CMS Physics Analysis Summary

Contact: cms-pag-conveners-higgs@cern.ch 2019/06/24

Measurement of $\mathrm{t\bar{t}H}$ production in the $\mathrm{H} \to \mathrm{b\bar{b}}$ decay channel in 41.5 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13\,\mathrm{TeV}$

The CMS Collaboration

Abstract

A measurement of the associated production of a standard model Higgs boson with a top quark-antiquark pair ($\mathrm{t\bar{t}H}$) in proton-proton collisions at $\sqrt{s} = 13\,\mathrm{TeV}$ is presented. The result is based on data recorded with the CMS detector at the CERN LHC in 2017 and corresponds to an integrated luminosity of 41.5 fb^{-1}. Candidate $\mathrm{t\bar{t}H}$ events are selected based on the number of leptons in the event, targeting all $\mathrm{t\bar{t}}$ decay channels, and are categorised according to the number of jets. Multivariate analysis techniques are employed to further categorise the events and eventually discriminate between signal and background. A combined fit of multivariate discriminant distributions in all categories results in a best fit value of the $\mathrm{t\bar{t}H}$ signal strength relative to the standard model cross section, $\mu = \sigma/\sigma_{\mathrm{SM}}$, of $\hat{\mu} = 1.49\,^{+0.21}_{-0.20}(\mathrm{stat})\,^{+0.39}_{-0.35}(\mathrm{syst})$, corresponding to an observed (expected) significance of 3.7 (2.6) standard deviations. Combined with previous results obtained with 35.9 fb^{-1} of data recorded in 2016, a best-fit value of $\hat{\mu} = 1.15\,^{+0.15}_{-0.15}(\mathrm{stat})\,^{+0.28}_{-0.25}(\mathrm{syst})$ is found, corresponding to an observed (expected) significance of 3.9 (3.5) standard deviations above the background-only hypothesis.

This document has been revised with respect to the version dated May 19, 2019.

1 Introduction

In the standard model (SM), 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 predict different coupling strengths. The associated production of a Higgs boson and a top quark-antiquark pair ($\mathrm{t\bar{t}H}$ production) provides a direct probe of the top-Higgs coupling and has recently been observed by the ATLAS and CMS Collaborations [1, 2].

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 [3] and is thus experimentally attractive as a final state. Existing searches for $\mathrm{t\bar{t}H}$ production in the $\mathrm{H} \to \mathrm{b\bar{b}}$ channel achieve sensitivities that correspond to observed (expected) significances of up to 1.6 (2.2) standard deviations [4–6]. This article describes an updated measurement of $\mathrm{t\bar{t}H}$ production in the $\mathrm{b\bar{b}}$ decay channel of the Higgs boson by the CMS Collaboration using the 2017 dataset of 41.5 fb^{-1} of pp collision data at 13 TeV centre-of-mass energy and covering all decay channels of the $\mathrm{t\bar{t}}$ system. The analysis follows the strategies developed in Refs. [5, 6] with the 2016 data. Several refinements over the previous analyses are included, which together lead to an improvement in expected sensitivity of approximately 20%: improvements in the multivariate classifiers and the parton-shower uncertainty modelling, a more efficient QCD rejection, improvements in the b tagging algorithm, and the combination of all three $\mathrm{t\bar{t}}$ decay channels.

The analysis begins by identifying pp collision events consistent with the production of a top quark-antiquark pair with additional b quark jets. The decays of the W bosons from the top quarks determine the specific $\mathrm{t\bar{t}}$ signatures recorded in the detector. This analysis covers all three $\mathrm{t\bar{t}}$ decay channels: the fully-hadronic channel, where both W bosons decay into quarks, the single-lepton channel, where one W boson decays into a charged lepton (electron or muon) and a neutrino and the other W boson decays into quarks, and the dilepton channel, where both W bosons decay into a charged lepton (electron or muon) and a neutrino. τ leptons from the W boson decays are not reconstructed explicitly: events with leptonically decaying τ leptons contribute to the leptonic channels, and events with hadronically decaying τ leptons contribute to the single-lepton or the fully-hadronic channel. The final states thus contain a large number of high-p_{T} jets as well as charged leptons and $\not\!\!E_{\mathrm{T}}$ arising from neutrinos in the semi-leptonic and dilepton cases.

The dominant background contributions arise from QCD multijet production in the fully-hadronic channel and from $\mathrm{t\bar{t}}$+jets production in all channels. The latter include $\mathrm{t\bar{t}}$ + light-flavour jets where one or more of the jets is misidentified as b quark jet, as well as $\mathrm{t\bar{t}+c\bar{c}}$ and $\mathrm{t\bar{t}+b\bar{b}}$ backgrounds, where additional b or c quarks can arise from QCD radiation or loop-induced QCD processes. The $\mathrm{t\bar{t}+b\bar{b}}$ background, in particular, 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. Minor background contributions arise from single-top quark, W+jets and Z+jets, $\mathrm{t\bar{t}}$+W and $\mathrm{t\bar{t}}$+Z, as well as diboson production.

The sensitivity of the analysis is limited by a combinatorial background due to multiple b quark 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 multivariate analysis techniques that use the differential distributions of several experimental observables simultaneously. These include boosted decision trees (BDTs) and artificial neural networks (ANNs) optimised on simulated events, complemented by a discriminant based on the direct evaluation of the leading-order $\mathrm{t\bar{t}H}$ and $\mathrm{t\bar{t}+b\bar{b}}$ matrix elements on an event-by-event basis, further

referred to as the matrix element method (MEM).

The analysis proceeds as follows: first, events are separated based the number of isolated, high-p_{T} leptons and a minimum number of jets and b-tagged jets into three independent channels targeting the decay channels of the $\mathrm{t\bar{t}}$ system. Within each channel, events are further categorised based on the jet multiplicity and b-tag information into sub-samples with varying signal purity and different background composition. Categories with low signal purity are useful for constraining background estimates and systematic uncertainties, while categories with higher signal purity provide sensitivity to $\mathrm{t\bar{t}H}$ production. Backgrounds are modelled using simulated samples corrected to account for known theoretical and experimental deficiencies and using data-driven methods.

The analysis strategy has been optimised individually in each channel. In the fully-hadronic channel, events are categorised by the jet and b-tag multiplicity. Background contributions from QCD-multijet production are suppressed exploiting angular variables that are sensitive to the spatial distribution of the jets in the event, and residual contributions are modelled from data. Eventually, the MEM is used to construct a final discriminant per category. In the single-lepton channel, ANNs are employed to perform a multi-classification of an event as either signal or any of five different $\mathrm{t\bar{t}}$+jets background processes. Events are consequently categorised by the jet multiplicity and the most-probable process according to the ANN classification, and the corresponding ANN classifier output is used as final discriminant. In the dilepton channel, events are categorised by the jet and b-tag multiplicity, and classification BDTs are used as final discriminant. The MEM output is used as input variable for the ANNs and BDTs in the single-lepton and dilepton channels, respectively. Finally, the signal is extracted in a simultaneous template fit of the signal and background rates in all the categories.

This document is structured as follows. The data and simulated samples are described in Section 2, followed by the object and event reconstruction and the event selection in Sections 3 and 4, respectively. The analysis strategy in the different channels is detailed in Section 5 and systematic uncertainties are discussed in Section 6. Finally, the results are presented in Section 7.

2 Data samples and Monte Carlo simulation

The analysis has been performed with pp collision data collected at 13 TeV centre-of-mass energy in 2017, corresponding to an integrated luminosity of 41.5 fb^{-1}. The data were selected using different trigger paths that require the presence of two or one leptons, one lepton and additional jets, or several jets in the events depending on the analysis channel, as detailed in Section 4.

Several Monte Carlo (MC) event generators, interfaced with a detailed detector simulation based on GEANT4 (v.9.4) [7], were used to model signal and background events. The $\mathrm{t\bar{t}H}$ signal is simulated at next-to-leading order (NLO) of QCD perturbation theory with the event generator POWHEG (v.2) [8–11]. All decay channels of the Higgs boson are considered. SM backgrounds are simulated at NLO with POWHEG or MADGRAPH5_aMC@NLO (v.2.2.2) [12], or at leading order (LO) using PYTHIA (v.8.230) [13], 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 functions (PDF) NNPDF3.1 [14]. Parton showering and hadronisation are simulated with PYTHIA [13]. The parameters for the underlying event description correspond to the CP5 tune [15] for all signal and background

processes. In case of the POWHEG samples, a damping parameter value of $\texttt{hdamp} = 237.9\,\text{GeV}$ has been used.

Major background contributions arise from $\mathrm{t\bar{t}}$ production. They are simulated using POWHEG. 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), $\mathrm{t\bar{t}}$ production in association with a W or Z boson (referred to as $\mathrm{t\bar{t}}$+V), and diboson (W W, W Z, and ZZ) processes. The single top quark processes in the *t*- and tW-channels are simulated with POWHEG [16, 17]. The *s*-channel single top quark processes as well as V+jets and $\mathrm{t\bar{t}}$+V processes are simulated with MADGRAPH5_aMC@NLO, where for the V+jets processes the matching of matrix-element (ME) jets to parton showers (PS) is performed using the FXFX [18] prescription. Diboson production is simulated using the PYTHIA event generator. In the fully-hadronic channel, the dominant background is due to QCD multijet production. Its contribution is estimated from data, as described in Section 5.

For comparison with the observed distributions, the event yields in the simulated samples are normalised to the same integrated luminosity of the data sample, according to their predicted cross sections. The $\mathrm{t\bar{t}H}$ signal cross section of $507^{+35}_{-50}\,\text{fb}$ and Higgs boson branching fractions are taken from calculations at NLO accuracy [3]. The $\mathrm{t\bar{t}}$ cross section of 831.76 pb corresponds to next-to-next-to-leading order (NNLO) with resummation to next-to-next-to-leading-logarithmic accuracy [19–25]. The cross sections of the other backgrounds are taken at NNLO (V+jets), approximate NNLO (single top quark tW channel [26]), and NLO (single top quark *t*- and *s*-channels [27, 28], $\mathrm{t\bar{t}}$+V [29], and diboson [30]) accuracy.

