Search for associated production of Higgs bosons and top quarks in multilepton final states at $\sqrt{s} = 13$ TeV

The CMS Collaboration

Abstract

We present a search for the associated production of a standard model Higgs boson and a top quark-anti quark pair ($t\bar{t}H$), using LHC pp collision data collected by the CMS experiment at a center of mass energy of $\sqrt{s} = 13$ TeV in 2016. The dataset corresponds to an integrated luminosity of 12.9 fb$^{-1}$. The analysis uses events with two leptons of the same charge or at least three charged leptons, produced together with b jets, targeting Higgs boson decay modes to $WW^*$, $ZZ^*$, and $\tau\tau$ and leptonic decays of at least one of the top quarks. The results are combined with the 2015 dataset and yield a $t\bar{t}H$ signal strength of $2.0^{+0.8}_{-0.7}$ times the standard model prediction. They are used to set a 95% confidence level upper limit on the signal production cross section of 3.4 times the standard model expectation, compared to an expected upper limit of $1.3^{+0.6}_{-0.4}$ in the absence of a signal.
1 Introduction

The first extended running period of the LHC brought about the anticipated discovery of the Higgs boson by the CMS and ATLAS collaborations [1, 2], the culmination of a decades-long effort. Since then, the two experiments have measured nearly all its accessible properties—its production and decay rates, its mass, its couplings to most other standard model (SM) particles—and set constraints on others [3–6]. So far, all of these have been in agreement with the predictions of the standard model. The currently ongoing second LHC run at an increased center-of-mass energy of $\sqrt{s} = 13$ TeV is now making an even larger sample of Higgs boson events available for analysis. These data will not only allow an increase in precision, but will open up new channels to study the interactions of Higgs bosons and standard model particles. Furthermore, any deviation from the standard model expectations is a potential indication for the presence of previously-unknown particles or interactions.

One channel becoming newly accessible is the associated production of top quarks and Higgs bosons—in particular the production of a pair of top quarks with a Higgs ($t\bar{t}H$). As the mass of the Higgs boson is far too small to allow a decay to a pair of top quarks, the $t\bar{t}H$ interaction vertex can only be studied in its production. Current constraints on the strength of the top-Higgs coupling—a particularly relevant parameter in the evolution of the Higgs potential towards the Planck scale due to the large mass of the top quark—have been derived from the study of the loop-induced production of Higgs bosons in gluon-gluon fusion, where virtual top quarks are the dominant contribution in the loop. Ambiguities remain, however, in particular in light of possible unknown contributions in the gluon fusion loop, and a direct observation of top-Higgs interactions using processes with both particles in the final state would provide a strong consistency check to the theoretical model.

Both ATLAS and CMS have used the 7 and 8 TeV LHC data and the first set of available 13 TeV data to search for $t\bar{t}H$ production, but were not expected to establish evidence for it with the available statistical power. Nevertheless, an accuracy of about 30% on the coupling strength was reached by analyzing various topologies [7–10].

This note reports the search for $t\bar{t}H$ production in final states with multiple leptons, targeting Higgs decays into pairs of W or Z bosons or $\tau$ leptons with subsequent leptonic decays, and additional leptons from top quark decays. The analysis uses events with either two electrons or muons of the same charge sign, or with three or more electrons or muons. The first data of the 2016 run of the LHC is used, corresponding to an integrated luminosity of 12.9 fb$^{-1}$ at a center-of-mass energy of 13 TeV, where the expected production cross section of $t\bar{t}H$ is increased by a factor of about 4, compared to 8 TeV. The dominant background processes ($t\bar{t}W$, $t\bar{t}Z$, $t\bar{t} +$jets) see a somewhat less pronounced increase of their production cross section.

The analysis is designed to efficiently identify and select prompt leptons from on-shell W and Z boson decays and to reject non-prompt leptons from b quark decays and spurious lepton signatures from hadronic jets. Events are then selected in the various lepton channels, and are required to contain hadronic jets, some of which must be consistent with b quark hadronization. Finally, the signal yield is extracted by simultaneously fitting the output of two dedicated multivariate discriminants (trained to separate the $t\bar{t}H$ signal from the two dominant backgrounds) in all categories.

