Available on the CERN CDS information server
CMS PAS HIG-16-038

CMS Physics Analysis Summary

Contact: cms-pag-conveners-higgs@cern.ch
2016/11/09

Search for $\mathrm{t\bar{t}H}$ production in the $\mathrm{H} \to \mathrm{b\bar{b}}$ decay channel with 2016 pp collision data at $\sqrt{s} = 13$ TeV

The CMS Collaboration

Abstract

The results of the search for the associated production of a Higgs boson with a top quark-antiquark pair ($\mathrm{t\bar{t}H}$) in proton-proton collisions at a center-of-mass energy of $\sqrt{s} = 13\,\mathrm{TeV}$ are presented. The data correspond to an integrated luminosity of up to 12.9 fb^{-1} recorded with the CMS experiment in 2016. Candidate $\mathrm{t\bar{t}H}$ events are selected with criteria enhancing the lepton+jets or dilepton decay-channels of the $\mathrm{t\bar{t}}$ system and the decay of the Higgs boson into a bottom quark-antiquark pair ($\mathrm{H} \to \mathrm{b\bar{b}}$). In order to increase the sensitivity of the search, selected events are split into several categories with different expected signal and background rates. In each category signal and background events are separated using a multivariate approach that combines a matrix element method with boosted decision trees. The results are characterized by an observed $\mathrm{t\bar{t}H}$ signal strength relative to the standard model cross section, $\mu = \sigma/\sigma_{\mathrm{SM}}$, under the assumption of $m_H = 125\,\mathrm{GeV}$. A combined fit of multivariate discriminant distributions in all categories results in an observed (expected) upper limit of $\mu < 1.5$ (1.7) at the 95% confidence level, and a best fit value of $\mu = -0.19\,^{+0.45}_{-0.44}(\mathrm{stat.})\,^{+0.66}_{-0.68}(\mathrm{syst.})$.

1 Introduction

The observation of a Higgs boson with a mass of approximately 125 GeV [1, 2] at the Large Hadron Collider (LHC) marked the starting point of a broad experimental program to determine the properties of the newly discovered particle. To date, the results of all measurements performed at the LHC are consistent with the expectations for a standard model (SM) Higgs boson. Decays into $\gamma\gamma$, ZZ and WW final states have been observed and there is evidence for the direct decay of the particle to fermions from the $\tau\tau$ and $\mathrm{b\bar{b}}$ decay channels [3, 4]. The measured rates of various production and decay channels agree with the SM expectations [5, 6] and the hypothesis of a spin-0 particle is favored over other hypotheses [7, 8].

In the SM the coupling of the Higgs boson to fermions is of Yukawa type, with a coupling strength proportional to the fermion mass. Probing the coupling of the Higgs boson to the heaviest known fermion, the top quark, is hence very important for testing the SM and for constraining models of physics beyond the SM (BSM). Indirect constraints on the top–Higgs coupling are available from processes including top-quark loops, for example Higgs boson production through gluon-gluon fusion [5, 6]. On the other hand, the associated production of a Higgs boson and a top quark-antiquark pair ($\mathrm{t\bar{t}H}$ production) is a direct probe of the top–Higgs coupling, as illustrated by the Feynman diagrams in Fig. 1. If observed it would prove the coupling of the Higgs boson to fermions with weak isospin $+1/2$ ("up-type") in addition to couplings to τ and b, which carry a weak isospin of $-1/2$ ("down-type"). The Higgs boson decay into bottom quark-antiquark pairs ($\mathrm{b\bar{b}}$), also shown in Fig. 1, is attractive as a final state because it features the largest branching fraction of 0.58 ± 0.02 for a 125 GeV Higgs boson [9].

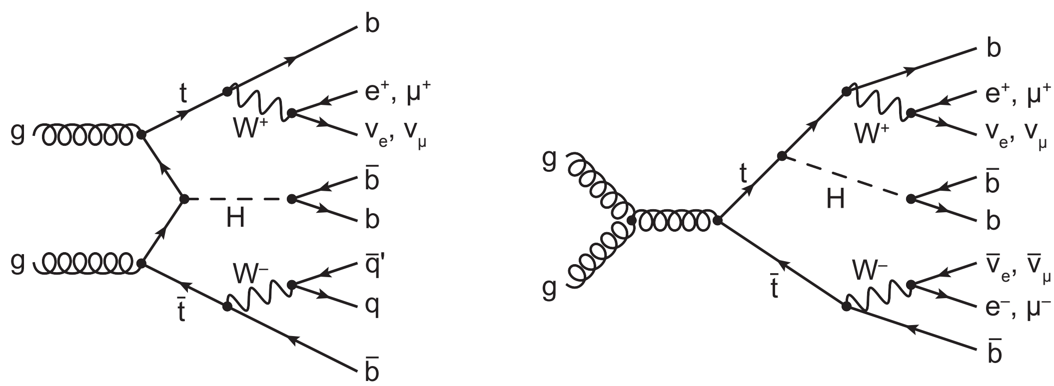

Figure 1: Exemplary leading-order Feynman diagrams for $\mathrm{t\bar{t}H}$ production, including the subsequent decays of the top quark-antiquark pair in the lepton+jets channel (left) and the dilepton channel (right) as well as the decay of the Higgs boson into a bottom quark-antiquark pair.

Several BSM physics models predict a significantly enhanced $\mathrm{t\bar{t}H}$ production rate while not modifying the branching fractions of Higgs boson decays by a measurable amount. For example, a number of BSM physics models predict vector-like partners of the top-quark (T) that decay into tH, bW and tZ final states [10–19]. The production and decay of $\mathrm{T\bar{T}}$ pairs would lead to final states indistinguishable from those of $\mathrm{t\bar{t}H}$ production. In this context, a measurement of the $\mathrm{t\bar{t}H}$ production cross section has the potential to distinguish the SM Higgs mechanism from alternative mechanisms to generate fermion mass.

Various dedicated searches for $\mathrm{t\bar{t}H}$ production have been conducted during Run I of the LHC. The CMS searches employ pp collision data corresponding to an integrated luminosity of $5\,\mathrm{fb}^{-1}$ at a center-of-mass energy of $\sqrt{s} = 7$ TeV and $19.5\,\mathrm{fb}^{-1}$ at $\sqrt{s} = 8$ TeV. These searches have been performed studying Higgs boson decays to hadrons, photons, and leptons using multivariate analysis (MVA) techniques, showing a mild excess of the observed $t\bar{t}H$ cross section relative to the SM expectation of $\mu = \sigma/\sigma_{SM} = 2.8 \pm 1.0$ [20]. A similar excess of $\mu = 2.1^{+1.4}_{-1.2}$ is observed in a search for $t\bar{t}H$ production in multilepton final states with 20.3 fb^{-1} of ATLAS data at $\sqrt{s} = 8$ TeV [21].

The CMS search results have also entered a comprehensive test of the compatibility of the Higgs boson couplings with SM predictions [6]. The sensitivity for the $t\bar{t}H$ process in the $H \to b\bar{b}$ decay channel was further increased by employing the matrix element method (MEM) [22], resulting in an observed (expected) upper limit of $\mu < 4.2\,(3.3)$ at 95% confidence level [23]. ATLAS obtained an observed (expected) upper limit on $t\bar{t}H$ production in the $H \to b\bar{b}$ decay channel of $\mu < 3.4\,(2.2)$ using 20.3 fb^{-1} of pp collision data at $\sqrt{s} = 8$ TeV [24].

Observation of $t\bar{t}H$ production is one of the major goals in Higgs boson physics for the LHC Run II. The increased center-of-mass energy of $\sqrt{s} = 13$ TeV results in a $t\bar{t}H$ production cross section 3.9 times larger than at $\sqrt{s} = 8$ TeV based on next-to-leading (NLO) calculations, while the cross section for the most important background, $t\bar{t}$ production, is only increased by a factor of 3.3 [25], resulting in a more favorable signal-to-background ratio. Latest CMS searches for $t\bar{t}H$ production with Run-II data in the diphoton and multilepton final-states of the Higgs boson observe signal strengths of $\mu = 1.9^{+1.5}_{-1.2}$ [26] and $\mu = 2.0^{+0.8}_{-0.7}$ [27], respectively. ATLAS finds $\mu = -0.3^{+1.3}_{-1.0}$ [28] and $\mu = 2.5^{+1.3}_{-1.1}$ [29], respectively.

In the $H \to b\bar{b}$ final state, CMS has performed a search for $t\bar{t}H$ production using 2.7 fb^{-1} of data recorded in 2015 [30]. Analysis methods established in Run I have been significantly improved, and novel methods have been added. In particular, the two multivariate techniques – namely MEM and boosted decision trees (BDT) [31–35] – using different information from each event and aiming at separating different background components, have both been employed to obtain a better performance than using one technique alone. The analysis obtains a value of $\mu = -2.0^{+1.8}_{-1.8}$. A new result by ATLAS in $H \to b\bar{b}$ channel, based on 13.2 fb^{-1}, measures $\mu = 2.1^{+1.0}_{-0.9}$ [29].

This document summarizes a search for the $t\bar{t}H$ production in the $H \to b\bar{b}$ final state performed with up to 12.9 fb^{-1} of data recorded with the CMS detector in 2016. It is an update of the above mentioned search [30]. One improvement is due to a refined Monte Carlo (MC) modeling, leading to a more accurate description of the data, in particular the jet-multiplicity spectrum. The event selection is adapted to $t\bar{t}H$ events with the decay of the Higgs boson into a $b\bar{b}$ pair and lepton+jets as well as dilepton decays of the $t\bar{t}$ pair, resulting in the final state $\ell\nu\, q\bar{q}'\, b\bar{b}$ ($\ell^+\nu\, \ell^-\bar{\nu}\, b\bar{b}$) for lepton+jets (dilepton) $t\bar{t}$ decays, where $\ell = e, \mu$. Events are split into mutually exclusive categories according to the number of reconstructed jets and the number of jets identified as coming from the hadronization of b quarks (b tagging). In each category, signal and background processes are separated employing BDTs which use the kinematic properties of jets and charged leptons, the b tagging probability, invariant masses and angular correlations of combinations of jets and leptons, as well as observables characterizing the event shape as inputs. Those categories with high number of jets or number of b-tagged jets are further subdivided into two sub-categories depending on the BDT discriminant output, and in each sub-category the MEM discriminant is used as final discriminant. From a combined profile-likelihood fit of the final discriminant output distributions to data in all categories, a best-fit value of the signal-strength modifier μ is obtained. In absence of a signal, an upper limit on μ is set.

This document is structured as follows: in Section 3, the data samples and MC simulated samples are described. The basic selection of analysis objects and events is discussed in Section 4. The general analysis strategy and background estimation methods are introduced in Section 5. The effect of systematic uncertainties is studied in Section 6. Results of the studies are presented in Section 7, followed by conclusions in Section 8.

2 The CMS Detector

The central feature of the CMS apparatus is a superconducting solenoid of 6 m internal diameter, providing a magnetic field of 3.8 T. Within the superconducting solenoid volume are a silicon pixel and strip tracker, a lead tungstate crystal electromagnetic calorimeter (ECAL), and a brass and scintillator hadron calorimeter (HCAL), each composed of a barrel and two endcap sections. Forward calorimeters extend the pseudorapidity coverage provided by the barrel and endcap detectors. Muons are measured in gas-ionization detectors embedded in the steel flux-return yoke outside the solenoid. A more detailed description of the CMS detector, together with a definition of the coordinate system used and the relevant kinematic variables, can be found in Ref. [36].

3 Data and Simulation Samples

This analysis is performed using proton-proton (pp) collision data at a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$, which were collected with the CMS detector in 2016 and correspond to a total integrated luminosity of $12.9\,\text{fb}^{-1}$ and $11.4 - 12.9\,\text{fb}^{-1}$ for the lepton+jets and dilepton channels, respectively. The different (lower) luminosity in the dilepton channel only affects the e^+e^- channel and is due to disabled trigger paths during parts of the data taking.

MC event generators, interfaced with a detailed detector simulation, are used to model experimental effects, such as reconstruction and selection efficiencies, as well as detector resolutions. The CMS detector response is simulated using GEANT4 (v. 9.4) [37].

For the simulation of the reference $t\bar{t}H$ signal sample, the next-to-leading-order (NLO) event generator POWHEG (v. 2) [38, 39] is used. The value of the Higgs boson mass is assumed to be 125 GeV, while the top quark mass value is set to 172.5 GeV. The proton structure is described by the parton distribution functions (PDF) NNPDF3.0 [40]. The generated events are subsequently processed with PYTHIA (v. 8.2) [41] for parton showering and hadronization.

Standard model backgrounds are simulated using POWHEG, MG5_aMC@NLO (v. 2.2.2) [42], or PYTHIA, depending on the process. The main background contribution originates from $t\bar{t}$ production, the production of W and Z/γ^* bosons with additional jets (referred to as W+jets and Z+jets or commonly as V+jets in the following), single top quark production (tW channel), and diboson (WW, WZ, and ZZ) processes, and $t\bar{t}$ production in association with a W or Z boson (referred to as $t\bar{t}$+W and $t\bar{t}$+Z or commonly as $t\bar{t}$+V in the following). Both the $t\bar{t}$ and the single top quark samples are simulated with POWHEG. The V+jets and $t\bar{t}$+V samples are simulated with the NLO generator MG5_aMC@NLO, where for the V+jets samples the matching of matrix-element jets to parton showers is performed using the FXFX [43] prescription. PYTHIA is used to simulate diboson events. Parton showering and hadronization are also simulated with PYTHIA in all the background samples. The PYTHIA CUETP8M1 tune [44, 45] was used in the past to characterize the underlying event in both the $t\bar{t}H$ signal and the background samples. However, our $t\bar{t}H$ search based on the 2015 data has shown that this tune results in a much harder jet-multiplicity spectrum in simulation with respect to data. Since $t\bar{t}H$ events typically show a high jet multiplicity, this data-simulation discrepancy will have a big impact on our search. In order to improve the modeling, a custom tune has been derived by CMS. For this new tune, the (mostly uncorrelated) parameters α_S^{ISR} and h_{damp} are optimized based on several measurements at $\sqrt{s} = 8\,\mathrm{TeV}$. The latter is the parameter that controls the matrix element and parton shower matching in POWHEG and effectively regulates the high-p_{T} radiation. Validation studies show that this new tune significantly improves the jet-multiplicity modeling. For this analysis, we used the $\mathrm{t\bar{t}H}$ signal and $\mathrm{t\bar{t}}$ background samples produced with this new tune.