The simulated $\mathrm{t\bar{t}}$ events are separated into the following processes, based on the flavour of additional jets that do not originate from the top quark decays: $\mathrm{t\bar{t}+b\bar{b}}$, defined at generator level as the events in which at least two additional jets are generated within the acceptance requirements (described in Section 3), each of which originates from one or more B hadrons; $\mathrm{t\bar{t}}$+b, for which only one additional jet within the acceptance originates from a single B hadron; $\mathrm{t\bar{t}}$+2b, which corresponds to events with two additional B hadrons that are close enough in direction to produce a single jet; $\mathrm{t\bar{t}+c\bar{c}}$, for which events have at least one additional jet from c hadrons within the acceptance and no additional jets from B hadrons; $\mathrm{t\bar{t}}$ + light flavour jets ($\mathrm{t\bar{t}}$+lf), which corresponds to events that do not belong to any of the above processes. The $\mathrm{t\bar{t}+b\bar{b}}$, $\mathrm{t\bar{t}}$+b, $\mathrm{t\bar{t}}$+2b, and $\mathrm{t\bar{t}+c\bar{c}}$ processes are collectively referred to as $\mathrm{t\bar{t}}$+hf in the following. This separation is important because the different processes are subject to different systematic uncertainties.

Effects from additional pp interactions in the same 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.

In a small fraction of events, deadtime of the triggers, due to incorrect assignment of trigger signals to previous bunch crossings, cause the event to be lost, leading effectively to an inefficiency of the trigger by 2–3%. This is taken into account by scaling down the event yields predicted in the simulation, based on the probability of the events to cause the prefiring.

Further correction factors described in Section 3 are applied to the simulation where necessary to improve the description of the data.

3 Object and event reconstruction

Events are reconstructed using a particle-flow (PF) technique [31], which combines information from all subdetectors to enhance the reconstruction performance by identifying individual particle candidates in the pp collisions. An interaction vertex [32] is required within 24 cm of the detector centre along the beamline direction, and within 2 cm of the beamline in the transverse plane. Among all reconstructed vertices, the one with the largest value of summed physics-object p_{T}^2 is taken to be the primary interaction vertex. The physics objects are the jets defined below, clustered with the tracks assigned to the vertex as inputs, and the associated missing transverse momentum, taken as the negative vector sum of the p_{T} of those jets. All other interaction vertices are considered as pileup vertices. Charged-particle tracks identified as hadrons from pileup vertices are omitted in the subsequent event reconstruction.

The missing transverse momentum vector $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ is defined as the projection of the negative vector sum of the momenta of all reconstructed PF objects in an event on the plane perpendicular to the beams. Its magnitude is referred to as $p_{\mathrm{T}}^{\mathrm{miss}}$.

Electron and muon candidates are required to be within $|\eta| < 2.4$, i.e. the detector region covered by the central tracking detector, and they have to be sufficiently isolated from nearby jet activity as follows. For each electron (muon) candidate, a cone of $\Delta R = 0.3\,(0.4)$ is constructed around the direction of the track at the event vertex, where ΔR is defined as $\sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$, and $\Delta\eta$ and $\Delta\phi$ are the distances in the pseudorapidity and azimuthal angle. Excluding the contribution from the lepton candidate, the scalar p_{T} sum of all particle candidates inside the cone consistent with arising from the chosen primary event vertex is calculated. The neutral component from pileup interactions is subtracted event-by-event, based on the average transverse energy deposited by neutral particles in the event in the case of electrons, and half the transverse momentum carried by charged-particles identified to come from pileup vertices in the case of muons. A relative isolation discriminant I_{rel} is defined as the ratio of this sum to the p_{T} of the lepton candidate. In all three $\mathrm{t\bar{t}}$ decay channels, electron candidates are required to fulfil the same I_{rel} criterion of either $I_{\mathrm{rel}} < 0.036$ or $I_{\mathrm{rel}} < 0.094$ depending on whether the electron was detected in the barrel ($|\eta| < 1.479$) or the endcap ($|\eta| > 1.479$) region of the detector. To reject misidentified electrons or electrons from identified photon conversions, electron candidates are required to fulfil quality criteria regarding the shape of the associated electromagnetic shower in the electromagnetic calorimeters and the number of hits associated with the electron track [33]. Furthermore candidates in the transition region between the barrel and endcap calorimeters, $1.4442 < |\eta| < 1.5560$, are excluded. Muon candidates are required to fulfil $I_{\mathrm{rel}} < 0.15$ in the fully-hadronic and single-lepton channels and $I_{\mathrm{rel}} < 0.25$ in the dilepton channel. To further increase the purity of muons originating from the primary interaction and to suppress misidentified muons or muons from decay-in-flight processes, additional quality criteria, such as a minimal number of hits associated with the muon track, are required in both the silicon tracker and the muon system [34].

Jets are reconstructed from the PF particle candidates using the anti-k_{T} clustering algorithm [35] with a distance parameter of 0.4, as implemented in FastJet [36]. Charged hadrons that are associated to pileup vertices are discarded from the clustering. The jet energy is corrected for the remaining neutral-hadron pileup component in a manner similar to that used to find the energy within the lepton isolation cone [37]. Jet energy corrections are also applied as a function of jet p_{T} and η [38] to data and simulation. All selected jets are required to satisfy $|\eta| < 2.4$.

Jets originating from the hadronisation of b quarks are identified using the deep combined secondary vertex algorithm (DeepCSV) [39], which provides a b tagging discriminant by combining identified secondary vertices and track-based lifetime information. Jets are considered

b tagged if they pass a discriminant value such that the probability of tagging jets originating from light-flavour quarks (u, d, or s) or gluons is approximately 1%, and the corresponding efficiency for tagging jets from b (c) quarks is approximately 76% (18%) [40]. The shape of the DeepCSV discriminant distribution in simulation is corrected by scale factors to better describe the data. This correction is derived separately for light-flavour and b quark jets with a tag-and-probe approach. Control samples enriched in events with a Z boson and exactly two jets where a b quark jet veto is applied are used to obtain the correction for light-flavour jets. The correction for b quark jets is estimated using a sample enriched in $\mathrm{t\bar{t}}$ events with no additional jets [39].

4 Event selection

The event selection targets events from the production of a Higgs boson in association with $\mathrm{t\bar{t}}$ events, where the Higgs boson decays into $\mathrm{b\bar{b}}$. All three $\mathrm{t\bar{t}}$ decay channels are considered: fully-hadronic, semi-leptonic, and dilepton decays. These signatures imply the presence of zero, one, or two isolated charged leptons ($\ell = \mathrm{e}, \mu$), missing transverse momentum due to the neutrinos from W boson decays, and jets with typical transverse momenta of several tens of GeV or more originating from the final-state quarks, several of which originate from b quarks.

During data taking, events in the fully-hadronic channel are selected by any of three dedicated multijet triggers, requiring either at least six or at least four jets. The triggers are complementary in that they trade off between kinematic conditions and b tagging requirements. Their selection requirements are: at least six jets with $p_\mathrm{T} > 40\,\mathrm{GeV}$ and $|\eta| < 2.6$, $H_\mathrm{T} > 430\,\mathrm{GeV}$, and at least one b-tagged jet; at least six jets with $p_\mathrm{T} > 32\,\mathrm{GeV}$ and $|\eta| < 2.6$, $H_\mathrm{T} > 380\,\mathrm{GeV}$, and at least two b-tagged jets; and at least four jets with $p_\mathrm{T} > 75, 60, 45, 40\,\mathrm{GeV}$ and $|\eta| < 2.6$, $H_\mathrm{T} > 300\,\mathrm{GeV}$ and at least three b-tagged jets. Events in the single-lepton channel were selected either by single-lepton triggers or by a single-electron trigger with additional H_T requirements. The single-lepton triggers require the presence of one muon with $p_\mathrm{T} > 24\,\mathrm{GeV}$ and $|\eta| < 2.1$ or with $p_\mathrm{T} > 27\,\mathrm{GeV}$, depending on the data-taking period, or of one electron with $p_\mathrm{T} > 35\,\mathrm{GeV}$ or one electron with $p_\mathrm{T} > 28\,\mathrm{GeV}$, $|\eta| < 2.1$, and $H_\mathrm{T} > 150\,\mathrm{GeV}$. Events in the dilepton channel were selected either by the single-lepton trigger (retaining events with an additional lepton) or by dilepton triggers that require the presence of two electrons or muons. The same-flavour dilepton triggers required two electrons with $p_\mathrm{T} > 23\,\mathrm{GeV}$ and 12 GeV, or two muons with $p_\mathrm{T} > 17\,\mathrm{GeV}$ and 8 GeV, respectively. The different-flavour dilepton triggers required either a muon with $p_\mathrm{T} > 23\,\mathrm{GeV}$ and an electron with $p_\mathrm{T} > 12\,\mathrm{GeV}$, or an electron with $p_\mathrm{T} > 23\,\mathrm{GeV}$ and a muon with $p_\mathrm{T} > 12\,\mathrm{GeV}$ or 8 GeV, depending on the data-taking period.

Offline events in the fully-hadronic channel are selected if they contain no lepton (e or μ) with $p_\mathrm{T} > 15\,\mathrm{GeV}$. Furthermore, events are required to pass a preselection which requires at least six jets with $p_\mathrm{T} > 40\,\mathrm{GeV}$, $H_\mathrm{T} > 500\,\mathrm{GeV}$, and at least 2 b-tagged jets. All other jets are required to have $p_\mathrm{T} > 30\,\mathrm{GeV}$ and $|\eta| < 2.4$.

In the single-lepton channel, events are selected if they contain exactly one lepton (e or μ) with $p_\mathrm{T} > 30(29)\,\mathrm{GeV}$ in the case of the electron (muon). The flavour of the lepton must match the flavour of the trigger that accepted the event (for example if an electron is identified, the single-electron trigger must have accepted the event). The single-lepton events are required to contain at least four jets with $p_\mathrm{T} > 30\,\mathrm{GeV}$, at least two of which are b tagged.

In the dilepton channel, events are selected if they contain exactly two, oppositely charged leptons ($\mathrm{e^+e^-}$, $\mu^\pm\mathrm{e}^\mp$, $\mu^+\mu^-$). The lepton with the highest p_T out of the pair is required to have $p_\mathrm{T} > 25\,\mathrm{GeV}$, and the other lepton $p_\mathrm{T} > 15\,\mathrm{GeV}$. Events must get accepted by a trigger based on

online leptons with a flavour content consistent with that of the two leptons selected offline (for example, $\mu^+\mu^-$ events are required to pass a dimuon or single-muon trigger). 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 the same-flavour channels, events are also rejected if $76\,\text{GeV} < m_{\ell\ell} < 106\,\text{GeV}$, thereby suppressing further contribution from Z+jets events. The dilepton events are required to contain at least two jets, where the leading two jets must have $p_\text{T} > 30\,\text{GeV}$ and all further jets $p_\text{T} > 20\,\text{GeV}$. At least one jet must be b tagged.