With respect to the 8 TeV analysis, the object selections have been adjusted to the updated LHC running conditions at 13 TeV, the lepton identification has been improved, and new kinematic event-observables have been included in the multivariate algorithm used for the signal extraction. Additionally, output weights from a matrix element method (MEM) are used as discrim-
inating variables in the signal extraction training against the \( t\bar{t}W \) and \( t\bar{t}Z \) backgrounds. As in the previous results based on 2015 data [11], a separate category for events with hadronically-decaying \( \tau \) leptons has also been included.

## 2 Object reconstruction and identification

Events are reconstructed using a particle-flow (PF) algorithm that optimally combines the information from all sub-detectors to identify individual particles and provide a global interpretation of the event [12, 13]. Particle flow candidates are classified into charged hadrons, neutral hadrons, photons, muons, and electrons.

Hadronic jets are reconstructed by clustering PF candidates in the anti-\( k_T \) algorithm with a distance parameter of 0.4, as implemented in the \textsc{fastjet} package [14, 15]. Charged hadrons that are not consistent with the selected primary interaction vertex are discarded from the clustering. The jet energy is then corrected for the varying response of the detector as a function of transverse momentum \( (p_T) \) and pseudorapidity \( (\eta) \) [16]. Jets are selected for use in the analysis only if they have \( p_T > 25 \text{ GeV}, |\eta| < 2.4 \), and are separated from any selected leptons by \( \Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} > 0.4 \). To select jets that are likely to have originated from the hadronization of a b quark, they are analyzed in a multivariate likelihood discriminant using track-based lifetime information and reconstructed secondary vertices (“combined secondary vertex” (CSV) algorithm) [17]. The efficiency to correctly tag b jets and the probability to misidentify jets from light quarks or gluons are measured in data as a function of the jet \( p_T \) and \( \eta \), and are used to correct for differences in the performance of the algorithm in simulated events. Two working points based on the algorithm output are used: “loose”, with a b signal tagging efficiency of about 83% and a mistagging rate of about 8%; and “medium”, with b efficiency of about 69% and mistagging rate of order 1% [18]. Tagging efficiencies for jets from charm quarks are about 40% for the loose, and about 18% for the medium working point.

In each event, the missing transverse momentum (\( E_{\text{miss}}^T \)) is calculated as the magnitude of the negative vector sum of the transverse momenta of all reconstructed PF candidates. The \( H_{\text{miss}}^T \) variable is defined as the magnitude of the negative vector sum of the transverse momenta of all selected leptons and jets in the event, i.e. it neglects unclustered energy deposits in the detector, making it robust against the pp collision remnants (underlying events) and additional pp interactions in a bunch crossing (pileup).

Muon candidates are reconstructed by combining information from the silicon tracker and the outer muon spectrometer of CMS in a global fit [19]. The quality of the geometrical matching between the individual measurements in the tracker and the muon system is used to discriminate genuine prompt muons from hadrons punching through the calorimetry and from muons produced in in-flight decays of kaons and pions. In the analysis, muon candidates are considered if they have \( p_T > 5 \text{ GeV} \) and \( |\eta| < 2.4 \). In the same-sign dilepton event categories, the relative uncertainty in the muon \( p_T \) from the fit is required to be better than 20% to ensure a high quality charge measurement.

Electrons are reconstructed using information from the tracker and from the electromagnetic calorimeter [20]. Genuine electrons are identified by a multivariate algorithm using the shape of the calorimetric shower and the quality of the reconstructed track. Furthermore, to reject electrons produced in photon conversions, candidates with missing hits in the innermost tracking layers or matched to a conversion secondary vertex are discarded. Electrons are selected for the analysis if they have \( p_T > 7 \text{ GeV} \) and \( |\eta| < 2.5 \). To suppress electrons with a mis-assigned electric charge in the same-sign dilepton categories, candidates are required to have consistent
charge measurements from three independent observables based on the calorimetry energy deposits and the track curvature.

Finally, hadronically decaying $\tau$ leptons ($\tau_h$) are reconstructed using the “hadron-plus-strip” algorithm [21]. Candidates are required to pass the “decay mode finding” discriminator and to be reconstructed either in the 1- or 3-prong decay modes, with or without additional neutral pions. They are selected for the analysis with $p_T > 20$ GeV, $|\eta| < 2.3$, and passing isolation requirements.

