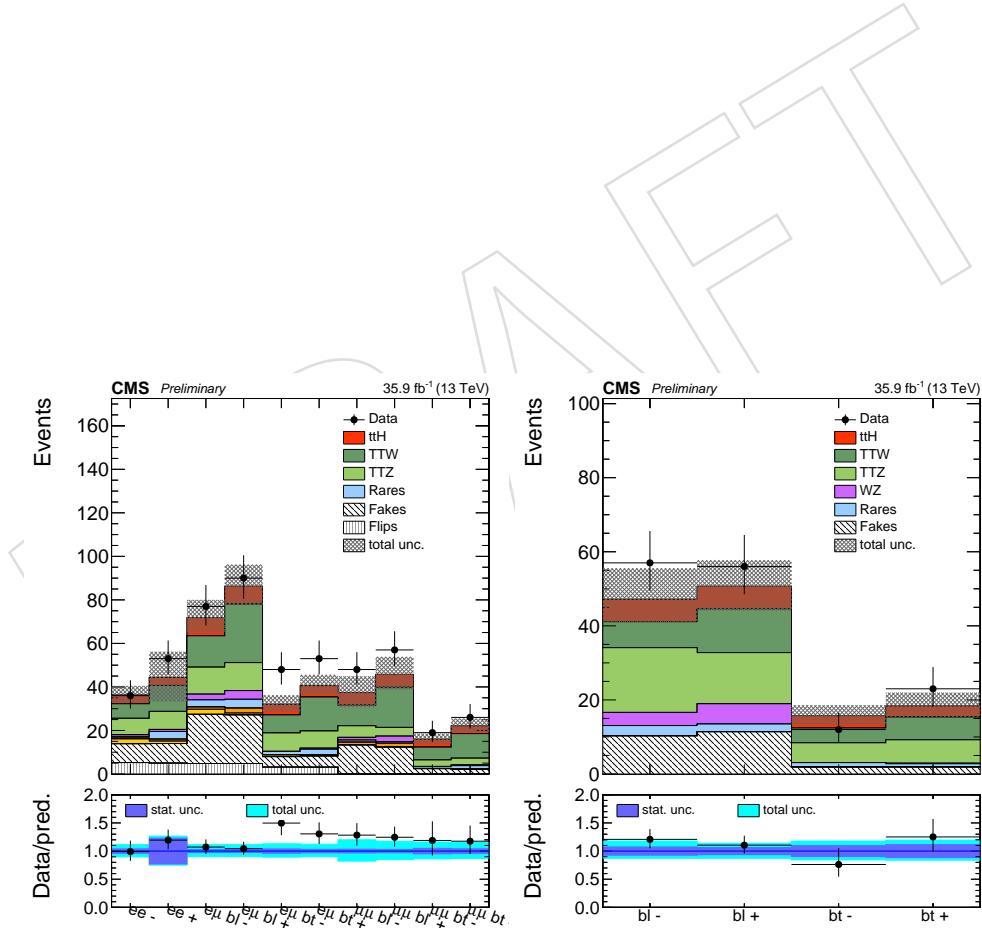


Figure 43: Splitting in categories for the 2lss (left) and 3l (right) channels.

621 6 Signal modeling

622 The signal is modelled using simulated events. In this version of the note, we present the
 623 same BDT outputs used in the 2015 analysis. The generator-level information is identical to the
 624 2015 samples, thus little change is expected. The plots in this section will be updated to reflect
 625 BDT output with 2016 samples in later versions of this note. The simulation has two different
 626 sources of systematic uncertainty. The first source of uncertainty is correction factors applied
 627 to the simulation in order to better reproduce the detector conditions and performance in data.
 628 The second source is assumptions made in the theoretical models that were used to produce
 629 the simulation. We account for uncertainties from both sources.

630 6.1 Correction factors and experimental uncertainties

631 As discussed in Section 3, we use scale factors to correct for differences in lepton performance
 632 between data and simulation. The scale factors account for the differences in the trigger, lepton
 633 Loose and Tight selections. Each of these scale factors has an uncertainty associated with it as
 634 discussed in that section, and it's propagated in the final uncertainties on signal yields.

635

636 In the RunI analysis we found that the uncertainty from the jet energy resolution plays a negligi-
 637 ble role in this analysis.

638 The uncertainties on the correction for the data/sim differences in the b-tagging performance
 639 described in 3 are parameterised as a function of p_T , η , and jet flavor. We assess their effect
 640 on the analysis by shifting the weight of each jet up and down by $\pm 1\sigma$ of the appropriate
 641 uncertainty and recalculating the overall event weight.

642 6.2 Theoretical uncertainties

643 We are completing the update of the impact of theoretical uncertainties on the signal prediction,
 644 it will be included in the next update of the AN.

645 The theoretical uncertainties on the NLO prediction for the inclusive $t\bar{t}H$ production cross sec-
 646 tion amount to $+5.8\text{--}9.2\%$ from unknown higher orders in the perturbative series and 3.6%
 647 from the knowledge of the parton distribution functions (PDFs) and α_s [29]. These uncertain-
 648 ties are propagated to the final normalization of the signal yields.

649 In addition to the overall normalisation, systematic uncertainties of theoretical origin on the
 650 distribution of the events in the final discriminating variables are considered, estimated con-
 651 ventionally by varying the normalisation and factorisation scales up and down by a factor of
 652 two and matching threshold between matrix element and parton shower (Fig. 44). In the cur-
 653 rent version of the results the shape uncertainties on the BDTs output are of the order of 2% to
 654 3%.

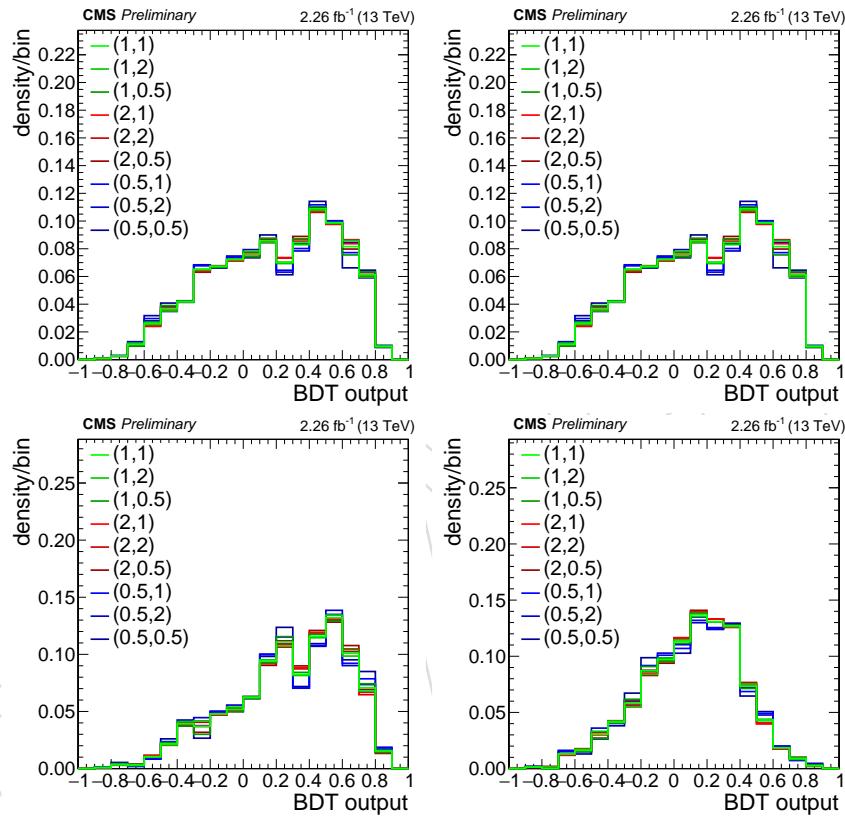


Figure 44: The BDT output distribution of the ttH signal, shown for the training against ttbar (left) and ttV (right) in the two same sign leptons (upper) and three lepton (lower) final state, with variations of the renormalization and factorization scale included in order to estimate the shape uncertainties.

655 7 Background predictions

656 7.1 Irreducible backgrounds

657 We are completing the update of the impact of theoretical uncertainties on the irreducible back-
 658 ground predictions, it will be included in the next update of the AN.

659 Irreducible backgrounds from $t\bar{t}W$, $t\bar{t}Z$ and $t\bar{t}WW$, are estimated from simulated events. The
 660 2015 BDT outputs are used here as for the signal modelling, and will be updated in later ver-
 661 sions of this note. Just like for the signal, corrections are applied for the different performance
 662 the individual physics objects between data and simulation measured in control regions in
 663 data. The effect of the JEC uncertainties on the final discriminator shapes is shown for $t\bar{t}W$, $t\bar{t}Z$
 664 in Fig. 45 and Fig. 46.

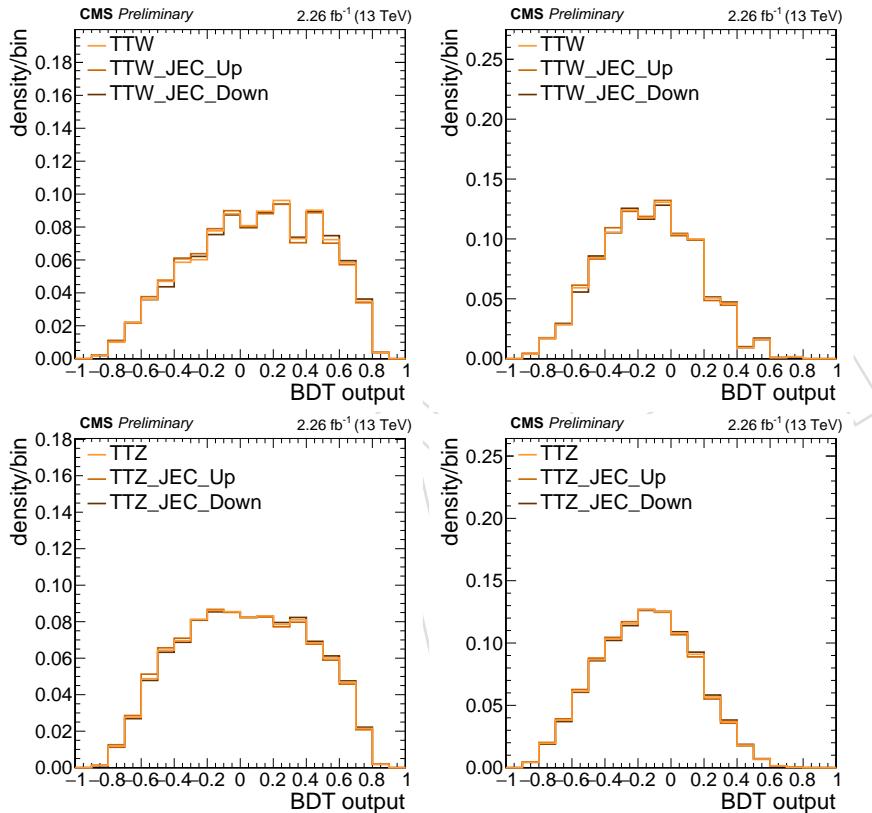


Figure 45: The BDT output distribution of the $t\bar{t}W$ and $t\bar{t}Z$, shown for the training against $t\bar{t}W$ (left) and $t\bar{t}Z$ (right) in the two same sign leptons final state, with the jet energy scale variations of one standard deviation included in order to estimate the shape uncertainties.

665 In the RunI analysis, the contribution of the $t\bar{t}WW$ process was found to be at least an order of
 666 magnitude smaller than $t\bar{t}W$ and $t\bar{t}Z$. At 13 TeV, NLO (arXiv 1405.0301 table 6), the cross section
 667 is a factor of 5 higher than at 8 TeV (partially because of the k factor as at 8 TeV we only had
 668 LO, and 13 TeV NLO/LO 1.5). Therefore if it was negligible at 8 TeV it still is.

669 The inclusive production cross sections for the $t\bar{t}W$ and $t\bar{t}Z$ processes are taken from the latest
 670 NLO computation, with theoretical uncertainties from unknown higher orders of 12% and 10%
 671 respectively, and uncertainties from the knowledge of the parton density functions and α_s of
 672 2% and 3% respectively [29].

673 In addition to the overall normalisation, systematic uncertainties of theoretical origin on the
 674 distribution of the events in the final discriminating variables are considered, estimated con-

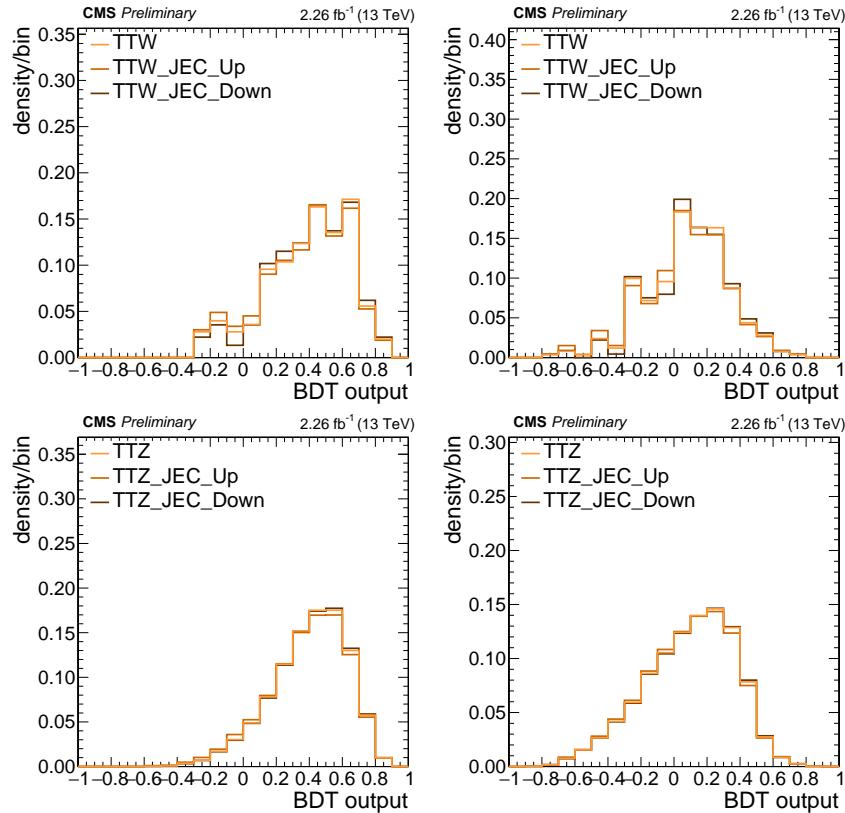


Figure 46: The BDT output distribution of the ttW and ttZ, shown for the training against ttbar (left) and ttV (right) in the three lepton final state, with the jet energy scale variations of one standard deviation included in order to estimate the shape uncertainties.

675 ventionally by varying the normalisation and factorisation scales up and down by a factor of
 676 two and matching threshold between matrix element and parton shower. These results are
 677 shown both for ttW and ttZ in Fig. 47 and 48. In the current version of the results the shape
 678 uncertainties on the BDTs output are of the order of 2% to 4%.

679 The cross section for the $t\bar{t}\gamma^*$ process with $\gamma^* \rightarrow \ell^+\ell^-$ process becomes large for decreasing
 680 virtuality of the γ^* , i.e. for small invariant masses of the dilepton pair. While in the analysis
 681 we reject events with low mass dileptons, the $t\bar{t}\gamma^*$ process can still contribute as a background
 682 when one of the two leptons is not reconstructed; this in particular can happen in kinematic
 683 configurations where the conversion is very asymmetric and one of the two leptons has trans-
 684 verse momentum below the acceptance.

685 Since the nominal ttZ MC sample is generated with the requirement $m_{\ell^+\ell^-} > 10 \text{ GeV}$, to es-
 686 timate this background we rely on an additional $t\bar{t}\gamma^*$ MC sample generated in the remaining
 687 part of the phase space. This additional sample is generated with LO MADGRAPH, and the
 688 details of the generation and normalization can be find here [30].

689 In the case of electrons, in addition to the $t\bar{t}\gamma^*$ background there is a similar topology of events
 690 from $t\bar{t}\gamma$ production where the photon converts early in the detector material, one conversion
 691 electron is not reconstructed and the remaining can then be misidentified as prompt electron¹.
 692 This background, despite being reducible, is not covered by the reducible background esti-
 693 mation obtained extrapolating from leptons failing the MVA requirement, described later in
 694 section 7.4, since the electron arising from the converted photon will be isolated, unlike non-

¹If both electrons are reconstructed, then the conversion veto applied in the electron selection will reject both.

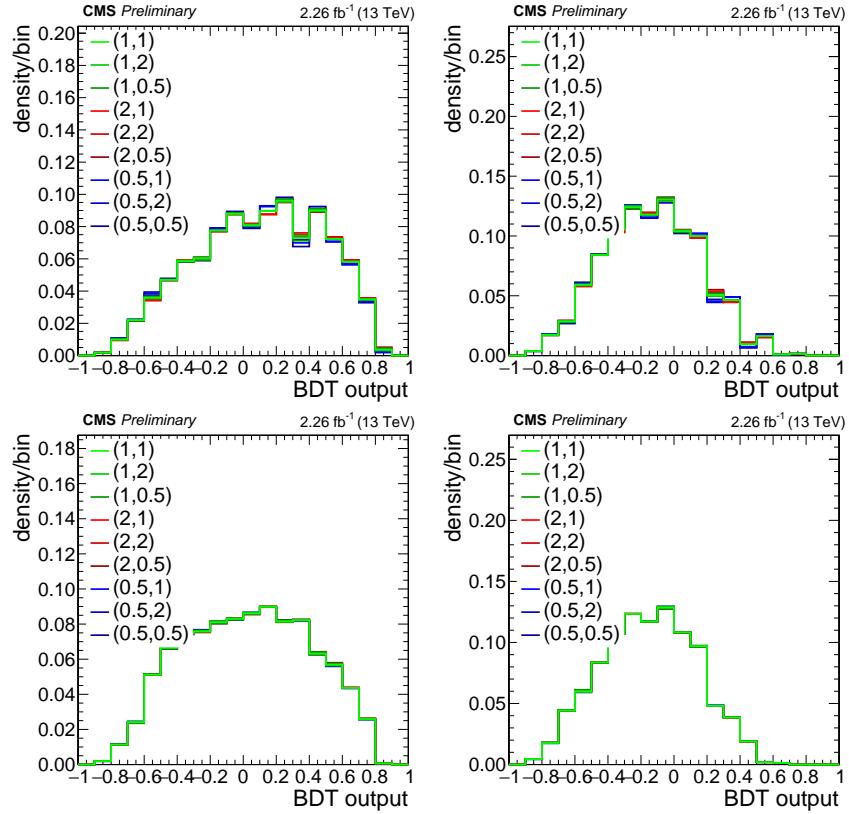


Figure 47: The BDT output distribution of the $t\bar{t}W$ and $t\bar{t}Z$, shown for the training against $t\bar{t}\bar{b}$ (left) and $t\bar{t}V$ (right) in the two same sign leptons final state, with variations of the renormalization and factorization scale included in order to estimate the shape uncertainties.

695 prompt electrons from hadron decays or misidentified charged hadrons. We therefore rely on
 696 simulations normalized to NLO QCD cross section from Madgraph5_aMC@NLO.

697

698 In addition to the experimental uncertainties, we assign a systematic uncertainty of 30% to the
 699 overall normalization of $t\bar{t}\gamma$ and a systematic uncertainty of 50% on the overall normalization
 700 of $t\bar{t}\gamma^*$.

701 7.2 Di-bosons backgrounds

702 WZ and ZZ production with the gauge boson decaying to electrons, muons or taus can yield
 703 the same leptonic final states as the signal, if considering also events where not all leptons
 704 are identified. While the ZZ background is greatly reduced by the cut on MET LD, the WZ
 705 background remains an important contribution to the three and more leptons signal region.

706 When not requiring additional hadronic jets in the final states, these processes are predicted
 707 theoretically at NLO accuracy, and the inclusive cross sections have been successfully mea-
 708 sured at the LHC. However these good agreement does not translate automatically to the signal
 709 regions used in this search, which always require the presence of at least one b-tagged jet.

710 Since dibosons are preferentially produced in association with jets from light quarks or glu-
 711 ons, it is possible to isolate a clean control region of WZ plus hadronic jets by inverting the
 712 b-tagging requirements of the signal region and also inverting the $Z \rightarrow ll$ veto. The approach
 713 chosen for estimating the background is therefore to use simulated events but normalizing the
 714 overall event yields in control regions of WZ plus two not b-tagged jets. This reduces the sys-

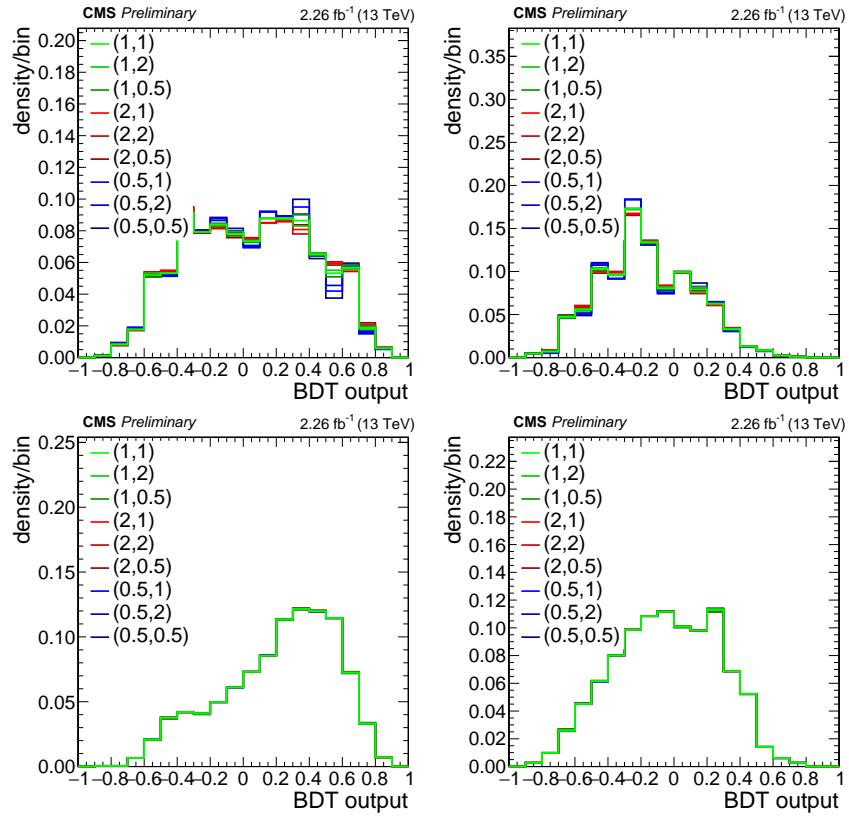


Figure 48: The BDT output distribution of the ttW and ttZ, shown for the training against ttbar (left) and ttV (right) in the three lepton final state, with the variations of the renormalization and factorization scale included in order to estimate the shape uncertainties.

715 systematic uncertainty on the prediction, since the theoretical uncertainty on the ratio of event
 716 yields between signal region and control region is much smaller than the uncertainty on the
 717 production cross section of diboson plus multijet. The majority of events from this background
 718 in the signal region contain jets from gluons or light quarks mistagged as b-jets, for which the
 719 extrapolation is affected only by uncertainties of experimental origin.

720 7.2.1 Measurement in data from events with no b-jets

721 This part is being updated with the full 2016 data. However, the plots in the $WZ \rightarrow 3l$ control
 722 region in Sec. A.3 show already a scale factor compatible with unity between data and simula-
 723 tion.

724 The extraction of the WZ yield in the control region is performed via a one dimensional neg-
 725 ative log likelihood fit of the shape of transverse mass of the lepton not associated to the Z
 726 boson. The shape and normalization of the residual backgrounds are fixed to the expectations
 727 from simulations. The measurement has been performed on 6.3 fb^{-1} data collected in 2016,
 728 and yields a scale factor of 1.2 ± 0.1 (stat.). Figures 49, 50, and 51 show the good agreement
 729 observed in the WZ control region for the following distributions:

- 730 • $m_T W(l)$
- 731 • E_T^{miss}
- 732 • selected jet multiplicity
- 733 • sum of lepton charges

- 734 • reconstructed Z invariant mass

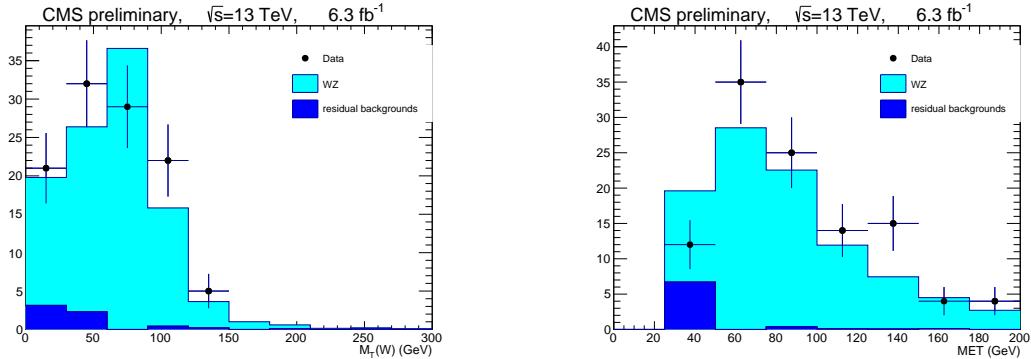


Figure 49: Distribution of the transverse mass of the lepton not associated to the reconstructed Z boson, $m_T(l)$, (left) and transverse missing energy, met , (right) after a fit of the WZ and background processes to the data.