In addition, events are required to fulfil $p_\text{T}^\text{miss} > 20\,\text{GeV}$ in the single-lepton and $p_\text{T}^\text{miss} > 40\,\text{GeV}$ in the dilepton same-flavour channels to suppress background contributions, for example from QCD multijet production, and to account for the neutrinos from the W boson decay.

The criteria of this selection, referred to as baseline event selection in the following, are summarised in Table 1.

Table 1: Baseline event selection criteria in the fully-hadronic (FH), single-lepton (SL), and dilepton (DL) channels.

	FH channel	SL channel	DL channel
Number of leptons	0	1	2
p_T of leptons (e/μ) [GeV]	—	$> 30/29$	$> 25/25\,\text{GeV}$
p_T of additional leptons [GeV]	< 15	< 15	< 15
$\lvert\eta\rvert$ of leptons	< 2.4	< 2.4	< 2.4
Number of jets	≥ 6	≥ 4	≥ 2
p_T of jets [GeV]	> 40	> 30	$> 30, 30, 20$
$\lvert\eta\rvert$ of jets	< 2.4	< 2.4	< 2.4
Number of b-tagged jets	≥ 2	≥ 2	≥ 1
p_T^miss	—	$> 20\,\text{GeV}$	$> 40\,\text{GeV}$

Event yields observed in data and predicted by the simulation after the baseline selection are listed in Table 2. The $\text{t}\bar{\text{t}}\text{H}$ signal includes $\text{H} \to \text{b}\bar{\text{b}}$ and all other Higgs boson decay modes. Background contributions from QCD multijet production, estimated using a low-p_T^miss control region in data, have been found to be negligible in the single-lepton and dilepton channels.

5 Analysis strategy and classification

The analysis strategy has been optimised separately in each channel, based on the expected sensitivity to a SM $\text{t}\bar{\text{t}}\text{H}$ signal evaluated with simulated data.

In each analysis channel, the selected events are further divided into categories with varying signal purity and different background composition, as described below. In each category, multivariate discriminants are optimised to separate signal from background. The signal is extracted in a simultaneous template fit of the expected background and signal discriminant output to the data across all the categories and channels, correlating corresponding processes. The systematic uncertainties discussed in Section 6 are taken into account using nuisance parameters and are correlated among the processes, categories, and channels where appropriate. The signal template includes contributions from all SM Higgs-boson decays to take into account contamination by other decay channels.

Table 2: Event yields observed in data and predicted by the simulation after the baseline selection in the fully-hadronic (FH), single-lepton (SL), and dilepton (DL) channels prior to the fit to data. Here, the QCD prediction is taken from simulation. The quoted uncertainties correspond to the total statistical and systematic uncertainties (excluding the 50% uncertainties on the normalisation of the $t\bar{t}$+hf processes).

Process	FH channel	SL channel	DL channel
QCD	2938305 ± 301286	—	—
$t\bar{t}$+lf	357488 ± 52694	718341 ± 83944	275407 ± 19610
$t\bar{t}$+c$\bar{c}$	93674 ± 11860	96581 ± 13795	24721 ± 3145
$t\bar{t}$+b	23737 ± 2892	27222 ± 3749	5613 ± 824
$t\bar{t}$+2b	14039 ± 2183	10537 ± 2206	1697 ± 351
$t\bar{t}$+b$\bar{b}$	19730 ± 2413	12770 ± 2050	1813 ± 262
Single t	24117 ± 1847	38170 ± 3720	14044 ± 1133
V+jets	31154 ± 2319	14491 ± 1754	2199 ± 264
$t\bar{t}$+V	2924 ± 228	2963 ± 286	1028 ± 99
Diboson	354 ± 40	503 ± 61	420 ± 55
Total bkg.	3505523 ± 339615	921576 ± 97714	326942 ± 23458
$t\bar{t}$H	2556 ± 164	1747 ± 167	363 ± 22
Data	3508079	923936	331055

5.1 Fully-hadronic channel

In the fully-hadronic channel, events are categorised according to their reconstructed jet and b-tagged jet multiplicities. Six categories are formed: seven reconstructed jets, out of which three are b-tagged jets (7 jets, 3 b-tags); seven jets with four or more b-tagged jets (7 jets, $\geq$ 4 b-tags); eight jets with three b-tagged jets (8 jets, 3 b-tags); eight jets with four or more b-tagged jets (8 jets, $\geq$ 4 b-tags); nine or more jets with three b-tagged jets ($\geq$ 9 jets, 3 b-tags); and nine or more jets with four or more b-tagged jets ($\geq$ 9 jets, $\geq$ 4 b-tags). Events with seven or more jets and two b-tagged jets are used to form control regions for the multijet background estimation that is described further below.

To reject events that are unlikely to include any W bosons from top quark decays, a requirement is placed on the invariant mass of a dijet pair, which is selected as follows: all jets that are not b tagged are considered, and the pair with invariant mass m_{qq} closest to the W boson mass is chosen as the W boson candidate in the event. The invariant mass is required to be $60\,\text{GeV} < m_{qq} < 100\,\text{GeV}$ in the 7- or 8-jet categories and $70\,\text{GeV} < m_{qq} < 92\,\text{GeV}$ in the $\geq$ 9-jet categories. The W mass distribution is broader in the 7- and 8-jet categories as not all events contain a correct W daughter jet-pair, either due to misidentification or detector acceptance.

In order to reduce the overwhelming QCD multijet background, an initial rejection is applied based on the angular information of the jets in the event. In general, the jets in QCD-multijet events are expected to be separated further in η than the jets in $t\bar{t}$H events. The variable $\Delta\eta_{\text{jets}}$ is defined by taking the average $\Delta\eta$ between each jet and the jet furthest apart from it in η. A simple selection on $\Delta\eta_{\text{jets}}$ is made in each category, such that the signal efficiency is 80%. The multijet background rejection efficiency is then between approximately 40 and 55% depending on the category. The $\Delta\eta_{\text{jets}}$ distribution in data and background expectation in the (8 jets, $\geq$ 4 b-tags) category is shown in Fig. 1 as an example; events are required to pass $\Delta\eta_{\text{jets}} \leq 2.52$ in this case.

Jets are further classified according to a quark-gluon likelihood (QGL) [41, 42] discriminant,

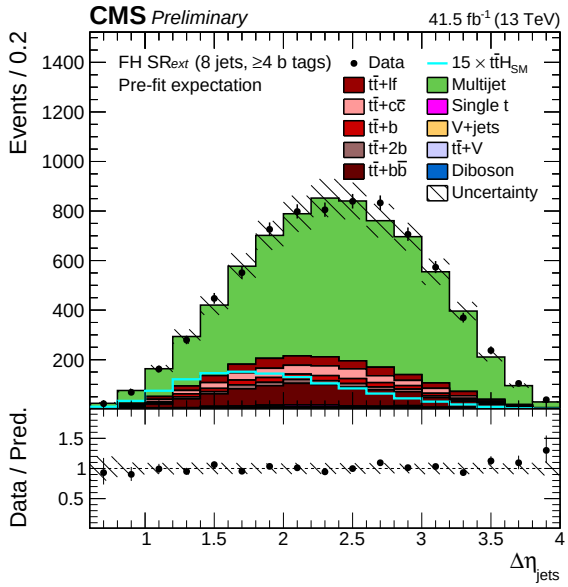

Figure 1: Distribution of the $\Delta\eta_{\text{jets}}$ for events with 8 jets and $\geq$ 4 b-tags in an extended signal region (SR*ext*), which corresponds to the regular SR but excluding the requirement of $\Delta\eta_{\text{jets}} \leq 2.52$ for this category. The expected background contributions (filled histograms) are stacked, and the expected signal distribution (line) is superimposed. Each contribution is normalised to an integrated luminosity of 41.5 fb^{-1}, and the signal distribution is additionally scaled by a factor of 15 for better visibility. The hatched uncertainty bands correspond to the total statistical and systematic uncertainties. The distributions observed in data (markers) are overlayed. The lower plot shows the ratio of the data to the background prediction.

which is used to separate events containing light-flavour jets from $q\bar{q}'$ decays of W bosons from events containing gluon jets produced in multijet interactions or via initial-state radiation. The QGL discriminant is applied only to non-b-tagged jets. To correct the modelling of the QGL distribution in simulation, a reweighting based on a control sample of $\mu^+\mu^-$+jet and dijet events is applied. A likelihood ratio (QGLR) is defined based on the QGL of N reconstructed light-flavour jets that arise from N quarks and the QGL of N reconstructed light-flavour jets that arise from N gluons. After excluding either the first three (in the categories with 3 b-tags) or four (in the categories with $\geq$ 4 b-tags) b-tagged jets in an event (ranked by the DeepCSV output value), up to N light-flavour jets with $N = 3$, 4, or 5 are used to calculate QGLR. The distributions of the QGLR after the baseline selection are shown in Fig. 2. The signal region (SR) is defined as QGLR > 0.5.

The main residual background to the fully-hadronic channel stems from multijet production, which is estimated from control regions in data, as described below. The approach uses a control region (CR) with low b tag multiplicity to estimate the contribution from multijet events in the SR. The CR is enriched in multijet events, and the remaining contribution from other backgrounds (mainly $t\bar{t}$+jets) is subtracted using simulation.