Electrons and muons passing the criteria described above are referred to as “loose leptons” in the following. A further discrimination between prompt signal leptons (i.e. from on-shell W and Z boson decays and from leptonic $\tau$ decays) and non-prompt and spurious leptons from b hadron decays, decays-in-flight, and photon conversions is crucial in light of the overwhelming background from t$\bar{t}$ production. To maximally exploit the available information in each event to that end, a multivariate discriminator based on a boosted decision tree (BDT) algorithm is built, taking as input not just observables related directly to the reconstructed leptons themselves, but also to the clustered energy deposits and charged particles in a cone around the lepton direction. The jet reconstruction and b-tagging algorithms are run on these, and their output is used to train the algorithm. In particular, the ratio between the lepton $p_T$ and the reconstructed jet $p_T$, and the transverse momentum of the lepton with respect to the jet axis provide good separation power in addition to more traditional observables like the relative isolation of the lepton (calculated in a variable cone size depending on the lepton $p_T$ [22, 23]), and the impact parameters of the lepton trajectory. The BDT algorithm is trained on prompt leptons in simulated t$\bar{t}$H signal and non-prompt leptons in t$\bar{t}$ background events and validated using data in various control regions. Leptons are then selected for the final analysis if they pass a given threshold of the BDT output, and are referred to as “tight leptons” in the following.

3 Event selection

This analysis is based on the 2016 run of the LHC at a center-of-mass energy of 13 TeV, and corresponds to an integrated luminosity of 12.9 fb$^{-1}$. Events are selected at the trigger level to contain either one, two, or three charged leptons (electrons or muons). The minimal transverse momentum thresholds for the leading leptons are 22 GeV for muons and 27 GeV for electrons for single lepton triggers. For double lepton triggers, the momentum thresholds on leading and sub-leading leptons are 17 and 8 GeV for muons and 23 and 12 GeV for electrons. The three lepton triggers apply a threshold on the third hardest lepton in the event of 5 and 9 GeV for muons and electrons, respectively.

At the analysis level, events are selected to reflect the final state signature of the signal process: two opposite sign W bosons and two b quark jets from the top quark decays in addition to the Higgs decay products. In case of a $H \rightarrow WW^*$ decay we therefore expect four W bosons (two pairs of opposite sign) and two b quarks in the final state. For the same-sign dilepton channels (2LSS) the signature comprises the two charged leptons (and their partner neutrinos), two hadronically decaying W bosons leading to four light quark jets, and the two b quark jets (see also Fig. 1). In the three and four lepton channels (3L), one or both of the two remaining W bosons will yield another pair of charged lepton and neutrino, reducing the number of light quark jets in the events.

To be included in the analysis, each event is required to contain at least two leptons with $p_T > 10$ GeV passing the tight requirements detailed above, of which the leading lepton must have $p_T > 25$ GeV. Events where a pair of leptons passing the loose selection has an invariant
mass of 12 GeV or less are rejected as they are not well-modeled by the simulation and do not significantly contribute to the signal acceptance. Furthermore, events are required to contain at least two hadronic jets, of which either two are to pass the loose working point of the b-tagging algorithm or at least one is to pass the medium working point.

**Same-sign dilepton channel**  
Events with exactly two leptons passing the tight criteria of which the leading lepton has $p_T > 25$ GeV and the sub-leading one has $p_T > 10$ GeV, and where they have the same charge, fall into the same-sign dilepton (2LSS) category of the analysis. In case the sub-leading lepton is an electron the $p_T$ threshold is increased to 15 GeV to comply with the trigger-level requirements. At least four hadronic jets are required for events in this category, which are then split into the three possible lepton flavor channels: ee, $\mu\mu$, or $e\mu$. To reduce backgrounds with mis-measured lepton charges and electrons from photon conversion, the two leptons are required to pass conversion rejection criteria and to have a well measured electric charge.

Even so, in the ee channel, a large contribution from charge mis-measured $Z \rightarrow ee$ events remains. To further suppress it, events in this channel are rejected if their dilepton invariant mass is within 10 GeV of the $Z$ boson mass. Finally, a minimum amount of missing momentum in the events is required to reject the neutrino-less $Z$ boson events. The applied criteria is for $0.6 \times E^{\text{miss}}_T + 0.4 \times H^{\text{miss}}_T$ to be larger than 30 GeV, which has a comparable signal efficiency to a simple $E^{\text{miss}}_T > 25$ GeV requirement, but was found to reject roughly a factor two more $Z$ background events.

