CMS Physics Analysis Summary

Contact: cms-pag-conveners-higgs@cern.ch 2023/09/12

Measurement of the $\mathrm{t\bar{t}H}$ and tH production rates in the $\mathrm{H} \to \mathrm{b\bar{b}}$ decay channel with 138 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13\,\mathrm{TeV}$

The CMS Collaboration

Abstract

An analysis of the production of a Higgs boson (H) in association with a top quark-antiquark pair ($\mathrm{t\bar{t}H}$) or a single top quark (tH) in the channel where the Higgs boson decays into a bottom quark-antiquark pair ($\mathrm{H} \to \mathrm{b\bar{b}}$) is presented. The analysis utilises proton-proton collision data collected at the CERN LHC with the CMS experiment at $\sqrt{s} = 13\,\mathrm{TeV}$ between 2016 and 2018, which correspond to an integrated luminosity of 138 fb^{-1}. All three decay channels of the $\mathrm{t\bar{t}}$ system are considered. Various signal interpretations are performed. The observed $\mathrm{t\bar{t}H}$ production rate relative to the standard model (SM) expectation is $0.33 \pm 0.26 = 0.33 \pm 0.17\,(\mathrm{stat}) \pm 0.21\,(\mathrm{syst})$. Additionally, the $\mathrm{t\bar{t}H}$ production rate is determined in intervals of Higgs boson p_{T}. An upper limit at 95% confidence level on the tH production rate of 14.6 times the SM expectation is observed, with an expectation of $19.3^{+9.2}_{-6.0}$. Finally, constraints are derived on the strength and structure of the coupling between the Higgs boson and the top quark from simultaneous extraction of the $\mathrm{t\bar{t}H}$ and tH production rates.

This document has been revised with respect to the version dated August 22nd, 2023.

1 Introduction

In the standard model (SM) of particle physics, the Higgs boson couples to fermions with a Yukawa-type interaction, with a coupling strength proportional to the fermion mass. Probing the coupling of the Higgs boson to the top quark, the heaviest known fermion, is therefore instrumental in testing the SM and constraining models of physics beyond the SM, which may predict different coupling strengths and structures. The associated production of a Higgs boson and a top quark-antiquark pair ($\mathrm{t\bar{t}H}$) provides a direct probe of the top-Higgs coupling as illustrated by the Feynman diagram in Fig. 1 (left), and has recently been observed by the ATLAS and CMS Collaborations [1, 2]. Moreover, Higgs boson decays to pairs of bottom quarks have also been observed [3, 4], thereby directly probing the Yukawa interactions between the Higgs boson and top as well as bottom quarks for the first time. The associated production of a single top quark and a Higgs boson (tH) proceeds via two groups of processes, in which the Higgs boson couples either to the top quark or to the W boson [5–8], as illustrated in Fig. 1 (centre and right). Therefore, tH production provides insight on the coupling of the Higgs boson to top quarks and to vector bosons, and also the interference of the two, and thus, tH production allows probing the relative sign of the two couplings. In the SM, the Higgs boson has a positive eigenvalue under combined charge-parity (CP) transformations ("CP even"). The combined measurement of $\mathrm{t\bar{t}H}$ and tH production provides a probe of CP-odd admixtures in the top-Higgs coupling through interfering contributions, which would be a clear sign of physics beyond the SM.

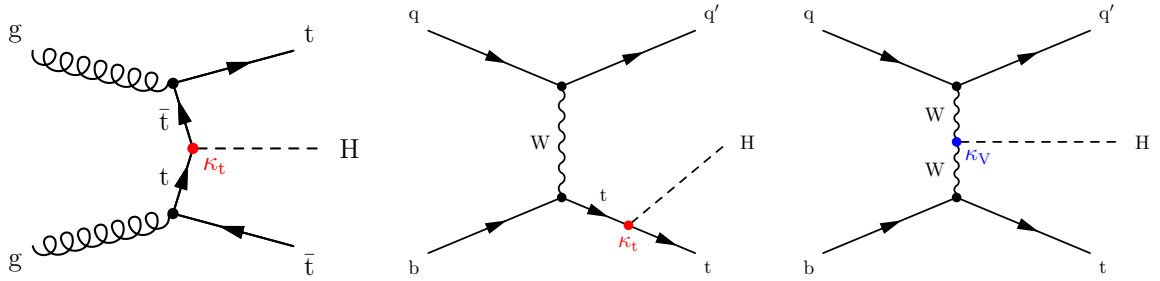


Figure 1: Representative leading-order Feynman diagrams for the associated production of a Higgs boson and a top quark-antiquark pair (left) and for the associated production of a single top quark and a Higgs boson in the *t* channel, where the Higgs boson couples to the top quark (centre) or the W boson (right).

In the SM, for a 125 GeV Higgs boson, the $\mathrm{H} \to \mathrm{b\bar{b}}$ decay has the largest branching fraction of 0.58 ± 0.02 [9] and a fully reconstructable final state. It is thus experimentally attractive as a final state, yet challenging due to considerable SM backgrounds at the LHC. Existing searches for $\mathrm{t\bar{t}H}$ and tH production in the $\mathrm{H} \to \mathrm{b\bar{b}}$ channel by the ATLAS and CMS Collaborations achieve sensitivities that correspond to observed (expected) significances of up to 1.6 (2.7) standard deviations (SD) for $\mathrm{t\bar{t}H}$ production [10–12] and observed (expected) upper limits at the 95% confidence level (CL) on SM tH production of approximately 90 (41) times the SM expectation [13]. Constraints on the CP nature of the top-Higgs coupling have been obtained in this channel from the combination of information from $\mathrm{t\bar{t}H}$ and tH production by the ATLAS Collaboration [14].

This note describes the combined analysis of $\mathrm{t\bar{t}H}$ and tH production in the $\mathrm{b\bar{b}}$ decay channel of the Higgs boson by the CMS Collaboration. Proton-proton (pp) collision data collected during LHC Run 2 from 2016 to 2018 at a centre-of-mass energy of 13 TeV are used, corresponding to an integrated luminosity of $138\,\mathrm{fb}^{-1}$. All decay channels of the $\mathrm{t\bar{t}}$ system are considered.

Several signal interpretations are performed. First, targeting solely the $\mathrm{t\bar{t}H}$ signal, the $\mathrm{t\bar{t}H}$ production rate is determined, both inclusively and in different regions of transverse momentum of the Higgs boson, following the Simplified Template Cross Section (STXS) approach [9]. Second, targeting the tH signal process, an upper limit on the tH production rate is determined assuming the $\mathrm{t\bar{t}H}$ production rate predicted in the SM. Third, targeting simultaneously the $\mathrm{t\bar{t}H}$ and tH production processes, the Higgs boson coupling to top quarks is analysed, and both the strength and the CP structure of the coupling are probed.

The analysis builds upon the strategies developed in previous analyses of $\mathrm{t\bar{t}H}$ production [11, 12] and tH production [13] using 2016 data. The event selection aims at identifying events in which a Higgs boson is produced in association with a $\mathrm{t\bar{t}}$ pair or a single top quark, and decays to $\mathrm{b\bar{b}}$. Three mutually exclusive channels are considered in the analysis, targeting the different $\mathrm{t\bar{t}}$ decay modes: the fully-hadronic (FH) channel, in which both W bosons decay to a $\mathrm{q\bar{q}}$ pair, the single-lepton (SL) channel, with one W boson decaying to a charged lepton and a neutrino, and the dilepton (DL) channel, with both W bosons decaying to a charged lepton and neutrino. The SL channel also includes dedicated event categories targeting tH production. Hadronically decaying tau leptons are not considered in this analysis, while tau leptons decaying to an electron or muon plus neutrinos can enter the selection in the SL and DL channels. The experimental signature of signal events is characterised by the presence of high-p_{T} b quark jets (b jets) from the Higgs boson and top quark decays and, depending on the channel, jets or isolated electrons and muons as well as missing transverse momentum arising from the presence of undetected neutrinos from the W boson decays. Dominant background contributions arise from quantum chromodynamics (QCD) multijet production (referred to as "QCD background") in the FH channel and from $\mathrm{t\bar{t}}+\text{jets}$ production in all channels. The latter includes in particular $\mathrm{t\bar{t}}+\mathrm{b\bar{b}}$ production, where additional b quarks can arise from QCD radiation or loop-induced QCD processes. The $\mathrm{t\bar{t}}+\mathrm{b\bar{b}}$ background remains almost irreducible with respect to $\mathrm{t\bar{t}H}$, $\mathrm{H} \to \mathrm{b\bar{b}}$, with both processes having four b quarks in the final state, and constitutes the critical background to the analysis.

The sensitivity of the analysis is limited by combinatorial background due to multiple b jets in the final state, with no unambiguous way of reconstructing the invariant mass peak of the Higgs boson. Therefore, the signal is extracted exploiting a multivariate analysis based on artificial neural networks (ANNs), which are optimised on simulated events or data events in control regions and which use the information of several experimental observables simultaneously. The ANNs are used to discriminate signal from background events and, in the SL and DL channels, to categorise events into signal and background control regions. The signal rate is obtained from a combined profile likelihood fit of distributions of the ANN score, distributions of ratios of ANN scores, or of the event yield to the data, depending on the channel and category, correlating processes and their uncertainties where appropriate.

Several changes in the analysis strategy have been adopted with respect to the previous analyses [11–13], including usage of state-of-the-art $\mathrm{t\bar{t}}+\mathrm{b\bar{b}}$ event simulation in the background modelling, updates in the multivariate analysis techniques, improvements in the b tagging algorithm, and the exploitation of all three $\mathrm{t\bar{t}}$ decay channels.

This document is structured as follows. The CMS detector is described in Section 2. In Section 3, the data samples and trigger selection are described, followed by a description of the simulated samples in Section 4. The object and event reconstruction are presented in Section 5 and the event selection in Section 6. The modelling of the critical $\mathrm{t\bar{t}}+\text{jets}$ background is detailed in Section 7. The analysis strategy in the different channels is detailed in Section 8 and systematic uncertainties are discussed in Section 9. Finally, details of the statistical analysis and the results

are presented in Section 10.

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 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-ionisation 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. [15].

Events of interest are selected using a two-tiered trigger system. The first level, composed of custom hardware processors, uses information from the calorimeters and muon detectors to select events at a rate of around 100 kHz within a fixed latency of about 4 μs [16]. The second level, known as the high-level trigger, consists of a farm of processors running a version of the full event reconstruction software optimised for fast processing, and reduces the event rate to around 1 kHz before data storage [17].

3 Data samples and trigger selection

The analysis uses pp collision data recorded at $\sqrt{s} = 13$ TeV at the LHC during 2016–2018. Only the data-taking periods during which the CMS detector was fully operational are included in the analysis. The total integrated luminosity of the analysed data amounts to 138 fb^{-1}, of which 36.3 fb^{-1}, 41.5 fb^{-1}, and 59.7 fb^{-1} have been recorded in 2016, 2017, and 2018, respectively [18–20]. The data were selected using different trigger paths that require the presence of several jets or one or two leptons and, in some cases, a large scalar sum of jet p_T (H_T) in the events, depending on the analysis channel.

Novel trigger paths were developed at the beginning of the LHC Run 2, targeting events in the FH channel. They require the presence of a high number of jets, large H_T, and one, two, or three jets tagged by the online b jet identification. In order to recover trigger inefficiencies, also events recorded by a single-jet trigger with a high p_T threshold or by a trigger with a high H_T threshold are considered. The selection criteria are detailed in Table 1.

In the leptonic channels, the trigger strategy relies on the presence of isolated muons and electrons in the final state. In the SL channel, events were selected by single-lepton triggers requiring the presence of one muon or electron exceeding a given p_T threshold. The threshold values span from 24 to 27 GeV for muons, and from 27 to 32 GeV for electrons, depending on the data-taking period and running conditions. In order to compensate for the increased p_T threshold of the single-electron trigger in 2017 and 2018, a trigger requiring both an electron with $p_T > 28$ GeV and $H_T > 150$ GeV was also used. In the DL channel, events were selected online by dilepton (ee, $\mu\mu$, eμ) triggers, complemented by the single-lepton triggers mentioned above to maximise the selection efficiency. The same-flavour dilepton triggers required the presence of two muons with $p_T > 17$ and 8 GeV and, for some running periods during 2017 and 2018, an invariant di-μ mass $m_{\mu\mu} > 3.8$ GeV, or two electrons with $p_T > 23$ and 12 GeV. The eμ triggers instead required either a muon with $p_T > 23$ and an electron with $p_T > 12$ GeV, or an electron with $p_T > 23$ GeV and a muon with $p_T > 8$ GeV. Lepton identification and isolation requirements are imposed for some of the single lepton and dilepton triggers, and some of the

Table 1: Trigger selection criteria in the fully-hadronic (FH) channel. Multiple triggers, each represented by one row, are used per year and combined with a logical OR. In the case of the four-jet trigger, the minimum jet p_T is different for each jet and separated by a slash (/).

Year	Min. number jets	Min. number b-tagged jets	Min. jet p_T (GeV)	Min. H_T (GeV)
2016	6	1	40	450
	6	2	30	400
	1	—	450	—
2017	6	1	40	430
	6	2	32	380
	4	3	75/60/45/40	300
	—	—	—	1050
2018	6	1	36	450
	6	2	32	400
	4	3	75/60/45/40	330
	—	—	—	1050

dilepton triggers also impose an impact parameter requirement.

During the 2016 and 2017 data-taking, a gradual shift in the timing of the inputs of the ECAL L1 trigger in the region at $|\eta| > 2$, referred to as "L1 prefiring", caused a specific trigger inefficiency. For events containing an electron (a jet) with $p_T \gtrsim 50\,\text{GeV}$ ($\gtrsim 100\,\text{GeV}$), the efficiency loss is approximately 10–20% in the region $2.5 < |\eta| < 3$, depending on p_T, η, and time. Correction factors were computed from data and applied to the acceptance evaluated by simulation.

4 Simulation samples

Several Monte Carlo (MC) event generators, interfaced to a detailed detector simulation based on GEANT4 (v. 9.4) [21], are used to model signal and background events. Separate samples corresponding to the 2016, 2017, and 2018 data-taking conditions were produced in order to match the different LHC and detector conditions and reconstruction efficiencies. Events are simulated at next-to-leading order (NLO) accuracy of QCD perturbation theory with the POWHEG (v. 2) [22–26] or MADGRAPH5_aMC@NLO (v. 2.4.2) [27] event generator, or at leading order (LO) accuracy using PYTHIA (v. 8.230) [28], depending on the process. 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 function (PDF) set NNPDF3.1 [29]. Parton showering (PS) and hadronisation are simulated with PYTHIA [28]. The parameters for the underlying event description correspond to the CP5 tune [30] for all signal and background processes, except for the minor background processes in the samples matching the 2016 conditions, where the CUETP8M1 [31] tune has been used.

The $t\bar{t}H$ signal is simulated at NLO accuracy with POWHEG and, solely for the coupling interpretation, at LO with MADGRAPH5_aMC@NLO. The tH signal is simulated at LO with MADGRAPH5_aMC@NLO. Different flavour schemes are chosen to simulate the tHq and tHW processes. In the five-flavour scheme (5 FS), bottom quarks are considered as massless sea quarks of the proton and may appear in the initial state of pp scattering processes. In the four-flavour scheme (4 FS), bottom quarks are considered massive and are produced by gluon splitting at the matrix-element (ME) level, and therefore appear only in the final state [32]. In the 5 FS, the distinction of tHq, *s* channel (tHb), and tHW contributions to tH production is well defined up to NLO, while at higher orders in perturbation theory the tHq and *s* channel production

processes start to interfere and can no longer be uniquely separated [8]. Similarly, in the same regime the tHW process starts to interfere with $t\bar{t}H$ production at NLO. In the 4 FS, the separation among the tHq, *s* channel, and tHW (if the W boson decays hadronically) processes holds only up to LO, and the tHW process starts to interfere with $t\bar{t}H$ production already at tree level [8]. The tHq process is simulated at LO in the 4 FS and the tHW process in the 5 FS such that interference contributions of the latter with $t\bar{t}H$ production are not present in the simulation. The interference effects are negligible in the phase space selected in this analysis [33]. The contribution from *s* channel tH production is negligible and is not considered in this analysis. In both the $t\bar{t}H$ and tH simulated samples, all decay channels of the Higgs boson are considered. In the case of the Higgs boson coupling analysis, the kinematic properties of the tH events vary depending on the probed coupling strength, and the kinematic properties of both the tH and the $t\bar{t}H$ events vary depending on the probed CP structure. This is accounted for by event weights, following the approach suggested in Refs. [34, 35].

Major background contributions arise from $t\bar{t}$ + jets production. They are modelled with simulated events obtained from two different generators implemented in the POWHEG framework and interfaced to PYTHIA for the PS simulation. The version and configuration of the two generators are summarised in Table 2. The POWHEG generator is used to simulate the $pp \rightarrow t\bar{t}$ process at NLO accuracy in the 5 FS (referred to as $t\bar{t}$ sample), while the dedicated POWHEG-BOX-RES program presented in Ref. [36] together with OPENLOOPS [37] is used to simulate the $pp \rightarrow t\bar{t}b\bar{b}$ process at NLO accuracy in the 4 FS (referred to as $t\bar{t}b\bar{b}$ sample). In the $t\bar{t}b\bar{b}$ sample, the additional b jets that do not stem from the top quark decays and that are critical for the background description are simulated at the ME level and thus expected to model more accurately the data. The choice of the renormalisation and factorisation scales (μ_R, μ_F) for the $t\bar{t}b\bar{b}$ sample follows Refs. [9, 36, 38] and is motivated by the presence of two very different relevant scales related to the $t\bar{t}$ and $b\bar{b}$ systems. Compared to Ref. [36], the values of the scales are reduced by a factor 0.5 in order to approximate the effect of missing higher-order corrections and attenuate theoretical uncertainties related to the ME-PS matching. This choice is motivated by the suggested scale options presented in Ref. [39] that were derived comparing the predictions of $t\bar{t}b\bar{b}$ calculations to those of $t\bar{t}b\bar{b}$ + 1 jet calculations at NLO in QCD. The PDF sets have been chosen consistently to the flavour scheme, with values of the strong coupling constant α_S matching the order of the ME generation in both cases. The same PYTHIA version and underlying event tune have been used for both samples. The combination of the two samples and the details of the $t\bar{t}$ + jets background model are described in Section 7.

Minor backgrounds originate from single top quark production (tW and *t* channel production), the production of W and Z/γ^* bosons with additional jets (referred to as V+jets), $t\bar{t}$ production in association with a W or Z boson (referred to as $t\bar{t}V$), and diboson (WW, WZ, and ZZ, referred to as VV) processes. The single top quark processes in the *t* and tW channels are simulated with POWHEG [40, 41]. The *s* channel single top quark processes as well as V+jets and $t\bar{t}V$ processes are simulated with MADGRAPH5_aMC@NLO, where for the V+jets processes the matching of ME to PS jets is performed using the FXFX [42] prescription. Diboson production is simulated using the PYTHIA event generator.

In the FH channel, the dominant background originates from QCD multijet production. Simulated QCD multijet events, generated with MADGRAPH5_aMC@NLO at LO accuracy using the MLM [43] prescription for the ME–PS matching, are employed solely for additional validation of observables and procedures. For the actual analysis, its validation and optimisation, the QCD background contribution is estimated from data as described in Section 8.

For comparison with the observed distributions, the event yields in the simulated samples are

Table 2: Generator version and configuration of the $\mathrm{t\bar{t}}$ and $\mathrm{t\bar{t}b\bar{b}}$ samples. The parameters m_t and m_b denote the top quark and bottom quark mass, respectively, $m_{\mathrm{T,t}}$, $m_{\mathrm{T,b}}$, and $m_{\mathrm{T,g}}$ the transverse mass of the top quark, the bottom quark, and additional gluons, respectively, and h_damp the parton-shower matching scale.