The CR is defined by events with exactly two b-tagged jets and one or more additional jets that are not b tagged but fulfil a requirement on the DeepCSV that corresponds to a 84% b tagging efficiency at a 11% light-flavour mistagging rate (referred to as "loose" b tag) [39]. We define a validation region (VR) using events with QGLR < 0.5. This definition provides four mutually exclusive regions, summarised in Table 3, two of which are used to obtain the multijet background estimate, while two are used to validate the method. The use of the Validation CR and the VR for validation is justified since the QGLR and the number of additional loose b-

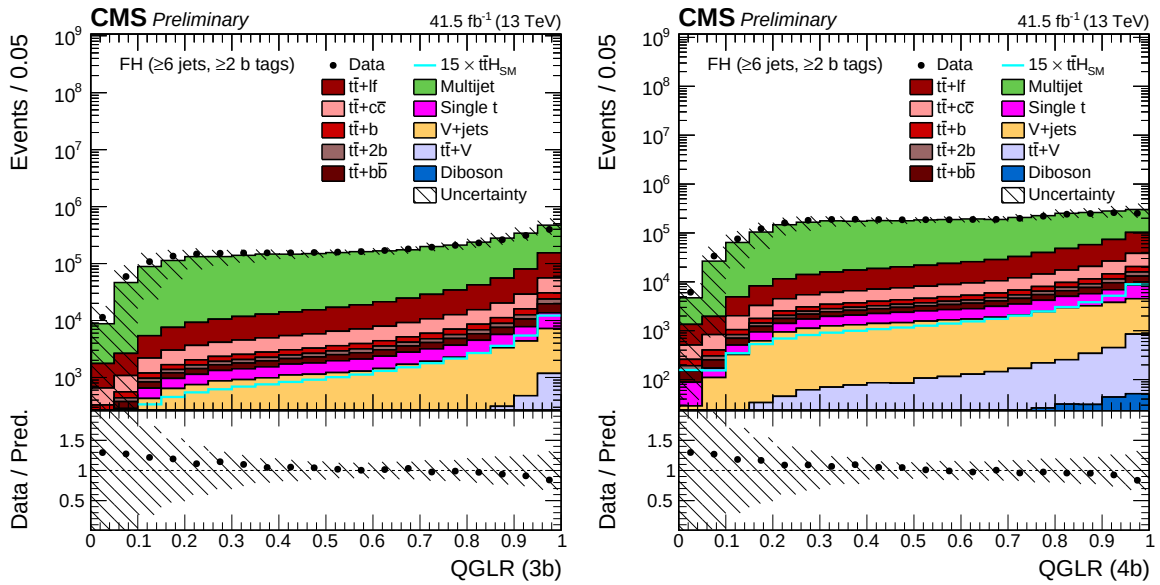

Figure 2: Distributions of the QGLR after excluding the first three (left) and first four (right) b-tagged jets (ranked by the DeepCSV output value) for the calculation in the fully-hadronic channel after the baseline selection. The expected background contributions (filled histograms) are stacked, and the expected signal distribution (line) is superimposed. Each contribution is normalised to an integrated luminosity of 41.5 fb^{-1}, and the signal distribution is additionally scaled by a factor of 15 for better visibility. The hatched uncertainty bands correspond to the total statistical and systematic uncertainties (excluding the 50% uncertainties on the normalisation of the $\text{t}\bar{\text{t}}$+hf processes) added in quadrature. The distributions observed in data (markers) are overlayed. The last bin includes overflow events. The lower plots show the ratio of the data to the background prediction.

tagged jets are uncorrelated by construction, since only the non-loose b-tagged jets are used in the calculation of the QGLR, except for the rare case of events with five or more loose b-tagged jets. However, there is a slight indirect correlation originating from different probabilties of quark- and gluon-like jets being promoted to loose b-tagged jets, which does not effect the validity of the method.

Table 3: Definition and description of the four mutually exclusive regions in the analysis.

	$N_{\text{b tag}} = 2$ $N_{\text{b tag loose}} \geq 3$	$N_{\text{b tag}} \geq 3$
QGLR > 0.5	CR (to extract distribution)	SR (final analysis)
QGLR < 0.5	Validation CR (to validate distribution)	VR (comparison with data)

The four mutually exclusive regions are used independently in each of the six analysis categories defined above. For a given variable, the distribution of the multijet events in the SR of each category is estimated from the data in the CR, after subtracting $\text{t}\bar{\text{t}}$+jets and other minor backgrounds. Since the kinematic properties of b-tagged and untagged jets differ in the CR and SR because of a different heavy-flavour composition, corrections (TF_{loose}) as a function of jet p_{T}, η, and the minimum distance between the jet and the first two b-tagged jets are applied to the loose b-tagged jets in the CR. The TF_{loose} correction is obtained from jets in events passing the baseline selection, excluding the first two jets, ordered according to their DeepCSV discrimi-

nant values. This is used to reweight the kinematic distributions of loose b-tagged jets to match those of b-tagged jets. The corrected multijet distribution in the CR is scaled to provide an estimate of the distribution in the SR. Specifically, the multijet yield in each category is left floating in the final fit discussed in Section 7.

A likelihood technique based on the LO matrix elements for the $t\bar{t}H$ signal and the $t\bar{t}+b\bar{b}$ background processes is used to extract the signal. This method utilises the full kinematic properties of each event to provide a MEM discriminant between the signal and background as documented in Ref. [6]. Although the discriminant is constructed to discriminate against the $t\bar{t}+b\bar{b}$ background, it performs well against $t\bar{t}$+lf jets and against multijet events, and the output distribution is used as final discriminant in each category.

5.2 Single-lepton channel

In the single-lepton channel, events are separated based on the jet multiplicity into three categories with $(4\,\text{jets}, \geq 3\,\text{b-tags})$, $(5\,\text{jets}, \geq 3\,\text{b-tags})$, and $(\geq 6\,\text{jets}, \geq 3\,\text{b-tags})$. Dedicated ANNs are trained in each category to perform a multi-classification of an event as either signal or any of the five $t\bar{t}$+jets background processes $t\bar{t}+b\bar{b}$, $t\bar{t}$+2b, $t\bar{t}$+b, $t\bar{t}+c\bar{c}$, or $t\bar{t}$+lf. The result of the ANN classification is used both to further categorise the events according to the most probable process as well as to construct the discriminant distribution that enters the final fit.

The ANNs utilise input variables related to kinematic properties of individual objects, event shape, and the jet b tagging discriminant, and additionally the MEM discriminant output. The input variables and their correlations have been verified to model the data well using a goodness-of-fit test. For each variable and each pair of variables, the one and two dimensional distributions, respectively, have been fit to the data in the three analysis categories using the full uncertainty model. The goodness-of-fit is quantified by a p value constructed from the post-fit uncertainty model. Only those variables for which the p value is greater than 5% in all of the tested combinations are considered in the further analysis. The distributions of several representative input variables are shown in Fig. 3 prior to the fit to data.

The ANNs are implemented in Keras [43] as feedforward neural networks with three hidden layers of 100 nodes each. They are trained using simulated $t\bar{t}H$ and $t\bar{t}$+jets events, amounting to at least 2100 training events for each of the six processes in each jet-multiplicity category. During the training, events are weighted such that the effective number of events is the same per process. Three independent sets of events are used to evaluate the performance over the validation set and optimise the parameters of the ANNs and to construct the discriminant distributions entering the final fit. The training is done over a maximum of 500 epochs, i.e. full passes over the training data, and it is terminated earlier if the performance does not improve significantly, which is the case in all trainings. The cost function that is minimised during the training is the cross entropy. To minimise overtraining, both L2 regularisation and dropout are used.

The values obtained in the six output nodes of the ANN are normalised to unity using a "softmax" function [44], and, as a result, can be interpreted as probabilities describing the likelihood of the event being a $t\bar{t}H$ signal or one of the five $t\bar{t}$+jets background processes. Events are then further divided into subcategories of the most probable process according to this ANN classification, leading to 18 jet-process categories in total. In each of the jet-process categories, the ANN classifier output distribution of the node that matches the process category is used as the final discriminant.

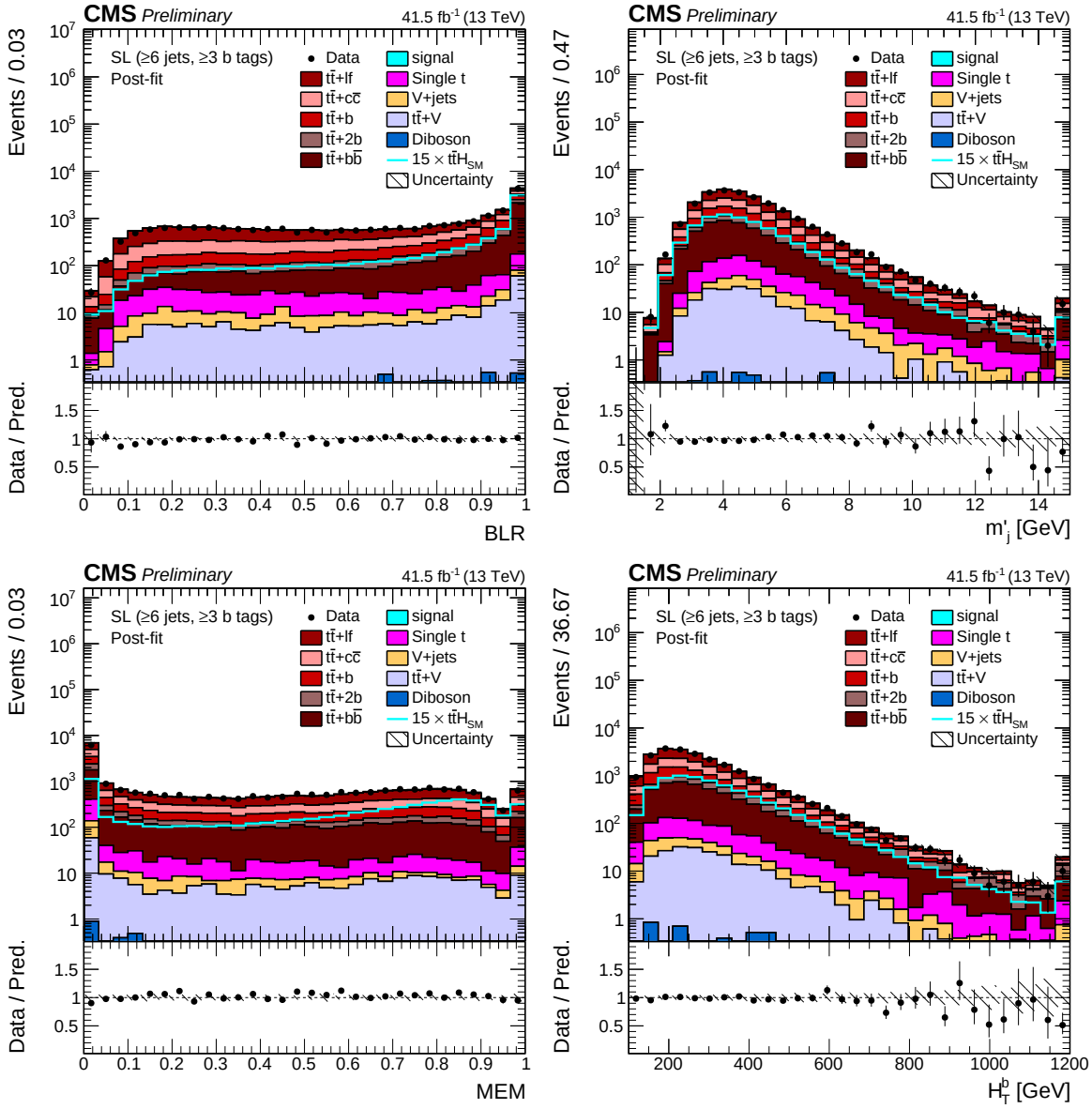

Figure 3: Distributions of representative variables used as input to the ANN in the (≥ 6 jets, ≥ 3 b-tags) category of the single-lepton (SL) channel: likelihood ratio discriminating between events with 4 b quark jets and 2 b quark jets (BLR), sum of the masses of all jets normalised to the number of dijet pairs in the event (m_j'), MEM discriminant (MEM), and scalar sum of p_T of b-tagged jets (H_T^b). The background and signal contributions (filled histograms) are stacked, and the hatched uncertainty bands correspond to the total statistical and systematic uncertainties. Shown are the post-fit contributions, where the model parameters are obtained from the final fit of the discriminant distributions to data, described in Section 7, and applied to the shown input variable distributions. The distributions observed in data (markers) are overlayed. In addition, the SM $t\bar{t}H$ signal expectation (line) is overlayed (scaled by a factor 15 for better visibility). The first and the last bins include underflow and overflow events, respectively. The lower plots show the ratio of the data to the background prediction.