**Three lepton channel**  
If the event contains more than two tight leptons, each with a $p_T$ greater than 10 GeV, the event falls into the three-lepton category. A large contribution from background processes containing $Z$ bosons is reduced by again rejecting events where a dilepton pair has invariant mass within 10 GeV of the $Z$ boson mass, and by requiring $0.6 \times E^{\text{miss}}_T + 0.4 \times H^{\text{miss}}_T > 30$ GeV. That threshold is tightened to 45 GeV in case the event contains a pair of opposite-sign and same-flavor leptons, but is not applied at all if the event contains four jets or more. Finally, events are rejected if any of the first three leptons do not pass the conversion rejection criteria.

The expected yields for these event selection categories for each process, and the observed yields in the data are shown in Tab. 1. Backgrounds from non-prompt leptons and charge
mis-measurements are estimated from the data itself, as described in Sec. 4.4. The Monte-Carlo (MC) tools used in the simulation of these processes, and the corrections applied to the generated event samples are discussed in the same section.

Table 1: Expected and observed yields after the selection in 2LSS and 3L final states. The rare SM backgrounds include $W^\pm W^\pm q\bar{q}'$, WW produced in double-parton interactions, and triboson production. Uncertainties are purely statistical. The backgrounds from non-prompt leptons and charge mis-measurements are extracted from data.

<table>
<thead>
<tr>
<th></th>
<th>$\mu\mu$</th>
<th>$ee$</th>
<th>$e\mu$</th>
<th>$3\ell$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t\bar{t}W$</td>
<td>18.3 ± 0.9</td>
<td>6.8 ± 0.6</td>
<td>24.5 ± 1.1</td>
<td>12.2 ± 0.7</td>
</tr>
<tr>
<td>$t\bar{t}Z/\gamma^*$</td>
<td>5.8 ± 0.6</td>
<td>7.4 ± 0.6</td>
<td>15.3 ± 1.3</td>
<td>22.6 ± 1.0</td>
</tr>
<tr>
<td>Di-boson</td>
<td>1.4 ± 0.2</td>
<td>1.1 ± 0.2</td>
<td>2.6 ± 0.3</td>
<td>5.7 ± 0.4</td>
</tr>
<tr>
<td>$t\bar{t}t$</td>
<td>0.8 ± 0.2</td>
<td>0.4 ± 0.1</td>
<td>1.5 ± 0.2</td>
<td>1.2 ± 0.1</td>
</tr>
<tr>
<td>$tqZ$</td>
<td>0.2 ± 0.3</td>
<td>0.4 ± 0.4</td>
<td>0.6 ± 0.6</td>
<td>2.7 ± 0.8</td>
</tr>
<tr>
<td>Rare SM bkg.</td>
<td>1.6 ± 0.3</td>
<td>0.5 ± 0.1</td>
<td>1.8 ± 0.1</td>
<td>0.3 ± 0.1</td>
</tr>
<tr>
<td>Charge mis-meas.</td>
<td>6.7 ± 0.1</td>
<td>10.0 ± 0.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-prompt leptons</td>
<td>33.4 ± 1.2</td>
<td>23.1 ± 1.1</td>
<td>61.9 ± 1.7</td>
<td>51.0 ± 1.8</td>
</tr>
<tr>
<td>All backgrounds</td>
<td>61.5 ± 1.7</td>
<td>46.4 ± 1.5</td>
<td>118.0 ± 2.5</td>
<td>95.7 ± 2.3</td>
</tr>
<tr>
<td>$t\bar{t}H (H \rightarrow WW^*)$</td>
<td>6.3 ± 0.2</td>
<td>2.6 ± 0.1</td>
<td>8.5 ± 0.2</td>
<td>8.0 ± 0.2</td>
</tr>
<tr>
<td>$t\bar{t}H (H \rightarrow \tau\tau)$</td>
<td>1.6 ± 0.1</td>
<td>0.7 ± 0.1</td>
<td>2.5 ± 0.1</td>
<td>2.1 ± 0.1</td>
</tr>
<tr>
<td>$t\bar{t}H (H \rightarrow ZZ^*)$</td>
<td>0.2 ± 0.0</td>
<td>0.1 ± 0.0</td>
<td>0.3 ± 0.0</td>
<td>0.5 ± 0.0</td>
</tr>
<tr>
<td>Data</td>
<td>74</td>
<td>45</td>
<td>154</td>
<td>105</td>
</tr>
</tbody>
</table>