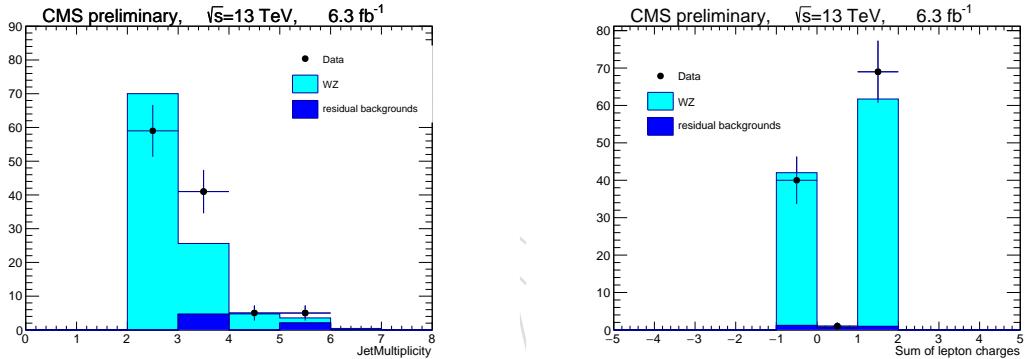


Figure 50: Distribution of the jet multiplicity (left) and sum of the leptons charges (right) after a fit of the WZ and background processes to the data

735 **7.2.2 Extrapolation to events with b-jets**

736 The ratio of the events yields between signal region and control region is measured in the
 737 simulation. The main systematics are expected to come from the b-tagging scale factors since
 738 most of the WZ plus two b-jets events are due to mistags:

$$\text{SR-b-loose/CR} : 0.0371 \pm 0.0041 \text{ (b-tagging)} \pm 0.0028 \text{ (theory)}$$

$$\text{SR-b-tight/CR} : 0.0015 \pm 0.0006 \text{ (b-tagging)} \pm 0.0001 \text{ (theory)}$$

739 Theoretical uncertainties arise from the modelling of the heavy flavour content of the jets in
 740 diboson plus multijet events. The expected flavour composition for WZ events passing the
 741 b-loose (resp. b-tight) jet selection is approximately 35% (13%) of events with mistagged jets
 742 from gluons or u,d,s quarks, 47% (50%) of events with a jet from a charm quark or antiquark, and
 743 the remaining fraction from events with at least one bottom quark or antiquark. Uncertainties
 744 on the extrapolation arising from the parton distribution functions are estimated by simulated
 745 reweighting of the events to different PDGs sets and all their associated eigenvectors or replicas.

746 The overall uncertainty on the normalization of the WZ background is composed by the sta-
 747 tistical uncertainty in the control region, from the residual backgrounds in the control region,

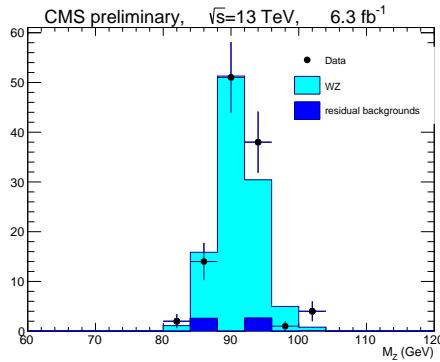


Figure 51: Distribution of the Z candidate invariant mass.

748 from the uncertainties on the b-tagging efficiencies, from the parton distribution functions and
 749 from the theoretical uncertainties on the extrapolation (dominated by the uncertainty on the
 750 flavor composition of the final state due to higher-order QCD terms).

751 7.3 Charge misassignment background

752 The background from processes with prompt opposite-sign lepton pairs like $t\bar{t}$ or DY+jets,
 753 where one of the two leptons has a wrongly assigned charge, is estimated from the measured
 754 charge misassignment probabilities and the events of a corresponding opposite-sign control re-
 755 gion. Naturally, this background is only relevant for the same-sign dilepton channels. Studies
 756 in MC show the charge misassignment probability for muons to be negligible, and we subse-
 757 quently restrict ourselves to electrons.

758 7.3.1 Measurement of the electron charge misassignment probabilities

759 The charge misassignment probability for electrons can be extracted from the data, in events
 760 with two same-sign electrons with invariant mass close to the mass of the Z boson. Electron
 761 pairs in the peak are sure to be from real opposite sign pairs with a wrongly assigned charge
 762 on one leg.

763 Charge misassignment probabilities are calculated for different bins of electron p_T and η by
 764 extracting same-sign and opposite-sign event yields categorized in the kinematics of the two
 765 lepton legs. In each category, the event yield of electron pairs from Z decays is determined
 766 from a fit to the invariant mass shape, and depends on the charge misassignment probabilities
 767 of each leg. The invariant mass shape is modeled with a crystal ball and Breit-Wigner function
 768 for the signal and an exponentially falling function for the backgrounds.

769 Electron kinematics are separated in three p_T (10–25 GeV, 25–50 GeV, and \geq 50 GeV) and two
 770 η bins (0–1.479 and 1.479–2.5), resulting in a total of 21 distinct categories of electron pairs.
 771 The six charge misassignment probabilities are then determined in a simultaneous fit to the 21
 772 same-sign and opposite-sign event yields.

773 The resulting misassignment probabilities range between about 0.03% in the barrel and about
 774 0.4% in the end caps and are shown in Tab. 8 and Fig. 52.

775 7.3.2 Background estimation

776 Contributions from opposite-sign prompt leptons with charge-misassigned electrons to the
 777 same-sign dilepton channels with electrons (ee, and e μ) are then estimated from the events
 778 of a control region with identical selection except for the requirement of equal charge of the

Table 8: Electron charge misassignment probabilities (in percent) as determined in data (top) and Drell–Yan MC (bottom).

Data	$10 \leq p_T < 25 \text{ GeV}$	$25 \leq p_T < 50 \text{ GeV}$	$50 \text{ GeV} \leq p_T$
$0 \leq \eta < 1.479$	0.0442 ± 0.0011	0.0179 ± 0.0004	0.0262 ± 0.0020
$1.479 \leq \eta < 2.5$	0.1329 ± 0.0066	0.1898 ± 0.0014	0.3067 ± 0.0113
MC			
$0 \leq \eta < 1.479$	0.0378 ± 0.0016	0.0222 ± 0.0003	0.0233 ± 0.0015
$1.479 \leq \eta < 2.5$	0.0956 ± 0.0044	0.2108 ± 0.0027	0.3157 ± 0.0018

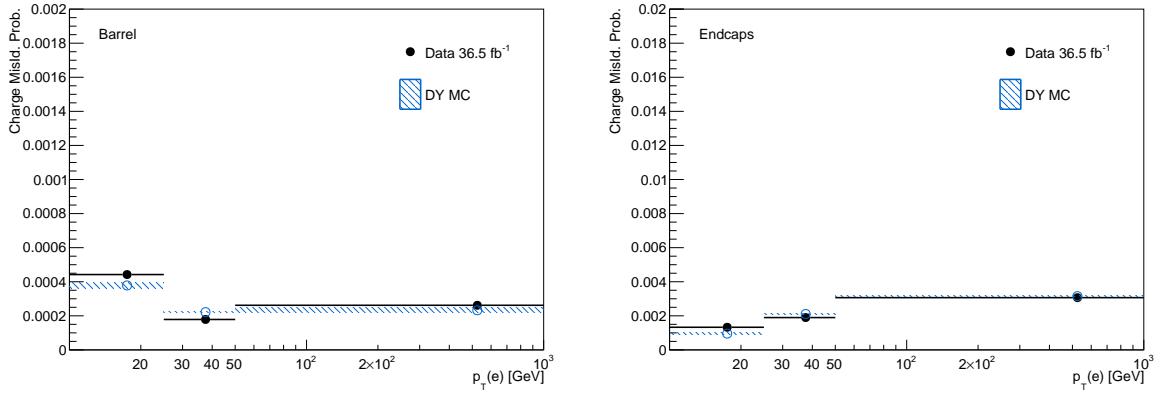


Figure 52: Electron charge misassignment probabilities as a function of p_T for electrons in the barrel (left) and endcaps (right). Note the change in y-axis scale.

779 lepton pair. Each event in the control region is assigned a weight of $P(p_T, \eta)$ for each electron
 780 with a given p_T and η in the event (i.e. ee events get a weight of $P_1 + P_2$ and $e\mu$ events get a
 781 weight of P), where P is the measured charge misassignment probability.

782 The procedure is tested in two control regions: once using the same events that were used
 783 to measure the probabilities, dominated by DY events, and once in a selection with at least
 784 one medium b-tagged jet or two loose b-tagged jet and between 2 and 3 hadronic jets, with a
 785 significant contribution from $t\bar{t}$ events. Event distributions in the two control regions where the
 786 background from charge misassigned electrons is estimated as described are shown in Fig. 53
 787 and 54.

788 From the statistical uncertainty of the measured probabilities and the good agreement of pre-
 789 dicted charge-flip yields with the observed data distributions in the control regions, we assign
 790 a generous 30% uncertainty on the predicted event yields from this background.

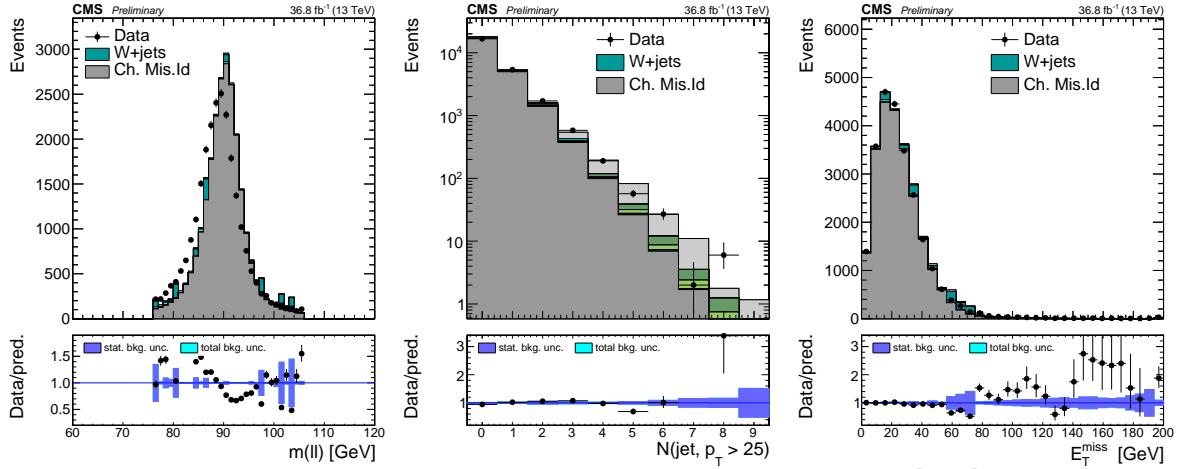


Figure 53: Charge misassignment closure test in DY dominated control region (where the misassignment probabilities are extracted from).

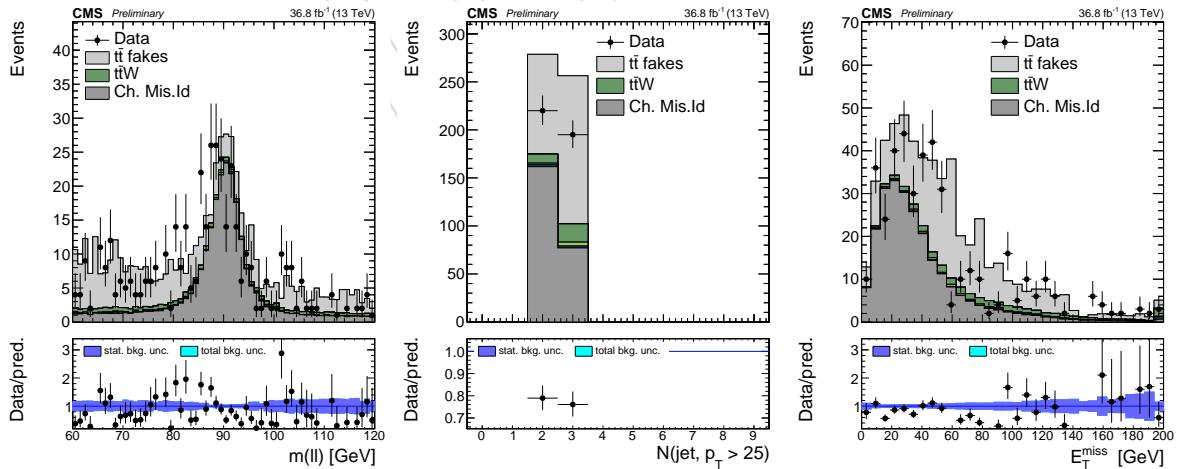


Figure 54: Charge misassignment closure test in a selection of exactly two same-sign electrons passing the full selection, between 2 and 3 hadronic jets, and at least one medium b-tagged jet or two loose b-tagged jets.

7.4 Fake lepton background

7.4.1 Fakeable object definition tuning

793 The main control region we use to measure the fake rate is enriched in QCD jet events, and
794 is obtained selecting events with one loose lepton and an hadronic jet well separated from the
795 lepton ($\Delta R > 0.7$).

796 For all muons and electrons with p_T above 30 GeV, events are selected at trigger level requiring
797 a prescaled single lepton trigger (with no isolation requirements) and a particle flow jet with p_T
798 30 GeV reconstructed at HLT (40 GeV for low- p_T muons). This implies that the fakeable objects
799 in the measurement region all pass the lepton trigger, while in the application region this is not
800 necessarily the case, as we include also events triggered by a single or double lepton trigger.
801 In order to avoid a bias in the background estimate, we therefore need to have a definition of
802 fakeable object so that the fake rate does not depend on whether the lepton passes the trigger
803 selection or not.

804 We assess this using non-prompt leptons in simulated lepton-enriched QCD events, where we
805 can compare the fake rate with and without requiring the lepton to pass the trigger, for dif-
806 ferent choices of the fakeable object selection. The study reveals two important features: first,
807 the trigger turn-on imposes a cut on the reconstructed lepton p_T , and so it limits the amount
808 of sideband available in the extrapolation for given bin in corrected lepton p_T . Because of
809 this, a trigger with a given threshold, can only be used to measure the fake rate for signifi-
810 cantly larger corrected p_T value, if we do not want the trigger threshold to bias the result. For
811 the definition used in this version of the analysis, the $p_T^{corr}/p_T \leq 0.9/0.5 = 1.8$, so e.g. the
812 HLT_Mu8 trigger can be used starting from a corrected p_T of 14.4 GeV. Because of these ef-
813 fects, we measure the fake rate for muons using a combination of the HLT_Mu3_PFJet40 trigger
814 (for corrected $p_T \in [10, 30]$ GeV), the HLT_Mu8 trigger (for corrected $p_T \in [15, 45]$ GeV), an the
815 HLT_Mu17 trigger (for corrected p_T above 30 GeV) and the HLT_Mu27 trigger (for corrected p_T
816 above 45 GeV, after checking on the MC that there is no visible bias even if the trigger is fully
817 efficieny only at corrected p_T of about 48 GeV); in the regions covered by multiple triggers, we
818 combine them since the prescales are large and uncorrelated and thus the datasets are indepen-
819 dent. For electrons, we use HLT_EleX_CaloIdM_TrackIdM_PFJet30 triggers, with thresholds of
820 8 GeV (for $p_T \in [15, 20]$ GeV), 12 GeV (for $p_T \in [20, 30]$ GeV), and 17 GeV (for $p_T > 30$ GeV)

821 Beyond the impact of the p_T threshold, the trigger can also bias the fake rate if it has require-
822 ments in identification or isolation that are not strictly looser than the selection used at the
823 denominator. For muons, as long as we use triggers with no isolation requirements this is not
824 the case. In the case of electrons, the identification criteria applied at HLT, cut-based, are differ-
825 ent from those offline (mva-based), and so we would have a mismatch even for electrons well
826 above the p_T threshold of the trigger.

827 We solve this by tightening the electron fakeable object definition including in it also some cut-
828 based electron identification criteria. The selection criteria to emulate the trigger are chosen
829 by comparing electron identification variables for non-prompt electrons in QCD MC events for
830 electrons that pass or fail the trigger. The cuts have to be tight enough that for events passing
831 them the fake rate does not depend strongly on whether the event passes or fails the trigger.
832 However, cuts need to be not too tight less they start causing a loss of signal efficiency, because
833 they use the identification variables in a less optimal way than the mva electron identification.
834 For the criteria we use in this analysis, the loss of efficiency introduced by these cuts for signal
835 MC events that pass the full analysis selection and trigger is about 3% in both the dielectron
836 and electron-muon final states. In this version of the analysis, we apply these additional criteria

837 on electrons irrespectively of their p_T , while in older versions of the analysis the request was relaxed for corrected $p_T < 30$ GeV.

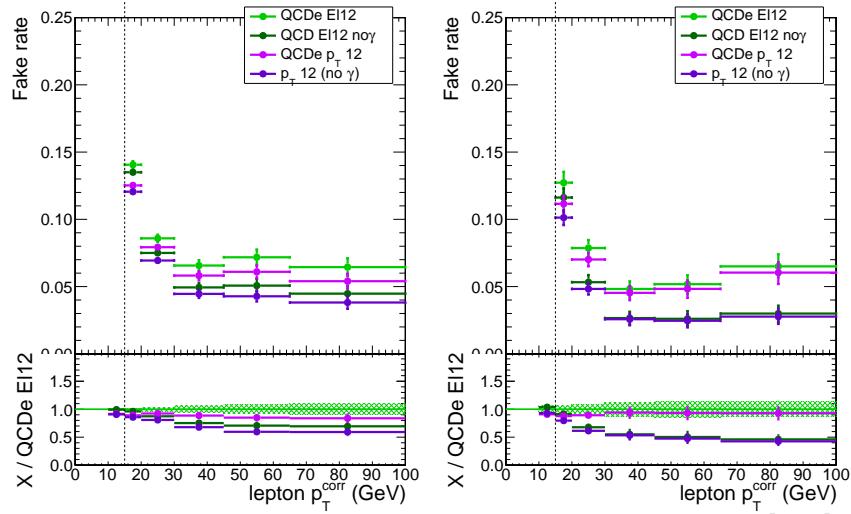


Figure 55: Fake rate for electrons in the barrel (left) and endcaps (right) from simulated QCD multi-jet events, before the trigger requirement (pink) and after (green). The lighter shades of pink and green are for all reconstructed electrons in the sample, while the darker shades are excluding the electrons that originate from the conversions of prompt photons (i.e. photons not hadron decays).

838

839 The choice of the lepton identification criteria on the fakeable object determine, together with
 840 the working point used for the tight lepton definition, the the fake rate for fake leptons, while
 841 they do not affect substantially the fake rate for non-prompt leptons originating from the decay
 842 of heavy flavour hadrons. On the other hand, the cut on the b-tagging discriminator of the jet
 843 associated to the lepton in the fakeable object definition can alter the fake rate for non-prompt
 844 leptons from heavy flavour without affecting fake leptons which are mostly originating from
 845 light jets. Thus, we can tune this cut to make the two fake rates more similar, and thus reduce
 846 the uncertainties associated to the flavour dependency of the fake rate. The quality of the
 847 tuning can be evaluated comparing e.g. the fake rates in $t\bar{t}$ events in the application region for
 848 the b-loose and b-tight category, that have different flavour compositions (Fig. 56).

849 7.4.2 QCD measurement region definition cuts

850 For a fixed choice of the fakeable object and numerator, we can then assess how the fake in QCD
 851 events agrees with that of $t\bar{t}$, and how it depends on the cuts on the tag jet used to select the
 852 events. A comparison of fake rates for background muons in $t\bar{t}$ and QCD is shown in Fig. 57,
 853 both inclusively and selecting only leptons from b-jets. Overall, QCD and $t\bar{t}$ are found to
 854 agree to better than 20%. For the central value of the measurement, for corrected $p_T < 30$ GeV
 855 we require one loose b-tagged jet away from the lepton, while we at higher p_T we drop this
 856 requirement.

857 A comparison of fake rates for background electrons in $t\bar{t}$ and QCD is shown in Fig. 58. The
 858 agreement is excellent in the barrel at low and moderate p_T , and in general within 20% every-
 859 where. Electrons from conversions are a sizeable contribution to the fake rate especially in the
 860 endcaps and at high p_T .

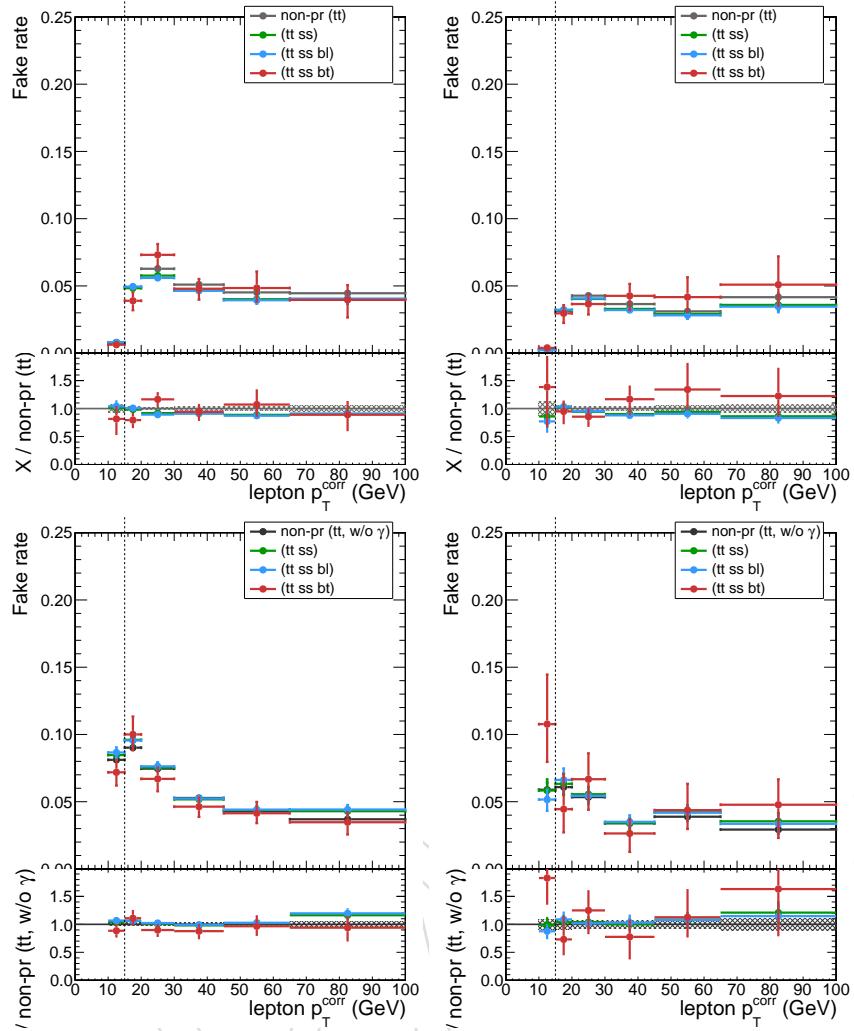


Figure 56: Fake rate for muons (top) and electrons (bottom) in the barrel (left) and endcaps (right), from simulated $t\bar{t}$ events, inclusively (gray), in same-sign events (green), in same-sign events in the b-loose category (blue), and in same sign-events in the b-tight category (red). Electrons originating from the conversion of prompt photons are excluded from the plots.