5.3 Dilepton channel

Dilepton events are separated into five exclusive categories based on the number of jets and b-tagged jets. The five categories correspond to events with $(3\,\text{jets}, 2\,\text{b-tags})$, $(3\,\text{jets}, 3\,\text{b-tags})$, $(\geq 4\,\text{jets}, 2\,\text{b-tags})$, $(\geq 4\,\text{jets}, 3\,\text{b-tags})$, $(\geq 4\,\text{jets}, \geq 4\,\text{b-tags})$.

Boosted Decision Trees (BDT) are used to construct, for each category, the final discriminant optimised to separate signal and background events. The BDT training is based on the *gradient boosting* algorithm, as implemented in the TMVA package [45].

The background sample used to construct the BDT corresponds to half of the available simulated $\mathrm{t\bar{t}}$ events with both top quarks decaying leptonically, while the other half of this MC sample is used solely to analyse the data. A dedicated MC sample of $\mathrm{t\bar{t}H}$ ($\mathrm{H} \to \mathrm{b\bar{b}}$) events in which both top quarks are forced to decay leptonically provides the signal events for the BDT training. For both signal and background, half of these events, corresponding to at least 1750 events per process and category, are used to train the BDT discriminant, while the other half is used to ensure that the BDTs are not affected by overtraining (test sample).

Similarly to the single-lepton channel, the BDT inputs correspond to quantities associated to individual physics objects, for example kinematic properties and jet b tagging discriminant, and event-shape variables sensitive to correlations between different reconstructed objects. For each category, only the quantities that are found to be well modelled in the data are considered and, amongst these, the twelve with the highest discrimination power are used as inputs to the BDT. The distributions of several representative input variables are shown in Fig. 4 prior to the fit to data.

The BDT hyperparameters are optimised with a procedure based on the *particle swarm* algorithm [46, 47], maximising the integral, or area-under-the-curve (AUC), of the receiver-operator-characteristic (ROC) curve associated to the BDT discriminant, which is used to approximate the final sensitivity of the discriminant. The selection of the input variables and optimisation of the BDT hyperparameters is performed independently for each of the five categories of the dilepton analysis.

Differently from the previous version of the dilepton analysis [5], the MEM discriminant is now used as an input to the BDT in events with $(\geq 4\,\text{jets}, 3\,\text{b-tags})$ and $(\geq 4\,\text{jets}, \geq 4\,\text{b-tags})$; this leads to an improvement of the expected sensitivity in the dilepton channel of approximately 10%, compared to the 2D fit approach followed in the 2016 analysis.

6 Systematic uncertainties

Several sources of systematic uncertainties are considered in the analysis. The uncertainties are taken into account via nuisance parameters in the final fit procedure described in Section 7 and affect either the rate or the discriminant shape of the signal or background processes or both. The effects from the same source are treated as fully correlated among the different categories.

The uncertainty in the integrated luminosity estimate is 2.3% [48]. The trigger efficiency uncertainty in the fully-hadronic channel is determined from the bin-by-bin uncertainties in the ratio of efficiency in data relative to simulation, and are 2.0% on average, with some being as large as 20%. The efficiencies and uncertainties of the single-electron and dilepton triggers are measured in data using reference triggers based on single-muon and $p_{\mathrm{T}}^{\mathrm{miss}}$ requirements, respectively, that are uncorrelated with those used in the analysis. The electron and muon identification efficiency uncertainties are estimated by comparing variations in measured ef-

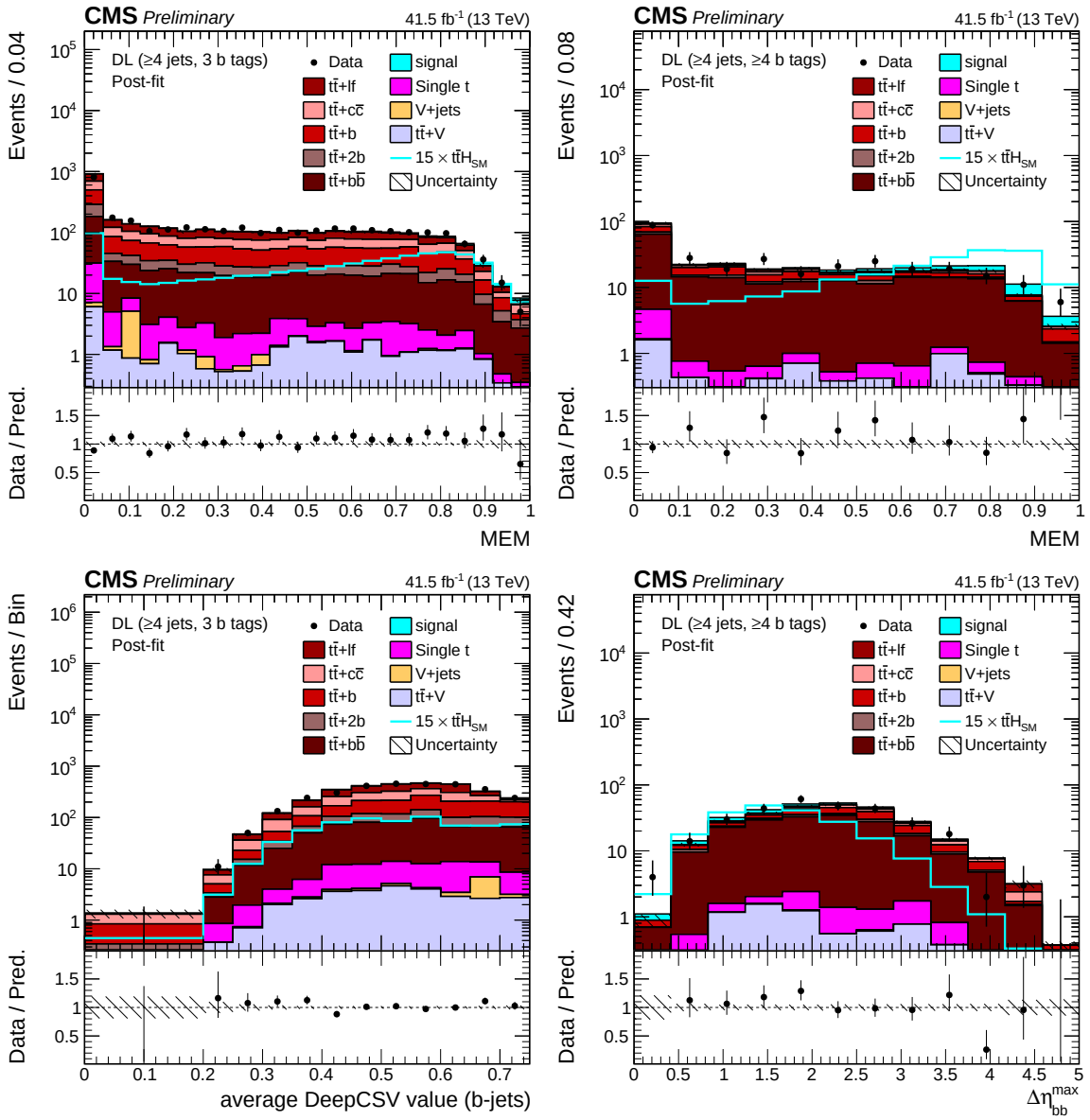

Figure 4: Distributions of representative variables used as input to the BDT in the $(\geq 4\,\text{jets}, 3\,\text{b-tags})$ (left) and $(\geq 4\,\text{jets}, \geq 4\,\text{b-tags})$ (right) categories of the dilepton channel: MEM discriminant (MEM), average b-tagging discriminant value of all b-tagged jets normalised to the total number of jets (average DeepCSV value (b-jets)), and maximum $\Delta\eta$ between any two b-tagged jets ($\Delta\eta_{\text{bb}}^{\text{max}}$). The background and signal contributions (filled histograms) are stacked, and the hatched uncertainty bands correspond to the total statistical and systematic uncertainties. Shown are the post-fit contributions, where the model parameters are obtained from the final fit of the discriminant distributions to data, described in Section 7, and applied to the shown input variable distributions. The distributions observed in data (markers) are overlayed. In addition, the SM $\text{t}\bar{\text{t}}\text{H}$ signal expectation (line) is overlayed (scaled by a factor 15 for better visibility). The first and the last bins include underflow and overflow events, respectively. The lower plots show the ratio of the data to the background prediction.

ficiency between data and simulation using a high-purity sample of Z boson decays. These uncertainties are found to be small, typically below 1–2%. The uncertainty of the 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 [49].

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 19 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 [38].

The DeepCSV b tagging scale factors receive uncertainties due to the contamination of background processes in the control samples, the jet energy scale uncertainty—which is correlated with the overall jet energy scale uncertainty—and the statistical uncertainty in the scale factor evaluation [39]. 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 systematic uncertainty from reweighting the simulated QGL distribution affects both the rate and the distribution in the MEM discriminant, and is taken into account by changing the function used in the reweighting procedure by the full amount. Many uncertainties that are related to the MC simulation of the multijet background are avoided by estimating this contribution from data. Nevertheless, a few small systematic uncertainties remain: an uncertainty 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; a reweighting based on the H_{T} distribution (considering the p_{T} of just the first six jets) in the categories with 3 and $\geq$4 b-tagged jets of the VR, to account for mismodelling at low H_{T}; and the total normalisation in each category that is left unconstrained in the final fit.

The theoretical uncertainties of the cross sections used to predict the rates of various processes, which arise primarily from PDFs and the factorisation and renormalisation scale choices, are propagated to the yield estimates. The cross section uncertainties are each separated into their PDF and scale components (renorm./fact. scales) and are correlated where appropriate between processes. The $\mathrm{t\bar{t}+b\bar{b}}$ process, and to a lesser extent the $\mathrm{t\bar{t}+2b}$, $\mathrm{t\bar{t}+b}$, and $\mathrm{t\bar{t}+c\bar{c}}$ production, represent important sources of irreducible background. Following the model established in previous analyses [5, 6], an additional 50% rate uncertainty is assigned separately to each of the $\mathrm{t\bar{t}+hf}$ processes. This approach provides sufficient flexibility to cover potential differences between the prediction and the data that are expected for example from comparing simulations based on the 4 and the 5 flavour scheme.