Signal extraction procedure  
After the selection of events based on the expected signal topology as described above, the event yields are still heavily dominated by background processes with similar final states and larger cross sections. The main backgrounds are twofold: $t\bar{t}W$ and $t\bar{t}Z$ production that have matching numbers of prompt leptons and b quarks in the final state; and $t\bar{t}$ events contributing due to non-prompt leptons. To separate the signal contribution from these two background classes, two boosted decision tree classifiers are trained, one for each category of backgrounds, exploiting further topological and kinematic differences between the signal and backgrounds. The two classifiers are trained using simulated events, separately in the 2LSS and 3L event categories and use the following input distributions:

- the maximum $|\eta|$ of the two leading leptons;
- the jet multiplicity;
- the minimum distances between the leading and trailing leptons and their closest jet;
- the transverse mass of the leading lepton and the $E_T^{miss}$;
- the $E_T^{miss}$ (or $H_T^{miss}$ in the 3L category, for training against $t\bar{t}$) or the leading lepton $p_T$ (for training against $t\bar{t}W/t\bar{t}Z$);
- the average separation between the two jets (against $t\bar{t}$) or the trailing lepton $p_T$ (against $t\bar{t}W/t\bar{t}Z$);
- and the logarithm of the three MEM weights for $t\bar{t}H$, $t\bar{t}W$, and $t\bar{t}Z$ (against $t\bar{t}W/t\bar{t}Z$ in the 3L category only).

For the 3L event category only, a matrix element weight $w_{i,\alpha}$ is computed for each hypothesis $\alpha$ (where $\alpha$ is either $t\bar{t}H$, $t\bar{t}W$, or $t\bar{t}Z$) and for the event $i$ as follows:

$$w_{i,\alpha}(\Phi') = \frac{1}{\sigma_\alpha} \int d\Phi_\alpha \cdot \delta^4\left(p_1^\mu + p_2^\mu - \sum_{k \geq 2} p_k^\mu\right) \cdot \frac{f(x_1, \mu_F) f(x_2, \mu_F)}{x_1 x_2 s} \cdot |M_\alpha(p_\mu)|^2 \cdot W(\Phi'|\Phi_\alpha),$$
where \( \sigma_\alpha \) is the cross section; \( \Phi' \) are the 4-momenta of the reconstructed particles; \( d\Phi_\alpha \) is the element of phase space corresponding to unmeasured quantities with momentum conservation enforced; \( f(x, \mu_F) \) are the parton density functions, computed using NNPDF2.3 LO [24]; \( |M_\alpha|^2 \) is the squared matrix element, computed with \textsc{MadGraph} 5 \textsc{Amc@Nlo} standalone [25] at LO in the narrow-width approximation for \( t, \bar{t} \) and \( H \); and \( W \) are the transfer functions for jet energy and \( E_{\text{miss}} \), relating parton to reconstructed quantities, estimated from simulated \( t\bar{t}H \) events. The two jets with the highest CSV tagging output are assigned to the two \( b \) quarks in the matrix element. Among the remaining jets, the pair with dijet mass closest to \( m_W \) is selected. (In the \( t\bar{t}H, H \rightarrow \ell\nu jj \) hypothesis, the pair with lowest dijet mass is selected.) In events where one or two jets needed to evaluate \( |M_\alpha|^2 \) fail to be reconstructed, the weight is recovered by extending the integration phase space for the missing jets. The final weight for each hypothesis \( \alpha \) is taken as the average of the weights computed for each lepton and jet permutation. Including the MEM weights in the BDT training against \( t\bar{t}W/t\bar{t}Z \) improves the background rejection power by about 10% for the three lepton category.

Before the final signal extraction, the selected events in the three 2LSS channels and the 3L channel are then further categorized. For 2LSS events, a separate category combining all three flavor channels is defined for events with a \( \tau_h \) in the final state—enriched in \( H \rightarrow \tau\tau \) decays. Events with no \( \tau_h \) (with the exception of the ee channel), are split according to whether they do or do not pass a more stringent b-tagging requirement where both jets pass the medium b-tagging working point (“b tight” for events that do, “b loose” for the remaining). The same categorization in b tight and b loose is done for 3L events, independent of the presence of a \( \tau_h \). Finally, all resulting categories except for the \( \tau_h \) category are split according by the sum of lepton charges to exploit the charge asymmetry in the production of several background processes.