	$\mathrm{t\bar{t}}$ sample	$\mathrm{t\bar{t}b\bar{b}}$ sample
POWHEG version	`Powheg v2`	`Powheg-Box-Res`
PYTHIA version	8.230	8.230
Flavour scheme	5	4
PDF set	NNPDF3.1	NNPDF3.1
m_t	172.5 GeV	172.5 GeV
m_b	0	4.75 GeV
μ_R	$\sqrt{\frac{1}{2}\left(m_{\mathrm{T,t}}^2 + m_{\mathrm{T,\bar{t}}}^2\right)}$	$\frac{1}{2}\sqrt[4]{m_{\mathrm{T,t}} \cdot m_{\mathrm{T,\bar{t}}} \cdot m_{\mathrm{T,b}} \cdot m_{\mathrm{T,\bar{b}}}}$
μ_F	μ_R	$\frac{1}{4}\left[m_{\mathrm{T,t}} + m_{\mathrm{T,\bar{t}}} + m_{\mathrm{T,b}} + m_{\mathrm{T,\bar{b}}} + m_{\mathrm{T,g}}\right]$
h_damp	$1.379 \cdot m_\mathrm{t}$	$1.379 \cdot m_\mathrm{t}$
Tune	CP5	CP5

normalised to the integrated luminosity of the corresponding data sample, according to their predicted cross sections. The SM $\mathrm{t\bar{t}H}$ cross section of 507^{+35}_{-50} fb is taken from calculations including QCD and electroweak corrections at NLO [9]. The SM cross section for tH production is computed in the 5 FS at NLO accuracy in QCD [9], which results in $74.3^{+5.6}_{-11.3}$ fb for tHq and $15.2^{+1.2}_{-1.4}$ fb for tHW production. The DR2 scheme [33] is employed in the calculation of tHW production to remove overlapping contributions between tHW and $\mathrm{t\bar{t}H}$ processes. Higgs boson branching fractions are obtained from calculations with at least NLO accuracy [9]. The calculated $\mathrm{t\bar{t}H}$ and tH cross sections and Higgs boson branching fractions are rescaled assuming a Higgs boson mass of 125.38 GeV, corresponding to the latest measurement by the CMS Collaboration [44]. The $\mathrm{t\bar{t}}$ cross section of 831.76 pb corresponds to the next-to-next-to-leading order (NNLO) calculation with resummation to next-to-next-to-leading-logarithmic (NNLL) accuracy [45–51]. The cross sections of the other background processes are taken at NNLO (V+jets), approximate NNLO (single top quark tW channel [52]), and NLO (single top quark t and s channels [53, 54], $\mathrm{t\bar{t}V}$ [55], and diboson [56]) accuracy.

Effects from additional pp interactions in the same or nearby bunch crossings (pileup) are modelled by adding simulated minimum-bias events to all simulated events. The pileup multiplicity distribution in simulation is reweighted to reflect the luminosity profile of the observed pp collisions. Further correction factors described in Section 5 are applied to the simulation where necessary to improve the description of the data.

5 Object and event reconstruction

Events are reconstructed offline based on a particle-flow (PF) technique [57], which aims to reconstruct and identify each individual particle produced in pp collisions by optimally combining information from the various elements of the CMS detector. The primary vertex is taken to be the vertex corresponding to the hardest scattering in the event, evaluated using tracking information alone, as described in Section 9.4.1 of Ref. [58].

The reconstruction of muons relies on a combination of measurements in the tracker and in the muon detectors [59], and electrons are reconstructed by combining the energy measurement in the ECAL with the momentum measurement in the tracker [60, 61]. Both electrons and muons are required to lie within the acceptance of the tracker, covering the region up to $|\eta| = 2.4$. Electrons reconstructed in the transition region $1.4442 < |\eta| < 1.556$ between the barrel and the endcap calorimeters are discarded. Isolation requirements based on the energy deposited in a cone of radius $\Delta R = 0.3$ (0.4) around the track direction of the electron (muon), where $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$, are defined as follows. The scalar p_T-sum of all charged particles within the cone and matched to the primary vertex is calculated and corrected for the neutral component expected from pileup interactions. The latter is taken as the average transverse energy deposited by neutral particles in the case of electrons, and half the transverse momentum carried by charged particles originating from pileup vertices in the case of muons, evaluated on an event-by-event basis. The isolation variable obtained in this way is divided by the lepton p_T to define a relative isolation discriminant I_{rel}. In the SL channel, muons are required to pass a tighter isolation requirement ($I_{rel} < 0.15$) than in the DL channel ($I_{rel} < 0.25$). The latter criterion is also applied in the FH channel to veto events with isolated muons. For electrons, isolation criteria are defined separately for the barrel and endcap regions, together with other quality criteria based on the energy measurement in the ECAL and HCAL, compatibility between the tracker and ECAL momentum measurement, missing tracker hits, as well as criteria to reject electrons from photon conversion. Tight identification requirements are imposed for electrons in all analysis channels, corresponding to an efficiency of approximately 70%.

Hadronic jets are reconstructed by clustering PF candidates using the anti-k_T algorithm [62, 63] with a distance parameter of 0.4, omitting charged particles matched to pileup vertices. The jet energy is corrected for the neutral-hadron contribution expected from pileup interactions [64]. Further energy corrections depending on the jet p_T and η are applied, which are derived in simulation such that the average measured energy of the jets becomes identical to that of the particle level jets. Residual differences between the jet energy scale in data and in simulation are measured using the momentum balance in dijet, γ + jets, Z + jets, and multijet events, and appropriate corrections are made. The jet energy resolution amounts typically to 15–20% at 30 GeV, 10% at 100 GeV, and 5% at 1 TeV [65]. Jets overlapping with an electron or muon passing the criteria described above within a cone of $\Delta R = 0.4$ are discarded. Further selection criteria are applied to each jet to remove jets potentially dominated by instrumental effects or reconstruction failures as well as jets arising from pileup interactions. Only jets within the tracker acceptance of $|\eta| < 2.4$ and with $p_T > 30$ GeV are considered in the analysis. In addition, jets with $p_T > 40$ GeV and $2.4 < |\eta| < 4.7$ (forward jets) are selected in the SL channel for the event reconstruction under the hypothesis of tH production, which is characterised by the presence of jets with large $|\eta|$ values.

Jets arising from the hadronisation of b quarks (b jets) are identified using b tagging techniques [66]. This analysis benefits from the DeepJet [67, 68] tagger, featuring an increase of the b tagging efficiency by 5–10% with respect to the tagger used in the previous analyses [11–13] for the same mistag rate. Jets are considered as b tagged if they fulfil a requirement on the DeepJet score that corresponds to a b tagging efficiency of about 75–80% and 1.5–2% mistag rate for light-flavour jets. Specifically for the estimation of the QCD background, a second requirement on the DeepJet output is utilised, which corresponds to approximately 90% b tagging efficiency at a 10% light-flavour mistagging rate, referred to as "loose" b tag.

The DeepJet discriminant value is further used in the analysis as one of the input variables for the ANN discriminants as well as in a b tagging likelihood ratio (BLR) criterion that is designed to discriminate between events compatible with the presence of four and two b jets. The

likelihoods are computed under either the hypothesis that four jets or that two jets in the event originate from b quarks, based on the expected b tagging discriminant probability densities from simulation. The BLR is computed as the ratio of the four-b-jets likelihood relative to the sum of the four- and the two-b-jets likelihoods.

The missing transverse momentum vector $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ is computed as the negative vector sum of the transverse momenta of all the PF candidates in an event, and its magnitude is denoted as $p_{\mathrm{T}}^{\mathrm{miss}}$ [69]. The $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ is recomputed to account for corrections to the energy scale of the reconstructed jets in the event.

6 Event selection

Events are divided into mutually exclusive categories according to the number of leptons, jets, and b-tagged jets. The first step of the classification, referred to as baseline selection, defines the events assigned to the DL, SL, and FH channels, and is summarised in Table 3 and described in the following.

Table 3: Baseline selection criteria in the fully-hadronic (FH), single-lepton (SL), and dilepton (DL) channels based on the observables defined in the text. Where the criteria differ per year of data taking, they are quoted as three values, corresponding to 2016/2017/2018, respectively.

	FH channel	SL channel	DL channel
Number of leptons	0	1	2
Sign and flavour of leptons	—	$\mathrm{e}^{\pm}$, $\mu^{\pm}$	$\mathrm{e}^{+}\mathrm{e}^{-}$, $\mu^{\pm}\mathrm{e}^{\mp}$, $\mu^{+}\mu^{-}$
Min. p_{T} of leading electron (GeV)	—	29/30/30	25
Min. p_{T} of leading muon (GeV)	—	26/29/26	25
Min. p_{T} of additional leptons (GeV)	—	—	15
Max. p_{T} of additional leptons (GeV)	15	15	—
Max. $\lvert\eta\rvert$ of leptons	2.4	2.4	2.4
Min. $m_{\ell\ell}$ (GeV)	—	—	20
$m_{\mathrm{ee}/\mu\mu}$ (GeV)	—	—	< 76 or > 106
Min. number of jets	7	5	3
Min. p_{T} of jets (GeV)	30	30	30
Min. p_{T} of 6^{th} jet (GeV)	40	—	—
Max. $\lvert\eta\rvert$ of jets	2.4	2.4	2.4
Min. number of b-tagged jets	2	4	3
m_{qq} (GeV)	> 30 and < 250	—	—
Min. H_{T} (GeV)	500	—	—
Min. $p_{\mathrm{T}}^{\mathrm{miss}}$ (GeV)	—	20	40

Events with exactly two opposite-sign leptons (electron or muon) with p_{T} of at least 25 GeV (15 GeV) for the leading (sub-leading) lepton and at least three b-tagged jets are assigned to the DL channel. The offline and online flavour content of the event must be consistent, meaning for example that $\mu\mu$ events are required to be recorded by single or double-muon triggers. The invariant mass of the selected lepton pair, $m_{\ell\ell}$, is required to be larger than 20 GeV to suppress events from heavy-flavour resonance decays and low-mass Drell–Yan processes. In order to reject Z + jets events, same-flavour events with 76 GeV $< m_{\ell\ell} <$ 106 GeV are discarded.

The baseline selection for the SL channel requires the presence of exactly one electron or muon, consistent with the trigger that accepted the event. The lepton p_T requirement is chosen based on the trigger thresholds, and therefore depends on the data-taking year. Muons must have p_T >26 (29) GeV for data recorded in 2016 and 2018 (2017), while the p_T thresholds for electrons are 29 (30) GeV in 2016 (2017 and 2018). Furthermore, only events with at least five jets, of which at least four are b tagged, are considered. To further suppress the background contribution, a minimum p_T^{miss} of 40 GeV and 20 GeV is required in the DL channels with same-flavour leptons and in the SL channel, respectively, as neutrinos from leptonic W boson decays remain undetected.

All remaining events with no leptons passing the selection criteria in the DL and SL channel and with at least seven jets, of which at least two are b tagged, are assigned to the FH channel. In order to match the online trigger selection, events with $H_T < 500$ GeV or less than six jets with $p_T > 40$ GeV are discarded. The invariant mass m_{qq} of the pair of non-b-tagged jets for which m_{qq} is closest to the nominal W boson mass is required to lie within 30 and 250 GeV. To perform a data-driven estimation of the dominant QCD background in this channel, events are further split into a signal region and several control and validations regions according to the value of m_{qq}, as described in Section 8.

Figure 2 shows the jet multiplicity distribution for the FH, SL, and DL channels after the baseline selection and prior to the final fit to data. The expected signal and background contributions are estimated using the simulations for the various processes introduced in Section 4. Good agreement between data and simulation is observed in the SL and DL channels within the uncertainties. In the FH channel, where the dominant background contribution stems from QCD, a significant trend in data is observed towards higher jet multiplicity; the poor modelling of the QCD background by the simulation is expected and the reason it is estimated directly from data.

The further event categorisation within each channel and the strategy adopted to extract the signal are described in Section 8. To maximise the analysis sensitivity, a likelihood ratio comparing the $t\bar{t}H$, $H \to b\bar{b}$ signal, and the $t\bar{t} + b\bar{b}$ background hypotheses is computed based on a matrix element method [70] and provided as input to the final ANN discriminants. Furthermore, in the SL channel boosted decision trees (BDTs), referred to as reconstruction BDTs, are used to assign the jets in an event to partons under the hypothesis of the event being either a $t\bar{t}H$, tHq, tHW, or $t\bar{t}$ event. The BDTs take as input the jet kinematic variables and are trained separately for each hypothesis, such that the most probable jet assignment under the given hypothesis yields the largest BDT output value. In case of the tHq and tHW hypotheses, the forward jets are also taken into account in the event reconstruction as candidates for the light-flavour quarks, which typically have large $|\eta|$ values in tHq and tHW events. The procedure achieves reconstruction efficiencies of approximately 43%, 60%, 52%, and 66% for $t\bar{t}H$, tHq, tHW, or $t\bar{t}$ events, respectively. The output values of the reconstruction BDT as well as several observables that are obtained based on the chosen jet assignments are among the input variables to the ANN used for the signal extraction in this channel.

7 $t\bar{t}$ + jets background model

After the baseline selection, major background contributions arise from $t\bar{t}$ + jets production, which constitute more than 95% of the total expected background events in the leptonic channels. Particularly critical among these are $t\bar{t}$ events with additional b jet production, since their topology is identical to that of the $t\bar{t}H$ signal events, the cross section is approximately 10–20 times larger than that of the signal in the phase space considered, and the events are difficult

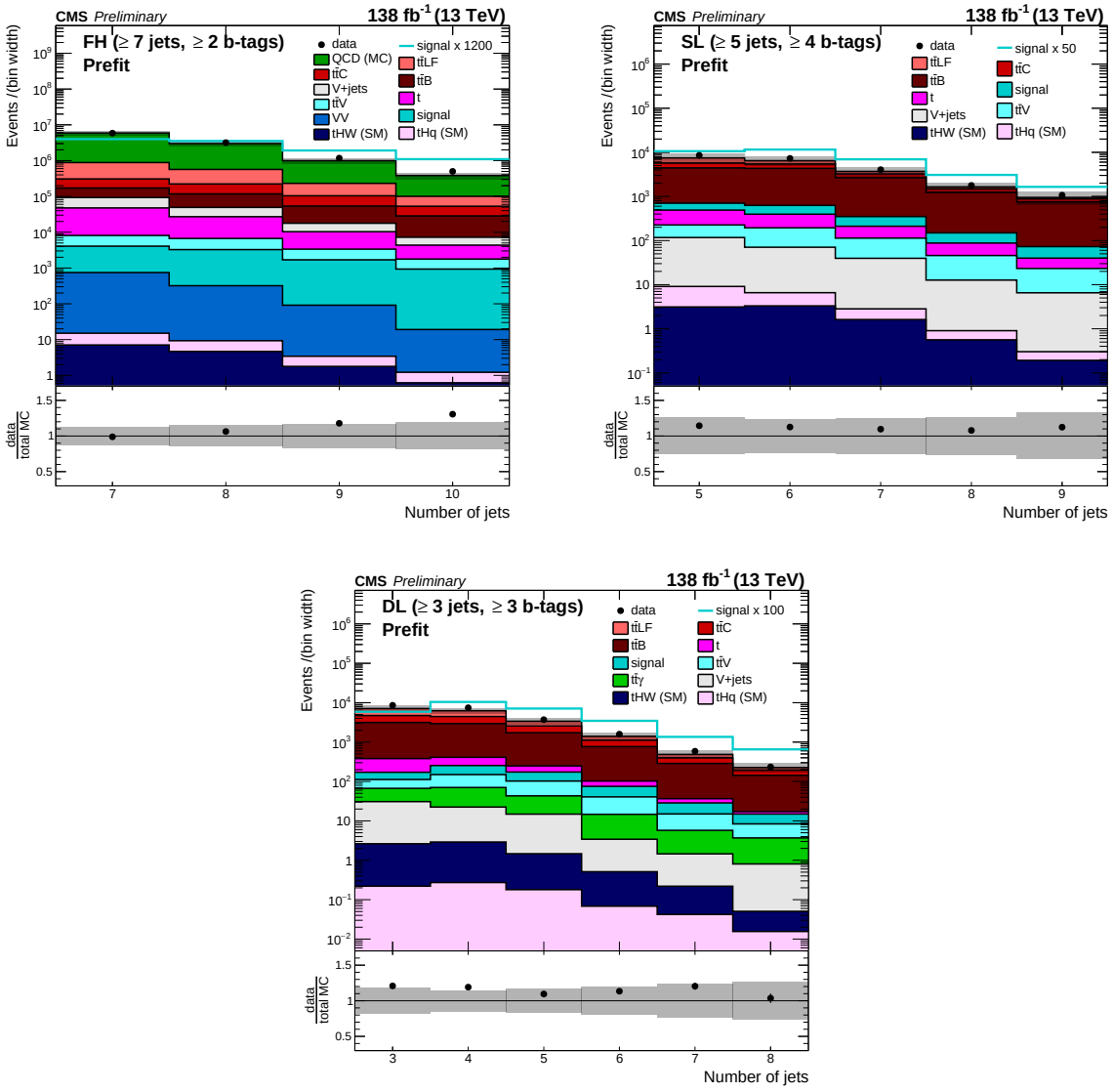


Figure 2: Jet multiplicity distribution in the FH (upper left), SL (upper right), and DL (lower) channels, after the baseline selection and prior to the fit to the data. Here, the QCD background prediction is taken from simulation. The expected $t\bar{t}H$ signal contribution, scaled as indicated in the legend for better visibility, is also overlayed (line). The uncertainty band represents the total uncertainty. The last bins include overflow events.

to model theoretically due to the multi-parton final state with two very different mass scales. Several dedicated measurements [71–73] as well as the observations in previous versions of this analysis [11] and the observations in the control regions studied in this analysis (see Section 8) show that the $t\bar{t}$ + b jets contribution present in data is larger than predicted by the simulation, by a factor of typically 1.2–1.4, depending on the MC event generator.

The $t\bar{t}$ + jets background is modelled using a combination of the $t\bar{t}$ and $t\bar{t}b\bar{b}$ samples described in Section 4. The simulated $t\bar{t}$ + jets events are separated into the following mutually exclusive processes, based on the flavour of the additional jets at the particle level that do not originate from the top quark decays and that fulfil the acceptance requirements of $p_T > 20\,\text{GeV}$ and of $|\eta| < 2.4$: $t\bar{t}B$, comprising events with at least one additional jet generated within the acceptance that contains one or more b hadrons; $t\bar{t}C$, for which events have at least one additional jet containing c hadrons within the acceptance and no additional jets containing b hadrons; or else $t\bar{t}$ + light flavour jets ($t\bar{t}LF$), which corresponds to events that do not belong to any of the above processes. Furthermore, out of the $t\bar{t}B$ events, the subset of events with exactly two additional b hadrons that are close enough in direction to be inside a single jet is denoted as $t\bar{t}$ + 2b. This separation is important because the processes are subject to different systematic uncertainties arising from the modelling of collinear gluon splitting.

The $t\bar{t}B$ background events, which represent the most critical background component, are taken from the $t\bar{t}b\bar{b}$ sample. Here, the additional b jets are modelled by the ME calculation and are subject, in particular, to the renormalisation and factorisation scale uncertainties. The $t\bar{t}C$ and $t\bar{t}LF$ events, which do not include additional b jets within acceptance but can still pass the event selection (e.g. due to mistagging), are taken from the $t\bar{t}$ sample. Prior to the fit to data, the fractional contributions of the $t\bar{t}B$, $t\bar{t}C$, and $t\bar{t}LF$ events are chosen according to the $t\bar{t}$ sample and the total yield of the $t\bar{t}$ + jets events is scaled to the NNLO+NNLL $t\bar{t}$ cross section of 831.76 pb. The final yields of the $t\bar{t}B$ and $t\bar{t}C$ contributions are free parameters in the fit to data, providing flexibility to adjust the predicted $t\bar{t}B$ yield to the one observed in data through dedicated control regions. The impact of potential mismodellings of the $t\bar{t}B$ background on the extracted signal was studied using pseudo experiments and found to be well covered by the systematic uncertainties, as described in Section 9.