Uncertainties arising from missing higher-order terms in the POWHEG $\mathrm{t\bar{t}}$+jets simulation at ME level are evaluated by a variation of the renormalisation and factorisation scales by factors of two up and down with respect to the nominal values. The corresponding uncertainty of 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 [50]. These variations are applied using event weights obtained directly from the generators. Uncertainties related to the matching scheme and the underlying event (UE) tune are evaluated by comparing the reference $t\bar{t}$+jets simulation with samples with varied *hdamp* parameter [51] and varied UE 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. For this reason, the uncertainties were estimated conservatively as the changes in the rates of the different $t\bar{t}$ subprocesses independently for each category. The derived rate uncertainties are treated as uncorrelated among the $t\bar{t}$ subprocesses. The uncertainty arising from the PDF set is determined from the PDF variations provided with the NNPDF set [52].

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 [53, 54].

The sources of uncertainties are listed in Table 4, and their impact on the final result is discussed in Section 7.

7 Results

The signal strength modifier $\mu = \sigma/\sigma_{\mathrm{SM}}$ of the $t\bar{t}H$ production cross-section is determined in a simultaneous binned profile likelihood fit to the data across all analysis categories. The systematic uncertainties described in Section 6 are taken into account via nuisance parameters, which allow for variations in the shape and normalisation of the final discriminant distributions during the fit. The discriminant distributions in the most sensitive categories in each channel are depicted in Fig. 5 before and after the fit.

The best fit values of the nuisance parameters are within 1 standard deviation of the prior uncertainty for more than 98% of the total number of nuisance parameters. As expected, the fit constrains the nuisance parameters related to the conservatively assigned 50% prior uncertainties on the $t\bar{t}$+hf cross section to 30% of the prior value. A few other nuisance parameters that are related to jet energy scale and b tagging uncertainties are constrained up to a factor of approximately 40%. These constraints are 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. The impact of the uncertainty source represented by 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 post-fit uncertainty. The best fit values and impacts of the 20 parameters ranked highest in impact for the fit to 2017 data are presented in Fig. 6. The nuisance parameters with the highest impact are related to the uncertainty in the QCD-multijet background prediction and to the uncertainty in the $t\bar{t}$+hf and signal cross sections as well as in the b tagging scale factors.

The obtained best fit value of the signal-strength modifier μ is $\hat{\mu} = 1.49\,^{+0.21}_{-0.20}\,(\mathrm{stat})\,^{+0.39}_{-0.35}\,(\mathrm{syst})$ with a total uncertainty of $+0.44/-0.40$ (at an expected uncertainty of $+0.38/-0.36$). This corresponds to an observed (expected) significance of 3.7 (2.6) standard deviations above the background-only hypothesis. The best fit values in each analysis channel separately and in the channel combination are listed in Table 5 and are shown in Fig. 7 (left). The goodness-of-fit is quantified using a p value that takes into account the post-fit uncertainty model and amounts to $p = 0.98$.

Table 4: Systematic uncertainties considered in the analysis.

Source	Type	Remarks
Integrated luminosity	rate	Signal and all backgrounds
Lepton identification/isolation	shape	Signal and all backgrounds
Trigger efficiency	shape	Signal and all backgrounds
Trigger prefiring correction	rate	Signal and all backgrounds
Pileup	shape	Signal and all backgrounds
Jet energy scale	shape	Signal and all backgrounds
Jet energy resolution	shape	Signal and all backgrounds
b tag hf fraction	shape	Signal and all backgrounds
b tag hf stats (linear)	shape	Signal and all backgrounds
b tag hf stats (quadratic)	shape	Signal and all backgrounds
b tag lf fraction	shape	Signal and all backgrounds
b tag lf stats (linear)	shape	Signal and all backgrounds
b tag lf stats (quadratic)	shape	Signal and all backgrounds
b tag charm (linear)	shape	Signal and all backgrounds
b tag charm (quadratic)	shape	Signal and all backgrounds
QGL reweighting	shape	Signal and all backgrounds
$\mathrm{TF}_{\mathrm{loose}}$ correction	shape	QCD multijet estimate
H_{T} reweighting	shape	QCD multijet estimate
Multijet normalisation	rate	QCD multijet estimate
Renorm./fact. scales ($\mathrm{t\bar{t}H}$)	rate	Scale uncertainty of NLO $\mathrm{t\bar{t}H}$ prediction
Renorm./fact. scales ($\mathrm{t\bar{t}}$)	rate	Scale uncertainty of NNLO $\mathrm{t\bar{t}}$ prediction
$\mathrm{t\bar{t}}$+hf cross sections	rate	Additional 50% rate uncertainty of $\mathrm{t\bar{t}}$+hf predictions
Renorm./fact. scales (t)	rate	Scale uncertainty of NLO single t prediction
Renorm./fact. scales (V)	rate	Scale uncertainty of NNLO W and Z prediction
Renorm./fact. scales (VV)	rate	Scale uncertainty of NLO diboson prediction
PDF (gg)	rate	PDF uncertainty for gg initiated processes except $\mathrm{t\bar{t}H}$
PDF (gg $\mathrm{t\bar{t}H}$)	rate	PDF uncertainty for $\mathrm{t\bar{t}H}$
PDF ($\mathrm{q\bar{q}}$)	rate	PDF uncertainty of $\mathrm{q\bar{q}}$ initiated processes ($\mathrm{t\bar{t}}$+W,W,Z)
PDF (qg)	rate	PDF uncertainty of qg initiated processes (single t)
PDF shape variations ($\mathrm{t\bar{t}H}$, $\mathrm{t\bar{t}}$)	shape	Based on the NNPDF variations, same for $\mathrm{t\bar{t}H}$ and additional jet flavours
μ_{R} scale ($\mathrm{t\bar{t}}$)	shape	Renormalisation scale uncertainty of the $\mathrm{t\bar{t}}$ ME generator (POWHEG), same for additional jet flavours
μ_{F} scale ($\mathrm{t\bar{t}}$)	shape	Factorisation scale uncertainty of the $\mathrm{t\bar{t}}$ ME generator (POWHEG), same for additional jet flavours
PS scale: ISR ($\mathrm{t\bar{t}}$)	shape	Initial state radiation uncertainty of the PS (for $\mathrm{t\bar{t}}$ events), independent for additional jet flavours
PS scale: FSR ($\mathrm{t\bar{t}}$)	shape	Final state radiation uncertainty of the PS (for $\mathrm{t\bar{t}}$ events), independent for additional jet flavours
ME-PS matching ($\mathrm{t\bar{t}}$)	rate	NLO ME to PS matching, *hdamp* [51] (for $\mathrm{t\bar{t}}$ events), independent for additional jet flavours
Underlying event ($\mathrm{t\bar{t}}$)	rate	Underlying event (for $\mathrm{t\bar{t}}$ events), independent for additional jet flavours
Bin-by-bin event count	shape	Statistical uncertainty of the signal and background prediction due to the limited sample size

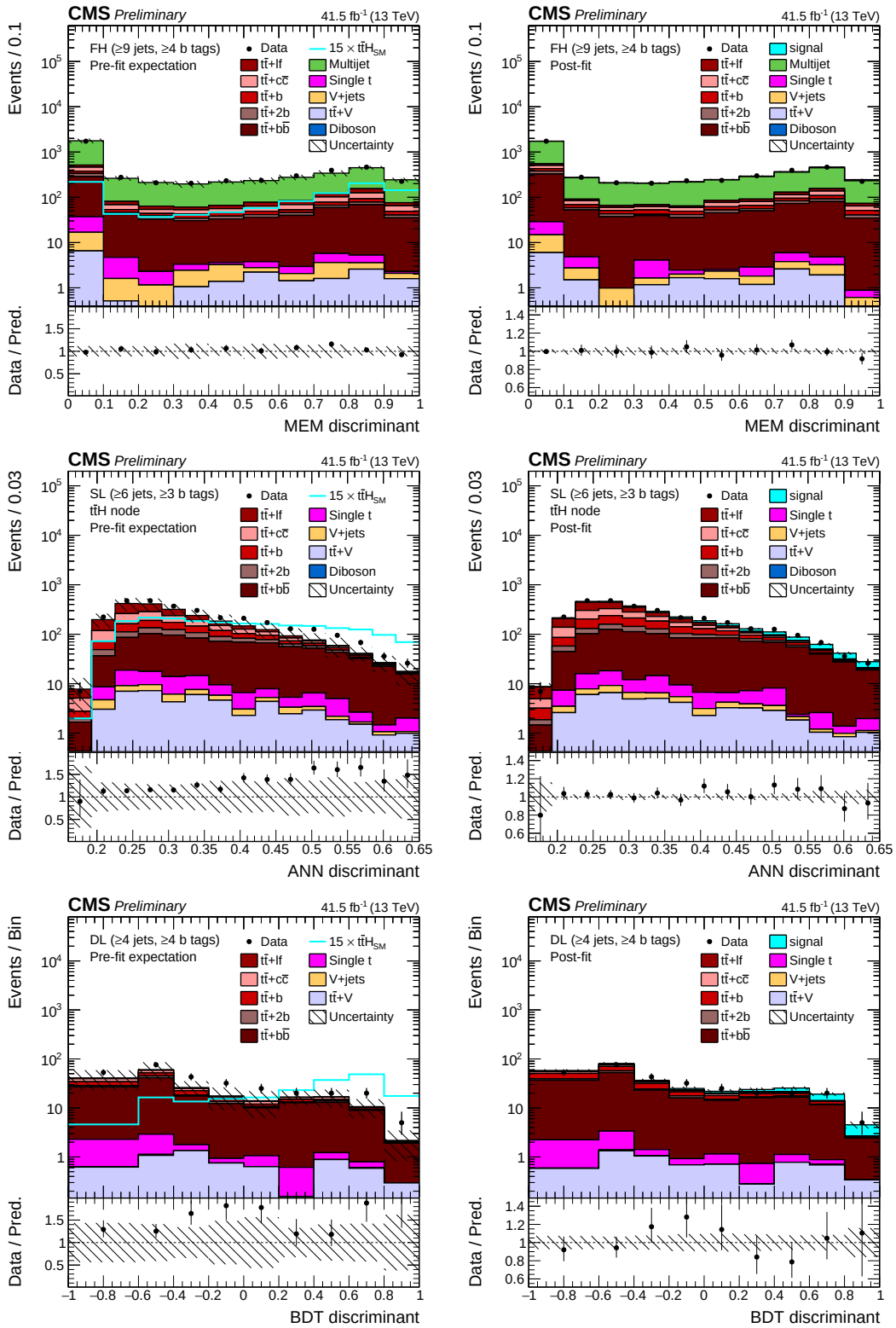

Figure 5: Final discriminant shapes in the categories with the highest sensitivity in fully-hadronic (top), semi-leptonic (middle), and dilepton (bottom) channels before (left) and after (right) the fit to data. The expected background contributions (filled histograms) are stacked. In the pre-fit case, the expected signal contribution (line), scaled by a factor 15, is superimposed. In the post-fit case, the fitted signal contribution is also stacked. The hatched uncertainty bands include the total uncertainty of the fit model. The distributions observed in data (markers) are overlayed. The first and the last bins include underflow and overflow events, respectively. The lower plots show the ratio of the data to the background (pre-fit) and signal+background (post-fit) prediction.