861 7.4.3 QCD measurement: prompt lepton contamination

862 An important challenge in measuring the fake rate in jet events in data is the contamination of
 863 prompt leptons, mostly from W and Z production in association with hadronic jets, but also
 864 from $t\bar{t}$. In order to suppress the Z contamination, events with more than one loose lepton are
 865 vetoed, leaving mostly events with one leptons outside the acceptance or from $Z \rightarrow \tau_\ell \tau_h$. A
 866 good discrimination between QCD events and W can be achieved from the transverse mass of
 867 the lepton and missing energy in the event, $M_T(\ell, E_T^{\text{miss}})$. The standard procedure is to apply a
 868 tight cut $M_T(\ell, E_T^{\text{miss}}) < 15$ GeV was applied, and the residual contamination was subtracted at
 869 numerator and denominator in each p_T bin using simulated $W/Z + \text{jets}$ events. The simulation
 870 was normalized to the data from a fit to $M_T(\ell, E_T^{\text{miss}})$, in the sample of events at the fake rate
 871 numerator (i.e. passing the tight requirements), before the cut at 15 GeV.

For this analysis, we implemented two improvements on that procedure. The first improvement is a change in the discriminating variable used: the traditional transverse mass

$$M_T(\ell, E_T^{\text{miss}}) = \sqrt{2p_{T\ell} E_T^{\text{miss}} (1 - \cos \Delta\phi)}$$

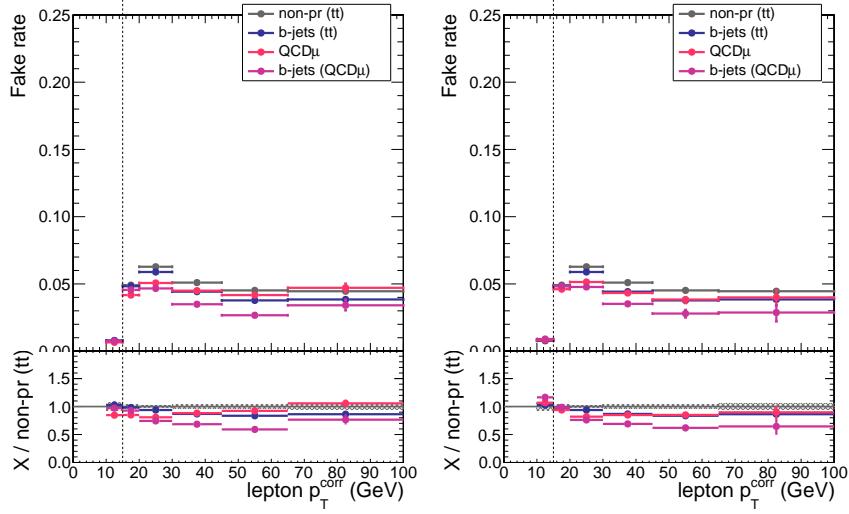


Figure 57: Fake rate for background muons in the barrel, from simulated $t\bar{t}$ and QCD events, inclusively or selecting only those from B hadron decays. The plot on the left is requiring only one away jet of $p_T > 30$ GeV, while the one on the right additionally require one loose b-tag.

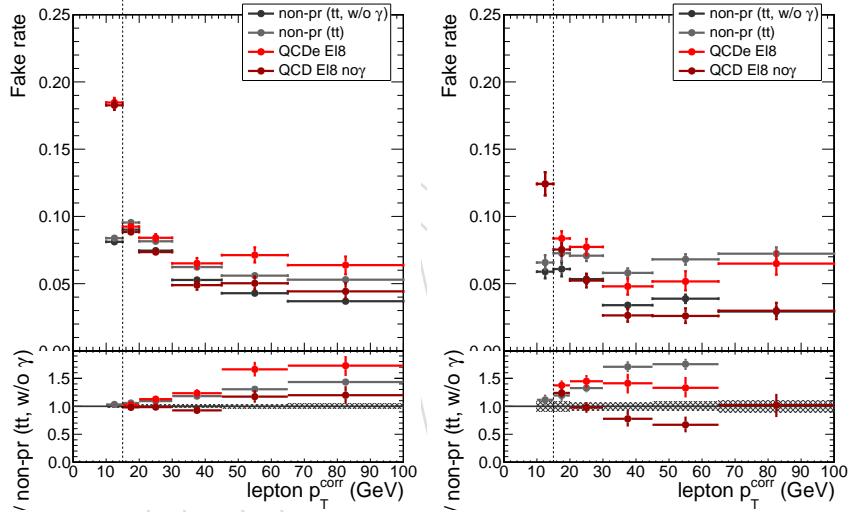


Figure 58: Fake rate for background electrons in the barrel (left) and endcaps (right), from simulated $t\bar{t}$ and QCD events, inclusively or excluding those from the conversions for prompt photons. Electrons in QCD events are required to pass the HLT_Ele8_CaloIdM_TrackIdM_PFJet30 trigger.

is obviously correlated with the lepton p_T , and so also with the lepton fake rate. To avoid this correlation, which can potentially introduce biases in the subtraction procedure, we define a new variable

$$M_T^{\text{fix}}(\ell, E_T^{\text{miss}}) := \sqrt{2p_{T\text{fix}} E_T^{\text{miss}} (1 - \cos \Delta\phi)}$$

replacing the lepton p_T with a fixed number (35 GeV), and thus relying only on the lepton direction. This variable still has a good discriminating power against $W + \text{jets}$ but is much less correlated with the lepton p_T and so with the fake rate.

The second improvement is the introduction of two alternative ways to implement the subtraction of the prompt contamination. The first alternative procedure is the one used in the run 1 analysis, and documented in detail in Section 7.4.2 of AN-13-159, except that we update it to use $M_T^{\text{fix}}(\ell, E_T^{\text{miss}})$ instead of E_T^{miss} . The procedure relies on two measurement of the fake rate

in data, one for small M_T^{fix} values and one for large and large M_T^{fix} values. Assuming the fake rate to be independent from M_T^{fix} , and taking from the simulation the ratio of $V + \text{jets}$ events expected in the two regions, it is possible to unfold the fake rate for QCD events from the two measurements:

$$f_{\text{QCD}} = \frac{f_S - r_{V+j}^{\text{SL}} f_L}{1 - r_{V+j}^{\text{SL}}} \quad \text{where} \quad r_{V+j}^{\text{SL}} = \left(\frac{N_{V+j}^S}{N_{V+j}^S} \right) / \left(\frac{N_{\text{data}}^S}{N_{\text{data}}^L} \right),$$

875 where f_i are the fake rates measured in data for small (S) and large (L) values of M_T^{fix} , N_{V+j}^i are
 876 the expected event yields from $V + \text{jets}$ and N_{data}^i are the observed events in data in the two
 877 regions at the denominator of the fake rate. This procedure can be performed separately in
 878 each bin of p_T , $|\eta|$.

In this version of the analysis, a small refinement of the procedure is done to account for the observed residual dependency of the fake rate on M_T^{fix} from MC: the formula is modified to

$$f_{\text{QCD}} = \frac{f_S - \mu r_{V+j}^{\text{SL}} f_L}{1 - \mu \gamma r_{V+j}^{\text{SL}} + \frac{N_{V+j}^S}{N_{V+j}^S + N_{\text{QCD}}^S} (\mu \gamma - 1)}$$

879 where $\mu = \epsilon_S^{V+j} / \epsilon_L^{V+j}$ and $\gamma = \epsilon_L^{\text{QCD}} / \epsilon_S^{\text{QCD}}$, and f_{QCD} is defined as ϵ_S^{QCD} since the bulk of
 880 the QCD events have low E_T^{miss} . In addition to the statistical uncertainties from data and
 881 MC, we assign a systematical uncertainty of 100% on the corrections (i.e on $\gamma - 1$, $\mu - 1$ and
 882 $N_{V+j}^S / (N_{V+j}^S + N_{\text{QCD}}^S)$) since they are derived from MC and not data.

883 A second alternative procedure relies on a simultaneous fit of the M_T^{fix} distribution for passing
 884 and failing probes, in a very similar way to the method used in the tag and probe method at
 885 the Z peak by fitting the invariant mass of the dilepton pair to extract efficiencies for the signal
 886 even in the presence of background. In our case, fit is done using templates from simulation
 887 for the QCD and $V + \text{jets}$ contributions. In addition to bin-by-bin statistical uncertainties on
 888 the templates, we include systematic shape uncertainties on the templates: we allow both a
 889 linear deformation of the template and a stretching of the template, as illustrated in Fig. 59.
 890 The shape systematics are assumed to be uncorrelated between QCD and $V + \text{jets}$, but totally
 891 correlated between passing and failing probes; the size of the deformation has been chosen to
 892 approximately cover the data to simulation differences observed across the various bins. The
 893 final uncertainty on the fake rate is obtained by profiling the likelihood of the simultaneous fit,
 894 and thus includes both the statistical and the systematical uncertainties.

895 Results of the measurement with all three subtraction methods are shown in Fig. 60. For the
 896 cut and subtraction method, the error bars include a systematical uncertainty of 5% on the sub-
 897 traction, which dominates over the statistical uncertainties at larger p_T . Within uncertainties,
 898 the three methods agree among themselves and also with the fake rate in MC. Since we do
 899 expect at least some correlation in the uncertainties of the three measurements, we opt for a
 900 conservative combination of the three by taking as uncertainty band the envelope of the three
 901 uncertainty bands and as central value the midpoint of the band.

902 7.4.4 Results

903 The final fake rates for electrons and muons are shown in Fig. 61. Overall, a good agreement
 904 between data and predictions from simulations is observed. Uncertainties are larger for high
 905 p_T leptons driven by the uncertainties on the subtraction of the prompt lepton contamination.

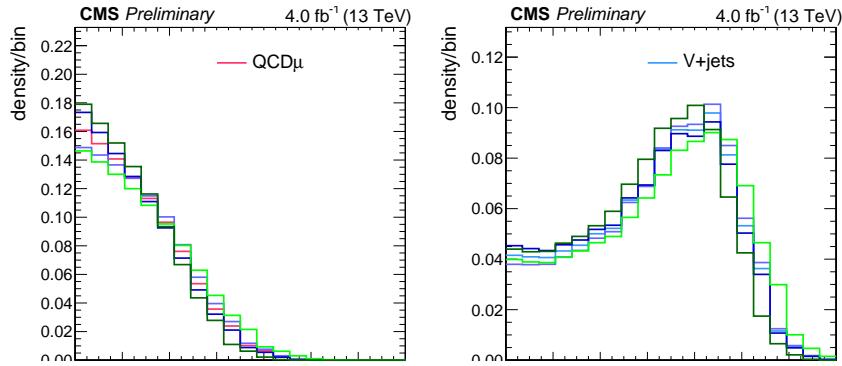


Figure 59: Shape uncertainties on the M_T^{fix} templates for muons in the barrel, in the bin of corrected p_T 20–30 GeV, for QCD events (left) and V + jets events (right)

906 For electrons, we estimate the fake rate for electrons not from prompt photon conversions by
 907 rescaling the measured fake rate in data by the ratio of the fake rates from QCD MC excluding
 908 and including electrons from conversions (from Fig. 58).

909 7.4.5 Fake rate application

910 The application of the fake rates follows the same principles already used for the analysis that
 911 was performed on the 8 TeV dataset.

912 In summary, a control region enriched in events with fake leptons, denoted as *application region*
 913 in the following, is selected by requiring that at least one of the selected leptons fails the tight
 914 lepton requirements.

915 The extrapolation from this control region to the signal region is performed by expressing the
 916 yields of events where k leptons pass the full selection and $n - k$ fail it in terms of the yields
 917 of events with n leptons among prompt and non-prompt ones, efficiencies and fake rates, for
 918 $n = 2$ or 3 and all possible values of k .

919
 920 In two lepton events ($n = 2$), the background contribution in the signal region can be expressed
 921 as:

$$N_{pp}^{\text{bkg}} = \frac{f_1}{1-f_1} N_{pf} + \frac{f_2}{1-f_2} N_{pf} - \frac{f_1 f_2}{(1-f_1)(1-f_2)} N_{ff}$$

922 under the approximation that the contribution of prompt leptons failing the selection can be
 923 neglected with respect to the contributions of non-prompt leptons. It is worth noting that the
 924 event yield observed in the signal region does not affect the background prediction.

925

926 Following the same logic, the background prediction in the three lepton category is obtained
 927 by weighting the events in the application region according to the following prescription:

- 928 • events with only one failing lepton are weighted by $f/(1-f)$, where f is the fake
 929 rate evaluated on the kinematic quantities of the failing lepton;

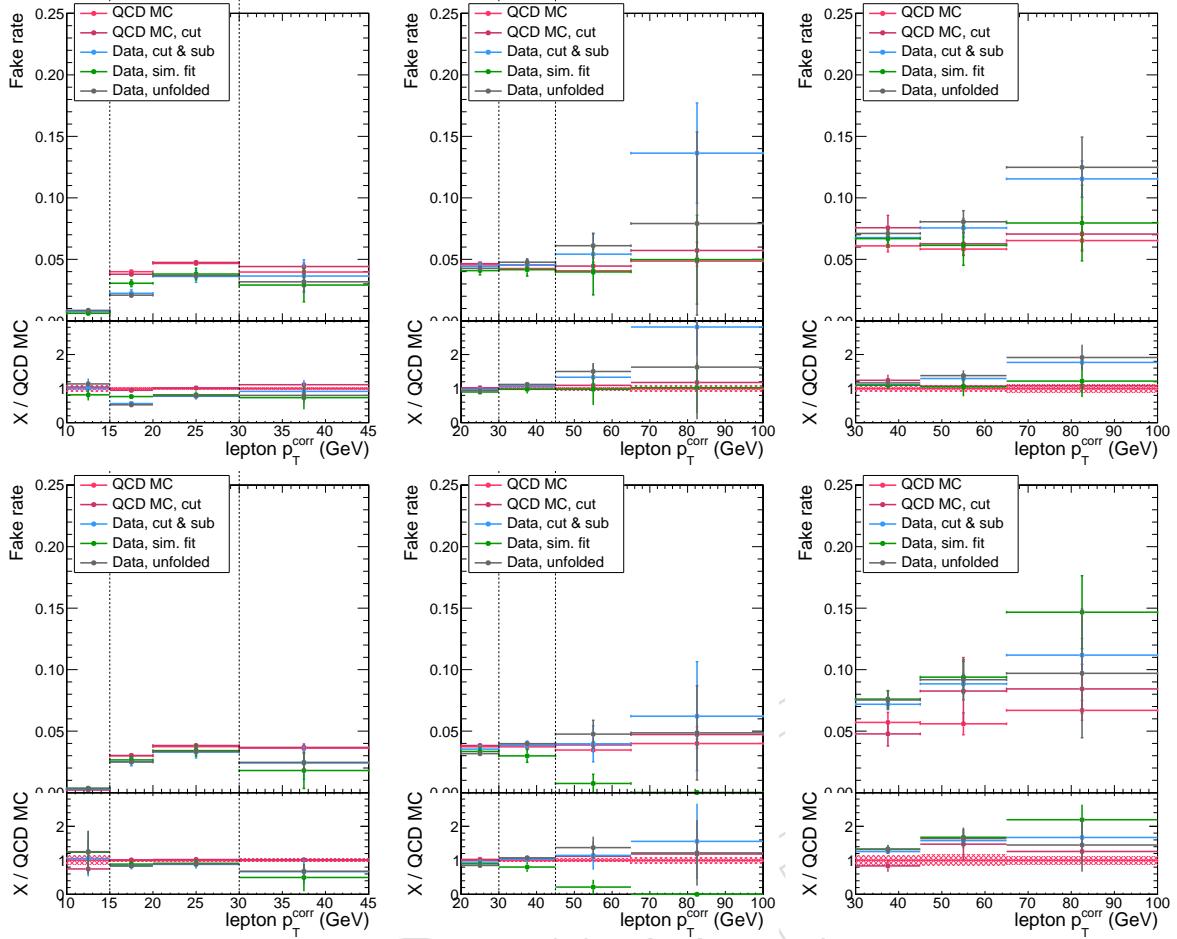


Figure 60: Fake rate measurement in data, for different subtraction methods, and compared with the predictions from simulation for non-prompt leptons from QCD MC. The top row is for the barrel, the bottom for the endcaps. The left column is for low p_T muons, the middle one for high p_T muons, and the right one for high p_T electrons.

- events with two failing leptons (i, j) are weighted by $-f_i f_j / ((1-f_i)(1-f_j))$;
- events with all three failing leptons are weighted by $f_1 f_2 f_3 / ((1-f_1)(1-f_2)(1-f_3))$.

A proof of the results presented above can be found in [31].

7.4.6 Closure tests and systematic uncertainties

Closure tests are performed on simulated events in order to confirm that the methods described in the present Section are well suited to predict the reducible background after the event selection requirements.

A fake rate extracted from non-prompt leptons in QCD MC, selected inclusively, is first applied to 2lss semi-leptonic $t\bar{t}$ simulated events where at least one of the two selected candidates fails the tight lepton requirements. Moreover, an alternative fake rate from non-prompt leptons in $t\bar{t}$ MC, selected inclusively as well, is applied to the same 2lss sample.

The test described above is performed in different event selections (separately for lepton flavors and analysis category). The difference in normalization between the two predictions, as well as their difference in the shape of the two kinematic discriminators against the $t\bar{t}$ and $t\bar{t}V$ backgrounds, are propagated as a systematic uncertainty to the fit used to extract the signal.

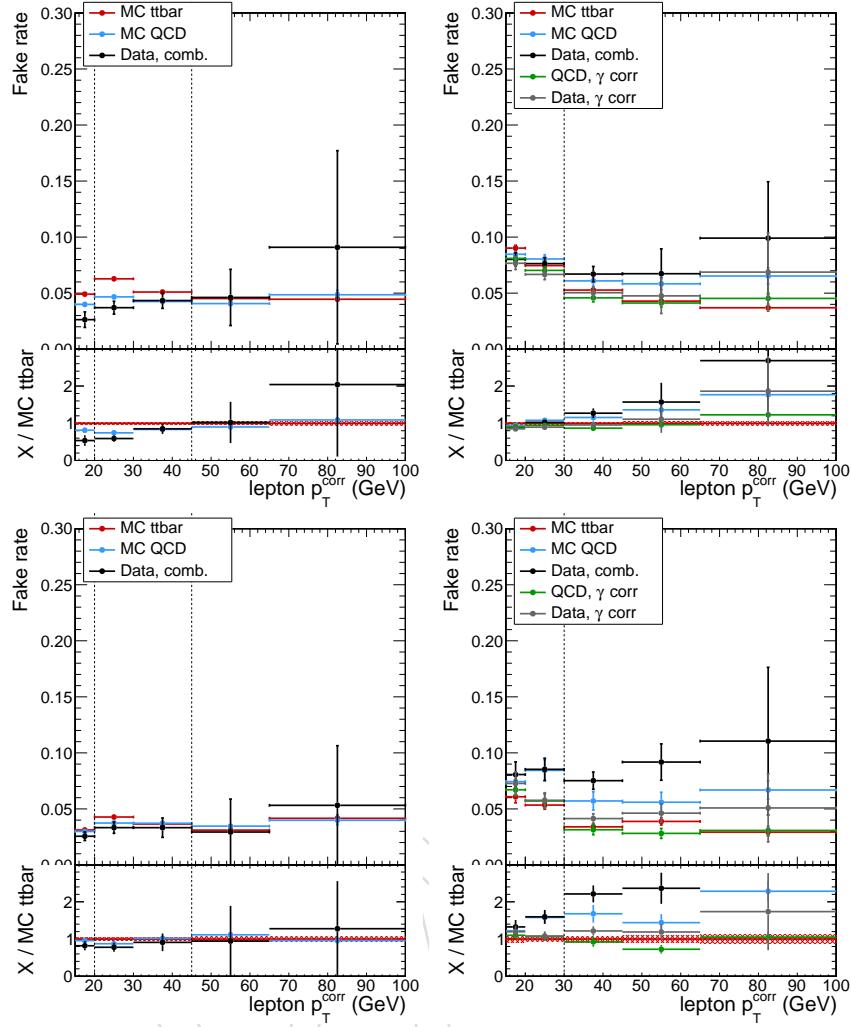


Figure 61: Summary of the fake rate measurement in events in data, compared with the predictions from simulated events in the measurement region (blue) and from non-prompt leptons in $t\bar{t}$ (red). Plots in the left column are for muons, those in the right column for electrons; the top row is for the barrel, the bottom for the endcaps.

945 Examples of the distributions can be found in Figures 62.

946 The normalization uncertainties induced by the closure of the electron (muon) fake rate range
 947 from 10% to 30% (from 20% to 40%) depending on the analysis category, being larger in the
 948 b-tight case. Separate nuisance parameters account for the normalization in the b-loose and
 949 b-tight categories, and for different lepton flavors.

950 Shape uncertainties are evaluated at fixed normalization, and accounted for in the fit with
 951 linear distortions acting simultaneously on the two axes of the 2D BDT plane. The slope of
 952 the distortion typically ranges from 0.1 to 0.4, with separate nuisances for b-loose and b-tight
 953 categories, and for different lepton flavors.

954 The uncertainty on the fake rate measurement is propagated to the background prediction
 955 introducing additional nuisances that vary both the background normalization via a coherent
 956 upwards or downwards shift of all bins in the fake rate map, and the background shape by
 957 introducing trends in the fake rate map as a function of the lepton p_T and $|\eta|$, within the fake
 958 rate map uncertainty (Fig. 63), at constant normalization. This is done separately for electrons

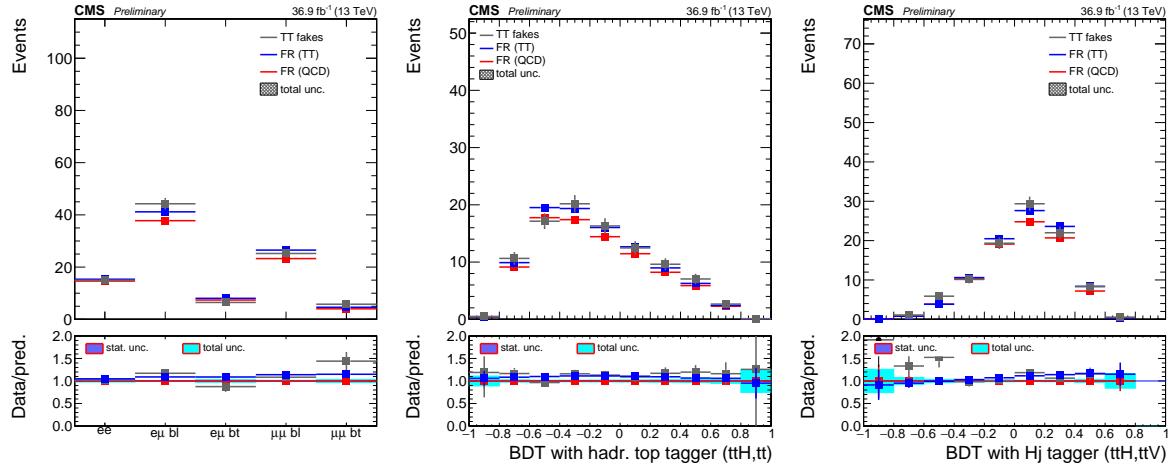


Figure 62: Normalization of event categories and distribution of background events in semi-leptonic $t\bar{t}$ events, compared with background predictions obtained in simulation using fake rates extracted in QCD and $t\bar{t}$ events.

959 and muons.