The modelling of the $t\bar{t}B$ background has been validated in data in various observables and relevant phase space regions. For example, the predicted jet multiplicity distributions, shown in Fig. 2 for the different analysis channels after the baseline selection and prior to the fit to data, agree well with the ones observed in data, within the uncertainties. Differences observed in the FH channel are attributed to mismodelling of the dominant QCD background contribution, which is taken from simulation in this figure. Further example distributions are shown in Figs. 3 and 4 for events in signal-enriched regions of high jet and b-tagged jet multiplicity: the average $\Delta\eta$ and the minimum ΔR between any two b-tagged jets, which are sensitive to the $t\bar{t}B$ modelling details; the MEM discriminant output, which is one of the most powerful single variables used in the analysis to separate the $t\bar{t}H$ signal from the $t\bar{t}B$ background; and the p_T of the Higgs boson candidate identified using the aforementioned reconstruction BDT or as the pair of b-tagged jets closest in ΔR, which is relevant specifically for the STXS measurement. Prior to the fit to data, the shape of the data distribution is well described by the simulation. Differences in the normalisation are attributed to the expected under-prediction of the $t\bar{t}B$ yield. With the postfit background model obtained from the final fit to data described in Section 10, the predicted distributions agree well with the ones observed in data. The effect of the fit on these observables mostly changes the relative normalisation of each component, reduces the corresponding uncertainties, and slightly modifies the shapes of the distributions within the associated systematic uncertainties.

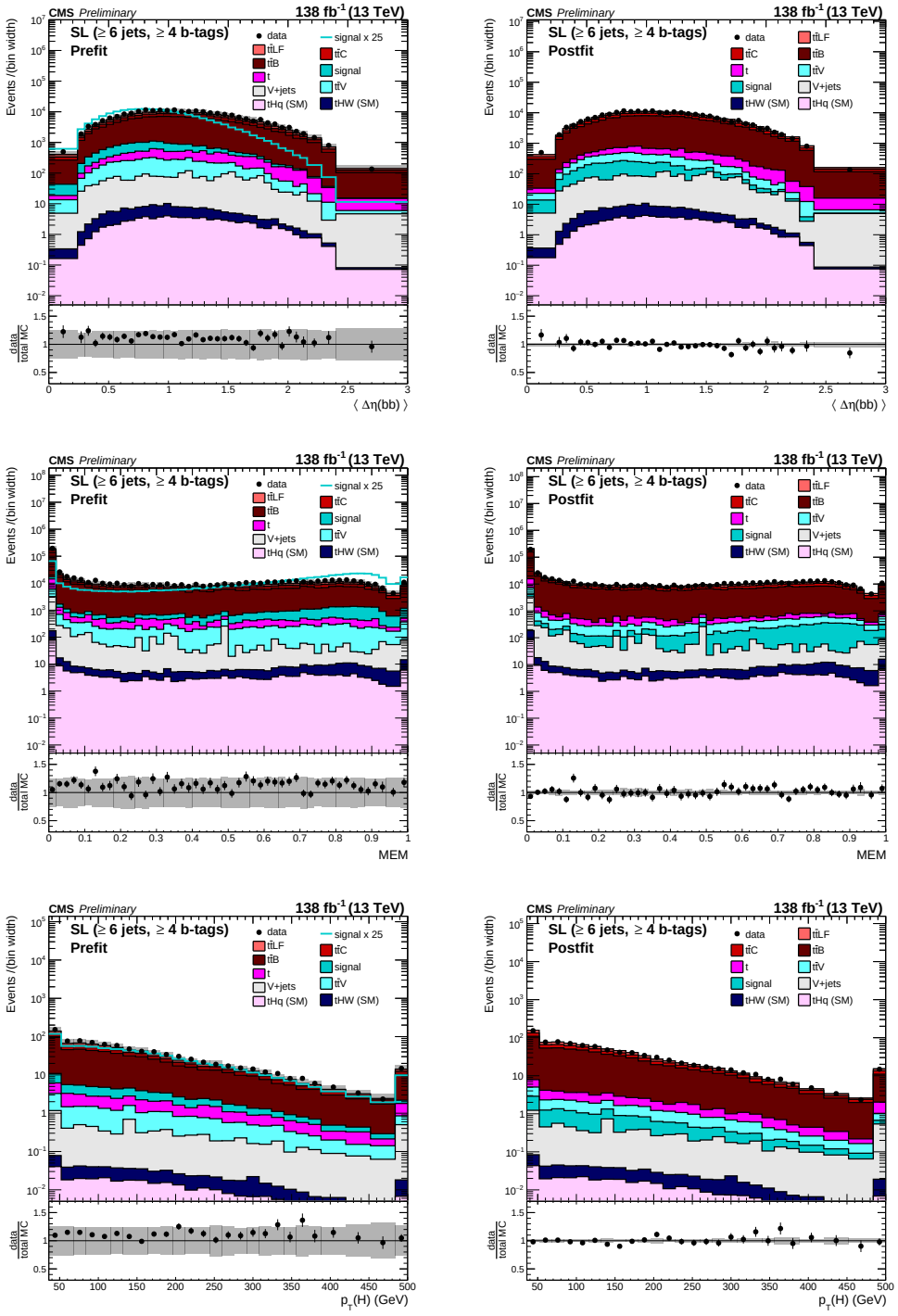


Figure 3: Average $\Delta\eta$ between any two b-tagged jets (upper), MEM discriminant output (middle), and p_{T} of the Higgs boson candidate identified with reconstruction BDT (lower) for events passing the baseline selection requirements and additionally $\geq$ 6 jets in the SL channel prefit (left) and with the postfit background model (right) obtained from the fit to data described in Section 10. In the prefit case, the $\mathrm{t\bar{t}H}$ signal contribution, scaled by a factor 25 for better visibility, is also overlayed (line). Where applicable, the last bins include overflow events.

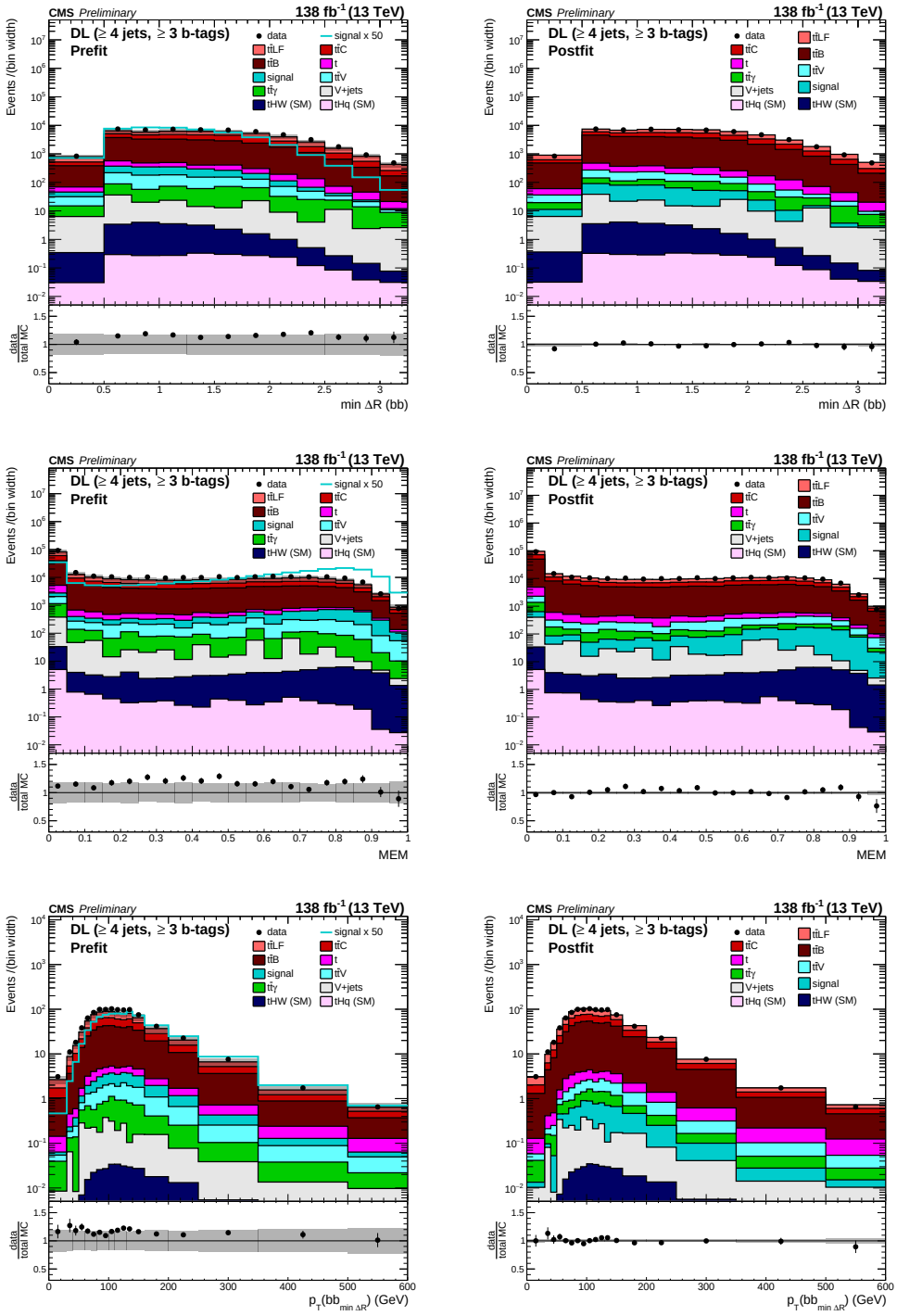


Figure 4: Minimum ΔR between any two b-tagged jets (upper), MEM discriminant output (middle), and p_{T} of the Higgs boson candidate identified as pair of b-tagged jets closest in ΔR (lower) for events passing the baseline selection requirements and additionally ≥ 4 jets in the DL channel prefit (left) and with the postfit background model (right) obtained from the fit to data described in Section 10. In the prefit case, the $t\bar{t}H$ signal contribution, scaled by a factor 50 for better visibility, is also overlayed (line). Where applicable, the last bins include overflow events.

8 Analysis strategy and classification

In each analysis channel, the selected events are divided into categories depending on the jet and b-tagged jet multiplicity. In each category, dedicated ANNs are trained and optimised to separate specific signal or background processes. The ANN output is used, in some cases, to categorise further the events and to construct the final discriminating observables. The events are categorised independently for the 2016, 2017, and 2018 data-taking periods, leading to an additional factor of three in the number of categories.

The signal is extracted in a simultaneous binned profile likelihood fit of the expected signal and background distributions of the discriminating observable or of the event yield, depending on the channel and category, to the data [74]. The fit is performed across all categories and channels. The systematic uncertainties discussed later in Section 9 are taken into account using nuisance parameters with appropriate constraints [74, 75], and are correlated among the processes, categories, channels, and data-taking periods where appropriate. The different background composition in the categories helps to constrain the uncertainties and thus to increase the overall sensitivity. The signal distributions include contributions from all SM Higgs boson decays to take into account contamination by other decay channels than $\mathrm{H} \to \mathrm{b\bar{b}}$.

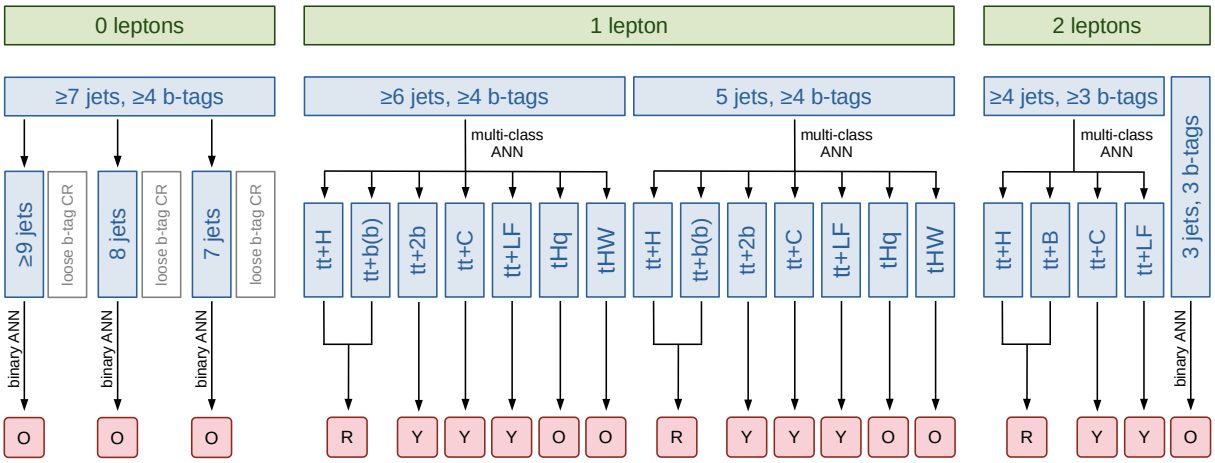
The categorisation of the signal events differs depending on the interpretation. As a first step, the signal categories are optimised for the inclusive $\mathrm{t\bar{t}H}$ and tH production rate measurements, as well as for the coupling and CP measurement. For the STXS measurement, the $\mathrm{t\bar{t}H}$ signal categories in the different channels are split further into five subcategories (STXS categories) that target different ranges of Higgs boson p_{T} (with lower boundaries at 0, 60, 120, 200, and 300 GeV). Independent signal templates are constructed per particle-level Higgs boson p_{T} bin and fitted simultaneously. In each subcategory, the contributions from all p_{T} bins are considered, thereby taking into account migration of events across bins due to reconstruction and resolution effects.

The analysis strategy is illustrated in Fig. 5. It has been optimised separately in each channel, as detailed below, based on the expected sensitivity to an SM signal evaluated with simulated events and, in the case of the FH channel, data for the QCD background determination.

8.1 Fully-hadronic channel

In the FH channel, events are categorised depending on the jet and b-tagged jet multiplicity, as well as m_{qq}. Three categories, referred to as "signal regions" (SR), with $(7\,\text{jets}, \geq 4\,\text{b tags})$, $(8\,\text{jets}, \geq 4\,\text{b tags})$, and $(\geq 9\,\text{jets}, \geq 4\,\text{b tags})$ are used for the signal extraction and enter the final fit. The events in the SR are in addition required to fulfil $60\,\text{GeV} < m_{\mathrm{qq}} < 100\,\text{GeV}$ for events with 7 or 8 jets, and $70\,\text{GeV} < m_{\mathrm{qq}} < 92\,\text{GeV}$ for events with ≥ 9 jets. Dedicated binary-classification ANNs are trained in each SR as detailed below to separate the $\mathrm{t\bar{t}H}$ signal from the background, and the ANN classifier output distribution is used as the final discriminating observable.

For each SR, two mutually exclusive control regions with looser b tagging requirements, referred to as "evaluation regions" (ER) and "training regions" (TR), are defined, which are used to estimate the QCD background and to train the ANNs, respectively. Events in the ER (TR) are required to contain at least four jets fulfilling the loose b tag requirement, out of which exactly three (two) are also b-tagged. Otherwise, events in the ER and TR have to pass the same selection criteria as events in the SR. For each SR, ER, and TR, further mutually exclusive regions, referred to as "validation regions" (VR-SR, VR-ER, VR-TR), are defined by inverting the criterion on m_{qq}. The validation regions are signal depleted, with an expected $\mathrm{t\bar{t}H}$ contri-



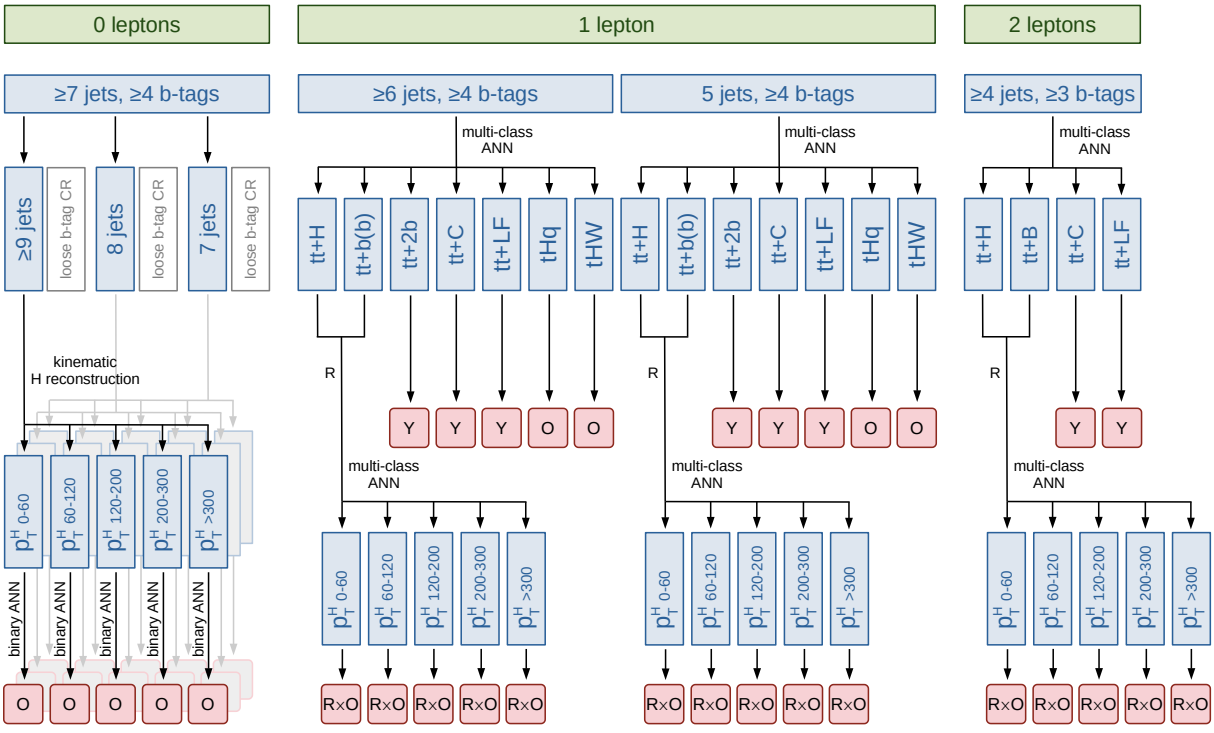


Figure 5: Illustration of the analysis strategy for the inclusive $t\bar{t}H$ and tH signal-strength modifier and coupling and CP measurements (upper) and for the $t\bar{t}H$ STXS measurement (lower). The procedure is applied separately for the three years of data taking.

bution below 0.6%, and used to test the background estimation and ANN training procedures described in the following. The categorisation scheme is summarised in Table 4.

Table 4: Categorisation scheme in the FH channel, applied independently in each jet-multiplicity category. The m_{qq} selection criteria refer to events with 7 or 8 (≥ 9) jets.