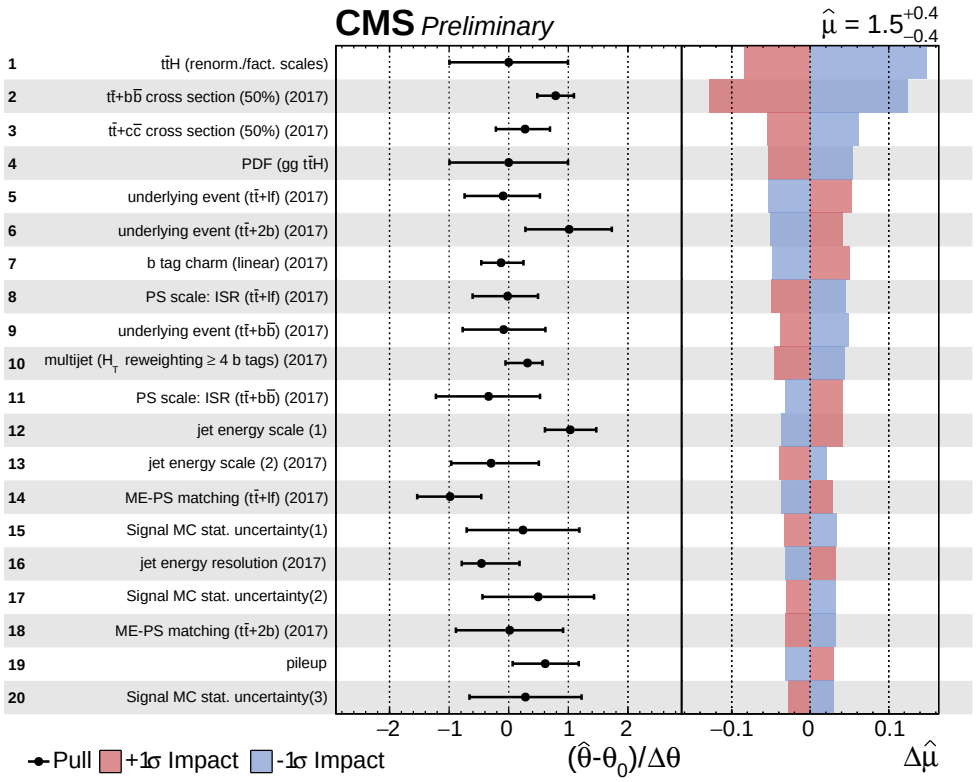

Figure 6: Post-fit pull of the nuisance parameters included in the fit to the 2017 data as well as their impact on the signal strength μ, ordered by their impact. Only the 20 highest ranked parameters are shown. The two highest-ranked nuisance parameters related to the jet energy scale uncertainty sources are shown as indicated in parentheses. The pulls of the nuisance parameters (black markers) are computed relative to their pre-fit values θ_0 and uncertainties $\Delta\theta$. The impact $\Delta\hat{\mu}$ 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 post-fit uncertainty (coloured areas).

Table 5: Best fit value of the signal strength modifier μ and the corresponding observed (obs) and expected (exp) significance in standard deviations in the fully-hadronic (FH), single-lepton (SL), and dilepton (DL) channels and in the channel combination.

	$\hat{\mu} \pm \text{tot}\,(\pm\text{stat} \pm \text{syst})$	significance obs (exp)
FH 3 b-tags	$1.36^{+3.57}_{-5.36}\ \left(^{+1.68\ +3.15}_{-1.69\ -5.09}\right)$	$0.3\,\sigma$ $(0.2\,\sigma)$
FH 4 b-tags	$-1.54^{+1.41}_{-1.45}\ \left(^{+0.91\ +1.08}_{-0.90\ -1.13}\right)$	— $(0.7\,\sigma)$
FH combined	$-1.69^{+1.43}_{-1.47}\ \left(^{+0.83\ +1.16}_{-0.83\ -1.22}\right)$	— $(0.7\,\sigma)$
SL 4 jets	$1.73^{+2.25}_{-2.21}\ \left(^{+0.88\ +2.07}_{-0.87\ -2.04}\right)$	$0.8\,\sigma$ $(0.5\,\sigma)$
SL 5 jets	$0.73^{+0.98}_{-0.97}\ \left(^{+0.47\ +0.86}_{-0.46\ -0.86}\right)$	$0.8\,\sigma$ $(1.0\,\sigma)$
SL $\geq$ 6 jets	$2.05^{+0.76}_{-0.69}\ \left(^{+0.31\ +0.69}_{-0.31\ -0.62}\right)$	$3.0\,\sigma$ $(1.6\,\sigma)$
SL combined	$1.84^{+0.62}_{-0.56}\ \left(^{+0.26\ +0.56}_{-0.26\ -0.50}\right)$	$3.3\,\sigma$ $(1.9\,\sigma)$
DL 3 jets	$-2.35^{+4.40}_{-2.65}\ \left(^{+2.13\ +3.85}_{-2.06\ -1.66}\right)$	— $(0.2\,\sigma)$
DL $\geq$ 4 jets	$1.57^{+1.02}_{-0.98}\ \left(^{+0.55\ +0.86}_{-0.53\ -0.82}\right)$	$1.6\,\sigma$ $(1.0\,\sigma)$
DL combined	$1.62^{+0.90}_{-0.85}\ \left(^{+0.50\ +0.76}_{-0.48\ -0.70}\right)$	$1.9\,\sigma$ $(1.2\,\sigma)$
FH+SL+DL combined	$1.49^{+0.44}_{-0.40}\ \left(^{+0.21\ +0.39}_{-0.20\ -0.35}\right)$	$3.7\,\sigma$ $(2.6\,\sigma)$
FH+SL+DL combined 2016+2017	$1.15^{+0.32}_{-0.29}\ \left(^{+0.15\ +0.28}_{-0.15\ -0.25}\right)$	$3.9\,\sigma$ $(3.5\,\sigma)$

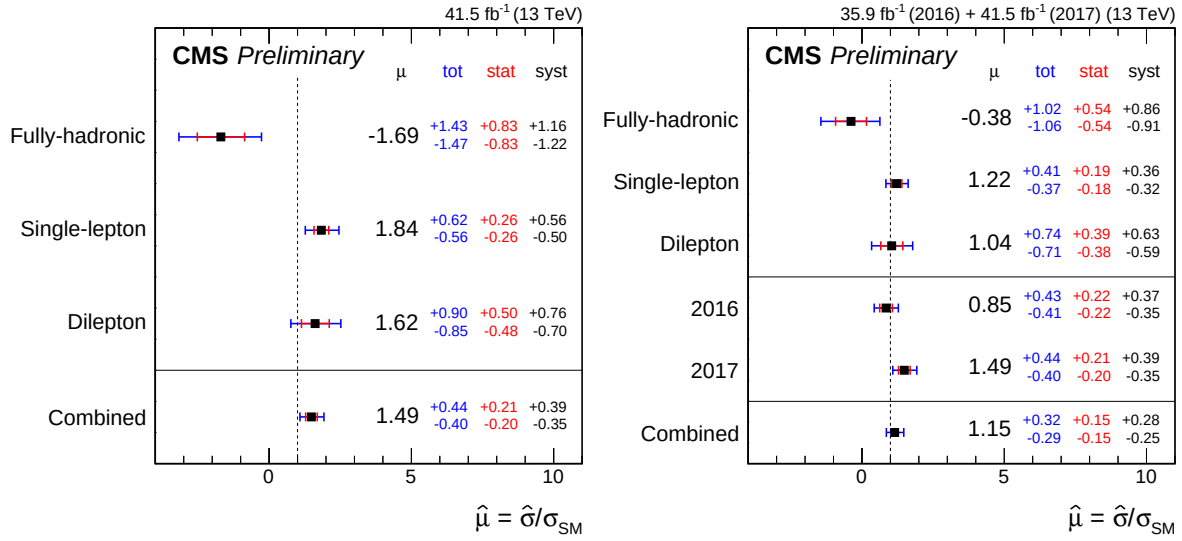

Figure 7: Best fit values of the signal strength modifiers μ obtained in the fit of the 2017 dataset (left) and in the combined fit of the 2016 and 2017 datasets (right) per channel and dataset and in the full combination. Also shown are the 68% expected confidence intervals (outer error bar), also split into their statistical (inner error bar) and systematic components.

The results are combined at the level of the profile likelihood with previous results obtained with 35.9 fb^{-1} of 2016 data [5, 6]. Several systematic uncertainties arise from theory calculations that are used in both analyses, and these are treated as fully correlated in the fit. This is the case for the uncertainties related to the inclusive cross-section calculations used to normalise the signal and background contributions, the renormalisation and factorisation scale uncertainties in the ME generators, as well as the pileup reweighting uncertainty. Uncertainties related to the PDF set used at sample generation as well as to the underlying event tune and matching scale are treated as uncorrelated because they have been updated for the 2017 analysis. Since the tune and matching scale parameters are different, also the additional 50% rate uncertainties per $\mathrm{t\bar{t}}$+hf process are treated as uncorrelated. All experimental systematic uncertainties are treated as uncorrelated because they arise to a large extent from the statistical uncertainties in independent control-sample measurements, with the exception of several jet-energy-scale related uncertainties, which are treated as fully correlated. Different choices of the correlation scheme have been tested on simulated data and found to have a negligible impact on the final expected sensitivity (less than 3%). The best fit values and impacts of the 20 nuisance parameters ranked highest in impact are presented in Fig. 8. The nuisance parameters with the highest impact are related to the uncertainty in the QCD-multijet background prediction, the $\mathrm{t\bar{t}}$+hf and signal cross sections, as well as in the b tagging scale factors.