For comparison with the measured distributions, the events in the simulated samples are normalized to the same integrated luminosity of the data according to their predicted cross sections. These are taken from theoretical calculations at next-to-next-to-leading order (NNLO, for V+jets production), approximate NNLO (single top quark tW channel [46]), and NLO (diboson production [47] and $\mathrm{t\bar{t}}$+V production [48]). The $\mathrm{t\bar{t}H}$ cross section [25, 49–52] and Higgs boson branching fractions [53–56] used in the analysis also have NLO accuracy. The $\mathrm{t\bar{t}}$ simulated sample is normalized to the full NNLO calculation with resummation to next-to-next-to-leading-logarithmic (NNLL) accuracy [57–63], assuming a top quark mass value of 172.5 GeV and using the NNPDF3.0 PDF set. This sample is further separated into the following processes based on the flavor of additional jets that do not originate from the top quark decays in the event: $\mathrm{t\bar{t}+b\bar{b}}$, defined at generator level as the events in which two additional b jets are generated within the acceptance requirements (cf. Section 4) and originate from one or more B hadrons; $\mathrm{t\bar{t}}$+b, for which only one additional b jet originates from a single B hadron; $\mathrm{t\bar{t}}$+2b, which corresponds to events with two additional B hadrons that are close enough in direction to produce a single b jet; $\mathrm{t\bar{t}+c\bar{c}}$, for which events have at least one c jet within acceptance and no additional b jets; $\mathrm{t\bar{t}}$ + light flavor ($\mathrm{t\bar{t}}$+LF), which correspond to events that do not belong to any of the above processes. The separation is motivated by the fact that different sub-samples originate from different physics processes and have different systematic uncertainties. A similar separation strategy has been followed by the ATLAS collaboration [24].

Effects from additional pp interactions in the same bunch crossing (pileup) are modeled by adding simulated minimum-bias events (generated with PYTHIA) to all simulated processes. The pileup multiplicity distribution in simulation is reweighted to reflect the luminosity profile of the observed pp collisions. Correction factors described in Section 4 are applied where necessary to improve the description of the data by the simulation.

4 Object and Event Selection

The event selection aims at selecting events from the production of a Higgs boson in association with a top quark-antiquark pair, where only the case in which the Higgs boson decays into a bottom quark-antiquark pair is considered. In the SM, the top quark is expected to decay into a W boson and a b quark nearly 100% of the time. Hence different $\mathrm{t\bar{t}}$ decay modes can be identified according to the subsequent decays of the W bosons. Two $\mathrm{t\bar{t}}$ decay modes are considered: the lepton+jets mode ($\mathrm{t\bar{t}} \to \ell\nu\,\mathrm{q\bar{q}'}\,\mathrm{b\bar{b}}$), where one W boson decays into a charged lepton and a neutrino, and the dilepton mode ($\mathrm{t\bar{t}} \to \ell^+\nu\,\ell^-\bar{\nu}\,\mathrm{b\bar{b}}$), where both W bosons decay into a charged lepton and a neutrino. These signatures imply the presence of isolated leptons ($\ell = \mathrm{e}, \mu$), missing transverse momentum owing to the neutrinos from W boson decays, and highly energetic jets originating from the final-state quarks. The heavy-quark content of the jets is identified through b tagging techniques.

At trigger level, events in the lepton+jets channel are required to contain an electron (muon) with transverse momentum (p_{T}) threshold of $p_{\mathrm{T}} > 27\,\mathrm{GeV}$ ($p_{\mathrm{T}} > 22\,\mathrm{GeV}$). For electrons a pseudorapidity range of $|\eta| < 2.1$ is required. Events in the dilepton channel are required to contain two leptons fulfilling the requirement of p_T thresholds between 8 GeV and 23 GeV and isolation criteria.

Events are reconstructed using a particle-flow (PF) technique [64, 65], which combines signals from all sub-detectors to enhance the reconstruction performance by identifying individual particle candidates in pp collisions. Charged hadrons from pileup events are omitted in the subsequent event reconstruction.

The electron and muon candidates are required to be sufficiently isolated from nearby jet activity as follows. For each electron (muon) candidate, a cone of $\Delta R = 0.3$ ($\Delta R = 0.4$) is constructed around the direction of the track at the event vertex, where ΔR is defined as $\sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$, and $\Delta\eta$ and $\Delta\phi$ are the distances in pseudorapidity and azimuthal angle. Excluding the contribution from the lepton candidate, the scalar sum of the p_T of all particle candidates inside the cone consistent with arising from the chosen primary event vertex is calculated. The neutral component from pileup events is subtracted event-by-event based on the average transverse energy deposited by neutral particles in the event, which is removed from the transverse energy in the isolation cone. A relative isolation discriminant, I_{rel}, is defined as the ratio of this sum to the p_T of the lepton candidate. Electron candidates are selected if they have values of $I_{rel} < 0.15$, while muons are selected if they fulfill the requirement of $I_{rel} < 0.15$ in the lepton+jets channel and $I_{rel} < 0.25$ in the dilepton channel. In addition, electrons from identified photon conversions are rejected. To further increase the purity of muons originating from the primary interaction and to suppress misidentified muons or muons from decay-in-flight processes, additional quality criteria, such as a minimal number of hits associated with the muon track, are required in both the silicon tracker and the muon system.

For the lepton+jets channel, events are selected containing exactly one energetic, isolated lepton (e or μ), which is required to have $p_T > 25$ GeV or $p_T > 30$ GeV in the case of the μ or e, respectively, and $|\eta| < 2.1$ (but excluding electrons within a small region of $|\eta|$ between the barrel and endcap sections of the ECAL). For the dilepton channel, events are required to have a pair of oppositely charged energetic leptons (e^+e^-, $\mu^+\mu^-$, $\mu^\pm e^\mp$). The leading lepton is required to have $p_T > 25$ GeV and the subleading lepton $p_T > 15$ GeV, and both leptons are required to fulfill the requirement of $|\eta| < 2.4$. The invariant mass of the selected lepton pair is required to be larger than 20 GeV to suppress events from heavy-flavor resonance decays and low-mass Drell-Yan processes. In the same-flavor channels, events are rejected if the dilepton invariant mass is within the region $76\,\text{GeV} < m^{\ell\ell} < 106\,\text{GeV}$, thereby suppressing further contribution from Z+jets processes.

Jets are reconstructed from the PF particle candidates using the anti-k_T clustering algorithm [66] implemented in FastJet [67] with a distance parameter of 0.4. The jet energy is corrected for the remaining neutral-hadron pileup component in a manner similar to that used to find the energy within the lepton isolation cone [68]. Jet energy corrections are also applied as a function of jet p_T and η [69] to data and simulation. Events in the lepton+jets channel are required to have at least four reconstructed jets with $p_T > 30$ GeV and $|\eta| < 2.4$. In the dilepton channels, at least three jets with $p_T > 20$ GeV and $|\eta| < 2.4$ are required, from which the two leading jets must satisfy $p_T > 30$ GeV.

Jets originating from the hadronization of b quarks are identified using a combined secondary vertex algorithm (CSVv2) [70], which provides a b tagging discriminant by combining identified secondary vertices and track-based lifetime information. A discriminant value is chosen such that the probability of tagging jets originating from light-flavor quarks (u, d, or s) or gluons is around 1%, and the corresponding efficiency for tagging jets from b (c) quarks is ≈70% (20%). The shape of the CSVv2 discriminant distribution in simulation is corrected by scale factors to better describe the jet CSVv2 shape observed in the data [71]. This correction is derived separately for light-flavor and b jets from a "tag-and-probe" approach using control samples enriched in events with a Z boson and exactly two jets, and $\mathrm{t\bar{t}}$ events with no additional jets.

The missing transverse momentum vector $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ is defined as the projection of the negative vector sum of the momenta of all reconstructed particles in an event on the plane perpendicular to the beams. Its magnitude is referred to as $E_{\mathrm{T}}^{\mathrm{miss}}$. In the dilepton same-flavor channels, events are required to fulfill the requirement of $E_{\mathrm{T}}^{\mathrm{miss}} > 40\,\mathrm{GeV}$.

Events from $\mathrm{t\bar{t}H}$ are generally characterized by having more jets and more b-tags than the background processes. Events are divided into categories based on the number of jets and the number of b-tagged jets. For the lepton+jets channel, events are separated into the following four categories: ≥ 6 jets, 3 b-tags; 4 jets, 4 b-tags; 5 jets, ≥ 4 b-tags and ≥ 6 jets, ≥ 4 b-tags. For the dilepton channel, events are divided into three categories: 3 jets, 3 b-tags; ≥ 4 jets, 3 b-tags and ≥ 4 jets, ≥ 4 b-tags.

Tables 1 and 2 show the predicted and observed event yields after the event selection in the lepton+jets and dilepton channels, respectively. The tables are sub-divided into the different jet and b-tag categories used in each channel. The expected and observed yields agree well in all final states across the different categories of jets and b-tags.

Table 1: $\mathrm{t\bar{t}H}$ and background event yields for lepton+jets categories. The processes and the separation of the $\mathrm{t\bar{t}}$ +jets sample are described in Section 3. The uncertainties in the expected yields include the statistical as well as all the systematic contributions. Cases where no events pass the event selection are marked as "—".

Process	≥ 6 jets, 3 b-tags	4 jets, 4 b-tags	5 jets, ≥ 4 b-tags	≥ 6 jets, ≥ 4 b-tags
$\mathrm{t\bar{t}}$+LF	2710.6 ± 1027.1	91.4 ± 69.9	96.9 ± 75.8	86.2 ± 71.9
$\mathrm{t\bar{t}}$+$\mathrm{c\bar{c}}$	1771.9 ± 1099.0	59.3 ± 47.0	112.5 ± 88.9	175.8 ± 141.6
$\mathrm{t\bar{t}}$+b	717.6 ± 406.8	37.6 ± 23.1	69.6 ± 42.4	86.3 ± 57.0
$\mathrm{t\bar{t}}$+2b	401.4 ± 237.1	16.0 ± 9.9	33.8 ± 20.6	54.9 ± 37.3
$\mathrm{t\bar{t}}$+$\mathrm{b\bar{b}}$	777.0 ± 468.4	42.2 ± 23.8	126.5 ± 71.6	306.3 ± 183.2
Single t	331.9 ± 77.2	19.2 ± 7.8	29.5 ± 11.4	30.0 ± 10.8
V+jets	79.5 ± 30.9	5.7 ± 6.8	7.7 ± 5.1	7.2 ± 3.9
$\mathrm{t\bar{t}}$+V	81.4 ± 17.0	2.2 ± 0.6	6.3 ± 1.7	18.1 ± 5.2
Diboson	4.3 ± 2.3	0.5 ± 0.5	—	—
Total bkg.	6875.6 ± 2776.8	274.1 ± 153.0	482.9 ± 239.7	764.8 ± 365.3
$\mathrm{t\bar{t}H}$	74.2 ± 9.7	4.0 ± 0.8	11.7 ± 2.2	26.9 ± 5.6
Data	6811	376	551	787

5 Analysis Strategy and Background Estimation

The BDT and MEM techniques are used to further improve the signal-to-background separation in both lepton+jets and dilepton channels of the analysis. The information of both techniques are used to derive a single discriminant, this way exploiting the strength of both methods.

The BDTs utilize information related to object kinematics, event shape, and the jet CSVv2 b-tag discriminant. A separate BDT is trained for each category, resulting in four BDTs in the lepton+jets and three in the dilepton channel. The training is performed using simulated $\mathrm{t\bar{t}H}$

Table 2: $t\bar{t}H$ and background event yields for dilepton categories. The processes and the separation of the $t\bar{t}$ +jets sample are described in Section 3. The uncertainties in the expected yields include the statistical as well as all the systematic contributions. Cases where no events pass the event selection are marked as "—".

Process	3 jets, 3 b-tags	$\geq$ 4 jets, 3 b-tags	$\geq$ 4 jets, $\geq$ 4 b-tags
$t\bar{t}$+LF	179.0 ± 68.7	390.1 ± 167.9	7.6 ± 3.6
$t\bar{t}$+$c\bar{c}$	117.5 ± 73.8	382.6 ± 237.7	19.4 ± 15.5
$t\bar{t}$+b	94.2 ± 51.9	228.0 ± 127.7	14.4 ± 9.2
$t\bar{t}$+2b	31.7 ± 17.3	99.1 ± 54.3	6.2 ± 3.8
$t\bar{t}$+$b\bar{b}$	17.1 ± 9.4	172.5 ± 92.9	57.9 ± 32.6
Single t	16.0 ± 4.6	38.4 ± 11.9	2.4 ± 1.2
V+jets	1.6 ± 2.2	1.6 ± 4.1	0.7 ± 0.5
$t\bar{t}$+V	1.4 ± 0.5	16.6 ± 3.4	2.6 ± 0.8
Diboson	—	0.4 ± 0.4	—
Total bkg.	458.3 ± 197.0	1329.3 ± 503.1	111.2 ± 49.8
$t\bar{t}H$	1.8 ± 0.4	16.5 ± 3.4	4.4 ± 1.3
Data	498	1469	146

and $t\bar{t}$+jets events as signal and background, respectively, which are weighted to achieve equal yields of signal and background events in each category. In order to avoid a biased performance estimate, the signal and background events are split in half: one half is used to perform the training, and the other half is used in the final analysis to monitor the performance and derive the final limits. The specific BDT boosting method used is the stochastic gradient boost [31, 72], available as part of the TMVA package [33] in ROOT. The choice of BDT input variables as well as the tree architecture are optimized separately in each category with a procedure based on the particle swarm algorithm [73]. A description of the input variables is provided in Appendices A and B for each category of the lepton+jets and the dilepton channel, respectively.