	$60(72)\,\text{GeV} < m_{qq} < 100(90)\,\text{GeV}$	$30\,\text{GeV} < m_{qq} < 60(72)\,\text{GeV}$ or $100(90)\,\text{GeV} < m_{qq} < 250\,\text{GeV}$
2 b tags	*training region (TR)*	*validation region (VR-TR)*
≥ 4 loose b tags	QCD events for ANN training	input variable validation
3 b tags	*evaluation region (ER)*	*validation region (VR-ER)*
≥ 4 loose b tags	discriminant shape for QCD	discriminant shape for QCD
≥ 4 b tags	*signal region (SR)*	*validation region (VR-SR)*
	analysis region	comparison of QCD shape with data

The dominant background contribution (approx. 72–80% of the events) in the SR originates from QCD. It is determined from the data to avoid effects from inaccurate modelling and insufficient number of events in the relevant phase space regions of the QCD simulation. The procedure is as follows: the ANNs are trained using the events from the TR, which are estimated to originate from QCD more than 80% of the time followed by $t\bar{t}$LF events, as background together with simulated signal events. The shape of the discriminant distribution is then determined from the events in the ER by evaluating the ANN output from these events, and then subtracting the remaining contributions from other backgrounds (mainly $t\bar{t}$ + jets) using simulation. The yield is left as free parameter in the final fit, independently in each SR. As initial value, the yield observed in data, after subtracting the other background contributions as expected from the simulation, is used.

Since the kinematic properties of b-tagged and untagged jets differ in the TR, ER, and SR due to the different heavy-flavour composition, event weights are applied to the events in the TR and ER such that the kinematic properties of the loose b-tagged jets match those of b-tagged jets. In the ER, a correction factor (TF_{loose}) is computed for the leading loose b-tagged jet that is not also b-tagged, as a function of the jet p_T, η, and the minimum distance between the jet and the first two b-tagged jets, and this factor is used as event weight. In the TR, the TF_{loose} correction factors are computed for the two leading loose b-tagged jets that are not also b tagged, and their product is used as event weight. The TF_{loose} correction is derived from the jets in events passing the baseline selection, excluding the first two jets, ordered according to their b tagging discriminant values. After the TF_{loose} correction, the shape of the discriminant distribution obtained in the ER is expected to represent well the shape of the QCD background in the SR. This is verified in data by applying the same procedure to the events in the VR-ER and then comparing the predicted shape to the one observed in the VR-SR. Additionally, the method is validated using the events of the ER itself by further separating this region into two depending on the b tagging discriminant value, and applying the procedure analogously.

In each SR, one single ANN, which is valid for all three data-taking periods, is constructed as binary classifier to separate $t\bar{t}$H events from QCD background events. Taking into account explicit separation against $t\bar{t}$ + jets events was found to not improve the results further. The ANN training is performed with simulated $t\bar{t}$H events, where only Higgs boson decays to $b\bar{b}$ are considered, and with the data events from the TR.

The ANNs utilise input variables related to the kinematic properties of individual objects and

event shape information as well as the MEM discriminant output, listed in Table 5. The input variable distributions are verified to agree between the SR, ER, and TR after application of the TF_{loose} correction using data in the VR and simulated $t\bar{t}H$ events. For example, variables that are directly correlated with the b tagging discriminant value are discarded. The remaining input variables have been further verified to model the data well based on a χ^2 criterion.

For the STXS measurement, a Higgs boson candidate is reconstructed from that pair of jets, out of the four jets with the highest b tagging discriminant value, whose invariant dijet mass is most compatible with that of a 125 GeV Higgs boson, based on a χ^2 criterion taking into account the experimental resolution of the jet p_T. Depending on the p_T of the reconstructed Higgs boson candidate, the SR events are categorised further and assigned to one of the five Higgs boson p_T ranges, resulting in 40–57% correct assignments for $t\bar{t}H$ events, as presented in Fig. 6. Thus, there are in total 15 STXS categories in the FH channel for each of the three data-taking periods. In each category, the ANN output distribution is used as final discriminant in the fit to extract the STXS.

8.2 Single-lepton and dilepton channels

In the channels with leptons, events are separated based on the jet and b-tagged jet multiplicity. In the SL channel, two categories with $(\geq 6\,\text{jets}, \geq 4\,\text{b tags})$ and $(5\,\text{jets}, \geq 4\,\text{b tags})$ are considered, and in the DL channel two categories with $(\geq 4\,\text{jets}, \geq 3\,\text{b tags})$ and $(3\,\text{jets}, 3\,\text{b tags})$.

In each of the two SL categories and in the $(\geq 4\,\text{jets}, \geq 3\,\text{b tags})$ DL category, events are further categorised based on the output of multi-classification ANNs, which are designed to separate between different signal and background processes. The values obtained in the output nodes of the ANNs are normalised to unity using a "soft-max" function [78], and, as a result, the output values $O(i)$ can be interpreted as probabilities describing the likelihood of the event being of a certain process i. Events are assigned to a category corresponding to the most probable process according to this ANN multi-classification. In the SL channel, the three signal processes $t\bar{t}H$, tHq, and tHW, and the four $t\bar{t}$ + jets background processes $t\bar{t}LF$, $t\bar{t}C$, $t\bar{t}+2b$, as well as $t\bar{t}B$ without $t\bar{t}+2b$ are considered in the ANN multi-classification. The dedicated class for $t\bar{t}+2b$ events out of the $t\bar{t}B$ background is designed in order to constrain the uncertainty related to collinear gluon splitting. In the DL channel, the $t\bar{t}H$ signal process and the three $t\bar{t}$ + jets background processes $t\bar{t}LF$, $t\bar{t}C$, and $t\bar{t}B$ are considered in the ANN multi-classification. No dedicated class for $t\bar{t}+2b$ events is included in the DL channel, motivated by the reduced statistical precision due to the lower $t\bar{t}$ + jets event rate in this channel.

The output values of the ANNs are subsequently used to compute the final discriminant observables. In the SL channel, for events in the $t\bar{t}H$ and $t\bar{t}+b(\bar{b})$ process categories, a likelihood ratio discriminant is computed from the ANN output values O as

$$R_{SL} = \frac{O(t\bar{t}H)}{O(t\bar{t}H) + O(t\bar{t}+b(\bar{b})) + O(t\bar{t}+2b)},$$

and used as the discriminating observable. This allows exploiting more information from the ANN multi-classification: for example, for an event that is categorised as $t\bar{t}H$, not only the output value of the $t\bar{t}H$ node is used but also the output values of the $t\bar{t}+b(\bar{b})$ and $t\bar{t}+2b$ nodes, which provides further information about the likelihood of the categorisation. In the tHq and tHW process categories, the $O(tHq)$ and $O(tHW)$ distributions are used as the final discriminating observables, respectively. The main purpose of the remaining process categories $t\bar{t}LF$, $t\bar{t}C$, and $t\bar{t}+2b$ is to constrain the normalisation of the corresponding background processes; therefore, in these categories only the event yield is considered instead of the full distribution.

Table 5: Observables used as input variables to the ANN per channel. Observables marked with a † are constructed using information from the BDT-based event reconstruction.

		FH			SL		DL	
Observable		(≥ 9 jets, ≥ 4 b tags)	(8 jets, ≥ 4 b tags)	(7 jets, ≥ 4 b tags)	(≥ 6 jets, ≥ 4 b tags)	(5 jets, ≥ 4 b tags)	(≥ 4 jets, ≥ 3 b tags)	(3 jets, 3 b tags)
MEM	matrix element method discriminant	✓	✓	✓	✓	✓	✓	
BLR	b tagging likelihood ratio discriminant						✓	
$\ln\left(\frac{\mathrm{BLR}}{1-\mathrm{BLR}}\right)$	transformed b tagging likelihood ratio discriminant				✓	✓		
$p_\mathrm{T}(\mathrm{j}^2)$	p_T of second leading jet, ranked in p_T						✓	
$p_\mathrm{T}(\mathrm{j}^3)$	p_T of third leading jet, ranked in p_T							✓
$p_\mathrm{T}(\mathrm{j}^7)$	p_T of seventh leading jet, ranked in p_T	✓						
$p_\mathrm{T}(\mathrm{b}^i)$	p_T of i^th, i =1–4, leading b-tagged jet, ranked in p_T						✓	
$\eta(\mathrm{j}^i)$	η of i^th, i =1–2, leading jet, ranked in b tagging discriminant value	✓	✓	✓				
$\langle d_\mathrm{b}(\mathrm{j})\rangle$	average b tagging discriminant value of all jets				✓	✓		
$\langle d_\mathrm{b}(\mathrm{b})\rangle$	average b tagging discriminant value of all b-tagged jets				✓	✓		
$d_\mathrm{b}^3(\mathrm{j})$	third highest b tagging discriminant value of all jets				✓	✓		
$\mathrm{Var}(d_\mathrm{b}(\mathrm{j}))$	variance of b tagging discriminant values of all jets				✓	✓		
$\langle \Delta R(\mathrm{bb})\rangle$	average of ΔR between two b-tagged jets						✓	
$\langle \Delta R(\mathrm{jj})\rangle$	average of ΔR between two jets	✓	✓					
min $\Delta R(\mathrm{jj})$	minimum of ΔR between two jets		✓	✓				✓
max $\Delta R(\mathrm{jj})$	maximum of ΔR between two jets	✓	✓	✓				
$\langle \Delta\eta(\mathrm{bb})\rangle$	average of $\Delta\eta$ between two b-tagged jets				✓	✓		
$\langle \Delta\eta(\mathrm{jj})\rangle$	average of $\Delta\eta$ between two jets	✓	✓	✓	✓	✓		
$\langle m(\mathrm{b})\rangle$	average invariant mass of all b-tagged jets				✓	✓		
$\langle m(\mathrm{j})\rangle$	average invariant mass of all jets				✓	✓		
$m(\mathrm{bb}_{\mathrm{min}\,\Delta R})$	invariant mass of pair of b-tagged jets closest in ΔR				✓	✓	✓	
$m(\mathrm{jb}_{\mathrm{min}\,\Delta R})$	invariant mass of pair of jet and b-tagged jet closest in ΔR						✓	
$m(\mathrm{jj}_{125\,\mathrm{GeV}})$	invariant mass of pair of jets with mass closest to 125 GeV	✓						
$m(\mathrm{bb}_{\mathrm{max}\,m})$	maximum invariant mass of pairs of b-tagged jets	✓	✓				✓	✓
$m(\mathrm{jbb}_{\mathrm{max}\,p_\mathrm{T}})$	invariant mass of jet and pair of b-tagged jets with highest p_T						✓	
$\langle p_\mathrm{T}(\mathrm{j})\rangle$	average p_T of all jets				✓	✓		
$\langle p_\mathrm{T}(\mathrm{b})\rangle$	average p_T of all b-tagged jets				✓	✓		
$p_\mathrm{T}(\mathrm{bb}_{\mathrm{min}\,\Delta R})$	p_T of pair of b-tagged jets closest in ΔR				✓	✓	✓	✓
$p_\mathrm{T}(\mathrm{jj}_{\mathrm{min}\,\Delta R})$	p_T of pair of jets closest in ΔR							✓
$p_\mathrm{T}(\mathrm{jb}_{\mathrm{min}\,\Delta R})$	p_T of pair of jet and b-tagged jet closest in ΔR							✓
$H_\mathrm{T}(\mathrm{j})$	scalar sum of p_T of all jets				✓	✓	✓	
$H_\mathrm{T}(\mathrm{b})$	scalar sum of p_T of all b-tagged jets				✓	✓	✓	
$N(\mathrm{j})$	number of jets				✓			
$N(\mathrm{b}^{\mathrm{loose}})$	number of jets with loose b tag						✓	
$d_\mathrm{b}(\mathrm{b}_{\mathrm{top}}^{\mathrm{tHW}})^\dagger$	b tagging discriminant value of b jet from t quark decay from tHW reconstruction				✓	✓		
$\lvert\eta(\mathrm{q}^{\mathrm{tHq}})\rvert^\dagger$	$\lvert\eta\rvert$ of light-quark jet from tHq reconstruction				✓	✓		
$m(\mathrm{t}_{\mathrm{lep}}^{\mathrm{t\bar{t}H}})^\dagger$	inv. mass of leptonically decaying t quark from $\mathrm{t\bar{t}H}$ reconstruction				✓	✓		
$\mathrm{BDT}^{i\,\dagger}$	reconstruction BDT output for tHq, $\mathrm{t\bar{t}H}$, $\mathrm{t\bar{t}}$ hypotheses				✓	✓		
A, S	event aplanarity and sphericity [76]	✓	✓	✓				
H_i^FW	i^th, i =0–5, Fox–Wolfram moment [77]	✓	✓	✓				
$H_i^\mathrm{FW}/H_0^\mathrm{FW}$	ratio of Fox–Wolfram moments, i =1–4	✓	✓	✓				

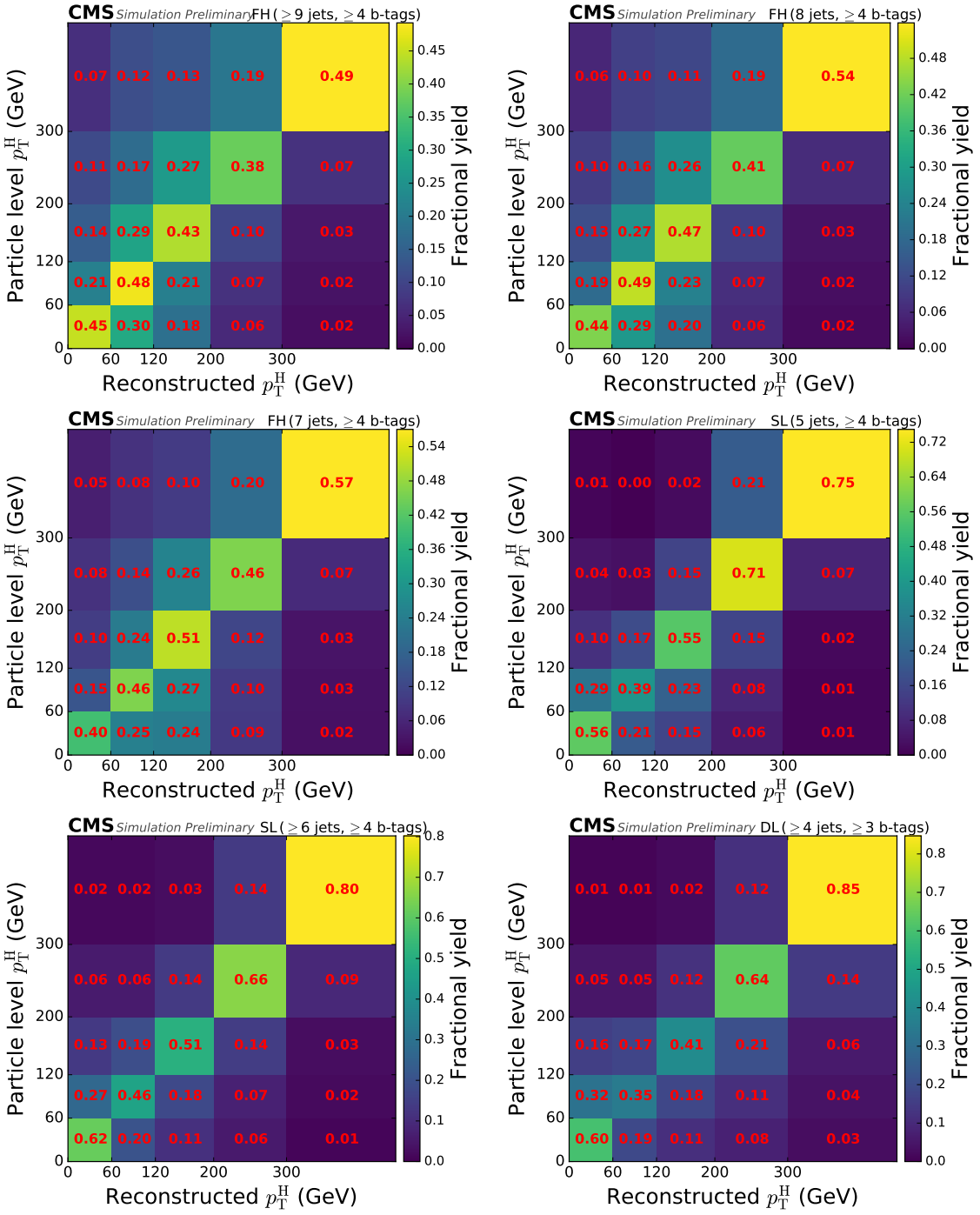


Figure 6: Categorisation efficiency of the $t\bar{t}H$ signal events in the STXS analysis in the different categories of the FH channel (upper row, middle row left), the SL channel (middle row right, lower row left), and the DL channel (lower row right).

Similarly, in the DL channel, a likelihood ratio discriminant is computed from the events in the $t\bar{t}H$ and $t\bar{t}B$ process categories as

$$R_{DL} = \frac{O(t\bar{t}H)}{O(t\bar{t}H) + O(t\bar{t}B)},$$

and used as the discriminating observable, while for the $t\bar{t}LF$ and $t\bar{t}C$ process categories, the event yield is used.

In the DL $(3\,\text{jets}, 3\,\text{b tags})$ category, a binary-classification ANN is designed to separate the $t\bar{t}H$ signal from the inclusive $t\bar{t}+\text{jets}$ background, and the ANN classifier output distribution is used as final discriminating observable.

In total, this leads to 12 categories in the SL channel and four categories in the DL channel per data-taking period, see Fig. 5.

As for the FH channel, the ANNs utilise input variables related to the kinematic properties of individual objects, the event shape, the MEM discriminant outputs, as well as the jet b tagging and the BLR discriminants, as listed in Table 5. In the SL channel, variables constructed using information from the BDT-based event reconstruction are also used (marked with a † in Table 5). The input variables and their correlations were verified to describe the data well, based on a goodness-of-fit test that takes into account the full uncertainty model, i.e. including statistical and systematic uncertainties, using the "saturated model" method [79]. For each variable and each pair of variables under scrutiny, the one and two dimensional distributions, respectively, were fitted to the data in the analysis categories, confirming they are well modelled.

The ANN are trained using simulated $t\bar{t}H$, tH, and $t\bar{t}+\text{jets}$ events. Only events with Higgs boson decays to $b\bar{b}$ are considered, and the $t\bar{t}$ system is required to have one or two leptons in the final state for the SL and the DL channels, respectively. For the tH events, only those with leptonic decays of the top quark are considered. The achieved classification accuracy ranges from 35–65% for $t\bar{t}H$ events to 40–70% for tH events, as well as 30–40% and 60–70% for $t\bar{t}B$ and $t\bar{t}LF$ events, respectively, depending on the channel.

For the STXS measurement, the most signal-like events are categorised further, targeting the five regions in Higgs boson p_T. Specifically, these are the events in the $t\bar{t}H$ and $t\bar{t}+b(\bar{b})$ process categories of the SL $(\geq 6\,\text{jets}, \geq 4\,\text{b tags})$ and $(5\,\text{jets}, \geq 4\,\text{b tags})$ categories and the $t\bar{t}H$ and $t\bar{t}B$ process categories of the DL $(\geq 4\,\text{jets}, \geq 3\,\text{b tags})$ category. The STXS categorisation is performed by additional multi-classification ANNs designed to classify $t\bar{t}H$ events by the Higgs boson p_T. Each event is assigned to the Higgs boson p_T bin with the highest probability according to the ANN multi-classification; the ANN output value in that node, multiplied by the observable value of the inclusive $t\bar{t}H$ measurement, i.e. the likelihood ratio value, is used as final discriminating observable. This approach was found to achieve superior sensitivity compared to other classification approaches studied. Thus, in total, there are ten STXS categories in the SL channel, corresponding to the five Higgs boson p_T ranges in each of the jet multiplicity categories, and five STXS categories in the DL channel. The categorisation of all other events remains the same as for the inclusive $t\bar{t}H$ measurement, with the exception of the events in the DL $(3\,\text{jets}, 3\,\text{b tags})$ category, which are not used since they do not contribute to improve the sensitivity in the STXS analysis.

A dedicated ANN is trained in each of the $(\geq 6\,\text{jets}, \geq 4\,\text{b tags})$ and $(5\,\text{jets}, \geq 4\,\text{b tags})$ categories of the SL channel and in the $(\geq 4\,\text{jets}, \geq 3\,\text{b tags})$ category of the DL channel, using the input variables listed in Table 5. For the training, only simulated $t\bar{t}H$ events are used, and from these only those where both b quarks are within the experimental acceptance defined by the

event selection described in Section 6. The achieved categorisation efficiency is presented in Fig. 6 and was found to be superior compared to a categorisation based on a kinematic reconstruction of the Higgs boson p_{T}.