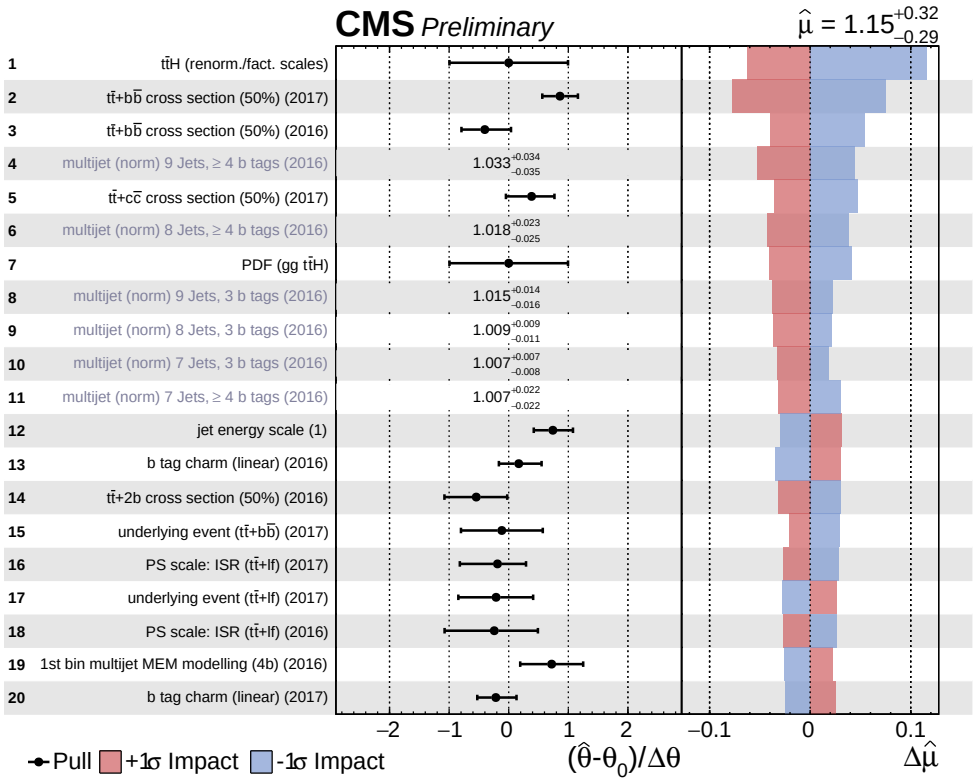

Figure 8: Post-fit pull of the constrained (text in black) and unconstrained (text in grey) nuisance parameters included in the fit to the 2016 plus 2017 data as well as their impact on the signal strength μ, ordered by their impact. Only the 20 highest ranked parameters are shown. The pulls of the nuisance parameters (black markers) are computed relative to their pre-fit values θ_0 and uncertainties $\Delta\theta$. The impact $\Delta\hat{\mu}$ 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 post-fit uncertainty (coloured areas).

The best fit value of μ in the combined 2016 and 2017 datasets is $\hat{\mu} = 1.15\,^{+0.15}_{-0.15}\,(\text{stat})\,^{+0.28}_{-0.25}\,(\text{syst})$

with a total uncertainty of $+0.32/-0.29$ (at an expected uncertainty of $+0.31/-0.29$). This corresponds to an observed (expected) significance of 3.9 (3.5) standard deviations above the background-only hypothesis. The best fit value is also listed in Table 5 and shown in Fig. 7 (right). The p value of the combined fit is $p = 0.68$.

The contributions of the statistical and various systematic uncertainties to the uncertainty in $\hat{\mu}$ are listed in Table 6. The statistical uncertainty is evaluated by fixing all nuisance parameters to their post-fit values. The impact of the systematic uncertainties is evaluated by repeating the fit fixing only the nuisance parameters related to the uncertainty under scrutiny to their post-fit values and subtracting the obtained uncertainty in quadrature from the total uncertainty of the fit where no parameters are fixed. The total uncertainty of the full fit ($^{+0.32}_{-0.29}$) is different from the quadratic sum of the listed contributions because of correlations between the nuisance parameters.

Table 6: Contributions of different sources of uncertainties to the result for the combined fit to the 2016 and 2017 datasets. The quoted uncertainties $\Delta\hat{\mu}$ in $\hat{\mu}$ are obtained by fixing the listed sources of uncertainties to their post-fit 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 post-fit values. The quadratic sum of the contributions is different from the total uncertainty because of correlations between the nuisance parameters.

Uncertainty source	$\Delta\hat{\mu}$
Total experimental	$+0.15/-0.13$
b tagging	$+0.08/-0.07$
jet energy scale and resolution	$+0.05/-0.04$
Total theory	$+0.23/-0.19$
signal	$+0.15/-0.06$
$\mathrm{t\bar{t}}$+hf modelling	$+0.14/-0.15$
QCD background prediction	$+0.10/-0.08$
Size of simulated samples	$+0.10/-0.10$
Total systematic	$+0.28/-0.25$
Statistical	$+0.15/-0.15$
Total	$+0.32/-0.29$

The total uncertainty of $+0.32/-0.29$ is dominated by contributions from systematic effects, while the statistical component is $+0.15/-0.15$. The largest contributions originate from the theoretical uncertainties, amounting to $+0.23/-0.19$, where the $\mathrm{t\bar{t}}$+hf modelling uncertainties have a major contribution. Uncertainties due to the QCD multijet background estimation contribute with $+0.10/-0.08$. Experimental uncertainties amount to $+0.15/-0.13$. 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 $+0.10/-0.10$.

8 Summary

A measurement of the associated production of a Higgs boson and a top quark-antiquark pair ($t\bar{t}H$) in the $b\bar{b}$ final state of the Higgs boson has been presented. All decay channels of the $t\bar{t}$ system are considered.

The analysis has been performed in 41.5 fb^{-1} of pp collision data recorded with the CMS detector at a centre-of-mass energy of 13 TeV in 2017. Candidate events are selected in mutually exclusive categories according to the $t\bar{t}$ decay channel and jet multiplicity. Multivariate discriminants are used to further categorise the events and to separate the $t\bar{t}H$ signal from the $t\bar{t}$-dominated background contributions. The signal is extracted in a simultaneous fit of the classifier distributions to the data across all categories and channels.

The best fit value of the $t\bar{t}H$ signal cross-section on the 2017 dataset is $\hat{\mu} = 1.49\,^{+0.21}_{-0.20}\,(\text{stat})\,^{+0.39}_{-0.35}\,(\text{syst})$ relative to the SM expectation, corresponding to an observed (expected) significance of 3.7 (2.6) standard deviations above the background-only hypothesis. Combined with previous results obtained with 35.9 fb^{-1} of data recorded in 2016, a best-fit value of $\hat{\mu} = 1.15\,^{+0.15}_{-0.15}\,(\text{stat})\,^{+0.28}_{-0.25}\,(\text{syst})$ is found, corresponding to an observed (expected) significance of 3.9 (3.5) standard deviations above the background-only hypothesis.

The presented result, which improves on previous CMS measurements in this channel owing to the increase in integrated luminosity and the usage of a more performant b tagging algorithm as well as refined analysis methods, constitutes the first evidence for $t\bar{t}H$ production in the $b\bar{b}$ decay mode of the Higgs boson.

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.* **B784** (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, 10.1130/PhysRevLett.120.231801, arXiv:1804.02610.

[3] LHC Higgs Cross Section Working Group, "Handbook of LHC Higgs cross sections: 4. deciphering the nature of the Higgs sector", *CERN* (2016) doi:10.23731/CYRM-2017-002, arXiv:1610.07922.

[4] ATLAS Collaboration, "Search for the standard model Higgs boson produced in association with top quarks and decaying into a $b\bar{b}$ pair in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector", *Phys. Rev.* **D97** (2018) 072016, doi:10.1103/PhysRevD.97.072016, arXiv:1712.08895.

[5] CMS Collaboration, "Search for $t\bar{t}H$ production in the $H \to b\bar{b}$ decay channel with leptonic $t\bar{t}$ decays in proton-proton collisions at $\sqrt{s} = 13$ TeV", arXiv:1804.03682.

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

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

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] T. Sjostrand 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.

[14] 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.

[15] CMS Collaboration, "Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements", (2019). arXiv:1903.12179. Submitted to EPJC.

[16] 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].

[17] 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.

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

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] 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.

[30] 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.

[31] 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.

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

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

[34] 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.

[35] 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.

[36] 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.

[37] 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.

[38] 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.

[39] 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.

[40] CMS Collaboration, "Performance of b tagging algorithms in proton-proton collisions at 13 TeV with Phase 1 CMS detector", CMS Detector Performance Report CMS-DP-2018-033, CERN, 2018.

[41] CMS Collaboration, "Performance of quark/gluon discrimination in 8 TeV pp data", CMS Physics Analysis Summary CMS-PAS-JME-13-002, CERN, 2013.

[42] CMS Collaboration, "Performance of quark/gluon discrimination in 13 TeV data", CMS Detector Performance Report CMS-DP-2016-070, CERN, 2016.

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

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

[45] A. Höcker et al., "TMVA: Toolkit for multivariate data analysis", *PoS* **ACAT** (2007) 040, arXiv:physics/0703039.

[46] J. Kennedy and R. Eberhart, "Particle swarm optimization", in *Proceedings of the IEEE International Conference on neural networks*, volume 4, p. 1942. 1995. doi:10.1109/ICNN.1995.488968.

[47] K. El Morabit, "A study of the multivariate analysis of Higgs boson production in association with a top quark-antiquark pair in the boosted regime at the CMS experiment", Master's thesis, Karlsruher Institut für Technologie (KIT), 2015. EKP-2016-00035.

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

[49] ATLAS Collaboration, "Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV with the ATLAS detector at the LHC", *Phys. Rev. Lett.* **117** (2016) 182002, doi:10.1103/PhysRevLett.117.182002, arXiv:1606.02625.

[50] 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.

[51] CMS Collaboration, "Investigations of the impact of the parton shower tuning in Pythia 8 in the modelling of $\mathrm{t\bar{t}}$ at $\sqrt{s} = 8$ and 13 TeV", CMS Physics Analysis Summary CMS-PAS-TOP-16-021, CERN, 2016.

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

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

[54] J. S. Conway, "Incorporating nuisance parameters in likelihoods for multisource spectra", in *Proceedings, PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding, CERN,Geneva, Switzerland 17-20 January 2011*. 2011. arXiv:1103.0354.