DRAFT

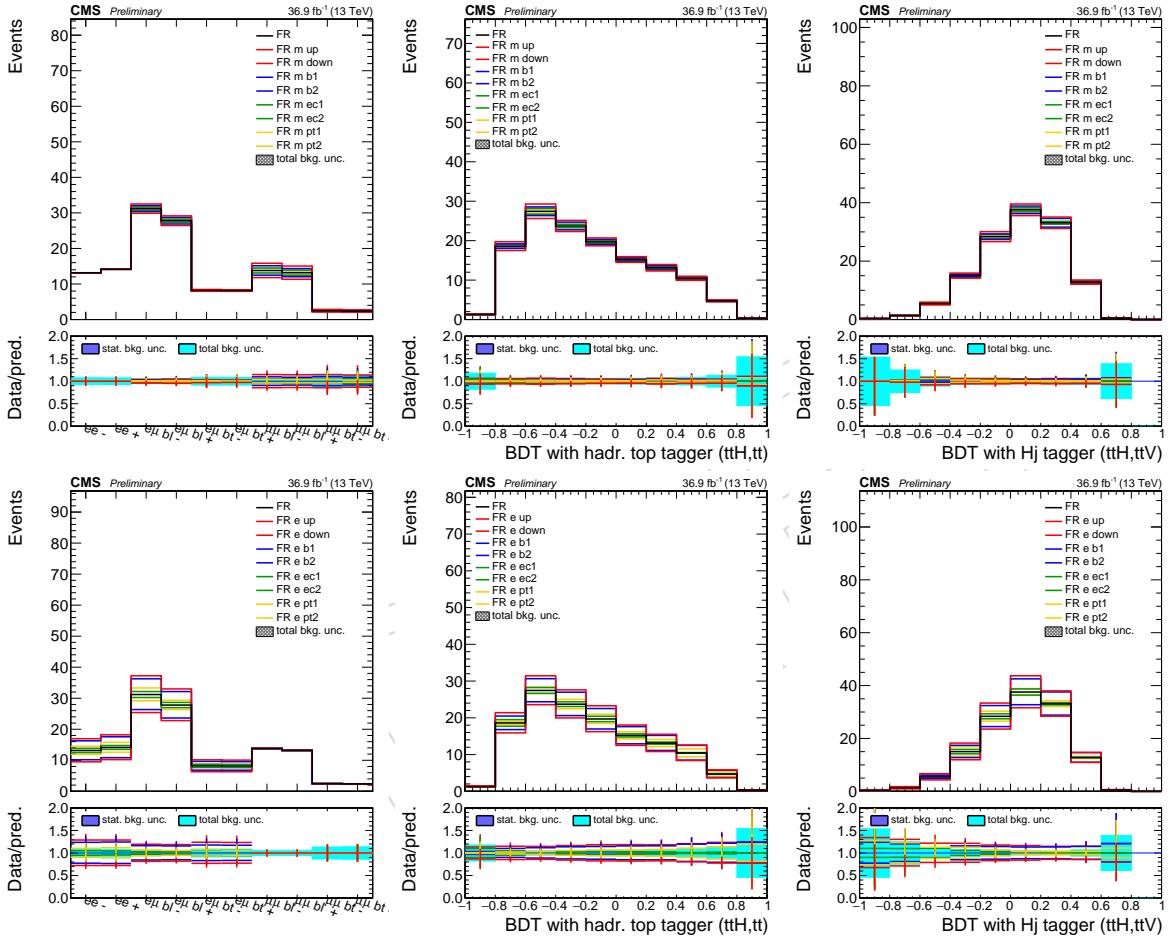


Figure 63: Variation of background prediction induced by shifts and distortions of the measured data fake rate maps within their uncertainties. Red: global shift, blue: shift in barrel region only, green: shift in endcap region only, yellow: trend as a function of the lepton p_T . Top row is for muons, bottom one for electrons.

960 8 Results

961 The results are interpreted by comparing the observed yields with the expectation from back-
 962 ground and a 125 GeV SM Higgs boson. We introduce a signal strength parameter $\mu = \sigma/\sigma_{\text{SM}}$,
 963 and we scale by that value the expected yields from $t\bar{t}H$ without altering the branching frac-
 964 tions or the kinematics of the events.

965 Results in terms of the asymptotic 95% CL upper limit on μ are presented in Table 9.

966 The observed (median expected in absence of signal) upper limit from the combination of all
 967 decay modes is 2.5 (0.8). The observed (expected) best fit signal strength for the SM Higgs
 968 hypothesis is $1.78^{+0.60}_{-0.54}$ ($1.00^{+0.46}_{-0.42}$) times the SM expectation, as shown in Table 10. The observed
 969 (expected) significance is 3.4σ (2.4σ).

970 The impact of statistical, theoretical and experimental sources of uncertainty is detailed in Ta-
 971 ble 11.

Category	Observed limit	Expected limit $\pm 1\sigma$
same-sign di-lepton	2.8	$0.86 (-0.25) (+0.39)$
three lepton	2.7	$1.34 (-0.41) (+0.64)$
four lepton	6.1	$4.70 (-1.66) (+2.96)$
combined	2.5	$0.76 (-0.23) (+0.34)$

Table 9: Asymptotic 95% CL upper limits on μ under the background-only hypothesis.

Category	Observed μ fit $\pm 1\sigma$	Expected μ fit $\pm 1\sigma$
same-sign di-lepton	$1.78 (-0.54) (+0.60)$	$1.00 (-0.47) (+0.51)$
three lepton	$1.16 (-0.76) (+0.84)$	$1.00 (-0.67) (+0.76)$
four lepton	$1.05 (-1.58) (+2.35)$	$1.00 (-1.56) (+2.29)$
combined	$1.56 (-0.48) (+0.54)$	$1.00 (-0.42) (+0.46)$

Table 10: Best fit of the signal strength parameter.

Category	Expected uncertainty on μ
Statistical sources	$(-0.26) (+0.27)$
Theoretical sources	$(-0.21) (+0.24)$
Experimental sources	$(-0.25) (+0.28)$
Total	$(-0.42) (+0.46)$

Table 11: Split of expected uncertainty in statistical, theoretical and experimental contributions.

972 Figure 64 show the post-fit distribution of the binned discriminating variables and the popula-
 973 tion of the categories used in the fit.

974 Figure 65 shows the post-fit values of the nuisances and their correlation with the fitted signal
 975 strength.

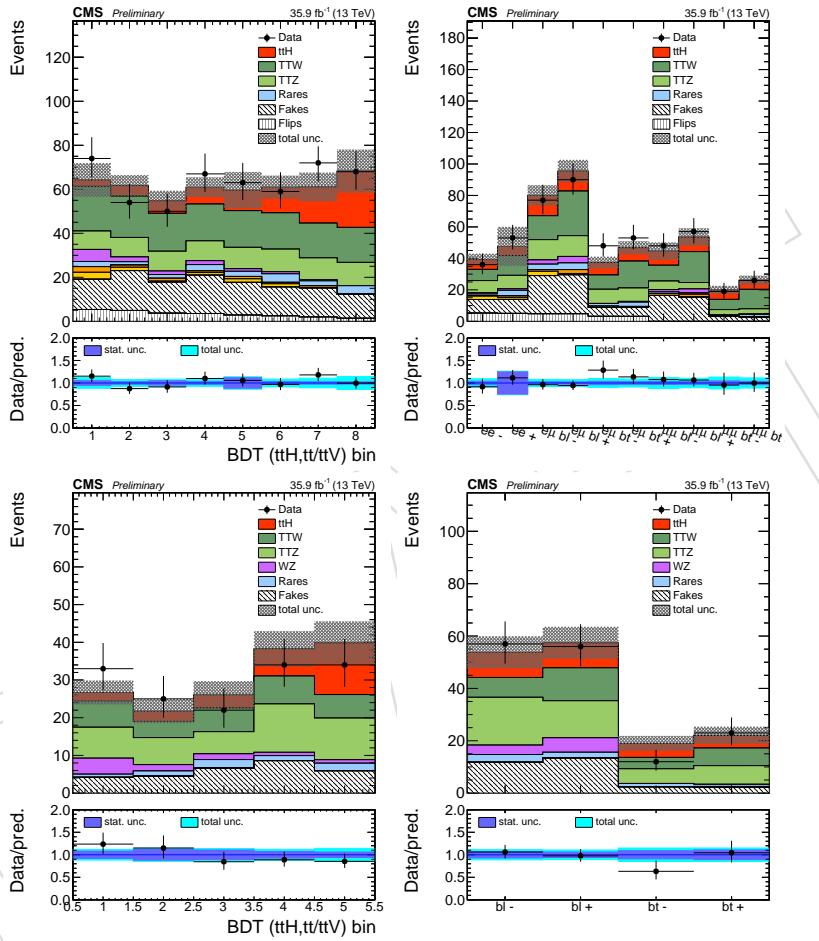


Figure 64: Post-fit distributions of discriminating variables and category population for 2lss (top row) and 3l (bottom row)

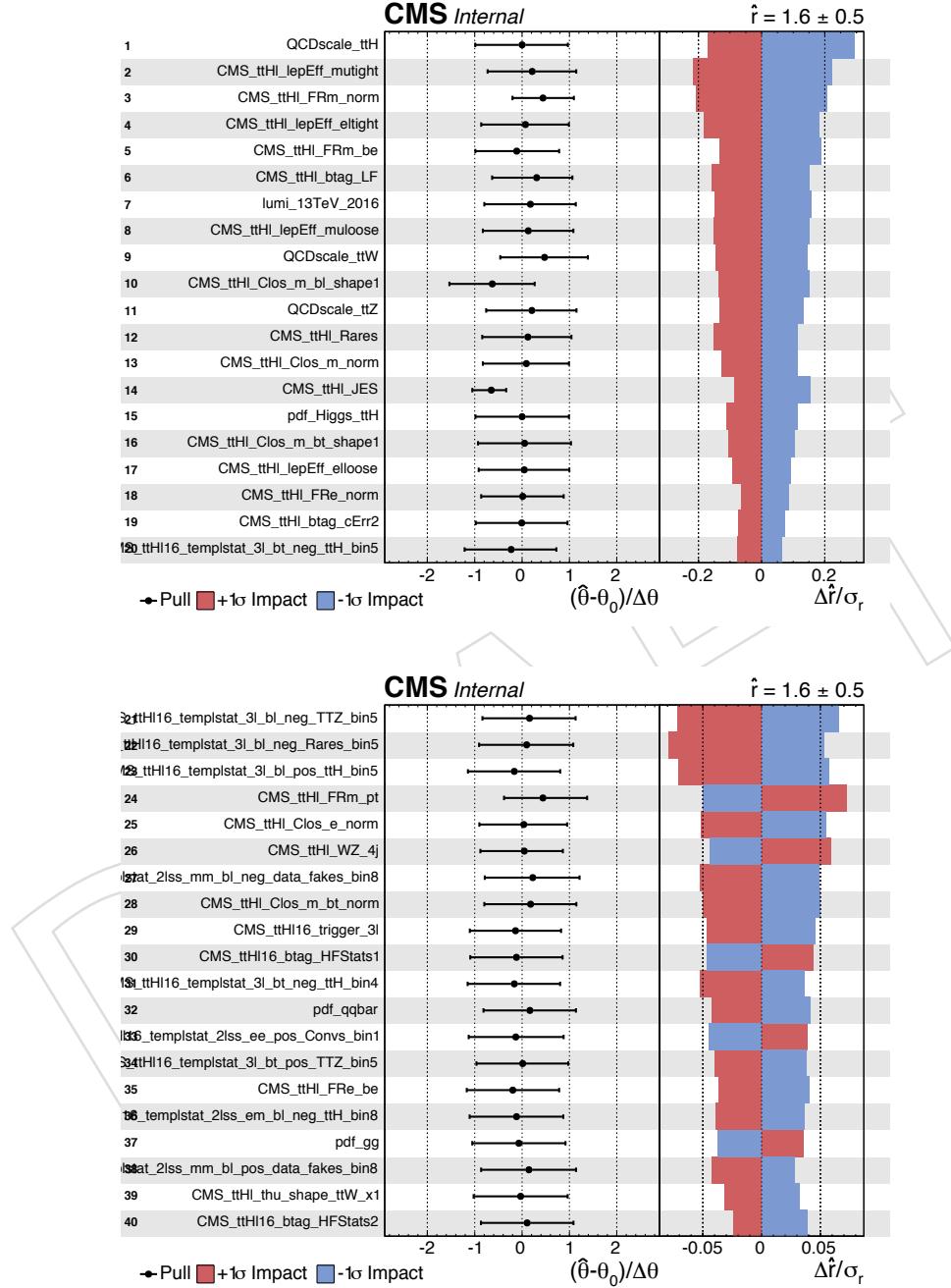


Figure 65: Impact plot showing the correlation between the main nuisance parameters and the best fit signal strength.

976 A Control region plots

977 A.1 Lepton MVA sideband region

978 The 2lss selection is modified by requiring that only one of the two selected leptons fails the
 979 tight lepton requirements, but still passes those for the fakeable object. In this way, we select a
 980 region enriched in $t\bar{t}$ events, where the lepton that fails the tight requirement is a fake lepton.

981 It is worth noting that the contamination from QCD events is not taken into account by the
 982 simulation. We observe a good agreement between simulation and data in terms of the shape of
 983 observables used for the selection and as inputs to the BDT discriminators. The latter variables
 984 are shown in Fig. 69.

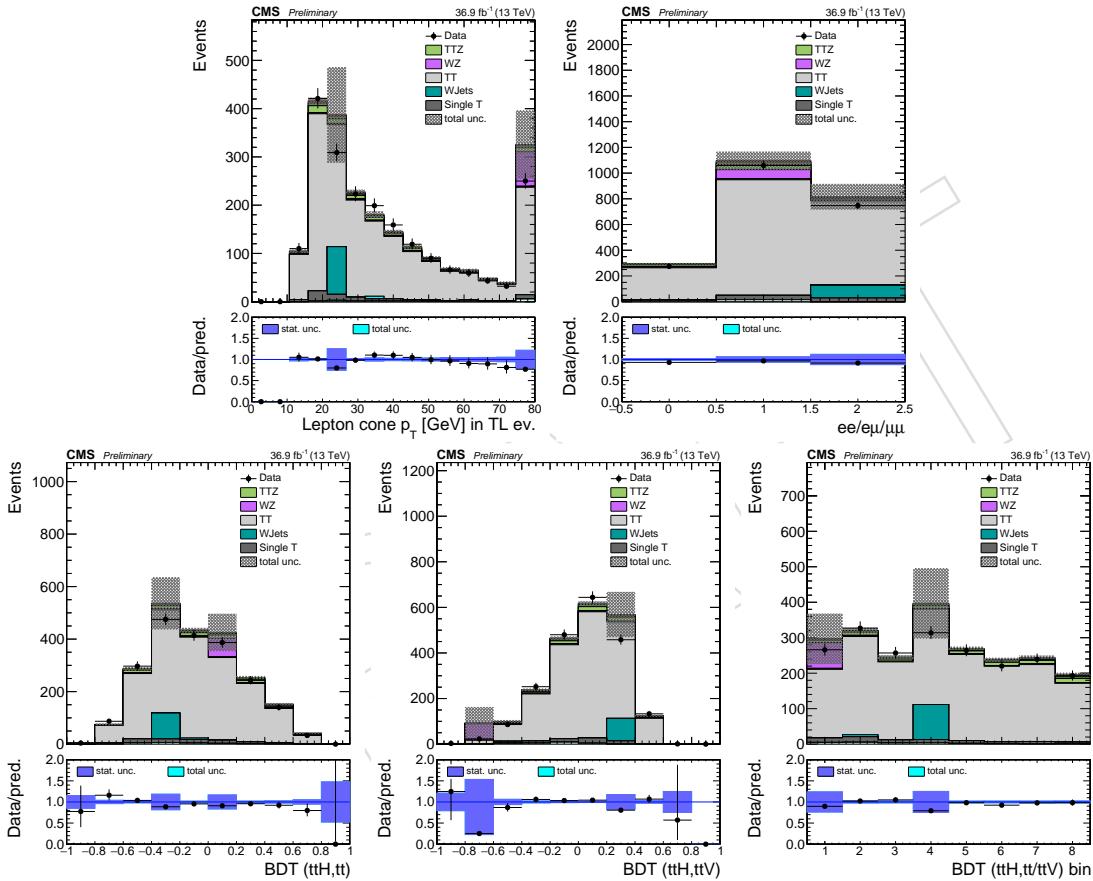


Figure 66: Data and simulation distributions in the 2lss control region with exactly one fakeable lepton failing the tight selection requirements. From top left to bottom right: the cone-corrected p_T of the failing lepton, the flavor of the lepton pair, the signal BDT discriminators against $t\bar{t}$ and $t\bar{t}V$ including the 2D-binned version as described in Section 5. Uncertainties are statistical only.

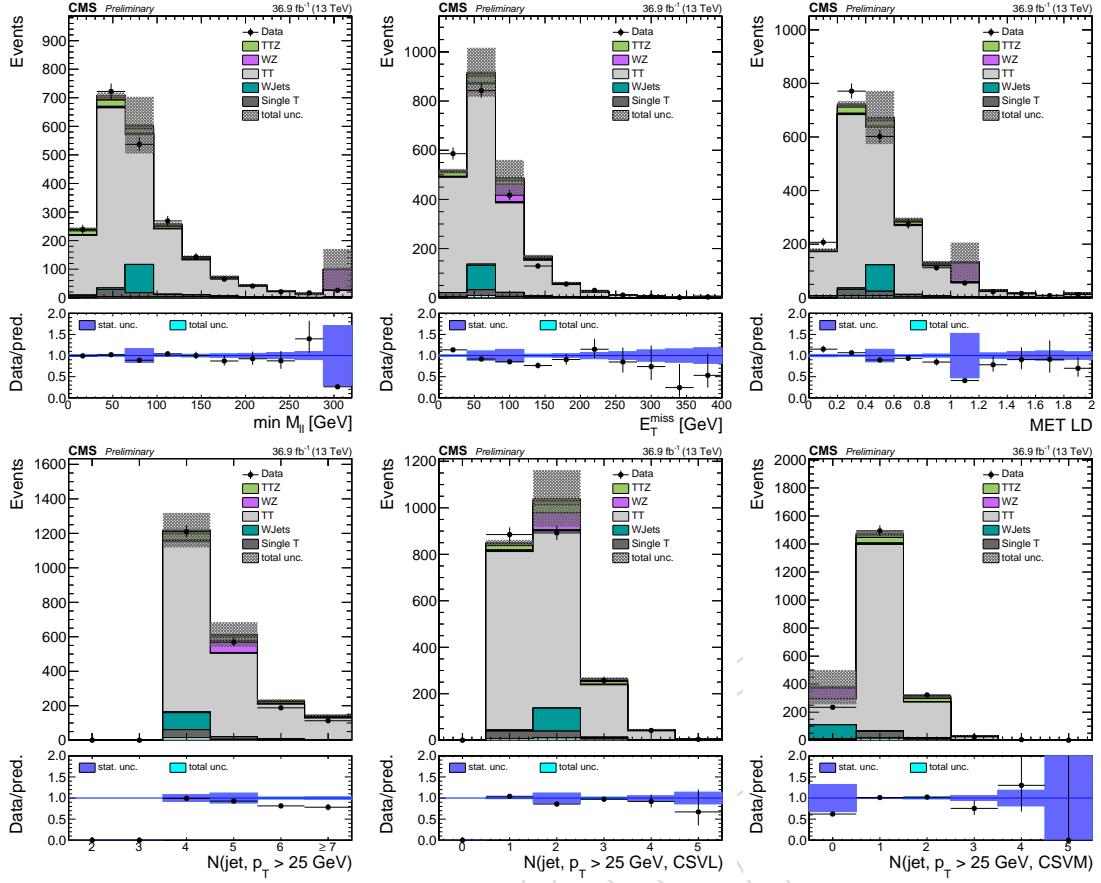


Figure 67: Data and simulation distributions in the 2lss control region with exactly one fakeable lepton failing the tight selection requirements. From top left to bottom right: the minimum invariant mass of loose di-lepton pairs, E_T^{miss} , $E_T^{\text{miss}}LD$, multiplicity of inclusive and b-tagged jets. cone-corrected p_T of the failing lepton, the flavor of the lepton pair, the signal BDT discriminators against $t\bar{t}$ and $t\bar{t}V$. Uncertainties are statistical only.

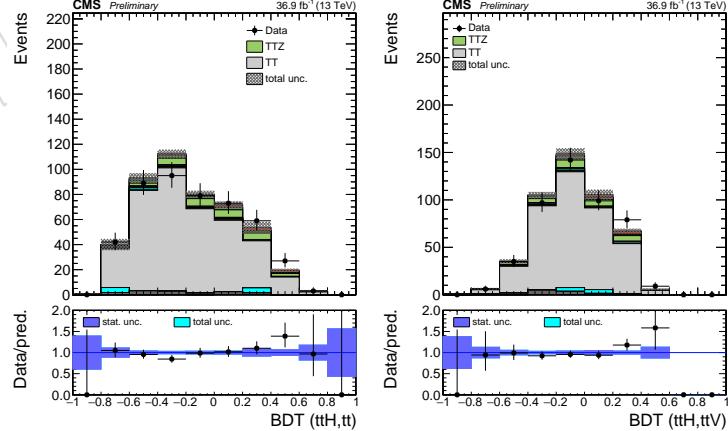


Figure 68: Same as Fig. 66, for the 3l category of the analysis.

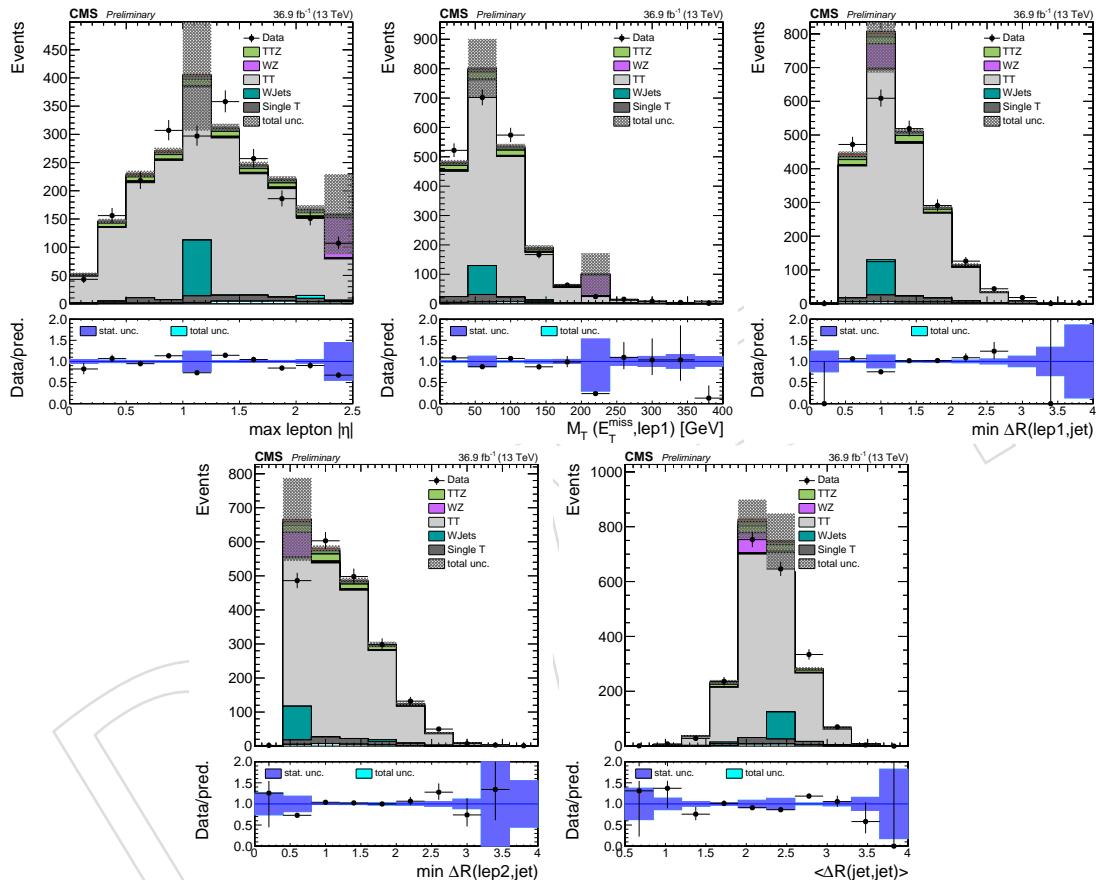


Figure 69: Distributions of several BDT input variables in the 2lss control region with exactly one fakeable lepton failing the tight selection requirements. Uncertainties are statistical only.

985 A.2 Jet multiplicity sideband region

986 This 2lss control region is enriched in fakes from $t\bar{t}$. It is obtained by requiring exactly three
 987 reconstructed jets in the final state, in the place of the requirement of at least four that is applied
 988 in the standard 2lss selection.

989 Fakes from $W+jets$ are estimated by the fake rate method described in Section 7.4, applied on
 990 MC events, while all other processes are predicted by the simulation. Distributions of event
 991 observables are shown in Fig. 70-72. In all cases we observe a satisfactory data/MC agreement,
 992 within the statistics currently available.