An alternative approach that targets the STXS bins in which the Higgs boson is produced at $p_{\mathrm{T}} > 300\,\mathrm{GeV}$ is also studied using techniques developed to address the boosted topology. In addition to the minimal selection requirements for jets, leptons, and $p_{\mathrm{T}}^{\mathrm{miss}}$ reported in Table 3, events must contain a high p_{T} large-cone-size jet with a radius parameter of 0.8, identified as a boosted $\mathrm{H} \to \mathrm{b\bar{b}}$ candidate through the DeepDoubleB algorithm [66, 80]. In this boosted category, the sum of all the SM background components is estimated directly from data by fitting a Bernstein polynomial to the softdrop mass distribution of the Higgs boson candidate jet. This method has not been included in the reported analysis due to the very small statistical sensitivity given by the limited number of events expected with the Run 2 luminosity, however its data-driven nature has potential to make it competitive on a longer timescale.

8.3 Neural network architecture and training

All ANNs are implemented in Keras [81] as feedforward neural networks. The architecture consists of three or four hidden layers with between 100 to 2048 nodes, depending on the channel and category. The hyperparameters have been optimised using a "Tree Parzen Estimator" [82, 83], which uses information from past trials when testing the next set of hyperparameters. The cost function that is minimised during the training is the "categorical cross entropy" in case of multi-classification ANNs, and the "squared hinge" function or the "binary cross entropy" in case of binary-classification ANNs. Potential overtraining is minimised using dropout and L2 regularisation or dropout and simultaneously L1 and L2 regularisation, depending on the channel. The training is terminated once the performance does not improve significantly after a full pass over the training data ("epoch").

In each category, one single ANN is trained that is valid for all three data-taking periods. Simulated data and, in case of the FH channel, data of all three periods, weighted to reflect the different integrated luminosities, are used in the training, thereby reducing statistical fluctuations. When evaluating the ANNs to obtain the final discriminant distributions in simulation and data, however, separate distributions are constructed per data-taking period in order to allow constraining uncertainties that are uncorrelated between the periods. It has been validated that the sensitivity of the analysis does not degrade compared to the case of training ANNs separately for each data-taking period or when using information about the year of data taking as an input feature to the ANN.

Depending on the category, between 0.5 and 1.5 million events are used for the training, with at least 16 000 events for the $\mathrm{t\bar{t}B}$ class in all cases. The events are further split into three independent subsamples used for the actual training (60%), for the optimisation of the hyperparameters (20%), and for validating the performance of the ANNs (20%). Additional weights are applied such that the effective number of events per class is the same in order to avoid that the ANN classification decision is biased by the relative frequency of the different processes or, in case of the STXS classification, the Higgs boson p_{T} bins. These events are used exclusively for training, optimisation, and validation, and another statistically independent set of events is used for the final analysis of the data to avoid biases due to potential overtraining.

9 Systematic uncertainties

Several sources of systematic uncertainties are considered in the analysis. The uncertainties are taken into account via nuisance parameters in the final profile-likelihood fit described in Section 10 and alter either the rate or both the rate and the discriminant shape of the signal or background processes. The effects from the same source are treated as fully correlated among the different categories. In general, and unless stated otherwise hereafter, theoretical uncertainties are treated as fully correlated among the different data-taking periods, while experimental uncertainties are treated as uncorrelated for the data recorded in 2016, 2017, and 2018. The latter is justified since the experimental uncertainties are mainly of statistical origin related to the limited size of the data and simulation samples used in auxiliary measurements, which are independent between the data-taking periods.

The theoretical uncertainties of the cross sections used to predict the rates of various processes, which arise primarily from the factorisation and renormalisation scale choices and the PDFs, are propagated to the yield estimates. The cross section uncertainties are each separated into their scale (renorm./fact. scales) and PDF components, and are correlated where appropriate among processes. In addition, the normalisation of the $t\bar{t}B$ and $t\bar{t}C$ background processes is left unconstrained in the final fit. To take into account additional uncertainties in the modelling of collinear gluon splitting, an additional 100% log-normal constrained rate uncertainty is assigned to the $t\bar{t} + 2b$ component relative to the normalisation of the overall $t\bar{t}B$ process. Since the shapes of the relevant distributions are different for the $t\bar{t} + 2b$ component and the overall $t\bar{t}B$ background, this leads effectively to a rate and shape variation of the $t\bar{t}B$ distributions.

Uncertainties arising from missing higher-order terms in the POWHEG $t\bar{t}b\bar{b}$ and $t\bar{t}$ simulations at ME level are evaluated by independent variations of the renormalisation and factorisation scales by factors of two up and down with respect to the nominal values, and two independent nuisance parameters are assigned in the fit (μ_R/μ_F scale). In addition, the uncertainties are treated as independent among the signal, the $t\bar{t}B$, and the other $t\bar{t}$ processes. The uncertainty arising from the PDF set is determined from the PDF variations provided with the NNPDF set [84], correlating processes for which the same FS and order in α_S are used in the PDF set. The corresponding uncertainty in the PYTHIA PS is determined by varying the parameters controlling the amount of initial- and final-state radiation independently by factors of two up and down [85] (PS scale ISR/FSR), separately for signal, $t\bar{t}B$, and other $t\bar{t}$. These variations are applied using event weights obtained directly from the generators. Uncertainties related to the ME-PS matching scheme and the underlying event tune are evaluated by comparing the reference $t\bar{t}b\bar{b}$ and $t\bar{t}$ simulation with samples with varied parton-shower matching scale (`hdamp` parameter) and varied tune parameters, respectively. The event count in these additional samples was small and induced changes to the discriminant distributions comparable in size to the statistical fluctuations of the additional samples and compatible with a pure rate variation. For this reason, the uncertainties were estimated as the changes in the rates of the different $t\bar{t} +$ jets subprocesses independently for each category. The derived rate variations amount typically to 5–10% and the uncertainties are treated as uncorrelated among the $t\bar{t} +$ jets subprocesses. This approach was verified with an independent estimation of the variations obtained from the same additional samples and applying the same analysis selection, except for using events with fewer b-tagged jets, thus reducing the statistical fluctuations. The effect of the selection with fewer b-tagged jets is accounted for via event weights that act as transfer factors, that depend on the p_T, η, and flavour of the jets in the event and encode the probability to observe jets originating from b quarks among the non-b-tagged jets. The impact of the mismodelling of the top quark p_T spectrum in the $t\bar{t}$ simulation [86] was found to be negligible.

The robustness of the background model against potential mismodelling of the $t\bar{t}B$ component has been confirmed using pseudo experiments. A large number of pseudo experiments were sampled from alternative $t\bar{t}B$ predictions, including the POWHEG $t\bar{t}$ sample and a $t\bar{t}B$ component scaled by a factor of 1.2, and fitted using the nominal background model, taking into account the systematic uncertainties. The fits were also performed including only the theoretical $t\bar{t}$ + jets modelling uncertainties. In all cases, the average fitted signal strength deviated from the injected value by at most 6%, which is well within the systematic uncertainty of the measurement, and the average fitted $t\bar{t}B$ normalisation corresponded to the applied scale of the $t\bar{t}B$ background.

For the STXS measurement, the $t\bar{t}$ + jets background uncertainty model is extended to provide further flexibility towards potential effects that depend on the reconstructed Higgs boson p_T. The $t\bar{t}B$ normalisation parameter, as well as the nuisance parameters associated with the uncertainties due to the modelling of collinear gluon splitting, the ISR and FSR PS scale, and the ME-PS matching scheme are partially decorrelated between each of the five STXS categories and the other categories.

Furthermore, the inclusive $t\bar{t}H$ signal cross section uncertainties (renorm./fact. scales and PDF) are omitted for the STXS measurement. Instead, two additional sources are considered to estimate the uncertainty in the description of migration and acceptance effects of signal events between the STXS bins, derived from factorisation and normalisation scale variations. The migration uncertainties are evaluated following an adaption of the method proposed in chapter IV.6 of Ref. [87] by studying the effect of all combinations of factorisation and normalisation scale variations across each $p_T(H)$ boundary on the total cross section above this boundary. The largest effect is taken as the absolute uncertainty, which is propagated by increasing the process normalisation in the STXS bins above the $p_T(H)$ boundary, and decreasing the process normalisation in the bin directly below the boundary. In addition, the uncertainties in the bins above $p_T(H) = 120\,\text{GeV}$ were multiplied by a tuning factor of 0.7 to ensure that they do not differ too much from the overall cross section uncertainty. The migration uncertainties are implemented as rate uncertainties per STXS bin and amount to 5–10%. Acceptance effects within each STXS bin are taken into account by varying the renormalisation and factorisation scales, taking into account a normalisation factor to ensure that the variations do not change the overall $t\bar{t}H$ cross section. The resulting shape variation is typically 1–3% and as large as up to 7% depending on the classifier bin. In addition, for the simulated $t\bar{t}H$ signal events, the renormalisation and factorisation scale variations (μ_R/μ_F scale) are performed simultaneously, and, as well as the parton-shower uncertainties (PS scale ISR/FSR), are treated as decorrelated for events with Higgs boson p_T below and above 300 GeV.

The integrated luminosities for the 2016, 2017, and 2018 data-taking years have 1.2–2.5% individual uncertainties [18–20], which are partially correlated to account for common sources of uncertainty in the luminosity measurement. They amount to an overall uncertainty of 1.6% for the 2016–2018 period. The trigger efficiency uncertainty in the FH channel is determined from the bin-by-bin uncertainties in the ratio of efficiency in data relative to simulation, and are 1–2% on average, with some being as large as 9%. The efficiencies of the single-electron and dilepton triggers are measured in data using reference triggers based on single-muon and p_T^{miss} requirements, respectively, that are uncorrelated with those used in the analysis; the uncertainties range up to 8%, dominated by statistical fluctuations in the data samples used in the auxiliary measurement. The electron and muon identification efficiency uncertainties are estimated by comparing variations in measured efficiency between data and simulation using a high-purity sample of Z boson decays. These uncertainties are found to be small, typically at the 1% level. The uncertainty of the L1 trigger prefiring correction is determined from the

uncertainty in the prefiring probability estimate and amounts to approximately 0.5%.

Effects of the uncertainty in the distribution of the number of pileup interactions are evaluated by varying the total inelastic cross section used to predict the number of pileup interactions in the simulated events by $\pm 4.6\%$ from its nominal value.

The uncertainty related to the jet energy scale (resolution) is determined by varying the energy scale (resolution) correction of all jets in the signal and background predictions by one standard deviation. The jet energy scale uncertainty is divided into 11 independent sources, which include uncertainties owing to the extrapolation between samples of different jet-flavour composition and the presence of pileup collisions in the derivation of the corrections, and which are treated as fully uncorrelated in the fit. While most of the sources are dominated by statistical fluctuations in auxiliary measurements and are treated as uncorrelated among the data-taking periods, some sources are related to theoretical predictions in the MC simulation used e.g. to extrapolate between samples of different jet-flavour composition, and are thus treated as correlated among the data-taking periods.

The b tagging scale factors receive uncertainties due to the contamination of background processes in the data samples used in the scale-factor measurements, 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 statistical uncertainty is parameterised as the sum of two contributions: one term with linear dependence on the b tagging discriminant value, allowing an overall tilt of the discriminant distribution, and another term with quadratic dependence, allowing an overall shift of the discriminant distribution. Each source of b tagging uncertainty is considered separately per jet flavour. The uncertainty related to the background contamination is treated as correlated among the data-taking periods of 2017 and 2018, and as uncorrelated with 2016 to allow for effects due to the upgraded pixel detector [88]. The statistical component is treated as uncorrelated among the data-taking periods.

Many uncertainties that are related to the MC simulation of the QCD background in the FH channel are avoided by estimating this contribution from data. Small uncertainties remain in the $\mathrm{TF}_{\mathrm{loose}}$ correction applied to the loose b-tagged jets, which is estimated by applying an additional η-dependent correction to $\mathrm{TF}_{\mathrm{loose}}$ to account for small effects of missing higher-order iterations in the correction procedure, and in the total normalisation in each category is left unconstrained in the final fit.

The impact of statistical fluctuations in the signal and background prediction due to the limited number of simulated events is accounted for using the Barlow–Beeston approach [89].

The described sources of uncertainty are summarised in Table 6 and their impact on the final result is discussed in Section 10.

10 Statistical analysis and results

The production rates of the $\mathrm{t\bar{t}H}$ and tH signal processes are determined in a simultaneous binned profile likelihood fit to the final discriminant distributions in all channels, categories, and data-taking periods. The rates of the $\mathrm{t\bar{t}B}$, $\mathrm{t\bar{t}C}$, as well as the QCD background, are separately left unconstrained in the fit. Several signal interpretations are performed and described below.

Table 6: Systematic uncertainties considered in the analysis. "Type" refers to rate (R) or rate and shape (S) altering uncertainties. "Correlation" indicates whether the uncertainty is treated as correlated, partially correlated (as detailed in the text), or uncorrelated across the years 2016–18. Uncertainties for $t\bar{t}$ + jets events marked with a † are treated as partially correlated between each of the STXS categories and the other categories in the STXS analysis.

Source	Type	Correlation	Remarks
Renorm./fact. scales	R	correlated	Scale uncertainty of (N)NLO prediction, independent for $t\bar{t}H$, tHq, tHW, $t\bar{t}$, t, V+jets, VV
PDF+α_S (gg)	R	correlated	PDF uncertainty for gg initiated processes, independent for $t\bar{t}H$, tHq, tHW, and others
PDF+α_S ($q\bar{q}$)	R	correlated	PDF uncertainty of $q\bar{q}$ initiated processes ($t\bar{t}W$,W,Z) except tHq
PDF+α_S (qg)	R	correlated	PDF uncertainty of qg initiated processes (single t) except tHW
Collinear gluon splitting†	S	correlated	Additional 100% rate uncertainty on $t\bar{t}$ + 2b component of $t\bar{t}B$ background
μ_R scale	S	correlated	Renormalisation scale uncertainty of the ME generator, independent for $t\bar{t}H$, tHq, tHW, $t\bar{t}B$ ($t\bar{t}b\bar{b}$ sample), other $t\bar{t}$ ($t\bar{t}$ sample)
μ_F scale	S	correlated	Factorisation scale uncertainty of the ME generator, independent for $t\bar{t}H$, tHq, tHW, $t\bar{t}B$ ($t\bar{t}b\bar{b}$ sample), other $t\bar{t}$ ($t\bar{t}$ sample)
PDF shape	S	correlated	From NNPDF variations, independent for tHq, tHW, $t\bar{t}B$ ($t\bar{t}b\bar{b}$ sample), other $t\bar{t}$ ($t\bar{t}$ sample) and $t\bar{t}H$
PS scale ISR†	S	correlated	Initial state radiation uncertainty of the PS (PYTHIA), independent for $t\bar{t}H$, $t\bar{t}B$ ($t\bar{t}b\bar{b}$ sample), other $t\bar{t}$ ($t\bar{t}$ sample)
PS scale FSR†	S	correlated	Final state radiation uncertainty of the PS (PYTHIA), independent for $t\bar{t}H$, $t\bar{t}B$ ($t\bar{t}b\bar{b}$ sample), other $t\bar{t}$ ($t\bar{t}$ sample)
ME-PS matching ($t\bar{t}$)†	R	correlated	NLO ME-PS matching (for $t\bar{t}$ + jets events), independent for $t\bar{t}B$, $t\bar{t}C$, $t\bar{t}LF$
Underlying event ($t\bar{t}$)	R	correlated	Underlying event (for all $t\bar{t}$ + jets events)
STXS migration	R	correlated	Signal, only in STXS measurement
STXS acceptance	S	correlated	Signal, only in STXS measurement
Integrated luminosity	R	partially	Signal and all backgrounds
Lepton ID/Iso (2 sources)	S	uncorrelated	Signal and all backgrounds
Trigger efficiency (4 sources)	S	uncorrelated	Signal and all backgrounds
L1 prefiring correction	S	uncorrelated	Signal and all backgrounds
Pileup	S	correlated	Signal and all backgrounds
Jet energy scale (11 sources)	S	partially	Signal, $t\bar{t}$ + jets and single t
Jet energy resolution	S	uncorrelated	Signal, $t\bar{t}$ + jets and single t
b tag bkg. contam. (2 sources)	S	partially	Signal and all backgrounds
b tag bkg. contam. stat. (4 sources)	S	uncorrelated	Signal and all backgrounds
b tag charm (2 sources)	S	partially	Signal and all backgrounds
TF_{loose} correction	S	uncorrelated	QCD background estimate
Size of the MC samples	S	uncorrelated	Statistical uncertainty of signal and background prediction due to limited sample size

10.1 $t\bar{t}H$ production rate

First, the $t\bar{t}H$ production rate is measured. For this interpretation, the tH contribution is assumed to conform to the SM expectation and is treated as background.

The observed yields in each bin of the final discriminant distributions in all channels and categories entering the fit are shown in Fig. 7, together with the fitted signal and background yields. The best-fit values of the inclusive $t\bar{t}H$ production rate relative to the SM expectation, denoted as the signal strength modifier $\mu_{t\bar{t}H}$, are listed in Table 7 and shown in Fig. 8. Results are shown for the combination of all channels and years, and for each channel and year alone when correlating the uncertainties. In the combination of all channels and years, a value of $\mu_{t\bar{t}H} = 0.33 \pm 0.26 = 0.33\,^{+0.17}_{-0.16}\,(\text{stat})\,^{+0.20}_{-0.21}\,(\text{syst})$ is obtained, with an expected uncertainty of $\pm 0.17\,(\text{stat})\,^{+0.23}_{-0.19}\,(\text{syst})$. The observed signal has a significance compared to the background-only hypothesis corresponding to 1.3 SD, with an expectation of 4.1 SD. The goodness-of-fit is quantified using a p value that takes into account the postfit uncertainty model and amounts to $p = 0.88$, indicating good description of the data by the fit model.

Table 7: Best-fit results of the $t\bar{t}H$ signal-strength modifier $\mu_{t\bar{t}H}$ in each channel and in their combination. Uncertainties are correlated between the channels.

	$\hat{\mu} \pm \text{tot}\,(\pm\text{stat} \pm \text{syst})$
FH	$+0.84^{+0.49}_{-0.46}$ $\begin{pmatrix}+0.24 & +0.42\\ -0.24 & -0.39\end{pmatrix}$
SL	$+0.46^{+0.33}_{-0.33}$ $\begin{pmatrix}+0.21 & +0.25\\ -0.21 & -0.26\end{pmatrix}$
DL	$-0.23^{+0.41}_{-0.42}$ $\begin{pmatrix}+0.31 & +0.26\\ -0.31 & -0.29\end{pmatrix}$
Combined	$+0.33^{+0.26}_{-0.26}$ $\begin{pmatrix}+0.17 & +0.20\\ -0.16 & -0.21\end{pmatrix}$

The postfit values and uncertainties of the $t\bar{t}B$ and $t\bar{t}C$ background normalisation parameters obtained in the combined fit of all channels are $1.19\,^{+0.13}_{-0.12}$ and $1.07\,^{+0.20}_{-0.19}$, respectively. These values are consistent with the results of dedicated inclusive $t\bar{t}b\bar{b}$ and $t\bar{t}c\bar{c}$ cross section measurements in similar phase space regions, which are at the level of 6–20% relative precision depending on the exact phase space and the size of the analysed dataset [71–73, 90], and reflect the known underprediction of the $t\bar{t}b\bar{b}$ cross section by the simulation. The anticorrelation between the $t\bar{t}H$ signal strength and the $t\bar{t}B$ background normalisation is visible in Fig. 9.