The outputs of the two BDT classifiers in each category are then used to define bins of different signal-to-background ratios. The contribution of the signal processes to the overall yields in each bin are then extracted by performing a simultaneous fit of the normalizations of each process in each category and bin. Figures 2, 3 and 4 show distributions of event observables and BDT classifier outputs for selected events in data, compared to the simulated and predicted background processes (described in Sec. 4.4).

4 Signal and background modeling and systematic uncertainties

Simulated Monte Carlo (MC) event samples are used to optimize the lepton selections (see Sec. 2), to determine the event selection efficiency, and to train the classifiers used to measure the signal contribution to the final selection. Variations of these samples are also used to estimate the effects of systematic uncertainties. Signal \( t\bar{t}H \) events are generated using the \textsc{MadGraph} 5 \textsc{Amc@Nlo} package (version 5.222) [25], which includes up to one additional hadronic jet at next-to-leading order (NLO) QCD accuracy. The same software is used for the main backgrounds: \( t\bar{t}W/t\bar{t}Z \), \( t\bar{t} + \text{jets} \), and \( t\bar{t}\gamma + \text{jets} \). Other minor backgrounds are simulated with different generators, such as \textsc{Powheg} [26–31] and \textsc{MadGraph} at leading order (LO) QCD accuracy. All generated events are interfaced to \textsc{Pythia8} (v8.205) [32] for the parton shower and hadronization steps. Pileup interactions are simulated to reflect the observed multiplicity in data. All events are finally passed through a full simulation of the CMS detector based on \textsc{Geant4} [33], and reconstructed using the same algorithms as used for the data.

Furthermore, the trigger selection is simulated and applied for generated signal events. Residual differences in the trigger efficiency between data and MC are studied and corrected for, using the measured trigger efficiencies of the data. Background samples without a simulation
The NLO calculation for the inclusive production cross section of the t\bar{t}H signal has uncertain-

of the trigger selection are corrected for the measured trigger inefficiency.

4.1 Signal model

Systematic uncertainties on the signal selection efficiency arise on the one hand from correction factors applied to the simulated events to better match the measured detector performance and on the other hand from theoretical uncertainties in the modeling of the signal process. Scale factors applied to correct for data/MC differences in the trigger efficiency, lepton reconstruction and identification performance, and lepton selection efficiency carry a combined uncertainty of about 5% per lepton. The impact of the uncertainty in the signal selection efficiency from jet energy corrections are evaluated by varying the correction factors within their uncertainty and propagating the effect to the final result by recalculating all kinematic quantities. Both effects on the overall normalization of event yields and on the shape of kinematic properties are taken into account. Jet energy resolution effect have negligible impact on this analysis. Correction factors for data/MC differences in the b-tagging performance are applied depending on the \( p_T \) and \( \eta \), and on the flavor of the jet, and their effect on the signal efficiency is evaluated by varying the factors within their measured uncertainty and recalculating the overall event scale factors.

The NLO calculation for the inclusive production cross section of the t\bar{t}H signal has uncertain-
of 2 to 4%, and an uncertainty in the shape of the classifier outputs is estimated by varying the renormalization and factorization scales in the final normalization of the signal yields. Additionally, an uncertainty in the shape of the final classifier output is estimated by varying the renormalization and factorization scales in generated events, and has an impact of about 2–3% on the normalization.

4.2 \( t\bar{t} + \text{vector boson backgrounds} \)

Backgrounds from associated production of \( t\bar{t} \) pairs and vector bosons (\( t\bar{t}W \) and \( t\bar{t}Z \)) are estimated directly from simulated events—corrected for data/MC differences and inefficiencies in the same way as signal events. Their production cross sections are calculated at NLO QCD and EWK, with theoretical uncertainties from unknown higher orders of 12% and 10%, respectively [34]. As for the signal, a further uncertainty arises from the knowledge of PDFs and \( \alpha_S \) of 2 to 4%, and an uncertainty in the shape of the classifier outputs is estimated by varying renormalization and factorization scales, leading to a variation of the classifier output shape of about 2 to 4% in amplitude.

Figure 3: Number of selected jets and distributions of the BDT classifier outputs for the three-lepton channel. Distributions are shown before the signal extraction fit.