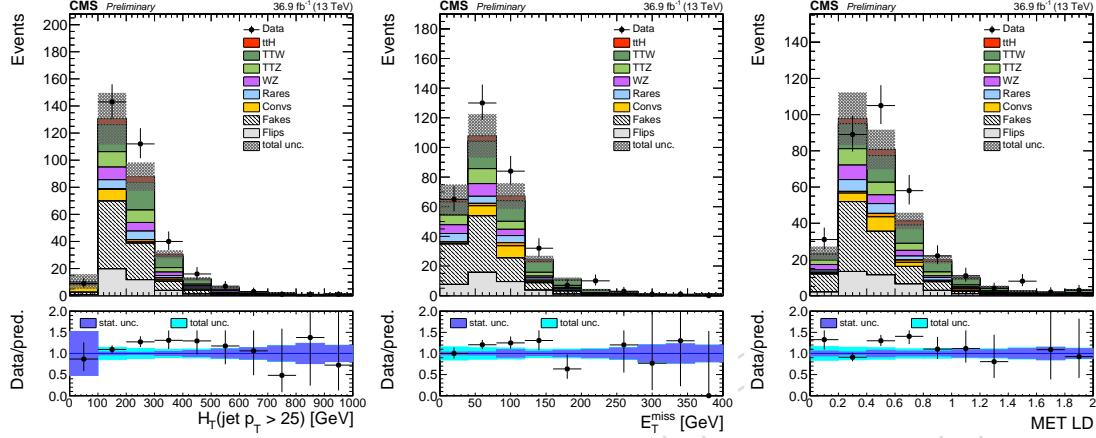


Figure 70: Distributions in the 2lss control region with exactly three jets in the final state. From left to right: the H_T , the E_T^{miss} , the $E_T^{\text{miss}} \text{LD}$. Uncertainties are statistical only.

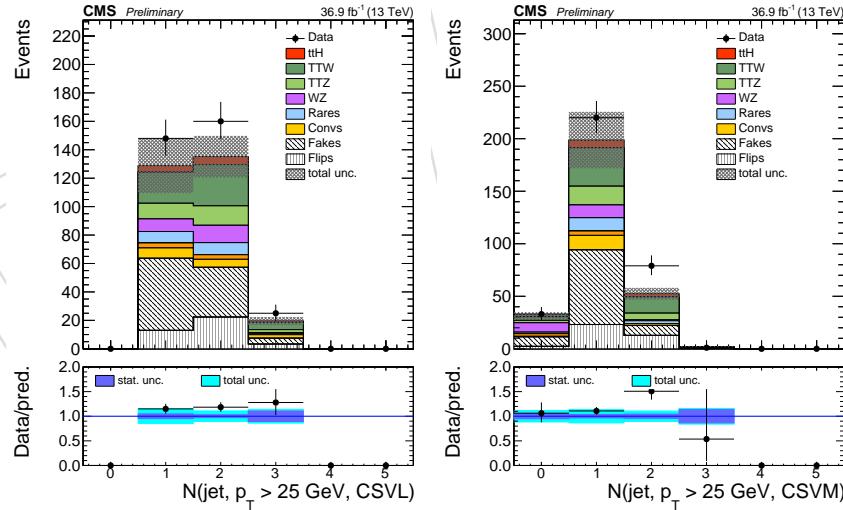


Figure 71: Distributions for the number of jets passing the loose and medium working points of the CSV b-tagger, in the 2lss control region with exactly three jets in the final state. Uncertainties are statistical only.

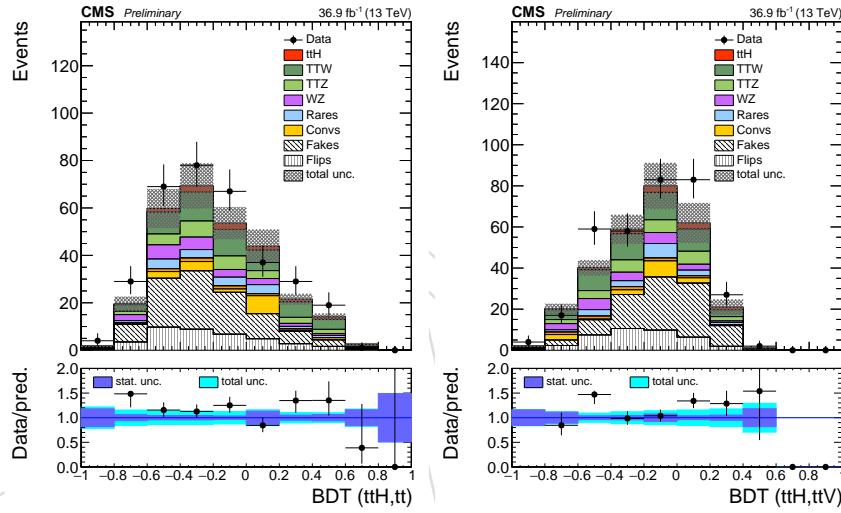


Figure 72: Distributions of the discriminators against $t\bar{t}$ and $t\bar{t}V$ in the 2lss control region with exactly three jets in the final state. Uncertainties are statistical only.

993 **A.3 $WZ \rightarrow 3\ell$**

994 With this control region we want to validate our objects (signal leptons, $E_T^{miss} LD$, jets) in the
 995 three lepton final state. A sample enriched in $WZ \rightarrow 3\ell$ events is selected modifying the 31
 996 selection in the following way:

- 997

 998 - the Z veto is inverted, i.e. we require the presence of a pair of loose opposite-sign
 999 same-flavor leptons whose invariant mass is within 10 GeV from the nominal Z boson mass;
 1000 - we require that no selected jets satisfy the medium working point of the CSV b-
 1001 tagging discriminator

1002 Some distributions are shown in Fig. 73.

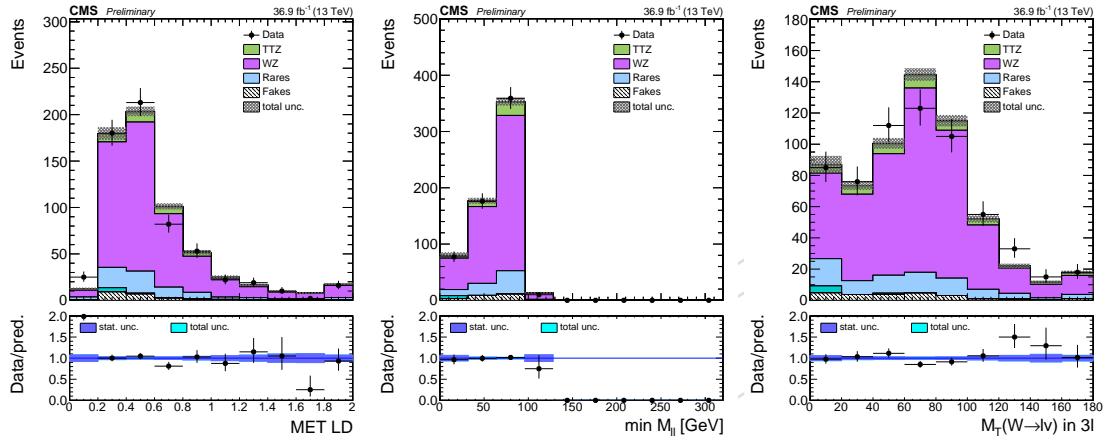


Figure 73: Data and simulation distributions in the $WZ \rightarrow 3\ell$ control region. From left to right: the $E_T^{miss} LD$, the minimum invariant mass of any $\ell\ell$ couples, M_T of the W boson candidate.

1003 A.4 $t\bar{t}Z \rightarrow 3\ell$

1004 The prediction for the $t\bar{t}Z$ process is tested directly in a trilepton control region requiring two
 1005 of the leptons to have the same flavour, opposite electrical charge and the invariant mass pair
 1006 of the pair to be within 10 GeV of the nominal Z boson mass.

1007 The definition of the control region differs from the one used for the 3l category of the analysis
 1008 in the following points:

- 1009 • the Z veto requirement is inverted, as described above;
- 1010 • the cut on the multiplicity b-tagged jets is tightened, requiring at least two loose and
 1011 one medium b-tagged jets

1012 The background from non-prompt leptons is estimated from data. Some distributions are
 1013 shown in Fig. 74.

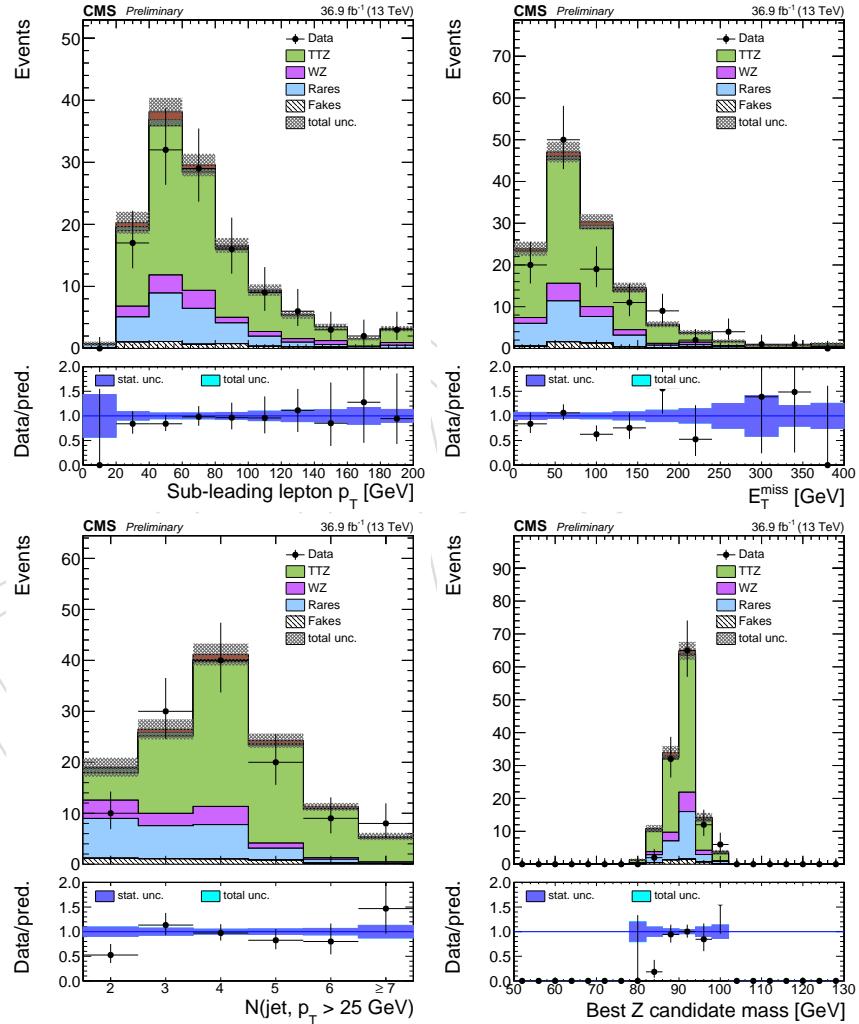


Figure 74: Data and simulation distributions in the $t\bar{t}Z \rightarrow 3\ell$ control region. From left to right: the p_T distribution of the second lepton ordered in p_T , the E_T^{miss} , the number of central jets with $p_T > 25$ GeV, the invariant mass of the best Z candidate.

1014 When requiring also the presence of at least four selected jets, as expected for a fully recon-
 1015 structed $t\bar{t}Z$ event, the control region becomes more pure in selecting $t\bar{t}Z$ events. This can
 1016 be seen in the distributions in Fig. 75.

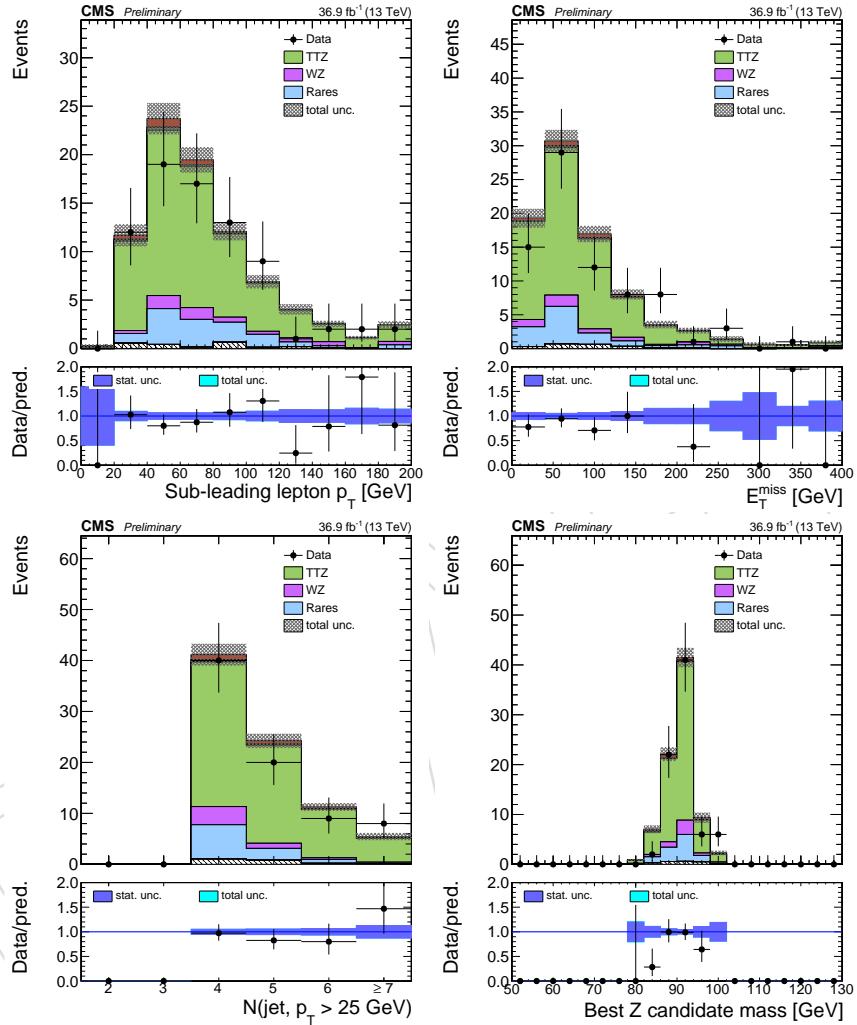


Figure 75: Data and simulation distributions in the $t\bar{t}Z \rightarrow 3\ell$ control region, with the additional requirement of at least four reconstructed jets. From left to right: the p_T distribution of the second lepton ordered in p_T , the E_T^{miss} , the number of central jets with $p_T > 25$ GeV, the invariant mass of the best Z candidate.

1017 B Matrix Element Method

1018 In this section we describe the Matrix Element Method (MEM) for the $t\bar{t}H$ multilepton analysis.

1019 B.1 The algorithm

1020 The matrix element method consists in estimating the probability of an event to be compatible
 1021 with the signal or background hypothesis, by computing the cross section of signal or back-
 1022 ground processes on a given phase-space point, corresponding to the reconstructed kinematic
 1023 configuration of the event.

For each hypothesis α , a weight $w_{i,\alpha}$ is computed for the event i using the following formula:

$$w_{i,\alpha}(\Phi') = \frac{1}{\sigma_\alpha} \int d\Phi_\alpha \cdot \delta^4\left(p_1^\mu + p_2^\mu - \sum_{k \geq 2} p_k^\mu\right) \cdot \frac{f(x_1, \mu_F) f(x_2, \mu_F)}{x_1 x_2 s} \cdot \left| \mathcal{M}_\alpha(p_k^\mu) \right|^2 \cdot W(\Phi' | \Phi_\alpha)$$

1024 where σ_α is the cross section of the process α , Φ' is the 4-momenta of the reconstructed par-
 1025 ticles in the event, $d\Phi_\alpha$ are the process-dependent integration variables, corresponding to the
 1026 4-momenta of all the particles at the vertex in the hypothesis α , the δ symbol represents the
 1027 momentum conservation between incoming and final state particles, $f(x, \mu_F)$ are the parton
 1028 density function in the proton, x_1, x_2 are the fraction of proton energy carried by the incoming
 1029 particles, $\left| \mathcal{M}_\alpha(p_k^\mu) \right|^2$ is the matrix element squared, and W are the transfer functions relating
 1030 the energy of particles at the vertex with their energy reconstructed with the detector.

1031 The MEM for $t\bar{t}H$ multilepton analysis is implemented in C++, thus can be easily interfaced with
 1032 analysis code. The integration is performed using VEGAS [32] stratified/importance sampling
 1033 implementation in ROOT. The matrix element squared is taken from Madgraph standalone
 1034 C++ code at LO. The parton distribution functions are taken from LHAPDF6 [33]. Transfer
 1035 functions are evaluated in CMS Run II Monte Carlo simulation. The phase space is analytic
 1036 and implemented as in Madweight [34].

1037 $t\bar{t}H$, and $t\bar{t}V$ and $t\bar{t}$ hypotheses

1038 In this analysis, the MEM for two lepton same-sign, three leptons and four leptons categories
 1039 are considered. Hypotheses corresponding to $t\bar{t}H$ signal (with semi-leptonic and fully leptonic
 1040 Higgs decay), and $t\bar{t}W$, $t\bar{t}\gamma^*/Z$ irreducible backgrounds, and $t\bar{t}$ reducible backgrounds (semi-
 1041 leptonic and fully leptonic top decay) are included.

1042 For all of the three processes, the W mass for W arising from top and Higgs decays is not
 1043 treated as fixed, and follows a Breit-Wigner as specified by the matrix element squared. The
 1044 W and γ^*/Z bosons produced in association with $t\bar{t}$ in $t\bar{t}V$ processes are also following a Breit
 1045 Wigner. Interference between γ^* and Z is included in the $t\bar{t}\gamma^*/Z$ matrix element.

1046 On the other hand, to decrease the number of integration variables, the masses of the top quark
 1047 and Higgs bosons are set to 173 GeV and 125 GeV respectively. In this narrow-width approx-
 1048 imation, matrix element and phase-space of top, anti-top and Higgs/ W/Z can be computed
 1049 independently:

$$\left| \mathcal{M}_\odot \right|^2 = \left| \mathcal{M}_{TTH} \right|^2 \cdot \left| \mathcal{M}_{Top} \right|^2 \cdot \left| \mathcal{M}_{Antitop} \right|^2 \cdot \left| \mathcal{M}_{Higgs} \right|^2$$

$$d\Phi_{tot} = d\Phi_{gg \rightarrow TTH} \cdot d\Phi_{Top} \cdot d\Phi_{Antitop} \cdot d\Phi_{Higgs}$$

The phase-space $d\Phi_\alpha$ is made of the product of all final state particle $dE d\theta d\phi$. Integration over some of these variables can be cancelled using momentum conservation formula. We implemented the phase-space parametrization proposed in Madweight MEM paper. This was compared with a custom parametrization and found to be faster. Changes of variables are performed to make explicit the W mass, which follows always a Breit Wigner and do not depends on event kinematics, such that VEGAS can treat it as an independent variable. We reproduce here the integration variables which are used:

$$d\Phi_{top,had} \propto dE_b d\theta_b d\phi_b \cdot d\theta_{j1} d\phi_{j1} \cdot d\theta_{j2} d\phi_{j2} \cdot dm_W$$

$$d\Phi_{top,lep} \propto dE_b d\theta_b d\phi_b \cdot dE_l d\theta_l d\phi_l \cdot d\phi_\nu dm_W$$

$$d\Phi_{H \rightarrow 2l2\nu} \propto dE_{l1} d\theta_{l1} d\phi_{l1} \cdot dE_{l2} d\theta_{l2} d\phi_{l2} \cdot dE_{\nu1} d\theta_{\nu1} d\phi_{\nu1} \cdot d\phi_{\nu2} dm_W$$

$$d\Phi_{H \rightarrow l\nu jj} \propto dE_{j1} d\theta_{j1} d\phi_{j1} \cdot dE_{j2} d\theta_{j2} d\phi_{j2} \cdot dE_{l1} d\theta_{l1} d\phi_{l1} \cdot d\phi_{\nu1} dm_W$$

$$d\Phi_Z \propto dE_{l1} d\theta_{l1} d\phi_{l1} \cdot dE_{l2} d\theta_{l2} d\phi_{l2}$$

$$d\Phi_W \propto dE_l d\theta_l d\phi_l \cdot d\theta_\nu d\phi_\nu dm_W$$

1050 For a given value of the integration variables, the momenta of all particles can be computed by
 1051 solving a set of linear and quadratic equations. The b -quark mass is set to 4.7 GeV while the
 1052 masses of the non- b quarks and leptons is set to 0.

1053 The matrix element squared provided by Madgraph is leading order in pQCD, i.e. with no
 1054 additional jets recoiling against the $t\bar{t}H/V$ system. However additional jets are present in data.
 1055 To evaluate the ME, the $t\bar{t}H/V$ system needs to be boosted back along the Px/Py direction such
 1056 that the incoming particles have no Px/Py component.

1057 The pdf chosen is NNPDF2.3 LO QED. The factorization scale in the pdf and matrix element are
 1058 chosen to be $\mu_F = (m_t + m_H)/2$ for $t\bar{t}H$, $\mu_F = (m_t + m_W)/2$ for $t\bar{t}W$, and $\mu_F = (m_t + m_{\gamma^*/Z})/2$
 1059 for $t\bar{t}\gamma^*/Z$.

1060 B.2 Treatment of jets and permutations

1061 To evaluate the matrix element, leptons, jets and b -jets need to be assigned to the ME leptons
 1062 and quarks. The jets assigned to the two b -quarks from tops are the two reconstructed jets with
 1063 highest CSV value (disregarding the analysis categories 1b tight / 2b loose).

1064 If all the needed jets are reconstructed, one has to choose 4 (2) jets among the remaining jets
 1065 in two lepton same-sign (three lepton) category. In the general case, the jets to be assigned
 1066 to the ME will arise from W decay from top or Higgs. For this reason, we select the dijet
 1067 pair with mass closest to m_W . In the three lepton categories, when attempting to evaluate the
 1068 $t\bar{t}H, H \rightarrow l\nu jj$ hypothesis, the dijet pair can also arise from W^* decay. Here we choose the dijet
 1069 pair with lowest dijet mass (dijet mass is bounded by the p_T cut on the jets to be mostly greater
 1070 than 50 GeV). Similarly in two lepton same-sign categories, once the jet pair with mass closest
 1071 to m_W are selected, two more jets with lowest dijet mass are selected. It was shown that this
 1072 "mixed" way of selecting jets performs better than selecting only jet pairs by mass closest to W
 1073 mass.

1074 If the needed jets are not all reconstructed (e.g. not passing jet identification or escaping detec-
 1075 tor acceptance), in principle the matrix element can not be evaluated. However, by expanding
 1076 the phase-space with additional integration variables for the missing jets, it is possible to cir-
 1077 convene this difficulty. This feature is implemented in the 1-missing and 2-missing jets cases.

1078 Thus the MEM can be computed for all events selected with the baseline analysis, requiring
 1079 $\geq 4(2)$ jets including b-jets, in 2lss (3l) category.

1080 Since we cannot know a priori what is the correct lepton and jet assignment to the ME partons,
 1081 the ME is evaluated for all possible permutations of leptons and jets.

1082 For all hypotheses and categories, there are 2 permutations arising from b-jet permutation (top
 1083 or anti-top). If only one b-jet is selected, the permutation still needs to be done with the tops
 1084 exchanging absence or presence of the b-jet.

1085 The $t\bar{t}Z$ hypothesis can be computed in 3l (with 1 lepton from leptonic decay and 2 leptons
 1086 from Z decay) and 4l categories (2 leptons from both leptonic top decay and 2 leptons from Z
 1087 decay). In 3l/4l categories, there can be 1 or 2 same flavour opposite sign pair to make a Z, thus
 1088 1 or 2 lepton permutations are allowed, the other lepton(s) being assigned to the top/anti-top
 1089 according to the sign of their charge. In 3l categories, 2 jet permutations are arising from the
 1090 hadronic top decay.

1091 The $t\bar{t}W$ ME does not have quarks in 3l categories (2 leptons from leptonic top decay and 1
 1092 lepton from associated W production), thus no jet permutations are allowed. In 2lss categories,
 1093 2 jets among 4 can be assigned to the ME (from a hadronic top decay), leading to 2 jet permu-
 1094 tations in any case.

1095 The $t\bar{t}H, H \rightarrow 2l2\nu$ hypothesis can be computed for 3l (with one leptonic top decay) and 4l
 1096 categories (two leptonic top decays) only.

1097 The $t\bar{t}H, H \rightarrow l\nu jj$ hypothesis can be computed for 2lss (with one hadronic top decay) and
 1098 3l (both leptonic top decays). The biggest number of permutations is found in the 2lss_2b_4j
 1099 category, where permutations have to be done within the 4 jets. To reduce the number of
 1100 permutations in the special case of 2lss_2b_3j category (time consuming since there are more
 1101 integration variables) the missing jet can only be assigned to the W from Higgs decay, assuming
 1102 that the jet is lost because of not passing the p_T requirement.

1103 The $t\bar{t}$ hypothesis is splitted in semi-leptonic and fully leptonic top decay. In that case, the
 1104 lepton permutation needs to be done while there are more reconstructed leptons available (2
 1105 same sign or 3) than in the matrix element (1 for semi-leptonic and 2 for fully leptonic).