Within the MEM, each event is assigned a probability density value computed from the four-momenta of the reconstructed particles, which is based on the differential cross section of the signal or background process. The MEM discriminant is constructed as ratio of the probability density values of the signal and background hypothesis. The deployed algorithm is an improved version of the method described in [23]. The probability density functions are constructed at LO, assuming gluon-gluon fusion production both for signal and background processes. The $t\bar{t} + b\bar{b}$ matrix elements have been found to provide comparable discrimination power against all background subprocesses and are solely used to model the background. Hadronization and detector effects are taken into account via transfer functions derived from simulation, which map the measured four-momenta to the final-state particles in the matrix element. In each event, the four jets that most likely originate from b quarks are considered explicitly as candidates for the b-quarks from the decay of the Higgs boson and the top quark, whereas light jets, if present, are permuted over as the candidates for the light quarks from the hadronic decay of the W-boson. All permutations are considered when associating the b-like jets to top quark or Higgs boson decays in the matrix element, similarly we permute over up to 4 additional light jets for the W decay candidates. The four b-like jets are selected using the likelihood ratio between the hypotheses that four or two jets in the event arose from b quarks and the rest from light quarks, based on the expected b tagging discriminant probability densities from simulation.

The BDT and MEM discriminants perform differently in terms of signal and background separation. While the BDT achieves a slightly better separation against the inclusive $t\bar{t}$ background, the MEM is by construction especially powerful in separating against the challenging $t\bar{t} + b\bar{b}$ background. The correlation between the BDT and MEM discriminants have been studied in different control regions in data and found to be well-modeled by the simulation. In this analysis, the two discriminants are utilized with the scheme described below. This results in the best sensitivity, and it is robust against effects due to the binning of the templates and overoptimization of the multivariate discriminants.

In the dilepton 3 jets, 3 b-tags category, a BDT output distribution is used as final discriminant that enters the fit. This category contains a relatively large number of events, which is a desirable situation for training the BDT. In the lepton+jets categories, as well as dilepton $\geq$ 4 jets, 3 b-tags and $\geq$ 4 jets, $\geq$ 4 b-tags categories, events are further separated into two sub-categories, one with low (background-like) and one with high (signal-like) BDT output, divided by the median of the BDT output distribution for simulated signal events. In each sub-category, the MEM is used as final discriminant. The high BDT output sub-category is expected to be enhanced with signal events, and the MEM discriminant achieves additional separation against the residual $t\bar{t} + b\bar{b}$ background contributions. The choice of the median contributes to a robust result by ensuring a sufficient number of events in each sub-category. Including the low BDT output sub-category constrains the background contributions and systematic uncertainties for each of the different event topologies.

The final discriminant outputs provide better discrimination between signal and background than any of the input variables individually. Utilizing both the BDT and MEM information also leads to better signal and background separation than using BDT-only or MEM-only information. The output distributions of the background and signal processes are fit to the data simultaneously in all channels and categories to set limits on the Higgs boson production cross section, as described in Section 7.

6 Systematic Uncertainties

In Table 3, all sources of systematic uncertainties considered in the analysis are listed. They affect either the yields of the signal or background processes, or the discriminant shape, or both. In the last case, the yield and shape effects are treated as entirely correlated and are varied simultaneously. The uncertainties are taken into account via nuisance parameters in the final fit procedure described in Section 7.

The effect of the uncertainties is evaluated individually in each category of each analysis channel, where the effects from the same source are treated as fully correlated. The impact of the systematic variations differs among the categories. As an example, the change in background and signal event yield due to the different uncertainties is listed in Table 4 for the $\geq$ 6 jets, 3 b-tags category of the lepton+jets channel, which shows high sensitivity and at the same time contains a relatively large number of events such that the variations are statistically significant.

The uncertainty in the luminosity estimate is 6.2% [74]. Electron and muon identification and trigger efficiency uncertainties were estimated by comparing variations in measured efficiency between data and MC simulation using a high-purity sample of Z-boson decays and are found to be 2–4%. Effects of the uncertainty in the distribution of the number of pileup interactions are evaluated by varying the cross section used to predict the number of pileup interactions in MC by $\pm$5% from its nominal value. The uncertainty of the jet energy scale [69] (resolution) is evaluated by varying the energy scale (resolution) correction of all jets in the signal and background

Table 3: Systematic uncertainties considered in the analysis.

Source	Type	Remarks
Luminosity	rate	Signal and all backgrounds
Lepton ID/Iso	shape	Signal and all backgrounds
Trigger efficiency	shape	Signal and all backgrounds
Pileup	shape	Signal and all backgrounds
Jet energy scale	shape	Signal and all backgrounds
Jet energy resolution	shape	Signal and all backgrounds
b-tag HF fraction	shape	Signal and all backgrounds
b-tag HF stats (linear)	shape	Signal and all backgrounds
b-tag HF stats (quadratic)	shape	Signal and all backgrounds
b-tag LF fraction	shape	Signal and all backgrounds
b-tag LF stats (linear)	shape	Signal and all backgrounds
b-tag LF stats (quadratic)	shape	Signal and all backgrounds
b-tag charm (linear)	shape	Signal and all backgrounds
b-tag charm (quadratic)	shape	Signal and all backgrounds
QCD scale ($t\bar{t}H$)	rate	Scale uncertainty of NLO $t\bar{t}H$ prediction
QCD scale ($t\bar{t}$)	rate	Scale uncertainty of NLO $t\bar{t}$ prediction
QCD scale ($t\bar{t}$+HF)	rate	Additional 50% rate uncertainty of $t\bar{t}$+HF predictions
QCD scale (t)	rate	Scale uncertainty of NLO single t prediction
QCD scale (V)	rate	Scale uncertainty of NNLO W and Z prediction
QCD scale (VV)	rate	Scale uncertainty of NLO diboson prediction
pdf (gg)	rate	PDF uncertainty for gg initiated processes except $t\bar{t}H$
pdf (gg $t\bar{t}H$)	rate	PDF uncertainty for $t\bar{t}H$
pdf ($q\bar{q}$)	rate	PDF uncertainty of $q\bar{q}$ initiated processes ($t\bar{t}$ W, W, Z)
pdf (qg)	rate	PDF uncertainty of qg initiated processes (single t)
Q^2 scale ($t\bar{t}$)	shape	Renormalization and factorization scale uncertainties of the $t\bar{t}$ ME generator, independent for additional jet flavors
PS Scale ($t\bar{t}$)	rate	Renormalization and factorization scale uncertainties of the parton shower (for $t\bar{t}$ events), independent for additional jet flavors
Bin-by-bin statistics	shape	statistical uncertainty of the signal and background prediction due to the limited sample size

Table 4: Specific effect of systematic uncertainties that affect the discriminant shape on the predicted background and signal yields for events in the $\geq$ 6 jets, 3 b-tags category of the lepton+jets channel. Here, only the sum of the largest background processes, $\mathrm{t\bar{t}}$+LF, $\mathrm{t\bar{t}}$+b, $\mathrm{t\bar{t}}$+2b, $\mathrm{t\bar{t}}$+$\mathrm{b\bar{b}}$, and $\mathrm{t\bar{t}}$+$\mathrm{c\bar{c}}$, are considered.

Process	$\mathrm{t\bar{t}}$ rate up/down [%]	$\mathrm{t\bar{t}H}$ rate up/down [%]
Jet energy scale	$+12.6/-11.8$	$+8.4/-8.0$
Jet energy resolution	$+0.2/-0.3$	$-0.0/-0.1$
Pile-up	$+0.1/-0.1$	$-0.2/+0.1$
Electron efficiency	$+0.5/-0.5$	$+0.5/-0.5$
Muon efficiency	$+0.4/-0.4$	$+0.4/-0.4$
Electron trigger efficiency	$+1.2/-1.2$	$+1.3/-1.3$
Muon trigger efficiency	$+0.8/-0.8$	$+0.9/-0.9$
b-Tag HF contamination	$-9.4/+9.8$	$-2.6/+2.8$
b-Tag HF stats (linear)	$-3.1/+3.3$	$-2.5/+2.7$
b-Tag HF stats (quadratic)	$+2.6/-2.4$	$+2.4/-2.2$
b-Tag LF contamination	$+7.1/-5.2$	$+5.8/-4.5$
b-Tag LF stats (linear)	$-2.0/+4.4$	$+0.5/+1.5$
b-Tag LF stats (quadratic)	$+2.1/+0.2$	$+1.5/+0.5$
b-Tag charm Uncertainty (linear)	$-11.1/+14.9$	$-3.1/+4.1$
b-Tag charm Uncertainty (quadratic)	$+0.5/-0.5$	$-0.0/+0.0$
Q^2 scale ($\mathrm{t\bar{t}}$+LF)	$-6.2/+7.5$	–
Q^2 scale ($\mathrm{t\bar{t}}$+b)	$-1.7/+2.0$	–
Q^2 scale ($\mathrm{t\bar{t}}$+2b)	$-1.1/+1.4$	–
Q^2 scale ($\mathrm{t\bar{t}}$+$\mathrm{b\bar{b}}$)	$-2.0/+2.5$	–
Q^2 scale ($\mathrm{t\bar{t}}$+$\mathrm{c\bar{c}}$)	$-4.3/+5.4$	–
PS scale ($\mathrm{t\bar{t}}$+LF)	$+4.8/-9.0$	–
PS scale ($\mathrm{t\bar{t}}$+b)	$-0.9/+0.7$	–
PS scale ($\mathrm{t\bar{t}}$+2b)	$-0.8/+0.9$	–
PS scale ($\mathrm{t\bar{t}}$+$\mathrm{b\bar{b}}$)	$-1.5/+2.7$	–
PS scale ($\mathrm{t\bar{t}}$+$\mathrm{c\bar{c}}$)	$-3.9/+3.0$	–

predictions by one standard deviation. The uncertainty of the CSVv2 b-tagging scale factors is evaluated by applying alternative scale factors based on varying the following systematic effects by one standard deviation, separately for the different jet flavors: the contamination of background processes in the control samples, the jet energy scale uncertainty — which is correlated with the overall jet energy scale uncertainty — and the statistical uncertainty in the scale factor evaluation. The impact of the latter is parametrized as the sum of two orthogonal contributions: a linear and a quadratic term, which allow an overall tilt and a shift of the center of the b-tagging discriminant distribution, respectively. Both for the jet energy scale and for the b-tagging scale factor uncertainties, the event categorization and successive evaluation of the discriminant is re-evaluated after applying the systematic variations to account for migration effects between categories.

Theoretical uncertainties of the cross sections used to predict the rates of various processes are propagated to the yield estimates. All rates are estimated using cross sections of at least NLO accuracy, which have uncertainties arising primarily from PDFs and the choice of factorization and renormalization scales (both in the matrix element and the parton shower). The cross section uncertainties are each separated into their PDF and scale components and correlated where appropriate between processes. For example, the PDF uncertainty for processes originating primarily from gluon-gluon initial states, such as $\mathrm{t\bar{t}}$ and $\mathrm{t\bar{t}H}$ production, are treated as 100% correlated. The $\mathrm{t\bar{t}}$+$\mathrm{b\bar{b}}$ process, and to lesser extent the $\mathrm{t\bar{t}}$+2b, $\mathrm{t\bar{t}}$+b, and $\mathrm{t\bar{t}}$+$\mathrm{c\bar{c}}$ production, represent important sources of irreducible background. Neither control region studies nor higher-order theoretical calculations can currently constrain the normalization of these contributions to better than 50% accuracy; therefore a conservative extra 50% rate uncertainty is assigned to the $\mathrm{t\bar{t}}$ +heavy-flavor processes. This rate uncertainty has the largest impact on the median expected limit; omitting it in the computation improves the limit by approximately 10%. The effect of the scale uncertainties on the discriminant output shape is also taken into account for the $\mathrm{t\bar{t}}$+jets production using event weights obtained directly from the MC generator in case of the matrix element and dedicated samples generated with different scale choices in case of the parton shower. The factorization and renormalization scales of the matrix element generator and also the scales of the initial-state radiation and the final-state radiation of the parton shower of the general-purpose MC event generator are varied simultaneously by a factor of 0.5 and 2. These scale variations are treated as uncorrelated between the matrix element generator and the general-purpose MC event generator. Possible shape variations of the final discriminant distributions due to the PDF uncertainty have been evaluated by comparing the results to those obtained when using the PDF replicas provided with the NNPDF set. The replicas parametrize the PDF uncertainties and are derived from re-sampling of the experimental data that are used to fit the nominal PDF [40]. The impact on the discriminant distributions has been found to be negligible, and therefore, is not propagated into the final analysis.

The impact of statistical fluctuations in the signal and background prediction due to the limited number of simulated events is accounted for using the approach described in [75, 76].

7 Results

The signal strength modifier $\mu = \sigma/\sigma_{\mathrm{SM}}$ of the $\mathrm{t\bar{t}H}$ production cross section is determined in a simultaneous binned maximum-likelihood fit to the data across all analysis categories, cf. Section 5. The fit procedure takes into account systematic uncertainties that modify the shape and normalization of the final discriminant distributions, as described in Section 6. The final discriminants in all categories before and after the fit to data are displayed in Figs. 2 to 5 and Figs. 6 to 9, respectively.