Figure 4: Combination of the BDT classifier outputs in the bins used for signal extraction, for the same-sign dilepton (left) and three-lepton (right) channels. Post-fit distributions and uncertainties are shown.
4.3 WZ and ZZ backgrounds

Diboson production with leptonic Z decays and additional jet radiation in the final state can lead to signatures very similar to that of the tH signal. Due to the larger cross section, the main contribution arises from WZ production. Inclusive production cross sections for both WZ and ZZ have been measured at the LHC and agree well with the NLO calculations. However, the good agreement in the inclusive phase space of the cross section measurements does not necessarily hold in the signal region of this analysis which requires the presence of hadronic jets, including b jets. Therefore, a dedicated control region dominated by WZ production is used to constrain the overall normalization of this process. It is defined by the presence of at least three leptons, of which one opposite-sign pair is to be compatible with a Z boson decay. Furthermore, at least two jets are required, with a veto on jets that pass the loose b tag selection to ensure exclusivity with the signal selection. A scale factor is then extracted from the predicted distribution of WZ events in the control region, and the observed data, keeping other processes fixed. Finally, this factor is used to scale the diboson prediction in the signal selection.

Of diboson events passing the signal selection, the majority contains jets from light quarks and gluons that are wrongly tagged as b jets, making this estimate mainly sensitive to the experimental uncertainty in the mis-tag rate rather than the theoretical uncertainty in the jet flavor composition. The overall uncertainty assigned to the diboson prediction is estimated from the statistical uncertainty due to the limited sample size in the control region (30%), the residual background in the control region (20%), the uncertainties on the b-tagging rate (between 10–40%), and from the knowledge of PDFs and the theoretical uncertainties of the extrapolation (up to 10%).

4.4 Non-prompt and charge mis-identified leptons

The main contribution to the overall event yield in the signal selection, and one that can be reduced up to a certain point by tighter lepton selections, comes from processes with comparatively large cross sections in which one of the leptons is produced inside a jet (i.e. it is non-prompt). These are mostly real leptons from b hadron decays but also contain hadronic jets misidentified as leptons. The yield of such events is estimated from a loose-to-tight extrapolation, in which a looser lepton selection is defined and the rate at which such leptons enter the tighter selection is measured in a control region and then used to extrapolate from a sideband with loose leptons to the signal selection with tight leptons.

The probability of a non-prompt lepton candidate passing a given loose selection to also pass the tight signal requirement is measured in a sample dominated by non-prompt leptons, as a function of \( p_T \) and \(|\eta|\) and separately for muons and electrons. The definitions of loose and tight leptons are given in Sec. 2. Two event samples are defined for the measurement of tight-to-loose ratios: one dominated by QCD multijet events, collected using single lepton triggers at relatively high \( p_T \) thresholds; and one dominated by Z +jets events, where the two high \( p_T \) leptons from the Z decay can be used to trigger the events without biasing a third lepton at low transverse momentum. The QCD-dominated sample is then used to extract ratios for lepton candidates with \( p_T \) above 30 GeV, whereas the ratios for low \( p_T \) leptons are determined in the Z +jets sample. For both regions, contributions from prompt leptons, mainly from W and Z +jet or from WZ and ZZ events, respectively, are first suppressed by vetoing additional leptons in the selection, and the residual contamination is then subtracted using the transverse mass as a discriminating variable.

A sideband control region is then defined by relaxing the lepton selection criteria to “loose” (see Sec. 2), while keeping all other selections equivalent to the full signal selection. By weighting
events in this expanded selection with a factor dependent on the measured tight-to-loose ratios, a fully data-driven estimation for the contribution of non-prompt leptons to the signal selection can be obtained. In events where just one of the two leptons fails the tight criteria, the applied event weight is \( f / (1 - f) \) (where \( f \) is the tight-to-loose ratio measured as described above), while events where both leptons fail the tight criteria are weighted by \(-f_1 f_2 / [(1 - f_1)(1 - f_2)]\). The resulting prediction of the event yield in the signal selection carries an uncertainty of between 30–50%, arising from the statistical uncertainty in the measurement of the tight-to-loose ratios, and from a systematical uncertainty derived by comparing alternative methods of subtracting prompt lepton backgrounds and from testing the closure of the method in simulated background events.