The best fit values and the impacts of the 20 nuisance parameters ranked highest in impact are presented in Fig. 10. The impact of each nuisance parameter is evaluated as the difference of the nominal best fit value of μ and the best fit value obtained when fixing the nuisance parameter under scrutiny to its best fit value plus/minus its postfit uncertainty. The nuisance parameters with the highest impact are related to the $t\bar{t}B$ background modelling, followed by the QCD background normalisation and, to a lesser extent, the jet energy scale, the uncertainty in the signal cross section ($\sigma_{t\bar{t}H}$), and the statistical uncertainty in the signal and background prediction due to the limited number of simulated events (MC stat.).

The best fit values of the nuisance parameters are within 1 SD of the prior uncertainty for more than 93% of the total number of nuisance parameters. As expected, significant shifts from the nominal value are observed for nuisance parameters related to the $t\bar{t}B$ background modelling, such as the $t\bar{t}B$ normalisation, the gluon-splitting uncertainty, and the PS modelling, since the a-priori knowledge does not reflect the data distributions. Consequently, the fit constrains

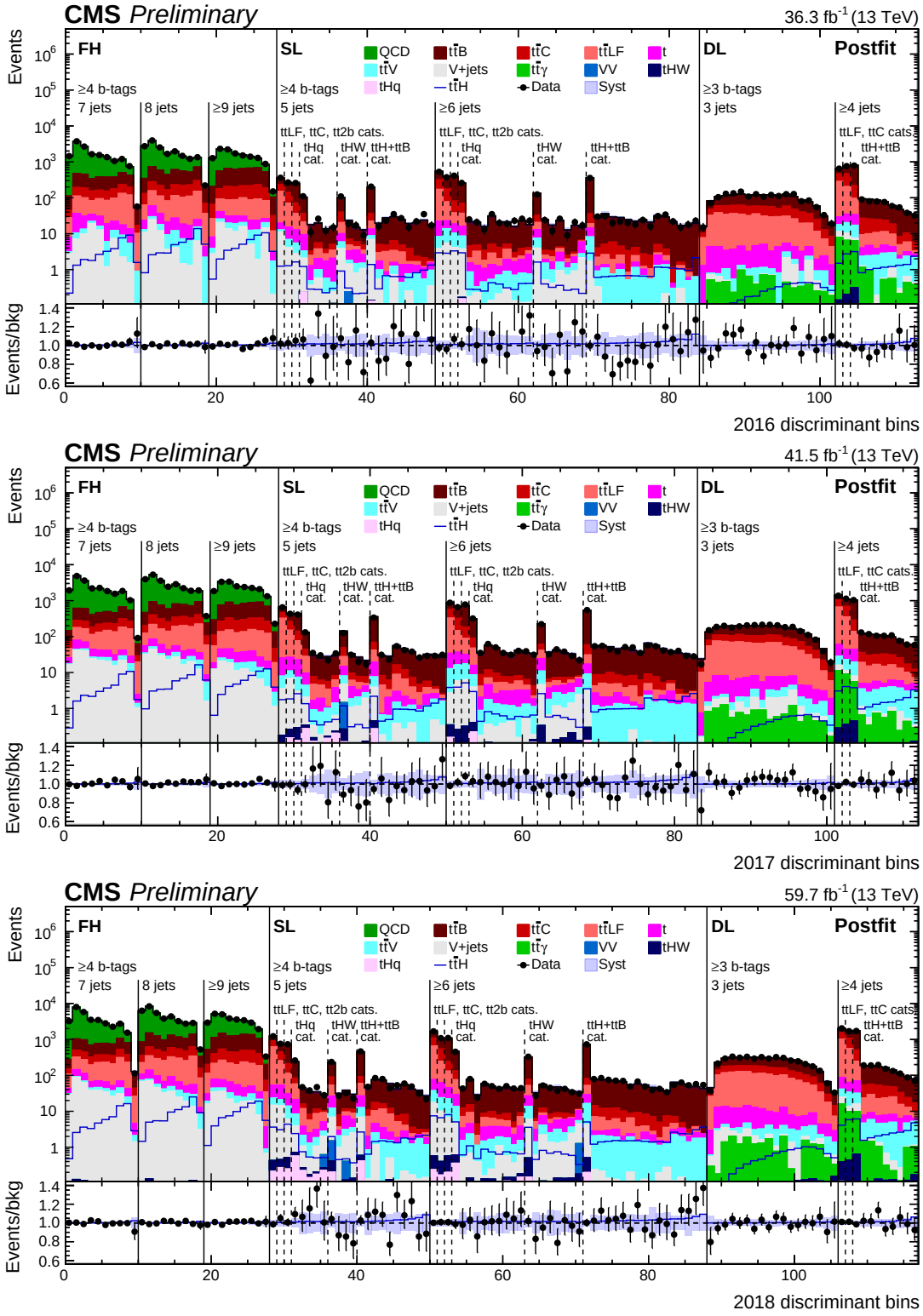


Figure 7: Observed (points) and postfit expected (filled histograms) yields in each discriminant (category yield, ANN score, or ratio of ANN scores) bin for the 2016 (upper), 2017 (middle), and 2018 (lower) data-taking periods. The uncertainty bands include the total uncertainty of the fit model. The lower pads show the ratio of the data to the background (points) and of the postfit expected signal+background to the background-only contribution (line).

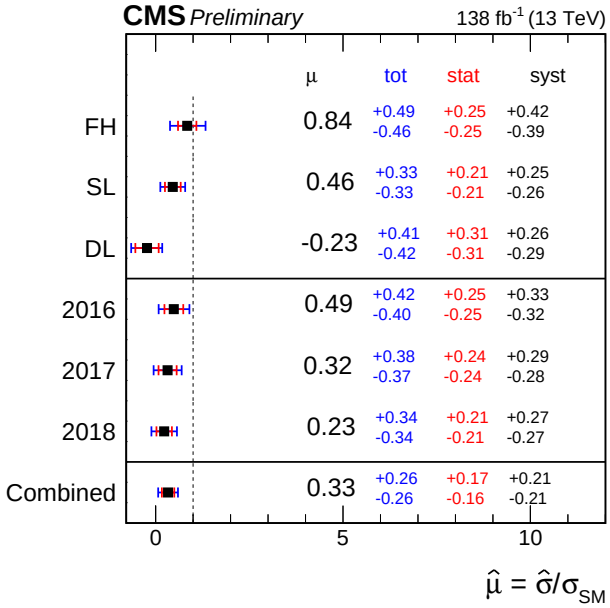


Figure 8: Best-fit results of the $t\bar{t}H$ signal-strength modifier $\mu_{t\bar{t}H}$ in each channel (upper three rows), in each year (middle three rows), and in the combination of all channels and years (lower row). Uncertainties are correlated between the channels and years.

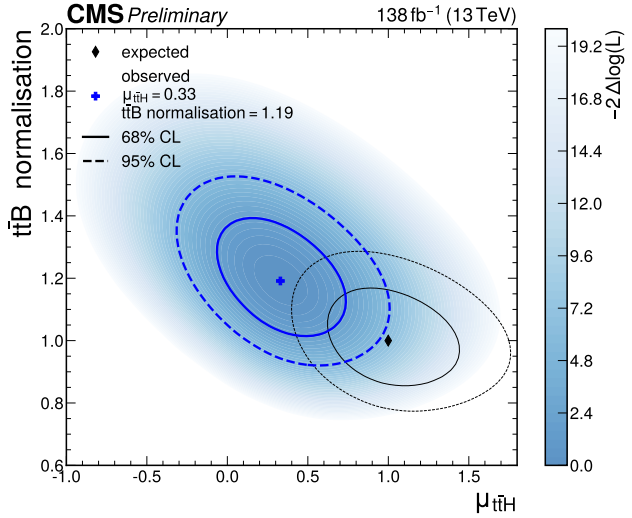


Figure 9: Observed likelihood-ratio test statistic (blue shading) as a function of the $t\bar{t}H$ signal-strength modifier $\mu_{t\bar{t}H}$ and the $t\bar{t}B$ background normalisation, together with the observed (blue) and SM expected (black) best-fit points (cross and diamond markers) as well as the 68% (solid lines) and 95% (dashed lines) CL regions. The $t\bar{t}C$ background normalisation and all other nuisance parameters are profiled such that the likelihood attains its minimum at each point in the plane.

these nuisance parameters relative to their prior values. Several other nuisance parameters, in particular those related to jet energy scale and b tagging uncertainties, are constrained. This is attributed to the fact that events are selected according to different, large multiplicities of jets and b-tagged jets, thus increasing the sensitivity of the analysis to changes of the jet energy scale and b tagging efficiency, for example by their effect on the event yield per analysis category, and that in several cases conservatively large prior uncertainties have been deliberately chosen.

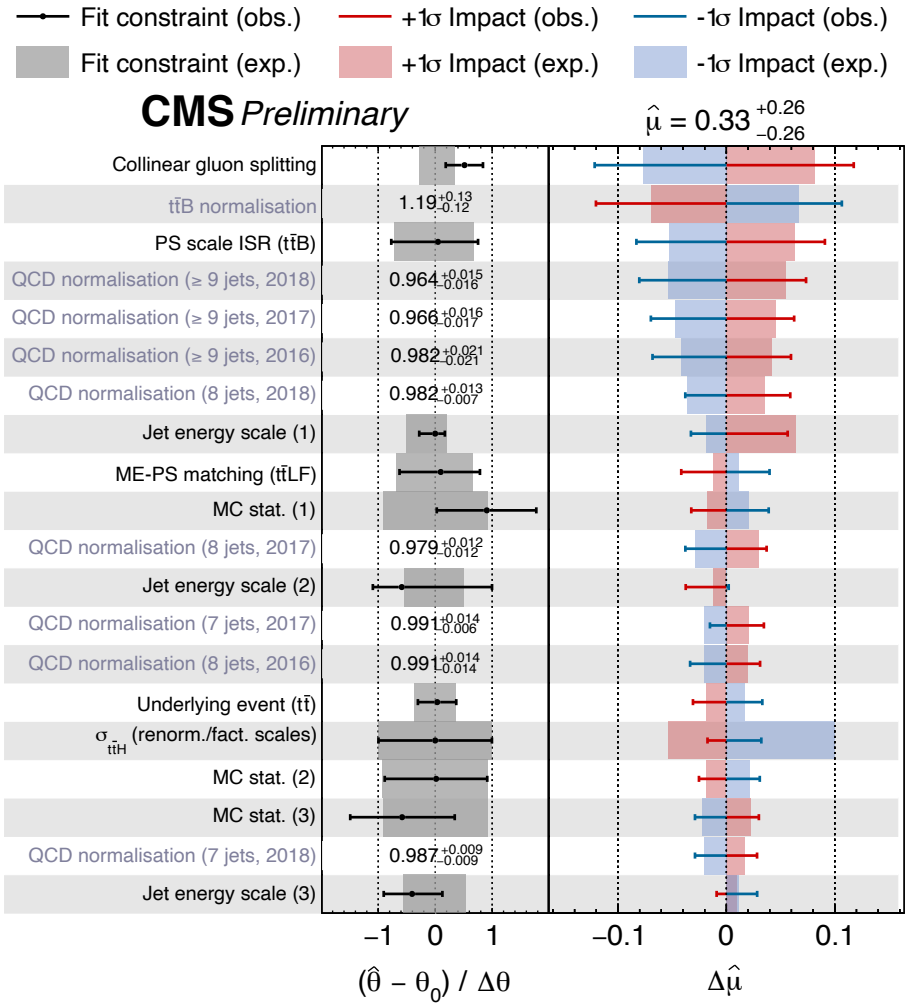


Figure 10: Postfit values of the nuisance parameters (black markers), shown as the difference of their best-fit values, $\hat{\theta}$, and prefit values, θ_0, relative to the prefit uncertainties $\Delta\theta$. The impact $\Delta\hat{\mu}$ of the nuisance parameters on the signal strength $\mu_{t\bar{t}H}$ is computed as the difference of the nominal best fit value of μ and the best fit value obtained when fixing the nuisance parameter under scrutiny to its best fit value $\hat{\theta}$ plus/minus its postfit uncertainty (coloured areas). The nuisance parameters are ordered by their impact, and only the 20 highest ranked parameters are shown.

The contributions of the statistical and various groups of systematic uncertainties to the uncertainty in $\mu_{t\bar{t}H}$ are listed in Table 8. The statistical uncertainty is evaluated by fixing all nuisance parameters to their postfit values and repeating the fit. The contribution by a group of systematic uncertainties is evaluated by repeating the fit fixing only the nuisance parameters related to the uncertainty under scrutiny to their postfit values, and subtracting the uncertainty ob-

tained in quadrature from the total uncertainty of the fit in which no parameters are fixed. The total uncertainty of the full fit on the signal strength (± 0.26) is different from the quadratic sum of the listed contributions because of correlations between the nuisance parameters. The statistical uncertainty also includes components from the background normalisations. The total contribution from the systematic uncertainties ($+0.20/-0.21$) is larger than from the statistical uncertainties ($+0.17/-0.16$), which include the uncertainties due to the freely-floating background normalisations, albeit at a similar level. The theoretical uncertainties amount to ± 0.16 and are dominated by the uncertainties of the $t\bar{t}$ + jets background modelling. Experimental uncertainties amount to ± 0.10, dominated by the jet energy scale and resolution as well as b tagging related uncertainties. Systematic uncertainties due to the size of the various simulated samples used to model the background and signal templates are at the same order and amount to $+0.13/-0.12$.

Table 8: Contributions of different sources of uncertainty to the result for the fit to the data (observed) and to the expectation from simulation (expected). The quoted uncertainties $\Delta\mu_{t\bar{t}H}$ in $\mu_{t\bar{t}H}$ are obtained by fixing the listed sources of uncertainties to their postfit values in the fit and subtracting the obtained result in quadrature from the result of the full fit. The statistical uncertainty is evaluated by fixing all nuisance parameters to their postfit values and repeating the fit. The quadratic sum of the contributions is different from the total uncertainty because of correlations between the nuisance parameters.

Uncertainty source	$\Delta\mu_{t\bar{t}H}$ (observed)	$\Delta\mu_{t\bar{t}H}$ (expected)
Total experimental	$+0.10/-0.10$	$+0.11/-0.10$
jet energy scale and resolution	$+0.08/-0.07$	$+0.09/-0.09$
b tagging	$+0.07/-0.06$	$+0.06/-0.02$
luminosity	$+0.02/-0.02$	$+0.01/-0.01$
Total theory	$+0.16/-0.16$	$+0.18/-0.14$
$t\bar{t}$ + jets background	$+0.15/-0.16$	$+0.12/-0.11$
signal modelling	$+0.06/-0.01$	$+0.13/-0.06$
Size of the simulated event samples	$+0.13/-0.12$	$+0.10/-0.10$
Total systematic	$+0.20/-0.21$	$+0.23/-0.19$
Statistical	$+0.17/-0.16$	$+0.17/-0.17$
background normalisation	$+0.13/-0.13$	$+0.13/-0.13$
$t\bar{t}$B and $t\bar{t}$C normalisation	$+0.12/-0.12$	$+0.12/-0.12$
QCD normalisation	$+0.01/-0.01$	$+0.01/-0.01$
Total	$+0.26/-0.26$	$+0.28/-0.25$

The single-channel best-fit results for $\mu_{t\bar{t}H}$ are compatible with the combined result at a level corresponding to a p value of 0.28 (1.1 SD), and the p-value compatibility of the combined result with the SM expectation of $\mu_{t\bar{t}H} = 1$ is 0.02 (2.4 SD). The result obtained in the combination of the leptonic channels with the 2016 data only is compatible to the central value of the earlier result in Ref. [11] obtained from a similar dataset at a p value of 0.41 (0.8 SD). For the p

value computation, the postfit uncertainty model was taken into account in each case.

The observed yields in each bin of the final discriminant distributions of the STXS measurement in the signal regions of the SL and DL channels are shown in Fig. 11, together with the fitted signal and background yields. The results of the STXS measurement are presented in Table 9, which lists the best-fit $\mathrm{t\bar{t}H}$ signal-strength modifier per region of Higgs boson p_{T}, obtained in the combination of all channels and years. The best-fit values and their correlations are also shown in Fig. 12. The goodness-of-fit p value is 0.89, indicating good description of the data by the fit model. The highest expected sensitivity is reached in the medium-p_{T} range between 120 GeV and 300 GeV.

Table 9: Best-fit results of the $\mathrm{t\bar{t}H}$ signal-strength modifier $\mu_{\mathrm{t\bar{t}H}}$ in the different bins of Higgs boson p_{T} of the STXS measurement.

$p_{\mathrm{T}}(\mathrm{H})$	$\hat{\mu} \pm \mathrm{tot}\,(\pm\mathrm{stat} \pm \mathrm{syst})$
0–60 GeV	$0.2^{+1.9}_{-1.8}\ \left(^{+1.2\ +1.4}_{-1.2\ -1.3}\right)$
60–120 GeV	$0.1^{+1.4}_{-1.4}\ \left(^{+1.0\ +0.91}_{-1.0\ -0.96}\right)$
120–200 GeV	$1.14^{+0.95}_{-0.86}\ \left(^{+0.69\ +0.65}_{-0.69\ -0.52}\right)$
200–300 GeV	$0.19^{+0.89}_{-0.90}\ \left(^{+0.65\ +0.60}_{-0.65\ -0.62}\right)$
$>$ 300 GeV	$-1.2^{+1.0}_{-1.1}\ \left(^{+0.80\ +0.61}_{-0.78\ -0.71}\right)$

The measured best-fit values are compatible with the central result of the inclusive $\mathrm{t\bar{t}H}$ production rate measurement, corresponding to an overall p value of 0.67 (0.4 SD). This p value takes into account correctly the correlations between the p_{T} bins shown in Fig. 12 and is thus different from the value obtained using a simple weighted average of the results quoted in Table 9. As an additional compatibility test, a combined fit of the STXS signal templates with a single parameter of interest was performed, resulting in a best-fit value that agrees within 3% to that of the inclusive result with uncertainties that are larger by 20%, which is expected due to the differences in the uncertainty model.

The compatibility with the SM expectation has a p value of 0.21 (1.3 SD). No significant trend in p_{T} compared to the SM expectation is observed. The largest deviation from the SM expectation is observed for $p_{\mathrm{T}} > 300$ GeV with a local significance of approximately 2 SD.

10.2 tH production rate

Second, the tH signal process is targeted. An upper limit at 95% CL on the tH production rate relative to the SM expectation, denoted as the signal strength modifier μ_{tH} of 14.6 is observed, with an expectation of $19.3^{+9.2}_{-6.0}$. Here, the $\mathrm{t\bar{t}H}$ contribution is assumed to conform to the SM expectation ($\mu_{\mathrm{t\bar{t}H}} = 1$) and is treated as background. The obtained limits per channel and year and in their combination are shown in Fig. 13.

Furthermore, a simultaneous fit of the $\mathrm{t\bar{t}H}$ and tH signal strength modifiers is performed. The observed and expected values of the likelihood-ratio test statistic are shown in Fig. 14, with best-fit values of $(\mu_{\mathrm{t\bar{t}H}}, \mu_{\mathrm{tH}})$ of $(0.35, -3.83)$. The correlation between the $\mathrm{t\bar{t}H}$ and tH signal strength modifiers is of moderate size, which demonstrates the discrimination between the two signal processes achieved in this analysis.