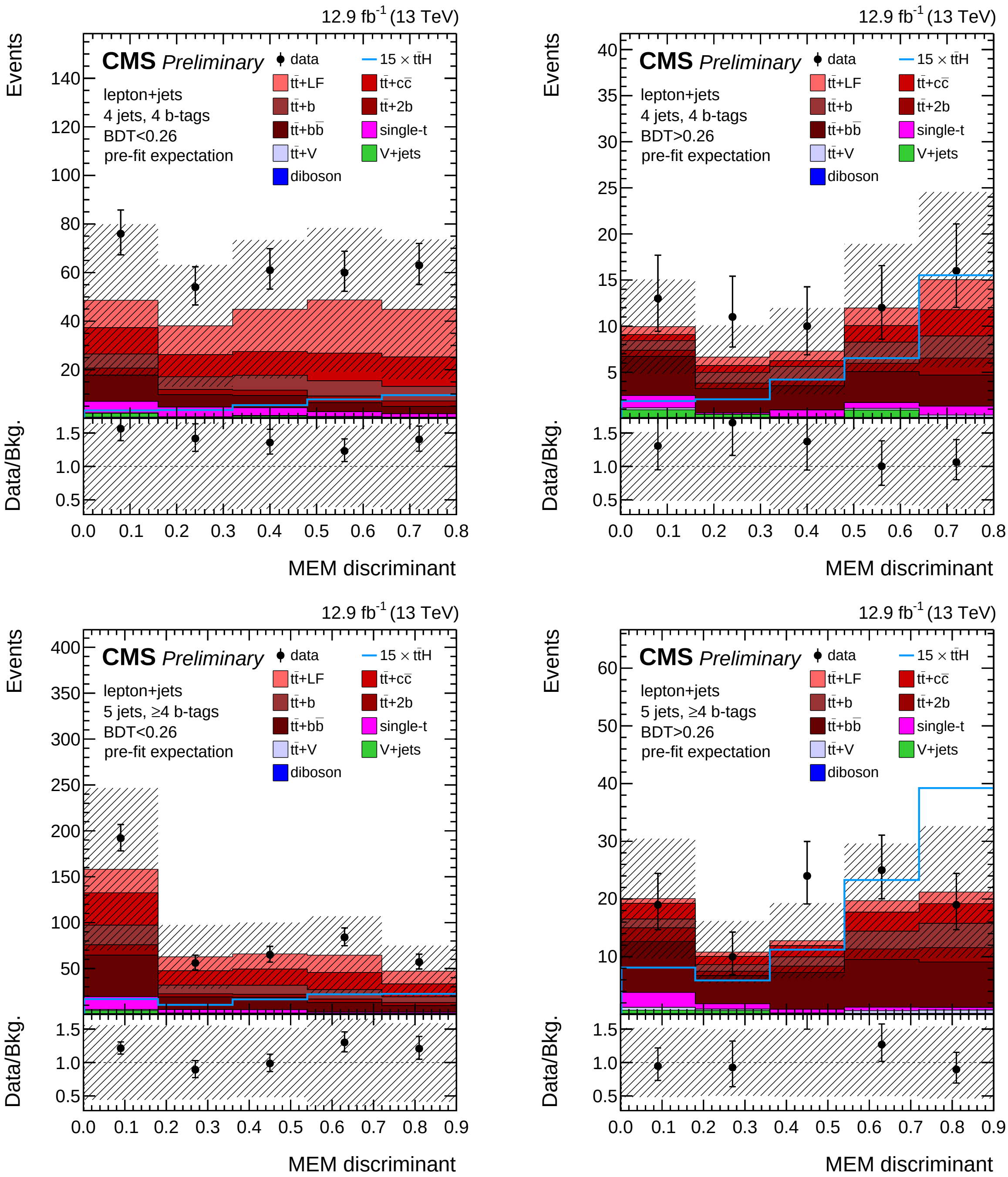

Figure 2: Final discriminant (MEM) shapes in the lepton+jets channel before the fit to data, in the analysis categories with 4 jets, 4 b-tags (top row) and 5 jets, $\geq$ 4 b-tags (bottom row) with low (left) and high (right) BDT output. The expected background contributions (filled histograms) are stacked, and the expected signal distribution (line) for a Higgs-boson mass of $m_H = 125\,\text{GeV}$ is superimposed. Each contribution is normalized to an integrated luminosity of $12.9\,\text{fb}^{-1}$, and the signal distribution is additionally scaled by a factor of 15 for better readability. The error bands include the total uncertainty of the fit model. The distributions observed in data (markers) are also shown.

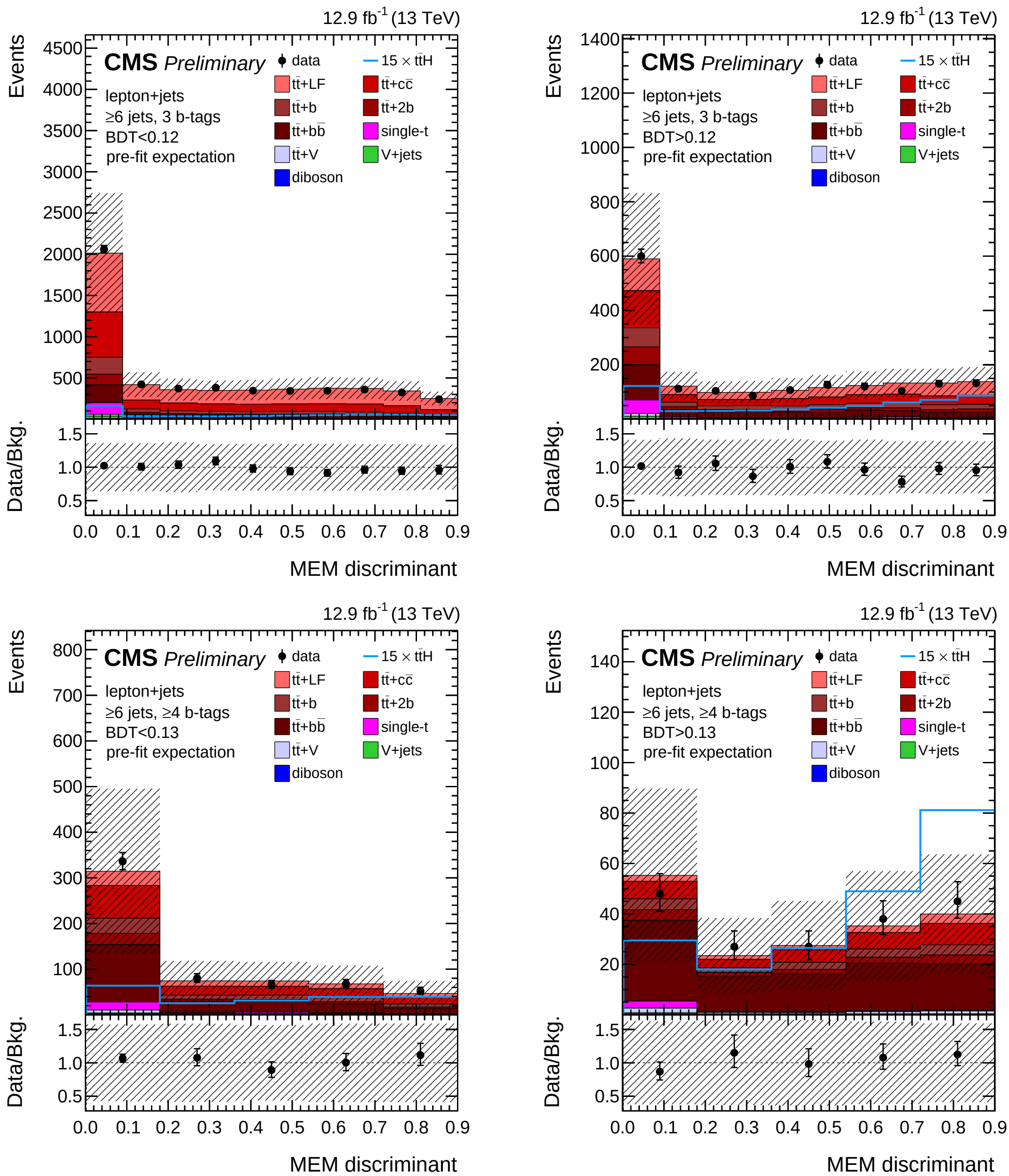

Figure 3: Final discriminant (MEM) shapes in the lepton+jets channel before the fit to data, in the analysis categories with $\geq$ 6 jets, 3 b-tags (top row) and $\geq$ 6 jets, $\geq$ 4 b-tags (bottom row) with low (left) and high (right) BDT output (continued from Fig. 2).

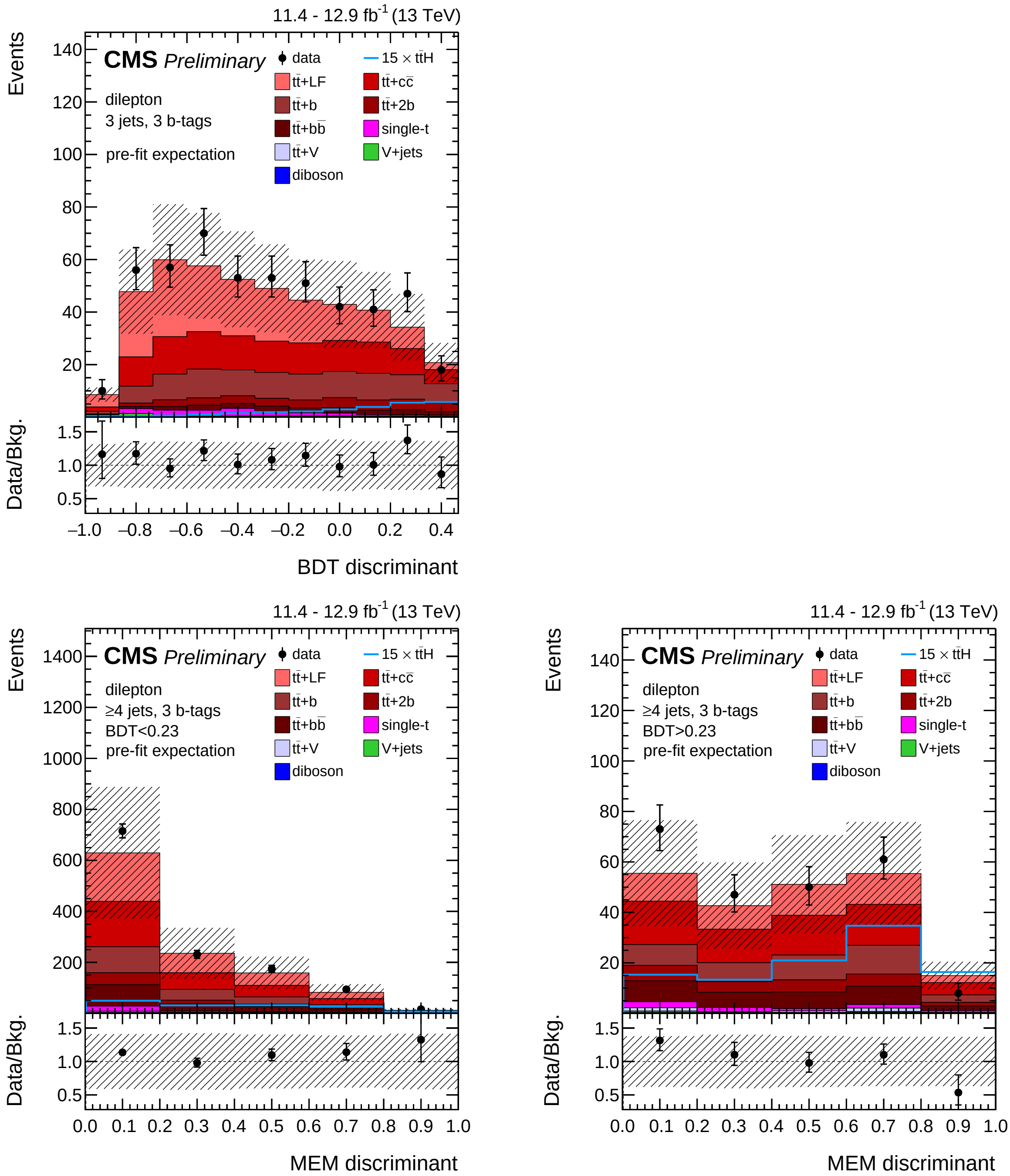

Figure 4: Final discriminant shapes (BDT or MEM) in the dilepton channel before the fit to data, in the analysis categories with 3 jets, 3 b-tags (top row) and $\geq$ 4 jets, 3 b-tags (bottom row) with low (left) and high (right) BDT output. The expected background contributions (filled histograms) are stacked, and the expected signal distribution (line) for a Higgs-boson mass of $m_H = 125\,\text{GeV}$ is superimposed. Each contribution is normalized to an integrated luminosity of $11.4 - 12.9\,\text{fb}^{-1}$, and the signal distribution is additionally scaled by a factor of 15 for better readability. The error bands include the total uncertainty of the fit model. The distributions observed in data (markers) are also shown.

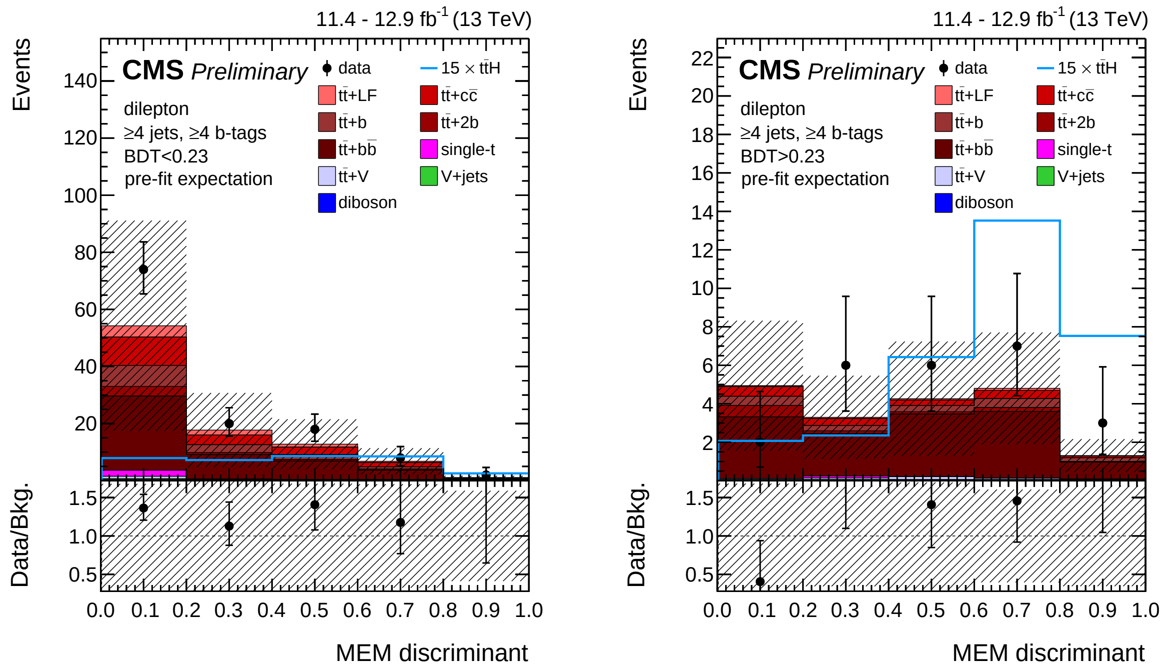

Figure 5: Final discriminant (MEM) shapes in the dilepton channel before the fit to data, in the analysis categories with ≥ 4 jets, ≥ 4 b-tags with low (left) and high (right) BDT output (continued from Fig. 4).