Once the MEM is computed for each of the permutations, an average weight is computed for
 each hypothesis. The weights with null value are excluded from the average (found to be
 slightly more discriminating than including them):

$$\begin{cases} w_\alpha = 10^{-300} & \text{if } \sum w_i = 0 \\ w_\alpha = \frac{1}{N_{w_i \neq 0}} \sum_{w_i \neq 0} w_i & \text{else} \end{cases}$$

1106 Transfer functions

1107 The transfer functions $W(\Phi'|\Phi_\alpha)$ give the probability density of measuring a set of observables
 1108 Φ' with the detector, given a phase space point Φ_α at ME level. For this analysis, the follow-
 1109 ing approximations are made. The lepton energy and its direction is assumed to be perfectly
 1110 measured. The direction of quarks is assumed to be perfectly measured by the direction of
 1111 the reconstructed jet. Thus no transfer functions are used for leptons, while jet energy transfer
 1112 functions are included in the MEM.

1113 Jets and b-jets energy transfer functions are evaluated in MC simulation with CMSSW_7_6_X.
 1114 The pdfs are histograms parameterized as a function of E_{rec}/E_{gen} , where E_{rec} is the jet re-
 1115 constructed energy after jet energy scale/resolution correction and E_{gen} is the energy of the

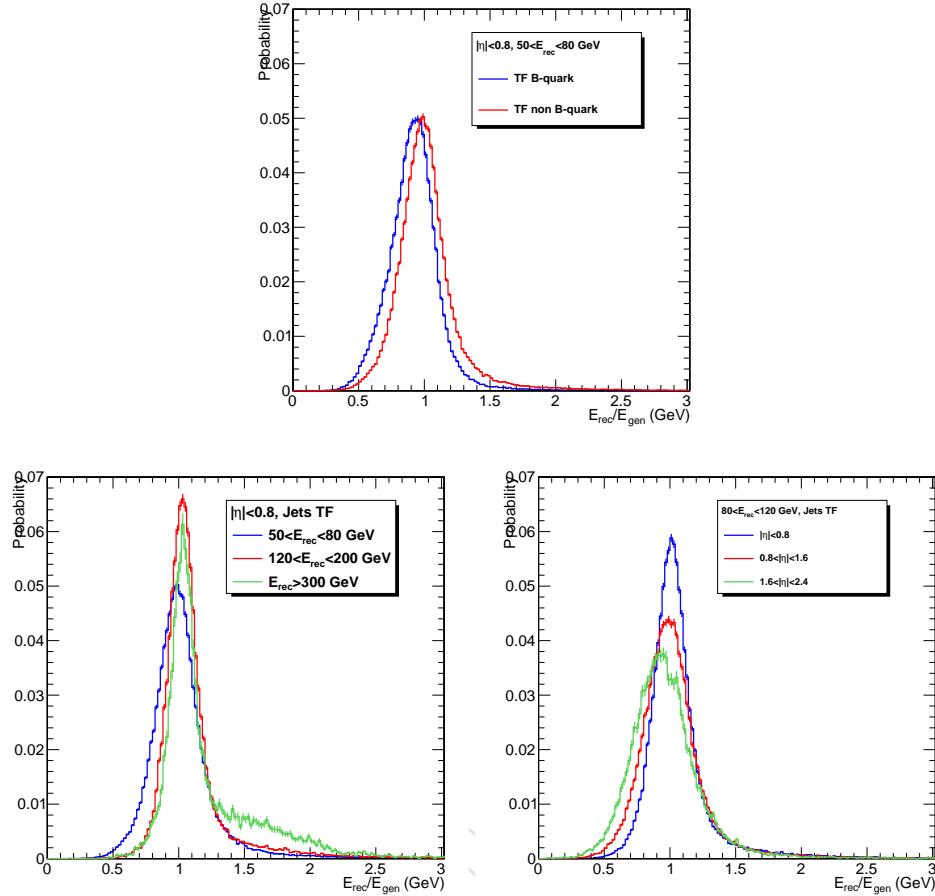


Figure 76: Example of jets transfer functions for (a) b-jets and non-b jets, (b) variation of jets TF with energy (c) variation of jets TF with η .

1116 matched parton. Transfer functions are defined in 3 bins of jet pseudorapidity ($|\eta| < 0.8$,

1117 $0.8 < |\eta| < 1.6$, $1.6 < |\eta| < 2.4$) and 6 bins of jet energy ($25 < E < 50$, $50 < E < 80$,

1118 $80 < E < 120$, $120 < E < 200$, $200 < E < 300$, $E > 300$ GeV). B-jet response is found to

1119 be slightly lower than jet response, as expected (due to missing momentum of neutrinos from

1120 B-hadron decays escaping detection). Transfer functions measured in $t\bar{t}H$ samples are used.

1121 Distributions were cross-checked in $t\bar{t}V$ and $t\bar{t}$ samples and found to behave similarly. Exam-

1122 ples of jets and b-jets transfer functions are shown fig. 76.

1123 If a jet is missing at reconstructed level, its transfer function is set to 0 if the associated MEM

1124 quark has $|\eta| > 2.4$, and to 1 if $|\eta| < 2.4$.

1125 Another set of transfer functions is used to constrain the total momentum of the $t\bar{t}H/V$ system.

1126 According to momentum conservation between intial state and final state particles, integration

1127 over the phase space of initial particles is cancelled. Despite the cancellation, one can con-

1128 strain the total momentum computed with the MEM at parton level with the total momentum

1129 reconstructed, by using a transfer function. For practical reason, this transfer function is approx-

1130 imated with a missing transverse energy transfer function. The parameterization is using the

1131 mET and $mET \phi$ distributions.

1132 **B.3 MEM discriminant**

1133 According to the Neyman-Person lemma, the likelihood of signal and background is the most
 1134 powerful test statistic for hypothesis testing. In the 2lss categories, a likelihood is built with the
 1135 $t\bar{t}H$ and $t\bar{t}W$ hypotheses, while in the 3l categories, a likelihood is built with $t\bar{t}H$ and $t\bar{t}W+t\bar{t}Z$
 1136 hypotheses as follows:

$$L_{2lss} = -\log\left(\frac{\sigma_{TTW}w_{TTW}}{\sigma_{TTH}w_{TTH} + k \cdot \sigma_{TTW}w_{TTW}}\right)$$

$$\begin{cases} L_{3l} = -\log\left(\frac{\sigma_{TTZ}w_{TTZ} + k \cdot \sigma_{TTW}w_{TTW}}{\sigma_{TTH}w_{TTH} + \sigma_{TTZ}w_{TTZ} + k \cdot \sigma_{TTW}w_{TTW}}\right) & \text{SFOS} \\ L_{3l} = -\log\left(\frac{k \cdot \sigma_{TTW}w_{TTW}}{\sigma_{TTH}w_{TTH} + k \cdot \sigma_{TTW}w_{TTW}}\right) & \text{no SFOS} \end{cases}$$

$$L_{4l} = -\log\left(\frac{\sigma_{TTZ}w_{TTZ}}{\sigma_{TTH}w_{TTH} + \sigma_{TTZ}w_{TTZ}}\right)$$

1137 MEM weights are weighted by relevant process cross section in the likelihood. Note that the
 1138 $t\bar{t}Z$ hypothesis can be included only if there is at least one same flavour opposite sign pair to
 1139 build a Z .

1140 A multiplicative factor k is included to counterbalance the missing phase space in the $t\bar{t}W$ hy-
 1141 pothesis with respect to the other processes (the $t\bar{t}W$ ME has 2 jets less in the matrix element
 1142 relative to $t\bar{t}H$ and $t\bar{t}Z$). In the case of 3l categories, k was tuned for each category (0/1/2 miss-
 1143 ing jets) to maximize the signal to background discrimination. In the 2lss categories, changing
 1144 k does not improve discrimination but allows easier fit of the final distributions.

1145 Final yields after event selection in CMSSW 7.6.X MC with 2015 luminosity assumed, are
 1146 shown fig. 77 for 2lss categories fig. 78 for 3l categories.

1147 **Comparison with TTV BDT**

1148 The performance of MEM discriminants is compared on fig. 79 for 2lss categories and fig. 80
 1149 for 3l categories. Performance is in any case comparable, slightly lower in 2lss categories, and
 1150 equivalent or slightly better in 3l categories. The performance is better for categories where all
 1151 the jets are reconstructed. There is almost no discrimination when 2 jets are missing, which is
 1152 also the case for the TTV BDT.

1153 **MEM as input variable to the TTV BDT**

1154 Given that the performances of TTV BDT and MEM discriminants are similar, it makes sense to
 1155 train BDT including the MEM. We train new BDT's using Madgraph $t\bar{t}W$ and $t\bar{t}Z$ large samples
 1156 with same setup used for training BDT TTV, but including MEM weights. Three trainings are
 1157 performed:

- 1158 • 2lss category: TTV BDT inputs + $\log(w_{TTH}) + \log(w_{TTW}) + \text{catJets}$, where catJets is
 1159 the jet category (0/1/2-missing jets)
- 1160 • 3l/4l categories with a SFOS lepton pair: TTV BDT inputs + $\log(w_{TTH}) + \log(w_{TTW})$
 1161 + $\log(w_{TTZ}) + \text{catJets}$
- 1162 • 3l categories without SFOS lepton pair: TTV BDT inputs + $\log(w_{TTH}) + \log(w_{TTW})$
 1163 + catJets

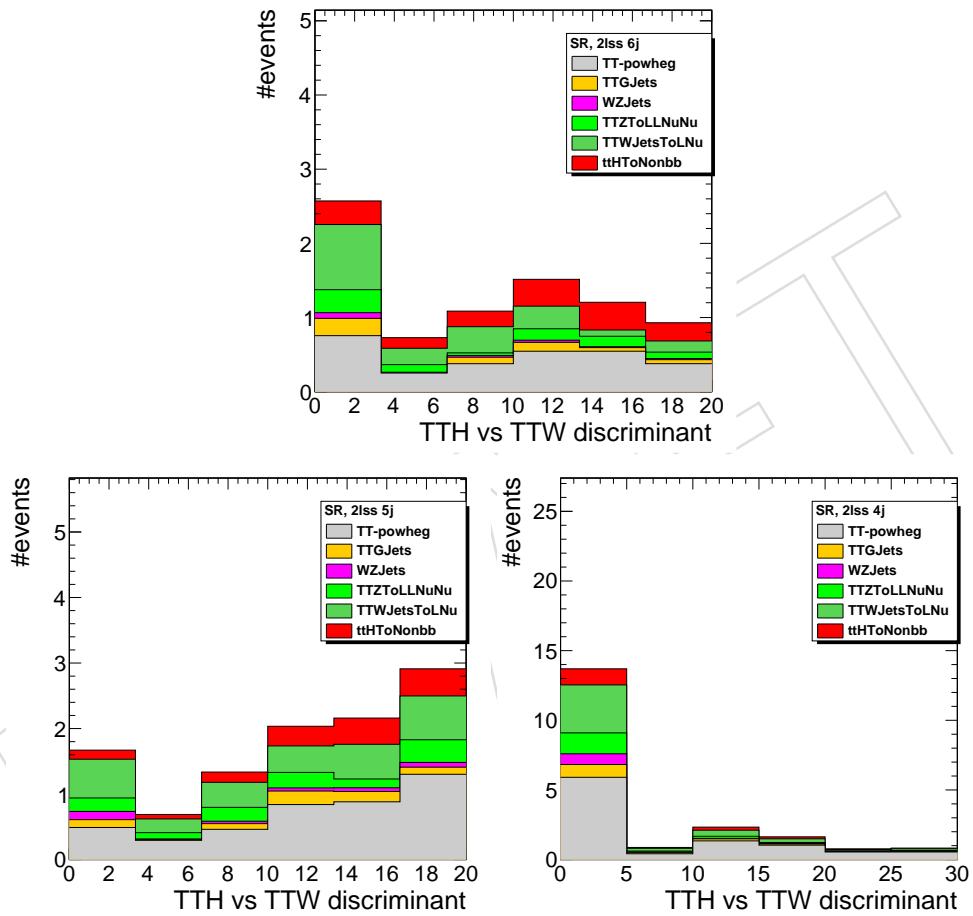


Figure 77: Final yields of MEM discriminants in 2lss signal region for (a) 0 missing jets, (b) 1 missing jets, (c) 2 missing jets.

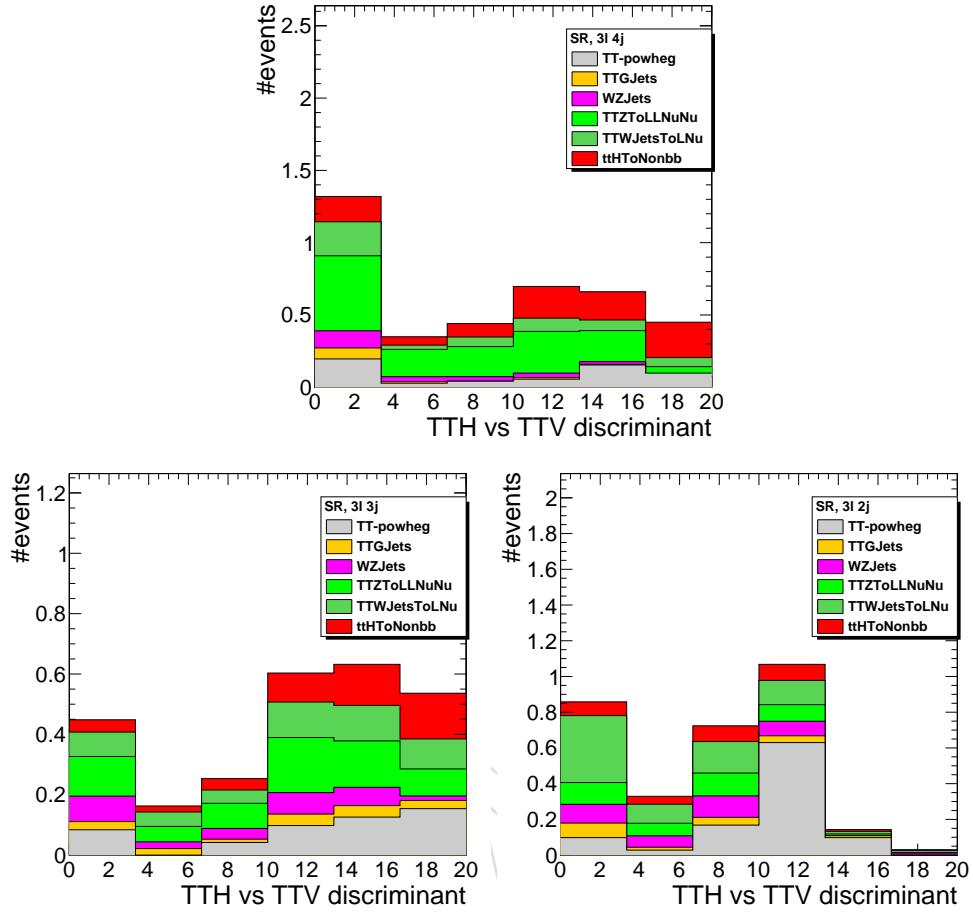


Figure 78: Final yields of MEM discriminants in 3l signal region for (a) 0 missing jets, (b) 1 missing jets, (c) 2 missing jets.

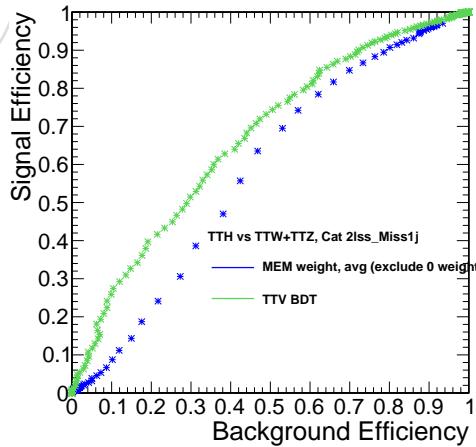


Figure 79: Not updated yet Comparison of MEM discriminants in 2lss signal region (merged).

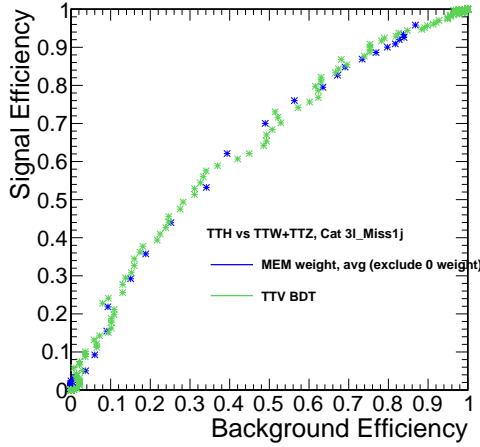


Figure 80: *Not updated yet* Comparison of MEM discriminants in 3l signal region (merged).

1164 The $\log(w)$ are the \log of average MEM weight excluding null weights. catJets variable is
 1165 included to make the BDT aware of the missing jet category.

1166 Results are shown on fig.87 for 2lss category, fig.88 for 3l without SFOS lepton pair, fig. 83 for 3l
 1167 with a SFOS lepton pair, and fig. 89 for 4l. Performance of BDT including MEM is greater than
 1168 the previous training of TTV BDT for all categories, by a few % in signal efficiency for a given
 1169 background rejection in 2lss categories and up to 10-15% in 3l and 4l categories.

1170 Comparison with TT BDT

1171 The performance of MEM discriminants is compared on fig. 85 for 2lss categories and fig. 86
 1172 for 3l categories. Performance is in any case comparable, slightly lower in 2lss categories, and
 1173 equivalent or slightly better in 3l categories. The performance is better for categories where all
 1174 the jets are reconstructed. There is almost no discrimination when 2 jets are missing, which is
 1175 also the case for the TTV BDT.

1176 MEM as input variable to the TT BDT

1177 Given that the performances of TT BDT and MEM discriminants are similar, it makes sense to
 1178 train BDT including the MEM. We train new BDT's using Madgraph $t\bar{t}$ large samples with same
 1179 setup used for training BDT TT, but including MEM weights. Three trainings are performed:

- 1180 • 2lss category: TT BDT inputs + $\log(w_{TTH}) + \log(w_{TTW})$ + weight of the $t\bar{t}H$ fully
 1181 leptonic and semi leptonic reconstruction
- 1182 • 3l/4l categories with a SFOS lepton pair: TT BDT inputs + $\log(w_{TTH}) + \log(w_{TT})$

1183 The $\log(w)$ are the \log of average MEM weight excluding null weights.

1184 Results are shown on fig.87 for 2lss category, fig.88 for 3l without SFOS lepton pair, fig. 83 for 3l
 1185 with a SFOS lepton pair, and fig. 89 for 4l. Performance of BDT including MEM is greater than
 1186 the previous training of TTV BDT for all categories, by a few % in signal efficiency for a given
 1187 background rejection in 2lss categories and up to 10-15% in 3l and 4l categories.

1188 B.4 MEM distributions after event selection

1189 Figures 90 and 91 show the distribution of log of MEM amplitudes after the full 3l event selec-
 1190 tion, in the signal and in the lepton MVA sideband region.

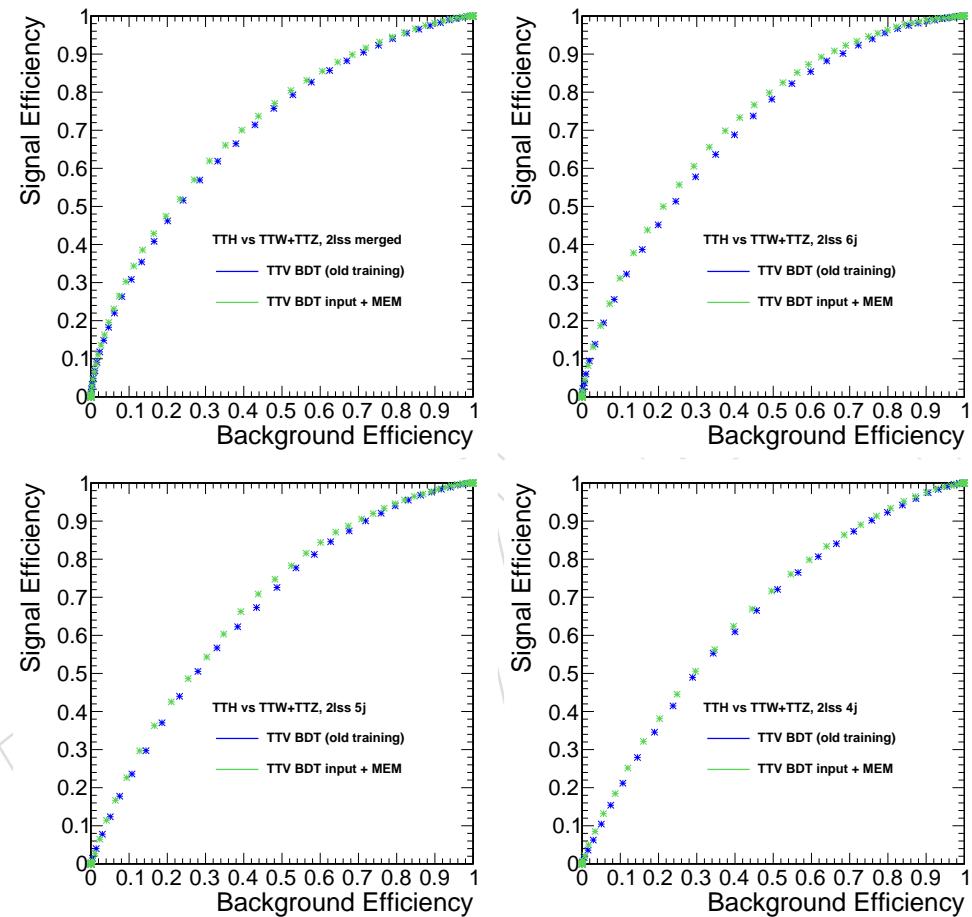


Figure 81: ~~Updated~~ Comparison of TTV BDT and new BDT including MEM in 2lss signal region for (a) all events (b) 0 missing jets, (c) 1 missing jets, (d) 2 missing jets.

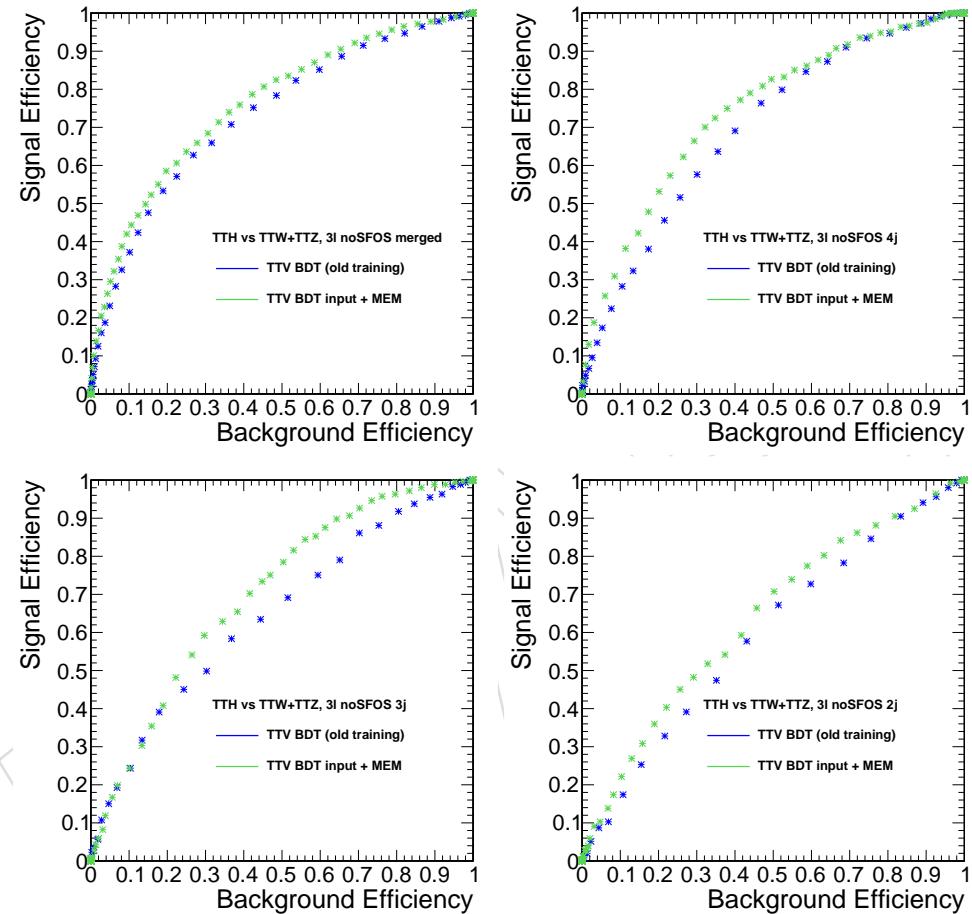


Figure 82: **Updated** Comparison of TTV BDT and new BDT including MEM in 3l signal region, without SFOS lepton pair, for (a) all events (b) 0 missing jets, (c) 1 missing jets, (d) 2 missing jets.