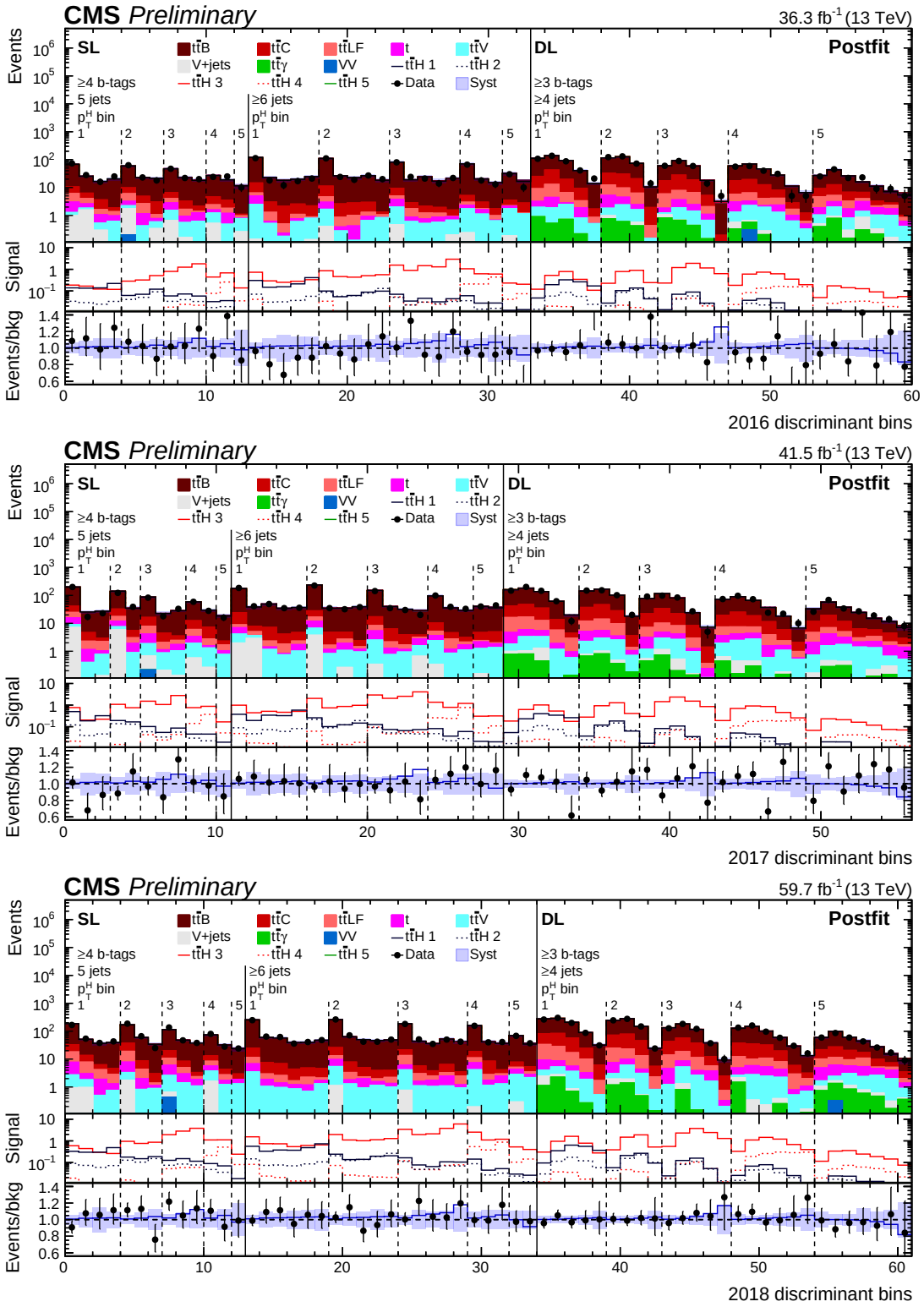


Figure 11: Observed (points) and postfit expected (filled histograms) yields in each STXS analysis discriminant bin in the signal regions of the SL and DL channels for the 2016 (upper), 2017 (middle), and 2018 (lower) data-taking periods. The fitted signal distributions (lines labelled $t\bar{t}H$ 1 to 5) in each Higgs boson p_T bin are shown in the middle pads. The lower pads show the ratio of the data to the background (points) and of the postfit expected total signal+background to the background-only contribution (line). The uncertainty bands include the total uncertainty of the fit model.

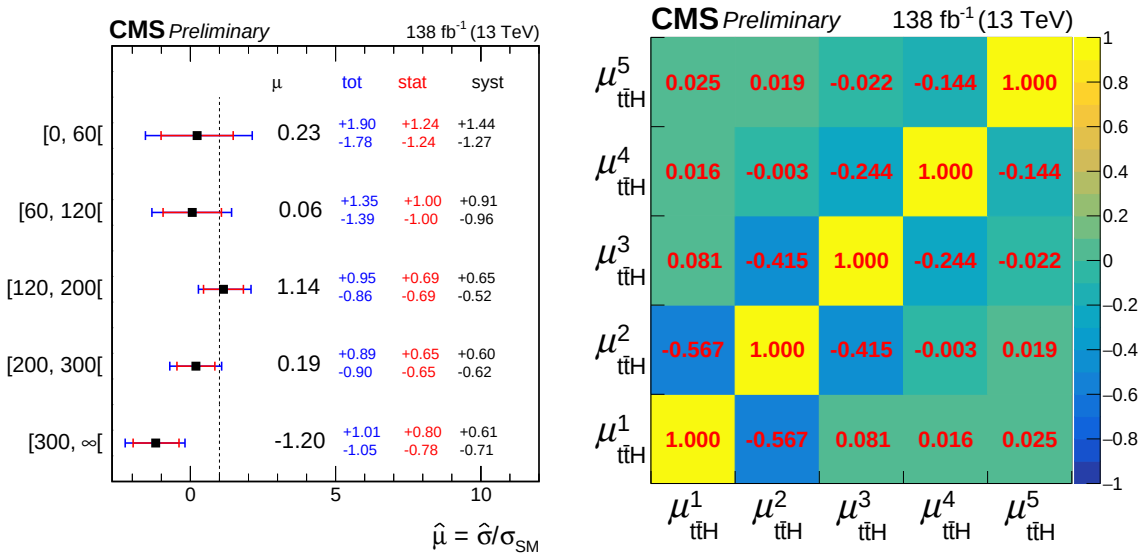


Figure 12: Best-fit results of the $t\bar{t}H$ signal-strength modifiers $\mu_{t\bar{t}H}$ in the different bins of Higgs boson p_T (left) and their correlations (right) of the STXS measurement.

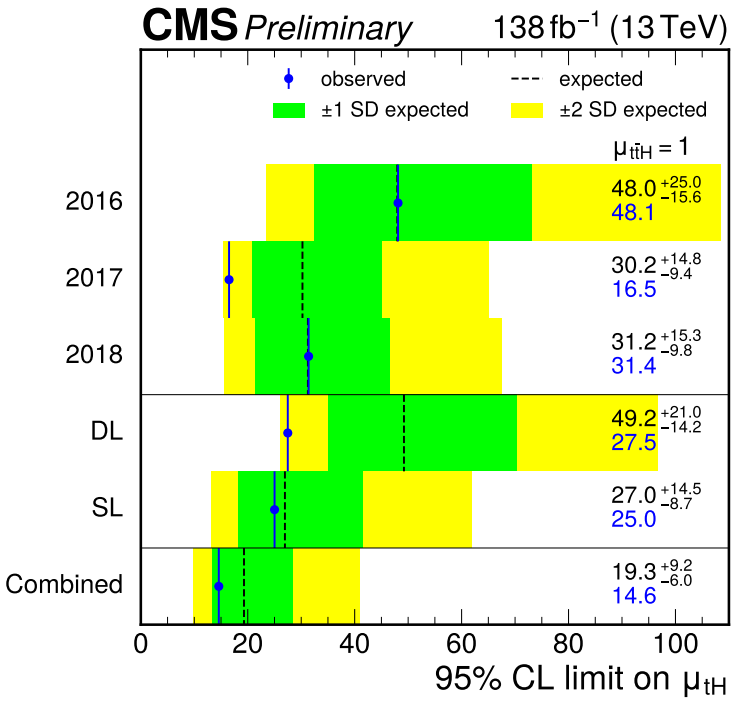


Figure 13: Observed (solid vertical line) and expected (dashed vertical line) upper 95% CL limit on the tH signal strength modifier μ_{tH} for different channels and years, where the uncertainties are uncorrelated between the channels and years, and in their combination. The green (yellow) areas indicate the one (two) standard deviation confidence intervals on the expected limit.

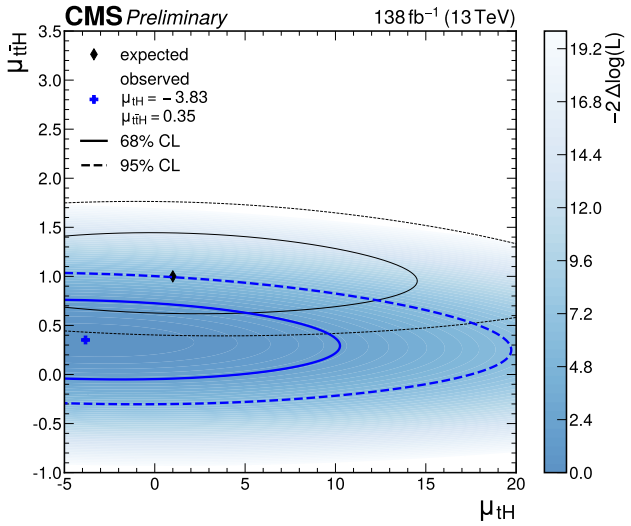


Figure 14: Observed likelihood-ratio test statistic (blue shading) as a function of the $\mathrm{t\bar{t}H}$ and tH signal strength modifiers $\mu_{\mathrm{t\bar{t}H}}$ and μ_{tH}, together with the observed (blue) and SM expected (black) best-fit points (cross and diamond markers) as well as the 68% (solid lines) and 95% (dashed lines) CL regions.

10.3 Coupling measurement

Third, the Higgs boson coupling is analysed in different models, where both $\mathrm{t\bar{t}H}$ and tH are treated as signal.

Assuming SM Higgs boson coupling structure, the coupling strength of the Higgs boson to top quarks and to vector bosons is allowed to vary. This is parameterised in terms of the coupling strength modifiers κ_t and κ_V, which denote the coupling strengths relative to the SM expectation following Ref. [91]. While the $\mathrm{t\bar{t}H}$ production rate is proportional to κ_t^2, for tH production interference occurs between processes in which the Higgs boson couples to the top quark or to the W boson, as shown in Fig. 1. As such, the tH production cross section $\sigma_{\mathrm{tHq/tHW}}$ is sensitive to the relative sign of κ_t and κ_V:

$$\begin{aligned}\sigma_{\mathrm{tHq}} &= \left(2.63 \cdot \kappa_\mathrm{t}^2 + 3.58 \cdot \kappa_\mathrm{V}^2 - 5.21 \cdot \kappa_\mathrm{t}\kappa_\mathrm{V}\right) \sigma_{\mathrm{tHq}}^{\mathrm{SM}} \,, \\ \sigma_{\mathrm{tHW}} &= \left(2.91 \cdot \kappa_\mathrm{t}^2 + 2.40 \cdot \kappa_\mathrm{V}^2 - 4.22 \cdot \kappa_\mathrm{t}\kappa_\mathrm{V}\right) \sigma_{\mathrm{tHW}}^{\mathrm{SM}} \,.\end{aligned}$$

The observed and expected values of the likelihood ratio test statistic for different values of κ_t and κ_V are shown in Fig. 15. Best-fit values of $(\kappa_\mathrm{t}, \kappa_\mathrm{V})$ of $(+0.59, +1.40)$ are observed, compatible with the SM expectation at the level of 2 SD. Assuming $\kappa_\mathrm{V} = 1$, a best-fit value of $+0.54^{+0.19}_{-0.34}$ is obtained.

Furthermore, the CP structure of the top-Higgs coupling is probed for potential non-SM contributions. For this, the amplitude of the top-Higgs interaction is parameterised as in Ref. [92] as

$$\mathcal{A}(\mathrm{Htt}) = -\frac{m_\mathrm{t}}{v}\overline{\psi}_\mathrm{t}\left(\kappa_\mathrm{t} + \mathrm{i}\tilde{\kappa}_\mathrm{t}\gamma_5\right)\psi_\mathrm{t} \,,$$

where ψ_t and $\overline{\psi}_\mathrm{t}$ are a Dirac spinor and its adjoint, respectively, m_t is the top quark mass, κ_t and $\tilde{\kappa}_\mathrm{t}$ denote the coupling strength modifiers to a purely CP-even and a purely CP-odd component, respectively, and v is the vacuum expectation value of the Higgs field. In the SM, $\kappa_\mathrm{t} = 1$ and $\tilde{\kappa}_\mathrm{t} = 0$.

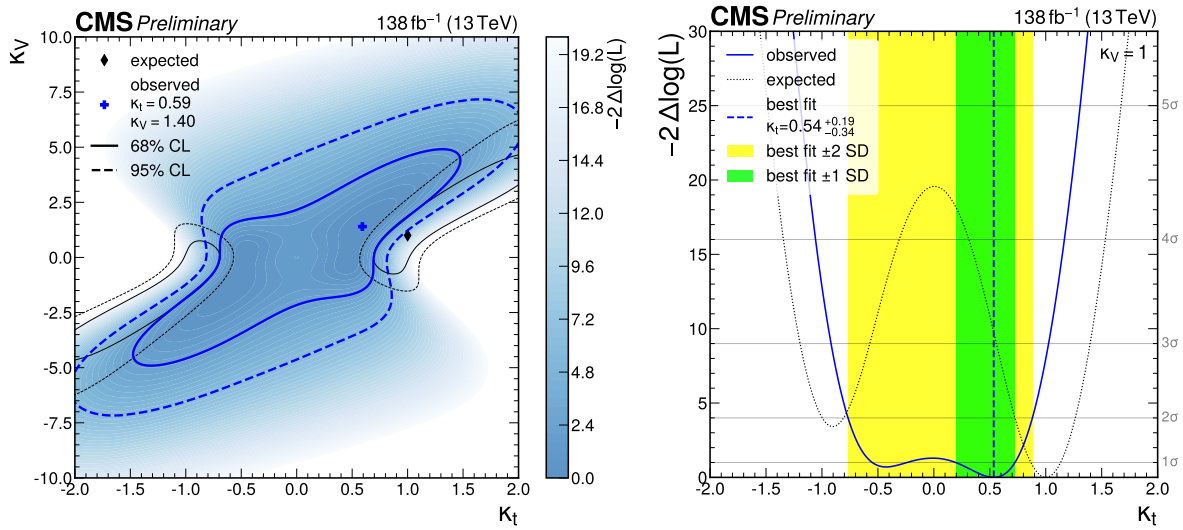


Figure 15: Observed likelihood ratio test statistic (blue shading) as a function of κ_t and κ_V, together with the observed (blue) and SM expected (black) best-fit points (cross and diamond markers) as well as the 68% (solid lines) and 95% (dashed lines) CL regions (left). The observed (solid blue line) and expected (dotted black line) values of the likelihood ratio for $\kappa_V = 1$ are also shown (right), together with the 1 (green area) and 2 (yellow area) standard deviations confidence intervals.

Figure 16 shows the observed and expected values of the likelihood ratio test statistic as a function of κ_t and $\tilde{\kappa}_t$, where κ_V is fixed to the SM value of 1. Best-fit values of $(\kappa_t, \tilde{\kappa}_t)$ of $(+0.53, -0.00)$ are observed, compatible with the SM expectation at the level of 2 SD. The results are also expressed in terms of the CP-odd fraction [92]

$$f_{CP} = \frac{\tilde{\kappa}_t^2}{\tilde{\kappa}_t^2 + \kappa_t^2} \cdot \text{sign}\left(\tilde{\kappa}_t / \kappa_t\right)$$

as well as the CP mixing angle [93]

$$\cos\alpha = \frac{\kappa_t}{\sqrt{\tilde{\kappa}_t^2 + \kappa_t^2}}$$

shown in Fig. 17. The reduction in observed sensitivity relative to the expectation is a consequence of the best-fit value of κ_t being smaller than 1, which leads to a shallower likelihood contour along a circle in $(\kappa_t, \tilde{\kappa}_t)$ space.

11 Summary

A combined analysis of the associated production of a Higgs boson (H) with a top quark-antiquark pair ($t\bar{t}H$) or a single top quark (tH) with the Higgs boson decaying into a bottom quark-antiquark pair has been presented. The analysis has been performed using pp collision data recorded with the CMS detector at a centre-of-mass energy of 13 TeV, corresponding to an integrated luminosity of 138 fb^{-1}. Candidate events are selected in mutually exclusive categories according to the lepton and jet multiplicity, targeting all decay channels of the $t\bar{t}$ system. Neural network discriminants are used to further categorise the events according to the most probable process, targeting the signal and different topologies of the dominant $t\bar{t}$ + jets background, as well as to separate the signal from the background. Compared to previous

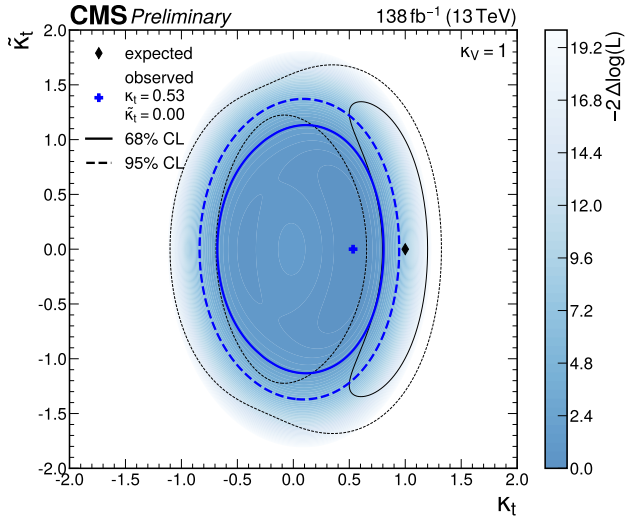


Figure 16: Observed likelihood ratio test statistic (blue shading) as a function of κ_t and $\tilde{\kappa}_t$, where $\kappa_V = 1$, together with the observed (blue) and SM expected (black) best-fit points (cross and diamond markers) as well as the 68% (solid lines) and 95% (dashed lines) CL regions (left).

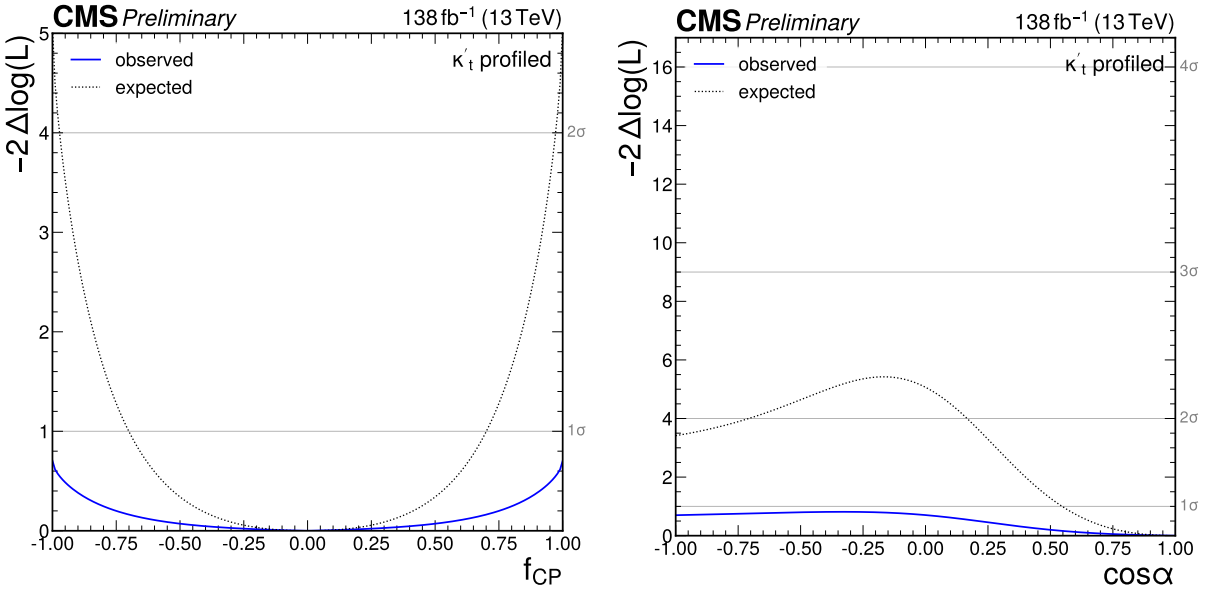


Figure 17: The observed (solid blue line) and expected (dotted black line) likelihood ratio test statistic as a function of f_{CP} (left) and $\cos\alpha$ (right), where κ_V is 1 and $\kappa'_t = \sqrt{\tilde{\kappa}_t^2 + \kappa_t^2}$, the overall modifier of the top-Higgs coupling strength, is profiled such that the likelihood attains its minimum at each point in the plane.