The best-fit value of μ is $-0.19\,^{+0.45}_{-0.44}(\text{stat.})\,^{+0.66}_{-0.68}(\text{syst.})$ with a total uncertainty of $^{+0.80}_{-0.81}$. This is 1.5 standard deviations from the standard model expectation of $\mu = 1$. The best-fit values in each analysis channel and in the combination are listed in Table 5 and displayed in Fig. 10 (left).

The value obtained for μ is both compatible with the SM expectation and no signal: an upper limit at 95% confidence level (CL) is determined using a modified frequentist CL_s method [77, 78]. When combining all categories and channels, an observed (expected) upper limit of $\mu <$ 1.5 (1.7) at the 95% CL is obtained. The expected and observed upper limits in each channel and in the combination are listed in Table 5 and visualized in Figure 10 (right). The limits in each individual category are listed in Appendices A and B for the lepton+jets and the dilepton channel, respectively.

Table 5: Best-fit value of the signal strength modifier μ and the median expected and observed 95% CL upper limits (UL) in the dilepton and the lepton+jets channels as well as the combined results. The one standard deviation ($\pm 1\sigma$) confidence intervals of the expected limit and the best-fit value are also quoted, split into the statistical and systematic components in the latter case. Expected limits are calculated with the asymptotic method [79].

Channel	Observed UL	Expected UL	Best-fit μ
Dilepton	3.2	$3.4^{+1.5}_{-1.0}$	$-0.04^{+1.50}_{-1.39}(\text{tot.})\ ^{+1.05}_{-0.96}(\text{stat.})\ ^{+1.01}_{-1.06}(\text{syst.})$
Lepton+jets	1.8	$2.1^{+1.0}_{-0.6}$	$-0.43^{+1.02}_{-1.02}(\text{tot.})\ ^{+0.51}_{-0.52}(\text{stat.})\ ^{+0.88}_{-0.87}(\text{syst.})$
Combined	1.5	$1.7^{+0.7}_{-0.5}$	$-0.19^{+0.80}_{-0.81}(\text{tot.})\ ^{+0.45}_{-0.44}(\text{stat.})\ ^{+0.66}_{-0.68}(\text{syst.})$

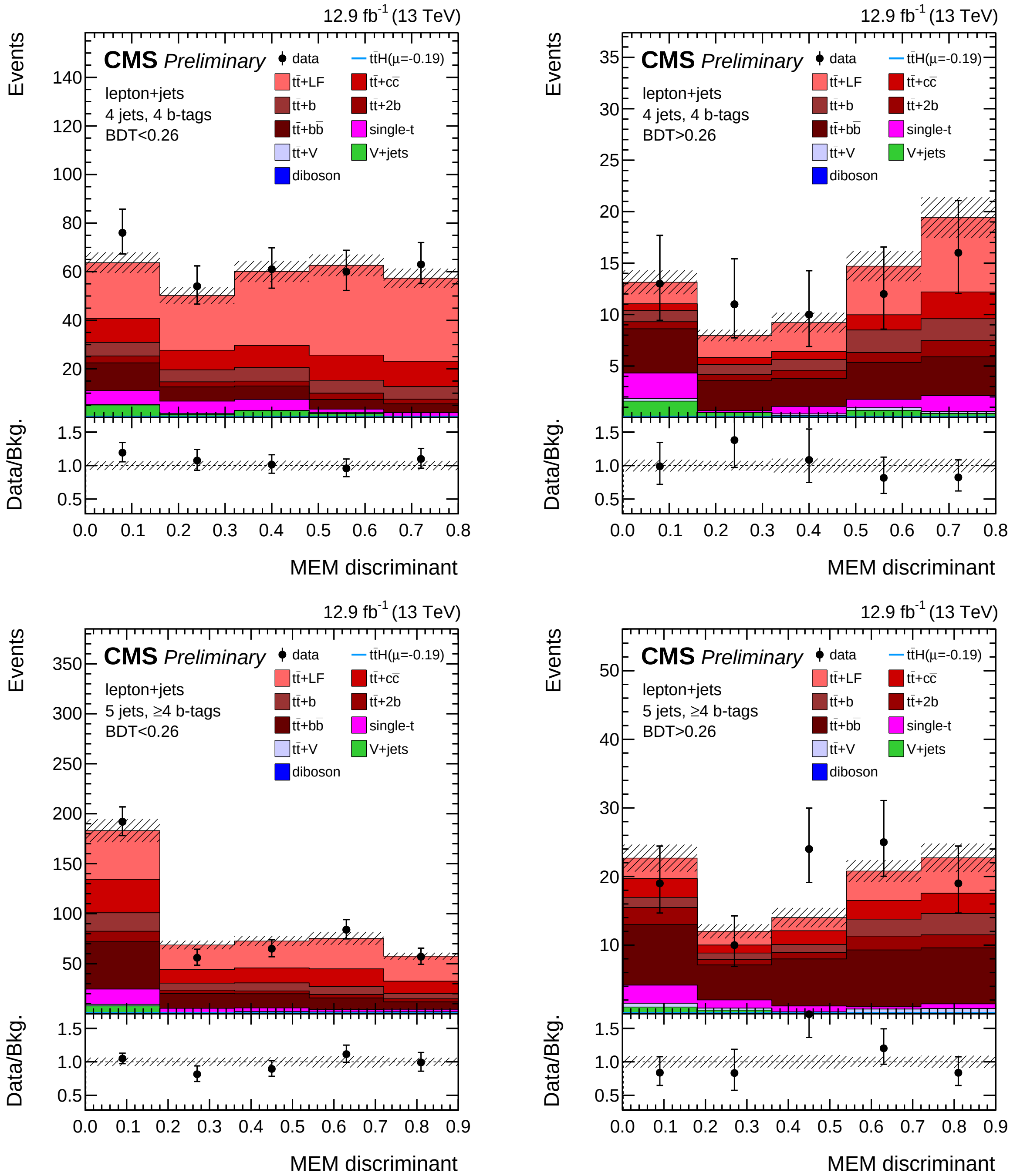

Figure 6: Final discriminant shapes (MEM) in the analysis categories with 4 jets, 4 b-tags (top row) and 5 jets, ≥ 4 b-tags (bottom row) with low (left) and high (right) BDT output in the lepton+jets channel after the fit to data.

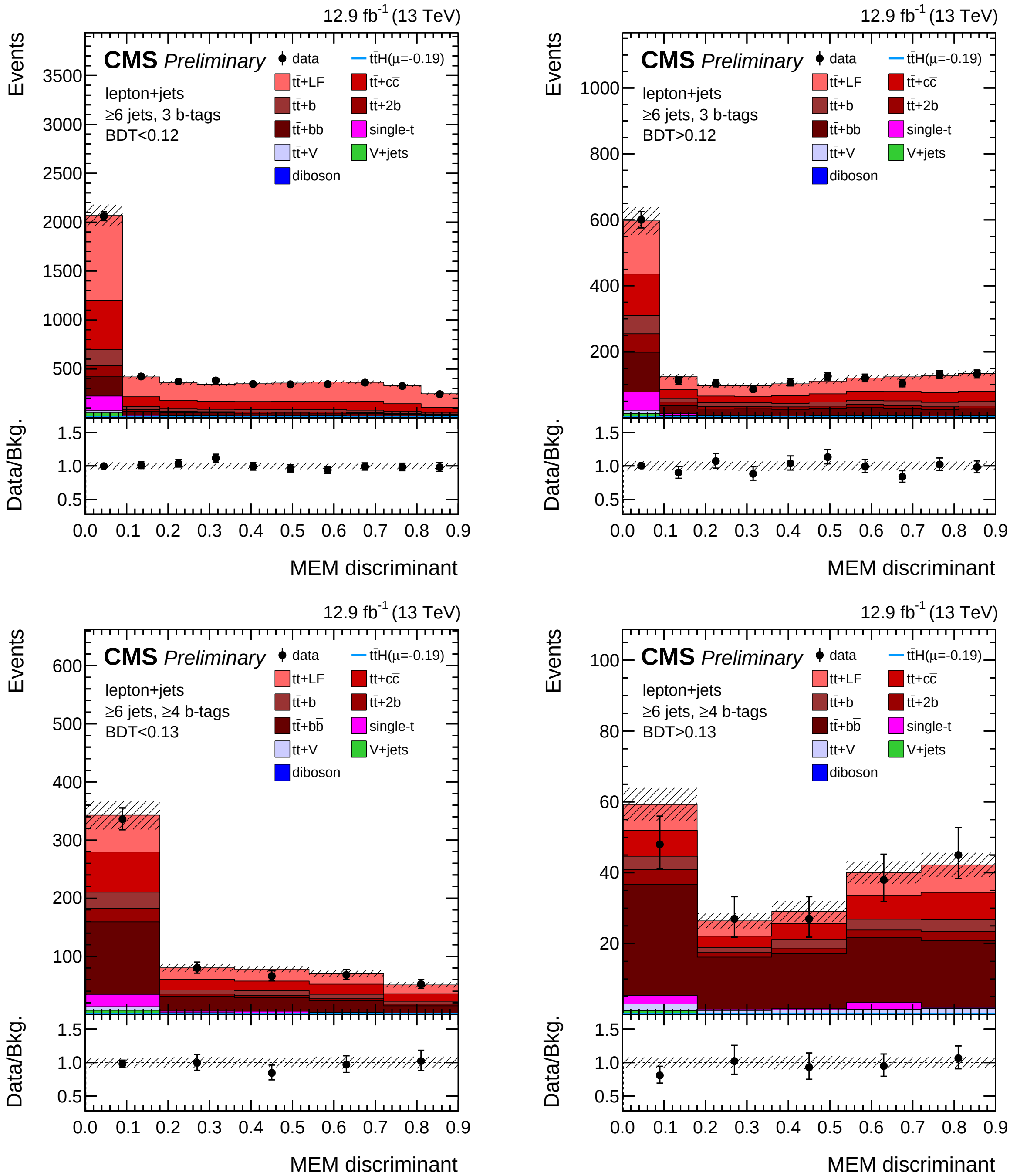

Figure 7: Final discriminant shapes (MEM) in the analysis categories with $\geq$ 6 jets, 3 b-tags (top row) and $\geq$ 6 jets, $\geq$ 4 b-tags (bottom row) with low (left) and high (right) BDT output in the lepton+jets channel after the fit to data (continued from Fig. 6).

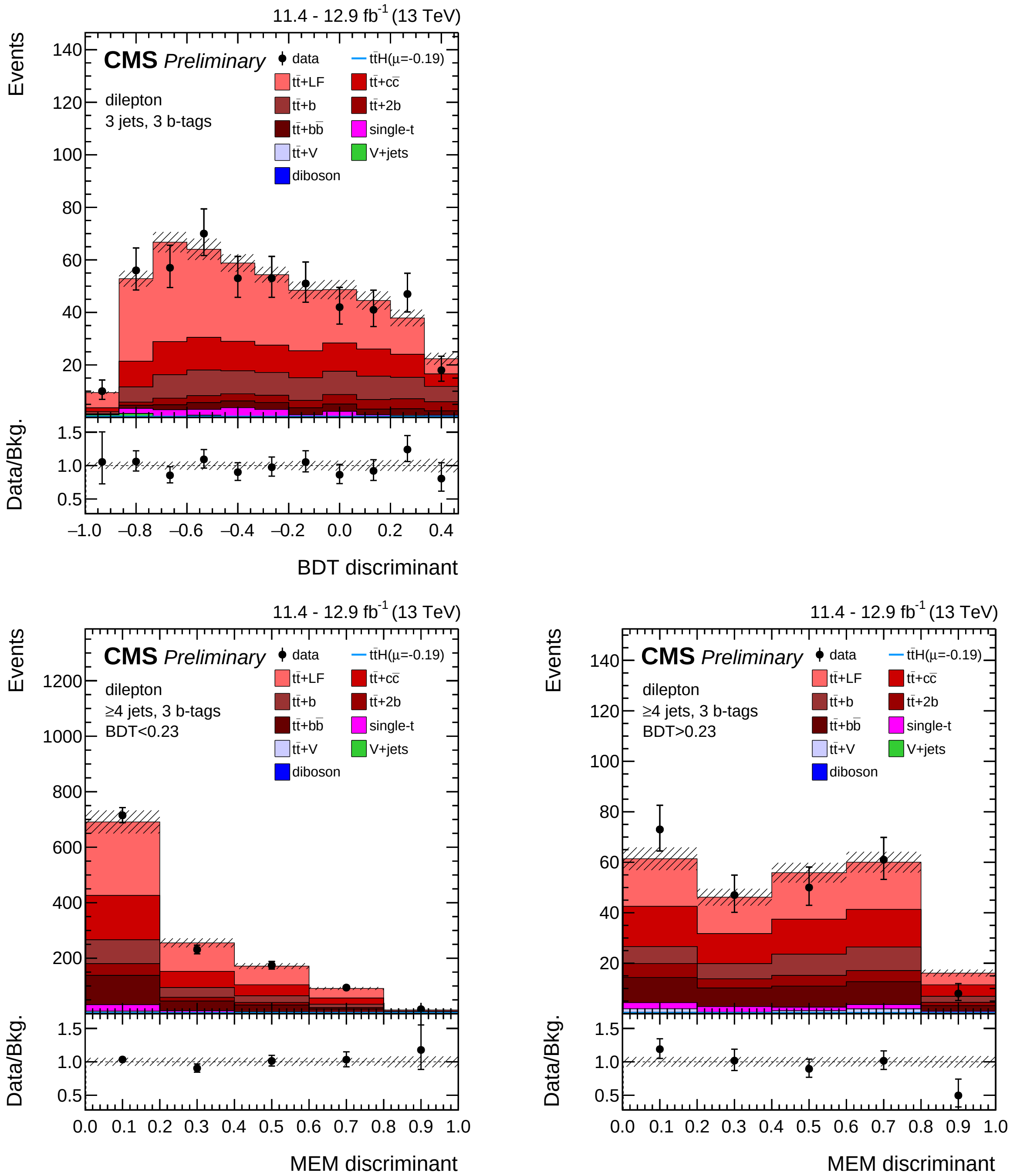

Figure 8: Final discriminant shapes (BDT or MEM) in the analysis categories with 3 jets, 3 b-tags (top row) and $\geq$ 4 jets, 3 b-tags (bottom row) with low (left) and high (right) BDT output in the dilepton channel after the fit to data.