Similarly, background from events where the charge of one of the leptons is wrongly assigned—relevant only in the same-sign dilepton channels—are determined by measuring the charge mis-assignment probability in a sample of same-sign dilepton event compatible with a Z boson decay and weighting events with opposite-sign leptons in the signal selection. The charge mis-assignment probability is found to be negligible for this analysis for muons, whereas for electrons it ranges from about 0.03% in the barrel section up to about 0.4% in the detector endcaps. It is measured separately in these two regions, and additionally as a function of the electron \( p_T \). A systematic uncertainty of 30% is assigned to the prediction from the statistical uncertainty of the probability measurement and from testing the performance of the method on simulated events.

## 5 Results

The signal and background event yields extracted in the simultaneous fit to the two classifier outputs are compared with the expectations for the spectrum of backgrounds and for a signal with a SM Higgs boson of 125 GeV. The relative strength of the \( t\bar{t}H \) signal with respect to the SM cross section calculated at NLO [34] is captured in a parameter \( \mu = \sigma / \sigma_{SM} \) (the signal strength). In the fit, the \( t\bar{t}H \) event yields are scaled by \( \mu \) without changing the Higgs branching fractions or the kinematic properties of the events.

The best fit signal strength on the combined categories amounts to \( 2.3^{+0.9}_{-0.8} \) times the standard model expectation, corresponding to an observed limit of \( \sigma < 3.9 \times \sigma_{SM} \) at 95% CL. The limit expected under a background-only hypothesis is \( 1.4^{+0.7}_{-0.4} \). The results are presented in Tab. 2 in terms of an asymptotic 95% confidence level (CL) upper limit on \( \mu \), and its best fit value [35–38]. When combining this result with the smaller 2015 dataset at the same center of mass energy [11], the best-fit signal strength is reduced to \( 2.0 \pm 0.4(\text{stat.})^{+0.7}_{-0.6}(\text{syst.}) \). The uncertainty on the result is currently limited by the systematic uncertainty in the estimation of backgrounds with non-prompt leptons, mainly related to limited statistics in the samples used to measure the tight/loose ratios and to the subtraction of processes with prompt leptons contaminating them. While the former is bound to improve with the addition of more data, the latter will require improvements in the analysis techniques to be significantly reduced.

The extracted signal strength in the combined result is compatible with a standard model Higgs hypothesis within 0.8\( \sigma \), while the observed significance in the context of the background-only hypothesis is 3.2\( \sigma \), with an expected significance of 1.7\( \sigma \).
Table 2: Observed and expected asymptotic 95% CL upper limits on and best fit value of the signal strength parameter ($\mu$).

<table>
<thead>
<tr>
<th>Category</th>
<th>Obs. limit</th>
<th>Exp. limit $\pm 1\sigma$</th>
<th>Best fit $\mu \pm 1\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Same-sign dileptons</td>
<td>4.6</td>
<td>$1.7^{+0.9}_{-0.5}$</td>
<td>$2.7^{+1.1}_{-1.0}$</td>
</tr>
<tr>
<td>Trileptons</td>
<td>3.7</td>
<td>$2.3^{+1.2}_{-0.7}$</td>
<td>$1.3^{+1.2}_{-1.0}$</td>
</tr>
<tr>
<td>Combined categories</td>
<td>3.9</td>
<td>$1.4^{+0.7}_{-0.4}$</td>
<td>$2.3^{+0.9}_{-0.8}$</td>
</tr>
<tr>
<td>Combined with 2015 data</td>
<td>3.4</td>
<td>$1.3^{+0.6}_{-0.4}$</td>
<td>$2.0^{+0.8}_{-0.7}$</td>
</tr>
</tbody>
</table>

6 Conclusions

A search for the associated production of a standard model Higgs boson and a top quark-anti quark pair has been performed using pp collision data collected by the CMS experiment in 2016 at a center of mass energy of $\sqrt{s} = 13$ TeV, and corresponding to an integrated luminosity of 12.9 fb$^{-1}$. The analysis targets Higgs boson decay modes to $WW^*$, $ZZ^*$, and $\tau\tau$ and leptonic decays of at least one of the top quarks. The 2016 dataset is combined with the smaller 2015 dataset at the same center of mass energy. We measure a signal strength of $\sigma/\sigma_{SM} = 2.0^{+0.8}_{-0.7}$ in the combined dataset, and set a 95% confidence level upper limit of $\sigma < 3.4 \times \sigma_{SM}$ on the signal cross section, compared to an expected upper limit of $1.3^{+0.8}_{-0.4}$ in absence of a signal.
References