CMS results in this channel, several updates of the analysis strategy as well as modelling of the $t\bar{t}$ + jets background based on state-of-the art $t\bar{t}b\bar{b}$ simulations have been adopted, and an extended set of interpretations is performed.

A best-fit value of the $t\bar{t}H$ production cross section relative to the standard model (SM) expectation of $0.33 \pm 0.26 = 0.33 \pm 0.17\,(\text{stat}) \pm 0.21\,(\text{syst})$ is obtained. The analysis is additionally performed within the Simplified Template Cross Section framework in five intervals of Higgs boson p_{T}, probing potential p_{T} dependent deviations from the SM expectation. An observed (expected) upper limit on the tH production cross section relative to the SM expectation of 14.6 (19.3) at 95% confidence level is derived. Information on the Higgs boson coupling strength is furthermore inferred from a simultaneous fit of the $t\bar{t}H$ and tH production rates, probing either the coupling strength of the Higgs boson to top quarks and to heavy vector bosons, or possible CP-odd admixtures in the coupling between the Higgs boson and top quarks.

References

[1] ATLAS Collaboration, "Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector", *Phys. Lett. B* **784** (2018) 159, doi:10.1016/j.physletb.2018.07.035, arXiv:1806.00425.

[2] CMS Collaboration, "Observation of $t\bar{t}H$ production", *Phys. Rev. Lett.* **120** (2018) 231801, doi:10.1103/PhysRevLett.120.231801, arXiv:1804.02610.

[3] ATLAS Collaboration, "Observation of $H \to b\bar{b}$ decays and VH production with the ATLAS detector", *Phys. Lett. B* **786** (2018) 59, doi:10.1016/j.physletb.2018.09.013, arXiv:1808.08238.

[4] CMS Collaboration, "Observation of Higgs boson decay to bottom quarks", *Phys. Rev. Lett.* **121** (2018) 121801, doi:10.1103/PhysRevLett.121.121801, arXiv:1808.08242.

[5] F. Maltoni, K. Paul, T. Stelzer, and S. Willenbrock, "Associated production of Higgs and single top at hadron colliders", *Phys. Rev. D* **64** (2001) 094023, doi:10.1103/PhysRevD.64.094023, arXiv:hep-ph/0106293.

[6] M. Farina et al., "Lifting degeneracies in Higgs couplings using single top production in association with a Higgs boson", *JHEP* **05** (2013) 022, doi:10.1007/JHEP05(2013)022, arXiv:1211.3736.

[7] P. Agrawal, S. Mitra, and A. Shivaji, "Effect of anomalous couplings on the associated production of a single top quark and a Higgs boson at the LHC", *JHEP* **12** (2013) 077, doi:10.1007/JHEP12(2013)077, arXiv:1211.4362.

[8] F. Demartin, F. Maltoni, K. Mawatari, and M. Zaro, "Higgs production in association with a single top quark at the LHC", *Eur. Phys. J. C* **75** (2015) 267, doi:10.1140/epjc/s10052-015-3475-9, arXiv:1504.00611.

[9] LHC Higgs Cross Section Working Group, "Handbook of LHC Higgs cross sections: 4. deciphering the nature of the Higgs sector", 2016. arXiv:1610.07922.

[10] ATLAS Collaboration, "Measurement of Higgs boson decay into b-quarks in associated production with a top-quark pair in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *JHEP* **06** (2022) 097, doi:10.1007/JHEP06(2022)097, arXiv:2111.06712.

[11] CMS Collaboration, "Search for $\mathrm{t\bar{t}H}$ production in the $\mathrm{H} \to \mathrm{b\bar{b}}$ decay channel with leptonic $\mathrm{t\bar{t}}$ decays in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *JHEP* **03** (2019) 026, doi:10.1007/JHEP03(2019)026, arXiv:1804.03682.

[12] CMS Collaboration, "Search for $\mathrm{t\bar{t}H}$ production in the all-jet final state in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *JHEP* **06** (2018) doi:10.1007/JHEP06(2018)101, arXiv:1803.06986.

[13] CMS Collaboration, "Search for associated production of a Higgs boson and a single top quark in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *Phys. Rev. D* **99** (2019) 092005, doi:10.1103/PhysRevD.99.092005, arXiv:1811.09696.

[14] ATLAS Collaboration, "Probing the *CP* nature of the top-Higgs Yukawa coupling in $\mathrm{t\bar{t}H}$ and tH events with $\mathrm{H} \to \mathrm{b\bar{b}}$ decays using the ATLAS detector at the LHC", 2023. arXiv:2303.05974. Submitted to PLB.

[15] CMS Collaboration, "The CMS experiment at the CERN LHC", *JINST* **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.

[16] CMS Collaboration, "Performance of the CMS level-1 trigger in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *JINST* **15** (2020) P10017, doi:10.1088/1748-0221/15/10/P10017, arXiv:2006.10165.

[17] CMS Collaboration, "The CMS trigger system", *JINST* **12** (2017) P01020, doi:10.1088/1748-0221/12/01/P01020, arXiv:1609.02366.

[18] CMS Collaboration, "Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ in 2015 and 2016 at CMS", *Eur. Phys. J. C* **81** (2021) 800, doi:10.1140/epjc/s10052-021-09538-2, arXiv:2104.01927.

[19] CMS Collaboration, "CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13\,\text{TeV}$", CMS Physics Analysis Summary CMS-PAS-LUM-17-004, 2017.

[20] CMS Collaboration, "CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13\,\text{TeV}$", CMS Physics Analysis Summary CMS-PAS-LUM-18-002, 2018.

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

[22] P. Nason, "A new method for combining NLO QCD with shower Monte Carlo algorithms", *JHEP* **11** (2004) 040, doi:10.1088/1126-6708/2004/11/040, arXiv:hep-ph/0409146.

[23] 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.

[24] S. Alioli, P. Nason, C. Oleari, and E. Re, "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX", *JHEP* **06** (2010) 043, doi:10.1007/JHEP06(2010)043, arXiv:1002.2581.

[25] T. Ježo and P. Nason, "On the treatment of resonances in next-to-leading order calculations matched to a parton shower", *JHEP* **12** (2015) 065, doi:10.1007/JHEP12(2015)065, arXiv:1509.09071.

[26] H. B. Hartanto, B. Jager, L. Reina, and D. Wackeroth, "Higgs boson production in association with top quarks in the POWHEG BOX", *Phys. Rev. D* **91** (2015) 094003, doi:10.1103/PhysRevD.91.094003, arXiv:1501.04498.

[27] 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.

[28] 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.

[29] NNPDF Collaboration, "Parton distributions from high-precision collider data", *Eur. Phys. J. C* **77** (2017) 663, doi:10.1140/epjc/s10052-017-5199-5, arXiv:1706.00428.

[30] CMS Collaboration, "Extraction and validation of a new set of CMS PYTHIA 8 tunes from underlying-event measurements", *Eur. Phys. J. C* **80** (2020) 4, doi:10.1140/epjc/s10052-019-7499-4, arXiv:1903.12179.

[31] CMS Collaboration, "Event generator tunes obtained from underlying event and multiparton scattering measurements", *Eur. Phys. J. C* **76** (2016) 155, doi:10.1140/epjc/s10052-016-3988-x, arXiv:1512.00815.

[32] F. Maltoni, G. Ridolfi, and M. Ubiali, "b-initiated processes at the LHC: a reappraisal", *JHEP* **07** (2012) 022, doi:10.1007/JHEP07(2012)022, arXiv:1203.6393. [Erratum: doi:10.1007/JHEP04(2013)095].

[33] F. Demartin et al., "tWH associated production at the LHC", *Eur. Phys. J. C* **77** (2017) 34, doi:10.1140/epjc/s10052-017-4601-7, arXiv:1607.05862.

[34] J. S. Gainer et al., "Exploring theory space with Monte Carlo reweighting", *JHEP* **10** (2014) 078, doi:10.1007/JHEP10(2014)078, arXiv:1404.7129.

[35] O. Mattelaer, "On the maximal use of Monte Carlo samples: re-weighting events at NLO accuracy", *Eur. Phys. J. C* **76** (2016) 674, doi:10.1140/epjc/s10052-016-4533-7, arXiv:1607.00763.

[36] T. Ježo, J. M. Lindert, N. Moretti, and S. Pozzorini, "New NLOPS predictions for $t\bar{t}$ + b-jet production at the LHC", *Eur. Phys. J. C* **78** (2018) 502, doi:10.1140/epjc/s10052-018-5956-0, arXiv:1802.00426.

[37] F. Buccioni et al., "OpenLoops 2", *Eur. Phys. J. C* **79** (2019) 866, doi:10.1140/epjc/s10052-019-7306-2, arXiv:1907.13071.

[38] F. Cascioli et al., "NLO matching for $t\bar{t}b\bar{b}$ production with massive b-quarks", *Phys. Lett. B* **734** (2014) 210, doi:10.1016/j.physletb.2014.05.040, arXiv:1309.5912.

[39] F. Buccioni, S. Kallweit, S. Pozzorini, and M. F. Zoller, "NLO QCD predictions for $t\bar{t}b\bar{b}$ production in association with a light jet at the LHC", *JHEP* **12** (2019) 015, doi:10.1007/JHEP12(2019)015, arXiv:1907.13624.

[40] S. Alioli, P. Nason, C. Oleari, and E. Re, "NLO single-top production matched with shower in POWHEG: *s*- and *t*-channel contributions", *JHEP* **09** (2009) 111, doi:10.1088/1126-6708/2009/09/111, arXiv:0907.4076. [Erratum: doi:10.1007/JHEP02(2010)011].

[41] 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.

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

[43] J. Alwall et al., "Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions", *Eur. Phys. J. C* **53** (2008) 473, doi:10.1140/epjc/s10052-007-0490-5, arXiv:0706.2569.

[44] CMS Collaboration, "A measurement of the Higgs boson mass in the diphoton decay channel", *Phys. Lett. B* **805** (2020) 135425, doi:10.1016/j.physletb.2020.135425, arXiv:2002.06398.

[45] 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) 612, doi:10.1016/j.physletb.2012.03.013, arXiv:1111.5869.

[46] P. Bärnreuther, M. Czakon, and A. Mitov, "Percent-level-precision physics at the Tevatron: next-to-next-to-leading order QCD corrections to $q\bar{q} \to t\bar{t}$+X", *Phys. Rev. Lett.* **109** (2012) 132001, doi:10.1103/PhysRevLett.109.132001, arXiv:1204.5201.

[47] M. Czakon and A. Mitov, "NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels", *JHEP* **12** (2012) 054, doi:10.1007/JHEP12(2012)054, arXiv:1207.0236.

[48] M. Czakon and A. Mitov, "NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction", *JHEP* **01** (2013) 080, doi:10.1007/JHEP01(2013)080, arXiv:1210.6832.

[49] M. Beneke, P. Falgari, S. Klein, and C. Schwinn, "Hadronic top-quark pair production with NNLL threshold resummation", *Nucl. Phys. B* **855** (2012) 695, doi:10.1016/j.nuclphysb.2011.10.021, arXiv:1109.1536.

[50] 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) 252004, doi:10.1103/PhysRevLett.110.252004, arXiv:1303.6254.

[51] 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) 2930, doi:10.1016/j.cpc.2014.06.021, arXiv:1112.5675.

[52] N. Kidonakis, "Two-loop soft anomalous dimensions for single top quark associated production with W^- or H^-", *Phys. Rev. D* **82** (2010) 054018, doi:10.1103/PhysRevD.82.054018, arXiv:1005.4451.

[53] M. Aliev et al., "HATHOR: HAdronic Top and Heavy quarks crOss section calculatoR", *Comput. Phys. Commun.* **182** (2011) 1034, doi:10.1016/j.cpc.2010.12.040, arXiv:1007.1327.

[54] P. Kant et al., "HatHor for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions", *Comput. Phys. Commun.* **191** (2015) 74, doi:10.1016/j.cpc.2015.02.001, arXiv:1406.4403.

[55] 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", *JHEP* **02** (2016) 113, doi:10.1007/JHEP02(2016)113, arXiv:1507.05640.

[56] 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.

[57] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector", *JINST* **12** (2017) P10003, doi:10.1088/1748-0221/12/10/P10003, arXiv:1706.04965.

[58] CMS Collaboration, "Technical proposal for the phase-2 upgrade of the Compact Muon Solenoid", CMS Technical Proposal CERN-LHCC-2015-010, CMS-TDR-15-02, 2015.

[59] CMS Collaboration, "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *JINST* **13** (2018) P06015, doi:10.1088/1748-0221/13/06/P06015, arXiv:1804.04528.

[60] CMS Collaboration, "Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC", *JINST* **16** (2021) P05014, doi:10.1088/1748-0221/16/05/P05014, arXiv:2012.06888.

[61] CMS Collaboration, "ECAL 2016 refined calibration and Run2 summary plots", CMS Detector Performance Summary CMS-DP-2020-021, 2020.

[62] 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.

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

[64] 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.

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

[66] CMS Collaboration, "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV", *JINST* **13** (2018) P05011, doi:10.1088/1748-0221/13/05/P05011, arXiv:1712.07158.

[67] E. Bols et al., "Jet flavour classification using DeepJet", *JINST* **15** (2020) P12012, doi:10.1088/1748-0221/15/12/P12012, arXiv:2008.10519.

[68] CMS Collaboration, "B-tagging performance of the CMS legacy dataset 2018", CMS Detector Performance Summary CMS-DP-2021-004, 2021.

[69] CMS Collaboration, "Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ using the CMS detector", *JINST* **14** (2019) P07004, doi:10.1088/1748-0221/14/07/P07004, arXiv:1903.06078.

[70] K. Kondo, "Dynamical likelihood method for reconstruction of events with missing momentum. 1: Method and toy models", *J. Phys. Soc. Jap.* **57** (1988) 4126, doi:10.1143/JPSJ.57.4126.

[71] CMS Collaboration, "Inclusive and differential cross section measurements of $\mathrm{t\bar{t}b\bar{b}}$ production in the lepton+jets channel at $\sqrt{s} = 13\,\text{TeV}$ with the CMS detector", CMS Physics Analysis Summary CMS-PAS-TOP-22-009, 2023.

[72] CMS Collaboration, "Measurement of the cross section for $\mathrm{t\bar{t}}$ production with additional jets and b jets in pp collisions at $\sqrt{s} = 13\,\text{TeV}$", *JHEP* **07** (2020) 125, doi:10.1007/JHEP07(2020)125, arXiv:2003.06467.

[73] CMS Collaboration, "Measurement of the $\mathrm{t\bar{t}b\bar{b}}$ production cross section in the all-jet final state in pp collisions at $\sqrt{s} = 13\,\text{TeV}$", *Phys. Lett. B* **803** (2020) 135285, doi:10.1016/j.physletb.2020.135285, arXiv:1909.05306.

[74] ATLAS and CMS Collaborations, and the LHC Higgs Combination Group, "Procedure for the LHC Higgs boson search combination in summer 2011", ATL-PHYS-PUB-2011-011, CMS NOTE-2011/005, 2011.

[75] J. S. Conway, "Incorporating nuisance parameters in likelihoods for multisource spectra", in *PHYSTAT 2011*. 3, 2011. arXiv:1103.0354. doi:10.5170/CERN-2011-006.115.

[76] J. D. Bjorken and S. J. Brodsky, "Statistical model for electron-positron annihilation into hadrons", *Phys. Rev. D* **1** (1970) 1416, doi:10.1103/PhysRevD.1.1416.

[77] G. C. Fox and S. Wolfram, "Event shapes in $\mathrm{e^+e^-}$ annihilation", *Nuclear Physics B* **157** (1979) 543, doi:10.1016/0550-3213(79)90120-2.

[78] I. Goodfellow, Y. Bengio, and A. Courville, "Deep Learning". MIT Press, 2016. http://www.deeplearningbook.org.

[79] J. K. Lindsey, "Parametric statistical inference". Clarendon Press, Oxford, England, 1996.

[80] CMS Collaboration, "Performance of deep tagging algorithms for boosted double quark jet topology in proton-proton collisions at 13 TeV with the phase-0 CMS detector", CMS Detector Performance Summary CMS-DP-2018-046, 2018.

[81] F. Chollet et al., "Keras". https://keras.io, 2015.

[82] S. Jasper, L. Hugo, and P. A. Ryan, "Practical bayesian optimization of machine learning algorithms", 2012. arXiv:1206.2944.

[83] F. Luca, D. Michele, F. Paolo, and P. Massimiliano, "Forward and reverse gradient-based hyperparameter optimization", in *Proceedings of the 34th International Conference on Machine Learning*, D. Precup and Y. W. Teh, eds., volume 70 of *Proceedings of Machine Learning Research*, p. 1165. PMLR, 2017.

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

[85] P. Skands, S. Carrazza, and J. Rojo, "Tuning PYTHIA 8.1: the Monash 2013 tune", *Eur. Phys. J. C* **74** (2014) 3024, doi:10.1140/epjc/s10052-014-3024-y, arXiv:1404.5630.

[86] CMS Collaboration, "Measurement of differential cross sections for the production of top quark pairs and of additional jets in lepton+jets events from pp collisions at $\sqrt{s} = 13\,\text{TeV}$", *Phys. Rev. D* **97** (2018), no. 11, 112003, doi:10.1103/PhysRevD.97.112003, arXiv:1803.08856.

[87] J. R. Andersen et al., "Les Houches 2017: Physics at TeV colliders standard model working group report", 2018. arXiv:1803.07977.

[88] CMS Tracker Group Collaboration, "The CMS phase-1 pixel detector upgrade", *JINST* **16** (2021) P02027, doi:10.1088/1748-0221/16/02/P02027, arXiv:2012.14304.

[89] R. Barlow and C. Beeston, "Fitting using finite Monte Carlo samples", *Comp. Phys. Commun.* **77** (1993) 219, doi:10.1016/0010-4655(93)90005-W.

[90] CMS Collaboration, "First measurement of the cross section for top quark pair production with additional charm jets using dileptonic final states in pp collisions at $\sqrt{s} = 13\,\text{TeV}$", *Phys. Lett. B* **820** (2021) 136565, doi:10.1016/j.physletb.2021.136565, arXiv:2012.09225.

[91] LHC Higgs Cross Section Working Group, "Handbook of LHC Higgs cross sections: 3. Higgs properties", 2013. arXiv:1307.1347.

[92] A. V. Gritsan, R. Röntsch, M. Schulze, and M. Xiao, "Constraining anomalous Higgs boson couplings to the heavy flavor fermions using matrix element techniques", *Phys. Rev. D* **94** (2016) 055023, doi:10.1103/PhysRevD.94.055023, arXiv:1606.03107.

[93] F. Demartin et al., "Higgs characterisation at NLO in QCD: CP properties of the top-quark Yukawa interaction", *Eur. Phys. J. C* **74** (2014) 3065, doi:10.1140/epjc/s10052-014-3065-2, arXiv:1407.5089.