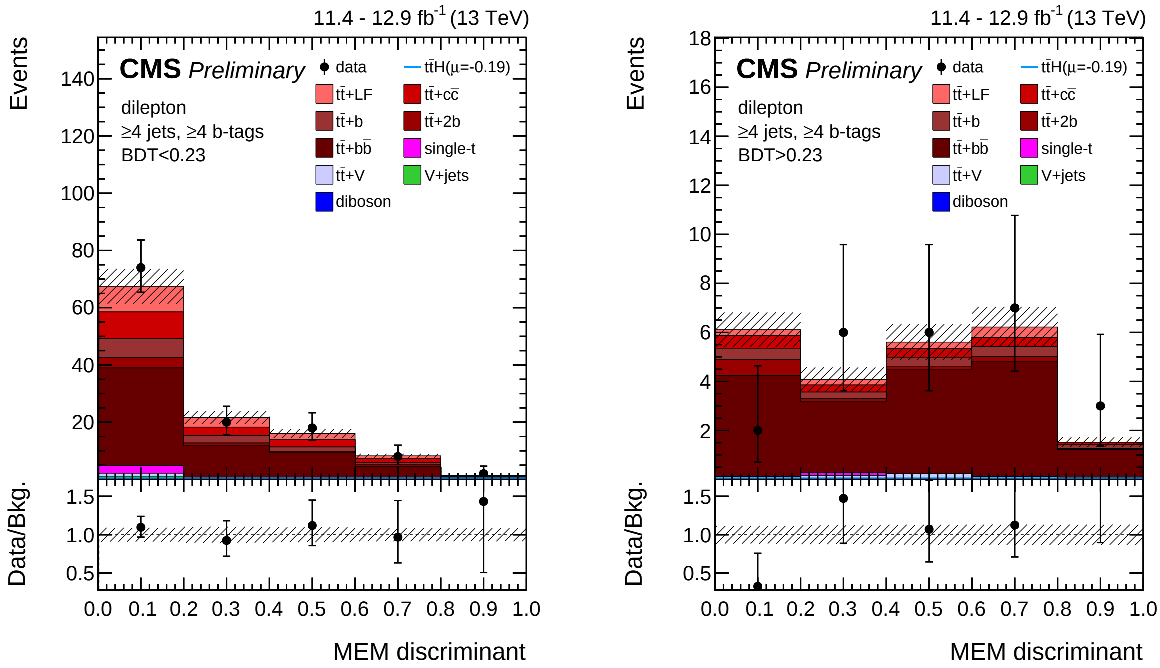

Figure 9: Final discriminant shapes (MEM) in the analysis categories with ≥ 4 jets, ≥ 4 b-tags with low (left) and high (right) BDT output in the dilepton channel after the fit to data (continued from Fig. 8).

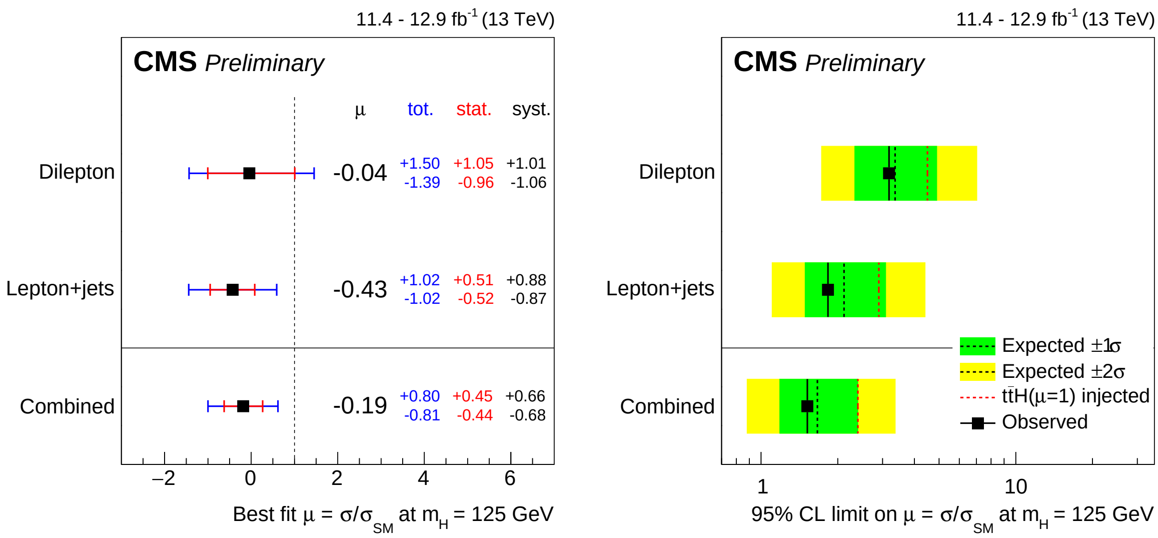

Figure 10: Best-fit values of the signal strength modifiers μ with their $\pm 1\sigma$ confidence intervals, also split into their statistical and systematic components (left), and median expected and observed 95% CL upper limits on μ (right). The expected limits are displayed together with $\pm 1\sigma$ and $\pm 2\sigma$ confidence intervals. Also shown are the limits in case of an injected signal of $\mu = 1$.

8 Summary

A search for the associated production of a Higgs boson and a top quark-antiquark pair is performed using up to 12.9 fb^{-1} of pp collision data recorded with the CMS detector at a center-of-mass energy of 13 TeV in 2016. Candidate events are selected in final states compatible with the Higgs boson decay $\mathrm{H} \rightarrow \mathrm{b\bar{b}}$ and the lepton+jets or dilepton decay channel of the $\mathrm{t\bar{t}}$ pair. Selected events are split into mutually exclusive categories according to their $\mathrm{t\bar{t}}$ decay channel and jet content. In each category a powerful discriminant is constructed to separate the $\mathrm{t\bar{t}H}$ signal from the $\mathrm{t\bar{t}}$-dominated background, based on boosted decision trees and the matrix element method. An observed (expected) upper limit on the $\mathrm{t\bar{t}H}$ production cross section relative to the SM expectations of $\mu = 1.5\,(1.7)$ at the 95% confidence level is obtained. The best-fit value of μ is $-0.19\,^{+0.45}_{-0.44}(\mathrm{stat.})\,^{+0.66}_{-0.68}(\mathrm{syst.})$. These results are compatible with SM expectations at the level of 1.5 standard deviations.

A Lepton+Jets Additional Material

In the following, the input variables used to train the BDTs in each category of the lepton+jets channel are presented. In Table 6, all variables used in any of the categories are described, and in Table 7, the variables used per category are listed. The observed and expected upper limits at 95% CL on the signal strength modifier μ under the background-only hypothesis in the lepton+jets channel are listed in Table 8 and displayed in Fig. 11 for the individual categories and for the combined fit in all categories.

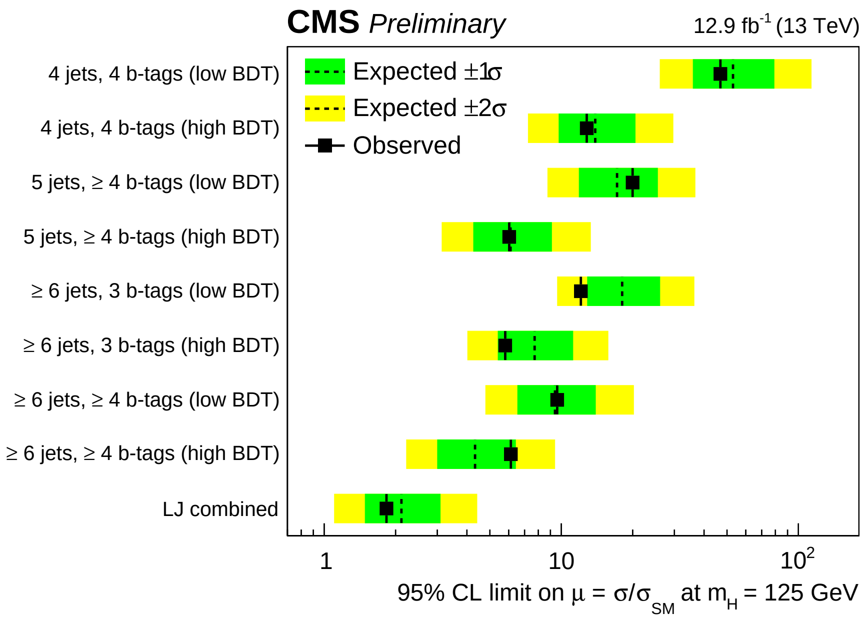

Figure 11: Observed and expected upper limits at 95% CL on μ in the lepton+jets channel. The limits are calculated with the asymptotic method.

Table 6: Variables used in the BDT training in the lepton+jets channel.

Object and event kinematics	
$p_{\mathrm{T}}(\text{jet } i)$	Transverse momentum of i-th jet, jets ordered in p_{T}
HT	Scalar sum of transverse momenta for all jets with $p_{\mathrm{T}} > 30$ GeV/c
MET	Missing transverse energy
$\sum p_{\mathrm{T}}(\text{jets, lepton, MET})$	Sum of the p_{T} of all jets, leptons, and MET
$M(\text{jets, lepton, MET})$	Invariant mass of the 4-vector sum of all jets, leptons, and MET
avg. $\Delta\mathrm{R}(\text{tag, tag})$	Average ΔR between b-tagged jets
avg. $\Delta\eta(\text{jet, jet})$	Average $\Delta\eta$ between jets
max $\Delta\|\eta\|$ (jet, avg. jet $\|\eta\|$)	Maximal $\Delta\|\eta\|$ between any jet and the average $\|\eta\|$ of all jets
max $\Delta\|\eta\|$ (tag, avg. jet $\|\eta\|$)	Maximal $\Delta\|\eta\|$ between any b-tagged jet and the average $\|\eta\|$ of all jets
max $\Delta\|\eta\|$ (tag, avg. tag $\|\eta\|$)	Maximal $\Delta\|\eta\|$ between any b-tagged jet and the average $\|\eta\|$ of all b-tagged jets
min $\Delta\mathrm{R}(\text{tag, tag})$	ΔR between the two closest b-tagged jets
M_3	Invariant mass of the 3-jet system with largest transverse momentum
min $\Delta\mathrm{R}(\text{lepton, jet})$	ΔR between the lepton and the closest jet
M_2 of min $\Delta\mathrm{R}(\text{tag, tag})$	Invariant mass of the two b-tagged jets that are closest in ΔR
$M_2(\text{tag, tag})$ closest to 125	Invariant mass of the two b-tagged jets with an invariant mass closest to 125 GeV/c^2
best Higgs mass	A minimum-chi-squared fit to event kinematics is used to select two b-tagged jets as top-decay products. Of the remaining b-tagged jets, the invariant mass of the two with highest E_{T} is saved as this quantity.
$\sqrt{\Delta\eta(\mathrm{t}^{\mathrm{lep}}, \mathrm{bb}) \times \Delta\eta(\mathrm{t}^{\mathrm{had}}, \mathrm{bb})}$	Square root of the product of $\|\Delta\eta\|(\text{leptonic top, bb})$ and $\|\Delta\eta\|(\text{hadronic top, bb})$, where the $\mathrm{b\bar{b}}$-system and the candidates for the leptonically and hadronically decaying top quarks are found with the best Higgs mass algorithm
$(\Sigma p_{\mathrm{T}}(\text{jet}))/(\Sigma E(\text{jet}))$	Ratio of the sum of the transverse momenta of all jets and the sum of the energies of all jets
CSVv2 b-tag	
1st- to 5th-highest CSVv2	First- to fifth-highest highest b-tag discriminant value of all jets
avg. CSVv2 of b-tagged jets	Average b-tag discriminant value of all b-tagged jets
2nd moment of b-tagged jets' CSVv2	Squared difference between the b-tag discriminant value of a given b-tagged jet and the average CSVv2 discriminant value of all b-tagged jets, summed over all b-tagged jets
b-tagging likelihood ratio	Ratio of the likelihood that the event contains four b-jets to the likelihood that it contains two b-jets. The likelihoods are constructed from the b-tag discriminant, the p_T and the η of the jets.
Event shape	
sphericity	The sphericity is defined as $\frac{3}{2}(\lambda_2 + \lambda_3)$ where the λ_{i} are the eigenvalues of the sphericity tensor as defined in [80]
aplanarity	The aplanarity is defined as $\frac{3}{2}\lambda_3$ where the λ_{i} are the eigenvalues of the sphericity tensor as defined in [80]
H_1–H_4	The first four Fox-Wolfram moments as defined in [81]

Table 7: BDT input variable assignment per category in the lepton+jets channel.