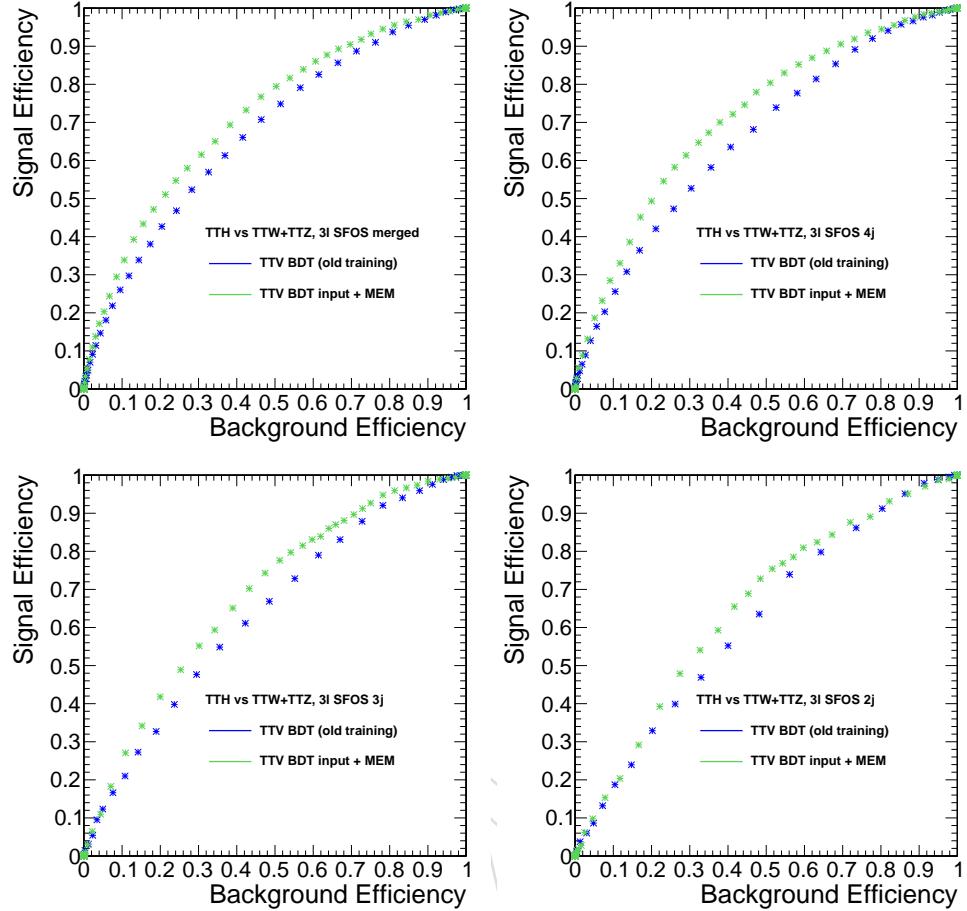


Figure 83: **Updated** Comparison of TTV BDT and new BDT including MEM in 3l signal region, with a SFOS lepton pair, for (a) all events (b) 0 missing jets, (c) 1 missing jets, (d) 2 missing jets.

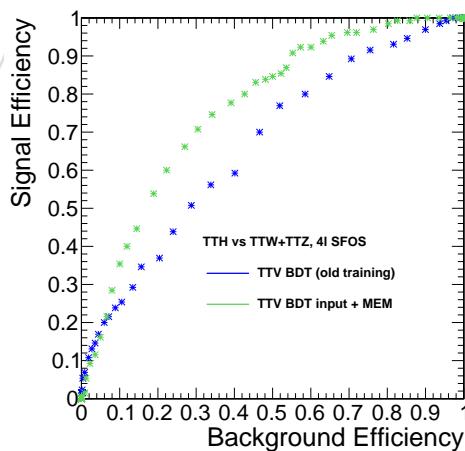


Figure 84: **Updated** Comparison of TTV BDT and new BDT including MEM in 4l signal region for all events

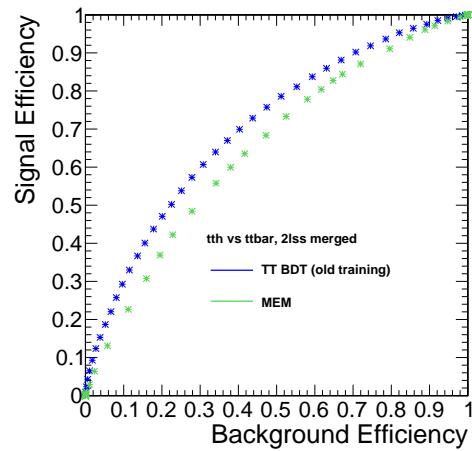


Figure 85: **Updated** Comparison of MEM discriminants in 2lss signal region (merged).

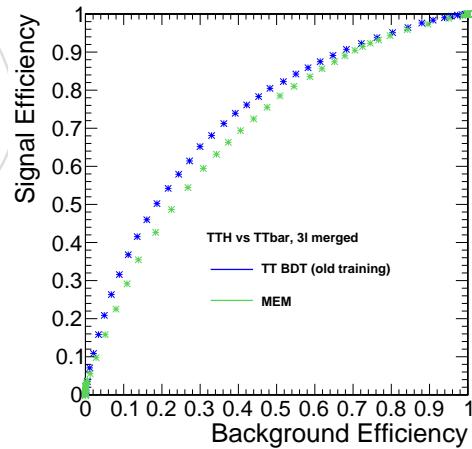


Figure 86: **Updated** Comparison of MEM discriminants in 3l signal region (merged).

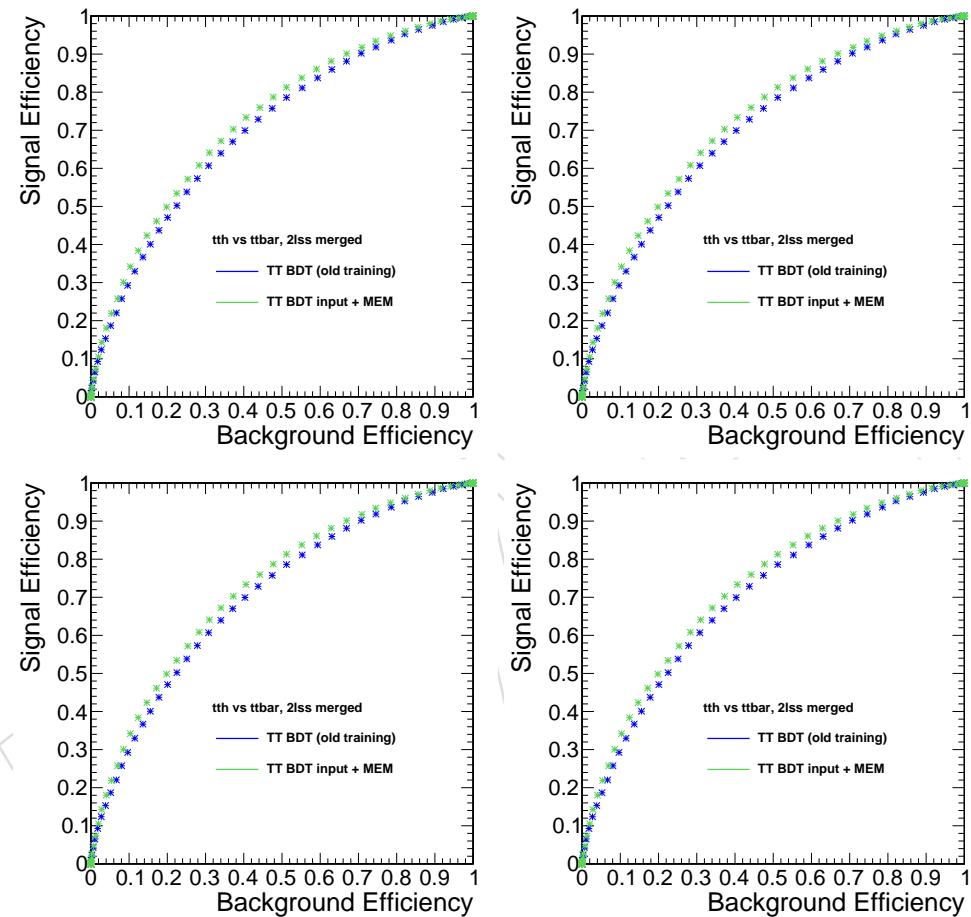


Figure 87: ~~Updated~~ Comparison of TT BDT and new BDT including MEM in 2lss signal region for (a) all events (b) 0 missing jets, (c) 1 missing jets, (d) 2 missing jets.

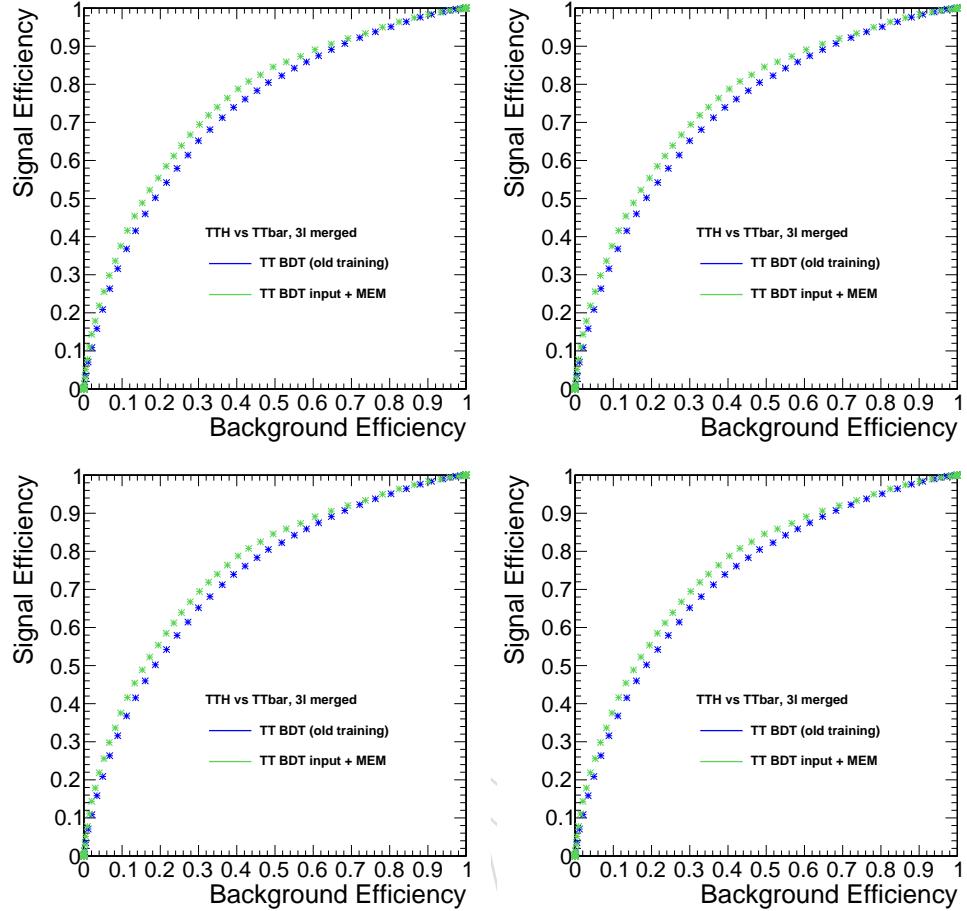


Figure 88: **Updated Comparison** of TT BDT and new BDT including MEM in 3l signal region for (a) all events (b) 0 missing jets, (c) 1 missing jets, (d) 2 missing jets.

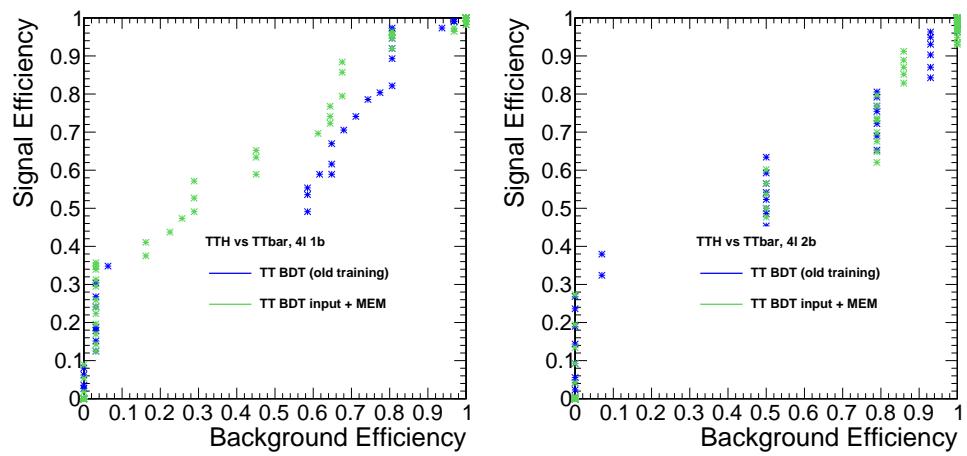


Figure 89: **Updated Comparison** of TT BDT and new BDT including MEM in 4l signal region for (a) 1b events, (b) 2b events

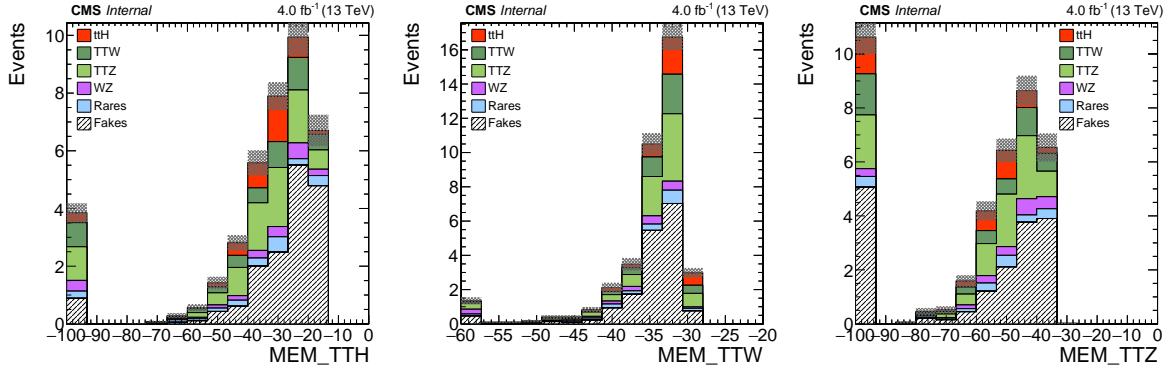


Figure 90: Log of MEM weights in the 3l signal region.

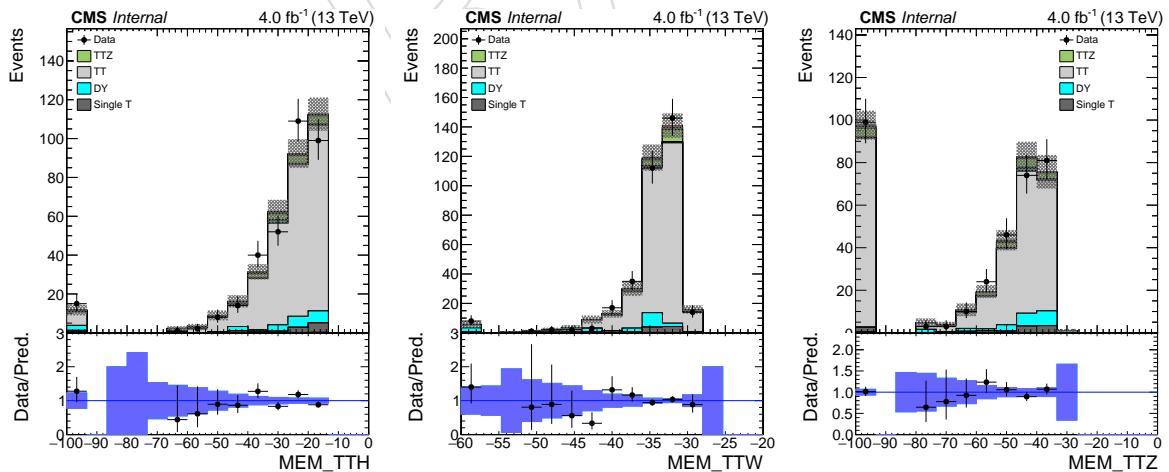


Figure 91: Log of MEM weights in the 3l application region.

1191 C The Hjj tagger

1192 In this section we introduce a discriminator aiming at identifying a dijet pair that originates
 1193 from a Higgs decaying to two Ws. This discriminator represents a possible extension of the
 1194 Hj tagger described in Sec. 5.1.2. We describe it below for completeness, despite we have not
 1195 found any direct improvement by using it in the main analysis.

We target the 2lss category in which the ttH signal decays, in the highest fraction of events, according to the following chain:

$$(t)(t)(H) \rightarrow (bW)(bW)(WW^*) \rightarrow (bjj)(b\ell\nu_\ell)(\ell\nu_\ell jj)$$

1196 We therefore expect 2 b quark jets, 4 jets, 2 same-sign leptons, and missing energy in the final
 1197 state, altough, in order to increase the signal acceptance, the analysis requires the presence of
 1198 at least 4 jets overall. This means that the Higgs jets do not necessarily enter the signal region.

1199 In order to deal with both the complicated jet combinatoric and the possibility that not all the
 1200 jets originating from Higgs are selected, a dedicated tagger is developed: the Higgs-dijet (Hjj)
 1201 tagger. The Hjj tagger is a discriminator that considers geometric and kinematic features in
 1202 order to estimate the likelihood of a jet pair originating again from $H \rightarrow WW^* \rightarrow \ell\nu_\ell jj$.

1203 The discriminator is developed considering the BDT multivariate technique. We rely on the
 1204 powheg ttH sample and ttV sample to define the signal and the background in the training.
 1205 For the Hjj tagger the signal is represented by a reconstructed dijet pair that is matched at gen-
 1206 level to the dijet pair of the process $H \rightarrow WW^* \rightarrow \ell\nu_\ell jj$, while the background is given by
 1207 the reconstructed dijet pairs in the ttV events. For the training, we consider the phase space of
 1208 events that enter the 2lss category, with 0 τ_h .

1209 In the following subsection we list the variables used for the Hjj tagger and their expected BDT
 1210 distributions.

1211 Hjj variables and performance

1212 The variables used for the Hjj tagger are:

- 1213 • sum of the Hj taggers for the two jets
- 1214 • dR of the two jets
- 1215 • minimum dR between the jet pair and another jet
- 1216 • ratio of the minimum and maximum dR between the jet pair and another jet
- 1217 • dijet mass
- 1218 • mass of the dijet plus the closest lepton

1219 The performances of the Hjj tagger are illustrated in Fig. 92.

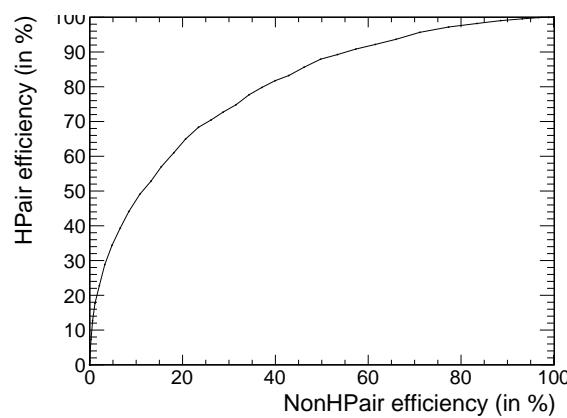


Figure 92: The H_{jj} ROC. Signal and background composition are described in the text.

1220 D Signal Extraction using the k -means algorithm

1221 The two-dimensional plane described in Sec. 5.3 is partitioned in several regions, depending
 1222 on their different signal or background composition, with the objective of maximizing the anal-
 1223 ysis sensitivity with the available luminosity. To do so, a recursive application of the k -means
 1224 algorithm [27, 28] is used.

1225 This algorithm consists of picking k random points in the space: these points are taken to be the
 1226 seed (“centroid”) for clustering the data into k sets. Each of the remaining data points is then
 1227 associated to the closest cluster; common implementations of the algorithm are often applied
 1228 preferentially to either sparse or dense data. For sparse data, and the original implementation
 1229 of the algorithm, each time a point is added to a set, the centroid of that group is recomputed
 1230 by taking into account the new data point. For dense data, all the points are associated to their
 1231 closest cluster. In this analysis, it is assumed that the data are sufficiently dense to permit using
 1232 the latter, which is computationally faster.

1233 The k -means algorithm is considered very stable, and indeed preliminary tests have shown
 1234 that, after fixing k , the algorithm converges to the same sets even when starting from many
 1235 different random initial centroids.

1236 However, the application of the algorithm to the problem of finding an effective binning to
 1237 optimize the sensitivity of the analysis depends on the arbitrary choice of number of clusters.

1238 We therefore devise a recursive version of the algorithm, which works as follows: an initial
 1239 clusterization into two sets ($k = 2$) is performed, and the resulting clusters are passed to the
 1240 next iteration as two independent clusters. The ($k = 2$)-means algorithm is applied to each
 1241 of these clusters, obtaining four subclusters. This procedure can in principle be repeated until
 1242 each cluster consists of a single data point: a meaningful set of clusters is instead characterized
 1243 by the presence of both signal and background, in different proportions in each cluster. A
 1244 stopping criterion is therefore devised, in order to stop the division of any given cluster.

1245 The criterion of choice states that a region will not be subdivided if any of its subregions would
 1246 contain less than 4 $t\bar{t}H$ or less than 3 $t\bar{t}$ expected events: this ensures a reasonable signal and
 1247 background population in each bin, thus protecting from hard fluctuations due to limited statis-
 1248 tics of the samples.

1249 This method is applied separately in the same-sign dileptonic and in the multileptonic inclusive
 1250 signal regions, using two sets of simulated events from the $t\bar{t}H$, $t\bar{t}W$, and $t\bar{t}$ samples.

1251 Figures 94 and 93 show – for the two same sign leptons and multileptonic inclusive regions,
 1252 respectively – the location of the different bins in the two-dimensional plane (left), as well as
 1253 the number of expected events in each bin (right).

1254 Similar results are obtained with an alternative method, in which each subdivision is obtained
 1255 by applying cuts of the type $a\text{MVA}_{t\bar{t}W} + b < \text{MVA}_{t\bar{t}}$, where a and b are parameters chosen
 1256 to optimize some figure-of-merit (FOM). The FOM chosen is $\mathcal{P}(s_1 + b_1 \parallel b_1)\mathcal{P}(s_2 + b_2 \parallel b_2)$,
 1257 where $\mathcal{P}(x \parallel \mu)$ represents the Poisson probability and $s_{1(2)}$, $b_{1(2)}$ are the number of signal and
 1258 background events in each subdivision.

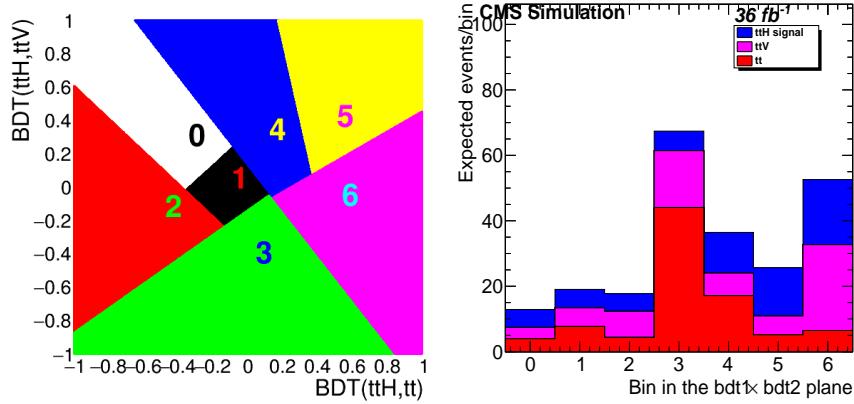


Figure 93: The location of the different bins in the two-dimensional plane (left), as well as the number of expected events in each bin (right) in the two same sign leptons channel, estimated from MC.