4 jets, 4 tags	5 jets, $\geq$ 4 tags
$\sum p_T$(jets, lepton, MET) avg. CSVv2 of b-tagged jets aplanarity H_3 $(\Sigma p_T(\text{jet}))/(\Sigma E(\text{jet}))$ M_2 of min ΔR(tag, tag)	avg. $\Delta\eta$(jet, jet) HT avg. CSVv2 of b-tagged jets M_2(tag, tag) closest to 125 M_3 $\sum p_T$(jets, lepton, MET) M_2 of min ΔR(tag, tag) aplanarity avg. ΔR(tag, tag)
$\geq$ 6 jets, 3 tags	$\geq$ 6 jets, $\geq$ 4 tags
aplanarity $\sqrt{\Delta\eta(\text{t}^{\text{lep}}, \text{bb}) \times \Delta\eta(\text{t}^{\text{had}}, \text{bb})}$ $(\Sigma p_T(\text{jet}))/(\Sigma E(\text{jet}))$ min ΔR(tag, tag) 2nd moment of b-tagged jets' CSVv2 $\sum p_T$(jets, lepton, MET) b-tagging likelihood ratio	best Higgs mass M_2(tag, tag) closest to 125 M(jets, lepton, MET) 4th highest CSVv2 $\sum p_T$(jets, lepton, MET) 5th highest CSVv2

Table 8: Observed and median expected 95% CLs upper limits on μ in the lepton+jets channel, calculated with the asymptotic method. The upper and lower range of the 1σ confidence interval is also quoted.

Category	Observed	Expected
4 jets, 4 b-tags (low BDT)	46.9	$53.0^{+26.0}_{-17.0}$
4 jets, 4 b-tags (high BDT)	12.8	$13.9^{+6.6}_{-4.1}$
5 jets, $\geq$ 4 b-tags (low BDT)	20.0	$17.2^{+8.3}_{-5.3}$
5 jets, $\geq$ 4 b-tags (high BDT)	6.0	$6.1^{+3.0}_{-1.8}$
$\geq$ 6 jets, 3 b-tags (low BDT)	12.1	$18.1^{+8.0}_{-5.2}$
$\geq$ 6 jets, 3 b-tags (high BDT)	5.8	$7.7^{+3.5}_{-2.3}$
$\geq$ 6 jets, $\geq$ 4 b-tags (low BDT)	9.6	$9.4^{+4.5}_{-2.9}$
$\geq$ 6 jets, $\geq$ 4 b-tags (high BDT)	6.1	$4.3^{+2.1}_{-1.3}$
lepton+jets combined	1.8	$2.1^{+1.0}_{-0.6}$

B Dilepton Additional Material

In the following, the input variables used to train the BDTs in each category of the dilepton channel are presented. In Table 9, all variables used in any of the categories are described, and in Table 10, the variables used per category are listed. The observed and expected upper limits at 95% CL on the signal strength modifier μ under the background-only hypothesis in the dilepton channel are listed in Table 11 and displayed in Fig. 12 for the individual categories and for the combined fit in all categories.

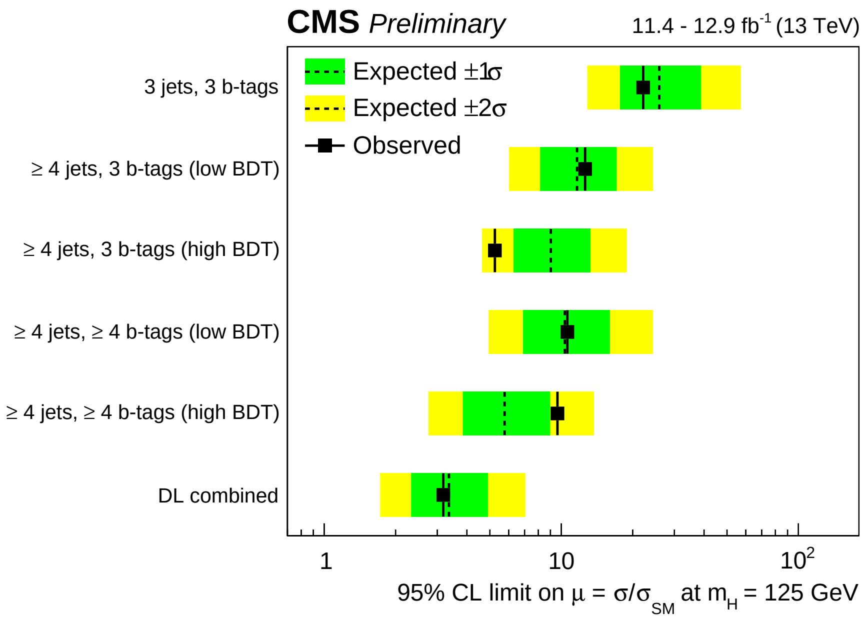

Figure 12: Observed and expected and upper limits at 95% CL on μ in the dilepton channel. The limits are calculated with the asymptotic method.

Table 9: Variables used in the BDT training in the dilepton channel.

Event variable	Description
Object and event kinematics	
$\langle \Delta R_{\text{tag,tag}} \rangle$	Average ΔR between b-tagged jets
$\sum p_{T\,\text{jets,leptons}}$	Sum of the p_T of all jets and leptons
$\tau_{\text{jet,jet}}^{\text{max mass}}$	Twist angle between jet pair
min $\Delta R_{\text{tag,tag}}$	ΔR between the two closest b-tagged jets
max $\Delta \eta_{\text{tag,tag}}$	$\Delta \eta$ between the two furthest b-tagged jets
$M_{\text{jet,jet}}^{\text{min}\Delta R}$	Invariant mass of jet pair with minimum ΔR
$M_{\text{higgs-like}}^{\text{jj}}$	Invariant mass of a jet pair ordered in closeness to the Higgs mass
$M_{\text{tag,tag}}^{\text{min}\Delta R}$	Mass of b-tagged jet pair with minimum ΔR
$p_{T\ \text{tag,tag}}^{\text{min}\Delta R}$	Sum of the p_T of b-tagged jet pair with minimum ΔR
Centrality (tags)	Ratio of the sum of the transverse momentum of all b-tagged jets and the sum of the energy of all b-tagged jets
Centrality (jets, leptons)	Ratio of the sum of the transverse momentum of all jets and leptons, and the sum of the energy of all jets and leptons
H_T	Scalar sum of transverse momentum for all jets
min $\Delta R_{\text{jet,jet}}$	ΔR between the two closest jets
median $M_{\text{jet,jet}}$	Median invariant mass of all combinations of jet pairs
$M_{\text{tag,tag}}^{\text{max mass}}$	Mass for b-tagged jet pair with maximum invariant mass combination
$\langle \Delta R_{jet,tag} \rangle$	Average ΔR between jets (with at least one b-tagged jet)
$p_{T\ jet,tag}^{\text{min}\Delta R}$	Sum of the p_T of jet pair with minimum ΔR (with at least one b-tagged jet)
$\tau_{\text{jet,tag}}^{\text{max mass}}$	Twist angle between jet pair (with at least one b-tagged jet)
$\text{m}_{\text{jet,jet,jet}}^{\text{max p}_\text{T}}$	Invariant mass of the 3-jet system with the largest transverse momentum.
$M_{\text{higgs-like}}^{\text{bj}}$	Invariant mass of a jet pair (with at least one b-tagged jet) ordered in closeness to the Higgs mass
CSVv2 b-tag	
$\langle d \rangle_{\text{tagged/untagged}}$	Average CSVv2 b-tag discriminant value for b-tagged/un-b-tagged jets
Event shape	
H_0, H_1, H_2, H_3, H_4	Fox-Wolfram moments [81]
C(jets)	$3\,(\lambda_1\lambda_2 + \lambda_1\lambda_3 + \lambda_2\lambda_3)$ [80]

Table 10: BDT input variable assignment per category in the dilepton channel.

3 jets, 3 tags	≥4 jets, 3 tags	≥4 jets, ≥4 tags
$\langle d \rangle_{\text{tagged}}$	Centrality(jets & leptons)	Centrality(jets & leptons)
H_1(jets)	C(jets)	Centrality(tags)
$M^{\text{bj}}_{\text{higgs-like}}$	H_2(tags)	H_T^{tags}
$M^{\text{max mass}}_{\text{tag,tag}}$	$M^{\text{jj}}_{\text{higgs-like}}$	$M^{\text{jj}}_{\text{higgs-like}}$
min $\Delta R_{\text{tag,tag}}$	$M^{\max p_T}_{\text{jet,jet,jet}}$	min $\Delta R_{\text{jet,jet}}$
max $\Delta\eta_{\text{jet,jet}}$	$M^{\min\Delta R}_{\text{tag,tag}}$	$M^{\min\Delta R}_{jet,tag}$
min $\Delta R_{\text{jet,jet}}$	min $\Delta R_{\text{tag,tag}}$	$M^{\text{max mass}}_{\text{tag,tag}}$
$\sum p_{T\text{jets,leptons}}$	max $\Delta\eta_{\text{tag,tag}}$	$M^{\min\Delta R}_{\text{tag,tag}}$
H_4/H_0(tags)	$\tau^{\text{max mass}}_{\text{tag,tag}}$	max $\Delta\eta_{\text{jet,jet}}$
		max $\Delta\eta_{\text{tag,tag}}$
		median $M_{\text{jet,jet}}$

Table 11: Observed and median expected and 95% CLs upper limits on μ in the dilepton channel, calculated with the asymptotic method. The upper and lower range of the 1σ confidence interval is also quoted.

Category	Observed	Expected
3 jets, 3 b-tags	22.2	$25.9^{+12.9}_{-8.2}$
$\geq$ 4 jets, 3 b-tags (low BDT)	12.6	$11.7^{+5.4}_{-3.5}$
$\geq$ 4 jets, 3 b-tags (high BDT)	5.2	$9.0^{+4.2}_{-2.7}$
$\geq$ 4 jets, $\geq$ 4 b-tags (low BDT)	10.6	$10.3^{+5.6}_{-3.4}$
$\geq$ 4 jets, $\geq$ 4 b-tags (high BDT)	9.6	$5.8^{+3.2}_{-1.9}$
dilepton combined	3.2	$3.4^{+1.5}_{-1.0}$

References

[1] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys. Lett. B* **716** (2012), no. 1, 1–29, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.

[2] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Phys. Lett. B* **716** (2012), no. 1, 30–61, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.

[3] CMS Collaboration, "Evidence for the direct decay of the 125 GeV Higgs boson to fermions", *Nature Phys.* **10** (2014), no. 5, 557–560, doi:10.1038/nphys3005, arXiv:1401.6527.

[4] ATLAS Collaboration, "Evidence for the Higgs-boson Yukawa coupling to tau leptons with the ATLAS detector", *JHEP* **04** (2015) 117, doi:10.1007/JHEP04(2015)117, arXiv:1501.04943.

[5] ATLAS Collaboration, "Measurements of Higgs boson production and couplings in diboson final states with the ATLAS detector at the LHC", *Phys. Lett. B* **726** (2013), no. 1-3, 88–119, doi:10.1016/j.physletb.2014.05.011, 10.1016/j.physletb.2013.08.010, arXiv:1307.1427. [Erratum: Phys. Lett.B734,406(2014)].

[6] CMS Collaboration, "Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV", *Eur. Phys. J. C.* **75** (2015), no. 5, 212, doi:10.1140/epjc/s10052-015-3351-7, arXiv:1412.8662.

[7] ATLAS Collaboration, "Evidence for the spin-0 nature of the Higgs boson using ATLAS data", *Phys. Lett. B* **726** (2013), no. 1-3, 120–144, doi:10.1016/j.physletb.2013.08.026, arXiv:1307.1432.

[8] CMS Collaboration, "Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at 7 and 8 TeV", *Phys. Rev. D* **92** (2015), no. 1, 012004, doi:10.1103/PhysRevD.92.012004, arXiv:1411.3441.

[9] LHC Higgs Cross Section Working Group Collaboration, "Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector", arXiv:1610.07922.

[10] G. Burdman, M. Perelstein, and A. Pierce, "Large Hadron Collider tests of a little Higgs model", *Phys. Rev. Lett.* **90** (2003), no. 24, 241802, doi:10.1103/PhysRevLett.90.241802, arXiv:hep-ph/0212228. [Erratum: Phys. Rev. Lett.92,049903(2004)].

[11] T. Han, H. E. Logan, B. McElrath, and L.-T. Wang, "Phenomenology of the little Higgs model", *Phys. Rev. D* **67** (2003), no. 9, 095004, doi:10.1103/PhysRevD.67.095004, arXiv:hep-ph/0301040.

[12] M. Perelstein, M. E. Peskin, and A. Pierce, "Top quarks and electroweak symmetry breaking in little Higgs models", *Phys. Rev. D* **69** (2004), no. 7, 075002, doi:10.1103/PhysRevD.69.075002, arXiv:hep-ph/0310039.

[13] H.-C. Cheng, I. Low, and L.-T. Wang, "Top partners in little Higgs theories with T-parity", *Phys. Rev. D* **74** (2006), no. 5, 055001, doi:10.1103/PhysRevD.74.055001, arXiv:hep-ph/0510225.

[14] H.-C. Cheng, B. A. Dobrescu, and C. T. Hill, "Electroweak symmetry breaking and extra dimensions", *Nucl. Phys. B.* **589** (2000), no. 1-3, 249–268, doi:10.1016/S0550-3213(00)00401-6, arXiv:hep-ph/9912343.

[15] M. Carena, E. Ponton, J. Santiago, and C. E. M. Wagner, "Light Kaluza Klein States in Randall-Sundrum Models with Custodial SU(2)", *Nucl. Phys. B.* **759** (2006), no. 1-2, 202–227, doi:10.1016/j.nuclphysb.2006.10.012, arXiv:hep-ph/0607106.

[16] R. Contino, L. Da Rold, and A. Pomarol, "Light custodians in natural composite Higgs models", *Phys. Rev. D* **75** (2007), no. 5, 055014, doi:10.1103/PhysRevD.75.055014, arXiv:hep-ph/0612048.

[17] G. Burdman and L. Da Rold, "Electroweak Symmetry Breaking from a Holographic Fourth Generation", *JHEP* **12** (2007) 086, doi:10.1088/1126-6708/2007/12/086, arXiv:0710.0623.

[18] C. T. Hill, "Topcolor: Top quark condensation in a gauge extension of the standard model", *Phys. Lett. B* **266** (1991), no. 3, 419–424, doi:10.1016/0370-2693(91)91061-Y.

[19] A. Carmona, M. Chala, and J. Santiago, "New Higgs Production Mechanism in Composite Higgs Models", *JHEP* **07** (2012) 049, doi:10.1007/JHEP07(2012)049, arXiv:1205.2378.

[20] CMS Collaboration, "Search for the associated production of the Higgs boson with a top-quark pair", *JHEP* **09** (2014) 087, doi:10.1007/JHEP09(2014)087,10.1007/JHEP10(2014)106, arXiv:1408.1682. [Erratum: JHEP10,106(2014)].

[21] ATLAS Collaboration, "Search for the associated production of the Higgs boson with a top quark pair in multilepton final states with the ATLAS detector", *Phys. Lett. B* **749** (2015) 519–541, doi:10.1016/j.physletb.2015.07.079, arXiv:1506.05988.

[22] J. M. Campbell et al., "The Matrix Element Method at Next-to-Leading Order", *JHEP* **11** (2012) 043.

[23] CMS Collaboration, "Search for a standard model Higgs boson produced in association with a top-quark pair and decaying to bottom quarks using a matrix element method", *Eur. Phys. J. C.* **75** (2015), no. 6, 251, doi:10.1140/epjc/s10052-015-3454-1, arXiv:1502.02485.

[24] ATLAS Collaboration, "Search for the Standard Model Higgs boson produced in association with top quarks and decaying into $b\bar{b}$ in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector", *Eur. Phys. J. C.* **75** (2015), no. 7, 349, doi:10.1140/epjc/s10052-015-3543-1, arXiv:1503.05066.

[25] LHC Higgs Cross Section Working Group Collaboration, "Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables", doi:10.5170/CERN-2011-002, arXiv:1101.0593.

[26] CMS Collaboration, "Updated measurements of Higgs boson production in the diphoton decay channel at $\sqrt{s} = 13$ TeV in pp collisions at CMS.", CMS Physics Analysis Summary CMS-PAS-HIG-16-020, 2016.

[27] CMS Collaboration, "Search for associated production of Higgs bosons and top quarks in multilepton final states at $\sqrt{s} = 13$ TeV", CMS Physics Analysis Summary CMS-PAS-HIG-16-022, 2016.

[28] ATLAS Collaboration, "Measurement of fiducial, differential and production cross sections in the $H \to \gamma\gamma$ decay channel with 13.3 fb^{-1} of 13 TeV proton-proton collision data with the ATLAS detector", Technical Report ATLAS-CONF-2016-067, 2016.

[29] ATLAS Collaboration, "Search for the Associated Production of a Higgs Boson and a Top Quark Pair in Multilepton Final States with the ATLAS Detector", Technical Report ATLAS-CONF-2016-058, 2016.

[30] CMS Collaboration, "Search for $\mathrm{t\bar{t}H}$ production in the $\mathrm{H} \to \mathrm{b\bar{b}}$ decay channel with $\sqrt{s} = 13$ TeV pp collisions at the CMS experiment", CMS Physics Analysis Summary CMS-PAS-HIG-16-004, 2016.

[31] T. J. Hastie, R. J. Tibshirani, and J. H. Friedman, "The elements of statistical learning : data mining, inference, and prediction". Springer series in statistics. Springer, New York, NY, 2. ed., corr. at 10. print. edition, 2013. ISBN 978-0-387-84857-0.

[32] P. C. Bhat, "Multivariate Analysis Methods in Particle Physics", *Annual Review of Nuclear and Particle Science* **61** (2011), no. 1, 281, `doi:10.1146/annurev.nucl.012809.104427`.

[33] A. Höcker et al., "TMVA: Toolkit for Multivariate Data Analysis", *PoS* **ACAT** (2007) 040, `arXiv:physics/0703039`.

[34] K. Kondo, "Dynamical Likelihood Method for Reconstruction of Events With Missing Momentum. 1: Method and Toy Models", *J. Phys. Soc. Jap.* **57** (1988) 4126–4140, `doi:10.1143/JPSJ.57.4126`.

[35] D0 Collaboration, "A precision measurement of the mass of the top quark", *Nature* **429** (2004) 638–642, `doi:10.1038/nature02589`, `arXiv:hep-ex/0406031`.

[36] CMS Collaboration, "The CMS experiment at the CERN LHC", *JINST* **3** (2008), no. 8, S08004, `doi:10.1088/1748-0221/3/08/S08004`.

[37] GEANT4 Collaboration, "GEANT4—a simulation toolkit", *Nucl. Instrum. Meth. A* **506** (2003), no. 3, 250, `doi:10.1016/S0168-9002(03)01368-8`.

[38] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with parton shower simulations: the POWHEG method", *JHEP* **11** (2007) 070, `doi:10.1088/1126-6708/2007/11/070`, `arXiv:0709.2092`.

[39] E. Re, "Single-top Wt-channel production matched with parton showers using the POWHEG method", *Eur. Phys. J. C* **71** (2011) 1547, `doi:10.1140/epjc/s10052-011-1547-z`, `arXiv:1009.2450`.

[40] NNPDF Collaboration, "Parton distributions for the LHC Run II", *JHEP* **04** (2015) 040, `doi:10.1007/JHEP04(2015)040`, `arXiv:1410.8849`.

[41] T. Sjöstrand et al., “An introduction to PYTHIA 8.2”, *Comput. Phys. Commun.* **191** (2015) 159, doi:10.1016/j.cpc.2015.01.024, arXiv:1410.3012.

[42] J. Alwall et al., “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”, *JHEP* **07** (2014) 079, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.

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

[44] CMS Collaboration, “Underlying event tunes and double parton scattering”, CMS Physics Analysis Summary CMS-PAS-GEN-14-001, 2014.

[45] P. Skands, S. Carrazza, and J. Rojo, “Tuning PYTHIA 8.1: the Monash 2013 Tune”, *Eur. Phys. J. C* **74** (2014), no. 8, 3024, doi:10.1140/epjc/s10052-014-3024-y, arXiv:1404.5630.

[46] N. Kidonakis, “Two-loop soft anomalous dimensions for single top quark associated production with W^- or H^-”, *Phys. Rev. D* **82** (2010), no. 5, 054018, doi:10.1103/PhysRevD.82.054018, arXiv:hep-ph/1005.4451.

[47] J. M. Campbell, R. K. Ellis, and C. Williams, “Vector boson pair production at the LHC”, *JHEP* **07** (2011) 018, doi:10.1007/JHEP07(2011)018, arXiv:1105.0020.

[48] F. Maltoni, D. Pagani, and I. Tsinikos, “Associated production of a top-quark pair with vector bosons at NLO in QCD: impact on $t\bar{t}H$ searches at the LHC”, arXiv:1507.05640.

[49] W. Beenakker et al., “Higgs radiation off top quarks at the Tevatron and the LHC”, *Phys. Rev. Lett.* **87** (2001), no. 20, 201805, doi:10.1103/PhysRevLett.87.201805, arXiv:hep-ph/0107081.

[50] W. Beenakker et al., “NLO QCD corrections to $\mathrm{t\bar{t}H}$ production in hadron collisions”, *Nucl. Phys. B* **653** (2003), no. 1-2, 151, doi:10.1016/S0550-3213(03)00044-0, arXiv:hep-ph/0211352.

[51] S. Dawson, L. H. Orr, L. Reina, and D. Wackeroth, “Associated top quark Higgs boson production at the LHC”, *Phys. Rev. D* **67** (2003), no. 7, 071503, doi:10.1103/PhysRevD.67.071503, arXiv:hep-ph/0211438.

[52] S. Dawson et al., “Associated Higgs production with top quarks at the large hadron collider: NLO QCD corrections”, *Phys. Rev. D* **68** (2003), no. 3, 034022, doi:10.1103/PhysRevD.68.034022, arXiv:hep-ph/0305087.

[53] A. Djouadi, J. Kalinowski, and M. Spira, “HDECAY: A program for Higgs boson decays in the standard model and its supersymmetric extension”, *Comput. Phys. Commun.* **108** (1998), no. 1, 56, doi:10.1016/S0010-4655(97)00123-9, arXiv:hep-ph/9704448.

[54] A. Djouadi, M. M. Mühlleitner, and M. Spira, “Decays of supersymmetric particles: The Program SUSY-HIT (SUspect-SdecaY-Hdecay-InTerface)”, *Acta Phys. Polon. B* **38** (2007) 635, arXiv:hep-ph/0609292.

[55] A. Bredenstein, A. Denner, S. Dittmaier, and M. M. Weber, "Precise predictions for the Higgs-boson decay $H \to WW/ZZ \to 4$ leptons", *Phys. Rev. D* **74** (2006), no. 1, 013004, doi:10.1103/PhysRevD.74.013004, arXiv:hep-ph/0604011.

[56] A. Bredenstein, A. Denner, S. Dittmaier, and M. M. Weber, "Radiative corrections to the semileptonic and hadronic Higgs-boson decays $H \to WW/ZZ \to 4$ fermions", *JHEP* **02** (2007) 080, doi:10.1088/1126-6708/2007/02/080, arXiv:hep-ph/0611234.

[57] M. Cacciari et al., "Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation", *Phys. Lett. B* **710** (2012), no. 4-5, 612, doi:10.1016/j.physletb.2012.03.013, arXiv:1111.5869.

[58] P. Baernreuther et al., "Percent Level Precision Physics at the Tevatron: First Genuine NNLO QCD Corrections to $q\bar{q} \to t\bar{t} + X$", *Phys. Rev. Lett.* **109** (2012), no. 13, 132001, doi:10.1103/PhysRevLett.109.132001, arXiv:1204.5201.

[59] M. Czakon and A. Mitov, "NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels", *JHEP* **12** (2012) 054, arXiv:1207.0236.

[60] M. Czakon and A. Mitov, "NNLO corrections to top-pair production at hadron colliders: the quark-gluon reaction", *JHEP* **01** (2013) 080, arXiv:1210.6832.

[61] M. Beneke et al., "Hadronic top-quark pair production with NNLL threshold resummation", *Nucl. Phys. B* **855** (2012), no. 3, 695, doi:10.1016/j.nuclphysb.2011.10.021, arXiv:1109.1536.

[62] M. Czakon, P. Fiedler, and A. Mitov, "Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\alpha_S{}^4)$", *Phys. Rev. Lett.* **110** (2013), no. 25, 252004, doi:10.1103/PhysRevLett.110.252004, arXiv:1303.6254.

[63] M. Czakon and A. Mitov, "Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders", *Comput. Phys. Commun.* **185** (2014), no. 11, 2930, doi:10.1016/j.cpc.2014.06.021, arXiv:1112.5675.

[64] CMS Collaboration, "Particle–flow event reconstruction in CMS and performance for jets, taus, and $E_{\mathrm{T}}^{\mathrm{miss}}$", CMS Physics Analysis Summary CMS-PAS-PFT-09-001, 2009.

[65] CMS Collaboration, "Commissioning of the particle-flow event reconstruction with the first LHC collisions recorded in the CMS detector", CMS Physics Analysis Summary CMS-PAS-PFT-10-001, 2010.

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

[67] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet User Manual", *Eur. Phys. J.* **C72** (2012) 1896, doi:10.1140/epjc/s10052-012-1896-2, arXiv:1111.6097.

[68] M. Cacciari, G. P. Salam, and G. Soyez, "The catchment area of jets", *JHEP* **04** (2008) 005, doi:10.1088/1126-6708/2008/04/005, arXiv:0802.1188.

[69] CMS Collaboration, "Determination of jet energy calibration and transverse momentum resolution in CMS", *JINST* **6** (2011), no. 11, P11002, doi:10.1088/1748-0221/6/11/P11002, arXiv:1107.4277.

[70] CMS Collaboration, “Identification of b-quark jets with the CMS experiment”, *JINST* **8** (2013), no. 4, P04013, doi:10.1088/1748-0221/8/04/P04013, arXiv:1211.4462.

[71] CMS Collaboration, “Identification of b quark jets at the CMS Experiment in the LHC Run 2”, CMS Physics Analysis Summary CMS-PAS-BTV-15-001, 2016.

[72] J. H. Friedman, “Stochastic gradient boosting”, *Computational Statistics & Data Analysis* **38** (2002), no. 4, 367, doi:http://dx.doi.org/10.1016/S0167-9473(01)00065-2. Nonlinear Methods and Data Mining.

[73] J. Kennedy and R. Eberhart, “Particle swarm optimization”, in *Proceedings of the IEEE International Conference on neural networks*, volume 4, pp. 1942–1948. Nov, 1995. doi:10.1109/ICNN.1995.488968.

[74] CMS Collaboration, “CMS Luminosity Measurement for the 2015 Data Taking Period”, CMS Physics Analysis Summary CMS-PAS-LUM-15-001, 2016.

[75] R. J. Barlow and C. Beeston, “Fitting using finite Monte Carlo samples”, *Comput. Phys. Commun.* **77** (1993), no. 2, 219–228, doi:10.1016/0010-4655(93)90005-W.

[76] J. S. Conway, “Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra”, in *Proceedings, PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding, CERN,Geneva, Switzerland 17-20 January 2011*. 2011. arXiv:1103.0354.

[77] A. Read, “Modified frequentist analysis of search results (the CL_s method)”, Technical Report CERN-OPEN-2000-005, CERN, 2000.

[78] T. Junk, “Confidence level computation for combining searches with small statistics”, *Nucl. Instrum. Meth. A* **434** (1999), no. 2-3, 435, doi:10.1016/S0168-9002(99)00498-2, arXiv:hep-ex/9902006.

[79] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, “Asymptotic formulae for likelihood-based tests of new physics”, *Eur. Phys. J.* **C71** (2011) 1554, doi:10.1140/epjc/s10052-011-1554-0, 10.1140/epjc/s10052-013-2501-z, arXiv:1007.1727. [Erratum: Eur. Phys. J.C73,2501(2013)].

[80] J. D. Bjorken and S. J. Brodsky, “Statistical Model for Electron-Positron Annihilation into Hadrons”, *Phys. Rev. D* **1** (Mar, 1970) 1416–1420, doi:10.1103/PhysRevD.1.1416.

[81] G. Fox and S. Wolfram, “Event shapes in e^+e^- annihilation”, *Nuclear Physics B* **157** (1979), no. 3, 543–544, doi:10.1016/0550-3213(79)90120-2.