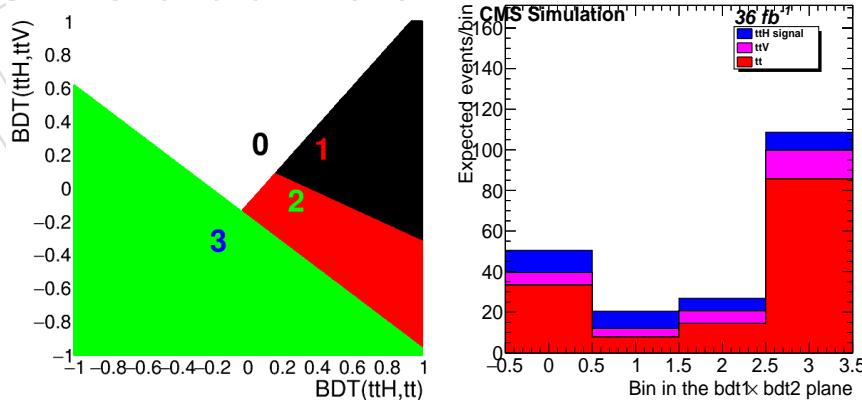


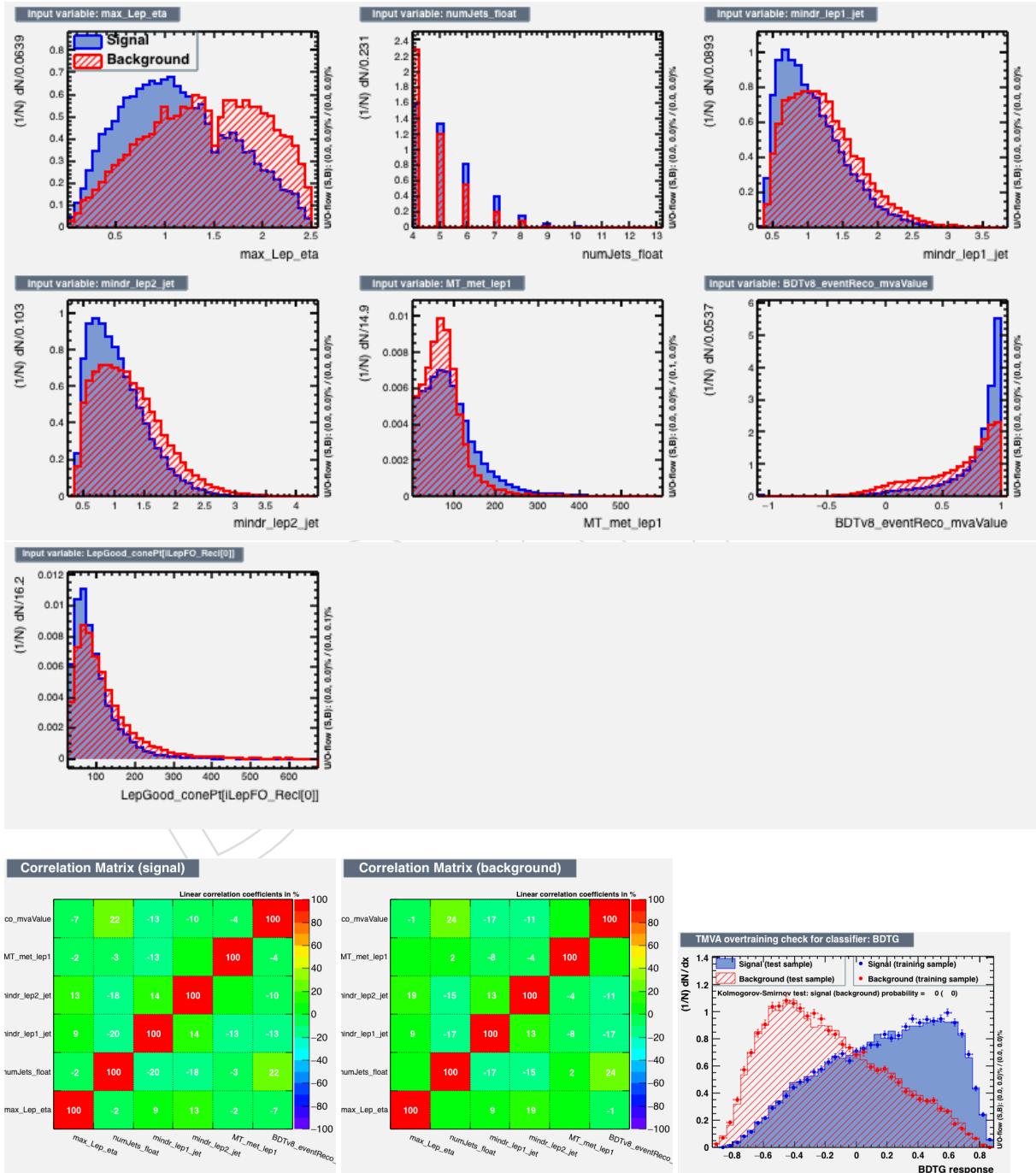
Figure 94: The location of the different bins in the two-dimensional plane (left), as well as the number of expected events in each bin (right) in the three leptons channel, estimated from MC.

1259 **E Training of the kinematic BDT discriminators**

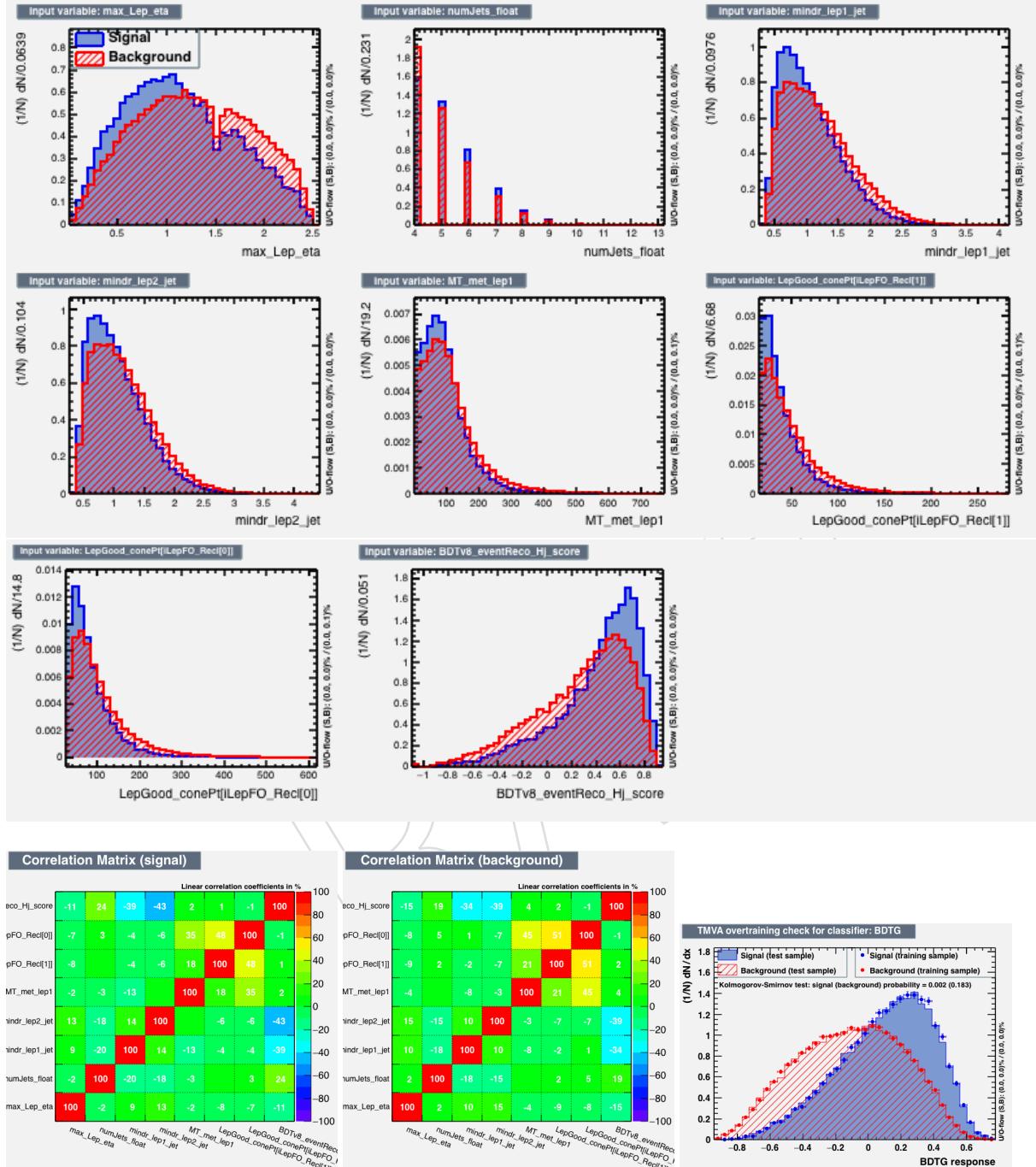
1260 We train two BDTs aiming at separating the signal from the $t\bar{t}V$ $t\bar{t}$ backgrounds respectively,
 1261 using the input variables described in Section 5.

1262 All trainings are performed on samples that are not used anywhere else in the signal extraction
 1263 (Powheg $t\bar{t}H$ leading order $t\bar{t}V$ and pure $t\bar{t}$ MC).

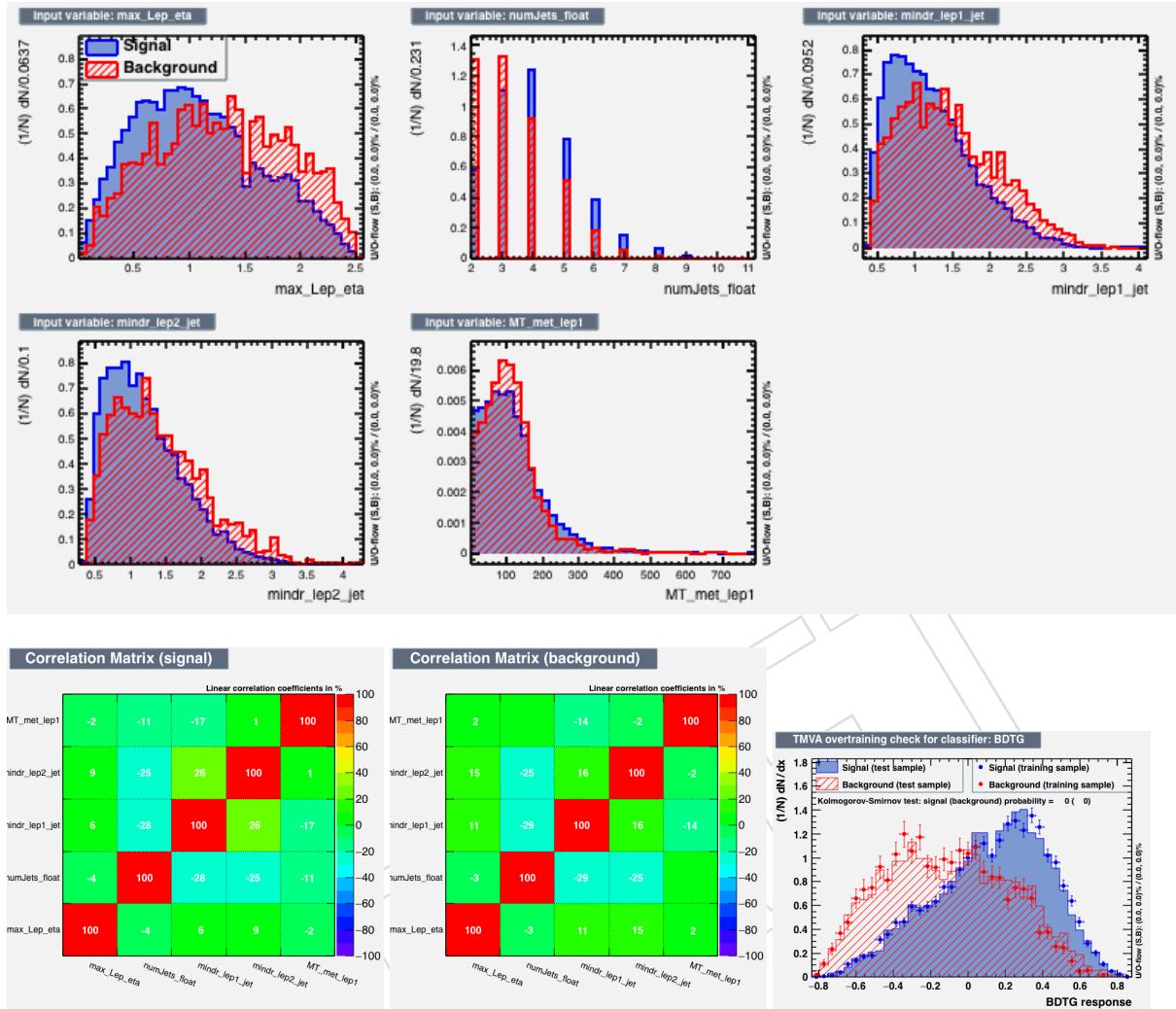
1264 **2lss event category, $t\bar{t}$ training**

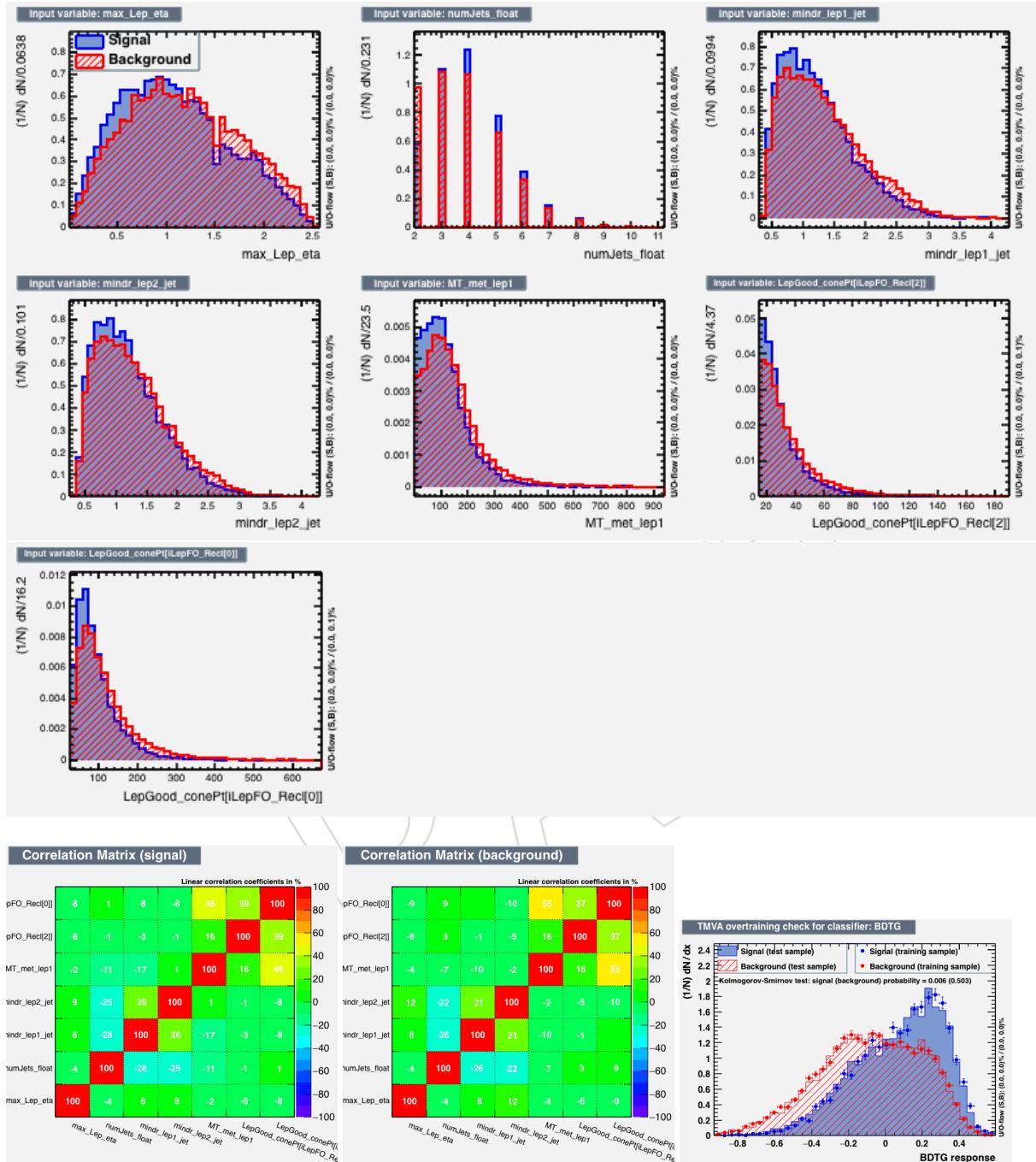


1265 2lss event category, $t\bar{t}V$ training



1266 3I event category, $t\bar{t}$ training



1267 3l event category, $t\bar{t}V$ training

1268 References

- 1269 [1] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS
1270 experiment at the LHC", *Phys. Lett. B* **716** (2012) 30,
1271 [doi:10.1016/j.physletb.2012.08.021](https://doi.org/10.1016/j.physletb.2012.08.021).
- 1272 [2] ATLAS Collaboration, "Observation of a new particle in the search for the Standard
1273 Model Higgs boson with the ATLAS detector at the LHC", *Phys. Lett. B* **716** (2012) 1,
1274 [doi:10.1016/j.physletb.2012.08.020](https://doi.org/10.1016/j.physletb.2012.08.020).
- 1275 [3] ATLAS and CMS Collaboration, "Combined measurement of the Higgs boson mass in pp
1276 collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS experiments", *Phys. Rev. Letter*
1277 **114** (2015) 45, [doi:htpp://dx.doi.org/10.1103/PhysRevLett.114.191803](http://dx.doi.org/10.1103/PhysRevLett.114.191803).
- 1278 [4] ATLAS and CMS Collaboration, "Combined measurement of the Higgs boson mass in pp
1279 collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS experiments", *Eur. Phys. J. C* **76**
1280 (2015) 64, [doi:10.1140/epjc/s10052-015-3769-y](https://doi.org/10.1140/epjc/s10052-015-3769-y).
- 1281 [5] CMS Collaboration, "Precise determination of the mass of the Higgs boson and tests of
1282 compatibility of its couplings with the standard model predictions using proton
1283 collisions at 7 and 8 TeV", *Eur. Phys. J. C* **75** (2015) 212,
1284 [doi:10.1140/epjc/s10052-015-3351-7](https://doi.org/10.1140/epjc/s10052-015-3351-7).
- 1285 [6] LHC Higgs Cross Section Working Group et al., "Handbook of LHC Higgs Cross
1286 Sections: 1. Inclusive Observables", *CERN-2011-002* (CERN, Geneva, 2011)
1287 [arXiv:1101.0593](https://arxiv.org/abs/1101.0593).
- 1288 [7] CMS Collaboration, "Search for the associated production of the Higgs boson with a
1289 top-quark pair", *JHEP* **09** (2015) doi:[10.1007/JHEP09\(2014\)087](https://doi.org/10.1007/JHEP09(2014)087).
- 1290 [8] CMS Collaboration, "Search for the SM Higgs boson produced in association with top
1291 quarks in multilepton final states", CMS Physics Analysis Summary
1292 CMS-PAS-HIG-13-020, 2013.
- 1293 [9] CMS Collaboration, "Search for ttH production in multilepton final states at $\sqrt{s} = 13$
1294 TeV", CMS Physics Analysis Summary CMS-PAS-HIG-15-008, 2015.
- 1295 [10] CMS Collaboration, "Search for associated production of Higgs bosons and top quarks in
1296 multilepton final states at $\sqrt{s} = 13$ TeV", CMS Physics Analysis Summary
1297 CMS-PAS-HIG-16-022, 2016.
- 1298 [11] C. Collaboration, "JSON files for 25ns data (CMS Hypernews)", 2016.
1299 <https://hypernews.cern.ch/HyperNews/CMS/get/physics-validation/2669.html>.
- 1300 [12] CMS Collaboration, "Dilepton trigger and lepton identification efficiencies for the top
1301 quark pair production cross section measurement at 8 TeV in the dilepton decay
1302 channel.", CMS AN AN-12-389, 2012.
- 1303 [13] CMS Collaboration, "Particle-Flow Event Reconstruction in CMS and Performance for
1304 Jets, Taus, and MET", CMS Physics Analysis Summary CMS-PAS-PFT-09-001, 2009.
- 1305 [14] CMS Collaboration, "Commissioning of the Particle-Flow reconstruction in
1306 Minimum-Bias and Jet Events from pp Collisions at 7 TeV", CMS Physics Analysis
1307 Summary CMS-PAS-PFT-10-002, 2010.

- 1308 [15] M. Cacciari, G. P. Salam, G. Soyez, “FastJet user manual”, (2011).
 1309 arXiv:hep-ph/1111.6097v1.
- 1310 [16] M. Cacciari, G. P. Salam, “Dispelling the N^3 myth for the k_t jet-finder”, *Phys. Lett. B* **641**
 1311 (2006) 57, doi:10.1016/j.physletb.2006.08.037, arXiv:hep-ph/0512210.
- 1312 [17] CMS Collaboration, “Determination of Jet Energy Calibration and Transverse Momentum
 1313 Resolution in CMS”, *JINST* **6** (2011) 11002, doi:10.1088/1748-0221/6/11/P11002.
- 1314 [18] CMS Collaboration Collaboration, “Identification of b-quark jets with the CMS
 1315 experiment”, *JINST* **8** (2013) P04013, doi:10.1088/1748-0221/8/04/P04013,
 1316 arXiv:1211.4462.
- 1317 [19] CMS Collaboration, “Performance of CMS muon reconstruction in pp collision events at
 1318 $\sqrt{s} = 7$ TeV”, *JINST* **7** (2012) P10002, doi:10.1088/1748-0221/7/10/P10002,
 1319 arXiv:1206.4071.
- 1320 [20] G. Abbiendi et al., “Baseline muon selections”, October, 2014. https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideMuonId#Tight_Muon.
- 1321
- 1322 [21] G. Petrucciani and C. Botta, “Two step prompt muon identification”, January, 2015.
 1323 <https://indico.cern.ch/event/368007/contribution/2/material/slides/0.pdf>.
- 1324
- 1325 [22] H. Brun and C. Ochando, “Updated Results on MVA eID with 13 TeV samples”, October,
 1326 2014. <https://indico.cern.ch/event/367861/contribution/1/material/slides/0.pdf>.
- 1327
- 1328 [23] A. Le Bihan et al., “Search for ttH in multilepton final states at 13 TeV”, CMS Analysis
 1329 Note CMS AN-15-321, 2015.
- 1330 [24] CMS Collaboration, “Studies of Tracker Material”, *CDS Record* **1279138** (2010).
- 1331 [25] CMS Collaboration, “Reconstruction and identification of τ lepton decays to hadrons and
 1332 ν_τ at CMS”, arXiv:1510.07488.
- 1333 [26] A. Nayak et al., “Multivariate τ -identification algorithms for 13 TeV data”, CMS Analysis
 1334 Note CMS AN-15-310, 2015.
- 1335 [27] H. Steinhaus, “Sur la division des corps matériels en parties”, *Bull. Acad. Polon. Sci. Cl.*
 1336 *III.* **4** (1956) 801–804 (1957).
- 1337 [28] J. MacQueen, “Some methods for classification and analysis of multivariate
 1338 observations”, in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and*
 1339 *Probability, Volume 1: Statistics*, pp. 281–297. University of California Press, Berkeley,
 1340 Calif., 1967.
- 1341 [29] Y. R. . in preparation. <http://indico.cern.ch/event/407347/session/3/contribution/10/attachments/1211342/1766869/hxswg16.pdf>.
- 1342
- 1343 [30] P. production of low m_{ll} ttZ sample.
 1344 <https://www.dropbox.com/s/fni1334aaa9cbfn/ttzlowmll.pdf?dl=0>.
- 1345 [31] C. Botta et al., “Search for the standard model Higgs boson produced in association with
 1346 top quarks and decaying to leptons”, CMS Analysis Note CMS AN-13-159, 2013.

- 1347 [32] G. Lepage, “A new algorithm for adaptive multidimensional integration”, *Journal of*
1348 *Computational Physics* **27** (1978) 192–203.
- 1349 [33] A. Buckley et al., “LHAPDF6: parton density access in the LHC precision era”, *Eur. Phys.*
1350 *J.* **C75** (2015) 132, doi:10.1140/epjc/s10052-015-3318-8, arXiv:1412.7420.
- 1351 [34] P. Artoisenet, V. Lemaitre, F. Maltoni, and O. Mattelaer, “Automation of the matrix
1352 element reweighting method”, *JHEP* **12** (2010) 068,
1353 doi:10.1007/JHEP12(2010)068, arXiv:1007.3300.

DRAFT